Emotion Based Music System
Nanda. M. B\textsuperscript{1}, Rithik. V Gopal\textsuperscript{2}, Soumyadeep Ghosh\textsuperscript{3}
Assistant Professor\textsuperscript{1}, UG Student\textsuperscript{2,3}
Department of Computer Science & Engineering
Sathagiri College of Engineering, Bangalore, India

Abstract: The most recognizable part of a human body is the human face itself. The human face also plays an extremely significant role in extraction of an individual’s current emotional state. Music can definitely be a mood changer, but manually segregating the list of songs and creating playlists according to the individual’s emotional state can be very exhaustive, time consuming, boring and labour intensive work. To automate the playlist generation as well as saving humans from a lot of boring hardship a number of algorithms have been proposed as well as developed. However these algorithms have their own drawbacks, they are computationally very slow, less accurate and some of them even require additional hardware like EEG, sensors etc. which can obviously add up to the overall cost. Through this paper we propose a system based on facial expression extraction which would automate the process of playlist generation thereby reducing the effort as well as time required to segregate the list of songs. And moreover the proposed system tends to reduce the computational time in obtaining the results, improving the overall accuracy as well as reducing the designing cost too.

Keywords: Audio Emotion Recognition, Music Information Retrieval, Emotion Extraction Module, Audio Feature Extraction Module, Artificial Neural Networks

I. INTRODUCTION
According to the psychologists the facial expression of an individual reveals information about the individual’s current emotional state. And they have categorized the facial expressions into 5 different types namely anger, joy, sad, surprise and excitement. Some of the psychologists have also found that the cognitive interpretations of emotions from facial expressions were innate and universal to all humans regardless of cultures. The below depicts the 5 types of facial expressions:

Music plays a very important role in improving the quality of an individual’s life as it is an important medium of entertainment and sometimes has its vital share in the therapeutic approach. In this 21\textsuperscript{st} century there has been a lot of development in the field of multimedia and technology. A number of music players have been developed with features like fast forward, reverse, variable playback speed, local playback, streaming playback with multicast streams. However these features satisfy the basic requirements but still there is a long way to achieve automatic segregation of songs and selection of playlist on the basis of the individual’s current mood or emotional status. Thus the introduction of Audio Emotion Recognition (AER) and Music Information Retrieval (MIR) techniques into traditional music players was done to provide automatic segregation of the playlist based on various classes of emotions and moods. AER classifies the received audio signals, by considering various audio features into diverse classes of emotions and moods, whereas MIR extracts some intensive information from an audio signal by examining some audio features like pitch, tempo, meter, timber, rhythm, energy, frequency etc. Though both AER and MIR have
lessened the work of humans by avoiding manual segregation of songs and generation of playlist, but still they are unable to incorporate fully a human emotion controlled music player. The music players available now are capable of classifying the music on the basis of artist, album, genre and number of times played, recently played etc. but not on the basis of the emotion conveyed by the music.

II. LITERATURE REVIEW

Music has been classified using lyrical analysis [1], [2] while this method is relatively quite easier to be implemented by its own but a major drawback is that it is not suitable to classify songs accurately. This method is not efficient due to the language barrier which restricts classification of songs to single language. Another method for music mood classification is using acoustic features like tempo, pitch and rhythm to identify the sentiment conveyed by the song. This method involves extracting a set of features and using those feature vectors to find patterns characteristic to a specific mood [3], [4]. As the technology advanced it made ways to build more complex intelligent systems in hardware constraint embedded platforms [5]. Recently there has been a boom in portable hand-held devices like tablet and smart phones. It would be good if these embedded systems can interact with a user using the emotional information, such as encoded in the human speech. These systems could have some sort of better intelligence that allows them to adapt and even “better” behave instead of just acting according to some given instructions. Some previous works that relates emotion and music applied to HRI applications [6], [7]. Qingli Zhang et. al. [8] used the technique based on Gaussian Mixture Model classifier for 6 emotions from Emo-DB and the algorithm provides 74.45% of accuracy. Thapanee Seehapoch et. al. [9], investigated on combined features of fundamental frequency, energy and MFCC using SVM linear kernel classifier giving 89.80%, 93.57% and 98% classification accuracy for Berlin, Japan and Thai emotions databases, respectively. Aditya Bihar et. al [10], presents a method based on GMM classifier and MFCC for recognizing the emotions from Assamese speeches and obtained an accuracy of 76.5%. A diverse number of approaches have been proposed and designed to extract facial features of the human face and generate an emotion based music playlist automatically. Additional hardware such as EEG, sensors etc. have been implemented into the design to automate the playlist generation which has indeed added up to the overall cost.

There are many drawbacks of the proposed systems

- It is very complex in terms of time and memory required for extracting facial features in real time.
- The proposed systems return less accurate playlist according to the current mood or emotional state of the individual.
- And with the use of additional hardware the overall cost has increased etc.

III. PROPOSED SYSTEM

This paper majorly focuses on introducing a system that is resolves most of the drawbacks of the existing system and generates playlists on the basis of the current emotional state of the individual using the facial expression of a human face therefore reducing the time and labour of the human as otherwise it needs to be done manually and thus avoiding the employment of additional hardware which in turn results in reduction in design cost. The aim is to build a system or player that has the capacity to segregate the list of songs and generate playlist according to the emotional state of the individual automatically with the need to be monitored by the humans. And which can process the query fast enough that is it is computationally fast and returns very accurate and precise results proving much superior efficient than the existing system.
A high accurate audio extraction technique is proposed that extracts significant, critical and relevant information from an audio signal based on certain audio features in a much lesser time. An emotion model is proposed that classifies a song based on any of the 8 classes of emotions viz sad, joy-anger, joy-surprise, joy-excitement, joy, anger, sad-anger and others. The emotion extraction module and audio feature extraction module is combined using an Emotion-Audio integration module. The proposed mechanism achieves a better efficiency and real time performance than the existing methodologies. SVM classifications are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. The classification of images takes place in SVM. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. This is also true of image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik.

IV. MODULE DESCRIPTION

The proposed system majorly involves 3 modules:

A. Emotion extraction module

B. Audio feature extraction module

C. Emotion-audio recognition module

Emotion extraction module and Audio feature extraction module are two separate modules where emotional expression is extracted and in latter audio corresponding to the emotion is extracted. Emotion-Audio recognition module performs the mapping of modules by querying the audio meta-data file. It is where the previous two modules are combined and the desired output is obtained that is the playlist according to the current mood or emotion of the individual.

Emotion extraction module

A static image of the human face can be obtained by capturing it by a webcam or can be accessed from the memory drive. The image in hand undergoes image enhancement in the form of tone mapping which restores the original contrast of the image. Once the image enhancement is done, the image is converted into binary image format. The face is detected using Viola and Jones algorithm where the Frontal Cart property of the algorithm is true of image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik.

Audio feature extraction module

In this module a list of songs are feed as input system. As songs are audio files, they require a certain amount of preprocessing Stereo signals obtained from the Internet are converted to 16 bit PCM mono signal around a variable sampling rate of 48.6 kHz. Audacity technique is used to convert the signal. The pre-processed signal obtained undergoes an audio feature extraction, where features like rhythm toning is extracted using MIR 1.5 Toolbox, pitch is extracted using Chroma Toolbox and other features like centroid, spectral flux, spectral roll off, kurtosis, 15 MFCC coefficients are extracted using Auditory Toolbox. Audio signals are categorized into 8 types viz. sad, joy-anger, joy-surprise, joy-excitement, joy, anger, sad-anger and others.

1. Songs that resemble cheerfulness, energetic and playfulness are classified under joy.

2. Songs that resemble very depressing are classified under the sad.

3. Songs that reflect mere attitude, revenge are classified under anger.

4. Songs with anger in playful is classified under Joy-anger category.

5. Songs with very depress mode and anger mood are classified under Sad-Anger category.

6. Songs which reflect excitement of joy is classified under Joy-Excitement category.

7. Songs which reflect surprise of joy is classified under Joy-surprise category.

8. All other songs fall under others category.

Emotion-audio recognition module

The extracted emotions are stored in the database as meta-data. Mapping is performed by querying the meta-data database. The emotion extraction module and audio feature extraction module is finally mapped and combined using an Emotion-Audio integration module. The below figure illustrates the mapping of facial features to audio features where there are 5 facial features namely sad, joy, anger, happy and excitement and 8 audio features namely joy, sad, anger, joy-anger, sad-anger, joy-excitement, joy-surprise and others. For example, if an input facial image is categorized under joy, the system will display songs under joy, joy-anger, Joy-Excitement, Joy-surprise category.
V. CONCLUSION

The Emotion-Based Music Player is used to automate and give a better music player experience for the end user. The application solves the basic needs of music listeners without troubling them as existing applications do: it uses technology to increase the interaction of the system with the user in many ways. It eases the work of the end-user by capturing the image using a camera, determining their emotion, and suggesting a customized play-list through a more advanced and interactive system. The user will also be noticed of songs that are not being played, to help them free up storage space. Testing of the system is done on both user dependent (dynamic) and user independent (static) dataset. Facial expressions are captured using an inbuilt camera. The accuracy of the emotion detection algorithm used in the system for real time images is around 85-90%, while for static images it is around 98-100%. The proposed algorithm on an average calculated estimation takes around 0.95-1.05 sec to generate an emotion based music playlist. Thus, it yields better accuracy in terms of performance and computational time and reduces the designing cost, compared to the algorithms used in the literature survey. The application can be improved by modifying and adding few functionalities. Current application uses Affectiva SDK that has a lot of limitations creating custom emotion recognition system that can be merged into the current application improves functionality and performance of the system.

- Making the application run without needing an internet connection.
- Including other emotions.
- Playing songs automatically.
- Optimizing the EMO-algorithm by including additional features which helps system to categorize user based on many other factors like location and suggesting the user to travel to that location and play songs accordingly.

VI. REFERENCES


[8]. Qingli Zhang, Ning An, Kunxia Wang, Fuji Ren and Lian Li, “Speech Emotion Recognition using Combination of Feature”,

Figure 2. Modules Mapping
