

Automatic Mood Detection of Music Audio Signals: An Overview

Sonal P.Sumare¹Mr. D.G.Bhalke²

1.(PG Student Department of Electronics and Telecommunication Rajarshi Shahu College of Engineering Pune)

2.(Faculty Department of Electronics and Telecommunication Rajarshi Shahu College of Engineering Pune)

ABSTRACT: Music mood describes the inherent emotional expression of a music clip. It is helpful in music understanding, music retrieval, and some other music related applications. 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, in an effective way. People select music compatibility to their moods and emotions, making the need to classify music in accordance to moods. In this paper different mood models are described which uses to detect different moods using different methods as hierarchical method and nonhierarchical method with GMM, SVM.

Keywords: Hierarchical framework, mood detection, mood tracking, music emotion, music information retrieval, music mood, GMM, SVM.

I. INTRODUCTION

Most people enjoy music in their leisure time. At present there is more and more music on personal computers, in music libraries, and on the Internet. 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. For example, when an individual comes back home from work, he may want to listen to some relaxing light music; while when he is at gymnasium, he may want to choose some exciting music with a strong beat and fast tempo. Music is not merely a form of entertainment but also the easiest way of communication among people, a medium to share emotions and a place to keep emotions and memories. Booming of the Internet technology, there is more and more music on personal computer, in the music libraries and on the Internet. Therefore, automatic music analysis system such as music classification, music browsing and play list generation system are urgently required for music management facility. Because of various listening objectives in different time concordance, music classification and retrieval based on perceived emotion is mightily powerful than other tagging such as artist, album, tempo and genre.

Some music genres such as classics usually contain more than one musical mood. Distinct musical features create different musical mood. For better accuracy, we use various low level musical features and detect musical mood changes based on them. For this purpose, we first divide music clips into segments based on such musical features and cluster them into groups with similar features. Beat and tempo detection and genre classification have been developed in a few research works, using different features and different models. It is noted that, in most psychology textbooks, "emotion" usually means a short but strong experience while "mood" is longer but less strong experiences. Therefore, we mainly choose to use the word "mood" in this paper. However, the words "affect," "emotion" and "emotional expression" are still used in order to keep the same usage as those used in the references.

II. LITERATURE SURVEY

Over the years, considerable work has been done in music mood detection. Literature survey has been carried out of last 20 years. These are listed below:

A.S. Bhat, Amith V. S., Namrata S. Prasad, Murali Mohan D. [1] describes an Efficient Classification Algorithm for Music Mood Detection in Western and Hindi Music using Audio Feature Extraction. This paper proposed an automated and efficient method to perceive the mood of any given music piece, or the emotions related to it. Features like rhythm, harmony, spectral feature, are studied in order to classify the songs according to its mood, based on Thayer's model. All the music composition signals used were sampled at 44100 Hz, and 16-bit quantized. The accuracy of classifying mood is as high as 94.44%.

Lie Lu, Dan Liu, and Hong-Jiang Zhang [2] described Automatic Mood Detection and tracking of Music Audio Signals. In this paper, a hierarchical framework is presented to automate the task of mood detection from acoustic music data. Music features such as intensity, timbre, and rhythm, were extracted to represent the characteristics of a music clip. The approach to mood detection is extended to mood tracking for a music piece. Thayer's model of mood is adopted, which is composed of four music moods, Contentment, Depression, Exuberance, and Anxious/Frantic. The average accuracy of mood detection is up to 86.3%.

Mark D. Korhonen, David A. Clausi, and M. Ed Jernigan [3] proposed modeling emotional content of music using system identification. This paper developed a methodology to model the emotional content of music. System-identification techniques are used to create the emotional content models. Emotion Space Lab is used to quantify emotions using the dimensions valence and arousal. Because Emotion Space Lab collects emotional appraisal data at 1 Hz. Results show that the system identification provides the emotional content for a genre of music.

Yi-Hsuan Yang, Yu-Ching Lin, YaFan Su, and Homer H. Chen [4] describes a regression approach to music emotion recognition. This paper proposed for recognizing the emotion content of music signals. Music emotion recognition

MER is formulated as a regression problem to predict the arousal and valence values (AV values). To improve the performance, principal component analysis is used and it reduce the correlation between arousal and valence. The best performance for arousal is 58.3% and for valence is 28.1% by employing support vector machine as the regressor.

George Tzanetakis, Perry Cook [5] explains Musical Genre classification of Audio Signals. Musical genres are categorized by humans to characterize pieces of music. The categorized characteristics typically related to the instrumentation, rhythmic structure, and harmonic content of the music. Timbral texture, rhythmic content and pitch content these three features are proposed in this paper. Training statistical pattern recognition classifiers is used to evaluate proposed features. Using the feature sets, 61% classification of ten musical genres is achieved.

Jong In Lee, Dong Gyu Yeo, Byeong Man Kim, and HaeYeoun Lee [6] introduces Automatic Music Mood Detection through Musical Structure Analysis. The mood variation in music makes their application more difficult. To cope with these problems, the author present an automatic method to classify the music mood. A modified Thayer's 2-dimensional mood model with AV model is used to detect the mood.

EiEiPeMyint, Moe Pwint [7] proposed An Approach for Multi Label Music Mood Classification. This paper presents self-colored music mood segmentation and a hierarchical framework based on new mood taxonomy model. The proposed mood taxonomy model combines Thayer's 2 Dimension (2D) model and Schubert's Updated Hevner adjective Model (UHM). FSVM has superior accuracy as compare with the SVM.

III. THEORETICAL BACKGROUND

The mood of the people can be recognized by the music which they are listening. The characteristics of music such as rhythm, melody, harmony, pitch and timbre play a significant role in human physiological and psychological functions, altering their mood. With the help of these music characteristics the music mood is divided in different types of mood as: Happy, Exuberant, energetic, depression, frantic, sad, calm and contentment[2].

There are number of music feature. From that some of acoustic features such as intensity, timbre, pitch and rhythm are given below:

1.1 Intensity Features

Intensity is an essential feature in music mood detection. For example, the intensity of Contentment and Depression is usually little, while that of Exuberance and Anxious/Frantic is usually large. It gives an indication of the degree of loudness or calmness of music.

1.2 Timbre Features

In music timbre also known as tone color or tone quality, it is the quality of musical note or sound or tone that distinguishes different types of sound production. For example, the brightness of Exuberance music is usually higher than that of Depression.

1.3 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.

1.4 Rhythm features :

In music rhythm refers to the placement of sounds intime. The sounds along with silences in between create a pattern, when these patterns are repeated they form rhythm. In general, three aspects of rhythm are related with people's mood response: rhythm strength, rhythm regularity, and tempo.

IV. MOOD MODELS

Human psychologists have done a great deal of work and proposed a number of models on human emotions.

4.1 Hevner's experiment

In music psychology, the earliest and best known systematic attempt at creating music mood taxonomy was by Kate Hevner . Hevner examined the affective value of six musical features such as tempo, mode, rhythm, pitch, harmony and melody and studied how they relate to mood. Based on the study 67 adjectives were categorized into eight different emotional groups with similar emotions.

4.2 Russell's model

Both Ekman's and Hevner's models belong to "Categorical Model" because the mood spaces consist of a set of discrete mood categories. On the contrary, James Russell came up with a circumplex model of emotions arranging 28 adjectives in a circle on two dimensional bipolar space (arousal - valence). This model helped in separating and keeping away the opposite emotions.

4.3 Thayer's model

Another well known dimensional model was proposed by Thayer . It describes the mood with two factors: Stress dimension (happy/anxious) and Energy dimension (calm/energetic), and divides music mood into four clusters according to the four quadrants in the two-dimensional space: Contentment, Depression, Exuberance and Anxious (Frantic), as shown in fig1[1][2].

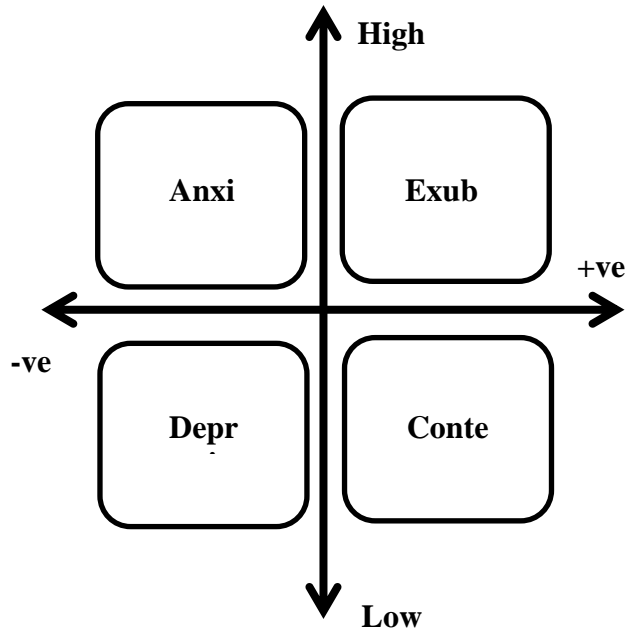


Fig.1: Thayer's Mood Model

V. MOOD DETECTION FRAMEWORK

5.1 Hierarchical Mood Detection Framework Using GMM

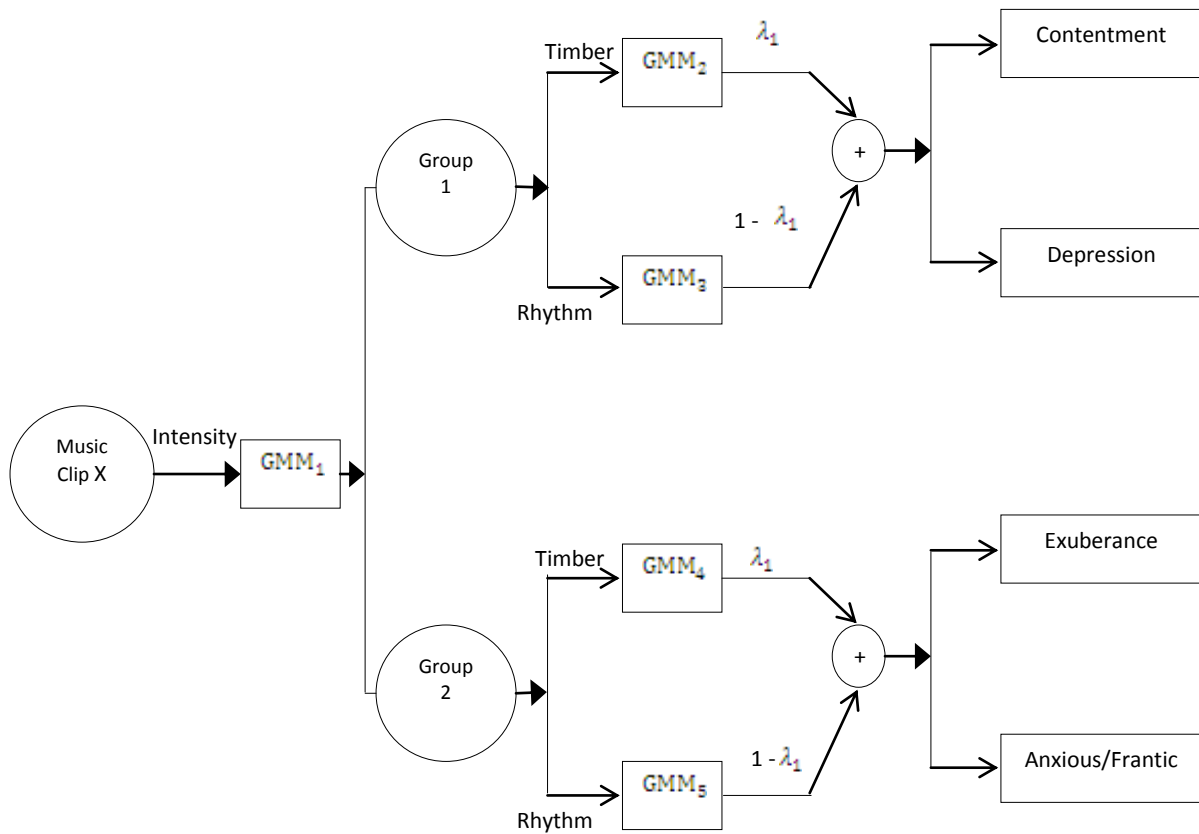


Fig. 2. Hierarchical mood detection framework[2]

Based on Thayer's model of mood, a hierarchical framework is proposed for mood detection, as illustrated in Fig.2. The intensity features are first used to classify a music clip into one of two mood groups.[2] The basic rule could be, if its energy is low, the music clip will be classified into Group 1(Contentment and Depression); otherwise, it is classified into Group 2 (Exuberance and Anxious/Frantic). Subsequently, the remaining features, including timbre and rhythm, are used to determine which exact mood the music clip is. With the obtained GMM models, the detailed mood classification can be performed in the following two steps. In the first step, a music clip is classified into different mood groups, i.e.Group 1 (Contentment and Depression) and Group 2 (Exuberance and Anxious/Frantic), by employing a simple hypothesis test with the intensity features, as

$$\lambda = \frac{P(G_1|I)}{P(G_2|I)} \begin{cases} > 1, \text{ Select } G_1 \\ < 1, \text{ Select } G_2 \end{cases} \quad (1)$$

Where λ is the likelihood ratio, G_i represents different mood group I is the intensity feature set. In the second step, the music clip in Group 1 is classified into Contentment and Depression, while that in Group 2 is classified into Exuberance and Anxious/Frantic, based on the timbre and rhythm features. In each group, the probability of the testing clip belonging to an exact mood can be calculated as,

$$P(M_{i,j} | G_i, T, R) = \lambda_i \times P(M_{i,j} | T) + (1 - \lambda_i) \times P(M_{i,j} | R), \quad i, j = 1, 2 \quad (2)$$

where $M_{i,j}$ is the j th mood cluster in i th mood group, T and R represent timbre and rhythm features, respectively, λ_1 and λ_2 are two weighting factors to represent different importance of timbre and rhythm features[2].

5.2 Nonhierarchical mood detection framework using GMM

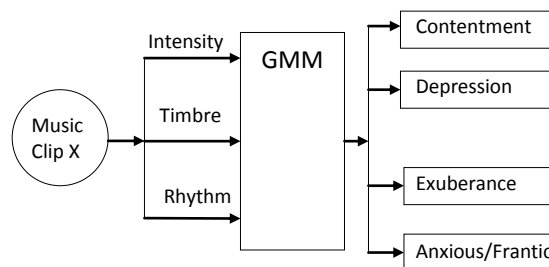


Fig. 3 Nonhierarchical mood detection framework[2]

Nonhierarchical framework is shown in fig.3. Comparing Nonhierarchical framework with its hierarchical part, the hierarchical framework can make better use of sparse training data, which is very important especially when the training data is limited.

In the framework, a Gaussian mixture model (GMM) with 16 mixtures is utilized to model each feature set regarding each mood cluster (group). In constructing each GMM, the Expectation Maximization (EM) algorithm is used to estimate the parameters of Gaussian components and mixture weights, and K -means are employed for initialization [2].

5.3 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 plane 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 nonlinear 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[10]. 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.

VI. CONCLUSION

This paper presents an approach to mood detection for acoustic recordings of music. A hierarchical framework is used to detect the mood in a music clip. In this intensity features, timbre and rhythm features are extracted. The hierarchical framework can utilize the most suitable features in different tasks and can perform better than its nonhierarchical framework. In SVM, a Mel frequency cepstral coefficient (MFCC) is extracted as a feature from the data collected. SVM Classifier performs better, which offers a new efficient way of solving problems.

ACKNOWLEDGMENTS

Any research or project is never an individual effort but contribution of many hands and brains. With great pleasure I express my gratitude to our Principal Prof. Dr. D.S. Bormane and Head Of Department Mr. D.G. Bhalke. I would like to place my thanks to all the faculty members of the Electronics and Telecommunication. At critical occasions their

affectionate and helping attitude helped me a lot in rectifying my mistakes and proved to be sources of unending inspiration, for which I am grateful to them. Their timely suggestions have helped me in completing this research work in time.

REFERENCES

- [1] Bhat, A.S.; Amith, V.S.; Prasad, N.S.; Mohan, D.M., "An Efficient Classification Algorithm for Music Mood Detection in Western and Hindi Music Using Audio Feature Extraction," *Signal and Image Processing (ICSIP), 2014 Fifth International Conference on* , vol., no., pp.359,364, 8-10 Jan. 2014
- [2] Lie Lu; Dan Liu; Hong-Jiang Zhang, "Automatic mood detection and tracking of music audio signals," *Audio, Speech, and Language Processing, IEEE Transactions on* , vol.14, no.1, pp.5,18, Jan. 2006
- [3] Korhonen, M.D.; Clausi, D.A; Jernigan, M.E., "Modeling emotional content of music using system identification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* , vol.36, no.3, pp.588,599, June 2005
- [4] Yi-Hsuan Yang; Yu-Ching Lin; Ya-Fan Su; Chen, H.H., "A Regression Approach to Music Emotion Recognition," *Audio, Speech, and Language Processing, IEEE Transactions on* , vol.16, no.2, pp.448,457, Feb. 2008
- [5] Tzanetakis, G.; Cook, P., "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE Transactions on* , vol.10, no.5, pp.293,302, Jul 2002
- [6] Lee, Jong In; Yeo, Dong-Gyu; Kim, Byeong Man; Hae-Yeoun Lee, "Automatic Music Mood Detection through Musical Structure Analysis," *Computer Science and its Applications, 2009. CSA '09. 2nd International Conference on* , vol., no., pp.1,6, 10-12 Dec. 2009
- [7] Myint, E.E.P.; Pwint, M., "An approach for multi-label music mood classification," *Signal Processing Systems (ICSPS), 2010 2nd International Conference on* , vol.1, no., pp.V1-290,V1-294, 5-7 July 2010
- [8] Miyoshi, M.; Tsuge, S.; Oyama, T.; Ito, M.; Fukumi, M., "Feature selection method for music mood score detection," *Modeling, Simulation and Applied Optimization (ICMSAO), 2011 4th International Conference on* , vol., no., pp.1,6, 19-21 April 2011
- [9] Bartoszewski, M.; Kwasnicka, H.; Markowska-Kaczmar, U.; Myszkowski, P.B., "Extraction of Emotional Content from Music Data," *Computer Information Systems and Industrial Management Applications, 2008. CISIM '08.7th* , vol., no., pp.293,299, 26-28 June 2008
- [10] E. Vijayavani1; P. Suganya; S.Lavanya; E.Elakiya , "Emotion Recognition Based on MFCC Features using SVM", *International Journal of Advance Research in Computer Science and Management Studies*, Volume 2, Issue 4, April 2014
- [11] A. McCallum *et al.*, "Improving text classification by shrinkage in a hierarchy of classes," in *Proc. Int. Conf. Machine Learning*, 1998, pp. 359–367.
- [12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer-Verlag, 2001.