Emotion Based Music Player

Vinayak Bali[1], Shubham Haval[2], Snehal Patil[3], R. Priyambiga[4]

[1,2,3]Student, Department of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur, India

[4]Assistant Professor, Dept. of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur, India

***Abstract*: Listening to music affects the human brain activities. Emotion based music player with automated playlist can help users to maintain a particular emotional state. This research proposes an emotion based music player that creates a playlists based on captured photos of the user. Manual sorting of a playlist and annotation of songs, in accordance with the current emotion, is more time consuming and quite tedious. Numerous algorithms have been implemented to automate this process. However, existing algorithms are slow, increase cost of the system by using additional hardware (e.g. EEG systems and sensors) and have quite very less accuracy. This paper presents an algorithm that not only automates the process of generating an audio playlist, but also to classify those songs which are newly added and the main task is to capture current mood of person and to play song accordingly. This enhances the system's efficiency, faster and automatic. The main goal is to reduce the overall computational time and the cost of the designed system. It also aims at increasing the accuracy of the system. The most important goal is to make change the mood of person if it is a negative one such as sad, depressed. This model is validated by testing the system against user dependent and user independent dataset.**

***Keywords*: Convolution neural network, Long Short term memory, Emotion detection, audio classification, hidden layers.**

1. Introduction

Expressing and recognizing emotions of human are very much important in communication system [3]. Human beings have the ability to express and recognize emotions. Computer seeks to identify the human emotions either by image analysis or through sensors [6]. In our day to day life and in our professional life we interact with many people face to face or indirectly by phone calls, sometimes it is necessary for people to be aware of their present emotions of the person with whom they are interacting. Human emotions are classified as: surprise, fear, anger, happy, sad, disgust and neutral [3].

Facial movement [1] and the tone of speech play a major role in expressing emotions. The physique and tone of the face tells the energy in the utterance of speech, which can be firstly modified to communicate different feelings. Humans can easily recognize these changes in signals along with the information felt by any other sensory organs. This project analyses the use of image or sensors or speech to capture the emotions.

Music plays [4] a vital role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners [15] and sometimes even imparts a therapeutic approach. In today's world, with lots of increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed, local playback, and streaming playback with multicast streams. Although these features satisfy the user's basic needs and requirements, yet the user has to face the task of manually browsing through the playlist of songs and select songs based on his current mood and behavior.

Emotions can be expressed through gestures, speech, facial expressions, body language etc. For the system to understand the user's mood, we use facial expression [3]. Using the mobile device's camera, we can capture the user's facial expression. There are many emotion recognition systems which take captured image as input and determine the emotion. For this application, we are using neural networks for recognition of emotion [9], [8].

2. Goal and objective

The main aim of our app is to play music [15] based on current emotion of the person and simultaneously change the negative emotions. The negative emotions can be changed [17] by playing appropriate music. This will also affect the thought process and induce positive emotions.

Why music?

1. All it takes 1 song to bring back 1000 memories.
2. Music is emotional life of most of the people.
3. Music is the movement of sound to reach the soul.
4. Sometimes music is the only thing that gets your mind off of everything else.
5. The most important things don't fit into a word, that's why there is music.

"Where words fail music speaks", and hence it can change person's negative emotion simultaneously and slowly into a positive mood [7].

We are detecting emotion through facial recognition and speech.

3. Literature survey

A. Existing work

The existing system [5] has only the boring traditional type music players in android, in which users can only simply listen to the songs that are already stored in the playlists and even if the user is in any sad or happy mood [7] all the songs in the playlist will be player by the player by the regular order, which the user may not like to listen to those songs and in that case, he/she has to manually change the song or skip the song.

Earlier the model was done using Machine learning and accuracy is less.

Currently, there are many existing music player applications. Some of the interesting applications among them are:

1) Saavan: These application gives good user accessibility features to play songs and recommends user with other songs of similar genre
2) Moodfuse: In this application, user should manually enter mood and genre that wants to be heard and mood fuse recommends the songs-list
3) Steromood: User should select his mood manually by selecting the moods from the list and the application plays music from YouTube.
4) Musicovery: This application has High quality songs and music recommendations. It also suggests predefined playlist for the user.
5) Gaana: It features music from 21 languages, user friendly and it allows users to make their playlists public so that they can be seen by other users. Supports all OS.

Fig. 1. Existing Applications

B. *Disadvantage of existing work*

The disadvantages of existing work are: The Application is manual. Whenever person is in bad mood, he doesn't want to speak to anyone, at that case the manually giving emotion doesn't work out. Accuracy is less. Person cannot able to hear songs which ever not present in the device. Newly added songs are not classified automatically.

C. *Proposed work*

This project is to capture the current emotion of the user and playing the corresponding song in the playlist according to the mood of the user and change the user's mood if it is negative emotion. The technique used is deep learning to recognize the facial expression and speech to detect the mood.

The main objective and our idea are to capture the emotions of human, to change human's negative emotion and also to categorize those songs which are not in the playlist, which are newly added.

D. *Advantages of proposed work*

The advantages of proposed system are: Extremely fast feature computation [11]. Provides efficient feature selection. Dynamically songs played. It provides increased Accuracy [8]. It is User-Friendly. Newly added songs are classified dynamically. Operation on speech of person is included. Changes negative emotion to positive one.

4. System architecture

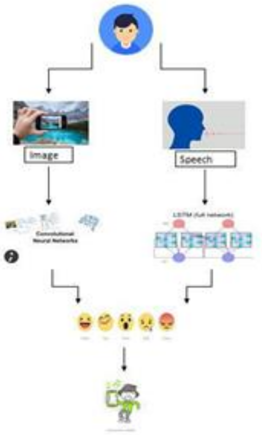


Fig. 2. Pictorial representation of system

The system architecture of Emotion-Based Music player is shown in Fig. 3. Application is built using the architectural pattern of Model-View-Controller. Here, the application is divided into three main components: Model, View and Controller.

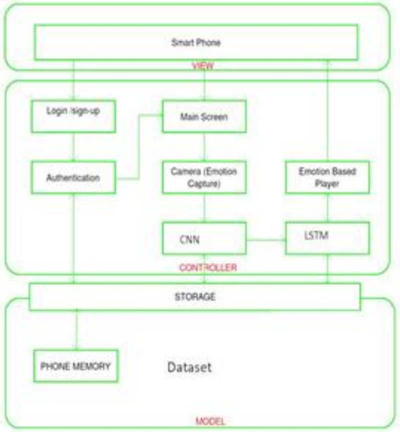


Fig. 3. System architecture

- *View:* The top layer is where the end-user communicates with the application by clicking buttons, typing details, accessing camera, selecting radio button, uploading songs, etc. This layer is meant for displaying all data or a portion of data to user on the requirement of application. This layer also acts as a bridge between the user and the application part. Android Application is used for displaying the output or response of the system to the user.

- *Controller:* This middle layer of consists of the business logic and the main functionality of application. Whenever the user interacts with the application, response is processed in this layer. From login to displaying the playlist, all the functions which run in the background belong to this layer. This mainly consists of all the important functions such as CNN [8] for emotion detection and LSTM [10] algorithm which helps in segregation of songs and sending output to output layer.
- *Model:* This layer is for maintaining the user's collection of data. Emotion Based Music Player uses trained dataset. This application also stores some temporary data on the device.

5. List of modules

Emotion based music system provides the generation of a customized playlist in accordance to the user's emotional state. The proposed system consists of three major modules:

A. Creation of music player

We are using Android studio, it is IDE for Google Android operating system. Android studio instant Run feature pushes code and resource changes to your running app. It clearly understands the changes and often sends them without restarting the app or rebuilding APK, so you can see the effects immediately. The code editor helps to write better code, work faster and more productive by offering advance code completion, refactoring, and code analysis. As we type, android studio provides suggestions in drop list. Simply press tab to insert in the code. Fast and feature- rich emulator installs your application faster than real device and allows you to prototype and test your app on various android studio.

B. Emotion extraction module

Image of a user is captured using camera or it can be accessed from the stored image in the phone. This acquired image undergoes image enhancement in the form of tone mapping in order to restore the original contrast of the image. After image enhancement all images are converted into binary image format and the face is detected. The Convolution Neural Network is used for facial emotion detection. It consists of hidden layers (n layers depending upon the dataset). The output layer shows the approximate result. Back propagation [14] is used for making the model error-free and producing the accurate result or match.

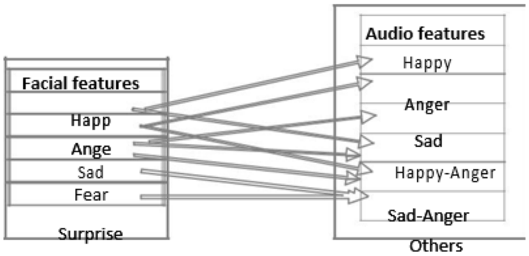
C. Audio feature extraction module

The songs which are stored in phone or songs which have been newly installing is need to be classified according to the mood or emotion. The lyrics of the song and also the frequency is extracted and processed to match the correct mood. It is done using LSTM (Long Short Term Memory) [10] Neural Network. It is a Sequence Algorithm and it consists of: Input gate, Forget gate, Output gate. It has special ability to remember and forget the contents, so making the model more efficient. Large sequential data can be able to classified or predicted using LSTM.

D. Emotion-audio integration module

The captured face of the person or user is firstly detected and the result is mapped or matched with the classified songs list and then the song is played accordingly.

Table 1
Mapping modules



6. Methodology

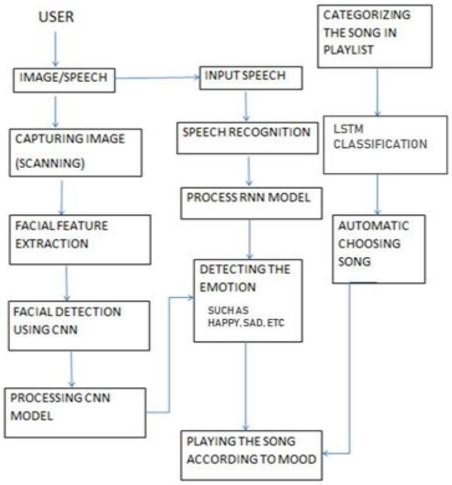


Fig. 4. Block diagram

The proposed algorithm revolves automated music recommendation system that plays song according to the mood or current emotion of the person. The person's photo is captured whenever the application gets open; hence current emotion is captured and detected. According to the information given by the image, the song is played related to the emotion. The songs whichever present in the phone are already classified into 7 different classes [7] such as happy, sad, anger, surprise, fear, disgust and neutral. The newly added songs are also classified dynamically to appropriate mood. It is composed of three modules: Facial expression recognition module, Song emotion recognition module and System integration module. Facial expression recognition and audio emotion recognition modules are two mutually exclusive modules. Hence, system integration module maps two modules to find the correct match of detected emotion.

A. Data set

The Raw dataset is downloaded one by one from Google images for seven emotions. Extra dataset is taken from Kaggle datasets for facial expression detection.

B. Trained dataset

Before processing the model, the training and testing phases [12] is undergone. Trained dataset are those which are taught to the model or which learns. Sample trained dataset is:

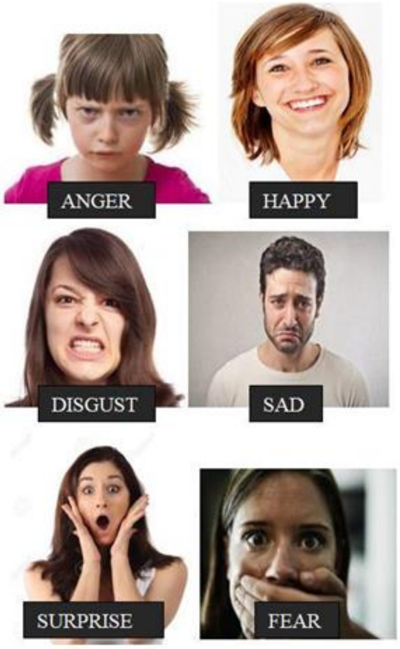


Fig. 5. Training dataset

At the time of training [12], system takes dataset of faces (images) with their respective expression; eye should be in centre location mostly and learns a set of weights, which splits the facial expressions for classification.
For training, the sequence is:

1. Spatial normalization
2. Synthetic samples generation
3. Image cropping
4. Down-sampling
5. Intensity normalization.

C. Test data

Fig. 6. Test dataset

At the time of testing, classifier takes images of face with respective eye center locations, and it gives output as predicted expression by using the weights learned during training.

For recognizing an unknown image (testing), the sequence is:

1. Spatial normalization
2. Image cropping
3. Down-sampling
4. Intensity normalization

D. Convolution neural network

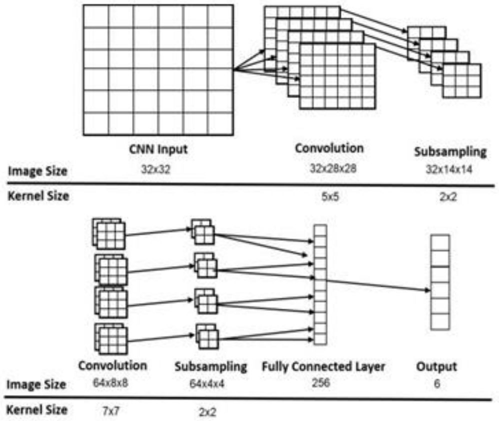


Fig. 7. Architecture of proposed convolution neutral network

It has five layers: first layer is convolution type, second layer is sub-sampling type reduces the map size by half, the third layer is convolution type, fourth layer is sub-sampling type reduces the map once more by half, fifth layer is fully connected type and final representing each one of the expression are responsible for classifying facial image.

7. Final resultant model

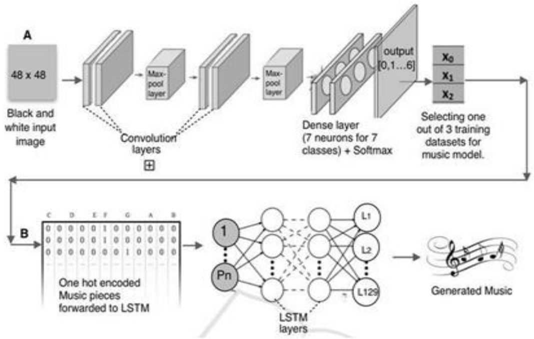

A
48 x 48
Black and white input image
Max-pool layer
Convolution layers
Max-pool layer
Dense layer (7 neurons for 7 classes) + Softmax
output [0,1…6]
X₀
X₁
X₂
Selecting one out of 3 training datasets for music model.
B
C D E F G A B
One hot encoded Music pieces forwarded to LSTM
1
Pn
L1
L2
L129
LSTM layers
Generated Music

Fig. 8. Final Model (Resultant)

All the photos present in dataset are firstly converted to grayscale, for making preprocessing and detection more efficiently, faster [9] and easier. Each input image is in form of pixels (e.g. 48x48). Now the pixel represented images are sent to the convolution layers (hidden layers) [8]. In Between each layer Maximum pooling is done, the purpose of doing so, is to down-sample the input data or image, reducing the dimensions and allows assumption to be made about features [11] contained in sub regions. This is done to avoid over-fitting [13]. As well as it reduces computational cost by reducing number of parameters to learn. Example, if input image is of matrix 4x4 representation and let's say output we want is in 2x2, then pooling is performed in between all hidden layers. After that data is sent to dense layer, to prevent over-fitting. Dropout technique is used to reduce over-fitting in neural networks. The output layer conveys the detected class. Let's say if the detected expression is happy, then the next step is to select anyone training dataset for music model. Now, the dataset is trained according to the match for playing music. LSTM [10] neural network is used for classifying the songs. One hot encoding is performed to represent categorical variables into binary vectors, so as to make the classification faster and better. Then the song is played according to the current mood of the person.

8. Conclusion

The Emotion Based Music Player is used to automate and give a best music player experience for end user. Application solves all 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 numerous ways. It eases the work of the user by capturing image and detecting the suggesting a customized playlist through more advanced and interactive features. The user's negative or bad thoughts are slowly converted to positive thoughts by changing the song from low tone to excited tone. Effectively categorize the songs based on the detected mood by using LSTM algorithm. The resultant system proposed to have increased accuracy in the range of 80% to 99%.

9. Future scope

In future Music Player can be enhanced with Google play music, so songs which are not present in local storage can also be played and to access the whole application in speech based. The Emotion Based Music System will be of great advantage to users looking for music based on their mood and emotional behavior [17]. It will help reduce the searching time for music thereby reducing unnecessary time and hence increasing the overall accuracy and efficiency of the system. The system will not only reduce physical stress but will also act as a boon for the music therapy systems and may also assist the music therapist to treat the patient. In future it can also be used to detect the sleepy mood of the driver, driving the car and many more uses. Also with its additional features mentioned above, it will be a complete system for music lovers and listeners.

References

[1] Y. Wu, H. Liu, and H. Zha, "Modeling facial expression space for recognition," in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005. (IROS 2005), 2005, pp. 1968–1973.

[2] S. Z. Li and A. K. Jain, Handbook of Face Recognition. Springer Science & Business Media, 2011.

[3] Mireille Besson, Frederique Faita, Isabelle Peretz, A- M Bonnel, and Jean Requin. Singing in the brain: Independence of lyrics and tunes. Psychological Sci- ence, 9(6):494–498, 1998.

[4] S. Dornbush, K. Fisher, K. McKay, A. Prikhodko and Z. Segall, Xpod- A Human Activity and Emotion Aware Mobile Music Player, UMBC Ebiquity, November 2005.

[5] G. Heamalathal, C.P. Sumathi, A Study of Techniques for Facial Detection and Expression Classification, International Journal of Computer Science & Engineering Survey(IJCSES) Vol.5, No. 2, April 2014.

[6] Alvin I. Goldmana, B. Chandra and SekharSripadab, "Simulationist models of face-based emotion recognition.

[7] Thayer, The biopsychology of mood & arousal, Oxford University Press, 1989.

[8] S. Demyanov, J. Bailey, R. Kotagiri, and C. Leckie, "Invariant back-propagation: how to train a transformation-invariant neural network," arXiv:1502.04434 [cs, stat], 2015.

[9] D. C. Cirean, U. Meier, J. Masci, L. M. Gambardella, and J. Schmid-huber,"Flexible, high performance convolutional neural networks for image classification," in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two, ser. IJCAI'11, AAAI Press, 2011, pp. 1237–1242.

[10] Colah's blog, Understanding LSTM Networks Posted on August 27, 2015.

[11] Debaditya, R., Sri, R. M. K., and Krishna, M. C. (2015). Feature selection using deep neural networks.

[12] Yadan O, Adams K, Taigman Y, et al. Multi-GPU Training of ConvNets [J]. Eprint Arxiv, 2013.

[13] J.E. Moody. The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In Advances in Neural Information Processing Systems, volume 4, pages 847-854. Morgan Kaufmann, 1992.

[14] Yan, P., Huang, R.: Artificial Neural Network — Model, Analysis and Application. Anhui Educational Publishing House, Hefei.

[15] "Impact of Music, Music Lyrics, and Music Videos on Children and Youth". Pediatrics. 124 (5): 1488–1494. 19 October 2009.

[16] Radford, C. (1989). "Emotions and music: A reply to the cognitivists". The Journal of Aesthetics and Art Criticism.

[17] Wigram T. Indications in music therapy: Evidence-based practice. British Journal of Music Therapy 2002.

[18] Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. (2012). "Improving neural networks by preventing co-adaptation of feature detectors".

[19] Startups, Requests for (2016-08-12). "Deep Learning in Healthcare: Challenges and Opportunities". Medium. Retrieved 2018-04-10.

[20] Bengio, Y. (1991). "Artificial Neural Networks and their Application to Speech/Sequence Recognition". McGill University Ph.D. thesis.

[21] Abdel-Hamid, O.; et al. (2014)."Convolutional Neural Networks forSpeechRecognition" (PDF). IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[22] B. Fasel, "Robust face analysis using convolutional neural networks," in 16th International Conference on Pattern Recognition, 2002. Proceedings, vol. 2, 2002, pp. 40–43 vol.2.

[23] F. Beat, "Head-pose invariant facial expression recognition using convolutional neural networks," in Fourth IEEE International Conference on Multimodal Interfaces, 2002. Proceedings, 2002, pp. 529–534.

[24] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," Neural Networks: The Official Journal of the International Neural Network Society, vol. 16, no. 5, pp. 555– 559, 2003.

[25] P. Zhao-yi, W. Zhi-qiang, and Z. Yu, "Application of mean shift algorithm in real-time facial expression recognition," in International Symposium on Computer Network and Multimedia Technology, 2009. CNMT 2009, 2009, pp. 1–4.

[26] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Work- shops (CVPRW), 2010, pp. 94–101.

[27] C. D. Caleanu, "Face expression recognition: A brief overview of the last decade," in 2013 IEEE 8th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2013, pp. 157–161.

[28] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial ex- pressions with gabor wavelets,” in Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings, 1998, pp. 200–205.

[29] M. Dahmane and J. Meunier, “Emotion recognition using dynamic grid-based HoG features,” in 2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011), 2011, pp. 884– 888.

[30] X. Glorot, “Understanding the difficulty of training deep feed forward neural networks,” (2010).