



Abstract of the PhD-thesis:

**Model-based feature enhancement
for robust automatic speech recognition
in presence of reverberation and background noise**

of Mr. Alexander Krüger

Using hands-free systems for automatic speech recognition may contribute to increased convenience and safety in many application areas. In this case, however, the problem occurs that, in addition to the desired speech signal, also reverberation and undesired background noise are captured by the microphone. These influences cause a degradation of the acoustic features, which are extracted from the microphone signal for the subsequent decoding. Since for the training of the recognizer usually clean speech signals are employed, the discrepancy between the training and testing conditions leads to an increased word error rate.

In this thesis a new technique for the enhancement of acoustic features for robust speech recognition in the presence of reverberation and noise is developed, which is based on the application of Bayesian inference and whose main focus is on the compensation of the effects of reverberation.

On the one hand, the technique involves a priori models to describe the time trajectory of the acoustic features belonging to the clean speech signal and background noise signal. In the former case switching linear dynamic models are employed to exploit correlations between successive features. This thesis concentrates on the training of the models as well as the initialization of the model parameters.

On the other hand, the feature enhancement technique uses an observation model, which relates the features of the reverberant and noisy speech signal to those of the clean speech signal. This relation depends on the room impulse response between the speaker and the microphone. As its blind estimation, which is required in an unknown environment, is extremely sensitive, the room impulse response is modeled statistically. The statistic model has only two parameters, which may be estimated from the captured microphone signal easier and more robust than the complete room impulse response.