A Novel Speaker Identification System using FeedForward Neural Networks

Seema Khanum
Department of Computer Science & Engineering,
Rajiv Gandhi University
Doimukh, Arunachal Pradesh, India
seema.khanum@rgu.ac.in

Firos A
Department of Computer Science & Engineering,
Rajiv Gandhi University
Doimukh, Arunachal Pradesh, India
firos.a@rgu.ac.in

***Abstract*—This paper proposes a novel speaker identification system which uses Mel Frequency Cepstral Coefficients (MFCC) and Feed Forward Neural Networks (FFNN) for feature extraction and speaker classification respectively. Fuzzy C Mean Clustering (FCM) method is also used against the extracted features from the speech, which facilitates in grouping large amount of data. The efficiency of the system is enhanced furthermore by identifying the gender of the speaker, before the actual speaker identification process, using another FFNN. As a result, the system shows better performance in terms of computational cost and real time identification.**

Keywords— Speaker Identification; MFCC; FFNN; FCM

I. INTRODUCTION

The main objective of the speaker identification system is to identify the voice of the speaker based on his/her previously stored voice samples. The designed system will work on text-independent speech as well as the speech samples containing background environmental noise. Literature survey shows that the gender identification plays a crucial role in carrying out efficiently other speech processing tasks such as speaker recognition, language identification, speech compression, and so on [1]. Thus, in the proposed system, the task of identifying gender prior to the speaker identification is accomplished to assist in achieving better performance and accuracy. The proposed speaker identification system has a wide range of applications such as forensic tests, remote access to computers, security control for confidential information, telephone shopping, banking over telephone network and so on.

The rest of this paper is organized as follows. Section II, describes about the methodology used in the proposed speaker identification system. Section III, describes about feature extraction from the input speech data. The classification technique used in the system is briefed in Section IV. Section V, shows the experiments and results obtained from the implemented system. Finally, Section VI concludes the paper.

II. METHODOLOGY

The speaker identification is carried out mainly by performing the tasks: speech database preparation, feature extraction and feature classification.

To conduct experiments and evaluate performance of the proposed system, a sample speech database is used. It contains around 25 sentences uttered by both male and female speakers. Recording was accomplished using high quality sound card, sound recording software and close talking microphone in a sound proof laboratory. The speech was recorded at a sampling frequency of 8 kHz and coded in 8 bits PCM. These recorded sentences are then corrupted by real world noises at different SNR levels.

The MFCC is one of the best feature extraction techniques exists particularly in case of speech processing. There exist many other feature extraction techniques such as LPC (Linear Prediction Cepstrum) and PLP (Perceptual Linear Prediction).

Classification is generally required at this stage to classify and match the speaker's speech models using classifier. These models are created from the extracted features of the speaker's utterance. To accomplished the task of classification the proposed system uses feed forward neural networks as a classifier.

III. MEL FREQUENCY CEPSTRAL COEFFICIENTS

The Extraction and selection of the best features from speech signal is very essential in speaker identification process as it directly affects the system performance. In such a scenario, MFCCs are proved to be more efficient [2]. The computation of the MFCCs includes the following steps.

1) Digitized speech at 16kHz is pre-emphasized using first order digital filter

$$H(z) = 1 - 0.9z^{-1}$$

2) The speech is then divided into frames of 20ms. Windowing is done using hamming window of window length 20ms.

3) The Fast Fourier Transform (FFT) transforms the windowed speech into frequency domain and short term power spectrum *P (f)* is calculated.

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

4) Obtained *P (f)* is then bent along its frequency axis *f* into the Mel frequency axis *M* as *P (M)* using the following equation as mention in [3],

$$M(f) = 2595 \log_{10}(1 + \frac{f}{700})$$

5) Obtained P *(M)* is then convolved with the triangular band pass filter into *θ (M)*.

$$\theta(M_K) = \sum_M P(M - M_k)\Psi(M), \quad k = 1..K$$

Where, Ψ *(M)* is the critical masking curve.

6) Then *K* outputs are obtained using the following equation.

$$X(k) = \ln\big(\theta(M_k)\big), \quad k = 1 \ldots K$$

7) The MFCC is calculated using the following equation

$$MFCC(d) = \sum_{k=1}^{K} X_k \cos\left[d(k - 0.5)\frac{\pi}{K}\right], \quad d = 1..D$$

Following Fig.1 shows the different stages involved in typical working of MFCC algorithm.

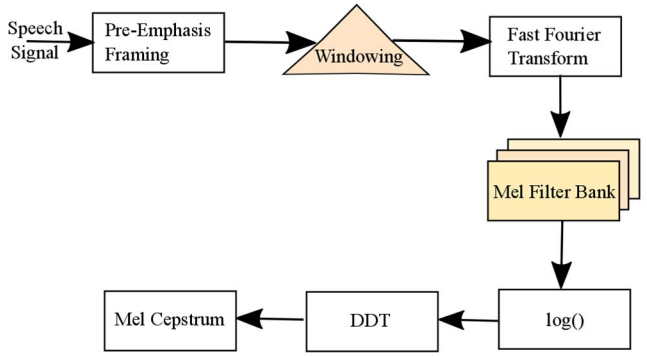

Fig.1 MFCC algorithm

IV. FEED FORWARD NEWRAL NETWORK

For classification, FFNN is used with only one hidden layer. The extracted MFCC features after going through FCM procedure, creates an output matrix of fixed size, which is then fed to the neurons of the input layer. The hidden layer can contain any number of hidden neurons, provided the resultant system performs better. Different trials can be conducted to fix the number of neurons in the hidden layer to achieve maximum accuracy. The general structure of layers present in feed forward neural networks is shown in the following Fig. 2.

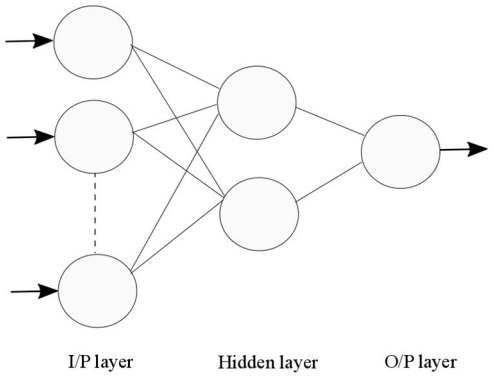

Fig.2. Layers in feed forward neural networks

For training neural network, a back propagation training algorithm called trainscg (Scaled Conjugate Gradient) is used, working of which is shown in the Fig.3. The derivative of the SEF (Squared Error Function) is calculated with respect to the weights of the network. The general equation to calculate SEF as mention in [4], uses the actual output of the output neuron *y* and target output of the training sample *t*,

$$E = \frac{1}{2}(t - y)^2$$

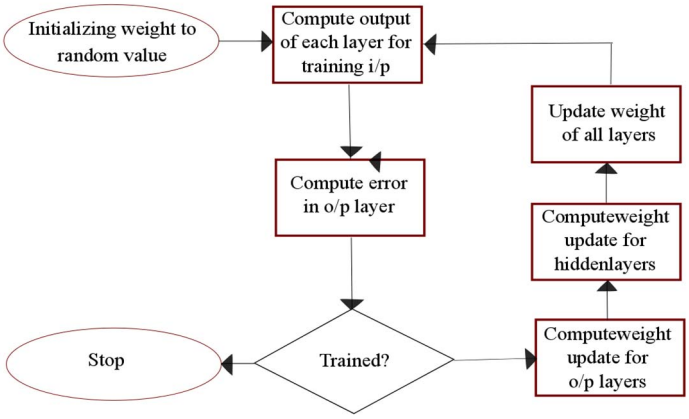

Fig.3. Back propagation training algorithm

V. EXPERIMENTS & RESULTS

The proposed system was developed using MATLAB 8 software. Experiments were conducted on the database created as mentioned in the earlier Section II. The following Fig. 4 shows the system accuracy obtained when taking different number of speakers. In each group of speakers, experiments were conducted varying number of neurons in the hidden layer of the neural network to identify for the best option. From the experimental analysis as shown in Fig. 4, it can be observed that the system with 10 hidden neurons gives the better

accuracy in identifying speaker for the test cases where the no. of speakers is set to 10, 20, 30, 40 and 50.

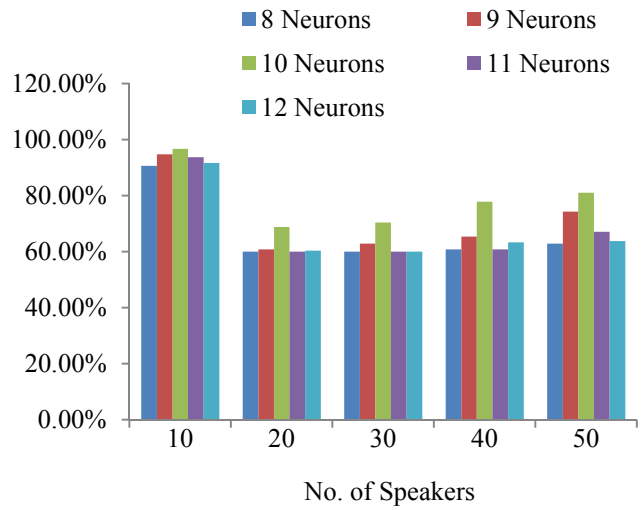

Fig.4. System accuracy for different number of speakers tested against different number of neurons in the hidden layer

VI. CONCLUSION

The purposed system for a novel speaker identification system using feed forward neural networks is implemented and experiments have been conducted to ensure its accuracy. From the experimental result, it can be seen that the accuracy of the system, considering different number of speakers for a particular instance with varying number of hidden neurons in the hidden layer, provides better accuracy compare to the other old techniques. The highest accuracy achieved in identification of speaker is 81.7%. Thus, it can be concluded that, MFCC in combination with artificial neural network can achieve better efficiency in terms of cost and time in voice based speaker identification.

REFERENCES

[1] Seema Khanum and Firos A, "Text Independent Gender Identification in Noisy Environmental Conditions", International Conference on Computing, Communication and Automation (ICCCA2017) , ISBN: 978-1-5090-6471-7/17/$31.00 ©2017 IEEE

[2] Davis, S.B. and Mermelstein, P. (1980), "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. on ASSP, Aug. 1980.

[3] Picone, J.W. (1993), "Signal modeling techniques in speech recognition," Proceedings of the IEEE, 1993, 81(9): 1215-1247

[4] Palit A. K, Popovic D, Computational Intelligence in Time Forecasting Theory and Engineering Applications, 2006, XXII, 372 p., Hardcover, ISBN: 978-1-85233-948-7