FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Elektrotechnik-Elektronik-Informationstechnik

EEI KOLLOQUIUM

Information Integration for Audio-visual Speech Processing

Prof. Dr. Dorothea Kolossa
Ruhr-Universität Bochum

Mittwoch, 06.02.2019, 11^{00} Uhr
Raum 05.025 / N5.17, Cauerstraße 7, 91058 Erlangen

Diskussionsleitung: Prof. Dr.-Ing. Walter Kellermann

Speech recognition has profited immensely from the recent developments in deep learning, and together with signal enhancement strategies for multiple microphones, it is now possible to successfully employ speech recognition even in difficult environments. This talk will focus on strategies for achieving even greater robustness by including visual information, e.g., lip movements, in addition to the acoustic channel alone. This is a strategy that is often employed by human listeners in noisy environments, and this talk will show how that capability can aid machine listening as well. Together with an appropriate stream weighting strategy, error rates of neural-network-based speech recognition can be cut in half in difficult situations by the addition of video information, while achieving reliable improvements even in good acoustic conditions. In turn, the recognition framework can also be used to aid speech enhancement, leading to significant improvements for speech intelligibility and quality under noisy conditions. Similarly, by means of adaptive stream weighting, it is also possible to achieve speaker localization in an optimally integrated fashion, pointing to the value of the visual modality for array processing and source separation. This talk will discuss the architecture of speech processing systems that enable such improvements, and present results that show the benefit of information integration within the contexts of speech recognition, speech enhancement and speaker localization.