*Proc. of Int. Conf. on Advances in Communication, Network, and Computing, CNC*

ELSEVIER

# Emotion Recognition System from Artificial Marathi Speech using MFCC and LDA Techniques

Vishal B. Waghmare, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal[1], Ganesh B. Janvale[2]
[1] Dr. Babasaheb Ambedkar Marathwada University, Department of Computer Science & IT, Aurangabad, India (M.S.)
Email: vishal_pri1@yahoo.co.in, ratnadeep_deshmukh@yahoo.co.in, pukhraj.shrishrimal@gmail.com
[2] Symbiosis Centre Information Technology, Symbiosis International University, Pune, India (M.S.)
Email: ganesh@scit.edu

***Abstract*— There are a variety of temporal and spectral features that can be extracted from human speech. These features are related to the pitch, Mel Frequency Cepstral Coefficients (MFCCs) and Formants of speech, can be classified using various algorithms. This study explores statistical features i.e. MFCCs and these features were classified with the help of Linear Discriminant Analaysis (LDA). This article also describes a database of artificial emotional Marathi speech. The data samples were collected from 5 Marathi movies (Actors and Actress) simulated the emotions producing the Marathi utterances which could be used in everyday communication and are interpretable in all applied emotions. The speech samples were distinguished by the various situations from the movie. The data samples were categorized in 5 basic categories that are Happy, Sad, Anger, Afraid and Surprise.**

***Index Terms*— Speech Database, Emotion, Speech, Emotional Speech database, MFCC, emotion recognition**

## I. INTRODUCTION

Marathi is an Indo-Aryan language spoken by about 71 million people mainly in the Indian state of Maharashtra and neighboring states. Marathi is also spoken in Israel and Mauritius. Marathi is thought to be a descendent of Maharashtri, one of the Prakrit languages which developed from Sanskrit. Marathi first appeared in writing during the 11$^{th}$ century in the form of inscriptions on stones and copper plates. From the 13$^{th}$ century until the mid 20$^{th}$ century, it was written with the Modi alphabet. Since 1950 it has been written with the Devanagari alphabet. There are 13 vowels and 36 consonants in Marathi language. The vowels and consonants along with their transliteration and International Phonetics Alphabetics are shown in figure 1.

The skill to recognize, interpret and express the emotion referred to emotional intelligent. The emotions recognition and expression are used in human computer interfacing [1]. The speech recognition understands basically what someone speak to a computer, asking a computer to translate speech into its corresponding textual message, where as in speech synthesis, a computer generate artificial spoken dialogs. Speech is the most prominent and natural form of communication between the humans. Speech would thus be a logical choice for man machine communication; hence there is growing interest in developing such machines that can accept speech as input. Given the substantial research efforts in speech recognition worldwide and the steady rate at which computers become faster and smaller [2, 3, 4, 5, 6]. The machine that accepts speech as

| अ | आ | इ | ई | उ | ऊ | ऋ | ए | ऐ | ओ | औ |
|---|---|---|---|---|---|---|---|---|---|---|
| A | Āa | i | ī | U | Ū | ṛ | e | Ai | O | Au |
| /ə/ | /a/ | /i/ | | /u/ | | /ru/ | /e/ | /əi/ | /o/ | /əu/ |
| अं | | | | | | अः | | | | |
| aṃ | | | | | | aḥ | | | | |
| /əⁿ/ | | | | | | /əh/ | | | | |

Figure 1 a . Vowels in Marathi language along with transliteration and IPA

| क | KA | /KƏ/ |
|---|---|---|
| ख | KHA | /K$^{h}$Ə/ |
| ग | GA | /gə/ |
| घ | GHA | /g$^{ɦ}$Ə/ |
| ङ | ṄA | /ŋƏ/ |

| च | Ca | /tsə/ |
|---|---|---|
| छ | Cha | /ts$^{h}$ə/ |
| ज | Ja | /zə/ |
| झ | Jha | /z$^{ɦ}$ə/ |
| ञ | ṅa | /ŋə/ |

| ट | ṭa | /ʈə/ |
|---|---|---|
| ठ | ṭha | /ʈ$^{h}$ə/ |
| ड | ḍa | /ɖə/ |
| ढ | ḍha | /ɖ$^{ɦ}$ə/ |
| ण | ṇa | /ɳə/ |

| ष | ṣa | /ʂə/ |
|---|---|---|
| स | sa | /sə/ |
| ह | ha | /hə/ |
| ळ | ḷa | /ɭə/ |
| क्ष | kṣa | /kʃə/ |
| ज्ञ | jña | /ʤɲə/ |

| प | PA | /PƏ/ |
|---|---|---|
| फ | PHA | /FƏ/ |
| ब | BA | /BƏ/ |
| भ | BHA | /B$^{ɦ}$Ə/ |
| म | MA | /MƏ/ |

| त | Ta | /t̪ə/ |
|---|---|---|
| थ | Tha | /t̪$^{h}$ə/ |
| द | Da | /d̪ə/ |
| ध | Dha | /d̪$^{ɦ}$ə/ |
| न | Na | /n̪ə/ |

| य | ya | /jə/ |
|---|---|---|
| र | ra | /rə/ |
| ल | la | /lə/ |
| व | va | /wə/ |
| श | śa | /ʃə/ |

Figure 1 b. consonants in Marathi Language along with Transliteration and IPA

input requires generally two stage interfacing. The first step requires an automatic recognition system (ASR) and the second step requires a system for speech understanding. The speech recognition systems that are on the market today are the embodiments of new algorithms that were once the province of the advanced laboratories. Many groups in India have also been engaged in speech related work like Tata Institute of Fundamental Research, Mumbai, Computer Vision and Pattern Recognition Unit at Indian Statistical Institute Kolkata, C-DAC Pune, Indian Institute of Technology, Madras, Indian Institute of Technology, Kanpur and BAMU [7].

There are the different ways to express the emotions by humans. Humans express their emotions by speech and actions like crying, yelling, dancing, laughing, stamping, and many other things [8]. But when it comes to speech human emotions affects the tone and the speaking style of the person. The emotion in speech sound affects the Speech recognition accuracy. The researchers around the globe are taking interest in detection of emotion in the Speech. In human computer interaction, many researchers are finding the depth of the area for emotion detection from speech. During the last few years, the research on speech emotion recognition has got much attention. Many emotional speech databases have been developed and the studies are carried on the developed emotional speech database around the world. [9]

The paper is organized as follows. In section 2 describes the development of emotional Marathi speech database; Section 3 explains what is meant by emotional feature extraction is explained. Section 4 describes the Mel Frequency Cepstral Coefficient. Speech emotion recognition based on MFCC is presented in section 5. An LDA technique is discussed in section 6. Finally, section 7 gives conclusions and future work of the work.

## II. Development Of Artificial Emotional Marathi Speech Database

We developed an emotional speech database using Marathi movies. For that purpose, we selected 5 Basic emotions i.e. Happy, Sad, Angry, Surprise and Afraid. It was decided to develop an artificial (from actors' actresses dialogue) emotional speech database using various Marathi movies in which the professional artists simulate the natural emotions. Capturing the natural emotions is not possible.

For the development of the artificial emotional speech database we have used 5 different Marathi movies. The movies were selected from different genre like drama, comedy & horror. The movies were averagely of two hour's duration. The file format of the movie used was MPEG-coded image segment with frame size of 720x480 pixels with frame rate of 24 fps (frame per second). The audio stream of the movie was sampled at 44 kHz.

The selected movies were first viewed and checked for the expressed emotions. The data samples were extracted from the movie using Cool Edit Pro version 2.0. As mentioned earlier the audio stream of the movie was sampled at 44 kHz so we down sampled the audio to 16 kHz 16 bit. Once the data samples were acquired we categorized the samples according to the emotions i.e. Happy, Sad, Anger, Afraid and Surprise. The extracted data samples were saved in .wav file format.

The database was developed after doing a comparative study of some of existed speech database in our earlier paper. The comparison was done on the initial standard input taken; the data collection from various situations, audio recordings, the instruments used for recordings, speakers and the percentage of the robustness comes after recognition. Some samples of happy, sad and angry along with their Transliterated (Translated in English) and IPA are shown in table I, II and III [10].

TABLE I. HAPPY WORDS IN MARATHI LANGUAGE ALONG WITH TRANSLITERATION AND IPA

| Devanagari | Transliterated (Translated In English) | IPA |
|---|---|---|
| कायहो | Kayho (Yes) | /kəajəhəo/ |
| खरंच | Kharach (Really) | /kʰərətʃə/ |
| शाब्बास | Shabbas (Good) | /ʃəabbəasə/ |
| अरेवा | Arewa (Well) | /ərəivəa/ |

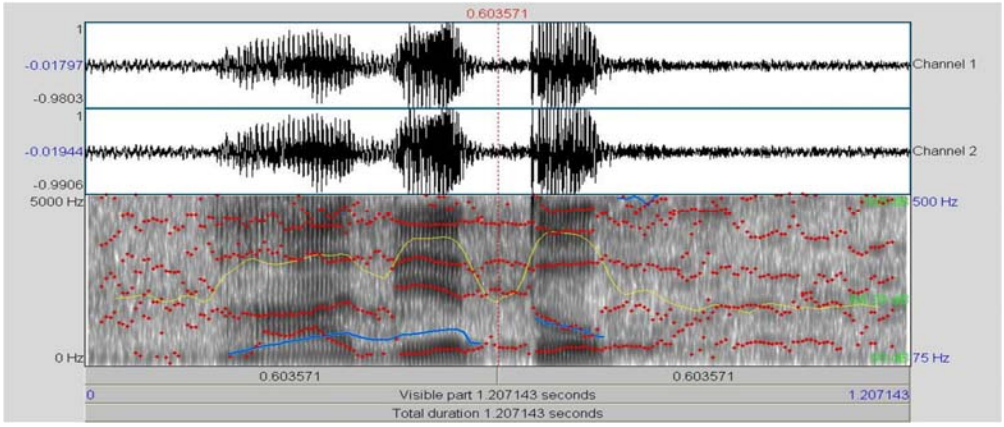
TABLE II. SAD WORDS IN MARATHI LANGUAGE ALONG WITH TRANSLITERATION AND IPA

| Devanagari | Transliterated (Translated In English) | IPA |
|---|---|---|
| काय झाल | Kay Zal (What Haapen) | /kəajə/ /dʒʱəalə/ |
| चुकल माझ | Chukla Maz (Sorry) | /tʃəukələ//məadʒʱə/ |
| माझ्या कर्मा | Mazya Karma(My Luck) | /məadʒʱəjə/ /kəarəmə/ |
| गेला र | Gelar (Pass away) | /gəəilərə/ |

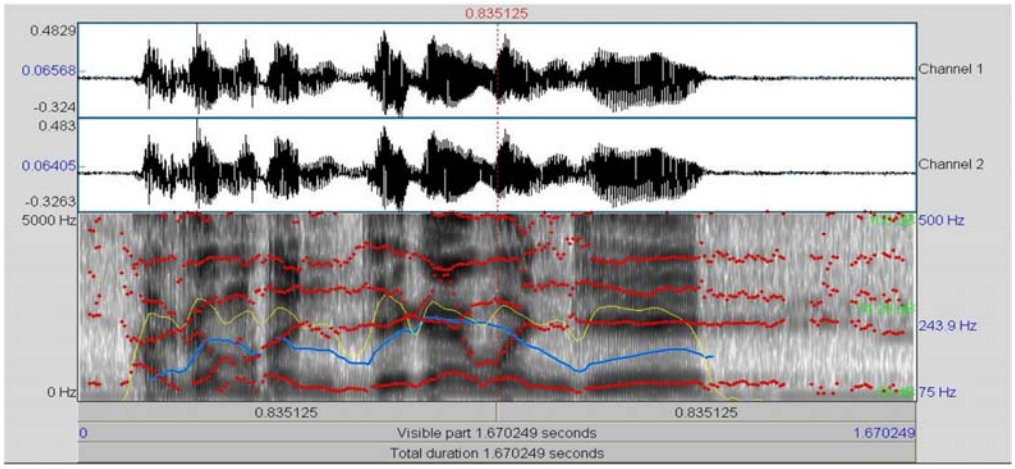
TABLE III. ANGRY WORDS IN MARATHI LANGUAGE ALONG WITH TRANSLITERATION AND IPA

| Devanagari | Transliterated (Translated In English) | IPA |
|---|---|---|
| ए गप्पे | Ai Gappey (Don't Speak) | /əi/ /gəppəi/ |
| चल निघ | Chal Nigh (Get out) | /tʃəalə/ /ɲəigʱə/ |
| उठ इथून | Utha Ithun (Stand up) | /uʈʰə/ /iɳʰəɳə/ |
| नकोय मला | Nakoy Mala (Don't Want) | /ɲəakəojə/ /mələa/ |

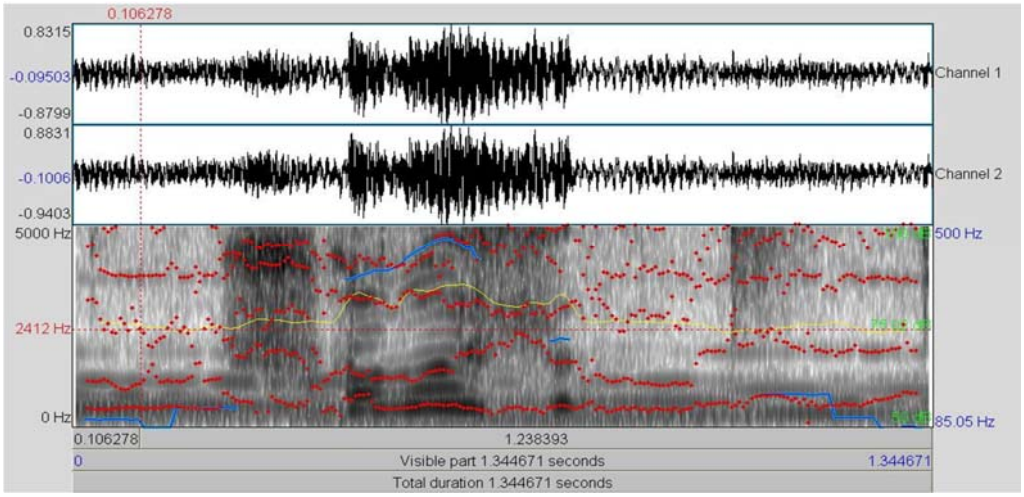
Each dialogue is uttered with one of the following emotions: angry, happy and sad and sample speech signals are given in figure 2.




Type of emotion: Anger




Type of emotion: Happy




Type of emotion: Sad

Figure 2. Speech Samples

## III. EMOTIONAL FEATURES

Feature extraction is a basic and fundamental pre processing step in pattern recognition and machine learning. It is a special form of dimensionality reduction technique used to reduce the data which is very large to be processed by an algorithm and extraction of specific properties from various features. In feature extraction, the provided input data is transformed into a set of features which provides the relevant information for performing a desired task without the need of the full size data but using the reduced set [11]. The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. To select suited features carrying information about emotion is necessary for emotion recognition. Studies on emotion of speech indicate that pitch, energy, formant, Feature extraction is process of extracting some valuable parameters for further processing of the input signals. We have studied various features extraction techniques as given below.

The prosodic features such as mean, maximum, minimum, standard – deviation of pitch and energy, and audible durations are extracted from four basic classes of emotions namely anger, sadness, happiness and neutral. These features are classified by fuzzy min-max neural network [12]. Kai-Tai Song, et. at. proposed a method of emotion recognition. Firstly, end-point detection and frame setting are done in pre-processing. Secondly, statistical features from pitch and energy are computed. Theses statistical features are classified by using support vector machine [13]. Stevros Ntalampiras and Nikus Fakutakis have used short-term statistics, spectral moments, and autoregressive models as a emotional features. Additionally, they have employed a

newly introduced groups of parameters based on wavelet decomposition [14]. It is shown from the above review spectral features are giving more recognition compatibly prosodic features of emotional speech. So we have decided MFCCs as features and LDA as classifier.

## IV. Mel Frequency Cepstral Coefficient (MFCC)

The Mel Frequency Cepstral Coefficient is the well known and most widely used feature extraction method in speech domain. The MFCC is based on the human auditory perception system. The human auditory perception system does not follow a linear scale of frequency. For each tone with actual frequency 'f' measured in Hz, a subjective pitch is calculated, is known as 'Mel Scale'. The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1000Hz. As a reference point, the pitch of a 1 KHz tone, 40 dB above the perceptual hearing threshold is defined as 1000 Mels [15].

There are various implementation of MFCC. These implementations differ mainly due to the number of filters, the shape of the filters, the way the filters are spaced, the bandwidth of the filter and the manner in which the spectrum is warped.

The main variations of the MFCC are as follows:

*A.* MFCC FB-20: introduced in1980 by Davis and Mermelstein [16].

*B.* MFCC FB-24 HTK: from the Cambridge HMM Toolkit by Young, 1995 [17].

*C.* MFCC FB-40: from Auditory Toolbox for MATLAB written by Slaney, 1998 [18].

*D.* HFCC-E FB29: (Human Factor Cepstral Coefficients) by Skowronski and Harris, 2004 [19].

The FB in the implementations of defines the number filters present in the Filter bank for the MFCC by the corresponding author. These implementations consider different sampling rates. To compute the features using MFCC the steps that are followed are Pre-emphasizing, Framing and Windowing, Fast Fourier Transform, Mel-Frequency Filter Bank, Logarithm and Discrete Cosine Transform.

The various steps involved in the calculation of MFCC are described below:

### *A. Pre-Emphasizing*

The speech signal is first pre-emphasized with the pre-emphasis filter 1-az-1 to spectrally flatten the signal.

### *B. Framing and Windowing*

A speech signal is assumed to remain stationary in periods of approximately 20 ms. Dividing a discrete signal s[n] into frames in the time domain truncating the signal with a window function w[n]. This is done by multiplying the signal, consisting of N samples. The signal is generally segmented in frame of 20 to 30 ms; then the frame is shifted by 10 ms so that the overlapping between two adjacent frames is 50% to avoid the risk of losing the information from the speech signal. After dividing the signal into frames that contain nearly stationary signal blocks, the windowing function is applied. For the proposed work the frame length was set to 25 ms and the frame was shifted by 10 ms.

### *C. Fast Fourier Transform*

Fast Fourier Transform converts each frame N samples from time domain into frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT), which is defined on the set of N samples $\{x_n\}$, as equation 1,

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} \quad \text{k=0, 1, 2 ... N-1} \qquad 1$$

In general $X_k$ are complex numbers and we only consider their absolute values (frequency magnitudes). The resulting sequence $\{X_k\}$ is interpreted as follow: positive frequencies $0 \le f < Fs/2$, correspond to values $0 \le n < N/2-1$, while negative frequencies $-Fs/2 < f < 0$ corresponds to $N/2+1 \le n \le N-1$. Here, Fs denote the sampling frequency. The result after this step is often referred to as spectrum or periodogram. To obtain a good frequency resolution, a 512 point Fast Fourier Transform (FFT) is used.

### *D. Mel Frequency Filter Bank*

A filter bank is created by calculating a number of peaks, uniformly spaced in the Mel-scale and then transforming back to normal frequency scale where they are used as peaks for the filter banks.

*E. Logarithm*

The logs of the powers at each of the Mel frequencies are calculated. The new array/vector of Mel log power is generated.

*F. Discrete Cosine Transform*

Discrete Cosine Transform (DCT) is being used to achieve the mel- cepstrum coefficients. In a frame, there are 24 Mel Cepstral coefficients, out of 24 only 13 coefficients have been selected for the recognition system.

## V. Speech Emotion Recognition Based On Mfcc

Overall the paper demonstrates the importance of the emotional speech database, for emotional speech recognition and synthesis. We have described briefly, the development of emotional speech databases in Marathi language. We have used the Mel Frequency Cepstral Coefficients (MFCCs) for the feature extraction purpose of this work. Mel Frequency Cepstral Coefficient (MFCC) is effective feature to distinguish certain emotions [20].

We have used only the pitch information extracted from the utterances for purposes of classification. Several studies indicate the importance of summary features, the fundamental pitch signal. To extract MFCC of artificial emotional speech, a set of speech samples are trained to learn the mapping between the acoustic signals.

In MFCC, the frequency scales are placed on a linear scale for frequency less than 1KHz and on a logarithmic scale for frequencies above 1 KHz. MFCC contain both time and frequency information of the signal which makes them ideal for Automatic Speech Recognition and Automatic Emotion Recognition [21]. MFCCs are the results of a cosine transform of the real logarithm of the short term energy spectrum expressed on a Mel–frequency scale. The MFCC tells about the short time energy migration in frequency domain. In MFCC a DFT spectrum of a signal is frequently warped through a Mel-frequency scale transformation by using the equation (1).

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

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

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

Where $d_i$ is the delta coefficient at frame i computed in terms of the corresponding basic coeffecients $c_{n+i}$ to $c_{n-i}$. The same equation is used to compute the acceleration coefficients by replacing the basic coefficients with the delta coefficients.

The Cepstral Mean Normalization is aimed at reducing the effect of multiplicative noise on the feature vectors. Mathematically it is:

$$c_i = c_i - \frac{1}{N}\sum_{k=1}^{N} c_{ik} \quad (3)$$

Where $c_i$ is the $i^{th}$ feature element in the feature vector and $c_{ik}$ is the $i^{th}$ feature element at frame k. N is the number of total input frames of data.

## VI. Linear Discriminant Analysis (LDA):

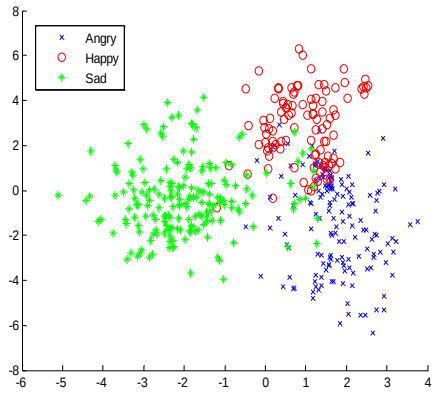
The matrix datasets were created into different three classes form features of angry happy and sad dialogs of all the samples for training purpose. Feature of the different emotion were subjected to the Linear Discriminant Analysis (LDA). The aim of LDA is used to reduce the dimensions of feature matrix and to clusters the data representing the different classes. The Linear discriminant function g(x) can be written as equation (4).

$$g(x) = \omega_0 + \sum_{i=1}^{d} \omega_i x_i \quad (4)$$

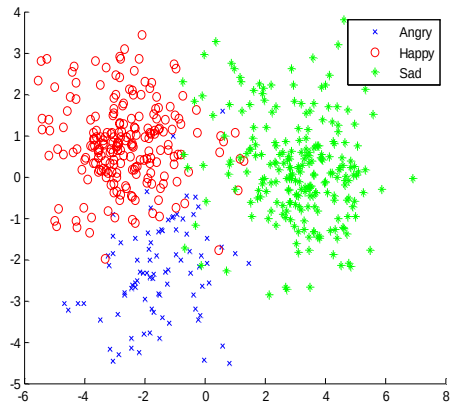
where the coefficients $\omega_i$ are the components of the weight i vector w. By involving the products of pairs of components of x the quadratic discriminant function is obtained and written as equation (5).

$$g(x) = \omega_0 + \sum_{i=1}^{d} \omega_i x_i + \sum_{i=1}^{d} \sum_{j=1}^{d} w_{ij} x_i x_j \quad (5)$$

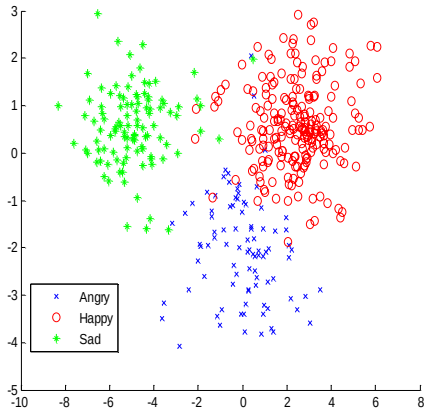
Since $x_i x_j = x_j x_i$ and $\omega_{ij} = \omega_{ji}$ with no lose thus, the quadratic descriminant function has an additional d(d+1)/2 coefficients at 8 its disposal with which complicated separating surface. For the classification and clustering purpose [22], we have created appropriate different classes of dataset. The three separate classes of datasets have been created according to three different emotions (angry, happy and sad) of all samples. 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. We have classified samples of 50 angry, 34 happy and 29 sad. Figure 3 shows graphical representation of sensitivity and clustering of 12 emotions out of 113.



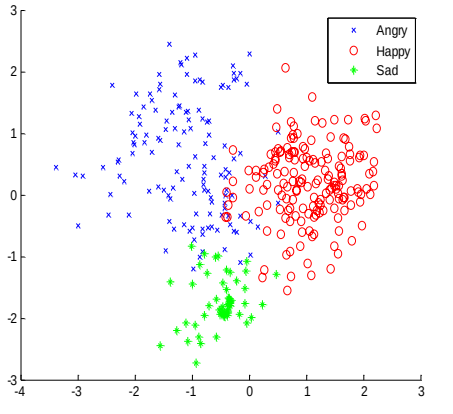

a. Clustering of dialogues: Ai Gappey (Don't Speak), Kayho (Yes), and Kay Zal (What Haapen) are angry, happy and sad emotions respectively




b. Clustering of dialogues: Chal Nigh (Get out), Kharach (Really) and Chukla Maz (Sorry) are angry, happy and sad emotions respectively




c. Clustering of dialogues: Utha Ithun (Stand up), Shabbas (Good) and Mazya Karma(My Luck) are angry, happy and sad emotions respectively




d. Clustering of dialogues: Nakoy Mala (Don't Want), Arewa (Well) and Gelar (Pass away) are angry, happy and sad emotions respectively

Figure 3: Clustering of three emotions: angry, happy and sad

## VI. Conclusion

In this study, we have analyzed the artificial emotional speech corpus by using Mel Frequency Cepstral Coefficients (MFCCs). We have conducted the experiment on our own developed database of 100 utterances by the male actors and female actress from Marathi movies. We have successively trained and tested the data samples. The samples are recorded by using a high quality Software's at a sampling rate of 16 KHz. This paper elaborates the development of Artificial Marathi emotional speech which was developed using the some Marathi Movies. It also describes recognition of Emotion from developed emotional speech database by using MFCCs. The 12 cepstral coefficients have been extracted from all emotional speech signals and these coefficients are classified with the help of Fisher linear discriminant analysis. Moreover, in the figure 2, all the features of a particular emotion are more crowded at a particular point is the sensitivity of emotion. More the distance among the clustering means more recognition. The emotions were recognized but the recognition rate for the developed database is low. The system is not able to recognize the happy emotion with more accuracy.

## Acknowledgement

This work is supported by University Grants Commission as Major Research Project. The authors would like to thank the University Authorities for providing the infrastructure to carry out the research.

## References

[1] W. Seok Lee, Yong-Wan Roh, Dong – Ju Kim, Jung – Hyun Kim, and Kwang – Seok Hong, “Speech Emotion Recognition using spectral Entropy”, ICIRA 2008, Part II LNAI 5315, PP 45-54, Springer – Verlag Berlin Heidelberg 2008.

[2] D. O’Shaughnessy, “Speech Communication: Human and Machine”, IEEE Proceeding Computer Society Press, Hardcover (1999), 742-770 D. O’ Shaughnessy 2006..

[3] L. Rabiner, B-H Juang, “Fundamental of Speech Recognition”, Prentice Hall International Inc.,ISBN 0-13-285826-6, April 22, 1993

[4] J. Deller, et. al., “Discrete – Time Processing of Speech signals”, MacMillan Publishing Co., ISBN 0-02-328301-7. (1993)

[5] C. Rowden, “Speech Processing”, (1992) McGraw- Hill, ISBN : 007707324X, 9780077073244

[6] D. Jurafsky and J. H. Martin, “Speech and Language processing”, (2009) Prentice Hall, ISBN 0131873210, 9780131873216.

[7] VishwaBharat@tdil ISSN No. 0972-645 TDIL PROGRAMME Ministry of Communications & Information Technology, Department of Information Technology Electronics Niketan, 6, CGO Complex, New Delhi-110003 in the year 1991-92

[8] Vishal B. Waghmare, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal, "A Comparative Study of Various Emotional Speech Databases", International Journal of Computer Science and Engineering (IJCSE), Vol 4, No. 6, June 2012 pp. 1236-1240, ISSN: 0975-3397

[9] Shashidhar G. Koolagudi, K. Sreenivasa Rao, “Emotion Recognition from speech: A Review”, International Journal of Speech Technology, 15 (2012), pp. 99–117

[10] Ganesh Janvale, Vishal Waghmare, Vijay Kale and Ajit Ghodke, “Recognition of Marathi Isolated Spoken words Using Interpolation and DTW techniques”, ICT and critical infrastructure: proceeding of the 48th Annual of Computer Society of India Vol I. Advances in Intelligent system 3-319-03107-1_3, Print ISBN 978-3-319-031066 Online ISBN 978-3-319-03107-1, January 2014.

[11] Pukhraj P. Shrishrimal, “Design and Development of Spoken Marathi Isolated Words Database for Agriculture Purpose and its Analysis”, M.Phil. Computer Science Thesis. May 2013.

[12] N. P. Jawarkar, “Emotion Recognition using Prosody Features and a fuzzy min-max neural classifier”, IETE Technical Review, Vol 24, No. 5, 2007, PP 369-373.

[13] Kai-Tai Song, Meng-Ju Han and Shih Chieh Wang, “Speech Signal based emotion recognition and its application to entertainment robots”, Journal of the Chinese Institute of Engineering, 2012, 1-12, iFirst Article, Taylor & Francis.

[14] Stavros Ntalampirs and Nikos Fakotaki, “Modeling the Temporal Evolution of Acoustic Parameters for speech Emotion Recognition”, IEEE transactions on Affective Computing, Vol. 3 No 1. January – March 2012.

[15] Vibha Tiwari, "MFCC and its application in speaker recognition", International Journal on Emerging Technologies, Vol. 1, No. 1, pp. 19-22 (2010).

[16] Davis S. B., Mermelstein P., “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences”, IEEE Transaction on Acoustic, Speech and Signal Processing, Vol. 28, No. 4, pp. 357-366 (1980).

[17] Young S.J., Odell J., Ollason D., Valtchev V., Woodland P., “The HTK Book. Version 2.1”, Department of Engineering, Cambridge University, UK, 1995.

[18] “The NIST Year 2001 Speaker Recognition Evaluation Plan”, The NIST of USA, 2001.
[19] Skowronski M.D., Harris J.G., “Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition”, Journal of the Acoustical Society of America, Vol. 116, No. 3, pp.1774–1780, Sept 2004.
[20] Xia Mao, Lijiang Chen, Bing Zhang, “Mandarin speech emotion recognition based on a hybrid of HMM/ANN”, International Journal of Computers, Issue 4, Volume 1, 2007.
[21] Yu Zhou, Yanqing Sun, Lin Yang, Yonghong Yan "Applying articulatory features to speech emotion recognition" ThinkIT Speech Lab., Institute of Acoustics, Chinese Academy of Sciences, Beijing, IEEE International Conference on Research Challenges in Computer Science, 2009
[22] Duda RO, Hart PE and Stork DG. “Pattern Classification”, john Wiley & 2nd edition 2001.