IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-10, Oct 2019 E-ISSN: 2347-2693

Emotion Recognition in Marathi Language by using Fast Fourier Transform

Shinde Ashok R.[1*], Agnihotri Prashant P.[2], Raut S.D.[3], Khanale Prakash B.[4]

[1]Punyashlok Ahilyadevi Holkar Solapur University, Solapur, India
[2]Swami Ramanand Teerth University, Sub-Centre Latur, India
[3]Punyashlok Ahilyadevi Holkar Solapur University, Solapur, India
[4]D.S.M. College, Parbhani, India

**Corresponding Author: shinde_ashok@redifmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i10.4347 | Available online at: www.ijcseonline.org

Accepted: 07/Oct/2019, Published: 31/Oct/2019

***Abstract*—** Emotion Recognition is a recent research area which can be applied in various applications. Speech Emotion is useful in E-learning, medical and in entertainment. The proposed work focuses on emotion recognition using Fast Fourier Transform and Marathi speech database. In this work six emotions are considered and Fast Fourier Transform is used for feature extraction and as a major feature for finding specific emotion. The Marathi words which represents the emotion like surprise and sad like *Are Bapre* (अरे बापरे !), *Kiti Wilakshan* (किती विलक्षण !), *Are Deva* (अरे देवा !) etc are used as a speech samples for analysis purpose. This paper highlights the overview of existing speech database as well as the newly developed Marathi emotional speech database. In this proposed work total 52 Marathi utterances are used in the experimental purpose. Fourier Transform is used to convert time domain signal into Frequency domain signal. The proposed experimental work gives 100% recognition rate for surprise and disgust emotion. Similarly 90% accuracy for sad emotion and 87.5% accuracy got for fear and angry emotion. The overall average recognition rate for six emotions is 93%. Very less accuracy rate i.e 62.5% got for happy emotion.

Keywords— Emotion Recognition, SER, AER

I. INTRODUCTION

Automatic Emotion Recognition (AER) is very recent research field. To make better communication with computer, the computer should able to communicate with human being. This can be done by recognizing human emotions. If computers are able to identify human emotions, it is more natural to communicate with computers. Broadly there are two approaches are considered by various researchers i.e. Facial expression and tone of the voice to recognize human emotions. Emotions like happy, sad, angry, fear, disgust and surprise are considered for the proposed work. As per literature review by using facial expression approach, the recognition rate is more as compare to speech.

Many researchers have proposed important speech features which contain emotion information, such as energy, pitch, formant frequency, Linear Prediction Cepstrum Coefficients (LPCC)[1]. The experimental results specifies that the feature combination of MFCC and MS has the highest accuracy rate on both Spanish emotional database using RNN classifier 90.05% and Berlin emotional database using MLR 82.41%[2]. Speech is one of the most natural communications between human beings. Humans also express their emotion via written or spoken languages. Speech emotion is an important role to express the feeling of one's expression.[3]

In the proposed work, some Marathi emotional words are used for experimental purpose. The Marathi emotional words *Are Bapre* (अरे बापरे !), *Kiti Wilakshan* (किती विलक्षण !), *how is boaring (kiti kantalwane,)* Oh my God! (अरे देवा !) etc.

Speech analysis can be done mainly in two domains namely as: Time domain and Frequency domain analysis. Time domain analysis tells mainly about the amplitude of the signal, on the other hand frequency domain analysis focused on individual frequency components present in the signal. Frequency domain analysis also highlights on the phase information of the system at different frequencies.

II. SPEECH DATABASE USED FOR EMOTION RECOGNITION

Two emotional speech databases are available as a standard for emotion recognition i.e. Berlin and Spanish database.

The Berlin speech database is frequently used in emotion recognition. These databases contain total 535 utterances spoken by 10 actors (5 female and 5 males) in 7 simulated emotions (anger, fear, boredom, sadness, joy and neutral).

The second speech database is Spanish emotional database which contains utterances from two professional actors (one male and one female). These databases recorded for six emotions i.e anger, joy, fear, sadness, surprise, disgust and normal. This database contains total 4528 utterances [1].

About Marathi Emotional Speech database:
In the development of emotional speech database of Marathi language, fifty speakers with different age groups, different sociolinguistic background and who is able to express his emotion have been selected to record the speech utterances. In Marathi language there are various words and phrases are available which can be useful for recognizing human emotions. The emotions such as happy, angry, sad, surprise, fear, disgust and neutral are the basic emotions. The words in Marathi such as "Aare Deva"(Oh God!), "Kiti Chan"(How Good!) are used to express human emotions. The words which strongly represent the emotions have been also selected for development of Marathi emotional speech database as shown in table 1and 2.

Table 1: Four Marathi Emotional Words/Phrases (for **Happy** Emotion)

Sr_No	Words/phrases in Marathi	Meaning in English	Written in Devanagari Word-net
1	Are Wa	Great !	अरे वा !
2	Kitti Chan	How good	किती छान
3	Kitti God	How nice !	किती गोड !
4	Mast	Superb !	मस्त !

Table 2: Four Marathi Emotional Words/Phrases (for **Angry** Emotion)

Sr_No	Words/phrases in Marathi	Meaning in English	Written in Devanagari Word-net
1	Gap Re	Don't talk	गपरे
2	Hat	Shut up	हट !
3	Nakoch mala	Don't need !	नकोच मला !
4	Chal Nigh	Gate out	चल निघ !

Feature Extraction:
The speech database contains a large number of features. These features reflect the emotional state. In recent research in Speech Emotion Recognition (SER), common features are extracted such as pitch, energy, formant and some spectrum features used for emotion recognition. Speech analysis can be done mainly in two domains namely as: Time domain and Frequency/Spectral domain analysis. Time domain analysis tells mainly about the amplitude of the signal, on the other hand frequency domain analysis focuses on individual frequency components present in the signal. Frequency domain analysis also highlights on the phase information of the system at different frequencies.

III. METHODOLOGY

The speech signal contains a large number of parameters that reflect the emotional characteristics. In recent research, the common features like energy, pitch, formant, and some spectrum features such as Linear Prediction Coefficients (LPC), Mel-Frequency Cepstrum Coefficients (MFCC) and Modulation spectral features. In this proposed work we have selected Fourier Transform for feature selection.

Fourier Transform:
Fourier Transform is also has several applications in speech analysis. It is also helpful in filter designing in order to remove any specific band or frequency component present in the speech signal. Fourier Transform can also act as feature vector for speaker recognition.

The mathematical formula known as Fourier Transform, in order to convert time domain signal into frequency domain signal, Discrete version of the same is known as Discrete Time Fourier Transform (DTFT). Mathematically it can be represented as follows:

$$X(w) = \sum_{n=-\infty}^{\infty} x(n)e^{-jwn}$$

Where x(n) is the discrete sequence and X(w) is the continuous function of the frequency. Due to discrete nature of computers this continuous function cannot be computed on digital processors and hence discrete version of X(w) is required in order to make it computable on digital machines. Discrete version of DTFT or X(w) is known as Discrete Fourier Transform (DFT) and it's obtained by uniform sampling of X(w). This is given as follows X(w) = X(wk). Where Wk = 2πk, k = 0, 1, ….. (N – 1), where N : no. of samples of X(w). In order to avoid aliasing in the reconstructed time domain signal x(n), N should be greater than or equal to the length of x(n). There are several algorithms available for fast computing of DFT and those are known as Fast Fourier Transform (FFT). It should be noted that FFT is complex quantity whose absolute value gives the information about the amplitude of individual frequency component while angle part gives the phase information of the signal.

In the experimental work firstly read the speech sample which is having 44100 sampling frequency and calculates the length of signal. Then we calculate FFT in which what is

next power of two which number is greater than number of samples. The NFFT can be mathematically represented as:

NFFT = 2^nextpow2 (l)

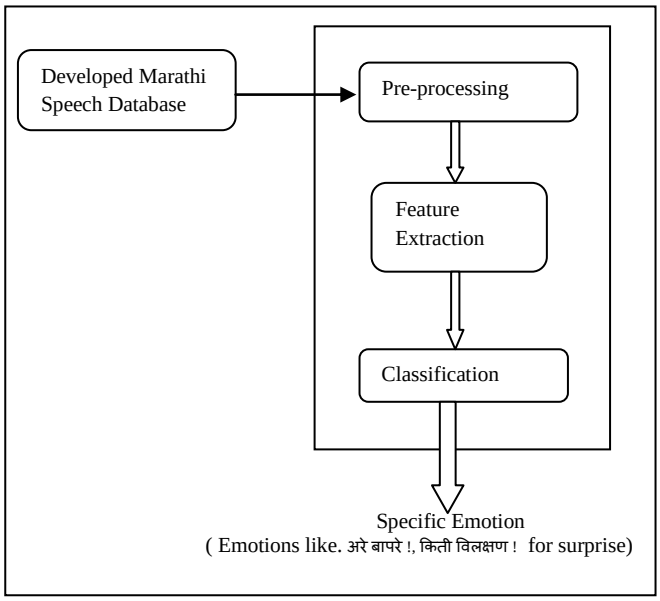
Then we will create frequency space f which is half of the sample frequency i.e.

F= fs/2*linspace(0,1,NFFT/2+1);

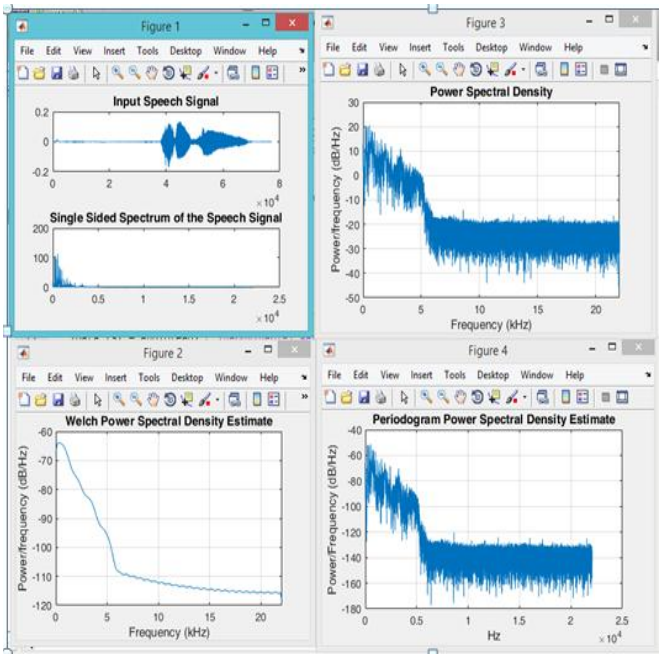
Now to take Fourier Transform we have to take absolute value,

xf = abs(fft(data,NFFT));

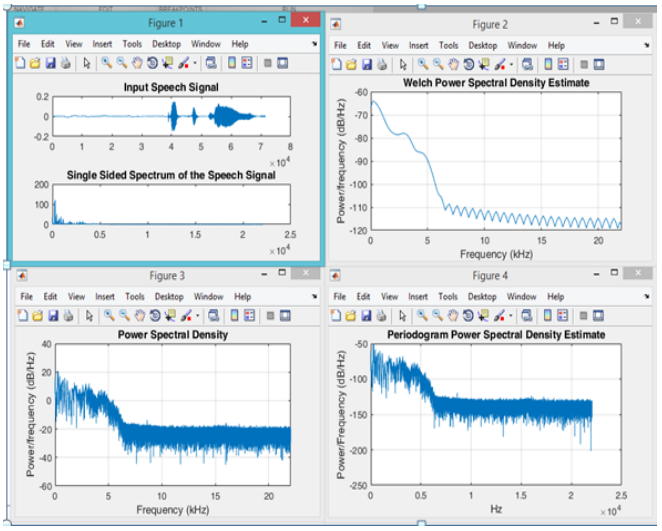
The flow of experiment is shown in the following block diagram of Speech Emotion Recognition System (SERS).



Block diagram of Speech Emotion Recognition System (SERS).



(Emotion : Happy (**अरे वा !**))



(Emotion: Happy (**किती छान**!)

Power Spectral Density: Power Spectral Density (PSD) describes that variation of power present in the signal as a function of frequency. In other words it tells which frequency components contain higher power and which frequency contain lower power. It shows the variation in the form of a graph with frequency of x-axis and power (db/Hz or watt/Hz) on y axis. PSD calculation can be done by simply by periodogram command in Matlab. It takes data, window and sampling frequency as input and returns the PSD and frequency of the data. Figure 1 shows the PSD estimate of the input speech signal based on periodogram command in matlab.

[psdestx, Fxx] = periodogram(data, rectwin(length(data)), length(data), fs);

Welch's Method for PSD estimate is an improved method to estimate the PSD of input speech. It is better than the periodogram method as it reduces the amount of noise in estimated PSD at the cost of frequency resolution. Figure 3 shows the PSD estimate of the input speech signal using Welch's method.

IV. RESULT

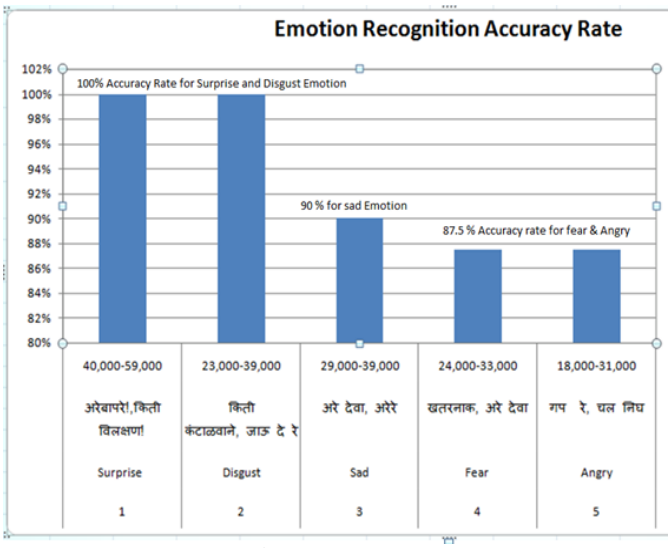
In this proposed, Fast Fourier Transform (FFT) is calculated on each input speech signal. The Marathi speech database is used for experiment purpose. Total six types of emotions are considered i.e. happy, angry, disgust, fear, sad and surprised. Total 52 Marathi emotional words are used. Total 10 Marathi words are used for sad & surprise emotions and 8 Marathi speech samples for angry, disgust fear and happy emotions.

After analysis of fxx values, it is observed that 100% recognition rate is achieved in surprise and disgust emotion.

90% accuracy rate is achieved for sad emotion, 87.5% for angry and fear. Very less i.e 62.5% accuracy rate got for happy emotion. The Marathi words for surprise like are (अरे बापरे !), *Kiti Wilakshan* (किती विलक्षण !), *Are Deva* (अरे देवा !) are used for experimental work.

(Table shows: Analysis and Accuracy Rate for five emotions :)

Sr.	Emotion	The Value of Fxx Ranges	Total Marathi Speech samples	Emotion Recognition Accuracy Rate
1	Surprise	40,000-59,000	10	100%
2	Disgust	23,000-39,000	08	100%
3	Sad	29,000-39,000	10	90%
4	Fear	24,000-33,000	08	87.50%
5	Angry	18,000-31,000	08	87.50%
Overall Average Recognition Rate				**93%**



(Graphical representation of Emotion Recognition Accuracy Rate)

V. CONCLUSION

The emotion recognition of Marathi emotional speech database using Fast Fourier Transform and fxx value gives 100% accuracy in Surprise and Disgust emotion. The recognition accuracy for Sad is 90% and the recognition accuracy for fear and angry is 87.5%. It is also observed that all six emotions has a fix range of fxx value. So the developed method is also helpful to recognize other emotions of the human beings also.

REFERENCES

[1] Surabhi Vaishnav and Saurabh Mitra, "Speech Emotion Recognition: A Review", International Journal of Engineering and Technology (IRJET), Vol. 03, Issue:04, Apr-2016.

[2] Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, Kosai Roof and Mohamed Ali Mahjoub, "Speech Emotion Recognition: Methods and Cases Study", in Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART 2018) – Vol.2,Pages 175-182.

[3] Dhruvi Desai, "Emotion Recognition using Speech Signal: A Review", International Research Journal of Engineering and Technology (IRJET), Vol. 05, Issue:04, Apr-2016.

[4] Shweta Sinha, "Analysis and Recognition of Dialects of Hindi Speech", International Journal of Scientific Research in Computer Science and Engineering (IJSRCSE), Volume-3, Issue-5, Oct 2015.

[5] V.K. Jain and N. Tripathi, "Speech Features Analysis and Biometric Person Identification in Multilingual Environment", International Journal of Scientific Research in Network Security and Cmmunications-6, Issue-1, February 2018.

[6] Agnihotri P.P., Shinde A.R. and Khanale P.B. "Development of Noise free Marathi Speech database using various Filtering Techniques", Indian Journal of Research, Volume-5, Issue-2, February 2016.

[7] Oh-Wook Kwon, Kwokleung Chan, Jiucang Hao, Te-Won Lee, "Emotion Recognition by Speech Signals" EUROSPEECH 2003 - INTERSPEECH 2003 8th European Conference on Speech Communication and Technology Geneva, Switzerland, September 1-4, 2003.

[8] S.Lalitha, Sbhishek Madhavan, Bharath Bhushan, Srinivas Saketh,"Speech Emotion Recognition", International Conference on Advances in Electronics, Computers and Communications(IJAECC), 2014.

[9] Nobuo Sato, Yasunari Obuchi, "Emotion Recognition using Mel-Frequency Cepstral Coefficients", Journal of Natural Language Processing, pp 83-96, 2007.

Authors Profile

Mr. Shinde Ashok R. has completed M.Sc. in Computer Science in 1997 and qualified UGC-SET in Computer Science & Applications in 2006. He is perusing his Ph.D. in Computer Science. He is currently working as Assistant Professor in Punyasholok Ahilyadevi Holkar Solapur University, Solapur. He has 19 years teaching experience at UG and PG level and 8 years of research experience. He has published 20 research articles in the reputed journals and conference proceedings. His area of interest is Digital Image Processing, Data Mining & Warehouse, and Network Security etc.

Dr. Prashant P. Agnihotri has completed M.Sc. in comptuer Science in 2002 and qualified UGC-SET in 2006 in Computer Science and Applcatons. He has awarded Ph.D. degree in Computer Science in August 2019. He is working as Assistant Professor in the School of Technology,

S.R.T.M. University Nanded, Sub-Centre, Latur. He has published more than 10 research articles in reputed journal. He has completed one minor research project. His areas of interest are Pattern Recognition (Speech), Signal Processing and Core Computer Science subjects.

Dr. Raut S.D. has completed M.Sc. in comptuer Science in 2011 and qualified UGC-SET in 2011 in Computer Science and Applcatons. He is completed his Ph.D. in 2018. Presently he is working as Assistant Professor in the Department of Computer Sceine, P.A.H. Solapur University, Solapur. He is author of book and published more than 30 research articles in reputed journal. His keen areas of research are pattern recognition, atificial intelligence and network security.

Dr. Prakash B. Khanale has completed M.Sc. in Physics-Electronics in 1986 from Marathwada University, Aurangabad (MS)-India. He completed his M.Phil. in Electronics in 1995 and Ph.D. in Electronics in 2005 from Shivaji University, Kolhapur, (MS)-India. He is currently working as a Professor in the department of Computer Science, Dnayanopasak Shikshan Mandal's College of Arts, Commerce and Science, Parbhani-431401 (M.S.) India. He has published more than 26 books on Computer Science and Electronics. He has published total 78 research papers in the National and International Journals. Under his guidance Six Research Students Awarded Ph.D. in Computer Science and seven students are awarded M.Phil Degree. He has completed 04 research projects sponsored by UGC, NCERT.

