ISSN 2278-3091
Volume 8, No.1.3, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse4181.32019.pdf
https://doi.org/10.30534/ijatcse/2019/4181.32019

A Model of Correction Mapping for Al-Quran Recitation Performance Evaluation Engine

Noraimi Shafie[1], Mohamad Zulkefli Adam[2], Salwani Mohd Daud[3] , Hafiza Abas[4]
[1,2,3,4] Advanced Informatics School, Universiti Teknologi Malaysia, noraimi.kl@utm.my

ABSTRACT

The main objective of the research is to develop the correction-mapping model for Al-Quran recitation performance evaluation engine. Machine learning and Digital Signal Processing techniques are applied in representing and analyzing the recitation speech signal. Consequently, a form of recitation correction results is derived and formulated for the final performance evaluation. The proposed corrective mapping model demonstrated in this paper confronted, but not limited to, with the challenging issues of variability of speaker recitation, recitation representation, speaker adaptation, feature extraction, parameters estimation and threshold process classification. The experimental results concluded the Al-Quran automated correction system knowns as Intelligent Quran Recitation Assistant (nur-IQRA) will be able to fulfil the current and future trends of digital society.

Key words : Digital Speech Processing; Machine Learning, Speaker Adaptation, Feature extraction, Parameter estimation, Correction Mapping Model.

1. INTRODUCTION

This A Muslim who practices Islam, must recite Al-Quran correctly, where the underlying Tajweed rules must be adequately followed. Indeed, according to Islam lesson, a Muslim who wrongly reciting the Al-Quran because of improper learning the Tajweed rules and without repenting from it, will bring a sin and punished by Allah (god). Technically, without the correct knowledge of proper recitation technique and rules will result the different pronunciation of Arabic words and leads to the different meaning for any particular verse in Al-Quran. Generally, the evaluation of Al-Quran recitation is based on *Makraj, Tajweed* rules and any of its derivatives [1]. It is an art and reciters who follow the Tajweed rules to build the recitation correctly will find satisfaction of doing the right way [2]. In this regard, having a systematic recitation practice is very much useful in order to guide the re-citer to recite Al-Quran correctly with respect to Makraj and Tajweed rules.

The process of recitation with Makraj and Tajweed rules and guidance involves specific pre-defined unique techniques that based on Arabic phonetic and languages. There are number of proposals of automated correction systems introduced in the literature that reveal the importance of intelligent *Tajweed* correction system[3]. From the perspective of human-machine interactive application, the design of the *Tajweed* correction engine should be able to support the recitation process without the presence of human experts. For instance, the speech recognition approach is used in checking the Tajweed rules to be followed accordingly [4]. Moreover, the discussion on the apps-based Tajweed correction have been given the special attention to help student fruitfully learn and revise Al-Quran recitation [5]-[6]. Technically, the discussion of intelligent *Tajweed* correction system encompasses a number of areas and factors such as data mining, interfacing, real time processing and also the media supports to name a few. Furthermore, the integration of computer and mobile application should be interestingly made available. Data mining will involve huge data based on audio, which contain enough information to reveal the performance of recitation.

2. QURANIC RECITATION: CHALLENGES TO DIGITAL SPEECH PROCESSING (DSP)

The guidelines of Makraj and Tajweed allude to elocution, diacritics, and articulation for every ayah (sentence) along the recitation procedure [24]. These are implanted inalienably in the Quranic discourse signal, created by any reciter. Ongoing preparing, for example, Digital Speech Processing ought to have the option to speak to, remove and order the uniqueness of recitation execution dependent on the discourse signal. Difficulties in managing the intricacy of discourse signal, particularly in choosing the reasonable methods of DSP to uncover the portrayal of recitation execution are in all respects very kind in this paper.

The discourse intricacy, for example, acoustic changeability, talking inconstancy, speaker fluctuation, language changeability, phonetic fluctuation, and the Lombard impact should influence the logical exhibition. The discovery of the error of recounting will be completed by contrasting the comparability properties of the discourse signal between the student and the master presenting signal. Then again, the stochastic of recitation discourse is a moving issue for the one to show or assess the recitation blunder, particularly in structuring the model or format for the miniatures highlight arrangement. Dealing with the issues of assessment in time arrangement for every recitation is likewise recommended in this exploration and is sensibly displayed to show the recitation blunder, incorrect spelling mistake or including/erasing word blunder. The procedure stream from chronicle the recitation to the effective, savvy amendment framework execution showed in this paper, for the most part, include the errands of speaking to, extricating the highlights (and small highlights), distinguishing, characterizing and perceiving forms that uncover the mistake of Al-Quran recitation with spotlight on Makraj and Tajweed rules.

3. EXPERIMENT DESIGN AND PROCEDURE

The core of experiment design is to deliberately present the training and the testing phases of the experiments. These involve the required stages of obtaining the optimal threshold range of corrective recitation signal that based on recitation weightage, defined by Tajweed expert (human) judgement. The threshold value or condition will then be used to perform the classification task and produce the recognition results. In confronting the variety of recitation signals, Natural Language Processing (NLP) is used solely of its distinguished characteristics in dealing with the process of applying suitable extraction techniques for automatically extracting the information from the speech (or text) [7]. As a result, the extracted informative value will enhance the design of an intelligent automated correction system and robustly assist the learner in this expert-template-based system. The process machine learning of recitation recognition demonstrated in this paper is divided into five (5) stages; Recitation Representation and Compensation, Speaker Adaptation, Miniatures Features Extraction, Parameters Estimation and Time Series Classification. Each of these will be explained throughout the paper, with the outcomes of experiments respectively.

3.1. Recitation Representation and Compensation

Quranic speech signals are complex by virtue of inherent information of Tajweed rules to be considered. As a matter of fact, the scenario is aggravated by the presence of various noises or simply signal variability as happen to be in any ordinary speech signals. Furthermore, the loss of information in a Quranic speech signal is also explicitly caused by the reciter behavioral variations. The complexity of variability will be compensated to gain the clean speech signals from each *ayah* (sentence/phrase) of a Quranic speech signal without losing the important features.

The process is firstly initiated by the pre-processing task with the variable compensation. Pre-processing is performed to avoid unnecessary waveform of signals such as noise, unwanted sound and unnecessary recorded signals. This includes silence removal and endpoint detection, pre-emphasis/noise filtering and channel normalization [9],[10],[11]. This is followed by the selection of techniques that should be able to represent the Quranic speech raw signals, which was transformed from the speech production signals.

In the silent trimming technique, the selected threshold is used to remove the silent. Therefore, the amplitude normalization is used to normalize the reciter's speech signals. While the endpoint detection is used to define the start point and endpoint of Quranic speech signals. Each of learners or experts have a different start point and endpoint while performing the recitation. Combined zero crossing and short term energy function are used to determine start point and end point [8].

3.2. Speaker Adaptation

Speech is a variable process and stochastic in which the duration of a word and its sub-words vary randomly. Time alignment or normalization can be applied in recitation, which is required to find the best alignment between the sequences of experts and learners recitation vector features. Both could be compared correspondingly at the time of occurrence, without having especially to concern on the reciter characteristics such as gender, age and health condition. In confronting the variability and complexity of the continuous speech recitation signals, the proposed approach introduced in this paper was able to manipulate the availability of signal properties of such recitation in order to measure the correct performance with respect to the expert recitation as a reference.

The Dynamic Time Warping (DTW) is used in this research as to measure the similarity between the two temporal sequences, which may vary in time. The technique is also able to find the optimal alignment between the two-time series if one-time series may be "warped" non-linear by stretching or shrinking it along its time axis [12]. Furthermore, the technique was used to re-map the recitation, such that all utterances considerably generated from same vocal tract [25]. DTW is well known technique in speech recognition to cope with different speaking speeds [13].

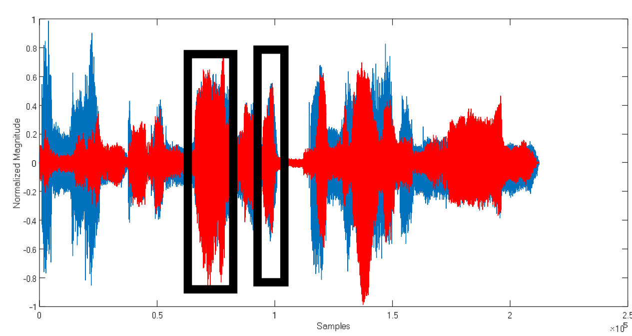


Figure 1: Same temporal utterance for expert and learner

As mentioned previously, the reciter variability is one of the difficulties in need of consideration to be tackled. For instance, the way the reciter expresses the speech, the start and endpoints are among the sources of variability. Figure 1 shows the representation of Quranic speech signals from both the expert and the learner. This correspondence has demonstrated the sequences of both have been synchronized and led to the utterances of each could be made the comparison within the same time framed. The temporal utterance in the Fig.1 is highlighted with the rectangular line of black in color. The feature vector extraction will then could be proceeded to the extent of obtaining the miniature features.

3.3 Miniatures features extraction

Most of modern speech recognition or pronunciation evaluation is used cepstral coefficient as acoustic features. The acoustic features can derive from the speech parameterization based on formant-like features. Formant-like features are used as acoustic and language model for Al-Quran recitation evaluation performance. The characteristic shape of the power spectrum can be aligned as acoustic and language model of Al-Quran recitation which representing the energy, rhythm and tone. While certain experiment suggested the frame size of between 20 ms and 40 ms [14], in this paper, the chosen shape is framed between 25 ms each based on phoneme formant representation, as depicted in Figure 2.

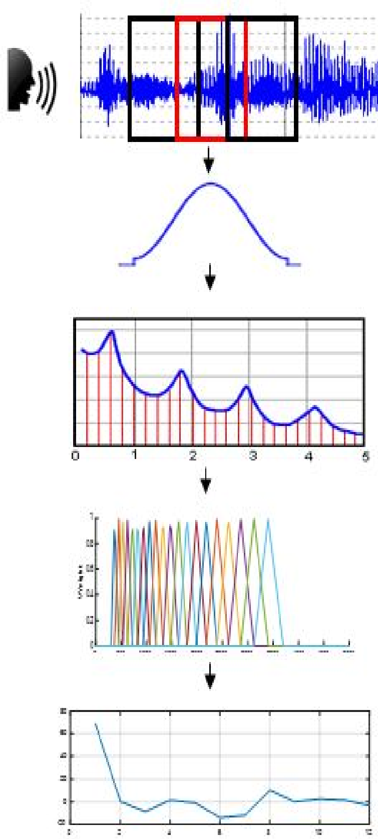


Figure 2 : MFCC features extraction

Fundamentally, include extraction utilized in this model improvement alluded to the way toward packing the sign without losing any significant data. There are a considerable lot of discourse handling systems, for example, Mel-Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPC) and Perceptual Linear Prediction (PLP)[3]. In any case, the most encouraging outcome for highlight extraction is the employments of MFCC [3,4,15,16,17]. The MFCC technique processes the largest vector of feature extraction that having the same values are grouped together as miniatures feature vectors [24]. MFCC feature vector can represent the transformation from an articulated system with speech acoustic features.

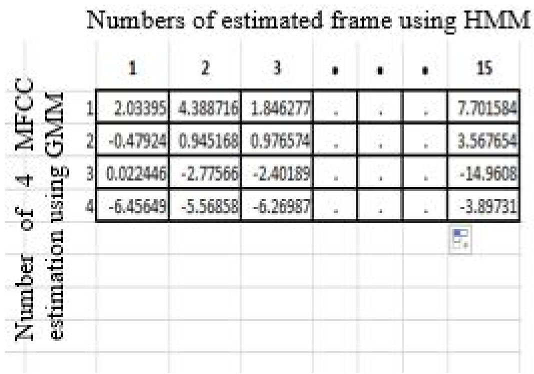
3.4. Parameters Estimation

The acoustic and language model or template will be used as machine learning unsupervised learning algorithm to represent the recitation feature vectors. The corrective mapping model and the practice of recitation are evaluated based on the level of threshold membership function of recitation of Al-Quran. The format-like features represent the Makraj and Tajweed as energy and prosody by using the Mel Frequency cepstral coefficient at the same frame. This will be modelled by Gaussian Mixture model (GMM) and Hidden Markov Model (HMM) as a static pattern and sequential pattern, respectively. Both of GMM-HMM parameter estimation are iteratively estimated by using expectation maximization (EM) algorithm. For GMM, the posterior probabilities of each bin being generated by the target mixtures are estimated, and then these values are used to calculate the Gaussian component parameters that represent the spectral-peak of formant frequencies. Whilst, for HMM, the forward-backward algorithm known as a Baum Welch algorithm was used to estimate the parameter that represents the sequences of utterance/word recitation. The acoustic and language model for expert recitation can be used as a template reference and can be compared with the acoustic model of learner to define the incorrectness of Al-Quran recitation [24]-[25]. Consequently, the representation of the acoustic model based on formant like-features can be used as a task of corrective measures that governed by these recitation speech representation challenges. Furthermore, the variability and continuous speech properties are embedded challenges in representing the specialty of Quranic speech recitation. The MFCC miniatures feature vector parameter estimation are illustrated in Figure 3.

Numbers of original segmented frame

Number of 12 MFCC	1	2	3	•	•	•	319
1	78.76649	83.6325	85.23154	.	.	.	75.68322
2	0.072707	1.876519	5.259163	.	.	.	-2.06523
3	-8.67339	-1.7077	-3.15294	.	.	.	-5.19583
4	3.17986	1.153416	4.01048	.	.	.	-2.14117
5	-4.06872	-5.66352	-0.15143	.	.	.	-0.76346
6	-10.496	-20.9392	-19.94	.	.	.	-13.3423
7	-1.98903	-7.59172	-12.7191	.	.	.	3.485182
8	-3.24579	-1.97554	-4.42254	.	.	.	13.87802
9	-4.83333	-5.63811	-6.08062	.	.	.	-34.3139
10	-0.77159	7.926194	5.797878	.	.	.	8.65769
11	2.970107	2.731584	3.751652	.	.	.	0.552538
12	5.996709	6.742986	10.40143	.	.	.	1.812557

(a)



Numbers of estimated frame using HMM

Number of 4 MFCC estimation using GMM

	1	2	3	.	.	.	15
1	2.03395	4.388716	1.846277	.	.	.	7.701584
2	-0.47924	0.945168	0.976574	.	.	.	3.567654
3	0.022446	-2.77566	-2.40189	.	.	.	-14.9608
4	-6.45649	-5.56858	-6.26987	.	.	.	-3.89731

(b)

Figure 3 : (a) Original 12-MFCC features vectors (b) 4-MFCC with 15 frame parameters estimation using GMM-HMM

3.5. Time Series Classification- Threshold Process

The nature of Fuzzy Logic is to provide a quantitative score [18], in which the measure of evaluation is based on satisfaction of all conditions. Practically, Fuzzy classification is the process of grouping elements into a fuzzy set in which each of the elements can be grouped based on membership function. Membership function is defined by the truth-value of a fuzzy propositional function [19]. The corresponding membership function can be derived from the Fuzzy set as to provide the means for capturing non-binary concepts and can be represented by various shapes for membership function such as Gaussian bells, trapezoids, triangles and etc. [20]. The classical fuzzy set (crisp set) is used to define the high, mid-high, mid, mid-low, and low vowels that indicate the acceptable values of each word or utterance for the Al-Quran recitation from the GMM-HMM parameter estimation.

Fuzzy correction mapping model in time series consists two main steps, which are structured identification and threshold parameter optimization [21]. For the structure identification, the number of membership functions and fuzzy if-then is applies to MFCC estimated parameters. From the fuzzy set, the similarity for each frame is measured by the distance-based similarity measure [22]. The calculation of similarity of fuzzy sets between expert template and learner is based on distance for each frame and the relationship between the similarity and distance can be represented as the degree of similarity for each frame. The Euclidean distance and distance-based assessment proposed by [22] are used to determine the degree of similarity.

The equation of The Euclidean Distance is given by

$$d_E(A,B) = \sqrt{\sum_{i=1}^{n}(A(x_i) - B(x_i))^2}. \quad (1)$$

The distance based assessment equation is given by

$$S(A,B) = \frac{1}{1 + d(A,B)}. \quad (2)$$

The performance of recitation evaluation accuracy is rated by word error rate (WER), The WER is derived from the Levenshtein Distance [23], working at the utterance/word level instead of the phoneme level. The utterance/word sequence of between expert and learner can be utilized and can be compared after the length of both recitations already aligned in the same frame. Word error rate can then be computed as:

$$\text{WER} = \frac{S+D+I}{N} \quad (3)$$

Where,
S is the number of substitutions,
D is the number of the deletions,
I is the number of the insertions,
N is the number of words in the reference.

4. CONCLUSION

Most of scientific approaches on digital apps of Al-Quran recitation system are still in the regular transition of traditional to the electronic version, especially to scaffold the *Talaqqi* and *Mushafahah* techniques of reciting. There is no intelligent *Tajweed* correction on recitation is applied in the real time applications. Most of the applications of digital Al-Quran only provide the recorded audio by the expert and learners may download the digital version of Al-Quran materials. The hybrid approach of DSP and Machine Learning (ML) revealed in this research paper, demonstrated a proper method on how to find the threshold range in determines the degree of recitation similarity (or error). The use of dynamic programming in that approach was to estimate the parameter and warp the recitation speech between expert and learner. As a result, a new dimension of representation (micro values) is usefully obtained to represent miniatures feature vector as recitation similarity (or error). Nevertheless, Muslims of all over the world, remain engaged with numerous daily activities and limited time for reciting the Al-Quran with the appropriate *Tajweed* teacher. It is hoped, with the invented automated correction mapping model as Intelligent Quran Recitation Assistant (nur-IQRA), demonstrated in this research paper, opens to a wide range of future interactive software development and availability for the Muslim to be able to properly recite Al-Quran and improve the quality of recitation, which is not limited to the time and place. The traditional methods are less excited and need specific time to perform the recitation in front of time-scheduled expert. With the "correction engine" demonstrated in this research, it is hoped that learners are able to recite Al-Quran frequently and perform self-assessment and practice without the presence of an expert. This tool can be used as a pre - evaluation before performing the actual correction with the expert.

.

ACKNOWLEDGEMENT

This work is financially supported by the AROMA Research Center and Universiti Teknologi Malaysia.

REFERENCES

1. 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."** International Symp. Compututer Applied Ind. Electronic. 2012. 209–212. https://doi.org/10.1109/ISCAIE.2012.6482098
2. H. Tabbal, W. El Falou, and B. Monla. **"Analysis and implementation of a 'Quranic' verses delimitation system in audio files using speech recognition techniques."** 2nd International Conference Informatic Communication Technology. 2006. 0–5.
3. A. H. Ahmed and P. Mohammed. **"Verification System for Quran Recitation Recordings."** International Journal Computer Application. 2017. 6–11. https://doi.org/10.5120/ijca2017913493
4. I. Ahsiah, N. M. Noor, and M. Y. I. Idris**. "Tajweed checking system to support recitation."** International Conference Advance Computer Science Informatic System. 2013.189–193. https://doi.org/10.1109/ICACSIS.2013.6761574
5. E. A. Z. E. Alwi, N. Anas, M. S. Ibrahim, A. F. M. Dahan, and Z. Yaacob. **"Digital Quran Applications on Smart Phones and Tablets: A Study of the Foundation Programme Students,"** Asian Social Science. 2014. 212–216.
6. N. Mustafa and M. Basri**."A preliminary Studyon Mobile Quranic Memorization for Remote Education Learning using RFID Technology:** KUIS as a Study." worldconferences.net. 2014. 1–6.
7. M. Zakariah, M. K. Khan, O. Tayan, and K. Salah. **"Digital Quran Computing: Review, Classification, and Trend Analysis".** Arabian Journal of Science and Engineering. 2017 3077–3102. https://doi.org/10.1007/s13369-017-2415-4
8. E. S. Gopi, **Digital Speech Processing Using Matlab**. 2014. https://doi.org/10.1007/978-81-322-1677-3
9. D. Abuzeina, W. Al-Khatib, M. Elshafei, and H. Al-Muhtaseb. **"Cross-word Arabic pronunciation variation modelling for speech recognition."** International of journal Speech Technology. 2011. 227–236. https://doi.org/10.1007/s10772-011-9098-0
10. E. Mourtaga, A. Sharieh, and M. Abdallah. **"Speaker Independent Quranic Recognizer Based on Maximum Likelihood Linear Regression."** World Academy of Science, Engineering and Technology. 2007. 61–67.
11. B. Abro, A. B. Naqvi, and A. Hussain. **"Qur'an recognition for the purpose of memorisation using Speech Recognition technique."** 15th International Multitopic Conference. 2012. 30–34. https://doi.org/10.1109/INMIC.2012.6511440
12. B. J. Mohan and R. Babu. **"Speech Recognition using MFCC and DTW"**. International Conference on Advances in Electrical Engineering (ICAEE). 2014. 1–4.
13. S. K. Gaikwad, B. W. Gawali, and P. Yannawar. **"A Review on Speech Recognition Technique."** International Journal of Computer Applications. 2010. 16–24. https://doi.org/10.5120/1462-1976
14. M. Bezoui, A. Elmoutaouakkil, and A. Beni-Hssane**."Feature extraction of some Quranic recitation using Mel-Frequency Cepstral Coefficients (MFCC)."** International Conference in Multimedia Computer System. 2017. 127–131. https://doi.org/10.1109/ICMCS.2016.7905619
15. N. S. Dey, R. Mohanty, and K. L. Chugh. **"Speech and Speaker Recognition System Using Artificial Neural Networks and Hidden Markov Model."** International Conference Communication System Network Technology.2012. 311–315.
16. M. Alsulaiman, G. Muhammad, and Z. Ali. "**Comparison of voice features for Arabic speech recognition."** 6th International Conference Digital Information Management. 2011.90–95. https://doi.org/10.1109/ICDIM.2011.6093369
17. A. Muhammad, Z. Ul Qayyum, M. Waqar Mirza, S. Tanveer, A. M. Martinez-Enriquez, and A. Z. Syed. **"E-Hafiz: Intelligent system to help Muslims in recitation and memorization of the Quran."** Life Science Journal, vol. 9, no. 1. 2012. 534–541.
18. P. F. Su, Q. C. Chen, and X. L. Wang. **"A fuzzy pronunciation evaluation model for English learning." International Conference on Machine Learning and Cybernetics.** 2006. 2598–2604.
19. H. Zimmermann, **Practical Applications of Fuzzy**. 2000. https://doi.org/10.1007/978-1-4615-4601-6
20. S. Raptis and G. Carayannis. **"Fuzzy Logic for Rule-based Formant Speech Synthesis**" 1997.
21. S. Silarbi, B. Abderrahmane, and A. Benyettou. **"Adaptive Network Based Fuzzy Inference System for Speech Recognition Through Subtractive**." vol. 5, no. 6. 2014.43–52. https://doi.org/10.5121/ijaia.2014.5604
22. I. Beg and S. Ashraf. **"Similarity measures for fuzzy sets."** Applied and Computer Math Journal. 2009. 192–202.
23. P. G. Scholar. **"Feature Selection Algorithms for Automatic Speech Recognition."** International Conference Computer Communication and Informatics. 2014.

24. N. Shafie, M. Z. Adam, and H. Abas**. “The model of Al-Quran recitation evaluation to support in Da’wah Technology media for self-learning of recitation using mobile apps.”** 3rd International Seminar on Da’wah. 2017. 1–10.
25. N. Shafie, M. Z. Adam, and H. Abas. **“Al-Quran Recitation Speech Signals Time Series Segmentation for Speaker Adaptation using Dynamic Time Warping**.” 7th International Conference of Information Technology and Multimedia.2017. 1-10..