Speech Recognition Using Recurrent Neural Network

Amitkumar O. Panchal
Master Of Engineering in Information Technology,
Gujarat Technological University, Gujarat, India

Abstract— The study on Speech Recognition (SR) and understanding has been done for many year. Speech is the vocalized form of human communication. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant speech sound units. SR is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine readable format. Today, however it uses continuous dictation, It is also become smarter with its own set of grammar rules to make out the meaning of what is being said. In this paper, we have a proposed alphabetical words of CORPUS database using MFCC (Mel Frequency Cepstral Coefficient) and Recurrent Neural Network method for a Speech Recogntion (SR).

***Index Terms*— Speech Recognition, Recurrent Neural Network, MFCC, Feature Extraction, Principal Component Analysis**

I. INTRODUCTION

Speech is a complex audio signal influenced by many factors, including speaker characteristics and environmental conditions [1][3]. As a pre-processing step before Automatic Speech Recognition (ASR), it is useful to determine which portions of audio contain speech.

Speech Recognition is the process of converting a speech signal to a sequence of words [7]. The standard approach to large vocabulary continuous SR is to assume a simple probabilistic model of speech production where by specified the word sequences. Speech is the primary means of communication between people. ASR is a process by which a computer takes a speech signal and converts it into words [6][8].

Here, the fully Recurrent Neural Network [14] is an Multilayer Perceptron (MLP) [9] with the previous set of hidden unit activations feeding back into the network along with the inputs as shown in figure 1.

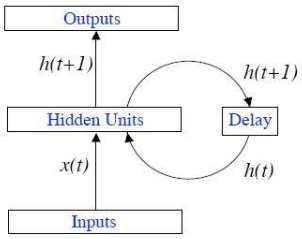

Figure 1 Fully Recurrent Neural Network

Note that the time t has to be discretized with the activations updated at each time step. The time scale might correspond to the operation of real neurons, or for artificial systems any time step size appropriate for given problem can be used. A delay unit needs to be introduce to hold activations until they are processed at the next time step [14].

II. RNN- BASED SPEECH RECOGNITION

We can show that the flow of speech recognition system in figure 2.

We propose approaches to improve speech recognition using Recurrent Neural Network. There are four different steps to find the speech recognition system.

1)Feature extraction
2)Artificial neural network based feature extraction
3)Acoustic model and Language model
4)Global search process

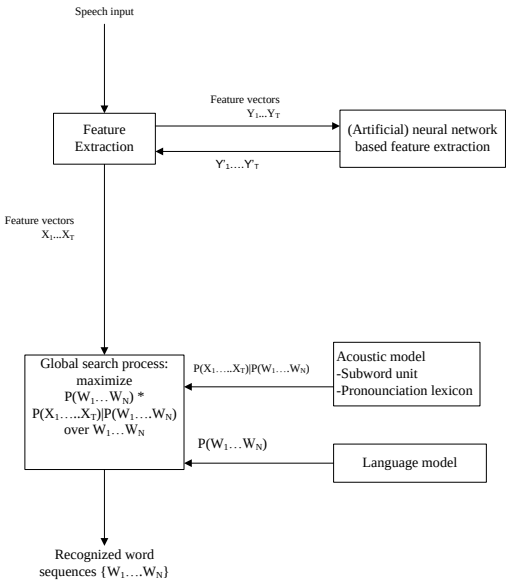

Figure 2 Speech recognition system

III. FEATURE EXTRACTION

A sequence of feature vectors is extracted. It refers to the process of conversion of sound signal to a form suitable for the next stages to use. MFCC are well known features used to describe speech signal other than PCA [10][11][12][13]. Technique of computing MFCC is based on the short term analysis and thus from each frame a MFCC vector is computed. There are some steps of MFCC feature extraction.

1)Pre-emphasis
2)Frame blocking
3)Hamming Windowing
4)Fast fourier transform
5)Mel-scale
6)Discrete cosine transform
7)Log energy
8)Delta cepstrum

IV. LANGUAGE MODEL AND ACOUSTIC MODEL

Language Model :It covers the syntax and semantics of the language implicitly and provides an a priori probability of the word sequence.

According to all model assumptions, the language model probability P(W_1^N) is expressed as:

$$P(W_1^N)=\prod_{n=1}^{N} p(W_n|W_1^{n-1}) \qquad (1)$$

Acoustic Model : The acoustic model is a statistical model which provides the likelihood $P(x_1^T|W_1^N)$ for a sequence of acoustic features x_1^T , given a word sequence W_1^N. Instead of modeling whole words, large vocabulary continuous speech recognition systems use sub-word models like syllables, phonemesor phonemes including context. A pronunciation lexicon provides the mapping of a sequence of sub-word units to whole words.

Whereas the observation sequence $x_1^T$of an hidden Markov model is visible, the state sequence S_1^T is unobservable.Therefore, the probability $P(x_1^T|W_1^N)$is extended by some (hidden) random variables representingthe states of the model [2][4][5] :

$$P(x_1^T|W_1^N)=\sum_{S_1^T:W_1^N}(X_1^T,S_1^T|W_1^N) \qquad (2)$$

Global search process : The search module is the most important part of the speech recognizer. As shown in Figure,the search combines all knowledge sources. The goal of the search module is to find the wordsequence which maximizes the posteriori probability $P(W_1^N|X_1^T)$ for a given feature vectorsequence .

$P(w_1….w_N)*P(x_1..x_T|w_1…w_N)$ over $w_1…w_N$ (3)

Then, finally recognized word sequences.

V. IMPLEMENTATION

Take the sound from given CORPUS database. We take an English alphabetic A to Z etc. as shown in figure 3

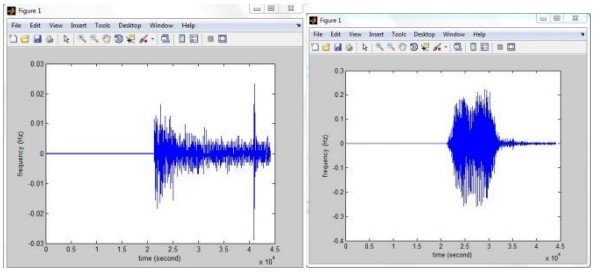

Figure 3 Sound MFCC feature extraction

VI. RESULT ANALYSIS

The following Figure 4 shows the regression graph for one of the sound result for training set, validation set, test set and all.

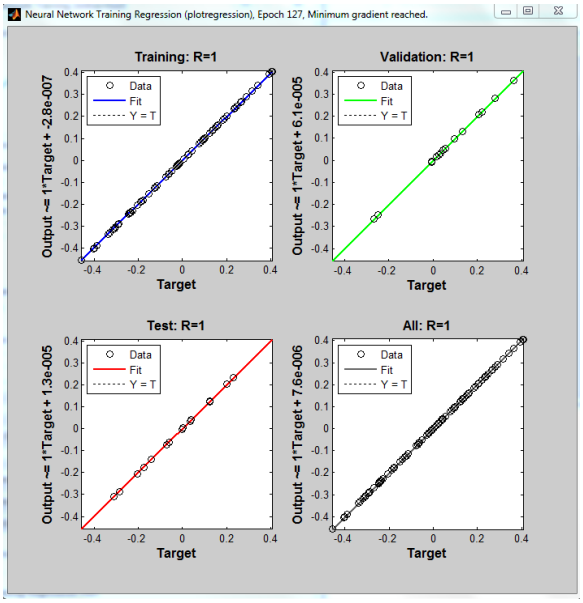

Figure 4 regression analysis

The following Figure 5 shows the performance graph for one of the data input speech like "A" which has best validation performance at 0.089797 at epoch 55.

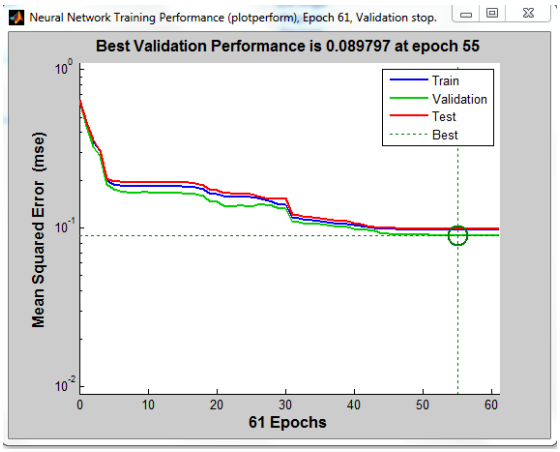

Figure 5 Performance analysis

The following Figure 6 shows the errors analysis.

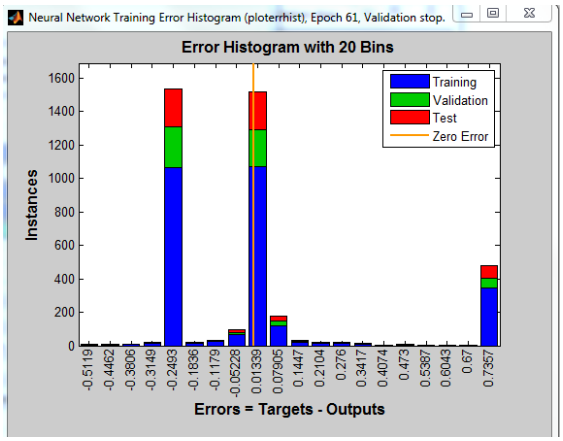

Figure 6 Error analysis

VII. CONCLUSION

We have presented Recurrent Neural Network and MFCC method are used to find recognize speech. MFCC helps to reduce noise frequency and reduce unknown noise. Using RNN, we found that RNN improve the speech.

In Future analysis, we take a different sound database which can provide result and improvement of the performance by decreasing the word error rate. Furthermore, the investigation for the improvement of the RNN proposed system in different modules and combined system analysis.

REFERENCES

[1] Thad Hughes and KeirMierle, "*Recurrent Neural Networks for Voice Activity Detection*," Google inc., USA. IEEE, 2013, pp. 7378-7382.

[2] G.Heigold, V. Vanhoucke, A. Senior, P.Nguyen, M. Ranzato, M. Devin and J. Dean, "*Multilingual Acoustic Models Using Distributed Deep Neural Networks*," Google Inc., USA. IEEE, 2013, pp. 8619-8623.

[3] Vincent Vanhoucke, Matthieu Devin, Georg Heigold, "*Multiframe Deep Neural Networks for Acoustic Modeling*," Google, Inc., USA. IEEE, 2013, pp. 7582-7585.

[4] Yik-Cheung Tam, Yun Lei, Jing Zheng and Wen Wang, "*ASR Error Detection Using Recurrent Neural Network Language Model and Complementary ASR*, " 2014 IEEE International

Conference on Acoustic, Speech and Signal Processing (ICASSP), 2014, pp. 2331-2335.
[5] Tomas Mikolov, Stefan Kombrink, Lukas Burget, Jan "Honza" Cernocky, SanjeevKhudanpur, "*Extensions of Recurrent Neural Network Language Model*, " 2011 IEEE ICASSP, 2011, pp. 5528-5531.
[6] Ms. Vrinda ,Mr. ChanderShekhar, "*Speech Recognition System for English Language*, " International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013, pp. 919-922, ISSN 2278-1021.
[7] Sanjivani S. Bhabad, Gajanan K. Kharate, "*An Overview of Technical Progress in Speech Recognition*, "International Journal of Advanced Research in Computer Science and Software Engineering Vol. 3, Issue 3, March 2013, pp 488-497, ISSN 2277 128X.
[8] IlyaSutskever, James Martens, Geoffrey Hinton, "*Generating Text with Recurrent Neural Network*, " Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA, USA, 2011, pp 1017-1024.
[9] Lawrence R. Rabiner, "*Applications of Speech Recognition in the area ofTelecommunications*, " AT&T Labs IEEE 1997, pp. 501-510.
[10] Abhishek Thakur, Rajesh Kumar, Naveen Kumar, "*Automatic Speech Recognition System for Hindi Utterances with Regional Indian Accents: A Review*, " IJECT Vol. 4, Issue Spl – 3, April – June 2013, pp. 38-43, ISSN 2230-7109.
[11] Andrew L. Maas, Quoc V. le, Tyler M. O'Neil, OriolVinyals, Patrick Nguyen, Andrew Y. Ng, "*Recurrent Neural Networks for Noise Reduction in Robust ASR*, " Google, Inc., USA 2013, pp. 1290-1293.
[12] M. Sundermeyer, I. Oparin, J.-L. Gauvain, B. Freiberg, R. Schluter, H. Ney, "*Comparison of Feedforward and Recurrent Neural Network Language Models*, " 2013 IEEE ICASSP, 2013, pp.8430-8434.
[13] Vidushi Sharma, SachinRai, AnuragDev, "*A Comprehensive Study of Artificial Neural Networks*, " IJARCSSE Vol. 2, Issue 10, October 2012, pp. 278-284, ISSN 2277 128X.
[14] Simon Haykin, "*Neural networks and learning machines*, " Pearson publications, third edition, 2013, pp.798, ISBN 971-81-203-4000-8.