

PROJECT 1 REPORT: SEPARATION OF DRUMS FROM MUSIC SIGNALS

Author: Hoanh Le - (student number: 281271)

1. Introduction

Lately, music signal has become one of the main focuses in the signal processing field with a wide range of tasks such as source separation or signal enhancement. In this report, I will write about one of them, which is separation of harmonic and percussive components from audio signal. In order to solve that problem, I observe that the harmonic component usually has a stable pitch and form parallel ridges with smooth temporal envelopes on the spectrogram, while the energy of a percussive tone is concentrated in a short time frame, which forms a vertical ridge with a wideband spectral envelope. Exploiting the anisotropy, we decompose the original power spectrogram into the harmonic component and the percussive component on the spectrogram. The algorithm is derived from the reference research paper.

2. Algorithm implementation

- Loading the music signal by the library Librosa.
- Calculate the STFT of the signal and range compressed power spectrum matrix W .
- Set initial values for harmonic component and percussive component H and P respectively, each equal half of the power spectrum.
- Calculate delta (updating value) using the equation (23) from the reference paper. [1]
- Update harmonic component and percussive component value, using the equations (26) and (27) from the reference paper. [1]
- Iterate the two above steps k times.
- Binarize the respiration result as the equation (28) in the reference paper. [1]
- Convert harmonic and percussive components into waveforms using the equation (29) and (30) from the reference paper. [1]
- The results are two files: "harmonic.wav" contains harmonic components and "percussive.wav" contains percussive components.

3. Result

- Experiment on the audio file "police03short.wav"

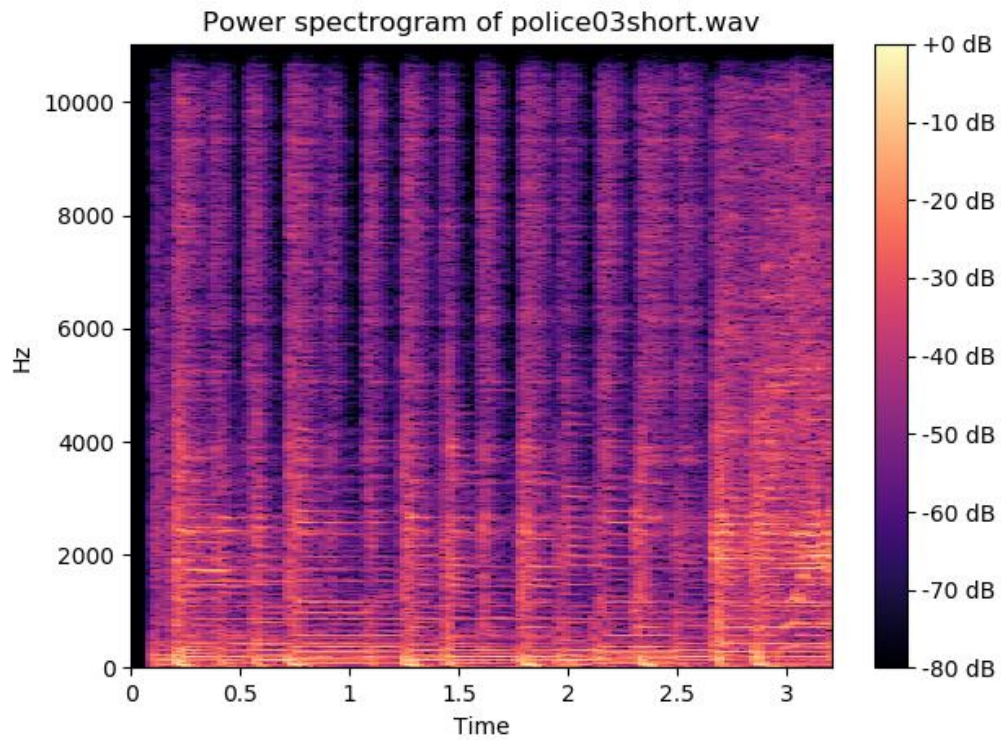


Figure 1: Power spectrogram of police03short.wav

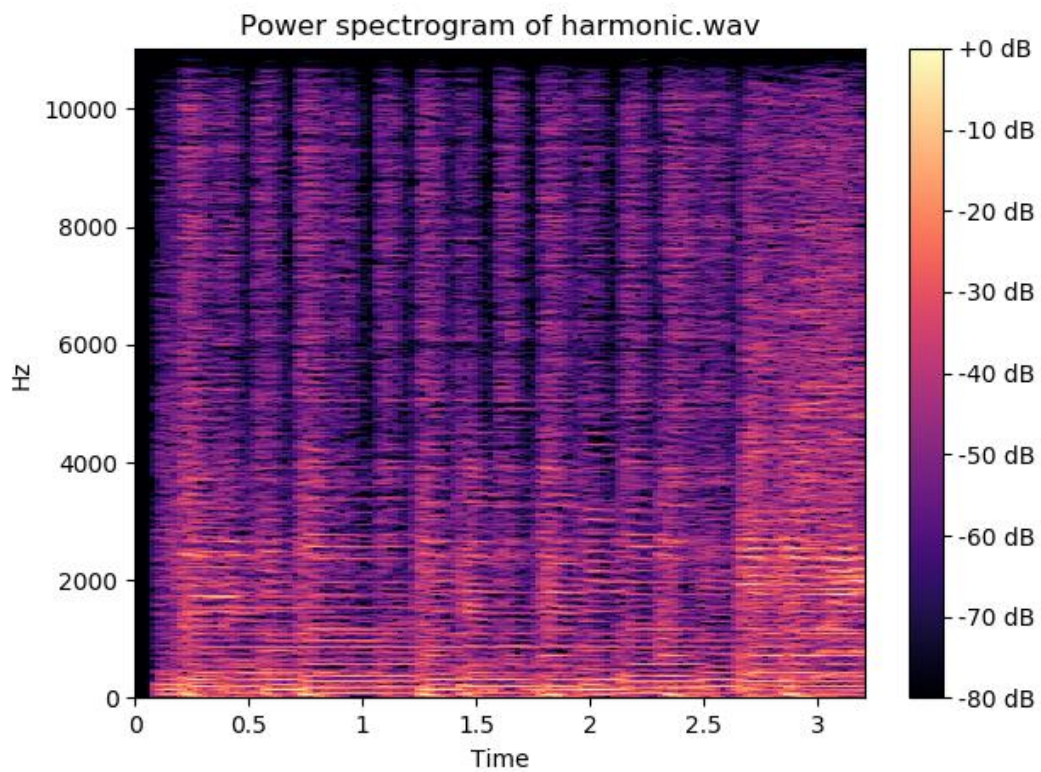


Figure 2: Power spectrogram of harmonic component of police03short.wav

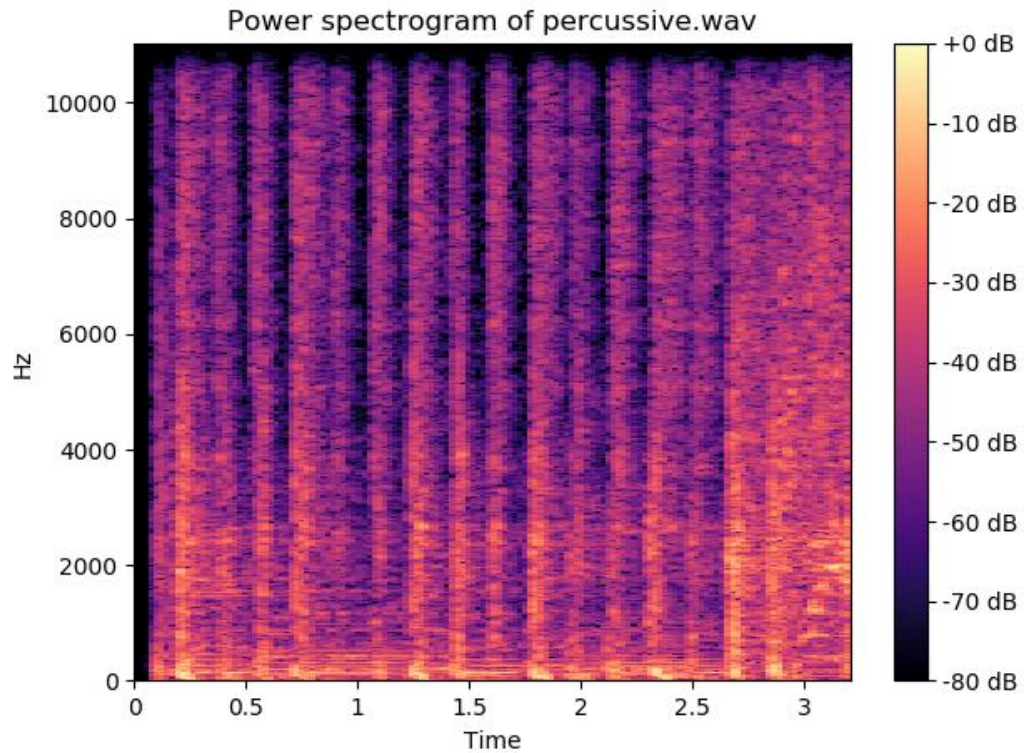


Figure 3: Power spectrogram of percussive component of police03short.wav

I observed that harmonic spectrogram contains parallel horizontal lines, and percussive spectrogram contains straight vertical lines, which is what we expected from the algorithm. Listening on the audio files confirmed the result of the algorithm.

- Experiment on audio file "project_test1.wav"

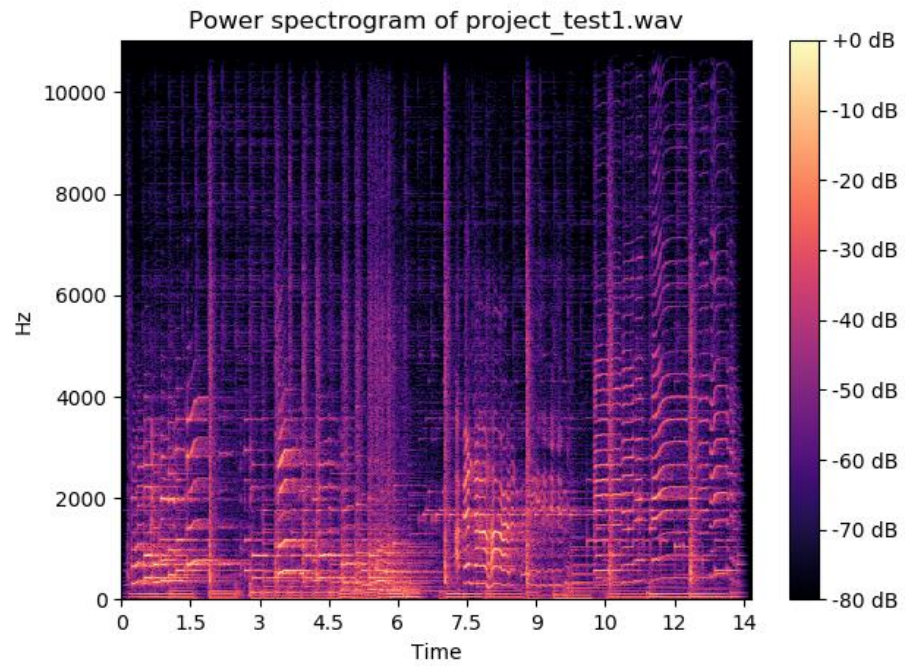


Figure 4: Power spectrogram of project_test1.wav

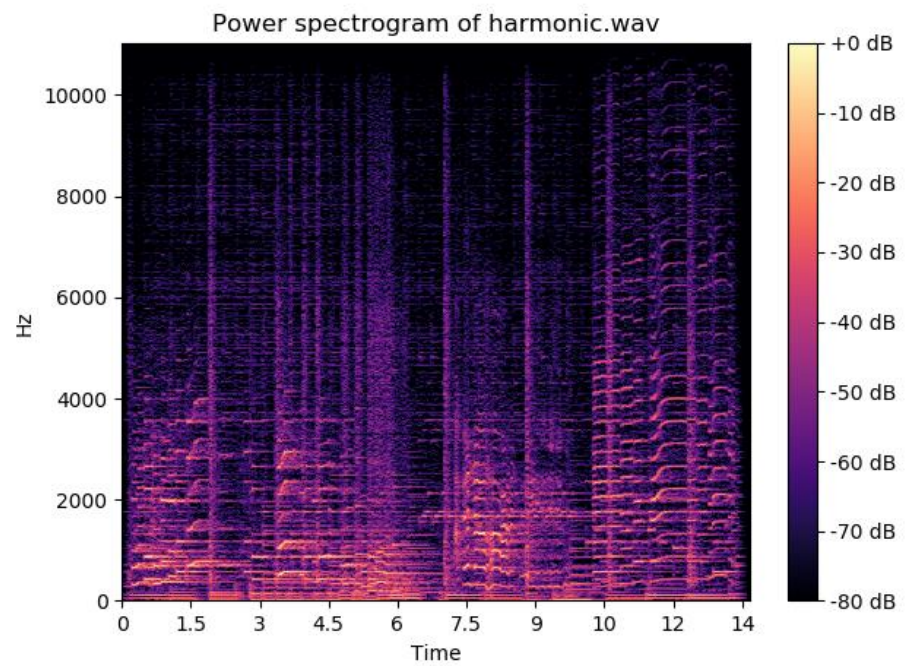


Figure 5: Power spectrogram of harmonic component of project_test1.wav

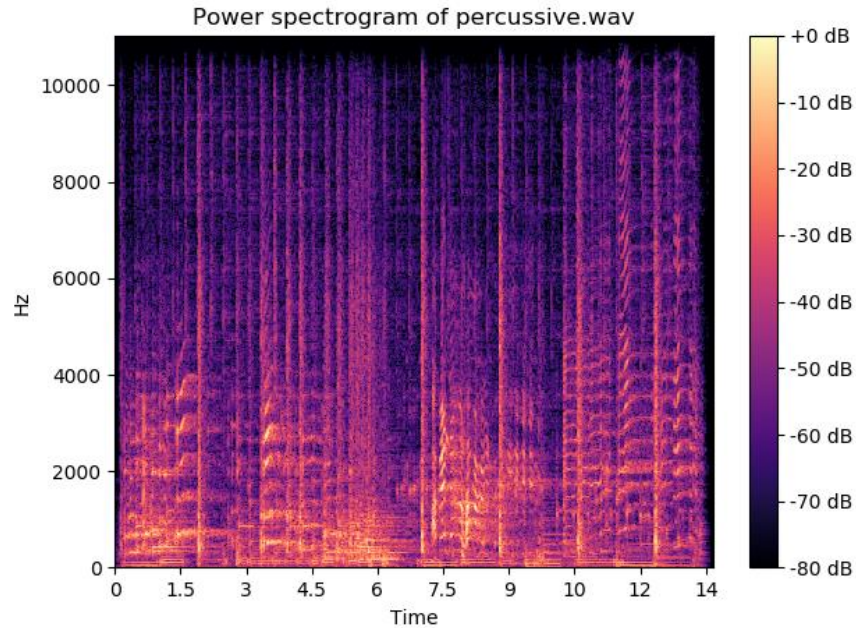


Figure 6: Power spectrogram of percussive component of project_test1.wav

I also observed the expected result in the experiment with the audio file "project_test1.wav". However, as we can see from the graph, the harmonic and percussive components are not separated fully. Listening to the audio files, I saw that the signals still contained some noises. This leads to evaluation of what audio files are suitable for the algorithm.

4. Suitable audio

As also discussed in the reference paper, the attack of pitched tone had tendency to belong to percussive component. A singing voice is also difficult to be perfectly classified to harmonic component as the nature of time-varying pitch. [1] (as we can see in the project_test1.wav). However, the bass drum in drum tracks are almost separated into harmonic component since it has a smooth temporal envelope. [1] (as we can see in the police03short.wav). Therefore, the algorithm is observed to be suitable for only drum tracks audio files.

5. Quality assessment

The criteria I used to evaluate the algorithm is signal-to-noise ratio (the ratio between a desired signal to the level of background noise). SNR is defined as the ratio of signal power to the noise power. A ratio higher than 1:1 indicates more signal than noise.[2]

Signal-to-noise original subtract harmonic component of police03short.wav

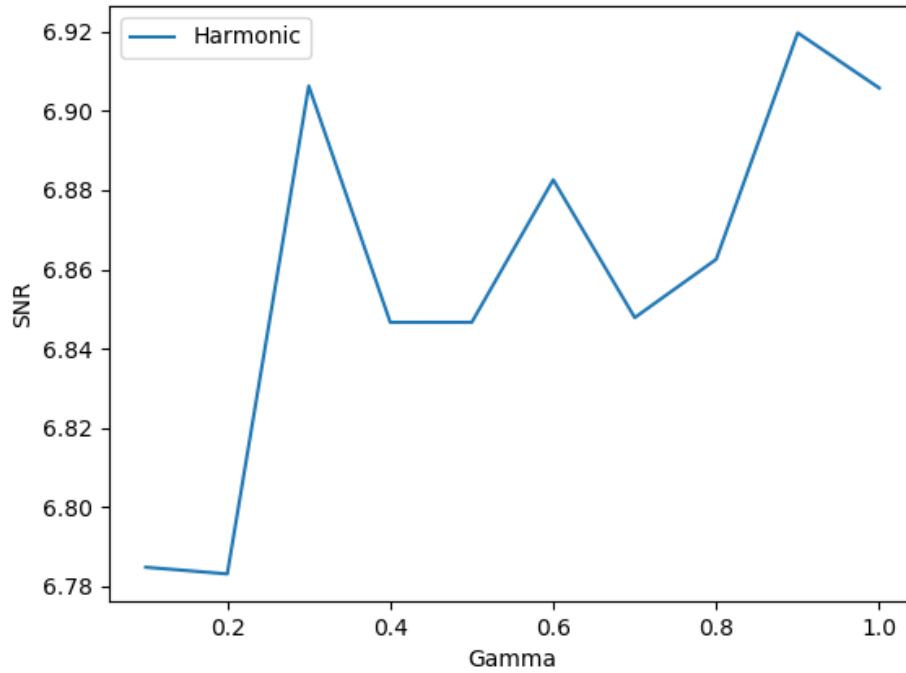


Figure 7: signal to noise of original signal after subtracted the harmonic component

Signal-to-noise original subtract percussive component of police03short.wav

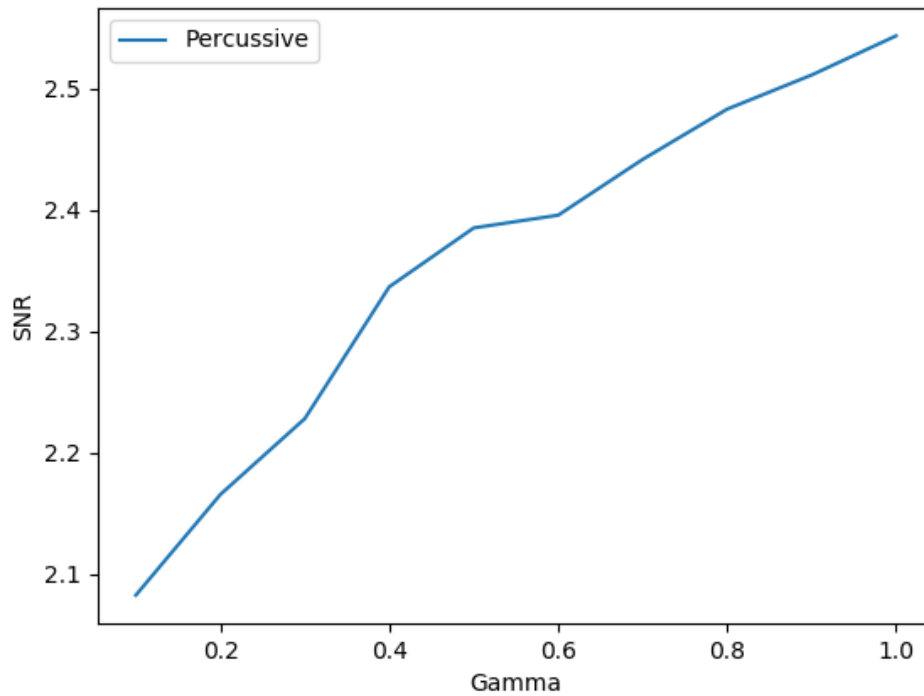


Figure 8: signal to noise of original signal after subtracted the percussive component

As we can see, the signal-to-noise ratio is always higher than 1 with all gammas, which is a good result.

6. References

- [1] Nobutaka Ono, Kenichi Miyamoto, Jonathan Le Roux, Hirokazu Kameoka and Shigeki Sagayama "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram" (EUSIPCO 2008)
- [2] Signal-to-noise ratio. Wikipedia: https://en.wikipedia.org/wiki/Signal-to-noise_ratio