

Zusammenfassung der Dissertation:

**Transmission Error Robust Speech Recognition
using Soft-Features
des Herrn Valentin Ion**

The fast-paced growth in cell phone usage experienced over the past few decades offers a huge potential market for speech enabled mobile services. A suitable technology is Remote Speech Recognition where the actual recognition task is carried out on a remote server in the network rather than on the mobile terminal. Despite the advantages of this client-server architecture, an inherent weakness is that the communication medium may introduce errors which impair recognition accuracy.

There are numerous research studies which have been concerned with methods aimed at the creation of remote speech recognition systems which are robust to transmission errors. A widely used error concealment technique is to replace the erroneously received speech feature by an estimate of the “true” transmitted one and the carrying out of recognition using the resulting point estimate. The improvement in recognition accuracy afforded by this technique has been limited, as the estimate does not perfectly match the transmitted value, i.e. is uncertain.

This thesis focuses on modification of the speech recognition framework to compensate for uncertain features. By reformulation of the classification problem we obtain a novel uncertainty decoding rule which, instead of a point estimate, employs the posterior probability density function of the clean feature. The conditional independence assumption, prevalent in Hidden Markov Model based ASR, is relaxed to obtain a feature posterior density that is conditioned on the complete feature vector sequence observed at the output of the communication channel. This is a more informative posterior than the one conditioned only on the current observation.

This novel decoding method is used to facilitate a transmission-error robust remote speech recognition system. It is shown how the clean feature posterior can be computed for communication links exhibiting either bit errors or packet loss. The probabilistic model which has been employed combines a priori knowledge about the clean features and bit reliability of the received data. The proposed techniques are evaluated in experiments measuring recognition accuracy of small- and medium-vocabulary recognition tasks under various channel conditions. Recognition results are presented for two types of remote recognition: Distributed and Network Speech Recognition. In the latter case common Voice-over-IP codecs are employed.