Extraction Of Audio Features For Emotion Recognition System Based On Music

Kee Moe Han, Theingi Zin, Hla Myo Tun

Abstract: Music is the combination of melody, linguistic information and the vocalist's emotion. Since music is a work of art, analyzing emotion in music by computer is a difficult task. Many approaches have been developed to detect the emotions included in music but the results are not satisfactory because emotion is very complex. In this paper, the evaluations of audio features from the music files are presented. The extracted features are used to classify the different emotion classes of the vocalists. Musical features extraction is done by using (Music Information Retrieval) MIR tool box in this paper. The database of 100 music clips are used to classify the emotions perceived in music clips. Music may contain many emotions according to the vocalist's mood such as happy, sad, nervous, bored, peace etc. In this paper, the audio features related to the emotions of the vocalists are extracted to use in emotion recognition system based on music.

Keywords: Vocalist's emotion, Audio features, MIR tool box, English songs, Emotion recognition system.

I. INTRODUCTION

Many people of different ages are fond of listening to music. Music can influence people's mind, that's why mothers sing songs to console their children. It is obvious that since music itself contains some expressions of the vocalists, it can manipulate our mind. From digital signal processing point of view, music is the combination of linguistic information, vocalist tone and emotion. In this paper, audio features from the input music files are extracted by using MIR toolbox [1] and these are intended to use in emotion classification system. There are different kinds of emotions: happy, sad, angry, excited, bored etc. Furthermore, emotion models can be divided into two types: categorical and dimensional. Categorical approach is describing several stages of emotion and dimensional approach presents several axes to map emotion into a plane. Both types can be adopted in classification of music signals and their emotion states can be classified by using the extracted features. Many audio features can be extracted by using several tool boxes and algorithms. Loudness, rhythm, MFCC (mel-frequency cepstral coefficients) and beat are the common audio features. The audio features extracted in this paper are show in Table 1. This paper is organized as follow: section 2 reviews related work, section 3 is the implementation plan, section 4 is about construction and preprocessing of training data set, section 5 is features extraction by using MIR tool box, section 6 is simulation results and the rest are discussions and conclusion.

TABLE 1
EXTRACTED AUDIO FEATURES

Features	No. of features	General description of audio features
Pitch	1	A term used to describe high or low of a note played by musical instruments
Tempo	1	The pulse of varying strength described in terms of tempo
Tonality	1	The relation between the notes of a scale or key
dynamic	1	The amplitude of a sound in rms energy
Timbre	13	Tone colour or tone quality of sound from its pitch and intensity
Total	17	

II. RELATED WORKS

Many classifiers and feature extraction tools can be used to classify emotions from the music files. Some used regression approach. Yi-Hsuan Yang [2] used multiple linear regression (MLR), support vector regression (SVR), and AdaBoost. RT (Boost.R).Many feature extraction algorithms can be used for audio feature extraction such as Psysound, Maryas , Spectural contrast and DWCH .Chia-Chu Liu[3] presented classification of emotion from popular music by using PsySound2 and Nearest Mean classifier. Chinese and Japanese 243 popular songs are chosen for database. Nadia Lachetar [4] used song title and lyrics as features and then Naive Bayes and Ant Colony algorithm as classifiers to classify emotion. Bram van de Laar [5] described emotion detection in music survey which compared many detection methods. Adit Jamdar, Jessica Abraham, Karishma Khanna [6] presented emotion analysis of songs based on lyrical and audio features tempo, mode, loudness, danceability and energy. Natural language processing is used for lyrics analysis. The classification is done by K-Nearest Neighbors algorithm.

III. IMPLEMENTATION PLAN

The system can be divided into two stages: the classifier training and testing. The system block diagram is shown in below Fig.1.The details of the system are described in following. First of all, music files (all types of English songs including pop, rock and classical etc) have to preprocess to make the training data set. After preprocessing, MIR (Music Information Retrieval) tool box is installed to extracted audio features. There are many tool boxes which can be used as feature extraction tools. MIR tool is chosen because of the following:

- Integrated set of function written in Mat lab
- Easy to use complex computation
- Reduce syntax complexity
- Batch of files can be processed

In subjective tests, 15 people, who are college students, are asked to listen to the songs and collected the advices which are one of the four emotions. Then, the average emotion class of the subject test is defined as the class label for the related audio signal. After that, a classifier is used to classify the emotions of the vocalists. From the review of the previous works, many classifiers can be used to classify emotion from music signal. SVM (support vector machine), MLR (Multi level

regressor), k-NN (k-nearest neighbor methods) and neural network are common types of classifiers.

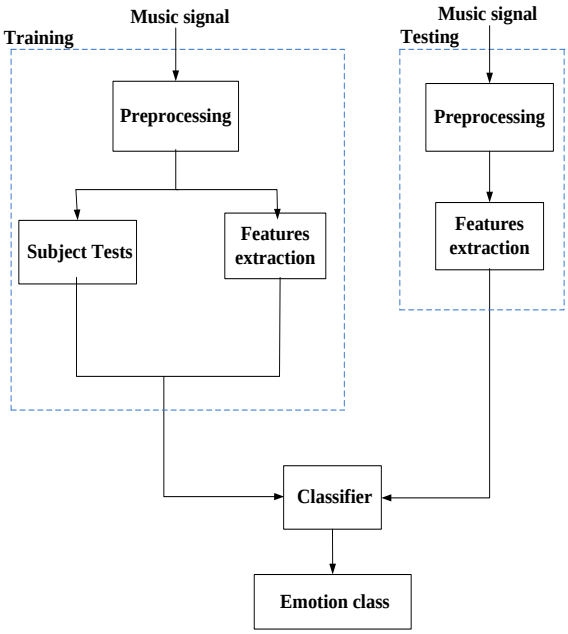

Fig. 1 *Training and testing system block diagram*

IV. PREPROCESSING OF TRAINING DATA SET

Firstly, all of the input songs are trimmed to two minutes and converted to a uniform format which is mono channel, wav files. After preprocessing, most college students are tested to listen to a subset of music files and to choose one of the four emotions. Then the ground truth is set as the average of these subjective tests. By this way, training data set is available by collecting the subjective tests and extracted audio features for each music sample.

V. EXTRACTED FEATURES FROM MIR TOOL BOX

MIR toolbox can easily be installed in Matlab. This toolbox is quite easy to use and reduces computation complexity.

A. Pitch

The frequency of a sound wave is what ears understand as pitch. A higher frequency sound has a higher pitch and a lower frequency sound has a lower pitch. The pitch frequency can be calculated by using auto correlation method in the tool box. The pitch periods of a given music file is computed by finding the time lag corresponds to the second largest peak from the central peak of autocorrelation sequence. Then pitch frequency is estimated from the pitch periods [7].

B. Tempo

The audio signal is first decomposed into auditory channels using bank of filters. The envelope of each channel is extracted. Then the envelopes are differentiated, half wave rectified, before being finally summed again. This gives a precise description of the variation of energy produced by each event from the different auditory channels. After onset detection, the periodicity is estimated through autocorrelation shown in Fig.2.The periodogram is filtered using a resonance curve, after which the best tempo is estimated through peak picking, and the result is converted into beat per minutes [7].

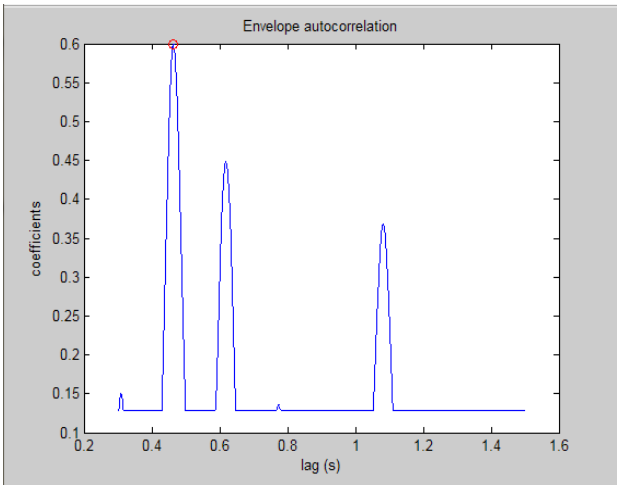

Fig .2 *Peak picking from envelope autocorrelation*

Tempo (bpm) =fs 60/ lag index,*
Where, fs=sampling frequency

C. Tonality

The spectrum is converted from the frequency domain to the pitch domain by applying a log-frequency transformation. The distribution of the along the pitches is called the chromagram. The chromagram is wrapped by fusing the pitches belonging to the same pitch classes. The wrapped chromagram shows a distribution of the energy with respect to the twelve possible pitch classes. Fig. 3 (a) describes the standard C major keys and Fig.3 (b) shows the rotating right shift of the standard keys to get the magnitudes of the other keys.Standard minor keys candidates can be seen in Fig.4(a) and Fig4.(b) is shifting right to the C minor keys candidates to get the other minor keys[8].Emilia Gomez Gutierrez, Krumhansl and Schmuckler (Krumhansl, 1990) [9] proposed a method for estimating the tonality of a musical piece by computing the cross correlation of its pitch class distribution through listening experiments. The most prevalent tonality is considered to be the tonality candidate with highest correlation.

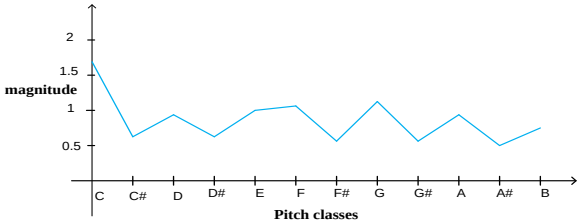

Fig.3(a) Standard C major keys candidates

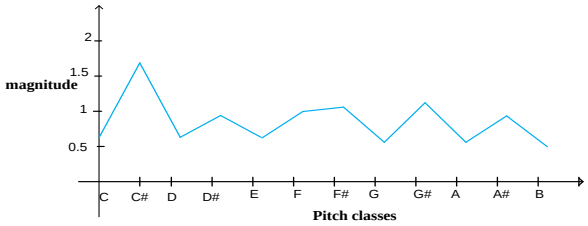

Fig .3(b) *Rotate right-shift of the standard C major key candidates*

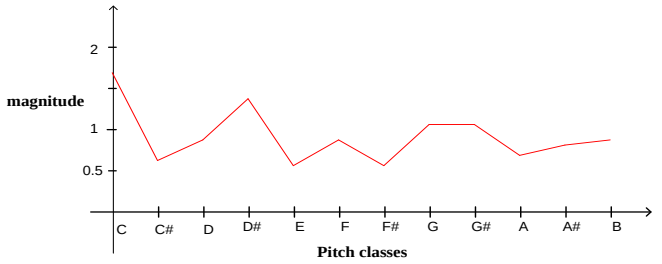

Fig.4(a) Standard C minor key candidates

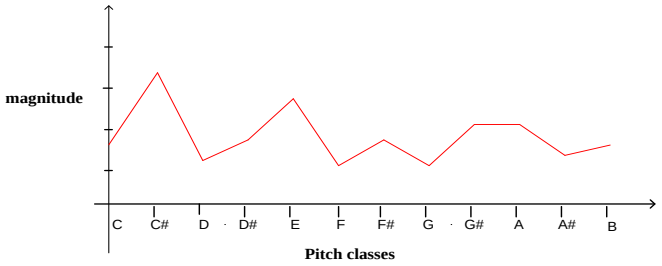

Fig .4(b)*Rotate right-shift of the Standard C minor key candidates*

D. Dynamics

The energy of the signal can be computed simply by taking the average of the square of the amplitude, also called root-mean-square (RMS) [10].Energy can be computed by using equation (1).

$$x_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \qquad (1)$$

Where x=input discrete signal
n=the length of x

E. Timbre

One common way of describing timbre is based on MFCCs, which are mel-frequency cepstral coefficients. First the audio sequence is described in the spectral domain to the Mel-scale domain: the frequencies are rearranged into 40 frequency bands called Mel-bands[11] .The conversion of frequency to mel-scale is shown in equation (2).

$$mel(f) = 2595\log(1+\frac{f}{700}) \qquad (2)$$

Where f=frequency (Hz)

According to the Mel-scale, a set of triangular band pass filters are used to compute a weighted sum of filter spectrum components, so the output of the process approximates to a Mel-scale [12]. The envelope of the Mel-scale spectrum is described through a Discrete Cosine Transform. The values obtained through this transform are the MFCCS. The 13 first ones are used because only a restricted number of them should be selected.

The formula of DCT is

$$C_m = \sum_{k=1}^{N} E_k.\cos[m*(k-\frac{1}{2})\frac{\pi}{N}], \qquad (3)$$

Where m=1,2,...,13.

VI. Simulation Results

The following Fig.5 is an example of extracted features and emotion from the subjective tests of a music file to make the training data set.

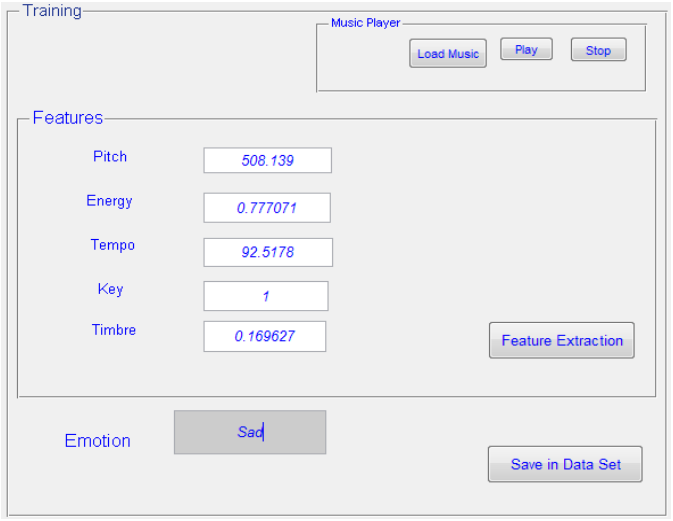

Fig.5 *Example of training data set of a music file*

TALBLE 2
SAMPLE TRAINING DATA SET

No.	Pitch	Energy	Tempo	Key	Timbre	Emotion
1	409.578	0.595149	118.129	10	0.0637567	sleepy
2	504.246	0.667847	118.129	9	0.315436	sleepy
3	922.301	0.553658	118.533	3	0.215653	excited
4	880.956	0.430718	179.333	11	0.0861522	excited
5	508.139	0.777071	92.5178	1	0.169627	sad
6	210.999	0.617754	100.048	1	0.0699997	happy
7	992.679	0.578646	127.007	8	0.13195	happy
8	721.62	0.561451	117.873	5	0.242666	happy
9	157.25	0.644389	115.792	1	0.0771521	excited
10	369.034	0.537607	99.8934	5	0.0903507	sad

The emotion of the music signal can be classified through the dataset with a probabilistic classifier. Adding more songs in the actual music database and using more emotional related features such as lyrics can improve classification accuracy.

VII. DISCUSSIONS

After the audio features are extracted by signal processing techniques and algorithms, the training data set is got to use in classification. In this paper, MIR tool box is used in musical features extraction. The sample data set is shown in Table 2. As can be seen in Table 2, most happy and excited songs have high pitches than the others. The sad songs correlate with slow tempo but happy songs are generally fast. The key is represented by 1 to 12 according to the 12 pitch classes. Timbre is the mean value of the 13 mel-frequency coefficients. Then, the data set can be used in classification tasks of music and emotion classification of songs.

VIII. CONCLUSION

Music emotion recognition system can be evaluated using various machine learning classification algorithms and digital signal processing techniques. Audio features can also be extracted by using many tool boxes and methods. Feature extraction from the music signal in emotion recognition system is important step because testing music signals are also needed to extract audio features and the performance of the classifier depends on the extracted features. Extraction of audio features using MIR tool box and making data set for classification are presented in this paper. Music dataset can be got from the extracted features and subjective test.

ACKNOWLEDGEMENT

The author is very pleased to express her deep gratitude to his Excellency, Minister Dr. Ko Ko Oo, Ministry of Science and Technology, for the opening of Master of Engineering Course at Mandalay Technological University. The author is greatly thankful to her supervisor, Daw Theingi Zin, Lecturer, Department of Electronic Engineering, Technological University, for her excellent supervision, true guideline and valuable suggestion in writing this paper.

REFERENCES

[1] (2011)[Online]Available: http://www.mathworks.com/matlabcentral/fileexchange/24583mirtoolbox/

[2] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen, **"**Music Emotion Classification: A Regression Approach**",** in Multimedia and Expo, IEEE International Conference, 2007, pp.208-211.

[3] C.C.Liu, Y.H .Yang, P.H. Wu, and H.H. Chen, "Detecting and Classifying Emotion in Popular Music" in JCIS.2006.

[4] Nadia Lachetar ,Halima Bahi, "Song classification", Computer Science Department.

[5] Bram van de Laar , " Emotion detection in music, a survey" ,in Twente Student Conference on IT, 2006,vol 1,p.700.

[6] Adit Jamdar, Jessica Abraham, Karishma Khanna and Rahul Dubey ,"emotion analysis of songs based on lyrical and audio features", International Journal of Artificial Intelligence & Applications (IJAIA) ,Vol. 6, No. 3, May 2015.

[7] Olivier Lartillot , Petri Toiviainen , "A matlab toolbox for musical feature extraction from audio", in Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15, 2007.

[8] Emilia Gomez Gutierrez "Tonal Description of music audio signals", PhD Thesis, University of Pompeu Febra, Barcelona, 2006.

[9] Erecetin, Sefika, Sule, Banerjee, Santo, "Chaos, Complexity and Leadership 2013", Springer.

[10] Christine Preisach, Hans Burkhardt, Lars Schmict-Thieme, Reinhold Decker, "Data Analysis, Machine Learning and Applications", Proc .of the 31st Annual Conference of the Gesellschaft fur Classification e..March 7-9, 2007.

[11] Hima Deepthi Vakayalapati, Koteswara Rao Anne and Swarna Kuchibhotla, "Acoustic Modeling for Emotion Recognition", Springer, March 14,2015.