International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Continuous Telugu Speech Recognition on T-LPC and DNN Techniques

N.Uma maheshwari,. Archek Praveen Kumar, K. Narmada, Affrose, B.Sneha

Abstract--- This paper deals with the basic application speech recognition. There are many languages in the world but one of the regional language is Telugu. Recognition of this language helps in many applications for 8 crores of people stay in AP and Telangana states. Recognition is done by recording the speech signals and database creation. Pre-processing is done by 2 stage DNN (seep neural networks) where denoising, framing is done. The preprocessed signal features are extracted using TLPC(teager energy operator linear prediction filter). The features extracted are classified using DNN which generates adequate results. The results are obtained for continuous speech of Telugu language.

Keywords— Continuous speech recognition, Telugu language, T-LPC, DNN.

I. INTRODUCTION

Communication is the process of exchanging information. Voice signals video signals and data are used for communication. Mainly communication is used to transmit voice signal from one place to another place. The different voice signals are generated by human beings depending on the languages, traditions, gender, emotions and psychology of person. There are different languages exist in different geographical areas over the world.

Gender also differentiates the voice features depending on the feelings, reactions, mindset, situation and position. Voice signal is a one-dimensional signal. This one-dimensional signal has temporal features and spectral features. Temporal features are also known as time domain features which are simple to extract. Time domain features extract amplitude, energy of signal. Spectral features are known as frequency domain features which are converted from time domain signal. Frequency domain features can be used to identify pitch, strength, intensity and rhythm of voice signal.

There are different types of speeches exist. Those are isolated words which has a single word, connected word in which different words are separated by some space and the continuous words in which words are connected together continuously as a text. Automatic speech recognition is an important technique which can be recognized as a voice or it can be converted into a stream of words by using different methods. Stream of word/ text can be a machine recognizable.

In former researches single words are recognized automatically, later it has developed to recognize a stream of words like different languages. Automatic recognition of a language is complex which took the attention of researchers for decades. Recognition of speech is started in 1950s and it is continuing till now to develop different speech recognition procedures. Speech recognition has different methodologies to convert it into text. Those include pre-processing of speech, feature extraction, and feature decoding and post-processing. Speech/ non-speech segmentation is a method of pre-processing, which removes the background noise while recording the voice signal.

It is also known as endpoint detection. After pre-processing features of voice are extracted by different methods like MFCC, LPC and PLP.

There are three major methods to recognize speeches are:

1. Acoustic phonetic approach
2. Pattern recognition approach
3. Artificial intelligence approach

Acoustic approach is used to recognize the sounds depending on the level of frequency by following the steps like firstly analyze the spectrum of speech then recognize the features of the speech followed by segmentation of features to recognize the speech. In pattern based approach a pattern is developed for both reference data and unknown speech. The developed pattern is compared with the unknown data patter and the difference between the patterns is used determined. A logical step is used to decide the unknown speech in corresponding to the matching of two patterns. In this paper the research is done for continuous speech signal.

II. BLOCK DESCRIPTION

This paper deals with recognition of continuous speech Telugu language. The created databases features are generated and various parameters are calculated. The extracted parameters are classified later. All this process can be seen in figure 1.

Manuscript published on 30 September 2019
* Correspondence Author
N.Uma maheshwari*, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Telangana, Hyderabad.
Dr. Archek Praveen Kumar, Professor, HOD, Department of ECE, Malla Reddy College of Engineering for Women, Telangana, Hyderabad.
K. Narmada, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Telangana, Hyderabad.
Affrose, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Telangana, Hyderabad.
B.Sneha, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Telangana, Hyderabad.

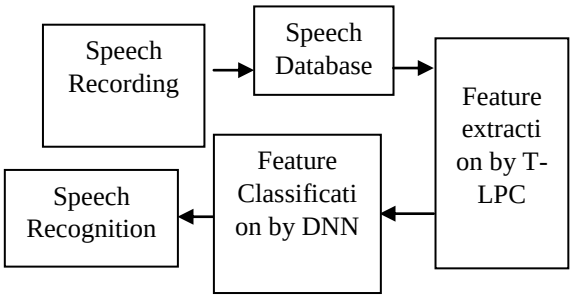

Fig.1 Block diagram

This recognition is done on continuous Telugu speech. Telugu is one of the sterile languages which is recognized on some basic parameters. Preprocessing is necessary to obtain proper recognition accuracy.

A. Pre-Processing

Two stage Deep Neural Network (2-DNN):

Two stages DNN preprocessing technique: In speech processing technique there is a problem of recognizing the main speech from noise produced by environment, vibrations and echoes. When processing an isolated speech data these disturbances are easily processed but for continuous speech data it is difficult to process the data and speech recognition.

Automatic speech identification as text or as voice is a major problem for noisy speech. So pre-processing of recorded speech signal to remove noise, vibrations and disturbances is an advantage for human beings and computers to recognize voice. There are different algorithms are developed for pre- processing of speech signal like Hidden markov mixing models, Gaussian mixture models. Later neural networks are developed to process the speech signal. Deep neural networks are has more advantages to process the complicated speech signal compared to other technologies. Deep neural network is artificial network which has large number of layers which also includes both input and output. In this paper a two stage DNN system to improve speech recognition by removing noise and vibration is developed. Two stage DNN algorithm removes noise in first stage and unwanted echoes can be removed in second stage.

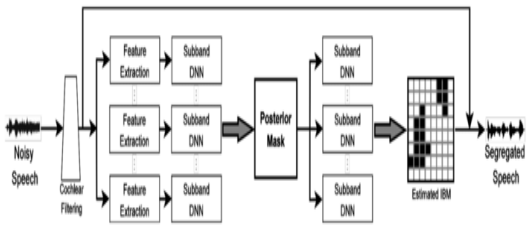

Fig.2 Block diagram

Linear prediction algorithm is used in subsequent stage to remove vibrations followed by Spectral deduction is used to remove noise in former stage. To get an improved speech magnitude spectrum noisy reverberant speech has been enhanced and combined with special speech. The detailed explanation is shown in figure 2. First a speech signal is recorded and enhanced in different segments, includes a segment of noise removing, a segment to remove vibrations and echoes and the last segment is to redesign the time frequency domain signal. By using masking techniques in different layers of DNN, noise is removed in first stage. by using spectral mapping techniques echoes are removed in second stage. in the last stage used time domain or frequency domain reconstruction methods to get back to the noise free speech signal. the pre-processing is shown in figure 3

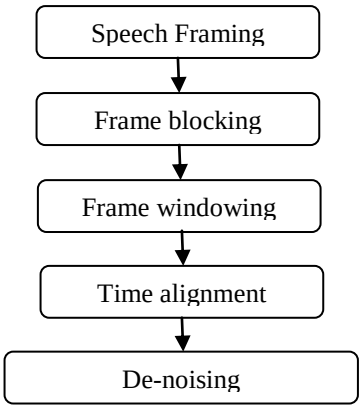

Fig 3 Pre processing flow diagram

B. Feature Extraction

The pre-processed speech signal is used to extract the features. Pre-processing includes filtering, framing and windowing. Speech signal has many features like pitch, intensity, energy, zero crossing rate, shaping and spectral. To identify the speech signal and to convert it into a text the features of speech signal to be detected first. There are many algorithms developed by researchers to extract the features of a speech signal.

The most widely used feature extraction methods are MFCC, PLD, RASTA and LPC. In this paper LPC method is used along with an energy operator called as teager energy operator which is used to extract the energy feature of speech. LPC method uses correlation transformation methods to extract the pitch or energy of a signal. To determine the auto correlated signal Levinson Durbin algorithm is used. A signal with less phase distortions is extracted by using this algorithm.

A speech signal is given to the teager energy operator to determine the energy feature of a signal. Later the signal is given to LPC extractor to determine the features like pitch. Then it is given to the classifier. Teager energy operator is define as an operator to determine the energy of a speech or any other signal.

Teager says that the pitch of a signal is generated by energy, the signal may be a time domain or frequency domain signal. Based on the frequency, amplitude, linear and nonlinear characteristics of energy of a speech signal is determined. It is also a end point detector to identify the zero crossings of frames in a speech signal. Later this is combined with Linear prediction coefficient algorithm to extract features of speech signal as shown in figure 4.

LPC is mostly used technique in which the speech signal is reduced its spectrum for efficiency of channel.

For this reduction the signal is sampled and quantized with some errors. An estimated speech signal is generated with this linear prediction.

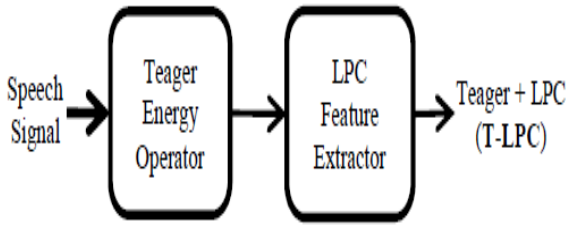

Fig 4 Teager-LPC feature extractor for speech signal

The Coefficients are the time or frequency with maximum amplitude or strength of the speech signal. The extracted speech signal features using T-LPC are classified in next stages. After extracting the features of a speech signal feature classification techniques are used. GMM, HMM and Neural Networks are most widely used classification methods.

C. Feature Classification

DNNs are predictably feed forward networks in which information flows from the input layer to the output layer without twisting back. At first, the DNN creates a map of essential neurons and assigns arbitrary numerical values, or "loads", to links between them. The loads and inputs are multiplied and coming back an output between 0 and 1. If the network didn't perfectly identify a certain pattern, an process would modify the loads. That way the process can make certain factors more influential, until it determines the correct mathematical operation to fully process the data.

Deep learning is a specific subfield of device learning, a new take on learning demonstration from information which puts an importance on learning successive **"layers**" of gradually meaningful demonstrations.

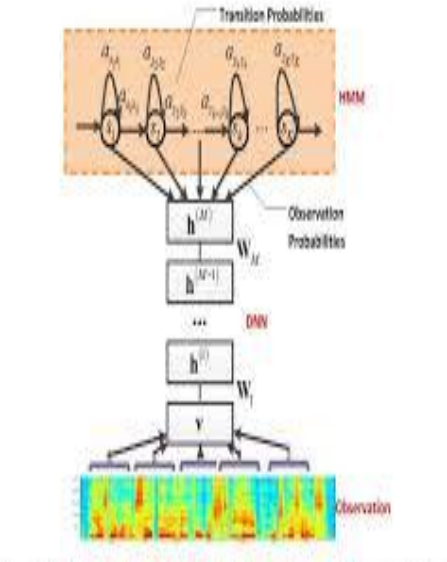

Fig 5 Pre processing flow diagram

III. PROPOSED ALGORITHM

This paper proposes an algorithm which recognizes the speech with greater accuracy. The detailed flow is shown in the figure 5

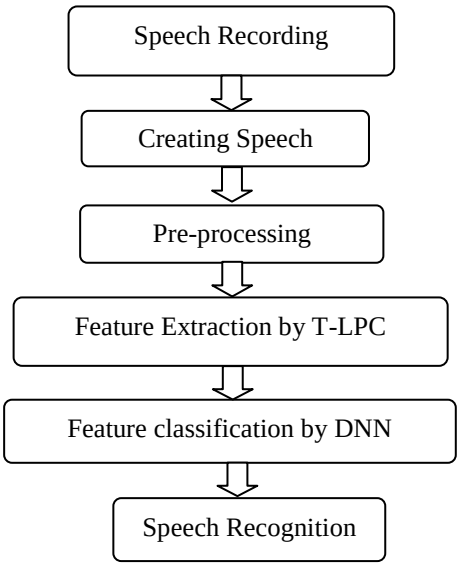

Fig 6 Proposed algorithm

IV. RESULTS AND DISCUSSION

The overall results are obtained starting by recording the speech. Actually the database is created with 500 speeches. The most famous dialog is spoken and recorded by 500 speakers in which 250 male speakers and 250 female speakers are considered. The signals are recorded in wave format and later converted to data sequence.

The data is preprocessed by two stage DNN technique where the framing, frame blocking, de-noising etc. are done. After pre-processing features are extracted by using T-LPC technique as shown below. Teager energy of male and female is shown in the figure 7 extracted various features using TLPC are shown.

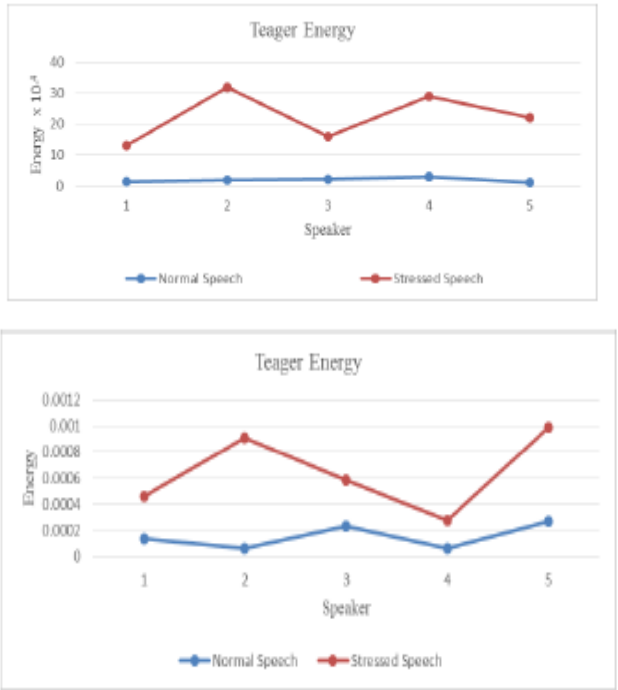

Fig 7 The Continuous speech Recognition

TABLE 1: Features extracted by TLPC

Features by TLPC	No Features
TLPC coefficients	12
Energy	1
Mean	1
Entropy of energy	1
Spectral centroid	1
Spectral spread	12
Spectral entropy	1
Spectral flux	1
Spectral roll off	1
Spectral density	1
Fundamental frequency	12
ZCR	1
Peak amplitude	1
Standard deviation	1
Total	47

The extracted features are classified by using DNN and recognition accuracy is 93.39%. recording can be seen in figure 8 and accuracy is seen in figure 9

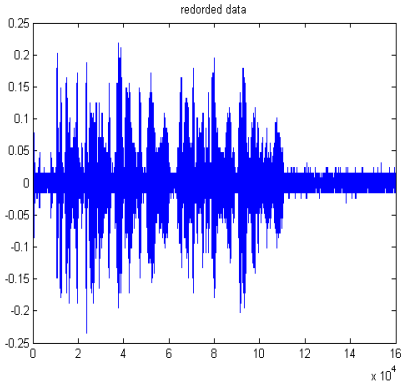

Fig 8 Recorded continuous speech

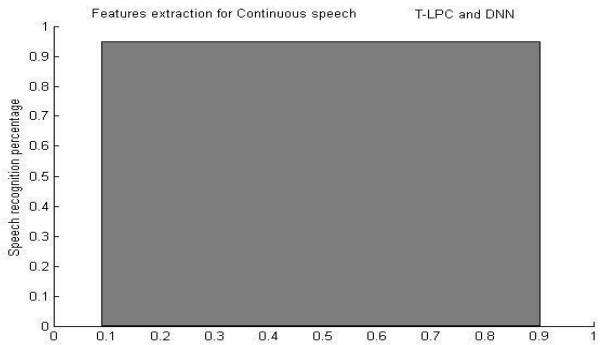

Fig 9 Continuous speech Recognition accuracy

V. CONCLUSION

To conclude this paper discussion of the flow is required. The overall recognition accuracy for continuous speech is calculated with excellent accuracy of 96.39%. The technique used for feature extraction is very much suitable, this technique TLPC is rarely used for Telugu continuous speech. This feature helps to give more accuracy in the classification. The DNN classifier is used for best accuracy. Two stage DNN also used in pre-processing which helps for good results in continuous Telugu speech recognition.

REFERENCES

1. A. K. Yadav, R. Roy, R. Kumar, C. S. Kumar and A. P. Kumar, "Algorithm for de-noising of color images based on median filter," *2015 Third International Conference on Image Information Processing (ICIIP)*, Waknaghat, 2015, pp. 428-432.
2. A. P. Kumar, N. Kumar, C. S. Kumar, A. K. Yadav and A. Sharma, "Speech recognition using arithmetic coding and MFCC for Telugu language," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2016, pp. 265-268.
3. A. P. Kumar, N. Kumar, C. S. Kumar, A. K. Yadav, "Speech compression by adap tive Huffman coding using Vitter algorithm", *International Journal of Innovative Sciences*, vol. 2, No. 5, May2015.
4. A. P. Kumar, R. Roy, S. Rawat, A. Sharma, "Telugu speech feature extraction by MODGDF and MFCC using Naïve Bayes classifier", *International Journal of Control Theory and Applications*, vol. 9, No. 21, Dec 2016.
5. A. K. Yadav, R. Roy, R. Kumar, and A. P. Kumar, "Survey on Content based image retrieval and texture applications", *International Journal of signal processing, image processing and pattern recognition.*
6. Kumar A.P., Roy R., Rawat S., Yadav A.K., Chaurasia A., Gupta R.K. (2018) Telugu Speech Recognition Using Combined MFCC, MODGDFFeature Extraction Techniques and MLP, TLRN Classifiers. In: Pant M., Ray K., Sharma T., Rawat S., Bandyopadhyay A. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 584. Springer, Singapore
7. VK Sharma, AP Kumar , "Continuous telugu speech recognition by joint feature extraction of mfcc, modgdf and dwpd techniques by pnn classifier", International Journal of Pure and Applied Mathematics,Vol. 118, No. 21, pp. 865-872, 2018
8. Kumar A.P., Roy R., Rawat S., Chaturvedi R., Sharma A., Kumar C.S. (2018) Speech Recognition with Combined MFCC, MODGDF and ZCPA Features Extraction Techniques Using NTN and MNTN Conventional Classifiers for Telugu Language. In: Pant M., Ray K., Sharma T., Rawat S., Bandyopadhyay A. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 584. Springer, Singapore
9. AP Kumar, R Roy, S Rawat, P Sudhakaran, "Continuous Telugu Speech Recognition through Combined Feature Extraction by MFCC and DWPD Using HMM based DNN Techniques", ", International Journal of Pure and Applied Mathematics,Vol. 114, No. 11, pp. 187-197, 2017