SGN-14007 Introduction to Audio Processing

Topic 1: Separation of drums from music signals

Petri Tikka - 206059

Joonas Linnosmaa - 218179

Division of work

Work was conducted in group sessions where both participants contributed to the outcome by sharing ideas and discussing and testing different approaches.

What problem is being solved in the project work?

Project work targets to solve the separation of harmonic and percussive components from monaural audio signal. Proposed solution is useful for multi-pitch analysis, automatic music transcription, drum detection, modification of music

How is the problem solved?

Problem is solved by implementing algorithm from publication: "Separation of monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram". Algorithm exploits differences in the spectrograms of harmonic and percussive components. The separation is done by complementary diffusion on spectrograms.

What assumptions were made?

Used audio signal must be of monaural nature. For the sake of using this algorithm and having proper results, the signal should include harmonic and percussive components, which are depicted as a vertical and horizontal component on spectrogram. Algorithm utilizes gamma (used in determining energy ratios of harmonic and percussive components) and alpha (balance parameter) parameters that are between [0 1]. Harmonic and percussive spectrogram gradients follow independent Gaussian distributions.

Short description of the implementation. What stages does the algorithm consist of?

- Calculating the STFT and power spectrogram of the input signal.
- Calculate Calculate range-compressed version of the power spectrogram.
- Set the initial values of harmonic and percussive components as ½ of the range-compressed power spectrogram.
- Calculate the update variables. The balance parameter alpha controls the strength of the diffusion along the vertical and horizontal directions.
- Update and harmonic and percussive spectrogram components. Updating harmonic and percussive variables were done in a loop structure where operations were done on matrices that required padding (including zeros on its sides).

- Increment constant to repeat loop structure so that harmonic and percussive components are found with enough accuracy.

- Binarize the separation result. Binarization is done so that range-compressed power spectrogram of the original signal will have all elements zero where harmonic component (H) >= percussive component (P).

- Finally, we convert harmonic and percussive components into waveforms, power spectrograms of the components are plotted and signal-to-noise ratio (SNR) is calculated. In calculation of SNR we utilized the information given by the course staff:

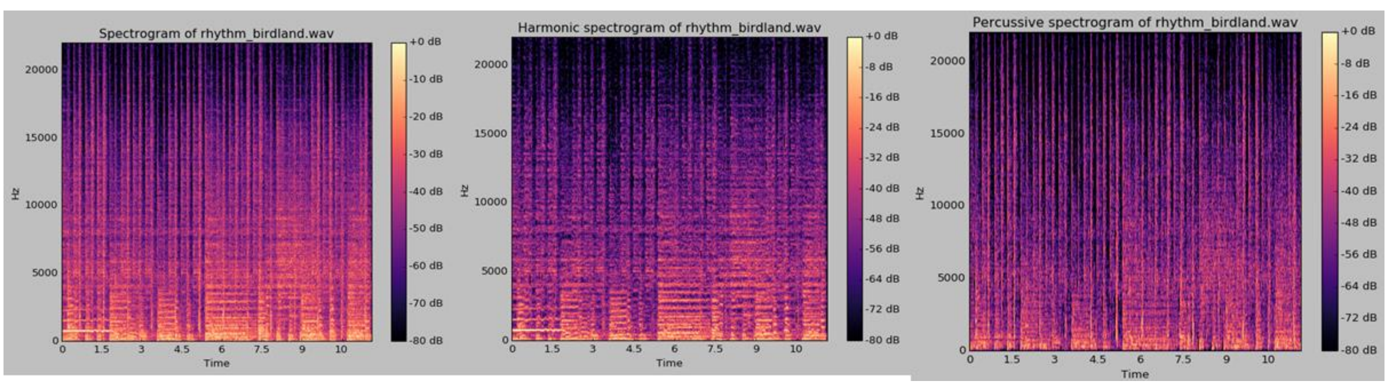
  Extra information for the separation project SNR evaluation: If

  - *s(t)* contains the monophonic mixture of the instruments and *y(t)* is the mixture of your separated harmonic components $y_h(t)$ and percussive $y_p(t)$ components, then *e(t)* is the difference between *s(t)* and *y(t),* i.e., *e(t)=s(t)-y(t).* Ideally *e(t)* should be close to zero all the time, and SNR very high, (+20 dB or better).

What is evaluated in your experiments, and how?

Evaluation consists of separating harmonic and percussive components from the original signal that are compared as spectrograms. Finally signal-to-noise ratio of the original and original minus the separated components are evaluated.

- Spectrograms of the original signal, harmonic and percussive components:



- Signal-to-noise ratio of harmonic and percussive components in comparison to the original signal as well as SNR of original signal and mixture of separated harmonic and percussive components:

```
Signal-to-noise ratio of original signal and harmonic: 1.2108017912373334
Signal-to-noise ratio of original signal and percussive: 7.586338931320293
Signal-to-noise ratio of original signal and mixture of separated harmonic and percussive components: 140.2410379415571
```

  SNR of original/harmonic: ~1.21 dB
  SNR of original/percussive: ~7.59 dB
  SNR of original/mixture of harmonic and percussive: ~140.24 dB

What kind of audio material is the algorithm limited to and why?

Algorithm is limited to utilize mono signals, and stereo signals can be only used if both channels are processed separately. Algorithm has difficulties to classify sound with time-varying pitch. Spectrograms with unclear distributions of harmonic and percussive components is also limiting aspect. This is because the algorithm searches for horizontal and vertical gradients from the spectrogram and if they are unclear the solution will also be.

How should the separation quality be measured and assessed?

Separation quality can be measured by listening the individual components of the harmonic and percussive parts of the signal and comparing them to the original audio file. Quality can be measured and assessed visually from the calculated spectrograms, where harmonic components should be observed as horizontal ridges and percussive components as vertical ridges.