The Use Of Wavelet Entropy in Conjuction with Neural Network for Arabic Vowels Recognition

*W. AL-SAWALMEH, **K.DAQROUQ AND ***O.DAOUD

*Department of Communication Engineering
Al-Hussein Bin Talal University
E-Mail: Walsawalmeh@ahu.edu.jo
P.O. Box 20, Ma'an
JORDAN
**Comp. & Electrical Eng. Dept,
King Abdulaziz Univ., Jeddah, Saudi Arabia,
Email:haleddaq@yahoo.com, P.O. Box 80204
*** Department of Communication and Electronics Engineering
Philadelphia University
P.O. Box 1, Amman 19392
JORDAN

Abstract— In this research paper, Arabic vowels recognition system using very promising techniques; wavelet packet transform (WT) with entropy and neural network was presented. Trying to enhance the recognition process, three types of entropies were applied for the wavelet packet (WP) of the speech signals. Moreover, different levels of WP were used in order to enhance the efficiency of the proposed work until level 7. To classify among the feature vectors; a probabilistic neural network (PNN) were used. A MATLAB program was used to build the model of the proposed work to show the powerfulness of 96.77% identification rate. This is due to that the functions of features extraction and classifications are performed using the entropy, wavelet packet and neural networks.

Key-words — Recognition, Wavelet; Entropy; Neural Network; and Arabic Vowels.

1 Introduction

Recently, 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. Text to speech (vice versa) as well as incredibly critical issues in many applications that are attracted the users. 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.

The major work of studying of speech recognition of Arabic language dealing with the morphological structure found in [1] or the phonetic features in order to recognize the distinct Arabic phonemes (pharyngeal, geminate and emphatic consonants) [2,3] and, discussed their further implication in a larger vocabulary speech system. This allocates and motivates interesting researchers of Arabic language with different dialect at various countries.

spoken separated 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 research of Arabic spoken with different dialect. [4] has studied the derivative scheme, named the Concurrent 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 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 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.

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].

From literatures papers, other researchers, neural networks were used 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.

Also, the noisy environments degrade the recognition performance considerably. Robustness to noise is then necessary for professional using recognition systems particularly in mobile networks context; [14]. A variety of studies have been conducted in this track [15,16,17]. Various 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 [18,19] remove ambient noise. The transmission effects are reduced using equalization techniques such as cepstral normalization and adaptive filtering [20, 21].

This paper presents a new combination of wavelet transform entropy and probabilistic neural network.

The main objective of such sophistication conjunction is to create a dialect-independent Arabic vowels classifier. The remainder of the paper is organized as follows: In section 2 we present the brief introduction to Arabic language. The proposed method is described in section 3 and section 4. The experimental results and discussion in section 5 followed in section 6 by conclusions.

2 Brief Introduction to Arabic Languages

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.

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 distinct dialects and socialists. Classical Arabic is an older, literary form of the language, exemplified by the type of Arabic utilized in the Quran. Modern Standard Arabic (MSA) is a version of Classical Arabic with a modern vocabulary. MSA is a formal standard common to all Arabic-speaking countries.

It is the language used in the newspapers, radio and TV, in official speeches, in courtrooms, and, usually 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 particular 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.

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	باً	/ban/
Tanween Aldam	بٌ	/bun/
Tanween Alkasr	بٍ	/bin/
Sokun	بْ	/b/

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 great number of TV programs which are created 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 preceding exposure to the other speaker's dialect. However, widely differing dialects, such as Moroccan Arabic, 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 carry difficulties for automatic speech recognition. Standard, language-independent techniques for acoustic and pronunciation modeling, such as context-dependent phones, may 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, 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 spoken Arabic; while, MSA data may readily be acquired from various media sources.

Table 2: Three examples of four different Arabic dialects

Gloss	MSA	EAD	JAD	PAD
'Three' ثلاث	thā-lā-thāh	tā-lā-tāh	thā-lā-thĕh	tā-lā-tĕh
'Eight' ثمانية	thā-mā-nê-yah	tā-mā-n-yah	thā-mā-n-yeh	tā-mā-n-yeh
'Two' أثنين	?ith-nān	te-nān	?ith-nen	?it-nān

In conclusion, morphological complexity is approved to present solemn problems for speech recognition. A high scale of affixation, derivation etc. contributes to the explosion of unlike word forms, making it difficult if not impossible to robustly estimate language model probabilities. Wealthy morphology also leads to elevated out-of-vocabulary rates and bigger search spaces during decoding, thus slowing down the recognition process [23].

3 Features extraction Method by Wavelet Packet Transform

The wavelet packet algorithm is a generalization of wavelet decomposition that offers a richer signal analysis. Wavelet packet atoms are waveforms indexed by means of three naturally interpreted parameters: position, scale and frequency. In the following, the wavelet transform is defined as the inner product of a signal $x(t)$ with the conjugate mother wavelet $\psi(t)$:

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right) \tag{1}$$

$$W_\psi x(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{2}$$

where a and b denotes the scale and shift parameters, respectively. The mother wavelet is dilated or translated by modulating a and b.

The wavelet packets transform performs the recursive decomposition of the signal obtained by the recursive binary tree (as seen at Fig.1). Basically, the WPT is very similar to Discrete Wavelet Transform

(DWT) but WPT decomposes both details and approximations unlike DWT that performs the decomposition process on approximations. The principle of WP is that, given a signal, a pair of low pass and high pass filters is used to yield two sequences to capture different frequency sub-band features of the original signal. The two wavelet orthogonal bases created form a previous node are defined as

$$\psi_{j+1}^{2p}(k) = \sum_{n=-\infty}^{\infty} h[n]\psi_j^p(k-2/n) \quad (1)$$

$$\psi_{j+1}^{2p}(k) = \sum_{n=-\infty}^{\infty} g[n]\psi_j^p(k-2/n) \quad (2)$$

where $h[n]$ and $g[n]$ are the low-pass and high-pass filters, respectively. In equations (2) and (3), $\psi[n]$ is the particular wavelet function. Parameters j and p are the number of decomposition levels and nodes of the previous node, respectively [24].

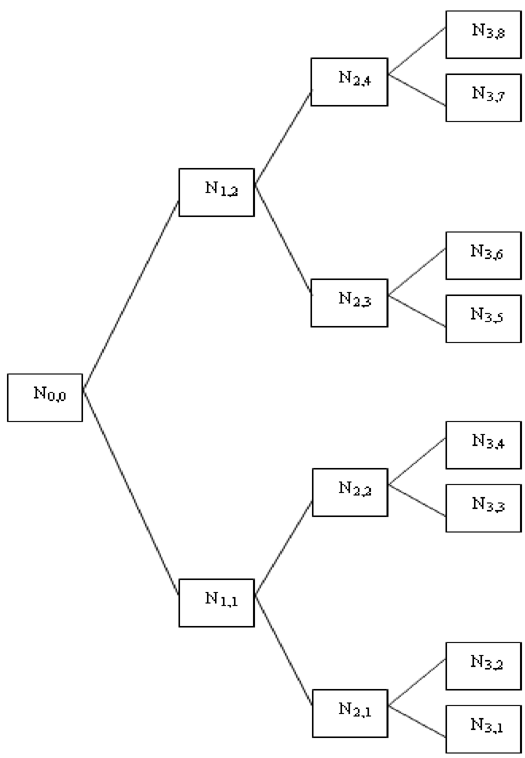


Fig. 1: Wavelet packet at depth 3

For a given 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 speech features. Previous studies proposed that the use of entropy of WP as features in recognition tasks is competent. [25] Suggested a method to calculate the entropy value of the wavelet norm in digital modulation recognition. [26] Proposed features extraction method for speaker recognition based on a combination of three entropies types (sure, logarithmic energy and norm). Lastly, [27] 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.4). In this paper, the entropy obtained from the WPT will be employed for Arabic vowels recognition. The wavelet packet features extraction method can be explained as follows:

- 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.
- Decomposing the speech signal by wavelet packet transform at level 7, with Daubechies type (db2).
- Calculating three entropy for all 256 nodes at depth 7 for wavelet packet using the equations [27]:

Shannon entropy:

$$E1(s) = -\sum_i s_i^2 \log(s_i^2) \quad (3)$$

entropy:

$$E1(s) = \sum_i \log(s_i^2) \quad (4)$$

Sure entropy:

$$|s_i| \le p \Rightarrow E(s) = \sum_i \min(s_i^2, p^2) \quad (5)$$

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.5 shows Shannon entropy calculated for WP at depth 7 for a-vowel and 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 Vary widely from class to class. 2) Stable over a long period of time. 3) Should not have correlation with other features (see Figure4).

- For a better demonstration of the sub-band signals, the energy of speech is commonly computed. Previous investigations showed that the utilization of an energy index as features in recognition roles is efficient [29]. In this work, the energy index of the WP is employed as additional features in conjunction with entropies for vowel recognition task.

3.1 Proposed Probabilistic Neural Networks Algorithm

We create a probabilistic neural network algorithm for classification problem (see Fig.2 and Fig.3):

$$Net = PNN\,(P, T, SPREAD),$$

where P is $4x\,2^{q+1} x\,24$ matrix of 24 input vowel feature vectors for net training, of 2^{q+1} (minus 2, repeated original node) WP nodes number;

$$P = \begin{bmatrix} WR_{11} & WR_{12}\,, ..., WR_{124} \\ WR_{21} & WR_{22}\,, ..., WR_{224} \\ . & . \quad\quad . \\ . & . \quad\quad . \\ . & . \quad\quad . \\ WR_{4x\,2^{q+1}1} & WR_{4x\,2^{q+1}2}, ..., WR_{4x\,2^{q+1}24} \end{bmatrix}, \tag{15}$$

T is the target class vector

T=[1,2,3, ...,24], (9)

and SPREAD is spread of radial basis functions. We employ a SPREAD value of 1 because that is a typical distance between the input vectors. If SPREAD is near zero the network acts as a nearest neighbour classifier. As SPREAD becomes larger the designed, network will take into account several nearby design vectors.

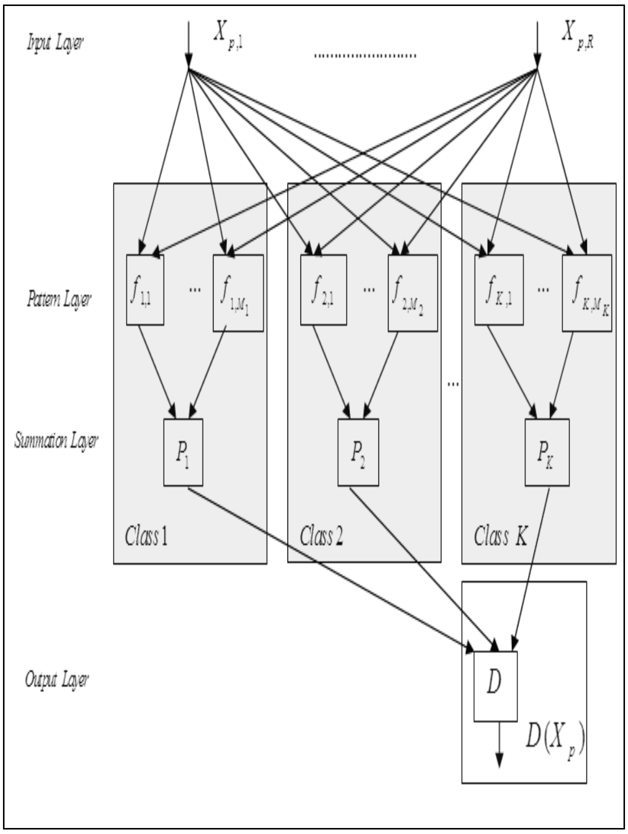


Fig. 2: Structure of the original probabilistic neural network

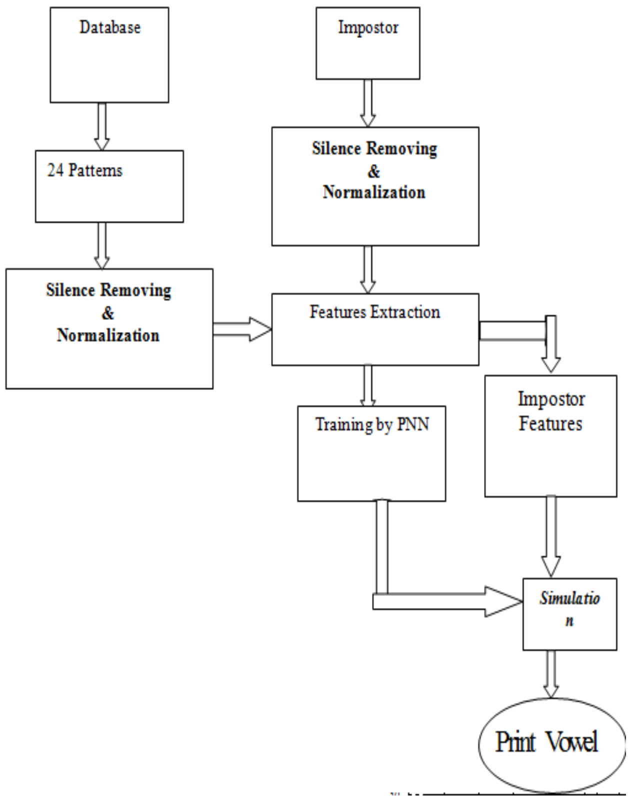


Fig. 3: Flow chart for proposed expert system

4 Results Analysis 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 20 speakers: 5 females, along with 15 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.

Experimental-1

We experimented 180 long Arabic vowels أ (pronounced as/a:/) signals of 20 speaker and 358 long Arabic vowels ي (pronounced as/e:/) signals. The results indicated that 96.77% were classified correctly. 85.75% of the signals were classified correctly for Arabic vowelsي. Tab.3 shows the results of recognition rates.

Experimental-2

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/). Table 4, reported the recognition rates. The results indicated 80.28% average recognition rate.

Experimental-3

In experiment-2, short Arabic vowels: fatha: represents the /a/ sound and kasra: represents the /i/ sound (pronounced as/e:/) for each vowel 300 signals
of 20 speaker results are reported. The recognition rates of above mentioned two short vowels connected with other letter such ل (pronounced as/l/) and ر (pronounced as/r/). The average recognition rate was 74.41% (Tab.5).

Table 3: The recognition rate results for long vowels

Long Vowels	Number of Signals	Accepted Signals	Rejected Signals	Recognition Rate [%]
Long A أ	186	180	6	96.77
Long E ي	358	307	51	85.75
			Avr. Recognition Rate	91.26

Table 4: The recognition rate results for long vowels connected with other letters

Short Vowels	Number of Signals	Recognized Signals	Not Recognized Signals	Recognition Rate [%]
Le لي	300	230	70	76.67
La لا	300	240	60	80.00
Re ري	300	251	51	83.66
Ra را	300	242	58	80.80
			Avr. Recognition Rate	80.28

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

Short Vowels	Number of Signals	Recognized Signals	Not Recognized Signals	Recognition Rate [%]
Le لي	300	224	76	74.66
La لا	300	222	38	74.00
Re ري	300	250	50	83.33
Ra را	300	197	103	65.66
			Avr. Recognition Rate	74.41

5 Conclusions

Probabilistic neural network based speaker verification system is proposed in this paper. This system was developed using a wavelet packet 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. Three types of entropy coefficients of WPT in conjunction with energy indexes of WPT are utilized. 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 speaker-independent Arabic vowels classifier system performance is performed via several experiments depending on vowel type. The declared results show that the proposed method can make an effectual analysis with identification rates may reach 96.77%.

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] Karray L. and Martin A., Towards improving speech detection robustness for speech recognition in adverse conditions, Speech Communication 40 (2003) 261–276.

[15] Savoji, M.H., 1989. A robust algorithm for accurate endpointing of speech signals. Speech Communication 8, 45–60.

[16] Mauuary, L., Monn_e, J., 1993. Speech/non-speech detection for voice response systems. In: European Conference on Speech Communication and Technology, Berlin, September 1993,pp. 1097–1100.

[17] Junqua, J.-C., Mak, B., Reaves, B., 1994. A robust algorithm for word boundary detection in the presence of noise. IEEE Trans. Speech Audio Process. 2 (3), 406–412.

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

[19] Mokbel, C., Mauuary, L., Karray, L., Jouvet, D., Monn_e, J., Simonin, J., Bartkova, K., 1997. Towards improving ASR robustness for PSN and GSM telephone applications. Speech Communication 23 (1–2), 141–159.

[20] Hermansky, H., Morgan, N., Hirsch, H.G., 1993. 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.

[21] Mokbel, C., Jouvet, D., Monn_e, J., 1995. Blind equalization using adaptive filtering for improving speech recognition over telephone. In: European Conference on Speech Communication

and Technology, pp. 141–1990.

[22] Zitouni I. and Sarikaya R., Arabic diacritic restoration approach based on maximum entropy models Computer Speech and Language, 23 (2009) 257–276

[23] Kirschhoff, K., Novel approach to Arabic speech recognition. Final Report from the JHU Summer School Workshop, (2002). Proceedings of the International Conference on ASSP (ICASSP'03), pp. 344–347.

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

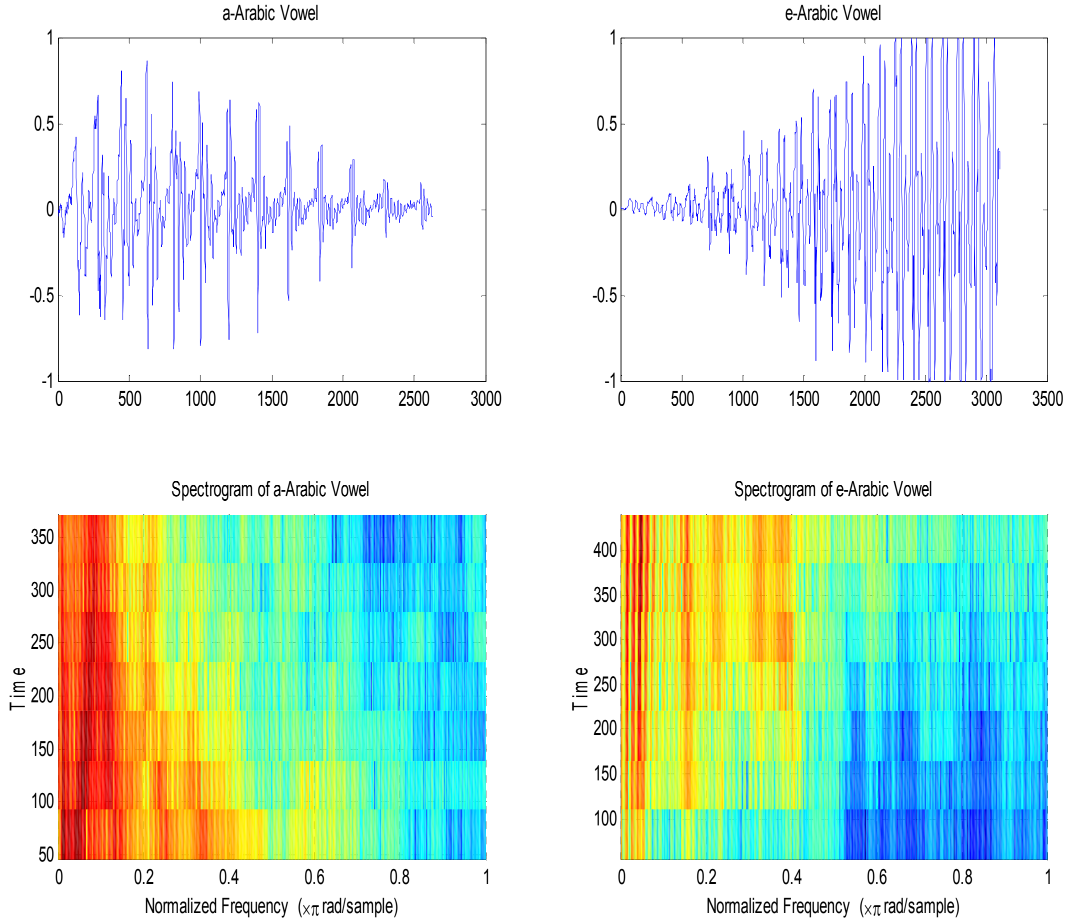
[25] Avci, E., Hanbay, D., & Varol, A. (2006). An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition. Expert System with Applications, 33, 582–589.

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

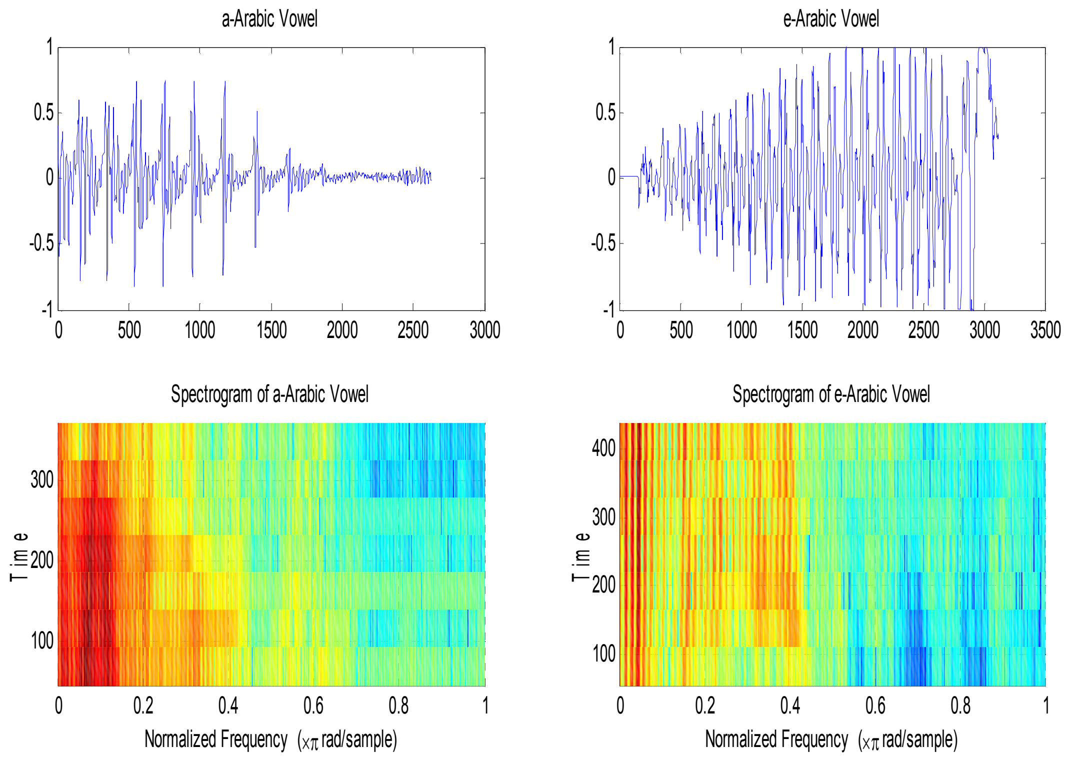
[27] Avci, D. (2009), An expert system for speaker identification using adaptive wavelet sure entropy, Expert Systems with Applications, 36, 6295–6300.

[28] Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. IEEE Transactions on Information Theory, 38(2), 713–718.

[29] Kotnik B., Kacic Z., Horvat B., Noise Robust Speech Parameterization Based on Joint Wavelet Packet Decomposition and Autoregressive Modeling, 8th European Conference on Speech Communication and Technology, Geneva, Sep. 1-4, 2003.



a.



b.

Fig. 4: Arabic vowels a. First Vowels of a speaker 1 with spectrogram b. First Vowels of a speaker 2 with spectrogram

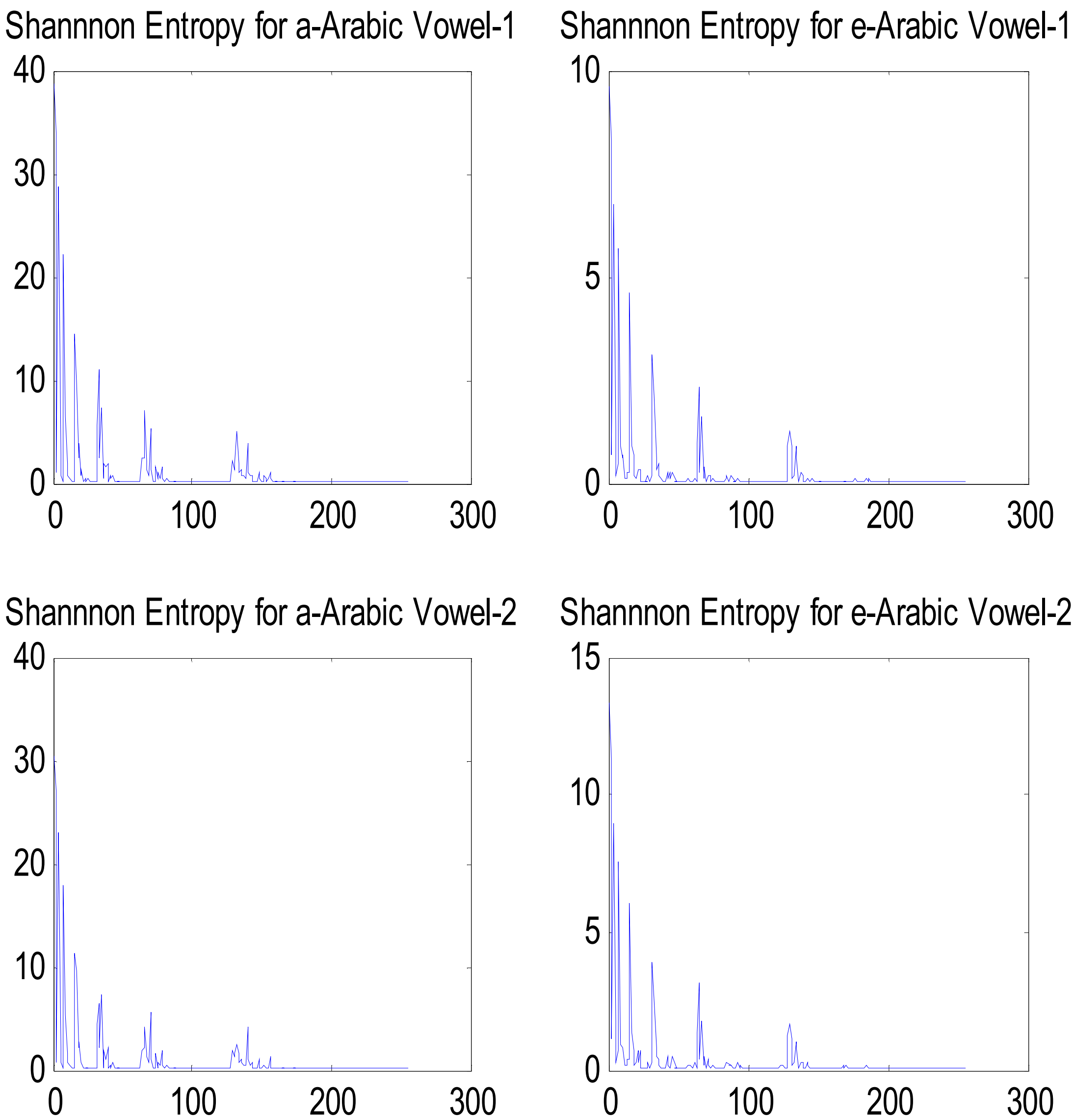


Fig. 5: Shannon entropy for Arabic vowels presented in Figure 4