Automatic segmentation of continuous speech at word- and phrase level by using suprasegmental parameters

Klára Vicsi, György Szaszák, Gábor Borostyán

Budapest University of Technology and Economics,
Department of Telecommunications and Media Informatics,
Laboratory of Speech Acoustics
{vicsi, szaszak}@tmit.bme.hu http://alpha.tmit.bme.hu/speech/

In our article we are searching for the answer of a question, whether it is possible to segment the continuous speech at the boundaries of words and phrases by examination of the change of fundamental frequency and energy level in time. We want to increase the robustness of the speech recognizers at linguistic level by the detection of boundaries of words and phrases. In this way we can significantly decrease the searching space during decoding.

In Hungarian language if stress is present, it marks always the first syllable of the word stressed. Thus if we can detect these stressed syllables, than we can detect the boundaries of the word. We describe the developed different searching algorithms.

For the evaluation of these algorithms we carried out examinations the BABEL Hungarian speech database. The results were the best, when the algorithm used the time series of the fundamental frequencies and the energies together. Perhaps the accuracy of the decisions by using these algorithms will decrease in spontaneous speech compared with the ones demonstrated here, but these results show that it is worth to continue our work on this field.