DOI: https://doi.org/10.33103/uot.ijccce.18.3.3

Improve the Recognition of Spoken Arabic Letter Based on Statistical Features

Jabbar Salman[1], Thamir Rashed Saeed[2,] and Alaa Hussein Ali[3]
[1]*College of Engineering,University of Karbala, Karbala, Iraq*
[2,3]*Department of Electrical Engineering, University of Technology, Baghdad, Iraq*
jabbae.salman@uokerbala.edu.iq, thamir_rashed@yahoo.com, alaa_eng1970@yahoo.com

Abstract — *The recognition and classification of languages represent a vital factor in the computer interaction. This paper presents Arabic Sign Language recognition, which is represented as an appealing application. The work in this paper is based on three steps; preprocessing, feature extraction and classification (Recognition). The statistical features have been used than the physical features, while Multilayer feed-forward neural network as classification methods. The recognition percent is 96.33% has been gained over-perform the earlier works. The simulation has been made by using Matlab 2015b.*

I. INTRODUCTION

Speech is the communication, or an expression tool carries the information and regarding the particular features of the speaker. Where, the features are the power spectrum density, and frequency at maximum power and others which is carrying speaker information. These features can be followed well by shifting the recurring qualities of the vocal tract and the variety in the excitation. In this context, the speech signal also carries the information of the particular speaker, including social factors, affective factor and the properties of the real voices production [1]. In effect, the speech has the potential of being an important mode of interaction with the computer. Speech processing is one of the exciting areas of signal processing.

Since the 1960s PC researchers have been looking into ways and intends to make PCs ready to record translate and comprehend human discourse. Correspondence among the person is overwhelmed by talking dialect. In this way, it is normal for individuals to expect discourse interfaces with a PC which can talk and perceive discourse in local dialect [2]. In this context and for the importance of the subject there are many types of research in this area, some of these;

In [3] an acoustic event detection system is proposed, which is consists of a noise reduction by using the power spectral density estimator. While the noise suppression is made by a Gabor filter bank, While using the hidden Markov model for feature extraction as a classifier. While [4] is used Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC) Analysis, Dynamic Time Wrapping (DTW), Relative Spectra Processing (RASTA) and Zero Crossings with Peak Amplitudes (ZCPA) as speech parameters and features. Where, RASTA and MFCC are considered as the nature of speech parameters, and they are extracted as the features, while the LPC predicts the future features based on previous features. A different Feature Extraction Methods for Linear and Non-Linear kernels has been presented in [5]; like frequency Cepstral coefficient (MFCC) and perceptual linear prediction (PLP) with several features normalization techniques like Rasta is filtering and Cepstral mean subtraction (CMS). In this context, when using the kernel (linear/ nonlinear) which is based on support vector machine in combination with the Gaussian mixture model, the identification of the speaker will be done. The gender identification system based on his\her speech is presented in [6], while its feature extraction is made by using a Fast Fourier Transform (FFT) algorithm. In the recognition phase, the classifier creates a gender model based on the back-end system to categorize gender from his/her speech. In this context, the present's system uses a threshold technique as an identification tool. The

Received 4 Mar 2018; Accepted 12 Sep 2018
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

recognition accuracy of this system is 80% on average [6]. In [7] a speech recognition system based on signal processing techniques is presented. Also, the difference between the speech recognition for the particular language and adopted an ASR system based on the adopted feature extraction technique has been presented through the comparison between them.

This paper gives an overview of speech features extraction and the proposed work which consisting of three steps; preprocessing, feature extraction, classification and finally the comparison with other works.

II. SPEECH FEATURES EXTRACTION

The speech creation is composed of four processes; these are; processing the language, motor generation, articulatory movement and air emission. Where the language processing is represented in the brain, while, the motor generates commands to the vocal organs. Therefore, the emission of air sent from the lungs in the form of speech has been controlled by the motor commands, while the articulatory movement for the production of speech by the vocal organs based on these motor commands [8]. Therefore, to the voice and unvoiced speech which is created by these structures can be categorized.

Therefore, two kinds of algorithms which are used for feature extraction; the first one is related to speech processes, while the second is related to its results. The feature vectors are equivalent to the vectors of explanatory variables used in statistical procedures such as linear regression [9]. In this context, the features are;

A. Articulatory features

The speech recognition community has attracted enthusiasm for Articulatory features, where these features highlight the design of the human vocal tract and the properties of speech creation. The essential thought of this approach is to bear a proclivity to the articulatory occasions fundamental the discourse flag. This portrayal is made out of classes depicting a basic articulatory properties of discourse sounds, for example, put, way, voicing, lip adjusting, the opening between the lips, and the position of the tongue.

B. Features based on perception system

The auditory system has been based on the sensory system for the feeling of hearing. The research in speech recognition is dealing with the way in which the human can recognize the speech and use the speech information to understand the spoken language [10].

III. PROPOSAL WORK

The proposed work has been based on three steps as in figure (1); these are; the first step is preprocessing, while the second step is the statistical feature extraction, and the third step is a classification of the signal. Beforehand, the data have been collected from different persons with different letter positions in the word, with indoor environments

A. Preprocessing.

The sound signal of a Ba (ب) letter from Arabic alphabet letters has been taken in real environments as an example, as in figure (2a). Where this signal has two problems; the salient period and a number of samples. Therefore, the first stage in this step is the removing of the salient period by checking the energy. In effect, there are many algorithms to remove the salient period, in this work the adaptive threshold has been used.[11,12] as in figure (2b).

To reduce the number of samples, the number of frames must be reduced, so the effective frames have been chosen by auto-correlation. Before that, the pre-emphasis process has been made to ensure high

Received 4 Mar 2018; Accepted 12 Sep 2018

 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

frequencies magnitude are not neglected and maximize the signal to noise ratio at high frequencies through some filters as in figures(2c-f)[13,14]

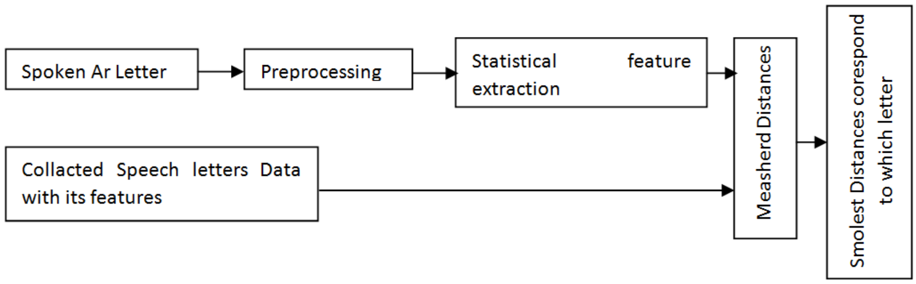
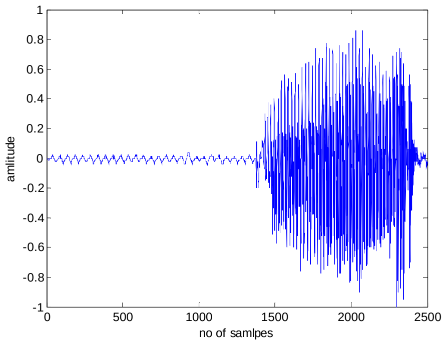
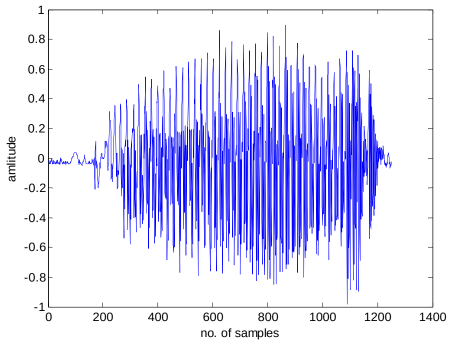


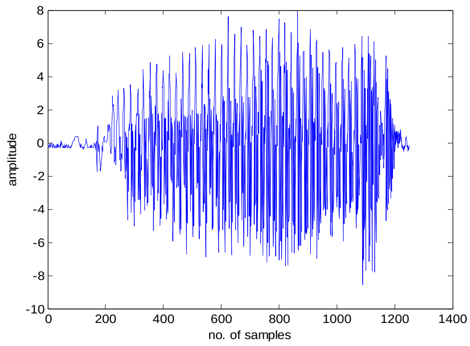
FIG. 1. BLOCK DIAGRAM OF PROPOSED WORK



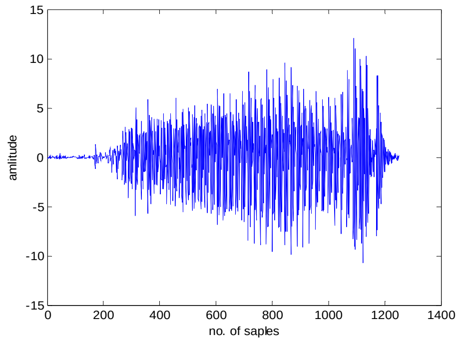
a) Original signal



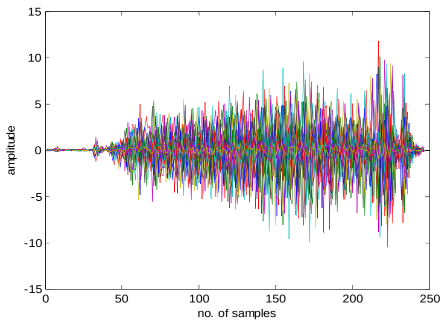
b) removing salience



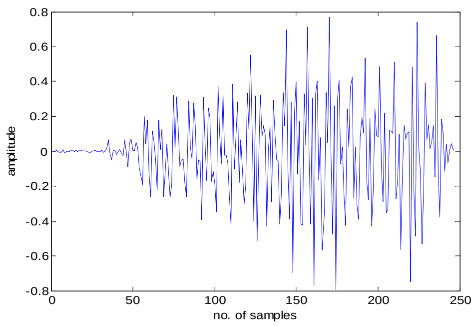
c) Normalized



d) after pre-emphases



e) Framing and windowing



f) windowing for one frame

FIG. 2. PREPROCESSING STEP

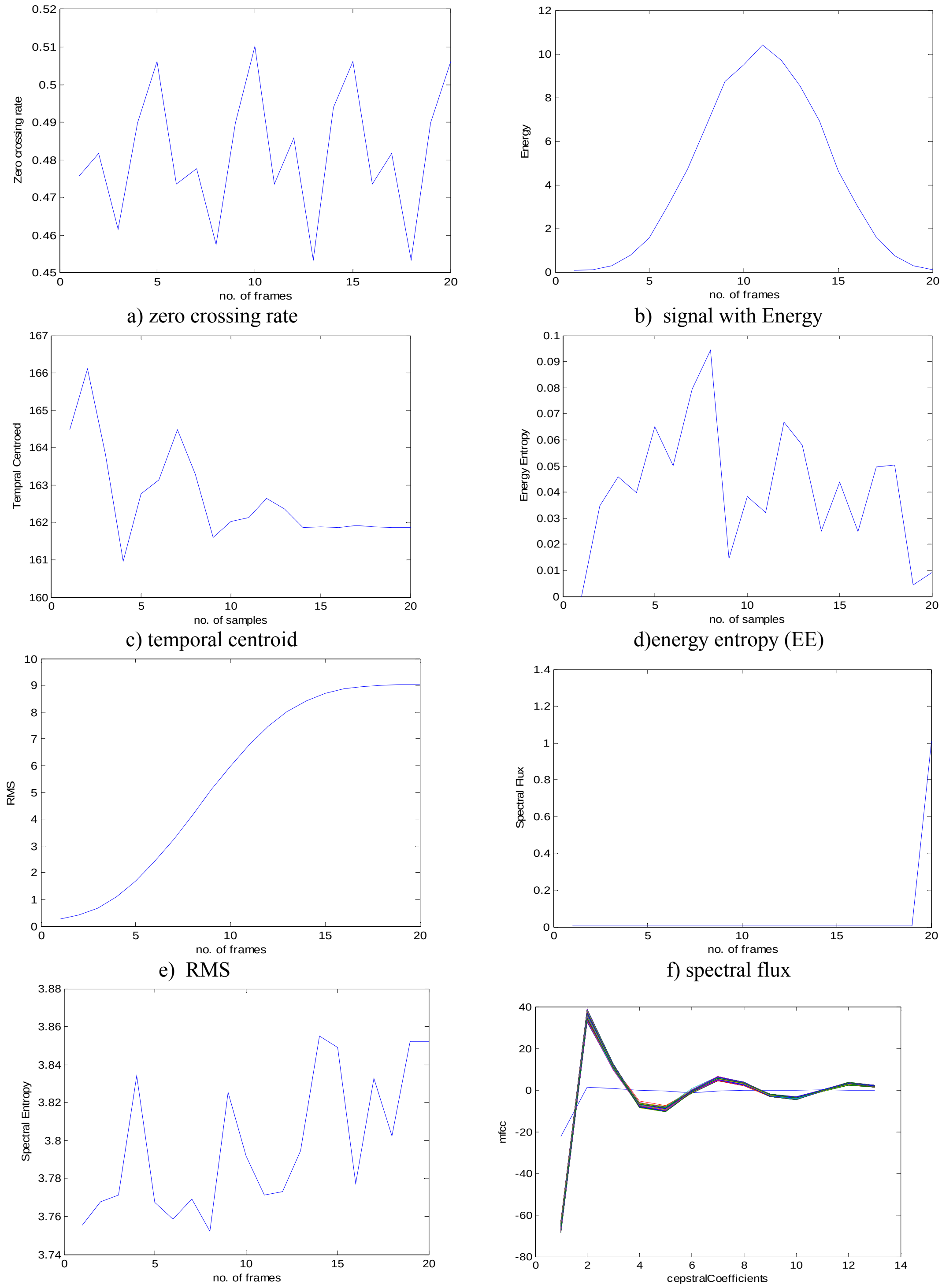
The preprocessing steps results have further strengthened our confidence in the statistical features as the classification tools, where the salience is removed, and normalization then framing and select one frame with a window, the effective frame in the window, which gives signal effect with decreasing the number of process samples as shown in the figure (2).

B. Statistical Features

Received 4 Mar 2018; Accepted 12 Sep 2018

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

Statistical properties have been representing the suitable features for sound signal, some of the statistical features are; zero crossing rate, signal energy, temporal centroid, energy entropy, RMS, spectral flux, Spectral energy and mel frequency Cepstral coefficients (mfcc) have been extracted as in figure (3)[15], where, these features contain the properties of the speech signal. After preprocessing step, the statistical feature selection has been found much more specific than the earlier used features, where, the statistical features represent the core of the signal and reduce the required size and then reduce the processing time.



a) zero crossing rate b) signal with Energy

c) temporal centroid d)energy entropy (EE)

e) RMS f) spectral flux

g) Spectral energy h) mfcc

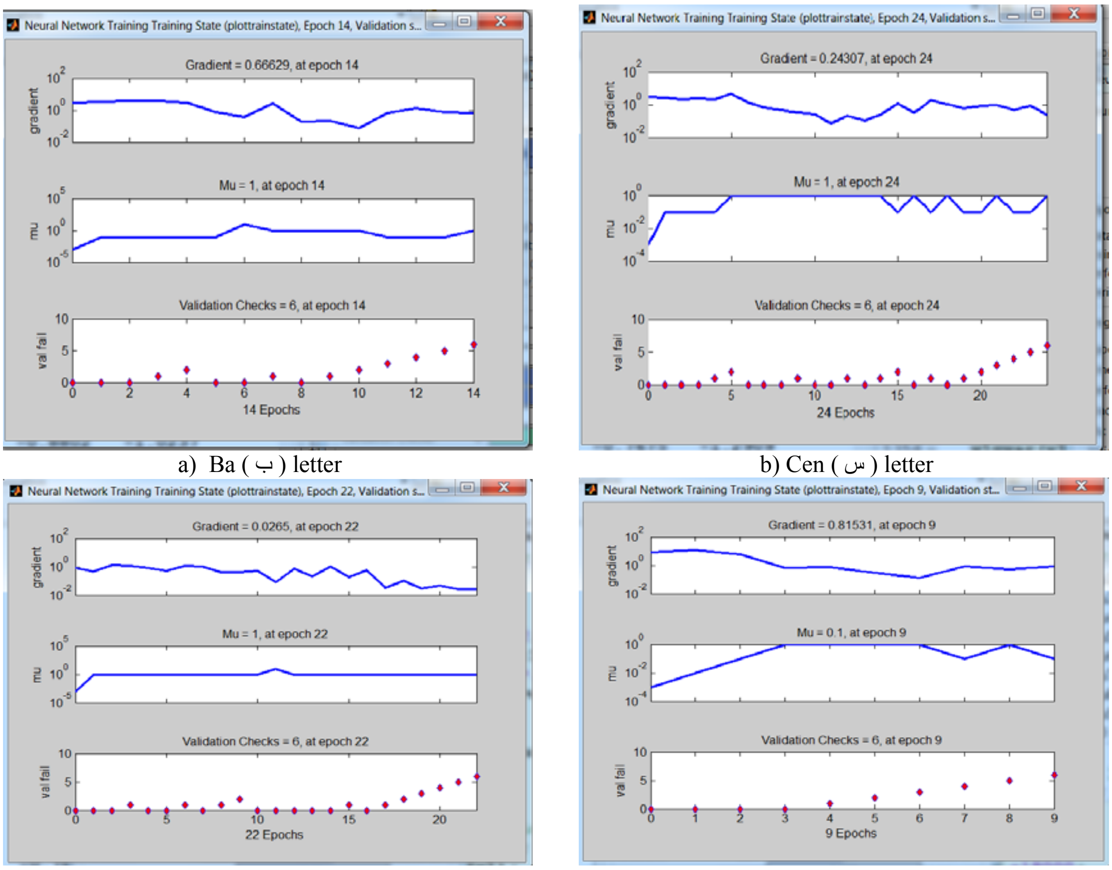
FIG. 3. STATISTICAL FEATURES OF THE SOUND SIGNAL (BA[ب])

Received 4 Mar 2018; Accepted 12 Sep 2018

 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

C. Classification

For satisfying the classification the gradient, mu and validation have been calculated to verify the proposed algorithm. Therefore, figure (4) represents these calculations for four alpha-beta Arabic letters.



a) Ba (ب) letter

b) Cen (س) letter

c) Non (ن) letter

d) Jem (ج) letter

FIG. 4. CALCULATION OF GRADIENT, MU, AND VALIDATION FOR FOUR LETTERS

While the classification step has been done by using multilayer feed-forward perceptron with the back-propagation neural network [16] with the 20 patterns for each letter (class). Therefore, the total training pattern for five letters is 50 patterns, while for testing is 10. The decision is according to the least distance criterion (Euclidian distance). The results showed acceptable results as shown in Table (1) with five letters with common classification is 96.33%.

Received 4 Mar 2018; Accepted 12 Sep 2018
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

TABLE (1) ARABIC LETTER CLASSIFICATION.

	ب	ص	س	ن	ك
ب	98.33	1.66	0	0	0
ص	5	95	0	0	0
س	0	1.66	98.33	0	0
ن	6.66	0	0	93.33	0
ك	6.66	1.66	0	0	91.66

The advancement of the proposed algorithm has been proved by the comparison between the results (classification) which has been gained with other work as presented in the Table (2). This comparison was made between the proposed work and other works for the classification of spoken language for Arabic and others due to a lack of Arabic-language works.

TABLE (2) COMPARISON PROPOSED WORK WITH OTHERS.

Ref	Classification Method	Recognition %
[6]	TMNN	90.7
[7]	MLP	96.3
[17]	PCA	96
Proposed work	MLFFNN	96.33

IV. Conclusions:

Our work has been led us to conclude the statistical features of the signal are over-performing than the physical features of that signal. Also, the multilayer feed-forward neural network classification methods are power tools, where a 96.33% has been gained as classification percent. Finally, the preprocessing step is essential for classification and then recognition goal.

REFERENCES

[1] Khalid M.O, Nahar, Moustafa, Elshafei, Wasfi G. Al-Khatib and Husni Al-Muhtaseb " Statistical Analysis of Arabic Phonemes for Continuous Arabic Speech Recognition " International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 01– Issue 02, November 2012.

[2] Satyanand Singh, " Forensic and Automatic Speaker Recognition System", IJECE, vol 8, No. 5, Oct., part1, 2018.

[3] Jens Schro¨der, Benjamin Cauchi, Marc Rene´ Scha¨dler, Niko Moritz, Kamil Adiloglu, Jo¨rn Anemu¨ller, Simon Doclo1 and Birger Kollmeier Stefan Goetze, "*Acoustic Event Detection Using Signal Enhancement and Spectro-Temporal Feature Extraction*", IEEE, Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013, New Paltz, NY, USA.

[4] Pratik K. Kurzekar, Ratnadeep R. Deshmukh and Vishal B. Waghmare, Pukhraj P. Shrishrimal,"*A Comparative Study of Feature Extraction Techniques for Speech Recognition System*", International Journal of Innovative Research in Science, Engineering and Technology, Vol. **3**, Issue 12, December 2014.

[5] Imen TRABELSI and Dorra BEN AYED, "*On the Use of Different Feature Extraction Methods for Linear and Non Linear kernels*", IEEE Sciences of Electronics, Technologies and Telecommunications, March 21-24, Tunisia, 2012.

[6] Sadek Ali, and Shariful Islam Alamgir Hossain, "*Gender Recognition System Using Speech Signal*", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.**2**, No.1, February 2012.

[7] Megha Agrawal and Tina Raikwar, "*Speech Recognition Using Signal Processing Techniques*", International Journal of Engineering and Innovative Technology (IJEIT) Volume **5**, Issue 8, February 2016.

[8] M. Honda "*Human Speech Production Mechanisms*". NTT Technical Review, Vol. **1** No. 2 May 2003.

[9] Liu, H. and Motoda H., "*Feature Selection for Knowledge Discovery and Data Mining*", Kluwer Academic Publishers, Norwell, MA, USA. The Springer International Series in Engineering and Computer Science book series (SECS, volume 454) 1998.

Received 4 Mar 2018; Accepted 12 Sep 2018

 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

[10] D. Kim and S. Lee "*Auditory processing of speech signals for robust speech recognition on realworld noisy environment*", IEEE Transaction on Speech and Audio Processing, vol. **7**, no. 1, 1999.

[11] Tae Hong Park, " Salient Feature Extraction of Musical Instrument Signals", MS.c in Electro-Acoustic Music, DARTMOUTH COLLEGE Hanover, New Hampshire, June 2, 2000.

[12] Ruairí de Fréin, " APHONIC: Adaptive Thresholding for Noise Cancellation in Smart Mobile Environments", IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Oct. 9th to Oct. 11th, 2017, Rome, Italy.

[13] R. Vergin and D. O'Shaughnessy, " Pre-emphasis and speech recognition", Canadian Conference on Electrical and Computer Engineering, 5-8 Sept. 1995.

[14] Manal El-Obaid and Amer Al-Nassir Iman Abuel Maaly, " *Arabic phoneme recognition using neural networks*", Proceeding SIP'06 Proceedings of the 5th WSEAS international conference on Signal processing Pages 99-104, Istanbul, Turkey - May 27 - 29, 2006.

[15] Gebremedhin T. Abreha, " An Environmental Audio{Based Context Recognition System Using Smartphones", MS.c in Embedded Systems University of Twente Faculty of Electrical Engineering, Mathematics and Computer Science, August 2014.

[16] Ghufran M. Hatem, Thamir R. Saeed, and Jafar W. Abdul Sadah, " Classification of Radar Non-homogenous Environment Based on Statistical Features using Artificial Neural Network", submitted to International Journal of Reasoning-based intelligent system for special issue " Advance in intelligent information communication technology, 2018.

[17] Moaz Abdulfattah Ahmad,Rasheed M. ElAwady, " Phonetic Recognition of Arabic Alphabet letters using Neural Networks", International Journal of Electric & Computer Sciences IJECS-IJENS Vol: 11 No: 01, Feb. 2011.

Received 4 Mar 2018; Accepted 12 Sep 2018

 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352