MIGAA - Music Interpreted as Graphical Art and Animation
Roseline Okpara | rco5918@rit.edu
Dr. Hans-Peter Bischof | hpb@cs.rit.edu

INTRODUCTION
Visual abstractions of data such as maps or diagrams have been used for centuries to aid human thinking and provide graphical assistants for learning. Music visualization is one of the most promising approaches to aid the learning curve of a general audience listening to music in regards to musical structure and common musical features. This project describes an audio signal processing and visualization library that extracts key features of a musical performance based on Fourier Analysis and represents them as visual components. A user study is conducted to validate the effectiveness of this library.

FOURIER ANALYSIS
Sound is fundamentally a waveform consisting of a weighted sum of sine and cosines that change over time. This is referred to as the time domain. The shape of the waveform is driven by the harmonic content of the sound. Fourier analysis is used to decompose a signal from the time domain into the frequency domain as well as calculate the frequency spectrum.

FEATURE EXTRACTION
Temporal (Time Domain) Features
- The Root Mean Square (RMS) is loosely corresponds to loudness and is used to detect rhythmic structure.
- The Zero Crossing Rate (ZCR) is correlated to pitch or the perceived fundamental frequency.

Frequency Domain Features

\[ \text{RMS}(f) = \sqrt{\frac{1}{n} \sum_{n} (f(n))^2} \]

\[ ZCR(f) = \frac{Z f_s}{2N} \]

BEAT DETECTION ALGORITHMS

Statistical Streaming
- Detect brutal variations in sound energy by computing the local average sound energy and threshold the instant energy against the average energy.
- Issues detecting “loud” music such as rock where there is a lack of variation in sound energy.

Frequency Selected Sound Energy
- Break the signal into sub-bands and compare it with its own recent energy average.
- Allows beat detectors on each sub-band to pick up different instruments.

Comb Filter Processing
- Select 2 - 5 second sample from the original signal and divide into several separate signals.
- Reduce signals down to their envelopes. This allows for the ability to differentiate when the sound amplitude changes. The largest change is a beat.
- Various comb filters are used to determine which signal yield the highest energy.
- Too computationally expensive to run on a streaming source.

USER STUDY
- 100 People with various musical experience were surveyed at Imagine RIT.
- Questions asked about the recognizability of specific audio features in the given animations.
- Three 30-second samples of music were played for each of the four visuals.

RESULTS

GENERATED VISUALIZATIONS

We presented a system that processes sound to extract features and then display them in a visual way. Four different visuals were created to take different approaches to displaying the same information. Despite the simplicity of the extracted features and visuals, it was enough to allow the user to pinpoint musical structures. A user study to help confirm the validity of these systems as a visual aid for music.

REFERENCES