Journal Of Al Azhar University Engineering Sector

Vol. 12, No. 44, July, 2017, 886-893

COMPUTATIONAL INTELLIGENT ALGORITHMS FOR ARABIC SPEECH RECOGNITION

Mohamed Sadek Ahmed Rashwan, Reda Hessian Abo Al Ez, Gamal Abd El reheem and GamalAbd El reheem

B.Sc. Of Computer and Systems Engineering Dept.

Al Azhar University, Faculty of Engineering, Computer and Systems Engineering

ABSTRACT

One of the most term in the biometric technologies is the voice recognition. Speech recognition is a complicate problem .But in today's society when technology is consistently striving for voice driven implementation, speech recognition can be a very useful system .This paper aims to build a system of voice recognition using two systems one of them is based on Fourier descriptors of speech and the other is Artificial neural network which is using back propagation algorithm in neural networks, The first speech recognition system is using Fourier descriptor of speech to get four formants of any speech signal. The second intelligent system is built neural network which is trained and tested using data sits of formants of speech signals. After we applied two systems for speech, the first method is not accurate because it depends on the human element in comparison and identification. This method makes it possible to have the error ratio greater than the second method, which is smart and accurate and depends on the intelligent system stable and continuous and has no life As well as not influenced by the external influences that might control the decision in the first method.

KEYWORDS : Voice Recognition .Fast Fourier Transform (FFT) ,Artificial Neural Network (ANN) ,Back-Propagation .

1. INTRODUCTION

Speech and voice recognition are the emerging scope of security and authentication for the future. Now a days text and image passwords are prone to attacks. In case of the most commonly used text passwords, users are required to handle different passwords for emails, internet banking, etc. Hence they tend to choose password such as they are easy to remember . But they are vulnerable in case of hackers. In case of image passwords, they are vulnerable to shoulder surfing and other hacking techniques. Advances in speech technology have created a larg interest in the practical application of speech recognition. Therefore this system provides the users with the appropriate and efficient method of authentication system based on voice recognition.

Voice is also physiological trait because every person has different pitch, but voice recognition is mainly based on the study of the way a person speaks, commonly classified as behavioral. Speaker verification focuses on the vocal characteristics that produce speech and not on the sound or the pronunciation of speech itself. The vocal characteristics depend on the dimension of the vocal tract, mouth, nasal cavities and the other speech processing mechanism of the human body .

Voice-biometrics system can be categorized as belonging in two industries : speech processing and biometric security. This dual parentage has strongly influenced how voice biometrics tools operate in the real world. Speech processing. Like other speech processing tools, voice biometrics extract information from the stream of speech to accomplish their work .

In this research we will use two techniques for speech recognition first using Fourier transform second using neural network .

2. The speech signal

The process from human speech production to human perception between the speaker and listener as shown:

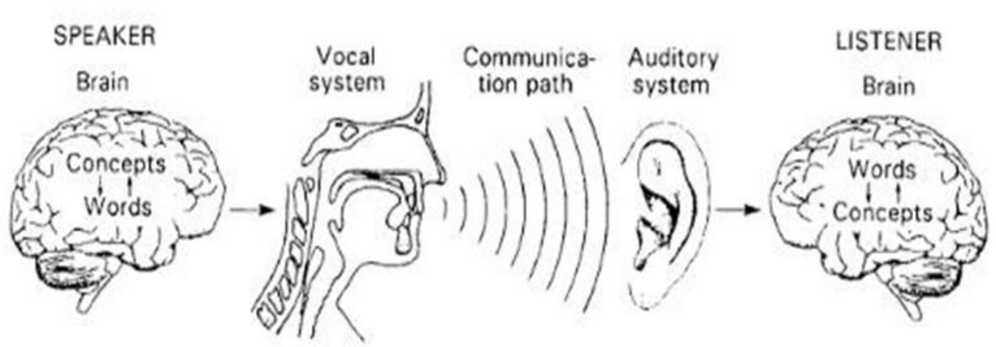


Fig (1) speech system

3. SIGNAL PRODUCTION

Speech is produce by air-pressure waves emanating from the mouth and the nostrils of speaker .The main organs involved into the speech production process are the lungs, larynx, nose and various parts of the mouth. From the lungs is modulated in different ways to produce the acoustic in the audio frequency range .

4. SPEECH REPRESENTATION

Representation of the speech signal based on short –time Fourier analysis is spectrogram . The spectral maxima of speech signal are called **formants**. They are formed because of the resonances which happen in the vocal tract during speech generation process. The formants (resonance frequencies) depend on the geometrical sizes and shape of vocal tract (head with all the cavities and organs) .

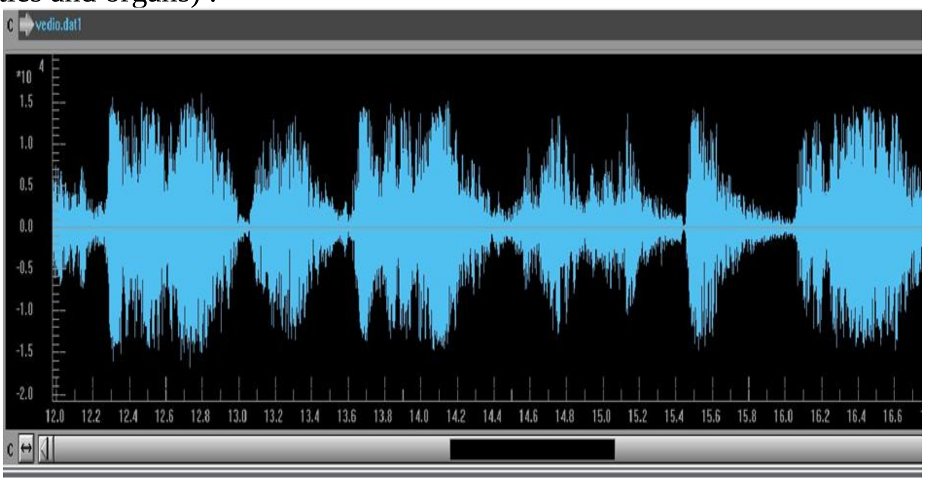
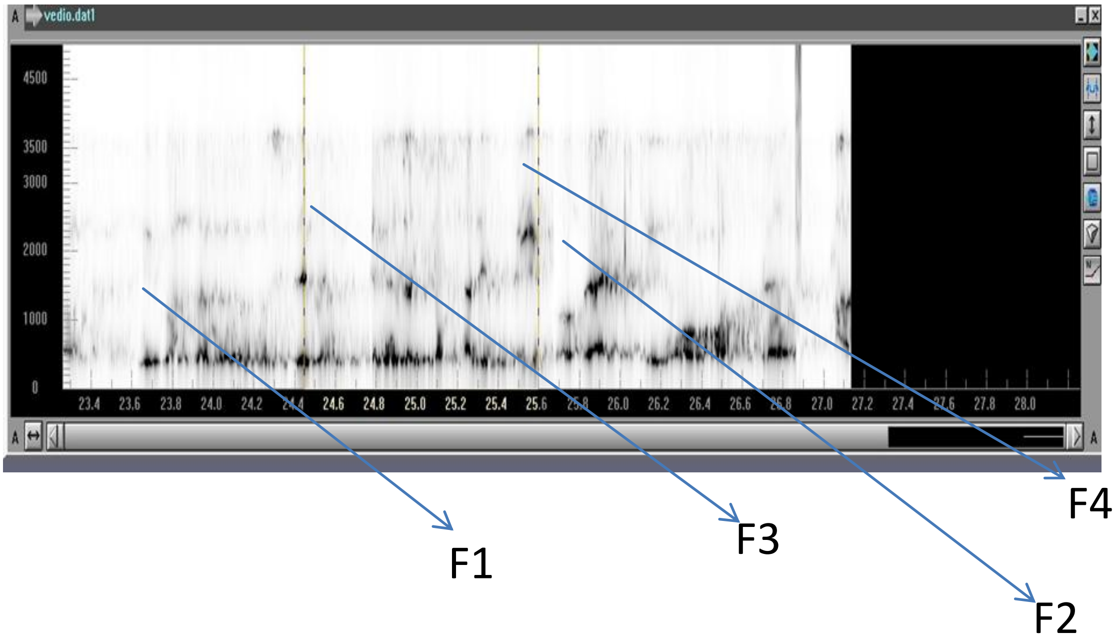


Fig (2) speech signal



Fig(3) The spectrogram of recorded speech

5. ARABIC SPEECH RECOGNITION SYSTEMS

1- **Speech experts in Egypt**
2- **FFT(Fast Fourier transform)**
3- **Neural Network**

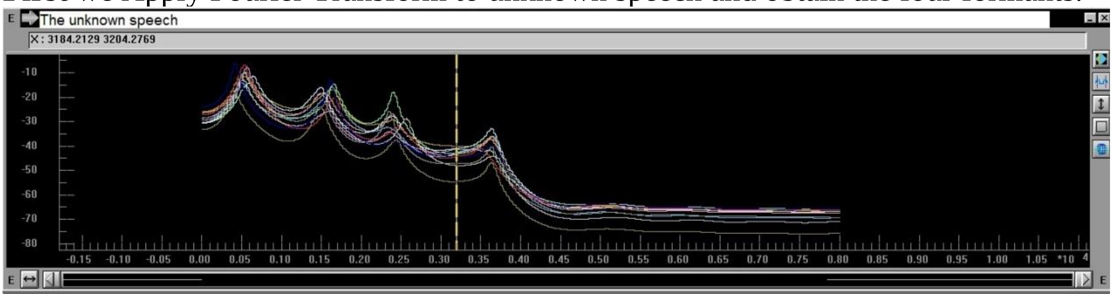
5.1 Speech Experts in Egypt

This method depends on the experience accumulated over the years and expert and increase its ability to distinguish human voices heard only through the establishment of a data base in his brain can identify people through their votes only. This way is the only way for a long time commonly referred to by the judicial bodies in the chapter on many issues based on the comparison of the votes.

5.2 FFT((Fast Fourier Transform)

Using Fourier Transform to transform Time domain to Frequency domain. We apply Fourier Transform to speech signal to find four formants of speech which is unique for everyone and we use this four formants to compare between speech signals .let us now apply Fourier Transform to case study to obtain the results . Here we have unknown speech and two suspect speech and we want to known which one of suspect belong the known speech.

First we Apply Fourier Transform to unknown speech and obtain the four formants.

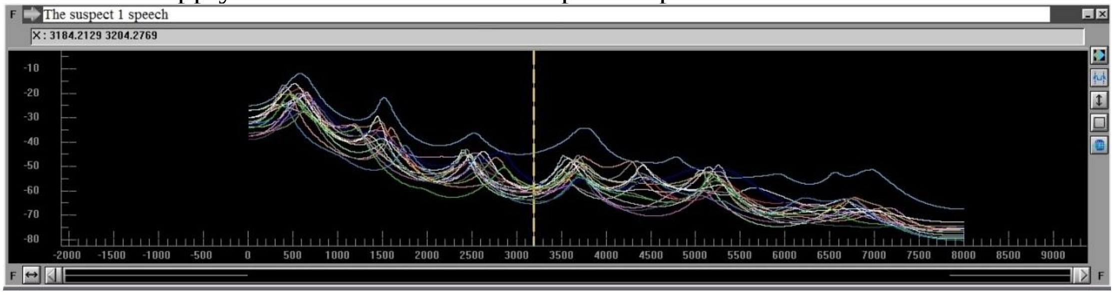


Fig(4) FFT for unknown speech

From this shape we obtain four formants:
F1= 500 Hz
F2=1500 Hz
F3=2500 Hz

F4=3500 Hz

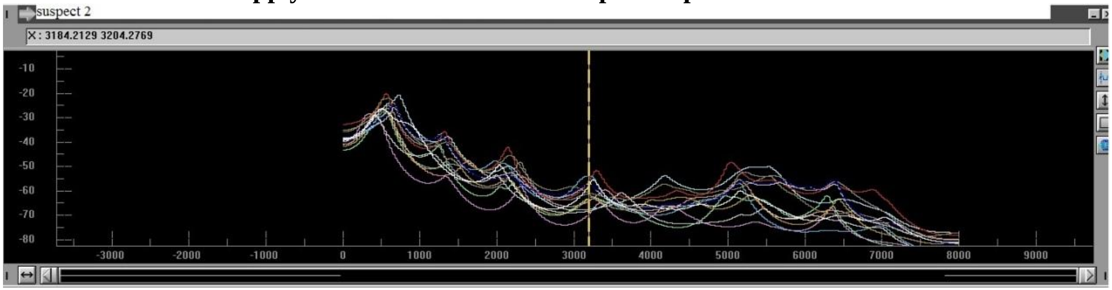
Second we Apply Fourier Transform to suspect 1 speech and obtain the four formants.



Fig(5) FFT for suspect 1

From this shape we obtain four formants:
F1= 500 Hz
F2=1500 Hz
F3=2500 Hz
F4=3500 Hz

Third we Apply Fourier Transform to suspect 2 speech and obtain the four formants.



From this shape we obtain four formants:
F1= 600 Hz
F2=1300 Hz
F3=2100 Hz
F4=3100 Hz

From the value of formants we find the value of four formants of suspect match with the four formants of unknown .

The results for speech recognition using Fourier Transform

The unknown formants speech match with suspect 1 .

5.3 Arabic speech recognition system using Neural Network

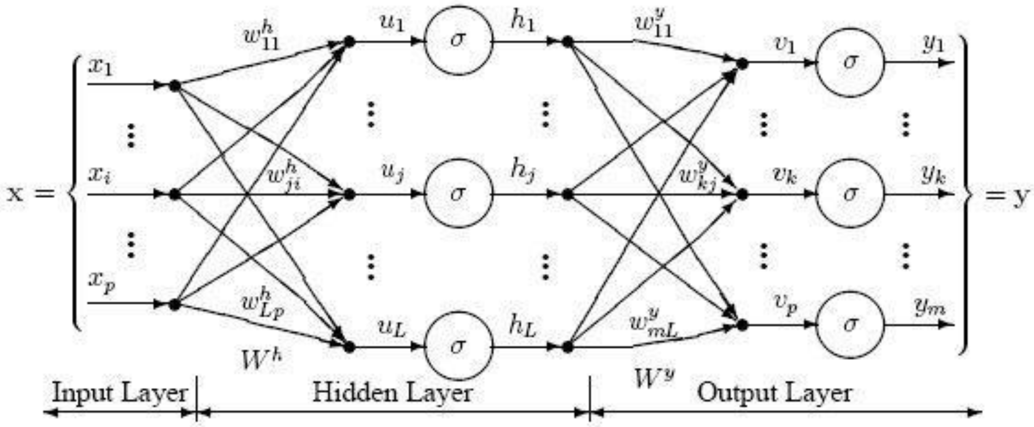
5.3.1 NEURAL NETWORK BASED MODEL

An Artificial Neural Network (ANN) is an information processing system that is inspired by the way biological nervous system works, such as the brain, process information. The key element of this paradigm is the structure of the information processing system. It consists of a large number of highly interconnected processing elements (neurons) work to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. A neural network is a system of parallel processors connected together as a directed graph. Schematically, each processing element (neuron) of the network is represented as a node. These connections provide a hierarchical structure trying to emulate the physiology of the brain for processing new models to solve specific problems in the real world. What is important in developing neural networks is their useful behavior by learning to recognize and apply relationships between objects and patterns of objects specific to the real world. In this respect neural networks are tools that can be used to solve difficult problems

.Artificial neural networks are inspired by the architecture of the biological nervous system, which consists of a large number of relatively simple neurons that work in parallel to facilitate rapid decision-making. The ANN is a universal computation algorithm that has the ability to compose complex hypotheses that can explain a high degree of correlation between features without any prior information from the data set A neural network is created and the Speech frequencies are given as the inputs. Here the Artificial Neural Network has to compare the scanned Speech frequencies with the stored frequencies and identify the person from the frequencies. For doing this the nodes in the neural network has to learn and store the features of the speech each time frequencies is being given as input. Aim of this study is to make Neural Network learn faster and accurately identify the person. In some cases if the nodes are forced with a mass data, the network instead of learning may adopt rote learning or memorize. This is just like the children resort to repeat what they read from the book verbatim from their memory instead of acquiring the abstract of the concept and interpret in their own way.

5.3.2 Input layer

The input layer or the processing stage before the input layer standardizes the input values so that the range of each value is -1 to 1. The predictor variable values (x1…x4) are given as input values. Then the values are distributed to each neurons of the hidden layer. In addition to the predictor values, a constant input of 1.0 is presented to each hidden layers which is called as bias. The bias is multiplied by a weight and added to the sum of values going into the neuron.



Fig(6) Arabic speech recognition system

5.3.3 Hidden Layer

The value from each input neuron is multiplied by a weight (wij), and the resulting weighted values are added together giving a combined value uj. Then the combined value (uj) is put into a transfer function σ which gives an output value hj. The output from the hidden layer is given to the output layer.

5.3.4 Output Layer

Given a neuron in the output layer, the outputs from each hidden layer neurons are multiplied by a weight (wkj) and the resulting weighted values are added together producing a combined value Vj. Then the weighted sum (Vj) is fed into a transfer function, σ, which gives an output value yk. The values are considered as the outputs of the network.

5.3.5 Training a Multilayer Neural Network

The goal of the training process is to find the set of weight values that will cause the output from the neural network to match the actual target values as closely as possible. There are several issues involved in designing and training a multilayer neural network:

- Selecting how many hidden layers to use in the network. We will use 2 Hidden layer .
- Deciding how many neurons to use in each hidden layer.we will use 10 Neurons .
- Validating the neural network to test for over fitting.

5.3.6 Selecting the Number of Hidden Layers

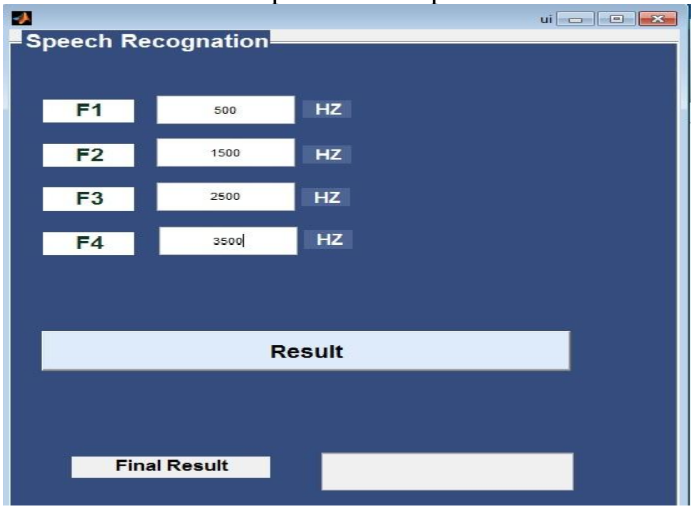
For most of the problems, one hidden layer is enough. For modeling data with discontinuities such as saw tooth patterns may require two hidden layers. Using two hidden layers may rarely improve the model.

5.3.7 Parameters

trainlm is a network training function that updates weight and bias values according to Levenberg-Marquardt optimization. trainlm is often the fastest backpropagation algorithm in the toolbox, and is highly recommended as a first-choice supervised algorithm, although it does require more memory than other algorithms.

- 10 neurons in the hidden layer.
- 2 hidden layers.

We will use 1000 person for training neural network. we will use 70% of data for training and 15% for validation and 15% for test it works and ready for use.
We will input four formants of unknown person in input view:



Fig(7) Input view for Neural Network

After that we will press on result button the neural network running as fig below:

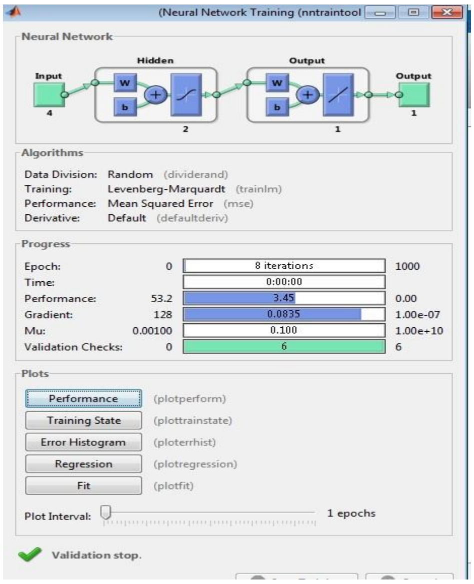
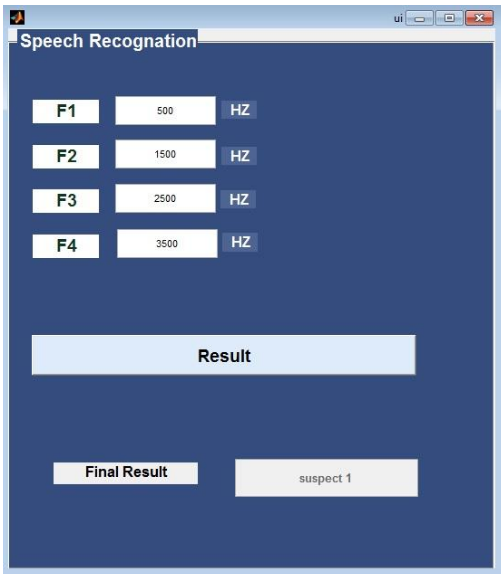


Fig (8) Neural Network running for recognition

We will get result as below:



Fig(9) The result of recognition

The formants match with suspect 1.

Results of speech recognition using Neural Network

From third techniques we obtain the result that the unknown speech for suspect 1 .

6. CONCLUSION

I started this thesis set of techniques of implementing a speech recognition system with Experts human , Fourier Transform and an artificial neural network. I wanted to research different techniques behind speech recognition, as well as techniques for using neural networks efficiently.The results were still very positive. Despite limiting the speech recognition side of the project, I gained an understanding of how neural networks can tackle a problem like speech recognition, as well as the benefits of certain structures and training algorithms. In the end, I accomplished what I set out to and successfully implemented speech recognition with a neural network and Fourier Transform .

REFERENCES

1. Fallahnezhad M, Moradi MH, Zaferanlouei S. A Hybrid Higher Order Neural Classifier for handling classification problems, Expert Systems with Applications. 2011; 38(1): 386-393.
2. I. R. Titze, Principles of Voice Production. Prentice Hall, 1994
3. B. Busse et al, "Single-Synapse Analysis of a Diverse Synapse Population: Proteomic Imaging Methods and Markers," Neuron, vol. 68, pp. 639-653. November, 2010.
4. Kiyoshi Kawaguchi. (2000, June, 17). Biological Neural Networks [Online]. Available: http://www.ece.utep.edu/research/webfuzzy/docs/kk-thesis/kk-thesis-html/node10.html
5. Albrecht Schmidt. (2000). Biological Neural Networks [Online]. Available: http://www.teco.edu/~albrecht/neuro/html/node7.html
6. Rodolfo R. Llinas, “The contribution of Santiago Ramon y Cajal to functional neuroscience,” Nature Reviews Neuroscience, vol. 4. January 2003.
7. Kiyoshi Kawaguchi. (2000, June, 17). The McCulloch-Pitts Model of Neuron [Online].Available:http://wwwold.ece.utep.edu/research/webfuzzy/docs/kkthesis/kk-thesis-html/node12.html
8. Informatics V – Scientific Computing. Pendulum Project [Online]. Available: http://www5.in.tum.de/wiki/index.php/Pendulum_Project
9. Andrew Blais, “An introduction to neural networks – Pattern learning with the back-propagation algorithm,” IBM Developer Works. July 2001.

10. Anil K, Jain, “Artificial Neural Networks: A Tutorial,” Michigan State University, MI.0018-9162/96/$5.00, March 1996.
11. Lawrence Davis, David J. Montana, “Training Feedforward Neural Networks Using Genetic Algorithms,” BBN Systems and Technologies Corp.
12. Fundamentals of Neural Networks. [Online] Available: http://www.myreaders.info/08_Neural_Networks.pdf
13. Roman Bertolami et al, “A Novel Connectionist System for Unconstrained Handwriting Recognition,” CH-6928 Manno-Lugano, Switzerland. May 9, 2008.
14. Andries P. Engelbrecht, Computation Intelligence: An Introduction, 2nd ed. England: John Wiley & Sons Ltd, 2007.
15. Simon Haykin, Neural Networks A Comprehensive Foundation, 2nd ed. Pearson Prentice Hall, 1999.
16. Speech Recognition. [Online] Available: http://www.learnartificialneuralnetworks.com/speechrecognition.html#neuralnetwork
17. Recurrent Neural Networks,” Department of Computer Science, Washington University. [17] Yuliang Feng; Xiange Sun, "Analysis and processing speech signal based on MATLAB,"
18. Electrical and Control Engineering (ICECE), 2011 International Conference on, vol., no., pp.555, 556, 16-18 Sept. 2011
19. Yasunari Yoshitomi, Taro Asada and Masayoshi Tabuse (2011). Vowel Judgment for Facial Expression Recognition of a Speaker, Speech Technologies, Prof. Ivo Ipsic (Ed.), ISBN: 978-953-307-996-7
20. Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit3, “Isolated Speech Recognition Using MFCC and DTW,” International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Vol. 2, Issue 8, August 2013.
21. Mark Gales, Steve Young, “The Application of Hidden Markov Models in Speech Recognition,” Foundations and Trends in Signal Processing. Vol. 1, No. 3, 2007.
22. Malcolm Slaney, Auditory Toolbox. [Online] Available: https://engineering.purdue.edu/~malcolm/interval/1998-010/
23. Jeff Heaton, Introduction to Neural Networks with Java, 2nd ed. Heaton Research, Inc. 2008.
24. Md. Sahidullah and Goutam Saha. 2012. “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition.” Speech Communication 54, 4 (May 2012), 543-565.