Separation of an audio signal into harmonic/percussive components

SGN-14006 Audio and Speech Processing

Md. Facihul Azam
student id: 238949
date: 03.11.2014

This project is a partial fulfillment of the course SGN-14006 Audio and Speech Processing

Abstract:

In this project the separation algorithm[1] was implemented with different parameter values and a few iterations. Here it is investigate the performance of the algorithm. It is proved that increasing the successive number iteration improves better separation results.

Introduction :

Music signal is consists of two components harmonic and percussive signal. In some music signal processing and analysis such as multi-pitch analysis percussive components are interfering. Many research has been done to separate these two components where suppressing of harmonic components allows to detect drum signals. This algorithm[1] is fast and efficiently detects drum signal. On the other hand can perfectly separate vocal tone from instruments.
In this project the goal in to implement the algorithm[1] which efficiently pre-processing the music signal. Here I observed how a monaural audio signal in separated in two components: harmonic and percussive that can be used to separate drum signal from a music. Here it is evaluated that how fast and optimized the algorithm is and how efficiently it can separate the drum signal. The evaluation was carried out by using some sort of short duration music signal[2,3]consists of drum,vocal, piano and guitar.

Methodology:

The separation algorithm [1] implemented here in this project work is as follows:

1. Calculate Fh,i , the STFT of an input signal f(t).
2. Calculate a range-compressed version of the power spectrogram by
$W_{h,i} = |F_{h,i}|^{2\gamma}$ $(0 < \gamma \leq 1)$.
3. Set initial values as
$H_{h,i}(0) = P_{h,i}(0) = (W_{h,i})/2$
for all h and i and set k = 0.
4. Calculate the update variables $\Delta(k)$ as
$\Delta(k) = (\alpha(H(k)_{h,i-1} - 2H(k)_{h,i} + H(k)_{h,i+1}))/4$
$-(1-\alpha)*(P(k)_{h-1,i} - 2P(k)_{h,i} + P(k)_{h+1,i}))/4$
5. Update Hh,i and P h,i as
$H(k+1)_{h,i} = \min(\max(H(k)_{h,i} + \Delta(k), 0), W_{h,i})$,
$P(k+1)_{h,i} = W_{h,i} - H(k+1)_{h,i}$
6. Increment k. If $k < k_{max} - 1$ (kmax: the maximum number of iterations), then, go to step 4, else, go to step 7.
7. Binarize the separation result as
$(H(k_{max})_{h,i}, P(k_{max})_{h,i}) =$

$(0, W_{h,i})$ $(H(k_{max}-1)_{h,i} < P(k_{max}-1)_{h,i})$
$(W_{h,i}, 0)$ $(H(k_{max}-1)_{h,i} \geq P(k_{max}-1)_{h,i})$

8. Convert H_h,i and P_h,i into waveforms by
$h(t) = ISTFT[(H(k_{max})_{h,i})^{1/2\gamma} e^{j*angle(F_{h,i})}]$
$p(t) = ISTFT[(P(k_{max})_{h,i})^{1/2\gamma} e^{j*angle(F_{h,i})}]$

where ISTFT represents the inverse STFT.Here I used spectrogram, matlab routin to do STFT. wavread function was used to read the sound signal. the parameter γ and α were choosen experimentally as 0.5 and 0.5 respectively. Others parameters are as follows:

window length= 1024
frame length = 1024
fftlength = 1024
number of iterations = 0, 3, 10 and 50
window = hann window

Evaluation and Discussion:

Here I used several music signals from music database[2,3].At first an audio song test1[2] was input to the system. It is observed that drum tone can be identified(not properly) separately after a few iterations but vocal cannot be separated alone. The optimum balance parameter α was chosen to 1 and γ =0.55 after several experiments. Next I used three different input as vocal , drum and guitar signal and mixed them. The experiment was carried out in two different way. First drum, dhol and guitar than vocal,drum,dhol and guitar mixture. Drum,dhol and guitar tone can be identified properly. Mixing vocal tone reduce snr and mixed with drum signal even though γ was changed from 0.1 to 1. better separation obtained after 10 iteration when the parameter γ and α were 0.4 and 0.5 respectively. Experimental results are given below:

experiment1: drum+ guitar signal+dhol(South Asian instrument):

Iteration	snr
1	1.8919
5	1.7989
10	1.6716
50	1.2812

Parameter γ = 0.5
α=0.5
'

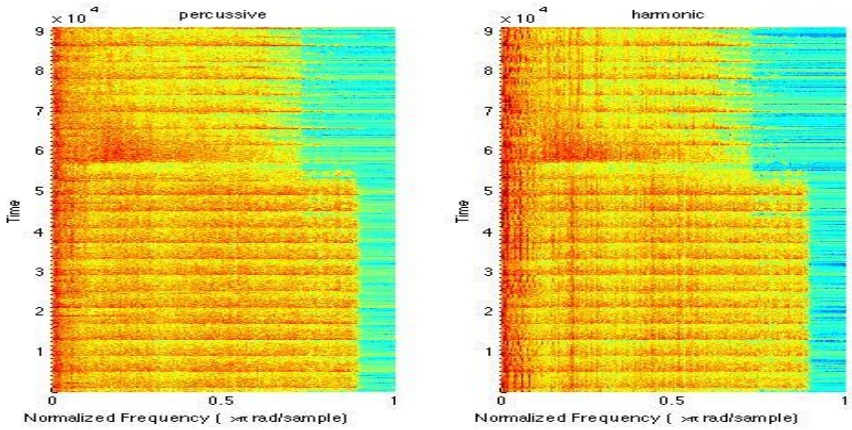


Figure shows percussive and harmonic component after 1 iteration.

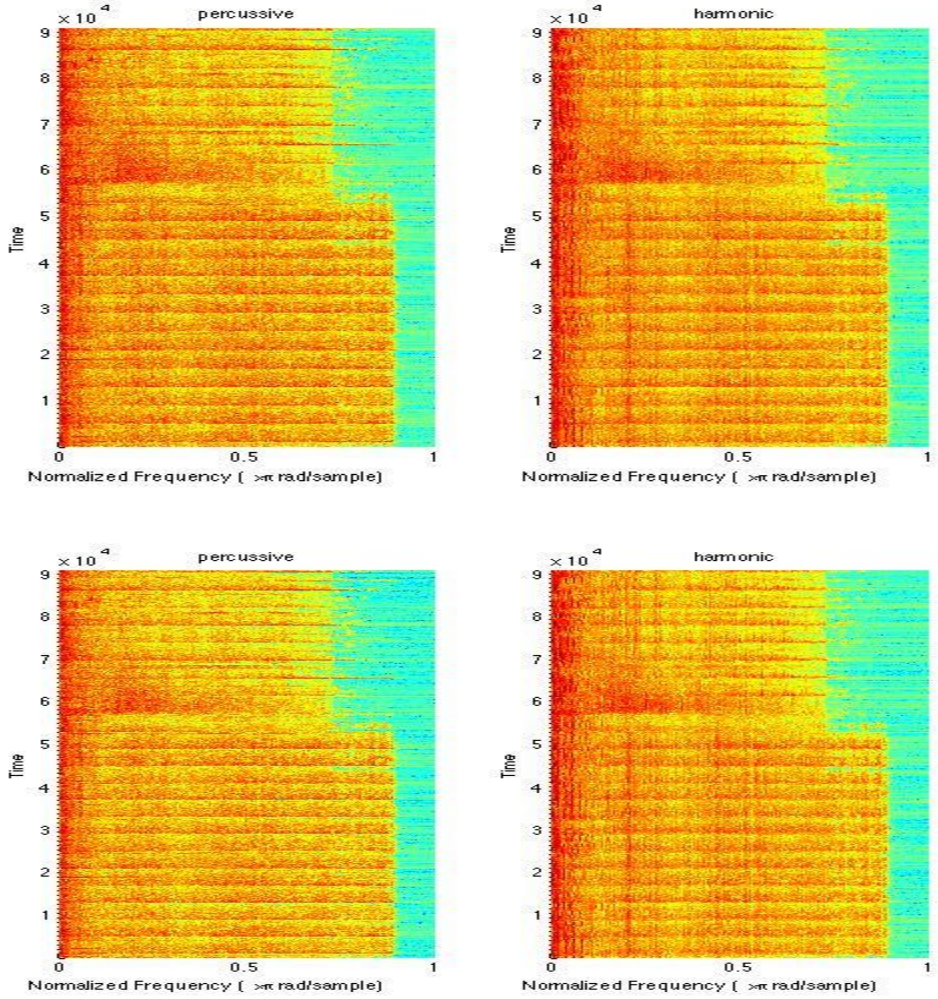
Experiment2: vocal+ drum+ dhol-guitar signal:

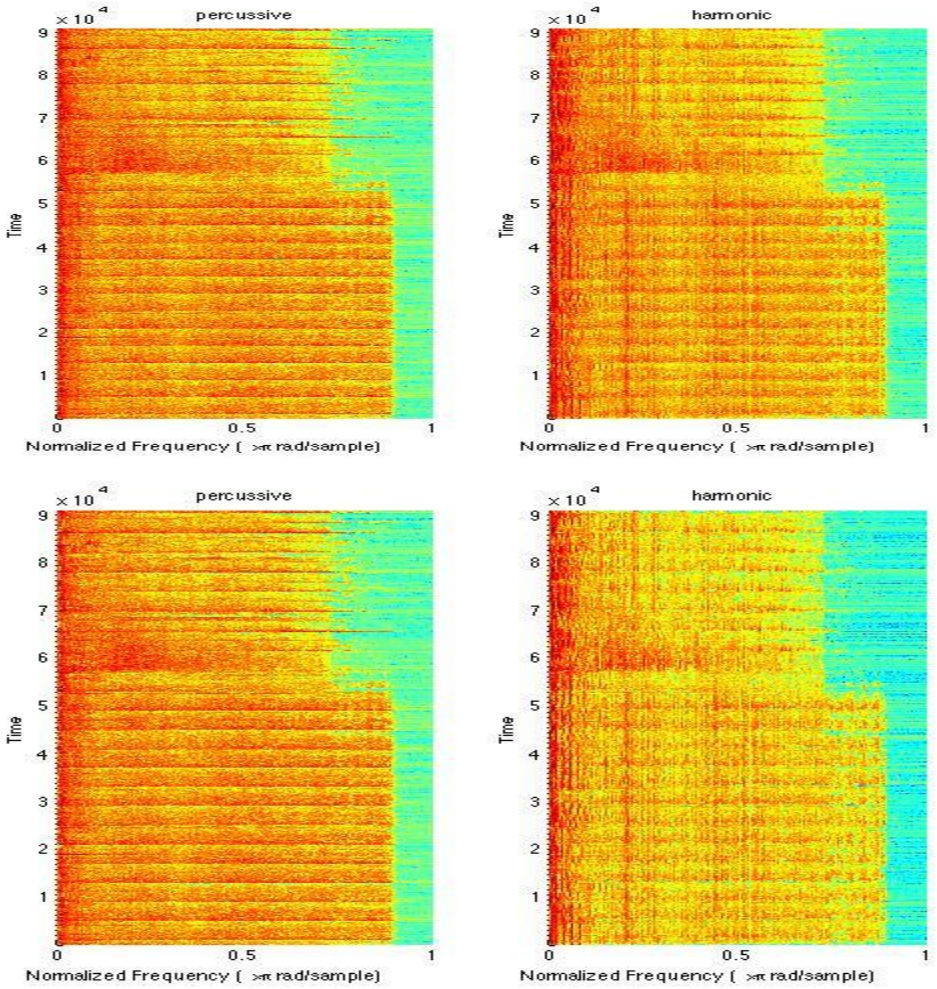
Iteration	snr
1	1.2960
5	1.2860
10	1.2528
50	1.1584

Parameter $\gamma = 0.4$

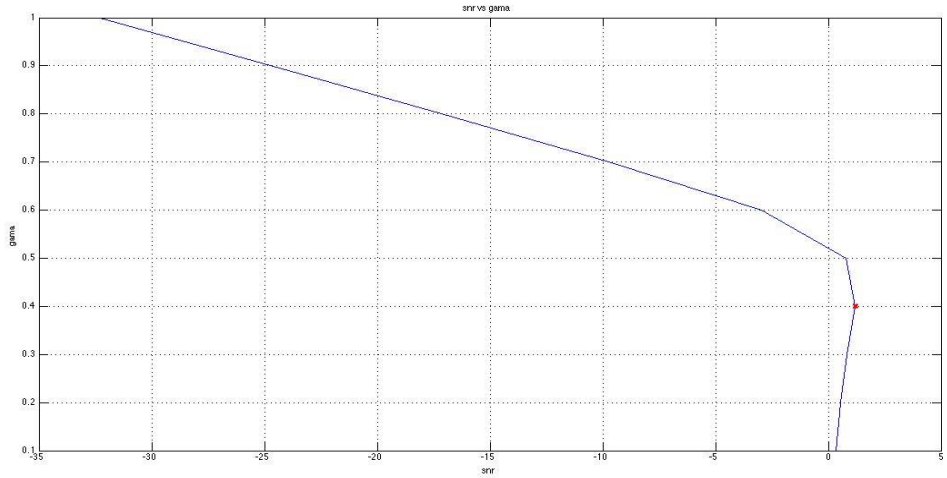
$\alpha=0.7$

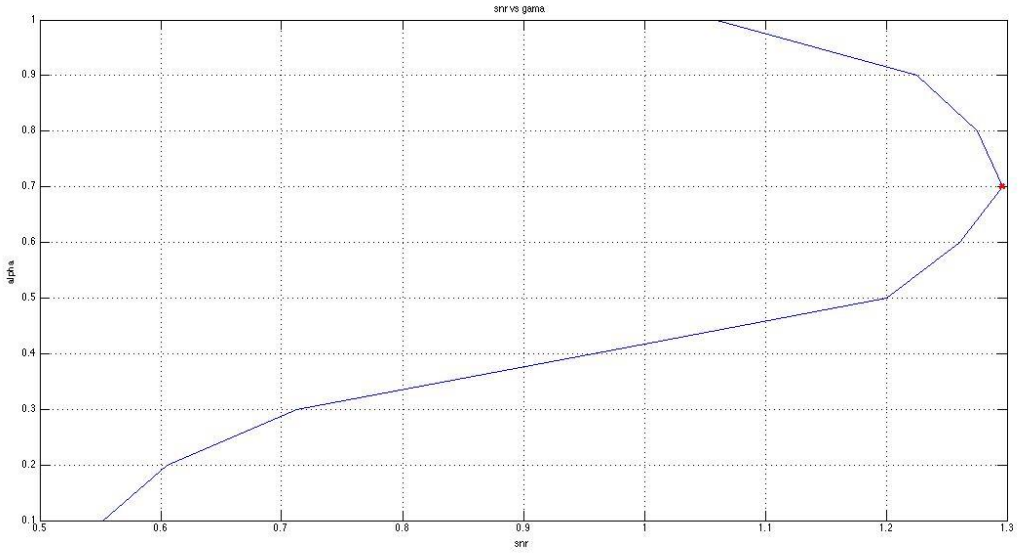
In the second experiment vocal signal was added with drum and guitar. Parameters are chosen after a few subjective evaluation. Noise still exists in output signals.





From top to bottom row represents harmonic and percussive components with iteration k=1,5,10,50 respectively. The snr vs gamma curve as follows (iteration = 1) :
Gamma and alpha was chosen objectively according to snr vs gamma and snr vs alpha curve as follows:





Conclusion:

The implementation results shows that drum and instrumental signal are well separated. Different music signal needs different values of parameters to separate drum signal from the music.

References:

[1] N. Ono, K. Miyamoto, J. Le 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.

[2] http://sisec.wiki.irisa.fr/tiki-index.php?page=Professionally+produced+music+recordings

[3] http://www.last.fm/music/+free-music-downloads/acoustic