EmoTunes: Your Emotional Soundtrack

Dhanashree Tamkhane
Department of Computer Engineering
Vishwakarma Institute of Technology , Pune.
Pune, India.
dhanashree.tamkhane@vit.edu

Tejas Phalke
Department of Computer Engineering
Vishwakarma Institute of Technology , Pune.
Pune, India.
tejas.phalke22@vit.edu

Ashutosh Thakre
Department of Computer Engineering
Vishwakarma Institute of Technology , Pune.
Pune, India.
ashutosh.thakre22@vit.edu

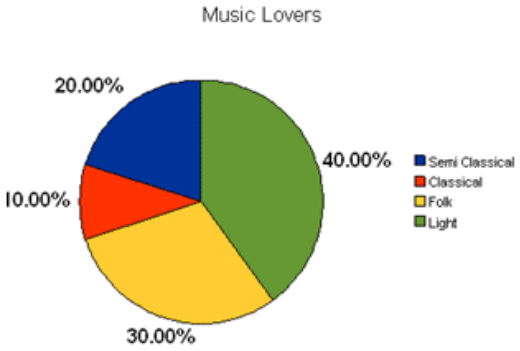
Nayana Thakur
Department of Computer Engineering
Vishwakarma Institute of Technology , Pune.
Pune, India.
nayana.thakur22@vit.edu

***Abstract*—The human face is an important organ of an individual's body and it especially plays an important role in extraction of an individual's behavior and emotional state. Manually segregating the list of songs and generating an appropriate playlist based on an individual's emotional features is a very tedious, time consuming, labor intensive and upheld task. Various algorithms have been proposed and developed for automating the playlist generation process.However the proposed existing algorithms in use are computationally slow, less accurate and sometimes even require use of additional hardware like EEG or sensors. This proposed system based on facial expression extracted will generate a playlist automatically thereby reducing the effort and time involved in rendering the process manually. Thus the proposed system tends to reduce the computational time involved in obtaining the results and the overall cost of the designed system, thereby increasing the overall accuracy of the system. 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.**

Keywords—Audio Emotion Recognition, Music Information Retrieval, Emotion Extraction Module, Audio Feature Extraction Module, Artificial Neural Networks, Confusion Matrix, Viola and Jones Face Detection

I. INTRODUCTION

Within the time of digital music consumption, personalized music suggestions have become integral to enhance user expe riences.
Conventional music proposal frameworks centre on compone nts like client inclinations, classes, and ubiquity, but they regularly disregard the enthusiastic perspective ofmusic.
[1]Recognizingthe importance of feelings in music delight,an Emotion-Based
Music Proposal Framework (EBMRS) develops as a novel approach.[6]



Graph 1. Music Lovers across the globe.

Byconsideringboth
the musical attributes and emotional signals,the
EBMRS promises to bridge the gap between innovation and human emotion, clearing the way for a more all-encompassing music suggestion approach.

II. LITERATURE SURVEY

Several methods have been proposed earlier to detect and recognize the facial features and audio features from an audio signal with certain algorithms. But there are very few systems that automatically generate a playlist based on the expressions detected which make use of some additional hardware like sensors or EEG. Some of which are discussed as below:

In Brain Computer Interface (BCI) - based mobile multimedia controller is proposed. It makes use of an external EEG hardware to monitor the cognitive state of mind. These systems require the user's active mental command continuously to control the multimedia. Most of the BCI systems require huge and costly EEG machines and commercial software's which limit the feasibility of the system for daily applications. [5] A MIDlet program that contains a cognitive detection algorithm is built in the mobiles that continuously monitor the EEG signals acquired from EEG machines, and then recognize the user's state of mind.

In the emotions are recognized by our voice/speech. The major concern of this system is the environment in which it is set up because if it is set up in an open environment then the surrounding noise will have a severe effect on the system's performance. To reduce the noise a voice activity detection algorithm has to be integrated to advance the system bandwidth efficiency which detects the speech from the input signal. It also reduces the occurrence of speech mismatch that can lead to a significant degradation of the emotion recognition rate.

In Electroencephalography (EEG) signals are used to detect the human emotions. EEG records the electrical activity of the brain within its neurons. EEG signals are mainly used due to the fact that it detects real emotions arising instantly from the mind ignoring all other outside characteristics like facial expressions or gestures. Two classifiers such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) are employed to categorize the EEG signals into seven emotions of an individual.[11] EEG signals are recorded by placing some electrodes on the scalp which are further processed to extract some features like Energy and Power Spectral Density (PSD) and are then fed to the two classifiers for the emotion detection. [4] Correlation Method, is one of the simplest methods also known as the nearest neighbor method. It returns the similarity score as the angle between two images. In this method the training images and test images are converted into column vectors. Comparison of the test image and gallery image is made in a high dimensional space rather in a low dimensional space which requires more storage space and hence the recognition time also increases leading to the disadvantage of correlation method.

In Geometric and Appearance Based Methods, it tracks the size and shape of the facial parts like eyes, eyebrows, lip corners, mouth, nose etc. To classify the expressions some shape models that are based on a set of characteristics points on the face are being used. However, the distance between the facial landmarks differs for different individuals which makes the system less reliable. Outward appearances include change in composition and hence filters like Gabor wavelets, Local Binary Pattern etc that are a portion of the appearance-based methods are applied either to whole of the face or to a particular face region to encode the surface. The facial patches vary according to the training data in these approaches based on their positions and sizes. Hence, it becomes hard to consider a standard system using these approaches. In PCA algorithm is used to extract the facial features and classify the expressions of an individual and from , the idea of automatically generating the song based on the expression detected is been used and combined to get the desired and efficient and robust system.

S.No.	Title	Tools Used
1.	BCI-based Mobile Multimedia Controller	EEG-based MIDlet Program
2.	Emotion Recognition from Voice/Speech	Voice Activity Detection Algorithm
3.	EEG-based Emotion Detection	E EG Signals, Support Vector Machine, Linear Discriminant Analysis
4.	Simplest Method for Emotion Recognition Using EEG Signals	Correlation Method
5.	Facial Expression Recognition Based on Landmark Tracking, Appearance-based Approaches	Geometric and Appearance-Based Methods
6.	Facial Feature Extraction and Expression Detection for Song Generation	PCA Algorithm
7.	Gaze-Based Music Playlist Generation	Eye-Tracking Technology, Machine Learning Algorithms
8.	Emotion-driven Music Selection System	Facial Expression Recognition System, Deep Learning Models

Table 1. Tools Used showcased by different research papers

III. METHODOLOGY

A. Need for the project

Music plays an important role in an individual's life. It is an important source of entertainment and is often associated with a therapeutic role. With the advent of technology and contiguous advancements in multimedia, sophisticated music players have been designed and have been enriched with numerous features, including volume modulation, genre classification etc. Although, these features successfully addressed the requirements of an individual, a user sporadically suffered through the need and desire of browsing through his playlist, according to his mood and emotions. 'This task was labour intensive and an individual often faced the dilemma of landing at appropriate list of songs. Hence, this application can be used which dynamically suggest songs from your playlist according to the mood the user is feeling for his/her satisfaction.Music plays an important role in an individuals life. It is an important source of entertainment and is often associated with a therapeutic role. With the advent of

technology and contiguous advancements in multimedia, sophisticated music players have been designed and have been enriched with numerous features, including volume modulation, genre classification etc. Although, these features successfully addressed the requirements of an individual, a user sporadically suffered through the need and desire of browsing through his playlist, according to his mood and emotions. 'This task was labour intensive and an individual often faced the dilemma of landing at appropriate list of songs. Hence, this application can be used which dynamically suggest songs from your playlist according to the mood the user is feeling for his/her satisfaction.

B. System Architecture

The system architecture for the proposed system is given in fig 3.1. The input image is loaded into the system in .jpg format. Then each image undergoes preprocessing i.e. removal of unwanted information like background colour, illumination and resizing of the images. Then the required features are extracted from the image and stored as useful information. These features are later added to the classifier where the expression is recognized with the help of Euclidean distance. Minimum the value of the distance calculated, the nearest the match will be found. Finally, a music track will be played based on the emotion detected of the user.

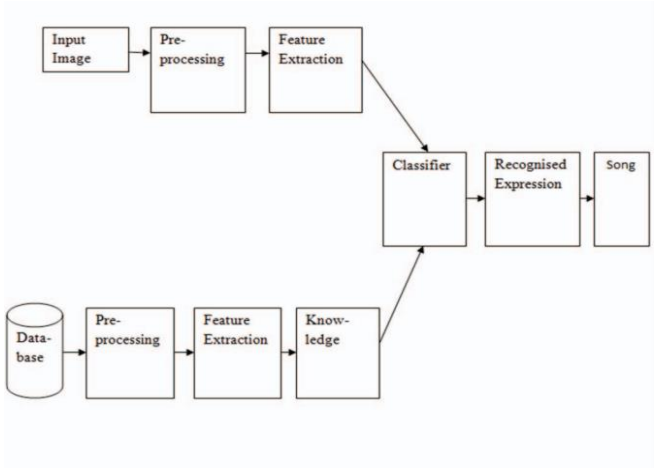


Fig. 1. System Architecture

C. Steps involved to design the system

To design the system, training dataset and test images are considered for which the following procedures are applied to get the desired results. The training set is the raw data which has large amount of data stored in it and the test set is the input given for recognition purpose. The whole system is designed in 5 steps:

1) Image Acquisition: In any of the image processing techniques, the first task is to acquire the image from the source. These images can be acquired either through camera or through standard datasets that are available online.

2) Pre-processing: During pre-processing, eyes, nose and mouth are considered to be the region of interest. It is detected by the cascade object detector which utilizes CNN and HAAR algorithm.

3) Facial Feature Extraction: After pre-processing, the next step is feature extraction. The extracted facial features are stored as the useful information in the form of vectors during training phase and testing phase. The following facial features can be considered "Mouth, forehead, eyes, complexion of skin, cheek and chin dimple, eyebrows, nose and wrinkles on the face". In this work, eyes, nose, mouth and forehead are considered for feature extraction purpose for the reason that these depict the most appealing expressions. With the wrinkles on the forehead or the mouth being opened one can easily recognise that the person is either surprised or is fearful. But with a person's complexion it can never be depicted. To extract the facial features CNN technique is used.

*4) Expression Recognition:*To recognize and classify the expressions of a person Euclidean distance classifier is used. It gets the nearest match for the test data from the training data set and hence gives a better match for the current expression detected. It is calculated from the mean of the eigenfaces of the training dataset. The training images that correspond to various distances from the mean image are labeled with expressions like happy, sad, fear, surprise, anger, disgust and neutral. When the Euclidean distance between the eigenfaces of the test image and mean image matches the distance of the mean image and eigenfaces of the training dataset the expression is classified and named as per the labeled trained images. Smaller the distance value obtained, the closest match will be found. If the distance value is large enough for an image then the system has to be trained for that individual. The equation to measure Euclidean distance between two points, say x and y is given as*:*

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}. \quad (1)$$

IV. RESULTS

Fig. 2. Suggesting songs depending on mood.

Fig. 3. Music Player

Fig. 4. Queuing up songs.

This music player records your emotions via camera and using the given techniques and algorithms, it suggests best suited songs for you. In this player, unlike other players, you can download songs. Also, you can queue up the songs accordingly.

Suppose mood detected is anger, then the player automatically suggests calm songs to you and vice versa.

Fig 5. Grayscale images processed by proposed algorithms

Images are converted to greyscale resized images and are given to the model in suitable format. Using haar – cascades, further procedure is undertaken.

Following table gives an idea of accuracy of mood detection using proposed method.

Emotion	Accuracy	
	Exact method	Average method
Neutral	55%	70%
Angry	60%	50%
Sad	40%	40%
Happy	100%	98%

V. FUTURE SCOPE

The future scope for the proposed system would be to implement it on mobiles. To design a mechanism that would help in the music therapy treatment for the music therapists to treat the patients suffering from mental stress, acute depression and trauma.

It can also be used to determine the mood of a physically challenged person. In the proposed work, only one emotion can be detected at a time so it can be extended to mixed mood detection by continuously recording the face of the user.

VI. CONCLUSION

The proposed work presents facial expression recognition system to play a song according to the expression detected. It uses PCA approach to extract features, and Euclidean distance classifier classifies these expressions. In this work, real images i.e. user dependent images are captured utilizing the in-built camera. The final result shows the accuracy level obtained is upto 84.82%.

VII. ACKNOWLEDGMENTS

We wish to express our sincere gratitude towards our mentor Manik Dhore Sir our professor-in-charge for their support and guidance in completing our project on the topic “EmoTunes: Your Emotional Soundtrack”. Without his superintendence, we would not have been able to create this project. Many people have directly and indirectly contributed in accomplishment of the project to whom we are grateful and lastly a thank you to all the team members for their constant presence and interest throughout the process.

REFERENCES

[1] A. Arora, A. Kaul and V. Mittal, "Mood Based Music Player," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 333-337, doi: 10.1109/ICSC45622.2019.8938384.

[2] Swathi Swaminathan and E. Glenn Schellenberg, “Current emotion research in music psychology,” Emotion Review, vol. 7, no. 2, pp. 189–197, Apr. 2015.

[3] K. S. Nathan, M. Arun and M. S. Kannan, "EMOSIC — An emotion based music player for Android," 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Bilbao, Spain, 2017, pp. 371-276, doi: 10.1109/ISSPIT.2017.8388671.

[4] S. G. Kamble and A. H. Kulkarni, "Facial expression based music player," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016, pp. 561-566, doi: 10.1109/ICACCI.2016.7732105.

[5] I. Cohen, A. Garg and T. S. Huang, Emotion Recognition from Facial Expressions using Multilevel HMM ,2015.

[6] Yading Song, Simon Dixon, and Marcus Pearce. Evaluation of Musical Features for Emotion Classification. In Proceedings of the 13th

International Society for Music Information Retrieval Conference, Portugal. pp. 523-528. October 2012.

[7] Thayer, R. E. The Biopsychology of Mood and Arousal. Oxford University Press, USA, 1989. [4] Nikalaou, N. Music emotion classification. PhD diss., Dissertation for the Diploma of Electronic and Computer Engineer, Technical University of Crete, 2011.

[8] Viola, Paul, and Michael Jones. ”Rapid object detection using a boosted cascade of simple features.” In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I-I. IEEE, 2001.

[9] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011) Florida, 2011.

[10] musiXmatch dataset, the official lyrics collection for the Million Song Dataset, available at: http://labrosa.ee.columbia.edu/millionsong/musixmatch

[11] McFee, Brian, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. ”librosa: Audio and music signal analysis in python.” In Proceedings of the 14th python in science conference, pp. 18-25. 2015.

[12] Giannakopoulos, Theodoros. ”pyaudioanalysis: An open-source python library for audio signal analysis.” PloS one 10, no. 12 (2015): e0144610.

[13] Kim, Youngmoo E., Erik M. Schmidt, Raymond Migneco, Brandon G. Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Turnbull. Music emotion recognition: A state of the art review. In Proc. ISMIR, Utrecht, pp. 255-266. 2010.

[14] Lucey, Patrick, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. ”The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression.” In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 94-101. IEEE, 2010.

[15] Jaiswal, Sushma. ”Comparison between face recognition algorithmeigenfaces, fisherfaces and elastic bunch graph matching.” Journal of Global Research in Computer Science 2, no. 7 (2011): 187-193.