I.J. Information Engineering and Electronic Business, 2017, 6, 1-9
Published Online November 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2017.06.01

ASR for Tajweed Rules: Integrated with Self-Learning Environments

Ahmed AbdulQader Al-Bakeri
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
Email: ahmedalbakiri@gmail.com

Abdullah Ahmad Basuhail
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
Email: abasuhail@kau.edu.sa

Received: 02 June 2017; Accepted: 01 August 2017; Published: 08 November 2017

***Abstract*—**Due to the recent progress in technology, the traditional learning setting in several fields has been renewed by different environments of learning, most of which involve the use of computers and networking to achieve a type of e-learning. With great interest surrounding the Holy Quran related research, only a few scientific research has been conducted on the rules of Tajweed (intonation) based on automatic speech recognition (ASR). In this research, the use of ASR and MVC design is proposed. This system enhances the learners' basic knowledge of Tajweed and facilitates self-learning. The learning process that is based on ASR ensures that the students have the proper pronunciation of the verses of the Holy Quran. However, the traditional method requires that both students and teacher meet face-to-face. This requirement is a limitation to enhancing individuals' learning. The purpose of this research is to use speech recognition techniques to correct students' recitation automatically, bearing in mind the rules of Tajweed. In the final steps, the system is integrated with self-learning environments which depend on MVC architectures.

***Index Terms*—**Automatic Speech Recognition (ASR), Acoustic model, Phonetic dictionary, Language model, Hidden Markov Model, Model View Controller (MVC).

I. Introduction

The Quran is the Holy book for the Muslims. The Quran contains guidance for life, which has to be applied by the Muslim people. In order to achieve this goal, it is important for the Muslims to understand the Quran clearly so they can be capable of applying it. Recitation is one of the Holy Quran related sciences. Previously, it was necessary to have a teacher of Quran and a student to meet face-to-face for the student to learn the recitation orally. It is the only certified way to guarantee that a certain verse of the Holy Quran is recited correctly.

Nowadays, because of the continuous increase and huge demand of people to learn the Quran, several organizations have started serving the learners by providing an online instructor in order to help them learn how to recite the verses of the Holy Quran correctly according to the rules of intonation (Tajweed science). Prophet Muhammad (peace be upon him) is the founder of the rules of this science, and certain Sahabah (companions of the prophet; may Allah be pleased with them) have learned from him the rules of pronunciation, and then those Sahabah have taught the second generation [1]. The process has continued up in the same manner till now.

The model proposed in this research facilitates teaching the recitation of the Holy Quran online so that students can practice Quran rules through using automatic speech recognition (ASR).

This approach of teaching has several benefits such as the delivery to individuals who cannot attend the Halaqat (sessions) held at the masjids (mosques), and it facilitates Quran teaching styles to receive more than the long-established model. The extreme importance of recitation and memorization of the Quran is due to the numerous benefits for readers and learners, as stated in Quran and Sunnah [2]. The learning of Quran is achieved by a qualified reciter (called sheikh qari) who has elder licensed linked to the transmission chain until it reaches the Messenger of Allah, Prophet Muhammad (peace be upon him). Detailed information about the Holy Quran and its sciences can be found in many resources; for example, see [3].

Due to the extensive use of the Internet and its availability, there is a strong need to develop a system that emulates the traditional way of the Quran teaching. There is some research that focuses on these issues, such as Miqra'ah, which is a server that uses virtually over the Internet.

The Holy Quran is written in the Arabic language, which is considered as a complex morphological language. From the perspective of ASR, the combination of letters is pronounced in the same way or different, depending on the Harakat used in upper and lower-case character [4]. Intrinsic motivation to develop ASR as participation to serve the Holy Quran sciences and its proposed approaches is needed to implement a system to

correct the pronunciation mistakes and integrate it within a self-learning environment. Therefore, we suggest the use of the Model View Controller (MVC) as a base structure that helps in massive development such as this research. Phonetic Quran is a special case of Arabic phonemes where there is a guttural letter followed by any other letter. This case is called guttural manifestation.

Gutturalness, in Quran, relates to the quality of being guttural (i.e., producing a particular sound that comes from the back of the throat). The articulation of Quran emphatics affects adjacent vowels.

There are many commercial packages that are available, such as audio applications to recite (Tarteel) the Holy Quran. One among these packages is the Quran Auto Reciter (QAR) [5]; however, this application does not support the rules of Tajweed to verify and validate the Quranic recitation.

The field of speech recognition in Quranic voice recognition is a significant field, where the processing and acoustic model has a relation with Arabic phonemes and articulation of each word; thus, the research in the recitation of Quran could be taken from a different aspect.

In general, and especially in computer science, there are substantial research achieves meant to produce worthy results in the correction of the pronunciation of the Quran words according to the rules of Tajweed. Hassan Tabbal has done research on the topic of automated delimiters, which extracts ayah (verse) from an audio file and then converts verses of Quran into an audio file using the technology of speech recognition tools. The developed system depends on the framework of Sphinx IV [6].

Putra, Atmaja, and Prananto developed a learning system that used speech recognition for the recitation of the Quranic verses to reduce obstacles in learning the Quran and to facilitate the learning process. Their implementation depended on the Gaussian Mixture Model (GMM) and the Mel Frequency Cepstral (MFCC) features. The system produces good results for an effective and flexible learning process. The method of template referencing was used in that research [7]. Noor Jamaliah, in her master thesis in the field of speech recognition, used Mel Frequency Cepstral Coefficients (MFCC) for extracting feature from the input sound, and she used the Hidden Markov Model (HMM) for recognition and training purposes. The engine showed recognition rates that exceeded 86.41% (phonemes), and 91.95% (ayates) [8].

Arabic sound is among the first of the world's languages that have been analyzed and described. The articulation manner and place of each sound in Arabic were documented and identified in the eighth century AD by a famous book written by Sibawayah called Al-Kitaab. Since then, not much work has been added to the treatise of Sibawayah. Recently, King Abdulaziz City for Science and Technology (KACST) has been doing research on the Arabic Phonetics using many tools that treated signals and captured images from glottis, which also include the air pressure, side and front facial images, airflow, lingual-palatal contact, perception and nasality. The raw data of KAPD are available on 3 CDs for researchers [9]. A research on e-Halagat is demonstrated in [10]. Noureddine Aloui with other researchers used Discrete Walsh Hadamard Transform (DWHT), where the original speech is converted into stationary frames, and then applied the DWHT to the output signal. The performance is evaluated by using some objective criteria such as NRMSE, SNR, CR, and PSNR [11].

Nijhawan and Soni used the MFCC for feature extraction to build Speaker Recognition System (SRS) [12]. The training phase was done by calculating MFCC, executing VQ, finding the nearest neighbor using Euclidean distance, and then computing centroid and creating codebook for each speaker. After the completion of the training process, the testing phase is achieved through calculating MFCC, finding the nearest neighbor, finding minimum distance and then decision making.

Reference [13] presented the use of DTW algorithm to compare between the MFCC features extraction of the learner and the MFCC features of the teacher, which was stored previously in the server. DTW is a technique used for measuring the distance between the student's speech signal and the exemplar's (teacher) speech signal. The results of DTW comparison is given the closest number to zero when the two words are similar and greater than zero if the two words are differentiated.

Carnegie Mellon University has developed a group of speech recognition systems called Sphinx. Sphinx 3 is among these systems [14]; speech recognition was written in C for a decoder; a modified speech recognition was written in Java, and the Pocketsphinx is a lightweight speech recognition library written in C as well [15]. Sphinx-II speech recognition can be used to construct medium, small, or large lexicon applications. Sphinx is a speaker-independent recognition system and continuous speech using statistical language n-gram and hidden Markov acoustic models (HMMs).

II. The Research Problem

Through progress in technology, the traditional learning environments in several fields have been renewed and now primarily use computer systems and networks to achieve a type of e-learning. With the great interest in the Holy Quran research, there is little scientific research that has been conducted in regard to the rules of Tajweed based on ASR and using a helpful architecture that could help to enhance the e-learning environment. Depending on speech recognition technology, open-source speech recognition tools are of interest in the research, not just in the Holy Quran sciences but also to build learning environments for different languages. It is important to use automatic speech recognition to train the system to recognize the Quranic Ayat (verses) recited by different reciters. When such a system is built, it will improve the level of learning one can achieve through reading the Holy Quran. At the same time, there are no time limits imposed on the student to learn. The learning will be dependent on his available time, because he can use the system whenever

he is free. This method establishes a great new environment for the learner to practice the recitation of the Quran based on the Tajweed rules.

There is no database for the Holy Quran, which can be engaged directly to the training process, so our goal is to implement this database and make it available to the other researchers.

III. Research Methodology

The recitation sound is unique, recognizable and reproducible according to specific pronunciation rules of Tajweed. The system's input is transcription phonetic of a speech utterance and a speech signal. Thus, this research requires having a reciter to take samples out of input speech, extraction of features, training features, pattern classification and matching. These stages are essential components of a verse recitation formulation for speech recognition architecture. Automated speech recognition for checking Tajweed rules is illustrated in Fig. 1. It demonstrates the correction of the learner's verse recitation. The training phase and matching phase are included in this system. The algorithm of Hidden Markov Model (HMM) was selected for feature training, feature extraction, and pattern recognition.

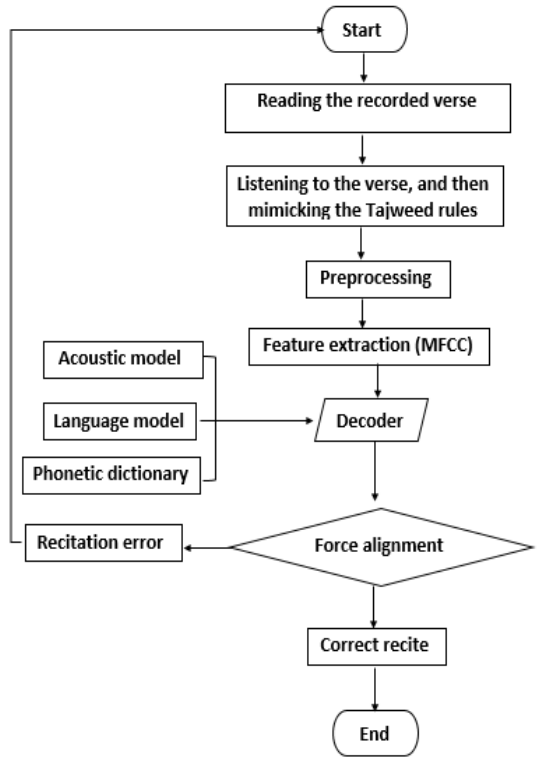

Fig.1. Automated speech recognition system for checking Tajweed rules.

The steps of the approach to speech recognition are to get a waveform, divide it on utterances by silences, and then attempt to recognize what the speaker said in each utterance. We try to match all possible combination of words with audio by selecting the best results of the match combinations. The essential components of the matching process include:

(1) Features: when the number of parameters is large, we attempt to enhance it. By splitting speech on frames, we can calculate the numbers from speech. The length of each frame is typically ten milliseconds.

(2) Model: here, the mathematical object describes the model and the commonly spoken word attributes are gathered using a mathematical object. Hidden Markov Model is the speech model. The process in this model is presented at sequential states which change in certain probability with one another. The speech is described through this sequential model.

(3) The process of matching itself, which compares all models with all feature vectors. At each stage, we get the best matching variants that we maintain and extend to produce the best results of matching in the next frame.

Speech recognition requires the combination of three entities in order to produce a speech recognition engine. These three entities are the acoustic model, phonetic dictionary, and language model.

The properties of the acoustic model for the atomic acoustic unit are also known as the senone. The phonetic dictionary includes a mapping from phones to words. To restrict word search in Hidden Markov Models, we use the language model. This model expresses the word which could follow previously recognized words. The matching is achieved in a sequential process and helps to restrict the process of matching by disrobing words that could not be probable. N-gram language models are the most popular language models where the finite state automation defines the speech sequences.

A. *Transcription file*

The link between the Quranic Ayat and their audio files is achieved through the transcription file. The delimiters <s> and </s> are used for the transcription of the audio file's contents which consist of Quranic Ayah, so each audio recorded and used in the ASR engine should be uniquely identified as it is written in the transcription file. Table 1 illustrates an example of the transcription file for one Surah (chapter) of the Holy Quran.

Currently, in the audio file, we have recorded one of the authors' voice as a reciter and the voices of 10 famous reciters of the Holy Quran.

Table 1. Transcription File for Surah Al-lkhlas

<s> بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ </s> (alhosari112_0)
<s> قُلْ هُوَ اللَّهُ أَحَدٌ </s>(alhosari112_1)
<s> اللَّهُ الصَّمَدُ </s>(alhosari112_2)
<s> لَمْ يَلِدْ وَلَمْ يُولَدْ </s>(alhosari112_3)
<s> وَلَمْ يَكُنْ لَهُ كُفُوًا أَحَدٌ </s>(alhosari112_4)

According to this template, the audio file is indexed as: first is the reciter, followed by chapter number, followed by an underscore, and finally the number of Ayah.

B. *Corpus*

The corpus consists of the voice of Quran reciter. This is a vocal database with a sample rate of 16 kHz and a mono wave format, as presented in Table 2. It's important

that the duration of the silence at the beginning and at the end is no more than 0.2 seconds.

Table 2. Recording Parameters

Parameters	Values
Sampling	16khz, 16-bit
wav format	Mono wav
Corpus	Al-Ikhlas, Alrahman
Speakers	10 reciters

The high recognition rate is based on the corpus preparation, where the chosen word should be selected carefully to be representative of the language and saved as high quality. The words should be exchanged with the selected language, but for our case, we have chosen two chapters of the Holy Quran, which can be extendable to train more chapters in the corpus.

C. Acoustic Model Training

The Hidden Markov Model is provided through the components of the acoustic models and uses the Quranic tri-phones to recognize an Ayah of the Holy Quran. The structure of HMM is presented in Fig. 2. The figure shows the basic structure of HMM. There are five states with three emitting states used to present the acoustic model of tri-phone. The Gaussian mixture density is used to train the state emission. In this representation, a letter followed by two numbers, such as a12, is the probability transition, and the b1, b2, and b3 are the emission probabilities. The Gaussian Mixture probabilities with Hidden Markov Model are called Continuous Hidden Markov Model (CHMM). The term P (xt | j) is the probability of xt observation with given transition state j; q0, q1, q3...qt is a state sequence; $N_{j,k}$ is a k-th Gaussian distribution, and $W_{j,k}$ is the mixture weights. Its equation is:

$$b_j(x_t) = P(x_t | q_t = j) = \sum_{x=2}^{M} W_{j,k} N_{j,k}(x_t) \quad (1)$$

The effective technique to build speech recognition with a large vocabulary is through the CHMM method.

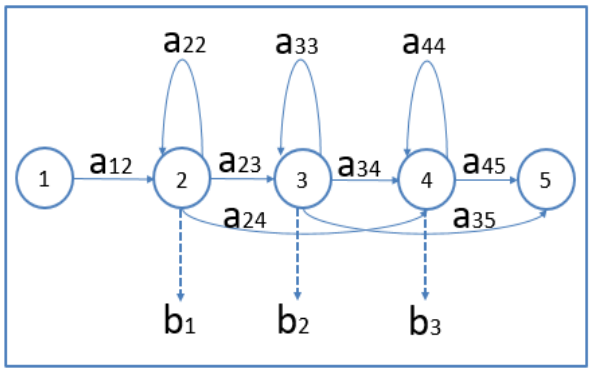

Fig.2. Bakis Model structure.

D. Quranic Language Model

The grammar used in the system is getting through processing the Quranic text in certain statistical steps that generate the Quran language model. The toolkit of cmuclmtk tools [16] is used here to get the uni-grams, bi-grams, and tri-grams. Fig. 3 illustrates the language model process. The steps to create the Quran language model is to first count the uni-gram words. The second step is to convert it to task vocabulary, the input as a word unigram file, which is the output of text2wfreq. The output is a vocabulary file, where the file contains the word corresponding to its number of occurrences. The third step is to produce the tri-grams and bi-grams based on the previous vocabulary. At the end, the output is converted to a binary format or to the ARPA (Advanced Research Projects Agency) format language model.

The language model format should be delimited between <s> and </s> tags. The text consists of diverse sentences, so each utterance is indicated by two signs; the first is <s>, which is bounded as the start of sentences, and </s> to mark the end of sentences.

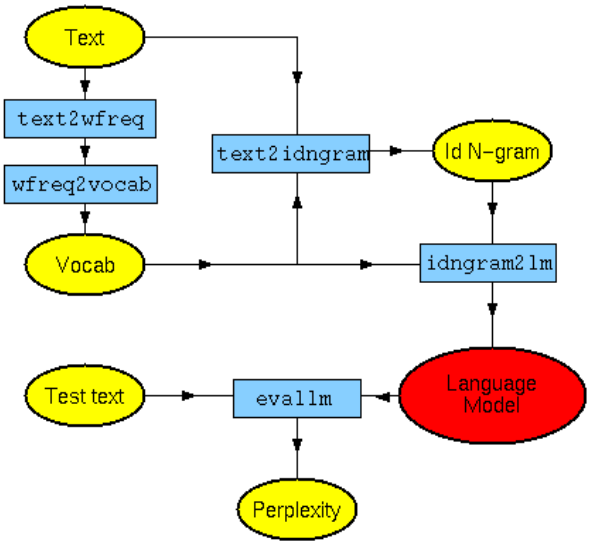

Fig.3. Language Model Creation.

The fundamental difference between n-gram models depends mainly on the N chosen. It's difficult to get the entire word history probability in a sentence, so in this research, we use the method of N-1 words such as the trigram model n=3, which takes the two previous words into account while the n=2 takes the two preceding words only, and the n=1 takes one word at a time.

E. Phonetic Dictionary

The phonetic dictionary involves all phonemes that are used in the transcription file, where the phonemes are the symbolic representation of spoken words in the audio files. The dictionary now is dynamically created through coding in Python language. The diacritic marks such as ˊ, and ˏ are considered in the process. Table 3 shows the mapping between a word and its phone. We have generated these marks using the syllable. The results of these two different schemes are shown in the results section.

Table 3. Phonetic Transcription for Surah Al-lkhlas

E AE: L AE: E IH	ٱلاءِ
E AE: N IH N	ٱنِ
E AE HH AE D UH N	أَحَدٌ
E AE F N AE: N IH N	أَفْنَانِ
E AE L L AE:	أَلَا
E AE Q TT AH: R IX	أَقْطَارِ
E AE N	أَنْ
E AE Y Y UH H AE	أَيُّهَ
E IH S T AE B R AA Q IX N	إِسْتَبْرَقٍ
…….	……

Several languages are not supported by CMUdic to produce the dictionary file, so we can do this in several ways. The Arabic language is one of those languages that are not supported by CMUdic, so we have created the dictionary of Arabic language, considering the rules on the lookup dictionary. Some languages provide a list of phonemes which the programmer can use to automatically generate the phoneme. In our case, we chose the dictionary building algorithm using three famous techniques used to produce the pronunciation dictionary. These techniques are:

1- Rule based
2- Recurrent neural network (RNN)
3- Lookup dictionary

There is difficulty in producing the pronunciation file due to some issues, such as irregular pronunciation. There is open-source software that can be used to produce a mapping between a word and its phonemes; espeak is a software that can be used to create a phonetic dictionary. Many languages have tools that can reduce the time needed to build the dictionary file, and then it can be used thereafter.

IV. Set of Arabic PhonemeS

Each Arabic phoneme corresponds to its English representation symbol, as is shown in Table 4. The chosen phoneme symbol is taken into consideration of the English ASR phoneme, and it's closely similar to Arabic phoneme. Specifically, the set of phonemes that we have used depends on the research that has been done by KACST about Text-to-Speech systems [17,18]. The /AE/, /UH/, and /IH/, are symbols of short vowels in the Arabic language, which represent the diacritical marks Fatha, Damma, and Kasra respectively. The pharyngealized allophone of the /AE/ is /AA/.

The pharyngealized allophone of Damma /UH/ is /UX/, and for Kasra /IH/, it's /IX/. The /UW/ is the long vowel for Damma, followed by 'و,' and /AE:/ for Fatha, followed by 'ا,' and /IY/ is for Kasra, followed by 'ي'; these /UW/, /AE:/, and /IY/ are considered long vowel allophones of the Arabic language.

The long vowel length is generally equivalent to two short vowels; /AW/ is the diphthong of Fatha and Damma, and the /AY/ is the diphthong of Fatha and Kasra. It comes when Fatha appears before undiacritized 'ي,' while the /AW/ acts when an undiacritized 'و' appears and Fatha comes before that. The /T/ and /K/ correspond to 'ت' and 'ك' respectively. They are counted as voiceless stops letters—closely to their English counterparts. The Dhad letter 'ض', corresponds to /DD/ in its English counterpart.

Table 4. Phoneme List for Arabic Letters

Phoneme	Arabic Letter	Phoneme	Arabic Letter	Phoneme	Arabic Letter
/B/	ب	/AE/	ـَ	/DD/	ض
/T/	ت	/AE:/	ـَا	/TT/	ط
/TH/	ث	/AA/	ـَ	/DH/	ظ
/JH/	ج	/AA:/	ـَ	/Al/	ع
/HH/	ح	/AH/	قَ	/GH/	غ
/KH/	خ	/AH:/	قَا	/F/	ف
/D/	د	/UH/	ـُ	/Q/	ق
/DH/	ذ	/UW/	ـُو	/K/	ك
/R/	ر	/UX/	غُ	/AY/	ـَي
/Z/	ز	/IH/	بِ	/M/	م
/S/	س	/IY/	بِي	/N/	ن
/SH/	ش	/IX/	غِ	/H/	هـ
/SS/	ص	/AW/	ـَو	/W/	و
/E/	ء	/L/	ل	/Y/	ي

The phone /Q/ is a representation for emphatic Arabic letter 'ق'; /E/ is a representation for plosive sound 'ء,' and phone /G/ represents the 'ج'. The representation phones for voiced fricative letters in Arabic are /DH/, /Z/, /GH/, and /AI/, which are 'ظ' 'ز' 'غ' 'ع'. The Arabic phones that

are similar to resonant of the English phones, are /R/ for 'ر,' /L/ for 'ل', /W/ for 'و', and /Y/ for 'ي'.

V. Integrated Environments

ASR redirects learners to its site when they log into the system as demonstrated in Fig. 4. There are several choices that appear on the page to help learner's follow-up with their learning progress and allow them to communicate with their instructor. In addition, the system provides the ASR to allow learners listen to the Quran reciter and then record their sounds so the ASR engine corrects them in the case of incorrect pronunciation. Moreover, a teacher can revise the progress history of each student (the student's scores will be displayed on his page), and they can offer advice to their students to improve their level of learning. The administrator is responsible for adding, deleting, updating the information of any staff in the system and managing the entire system. The administrator is also responsible for the assignment of students to certain teachers and allocating the proportion of students to each instructor. In addition, he can assign the students to groups according to their ages, such as students under the age of 20 being allocated through the administrator and moved to the proper class.

The system can make rulers (parents) monitor the progress of their children.

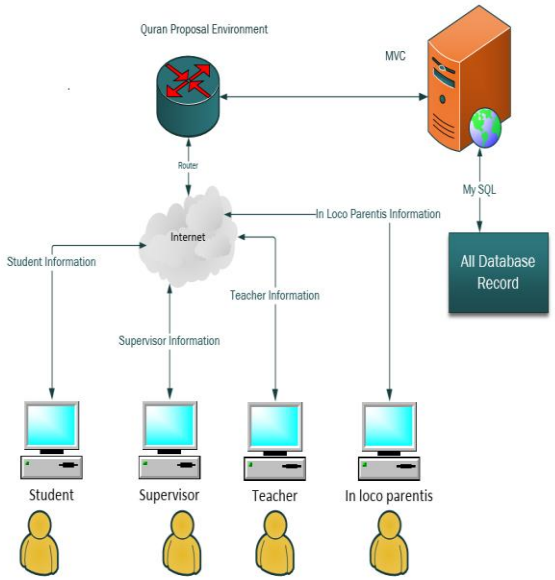

Fig.4. Integrated Environments with ASR.

VI. Results and Discussion

The word error rate (WER) is metric that assesses the ASR performance. In the ASR results, there is percentage of error, which refers to the number of errors in the misrecognized words that occurred in the speech. Thus, the WER is a measurement of the performance of ASR. The situation regarding the continuous speech recognition has some differences. In continuous speech recognition, the WER is not efficient enough to measure the performance of ASR because the sequence of words in continuous speech recognition has additional errors that could occur in the results of ASR. First is word substitution; it occurs when there is a replacement of a word and when an incorrect word is put in place of a correct word, such as when exactly the speaker speaks a word that the ASR engine recognizes as another word. The second error that could happen in the continuous speech recognition is the deletion of words. Word deletion happens when the speaker spells a word, but this word is not recognized in the results of ASR. At the end, the last error is an insertion—where the actual spoken word is recognized, and there is an extra word not spoken, but the ASR system recognized it as spoken word.

We used the phonemes described in the previous section with Surah (chapter) 112 from the Holy Quran. The number of tried states is static for now, and the Gaussian Mixtures dimension has a straight effect on the speech recognition performance. The training for data brings two choices. The first training is context-dependent, which is used for large data, and the second is context-independent, which is used to train the system that has a short data. In other words, when we have a small amount of data, we can use the context-independent training more effectively.

In the training process, the first scenario we have followed is the use of phonemes, and the second scenario is the use of the syllables to train the system. We found that using the syllable to train the data is workable for small data where we get 100% accuracy of the system with a syllable as shown in Table 6. These results were achieved without any insertion, deletion, or substitution errors, but when we increased the amount of data, this accuracy decreased. So, we decided to work on phonemes rather than syllables.

As shown in Table 5, the total training words is 19, the correct words are 18, and there are two error words. The total percentage of correct words is 94.74%, while the rate of error is 10.53%, and the accuracy rate is 89.47%. The insertions are one, and the deletions are one, while the substitutions are zero.

Table 5. Quran Automatic Speech Recognition for Surah Al-lkhlas

Ayah	Words	Subs	Dels	Ins	%Accuracy
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ	4	0	0	0	100
قُلْ هُوَ اللَّهُ أَحَدٌ	4	0	1	0	75
اللَّهُ الصَّمَدُ	2	0	0	1	50
لَمْ يَلِدْ وَلَمْ يُولَدْ	4	0	0	0	100
وَلَمْ يَكُنْ لَهُ كُفُوًا أَحَدٌ	5	0	0	0	100

Building the ASR based on the syllable is an alternative method where there are two approaches that can be followed to segment the speech into units. Using syllables to segment speech brings greater results than using phonemes. The testing word for phonemes and syllable is 19 words. The accuracy of the syllable in ASR is 100%, and 89.47% for using phonemes. Using syllables is a better choice for ASR in some cases, such as when the training data is not large. If there are many utterances needed to be converted to their corresponding syllable, then the conversion process is going to be more complex due to the absence of rules for the creation of syllable units. The limitation on the number of syllables decreases the accuracy of the system. The determination of the syllable's boundary is a difficult process as well.

Table 6. Automatic Speech Recognition for Surah Al-lkhlas using Syllable

Ayah	Words	Subs	Dels	Ins	%Accuracy
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ	4	4	4	0	100
قُلْ هُوَ اللَّهُ أَحَدٌ	4	4	4	0	100
اللَّهُ الصَّمَدُ	2	2	2	0	100
لَمْ يَلِدْ وَلَمْ يُولَدْ	4	4	4	0	100
وَلَمْ يَكُنْ لَهُ كُفُوًا أَحَدٌ	5	5	5	0	100

Table 7 shows comparisons between the syllables and the results of the phonemes. The process of training is based on the Hidden Markov Model for syllable and phoneme. The changing is done on the pronunciation dictionary to test the system based on two different approaches that can be used to build the dictionary.

Tables 8 and 9 help to take the proper combinations of different reciters, where we can exclude the inapplicable

results of recognition to increase the accuracy of the system. The chosen word and its right pronunciation is a reflective job, where the results we have is affected by the way of building the pronunciation file for each word. The substitution is seen as complex work due to its need for more states than insertion and deletion. The Percentage Correct and Word Accuracy are calculated using the following equations:

$$Percentage\_Correct = 100 * \left(\frac{Words\_Correct}{Correct\_length} \right) \quad (2)$$

$$Word\_Accuracy = 100 * \left(\frac{Correct\_Length - (Subs - Dels - Ins)}{Correct\_length} \right) \quad (3)$$

Table 7. Comparison between Syllable and Phonemes

	1 Surah		2 Surahs	
	Phonemes	Syllable	Phonemes	Syllable
Words	19	19	383	383
Correct	18	19	288	258
Errors	2	0	109	151
% Correct	94.74	100	75.20	67.36
% Error	10.53	0	28.46	39.43
% Accuracy	89.47	100	71.54	60.57
Insertions	1	0	14	26
Deletions	1	0	51	48
Substitution	0	0	44	77

Table 8. Quran ASR Using Phonemes (Second Five Reciters)

Reciter	1	2	3	4	5
Total Words	383	383	375	378	375
Correct	258	288	219	262	184
Errors	151	109	219	130	195
% Correct	67.36	75.20	47.47	69.87	49.07
% Error	39.43	28.46	58.40	34.67	52.00
Accuracy	60.57	71.54	41.60	65.33	48.00
Insertions	26	14	22	17	4
Deletions	48	51	104	45	66
Substitution	77	44	93	68	125

Table 9. Quran Automatic Speech Recognition Using Phonemes (Second Five Reciters)

Reciter	6	7	8	9	10
Total Words	375	375	375	375	375
Correct	221	275	278	236	277
Errors	173	127	105	157	109
% Correct	58.93	73.33	74.13	62.93	73.87
% Error	46.13	33.87	28.00	41.87	29.07
Accuracy	53.87	66.13	72.00	58.13	70.93
Insertions	19	27	8	18	11
Deletions	67	19	54	48	25
Substitution	77	44	93	68	125

To determine the accuracy of the system, we need the information about insertion, deletion, and substitution. This type of information is computed by aligning the hypothesis string with the correct utterance using a string match algorithm.

We found the best recognition result with the Qari Swed, where the correction rate is 74.13%, and the accuracy rate is 72%. The second-best results are with the Qari Alhosari, where the correction rate is 75.2%, and the accuracy rate is 71.5. The lowest correction rate of accuracy comes from the Qari Ayyub with 375 words while the Qari Alkalbani comes before the last with a correction rate of 49.07%.

We have tested our training data on two chapters of the Holy Quran (Suras: Al-lkhlas and Alrahman). We have used 10 reciters, namely: Ahmed, Alhosari, Ayyub, Alhuthaify, Alkalbani, Alakhdar, Altablawy, Swed, Abdulbaset, and Elsayed.

Fig. 5 below demonstrates the highest and the lowest numbers of insertion, deletion, or substitution for each reciter. The total number of words for each Qari is different. It is dependent on the number of utterances used for each Qari, where some Qari will repeat some words to stop in the proper time to avoid giving meaningless Aya. Ahmed and Alhosari have 383 words, while the remaining Qura'a (reciters) have 375 words. Alkalbani got the highest value in substitution, and Ayyub got the highest value in deletion, while the highest value for insertion was for Altablawy. The Qari Alkalbani has the highest value in substitution, and although this result is not sufficient for the recognition process, it's better than the insertion and substitution errors. If the reciter (Qari) has read the Holy Quran text using a variety of rhythms, then the recognition process becomes complex due to how the representation of the words will yield similar phones as it happens with the Qari (Alkalbani).

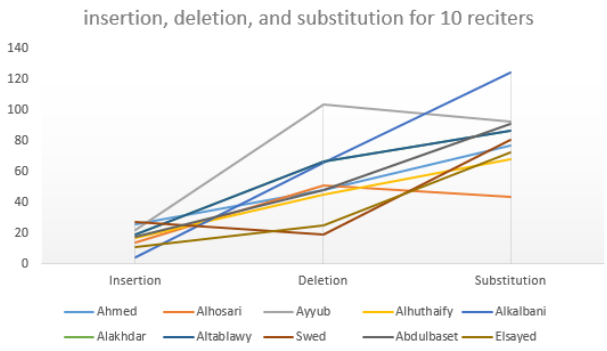

Fig.5. Insertion Deletion, and Substitution for 10 reciters.

The results below show the correction percent of each utterance after it's converted to its phonemes using the training pronunciation dictionary compared to grammar and language models that we created to generate the hypothesis. As shown in Table 10, the correction percentage in the first two hypotheses is 100%. In the third hypothesis, the correction percentage is 83.33% due to the deletion in the recognized utterance as it appears. When we compare this to the original text, we notice that the word /ربك/ is dropped out, so the total number of words is 6, while in the hypothesis, it's just 5 words. The 5 is divided by 6 and multiplied by 100, which equals the

83.33% correct word. The next verse is / فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ /. The number of words here is 4, where there is one error that occurs as a substitution error, so the number of corrected words in the hypothesis divided by the number of words in the original text and multiplied by 100 gives a 75% correction in the recognized text.

To find the percentage of error for one hypothesis, we need to divide the number of errors in the recognized text by the total number of words in the original text.

Table 10. Sample of Correct Percentage for Each Utterance

Original Text	Recognized Text	% Correct
قُلْ هُوَ اللَّهُ أَحَدٌ	قُلْ هُوَ اللَّهُ أَحَدٌ	100.00
فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ	فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ	100.00
تَبَارَكَ اسْمُ رَبِّكَ ذِي الْجَلَالِ وَالْإِكْرَامِ	تَبَارَكَ اسْمُ رَبِّكَ الْجَلَالِ وَالْإِكْرَامِ	83.33
فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ	فَبِأَيِّ آلَاءِ رَبِّكُمَا لَمْ	75.00
لَمْ يَطْمِثْهُنَّ إِنْسٌ قَبْلَهُمْ وَلَا جَانٌّ	أَنْ يَطْمِثْهُنَّ إِنْسٌ قَبْلَهُمْ وَلَا	66.67
فِيهِنَّ قَاصِرَاتُ الطَّرْفِ لَمْ يَطْمِثْهُنَّ إِنْسٌ قَبْلَهُمْ	فِيهِنَّ قَاصِرَاتُ قَبْلَهُمْ أَنْ يَطْمِثْهُنَّ إِنْسٌ قَبْلَهُمْ	71.43

VII. Conclusion

In this research, we have developed an automatic system using the phonemes with the Arabic language to train the system to get ASR for the Holy Quran. The system is based on an open-source tool using speech recognition CMU Sphinx tools that also contains the HMM model code which was written in C-language. We have wrapped the C code to access library by using the Node.js. The results obtained here are encouraging in proceeding further with this research.

The Quran speech recognition was created using our phonemes model, which was designed by using the Lookup Dictionary to test the accuracy of automatic speech recognition. The training process was achieved by using diacritical marks in the training file as was presented in the results section. The flexibility of CMU Sphinx tools is helpful when it is allowed to use the Arabic characters to create the pronunciation dictionary or create it by using the English characters. The Quran speech data needs to be modeled, so we have generated the language model and trained the acoustic to be used for the decoding process.

The Decoder is needed and has three main parts to decode the input sound. The first part is the acoustic model, which needs to be trained previously; the second one is the language model, which is created from Quranic text, and the third one is the pronunciation Dictionary, which was generated by developing our code written in Python.

The speech recognition has gotten complex or has a bad quality of recognition based on the units of sound that was used in the dictionary file. Each diaphone, phoneme, or diphthong has its own use in the dictionary and has pros and cons for the recognition process. We have explained the phonetic properties for the Holy Quran and for the Arabic language in general. We have used the phonemes as the smallest sound units to present the Quranic words. In our examination of our code to get the times needed for translating all the Holy Quran to its corresponding phonemes, our results showed that about 10 minutes was needed to translate the entire Holy Quran into its phonemes.

The HMM is used to characterize the features of the Holy Quran signal, which is the famous statistical model used for speech recognition. In HMM, the assessing parameters were covariance probability and means for each Quranic phoneme. The identification of utterance is determined by the phoneme that has the highest probability score. The benefits of using Sphinx tools include the fact that its program contains many tools involving the language model tools, the acoustic model, and the sphinxtrain toolkit as well. The software is open-source so that MATLAB can be used. However, it does not have support for the Android platform, as the sphinx has other tools called pocketsphinx.

The alignment of identified words against the correct word of Aya makes the result to be produced by three terms, namely, how many insertions, deletions, and substitution, so the correctness is presented as a number. The Quranic speech is processed as a continuous speech recognition, and its process matches the request parameters. Because there is no available acoustic model for the researchers, we made our acoustic model using 10 reciters for two chapters of the Holy Quran. The system can be enhanced with more reciters involved in the training data—as we intended to perform in future studies.

References

[1] سنن سعيد بن منصور، دراسة وتحقيق د. سعد بن عبدالله بن عبدالعزيز آل حميد، دار الصميعي، المملكة العربية السعودية، ج: 5 ص: 257

[2] Islam City Website. [Online]. http://www.islamicity.com/mosque/quran/

[3] Islam Way Website. [Online]. The Most Famous Broadcast Website about Islam. http://en.islamway.net/

[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.

[5] http://www.searchtruth.com/download.php.

[6] A.-F. W. a. M. B. Tabbal Hassan" ،Analysis and Implementation of an Automated Delimiter of "Quranic" Verses in Audio Files using Speech Recognition Techniques " ،"Robust Speech Recognition and Understanding: 351.

[7] B.Putra , B.T. Atmaja , D.Prananto,’’ Developing Speech Recognition System for Quranic Verse Recitation Learning Software’’, International Journal on Informatics for Development (IJID), Vol. 1, No. 2, 2012.

[8] N. Jamaliah, I. “Automated Tajweed Checking rules engine for Quranic verse Recitation”, MCS Thesis, University of Malaya, Faculty of Computer Science & Information Technology, Department of Computer System & Technology, Kuala Lumpur, April 2010.

[9] Alghmadi M., "KACST Arabic Phonetic Database, "The Fifteenth International Congress of Phonetics Science, Barcelona 2003, pp 3109-3112.

[10] Yahya O. Mohamed ELHADJ,’’ E-HALAGAT: AN E-LEARNING SYSTEM FOR TEACHING THE HOLY QURAN’’, TOJET: The Turkish Online Journal of Educational Technology – January 2010, volume 9 Issue 1.

[11] Noureddine Aloui, Souha Bousselmi, Adnane Cherif," Speech Compression Based on Discrete Walsh Hadamard Transform", IJIEEB, vol.5, no.3, pp.59-65, 2013. DOI: 10.5815/ijieeb.2013.03.07
[12] Geeta Nijhawan, M.K Soni,"Real Time Speaker Recognition System for Hindi Words", IJIEEB, vol.6, no.2, pp.35-40, 2014. DOI: 10.5815/ijieeb.2014.02.04
[13] Alkhatib. B, Kawas. M, Alnahhas. A, Bondok. R, Kannous R,"BUILDING AN ASSISTANT MOBILE APPLICATION FOR TEACHING ARABIC PRONUNCIATION USING A NEW APPROACH FOR ARABIC SPEECH RECOGNITION", Journal of Theoretical and Applied Information Technology, Vol.95, No.3, pp.478-489, 2017.
[14] K. Seymore, S. Chen, S. Doh, M. Eskenazi, E. Gouvea, B. Raj, M. Ravishankar, R. Rosen-feld, M. Siegler, R. Stern, and E. Thayer, "The 1997 CMU Sphinx-3 English BroadcastNews Transcription System," in Proc. of the DARPA Broadcast News Transcription andUnderstanding Workshop, Lansdowne, USA, 1998.
[15] D. Huggins-Daines, M. Kumar, A. Chan, A. Black, M. Ravishankar, and A. Rudnicky, "Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand Held Devices," in Proc. of the ICASSP, Toulouse, France, 2006.
[16] The CMU-Cambridge Statistical Language Modeling Toolkit v2, [Online], http://www.speech.cs.cmu.edu/SLM/toolkit_documentation.html.
[17] M. Elshafei Ahmed, " Toward an Arabic Text-to-Speech System", in the special issue on Arabization, the Arabian Journal of Science and Engineering, Vol. 16, No. 4B, pp.565-583, October 1991.
[18] Moustafa, Al-Muhtaseb Husni, Al-Ghamdi Mansour," Techniques for high quality Arabic speech synthesis ", Information Sciences 140(3-4): 255-267 (2002).

Authors' Profiles

Abdullah Basuhail, received the Ph.D. degree in computer engineering from Florida Institute of Technology, Melbourne, FL, USA in 1419H/1998G. His research interests include: digital image processing, computer vision, the use of computer technologies, applications, information technology in e-teaching, e-learning, e-training and e-management supportive systems. Dr. Basuhail was an ex-member of the Saudi Computer Society, the IEEE, and the IEEE Computer Society.

Ahmed Al-bakeri, received the BSc degree from Taibah University, Madinah, Saudi Arabia, in 2013,and then he worked as a programmer in (Cooperative Office for Call & Guidance) at Al-Madinah AL munawwarah, currently he working to toward the MSc degree in the department of computer sciences at the University of King Abdul-Aziz. His current research interests are in the areas of speech recognition (ASR), and human computer interaction.

How to cite this paper: Ahmed AbdulQader Al-Bakeri, Abdullah Ahmad Basuhail," ASR for Tajweed Rules: Integrated with Self-Learning Environments", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.9, No.6, pp.1-9, 2017. DOI: 10.5815/ijieeb.2017.06.01

