ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Audio Source Separation using Language Models

Many applications rely on separating audio into their individual sources, this is the task of audio source separation. Applications include background noise removal, separating speaker voices, or separating musical instruments and vocals in music, amongst many others. This is a challenging task and while large leaps have been made it is yet to be fully solved. Historically, source separation approaches have used audio spectrograms and various heuristics to determine the different sources. Newer approaches use audio spectrograms and audio waveforms in autoencoder settings.

In this thesis we want to evaluate if source separation can be tackled using language models. We take inspiration from recent advances in audio and music generation that leverage language models on tokenized snippets of audio to synthesize novel music (e.g. MusicLM, MusicGen). We will discuss details, related work, and ideas in a first meeting.

For this thesis we are looking for a student that is eager to work in the field of audio processing, source separation, and language modeling. The student should be highly motivated to publish their work in a renowned conference.

Requirements: Knowledge in Python, Machine Learning, PyTorch, (Large) Language Models, Audio Processing.

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

Contact

- Luca Lanzendörfer : lanzendoerfer@ethz.ch, ETZ G93