# A Survey on MIR for Music And Mood Detection

**Harshali Nemade[1], Deipali Gore[2]**
Department of Computer Engineering
[1]PESs MCOE, Savitribai Phule Pune University, Pune, Maharashtra, India
[2] Assistant Professor, PESs MCOE, Savitribai Phule Pune University, Pune, Maharashtra, India

***Abstract-*** *Music Information Retrieval (MIR) is the task of retrieving information from the music and it is the fastest growing area of music industry. Music Information Retrieval basically deals with the problems of querying and retrieving certain types of information from audio files with the help of large data set. Digital music is widely available in different digital formats due to explosive growth of information and multimedia technologies. Thus, the management and retrieval of such music is necessary for accessing music according to their meanings in respective songs. A lot of research and study has been going on in the field of music genre classification and music mood detection in the recent years. MIR deals with different audio features like spectrogram, mfcc, timbre, pitch etc. Audio features are necessary in various domains of data mining like speech recognition, Music information retrieval, and environmental sound recognition and so on. It is nothing but mathematical functions calculated over the audio data, in order to describe some unique aspect of that data. In the last decades a huge number of features were developed for the analysis of audio content. Music is available everywhere and with the development of multimedia technology digital music has reached almost every individual's personal gadgets like laptop, music player, cell phones and so on. Today, the overall collection of songs is nearly few millions. With so much variety of songs it is necessary to query and retrieve certain type of song from a large dataset. Classification is the fundamental problem in MIR which includes assigning of labels to each songs based on genre, mood, artist etc. Music classification can help end users to select a particular song of their interest as well as on the other hand it can also help in managing different types of music more effectively and efficiently once they are categorized into different groups*

***Keywords***- Music Information retrieval,audio signal processing, classification algorithms, clustering algorithms, feature analysis , feature extraction,mood models

## I. INTRODUCTION

### *A. Introduction to music information retrieval*

Music Information Retrieval (MIR) is the science of retrieving information from the music. It deals with different audio features like spectrogram, mfcc, timbre, pitch etc. Audio features are necessary in various domains of data mining like speech recognition, Music information retrieval, and environmental sound recognition and so on. It is nothing but mathematical functions calculated over the audio data, in order to describe some unique aspect of that data. In the last decades a huge number of features were developed for the analysis of audio content. Music is available everywhere and with the development of multimedia technology digital music has reached almost every individual's personal gadgets like laptop, music player, cell phones and so on. Today, the overall collection of songs is nearly few millions. With so much variety of songs it is necessary to query and retrieve certain type of song from a large dataset. Music Information Retrieval (MIR) is an emerging research area in multimedia to cope with such necessity. It is the science of retrieving information from the music. Classification is the fundamental problem in MIR which includes assigning of labels to each songs based on genre, mood, artist etc. Music classification can help end users to select a particular song of their interest as well as on the other hand it can also help in managing different types of music more effectively and efficiently once they are categorized into different groups.

### *B. Introduction to audio features*

It is very essential to know what are the intrinsic factors present in music which associate it with a particular mood or emotion. A lot of research has been done and still going on in capturing various features from the audio file based on which we can analyze and classify a list of audio files. Audio feat-ures are nothing but mathematical functions calculated over the audio data, in order to describe some unique aspect of that data. In the last decades a huge number of features were developed for the analysis of audio content. Different taxonomies exist for the categorization of audio features. Weihs et al. have categorized the audio features into four subcategories, namely short-term features, long-term features, semantic features, and compositional features. Scaringella followed a more standard taxonomy by dividing audio features used for genre classification into three groups based on timbre, rhythm, and pitch information, respectively. Timbre features capture the tonal quality of sound that is related to different instrumentation, whereas temporal features capture the variation and evolution of timbre over time. Low-level features are the basis description of the audio data, for instance, tempo, beats per minute and so on. On the contrary, mid-level features are derived by using these basic features to provide the music related technical understanding such as rhythm,pitch which in turn is perceived by the humans as genre, mood, which form the top-level of the taxonomy.

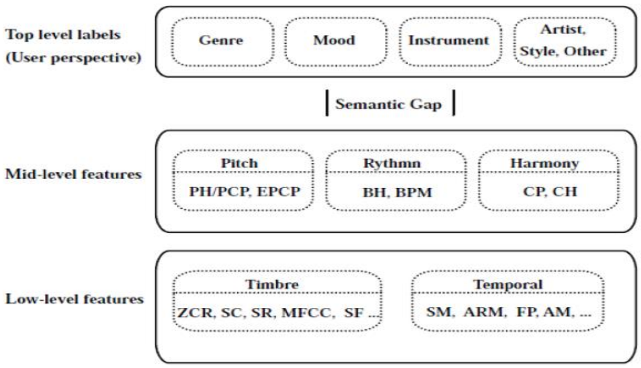



Fig. 1. Audio features categories

## II. OBJECTIVE

Objectives of the Music Information Retrieval are:

1) Identifying the various types of songs.
2) Learning different feature extraction techniques and their suitable representation for the feature vector.
3) Studying different classifiers which are available to perform music classification tasks more efficiently.

## III. KEY COMPONENTS OF MIR

The key components of a music classification system are
1 ) Feature extraction and representation
2) Music classification.
Following figure 2 shows the categorization of key components that are involved in MIR:

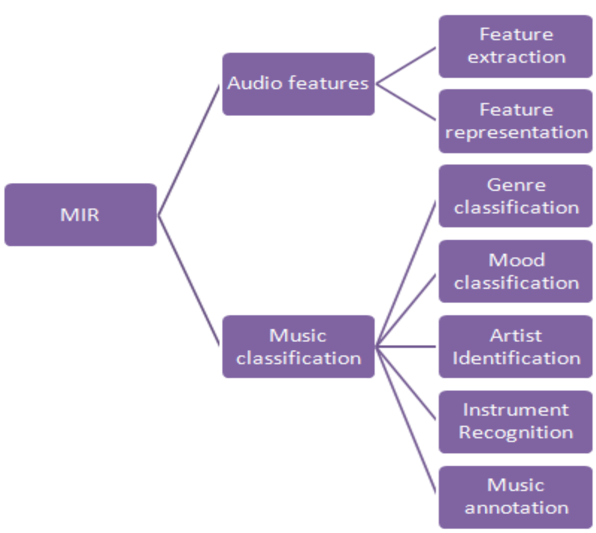



Fig.2 Key components of MIR

1. Audio features

### *A. Feature extraction*

Feature extraction addresses the problem of how to represent the music pieces to be classified in terms of feature vectors or pair-wise similarities [10].Different taxonomies exist for the categorization of audio features. Weighs et al.[11] has categorized the audio features into four subcategories, namely short-term features, long-term features, semantic features and compositional features.Scaringella [12] followed a more standard taxonomy by dividing audio features used for genre classification into three groups based on timbre,rhythm and pitch information. Each taxonomy attempts to capture audio features from certain perspective. As shown in figure 1, Low level features can be further divided into two classes Timbre and temporal features. Timbre features captures the tonal quality of a sound whereas temporal features capture the variation and evolution of timbre over time. Low level features are obtained directly from various signal processing techniques like Fourier transform, spectral analysis or ceptral analysis etc.Mid level features include rhythm, pitch and harmony. It relates to the strength, tempo and regularity of the music. From user perspective, top level assign labels to the audio file.

Following table shows summary of low level features:-

| Class | Feature type |
|---|---|
| Timbre | • Zero crossing rate<br>• Spectral roll off<br>• Spectral centroid<br>• Spectral flux<br>• Spectral bandwidth<br>• Spectral Crest factor<br>• Octave based spectral contrast<br>• Mel-frequency cepstrum coefficient<br>• Stereo Panning Spectrum Features<br>• Linear predictive Cepstrum Coefficient |
| Temporal | • Statistical momnets<br>• Amplitude modulation<br>• Auto regressive modelling |

Table 1: Low level features

### *B. Feature represenation*

Feature representation of a song can take any one out of the three possible forms.

1.Single vector for each song

It is a straight forward approach. It takes the average or median values for each feature attribute over all segments so as to construct a song level summary. The global codebook is constructed by clustering which is used as reference [1].Each local vector in the song is then mapped to the nearest codebook vector.

2.A similarity measure turned for each pair of songs.

In this type of representation pair-wise similarity is used.A probability model is used based on some divergence criterion such as kullback-leibler divergence or standard parametric models like single Gaussian with full covariance matrix or Gaussian mixture model and non-parametric models like k-means cluster model.

3.A feature set of local feature vectors. It keeps the feature set of local feature vectors for each song and use them directly for classification. Two classifiers that can be used in this way are: GMM classifier and Convolutional neural network.

2. Music classification

Music classification is a fundamental problem in MIR.Classifer learning is one of the main key component in classification process.The purpose of classifier learning is to find a mapping from the feature space to output labels so as to minimize the prediction error.Most of the high level tasks involved in the music classification are as explained below:

*A. Genre Classification*

Music can be divided into many categories based on styles, rhythm and even cultural background. The styles are what we call the "GENRES". It is the most widely studied area in Music Information Retrieval.Traditionally, the genres of the music are tagged by musical experts who may be musicians,professors or artists etc. With the wide development in the multimedia, it becomes necessary to automate a system for automatic genre classification.Following figure 3 shows the flow chart of a music genre classification.

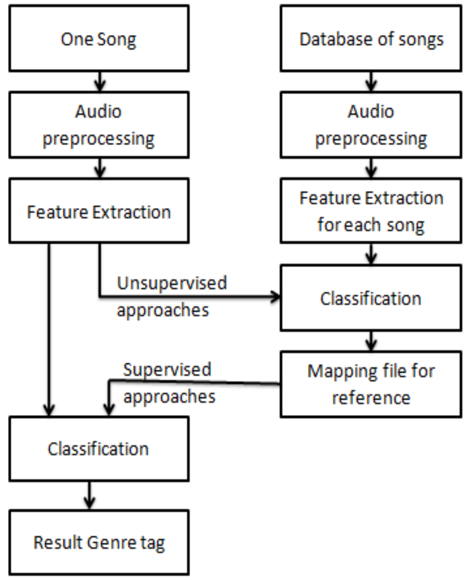


Figure 3. Genre classification

Data mining algorithms classify the data with unsupervised or supervised approaches as shown in following figure 4.

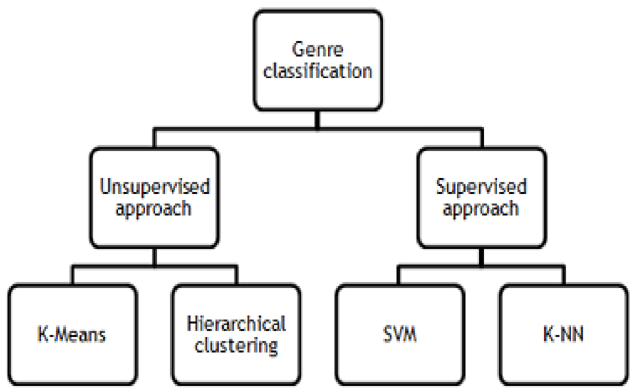


Figure 4. Genre classification techniques

i) Unsupervised Genre Classification

Unsupervised approaches classify data based on no knowledge about the genre clusters. Classifier can observe only the data position in feature space, but do not know what the genre of data is. When new data comes, the classifier implements an unsupervised approach again to get the clustering result. The approaches measure similarity among songs rather than genre of them. Two important unsupervised approaches are
introduced here.

- K-means [9]

When features span a high dimensional space, each song is seen as a feature set i.e a position in the feature space as figure 5.The operations of K-means are shown in figure 5.

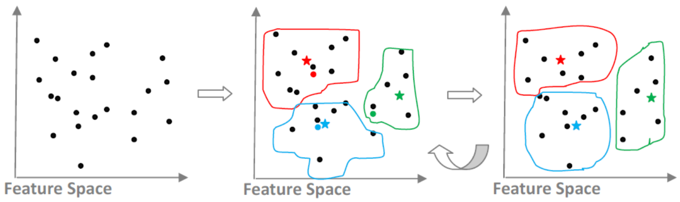


Figure 5.K-means example

K-means take k points(for eg,k=3) as the initialized centroid,and the k centroid are the red, blue and green dots in 5(b). Then, each point in feature space is classified into one of the 3 clusters by minimizing the root mean square distance between the centroid and the point. The result of the first iteration is as the red, blue and green boundary in 5(b), and this is the new clustering result. Next, we calculate each centroids of the new clusters,repeat the operations of 5(b) until the centoids of clusters do not change any more. The resulting clusters describe the similarity among songs, that is, the songs belong to the same cluster may sound more similar as shown in fig 5(c).The drawback of k-means is that k must be decided in advance, and how to choose a proper k is still an intractable issue.

- Agglomerative hierarchical classification [13,14]

Agglomerative hierarchical classification agglomerates clusters with a tree structure, as figure 6. Initially, we take each data point in feature space as a cluster Ci. Find out the cluster Cj with the minimal distance between Ci, and then agglomerate Ci and Cj into a new cluster. Repeat the operations describe above to build up an agglomerative tree as figure 6. Repeat the operations until the distance between clusters is smaller than the threshold or the clusters number is enough. The advantage of agglomerative hierarchical classification is that we can realize the similarity among songs in depth. For music recommendation, we can recommend the song within the same sub-cluster first.

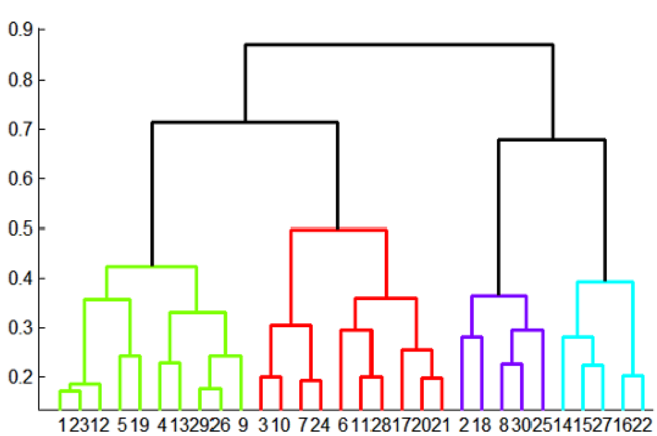


Figure 6. Agglomerative hierarchical classification example

ii) Supervised Genre classification

In Supervised approach, the system is trained by manually labeled data first, that is, supervised approach knows the genres of the songs. When unlabeled data comes, the trained system is used to classify it into a known genre. K-NN and SVM are the two most popular classifiers used for both general classification problems and in music classification.

- Support Vector machines[15]

SVM is the state-of-art binary classifier based on the larger margin principle.The goal of SVM is to find a classification hyper-plane that maximizes the margin among different genre data, and the basic idea is shown in figure 7.

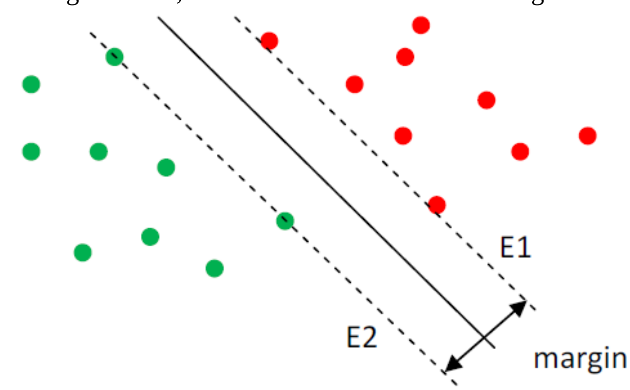


Figure 7. Support vector machine example

- KNN Algorithm

K-nearest neighbor is a supervised classifying algorithm where the result of new coming data is classified based on majority of K-nearest neighbor category. The operations of KNN are shown in figure 8.

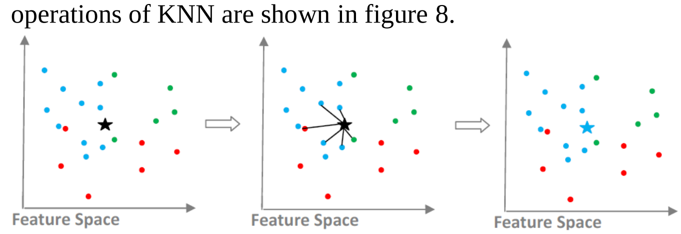


Figure 8. KNN example

For example, in figure 8, data points with known genres (red, green, and blue) are scattered in the high-dimension (2-D in fig. 8) feature space. A to-be-classified song comes, and the position of it is marked as a star in 8(a). Then, we decide k =6, that is, we take six nearest neighbors for classification. The distance between positions is commonly measured by equation which is the Minkowski metric.The five nearest neighbors are then linked to the star in figure 8(b). The six nearest neighbors then vote for the genre of the new coming song. As in 8(b), four neighbors are blue genre, one is red, and one is green, so the genre of the new coming song is classified as blue.

*B. Mood Classification*

- Introduction

Today, the overall collection of songs is nearly few millions.With so much variety of music easily available,we humans do not always listen to the same type of music all the time. We have our own interests, favorite artists, album or music type. To put it simply, We have our personal choices and more importantly even our choices might differ from time to time. Choosing a song or music piece suiting out mood from a large dataset is difficult and time consuming. Hence it is a high time to introduce a parameter mood. Following figure 4 shows the steps involved in mood classification system [10].

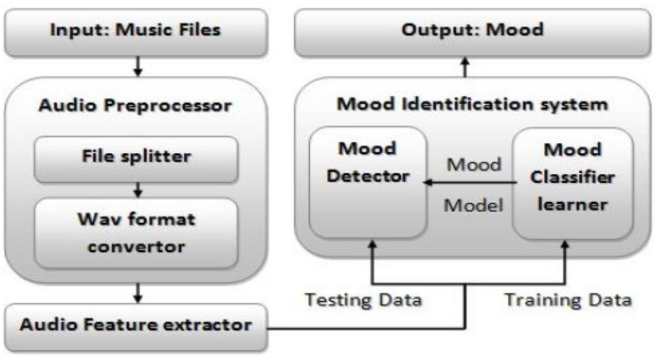



Figure 9. Mood detection flow system

- Classification methods for Mood categorization

Tagging a song with a mood label,Various mood models are available.Among those mood models Thayer's mood model is best and simple to use.It describes the mood with two factors as shown in fig 10: Stress dimension (happy/anxious) and Energy dimension (calm/energetic).

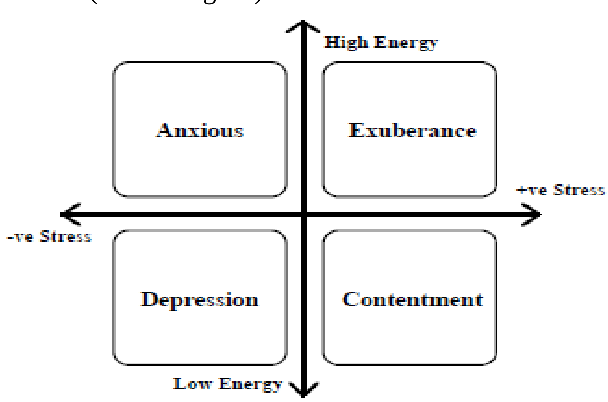



Figure10. Thayer's mood model

Thayer's mood model divides music mood into four clusters according to the four quadrants in the two- dimensional space:

1) Contentment
2) Depression
3) Exuberance
4) Anxious (Frantic).

*C. Artist Identification*

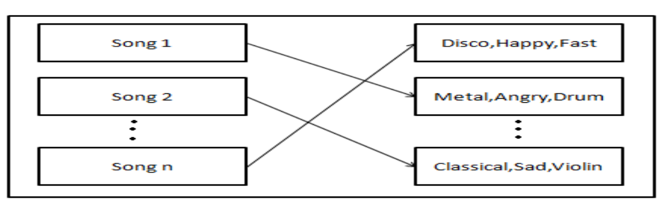
Artist identification includes several subtasks such as artist identification, Singer Recognition. Minnowmatch (Mima) automatically determines various meta-data and makes classifications concerning a piece of audio using neural networks and support vector machines[27]. The unique qualities of a singer's voice make it relatively easy for us to identify a song as belonging to that particular artist [28].In singer recognition system, Vocal part of the music is focused.GMM and SVM classifiers can be used efficiently in Singer recognition systems.

*D. Instrument Recogintion*

Genre,mood,artist etc classification is done at song level.The result is returned for a song while Instrument recognition is done at segment level because different instruments are played at different intervals in the audio clip.Instrument recognition problem can be categorized into 2 sub-problem as : Solo music recognition where a single instrument is played and Polyphonic music recognition where multiple instrument constitutes an audio file.

*E. Music Annotation*

Music annotation has recently gained much popularity in MIR related tasks.The main goal of music annotation is to annotate each piece of song with a set of semantically meaningful text annotations called tags.The music annotation problem can be viewed as a multi-label annotation problem as shown in following figure 4.




## AUTHOR PROFILE

**Harshali Nemade** received the B.E degree in Computer Engineering from AISSMS,Institute of Information Technology,Pune in 2011. She is now pursuing M.E. degree in

Computer Engineering from P.E.S.'s Modern College of Engineering, Savitribai Phule Pune University, Pune.

**Deipali Gore** received the M.E degree in Computer Engineering from D.Y. Patil, Akurdi College under Savitribai Phule Pune University in 2006. Her current research area includes Information Retrieval and Web Mining. She is currently working as an Assistant Professor in P.E.S.'s Modern College of Engineering, Savitribai Phule Pune University, Pune.

## REFERENCES


[1] Li Su,Chin-Chia Michael Yeh, Jen-Yu Liu, Ju-Chiang Wang, and Yi-Hsuan Yang,"A Systematic Evaluation of the Bag-of-Frames Representation for Music Information Retrieval",IEEE ,Aug 2014.

[2] G. Tzanetakis, and P. Cook,"Musical genre classification of audio signals ",Speech and Audio Processing, IEEE Transactions on, vol. 10, no. 5, pp.293-302, 2002.

[3] M. Henaff, K. Jarrett, K. Kavukcuoglu, and Y. LeCun,"Unsupervised learning of sparse features for scalable audio classification",ISMIR, 2011.

[4] P. Hamel and D. Eck,"Learning features from music audio with deep belief networks",ISMIR, 2010.

[5] M. Riley, E. Heinen, and J. Ghosh"A text retrieval approach to content-based audio retrieval",ISMIR, 2008.

[6] Velankar Makarand and Dr. Sahasrabuddhe H.V.,"Novel Approach for Music search using music contents and human perception",IEEE,2014.

[7] Aniruddha M Ujlambkar,"Automatic Mood Classification Model For Indian Popular Music",IEEE, 2012.

[8] Vallabha Hampiholi,"A method for classification based on Perceived mood detection for Indian Bollywood Music",World Academy of Science, Engineering and Technology, Vol:72 2012-12-25.

[9] Vallabha Hampiholi,"A method for classification based on Perceived mood detection for Indian Bollywood Music",World Academy of Science, Engineering and Technology, Vol:72 2012-12-25.

[10] Aniruddha M Ujlambkar,"Automatic Mood Classification Model For Indian Popular Music",IEEE, 2012.

[11] ]Weihs C., Ligges U., Morchen F., Mullensiefen D. ,"Classiffication in music research",Advance Data Analysis Classiffication, vol. 1, no. 3, pp,2007.

[12] Scaringella N., Zoia G., Mlynek D.,"Automatic genre classiffication of music content, A survey",IEEE Signal Processing Magazine, vol. 23, no.2, pp. 133141,2006.

[13] S. L. Aristomenis, and A. T. George,"Agglomerative Hierarchical Clustering For Musical Database Visualization and Browsing", 2004.

[14] S. Xi, X. Changsheng, and M. S. Kankanhalli, "Unsupervised classification of music genre using hidden Markov model."pp. 2023-2026 Vol.3.

[15] C. Xu, N. C. Maddage, X. Shao et al., "Musical genre classification using support vector machines." pp. 429-432.

[16] E. Pampalk, S. Dixon, and G.Widmer, "On the evaluation of perceptual similarity measures for music", in Proc. Int. Conf. Music Information Retrieval, 2003.

[17] T. Li and M. Ogihara, " Toward intelligent music information retrieval," IEEE Trans. Multimedia, vol. 8, no. 3, pp. 564–574, 2006.

[18] M. Mandel and D. Ellis, "Song-level features and SVMs for music classification ",in Proc. Int. Conf. Music Information Retrieval, 2005.

[19] G. Tzanetakis, "Marsyas-0.2: A case study in implementing music information retrieval system," , in Proc. Intelligent Mobile Information Systems.

[20] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kegl, "Aggregate features and ada boost for music classification" Mach. Learn., vol. 65, no. 2–3, pp. 473–484, 2006.

[21] I. Panagakis, E. Benetos, and C. Kotropoulos, "Music genre classification: A multilinear approach" in Proc. Int. Conf. Music Information Retrieval, 2008.

[22] C.-H. Lin, J.-L. Shih, K.-M. Yu, and H.-S. Lin, "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features" IEEE Trans. Multimedia, vol. 11, no. 4, pp. 670–682,2009.

[23] I. Panagakis, C. Kotropoulos, and G. R. Arce, "Music genre classification using locality preserving non-negative tensor factorization and sparse representations" in Proc. Int. Conf. Music Information Retrieval, 2009.

[24] Z. Fu, G. Lu, K. Ting, and D. Zhang, "On feature combination for music classification" in Proc. Int. Workshop Statistical Pattern Recognition, 2010.

[25] http://www.musicgenreslist.com/, Musicgenrelist

[26] L. Lu, D. Liu, and H.-J. Zhang, " Automatic mood detection and tracking of music audio signals " IEEE Trans. Speech Audio Process.,vol. 14, no. 1, pp. 5–18, 2006.

[27] W.-H. Tsai and H.-M.Wang, "Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals" IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 1, pp. 330–341,2006.

[28] Y. E. Kim and B. Whitman, "Singer identification in popular music recordings using voice coding features,"in Proc. Int. Conf. Music Information Retrieval, 2002.