Pronunciation variation modeling of Hungarian language for CSR

György Szaszák[1] – Klára Vicsi[1]

[1] Budapest University of Technology and Economics, Department of Telecommunication and Mediainformatics, Laboratory of Speech Acoustics. 1111 Budapest, Sztoczek u. 2.
`szaszak@tmit.bme.hu` , `vicsi@tmit.bme.hu`
`http://alpha.ttt.bme.hu/speech/`

Keywords. speech recognition, pronunciation variation, language modeling

In our presentation we provide an overview about pronunciation variation modeling used in Continuous Speech Recogniser (CSR) systems at the phonological level. Human speech is caracterized by a high degree of variability with regard to pronunciation forms, mainly in case of spontaneous, everyday speech. This means that the same text can be uttered in more then one correct manner. Pronunciation variability in speech is a not negligible source of misrecognitions in CSR systems, attempts to model pronunciation variation envisage the reduction of the error rates and to get more insight into linguistic mechanisms which control pronunciation variation during speech production.

By examining pronunciation variation in Hungarian language, we used a data-driven method, but we are planning to construct a knowledge base from our current results. The examination procedure was as follows: we examine whether the actual pronunciation of a phone corresponds to its canonical pronunciation, that is regarded as a reference one, obtained by rule based automatic phonotypical transcription of the corpus. The actual pronunciation was obtained by audio-visual transcription of the utterances. By comparing these two streams we get statistics about pronunciation variants of each sound, presented in matrice form. This examination process strated with separate phones, then it was extended to biphones or triphones too.

The importance and further benefits of pronunciation variation examinations are: we get better insight into the governing mechanisms of pronunciation variation so we can investigate the relationship among linguistic or grammatical aspects of speech and pronunciation variation. More practical advantages are the feedback to automatic rule based phonotypical transcription, which means the control and the correction of the phone set or the rules applied. Finally the integration of pronunciation variation models into CSR systems could improve recognition performance by ensuring a more robust and flexible background for speech recognisers mainly on the language modeling level which needs exact knowledge about the different pronunciation forms either in the form of pure statistics or as a knowledge base.