Wavelet LPC With Neural Network for Speaker Identification System

KHALED DAQROUQ[#1], ALI MORFEQ, MOHAMMAD AJOUR, AND ABDULHAMEED ALKHATEEB
Electrical and Computer Engineering Department, King Abdulaziz University
[#1]P.O. Box 80204, Jeddah 21589, Tel +966 5 66 980400, Fax: +966-5-695-2686,
E-Mail:haleddaq@yahoo.com.

Abstract: In this study, an average framing linear prediction coding (AFLPC) technique for text-independent speaker identification systems is proposed.The study of the combination of modified LPC with wavelet transform (WT), termed AFLPC, is presented for speaker identification based on our previous paper. The study procedure is based on feature extraction and voice classification. In the phase of classification, feed forward backprobagationneural network (FFBPN) is applied because of its rapid response and ease in implementation. In the practical investigation, performance of different wavelet transforms in conjunction with AFLPC were compared with one another. In addition, the capability analysis on the proposed system was examined by comparing it with other systems proposed in literature. Consequently, the FFBPNclassifier achieves a better recognition rate (97.36%) with the wavelet packet (WP) and AFLPC termed WPLPCF feature extraction method. It is also suggested to analyze the proposed system in additive white Gaussian noise (AWGN) and real noise environments.

Keywords: Speech; LPC; Average framing; Wavelet; Neural network.

1. Introduction

Speaker recognition (SR) has been studied by a large number of researchers for about four decades [1]. From a commercial viewpoint, SR is a tool with a potentially large market due to its wide range of application from the automation of operator-assisted service to speech-to-text aiding systems for hearing-impaired individuals [2].

A commonly used technique for feature extraction is based on the Karhunen–Loeve transform (KLT). These models have been applied to text-independent speaker recognition cases [3] with exceptional results. Karhunen–Loeve transform is the optimal transform according to minimum mean square error (MMSE) and maximal energy packing. Most of the suggested speaker identification systems use mel frequency cepstral coefficient (MFCC) [5] and linear predictive cepstral coefficient (LPCC) [6] as features. Although MFCC and LPCC have proved to be two very good features in speech recognition,the disadvantage of the MFCC is that it uses short time Fourier transform, which has a weak time-frequency resolution and an assumption that the signal is stationary. Therefore it is relatively difficult to recognize plosive phonemes by these features. Currently, some researches [7], [8], [9] are focusing on the wavelet transform for speaker feature extraction.

Wavelet transform [4], [3], [11] has been extensively considered in the last two decades and has been widely utilized in various areas of science and engineering. The wavelet analysis process is implemented with dilated and translated versions of a mother wavelet. Since signals of interest can generally be expressed using wavelet decompositions, signal processing algorithms can be implemented by adjusting only the corresponding 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 [1]. From a practical point of view, however, in order to improve computation efficiency, the values of the shift and scale parameters are often limited to some discrete lattices [12], [13].

Wavelet and WP analysis have been proven as effectual signal processing techniques for a variety of digital signal processing problems. Wavelets have been used in two different methods in feature extraction plans designed for the task of speech/voice recognition. Discrete wavelet transform in place of Discrete Cosine Transform is utilized for the feature extraction period in the first method [16]. In the second method, wavelet transform is used directly on the speech/voice signals and either wavelet coefficients containing high energy are extracted as features [8] but suffer from shift variance, or sub band energies are used instead of the Mel filter-bank sub band energies proposed in [17]. Particularly, WP bases are used in [18] as close approximations of the Mel-frequency division using Daubechies orthogonal filters. In [19], a feature extraction method based on the wavelet Eigen function was proposed. Wavelets can offer a significant computational benefit by reducing the dimensionality of the Eigen value problem. A text-independent speaker identification system based on improved wavelet transform is proposed in [9], where learning of the correlation between the wavelet transform and the expression vector is performed by kernel canonical correlation analysis.

The wavelet packets transform (WPT) performs the recursive decomposition of the speech signal obtained by the recursive binary tree. Basically, the WPT is very similar to discrete wavelet transform (DWT).However, WPT decomposes both details and approximations instead of only performing the decomposition process on approximations. WPT features have superior presentation than those of the DWT [19]. Nevertheless, as the number of wavelet packet bases grows, the time required to appropriately classify the database will become nonlinear. Consequently, dimensionality decreasing 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 [20] & [21]. 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 perceptual decomposition tree was first proposed by R. Sarikaya [22] and yields the wavelet packet parameters (WPP). In [24], the energy indexes of DWT or WPT were proposed for speaker identification, where WPT was superior in terms of recognition rate. Sure entropy was calculated for the waveforms at the terminal node signals obtained from DWT [25] [56, 57] for speaker identification.

Neural network applications for classification have been considered in recent years [30], [15]. They are widely applied in data analysis and speaker identification. The advantage of the 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 mainly on the size and quality of training samples [28], [29]). When the number of training data is small, not representative of the possible space, standard neural network results are poor. Fuzzy theory has been used successfully in many applications to reduce the dimensionality of feature vector [31]. There are many kinds of artificial neural network models, among which the back-propagation neural network (BPNN) model is the most widely used [32]. The generalized regression neural network (GRN) was introduced by [32].Daqrouq[33] proposed a probabilistic neural network for speaker identification.

In fact, LPC is popular and widely used because its coefficients representing 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 [35], [36], [23]. The predictor coefficients themselves are rarely utilized as features, but they are transformed into robust and less correlated features such as linear predictive cepstral coefficients (LPCCs) [37], line spectral frequencies (LSFs) [38], and perceptual linear prediction (PLP) coefficients [49]. PLP is known as a state of the art for speech recognition task. Other, somewhat less effective features include partial correlation coefficients (PARCORs), log area ratios (LARs) and formant frequencies and bandwidths [42], [51]. In the present work, the focus will be on modifying LPC coefficients and reducing the dimensionality of feature vectors.

In this research, 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 quality of the training set. For this reason, the presented features extraction method offers 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.

2. Problem Definition

In the presented study, an average framing linear prediction coding (AFLPC) method for text-independent speaker identification task is investigated.The study of the combination of modified LPC with wavelet transform (WT), termed AFLPC, is presented for speaker identification based on our previous paper. In the phase of classification, feed forward backprobagationneural network (FFBPN) is applied because of its rapid response and ease in implementation.

3. Methodology

3.1 Wavelet Packet Transform Feature Extraction Method

To decompose the speech signal into wavelet packet transform (WPT), we start from the common form of the equivalent low pass of discrete time speech signal [33]

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

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 $\phi(t-mT)$ 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 [47], [48].

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

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 $\{\psi_n(t)\}_{n\in\mathrm{N}}$ is called the wavelet packet, considered by the scaling function ϕ. Therefore, the speech signal model in (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 (3) is the basic form of wavelet packet transform, which is used in signal decomposition. The signal is carried by orthogonal functions, which shape a wavelet packet composition in $W_{2^N}^{0}$ space. We may use the discrete wavelet packet transforms (DWPT) procedure 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 (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, we can reconstruct $\phi_0^1(i)$ via the terminal functions of an arbitrary tree-structured decomposition:

$$\phi_0^1(i)=\sum_{l\in L, n\in C_l}\sum_{k\in Z} f_{ln}(i-2^l k)\phi_l^n(k), \tag{6}$$

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.,

$$\phi_l^n(i)=\sum_{k\in Z} f_{ln}(k-2^l i)\phi_0^1(k). \tag{7}$$

For a certain tree structure, the function ϕ_l^n in (7) is called the constituent terminal function of ϕ_0^1. In this work, 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. In this study, WPT is applied at the stage of feature extraction, but these data are not proper for classification due to a great amount of data length (for example, a speech signal with a number of 35582 samples will reach 71166 after WPT decomposition at level two). Thus, we have to seek for a better representation of the speech features. [27] proposed a method to calculate the entropy value of the wavelet norm in digital modulation recognition. In the biomedical field, [46] 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. [47] proposed a robust speech recognition scheme in a noisy environment by using wavelet-based energy as a threshold for denoising estimation. In [24], the energy indexes of WP were proposed for speaker identification. Sure entropy is calculated for the waveforms at the terminal node signals obtained from DWT [25] for speaker identification. [26] proposeda features extraction method for speaker recognition based on a combination of three entropies types (sure, logarithmic energy and norm). In this paper, we use LPCC obtained from WP tree nodes for speaker feature vector constructing to be used for speaker identification [33].

3.2 Discrete Wavelet Transform Feature Extraction Method

The DWT indicates an arbitrary square integrablefunction 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[14]. 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 (see Fig.2) [14].

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 let the 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 $2^1, 2^2, ..., 2^m$. As described by theMallat pyramid algorithm (Fig.1), the DWT coefficients of the previous stage are expressed as follows [33, 51]:

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

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

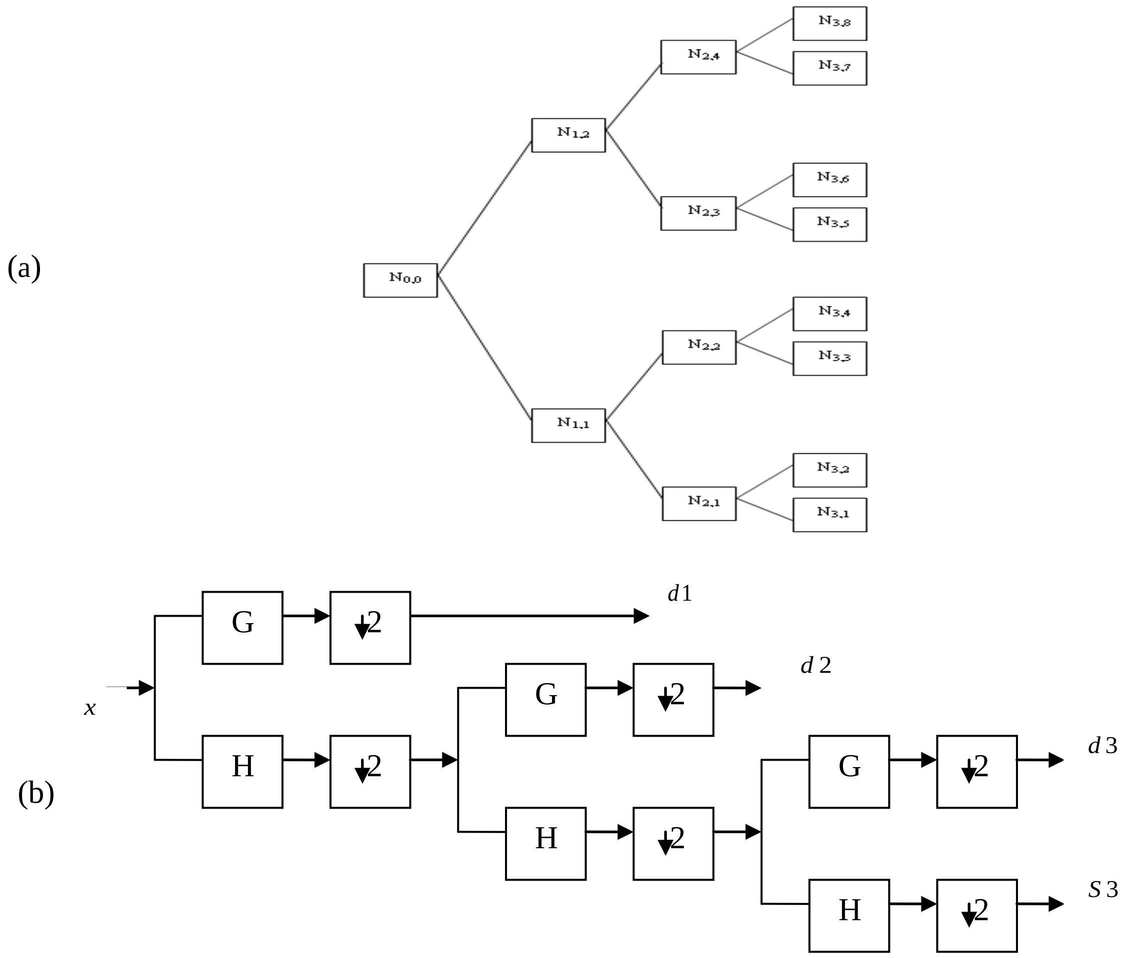


Fig. 1 (a) Wavelet packet at depth 3, (b) DWT-tree by Mallat's Algorithm where $W_L(p, j)$ is the pth scaling coefficient at the jth stage, $W_H(p, j)$ is the pth wavelet coefficient at the jth stage, and *h(n), g(n)* are the dilation coefficients relating to the scaling and wavelet functions, respectively.

In the last decade, there has been an enormous increase in the applications of wavelets in various scientific fields [54]. 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.

Delac[49] proposed DWT for face recognition. In [16] and [31], 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 windowboundaries. The features based on DWT and WPT were chosen to evaluate the effectiveness of the selected feature for speaker identification [23]. [28] 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, we use LPCC obtained from DWT tree nodes for speaker feature vector constructing to be 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 the low frequency. This assists greatly in expanding the utility of DWT via a high pass band of frequency.

3.3 Average Framing LPC Feature Extraction Method

Before the stage of features extraction, the speech data are processed by a silence removing algorithm followed by the application of a pre-processing, which is achieved by applying the normalization on speech signals to make the signals comparable regardless of differences in magnitude, because the distribution of these magnitudes is closely related to the volume of the speakers. The signals are normalized by using the following formula [33, 23, 53]:

$$S_{Ni} = \frac{S_i - \dddot{S}}{\sigma} \tag{9}$$

where S_i is the ith element of the signal S, $\dddot{S}$ 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 [50] but is admired and widely used to thisday because the LPC coefficients representing a speaker by modeling vocal tract parameters and the data size are very suitable for speech compression throughout the digital channel [23]. In the proposed study, the focus will be on modifying LPC coefficients for reducing the size of feature vectors based on our previous paper [33]. In our work, we propose the AFLPC to extract features from Z frames of each WT speech sub signal:

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

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:

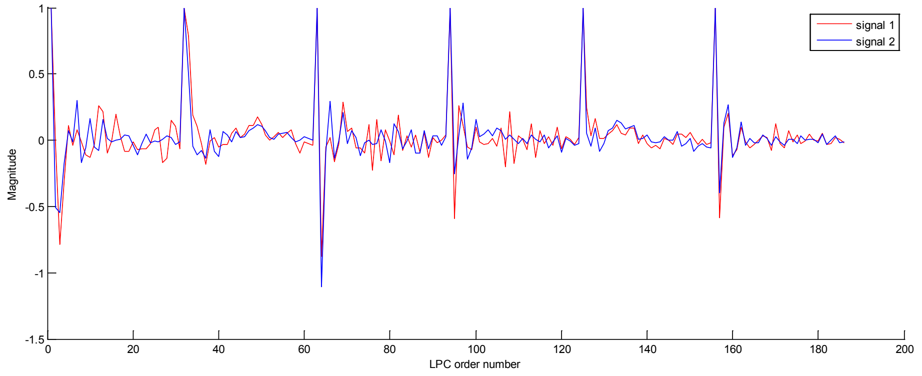
$$aflpc_q = \sum_{z=1}^{Z} LPC\,(u_{qz}(t))\frac{1}{Z} \tag{11}$$

The feature vector of the whole given speech signal is

$$AFLPC = \left\{aflpc_1, aflpc_2, ..., aflpc_Q\right\} \tag{12}$$

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. Fig. 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.

(a)



(b)

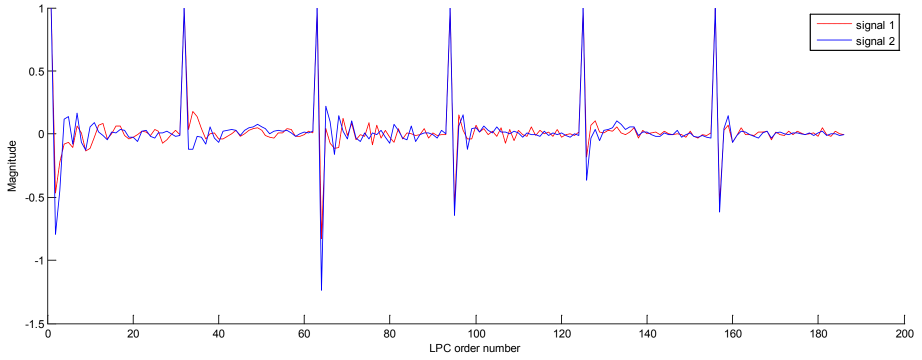


Fig.2 Two feature vectors taken for a single speaker:
(a) illustrates feature vectors using LPC from WP at level two, and
(b) illustrates feature vectors using AFLPC from WP at level two.

3.4 Classifications

This operation performs the intelligent classification by means of features obtained from feature extraction phase. In this study FFBPNN is utilized [33, 52]. The training specifications and the structure of the NNT used in this paper are as tabulated in Tab.1. These were selected for the best performance. That is accomplished after several experiments, such as the number of hidden layers, the size of the hidden layers, value of the moment constant, and type of the activation functions (transfer functions). 35x18feature matrix which isobtained in features extraction stage is given to the input of the feed-forward networks consist of many layers using the DOTPRODweight function, NETSUM net input function, and the specifiedtransfer functions.The weights of the first layer come from the input. Each network layer has a weight coming from the previous layer. All layershave biases. The last layer is the network output (target). In this paper target (T) is designed as a 47 binary digits for each features vector.

Table 1 Parameters used for the Network

Functions	Description
Network Type	Feed Forward Back Propagation
No. of Layers	Four Layers: Input, Two Hidden & Output
No. of neurons in Layers	500- Input, 20-Hidden &5-Output
Weight Function	DOTPROD
Training Function	Levenberg-Marquardt Backpropagation
Activation functions	Log- sigmoid
Performance Function (mse)	10^{-5}
No. of Epochs	200

The mean square error of the ANN is obtained at the final of the training of the ANN classifier by means of Levenberg-MarquardtBackpropagation.
Backpropagation is used to calculate the JacobianjX of performance with respect to the weight and bias variables X. Eachvariable is accustomed according to Levenberg-Marquardt,

$$jj = jX * jX$$
$$je = jX * E \tag{13}$$
$$dX = -(jj + I * Mu) \backslash je$$

Where E is all errors and I is the identity matrix.The adaptive value Mu is increased by 10 Mu increase factor until the change aboveresults in a reduced performance value. The change is then made tothe network and Mu is decreased by 0.1 Mu decrease factor. After training the 18 speaker features, imposter simulation is performed. The imposter simulation result (SR) is compared with each of the 18 patterns target (P_n, n =1,2,…,18) in order to determine the decision by

$$C_n = 100 - [100 * \sqrt{(\sum (P_n - SR)^2 / \sum P_n^{\ 2})}] \tag{14}$$

where C_n is the similarity percent between imposter simulation results and pattern target P_n. The speaker is identified as a pattern of maximum similarity percent.

4. Results and Discussion

To examine the presented text-independent speaker identification system, a testing database was created from the Arabic language. The recording environment is a normal office setting via PC-sound card, with original frequency of 4 KHz and a sampling frequency of 16 KHz.These utterances are Arabic spoken digits from 0 to 14,Each speaker also distinctly reads 30 seconds worth of different Arabic texts ten separate times. A total of 47 individual speakers (19 to 40 years old), of whom are 31 individual males and 16 individual females, spoke these Arabic words and texts for training and testing modes. The total number of tokens considered for training and testing was 1128.

Some experiments were performed using all of the 1128 Arabic utterances from these 47 individual speakers. For each of these speakers, 24 speech signals were used, of which 6 were used for the training mode and 18 for testing. The proposed system was tested by utilizing all of these speakers.

Based on stated results in [33], an LPC order of 30 for each frame will be used. It was determined based on the GA and empirically as a tradeoff between the recognition rate and the feature vector length.

The complete analysis flowchart is shown in Fig. 3, indicating out that the speech signals are processed by a 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 real-time applications after off-line training stage.

In the first experiment, AFLPC with WP is applied to reveal the correlation between the WPlevel and the recognition rate. We examined the WP with an upper limit of 7 in order to determine the feature vector of lower dimensionality. Four WPlevels were determined: 2, 4, 5 and 7 in term of the recognition rate (presents identification accuracy). Table 2gives the results ofthe recognition rate by means of the proposed method for the four WP levels. In all cases it was found that the recognition rate is proportional to WPlevel. With more coefficients, the higher recognition rate was acquired, and, the increase of WP leveldid not tremendously burden the system load. However, the use of these parameters still has its limitation since the number of parameters slightly affects the recognition rate.When the recognition rate reached over 96%, it did not produceessential improvement in the performance even though double the amount of WP coefficients (from level 6 to 7) was used. Moreover, for FFBPN, the increase in parameters also affects the training time.

Table 2: WP different levels results

No. of Speakers	WP Levels	2	4	5	7
47	Recognition Rate	91.04	92.62	96.45	96.72

Table 3: Comparison between different feature extraction methods

Identification Method	**Recognition Rate[%]**
WPID	93.11
GWPNN	89.23
MDWTLPC	91.34
EWPLPC	94.01
EDWTLPC	90.21
Shannon & WP	88.31
Sure & WP	51.98
MFGMM	90.42
Log Energy & WP	76.01
WPLPCF	96.45

In the experiments, several feature extraction methods were analyzed to expose the efficiency of thepresented system. The following experiment examines the proposed method in terms of the recognition rate. This can be concluded after interpretation of the results, where the results of DWT with conventional LPC (DWTLPC), DWT with AFLPC (DWTLPCF), WP with conventional LPC (WPLPC) and WP with AFLPC (WPLPCF) are tabulated. DWT was processed at level 5 with 6 sub signals while WP was processed at level 5 with 64 sub signals. It was found that the recognition rates of WP methods are superior (95.24&96.45) when compared with DWT methods (93.32& 95.95). The same conclusion was derived by means of the correlation coefficient method being taken for 150 different signals of 15 speakers instead of the FFBPN classifier.

A comparative study of the proposed feature extraction method with other feature extraction methods was performed. The Wavelet packet energy index distribution method (WPID) [23], genetic wavelet packet neural network (GWPNN) [45], Modified DWT with conventional LPC (MDWTLPC), Eigen vector with conventional LPC [46] in conjunction with WP (EWPLPC) or with DWT (EDWTLPC), Shannon [28], sure [25], MFCC with Gaussian mixture model (GMM) (MFGMM) [30], [34] and log energy [45] entropies methods taken for WP are employed for comparison. The results are presented in table 3. To choose the optimal WP level used for entropies and energy index methods to be used in comparison, investigation results were presented in our previous study[33]. For all these methods,FFBPN classifier is utilized. The best recognition rate selection obtained was 96.45% for WPLPCF (table 3).

4.1 Performance of the System in Noisy Environments

Another experiment was conducted to evaluate the performance of the system in noisy environments. Table 4 summarizes the results of the speaker identification corresponding to white Gaussian noise and real noise (restaurant noise, which seems like babbling) with the Signal-to-Noise ratio (SNR) of 0dB and 5dB references. SNR was calculated as follows:

$SNR = 20\log 10 \frac{\sum s}{\sum (s - s_n)}$, where s is free of a noise speech signal and s_n is a noisy

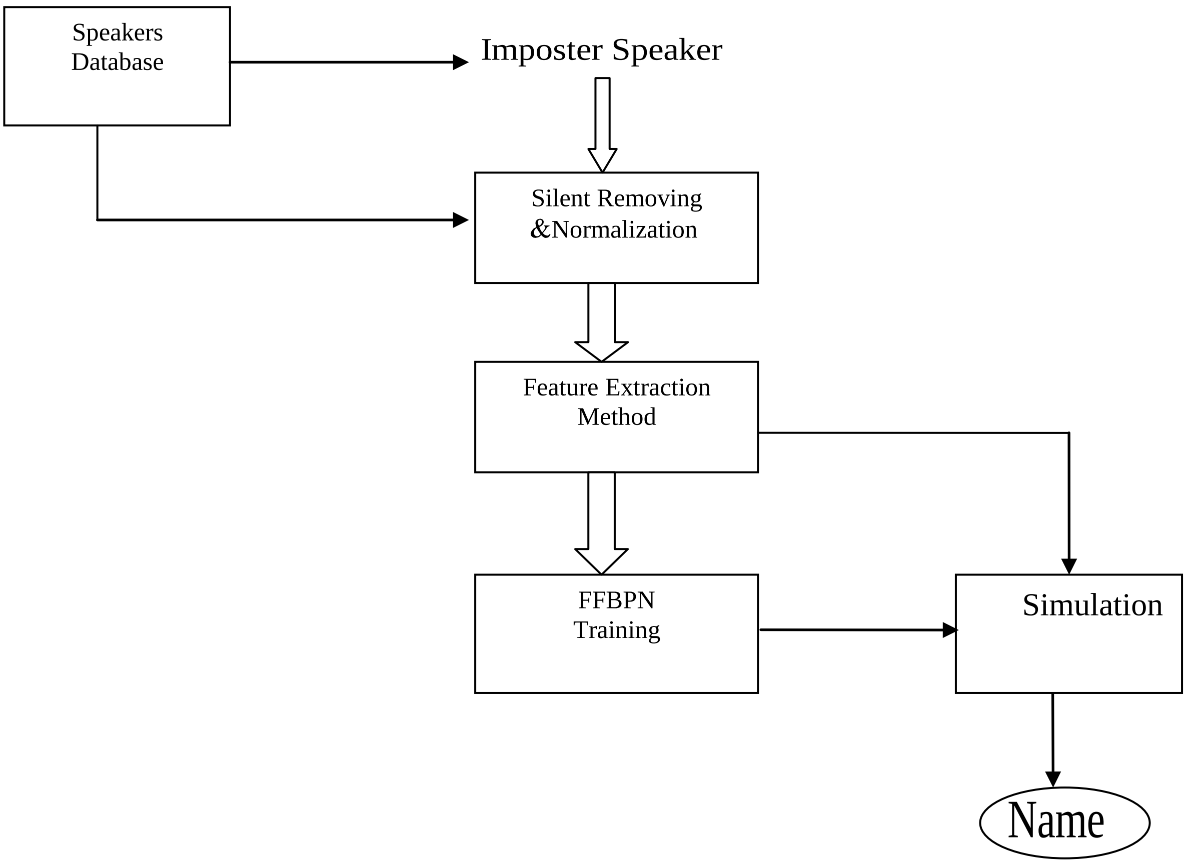


Fig. 3 Flowchart of the presented system

speechsignal.Three approaches used in the experimental investigation for comparison: WPLPCF and DWTLPCF. The recognition rate of DWTLPCF reached the lowest value (table 4). The best recognition rate selections obtained were67.51% (with 0dB) and 75.09% (with 5dB) forWPLPCF. The reason forWP success over DWT is that the feature vector is obtained from level 5 from two direction (detailed and approximation subsignals) not only one direction like in DWT (approximationsubsignal).

Table 4: Comparison between DWT and WP with AWGN

Identification Method	**Recognition Rate[%] AWGN**		**Recognition [%]Rate Restaurant Noise**	
	0dB	**5dB**	**0dB**	**5dB**
WPLPCF	67.51	75.09	47.34	67.02
DWTLPCF	56.54	61.46	44.94	50.78

5. Conclusion

This work proposed a speaker identification system based AFLPC. The benefit of AFLPC is its capability to reduce the huge speech data into a few values, and the computing speed is also accomplished. 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 is applied. The speaker identification performance of this method was demonstrated on a total of 47 individual speakers (31 male speakers and 16 female speakers). Experimental results showed that WP resulted in better performance in terms of recognition rate (96.45%). As a comparison with other published methods, WPLPCF produced a higher recognition rate. The experimental results revealed the proposed AFLPC technique with WP at level 5 can accomplish better results for a speaker identification system in an AWGN environment.

ACKNOWLEDGEMENT: THIS ARTICLE 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

[1] Antonini, Marc, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies. "Image coding using wavelet transform." *Image Processing, IEEE Transactions on* 1, no. 2 (1992): 205-220.

[2] Quiroga, Rodrigo Quian. "Quantitative analysis of EEG signals: time-frequency methods and chaos theory." *Institute of Physiology-Medical University Lubeck and Institute of Signal Processing-Medical University Lubeck* (1998).

[4] Chen, C-CT, Chin-Ta Chen, and Chih-Ming Tsai. "Hard-limited Karhunen-Loeve transform for text independent speaker recognition." *Electronics Letters* 33, no. 24 (1997): 2014-2016.

[3] Lung, Shung-Yung, and C-CT Chen. "Further reduced form of Karhunen-Loeve transform for text independent speaker recognition." *Electronics Letters* 34, no. 14 (1998): 1380-1382.

[4] Mallat S. (1998) , A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA (1998).

[5] Hosseinzadeh D., Krishnan S. (2007), Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs, in: *Proc. Int. Conf. on Signal Processing*, pp. 365–368.

[6] Afify, Mohamed, and Olivier Siohan. "Comments on vocal tract length normalization equals linear transformation in cepstral space." *Audio, Speech, and Language Processing, IEEE Transactions on* 15, no. 5 (2007): 1731-1732.

[7] Rioul O., M. Vetterli V. "Wavelets and signal processing".IEEE Trans. Signal Process. Mag. (1991): 14–38.

[8] Long, C. J., &Datta, S. "Wavelet based feature extraction for phoneme recognition". *Proceeding of the Fourth International Conference of Spoken Language Processing*, Philadelphia, USA, Vol. 1 (1996): 264–267.

[9] Lung S-Y. "Improved wavelet feature extraction using kernel analysis for text independent speaker recognition".*Digital Signal Processing*, 20, (2010): 1400–1407.

[11] Vetterli M., Kovacevic J. "Wavelets and Sub band Coding". Prentice Hall, Englewood Cli4s, NJ. (1995)

[12] Mallat, S., &Zhong, S. "Characterization of signals from multiscale edges". *IEEE Transactions on Pattern Analysis Machine Intelligence*, 14, (1992): 710–732.

[13] Mallat, S. "Zero-crossings of a wavelet transform". IEEE Transactions on Information Theory, 37, (1991): 1019–1033.

[14] Mallat S.“A Theory for Multiresolution Signal Decomposition: the Wavelet Representation" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, (1989): 674-693.

[15] Specht, D. F. “A general regression neural network.”IEEE Transactions on Neural Networks, 19(4), 1560–1568, 1991.

[16] Tufekci, Z., and J. N. Gowdy. "Feature extraction using discrete wavelet transform for speech recognition." In *Southeastcon 2000. Proceedings of the IEEE*, pp. 116-123. IEEE, 2000.

[17] Davis, Steven, and Paul Mermelstein. "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences."*Acoustics, Speech and Signal Processing, IEEE Transactions on* 28, no. 4 (1980): 357-366.

[18] Mitchell, R. A. "Hybrid statistical recognition algorithm for aircraft identification." PhD diss., Dissertation Proposal, University of Dayton, 1997.

[19] Lung, Shung-Yung. "Feature extracted from wavelet eigenfunction estimation for text-independent speaker recognition." *Pattern recognition* 37, no. 7 (2004): 1543-1544.

[20] Chen, Chin-Ta, Shung-Yung Lung, Cheng-Fu Yang, and Mo-Chung Lee. "Speaker recognition based on 80/20 genetic algorithm."In *IASTED International Conference on Signal Processing, Pattern Recognition, and Application, Greece*, pp. 547-549. 2002.

[21] Nathan, K. S., and H. F. Silverman. "Time-varying feature selection and classification of unvoiced stop consonants." *Speech and Audio Processing, IEEE Transactions on* 2, no. 3 (1994): 395-405.

[22] Sarikaya, Ruhi, Bryan L. Pellom, and John HL Hansen. "Wavelet packet transform features with application to speaker identification." In *IEEE Nordic Signal Processing Symposium*, (1998): 81-84.

[23] Wu, Jian-Da, and Bing-Fu Lin. "Speaker identification based on the frame linear predictive coding spectrum technique." *Expert Systems With Applications* 36, no. 4 (2009): 8056-8063.

[24] Wu, Jian-Da, and Bing-Fu Lin. "Speaker identification using discrete wavelet packet transform technique with irregular decomposition." *Expert Systems with Applications* 36, no. 2 (2009): 3136-3143.

[25] Avci, Derya. "An expert system for speaker identification using adaptive wavelet sure entropy."*Expert Systems with Applications* 36, no. 3 (2009): 6295-6300.

[26] Avci, Engin. "A new optimum feature extraction and classification method for speaker recognition: GWPNN." *Expert Systems with Applications* 32, no. 2 (2007): 485-498.

[27] Avci, Engin, DavutHanbay, and AsafVarol. "An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition." *Expert Systems with Applications* 33, no. 3 (2007): 582-589.

[28] Daqrouq, Khaled. "Wavelet entropy and neural network for text-independent speaker identification."*Engineering Applications of Artificial Intelligence* 24, no. 5 (2011): 796-802.

[29] Daqrouq, Khaled, Ibrahim N. Abu-Isbeih, Omar Daoud, and EmadKhalaf. "An investigation of speech enhancement using wavelet filtering method."*International Journal of Speech Technology* 13, no. 2 (2010): 101-115.

[30] Xiang, Bing, and Toby Berger. "Efficient text-independent speaker verification with structural gaussian mixture models and neural network."*Speech and Audio Processing, IEEE Transactions on* 11, no. 5 (2003): 447-456.

[31] Gowdy, J. N., and Z. Tufekci. "Mel-scaled discrete wavelet coefficients for speech recognition."In *Acoustics, Speech, and Signal Processing, 2000.ICASSP'00.Proceedings. 2000 IEEE International Conference on*, vol. 3, pp. 1351-1354. IEEE, 2000.

[32] Yuanyou, Xia, XieYanming, and Zhu Ruigeng. "An engineering geology evaluation method based on an artificial neural network and its application." *Engineering geology* 47, no. 1 (1997): 149-156.

[33]Daqrouq, Khaled, and Khalooq Y. Al Azzawi. "Average framing linear prediction coding with wavelet transform for text-independent speaker identification system." *Computers & Electrical Engineering* (2012).http://dx.doi.org/10.1016/j.compeleceng.2012.04.014, (Elsevier).

[34]Ganchev, Todor, Nikos Fakotakis, and George Kokkinakis. "Comparative evaluation of various MFCC implementations on the speaker verification task."In *Proceedings of the SPECOM*, vol. 1, pp. 191-194. 2005.

[35] Adami, André Gustavo, and Dante Augusto CoutoBarone. "A speaker identification system using a model of artificial neural networks for an elevator application."*Information Sciences* 138, no. 1 (2001): 1-5.

[36] Haydar, A., M. Demirekler, and M. K. Yurtseven. "Speaker identification through use of features selected using genetic algorithm." *Electronics Letters* 34, no. 1 (1998): 39-40.

[37] Huang, Xuedong, Alex Acero, Hsiao-Wuen Hon, and R. Reddy. "Spoken Language Processing: A Guide to Theory, Algorithm & System Development", Prentice-Hall, New Jersey (2001)

[38] Huang, Yi-Chia, Song-Lin Wueng, Chu-ChynOu, Chien-Hsiang Cheng, and Kuo-Hsiung Su. "Nutritional status of functionally dependent and nonfunctionally dependent elderly in Taiwan."*Journal of the American College of Nutrition* 20, no. 2 (2001): 135-142.

[39] Hermansky, Hynek. "Perceptual linear predictive (PLP) analysis of speech."*The Journal of the Acoustical Society of America* 87 (1990): 1738-1752.

[40] Rabiner, L., Juang, B.-H. Fundamentals of Speech Recognition.Prentice-Hall, Englewood Cliffs, New Jersey. (1993).

[41] Kwong, S., C. W. Chau, K. F. Man, and K. S. Tang. "Optimisation of HMM topology and its model parameters by genetic algorithms."*Pattern Recognition* 34, no. 2 (2001): 509-522.

[42] Avci, Engin. "A new optimum feature extraction and classification method for speaker recognition: GWPNN." *Expert Systems with Applications* 32, no. 2 (2007): 485-498.

[43] Uchida, Seiichi, Mohammad AsadRonee, and Hiroaki Sakoe. "Using eigen-deformations in handwritten character recognition."In *Pattern Recognition, 2002.Proceedings. 16th International Conference on*, vol. 1, pp. 572-575. IEEE, 2002.

[44] Daubechies, Ingrid. "Orthonormal bases of compactly supported wavelets." *Communications on pure and applied mathematics* 41, no. 7 (1988): 909-996.

[45] Lei, Zhou, Li Jiandong, Liu Jing, and Zhang Guanghui. "A novel wavelet packet division multiplexing based on maximum likelihood algorithm and optimum pilot symbol assisted modulation for Rayleigh fading channels." *Circuits, Systems and Signal Processing* 24, no. 3 (2005): 287-302.

[46] Behroozmand, Roozbeh, and FarshadAlmasganj. "Optimal selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patients' speech signal with unilateral vocal fold paralysis." *Computers in Biology and Medicine* 37, no. 4 (2007): 474-485.

[47] Sarikaya, Ruhi, and John HL Hansen. "High resolution speech feature parametrization for monophone-based stressed speech recognition." *Signal Processing Letters, IEEE* 7, no. 7 (2000): 182-185.

[48] Souani, Chokri, Mohamed Abid, KholdounTorki, and RachedTourki. "VLSI design of 1-D DWT architecture with parallel filters."*INTEGRATION, the VLSI journal* 29, no. 2 (2000): 181-207.

[49] Delac, Kresimir, MislavGrgic, and Sonja Grgic. "Face recognition in JPEG and JPEG2000 compressed domain." *Image and Vision Computing* 27, no. 8 (2009): 1108-1120.

[50] Atal, Bishnu S. "The history of linear prediction." *IEEE Signal Processing Magazine* 23, no. 2 (2006): 154-161.

[51] Chau, C. W., S. Kwong, C. K. Diu, and W. R. Fahrner. "Optimization of HMM by a genetic algorithm."In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 3, pp. 1727-1730. IEEE, 1997.

[52] Bennani, Younès, and Patrick Gallinari. "Neural networks for discrimination and modelization of speakers." *Speech Communication* 17, no. 1 (1995): 159-175.

[53] Kinnunen, Tomi, and Haizhou Li. "An overview of text-independent speaker recognition: From features to supervectors." *Speech communication* 52, no. 1 (2010): 12-40.

[54] Ghaffari, A., H. Golbayani, and M. Ghasemi. "A new mathematical based QRS detector using continuous wavelet transform." *Computers & Electrical Engineering* 34, no. 2 (2008): 81-91.

[55] Lu, Wei, Wei Sun, and Hongtao Lu. "Robust watermarking based on DWT and nonnegative matrix factorization." *Computers & Electrical Engineering* 35, no. 1 (2009): 183-188.