Combined Fluctuation Features For Kid’s Song Classification Based on Mood Parameters

Kadek Cahya Dewi

STMIK STIKOM BALI, Indonesia

cahya.amchy@gmail.com

Abstract

Music is closely related to human psychology. A piece of music often associated with certain adjectives such as happy, sad, romantic, and many more. The linkage between the music with a certain mood has been widely used in various occasions by people and music classification based on relevance to a particular emotion is important. This research concerns in music classification system based on mood parameters with combined fluctuation features. The mood parameters used is based on Robert Thayer's energy-stress model which are exuberance / happy, contentment / relax, anxious and depression. All feature sets are based on fluctuation of modulation amplitudes in psychoacoustically transformed spectrum data, namely the combination of rhythm patterns, rhythm histograms and statistical spectrum descriptors of the music. The system is tested using a set of song with various genre and the classification results are compared with the mood obtained by child psychology experts. Clustering and classification method obtained by Self Organizing Map method.

Keywords: Self Organizing Map, music classification, mood classification, rhythm patterns, rhythm histograms, statistical spectrum descriptor

1. INTRODUCTION

Music is an art, entertainment and human activities that involve the voices of regular. Music is all the possibilities that could happen to the voices / sounds and silence to be organized into a series of meaningful hearing. The meaning is not acquired verbal meaning, but the aural sense. The meaning of aural harmony means the perceived sound when listening to music. For example when listening a traditional song 'Yamko Rambe Yamko' which originated from Irian Jaya, listeners may not understand the intent of the song because they do not know the language that used in the lyrics or do not understand what's played instruments. However, when hear it is very possible as a matter of making nice to hear, making enthusiastic, sadness, grief or perhaps touching. Another example of the aural sense is when listening to an instrumental song.

Music is closely related to human psychology. A piece of music often associated with certain adjectives such as happy, sad, romantic, etc. The linkage between the music with a certain mood has been widely used in various occasions by men. For example, in a musical film is used to reinforce the atmosphere of the specific scene, the dramatic music used for background suspense scene, music scene eager for war, the music is fun to use as a background scene of humor, etc. In addition to support smart parenting programs will be very useful if the mothers know and understand the classification of music based on mood parameters, so they can do song selection in accordance with the desired conditions / moods. For example when the children wake up just choose the happy song (full of spirit). There are many more examples that can not be mentioned here.

A number of researches on music classification based on mood have been conducted. Feng et.al. classified music based on Dixon’s beat detection [3]. Li and Oghihara classified music using Support Vector Machines (SVM) [7], whereas Yang and Lee using agent system [10]. Leman et.al. based his classification on three level analysis: subjective judgments to manual-based musical analysis to acoustical-based feature analysis[6]. Other researchers include Wang et.al. which uses Support Vector Machines (SVM) [20], Wieczorkowska et.al. with K-Nearest Neighbor
This paper will discuss the classification of music based on relevance to a particular emotion/mood. It will be compared with the combined fluctuation features. All feature sets are based on fluctuation of modulation amplitudes in psychoacoustically transformed spectrum data, namely the combination of rhythm patterns, rhythm histograms and statistical spectrum descriptors of the music. The system is tested using a set of songs with various genres and in classification results are compared with the mood obtained by child psychology experts. Clustering and classification method obtained by Self Organizing Maps method.

2. IMPLEMENTATION

2.1 Emotion Model

The emotion model used is the Robert Thayer’s energy-stress model [10] which consists of exuberance / happy, contentment / relax, anxious / anxiety and depression. Here is in Figure 1 is the Thayer’s two dimensional model of emotion.

![Thayer's two-dimensional model of emotion](image)

2.2 Feature Extraction

The feature extraction process is the process to get the pattern of a song. The following is the feature extraction process in accordance with Rauber’s research [14] and the block diagram of audio extraction is shown by Figure 2.

2.2.1. Preprocessing:

b. Audio quality reduction. Audio quality is reduced from stereo to mono (sound exchange). Decreasing the number of channels from two to one and is done by taking the average of the two channels. Audio sampling rate is changed from 44 KHz to 11 KHz.
c. Splitting music into segments with each segment of size 6s in. Because of the time 6s deemed to have enough to get the impression from the style of a piece of music.

2.2.2. Feature Extraction for Rhythm Patterns:

a. Transformation into a spectrogram, by first performing an FFT. The FFT window size used is 256 samples to meet the 23ms sampling of mp3 (253 samples).
b. Groups frequencies into 24 critical frequency bands to meet the Bark scale. Bark scale ranges from 1 to 24 Barks.
c. Calculating the spectral masking effects with spreading function.
d. Transform into decibel to form a base 10 logarithmic scale.
e. Calculate the equal loudness level in Phon. Forty Phon = 40 db-SPL tone at 1 kHz frequency.
f. Calculating loudness sensation. One SONE = 1 kHz tone at 40 db-SPL.
g. Perform SONE transformation into rhythm patterns with the FFT.
h. Limit the amplitude modulation to 60. So that for every 24 critical bands 60 values is obtained for modulation frequencies between 0 to 10Hz. This results in 1440 values representing the fluctuation strength.
i. Filtering rhythm patterns using the Gaussian and the gradient.

2.2.3 Statistical Spectrum Descriptors

During feature extraction for the Rhythm Patterns, it was computed a Statistical Spectrum Descriptor (SSD) for the 24 critical bands. From the SONE representation of the spectrum (Sonogram), we compute the following statistical moments for each critical band: mean, median, variance, skewness, kurtosis, min-value and max-value, resulting in a 168-dimensional feature vector.
2.2.4 Rhythm Histograms

Contrary to the Rhythm Patterns and the SSD, this feature set does not contain information per critical band. The magnitudes of each modulation frequency bin of all 24 critical bands are summed up in order to form a histogram of modulation magnitude per modulation frequency. This feature set contains 60 attributes, according to modulation frequencies between 0.168 and 10 Hz.

2.2.5 Combination of Feature sets

In order to evaluate the results on my past research entitle Kid's Song Classification Based On Mood Parameters Using Rhythm Patterns Features, now we try to use the combination of feature sets. The different feature sets achieve largely different results depending on the database, i.e. the type of music contained in the collection. As a consequence we are interested in the performance of combined approaches, especially of the two sets with contrary results: SSD and Rhythm Histograms. The combination is expected to represent a more generalized feature set with potentially better results in a broader variety of musical styles. Moreover, we wanted to evaluate, whether classification without the much higher-dimensional Rhythm Patterns feature set could achieve comparable results. The following combinations of feature sets have been submitted to MIREX 2005 by Lidy and Rauber [8]:

- Rhythm Patterns + SSD (1608 dimensions)
- SSD + Rhythm Histograms (228 dimensions)
- Rhythm Patterns + SSD + Rhythm Histograms (1668 dimensions)

2.3 Classification

Classification is carried out using the Self Organizing Map. Classification is carried out after the feature extraction and clustering. Clustering is done by Self Organizing Map method and is carried out after the feature extraction process. The SOMeJB toolbox is used for clustering and visualization of the clustering results. Here are the parameters required for clustering:

1. The input is a 2-dimensional matrix \( n \times d \), where \( n \) is the number of data and \( d \) is the data dimensions resulted from feature extraction.
2. The output layer is a 2-dimensional matrix with the number of map units equals the number of terms / words on the Thayer emotion model.
3. Calculate the Euclidian distance.
4. Models the emotion as Thayer’s emotion model which consists of exuberance, contentment, anxious and depression. The terms are: exuberant, triumphant, carefree, anxious, frantic, terror, content, serene, ominous, and depression.
5. Visualization of the clustering results with smoothed data histogram.

Classification with the SOM method is performed by entering testing data on the network and labeling moods according to the results of the clustering. Classification process is done by SOMeJB Toolbox. The results of the classification
using the SOM methods were compared with the results obtained by child psychology experts.

3. RESEARCH RESULTS AND DISCUSSION

Clustering process was tested on 120 songs in Indonesian language and in English with a variety of genres using SOM method. Clustering performed three times in accordance with the number of feature combinations. First, clustering based on combination of rhythm patterns (RP) and statistical spectrum descriptors (SSD) with 1608 dimensions generate by 1440 dimensions from RP and 168 dimensions from SSD. Second, clustering based on combination of statistical spectrum descriptors (SSD) and rhythm histograms (RH) with 228 dimensions generate by 168 dimensions from SSD and 60 dimensions from RH. Third, clustering based on combination of rhythm patterns (RP), statistical spectrum descriptors (SSD) and rhythm histograms (RH) with 1668 dimensions generate by 1440 dimensions from RP, 168 dimensions from SSD and 60 dimensions from RH.

The clustering results are visualized in the form of islands of music. The songs are placed according to the level of sound similarity and islands formed in accordance with the cluster density. Islands of music using a color code with color levels range from dark blue (deep sea) to light blue (shallow water) to light green (hills) to gray (rocks) and white (mountain). Each island which is marked by passage of land / forest represents a different style of music (different cluster). Mountains indicate the center of the cluster, so if there is an island with more than one mountain, it can be said as a style in a style (sub-style).

Classification process was tested on 25 songs in Indonesian and English language with a variety of genres using SOM method. Classification results done by three combinations of features and then are compared with data mood from the psychological expert. The classification results in this research achieved very similar results for all three combinations of features, but the combination of rhythm patterns and statistical spectrum descriptors is the best one. The result accuracy that we get by comparing the number of correct results with the number of songs showing the data for the RP-SSD feature yields 80%, SSD-RH yields 76% and RP-SSD-RH yields 76%. The comparison of classification results can be seen in table 1. Here is the explanation of symbols on the table 1:

- 1 for mood “exuberance”
- 2 for mood “contentment”
- 3 for mood “anxious”
- 4 for mood “depression”

From table 1 also shows that the most misclassification occurs in moods “exuberance”. Level of accuracy for the mood of “exuberance” in all combination of features is 54%. The accuracy on the mood of contentment in all features combination is 88.89%. The level of accuracy in the mood "anxious” on the RP-SSD features is 100%, while in SSD-RH and RP-SSD-RH is 75%. The accuracy of the mood depression on all combination of features is 100%.

<table>
<thead>
<tr>
<th>No</th>
<th>Song’s Title</th>
<th>RP-SSD</th>
<th>SSD-RH</th>
<th>RP-SSD-RH</th>
<th>Mood Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>children music - london bridge.mp3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Tasya - Ambilikan Bulan Bu.mp3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Alphabet song.mp3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Twinkle2.mp3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Ibu Pertiwi.mp3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>The Lion king -06- This land.MP3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>01. Two Voice, One Song.mp3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>02. Connected – Melody.mp3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>Anastasia-10-Once upon A December (Deana Carter).mp3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>02- Reflection.mp3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>01 Armada Masa Depan.MP3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>03. Believe - Melody.mp3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>04. Wonderful Me.mp3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1. Comparison of classification results
4. Conclusions

The research results show that the system for music classification based on mood parameters could be developed by Self Organizing Map method using with combined fluctuation features. In this research music is classified by sound similarity obtained from the combination of rhythm patterns, rhythm histograms and statistical spectrum descriptors based on mood parameters according to Thayer's theory. The classification results in this research achieved very similar results for all three combinations of features, but the combination of rhythm patterns and statistical spectrum descriptors is the best one. The result accuracy for the RP-SSD feature yields 80%, SSD-RH yields 76% and RP-SSD-RH yields 76%.

5. References


