International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Text Dependent Speaker Recognition with Back Propagation Neural Network

N K Kaphungkui, Aditya Bihar Kandali,

Abstract: ***Speaker recognition system follows the procedure of consequently perceiving who is talking or speaking by utilizing the speaker's particular data incorporated into the speech waveform to confirm the identity of a person. By checking the voice attributes of expression, this framework makes it conceivable to validate their character and control access to various database administrations system. This will add additional level of security to any system where security is the main concern. The primary aim of this work is to verify the speaker by extraction of speech features using MFCC and Back Propagation Neural Network as speech classifier. Voice sample of a group of four male and three female uttering the same sentence "This Voice is my password" repeatedly are collected and trained with neural network and testing the network for recognizing are done with untrained new data set with the same utterance spoken once . A specific code or target is assigned for each speakers and recognition is based on how close the network output is to the assigned code for each speaker. Recognition is based on the minimum positive error generation between the code and the actual network output. The tool for simulation is MATLAB R2013a.***

Index Terms: Speaker recognition, MFCC, text dependent, positive error, BPNN, training, testing.

I. INTRODUCTION

Speaker recognition system is classified into many categories [12] as shown in the Figure 1. The system which has any number of trained speakers is an open set system and the number of speakers is always greater than one. Whereas in a closed set recognition system, the system has a specified number of person which is registered in the system. Again in Text-Dependent type (constrained mode), the test utterance is the same to the text which is used during the training session.

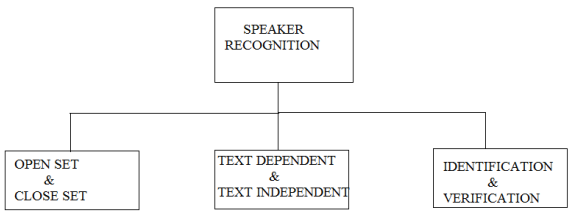

Figure 1: Types of Speaker Recognition system

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
N K Kaphungkui, Electronics and Communication Department, Dibrugarh University, Assam.
Dr Aditya Bihar Kandali, Electrical Department, Jorhat Engineering College, Assam.

Here the test speaker has earlier information of the framework. In contrast to Text-dependent, Text independent system (unconstrained mode) is a type where the test speaker doesn't have earlier information about the content of the training session and can talk or speak anything [15], [11]. Finally Speaker identification is the way toward figuring out which enlisted speaker gives a given articulation and Speaker verification deals with the way toward accepting or dismissing the personality guarantee of as a speaker. It is a 1:1 matching process [3].Training and testing are the two main module or phases which governs speaker recognition system [10], [14] i.e. speech features extraction module and speech features matching model. The various Techniques which are used for features extraction are MFCC, RCC (real cepstral coefficient), LPC (linear prediction coding), LPCC (linear prediction cepstral coding), and PLPC (Perceptual Linear Predictive Cepstral Coefficients). Likewise for Speech Feature matching, various classifiers such as DWT, GMM, SVM, HMM and VQ are used and in recent years neural network is also employed.

II. MEL FREQUENCY CEPSTRAL COEFFICIENT

Researcher have reported that MFCC method is popular and becoming successful for speech features extraction as it is being modeled as human auditory system and also of its high accuracy[9], [10]. Because of its popularity many work have been carried out on speaker recognition based on MFCC. A new approach using weighted MFCC is already presented for speaker recognition. Here the recognition rate is found to be superior to non-weighted MFCC [2]. Speaker recognition with 16-order MFCC coefficient with a combination of LPCC is also reported [7]. Mel Frequency Cepstral Coefficients (MFCC) algorithm is generally preferred as a feature extraction technique to perform voice recognition as it involves generation of coefficients from the voice of the user that are unique to every user. [8]. The various MFCC steps involving for speech feature extraction are listed below [1].

1. Pre emphesis the speech signal
2. Divide the signal into short duration in terms of millisecond.
3. Find the power spectrum for each frame
4. Apply the melfilter bank to the power spectra, sum the energy in each filter.

5. Obtain the logarithm of all filterbank energies.
6. Take the DCT of the log filterbank energies.

The aim of this work is to extract the speech features with MFCC from any speech wav file and represent with unique 13 coefficients for each speakers. The parameters of MATLAB code for speech feature extraction is set at frame duration of 26msec, frame shift 10msec, pre emphasis coefficient of 0.97, filter bank channel 20, lower frequency cut off of filter 300Hz and upper frequency cut of 3750 Hz. The resultant 13 coefficient matrix will have variable column depending on the speech length. The sentence "This voice is my password" is uttered 5 times repeatedly. The voice is sampled at 48 KHz with a bit depth of 16 bit. The MFCC feature is in the form of matrix consisting of fix 13 rows and having different column for different speaker. The longer the speech duration, the more the number of column and vice versa. Data's are collected for four male M1, M2, M3, M4 and three female F1, F2 and F3 in a relatively noise free environment with the same recording device. The speech length is different for different person depending upon how slow or how fast the person utters the sentence. The resultants of MFCC are then used for training the neural network. For better performance of the network, more number of input data is preferred. For this work, the input of the network is obtained from the sentence "This voice is my password" which is uttered repeatedly five times. The simulated speech waveform and its cepstrum of a particular person is shown in Figure 2

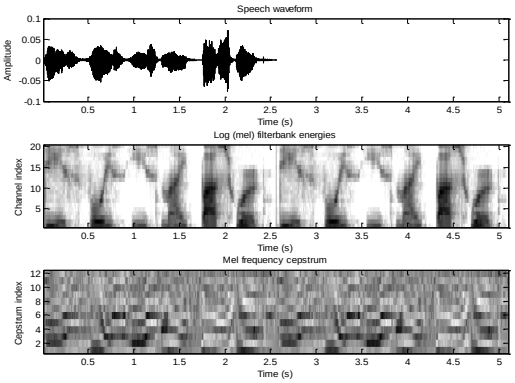

Figure 2 Simulation result of MFCC

III. BACK PROPAGATION NEURAL NETWORK

Two basic steps govern the Back propagation neural network system i.e. feed forward direction which propagates from inputs towards output nodes and back propagation which propagates from output towards input nodes. In the feed forward direction, the output is produced when the input data is applied to the input layer which again passes through the hidden layer until it reach the output node as shown in Figure 3. The error signal is computed, which is the variation between the network's real output and the set network target. Since the hidden layers have contribution to the network output errors, the error signals are then send in reverse direction i.e from output towards input nodes through the entire hidden layer. when the error signal for all the neural node has been calculated, the errors are then used by the nodes to update the values of weights until the network outputs converges to the set target. The Back propagation algorithm aim to minimize the value of the mean square error function by updating its weight until the actual output is so close to its desire output or target value [4], [5]. The weights which minimize the error function is the solution to the learning problem. The more the number of input data's, accuracy will be more and performance of the network will be better [13].

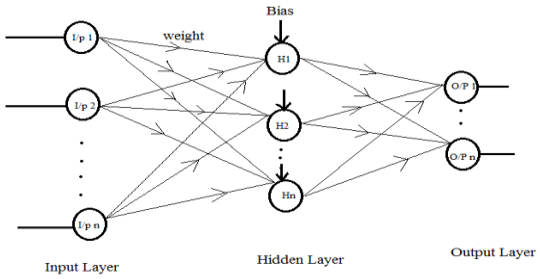

Fig. 3 Multilayer Perceptron Neural Network Structure

The sequences of procedure for Back Propagation Algorithm is as follows [6]
1. Initialization of inputs data at input nodes and set the desired target of the network.
2. Initialization of Weights and biases of network.
3. Divide the data sets.
4. Selecting the no. of Hidden Layers.
5. Network Training to make the output equal to the set target
6. If the output is not match with the target value, update the Weight and bias of network and repeat the process of retrain.
7. Stop the training when target value match with output value.

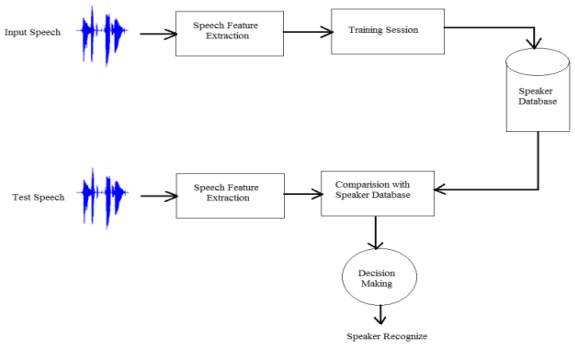

Figure. 4 Structure of speaker verification

The general structure of speaker verification is shown in Figure. 4 where a speaker is either dismiss or accepted which is carried out by the decision making block. The propose network is designed for one to one mapping as shown in Figure. 5. Each of the networks is assigned with a specific code exclusively for a specific speaker. After training the network with sufficient number of input data from speech features extracted with MFCC method, the network will recognized only the speaker which is trained for, based on the error generated by the network which is the difference between the actual network output and the code assigned to the network

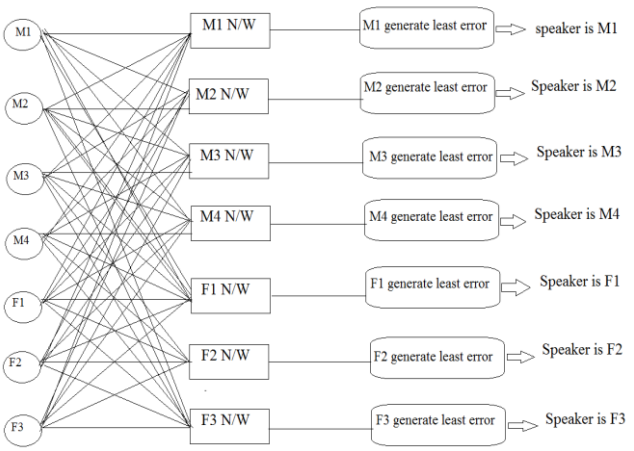

Figure. 5 The propose structure of Speaker recognition

The model is built with a multi-layer perceptron neural network consisting of 13 inputs nodes, 30 hidden neurons with 1 output node. The epoch is set at 500 with a learning rate of 0.35. The 13 coefficient features obtained from MFCC computation is given as inputs to the inputs of neural network.

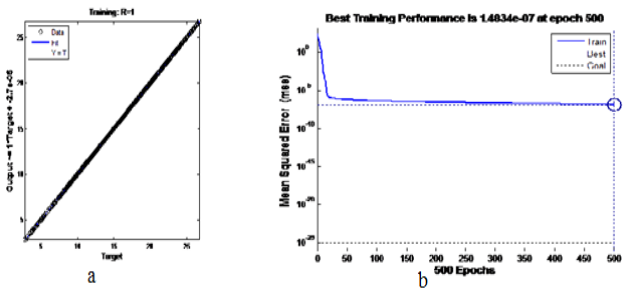

Figure. 6 Regression plot and performance of the neural network

The Training phase of the neural network is carried out for each speaker uttering the same sentence "This voice is my password" for five times repeatedly. The desire target is set for the given inputs and simulation is carried out. The best result is obtained at 500 epochs with mean square error zero as shown in Figure. 6 (b). The fitting plot and the actual output with the set target is also shown in Figure. 6 (a) and Figure. 7 (a) respectively.

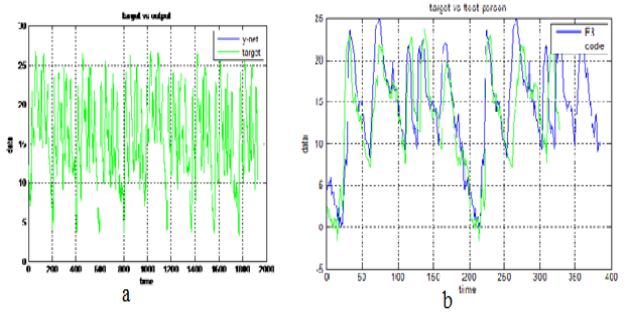

Figure. 7 simulation result of Training and Testing with new speech input

The testing phase is carried out with new inputs for the same sentence spoken once and the simulation result of a particular speaker is shown in Figure. 7 (b). A specific tag or code is assigned to each speaker. The neural network target for each speaker is a set of different numeric number from negative to positive values. The network is trained to produce optimum output which is as close as to the target. The variation between the target set and the actual output is the network error. In order to represent the error with a single value, the algebraic sum of all the target's values and output's values are computed and the difference between the two will represent as error. The network output's which is trained for a particular speaker say F3 will match the code closely for that speaker only and for other speaker there will be a wide variation between the output and the target or code assigned. The positive error between the actual output and the code is calculated and found that the error is always less for the right speaker and for other speakers' error will be more as shown in Figure.8. Therefore based on the least positive error generated the speaker can be recognized.

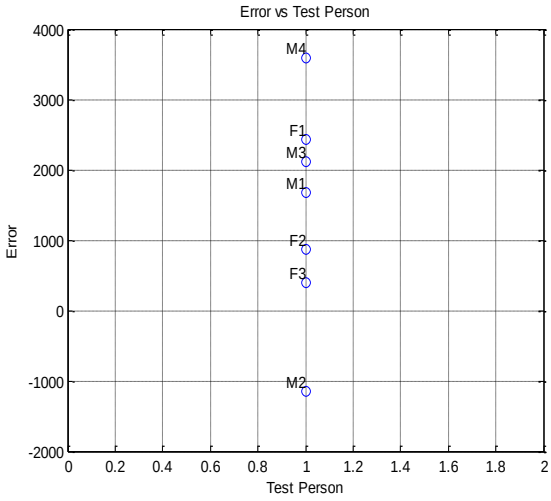

Figure. 8 Test person Vs error plot

In a similar manner the recognize speakers based on the least positive error generated of four male M1, M2, M3, M4 and other two female F1 and F2 is shown in the simulation result in Figure. 9 and Figure. 10 respectively.

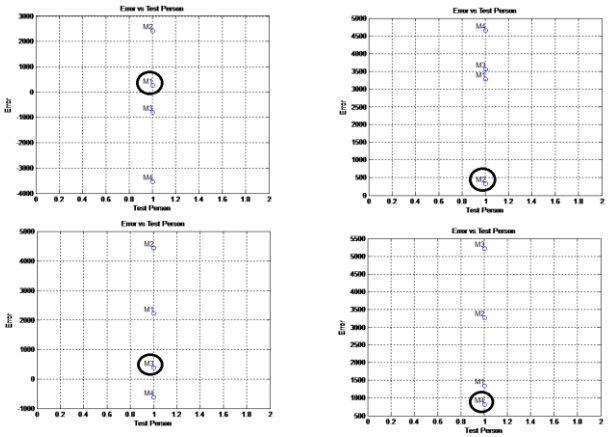

Figure.9 simulation result generating the least positive error of M1, M2, M3 and M4

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

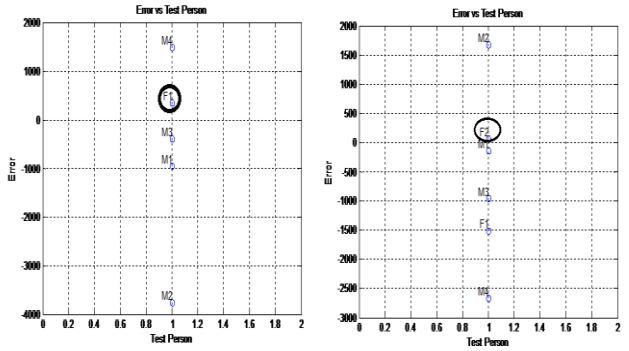

Figure.10 simulation result generating the least positive error of F1 and F2

IV. CONCLUSION

A positive error threshold value is set and by applying conditional property, the speaker claimed to be can be either rejected or accepted. A speaker is accepted if the error generated between the specific codes assigned for his/her voice and the network output is greater than zero and less than or equal to the threshold error value otherwise the speaker is rejected. The more the number of training set data, the more the accuracy of the network performance. The propose work is carried out for seven person i.e four male M1, M2, M3, M4 and three female F1, F2 and F3. The simulation result generating the least error is also shown in Figure. 9 and Figure. 10. This can be extended for N number of speakers applying the same concept for its recognition. The advantage of this network is less simulation time as each network is trained only for a specific user for one to one mapping.

REFERENCES

1. "Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques" Jorge MARTINEZ*, Hector PEREZ, Enrique ESCAMILLA, Masahisa Mabo SUZUKI, CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers, 27-29 Feb. 2012 pages: 248 - 251,IEEE Conference Publications
2. "The Research of Feature Extraction Based on MFCC for Speaker Recognition" Zhang Wanli, Li Guoxin, Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, 12-13 Oct. 2013, Pages: 1074 - 107IEEE Conference Publications
3. "A Review On Speaker Recognition Approaches And Challenges" Varun Sharma, Dr. P K Bansal, International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 5, May – 2013 page 1581-1588.
4. "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition" Md. Ali Hossain1, Md. Mijanur Rahman2, Uzzal Kumar Prodhan3, Md. Farukuzzaman Khan4 International Journal of Information Sciences and Techniques (IJIST) Vol.3, No.4, pp 1-9 July 2013
5. "MATLAB Based Back-Propagation Neural Network for Automatic Speech Recognition" Siddhant C. Joshi1, Dr. A.N.Cheeran2, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 7, pp 10498-10504, July 2014.
6. "FEED FORWARD BACK PROPAGATION NEURAL NETWORK FOR SPEAKER INDEPENDENT SPEECH RECOGNITION" N.AYSHWARYA[1], G.LOGESHWARI[2], G.S.ANANDHA MALA[3], International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, pp 36-39, Aug.-2014
7. "Speaker Recognition Based on Principal Component Analysis of LPCC and MFCC"XinxingJing1 ,Jinlong Ma2, Jing Zhao3 , Haiyan Yang4, 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC) 5-8 Aug. 2014, Pages: 403 - 408 IEEE Conference Publications
8. "Voice Recognition Using MFCC Algorithm"KoustavChakrabortyAsmitaTaleleProf.SavithaUpadhya, International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 10 (November 2014), page 158-161.
9. "A Unique Approach in Text Independent Speaker Recognition using MFCC Feature Sets and Probabilistic Neural Network" Khan Suhail Ahmad1, Anil S. Thosar2, Jagannath H. Nirmal3 and Vinay S. Pande42015 Eighth International Conference on Advances in Pattern Recognition (ICAPR) Year: 5-7 June2015 Pages: 1 - 6, IEEE Conference Publications
10. "Voice recognition Using back propagation algorithm in neural networks" Abdelmajid Hassan Mansour[1], Gafar Zen Alabdeen Salh[2], HozayfaHayderZeen Alabdeen[3] International Journal of Computer Trends and Technology (IJCTT) ISSN: 2231-2803 volume 23 Number 3, pp 132-139 May 2015
11. "Text-Independent Speaker Recognition for Ambient Intelligence Applications by Using Information Set Features" AbhinavAnand, RuggeroDonidaLabati, MadasuHanmandluy, Vincenzo Piuri_, Fabio Scotti, IEEE International Conference on Computational Intelligence and virtual Environments for Measurement system and Applications (CIVEMSA),Year: 26-28 June, 2017, Pages: 30 – 35
12. "A Review Article on Speaker Recognition with Feature Extraction" Parvati J.Chaudhary1, Kinjal M. Vagadia2, International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015 page 94-97
13. "Voice Identity Finder Using the Back Propagation Algorithm of an Artificial Neural Network" Roger Achkar*, Mustafa El-Halabi*, ElieBassil*, Rayan Fakhro*, Marny Khalil* Complex Adaptive Systems, Publication 6, Conference Organized by Missouri University of Science and Technology 2016 - Los Angeles, CA
14. "Speaker Recognition Based on MFCC and BP Neural Networks" Yi Wang, Dr. Bob Lawlor, 28th Irish Signals And systems Conference Year: 20-21 June 2017, Pages: 1 - 4, IEEE Conference publication
15. "A Text-dependent Speaker-Recognition System" Dany Ishac1, 3, Antoine Abche2 and Elie Karam2, Georges Nassar3, Dorothée Callens3, 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Year: 22-25 May 2017 page 1-6, IEEE Conference Publications