# Performance Measurement Of *Mel Frequency Ceptral Coefficient* (MFCC) Method In Learning System Of Al- Qur'an Based In *Nagham* Pattern Recognition

# Performance Measurement Of *Mel Frequency CeptralCoefficient* (MFCC) Method In Learning System Of AlQur'anBasedIn*Nagham*PatternRecognition

**Yesy Afrillia[1], Herman Mawengkang[2], [3]Marwan Ramli, [4]Fadlisyahand [5]RizkyPutra Fhonna**

[1]Department of Informatics Engineering, Universitas Sumatera Utara, Medan, 20155.
[2]Department of Mathematics, Universitas Sumatera Utara, Medan, 20155.
[3]Department of Mathematics, Universitas Syiah Kuala, Aceh.
[4]Department of Informatics Engineering, Universitas Malikussaleh, Aceh.
[5]Department of Informatics Engineering, Universitas Sumatera Utara, Medan, 20155.

Email : [1]yesy.afrillia@gmail.com, [2]ramlimarwan@gmail.com,[5]rizkyputrafhonna@gmail.com.

**Abstract**. Most of research have used signal and speech processing in order to recognize*makhraj* pattern and*tajwid* reading in Al-Quran by exploring the *mel frequency ceptralcoefficient* (MFCC). However, to our knowledge so far there is no research has been conductedto recognize the chanting of Al-Quran verse using MFCC. This term is also well-known as*nagham* Al-Quran. The characteristics of *nagham* Al-Quran pattern is much more complexthen *makhraj* and *tajwid* pattern. In *nagham* the wave of the sound has more variation whichimplies the level of noice is much higher and has sound duration longer. The data testing inthis research was taken term by real-time recording. The evaluation measurement in the systemperformance of *nagham* Al-Quran pattern is based on true and false detection parameter withaccuracy 80%. To measure this accuracy it is necessary to modify the MFCC or to give moredata learning process with more variation.
**Key words** : speech recognition, *nagham* Al-Quran pattern, *mel frequency ceptral coefficient* (MFCC).

## 1. Introduction

Speaker awareness, from time to time referred to as speaker biometrics, comprises identity, verification (authentication), different types, and by extension, segmentation, Tracking and detection of speakers. It is a generic term used for any procedure which Involves knowledge of the identity or pattern recognition of a based on voice [1]. Biometrics technology is a self-recognition technique using body parts or human behavior. This technology fulfills two important functions of identification and verification. The identification system aims to solve one's identity. While the verification system aims to refuse or accept identity that someone claimed [2]. A speaker recognition system first attemptsto model the vocal tract traits of a person. This may be a mathematical model of the physiologicalsystem generating the human speech.However, the growing technology of many researchers build amore interesting and specific model on voice recognition systems [1].

Research on voice recognition has been largely done with various methods of signal processing, such as *Linier Prediction Coding* (*LPC*), *Mel Frequency Cepstrum Coefficients* (*MFCC*), *Neural PredictiveCoding* (*NPC*), etc. Which all of the above methods are based on *fourier* transform as a signalcharacteristic extraction method [3]. Feature and value extraction methods that have a direct effect inthe sound system [4]. The pattern recognition rate be effected by the method of extraction andclassification used. So far the method of extraction of Mel-Frequency Cepstral Coefficients or often called MFCC is more commonly used for speech recognition, as has been done by [5][6][7]. MFCC

uses discrete cosine transform (DCT) to calculate coefficients. by means of *mel filter* which is based on the human hearing system, MFCC Optimized for speech recognition.

## 2. Research purposes

The aim of this research isgive information how performance *mel frequecy ceptral coefficient*(MFCC) on the memorization system Qur'an based on pattern recognition *nagham* pattern to simplify *tahfidz* in memorization Al Qur'an to be more efficient so that minimizes the length of time learning inmemorizing the Qur'an.

## 3. Speech

Speech is useful for communication between human beings. Where each sound has the characteristics and levels of different frequency. Sound is a unique thing and have *range* specific frequency and the intensity of the sound which can and can't be heard by humans. The units of intensity to measure the intensity of the sound that is *desibel (dB)* taken from the name of the inventor, namely *Alexander Graham Bell.* Known as the inventor of the phone, While the unit of the sound frequency is *Hertz*, Taken from the name of a Physicist, *Heinrich Rudolf Hertz* to appreciate his contribution in the electromagnetic field [1].

## 4. Speech Recognition

Speech recognition is a process to recognize letters, words or sentences that are spoken. Speech recognition is better known by the term *Automatic SpeechRecognition* or *Computer SpeechRecognition* Where the use of a machine/computer for Recognize a person's voice or identificationfrom the spoken voice. Generally speaker talking in front of computer/machine then thecomputer/machine recognizes a person's voice exactly as it is spoken. Speech recognitionrecognizable into various task levels, recognition in the level of acoustic signal in the form of a level
test in the composition of sub-word units in the form of phonemes, words, phrases and sentences [8].

## 5. Feature Extraction

Speech feature extraction is tasked with changing voice signals to the coefficient vector feature that contains only the information needed for identity of the given phrase. Because speech has unique and different attributes contained in the spoken words of this attribute can be extracted from a lot of feature extraction recommendations and can be used for speech recognition tasks[9].
The purpose of feature extraction is to minimize the size of the data without change the characteristicsof the voice signal in each frame that can be used as a characteristic. Feature extraction is obtainedfrom converting the form of voice signal into parameterized representation [1].

## 6. *Mel Frequency Ceptral Coefficient* (MFCC)

The MFCC characteristic extraction process is a feature-taking process based on *discrete fouriertransforms*. MFCC is the most generally used method for the speech processing, because it isconsidered good enough to present feature and signal. The workings of MFCC in capturing sound and
frequency signals exactly as humans recognize sound characteristics [10]. The MFCC process diagramis shown in the following figure :

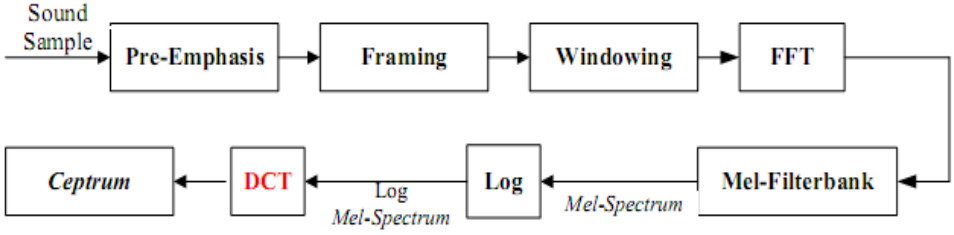


**Figure 1.** Diagram Process MFCC

### 6.1. Pre-Emphasis

Sound signal filtering process required recording or sampling process. The purpose of filtering is to get a smoother spectral form of the sound frequency. Filter pre-emphasis Is based on the input/output relation in the time-domain, expressed in the equation [1] :

$$y[n] = s[n] - \propto s[n-1], 0.9 \leq \propto \leq 1.0$$

Where :
y [n] = result signal pre-emphasize filter
s [n] = before signal pre-emphasize filter

### 6.2. Framing

Sound signals are non-stationary signals, meaning their statistical properties are always changing with time. So it is not possible to extract the spectral features of speech at once. Therefore, the spectral character of sound signals is extracted through a window as a sound signal characterizing a particular sound part so that it can be made an assumption that the sound signals is stationary [1].

### 6.3. Windowing

The process of windowing is the process of filtering each frame by multiplying each frame with certain windows functions the same size as the frame. Windowing is also used to ensure continuous sound from the initial frame to the end [11]. A good window function should be narrowed to *the mainlobe*andextendstoits*side-lobe*.

### 6.4. Fast Fourier Transform (FFT)

The purpose of *Fast fourier transformation* (FFT) is to decompose the signal into a *sinusoidal* signal, Which consists of two units, namely real units and imaginary units. FFT Used for frequency analysis, thus simplifying voice processing as it corresponds to human hearing. FFT Is an algorithm that implements *discrete fouries transform* (DFT). DFT is the transformation of each *frame* with *N sample* From time domain to frequency domain.

$$X_n = \sum_{k=0}^{N-1} X_k \, e^{-2\pi j \, k \, n/N}$$

Defined :
N= The number of frame segments
X= Data value to *k*
N = 0,1,2,3,…, N-1 and k=0,1,2,3,…, N-1
k
j =-1

### 6.5. *Mel Frequency Wrapping*

*Mel Frequency Wrapping* is commonly carried out by using *filterbank*. *Filterbank* is one form of filter made with purpose to determine the energy size of a particular frequency-band in frequency, but for MFCC purposes, the *filterbank* should be applied in the frequency domain. *Filterbank* using convolution representation in doing filter to signal. *Convolution* can be done by multiplication between signal spectrum with filter bank coefficient.

### 6.6. Log

*Logarithm* values are obtained by converting DFT values into one value. Reduce the value of the mel filter bank by substituting each base log value. By using the *matlab* command "*log*" to extract the logarithm value of the mel-filter on the sound segment. human ear is less delicate in hearing sounds that have low amplitude and height [12].

### 6.7. Discrete Cosinus Transform (DCT)

DCT is the last step of the main process of MFCC feature extraction. The basic concept of DCT is correlating value of *mel spectrum* so as to produce a good representation of *property spectral local*. Basically the concept of DCT is the same as inverse *fourier transform*.

### 6.8. Ceptrum

*Ceptrum* is the inverse value of *spectrum.* Cepstrum Commonly used to obtain information from a voice signal spoken by humans. In this last step, *spectrum log mel* converted to ceptrums use Discrete Cosine Transform (DCT) Which is the value of the resulting mel frequency being converted into a time domain.

## 7. *Nagham*

*Nagham* is a word derived from Arabic which means song/rhythm. The popular term Nagham comes from the *qori* and *qoriah* of Egypt was teached his knowledge in Indonesia in the year 1973. Which mean chanting the Qur'an or make sounds in reading the Qur'an. Nagham then known as the Art of Read Al-Quran. Arts read Al-Quran is a science of sound art that apply the tone, tempo and rhythm inthe recitation of the Qur'an.

## 8. Results and Discussion

The scheme of the nagham al-qur'an pattern recognition system developed in this examine is illustrated in the following figure:

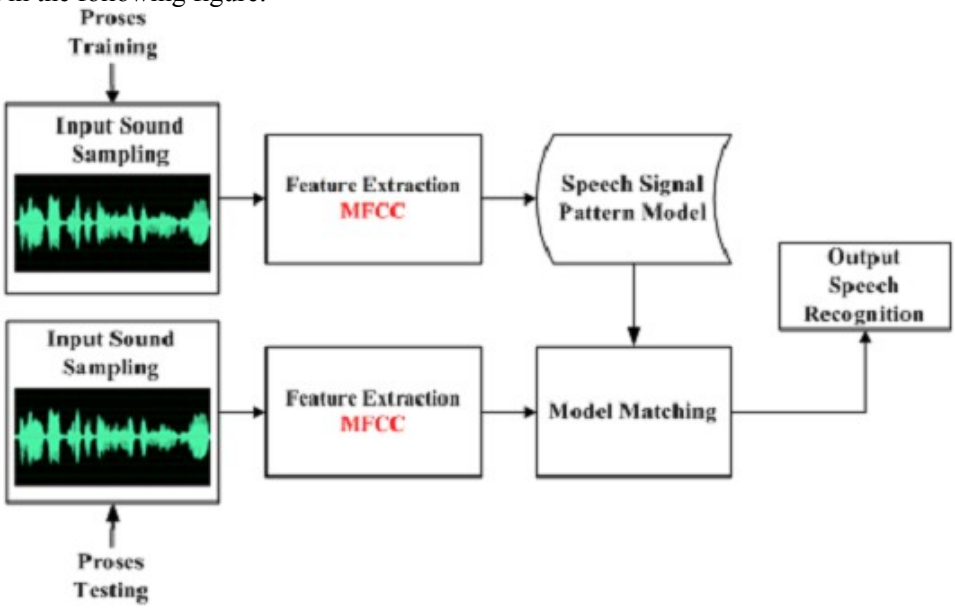


**Figure 2.** The performance scheme of the al-qur'an memorizing systemon the introduction of the nagham pattern**4**

There are two steps performed on al-qur'an memorizing system on nagham pattern recognition that is training process and testing process. In the training process, after the system receives voice input nagham al-qur'an, the next step is feature extraction using mel-frequency cepstral coefficient (MFCC), next model sound signal pattern will be stored in the database. In the testing process through the same

steps as the training process, but the sound will be processed taking voice in real-time then theprocess of matching the model that has been stored in the previous database in the training process to get the recognition of the sound pattern of *nagham* Al-Quran.
The sample test used in this research amounted to 10 sound an example of the pattern of the Qur'an *nagham* on the Al-Fatihah letter is an example of the characteristics of the 7 samples of the *nagham* pattern. The evaluation measurement in the system performance of *nagham* Al-Quran pattern is based on *true* and *false* detection parameter. *Detection rate* is the ratio between the number of successful *nagham* pattern recognition. While *false positive rate* Is the number of unrecognized sounds.
The evaluation measurement in the system performance of *nagham* Al-Quran pattern is based on true and false detection parameter with accuracy 80%. Table 1. illustrates some performance measurement results of the MFCC method in the al-qur'an recitation system based on the recognition of the nagham pattern.

**Table 1.** The results of the MFCC method show on therecitation system of al-qur'an based on the recognition of the nagham pattern

| Data Speech Sampling Training | Data Speech Sampling Testing | Detection True | *False positive rate* | *Detection rate* |
|---|---|---|---|---|
| 20 | 100 | 20 | 0.8 | 0.2 |
| 40 | 100 | 50 | 0.5 | 0.5 |
| 60 | 100 | 70 | 0.3 | 0.7 |
| 80 | 100 | 80 | 0.2 | 0.8 |

The results of the system evaluation show that the detection rate increase is intently associated to the expanding number of training. Figure 3 shows a graph of the results of the recitation system of AlQur'anbasedontherecognitionof the *nagham* pattern.

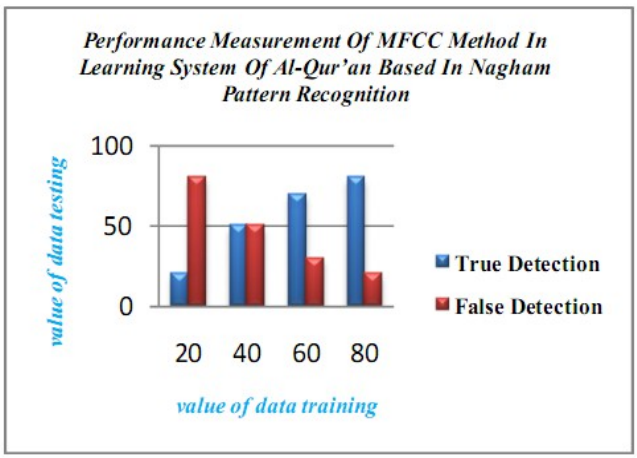


**Figure 3.** Graph of MFCC performance

Based on figure 3 the graph illustrates the measurement of performance performed after 2 samples of trained *nagham* Al-Quran pattern for the overall sound characteristics of the *nagham* pattern using the MFCC method. Amount of test sounds 10 samples of voice pattern *nagham* Al-Quran, at the beginning of the testing phase is done on a system that has been trained 2 samples of *nagham* pattern using MFCC method and obtained detection rate value = 0.2 or 20%. In the second step, the test is

performed on a system that has trained 4 sound samples of *nagham* Al-Qur'an pattern and obtained detection rate value = 0.5 or 50%. In the third step, testing is done on a system that has trained 6 soundsamples of *nagham* Al-Qur'an pattern and obtained detection rate value = 0.7 or 70%. And the last stepof testing accomplished on a system that has been trained 8 sound samples pattern *nagham* Al-Qur'anand obtained detection rate value = 0.8 or 80%.

## 9. Conclusion and Future Work

In this research try to build voice recognition system based on *nagham* Al-Qur'an pattern using MFCCBased on signal processing techniques and sound. MFCC feature extraction applied in this research tocalculate a significant coefficient value for each input data in the form of sound. The results of thesystem evaluation show that the detection rate increase is closely related to the increasing number oftraining. Based on the detection rate of *nagham* pattern recognition on the recitation system of theQur'an using MFCC the accuracy rate reaches 80%. The future of this journal can still be developed that is by modifying the MFCC algorithm to optimizethe coefficient value in covering the weakness for sound input that has a more variable sound waveswith high noise level. In addition, other techniques for pattern matching should be attempted forsubsequent research.

## 10. Reference

[1] Beigi, H. 2011. *Fundamentals Of Speaker Recognition*. Spinger : New York.
[2] Putra, D. 2010. *Pengolahan Citra Digital.* Andi. Yogyakarta.
[3] I.Daubechies. 1988. *Ortonormal bases of Compactly Supported Wavelet*. Comm.Pure Apple. Math. PP : 909- 996.
[4] Al-Irhaim, Y.F. & Saeed, E.G. 2010. *Arabic Word Recognition Using Wavelet Neural Network*. Third Scientific Conference in Information Technology. PP: 416-420.
[5] Martinez, J., Perez, H., Escamilla, E. & Suzuki, M.M. 2012. *Speaker Recognition Using Mel-Frequency Ceptral Coefficient (MFCC) and Vector Quantization (VQ) Techniques*.International Confernce on Electrical Communication and Computers Vol.10. PP : 248251.
[6] Ismail, A., Idris, M.Y.I., Noor, N.M., Razak, Z. & Yusof, Z. 2014. *MFCC-VQ Approach forQalqalah Tajweed Rule Checking*. Malaysian Journal Of Computer Science Vol. 27 (4). PP: 275-293.
[7] Logan, B. 2000. *Mel Frequency Cepstral Coefficient for Music Modeling*. InternationalSymposium in Music Information Ritival.
[8] Tavanaei, A., Manzuri, M.T. & Sameti, H. 2011. *Mel-Scaled Discrete Wavelet Transform andDynamic Features for the Persian Phoneme Recognition*. Internasional Symposium onArtifical Intelligance and Signal Processing (AISP) (2). PP : 138-140.
[9] Kamarudin, N. Nematollahi, M.A. & Abd Rauf bin Hasan. 2014. *Feature Extraction UsingSpectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent AutomaticIdentification*. IEEE Student Conference on Research and Development (SCOReD).
[10] Abro, B., Naqvi, A.B. & Hussain, A. 2012. *Qur'an Recognition for the Purpose of Memorisation Using Speech Recognition Technique*. International Multitopic Conference (INIMIC).
[11] Arshad, N.W., Sukri, S.M., Muhammad, L.N., Ahmad, H., Hamid, R., Naim, F. & Naharuddin, N.Z.A. 2013. *Makhraj Recognition For Al Quran Recitation using MFCC*. InternationalJournal Of Intelligent Information Processing (IJIIP) vol.4 (2). PP : 45-53.
[12] Ismail, A., Idris, M.Y.I., Noor, N.M., Razak, Z. & Yusof, Z. 2014. *MFCC-VQ Approach for Qalqalah Tajweed Rule Checking*. Malaysian Journal Of Computer Science Vol. 27 (4). PP: 275-293.