The perception of complex sounds like speech is critically dependent on the faithful representation of the signal’s spectral and temporal modulations in the auditory system. Several stages of auditory processing are considered to be crucial for a robust representation of such spectro-temporal modulations and a deficiency in any of these processing stages is likely to result in a deterioration of the entire system’s performance.

This presentation considers models of auditory processing and perception of modulated sounds. The modelling is inspired by neurophysiological findings but reflects an “effective” modelling strategy that does not allow conclusions about details of signal processing at a neuronal level. On the other hand, since the effective model accounts for a large variety of perceptual data, such as spectro-temporal masking patterns and speech intelligibility results, this suggests certain processing principles which in turn motivate the search for neural circuits in corresponding physiological studies. The modelling assumes as one of the key elements an amplitude modulation filterbank at the output of each cochlear filter (Dau et al., 1997; Jepsen et al., 2008). The modulation filterbank realizes a limited-resolution decomposition of the temporal modulations whereby the parameters of the filterbank are not directly related to the parameters from physiological models that describe the transformation from a temporal neural code into a rate-based representation of AM in the auditory brainstem and cortex (e.g., Nelson and Carney, 2004). The output of the preprocessing, i.e., the “internal representation” of the acoustical input signal, has been used in a variety of applications, e.g., for assessing speech quality, predicting speech intelligibility and as a front-end for automatic speech recognition.

A conceptually similar approach has been presented by Shamma and co-workers (e.g., Elhilali et al., 2003; Chi et al., 2005). They described a model that includes an additional “dimension” in the signal analysis. They suggested a spectrotemporal analysis of the envelope, motivated by neurophysiological findings in the auditory cortex (e.g., Schreiner and Calhoun, 1995). In their model, a “spectral” modulation filterbank is combined with the temporal modulation analysis, resulting in two-dimensional spectro-temporal filters. Thus, in contrast to the implementation presented above, their model contains joint (and inseparable) spectro-temporal modulations. In conditions where both temporal and spectral features of the input are manipulated, the two models respond differently. The model of Shamma and co-workers has been utilized to account for spectro-temporal modulation transfer functions, the assessment of speech intelligibility as well as for the prediction of musical timbre.


