iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52105

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

A Review on Speech Emotion Classification using Linear PredictiveCoding and Neural Networks

Dr. C M Patil[1], Aishwarya M[2], Nisarga K R[3], Lokasharadhi S[4], Uma D[5]
[1, 2, 3, 4, 5]*Department of Electronics and Communication Vidyavardhaka College of Engineering Mysuru, India*

Abstract: ***Speech is one of the significant method for communicating their thoughts for human beings. Using the intelligence of computing devices to understand human emotions from language has become an interesting area. In the recent years this topic has grabbed so much attention. Speech emotion recognition (SER) has made significant paces with the evolution of hardware and software systems in digital signal processing field. SER is a key feature of human-computer interaction systems that are mostly employed in various fields such as healthcare, automated call centres, and distance learning. SER comprises of the depth study about the signal and also recognizing the appropriate emotions in relation to the pre-determined dataset using the extracted features.***

The proposed system consists of speech signal, feature extraction module, pre-determined dataset, classifier, and lastly the classified emotions. Initially, the speech signal will be pre-processed which helps to eliminate the unwanted noise signal present in the speech signal. This pre-processed signal will be sent to the feature extraction module for the extraction of different types of features contained in a speech. Lastly, mapping only the required features corresponding to different types of emotions is done. This mapping is done by using classifiers. In SER, several methods are employed to obtain the different types of emotions present in speech signals. Recently, Deep learning techniques forSER were offered as an alternative to traditional methods.

Keywords: Neural networks, Linear Predictive coding, Speech Emotion Recognition, Speech emotionclassification.

I. INTRODUCTION

Speech has become a vital means of communication between the people. It is the capacity to communicate ideas and feelings using a range of vocal tones and body gestures. The piece of doing this is also known as speech. The job of SER is to extract the speaker's emotion from the speaker's audio signal.

Identifying these emotions provides an image into the deeper complications that supports to navigate over actual circumstances. Speech processing mainly focuses on the analysis of voice signals and the signal processing techniques, to analyse speech for automatic recognition of emotions and extracting the data.

Speech expresses human feelings. There are many kinds of emotions such as happy, sad, surprise, fear, bore etc., Emotion can play an important role in decision making. Recognizing the different emotions from the speech signal has a various application and are active in several fields, such as healthcare, medicine, entertainment, etc.,.

Presently, SER is primarilyutilized to facilitate interaction between humans, computers, and robots. In our project we will be using Linear predictive coding (LPC).

It is one of the most widely used method. It is a speech analysis technique and it’s also a technique that helps humans encode high-quality audio at low bit- rates using computers and robots.

II. METHODOLOGY

The proposed system consists of speech signal, feature extraction module, predetermined dataset, classifier, and lastly the classified emotions. The speech or the audio signals are generally pre-processed in prior to the feature extraction in order to enhance the energy level present in higher order frequencies relating to the lower order frequencies.

This process is called pre-emphasis. Speech encoding, speech segmentation, and removal of noise in speech are all included in speech pre- processing. Feature extraction makes use of computational procedure to detect the distinguishing characteristics present in the samples of the speech signal. In the classification phase, various typesof classifiers will be applied to classify the emotions in the speech signal such as SVM, HMM, Neural Networks and many more. Below figure represents the block diagram of the proposed system.

©IJRASET: All Rights are Reserved

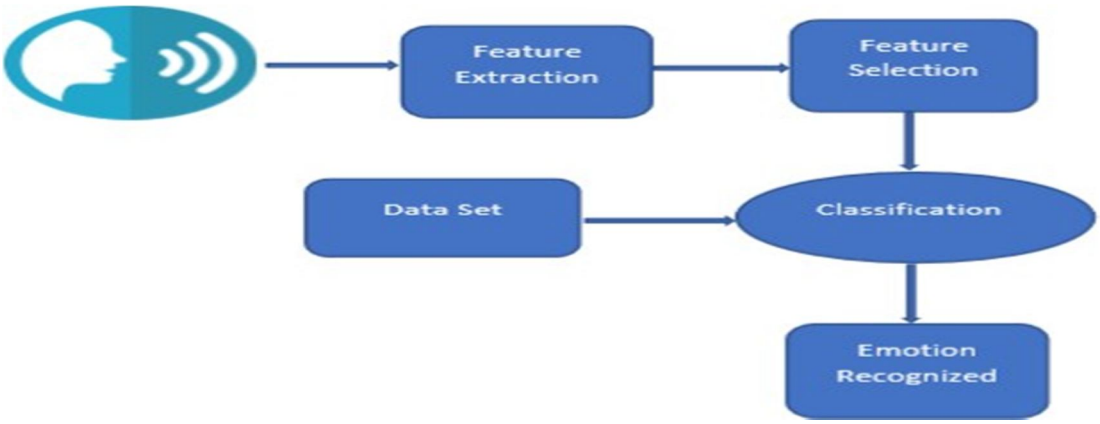


Figure 1: Block Diagram of Speech Emotion Classification

III. LITERATURE REVIEW

Manish P. Kesarkar [1] proposed a Feature extraction for Speech Emotion Recognition using the methods Cepstral Analysis, Mel Cepstrum Analysis, Linear Predictive Coding (LPC) Analysis, Perceptually Based Linear Predictive Analysis (PLP). Cepstral analysis provides information about pitch and vocal tract configuration. LPC analysis provides depiction of vocal tract configuration by the simple computation when compared to the cepstral analysis. The emotions recognized are happy, sad and fear. Results can be improved in case of vowels but still acceptable if order of model is sufficiently high (80%).

Soham Chattopadhyay et.al. [2] have claimed Optimizing Speech Emotion Recognition usingMantar-Ray Based Feature Selection by the methods LPC, PLP and MFCC and also use the Multilayer Perception classifier. The emotions recognized are sorrow, anger, hatred. This method gives the two-emotion recognition database, two feature extraction methods and achieves the best performance about 83.5% on both datasets. The datasets are Emo-DB and SAVEE datasets.

R. Subhashree and G. N. Rathna [3] proposed a Speech Emotion Recognition performance analysis based on Fused Algorithms and GMM modelling. Method used is a combination of five algorithms such as MFCC, LPC, LPCC, LFCC and OSALPC and Gaussian Mixture Model classifier. This creates a dataset Emo-DB, a German database. The emotions recognized are happy, angry, fear, sad, disgust, boredom. Overall efficiency of 89% was achieved using the fused algorithms which was higher than that of the individual algorithms.

Rashmirekha Ram et.al. [4] proposed an Emotion Recognition with Speech for Call Centres using a methods LPC and Spectral analysis. The dataset used a speech database containing speech sample from 25 different females. The emotions are recognized are happy, sad, angry, surprise and bore. This gives the result of ‘sad’ emotion has lowest prediction error rate (less than 50%) and ‘surprise’ has highest prediction error rate (more than 50%).

Mr. Kenaz K Babu and Mr. Valanto Alappat [5] proposed a Speech Emotion Recognition using a method MFCC and classified using Deep Learning Technique (RNN and LSTM). TESS- Toronto emotional speech dataset is created for recognition of emotions. The emotions recognized are anger, disgust, fear, happy, surprise, sadness, neutral. 95% of accuracy was achieved. Usage of LSTM provides maximum accuracy about 98% for feature extraction using MFCC method.

K. Pavan Raju et.al. [6] proposed an Automatic Speech Recognition System Using MFCC based LPC Approach with Back Propagated Artificial Neural Networks using a method that is a combination of LPC and MFCC and is classified using Artificial Neural Networks (ANN). The emotions that are observed with the combination of LPC and MFCC has the highest accuracy when compared to the individual methods. 97% of accuracy was achieved. Proposed ANN deep learning model requires less training period and has faster training performance compared to other classifiers such as SVM, RNN etc.

Ruhul Amin Khalil et.al. [7] proposed a Speech Emotion Recognition classified using Deep Learning Technique (RNN, DBN, CNN). The emotions observed are Panic, Joy, Happy, Surprise. The methods of Deep Learning and their layer wise architectures have been developed on the basis of the classification of several real emotions such as happiness, sadness, surprise etc. Such approaches provide simple model training as well as the effectiveness in shared weights.

Apoorv Singh et.al. [8] proposed a Speech Emotion Recognition using a method MFCC and classified using Convolution Neural Network (CNN). The emotions are observed that are Happy, Sad, Angry, Fear, Calm, Disgust and Surprise. 71% of accuracy was achieved. Model would have executed well with more data. Also, it has succeeded in distinguishing masculine and feminine voice.

Babak Basharirad and Mohammadreza Moradhaseli [9] proposed a Speech Emotion Recognition using a method MFCC, LFPC, LPCC and classified using GMM, HMM, SVM. Berlin database is created for the recognition of emotions. The emotions recognized are happy, sad, angry and fear. HMM with adopting short time LFPC as a feature proves an 80% of accuracy for a Berlin dataset.

Mehmet Cenk Sezgin et.al. [10] proposed a Perceptual Audio Features for Speech Emotion Detection are classified using GMM/SVM. EMO-DB and VAM database are created for recognition of emotions. The emotions are recognized are sadness, surprise, angry and fear. The performance is valid both in natural emotions and in acted emotions assessed on the EMO-DB and VAM corpus.

IV. CONCLUSION

The widely used feature extraction technique is MFCC, LPC (Linear Predictive Coding) feature extraction technique is not used much, so we are utilizing LPC as a feature extraction technique for our project.

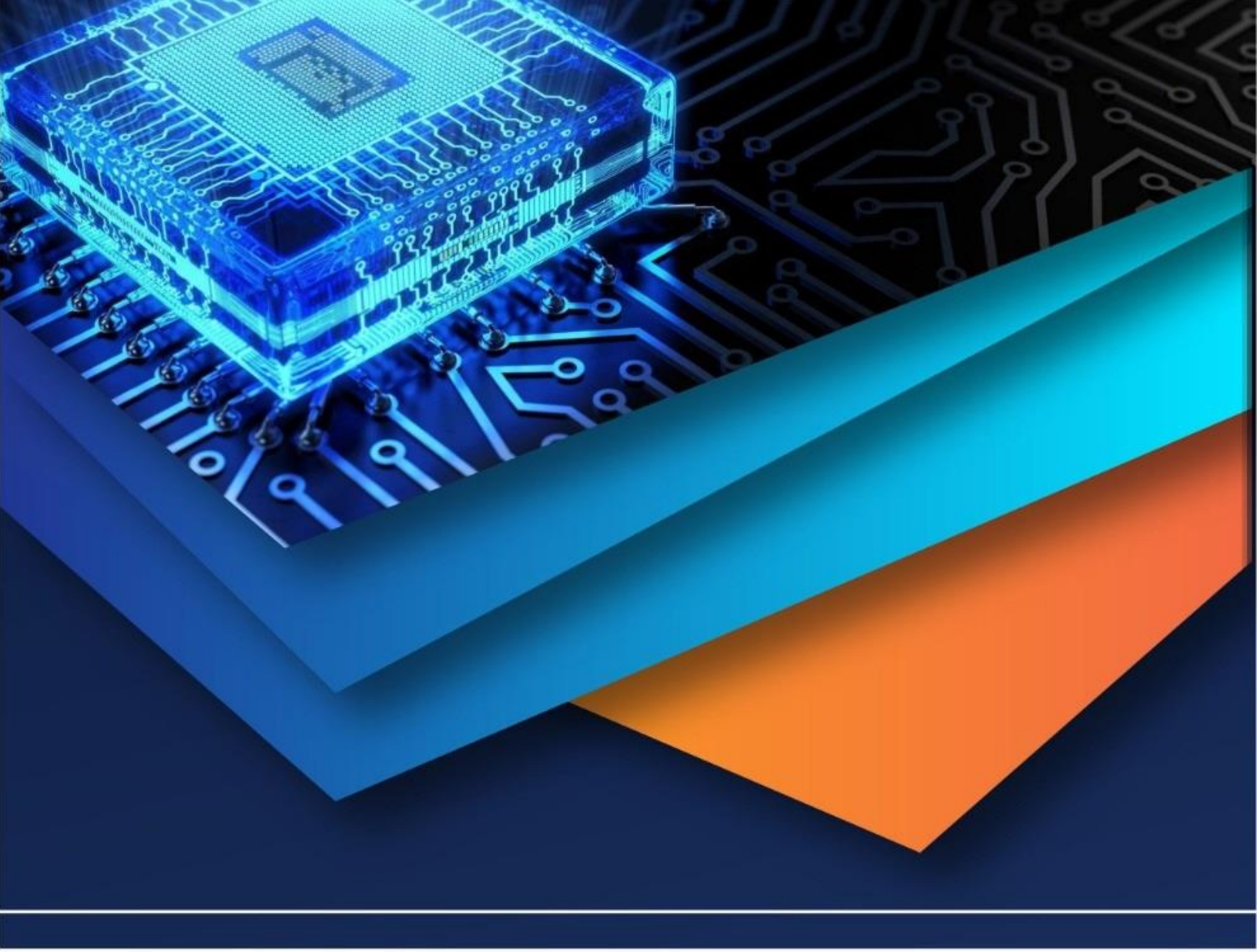
Recognizing the emotions from input speech signals is a significant but it is also a challenging aspectof interaction between humans and computers. Here, the importance and working of Speech Emotion Recognition (SER) system is provided, in addition to the comprehensive overview on various types of SER system was discussed. This paper represents the survey of the work done in recent days w.r.t., find best feature extraction technique and classifier to get better accuracy in finding emotions from personalized speech signal.

In all these papers, they have tested their models against the existing datasets, so we are aiming to create a new dataset.

REFERENCES

[1] Manish P. Kesarkar, “Feature Extraction for Speech Recognition”, Electronic Systems Group, EE.Dept, IIT Bombay, November 2003.
[2] Soham Chattopadhyay, Arijit Dey & Hritam Basak, “Optimising Speech Emotion Recognition usingManta-Ray Based Feature Selection”, 18 September 2020.
[3] R. Subhashree and G. N. Rathna, “Speech Emotion Recognition: Performance Analysis based onFused Algorithms and GMM Modeling”, Vol 9(11) Indian journal Technology, March 2016.
[4] Rashmirekha Ram, Hemanta Kumar Palo, Mihir Narayan Mohanty, “Emotion Recognition with Speech for Call Centres using LPC and Spectral Analysis”, International Journal of Advanced Computer Research”, Volume-3 Number-3 Issue-11, Septemper-2013.
[5] Mr. Kenaz K Babu & Mr. Valanto Alappat, “Speech Emotion Recognition Using Deep Learning Technique”, International Journal of Research in Engineering and Science, 2320-9356 www.ijres.orgVolume 10 Issue 7, pp.436.440, July 2022.
[6] K. Pavan Raju, A. Sri Krishna and M. Murali, “Automatic Speech Recognition System Using MFCCbased LPC Approach with Back Propagated Artificial Neural Networks”, ICTACT Journal on Soft Computing, July 2020.
[7] Ruhul Amin Khalil, Edward Jones, Mohammad Inayatullah Babar, Tariqullah Jan, Mohammad Haseeb Zafar and Thamer Al Hussain, “Speech Emotion Recognition Using Deep Learning Techniques”, July 25 2019, accepted August 5 2019, date of publication August 19 2019, date of current version September 4 2019.
[8] Apoorv Singh, Kshitij Kumar Srivastava, Harini Murugan, “Speech Emotion Recognition Using Convolution Neural Network (CNN)”, International Journal of Psychosocial Rehabilitation, Volume 24, Issue 08 2020.
[9] Babak Basharirad and Mohammadreza Moradhaseli, “Speech Emotion Recognition Methods”, AIP Conference, 03 October 2017.
[10] Mehmet Cenk Sezgin, Bilge Gunsel and Gunes Karabulut Kurt, “Perceptual Audio Features for Emotion Detection”, Sezgin et al. EURASIP Journal on Audio, Speech, and Music Processing 2012.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)