Separation of drums from music signals

Introduction

This project is about separating the sound of drums from music signals. Monaural audio signal is separated into harmonic and percussive components. Harmonic component contains pitched instrument tracks and percussive component contains drum tracks. Method is useful for multipitch analysis, automatic music transcription, drum detection and modification of music. [1]

Implementation

First part of the implementation is to load the audio file and take the shot time Fourier transform of it in frames. Every frame is windowed with Hanning window. I used frame length of 1024 and overlapping of 512. Sampling rate of the audio was converted to 16 kHz. I took the values from the experiment in the supporting material [1].

Next part is to calculate range-compressed version of power spectrogram with $\gamma = 0.3$, which was given in the material's experiment. After that iteration times "*kmax*" is chosen and a loop is formed to calculate updates for harmonic and percussive components in every iteration with $\Delta^{(k)}$. After the iterations, harmonic and percussive component are binarized.

Then components are turned into waveforms using inverse short time Fourier transform. Then spectrograms of original signal, harmonic signal and percussive signal are formed to evaluate the implementation (figures 1-3). It is seen from the spectrograms that the energy is splitting to two components. The spectrogram of the harmonic component forms horizontal ridges of energy and the spectrogram of the percussive component forms vertical ridges.

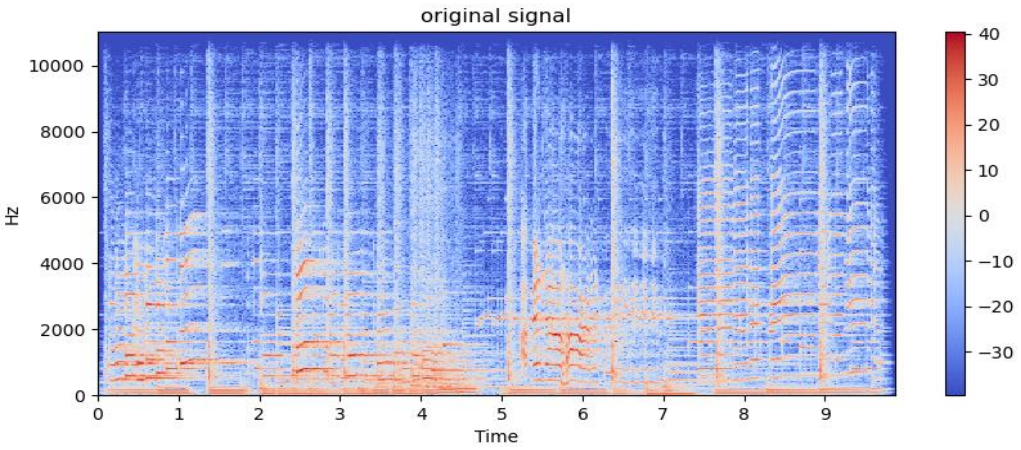

Figure 1: Spectrogram of the loaded music signal

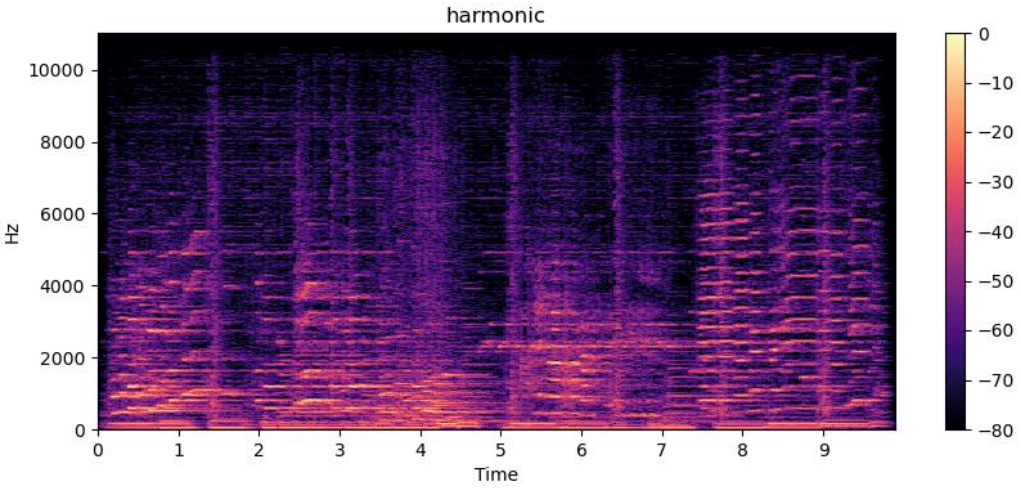

Figure 2: Spectrogram of the harmonic component of the signal

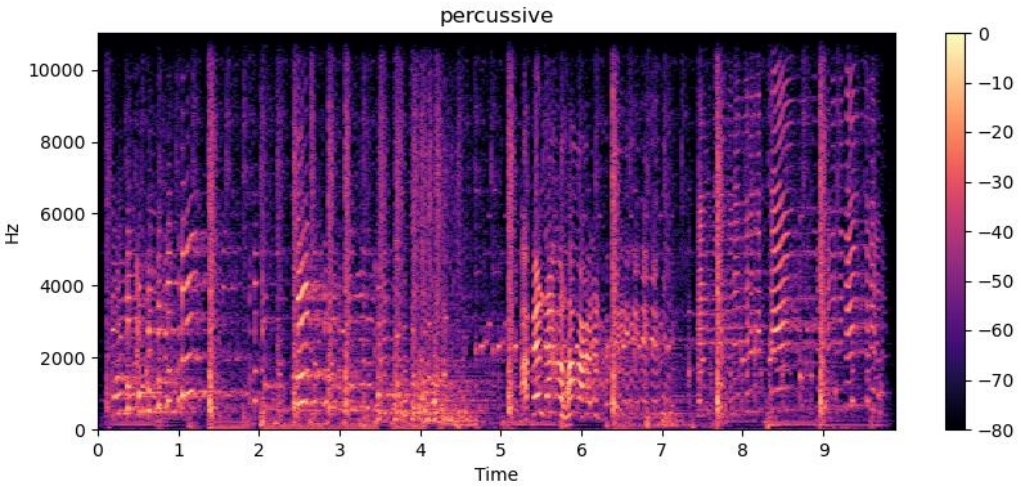

Figure 3: Spectrogram of the percussive component of the signal

Summary

Drum sounds were separated well to percussive component as can be heard in music samples of components. Only the bass drum was almost separated into harmonic component. On the other hand the attack of pitched tone was separated into percussive component. The singing voice wasn't separated well but most of it was separated into percussive component with "project_test1.wav" music, although it should be harmonic.

The separation quality should be measured by calculating energy ratio in the components with different values of γ. In this project the evaluation was measured with signal-to-noise ratio, which gave the value of 10.34 dB. The ratio is good enough because the reconstructed music sounded good. The algorithm could be used as a pre-process to different tasks of music signal analysis. [1]

References

[1] N. Ono, K. Miyamoto, J. L. Roux, H. Kameoka and S. Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram," in Proc. EUSIPCO, 2008