

Deep Learning for Spech-Rap-Singing Audio Classification

Axel Ind, Msc. Computer Science

July 14, 2021

Abstract

I have implemented three different deep learning architecture to learn the 3-output classification of 3 second music clips with a $16kHz$ sample rate. Implementations make use of Librosa for audio-preprocessing, and the Keras environment for model implementation. Labels are learned with a 13-variable MFCC as input data. Results achieved are: @TODO

1 Audio-Preprocessing

2 Architectures

2.1 Classic Dense Network

2.2 Convolutional Neural Network

2.3 Long Term Short Term Network