ISSN: 2277-9655
Impact Factor: 4.116

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

TAJWEED UTOMATION SYSTEM USING HIDDEN MARKOUV MODEL AND NURAL NETWORK

Safaa Omer Mohammed Nssr*, Hoida Ali Abdelgader
SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY FACULTY OF COMPUTER SCIENCE, SUDAN

DOI: 10.5281/zenodo.57006

ABSTRACT

Voice recognition is considered as one of the most important aspects of machine learning and artificial intelligence engineering domain. But it still has limited and modest applications in Arabic language. The Holy Quran is the largest container of Arabic language grammar in terms of speaking and utterance as it is considered as a message for all humanity. However, we present within this study a classification model for four different altajweed rules like the Allah name(mofakham, morakaq) and moon and sun L(لام),as we depended on three different kinds of voice features LPC(liner predictive coding),MFCC(Mel-frequency cepstrum),FFT(Fast Fourier transform),where those three types of features are the most used within the domain of processing voice signaldomain.as we depended on two classifying mechanisms (neural networks and hidden Markov model(HMM)) in order to study all possible cases of those studied rules, then we extracted those features of two different readers(males).each of Markov hidden model and neural networks have been trained by using three different types of extracted features and then we tested those trained models in order to obtain final results as to evaluate them.

MATLAB version (0.8.1) , Audacity 2.1.2 cutting the samples voices .will be used to implement this concept to achieve further understanding.

KEYWORDS: Speech recognition, , MFCC , LPC , FFT, (NN) ,(HMM))

INTRODUCTION

Computer scientists have shown interest since more than four centuries in order to make humans capable of communicating with computers. The need to communicate with computers have emerged due to the increase of computer development, besides human is still in continuous search to find the simplest methods to communicate with it, speech recognition was the main interest of those studies for more than four centuries, along with the appearance of computing science

And digital signals, this technique have become most common because that technique that has the feature of speech recognition is more usable. Speech recognition apps have developed recently and entered many daily life domains and since the Holy Quraan is one of the higher sciences,thus many have shown interest in tajweed the holy Quraan since it came to Allah messenger(peace be upon him) as to memorize it by many of the messenger's followers (may Allah be pleased with them),where many of the messenger's followers took on their shoulders the mission of saving and writing and learning it to the new generation of followers(may Allah be pleased with them).scientists studied it where some of them studied phonetics phenomenon's and set rules for altajweed science

MEL FREQUENCY CEPSTRAL COEFICIENTS PROCESSOR (MFCC):

MFCC's is a type of algorithm i.e. basically used to define relationship between human ear's critical bandwidths with frequency. This method is basically used for analyzing and extraction of pitch vectors. [3]

Figure:

ISSN: 2277-9655
Impact Factor: 4.116

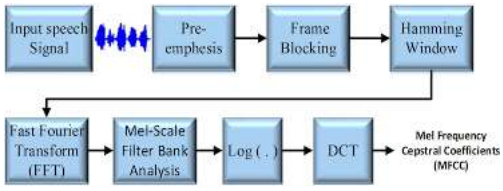

Fig. 2: MFCC Block Diagram

As shown in Figure 2, MFCC consists of seven computational steps. Each step has its function and mathematical. Figure 3 represents Extraction of MFCC Feature for a Frame. [3]

Figure:

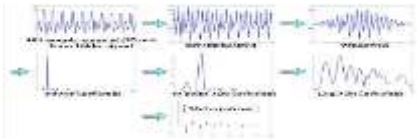

Fig 3:. Extraction of MFCC Feature for a Frame

(a) Pre-emphasis

The speech signal x(n) is sent to a high-pass filter

Formulae:

$y(n) = x(n) - a*x(n-1)$ (1)

Where s2(n) is the output signal and the value of a is typically between 0.9 and 0.99.

. (b) Frame blocking

The input speech signal is segmented into frames of 20~30 ms. Usually the frame size (in terms of sample points) is equal to power of two in order to facilitate the use of FFT. [3]

Figure:

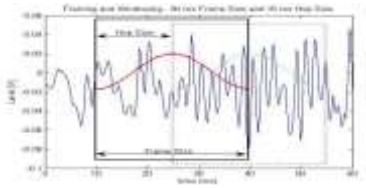

Fig 4:. Framing and overlapping

. (c) Hamming windowing

Each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame. If the signal in a frame is denoted by x(n), n = 0,…N-1, then the signal after Hamming windowing is

Formulae:

x(n)*w(n), where w(n) is the Hamming window defined by:

$w(n, a)=(1- a)- a \cos(2pn/(N-1)), \quad 0 \leqq n \leqq N-1$ (2) [4].

Figure:

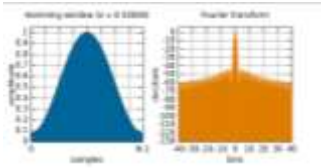

Fig 5: Hamming window and its frequency spectrum

ISSN: 2277-9655
Impact Factor: 4.116

(d) Fast Fourier Transform or FFT

Spectral analysis of different pitches in speech signals corresponds to different energy distribution on frequency scale.

Therefore FFT is used to obtain the magnitude frequency response of each frame.

(e) Mel-frequency wrapping

Human perception of frequency contents of sounds for speech signal does not follow a linear scale. Therefore for eacht one with an actual frequency is measured in Hz,. we can use the following approximate formula to compute the mels for a given frequency f in Hz.

Formulae:

Mel(f) = 2595*log10(1 + f/700) (3)

The mel scale filter bank is a series of triangular band pass filters that have been designed to simulate the band pass Filtering believed to occur in the audible system.. [4].

Figure:

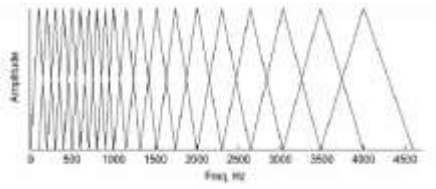

Fig 6: Filter bank in Mel frequency scale

(f) Discrete cosine transform or DCT

In this step apply DCT on the 20 log energy Ek obtained from the triangular band pass filters to have L mel-scale Cepstral coefficients.

Formulae:

Cm = Sk=1

N cos [m*(k-0.5)*p/N]* Ek, m=1,2, ..., L (4)

Where N is the number of triangular band pass filters, L is the number of Mel Scale Cepstral Coefficients. Set N=20 and L=12 performed FFT. DCT converts the frequency domain into a time domain called quefrency domain. The obtained features are same as cepstrum, thus it is referred to as the mel-scale Cepstral coefficients (MFCC). [7]

2.2 LINEAR PREDICTIVE CODING (LPC)

A very powerful method for speech analysis is based on linear predictive coding (LPC), also known as LPC analysis or auto-regressive (AR) modeling. This method is widely used because it is fast and simple, yet an effective way of estimating the main parameters of speech signals. All-pole filter with a sufficient number of poles is a good approximation for speech signals. Thus, we could model the filter H(z) in Figure 6.15 as

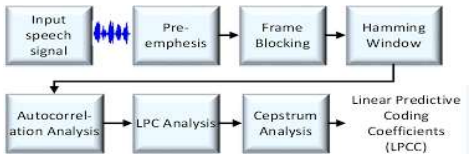

Fig 7:. Block diagram for LPC processor for Speech Recognition

(a)Pre – emphasis:

From the speech production model it is known that the speech undergoes a spectral tilt of -6dB/Oct. To counteract this fact a pre-emphasis filter is used. The main goal of the pre-emphasis filter is to boost the higher frequencies in order to flatten the spectrum. Pre emphasis follows a 6 dB per octave rate. This means that as the frequency doubles, the amplitude increases 6 dB. This is usually done between 300 - 3000 cycles. Pre emphasis is needed in FM to maintain good signal to noise ratio. Perhaps the most widely used pre emphasis network is the fixed first-order system: [4].

Formulae:

$H(z) = 1 - \alpha z^{-1}, \quad 0.9 \le \alpha \le 1.0$ (1)

ISSN: 2277-9655
Impact Factor: 4.116

(b) Frame – Blocking:

In this step the pre-emphasized speech signal is blocked into frames of N samples, with adjacent frames being separated by M samples .Thus frame blocking is done to reduce the mean squared predication error over a short segment of the speech wave form. In this step the pre emphasized speech signal, S(n) is blocked into frames of N samples, with adjacent frames being separated by M samples[4].

Figure:

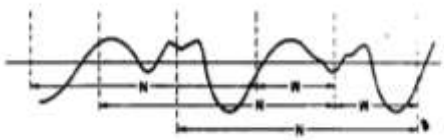

Fig 8. Blocking of speech into overlapping frames

Typical values for N and M are 256 and 128 when the sampling rate of the speech is 6.67 kHz. These correspond to 45-msec frames, separated by 15-msec, or a 66.7-Hz frame rate. [4].

. (c) Windowing:

Here we want to extract spectral features of entire utterance or conversation, but the spectrum changes very quickly. Technically, we say that speech is a non-stationary signal, meaning that its statistical properties are not constant across time. Instead, we want or extract spectral features from a small window of speech. [4].

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \quad (2.2)$$
$$0, \quad Else\ where \qquad (2)$$

(d) Autocorrelation analysis:

Each frame of windowing signal is next auto correlated to give

$$r_l(m) = \sum_{n=0}^{N-1-m} \bar{x}_l(n)\bar{x}_l(n+m), m = 0,1,2,\ldots\ldots\ldots,P, \qquad (3)$$

Where the highest autocorrelation value, p, is the order of the LPC analysis. Typically, values of p from 8 to 16 have been used, with p = 10 being the value used for this systems. A side benefit of the autocorrelation analysis is that the zeros autocorre$^{R_l(0)}$on,$R_l(0)$, is the energy, l^{th}he, l^{th} frame.

(e) LPC Analysis:

The next processing step is the LPC analysis, which converts each frame of P+1 autocorrelations into an LPC parameter set in which the set might be the LPC coefficients, [1]

(g) LPC parameter conversion to Cepstral coefficients:

A very important LPC parameter set, which can be derived directly from the LPC coefficients set, is the LPC Cepstral coefficients $c(m)$, recursion method is used. The Cepstral coefficients, which are the coefficients of the Fourier transform representation of the log magnitude spectrum [1]

2.3 Fast Fourier Transform or FFT

Fast Fourier transform (FFT) is an efficient implementation of the discrete fourier transform (DFT) of all the discrete transforms,(DFT) is most widely used in digital signal processing DFT maps a sequence x(n) into the frequency domain , the Fourier transform of x(t) and is represented by X(ω), that is,

$$X(\omega) = \mathcal{F}[x(t)] = \int x(t)e^{-j\omega t}dt$$

where $\mathcal{F}$ is the Fourier transform operator.

The Fourier transform of a signal x(t) is the integration of the product of x(t) and

Over the interval from $-\infty$ to $+\infty$.

$e^{-j\omega t}$ X(ω) is an integral transformation of x(t) from the time-domain to the frequency domain and is generally a complex function. X(ω) is known as the spectrum of x(t).[10][11]

ISSN: 2277-9655
Impact Factor: 4.116

CLASSIFICATION METHODS

There are two major types of models for classification: stochastic models (parametric) and template models (non-parametric) .

In stochastic models, the pattern matching is probabilistic (evaluating probabilities) and results in a measure of the likelihood, or conditional probability, of the observation given the model. Here, a certain type of distribution is fitted to the training data by searching the parameters of the distribution that maximize some criterion. Stochastic models provide more flexibility and better results. [4].

3.1 Hidden Markov Model (HMM)

In speech proceeding, both deterministic and stochastic models have had good success, one type of stochastic model, namely hidden Markov model (HMM) .(these models are referred to as Markov source s or probabilistic functions of markov chains in the communications literature .[6]

Until now, this is the most successful and most used

pattern recognition method for speech recognition.

.[10]

3.2 Basic Concepts

A Hidden Markov Model is a collection of states connected by transitions, as illustrated in Figure 9. It begins in a designated initial state. In each discrete time step, a transition is taken into a new state, and then one output symbol is generated in that state. The choice of transition and output symbol are both random, governed by probability Distributions.

Figure:

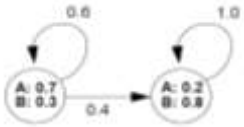

Fig9: A simple Hidden Markov Model, with two states and two output symbols, A and B.

Speech always goes forward in time, transitions in a speech application always go forward (or make a self-loop, allowing a state to have arbitrary duration). Figure 10 illustrates how states and transitions in an HMM can be structured hierarchically,. .[5]

Figure:

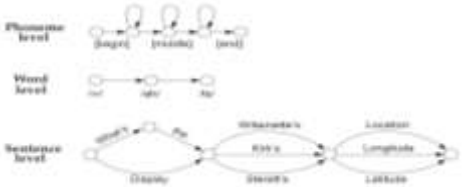

Fig10: illustrates how states and transitions in an HMM can be structured hierarchically

In using this notation we implicitly confine our attention to First-Order HMMs, in which a and b depend only on the current state, independent of the previous history of the state sequence. This assumption, almost universally observed, limits the number of trainable parameters and makes the

3.3 The forward algorithm

Formulae:

Step 1: Initialization

$$\alpha_1(i) = \pi_i b_i(X_1) \qquad 1 \le i \le N$$

Step 2: Induction

$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij}\right] b_j(X_t) \qquad 2 \le t \le T;\ 1 \le j \le N$$

Step 3: Termination

$$P(\mathbf{X}|\Phi) = \sum_{i=1}^{N} \alpha_T(i) \qquad \text{If it is required to end in the final state, } P(\mathbf{X}|\Phi) = \alpha_T(s_T)$$

Figure:

ISSN: 2277-9655
Impact Factor: 4.116

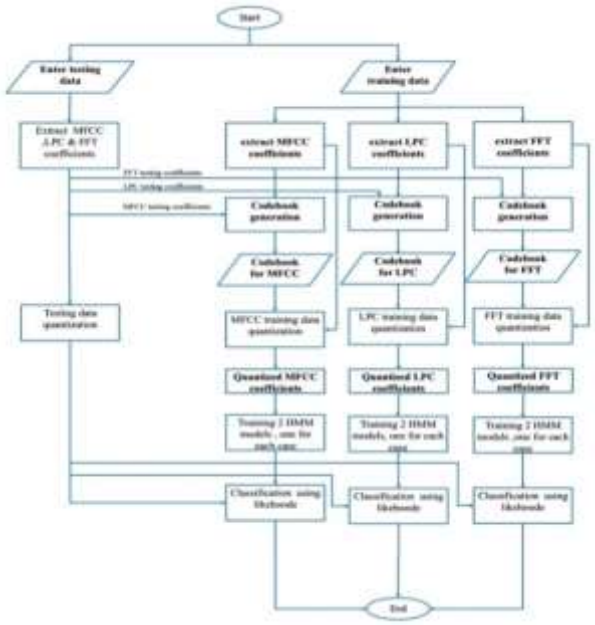

Fig 12: diagram represents the process of training hidden markov models HMM of classification data models

3.4 Neural networks

Artificial neural networks are relatively crude electronic networks of neurons based on the neural structure of the brain. They process records one at a time, and learn by comparing their classification of the record (i.e., largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm for further iterations. Roughly speaking, a neuron in an artificial neural network is

1. A set of input values (xi) and associated weights (wi).
2. A function (g) that sums the weights and maps the results to an output (y).[7]

Figure:

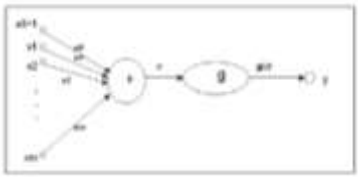

Fig 12. Neurons in Neural network

Pattern matching problem that is amenable to neural networks; therefore we use neural networks for acoustic modeling, while we rely on conventional Hidden Markov Models for temporal modeling, two different ways to use neural networks for acoustic modeling, namely prediction and classification of the speech patterns. Prediction is shown to be a weak approach because it lacks discrimination, while classification is shown to be a much stronger approach..[8]

Neural networks have many similarities with Markov

Models. Both are statistical models which are represented as graphs. Where Markov models use probabilities for state transitions, neural networks use connection strengths and functions. A key difference is that neural networks are fundamentally parallel while Markov chains are serial. Frequencies in speech, occur in parallel, while syllable series and words are essentially serial. .[9]

3.5 Patternnet

Syntax : patternnet (hidden sizes, trainfcn)

Pattern recognition networks are feed forward networks that can be trained to classify inputs according to target classes .the target data for pattern recognition networks should consist of vectors of all zero values except for a1 in

ISSN: 2277-9655
Impact Factor: 4.116

element I, where is the class they are to be represent , and takes this arguments, hidden sizes : Row vector of one or more hidden layer sizes(default =10), trainfcn: training function(default =train scg) , and return pattern recognition neural network .

Figure:

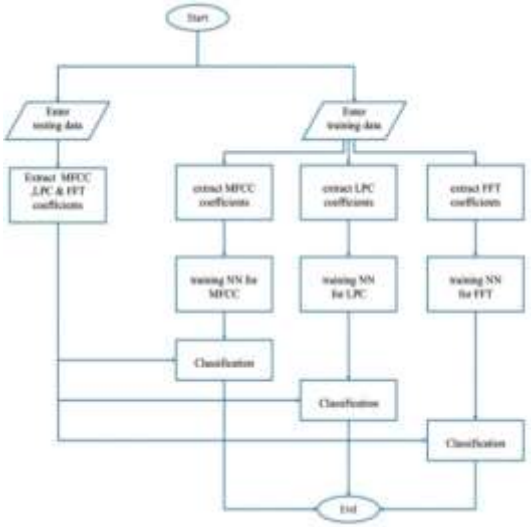

Fig 1.3: diagram represents the process of training neural networks for the purpose of uttering each of the holy name and (اللام)

DATA SET

There are thirteen cases of sun L(لام), ten cases of moon sun L(لام), and three cases for each of the Allah name(mofakham, morakaq) in the altajweed rules In The Holy Quran We have collected from register voices for two readers (Al sudis and Al huzafi), and we cut it with audacity program , the size of data set for Al sudes reader it's 40 samples (moon / sun), and 30 samples for Allah (mofakham, morakaq),and Al huzafi reader 50 samples for (moon / sun) , 50 samples for Allah (large / small) it was done by a Audacity program. , selected 50% of the sample data to training the system, and 50% of the sample data for testing the system.

RESULTS AND DISCUSSION

We choose MATLAB version (0.8.1) as our programming environment as it offers many advantages. It contains a variety of signal processing and statistical tools, which help users in generating a variety of signals and plotting them. And through this special education system of altajweed which can classify the holy name into minimized and maximized names, as to classifying the لام into sunny or moony one. We used hidden Markov model and neural network to recognize these rules, through them have been training, analyze the words, these program serves a lot in the areas of education and language schools and teach the Quran and others.

The following is samples of altajweed words:

ISSN: 2277-9655
Impact Factor: 4.116

Figure:

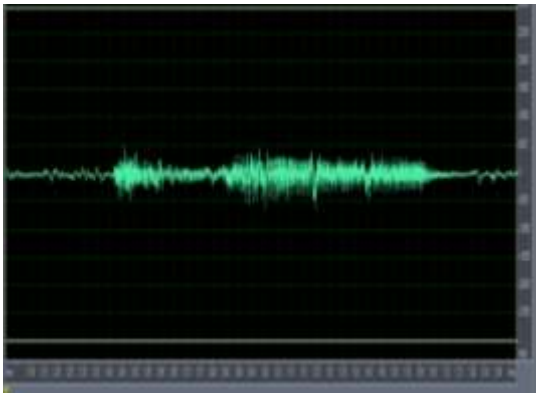

Fig 14: Signal for Allah name (mofakham)

:

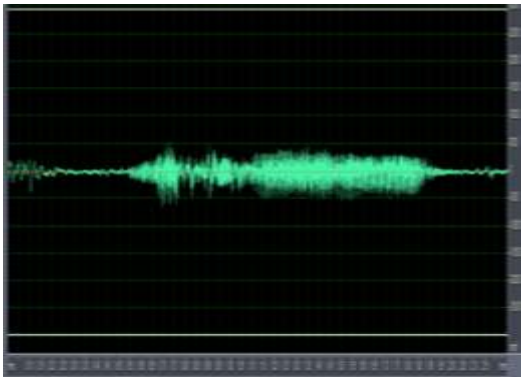

Fig 15: Signal for Allah name (morakaq)

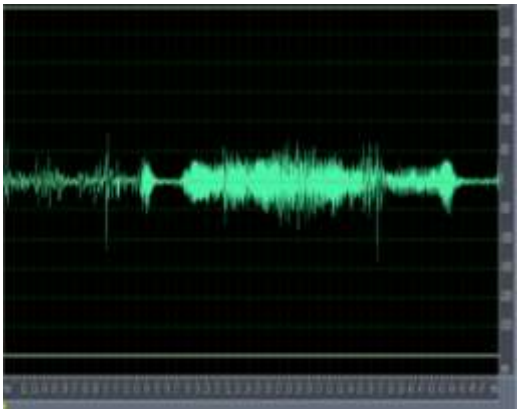

Fig 16: Signal for sunny L (لام)

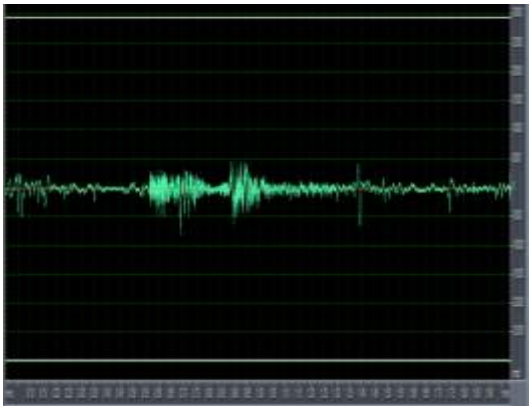

Fig 17 Signal for moony L (لام)

For 20 Large testing sample and 15 small testing sample for al Sudes reader using hidden Markov model

ISSN: 2277-9655
Impact Factor: 4.116

Tables:

Table1 for Allah large and small with multi value for Number of quantization levels and number of states:

For 25 moon testing sample and 25 sun testing sample for al Sudes reader using hidden Markov model

Reader	Number of quantization levels=32									
Sudes	Num_stats		3	4	5	6	7	8	9	10
	MFCC	large	100	95	100	100	100	100	100	100
		small	46.667	100	100	100	100	100	100	100
	LPC	large	45	75	90	90	90	95	95	85
		small	0	0	0	0	0	0	100	0
	FFT	large	70	100	90	100	90	100	100	95
		small	0	0	0	0	0	0	0	0

Table2 A moon and sun for sudes with multi values for number of quantization levels and number of states:

Reader	Number of quantization levels=32									
Sudes	Num_stats		3	4	5	6	7	8	9	10
	MFCC	moon	63	88	100	100	100	100	100	100
		sun	46	88	100	100	100	100	100	100
	LPC	moon	64	76	92	92	88	100	92	96
		sun	0	0	0	52	0	0	0	0
	FFT	moon	44	92	88	88	92	92	92	96
		sun	100	100	0	0	100	0	0	100

Allah: 20 large testing samples. , 15 small testing sample and A: 25 moon testing sample. 25 sun testing sample

Table3 Results for ANN for al_sudes:

Reader	Number of quantization levels=32									
Huzaifi	Num_stats		3	4	5	6	7	8	9	10
	MFCC	Large	95	100	100	100	100	100	100	100
		Small	50	100	100	100	100	100	100	100
	LPC	Large	70	100	100	85	100	85	60	100
		Small	90	0	0	0	0	0	0	0
	FFT	Large	60	90	95	95	100	100	100	100
		Small	0	0	10	10	0	10	0	0

For 20 large testing sample and 10 small testing sample for al huzafi reader using hidden Markov model

Table4 for Allah large and small with multi value for Number of quantization levels and number of states:

Tested case		MFCC	LPC	FFT
Allah	large	100	95	95
	small	100	46.667	73.333
A	moon	92	72	80
	sun	80	68	84

ISSN: 2277-9655
Impact Factor: 4.116

For 10 moon testing sample and 10 sun testing sample for al huzafi reader using hidden Markov model

Table5 A moon and sun for al huzafi with multi values for number of quantization levels and number of states:

Reader	Number of quantization levels=32									
Huzaifi	Num_stats		3	4	5	6	7	8	9	10
	MFCC	moon	50	100	100	100	100	100	100	100
		sun	40	100	100	100	100	100	100	100
	LPC	moon	60	80	100	90	100	90	100	100
		sun	10	10	0	0	0	0	0	0
	FFT	moon	60	100	100	100	100	100	100	100
		sun	60	0	0	0	40	40	0	0

Allah: 20 large testing samples. , 10 small testing sample and A: 10 moon testing sample. 10 sun testing sample

Table6 Results for ANN for al_huzafi:

Tested case		MFCC	LPC	FFT
Allah	large	100	100	95
	small	100	70	70
A	moon	90	70	70
	sun	90	60	90

CONCLUSIONS

The paper presents a special education system of altajweed which can classify the holy name Allah into minimized and maximized names, as to classifying the لام into sunny or moony one.

We depended on three different kinds of extractions features LPC (liner predictive coding), MFCC (Mel-frequency cepstrum), and FFT (Fast Fourier transform), where those three types of features of speech in a simple and efficient way,the algorithm shows good results in classifying the speech we depended on two classifying mechanisms (neural networks(NN) and hidden Markov model (HMM)),It presents the comparison among these techniques From tables it was observed that recognition rate of Allah's name mofakham(maximized), Allah's name morakaq(minimized) for two readers with hidden markov and MFCC the rate recognition increasing , compared with the neural networks ,FFT and LPC, the recognition rates found to be decreased These is a scope in increasing the recognition rate of hidden markov. is found to be efficient, we can observed that recognition of moony and sunny L (لام),found to be decreased among all techniques because the samples of moony and sunny L (لام) is too short .

In many speech recognition system using modern microphones to eliminate noises but expensive, in this work we have been take registration Quran voices with high quality equipment's in studio's for registration we faced problems with old data especially with moony and sunny L (لام), its two short and there was some noise in recording, so less recognition rate in the results , for that we use data more efficient recorded by modern mice and special place, we achieved better results .

ACKNOWLEDGEMENTS

I extend my sincere thanks and gratitude and appreciation to all those who contributed effort, or gave me help to accomplish this search. I especially thank the Sudan University of science For giving me chance to study with them

REFERENCES

[1] Lawrence Rabiner , "Fudementals Of Speech Recognition, ", Inc Asimon & Schuster Company,1993., PP 96–102

[2] Thomas F. Quatieri,N. Chandrasekaran,Ting-Peng Liang, " Discrete Time Speech Signal Processing Principles and Practice", : Pearson Education, Academic Texts, 2003, PP 55–62

[3] Bidoor Noori Ishaq, Bharti W. Gawali, " Comparative Analysis of MFCC, DTW&ANN for Arabic Speech Recognition" INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH International

ISSN: 2277-9655
Impact Factor: 4.116

Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163,VOL 1 ,Issue 11, NOVEMBER 2014 , PP 57.

[4] Anu L B , Dr Suresh D , Sanjeev kubakaddi, " Person Identification using MFCC and Vector Quantization" IPASJ INTERNATIONAL JOURNAL OF ELCTRONICS & COMMUNICATION (IIJEC) Volume 3, Issue 6, June 2015, PP 20.

[5] Joe Tebelskis, "Speech Recognition using Neural Networks", Carnegie Mellon University , Pittsburgh, Pennsylvania 15213-3890, MAY 1995, PP 16–18

[6] Dr.Raj Reddy, "Spoken Language Processing", PTR Prentice = Hall, 211, PP 383–385.

[7] Dr.S.P.Victor, C.RajKumar, , "Modular Implementation of Neural network Structures in Marketing Domain Using Data Mining Technique", INTERNATIONAL JOURNAL OF ENGINEERING AND COMPUTER SCIENCE ,ISSN: 2319-7242 VOL 5 ISSUE 1,JANUARY 2016, PP. 15624–15630

[8] Joe Tebelskis, "Speech Recognition using Neural Networks", Carnegie Mellon University , Pittsburgh, Pennsylvania 15213-3890, MAY 1995, PP 7

[9] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, Senior Member, 'Neural Networks used for Speech Recognition', JOURNAL OF AUTOMATIC CONTROL, UNIVERSITY OF BELGRADE, VOL. 20:1-7, 2010

[10] MATTHEW N. O. SADIKU,WARSAME H. ALI'signals and systems',CRC Press, 2016 , PP 222.

[11] K.R.Rao,D.N.Kim,J.J.H, , "Fast Fourier transforms", CRC Press, PP 1