Design of Optimal MLP NN for Speaker Dependent Spoken Words Recognition Application

Sneha B. Lonkar
PG Student M.E. (Electronics and Telecom)
Vivekanand Education Society's Institute of Technology, Mumbai

Nadir N. Charniya
Professor (Electronics and Telecom)
Vivekanand Education Society's Institute of Technology, Mumbai

ABSTRACT

Spoken words recognition provides applications like spoken commands recognitions in robotics command, speech based number dialing for phones and mobiles, etc. It also provides applications in railway and banking areas. This work aims at designing of optimal Multilayer Perceptron Neural Network (MLP NN) based classifiers for speaker dependent spoken digits recognition. The classifier attempted as optimal leading to less number of computations and few components requirement for its future implementation in hardware leading to a low cost speech recognition system. Isolated spoken digits were used as an input data to the neural networks based classifiers. Each spoken word was analyzed for the feature like Mel Frequency Cepstral Coefficients (MFCC). The MLP NN based classifier was designed meticulously with the condition of minimum components and attempting maximum classification accuracy.

General Terms

Speech recognition, Back-propagation algorithm

Keywords

Neural Network, Multilayer Perceptron Neural Network, Speech Recognition, Mel Frequency Cepstral Coefficients

1. INTRODUCTION

The objective of this work is to design maximize accuracy neural network for speaker dependent isolated spoken digits recognition under the constraints of minimum network dimension. While applying neural network to any application there are chances of overloading the system with computation time. Therefore, neural networks design attempted with optimal network design configuration. The design comprised of minimum number of weights and neuron components leading to a network with fast speech recognition. Significant work has already been done in the neural network and speech recognition [1] – [27]. Main advantages of neural networks are the ability to generalize results obtained from known situations to unforeseen situations, fast response time in operational phase due to a high degree of structural parallelism, reliability, and efficiency [1]. In the network structure if number of neurons in the hidden layer approaches to the number of input training vectors, it becomes an associative memory and generalization property of such neural networks becomes worst [2]-[4]. Moreover, computation time of the neural network increases dramatically with the number of the hidden neurons [5]. The activation functions and the thresholds are defined by a recursive optimization procedure [6]. It has been reported in literature that three-layer network with sigmoid transfer function in the hidden layer and linear transfer function in the output layer can virtually approximate any nonlinear function to any degree of accuracy provided sufficient number of neurons in the hidden layer is available [7], [8]. It is found that the choice of learning algorithm and activation function is important to improve network performance [9]. The point where the validation set starts to decrease in performance, while the training set continues to increase is the point that gives the maximum training and validation performance [10]. Methodology for design of near-optimal MLP NN based classifier based stopping point from learning curve for training and validation sets is presented by [11]. S V Dudul [12] described training procedure of MLP NN. Meng Joo Er *et al.* [13] presented a general design approach using neural network based classifier for face recognition to cope with small training sets of high-dimensional problem.

Several researchers have used neural network for speech recognition applications. Judith Justin and Ila Vennila [14] analyzed the performance of continuous speech recognition of a speaker independent system using Artificial Neural Network (ANN). Bachu R.G. *et al.* [15] presented his work on speech recognition mentioning Zero Crossing Rate (ZCR) and energy of speech signal as important features of speech signal and classified voiced and unvoiced signals using Energy and ZCR. Bishnu Prasad Das and Ranjan Parekh [16] developed speaker independent isolated English words corresponding to digits zero to nine classification using ANN with feed-forward back-propagation architectures. Maruti Limkar *et al.* [17] presented recognition of spoken English words corresponding to digits zero to nine in an isolated way by different male and female speakers. Md. Ali Hossain *et al.* [18] developed back-propagation neural network for Bangla Speech Recognition. Geeta Nijhawan and M.K. Soni [19] first gave a brief overview of automatic speech recognition (ASR) and then described the use of ANN's as statistical estimators, then, back-propagation neural network performance as applied to the speaker recognition. D. B. Hanchate *et al.* [20] described number recognition system based on ANN. The base system employs MFCC. Q. Ibrahim and N. Abdulghani [21] first described a complete speech recognition system based on Linear Prediction Coefficient (LPC) feature extraction and decision based on ANN. Lakshmi Kanaka Venkateswarlu Revada, *et al.* [22] proposed Network that has been tested on one digit numbers dataset and produced significantly lower recognition error rate in comparison with common pattern classifiers. Wouter Gevaert *et al.* [23] investigated on the speech recognition classification performance which is performed using two standard neural networks structures as the classifier.

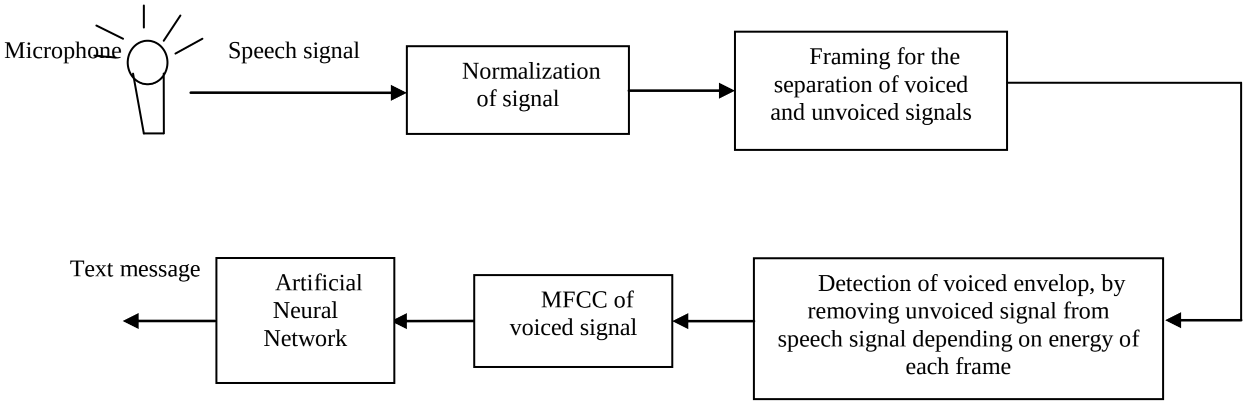
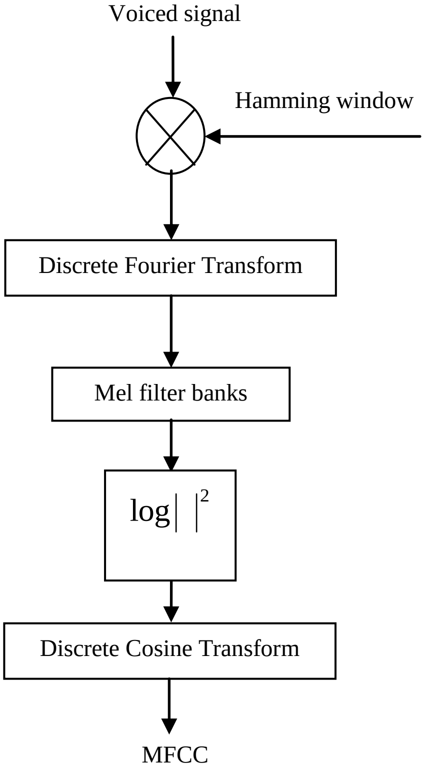

Microphone
Speech signal
Normalization of signal
Framing for the separation of voiced and unvoiced signals
Detection of voiced envelop, by removing unvoiced signal from speech signal depending on energy of each frame
MFCC of voiced signal
Artificial Neural Network
Text message

Fig 1: If necessary, the images can be extended both columns

Chin Luh Tan and Adznan Jantan [24] investigated the use of feed-forward multi-layer perceptrons trained by back-propagation in speech recognition and approached use of neural networks for speaker independent isolated word recognition. Mondher Frikha and Ahmed Ben Hamida [25] described modeling approach ANN used in state of the art speech recognition systems and then provided survey of ANN architecture for robust Speech Recognition. Howard Demuth and Mark Beale [26], [27] presented ANN models of with abbreviated notations.

This paper work aims at classification and recognition of spoken digit from 0 to 9 in a normal minimum noise environment. Features of audio signals like MFCC attempted as inputs patterns to the neural networks. Designing of MLP NN carried out to recognize spoken digits.



\

Fig 2: Block diagram of MFCC

2. METHODOLOGY

Ten records of each spoken digit from 0 to 9 were collected as a database. Block diagram of Speech Recognition System is as shown in Fig. 1. Speech signal from microphone is normalized in between the amplitude of -1 to 1. Normalized signal is divided into frames. Generally, the amplitude of unvoiced speech Segments are much lower than the amplitude of voiced segments. The energy of the speech signal provides a representation that reflects these amplitude variations. Short-time energy of signal *x(n)* can define [15] as:

$$E = \sum_{n=-\infty}^{\infty} [x(n)]^2 \qquad (1)$$

Energy of each frame is calculated. As energy of unvoiced signal is low and energy of voiced signal is high, unvoiced signal removed from speech signal. Voiced signal envelop is detected. As per the literature MFCC is the important features of speech signal. MFCC of voiced signal can be calculated as shown in Fig. 2.

First step to calculate MFCC is windowing. Normally Hamming Window is used [14], [17]. Hamming window can define as:

$$w(n) = 0.54 - 0.46 cos\left(\frac{2\pi n}{L-1}\right) \qquad 0 \le n \le L-1$$

$$= 0 \qquad\qquad otherwise \qquad (2)$$

Where, *L* is window length

After windowing, second step is to take Discrete Fourier Transform to transfer the signal from time domain to frequency domain. Discrete Fourier Transform of signal *x(n)* can define as:

$$X[K] = \sum_{n=0}^{N-1} x(n) \cdot e^{-j2kn\frac{\pi}{N}}$$

$$, 0 \le k \le N-1 \qquad (3)$$

Where, N is number of samples.

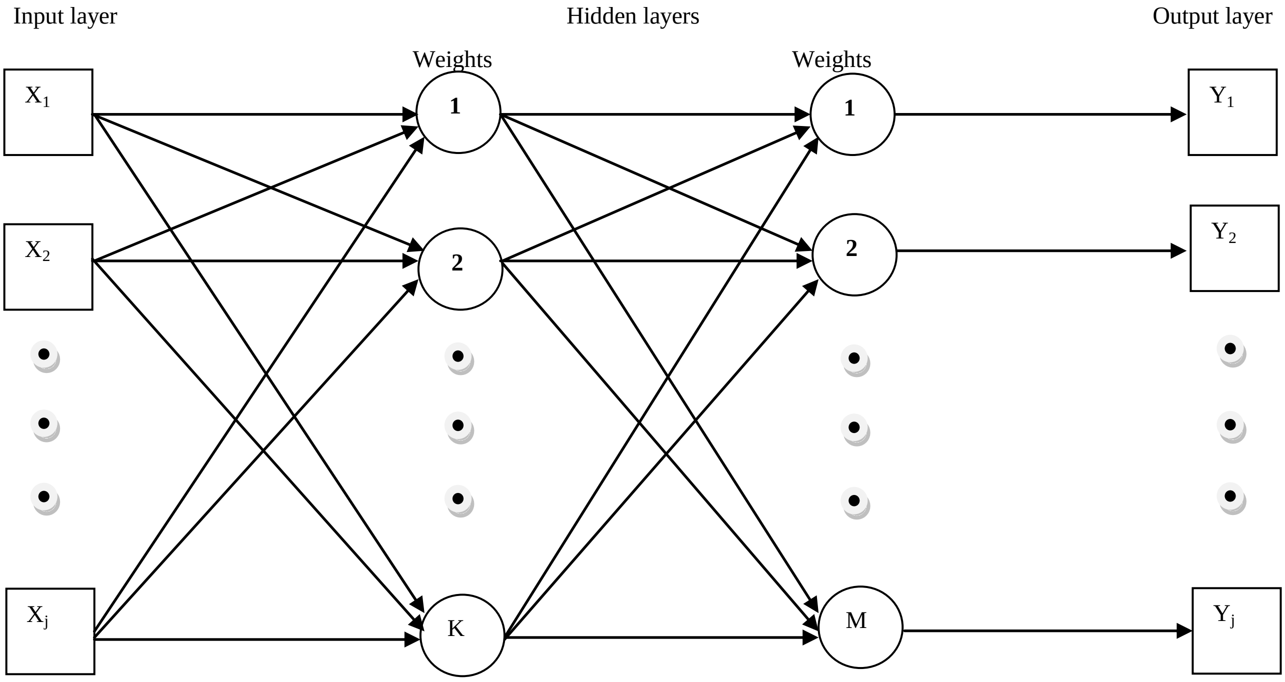


Fig 3: Multilayer neural network

Third step is Mel filter banks. Overlapping triangular windows for frequency domain analysis is defined. Linear to Mel frequency scale conversion can define as:

$$Mel = 2595\log_{10}\left(1+\frac{f}{700}\right) \quad (4)$$

A set of filters with triangular band pass frequency response believed to occur in the auditory system. One filter is assigned for each desired mel-frequency component. Spacing and bandwidth is determined by a constant mel-frequency interval [17].

Next step is to determine log of the energy of signal within each window by summing square of the magnitude of the spectrum.

Final step is to take Discrete Cosine Transform. Discrete Cosine Transform of signal *x(n)* can define as:

$$y(k) = w(k)\sum_{n=1}^{N} x(n)\cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right) \quad (5)$$

Where, range of *k* is from 1 to *N* and *w(k)* can be define as:

$$w(k) = \frac{1}{\sqrt{N}} \qquad \text{for } k = 1$$

$$= \sqrt{\frac{2}{N}} \qquad \text{for } 2 \le k \le N \quad (6)$$

Output of Discrete Cosine Transform is MFCC. Number of MFCC for each record was typically chosen as 20. These MFCC features are given as training input data sets to ANN. ANN is a massively connected structure of artificial neurons that has a natural tendency for strong experiential knowledge and making it available for use. Neural networks are Data Dependent and are suited for non linear applications. An ANN is mans crude way of trying to simulate the brain electronically by using Hardware or Software. There are two aspects. One is knowledge Acquisition through learning and another is knowledge is stored in weights i.e. coefficients. These are two aspects in which ANN resembles the brain. By using ANN it is possible to developed powerful mathematical algorithms, which mimic the human brain's ability to learn. ANN is used when first principle equations of system cannot easily be developed. ANNs are used for Non-linear Regression, Clustering and Classification. It has a capability to learn from examples. For designing neural network data-collection phase must be carefully planned to ensure that data should be sufficient and it should capture the fundamental principles at work and data should be free as far as possible from noise. There are many features of neural networks. Neural network is intrinsically capable of recognizing a complex pattern and modeling non-linear system. It also gives fast response due to a high degree of structural parallelism.

Multilayer neural networks are used. Multilayer neural network is as shown in Fig. 3. These were adapted by back-propagation. Usually a fully connected variant is used, so that each neuron from the n-th layer is connected to all neurons in the (n+1)-th layer, but there are no connections between neurons of the same layer. A subset of input units has no input connections from other units; their states are fixed by the problem. Another subset of units is designated as output units; their states are considered the result of the computation. Units that are neither input nor output are known as hidden units.

In MATLAB, coding is done for designing MLP NN to classify digits from 0 to 9 based on their features. Here, MLP NN consists of 20 inputs and 10 output neurons for 10 digits from 0 to 9. Number of hidden layer neurons was estimated systematically by using following technique. Single hidden layer MLP NN designed by using following technique:

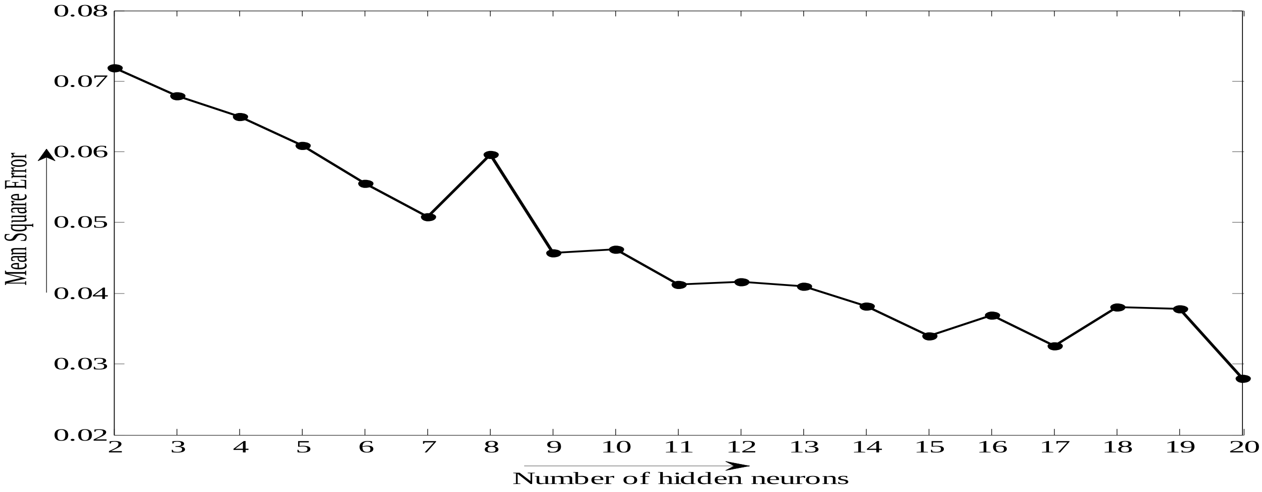


Fig 4: Plot of Mean Square Error verses Number of hidden neurons

- tansigmoidal function for the neurons of MLP NN.
- No of neurons in hidden layer varied from 2 to 20.
- Train the network number of times with different random initialization of connection weights (small values) to avoid biasing.
- Network trained number of times with different random initialization of connection weights (small values) to avoid biasing.

After training, network was validated for testing data sets on the basis of classification accuracy. From the Fig. 4 and Fig. 5 it is found that hidden layer with 12 neurons gave maximum accuracy of 100 percent on training data. When number of hidden neurons is 12 the Mean Square Error is also reasonably low. Therefore, optimal configuration for the isolated spoken digits recognition MLP NN is 20-12-10 with tansigmoidal activation functions in the hidden layer and the output layer. The number of training epochs for the minimal configuration was 89.

3. RESULTS AND CONCLUSION

Training of neural network was set to maximum of 5000 epochs. The transfer function of hidden and output neuron was set to tansigmoid. The number of hidden layer neurons was varied from 2 to 20. Each configuration of neural network was trained three times with random initial weights. The best network with maximum classification accuracy was considered. The plot of Mean Square Error verses number of neurons and plot of accuracy verses number of neurons were obtained as shown in Fig. 4 and Fig. 5 respectively. Our objective is to get maximum accuracy with reasonably low Mean Square Error. The optimal configuration thus obtained was validated for testing data set of the same size. The accuracy on testing data set was found to be 97 percent.

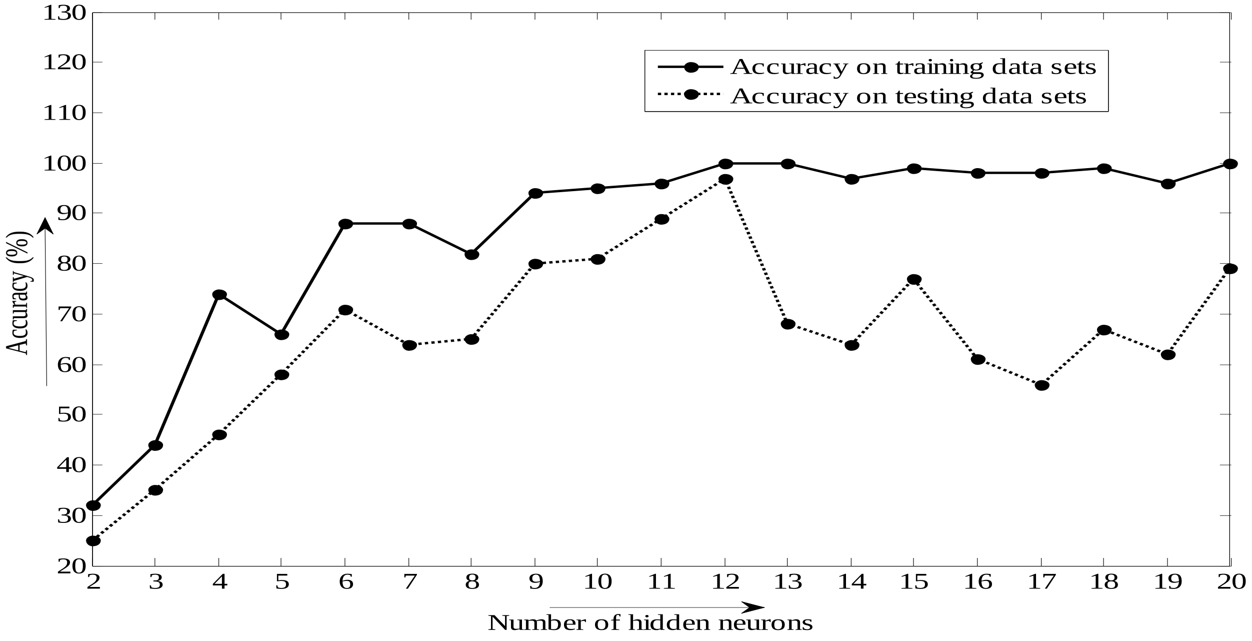


Fig 5: Plot of Accuracy verses Number of neurons on training and testing data sets

4. REFERENCES

[1] J. M. Dias Pereira, Octavian Postolache, P. M. B. Silva Girao, and Mihai Cretu, "Minimizing Temperature Drift Errors of Conditioning Circuits Using Artificial Neural Networks", IEEE transactions on instrumentation and measurement, Vol. 49, pp. 1122-1127, October 2000.

[2] Iryna V. Turchenko, "Simulation Modeling of Mullti-Parameter Sensor Signal Identification Using Neural Networks", second IEEE international conference on intelligent systems, pp. 48-53, June 2004.

[3] Ali Gulbag, Fevzullah Temurtas, Cihat Tasaltin, Zafer Ziya Ozturk, "A study on radial basis function neural network size reduction for quantitative identification of individual gas concentrations in their gas mixtures", ScienceDirect, January 2007.

[4] Henry Leung, Titus Lo and Sichun Wang, "Prediction of Noisy Chaotic Time Series Using an Optimal Radial Basis Function Neural Network", IEEE transactions on Neural Networks, Vol. 12, pp. 1163-1172, September 2001.

[5] Huien Han and Peter Felker, "Estimation of daily soil water evaporation using an artificial neural network", Journal of Arid Environments, pp. 251-260, April 1997.

[6] Pasquale Arpaia, Pasquale Daponte, Domenico Grimaldi and Linus Michaeli, "ANN-Based Error Reduction for Experimentally Modeled Sensors", IEEE transactions on instrumentation and measurement, Vol. 51, pp. 23-30, February 2002.

[7] Guang-Bin Huang, Yan-Qiu Chen and Haroon A. Babri, "Classification Ability of Single Hidden Layer Feedforward Neural Networks", IEEE transactions on Neural Networks, Vol. 11, pp. 799-801, May 2000.

[8] Maxwell Stinchcombe and Halbert White, "Universal approximation using feedforward networks with non-sigmoid hidden layer activation functions", pp. I613-I617.

[9] C.R. Chen, H.S. Ramaswamy, "A neuro-computing approach for modeling of residence time distribution (RTD) of carrot cubes in a vertical scraped surface heat exchanger (SSHE)", Food Research International 33, pp. 549-556, February 2000.

[10] .T. Lewicke, E.S. Sazonov, M.J. Corwin, S.A.C. Schuckers, "Reliable Determnination of Sleep Versus Wake from Heart Rate Variability Using Neural Networks", Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, pp. 2394-2399, August 2005.

[11] Nadir N. Charniya, "Design of Near-Optimal Classifier Using Multi-Layer Perceptron Neural Networks for Intelligent Sensors", International Journal of Modeling and Optimization, Vol. 3, No. 1, pp. 56-60, February 2013.

[12] S V Dudul, "Classification of Radar Returns from the Ionosphere using RBF Neural Network", IE(I) Journal-ET, Vol. 88, pp. 26-33, July 2007.

[13] Christopher M. Bishop, "Neural Network for Pattern Recognition", Indian Edition, Year 2007.

[14] Judith Justin and Ila Vennila, "A hybrid speech recognition system with Hidden Markov Model and Radial Basis Function Neural Network", American Journal of Applied Sciences, pp. 1148-1153, Year 2013.

[15] Bachu R.G., Kopparthi S., Adapa B., Barkana B.D., "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal", pp. 1-7.

[16] Bishnu Prasad Das, Ranjan Parekh, " Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers", International Journal of Modern Engineering Research (IJMER), pp. 854-858, May-June 2012.

[17] Maruti Limkar, RamaRao & VidyaSagvekar, "Isolated Digit Recognition Using MFCC AND DTW", International Journal on Advanced Electrical and Electronics Engineering, (IJAEEE), pp. 59-64, Year 2012.

[18] Md. Ali Hossain, Md. Mijanur Rahman, Uzzal Kumar Prodhan, Md. Farukuzzaman Khan, "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition", International Journal of Information Sciences and Techniques (IJIST), Vol.3, pp. 1-9, July 2013.

[19] Geeta Nijhawan, M.K. Soni, "A Comparative Study of Two Different Neural Models For Speaker Recognition Systems", International Journal of Innovative Technology and Exploring Engineering, Vol.1, pp. 67-72, June 2012.

[20] D. B. Hanchate Mohini Nalawade, Manoj Pawar, Vijay Pophale, Prabhat Kumar Maurya, "Vocal Digit Recognition using Artificial Neural Network", second International Conference on Computer Engineering and Technology, Vol.6, pp. 88-91, Year 2010.

[21] Q. Ibrahim, N. Abdulghani, "Security enhancement of voice over Internet protocol using speaker recognition technique", IET Communications, Vol.6, pp. 604-612, Year 2012.

[22] Lakshmi Kanaka Venkateswarlu Revada, Vasantha Kumari Rambatla and Koti Verra Nagayya Ande, "A Novel Approach to Speech Recognition by Using Generalized Regression Neural Networks", International Journal of Computer Science Issues, Vol. 8, pp. 484-489, March 2011.

[23] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "Neural Networks used for Speech Recognition", Journal of Automatic Control, University of Belgrade, Vol. 20, pp. 1-7, Year 2010.

[24] Chin Luh Tan and Adznan Jantan, "Digit Recognition using Neural Networks", Malaysian Journal of Computer Science, Vol. 17, pp. 40-54, December 2004.

[25] Mondher Frikha, Ahmed Ben Hamida, "A Comparitive Survey of ANN and Hybrid HMM/ANN Architectures for Robust Speech Recognition", American Journal of Intelligent Systems, pp. 1-8, 2012.

[26] Martin T. Hagan, Howard B. Demuth, Mark H. Beal, "Neural Network Design", Campus Pub. Service, University of Colorado Bookstore, Year 2002.

[27] Howard Demuth, Mark Beale, "Neural Network Toolbox, version4", The MathWorks, Year 2012.