0.1   Background and Motivation

For the last two decades, machine learning and deep learning in particular has greatly improved and developed. It became relevant in almost all fields of study, including audio recognition. As deep learning evolved and new concepts such as Long Short-Term Memory (LSTM) were being theorized, researchers began to move the audio source/signal separation problem from a signal processing problem to a deep learning one.

There are a lot of reasons for implementing a software that is capable of separating audio sources. Needs such as post production of raw recording, automatic karaoke song maker, automatic soloist, and other come to mind. Music signals are mostly a mixture of several sources active simultaneously (voice, musical instrument or synthetic sounds) thus the previously mentioned needs could be solved by unmixing these sources and labeling them. We will next try to come up with a convenient solution using state of the art deep learning and signal processing techniques.

0.2   Problem Statement

Music is a complex thing. Trying to distinguish the sound of an instrument in a song is a task that only humans can do it successfully. Finding patterns of the sound the instrument makes doesn't seem possible using basic signal processing. Using AI to find such patterns might however work, if a good model and enough training data is provided.

My main goal is to design a software that is able to extract audio coming from a specified musical source (instrument or voice). Such software has been previously developed. Unfortunately, these either have poor accuracy or are only capable of extracting vocals. By taking advantage of the latest discoveries in deep learning, we will try to design and train a deep network that can achieve preffesional separation of music sources from a (monaural) recording. The model will be described below, along with other techniques for processing and filtering the audio.