DOI: http://dx.doi.org/10.26483/ijarcs.v11i2.6519

Volume 11, No. 2, March-April 2020

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

ARABIC SPEAKER RECOGNITION SYSTEM USING GAUSSIAN MIXTURE MODEL AND EM ALGORITHM

Dr. Mowaffak O. A. Albaraq
Dean faculty of Science and Engineering,
National University; Sana'a; Yemen
Assistant Professor of Alhajar (Qabaitah) Community College; Yemen

Abstract: Arabic language is a semantic language that has complicated difficulties when compared to English and other languages. In this paper an Arabic speaker recognition system has been developed for introducing conversion of the uttered Arabic speaker instantly after the utterance. The voice samples were recorded, the pre-processing activity detected to evaluate the voice parts from unvoiced, framing and rectangular window slides techniques has been used for segmentation of the Arabic Speech signals, followed by Mel Frequency Spectrum Coefficients (MFCC) for features extractions, The feature vectors are grouped for each spoken sample using VQLBG Algorithm and Gaussian Mixer Model (GMM) applied for classification and recognition an unknowing speaker through his uttered words which belong to specific cluster that is differenced form others clusters related to others Arabic speakers. This approach reported in providing 95.5% of recognition rate.

Keywords: ASRS, Linguistics, Forensic Speaker Recognition, Utterance Arabic Word, MFCC, VQLBG, GMM and EM Algorithm.

INTRODUCTION

Arabic is a semantic language that has complicated difficulties when compared to English language. Some of the difficulties encountered by a speech recognition system that are related to the Arabic language are fully described in literature references such as in [1],[2],[15],[41],[42]. Arabic language inherent mismatch between spoken and written language. Standard Arabic has 34 basic phonemes, of which six are vowels, and 28 are consonants [30]. Arabic has fewer vowels than English. It has three long and three short vowels, while American English has at least 12 vowels [41]. Arabic phonemes contain two distinctive classes, which are named pharyngeal and emphatic phonemes. These two classes can be found only in Semitic languages like Hebrew [42]. Great difficulties occur when several speakers with different dialects are to be recognized. Because the lack of standardization and lack of rules caused the spoken Arabic to be considerably varietal from one region to another. Arabic used in daily informal communication is not the same form of Arabic that is used in books, magazines, newspapers and on TV to broadcast the news. Isolate Arabic alphabet pronunciation is different from pronunciation the same alphabet connected in words. Lack of spoken and written training data is one of the main issues encountered by Arabic ASR researchers. These problems can be minimized by restricting the number of speakers, words and working with good acoustic condition. Also, by avoiding the complexities of fluent speech and working on modern standard Arabic to overcome different dialects [1],[15],[16],[28],[29].

Arabic speaker recognition have important application in daily life, there is a need for controlled access to certain information or places for security. For instances, users have to speak a PIN (Personal Identification Number) in order to gain access to the laboratory door, or users have to speak their credit card number over the telephone line to verify their identity. Some of the possible applications of biometric systems include user-interface customization and access control such as airport check in, building access control, telephone banking or remote credit card purchases. Speech technology offers many possibilities for personal identification that is natural and nonintrusive[1],[2],[30].

Speech recognition has created a technological impact on society and is expected to flourish further in the area of human machine interaction. By checking the voice characteristics of the input utterance, using an automatic speaker recognition system similar to the one that this paper describe, the system is able to add an extra level of security and other applications .

A conversation between people contains a lot of information besides just the communication of ideas. Speech also conveys information such as gender, emotion, attitude, health situation and identity of a speaker. The desire for a more secure identification system leads to the research in the of biometric recognition systems. There are two main properties of biometric features. Behavioral characteristics such as voice, signature are the result of body part movements. In the case of voice it merely shows the physical properties of the voice production organs. The articulatory process and the subsequent speech produced are never exactly same even when the same person utters the same sentence. Physiological characteristics refer to the actual physical properties of a person such as fingerprint, iris and hand geometry measurement[29],[30],[31].

Different approaches can be used in speech recognition such as HMM, ANN, SVM, GMM, Fuzzy Logic, hybrid systems and Combined Classifiers. The topic of this paper deals with speaker recognition that refers to the task of recognizing people by their voices[20],[22].

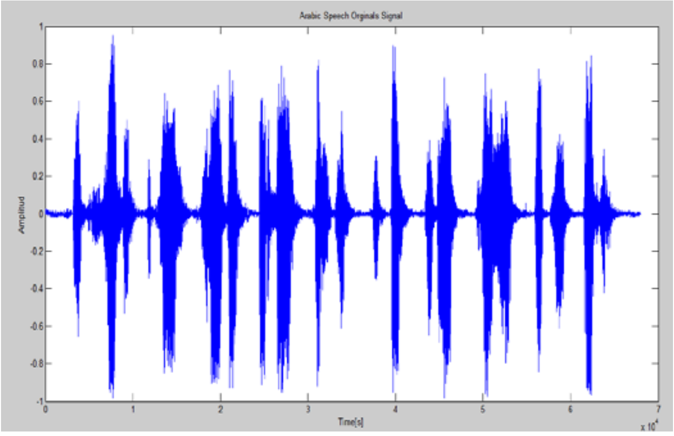
The remaining of this paper will discuss the System Over View in section II, system architecture in section III, the section IV deals with the results and experiment , the conclusion presented in section V and finally the references is assigned for section VI.

Table I: Arabic Digits and It's Associated Words

Sr. No.	*Arabic Words*	*English Arabic*	*Translate to English*
1	واحد	Wahid	One
2	اثنان	Ethnan	Two
3	ثلاثة	Thalatheh	Three
4	اربعة	Arbaah	Four
5	خمسة	Khamsah	Five
6	ستة	Setah	Six
7	سبعة	Sabaah	Seven
8	ثمانية	Thamaniah	Eight
9	تسعة	Tesaah	Nine
10	عشرة	Ashrah	Ten

Arabic Digits and it's corresponding Meaning

Figure 1: Arabic Speech words (Zero to ten) Original Signal



II. System Over View

Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to many services.

The speech wave itself contains linguistic information that includes meaning the speaker wishes to impart, the speaker’s vocal characteristics and the speaker’s emotion. Speech recognition is the process of automatically extracting and determining linguistic information conveyed by a speech wave using computers or electronic circuits.

III. ASR System Architecher

The system describes how to build a simple, yet complete and representative Arabic speaker recognition system. Such a speaker recognition system has potential in many applications. By checking the voice characteristics of the input utterance, using an automatic speaker recognition system similar to the one that this in this paper built, the system is able to add an extra level of many knowledge in the Arabic speech field. The ASR System Architected represented here is included many phases that are described in details as follows:

A. Data Acquisition (Arabic Voiced Recorded)

Arabic Speech databases does not exist so our own database has been collected from Arabic speakers whose can speaks Arabic Language fluently and they recorded by the same recorder one by one to speak Arabic digits words from (whahid to ashrah) meaning (zero to ten) in the same environment and equipment. There are about *5000(10 words X 10 repetitions X 50 speakers)* time series 13 frequencies. Moreover the goal is to create sufficient data for each Arabic Speaker speech samples (Speaker1, Speaker2,….…. Speaker50).

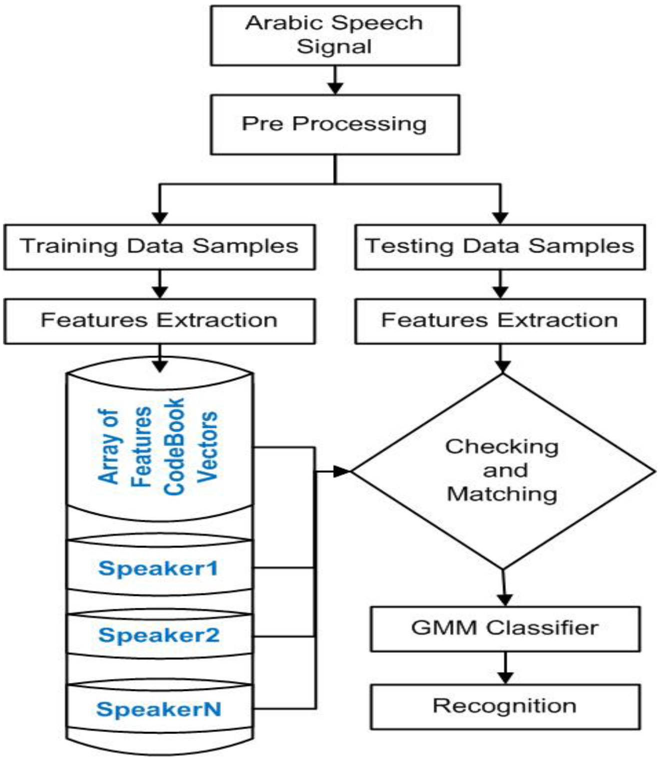


Figure 2: Speaker Identification & Verification Architecture

Each sample is labeled after the Identify (ID) of the speaker and the files were recorded and stored in **.WAV format, using Matlab Language 2014. MFCC's are taken from 50 male Arabic native speakers between the ages of 19 and 25 to represent Arabic spoken words.

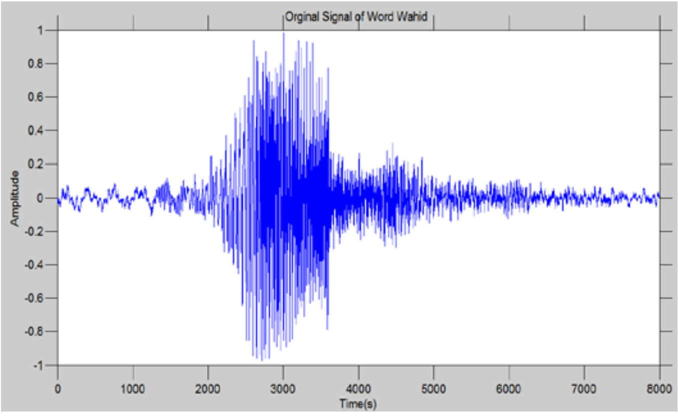


Figure 3: The utterance Wahid Original Signal

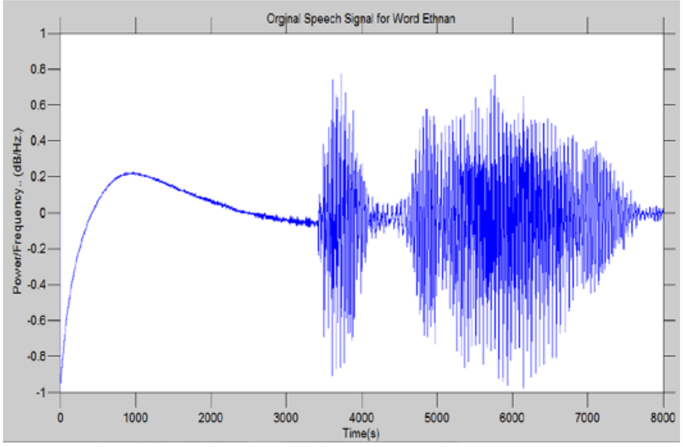


Figure 4: The utterance Ethnan Original Signal

B. Arabic Signal Pre-processing

Voice signal samples into the recognizer to recognize the speech directly, because of the non-stationary of the speech signal and high redundancy of the samples, thus it is very important to pre-process the speech signal for eliminating redundant information and extracting useful information. The speech signal pre-process step can improve the performance of speech recognition and enhance recognition robustness. There are five pre-processing techniques that can be used to enhance feature extraction. These include endpoint detection, pre-emphasis, silence removal, windowing and autocorrelation.

1. Pre emphasis

The speech generated from the mouth will loss the information at high frequency, thus it need the pre emphasis process in order to compensate the high frequency loss. Each frame need to be emphasized by a high frequency filter. The pre emphasis is a 1st order high pass filter. The speech will only remain the track section; it will be very simple for analyzing the speech parameters [10].

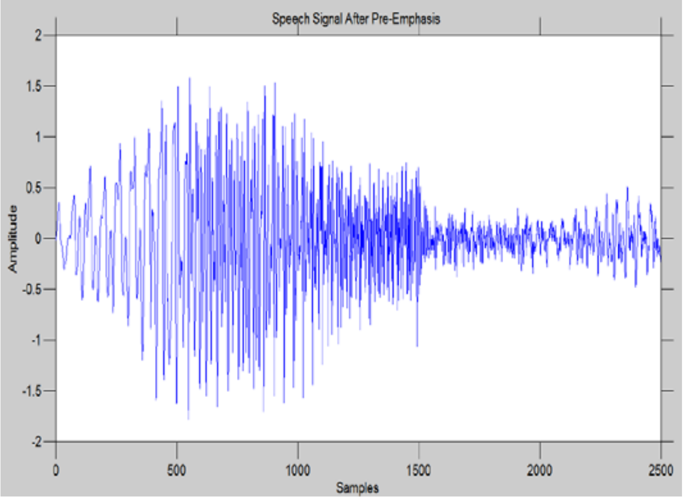


Figure 5: The Speech Arabic Word after Pre emphases process

2. Dynamic time warping (DTW)

The warping between two time series can then be used to find corresponding regions between the two time series or to determine the similarity between the two time series. We desire to develop a dynamic time warping algorithm that is linear in both time and space complexity and can find a warp path between two time series that is nearly optimal[33],[34]. This paper introduced the fast DTW algorithm, which is able to find an accurate approximation of the optimal warp path between two time series. The time series are initially sampled down to a very low resolution. A warp path is found for the lowest resolution.

3. Noise Elimination

The biggest problem ever been in speech recognition systems is the noises in the environment. The pre-trained model for test might be inaccurate; the best result is got when we do the test in exactly the same room as we record the training data.

4. Silence Detection and Removal

Voice Activity Detection (VAD), is the technique used to scan the speech signal from the beginning and to its end for deleting all values under some specified value which is the noise values. Detecting the end of the word generally, it contains two methods in end point detection, one is based on entropy-spectral properties and another is according to double threshold method. In this paper we used double threshold technique.

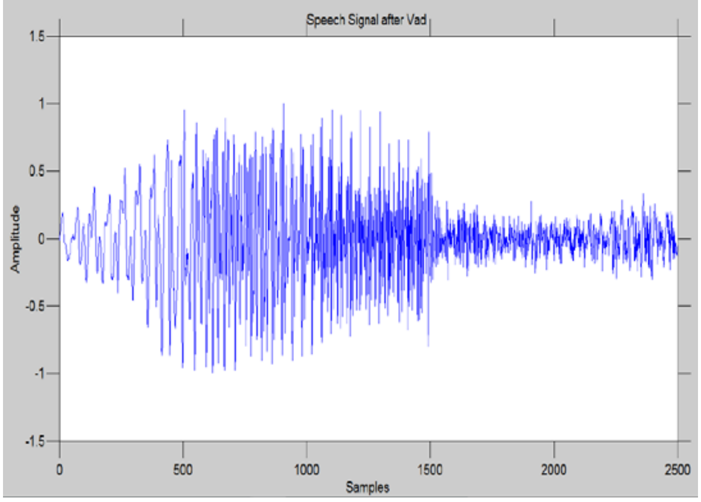


Figure 6: The word "wahad" after VAD detecting

5. Double Threshold

The Double threshold techniques is used for detecting endpoints of speech signal. Because the technique can detect a speech voice or unvoiced, if theshold1 > ratio(ration is a presetting Zero crossing rate) , then it's a speech signal , namely , it's been found the speech head . Vice versa, if theshold2 < ratio, then the speech signal is over, which means speech tail will be found. The signal between head and tail is the useful signal and thus the threshold in a big noise environment is adjustable as shown in figure (5). The more Generally speaking, author will check the endpoint of speech voice by average energy or the product of average amplitude value and zero crossing rate with the following equation(6). An average energy can be defined as:

$$E_n = \sum_{m=0}^{N-1}[w(m)x(n-m)]^2, \quad 0 \le m \le N-1 \qquad (1)$$

where x(n) is the speech signal, N the length of frame, m is the frame shift, w (m) is the windows function which expressed as:

$$\mathrm{W(m)} = \begin{cases} 1, M = 0 \sim N-1 \\ 0, M = other \end{cases} \qquad (2)$$

The signal windowing is to avoid truncation effect when framing, so windowing is necessary when extract every frames of sound signal. Windowing detailed will described in next section[6].

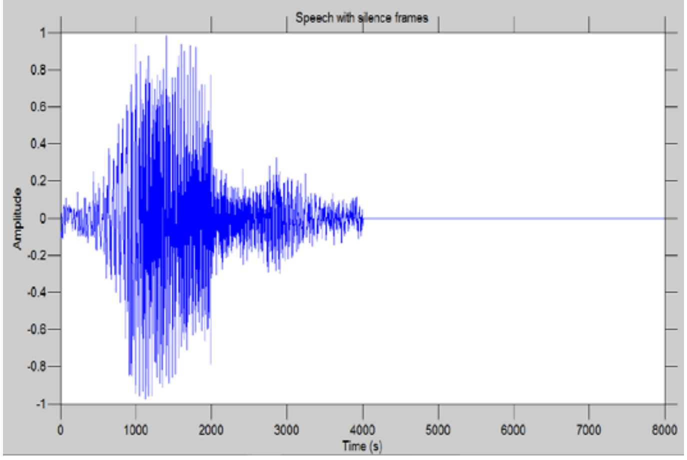


Figure 7: Signal Wahad after detecting the additional noises

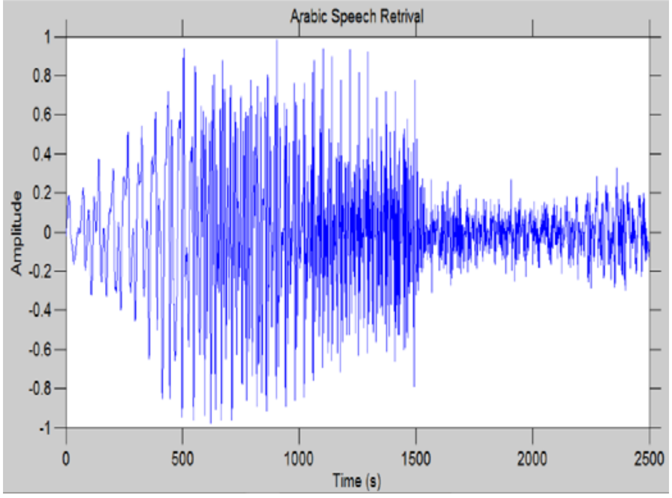


Figure 8: The Signal wahad After Unvoiced Removal

Zero crossing rate is another equation has been used during the detection, it indicates number of times that a frame of speech signal waveform cross through the horizontal axis. Zero crossing analysis is one of the simplest method in time domain speech analysis [9]. It can be defined as:

$$Z_n = \frac{1}{2}\sum_{m=0}^{n-1} \left| sig[x(m) - sig[x(m-1) \right| * w(n-m) \qquad (3)$$

C. Feature Extraction

In speaker recognition technology, feature extraction is mainly used. Extracting features is a process of holding useful statistics of data from a speech signal while eliminating unwanted signals such as noise. Here, the conversion of the original acoustic wave into a tightly packed representation of the signal feature selection technique. The series of eigenvectors representing a close-packed speech signal is determined by a feature extraction method. The feature vectors extracted from the original signal in the feature extraction module prominence speaker-specific attributes and vanquish statistical redundancy [9]. This system will perform operations in three phases which are a preprocessing phase, the training phase, and decision phase.

The paper proposed Mel Frequency Cepstral Coefficients (MFCCS) for features extraction, it is perhaps the best known and most popular, recently used in most research. The popularity of this method can be explained by the low computational cost compared to FFT, LPCC and LPC based techniques [1], [2], [4], [6], [10], [19].

MFCC's are based on the known variation of the human ear's critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the important phonetically characteristics of speech and speaker. The steps of computing MFCCs is described in more detail as follows:

1. Frame Blocking

In this step the continuous speech signal is blocked into frames of N samples, with adjacent frames being separated by M ($M < N$). The first frame consists of the first N samples. The second frame begins M samples after the first frame, and overlaps it by N - M samples and so on. This process continues until all the speech is accounted for within one or more frames. Typical values for N and M are N = 256 (which is equivalent to ~ 20 ml sec windowing and facilitate the fast radix-2 FFT) and M = 30.

2. Rectangular Windows (Windowing)

The selection of different windows will determine the nature of the speech signal short-time average energy. During the increment the study found the length of window played very important role in the design filter. If the length of window is too long, the pass band of filter will be narrow. Otherwise, if the length of window is too small , the pass band of filter will be wide, and the signal can be represented sufficiently equally distributed. [19], [20]. The main lobe of hamming window is the widest, and it has the lowest side lobe level. The choice of the window is critical for analysis of speech signal, utilizing rectangular window is easily loss the details of the waveform, on the contrary, hamming window is more effective to decrease frequency spectrum leakage with the smoother low pass effect.

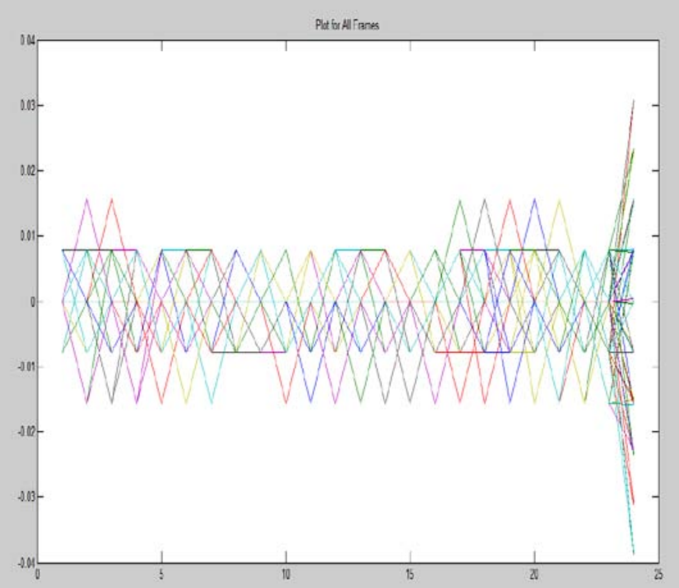


Figure 9: Most Features of Frames

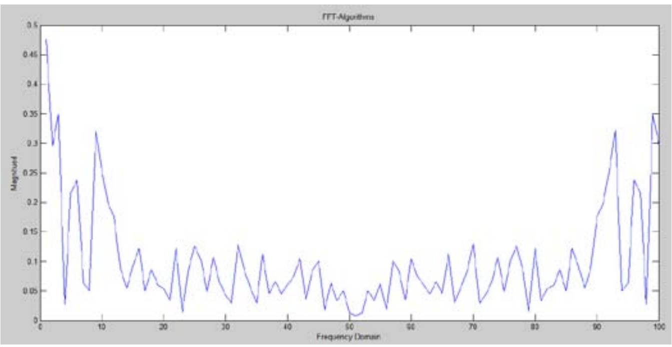


Figure 10: Transform part of Signal to Frequency Domain

Therefore, Rectangular window is more fitting for processing signals in time domain and hamming window is more used in frequency domain. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. The study defined the window as $w(n), 0 \le n \le N-1$, where N is the number of samples in each frame, then the result of windowing is the signal as follows:

$$\left[y_l(n) = x_l(n)w(n), \quad 0 \le n \le N-1\right] \qquad (4)$$

Typically the *Hamming* window is used, which has the form:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \quad (5)$$

3. Fast Fourier Transform (FFT)

The FFT is used to convert each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT), which is defined on the set of N samples $\{x_n\}$, as follow:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N}, \left[k = 0,1,\ldots N-1\right] \qquad (6)$$

In general X_k's are complex numbers and we only consider their absolute values (frequency magnitudes). The resulting sequence $\{X_k\}$ is interpreted as follow:

positive frequencies $0 \le f < F_s/2$ correspond to values

$0 \le n \le N/2-1$, while

negative frequencies $-F_s/2 < f < 0$ correspond to

$N/2+1 \le n \le N-1$.

Here, F_s denotes the sampling frequency.

The result after this step is often referred to as *spectrum* or *periodogram*.

4. Mel-frequency Wrapping

As psychophysical studies have shown that human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency, *f*, measured in Hz, a subjective pitch is measured on a scale called the 'mel' scale. The *mel-frequency* scale is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz.[20], [24],[25],[26],[27].

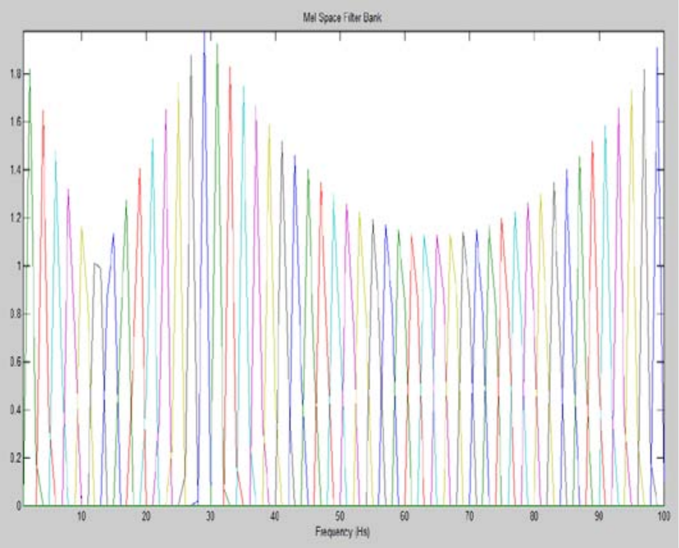


Figure 11: Mel Spaced Filter Bank

One approach to simulating the subjective spectrum is to use a filter bank, spaced uniformly on the mel scale. That filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant mel frequency interval. The number of mel spectrum coefficients, *K*, is typically chosen as 20. For many reasons the paper applied the filter bank in the frequency domain, thus it simply amounts to applying the triangle-shape windows as in the figure No.(12) to the spectrum. A useful way of thinking about this mel wrapping filter bank is to view each filter as a histogram bin (where bins have overlap) in the frequency domain.

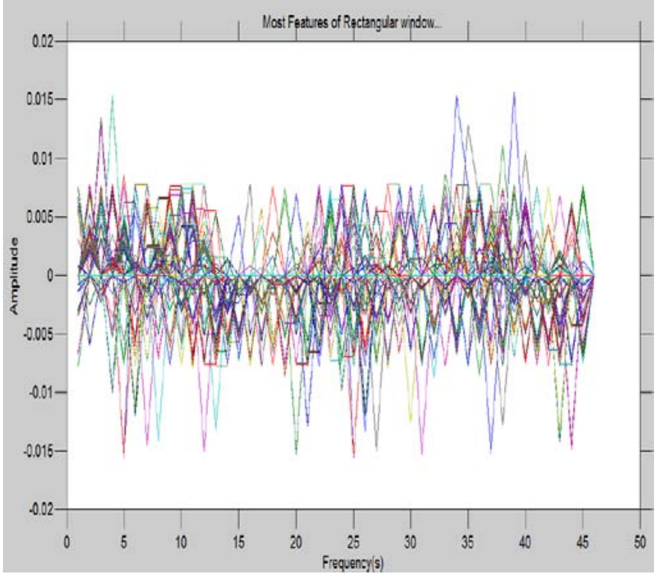


Figure 12: Most Features After Rectangular Window

5. Cepstrum

In final step, the study the log mel spectrum back to time. The result is called the mel frequency cepstrum coefficients (MFCC). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the mel spectrum coefficients (and so their logarithm) are real numbers, we can convert them to the time domain using the Discrete Cosine Transform (DCT).

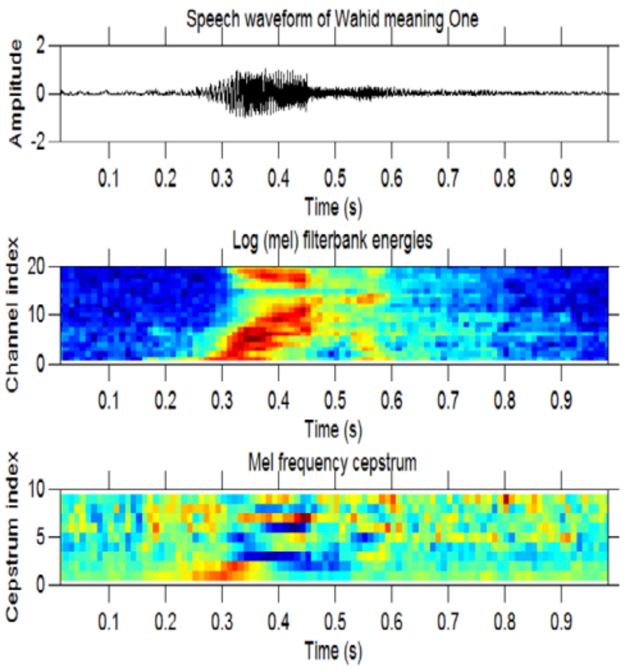


Figure 13: Mel Frequency Cepstrum for Speech Wahid

Therefore if the study denoted the mel power spectrum coefficients that are the result of the last step are $\tilde{S}_0, k = 0,2,..., K-1$, they can calculate the MFCC's, $\tilde{c}_n$, as:

$$\tilde{C}_n = \sum_{k=1}^{K} (\log \tilde{S}_k) \cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right], \qquad n = 0,...K-1 \qquad (7)$$

Note that the study excluded the first component, $\tilde{c}_0$, from the DCT since it represents the mean value of the input signal, which carried little speaker specific information[19].

6. *Training Phase*

Vector Quantization must able to estimate of the computed feature vectors. Storing every single vector that generate from the training mode is impossible, since these vectors are defined over a high dimensional space. It is often easier to start by quantizing each feature vector to one of a relatively small number of template vectors, with a process called vector quantization. VQ is a process of taking a large set of feature vectors and producing a smaller set of measure vectors that represents the centroids of the features. Furthermore of VQ, storing every single vector that we generate from the training is impossible. By using these training data features are clustered to form a codebook for each acoustic of word [12]. Finally Saving Trained Features with specific intent for using and editing without the aid of any programming algorithms.

7. Feature Matching

The VQLGB technique used for mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a *cluster* and can be represented by its center called a *codeword* and the combination of are called a *codebook* [1],[2],[8],[21].In feature matching of speech signal, Vector Quantization (VQ) technique that included plotting VQ codebook and also implementing a well-known algorithm developed by Linde, Buzo and Gray which was called LBG algorithm was used[1].

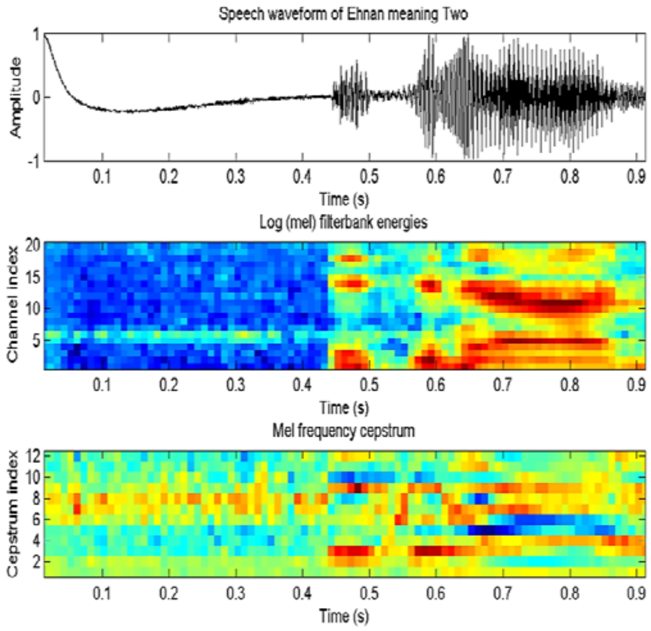


Figure 14: Mel Frequency Cepstrum for Speech Ethnan

The Euclidian distance is calculated for a given codebook to find least distance found denoted as VQ distortion. Similarly distortions are computed for the remaining feature vectors and a summed up. Same procedure is repeated for rest of the speakers. The least summation of the VQ distortions will identify the desired Speaker[7],[8],[9].

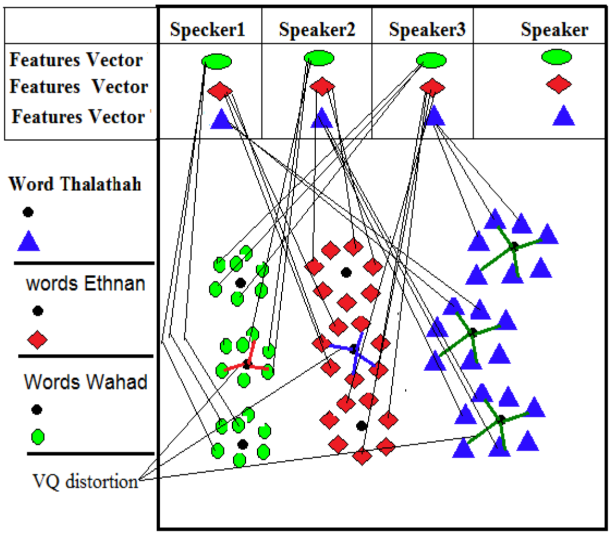


Figure 15 : VQ codebook information for Each Speaker

D. Gaussian Mixture Model GMM

A Gaussian mixture model (GMM) forms clusters as a mixture of multivariate normal density components. For a given observation, the GMM assigns posterior probabilities to each component density (or cluster). The posterior probabilities indicate that the observation has some probability of belonging to each cluster. A GMM can perform hard clustering by selecting the component that maximizes the posterior probability as the assigned cluster for the observation. GMM is appropriate method than k-means clustering when clusters have different sizes and different correlation structures within them. The paper designed a GMM model which used for Arabic speaker recognition with mixtures and diagonal covariance matrices. Gaussian mixtures are combinations of Gaussians, or–normal distributions. Feature vectors are displayed in d-dimensional feature space after clustering, they somehow look like Gaussian distribution. It means each matching cluster can be viewed as a Gaussian probability distribution and features fitting to the clusters can be best characterized by their probability values [35],[36],[37],[38].

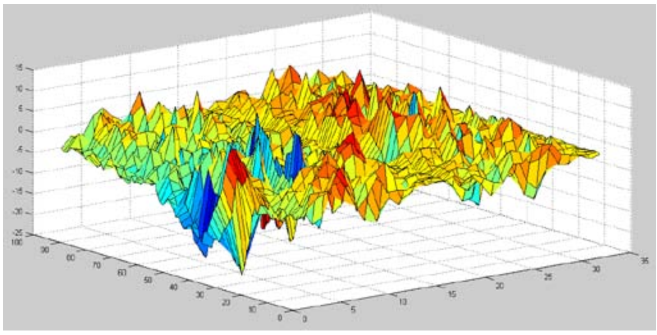


Figure 16: Features of Arabic speaker1 uttered Word Wahad

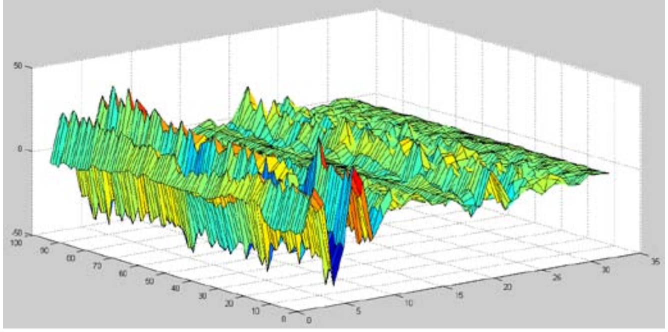


Figure 17: Features of Arabic speaker1 uttered Word Ethnan

The use of Gaussian mixture density for Arabic speaker identification can derived by two facts as follows[14],[15]:

- Individual Gaussian classes are interpreted to represents set of acoustic classes. These acoustic classes represent speaker vocal tract information.
- Gaussian mixture density provides smooth approximation to distribution of feature vectors in multi-dimensional feature space.

A mixture of Gaussians can be written as a weighted sum of Gaussian densities. Recall the d-dimensional Gaussian probability density function (pdf) for the d-dimensional random vector x and given by the equation[26]:

$$g(u,\Sigma)(x) = \frac{1}{(2\pi)^{\pi/2}\,|\Sigma|^{1/2}} \exp^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)} \qquad (8)$$

A weighted mixture of K Gaussians can be written as

$$GM(x) = \sum_{k=1}^{k} w_k\, g(u_k, \Sigma_k)(x) \qquad (9)$$

where all the weights are positive and add up to one:

$$\sum_{k=1}^{k} w_k = 1 \quad and \quad w_k > 0 \; for \quad k \in \{1,2,....k\} \quad (10)$$

The parameters of this probability density function are the number of Gaussians, their weighting factors, and the mean vector and covariance matrix of each Gaussian function. To find these parameters and optimallyfit of probability density function for a set of data, an iterative algorithm, the expectation–maximization (EM) algorithm can be used [3],[7],[19],[26].

E. RECOGNITION

GMM assumes vector space to be divided into specific components depending on clustering of feature vectors and frames the feature vector distribution in each component to be Gaussian. As initially the study has no idea about which vector belongs to which component a likelihood maximization algorithm is followed for optimal classification. For testing purpose the calculated posteriori probability of test utterance and the reference speaker maximizing Gaussian distribution is termed as identified of unknown speaker[17],[35],[38].

The words uttered by any Arabic speaker will belong to specific cluster that is differenced form others clusters related to others Arabic speakers. This is the base techniques chosen to be verified and recognized as it is assigned to the known speaker.

IV. EXPERIMENT AND RESULT

The study applied pattern recognition techniques to design speaker identification reference models for trained features and then can be recognize any sequences of acoustic vectors uttered by unknown speaker. VQLBG-based pattern recognition technique used to build speaker reference models from their vectors in the training phase and then can identify any sequences of acoustic vectors uttered by unknown speaker[1],[2],[3][14],[18].

The GMM models used to compute the pairwise between the codewords for each speaker and trained vectors in the iterative process classifier. Train and test programs (which require three functions MFCC, VQLBG and GMM to simulate the training, testing and recognition procedures in Arabic speaker recognition system, respectively has been implemented effectively. The results compared in between and evaluated for gaining more efficient rate[8],[12],[14], [19],[21]. All experiments were implemented in Matlab 2014 and using Intel(R) Core(TM) i5 CPU M 370 @ 2.40GHz.

V. CONCLUSION

The results obtained using MFCC and VQLBG algorithm are evaluated carefully. An accuracy reported for VQLBG was 75.6%. Then the study applied GMM method for Arabic speaker identification and recognition an accuracy gained was 95.5%. It can be seen that the GMM model is the most attractive when compare with VQLBG Algorithm. The study concluded that the efficiency results has been obtained by GMM model comparing with VQLBG algorithm [14],[20],[35],[36],[37],[38].

VI. REFERENCES

[1] Mowaffak O. A. Albaraq, "Arabic Speech Recognition System through VQLBG and Euclidean Distance Algorithms using Matlab", IJCA., (0975 – 8887) Volume 177 – No. 39, February 2020.

[2] Nilu Singh, Alka Agrawal, and R. A. Khan, "Voice Biometric: A Technology for Voice Based Authentication", Advanced Science Engineering and Medicine Vol. 10, 1–6, 2018.

[3] Ajinkya N. Jadhav, Nagaraj V. Dharwadkar, " A Speaker Recognition System Using Gaussian Mixture Model, EM Algorithm and K-Means Clustering", International Journal of Modern Education and Computer Science(IJMECS), Vol.10, No.11, DOI: 10.5815/ijmecs.2018.

[4] Nikita Dhanvijay *, Prof. P. R. Badadapure, "Hindi Speech Recognition Technique Using Htk", IJESRT International Journal Of Engineering Sciences & Research Technology, ISSN: 2277-9655, 2016.

[5] Neha Sharma and Shipra Sardana, “Designing a Real Time Speech Recognition System using MATLAB”, IJ CA(0975 – 8887) National Conference on Latest Initiatives& Innovations,.. (IICE 2016).

[6] Sreelakshmi V and Dr. Gnana Sheela K., "Design of an Intelligent Speaker Recognition System using Mel Frequency Cepstrum Coefficients and Vector Quantization for Biometric Authentication", IJCSN International Journal of CSN., Vol. 4, Issue 6, 2015.

[7] Nisha N. Nichat and P. C. Latane, ”Real Time Speaker Recognition using Mel- Frequency Cepstral Coefficients (MFCC),VQLBG & GMM Techniques”, Vol. 5, Issue 6, June 2016, IJIRSET DOI:10.15680.2015.

[8] Mr. P, Kumar, Dr. S. L. Lahudkar, Automatic Speaker Recognition using LPCC and MFCC", IJRITCC, Volume: 3 Issue: 4 | April 2015.

[9] Roma Bharti Mtech, Priyanka Bansal, "Real Time Speaker Recognition System using MFCC and Vector Quantization Technique", International Journal of Computer Applic. Volume 117 – No. 1, 2015.

[10] Naoki H., Koichiro Y., Katsutoshi I., Shinsuke M., and Hiroshi G. O., “Automatic Speech Recognition for Mixed Dialect Utterances by Mixing Dialect Language Models”, IEEE/ACM transactions on Audio, Speech, And Language Processing, vol. 23, no. 2, february 2015.

[11] Deepak Baby, Tuomas Virtanen, Jort F. Gemmeke, Hugo van hamme, “Coupled dictionaries for Exampler- Based Speech Enhancement and Automatic Speech Recognition”, IEEE/ACM Tans. On Audio, speech and Language processing, vol. 23, No. 11, 2015.

[12] Dr. R.K. Prasad and Mr. K. Patel, “Speech Recognition and Verification Using MFCC & VQ”, IJARCSSE, Volume 3, Issue 5, May 2013.

[13] Aaron Nichie, "Voice Recognition Using Artificial Neural Networks And Gaussian Mixture Models", International Journal Of Engineering Science And Technology (Ijest), Vol. 5 No.05 May 2013.

[14] Jorge Martinez, Hector Perez, Enrique Escamilla, M. Mabo Suzuk, "Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques", 978-1-61284-1325-5/12, 2012 IEEE.

[15] Mohammad A., Raja A., Roziati Z., Moustafa E., and Othman K., "Arabic Speaker-Independent Continuous Automatic Speech Recognition Based on a Phonetically Rich and Balanced Speech Corpus", The International Arab Journal of Information Tech., Vol. 9, No. 1, 2012.

[16] N.hammami, M. Bedda and N. frah, "Spoken Arabic Digits Recognition Using MFCC Based on GMM", Malisia Conference IEEE, 2012.

[17] Ms. Arundhati S. Mehendale and Mrs. M.R. Dixit "Speaker Identification" Signals and Image processing : An International Journal (SIPIJ) Vol. 2, No. 2, June 2011.

[18] Prof. Ch.Srinivasa Kumar, "Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm", International Journal on Computer SE, (IJCSE), Vol. 3 No. 8 August 2011.

[19] G.S. Kumar, K.A.P. Raju, Dr. Mohan R. C. and P. Satheesh, "Speaker Recognition Using Gmm", International Journal of Engineering Science and Technology, Vol. 2(6), 2010.

[20] Vibha Tiwari, "MFCC and its applications in speaker recognition", International Journal on Emerging Technologies 1(1): 19-22(2010).

[21] Mahdi Shaneh and Azizollah Taheri, " Voice Command Recognition System Based on MFCC and VQ algorithms" World Academy of Science, Engineering and Tech. 2009.

[22] M.A.Anusuya, S.K. Katti, "Speech Recognition by Machine: A Review", (IJCSIS), Vol. 6, No. 3, 2009.

[23] Suliman S. Al-Dahri, YoussafH. Al-Jassar, YousefA. Alotaibi, Mansour M. Alsulaiman, Khondaker Abdullah-Al-Mamun, "A Word-Dependent Automatic Arabic Speaker Identification System", 2008 IEEE.

[24] M. F.-Zanuya, M. Hagmüllerb, G. Kubinb, "Speaker Identification Security Improvement Bymeans Of Speech Watermarking", Elsevier, Science Direct, Pattern Recognition 40 (2007) 3027–3034.

[25] Ramzi A. Haraty and Omar El Ariss, " CASRA+: A Colloquial Arabic Speech Recognition Application", Lebanese American University, Beirut, AJAS 4 (1): 23-32, 2007, ISSN 1546-9239.

[26] Xiaodong Lui, et.al., A study of variable parameter Gaussian mixture HMM modeling fro Noisy speech recognition , IEEE Transactions on Audio, Speech and Language processing, Vol.15,No.1, Jan. 2007.

[27] M. Faundez-Zanuy, M. Hagmüller, G. Kubin, "Speaker verification security improvement by means of speech watermarking", Speech Commun., issue Dece. 2006.

[28] Bahi, H. and M. Sellami, “A connectionist expert approach for speech recognition”, The International Arabic Journal of Information Technology, 2004.

[29] Lazli, L. and M. Sellami, “Speaker independent isolated speech recognition for Arabic language using hybrid HMM-MLP-FCM system”, AICCSA, Tunisia, 2003.

[30] El Choubassi, M.M. et al., “Arabic speech recognition using recurrent neural networks”, IEEE, Intl. Symp. Signal Processing and Information, 2003.

[31] Kirchhoff, K., et al., “Novel approaches to Arabic speech recognition”, Final Report from the JHU Summer Workshop, Tech. Rep., John , Hopkins University 2002.

[32] E. Darren. Ellis "Design of a Speaker Recognition Code using MATLAB "Department of Computer and Electrical Engineering University of Tennessee, Knoxville Tennessee 37996. 9th May 2001.

[33] S. Furui, “An overview of speaker recognition technology”, ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, 1994.

[34] L.R. Rabiner and B.H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, N.J., 1993.

[35] D. A. Reynolds, ”Robust text-independent speaker identification using Gaussian mixture speaker models,” Speech and Audio P., IEEE Trans. on, p. vol 2 (1), 2002.

[36] J. Lan, Gaussian Mixture Model Based System Identification and Control, University of Florida, 2006.

[37] G. Xuan, W. Zhang och P. Chai, ”EM algorithms of Gaussian mixture model and hidden Markov model,” Image Processing, 2001. Proceedings. 2001 International Conference on , p. Vol 1, 2001.

[38] Douglas A. Reynolds and Richard C. Rose, “Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models“, Ieee Transactions On Speech And Audio Processing, VOL. 3, NO. 1, January 1995.

[39] D.A. Reynolds, R.C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models", IEEE Trans. Speech Audio Proce. 1995.

[40] V.D.Bhagile , Mowaffak Al-Baraq, R.J. Ramteke, S.C.Mehrotra, “Recognition System for Arabic Printed Numeral :A size Independent Approach”, IEEE IDICON 2007&16th A. S. of IEEE Bangalore,6-8 sept.,2007.

[41] System :A Block Based Approach” , IEEE, International Conference on ,ACVIT-07,28th – 30th Nov. 2007.

[42] Mowaffak Al_Barraq And S.C.Mehrotra “Handwritten Arabic Text Recognition System Using Window Based Moment Invariant Method”, IJAR in Computer Science, ISSN No. 0976-5697, Volume 2, No. 1, Jan-Feb 2011.

[43] Mowaffak Al-Baraq And S. C. Mehrotra “Arabic Handwritten Amount in Cheque Through Windowing Approach”, IJCA., (P-ID: pxc3902420) , Appirl, 2015.

Dr. Mowaffak O. A. Albaraq, 1972, Assistant Professor & Dean of Almafa'ar Community College, Taiz, Yemen, has received his M.C.A. (Computer Application), 2006 and Ph.D. (Computer Science), 2010 from Dr. Baba saheb Ambedkar Marathwada University, Aurangabad (India) respectively. Presently working as a Dean & Assistant Professor in College of Science and Engineering. National University, Sana'a Yemen and Assistant professor in Department of Information Technology Alhajar (Qubaitah) Community College, Yemen. His areas of interest are pattern recognition, Image Processing, Document Analysis, Soft computing, Machine Learning, Human Computer Interaction. He has published more than 10 research articles in National/International journals &conferences. He has participated in various academic courses, International & National Conferences/Seminars and workshops. Dr. Mowaffak has received scholarship from Ministry of Higher Education and Scientific Research Yemen 1990-to-1994 and as well has received scholarship form Ministry of Education and Vocational Training Yemen 2003-to-2009.