Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 14 (19): 2341-2349, 2014
ISSN 1812-5654 / DOI: 10.3923/jas.2014.2341.2349

Wavelet Linear Prediction Coding with Feed Forward Backpropagation Neural Network for Noisy Speaker Identification System

Khaled Daqrouq, Abdulhameed Alkhateeb, Mohammad Ajour and Ali Morfeq
Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract: Recently, speaker identification systems become very attractive for researchers. As a matter of fact, the results of new identification systems have been very crucial in the academic community. In the presented study, An average Framing Linear Prediction Coding (AFLPC) technique for noisy speaker identification systems is proposed. The work of the modified LPC with Wavelet Transform (WT), termed AFLPC, is investigated for speaker identification based on previous study by the author. The study procedure is based on feature extraction and voice classification. In the classification phase, Feed Forward Back-Propagation Neural network (FFBPN) is applied because of its rapid response and ease in implementation. In the practical investigation, performances of different wavelet transforms in conjunction with AFLPC were compared with one another. In addition, white Gaussian (AWGN), restaurant, babble and train station noises with 5 and 0 dB were examined for proposed system and other systems presented in the literature. Consequently, the FFBPN classifier achieves a better recognition rate (96.87%) with the Wavelet Packet (WP) and AFLPC termed WPLPCF feature extraction method.

Key words: Speech, LPC, average framing, wavelet, neural network, noise

INTRODUCTION

Speech processing applications include speech recognition and speaker identification. Speaker identification procedure is a technology with possibly big market because of its broad applications that varies from automation of operator-assisted service to speech-to-text aiding system (Antonini *et al*., 1992; Quiroga, 1998).

A commonly used technique for feature extraction is based on the Karhunen-Loeve Transform (KLT). These methods have been used for text-independent speaker recognition tasks (Lung and Chen, 1998) with excellent outcomes. Karhunen-Loeve transform is the optimal transform regarding to Minimum Mean Square Error (MMSE) as well as maximal energy packing. Most of the proposed speaker identification systems utilize Mel Frequency Cepstral Coefficient (MFCC) (Hosseinzadeh and Krishnan, 2007) and Linear Predictive Cepstral Coefficient (LPCC) (Afify and Siohan, 2007) as feature vector. Although MFCC and LPCC have proved to be two exceptional features in speech recognition, the weakness of the MFCC is that it produces short time Fourier transform which has a very weak time-frequency resolution as well as its inherent pre-assumption that the signal is stationary. Consequently, it is fairly problematic to recognize the phonemes by these features. Presently, some researches (Afify and Siohan, 2007; Long and Datta, 1996; Lung, 2010) are concentrating on the wavelet transform for speaker feature extraction stage.

Wavelet transform (Lung and Chen, 1998; Mallat, 1998; Vitterli and Kovacevic, 1995) has been widely presented in the last two decades and has been widely exploited in various fields of science and engineering. The wavelet analysis process is applied with expanded and translated versions of a mother wavelet function. Meanwhile, signals of interest can usually be expressed using wavelet decompositions, signal processing methods may be implemented by fine-tuning only the matching wavelet coefficients. From a mathematical point of view, the scale parameter of a wavelet can be a positive real value and the translation can be an arbitrary real number (Antonini *et al*., 1992). From a practical point of view, in order to improve computation efficacy, the values of the shift and scale parameters are frequently restricted to some discrete lattices (Mallat and Zhong, 1992; Mallat, 1991).

Discrete wavelet transform and WP analysis have been proven as useful signal processing techniques for a several digital signal processing problems. Wavelets have been used in two different methods in feature extraction procedures designed for the purpose of speech/voice recognition. Discrete wavelet transform is applied instead

Corresponding Author: Khaled Daqrouq, Department of Electrical and Computer Engineering, King Abdulaziz University, P.O. Box 80204, 21589, Jeddah, Saudi Arabia
Tel: +966 5 66980400 Fax: +96656952686

of Discrete Cosine Transform (DCT) that is employed for the feature extraction stage in the first method (Tufekci and Gowdy, 2000). Wavelet transform is applied directly on the speech/voice signals where, wavelet coefficients containing high energy are obtained as the features (Long and Datta, 1996).

Sub band energies that are exploited in place of the Mel filter-bank sub band energies suggested in (Davis and Mermelstein, 1980). Mostly, WP bases are utilized in (Mitchell, 1997) as close rough calculation of the Mel-frequency division by means of Daubechies orthogonal filters. In (Lung, 2004), a feature extraction method based on the wavelet Eigen function was suggested. Wavelets can offer a significant computational benefit by reducing the dimensionality of the Eigen value problem. A text-independent speaker identification scheme dependent on improved wavelet transform is offered in (Lung, 2010) where learning the correlation between the wavelet transform and the expression vector is obtained by kernel canonical correlation analysis.

The Wavelet Packets Transform (WPT) performs the recursive disintegration of the speech signal gained by the recursive binary tree. Fundamentally, the WPT is very comparable to Discrete Wavelet Transform (DWT). Nevertheless, WPT decomposes both details and approximations by only using the decomposition process on approximations. WPT features have superior representation than those of the DWT (Lung, 2004). Moreover, as the number of wavelet packet bases grows, the time required to appropriately classify the database will become non-linear. Consequently, the decreasing of the dimensionality becomes a significant issue. Selecting a beneficial and relevant subset of features from a larger set is crucial to enhance the performance of speaker recognition (Chen *et al*., 2002; Nathan and Silverman, 1994). A feature selection scheme is, therefore, needed to choose the most valuable information from the complete feature space to form a feature vector in a lower-dimensionality and take away any redundant information that may have disadvantageous effects on the classification quality. To select an appropriate set of features, a criterion function can be used to provide the discriminatory power of the individual features.

The wavelet packet decomposition tree was first suggested by Sarikaya *et al*. (1998) and produces the Wavelet Packet Parameters (WPP). Wu and Lin (2009b) proposed the energy indexes of DWT or WPT for speaker identification where, WPT was superior in terms of recognition rate. Sure entropy was considered for the waveforms at the terminal node signals acquired from DWT (Avci, 2009) for speaker identification.

Neural network applications for classification have been offered in recent years (Xiang and Berger, 2003; Specht, 1991). They are widely applied in data analysis and speaker identification. The advantage of artificial neural network is that the transfer function between the input vectors and the target matrix (output) does not have to be predicted in advance. Artificial neural network performance depends mostly on the size and quality of the training examples (Daqrouq, 2011; Daqrouq *et al*., 2010). When the number of training data is not big and not representative of the possible space, standard neural network results are bad. Fuzzy theory has been utilized effectively in numerous applications to reduce the dimensionality of feature vector of the pattern (Gowdy and Tufekci, 2000). There are many kinds of artificial neural network models among which the Back-Propagation Neural Network (BPNN) model is the most widely used (Yuanyou *et al*., 1997). The Generalized Regression Neural network (GRN) was introduced by Yuanyou *et al*. (1997) proposed a probabilistic neural network for speaker identification.

In fact, LPC is popular and widely used because it's coefficients represent a speaker by modeling vocal tract parameters and the data size are very suitable for speaker and speech recognition. Many algorithms were developed to find a better representation of a speaker by means of a linear predictive coding technique (Adami and Barone, 2001; Haydar *et al*., 1998; Wu and Lin, 2009a). The predictor coefficients themselves are hardly exploited as features but they are changed into robust and less correlated features such as Linear Predictive Cepstral Coefficients (LPCCs) (Huang *et al*., 2001a), Line Spectral Frequencies (LSFs) (Huang *et al*., 2001b) and Perceptual Linear Prediction Coefficients (PLPC) (Delac *et al*., 2009). PLP is common as a state of the art for speech recognition duty. Other somewhat less operational features include partial correlation coefficients (PARCORs), log area ratios (LARs) and speech formant frequencies and bandwidths (Avci, 2007; Chau *et al*., 1997). In the present study, the focus will be on modifying LPC coefficients and reducing the dimensionality of feature vectors.

In this study, the authors improve an effectual and a novel feature extraction method for text-independent systems, taking in consideration that the size of neural network input is a very crucial issue. This affects superiority of the training set. Hence, the presented features extraction method suggests a reduction in the dimensionality of speech signals. The proposed method is based on average framing LPC in conjunction with WT upon suitable level with an appropriate wavelet function

(Daubechies-type1 which is known as Haar function). For classification, FFBPN is proposed to accomplish online operations in a speedy manner.

MATERIALS AND METHODS

Wavelet packet transform feature extraction method: To decompose a speech signal into Wavelet Packet Transform (WPT), the common form of the equivalent low pass of discrete time speech signal (Daqrouq and Al Azzawi, 2012) is used:

$$u(t)=\sum_{m} X_m\, p(t-mT) \tag{1}$$

where, X_m is a sequence of discrete speech signal values which are obtained by a data acquisition stage; the signal p(t) is a pulse whose figure represents an important signal design problem when there is a bandwidth restriction on the channel and T is the sampling time. Considering that ϕ(t-mT) is a scaling function of a wavelet packet, i.e., $\phi \in W_{2^N}^0$, then a finite set of orthogonal subspaces can be constructed as (Sarikaya and Hansen, 2000; Souani *et al*., 2000):

$$W_{2^N}^0=\underset{(l,n)\in \rho N}{\oplus} W_{2^l}^0 \tag{2}$$

where, $W_{2^N}^0 \subset L^2(R), \rho N=\{(l,n)\}$ is a dyadic interval that forms a disjoint covering of $[0,2^N]$, $W_{2^l}^n$ denoting the closed linear span of process $\sqrt{2^l}\psi_n(2^l t-m), m\in Z$ and $\{\psi_n(t)\}_{n\in N}$ is called the wavelet packet considered by the scaling function ϕ. Therefore, the speech signal model in Eq. 1 is customized as:

$$u(t)=\sum_{m}\sum_{(l,n)\in \rho N} X_m \sqrt{2^l}\psi_n(2^l t-m) \tag{3}$$

The speech signal model in Eq. 3 is the basic form of wavelet packet transform which is used in signal decomposition. The signal is represented by orthogonal functions which shape a wavelet packet composition in $W_{2^N}^0$ space. The Discrete Wavelet Packet Transforms (DWPT) procedure is used as:

$$\phi_{l+1}^{2n}(i)=\sum_{k\in Z} h(k-2i)\phi_l^n(k) \tag{4}$$

$$\phi_{l+1}^{2n+1}(i)=\sum_{k\in Z} g(k-2i)\phi_l^n(k) \tag{5}$$

where, $\phi_{l+1}^n \in W_{2^{l+1}}^n$ and $\phi_l^n \in W_{2^l}^n$. These two processes can be carried out recursively by proceeding through the binary tree structure with 0(N log N) computational complexity. Using Eq. 3, 4 and 5, the coefficients of the linear combination may be shown to be the reversed versions of the decomposition sequences h(k) and g(k) (with zero padding), respectively. Continuously, $\phi_0^l(i)$ is reconstructed via the terminal functions of an arbitrary tree-structured decomposition:

$$\phi_0^l(i)=\sum_{l\in L,\, n\in C_l}\sum_{k\in Z} f_{ln}(i-2^l k)\phi_l^n(k) \tag{6}$$

where, L is the set of levels having the terminals of a given tree, C_l is the set of indices of the terminals at the lth level and f_{ln} (i) is the equivalent sequence generated from the combination of h(k), g(k) and decimation operation which leads from the root to the (l, n)th terminal, i.e:

$$\phi_l^n(i)=\sum_{k\in Z} f_{ln}(k-2^l i)\phi_0^l(k) \tag{7}$$

For a certain tree structure, the function ϕ_l^n in Eq. 7 is called the constituent terminal function of ϕ_0^l. In this study, the tree consists of two stages and therefore has three high pass nodes and three low pass nodes.

The wavelet packet is used to extract additional features to guarantee a higher recognition rate.

WPT data is not proper for classification due to the large amount of data length (for example, a speech signal with 35582 samples will reach 71166 samples after WPT decomposition at level two and double that at level three and so on). Thus, a better representation of the speech features is needed. Avci (2009) proposed a method to calculate the entropy value of the wavelet norm in digital modulation recognition. In the biomedical field, Behroozmand and Almasganj (2007) presented a combination of genetic algorithm and wavelet packet transform used in the pathological evaluation and the energy features are determined from a group of wavelet packet coefficients. Sarikaya and Hansen (2000) proposed a robust speech recognition scheme in a noisy environment by using wavelet-based energy as a threshold for denoising estimation. Wu and Lin (2009b) proposed the energy indexes of WP for speaker identification. Sure entropy is calculated for the waveforms at the terminal node signals obtained from DWT (Avci, 2009) for speaker identification. Avci (2007) proposed a features extraction method for speaker recognition based on a combination of three entropy types (sure, logarithmic energy and norm). In this study, LPCC obtained from WP tree nodes for speaker feature vector constructing is used for speaker identification (Mallat, 1998; Daqrouq and Al Azzawi, 2012).

Discrete wavelet transform feature extraction method: The DWT indicates an arbitrary square integrable function as a superposition of a family of basic functions. These functions are wavelet functions. A family of wavelet basis functions can be produced by translating and dilating the mother wavelet (Mallat, 1989). The DWT coefficients can be generated by taking the inner product between the original signal and the wavelet functions. Since the wavelet functions are translated and dilated versions of each other, a simpler algorithm, known as Mallat's pyramid tree algorithm, has been proposed (Mallat, 1989).

The DWT can be utilized as the multi-resolution decomposition of a sequence. It takes a length N sequence a(n) as the input and produces a length N sequence as the output. The output N/2 has values at the highest resolution (level 1) and N/4 values at the next resolution (level 2) and so on. Let $N = 2^m$ and number of frequencies or resolutions be m while, bearing in mind that m=log N octaves. So the frequency index k varies as 1, 2,…, m corresponds to the scales 2^1, 2^2, ..., 2^m. In Fig. 1, Mallat's pyramid algorithm is illustrated which presents the DWT sub signal generation at each level. DWT coefficients of the previous stage are expressed as follows (Daqrouq and Al Azzawi, 2012; Tang *et al*., 1996):

$$W_L(n,k) = \sum_i W_L(i,k-1)h(i-2n) \quad (8a)$$

$$W_H(n,k) = \sum_i W_L(i,k-1)g(i-2n) \quad (8b)$$

where, $W_L(p, j)$ is the pth scaling coefficient at the jth stage, $W_H(p, j)$ is the pth wavelet coefficient at the jth stage and h(n), g(n) are the dilation coefficients relating to the scaling and wavelet functions, respectively.

In the last decade, there has been an enormous increase in the applications of wavelets in various scientific fields. Typical applications of wavelets include signal processing, image processing, security systems, numerical analysis, statistics, biomedicine, etc. Wavelet transform tenders a wide variety of useful features, on the contrary to other transforms, such as Fourier transform or cosine transform. Some of these are as follows:

- Adaptive time-frequency windows
- Lower aliasing distortion for signal processing applications
- Computational complexity of O(N) where, N is the length of data
- Inherent scalability

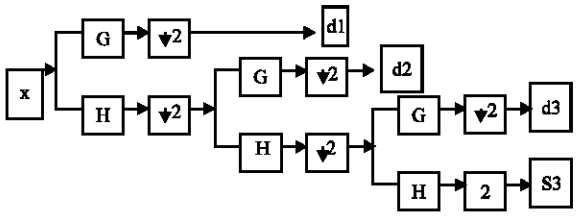

Fig. 1: DWT-tree by Mallat's algorithm

Delac *et al.* (2009) proposed DWT for face recognition. In Tufekci and Gowdy, 2000 and Gowdy and Tufekci, 2000, the use of DWT for speech recognition which has a good time and frequency resolution is proposed instead of the Discrete Cosine Transform (DCT) to solve the problem of high frequency artifacts being introduced due to abrupt changes at window boundaries. The features based on DWT and WPT were chosen to evaluate the effectiveness of the selected feature for speaker identification (Wu and Lin, 2009a). Daqrouq (2011) stated that the use of a DWT approximation sub signal via several levels instead of the original imposter had good performance on AWGN facing, particularly on levels 3 and 4 in the text-independent speaker identification system. Therefore, LPCC obtained from DWT tree nodes for speaker feature vector construction is used for text-independent speaker identification.

Modified DWT (MDWT) is proposed in this text for comparison with the proposed method which is achieved by applying the same Mallat operation to the high frequency sub signal (d_1) as well as to the low frequency. This assists greatly in expanding the utility of DWT via a high pass band of frequency.

Average framing LPC feature extraction method: Before the stage of features extraction, the speech signals are processed by a silence removing algorithm followed by the normalization of the speech signals to make the signals comparable regardless of differences in magnitude. The signals are normalized by using the following formula (Daqrouq and Al Azzawi, 2012; Wu and Lin, 2009a):

$$S_{Ni} = \frac{S_i - \bar{S}}{\sigma} \quad (9)$$

where, S_i is the ith element of the signal, $s, \bar{s}$ and σ are the mean and standard deviation of the vector S, respectively and S_{Ni} is the ith element of the signal series S_N after normalization.

LPC is not a new technique. It was developed in the 1960s by Atal (2006) but is admired and widely used to

this day because the LPC coefficients representing a speaker by modeling vocal tract parameters and the data size are very suitable (Wu and Lin, 2009b). In the proposed study, the focus will be on modifying LPC coefficients for reducing the size of feature vectors based on the author's previous study (Daqrouq and Al Azzawi, 2012). In this study, it is proposed to use the AFLPC to extract features from Z frames of each WT speech sub signal:

$$\{u_q(t)\} = \{u_{q1}(t),\ u_{q2}(t),\ ...,\ u_{qZ}(t)\} \quad (10)$$

where, Z is the number of considered frames (each frame of 20 ms duration) for the qth WT sub signal $u_q(t)$. The average of LPC coefficients calculated for Z frames of $u_q(t)$ is utilized to extract a wavelet sub signal feature vector as follows:

$$aflpc_q = \sum_{z=1}^{Z} LPC(u_{qz}(t))\frac{1}{Z} \quad (11)$$

The feature vector of the whole given speech signal is:

$$AFLPC = \{aflpc_1,\ aflpc_2,...,\ aflpc_Q\} \quad (12)$$

The superiority of the proposed feature extraction method over a conventional one is shown in Fig. 2 where, Fig. 2a illustrates two feature vectors taken for a single speaker using LPC from WP at level two. It can be seen that the LPC coefficients have similar shape but are dispersedly distributed. Figure 2b illustrates two feature vectors taken for the same speaker using AFLPC from WP at level two. This figure shows these coefficients distributed very well after using AFLPC.

Classifications: Feed-forward networks are typically composed of multi-layer nodes (Fig. 3) (Daqrouq, 2011). The feed-forward network is illustrated in Fig. 3. The direction of the data goes only in one-way (i.e., forward). Consider a three-layer neural network which receives an input vector signal X, processes it to the hidden layer and then to the output layer to give an output vector Z.

The connecting arrows between the nodes have weights (the network variables) and therefore the output of each node is related to linear summation of the nodes inputs multiplied by their connecting weights. An

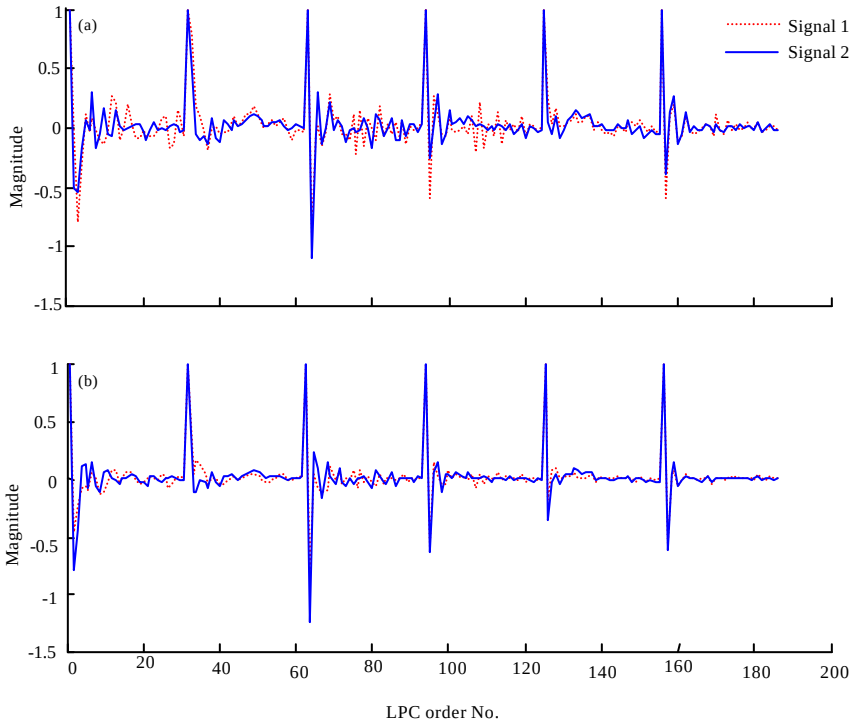

Fig. 2(a-b): Two feature vectors taken for a single speaker illustrating feature vector at level two, (a) Using LPC from WP and (b) Using AFLPC from WP

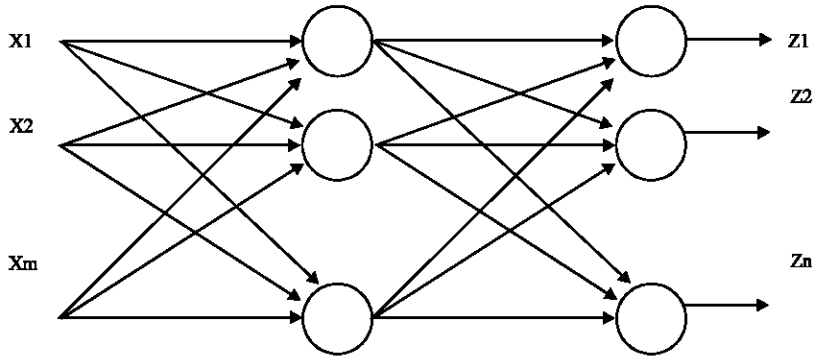

Fig. 3: Feed-forward multi-layer architecture

activation function is used to fire the neurons. A typical node output is calculated as follows:

$$\text{Node Output} = f(\Sigma(\text{Weights} \times \text{Inputs})) \quad (13)$$

where, f is the activation function. The most commonly used activation functions are the sigmoidal and hyper-tangent.

Let, D be the target vector representing the desired output of the network. The learning objective is to determine the weights values that minimize the difference between the desired output D and the computed output Z for all the patterns. Let the error criterion be defined as follows:

$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{k=1}^{n}\left(z_{pk} - d_{pk}\right)^2 \quad (14)$$

where, p refers to the pattern number and k refers to the output node number. The weights are updated, recursively:

$$w^{iter+1} = w^{iter} + D_w^{iter} \quad (15)$$

Here, D_w^{iter} are the update directions. One update option is to use the steepest descent where:

$$D_w^{iter} = -\alpha \nabla E_w^{iter}$$

where, ∇E_w^{iter} is the error gradient and α is the step size. Since the target values at the hidden layer are not available, a chain rule is used to approximate the hidden error. This algorithm is called backpropagation (Daqrouq and Al Azzawi, 2012).

A more efficient update is the use of Levenburg-Marquardt method (rather than the steepest descent). This is given in the form:

$$D_w^{iter} = -\left[\varepsilon^{iter} I + H\left(w^{iter}\right)\right]^{-1} \nabla E_w^{iter} \quad (16a)$$

where, H is the Hessian matrix and $\varepsilon^{iter}>0$ is used for conditioning the matrix inversion.

A major application for neural networks is pattern classification. In this study, the formant and wavelet information presented in the two previous sections were used as input/output data for the neural network for classification. The total number of inputs used was 12 (five formants and seven entropies).

In the context of the proposed study, the extracted features (five formants and seven entropies) are to be calculated for each person from filtered speech signals. For filtration, multistage wavelet enhancement method is used. The input matrix, X, will contain n columns (that represent the number of speakers). Each column should have 12 entrees: Five formants and seven entropies:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{12,1} & x_{12,2} & \cdots & x_{12,n} \end{bmatrix} \quad (16b)$$

The target output matrix is a binary-decoded matrix. As an example, for six speakers, the desired output is

$$D = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (17)$$

However, in order to improve the performance of the network, several patterns will be recorded for each person (not just one column). Therefore, the final input and output matrices will be of the form:

$$X = [X_1, X_2, \ldots, X_t] \quad (18)$$

$$D = [D_1, D_2, ..., D_r] \quad (19)$$

where, r is the number of recordings for each person.

RESULTS

The experimental setup was as follow: Speech signals were recorded via a PC sound card with a spectral frequency of 4000 Hz and sampling frequency of 8000 Hz. Fifty persons participated in the recordings. Each participant recorded a minimum of 20 different utterances in Arabic language. The age of the speakers varied from 20 to 45 years included 28 males and 22 females. The recording process was done under normal university office conditions.

Based on stated results in Daqrouq and Al Azzawi, 2012, an LPC order of 30 for each frame will be used. It was determined based on the Genetic Algorithm (GA) and empirically as a tradeoff between the recognition rate and the feature vector length. The first step was speech signals silence removing algorithm. This process is followed by the application of a pre-process by applying the normalization on speech signals. This stage makes the signals comparable regardless of differences in magnitude before extracting the feature vector. The performance of the AFLPC method was evaluated by FFBPN classifier which is not only rapid in the training procedure but also has the potential for the real-time applications after the off-line training stage. Table 1, presents the parameters of the FFBPN classifier for the best performance.

In the first experiment, AFLPC with WP is applied to reveal the correlation between the wavelet function and the recognition rate. Four wavelet functions were determined: db 1, 2, 3 and 4 in term of the recognition rate. The next experiment gave the results of the recognition rate by means of the proposed method for the WP level 4. When the recognition rate exceeds 96%, it did not produce essential improvement in the performance by using different wavelet families. The results were as follows: 96.04, 96.87, 96.53 and 95.22% for db 1, 2, 3 and 4, respectively. However, db 2 has the best results.

A comparative study of the proposed feature extraction method with other feature extraction methods was performed. All the signals were contaminated with Additive White Gaussian Noise (AWGN) at 10 dB. Genetic Wavelet Packet Neural Network (GWPNN) (Lei *et al*., 2005), modified DWT with conventional LPC (MDWTLPC), Eigen vector with conventional LPC (Behroozmand and Almasganj, 2007) in conjunction with WP (EWPLPC) or with DWT (EDWTLPC) are employed for comparison. The results are presented in Table 2. For all these methods, FFBPN classifier is utilized. The best recognition rate selection obtained was 87.94% for WPLPCF (Table 2).

The following experiment investigates the proposed method in term of recognition rate in additive white Gaussian (AWGN), restaurant, babble and train station noises with 5 and 0 dB. The results of WPLPCF and

Table 1: Used parameters in the proposed network that are selected empirically for the best performance of the neural network

Description	Functions
Feed forward back propagation	Network type
Four layers: Input, two hidden and output	No. of layers
500-Input, 20-hidden and 5-output	No. of neurons in layers
DOTPROD	Weight function
Levenberg-marquardt backpropagation	Training function
Log-sigmoid	Activation functions
10^{-5}	Performance function (ms)
200	No. of epochs

Table 2: Recognition results for comparison with AWGN at 10 d B using different identification methods

Identification methods	Recognition rate (%)
Genetic wavelet packet neural network (GWPNN)	77.56
Modified DWT with conventional LPC (MDWTLPC)	81.46
Eigen vector with conventional LPC with WP (EWPLPC)	83.78
Eigen vector with conventional LPC with DWT (EDWTLPC)	81.37
WP and AFLPC (WPLPCF)	87.94

Table 3: Comparison between DWT and WP with AWGN and restaurant noise

	Recognition rate (%) AWGN (dB)		Recognition rate (%) restaurant noise (dB)	
Identification methods	0	5	0	5
WP and AFLPC (WPLPCF)	61.22	75.45	44.56	65.43
DWT and AFLPC (DWTLPCF)	55.57	60.08	44.32	54.90

Table 4: Comparison between DWT and WP with babble and train noise

	Recognition rate (%) babble noise (dB)		Recognition rate (%) train station noise (dB)	
Identification methods	0	5	0	5
WP and AFLPC (WPLPCF)	56.43	66.78	57.36	69.87
DWT and AFLPC (DWTLPCF)	46.32	56.56	53.67	65.65

Table 5: Recognition results for comparison with babble noise for 5 dB

Identification methods	Recognition rate (%)
Genetic wavelet packet neural network (GWPNN)	50.43
Modified DWT with conventional LPC (MDWTLPC)	54.64
Eigen vector with conventional LPC with WP (EWPLPC)	64.58
Eigen vector with conventional LPC with DWT (EDWTLPC)	60.35
WP and AFLPC (WPLPCF)	66.78

DWLPCF are tabulated in Table 3 and 4. DWT was processed at level 5 with 6 sub signals while WP was processed at level 4 with bigger number of sub signals. It was found that the recognition rates of WPLPCF were the best.

A comparative study between proposed feature extraction method and other feature extraction methods was performed. GWPNN, MDWTLPC, EWPLPC and EDWTLPC are employed for comparison in babble noise environment for 5 dB. The results are presented in Table 5. For all these methods, FFBPN classifier is utilized. The best recognition rate selection obtained was 66.78% for WPLPCF (Table 5).

CONCLUSION

This study presented a speaker identification system based AFLPC in noise environments. The benefit of AFLPC is its capability to reduce the huge speech data into a fewer values while accomplishing good computing speed. In the beginning of feature extraction, WT is applied with LPC coefficients by analyzing the vocal tract parameters of a speaker. Then AFLPC coefficients are extracted from LPC obtained from wavelet coefficients and used as a representative speaker feature vector. For classification, FFBPN was applied. The speaker identification performance of this method was demonstrated on a total of 50 individual speakers. Four different noise types are investigated. Experimental results showed that WP resulted in better performance in terms of recognition rate. As a comparison with other published methods in noisy environments, WPLPCF produced a higher recognition rate. The experimental results revealed the proposed AFLPC technique with WP at level 4 can accomplish better results for a speaker identification system in noisy environments.

ACKNOWLEDGEMENT

This study was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah. The authors, therefore, acknowledge with thanks DSR technical and financial support.

REFERENCES

Adami, A.G. and D.A.C. Barone, 2001. A speaker identification system using a model of artificial neural networks for an elevator application. Inform. Sci., 138: 1-5.

Afify, M. and O. Siohan, 2007. Comments on vocal tract length normalization equals linear transformation in cepstral space. IEEE Trans. Audio Speech Lang. Process., 15: 1731-1732.

Antonini, M., M. Barlaud, P. Mathieu and I. Daubechies, 1992. Image coding using wavelet transform. IEEE Trans. Image Process., 1: 205-220.

Atal, B.S., 2006. The history of linear prediction. IEEE Signal Process. Mag., 23: 154--161.

Avci, E., 2007. A new optimum feature extraction and classification method for speaker recognition: GWPNN. Exp. Syst. Appl., 32: 485-498.

Avci, D., 2009. An expert system for speaker identification using adaptive wavelet sure entropy. Exp. Syst. Appl., 36: 6295-6300.

Behroozmand, R. and F. Almasganj, 2007. Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis. Comput. Biol. Med., 37: 474-485.

Chau, C.W., S. Kwong, C.K. Diu and W.R. Fahrner, 1997. Optimization of HMM by a genetic algorithm. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Volume 1, April 21-24, 1997, Munich, pp: 1727-1730.

Chen, C.T., S.Y. Lung, C.F. Yang and M.C. Lee, 2002. Speaker recognition based on 80/20 genetic algorithm. Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Application, June 25-28, 2002, Greece, pp: 547-549.

Daqrouq, K., I.N. Abu-Isbeih, O. Daoud and E. Khalaf, 2010. An investigation of speech enhancement using wavelet filtering method. Int. J. Speech Technol., 13: 101-115.

Daqrouq, K., 2011. Wavelet entropy and neural network for text-independent speaker identification. Eng. Appl. Artif. Intell., 24: 796-802.

Daqrouq, K. and K.Y. Al Azzawi, 2012. Average framing linear prediction coding with wavelet transform for text-independent speaker identification system. Comput. Electr. Eng., 38: 1467-1479.

Davis, S. and P. Mermelstein, 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process., 28: 357-366.

Delac, K., M. Grgic and S. Grgic, 2009. Face recognition in JPEG and JPEG2000 compressed domain. Image Vision Comput., 27: 1108-1120.

Gowdy, J. and Z. Tufekci, 2000. Mel-scaled discrete wavelet coefficients for speech recognition. Acoustics Speech Signal Process., 3: 1351-1354.

Haydar, A., M. Demirekler and M.K. Yurtseven, 1998. Speaker identification through use of features selected using genetic algorithm. Elect. Lett., 34: 39-40.

Hosseinzadeh, D. and S. Krishnan, 2007. Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, April 15-20, 2007, Honolulu, HI., pp: 365-368.

Huang, X., A. Acero and H.W. Hon, 2001a. Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Prentice-Hall, New Jersey, ISBN-13: 9780130226167, Pages: 980.

Huang, Y.C., S.L. Wueng, C.C. Ou, C.H. Cheng and K.H. Su, 2001b. Nutritional status of functionally dependent and nonfunctionally dependent elderly in Taiwan. J. Am. Coll. Nutr., 20: 135-142.

Lei, Z., L. Jiandong, L. Jing and Z. Guanghui, 2005. A novel wavelet packet division multiplexing based on maximum likelihood algorithm and optimum pilot symbol assisted modulation for Rayleigh fading channels. Circuits Syst. Signal Process., 24: 287-302.

Long, C.J. and S. Datta, 1996. Wavelet based feature extraction for phoneme recognition. Proceeding of the 4th International Conference of Spoken Language Processing, October 3-6, 1996, Philadelphia, USA., pp: 264-267.

Lung, S.Y. and C.C.T. Chen, 1998. Further reduced form of Karhunen-Loeve transform for text independent speaker recognition. Electron. Lett., 34: 1380-1382.

Lung, S.Y., 2004. Feature extracted from wavelet eigenfunction estimation for text-independent speaker recognition. Pattern Recognition, 37: 1543-1544.

Lung, S.Y., 2010. Improved wavelet feature extraction using kernel analysis for text independent speaker recognition. Digital Signal Process., 20: 1400-1407.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell., 11: 674-693.

Mallat, S., 1991. Zero-crossings of a wavelet transform. IEEE Trans. Inform. Theory, 37: 1019-1033.

Mallat, S. and S. Zhong, 1992. Characterization of signals from multiscale edges. IEEE Trans. Pattern Anal. Machine Intell., 14: 710-732.

Mallat, S., 1998. A Wavelet Tour of Signal Processing. 2nd Edn., Academic Press, San Diego, CA pp: 250-252.

Mitchell, R.A., 1997. Hybrid statistical recognition algorithm for aircraft identification. Ph.D. Thesis, University of Dayton, Dayton, OH.

Nathan, K.S. and H.F. Silverman, 1994. Time-varying feature selection and classification of unvoiced stop consonants. IEEE Trans. Speech Audio Process., 2: 395-405.

Quiroga, R.Q., 1998. Quantitative analysis of EEG signals: Time-frequency methods and chaos theory. Ph.D. Thesis, Institute of Physiology, Medical University Lubeck, Lubeck, Germany.

Sarikaya, R. and J.H.L. Hansen, 2000. High resolution speech feature parametrization for monophone-based stressed speech recognition. IEEE Signal Process. Lett., 7: 182-185.

Sarikaya, R., B.L. Pellom and J.H.L. Hansen, 1998. Wavelet packet transform features with application to speaker identification. Proceedings of the IEEE Nordic Signal Processing Symposium, June 1998, Vigso, Denmark, pp: 81-84.

Souani, C., M. Abid, K. Torki and R. Tourki, 2000. VLSI design of 1-D DWT architecture with parallel filters. Integration, VLSI J., 29: 181-207.

Specht, D.F., 1991. A general regression neural network. Trans. Neural Network, 2: 568-576.

Tang, K.S., K.F. Man, S. Kwong and Q. He, 1996. Genetic algorithms and their applications. IEEE Signal Process. Magazine, 13: 22-37.

Tufekci, Z. and J.N. Gowdy, 2000. Feature extraction using discrete wavelet transform for speech recognition. Proceedings of the SOUTHEASTCON, April 7-9, 2000, Nashville, TN., USA., pp: 116-123.

Vitterli, M. and J. Kovacevic, 1995. Wavelets and Subband Coding. Vol. 87, Prentice Hall, Englewood Cliffs, New Jersey,.

Wu, J.D. and B.F. Lin, 2009a. Speaker identification based on the frame linear predictive coding spectrum technique. Exp. Syst. Appl., 36: 8056-8063.

Wu, J.D. and B.F. Lin, 2009b. Speaker identification using discrete wavelet packet transform technique with irregular decomposition. Expert Syst. Appl., 36: 3136-3143.

Xiang, B. and T. Berger, 2003. Efficient text-independent speaker verification with structural *Gaussian mixture* models and neural network. IEEE Trans. Speech Audio Proc., 11: 447-456.

Yuanyou, X., X. Yanming and Z. Ruigeng, 1997. An engineering geology evaluation method based on an artificial neural network and its application. Eng. Geol., 47: 149-156.