Verification System for Quran Recitation Recordings

Ayat Hafzalla Ahmed
College of Computer
Science and Information Technology
Sudan University of
Science and Technology

Sherif Mahdi Abdo
Faculty of Computers at
Cairo University Egypt
The Research and
Development International (rdI)

ABSTRACT

Quran is the holy book of Allah which was revealed to Prophet Mohammed.

Quran 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.

The Quran should be recited in Arabic language as it is with neither additions nor subtractions.

When the Arabs started to mix with the non Arabs as Islam spread, mistakes in Quran recitation started to appear, so the scholars had to record the rules of tajweed and write them down in order to preserve the Qur'an recitation as revealed by Allah. In this regard, it is necessary to preserve the authenticity and integrity of the Quran from all sorts of corruption or deletion. This paper provides an overview of the techniques used in voice recognition in the Quran recitation focusing on the techniques used, the advantages, and drawbacks. And proposed model of verification system for Quran verses.

General Terms

Artificial intelligent, Speech Recognition

Keywords

Quran, Arabic Language, Quran recitation, Computer Aided Pronunciation Learning.

1. INTRODUCTION

The enormous advances in the computer technologies in the twentieth century permitted computers to effectively contribute in various fields of human life. The Holy Quran is the book that was revealed to the prophet of Islam Muhammad. More than one billion people in the world today recite verses of the Quran everyday during praying. Recitation should be done according to rules of pronunciation, intonation and caesuras established by the Islamic prophet Muhammad, these rules are called Tajweed. Programs in the field of second language learning took huge strides in recent years. Very few human tutors have the unlimited amount of time, patience and flexibility to practice individually at any hour that the recitation of the holy Quran differs from the normal reading of Arabic text due to special rules called Tajweed.So it is upon every Muslims to observe those rules to recite Quran in the way it was revealed. Another word to describe the recitation of Quran is Tarteel. While obeying to the same laws of tajweed, Tarteel is generally identified by a faster reading in contrast with tajweed where reading is in slower way and there is more focus on the artistic aspect of reading the holy Quran [14]. Creating a computer aided system for automatic learning f the holy Quran recitation is practical way

of learning because it is very difficult to provide human teacher for each learner or group of learners. Computer Aided Pronunciation Learning (CAPL) has received a considerable attention in recent years. Many research efforts have been done for improving such systems especially in the field of second language teaching [1][7][17][15][6]. Computer programs can be an important complement to language teachers for pronunciation training and accent improvements in foreign language classes[16]. No one can deny the use of computer in automated tutor has. Computer aided pronunciation learning allows the possibility of acquiring knowledge without leaving home or bedroom. A challenging application for CAPL is the automatic training for correct recitation of the Holy Quran. In contrast to the foreign language training task where a wide variety of pronunciations can be accepted by native speakers as being correct. The Holy Quran has to be recited the same way as in the classical Arabic dialect and the tolerance for allowed variation is very fine[1][3]. This paper provides an overview of the techniques used in voice recognition in the Quran recitation and proposed model for Quran recitation verification system.

2. LITERATURE REVIEW

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

2.1 Definition of Holly Quran

Quran or Holy Quran is holy book for Muslims its revelation from the god sent to all people through prophet Mohamed. Quran explains all things about Muslims day life so studying, memorizing and understanding Quran play central role in Muslims life. With the proliferation of information technology a huge number of applications were developed to help Muslims learning the Holy Quran .These applications cover all the Quran knowledge like reciting the Quran with Tajweed rules, understanding the meaning of its verses (Tafseer) and memorizing it. Quran is divided to 114 chapters each of it is called Sorra. Each Sorra contains number of verses. Holly Quran consists of 6236 verses. Each Sorra also differs from one another in the number of its verses [31].

1) The Quran knowledge includes the following fields: Reciting the Quran with Tajweed (rules of recitation of the Quran.

2) Tafseer (explain the meaning of the Quran verses).

3) Memorization of the Quran verses.

4) Searching in Quran for verses or words including semantic search.

5) Quran Recitation and Bookmarks.

6) Authentication of Quran verses available in various online documents.

7) Speech Recognition technologies for learning the Quran recitations [2].

2.2 Arabic Language

Arabic is a Semitic language and it is one of the oldest languages in the world. It is the fifth widely used language nowadays [20].Arabic Language is a language of Quran .it is the second language in the world in terms of the numbers of its speakers. Arabic language is not a single linguistic variety; rather, it is a collection of different dialects (e.g., Gulf Arabic, Middle East Arabic, and North Africa Arabic). In general, there are two classes of Arabic language:

1) The classical Arabic which used in the Quran.

2) Modern standard Arabic used as an academic language (in schools and universities).

Phonologically, the Arabic language has 34 phonemes: 25 of them are consonants and 3 are semivowels, and 3 long vowels and three short vowels not represented by letters but marked by the so-called diacritics (short strokes placed either above or below the preceding consonant). So Arabic has fewer vowels than English. It has three long and three short vowels, while American English consists from at least 12 vowels [21].Arabic is a Semitic language whose main characteristic feature is that most words are built up from roots by following certain fixed patterns and adding infixes, prefixes, and suffixes [8].

2.3 Speech Recognition System

According to the existing studies each speech recognition system consists of four major stages. Under the same techniques of speech recognition, the Quran Arabic recitation recognition system 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.

2.3.1 Pre processing:

The main benefit of pre processing step of speech recognition is to organize the information and simplify the data (input speech signal) and simplify the recognition. Pre-processing steps mainly consist of the following

o Silence removal: in most cases reciters do not recite the holy Quran rhythmically and their utterance of words is slow. So there will be silences between uttered words. W.M. Muhammad et al.[10] suggested calculating the energy of each frame and removing the frames which have energy extended to Zero so this gives a rich stream (words in the speech) for processing. So, the overall time length of the signal is reduced, since blank spaces are removed. In order to remove silences, the short-term energy method (Mark Greenwood et al., 1999) can be used [27].

• Pre Emphasis: after removing silence, pre emphasis is performed on signals, giving rise to higher frequencies with respect to magnitude of lower frequency, improving the signal to noise ratio. The present echoes within the signal are also eliminated because of that it is also known as a noise cancelling filter [9].

2.3.2 Feature Extraction:

The goal of feature extraction stage is to extract important characteristics from the speech signal that are unique, discriminative, robust and computationally efficient to each word, which then used to differentiate between different words [11]. According to J. P. Martens (2000)[9] There are various speech features extraction techniques as listed below:

1) Linear Predictive Coding (LPC).

2) Perceptual Linear Prediction (PLP).

3) Mel-Frequency Cepstral Coefficient (MFCC) [9].

4) Spectrographic Analysis.

Linear Predictive Coding (LPC) is not considered as a good method, since it reduces high and low order Cepstral coefficient into noise when coefficient are transferred into Cepstral coefficient.

Perceptual Linear Prediction (PLP) is first introduced by H.Hermansky (1990)[28] it is based on the Nonlinear Bark scale. The PLP is designed for speech recognition with removing of speaker dependent characteristics [9]. (PLP) is better than LPC, since the spectral features remains smooth within the frequency band in PLP and the spectral scale is non-linear Bark scale [5] .

Spectrographic Analysis is used for Arabic language phoneme identification. Arabic phonemes are identified by spectrograms that are represented by distinct bands [5].

According to Ursin, M. (2002), 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 is able to obtain a better accuracy with a minor computational complexity, respect to alternative processing as compared to other feature extraction techniques. The MFCC Processes are:

1) Take short-time Fourier transform for the signal every 10 ms with 20 ms Hamming window,

2) Map the power obtained in 1. Into the Mel scale using triangular overlapping windows.

3) Compute the log-energy of each filter output.

4) Applied Discrete Cosine Transform (DCT) to the filter bank output.

2.3.3 Training and Testing

The training is a process of enrolling 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, Testing is a process of computing a matching score, which is determining the measure of similarity of feature extracted from unknown word and stored model in the database. Three methods for training and testing are listed below:

1) Hidden Markov Model (HMM).

2) Artificial Neural Network (ANN).

3) Vector Quantization (VQ).

Artificial Neural Network (ANN) often called as Neural Network (NN). It is a mathematical model based on biological neural networks.ANN is used to gain an understanding of

biological neural networks, or to solve artificial intelligence problems. Artificial Neural Network (ANN) is one of the Artificial intelligence approaches, which attempt to computerize the recognition procedure [29].

X.Huang et al (2001) [30] had described the vector quantizer as the codebook. VQ is a set of fixed prototype vectors. Each prototype vectors also known as a codeword. Quantization is performed by matching input vector against each codeword using distortion measure. The goal of VQ is definitely on how to minimize the distortion [30].

2.3.4 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. The goal is to know patterns and classes referred to individual words.

1) There are three methods used for pattern matching, classification as well as recognition. These methods are: 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[13].

3. CAPL System Design

The CAPL system consists of six main blocks:

1) Verification HMM models.

2) Speaker Adaptation.

3) Pronunciation hypotheses generator.

4) Confidence Score Analysis.

5) Phoneme duration analysis.

6) Feedback Generator .

4. QURAN VERIFICATION SYSTEM STRUCTURE DESCRIPTION

The Basic design aspects of speech verification systems used to verify the recitation of the Holy Qur'an Its main blocks are:

1) Verification HMM models: Is the acoustic HMM models for the system.

2) Speaker Adaptation: Is used to adapt acoustic models to each user acoustic Properties in order to boost system performance.

3) Pronunciation hypotheses generator: It analyzes Holy Qur'an current prompt and generates all possible pronunciation variants that are fed to the speech recognizer in order to test them against the spoken utterance. It also generates mapping information to determine the appropriate location for the feedback of each candidate hypothesis [24].

4) Confidence Score Analysis: It receives n-best decoded word sequence form the decoder, then analyzes their scores to determine whether to report that result or not.

5) Phoneme duration analysis: For phonemes that have variable duration according to its location in the Holy Quran, this layer determines whether these phonemes have correct lengths or not.

6) Feedback Generator: Analyze results from the speech recognizer and user selectable options to produce useful feedback messages to the user [3][25].
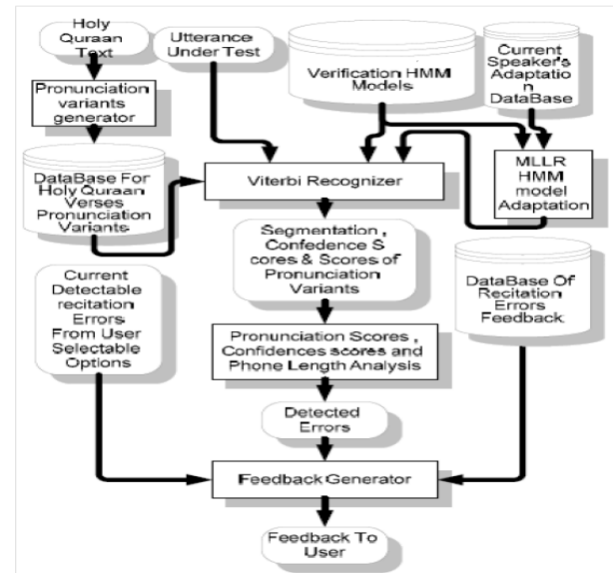

Figure1: Block diagram of the Quran verification system

5. RELATED WORKS

Recently there are many researchers focused on holly Quran recitation in terms of automatic speech recognition (ASR) such as in [9].Verification of Quranic Verses in Audio Files using Speech Recognition Techniques to preserve the authenticity and integrity of the Quran from all sorts of corruption by identify all sort of errors in the Quranic audio files. In this paper an overview of all techniques used in voice recognition in Quran recitation are explained .The paper also proposed a model for a system which will be capable for identifying errors or mistakes in Quran recitation. This system will help non Arabic speaking Muslims to verify Quran verses in Audio Files. In their system they will use MFCC as feature extraction technique and will use the HMMs with 3 states in tow main stages in a speech recognition system, which are training and recognition stages.

Another system in [4] tajweed checking rule system support recitation aims to Support learners reciting the holy Quran without the need for teacher. The proposed system provides alternative methods for learning Al Quran and act as a supplements to the existing conventional method of learning Al-Quran recitation (Talaqqi and Musyafahah).,however the verification algorithm of proposed Focused only on the selected law of Tajweed rules of selected verse. This system use MFCC algorithm for feature extraction and HMM for feature classification.

.Another system for speech recognition technique is in [5] E-Hafiz: This is an Intelligent System to facilitate the learning and Memorization of Holly Quran. The system also helps in Quran recitation learning by minimizing errors or mistakes during recitation of Holly Quran (These mistakes may include the wrong utterance of words, misreading words, and punctuation and pronunciation mistakes.). The current system

has the ability to detect mistakes on word level as the back end database is made at word level. The dataset of E-Hafiz consists of 10 experts with database of first 5 Sorras of the Holy Quran.

The E-Hafiz is the expert reader that pointing out errors made during Quran recitation process; however, the E-hafiz system is only for people who are well versed with Tajweed rules for Quran recitation.

The principal phases in E-Hafiz architecture are: data preparation, feature extraction, and modeling, storing and comparison. One of the famous research in the field of Quranic voice recognition is introduced by Hassan tabbal in [4].that is an automated delimiter which can extract the verses from the audio file using the open source Sphinx framework and speech recognition techniques. The sphinx framework must be configured using an xml based configuration file. Sphinx core recognizer uses HMM as a recognizer tool. In whole processing, two models acoustic and language are discussed. A set of phonemes symbols was used to train the state of HMM that corresponds to the acoustic model.

For this system a language model based on the Java Speech Grammar Format (JSGF) specification was taken that is compatible with sphinx framework and fulfills all the requirements of the system. These models take parts with feature vectors to generate search space for HMM nodes.

The system can help users to extract a required verse form audio files however it is not useful for the Muslims who want to learn the recitation of Holly Quran (users who are not Arabic speakers don't benefit from this system) the system is useful for Muslims who know tajweed rules well. The system used MFCC feature extraction techniques for its implementation.

In [18] a makhraj recognition system is recitation verification Application based on correct makhraj. The system is proposed to reduce the duration of learning al Quran from the expert. A system using Mel Frequency Cepstrum Coefficient (MFCC) as feature extraction and Mean Square Error (MSE) as a pattern matching technique is considered in order to develop a makhraj recognition system. An experiment has been setup to measure the system performance in terms of accuracy based on False Reject Rate (FRR) and Wrong Recognition. This system has successfully recognized the correct makhraj for one-to-one mode which give an average of accuracy of 100 which is promising result. However, the system has an obstacle for one to many recognition. Besides, the system is only focusing on the makhraj part of the Al-Quran recitation.

In [19] the system here 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) the system is implemented by using Spectral Subtraction for pre processing of speech signal, Mel Frequency Cepstral Coefficient (MFCC) as feature extraction and Artificial Neural Networks (ANN) for acoustic modelling and pattern matching. The system is implemented in MATLAB. The system achieves 88 accuracy on sentence recognition. However, the system was trained on distinct sentences only. The dataset consists of few small Quranic verses (only the verses of Sorrat alnas).

6. PROPOSED SYSTEM

According to the existing studies, the process of recitation recognition is commonly divided into; pre-processing, feature extraction, training and testing, feature classification, as well as pattern recognition.

The proposed method used in this research is that input speech signal for the Quran recitation from the scholar is sampled and extracts through segmented features which were gain from the input speech signal and then this features will be processed.

In the Training phase the input voice for the correct Quran recitation recorded from the verified scholar will be stored in the Database as reference recitation.

In the Testing phase the step of our proposed system is occur This step is Audio segmentation here in this step the recitation audio file of the scholar will be segmented according to the places of silence the next step is to align the Quran text with segmented audio and this is very challenging step since we will have undetermined begin and end for the specified part which the scholar is reciting .there is many possibilities ,In the start the scholar can repeat part of the previous verses or only read next verses without repetition other issue is that the place in which he stop reading depend on the length of his breathe and the last issue is that his reading may include some errors (it may be Tajweed error or misspelling error or adding or deleting word error) the last step in the proposed model is the verification for reading error in this step the recitation of the scholar will be compared with the reference recitation in order to locate the mistake type (can be Tajweed error or misspelling error or adding or deleting word) in the recitation of the scholar.

Then the difference between the test segment and the set of database will be calculated and this test segment later will verified. Mel-Frequency Cepstral Coefficient (MFCC) technique will be used. The features of recorded voices will be extracted using MFCC and will be compared with correct recitation stored in database. Any mismatch between the scholar recitation and correct recitation will be showed to the scholar in order to know his mistake and correct it.
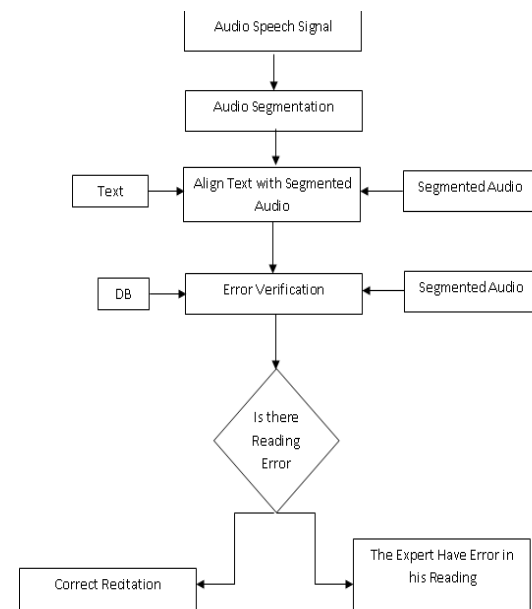

Figure 2: The proposed system

7. CONCLUSION

Errors in Quran recitation are forbidden because Quran is book revealed from god, so all Muslims are responsible to

maintain the holy Quran against distortions and alternation so the scholar's recitation must be verified to maintain the Quran as it is. Here a model of Quran recitation Verification is proposed. The model will be design to gain high accuracy in verification of scholar Quran recitation. This model will depend on robust technologies in Arabic speech recognition, such as MFCC for the Feature Extraction phase and HMMs in recognition phase and matching phase.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Abdou, Sherif Mahdy, and Mohsen Rashwan. "A Computer Aided Pronunciation Learning system for teaching the holy quran Recitation rules." 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA). IEEE, 2014.

[2] [2] Zameer Ahmed Adhoni1, Husam Al Hamad1, Abdul Ahad Siddiqi2, Mohammd Parvez1, Zaheer Ahmed Adhoni3 "Cloud-based Online Portal and Mobile Friendly Application for the Holy Qur'an". Life Science Journal (2013); 10(12s) .http://www.lifesciencesite.com.

[3] [3] S. Hamid "Computer aided pronunciation learning system using statistical based automatic speech recognition". PhD thesis, Cairo University, Cairo, Egypt, 2005.

[4] [4] Ahsiah. I, Noor. N.M, Idris. M.Y.I." Tajweed checking system to support recitation." ICACSIS (2013). ISBN: 978-979-1421-19-5

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

[6] Silke M. W, (2012). Automatic Error Detection in Pronunciation Training: Where we are and where we need to go. Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training June 6 - 8, 2012 KTH, Stockholm, Sweden.

[7] H. Franco et al., "The SRI EduSpeak system: Recognition and pronunciation scoring for language learning", Proc. of InSTIL, Scotland, 123-128, 2000.

[8] Harrag, A. and Mohamadi T., " QSDAS: New Quranic Speech Database for Arabic Speaker Recognition", the Arabian Journal for Science and Engineering, vol. 35, no. 2C, pp. 7-13, December (2010).

[9] Ammar Mohammed 1*, Mohd Shahrizal Bin Sunar 2,." Quranic Verses Verification using Speech Recognition Techniques" Jurnal Teknologi (Sciences & Engineering) 73:2 (2015), 99–106.

[10] N. J. I. M. Y. I. I. E. M. T. M. Y. Zaidi Razak " ‹†Quranic Verse Recitation Recognition Module for Support in j-QAF Learning: A Review "‹IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August (2008).2118.

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

[12] Nengheng Zheng, Tan Lee and P. C. Ching, "Integration of Complementary Acoustic Features for Speaker Recognition", International Journal of Signal Processing Letter, IEEE, vol. ASSP-14, no.3, pp. 181-184, (2007).

[13]]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.

[14] H. Tabbal, W. El-Falou and B. Monla, Analysis and Implementation of a "Quranic verses delimitation system in audio files using speech recognition techniques. Proc. of the IEEE Conf. of 2nd Information and Communication Technologies 2006; 2(1): 2979 – 2984.

[15] A. Hagen and B. Pellom, "Data driven sub-word unit modeling for speech recognition and its application to interactive reading tutors," in Proc. European Conference on Speech Communication and Technology, Lisbon, Portugal, Sept. 2005, pp. 2757–2760.

[16] J. Mostow, et al., "Evaluation of an automated Reading Tutor that listens: Comparison to human tutoring and classroom instruction". Journal of Educational Computing Research, 29(1), 61-117, 2003.

[17] D. A. Williams, "Knowing what you don't know: roles for confidence measures in automatic speech recognition", Ph.D. thesis, Department of Computer Sciences, University of Sheffield, Sheffield, United Kingdom, 1999.

[18] Nurul Wahidah Arshad, Suriazalmi Mohd Sukri, Lailatul Niza Muhammad, Hasan Ahmad, Rosyati Hamid, Faradila Naim, Noor Zirwatul Ahlam Naharuddin "Makhraj Recognition for Al-Quran Recitation using MFCC". International Journal of Intelligent Information Processing (IJIIP).Volume4, Number2, June 2013.

[19] Bushra Abro, Asma Batool Naqvi and Ayyaz Hussain, "Qur'an Recognition For The Purpose Of Memorisation Using Speech Recognition Technique", 978-1-4673-2252-2/12/$31.00 © (2012) IEEE.

[20] M. Al-Zabibi, "An Acoustic–Phonetic Approach in Automatic Arabic Speech Recognition," The British Library in Association with UMI, 1990.

[21] J. Deller, J. Proakis, J.H. Hansen, "Discrete- Time Processing of Speech Signal," Macmillan, NY, 1993.

[22] Bassam A. Q. Al-Qatab , Raja N. Ainon, "Arabic Speech Recognition Using Hidden Markov Model Toolkit(HTK)", 978-1-4244-6716-711 0/$26.00 ©(2010) IEEE.

[23] J. H. M. Daniel Jurafsky. Speech and Language Processing, An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall, Upper Saddle River, New Jersey 07458,(2000).

[24] S. Hamed, M. Rashwan, "Automatic Generation of Hypotheses for Automatic Diagnosis of Pronunciation Errors" First International Conference on Arabic Language Resources and Tools, Cairo, Egypt, 22-23 April 2004.

[25] Abdou Sherif Mahdy and Mohsen Rashwan "Performance Evaluations for a Computer Aided Pronunciation Learning System".

[26] Noor Jamaliah Ibrahim, Zaidi Razak, Mohd Yakub, Zulkifli Mohd Yusoff, Mohd Yamani Idna Idris and Emran Mohd Tamil, Quranic verse Recitation feature extraction using Mel-Frequenc Cepstral Coefficients (MFCC). Proc. the 4th IEEE Int. Colloquium on Signal Processing and its Application (CSPA) 2008, Kuala Lumpur, MALAYSIA. (2008).

[27] SUVing, Automatic Silence/Unvoiced/Voiced Classification of Speech. Undergraduate Coursework, Department of Computer Science, the University of Sheffield, UK. (1999).

[28] 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.

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

[30] 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.

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