

EEI KOLLOQUIUM

Divide and Enhance: Mixtures of Gaussians/Phonemes/Experts in Speech Enhancement

Prof. Sharon Gannot

Bar-Ilan University (Ramat-Gan, Israel)

Donnerstag, der 21.07.2022, 10:00 Uhr

Präsenz (Raum 05.025, Cauerstr. 7) + Zoom-Meeting

<https://fau.zoom.us/j/67196492193>

For decades, model-based approaches governed the speech enhancement landscape. In this talk, we explore a line of work that starts with a data-driven approach, still rooted in statistical modeling, and gradually progresses towards a full-fledged deep neural network (DNN) approach.

We start with an early data-driven approach, the MixMax algorithm. A Mixture of Gaussians (MoG), whose parameters are inferred using the expectation-maximization procedure, is adopted to model the log-spectral features of the speech frames. The algorithm then applies the most suitable enhancer to each of the speech classes. We then continue with a hybrid approach, merging the generative MoG model and a discriminative (DNN).

Later, entirely abandoning the MoG model, we propose a Mixture of Phonemes (MoP) framework comprising a set of phoneme-specific DNNs (pDNNs), each of which estimates a phoneme-specific speech presence probability (pSPP), while a DNN-based phoneme-classifier, selects the most probable phoneme(s). Finally, the MoP model is substituted by a Mixture of Deep Experts (MoDE) model. This architecture comprises a set of DNNs, each of which specializes in a different spectral pattern of the speech signal and estimates the corresponding expert-specific speech presence probability (eSPP). In parallel, a gating DNN weights the eSPP estimates to obtain a compound SPP estimate used in the enhancement stage. The talk is accompanied by sound examples.