Project Report

Separation of drums from music signals

Name: Ishan Sen

Student Id: 150220879

Table of Content:

Overview of the Project:

In this project, the separation of percussive components (drum) from music signal is being solved. This work observes an anisotropy in percussive and harmonic components and uses the formulas and algorithm used in the Nobutaka et al paper [1]. The harmonic component has a stable pitch and shapes parallel edges with smooth temporal envelopes on the spectrogram. In contrast, the vitality of a percussive tone is concentrated in a brief time outline, which forms a vertical edge with a wideband spectral envelope [1].

Algorithm implementation: The steps of the algorithm are:

- Firstly, Need to make sure the music audios are in proper format (.wav).
- Read the audio file and gain the sampling rate and perform
- Short-Time Fourier Transform on the signal.
- A range-compressed adaptation of the power spectrogram determined from the sequence gotten within the previous step is calculated.
- Some constants are then initialized:
  - kmax: number of iterations the algorithm will search for the optimally separated power spectrograms. σH and σP: control the vertical and horizontal smoothness weights, set to 1.
- The algorithm then goes into a while loop, repeatedly updating the values of H and P with the changing value of delta. The delta value is found iteratively by the function find_delta. The values of H and P are then changed according to equations (21) and (22) from [1].
- After the while loop is completed, the outputs in terms of H and P are binarized to be either 0 or the value from the range-compressed version of the power spectrogram calculated earlier. Hence, successfully isolating the harmonic and percussive components within the starting signal.
- The Inverse Short Time Fourier Transform is performed, and the components are converted back into the WAV signals: “H.wav” for harmonic audio and “P.wav” for percussion audio.

Evaluation:

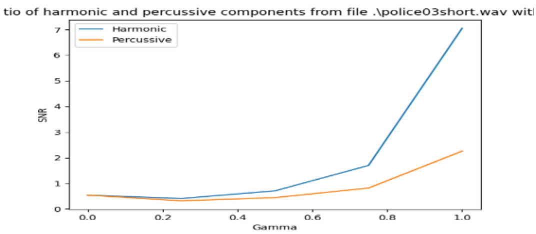
Three distinctive assessment estimations were utilized: signal-to-noise ratio, spectrogram, and sound-related assessment on two sound tests given by the instructor: police03short.wav and project_test1.wav.

Signal-to-noise ratio

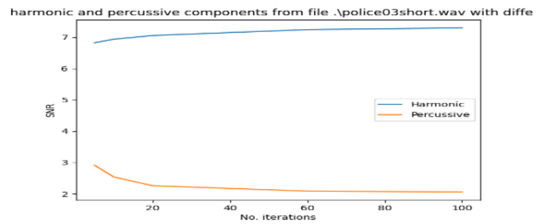
The signal-to-noise ratio is defined as

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

Here, s(t) is the original signal, and e(t) is the initial minus the separated signal. A ratio higher than 1 (greater than 0 dB) indicates more signal than noise.



a



b

Figure 1: Different Gamma (a) and iteration (b) effects on the separation signal-to-noise

These graphs show that setting the extended gamma value to 1 (the highest possible value) will give the most increasing number of SNR, which reduces the impact of noise. Changing the number of iterations, in any case, does not altogether influence the SNR. In project_test1.wav similar situation is being observed. In this case, it should be recommended by setting the value of γ(gamma) to 1, and the number of iterations should be at the slightest 20, but increasing it will not drastically affect SNR.

Spectrogram

Comparative to SNR, spectrograms of the separated sound tests were drawn to examine the impacts of γ and number of iterations in separation quality.

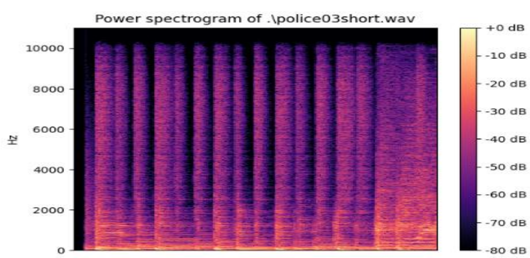


Figure 2. Spectrogram of police03short.wav

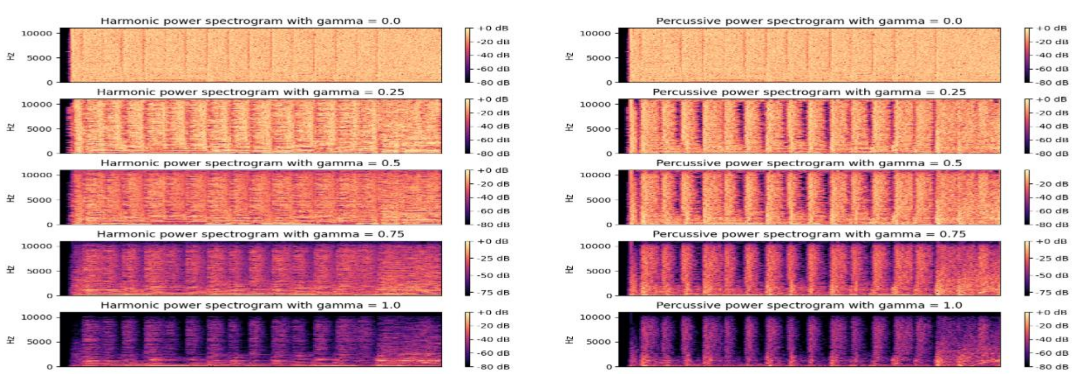


Figure 3: Spectrograms of police03short.wav describing the effect of different γ values

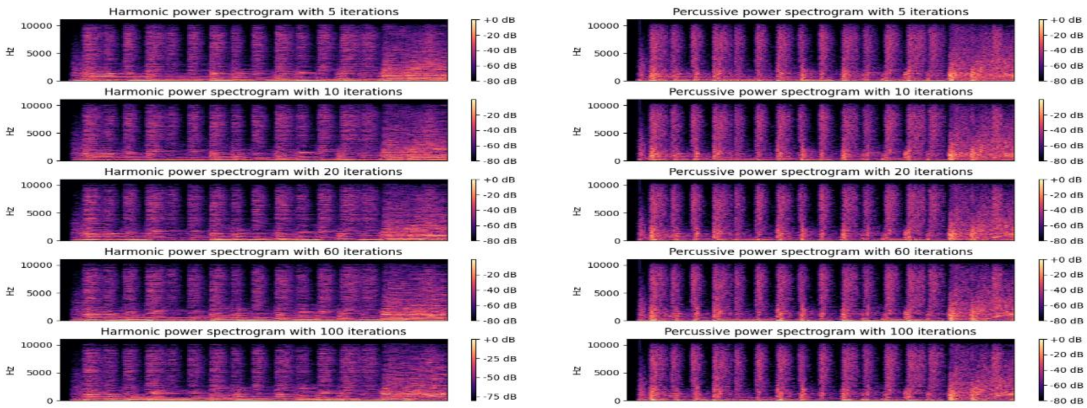
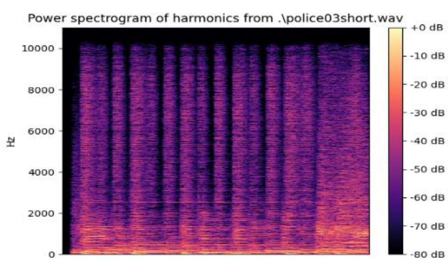
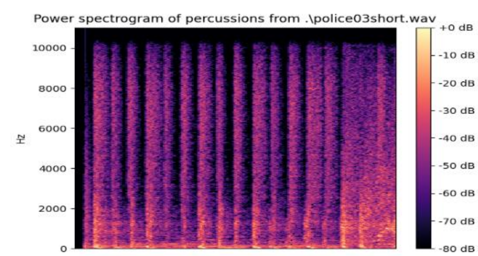


Figure 4. Spectrograms of police03short.wav describing the effect of different no. iterations

It can be seen that from Figure 3, increasing the gamma (γ) values will give a better result, whereas by decreasing the value, the power spectrogram has a noisy effect. However, in Figure 4, there are not much noticeable differences. Below the separated spectrograms of police03short.wav from H.wav and P.wav are shown for better understanding.



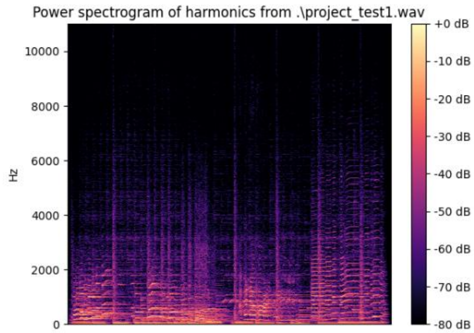
a



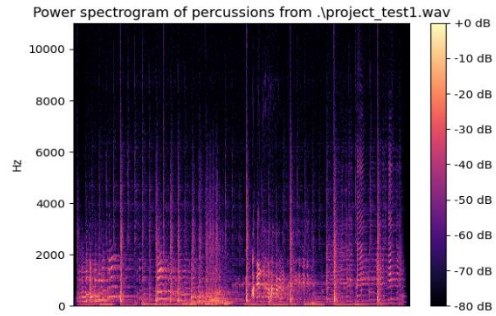
b

Figure 5. Power spectrogram of harmonics(a) and percussions(b) in police03short.wav

If this Figure 5 is compared with the original signal of the power spectrogram, then it can see that the harmonics in the spectrogram (horizontal lines) in (a) as well as the percussions (straight, vertical lines) in (b). Therefore, it can be said that the algorithm works very well. On the other hand, the project_test1.wav does not give a positive result. Whereas in Figure 6, it can see the contrasts between the two separated power spectrograms, appearances of level and vertical lines are displayed in both graphs, indicating that the algorithm does not partition the music signal.



a



b

Figure 6. Power spectrogram of harmonics (a) and percussions in project_test1.wav

Usage for different types of audios

It can be observed from this project; If the test audio is music audio with singing and specific instruments which have a time-varying pitch, then the algorithm will not work well. Therefore, in project_test1.wav, the so-called “attack of the pitched tone” tends to place percussion-only sound [1]; since harmonic-only sound will contain time-continuous recurrence components. Whereas police03short.wav has only pitched instruments and drums, which will be correctly separated. However, at the end of police03short.wav, some singing presents can be heard even after separation.

Conclusion

To summarize, the algorithm works well in pitched only sound like a drum, but with singing present with the audio, this method does not give accurate results. Last but not least, using differents audio samples, this work observed that the algorithm provides similar results: decreasing the range compression coefficient γ increases noise significantly, increasing the number of iterations does not affect; however, the iteration number should be at least 20 for acceptable quality.

References

[1] Nobutaka Dno, 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." Proc. EUSIPCO, Aug2008