Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6, July, 2021: 10158 - 10168

Research Article

Implementation of Novel Music Player Based on Speech and Text Emotion Recognition for Mood Uplift

Shrikala Deshmukh [a], Preeti Gupta [b]

[a] Research Scholar, Amity University, Mumbai, India
[b] Assistant Professor, College of Engineering, Bharati Vidyapeeth, India

Abstract

It is observed that music possesses human mood regulation ability. Establishing congruence between music and mood uplift is an important domain of study for facilitating emotional elevation. The work implements a novel speech and text emotion recognition based music player for mood uplift. The study takes into consideration five important human emotions namely Happy, Sad, Angry, Fear and Bored. The novelty is introduced in the implementation of music player due to its ability to recognize emotions through speech and text inputs provided by the user. Emotion recognition through speech is based on identification of pitch and amplitude of the speech and later subjecting it to algorithmic technique of Probabilistic Neural Network (PNN). Text input is processed using Support Vector Machine (SVM) algorithm for analysing emotions in the text. The working database consists of about 100 songs per emotion in Hindi Language. The most widely used scales to measure mood, Positive and Negative Affect Schedule (PANAS) mood scale has been employed to design the database of around 500 Hindi Songs. Result reveals accuracy around 95.76% of the music player for recognizing human emotion based on speech and text and playing the songs accordingly for mood uplift. On comparison with other algorithms, it is statistically revealed that the proposed model outperforms the other existing techniques used for speech and text emotion recognition.

Keywords: speech emotion recognition, text emotion recognition, probabilistic neural network, support vector machine, positive and negative affect schedule mood scale

1. Introduction

Music has a high impact on people's brain activity. People listen to music based on their mood and interests. Music has become a part of human life as it plays a key role in expressing feelings.

In the recent years, there has been a tremendous increase of computerized music libraries. An ever increasing number of individuals these days have increased simple access to computerized melodic substance while being out and about, or in their homes through their advanced mobile phones, tablets, savvy TVs, and so on. Administrations like Spotify, Pandora or Lastfm, offer an abundance of music content that clients can peruse and arrange dependent on their sort, ethnicity, time, feeling, and so on. It is in this way there is an enormous need to create computerized frameworks that will help look, compose, and sort music content and related information.

Availability of online music libraries has expanded approaches for selecting the songs and has also made it complex as user might get confused or lost in the large dataset. In the presented work an attempt is made to categorize Hindi popular music area under specific moods.

Segregation of songs and generation of an appropriate playlist based on an individual's emotional features is a very tedious and time consuming task. Emotion recognition is the process of identifying human emotion. The human emotions – Happy, Sad, Angry, Fear and Bored - can be recognized through analysis of Visual/facial

expressions, Speech or Text inputs. Here the use of speech and text for emotion recognition is proposed and presented.

The task of speech emotion recognition is very challenging for the reasons like variety of sentences, different speakers, various speaking styles, and speaking rates (Basharirad and Mohammadreza, 2017; Tamulevičius et al., 2020). Another challenging issue is the cultural differences in how a certain emotion is expressed by a speaker. There are mainly two types of speech emotion recognition: a) Speaker Dependent, wherein energy and pitch are used as features for emotion recognition and b) Speaker Independent, in which the focus is on "what was said" regardless "who said it" (Deshmukh and Devulapalli, 2020).

Detection of human emotions in text is becoming increasingly important for various applications. Text emotion recognition can be treated as content-based classification, involving concepts of natural language processing and machine learning. Recognizing and analysing emotion from text is very challenging task involving various steps.

In the present work a novel speech and text emotion recognition based music player that will assist for mood uplift is proposed. The paper outline consists of Section 2 discusses literature survey of music player based on emotion recognition; Section 3 elaborates motivation. Section 4 elaborates proposed system on mood enhancing music player using speech emotion recognition and text emotion recognition; It also consists of database creation of songs using PANAS scale that will uplift the mood. Section 5 highlights the implementation of proposed system; Section 6 discusses on results and analysis. Section 7 describes conclusion and future scope.

2. Literature survey

Emotion recognition is mainly divided into various aspects such as face emotion recognition, speech emotion recognition and text emotion recognition. Music player using face emotion recognition is implemented by various people. In this work a new implementation of music player using speech emotion recognition and text emotion recognition is demonstrated.

A music player is implemented by (Kabani H, et al., 2015) - for emotion recognition they had used face emotion recognition system. Emotion based music player is implemented by (Hemanth P, et al., 2018) – here again for emotion recognition they had used face emotion recognition system. The emotions can be recognized using Support Vector Machine (SVM) algorithm of machine learning. Eye, eyebrow and lip features of face emotion recognition is used by (Ghule V, et al., 2017). Neural network is also now used for emotion recognition. Real-time extraction of facial expressions as well as extracting audio features from songs to classify into a specific emotion that will generate a playlist was implemented in EMOSIC - a mobile application developed by (Nathan, et al., 2017). Facial expression based music player is implemented by (Vijay Chakole, et al., 2018). Implementation of music player based on face emotion recognition is implemented by (Mistry, et al., 2017) – here input is directly extracted from the human face employing a camera. A music player based on facial expression is proposed by (Prof. Jayshree Jha, et al., 2015). This proposed model uses Viola-Jones algorithm. SVM classifier is used for feature extraction and classification. An effective music player which is integrated on user's mood is implemented by (Rohit Kadam, et al., 2018). Human expressions are extracted by facial features using inbuilt camera and plays song according to the user's mood. A mood based music system is implemented by (A.S.Mali, et al., 2018). The emotions are interpreted from facial expressions using webcam. CNN classifier is used to build a neural network model.

3. Motivation

As it is highlighted that most work done on emotion recognition is using face/facial expressions (Nathan, et al., 2017; Prof. Jayshree Jha, et al., 2015) there are no instances where speech and text emotion recognized music player are implemented. Also, existing methods for automating the playlist generation process are computationally slow, less accurate and sometimes even require use of additional hardware like EEG or sensors. There are some systems which require manual selection of current emotion from list of predefined emotions.

The motivation for the proposed work is to - correctly identify user's emotion by using speech emotion recognition & text emotion recognition and play song according to the emotion of user. Here the mood is uplifted by playing motivational songs.

4. Proposed system

Proposed system is a music player which uplifts mood based on speech emotion and text emotion recognition. This system plays music as per predicted emotion. Input from user is accepted in the form of speech or text and processed using speech emotion and text emotion recognition. The Figure. 1 shows the proposed system architecture.

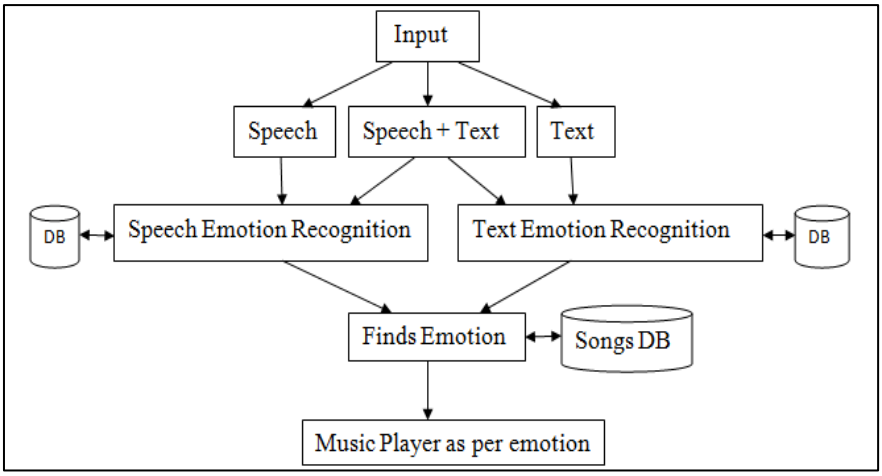


Figure 1. Proposed system architecture.

The mood uplifting technique is used for enhancing mood. In mood uplifting technique, motivational songs are used to change feeling/mood of a person from sad/bored to happy. Motivational songs to change mood is based on songs mapped to elevate the mood are used. Here, 5 emotions - happy, sad, angry, fear and bored are considered. In this system, the mood of person is uplifted from sad to happy, bored to excited, angry to calm and fear to strong. A database of songs has been created based on the empirical method of classification by studying the nature of songs and what people listen to for several emotions.

Speech emotion recognition

Speech emotion recognition is the act of attempting to recognize human emotion. This is based on the fact that voice generally reflects underlying emotion through pitch and frequency.

A python model is built to recognize emotion from speech by means of the librosa and sklearn libraries and the EMO_DB dataset. This dataset is having various actors each speaking a number of sentences for several emotions.

The Figure. 2 highlights the zcr (zero crossing rate), rolloff, spec_bw(spectral bandwidth), spec_cent(spectral centroid), chroma_stft(Short time fourier transform) features that are extracted from recorded audio file.

Date	Time	chroma_stft	spec_cent	spec_bw	rolloff	zcr
24-06-2020	21:20	0.486079763	1432.903265	1675.909102	2643.949355	0.072844448
25-06-2020	11:18	0.635418368	879.2331573	1525.811363	1845.580893	0.022910415
25-06-2020	11:19	0.637715974	934.9113867	1566.702291	1983.871116	0.02435424
25-06-2020	11:19	0.600572748	991.6459342	1615.385619	2130.182814	0.028334864
25-06-2020	11:20	0.566136153	1093.573937	1640.411496	2187.901781	0.039423052
25-06-2020	11:20	0.634012043	1000.632419	1572.7018	2058.417354	0.032241114
25-06-2020	11:21	0.486524434	1222.667424	1741.313033	2478.956533	0.043685004
25-06-2020	11:50	0.717211182	671.2725199	1296.960534	1327.989108	0.014728874
25-06-2020	13:33	0.715189276	557.1003013	1166.732795	1022.179159	0.0125642
28-12-2020	17:59	0.88677657	2586.710653	2258.971064	5025.460815	0.103366428
28-12-2020	18:14	0.88138145	4515.910305	3243.468252	8564.632161	0.178148058
29-12-2020	12:08	0.89029354	2500.158994	2354.385986	5070.820109	0.058666088
29-12-2020	12:14	0.87665313	4502.039201	3237.096973	8549.379476	0.180467394
29-12-2020	12:40	0.8768371	4534.396098	3226.082283	8552.868652	0.202110008
29-12-2020	12:44	0.8795903	4991.288079	3091.738444	8752.698771	0.393373843
29-12-2020	12:59	0.8711997	4684.216398	3196.481801	8636.758423	0.23232241
05-01-2021	12:59	0.5330015	2220.303169	2197.171782	4276.384481	0.127762406
05-01-2021	16:54	0.58543676	1937.776228	1506.055878	3551.333618	0.142763491
05-01-2021	16:55	0.54916126	1970.098995	1565.998352	3597.839355	0.135357892
05-01-2021	16:56	0.48528093	2410.226059	2183.352221	4765.068563	0.144404659

Figure 2. Features extracted from sound file.

The model is then trained and tested. From the input dataset or the input speech the noise removal pre-processing is initially done. The number of features are extracted and using the Probabilistic Neural Network algorithm the emotion is predicted. The Figure. 3 shows the flowchart for training and testing with PNN as classifier.

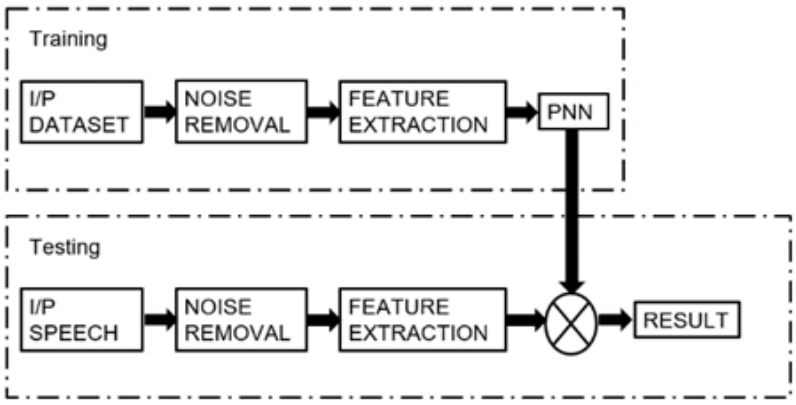


Figure 3. Speech emotion recognition flowchart.

Probabilistic neural network

Probabilistic neural network (Specht, 1990; Mao et al., 2000) is a feedforward neural network widely used in classification problems. PNN is used since the recognition is a difficult task when single emotion conveys multiple information. PNN has faster training ability with continuous class probability density functions. It provides better classification even with reduced feature set. Result shows remarkable classification accuracy for these classes of emotions.

When an input is given, the input layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The pattern layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities and produces a 1 (positive identification) for targeted class and a 0 (negative identification) for non-targeted classes as highlighted in Figure. 4.

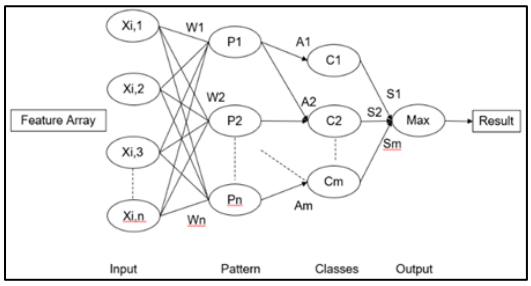


Figure 4. PNN classifier.

$S \in \{A, B, C, D, E\}$

$F = P(A \cap B \cap C \cap D \cap E) = 0$ ideally

$G = P(A \cap B \cap C \cap D \cap E) < 0.01$ practically

$H = P(A \cup B \cup C \cup D \cup E)$

$H = P(A) + P(B) + P(C) + P(D) + P(E) - G$

$$P(K) = \sum_{n=A}^{E} P(n \neq K) - G$$

where K – A, B, C, D, E

A – feature set of Happy

B – feature set of Sad

C – feature set of Anger

D – feature set of Bored

E – feature set of Fear

Text emotion recognition

A python model is built to recognize emotion from text by means of sklearn libraries. Databases used are lexicon dictionary & BBC database.

The model is then trained and tested. From the input dataset or the input text the stop words pre-processing is initially done. The number of features are extracted and using the Support vector machine (SVM) algorithm the emotion is predicted. The Figure. 5 shows the flowchart for training and testing with SVM as classifier.

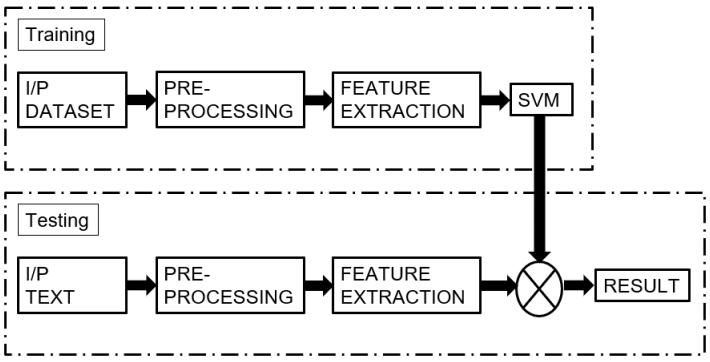


Figure 5. Text emotion recognition flowchart.

For text emotion recognition (Calefato et al., 2017; Shivhare and Khethawat, 2012; Chopade, 2015), support vector machine algorithm is used with Bag of Words (BOW) (Soumya and Shibily, 2014).

Step 1: Import the data

Step 2: Pre-processing the data

Remove non-alphabetic characters - remove unwanted characters such as numbers and punctuation marks.

Remove the stop words - stop words in text classification are words that doesn't have any impact on deciding the class of the text. For example, words like the, we, a, will, and etc.

Word tokenization - this tokenizer will tokenize the text and create a list of words.

Stemming - stemming is the process of finding the base word.

Step 3: Feature extraction

Step 4: Use a model for classification and find accuracy

Support vector machines

A Support Vector Machine (SVM) is used for text emotion recognition (Batoul and Nishith, 2020; Sarkar et al., 2015; Basu et al., 2003; Joachims, 1998). A type of supervised machine learning classification algorithm is used. SVM differs from the other classification algorithms in the way that it chooses the decision boundary that maximizes the distance from the nearest data points of all the classes. The SVM finds the most optimal decision boundary.

The most optimal decision boundary is the one which has maximum margin from the nearest points of all the classes. The nearest points from the decision boundary that maximize the distance between the decision boundary and the points are called support vectors as seen in Figure. 6.

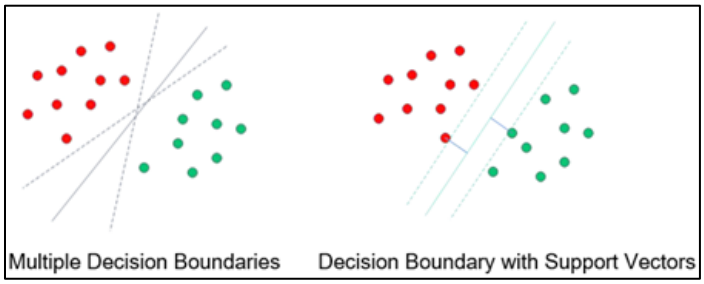


Figure 6. SVM classifier.

The formula of linear kernel is as below:

k(x,xi)=sum(x*xi)

From the above formula, we can see that the product between two vectors say $x\&xi$ is the sum of the multiplication of each pair of input values.

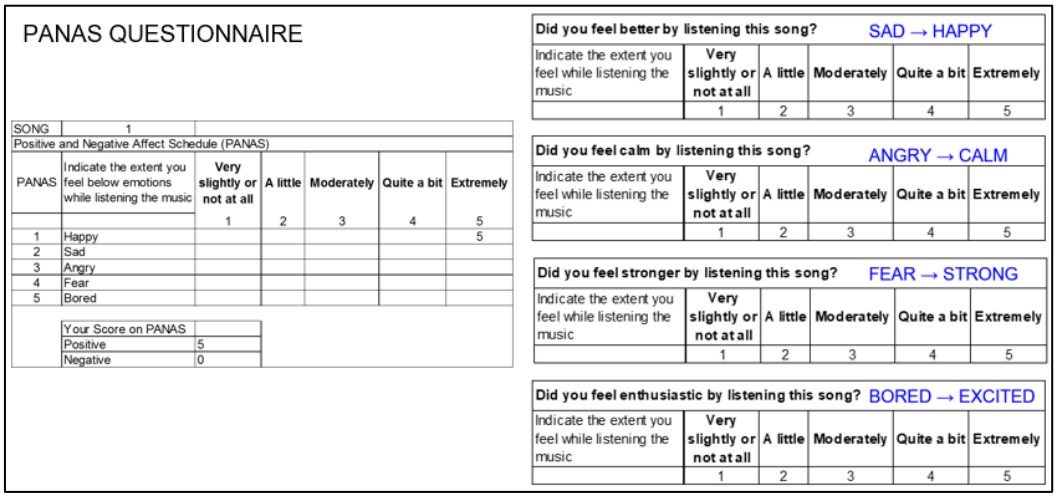
Creation of songs database

A database of Hindi songs for selected emotions is created. The database consists of 100 songs per emotion. Since, music is unquestionably important in shaping moods, the database is created using mood uplifting technique. If the user is sad/ bored, then instead of playing sad/boring songs, the music player will play randomly selected motivational music, which will motivate the user and assist in changing his/her mood. For creation of songs database around 700 Hindi songs are selected randomly from various music libraries. These songs are divided based on selected emotions as – Happy, Sad, Angry, Fear, Bored.

In order to determine that the other users also feel the same emotions while listening to the selected songs; The Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988; Crawford and Julie, 2004) is used for songs database creation. A questionnaire was shared with around 100 participants along with the songs database. PANAS is one of the most widely used scales to measure mood or emotion.

PANAS questionnaire shared consisted of a 5-point Likert scale in order to record the response from participants - 1. Very Slightly or Not at All, 2. A Little, 3. Moderately, 4. Quite a Bit, 5. Extremely

The Figure. 7 highlights the PANAS questionnaire. The participants were requested to indicate the extent they felt the selected 5 emotions while listening to the song. The questionnaire also captures that fact whether the participant felt mood change from sad→happy, angry→calm, fear→strong, bored→excited emotion. The participants were requested to share their feedback, if their mood is uplifted.



PANAS QUESTIONNAIRE

SONG	1					
Positive and Negative Affect Schedule (PANAS)						
PANAS	Indicate the extent you feel below emotions while listening the music	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
		1	2	3	4	5
1	Happy					5
2	Sad					
3	Angry					
4	Fear					
5	Bored					

Your Score on PANAS	
Positive	5
Negative	0

Did you feel better by listening this song? SAD → HAPPY					
Indicate the extent you feel while listening the music	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
	1	2	3	4	5

Did you feel calm by listening this song? ANGRY → CALM					
Indicate the extent you feel while listening the music	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
	1	2	3	4	5

Did you feel stronger by listening this song? FEAR → STRONG					
Indicate the extent you feel while listening the music	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
	1	2	3	4	5

Did you feel enthusiastic by listening this song? BORED → EXCITED					
Indicate the extent you feel while listening the music	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
	1	2	3	4	5

Figure 7. PANAS questionnaire.

5. Implementation

In this section the implementation of a novel speech and text emotion recognition based music player for mood uplift is discussed.

The Figure. 8 shows the welcome screen; registration; login and select input mode of the music player which is based on python.

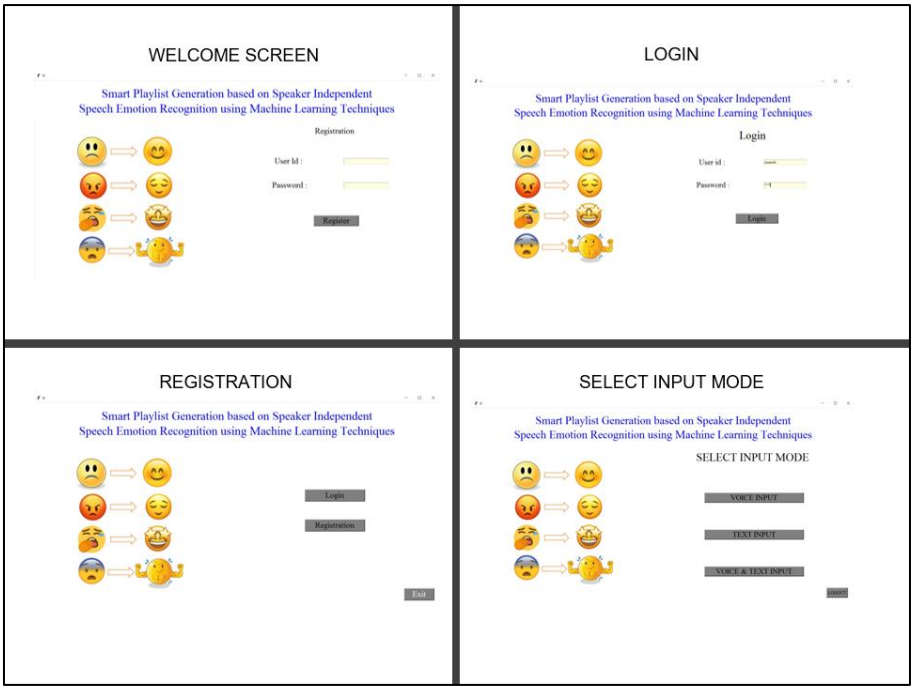


Figure 8. Implementation screen.

The purpose of the registration is to prevent unintended login and usage. The registration is one time activity. Once the registration is successfully done; with the login tab and credentials the music player is started.

Once in the music player, there are 3 input modes available for user –

Voice input – speech emotion recognition

Text input – text emotion recognition

Voice & text input – combination of speech and text emotion recognition

The Figure. 9 shows the voice; text; voice + text input page. For the voice input, the pitch of voice is shown along with predicted emotion.

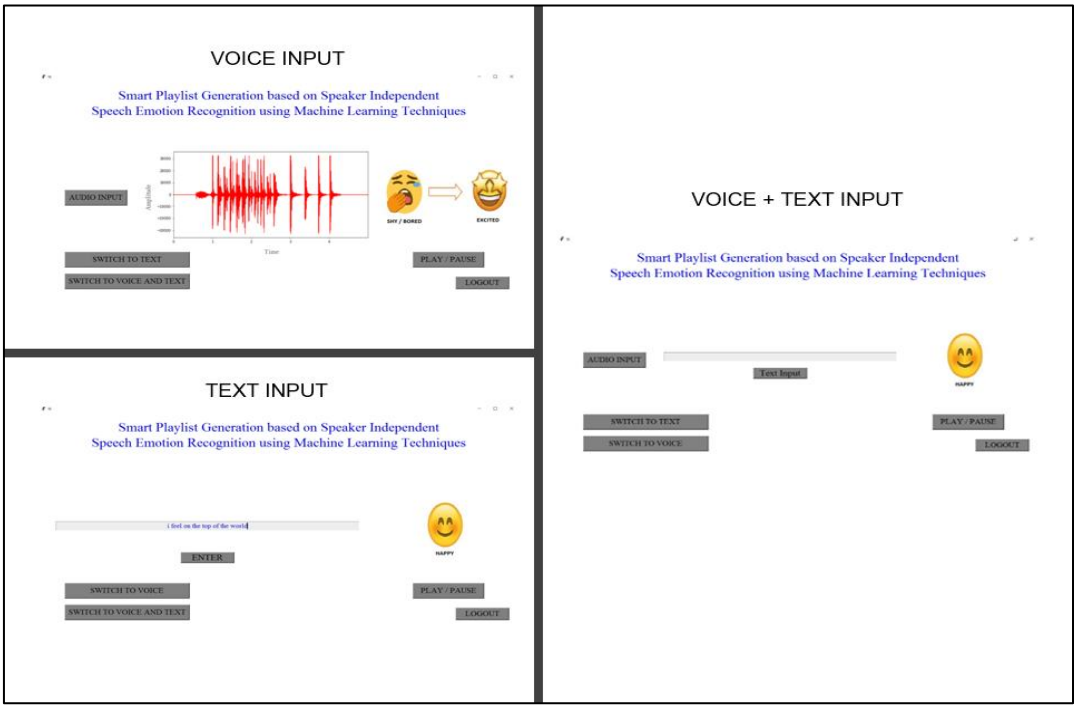


Figure 9. Input mode page.

6. Results and discussion

In this section the implementation of a novel speech and text emotion recognition based music player for mood uplift is discussed.

Music player accuracy

The music player with voice input shows overall accuracy of average 94.56%. The voice input is based on PNN classifier. The Figure. 10 shows the confusion matrix for speech; text; speech + text emotion recognition. The music player with text input shows overall accuracy of average 91.28%. The text input is based on SVM classifier. In speech + text method, above both approaches of speech and text emotion recognition are combined. It shows improved accuracy of average 95.76%.

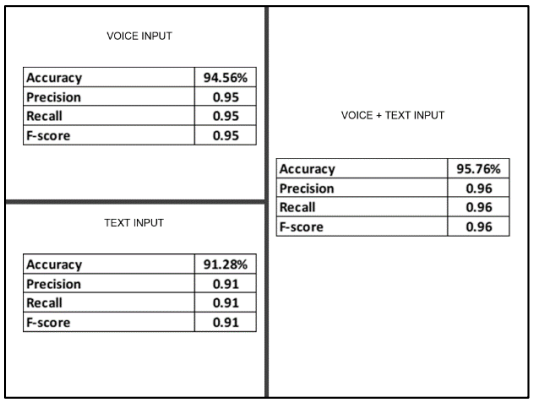


Figure 10. Confusion matrix.

The Fig. 11 highlights the emotion based accuracy of the music player.

with voice input - The accuracy for happy emotion is 96%, for sad emotion is 92%, for angry emotion is 94%, for fear emotion is 95%, and bored emotion is 96%.

with text input - The accuracy for happy emotion is 94%, for sad emotion is 89%, for angry emotion is 91%, for fear emotion is 90%, for bored emotion is 93%.

with voice + text input - The accuracy for happy emotion is 96%, for sad emotion is 95%, for angry emotion is 95%, for fear emotion is 95%, for bored emotion is 97%.

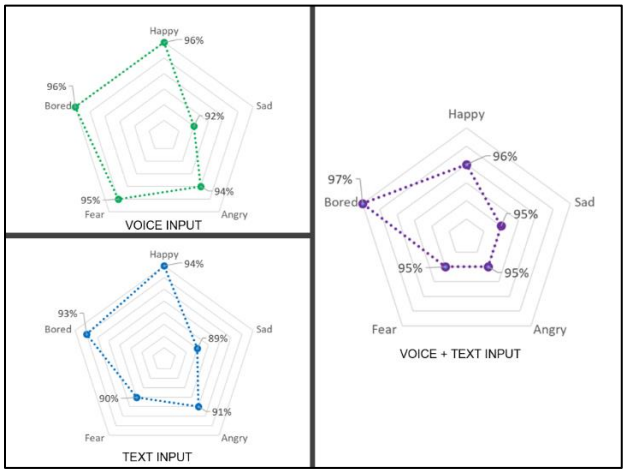


Figure 11. Emotion based accuracy.

Feedback of mood upliftment

The Figure. 12 shows the feedback from 250 participants whether their mood uplifted after listening the song. 84% of the participants reported they felt better by listening the song.

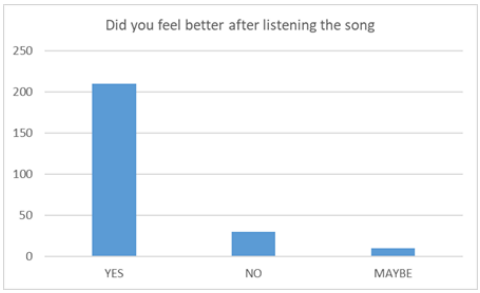


Figure 12. Mood upliftment feedback.

Comparison with existing system

The Figure. 13 shows the feedback from 250 participants whether their mood uplifted after listening the song. 84% of the participants reported they felt better by listening the song.

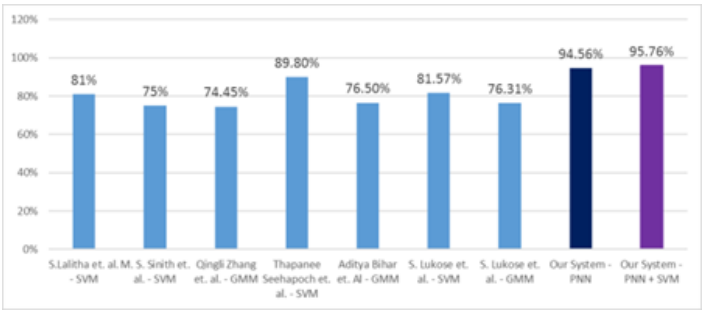


Figure 13. Comparison of proposed system with existing systems.

Uniqueness of proposed work

The results show remarkable classification accuracy for the selected classes of emotions. The proposed system helps in uplifting the user's mood. The system provides multiple input options such as – voice, text and voice + text; thereby even assisting people who are unable to speak. A database of Hindi songs consisting of 100 songs per emotion that are selected is now available at request.

7. Conclusion and future scope

In this paper, an implementation of a novel speech and text emotion recognition based music player for mood upliftment is presented. The output of the system is a song played based on the emotion predicted from either speech or text or both. With the input as voice PNN classifier is employed for emotion prediction. The overall accuracy reported in this case is around 94.56%. With the input as text SVM classifier is employed for emotion prediction. The overall accuracy reported in this case is around 91.28%. With the input as voice and text both approaches are combined for emotion prediction. The overall accuracy reported in this case is around 95.76%. Mood uplifting technique is used for enhancing mood. A person's mood is changed from sad to happy, bored to excited, angry to calm and fear to strong.

For future work, the segregated songs can be exported to a cloud database which then can be made available for users to download based on request. Also, the proposed system can be further developed to predict more emotions..

References

[1] Babak Basharirad, and Mohammadreza Moradhaseli (2017). Speech emotion recognition methods: A literature review. In AIP Conference Proceedings, vol. 1891, no. 1, p. 020105.

[2] Tamulevičius Gintautas, Gražina Korvel, Anil Bora Yayak, Povilas Treigys, Jolita Bernatavičienė, and Bożena Kostek (2020). A Study of Cross-Linguistic Speech Emotion Recognition Based on 2D Feature Spaces. Electronics 2020, 9, no. 10: 1725.

[3] Deshmukh Shrikala Madhav, and Devulapalli Sita (2020). Mood Enhancing Music Player Based on Speech Emotion Recognition and Text Emotion Recognition. International Journal of Emerging Trends in Engineering Research. Vol. 8, No. 6.

[4] Kabani H., Khan S., Khan O. and Tadvi S. (2015). Emotion based music player. International Journal of Engineering Research and General Science, 3(1), pp.2091-2730.

[5] Hemanth P, Adarsh, Aswani C.B, Ajith P, Veena A Kumar (2018), EMO PLAYER: Emotion Based Music Player. International Research Journal of Engineering and Technology (IRJET). Volume: 05 Issue: 04, pp-4822-4827.

[6] Ghule V.R., Benke A.B., Jadhav S.S. and Joshi S.A. (2017). Emotion Based Music Player Using Facial Recognition. International Journal of Innovative Research in Computer and Communication Engineering, 5(2), pp 2188-2194.

[7] Nathan Karthik Subramanian, Manasi Arun, and Megala S. Kannan (2017). "EMOSIC—An emotion based music player for Android." In 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 371-276.

[8] Vijay Chakole, Aniket Choudhary, Kalyani Trivedi, Kshitija Bhoyar, Ruchita Bodele, Sayali Karmore (2018). Emotion Based Music Player. International Journal of Engineering Science and Computing, Volume 8 Issue No.3, pp-16322-16324.

[9] Mistry Karan, Prof Prince Pathak, and Suvarna Aranjo (2017). Mood based Music Player. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 3, pp 779-781.

[10] Prof. Jayshree Jha, Akshay Mangaonkar, Deep Mistry, Nipun Jambaulikar, Prathamesh Kolhatkar (2015). Facial Expression Based Music Player. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 10, pp 331-334.

[11] Rohit Kadam, Darshan Gosavi, Chaitanya Sankhe, Amruta Mhatre (2018). Effective Music Player Integrated On User's Mood. IOSR Journal of Engineering (IOSRJEN), Volume 13, pp 63-69.

[12] A. S. Mali, A. A. Kenjale, P. M. Ghatage, A. G. Deshpande (2018). Mood based Music System, International Journal of Scientific Research in Research Paper Computer Science and Engineering, Vol.6, Issue.3, pp.27-30.

[13] Specht, Donald F (1990). Probabilistic neural networks. Neural networks 3, no. 1: 109-118.

[14] Mao Ke Zhi, K-C. Tan, and Wee Ser (2000). Probabilistic neural-network structure determination for pattern classification. IEEE Transactions on neural networks, Vol. 11, no. 4: 1009-1016.

[15] Calefato F., Lanubile F., Novielli N. (2017). EmoTxt: a toolkit for emotion recognition from text. In 2017 seventh international conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), IEEE, pp. 79-80.

[16] S. N. Shivhare, S. Khethawat (2012). Emotion Detection From Text, CS & IT-CSCP, pp. 371–377, 2012. https://doi.org/10.5121/csit.2012.2237

[17] C.R. Chopade (2015). Text based emotion recognition: A survey. International Journal of Science and Research, vol. 4, no. 6, pp. 409-414.

[18] Soumya George K., and Shibily Joseph (2014). Text classification by augmenting bag of words (BOW) representation with co-occurrence feature. IOSR Journal of Computer Engineering, 16, no. 1: 34-38.

[19] Batoul Aljaddouh, Nishith A. Kotak (2020). Document Text Classification Using Support Vector Machine. International Journal of Engineering Development and Research, ISSN: 2321-9939, Volume 8, Issue 1, pp 138-142.

[20] Sarkar Anurag, Saptarshi Chatterjee, Writayan Das, and Debabrata Datta (2015). Text classification using support vector machine. International Journal of Engineering Science Invention, Vol 4, no. 11: 33-37.

[21] Basu Atreya, Christine Watters, and M. Shepherd (2003). Support vector machines for text categorization. In 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, pp. 7.

[22] Joachims Thorsten (1998). Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning, Springer, Berlin, Heidelberg, pp. 137-142.

[23] Watson David, Lee Anna Clark, and Auke Tellegen (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology 54, no. 6, pp 1063.

[24] Crawford John R., and Julie D. Henry (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. British journal of clinical psychology 43, no. 3, pp 245-265.