Report for
ECE457B Course Project, Winter 2017
Title_Of_Project

Daniel Cardoza/dpmcardo/20471664
Lara Janecka/lajaneck/20460089
Hong Wang/hmwang/20469058

Due Date: March 30, 2017

1 Abstract

2 Introduction

3 Background

3.1 Music Files

The files used for analysis were primarily wav files. To use another type of file required conversion into a wav file. Wav files consist of a header containing relevant information such as sample rate and the number of channels, and the sound data for each sample. This format does not compress the data and thus was chosen of the most accurate results. A parser was written to extract the sample rate and a waveform for the sound data. The sample rate is the number of samples taken per second. For CD quality (the most common type of file looked at) this value is 44,100. The waveform is the list of data for the sound data at each sample point for each channel.

3.2 Stepmania Files

3.3 Beat Detection

Notes:

- humans have temporal masking of about 3ms
- humans hear frequency range of 20-20000Hz
- humans have frequency accuracy to 3.6Hz
- music contains sounds from 100-3200Hz (most often)

4 Solution

4.1 Feature Extraction

Three kinds of features were used in the scope of this project. There were many other potentially useful features that could be extracted from audio files for use in beat detection, but many of them required complex audio processing beyond the knowledge of the authors of this report. This project focused on power variance, bandwidth power variance, and change in peak frequency. To accommodate for audio temporal masking in human hearing the sample rate was decreased to a sample every three milliseconds. This was done by aggregating samples into a chunk and comparing that with the other aggregated samples in its neighborhood. A chunk of data is the number of samples calculated to span three milliseconds of time.

$$\text{chunk size} = \frac{\text{number of samples}}{300}$$

Each type of feature extraction generated data relative to the song it is in. This is due to the range of song genres considered in the scope of this project. Data comparisons using absolute values would confuse the training data when transitioning between songs, for example the electronic music has a much higher average frequency than dubstep. The solution to this was to compare features to a neighborhood of data to get relative values.

Since most songs go through various phases in which the pitch, tempo, and power levels can be drastically different the neighborhood of comparison was limited to one seconds worth of data. This value was chosen because it was small enough to capture the changes in features while still being large enough to contain data about the current features. One second is also the unit used when describing the tempo of a song when making a rough estimate. This allowed for a benchmark of what would be a reasonable number of beats in a second when comparing to the genre of the song.

4.1.1 Power Variance

The waveform of a sound file represents its data as energy levels within a channel at a given sample. The total power level of a sample can be calculated by summing the square over all channels in the waveform. This value is calculated as the average over a single chunk of data

5 Results

6 Conclusion

7 References