Combining the Spectral Features To identify the musical Instruments and Recognize the emotion from music

[1]Janani.S, [1]Iyswarya.K, [1]Krishna Priya.K.P
Computer Science and Engineering
Velammal Institute of Technology
Chennai, Tamil Nadu
jananimpm@gmail.com, iyswaryak@gmail.com, krishnapriyakpkp@gmail.com

[2]Maria Michael Visuwasam.L., IEEE Member
Assistant Professor, Department of Computer Science and Engineering
Velammal Institute of Technology
Chennai, Tamil Nadu
Micael_vm@yahoo.co.in

***Abstract*— Music can influence Human pervasive that can console, motivate, feel the love and hate or even bring us tears. Instrument plays a vital role in Musical Composition. 'Combining the Spectral Features To Identify the Musical Instruments and Recognize the Emotion from a Music' aims at providing the most easy and efficient method to identify the emotion of the song which can be used for Music Therapy. Our proposed work includes identifying the Musical Instruments using Dynamic Time Warping (DTW) Technique [3] which is a time alignment method. Since Spectrogram features are combined with MFCC the Musical Instruments can be effectively identified. Emotion Recognition estimates the mood of the Musical song which becomes an important aspect of Music Information Retrieval. This Musical Information can be determined by extracting the Features of Dynamics, Timbre, Harmony, Rhythm and Articulation. Using these features the Emotional Values are estimated by a Three-Dimensional Emotional Space which involves Valence, Activity and Tension which is analogous to negatively excited, positively excited and calm neutral space. The effect of combining the Spectral Features degrades the performance of the system, which can be resolved by applying Dimensionality Reduction Process. This provides very stable and successful emotional classification.**

Keywords- DTW, MFCC, Emotion, Timbre, Valence.

I. INTRODUCTION

With digital music becoming more and more popular (such as music CDs and MP3 music downloadable from the internet), music databases, both professional and personal, are growing rapidly. Technologies are demanded for efficient categorization and retrieval of these music collections, so that consumer can be provided with powerful functions for browsing and searching musical content. With this capability provided in a music system, the user can easily get to know about the particular music instrument songs, or retrieve all songs contained such instrument music in a distributed music database. Among such technologies, is the music instrument identification of a song, i.e. to recognize the music instrument of a song by analyzing music features of the music signal. One of the most appealing functions of music is that it can convey emotion and modulate a listener's mood. It is generally believed that music cannot be composed, performed, or listened to without affection involvement. Music can bring us to tears, console us when we are grieving, and drive us to love. Music information behavior studies have also identified emotion as an important criterion used by people in music searching and organization.

II. LITERATURE REVIEW

The major challenges for this study arise from the fact that a music signal tends to be arbitrarily altered from time to time and is inextricably intertwined with the signal of the background accompaniment.

This paper[1] surveys the various aspects of automatic emotion recognition in music. Music is natural for us to categorize music in terms of its emotional associations. Myriad features, such as harmony, timbre, interpretation, and lyrics affect emotion, and the mood of a piece may also change over its duration where the music emotion recognition technique particularly focuses on the methods that use contextual text information (e.g., websites, tags, and lyrics) and content-based approaches.

The Recognition of Instrument for a polyphonic background [5] which extracts the most prominent fundamental frequency and its harmonic overtone series which provides high recognition of instrument accuracies even when the music is accompanied by a keyboard instrument or a complete orchestra. The drawback in this paper is that Fundamental Frequency is to be computed for each note which increases Time Complexity.

The voice recognition [4] is performed by converting the speech waveform into MFCC features and then these features are subjected to feature matching by DTW algorithm using the reference. The drawback is that the converted digitized signals are then converted back into the waveform which does not produce accurate results.

Music Instrument Recognition by using the Dynamic Time Warping (DTW) [3] technique which can be done efficiently by classifying the musical instruments using the MFCC features which involves warping between two time series by stretching or shrinking along its axis to determine the similarity between the two time series music signals. In this paper Music Instrument is recognized for Isolated notes only.

Mood Classification in Music [6] is performed by developing a framework for studying the wrapper selection in music classification based on cross-indexing algorithm. This recognition of expressed emotions in music leads to an efficient and interpretable models which are simplified and generalized. This method cannot be applied for complex tasks and does not produce accurate results.

The emotion is recognized [2] by determining the coordinates of the song which is represented as a point in the Cartesian space with valence and arousal as the dimensions. Ranking-Based objective function is applied by ranking the collection of music pieces by emotion and then determines the emotional values of each music piece by using RBF-ListNet Algorithm. Ranking the music piece should be taken care such that it perfectly determines the emotions. This is done in 2-Dimensional space to map the emotion..

III. PROPOSED WORK

In this paper, the emotion of the song and its instruments used are identified from the input music signal. The input is given in the form of the wave file and the output the identification of the music instrument and its characteristics from the wave file. The songs are analyzed for the identification of the music instrument; once the instrument is identified the emotion of the song is also identified using the instrument and its related features. The basic block diagram given in fig.1.1 shows the overview of this paper. The processes involved in this paper are Preprocessing, Feature extraction, Dimensionality Reduction, Instrument Identification and Emotion Recognition of the input music signal.

In the Preprocessing stage the input signal is noise removed, sampled and segmented. The preprocessed segments are given to the Feature Extraction process, where the segments are converted into frames by applying an overlap. Finally, MFCC, Temporal, Energy, Rhythm, Spectral and pitch determination features are determined. MFCC and Spectral features are used to recognize the instrument in the audio signal whereas Energy, Spectral, Rhythm and Temporal features are used to recognize the Emotion of the music signal.

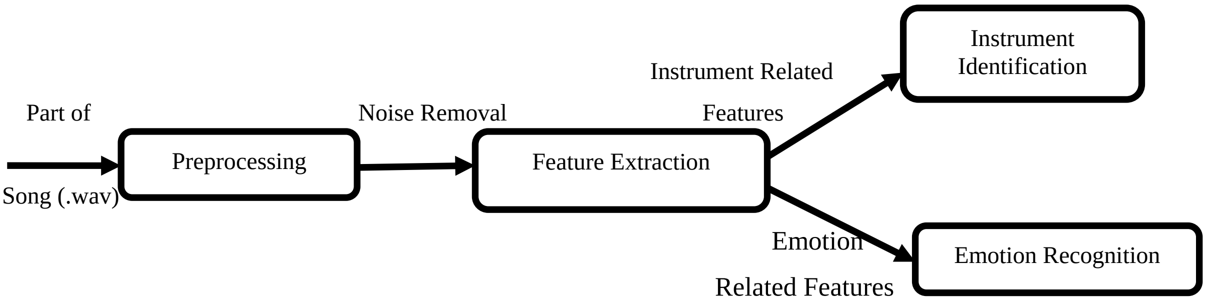

Figure 1. System Overview

A. *Pre-processing*

The input music song will we segmented to different number of frames and the external noise will be removed from the given signal. It is the process of removing the noise and also the silence period available in the music signal. The silence part in a music signal can be removed by detaching the zero crossings in the signal. Noise removal involves the removal of energy signals that acts as noise to the music. This can be done by using calculating RMS value for each frame of the signal.

INPUT: Music signal
OUTPUT: Noise and silence period removed analog music signal.
Algorithm of the module:

i. Start

ii. Fix the threshold value.
iii. Determine the harmonic structure stability analysis.
iv. Segment the input music signal.
v. Design the wiener Filter for noise removal
vi. Stop

B. Feature Extraction

In this module, Cepstrum analysis is done. The MFCC co-efficient are determined in this module. These are considered as the feature vectors of this module. Hamming window is used to minimize the spectral distortion of each frame.

INPUT: Preprocessed noise removed signal

OUTPUT: Features Extracted Signal

Algorithm of the module:

i. Start
ii. Convert the segments into frames
iii. Determine the FFT of the signal
iv. Apply windowing to smooth the edges of each frames
v. Determine DCT to calculate the Feature vector of each frames
vi. Stop

In this process, the features are extracted from the noise and silence removed music signal. We extract the features in order to determine the instrument and also the respective emotion. Hence it is necessary to extract both the emotion and also instrument related features.

C. Dimensionality Reduction

INPUT: Extracted features from the music signal

OUTPUT: Dimensionality reduced feature set

DATABASE:

It has two sets in it. They are,

- Training set
- Evaluation set

Algorithm of the module:

i. Start
ii. The extracted feature values are stored in the database.
iii. The features are clustered based on the emotion.
iv. Probability is identified for each feature set.
v. Depending on the priority of each feature the unrelated features are abandoned.

In this process, the feature extracted music signal is clustered together according to the features based on emotion. Once it is clustered, it is necessary to find the probability for the music signal in order to find the related emotion of the given music signal. Dimensionality reduction is mainly used to reduce the number of random variables and to increase the performance of the process. Hence, Principle Component Analysis (PCA) algorithm is used in-order to perform Dimensionality Reduction process. The output will be dimensionality reduced feature set. It is sent to the database for future use.

D. Instrument Identification

With the clustering values the database contains different set of files for training violin, tabla and flute. The new song feature vectors are found using MFCC and this is used for clustering. Again this value will be compared with the reference file. The distance between the two files will be computed. The minimum distance between these will give the threshold value. The threshold value is found in the training phase. In corresponds to this value the instrument will be found.

INPUT: Dimensionality Reduced feature set

OUTPUT: Instruments like violin, flute and tabla are identified.

Algorithm of the module:

i. Compute Euclidean distance from target plot to those that were sampled.
ii. Order samples for calculated distances.
iii. Choose optimal k nearest neighbor by cross validation technique.
iv. Calculate an inverse distance weighted average.

v. Using a weighted k-NN also significantly improves the results.
vi. Apply DTW technique by using distance matrix
vii. The calculated distance is analyzed and the instrument is identified.

Dimensionality reduced feature set is given as an input to the classifier. The classifier uses the k-nearest neighbor algorithm and also gets the information from the training set of the database to classify the available features. Once it is identified, it is sent to the evaluation set in the database to perform Instrument Recognition. In the instrument recognition process, the evaluation set input signal is sent. Using the Dynamic Time Warping algorithm, the instrument used in the music signal is identified. If there are more than one instrument in the signal, the instrument with dominant frequency will be considered. Instrument is recognized using Dynamic Time Warping Algorithm. In this technique, 39 MFCC features are used. For accuracy Spectral Features are combined with the MFCC features. The features are extracted and a reference template is created. By the algorithm, the feature extracted is compared with the reference template and stretching and shrinking is applied and the respective instrument is recognized.

E. Emotion recognition

INPUT: Instrument recognized music signal
OUTPUT: Emotion recognized music signal
Algorithm of the module:

i. Start
ii. The reduced features by dimensionality reduction are extracted and analyzed from the database using the classifier.
iii. The classifier creates a hyper-plane of which each music piece is segregated to the respective emotion through the SVM classifier.
iv. Based on this values as priority the emotion for the particular song is plotted into the 3-D space of which the emotion of the song can be recognized
v. Stop.

PROCESS:
The classifier uses the Support Vector Machine Algorithm to classify the music signal according to the respective emotion using the training set in the database. The computed values using the algorithm are stored in the evaluation set in the database. Using the emotion Recognizer, the available emotion in the music signal is identified using the Ranking based technique with the help of valence, activity and tension values. Finally the emotion is recognized from the available datasets and also the instrument recognized signal values.

IV. RESULTS AND COMPARISON

We have conducted experiments on three musical instruments which are Tabla, Flute and Violin using MFCC, spectrum features and DTW algorithms. Algorithms are implemented in MATLAB to get the identified instrument. Also results for only training dataset is more promising than testing dataset. DTW score for all templates are determined and depending on the score the instrument is identified.

In this paper, it is proposed that the musical instrument can be better recognized combining the spectrum features with MFCC features. Every distortion measure should be based on DTW for better recognition accuracy. It is also shown that 13 MFCC coefficients (plus delta and double delta) and Spectral features represent the acoustic model of musical instruments. The emotion is recognized from the spectral features of the instruments used in the song. The emotion related features generated can be used for finding the emotion of the song by using the dimensions valence, activity and tension which provides precise values for emotion of the song.

The correctness of the module is valuated by comparing the obtained output with that of the actual output.

TABLE I. TEST CASE

S.NO	TEST CONDITION	INPUT	EXPECTED RESULTS
1	Input Music Signal	Proper format has to be followed. Wave file should be given.	Supports only Wave files other formats are not supported.
2	Pre-processing	Song (.wav files)	Segmented and noise removed signal.
3	Threshold Value	Pre-defined value for noise removal and Peak Finding	Separated segments based on the threshold value
4	Feature Extraction	Music Segments.	Related features are extracted for the music segments
5	Dimensionality Reduction	Music Frames	Reduce the feature set by using clustering technique
6	Instrument Identification	Feature Values as coefficients	Distance matrix is applied and related instrument is identified
7	Emotion Recognition	Frequency Response of features	Emotional Values of music signal id determined

Performance evaluation is a necessary and beneficial process, which provides annual feedback to staff members about job effectiveness and career guidance. The performance review is intended to be a fair and balanced assessment of an employee's performance.

TABLE II. FEATURE RELEVANCY

S.NO	EMOTION	NO. OF SONGS SELECTED	FEATURES	DESCRIPTION
1.	Happy	50	1.Peak Features	High
			2.Spectrum Features	Provides the fundamental frequency
			3.Chromogram features	High
			4.Key Features	Provides the key note values.
			5.Key strength	Gives the pitch of the song and it must be high.
2.	Sad	50	1.Spectrum Features	Provides the fundamental frequency.
			2.Peak Features	Low
			3.Key Strength	Very Low
			4.Harmonic change detection features	Provides the harmonic change and determine its standard deviation.
3.	Tender	50	1.Spectral Centroid Features	Provides the centroid values which determine the tender.
			2.Roughness Features	The roughness is low for tender
			3.Key strength	Medium
			4.Harmonic change detection features	Harmonic changes are known and standard deviation is calculated.
4.	Anger	50	1.Roughness Features	Roughness is high
			2.Key strength	It is negatively high
			3.Spectral entropy features	Entropy values for determining anger.
			4.Spectral novelty features	The novelty values are low.
5.	Fear	50	1.RMS features	RMS values will be low
			2.Attack time features	The Attack time will also be low
			3.Peak feature	The peak features will be very low.
			4.Key feature and Key strength	Key strength will be low.

The emotion of each song is tested and the performance is calculated for the songs. We consider around 50 songs for each emotion and the project is tested for each input file. The songs which illustrate the approved emotion is considered as classified and the songs which does not match is regarded as misclassified.

TABLE III. OVERALL PERFORMANCE

Related Emotion	No. of input songs	No. of songs classified	No. of songs misclassified	Emotional Performance
Happy	50	42	8	84%
Sad	55	52	3	94.5 %
Tender	53	46	7	86.8%
Anger	53	49	4	92.5%
Fear	56	51	5	91.1%
Overall Performance				89.8%

The overall performance is evaluated for the classified and misclassified songs. This system provides outcome of 90% which provides stable results.

V. CONCLUSION

In this paper, it is proposed that the musical instrument can be better recognized combining MFCC and Spectral features. DTW works better for time varying signals. Every distortion measure should be based on DTW for better recognition accuracy which is represented as the acoustic model of musical instruments.
Emotion recognition technique is estimated based on the 3 dimensions Valence, Activity and Tension. This provides the emotion of the song which can be used for listening to songs based on the mood of a person. We have used Dimensionality Reduction process to reduce the large set of feature set. This improves the performance of the system by reducing the time and space complexity of the process.

VI. FURTHER RESEARCH

1. Compatibility with all kinds of input audio files.
2. A larger database for storing the feature set for emotions.
3. Enhanced design with automatic music information retrieval process and music documents for the respective constraint

REFERENCES

[1] S. Janani, K. Iyswarya and L. Maria Michael Visuwasam, “A Critical Survey on Music Emotion Recognition Techniques for Music Information Retrieval”, CiiT International Journal of Programmable Device Circuits and Systems, PDCS102011006, October 2011.

[2] Yi-Hsuan Yang and Homer H. Chen, Fellow, “Ranking-Based Emotion Recognition for Music Organization and Retrieval”, IEEE Transactions on Audio, Speech, and Language Processing, Volume 19,No.4, May 2011.

[3] D. G. Bhalke, C. B. Rama Rao and D. S. Bormane,” Dynamic Time Warping Technique for Musical Instrument Recognition for Isolated Notes”IEEE International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), May 2011.

[4] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, “Voice Recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques “Journal of computing , Volume 2, issue 3, March 2010, ISSN 2151-9617

[5] Jana Eggink and Guy J. Brown, “INSTRUMENT RECOGNITION IN ACCOMPANIED SONATAS AND CONCERTOS” , ICASSP 2004, International Conference in IEEE, Volume 4, May 2004, ISSN 1520-6149.

[6] Pasi Saari, Tuomas Eerola, and Olivier Lartillot, Member,” Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 6, AUGUST 2011

[7] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang, “A Survey of Audio-Based Music Classification and Annotation”, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 2, APRIL 2011

[8] Mahmoud I. Abdalla and Hanaa S. Ali “Wavelet-Based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models “Journal of telecommunications , volume 1, Issue 2, March 2010

[9] Y.-C. Lin, Y.-H. Yang, and H.-H. Chen, "Exploiting genre for music emotion classification," in Proc. IEEE Int. Conf. Multimedia Expo.,2009, pp. 618–621.

[10] Y.-H. Yang, Y.-C. Lin, and H. H. Chen, "Personalized music emotion recognition," in Proc. ACMInt. Conf. Inf. Retrieval, 2009, pp. 748–749.

[11] X. Hu, J. S. Downie, C. Laurier,M. Bay, and A. F. Ehmann, "The 2007 MIREX audio mood classification task: Lessons learned," in Proc. Int. Conf. Music Inf. Retrieval, 2008, pp. 462–467.

[12] Y.-H. Yang, Y.-C. Lin, H.-T. Cheng, I.-B. Liao, Y.-C. Ho, and H.-H.Chen, "Toward multi-modal music emotion classification," in Proc. Pacific-Rim Conf. Multimedia, 2008, pp. 70–79.

[13] L. Lu, D. Liu, and H. Zhang, "Automatic mood detection and tracking of music audio signals," IEEE Trans.

[14] Audio, Speech, Lang. Process., vol. 14, no. 1, pp. 5–18, Jan. 2006.

[15] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman," Speaker Identification using Mel frequency cepstral coefficients " ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh

[16] Mark D. Skowronski and John G. Harris," Improving the filter bank of a classic speech feature extraction algorithm "IEEE Intl Symposium on Circuits and Systems, Bangkok, Thailand, vol IV, pp 281-284, May 25 - 28, 2003, ISBN: 0-7803-7761-31

[17] Waibel , B. Yegnanarayana , " Comparative Study of Nonlinear Time Warping Techniques in Isolated Word Speech Recognition Systems "IEEE transaction on Acoustics , speech and signal processing Vol. . ASSP-31, NO. 6, December 1983

AUTHORS PROFILE

L. Maria Michael Visuwasam born on 5th September, 1981 near Kovilpatti. He received the B.Tech degree in Information Technology from Anna University, Chennai with Distinction in 2005. He has received M.E degree in Computer Science and Engineering specialization with Knowledge Engineering from College of Engineering, Anna University, Chennai, India in 2008. He received MBA degree in Education Management from Alagappa University, Karaikudi, India in 2010 and registered Ph. D degree in Anna University, Chennai in 2010. He is doing research in the area of Music Emotion Recognition. He has been with the Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai as Assistant Professor. He has published two international journals, six international conferences, ten national conference papers and got best paper award from NCRTIS-2K11. His research includes emotion recognition, Ad-hoc Networks and web security.

S. Janani born on 27th October, 1990 at Chennai. She is currently pursuing her B.E computer science and engineering in Velammal Institute of Technology, Chennai. She has successfully completed Diploma in UNIX and C (DUC) and in CSC Computer Education during 2008-2009 with excellent grade. She had presented a paper in the National Conference on Emerging Research in Engineering & Technology for 'A Survey on Music Emotion Recognition Techniques for MIR' and also published one International Journal and two International Conferences in various leading institutes. She had also participated in an Academic Developer Conference'11, Microsoft DreamSpark Yatra – Chennai.

K. Iyswarya born on 16 th October, 1990. She had completed a database course on SQLServer and a database administration course on ORACLE in 2008. She had been a co-ordinator and student vision leader of Women Empowerment Cell in Velammal Institute of Technology. She had secured Ist class in the Mathematics Talent Test conducted by the District Maths Club, Villupuram.

K.P. Krishna Priya born on 1st October, 1990. She is doing her B.E. Computer Science and Engineering in Velammal Institute of Technology, Chennai. She had presented a paper on the topic 'Combining Phoneme And Spectral Features For Emotion Recognition' at the conference ICCCS at Saveetha Engineering College.