An Efficient Classification Algorithm For Music Mood Detection In Western and Hindi Music Using Audio Feature Extraction

Aathreya S. Bhat
Electronics and Communication Engineering
BNM Institute of Technology
Bangalore, India
s_aathreya@yahoo.com

Amith V. S.
Electronics and Communication Engineering
BNM Institute of Technology
Bangalore, India
amith.vs29@gmail.com

Namrata S. Prasad
Electronics and Communication Engineering
BNM Institute of Technology
Bangalore, India
namratasp3@gmail.com

Murali Mohan D.
Audience,
Bangalore, India
dmmohan@gmail.com

***Abstract*— Over the past decade, a lot of research has been done in audio content analysis for extracting various kinds of information, especially the moods it denotes, from an audio signal, because music expresses emotions in a concise and succinct way, yet in an effective way. People select music in congruence to their moods and emotions, making the need to classify music in accordance to moods more of a demand. Since different individuals have different perceptions about classifying music according to mood, it becomes a much more difficult task. This paper proposes an automated and efficient method to perceive the mood of any given music piece, or the "emotions" related to it, by drawing out a link between the spectral and harmonic features and human perception of music and moods. Features such as rhythm, harmony, spectral feature, and so on, are studied in order to classify the songs according to its mood, based on Thayer's model. The values of the quantified features are then compared against the threshold value using neural networks before classifying them according to different mood labels. The method analyzes many different features of the music piece, including spectra of beat and roughness, before classifying it under any mood. A total of 8 different moods are considered. In particular, the paper classifies both western and Indian Hindi film music, taking into consideration, a database of over 100 songs in total. The efficiency of this method was found to reach 94.44% at the best.**

Keywords- Mood detection, music emotion, Thayer's model, feature extraction.

I. INTRODUCTION

Music is considered as the best form of expression of emotions. The music that people listen to is governed by what mood they are in. The characteristics of music such as rhythm, melody, harmony, pitch and timbre play a significant role in human physiological and psychological functions, thus altering their mood.[1] By studying the various characteristics of the song, it is possible to group them according to different moods specified in Thayer's mood model. The model is depicted in Fig 1.

Although some information like album name, artist name, genre, are present in the meta-data of the music clip, their significance is limited when it comes to music related applications like creation of play-list based on the mood of the listener. To address such scenarios, properties such as beat, mood and tempo need to be studied.[2] Structural features are divided into two parts, segmental features and suprasegmental features. Segmental features are the individual sounds or tones that make up the music; this includes acoustic structures such as amplitude, and pitch. Suprasegmental features are the foundational structures of a piece, such as melody, tempo and rhythm.

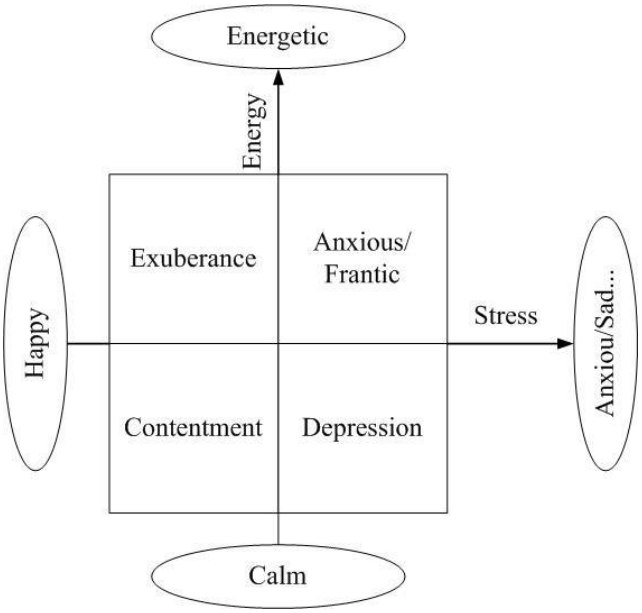

Fig 1: Thayer's mood model

Fig 1 is the model presented by Thayer for mood classification.

978-0-7695-5100-5/13 $31.00 © 2013 IEEE
DOI 10.1109/ICSIP.2014.63

The variation in these parameters influences the mood of the listener in the following ways:

Fast tempo is associated with happiness or excitement (or even anger). Slow tempo may be associated with sadness.[3][4][5][6][7] Loudness, or the physical strength and amplitude of a sound, may be perceived as intensity, power, or anger; while soft music is associated with tenderness, sadness, or fear.[3] Rapid changes in loudness may connote playfulness or pleading, whereas few or no changes can indicate peace and sadness.[3] High pitch implies light, happy, carefree, funny moods while a low one indicates dark, sad, ominous and serious situations. Bright timbres coupled with loud dynamics affect moods of vigor, turmoil, conflict, and valor. The same timbres coupled with soft dynamics affect moods of sensuality, passion and compassion. Dark timbres coupled with loud dynamics are encountered only occasionally in music and affect moods of starkness. Timbre stimulates human energy levels without regard to rhythmic or harmonic saturation.[6] Sound sources that have simple harmonic profiles have "darker" timbres and tend to soothe human emotions.

The summary of mood influenced by various features of a music wave is shown in Table 1 which gives the basic feature ranges for the moods under consideration.

Mood	Intensity	Timbre	Pitch	Rhythm
Happy	Medium	Medium	Very High	Very High
Exuberent	High	Medium	High	High
Energetic	Very High	Medium	Medium	High
Frantic	High	Very High	Low	Very High
Sad	Medium	Very Low	Very Low	Low
Deppression	Low	Low	Low	Low
Calm	Very Low	Very Low	Medium	Very Low
Contentment	Low	Low	High	Low

Table 1: Moods classified in accordance to audio features

This paper puts forth a method of classification of music in accordance to the mood it denotes, by extracting key features and comparing the values with the values of the most ideal song fitting in that mood. The comparison is done by training multi layered ANNs (artificial neural networks) and classifying the results based on if a certain threshold is exceeded. Classification is done only when a majority of the characteristics match that of any particular mood. This eliminates the ambiguity caused when one or few of the features don't agree with the results of the listening tests.

The feature extraction technique was implemented on more than 100 music pieces, selected from a western collection of songs as well as from Hindi films. This was done because moods are depicted in different fashions in the two genres. As expected, the threshold values of the features are found to vary. Depending on the type of input music composition, the proposed method is designed to accurately indicate the mood of the song. The following sections of the paper discuss in detail about the approach.

II. MOOD DETECTION USING FEATURE EXTRACTION AND MUSIC INFORMATION RETRIEVAL

In this section, the paper proposes a method to extract the mood of the music piece by extracting various features from it. The classification of moods is based on the Thayer's model of moods. First, the features are extracted for determining the mood. Since considering only one feature and deducing conclusions on the mood of the song is not reliable, and the values of one feature may wrongly represent the mood of the music piece, all the described features are considered. Next, the mood of the music piece is selected by neural networks that have been trained to select a particular mood depending on the threshold level determined. The threshold values of the features were determined to represent a particular mood by a series of listening tests and making use of music pieces from a website, stereomood.com, where the music pieces are enlisted according to ratings provided by listeners worldwide. Machine learning, in this case, neural networks, is then implemented to construct a computational model using listening test results and measurable audio features of the musical pieces. This approach is illustrated in the flowchart in Fig 2.

The details of the processes enlisted are given in the proceeding sections.

A. Feature extraction

Acoustic features such as intensity, timbre, pitch and rhythm are extracted from the music piece and measures derived from them. Each individual feature is described as follows:

a) Intensity features

Intensity features of a music piece, also known as dynamics clearly gives an indication of the degree of loudness or calmness of music and also is a pointer to measure stress in the composition of music. Two important entities that help to elicit the intensity features from a song are:

- *Amplitude*

Amplitude of the sound waveform in time domain can be quantized by the root mean square (RMS) of the signal. The energy of a music piece x is computed by taking the root average of the square of amplitude.[12] This is computed by first segmenting the audio clip followed by decomposition into frames of length 50ms and half-overlapping and further RMS is found. RMS energy of a calm and serene track of music is found to be lower than that of a high-energy music.

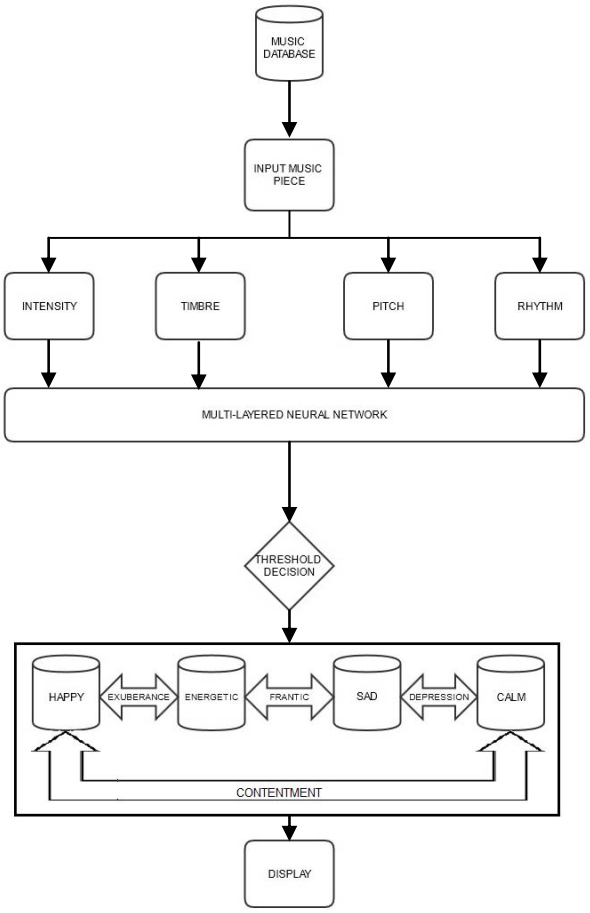

Fig 2: Flowchart of the method used in the paper to extract audio features and classify the musical pieces in accordance to moods.

- *Low energy*

As fast changes in amplitude relates to energy or exuberance and none such corresponds to low energy or peace, it is necessary to evaluate the consistency of energy or RMS values assessed through all the segments or frames and its matching with the value obtained for the whole piece of music.[13] Hence, to observe the contrast, if any, in different segments, the ratio of segments showing lower energy than the average RMS energy to the total number of frames is to be evaluated.[14]

b) Timbre features

The harmonic profile or sound quality of a sound source is timbre or tone color in music. Timbre is determined by the harmonic profile of the sound source. To quantize the value of timbre in the music pieces, the following factors were computed.

- *Zero Crossing:*

The zero crossing rate is the rate at which a signal crosses the zero-line. The value of zero crossing can be used in music information retrieval in order to classify percussive sounds.[8] It is a simple indicator of noisiness of the signal. This factor is also used as a primitive in pitch detection.

First, frame decomposition is performed with a default frame of 50ms and half overlapping before the zero crossing rate is determined.

- *Spectral Irregularity*

The irregularity of a spectrum is the degree of variation of the successive peaks of the spectrum. It is calculated based on the sum of the square of the difference in amplitude between adjoining partials.[9]

The audio signal is first decomposed into different channels of different frequencies, by using a number of filters with varying frequencies. This transformation models an actual process of human perception, corresponding to the distribution of frequencies into critical bands in the cochlea. The Discrete Fourier Transform of the decomposed signal is calculated to decompose the energy of the signal along frequencies. The peaks of the curve are found, before using the Jensen's formula to compute the irregularity.

- *Roughness*

Roughness, or the sensory dissonance, related to the beating phenomenon whenever pair of sinusoids are closed in frequency, can be estimated.[10] The estimation is based on the frequency ratio of each pair of sinusoids represented.

The estimation of the total roughness is computed by computing the peaks of the spectrum and taking the average of all the dissonance between all possible pairs of peaks.[11]

c) Pitch features

The pitch of a sound is dependent on the frequency of vibration and the size of the vibrating object. This feature corresponds to the relative lowness or highness that can be heard in a song. The two components that are considered for the purpose of mood taxonomy are the pitch itself and inharmonicity.

- *Pitch*

To get the pitch of a track, autocorrelation of the audio song is performed and FFT is computed for the pitch contents of the segments. Again, autocorrelation is performed for these values and the resulting representations, all in frequency domain are multiplied.

- *Inharmonicity*

Inharmonicity defines the number of components that are not multiples of fundamental frequency and estimates the energy outside the ideal threshold of

harmonic series. These are highly relevant to tuned percussion instruments.

d) Rhythm features

Rhythm, in music, refers to the placement of sounds in time. The sounds along with silences in between create a pattern, when these patterns are repeated they form Rhythm. Since rhythm is repetitive, it also relates to periodicity and frequency. Human mood is dictated by the rhythm and its periodicity and hence the following factors play a vital role in determining the mood of the musical piece [15].

- *Beat/Beat Spectrum*

Beat is one of the notes in a musical piece that sounds stronger than the other notes. A song usually contains a number of beats that are repeated periodically. These beats are of various amplitudes and frequencies. A spectral analysis of the beats provides an insight into the mood of a musical piece.

The beat spectrum is extracted by performing three operations on the musical piece. Firstly the audio is parameterized which results in a sequence of feature vectors. Later a distance measure is done to calculate the similarity between two matching feature vectors. Lastly this similarity is used to find the periodicities in the music which results in the beat spectrum [16].

- *Tempo*

Tempo is the speed or pace of the musical piece. The pace is determined by the frequency of the beats; hence tempo is measured in BPM (beats –per-minute). It indicates a measure of acoustic self-similarity as a function of time. A particular piece of music can have a range of tempi and usually two tempi are considered, the slowest tempo and the fastest tempo, to represent the rhythm of the music. In this particular method of music mood detection the fastest tempo is considered as a factor to decide the mood of the musical piece.

From the global shape of the audio signal, the successive bursts of energy are retrieved. Autocorrelation is performed on the data retrieved and the summation of the result yields the tempo of the audio clip.

- *Fluctuations*

Fluctuations in music indicate the rhythmic periodicities. One way of computing the rhythmic periodicity is based on the spectrogram computation transformed by auditory modeling [17].

The fluctuations are estimated by first computing a spectrogram and Terhardt outer ear modeling, with Bark-band redistribution of energy. Then FFT is performed on each Bark-band. The sum of the resulting spectrum leads to a global repartition of rhythmic periodicities i. e. fluctuations.

B. Machine learning using artificial neural networks (ANN)

Multilayered neural networks are explicitly programmed to evaluate the mood of a track and they are trained using Back Propagation algorithm. The network is first initialized by setting up all its weights to be small random numbers and the input is applied from the stereomood.com consisting of the four extracted features of the songs: pitch, rhythm, intensity and timbre that depict various moods that have been accepted by the people all over the world. Hence, there are 4 layers at the input and 8 layers to depict 8 mood scores at the output. Later, the features of the songs from both Indian and western genres are tested given as inputs to the neural network and are being classified into different moods based on the mood scores at the output. The highest mood score defines the mood of the song.

III. EXPERIMENTAL CONDITIONS AND RESULTS

In this section, the results of the proposed mechanism and the conditions of testing are mentioned in order to evaluate the performance of the mechanism.

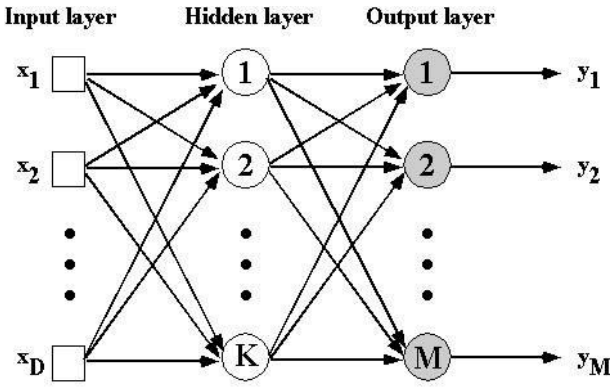

Fig 3: The general depiction of MLNN.

A. Experimental conditions

All the music composition signals used were sampled at 44100 Hz, and 16-bit quantized. Frame decomposition is performed with default frame length of 50ms and half overlapping. This frame length is used in calculations of timbre, intensity, pitch and rhythm features.

In this paper, music segments from 110 data are selected, with an average time length of 45 seconds. The mood scores of all the different music patterns are compared against moods such as happy, exuberance, energy, sad, depression, calm, anxiety, and contentment.

B. Experimental results

The experimental results are summarized in Table 3 and Table 4. From Table 2, it is seen that highest efficiency in

mood detection is obtained in energetic, calm and happy moods. The lowest efficiency is seen in frantic mood, which suggests that more factors need to be accurately considered in order to increase the efficiency in classifying under that category.

Table 2 shows the mean values of the important features observed in this paper. It is seen that most of these values are in high corroboration with the information conveyed in Table 1. Also, the method has been successful to an extent in classifying both western music and Hindi film music.

Mood	No of Songs	% Efficiency
Happy	13	92.31
Exuberent	14	85.71
Energetic	18	94.44
Frantic	6	66.67
Sad	9	77.78
Depression	10	90
Calm	17	94.11
Contentment	20	85

Table 2: Success rate of the proposed algorithm

Some of the beatspectra obtained for ideal musical pieces are shown below in Fig 4.

Each of the beatspectra represented in Fig 4, has a different shape and represents different moods of a musical piece. The spectrums are examples of musical pieces of moods Calm, Depression, Energetic, Exuberance, Frantic, Happy, Contentment and Sad respectively.

The spectra of calm and contentment are seen to be smooth, with little sudden variations in the amplitude, while those of energy, exuberance and frantic have small peaks rising up in either a predictable manner (energy and exuberance) or a non-uniform manner (in frantic). Similarly, the spectra of happy and sad songs are found to follow respective common patterns as is evident from Fig 4.

Mood of Western songs	Mean Intensity	Mean Timbre	Mean Pitch	Mean Rhythm
Happy	0.2055	0.4418	967.47	209.01
Exuberent	0.3170	0.4265	611.94	177.70
Energetic	0.4564	0.3190	381.65	163.14
Frantic	0.2827	0.6376	239.78	189.03
Sad	0.2245	0.1572	95.654	137.23
Depression	0.1177	0.2280	212.65	122.65
Calm	0.0658	0.1049	383.49	72.23
Contentment	0.1482	0.2114	756.65	101.73

Table 3: The mean values of extracted audio features in Western songs

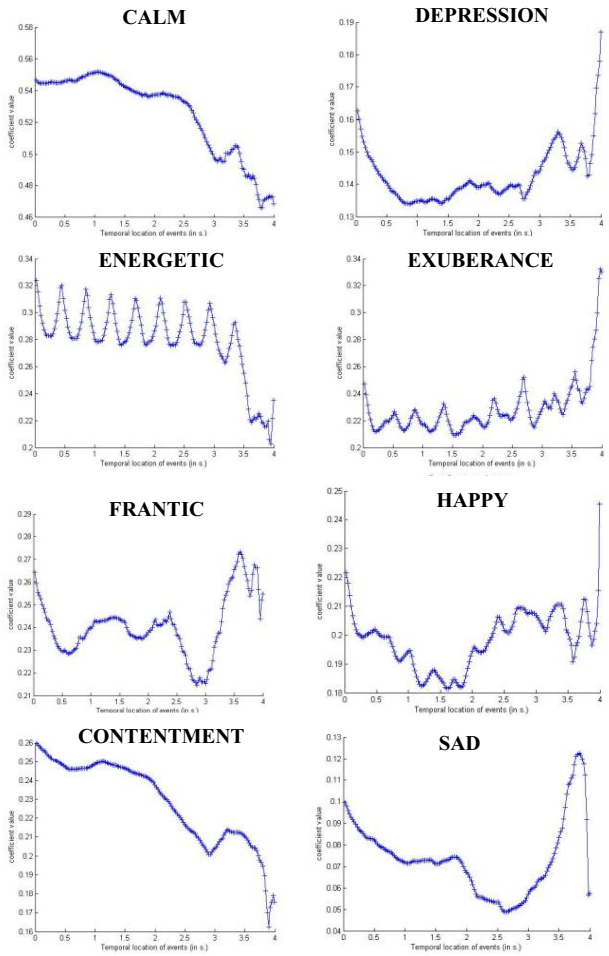

Fig 4: Beat-spectrum graphs of music pieces conforming to various moods

Mood of Hindi songs	Mean Intensity	Mean Timbre	Mean Pitch	Mean Rhythm
Happy	0.1806	0.5233	604.30	158.19
Exuberent	0.2643	0.3845	538.25	134.70
Energetic	0.3986	0.3799	380.79	150.55
Frantic	0.2888	0.6147	212.40	187.15
Sad	0.1640	0.1020	132.18	120.77
Depression	0.1919	0.2090	160.15	122.40
Calm	0.1160	0.1315	361.45	92.39
Contentment	0.1768	0.2689	411.25	125.53

Table 4: The mean values of extracted audio features in Hindi film songs

IV. CONCLUSIONS

Mood of the music is a basic aspect and finds its usefulness in music retrieval systems or mood taxonomy applications. A simple algorithm is proposed here for the mood detection by extracting four important features of music: pitch, timbre, intensity and rhythm. They are used to classify the moods Indian music as well as Western music on an accuracy level as high as 94.44% in the best case, as the results are not deduced by studying just one feature.

Rather, it is confirmed by other features as well and is well supported by the beatspectrum and roughness too.

High levels of accuracy are also attributed to the use of simple back propagation algorithm which has multi layered neural artificial network trained many times to produce the best matched results, consisting of 4 layers at the input and 8 layers depicting the mood scores of 8 different moods. From the experimental results, the optimal threshold values were different for every mood label.

As can be seen from Table 2, there is scope for further improvement. More efficient ways to integrate various features should still be explored due to the unbalanced nature of each feature set. More audio features could help improve the accuracy of the system. It is also observed that the data set used to build the classification model could be increased further to improve the accuracy of the classification system. The classification of Hindi music can be extended further to include music pieces from various forms of abundant traditional music. This would need more qualified listeners and a larger database. The same extension can be achieved with Western music, by including a larger database of songs. These mentioned points can be the focus of future work.

REFERENCES

[1] Carolyn J. Murrock,"Music and Mood," in Psychology of Moods 2005

[2] D. Huron, "Perceptual and cognitive applications in music information retrieval," in Proc. Int. Symp. Music Information Retrieval (ISMIR),2000

[3] Gabrielsson, A.; Lindstrom, E. (2001). "The influence of musical structure on emotional expression". Music and Emotion: Theory and Research: 223–243.

[4] Hunter, P. G.; Schellenburg, E. G., & Schimmack, U. (2010). "Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions". Psychology of Aesthetics, Creativity, and the Arts 4: 47–56.

[5] Hunter, P. G.; Schellenburg, E. G., & Schimmack, U. (2008). "Mixed affective responses to music with conflicting cues". Cognition and Emotion 22: 327–352.

[6] Ali, S. O.; Peynircioglu, Z. F. (2010). "Intensity of emotions conveyed and elicited by familiar and unfamiliar music". Music Perception: An Interdisciplinary Journal 27: 177–182.

[7] Webster, G. D.; Weir, C. G. (2005). "Emotional responses to music: Interactive effects of mode, texture, and tempo". Motivation and Emotion 29: 19–39.

[8] Gouyon F., Pachet F., Delerue O. (2000),Classifying percussive sounds: a matter of zero-crossing rate?, in Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7–9, 2000. Accessed 26th April 2011.

[9] Jensen, 1999.

[10] Plomp and Levelt, 1965

[11] Sethares, 1998.

[12] Department of Communicative Disorders University of Wisconsin–Madison. RMS Amplitude. Retrieved 2008-08-22.

[13] Gabrielsson, A.; Lindstrom, E. (2001). "The influence of musical structure on emotional expression". Music and Emotion: Theory and Research: 223–243.

[14] "Musical Genre Classification of Audio Signals" George Tzanetakis and Perry Cook IEEE Transactions on Speech and Audio Processing, 10(5), July 2002.

[15] P. N. Juslin, "Cue utilization in communication of emotion in music performance: relating performance to perception," J. Exper. Psychol.:Human Percept. Perf.,vol. 16, no. 6, pp. 17971813,2000.

[16] Jonathan Foote , Shingo Uchihashi , "The beat spectrum: a new approach to rhythm analysis".

[17] E. Pampalk, A. Rauber, D. Merkl, "Content-based Organization and Visualization of Music Archives", *ACM Multimedia* 2002, pp. 570-579.