Speaker Identification by using Artificial Neural Network in MATLAB

Priya Dubey, K.K. Nayak

Mtech Scholar, B.I.S.T. Bhopal

Abstract - The term speaker identification or voice recognition refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process. Implemented research work yields with a novel, robust, secure and highly efficient machine learning based approach for speech recognition. Given approach takes the advantage of MFCC algorithm in Artificial Neural network for classifying the voices. It works smoothly with multiple voice samples and also with ambiguous samples. The recognition process is followed by recording of 32 samples each of 5 seconds and feature extraction of that voice samples. There after training of neural network takes place and the resulted trained network is used for recognizing individual speeches. Whole work has been simulated and demonstrated with MATLAB. The method implemented here worked perfectly and efficiently with improved security features (i.e. it is based on learning not on comparison). In this section the outcomes of the implementation has been explained on implementing speaker identification by using MFCC Algorithm in Artificial Neural Network. Different techniques can be analysed behind speech recognition, as well as techniques for using neural networks efficiently. While the scope of this work has been reduced to an isolated word recognition network, the results were still very positive. Despite limiting the speech recognition side of the dissertation, Here work have given an understanding of how neural networks can tackle a problem like speaker recognition, as well as the benefits of certain structures and training algorithms. In the end, this work is been accomplished successfully and implemented speaker identification with a neural network.

Keywords: Artificial Intelligence (AI), Automatic Speech Recognition (ASR), Artificial Neural Network (ANN), Mel Frequency Cepstral Coefficient Algorithm (MFCC), Machine Intelligence (MI), MATLAB, Speech Recognition (SR).

I. INTRODUCTION

Speech recognition is a very popular research area in the field of machine intelligence. There are many reasons for automatic speech recognition being widely developed by engineers and scientists around the world. Human-machine interaction is one of the most important reasons. We always dream of ordering machines such as the TV to turn itself on and change channels per our orders, thermostats to adjusting the temperature by them to adapt to a human's preferences, or even a robot babysitter to do all the house tasks fast and efficiently. The basic sensory stages of the human-machine interaction are vision recognition and speech recognition. Voice recognition, which is a special kind of speech recognition, is widely used in high security locations. Due to the high demand in the current market, many corporations have already built some automatic speech recognition systems: like the dictation system used by IBM and the telephone transaction system used by T-Mobile, AT&T and Philips. Some of the "smart" recognition systems can recognize a word, a sentence or even a paragraph but require to be adapted to every new user, so every new user needs to train the system to recognize his/her specific voice. This approach is not suitable or feasible for a commercial use. These problems lead researchers and scientists to improve the speech recognition systems.

II. SYSTEM MODEL

Speech recognition is a multileveled pattern recognition task, in which acoustical signals are examined and structured into a hierarchy of sub word units (e.g., phonemes), words, phrases, and sentences. Each level may provide additional temporal constraints, e.g., known word pronunciations or legal word sequences, which can compensate for errors or uncertainties at lower levels. This hierarchy of constraints can best be exploited by combining decisions probabilistically at all lower levels, and making discrete decisions only at the highest level.

For designing of neural network we have sampled four words North, South, East and West and we extracted 1274 different inputs from those four words and then trained them by using Mel frequency Cepstral Coefficients algorithm.

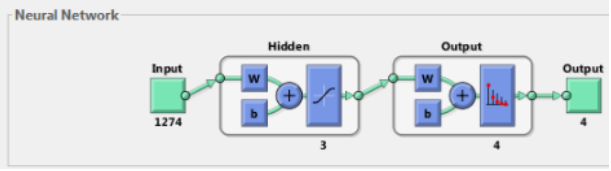

Fig. 1 Standard Speech Recognition Model

The elements are as follows:

Raw speech: Speech is sampled at a standard frequency of 16 KHz over a microphone. This sampling yields a sequence of amplitude values over time [16].

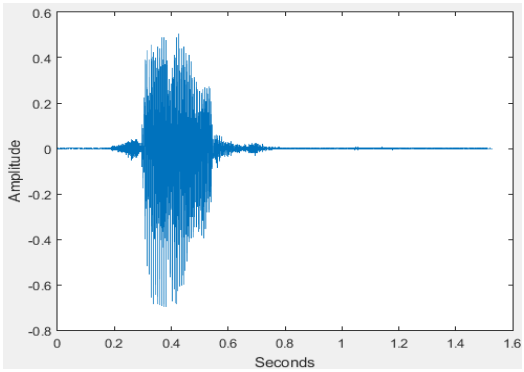

Fig. 2 Raw speech of word East

Signal analysis: Sampled raw speech must be transformed and compressed in order to simplify the recognition process. There are several popular techniques that extract features from raw speech and compress the data without loss of data. Fourier analysis, Perceptual Linear Prediction, Linear Predictive Coding and Cepstral analysis all properly process the raw speech into a more usable state [17].

Speech frames: Once raw speech is processed and analysed, the audio is broken up into speech frames. These frames are typically 10ms intervals of the processed audio and provide unique information relative to the speech recognition process [18].

Performance Plot: This topic presents part of a typical multilayer network workflow. When the training in train and apply multilayer neural networks is complete, you can check the network performance and determine if any changes need to be made to the training process, the network architecture, or the data sets. First check the training record and second argument returned from the training function.

Training State: It is very difficult to know which training algorithm will be the fastest for a given problem. It depends on many factors, including the complexity of the problem, the number of data points in the training set, the number of weights and biases in the network, the error goal, and whether the network is being used for pattern recognition (discriminant analysis) or function approximation (regression). This section compares the various training algorithms. Feed forward networks are trained on six different problems. Three of the problems fall in the pattern recognition category and the three others fall in the function approximation category. Two of the problems are simple "toy" problems, while the other four are "real world" problems. Networks with a variety of different architectures and complexities are used, and the networks are trained to a variety of different accuracy levels.

Error Histogram: A two layer feed forward network with sigmoid hidden neurons and linear output neurons can fit multi-dimensional mapping problems arbitrarily well, given consistent data and enough neurons in its hidden layer.

Confusion Matrix: In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, it is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class (or vice versa). The name stems from the fact that it makes it easy to see if the system is confusing two classes.

Fig. 3 Confusion Matrix

Receiver Operating Characteristic: In statistics, a receiver operating characteristic (ROC), or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection in machine learning.

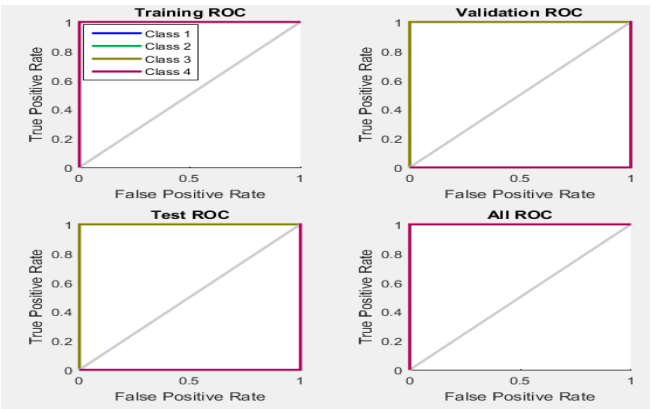

Fig. 4 Receiver Operating Characterstics

The false-positive rate is also known as the fall-out or probability of false alarm and can be calculated as (1 − specificity). The ROC curve is thus the sensitivity as a function of fall-out. In general, if the probability distributions for both detection and false alarm are known, the ROC curve can be generated by plotting the cumulative distribution function (area under the probability distribution from -∞ to the discrimination threshold) of the detection probability in the y-axis versus the cumulative distribution function of the false-alarm probability in x-axis.

III. PREVIOUS WORK

In this previous work the system is less efficient due to hidden model but in implemented paper proposes the classification model which serves better results than older one. Mel Frequency Cepstral Coefficient Algorithm serves for best results in the implemented work.

IV. PROPOSED METHODOLOGY

Proposed Methodology Gives the advantage of artificial neural network for classifying the voices. It works smoothly with multiple voice samples. The recognition process is followed by recording of 32 samples each of 5 seconds and feature extraction from that voice samples. Thereafter training of neural network takes place and the resulted trained network is used for recognizing individual speeches. Whole work has been simulated and demonstrated with MATLAB® 2014a. The method implemented here worked perfectly and efficiently with improved security features.

V. SIMULATION/EXPERIMENTAL RESULTS

In the following section, we have described Result analysis which illustrates the simulation graph of the word North, South, East and West by different speakers. This voice samples are firstly recorded by recorder then converted into .wav format and then preprocessed into MFCC algorithm. For training purpose after it preprocessed it goes into classification model for further finalization. In figure 5 shown below the Performance of network for training, validation and test sets. The trained network that did best on the validation set.

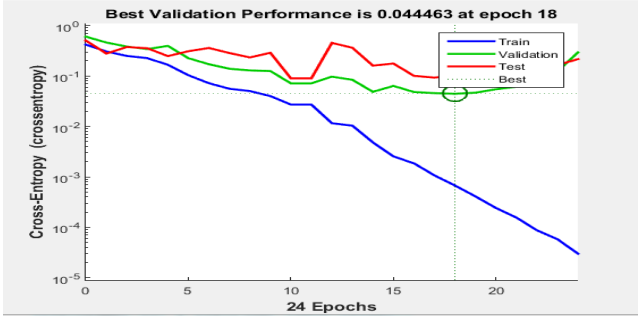

Fig. 5 Performance Graph

This will give us a sense of how well the network will do when applied to data from the real world. In figure 5 shown below the word east is been revealed by speaker 1. In figure 6 shown below its clearly represented how the variations of word east is been look after processing though MFCC algorithm.

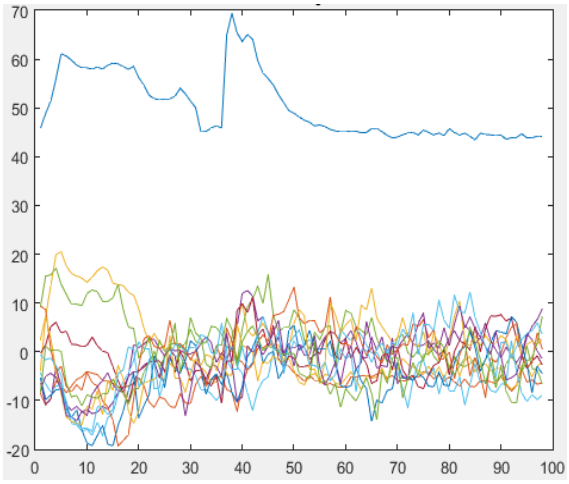

Fig. 6 MFCC of East Speaker 1

VI. CONCLUSION

In this section the outcomes of the implementation has been explained on implementing speaker identification by using MFCC Algorithm in Artificial Neural Network. Different techniques can be analyzed behind speech recognition, as well as techniques for using neural networks efficiently. While the scope of this work has been reduced to an isolated word recognition network, the results were still very positive. Despite limiting the speech recognition side of the dissertation, Here work have given an understanding of how neural networks can tackle a problem like speaker recognition, as well as the benefits of certain structures and training algorithms. In the end, this work is been accomplished successfully and implemented speaker identification with a neural network.

VII. FUTURE SCOPES

In future an understanding of how speech recognition works with artificial neural networks can expand the implementation. The developed algorithm in this dissertation can be further extended to implement some more difficult tasks like recognizing phonemes or pattern. Overall, results are good but can be further be improved by considering more features.

REFERENCES

[1] Nawel Souissi and Adnane Cherif, “Speech Recognition System based on Short-Term Cepstral Parameters, Feature Reduction Method and Artificial Neural Networks” IEEE 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2016

[2] Deepak Baby, Tuomas Virtanen, Jort F. Gemmeke and Hugo Van Hamme, “Coupled Dictionaries for Exemplar-

Based Speech Enhancement and Automatic Speech Recognition" IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23, No. 11, November 2015

[3] Dipali Bansal, Neelam Turk and Sunanda Mendiratta, "Automatic Speech Recognition by Cuckoo Search Optimization based Artificial Neural Network Classifier" IEEE International Conference on Soft Computing Techniques and Implementations (ICSCTI), 2015

[4] Bassam M. El-Zaghmouri, "Speech Recognition Using Neural Networks" International Conference on Computing, Communication and Control Engineering, 2015

[5] Shaofei Xue, Ossama Abdel-Hamid, Hui Jiang, Lirong Dai and Qingfeng Liu, "Fast Adaptation of Deep Neural Network based on Discriminant Codes for Speech Recognition" IEEE/ACM Transactions on Audio, Speech and Language Processing, VOL. 22, No. 12, 2014

[6] Ossama Abdel-Hamid, Abdel-Rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu, "Convolutional Neural Networks for Speech Recognition" IEEE Transactions on Audio, Speech, and Language Processing, Vol. 22, No. 10, October 2014

[7] Seyed Reza Shahamiri, and Siti Salwah Binti Salim, "A Multi-Views Multi-Learners Approach Towards Dysarthric Speech Recognition Using Multi-Nets Artificial Neural Networks" IEEE Transactions on Neural Systems and Rehabilitation Engineering, VOL. 22, NO. 5, 2014

[8] Bo Li, and Khe Chai Sim, "A Spectral Masking Approach to Noise-Robust Speech Recognition Using Deep Neural Networks" IEEE/ACM Transactions on Audio, Speech, and Language Processing, VOL. 22, No. 8, 2014

[9] Qirong Mao, Mingdong, Zhengwei Huang and Yongzhao Zhan, "Learning Salient Features for Speech Emotion Recognition using Convolutional Neural Networks" IEEE Transactions on Multimedia, VOL. 16, No. 8, 2014

[10] Li Deng, Geoffrey Hinton and Brian Kingsbury, "New Types of Deep Neural Network Learning for Speech Recognition and Related Applications: An Overview" IEEE International Conference on Acoustics, Speech and Signal Processing, 2013

[11] Jasdeep Singh Bhalla and Anmol Aggarwal, "Using ADABOOST Algorithm along with Artificial Neural Networks for Efficient Human Emotion Recognition from Speech**"** IEEE International Conference on Control**,** Automation**,** Robotics and Embedded Systems, 2013

[12] Shivam Jain, Preeti Jha and Suresh. R, "Design and Implementation of an a Automatic Speaker Recognition System using Neural and Fuzzy Logic in Matlab**"** International Conference on Signal Processing and Communication **(**ICSC), 2013

[13] George E. Dahl, Dong Yu, Li Deng and Alex Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition" IEEE Transactions on Audio, Speech, and Language Processing, VOL. 20, No. 1, 2012

[14] Ossama Abdel-Hamid, Abdel-Rahman Mohamed, Hui Jiang and Gerald Penn, "Applying Convolutional Neural Networks Concepts to Hybrid NN-Hmm Model for Speech Recognition" IEEE International Conference on Acoustics, Speech and Signal Processing, 2012

[15] Vikramjit Mitra, Hosung Nam, Carol Y. Espy-Wilson, Elliot Saltzman, and Louis Goldstein, "Articulatory Information for Noise Robust Speech Recognition" IEEE Transactions on Audio, Speech, and Language Processing, Vol. 19, No. 7, September 2011

[16] Chenghui Yang, Weixin Yang and Shuwen Wang, "Based on Artificial Neural Networks for Voice Recognition Word Segment" IEEE 3rd Conference on Communication Software and Networks (ICCSN), 2011

[17] Georgi T. Tsenov and Valeri M. Mladenov, "Speech Recognition Using Neural Networks" IEEE 10th Symposium on Neural Network Application in Electrical Engineering, 2010

[18] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "Neural Networks used for Speech Recognition" IEEE Journal of Automatic Control, Vol. 20:1-7, 2010

[19] Gulin Dede and Murat Husnu Sazlı, "Speech recognition with artificial neural networks" Digital Signal Processing 20 (2010) 763–768, Elsevier, 2009

[20] Min-Lun Lan, Shing-Tai Pan and Chih-Chin Lai, "Using Genetic Algorithm to Improve the Performance of Speech Recognition Based on Artificial Neural Network**"** First International Conference on Innovative Computing**,** Information and Control - Volume I, 2006

[21] Yonghong Yan, "Understanding Speech Recognition using Correlation-Generated Neural Network Targets" IEEE Transactions on Speech and Audio Processing, VOL. 7, No. 3, 1999