ISSN: 2321-7782 (Online)
Volume 2, Issue 4, April 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Emotion Recognition Based on MFCC Features using SVM

E. Vijayavani[1]
Department of Information Technology
E. G. S. Pillay Engineering College
Nagapattinam, Tamil Nadu – India

S.Lavanya[2]
Department of Information Technology
E. G. S. Pillay Engineering College
Nagapattinam, Tamil Nadu – India

P. Suganya[3]
Department of Information Technology
E. G. S. Pillay Engineering College
Nagapattinam, Tamil Nadu – India

E. Elakiya[4]
Department of Computer Science and Engineering
E. G. S. Pillay Engineering College
Nagapattinam, Tamil Nadu – India

Abstract: ***Music oftentimes referred to as a language of emotion and hence music emotion could be useful in music understanding, retrieval and some other musical related applications. This paper discusses the method to extract features from samples, and using those features, to detect the emotion. we focus on challenging issue of recognizing music emotions such as happy, sad, anger, fear, and neutral. Musical data is collected from various areas. A mel frequency cepstral coefficient (MFCC) is extracted as a feature from the data collected. These features result in different MFCC coefficients that are input to the support vector machine (SVM), which will analyze them with the stored database recognize the emotion. Data are collected from various websites and referred using recorded data.***

Keywords: ***Mel Frequency cepstral coefficient, support vector machine, Thayer's model.***

I. INTRODUCTION

Many issues for music emotion recognition have been addressed by different disciplines such as physiology, psychology, and musicology. In this paper, the challenging issue of recognizing music emotions based on subjective human emotions and acoustic music signal features and present an intelligent music emotion recognition system is focused.

Hence, one of the most important prerequisites for realizing such an advanced user interface is a reliable emotion recognition system that guarantees acceptable recognition, robustness, and adaptability to practical applications. To develop such a system requires the following stages: modelling, analysing, processing, training, and classifying emotional features measured from the implicit emotion channels of human communication, such as speech, facial expression, physiological responses, etc.

A. Music and Emotion

Automatic emotion detection and recognition in speech and music is growing rapidly with the technological advances of digital signal processing and various effective feature extraction methods. Emotion recognition can play an important role in many other potential applications such as music entertainment and human-computer interaction systems. Many researchers have explored models of emotions and factors that give rise to the perception of emotion in music. Many other researchers investigate the problem of automatically recognizing emotion in music. Traditional mood and emotion research in music has focused on finding psychological and physiological factors that influence emotion recognition and classification. During the 1980s, several emotion models were proposed, which were largely based on the dimensional approach for emotion rating.

The dimensional approach focuses on identifying emotions of dimensions such as valence and activity. Thayer suggested a two dimensional emotion model that is simple but powerful in organizing different emotion responses. The dimension of stress

is called valence while the dimension of energy is called arousal. During the last decade, many researchers have investigated the influence of music factors like loudness and tonality on the perceived emotional expression.

The two major approaches to emotional modelling that exist in the field are categorical and dimensional. A dimensional approach classifies emotions along several axes, such as valence (pleasure), arousal (activity), and potency (dominance).

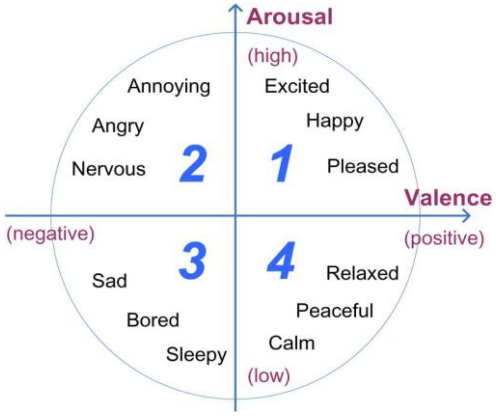


Fig 1 Thayer's arousal-valence emotion plane.

B. Musical Features

Scale is an overall rule of tonic formation of music. A set of key, mode, and tonality is defined as scale. For accurate scale features, first analyse the chromagram for representing the frequencies in musical scales. After that, apply the key profile matrix. The following equations show the process of combining chromagram and key characterization

$$\text{Tonality} = C\_KeyProfileMatrix \quad (1)$$

$$\text{Key} = \max(\text{Tonality}(\text{Idx})) \quad (2)$$

key Index

Where vector C has 12 elements and represents the summed chromagram analyzed for each acoustic frame. KeyProfileMatrix is composed of 12-by-24 elements. Key Index indexes KeyProfileMatrix, where KeyIndex=1, 2,…,24. After the inner product of C and KeyProfileMatrix in Equation (1), we obtain a tonality score for each key. Finally, the most appropriate key by picking the key having maximum tonality in Equation (2) is obtained.

Average energy (AE) of the overall wave sequence is widely adopted to measure the loudness of music. Also, standard deviation (σ) of AE measures the regularity of loudness. Those are defined as:

$$AE(x)=\frac{1}{N}\sum_{i=0}^{N} x(t)^2 \,, \sigma\big(AE(x)\big) = \sqrt{\frac{1}{N}\sum_{t=0}^{N}(AE(x)-x(t))^2}$$

where x is an input discrete signal, t is the time in samples, and N is the length of x in samples.

Rhythm, which is composed of rhythmic features such as tempo and beat. Beat is a fundamental rhythmic element of music. Tempo is usually defined as the beats per a minute (BPM) which is used to represent the global rhythmic feature of music. Tempo and regularity of beats can be measured in various ways. For beat tracking and tempo analysis, the algorithm by Ellis is used. The features are overall tempo (in beats per minute) and the standard deviation of beat intervals, which indicates tempo regularity.

Harmonics can be observed in musical tones. Harmonics are easily observed in the spectrogram in monophonic music,. However, in polyphony it is hard to find harmonics, because many instruments and voices are performed at once. To solve this problem, a method to compute harmonic distribution yields

$$HS(f)=\sum_{k=1}^{M} \min\left(\|x(f)\|, \|x(kf)\|\right)$$

Here, M denotes the maximum number of harmonics considered, f is the fundamental frequency, X is the short-time Fourier transform (STFT) of the source signal. In the equation, the min function is used in such a way that only the strong fundamental and strong harmonics result in a large value for HS.

II. Database Collection

The detection of emotion in music is modelled as a classification task. The musical database is collected from various websites. To classify the data sets of music clips have taken 20 seconds in duration is used .In each category we have taken 5 samples. The following sub sections present the features that were extracted from each wave file and the emotion labelling process. To compare the segments fairly the music samples are converted to a uniform format.

III. Feature Extraction

Transforming the input data into the set of features is called feature extraction. If the features extracted are chosen carefully it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction includes feature construction, space dimensionality reduction, sparse representations, and feature selection. In order to better classification of emotion in the music data set, we consider the basic features such as intensity, scale, harmony, energy, pitch, formant frequencies, etc. all these are prosodic features. Here in feature extraction process an extracted feature is Mel Frequency Cepstral Coefficient (MFCC).

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum.

In viewing the Mel Frequency Cepstral Coefficients (MFCCs) the dominant features used for speech recognition and investigate their applicability to modelling music. In particular, two of the main assumptions of the process of forming MFCCs: the use of the Mel frequency scale to model the spectra; and the use of the Discrete Cosine Transform (DCT) to decorrelate the Mel-spectral vectors.

A. Mel-Frequency Cepstral Coefficient (Mfcc)

All the human generated sounds which influence our lives, speech and music are arguably the most prolific. Speech has received much focused attention and decades of research in this community have led to usable systems and convergence of the features used for s. In the music community however, although the field of synthesis is very mature, a dominant paradigm has yet to emerge to solve other problems such as music classification or transcription.

MFCCs are short-term spectral features. Which are calculated as follows

1. Divide signal into frames.
2. For each frame, obtain the amplitude spectrum.
3. Take the logarithm.
4. Convert to Mel (a perceptually-based) spectrum.
5. Take the discrete cosine transform (DCT).

In order to better classification of emotion in the music data set, we consider the basic features such as intensity, scale, harmony, energy, pitch, formant frequencies, etc. all these are prosodic features. Here in feature extraction process extracted features are Mel Frequency Cepstral Coefficient (MFCC). Fig. 2 shows the MFCC feature extraction process. Feature extraction process contains following steps:

• **Preprocessing:** The continuous time signal is sampled at sampling frequency. At the first stage in MFCC feature extraction is to boost the amount of energy in the high frequencies. This pre emphasis is done by using a filter.

• **Framing**: It is a process of segmenting the music samples obtained from the analog to digital conversion (ADC), into the small frames with the time length within the range of 20-40 ms. Framing enables the non stationary music signal to be segmented into quasi-stationary frames, and enables Fourier Transformation of the audio signal. It is because, audio signal is known to exhibit quasi-stationary behavior within the short time period of 20-40 ms

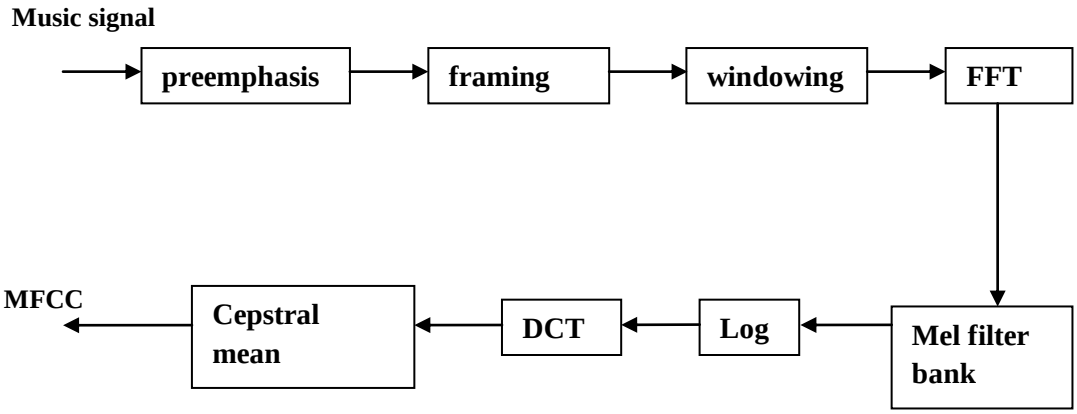


Fig 2 MFCC Block Diagram

• **Windowing**: It is the process to window each individual frame, in order to minimize the signal discontinuities at the beginning and the end of each frame.

• **FFT**: Fast Fourier Transform (FFT) algorithm is ideally used for evaluating the frequency spectrum. FFT converts each frame of N samples from the time domain into the frequency domain.

• **Mel Filterbank** and Frequency wrapping: The Mel filter bank consists of overlapping triangular filters with the cutoff frequencies determined by the center frequencies of the two adjacent filters. The filters have fixed bandwidth and linearly spaced centre frequencies on the Mel scale.

• **Log**: The logarithm has the effect of changing multiplication into addition. Therefore, this step simply converts the multiplication of the magnitude in the Fourier transform into addition.

• **Discrete Cosine Transform**: It is used to orthogonalise the filter energy vectors. Because of this orthogonalization step, the information of the filter energy vector is compacted into the first number of components and shortens the vector to number of components.

IV. CLASSIFICATION USING SVM

Support vector machines (SVM) is based on the principle of empirical risk minimization i.e., minimization of error on training data. For linear separable data SVM finds a separating hyper lane which separates the data with the largest margin. For linearly separable data, it maps the input pattern space X to a high dimensional feature space Z using a non linear function. Then the SVM finds optimal hyper plane as the decision surface to separate the examples of two classes in the feature space. The SVM in particular defines the criterion to be looking for a decision surface that is maximally far away from any data point. This distance from the decision surface to the closest data point determines the margin of the classifier. This method of construction necessarily means that the decision function for an SVM is fully specified by a (usually small) subset of the data which defines the position of the separator.

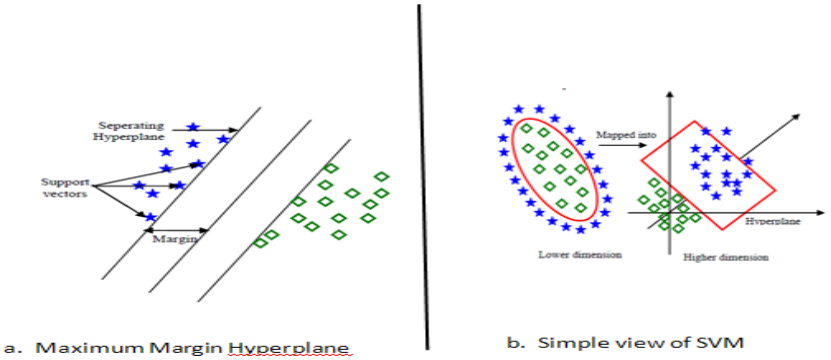


Fig: Classification using SVM

V. Experimental Results and Discussion

Several machine learning methods have been applied to this task; use of SVMs has been prominent. Support vector machines are not necessarily better than other machine learning methods (except perhaps in situations with little training data), but they perform at the state-of-the-art level and have much current theoretical and empirical appeal. Discovering emotions in music is a difficult issue for many reasons. First of all, emotions perceived with music are subjective and depend on numerous factors. Additionally, the same piece may evoke various emotions may feel various emotions when listening to this same piece of music. The purpose of our research was to perform automatic recognition of emotions such as happy, sad, anger, fear, and neutral in music, and the special collection of music pieces was gathered and turned into a data set. In order to test the performance of the emotion classification in music, several audio records are considered. A total of 25 samples of music clips are taken as dataset were used in our studies. The MFCC feature vectors are extracted for all the music samples. For each analysis the distribution of the feature vectors is captured using an SVM model as described in Section 4. The feature vectors are given as input to the SVM and trained for 2000 epochs. One epoch of training is a single presentation of all the training vectors to the network. The feature vectors are given as input to the SVM model and the average confidence score is calculated. An experimental result shows SVMN Classifier performs better, which offers a new efficient way of solving problems.

VI. Conclusion

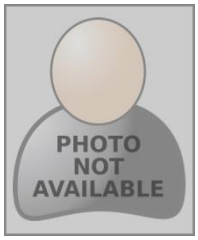
Recognizing musical mood remains a challenging problem primarily due to the inherent ambiguities of human emotions. Accurate detection of emotion from music has clear benefits for the design of music related application or for the extraction of useful information from large quantities of music data. It is also known that each classifier has its own advantages and disadvantages. A Mel frequency cepstral coefficient (MFCC) is extracted as a feature from the data collected. A neural network classifier was applied to evaluate the classification performance and recognize different emotional states of these utterances. An experimental result shows SVM Classifier performs better, which offers a new efficient way of solving problems. Further research should include precisely and efficiently add more features and a more detailed evaluation of system performance.

References

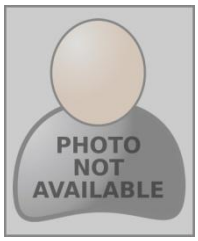
1. Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, “A Regression Approach to Music Emotion Recognition,” IEEE Trans. on Audio, Speech, and Language Processing, vol. 16, no. 2, pp. 448-457, Feb. 2008.
2. K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas,“Multilabel classification of music into emotion,” in Proc. Of the Intl. Conf. on Music Information Retrieval, Philadelphia,PA, 2008.
3. B. Logan, “Mel frequency cepstral coefficients for music modeling,” in Proc. of the Intl. Symposium on Music Information Retrieval, Plymouth, MA, September 2000.
4. T. Li and M. Ogihara, “Detecting emotion in music,” in Proc. of the Intl. Conf. on Music Information Retrieval, Baltimore,MD, October 2003.
5. Yegnanarayana, B. and Kishore, S.P., “AANN: an alternative to GMM for pattern recognition”, Neural Networks, 15(3):459–469,2002.
6. B. Han, S. Rho, R. B. Dannenberg, and E. Hwang, “SMERS:Music emotion recognition using support vector regression,”in Proc. of the Intl. Society for Music Information Conf.,Kobe, Japan, 2009.
7. Jaume Padrell-Sendra, Dar´ıo Mart´ın-Iglesias and Fernando D´ıaz-de-Mar´ıa., “Support Vector Machines for continuous speech recognition ,” in Proc. of the 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP 2006, vol. 4, pp. 504–507.
8. Oh-Wook Kwon, Kwokleung Chan, Jiucang Hao, Te-Won Lee,” Emotion Recognition by Speech Signals,” Institute for Neural Computatio University of California, San Diego, USA 2003.

9. Xia Mao, Lijiang Chen, Bing Zhang," Mandarin speech emotion recognition based on a hybrid of HMM/ANN," international journal of computers Issue 4, Volume 1, 2007.
10. L. Lie, D. Liu and Hong-Jiang Zhang: "Automatic Mood Detection and Tracking of Music Audio Signals," IEEE Trans. on ASLP, Vol. 14(1), 2006.
11. Y.-H. Yang, C.-C Liu, and H. H. Chen, "Music emotion classification: A fuzzy approach," Proc. ACM Multimedia, Santa Barbara, USA, pp. 81–84, 2006.

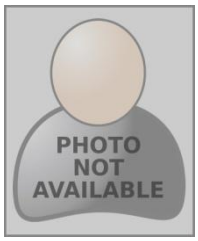
AUTHOR(S) PROFILE



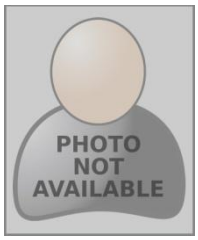
E. Vijayavani received her B. Tech., IT from Anna University, Chennai in 2010 and M.E Computer Science and Engineering from Annamalai University, Chidambaram in 2012. At present working as Assistant Professor at E. G. S. Pillay Engineering College, Nagapattinam. Her research interest includes image processing and speech recognition in general.



S. Lavanya received her B. Tech., IT from Anna University, Chennai in 2010 and M.E Computer Science and Engineering from Anna University, Chennai in 2012. At present working as Assistant Professor at E. G. S. Pillay Engineering College, Nagapattinam. Her research interest includes text classification and information security in general.



P. Suganya received her B. E., CSE from Anna University, Chennai in 2010 and M.E Computer Science and Engineering from Anna University, Chennai in 2012. At present working as Assistant Professor at E. G. S. Pillay Engineering College, Nagapattinam. Her research interest includes web services, SOA and Mashup in general.



E. Elakiya received her B. Tech., IT from Anna University, Chennai in 2010 and M.E Software Engineering from Anna University, Chennai in 2012. At present doing Ph.D at Anna University and working as Assistant Professor at E. G. S. Pillay Engineering College, Nagapattinam. Her research interest includes information retrieval, text classification and text mining in general.

