International Institution for Technological Research and Development

Volume 1, Issue 3, 2016

Review on Speech Recognition using Neuro-fuzzy system

Aksha D. Patel*,Vimal Nayak**
*(Department of E.C., Silver oak college of Engineering and Technology, India
Email: paksha911@gmail.com)
** (Department of E.C., Silver oak college of Engineering and Technology, India
Email: vimalnayak.ec@socet.edu.in)

Abstract:- Speech recognition system is used to extract the words from the input speech and tries to match it with the database stored for a particular speaker. There are many popular algorithms to recognize the human voice. However, a literature survey of existing algorithms shows that none of the present day algorithm results in a good recognition accuracy and also unable to cope with background noise. We proposed a speech recognition system based on neuro-fuzzy logic which combine pros and cons of neural networks and fuzzy logic. Several experiments can be use to verify the performance of the Neuro-fuzzy system to recognize the human voice. English language record in a different environment, syllable and pronunciations are use as a data set in the experiments.

Keywords: English language, Artificial neural network, fuzzy logic , neuro-fuzzy technique

1. INTRODUCTION

Speech processing is the study of speech signals, and the various methods which are used to process them. In this process various applications such as speech coding, speech synthesis, speech recognition and speaker recognition technologies; speech processing is employed . Among the above, speech recognition is the most important one. The main purpose of speech recognition is to convert the acoustic signal obtained from a microphone or a telephone to generate a set of words . In order to extract and determine the linguistic information conveyed by a speech wave we have to employ computers or electronic circuits . This process is performed for several applications such as security device, household appliances, cellular phones ATM machines and computers .

The speech is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. So, it is only logical that the next technological development to be natural language speech recognition. Speech Recognition can be defined as the process of converting speech signal to a sequence of
words by means Algorithm implemented as a computer program. Speech processing is one of the exciting areas of signal processing.The goal of speech recognition area is to developed technique and system to developed for speech input to machine based on major advanced in statically modelling of speech ,automatic speech recognition today find widespread application in task that require human machine interface such as automatic call processing. Since the 1960s computer scientists have been researching ways and means to make computers able to record interpret and understand human speech. Throughout the decades this has been a daunting task.

In classification, there are many algorithms which have high accuracy such as the results from . The Neuro-Fuzzy system is the algorithms which having high efficiency and accuracy because it combines strong point from Neural Network and Fuzzy Logic. The structure of Neuro-Fuzzy system proposed in results the best accuracy over other algorithms of infant cries recognition. This work uses the same structure of the Neuro-Fuzzy system to classify the English language machine command data set. In this work, 3 factors are considered, including the types of noise, syllable, and pronunciation to test the robustness of the algorithm.

The paper aims to describe proposed method in detail Section II shows the survey done to know the work done and history. Section III shows different techniques for Speech Recognition. Section IV describes proposed method in detail.

2. LITERATURE SURVEY

It shows an investigation of the speech recognition classification performance and speech features (spectrogram and cepstrum) will be sequentially presented at neural network inputs and be classified at the output of the network. [1]

It shows the proposed techniques designed to increase the noise robustness of the systems, a series of experiments were performed using the 7-layer DNN with 3202 senones and FBANK features . It is investigate the noise robustness performance of DNN-based acoustic models and propose three methods to improve accuracy. [2]

It shows the comparative analysis of HMM and ISRS with two, three and four parameters for speech recognition. With proper SNR in process of training and testing. ISRS improve performance Increase the speech bank with different accents Genetic algorithm to 100% recognition in the training process. Use nonlinear predictive coding for feature extraction in SR.[3]

It shows the Result of various syllable data ,Continuous and easy control noise , Normal pronunciation data , Various pronunciation data, Loud and unpredictable noise. Neuro-fuzzy system proposed in result the best accuracy over other algorithm of infant cries algorithm. [4]

3. DIFFERENT TECHNIQUES FOR SPEECH RECOGNITION

There are several techniques that use in speech recognition. There are three existing techniques neural network , hidden markov model , dynamic time rapping respectively.

A. Dynamic Time Warping (DTW)

This technique compares words with reference words. Every reference word has a set of spectra; but there is no distinction between separate sounds in the word. Because a word can be pronounced at different speeds, a time normalization will be necessary. Dynamic Time Warping is a programming technique where the time dimension of the unknown word is changed (stretched and shrunk) until there is a similarity with a reference word.

B. Hidden Markov Modeling (HMM)

Until now, this is the most successful and most used pattern recognition method for speech recognition. It's a mathematical model derived from a Markov Model. Speech recognition uses a slightly adapted Markov Model. Speech is split into the smallest audible entities (not only vowels and consonants but also conjugated sound like ou, ea, eu,...). All these entities are represented as states in the Markov Model. As a word enters the Hidden Markov Model it is compared to the best suited model (entity). According to transition probabilities there exist a transition from one state to another. For example: the probability of a word starting with xq is almost zero. A state can also have a transition to it's own if the sound repeats itself. Markov Models seems to perform quite well in noisy environments because every sound entity is treated separately. If a sound entity is lost in the noise, the model might be able to guess that entity based on the probability of going from one sound entity to another.

C. Neural Network

Neural networks have many similarities with Markov models. Both are statistical models which are represented as graphs. Where Markov models use probabilities for state transitions, neural networks use connection strengths and functions. A key difference is that neural networks are fundamentally parallel while Markov chains are serial. Frequencies in speech, occur in parallel, while syllable series and words are essentially serial. This means that both techniques are very powerful in a different context. As in the neural network, the challenge is to set the appropriate weights of the connection, the Markov model challenge is finding the appropriate transition and observation probabilities. In many speech recognition systems, both techniques are implemented together and work in a symbiotic relationship. Neural networks perform very well at learning phoneme probability from highly parallel audio input, while Markov models can use the phoneme observation probabilities that neural networks provide to produce the likeliest phoneme sequence or word. This is at the core of a hybrid approach to natural language understanding. In this paper speech features (spectrogram and cepstrum) will be sequentially presented at neural network inputs and will be classified at the output of the network.

Fuzzy Logic:-

Fuzzy logic can be used as an interpretation model for the properties of neural networks, as well as for

giving a more precise description of their performance. We will show that fuzzy operators can be conceived as generalized output functions of computing units. Fuzzy logic can also be used to specify networks directly without having to apply a learning algorithm. An expert in a certain field can sometimes produce a simple set of control rules for a dynamical system with less effort than the work involved in training a neural network.

4. NEURO-FUZZY SYSTEM

Fuzzy logic and neural networks are natural complementary tools in building intelligent systems. While neural networks are low-level computational structures that perform well when dealing with raw data, fuzzy logic deals with reasoning on a higher level, using linguistic information acquired from domain experts. However, fuzzy systems lack the ability to learn and cannot adjust themselves to a new environment. On the other hand, although neural networks can learn, they are$_i$ opaque to the user. Integrated neuro-fuzzy systems can combine the parallel computation and learning abilities of neural networks with the human-like knowledge representation and explanation abilities of fuzzy systems. As a result, neural networks become more transparent, while fuzzy systems become capable of learning**.** A neuro-fuzzy system is a neural network which is functionally equivalent to a fuzzy inference model. Neuro-Fuzzy System can be trained to develop IF-THEN fuzzy rules and determine membership functions for input and output variables of the system. Expert knowledge can be incorporated into the structure of the neuro-fuzzy system. At the same time, the connectionist structure avoids fuzzy inference, which entails a substantial computational burden. The structure of a neuro-fuzzy system is similar to a multi-layer neural network. In general, a neuro-fuzzy system has: input and output layers, and three hidden layers that represent membership functions and fuzzy rules.

In neuro-fuzzy logic there are five layers. There are input layer and output layer and three hidden layers namely membership function and fuzzy rules.

1) Layer 1:- Input layer

In this layer each neuron transmits external crips signals directly to next layer. That is

$$Y_i^{(1)} = x_i^{(1)}$$

2) Layer 2:- Fuzzification layer

In this layer neurons represent fuzzy sets used in antecedents of fuzzy rules. A fuzzification neuron receives a crisp input and determines the degree to which this input belongs to the neurons fuzzy set.

3) Layer 3 :- Fuzzy rule layer

In this layer each neuron corresponds to a single fuzzy rule. A fuzzy rule neuron receives inputs from the fuzzification neurons that represents fuzzy sets in the rule antecedents.

4) Layer 4:- Membership layer

In this neurons in this layer represents fuzzy sets used in the consequent of fuzzy rules. An output membership neuron combines all its inputs by using the fuzzy operation union. This operation can be implemented by the probabilistic.

5) Layer 5:- Defuzzification

In this layer each neuron represents a single output of the neuro-fuzzy system. It takes the outputs fuzzy sets clipped by the respective integrated firing strengths and combines them into a single fuzzy set.

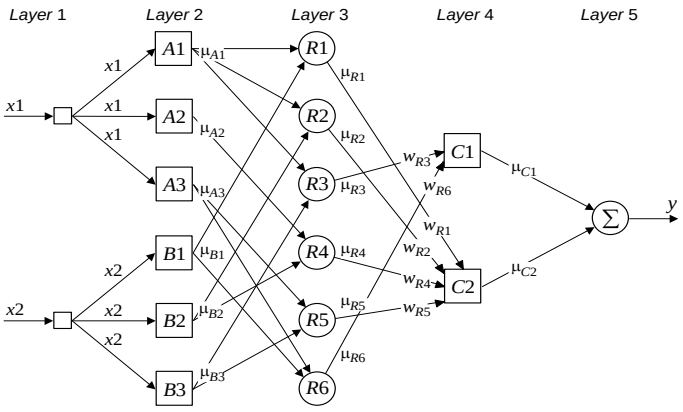

fig.1 Neuro-fuzzy system

5. RESULT AND CONCLUSION

The figure shows that the simulation result of basic speech signal. In which they are separated with different regions. They are namely voiced region, unvoiced region and silent region. It is the basic speech wave signal which is defined by used MATLAB.

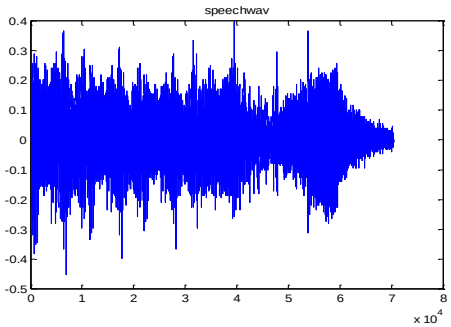

Fig. 2 Speechwave signal

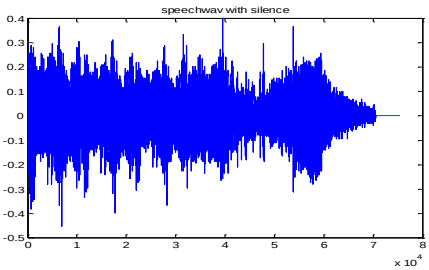

Fig. 3 Speechwave with silence

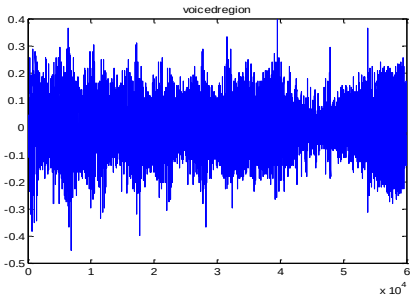

Fig.4 Voiced region

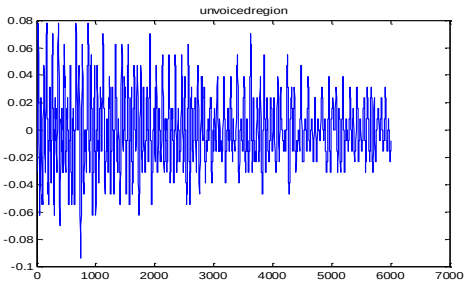

Fig. 5 Unvoiced region

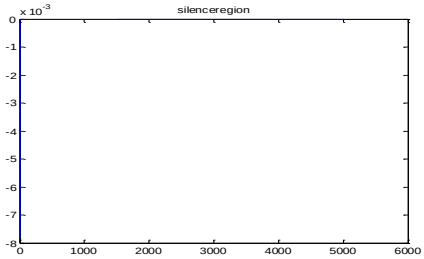

Fig. 6 Silent region

The results of speech recognition using neural network with different letters convertor shows.
Here as we see the sound wave of letters ‘A’ , ‘E’ and ‘O’ are display in figures which are shown as below. We give the trained input which compare with this letters sound wave and gave the output of trained network. And also see as in figure the input wave of trained sound and output of trained network seen .

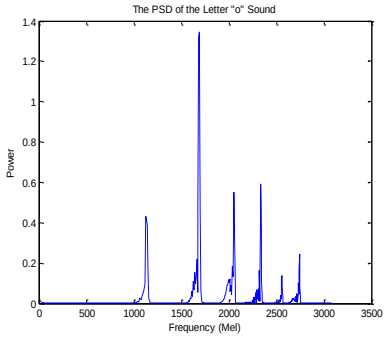

Fig 7 ‘A’ Letter Sound wave

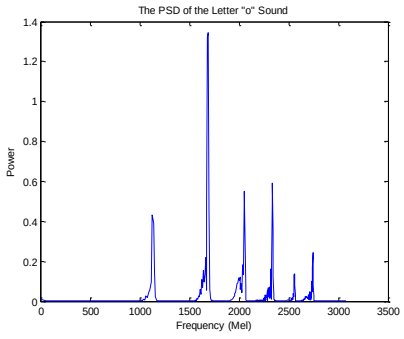

Fig 8 'O' Letter sound wave

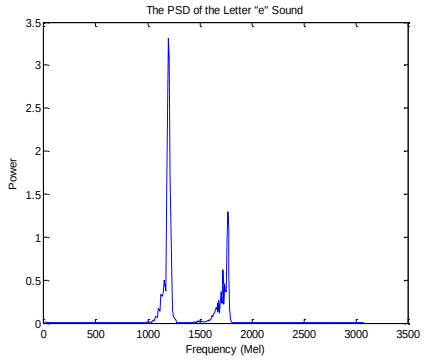

Fig. 9 'E' letter sound wave

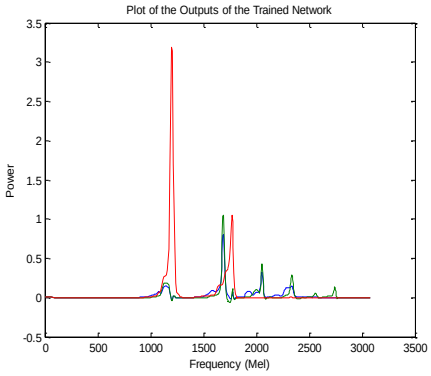

Fig 10 Output of trained network

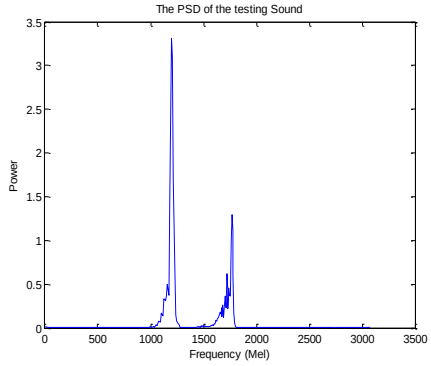

Fig 11 the testing sound wave

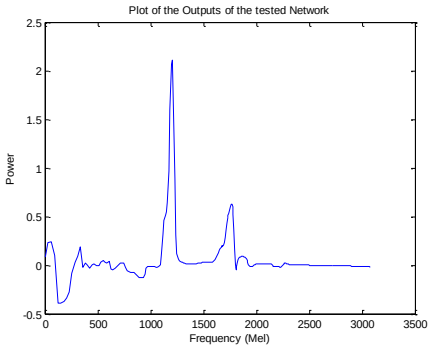

Fig 12 output of the testing network

6. CONCLUSION

The existing literature shows that till date the Speech recognition was done with the help of various methods such as DNN,ANN , RNN, HMM and many more. Fuzzy logic efficiently handles uncertainity and it will increase accuracy. Various papers are surveyed for the method and to know the work done and history. There are some advantages and disadvantages over the neural network so from this conclude the new technique which is the combination of neural network and fuzzy logic is called "neuro-fuzzy". The neural network can be very powerful in classifying speech signals. Find the parameters of speech signal wave from the simulation result. Which can enhance the speech recognition process from this technique. Here the speech recognition using neural network is performed.

REFERENCES

[1] Georgi T. Tsenov, and Valeri M. Mladenov, Senior Member, IEEE " Speech recognition using Neural Network", 2010 IEEE

[2] Michael L. Seltzer, Dong Yu, Yongqiang Wang "An Investigation of Deep Neural Networks for Noise Robust Speech Recognition", 2013 IEEE

[3] Washington Luis Santos Silva, Ginalber Luiz de Oliveira Serra " Proposal of an intelligent speech recognition system ", 2012 IEEE

[4] Krittakom Srijiranon , Narissara Eiamkanitchat "Thai speech recognition using neuro-fuzzy",2015 IEEE

[5] C. Y. Fook, M. Hariharan, Sazali Yaacob, Adom "Malay Speech Recognition and Audio Visual Speech Recognition",2012 IEEE

[6] Amane TALEB "Speech Recognition by Fuzzy-Neuro ANFIS Network and Genetic Algorithms", International Conference on Intelligent Computational Systems (ICICS'2012) Jan. 7-8, 2012 Dubai

[7] S.Matsunaga, M. Yamaguchi, K. Yamauchi, and M. Yamashita, "Sound Source Detection using Multiple Noise Models ," in Acoustics, Speech and Signal Processing 2008 (ICASSP 2008), pp. 2025 - 2028, USA, April 2008.

[8] Joao P. Carvalho, Fernando Batista, Luisa Coheur "A Critical Survey on the use of Fuzzy Sets in Speech
and Natural Language Processing",2012 IEEE

[9] X. Lv, M. Zhang, and H. Li, " Robot Control Based on Voice Command," in, 2008 IEEE International Conference on Automation and Logistics (ICAL 2008), pp. 2490 - 2494, China, September 2008.

[10] J.H. McClellan, R.W. Schafer, M.A. Yoder, "Signal Processing First", Prentice Hall, 2003, pp. 415-426.

[11] Dou-Suk Kim and Soo-Young Lee, "Intelligent judge neural network for speech recognition", Neural Processing Letters , Vol 1.

[12] Chee Peng Lim, Siew Chan Woo, Aun Sim Loh, Rohaizan Osman,"Speech Recognition Using Artificial Neural Networks," wise, vol. 1,pp.0419, *First International Conference on Web Information Systems Engineering* (WISE'00)-Volume 1, 2000.