The combination of advanced Machine Intelligence and Signal Processing techniques holds promises far beyond today's audio processing systems. Besides Automatic Speech Recognition, recently an increasing number of further speech and speaker characterisation tasks are pursued targeting technical systems' social competence. In addition, the younger field of Music Information Retrieval is growing and there is emerging interest in the computationally `intelligent' analysis of general sound events. Fields of application comprise audio coding, edition, interaction, search, surveillance as well as coaching and entertainment applications.

This seminar first propagates a unified view on the multiplicity of resulting tasks. It further provides a broad overview on the field enriched by extensive research and project results of the presenter's latest work.  The focus thereby lies on realistic conditions and standardisation by open-source software implementations and comparative evaluations. Robustness is advanced by latest and innovative methods such as automated data-acquisition by active- and semi-supervised learning, signal enhancement by non-negative matrix factorisation, analytical feature brute-forcing, and memory-enhanced deep learning - for example in combination with tailored graphical model structures. Machine-based recognition of speech, non-linguistic vocalisations, and paralinguistic speaker states and traits serve as examples of applied speech processing.

As for music processing, examples include blind separation of instruments, determination of tempo, metre and ballroom dance style, as well as analysis of musical key, chord progression, and structure, next to estimation of music mood and singer traits. Finally, examples are complemented by the recognition of general sound events along with their emotional connotation. Synergistic fusion with vision-based and further information is shown at different stages.

In the outlook, avenues towards evolutionary, unsupervised, and holistic signal processing are shown.