International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Text Dependent Speakers Pattern Classification with Back Propagation Neural Network

N K Kaphungkui, Gurumayum Robert Michael, Aditya Bihar Kandali

Abstract: ***Speaker Recognition is the procedure of validating a speaker's claimed identity using his/her speech characteristics which is unique to each individual. The primary objective of all speech recognition system is a man-machine interface which grants access into the system with the voice characteristics. This will served as a highly secure biometric system where security is the primary concern. The primary aim of this paper is to classify each speaker accurately with MFCC and Back Propagation Neural Network. Scaled conjugate gradient training function is used for back propagation neural network. A small database of 10 people is created from a group of five male and five female uttering the same sentence five times repeatedly. The sentence consists of five different words. The numbers of data set for classification is 22182.The accuracy obtained from the classification is 92.1% with small percentage of 7.9% misclassification which is acceptable good. The tool for simulation is MATLAB.***

Keywords : Speaker recognition, MFCC, text dependent, confusion matrix, training, validation, testing.

I. INTRODUCTION

Speaker recognition system is broadly classified into different types as shown in Fig.1 [12]. The system having many trained speakers is an open set system. Whereas a closed set recognition system is the system which has specified registered speakers in the system. Again in Text-Dependent type (constrained mode), the same test utterance (words or sentences) is used during training and testing session. As such, the speaker has prior knowledge of the system and the system need cooperation from the speaker. But in text independent system (unconstrained mode) the test speakers have no information about the words or sentences used for training the system and have the liberty of speaking anything [15], [11].

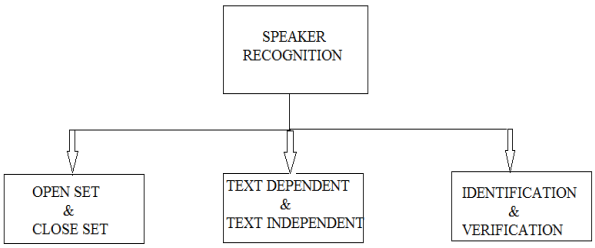

Fig1. Types of Speaker Recognition system

Finally Speaker identification is the process of figuring out that a given articulation belongs to which of the enlisted speaker and Speaker verification is a framework of either accepting or dismissing the personality claimed of as a speaker. It is a 1:1 mapping process [3]. System's training and testing are two important phases which governs all types of speaker recognition system [10], [14]. The various Techniques which are used for unique speech features extraction are MFCC, LPCC, RCC, LPC and PLPC. Likewise for Speech classification various classifiers such as Neural Network DWT, HMM, GMM, SVM and VQ are used. Researcher have reported that when 70% of total samples are used for training and 30% of total samples are used for system testing, the highest efficiency of 81.8% for 10 person is obtained using combination of MFCC, pitch and rms with feed forward neural network (FFNN) [16]. Other have proposed that ANN can give results better than fuzzy logic based systems in terms of accuracy rate with the condition that the samples is recorded in a controlled and clean environment. 74% accuracy is obtained with ANN against 72% accuracy rate with fuzzy logic [17]. Mel Frequency Cepstrum Co-efficients) Based Text-Dependent Speaker Identification Using BPNN is already presented. The highest accuracy of 92% is achieved for 10 users [18]. A new method where both LPC and MFCC are used in parallel is also proposed for Assamese Speaker Recognition Using Artificial Neural Network for 10 speakers where each single word is uttered twenty times by each speaker. A moderately high accuracy is obtained [19]. This work will be carried out with basic 13 coefficients MFCC to represent speech features and BPNN as classifier.

II. MEL FREQUENCY CEPSTRAL COEFFICIENT

In recent years many have reported that speech features extraction based on MFCC method is becoming popular and successful as it is being modeled as human auditory system with high accuracy rate [9], [10]. Some have already used a new way of using weighted MFCC for speaker recognition and found that the recognition rate is superior to non-weighted MFCC [2].

Manuscript published on November 30, 2019.
* Correspondence Author
N K Kaphungkui, Department Of Electronics and Communication, Dibrugarh University, Assam. India. Email: pipizs.kaps@gmail.com
Gurumayum Robert Michael, Department Of Electronics and Communication, Dibrugarh University, Assam, India. Email: roberteld008@gmail.com
Dr Aditya Bihar Kandali, Electrical Department, Jorhat Engineering College, Assam, India. Email: abkandali@gmail.com

A 16-order MFCC coefficient with a combination of LPCC as classifier for speaker recognition is also reported [7]. Mel Frequency Cepstral Coefficients (MFCC) algorithm is more preferred to extract speech features for performing voice recognition due to the generation of unique coefficients from the user's voice [8]. Fig. 2 shows the basic steps involving for speech feature extraction based on MFCC [1].

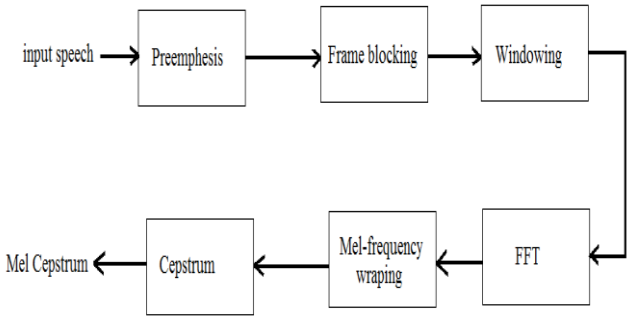

Fig. 2 MFCC blocks

This work's aim is the extraction of the speech features with MFCC algorithm from any speech wav file and represent with a unique 13 coefficients for each speakers. The MATLAB code parameters for speech feature extraction is set as follows

a) Frame duration = 26msec
b) Frame shift = 10msec
c) Pre-emphasis coefficient = 0.97
d) Number of filter bank channel = 20
e) Filter's Lower cut off frequency= 300Hz
f) Filter's Upper cut off frequency = 3750 Hz.

A five words sentence "This voice is my password" is uttered five times repeatedly. The sampling of the voice is 48 KHz with a 16 bit, bit depth. The resultant MFCC features are in matrix form which consists of fix 13 rows and having different number of column for different speaker. It can be represented as M x N where M is fix rows and N is variable columns. A small database is created for five female F1, F2, F3, F4, F5 and five male M1, M2, M3, M4 and M5 in a quite noise free environment with the same recording device. The length of the speech will be different for different person as it depends upon how fast or how slow the person utters the sentence. The more the time is taken while uttering the sentence the more the speech length is and vice versa. The resultants 13 coefficients representing the speech are then used for training the neural network. The more the number of input data is used for training, the better the performance of the network will be. The pictorial representation of the simulated speech waveform and its cepstrum for a particular person is shown in Fig 3.

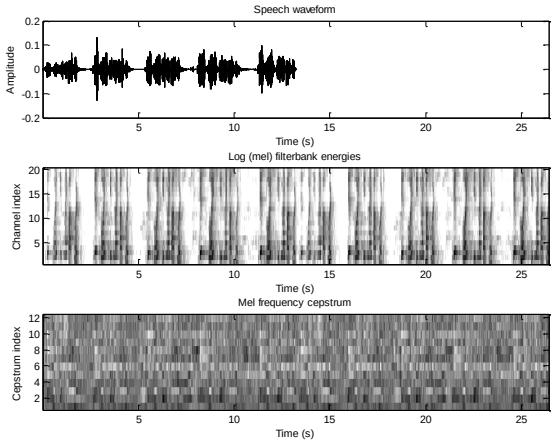

Fig. 3 Simulation result of MFCC

III. BACK PROPAGATION NEURAL NETWORK

The structure of multi layer Perceptron neural network is shown in Fig. 4. There are two basic steps that governs the Back propagation algorithm i.e. feed forward direction which propagates from inputs towards output nodes and back propagation which propagates from output towards input nodes. A back propagation algorithm is a method for training Multi-layer Artificial Neural Networks. Initially while neural network is designed, random values of weight are assigned for the model to predict the correct output. If there is a huge variation between the output and the actual target, the weights are updated to minimize the model error. The solution of the leaning problem lies with the weight that minimizes the error function. The basic steps of the algorithm is as follows

- Error calculation – how much difference between network output from actual output
- Error minimizing – model error minimum or not
- Change the network parameters – if there is huge error, update the weights and bias and check the error again. The whole process is repeated to obtain minimum error.
- Network ready for prediction – after obtaining least error, the training stop and the model is ready for the inputs to predict the actual output.

Minimizing the value of the mean square error function so that the output is closely following the desired targets by updating the weights of the network is the algorithm of back propagation in machine learning. [4], [5], [6]. The weights minimizing the error function is the solution to the neural network learning problem. The more the number of training data set, high accuracy is obtained and the network's performance will improve [13].

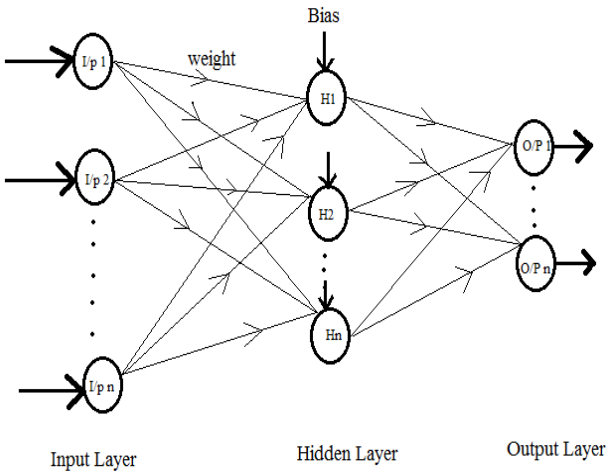

Fig. 4 Multilayer Perceptron Neural Network Structure

The Back Propagation Algorithm is shown in Fig.5

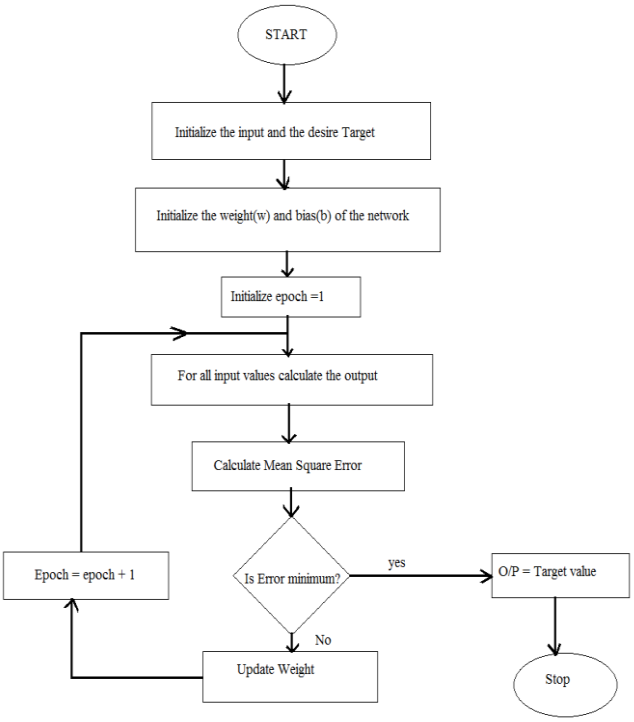

Fig.5 Flowchart of BPNN

Training of the network only stop when the output is almost equal to the target value ideally. To achieved this the loop will run by updating the weights and incrementing the number of epoch.

IV. CONFUSION MATRIX

Confusion matrix is the matrix which is used for different pattern classification as shown in Fig. 6. The sum of a particular row represents the total number of test examples of that class i.e the sum of true positive (TP) and false negative (FN) of that class. For example, total number of test examples of class 1 is the sum of TP1+E12+E13+E14…+E1n. The total number of FN's of a certain class is the sum of all the errors in a row excluding the TP. Likewise the total number of FP's of a particular class is the sum of all the errors in a column excluding the TP. And the total number of TN's for a specific class will be the sum of all the rows and columns by excluding or masking that class's row and column.

	PREDICTED					
ACTUAL		1	2	3	4 •••	n
	1	**TP1**	E12	E13	E14•••	E1n
	2	E21	**TP2**	E23	E24•••	E2n
	3	E31	E32	**TP3**	E34•••	E3n
	4	E41	E42	E43	**TP4**•••	E4n
	n	En1	En2	En3	En4•••	**TPn**

Fig. 6 Confusion matrix of multiple class

The various parameters of confusion matrix namely Accuracy, Precision, Sensitivity and Specificity are expressed as under.

Accuracy is defined as how close the predicted values are to the actual or true values. Accuracy of the confusion matrix is given by the sum of correct classifications or the true positive values divided by the total number of classifications.

Accuracy = (Sum of diagonal elements)/(Sum of all elements) (1)

Precision refers to how close two or more measurements are to each other. Precision of a certain class is calculated as the ratio of true positive value of that class to the corresponding sum of true positive value and the false positive value of that class.

Precision = TP/(TP + FP). (2)

Precision of class 1 is expressed as, Precision1 = TP1/(TP1 + E21 + E31 + E41…+ En1)

Sensitivity (which is also known as a true positive rate, **TPR**) is the proportion of data's that are genuinely positive and are correctly classified. Sensitivity or recall is calculated as the ratio of true positive value of that class to the corresponding sum of true positive value and the false negative value of that class.

Sensitivity = TP/(TP + FN) (3)

Sensitivity of class 2 is expressed as, Sensitivity 2 = TP2/(TP2 + E21 + E23 +E24…+E2n)

The specificity (also called as the true negative rate, **TNR**) is the proportion of actual negative data's that are correctly classified. Specificity is given by the ratio of true negative value to the sum of true negative value and false positive value.

Specificity = TN/(TN + FN) (4)

Specificity of class 1 is calculated as, Specificity1= TN1/(TN1 + E21 +E31 +E41 +…+En1)

V. IMPLEMENTATION AND SIMULATION RESULT

The model is built with a multi-layer perceptron neural network consisting of 13 inputs nodes, 350 hidden neurons with 10 output node for ten class classification. The basic 13 coefficient features obtained from MFCC computation of ten person are concatenated and given as inputs to the neural network. The inputs given to the neural network is 13x22182 and the target is set at 10x22182.

The total number of data set for classification is 22182. The simulation result of ten class's pattern classification is shown in the confusion matrix of Fig. 7.

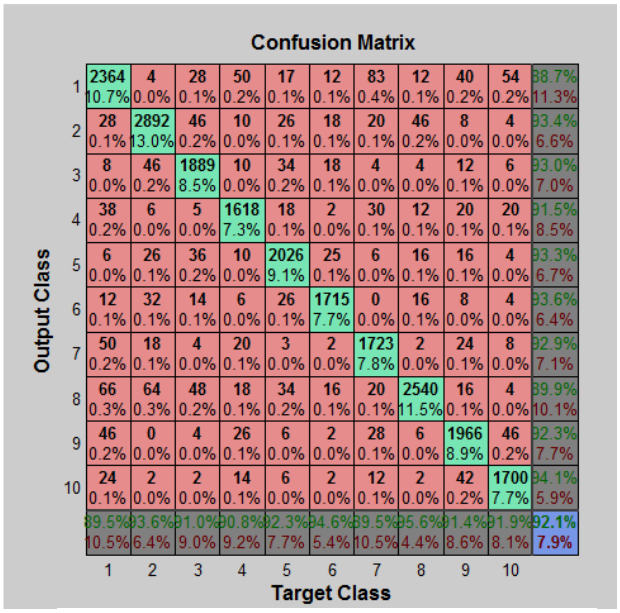

Fig. 7 Simulation result of pattern classification

Accuracy of the classifications is found to be 92.1% and the corresponding Precision and Sensitivity of various classes is also summarized in the table as shown in Fig.8. Precision varies between 88.7% and 94.1% and sensitivity between 89.5% and 95.6%.

1	Accuracy	92.1%
2	Precision of F1 (class1)	88.7%
3	Precision of M1 (class2)	93.4%
4	Precision of M2 (class3)	93%
5	Precision of F2 (class4)	91.5%
6	Precision of M3 (class5)	93.3%
7	Precision of M4 (class6)	93.6%
8	Precision of F3 (class7)	92.9%
9	Precision of M5 (class8)	89.9%
10	Precision of F4 (class9)	92.3%
11	Precision of F5 (class10)	94.1%

1	Sensitivity of F1	89.5%
2	Sensitivity of M1	93.6%
3	Sensitivity of M2	91%
4	Sensitivity of F2	90.8%
5	Sensitivity of M3	92.3%
6	Sensitivity of M4	94.6%
7	Sensitivity of F3	89.5%
8	Sensitivity of M5	95.6%
9	Sensitivity of F4	91.4%
10	Sensitivity of F5	91.9%

Fig.8 Table of Accuracy, Precision and Sensitivity

The specificity of different classes is also calculated and listed as shown in the Fig.9. It varies from 98.5% to 99.5% which is acceptable good

Sl	Specificity	TN/(TN+FN)	%
1	Specificity F1	19240/19518	98.5%
2	Specificity M1	18886/19084	98.9%
3	Specificity M2	19964/20151	99%
4	Specificity F2	20249/20413	99.1%
5	Specificity M3	19841/20011	99.1%
6	Specificity M4	20252/20349	99.5%
7	Specificity F3	20125/20328	99%
8	Specificity M5	19240/19356	99.4%
9	Specificity F4	19866/20052	99%
10	Specificity F5	20226/20376	99.2%

Fig. 9 Specificity of ten classes

The ROC is a graph which summarizes the classifier performance and it is generated by plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis). The more the results lean towards the true positive rate the better is the classification. The simulated result of all the ten classes classification scoring the maximum true positive rate and the validation performance at 832 epochs is also shown in Fig.10.

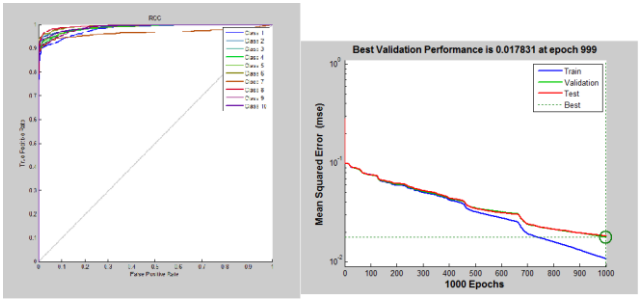

Fig. 10 Receiver Operating Characteristic (ROC) and Validation Performance

VI. CONCLUSION

When more number of data set is given to the network for adequate training, the network performance will improve and achieved high accuracy. 75 % of the data's are use for training, 15% for validation and another 15% for testing the network. Network classification performance in terms of accuracy, precision, sensitivity and specificity all score above 90% with less percentage of misclassification i.e only 7.9%. This model is built for a small number of database and it can be extended for N number of users applying the same methodology. The simulation result of various parameters and the calculated values are all shown in the implementation and simulation section.

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
16. "Speaker recognition using pattern recognition neural network and feedforward neural network" Neha Chauhan, International Journal of Scientific & Engineering Research, Volume 8, Issue 3, pp 1444-1446, March-2017.
17. "Language and Text Independent Speaker Recognition System using Artificial Neural Networks and Fuzzy Logic" J Sirisha Devi International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-6, pp 327-330, March 2019.
18. "MFCC Based Text-Dependent Speaker Identification Using BPNN" S. S.Wali, S.M. Hatture, S. Nandyal, International Journal of Signal Processing System VOL.3, No.1, pp 30-34, June 2015.
19. "Assamese Speaker Recognition Using Artificial Neural Network" Bhargab Medhi1, Prof. P.H.Talukdar2, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, pp 321-324, March 2015