An Accurate Algorithm for Generating a Music Playlist based on Facial Expressions

Anukriti Dureha
Computer Science and Engineering Department
Amity School of Engineering & Technology,
Amity University, Noida, India

ABSTRACT

Manual segregation of a playlist and annotation of songs, in accordance with the current emotional state of a user, is labor intensive and time consuming. Numerous algorithms have been proposed to automate this process. However the existing algorithms are slow, increase the overall cost of the system by using additional hardware (e.g. EEG systems and sensors) and have less accuracy. This paper presents an algorithm that automates the process of generating an audio playlist, based on the facial expressions of a user, for rendering salvage of time and labor, invested in performing the process manually. The algorithm proposed in this paper aspires to reduce the overall computational time and the cost of the designed system. It also aims at increasing the accuracy of the designed system. The facial expression recognition module of the proposed algorithm is validated by testing the system against user dependent and user independent dataset. Experimental results indicate that the user dependent results give 100% accuracy, while user independent results for joy and surprise are 100 %, but for sad, anger and fear are 84.3 %, 80 % and is 66% respectively. The overall accuracy of the emotion recognition algorithm, for user independent dataset is 86%. In audio, 100 % recognition rates are obtained for sad, sad-anger and joy-anger but for joy and anger, recognition rates obtained are 95.4% and 90 % respectively. The overall accuracy of the audio emotion recognition algorithm is 98%. Implementation and testing of the proposed algorithm is carried out using an inbuilt camera. Hence, the proposed algorithm reduces the overall cost of the system successfully. Also, on average, the proposed algorithm takes 1.10 sec to generate a playlist based on facial expression. Thus, it yields better performance, in terms of computational time, as compared to the algorithms in the existing literature.

Keywords

Audio Emotion Recognition, Music Information Retrieval, Facial Expression Recognition, Music Recommendation systems, Audio Feature Extraction.

1 INTRODUCTION

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. Using traditional music players, a user had to manually browse through his playlist and select songs that would soothe his mood and emotional experience. This task was labor intensive and an individual often faced the dilemma of

landing at an appropriate list of songs. The advent of Audio Emotion Recognition (AER) and Music Information Retrieval (MIR) equipped the traditional systems with a feature that automatically parsed the playlist, based on different classes of emotions.

While AER deals with categorizing an audio signal under various classes of emotions, based on certain audio features , MIR is a field that relies upon exploring crucial information and extracting various audio features (e.g. pitch, energy, flux, MFCC, kurtosis etc.) from an audio signal.

Although AER and MIR augmented the capabilities of traditional music players by eradicating the need of manual segregation of playlist and annotation of songs, based on a user's emotion, yet, such systems did not incorporate mechanisms that enabled a music player to be fully controlled by human emotions. Emotions are synonymous with the aftermath of an interplay between an individual's cognitive gauging of an event and the corresponding physical response towards it. Among the various ways of expressing emotions, including human speech and gesture, a facial expression is the most natural way of relaying them.

A facial expression is a discernible manifestation of the emotive state, cognitive activity, motive, and psychopathology of a person. Homo sapiens have been blessed with an ability to interpret and analyze an individual's emotional state. However, machines were deprived of a complex brain like a human's, which could recognize, distinguish, and perceive different emotions accurately. Ergo, an urge to craft sophisticated intelligent systems, equipped with such skills persisted. The field of Facial Expression recognition (FER) synthesized algorithm that excelled in furnishing such demands. FER enabled the computer systems to monitor an individual's emotional state effectively and react appropriately.

While various systems have been designed to recognize facial expressions, their time and memory complexity is relatively high and hence fail in achieving a real time performance. Their feature extraction techniques are also less reliable and are responsible for reducing the overall accuracy of the system. Numerous algorithms have been published to recognize emotions in an audio signal. However, the existing algorithms are less accurate. They yield unpredictable results and often increase the overall memory overheads of the system. Their information retrieval algorithms are less efficient. They lack the capability of extracting significant and relevant information from an audio signal, in minimal time. The existing audio emotion recognition algorithms employ mood models that are loosely coupled with the perception of a user. Also, the state-of-the art is deprived of designs, capable of fostering a customized playlist by inferring human emotions, conveyed by a facial image, without exhausting

additional resources. The existing designs either employ additional hardware (like EEG systems and sensors) or uses human speech. Hence this paper proposes a methodology that aims at eradicating the drawbacks and shortcomings of the existing technology.

The underlying objective of this paper is to design an accurate algorithm that would yield a list of songs from a user's playlist in conformance with a user's emotional state. The algorithm designed, requires less computational time, storage and reduces the cost incurred in employing additional hardware. This paper proposes a highly accurate mechanism for recognizing facial expressions, owing to its accurate and reliable feature extraction technique. It categorizes a facial image under 5 different facial expressions viz. Joy, Sad, Anger, Surprise and Fear. The proposed mechanism is capable of achieving real time performance. This paper also proposes a highly accurate audio information retrieval approach that extracts significant and relevant information from an audio signal in minimal time. It employs a mood model that captures the perception of a user accurately. The mood model categorizes a song under one of the five classes of emotions including Joy, Sad, Anger, Joy-Anger and Sad-Anger. The integration of both the methods is performed using an efficient system integration module. The algorithm proposed in this paper yields better performance than the existing state-of -the-art methodologies.

This paper is organized as follows: Section II gives the literature review. Section III gives the methodology. Section IV describes the experiments performed and the results obtained and finally Section V concludes the paper and gives the future scope.

2 LITERATURE SURVEY

Several approaches have been proposed and have been adopted to classify human emotions successfully. Most of these approaches have laid their focus on seven basic emotions, attributed to their stability over culture, age and other identities. Facial features, for the purpose of feature recognition, have been classified by zheng et. al [20] under two broad categories viz. Appearance-based features and Geometric features. The geometric features were derived from shape or prominent points of some important facial features such as mouth and eyes. In the work of Changbo et. al [2], 58 landmark points were considered to craft an ASM. The appearance based features, such as texture, have also been employed in different works. Amongst them Michael lyons [4] et. al proposed a methodology for coding facial expressions with multi-orientation and multi-resolution set of Gabor filters, that were ordered topographically and were aligned approximately, with the face. The degree of correlation obtained was significantly high, but the overall computational complexity increased exponentially.

Renuka R. Londhe et al. [16], proposed a statistical based approach for analyzing facial expression. In this paper, they studied the changes in curvatures on the face and intensities of corresponding pixels of images. ANN was used to classify these features in to six universal emotions such as anger, disgust, fear, happy, sad, and surprise. Two-layered feed forward neural network was trained and tested using Scaled Conjugate Gradient back-propagation algorithm and obtained 92.2 % recognition rate.

To eradicate the need and reduce the labor required for manual annotation of songs in a playlist, in accordance with different categories of emotions, several approaches and designs have been proposed. Most of these approaches rely on

an Arousal-valence model proposed by Jung Hyun Kim or the 2-dimenesional (Stress vs energy) model proposed by Thayer. In Jung Hyun Kim's[7] work, the collected music mood tags and A-V values from 20 subjects were analyzed and the A-V plane was classified into 8 regions, depicting mood by using k-means clustering algorithm. Thayer [19] came up with a dimensional model, plotted along two axes (Stress versus energy), with mood represented by a two-dimensional co-ordinate system, lying on either of the two axes or the four quadrants formed by the two-dimensional plot.

The field of human emotion aware music player hasn't enticed much attention and holds a massive amount of scope for research. K.McKay et. al [17] designed xpod-a human activity and emotion aware music player. Their system employed sensors to collect information related to a user's emotions and activities for music recommendation. Their system was based upon a client/server architecture. Karlos A. Cervantes[5] proposed an embedded design of an emotion aware music player, that used emotions in speech to control a music player. Some existing designs have also used EEG-based systems to measure human emotions and have used it to control a music player.

While various approaches have been proposed to recognize facial emotions and emotions in an audio signal, very few systems have been designed to control the generation of a music playlist using human emotions. The existing designs developed and researches published to control playlist generation through human emotions, either make use of additional hardware like EEG systems and sensors or uses human speech. The work presented in this paper aims to resolve various drawbacks in the existing literature and are as follows

i. Existing systems are highly complex in terms of time and storage for recognizing facial expressions in a real environment.

ii. Existing systems lack accuracy in generating a playlist based on the current emotional experience of a user.

iii. Existing systems employed additional hardware like EEG systems and sensors that increased the overall cost of the system.

iv. Some of the existing system imposed a requirement of human speech for generating a playlist, in accordance with a user's facial expressions.

This paper aims to resolve these drawbacks by designing an automated music recommendation system that would generate a customized playlist based on a user's facial expression without using any additional hardware.

3 METHODOLOGY

In this paper, the proposed algorithm revolves around an automated music recommendation system that generates a subset of the original playlist or a customized playlist in accordance with a user's facial expressions. A User's facial expression helps the music recommendation system to decipher the current emotional state of a user and recommend a corresponding subset of the original playlist. It is composed of three main modules: Facial expression recognition module, Audio emotion recognition module and a System integration module. Facial expression recognition and audio emotion recognition modules are two mutually exclusive modules. Hence, the system integration module maps the two sub-spaces by constructing and querying a meta-data file. It is composed of meta-data corresponding to each audio file. Figure 1 depicts the flowchart of the proposed algorithm.
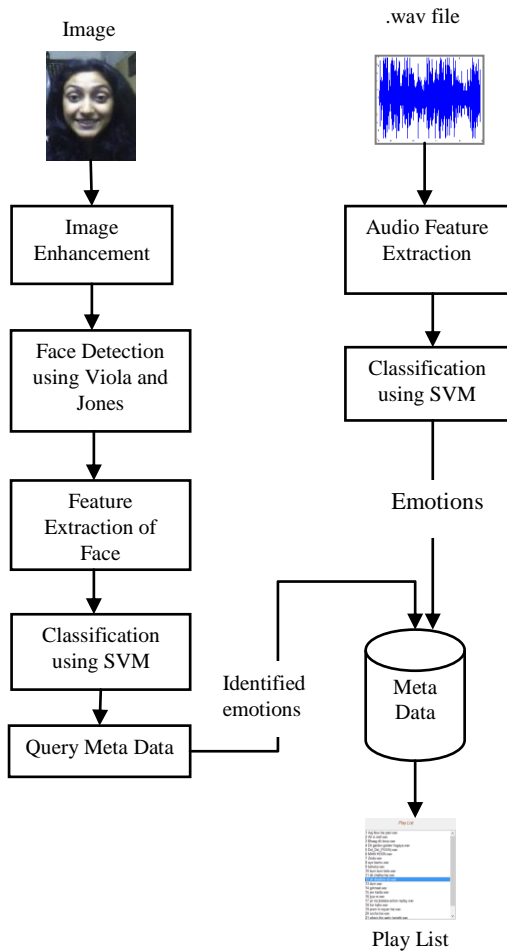
Figure 1 Flowchart of Proposed Algorithm

3.1 Facial Expression Recognition

The input image to the system can be captured using a web cam or can be acquired from the hard disk. This image undergoes image enhancement, where tone mapping is applied to images with low contrast to restore the original contrast of the image.
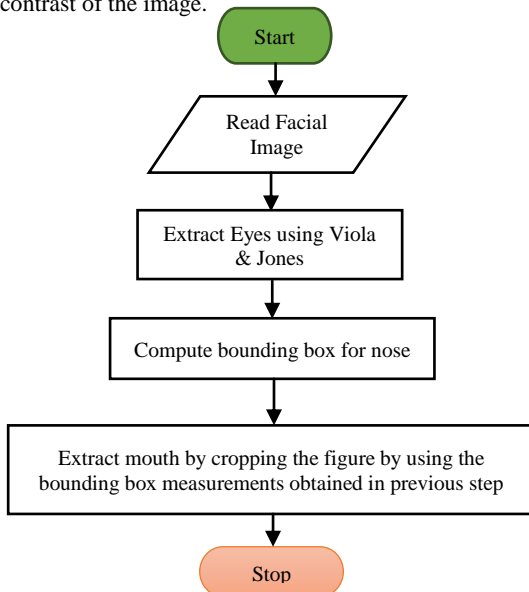

Figure 2 Feature Extraction from Facial Image

All RGB and gray scale images are converted into a binary image. This preprocessed image is fed into the face detection block. Face detection is carried out using viola and jones algorithm. The default property of 'Frontal Cart' with a merging threshold of 16 is employed to carry out this process. The 'frontal cart' property only detects the frontal face that are upright and forward facing. Since, viola and jones, by default, produces multiple bounding boxes around the face, merging threshold of 16 helps in coagulating these multiple boxes into a single bounding box. In-case of multiple faces, this threshold detects the face closest to the camera and filters out all the distant images. The facial image obtained from the face detection stage forms an input to the feature extraction stage. To obtain real time performance and to reduce time complexity, for the intent of expression recognition, only eyes and mouth are considered. The combination of two features is adequate to convey emotions accurately. Figure 2 depicts the schematic for feature extraction.

The binary image obtained from the face detection block forms an input to the feature extraction block, where eyes and mouth are extracted. Eyes are extracted using the viola and jones method, however, to extract mouth, certain measurements were considered. To extract mouth, first bounding box for nose is calculated using viola and jones and then using the bounding box for nose, bounding box of mouth is deduced. Equations (i), (ii), (iii) and (iv) depict the bounding box calculations employed for extracting mouth from a facial image

$$X_{start_{mouth}} = X_{Mid_{Nose}} - \left(X_{end_{nose}} - X_{start_{nose}}\right) \qquad \text{(i)}$$

$$X_{end_{mouth}} = X_{Mid_{Nose}} + \left(X_{end_{nose}} - X_{start_{nose}}\right) \qquad \text{(ii)}$$

$$Y_{start_{mouth}} = Y_{Mid_{Nose}} + 5 \qquad \text{(iii)}$$

$$Y_{end_{mouth}} = Y_{start_{mouth}} + 90 \qquad \text{(iv)}$$

Where ($X_{start_{mouth}}$, $Y_{start_{mouth}}$) and ($X_{end_{mouth}}, Y_{end_{mouth}}$) are starting and ending coordinates of the bounding box for mouth respectively,($X_{Mid_{Nose}}, Y_{Mid_{Nose}}$) are coordinates of mid -point of nose and $X_{end_{nose}}, X_{start_{nose}}$ are x coordinates of ending and starting of nose respectively.

Figure 3 Annotations for Bounding Box Calculation

Figure 3 is an example depicting various annotations including $X_{start_{mouth}}, Y_{start_{mouth}}, X_{Mid_{Nose}}, Y_{Mid_{Nose}}$, $X_{end_{nose}}, X_{start_{nose}}$ and bounding box calculations. Training and classification is carried out using support vector machine. Since, SVM was originally designed for binary-classification, to classify faces among 5 classes of emotions, a multi-class

classifier was desired. Hence, training and classification is performed using 'one-vs-all' approach of SVM which successfully facilitates multi-class classification.

3.2 Audio Emotion Recognition

In Music Emotion recognition block, the playlist of a user forms the input. An audio signal initially undergoes certain amount of preprocessing. Since, music files acquired from the internet are usually stereo signals, all stereo signals are converted to 16 bit PCM mono signal at a sampling rate of 44.1 kHz. The conversion is performed using Audacity. Conversion of a stereo signal into mono signal is crucial to reduce the mathematical complexity of processing the similar content of both the channels of a stereo signal. An 80 second window is then extracted from the entire audio signal. Audio files are usually very large in size. They are computationally expensive in terms of memory. Hence, to reduce the memory complexity of an audio file incurred during audio feature extraction, an 80 sec window is extracted from each audio file. This also ensures uniformity in terms of size of each audio file. During extraction, first 20 seconds of an audio file are discarded and are not considered. This helps in ensuring an efficient retrieval of significant information from an audio file. Figure 4 depicts the process of extraction of an 80 second window.
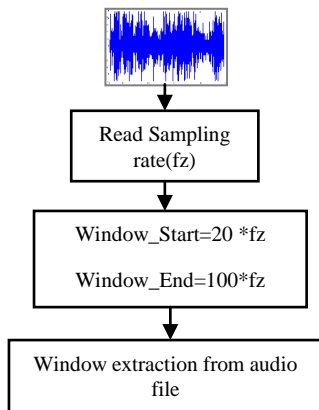
```
┌─────────────────────────┐
│   Read Sampling         │
│      rate(fz)           │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│  Window_Start=20 *fz    │
│                         │
│  Window_End=100*fz      │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│  Window extraction from │
│       audio file        │
└─────────────────────────┘
```

Figure 4 Audio Preprocessing

This pre-processed signal then undergoes audio feature extraction, where centroid, spectral flux, spectral roll off, kurtosis, zero-crossing rate and 13 MFCC coefficients are extracted. Toolboxes used for audio feature extraction includes MIR 1.5, Auditory Toolbox and Chroma toolbox. Music based emotion recognition is then carried out using Support Vector Machine's 'one vs other' approach. Audio signals are classified under 6 categories viz. Joy, Sad, Anger, Sad-Anger, Joy-Anger and Others. While various mood models have been proposed in the literature, they failed in capturing the perceptions of a user in real time. They are figments of theoretical aspects that researchers have associated with different audio signals. The mood model and the classes considered in this paper takes into account, how a user may associate different moods with an audio signal and a song. While composing a song, an artist may not maintain the uniformity of a mood across the entire excerpt of a song. Songs are usually based on a theme or a situation. A song may boast sadness in first half while the second half may become cheerful. Hence the model adopted in this paper takes into consideration all these aspects and generates paired classes for such songs, apart from the individual classes. The model adopted is as follows

i.   All those songs that are cheerful, energetic and playful are classified under joy.
ii. Songs that are very depressing are classified under the class sad.
iii. Songs that reflect attitude, 'anger associated with patriotism', and are revengeful are classified under anger.
iv. The category Joy-Anger is associated with songs that possess anger in a playful mode.
v.   Sad-anger category is composed of all those songs that revolve around the theme of being extremely depressed and angry.
vi. All other songs apart from these general categories falls under the 'other' category.
vii. When a user is detected with emotions such as surprise and fear, the songs from the other category are suggested.
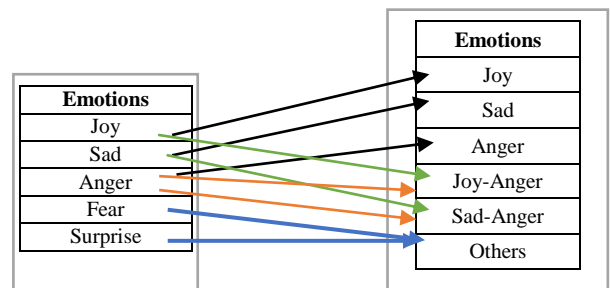

Figure 5   Mapping of Modules

Figure 5 depicts the mapping of facial emotion recognition module and audio emotion recognition module. The name of the file and emotions associated with the song is then recorded in a database as its meta-data. The final mapping between the two blocks is carried out by querying the meta-data database. Since, Facial Emotion recognition and Audio emotion recognition modules are two mutually exclusive components, system integration is performed using an intermediate mapping module that relates the two mutually exclusive blocks with the help of the audio Meta data file. Figure below depicts the mapping of Facial emotions and Audio emotions. While fear and surprise categories of facial emotions are mapped onto the 'Others' category of audio emotions, each of joy, sad and anger categories of facial emotions are mapped onto two categories in audio emotions as shown in diagram below. During play list generation, when the system classifies a facial image under one of the three emotions, it considers both individual as well as paired classes. For example, if a facial image is classified under joy, the system will display songs under joy and joy-anger categories

4   RESULTS AND EXPERIMENTS

Implementation and experimentation in this paper is carried out using MATLAB R2013a on Windows8, 32 bit operating system and Intel's i5 M460(2.53 GHz) processor . For facial emotion recognition, 2 experiments were carried out.

i.   Using user independent dataset
ii. Using user dependent dataset.

Testing was first carried out using Cohn- Kanade dataset and then to achieve real time performance, self-annotated dataset was used. In the self-annotated dataset, images clicked using web cams of various users were collected. User independent dataset comprised of 30 individuals from the Cohn-Kanade dataset, while real time performance of the system was

achieved by testing the system against 15 individuals. User dependent dataset comprised of 10 individuals.

Table 1 Image Type and Size Used for Testing

Type of Dataset	Image Type	Image Size
Cohn-Kanade	Gray scale	490 X640
Self-Annotated	RGB	240X320,480X648, 480X340 & 318X284

Table 1 gives the type and size of the images used for testing and experimentation for Cohn-Kanade dataset and the self-annotated dataset.

4.1 Facial Emotion Recognition

Various types of experiments were carried out to evaluate the performance of the facial emotion recognition module. These experiments were broadly classified under two types: user independent and user dependent emotion classification. User independent experiments were carried out for 30 individuals. For joy, all 30 images were classified under joy, for sad, 25 images were classified under sad, while 5 were classified under anger. For anger, 25 were classified under anger, 4 under sad and 2 were classified under fear. For fear, 20 images were classified under fear and 10 were classified under anger.

Table 2 Confusion Matrix for User-Independent Experiments

Emotions	Joy	Sad	Anger	Surprise	Fear
Joy	30	0	0	0	0
Sad	0	25	5	0	0
Anger	0	4	24	0	2
Surprise	0	0	0	30	0
Fear	0	0	10	0	20

Table 2 depicts user independent results for 30 images. While joy and surprise yielded 100 % recognition rates, sad, anger and fear yielded 83 %, 80 % and 67 % recognition rates respectively.

User dependent experiments were carried out on static images and images captured using web-cam. In these experiments, classifiers were separately trained for each user. Static Images were collected from Cohn-Kanade dataset and self-annotated dataset.

Table 3 Confusion Matrix for User Dependent Experiments for User S011

Emotions	Joy	Sad	Anger	Surprise	Fear
Joy	6	0	0	0	0
Sad	0	6	0	0	0
Anger	0	0	6	0	0
Surprise	0	0	0	6	0
fear	0	0	0	0	6

Table 4 Confusion Matrix for User Dependent Experiments for User S026

Emotions	Joy	Sad	Anger	Surprise	Fear
Joy	5	0	0	0	0
Sad	0	5	0	0	0
Anger	0	0	5	0	0
Surprise	0	0	0	5	0
fear	0	0	0	0	5

Table 5 Confusion Matrix for User Dependent Experiments for User S136

Emotions	Joy	Sad	Anger	Surprise	Fear
Joy	8	0	0	0	0
Sad	0	8	0	0	0
Anger	0	0	8	0	0
Surprise	0	0	0	8	0
fear	0	0	0	0	8

Table 3, 4 and 5 gives user dependent results for 3 users. These are based on static images obtained from Cohn-Kanade database. For all users, all emotions produces 100% recognition rates, i.e. all 6 images for each emotion for user S011 were classified successfully under their respective emotion, similarly all 5 images for user S026 were classified successfully and all 8 images for user S136 were classified successfully.

Table 6 Web Cam results for Facial Emotion Recognition

Parameters	Joy	Sad	Anger	Surprise	Fear
Similar Image as database	15	15	15	15	2
Head tilted towards right	10	0	0	3	0
Head tilted towards left	13	0	0	1	0
Slightly Less intense	15	0	12	15	0

Table 6 depicts the web cam results for facial emotion recognition. These results are calculated based on different parameters for a self -annotated dataset for 15 individuals. Various parameters have been evaluated to determine the accuracy of the system.

Table 7 Time Taken by Various Modules of Facial Expression Recognition Algorithm

Module	Time Taken(sec)
Feature Extraction	0.9896
Classification	0.0231
Emotion Recognition	1.0144

Table 7 gives the average time taken by the facial expression recognition algorithm and its feature extraction and classification modules. The time lag between the time taken

by the emotion recognition algorithm and the total time taken by the feature extraction and classification modules is due to certain experimental errors and the time taken by certain auxiliary processes, such as loading of a mat file.

4.2 Audio Emotion Recognition

For Audio emotion recognition, a dataset was carved out of 100 songs for training and 200 songs for testing. These songs were collected from various sites for Bollywood music, including Songs.pk, Mp3 skull etc. Later these were self-annotated under different categories of emotions.

Table 8 Confusion Matrix for Audio Emotion Recognition

Emotions	Joy	Sad	Anger	Sad-Anger	Joy-Anger
Joy	37	0	0	0	3
Sad	0	40	0	0	0
Anger	0	1	39	0	0
Sad-Anger	0	0	0	40	0
Joy-Anger	0	0	0	0	40

Table 8 depicts the confusion matrix for audio emotion recognition.

Table 9 Accuracy of Different Classes of Audio Emotion Recognition

Emotions	Accuracy
Joy	92%
Sad	100%
Anger	97.5%
Sad-Anger	100%
Joy -Anger	100%

Table 9 depicts the accuracy of different classes of audio emotion recognition. While 100 % recognition rates were achieved for sad, sad-anger and joy-anger, 92% and 97.5 % recognition rates were achieved for joy and anger respectively.

4.3 Human Emotion Aware Music Player

Testing and experimentation of the designed system is carried out using the inbuilt webcam (USB 2.0 UVC VGA Webcam) and hence, the total cost incurred in automating the system is nil.

Table 10 Time Taken by Various Modules of the Proposed Algorithm

Module	Time Taken (Sec)
Facial Emotion Recognition	1.0144
System Integration	0.0008
Proposed Algorithm	1.10

Table 10 gives the average time taken by the proposed algorithm (i.e. algorithm to generate playlist, based on facial

expressions) and its facial expression recognition and system integration modules.

5 CONCLUSION AND FUTURE WORK

The algorithm proposed in this paper aims to control playlist generation based on facial expressions. Experimental results indicate that the proposed algorithm was successful in automating playlist generation on the basis of facial expressions and hence reduced the amount of labor and time, incurred in performing the task manually. The use of web cams helped in eradicating the requirement of any additional hardware, such as EEG systems and sensors, and thus helped in curtailing the cost involved. Since, audio emotion recognition of songs is not performed in real time and a meta data for all the audio file was deduced, the total time taken by the algorithm is equal to the amount of time taken by the algorithm to recognize facial expressions and the amount of time taken by the algorithm to query the meta data file. Hence, the proposed algorithm yields better performance, in terms of computational time, than the algorithms reported in the existing literature. Also, since the time taken by the algorithm to query the Meta data file is negligible (0.0008 sec), the total time taken by the algorithm is proportional to the time taken to recognize facial expressions.

Experimental results of the facial emotion recognition algorithm indicate that training the system with a user dependent dataset gives 100% accuracy, whilst, recognition rates with user independent dataset for joy and surprise are 100 % , for sad, anger and fear, the recognition rates are 84.3 %, 80 % and 66% respectively. Also, table 7 indicates that the average time taken by the classification algorithm is relatively less as compared to the feature extraction algorithm. Hence, the time taken by the emotion recognition algorithm is approximately equal to the time taken by the feature extraction algorithm (neglecting the time lag incurred due to experimental results). The total accuracy of the emotion recognition algorithm for user dependent dataset and user independent dataset is 100%and 86 % respectively. The high accuracy of the algorithm suggests that the feature extraction algorithm employed by the system is highly reliable and outperforms the existing algorithms. Since, the time taken by the emotion recognition is less (1.0144 sec), the proposed algorithm excelled in achieving a real time performance.

Since, the first 20 seconds of an audio file, during training and classification, were discarded, the proposed algorithm was successful in extracting relevant and significant information from an audio signal. The overall accuracy of the audio emotion recognition algorithm is 98%. This indicates that the information retrieval mechanism employed by the audio emotion recognition algorithm is highly efficient. Further, the mood model, employed in the algorithm, excelled in capturing the perception of a user in real time.

The proposed algorithm was successful in crafting a mechanism that can find its application in music therapy systems and can help a music therapist to therapize a patient, suffering from disorders like acute depression, stress or aggression.

The system is prone to give unpredictable results in difficult light conditions, hence as part of the future work, removing such a drawback from the system is intuited.

6 REFERENCES

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

[2] A. habibzad, ninavin, Mir kamalMirnia," A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization."

[3] Byeong-jun Han, Seungmin Rho, Roger B. Dannenberg and Eenjun Hwang, "SMERS: music emotion recognition using support vector regression", 10[th]ISMIR , 2009.

[4] Chang, C. Hu, R. Feris, and M. Turk, "Manifold based analysis of facial expression," Image Vision Comput ,IEEE Trans. Pattern Anal. Mach. Intell. vol. 24, pp. 05–614, June 2006.

[5] Carlos A. Cervantes and Kai-Tai Song , "Embedded Design of an Emotion-Aware Music Player", IEEE International Conference on Systems, Man, and Cybernetics, pp 2528-2533 ,2013.

[6] Fatma Guney, "Emotion Recognition using Face Images", Bogazici University, Istanbul, Turkey 34342.

[7] Jia-Jun Wong, Siu-Yeung Cho," Facial emotion recognition by adaptive processing of tree structures".

[8] K.Hevener,"The affective Character of the major and minor modes in music", The American Journal of Psychology,Vol 47(1) pp 103-118,1935.

[9] Kuan-Chieh Huang, Yau-Hwang Kuo, Mong-Fong Horng ,"Emotion Recognition by a novel triangular facial feature extraction method".

[10] Michael lyon and Shigeru Akamatsu, "Coding Facial expression with Gabor wavelets.", IEEE conf. on Automatic face and gesture recognition, March 2000.

[11] P. Ekman, W.V. Friesen and J.C. Hager, "The Facial Action Coding System: A Technique for the Measurement of Facial Movement",2002.

[12] Simon baker and Iain Matthews," Lucas-Kanade 20 Years On: A Unifying Framework", International Journal of Computer Vision,,*vol* 56(3), pp 221–255, 2004.

[13] Spirosv. Ionnau, Amaryllis T.Raouzaiou, VazilisA.tzouvaras,"Emotion Recognition though facial expression analysis based on neurofuzzy network".

[14] Samuel Strupp, Norbert Schmitz, and KarstenBerns, "Visual-Based Emotion Detection for Natural Man-Machine Interaction".

[15] Russell, "A circumplex model of affect", Journal of Personality and Social Psychology Vol-39(6), pp1161–1178 , 1980.

[16] Renuka R. Londhe, Dr. Vrushshen P. Pawar, "Analysis of Facial Expression and Recognition Based On Statistical Approach", International Journal of Soft Computing and Engineering (IJSCE) Volume-2, May 2012.

[17] 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.

[18] Sanghoon Jun, Seungmin Rho, Byeong-jun Han and Eenjun Hwang, "A fuzzy inference-based music emotion recognition system",VIE,2008.

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

[20] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE. Transaction Pattern Analysis, vol 31, January 2009.