



# *Iqro* Reading Learning System through Speech Recognition Using Mel Frequency Cepstral Coefficient (MFCC) and Vector Quantization (VQ) Method

Youllia Indrawaty Nurhasanah<sup>a,\*</sup>, Irma Amelia Dewi<sup>b</sup>, Bagus Ade Saputro<sup>c</sup>

<sup>a,b,c</sup> Department of Informatics, Faculty of Industrial Technology, Institut Teknologi Nasional (Itenas), Bandung, West Java, Indonesia.

## ARTICLE INFO

Received 23 February 2018  
Revised 04 April 2018  
Accepted 11 April 2018  
Available online 30 May 2018

### Keywords

*Iqro* reading,  
speech recognition,  
Mel Frequency Cepstral Coefficient (MFCC) and Vector Quantization (VQ)

## ABSTRACT

Historically, the study of Qur'an in Indonesia evolved along with the spread of Islam. Learning methods of reading the Qur'an have been found ranging from al-Baghdadi, al-Barqi, Qiraati, *Iqro'*, Human, Tartila, and others, which can make it easier to learn to read the Qur'an. Currently, the development of speech recognition technology can be used for the detection of *Iqro* vol 3 reading pronunciations. Speech recognition consists of two general stages of feature extraction and speech matching. The feature extraction step is used to derive speech-feature and speech-matching stages to compare compatibility between test sound and train voice. The speech recognition method used to recognize *Iqro* readings is extracting speech signal features using Mel Frequency Cepstral Coefficient (MFCC) and classifying them using Vector Quantization (VQ) to get the appropriate speech results. The result of testing for speech recognition system of *Iqro* reading has been tested for 30 peoples as a sample of data and there are 6 utterances indicating the information failed, so the system has a success rate of 80%.

• Corresponding author at:  
Bachelor Degree in Informatics, Faculty of Industrial Technology, Institut Teknologi Nasional (Itenas)  
Jl. PH. H Mustopa No.23 Bandung, 40124, West Java, Indonesia.  
E-mail address: youllia1975@gmail.com

ORCID ID:  
First Author: 0000-0002-6497-7932

<https://doi.org/10.25124/ijait.v2i01.1173>

Paper\_reg\_number IJAIT000020105 2018 © The Authors. Published by School of Applied Science, Telkom University.  
This is an open-access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>)

## 1. Introduction

Historically, the study of Qur'an in Indonesia evolved along with the spread of Islam. The establishment of a Mosque or Mushalla as a place of worship and as a central study for both children, adolescents, adults, parents, and public education. Learning methods of reading the Qur'an have been found ranging from al-Baghdadi, al-Barqi, Qiraati, *Iqro'* [1], Human, Tartila and others, which can make it easier to learn to read the Qur'an. This study identifies the reading sound of *Iqro*.

The previous research conducted by Heriyanto and N, Azhari S. They have studied about the reading of *Iqro* volume one, which consisted of 29 letters hijaiyah. The research's results obtained the conclusion as follows, *Iqro* reading detection through voice recognition with sampling method of energy and wave deviation get 1 wave analysis value, 2 wave analysis and so on has accuracy approximately 45%. In the same person has done testing for indexed first, there is a precision of 60% [1]. Based on the previous research, our research was inspired by that research to build applications for learning *Iqro* reading through voice recognition using MFCC and VQ method. In this research used *Iqro* volume 3, which introducing another harokat, such as fathah, kasroh and dhomah.

Speech recognition consists of two general stages of feature extraction and voice matching. The feature extraction step is used to determine the sound characteristic. Then, the voice matching will find out the suitability of test sound and voice training. Two widely feature extraction techniques are Mel-Frequency Cepstral Coefficients (MFCC) and Linear Prediction Coefficients (LPC). MFCC is more accurate than LPC [2]. The MFCC algorithm is used to simulate the feature extraction module. Using the MFCC algorithm, co-efficient cepstral is calculated from the Mel frequency scale. The VQ (Vector Quantization) method is used to reduce the amount of data, thus to reduce the computation time. In matching stage, Euclidean distance applied for calculating the similarity in criterion [3].

Therefore, in this study, the authors will design a system that can recognize the *Iqro* readings correctly. Voice training for pronunciation fluent is extracted characteristics using MFCC and matched using VQ.

**Table 1** *Iqro* volume 3 used in this research

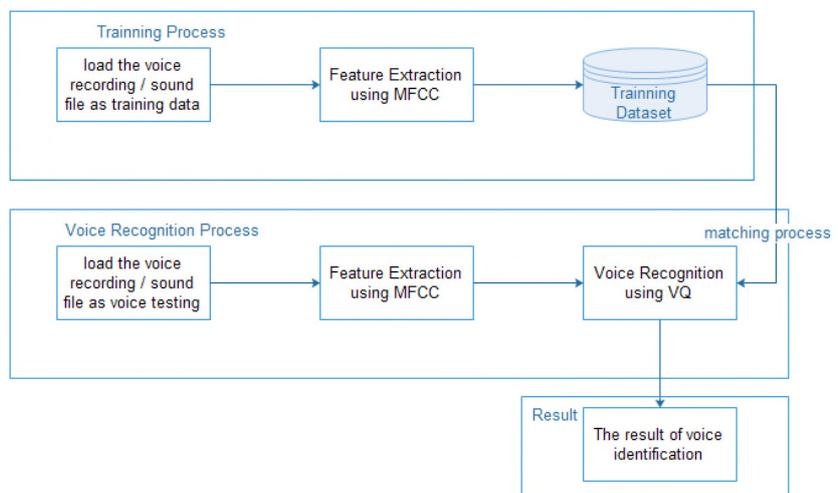
| Reading <i>Iqro</i> Vol 3             |
|---------------------------------------|
| قَادِرَ                               |
| بَاطِلَ                               |
| عَاصِفَ                               |
| كَوَاكِبَ                             |
| قَالُوا بَلَىٰ                        |
| بِمَا خَلَقَ وَعَلَا                  |
| وَجَعَلَ فِيهَا رُوسِي                |
| تَرَكَ ءَالَ مُوسَىٰ وَءَالَ هَارُونَ |

1. Training data. Voice training has been taken from Ustadz and Ustadzah as well as the Santri. Sound samples of reference using 6 different voices that are 3 women and 3 men. The words contained in *Iqro* volume 3 are shown in Table 1 as follows [4].
  - a. Each Ustadz and Ustadzah pronounce every word contained in *Iqro* volume 3 are 10 times. So that obtained 60 training data for each word. The training data is used as a template or reference data to be compared with 30 test sounds.
  - b. After recording the reference data is complete then the feature extraction is processed by using the MFCC method to obtain a reference value, and the value of feature extraction is stored in the database as training data or reference data.
2. Test data. For the sample sound of the test is as many as 30 people consisting of 15 women and 15 men. The test sample is taken from Student and Elementary Student (SD), the students will be recorded according to *Iqro* reading after the recording process is completed will be processed feature extraction using MFCC in order to obtain a value corresponding to training data. Then the test data will be matching using the VQ method to see the results of matching test data with the train data, whether 'appropriate or not appropriate'.

## 2. Discussion

Speech recognition is defined as the process of converting voice signals to machine language in the form of digital data [5]. The voice recognition system is a combination of speaker recognition and speech recognition [6].

In Figure 1, the initial step to creating a system is sound recording. In this system, there are two sub-processes, namely processing training data and testing data. The initial process is undertaken to detect *Iqro* readings to perform data processing. The training data is used as a parameter or reference to detect *Iqro* readings while the second process is used for the adjustment of test data with training data from this application. Training data is taken from several sound samples based on the *Iqro* reading by recording sound and then stored as reference data while test data is taken by performing the process of recording the user's voice directly or by taking recorded data (voice already saved).



**Figure 1** Block diagram of training and recognition process

Figure 1 is the steps of learning process *Iqro* reading system, and the steps as follows.

### 2.1. Voice Acquisition Process

Spoken words are transformed into digital signals by converting sound waves into a set of numbers that are then adapted to specific codes to identify the words [7]. The results of the spoken word identification can be displayed in a written form or can be read by a technology device as a command to perform the job, such as pressing a button on a handheld phone that is done automatically with a voice command. In the process of a voice data will be explained in the data acquisition.

Data acquisition system is a system that serves to retrieve, collect and prepare data, process it to produce the desired data. This method aims to simplify every step taken in the whole process.

Sound signal data is obtained by recording sound through a microphone connected to the computer. Voice recording is performed on the application using the standard 44100 Hz sampling frequency. Sound signal can use 16 bits/sample and 1 for the mono channel. Shorter recording duration in a short time is easier to take the difference between its features. In this analysis, we use a 5-second sample. The calculation to determine the sampling data can use Equation 1 as follows.

$$X = Fs \times dt(seconds) \times \left(\frac{bit}{8}\right) \times j \tag{Equation 1}$$

- X: signal sampling data
- bit: number of bits of resolution
- j: channel (mono = 1)
- Fs: sampling frequency
- dt: recording duration (sec)

Calculations on the data acquisition process for sampling are:

$$X = 44100 \times 5 \times \left(\frac{16}{8}\right) \times 1 = 441000 \text{ byte}$$

To calculate sample rate, use it in a way  $\frac{Fs}{Ts}$  and time 5 seconds then the result.

$$\frac{Fs}{Ts} = \frac{441000}{5} = 88200 \text{ Hz}$$

### 2.2. Feature Extraction

After obtaining the digital voice through data acquisition process, next process is feature extraction. In this research, Mel Frequency Cepstral Coefficient (MFCC) is used to extract the characteristics of a sound signal. Sound recordings that have been entered into the system are extracted using MFCC method to be recognized by the computer [8]. Steps of MFCC process [9] are described in Figure 2 as follows.

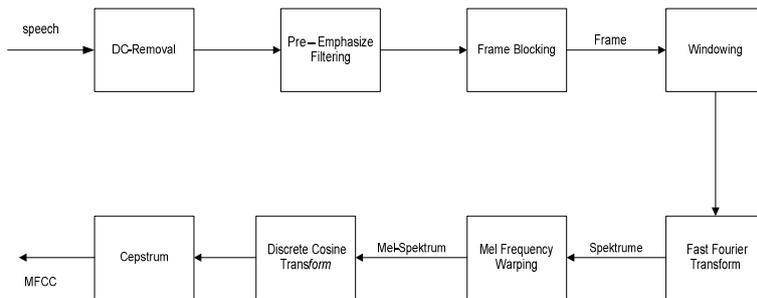


Figure 2 MFCC Diagram Block

*a. DC - Removal*

DC-Removal aims to calculate the average voice sample data and subtract the value of each sound sample by an average value. The goal is to get the normalization of voice input. The input of this process is the sample sound in the form of an array and the result is the normalization of array sound samples.

Based on Equation 2 for DC Removal calculations, the following is an example of some sample data that will be used for calculation using formulas contained in the MFCC method. The sample data consists of 10, 11, 13, 14, 16, 18, 20, 9, 6, 4 are calculated as reference data.

$$\bar{x} = \frac{10+11+13+14+16+18+20+9+6+4}{10} = 12.1 \tag{Equation 2}$$

After the average value is known in Equation 3, the initial signal value is subtracted with DC Removal result, so resulting from the signal value as follows.

$$y[n] = x[n] - \bar{x} ; 0 \leq n \leq N - 1 \tag{Equation 3}$$

- y[n]: signal sample from DC-Removal
- x[n]: original signal samples
- $\bar{x}$ : the average value of the original signal sample
- N: signal length

|                        |                       |
|------------------------|-----------------------|
| y0 = 10 - 12.1 = - 2.1 | y5 = 18 - 12.1 = 5.9  |
| y1 = 11 - 12.1 = - 1.1 | y6 = 20 - 12.1 = 7.9  |
| y2 = 13 - 12.1 = 0.9   | y7 = 9 - 12.1 = - 3.1 |
| y3 = 14 - 12.1 = 1.9   | y8 = 6 - 12.1 = - 6.1 |
| y4 = 16 - 12.1 = 3.9   | y9 = 4 - 12.1 = - 8.1 |

So the result of signal data after DC Removal is (-2.1, -1.1, 0.9, 1.9, 3.9, 5.9, 7.9, -3.1, -6.1, -8.1).

*b. Pre-Emphasize*

Pre-emphasize is used to correct the signal from noise interference, thus increasing the accuracy of the extraction process. The default alpha value is 0.97. If the result of DC removal is (-2.1, -1.1, 0.9, 1.9, 3.9, 5.9, 7.9, -3.1, -6.1, -8.1), so pre-emphasize is calculated using Equation 4 and value  $\alpha = 0,97$  as follows.

$$y[n] = s[n] - \alpha s[n - 1] ; 0.9 \leq \alpha \leq 1.0 \tag{Equation 4}$$

- y[n]: signal results of pre-emphasize
- s[n]: signal before pre-emphasize

|                                     |                                      |
|-------------------------------------|--------------------------------------|
| Y0 = -2.1                           | Y5 = (5.9 - (3.9 * 0.97)) = 2.117    |
| Y1 = (-1.1 - (-2.1 * 0.97)) = 0.937 | Y6 = (7.9 - (5.9 * 0.97)) = 2.167    |
| Y2 = (0.9 - (-1.1 * 0.97)) = 1.967  | Y7 = (-3.1 - (7.9 * 0.97)) = -10.763 |
| Y3 = (1.9 - (0.9 * 0.97)) = 1.027   | Y8 = (-6.1 - (-3.1 * 0.97)) = -3.093 |
| Y4 = (3.9 - (1.9 * 0.97)) = 2.057   | Y9 = (-8.1 - (-6.1 * 0.97)) = -2.183 |

The new signal after pre-emphasize is obtained from DC Removal data added with pre-emphasize data from Equation 4. So signal after pre-emphasis is:



The result of pre-emphasize process (-4.2, -0.163, 2.867, 2.927, 5.957, 8.017, 10.067, -13.863, -9.193, -10.283) will be done windowing process using Equation 8.

$$x(n) = xi(n) * w(n) \quad ; \quad n = 0.1, \dots, N - 1 \quad \text{Equation 8}$$

x(n): signal sample value

N: frame size, is a multiple of 2

xi(n): sample value of signal frame 1

w(n): windowing function

$$X_0 = -4.1 * 0.08 = -0.33$$

$$X_5 = 8.017 * 0.08 = 0.64$$

$$X_1 = -0.163 * 0.08 = -0.01$$

$$X_6 = 10.067 * 0.08 = 0.80$$

$$X_2 = 2.867 * 0.08 = 0.23$$

$$X_7 = -13.863 * 0.08 = -1.11$$

$$X_3 = 2.927 * 0.08 = 0.23$$

$$X_8 = -9.193 * 0.08 = -0.73$$

$$X_4 = 5.957 * 0.08 = 0.48$$

$$X_9 = -10.283 * 0.08 = -0.82$$

In the windowing process is obtained values as follows. (-0.33, -0.01, 0.23, 0.23, 0.48, 0.64, 0.80, -1.11, -0.73, -0.82).

#### e. *Fast Fourier Transform (FFT)*

FFT is one of the methods to convert a voice signal into a frequency signal. Equation 9.

$$F(K) = \sum_{n=1}^N f(n) \cos\left(\frac{2\pi n k T}{N}\right) - j \sum_{n=1}^N f(n) \sin\left(\frac{2\pi n k T}{N}\right) \quad \text{Equation 9}$$

$$\begin{aligned} F_0 &= \frac{1}{4} [-0.33 \left( \cos\left(\frac{2\phi i * 0 * 0}{10}\right) \right)] - j \sin\left(\cos\frac{2\phi i * 0 * 0}{10}\right) + \\ &[-0.01 \left( \cos\left(\frac{2\phi i * 0 * 1}{10}\right) \right)] - j \sin\left(\cos\frac{2\phi i * 0 * 1}{10}\right) + [0.23 \left( \cos\left(\frac{2\phi i * 0 * 2}{10}\right) \right)] - \\ &j \sin\left(\cos\frac{2\phi i * 0 * 2}{10}\right) + [0.23 \left( \cos\left(\frac{2\phi i * 0 * 3}{10}\right) \right)] - \\ &j \sin\left(\cos\frac{2\phi i * 0 * 3}{10}\right) [0.48 \left( \cos\left(\frac{2\phi i * 0 * 4}{10}\right) \right)] - \\ &j \sin\left(\cos\frac{2\phi i * 0 * 4}{10}\right) [0.64 \left( \cos\left(\frac{2\phi i * 0 * 5}{10}\right) \right)] - j \sin\left(\cos\frac{2\phi i * 0 * 5}{10}\right) + \\ &[0.80 \left( \cos\left(\frac{2\phi i * 0 * 6}{10}\right) \right)] - j \sin\left(\cos\frac{2\phi i * 0 * 6}{10}\right) + [-1.11 \left( \cos\left(\frac{2\phi i * 0 * 7}{10}\right) \right)] - \\ &j \sin\left(\cos\frac{2\phi i * 0 * 7}{10}\right) + [-0.73 \left( \cos\left(\frac{2\phi i * 0 * 8}{10}\right) \right)] - j \sin\left(\cos\frac{2\phi i * 0 * 8}{10}\right) + \\ &[-0.82 \left( \cos\left(\frac{2\phi i * 0 * 9}{10}\right) \right)] - j \sin\left(\cos\frac{2\phi i * 0 * 9}{10}\right) = -0.16 + 0j = -0.16 \end{aligned}$$

For the calculation of F1 to F9 is equal to  $F_0 = \frac{1}{4}$  just different in part f as follows.

$$F_1 = \frac{2}{4}, F_2 = \frac{3}{4}, F_3 = \frac{4}{4}, F_4 = \frac{5}{4}, F_5 = \frac{6}{4}, F_6 = \frac{7}{4}, F_7 = \frac{8}{4}, F_8 = \frac{9}{4}, F_9 = \frac{10}{4}$$

So, data is obtained from the FFT signal results as follows. (F0 = -0.16, F1 = -0.32, F2 = -0.48, F3 = -0.64, F4 = -0.8, F5 = -0.96, F6 = -1.12, F7 = -1.28, F8 = -1.44, F9 = -1.6)

#### f. *Mel Frequency Warping*

Mel Frequency Warping is generally using filterbank. Filterbank is one of the form of filter has aimed to find out size of a particular band frequency on the sound signal. To get  $H_i$ , use Equation 10:

$$H_i = 2595 * \log(1 + 1000/700) / S_i / 2 \quad \text{Equation 10}$$

$S_i$ : results from FFT

$$H_0 = 2595 * \log(1 + 1000/700) / -0.16 / 2 = -3124.95$$

$$H_1 = 2595 * \log(1 + 1000/700) / -0.32 / 2 = -1562.47$$

$$H_2 = 2595 * \log(1 + 1000/700) / -0.48 / 2 = -1041.65$$

$$H_3 = 2595 * \log(1 + 1000/700) / -0.64 / 2 = -781.24$$

$$H_4 = 2595 * \log(1 + 1000/700) / -0.8 / 2 = -624.99$$

$$H_5 = 2595 * \log(1 + 1000/700) / -0.96 / 2 = -520.82$$

$$H_6 = 2595 * \log(1 + 1000/700) / -1.12 / 2 = -446.42$$

$$H_7 = 2595 * \log(1 + 1000/700) / -1.28 / 2 = -390.62$$

$$H_8 = 2595 * \log(1 + 1000/700) / -1.44 / 2 = -347.22$$

$$H_9 = 2595 * \log(1 + 1000/700) / -1.6 / 2 = -312.49$$

Then the filterbank calculation can use Equation 11 as follows.

$$Y[t] = \sum_{j=1}^n S[j] H_i[j] \quad \text{Equation 11}$$

N : amount magnitude spectrum

$S[j]$ : Magnitude spectrum on frequency j

$H_i[j]$ : filterbank coefficient on frequency j ( $1 \leq i \leq M$ )

M: number of channels in the filterbank

$$Y_0 = -0.16 * -3124.95 = 499.99$$

$$Y_5 = -0.96 * -520.82 = 499.98$$

$$Y_1 = -0.32 * -1562.47 = 499.99$$

$$Y_6 = -1.12 * -446.42 = 499.99$$

$$Y_2 = -0.48 * -1041.65 = 499.99$$

$$Y_7 = -1.28 * -390.62 = 499.99$$

$$Y_3 = -0.64 * -781.24 = 499.99$$

$$Y_8 = -1.44 * -347.22 = 499.10$$

$$Y_4 = -0.8 * -624.99 = 499.99$$

$$Y_9 = -1.6 * -312.49 = 499.98$$

The result of filterbank value = (499.99, 499.99, 499.99, 499.99, 499.98, 499.99, 499.99, 499.10, 499.98).

#### g. *Discrete Cosine Transform (DCT)*

The DCT process is the last step of the feature extraction process. DCT is used to obtain some dimensional value of the corresponding vector. DCT is calculated by the Equation 12.

Coefficient (k) = 10, for n = 0.

$$C_n = \sum_{k=1}^k (\log SK) \cos \left[ n \left( K - \frac{1}{2} \right) \frac{\pi}{k} \right], \quad n = 1, 2 \dots K \quad \text{Equation 12}$$

SK = the output of the filterbank process on index K

K = number of expected coefficients

$$C_0 = \log (499.99) \cos[0(0-1/2)3.14/10] + \log (499.99) \cos[0(1-1/2)3.14/10] + \log (499.99) \cos[0(2-1/2)3.14/10] + \log (499.99) \cos[0(3-1/2)3.14/10] + \log (499.98) \cos[0(0-4/2)3.14/10] + \log (499.99) \cos[0(5-1/2)3.14/10] + \log (499.99) \cos[0(6-1/2)3.14/10] + \log (499.99) \cos[0(7-1/2)3.14/10] + \log (499.10) \cos[0(8-1/2)3.14/10] + \log (499.98) \cos[0(9-1/2)3.14/10] = 26.987$$

So the DCT value = 26.987

### h. Cepstrum (Cepstral Liftering)

The result of DCT process is Cepstrum which is the end result of feature extraction process. Cepstrum is the inverse term for the spectrum. It is used to obtain information from human speech sound signals. To be able to improve the quality of the recognition process, then the result of the DCT process must be through the process of Cepstral Liftering. Cepstral Liftering is calculated by the Equation 13.

$$w[n] = \begin{cases} 1 + \frac{L}{2} \sin\left(\frac{n\pi}{L}\right); & n = 1, 2, \dots, L \\ 0 & \end{cases} \quad \text{Equation 13}$$

L = number of cepstral coefficients

N = index of cepstral coefficients

### 2.3. Voice Matching Process

The output obtained from the feature extraction are used as input data in the sound matching process. Vector Quantization (VQ) is the template matching methods. VQ is a vector mapping process of many vectors become certain vector only. In speech pronunciation, the vector is the characteristic of each utterance derived from the feature extraction process.

VQ has produced a vector representation of the characteristics of each utter with a number of vectors. The vector is referred to as the codebook of each utterance. The algorithm used to construct the codebook is the LBG algorithm (*algorithm Linde Buzo Grey*) [10] shown in Figure 3.

The LBG algorithm [11] is implemented in the following repetitive procedure.

1. Specify the vector codebook, which is the centroid of the overall nature of the vector (no repetition at this stage).
2. Doubling the number of code books by dividing each vector  $y_n$  of the code book that has been formed, with the rules in Equation 14 and Equation 15.

$$y_n^+ = y_n(1+\epsilon) \quad \text{Equation 14}$$

$$y_n^- = y_n(1-\epsilon) \quad \text{Equation 15}$$

Where  $n$  has a value between 1 to  $M$  (desired codebook size), and  $\epsilon$  is the divisor parameter (where  $\epsilon = 0.01$ ).

3. Searching Nearest - Neighbors for a feature vector, find the codeword in the codebook (the closest codebook) (distortion at least) and place the vector in the codeword group.

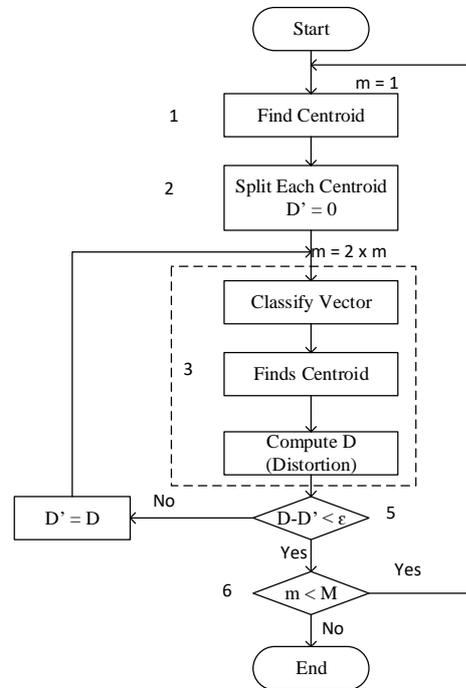


Figure 3 LBG algorithm flow diagram

4. Centroid renewal. Update the codeword in each group by using the centroid of the feature vector present in the cell.
5. Iteration I. Repeat steps 3 and 4 until the mean deviation distance ( $D$ ) is obtained that is below the specified limit ( $\epsilon$ ).  $D'$  is the initial distortion value whose value is determined at initialization at the beginning of the program.
6. Iteration II. Repeat steps 2, 3 and 4 until the code book of size  $M$  is obtained.

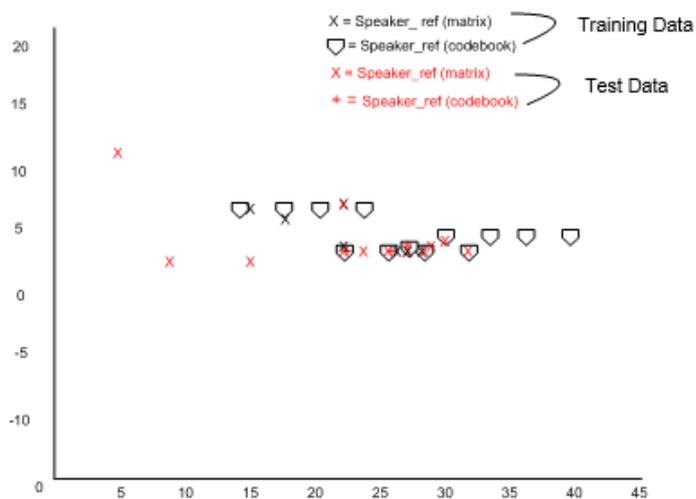


Figure 4 Match Result (codebook)

After the extraction process, the sound characteristics are stored in the application database as training data.

VQ method has checked the suitability by calculated the distance of the test data value with the training data. The results of matching data can be seen in Figure 4.

### 3. Result

In the application of "Iqro reading system through voice recognition using the MFCC and VQ methods" can identify with the closest distance. System process can be seen in Figure 5.

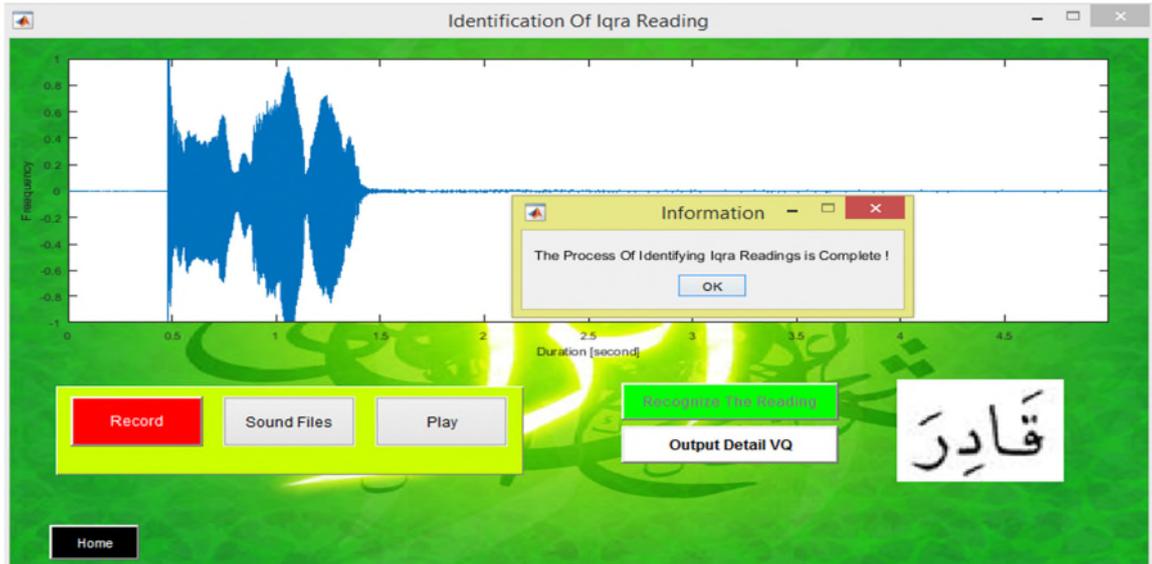


Figure 5 Display the results of Iqro reading identification

Identification process can be done by pressing the record button or sound file (retrieve existing voice data), then pressed recognize the reading button. If the user’s utterance is in accordance with training data, so system displayed the appropriate image Iqro. But, if the user’s utterance is not in accordance with the training data, the system displays the information cannot be detected or displayed another image that has the closest distance.

Table 2 Iqro Reading Matching Results (reading "Qoodiro")

| Input of tested utterance (Iqro' vol. 3) | Test speech Input from the User | Output from application | Result        | Percentage |
|--|---------------------------------|-------------------------|---------------|------------|
| قَادِر                                   | Aji                             | قَادِر                  | corresponding | 80%        |
|  | Cross                           | قَادِر                  | corresponding |            |
|  | Debi                            | قَادِر                  | corresponding |            |
|  | Dindi                           | كَوَاكِب                | not suitable  |            |
|  | Fajri                           | قَادِر                  | corresponding |            |
|  | Gilang                          | قَادِر                  | corresponding |            |

| Input of tested utterance<br>( <i>Iqro'</i> vol. 3) | Test speech<br>Input from<br>the User | Output from application | Result        | Percentage |
|---|---------------------------------------|-------------------------|---------------|------------|
|   | Imam                                  | قَادِرَ                 | corresponding |            |
|   | Khemal                                | قَادِرَ                 | corresponding |            |
|   | Rizki                                 | قَادِرَ                 | corresponding |            |
|   | Wira                                  | قَادِرَ                 | corresponding |            |
|   | Gilang                                | قَادِرَ                 | corresponding |            |
|   | Ismail                                | كَوَاكِبُ               | not suitable  |            |
|   | Rafa                                  | بَاطِلَ                 | not suitable  |            |
|   | Raga                                  | قَادِرَ                 | corresponding |            |
|   | Zainal                                | قَادِرَ                 | corresponding |            |
|   | Anisa                                 | قَادِرَ                 | corresponding |            |
|   | Audhya                                | قَادِرَ                 | corresponding |            |
|   | Bintang                               | قَادِرَ                 | corresponding |            |
|   | Dewi                                  | قَادِرَ                 | corresponding |            |
|   | Finka                                 | قَادِرَ                 | corresponding |            |
|   | Inok                                  | قَادِرَ                 | corresponding |            |
|   | Lela                                  | بَاطِلَ                 | not suitable  |            |
|   | Nanda                                 | قَادِرَ                 | corresponding |            |
|   | Rizka                                 | قَادِرَ                 | corresponding |            |
|   | Sindri                                | بَاطِلَ                 | not suitable  |            |
|   | Windi                                 | قَادِرَ                 | corresponding |            |
|   | Adkya                                 | قَادِرَ                 | corresponding |            |
|   | Kayla                                 | قَادِرَ                 | corresponding |            |
|   | Nayaka                                | قَادِرَ                 | corresponding |            |

| Input of tested utterance<br>( <i>Iqro'</i> vol. 3) | Test speech<br>Input from<br>the User | Output from application   | Result       | Percentage |
|---|---------------------------------------|---|--------------|------------|
|   | Silmi                                 |  | not suitable |            |

Success rate is calculated using Equation 16.

$$\frac{\text{matching results}}{\text{total all trials}} \times 100\% \quad \text{Equation 16}$$

As shown in Table 2, there are 30 data for pronunciation "qodiro" from male and female training voice. Of the 30 test data, there are 6 pronunciations indicated that the information failed. In the test of Table 2, the voice failed is an unsuitable sound between the training sound and the detection result where the test sound is not in accordance with the Tartila reading. The percentage of accuracy was calculated using the Equation 16.

$$\frac{24}{30} \times 100\% = 80\%$$

#### 4. Conclusion

Based on research "Iqro reading learning application through speech recognition using Mel Frequency Cepstrum Coefficient (MFCC) and Vector Quantization (VQ) method", the conclusion is as follows.

The application can be used to identify the pronunciation of words on *Iqro* vol 3 by checking the suitability of the pronunciation derived from using MFCC feature extraction and matching using VQ. Application tests results reached 80%. In terms of the accuracy and appropriateness of pronunciation in *Iqro* vol 3, this application can help to learn *Iqro* vol 3 which consists of words that have a harokat, such as fathah, kasroh and dhomah. This application can be used to improve the speech of the user to study reading *Iqro* vol 3.

#### Bibliography

- [1] Heriyanto and Azhari, S. N., "Rekayasa Perangkat Lunak Untuk Klasifikasi Bacaan Iqra Melalui Voice Recognition Menggunakan Metode Sampling Deviasi Everage Energy Dan Deviasi Wave," *Telematika*, vol. 10, no. 2, p. 119 – 128, 2014.
- [2] C. B. Soltane and I. Y. Kelbesa, "An Intelligent Text Independent Speaker Identification Using VQ-GMM Model Based Multiple Classifier System," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 8, no. 10, p. 1949, 2014.
- [3] K. Patel and R. K. Prasad, "Speech Recognition and Verification Using MFCC & VQ," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 5, pp. 478-483, 2013.
- [4] K. A. Humam, *Buku Iqro, Cara cepat membaca Alquran, Jilid 3*, Yogyakarta: Balai Litbang LPTQ Nasional, Team Tadarus AMM, 2000.
- [5] A. Y. Idwal, Y. I. Nurhasanah and D. B. Utami, "Pengembangan Aplikasi Aksent Daerah Bahasa Indonesia," *Jutisi Maranatha*, vol. 3, no. 3, p. 461, 2017.
- [6] R. S. Azizah, D. Nurjanah and F. D. Sari, "Sistem Automatic Speech Recognition menggunakan Metode MFCC dan HMMS untuk Deteksi Kesalahan Pengucapan Kata Bahasa Inggris," *e-Proceedings of Engineering*, vol. 2, no. 3, 2015.
- [7] A. Charisma, "Sistem Verifikasi Penutur Kata," *Jurnal Teknik Elektro (JTE) -ITP*, vol. 2, 2013.

- [8] F. Darmadi, A. Rizal and U. Sunardi, "Detection Of Sleep Apnea By Analysis Of Snore Sounds Using Mel Frequency Cepstrum Coefficient Method," *e-Proceedings of Engineering*, vol. 2, no. 2, p. 2681, August 2015.
- [9] D. Putra and A. Resmawan, "Verifikasi Biometrika suara menggunakan metoda MFCC dan DTW," *Lontar Komputer*, vol. 2, no. 1, p. 8, 2011.
- [10] L. Rabiner and B.-H. Juang, "Fundamentals Of Speech Recognition," New Jersey, Prentice Hall, 1993, pp. 126-127.
- [11] R. A. Sadewa, T. A. B. Wirayuda and S. Sa'adah, "Speaker Recognition Implementasi For Authentication Using Modified MFCC - Vector Quantization LGB Algorithm," *e-Proceeding of Engineering*, vol. 2, no. 1, p. 1453, April 2015.