Towards the integrations of stochastic information in speech technologies: the case of suprasegmentals

Irina Nesterenko

St. Petersburg State University, Department of Phonetics, 11, University embankment, St. Petersburg, Russia
Laboratoire Parole et Langage, Univeriste de Provence, Laboratoire Parole et Langage
Université de Provence, 29 av. R. Schuman – 13621 Aix-en-Provence Cedex 1, France
Irina.Nesterenko@lpl.univ-aix.fr

Abstract

In the present study, we defend the thesis that numerous studies of extensive spontaneous speech corpora are needed to improve the performance of Natural Language Processing systems (NLP). We consider that at the present stage of research, more attention should be paid to the variability of human discourse and the probabilistic approach ([3], [4]) which will contribute significantly to the better intelligibility and naturalness of synthesized speech. In our contribution, we will address the question of automatic prosodic annotation of speech corpora; consider the problem of choosing appropriate units of description and of the automatic extraction and coding of these units.

Our theoretical assumptions are based on the work of Di Cristo et al. [2] (the multilinear approach to the intonation and discourse) and in Blache & Di Cristo [1] (the competitive approach to the generation of meaning with the primary information being provided by different linguistic components). The ideology of these approaches determines the choice of the speech-in-interaction corpus for the present study.

We develop the question of automatic prosodic annotation in its relation to the implementation of prosodic phonological models in the NLP systems. Phonological models propose a hierarchy of phonological units that determine the prosodic organization of an utterance and all of them could be used as primary description units in the technological implementation. We have tested the validity of the concept of the Intonational Unit (IU) in a perceptual experiment. The IU is defined with reference to its boundaries and its internal organisation: it is a unit that comprises at least one nuclear accent and is delimited from the rest of the utterance by the boundary tone. We have studied the pertinence of this concept via the measure of listeners’ agreement in the segmentation task. Secondly, we analysed the relationship between the subjective judgements and the objective acoustic cues in order to establish the hierarchy of the acoustic parameters that cue the perception of a IU boundary. The results suggested a very complex relationship between prosody and syntax and prosody and discourse structure.

In the next stage, we studied the concept of linguistic prominence: the prominences were coded manually by the author according to linguistic categories (unaccented syllable, accented syllable, nuclear accent). A set of acoustic parameters was extracted from the corpus and subjected to discriminant analysis. We studied the ranking of acoustic parameters and their predictive strength for the prominence category.

We conclude with the probabilistic study of pitch contours associated with the chosen description units and we discuss the application of this algorithm in the speech technologies.

References