Tampere University

COMP.SGN.120

Introduction to Audio Processing

Project Report

SEPARATION OF DRUMS FROM MUSIC SIGNALS

Wajeeha Jamil
(150209683)

Umar Zubair
(050519664)

December 24, 2021

Problem

A music signal often consists of two different components: harmonic one and percussive one. The simultaneous presence of them makes some tasks harder because of their much different spectral structures. For instance, most of the multi-pitch analysis are interfered by percussive tones, while suppression of harmonic components will facilitate the drum detection or the rhythm analysis. So, there would be a need to separate harmonic component from percussive component in a music signal.

Proposed Solution

Following the reference paper [1] which proposed to derive a simple and fast algorithm specifically for the harmonic/percussive separation based on the anisotropy of them on spectrogram. The first step is to calculate the STFT of an input signal and then apply the algorithm which executes the diffusion-like process of harmonic and percussive energy distribution. Afterwards, we transform the binarized results into waveform with inverse STFT through which, the separated results would be obtained.

Implementation

The implementation can be divided into four steps - STFT transformation, initialization, iteration and binarization, and ISTFT convention.

1. STFT transformation: We calculate the STFT of the input signal first. When calculating the STFT of the input signal, the choosen window and its length are 1024 and hanning window respectively.

2. Initialization: Calculate a range-compressed version of the power spectrogram by taking gamma value 0.3 and then initialize the harmonic and percussive components with half of this calculated power spectrogram.

3. Iteration and binarization: Iteratively update the harmonic and percussive components with the gradient which is the second order derivation of them. After the iterations, take the final updated results and binarize them. In this project, I had used the following mentioned formulas to update harmonic and percussive components.

$$
\begin{aligned}
H_{h,i}^{(k+1)} &= \min(\max(H_{h,i}^{(k)} + \Delta^{(k)}, 0), W_{h,i}), \\
P_{h,i}^{(k+1)} &= W_{h,i} - H_{w,i}^{(k+1)}.
\end{aligned}
$$

   For binarization, I used the following formula:

$$
\begin{aligned}
&(H_{h,i}^{(k_{max})}, P_{h,i}^{(k_{max})}) \\
= &\begin{cases} (0, W_{h,i}) & (H_{h,i}^{(k_{max}-1)} < P_{h,i}^{(k_{max}-1)}) \\ (W_{h,i}, 0) & (H_{h,i}^{(k_{max}-1)} \geq P_{h,i}^{(k_{max}-1)}) \end{cases}
\end{aligned}
$$

4. ISTFT convention: Finally, convert the binarized results into waveform with inverse STFT with the same parameter as the STFT function

Evaluation

The above mentioned algorithm is implemented on the given "project1-test.wav" input signal. The following separation results have been acheived at k=0, k=5, k=10 and k=50.

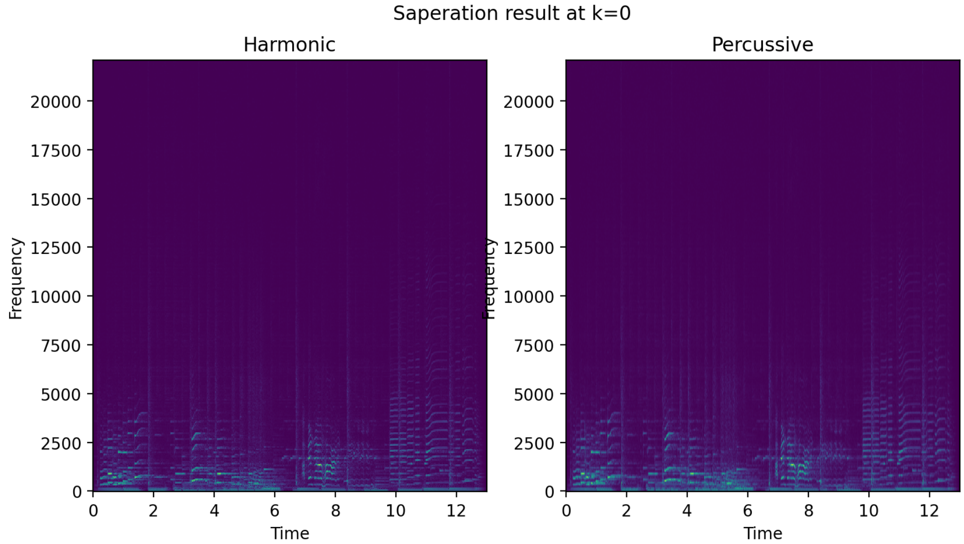

Figure 1: The spectrograms of the iteratively-updated harmonic component (left) and the percussive component (right) at k=0

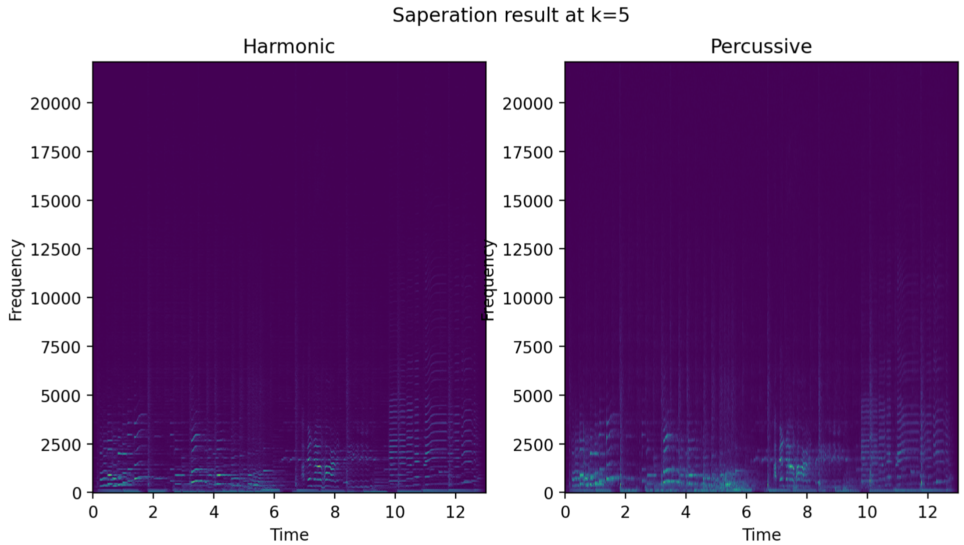

Figure 2: The spectrograms of the iteratively-updated harmonic component (left) and the percussive component (right) at k=5

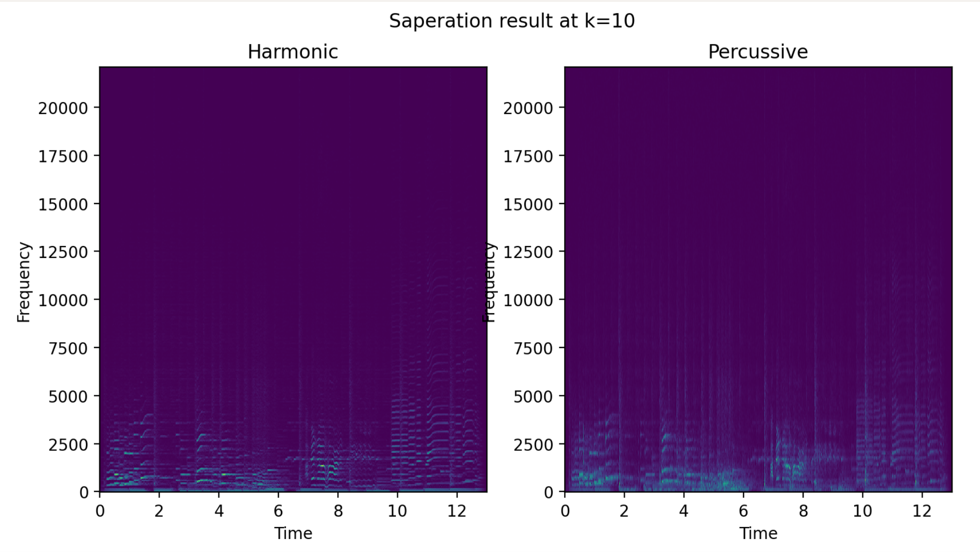

Figure 3: The spectrograms of the iteratively-updated harmonic component (left) and the percussive component (right) at k=10

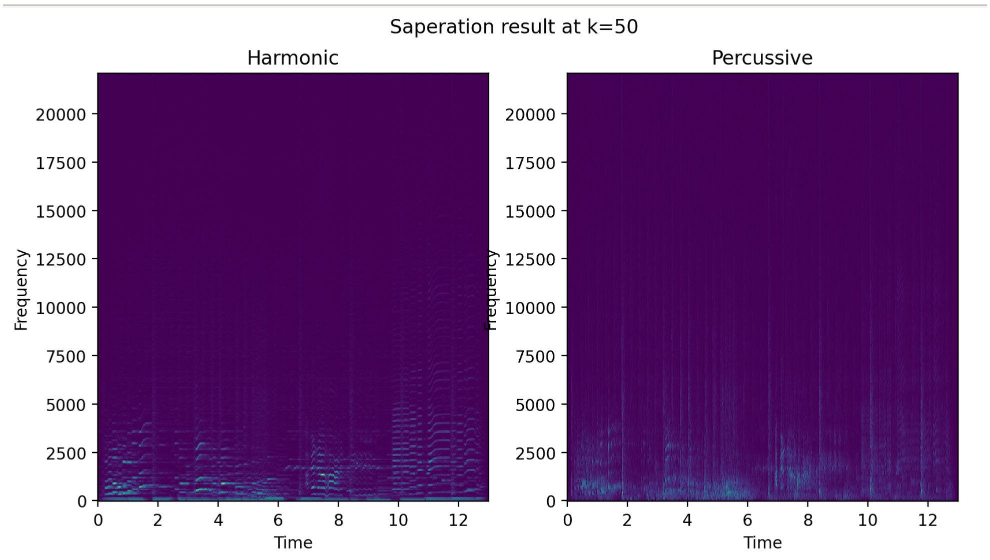

Figure 4: The spectrograms of the iteratively-updated harmonic component (left) and the percussive component (right) at k=50

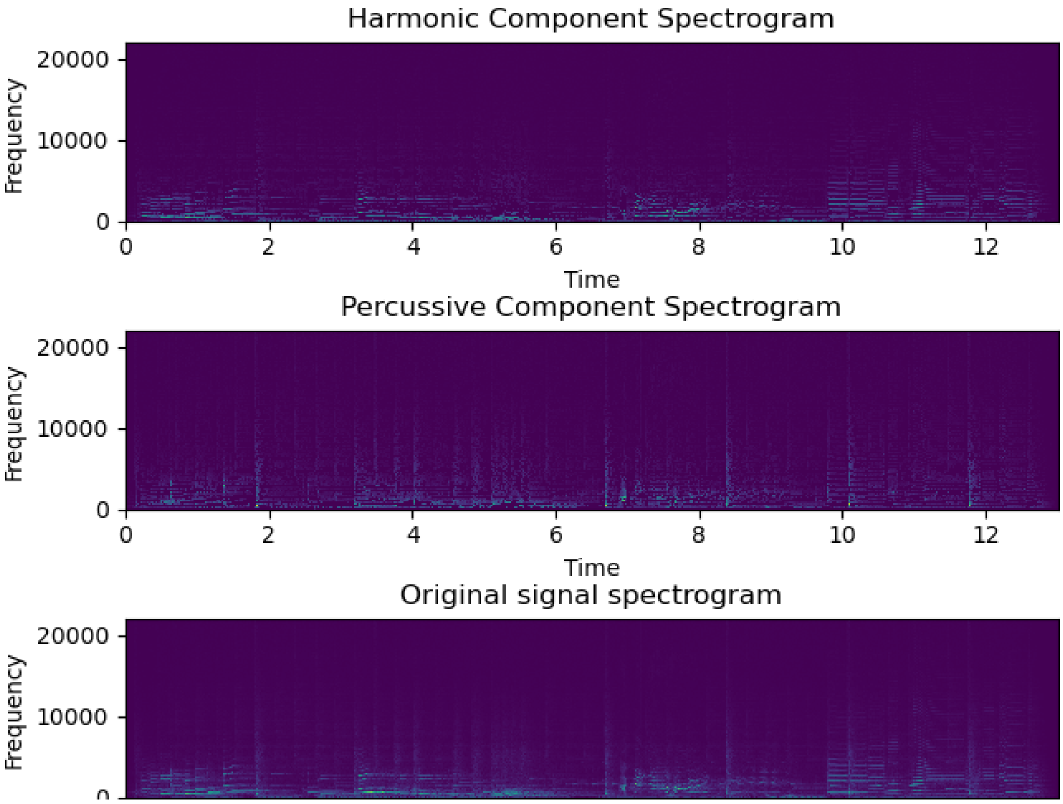

Figure 5: The spectrograms of the harmonic component, the percussive component and original signal with SNR ratio is 311.331

Analysis

What kind of audio material is the algorithm suitable for and why?

The algorithm works only for mono signals. It can be suitable for stereo signals where both channels have to processed separately. However, it would lead to the wrong result for audio materials which includes a lot of time-varying pitch, such as the base drum or singing audio. The reference paper suggests that utilizing wavelet transform instead of STFT might has some potential to improve the performance for pitch-varying components.

How should the separation quality be measured and assessed?

Separation quality can be measured and assessed by a number of ways. Some of the ways can be:

- Compare the SNR. If its value is larger, the result is more precise.
- Use energy radio of each track included in harmonic and percussive signal and draw figure to observe the performance as described in the paper [1].
- Compare the harmonic and percussive signal spectrums. They would be evaluated according to their properties like harmonic components should contain stable tones that produce parallel ridges with smooth time envelopes, but the pulse tones form a vertical ridge with the broadband spectral envelope.

References

[1] Nobutaka Ono et al. "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram". In: *2008 16th European Signal Processing Conference*. Aug. 2008, pp. 1–4.