International Journal on Applications in Information and Communication Engineering
Volume 6 : Issue 3 : September 2020, pp 17 – 20 www.aetsjournal.com ISSN (Online) : 2394 - 6237

An Efficient Convolutional Neural Network based Speech Recognition Technique for Tamil Speech Automatic Recognition

Dr. S. Subasree , Dr.D.SatishKumar, Mr.S.Mani, Mr.L.Krishna Kumar, R. Karthika, S. Pooja ,

***Abstract*—** Speech and Voice Recognition, sometimes known as Voice Biometrics, is used for the recognising an individual Voice or Speech, This is the systems for finding various applications in a number of different domains including Telephone Banking • E Commerce and Forensics, Human Voice is also being incorporated into user interfaces for accessing and controlling smartphones and personal virtual assistants in the form of voice-controlled user interfaces. Thus the voice recognition is an prime importance to access and control as well of our Smart Devices. Thus more researchers are focusing Speech Recognition Technique to achieve better recognition with higher accuracy. The Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) are considered as the best Recognition Approaches for Voice and Speech Recognition. This existing work computed average weight of MFCC and LPC for predicting and recognising Speech or Voice Recognition. This project work is identified that the Hybrid Model of MFCC and LPC through Convolutional Neural Network(CNN) is facilitating to achieve better classification accuracy. Thus this work is planned to design an efficient CNN based Speech Recogniser for Speech Recognition. This is partially implemented and studied for Tamil Speech Recognition

***Keywords*—**Speech , Voice, Smart Devices, Voice Controller, etc.

I. Introduction

Speech is the most primary,efficient and widely used mode of communication between humans beings. There are large number of different spoken languages which are used throughout the world. The communication among the human is mostly done by vocally for information exchange[1]. Thus it is natural for people to hope speech interfaces with computer. For a real-time a intelligent applications, it is essential that the machine can hear, interpret , analysis and act upon input information from speaker, and also give immediate response to complete the information transfer .This can be carried out by developing an Automatic Speech Recognition (ASR) system which is a procedure of translating an acoustic signal into a written text or a command without understanding what has been recognized[2].

Since, 1960s many computer scientists have been researching on different means to make computer record, interpret, analys and understand human speech. Research on automatic speech recognition by machine has attracted much attention over the last five decades. The survey shows that the agencies like AT & T Bell Labs, DARPA, IBM, and Microsoft have sponsored number of programs for research in this area in the last 50 years [3]. Still a lot of research work is being done in this area. Although ASR is still lagging far behind commercial speech recognition systems are being used in many applications, like , auto-attendants, virtual reality, electronic devices, dictation, controlling the various programs, automatic telephone call processing system and query based information system such as travel information system , weather forecast information system etc. Though Speech recognition systems has gained new heights and era but robustness and noise tolerant recognition systems are few of the problems which make them difficult to handle. Many researcher around the world are trying to develop a robust and noise tolerant speech recognition systems.

II. Existing System

A. *Mel Frequency Cepstral Coefficients (MFCC)*

- Mel-cepstral Frequency Coefficients (MFCC) have been used in literature [1], for modelling the human auditory perception system
- MFCC feature is also a very popular choice for performing Voice Recognition
- MFCC takes human perception sensitivity with respect to frequencies into consideration, and therefore are best for speech/speaker recognition

Dr. S. Subasree , Professor, Department of Computer Science and Engineering, , Nehru Insitute of Engineering and Technology, Coimbatore, Tamilnadu, India..

Dr. D.Satishkumar , Associate Professor, Department of Computer Science and Engineering, , Nehru Insitute of Engineering and Technology, Coimbatore, Tamilnadu, India..

Mr.S.Mani , Assistant Professor, Department of Computer Science and Engineering, , Nehru Insitute of Engineering and Technology, Coimbatore, Tamilnadu, India..

Mr.L.Krishna Kumar, Assistant Professor, Department of Computer Science and Engineering, , Nehru Insitute of Engineering and Technology, Coimbatore, Tamilnadu, India.

R. Karthika, UG Scholar Department of Computer Science and Engineering, , Nehru Insitute of Engineering and Technology, Coimbatore, Tamilnadu, India.

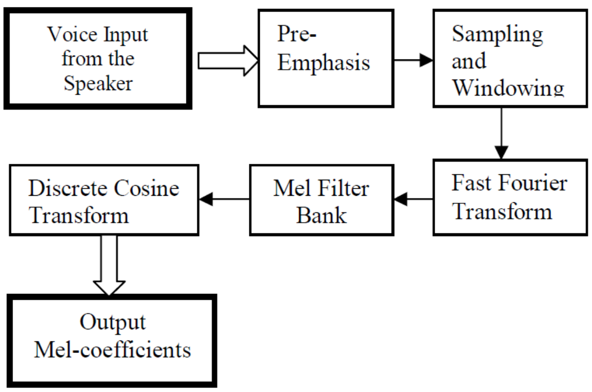
S. Pooja , UG scholar, Department of Computer Science and Engineering, , Nehru Insitute of Engineering and Technology, Coimbatore, Tamilnadu, India.

- The step-by-step computation of MFCC is shown in the Figure.

B. Linear Predictive Coding (LPC)

- The transfer function of the digital filter equivalent of vocal tract can be given by

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}}.$$

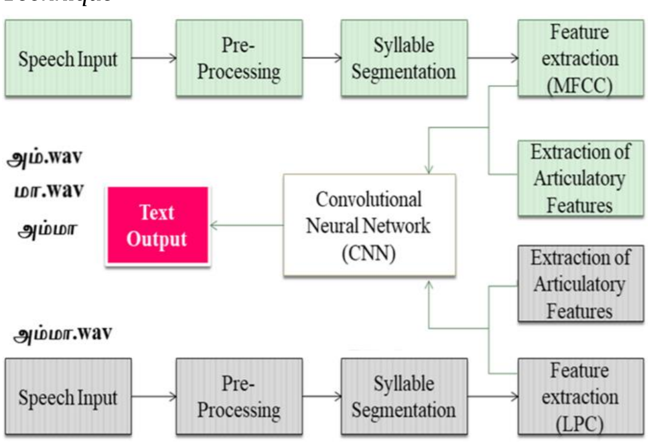


- Speech data is sequential in nature and, for modeling the vocal tract, assume that the voice acoustics of the nth speech sample (S[n])
- It can be viewed as a combination of p past speech samples
- Thus, the nth speech sample (S[n]) can be written as:

$$S[n] = \sum_{k=1}^{p} \alpha_k S[n-k] + G.u[n]$$

III. Proposed System

A. Convolutional Neural Network based Speech Recognition Technique



B. Modules

This Project Work consists of the following Modules. They are

- *Pre-Processing*
- *Syllable Segmentation*
- *Feature Selection*
- *Training and Pattern Recognition*
- *Speech Recognition*

C. Pre-Processing

The pre-processing is performed on input wave files. In the analysis, the speech patterns presence and absence were recorded Error Removal and Normalization

D. Syllable Segmentation

This Module is used to perform on input wave files to split Wave Forms Syllable wise, so that the Speech or Voice can be Recognized. The speech patterns presence and absence were recorded

E. Feature Selection

This Module will collect information from Syllable Segmentation Module and extract all the wave patterns with the help of Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC). The extracted values will ne recorded on Excel Sheet for further analysis

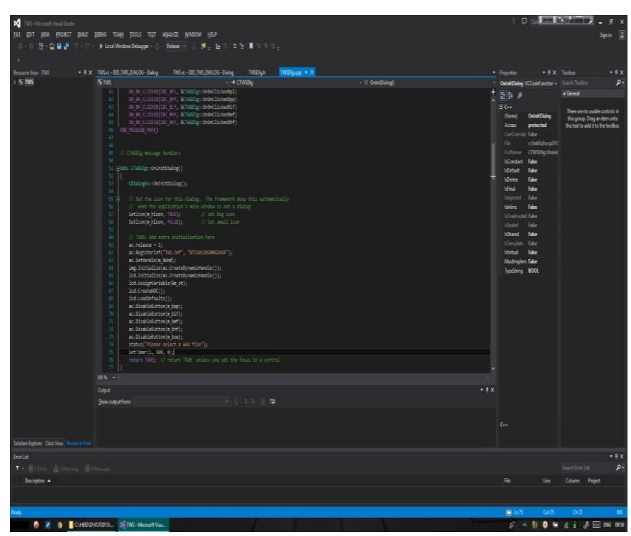
F. Training and Pattern Recognition

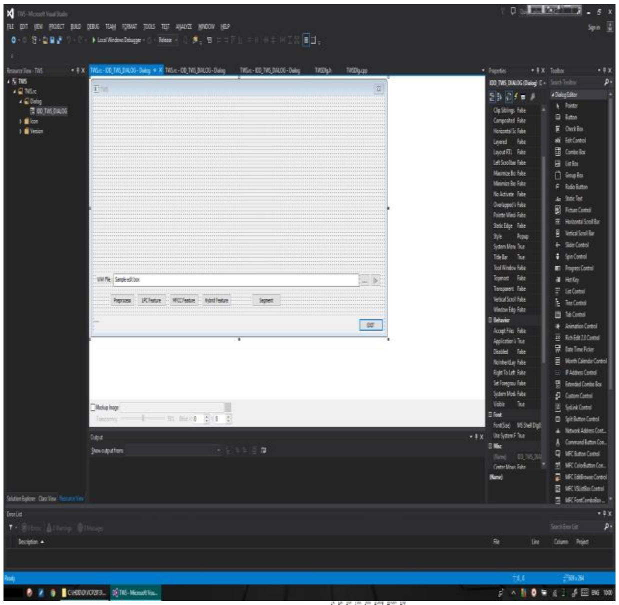
The output of the Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) is trained by Convolutional Neural Network for better Pattern Recognition and Speech Recognition

G. Speech Recognition

This Module will print the Speech Pattern in Text Pattern that will ensure that the Speech is recognized and confirmed. The Accuracy and Error Rate also will be measured to understand the efficiency of the proposed model

IV. Implementation





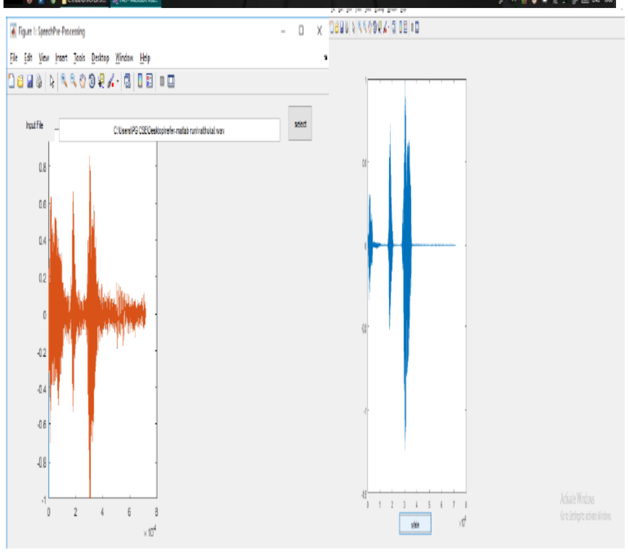
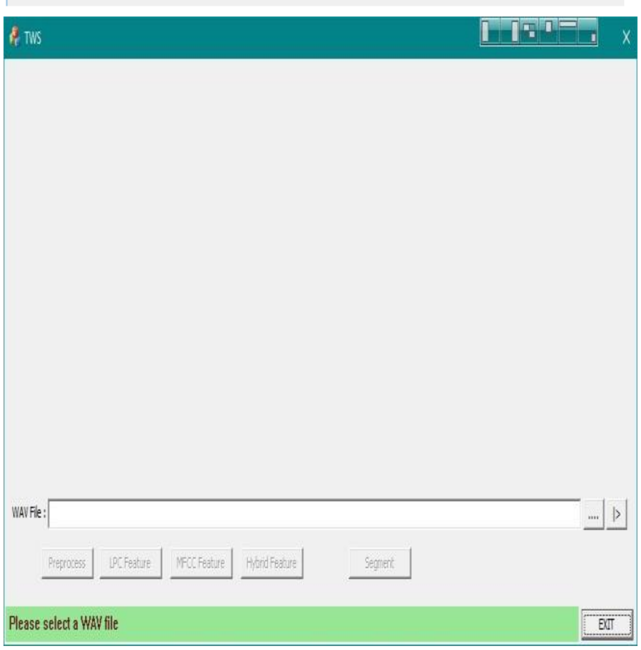
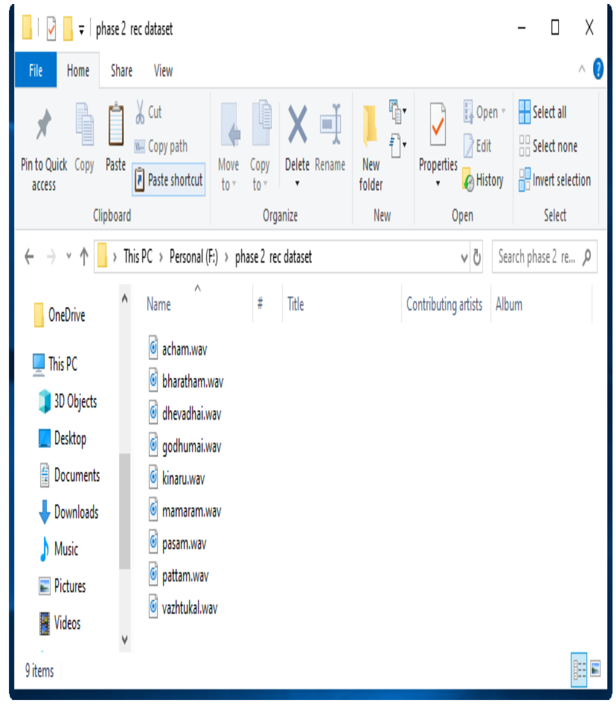
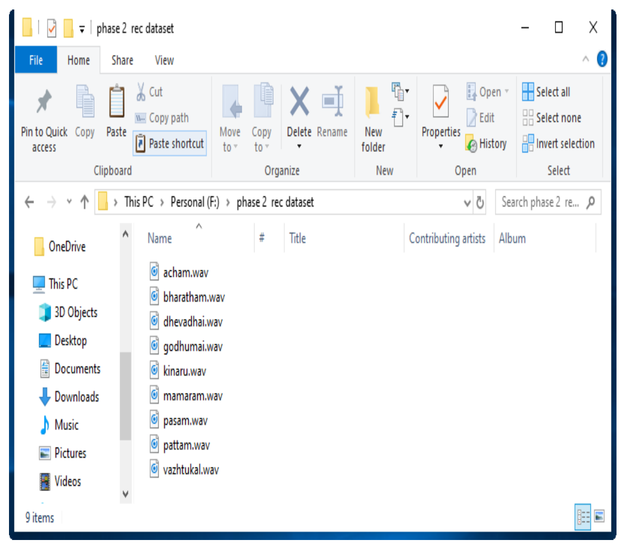


Figure 1: Speech Pre Processing
Input File
select


TWS
WAV File :
Preprocess
LPC Feature
MFCC Feature
Hybrid Feature
Segment
Please select a WAV file
EXIT


phase 2 rec dataset
This PC › Personal (F:) › phase 2 rec dataset
Name
acham.wav
bharatham.wav
dhevadhai.wav
godhumai.wav
kinaru.wav
mamaram.wav
pasam.wav
pattam.wav
vazhtukal.wav
9 items


phase 2 rec dataset
This PC › Personal (F:) › phase 2 rec dataset
Name
acham.wav
bharatham.wav
dhevadhai.wav
godhumai.wav
kinaru.wav
mamaram.wav
pasam.wav
pattam.wav
vazhtukal.wav
9 items

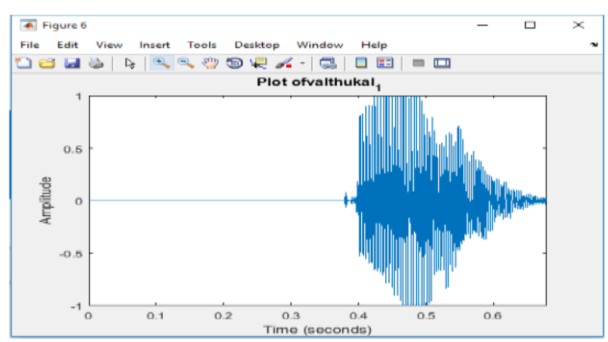
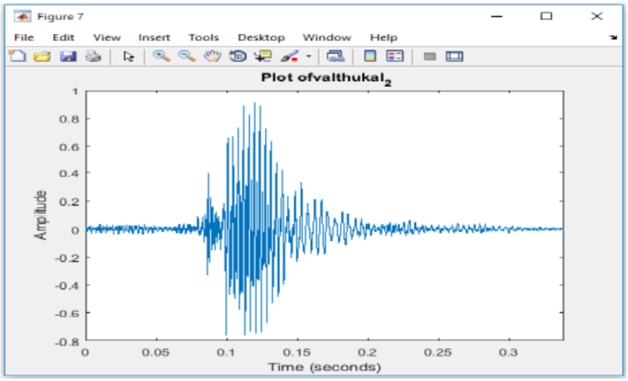
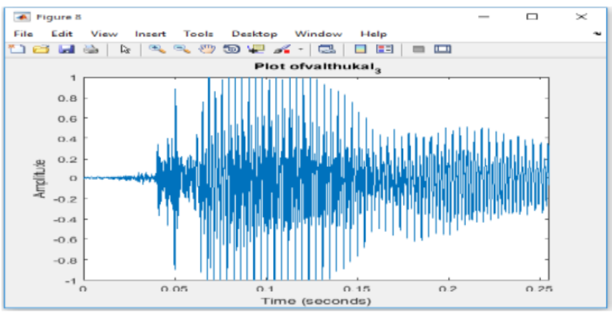


Figure 6
Plot ofvalthukal
Amplitude
Time (seconds)
1
0.5
0
-0.5
-1
0
0.1
0.2
0.3
0.4
0.5
0.6





V. Conclusion

The Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) are considered as the best Recognition Approaches for Voice and Speech Recognition. This work computed average weight of MFCC and LPC for predicting and recognising Speech or Voice Recognition. This project work is identified that the Hybrid Model of MFCC and LPC with the help of Convolutional Neural Network, it can be trained in a better way that will achieve better classification accuracy

Thus this work is implemented the CNN based Speech Recogniser for Speech Recognition. This work studied for Tamil Speech Recognition Successfully

References

1. Anurag Chowdhury, and Arun Ross, “Fusing MFCC and LPC Features using 1D Triplet CNN for Speaker Recognition in Severely Degraded Audio Signals”, IEEE Transactions on Information Forensics and Security, 2020.
2. Z. Zhang, L. Wang, A. Kai, T. Yamada, W. Li, and M. Iwahashi. Deep neural network-based bottleneck feature and denoising autoencoder-based dereverberation for distant-talking speaker identification. EURASIP Journal on Audio, Speech, and Music Processing, 2015.
3. X. Zhao, Y. Wang, and D. Wang. Robust speaker identification in noisy and reverberant conditions. IEEE/ACM TASLP, 2014.
4. A. Gibiansky, S. Arik, G. Diamos, J. Miller, K. Peng, W. Ping, J. Raiman, and Y. Zhou. Deep voice 2: Multi-speaker neural text-tospeech. In NIPS, 2017.
5. J. Guo, R. Yang, H. Arsikere, and A. Alwan. Robust speaker identification via fusion of subglottal resonances and cepstral features. The Journal of the Acoustical Society of America, 2017.
7. HSBC voice id making telephone banking safer than ever. https://www.hsbc.co.uk/1/2/voice-id. Accessed: 2017-12-29.
8. Morpho and agnitio partner, bring voice biometricsto criminal id. https://findbiometrics.com/morpho-and-agnitio-partner-bring-voice-biometrics-to-criminal-id-21261/.Accessed: 2018-06-13.
9. Siri background noise. https://support.apple.com/en-us/HT204389. Accessed:2017-12-29.
10. Voicevault biometrics to protect payments. https://findbiometrics.com/voicevault-biometrics-to-protect-payments-25131/. Accessed: 2018-06-13.