FEED FORWARD BACK PROPAGATION NEURAL NETWORK METHOD FOR ARABIC VOWEL RECOGNITION BASED ON WAVELET LINEAR PREDICTION CODING

Khalooq Y. Al Azzawi[1], Khaled Daqrouq[2]
[1] Electromechanical Engineering Dept., Univ. of Technology, Baghdad, Iraq.
[2] Communication and Electronics Engineering Dept., Philadelphia Univ., Amman, Jordan.

ABSTRACT

A novel vowel feature extraction method via hybrid wavelet and linear prediction coding (LPC) is presented here. The proposed Arabic vowels recognition system is composed of very promising techniques; wavelet transform (WT) with linear prediction coding (LPC) for feature extraction and feed forward backpropagation neural network (FFBPNN) for classification. Trying to enhance the recognition process and for comparison purposes, three techniques of WT were applied for the feature extraction stage: Wavelet packet transform (WPT) with LPC, discrete wavelet transform (DWT) with LPC, and WP with entropy (WPE). Moreover, different levels of WT were used in order to enhance the efficiency of the proposed method. Level 2 until level 7 were studied. A MATLAB program was utilised to build the model of the proposed work. The performance of 82.47% recognition rate was established. The mentioned above methods were investigated for comparison. The best recognition rate selection obtained was for DWT.

KEYWORDS: *Wavelet; Entropy; Neural Network; Arabic Vowels.*

I. INTRODUCTION

Unlike the English language, Arabic language recognition has the lowest share of attraction; this is due to its nature, in terms of, various dialects and several alphabets forms. But because of an increase of loudening activity in mobile communication domain draw new opportunities and shed some lights for applications of speech recognition including words and sentences in English as well as in Arabic. So, the Arabic text to speech and vice versa as well as incredibly critical issues in many applications that are attracted the users.

Numerous researchers have contributed in speech recognition, particularly in Arabic language recognition. The major work of studying speech recognition for Arabic language dealing with the morphological structure is presented in [1]. To recognize the distinct Arabic phonemes (pharyngeal, geminate and emphatic consonants) [2,3], the phonetic features is discussed. This allocates and motivates interesting researchers of Arabic language with different dialect at various countries. The applications in term of implementation of recognition system devoted to spoken separated words or continuous speech are not extensively conducted. [4] has studied the derivative scheme, named the concurrent general recursive neural network (GRNN), implemented for accurate Arabic phonemes recognition in order to automate the intensity and formants-based feature extraction. The validation tests expressed in terms of recognition rate obtained with free of noise speech signals were up to 93.37%. [5] has investigated an isolated word speech recognition by means of the recurrent neural network (RNN). The achieved accuracy was 94.5% in term of recognition rate in speaker-independent mode and 99.5% in speaker-dependent mode. [6] discussed a set of Arabic speech recognition systems also.

The Fuzzy C-Means method has been added to the traditional ANN/HMM speech recognizer using RASTA-PLP features vectors. The Word Error Rate (WER) is over 14.4%. With the same way, an approach using data fusion gave a WER of 0.8%. However, this method was tested only on one personal corpus and the authors showed that the obtained improvement needed the use of three neural networks running in parallel. Another alternative hybrid method was suggested [7], where the Support Vector Machine (SVM) and the K nearest neighbour (KNN) were substituted to the ANN in the traditional hybrid system, but the recognition rate, did not exceed 92.72% for KNN/HMM and 90.62% for SVM/HMM.

Saeed and Nammous [8] presented a novel Algorithm to recognize separate voices of some Arabic words, the digits from zero to ten. For feature extraction, transformation and hence recognition, the algorithm of minimal eigenvalues of Toeplitz matrices together with other methods of speech processing and recognition were used. The success rate obtained in the presented experiments was almost ideal and exceeded 98% for many cases. A hybrid method has been applied to Arabic digits recognition [9].

In literature papers, other researchers used neural networks to recognize features of Arabic language such as emphasis, gemination and related vowel lengthening. This was studied using ANN and other techniques [10], where many systems and configurations were considered including time delay neural networks (TDNNs). Again ANNs were used to identify the 10 Malay digits [11, 12] has anticipated a heuristic method of Arabic digit recognition, by means of the Probabilistic Neural Network (PNN). The use of a neural network recognizer, with a nonparametric activation function, presents a promising solution to increase the performances of speech recognition systems, particularly in the case of Arabic language. [13] demonstrated the advantages of the GRNN speech recognizer over the MLP and the HMM in calm environment.

Unfortunately, formants of Arabic vowels are not sufficiently tackled in the literature. Other studies that addressed formant frequencies in Arabic were not directed toward obtaining norms or comparing these frequencies to frequencies of vowels spoken by other populations. As an alternative, studies were directed toward speech perception, recognition, or speech analysis in Arabic [19,20,21,22]. These studies scheduled a range of formant frequency values. The presented research paper introduces a novel combination of wavelet transform, LPC and FFBPNN. The benefit of such sophistication conjunction is to create a dialect-independent Arabic vowels classifier. The remainder of the paper is organized as follows: a brief introduction to Arabic language is presented in section 2. The proposed method is described in section 3. The experimental results and discussion is introduced in section 4 followed in section 5 by conclusions.

II. ARABIC LANGUAGE

Recently, Arabic language became one of the most significant and broadly spoken languages in the world, with an expected number of 350 millions speakers distributed all over the world and mostly covering 22 Arabic countries. Arabic is Semitic language that characterizes by the existence of particular consonants like pharyngeal, glottal and emphatic consonants. Furthermore, it presents some phonetics and morpho-syntactic particularities. The morpho-syntactic structure built, around pattern roots (CVCVCV, CVCCVC, etc.) [22]. The Arabic alphabet consists of 28 letters that can be expanded to a set of 90 by additional shapes, marks, and vowels. The 28 letters represent the consonants and long vowels such as ى and أ (both pronounced as/a:/), ي (pronounced as/i:/), and و (pronounced as/u:/). The short vowels and certain other phonetic information such as consonant doubling (shadda) are not represented by letters directly, but by diacritics. A diacritic is a short stroke located above or below the consonant. Table 1 shows the complete set of Arabic diacritics. We split the Arabic diacritics into three sets: short vowels, doubled case endings, and syllabification marks. Short vowels are written as symbols either above or below the letter in text with diacritics, and dropped all together in text without diacritics. We get three short vowels: fatha: it represents the /a/ sound and is an oblique dash over a letter, damma: it represents the /u/ sound and has shape of a comma over a letter and kasra: it represents the /i/ sound and is an oblique dash under a letter as reported in Table 1.

Table 1. Diacritics above or below consonant letter

Short Vowel Name (Diacritics)	Diacritics above or below letter 'ب' (sounds B)	Pronunciation
Fatha	بَ	/ba/
Damma	بُ	/bu/
Kasra	بِ	/bi/
Tanween Alfath	بًا	/ban/
Tanween Aldam	بٌ	/bun/
Tanween Alkasr	بٍ	/bin/
Sokun	بْ	/b/

III. FEATURES EXTRACTION BY WAVELET TRANSFORM

Before the stage of features extraction, the speech data are processed by a silence removing algorithm followed by the application of a pre-processed by applying the normalization on speech signals to make the signals comparable regardless of differences in magnitude. In this study three feature extraction methods based on wavelet transform are discussed in the following part of the paper.

3.1 Wavelet Packet Method with LPC

For an orthogonal wavelet function, a library of wavelet packet bases is generated. Each of these bases offers a particular way of coding signals, preserving global energy and reconstructing exact features. The wavelet packet is used to extract additional features to guarantee higher recognition rate. In this study, WPT is applied at the stage of feature extraction, but these data are not proper for classifier due to a great amount of data length. Thus, we have to seek for a better representation for the vowel features. Previous studies proposed that the use of LPC of WP as features in recognition tasks is competent. [18] Suggested a method to calculate the LPC orders of wavelet transform for speaker recognition. This method may be utilized for Arabic vowel classification. This is possible because each Arabic vowel has distinct energy (Fig.2). Fig.4 shows LPC orders calculated for WP at depth 2 for three different utterances of Arabic a-vowel for the same person. We can notice that the feature vector extracted by WP and LPC is appropriate for vowel recognition.

3.2 Discrete Wavelet Transform Method with LPC

The additional proposed method is DWT combined with LPC. In this method the LPC is obtained from DWT Sub signals. The DWT at level three is generated and then 30 LPC orders are obtained for each sub signals to be combined in one feature vector. The main advantage of such sophisticated feature method is to extract different LPC impact based on multi resolution of DWT capability [14]. LPC orders sequence will contain distinguishable information as well as wavelet transform. Fig.4 shows LPC coefficients calculated for DWT at depth 3 for three different utterances of Arabic a-vowel for the same person. We may notice that the feature vector extracted by DWT and LPC is appropriate for vowel recognition.

3.3 Wavelet Packet Entropy Method

[15] Suggested a method to calculate the entropy value of the wavelet norm in digital modulation recognition. [16] Proposed features extraction method for speaker recognition based on a combination of three entropy types (sure, logarithmic energy and norm). Lastly, [17] investigated a speaker identification system using adaptive wavelet sure entropy.

As seen in above studies, the entropy of the specific sub-band signal may be employed as features for recognition tasks. This is possible because each Arabic vowel has distinct energy (see Fig.2). In this paper, the entropy obtained from the WPT will be employed for Arabic vowels recognition. The features extraction method can be explained as follows:

- Decomposing the speech signal by wavelet packet transform at level 7, with Daubechies type (db2).
- Calculating three entropy types for all 256 nodes at depth 7 for wavelet packet using the following equations:

Shannon entropy:

$$E1(s) = -\sum_i s_i^2 \log(s_i^2) \quad (1)$$

Log energy entropy:

$$E1(s) = \sum_i \log(s_i^2) \quad (2)$$

Sure entropy:

$$|s_i| \le p \Rightarrow E(s) = \sum_i \min(s_i^2, p^2) \quad (3)$$

where s is the signal, s_i are the WPT coefficients and p is a positive threshold. Entropy is a common concept in many fields, mainly in signal processing. Classical entropy-based criterion describes information-related properties for a precise representation of a given signal. Entropy is commonly used in image processing; it posses information about the concentration of the image. On the other hand, a method for measuring the entropy appears as a supreme tool for quantifying the ordering of non-stationary signals. Fig.3 shows Shannon entropy calculated for WP at depth 7 for Arabic a-vowel and Arabic e-vowels for two persons. For each person two different utterances were used, we can notice that the feature vector extracted by Shannon entropy is appropriate for vowel recognition. This conclusion has been obtained by interpretation the following criterion: the feature vector extracted should possess the following properties:
1) Vary widely from class to class.
2) Stable over a long period of time.
3) Should not have correlation with other features (see Fig.3 and 4).

3.4 Classification

Speech recognition with NN has recently undergone a significant development. Early experiments have exposed the potential of these methods for tasks with limited complexity. Many experiments have then been performed to test the ability of several NN models or approaches to the problem. Although most of these preliminary studies deal with a small number of signals, they have shown that NN models were serious candidates for speaker identification or speech recognition tasks. NN classifiers like FFBPNN may lead to very good performances because they allow to take into account speech features information and to build complex decision regions. However, the complexity of classification training procedures forbids the use of this simple approach when dealing with a large number of patterns. Two solutions do emerge for managing large databases: modular classification systems which a how to break the complexity of single NN architectures, or NN predictive models which tender a large variety of possible implementations.
Classification operation performs the intelligent discrimination by means of features obtained from feature extraction phase. In this study FFBPNN is used. The training condition and the structure of the NN used in this paper are as tabulated in Tab.2. These were selected empirically for the best performance selected for 10^{-5} of mse. 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 or transfer functions. 180x24 feature matrix which is obtained in features extraction stage for 24 vowel patterns (see flow chart at Fig.1) is given to the input of the Feed-forward networks consist of several layers using the DOTPROD weight function, NETSUM net input function, and the particular transfer functions. The weights of the first layer come from the input. Each network layer has a weight coming from the previous layer. All layers have biases. The last layer is the network output, which we call target (T). In this paper target is designed as a six binary digits for each features vector:

$$T = \begin{bmatrix} 0\ 0\ 0 \ldots 1 \\ 0\ 0\ 0 \ldots 0 \\ 0\ 0\ 0 \ldots 0 \\ 0\ 1\ \ 1 \ldots 1 \\ 1\ 0\ \ 1 \ldots 0 \end{bmatrix} \tag{4}$$

Table 2 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	128- Input, 30-Hidden & -4 Output
Weight Function	DOTPROD
Training Function	Levenberg-Marquardt Backpropagation
Activation functions	Log- sigmoid
Performance Function (mse)	10^{-5}
No. of Epochs	200

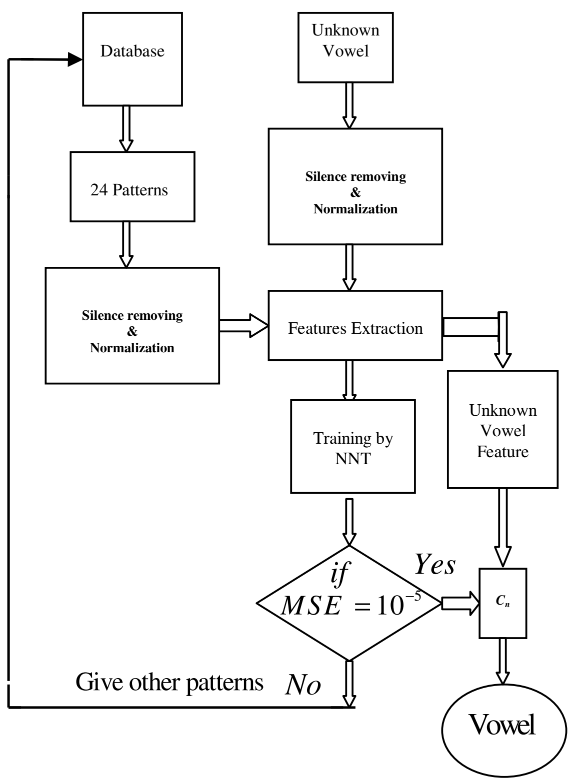

Fig. 1. Proposed expert system flow diagram of the proposed system

The mean square error of the NN is achieved at the final of the training of the ANN classifier by means of Levenberg-Marquardt Backpropagation. Backpropagation is used to compute the Jacobian jX of performance with respect to the weight and bias variables X. Each variable is adapted according to Levenberg-Marquardt,

$$jj = jX * jX$$
$$je = jX * E \quad (5)$$
$$dX = -(jj + I * Mu) \backslash je$$

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 above outcomes in a reduced performance value. The change is then made to the network and Mu is decreased by 0.1 Mu decrease factor. After training the 24 (12 male and 12 female) speakers a feature, imposter simulation is performed. The unknown vowel simulation result (SR) is compared with each of the 24 patterns target (P_n, n =1,2,…,24) in order to determine the decision by

$$C_n = 100 - [100 * \sqrt{(\sum (P_n - SR)^2 / \sum P_n^2)}] \quad (6)$$

where C_n is the similarity percent between unknown vowel simulation results and pattern target P_n. The vowel is identified as patterns of maximum similarity percent. For instant, when most higher magnitudes of C_n belong to given type patterns then decision is this type.

IV. RESULTS AND DISCUSSION

In this research paper, speech signals were recorded via PC-sound card, with a sampling frequency of 16000 Hz. The Arabic vowels were recorded by 27 speakers of different Arabic dialects (Jordanian, Palestinian and Egyptian: 5 females, along with 22 males. The recording process was provided in normal university office conditions. Our investigation of speaker-independent Arabic vowels classifier system performance is performed via several experiments depending on vowel type. In the following three experiments the used feature extraction method is WP and LPC.

Experiment-1

We experimented 95 long Arabic vowels أ (pronounced as/a:/) signals, 354 long Arabic vowels ي (pronounced as/e:/) signals and 88 long Arabic vowels و (pronounced as/u:/) signals. The results indicated that 84.44% were classified correctly for Arabic vowels أ, 71.47% of the signals were classified correctly for Arabic vowel ي, and 72.72% of the signals were classified correctly for Arabic vowel و. Tab.3 shows the results of Recognition Rates.

Experiment-2

We experimented 90 short Arabic vowels أَ (fatha) (pronounced as/a:/) signals, 45 short Arabic vowels يِ (kasra) (pronounced as/e:/) signals and 45 long Arabic vowels وُ (damma) (pronounced as/u:/) signals. The results indicated that 100% were classified correctly for short Arabic vowels أ, 84.44% of the signals were classified correctly for short Arabic vowel ي, and 91.11% of the signals were classified correctly for short Arabic vowel و. Tab.4 shows the results of Recognition Rates.

Experiment-3

In this experiment we study the recognition rates for long vowels connected with other letter such ل (pronounced as/l/) and ر (pronounced as/r/). Tab. 5, reported the recognition rates. The results indicated 82.89% average recognition rate.

Experiment-4

In experiment-4, short Arabic vowels: fatha: represents the short أ (pronounced as short /a/), kasra: represents the short ي (pronounced as short /e:/) and damma represents short و (pronoused as short /u/) for each vowel a number of signals of 20 speakers results are reported in tab. 6 . The recognition rates of above mentioned three short vowels connected with other letter such ل (pronounced as/l/) and ر (pronounced as/r/) are studied and their results are tabulated in table 6. The average recognition rate was 88.96%.

Table 3: The recognition rate results for long vowels

Long Vowels	Number of Signals	Recognized Signals	Not Recognized Signals	Recognition Rate [%]
Long A أ	90	76	14	84.44
Long E ي	354	253	101	71.47
Long O و	88	64	24	72.72
			Avr. Recognition Rate	**76.21**

Table 4: The recognition rate results for short vowels

Short Vowels	Number of Signals	Recognized Signals	Not Recognized Signals	Recognition Rate [%]
Short A أ	95	95	0	100
Short E ي	45	38	7	84.44
Short O و	45	41	4	91.11
			Avr. Recognition Rate	**91.85**

Table 5: The recognition rate results for long vowels connected with other letters

Long Vowels	Number of Signals	Recognized Signals	Not Recognized Signals	Recognition Rate [%]
La لا	54	46	8	85.19
Le لي	54	52	2	96.30
Lo لو	54	32	22	59.26
Ra را	48	44	4	91.67
Re ري	46	40	6	89.96
Ro رو	48	36	12	75.00
			Avr. Recognition Rate	**82.89**

Table 6: The recognition rate results for short vowels connected with other letters

Short Vowels	Number of Signals	Recognized Signals	Not Recognized Signals	Recognition Rate [%]
La لا	54	50	4	92.59
Le لي	54	50	4	92.59
Lo لو	54	48	6	88.89
Ra را	46	38	8	82.61
Re ري	48	44	4	91.67
Ro رو	48	41	9	85.42
			Avr. Recognition Rate	**88.96**

In the next experiment, the performances of the three WT Arabic vowels recognition systems (proposed in section 3) are compared with each other under the recorded database. The results of these experiments are summarized in Tab. 7. The best results were achieved by DWT and LPC.

Table 7: The recognition rate results for the three proposed systems

Recognition method	Number of Signals	Recognition Rate [%]
WP	1356	80.23
DWT	1356	82.47
WPE	1356	72.9

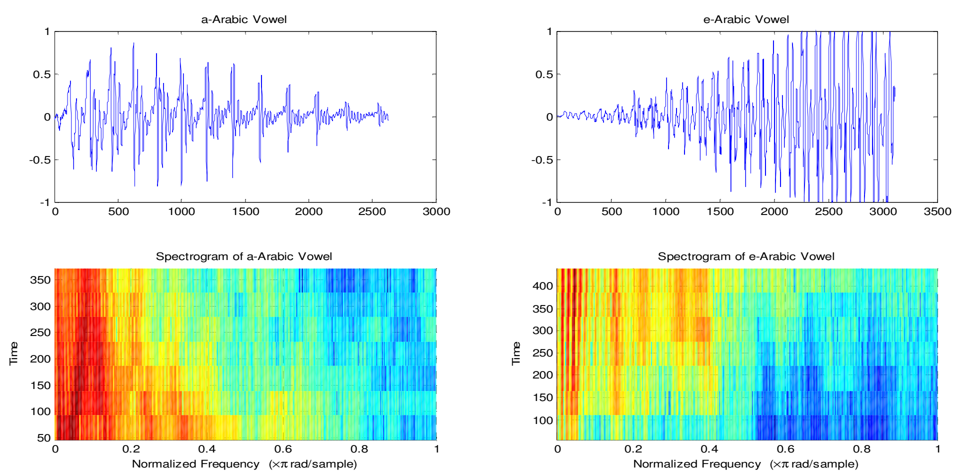

Figure 2.a. First Arabic Vowels of a speaker 1 with spectrogram

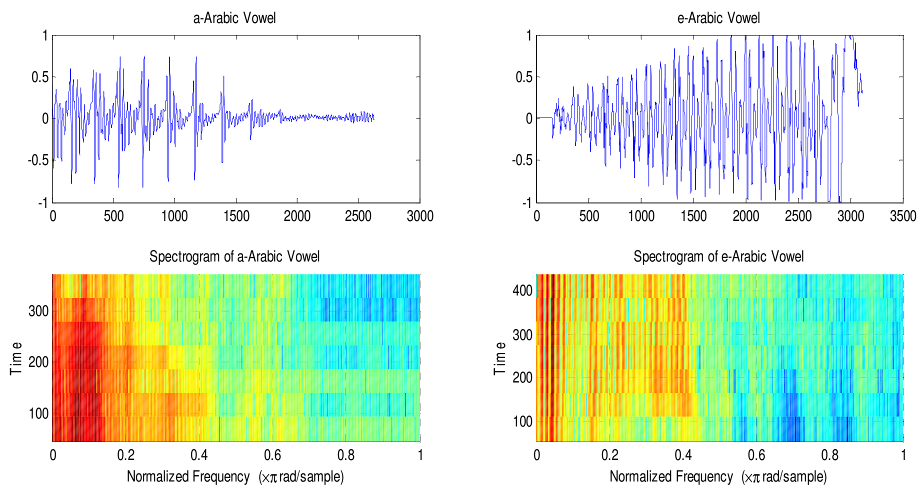

Figure 2. First Arabic Vowels of a speaker 2 with spectrogram

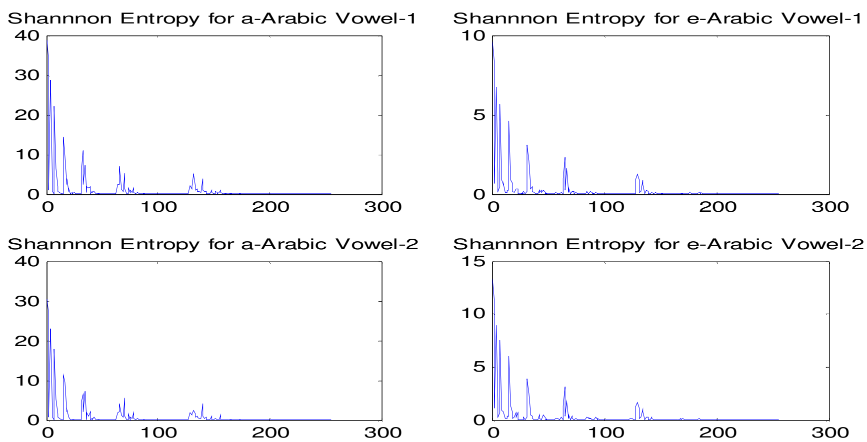

Figure 3. Shannon entropy for Arabic vowels presented in Figure 2

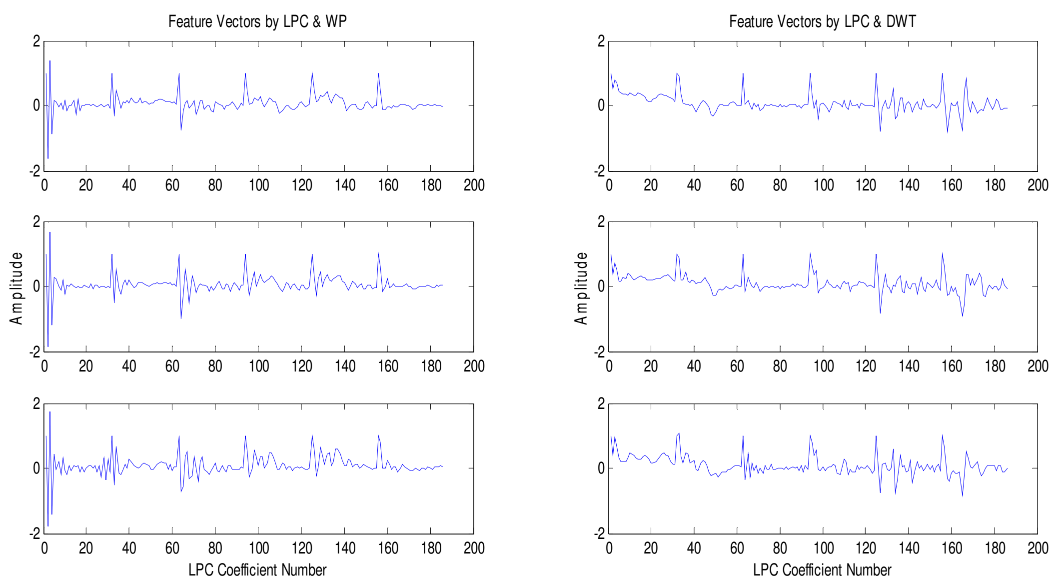

Figure 4. WP and DWT with LPC for three utterances of Arabic a-vowel for the same speaker.

V. CONCLUSION

Feed forward backpropagation neural network based speech recognition system is proposed in this paper. This system was developed using a wavelet feature extraction method. In this work, effective feature extraction method for Arabic vowels system is developed, taking in consideration that the computational complexity is very crucial issue. Trying to enhance the recognition process, three techniques of WT were applied for the feature extraction stage: WP with LPC, DWT with LPC, and WPE. The experimental results on a subset of recorded database showed that feature extraction method proposed in this paper is appropriate for Arabic recognition system. Our investigation of dialect-independent Arabic vowels classifier system performance is performed via several experiments depending on vowel type. The declared results showed that the proposed method can make an effectual analysis with identification rates may reach 100% in some cases.

REFERENCES

[1]Datta, S., Al Zabibi, M., Farook, O., 2005. Exploitation of morphological in large vocabulary Arabic speech recognition. International Journal of Computer Processing of Oriental Language 18 (4), 291–302.

[2]Selouani, S.A., Caelen, J., (1999). Recognition of Arabic phonetic features using neural networks and knowledge-based system: a comparative study. International Journal of Artificial Intelligence Tools (IJAIT) 8 (1), 73–103.

[3]Debyeche, M., Haton, J.P., Houacine, A., (2006). A new vector quantization approach for discrete HMM speech recognition system. International Scientific Journal of Computing 5 (1), 72–78.

[4]Shoaib, M., Awais, M., Masud, S., Shamail, S., Akhbar, J., (2004). Application of concurrent generalized regression neural networks for Arabic speech recognition. Proceedings of the IASTED International Conference on Neural Networks and Computational Intelligence, (NCI 2004), 206–210.

[5]Alotaibi Y.A., Investigating spoken Arabic digits in speech recognition setting, Information Sciences 173 (2005) 115–139.

[6]Amrouche, A., et al., An efficient speech recognition system in adverse conditions using the nonparametric regression. Engineering Applications of Artificial Intelligence (2009).

[7]Bourouba, H., Djemili, R., Bedda, M., Snani, C., (2006). New hybrid system (supervised classifier/HMM) for isolated Arabic speech recognition. Proceed- ings of the Second IEEE International Conference on Information and Communication Technologies (ICTTA'06), 1264–1269.

[8]Saeed, K., Nammous, M., (2005), A New Step in Arabic Speech Identification: Spoken Digit Recognition b.

[9]Lazli, L., Sellami, M., 2003. Connectionist probability estimation in HMM Arabic speech recognition using fuzzy logic. Lectures Notes in LNCS 2734, 379–388.

[10]Selouani S.A., Douglas O., Hybrid architectures for complex phonetic features classification: a unified approach, in: International Symposium on Signal Processing and its Applications (ASSPA), Kuala Lumpur, Malaysia, August 2001, pp. 719–722.

[11]Salam M. , Mohamad D. , Salleh S., Neural Network speaker dependent isolated malay speech recognition system: handcrafted vs. genetic algorithm, in: International Symposium on Signal Processing and its Application (ISSPA), Kuala Lumpur, Malaysia, August (2001), pp. 731–734.

[12]Saeed, K., Nammous, M., 2005. Heuristic method of Arabic speech recognition. Proceedings of the IEEE International Conference on Digital Signal Processing and its Applications (IEEE DSPA'05), 528–530a.

[13]Amrouche, A., Rouvaen, J.M., 2003. Arabic isolated word recognition using general regression neural network. Proceedings of the 46th IEEE MWSCAS, 689–692.

[14] Wu, J.-D. & Lin B.-F. (2009), Speaker identification using discrete wavelet packet transform technique with irregular decomposition Expert Systems with Applications 363136–3143.

[15]Avci, E., Hanbay, D., & Varol, A. (2006). An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition. Expert System with Applications, 33, 582–589.

[16]Avci, E. (2007), A new optimum feature extraction and classification method for speaker recognition: GWPNN, Expert Systems with Applications 32, 485–498.

[17]Avci, D. (2009), An expert system for speaker identification using adaptive wavelet sure entropy, Expert Systems with Applications, 36, 6295–6300.

[18] Daqrouq, K. Al-Qawasmi, A.-R. Al-Sawalmeh, W. Hilal, T.A., Wavelet Transform based multistage speaker feature tracking identification system using Linear Prediction Coefficient, ACTEA-IEEE Explorer, 2009.

[19]Anani M. (1999), Arabic vowel formant frequencies. In: Proceedings of the 14th International Congress of Phonetic Sciences, Vol.9. San Francisco, CA;2117–2119.

[20]Cherif A, Bouafif L, Dabbabi T. Pitch detection and formant analysis of Arabic speech processing. Applied Acoustics. 2001;62:1129–1140.

[21]Alghamdi M. A spectrographic analysis of Arabic vowels: a cross-dialect study. J King Saud Univ. 1998;10:3–24.

[22]Alotaibi Y, Hussain A. Speech recognition system and formant based analysis of spoken Arabic vowels. In: Proceedings of the First International.

Authors Biographies:

Khaled Daqrouq received the B.S. and M.S. degrees in biomedical engineering from the Wroclaw University of Technology in Poland, in 1995, as one certificate, and the Ph.D. degree in electronics engineering from the Wroclaw University of Technology, in Poland, in 2001. He is currently associate professor at the Philadelphia University, in Jordan. His research interests are ECG signal processing, wavelet transform applications in speech recognition and the general area of speech and audio signal processing and improving auditory prostheses in noisy environments.

Khalooq Y. Al Azzawi received the B.SC. in Electrical Engineering from University of Mosul in 1970 , A Post Graduate Diploma in Communication System from Manchester University of Technology in England in 1976, and M.Sc. degrees in Communication Engineering & Electronics from Loughborough University of Technology in England in 1977.,. He is currently associate professor at the Philadelphia University, in Jordan working in a Sabbatical year . He is an Ass. Prof. in Comm. Eng. & Electronics in Baghdad University of Technology. His research interests are FDNR Networks in filters. wavelet transform applications in speech recognition.