7

Arabic Speech Recognition System Trough VQLBG Algorithm Using Matlab

Dr.Mowaffak O. A. Al baraq
Dean faculty of Science and Engineering,
National University; Sana'a; Yemen
Professor of Alhajar Community College, Taiz; Yemen
Email:dr.mowaffak.albaraq@gmail.com

Abstract

A rectangular window techniques has been used for segmentation of the Arabic Speech uttered words to atomic or sub atomic samples. Real time Arabic Speech Recognition system has been developed for introducing recognition of the uttered Arabic words instantly after the utterance. This paper introduces a unique technique making interaction of human with a computer for natural language processing which is basically a speech recognition system. In this paper 100 Arabic voice samples were recorded through a microphone and the system was trained according to the recoded voice samples. MFCC features of speech sample were calculated, Vector Quantization for mapping large feature vectors to finite cluster codewords, build trained codebook model for each word and Euclidean Distances used for recognition word according to distortions associated with each features. This system provides a high accuracy in case of Arabic speech words.

Keywords: A rectangular window, ASR System, Feature Extraction, Arabic word, NLP, DTW, FFT, Mel frequency scale, Mel power spectrum, MFCC, CodeBook, VQLBG and Euclidean Distances Algorithm.

1. INTRODUCTION:

Arabic Speech Recognition System is useful in a large variety of applications banking, business applications, postal zip code reading, security affairs, controlling machine, operate equipment, robotics conversation, communication, expert systems and data entry applications….etc. Arabic is a language spoken by Arabs in over to 22 countries, and roughly associated with the geographic region of the Middle East and North Africa as first language (Mother Tongue). It is also spoken as a second language by several Asian countries, (e.g. Iran, Indonesia, India, Malaysia, Pakistan, ..etc), in which Islam is the principle religion.

Arabic is a Semitic language, and it is one of the oldest languages in the world. It is the fifth widely used language nowadays. Non-Semitic Languages such as Farsi, Urdu, Malay, and some West African languages such as Hausa have also adopted the Arabic alphabet for writing. Due to the cursive nature of the script, there are several characteristics that make recognition of Arabic distinct from the recognition of Latin scripts or Chinese.

Due to the cursive nature of the script, there are several characteristics that make recognition of Arabic distinct from the recognition of Latin scripts or Chinese. There are difficulties in this script [3,4,5,6,7,8,25]. Not much work has been done in Arabic Speech Recognition (ASR) as compared to other languages.

The main problems in Arabic Speech Recognition Online or Offline is due to property of Arabic Language along to Arabic generating voices [7,8,9,10,11,12,16,15,25].

1- Standard Arabic has 34 basic phonemes[1], of which six are vowels, and 28 are consonants. Arabic has fewer vowels than English. It has three long and three short vowels, while American English has at least 12 vowels.
2- Arabic phonemes contain two distinctive classes, which are named pharyngeal(بلعومي) and emphatic phonemes. These two classes can be found only in Semitic languages like Hebrew [20-22].
3- The allowed syllables in Arabic language are: CV, CVC, and CVCC where V indicates a (long or short) vowel while C indicates a consonant.

Arabic utterances can only start with a consonant. All Arabic syllables must contain at least one vowel. Also Arabic vowels cannot be initials and can occur either between two consonants or final in a word. Arabic syllables can be classified as short or long.

4- The CV type is a short one while all others are long. Syllables can also be classified as open or closed. An open syllable ends with a vowel, while a closed syllable ends with a consonant. For Arabic, a vowel always forms a syllable nucleus, and there are as many syllables in a word as vowels in it .

5- Great difficulties occur when several speakers with different dialects are to be recognized.

6- Homophone is a word that is pronounced the same as another word but differs in meaning. For example: The word علم mean flag, the word علم that mean understood and the word علم that means science. The word ساعة that means clock or time and the same word ساعة that means a day of the judgment.

7- Arabic language is morphologically rich which causes a high vocabulary growth rate. This high growth rate is problematic for language models by causing a large number of out-of-vocabulary words [25].

8- Arabic Language is cursive in general spoken and written from right to left. Arabic letters are normally connected to each other on the baseline and recognition must consider this aspect.

9- Isolate Arabic alphabet pronunciation is different from pronunciation the same alphabet connected in words.

Table 1: Arabic Digits from (0 to 10) with corresponding words and English meaning

Arabic digits	0	1	2	3	4	5	6	7	8	9	10
Latin Digits	0	1	2	3	4	5	6	7	8	9	10
Arabic Words	صفر	واحد	اثنان	ثلاثة	اربعة	خمسة	ستة	سبعة	ثمانية	تسعة	عشرة
English Words	Zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten

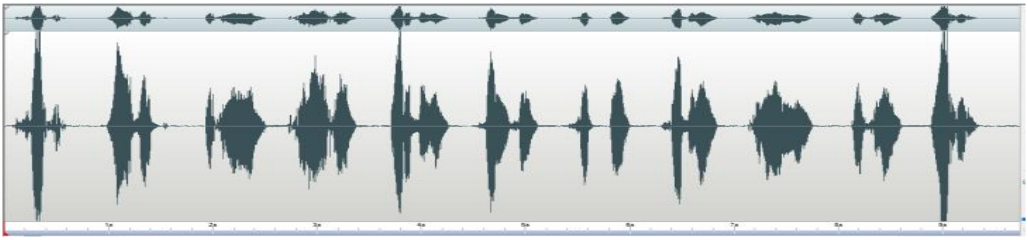


Figure 1-a: Arabic Speaker for Arabic Speech words (wahad to asrah)

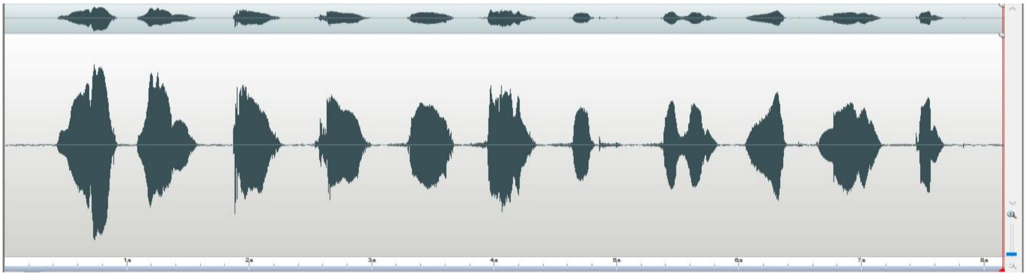


Figure 1-b: Arabic speaker for English speech words from (zero to ten)

10- Arabic speech is more difficult than English speech as comparison in figure 1-a,1-b above which appear from utterance Arabic energy and long which need more pre processing for fixing and adapting for longer processing

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. Different approaches can be used in speech recognition such as HMM, ANN, SVM, GMM, Bayes, Fuzzy Logic, hybrid systems and Combined Classifiers.

2. SYSTEM OVERVIEW

Speech recognition is an important application of Natural Language Processing (NLP). Speech is the most important part of communication. We express our ideas through a specific language. Computers understand our language (natural language) by speech recognition. Speech or word by word recognition is the process of automatically extracting and determining linguistic information conveyed by a speech wave using computers. Linguistic information, the most important information in a speech wave, is called phonetic information. The term speech recognition means the recognizing the spoken words only. However, the recognition system has no idea what those words mean. It only knows that they are words and what words they are. To be of any use, these words must be passed on to higher-level software for syntactic and semantic analysis. It is a technique of pattern recognition, where acoustic signals are tested and framed into phonetics (number of words, phrases and sentences)[1,2,3,4,24,25].

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 services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers.

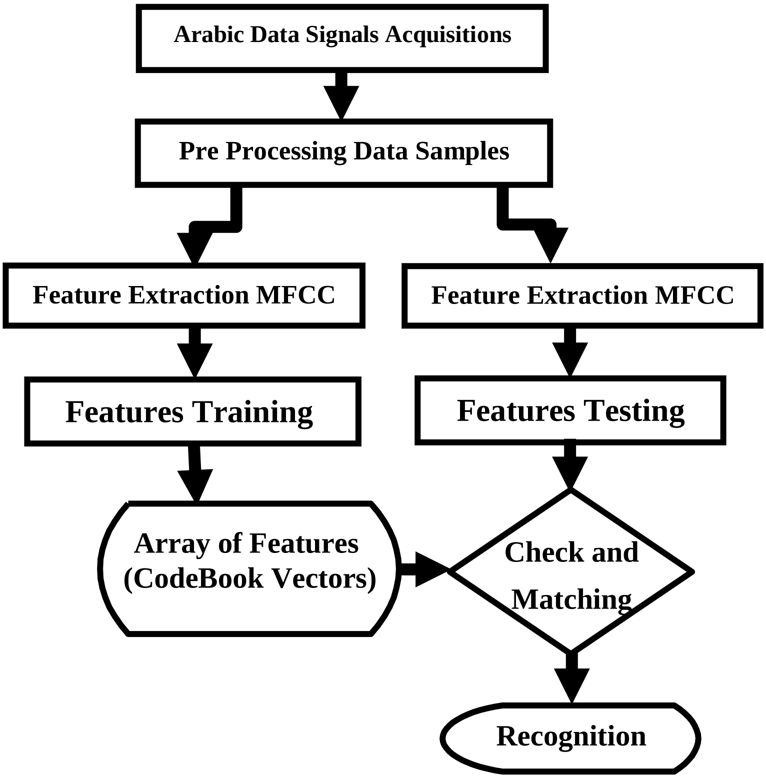


Figure 2: ASR System Architecture

The speech wave itself contains linguistic information that includes the 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. Only the linguistic information is needed from the speech wave, while the rest of the information is used in other fields of signal processing. There are basically two types of speech [3]:

1. Continuous speech
2. Discrete speech.

Discrete speech consists of isolated words that are separated by silences. The advantage of discrete speech is that word boundaries can be set exactly while with continuous speech; words will be spoken without silences[5,6].

The remaining of this paper will discuss the system architecture in section 2 and the phases of the ASR System Architecture are presented in section 3, section 4 deals with the results and experiment , conclusion is presented in section 5 and finally section 6 for references.

3. SYSTEM ARCHITECTURE

The block diagram of ASR System in figure 2 is have the following components: Data acquisition, Pre processing Features extraction, Features Training, CodeBook Vectors, as well as Features Testing, Check and Matching and Recognition for acoustics words. The steps will discuss in details as follows:

3.1 DATA ACQUISITION

Own Data samples has been collected from Arabic students speakers whose can speaks Arabic language fluently and they recorded by the same recorder one by one to speaks Arabic numerals words from (whahad to ashrah) (zero to ten) . Our speech data samples consists of about 100 samples on each speaker recorded five samples each sample including the same words for more efficiency. So we have five hundred speech simples versions of each word collected from those speakers. Though this might seem a lot, it is probably is sufficient to obtain a speaker independent ASR System. More over the speech database contains only male speakers. There is however a large database of speech samples available on the Internet but unfortunately it contains only English samples. So that a database is created by recording some speech samples and the speech samples are stored into (namfiles.wav) sound files to manageable sampling the input, which results in a discrete-time speech signal.

3.2 PRE-PROCESSING

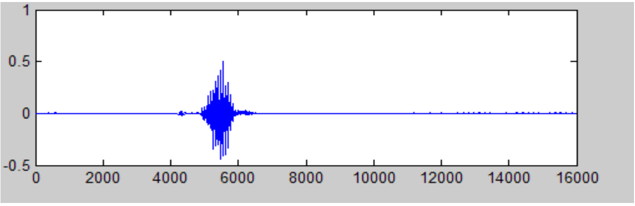
Pre-processing was used to make the discrete-time speech signal more amendable for the processes that followed. There are five pre-processing techniques that can be used to enhance feature extraction. These include DC offset removal, silence removal, pre-emphasis, windowing and autocorrelation [15,18,22,23,24].

3.2.1. Dynamic time warping (DTW)

DTW is a technique that finds the optimal alignment between two time series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis. This 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.

Problem: 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.

Approach: In this paper we introduce 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.



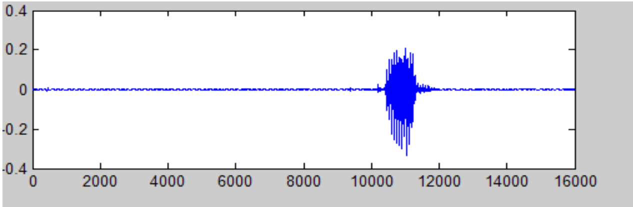


Figure 3: Wahed Utterances of the same word "One" at different times

Figure 3 shows two utterances of the word “One” but they start in different times, thus causing mismatching in recognition process although they are the same word.

Solution: Start recording when detecting voice only... We will do it through data acquisition toolbox. For instance, we have to set a threshold to trigger the data acquisition; otherwise, the program would constantly save the analog input in the memory careless about it is our voice or noise, and finally result in a dead loop or memory exceed.

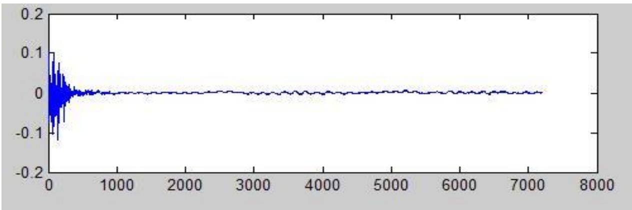


Figure 4-a: The word "One" detected at the first of time frame

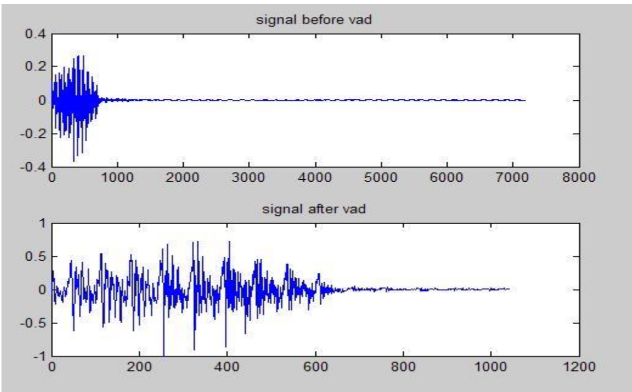


Figure 1-b: The word "One" after cutting the additional noise at the end

3.2.2. Noise elimination

The biggest problem ever been in speech recognition systems is the noise 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.

Solution: There is still no way to finally remove or eliminate the noise in the environment from the speech, but there are some ways to get around this problem as recording many samples of speech in different environments, recoding in quiet and noisy environments, recording many samples to different persons.

1- **Speaking way during training**
This also affects the result a little bit, since we can not make sure people say the same word always in the same way.

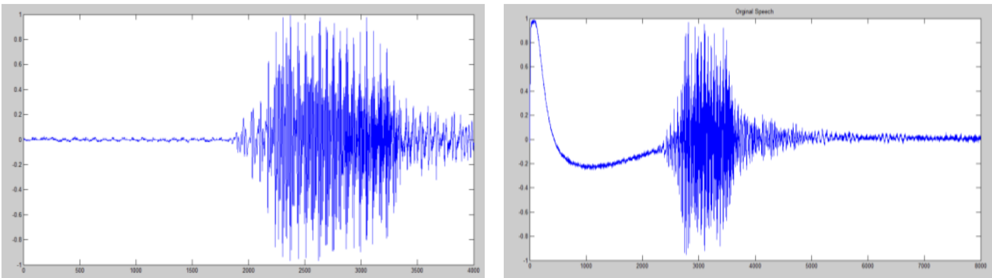


Figure 4-c: The word whahad from same speaker in different time

Solution: It is necessary to record different kinds of utterances for wahad word from the same person, to make sure the test utterance can still be recognized even if it is spoken in a weird way as well as in different times.

3.2.3. Silence Removal

This signal still has noise after the word utterance finished. The form of noise is the speech is voiced or unvoiced.

Voice Activity Detection (VAD), is the technique we used to scan the speech signal from the end to its beginning to determine the presence or absence of speech signal. The main uses of VAD are in speech coding and speech recognition. A VAD may not just indicate the presence or absence of speech, but also whether the speech is voiced or unvoiced, sustained or early, etc. The technique is also used for deleting all values under some specified value which is the noise value.

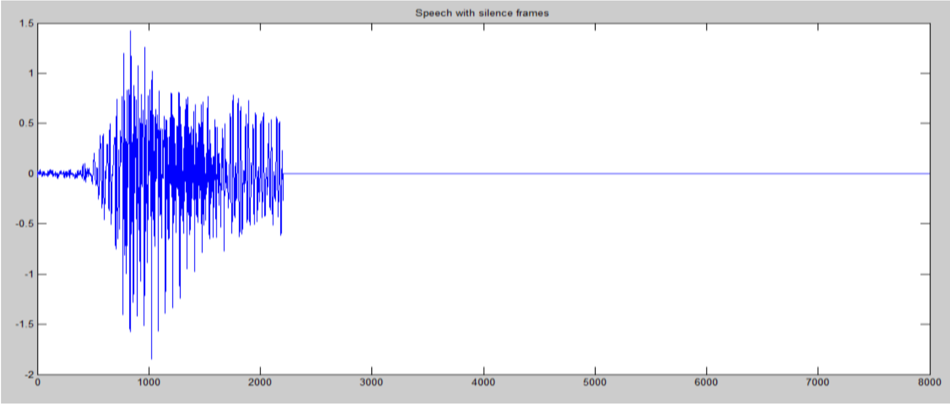


Figure 5-a: The word Waheed after detecting and determine the additional noise

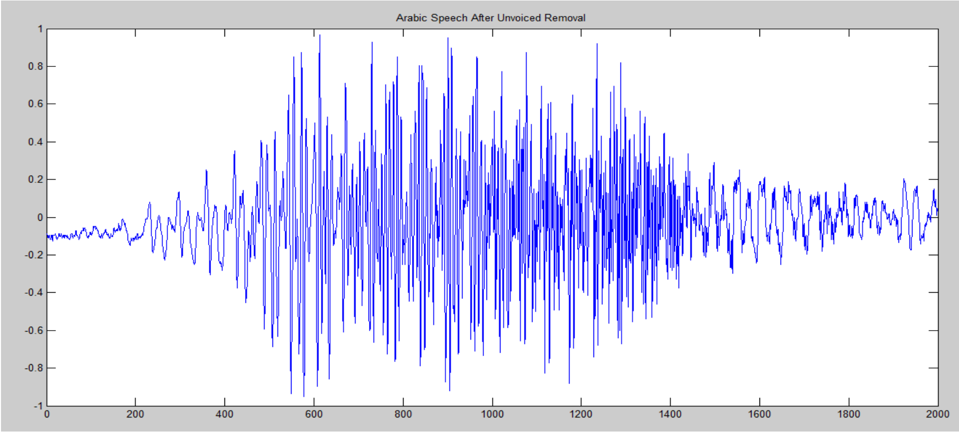


Figure 5-b : The word whaheed After Unvoiced Removal

3.3. FEATURE EXTRACTION

There are a wide range techniques available for features extraction of speech signals. The most prominent ones are Linear Predictive Coding (LPC) and Fast Fourier transforms (FFT) and Mel Frequency Cepestral Coefficients (MFCC). The LPC technique is based on finding the coefficients of a linear predictive filter. MFCC is perhaps the best known and most popular, and will be described in this paper. The popularity of this method can be explained by the low computational cost compared to FFT and LPC based techniques [13,14,15,16,17,18,19].
The main objective of this stage is to extract the important features that are enough for the Recognizer to recognize the words. The Algorithm we will use in our project is MFCC. Stages of feature extraction:

1) Rectangular window (Hamming window) filtering.
2) Spectrum computation (using fast Fourier transform [FFT]).
3) Power spectrum computation.
4) Mel frequency scale mapping.
5) Mel power spectrum computation.
6) Cepstrum computation.
7) Cepstral filtering.

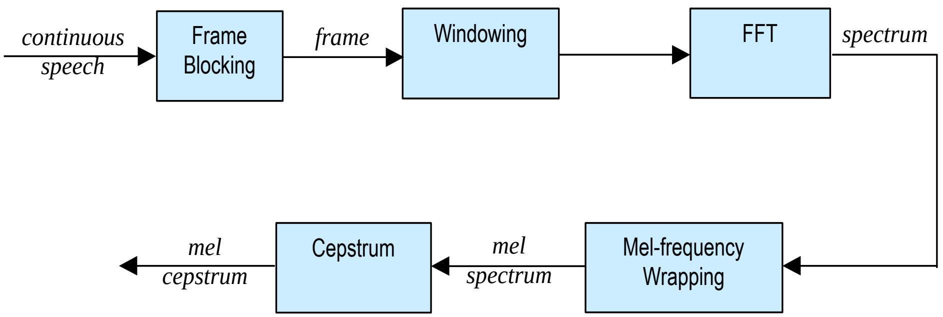


Figure 6: The block diagram of MFCC algorithm

1) Rectangular window (Hamming window) filtering

Multiplication of the signal by a window function in the time domain is the same as convolving the signal in the frequency domain. Rectangular window gives maximum sharpness but large side-lobes (ripples) - hamming window blurs in frequency but produces much less leakage.

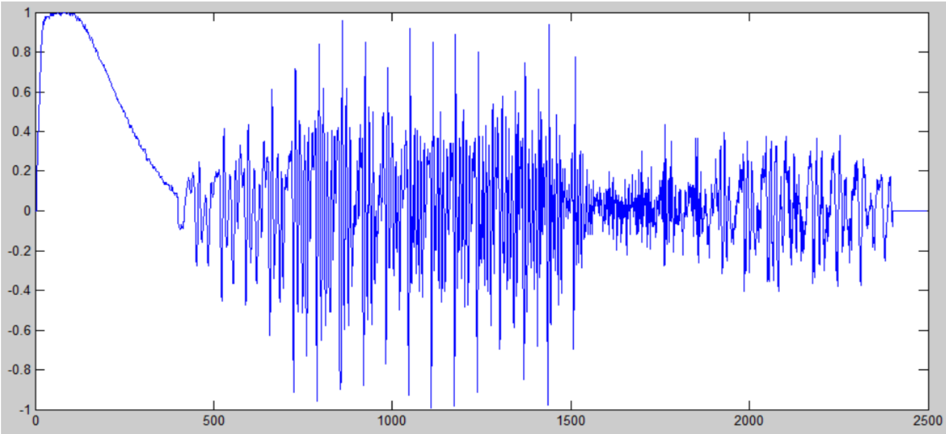


Figure 7: speech Vectors after hamming appling

2) Spectrum computation (using fast Fourier transform [FFT])

We computed the Fast Fourier transform (FFT, which gives us the discrete, complex–valued short term spectrum of the speech signal), the FFT is more faster and accurate than the discrete Fourier transfer.

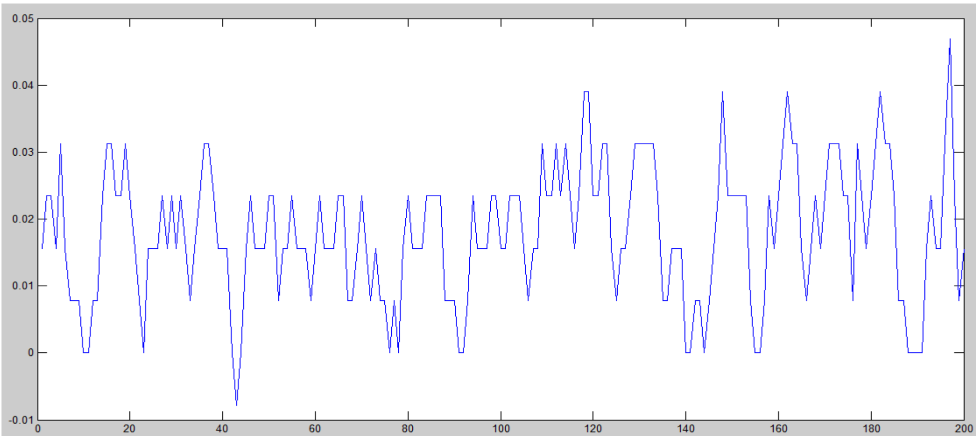


Figure 8: Frame Length and plot FFT Features

3) Power spectrum computation

After computing the V(n) we take the absolute value |V(n)| and then take the power $|V(n)|^2$. Then we have the Power Spectrum = |V (n)|2.

4) Mel frequency scale mapping.

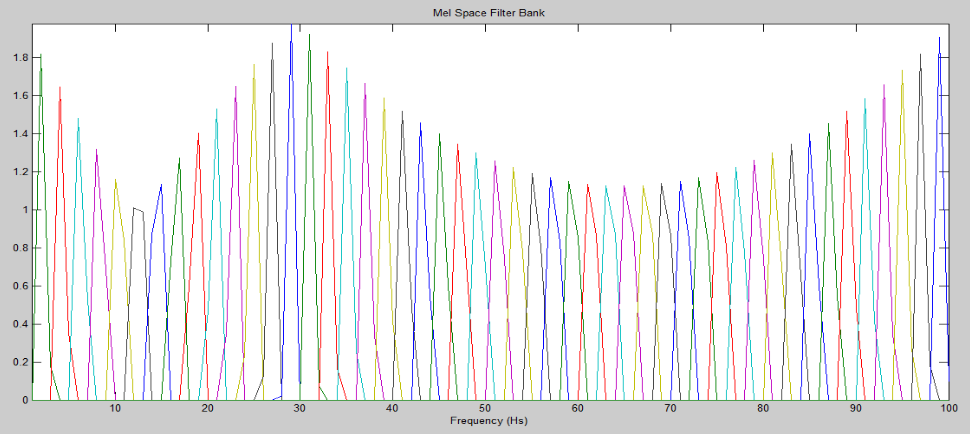


Figure 9: Mel Space Filter Bank

As was shown in perception experiments, the human ear does not show a linear frequency resolution but builds several groups of frequencies and integrates the spectral energies within a given group. Furthermore, the mid-frequency and bandwidth of these groups are non–linearly distributed. The non–linear warping of the frequency axis can be modelled by the so–called Mel-scale. The frequency groups are assumed to be linearly distributed along the Mel-scale.

5) Mel power spectrum computation

Mel-filter coefficients: Construction of filter channels with center frequencies linearly distributed along Mel scale. We are going to deal with Mel power spectrum not the original power spectrum, so we will multiply the original power spectrum by the Mel filter coefficients to get Mel power spectrum.

6) Cepstrum computation

Cepstrum is the inverse of spectrum, Cepstrum is in time domain, Cepstrum is obtained by inverse Fourier transform to power spectrum. But we want the Mel Cepstrum, so we will take the log of Mel power spectrum instead of the power spectrum itself and transform it. Since the Mel power spectrum is symmetric "as we explained earlier", the Fourier-Transform can be replaced by a simple cosine transform

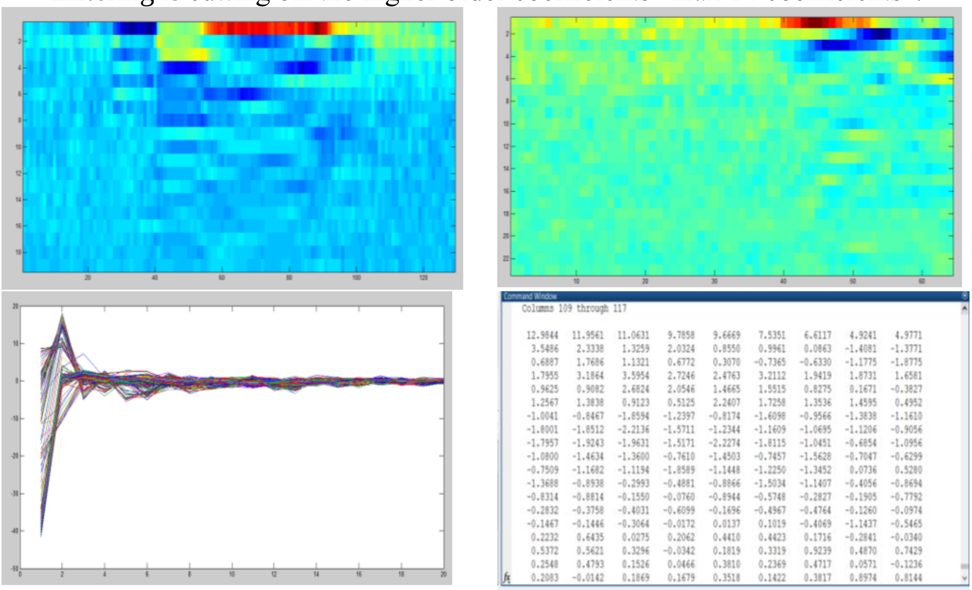
$$C(q) = \sum_{k=0}^{k-1} \log(G(k)) * \cos\left(\frac{\pi * q(2k+1)}{2k}\right) \qquad 8$$

Where C(q) is cepstrum in time domain

7) Cepstral filtering "liftering"

The MFCC are used directly for further processing in the speech recognition system instead of transforming them back to the frequency domain, but we will first apply the liftering process.

Liftering is cutting off the higher order coefficients "max 14 coefficients".



12.9844	11.9561	11.0631	9.7858	9.6669	7.5351	6.6117	4.9241	4.9771
3.5486	2.3338	1.3259	2.0324	0.8550	0.9961	0.0863	-1.4081	-1.3771
0.6887	1.7686	1.1321	0.6772	0.3070	-0.7365	-0.6330	-1.1775	-1.8775
1.7955	3.1864	3.5954	2.7246	2.4763	3.2112	1.9419	1.8731	1.6581
0.9625	0.9082	2.6824	2.0546	1.4665	1.5515	0.8275	0.1671	-0.3827
1.2567	1.3838	0.9123	0.5125	2.2407	1.7258	1.3536	1.4595	0.4952
-1.0041	-0.8467	-1.8594	-1.2397	-0.8174	-1.6098	-0.9566	-1.3838	-1.1610
-1.8001	-1.8512	-2.2136	-1.5711	-1.2344	-1.1609	-1.0695	-1.1206	-0.9056
-1.7957	-1.9243	-1.9631	-1.5171	-2.2274	-1.8115	-1.0451	-0.6854	-1.0956
-1.0800	-1.4634	-1.3600	-0.7610	-1.4503	-0.7457	-1.5628	-0.7047	-0.6299
-0.7509	-1.1682	-1.1194	-1.8589	-1.1448	-1.2250	-1.3452	0.0736	0.5280
-1.3688	-0.8938	-0.2993	-0.4881	-0.8866	-1.5034	-1.1407	-0.4056	-0.8694
-0.8314	-0.8814	-0.1550	-0.0760	-0.8944	-0.5748	-0.2827	-0.1905	-0.7792
-0.2832	-0.3758	-0.4031	-0.6099	-0.1696	-0.4967	-0.4764	-0.1260	-0.0974
-0.1467	-0.1446	-0.3064	-0.0172	0.0137	0.1019	-0.4069	-1.1437	-0.5465
0.2232	0.6435	0.0275	0.2062	0.4410	0.4423	0.1716	-0.2841	-0.0340
0.5372	0.5621	0.3296	-0.0342	0.1819	0.3319	0.9239	0.4870	0.7429
0.2548	0.4793	0.1526	0.0466	0.3810	0.2369	0.4717	0.0571	-0.1236
0.2083	-0.0142	0.1869	0.1679	0.3518	0.1422	0.3817	0.8974	0.8144

Figure 10: Computing MFCC Features in deferent types

3.4. FEATURES TRAINING

Vector Quantization Arabic speech recognition system must able to estimate probability distributions of the computed feature vectors. Storing every single vector that generate from the training mode is impossible, since these distributions 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 distribution. The technique of VQ consists of extracting a small number of representative feature vectors as an efficient means of characterizing the word specific features. By means 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 word acoustic[26].

3.4.1. Saving Trained Features to Database

Saving the feature after trained, the user entered word and the feature vectors saved to our database with specific for using and editing without the aid of any programming algorithms.

3.4.2. Feature Matching

The goal of pattern recognition is to classify objects of interest into one of a number of categories or classes. The objects of interest are generically called *patterns* and in our case are sequences of acoustic vectors that are extracted from an input speech using many techniques described in the previous sections. The classes here refer to similar utterance words. Since the classification procedure in our case is applied on extracted features, it can be also referred to as *feature matching*.

For more efficient we divided data samples to the training set used to derive trained algorithm. The remaining testing set they used for testing and classification algorithm. If the correct classes of the individual patterns in the test set are also known, then one can evaluate the performance of the algorithm. In this section, the VQ techniques will be used, due to ease of implementation and high accuracy. VQ is a process of 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. The collection of all code words is called a codebook [1,2,3,4].

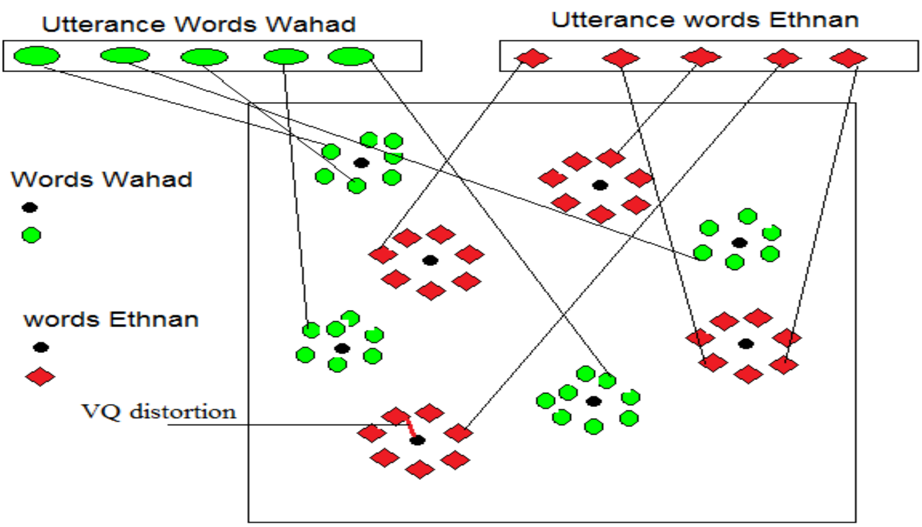


Figure 9. Vector quantization codebook information for Speech Words Features

3.4.3. Clustering the Training Vectors

After the enrolment session, the acoustic vectors extracted from input speech of each features word provide a set of training vectors for that speech. As described above, the next important step is to build a word-specific VQ codebook for each word using those trained vectors. There is a well-know algorithm, namely LBG algorithm [Linde, Buzo and Gray, 1980], for clustering a set of *Xi* trained vectors into a set of *Y*i codebook vectors [2,3].

3.4.4. Check and Matching using (Euclidean Distance)

Euclidean Minimum Distance is more efficient used in the recognition phase, an unknown words's voice is represented by a sequence of feature vector (w1, w2 ….wi), and then it is compared with the codebooks from the database. In order to identify the unknown words, this can be done by measuring the distortion distance of two vector sets based on minimizing the Euclidean distance. The Euclidean distance is the "ordinary" distance between the two points that one would measure with a ruler, which can be proven by repeated application of the Pythagorean theorem. The Euclidean minimum distance between two points P = (p1, p2…pn) and Q = (q1, q2...qn), is given by the formula as following:

$$\sqrt{\sum_{i=1}^{n}(pi-qi)^2} = \sqrt{(p1-q1)^2+(p2-q2)^2+......(pn-qn)^2} \quad 9$$

The word with the lowest distortion distance is chosen to be recognized as it is assigned to the known class.

3.5. RECOGNITION

It is very important to create an acoustical model for the detection of each uttered words. So we created an acoustical model. It is known that different Arabic words are produced by Arabic Speakers vocal cord and different sounds can have different frequencies, power spectral, Cepstrum density measure and MFCC which used to predict for computing different features on each class belong. Speech can be termed as short term stationary so MFCC features were again extracted and words pronounced by the Arabic speakers were detected and classified according to each similarities classes of features the same words and differences of the feature for the different words. In this stage, the data from the tested speaker is compared with codebook of each word vector and measure the difference. These differences are then use to make the recognition decision [26].

4. EXPERIMENT AND RESULT

The Matlab environment offers the several helper functions to ease the development process and functions: wavread, hamming window, fft, Mel_Filter_Bank.m, DCT, MFCC (acoustic vectors), vqlbg.m and Euclidean distance.m, … etc[20]. Next cut the speech signal (a vector) into frames with overlap. The result is a matrix where each column is a frame of N samples from original speech signal. Applying the steps “Windowing” and “FFT” to transform the signal into the frequency domain. Then Computed Mel frequency scale mapping, Mel power spectrum computation and finally Cepstrum computation which means MFCC calculated as the features result. The last result transform to apply the VQ-based pattern recognition technique to build words reference models from those vectors in the training phase and then can be recognize any sequences of acoustic vectors uttered by unknown uttered word.

The distance method used to compute the pairwise Euclidean distances between the code words and training vectors in the iterative process. Train and test programs (which require three functions mfcc, vqlbg and Euclidean distance) to simulate the training and testing procedure in Arabic speech recognition system, respectively.

5. CONCLUSION

The goal of this paper was to create a gender and Arabic speech recognition system, and apply it to a speech of an unknown words. By investigating the extracted features of the unknown speech and then compare them to the stored extracted features for each different word in order to identify the unknown words. The results obtained using MFCC and VQLBG algorithm are considerable for training and build model for each features word. Euclidean distances algorithm computed distortions used for recognition and accuracy obtained was 82.5%. It can be improved by taking voice sample using high quality audio devices in a noise free environment. Use of more number of centroids increases the performance factor but degrades the computational efficiency. Hence an economical trade-off between code vectors and number of computation is required for optimized performance of VQLBG with Euclidean distances algorithms.

6. REFERENCES

[1] 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, Copyright to IJIRSET DOI:10.15680/IJIRSET.2015.0506055 9923

[2] Dr. R.K. Prasad and Mr. Kashyap Patel, "Speech Recognition and Verification Using MFCC & VQ", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013 ISSN: 2277 128X.

[3] Ms. Vrinda ,Mr. Chander Shekhar, "SPEECH RECOGNITION SYSTEM FOR ENGLISH LANGUAGE" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013.

[4] J.S Chitode, Anuradha S. Nigade " Throat Microphone Signals for Isolated Word Recognition Using LPC " International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 8, August 2012. ISSN: 2277 128X.

[5] 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.

[6] Mahdi Shaneh and Azizollah Taheri, " Voice Command Recognition System Based on MFCC and VQ algorithms" World Academy of Science, Engineering and Technology 33 2009.

[7] Vergyri, D., K. Katrin, D. Kevin and A. Stolcke, " Morphology-based language modeling for Arabic speech recognition" Proc. ICSLP, Jeju, 2004, South Korea.

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

[9] Billa, J. et al. Arabic speech and text in tides ontap" , Proc. 2001, HLT.

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

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

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

[13] S.B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", IEEE Transactions on Acoustics, Speech, Signal Processing, Vol. ASSP-28, No. 4, August 1980.

[14] Y. Linde, A. Buzo & R. Gray, "An algorithm for vector quantizer design", IEEE Transactions on Communications, Vol. 28, pp.84-95, 1980.

[15] S. Furui, "Speaker independent isolated word recognition using dynamic features of speech spectrum", IEEE Transactions on Acoustic, Speech, Signal Processing, Vol. ASSP-34, No. 1, pp. 52-59, February 1986.

[16] S. Furui, "An overview of speaker recognition technology", ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 1-9, 1994.

[17] F.K. Song, A.E. Rosenberg and B.H. Juang, "A vector quantisation approach to speaker recognition", AT&T Technical Journal, Vol. 66-2, pp. 14-26, March 1987.

[18] Jamel Price, Sophomore student, Dr. Ali Eydgahi "Design of an Automatic Speech Recognition System Using MATLAB" Chesapeake Information Based Aeronautics Consortium August 2005.

[19] 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.

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

[21] 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), 2428-2436

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

[23] L.R Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, N.J., 1978.

[24] Lubna Eljawad, Rami Aljamaeen, Mutasem K. Alsmadi, Ibrahim Al-Marashdeh, Hayam Abouelmagd, Sanaa Alsmadi, Firas Haddad, Raed A. Alkhasawneh, Mohmmed Alzughoul and Malik B. Alazzam, "Arabic Voice Recognition Using Fuzzy Logic and Neural Network", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 3 (2019) pp. 651-662.

[25] Mr. Kashyap Patel, "Speech Recognition and Verification Using MFCC & VQ", IJAR in Computer Science and Software Engineering , Volume 3, Issue 5, May (2013), ISSN: 2277 128X.