International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Telugu Speech Recognition on LSF and DNN Techniques

Y.Sangeetha, Archek Praveen Kumar, Neerudu Uma Maheshwari, Rodda Srinivas, P. Jyothi

Abstract— This fast world is running with machine and human interaction. This kind of interaction is not an easy task. For proper interaction between human and machine speech recognition is major area where the machine should understand the speech properly to perform the tasks. So ASR have been developed which improvised the HMIS ("Human Machine Interaction systems") technology in to the deep level. This research focuses on speech recognition over "Telugu language", which is used in Telugu HMI systems. This paper uses LSF (linear spectral frequencies) technique for feature extraction and DNN for feature classification which finally produced the effective results. Many other recognition systems also used these techniques but for Telugu language this are the most suitable techniques.

Keywords: speech recognition, Telugu language, LSF, DNN..

I. INTRODUCTION

Automatic speech recognition system. Speech signal is 1 dimensional signal. this is an dependent signal and depended on time. For every communication speech processing is the main step [1]. Similarly for the communication between human and computer or machine speech processing plays vital role. Storage is one of the important problem where the technology is facing. So instead of storage the transmitting the speech is further easy.

Speech signals will corrupt easily, so proper algorithms with powerful techniques are required for the design. All the speech signals are in wave format after recording [2]. This wave signals are converted to analog signals for further processing. Since analog signals are not processed easily the analog signals are converted to digital and processed. Due to the digital signal processors available in the market very easily the signals are processed. There are few steps to be followed for speech recognition

Manuscript published on November 30, 2019.
* Correspondence Author
Y. Sangeetha*, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.
Dr. Archek Praveen Kumar, Professor, HOD, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.
Neerudu Uma Maheshwari, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.
Rodda Srinivas Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.
P. Jyothi, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.

- Record the speech
- Speech pre-processing
- Feature extraction
- Feature classification

When these four steps are done perfectly the speech is recognized [3]. But the question is what techniques to be used. The user should think about many factors before choosing the technique.

- What language
- How many vowels
- How many alphabets
- Words to recognize
- Emotion
- Time alignment
- Gender
- Age of the speaker

By considering all these factors for pre-processing 2 stage DNN is the better technique and for feature extraction LSF is the suitable technique and for feature classification DNN is the best applicable technique [4]. This research shows what the results are by applying these techniques

II. BLOCK DESCRIPTION

This basic steps need to be followed are shown in figure 1. The speech signals are recognized for Telugu language which is a Dravidian language.

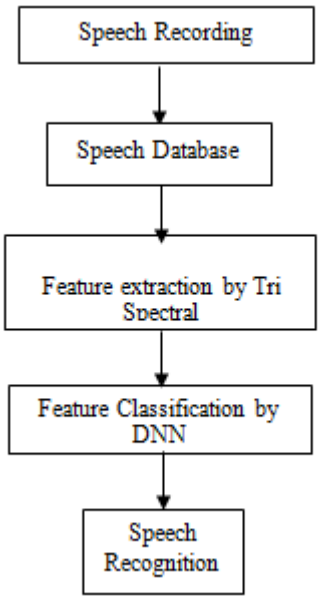

Fig 1 Block diagram

Retrieval Number: D5257118419/2019©BEIESP
DOI:10.35940/ijrte.D5257.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

As shown in figure speech is recorded with a perfect microphone which generates minimal noise [5].

A. Pre-Processing

Speech recognition can be improvised by using this technique. This works on Mel scale which is the cepstral domain obtained by using DNN. This actually provides the ideal binary mask [6]. DNN deals with the advanced ceptstral features which are classified later. The pre-processing using DNN generates a denoised speech signal.

The vocal chords vibration produces a periodic sounds caused by forcing air in the vocal tract. Acoustic tube. Speech –speech pressure waves. The technique is shown in equation (1). The speech signals are easily de-noised by using this technique. The sub bands are also shown in figure 2.

$$W=w^{max}\{p(\frac{w}{0})\} \quad (1)$$

The maximum values are replicated through the domain which is easily featured with various parameters. This makes the user to simulate the code.

Pre-processing is a combination of few steps starting with framing followed by blocking, windowing. The signals which are windowed performs time alignment and finally De-noise is done as shown in figure 3

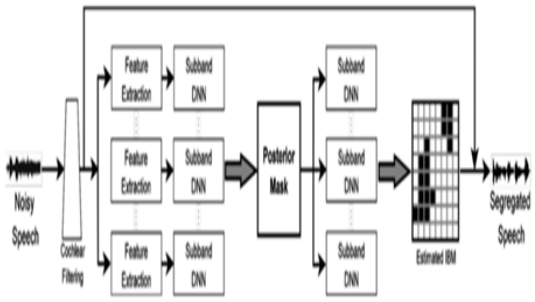

Fig 2 Block diagram

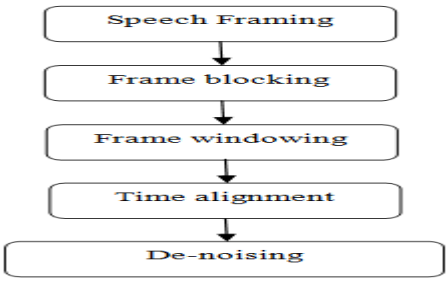

Fig 3 Flow chart for pre-processing

B. Feature Extraction

This ASR uses LSF technique for extraction of parameters. LSF is linear spectral frequencies.

LSP is an important technology for speech synthesis and coding [7]. Speech coding is used as an application for compression of digital audio signals that contain speech. Important applications of LSP are:

1: Mobile telephony

2:Voice over internet protocol (VOIP)

LSP decomposition is used in speech coding for quantification of LP parameters which represent spectral envelop of a signal. It is widely used due to its robustness to quantification and guarantees model stability. We use linear prediction polynomials to represent the linear prediction Co efficients [LPC] We use 31 bit representation instead of 41 bit representation for better outcomes.

Now LP polynomial is given by equation (2)

$$A(z)=0.5\,[P(z)+Q(z) \quad (2)$$

Here P=palindrome Q=anti palindrome P(z) is the response when glottis is closed Q(z) is the response when glottis is opened. By solving the polynomial equations, we get roots of P and Q 1. Roots of P and Q lie in unit circle in a complex plane 2 [8]. Roots of P which do not include roots of Q travel around the unit circle 3. As the coefficients are real, roots occur in conjugate pairs To concert LSP back to LPC we evaluate A(z) by clicking an impulse through it at an Order of N times.

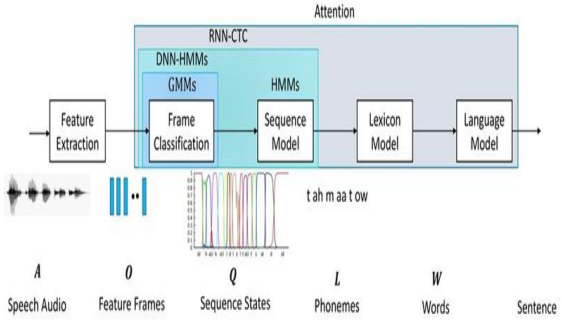

Fig 4 DNN Diagram

C. Feature Classification

Usually having two or more hidden layers counts as deep. But any network with only a single hidden layer is called 'Shallow'. Recurrent neural network is the bad boy that can best process speech. DNN is a network which works on a forward feed of ANN. DNN have excess of layers than ANN. At least two layers are present as a hidden layer between outputs and inputs as shown in figure 4 [9].

III. PROPOSED ALGORITHM

The proposed algorithm is shown in the figure 5 which briefs all the techniques used for this research.

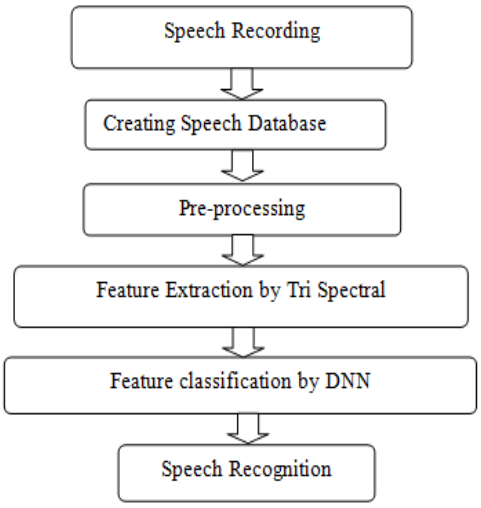

Fig 5 Proposed algorithm

Retrieval Number: D5257118419/2019©BEIESP
DOI:10.35940/ijrte.D5257.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IV. RESULTS AND DISCUSSION

Speech signal recorded at 8kHz frequency is sampled with 16kHz frequency and converted to discrete signal. that discrete signal is framed as 20 frames per second and each frame is processed. The recorded speech is shown in figure 6

Speech Recording

Database is created for 2000 words where 20 word are spoken by 100 speakers where the male and female ratio is same. For example one speech word called "MAMA" recorded graph is shown in figure 6.

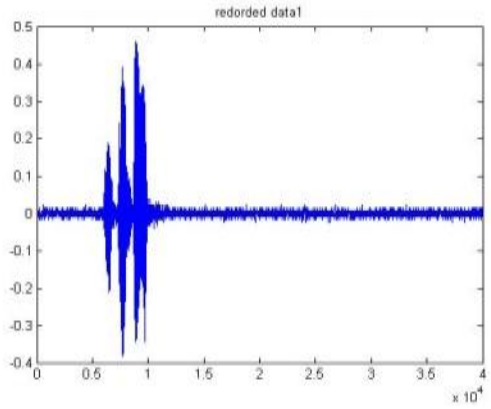

Fig 6 Recorded Speech

Features Extracted

The spectral features or parameters extracted are shown in table 1 and figure 7. Each parameters are extracted with respect to spectral domain. Total 36 features are extracted.

Table 1. Recorded Speech

Features by TLPC	No Features
LFS coefficients	12
Energy	1
Mean	2
Entropy of energy	1
Spectral density	4
Fundamental frequency	12
ZCR	1
Peak amplitude	1
Standard deviation	1
Total	36

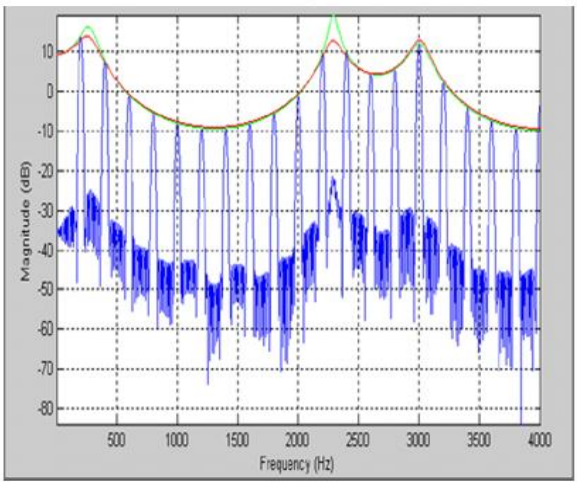

Fig 7 Features Extracted

Features Classification

DNN is used and classified with a vector space and the recognition accuracy is 95.89%

V. CONCLUSION

This research shows automatic speech recognition for Telugu language. This paper uses, two stage DNN for preprocessing, LSF for feature extraction and DNN for feature classification. Using this techniques promising result is obtained with a recognition accuracy of 95.89%. Any how the results may be improved if different combinations are used for the 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, 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.
4. 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.
5. Kumar A.P., Roy R., Rawat S., Yadav A.K., Chaurasia A., Gupta R.K. (2018) Telugu Speech Recognition Using Combined MFCC, MODGDF Feature 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
6. 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
7. 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
8. 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