Emotion Recognition System for Therapeutic cause through identification of Indian Music

Paromita Das, Assistant Professor, Amity University Kolkata INDIA, rchparomita@gmail.com

Somsubhra Gupta, Associate Professor, Amity University Kolkata INDIA, gsomsubhra@gmail.com

Biswarup Neogi, Associate Professor, JIS College of Engineering Kalyani INDIA, biswarupneogi@gmail.com

Abstract Guided by the Artificial Intelligence inspired System to make use of music for therapeutic cause of individual to get rid of stress and tension in extension to relaxation, recreation, the present work is an attempt to develop a production system to extend further the pattern identification of regional music in this part of the globe from the erstwhile published work on measurement of recorded effect of sound wave on human brain inquest of improved attentiveness and concentration [1]. Music combines melody, linguistic knowledge, and singer's emotion. Music may contain many emotions according to the vocalist's mood such as glad, sorrow, anxious, tired, peace which in turn may be treated in therapeutic approaches. Music classification and retrieval by perceived emotion is a powerful approach. As a therapeutic part, analysis of emotion involved in the music is an essential task though the same by computer is a difficult mission. Emotion, one of the most complex part to study proposed to adapt the evaluations of audio features from the music files. The extracted features help to categorized different emotion attributes of the vocalists. Musical features extraction using (Music Information Retrieval) MIR toolbox is one the most efficient approach. In this paper, the audio features related to the emotions of the vocalists are extracted to use in emotion recognition system based on music.

Keywords — ***Emotion, Music detection, Perception, Recognition, Signal Processing***

I. Introduction

All of us are fond of music, music tunes our heart control our mind state by making emotional adjustment. It starts when mother sings song to console her child [2]. Since song involve the voice modulation of the vocalist it easily controls our sentiments. As the advancement of technology continues, it becomes important to develop computers that can appraise the emotion of multimedia content to provide information access [3,4]. Enabling computers to recognize the emotion of music helps to enhance the way human and computer interacts. For example, an electronic device such as MP3 player or mobile phone equipped with MER function can play a song most suitable for the emotion of the user [5,6] some common spaces like restaurant, conference room, residence can play background music best suited the people inside it. Automatic prediction of emotion in music is referred to as music emotion recognition (MER). Usually, MIR toolbox used to extract different features of music sample under consideration. Two types of emotional models are considered i.e categorical and dimensional. Categorical approach describes several stages of emotion and dimensional approach offers several axes to map emotion into a plane. Both are adopted in classification of music signals and their emotion states can be classified by using the extracted features [2]. The extracted features involved in Indian classical music are shown as a sample are depicted in Table 1 at the Section III subsection E.

A. Present State of Art

Most prior works on MER [7,8,9,10,11] categorize emotions into numerous classes and apply the standard pattern recognition technique to train a classification model. Typically, the emotion classes are defined in terms of arousal (how exciting or calming) and valence (how positive or negative).

Many of the approaches involved multiple linear regression (MLR), support vector regression (SVR), and AdaBoost. RT (Boost.R) [3]. In [13], the multi-label classifiers, which allow allocating more than one emotion class to the similar song, are accepted in response to the fact that human perception of music emotion is not same. In [3], where the continuous perspective is elaborated, MER is formulated as a regression problem and SVR is Fig. 1. Thayer's arousal-valence emotion plane. employed to predict the AV values. With this regression approach, the problems intrinsic to categorical approaches are evaded. In addition, because there is more scope to describe a song, the subjectivity

 DOI : 10.35291/2454-9150.2021.0355

ISSN : 2454-9150 Vol-07, Issue-03, JUNE 2021

problem can be avoided. For instance, apart from the quadrant to which the song belongs, one can further recognize the emotion intensity the song expresses by examining its AV values. Nadia Lachetar [14] considered song title and lyrics as features and then Naive Bayes and Ant Colony algorithm as classifiers to classify emotion. Bram van de Laar [15] described emotion detection in music survey which compared many detection methods. Adit Jamdar, Jessica Abraham, Karishma Khanna [16] presented emotion analysis of songs based on lyrical and audio features like tempo, mode, loudness, danceability and energy. Natural language processing involved in lyrics analysis. The classification is done by K-Nearest Neighbors algorithm.

A four quadrants representation of emotions classes is depicted in fig. 1 known as Thayer's arousal valence emotion plane [12] for clear reference.

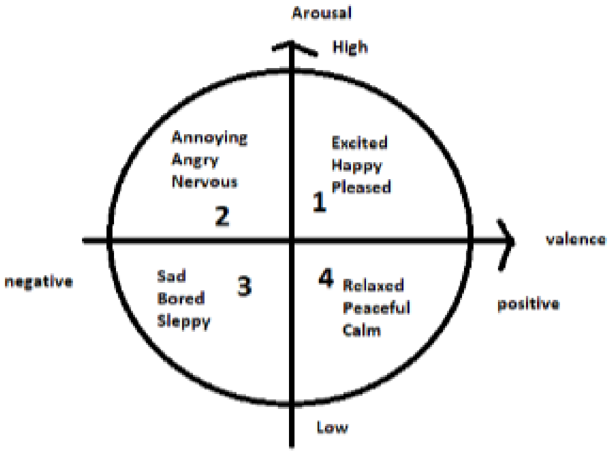

Fig. 1 *Thayer's arousal-valence emotion plane*

The methodological aspect is now presented in the next Section.

II. METHODOLOGICAL ASPECTS

The proposed system consists of classifier training and Testing procedure. The music emotion classification (MEC)[16] implementing Machine Learning is depicted in the block diagram shown in Fig.2

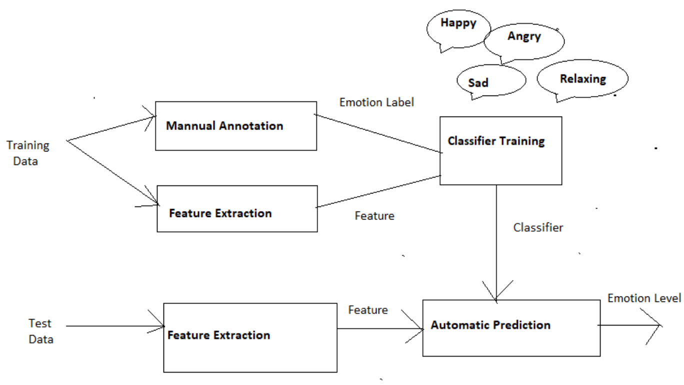

Fig. 2 *Block diagram of MEC.*

This project takes song as input, disintegrate the vocal from instrumental and background noise, extract the features of vocal using DSP system toolbox and identify the notes each with a duration [18].

III. EMOTION RECOGNITION SYSTEM DESCRIPTION

The proposed system is represented in subsections with a goal to provide both the data analytics and process. These are gradually presented below.

A. Data Collection

To study the role of individual, each song should be annotated by enough participants. Therefore, we collect 50 famous popular Indian songs, and make each song annotated by 50 participants. The collected music samples are trimmed to 25 seconds by manually trimming. The emotions of these songs distribute roughly uniformly in each quadrant of the emotion plane.

B. Methodology

The methodological aspects better represented in the following diagram in Fig. 3 in which the work clusters of detection and identification has been presented. Following the normal process of detection averaging, thresholding, width selection, finding particular or trivial instant has been taken into consideration. The triviality is given due information is decision making by following the thumb rule that "exceptions are not examples and examples are not rules". On the other hand, Identification incorporates stepwise the padding, followed by Discrete Fourier Transform the details of which are discussed in the next.

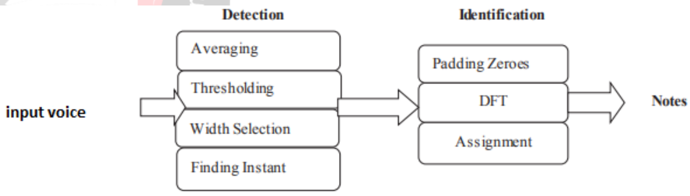

Fig. 3 *Flow Chart*

The experiments are done on different songs. Here each note follows a kind of similar pattern as shown in the fig 3. The moment we press one note to the immediate other note, the amplitude is initially high enough and decreases with time. If we can detect the duration of each note from the time domain characteristics, we can discover and identify the frequency.

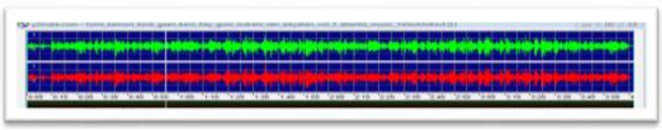

Fig. 4 *Vocal sample*

These experiments are mainly done based on the framework defined early [19] under the framework of Machine Intelligence as follows:

The entire technique is clustered into steps as described in the next Subsection.

 DOI : 10.35291/2454-9150.2021.0355

C. Proposed Technique

Step 1: The machine will either interact directly with the patient (in the case of a non-psychological patient) or counsel via a single member patient party to get the list of symptoms of the disease(s) that the patient is suffering from.
Step 2: From the existing Database (1) of symptom to disease it will analyze (through reasoning) what can be the possible disease(s). In this point the technique will use a checkpoint to examine how relevant is the patient's feedback through backward reasoning.

If some symptoms are beyond its knowledge base, then those will be added to its knowledge base for further up-gradation of its standard. It will be practically implemented by keeping some weights to each symptom to determine precedence as number of diseases can have same symptoms with different intensity.

Step 3: Once the disease(s) is determined, the second Database (2) of disease to music domain will come into effect to conclude about the kind of treatment the patient should have. The music domain will contain several music samples.

Step 4: The interaction between machine and patient will occur in two stages of which the first is already discussed at step 1. This step will host the second phase of interaction through which the patients taste and likeness of the music will be analyzed. This final step will be concluded through heuristic determination of sample music medicine that already exists in the music domain.

Meaning of the symbols symbol used to represent Production system pictorially -

- ❖ Knowledge base Modification
- ❖ Any set of symptoms, diseases, or music domains
- ❖ Interaction with Patient
- ❖ Conclusion
- ❖ Correspondence

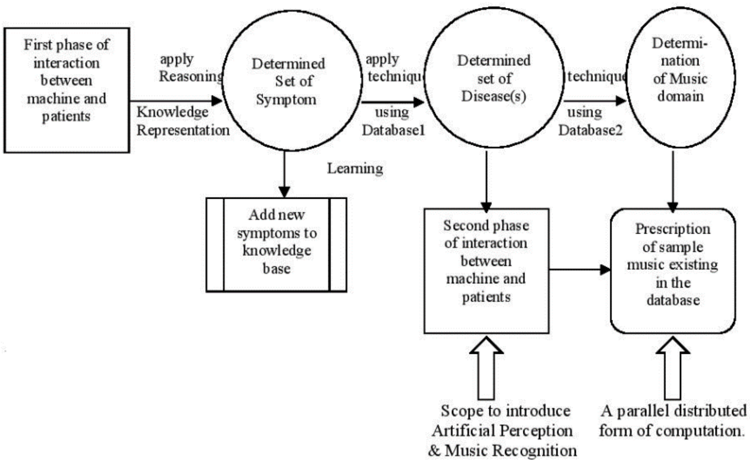

Fig. 5 *Complete Intelligent System Control flow*

The symptom to disease mapping in the above figure 5 is clearly built upon domain specific knowledge in medical science in the category of any i.e., considered to be application domain of this concept. Surely surgical cases are beyond the scope. Similarly, to build initial disease to music domain database (2) we should have explicit domain specific knowledge about music therapy. Both databases will be enriched with more and more trial using learning concepts.

We can consider branching the entire world of music into number of categories. However, any starting database will be too tiny as compared to the total number of music variations that the real world possesses.

Here is the illustration of a disease to music domain database in the trivial sub-field of classical Indian Ragas-

Disease	Raga
Indigestion	Ahir Bhairab
lack of Confidence	Asavari
Insomnia	Bagesri
Gall-stone	Basant-bahar
Rheumatic Arthritis	Bhairavi
Anxiety	Bhim Palasri
Anorexia	Chandrakauns
High blood pressure	Tori/Bhupali
Low blood pressure	Malkauns
Heart ailments	Chandrakauns
Easing tension	Hanswadhani /Kalavati

and many more.

To create some specific state of vibration for example –
Humid, cool, soothing, and deep mood – Raga Kafi
Sweet, deep, heavy, and cloudy mood – Raga Megha / Poorya Dhanasri.
Pleasing, refreshing, light and sweet touch – Raga Mishra Mand.

In practical, several such sub-fields of music domain are to be considered to enable effective treatment irrespective of nationality, race and religion. Ultimately it may be required to incorporate Data Mining technique to find applicable sub-field of music domain.

D. Detection

Averaging

As there is large number of samples for a song and many fluctuations are also present, first step is averaging, where average of every 100 samples is assigned to first sample and the average of next 100 samples to 2nd sample. This reduces the number of samples and fluctuations. When the decay in the signal is slow, the averaged signal is denser. But in case of fast decaying the averaged signal represents the envelope of the original signal.

Thresholding

Constant thresholding: Next stage of averaging is detection of the peaks from the averaged signal. As a part of

 DOI : 10.35291/2454-9150.2021.0355

constant thresholding one optimum value is decided for which maximum number of peaks are obtained.

Adaptive thresholding: Adaptive thresholding is applicable when we take some constant threshold value, for some notes it may be higher than the maximum value of the note and for some notes it is low enough such that two peaks of the note get merged and only one peak (one note) could be detected.

Even in adaptive thresholding one problem is, when there is long silence between two notes, the threshold value for the silence will be very low that supposed to be discarded.

Width Selection for Finding instant

Once the thresholding is done, our aim is to detect the occurrences of the peaks from the signal we get after thresholding. There are the possibilities when we take some constant threshold value, for some notes it may be higher than the maximum value of the note and for some notes it is low enough such that two peaks of the note get merged and only one peak (one note) could be detected. To overcome this problem the concept of adaptive thresholding came into picture.

The width of the window contains all zeroes is selected on the basis of worst case scenario. For our song it is found as 57, so the width was chosen as 50 for safety. The worst case depends on the speed for a given sampling frequency. For slow songs it is more and for faster one it is less. So irrespective of the song, the minimum width was decided as per following equation:

$$W_{min} = \frac{fs}{(A*20)}$$

In which fs = sampling frequency, A = Total number of samples in the original signal that corresponds to one sample in the averaged signal.

E. Identification

So far we have decided the instants at which the notes are played (key is pressed). The time duration from the first instant to the second instant is the duration of the first note being played, from second to third instant, it's second note being played and so on. Now our aim is to identify these notes based.

Padding Zeroes

For a given note duration, if crop the corresponding instants of the original signal and find the discrete Fourier transform, the results are not close enough as required. So we need to pad the zeroes. We can pad the zeroes with different length and different part of the cropped signal and find the DFT that will give the different results. Different part means it can be either only before the cropped version, after it or both the sides with different length. From these different variations the closest results are obtained in the case where the section of zeroes with the same length is padded both the sides of the cropped version. So DFT of the resultant signal is found to determine the notation.

Finding Frequency and Assigning Notation

After padding zeroes, the DFT of the resultant signal is found. Then the corresponding frequency of a particular note is found using equation (1) and the actual note is assigned using table 1.

TABLE I. AUDIO FEATURE UNDER CONSIDERATION

Song Type							
	Time	Peak frequency (Hz)	Bin frequency (Hz)	Peak pitch	Bin pitch	Power in dB	phase
Thumri 1	5:10 - 5:11	1536-1553	1539-1550	G6-35c---G6-17C	G6-32C---G6-20C	-50--29	-3.09--3.005
Thumri 2	5:10 - 5:11	993.147	979-998	B5+9C	B5-14C—B5-19C	-35--24	-2.38—2.74
Gazal 1	5:10 - 5:11	13867-13870	13824-13910	A9-26C	A9-32C—A9-21C	-54--43	.853—3.133
Gazal 2	5:10 - 5:11	1722-1728	1701-1722	A6-38C—A6-32C	A6-38C—A6-32C	-36--27	-2.64-2.97

IV. CONCLUSION

This paper presents a methodology proposed an approach to identify Indian music for therapeutic cause with an extension to identify vocalists. This is an extension works to Measurement of effect of music on human brain and consequent impact on attentiveness and concentration during reading. In this paper mainly deals with song and singer both on music perspective. Therapeutic cause mainly incorporates relief from stress and strain. Any further extension however needs further investigation. The tested sample that has been mentioned early is used to draw systematic inference. However, further investigation is needed before making specific conclusion under the framework of machine intelligence. The initiative was made much earlier in 2007 in [19] and is still under process. The erstwhile work was mainly based on Indian classical ragas. However, there was no specific DSP tools, simulations and application. Beginning was only based on formation of production Systems. Now, technological analysis using DSP is in place however, to have a concrete and complete conclusion, further investigation is needed. However, our initiative so far based on the specific roadmap are inline and expected to produce some more interesting result to validate the tile of the research proposal.

REFERENCES

[1] P. Das, S. Gupta, B. Neogi, "Measurement of effect

 DOI : 10.35291/2454-9150.2021.0355

of music on human brain and consequent impact on attentiveness and concentration during reading", Procedia Computer Science, Elsevier Science, ISSN:1877-0509, Vol. 172, pp. 1033-1038, 2020. (https://doi.org/10.1016/j.procs.2020.05.151)

[2] Kee Moe Han, Theingi Zin, Hla Myo Tun "Extraction Of Audio Features For Emotion Recognition System Based On Music", international journal of scientific & technology research volume 5, issue 06, june 2016

[3] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen, "Music Emotion Classification: A Regression Approach", in Multimedia and Expo, IEEE International Conference, 2007, pp.208-211.

[4] *Yi-Hsuan Yang*, *Ya-Fan Su*, *Yu-Ching Lin*, and *Homer H. Chen*," *Music Emotion Recognition*: The *Role* of *Individuality*." HCM'07, September 28, 2007, Augsburg, Bavaria, Germany.

[5] S. Reddy and J. Mascia, "Lifetrak: music in tune with your life," Proc. Human-centered multimedia, pp. 25–34, 2006.

[6] S. Dornbush, K. Fisher, K. Mckay, A. Prikhodko, and Z. Segall, "XPOD – A human activity and emotion aware mobile music player," Proc. Int. Conf. Mobile Technology, Applications and Systems, 2005.

[7] E. Schubert, "Measurement and time series analysis of emotion in music," Ph.D. dissertation, School of Music & Music Education, Univ. New South Wales, Sydney, Australia, 1999.

[8] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," Proc. ACM SIGIR, pp. 375–376, 2003.

[9] D. Yang and W. Lee, "Disambiguating music emotion using software agents," Proc. Int. Conf. Music Information Retrieval, pp. 52–58, 2004.

[10] T. Li and M. Ogihara, "Content-based music similarity search and emotion detection," Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing, pp. 17–21, 2006.

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

[12] R. E. Thayer, The Biopsychology of Mood and Arousal, New York, Oxford University Press, 1989.

[13] P. Synak and A, Wieczorkowska, "Some issues on detecting emotions in music," Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Springer, pp. 314–322, 2005.

[14] Nadia Lachetar ,Halima Bahi, "Song classification", Computer Science Department.

[15] Bram van de Laar , " Emotion detection in music, a survey" ,in Twente Student Conference on IT, 2006,vol 1,p.700

[16] Adit Jamdar, Jessica Abraham, Karishma Khanna and Rahul Dubey ,"emotion analysis of songs based on lyrical and audio features", International Journal of Artificial Intelligence & Applications (IJAIA) ,Vol. 6, No. 3, May 2015.

[17] C. Ho, W.-T. Tsai, K.-S. Lin, and H. H. Chen, "Extraction and alignment and evaluation of motion beats for street dance," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2013.

[18] Jay K. Patela , E.S.Gopia, Musical Notes Identification using Digital Signal Processing, Proceeding of 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)

[19] S. Gupta, "Music Perception, Cognition and Effect: A case study of Artificial Intelligence illustrating Disease Diagnosis", Proceedings of the 5th International Conference on Information Science, Technology and Management (CISTM) 2007, ISBN: 978-0-9772107-8-7, 50, pp. 1-16, 2007

 DOI : 10.35291/2454-9150.2021.0355