Arabic Word Recognition Using Wavelet Neural Network

Yousra F. Al-Irhaim   **Enaam Ghanem Saeed**

enaamghanim@uomosul.edu.iq

College of Computer Sciences and Mathematics
University of Mosul

ABSTRACT

In this paper, a system is presented for word recognition using Arabic word signals. The aim of the paper is to improve the recognition rate by finding out good feature parameters based on discrete wavelet transform. We have used Daubechies wavelet for the experiment. The back propagation neural network is used for classification. Test results showing the effectiveness of the proposed system are presented in this paper, A recognition accuracy of 77%.

Keywords:signal,classification,word

تمييز الكلمة العربية باستخدام التحويل المويجي المتقطع

يسرى فيصل الارحيم   **انعام غانم سعيد**

enaamghanim@uomosul.edu.iq

كلية علوم الحاسوب والرياضيات

جامعة الموصل

تاريخ استلام البحث: 2010/10/1   **تاريخ قبول البحث: 2010/11/10**

الملخص

في هذا البحث , تم وضع نظام لتمييز الكلمات العربية باستخدام إشارات الكلمة العربية. الهدف من البحث برهنة نسبة التمييز بالبحث عن القيم الجيدة للصفات بالاعتماد على التحويل المويجي المتقطع. لقد تم استخدام مويحة Daubechies في التجارب. تم استخدام شبكة الانتشار العكسي في التصنيف. نتائج التمييز أظهرت كفاءة النظام وتم عرضها في البحث، وكانت نسبة التمييز ما يقارب 77%.

الكلمات المفتاحية: اشارة, تصنيف,كلمة

1. Introduction:

Speech is the predominant mode of human communication for every day interaction. Speech will also be the preferred mode for human-machine interaction. [11] Speech recognition is becoming a very important concept for any type of system requiring human interaction in todays hi-tech pervasive services. Controlling a system with speech rather than using hardware e.g. keyboard or keypad definitely much more easy and appealing. Recognition should be accurate and quick. Consequently, the feature extraction method and the classifier have a direct influence in speech recognition systems. [15]

Speech recognition systems falls into two classes isolated word recognition and continuous speech recognition. [11]

Over the past many decades the researchers are trying to come out with new feature parameters which give good recognition result for computer speech recognition. Majority of the research activities are focusing on some of the conventional transform techniques like FFT, MFCC, LPC, and STFT etc. Speech signals from human are considered to be non stationary in nature. It is very difficult to analyze these non stationary signals by using these conventional transform techniques. [17] However, these methods have some disadvantages. These methods accept signal stationary within a given time frame and may therefore lack the ability to analyze localized events correctly. Moreover, the LPC method accepts a particular linear (all-pole) model of speech word production which strictly speaking is not the case. [3, 4,]

The Wavelet Transform overcomes some of these limitations; it can provide a constant-Q analysis of a given signal by projection onto a set of basis functions that are scale variant with frequency. Each waveletis a shifted scaled version of an original or mother wavelet. These families are usually orthogonal to one another, important since this yields computational efficiency and ease of numerical implementation. Other factors influencing the choice of Wavelet Transform over conventional methods include their ability to capture localized features. Also, developments aimed at generalization such as the Best-Basis Paradigm of Coifman and Wickerhauser make for more flexible and useful representation.[3,15,4,17]

Neural Network is well-known as a technique that has the ability to classified nonlinear problem. Today, lots of researches have been done in applying Neural Network towards the solution of speech recognition such as Arabic. The Arabic language offers a number of challenges for speech recognition. [1]

ANN is a fast emerging technology. Its ability to compute complex decision surfaces and its numerous processing elements have given it the ability to classify objects and make complex decisions. [7, 9]

Generally, there are three usual methods in speech recognition: Dynamic Time Warping (DTW) Model, Hidden Markov Model (HMM) and Artificial Neural Networks (ANNs). Nowadays, ANNs are utilized in wide ranges for their parallel distributed processing, distributed memories, error stability, and pattern learning distinguishing ability . [10]

The Wavelet Transform theory and its implementation are limited to the wavelets of the small dimension, where as the ANN are the powerful tools for handling the problem of higher dimension. Combination of both results in Wavelet network. The weakness of each other compensates and it can handle problems of larger dimension and it also shows efficient network construction methods.[17,14] In several studies, a wavelet neural network was used for speech recognition.[4]

This paper presents a method of the wavelet-neural network for Arabic word recognition. The discrete wavelet transform is used to extract features of analyzed speech signals. Then based on extracted features, a neural network is used for pattern recognition approximation. So Arabic word signals are used for obtaining the data sets. The word signals are transmitted to the computer by using a microphone and an audio card which has 11 KHz sampling frequency.
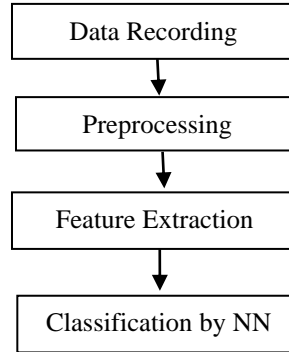

2. Architecture of System:

A speech recognition system consists of two main parts: training unit and testing unit. Training speech data is input in the training unit which generates a model. This model then used by testing unit. The testing speech data is fed to the testing unit which

performs pattern matching using the model obtained from the training unit. The speech data is pre-processed and set of features are extracted from the speech data. [16] The proposed system here is implemented using matlab 6.1. The architecture of our speech recognition system has been shown below in figure(1) .

Our speech recognition process contains three main stages:
1- Preprocessing.
2- Feature extraction from wavelet transforms coefficients.
3- Classification and recognition using back propagation learning algorithm.

```
┌─────────────────────┐
│   Data Recording    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Preprocessing    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│  Feature Extraction │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│ Classification by NN│
└─────────────────────┘
```

Fig(1): System Architecture

2.1. Preprocessing:

The analog speech signals are recorded using microphone, converted and stored into digital speech signal. The stored speech signal is in the form of wave files as shown in figure(2). The speech samples thus obtained are stored for further computation.

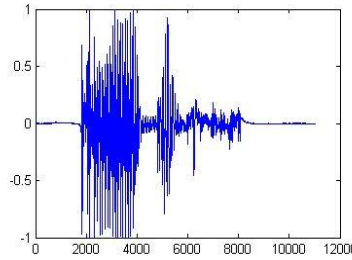Audio sampling rate 11 kHz
Audio sampling rate size 16 bit

Fig (2):The wav file of the word "قام" for the male speaker

2.1.1. Windowing:

The Speech signal is multiplied by an appropriate times window when dividing it into frames. Windowing process gradually attenuates the amplitude at both ends of the extraction interval to prevent an abrupt change at the endpoints . A Rectangular Window is used [5,2]

$$w(n) = \begin{cases} 1, & 0 \le n \le N \\ 0, & otherwise \end{cases} \qquad \dots(1)$$

2.1.2. Edge Detection:

A simple algorithm based on amplitude detection is used here. A word is considered as started if the amplitude crosses over a pre-defined threshold value. Until the amplitude of the speech signal remains over this threshold value, the signal is considered to be in the voiced region. When the signal amplitude stays below the threshold for a predefined time, the end of the signal is detected. This value is selected in such a way that it would not mistakenly cut off the speech signal in an intermediate point.[12] As shown in figure(3).
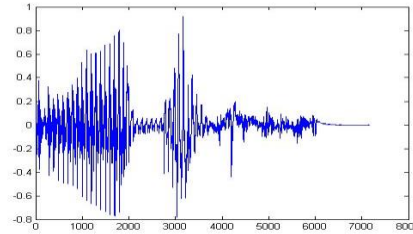

Fig (3): The Signal after preprocessing

3. Structure of The Wavelet Neural Network:

The Wavelet has generated a tremendous interest in both applied and theoretical areas. The wavelet transform theory provides an alternative tool for short time analysis of quasi stationary signal such as Speech as opposed to traditional transforms like FFT.[18]

Artificial neural networks (ANN's) are systems consisting of interconnected computational nodes working somewhat similarly to human neurons.[5]
The combination of wavelet theory and neural network has lead to the development of wavelet networks. [8]

3.1. Discrete Wavelet Transform:

Discrete Wavelet Transforms (DWTs) are orthogonal functions which can be implemented through digital filtering techniques and are basically originates from Gabor wavelets. Wavelets have energy concentrations in time and are useful for the analysis of transient signals such as speech signals. DWT is the most promising mathematical transformation which provides both the time –frequency information of the signal and is computed by successive low pass filtering and high pass filtering to construct a multi resolution time-frequency plane . In DWT a discrete signal x[k] is filtered by using a high pass filter and a low pass filter, which will separate the signals to high frequency and low frequency components. To reduce the number of samples in the resultant output we apply a down sampling factor of 2.[17,16,18,6] The Discrete Wavelet Transform is defined by the following equation:

$$W(j,k) = \sum_{j}\sum_{k} X(k)\, 2^{-\frac{j}{2}}\, \psi(2^{-j}n - k) \qquad \ldots(2)$$

Where Ψ (t) is the basic analyzing function called the mother wavelet .The digital filtering technique can be expressed by the following equations :

$$Y_{high}[k] = \sum_{n} X[n]\, g[2k-1] \qquad \ldots(3)$$

$$Y_{low}[k] = \sum_{n} X[n]\, h[2k-1] \qquad \ldots(4)$$

Where Y high and Y low are the outputs of the high pass and low pass filters.

## 4. Feature Extraction and Classification Using DWNN:

Figure (4) shows the Discrete Wavelet Neural Network (DWNN) structure for classification of speech signal waveform patterns from the Speech file set. Feature extraction is the key for the system, so that it is arguably the most important component of designing an intelligent system based on speech recognition, since the best classifier will perform poorly if the features are not chosen well. A feature extractor should reduce the pattern vector (i.e. the original waveform) to a lower dimension, which contains most of the useful information from the original vector.[4] In all the cases we have taken the approximation coefficients value.
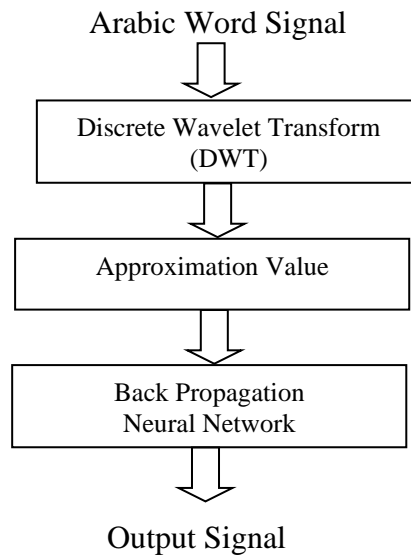
Fig(4) The structure of DWNN for Arabic word

4.1. Feature Extraction Using Discrete Wavelet:

Discrete Wavelet transform is used for feature extraction from word signals. For DWT of the word signals, the tree structure is used with m=7 as the level [3]. In this research, 5 individual speakers are used to obtaining the word signals. Two of these speakers are male, and three are female. Each of these speakers is asked to utter all used Arabic words twice. For DWT of the word signals the decomposition structure and reconstruction tree at level 7, as shown in figure(5) is used. DWT is applied to the word signal using Daubechies-10 (db10) wavelet decomposition filters. Thus we obtain two types of coefficients: one approximation coefficients cA and seven detail coefficients cD. A representative example of the speech signal of a male speaker for the Arabic word "قام " and the DWT of the speech signals of a male speaker are shown in Figure (6).
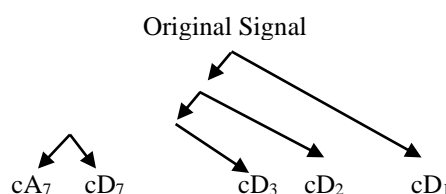
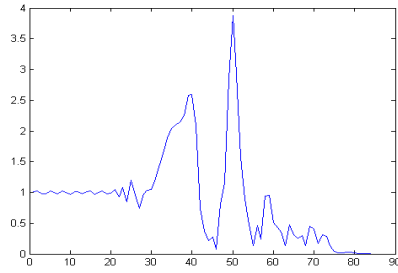**Fig(5) The decomposition structure at level-7**

Fig (6): The Approximation coefficients for the word "قام"

4.2. Classification:

In a general sense, a neural network is a system that emulates the optimal processor for a particular task, something which cannot be done using a conventional digital computer, except with a lot of user input. Optimal processors are sometimes highly complex, nonlinear and parallel information processing systems. [17]
Back propagation Neural Network are one of the most common neural network structures, as they are simple and effective, and have been used widely in assortment of machine learning applications. [13]

The Back propagation realizes the classification using features obtained from the discrete wavelet transform. Figure(7) shows the Feedforword Backpropagation Network with 3 layers for input, hidden and output. In training stage, we have used 7 neuron in the input layer and 7 neuron in the output layer for each speaker. The training parameters and the structure of the network used in this research are as shown in table 1. These were selected for the best performance, after several different experiments, such as the number of hidden layers, the size of the hidden layers and type of activation function.
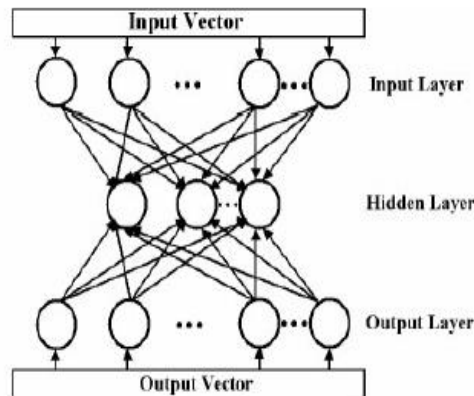
Fig(7): An Example Feed forward Back-propagation Network

Table 1. Architecture and Training Parameters

Architecture	
Number of layers	3
Number of neurons on the layers	Input: 7 Hidden: 1 Output: 7
Activation function	Tansig

5. Experiment and Result:

In this experiment , the source of data is a database consists of 7 Arabic words ("شب","كتب","مكتب","سد","عد","قام","قوم"), spoken 2 times by 5 speakers; those are 3 male and 2 females of various ages, so we have 70 files. The data which is speaker dependent , will be used for training and testing form. In speaker dependent form, the first utterance of each of the 7 words spoken by every speaker are used to train the network and the remaining utterance are used to test the network. Therefore, the speech data base contains 35 utterances, which can be used for training the network, and 35 utterances, which are available for testing. Table(2) contains the performance for the test phase for each speaker.

Performance = Total succeeded number of testing words *100
Total number of words

Table 2. The Performance For The Test For each Speaker

Speaker	Performance
1(male)	75%
2(male)	80%
3(female)	70%
4(female)	76%
5(female)	83%

6. Conclusion:

From this study we could understand and experience the effectiveness of discrete wavelet transform in feature extraction. We have also observed that Neural Network is an effective tool which can be embedded successfully with wavelet. However, the result is encouraging one. Even though the Discrete Wavelet Based Transform technique with ANN classifier gives a very good recognition result, the efficiency of the method is to be verified with very large database (number of words, number of speakers).

REFERENCE

[1]. Abdul Manan Ahmad, Saliza Ismail, Den Fairol Samaon,[2004],"Recurrent Neural Network with Back propagation through Time for Speech Recognition", International Symposium on Communications and Information Technologies 2004 (ISCIT 2004) Sapporo, Japan October 26-29.

[2]. Akram M. Othman, and May H. Riadh,[2008],"Speech Recognition Using Scaly Neural Networks", World Academy of Science,Engineering and Technology 38.

[3]. Engin Avci,[2007],"An Automatic System for Turkish Word Recognition Using Discrete Wavelet Neural Network Based On Adaptive Entropy", The Arabian Journal for Science and Engineering, Vol. 32, No. 2B, October .

[4]. E. M. Essa, A. S. Tolba and S. Elmougy,[2008]," Combined Classifier Based Arabic Speech Recognition", INFOS2008, March 27-29, Faculty of Computers & Information-Cairo University.

[5]. Firoz Shah. A, Raji Sukumar. A and Babu Anto. P,[2010]," Discrete Wavelet Transforms and Artificial Neural Networks for Speech Emotion Recognition", International Journal of Computer Theory and Engineering, Vol. 2, No. 3, June, 2010, 1793-8201.

[6]. Geok See Ng, S. S. Erdogan, Pan Wei Ng ,"Neural Network For Voice Recognition",.

[7]. Marc Thuillard,[2000],"A Review of Wavelet Network, Wavenets, Fuzzy Waventes and Their Application", ESIT 2000, 14-15 September, Achen, Germany.

[8]. Mutasem Khalil Sari Alsmadi, Khairuddin Bin Omer and Shahrul Azman Noah,[2009],"Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm", IJCSNS International Journal of Computer Science and Network Security, Vol.9, No. 4, April.

[9]. Meysam Mohamad pour, Fardad Farokhi,,[2009],"An Advanced Method for Speech Recognition", World Academy of Science, Engineering and Technology 49 .

[10]. Raji Sukumar A., Firoz Shah. A, Babu Anto.P,[2010],"Query Type Identification Using Isolated Question Word Recognition", International Journal of Managing Information Technology(IJMIT), Vol.2, No.1.

[11]. Perera K.A.D., Ranathunga R.A.D.S., Welivyitigoda I.P., and Withanawasaw R.M.,[2005],"Isolated Word Recognition", International Conference on Information and Automation, December15-18,2005,Colombo,SriLanka.

[12]. Rubita Chee Mimg ,"NN Speech Recognition Utilizing Aligned DTW Local Distance Scores".

[13]. S. Mohanty, S. Bhattacharya ,"Recognition of Voice signals for Oriva Language using wavelet NueralNetwork",.[http://www.cse.iitk.ac.iu/ users/langtech/ strans 200.

[14]. Serkan Gunal and Rifat Edizkau,[2007],"Use of Novel Feature Extraction Technique with Subspace Classifiers for Speech Recognition", IEEE Inernational Conference on Perasive Services, pp80-83.

[14]. Sukminder Singh Grewal, Dinesh,[2010],"Isolated Word Recognition System for English Language",International Journal of Information Technology and Knowledge Management, July-December, Vol. 2, No. 2, pp. 447-450.

[15]. Vimal Krishnan V.R, Babu Anto P,[2009],"Feature Parameter Extraction from Wavelet Sub-band Analysis for the Recognition of Isolated Malayam Spoken Word", (IJCNS) International Journal of Computer and Network Security, Vol. 1, No. 1, October.

[16]. Vimal Krishnan V.R, Babu Anto P,[2009],"Features of Wavelet Packet Decomposition and Discrete Wavelet Transform for Malayalam Speech Recognition", International Journal of Recent Trends in Engineering,Vol. 1, No. 2, May .