Project Report

Separation of Drums from Music Signals

Student:

Rafia Bushra

Student ID:

268449

Course:

SGN-14007 Introduction to Audio Processing

Introduction

Music signals tend to have two components: harmonic and percussive. These components, despite co-existing in most music signals, have quite variable spectral structures. This makes it difficult to analyze the signals. In order to detect drums from music signals, the harmonic components of the signal need to be suppressed.

In this project work, I created a Python program that implements the algorithms formulated in the paper[1] to separate a monoaural audio signal into its respective harmonic and percussive components.

Implementation

The algorithm is as follows:

1. Calculate $F_{h,i}$, the STFT of the input signal $f(t)$.
2. Calculate a range-compressed version of the power spectrogram with the following equation:

$$W_{h,i} = \left|F_{h,i}\right|^{2\gamma} \qquad (0 < \gamma < 1)$$

3. Set initial values of the Harmonic and Percussive components as follows:

$$H_{h,i}^{(0)} = P_{h,i}^{(0)} = \frac{1}{2}W_{h,i} \qquad (\forall h, i; k = 0)$$

4. Calculate the update variables $\Delta^{(k)}$ with the following equation:

$$\Delta^{(k)} = \alpha\left(\frac{H_{h,i-1}^{(k)} - 2H_{h,i}^{(k)} + H_{h,i+1}^{(k)}}{4}\right) - (1-\alpha)\left(\frac{P_{h-1,i}^{(k)} - 2P_{h,i}^{(k)} + P_{h+1,i}^{(k)}}{4}\right)$$

The values of $\left(\frac{H_{h,i-1}^{(k)} - 2H_{h,i}^{(k)} + H_{h,i+1}^{(k)}}{4}\right)$ and $\left(\frac{P_{h-1,i}^{(k)} - 2P_{h,i}^{(k)} + P_{h+1,i}^{(k)}}{4}\right)$ are calculated in the code by convolving. Since the sum is divided by 4, the co-vector has the value $[\frac{1}{4}, \frac{-1}{2}, \frac{1}{4}]$.

[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

5. Update $H_{h,i}$ and $P_{h,i}$ with the following the equations:

$$H_{h,i}^{(k+1)} = \min\left(\max\left(H_{h,i}^{(k)} + \Delta^{(k)}, 0\right), W_{h,i}\right)$$

$$P_{h,i}^{(k+1)} = W_{h,i} - H_{h,i}^{(k+1)}$$

6. Increment the value of k while repeating steps 4 and 5 until $k < k_{max} - 1$ where k_{max} is the maximum number of iterations.

7. Binarize the separation result as follows:

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

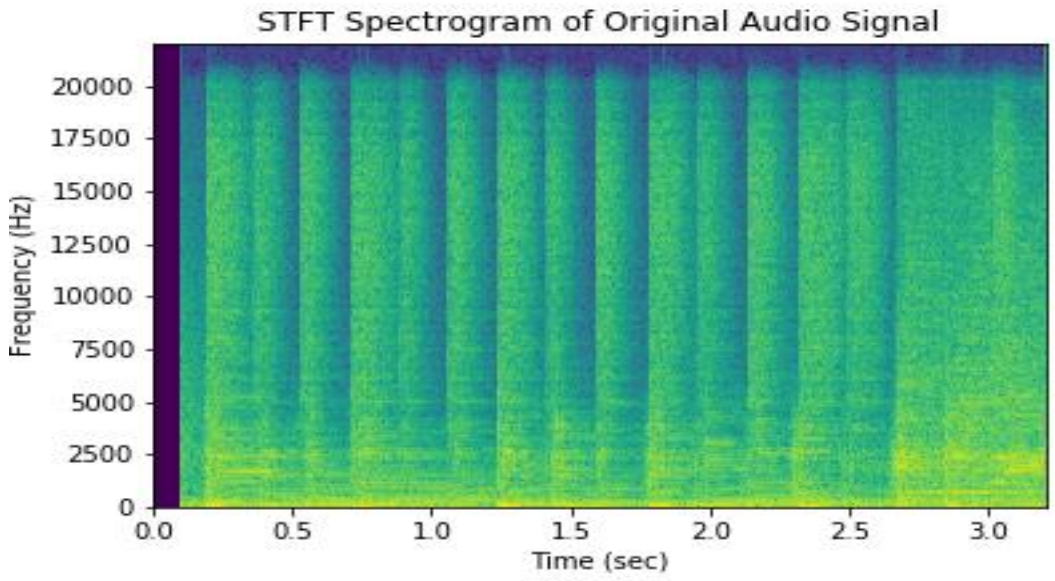
8. Convert $H_{h,i}^{(k_{max})}$ and $P_{h,i}^{(k_{max})}$ into waveforms by getting the inverse STFT's in the following method:

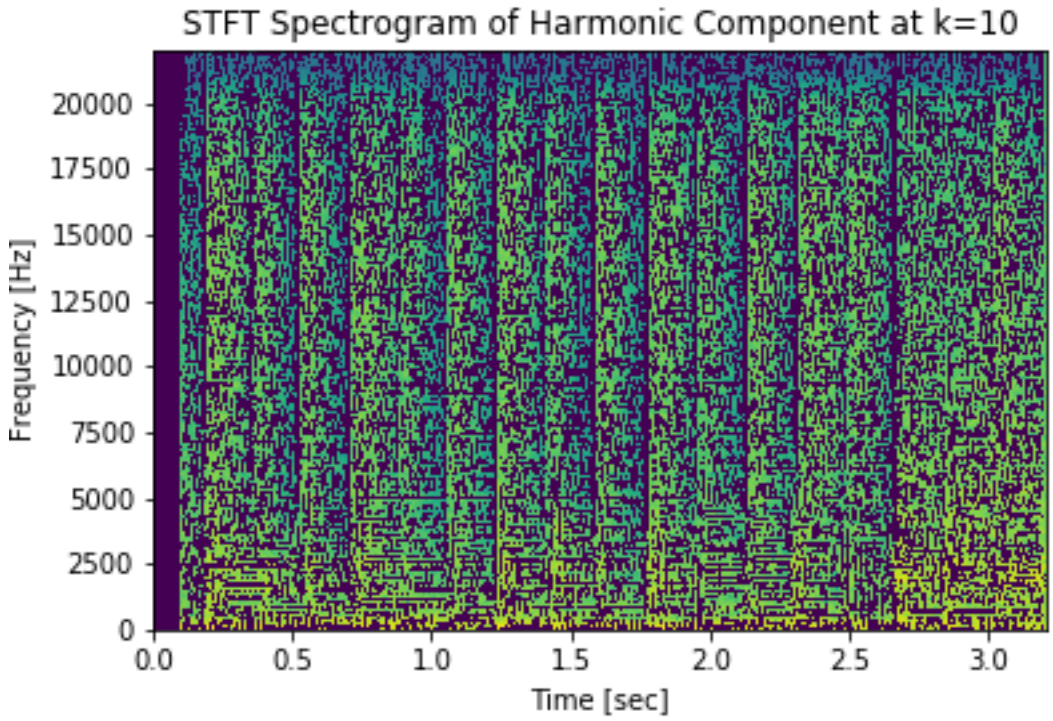
$$h(t) = \text{ISTFT}\left((H_{h,i}^{(k_{max})})^{\frac{1}{2\gamma}} e^{j\angle F_{h,i}}\right)$$

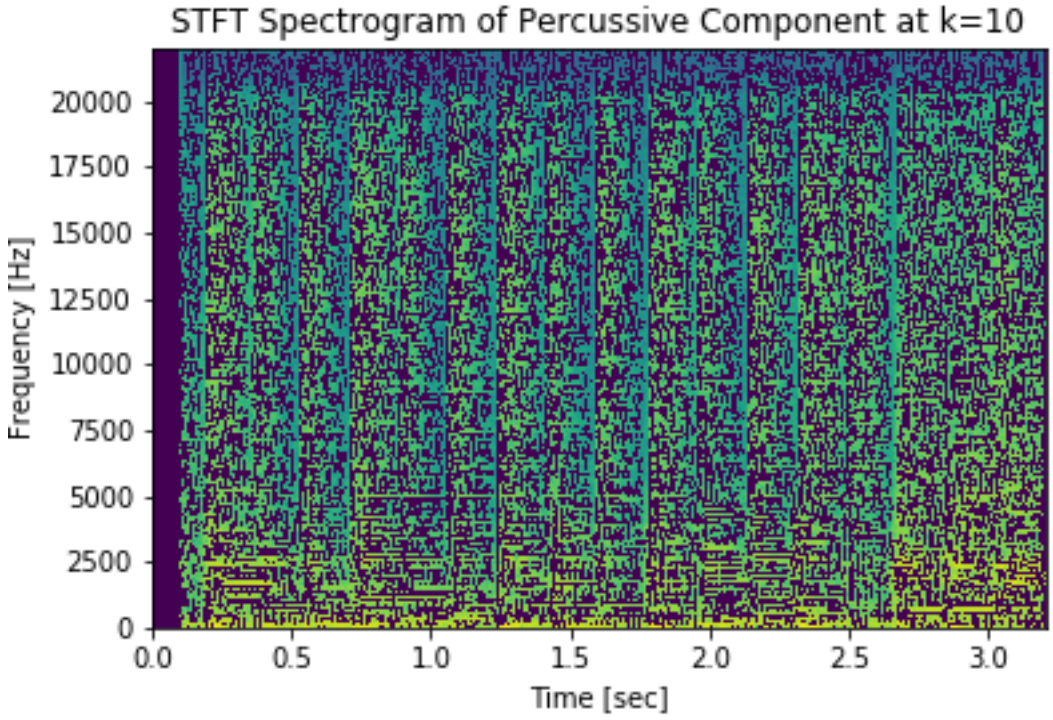
$$p(t) = \text{ISTFT}\left((P_{h,i}^{(k_{max})})^{\frac{1}{2\gamma}} e^{j\angle F_{h,i}}\right)$$

The data in $h(t)$ and $p(t)$ can then be used for multipitch analysis and drum detection.

Plot Results




STFT Spectrogram of Harmonic Component at k=10
Frequency [Hz]
20000
17500
15000
12500
10000
7500
5000
2500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time [sec]


STFT Spectrogram of Percussive Component at k=10
Frequency [Hz]
20000
17500
15000
12500
10000
7500
5000
2500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Time [sec]

Result Evaluation

Calculating the Signal-to-Noise Ratio (SNR)

The result of the separation algorithm can be evaluated by finding the SNR for different iteration values. SNR is calculated with the following equation:

$$SNR = 10 \log_{10} \left(\frac{\sum_t s(t)^2}{\sum_t e(t)^2} \right)$$

Where $s(t)$ = original signal, $e(t)$ = original – separated signal.

I evaluated an audio signal for $k_{max} = [5, 10, 100]$ and received the following ratio values:

SNR = 88.51355601928951 when $k_{max} = 5$

SNR = 88.50772427204126 when $k_{max} = 10$

SNR = 88.4407231722125 when $k_{max} = 100$

Clearly the ratios are similar. This is because the error deviation in the separation is very small.

What kind of audio material is the algorithm suitable for?

This algorithm is intended to be used on monoaural audio signals.

How should the separation quality be measured and assessed?

The SNR values give an estimate for the separation quality. The signal-to-noise ratio of the separation is an error estimation method. It estimates the deviation of the original signal from the combined harmonic and percussive signal.