

## Development of Quranic Reciter Identification System using MFCC and GMM Classifier

Teddy Surya Gunawan<sup>1</sup>, Nur Atikah Muhamat Saleh<sup>2</sup>, Mira Kartiwi<sup>3</sup>

<sup>1,2</sup>Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia

<sup>3</sup>Department of Information Systems, International Islamic University Malaysia, Malaysia

---

### Article Info

#### Article history:

Received Oct 2, 2017

Revised Dec 2, 2017

Accepted Dec 16, 2017

#### Keyword:

GMM

MFCC

Quranic recitation

Speaker identification

---

### ABSTRACT

Nowadays, there are many beautiful recitation of Al-Quran available. Quranic recitation has its own characteristics, and the problem to identify the reciter is similar to the speaker recognition/identification problem. The objective of this paper is to develop Quran reciter identification system using Mel-frequency Cepstral Coefficient (MFCC) and Gaussian Mixture Model (GMM). In this paper, a database of five Quranic reciters is developed and used in training and testing phases. We carefully randomized the database from various surah in the Quran so that the proposed system will not prone to the recited verses but only to the reciter. Around 15 Quranic audio samples from 5 reciters were collected and randomized, in which 10 samples were used for training the GMM and 5 samples were used for testing. Results showed that our proposed system has 100% recognition rate for the five reciters tested. Even when tested with unknown samples, the proposed system is able to reject it.

Copyright © 2018 Institute of Advanced Engineering and Science.

All rights reserved.

---

### Corresponding Author:

Teddy Surya Gunawan,  
Department of Electrical and Computer Engineering,  
International Islamic University Malaysia,  
Jalan Gombak, 53100 Kuala Lumpur, Malaysia.  
Email: tsgunawan@iiu.edu.my

---

## 1. INTRODUCTION

Al-Quran is the holy book of Muslims which is written and recited in Arabic language. Interestingly, Al-Quran is the most popular and most recited book of all time [1], [2]. Muslim should try their best to avoid mistakes in reciting the Quran, such as reciting rules (*tajwid*), missing words, verses, misreading vowel pronunciations, punctuations, and accents [3]. Recitation should follow the rules of pronunciation, intonation, and caesuras established by the the Islamic prophet Muhammad (PBUH). The rules and guidance to read Quran is propagated from the prophet Muhammad until the Quranic reciter through a verified chain of transmission (*sanad*). Many non-Arabic people studied and learnt Al-Quran by listening to the well known Quranic reciters (*qari*). Although each reciter recited the same Quranic verses, but it has differences due to their unique voice and characteristics. To identify the Quranic reciter, the problem is similar to the speaker recognition [4].

Typical speaker recognition system includes pre-processing, feature extraction, and classification [5]. Many features and classifiers have been used in the speaker recognition research. Audio features such as Mel-frequency Cepstral Coefficients [4], [6], linear-frequency cepstral coefficients (LFCC), and linear predictive coefficients (LPC). LFCC is similar to MFCC except that their frequencies is not warped by a non-linear frequency scale and it has been found that LFCC performed better than MFCC in female trials [7]. As stated by [8] is the most commonly used features in speaker recognition.

Given a set of feature vectors, each speaker model will be built so that a vector from the same speaker has higher probability compared to any other models. Several classifiers have been used, such as *k*-

nearest neighbors, vector quantization [6], hidden Markov model (HMM) [9], Gaussian mixture model (GMM) [10], artificial neural network [4], and deep neural network (DNN) [11]. Of the various classifiers available, in this research we selected GMM as our baseline for speaker recognition.

Although many researches have been conducted on speaker recognition, but very limited are targeted on the Quranic reciter recognition. Recent research conducted by [12] stated that the Quranic recitation has different characteristics compared to the English spoken language. The Quranic recitation is predominantly voiced speech, in which it could potentially be exploited to build more efficient speaker models. Therefore, the objective of this research is to develop a Quranic recitation identification using MFCC and GMM, and to evaluate its performance. The rest of the paper is organized as follows: Section 2 describes the typical components in a speaker recognition system. Section 3 explains the proposed Quranic reciter identification system. Section 4 evaluates its performance in terms of recognition rate, while Section 5 concludes this paper.

## 2. SPEAKER RECOGNITION

The flow chart of basic model for recognition speaker as shown in Figure 1. First, the audio signal is going through the front-end processing, in which the features that could uniquely represent the speaker information are extracted. The short-time spectral is the most-frequently used typed of features [5]. The front-end may also include pre-processing modules, such as voice activity detection to remove silence from the input, or a channel compensation module to normalize the effect of the recording channel [5], [13].

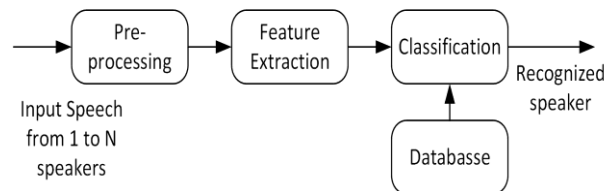


Figure 1. Typical Speaker Recognition System

Currently, there are many methods that can be used to verify a speaker identity and the most two known methods are linear predictive coding (LPC) and Mel frequency cepstrum (MFCC) [4], [6]. However, in this paper MFCC methods is chosen as the feature extraction since the system give higher accuracy. MFCC is the most popular method due to it is easy to moderate and can handle multiple speakers or multiple languages.

A vector of features acquired from the previous step is then compared against a set of speaker models. The identity of the test speaker is associated with the identity of the highest scoring model. A speaker model is a statistical model that represents speaker-dependent information, and can be used to predict new data. Any modeling techniques can be used, but the most popular techniques are: clustering, hidden Markov model, artificial neural network, and Gaussian mixture model [4], [5], [14]. In this research, we used GMM as it is one of the most effective techniques in speaker recognition [5]. GMM used estimation maximum log-likelihood algorithm to find the pattern matching and is able to form smooth approximation for arbitrarily shaped densities.

## 3. PROPOSED QURANIC RECITER IDENTIFICATION SYSTEM

Figure 2 illustrates our proposed system for Quranic recitation identification. We used MFCC for feature extraction and GMM for classifier due to its popularity and effectiveness for speaker recognition.

### 3.1. Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs use a non-linear frequency scale, i.e. mel scale, based on the auditory perception. A *mel* is a unit of measure of perceived pitch or frequency of a tone. Equation (1) can be used to convert frequency scale to mel scale.

$$f_{mel} = 117 \ln \left( 1 + \frac{f_{Hz}}{700} \right) \quad (1)$$

where  $f_{mel}$  is the frequency in mels and  $f_{Hz}$  is the normal frequency in Hz. MFCCs are often calculated using a filter bank of  $M$  filters, in which each filter has a triangular shape and is spaced uniformly on the mel scale as shown in Equation (2).

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \frac{k - f[m-1]}{f[m] - f[m-1]} & f[m-1] < k \leq f[m] \\ \frac{f[m+1] - k}{f[m+1] - f[m]} & f[m] \leq k < f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (2)$$

where  $m = 0, 1, \dots, M - 1$ . The log-energy mel spectrum is then calculated as follows:

$$S[m] = \ln \left[ \sum_{k=0}^{N-1} |X[k]|^2 H_m[k] \right] \quad m = 0, 1, \dots, M - 1 \quad (3)$$

where  $X[k]$  is the discrete Fourier transform (DFT) of a speech input  $x[n]$ .

Although traditional cepstrum uses inverse discrete Fourier transform (IDFT), mel frequency cepstrum is normally implemented using discrete cosine transform (DCT) since  $S[m]$  is even as shown in Equation (4), as follows:

$$\hat{x}[n] = \sum_{m=0}^{M-1} S[m] \cos \left[ \left( m + \frac{1}{2} \right) \frac{\pi n}{M} \right] \quad n = 0, 1, \dots, M - 1 \quad (4)$$

Typically, the number of filters  $M$  ranges from 20 to 40, and the number of kept coefficients is 13. Some research reported that the performance of speech recognition and speaker identification systems reached peak with 32-35 filters [8].

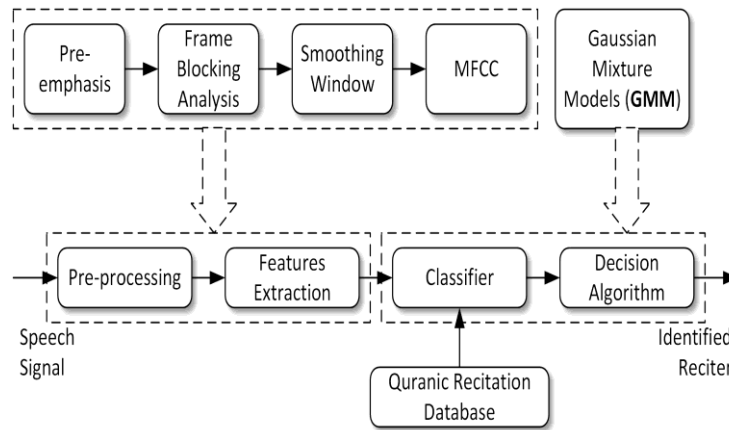


Figure 2. The proposed Quranic reciter identification system

### 3.2. Gaussian Mixture Model (GMM)

GMM provides a probabilistic model of a speaker’s voice. A Gaussian mixture distribution is a weighted sum of  $M$  densities:

$$p(\vec{x} | \lambda) = \sum_{i=1}^M p_i b_i(\vec{x}) \quad (5)$$

where  $\vec{x}$  is a  $D$ -dimensional vector,  $b_i(\vec{x})$  is the  $i$ -th component density, and  $p_i$  is the weight of the  $i$ -th component. The mixture weights satisfy  $\sum_{i=1}^M p_i = 1$ . Each mixture component is a  $D$ -variate Gaussian density function:

$$b_i(\vec{x}) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|^{\frac{1}{2}}}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)\right) \quad (6)$$

where  $\mu_i$  is the mean vector, and  $\Sigma_i$  is the covariance matrix. A GMM is characterized by the mean vector, covariance matrix, and weight from all components. So, we can represent it in a compact notation as follows:

$$\lambda = (p_i, \vec{\mu}_i, \Sigma_i) \quad i = 1, 2, \dots, M \quad (7)$$

### 3.3. Maximum Likelihood Estimation

Given a set of training samples  $X$ , the most popular method to train a GMM is maximum likelihood estimation. The likelihood of a GMM can be defined as:

$$p(X | \lambda) = \prod_{t=1}^T p(\vec{x}_t | \lambda) \quad (8)$$

Maximum likelihood parameters are normally estimated using the expectation maximization (EM) algorithm. Among a set of speakers characterized by parameters  $\lambda_1, \lambda_2, \dots, \lambda_n$ , a GMM system makes its prediction by returning the speaker that maximizes the *a posteriori* probability given an utterance  $X$  as follows:

$$\hat{s} = \arg \max_{1 \leq k \leq n} P(X | \lambda_k) = \frac{P(X | \lambda_k) P(\lambda_k)}{P(X)} \quad (9)$$

If prior probabilities of all speakers are equal,  $P(\lambda_k) = \frac{1}{n}, \forall k$ , since  $P(X)$  is the same for all speakers, we can rewrite Equation (10) as follows:

$$\hat{s} = \arg \max_{1 \leq k \leq n} \sum_{t=1}^T \log p(\vec{x}_t | \lambda_k) \quad (10)$$

## 4. RESULTS AND DISCUSSION

In this section, we will present the experimental setup and Quranic audio database, experiment with training samples, experiment with testing samples, experiment with unknown samples, and recognition rate evaluation. We carefully randomized the audio database so that the proposed system is not prone to the recited verses but to the reciter voice only.

### 4.1. Experimental Setup and Quranic Audio Database

A high performance system was used for processing, i.e. a multicore system with Intel Core i7 6700 K 4.00 GHz (4 cores with 8 threads), 32 GBytes RAM, 256 GBytes SSD and 2 TBytes hard disk, installed with the latest version of Windows 10 64-bits operating system and Matlab 2017b with Signal Processing and Neural Network Toolboxes.

The audio database in this research were downloaded in form of MP3 from the internet or originally from CD audio for five reciters, i.e. Abdul Basit (Reciter A), Abdurahman As-Sudais (Reciter B), Saud Ash-Shuraym (Reciter C), Sheikh Ali Abdurahman (Reciter D) and Sheikh Said al-Ghamdi (Reciter E). Using Audacity software, the audio files were converted from MP3 to WAV files with 8000Hz sampling frequency, mono, and were cut into duration of 60 seconds for each sample. The database is divided into two

parts, including testing and training samples. The training samples starting from samples 01 until 10, while testing samples from number 11 until 15. Table 1 shows the Quranic audio database for five reciters.

Table 1. Quranic Audio Database for Five Reciters

Samples	Reciter A	Reciter B	Reciter C	Reciter D	Reciter E
01	Al-Fatihah	Al-Munafiqun	Al-Baqarah	An-Naba'	Al-Fatihah
02	Al-Baqarah	At-Talaq	Al-Fajr	An-Nazi'at	Al-Lail
03	An-Nisa'	At-Tahrim	Al-Ghasyiyah	'Abasa	At-Takwir
04	Al-Imran	Al-Mursalat	Al-Buruj	At-Takwir	Al-Infitar
05	Al-Ma'idah	Al-Insan	Al-Insyiqaq	Al-Infitar	Al-Mutaffifin
06	Al-An'am	Al-Muddaththir	Al-Infitar	Al-Mumtahanah	Al-Insyiqaq
07	Ar-Ra'd	Al-Muzzamil	'Abasa	Al-Insyiqaq	Al-Buruj
08	Ibrahim	Jin	An-Nazi'at	Al-Buruj	At-Tariq
09	Al-Hijr	Nuh	Nuh	At-Tariq	Nuh
10	Maryam	Al-Haqqah	Al-Haqqah	Al-A'la	'Abasa
11	Al-Ghashiyah	Al-Qalam	Al-Qalam	Al-Ghashiyah	Al-Muzammil
12	Al-Kahfi	At-Taghabun	At-Tahrim	Al-Fajr	Al-Qiyamah
13	Yasin	Al-Jumu'ah	At-Talaq	Al-Balad	Al-Insan
14	Al-Mulk	As-Saff	At-Taghabun	Ad-Dhuha	An-Naba'
15	Yunus	Al-Mulk	Al-Mulk	Ash-Shams	An-Nazi'at

#### 4.2. Experiment with the Training Samples

In this experiment, the samples 01 until 10 for each reciter were used to train the GMM. The same samples were then used to evaluate the recognition rate of the trained GMM. The log-likelihood was calculated for each reciter and each training samples, in which the highest likelihood as selected as the recognized reciter. Table 2 shows the recognition rate of the training samples. It shows that each samples for each reciters were identified correctly. It means that the recognition rate for the training phase was 100%.

Table 2. Recognition Rate of Training Samples

Reciter	Samples									
	01	02	03	04	05	06	07	08	09	10
A	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
B	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
C	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
D	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
E	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

\*Y = matched, N= unmatched

#### 4.3. Experiment with the Testing Samples

In this experiment, the previous trained GMM in section 4.2 was used to test different samples. The samples 11 until 15 were used to evaluate the recognition rate of the trained GMM when tested with different samples from the same reciters. The log-likelihood was calculated for each reciter and each testing samples, in which the highest likelihood as selected as the recognized reciter. Table 3 shows the recognition rate of the testing samples. It shows that each samples for each reciters were identified correctly. It means that the recognition rate for the testing phase was 100% as well.

Table 3. Recognition Rate of Testing Samples

Reciter	Samples				
	11	12	13	14	15
A	Y	Y	Y	Y	Y
B	Y	Y	Y	Y	Y
C	Y	Y	Y	Y	Y
D	Y	Y	Y	Y	Y
E	Y	Y	Y	Y	Y

\*Y = matched, N= unmatched

#### 4.4. Experiment with Unknown Samples

The last experiment is conducted to evaluate the proposed system whether it can detect unknown speaker which is not in the database. The database are collected for five reciter only which are Abdul Basit,

Abdurahman As-Sudais, Saud Ash-Shuraym, Sheikh Ali Abdurahman and Sheikh Said al-Ghamdi. For this purpose, one unknown reciter named Fatih is tested using the trained GMM and the training samples. The maximum estimation log-likelihood of unknown speaker does not matched with any parameter of the reciter in the database. Because of that, the result shows that Fatih is recognized as an unknown speaker as shown in Figure 4.

Table 4. Result of testing with unknown reciter

Speaker	Samples									
	01	02	03	04	05	06	07	08	09	10
A	N	N	N	N	N	N	N	N	N	N
B	N	N	N	N	N	N	N	N	N	N
C	N	N	N	N	N	N	N	N	N	N
D	N	N	N	N	N	N	N	N	N	N
E	N	N	N	N	N	N	N	N	N	N

\*Y = matched, N= unmatched

#### 4.5. Recognition Rate Evaluation

It can be concluded that, three of the experiments are successfully conducted. This results showed that the proposed system was able to verify and identify the trained reciters. Although the reciter was randomly reciting different Surah in each samples, but the system still can recognize the pattern of the reciter's recitation. Furthermore, the proposed system was also able to reject the unknown reciter tested. Table 5 show the recognition rate for each experiments. This result is better than the result reported in [4], in which they obtained around 91% accuracy using MFCC and ANN. Better results achieved in this paper could be due to the use of GMM instead of ANN, and the randomize Quranic audio database so that the proposed system is not prone to the recited Quranic verses but able to distinguish the characteristics of individual reciter.

Table 5. Performance of the system with different samples

Experiment	Recognition Rate
Testing with training samples	100% match, 0% mismatch
Testing with testing samples	100% match, 0% mismatch
Testing with unknown reciter	0% match, 100% mismatch

## 5. CONCLUSIONS AND FUTURE WORKS

This paper has presented the development of Quranic reciter identification system using MFCC and GMM. MFCC was selected as the features, while GMM is selected as the classifier. First, we build a Quranic audio database from five reciters, in which they recite different surah from Al-Quran. Altogether, there are 15 samples collected for each reciter, in which 10 samples were used to train the GMM and 5 samples were used for testing. Furthermore, we use another unknown reciter to evaluate the performance of the proposed system. Results showed that our proposed system achieved 100% accuracy in the training and testing phase. The unknown samples were also achieved 100% rejection rate. Further research includes variation of shorter utterance of the recited Quranic verses, different reciters, different features, or different classifier.

## ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Malaysian Ministry of Higher Education (MOHE), which has provided funding for the research through the Fundamental Research Grant Scheme, FRGS15-194-0435.

## REFERENCES

- [1] B. Lawrence, *The Quran: A Biography (A Book that Shook the World)*, Atlantic Books Ltd, 2014.
- [2] T. Economist, "The Bible vs the Koran: The battle of the books," [<http://www.economist.com/node/10311317>], Retrieved on: 13 Nov 2017.
- [3] I. Alsmadi and M. Zarour, "Online integrity and authentication checking for Quran electronic versions," *Applied Computing and Informatics*, vol. 13, pp. 38-46, 2017.

- [4] T. M. H. Asda, T. S. Gunawan, M. Kartiwi, and H. Mansor, "Development of Quran Reciter Identification System Using MFCC and Neural Network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 1, pp. 168-175, 2016.
- [5] H. Beigi, *Fundamentals of speaker recognition*, Springer Science & Business Media, 2011.
- [6] P. Bansal, S. A. Imam, and R. Bharti, "Speaker recognition using MFCC, shifted MFCC with vector quantization and fuzzy," in *Soft Computing Techniques and Implementations (ICSCIT)*, 2015 International Conference on, pp. 41-44, 2015.
- [7] X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson, and S. Shamma, "Linear versus mel frequency cepstral coefficients for speaker recognition," in *Automatic Speech Recognition and Understanding (ASRU)*, 2011 IEEE Workshop on, pp. 559-564, 2011.
- [8] V. Tiwari, "MFCC and its applications in speaker recognition," *International journal on emerging technologies*, vol. 1, pp. 19-22, 2010.
- [9] A. Poritz, "Linear predictive hidden Markov models and the speech signal," in *Acoustics, Speech, and Signal Processing*, IEEE International Conference on ICASSP'82., pp. 1291-1294, 1982.
- [10] L. Li, D. Wang, C. Zhang, and T. F. Zheng, "Improving short utterance speaker recognition by modeling speech unit classes," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 24, pp. 1129-1139, 2016.
- [11] F. Richardson, D. Reynolds, and N. Dehak, "Deep neural network approaches to speaker and language recognition," *IEEE Signal Processing Letters*, vol. 22, pp. 1671-1675, 2015.
- [12] T. S. Gunawan and M. Kartiwi, "On the Characteristics of Various Quranic Recitation for Lossless Audio Coding Application," in *Computer and Communication Engineering (ICCCE)*, 2016 International Conference on, pp. 121-125, 2016.
- [13] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 788-798, 2011.
- [14] J. P. Campbell, "Speaker recognition: A tutorial," *Proceedings of the IEEE*, vol. 85, pp. 1437-1462, 1997

## BIOGRAPHIES OF AUTHORS



**Teddy Surya Gunawan** received his BEng degree in Electrical Engineering with cum laude award from Institut Teknologi Bandung (ITB), Indonesia in 1998. He obtained his M.Eng degree in 2001 from the School of Computer Engineering at Nanyang Technological University, Singapore, and PhD degree in 2007 from the School of Electrical Engineering and Telecommunications, The University of New South Wales, Australia. His research interests are in speech and audio processing, biomedical signal processing and instrumentation, image and video processing, and parallel computing. He is currently an IEEE Senior Member (since 2012), was chairman of IEEE Instrumentation and Measurement Society – Malaysia Section (2013 and 2014), Associate Professor (since 2012), Head of Department (2015-2016) at Department of Electrical and Computer Engineering, and Head of Programme Accreditation and Quality Assurance for Faculty of Engineering (since 2017), International Islamic University Malaysia. He is Chartered Engineer (IET, UK) and Insinyur Profesional Madya (PII, Indonesia) since 2016.



**Nur Atikah Muhamat Saleh** received her BEng degree in Communication Engineering in 2017 from the Department of Electrical and Computer Engineering, International Islamic University Malaysia. Her research interests include speech processing and signal processing. Currently, she is pursuing Master of Engineering in Communication at the same university working on speech emotion recognition.



**Mira Kartiwi** completed her studies at the University of Wollongong, Australia resulting in the following degrees being conferred: Bachelor of Commerce in Business Information Systems, Master in Information Systems in 2001 and her Doctor of Philosophy in 2009. She is currently an Associate Professor in Department of Information Systems, Kuliyyah of Information and Communication Technology, International Islamic University Malaysia. Her research interests include electronic commerce, data mining, e-health and mobile applications development.