

Preliminary Report from the Project, "Prosodic Segmentation and Structuring of Dialogue"

Gösta Bruce*, Björn Granström**, Kjell Gustafson**, David House* and Paul Touati*
 *Dept of Linguistics and Phonetics **Dept of Speech Comm. and Music Acoustics
 Helgonabacken 12, S-22362 Lund KTH, Box 70014, S-10044 Stockholm
 (names in alphabetic order)

ABSTRACT

The research reported here is conducted within the recently initiated project 'Prosodic Segmentation and Structuring of Dialogue'. The object of study in the project is the prosody of dialogue in a language technology framework. The specific goal of our research is to increase our understanding of how the prosodic aspects of speech are exploited interactively in dialogue – the genuine environment for prosody – and on the basis of this increased knowledge to be able to create a more powerful prosody model. In this paper we present an overview of project design and methods.

INTRODUCTION

It is our conviction that prosody is of vital importance in the structuring of spoken dialogue. This can be attested for example by listening to the output of speech synthesizers of the early days characterized by complete monotony and absence of prosodic variation and consequently with no attempt of simulating interaction with the listener. The object of study in the project *Prosodic Segmentation and Structuring of Dialogue* is the prosody of dialogue in a language technology framework. The project represents cooperation between Phonetics at Lund University and Speech Communication at KTH, Stockholm and is part of the Swedish Language Technology Programme. Related projects within the Language Technology framework are *Intonation in Restrictive Texts: Modelling and Synthesis* (Home et al. 1993), *Interaction in Speech between Prosody, Syntax, Semantics and Pragmatics* (Strangert et al. 1993) and also *Language Technology for Spoken Dialogue Systems* (Blomberg et al. 1993).

BACKGROUND

Research within our project *Prosodic Segmentation and Structuring of Dialogue* is based on earlier work on prosody from different perspectives. One starting point is research conducted within the project *Contrastive Interactive Prosody* (KIPROS) at Lund supported by the Bank of Sweden between 1988-90. The object of study of KIPROS was dialogue prosody in a contrastive perspective in French, Greek and Swedish. We conducted three types of analysis: analysis of dialogue structure, auditory (prosodic) analysis, and acoustic-phonetic analysis. This project was our first large-scale confrontation with spontaneous speech and dialogue and comprised exploratory testing of the prosody model which was based on experience from extensive work with laboratory speech (see also Gårding 1967). The focus of the KIPROS project was largely on methodology, which resulted in the development of tools and conventions for prosodic transcription of Swedish and French (Bruce & Touati 1990, 1992). Experience from the project also made apparent the main difficulties involved in analyzing spontaneous speech where experimental control is low.

The second point of departure for the current project is work carried out within the project *Prosodic Phrasing in Swedish* which was also a joint effort between Phonetics in Lund and Speech Communication at KTH, Stockholm, and part of the Language Technology Programme 1990-93. Our cooperation relates to two different research traditions: work in Lund aimed at developing a model for Swedish prosody and work in Stockholm directed towards the development of the prosodic component of a text-to-speech system. The main orientation of this project was directed towards studying how

prosody signals phrasing, i.e. grouping of words into phrases. The Prosodic Phrasing project represented a return to the phonetics laboratory and more controlled conditions in the form of analyses of read speech (Bruce, Granström & House 1992; Bruce, Granström, Gustafson & House 1993a, 1993b).

GOAL AND METHODOLOGY

The primary goal of the new project is to increase our understanding of how the prosodic aspects of speech are exploited interactively in dialogue - the genuine environment for prosody - and on the basis of this increased knowledge to be able to create a more powerful prosody model. To be able to achieve this goal the following methodology is being employed:

- analysis of dialogue structure (independent of prosody)
- auditory analysis in the form of prosodic transcription
- acoustic-phonetic analysis (based on F0 and waveform information)

We are exploiting speech material from the national Swedish prosodic database under development. The dialogues under study cover true spontaneous conversation, spontaneous but more restricted and well controlled dialogues, as well as acted dialogues from scripts and dialogues simulated using text-to-speech synthesis.

ANALYSIS OF DIALOGUE STRUCTURE

The ultimate goal for prosody research within language technology is to be able to combine phonetic knowledge about prosody with linguistic and other contextual information. It is therefore important that the analysis of the dialogue structure itself is carried out independently of prosody. Although our method of analyzing dialogue structure is preliminary, we have been working with three basic, interactive dimensions, namely textual aspects, turn regulating aspects and aspects of initiative/response structure.

The *textual* aspect concerns division into conversation topics involving grouping into 'speech paragraphs' (Brown et al. 1980). This applies to discourse both in the form of dialogue and monologue. It is clear that prosody plays an important role in signalling topic structure, even if different studies show different types of relationships.

The *turn regulating* aspect involves e.g. taking, keeping, yielding and competing for the floor in a dialogue (cf. Cutler & Pearson 1986). It is apparent that this aspect is signalled by different means (verbal, non-verbal, prosodic). The exact contribution of prosody here is not fully understood.

Aspects of *initiative/response* structure concern the contribution of the speakers to the development of the dialogue through taking or refraining from taking initiative, responding to initiatives and making reference to what has been said. Prosody plays an important role here, although it is clear that there is a considerable degree of freedom in the way that it is used to signal this aspect of dialogue structure.

In addition to the above, there is also a *feedback* dimension, indicating the way in which speakers give and seek feedback in a dialogue. Feedback giving (backchannelling) is often noted in dialogue studies while the speaker's feedback seeking (seeking feedback from the listener) has not been given as much attention. It is possible that the feedback dimension can be seen as a subdivision of the initiative/response structure, although we have chosen to regard it as a separate dimension for the present time. We believe that prosody plays an important role in signalling both feedback giving and seeking.

Other interactive dimensions which can easily be expressed prosodically are the signaling of *attitudes/emotions* and *rhetoric activity* (Touati 1993).

AUDITORY ANALYSIS

An independent auditory analysis of prosody is made in the form of a prosodic transcription. This transcription is tied to the orthographic representation of the dialogue and thus contains symbolization of selected prosodic categories.

We have witnessed a marked increase in interest in transcription, including prosodic transcription, during the last five year period. One important reason for this newborn interest arises from new needs for annotation of large speech databases. A starting point was the 1989 IPA Convention in Kiel for the revision of the International Phonetic Alphabet, the first substantial revision in 40 years. The new version of the IPA (cf. I.P.A. 1989) does not, however, contain any specific symbolization of discourse prosody.

Another example of this transcription wave is ToBI (tones and break indices), a system which has recently been developed for the prosodic transcription of American English (Silverman et al. 1992). This transcription system provides symbols mainly for prominence and grouping. An innovation in ToBI is the combined auditory and acoustic (F0, waveform) analysis, where both types of information are integrated in the prosodic symbolization.

Unlike ToBI we have chosen to rely on a purely auditory analysis. Our transcription is intended to be phonological rather than a narrow phonetic transcription. The prosodic transcription that was developed within the KIPROS project consists of symbols for the following prosodic categories: prominence, grouping, pausing, pitch range and boundary tones. There are, however, other potentially interesting categories such as voice intensity, voice quality and speech tempo which have not been included here. In the KIPROS transcription system IPA symbols for prominence, grouping and pausing are used as well as special symbols for pitch range and boundary tones. IPA symbolization of prominence, grouping and pausing is abstract but well established and relatively simple, while pitch range and boundary tones are represented by more iconic and transparent symbols (cf. Bruce & Touati 1990).

A vital issue for the construction of a national Swedish prosody database within the Language Technology framework is the choice of prosodic transcription. Discussions of this issue have resulted in an agreement whereby a base module for prosodic transcription based on the IPA comprises a standard for the phonological symbolization of the categories prominence and grouping (see further Bruce 1994). In addition to this base module different prosody projects within Language Technology are expected to create their own modules according to existing needs. In our new Prosodic Segmentation project we intend to add a module containing symbols from the KIPROS transcription system. Moreover we have also begun development of a module for tonal analysis using notation not unlike ToBI.

ACOUSTIC-PHONETIC ANALYSIS

Our acoustic-phonetic analysis comprises standard F0 extraction and spectral information in addition to the speech waveform. The analysis is carried out in the ESFS/Waves environment which includes transcription and labeling in multiple tiers (Ayers 1994). This enables an automatic processing of possible relationships between, for example, prosodic and discourse categories.

An important part of the analysis of F0 is the intonation model which has been developed from extensive studies of laboratory speech (cf. Bruce 1977, Bruce & Granström 1993). The intonation model involves categorization with respect to accentuation (prominence) and phrasing (grouping), including boundary signalling and other intonation features. The categories are expressed using tonal turning points (H/L) with association to stressed syllables or boundaries.

ACKNOWLEDGEMENTS

This work was carried out under a contract from the Swedish Language Technology Programme (HSFR-NUTEK). Gayle Ayers, Dept of Linguistics, Ohio State University was a guest researcher in Lund (autumn 1993) and has contributed to the project.

REFERENCES

- Ayers, G. 1994. 'Discourse functions of pitch range in spontaneous and read speech'. To appear in *Working Papers from the Dept of Linguistics*. Ohio State University.
- Blomberg, M., R. Carlson, K. Elenius, B. Granström, S. Hunnicutt, R. Lindell & L. Neovius. 1993. 'An experimental dialogue system: Waxholm'. *RUUL 23, Fonetik -93*, 49-52. Department of Linguistics, Uppsala University.
- Brown, G., K. Currie & J. Kenworthy. 1980. *Questions of intonation*. London: Croom Helm.
- Bruce, G. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerups.
- Bruce, G. 1994. 'Prosodisk strukturering i dialog'. To appear in *Svenskans Beskrivning 20*. Umeå University.
- Bruce, G. & B. Granström. 1993. 'Prosodic modelling in Swedish speech synthesis'. *Speech Communication* 13, 63-73.
- Bruce, G., B. Granström & D. House. 1992. 'Prosodic phrasing in Swedish speech synthesis'. In G. Bailly, C. Benoît and T.R. Sawallis (eds.), *Talking Machines: Theories, Models, and Designs*, 113-125. Amsterdam: Elsevier Science Publishers. B.V.
- Bruce, G., B. Granström, K. Gustafson & D. House. 1993a. 'Interaction of F0 and duration in the perception of prosodic phrasing in Swedish'. In B. Granström & L. Nord (eds.), *Nordic Prosody VI*, 7-22. Stockholm: Almqvist & Wiksell International.
- Bruce, G., B. Granström, K. Gustafson & D. House. 1993b. 'Prosodic modelling of phrasing in Swedish'. *Working Papers 41, Proceedings of an ESCA workshop on prosody*, 180-183. Department of Linguistics, Lund University.
- Bruce, G. & P. Touati. 1990. 'On the analysis of prosody in spontaneous dialogue'. *Working Papers 36*, 37-55. Department of Linguistics, Lund University.
- Bruce, G. & P. Touati. 1992. 'On the analysis of prosody in spontaneous speech with exemplifications from Swedish and French'. *Speech Communication* 11, 453-458.
- Carlson, R., B. Granström & S. Hunnicutt. 1991. 'Multilingual text-to-speech development and applications'. In W. Ainsworth (ed.), *Advances in speech, hearing and language processing*, 269-296. London: JAI Press.
- Cutler, A. & M. Pearson. 1986. 'On the analysis of prosodic turn-taking cues'. In C. Johns-Lewis (ed.), *Intonation in Discourse*, 139-155. London: Croom Helm.
- Gårding, E. 1967. 'Prosodiska drag i spontant och uppläst tal'. In G. Holm (ed.), *Svenskt talspråk*, 40-85. Uppsala: Almqvist & Wiksell.
- Horne, M., M. Filipsson, C. Johansson, M. Ljungqvist & A. Lindström. 1993. 'Improving the prosody in TTS systems: Morphological and lexical-semantic methods for tracking "new" vs. "given" information'. *Working Papers 41, Proceedings of an ESCA workshop on prosody*, 208-211. Department of Linguistics, Lund University.
- I.P.A. 1989. 'Report on the 1989 Kiel convention'. *Journal of the International Phonetic Association* 19 (2), 67-80.
- Linell, P. & L. Gustavsson. 1987. *Initiativ och respons. Om dialogens dynamik, dominans och koherens*. SIC 15. Studies in Communication. University of Linköping.
- Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, P. Price, J. Pierrehumbert & J. Hirschberg. 1992. 'TOBI: A standard for labeling English prosody'. In *Proceedings of the Second International Conference on Spoken Language Processings 2*, 867-870. Banff, Canada.
- Strangert, E., E. Ejerhed & D. Huber. 1993. 'Clause structure and prosodic segmentation'. *RUUL 23, Fonetik -93*, 81-84. Department of Linguistics, Uppsala University.
- Touati, P. 1993. 'Overall pitch and direct quote-comment structure in French political rhetoric'. *RUUL 23, Fonetik -93*, 98-101. Department of Linguistics, Uppsala University.