

Telugu Speech Recognition on TRI-SPECTRAL and DNN Techniques



Chunchu Raj Kumar, Archek Praveen Kumar, B Sheha Priya, Affrose, A. Haseena

Abstract— This Research focus on the recognition of speech signals for Telugu language. The data of Telugu language considered is in isolated format. 10 isolated words are considered which are frequently spoken and recognized. Advanced technique named Tri spectral technique and DNN is used for this recognition. Tri spectral is a feature extraction technique. DNN is a feature classification technique. This research can be used in many interfacing systems which helps the humans to interact with the hardware or software systems easily. Design of ASR (“Automatic Speech Recognition System”) deals with many parameters which should finally conclude with promising recognition results. This techniques used in this research has given a better result with the accuracy of approximately 96.27%.

Keywords: speech recognition, Telugu language, Tri spectral, DNN.

I. INTRODUCTION

As the technology is getting improved day by day, speech recognition plays a vital role in every aspect. For example a human not even want to type the date for searching and they want to speak and automatically the data should be searched immediately. Similarly many areas automation is used like automatic cars, automatic clothes, automatic home appliances, automatic shoes, automatic lifts etc. where ever the automation is there speech recognition or speech processing follows [1]. Previously we used to enter the codes or give the commands by typing but not its totally automation, the user speaks and the output is generated. ASR started in 1970’s but the actual interaction with this topic is started in 1990’s after the digital technology is evolved and got in to practice.

Firstly they used telepathy for recognition but now many and many algorithms are developed for greater recognition

accuracy. Speech recognition is not that easy since every individual person have unique identity. So for this much population, equating through proper links and obtaining the greater accuracy is a huge task. The sound of individual also changes with their emotion. The emotions of the female speech variations are higher than the male speech variations. Speech processing or speech recognition is a branch of signal processing. Speech recognition basically consists of four steps

- Pre-processing
- Feature extraction
- Feature reduction
- Feature classification

If all these four steps are perfectly done, then the accuracy will be high. There are hundreds of techniques used in these four steps [2]. Suitable techniques should be chosen with respect to the language. One dimensional signal is a speech signal. There are 3 different types of speeches

- Alphabets
- Isolated words
- Connected words
- Continuous words

Alphabets are single letters. Alphabets are totally 52 in Telugu. Isolated words are individual words which are spoken very frequently, connected words are small sentences spoken very regularly and the last is continuous words which is a paragraph which is combination of few sentences. This research is done on the isolated words.

II. BLOCK DESCRIPTION

This Research is done for isolated words of Telugu language the block diagram shown in fig 1 explains in detail about how the speech recognition is done.

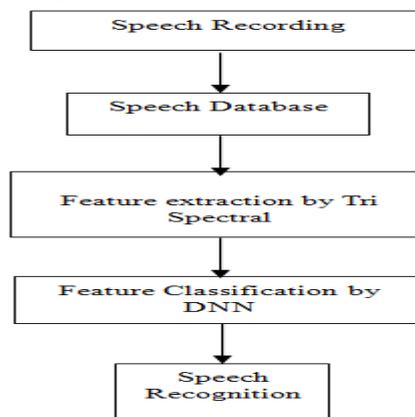


Fig 1 :Block diagram

Manuscript published on November 30, 2019.

* Correspondence Author

Chunchu Raj Kumar*, 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.

B Sheha Priya Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.

Affrose, Assistant Professor, Department of ECE, Malla Reddy College of Engineering for Women, Hyderabad, Telangana, India.

A. Haseena, Assistant Professor, Department of ECE, Amity University, Jaipur, Rajasthan, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Firstly the speech is recorded [3]. The speech should be recorded carefully since the noise corruption is more in this case. There is chance of thermal noise generation so the recording should be done in a quite environment. The recorded speech is saved as a .wave format. Later the wave format is converted to analog signal and again to digital for easy processing.

The speech signal is the signal which is time dependent, always varies with respect to time. Now the recorded speech is stored in a place called as database. The database is created for 1000 words with various gender and age group. Next important task is to pre-process the data.

A. Pre-Processing

Isolated data which is recorded consists of various types of internal and external noises. So pre-processing is required [4]. Preprocessing is done by two stage deep neural network. This is similar to artificial neural network. Two stages are considered as shown in figure 2. Two stages provides more advantages compared to single stage. Two stages are shown as

- First stage removes the noise
- Second stage removes the unwanted echos

Echos are also a major problem in preprocessing signals. For isolated data the two stages can solve easily the representation of noise removal.

Many algorithms are developed for pre-processing like HMM, DWT, DCT, ANN etc, but the deep neural networks plays a vital role in the removal of noise. The sub bands are differentiated among the speech signals. All the sub bands are independent to each other [5]. This technique not only preprocesses the signals but also do the enhancement of the speech signals.

There are many numbers of layers which are interconnected with neurons as a network. The advanced version of ANN is DNN. This technique is based on the pipeline process. The discriminations of the bands are processed according to the time.

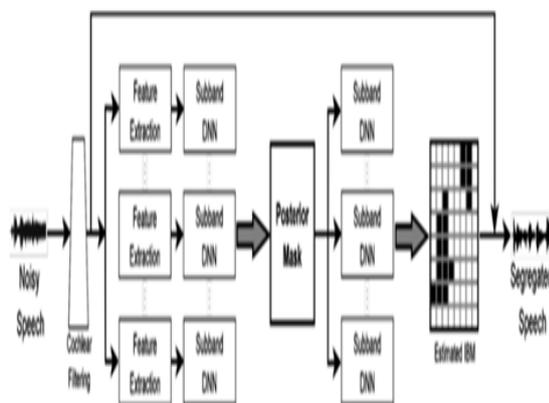


Fig 2 Block diagram

Preprocessing inbuilt process is shown in figure 3. as the signal is recorded and database is generated then framing is done where the speech signals are broken in to the number of frames which makes the user for easy process.

After framing the next procedure is frame blocking which

is done by endpoint detection. The end point detection is done by ZCR process which is zero crossing rate [6]. Later hamming window is used for windowing the signals. Time alignment is one of the important process to de-noise the speech signals.

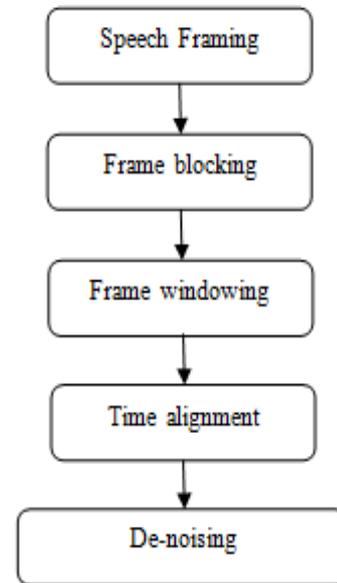


Fig 3 Pre processing flow diagram

B. Feature Extraction

This research uses Tri spectral technique for the extraction of the speech features. This is one of the most powerful techniques which is used in automatic speech recognition systems [7]. After preprocessing the feature extraction is made. The Tri spectral equations are shown in the equation 1 and 2. Generally, the speech signals are recorded and then the tri-spectrum features are analyzed.

C. Feature Classification

Deep Neural network is a neural network. It is a technology which built the simulation of the activity of the human brain [8]. Deep Neural Network specifically simulates the pattern recognition and passage of input through various layers of neural connections.

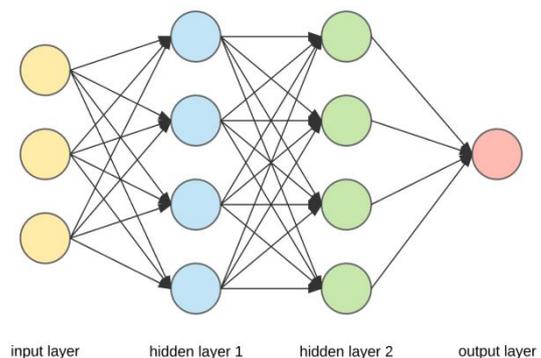


Fig 4 DNN flow diagram

It contains an input layer, an output layer and at least one hidden layer in between them as shown in the figure 4. The extracted features are classified by using DNN technique which is more suitable for Telugu language [9].

III. PROPOSED ALGORITHM

The detailed proposed algorithm is shown in the figure 5 with what techniques are used. The data to be recognized should follow all these methods mentioned in the algorithm.

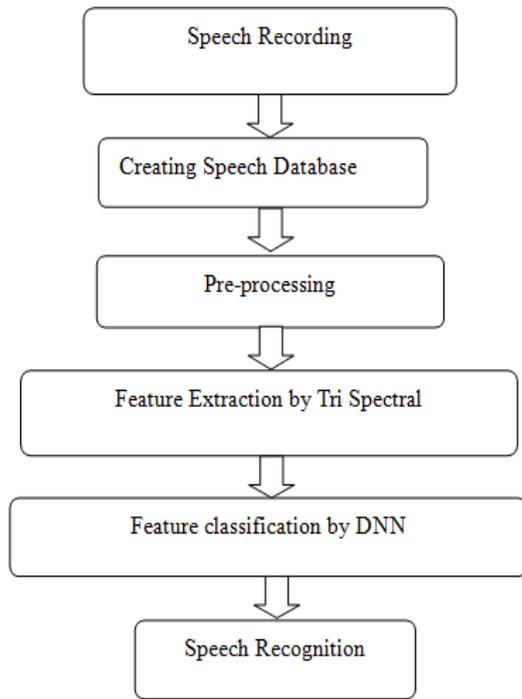


Fig 5 Proposed algorithm

IV RESULTS AND DISCUSSION

The overall results are obtained by using MATLAB software tool. The data is first recorded and database is created which is preprocessed and de-noised. Then the features are extracted and classified for isolated words which are very frequently spoken. The graphs with detailed results are shown step by step

Speech Recording

Totally 10 words are spoken by 50 male and 50 female with the variations of age from 15 years to 50 years and database is created. For example one speech word called “navvu” recorded graph is shown in figure 6.

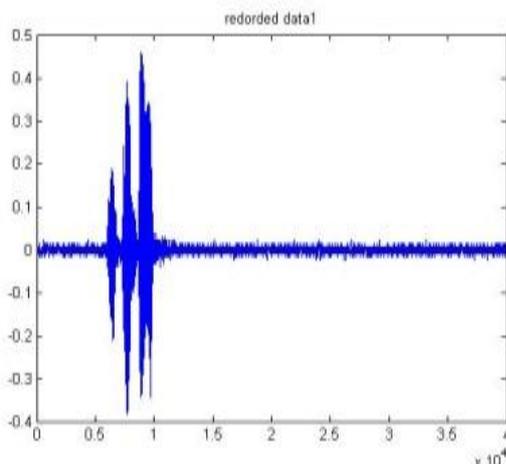


Fig 6 Recorded Speech

Features Extracted

The recorded speech is preprocessed and noise is removed and features are extracted. The tri spectral features extracted are shown in figure 7. There are two graphs which are Tri state with Mel filter bank and without Mel filter bank.

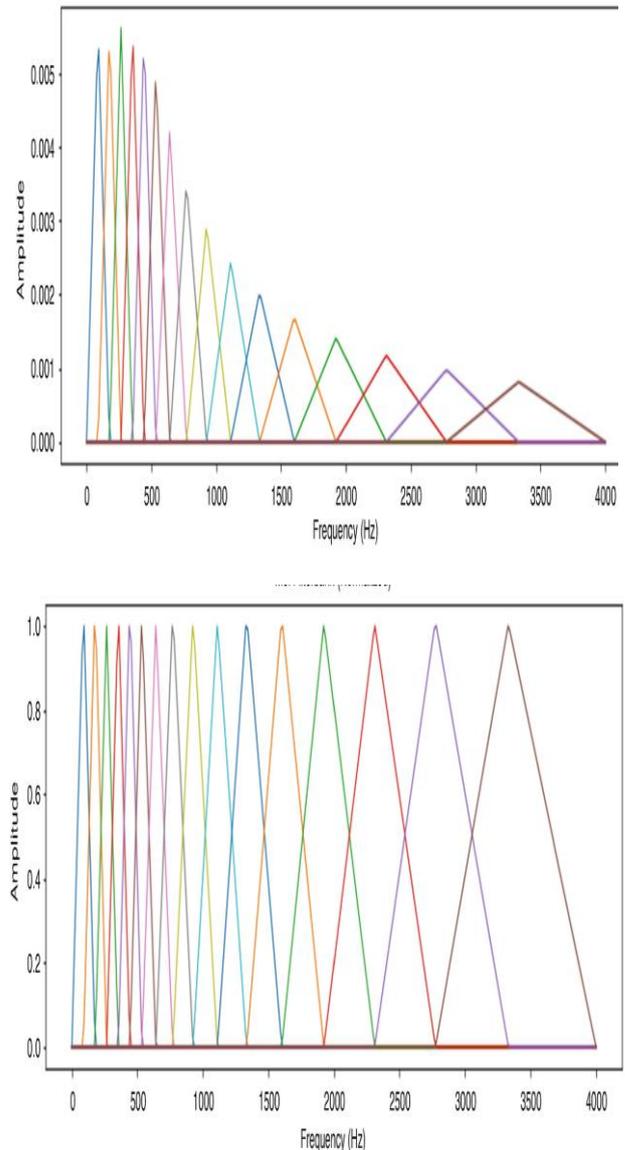


Fig 7 Features Extracted

Features Classification

DNN is used and classified with a vector space and the recognition accuracy is 96.27%

IV CONCLUSION

As per the research discussion, the isolated 10 words which are recognized with a promising percentage of 96.27 which can further improved by changing the feature extraction or classification techniques. This research uses 1000 words databases which is created by user. To conclude finally Telugu speech recognition can be done by Tri spectral and DNN techniques.



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), Wagnaghat, 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