Jurnal
Teknologi

Full paper

Quranic Verses Verification using Speech Recognition Techniques

Ammar Mohammed[a*], Mohd Shahrizal Sunar[a], Md. Sah Hj Salam[b]

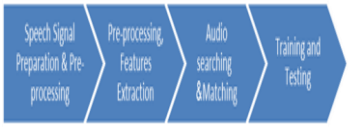
[a]*MaGIC-X UTM-IRDA, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*
[b]*Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*

*Corresponding author: ammar@magicx.my

Article history

Received :10 December 2014
Received in revised form :
1 February 2015
Accepted :12 February 2015

Graphical abstract


Speech Signal Preparation & Pre-processing
Pre-processing, Features Extraction
Audio searching &Matching
Training and Testing

Abstract

Al-Quran is the holy book of Muslims which is written and recited in Arabic language, the language in which it was revealed. Muslims believe that the Quran is neither corrupted nor altered this is mainly due to maintaining its original text. It is forbidden to recite the Quran in any other language apart from Arabic with neither additions nor subtractions. However with the proliferation of technology especially the Internet and social media sites such as Facebook and Twitter, the spread of mistakenly or deliberately distorted audio clips are witnessed regularly. In this regard, it is necessary to preserve the authenticity and integrity of the Quran from all sorts of corruption. This paper describes challenges and solutions for building a successful verification system of the Quran verses online. The paper describes the techniques used to deal with a finite vocabulary how modelling completely in the voice domain for language model and dictionary can avoid some system complexity, and how we built dictionaries, language and acoustic models in the framework.

Keywords: Speech recognition; quranic recitation recognition; speech to text; verses verification

■1.0 INTRODUCTION

The Mistakes in the recitation of holy Quran are forbidden. These errors may include; missing words, verse, misreading vowel pronunciations, punctuations, and accents [1].

The proliferation of online media especially the social media has played a big role in distributing Quran clips all over the world but many have been found carrying a lot of errors. These errors are sometimes intentional spread by people whose aim is to distort the Quran, and others are done unintentionally.

To solve this problem, there is a need for a technique which will be able to spot the errors online in the recitations carried by the clips. This will be of a great importance to the users who access these clips as well as maintaining the integrity and authenticity of the Quran. Presently, there is no automated application that can recognize Quranic recitation using speech recognition techniques.

Recitation of the Quran is not similar to a normal reading of Arabic text or book. There are rules that must be obeyed in recitation. These patterns are built on recitation artistically, they have been organized by Quran scholars and are called TAJWEED/TAJWID [1]. This illustrates the complexity of the problem.

This paper provides an overview of the techniques used in voice recognition in the Quran or Quranic recitation. It also provides a proposal for a system that will be capable of identifying errors in Quran recitation and be able to show where exactly errors have occurred [2]. Some of the techniques proposed in this study are; Mel Frequency Spectral Coefficients (MFFC) a technique that will extract important features in a speech signal for Quranic recitation and The Phonetic Search Engine (PSE) technique that will be used to search in the Quran database. The Dynamic Time Warping (DTW) is the technique used to recognize user recitation by comparing the recitation features and reference templates of speech features that are stored in the system.

■2.0 PROBLEM STATEMENT

The Holy Quran is a holy book for Muslims. It is for every Muslim to believe that there is no any kind of distortion in its text from the day it was revealed, because of maintaining its original text. However, with the spread multimedia in social networks individuals with different motives have used this advantage to spread the Quran. The motives are both negative and positive. The positive side is to teach users all over the world the real message the Quran carries, though there are people with the negative motivation like alteration and deformation of the Quran.

Apparently, there is no specific identification system of the Quranic based on its audio clips. The existing systems are the systems that are used in education in correcting Tajweed. This paper proposes a model which will identify errors in the Quranic audio files and subsequently distinguish incorrect recitation from the correct recitation.

■3.0 LITERATURE REVIEW

Holy Quran is written in classical Arabic language, and for this reason in this section will investigate the previous studies in speech recognition on the Holy Quran in addition to the Arabic language.

3.1 Background Arabic Language

This section focuses on a general description of the Arabic language and history and highlights some of the potential challenges for speech recognition.

3.1.1 Description of the Language

There are 22 Arabic countries with around 350 million Arabic speakers living in it or distributed all over the world. For this reason, Arabic language considered one of the most important and widely spoken languages in the world. Arabic is Semitic language that is characterized by the existence of particular consonants like pharyngeal, glottal and emphatic consonants [3]. As well as Arabic language has some phonetics characteristics that are built, around pattern roots (CVCVCV, CVCCVC, etc.) [4]. The 28 letters can be used in a set of 90 additional combinations, shapes, and vowels [4]. The 28 letters enclose consonants and long vowels such as ى and ٱ (both pronounced as/a:/), ي (pronounced as/i:/), and و (pronounced as/u:/). The short vowels and some other phonetic pronouncing like consonant doubling (shadda) are not introduced using letters directly, but by diacritics.

The diacritics are short strokes, where each can be located above or below the consonant. Complete set of Arabic diacritics. Arabic discretization is interpreted by three groups: short vowels, doubled case endings form, and syllabification marks.

3.1.2 Definition Of The Holy Quran

The Quran is the holy book of Islam, originally written in Classical Arabic language, consists of 6236 verses divided into 114 chapters called suras. Each surah also differs from one another in terms of the number of verses (ayat) [5, 6].

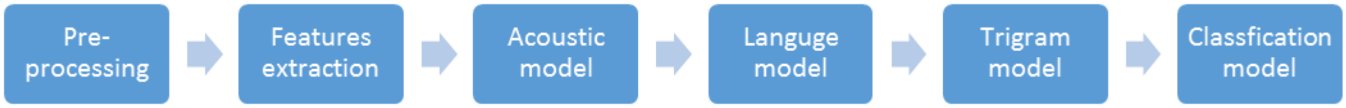


Figure 1 Speech recognition process

3.1.3 Challenges Facing The Recognition Of Arabic

Arabic language is not similar to US English, it is a semantic language with a composite morphology.it is one of the languages that are often described as morphologically complex and the problem of language modeling for Arabic are multipart by the variation of dialectal. there are many difficulties begin when dealing with the specialties of the Arabic language in Al-Quran, due to the differences between written and recite Al-Quran [1, 5, 7, 8].

Unlike most western languages, Arabic script writing orientation is from right to left. There are 28 characters in Arabic. The characters are connected and do not start with capital letter as in English. Most of the characters differ in shape based in their position in the sentence and adjunct letters, thus the Quranic Arabic alphabets consist of 28 letters, known as hijaiyah letters (from alif (ا)…until ya (ي)) [1, 9]. Those letters includes 25 letters, which represent consonants and 3 letters for vowels (/i: /, /a: /, /u :/) and the corresponding semivowels (/y/ and /w/), if applicable. A letter can have two to four different shapes: Isolated, beginning of a (sub) word, middle of a (sub) word and end of a (sub) word. Letters are mostly connected and there is no capitalization. The letter is represented as below at table, in their various forms.

3.2 Quranic Search Systems:

Most of researchers have interested on the development of search techniques for the Quranic text, but the principle can be applied in the development of voice search systems, for example,

Kadri *et al*. [10] proposed a new stemming method that tries to determine the core of a word according to linguistic rules. The new method shows the best retrieval effectiveness. The linguistic-based method can better determine the semantic core of a word.

Naglaa Thabet [11] proposed a new light stemming approach that gives better results, when applied to a rich vocalized text as the Quran. The stemmer is basically a light stemmer to remove prefixes and suffixes and is applied to a version of the Quran transliterated into western script.

Riyad Alshalabi [10] provided a technique for extracting the trilateral Arabic root for an unvocalized Arabic corpus. It provides an efficient way to remove suffixes and prefixes from the inflected words. Then it matches the resulting word with the available patterns to find the suitable one and then extracts the three letters of the root by removing all infixes in that pattern.

3.3 Process Speech Recognition System

Many speech recognition systems are used isolated word speech recognition systems require that speak most paused briefly between the words while continuous speech recognized system does not spontaneous speech may have disfluencies and is difficult to recognize.

Some speech recognizes systems require speech enrollment that is user has to provide sample of his/her speech before using the system while some systems are speaker independent.

3.4 Diagrammatic Representation of Speech Recognition

In this section the speech recognition process will represent as shown in Figure 1.

3.4.1 Pre-processing

In order to coordinate and simplify the data (input speech signals) and simplify feature extraction process, the data preparation consists of:

Silence removal: The holy Quran verse is uttered as an input audio, learners in most cases do not recite the holy Quran rhythmically and their utterance of words is slow. Therefore there are many chances of silences between uttered words. W.M. Muhammad *et al*. [3] proposed to calculate the energy of each frame and remove the frames which have energy extended to zero so that the system is bale to capture only words/content available in the speech. Thus, the overall time length of the signal is reduced, since blank spaces are removed.

Pre Emphasis: Once silence is removed, pre-emphasis is performed on signals, giving rise to higher frequencies with respect to magnitude of lower frequency, improving the signal to noise ratio. It is also known as a noise cancelling filter, because present echoes within the signal are also eliminated

3.4.2 Feature Extraction

The goal of feature extraction is to find a set of properties of an utterance that have acoustic correlations in the speech signal i.e. parameters that can somewhat be estimated through processing of the signal waveform. Such parameters are termed as features [4].

Several different feature extraction algorithms exist, namely Linear Predictive Cepstral Coefficients (LPCC): computes Spectral envelop before converting it into Cepstral coefficient.

Perceptual Linear Prediction (PLP) Cepstra: it is based on the Nonlinear Bark scale. The PLP is designed for speech recognition with removing of speaker dependent characteristics.

Mel-Frequency Cepstral Coefficients (MFCC) MFCC are extensively in ASR.MFCC is based on signal decomposition with the help of a filter bank, which uses the Mel scale. The MFCC results on Discrete Cosine Transform (DCT) of a real logarithm of a short-time energy expressed in the Mel frequency scale.

MFFC is the most widely used techniques in the Arab speech recognition because it is effective in noisy, vocal tract, and provide higher result of low bandwidth. The MFCC Processes are; frame blocking, windowing, DFT, Mel Scale Filter, Inverse Discrete Fourier Transform (IDFT) block. In the frame, blocking the speech waveform is cropped to remove silence or acoustical interference that may be present at the beginning or end of the sound file. As an outcome of this process Fourier transformation process is enabled.

Windowing: In order to minimize and eliminate discontinuity from the start and end of each frame of the signal, hamming window process is applied. It is the most commonly used in MFCC to minimize the discontinuities of the signal by tapering the beginning and end of each frame to zero.

Discrete Fourier Transform (DFT): Due to the speech signal form is a set of N discrete number of samples (windowed signal Y1 [k]… Y1 [m]), each frame sample is converted from time domain into the frequency domain (a complex number Y2 [k]). DFT is normally computed through Fast Fourier Transformation (FFT) algorithm.

Mel Filter-bank: Low frequency component in speech contains useful information as compared to high frequency. It represents the relationship between the frequency in Hz and Mel scale frequency. In order to perform Mel-scaling, a number of triangular filter-bank is used and therefore, a bank of triangular filters is created during MFCC calculation, collecting energy from each frequency band.

Inverse Discrete Fourier Transformation (IDFT): The final step of MFCC feature extraction is to take inverse of DFT. As an output of this step we get features of speech in vector format called feature vectors. The feature vector is obtained. This feature vector is used as input of the next phase. MFCC feature is considered for speaker–independent speech recognition and for the speaker recognition tasks as well.

Training and testing features training is a process of enrolling or registering a new speech sample of a distinct word to the identification system database, by constructing a model of the word based on the features extracted of word input speech. There are three methods for this purpose: HMM, VQ, and ANN.

The author recommends HMM the best approach for feature extraction and HMM or VQ is for training and testing. HMM is used when Arabic language recognition has to perform and VQ for English language [3] HMM had introduced the Viterbi algorithm for decoding HMMs, and the Baum-Welch or Forward-Backward algorithm for training HMMs. All the algorithm of HMM play a crucial role in ASR. It involved with states, transitions, and observations map into the speech recognition task.

The extensions to the Baum-Welch algorithms needed to deal with spoken language. These methods had been implemented by D. Jurafsky and J. H. Martin (2007) in their research. Here, speech recognition systems train each phone HMM embedded in an entire sentence. Hence, the segmentation and phone alignment are performed automatically as parts of the training procedure [5]. It consists of two interrelated stochastic processes common to describe the statistical characteristics of the signal. One of which is hidden (unobserved) finite-state Markov chain, and the other is the observation vector associated with each state of the Markov chain stochastic process (observable) [6].

Artificial Neural Network (ANN) is a computational model or mathematical model based on biological neural networks. The procedure depends on the way a person applies intelligence in visualizing, analyzing and characterized the speech based on a set of measured acoustic features [3], but the basic neural networks are not well equipped to address these problems as compared to HMM’s.

Vector Quantization (VQ) Quantization is the process of approximating continuous amplitude signals by discrete symbols. It can be quantized on a single signal value or parameter known as scalar quantization, vector quantization or others. VQ is divided into 2 parts, known as features training and matching features. Features training is mainly concerned with randomly selecting feature vectors and perform training for the codebook using vector quantization (VQ) algorithm.

3.4.3 Acoustic Model

Acoustic model represents the acoustic sounds of a language and can be trained to recognize the char of a particular user’s speech patter and acoustic environment.

Lexical model gives a list of large no. of words in a language along with how to pronounce each word.

The acoustic modelling will be done by HMMs, Nevertheless there are three methods for: Hidden Markov Models HMM, Vector Quantization (VQ), Artificial Neural Network (ANN) [6] and Hybird Model [12].

HMM or VQ can be apply for training and testing. HMM is used when Arabic language recognition has to perform and VQ for English language [13] HMM had introduced the Viterbi algorithm for decoding HMMs, and the Baum-Welch or Forward-Backward algorithm for training HMMs. All the algorithm of HMM play a crucial role in ASR. It involved with states, transitions, and observations map into the speech recognition task.

The extensions to the Baum-Welch algorithms needed to deal with spoken language. These methods had been implemented by D. Jurafsky and J. H. Martin (2007) in their research. Here, speech recognition systems train each phone HMM embedded in an entire sentence. Hence, the segmentation and phone alignment are performed automatically as parts of the training procedure [7]. It consists of two interrelated stochastic processes common to describe the statistical characteristics of the signal. One of which is hidden (unobserved) finite-state Markov chain, and the other is the observation vector associated with each state of the Markov chain stochastic process (observable) [14].

Artificial Neural Network (ANN) is a computational model or mathematical model based on biological neural networks. The procedure depends on the way a person applies intelligence in visualizing, analyzing and characterized the speech based on a set of measured acoustic features [7], but the basic neural networks are not well equipped to address these problems as compared to HMM’s.

Vector Quantization (VQ) Quantization is the process of approximating continuous amplitude signals by discrete symbols. It can be quantized on a single signal value or parameter known as scalar quantization, vector quantization or others. VQ is divided into 2 parts, known as features training and matching features. Features training is mainly concerned with randomly selecting feature vectors and perform training for the codebook using vector quantization (VQ) algorithm[7].

3.4.4 Language Model

Language model gives the way in which different words of a language are combined. In order to recognized a word the recognizer chooses it is guess from a finite vocabulary as the word is uniquely identified by it is spelling different models are used for this purpose.

As it is known that the Quran is written in the original classical Arabic. And Arabic is one of the languages that are often described as morphologically complex and the problem of language modeling for Quranic recitation are multipart by the methods and speed of recitation.It is also there are many difficulties begin when dealing with the specialties of the Arabic language in Al-Quran, due to the differences between written and recite Al-Quran [1, 15, 16]. The Quranic Arabic alphabets consist of 28 letters as shown in Table 1, known as hijaiyah letters (from alif (ا)…until ya (ي)) [5, 6, 16]. Those letters includes 25 letters, which represent consonants and 3 letters for vowels (/i: /, /a: /, /u :/) and the corresponding semivowels (/y/ and /w/), if applicable. A letter can have two to four different shapes: Isolated, beginning of a (sub) word, middle of a (sub) word and end of a (sub) word. Letters are mostly connected and there is no capitalization. The letter is represented as below at table, in their various forms.

Table 1 The various forms of Arabic letter

Character	Name	Isolated	Initial	Middle	Final
Alif	ألف	ا	ا	ـا	ـا
Ba'	باء	ب	بـ	ـبـ	ـب
Ta'	تاء	ت	تـ	ـتـ	ـت
Tha'	ثاء	ث	ثـ	ـتـ	ـث
Jeem	جيم	ج	جـ	ـجـ	ـج
H'a'	حاء	ح	حـ	ـحـ	ـح
Kha'	خاء	خ	خـ	ـخـ	ـخ
Dal	دال	د	د	ـد	ـد
Thal	ذال	ذ	ذ	ـذ	ـذ
Rai	راي	ر	ر	ـر	ـر
Zai	زاي	ز	ز	ـز	ـز
Seen	سين	س	سـ	ـسـ	ـس
Sheen	شين	ش	شـ	ـشـ	ـش
Sad	صاد	ص	صـ	ـصـ	ـص
Dhad	ضاد	ض	ضـ	ـضـ	ـض
Tta'	طاء	ط	طـ	ـطـ	ـط
Dha'	ظاء	ظ	ظـ	ـظـ	ـظ
A'in	عين	ع	عـ	ـعـ	ـع
Ghain	غين	غ	غـ	ـغـ	ـغ
Fa'	فاء	ف	فـ	ـفـ	ـف
Qaf	قاف	ق	قـ	ـقـ	ـق
Kaf	كاف	ك	كـ	ـكـ	ـك
Lam	لام	ل	لـ	ـلـ	ـل
Meem	ميم	م	مـ	ـمـ	ـم
Noon	نون	ن	نـ	ـنـ	ـن
Ha'	هاء	ه	هـ	ـهـ	ـه
Waw	واو	و	و	ـو	ـو
Ya'	ياء	ي	يـ	ـيـ	ـي

3.4.5 Features Classification and Pattern Recognition

The main objective of pattern recognition is to classify the object of interest into one of a number of categories or classes.
There are many methods used for pattern matching, classification as well as recognition. Under the same techniques of speech recognition, the normally methods used nowadays listed as below [5]:

1. Hidden Markov Model (HMM)
2. Vector Quantization (VQ)
3. Artificial Neural Network (ANN).

HMM is a pure expression of acoustic model of the voice because the simplicity and accessibility of training algorithms for estimating the parameters of the models from finite training sets of speech data; and the ease of implementation of the overall recognition system. [7] according the audio search, there are Several different methods have commonly been applied to the speech retrieval problem. One approach is to employ a Large Vocabulary Continuous Speech Recognizer (LVCSR, also known as “speech-to-text”). The speech is converted to text so that it can be searched very quickly for occurrences of a specified keyword or keywords.

In that case, these results are in a closed vocabulary and the other drawback here is that it is a tough decision on the word’s existence which must be achieved during the realization phase. In another retrieval technique, called word spotting, the search is performed on the speech after a keyword is presented. This creates an open vocabulary. The disadvantage to this process is the difficulty of searching much faster than real time.

3.5 Speech Recognition Technique for Quranic Recitation

In general there are several research in the field of Quranic voice recognition, of the most famous of such research is an automated delimiter introduced by Hassan Tabbal *et al.* [1] which extracts verses from an audio file and coverts Quran verses in audio file, using speech recognition technique. The Sphinx IV framework is used to develop this system.

The core recognition process is provided automatically by the sphinx engine using the appropriate language and acoustic models. The sphinx framework must be configured using an xml based configuration file.

The recognition ratio in the case of Tarteel is slightly better than in the case of TAJWEED. One possible reason for this could be that the majority of the Tarteel recitations available now follow the same monotony and the duration (in time) of each phoneme differs slightly from one reciter to another [1]. There is also the extra noise that is caused by the compression of the audio files and the low quality of the recordings.

Although we had anticipated this by using noisy audio files during the training, but the differences in compression ratios between the files add a lot of variation of the added noise and thus causing extra errors. When unskilled persons tested the system (we even tested it on children), it behaved astonishingly well even when the reciter was a woman, a case that cannot be encountered in real life because it's not usual to have a woman reciting the Holy Quran. There is also an interesting observation drawn from these tests: It is always recommended [6] to train the system with more than 500 different voices in order to reach speaker independence. But we didn't train our system with this relatively large number and still we were able to have remarkable speaker independence results.

The system introduced by Hassan Tabba [1], an automated delimiter, which extracts verses from the audio files, using MFCC feature extraction. This system is useful for people who are well versed with Tajweed rules. However, users who are not Arabic speakers don't benefit from this system. In addition, it may also not help reciters to improve recitation abilities. The system is useful to those people who already know the correct recitation of the holy Quran and the subsequent rules and not suitable for none Arabic speakers. A system that will be able to help users to know recitation rules (TAJWEED), pointing out mistakes made during recitation is a necessity and a task achieved by the E - Hafiz system.

The other system introduced by Bushra Abro [2]. Arabic language which included isolated words and sentences. The dataset comprised of few Arabic sentences and words. The system was built on Al-Alaoui algorithm that trains Neural Networks (NN) and has been able to achieve 88% accuracy on sentence recognition. But the drawback of this technique is that the system was trained on distinct sentences.

Similarities in sentences separate NNs and therefore are needed to be trained which is computationally very expensive, for the purpose of Qur'an memorization, MFCC was used to extract features. The dataset consists of few small Quranic verses and pattern matching was done by Vector Quantization. The system gained good recognition rate but the approach is statistical and therefore difficult to scale up for larger system to be built on complete Qur'an. It is also known to take much time and space complexity which is a shortfall of a real time system.

Zaidi Razzak [7] presents different recognition techniques used for the recitation of the Quran verses in Arabic verse pointing out the advantages and the drawbacks. The most useful method for the project "Quranic verse Recitation Recognition for support j-QAF learning" is explained therein. J-QAF is a pilot program, which aims to encourage learners to learn Quran reading skills, understanding Tajweed and Islamic obligations. The method of teaching j-QAF (teacher and student) is still handled manually. One basic goal of this paper is to automate the learning process.

■4.0 METHODOLOGY

There are four major stages in the development of the Verification of "Quranic" Verses in Audio Files System. The phases are sequential as described in Figure 2. The stages are as follows; Speech Signal Preparation & Pre-processing, Features Extraction, Audio searching &Matching, Training and Testing

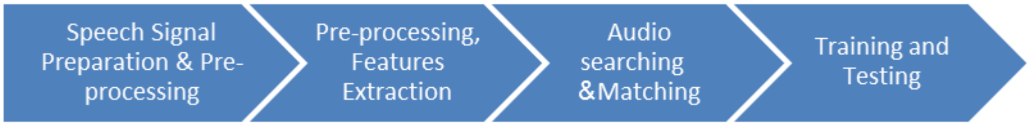

Speech Signal Preparation & Pre-processing
Pre-processing, Features Extraction
Audio searching &Matching
Training and Testing

Figure 2 Phases of Quranic verification system

4.1 Speech Signal Preparation & Pre-processing:

In this part, recording process will be executed, due to collect the Quranic recitation of speech samples from different Qari (Recitor). According to Rabiner and Juang (1993), there are 4 main factors need to be considered while collecting the speech sample, such as: Who are the talkers, The speaking condition, The transducers & transmission systems and The speech unit.

These 4 factors need to be identified first, before any process of recording executed.

It is because; these factors will affect the performance and the output result, especially the training set vectors that will be used in training and testing process.

This research will used a simple MATLAB function for recording the speech samples (Quran verses). And the data set will be collected from ten of expert Qari (They have Ejazah in Hafss), each of them will be recite suratu Al-Nnass ten times correct and ten times with mistakes, this mistake should be different types as mistakes in Makhraj, mistakes in Tajweed, mistakes by words missing), this data set will be used in training and testing phase.

After data recording, the process of speech signal processing consists of Sampling, Remove the noise and Segmentation. Sampling needed because that Human voices will generate continuous analog signals. Therefore, the analog signal is chopped in certain interval of time. Discrete series sample x

[n] is obtained from continuous signal x(t), x[n] = x(nT), Where T is sampling period and i/T = Fs is sampling frequency in unit of sample/second. The value of n is the number of samples. According to the Nyquist sampling theory, minimal sampling frequency required is twice of original maximal signal.

4.2 Feature Extraction

The main objective of feature extraction is to extract the important characteristics from the speech signal, that are unique for each word, due to differentiate between a wide set of distinct words. According to Ursin [17], MFCC is considered as the standard method for feature extraction in speech recognition and perhaps, the most popular feature extraction technique used nowadays, This is because MFCC able to obtain a better accuracy with a minor computational complexity, respect to alternative processing as compared to other feature extraction techniques.

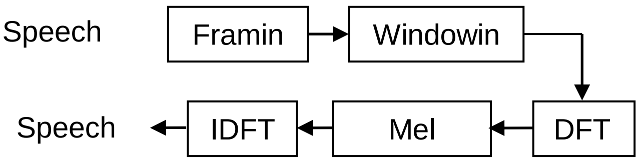


Figure 3 Block diagram of the MFCC processes

The Mel-Frequency Cepstral Coefficients (MFCC) is frequently used for feature extraction technique in speech processing. In this technique, the used of Mel scale in the derivation of cepstrum coefficients was introduced. The Mel scale is a mapping of the linear frequency scale based on human auditory perception[18].

The MFCC Processes are described in Figure 3; frame blocking, windowing, DFT, Mel Scale Filter, Inverse Discrete Fourier Transform (IDFT) block. In the frame, blocking the speech waveform is cropped to remove silence or acoustical interference that may be present at the beginning or end of the sound file. As an outcome of this process Fourier transformation process is enabled.

Windowing: In order to minimize and eliminate discontinuity from the start and end of each frame of the signal, hamming window process is applied. It is the most commonly used in MFCC to minimize the discontinuities of the signal by tapering the beginning and end of each frame to zero. Windowing which is used in this research is Hamming windowing.

Discrete Fourier Transform (DFT): Due to the speech signal form is a set of N discrete number of samples (windowed signal Y1 [k]… Y1 [m]), each frame sample is converted from time domain into the frequency domain (a complex number Y2 [k]). DFT is normally computed through Fast Fourier Transformation (FFT) algorithm. The FFT is defined on the set of N samples {Xn}

Mel Filter-bank: Low frequency component in speech contains useful information as compared to high frequency. It represents the relationship between the frequency in Hz and Mel scale frequency. In order to perform Mel-scaling, a number of triangular filter-bank is used and therefore, a bank of triangular filters is created during MFCC calculation, collecting energy from each frequency band.

Inverse Discrete Fourier Transformation (IDFT): The final step of MFCC feature extraction is to take inverse of DFT. As an output of this step we get features of speech in vector format called feature vectors. The feature vector is obtained. This feature vector is used as input of the next phase. MFCC feature is considered for speaker–independent speech recognition and for the speaker recognition tasks as well.

In this research the incoming speech signal is sampled with sampling frequency of 8000 Hz, accordance with Nyquist rule, and then divided into time slots (framing) with frame time 40 ms and overlapping time 20 ms or about 50%. The number of frame is separated depend on the recitation word number. Then, each frame is passed through Hamming window to reduce signals discontinuity after chopping. The signal is then transformed to frequency domain by using DFT with N = 1024. The signal is passed through the Filter bank of 24 triangle filter. The DCT with MFCC coefficient of 14 is done after filtering process. Other feature calculation like signal energy, delta-MFCC and delta-delta MFCC will be done after DCT process.

4.3 Training and Testing Phase (HMMs):

Hidden Markov Model (HMM) is a statistical model of system that is used in pattern recognition field, especially in speech recognition. It is widely used for characterizing the spectral properties of the frames for a certain pattern. Using the HMM, the input of speech signal is well characterized as a parametric random process and the parameters of stochastic process can be determined in precise and well-defined manner.

The parameter of HMM model need to update regularly, due to make the system able to fit a sequence for particular application. Thus, the training of the HMM model is so important, due to represent the utterances of words. This model is used later on in the testing of utterances and calculating the probability of HMM model, in order to create the sequence of vectors.

In HMM statistical approach, the Quranic recitation of input speech is represented accordingly with some probability distributions. According to Markov models, if the observation is a probabilistic function of state, it is called as Hidden Markov Model. It is because, it consist of doubly embedded stochastic process with underlying, that is not directly observable (hidden), but can be observed through another set of stochastic process only, that may produce the sequence of observations [19].

This research will be used the HMMs with 3 state in tow main stages in a speech recognition system, which are training and recognition stages. Under the training stage, models (patterns) are generated from the input of speech samples, after the feature extraction process and modeling techniques. Meanwhile, in the recognition stage, features vector will be generated from the input speech samples with the same extraction procedures in the training stage, mentioned earlier. After that classification process, as well as the decision process was made and executed with some matching techniques. Under the classification type, the recognition task can be divided either identification or verification process.

The continuous speech recognition mainly for Tajweed roles that consist of the following 3 major steps, which are:

(1) Training/Modeling: Each word in the vocabulary build an HMM model and estimate the model parameters of □ □ (A, pi0,mu, sigma) , which represent the likelihood of the training set observation vectors.

(2) Identification: Each unknown words to be recognized and measurement of the observation sequence through the feature analysis of the speech, corresponding to the word were made. Lastly, the word is selected using the Viterbi algorithm, which the model likelihood is maximum

(3) Verification: The input features were compared with the registered pattern, and any features that giving the highest score is identified as the selected/target speaker (recitor) and recitation results. Then, these input features are compared with the claimed speaker (recitor) and decision is made either to accept or reject the claimed/results.

According to these 3 major steps listed above, the training/modeling step was executed during HMM training, while the identification and verification steps were carried out during HMM testing/matching.

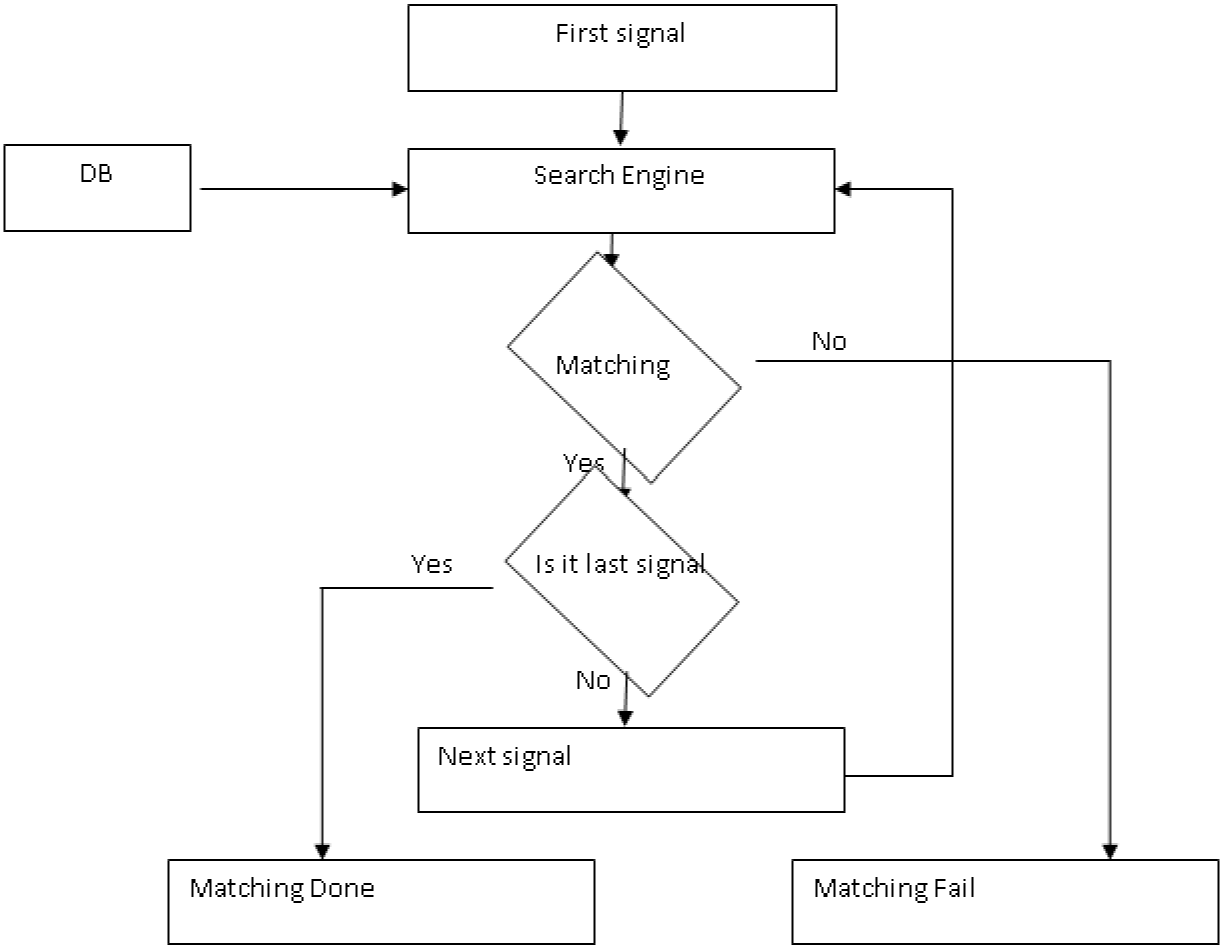


Figure 4 Block diagram of the search & matching steps

4.4 Modeling and Storing

This stage uses the output of the second stage to form a model as a feature vector of recitation [8], and stored in the database, which also contains a large number of vectors obtained from All Quranic verses. Essentially, it is a vector of MFCC features. When the user enters any verse, it is compared with the verses that are stored in the several Hafizes of Quran. The verses that do not match registers as an error, and pointed to the user.

4.5 Search and Matching:

At this point the audio signals are captured and searched in the database that contains all the Quranic Verses that are stored as MFCC factors Form in the third phase [9]. This phase is further elaborated in Figure 4.

■5.0 EXPECTED RESULT

This pedagogical model will provide the following:

1. Develop a new model of Quran verification, verification of Verses in Audio Files using Speech Recognition Techniques.
2. Help non-Arabic speaking Muslims to Verify Quran verses in Audio Files.
3. Help to develop Audio Search Engine in Quran Recitation

■6.0 CONCLUSION

Mistakes in the Quran are forbidden, and all Muslims are supposed to work towards maintaining the authenticity and integrity of the holy Quran against distortions and corruption. The systems in existence towards this endeavour have been discussed in this paper and present the model of Verification of Quranic recitation of sound files and media. By the end of this study it is expected that the existing models will be enhanced and improved for more accuracy in recognition of Quran verses especially online. This model will relay on proven technologies in Arab speech recognition, such as MFCC for the Feature

Extraction phase and HMMs in recognition and matching phase in addition a model to search for the Quranic verses as a signal will also be utilized.

Acknowledgement

This paper is supported by MaGICX UTM-IRDA Universiti Teknologi Malaysia.

References

[1] Hassan, T., Wassim, A.-F., and Bassem, M. 2007. Analysis and Implementation of an Automated Delimiter of "Quranic" Verses in Audio Files using Speech Recognition Techniques. *Robust Speech Recognition and Understanding*. 351.

[2] Abro, B., Naqvi, A. B., and Hussain, A. 'Qur'an Recognition For The Purpose Of Memorisation Using Speech Recognition Technique'. In Editor (Ed.)^(Eds.): 'Book Qur'an Recognition for the Purpose of Memorisation Using Speech Recognition Technique' (IEEE, 2012, edn.). 30–34

[3] Khalaf, E. F., Daqrouq, K., and Morfeq. 2014. A. 'Arabic Vowels Recognition by Modular Arithmetic and Wavelets using Neural Network. *Life Science Journal*. 11(3).

[4] Azmi, A. M., and Almajed, R. S. 2013. A Survey of Automatic Arabic Diacritization Techniques. *Natural Language Engineering*. 1–19.

[5] Ibrahim, N. J. 2010. Automated TAJWEED Checking Rules Engine for Quranic Verse Recitation.

[6] Ibrahim, N. J., Idris, M. Y. I., Razak, Z., and Rahman, N. N. A. 2013. Automated Tajweed Checking Rules Engine for Quranic Learning. *Multicultural Education & Technology Journa*l. 7(4): 275–287.

[7] Zaidi Razak†, N. J. I., Mohd Yamani Idna Idris, Emran Mohd Tamil, Mohd Yakub @ Zulkifli Mohd Yusoff and Noor Naemah Abdul Rahman. 2008. Quranic Verse Recitation Recognition Module for Support in j-QAF Learning: A Review. *IJCSNS International Journal of Computer Science and Network Security*. 8(8): August 2008.

[8] Muhammad, A., ul Qayyum, Z., Tanveer, W. M. M. S., and Syed, A. Z. E-Hafiz. 2012. Intelligent System to Help Muslims in Recitation and Memorization of Quran. *Life Science Journal*. 9(1): 534–541.

[9] Djemili, R., Bedda, M., and Bourouba, H. 2004. Recognition of Spoken Arabic Digits Using Neural Predictive Hidden Markov Models. *Int. Arab J. Inf. Technol*. 1(2): 226–233.

[10] Al-Taani, A. T., and Al-Gharaibeh, A. M. 2010. *Searching Concepts and Keywords in the Holy Quran*.

[11] Thabet, N. 2004. Stemming the Qur'an. In Editor (Ed.)^(Eds.): 'Book Stemming the Qur'an' (Association for Computational Linguistics, edn.). 85–88.

[12] Zarrouk, E., Ayed, Y. B., and Gargouri, F. 2014. Hybrid Continuous Speech Recognition Systems By HMM, MLP and SVM: A Comparative Study. *International Journal of Speech Technology*. 1–11.

[13] Muhammad, W. M., Muhammad, R., Muhammad, A., and Martinez-Enriquez, A. 2010. Voice Content Matching System for Quran Readers. In Editor (Ed.)^(Eds.): Book Voice Content Matching System for Quran Readers. 148–153.

[14] Meng, J., Zhang, J., and Zhao, H. 2012. Overview of the Speech Recognition Technology. In Editor (Ed.)^(Eds.): Book Overview of the Speech Recognition Technology. 199–202.

[15] Truong, K. 2004 'Automatic pronunciation error detection in Dutch as a second language: an acoustic-phonetic approach',

[16] Ahsiah, I., Noor, N., and Idris, M. 2013. Tajweed Checking System to Support Recitation. In Editor (Ed.)^(Eds.). Book Tajweed Checking System to Support Recitation. 189–193.

[17] Ursin, M. 2002 Triphone Clustering in Finnish Continuous Speech Recognition. Diplomityö, Teknillinen korkeakoulu.

[18] Arslan, L. M. 1996. Foreign Accent Classification in American English.

[19] Juang, B. H., and Rabiner, L. R. 1991. Hidden Markov Models for Speech Recognition. *Technometrics*. 33(3): 251–272.