International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

Modeling of Speaker Recognition with MFCC and Neural Network

N K Kaphungkui[1], Dr Aditya Bihar Kandali[2]

[1]*Department of Electronics and Communication, Dibrugarh University, Assam, India.*

[2]*Department of Electrical Engineering, Jorhat Engineering College, Assam, India.*

ABSTRACT:

This paper will present the use of MFCC and neural network to model an automatic Speaker Recognition system. The objective of this work is to classify the different speakers' pattern with acceptable accuracy rate and to verify each speaker correctly without any unambiguity. Mel frequency Cepstral Coefficient (MFCC) will be used for the extraction of speech features from the voice signal and Back Propagation Neural Network for identification of the speaker. After the speakers' patterns are correctly classified, the classifier model is tested with the voice of a registered speaker's and found that it successfully recognized without any obscurity. Resilient Back propagation training function is used for training the BPNN. A speech database consisting of 20 speakers is created from a group of ten male and ten female with the same utterance. The number of data set for 20 speakers' classification is 28,414. The accuracy obtained from the classification is 83.6%. The overall precision and sensitivity scores are 84.21% and 83.88% respectively which are good enough. Matlab tool is used for the entire simulation.

Keywords: *Speaker recognition, MFCC, Back Propagation Neural Network, Accuracy, training, testing.*

I. INTRODUCTION

Speech is the most effective way of communication between human. In the same way many researchers have already implemented efficient method for the interaction of man and machine successful. Speaker recognition system is one of its application to identify and verify a speaker by the system. In any automatic speaker recognition system there are two main modules which govern the whole system, they are the speech features extraction module and speech feature matching module. Several methods can be used for the extraction of speech features such as Linear Prediction coefficients, Mel-Frequency Cepstral Coefficients, Gammatone Frequency Cepstral Coefficients, Linear Predictive Cepstral Coefficients and Perceptual Linear Predictive. In the same way for Speech features matching methods, different classifiers such as Artificial neural network (ANN), Dynamic Time Warping (DWT), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), K-Mean Clustering (K-mean) and Vector Quantization (VQ) can be used. The two important phases that a Speaker Recognition system undergo are the training phase and testing phase [12]. A work has already reported that for class ten classification. Here an accuracy of 81.8% is obtained with the combination of MFCC, pitch and rms in

feed forward neural network (FFNN) [9]. Artificial Neural Network shows better result in terms of accuracy than fuzzy logic based systems when a speech is recorded in a noiseless environment. Accuracy obtained with ANN is 74% against 72% with fuzzy logic [3]. In text dependent speaker recognition system of 10 speakers' accuracy of 92% is achieved with the combination of MFCC and BPNN [11]. A moderate accuracy for 10 speakers is also achieved with the combination of LPC and MFCC using Artificial Neural Network for Assamese Speaker Recognition [1]. This paper will be implemented for 20 speakers with the combination of MFCC and BPNN to achieve an accuracy of 83.6% and to show that this accuracy can also suffice for the identification of an unknown registered speaker correctly. This work will be organized as follows, literature survey is given in this section, in section 2 MFCC will be briefly discuss, section 3 will give the working of back propagation neural network, section 4 will show the implementation and result and section 5 will give the conclusion of the work.

II. MEL FREQUENCY CEPSTRAL COEFFICIENT

MFCC is one of the most widely used and common for the extraction of speech features. Higher success rate is obtained with MFCC due to the fact that it is modelled as human auditory system. It fails to perceived signal over 1KHz and showing more robust against noisy environment [2], [5]. Mel Frequency Cepstral Coefficients generates the voice signal coefficients which are unique to every individual speaker [6]. The overall MFCC's steps for the extraction of speech feature are shown in Figure. 1 [4].

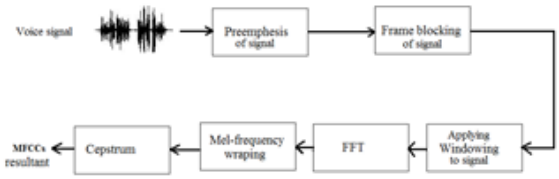

Figure. 1 Block diagram of MFCC computation.

A speech data base is initially created from 20 different speakers with the same utterance speaking repeatedly. The voice data base consist of ten female speakers label as f1, f2, f3, f4, f5, f6, f7, f8, f9,f10 and ten male speakers as m1, m2, m3, m4, m5, m6, m7, m8, m9 and m10. The voice samples are recorded in a relatively quiet and noise free environment. All the speech length will vary for different person. The speech length of a particular speaker will be more if more time is taken to complete the utterance and vice versa. The sampling frequency of the voice signal is at 48 KHz with a 16 bit, bit depth. All the collected speech samples are process with MFCC methods and represented with a unique 13 coefficients for each speaker following a 13-order MFCC. The MFCC resultants which is in matrix form consisting of fix 13 rows and variable column for all the speakers as shown in Figure. 2. This can be represented as M x N where M is fix 13 rows and N is variable columns

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

Sl	Male and Female speakers		MFCC result (M x N)
1	f1	s1	13 x 2064
2	f2	s2	13 x 1615
3	m1	s3	13 x 1050
4	m2	s4	13 x 1096
5	f3	s5	13 x 1426
6	m3	s6	13 x 1085
7	f4	s7	13 x 1544
8	f5	s8	13 x 1473
9	m4	s9	13 x 1419
10	m5	s10	13 x 1918
11	m6	s11	13 x 1242
12	f6	s12	13 x 1686
13	f7	s13	13 x 1302
14	m7	s14	13 x 1397
15	f8	s15	13 x 1408
16	f9	s16	13 x 1556
17	m8	s17	13 x 1226
18	m9	s18	13 x 1094
19	m10	s19	13 x 1466
20	f10	s20	13 x 1347

Figure.2 MFCC result for different speakers

The N values will be more when a speaker takes more time to utter the given sentence and vice versa. The pictorial representation of a particular speaker's voice signal and its MFCCs is shown in Figure 3.

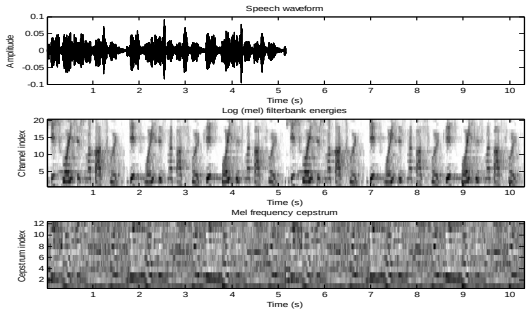

Figure. 3. The Plot of voice signal and its MFCC's resultant

III. BACK PROPAGATION NEURAL NETWORK

The basic structure of multi-layer Perceptron neural network is shown in Figure. 4. It consists of three layers as input layer, hidden layer and output layer.

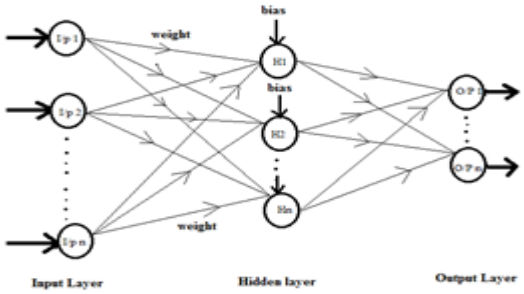

Figure. 4. The network of Multilayer Perceptron Neural Network.

The Back propagation algorithm is used to train a neural network through a chain rule method. It consists of two basic steps which are forward pass of the inputs through the network and back propagation which execute a backward pass by adjusting the parameters of the network. In feed forward direction the input data propagates towards the output node through the hidden layer along with the assigned initial model's parameters i.e. weights

and bias. If the output produced by the network is not equal to the set target then back propagation process will takes place by updating the network parameters backward from output node towards input node. This is an algorithm for supervised learning of Artificial Neural Network. Initially when the model is designed, any random values of weights and bias are assigned to predict the set target. After the forward pass execution if there exist a variation between the network's output and the set target, the parameters are updated backward to minimize this huge error. The particular weights and bias which result in minimizing the error function between the output and the set target is the solution of the network leaning.

The flowchart of a Back Propagation Algorithm is shown in Figure. 5. The objective of a back propagation algorithm is to minimize the mean square error (MSE) functions between the actual output and the desire set target by updating the parameters of the network [7], [8]. The final network parameters which minimize the MSE function is the solution of neural network learning algorithm. Accuracy and performance of the network is better when more number of training data set is used for training the model [10].

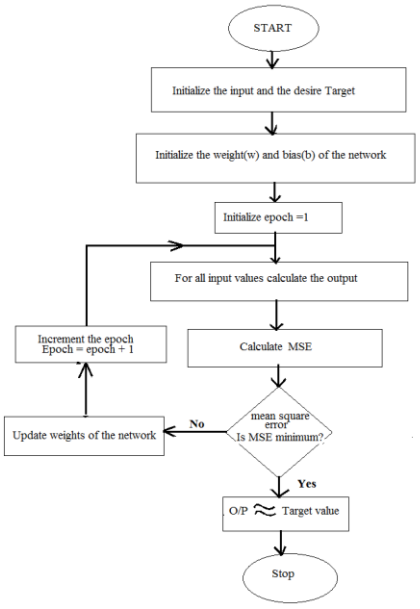

Figure.5 Flowchart of BPNN

In ideal condition, the training session of network will stop when the network's actual output is equal to the desired set target value. To attain this condition, the loop will run continuously by increment the epoch number along with the updation of the network's parameters.

IV. IMPLEMENTATION AND SIMULATION RESULT

The classifier model is implemented with multi-layer perceptron neural network consisting of 13 input nodes, 362 hidden neurons and 20 output nodes for twenty class's classification as shown in Figure 6. The input datas for the neural network is fed from the MFCCs result of twenty different persons. The 13 order MFCC of each speaker are concatenated and gave as input data to the neural network. The simulation result of twenty class's pattern classification is given in Figure. 7.

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

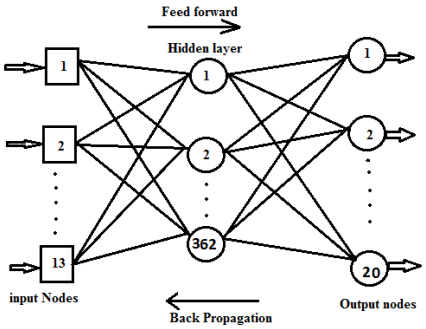

Figure. 6. Speech Classifier Model

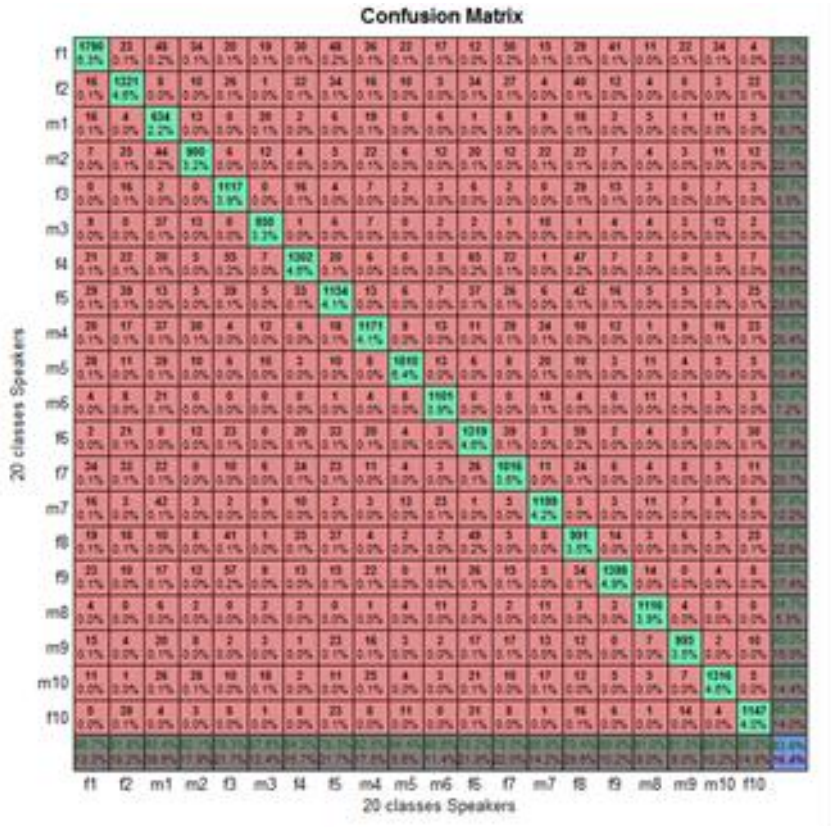

Figure. 7 Confusion matrix of 20 classes classification

The characteristics of confusion matrix are defined by the following terminoly.

- *Accuracy = (Sum of Diagonal elements)/(Sum of all elements)*
- *Precision = True Positive/(True positive+ False Positive).*
- *Sensitivity = True Positive/(True Positive + False Negative)*

The overall accuracy obtained from the classification is found to be 83.6%. And the overall Precision and Sensitivity are 84.21 % and 83.88 % respectively. The Receiver output characteristic (ROC) is used for analysing the classification accuracy. It summarizes the overall performance of the neural network model. It is a graphical representation of the True Positive Rate along the y-axis against the False Positive Rate along x-axis. When the resultants lean sharply towards the true positive rate, the accuracy obtained from the classification will

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

be high and better will be the classifier's performance. The plot of ROC and the best validation performance at 836 epochs is also shown in Figure. 8.

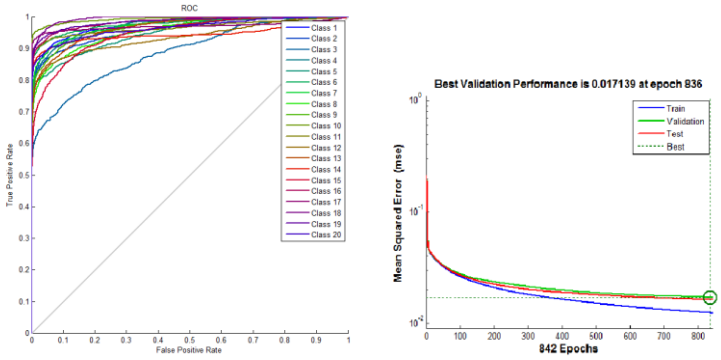

Figure. 8 Classifier's ROC and Validation Performance

After classification of the registered speakers, the classifier model is now ready for testing with new input data's and recognize the speaker. For this process, the voice samples of all registered speakers are collected again with the same utterance. The MFCCs for each speaker are computed again and given as new testing input data to the model. The diagonal elements score of the confusion matrix while testing with different speakers is shown from Figure. 9 to Figure. 12. The score of diagonal element corresponding to the right speaker will always be maximum while all other score will be minimum. Therefore, by examining the row and column to which speaker this maximum score belongs, the speaker can be successfully recognized without any unambiguity.

Speakers	Diagonal Score
f1	1461
f2	49
m1	7
m2	3
f3	0
m3	0
f4	31
f5	111
m4	41
m5	52
m6	0
f6	0
f7	33
m7	0
f8	44
f9	29
m8	25
m9	0
m10	16
f10	4

Speakers	Diagonal Score
f1	54
f2	1065
m1	0
m2	22
f3	45
m3	0
f4	34
f5	107
m4	39
m5	21
m6	0
f6	23
f7	35
m7	0
f8	65
f9	16
m8	0
m9	0
m10	9
f10	43

Speakers	Diagonal Score
f1	241
f2	53
m1	326
m2	95
f3	0
m3	145
f4	145
f5	10
m4	113
m5	26
m6	18
f6	0
f7	6
m7	10
f8	25
f9	0
m8	0
m9	6
m10	95
f10	19

Speakers	Diagonal Score
f1	254
f2	43
m1	39
m2	576
f3	26
m3	30
f4	55
f5	15
m4	83
m5	13
m6	16
f6	0
f7	83
m7	14
f8	1
f9	42
m8	12
m9	0
m10	23
f10	0

Speakers	Diagonal Score
f1	28
f2	81
m1	0
m2	5
f3	869
m3	0
f4	117
f5	114
m4	4
m5	0
m6	3
f6	0
f7	0
m7	0
f8	116
f9	82
m8	5
m9	0
m10	15
f10	0

Figure. 9 The diagonal score of confusion matrix while testing the model with f1, f2, m1, m2 and f3 speakers.

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

Speakers	Diagonal Score
f1	71
f2	0
m1	47
m2	23
f3	0
m3	802
f4	0
f5	0
m4	33
m5	26
m6	7
f6	0
f7	8
m7	9
f8	0
f9	36
m8	3
m9	0
m10	115
f10	0

Speakers	Diagonal Score
f1	69
f2	74
m1	0
m2	0
f3	29
m3	0
f4	1157
f5	63
m4	5
m5	0
m6	0
f6	49
f7	21
m7	0
f8	41
f9	10
m8	4
m9	0
m10	0
f10	11

Speakers	Diagonal Score
f1	72
f2	105
m1	28
m2	37
f3	34
m3	0
f4	73
f5	463
m4	40
m5	28
m6	0
f6	271
f7	74
m7	0
f8	84
f9	102
m8	3
m9	14
m10	0
f10	98

Speakers	Diagonal Score
f1	100
f2	8
m1	36
m2	26
f3	9
m3	7
f4	0
f5	12
m4	980
m5	2
m6	1
f6	0
f7	61
m7	4
f8	4
f9	0
m8	0
m9	46
m10	81
f10	23

Speakers	Diagonal Score
f1	109
f2	11
m1	0
m2	0
f3	0
m3	0
f4	15
f5	0
m4	5
m5	1600
m6	12
f6	8
f7	2
m7	8
f8	20
f9	22
m8	19
m9	0
m10	24
f10	0

Figure. 10 The diagonal score of confusion matrix while testing the model with m3, f4, f5, m4 and m5 speakers.

Speakers	Diagonal Score
f1	77
f2	0
m1	22
m2	12
f3	0
m3	0
f4	0
f5	0
m4	50
m5	74
m6	773
f6	36
f7	0
m7	231
f8	0
f9	33
m8	24
m9	8
m10	0
f10	8

Speakers	Diagonal Score
f1	20
f2	35
m1	0
m2	34
f3	22
m3	0
f4	33
f5	50
m4	0
m5	0
m6	0
f6	1165
f7	20
m7	5
f8	83
f9	74
m8	0
m9	0
m10	4
f10	58

Speakers	Diagonal Score
f1	114
f2	47
m1	14
m2	0
f3	10
m3	0
f4	48
f5	27
m4	7
m5	22
m6	0
f6	24
f7	947
m7	9
f8	11
f9	50
m8	5
m9	0
m10	51
f10	18

Speakers	Diagonal Score
f1	45
f2	46
m1	19
m2	42
f3	5
m3	28
f4	29
f5	10
m4	41
m5	29
m6	48
f6	19
f7	22
m7	880
f8	26
f9	16
m8	16
m9	16
m10	18
f10	5

Speakers	Diagonal Score
f1	30
f2	42
m1	0
m2	1
f3	56
m3	7
f4	129
f5	38
m4	5
m5	7
m6	0
f6	138
f7	6
m7	0
f8	879
f9	33
m8	13
m9	7
m10	40
f10	32

Figure. 11 The diagonal score of confusion matrix while testing the model with m6, f6, f7, m7 and f8 speakers

Speakers	Diagonal Score
f1	53
f2	65
m1	7
m2	13
f3	28
m3	7
f4	5
f5	41
m4	6
m5	2
m6	0
f6	21
f7	0
m7	0
f8	49
f9	1211
m8	0
m9	0
m10	21
f10	15

Speakers	Diagonal Score
f1	35
f2	45
m1	36
m2	31
f3	0
m3	0
f4	0
f5	20
m4	0
m5	26
m6	50
f6	61
f7	14
m7	142
f8	10
f9	51
m8	806
m9	0
m10	10
f10	9

Speakers	Diagonal Score
f1	111
f2	0
m1	7
m2	0
f3	0
m3	15
f4	0
f5	28
m4	12
m5	26
m6	71
f6	4
f7	2
m7	10
f8	0
f9	0
m8	8
m9	810
m10	5
f10	91

Speakers	Diagonal Score
f1	57
f2	5
m1	21
m2	64
f3	0
m3	22
f4	0
f5	21
m4	24
m5	33
m6	3
f6	0
f7	6
m7	0
f8	3
f9	25
m8	0
m9	4
m10	1129
f10	5

Speakers	Diagonal Score
f1	0
f2	17
m1	7
m2	0
f3	10
m3	0
f4	54
f5	66
m4	33
m5	1
m6	4
f6	50
f7	26
m7	0
f8	17
f9	0
m8	0
m9	41
m10	5
f10	1028

Figure. 12 The diagonal score of confusion matrix while testing the model with f9, m8, m9, m10 and f10 speakers.

V. CONCLUSION

Training of the neural network with sufficient number of training data is required to improve its classification accuracy. Hence, more training data set is used for training the network. In this work, the total data set is divided into three set as training data, validation data and testing data. Out of the total data set 80 % is reserved and used for training the model and 10% each for validation and testing the network. The classification accuracy obtained is 83.6%**.** The classifier's performance is acceptable good as the overall precision and sensitivity of all the classes score 84.21 % and 83.88% respectively. The classifier model which correctly recognizing the unknown speaker when testing with new input speech is also shown in the simulation result. The maximum score in the diagonal element of confusion matrix will always corresponds to the right speaker as shown from Figure 9 to Figure 12. The objective of this paper is to recognize a speaker successfully even with 83.6% classification accuracy and hence it is achieved.

REFERENCES

[1]. Bhargab Medhi1, Prof. P.H.Talukdar2 "Assamese Speaker Recognition Using ArtificialNeural Network" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, pp 321-324, March 2015.

[2]. E.C. Gordon,Signal and Linear System Analysis.John Wiley & Sons Ltd., New York, USA,1998.

[3]. J Sirisha Devi "Language and Text Independent Speaker Recognition System using Artificial Neural Networks and Fuzzy Logic" International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-6, pp 327-330, March 2019.

[4]. Jorge MARTINEZ*, Hector PEREZ, Enrique ESCAMILLA, Masahisa Mabo SUZUKI, "Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques" CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers, 27-29 Feb. 2012 pages: 248 - 251,IEEE Conference Publications.

[5]. Khan Suhail Ahmad1, Anil S. Thosar2, Jagannath H. Nirmal3 and Vinay S. Pande4 "A Unique Approach in Text Independent Speaker Recognition using MFCC Feature Sets and Probabilistic Neural Network" 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR) Year: 5-7 June2015 Pages: 1 - 6, IEEE Conference Publications.

[6]. Koustav Chakraborty, AsmitaTaleleProf.SavithaUpadhya, "Voice Recognition Using MFCC Algorithm" International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 10 (November 2014), page 158-161.

[7]. Md. Ali Hossain1, Md. Mijanur Rahman2, Uzzal Kumar Prodhan3, Md. Farukuzzaman Khan4 "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition" International Journal of Information Sciences and Techniques (IJIST) Vol.3, No.4, pp 1-9 July 2013.

[8]. N.AYSHWARYA[1], G.LOGESHWARI[2], G.S.ANANDHA MALA[3] "FEED FORWARD BACK PROPAGATION NEURAL NETWORK FOR SPEAKER INDEPENDENT SPEECH RECOGNITION", International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, pp 36-39, Aug.-2014.

ISSN: 2454-9665

[9]. Neha Chauhan "Speaker recognition using pattern recognitionneural network and feedforward neural network, International Journal of Scientific & Engineering Research, Volume 8, Issue 3, pp 1444-1446, March-2017.

[10]. Roger Achkar*, Mustafa El-Halabi*, ElieBassil*, Rayan Fakhro*, Marny Khalil* "Voice Identity Finder Using the Back Propagation Algorithm of an Artificial Neural Network" Complex Adaptive Systems, Publication 6, Conference Organized by Missouri University of Science and Technology 2016 - Los Angeles, CA.

[11]. S. S.Wali, S.M. Hatture, S. Nandyal "MFCC Based Text-Dependent Speaker Identification Using BPNN" International Journal of Signal Processing System VOL.3, No.1, pp 30-34, June 2015.

[12]. Yi Wang, Dr. Bob Lawlor, "Speaker Recognition Based on MFCC and BP Neural Networks" 28th Irish Signals And systems Conference Year: 20-21 June 2017, Pages: 1 - 4, IEEE Conference publication.