Speech Classification Based on FFT and ANN Algorithms

Shaima A. Abushaala, Salma N. Ben Hmeida, Laila Y. Fannas
Department of Information Technology
Libya
sh3ah_it@yahoo.com

ABSTRACT: *This research presents a speech classification based on feature extraction using Fast Fourier Transform (FFT) and classified by Artificial Neural Network (ANN). Each speech signal will be represented by a vector. The feature vector will constitute the input to the ANN. The collection of speech signal will be divided into two sets. One set will be used for training the ANN in a supervised fashion. The other set which is never seen by the ANN will be used for testing. After training, the ANN will be tested for classification of the speech When the speech is classified correctly.*

Keywords: Speech Classification, Artificial Neural Network, Fast Fourier Transform

Received: 2 March 2014, Revised 7 April 2014, Accepted 10 April 2014

1. Introduction

The speech is defined as: Human communication through audible language. Speech sounds are made with air exhaled from the lungs, which passes between the vocal cords in the larynx and out through the vocal tract (pharynx and oral and nasal cavities). This airstream is shaped into different sounds by the articulators, mainly the tongue, palate, and lips. Articulatory phonetics describes each sound in terms of the position and action of the articulators used to make it. Speech is also described in terms of syntax, lexicon (inventory of words or morphemes), and phonology (sounds) [1]. Speech classification is defined as: embedded in voice-activated routing systems at customer call centers, voice dialing on mobile phones, and many other everyday applications. A robust speech-recognition system combines accuracy of identification with the ability to filter out noise and adapt to other acoustic conditions, such as the speaker's speech rate and accent. Designing a robust speech-recognition algorithm is a complex task requiring detailed knowledge of signal processing and statistical modeling [4].

The classifiers available are: Speech detection, Speech enhancement, Language Identification, Speaker identification that are need training material according to the classes that are to be distinguished. Each classifier will be described according to its properties and some typical results given [2].

This paper presents a speaker identification classifier which determine the speaker in speech signals. The result is the decision for a certain speaker or a rejection in case of an unknown speaker. A set of speech signals for each of the speakers to be identified is needed for training. This algorithm is language independent so the speaker will be identified irrespective of the language used [2]. The neural network is used by dividing the database of voices database into two groups: one for training the GRNN neural network and the other group for testing.

2. Proposed System

The previous researches used an algorithm of feature extraction with Mel frequency cepstrum coefficient (MFCC) [3]. However, this research proposed to use feature extraction with FFT algorithm.

The system algorithm is described in the following steps:

1) Collecting a database of recorded voices with format (.wav) that will be used for feature extraction algorithm, and the database consists of 200voices: 10 voices for 20 persons.

2) Dividing the database into two sets, first set consists of 140 voices for the training phase and the second set consists of 60 voices for the testing phase.

3) Implementing pre-processing steps on all voices to be ready to extract features, These steps are: convert voices from analog to digital forms and noise remove.

4) Implementing a FFT algorithm on all voices to obtain the features.

5) Training the Generalized Regression Neural Network (GRNN) on the training set of the voices.

6) Testing the GRNN using the testing set, and the resulting score of this phase represents the rate.

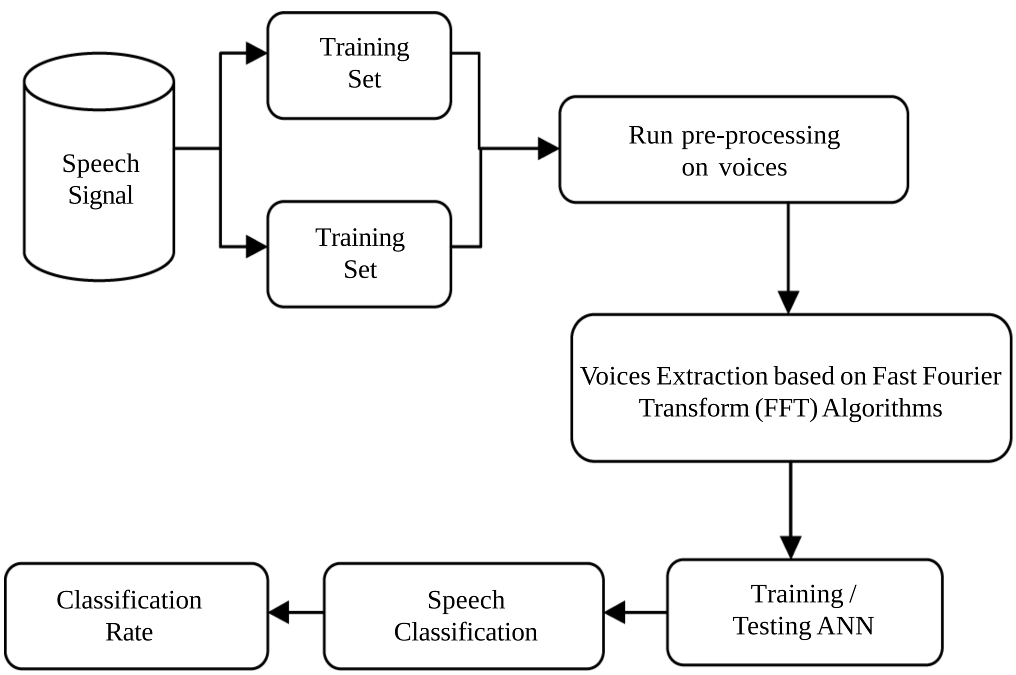


Figure 1. General Flow Chart of the Proposed Method

2. Implemented System

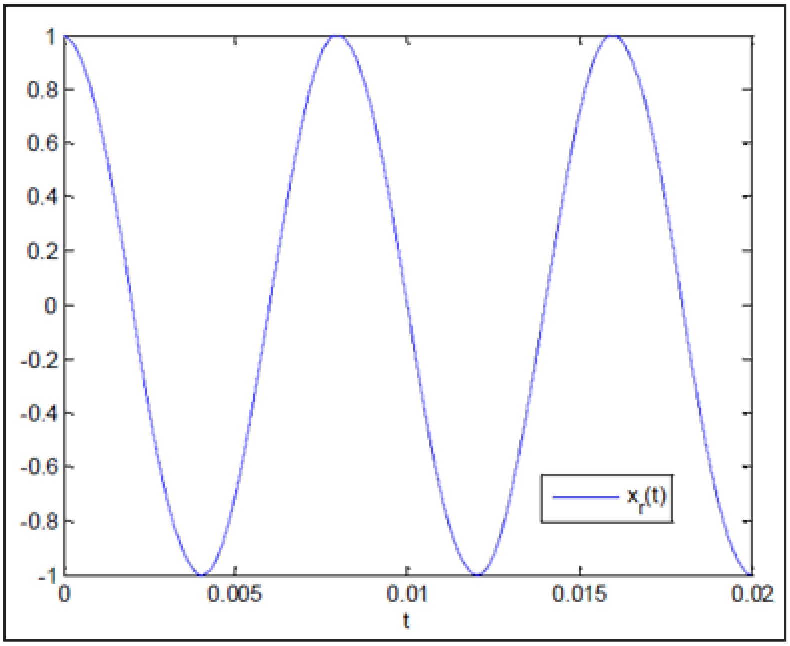
This section explains the implemented system that was used to perform the aim of this research.

2.1 Pre-Processing signal

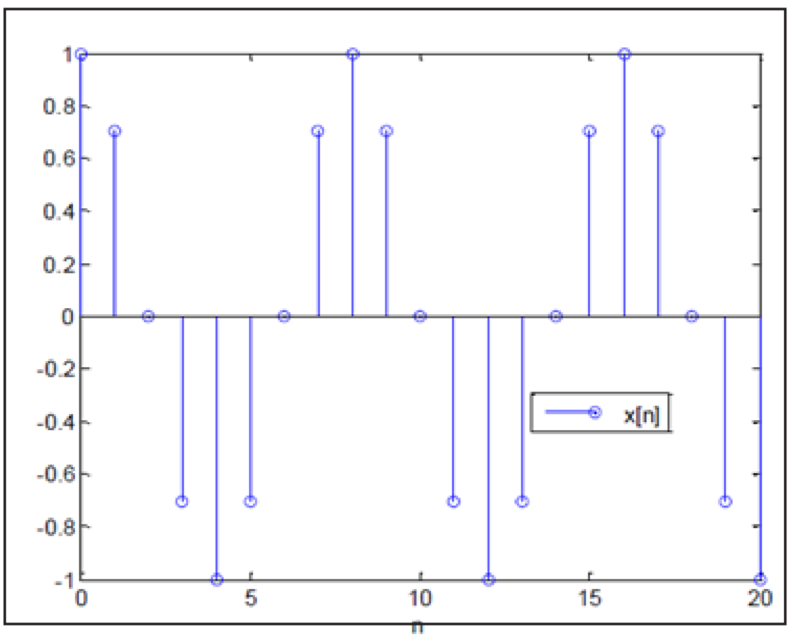
In this phase the proposed system executes two steps to format the voice signal before it extracts the features. These steps are, as shown in Figure 2:

1) Noise removal: To remove the noise from signal.

2) Analog to digital convert: The computers can not deal with analog signal, so we must convert the analog signal input to digital signal, we used the (wavread) function which is included in MATLAB tools, that takes the analog signal as input then convert it to digital signal.



Analog signal



Digital signal

Figure 2. Pre-processing Signal

2.2 Extract Features

After executing pre-processing, the signal is ready to extract features by Fast Fourier Transform (FFT) algorithm, The FFT is a faster version of the Discrete Fourier Transform (DFT). The FFT utilizes some clever algorithms to do the same thing as the DTF, but in much less time [5]. The, features are extracted as follows equations:

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-\frac{2x}{nk}/N}$$

As

$$X[k] = \sum_{n=0}^{N-1} x[n]\, W_N^{nk}$$

, $k = 0, 1, \ldots\ldots, N-1$ superscript (N) is to show length of DFT. For each value of k, computation of $X[k]$ requires:

N complex multiplications

$N-1$ complex additions

It is easy to realize that the same values of W_N^{nk} are calculated many times as the computation proceeds Using the symmetric property of the twiddle factor, we can save lots of computations.

$$X[k] = \sum_{n=0}^{N-1} x[n]\, W_N^{nk} = X[k] = \sum_{\substack{n=0 \\ \text{even } n}}^{N-1} x(n)\, W_N^{nk} + X[k] = \sum_{\substack{n=0 \\ \text{odd } n}}^{N-1} x(n)\, W_N^{nk}$$

$$= \sum_{r=0}^{N/2-1} x(2r)\, W_N^{2nk} + \sum_{r=0}^{N/2-1} x(2r+1)\, W_N^{k(2r+1)}$$

$$= \sum_{r=0}^{N/2-1} x_1(r)\, W_{N/2}^{rk} + W^{\frac{k}{N}} \sum_{r=0}^{N/2-1} x_2(r)\, W_{N/2}^{rk}$$

$$= X_1[k] + W_N^k X_2(k)$$

Thus, the N-point DFT can be obtained from two $N/2$-point transforms, one on even input data, and one on odd input data.

3.3 Artificial Neural Networks

The type of neural network that was used in this system is Generalized Regression Neural Network (GRNN). A Generalized Regression Neural Network (GRNN) is often used for function approximation. As shown in Figure 3, it has a radial basis layer and a special linear layer.

A GRNN is a variation of the radial basis neural networks. A GRNN does not require an iterative training procedure as back propagation networks. It approximates any arbitrary function between input and output vectors, drawing the function estimate directly from the training data. In addition, it is consistent that as the training set size becomes large, the estimation error approaches zero, with only mild restrictions on the function.

A GRNN consists of four layers: input layer, pattern layer, summation layer and output layer as shown in Figure 3.

1) The first layer is input layer that is connected to the pattern layer and in this layer each neuron presents a training pattern and its output.

2) The pattern layer is connected to the summation layer.

3) The summation layer has two different types of summation, which are a single division unit and summation units. Where the summation and output layer together perform a normalization of output set. In training of network, radial basis and linear activation functions are used in hidden and output layers.

Each pattern layer unit is connected to the two neurons in the summation layer, *S* and *D* summation neurons. *S* summation neuron computes the sum of weighted responses of the pattern layer. On the other hand, *D* summation neuron is used to calculate unweighted outputs of pattern neurons.

4) The output layer merely divides the output of each *S* summation neuron by that of each *D*-summation neuron, yielding the

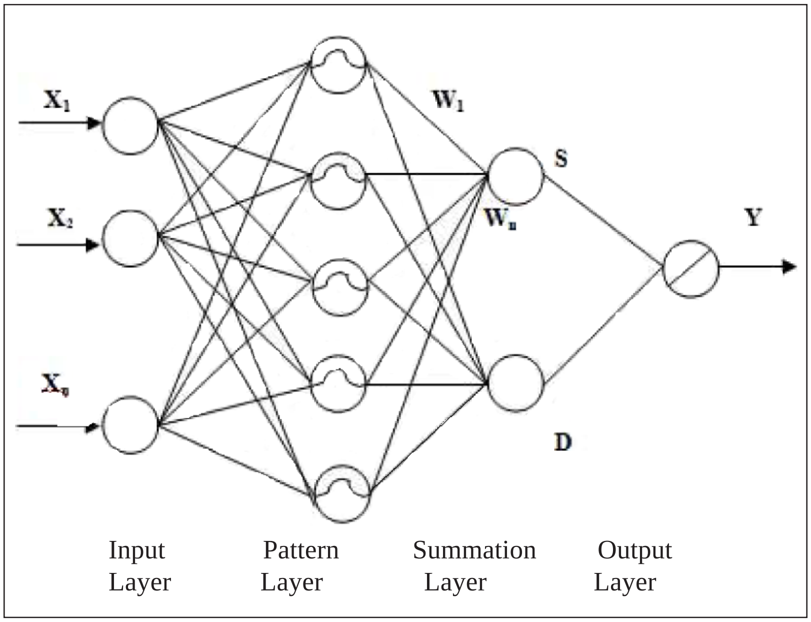


Figure 3. GRNN Architecture

predicted value *Y* to an unknown input vector *x*.

$$Y = \frac{\sum_{i=1}^{n} W_i \; exp - D_i}{\sum_{i=1}^{n} exp - D_i} \qquad (1)$$

where:

$$D_i = \sum_{k=1}^{m} \left(\frac{x_i - x_{ik}}{\sigma} \right)^2 \qquad (2)$$

$$\sigma = \frac{d_{max}}{\sqrt{m}} \qquad (3)$$

Where w_i is the weight connection between the i^{th} neuron in the pattern layer and the S-summation neuron, *n* is the number of the training patterns, D_i is the Euclidean distance with Gaussian Function. This function calculates the distance between an input (x_i) and the mean (x_{ik}) of those inputs, d_{max} is the maximum distance between the chosen centers, *m* is the number of elements of an input vector.

And this network has certain characteristics:

1) Fast learning.

2) Good convergence with a large number of training examples.

3) Handling of sparse data well [6].

4. Conclusion

This study has presented a hybrid method for the classification of speech. The technique of feature extraction using the (FFT) algorithm, calculation to extract features has been used, followed by Generalized Regression Neural Network (GRNN) for classification which has been implemented as well.

From this research the classification rate depends on the method that is used to extract the features of a speech and the technique that used as a classifier. Generalized Regression Neural Network (GRNN) is used as a classifier and extracted features with the (FFT) Features Algorithm gave 77.7% classification rate.

5. Acknowledgment

The authors would like to thank her supervisor, Ms. Laila, for her guidance and support all the way through the postgraduate research.

References

[1] Britannica, Encyclopedia. (2002). *Britannica concise encyclopedia*, Chicago: Encyclopaedia Britannica.

[2] Ulla Uebler. *A Speech Classification System*, MEDAV GmbH Gräfenberger Str. 32-34 D-91080 Uttenreuth Germany.

[3] Saurabh Padmawar, Deshpande, P. S. (2013). *Classification Of Speech Using Mfcc And Power Spectrum*, *International Journal of Engineering Research and Applications* (IJERA), 3 (1) 1451-1454, January -February.

[4] Ning, D. *Developing an Isolated Word Recognition System in MATLAB*.

[5] *FFT Tutorial*, University of Rhode Island Department of Electrical and Computer Engineering.

[5] Fannas, L. Y., Ahmed Y. Ben Sasi. (2014). *Off-Line Signature Recognition Based on Angle Features and Artificial Neural Networks Algorithm*, 12 (1) 85-89.