ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022

Emotion Recognition from Non-Native Marathi Speech using MFCC and LPC Techniques

Bharati D.Borade* and Ratnadeep R. Deshmukh

Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad(MS), India

Abstract- Automatic emotion detection from human speech is becoming more prevalent today because it improves interactions between humans and machines. Human speech can be used to extract a range of temporal and spectral properties. Pitch-related characteristics, Mel Frequency Cepstral Coefficients (MFCCs), and speech formants can all be categorised using different techniques. This study examines statistical characteristics, such as Linear Discriminant Analysis was used to classify these features and MFCCs (LDA). A database of artificially emotional Marathi speech is also described in this article. The data samples were taken from male and female Marathi speeches that mimicked the emotions that gave rise to the Marathi utterances that could be employed in ordinary communication & are interpreted in all considered emotions. Three fundamental categories—happy, sad, and angry—were used to classify data samples. The training accuracy and testing accuracy for MFCC and LPC are 98, 82 and 85,82 respectively.

Keywords : Non- Native Speech Database; Emotional Speech database; MFCC; LPC; emotion recognition;

Automatic emotion detection from human speech is becoming more prevalent today because it improves interactions between humans and machines. About 71 million people, mostly in the Indian state of Maharashtra and its bordering states, speak Marathi, an Indo-Aryan tongue. Israel and Mauritius both have native speakers of Marathi. It is believed that Maharashtri, one of the Prakrit languages that emerged from Sanskrit, is related to Marathi. Inscriptions on copper plates and stones from the 11th century were the first written records of Marathi. It was written using the Modi alphabet from the 13th century till the middle of the 20th century. It has been written using the Devanagari alphabet since 1950.

The Marathi language has 36 consonants and 13 vowels. Emotional intelligence is the capacity for emotion recognition, interpretation, and expression. In human computer interfaces, emotions are recognised and expressed [1]. When someone speaks to a computer, voice recognition basically recognises what they are saying and asks the computer to convert that speech into a text message. In contrast, speech synthesis uses a computer to create false spoken dialogue. The most common and organic method of human communication is speech. There is considerable interest in creating machines that can accept voice as input since speech would be a reasonable choice for human-machine communication. Given the significant global research efforts in speech recognition and the steady rate of advancement of computer speed and size[2- 6].

The interface for a machine that receives speech input typically consists of two stages. An automatic recognition system (ASR) is needed for the first phase, and a speech understanding system is needed for the second. Humans can communicate their emotions in a variety of ways. Humans can cry, yell, dance, laugh, stomp, and do a variety of other activities to show their emotions[8]. However, when it comes to communication, human emotions have an impact on a person's tone and speaking manner[22,24]. Spoken recognition accuracy is impacted by the emotion in the speech sound. Researchers from all across the world are interested in learning how to identify emotions in speech. Many academics are exploring the breadth of the field of emotion detection from speech in human-computer interaction. The study of voice emotion recognition has drawn a lot of interest in recent years. Numerous emotional speech databases have been created, and researchers from all over the world have conducted studies using these datasets[26].

LITERATURE SURVEY

The matrix datasets were created into different three classes form features of angry happy and sad dialogs of all the samples for training purpose. Feature of the different emotion subjected to Linear Discriminant Analysis (LDA). Aim of LDA is used to reduce dimensions of feature matrix and to clusters data representing the different classes. For the classification and clustering purpose[6], they have created appropriate different classes of dataset. The three separate classes of datasets have been created according to three different emotions (angry, happy and sad) of all samples.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022

The Linear Discriminant (Fishers Algorithm) has been implemented on class-within class matrix dataset. Results of the LDA are made three groups of all emotions.

After the Marathi language's emotive speech data banks were developed, the experiment was conducted. For the aim of feature extraction, Linear Predictive Coding (LPC) was utilised. LPC characteristics include specific emotional information. The recursion method is used to convert an extremely crucial set of LPC parameters, known as the LPC Cepstral coefficients, it was obtained directly from the set of LPC coefficients. It has been demonstrated that the Cepstral coefficients, which are the log magnitude spectrum's Fourier transform coefficients, are a more durable and trustworthy feature set for speech recognition than LPC coefficients[10].

A suggested framework that integrates Principal Component Analysis[19,20] and five common emotion categories—happiness, sadness, anger, fear, and surprise—recognized the mantic information of these emotions (PCA). The goal of the principle component analysis experiment is to produce a limited number of linear combinations, or principal components, of a set of variables while preserving as much of the original variables' information as feasible[22] We have chosen to use the PCA approach to reduce the number of feature parameters and get rid of duplication thanks to the general The prosodic qualities pertain to musical aspects of speech. The key components for speech emotion prosody are the fundamental frequency, duration, and energy qualities. Information concerning intonation, accent, and rhythm is conveyed via prosodic qualities. The intensity contours and a linear approximation of F0 serve as the foundation for the prosodic features. Articulatory features must be retrieved in order to identify the speaker characteristics related to pronunciation. The neural network was trained to accomplish this. By manually extracting the lexical features using the speech corpus as a training set[11]. For the purpose of classifying various emotional states, a neural network was employed. Since neural networks are fundamentally parallel, a neural network makes use of link.

strengths and functions. Speech frequencies happen in parallel, whereas word and syllable sequences are essentially serial. As a result, neural network techniques are extremely potent in other contexts[12].

1. PROPOSED METHODOLOGY

The analysis demonstrates the various methods used to extract emotions from speech for the created emotional speech database. The used strategies are then subjected to benchmark tests in order to identify which method is best for recognising emotions. It also provides results for the application of these techniques to voice emotion recognition. The basic block diagram for emotion recognition from speech signal (Figure 1).

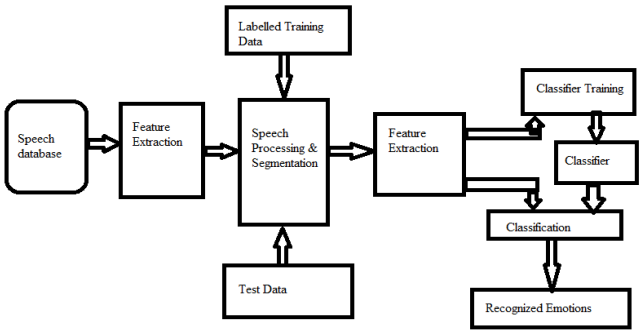


Figure 1 Speech Emotion Recognition

The Native and Non- native speaker audio files are used to do feature extraction.

a. Mel Frequency Cepstral Coefficient (MFCC)

The MFCC is implemented in several ways. These methods differ primarily in terms of the quantity of filters used, their design, their spacing, their bandwidth, and how the spectrum is distorted. The number of filters included in Filter bank (FB) for the MFCC by associated author is defined by the FB in the implementations. These implementations take various sample rates into account. Figure 2 depicts the the steps that are taken to compute the features using MFCC[2].

The frequency scales in MFCC are laid-out on logarithmic scale for frequencies above 1KHz and a linear scale for frequencies below 1KHz. MFCCs are the best for involuntary speech recognition & involuntary emotion recognition because they contain both temporal and frequency information of signal[8]. The logarithm of short-term energy spectrum expressed on Mel-frequency scale underwent a cosine transform to produce MFCCs. The MFCC describes the fast frequency domain energy migration. In MFCC, a signal's DFT spectrum is frequently warped using the equation for the Mel-frequency scale transformation.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022

Figure 2 - MFCC Feature Extraction steps

$$\text{Mel}(f)=2595\log_{10}(1+f/700) \quad (1)$$

The first order regression coefficients (delta coefficients) are computed by the following regression equation:

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2} \quad (2)$$

The values initialised during computation of MFCC are as follows:
Sampling frequency: 22000 ;
Window type: Hamming Window;
Window length: 25 millisecond Step ;
time: 10 millisecond;
Number of coefficient :13 (1 Energy and 12 standard coefficient) ;
Min Frequency: 0 (lowest band edge Mel filters (Hz));
Max Frequency: 4000 (The highest band edge of Mel filters (Hz) set); FFT: 512 point FFT;

b. Linear Predictive Coding (LPC)

A specific set of predictor coefficients is discovered by minimising sum of squared differences between real speech samples & linearly predicted ones over a finite period. As a result, a speech sample can be roughly represented as a linear combination of preceding speech samples. The output of a linear, time-varying system that, depending on the situation, is either driven by a quasi-periodic pulse for voiced speech or by random noise for unvoiced speech is how speech is simulated. For predicting the parameters that define the linear time-varying system representing the vocal tract, which is useful for emotion recognition, the linear prediction approach offers a strong, dependable, and accurate method. After the creation of emotional speech databases in the Marathi language, the experiment was conducted. For the feature extraction in this work, we used linear predictive coding (LPC). Features of Linear Predictive Coding (LPC) transmit specific emotional information.

The strongest, most powerful frequencies form the basis of how humans perceive speech. As a result, the vocal tract is frequently defined in terms of its formants, or resonant frequencies. The poles of vocal tract transfer function are what create these resonances. These formants are denoted by the symbol Fi, where I denotes the formant number, such as F_1,F_2,...,Fn. In the 0 to 4000Hz range of human speech, there are typically four formants present. In order to improve the performance of the system utilising solely LPC features, it was discovered that LPC is quite good for speech emotion recognition. Figure 3 depicts an overall technique for recognising emotions.

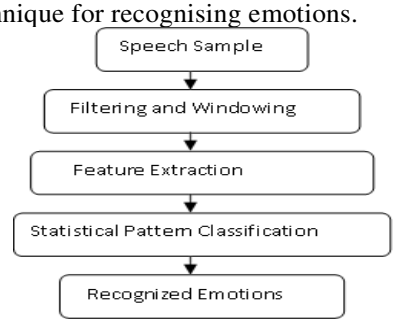


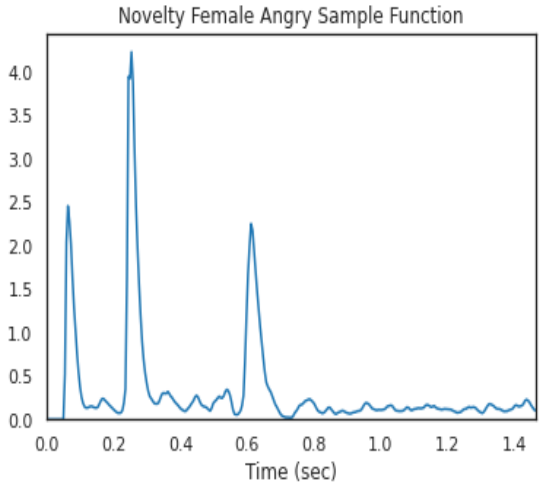
Figure 3-Steps in emotion recognition using LPC

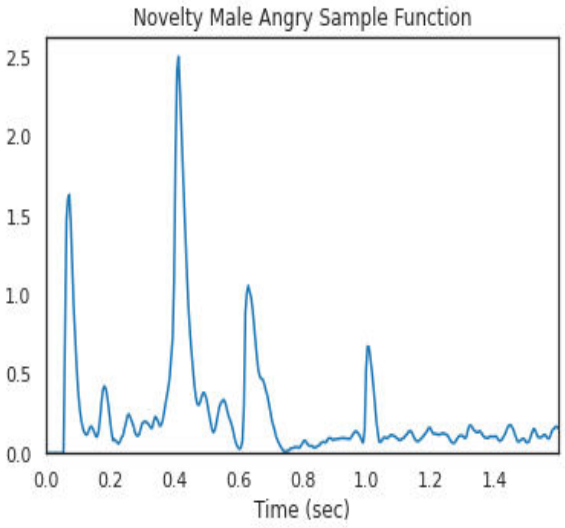
C. Emotions dataset-

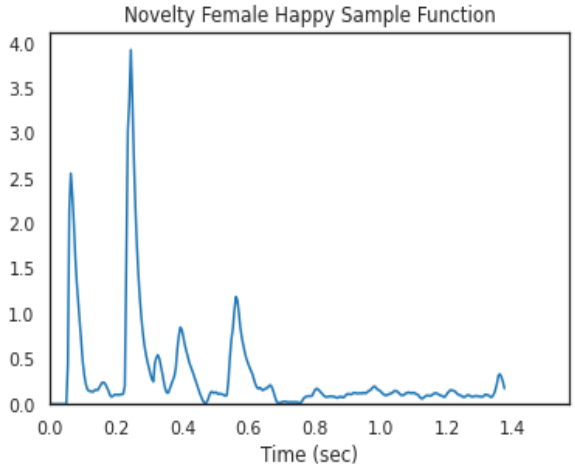
The different functions of emotions are shown below-

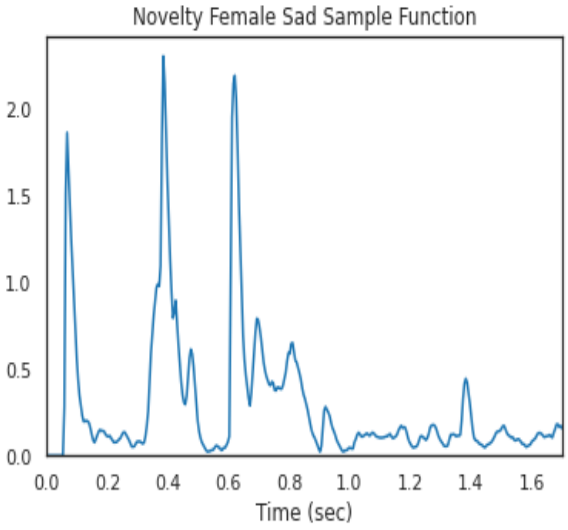
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022



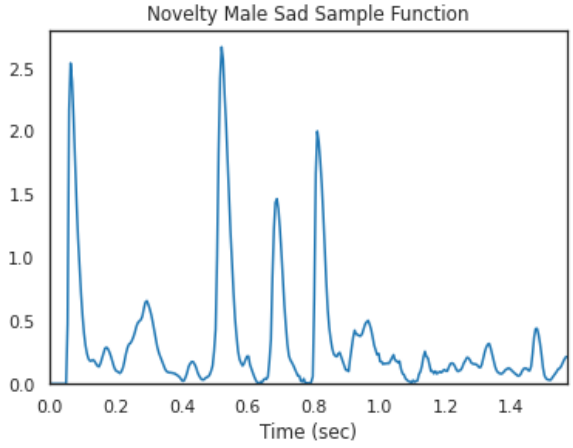


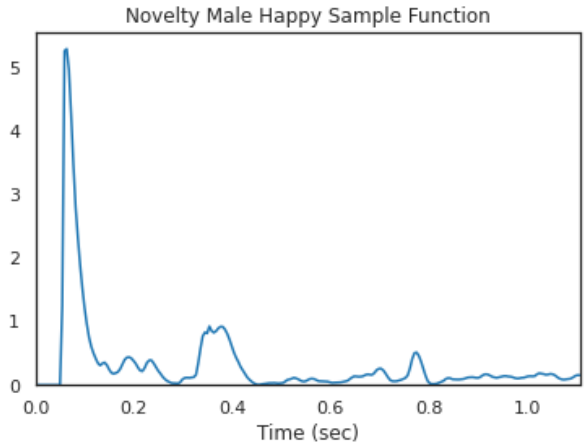


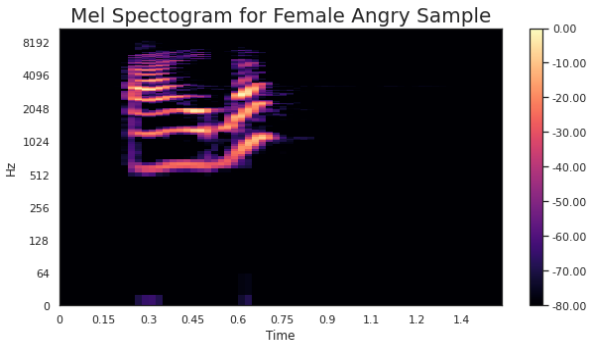


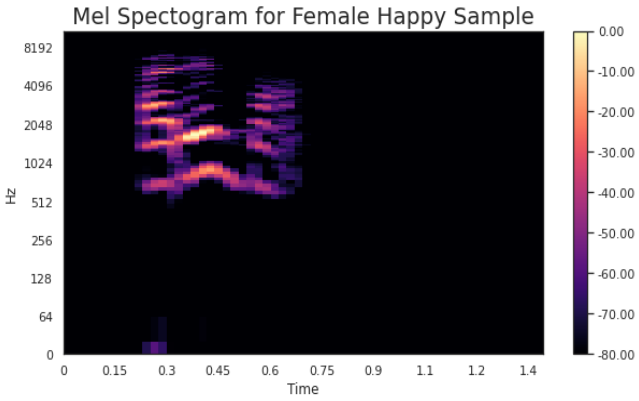
ISSN PRINT 2319 1775 Online 2320 7876

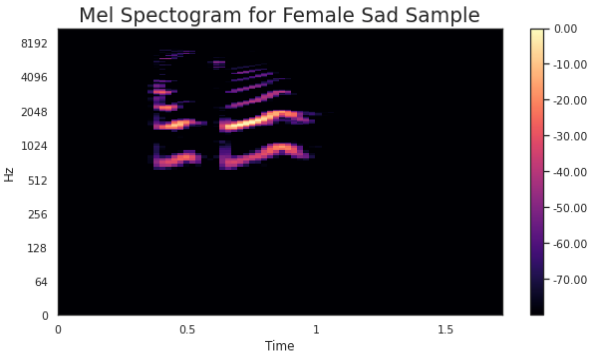
Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022









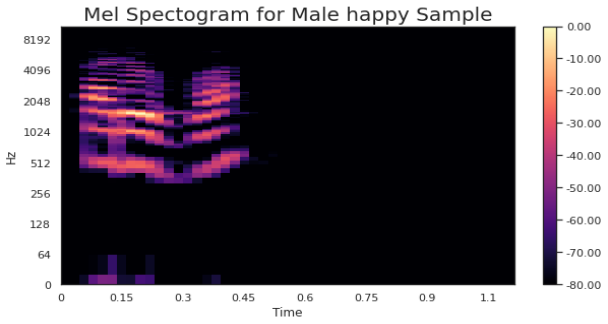


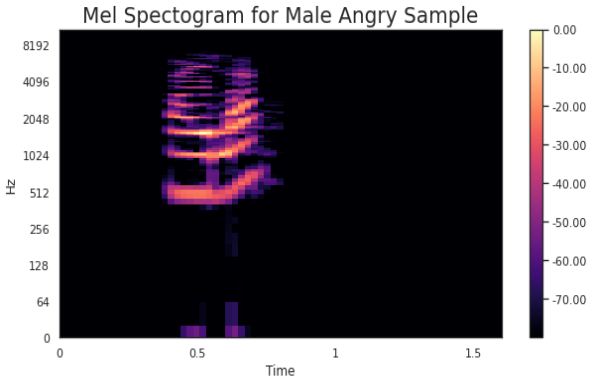
RESULTS AND DISCUSSIONS

In the study, we used MFCCs and LPC to assess the corpus of emotional speech. On our own built database, we ran an experiment with male and female participants using the Marathi words for sad, happy, and angry. The data samples were gradually trained and tested. Figure 5 displays the Mel Spectrum for various audios representing the aforementioned emotions.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022





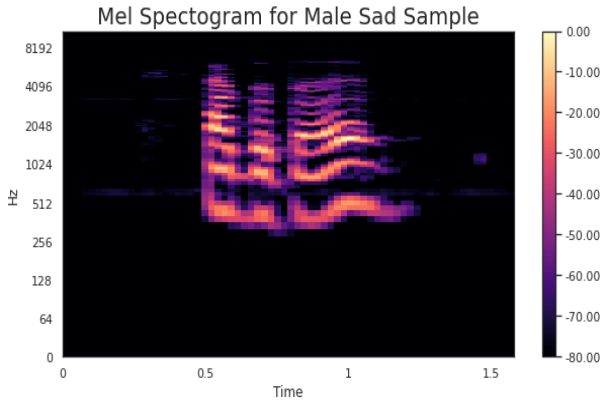


Figure 6 shows the confusion matrix for MFCC and figure 7 shows the MFCC features. According to the observation, pointed diagonal values ought to be less than non-diagonal values but the distance is much higher than non diagonal member hence this technique is not suitable for classification so we have used confusion matrix for the classification purpose.

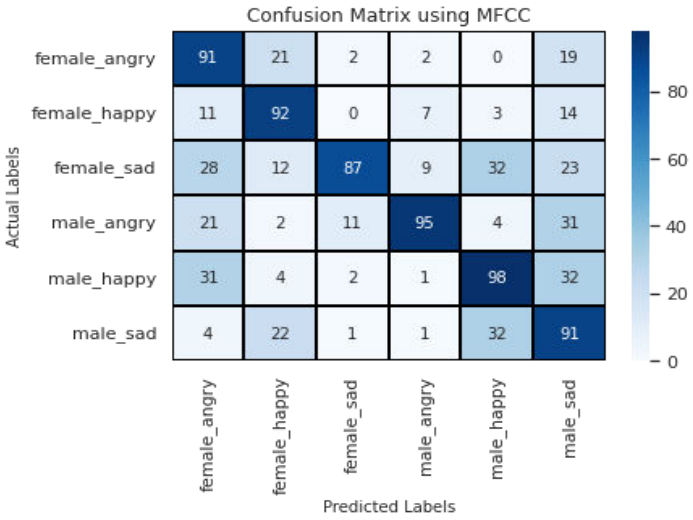
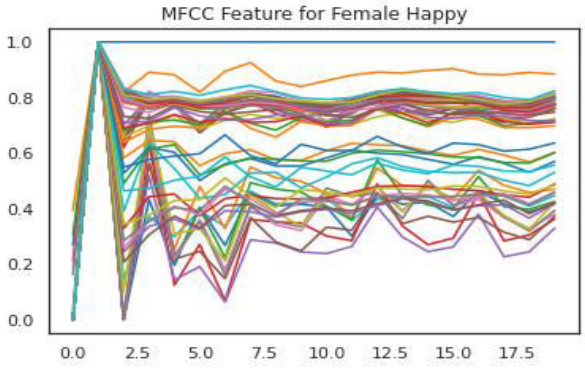
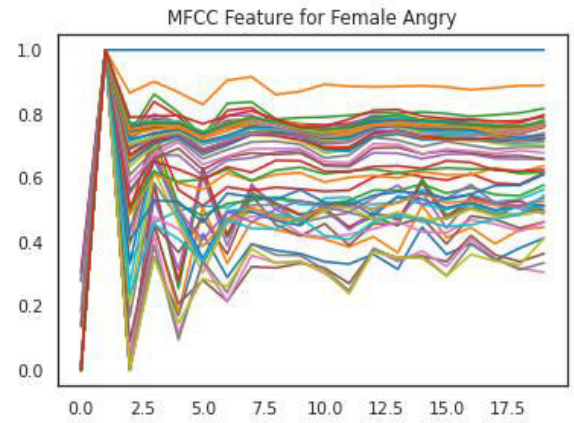


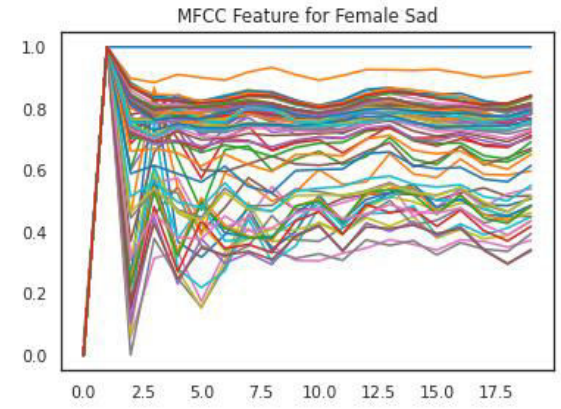
Figure 6- Confusion Matrix for MFCC

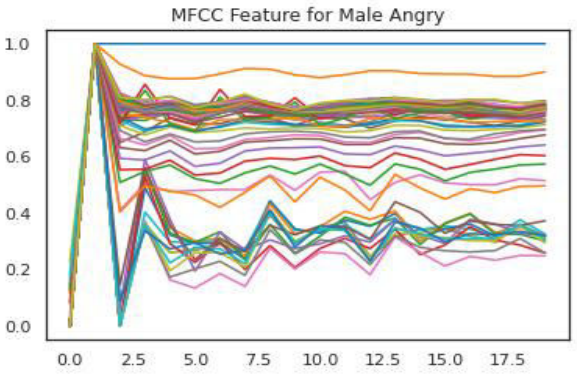


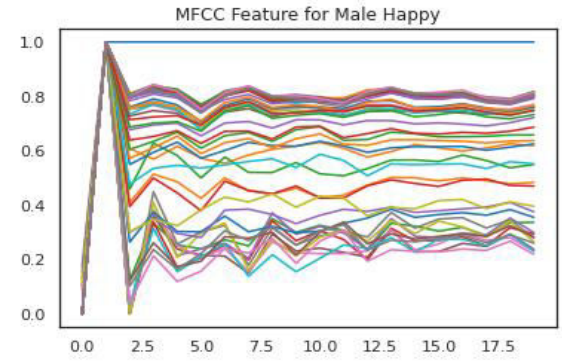


ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022







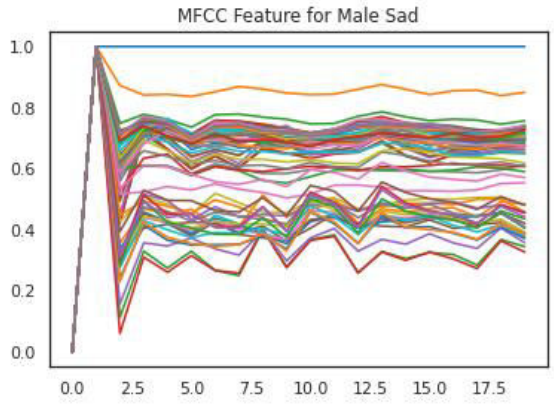


Figure 7 - MFCC Features

```
train_mfcc.describe()
```

	y
count	498.000000
mean	2.654618
std	1.703703
min	0.000000
25%	1.000000
50%	3.000000
75%	4.000000
max	5.000000

Figure 8 - Feature stats of Training MFCC

The training of dataset by using MFCC has feature set as shown in figure 8.

```
Best parameters: {'C': 1, 'degree': 1, 'gamma': 0.1, 'kernel': 'poly'}
MFCC:
Train accuracy: 98.0%
Test accuracy:  82.0%
```

Figure 9- Accuracy with MFCC

The confusion varied with the help of LPC features is shown in figure 10, 11 shows the feature stats of LPC training.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022

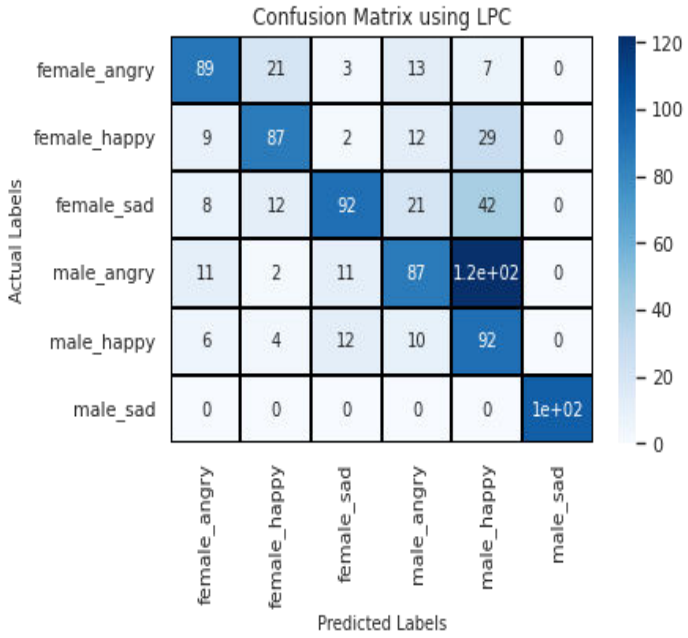
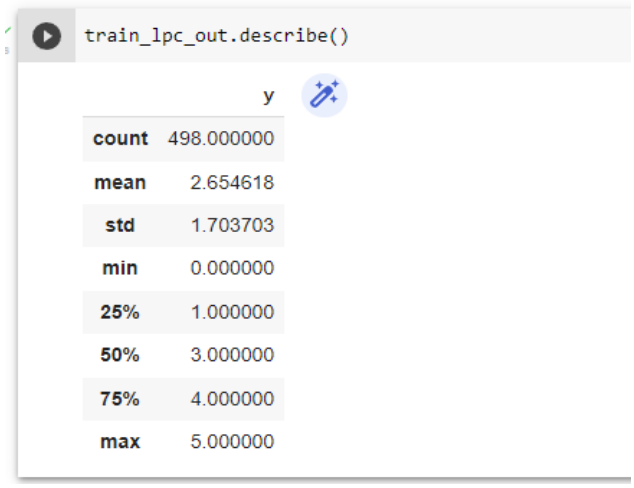


Figure 10- Confusion Matrix LPC



```
train_lpc_out.describe()
```

	y
count	498.000000
mean	2.654618
std	1.703703
min	0.000000
25%	1.000000
50%	3.000000
75%	4.000000
max	5.000000

Figure 11- Feature stats of Training of LPC

```
Best parameters: {'C': 1, 'degree': 1, 'gamma': 0.1, 'kernel': 'poly'}
LPC:
Train accuracy: 85.0%
Test accuracy:  82.0%
```

Figure 12 - Accuracy with LPC

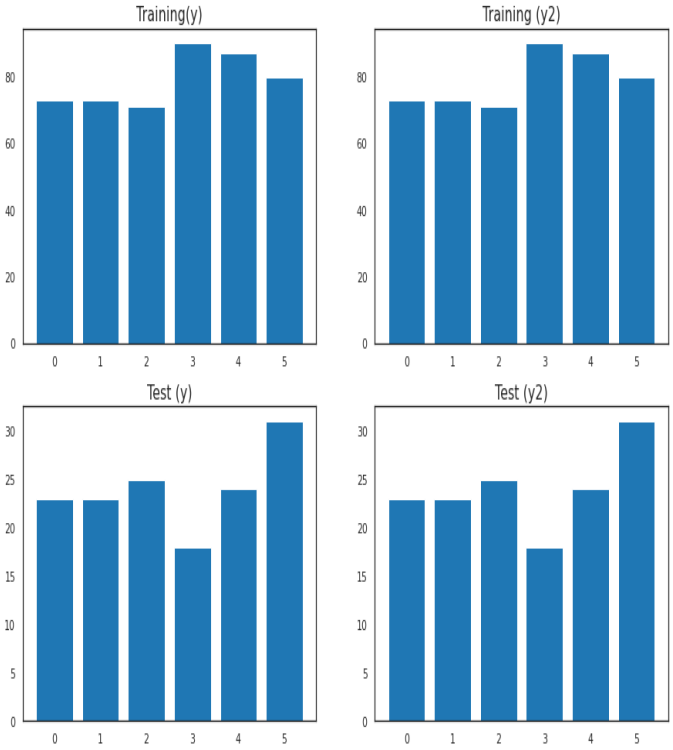


Figure 13- Training and Testing Features

CONCLUSION

Automatic emotion detection from human speech is becoming more prevalent today because it improves interactions between humans and machines. In this paper, we used a few key emotion recognition methods based on feature extraction strategies and classifiers. Also observed that the proper feature extraction and classifier selection are essential to the accuracy and effectiveness of an emotion identification system. In this study, we used Mel Frequency Cepstral Coefficients to assess the artificial emotional speech corpus (MFCCs). We tested the hypothesis using quotes from Marathi movie actors and actresses from our own, in-house

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8, Dec 2022

developed database. The data samples were gradually trained and tested. It explains how to recognise emotions using MFCCs and LPC and database of emotional speech. The training accuracy and testing accuracy for MFCC and LPC are 98, 82 and 85,82 respectively as shown in figure 9,12 above.

Acknowledgment

Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India is supported by the Department of Science and Technology under the Funds for Infrastructure under Science and Technology (DST-FIST) with sanction no. SR/FST/ETI-340/2013.The authors would like to express their gratitude to the Department and University Authorities for providing the required infrastructure and support for the research.Thank you very much to SARATHI for financing my M.Phil dissertation.

REFERENCES

1. Pukhraj P. Shrishrimal, “Design and Development of Spoken Marathi Isolated Words Database for Agriculture Purpose and its Analysis”, M.Phil. Computer ScienceThesis. May 2013.
2. Vibha Tiwari, "MFCC and its application in speaker recognition", International Journal on Emerging Technologies, 2010, Vol.1,No.1,pp.19-22.
3. Davis S. B., Mermelstein P., “Comparison of Parametric Representations for Mono syllabic Word Recognition in Continuously Spoken Sentences”, IEEE Transaction on Acoustic, Speech and Signal Processing, 1980, Vol. 28, No. 4, pp. 357-366.
4. Young S. J., Odell J., Ollason D., Valtchev V., Woodland P., “The HTK Book.Version2.1”,DepartmentofEngineering,CambridgeUniversity,UK,1995.
5. “TheNISTYear2001SpeakerRecognitionEvaluationPlan”,TheNISTofUSA,2001.Available: http://www.nist.gov/speech/tests/spk/2001/doc/2001-spkrec-evalplan-v05.9.pdf.
6. Skowronski M. D., Harris J. G., “Exploiting independent filter bandwidth ofhumanfactorcepstralcoefficientsinautomaticspeechrecognition”,JournaloftheAcoustical Society of America,2004, Vol.116,No.3,pp.1774–1780.
7. Xia Mao, Lijiang Chen, Bing Zhang, “Mandarin speech emotion recognition based on a hybrid of HMM/ANN”, International Journal of Computers, Vol 1, Issue 4.
8. Yu Zhou, Yanqing Sun, Lin Yang, Yonghong Yan, "Applying articulatory features to speech emotion recognition"m ThinkIT Speech Lab., Institute of Acoustics, Chinese Academy of Sciences, Beijing, IEEE International Conference on Research Challenges in Computer Science,2009
9. Vishal Waghmare, Ratnadeep Deshmukh, Pukhraj Shrishrimal, Ganesh Janvale,“Emotion Recognition System from Artificial Marathi Speech using MFCC and LDA Techniques” Proc. Of Int. Conf. on Advances in Communication, Network, and Computing, CNC-2014, Chennai 2014. Pp 408-416
10. Xia Mao, Lijiang Chen, Bing Zhang, “Mandarin speech emotion recognition based on a hybrid of HMM/ANN”, International Journal of Computers, 2007, Vol 1, Issue4.
11. Szymon Drgas, Adam Dabrowski, "Speaker Recognition Based on Multi level Speech Signal Analysis on Polish Corpus", 2012, CCIS 287, pp. 85–94
12. Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "Neural Networks used for Speech Recognition", Journal Of Automatic Control, University Of Belgrade, 2010, Vol 2, Issue 1
13. Mao, Xia, Lijiang Chen, L. Fu. "Mandarin speech emotion recognition based on a hybrid of HMM/ANN." International Journal of Computers, 2007, Vol 1, Issue 4, pp. 321-324.
14. Vemula Yakub Reddy1, Mangipudi Pavan Kumar2, Mankala Sushma,Gurindagunta Kiran4,Vijaya Kumar Gurrala , “SPEECH BASED EMOTION DETECTION SYSTEM USING MFCC”, International Research Journal of Engineering and Technology (IRJET), 2020, vol 7, Issue 5, pp. 4329-4332.
15. Somi Kolita, “ Speech Emotion Recognition using Non-linear Classifier-A Review”, International Journal of Engineering Research & Technology (IJERT), 2019, vol 8, Issue 5, pp. 207-210
16. Ms. Machha Babitha, C Sushma, et al, “Trends of Artificial Intelligence for online exams in education”, *International journal of Early Childhood special Education*, 2022, Vol 14, Issue 01, pp. 2457-2463.
17. Dr. J. Sirisha Devi, Mr. B. Sreedhar, et al, “A path towards child-centric Artificial Intelligence based Education”, *International journal of Early Childhood special Education*, 2022, Vol 14, Issue 03, pp. 9915-9922.
18. Mr. D. Sreenivasulu, Dr. J. Sirishadevi, et al, “Implementation of Latest machine learning approaches for students Grade Prediction”, *International journal of Early Childhood special Education*, June 2022, Vol 14, Issue 03, pp. 9887-9894.
19. Kazi K. S., Shirgan S S, “ Face Recognition based on Principal Component Analysis and Feed Forward Neural Network”, *National Conference on Emerging trends in Engineering, Technology, Architecture*, Dec 2010, pp. 250-253.
20. Dr. Kazi Kutubuddin, V A Mane, Dr K P Pardeshi, Dr. D.B Kadam, Dr. Pandyaji K K, “Development of Pose invariant Face Recognition method based on PCA and Artificial Neural Network”, *Journal of Algebraic Statistics*, 2022, Vol 13, issue 3, pp. 3676-3684.
21. Ravi Aavula, Amar Deshmukh, V A Mane, et al, “Design and Implementation of sensor and IoT based Remembrance system for closed one”, *Telematique*, 2022, Vol 21, Issue 1, pp. 2769- 2778.
22. Salunke Nikita, et al, “Announcement system in Bus”, Journal of Image Processing and Intelligent remote sensing, 2022, Vol 2, issue 6
23. Satpute Pratiskha Vaijnath, Mali Prajakta et al. “Smart safty Device for Women”, International Journal of Aquatic Science, 2022, Vol 13, Issue 1, pp. 556- 560
24. Miss. Priyanka M Tadlgi, et al, “Depression Detection”, Journal of Mental Health Issues and Behavior (JHMIB), 2022, Vol 2, Issue 6, pp. 1-7
25. Waghmare Maithili, et al, “Smart watch system”, International journal of information Technology and computer engineering (IJITC), 2022, Vol 2, issue 6, pp. 1- 9.
26. Divya Swami, et al, “Sending notification to someone missing you through smart watch”, International journal of information Technology and computer engineering (IJITC), 2022, Vol 2, issue 8, pp. 19-24