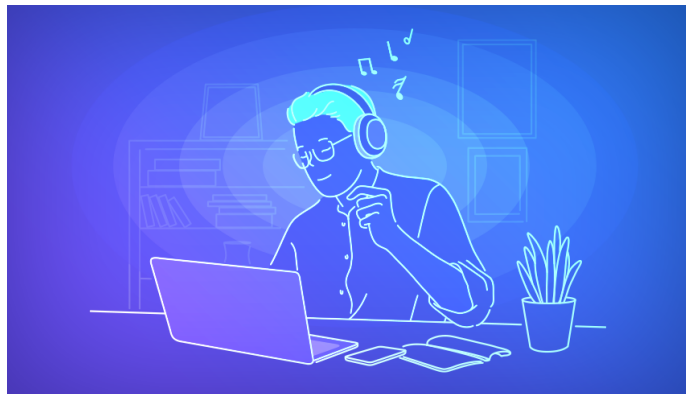




TF-Codec: High Fidelity Neural Audio Compression using Time-Frequency Representation

Despite impressive recent advances in Deep Learning, synthesizing a novel audio waveform remains a challenge in machine learning. As a result, recent work typically uses specific encoders, such as EnCodec or SoundStream, to synthesize sound and avoid generating the waveform. Training audio encoders, however, is challenging due to the oscillatory and multiscale nature of sound, making it data and computation heavy. In this project, we will investigate an alternative technique to build an encoder that requires less memory and computation.

We will build an Autoencoder, similar in structure to EnCodec, but based on the spectrogram representation. We believe this Time-Frequency representation will allow us to create a simpler model, and give us better results without requiring the extensive compute used to train EnCodec.



Requirements: Ability to work independently and determined to obtain results, creative thinking, knowledge with Machine Learning and Python.

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

Contact

- Luca Lanzendörfer : lanzendoerfer@ethz.ch, ETZ G93
- Nathanaël Perraudin : nathanael.perraudin@sdsc.ethz.ch