Master Thesis

Group Biomedical Devices and Systems from OFFIS R&D division Health has an announcement for an immediate master thesis.

Quantitative Evaluation of Spatial Features for Automatic Speech Recognition

Speech Recognition, Speech enhancement, Machine Learning

BACKGROUND:

Accuracy of ASR (Automatic Speech Recognition) systems and the speech intelligibility perceived by a human listener are defined in a similar way, i.e., as the ratio between the number of words correctly identified and the number of words present in the target signal. Additionally, it is well known that binaural cues have a large impact on speech intelligibility [1] and that preprocessing using spatial characteristics of the signal, e.g. beamforming, can greatly improve the performance of ASR systems [2].
Speech enhancement, which aim at improving speech intelligibility, and preprocessing, which aims at improving ASR accuracy are often based on similar processing schemes and approaches designed for one application can be exploited by the other. For example, features used by ASR systems are often derived from psychoacoustic models [3] while recent approaches have used ASR systems to predict speech intelligibility [4].

[1] J. Blauert, „Spatial Hearing: The Psychophysics of Human Sound Localization.“, Cambridge, MA, USA: MIT Press, 1997.
[2] Results of the 4th CHiME Speech Separation and Recognition Challenge http://spandh.dcs.shef.ac.uk/chime_challenge /chime2016/ results.html“
[3] N. Moritz et al., „An Auditory Inspired Amplitude Modulation Filter Bank for Robust Feature Extraction in Automatic Speech Recognition.“, in IEEE/ACM Trans. on Audio, Speech, and Language Processing, 2015.
[4] B. Kollmeier et al., „Sentence Recognition Prediction for Hearing-impaired Listeners in Stationary and Fluctuation Noise With FADE: Empowering the Attenuation and Distortion Concept by Plomp With a Quantitative Processing Model“, in Trends in Hearing, 2016

OBJECTIVE:

The aim of this work is to quantify the impact of binaural cues, used in binaural speech enhancement schemes, on the performance of an ASR system. The first steps will consist in getting familiar with the literature and in extracting standard spatial features using Matlab. These features will then have to be integrated into a provided ASR framework (Based on Kaldi), which aims at recognizing speech sentences from binaural signals, in order to quantify the benefits of the considered spatial features.

YOUR PROFILE:

- You are currently studying in a University or other institution for higher education.
- You are familiar with signal processing for speech enhancement or other audio applications.
- You are interested in machine learning and speech recognition.
- You have experience programming in Matlab and C or C++.

CONTACT:

Benjamin Cauchi
OFFIS - Institute for Information Technology
Escherweg 2, 26121 Oldenburg
Tel: 0441 9722-183
Mail: benjamin.cauchi@offis.de