



## Mel Spectrogram Inversion

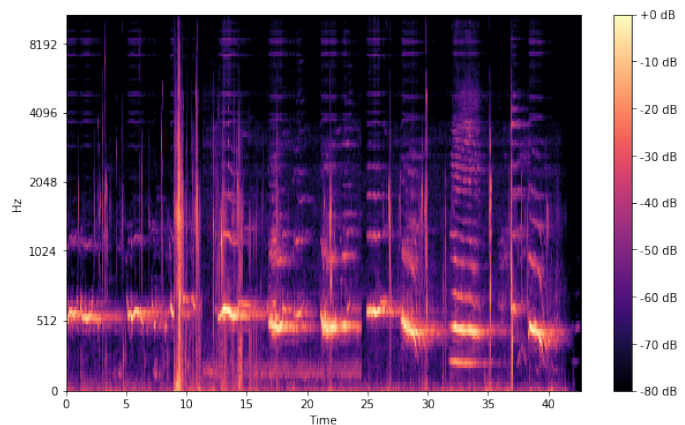
Mel Spectrograms are widely used in audio processing. They are a variant of the standard spectrogram, converting the frequency scale into mel scale. The mel scale is a perceptual scale of pitches. It is calibrated such that for human hearing each pitch is audibly equally spaced from one another.

One drawback is that the Mel Spectrogram does not have a well-defined inverse and is therefore only invertible using approximations. Various approaches have been proposed over time on how to invert Mel Spectrograms, especially in the speech domain. These speech vocoders, however, do not work well for polyphonic music.

Researchers from Apple have recently published a new method for inverting Mel Spectrograms back into the time domain, specifically targeting polyphonic audio. Since their work is not open-source we are looking for a student to implement their work and improve it.

**Requirements:** Knowledge in Python and Machine Learning. Experience with signal processing and Pytorch is an advantage.

We will have weekly meetings to address questions, discuss progress and think about future ideas.



### Contact

- Luca Lanzendörfer : [lanzenoerfer@ethz.ch](mailto:lanzenoerfer@ethz.ch),  
ETZ G93