

Automatic Speech Recognition using Multinet Neural Network

¹Akshay Kumar, ²Parisha Sharma, ³Vishal Koul

¹Dept. of E&TC, SPPU, DYPIET, Pune, India

²Dept. of E&TC, SPPU, DYPIET, Pune, India

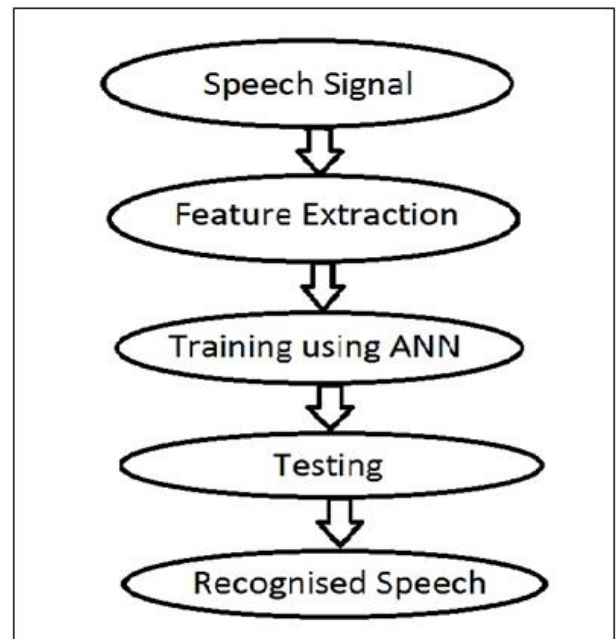
³Dept. of E&TC, SPPU, DYPIET, Pune, India

Abstract:- The work of an Automatic Speech Recognition system is to convert speech signal into a text message accurately. Transformation of the spoken words independent of the speaker, we need a microphone device to record the speech and speaker decides what to say and actually speaks a word or sentence. The Matrix laboratory (MATLAB) software then produces a speech wave form. In this project, artificial neural networks are used to do speech recognition. The paper will be investigated in two steps, consisting of the pre-processing part and the post-processing part with Artificial Neural Networks (ANN). These two parts are explained in detail and speech recognizers using different ANN architectures will be implemented on Matlab. Speech processing is one of the most important application area in the field of digital signal processing. Different fields for research in speech processing are speaker recognition, speech recognition and speech coding. In this project, the Mel Frequency Cepstrum Coefficient (MFCC) feature has taken for constructing a text based speech recognition system. Few changes to the feature extraction technique of MFCC are also suggested to improve the speech recognition efficiency. Although speech recognition devices are already available, now in the market, their development is based on statistical techniques.

I. INTRODUCTION

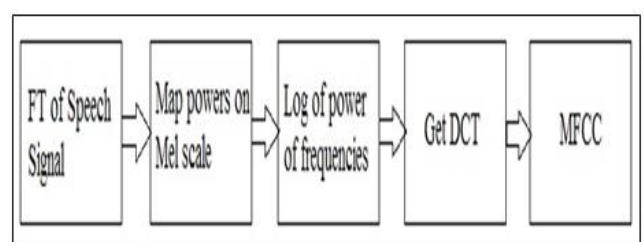
Speech is an art to communicate with other. This is a useful interface to communicate with machines. Many investigation has been done to improve the way of communication. There are many examples like invention of the telephone and loudhailer etc. we learn all the basic skills during childhood, without any guidance, and they continue with us throughout our entire life. we also want to have a simple and efficient way of communication with machines. Therefore, we prefer speech as an interface as compare to using any other interfaces like computer mouse and keyboards. speech is a different phenomenon as the human vocal codes and articulators, are biological organs and they are not under our control. There is different kind of speech in different area which is greatly affected by accents, articulation, pronunciation, roughness, emotional state, gender, pitch, speed, volume, background noise and echoes.

II. SYSTEM SPECIFICATION AND BLOCK SCHEMATIC



A. Feature Extraction

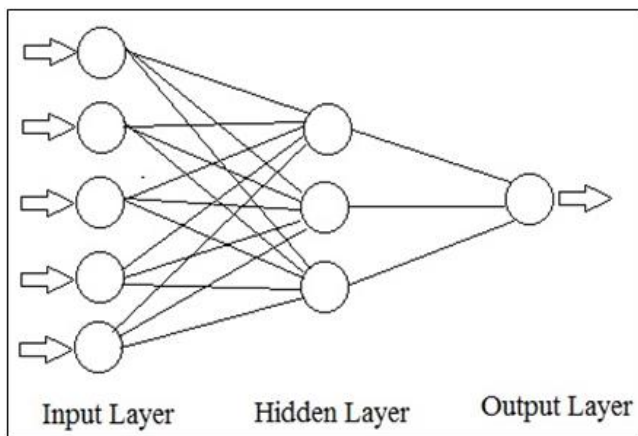
As the speech signal that we are working on in this model contains a large amount of information. This information to be understood in order to be understandable by the developer and the machines are classified as features. Thus features are the different types of requisite information contained in the signal or in the pool of data that can be extracted from the signal but only certain data is needed are thus called as features. Next come the stage of selection the features that will be required in the following stages, thus the term extraction. There are many algorithms used for feature extraction and here we will be using MFCC. Mel Frequency Cepstrum Coefficients are based on the known variations of the human ear’s critical bandwidths with frequencies which are below a 1000 Hz. The main purpose of the MFCC processor is to copy the behaviour of human ears. The derivation of MFCCs is done by the following steps.



B. Artificial Neural Network

What is an artificial neural network?

It is the man-made design of the human brain with the help of the elements made from the electronic model. The ANN finds its base from the working principle of the brain where all the neurons form a dense network and perform different functions over the period of time of their work. Thus a neuron performs as input operator and as processing unit and as output operator. The artificial neurons are built in the same manner keeping in view the working of the biological neurons. To understand working of ANN we need to understand the working of the human brain, where a brain works on three basic functions which are learning, storing and problem solving. Learning includes the time when the brain experiences new situation and thus learns about the problems regarding the situation, and some is to teach or accidentally the solution is found which is near perfect for the particular situation and then the brain stores the experience with its solution. Similarly, the ANN is put through all these steps by the developer thus ANN works efficiently. Also it adapts over the course of it working time as it is flexible in training over the same problem faced with different solutions.



Structure of an artificial neural network.

C. Training

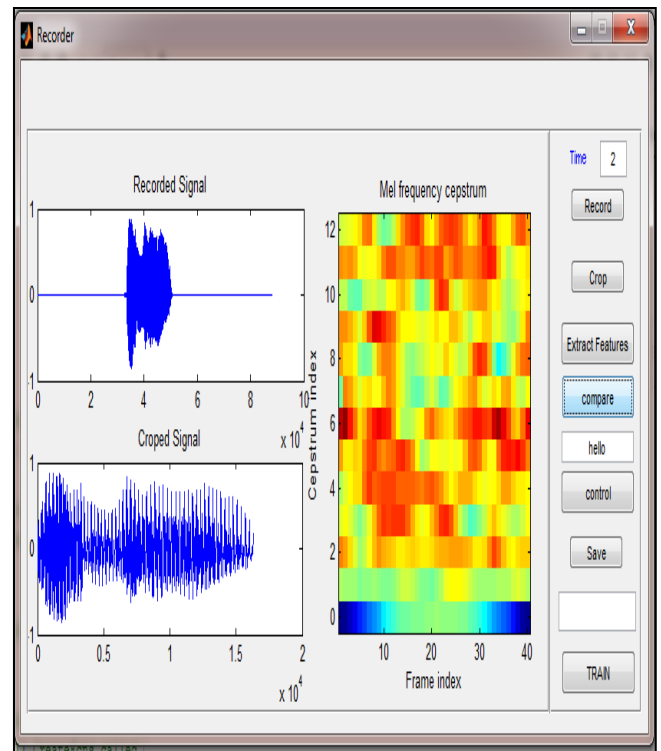
The artificial neural network is trained on a training set of the words that may be spoken in any language. The target file gets created by assigning a particular string. The strings that are used in stored the trained word are accumulated to form a database. Thus every time we train the network it stores the features extracted from the speech signal at the time of training into a particular string designated by the developer.

D. Testing.

At the time of testing the user is allowed to speak the word which he or she desires to be recognised, the word spoken by the user is recorded over the particular period of time according to the word's length in order to reduce the redundancy in the signal. Then the spoken word is put through the feature extractor, same that was used at time of training, which extracts the features from the speech, these features are then given to ANN to perform classification (matching) function in order to recognise the spoken word.

III. RESULT AND CONCLUSION

The main aim of this project was to identify the speech of a person using Multinet neural network,



Lastly by concluding we can say that with increase in the number of samples, the rate of recognition increases

IV. FUTURE SCOPE

- As we know, the technology is improving everyday So, Accuracy level will increase.
- Dictation speech recognition will be accepted gradually.
- Small writing tablets for speech recognition dictation and data entry will be developed.
- To adapt more quickly to the changing background noise levels, different environment, Microphone and sound systems will be designed.
- It will be helpful for disable.

REFERENCES

- [1]. WouterGevaert, Georgi Tsenov, Valeri Mladenov, "Neural Network used for Speech Recognition", Journal of Automatic Control, University of Belgrade, Vol. 20, pp. 1-7, 2010. <http://dx.doi.org/10.2298/JAC1001001G>.
- [2]. Bhushan C. Kamble, "Speech Recognition Using Artificial Neural Network – A Review", Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE) Vol. 3, Issue 1 (2016) ISSN 2349-1469 EISSN 2349-1477.
- [3]. Gulin Dede, Murat HusnuSazl, "Speech recognition with artificial neural networks", Ankara University, Faculty of Engineering, Electronics Engineering Department, Tandogan 06100, Ankara, Turkey.

- [4]. <http://www.guidogybels.eu/asrp4.html>.
- [5]. BhadragiriJagan Mohan, Ramesh Babu. N, “Speech Recognition using MFCC and DTW”, School of Electrical Engineering, ITUniversity, Vellore, India, jagan968@gmail.com, nrameshbabu@vit.ac.in.
- [6]. Seyed Reza Shahamiri, Siti Salwah Binti Salim, “Real-time frequency-based noise-robust Automatic Speech Recognition using Multi-Nets Artificial Neural Networks: A multi-views multi-learners approach”, Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Lembah Pantai, Kuala Lumpur, Malaysia.