QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

NEURAL FUZZY PETRI NET BASED ARABIC PHONEME CLASSIFIER WITH MFCC FEATURE EXTRACTION

Abduladhem Abdulkareem Ali
Department of Computer Engineering , College of Engineering , University of Basrah, IRAQ.
abduladhem@ieee.org
Ghassaq S. Mosa
Department of Computer Engineering , College of Engineering , University of Basrah, IRAQ

ARTICLE INFO

Article History:

Received: 8 March 2017

Accepted: 1 April 2017

Published: 10 April 2017

DOI:

10.25212/lfu.qzj.2.2.38

Keywords: neural fuzzy Petri net, Phoneme recognition, speech recognition, Mel Frequency Cepstral Coefficient, pattern recognition.

ABSTRACT

In this paper Arabic phoneme classification is employed using Mel Frequency Cepstral Coefficient (MFCC) as the basic recognition features. These features are first calculated, then used as an input to fuzzy neural Petri net. One network are used for each phoneme. The network was first trained based on a set of training recoded data, then the network are validated based on another set of data. Classification accuracy were then calculated and it have been found that the resulting total accuracy reached 74.94%.

1. INTRODUCTION

ARabic language are the 5th spoken language. Its importance also comes from that many non-Arabic origin people speaks this language. It is the language of Quran. All Muslims who reads Quran have to read it in this Language. Arabic phoneme recognition attracts many researchers in the world. Many techniques have been used for this purpose. Ahmad etal [1] employed neural network with Mile Frequency Cepstral Coefficient (MFCC), and Linear Predictive Coding (LPC) for the recognition of isolated Arabic words. Ali etal [2] used neural network with Fast Fourier Transform(FFT), Wavelet Transform (WT) and a combination of both for the recognition of Arabic phonemes. MFCC have also been adopted in [3] using a decision tree based on hierarchical neural networks. Fuzzy Systems have also been investigated for such applications. Debyeche etal [4] used fuzzy vector quantization for the Arabic speech recognition. Taleb and Benyettou [5] employed fuzzy neural systems and frequency parameters to identify Vowels. In [6] and [7] Hidden Markov Model (HMM) are employed for the Arabic phoneme and speech recognition. Fuzzy Neural Petri net have also been used for this purpose but with LPC

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

features only[8]. This paper investigates farther the use of Fuzzy Neural Network as a decision making logic but with MFCC as a feature extraction technique. In MFCC the features mimics the Human ear by adding a mile scale filter.

Figure (1) General Speech Recognition System Module

Figure (1) shows the block diagram of a phoneme classification system. At first the speech signal is recorded via a microphone. The recorded signal is then passed into a band pass filter to limit its frequency and avoids aliasing during sampling. The signal is then sampled and digitized using analogue to digital converter. After digitization the signal is segmented to extract and isolate the phonemes that have to be classified. Segmentation into phonemes avoids the recognition of huge amount of vocabularies exist in the language if isolated word recognition is to be used. The isolated phonemes are then feed into filtering to reduce noise and then the signal is scaled and unified in order to remove the effect of recording gain and speaker voice level variations. The features are then extracted from the phonemes signal and then feed into the decision making in order to classify the phoneme.

2.MEL FREQUENCY CEPSTRAL COEFFICIENT

The MFCC is the mostly used feature extraction technique in speech recognition system. This feature extraction technique was developed in 1940 by Stevens and Volkman and employed by Mermelstein and Davis in 1980 to extract the features of the speech signal [9]. This technique is based on the modeling of human auditory system. The study of human ear shows that its sensitivity to low frequency components of the speech signal are linear. If the frequency increases the ear sensitivity change to vary in logarithmic scale to the variation of audio signal frequency to capture only important features of the speech signal. This phenomena is implemented in MFCC scale through linear scale below 1kHz and logarithmic scale above 1kHz. Figure (2) shows the relation between the mile scale and frequency physical scale. This relation can be expressed mathematically as:

$$Mel\,(f) = 2595\,log10\,(1 + f/700) \qquad (1)$$

In which (f) represents the signal frequency

and the "Mel" represent the unit that measures the frequency or the perceived pitch or of a tone. Figure(3) shows the steps required for the calculation of MFCC. This calculation requires seven steps they are:

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

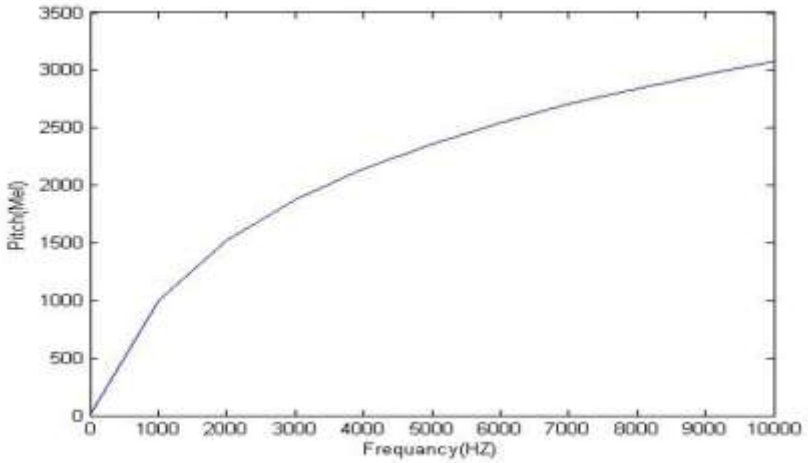

Fig. (2) Relation between Mel scale & physical scale

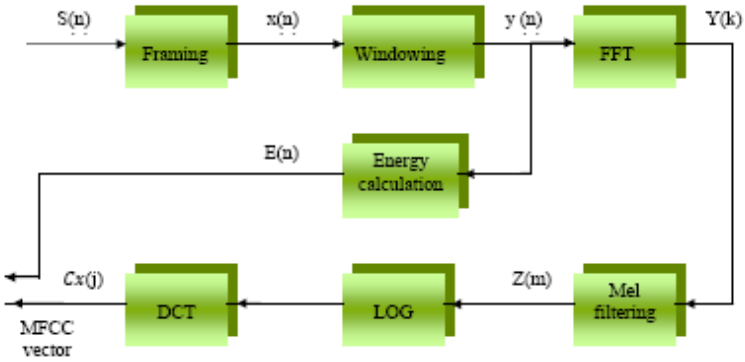

Fig.(3) The block diagram of MFCC calculation

A. Frame Blocking: Frame the speech signal into blocks of approximately (10 to 30 msec) in length (at sampling rate 16 kHz) over which the characteristics of the speech Signal are fairly stationary. Results of the framed signals (blocking into blocks of N sample) [10].

B. Windowing: The samples of speech within each frame are multiplied by a finite-duration window. The shape of duration of the window determines the time and frequency resolution that can be represented in the data. One of the most popular window function in ASR systems is the hamming window defined in Equation (2). The effect of widowing is shown in figure (4). [10].

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

$$w(n) = 0.54 - 0.45\cos(\frac{2\pi n}{N-1})$$
$$0 \leq n \leq N-1 \qquad (2)$$

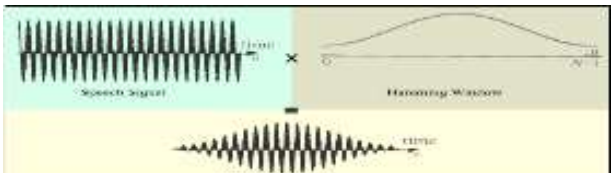

Fig.(4) Effect of window on signal

.

C. Fast Fourier Transform: After windowing, the power spectrum of each data frame is calculated. The most commonly used spectral estimation technique is the Fast Fourier Transform FFT. The vocal tract shape is revealed by the spectral analysis.

For the finite duration sequence {y(n)}, the Discrete Fourier Transform is defined as:

$$Y(k) = \sum_{n=0}^{N-1} y(n)e^{-j\pi kn/N} = \sum_{n=0}^{N-1} y(n)W^{nk}$$
In which $(0 \leq n, k \leq N-1)$ (3)

D. Mel Filter Bank Processing: This Mel filter function is to smooth the frequency spectrum, such that it closely model the human ear sensitivity. The Mel frequency scale calculation is performed using a set of band-pass filters varies between 13 and 24, depending on the bandwidth of the signals that are being processed. The filter output Z(m) can be given as[9]:

$$Z(m) = \sum_{k=0}^{\frac{k}{2}-1} |Y(k)|S_m(k)$$

$$m = 1 \ldots M \qquad (4)$$

In which M: is the number of the filter banks.

Sm(k): the triangular filter banks as shown in figure (5).

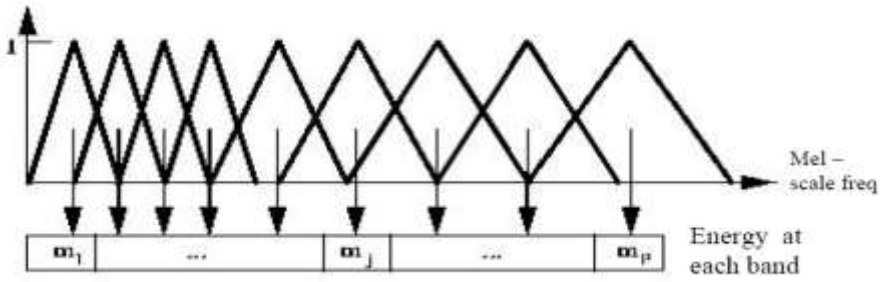

Fig. (5) Mel scale filter bank

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

E. Log Magnitude: After Mel –scale filtering the logarithm of the output of each filter bank is calculated. The logarithm can be regarded as a crude approximation of the non-linear intensity-loudness relationship associated with the human auditory system. An important advantage of this processing step is that it makes the shape of speech spectra insensitive to different loudness levels amplification factors in the linear energy domain become constant offsets in the log-energy domain and if these are properly compensated for, the resulting spectra are independent of differences in absolute loudness [11].

F. Discrete Cosine Transform or DCT: The discrete cosine transform (DCT) is applied to Mel-scaled log energy vectors in order to obtain features that are statically independent. The resulting Mel-scaled cepstral coefficients (MFCCs) are ordered such that higher order coefficients explain less variance [10,11].Because the higher order coefficient contain less information after the application of the DCT, the cepstral vector can be truncated to Contain fewer components than the number of filters in the Mel-scaled filter bank.

$$C_x(j) = \sum_{m=1}^{M} (\log(Z(m)) \cos\left[\frac{(2m-1)\pi j}{2M}\right]$$

$$j = 1,2,..J-1 \qquad (5)$$

Where $C_x(j)$: is the jth Mel Frequency Cepstrum Coefficients.

J: is the desired length of the cepstrum.

M: is the number of filter banks.

G. Energy Features: In addition to the features describing the spectral envelop, a caustic feature vectors usually also contain features that are related to the energy in the signal. The cepstral coefficient, c is often used as a measure of the mean log-energy in each frame. It is sometimes replaced by (or used in addition to) the log-energy that is directly computed from the signal at frame level (Log E) [10].

$$E_n = \sum_{-\infty}^{\infty}[x(m).w(n-m)]^2 \qquad (6)$$

where, x(n) represent the signal, w(n) represent a window function (such as Hamming window) and E_n is the short time energy at discrete time n.

Another time dependant sequence, which can serve as representation of speech signal is the short-time average zero-crossing rate. A simple technique to measure the frequency contents of a signal can be found by the rate at which zero crossing occurs. The short-time average zero-crossing can be given by the following Equation [10].

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

$$Z_n = \frac{1}{2N}\sum_{-\infty}^{\infty}|\text{sgn}[x(m)] - \text{sgn}[x(m-1)]w(n-m)| \quad (7)$$

$$\text{sgn}[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ 0 & x(n) < 0 \end{cases} \quad (8)$$

w (n): is a window function.

N: is the window length.

According to the MFCC algorithm, the performance of MFCC might be affected with the following factors: The first factor is number of the filters used in the filter bank. The second is the shape of the filters used (triangular, rectangular or Schroeder). The third factor is the spacing of these filters, whether they are overlapped or not. And the fourth factor is the way that power spectrum is warped.

3.FUZZY NEURAL PETRI NET

The fuzzy neural Petri net uses the Petri net stricture, embeds the fuzzyfication in the input place and employs the back propagation algorithm as a learning role. Figure (6) shows the basic structure of the fuzzy Petri net. In general Petri net consists of a number of places which are input, hidden or output places. These places are connected by arc and separated by transitions. In the fuzzy neural Petri net the input places are the input layer, the transitions are the hidden layer and the output places are the output layer.

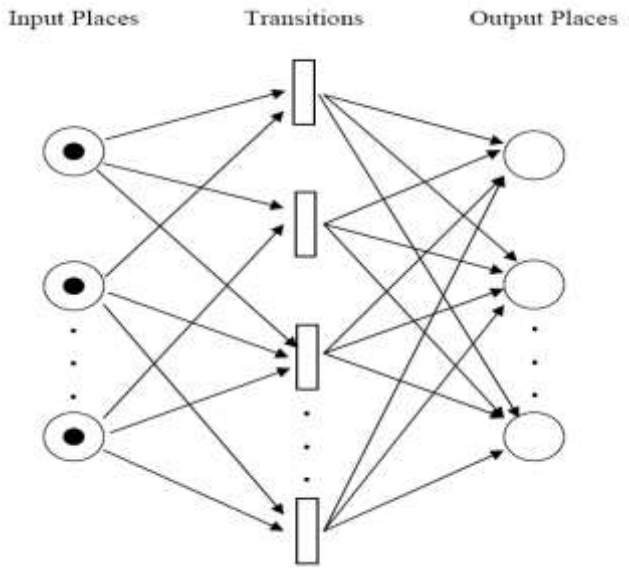

Fig. (6) The structure of the Neural Fuzzy Petri Net.

The features are the MFCCs are fuzzyfied and marked in the input places. The transition are fired based on the value calculated based on the weights connecting input places and

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

transitions and also based on threshold limit of each transition. The output places are marked based on the weights connecting transitions with the output place. Figure (7) shows a typical of the net showing the notation used. The equations relating different parts of the network are as follows:

$$x_j = f(input(j)) \tag{9}$$

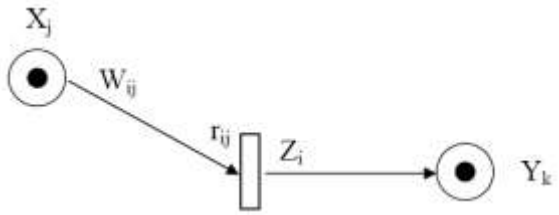

Fig.(7) One section of the Fuzzy Neural Petri Net

Where f is a triangular mapping function defined by

$$f(x) = \begin{cases} \dfrac{x - \min(x)}{average(x) - \min(x)}, & if \quad x < average(x) \\ \dfrac{\max(x) - x}{\max(x) - average(x)}, & if \quad x > average(x) \\ 1, & if \quad x = average(x) \end{cases} \tag{10}$$

$$Z_i = \overset{n}{\underset{j=1}{T}}[W_{ij} S(r_{ij} \rightarrow X_j)] \tag{11}$$

Wij is the weight between the i-th transition and the j-th input place;

- rij is a threshold level associated with the level of marking of the j-th input place and the i-th transition;

- Zi is the activation level of i-th transition

" T " is a t-norm, " S " denotes an s-norm, while $\rightarrow$ stands for an

Implication operation expressed in the form

$$a \rightarrow b = \sup\{c \in [0,1], aTc \leq b\} \tag{12}$$

Where a, b are the arguments of the implication operator confined to the unit interval.

If t-norm is defined as a multiplication operator (Π) then

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

$$r_{ij} \to x_j = \begin{cases} \dfrac{x_j}{r_{ij}}, & if \quad r_{ij} > x_j \\ 1, otherwise \end{cases} \tag{13}$$

$$Z_i = \prod_{j=1}^{n} W_{ij} \vee \begin{cases} \dfrac{x_j}{r_{ij}}, & if \quad r_{ij} > x_j \\ 1, otherwise \end{cases} \tag{14}$$

- Yk is the marking level of the k-th output place produced by the transition layer and performs a nonlinear mapping of the weighted sum of the activation levels of these transitions (Zi) and the associated connections Vki

$$Y_k = f\left(\sum_{i=1}^{No.of Transitions} V_{ki} Z_i\right), \qquad k = 1,2,...,m \tag{15}$$

Where " f " is a nonlinear monotonically increasing function from R to [0,1].

The parameter update is derived in a similar way to back propagation algorithm. The sum squared error is used as an optimization criteria. The update is based on minimizing this measure.

$$E = \frac{1}{2}\sum_{k=1}^{m}(t_k - y_k)^2 \tag{16}$$

tk is the k-th target;

yk is the k-th output.

The updates of the parameters are performed according to the gradient method

$$param(iter+1) = param(iter) - \alpha \nabla_{param} E \tag{17}$$

Where ∇ Eparam E is a gradient of the performance index E with respect to the network parameters, α is the learning rate coefficient, and iter is the iteration counter.

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

The nonlinear function associated with the output place is a standard sigmoid described as

$$y_k = \frac{1}{1+\exp(-\sum Z_i V_{ki})} \tag{18}$$

Hence, the parameter update is derived as follows:

1- $$\Delta V_{ki} = -\zeta \frac{\partial E}{\partial V_{ki}} \tag{19}$$

$$\frac{\partial E}{\partial V_{ki}} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial V_{ki}} \tag{20}$$

$$\frac{\partial E}{\partial y_k} = -(t_k - y_k) \tag{21}$$

$$\frac{\partial y_k}{\partial V_{ki}} = y_k(1-y_k)Z_i$$

$$= \frac{-\exp(-Z_i V_{ki})Z_i}{1+\exp(-Z_i V_{ki})} \tag{22}$$

$$\Delta V_{ki} = -\zeta \frac{\exp(-Z_i V_{ki})Z_i(t_k - y_k)}{1+\exp(-Z_i V_{ki})} \tag{23}$$

$$V_{ki}(iter+1) = V_{ki}(iter) - \alpha \Delta V_{ki} \tag{24}$$

for k=1,2,…,m, i=1,2,…, no. of transitions

2- $$\Delta r_{ij} = -\zeta \frac{\partial E}{\partial r_{ij}} \tag{25}$$

$$\frac{\partial E}{\partial r_{ij}} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial Z_i}\frac{\partial Z_i}{\partial r_{ij}} \tag{26}$$

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

$$\frac{\partial E}{\partial y_k} = -(t_k - y_k) \tag{27}$$

$$\frac{\partial y_k}{\partial Z_i} = y_k(1-y_k)V_{ki} \tag{28}$$

$$Z_i = \sum_{j=1}^{n}[W_{ij}S(r_{ij} \to x_j)] \tag{29}$$

$$\frac{\partial Z_i}{\partial r_{ij}} = A\frac{\partial}{\partial r_{ij}}(W_{ij} + (r_{ij} \to x_j) - W_{ij}(r_{ij} \to x_j))$$

$$= A(1-W_{ij})\frac{\partial}{\partial r_{ij}}(r_{ij} \to x_j) \tag{30}$$

Where

$$A = \sum_{\substack{\lambda=1 \\ \lambda \neq j}}^{n}[W_{i\lambda}S(r_{i\lambda} \to x_\lambda)] \tag{31}$$

$$\frac{\partial}{\partial r_{ij}}(r_{ij} \to x_j) = \frac{\partial}{\partial r_{ij}}\begin{cases}\frac{x_j}{r_{ij}}, if \quad r_{ij} > x_j \\ 1, otherwise\end{cases} = \begin{cases}\frac{-x_j}{r_{ij}^2}, if \quad r_{ij} > x_j \\ 0, otherwise\end{cases} \tag{32}$$

$$\Delta r_{ij} = -\zeta(t_k - y_k)\frac{\exp(-Z_iV_{ki})V_{ki}}{1+\exp(-Z_iV_{ki})}\sum_{\substack{\lambda=1 \\ \lambda \neq j}}^{n}[W_{i\lambda}S(r_{i\lambda} \to x_\lambda)](1-W_{ij})\begin{cases}\frac{-x_j}{r_{ij}^2}, if \quad r_{ij} > x_j \\ 0, otherwise\end{cases}$$

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

$$= \Delta V_{ki} * \prod_{\substack{\lambda=1 \\ \lambda \neq j}}^{n} W_{i\lambda} \vee \left\{ \begin{array}{l} \frac{x_j}{r_{i\lambda}}, \quad if \quad r_{i\lambda} > x_\lambda \\ 1, otherwise \end{array} \right\} * (1 - W_{ij}) \left\{ \begin{array}{l} \frac{-x_j}{r_{ij}^2}, if \quad r_{ij} > x_j \\ 0, otherwise \end{array} \right.$$

(33)

$$r_{ij}(iter+1) = r_{ij}(iter) - \alpha \Delta r_{ij} \tag{34}$$

3- $$\Delta W_{ij} = -\zeta \frac{\partial E}{\partial W_{ij}} \tag{35}$$

$$\frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial Z_i} \frac{\partial Z_i}{\partial W_{ij}}$$, k=1,2,…,m, i=1,2,…, hidden, j=1,2,…, n.

(36)

$$\frac{\partial E}{\partial y_k} = -(t_k - y_k) \tag{37}$$

$$\frac{\partial y_k}{\partial Z_i} = y_k(1 - y_k)V_{ki}$$

$$= \frac{-\exp(-Z_i V_{ki})V_{ki}}{1 + \exp(-Z_i V_{ki})} \tag{38}$$

$$\frac{\partial Z_i}{\partial W_{ij}} = A \frac{\partial}{\partial W_{ij}} (W_{ij} + (r_{ij} \to x_i) - W_{ij}(r_{ij} \to x_j))$$

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

$$= A(1-(r_{ij} \rightarrow x_j)) \tag{39}$$

$$A = \prod_{\substack{\lambda=1 \\ \lambda \neq j}}^{n} [W_{i\lambda} S(r_{i\lambda} \rightarrow x_\lambda)]$$

$$(r_{ij} \rightarrow x_j) = \begin{cases} \dfrac{x_j}{r_{ij}}, & if \quad r_{ij} > x_j \\ 1, otherwise \end{cases} \tag{40}$$

$$\Delta W_{ij} = -\zeta(t_k - y_k) \frac{\exp(-Z_i V_{ki}) V_{ki}}{1+\exp(-Z_i V_{ki})} \prod_{\substack{\lambda=1 \\ \lambda \neq j}}^{n} [W_{i\lambda} S(r_{i\lambda} \rightarrow x_\lambda)](1-(r_{ij} \rightarrow x_j))$$

$$= \Delta V_{ki} * \prod_{\substack{\lambda=1 \\ \lambda \neq j}}^{n} W_{i\lambda} * \begin{Bmatrix} \dfrac{x_j}{r_{i\lambda}}, & if \quad r_{i\lambda} > x_\lambda \\ 1, otherwise \end{Bmatrix} * (1 - \begin{Bmatrix} \dfrac{-x_j}{r_{ij}^2}, if \quad r_{ij} > x_j \\ 0, otherwise \end{Bmatrix}) \tag{41}$$

$$W_{ij}(iter+1) = W_{ij}(iter) - \alpha \Delta W_{ij} \tag{42}$$

Hence, the update of parameter learning is performed using equations (23 and 24) for V_{ki} equations (33 and 34) for the update of r_{ij}, and equations (41and 42) for the update of W_{ij}

4. EXPERMENTAL RESULTS

In the recognition system proposed her, 33 Neural Fuzzy Petri Nets NFPN are used. One network for each phoneme. 23 input places are used for each network, corresponding to the input feature for MFCC. The feature vector is first fuzzified, then used as input to each of the NFPN the number of transitions was selected between 22 to 33 based on the phoneme. Each NFPN has one output place. Twelve cases are token for each phoneme and applied to the network. The network is trained such that its output place is one corresponding to the phoneme that it represents as zero otherwise. (1*10-5) error goal was selected as stopping criteria for the training phase of the network. The network is then tested using the twelve set for each

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

phoneme (testing phonemes) as input to the network and total recognition accuracy is calculated to be as:

$$\text{phoneme classification accuracy} = \frac{\text{Number of correctly classified phonemes}}{\text{Total number of test phonemes}} * 100\%$$

Table (1) shows the resulting total classification accuracy obtained in the testing phase. As can be seen from the results, some of the phonemes classification reaches 100%. However, others have poor classification. The total classification accuracy for all phonemes are 74.93%.

5.CONCLUSIONS

This work proposes Arabic phoneme classification technique uses MFCC and FNPN. The proposed system was tested and it have been found that the total classification accuracy is 74.93% comparing the results with that obtained in [8] that uses linear predictive coding which have accuracy of 71% and most of the phonemes, have improved accuracy.

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Table (1) Classification accuracy

NO	Phonemes	IPA symbols	Phoneme recognition accuracy%
1	ء	/ʔ/	52.1717
2	ب	/b/	66.667
3	ت	/t/	75
4	ث	/θ/	75
5	ج	/d3/	41.667
6	ح	/ħ/	91.667
7	خ	/x/	58.33
8	د	/d/	83.33
9	ذ	/ð/	41.667
10	ر	/r/	83.33
11	ز	/z/	87.878
12	س	/s/	100
13	ش	/ʃ/	100
14	ص	/s/	91.667
15	ض	/d̪/	66.667
16	ط	/t̪/	91.667
17	ع	/ʕ/	91.667
18	غ	/ɣ/	91.667
19	ف	/f/	75
20	ق	/q/	58.33
21	ك	/k/	62.056
22	ل	/l/	78.5
23	م	/m/	50
24	ن	/n/	50
25	ه	/h/	81.73
26	و	/w/	83.33
27	ي	/j/	50
28	ا علة	/a:/	91.92
29	و علة	/u:/	83.33
30	ي علة	/i:/	100
31	فتحة	/a/	75
32	ضمة	/u/	75
33	كسرة	/i/	68.44
Total accuracy %			74.93

QALAAI ZANIST JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil – Kurdistan – Iraq
Vol. (2), No. (2), April 2017
Speical Issue : The 1st International Conference on Information Technology (ICoIT'17)
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

References:

[1] A. M. Ahmad, S. Ismail and D. F. Samaon, "Recurrent neural network with back propagation through time for speech recognition," IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004., , pp. 98-102 vol.1. 2004.

[2] A. A. Ali, M. A. Alwan, and A. A. Jasim. "Hybrid Wavelet-Network Neural/FFT Neural Phoneme Recognition." Proceedings of The 2nd International Conference on Information Technology. pp. 39-47, Amman Jordan. 2005.

[3] Abduladhem A. Ali and I. T. Hwaidy "Hierarchical Arabic Phoneme Recognition Using MFCC Analysis". Iraq J. Electrical and Electronic Engineering, vol.3, pp.97-106, 2007.

[4] M. Debyeche, A. Amrouche and J. P. Haton, "Distributed TDNN-Fuzzy Vector Quantization for HMM speech recognition," *2009 International Conference on Multimedia Computing and Systems*, Ouarzazate, pp. 72-76, 2009.

[5] A. Taleb and A. Benyettou, "Arabic Vowels Fuzzy Neural Network Recognition". Journal of Applied Sciences, vol.10, pp.848-851, 2010.

[6] A. A. Abushariah et al. "Arabic speaker-independent continuous automatic speech recognition based on a phonetically rich and balanced speech corpus." International Arab Journal of Information Technology (IAJIT) vol.9, no.1, pp.84-93, 2012.

[7] Gadeed, Ashwag, and Talaat Wahbi. "The Recognition of Holy Quran Reading types "Rewaih"." International Journal of Advanced Research in Computer Science vol.5. no.3,pp.37-40, 2017.

[8] Ghassaq S. Mosa, and Abduladhem Abdulkareem Ali. " Arabic Phoneme Recognition Using Neural Fuzzy Petri Net and Lpc Feature Extracting." The International Arab Conference on Information Technology (ICIT-2009).

[9] R. M. Hegde, "Fourier Transform Phased-Based Features for Speech Recognition," Ph.D. Thesis, Department of Computer Science Engineering, Institute of Technology, India, 2005.

[10] L. Rabinar and R.W. Schafar "Fundamental of Speech Recognition ",Prentice Hall, 1993.

[11] H. Xie ,"Speech Analyzer in an ICAI System for TESOL ",M.Sc. Thesis, Computer science, Victoria University of Wellington, 2004.