

# Holy Qur'an Speech Recognition System Distinguishing the Type of prolongation

Bilal Yousfi<sup>1</sup>, Akram M. Zeki<sup>2</sup>, Aminah Haji<sup>1</sup>

## Abstract:

The act of learning and teaching of the Holy Quran has become a scientific practice to Muslims around. The stakeholders are faced with a huge challenge when it comes to the principle of application of Tajweed (that is, the rules guiding the pronunciation during the recitation of the Quran). There are several efforts made by previous systems on the development of feasible guiding techniques to the act of Tajweed. Unfortunately, liking the major control variables of the practices of Tajweed in those approaches were neglected. In order to fill this gap, this research presents a speech recognition system that distinguishes the types of Madd (elongated tone) or prolongation and the type of Qira'at (method of recitation) related to Madd. The proposed system is capable of recognising, identifying, pointing out the mismatch and discrimination between two types of Madd namely, The greater connective prolongation and The Exchange Prolongation rules for Hafss and Warsh for the verses that contains the two rules, that were made by the expert found in a database. Furthermore, this study used Mel-Frequency Cepstral Coefficient (MFCC) and Hidden Markov Models (HMM) as feature extraction and feature classification respectively.

**Keywords:** *Holy Quran; Tajweed; Qira'at; Sound, Mel-Frequency; Hidden Markov Models.*

## 1. Introduction

The Holy Qur'an was revealed with Tajweed rules, and it's important for readers to apply those rules during recitation. According to Qira'at science, each Qira'at has its own rules of Tajweed. Table 1 shows the difference between Hafss and Warsh in terms of the greater connective prolongation and the exchange prolongation. The act of Tajweed is the body of knowledge perfecting and laying the path to the understanding of the articulation of the Holy Quran letters and reaching the utmost level in pronouncing them properly. It is quite obvious that the application of the rules of recitation of Holy Quran can be performed by giving every letter of the Qur'an its rights and dues. Various lexical characteristics emerge when reciting the Qur'an and observing the rules that apply to those letters in different situations. For instance, an onwards break and the need for reciters to halt for either

a short period of time or longer. This is a unique attribute that influences people. This tends to be habitual and in some certain situation, even without due consideration of rules, it becomes an important part of learning the Holy Quran. However, the facts still lie with variation of short or long period needed. Many other features and rules that need to be taken in to consideration make it necessary for researchers to formulate techniques to ease the learning and teaching the act of Tajweed.

The act of Tajweed brings along some well-defined rules of recitation of Al-Quran. Noticeably, those rules create a big difference between normal Arabic speeches and the Quranic verses. Madd or prolongation is one of the significant tajwid rules. It stands for extending or prolonging sound with a letter of the Madd. It is divided into two groups:

The Original Madd and The Secondary Madd. The later presentations do not involve a

<sup>1</sup> Kulliyah of Information and Communication Technology (KICT)

<sup>2</sup> International Islamic University Malaysia (IIUM)

Corresponding Email: : [yousfi.bilal@hotmail.fr](mailto:yousfi.bilal@hotmail.fr)

hamzah before it, and there should not be a hamzah or sukoon after it. Whereas the former has a longer timing (or the possibility of longer timing) than that of the natural Madd. This comes into being because of the present of a hamzah or a sukoon before and after it. Considering these presentations and with the availabilities of two rules of prolongation (mudud);the greater connective prolongation and the exchange prolongation rules under The Secondary Madd. This research focuses on their dynamics.

The Exchange Prolongation: rule substitute prolongation which occurs when a hamza (ء) precedes a half Madd (ا or ي or و). This Madd is only found within one word and occurs when the hamza has the respective diacritic on it, for example. if the harf Madd 'waaw' follows a hamza, the hamza has a dammah on it [1].

The Greater Connective Prolongation : rule of the greater connective prolongation occurs If the pronoun/possessive pronoun (هـ) is at the end of a word and it has a vowel of a dhammah or a kasrah, is between two voweled letters, and the first letter of the next word is a hamzah, it is permissible to lengthen according to the type of recitation [1].

The rules of prolongation are directly attached to Tajweed rules, these are mostly studied independently, however, their correct application is most reliable when performing subjective assessment in the present of Tajweed experts. That is, where experts of Tajweed are involved in the treatment of the rules governing the Tajweed application. Unfortunately, it's difficult to get experts at any time while in Quranic learning and teaching application practices. In most cases, many people will like to find Tajweed expert who will listen to them and point out mistakes if any while in Quranic learning session. Thus, it has become an important task to develop a learning software that will aid the practices of Tajweed in the act of learning Quran. This will guide people to practise reading of the Quran in correct way of Mudud spelling. Crucial to this is utilizing a speech recognition system for

distinguishing the type of Madd as well as the type of Qira'at.

The remaining part of this paper is organized as follows. The current section is section one, followed by section two which represents this research related work. Section three describes the speech recognition approach, section four present the research methodology and section five presents the outcomes of this research. Finally, section six presents the conclusion.

**TABLE 1.** The difference between the Greater Connective Prolongation and the Exchange Prolongation according to Hafss and Warsh Qira'at.

| The Type of mudd                                     | Hafss                    | Warsh                        |
|--|--------------------------|------------------------------|
| The Exchange lengthening مد البدل                    | lengthened 2 counts      | lengthened 2, 4, or 6 counts |
| The greater connective prolongation م د الصلة الكبرى | lengthened 4 or 5 counts | lengthened 6counts           |

## 2. Previous Work

Previously, great effort has been made mostly for the study of Holly Quran Arabic speech recognition. There are many reviews of the evolution of Arabic Holy Qur'an ASR. Several theories were proposed for evaluating high accuracy for Arabic Qur'an speech recognition [2].

One of the most important researches in this area presented by [3]. The use of Multilayer Perceptron for classification of the pronunciation of Qalqalah Kubra (a Tajweed rule) has been presented. Feature extraction technique using MFCC has been utilized to extract the characteristics from Quranic verses' recitation. The technique used was able to achieve recognition rate within the range of (95% to 100%). Thus, the study has contributed to identifying correct and incorrect Qalqalah Kubra pronunciation.

The needs for people to be checking Tajweed rules in Quran verses by using

interactive way of learning without the guidance and presence of an expert has been emphasized in [4]. Thus, the design, implementation and evaluation of an automated Tajweed checking rules engine for Qur'anic learning were also presented. This has been validated with MFCC features and HMM model, where an accuracy of 91.95% (ayates) and 86.41% (phonemes) were obtained.

Generally, people were mostly involved in independent practice of listening of recitation of famous reciters with the aim of learning from that. This is appropriate but Makhraj might be missing, thus a novel technique based on correct Makhraj has been proposed in [5]. The technique was evaluated with s MFCC as feature extraction and Mean Square Error (MSE) as a pattern matching technique. Thus, accuracy of the approach based on False Reject Rate (FRR) and Wrong Recognition (WR) has been obtained, where the percentage of FRR for all recitation is 0% and the accuracy of the system is 100%.

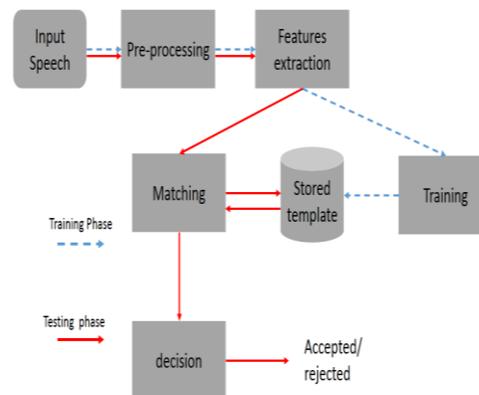
A novel model which distinguishes the type of recitation of holy Quran has been proposed in [6]-[7]. Feature extraction technique used is Mel-frequency Cepstral Coefficients (MFCC). Hidden Markov Model (HMM) is also employed for classification. Similar to other techniques, this has also aimed at improving learners techniques of recitation of the Holy Quran [8]-[9].

### 3. Speech Recognition for Distinguishing the type of mud according HAFSS or WARSH

Speech recognition technique involve the process of recording speech or acoustic signal which will be accurately and efficiently convert into a set of words [10]. The steps involve producing a speech recognition system is presented in Figure 1.

The aim of ASR is to extract, characterise, and recognise, the information about speech identification. The system consists of three basic stages as shown in fig 1.: pre-processing, where the recording speech (verses) signals is

passed through the pre-processing block to remove the noise and separate desirable voice from undesirable once and detect the start point and end points of verses.



**Fig. 1.** Block Diagram of speech Recognition System.

Feature extraction is the process of extracting parameters that are unique to each word from the input sample of speech. This can be used to differentiate between a wide set of distinct words. The Mel-Frequency Cepstral Coefficient (MFCC) is considered the most evident example of a feature set [11]. This is widely utilized in speech related studies. However, it is closely related to the logarithmic perceptual ability of the humans. As a way to extract the coefficients, the speech sample is taken as the input. Pre-emphasis is applied to pass the signals through a filter which emphasises higher frequencies. This process will increase the energy of signal at higher frequency. After pre-emphasis, the signals are directed for frame blocking and windowing. Frame blocking is the process of segmenting the speech samples obtained into frames with the length within the range of 20 to 40 msec of N samples, with adjacent frames. Windowing is aimed at minimising the discontinuities of a signal at the beginning and at the end of each frame. This step follows by converting each frame from the time domain into the frequency domain by utilizing DFT. Then to generate the Mel filter bank. This is

done by a set of triangular filters that are used for each frame with actual frequency with Mel frequency as middle frequency. The triangular filter represents the process of Mel scaling in the signal. The next step is the computation of logarithmic of signal energy. The goal of logarithmic signal energy process is to adapt with the system just like human ear. In order to obtain the MFCC, the result of energy logarithmic is processed with Discrete Cosine Transform (DCT). Equation 1 present the approximate empirical relationship to compute the Mel frequencies for a given frequency  $f$  expressed in Hz:

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

Figure 2 shows the steps involved in MFCC feature extraction.

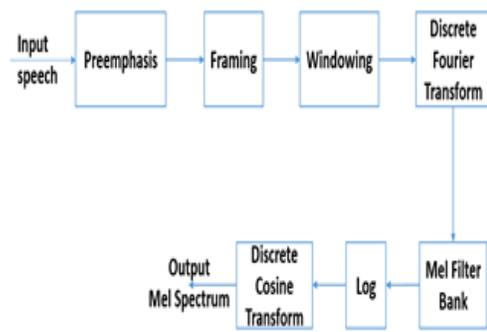


Fig. 2. Block diagram of the computation steps of MFCC.

The features classification or pattern recognition is the process of identifying similarities of spoken words between an extract feature from the input signal and set of acoustic models stored in the database. HMM [12] techniques were used by many researchers for speech recognition.

#### 4. Methodology

The research methodological approach focus on prolongation type recognition and is presented in Figure 3.

The first step involves data collection of the Quranic recitation samples from different experts Qari (Reciter). These experts were known to have Ejazah in Hafss and Warsh.

Each of them recite specific verses that contained the two kind of prolongation (mad): The greater connective prolongation and The Exchange Prolongation rules for Hafss and Warsh. The samples were for collected many times in correct way. Thus, the entire samples are stored as the raw dataset that are prepared for pre-processing.

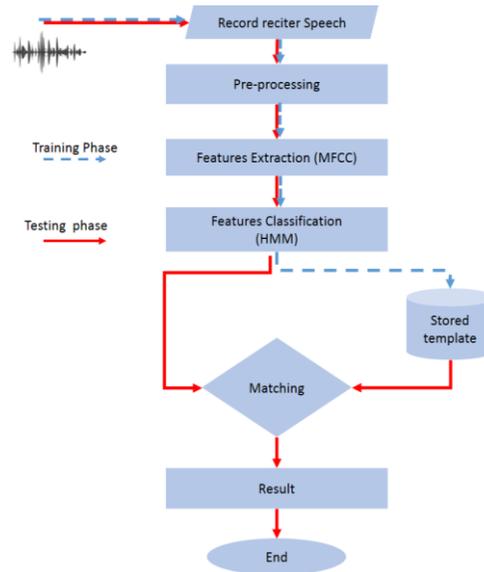


Fig. 3. Prolongation type identification for Qur'an flow chart.

The collected raw data are pre-processing to remove the noise contained in the speech signal and separate the desirable voice from undesirable once. This reduce the group of attributes which assure only the information that wants to be conveyed. The MFCC algorithm is applied to extract and generate features vector which is extensively used as an input for recognition purposes. The classification is done by HMM model. It calculates the HMM parameters. Training phase is characterised by extracting features using large number of samples "training data", and testing phase is characterised by extracting features from testing data "data speech". Testing data (the user recorded) are matched with voice features stored in the database, to

provide responses based on whether they recited correctly or incorrectly. Then the comparison acknowledges the users' level of accuracy. A codebook models (stored template) in the database that is constructed from training data is used for the experimental records.

## 5. Experiment and Results

An experiment was carried out intended to present some of the recognition scenarios. The acoustic model of some Holy Qur'an verses speech signal contained the two kinds of prolongation were used to show the differences between that exist among them. This is based on the different type of recitation of Hafss and Warsh as shown in Figure 4, 5 and 6, and Figure. 7 and Figure. 8. The recognition was carried out based on the guidelines presented in section 4.

This research focus on five words from verses of the holy Qur'an. These verses have been chosen for each of the greater connective prolongation and The Exchange Prolongation as shown in table 2 and table 3, which was recited by the two famous types of Qira'at; namely the Qira'at of Hafss from Asim and the Qira'at of Warsh from Nafi.

**TABLE 2.** Verses selected for the Greater Connective Prolongation.

| VERSES  | Surah                      |
|---|----------------------------|
| "خَيْرَانَ لَهُ أَصْحَابٌ [6:71]"             | <i>Al-An'am</i><br>الإنعام |
| "يَدْعُونَهُ إِلَى الْهُدَى [6:71]"           | <i>Al-An'am</i><br>الإنعام |
| يُشْرِكْ بِعِبَادَةِ رَبِّهِ أَحَدًا [18:110] | <i>Al-Kahf</i><br>الكهف    |
| أَيُّحْسِنُ أَنْ لَمْ يَزِدْهُ أَحَدٌ [90:7]  | <i>Al-Balad</i><br>البلد   |
| فَتَمَّ مِيقَاتُ رَبِّهِ أَرْبَعِينَ [7:142]  | <i>Al-Araf</i><br>الأعراف  |

The experimental test and results of this research is presented in this section through the following:

- Data collection or training phase.
- Recognition phase.

The first phase involves collection of the recitation of the sample data, that will aid in

extracting and training the features. The data are found at the Reciter's database. The database is selected from Internet. Verses of the Holy Qur'an for each of exchange prolongation and the greater connective prolongation are the key objects used. This are tested on the most popular reciters such as Sheikh Al-Hosry. Five (5) verses are recited by three (3) reciters with two (2) categories of Qira'at which are Warsh and Hafss on The Exchange Prolongation as well as The Greater Connective Prolongation respectively. A total of sixty (60) of data samples are obtained. All the samples are passed through the extraction stage in order to extract and represent the features in the form of frequency on Mel scale. Delta coefficients of Mel Coefficients are calculated and then, trained and recognized using HMM. This are used as reference patterns and stored as Reciter's database.

Recognition phase involves verifying the recitation of new reciters to the pre-stored value against the entire reciter in reciter's database. The results are tested against the specified objectives of proposed system. The developed system is tested by performing the MFCC algorithm for features extraction from the Qur'anic recitation of samples data used and then, matching/testing against the trained HMM model of data templates, using the same classification of HMM method. The HMM algorithm is anticipated to get the best results of identification system.

The recognition accuracy rate is calculated using equation 2:

$$\text{Accuracy} = (\text{number of correct samples} / \text{total samples}) \times 100 \quad (2)$$

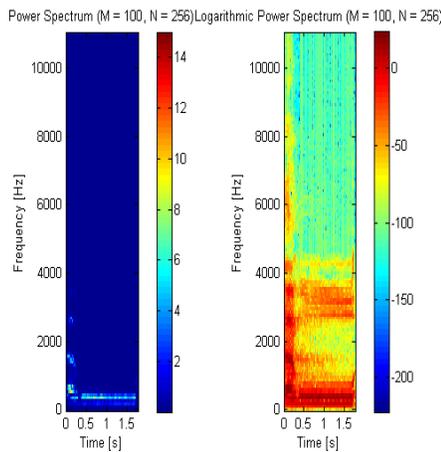
The experimental results of the testing process are presented here. The experiment reveals the extracted features of 10 verses of the Qur'anic recitation which were directly compared with the data based on the Model. As a result, the test result on the training data obtained for this study is at 60% and 50% for The Exchange Prolongation according to Hafss and Warsh and 40%, 70% for the greater connective prolongation according to Hafss and Warsh respectively (see Table 4).

The gathered results have shown some enhancements compared to the previous findings. The research contributions lie with the improve performance and efficiency of the proposed technique.

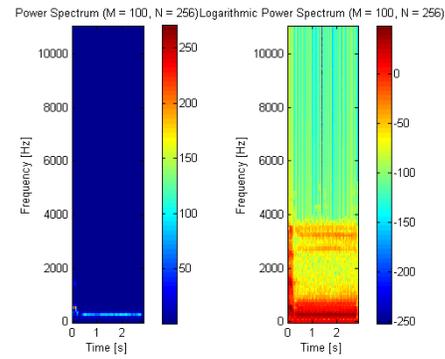
Although, the system faces some drawbacks, with the extra noise due to audio file compression and poor quality during the recording process. Yet, high-performance measure was achieved.

**TABLE 3.** The verses selected for the Exchange Prolongation.

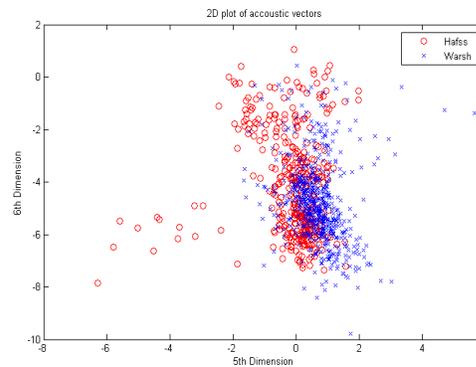
| VERSES  | Surah                       |
|---|-----------------------------|
| “يَا أَيُّهَا الَّذِينَ آمَنُوا اذْكُرُوا اللَّهَ ذِكْرًا كَثِيرًا [33:41]”                   | <i>Al-Ahzab</i><br>الاحزاب  |
| “ وَمَا تَشَاؤُونَ إِلَّا أَنْ يَشَاءَ اللَّهُ رَبُّ الْعَالَمِينَ [81:29]”                   | <i>Al-Takwir</i><br>التكوير |
| “الْمُؤْمِنِينَ لِيَزِدُوا إِيمَانًا مَعَ إِيْمَانِهِمْ” [48:4]                               | <i>Al-Fath</i><br>الفتح     |
| “وَإِنَّ الَّذِينَ أُوْتُوا الْكِتَابَ لَيَعْلَمُونَ أَنَّهُ الْحَقُّ مِنْ رَبِّهِمْ [2:144]” | <i>At-Tawba</i><br>التوبة   |
| “وَجَاؤُوا أَبَاهُمْ عِشَاءً يَبْكُونَ [12:16]”   | <i>Yusuf</i><br>يوسف        |



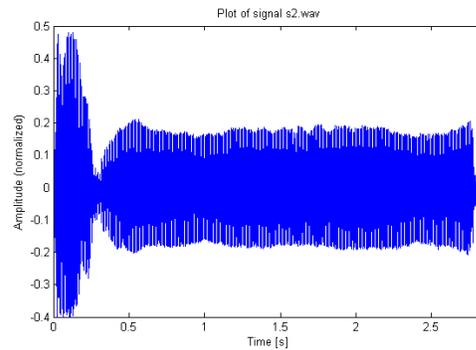
**Fig. 4.** Power spectrum plot of spoken word له خَيْرَانِ لَهُ أَصْحَابٌ the Greater Connective Prolongation According to Warsh.



**Fig. 5.** Power spectrum plot of spoken word له خَيْرَانِ لَهُ أَصْحَابٌ The Greater Connective Prolongation According to Hafss.



**Fig. 6.** 2D Plot of acoustic vector of spoken word له خَيْرَانِ لَهُ أَصْحَابٌ The Greater Connective Prolongation According to Warsh and Hafss.



**Fig. 7.** Speech signals of spoken word له خَيْرَانِ لَهُ أَصْحَابٌ The Greater Connective Prolongation According to Warsh.

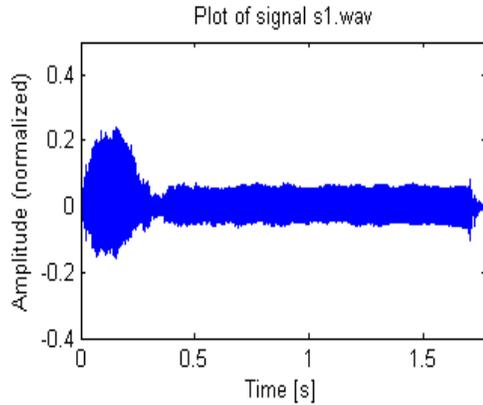


Fig. 8. Speech signals of spoken word خَيْرَانَ لَهُ، له "أصْحَابُ" The Greater Connective Prolongation According to Hafss.

TABLE 4. Model tuning results.

| Prolongation type      | The Exchange Prolongation |       | The Greater Connective Prolongation |       |
|------------------------|---------------------------|-------|-------------------------------------|-------|
|                        | Warsh                     | Hafss | Warsh                               | Hafss |
| <i>Qira'at type</i>    | Warsh                     | Hafss | Warsh                               | Hafss |
| <i># of utterances</i> | 10                        | 10    | 10                                  | 10    |
| <i>Correct</i>         | 06                        | 05    | 04                                  | 07    |
| <i>Wrong</i>           | 04                        | 05    | 06                                  | 03    |
| <i>% Accuracy</i>      | 60%                       | 50%   | 40%                                 | 70%   |

Figure 9 shows the recognition accuracy rate of each kind of prolongation type where y-axis contains results and x-axis contains the types of Madd.

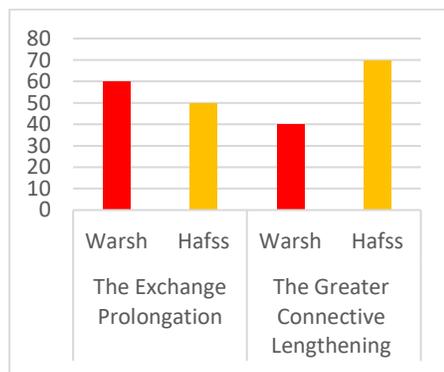


Fig. 9. The accuracy rate of proposed system.

## 6. Conclusion

This paper has developed on theory and practice based for developing a high-performance Tajweed system that assist in proper Tajweed Qur'anic recitation based on the automatic speech recognition system. The research utilized, Mel-frequency Cepstral Coefficients (MFCC) and HMM (Hidden Markov Model) algorithms to enable the validation of the proposed system. Several experiments were carried out. The experimental results on a database indicate that the feature extraction method and recognition method used for this research appropriate for Arabic recognition system are feasible.

There are other several techniques such as Liner Predictive Coding (LPC) and Artificial Neural Network (ANN) could also be used for similar research approach. The findings from those might be different, therefore, this research recommend future work to focus on using discriminative training techniques which might improve the discrimination between some confusable pronunciation alternatives.

## ACKNOWLEDGMENT

The authors would like to thank the Research Management Centre and the Faculty of Information and Communication Technology, the International Islamic University Malaysia for their supports.

## REFERENCES

- [1] K. C. Czerepinski and A. D. A. R. Swayd, Tajweed Rules of the Qur'an. *Dar Al-Khair Islamic Books Publisher*, 2006.
- [2] B. Yousfi and A. M. Zeki, "Automatic Speech Recognition for the Holy Qur'an, A Review," in *The International Conference on Data Mining, Multimedia, Image Processing and their Applications (ICDMMIPA2016)*, 2016, p. 23.
- [3] H. A. Hassan, N. H. Nasrudin, M. N. M. Khalid, A. Zabidi, and A. I. Yassin, "Pattern classification in recognizing Qalqalah Kubra pronunciation using

- multilayer perceptrons,” *IEEE symposium on Computer Applications and Industrial Electronics (ISCAIE)*, 2012, pp. 209–212.
- [4] D. Raja-Jamilah Raja-Yusof , Fadila Grine, N. Jamaliah Ibrahim, M. Yamani Idna Idris, Z. Razak, and N. Naemah Abdul Rahman, “Automated tajweed checking rules engine for Quranic learning,” *Multicult. Educ. Technol. J.*, vol. 7, no. 4, pp. 275–287, 2013.
- [5] A. N. Wahidah et al., “Makhraj recognition using speech processing,” *7th International Conference on Computing and Convergence Technology (ICCCT)*, 2012, pp. 689–693.
- [6] B. Yousfi and A. M. Zeki, “Holy Qur’an speech recognition system distinguishing the type of recitation,” *7th International Conference on Computer Science and Information Technology (CSIT)*, 2016, pp. 1–6.
- [7] B. Yousfi, A. M. Zeki, and A. Haji, “Holy Qur’an Speech Recognition System Mudud Tajweed Rule Checking,” *Int. J. Islam. Appl. Comput. Sci. Technol*, pp. 10–18, 2016.
- [8] B. Yousfi and A. M. Zeki, “Holy Qur’an speech recognition system Imaalah checking rule for warsh recitation,” *IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, 2017, pp. 258–263.
- [9] B. Yousfi, A. M. Zeki, and A. Haji, “Isolated Iqlab checking rules based on speech recognition system,” *8th International Conference on Information Technology (ICIT)*, 2017, pp. 619–624.
- [10] N. Zerari, B. Yousfi, and S. Abdelhamid, “Automatic Speech Recognition: A Review,” *Int. Acad. Res. J. Bus. Technol.*, vol. 2, no. 2, pp. 63–68, 2016.
- [11] S. B. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” in *Readings in speech recognition*, Elsevier, 1990, pp. 65–74.
- [12] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.