MFCC-VQ APPROACHFOR QALQALAH TAJWEED RULE CHECKING

Ahsiah Ismail [1], Mohd Yamani Idna Idris [2], Noorzaily Mohamed Noor [3], Zaidi Razak[4], ZulkifliMohd Yusoff[5]

[1,2,3,4] Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
[5]Academy of Islamic Studies, University of Malaya, Malaysia

Email: [1]ahsiahismail@um.edu.my, [2]yamani@um.edu.my, [3]zaily@um.edu.my, [4]zaidi@um.edu.my, [5]zulkifliy@um.edu.my

ABSTRACT

In this paper, we investigate the speech recognition system for Tajweed Rule Checking Tool. We propose a novel Mel-Frequency Cepstral Coefficient andVector Quantization (MFCC-VQ) hybridalgorithm to help students to learn and revise proper Al-Quran recitation by themselves. We describe a hybridMFCC-VQ architecture toautomatically point out the mismatch between the students'recitationsandthecorrect recitationverified by the expert. The vectoralgorithm is chosen due to its data reduction capabilities and computationally efficient characteristics.We illustrate our component model and describe the MFCC-VQ proceduretodevelop theTajweed Rule CheckingTool.Two features, i.e., a hybrid algorithm and solely Mel-Frequency Cepstral Coefficientare compared to investigate their effect on the Tajweed Rule CheckingToolperformance. Experiments arecarried out on a dataset to demonstrate that the speed performance of a hybrid MFCC-VQis86.928%, 94.495% and 64.683% faster than theMel-FrequencyCepstral Coefficient for male, female and children respectively.

Keywords: Speech Recognition, Tajweed Recognition, MFCC, VQ

1.0 INTRODUCTION

There have been some notable tools in the market to help proper Quranrecitation, suchas Quran Auto Reciter (QAR) and Quran Explorer. However, these existing tools only displayed the Al-Quran text while the recitation is played with limited functionality such as play, forward, backward and stop. Moreover, these tools are incapable to evaluate the recitation made by the users. In learning the Quran recitation, feedback is integral to the learning process to ensure that recitation and pronunciation are conducted correctly. Therefore, an automated tool that is capable to evaluate the users' recitation will promote the learning of Quran recitation. To the best of the author's knowledge, no real time and automated tools for checking the *Tajweed* were availableon the Internet to assess reading and performance of the userson real time[1]. Thus, there is a deep beliefffor theneed to investigate a suitable speech recognition algorithm for real-time correcting capability towards user recitation of Tajweed Rule Checking.

Speech recognition has been widely studied, not only for English, but also for other languages like Chinese, Thai, Portuguese, Malay and Arabic including the audio recitations of the Holy Al-Quran [2][3]. Variation in the structures of different spoken languages requires different approaches for recognition. Hence, this requires continued vibrant research.It is a known fact that the recitation of Al-Quran is different from normal Arabic readings. Unlike normal Arabic readings, the Quran recitation must follow a set of *Tajweed* rules such as *Harakat*. The *Harakat* includes pronunciations, punctuations and accents components unique from other languages.The characters in the Quran may have a stroke diacritic written either above or below a character,which specifiesthe *Tajweed* rules. This stroke can change the pronunciation and meaning of a word. For example, the *tanween* (ً ٍ ٌ) will indicate that the three vowels (/ a: /, / i :/, / u :/) pronunciation should be followed by the consonant /n/ (/ an: /, / in: /, / un :/). Diacritics are important in forming short vowels[4] and these characters form a very large Arabic vocabulary. The main diacritic symbols are outlined inTable 1. *Tajweed* rules ensure that the recitation is performed with correct pronunciation at a moderate speed without changing the meaning. The pronunciation and recitation for each letter of the Al-Quran should follow its proper characteristic, such as prolongation (*isti'laa*), shorten (*istifāl*), thinning (*tarqeeq*), thickening (*tafkheem*) and other such phonetic methods of absorbing/merging (*idghaam*), clearly mentions (*Idh-haar*), change (*iqlab*), hide

(*ikhfa*) [5]. In addition, a unique *Tajweed* rule known as *Qalqalah* rule requires a slight echo or bouncing sound during recitation [6].This unique pronunciation characteristic has led to additional difficulties and a great challengeto develop automated speech recognition for the Tajweed Rule Checking. Since *Tajweed* is a special art of reading, the recitation of the same verse between reciters may be different. Furthermore, the reciters tend to change their reading by the musical tone *maqams* and echoing sound with some unrest letter produced during the recitation.

There have been some prominent techniques being used by the researcher for feature extraction[7]. Mel-Frequency Cepstral Coefficient (MFCC) is a well-known feature extraction technique in speech recognition and has been successfully adopted in many practical applications[8]. However, in the implementation of the real time applications, performance of the MFCC degrades due to its high computational time [9].The advance modification towards the conventional has to be done in order to improve the speed. In contrast, Vector Quantization (VQ) speeds up the computational time by reducing the data while maintaining the same quality. Since a collection of 6238 verses varies in length were recorded in Al-Quran [10], a fast and robust processing technique is required in the Tajweed Rule CheckingTool to analyse long verses. Therefore, in this study, we proposed thenovel acoustic model,a hybrid MFCC-VQ algorithm for theTajweed Rule CheckingTool. The hybrid MFCC-VQ technique can accelerate the process of the Tajweed Rule Checking as well as maintaining its robustness. We analyseand comparethe hybrid MFCC-VQ with Mel-Frequency Cepstral Coefficient to investigate their performance on speed and accuracy.

This paper is organised as follows. Section 2presents the related works for recognition of Al-Quran recitation.Section 3describesthe proposed method of feature extraction and the reference model. Section 4 explains the data description used in the codebook.The experimental result and analysis are discussed in section 5. Finally, section 6 draws the conclusions.

Table 1. The eight main Arabic diacritics [11].

Symbol	Name
ـَ	*fatha*
ـُ	*dhammah*
ـّ	*shaddah*
ـٍ	*tanween kasr*
ـِ	*kasrah*
ـْ	*sukoon*
ـً	*tanween fath*

2.0 RELATED WORK

Automatic speech recognition (ASR) has been an important area of interest over the past several decades. The ASR is a computer programed or tool to recognise and respond towards the spoken words uttered by the user[12].Even after decades of research and a lot of successfully commercial products deployed, the performance of ASR systems in real deployment scenarios lags behind the performance of a human[13].

Speech recognition technology has been applied in many domains including audio recitations of Holy Al-Quran [3]. Converting existing static based Islamic materials into software and developing a computer based training program becomes a *fardu kifayah* to Muslim [14]. Even though learning of the Al-Quran through the conventional method is the best and the most efficientpractise;however thelearners are restricted with the time constraint to learn through conventional methods. Conventional practices need more time allocation and proper arrangement for the classes, which made it less-efficient for modern people that are mostly mobile[15].

There are numerous feature extraction and feature classification techniques for speech extraction. A variety of different algorithms for feature extraction has been used by the researcher such as Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP) and Mel-Frequency Cepstral Coefficient (MFCC). Algorithms that have been widely used for feature classification include Artificial Neural Network (ANN), Hidden Markov Model (HMM) and Vector Quantization (VQ), with adequate accuracy in the classification process for speech [16].Razak etal., (2008)discuss a number of advantages and disadvantages for each method. Their study showed that the most suitable technique for feature extraction is MFCC. MFCC are cepstral coefficients based on known human auditory perception [17]**.** The MFCC had been the most accurate technique for speech recognition and speaker verification[8][18].

Recently, there has been several research related to Al-Quran or Arabic speech recognitionthat implements the LPC techniques to recognise the Arabic phoneme. LPC analyses the speech signal by estimating the formant frequencies[19]. The main advantages of LPC is thatthe estimation of the speech parameter is accurate and efficient for a computational model of speech[20]. However,G. Mosa*et.al*found that the implementation of LPC techniques towards the Arabic phoneme had not produced the desired recognition accuracy[21]. This is due to the characteristic of the LPC method, which is too sensitive towards quantization noise [17]. In LPC algorithm, converting LPC coefficients to cepstral coefficient process can reduce the sensitivity of high order and low order cepstral coefficient to noise. In addition, due to deduce the prosodic rules, LPC synthetic model may affect and bring a negativeresult for the research [1].

Another technique used for feature extraction is Perceptual Linear Prediction(PLP) analysis using the Bark scale and autoregressive all poles modelling. The spectral scale is the non-linear Bark scale. PLP combines some of the engineering calculations to select characteristics of human hearing [22]. The characteristic of the spectral features is smoothed within the frequency bands [1]. However, the time scale of several interesting speech features is not uniform. The PLP algorithm isnearly similar to the LPC in terms of performance in accuracy [23].

Among these techniques, the most successful is Mel-Frequency Cepstral Coefficient (MFCC). MFCC is the most popular, effective prevalent and dominant for spectral features extraction method and most of the researchers had used this method as their feature extraction[24][25][26]. MFCC are cepstral coefficients is based on known human auditory perception [17]. MFCC applies FFT and DFT algorithm [1]. MFCC hasbeen used as feature extraction techniques to convert voice signals into a sequence of acoustic feature vectors.

Most research has adopted MFCC. The Automated Tajweed Checking Rules Engine for Quranic Learning had been introduced using the MFCC and HMM to help the learner to recite Al-Quran by using an interactive way of learning [27]. This system will assist the student to recite the Al-Quran without the presence of a teacher. The recitation of the user will be processed through the system and the recitation will be revised. However, the engine of the system can be tested only on selected *Tajweed* rule of sourate*Al-Fatihah*.

The E Hafiz system acts like a Hafiz expert to assist Al-Quran recitation learning by minimizing mistakes and errors during the recitation process. The Hafiz is the expert reader that acts like a teacher to listen and teach the right *Tajweed* recitation [28]. However, this system can only be used for those who already know *Tajweed*. The system currently operates in offline mode. During the recitation process, if the user makes any mistakes, the system will not identify it concurrently [3].

Wahidah *et.al* proposed the application of *makhraj* recognition[29]. This system uses MFCC for feature extraction and Mean Square Error (MSE) for feature classification. Even though the system shows promising results in recognizing the correct pronunciation of*makhraj* in one to one mode, however it has an obstacle for one to many recognitions. Besides, the system only focuseson the *makhraj* part of the Al-Quran recitation.

T. Hassan*et.al* developed another system for *Tajweed* Al-Quran recognition using the sphinx tools [30]. In this system, MFCC is used as feature extraction to extract verses from the audio files and HMM as feature classification. Overall, the system recorded 85-90% accuracy when tested offline on a small chapter of Al-Quran.Generally, the errors obtained in this system were due to audio file compression and poor output quality during recording process. This work has demonstratedthat the Al-Quran automated delimiter is possible to be developed but it is time consuming to be implemented for all sourate in Al-Quran andfor various reciters.

Although the use of MFCC in this earlier work shownremarkable results and the most accurate method [8], however, in terms of speed performance, the MFCC method is slow [9]. Based on the literature review conducted, prior research on the Al-Quran recitation checking tools is mostly concentrated solely on accuracy performance.There is no evaluation on the processing time has been reported in the earlier work.

In our work, we try to improve the recent work related to Quranic speech recognition using the hybrid approach. The response time is very important factor for real time application [31]. Since the processing time is slow for the conventional MFCC algorithm, we investigate the data reduction algorithm toaccelerate the execution time by diminishing the quantity of information.In order to maintain or at least without a considerable decrease the classification accuracy, we try to find the most optimum algorithm that can represent the complete training set by some representatives as effectively as possible.

There have been numerous data reduction techniques such as Hart's Condensing,Chen's Algorithm, Reduction by Space Partition,Vector Quantization(VQ),Nearest Neighbour Efficient Search Techniques, the k-Dimensional Tree,Vantage Point Tree, Fukunaga and Narendra Algorithm, Geometric Near-neighbour Access Tree and Linear Approximation and Elimination Search Algorithm[32]. Among these techniques, the most widely used in speech recognition is VQ [32]. 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 centre called a codeword. The collection of all codewords is called a codebook. VQ can reduce the size byremoving the redundancy that makes it effective use for nonlinear dependency and dimensionality by compression of speech spectral parameters [33]. Generally, the use of VQ in a lower distortion than the use of scalar quantization isat the same rate. Besides, VQ is an efficient technique for data compression and good strategy in reducing the data by approximately minimise the misclassification error. Moreover, VQ techniques give more data compression while maintaining the same quality[31]. Further, it saves time during the testing phase[34]. Hence it is a suitable algorithm for accelerating the process of finding the best matching codeword to be used for real time applications[35].

Based on the review conducted, the most relevant techniques to be used for the Tajweed Rule Checking are the hybrid MFCC and VQ which have been referred as MFCC-VQ hybrid models. We proposed a hybrid MFCC-VQ algorithm to speed up the execution time while maintain or at least without a considerable decrease the classification accuracy of the Tajweed Rule Checking. In our work, since the MFCC is the most accurate technique, we used the MFCC to convert voice signals into a sequence of acoustic feature vectors and the VQ is used to cluster the coefficient feature vector extracted gain to their sound class in order to reduce the execution time.The best of our knowledge, this is the first time that a hybrid MFCC and VQ are successfully applied to *Tajweed* speech recognition problems. We choose the VQ algorithm due to its data reduction capabilities and computationally efficient characteristics. VQ is a lossy data compression, method based on the principle of block coding [36]. It is utilised to preserve the prominent characteristic of data. VQ is one of the ideal methods to map huge amount of vectors from a space to a predefined number of clusters, each of which is defined by its central vector or centroid. In this paper, the performance of speed and accuracy of the proposed hybrid MFCC-VQ and solely MFCC algorithm is evaluated.

3.0 METHODOLOGY

This research is concentrated on the *TajweedQalqalah* rule. The *Qalqalah* is pronounced with an echoing or bouncing sound when the letter carries a small circle over the letter (*sukoon*). The quality of *Qalqalah* is found in the five letters which is *Qaaf*, *taa*, *baa*, *jeem* and *daal* (ق, ط, ب, ج and د) when they carry a *sukoon*. There are two types of *Qalqalah:Sughrah* (minor)and *Kubrah* (major). For the *QalqalahSughrah*, the *Qalqalah* letter appears in the middle or end of a word, while for the *QalqalahKubrah*, the *Qalqalah* letter appears at the end of a word and the reader stops on it (for whatever reason), the *Qalqalah* sound is at its clearest or strongest.As we mentioned previously, the pronunciation of *Qalqalah* needs to produce some echoing or bouncing sound. This leads to the frequency variation across a phoneme. Therefore, we extract frequency information of the *Qalqalah*phoneme at discrete time instants. The signal is split into 10 non-overlapping segments with approximately the same length. For each segment, a pitch frequency is calculated to serve as a feature. Therefore,it will end up with 10 features for each phoneme. The initial segment of*Qalqalah*signal contains either low or high frequency components that arebeing dependent on the phoneme. However, the final segment of the *Qalqalah*phonemecontains only low frequency component due to theslight echoing bouncing sound. The feature of the phoneme *Qalqalah* is achieved by using the MFCC with the 10 coefficients.

3.1 Block Diagram of the System

Tajweed Rule CheckingToolconsists of four stages. The first stage is pre-processing, feature extraction, training and testing, feature classification and pattern recognition. The architecture of the Tajweed Rule CheckingToolwhich is based on a hybrid MFCC-VQ,is illustrated in Fig. 1and Fig. 2

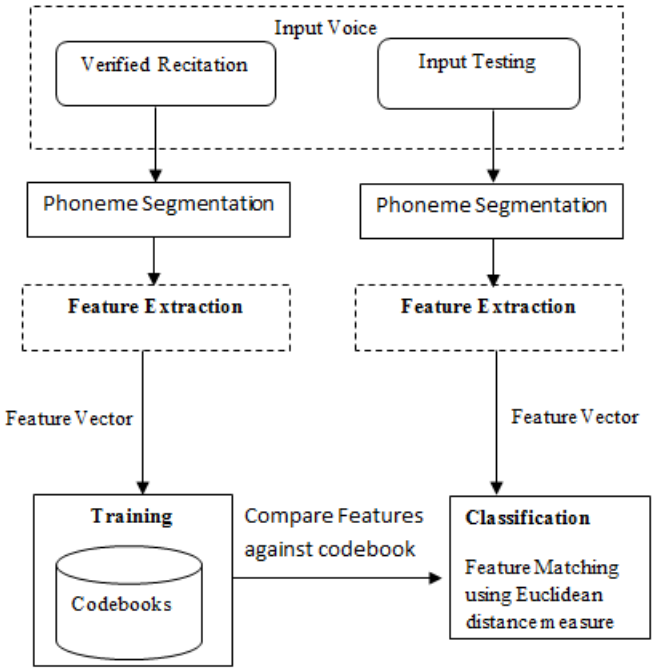

Fig. 1. Tajweed rule checking tool.

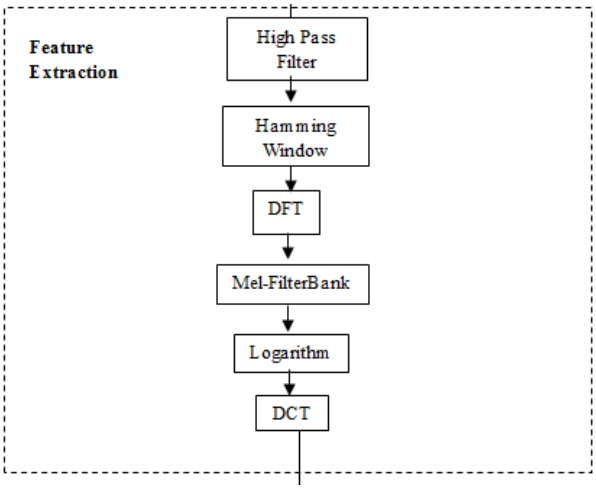

Fig. 2. Feature extraction.

Tajweed Rule CheckingTool consists of four main modules:

1) Input Module: The input speech signal for correct and incorrect recitation of Al-Quran recitation verse from the users is sampled and extracts through segmented features which were gained from the input. Tajweed Rule CheckingTool have two types of input which are verified recitation and input testing.
2) Training Module: The input voice for the verified correct Al-Quran recitation recorded will be stored as a reference model.

3) Testing Module: In the testing module, the input testing will be compared with the reference model and produce the result.
4) Analyse Module: To obtain and calculate the difference range between the test segment and the set of database, which will later be verified for score display.

3.2 Feature Extraction

In this section, we will introduce aMFCC algorithm used in our proposed tool to extract a feature vector. MFCC is a based linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. Generally MFCC consist of seven important steps[37]. Thisincludes Pre-processing, Framing, Windowing, DFT,Mel-Filterbank, Logarithm and Inverse DFT.

3.2.1 Pre-processing

Before performing the recognition step, there are three steps involved in our system during the Pre-Processing stage in order to simplify the recognition task and organisethe data. These three steps are the end point detection, pre-emphasis filtering and channel normalization. For pre-emphasis, we used the high pass filter in MFCC to pass the high frequency signal but reduce the amplitude of the signals with the frequency than cut off frequency. This process will reduce the noise in order to deal with the echoing sound with some unrest letter produced during the Al-Quran recitation process. There were numbers of filtering processes done to eliminate the noise so that it does not affect the recognition process. The high pass filter used in this work is given by the equation below:

$$\tilde{S}(n) = s(n) - as(n-1) \tag{3.1}$$

The $\tilde{S}(n)$ in 1 is the output of the signal where n is a speech signal and a is the pre-emphasis factor. The value of a is typically between zero to one. In our case, the value of 0.95 boosts the higher frequency spectrum in the MFCC model.

3.2.1.1 End point detection

In our proposed system, the end point detection is used to identify the start point and end points of the recorded Al-Quran recitation verse. We used the end point detection in our proposed system to reduce the pre-processing time and remove the noise from the part of silence to assure the perfect recognition performance of our system[38]. During the recording session, after the recording started, the reciter will recite the verse of Al-Quran, then for a few seconds the reciter will leave a silent and noise frame at the beginning and end of a spoken word. Here, the processing is only on the speech samples corresponding to the spoken word, the unnecessary frame was eliminated so that the analysis of the speech signal is more accurate.

3.2.1.2 Pre–emphasis

After extracting all available input for each verse in Al-Quran, the uttered speech goes to a pre-emphasis stage. The purpose of this step is to decrease the noise by increasing the high-frequency signal and altering flat spectrum. The pre-emphasis process will take place to amplifythe area of spectrum to increase the efficiency ofthe spectral analysis[39].

3.2.1.3 Channel Normalized

The channel normalization technique developed using a different application domain. In our work, the voice channel is assumed to be a linear time invariant system, a speech signal, $y(t)$ (referred to as the noisy signal), that has been transmitted over a given voice communication channel can be expressed as:

$$y(t) = s(t) * -h(t) + n(t) \tag{3.2}$$

Where * is a convolution, $s(t)$ the original speech that was sent, $h(t)$ the linear time-invariant system of the channel and $n(t)$ additive noise assumed stationary. The z -transformation of this equation gives:

$$Y(z) = s(z)^{*} - H(z) + N(z) \tag{3.3}$$

In equation(3.3), if the additive noise is known, and is stationary and uncorrelated with the signal, its power spectrum ($|N(z)|^2$) can be subtracted from the power spectrum of the noisy signal ($|Y(z)|^2$). This is the approach that spectral subtraction channel normalization techniques take. In this technique, the additive noise power spectrum at the intervals between the speeches, is estimated then the power spectrum of the signal is subtracted from the estimate value.

3.2.2 Framing

After executing the pre-processing filtering, the filtered input speech is framed. Then from the speech input, we determined the columns of data. Here, the Fourier Transform only reliable when used if the signal is in a stationary position. In our case, the speech implementation holds only within a short time interval which is less than 100 milliseconds of frame rate. Thus, the speech signal will be decomposed into a series of short segments. We analysed each of the frames and extract only the useful ones. In our work, we choose 256 sizefor the window frame.

3.2.3 Windowing

The purpose of windowing is to minimise the spectral distortion and to taper the signal to zero at the start and end of each frame. In the windowing step;for each speech signal framed is windowed in order to minimise the signal discontinuities at the start and end of the frame. The windowing equation used in this work is defined as below:

$$w(n) = 0.54 - 0.46\cos{-(2\Pi n/N - 1)} \tag{3.4}$$

$$where \ldots 0 \le n \le N-1$$

Where $0 \le n \le N-1$and N is the speech sample number in frame, and n is the n th number of speech sample in the frame.

3.2.4 Dft

Normally, the Discrete Fourier Transform (DFT) is derived via the Fast Fourier Transform (FFT) algorithm. In our work, we used this algorithm to evaluate the frequency spectrum of the speech and convertthe time domain to the frequency domain of each frame for N samples. The FFT is defined on the set of N samples $\{X_n\}$as:

$$Y_2[n] = \sum_{k=0}^{N-1} Y_1[K] e^{-2\Pi jkn/N} \tag{3.5}$$

$$where \ldots 0 \le k \le N-1$$

The $Y_2[n]$in the equation above is the Fourier Transform of$Y_1[K]$.

3.2.5 Mel Filterbank

We applied the Mel scale to emphasise the low frequency components of the bouncing sound of the *Qalqalah*. The low frequency component here is referring to the slight echoing or bouncing sound of the *Qalqalah*, which carried more important information compared to the high frequency. Mel scale is a special measure unit of perceived pitch of tone. Mel Filterbank does not linearly correspond to the normal frequency, but it behaves linearly below1000Hz and a logarithmic spacing above 1000Hz. The equation below shows the approximate empirical relationship to compute the Mel frequencies for a frequency f expressed in Hz:

$$\text{Mel}(f) = 2592 * \log 10(1 + f/700) \tag{3.6}$$

To implement the filterbanks, the magnitude coefficients of each Fourier transform of speech segment is binned by correlating them with triangular filter in the filterbank. Mel-scaling is performed using a number of triangular filters or filterbanks.

3.2.6 Logarithm

We obtain the logarithm by converting the values of DFT multiplication into an addition one. We reduced the Mel Filter-bank values by replacing each value by its natural log. We used the Matlab command 'log' to take the logarithm of Mel filtered speech segments. As the human ear is less sensitive in a slight variation of low and high amplitudes, we used a log feature to estimate entail, less sensitiveness in the variation of the input (such as loudness or slower speech).

3.2.7 Inverse DFT

The final step in MFCC is the inverse DFT. Inverse DFT converts from frequency of the speech signal back to time domain. Finally, we get the output of the Mel-Frequency Cepstral Coefficients (MFCC) in the form of feature vector. In order to get the MFCCs, we used the equation below:

$$Y[n] = \sum_{k=0}^{N-1} X[K] \cos[n(k - 1/2)\Pi / N] \tag{3.7}$$

$$n = 1,2,3\ldots, N$$

In the equation above, x[k] is the value for of each Melfiltered speech segment logged obtain from the previous step.

3.3 Reference Model and Comparison of Codebook

Instead of using the conventional MFCC, we used the hybrid Vector Quantization in MFCC algorithm.Here, the Vector Quantization is applied to quantize signal values separately for each of a set or sequence. Vector quantization is used in this work to cluster the coefficient vector extracted from the speech samples on their sound class for each of *Qalqalah* letter, for both *Sughrah* and *Kubrah* (**ق ط د ب ج**). VQ algorithm contains the encoder, decoder and transmission channel. For each of the encoder and decoder, it has an access to a codebook. We used the MFCC feature vector array obtained in the previous phase to generate the reference model (a codebook features) of the speech, and stored it in a database.

A single codebook is designed from a training sequence. The training sequence represents the speech which is supposed to be encoded by the system [7]. For the training data, we sampled the input signal for the verified correct recitation of the *Al-Ikhlas* verse and ten *Qalqalah* phonemeby the expert and extracts through segmented features, which were gained from the input utterances. In our work,a codebook is applied to *Al-Ikhlas*verse and ten phoneme in the recognition database. Each codebook is formed from a training sequence which contains several samples of a single word. The uttered phoneme is passed through each of the vector quantizer in the recognition database. The quantizers will define a set of distortions which is *d1 (t), d2 (t), ... dM(t).* After forming the codebook, wecompared the signal input against each vector in the codebook through a distortion measure. We used theLinde-Buzo-Gray (LBG)algorithm to group the training vectors set into the codebook vectors, set to perform the iterative procedure. The main steps in LBG algorithm used in this work are:

a. Determine the size of codebook (number of codebook vectors, N).
b. Randomly select N codebook vectors.
c. To cluster the vectors, the Euclidean distance measure is used.
d. The clustered value is compared with the input vector iteratively to produce distortion value.
e. The word with the smallest distortion is chosen as the recognised word.

In the testing phase, we extract features from each of the verses recited by the reader. This feature is matched with each of the verses verified recitation which is stored in the database. The average value of each of the expert's recitation (recorded and stored) is calculated using features vector. Then, we calculate the distance between recorded and stored, by taking the difference of these two averages. In order to check the correctness of each verse, the threshold value is matched with each distance. If the distance is less than threshold, the recorded verse is considered correct.

4.0 DATA DESCRIPTION

In this study, the data are recorded from 45 speakers in three categories of reciters which are 20 males, 20 females and 5 childrens. The evaluation of the proposed tool is performed on the recitation of sourate *Al-Ikhlas* and ten phonemes of all the *Qalqalah* letters for both *Sughrah* and *Kubrah*as listed in Table 2 and

Table 3. Since *Qalqalah Tajweed* in sourate *Al-Ikhlas* only involved a letter of *Dal* (د), the remaining letters (ق ب د ط ج) are obtained from phonemes in other sourates.

Table 2.Excerpt from the dictionary of sourate *Al-Ikhlas*

Ayates	Phoneme	The ayates in the Al-Quran	*Qalqalah Tajweed*
*Al-Ikhlas*1	*Qul huwa Allāhu aḥad*	قُلْ هُوَ اللَّهُ أَحَدٌ	*Kubrah*(د)
*Al-Ikhlas*2	*Allahu -ṣ-ṣamad*	اللَّهُ الصَّمَدُ	*Kubrah*(د)
*Al-Ikhlas*3	*Lam yalid wa lam yūlad*	لَمْ يَلِدْ وَلَمْ يُولَدْ	*Sughrah&Kubrah*(د)
*Al-Ikhlas*4	*Wa lam yaku(n)l lahu kufuwan aḥad*	وَلَمْ يَكُنْ لَهُ كُفُوًا أَحَدٌ	*Kubrah*(د)

Table 3.Dataset for Testing *QalqalahSughrah* and *Kubrah.*

Phoneme	*Qalqalah Tajweed*	Sourate
إِذَا وَقَبَ	*Kubrah*(ب)	*Al-Falaq*
قَبْلِكَ	*Sughrah*(ب)	*Baqarah*
الْبُرُوجِ	*Kubrah*(ج)	*Al-Buruj*
تَجْرِي	*Sughrah*(ج)	*Baqarah*
إِذَا حَسَدَ	*Kubrah*(د)	*Al-Falaq*
يَلِدْ	*Sughrah*(د)	*Al-Ikhlas*
مُّحِيطٌ	*Kubrah*(ط)	*Al-Buruj*
أَفَتَطْمَعُونَ	*Sughrah*(ط)	*Baqarah*
مَا خَلَقَ	*Kubrah*(ق)	*Al-Falaq*
رَزَقْنَاهُمْ	*Sughrah*(ق)	*Baqarah*

In data collection, a total of 1740 samples consisting of correct and incorrect recitation for sourate Al-Ikhlas and ten Qalqalahphoneme are collected. Initially, the proposed MFCC-VQ method is trained with 1310 samples of correct recitation. Then, the evaluation of this study is tested using 580 samples consists of correct and incorrect recitations with 180 samples are from Al-Ikhlas verses and the remaining samples are from Qalqalah phonemes.

Further, the samples of Al-Ikhlas verses are recorded from male, female and children. However, children are excluded from the data collection of Qalqalah phonemes due to the pronunciation of these phonemes requires a certain extent of knowledge and skills.

5.0 RESULTS AND DISCUSSION

In this section, results of a hybrid MFCC-VQ and solely MFCC were discussed under two types of conditions, which are evaluation based on speed performance and accuracy. We test our proposed tool using the real-time factor to measure the speed. We choose the real-time factor (RTF) since it has been a widely used metric system for measuring the speed of automated speech recognition. However, the performance analysis for the accuracy is a test using the sensitivity and specificity. This test is chosen because it is more specific, where we can identify the actual positives and negatives that are correctly identified by the proposed tool. The sensitivity is the true positive rate, while the specificity is the true negative rate. In our case, the True Positive (TP) is for the wrong recitation correctly identified as wrong, False Positive (FP) is a right recitation incorrectly identified as wrong, True Negative (TN) is a right recitation correctly identified as right and False Negative (FN) is a wrong recitation incorrectly identified as right. These four values are used to compute performance parameters like Specificity, Sensitivity (SE), Accuracy and Precision. All the formula used is shown in the equation given below:

Real-Time Factor (RTF) = EXECUTION TIME/ RECORDING DURATION (5.1)

Specificity (SP) = TN/ (FP + TN) (5.2)

Sensitivity (SE) = TP/ (TP + FN) (5.3)

Accuracy (Acc) = TP +TN/ (TP+FP+TN+FN) (5.4)

Precision (Pre) = TP/ (TP+FP) (5.5)

Improvement Real-Time Factor (%) = ((RTF MFCC-VQ – RTF MFCC) / RTF MFCC)* 100 (5.6)

To evaluate the proposed MFCC -VQ algorithm , we conducted a series of experiments on different type of reciters. The result for the acoustic model of the speech signal for the verse قُلْهُوَاللَّهُأَحَد (QulhuwaAllāhuaḥad) before and after applying a high pass filter are shown in Fig. 3. The figure shows that after the pre-emphasis sounds, the speech is sharper with a smaller frequency volume. While Fig. 4 shows the logarithmic power spectrum. In this figure, the areas containing the highest level of energy are displayed in red. As we can see on this figure, the red area is located between 0.5 and 2.5 seconds. The figure also shows that most of the energy is concentrated in the lower frequencies (between 50 Hz and 1 kHz). It shows that the majority of the information contained the slight echoing bouncing sound in the lower frequencies. The result for 2D acoustic vectors for sourates Al-Ikhlas and each of Qalqalah sound (ق د ط ب ج) for both QalqalahSughrah and Kubrah are presented in Fig. 5 and Fig. 6.The figure below shows that the recitation for verse Al-Ikhlas and ten Qalqalah phoneme recited by one of the male reciters. From these two figures, it can be seen that most of the areas are not overlapping exclusively, but certain regions seem to be overlapped by one or the other verse and phoneme. This allows us to distinguish the difference of each of the verse and phoneme. On the other hand, for the VQ codebook for Al-Ikhlas recitation can be seen Fig. 7. Here, there are four codeword vectors in the Al-Ikhlas codebook. Feature vectors for each verse are stored in its codeword vector inside the codebook. Fig. 8 shows the VQ codebook for the pronunciations of ten phoneme for each of Qalqalah sound (ق د ط ب ج)for both QalqalahSughrah and Kubrah. From this figure, it shows that phoneme Qalqalah codebook consists of ten codeword vectors. For each of the Qalqalah letter it consists of two codeword vectors which are for QalqalahSughrah and Kubrah. Since there is five letters of Qalqalah and each of the letters consist of two types of rule, which are Sughrah and Kubrah, so there will be ten of the codewords involved in the phoneme Qalqalah codebook, as shown in Fig. 8. Here, in these two figures, they show that each of the large feature vectors is grouped approximately to the same number of points that are close to it. This process is done in order to speed

up the execution time by reducing the storage and the computation for determining similarity of spectral analysis.

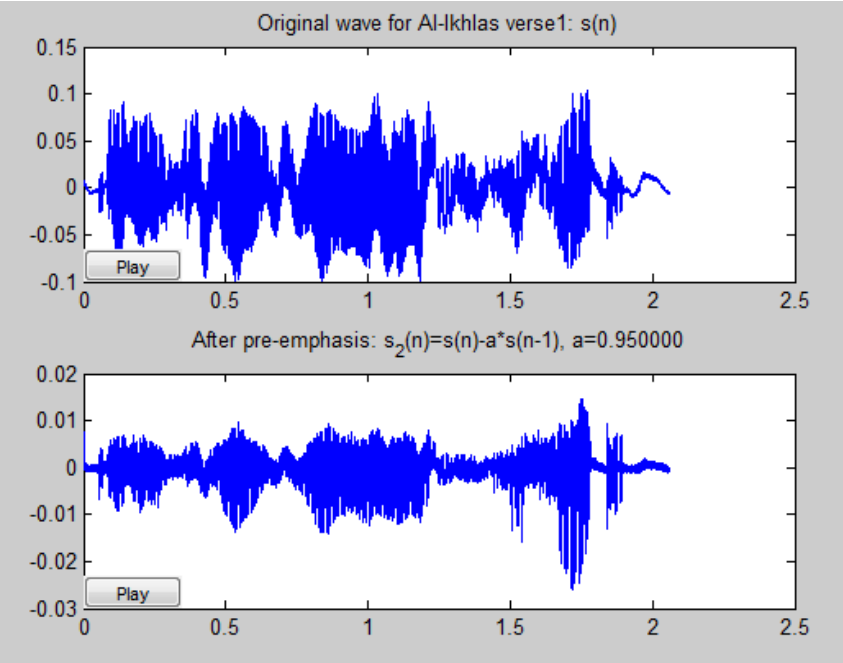

Fig. 3.Example of high pass filter of voicing for the verse قُلْ هُوَ اللَّهُ أَحَد(*QulhuwaAllāhuaḥad)*.

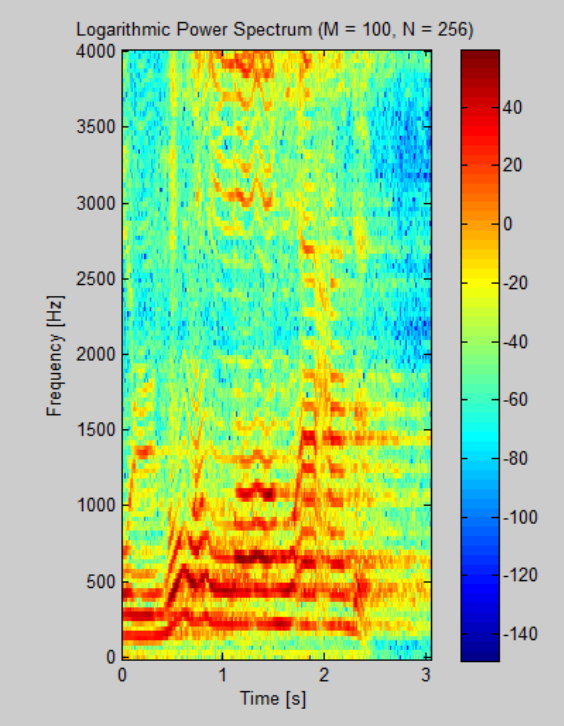

Fig. 4. Logarithmic power spectrum.

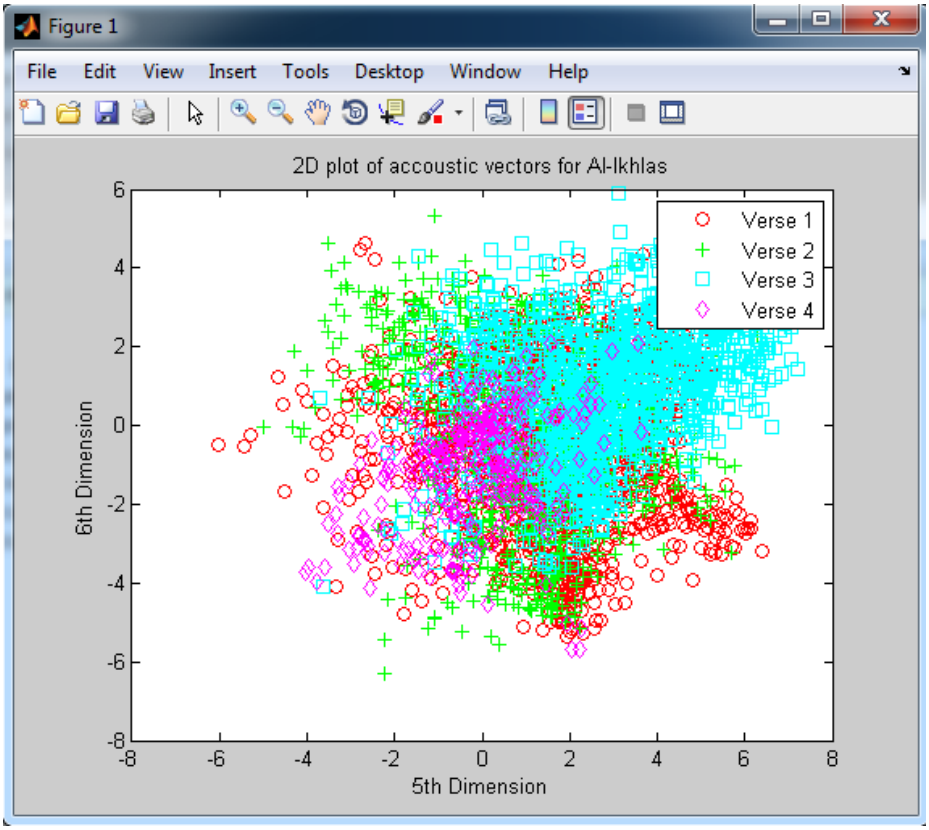

Fig. 5. The 2D acoustic vectors for *Al-Ikhlas.*

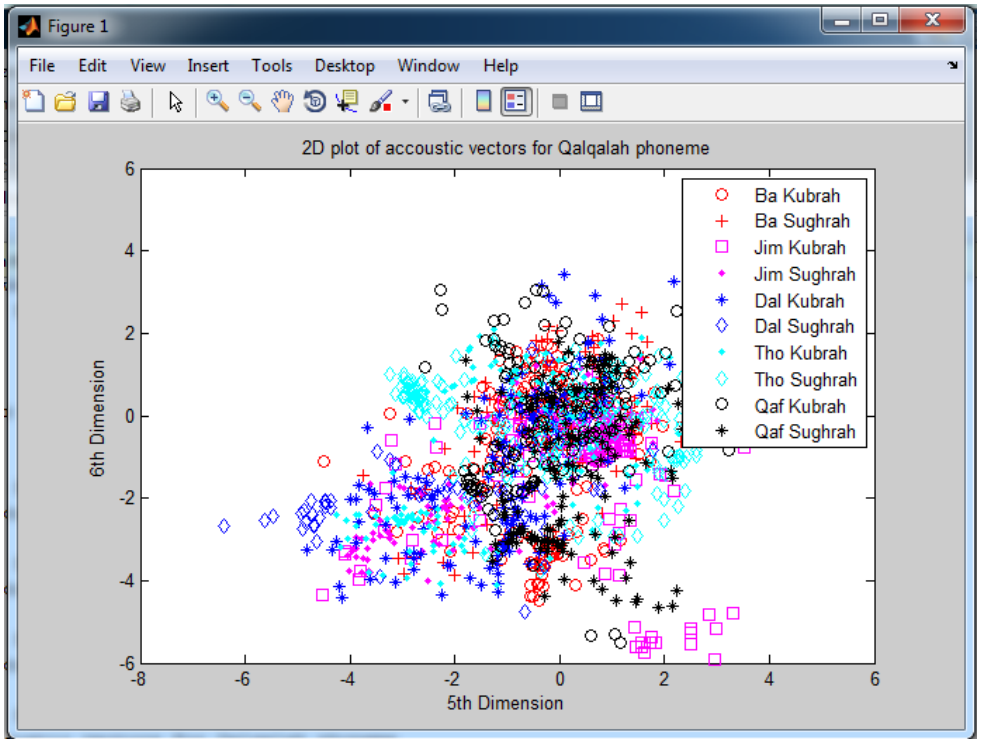

Fig. 6. The 2D acoustic vectors for *Qalqalah* phoneme.

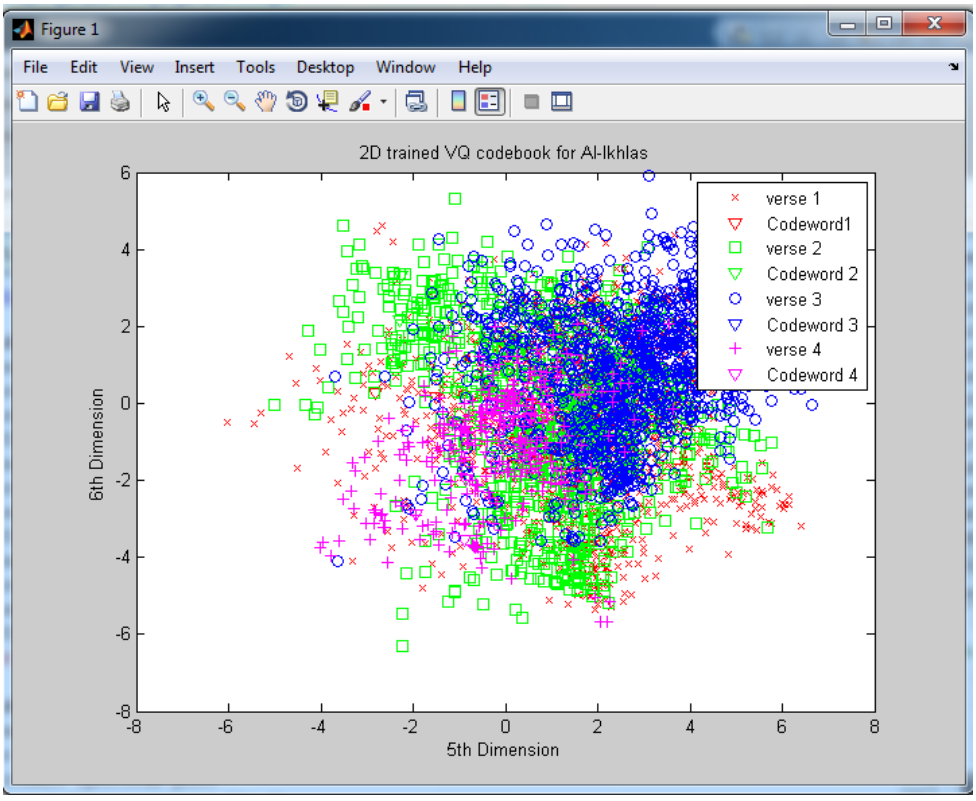

Fig. 7.The 2D VQ codebook for *Al-Ikhlas.*

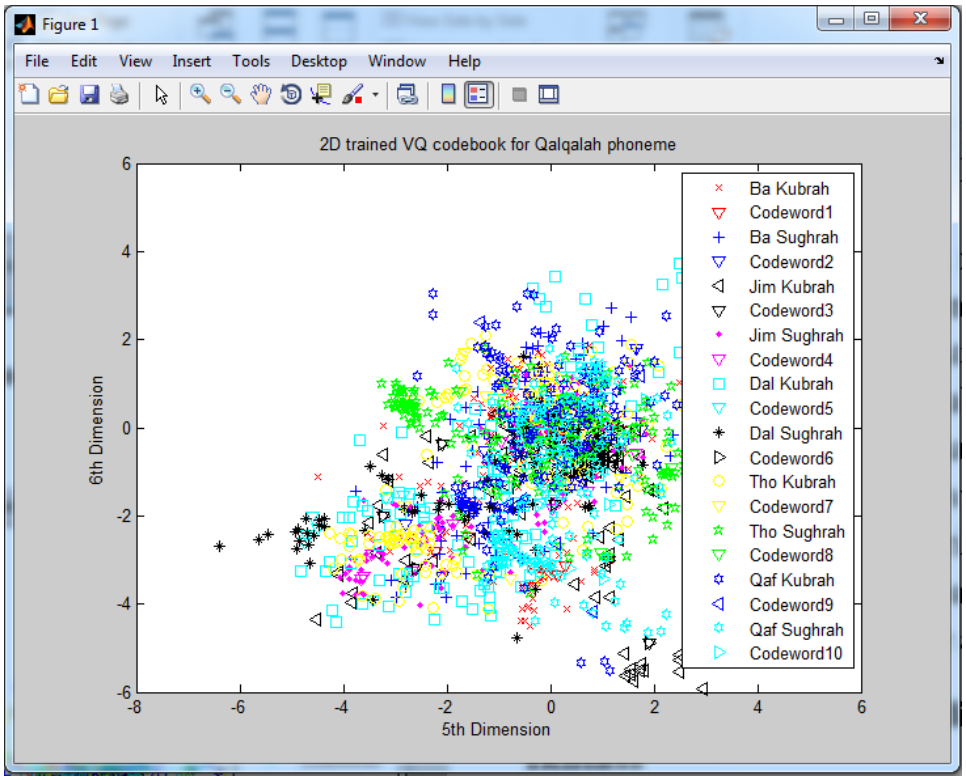

Fig. 8. The 2D VQ codebook for *Qalqalah* phoneme.

For further analysis, we evaluated both hybrid MFCC-VQ and a conventional MFCC algorithm. Their performances arecompared using the metrics defined in (5.1) - (5.6) for each *QalqalahKubrah* and *Sughrah* verses and phonemes as shown in Table 4. In this analysis, the execution time recorded from the proposed MFCC-VQ outperform the conventional MFCC for all categories of reciters. This can be described by the low real-time factor obtained for the proposed MFCC-VQ indicating the proposed study is processed in real-time [40].The execution time recorded in the Table 4 is based on the execution time written in Matlab.

The overall real-timefactor obtained in our proposed method are 0.156, 0.161 and 0.261 for male, female and children respectively. This implies that the proposed tool is processed in real-time. Further, the real-time factor of our proposed hybrid MFCC-VQ is compared with a conventional MFCC algorithm whereby the real-time factor recorded for male, female and children are 1.192, 2.928 and 0.740 respectively. From these table, it can be observed thatthe speed performance of the proposed MFCC-VQ method is faster than the conventional MFCC by 86.928%, 94.495% and 64.683% of real-time factor improvement for male, female and children respectively. However, variation of real-time factor improvement can be observed for different categories of reciters.

In our case, the recitation pace of each reciter may influence the real-time factor. As can be seen in Table 4, there is a slightly different in real-time factor improvement between the male and female reciter. Even though number of the testing sample for male and female reciter are the same, however the recording duration for female reciters is 1.6 times longer compared to the male reciter. Most of the female reciters tend to recite at a very slow pace compared to male reciter. The major difference in the recording duration for the same dataset*Al-Ikhlas* and phoneme is due to the flexibility in reciting of Al-Quran at the moderate pace speed[1].

On the other hand, as can be seen inTable 4, the real-time factor improvement for the children reciter is a bit lower compared to male and female reciter. This is because the dataset of testing sample for the children reciter is different compared to male and female reciter. The dataset for the testing phase for the children reciter only be tested using sourate *Al-Ikhlas*. This may also affect the real-factor obtained in this study. However, in our work, we only observed and investigated the comparison result between the proposed hybrid MFCC-VQ and the conventional MFCC. As can be seen clearly in Table 4, our proposed method is faster than the conventional method for all categories of reciter in classifying the *Qalqalah* rule. Hence, the variations in the results for these three categories of reciters does not affect our findings.

Table 4. Comparison of hybrid MFCC-VQ and MFCC.

METRICS	**MALE**		**FEMALE**		**CHILDREN**	
	MFCC-VQ	**MFCC**	**MFCC-VQ**	**MFCC**	**MFCC-VQ**	**MFCC**
TN	155	151	162	166	15	15
TP	80	116	68	93	4	2
FP	3	7	15	8	0	0
FN	42	6	35	13	1	3
Specificity	0.981	0.956	0.915	0.954	1.000	1.000
Sensitivity	0.656	0.951	0.660	0.877	0.800	0.400
Accuracy	0.839	0.954	0.821	0.925	0.950	0.850
Precision	0.964	0.943	0.819	0.921	1.000	1.000
Execution Time (s)	46.259	353.890	75.606	1373.394	14.109	39.950
Recording Duration (s)	297.000	297.000	469.000	469.000	54.000	54.000
Real-Time Factor	0.156	1.192	0.161	2.928	0.261	0.740
Real-Time Factor Improved (%)	86.928		94.495		64.683	

Besides the real-time factor evaluation, the analysis on the accuracy towards the proposed MFCC-VQ algorithm also shows the promising recognition rate for all categories of reciter as summarize in Table 4. The overall recognition accuracy achieved in our proposed MFCC-VQ method are 83.9%, 82.1%, and 95.0%. These results indicate that the proposed MFCC-VQ method can be used for recognition of TajweedQalqalahSughrah and Kubrah for all categories reciter.

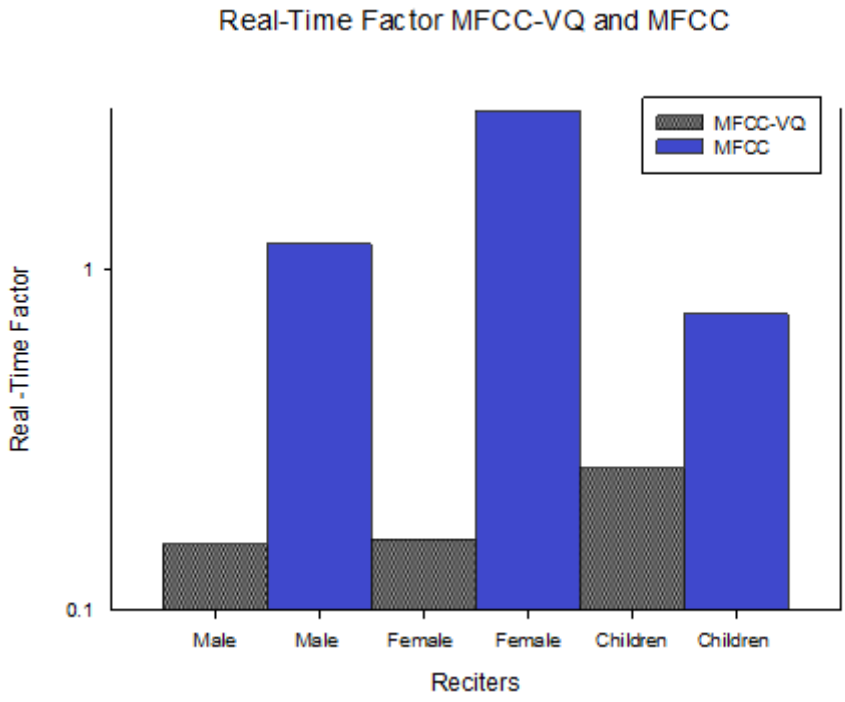

Fig. 9. Comparison graph real-time factor for MFCC-VQ and MFCC.

The graph of a real-time factor and accuracy values for both QalqalahSughrah and Kubrah obtained are shown in
Fig. 9 and Fig. 10. From the comparison results as shown in the graph below for both types of evaluation, it is clear that our proposed algorithm MFCC-VQ is better than the conventional MFCC in terms of its speed performance.

To further understand the effect of the different types hybrid approaches towards the Quranic recitation, a comparison of the similar studies involving hybrid method for Quranic recitation is summarized in Table 5. However, none of these studies reported the speed performance of their approach. Most of the Quranic tools reported in Table 5 show a high recognition accuracy. A hybrid method of the MFCC and MSE for the Makhraj recognition obtained the higher accuracy which is 100% when it is being tested to one to one mode reciter. While the MFCC and MLP hybrid method which being tested in the TajweedQalqalahKubrahobtain the accuracy of 95 to 100%. Besides, the MFCC and HMM obtained the accuracy between the 86.4 to 91.95% when it is being tested in the Tajweed rule in Sourate Al-Fatihah.

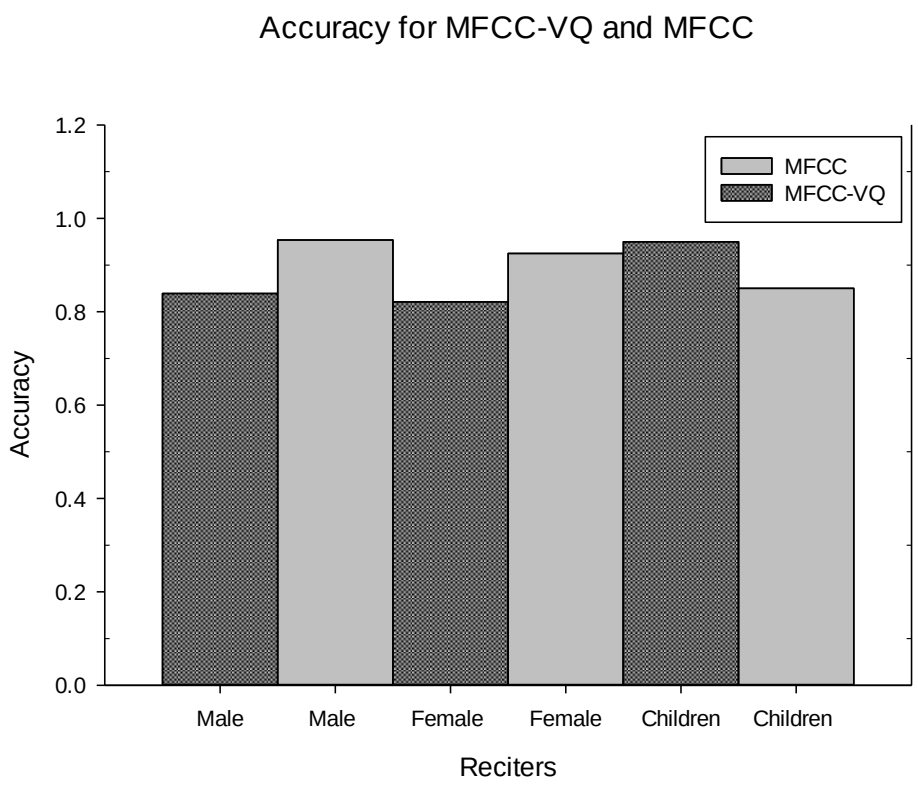

Fig. 10. Comparison accuracy graph for MFCC-VQ and MFCC.

As can be seen clearly in the Table 5, there is only one of the previous study that are closely related to our work [6]. However in this study, the tool is only tested on the five phoneme of the QalqalahKubrah dataset. Note that our proposed hybrid MFCC-VQ method of the accuracy between 82.1 to 95% are being tested using the TajweedQalqalahSughrah and Kubrah for sourate Al-Ikhlas and ten Qalqalah phoneme. Even though there is a slightly lower in accuracy on our MFCC-VQ proposed method, unlike the prior research which only evaluate on the accuracy performance, our proposed hybrid method investigate on both accuracy and the speed performance to ensure the proposed tool are process in a real-time mode. The hybrid MFCC-VQ proposed tools show the promising result as shown in Table 5 with the real-time factor obtained between 0.156 to 0.261. This indicates that the proposed tool using the hybrid MFCC-VQ are processed in a real-time mode.

Table 5. Performance of Quranic Recitation using other hybrid methods.

Hybrid Method	**Dataset Tested**	**Accuracy**	**Real-Time Factor**
MFCC and MSE [29]	*Makhraj*	100%	Non Reported
MFCC and MLP [6]	*Tajweed Qalqalah Kubrah*	95-100%	Non Reported
MFCC and HMM [27]	*Tajweed* in Sourate *Al-Fatihah*	86.4 - 91.95%	Non Reported
MFCC-VQ (proposed method)	*Tajweed Qalqalah Sughrah* and *Kubrah* in Sourate *Al-Ikhlas* and *Qalqalah* phoneme	82.1-95%	0.156-0.261
MFCC and ANN [41]	*Tajweed* in Sourate *An-Nas*	72-93%	Non Reported

6.0 CONCLUSION

In this paper, we investigate the hybrid MFCC-VQ techniques for Tajweed Rule Checking Tool to recognise the types of bouncing sound in *QalqalahTajweed* rule, i.e., the *QalqalahKubrah* and *QalqalahSughrah* for each **ق ط د ب ج**. The result for hybrid MFCC-VQ is compared with conventional MFCC. Overall, using the hybrid MFCC-VQ provides us with 86.928%, 94.495% and 64.683% real-time factor improvement over the conventional MFCC when the same testing sample is used for male, female and children. Although there is a slight difference in the percentage of improvement for the real-time factor between these three categories of reciter, in our work, we restricted ourselves to observe only the comparison of the real-time factor between the hybrid MFCC-VQ and conventional MFCC. From the result, it shows that the MFCC-VQ achieves substantially better results in terms of speed performance than conventional MFCC. We demonstrate how this algorithm can recognise sequences and perform recognition. We believe our work on a hybrid MFCC-VQ is only the first step towards a more powerful acoustic model for Tajweed Rule Checking Tool. There is still room for the improvement in the future work. Another hybrid technique should be tested on the *Qalqalah Tajweed*rule for the future work.

ACKNOWLEDGEMENT

Special thanks to University of Malaya for providing support and material related to educational research. This research was supported by UMRGgrant RP004A-13HNE.

REFERENCES

[1] Z. Razak, N. J. Ibrahim, MYI, Idris, E. M. Tamil, and M. Yakub, "Quranic Verse Recitation Recognition Module for Support in j-QAF Learning : A Review," vol. 8, no. 8, 2008.

[2] M. Mustafa and R. Ainon, "Prosodic Analysis And Modelling For Malay Emotional Speech Synthesis," *Malaysian J. Comput. Sci.*, vol. 23, no. 2, pp. 102–110, 2010.

[3] W. M. Muhammad, R. Muhammad, A. Muhammad, and A. M. Martinez-Enriquez, "Voice Content Matching System for Quran Readers," *2010 Ninth Mex. Int. Conf. Artif. Intell.*, pp. 148–153, Nov. 2010.

[4] R. A. Haraty and O. El Ariss, "CASRA+: A Colloquial Arabic Speech Recognition Application," *Am. J. Appl. Sci.*, vol. 4, no. 1, pp. 23–32, Jan. 2007.

[5] F. Barakatullah, "Q Read A step by step guide to learning how to Understanding Tajweed," 2013. [Online]. Available: http://www.qfatima.com/.

[6] H. A. Hassan, N. H. Nasrudin, M. N. M. Khalid, A. Zabidi, and A. I. Yassin, "Pattern classification in recognizing Qalqalah Kubra pronuncation using multilayer perceptrons," in *2012 International Symposium on Computer Applications and Industrial Electronics (ISCAIE)*, 2012, no. Iscaie, pp. 209–212.

[7] Z. A. Othman, "Malay speech to Jawi text engine," no. 2011, 2011.

[8] M. a. N. R. Rahaman, a. Das, M. Z. Nayen, and M. S. Rahman, "Special feature extraction techniques for Bangla speech," *2010 13th Int. Conf. Comput. Inf. Technol.*, no. Iccit, pp. 114–119, Dec. 2010.

[9] C. Goh and K. Leon, "Robust Computer Voice Recognition Using Improved MFCC Algorithm," *2009 Int. Conf. New Trends Inf. Serv. Sci.*, pp. 835–840, Jun. 2009.

[10] A. H. Fazli, "Number of Verses of the Qur ā n (Index and Argument)," *Int. J. Humanit. Soc. Sci.*, vol. 2, no. 19, 2012.

[11] M. A. Aabed, S. M. Awaideh, A.-R. M. Elshafei, and A. A. Gutub, "Arabic Diacritics based Steganography," in *2007 IEEE International Conference on Signal Processing and Communications*, 2007, pp. 756–759.

[12] M. B. Mustafa, S. S. Salim, N. Mohamed, B. Al-Qatab, and C. E. Siong, "Severity-based adaptation with limited data for ASR to aid dysarthric speakers.," *PLoS One*, vol. 9, no. 1, p. e86285, Jan. 2014.

[13] M. B. Janet, D. Li, J. Glass, S. Khudanpur, C. Lee, N. Morgan, and D. O. Shaughnessy, "Research Developments and Directions in Speech Recognition and Understanding, Part 1," 2009.

[14] M. Ismail, N. M. Diah, S. Ahmad, and A. A. Rahman, "Engaging Learners to Learn Tajweed through Active Participation in a Multimedia Application (TaLA)," pp. 88–91, 2011.

[15] S. Sholehuddin, "INDEPENDENT LEARNING OF QURAN (ILoQ) -ALPHABET USING SPEECH RECOGNITION," 2011.

[16] Z. Othman, "Malay speech to Jawi text engine," 2011.

[17] A. Ahmad, S. Ismail, and D. Samaon, "Recurrent neural network with backpropagation through time for speech recognition," in *IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004.*, 2004, vol. 1, pp. 98–102.

[18] M. Abd-AlRahman and M. Abushariah, "A Vector quantization approach to isolated-word automatic speech recognition," no. November, 2006.

[19] X. Yap, A. W. H. Khong, and W.-S. Gan, "Localization of acoustic source on solids: A linear predictive coding based algorithm for location template matching," *2010 IEEE Int. Conf. Acoust. Speech Signal Process.*, no. 1, pp. 2490–2493, 2010.

[20] V. Radha and C. Vimala, "A Review on Speech Recognition Challenges and Approaches," *doaj.org*, vol. 2, no. 1, pp. 1–7, 2012.

[21] G. Mosa and A. Ali, "Arabic phoneme recognition using hierarchical neural fuzzy petri net and LPC feature extraction," *Signal Process. An Int. J.*, no. 3, pp. 161–171, 2009.

[22] J.-P. Martens, *Continuous Speech Recognition over the Telephone*, no. May. 2000.

[23] A. Protopapas, "Performance of the LEMS HMM speech recognizer with PLP features and with multiple-window log-spaced DFT spectra," no. May, 1995.

[24] M. Jahanirad, A. W. A. Wahab, N. B. Anuar, M. Y. I. Idris, and M. N. Ayub, "Blind source mobile device identification based on recorded call," *Eng. Appl. Artif. Intell.*, vol. 36, pp. 320–331, Nov. 2014.

[25] H.-N. Ting, B.-F. Yong, and S. M. Mirhassani, "Self-Adjustable Neural Network for speech recognition," *Eng. Appl. Artif. Intell.*, vol. 26, no. 9, pp. 2022–2027, Oct. 2013.

[26] R. Rajavel and P. S. Sathidevi, "Adaptive Reliability Measure and Optimum Integration Weight for Decision Fusion Audio-visual Speech Recognition," *J. Signal Process. Syst.*, vol. 68, no. 1, pp. 83–93, Feb. 2011.

[27] N. Jamaliah Ibrahim, M. Yamani Idna Idris, Z. Razak, and N. Naemah Abdul Rahman, "Automated tajweed checking rules engine for Quranic learning," *Multicult. Educ. Technol. J.*, vol. 7, no. 4, pp. 275–287, Nov. 2013.

[28] A. Muhammad, "E-Hafiz: Intelligent System to Help Muslims in Recitation and Memorization of Quran," in *Life Science Journal*, 2012, vol. 9, no. 1, pp. 534–541.

[29] A. Wahidah and M. Suriazalmi, "Makhraj recognition using speech processing," *Comput. Converg. Technol. (ICCCT), 2012 7th Int. Conf.*, pp. 689–693, 2012.

[30] T. Hassan, A. Wassim, and M. Bassem, *Analysis and Implementation of an Automated Delimiter of" Quranic" Verses in Audio Files using Speech Recognition Techniques*, no. June. 2007.

[31] H. B. Kekre and T. K. Sarode, "Speech Data Compression using Vector Quantization," pp. 603–606, 2008.

[32] M. L. Albalate, "Data reduction techniques in classification processes," 2007.

[33] K. Sayood, *Introduction to data compression*, 3rd ed., vol. 114, no. 3. 2006.

[34] P. G. S. Rath, "REAL TIME SPEAKER RECOGNITION USING MFCC AND VQ National Institute of Technology," 2008.

[35] C.-C. Chang and W.-C. Wu, "Fast planar-oriented ripple search algorithm for hyperspace VQ codebook.," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1538–47, Jun. 2007.

[36] Y. Linde, a. Buzo, and R. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[37] N. J. Ibrahim, "Automated tajweed checking rules engine for quranic verse recitation," 2010.

[38] C. Tan and A. Jantan, "Digit recognition using neural networks," *Malaysian J. Comput. Sci.*, vol. 17, no. 2, pp. 40–54, 2004.

[39] G. Wang, Y. Guan, and Y. Zhang, "The MATLAB Simulation for More Generic Speech Recognizing Chinese Isolated Word System," *2009 Int. Conf. New Trends Inf. Serv. Sci.*, pp. 1145–1149, Jun. 2009.

[40] W. Ghai and N. Singh, "Literature review on automatic speech recognition," *Int. J. Comput. Appl.*, vol. 41, no. 8, pp. 42–50, 2012.

[41] B. Abro, A. B. Naqvi, and A. Hussain, "Qur'an recognition for the purpose of memorisation using Speech Recognition technique," in *2012 15th International Multitopic Conference (INMIC)*, 2012, pp. 30–34.