Quranic Verse Recitation Recognition Module for Support in j-QAF Learning: A Review

Zaidi Razak[†], Noor Jamaliah Ibrahim[†], Mohd Yamani Idna Idris[†], Emran Mohd Tamil[†], Mohd Yakub @ Zulkifli Mohd Yusoff[††] and Noor Naemah Abdul Rahman[††]

[†]Faculty of Science Computer and Information Technology, University of Malaya, Kuala Lumpur, Malaysia.
[††]Academy of Islamic Studies, University of Malaya, Kuala Lumpur, Malaysia.

Summary
Each person's voice is different. Thus, the Quran sound, which had been recited by most of recitors will probably tend to differ a lot from one person to another. Although those Quranic sentence were particularly taken from the same verse, but the way of the sentence in Al-Quran been recited or delivered may be different. It may produce the difference sounds for the different recitors. Those same combinations of letters may be pronounced differently due to the use of harakates. This paper seeks to provide a comprehensive review of Quran Arabic verse recitation recognition focusing on the techniques used, the advantages, and drawbacks. Areas with potential of further expansion are identified for future research for support in j-QAF learning.

Key words:
Quranic verse recitation recognition, j-QAF and speech recognition.

1. Introduction

Malaysia government had allocated RM150 million under the Ninth Malaysia Plan to build j-QAF (Jawi, Quran, Arab Language and Fardu Ain) laboratories in primary schools nationwide. In order to inculcate moral and ethical values as well as nurture responsible citizens, various programmes were introduced, and the j-QAF is one of them. The j-QAF programme was introduced in 2005 to Muslim students at the primary level, starting with Year 1 [2]. j-QAF is a pilot programme, which mainly focus to promote a better skills and understanding of Jawi, Quran reading, Arabic and Islam's obligatory duties (fardhu ain) [6]. Observations of the j-QAF programme always been done by JAPIM, due to increase the effectiveness of j-QAF program. This j-QAF program consists of 5 teaching models and will gradually implement to each level until 2010. Those 5 teaching models are JAWI recovery class model, 'Tasmik' model, 'Khatam Al-Quran' model, Arabic language communication (BAK) model and another curriculum model which is 'Bestari Solat' model [3].

Focus on the Quran learning process under j-QAF programme; it is believed that this programme required the special and effective way to recite Quran [4]. Furthermore, j-QAF learning process for Quran is still handled with manual method between j-QAF teachers and students [3]. It is based on 'Khatam Al-Quran' model and 'Tasmik' model [6], which focus on reading the Al-Quran skills through *talaqqi* and *musyafahah* methods [3]. These methods are described as face to face learning process between students and teachers, where listening, correction of Al-Quran recitation and repetition of the correct Al-Quran recitation took place [3]. This factor is important, so that students will know how the *hijaiyah* letters are pronouncing correctly. The process only can be done, if the teachers and students follow the art, rules and regulations while reading the Al-Quran, known as "Rules of Tajweed" [4].

2. Holy Quran Recitations and Acoustic Model

2.1 The "Art of Tajweed"

"Tajweed" is an Arabic word meaning proper pronunciation during recitation, as well as recitation at a moderate speed. It is a set of rules which govern how the Quran should be read [5]. It is considered as an art because not all recitors will perform the same Quran verse in the same way [4]. The 'art of tajweed' defines some of flexible well-defines rules to recite Quran. Those rules create a big difference between normal Arabic speech and recited Quranic verses, which may produce the interesting result, based on the impact of "art" analysis for automatic recognition process especially on the acoustic model. Furthermore, the "art of tajweed" is the manual methods that need a lot of work and proved to be unable to adapt to new recitors. But, it is still believe that, the special way to recite Quran is by looking forward the art of tajweed [5].

Manuscript received August 5, 2008.
Manuscript revised August 20, 2008.

2.2 Effect of the "Art of Tajweed" on the acoustic Model

As we already know, each person's voice is different. Therefore, the Quran sound, which had been recited by most of recitors were totally different from one person to another. Although those Quranic sentence were particularly taken from the same verse, but the way of the sentence in Al-Quran been recited or delivered may be different [4]. It may produce the difference sounds for the different recitors. Moreover, there are many difficulties arise when dealing with the specialties of the Arabic language in Holy Quran, due to the differences between written and recite the Al-Quran. Those same combinations of letters may be pronounced differently due to the use of harakattes [4]. The most important tajweed rules that believed can influence the recitation recognition aspect were stated as below:

i) Necessary prolongation of 6 vowels.
ii) Obligatory prolongation of 4 or 5 vowels.
iii) Permissible prolongation of 2, 4 or 6 vowels.
iv) Normal prolongation of 2 vowels.
v) Nasalization (ghunnah) of 2 vowels.
vi) Silent unannounced letters.
vii) Emphatic pronunciation of the letter R.

The above laws are based from the specific recitation rules. Moreover, the predefined "maqams" also been used by recitors to vary the tone of their recitations [4]. There is 10 different laws set according to the 10 certified scholars [Hafs, Kaloun, Warsh…] who teached the recitation of the Holy Quran [8] [1]. In order to deal with these laws, the prolongations as the repetition of the vowel n-corresponding times need to be considered, same as well as the nasalization [4]. This rule governs the consonants/vowel combinations, usage of short and long vowels, the co-articulation effect of emphatics and pharyngeals, pronounciation, *Tanween* and *Ghonna* rules and rules for combining words [9]. Note that, if there any echoing sound produced during the Quranic recitation recording process, the echo will be considered as noise. That noise can be eliminated using the noise-canceling filter [4].

2.3 Linguistic properties of Arabic

Arabic is an official language in more than 22 countries. Since it is also the language of religious instruction in Islam, many more speakers have at least a passive knowledge of the language. Arabic 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 [12] [13] [14]. But, only Modern Standard Arabic (MSA) is used for written and formal communication. It is because, only MSA has a universally agreed upon the writing standard as well as for communication purposes [12] [13] [14] [23].

As mentioned earlier in part 2.2, there are many difficulties begin when dealing with the specialties of the Arabic language in Holy Quran, due to the differences between written and recite the Al-Quran [4] [13] [14]. The Quranic Arabic alphabets consist of 28 letters, known as hijaiyah letters (from alif until ya) [7] [10] [12] [14]. Those letters includes 25 letters, which represent consonant and 3 letters for vowels (/i: /, /a: /, /u :/) [12] [14] and the corresponding semivowels (/y/ and /w/), if applicable [23]. 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 [23]. Letters are mostly connected and there is no capitalization. The letters are represented as below at Table 1, in their various forms.

Table 1: Letters of Arabic alphabets [1]

Isolated	Beginning	Middle	End	Name	Phoneme
ا	ا	ا	ا	ʾalif	/a:/
ب	بـ	ـبـ	ـب	baaʾ	/b/
ت	تـ	ـتـ	ـت	taaʾ	/t/
ث	ثـ	ـثـ	ـث	thaaʾ	/θ/
ج	جـ	ـجـ	ـج	gym	/ʤ/
ح	حـ	ـحـ	ـح	Haaʾ	/ħ/
خ	خـ	ـخـ	ـخ	khaaʾ	/x/
د	د	ـدـ	ـد	daal	/d/
ذ	ذ	ـذـ	ـذ	dhaal	/ð/
ز	ز	ـزـ	ـز	zayn	/z/
ر	ر	ـرـ	ـر	raa	/r/
س	سـ	ـسـ	ـس	syn	/s/
ش	شـ	ـشـ	ـش	shyn	/ʃ/
ص	صـ	ـصـ	ـص	Saad	/ʂ/
ض	ضـ	ـضـ	ـض	Daad	/ɖ/
ط	طـ	ـطـ	ـط	Taaʾ	/ʈ/
ظ	ظـ	ـظـ	ـظ	Zaaʾ	/ʐ/
ع	عـ	ـعـ	ـع	ʾayn	/ʕ/
غ	غـ	ـغـ	ـغ	ghayn	/ɤ/
ك	كـ	ـكـ	ـك	kaaf	/k/
ق	قـ	ـقـ	ـق	qaaf	/q/
ف	فـ	ـفـ	ـف	faaʾ	/f/
ل	لـ	ـلـ	ـل	laam	/l/
ن	نـ	ـنـ	ـن	nuwn	/n/
م	مـ	ـمـ	ـم	mym	/m/
ه	هـ	ـهـ	ـه	haaʾ	/h/
و	و	ـو	ـو	waaw	/u:/
ي	يـ	ـيـ	ـي	yaaʾ	/i:/
ء	أ	ئ	ء	hamza	/ʔ/

Furthermore, other phonemes of pronunciation are marked by diacritics, such as consonant doubling (phonemic in Arabic). It is indicated by the "shadda" sign and the "tanween", word final adverbial markers which add /n/ to the pronunciation [13] [23]. Those sign can reflect the differences of pronunciation [13]. Moreover, the diacritic is really important in setting up the grammatical functions, which leading to the acceptable text understanding and correct reading or analysis [13]. The entire set of diacritics is listed in Table 2 below [12] [23].

Table 2: Arabic diacritics (from D. Vergyri and K. Kirchhoff [12])

Example	Symbol Name	Meaning
أَ	fatHa	/a/
إِ	kasra	/i/
أُ	Damma	/u/
رّ	shadda	consonant doubling
درْس	sukuwn	absence of vowel after consonant
أً	tanwyn al-fatHa	/an/
إٍ	tanwyn al-kasr	/in/
أٌ	tanwyn aD-Damm	/un/
ى	'alif maqsuwra	/a:/ sound, historical
هٰذه	dagger 'alif	/a:/ sound, historical
آ	madda	double alif
في ٱلبيت	waSla	on 'alif in *al*
لا	laam 'alif	combination of laam and 'alif
ة	taa marbuwta	morphophonemic marker

Letter to sound conversion for Arabic usually has simple one to one mapping between orthography and phonetic transcription for given correct diacritics. 14 vowels had been used to accommodate for short and long vowels, same as well as the emphatic vowels. Each syllable begins with a consonant followed by a vowel, which are limited and easily detectable. Short vowels are denoted by "V" and long vowels are denoted as "V:" [9] [11] [19]. Those syllable can be classified according to the length of the syllable, which also known as *harakattes* [4].

CV	Short	;	open
CV:	Long	;	open
CVC	Long	;	closed
CV: C	Long	;	closed
CVCC	Long	;	closed
CV: C	Long	;	closed

3. Quranic Verse Recitation Recognition Systems

H. Tabbal et al. (2006) [4] go through the Quranic verse recitation recognition, which covered the Quran verse delimitation system in audio files using speech recognition techniques. Here, the Holy Quran recitation and pronunciation as well as software used for recognition purposes had been discussed. The Automatic Speech Recognizer (ASR) has been developed by using the open source Sphinx framework as the basis of this research. The scope of this project more focus into the automated delimiter, which can extract the verse from the audio files. Research techniques for each phase were discussed and evaluated using implementation of various techniques for different recitors, which recite sourat "Al-Ikhlas". Here, the most important tajweed rules and tarteel, which can influence the recognition of a specific recitation, can be specified.

A comprehensive evaluation of Quran recitation recognition techniques was provided by A.M. Ahmad et al. in 2004 [7]. The survey provides recognition rates and descriptions of test data for the approaches considered. Focusing on Quran Arabic recitation recognition, it incorporates background on the area, discussion of the techniques, and potential research directions.

In general there are four major stages in speech recognition system. Under the same techniques of speech recognition, the Quranic Arabic recitation recognition also can be implemented based on these techniques specified:

1. Pre-processing,
2. Feature Extraction,
3. Training and Testing
4. Features Classification and Pattern Recognition

It can be described based on the system architecture shown below:

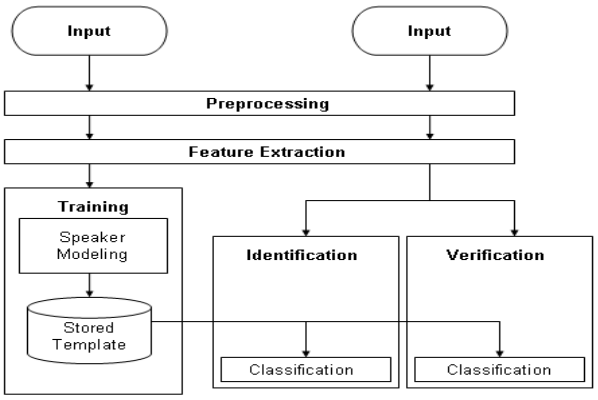

Fig. 1: System architecture

3.1 Pre-processing

In order to improve the readability and the automatic recognition of speech processing, pre-processing steps are essential. The main benefit in pre-processing of speech recognition is to organize the information and simplify the following task of recognition. Pre-processing steps mainly consist of the following:

1. Endpoint detection [7].
2. Pre-emphasis filtering/Noise filtering/Smoothing [7].
3. Channel Normalization [40].

3.1.1 Endpoint Detection

Short-time energy or spectral energy is usually used as the primary feature parameter with the augmentation of zero-crossing rate, pitch and duration information in endpoint detection algorithms. But recently, the endpoint detection features become less reliable in the presence of non-stationary noise and various type of sound artifact [30]. It is because; endpoint detection and verification of speech segments become relatively difficult in noisy environment.

3.1.2 Pre-emphasis /Noise filtering/Smoothing

The purpose of the smoothing stage is to decrease the noise and regularize the word contours. A. M. Ahmad et al. (2004) [7] are also digitized the Arabic's alphabet from speaker, as well as digital filtering. The digital filtering may emphasize the important frequency component in signal. Then the start-end point can be analyzed based on the signal of the phonemes. GoldWave was responsible to filter from analog signal to digital signals, due to analyze the start-end points that contain information of speeches. According to the H. Tabbal et al. (2006) [4], the use of 2-stage pre-emphasis filter with the different factor value (0.92 and 0.97) could increase the recognition ratio of some audio files. It is due to the speech frame of 10ms and a threshold of 10dB for the speech extractor chosen. It also can consider as the noise canceling filter, due to eliminate echo (noise).

Besides the pre-emphasis filtering, there are another techniques used by K.Kirchhoff et al. (2004) [14], which is Kneser-Ney smoothing. Kneser-Ney smoothing able to built trigram models for each of the stream with different morphology. It believes can outperform other smoothing method consistently include in noisy environment [35].

3.1.3 Channel Normalization

Another approaches used for pre-processing method was also known as channel normalization. According to J. de Veth and L. Boves (1998) [40], channel normalization (CN) techniques have been developed with the different applications domains, where a particular recognizer is trained with speech recorded using the microphone. The recognition is attempted based on speech recorded with the different microphone. Here, the contribution of the channel normalization during the training is still unknown in details but, it still constant although during the test time.

3.2 Feature Extraction

Feature extraction is the process of extracting measurements from the input to differentiate among classes. The main objective of feature extraction is to extract characteristics from the speech signal that are unique, discriminative, robust and computationally efficient to each word, which then used to differentiate between different words [26].

According to J. P. Martens (2000) [27], there are various speech features extraction techniques, stated as below:

1. Linear Predictive Coding (LPC) [7] [9]
2. Perceptual Linear Prediction (PLP) [38] [39]
3. Mel-Frequency Cepstral Coefficient (MFCC) [7] [4] [47]
4. Spectrographic Analysis [5]

3.2.1 Linear Predictive Coding (LPC)

A. M. Ahmad et al. (2004) [7] used this type of extraction technique to extract the LPCC coefficients from the speech token. The coefficients are then converted to cepstral coefficient that served as the input to the neural networks. The drawback of LPCC may estimate the high sensitivity to quantization noise. By converting the LPCC coefficients back into cepstral coefficient, it can decrease the sensitivity of high and low order cepstral coefficient to noise.

According to the M. E. Ahmed (1991) [9], LPC model had been replaced with a formant that is has much wider frequency spectrum. It is believe that, the LPC synthetic model give a bad outcome for the research, due to deduce the prosodic rules. This rule is very important rules of missing blocks, in order to construct an allophone based Arabic text-to-speech by rules.

3.2.2 Perceptual Linear Prediction (PLP)

Another popular feature set is the set of perceptual linear prediction (PLP) coefficients, which had been used by S.V.Vuuren (1996) [38] in his research. In the research, S.V.Vuuren compared the discriminability and robustness to noise of Perceptual Prediction (PLP) and Linear Prediction (LPC). Particularly for PLP, the spectral scale is the non-linear Bark scale and the spectral features are smoothed within the frequency bands.

PLP is first introduced by H.Hermansky (1990) [39], who formulated PLP feature extraction as a method for deriving a more auditory-like spectrum based on linear predictive analysis that achieved by making some engineering approximations of the psychophysical attributes of the human hearing process.

3.2.3 Mel-Frequency Cepstral Coefficient (MFCC)

The most prevalent and dominant method used to extract spectral features is calculating Mel-Frequency Cepstral Coefficients (MFCC). MFCC is one of the most popular feature extraction techniques used in speech recognition, whereby it is based on the frequency domain of Mel scale for human ear scale [7] [28]. Based on research of A. M. Ahmad et al. (2004) [7], Mel-scale has been used to perform filter bank processing to the power spectrum. It had been performed after windowing process and FFT had been implemented.

Similar approaches also had been implemented by H. Tabbal et al. (2006) [4]. The use of the MFCC has proven the remarkable result in the field of speech recognition. It is because, the behavior of the auditory system had been tried to emulate by transforming the frequency from a linear scale to a non-linear one.

According to the A.Youssef & O.Emam (2004) [11], 12 dimensional mel frequency cepstral coefficients (MFCCs) is been coded for recorded speech data. Pitch marks were produced using Wavelet transform approach, by using the glottal closure signal. This signal is obtained from the professional speaker during the recording.

O.Khalifa et al. (2004) [41], had identified the main steps for MFCCs are clearly shown in **Figure 2**. The main steps include the followings:

1. Preprocessing
2. Framing
3. Windowing [47]
4. DFT
5. Mel Filterbank
6. Logarithm
7. Inverse DFT.

It can be described clearly based on the block diagram shown below:

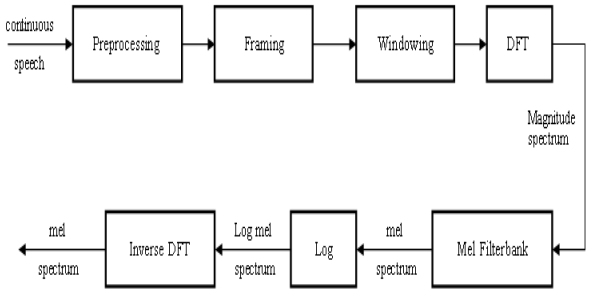

Fig. 2: Block diagram of the computation steps of MFCC (from O .Khalifa et al. [41])

Same as well as M.R. Hasan et al. (2004) [42], MFCCs have been used for feature extraction for security system based on speaker identification. Here, the pitch tone of the speech signal is measured on the 'Mel' scale. The mel-frequency scale formula based on mathematical equation is shown below:

$$\text{Mel}(f) = 2595 * \log_{10}(1 + f/700) \quad \ldots\ldots\ldots\ldots (1)$$

3.2.3 Spectrographic Analysis

The objective research of M. S. Bashir et al. (2003) [5] is to implement one such feature extraction strategy for Arabic language phoneme identification through spectrographic analysis. Based on the research, the different spectrograms represented by particular distinct bands within the spectrogram can be identified for each phoneme of Arabic language.

3.3 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. In other hand, feature matching/testing is a process of computing a matching score, which is the measure of similarity of feature extracted from unknown word and stored model in the database. For training and testing purposes, it is divided into 3 types, which are:

1. Hidden Markov Model (HMM)

2. Artificial Neural Network (ANN).
3. Vector Quantization (VQ)

3.3.1 Hidden Markov Model (HMM)

Training the HMM is the most crucial and important task for developing a robust HMM based Automatic Speech Recognition system. The training set for each model consists of several utterances, where each word model needs to be trained independently, due to generate the best likelihood parameters. Here, HMM trained is used by O.Essa [20] using the optimized segmentation, in order to detect phoneme boundaries.

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 method had been implemented by D.Jurafsky and J.H.Martin (2007) [47] in their research. Here, speech recognition systems train each phone HMM embedded in an entire sentence. Thus, the segmentation and phone alignment are done automatically as parts of the training procedure.

In other hand, the acoustic decisions trees used in synthesis are built from the HMM alignment. The HMM alignment is done by A.Youssef & O.Emam (2004) [11], where acoustic, energy, pitch and duration trees have been developed and executed, with the efficient maximum-likelihood algorithms existed for HMM training and recognition [48].

3.3.2 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) often called as Neural Network (NN). It is a computational model or mathematical model based on biological neural networks. ANN are made up from the artificial neurons interconnecting and it may either used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system.Artificial Neural Network (ANN) belongs to the Artificial intelligence approaches, which attempt to mechanize the recognition procedure. The procedure is depend to the way a person applies intelligence in visualizing, analyzing and characterized the speech based on a set of measured acoustic features [34].

According to X.Huang et al. (2001) [29], deal with non-stationary signals need us to address on how to map an input sequences to an output sequences properly. It could be happen when 2 sequences are not synchronous, where the proper alignment, segmentation and classification should be included. Thus, the basic neural networks are not well equipped to address these problems as compared to HMM's.

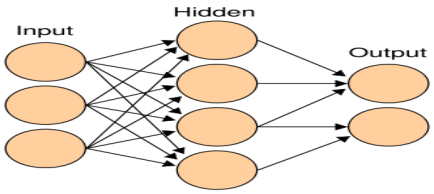

Fig. 3: Interconnected group of nodes in ANN (from X.Huang et al. [29])

3.3.3 Vector Quantization (VQ)

Quantization is the process of approximating continuous amplitude signals by discrete symbols. It can be quantizes on a single signal value or parameter known as scalar quantization, vector quantization or others [29].

Related to this topic, X.Huang et al (2001) [29] had described the vector quantizer as the codebook, which is a set of fixed prototype vectors or reproduction vectors. Each prototype vectors also known as a codeword. Due to perform the quantization process, each of input vector need to be matched against each codeword in the codebook, using distortion measure. Thus, the VQ process includes the distortion measure and the generation of each codeword for particular codebook involved. The goal of VQ is definitely on how to minimize the distortion [38].

VQ is divided into 2 parts, known as features training and features matching. Features training are mainly concerned with randomly selecting feature vectors and perform training for the codebook using vector quantization (VQ) algorithm.

Besides, the feature training is also involved with Vector quantization (VQ). The training process of the VQ codebook applies an important algorithm known as the LBG VQ algorithm, which is used for clustering a set of L training vectors into a set of M codebook vector. This algorithm is formally implemented by the recursive procedure: (Y. Linde et al. (1980) [45]. The following steps are required for the training of the VQ codebook using the LBG algorithm described by L. Rabiner and B.H. Juang (1993) [46]. VQ uses the Euclidean distance measure in matching/testing part, for comparing an unknown speech signal against the VQ codebook.

Figure 4 shows a block diagram of a vector quantizer, which consist of two main parts known as encoder and decoder.

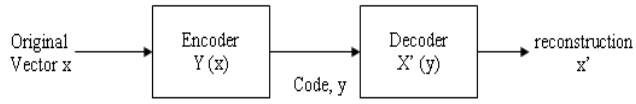

Fig. 4: The Encoding-Decoding in VQ (from C.C.Wai [37])

The task of encoder is to identify in which of N geometrically specified regions the input vector lays. Then, the decoder refers to the table lookup and is fully determined by specifying the codebook [37].

3.4 Features Classification and Pattern Recognition

According to the X.Huang et al. (2001) [29], spoken language processing relies heavily on pattern recognition, which is one of the most challenging problems for machines. The main objective of pattern recognition is to classify the object of interest into one of a number of categories or classes. The object of interest known as *patterns* and the classes here, refer to the individual words. Since the classification procedure applied on this research is applied on extracted features, thus it also can refer as feature matching. 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:

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

3.4.1 Hidden Markov Model (HMM)

K. Nathan et al. (1995) [21] had implemented the HMM's for recognizing handwritten words captured from a tablet. It is because Hidden Markov Model (HMM's) had been successfully applied for speech recognition system. Moreover, the output of the front-end was then used to feed the sphinx core recognizer, which uses the Hidden Markov Model (HMM) as the recognition tool. This recognition method, had been implemented by H.Tabbal et al. (2006) [4] in their research. The results of the recognizer have been used by a hash map to translate into the common Arabic words. An HMM generates a discrete time random process consisting of two sequences of random variables which hidden and the known observations. The underlying structure of HMM is the set of states and associated with the probabilities of transitions between states, known as Markov chain [32].

As described at section 3.3.1 under the HMM classification, HMM method had been fully implemented for both recognition and training purposes [48]. Under same research handled by D.Jurafsky and J.H.Martin (2007) [47], digit recognition task for HMM recognition have been used. A lexicon specifies the phone sequence, and each phone HMM is composed of three subphones with a Gaussian emission likelihood model. The observation of likelihood is computed by the acoustic model. Combining these and adding an optional silence at the end of each word will results in a single HMM for the whole task. Note that, the transition from the End state to the Start state to allow digit sequences of arbitrary length.

In order hand, recognition is also had been carried out by A.J. Viterbi (1967) [49] search in a large HMM. For context-independent phone recognition, an initial and a final state are created. The initial state is connected with null arcs to the initial state of each phonetic HMM, and null arcs connect the final state of each phonetic HMM to the final state. The final state is also connected to the initial state.

3.4.2 Vector Quantization (VQ)

The most successful text-independent recognition method is based on VQ. In this method, VQ codebook consists of a small number of representative feature vectors, which used as an efficient means of characterizing speaker-specific features. A speaker-specific codebook is generated by clustering the training feature vectors of each speaker, which described at part 3.3.3. In the recognition stage, an input utterance is vector-quantized using the codebook of each reference speaker and the VQ distortion accumulated over the entire input utterance is used to make the recognition decision. It is believe that, VQ-based method is more robust than a continuous HMM method, which had been stated by T. Matsui and S. Furui (1993) [51] in their research.

3.4.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN), also known as Neural Network (NN). ANN also mainly used as feature matching or recognition for speech processing. It normally used to classify a set of features, which represent the spectral-domain content of the speech (regions of strong energy at particular frequencies). The features then will be converted into phonetic-based categories at each frame. Then, Viterbi search is used to match the neural-network output scores to the target words (the words that are assumed to be in the input speech), in order to determine the word that was most likely uttered [52].

4. Conclusion

In this research, all different methods or approaches have been discussed, in order to find the most suitable method to be used in this project entitle "Quranic verse recitation recognition for support in j-QAF learning". After a while, method or approaches which logically can be used in this project had been decided. MFCCs method was decided to be used in feature extraction, because it implements the DFT and FFT algorithm. Moreover, majority of researches had used MFCCs, as their main features for extraction purposes.

In order hand, the training as well as recognition part will be conducted either using the HMM or VQ. Those 2 methods normally used in speech recognition purposes, recently. Moreover, these methods have shown great performance equally, through the different ways and expectations. Both methods have their own benefits and weaknesses. From the point of view, HMM is the most suitable methods used and this method have been implemented by most of researches in Arabic speech recognition. But, these methods had been implemented with speaker-dependent and not speaker-independent, with low percentage of accuracy. It totally different with VQ, which mostly used by the researches through their project of Speech Recognition, which related to English language.

Table 3: Approaches use by Quran Arabic Recitation recognition using speech Recognition techniques.

Reference	Pre-Processing	Feature extraction method	Classification/Recognition Technique	Performance
[H.Tabbal et al. '06]	Pre-emphasis filter	MFCC	Hidden Markov Model (HMM)	85% - 92%
[A.Youssef & O.Emam '04]	-	MFCC	Hidden Markov Model (HMM)	90.2%
[A.M.Ahmad et al. '04]	Digital filtering	MFCC LPCC	Recurrent Neural Network (RNN)	MFCC 95.9%– 8.6% LPCC 94.5%- 99.3%
[D.Vergyri & K.Kirchhoff '04]	Not stated	Not stated	Hidden Markov Model (HMM)	Not stated
[M.S.Bashir et al. '03]	Pre-emphasis filtering (Bandpass Filter)	Spectrographic Analysis	Spectrographic Analysis based on different frequency band of intensity.	93.33%
[K.Kirchhoff et al. '04]	Kneser-Ney smoothing	Not stated	Hidden Markov Model (HMM)	Not stated
[M.R.Hasan et al. '04]	-	MFCC	Vector Quantization (VQ)	57% - 100%
[S.K.Podder '97]		LPC	VQ and HMM	62% - 96%
[M.Z.A.Bhotto & M.R.Amin '04]	-	MFCC	Vector Quantization (VQ)	70% - 85%

5. References

[1] The *Holy Quran*

[2] Ninth Malaysia Plan 2006 – 2010, Chapter 11:"Enhancing Human Capital", Prime Minister's Office.

[3] Program j-QAF sentiasa dipantau", Berita Harian Online –10 Mei 2005.

[4] H. Tabbal, W. El-Falou, B. Monla, 2006. "Analysis and implementation of a "Quranic" verses delimitation system in audio files using speech recognition techniques".In: Proceeding of the IEEE Conference of 2nd Information and Communication Technologies, 2006. ICTTA '06.Volume 2, pp. 2979 – 2984.

[5] M.S. Bashir, S.F. Rasheed, M.M.Awais, S. Masud, S.Shamail,2003."Simulation of Arabic Phoneme Identification through Spectrographic Analysis." Department of Computer Science LUMS, Lahore Pakistan.

[6] "Institute for Research in Islamic education", The New Strait Times online – 26 September 2007.

[7] A. M. Ahmad, S. Ismail, D.F. Samaon, 2004." Recurrent Neural Network with Backpropagation through Time for Speech Recognition." IEEE International Symposium on Communications & Information Technology, 2004. ISCIT '04. Volume 1,pp.98-102.

[8] M. Habash, "How to memorize the Quran", Dar al-Khayr, Beirut 1986.

[9] M. E. Ahmed, 1991." Toward an Arabic Text-To-Speech system." The Arabic Journal Science and Engine, 1991.

[10] "How to read Quran/Arabic-Some Conversions" Citing Internet sources URL http://easyreadwritearabic.freeweb7.com/read_arabic_practice_1.html

[11] A. Youssef, O. Emam, 2004." An Arabic TTS based on the IBM Trainable Speech Sythesizer." Department of Electronics & Communication Engineering, Cairo University, Giza, Egypt.

[12] D. Vergyri, K. Kirchhoff, 2004." Automatic Diacritization of Arabic for Acoustic Modeling in Speech Recognition." COLING Workshop on Arabic-script Based Languages, Geneva, 2004.

[13] M. Maamouri, A. Bies, S. Kulick, 2006. "Diacritization to Arabic Treebank Annotation and Parsing." Proceedings of the Conference of the Machine Translation SIG, 2006.

[14] K. Kirchhoff, D. Vergyri, J. Bilmes, K. Duh, A. Stolcke, 2004." Morphology-based language modeling for conversational Arabic speech recognition." Eighth International Conference on Spoken Language ISCA, 2004.

[15] M. J. Anwar, M.M. Awais, S. Masud, Shafay Shamail," Automatic Arabic Speech Segmentation System." Department of Computer Science, Lahore University of Management Sciences, Lahore, Pakistan.

[16] A. El-Imam, 1989. "An Unrestricted Vocabulary Arabic Speech Synthesis system." IEEE Transactions on Acoustic, Speech and Signal Processing. Vol. 37. No. 12. December 1989.

[17] M. M. Assaf, 2005. "A Prototype of an Arabic Diphone Speech Synthesizer in Festival." Department of Linguistics and Philology, Uppsala University.

[18] K. Kirchhoff, J. Bilmes,S. Das, N. Duta, M. Egan, G. Ji, F. He, J. Henderson, J, D. Liu, M. Noamany, P. Schone, R. Schwartz, D. Vergyri, 2003."Novel approaches to Arabic speech recognition: report from the 2002 Johns-Hopkins Summer Workshop." IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003 (ICASSP '03). Volume 1, 6-10 April 2003, pp. I-344 - I-347 vol.1

[19] O. Essa, 1998. "Using Prosody in Automatic Segmentation of Speech." Proceeding 36th ACM Southeast Regional Conference, pp. 44 - 49, April 1998.

[20] O. Essa,"Using Suprasegmentals in Training Hidden Markov Models for Arabic."Computer Science Department, University of South Carolina, Columbia.

[21] H.S.M. Beigi,K. Nathan, J. Subrahmonia,"On-line Unconstrained Handwriting Recognition Based On Probabilistic Techniques."

[22] L. Josifovski, 1998."Robust Automatic Speech Recognition with Unreliable Data."

[23] K. Kirchhoff, 2002."Novel Speech Recognition Models for Arabic."Johns-Hopkins University Summer Research Workshop 2002.

[24] T.Pruthi, Ph.D, 2007."Analysis, vocal-tract modeling and automatic detection of vowel nasalization." University of Maryland, College Park, 2007.pp. 215 pages.

[25] T. Sari, L. Souici, and M. Sellami, "Off-Line Handwritten Arabic Character Segmentation Algorithm: ACSA," *Proc. Int'l Workshop Frontiers in Handwriting Recognition*, pp. 452-457, 2002.

[26] M. Ursin (2002). "Triphone Clustering in Finnish Continuous Speech Recognition." Master Thesis, Department of Computer Science, Helsinki University of Technology, Finland.

[27] J.P. Martens (2002)."Continuous Speech Recognition over the Telephone." Electronics & Information Systems, Ghent University, Belgium.

[28] M. Chetouani, B. Gas, J.L. Zarader and C. Chavy, 2002. "Neural Predictive Coding for speech Discriminant Feature Extraction: The DFE-NPC." ESANN'2002 Proceedings – European Symposium on Artificial Neural Network, Bruges, Belgium, pp. 275-280.

[29] X. Huang, A. Acero and H. W. Hon, 2001." Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Prentice Hall, Upper Saddle River, NJ, USA.

[30] J. Shen, J. Hung, L. Lee, 1998."Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments." 5th International conference ICSLP '98 Sydney, Australia, 1998."

[31] C. Avendano, S.V. Vuuren and H. Hermansky, 1996. "Data Based Filter Design for {RASTA}-like Channel Normalization in {ASR}."Proc. {ICSLP} '96", Philadelphia, PA. Volume-4, pp.2087-2090.

[32] J.C. Hansen, 2003." Modulation based parameter for Automatic Speech Recognition." Master Thesis of Department of Electrical Engineering, University of Rhode Island, USA.

[33] D.J.Kershaw, 1996."Phonetic Context-Dependency In a Hybrid ANN/HMM Speech Recognition System."PhD thesis,Cambridge University Engineering Department, September, 1996.

[34] V.K. Medisetti and D.B. William, 1999." Digital Signal Processing Handbook." CRC Press LLC.

[35] V. Siivola, T. Hirsimäki, and S. Virpioja, 2007." On Growing and Pruning Kneser–Ney Smoothed N-Gram Models." IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 5, July 2007.

[36] A. Stolcke, L. Ferrer,S. Kajarekar, E. Shriberg, A. Venkataraman," MLLR Transforms as Features in Speaker Recognition." Speech Technology and Research Laboratory, SRI International, Menlo Park, CA, USA and Department of Electrical Engineering, Stanford University, Stanford, CA, USA.

[37] C.C. Wai, 2003."Speech Coding Algorithm foundations and evolution of standardized Coders."John Wiley & Sons Inc.,NJ, USA.

[38] S.V. Vuuren, 1996."Comparison of Text-Independent Speaker Recognition Methods on Telephone Speech with Acoustic Mismatch". Proceeding (ICSLP)96, Vol:3, Philadelphia, PA. pp. 1788-1791.

[39] H.Hermansky,1990." Perceptual linear predictive (PLP) analysis of speech". The Journal of the Acoustical Society of America -April 1990. Volume 87, Issue 4, pp. 1738-1752.

[40] J. de Veth and L. Boves, 1998." Channel normalization techniques for automatic speech recognition over the telephone". Speech Communication 25 (1998) 149-164.

[41] O.Khalifa, S.Khan, M.R.Islam, M.Faizal and D.Dol, 2004."Text Independent Automatic Speaker Recognition".3rd International Conference on Electrical & Computer Engineering, Dhaka, Bangladesh,pp.561-564.

[42] M.R. Hasan, M.Jamil, M. G. Rabbani, M. S. Rahman, 2004." Speaker Identification Using Mel Frequency Cepstral Coefficients". **3**rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh ISBN 984-32-1804-4 565.

[43] S.K.Podder, 1997." Segment-based Stochastic Modeling for Speech Recognition". PhD Thesis. Department of

Electrical& Electronics Engineering, Ehime University, Matsuyama, pp. 790-77, Japan.

[44] M.Z.A.Bhotto and M.R.Amin, 2004."Bengali Text Dependent Speaker Identification Using Mel-frequency Cepstrum Coefficient and Vector Quantization".**3**rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh. ISBN 984-32-1804-4 569.

[45] Y.Linde, A.Buzo and R.M.Gray,1980." An algorithm for Vector Quantizer Design". IEEE Transactions on Communications, Vol.COM28,no 1,pp.84-95.

[46] L.Rabiner and B.H. Juang ,1993."Fundamentals of Speech Recognition".Prentice Hall, NJ, USA.

[47] D.Jurafsky and J.H.Martin, 2007."Automatic Speech Recognition".Speech and Language Processing: An Introduction to natural language processing, computational linguistics, and speech recognition.

[48] K.F. Lee and H.W.Hon, 1989." Speaker-Independent Phone Recognition Using Hidden Markov Models". IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 31, pp.1641-1648.

[49] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE Trans. Inform. Theory, vol. IT-13, pp. 260-269, Apr. 1967.

[50] P.Franti, T.Kaukoranta, and O.Nevalainen, 1997." On the Splitting Method for Vector Quantization Codebook Generation Codebook Generation". Optical Engineering, 36(11),pp.3043-3051.

[51] T. Matsui and S. Furui,1993." Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs". Proceedings of the 1993 International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Institute of Electrical and Electronic Engineers. Minneapolis, Minnesota, pp.157 –160.

[52] J.P.Hosom, R.Cole, and M. Fanty, 1999."Speech Spoken Language Understanding". Center for Spoken Language Understanding (CSLU) Oregon Graduate Institute of Science and Technology, July 6, 1999.

Zaidi Razak obtained his Bachelor's degree in Computer Science and Master in Chip Design from University of Malaya in 2000. Currently, he is a lecturer at the Faculty of Computer Science and Information Technology, University of Malaya. His research areas include image processing, Jawi character recognition and System on Chip (SoC) design. He has published a number of papers related to these areas.

Noor Jamaliah Ibrahim obtained her Bachelor's degree in Electrical Engineering from University Malaysia Pahang (UMP). Currently she is a research assistant at the Faculty of Computer Science and Information Technology, University of Malaya. Her research areas include speech processing, and Quranic recitation recognition.

Mohd Yamani Idna Idris obtained his Bachelor's degree in Electrical Engineering and Master's degree in Computer Science from University of Malaya. Currently, he is a lecturer at the Faculty of Computer Science and Information Technology, University of Malaya. His research areas include System on Chip, SCADA, image processing.

Emran Mohd Tamil obtained his Bachelor's degree in Electrical-Robotic Engineering from University Technology Malaysia and Master's degree in Information Technology from University Technology MARA. Currently, he is a lecturer at the Faculty of Computer Science and Information Technology, University of Malaya. His specializations are system and network and current research interests are embedded systems, network security, SCADA and chip design.

Mohd Yakub @ Zulkifli Mohd Yusoff obtained his Bachelor's degree, Master's degree and PhD from University of Malaya, University of Jordan and University of Wales, respectively. Currently, he is a lecturer at the Academy of Islamic Studies, University of Malaya. His research areas include Quranic Exegesis, Quranic Studies and Methodology of Quranic Exegesis.

Noor Naemah Abdul Rahman obtained her Bachelor's degree, Master's degree and PhD from University of Malaya, University of Jordan and University of Malaya respectively. Currently, she is a lecturer at the Academy of Islamic Studies, University of Malaya. Her research areas include Fatwa, Principles of Islamic Jurisprudence and Contemporary Fiqh.