Inversion of Auditory Spectrograms, Traditional Spectrograms, and Other Envelope Representations - DTU Orbit (09/12/2018)

Inversion of Auditory Spectrograms, Traditional Spectrograms, and Other Envelope Representations

Envelope representations such as the auditory or traditional spectrogram can be defined by the set of envelopes from the outputs of a filterbank. Common envelope extraction methods discard information regarding the fast fluctuations, or phase, of the signal. Thus, it is difficult to invert, or reconstruct a time-domain signal from, an arbitrary envelope representation.

To address this problem, a general optimization approach in the time domain is proposed here, which iteratively minimizes the distance between a target envelope representation and that of a reconstructed time-domain signal. Two implementations of this framework are presented for auditory spectrograms, where the filterbank is based on the behavior of the basilar membrane and envelope extraction is modeled on the response of inner hair cells. One implementation is direct while the other is a two-stage approach that is computationally simpler. While both can accurately invert an auditory spectrogram, the two-stage approach performs better on time-domain metrics. The same framework is applied to traditional spectrograms based on the magnitude of the short-time Fourier transform. Inspired by human perception of loudness, a modification to the framework is proposed, which leads to a more accurate inversion of traditional spectrograms.