

## Proposal for a Project thesis

**Topic:** Uncertainty decoding for robust speech recognition in everyday environments

**Description:** Hands-free voice control of entertainment devices (e.g., an interactive TV) has become of increasing interest in recent years. However, in distant-talking scenarios, where the speaker is located several meters away from the microphones, the multiple reflections of acoustic sound waves on walls or object inside a room alter the characteristics of the recorded speech signal. This limits the performance of automatic speech recognition (ASR) systems and motivated various concepts for environmentally-robust ASR. For instance, front-end techniques aim at suppressing background noise and reverberation, while back-end techniques adjust the acoustic model of a speech recognition system.

Uncertainty decoding bridges front-end and back-end techniques by modeling acoustic features as random variables. To take this probabilistic distortion model into account, we modify the decoding procedure by drawing a finite set of feature samples (realizations of a random variable) and averaging the resulting outputs of the acoustic model [1].

This thesis focuses on improving uncertainty decoding for environmentally-robust ASR by evaluating different numerical sampling strategies. The implementation is to be done with the open-source Kaldi Toolkit using state-of-the-art ASR systems based on deep feed-forward and convolutional neural networks. As prerequisites, the student should be interested in speech recognition and to have basic shell programming experience.

[1]: Huemmer C., Maas R., Schwarz A., Kellermann W. (Interspeech 2015): *Uncertainty decoding for DNN-HMM hybrid systems based on numerical sampling*



Interactive TV scenario

**Supervisor:** M.Sc. Christian Hümmel, room 01.178 (Wetterkreuz 15), huemmer@Int.de

**Professor:** Prof. Dr.-Ing. Walter Kellermann

**Available:** Immediately