Information Sciences and Computer Engineering, Vol. 1, No. 1, (2010) 16–25

PHARUS

International Journal of
Information Sciences and Computer Engineering

journal homepage: http://www.ijisce.org

# Spoken Arabic Digits Classifier via Sophisticated Wavelet Transform Features Extraction Method[☆]

Khaled Daqrouq[*], Mikhled Alfaouri

*Department of Communication and Electronics Engineering, Philadelphia University, State province, Jordan*

***Abstract***– **The essential problem of Arabic recognition systems is the several of Arabic language dialects, especially along with associated noise. Therefore, low recognition rate is encountered, as a result of such an environment. In this research paper, the authors presented dialect-independent via sophisticated wavelet transform-based Arabic digits classifiers (SWADC). The proposed classifier is divided into three main blocks: 1) Filtration and widowing. 2) Sophisticated Features Extraction Method by combining Continuous Wavelet Transform (CWT) with Linear Prediction Coefficient (LPC) and Mel Frequency Cepstral Coefficient (MFCC). 3) Classification by Root Mean Square Difference Similarity Measure (RDSM) and Feed Forward Back Propagation Neural Network Classification (FFBPNC). The proposed classifier provided a high Recognition Rate reaches up to 100%, in some cases, and an average cases up to 95.9%, for about 450 tested individual digits, based on speaker-independent system.**

**Keyword:** Arabic digits; Wavelet transform; Speech recognition; Cepstral coefficients; and Neural network.

## 1. Introduction

At this moment, interactive voice response systems are increasingly and widely used, especially involving speaker-independent recognition of a given vocabularies conveyed over the telephone network or microphone [1]. The recent rising increase of crescendo activity in mobile communication domain allure new opportunities and shed some lights for applications of speech recognition including digits and sentences. Text to speech, or vice versa, as well as incredibly vital issues in many computer applications, where English language has achieved immense success of the major part of interest. On the other hand, Arabic language speech recognition has slight attraction; due to its many nature difficulties, in term of, various dialects and several alphabets forms.

The major work of investigation of speech recognition of Arabic language dealing with the morphological structure found in [2, 3], or the phonetic features in order to recognize the distinct Arabic phonemes (pharyngeal, geminate and emphatic consonants), and discussed their further implication in a larger vocabulary speech system [4, 5]. This allocate and motivate interesting researchers of Arabic language with different dialect at different countries, the applications in term of implementation of recognition system devoted to spoken isolated words or continuous speech are not extensively conducted, and only few examples have been showed, improved and ameliorated in this research paper to establish a new path of investigation of Arabic spoken with different dialect. Shoaib in [6] has studied the derivative scheme, named the Concurrent GRNN, implemented for accurate Arabic phonemes identification 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%. Alotaibi in [7] has investigated an isolated word speech recognition using the RNN. The achieved accuracy was 94.5% in term of recognition rate in speaker-independent mode and 99.5% in speaker-dependent mode. Amrouche in [8], discussed a lot 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 approach, a method using data fusion gave a WER of 0.8%. However, this method was tested only on one personal corpus and the authors indicated that the obtained improvement needed the use of three neural networks working in parallel. Another alternative hybrid method was proposed by [9], where the Support Vector Machine (SVM) and the K nearest neighbor (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.

[10] presented a new 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 by [11].

From literatures papers, other researchers, neural networks were used to identify features of Arabic language such as emphasis, gemination and relevant vowel lengthening [7]. This was studied using ANN and other techniques [12], where many sys-

[☆]This work was supported by Philadelphia University

[*]Corresponding author:
Email address: haleddaq@yahoo.com, Ph: +96 265352888

tems and configurations were considered including time delay neural networks (TDNNs). Again ANNs were used to recognize the 10 Malay digits [13] Saeed and Nammous in [14] has proposed a heuristic method of Arabic digit recognition, using the Probabilistic Neural Network (PNN). The use of a neural network recognizer, with a nonparametric activation function, constitutes a promising solution to increase the performances of speech recognition systems, particularly in the case of Arabic language. [15, 16] demonstrated the advantages of the GRNN speech recognizer over the MLP and the HMM in quiet environment.

Also, the lake of an extremely very noisy environments investegation, the recognition performance degraded considerably. Robustness to noise is then essential for professional using recognition systems particularly in mobile networks context [17]. Various studies have been conducted in this track [18, 19]. Numerous pre-processing techniques have been developed in order to reduce or eliminate the noise effects in the speech before adding to a recognizer. Enhancement procedures like spectral subtraction remove ambient noise [20, 21]. The transmission effects are reduced using equalization techniques such as cepstral normalization and adaptive filtering [21, 22, 23].

This paper presents a new combination of wavelet Transform and most popular feature extraction methods: MFCC and LPC. The main objective of such sophistication conjunction is to create a dialect-independent and robust to noise spoken Arabic digits classifier. The remainder of the paper is organized as follows: In Section 2 we present the brief introduction to Arabic language and Arabic digits. The proposed method is described in Section 3. The experimental results and discussion in Section 4 followed in Section 5 by conclusions.

## 2. Arabic Language

Recently, Arabic language, became one of the most important and broadly spoken languages in the world, with an expected number of 350 millions speakers distributed all over the world and mainly 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) [24].

The Arabic alphabet consists of 28 letters that can be extended to a set of 90 by additional shapes, marks, and vowels [25]. 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, but by diacritics. A diacritic is a short stroke placed 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 find 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 explained in Table 1 [24].

**Table 1.** Diacritics above or below consonant letter.

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

Consequently, it is important to realize that, what we typically refer to as "Arabic" is not single linguistic variety; rather, it is a collection of different dialects and socialists. Classical Arabic is an older, literary form of the language, exemplified by the type of Arabic used in the Quran. Modern Standard Arabic (MSA) is a version of Classical Arabic with a modernized vocabulary. MSA is a formal standard common to all Arabic-speaking countries. It is the language used in the media (newspapers, radio, TV), in official speeches, in courtrooms, and, generally speaking, in any kind of formal communication. However, it is *not* used for everyday, informal communication, which is typically carried out in one of the local dialects.

The dialects of Arabic can roughly be divided into two groups: Western Arabic, which includes the dialects spoken in Morocco, Algeria, Tunisia, and Libya, and Eastern Arabic, which can be further subdivided into Egyptian, Levantine, and Gulf Arabic. These various dialects differ considerably from each other and from Modern Standard Arabic. Differences affect all levels of language, i.e. pronunciation, phonology, vocabulary, morphology, and syntax. Table 1 lists examples of the differences between Egyptianl Arabic Dialect (EAD) and Modern Standard Arabic. EAD is that dialect which is most widely understood through-out the Arabic-speaking world, due to a large number of TV programmes which are produced in Egypt and exported to other Arabic countries. Native speakers from different dialect regions are for the most part capable of communicating with each other, especially if they have had some previous exposure to the other speaker's dialect. However, widely differing dialects, such as Moroccan Arabic and the Iraqi dialect, may hinder communication to the extent that speakers adopt Modern Standard Arabic as a lingua franca.

Many issues of Arabic language, such as the phonology and the syntax, do not present difficulty for automatic speech recognition. Standard, language-independent techniques for acoustic and pronunciation modeling, such as context-dependent phones, can easily be applied to model of the acoustic-phonetic properties of Arabic. The most difficult problems in developing high-accuracy speech recognition systems to Arabic language are the predominance of non-diacritized text material, the enormous dialectal variety, and the morphological complexity.

The principally problem of the dialectal variety, is due to a current lack of training data for conversational Arabic; while, MSA data can readily be acquired from various media sources.

Finally, morphological complexity is approved to present solemn problems for speech recognition. A high scale of affixa-

**Table 2.** Three examples of four different Arabic dialects.

| Gloss | MSA | EAD | JAD | PAD |
|---|---|---|---|---|
| **'Three'** ثلاث | **th-l-thh** | **t-l-th** | **th-l-theh** | **t-l-teh** |
| **'Eight'** ثمانية | **th-m-nê-yah** | **t-m-n-yah** | **th-m-n-yeh** | **t-m-n-yeh** |
| **'Two'** أثنين | **?ith-nn** | **te-nn** | **?ith-nen** | **?it-nn** |

tion, derivation etc. contributes to the explosion of unlike word forms, making if difficult if not impossible to robustly estimate language model probabilities. Prosperous morphology also leads to elevated out-of-vocabulary rates and bigger search spaces during decoding, thus slowing down the recognition process [3].

*Arabic Digits*

Arabic digits are from zero to nine are polysyllabic words except the first one, zero, which is a monosyllable word [2]. Table 2 shows the 10 Arabic digits along with pronunciation, signals and number of syllable [7]

Compared to other languages, Arabic digits are much more elongated. They include two to four syllables, while French, English and Mandarin digits are single or double syllables. Arabic digits can be considered as representative elements of language, because more than half of the phonemes of the Arabic language are included in the 10 digits. The fricative and plosive consonants are more dominants and characterized by the presence of noise in the high-frequency band spectrum. In fact, these consonants are easily corrupted by noise sources making. Therefore, speech recognition systems usually fall to identify them in adverse conditions [24].

*Similarities between spoken arabic digits*

The similarity between Arabic digits, in term of pronunciation and signal morphology, may leads to big recognition confusion rate [7]. In this research paper, the authors, presented some of these Arabic digits similarities to cover the lake of investigations of similarities of Arabic digits:

- When digit 0 is investigated against digit 1, we can observe that the second phonemes in both digits 0 and 1 are vowels /i/ and /a:/, respectively, and they have high similarity depending on their spectrograms. Power spectral Density (PSD) of the two digits contains some common maximum peaks (see Fig.1). An overlap between these phonemes may occur, hence causing a misleading match between these digits.
- The similarities between digits 0 and 2 are very little and this result is evident when spectrograms and PSD are studied. This is also confirmed by the results of digit recognition system, except noise contaminated digits.
- By investigating digits 1 and 2, we can encounter that there is a large dissimilarity between these two digits especially in the second partially of their spectrograms. Digit 2 has a long vowel in the second syllable and the same syllable starts with the nasal phoneme /n/ and ends with the same nasal phoneme.
- Spectrograms of digits 1 and 3 contain big similarities. PSD of the two digits have common two core peaks at 40 and 10 in the frequency scale (see Fig.1). The digit recognition systems, always produces immense confusions [7, 14]
- Digit 1 and 4 have the same penultimate phoneme, a short vowel /a/. There are moderate common peaks in PSD curves. On the other hand, there were small spectrogram similarities.
- There is little similarity between digits 2 and 3 in number and type of syllable.
- Digits 3 and 8 have high pronunciation similarities; the sounds /h/ and /a/ are the first and second phonemes in both digits (i.e., the first syllable in both digits are exactly the same).
- There is a similarity between digits 4 and 5 in the last two phonemes. Phonemes /a/ and /h/ are the final two phonemes in both digits. Also the second phonemes in each are also the same
- There is a large pronunciation dissimilarity between digits 4 and 6. Digit 6 consists mostly of unvoiced. Consonants namely /s/ and /t/ (twice), while digit 4 consists mostly of voiced phonemes namely vowels and /r/, /b/, and /?/ consonants. Low similarity between these digits in term of PSD and spectrogram. There are no recognition system confusions.
- Digits 4 and 7 have a high similarity in term of pronunciation but are different in term of PSD and spectrogram.
- Digits 6 and 7 have the identical pattern of syllables, CVC-CVC, and they have the same first phoneme, /s/, but are different in term of PSD and spectrogram.

## 3. Proposed Methods

In any Arabic recognition system, Various Arabic language dialects and association noise are the most essential problems. Thus, low recognition rate is accomplished, as a result of such limitations [26, 27]. In this paper, we present a sophisticated wavelet transform-based Arabic digits classifiers. The proposed classifier is divided into three main blocks as shown in Fig. 3:

• Filtration and widowing:

After the original signal acquisition, Wavelet Filtration (WF) is performed. For wavelet filters, we start with the scaling function $\phi$. If $w_n$ is the coefficient of the linear combination in Eq. 1, $w_n$ can be generated by the integration in Eq. 2 [28, 29].

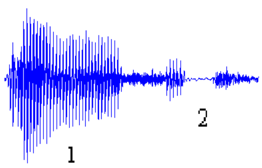
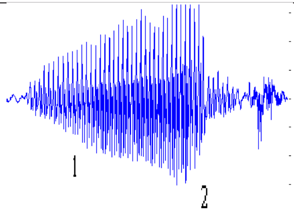
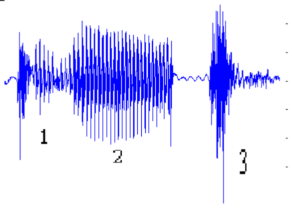
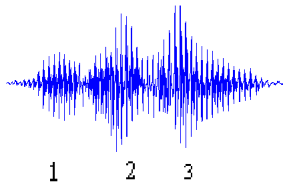
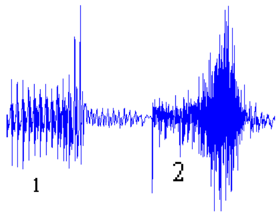
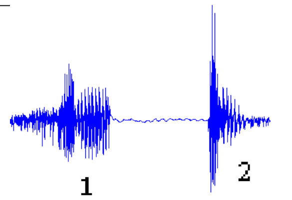
$$\phi\left(\frac{x}{2}\right) = 2^{1/2} \sum_n w_n \phi(x-n) \tag{1}$$

$$w_n = \frac{1}{2^{1/2}} \int \phi(\frac{x}{2})\phi(x-n)dx \tag{2}$$

Clearly if $\phi$ is compactly supported, the sequence $w_n$ is finite and may be viewed as a filter. The filter $w_n$ (scaling filter) is a low-pass Finite Impulse Response (FIR) filter, of length $2N$. A low digital filters output $y(k)$ is interrelated to its input $s(k)$ by convolution with its impulse response$w(k)$.

$$y(k) = w(k) * s(k) = \sum_{\tau=-\infty}^{\infty} w(k-\tau)s(\tau) \tag{3}$$

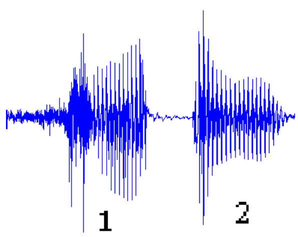
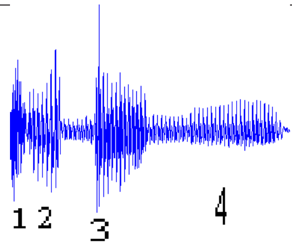
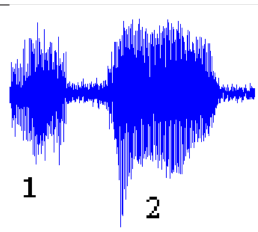
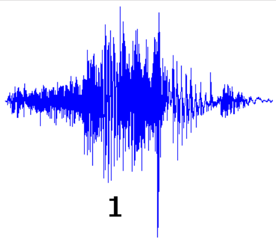
**Table 3.** Arabic spoken digits: writing, pronunciation, signal and number of syllables.

| Digit | Digit Writing | Pronunciation | Speech Signal | No. of Syllables |
|---|---|---|---|---|
| 1 | واحد | w-hid |  | 2 |
| 2 | أثنين | ?ith-nn |  | 2 |
| 3 | ثلاثة | th-l-thh |  | 3 |
| 4 | أربعة | ʼaâr-bâ-?âh |  | 3 |
| 5 | خمسة | khm-sâh |  | 2 |
| 6 | ستة | set-th |  | 2 |

Reverse Biorthogonal Wavelet is proposed, because of its excellent smoothing capability [16]. This is equivalent to decomposing the signal into discrete wavelet transform approximation coefficients. The proposed low pass filter makes the system more robust to noise. At this time, the output signal $y(k)$ is windowed for 20ms segments.

• **Sophisticated Extraction:**

At this stage, we investigate the utilization of three types of features extraction methods. This sophisticated approach is pro-

| | | | | |
|---|---|---|---|---|
| **7** | سبعة | **sûb-?âh** |  | **2** |
| **8** | ثمانية | **th-m-nê-yah** |  | **4** |
| **9** | تسعة | **tes-âh** |  | **2** |
| **0** | صفر | **sefr** |  | **1** |

posed to face the dialect-independent and speaker-independent difficulties. Types are distinguished as:

1. Continuous Wavelet Transform (CWT) and Standard Deviation Features Extraction Method; this method is implemented by calculating CWT of each window of $y(k)$ signal over the three levels: low, medium and high. And then, standard deviation is obtained:

$$\sigma_{CWT(2^j)} = \sqrt{E\left[(W_y(n))^2\right]} \tag{4}$$

for $j = 5, 10$ and $15$ Where $W_y = CWT_y(n)$, and $n = 1, 2, \ldots, N$ is the window number. The level determination as 5, 10 and 15 is according to the sampling frequency of the speech signal [25, 30]. These levels present low, medium and high pass bands of the signal frequency. Thus, the utilizing of this method helps extracting the several signal features via three frequency bands, instead of whole signal overlapped bands. Then, three vectors of standard deviations are accomplished.

2. Linear Prediction Coefficient (LPC); LPC of number of coefficients equal to 20 for each window is applied. This number is determined empirically. The reason of utilizing this method is to exploit the distinct LPC features for such difficult nature in term of dialect-independent and speaker-independent.

3. CWT and Mel Frequency Cepstrum Features Extraction Method: this method is implemented by calculating CWT of each window of $y(k)$ signal over three levels: low, medium and high. And then, Mel Frequency Cepstrum Coefficient is obtained, three matrices are obtained, where each column in the matrices contains 13 coefficients calculated for each window:

$$MFCC_{CWT(2^j)} = MFCC(W_y(n)) \tag{5}$$

Where,

$$MFCC\ (W_Y(n)) = \begin{bmatrix} C1 & C1 & \ldots & C1_N \\ C2 & C2 & \ldots & C2_N \\ & & & \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ & & & \\ C13 & C13 & \ldots & C13_N \end{bmatrix} \tag{6}$$

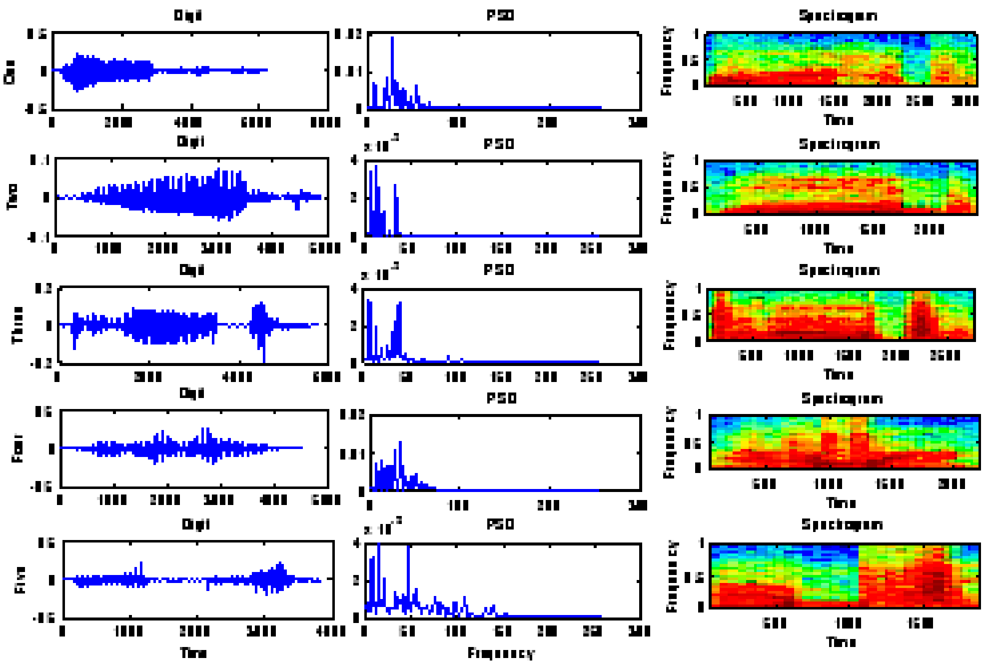


**Fig. 1.** Spectrogram PSD of Spoken Arabic digits (from one to five).

• **Classification:**

The third stage of our classifier system is the classification stage where the digit features are distinguished based on calculating maximum similarity between input digit features and the model digits features, stored in the system memory. Verification between seven input digit features (three of standard deviations vectors, one LPC coefficients vector and three of MFCC matrices) and the same model digits features is accomplished by Root Mean Square Difference Similarity Measure (RDSM):

$$RDSM_i = 100 - \sqrt{\left[\sum_{i=1}^{7} \frac{(\vec{Fin}_i - \vec{Fm}_i)^2}{\vec{Fin}_i^2}\right]} \times 100 \qquad (7)$$

Where $i = 1, 2, \dots 7$, $\vec{Fin}_i$ is input digit features vector and $\vec{Fm}_i$ same model digit features vector. This measure is calculated for each features vector and matrix of those seven features vectors and matrices, for input digit at each time, one of stored models. So that, seven magnitudes are achieved for an individual model.

$$RDSM_i = [RDSM_1 RDSM_2 \dots, RDSM_7] \qquad (8)$$

Then mean value (Expectation) of $RDSM_i$ is determined for an individual model

$$E[RDSM_i] = (\sum_{i=1}^{7} RDSM_i)/7 \qquad (9)$$

The classifier decision is taken by determining the maximum of all $E[RDSM_i]$ calculated for all models stored in the system

$$\max_{1 \leqslant m \leqslant M}^{E} [RDSM_i]_m \qquad \text{for } m = 1, 2, \dots, M \qquad (10)$$

Where $M$ is referred to the number of stored digits models.

**Alternative Classification Method:**

The Feed Forward Back Propagation Neural Network Classifier (FFBPNC) for each features vector of those explained above, is also investigated. Input digit Features vectors and matrices are compared with one of stored models features vectors at each time. So that, the first three input digit features vectors are the three column vectors matrix input in FFBPNC.

$$P = \begin{bmatrix} \sigma_{CWT(2^5)}(1)\ \sigma_{CWT(2^{10})}(1)\ \sigma_{CWT(2^{15})}(1) \\ \sigma_{CWT(2^5)}(1)\ \sigma_{CWT(2^{10})}(1)\ \sigma_{CWT(2^{15})}(1) \\ \dots \\ \dots \\ \dots \\ \sigma_{CWT(2^5)}(N)\ \sigma_{CWT(2^{10})}(N)\ \sigma_{CWT(2^{15})}(N) \end{bmatrix} \qquad (11)$$

With the following Target

$$T = \begin{bmatrix} 1\,1\,1 \\ 1\,1\,0 \\ 1\,0\,1 \\ 0\,1\,1 \\ 1\,1\,1 \end{bmatrix} \qquad (12)$$

Then the same model features are simulated with the output network parameters (weights and biases). This gives a Simulated Target ($ST$) for each model digit, which in the ideal case (input digit equal to the model digit) should be equal to the Target. To accomplish the recognition rate assessment measure we use

$$RecRate = \left(\frac{B - (nnz(T - ST)}{B}\right) \qquad (13)$$

Where $B$ is the number of elements in $T$ and $nnz$ is the nun zeros number, which represent the error between $T$ and $ST$.

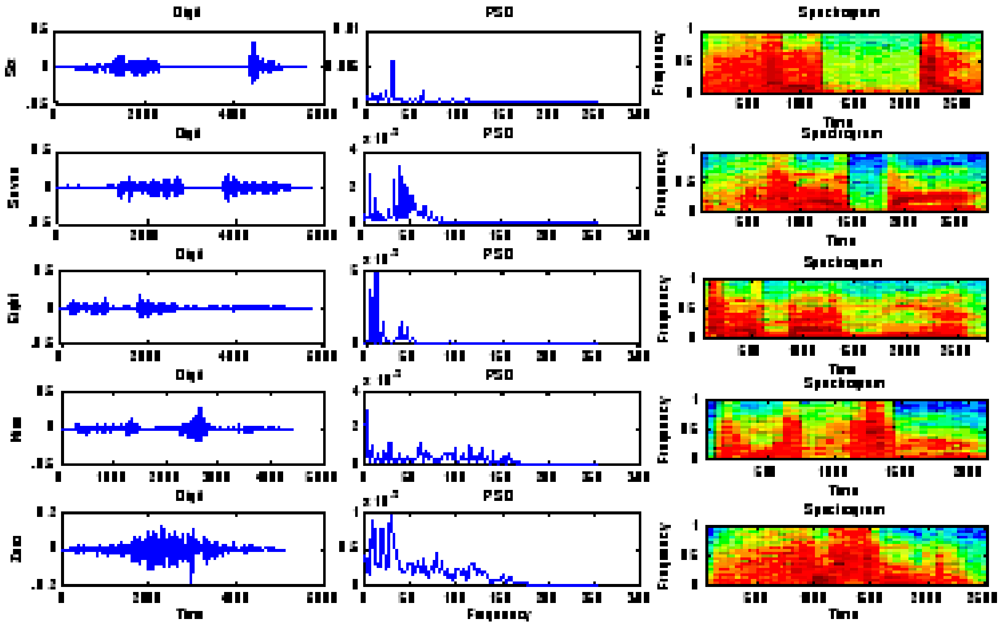


**Fig. 2.** Spectrogram PSD of Spoken Arabic digits (from six to nine and zero).

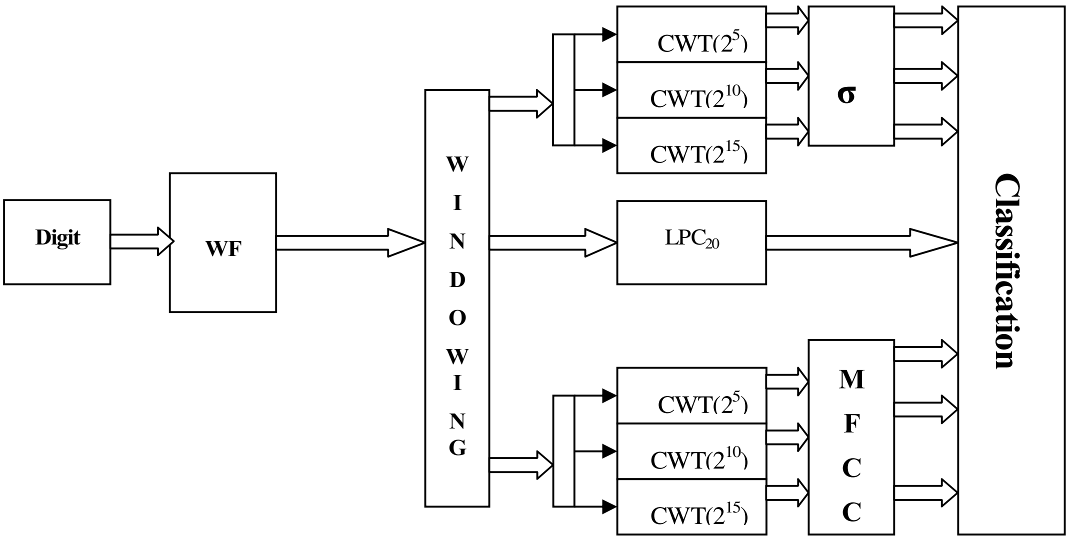



**Fig. 3.** Block diagram of the SWADC.

Same method is used for defined above each of three $MFCC_{CWT(2^j)}$ matrices. For LPC measure, tow additional vector of 25 and 30 coefficients are calculated, in order to be used as three column vectors matrix input to FFBPNC, because at least, three column vectors matrix input to FFBPNC is required. As a result we have three neural networks for each digit to be classified, which are used for models features simulated.

## 4. Experimental Results and Discussion

In this research paper, speech signals were recorded via PC-sound card, with a spectral frequency of 4000 Hz and sampling frequency of 8000 Hz. For each speaker, the Arabic digits from zero to nine were recorded 3 times by, in three Arabic dialects: Egyptian Arabic Dialect (EAD), Jordanian Arabic Dialect (JAD) and Palestinian Arabic Dialect PAD (tabulated in Tab.2); 3 females, aged one and four years old along with 11 males participated in speech digits recording. The recording process was provided in normal university office conditions.

Our investigation of Dialect-independent classifier system performance with robustness to noise is presenting some experiments depending on several considered aspects, are studied in the following part of the Section 4.

*Experimental-1*

Tab.4 shows the results of Recognition Rate, which we defined as Classification Rate in this research paper of the dialect-independent digits database of the recorded spoken Arabic digits. We experimented 450 different signals of different individuals.

**Table 4.** Recognition Rate of individual Arabic digits and mixed recognized digits.

| ***Digit*** | | ***1*** | ***2*** | ***3*** | ***4*** | ***5*** | ***6*** | ***7*** | ***8*** | ***9*** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 45 | | | | | | | | | | |
| ***1*** | | 43 | | 2 | 1 | | | | | | |
| ***2*** | | | 31 | | 1 | | | | | | |
| ***3*** | | | | 37 | | | | | 1 | | |
| ***4*** | 1 | | | | 42 | | | 3 | | | |
| ***5*** | | | | | | 40 | | | | | |
| ***6*** | | | | | | | 36 | | | | |
| ***7*** | | | | 1 | | | 1 | 43 | | | |
| ***8*** | 1 | | | 1 | | | | | 44 | | |
| ***9*** | | | | | 1 | | | | | 42 | |
| ***Rec. Rate [%]*** | 95 | 100 | 100 | 81 | 95 | 100 | 97 | 93 | 98 | 100 | **Total 95.9 [%]** |

**Table 5.** Recognition individual Arabic digits (0, 1, 2, 3, and 4) and mixed recognized digits via WGN with WF or without.

| | | | ***1*** | | ***2*** | | ***3*** | | ***4*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **SNR [dB]** | **With WF** | **Without WF** | **With WF** | **Without WF** | **With WF** | **Without WF** | **With WF** | **Without WF** | **With WF** | **Without WF** |
| −2.7302 | | | | 9 | 0 | 0 | **3** | | 3 | 3 |
| −0.4221 | | | 3 | 9 | 0 | 0 | **3** | | 3 | 3 |
| 1.8427 | | | 3 | 9 | 0 | 0 | **3** | | 3 | 3 |
| 4.2329 | | | 9 | 9 | 0 | 0 | **3** | **3** | 3 | 3 |
| 6.5969 | | | 9 | 9 | 0 | 0 | **3** | **3** | **4** | 3 |
| 8.8047 | | | 9 | 9 | **2** | | **3** | **3** | **4** | 3 |
| 11.1756 | | | **1** | 9 | **2** | **2** | **3** | **3** | **4** | **4** |
| 13.3849 | | | **1** | 9 | **2** | **2** | **3** | **3** | **4** | **4** |

The involvement of the experiment provides a study of classification capability via RDSM and FFBPNC. The results, contained in Tab.4, showed the Recognition Rate for every individual spoken digit. Table 4, also showed the mixed recognition digit for certain cases. Confused similarity is seen for Arabic digit 3, where was mixed with digits 1, 7 and 8. So that, less Recognition Rate was obtained for digit 3. Large mixed recognition between digits 7 and 3 found because of the similarities between these two digits as shown in Fig. 2. Some confusing results were observed between 4 and 1, and between 2 and 9. For example, Digits 4 and 1 have the same penultimate phoneme a short vowel /a/. There are number of common peaks in PSD curves for the digits given above. Despite the fact of, digits 4 and 2 have no important phonetic or syllable morphological similarities, same mixed recognition results were observed. A comparison Between 4 and 9, was found with some confusing results as shown in table 4. This occurs because of large morphological and pronunciation similarities (4-ʼaâr-bâ-?âh and 9- tes-âh).

Two kinds of classification methods were used, FFBPNC and RDSM for classification extracted features. It important to state that, these two methods have nearly same Recognition Rate, but different computational time needed for FFBPNC over the RDSM method by 3.

*Experimental-2*

In experiment-2, wavelet transform was added to WF block to compare the effect of utilizing feature extracting with and without as shown in Tab.5 in addition with the CWT as shown in Tab.6. The system performance associated with noise is examined using White Gaussian Noise (WGN). WGN is added to the digits speech signals based on Gaussian Density defined mathematically as:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{\frac{-(x-\mu_X)^2}{2\sigma_X^2}}$$

Where $\mu$ is mean value and $\sigma$ is standard deviation of a random variable$X$. The noise process $N(t)$ is called white noise if the power density spectrum is a constant at all frequencies$P_{NN}(\omega) = N_o/2$, and autocorrelation $R_{NN} = (N_O/2)\delta(\tau)$, where $N_O$ is a real positive constant and $\delta(\tau)$ is unit-impulse function. The results were obtained at different SNR levels (Tab.5). The results prove the effect of such utilization on robustness to noise.

*Experimental-3*

In experiment-3, A comparison between the proposed classifier (SWADC) and Alotaibi classifier [7] was investigated. The aspect of comparison is based on adding White Gaussian Noise to five Arabic digits signals, in different magnitude, as shown in Tab.7. The two classifiers are implemented to the same signals via several SNR. The aim of this implementation is to investigate the classifiers robustness to noise degree. We conclude according to the results contained in Tab.7 that the proposed classifier has superior performance. In term of Recognition Rate in dialect-independent system, and according to our results, Alotaibi classifier has considerable results about 80%. Taking in consideration the proposed classifier Recognition Rate contained in Tabl.4 was

**Table 6.** Recognition individual Arabic digits (2, 6, 7, 8, and 9) and mixed recognized digits via WGN with CWT or without.

| | *2* | | *6* | | *7* | | *8* | | *9* | |
|---|---|---|---|---|---|---|---|---|---|---|
| **SNR** | **With CWT** | **Without CWT** | **With CWT** | **Without CWT** | **With CWT** | **Without CWT** | **With CWT** | **Without CWT** | **With CWT** | **Without CWT** |
| −4.5718 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| −2.2667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0.0372 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2.2626 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4.6439 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7.1312 | **2** | | | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9.3465 | **2** | | | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11.4867 | **2** | | | 3 | 0 | | **8** | | | |
| 14.2551 | **2** | **2** | 3 | 1 | 0 | 0 | **8** | | **9** | |
| 16.0557 | **2** | | **6** | 1 | 0 | 0 | **8** | **8** | **9** | **9** |
| 20.0322 | **2** | **2** | **6** | **6** | | | **8** | **8** | **9** | **9** |
| 23.4214 | **2** | **2** | **6** | **6** | **7** | 1 | **8** | **8** | **9** | **9** |

**Table 7.** Comparison between SWADC and Alotaibi classifiers in term of noise robustness.

| | | *2* | *6* | | *7* | | *8* | | *9* | |
|---|---|---|---|---|---|---|---|---|---|---|
| **SNR** | **Alotaibi** | **SWADC** | **Alotaibi** | **SWADC** | **Alotaibi** | **SWADC** | **Alotaibi** | **SWADC** | **Alotaibi** | **SWADC** |
| −4.5718 | 6 | 0 | 8 | 3 | 0 | 0 | 0 | 0 | **9** | |
| −2.2667 | 6 | 0 | 8 | 3 | 0 | 0 | 0 | 1 | **9** | |
| 0.0372 | 6 | 0 | 8 | 3 | 0 | 0 | 0 | 3 | **9** | **9** |
| 2.2626 | 6 | 0 | 8 | 3 | 0 | 0 | 0 | **8** | **9** | **9** |
| 4.6439 | 6 | 0 | 8 | 6 | 9 | 0 | 0 | **8** | **9** | **9** |
| 7.1312 | 6 | **2** | 8 | **6** | 9 | 0 | 0 | **8** | **9** | **9** |
| 9.3465 | **2** | **2** | 8 | **6** | 9 | 0 | 0 | **8** | **9** | **9** |
| 11.4867 | **2** | **2** | 8 | **6** | 5 | | | **8** | **9** | **9** |
| 14.2551 | **2** | **2** | **6** | **6** | 5 | 0 | 0 | **8** | **9** | **9** |
| 16.0557 | **2** | **2** | **6** | **6** | 5 | 0 | 0 | **8** | **9** | **9** |
| 20.0322 | **2** | **2** | **6** | **6** | 5 | | | **8** | **9** | **9** |
| 23.4214 | **2** | **2** | **6** | **6** | 5 | **7** | **8** | **8** | **9** | **9** |

95.9%, we state that it has superior performance than Alotaibi classifier. The results were obtained at different SNR levels (shown in Tab.5) provides the benefits of WF and CWT in facing the noise difficulties at considerable degree. The reason of this success is based on the utilization of the sophisticated features extraction method, with several mathematical aspects proposed in this research paper.

## 5. Conclusion

The sophisticated features extraction method with several mathematical aspects (Continuous Wavelet Transform with LPC and MFCC) is proposed. This sophisticated approach is proposed to face the dialect-independent and speaker-independent difficulties. Root Mean Square Difference Similarity Measure and Feed Forward Back Propagation Neural Network Classification are utilized in the classification part of the proposed classifier. The proposed classifier has high Recognition Rate reached up to 100%, in some cases, and the average rate reached up to 95.9%, for about 450 tested digits signals. The results also provided the effect of such utilization on robustness to noise. The comparison between the proposed classifier and Alotaibi classifier was investigated. We concluded, according to the results contained in Tab.7 that proposed classifier has superior performance. In term of Recognition Rate in dialect-independent system according to our implementation results, Alotaibi classifier has considerable results about 80%, less than proposed classifier. The reason of this success is the utilization of the sophisticated extraction based on Wavelet Transform in conjunction with LPC and MFCC.

## References


[1] O. Douglas, "Interacting with computers by voice: Automatic speech recognition and synthesis," *Proceeding of the IEEE*, vol. 91, no. 9, pp. 141–159, 2003.

[2] S. Datta, M. A. Zabibi, and O. Farook, "Exploitation of morphological in large vocabulary arabic speech recognition," *International Journal of Computer Processing of Oriental Language*, vol. 18, pp. 291–302, 2005.

[3] K. Kirschhoff, "Novel approach to arabic speech recognition. final report from the jhu summer school workshop," in *Proceedings of the International Conference on ASSP*, pp. 344–347, 2002.

[4] S. A. Selouani and J. Caelen, "Recognition of arabic phonetic features using neural networks and knowledge-based system: a

comparative study," *International Journal of Artificial Intelligence Tools*, vol. 8, no. 1, pp. 73–103, 1999.

[5] M. Debyeche, J. P. Haton, and A. Houacine, "A new vector quantization approach for discrete hmm speech recognition system," *International Scientific Journal of Computing*, vol. 5, no. 1, pp. 72–78, 2006.

[6] M. Shoaib, M. Awais, S. Masud, S. Shamail, and J. Akhbar, "Application of concurrent generalized regression neural networks for arabic speech recognition," in *Proceedings of the IASTED International Conference on Neural Networks and Computational Intelligence*, pp. 206–210, 2004.

[7] Y. A. Alotaibi, "Investigating spoken arabic digits in speech recognition setting," *Information Sciences*, vol. 173, pp. 115–139, 2005.

[8] A. Amrouche, M. Debyeche, A. T-Ahmed, ean M. Rouvaen, and M. C. E. Yagoub, "An efficient speech recognition system in adverse conditions using the nonparametric regression," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 1, pp. 85–94, 2009.

[9] H. Bourouba, R. Djemili, M. Bedda, and C. Snani, "New hybrid system (supervised classifier/HMM) for isolated arabic speech recognition," in *Proceedings of the Second IEEE International Conference on Information and Communication Technologies*, pp. 1264–1269, 2006.

[10] K. Saeed and M. Nammous *Information Processing and Security Systems*, pp. 55–66, 2005. A New Step in Arabic Speech Identification: Spoken Digit Recognition.

[11] L. Lazli and M. Sellami, "Connectionist probability estimation in hmm arabic speech recognition using fuzzy logic," *Lectures Notes in LNCS*, vol. 2734, pp. 379–388, 2003.

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

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

[14] K. Saeed and M. Nammous, "Heuristic method of arabic speech recognition," in *Proceedings of the IEEE International Conference on Digital Signal Processing and its Applications*, pp. 528–530, 2005.

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

[16] K. Daqrouq and A. Al-Qawasmi, "The study of wavelet filters speech enhancement method," *Third Mosharaka International Conference on Communications, Computers and Applications*, pp. 26–28, Oct. 2009.

[17] L. Karray and A. Martin, "Towards improving speech detection robustness for speech recognition in adverse conditions," *Speech Communication*, vol. 40, pp. 261–276, 2003.

[18] M. H. Savoji, "A robust algorithm for accurate endpointing of speech signals," *Speech Communication*, vol. 8, pp. 45–60, 1989.

[19] J.-C. Junqua, B. Mak, and B. Reaves, "A robust algorithm for word boundary detection in the presence of noise," *IEEE Transactions Speech Audio Process*, vol. 2, no. 3, pp. 406–412, 1994.

[20] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *International Conference on Acoustics, Speech, and Signal Processing*, pp. 208–211, 1979.

[21] C. Mokbel, D. Jouvet, and J. Monn_e, "Blind equalization using adaptive filtering for improving speech recognition over telephone," in *European Conference on Speech Communication and Technology*, pp. 141–1990, 1995.

[22] H. Hermansky, N. Morgan, and H. G. Hirsch, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing," in *International Conference on Acoustics, Speech, and Signal Processing*, pp. 83–86, 1993.

[23] M. R. Mashinchi, M. H. Mashinchi, and A. Selamat, "New approach for language identification based on DNA computing," *International Conference on Bioinformatics & Computational*, pp. 748–752, 2007.

[24] I. Zitouni and R. Sarikaya, "Arabic diacritic restoration approach based on maximum entropy models," *Computer Speech and Language*, vol. 23, pp. 257–276, 2009.

[25] K. Daqrouq and N. M. Abu-Sheikha, "Heart rate variability analysis using wavelet transform," *Asian Journal for Information Technology*, vol. 4, no. 4, 2005.

[26] P. Li, Y. Guan, S. Wang, B. Xu, and W. Liu, "Monaural speech separation based on maxvq and casa for robust speech recognition," *Computer Speech & Language*, vol. 24, pp. 30–44, January 2010.

[27] M. J. F. Gales and F. Flego, "Discriminative classifiers with adaptive kernels for noise robust speech recognition," *Computer Speech & Language*. In Press, Corrected. Proof, Available online 27 September 2009.

[28] S. Mallat, "A theory for multiresolution signal decomposition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[29] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 37, no. 12, pp. 2091–2110, 1989.

[30] K. Daqrouq and I. N. Abu-Isbeih, "Arrhythmia detection using wavelet transform," in *IEEE Region 8 EUROCON 2007 International Conference*, (Warsaw, Poland), 2007.

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. He is currently a lecturer researcher 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.

Mikhled Alfaouri, Associate Professor, Department of Communication and Electronics Engineering, Philadelphia University, Amman Jordan. He received his B. S., Electrical Engineering, New Mexico State University, 1987, M.S. (Electromagnetics and Communication systems), New Mexico State University, 1988, Ph.D. (Antenna Theory & Design), New Mexico State University, USA, 1991. He published 25 papers in the national and international conferences. He published 30 papers in the national and international journals. His areas of interests are Application of Signal Processing Techniques to Electromagnetics problems and Using Wavelet Transform for data Recognition and Identifications. He is a member of many professional societies. Dr. Mikhled Alfaouri is a writer of a political column at exogenesis and indigenous newspapers.