**Automatic extraction of prosodic features and their employment in the analysis of speech corpora**

István Szekrényes

The aim of the dissertation is to contribute to the research of spoken language with new methodological tools and experiments that represent an approach to the prosody of speech not only at the level of utterances but also at that of the micro- and macrostructure of social interaction, as patterns of acts informative by themselves.

At the beginning of the dissertation, from a wider perspective, I analyze what aspects of prosody are treated by various schools of linguistics (generative grammar, speech act theory, conversation analysis), to be followed by specifying those concrete questions and hypotheses which motivated my research carried out within the frameworks of the dissertation. Before describing my own developments and experiments I give a short overview of related fields of language technology (speech detection and speaker diarization), and a summary of solutions and frameworks available for the modeling and computational processing of intonation (ToBI, Tilt, Prosogram).

The most detailed discussion in the dissertation is related to my methodology developed for the automatic annotation of prosody, whose aim it is to describe the modulations of speech melody, intensity and tempo as sequences of adjacent events characterized by categorical labels. For a technical point of view, it consisted of the stylization of the measurable values of the physical parameters of the given feature (as in the case of tone: the fundamental frequency of the speech signal), as well as the analysis of the trends resulting from the stylization, expressing the main direction of the modulation. According to my hypothesis, the resulting sequences of events discover those modulations relevant to perception which can serve as indicators for the interpretation of the attitudinal content of an utterance, the segmentation of a conversation, or, in longer samples, the characterization of the context of a given interaction. In addition, I implemented a web based visualization of the textual output of the analysis of speech melody. It was partly done for the purpose of validation so that the output tonal contours could be more easily comparable with the perceived sound experience.

I carried out a number of experiments for the validation of the hypotheses. Through them I wished to find an answer to the questions whether the subjects (1) can determine the situation of the conversation, (2) sense if the topic has changed, (3) can give a judgement about the relation between participants of the interaction, (4) can position within the conversation one of its given segments, (5) can identify the attitudinal content of an utterance. I also tested the task of the classification of the conversation situation training deep neuronal networks (DNNs) based on prosodic features included in the HuComTech corpus.