A WAVELET NEURAL NETW0RK RAMWORK FOR SPEAKER IDNTIFCATION

Prof. Dr. W. A. Mahmoud **Dhiadeen.M.Salih** **Prof. Saleem M-R.Taha**
College of Engineering - University of Baghdad
Baghdad - Iraq

ABSTRACT

This paper introduces a new model-free identification methodology to detect and identify speakers and recognize them. The basic module of the methodology is a novel multi-dimensional wavelet neural network. The WNN approach include: a universal approximator; the time - frequency localization; property of wavelets leads to reduced networks at a given level of performance; The construct used as the feature mode classifier. Wavelet transform has been successfully applied to the processing of non - stationary speech signal and the feature vector that obtained becomes the input to the wavelet neural network which is trained off-line to map features to used for the classification procedure. An example is employed to illustrate the robustness and effectiveness of the proposed scheme.

الخلاصة

في هذا البحث تم اقتراح طريقة لنظام تمييز تعتمد على شبكة العصبية للتحويل المتموج ذات متعددة الأبعاد (wavelet neural network) حيث إن نظرية(WN) يتضمن التحديد الزمني والترددي والتابع للتحويل التموجي مساعدا بتقليل نسبة تعقيد الشبكة و على هذا الأساس استخدم هذه الشبكة كمصنف لخصائص لنماذج معينة من صوت كل متكلم حيث يستخلص بطريقة التحويل المتموج المتقطع (Discreet wavelet transform) لعدة مستويات بعد تقسيم كل صوت إلى عدد من مقاطع متساوية ومن ثم اخذ الطاقة المعدلة لكل مستوى حيث يتحصل بذلك على متجه ذات معاملات تدل لخصائص الكلمة للمتكلم وبعده يطبق جميع المتجهات المستحصلة لكل متكلم على شبكة التحويل المتموج (WN) وذلك لغرض تعليم الشبكة (Learning face) ومن ثم تطبيق صوت متكلم مجهول على الشبكة للتعرف عليه وقد أعطت هذه الطريقة عدد أوطئ من الحسابات وبذلك يزيد من كفاءة النظام ويقلل من وقت التنفيذ مقارنة لبقية الشبكات العصبية المستخدمة مسبقا . هذه الطريقة تم تطبيقه على حاسبة سرعة معالجها (850 MHz Celeron) و(RAM 128 MB) ولغة برنامج هي MATLAB 6 . أما قاعدة البيانات فهي مكونة من خمسة و عشرين شخص (12 ذكور و 13 إناث) وقد كانت نسبة التمييز هي 82% مع زمن تعلم للشبكة لا يتجاوز 47 ثانية في حالة النص المستقل ونسبة 100% مع زمن تعلم للشبكة تصل إلى 55 ثانية في حالة النص المعتمد .

KEYWORDS
Speaker identification, speaker recognition, wavelet neural network wavelet transform, discrete wavelet transform, neural network, back – propagation algorithm.

INTRODUCTION
Recently, some strategic issues and approaches for speaker identification (SI) have been addressed by several investigators. The issue is the performance of SI so that recognition delays and false identify may be avoided .The complexity of the SI task lies in the fact that given utterance can be represented by an effectively infinite number of time - frequency pattern . typical classification problem, which generally include tow main modules: feature selection and classification where the second part i.e., classifier design have their own disadvantages due to the complex distribution of the feature vectors [Hex 2001]. Wavelet neural network (WNN) have recently attracted great interest ,because their advantages over radial base function network as thy are universal approximators but achieve faster convergence, Furthermore, WNNs possess a unique attribute: In addition to forming an orthogonal basis are also capable of explicitly representing the behavior of a function at various resolutions of input variables[George 2000].For instance, the task of pattern recognition is function mapping whose objective is to assign each pattern in a feature space to a specific label in a class space.

This paper is organized as follows the next section introduces some basic concepts in wavelets and wavelet neural networks; we describe next the general identification and classification architectures; focused attention is paid to the wavelet neural network; our example is used to illustrate the main features of the scheme; the paper concludes preprocess the speech signals (16 bit sampled in 8khz).then extract features vectors with discrete wavelet transform (DWT) to be trained off-line by WNN with different selection errors to get data base of speakers, then applied unknown speaker vector to the WNN to be classified and identify the speaker.

DISCRETE WAVELET TRANSFORM FOR REATURE EXTRACTING
The Discrete Wavelet Transform (DWT) is more popular in the field of signal digital processing. We thus introduce a simple feature extraction model based on the result of DWT. In order to parameterize the speech signal, we should first decompose the signal in the dyadic form using the Mallat's algorithm [Mallat 1989].

The ability ofDWT to extract features from the signal is dependent on the appropriate choice of the mother wavelet function [Burrus 1998]. Some of the popular families of wavelet bases functions are Harr, Daubechies. Coiflet, Symlet, Morlet, and Mexican Hat. The properties of the wavelet functions and the characteristics of the signal being analyzed need to be matched [Khalaf 2003] . The properties of wavelet function are tabulated in **Table (1).**

Table (1) Properties of Wavelet Functions

[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

The objective of this module is to determine and extract appropriatpJeatures for the fault or defect classification task. An additional objective is to reduce the search space and to speed up the computation.. In preparation for feature extraction, a windowing operation is applied to the I-D signals in order to reduce the search space and facilitate the selection of appropriate features [Khalaf 2003].

WAVELET NEURAL NETWORKS

A neural network is composed of multiple layers of interconnected nodes with an activation function in each node and weights on the edges or arcs connecting the nodes of the network. The output of each node is a nonlinear function of all its inputs and the network represents an expansion of the unknown nonlinear relationship between inputs, x, and outputs, F (or y), into a space spanned by the functions represented by the activation functionsofthe network's nodes. Learning is viewed as synthesizing an approximation of a multidimensional function, over a space spanned by the activation functions $\Sigma\Phi(x)$,i = 1,2,...,m, i.e.

[illegible]

where .Np is the number of wavelet nodes in the hidden layer and WI is the synaptic weight of WNN. The additional parameter c_i is introduce to help dealing with nonezero average .since wavelet If/"is zero mean. A. WNN can be regarded as function aproximator ,which estimate an unknown function mapping [Q. Zhang 1992].This network structure is shown in **Fig .(1).**

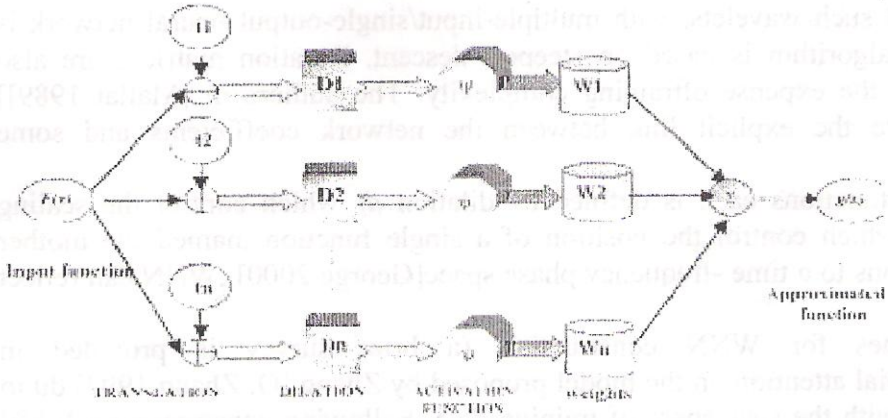


Fig (1) WNN Structure for approximation The combination of translation ,dilation ,and wavelet lying on the same line will be called a *wave/on* in the sequel.

The authors in [Oubez 1994], and independently, in [Bakshi 1992], arrive at very similar formulations of the wavelet network that are closer to the wavelet expansion than to neural networks. The wavelet
parameters are neither adapted as in [Q. Zhang 1992] .nor computed from prior Fourier data analysis as in [Pati 1993], but are taken incrementally from a predefined space-frequency grid of orthogonal wavelets.
This approach prescribes learning as a multiresolution, hierarchical procedure, and brings about the possibility of a type of network growth.

Wnn Initialization

The initialization of WNN consists in the evaluation of the param(;;eers f.J, *Wi* ,*tli* and *di* =1/sj for i =1,2 ...,*N*. To initialize f.J we need to estimate the mean of the function f(x) (from its available observation) and set f.J to the estimated mean. *WTS* are simply set to zero .the rest of problem is how to initialize *tli* 's and *s;* 's . The approximation error is minimized by adjusting the activation function and network parameters using empirical (experimental) data. Two types of activation functions are commonly used: global and local. Global activation functions are active over a large range of input values and provide a global approximation to the empirical data. Local activation functions are active only in the immediate vicinity of the given input value[Bakshi 1994].

It is well known that functions can be represented as a weighted sum of orthogonal basis functions. Such expansions can be easily represented as neural nets by having the selected basis functions as activation functions in each node, and the coefficients of the expansion as the weights on each output edge. Several classical orthogonal functions, such as sinusoids, Walsh functions. etc., but, unfortunately, most of them are global approximators and suffer, therefore, from the disadvantages of approximation using global functions. What is :needed is a set of basis functions which are local and orthogonal. A special class of functions, known as wavelets, possess good localization properties while they are simple orthonormal bases. Thus, they may be employed as the activation functions of a neural network known as the Wavelet Neural Network (WNN). WNNs possess a unique attribute: In addition to forming an orthogonal basis are also capable of explicitly representing the behavior of a function at various resolutions of input variables. The pivotal concept, in the formulation and design of neural networks with wavelets as basis functions, is the multiresolution representation of functions using wavelets. It provides the essential framework for the completely localized and hierarchical training afforded by Wavelet Neural Networks [George 2000]].

By linearly combining several such wavelets, with multiple-input/single-output neural network is obtained. The basic training algorithm is based on steepest descent. Rotation matrices are also incorporated for versatility at the expense oftraining complexity. The authors in [Mallat 1989]] demonstrate the way to have the explicit link between the network coefficients and some appropriate wavelet transform.

Wavelets occur in family of functions each is defined by dilation *dj* which control the scaling Parameter and translation *tj* which control the position of a single function ,named the mother wavelet *lj/(x)*. Mapping functions to a time -frequency phase space[George 2000] .WNN can reflect properties more accurately.

There are several approaches for WNN construction (a brief survey is provided in [Q. Zhang 1992]),we pay special attention on the model proposed by Zhang [Q. Zhang 1992] du to its notable feature in dealing with the sparseness of training data. Following constructing a WNN involves tow stages: First, construct a wavelet library *W* of discretely dilated and translated version of wavelet mother function lj/:

[illegible]

where Xk is the sampled input, and L is the number of wavelets in W, seconed select the best M wavelet based on the training data form wavelet library W, in ordex to build the regression Based on the previous discession we propose a network structure ,Given an n-element training set of the form:

[illegible]

To initialize tJ and Sj select a point p between interval of function a and b " $a < p < b$ The choice of this point will be detailed later. Then we set

Where ζ >0 is properly selected constant (the typical value of ζ is 0.5). the interval divided into two parts by p . in each sub interval we repeat the same procedure which will initialize 12, S2and IJ . s] and so on , until all wavelet will initialize. \ This procedure applies in this form when a number of wavelons are used which is a power of 2. then we take the point p to be the center of gravity of (a, b). There are several mother wavelets that could be useful in ore project .The continuos wavelet transform theory in the Morlet- Grossmann sense provides us with considerable flexibility in designing our networks **If** = *(x exp* (-1/2 *X2)* [Q. Zhang 1992] . shown in fig 2. There is Mexican - Hat. The mother wavelet If= (1- *X2) exp* (-1/2 *X2*), shown in **Fig .(3)**.

This function, shown in **Fig. (4)**, consists of two cycles of the cosine function, windowed by a trapezoid that linearly tapers two thirds of the endpoints to zero[George 2000], these function will be used in training WNN.

Wavelet Neural Network Classifier

The MIMI WNN, depicted schematically in **Fig. (5)**, is used as the classifier. Potential dvantages of the WNN as a universal approximator and the time - frequency localization roperty of wavelets leads to reduced networks at a given level of performance ,so WNNs offer a ood compromise between robust implementations and efficient functional representations; the ulti-resolution organization of wavelets provides a heuristic *for* neural network growth.

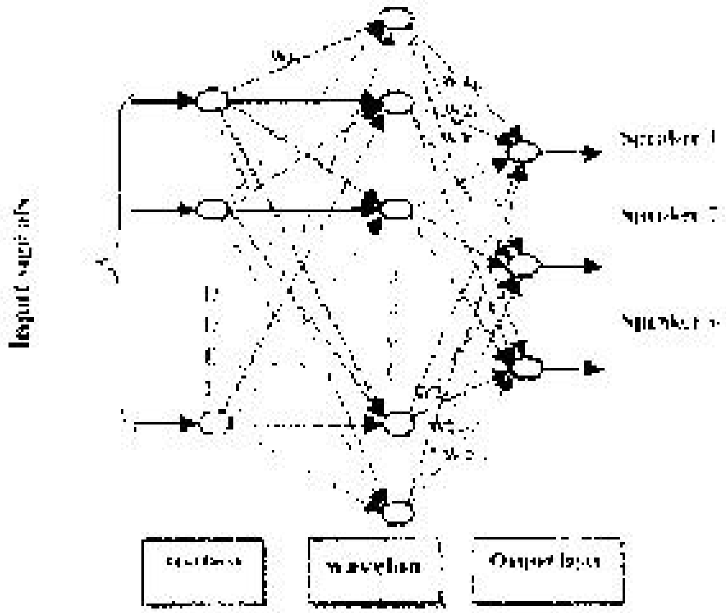
Furthermore, WNNs may be optimized with respect to structure (nm-'berof nodes) and their arameters using a Genetic Algorithm as the optimization tool. The structure and the parameters f the network are determined iteratively until a performance metric is satisfied. The WNN onstruct suggests a means to parallel-process multiple signals in a multi-tasking environment, hus expediting considerably processing times. Finally, it offers an easy and user-friendly way to learn" new signal patterns, as long as training data is available.

This algorithm modifies the parameters vector *e* after each measurement *(Xk,Yk)* in the opposite irection of the gradient of the functional

[illegible] (4)

As is the case for backprobagation algorithm for neural network learning [Q. Zhang 1992] . The objective unction (4) is likely to be highly nonconvex ,so local minima are expected. To improve the ituation ,careful initialization of the algorithm is preformed and appropriate constraints ar~ set n the adjusted parameters.

- [illegible]
  [illegible]
- [illegible] (6)
  [illegible]
  [illegible]
  [illegible]
- [illegible] (7)
- [illegible] (8)
- [illegible] (9)
- [illegible] (10)
- [illegible]



[illegible]

Wavelet Neural Network Learning Algorithm:

The method of setting the values of weights (in training phase) is an important istinguishes characteristics of different neural networks. There are two common types of raining algorithms, supervised and unsupervised, sometimes there is a third method, i.e. elf-supervised or reinforcement training method [[Zhanshou 200].

The learning is based a Stochastic gradient type algorithm **Fig. (6)** which very imilar to the backprobagation algorithm for neural network ,first collect all the parameters *go,Wi,tidi* In a vector *e* and write *fa (-y)* to refer to the network defined by Eg.(3) whith the parameter vector *e* . The objective function to be minimized is

[illegible] 3

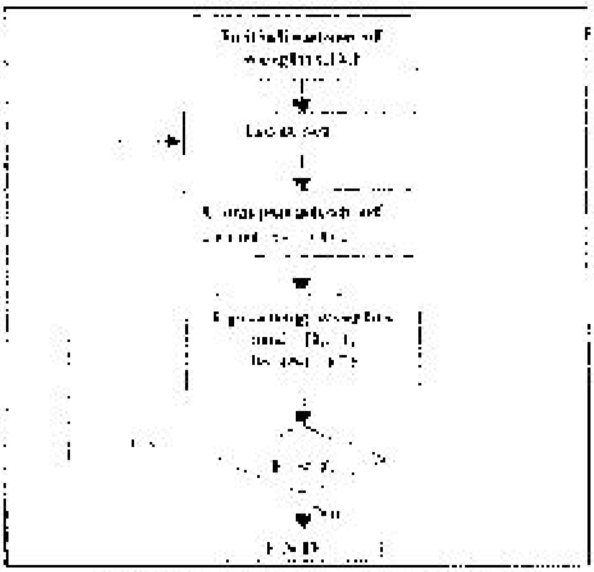


Fig (6) [illegible]

SPEAKER IDENTIFICATION AND CLASSIFICATION METHODOLOGY

Fig. (7) depicts the major modules of the identification and classification methodology. Sensor data are used first off-line to generate a feature base. From a feature library, those features are selected that provide a good match with the failure modes to be detected and identified. An incoming sensor signal is fed on-line in real-time to the feature extractor which attempts to extract a set of features. This feature vector is provided next as an input to the wavelet neural network; the latter is accompanied with an appropriate decision logic that decided upon the particular speaker class that the features (symptoms) belong to.

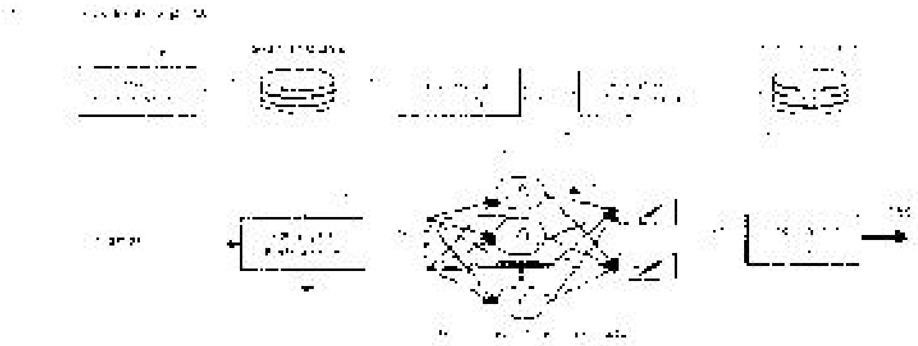


Fig. 7

EVALUATION TEST OF THE PROPOSED SPEAKER IDENTIFICATION SYSTEM

In this section the experimental result will be given.

Database

[illegible]

The Preprocessing

- The speech signals used in this work are sampled with a sampling frequency of44.1 kHz.
- The speech signal is segmented into 256 samples per segment (frames), the overlapbetween frames is 128 samples per segment.

Feature Extraction

- Each frame of the spoken words is now expanded using the Discrete Wavelet Transform (DWT) up to 8 levels of decomposition.
- By computing the power in each segment in each level of the decomposition, a feature vectors 'Yill be obtained that describes the power distribution over the timefrequencyplane. IThis scale power density along every segment describes the power variation in each scale.
- The variance of the power overall the segments and for each of the 8 levels is computed, leading in a vector called (normalized power vector).

These steps will be shown in the **Fig (7).**

The proposed method is first studied with changing the wavelet functions and using different numbers of wavelones, because these two parameter influence directly on the speaker recognition accuracy. The Daubechies wavelet of order 4 (Db4) is chosen in the processing phases and. The Daubechies wavelets have some characteristics that are useful for speaker recognition

[Khalaf 2003]. **Table (2) and (3)** shows the percentage of correct classification for text - independent and text - dependent, respectively.

[illegible]

[illegible]	Accept	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

[illegible]

[illegible]	Accept	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

CONCLUSIONS

A model-free approach to the problem of speaker identification conditions hasbt-enpresented. The multi-dimensional WNN is an effective and efficient tool for classification. The computational to the feature extraction step where appropriate features must be computed from signal data that comprise eventually the input vector to the network. The WNN approach offers additional advantages in terms of learning and optimization functions that may be carried out offline or on-line. FUlibermore, the neural net topology suggests means for parallel processing - useful in high frequency processes because of fast learning time. These shows promise as an effective model for the analysis of process data for many industrial and other engineered systems.

REFERENCES

He Xuming, Hu Guangrui and Tan Zhenghua (2001), Coevolutionary Approach To Speaker Identification Using Neural Networks" Proc. ICASSP 2001.

George Vachtsevanos, Peng Wang (2000), A WNN frame work for diagnostic complex engineered stems. Atlanta, GA 30332, USA.

Mallat, (1989). A Theory of Multi - Resolution Signal Decomposition, proceeding of IEEE.

Burrus C.S., Gopinath, R.A., and Gou, H., (1998), Introduction to Wavelet and Wavelet Transform.

prof. Dr. W. A. Mahmoud, R. F. Khalaf, Safiea K. Omran (2003). Speaker Identification Based On Wavelet And Neural Network, College Of Engineering - University of Baghdad ..

Q. Zhang and A. Benveniste. (1992), Wavelet networks, IEEE Transactions on Neural Networks, vol. 3, no. 6, pp. 889-898, Nov..

Pati and P.S. Krishnaprasad. (1993), Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations, IEEE Transactions on Neural Networks, vol. 4, no. 1, pp. 73-85, Jan..

Oubez and R.L. Peskin, (1994), Multiresolution neural networks. in SPIE.,vol. 242. pp. 649-659.

Bakshi and G. Stephanopoulos, (1992). Wavelets as basis functions for localized learning in a multi-resolution hierarchy, in Proc. IJCNN MD, , vol. 2, pp. 140-145.

Bakshi A.K. Koulouris, and G. Stephanopoulos, (1994), Wave-Nets: Novel learning techniques, and the induction of physically interpretable models." in SPIE., vol. 2242, pp. 637-648.

Zhanshou Y., (2000), Feed Forward Neural Networks And Their Applications In Forecasting, Msc. Thesis. Department of Computer Science, University of Housten, USA December