Analysis of two adjacent articulation Quranic letters based on MFCC and DTW

Tareq AlTalmas
Department of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
Tareq.ma@live.iium.edu.my

Wahju Sediono*
Department of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
wsediono@iium.edu.my

Nik Nur Wahidah Nik Hashim
Department of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
niknurwahidah@iium.edu.my

Salmiah Ahmad
Department of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
sbintiahmad@kau.edu.sa

Safiah Khairuddin
Department of Mechatronics Engineering
International Islamic University Malaysia
Kuala Lumpur, Malaysia
safiahkd@gmail.com

Abstract—Reciting al-Quran in the correct way is an obligatory duty for Muslims, and therefore learning al-Quran is a continuous education until the correct recitation is achieved. It is important to learn Tajweed rules to master the recitation of Quranic verses. Moreover, mastering the pronunciation of Arabic sounds is the first and key step to achieve accurate recitation of al-Quran. The rules were guided by the Islamic Scholars in fields related to al-Quran from their knowledge and experiences. Very limited researches were found in the perspective of sciences and engineering. In this paper two Quranic letters (ذ and ز) that are articulated from adjacent points of articulation were analyzed using Mel- frequency coefficient analysis. MFCCs matrices were calculated then compared using the dynamic time warping DTW technique to calculate the similarity matrices and find the similarity distance. Results show that letters from the same point of articulation have less similarity distance compared to the letters from different point of articulation.

Keywords—Feature extraction, speech processing, MFCC, DTW, Tajweed rules;

I. Introduction

The pronunciation of Quranic letters in the correct way is very important to recite al-Quran. In fact, that was among the introductory topic learnt before higher level ones. Every Quranic letter has a way to be pronounced from its right points of articulation (Makharij) in the vocal tract. People who are not from the Arab background face difficulties in pronouncing Quranic letters correctly specially if the letters are emitted from adjacent points of articulation (Makharij)[1]. Articulation process of sounds includes two types of articulators; active articulators (moving parts in the vocal tract like tongue), and passive articulators (static parts of the vocal tract like the teeth)[2]. The letters ز and ذ are both articulated from the tip of the tongue but with different passive articulator. ز is a result of touching between tip of the tongue with the base of the frontal teeth, while ذ with tip of the frontal teeth, [3]. Due to this fact (closed points of articulation) most of the people including some Arabs are pronouncing both as ز, which may lead to change of the meaning of Quranic words and this is seriously prohibited.

Therefore there is an urgent needs in the study of pronouncing both letters correctly from engineering and technology perspective, such that all other similar pronunciation cases could be solved appropriately

Digital processing of speech signals is widely used in daily applications around us. Speech aid and training applications one of these potential applications that utilize digital signal processing. Various techniques are being used to analyse the speech data to extract distinguishing features that can help in detecting the wrong pronunciations, such as, Linear prediction coding (LPC), linear prediction cepstral analysis (LPCC), perceptual linear prediction cepstral analysis (PLPC), and Mel-frequency cepstral analysis (MFCC)[4]. MFCC has shown good efficiency in the system of isolated word speech recognition systems[5]. The step of speech analysis to extract feature vector of the speech is followed by the machine learning step to decide whether the pronunciation is correct or wrong. MFCC was implemented as a feature extraction for Makharij recognition with support vector machine as a classifier [6]. A speaker dependent system to recognize isolated Arabic words has developed. Many feature extraction and classification methods were tested, MFCC with DTW have shown the highest recognition rate compared to other techniques. 90% recognition rate has been achieved [7].

This paper has studied two letters that have common active articulator, and different passive articulator, which are ز and ذ. The Mel-frequency cepstral coefficients MFCCs of these letters were extracted from two expert samples in reciting al-Quran. The MFCCs matrices were then compared to extract the similarities and distances between it using DTW.

II. Method

A. Recording the pronunciation of the quranic letters

The Quranic letters' pronunciation were recorded from the certified and expert reciters of different backgrounds to ensure the accuracy of the data. 28 letters with sukun pronunciation were recorded from all experts. The approved Tajweed method to recite the letters from its articulations (Makharij) was used to record the letters. This method is implemented by starting the pronunciation by hamzah and fathah diacritic followed by the letter itself using sukun

*Corresponding author: Email: wsediono@iium.edu.my

978-1-5386-6992-1/18/$31.00

diacritic as shown in Table1, [8]. The portable high-quality field recorder (TASCAM DR-05) was used to collect the audio data in the normal learning environment[9].

TABLE I. THE RIGHT WAY TO PRONOUNCE QURANIC LETTERS

أءْ	أبْ	أتْ	أثْ	أجْ	أحْ	أخْ
أدْ	أذْ	أرْ	أزْ	أسْ	أشْ	أصْ
أضْ	أطْ	أظْ	أعْ	أغْ	أفْ	أقْ
أكْ	ألْ	أمْ	أنْ	أهْ	أوْ	أيْ

The data were recorded using high sampling frequency 99.6kHz and then down-sampled to 8khz. The amplitude of the data were normalized to be in between 1 and -1. For simplicity, the stereo-channel data were converted to mono-channel as well.

B. Feature Extraction (The Mel-Frequency Cepstral Coefficients)

Speech signal analysis has various techniques to translate the time domain representation of the signal into less size and parametric representation. These techniques have shown varying success rate. The step of analysing the signal and extract its parameters is called feature extraction. Selecting the parameters to represent the acoustic signal play a key role in the overall system performance. The Mel-Frequency Cepstral Coefficients (MFCCs) is one of the common used technique in representing the envelope of the short time power spectrum. This envelope represents the shape of the vocal tract. The process of calculating the MFCCs is shown in Fig. 1.

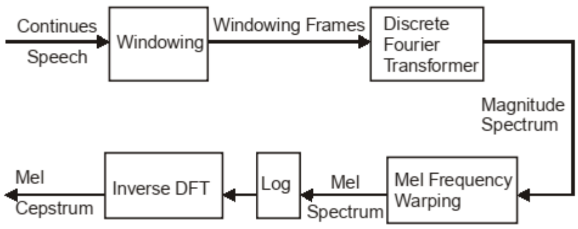

Fig. 1. The Mel-frequency cepstral coefficienrs block diagram[10].

During the speech production some parts of the high frequency components are being suppressed. Pre-emphasizing speech signals is important to amplify the high frequencies, and to improve the signal-to-noise ratio (SNR). A high-pass FIR filter is commonly used as a pre-emphasize filter in the speech processing. Equation (1) shows the pre-emphasize filter and the coefficient value is mostly in between 0.95 to 0.97.

$$H(z) = 1 - aZ^{-1} \tag{1}$$

Speech signals are a result of the continuous changing of the vocal tract structure over the time. Framing the speech signal helps in reducing the non-stationary natural of the signals. Speech signals can be considered a quasi-stationary over a short time period (20 to 40ms). Based on that nature of the speech signal, therefore the speech signal must be divided into overlapped short frames to be analyzed while it holds the quasi-stationary property. The selection of the frame width and the overlapping amount between the subsequent frames is a trade-off process to preserve the speech properties and reduce the analysis complexity. Each frame is multiplied by a window function in order to reduce the discontinuities of the speech signal at the edges of each frame. Hamming window is one of the common used windows in the speech processing[11][12].

Some dynamic features can be extracted from the MFCCs to enhance the system performance. These features are the temporal derivatives to the MFCCs vectors; delta coefficients and acceleration coefficients[13].

C. Feature vectors comparision

Every audio file gave a matrix of MFFCs, these matrices must be compared to calculate the similarities. DTW is a well-known method to compare sequences and extract the similarities between parameters. It has been used widely in the automatic speech recognition field to compare the feature vectors and calculate the similarities between it, [5]. It is therefore used in this paper to compare and calculate the similarity matrix between the matrices. Distances between matrices were calculated from the similarity matrices. DTW was then used to find the optimal alignment between two feature vectors by expanding or shrinking these vector on the time axis.

III. EXPERIMENT AND RESULTS

This paper investigates 2 Quranic letters acoustic features. These two letters are ز and ذ. These letters are articulated from point of articulations (Makharij) that are closed to each other’s. The ز is articulated by the touching that happens between the tip of the tongue with the base of the front two teeth, while the ذ is articulated by the tip of the tongue touching the tip of the frontal two teeth as it is shown in Fig. 1 and Fig. 2 respectively. The closed point of articulation of these two letters lead to mispronunciation of it, especially of people whose mother tongue not the Arabic language. The recordings from two experts was tested in this paper.

Fig. 2. Articulation point (Makhraj) of the letter ز

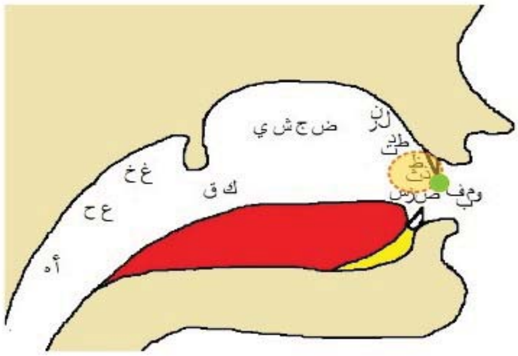

Fig. 3. Articulation point (Makhraj) of the letter ذ

The MFCCs of the audio files were extracted from the pre-processed data. Then the Delta coefficients (the time derivative of the MFCC matrix) were calculated from the MFCCs. Fig. 4 and Fig. 6 show the time domain graph of the letter ذ and ز from the first expert (Sample 1). Fig. 5 and Fig. 7 show the time domain graph of the letter ذ and ز from the second expert (Sample 2). It is obvious that the time domain representation of the speech signal does not convey enough information about the speech and its acoustic characteristics.

Matlab environment was used to extract the MFCCs. MFCCs matrix for every audio file was calculated and the Delta coefficients matrix was extracted. The analysis parameters are shown in Table II.

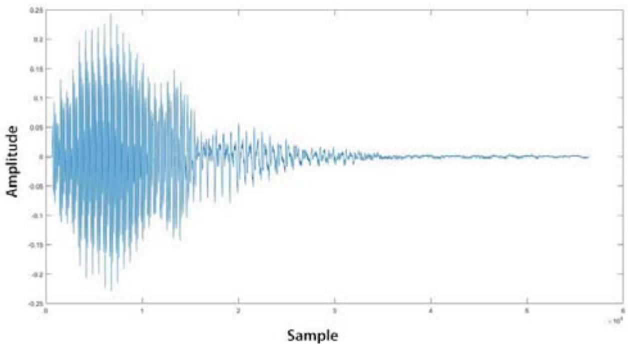

Fig. 4. Time domain representation of the letter ذ from sample1

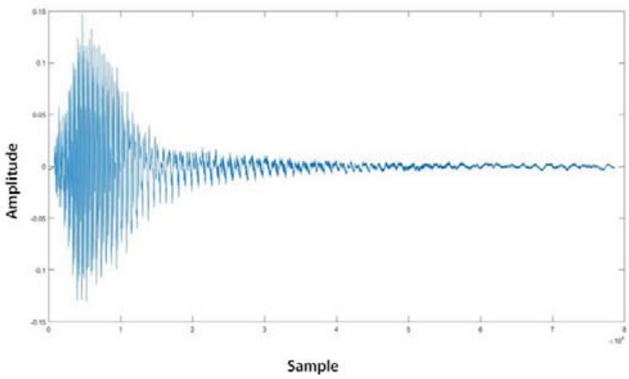

Fig. 5. Time domain representation of the letter ذ from sample2

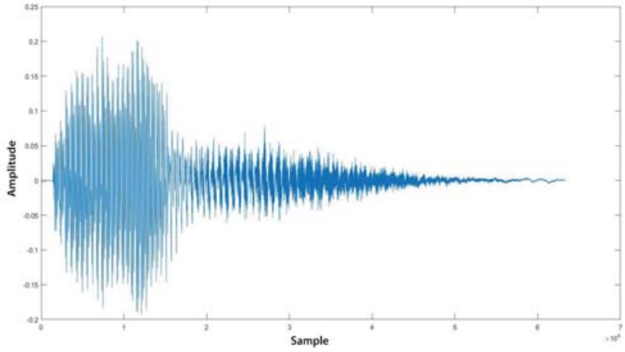

Fig. 6. Time domain representation of the letter ز from sample1

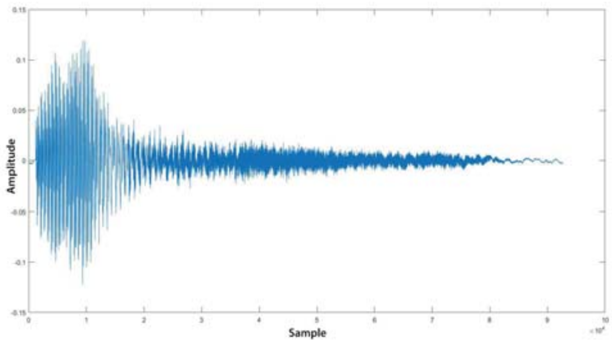

Fig. 7. Time domain representation of the letter ز from sample2

TABLE II. MFCC ANALYSIS PARAMETERS

Parameter	Value
Sampling Frequency	8000Hz
Frame width	25ms
Frame shift	10ms
Frequency range:	300 to 3700Hz
Number of the filter bank channels	20
Number of the cepstral coefficients	12

Fig. 8 and Fig. 9 show the MFCCs of the Letter ذ and ز respectively, from Sample 1. Fig. 10 and Fig. 11 show the MFCCs of the Letter ذ and ز respectively, from Sample 2. These figures are drawn from the extracted feature matrices of the letters. Table III shows the distance between the extracted MFFCs matrices of the letters from the two experts. Comparing the same letter pronunciation from the two experts, it gives the small distance, while comparing the different letters pronunciation pronounced by the same experts produce high distance value but it is less than comparing the different letter pronunciation from different experts. Comparing the Delta coefficient matrices gives the same results. Comparing the same pronunciation letter from two different experts the distance value is 325.8. But comparing the two letters pronunciations from same expert, the distance value is 347.3 while from the different experts the distance value is 392.6.

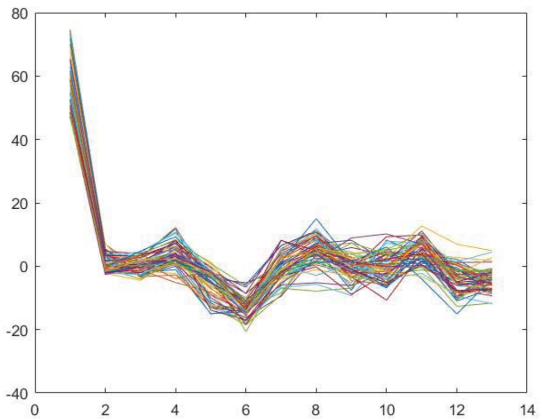

Fig. 8. MFCCs of the Letter ذ (sample1)

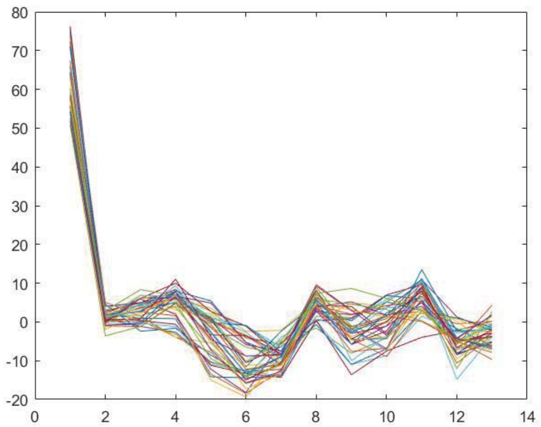

Fig. 9. MFCCs of the Letter ز (sample1)

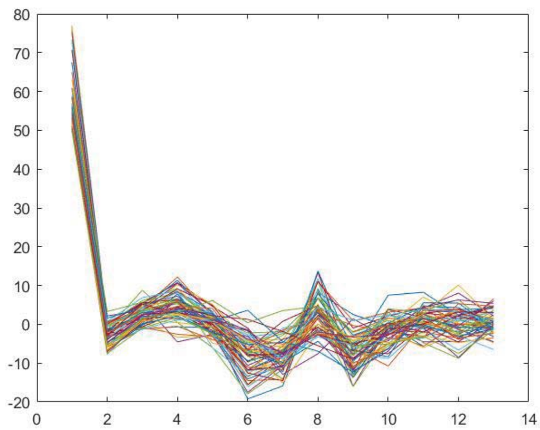

Fig. 10. MFCCs of the Letter ذ (sample2)

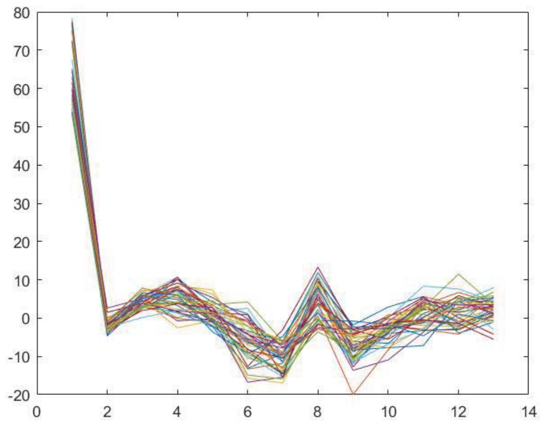

Fig. 11. MFCCs of the Letter ز (sample2)

TABLE III. CALCULATED DISTANCES BETWEEN FEATURE MATRICES

Letters to be compared		Distance between MFCCs Matrices
Sample 1 Letter ذ	Sample 1 Letter ز	11077
Sample 1 Letter ذ	Sample 2 Letter ذ	855.0093
Sample 1 Letter ذ	Sample 2 Letter ز	1959.7

IV. CONCLUSION

By using Mel-Frequency cepstral coefficients analysis, the feature matrices of two letters that are from two adjacent points of articulation (Makharij) were produced. Then they were compared by the dynamic time warping technique and the similarity distances between the matrices were calculated. Results show that the pronunciation of the same letter from two different experts produces less similarity distance, and the pronunciation of two different letters from the same expert produces high distance between the feature matrices. More letters pronunciation will be tested from more experts to ensure the MFCC ability to be used in Makharij checking system.

REFERENCES

[1] Tareq Altalmas et al., “Lips Tracking Identification of a Correct Pronunciation of Quranic Alphabets for Tajweed Teaching and Learning,” IIUM Eng. J., 2017.

[2] “Place of Articulation.” [Online]. Available: http://clas.mq.edu.au/speech/phonetics/phonetics/consonants/place.html. [Accessed: 30-Jul-2018].

[3] “about Tajweed,” 2007. [Online]. Available: http://www.abouttajweed.com/index.php/tajweed-rules/32-articulation-points-and-articulation-areas-of-the-arabic-letters/19-places-of-articulation. [Accessed: 01-Mar-2018].

[4] U. Shrawankar and V. M. Thakare, “Techniques for Feature Extraction In Speech Recognition System : A Comparative Study,” 2013.

[5] B. J. Mohan and N. Ramesh Babu, “Speech recognition using MFCC and DTW,” in 2014 International Conference on Advances in Electrical Engineering, ICAEE 2014, 2014.

[6] L. Marlina et al., “Makhraj recognition of Hijaiyah letter for children based on Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machines (SVM) method,” in 2018 International Conference on Information and Communications Technology, ICOIACT 2018, 2018.

[7] A. Elkourd and K. El Kourd, Arabic isolated word speaker dependent recognition system. 2017.

[8] S. Nuratikah et al., “Analysis of Formant Frequencies of the Correct Pronunciation of Quranic Alphabets Between Kids and Adults,” vol. 25, no. 6, pp. 1370–1379, 2017.

[9] “Product: DR-05 | TASCAM.” [Online]. Available: http://tascam.com/product/dr-05/. [Accessed: 30-Jul-2018].

[10] V. Tiwari, “MFCC and its applications in speaker recognition,” Int. J. Emerg. Technol., 2010.

[11] K. Paliwal and K. Wojcicki, “Effect of analysis window duration on speech intelligibility,” IEEE Signal Process. Lett., vol. 15, pp. 785–788, 2008.

[12] D. Eringis and G. Tamulevičius, “Improving Speech Recognition Rate through Analysis Parameters,” Electr. Control Commun. Eng., vol. 5, no. 1, pp. 61–66, 2014.

[13] K. Kumar, C. Kim, and R. Stern, “Delta-Spectral Cepstral Coefficients for Robust Speech Recognition,” IEEE Trans. Audio, Speech Lang. Process., vol. X, no. X, pp. 1–4, 2011.