Available Online : https://proceeding.researchsynergypress.com/index.php/cset/index
RSF Conference Series: Engineering and Technology
ISSN XXXX-XXXX (Online) | XXXX-XXXX (Print)
Volume 1 Number 1 (2021): 453-478

The Implementation Of Mfcc Feature Extraction And Selection of Cepstral Coefficient for Qur'an Recitation in TPA (Qur'an Learning Center) Nurul Huda Plus Purbayan

Heriyanto[1], Herlina Jayadianti[2], Juwairiah[3]
[1,2,3]Informatics Department, Universitas Pembangunan Nasional "Veteran" Yogyakarta, Indonesia

Abstract
There are two approaches to Qur'an recitation, namely *talaqqi* and *qira'ati*. Both approaches use the science of recitation containing knowledge of the rules and procedures for reading the Qur'an properly. *Talaqqi* requires the teacher and students to sit facing each other while *qira'ati* is the recitation of the Qur'an with rhythms and tones. Many studies have developed an automatic speech recognition system for Qur'an recitation to help the learning process. Feature extraction model using Mel Frequency Cepstral Coefficient (MFCC) and Linear Predictive Code (LPC). The MFCC method has an accuracy of 50% to 60% while the accuracy of Linear Predictive Code (LPC) is only 45% to 50%, so the non-linear MFCC method has higher accuracy than the linear approach method. The cepstral coefficient feature that is used starts from 0 to 23 or 24 cepstral coefficients. Meanwhile, the frame taken consists of 0 to 10 frames or eleven frames. Voting for 300 recorded voice samples was tested against 200 voice recordings, both male and female voices. The frequency used was 44.100 kHz stereo 16 bit. This study aims to obtain good accuracy by selecting the right feature on the cepstral coefficient using MFCC feature extraction and matching accuracy through the selection of the cepstral coefficient feature with Dominant Weight Normalization (NBD) at TPA Nurul Huda Plus Purbayan. Accuracy results showed that the MFCC method with the selection of the 23rd cepstral coefficient has a higher accuracy rate of 90.2% compared to the others. It can be concluded that the selection of the right features on the 23rd cepstral coefficient affects the accuracy of the voice of Qur'an recitation.

Keywords: Extraction, Feature, Accuracy, Cepstral, Coefficient

INTRODUCTION

Qur'an is studied through two approaches in its recitation, namely *talaqqi* and *qira'ati* (Yuwan & Lestari, 2005). *Talaqqi* is a method that requires teachers and students to sit opposite each other (Bustami et al., 2017). Meanwhile, *qira'ati* method is the Qur'an recitation using rhythms and tones. Both approaches use the science of recitation (*tajweed*) containing the rules and procedures for Qur'an recitation as properly as possible (Zarkasyi, 1995).

Talaqqi and *qira'ati* methods taught by the teacher have differences between one teacher and another. These differences can be categorized in pronunciation (Bethaningtyas, 2017) and speech (Chamidy, 2016). The difference in pronunciation is reflected in the way of pronouncing Arabic letters that are different from Indonesian (Bethaningtyas, 2017). Meanwhile, the difference in speech is in the Indonesian dialect (Chamidy, 2016).

Many studies have developed an automatic speech recognition system for Qur'an recitation to help the Qur'an learning process (Muhammad et al., 2010; Hassan et al., 2007; Yuwan and Lestari, 2005). One of the studies that use a speech recognition system for Qur'an recitation is the Automatic Delimiter Quranic Verse (ADQV) developed by Hassan et al. (2007). ADQV system uses the Sphinx IV framework, rule-based HMM, and speech recognition techniques. The goal of ADQV system is to automatically delimit and snip each verse of an audio file.

Corresponding author:
heriyanto@upnyk.ac.id
DOI: *to be assigned*

Another speech recognition system of Qur'an recitation that has been developed is the E-Hafizh system (Muhammad et al., 2012). E-Hafiz is a voice content matching system that applies a speech recognition technique. The system extracts the audio file of Qur'an recitation to get the expert's voice feature vector, then saves it into the database. Feature extraction with expert's feature vector testing is matched with the voice feature vectors of novice readers who have a close value to at least the voice features by three experts (Muhammad et al., 2010). In addition, the study also examined the phoneme content based on the standard of *qira'ati ashim riwayat hafsh*, and the results showed that there were phonemes that were not included in the ADQV and E-Hafizh training data, namely the phoneme "e" in the recitation of the 11th surah, Hud, verse 41.

The introduction of other Qur'an recitation was developed, utilizing a phonetically rich and balanced corpus (Yuwan & Lestari, 2005). This method recognizes acoustic phenomena that arise when the Qur'an is read well even though it is trained with little training data (Yuwan & Lestari, 2005). Research on Qur'an recitation was also developed using the Viterbi method approach to detect errors and pattern recognition in the Qur'an recitation (Bustami et al., 2017). The samples of voice used for training through segment detection and frame size can affect a high percentage of success (Bustami et al., 2017). In addition, the voice samples become an important reference material in checking the suitability of Qur'an recitation.

The voice samples that are used as references in the pronunciation and speech of Qur'an recitation often have differences or changes. Therefore, changes in the Qur'an recitation can affect changes in translation and interpretation (Subali et al., 2015). Thus, the sample of Qur'an recitation requires repetition of recitation more than once to choose the most appropriate one as a reference for Qur'an recitation. Feature extraction, reference selection, and proper feature selection are needed in this study (Heriyanto et al., 2018). In addition, the problem of the quality of the speech recognition system is also influenced by the length of the frame, the length of the overlap, the number of filter banks and coefficients (Putra, 2011).

Problem Statement

Based on the description in the introduction, the problem statement of this research is how accurate the feature extraction, the selection of the right features, the cepstral coefficient are for checking the suitability of the Qur'an recitation according to *tajweed* implemented at TPA Nurul Huda Plus.

Research Objective

This study aims to obtain more accurate results of MFCC feature extraction and the right feature selection method in checking the suitability of Qur'an recitation according to *tajweed*.

Research Benefit

The benefit of the research is to contribute to science and technology in the field of speech recognition by using MFCC feature extraction and alternative method of selecting the right features so that readers know how to recite the Qur'an according to tajweed, which can be applied at the TPA Nurul Huda Plus Purbayan, Kotagede.

LITERATURE REVIEW

Voice Research of Qur'an Recitation

Research on the speech recognition system for Qur'an recitation has been widely developed as an automatic tool to help in the Qur'an recitation learning. One of the studies on the Qur'an using the Viterbi method stated that the voice samples used for training had a very high impact on the high percentage of success (Bustami et al., 2017). Meanwhile, the percentage of success is also influenced by pattern planting if the sizes are not the same, segment detection, and frame size (Bustami et al., 2017).

Research on the speech recognition system for Qur'an recitation that has been developed by utilizing a phonetically rich and balanced corpus as training data on an acoustic model is very good for supporting a system of speaker-dependent (Yuwan & Lestari, 2005). On the other hand, due to the lack of voice sources of training data, the system does not fully support the introduction of speaker-independent Qur'an recitation.

Another study of the speech recognition system for Qur'an recitation uses the Automatic Delimiter Quranic Verse (ADQV). ADQV has the goal of automatically determining the limits and deduction for each verse (Hassan et al., 2007). This model was developed with a language approach in the form of grammar applied to a limited vocabulary speech recognition system, but it is not suitable to recognize many vocabularies. It is because the connection and termination of recitation in each verse have quite flexible rules. The rules are flexible because they depend on the length of the breath and the rhythm of the reader's recitation. The rules listed to form this language model are becoming more numerous and very complicated.

Another speech recognition system of Qur'an recitation that has been developed is E-Hafizh (Muhammad et al., 2012). E-Hafizh is a voice content matching system and uses speech recognition techniques. A novice reader is classified as being able to recite well by the system if the voice feature vector has close value to at least three experts (Muhammad et al., 2010).

Research on Voice Feature Extraction

The speech recognition method that uses feature extraction is very important because the results of the feature extraction in the form of features significantly affect the results of matching and checking pattern recognition. Research that examines the use of feature extraction methods includes Mel Frequency Ceptrum Coefficients (MFCC) and Linear Predictive Code (LPC) (Abriyono & Harjoko, 2012). Both methods have strengths and weaknesses in feature extraction that produces features.

The weaknesses of MFCC include low frequency, environmental noise, sensitivity, almost similar sound patterns, and classification (Syafria et al., 2014). Meanwhile, the strength of MFCC includes being able to capture voice characteristics that are important in recognition, capturing important information in the voice, producing minimal data without losing information, and replicating human auditory sounds (Manunggal, 2005). Furthermore, feature extraction using MFCC is widely used for speech recognition because it is more precise in various conditions (Chamidy, 2016). Feature extraction using MFCC has an accuracy value between 58-60% (Aibinu et al., 2011b).

The feature extraction method that uses Linear Predictive Code (LPC) has weaknesses including noise, fluctuating speech frequencies, and classification (Irmawan et al., 2014). This method has the advantage of autocorrelation (Abriyono & Harjoko, 2012; Thiang, 2005).

The voice feature extraction research using both MFCC and LPC has the same weaknesses, including noise, almost similar speech frequencies, frequently changing frequencies, and classification. The weaknesses of the two methods are also expressed by Abriyono and Harjoko (Abriyono & Harjoko, 2012) that feature extraction using MFCC and LPC is not suitable for recognizing very large numbers of voices, so classification is needed.

Based on the weaknesses and strengths of the two methods, both feature extraction using MFCC and using LPC, feature extraction using MFCC from the level of accuracy is better than LPC (Abriyono and Harjoko, 2012; Aibinu et al., 2011b; Hidayat et al., 2015). In addition, the LPC method, according to Widodo et al., (Widodo et al., 2016), is more suitable for linear computing, while the human voice is basically nonlinear.

Research on speech recognition using MFCC has been widely carried out in all fields, including being applied in the field of language. Speech recognition research in the Arabic language by Chamidy (Chamidy, 2016) stated that the extraction of Mel Frequency Cepstral Coefficients (MFCC) in the form of features to obtain the suitability value of Indonesian speakers towards native speakers is classified using the Hidden Markov Model (HMM).

MFCC is applied in other language fields, such as Indonesian, by identifying speech signals into vocabulary that produces Phoneme and Syllable Models and segmentation (Suyanto & Hartati, 2013). Similar research was conducted by Suyanto and Putra (Suyanto & Putra, 2014) using Mel Frequency Ceptral Coefficient (MFCC) and Hidden Markov Model (HMM), which are able to recognize phoneme segmentation in Indonesian. A similar study on phonemes was also conducted by Cahyarini et al., (Cahyarini et al., 2013), which are able to identify speech pauses between phonemes.

Another study related to language regarding the introduction of *hijaiyyah* letters by Bethaningtyas (Hertiana Bethaningtyas, 2017) used MFCC by comparing the use of 3, 6, 9 and 12 channels from the training data model and the deviation value. Another study related to *hijaiyah* letters by Heriyanto (Heriyanto, 2015) used the average energy and wave deviation methods as a comparison. Meanwhile, another study related to the *hijaiyyah* letter phoneme by Subali et al., (Subali et al., 2015) using the LPC and DTW methods resulted in the speaker's formant frequency in pronunciation and DTW has the strength in terms of autocorrelation.

Another research on MFCC through modification was carried out by Leon (Leon, 2009) in the windowing section. Other studies have also focused on modified MFCC to produce acoustic signal analysis with the stages of preemphasis, frame blocking, hamming windowing, Fast Fourier Transform, Mel Filterbank, Discrete Cosine Transform (DCT), Delta energy, and delta spectrum (Muda et al., 2010).

Research on Speech recognition

Research on speech recognition using different methods produces different outputs, namely using artificial neural networks (Sanjaya & Salleh, 2014), Hidden Markov Model (HMM) (Chamidy, 2016), and Dynamic Time Wrapping (DTW) (Miftahuddin & Hakim, 2017).

Speech recognition using DTW method is carried out to calculate the distance between two-time series data (Putra et al., 2011). This method has the advantage of being able to calculate the distance between two data vectors with different lengths or to know the value of the smallest distance matching between the voice of novice speakers and expert speakers. DTW, according to Miftahuddin and Hakim (Miftahuddin & Hakim, 2017), is an algorithm as a non-linear sequence alignment, which is used to measure the similarity of a pattern in a data series area with time varies and is more realistic.

DTW has a weakness in terms of accuracy, namely the results are varied (Novianto & Yuliantari, 2017) and still match the level of accuracy of HMM (Chamidy, 2016). Meanwhile, the use of the HMM method, according to Suyanto and Putra (Suyanto & Putra, 2014), has weaknesses in terms of being less resistant or robust.

Another speech recognition method using Neural Network (NN) has advantages in terms of learning systems, knowledge acquisition, classification, and generalization of a pattern (Sanjaya & Salleh, 2014). NN, according to Martyna and Sudaryanto (Martyna & Sudaryanto, 2011), has a weakness in terms of the training process that requires a long time with a large amount of data. The same statement is uttered by Aibinu et al., (Aibinu et al., 2011) who identified pronunciation of number one to nine, and it has a problem when the training process with very large data, which requires a very long processing time as well.

State of the Art

Research on speech recognition of Qur'an recitation using MFCC still has problems in selecting the right features. Reference constraints from voice samples used as training have a very high impact on the high percentage of success (Bustami et al., 2017). Constraints on reference success and feature selection are influenced by pattern planting if the size is not the same, segment detection, and frame size (Bustami et al., 2017).

The researcher's contribution is in the method of selecting the right features with one expert for beginners and one expert for advanced levels, while in previous studies, this process used three experts as a reference.

Contribution of Research

The contribution of the research is in the method of selecting the right features to improve the suitability of Qur'an recitation using only one expert for beginners and one expert for advanced level, and it has been able to increase accuracy, while previous research used three experts as reference features for Qur'an recitation.

RESEARCH METHOD

This research consists of three stages. The first stage is the extraction of recitation voice features using Mel Frequency Cepstral Coefficient (MFCC). The second stage is the selection of features that will be used as a feature table using the proposed model of Dominant Weight Normalization (NBD) with the same threshold, range, filtering, weight normalization, and dominant weight. The third stage is the testing process by checking the suitability of Qur'an recitation and applying it in learning tajweed, starting from the easiest lesson to the hardest one, for Qur'an recitation at TPA Nurul Huda Plus. Testing of checking the suitability of selecting the right features towards the number of cepstral coefficients and frames and getting the ease of tajweed that can be applied in learning Qur'an recitation at TPA Nurul Huda Plus.

The MFCC method was first introduced by Davis and Mermelstein around 1980. MFCC is a method that is quite good in speech recognition (Davis & Mermelstein, 1980). MFCC is the most widely used feature extraction in the field of speaker recognition and speech recognition.

MFCC is a feature extraction that produces features or characteristics that distinguish one another in the form of cepstral coefficient parameters (Abriyono & Harjoko, 2012). Feature extraction of Mel Frequency Cepstral Coefficient (MFCC) converts sound waves into several types of parameters, such as the cepstral coefficient, which represents the audio file (Chamidy, 2016). In addition, MFCC produces feature vectors that convert voice signals into several vectors for speech feature recognition (Putra et al., 2011).

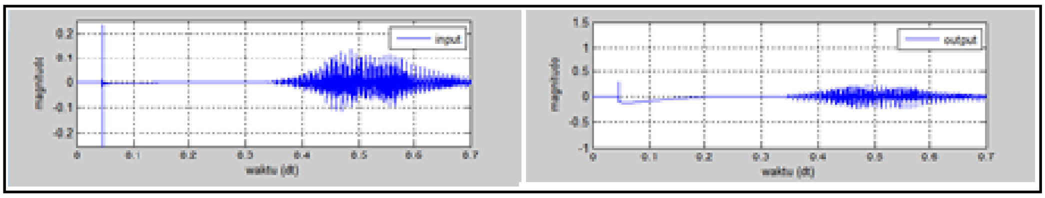
MFCC has some stages, namely pre-emphasis, frame blocking, windowing, Fast Fourier Transform (FFT), Mel Frequency Wrapping (MFW), Discrete Cosine Transform (DCT) and cepstral lifting, which produce parameters as features, namely frames and cepstral coefficient (Putra, 2011).).

Pre-emphasis

Pre-emphasis is the initial process of feature extraction to cause the baseband level at high frequencies to have a good signal quality. The pre-emphasis process, according to Proakis and Manolakis (Proakis & Manolakis, 1996), has a value between 0 to 1 or between $0.9 \le \alpha \le 1.0$ using equation (1).

$$y(n) = s(n) - \alpha\, s(n-1). \qquad (1)$$

In this case, *y(n)* is the pre-emphasis signal, while *s(n)* is the pre-emphasis signal, symbol *n* is the serial number of the signal, α is the pre-emphasis filter constant between 0.9-1.0 and *s* is the signal. The description of the pre-emphasis process is shown in Figure 1. Figure 1 in part (a) is the input sound before pre-emphasis is performed, while Figure 1 part (b) is the output result after processing the pre-emphasis signal.



(a) (b)

Figure 1. *Pre-emphasis* (a) before and (b) after (Putra, 2011)

The signal is taken in *n* in the pre-emphasis with a recitation of one word or two words in one to three seconds.

Frame blocking

A good frame blocking process is taken as long as possible to get the best frequency resolution, and preferably with the shortest possible time in order to get the best time domain process. The number of blocking frames is calculated using equation (2).

$$f_l(n) = y(Ml + n). \qquad (2)$$

In this case, f_l(n) is the result of frame blocking, symbol *n* is 0.1,...N-1. The symbol *N* is the number of samples, *M* is the length of the frame, *l* is 0.1,...L-1. The symbol *L* is the entire signal and *y* is the pre-emphasis result.

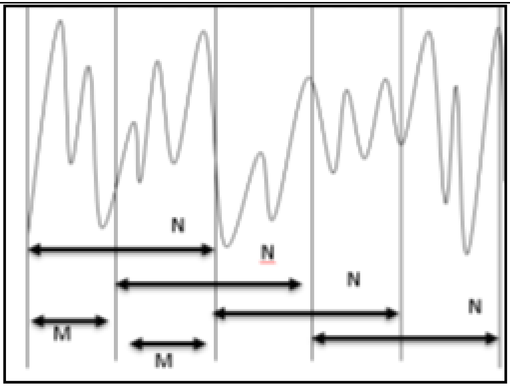


Figure 2. Ilustration of *Frame blocking* (Abriyono & Harjoko, 2013)

Figure 2 shows that *M* is the first frame with the voice signal in the formula symbolized by f_l then *M+M=N*.

Windowing

The process after frame blocking is Windowing. Windowing aims to reduce the effect of discontinuity at the edges of the frame. The frame generated by the Windowing process uses a formula, namely Rectangular Window, Hamming Window, and Hanning Window (Chamidy, 2016). Among the three windowing functions, the Hanning windowing process is better because it is smoother than the others (Putra, 2008). Representation of windowing function uses equation (3).

$$X(n) = f_l(n)w(n). \tag{3}$$

In this case, the function *X(n)* is the windowing signal, where f_l is the frame blocking result, where n is 0.1,...,N-1. The symbol *N* is the number of samples in each frame and *w(n)* is the window function. Meanwhile, Hanning's windowing function uses equation (4)

$$w(n) = 0{,}5\left(1 - \cos\left(\frac{2\pi n}{M-1}\right)\right). \tag{4}$$

In this case, *w(n)* is the window function using hanning, where *n* is 0.1,...,M-1, *M* is the frame length. Figure 3 describes the results of the windowing process using the Hanning window.

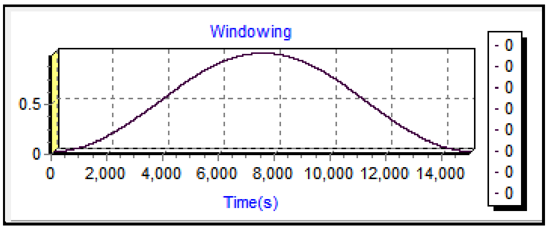


Figure 3. Illustration of *windowing* (Proakis & Manolakis, 1996)

Fast Fourier Transform (FFT)

The Fast Fourier Transform process is the development of the Discrete Fourier Transform (DFT) algorithm which is used to convert signals from digital signals in the time domain to the digital signal domain to the frequency domain (Abriyono & Harjoko, 2012). The principle of FFT converts signal from time to frequency domain.

The computational time calculation process with Discrete Fourier Transform (DFT) takes time and is inefficient so that FFT can perform computational efficiency, as stated by Proakis and Manolakis (Proakis & Manolakis, 1996) that the FFT method is a more efficient to calculate DFT. Discrete Fourier Transform (DFT) is calculated using equation (5).

$$d[m]=\sum_{n=0}^{N-1} X(n)e^{-j\frac{2\pi}{N}nm};m=0,1,2,\ldots,N-1. \quad (5)$$

In this case*, d[k]* is the result of the DFT calculation, the symbol *X(n)* is the windowing result. The symbol *N* is a natural number, *N* is the number of samples to be processed ($N\in$ N). The symbol *k* is a variable frequency discrete value (m=N/2, $m \in$ N). Fast Fourier Transform aims to decompose the signal into a sinusoidal signal in the form of real units and imaginary units. Fast Fourier Transform is calculated using equation (6)

$$T(m)=\sum_{n=0}^{N-1} X(n)\cos\left(\frac{2\pi mn}{N}\right)-\sum_{n=0}^{N-1} X(n)\sin\left(\frac{2\pi mn}{N}\right) \quad (6)$$

In this case, the function *T(m)* is the result of the Fast Fourier Transform calculation in *m*, the symbol *X(n)* is the result of the windowing calculation in *n*. The symbol *n* is the serial number of the signal. The symbol *m* is the index of the frequency (1, 2,…*N*). Figure 4 shows the spectrum results with FFT.

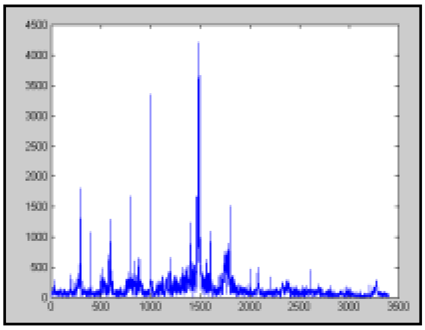


Figure 4. FFT generates spectrum (Kumar, 2013)

Mel Frequency Wrapping (MFW)

MFW is a filterbank on the email scale. This is as stated by Tshilidzi Marwala (Tshilidzi Marwala, 2012) that MFW contains filter banks, which are spaced on the mel scale. Filterbank has a frequency response by going through a triangular path of which distance and magnitude are determined with constant frequency intervals. The process output obtained from the filter is known as the mel spectrum. MFW has the goal of producing a mel spectrum using equation (7).

$$Y[i] = \sum_{j=1}^{G} T[j] H_i[j] \tag{7}$$

In this case, *Y[i]* is the result of the calculation of the frequency wrapping in *i* where G is the number of magnitude spectrum ($G \in$ N), symbol *T[j]* is the result of FFT, *Hi[j]* is the filterbank coefficient at frequency *j* ($1 \leq i \leq E$), and *E* is the number of channels in the filterbank. The approach used in the form of mel uses equation (8).

$$mel(f) = 2595\ log_{10}\left(1 + \frac{f}{700}\right). \tag{8}$$

In this case, mel uses a frequency with a mel scale, *f* as *frequency*. MFW produces a mel spectrum. Mel frequency scale is a linear frequency scale at frequencies below 1,000 Hz and is a logarithmic scale at frequencies above 1,000 Hz (Putra, 2011). Figure 5 shows the mel spectrum process.

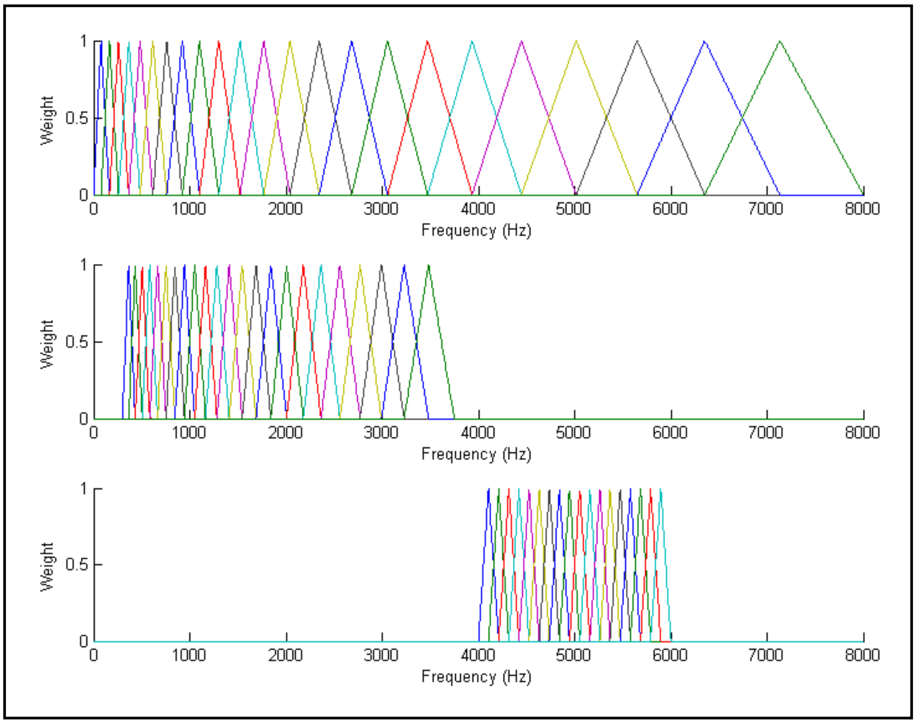


Figure 5. Mel Spectrum (Hidayat et al., 2015; Hassan et al., 2007)

Figure 5 shows the mel scale made from the filter bank using a triangular filter type in color while the weight in this case *db* is the amplitude.

Discrete Cosine Transform (DCT)

The process after the mel scale on the filter bank is DCT. DCT generates mel septrum to improve recognition quality. DCT uses equation (9).

$$C_r = \sum_{k=1}^{K} (\log_{10} Y[i] \cos\left[r(i - \frac{1}{2}) \frac{\pi}{K} \right]; r = 1,2,\ldots,K. \tag{9}$$

In this case, C_m is the *Coefficient*, where *Y[i]* is the output of the filterbank process on the index, *r* is the number of coefficients and *K* is the expected number of coefficients. The DCT process produces a mel septrum. DCT can also be used for image processing.

Cepstral Liftering

The last process in MFCC is ceptral liftering. Cepstral liftering aims to improve the accuracy used to recognize pattern matching, both speaker recognition and speech recognition (Putra, 2011). Cepstral coefficient uses equation (10)

$$w(k) = 1 + \frac{C}{2}\sin\left(\frac{b\pi}{C}\right); b = 1,2,\ldots C \tag{10}$$

In this case, *w(k)* is the window function of the cepstral features, *C* is the cepstral coefficients, the symbol *k* is the index of the cepstral coefficients. Processing of cepstral liftering has results in the form of frames and cepstral coefficients which are then processed into feature selection.

Threshold

Threshold is used to limit the maximum and minimum portion. Calculation of min plus max and divided by two for each part. Threshold is determined based on MFCC results in the form of cepstral coefficient (c)=w(k) parameters and the frame formed produces 1st to 6th thresholds (parameters *b1* to *b6*). Threshold uses equations (11) to (16)

$$b_1 = min((w(k)), \tag{11}$$

$$b_2 = \frac{\min(w(k)) + \left(\frac{min(w(k))+max(w(k))}{2}\right)}{2}, \tag{12}$$

$$b_3 = \frac{min(w(k))+max(w(k))}{2}, \tag{13}$$

$$b_4 = average(w(k)), \tag{14}$$

$$b_5 = \frac{\left(\frac{min(w(k))+max(w(k))}{2}\right)+max(w(k))}{2}, \tag{15}$$

$$b_6 = max\,\left(w(k)\right). \tag{16}$$

In this case, *min* is the minimum threshold, while *max* is the maximum threshold value. The *min* and *max* threshold values are taken from the MFCC features, namely the frame and cepstral coefficient. The threshold is taken based on the research of Sari, et al., (2013) which states that the signal sample has certain threshold values that produce maximum accuracy.

Range Similarity

After the threshold area has determined maximum and minimum value, then the range similarity is sought within the threshold. The range similarity is determined based on a predetermined threshold and with the threshold (parameters *b1* to *b6*), a rule is made with various conditions. The rule checks in terms of the condition if from the cepstral coefficients (c) with the rule if it meets, the weight becomes one or rule if it does not meet then 0 and uses (parameters *r1* to *r7*).

Range Suitability Check
Range $b_1 - b_6$ by checking the condition is in each frame $(i=0,1,\ldots,F)$
using rule r_1-r_7 *Filtering* accumulated with R_{ij} with j=1,2, 3, 5, 6,7 *Weight=1*
1.Condition r_1 is if (b_1) min=$w(k)$ then R_{i1} = weight elseif weight=0
2.Condition r_2 is if $(b_1) \geq w(k)$ and $(w(k) < b_2)$then R_{i2} =weight elseif weight=0
3.Condition r_3 is if $(b_2) \geq w(k)$ and $(w(k) < b_3)$then R_{i3} =weight elseif weight=0
4.Condition r_4 is if $(b_3) \geq w(k)$ and $(w(k) < b_4)$then R_{i4} =weight elseif weight=0
5.Condition r_5 is if $(b_4) \geq w(k)$ and $(w(k) < b_5)$then R_{i5} =weight elseif weight=0
6.Condition r_6 is if $(b_5) \geq w(k)$ and $(w(k) < b_6)$then R_{i6} =weight elseif weight=0
7.Condition r_7 is if (b_6) max=$w(k)$ then R_{i7} =weight elseif weight=0

Figure 7 Range Suitability Check and Filtering

Figure 7 shows the cepstral coefficient (c) or w(k) check with range and filtering.

Filtering

After the range of *r* is obtained, then the selection is carried out. The selection is in the form of filters. Filtering aims to generate weights in each recitation and frame. Filtering, according to Rizal (2014), is a model to represent each frame. The filtering uses rules that are formed r_1, r_2, r_3, r_4, r_5, r_6, and r_7 "if it meets the threshold or does not meet the cepstral coefficient $(c) = w(k)$ according to the rules for the range of Figure 4.5, then it is given weight of one in each frame and recitation" so that they are summed or accumulated. Filtering is done to select by separating or selecting or filtering, so that the filtering results become P_{ij} parameters in the feature table in the form of p_1 to p_7. The sum of the filtering results uses equations (17) and (18).

$$G_j = \sum_{i=0}^{F} p_{ij}, \quad (17)$$

$$U = \sum_{i=1}^{a} G_j. \quad (18)$$

In this case, *U* is the total number of *Gj*. The symbol *Gj* is the result of the number of p_{ij}, overall, the symbol *j* is 1 to 7, while the frame formed in the form of *i* is the 0 frame to the F frame. The symbol p_{ij} is the result of filtering or weight accumulation. Symbol a =1, 2, …, 7.

Eliminating Weight Duplication

The filtering process in the form of weights produces duplicated data, so it is necessary to eliminate the duplication of weights so that each frame and recitation will be different. According to Bender, et al. (1996), duplication occurs because the results of the cepstrum have duplicates. Duplication of weights can be eliminated by looking for similarities and then eliminating weights=0. Determine that $Q_{ij}=p_{ij}$ and check for duplication.

if $Q_{i0}=p_{i1}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i1}=p_{i2}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i2}=p_{i3}$ ***then*** $Q_{ij}=0$, … ***if*** $Q_{i9}=p_{i10}$ ***then*** $Q_{ij}=0$
if $Q_{i0}=p_{i2}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i1}=p_{i3}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i2}=p_{i4}$ ***then*** $Q_{ij}=0$, …
if $Q_{i0}=p_{i3}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i1}=p_{i4}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i2}=p_{i5}$ ***then*** $Q_{ij}=0$, …
if $Q_{i0}=p_{i10}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i1}=p_{i10}$ ***then*** $Q_{ij}=0$, ***if*** $Q_{i2}=p_{i10}$ ***then*** $Q_{ij}=0$,…

The result of eliminating the duplication of weights on the Q_{ij} parameter if it is equal to p_{ij} with the same check is given a weight = 0. Duplicate weights are removed and then the total weights are calculated using equation (19).

$$Z_j = \sum_{i=0}^{F} Q_{ij} \tag{19}$$

In this case, the symbol Z is the calculation of the total weight Q. Weight Q_{ij} is the result of filtering the weights that have been removed from duplication, j is 1 to 7, i is the 0 frame to the F frame.

Weight Normalization

The next process is weight normalization. Weight normalization is a process of equalizing and aligning the weights in a more balanced way so that they are more proportional. The result of weight normalization (S_j) is the *npf* parameter in the feature table. Weight normalization is calculated using equation (20).

$$S_j = \sum_{i=0}^{F} \frac{Q_{ij}}{Z_j} \tag{20}$$

In this case, S is the result of weight normalization. The symbol $_j$ with j = 2, 3, 5 and 6, while Z is the calculation of the number of weights. Weight Q_{ij} is the result of weight filtering that has been removed from duplication. The symbol i is the frame. Frames are taken from $i=0$ to F.

Sequential multiplication

Each weighted result is multiplied sequentially. Sequential multiplication was performed on the feature table of the total weight and dominant weight. This multiplication is used because multiplication has the largest value (Marlina et al., 2017). The result of sequential multiplication in the parameter P_j. Sequential multiplication is calculated using equation (21).

$$P_j = \sum_{i=0}^{F} (R_{ij} * B_{ij} * Z_{ij}^2). \tag{21}$$

In this case, P_j is the result of sequential multiplication, the symbol R_{ij} is the result of the range of the filtering process in the form of accumulated weights. Symbol j is 2, 3, 5 and 6. Symbol B is the normalized dominant weight. Symbol Z is the calculation of the total weight Q, while Q is the result of filtering the weights that have been removed from duplication. Symbol i is the frame from $i=0$ to F.

Pattern Uniformity Suitability (KKP)

The final process of calculating the sequential prose is checking the suitability of the recitation by calculating the KKP. KKP is to approach the results to the feature table whether it is appropriate or not. According to Yuwono and Antonio (2015), more complex signals are used with an average frequency. The results of the KKP calculation using Kj are parameters j=2, 3, 5 and 6. The KKP calculation uses equations (22) and (23).

$$Kj = \frac{\overline{P_j}\,\overline{U}}{\overline{Z_j^2}}, \qquad (22)$$

$$K = \frac{(\overline{P_2} + \overline{P_3} + \overline{P_5} + \overline{P_6})\overline{U}}{\overline{Z_2^2} + \overline{Z_3^2} + \overline{Z_5^2} + \overline{Z_6^2}} \qquad (23)$$

In this case, P_j is the result of sequential product. Symbol U is the sum total of P_2, P_3, P_5 and P_6. Symbol K is the overall pattern suitability (KKP). Symbol Z is the number of weights. Symbols j are 2, 3, 5, and 6.

TESTING

Tests in matching the results of accurate distributions used the average calculation in the form of percentages. Percentage, according to Putra (Darma Putra, 2011), is the correct average calculation that is obtained by dividing the correct number by the total number of matches and multiplying by 100% as well as for the average percentage of errors. The average correct percentage and the average wrong percentage is calculated using equations (11) and (12)

$$h = \frac{g}{o} 100\%. \qquad (21)$$

In this case, h is the true mean, symbol g is the number of correct percentage and o is the total number of matches.

$$A = \frac{q}{o}\, 100\%. \qquad (22)$$

In this case, A is the average of the errors, symbol q is the number of errors, and o is the total number of matches.

Testing and Indicators

Testing and indicators are carried out in the process of checking the suitability of Qur'an recitation with indicators of research achievement results in the form of percentages in Table 1. The tests carried out consist of three stages, namely testing of references, selecting the right feature, the number of cepstral coefficients, and testing *tajweed* by looking for the easiest and the hardest part.

Table 1. Testing and Indicators

No.	Testing	Goals	Indicators
1	The right reference	Looking for the correct average	85-90%
2	Selection of the right features		
	The number of cepstral coefficient	Looking for the correct average	85-90%
3	The easiest to the hardest *tajweed*	Looking for the correct average	75%

.

Research Stage Model

This research framework is used to find solutions to problems in the research. The research problems globally consist of feature extraction, feature selection, and checking with recitation suitability testing.

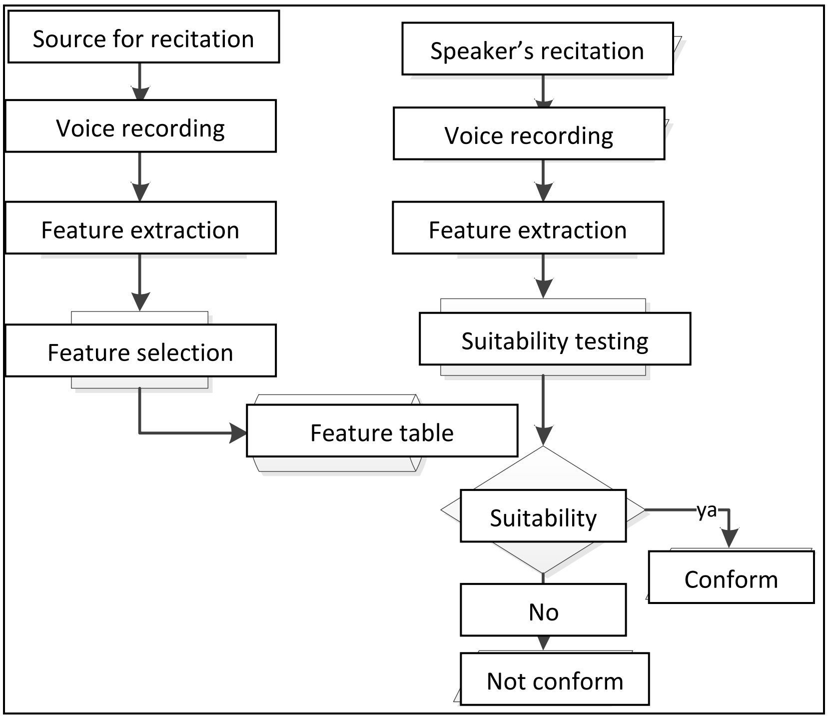


Figure 6. The general model of research stages

Figure 6 shows the general model of the research stages starting from the stage of the feature extraction process and feature selection as well as the stage of testing the suitability of Qur'an recitation. The detailed research was carried out in six stages. Figure 6 describes the six stages of the proposed method. The first stage includes recording voice samples, taking the voice of an expert's reference recitation or the voice of the source as many as eleven letters in the Qur'an. The process of recording Qur'an recitation is done per word or two words. The second stage includes carrying out feature extraction starting from taking the voice recording of Qur'an recitation. The feature extraction process in this study uses the Mel Frequency Cepstral Coefficient (MFCC) method to produce frame features and cepstral coefficients. The MFCC process in this study was carried out by pre-emphasis, frame blocking, windowing using the Hanning Window method, Fast Fourier Transform (FFT), Mel Frequency Wrapping (MFW), Discrete Cosine Transform (DCT), and cepstral liftering.

In the third stage, feature selection is carried out using the Dominant Weight Normalization (NBD) model which consists of determining the threshold, creating ranges, filtering, eliminating weight duplication, and normalizing weights and dominant weights. The feature selection takes the results of the MFCC feature extraction in the form of frames and cepstral coefficients to be processed to NBD, the result of which is a feature table.

RESEARCH RESULTS

Feature Selection

The selection of the proposed features has six stages, namely determining the same threshold, making the same range, filtering, eliminating duplication of weights, and normalizing weights and dominant weights. All these stages are applied in the Dominant Weight Normalization (NBD) feature selection algorithm to generate a feature table.

Table 2. Results of MFCC of "iqra" dan "abbdan" recitation

File Name	Recitation	*Frame*	*Coefficient*	*Cepstral Coefficient (c)*
Barkoni-iqq-ro01.wave	iqrro	0	0	20.5
Barkoni-iqq-ro01.wave	iqrro	1	0	27.8
Barkoni-iqq-ro01.wave	iqrro	10	0	24.3
Barkoni-iqq-ro01.wave	iqrro	0	1	-0.26
Barkoni-iqq-ro01.wave	iqrro	2	1	-5.60
Barkoni-iqq-ro01.wave	iqrro	10	1	-1.14
The number of *frame*, *cepstral coefficient*	264	0-10 (11)	0-23(24)	
Barkoni--Abb-dan.wave	abbdan	0	0	26.7
Barkoni--Abb-dan.wave	abbdan	1	0	25.7
Barkoni--Abb-dan.wave	abbdan	10	0	22.1
Barkoni--Abb-dan.wave	abbdan	0	1	-8.30
Barkoni--Abb-dan.wave	abbdan	1	1	-9.34
Barkoni--Abb-dan.wave	abbdan	2	1	-4.42
Barkoni--Abb-dan.wave	abbdan	10	1	-2.86
The number of *frame*, *cepstral coefficient*	264	0-10 (11)	0-23(24)	

Table 1 shows the results of feature extraction using MFCC for "iqra" and "abbdan" recitation in Al-Alaq with expert voice references. The result of feature extraction is in the form of a frame consisting of eleven frames and twenty-four cepstral coefficients.

Feature Selection Algorithm Using
Dominant Weight Normalization (DWN)

0. Start
1. Take the recitation voice
2. MFCC function with *frame*=f, *cepstral coefficient* ***(c)***=***w(k)***
3. **Determining *threshold* 1 to *threshold* 6 *(b_1-b_6)***
   For $b_1 = min(w(k))$,
   For $b_2 = \frac{\min(w(k)) + \left(\frac{\min(w(k))+\max(w(k))}{2}\right)}{2}$
   For $b_3 = average(w(k))$,
   For $b_4 = \frac{\min(w(k))+\max(w(k))}{2}$
   For $b_5 = \frac{\left(\frac{\min(w(k))+\max(w(k))}{2}\right)+\max(w(k))}{2}$,
   For $b_6 = max(w(k))$.
4. **Making range**
   Range by checking condition in every frame with *cepstral coefficient(c)=w(k)*
   Weight =1
   1.Rule$_1$ is if (b_1) *min=w(k)* then p_{i1} =bobot elseif bobot=0
   2.Rule$_2$ is if $(b_1) \geq = w(k)$ and ($w(k) < b_2$) then p_{i2} =weight elseif weight=0
   3.Rule$_3$ is if $(b_2) \geq = w(k)$ and ($w(k) < b_3$) then p_{i3} =weight elseif weight=0
   4 Rule$_4$ is if $(b_3) \geq = w(k)$ and ($w(k) < b_4$) then p_{i4} =weight elseif weight=0
   5.Rule$_5$ is if $(b_4) \geq = w(k)$ and ($w(k) < b_5$) then p_{i5} =weight elseif weight=0
   6.Rule$_6$ is if $(b_5) \geq = w(k)$ and ($w(k) < b_6$) then p_{i6}=weight elseif weight=0
   7 Rule$_7$ is if (b_6) *max*= *w(k)* then p_{i7} =weight elseif weight=0
5. ***Filtering***
   Result of filtering is from p_{i1} to p_{i7}
   Calculate the number of p_{i1} to p_{i7} using $G_j = \sum_{i=0}^{F} p_{ij}$
   Calculate the total pattern using $U = \sum_{j=0}^{a} G_j$
6. **Eliminating duplication of weights**
   Determine that $Q_{ij}=p_{ij}$ and looking for suitability eliminates duplication.
   if $Q_{i0}=p_{i1}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i2}$ then $Q_{jj}=0$, if $Q_{i2}=p_{i3}$ then $Q_{ij}=0$
   if $Q_{i0}=p_{i2}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i3}$ then $Q_{jj}=0$, if $Q_{i2}=p_{i4}$ then $Q_{ij}=0$
   if $Q_{i0}=p_{i10}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i10}$ then $Q_{jj}=0$, if $Q_{i2}=p_{i10}$ then $Q_{ij}=0$
   calculating the total number using $Z_j = \sum_{i=0}^{F} Q_{ij}$
7. **Weight Normalization**
   Calculating weight normalization using $S_j = \sum_{i=0}^{F} \frac{Q_{ij}}{Z_j}$
8. **Dominant weight**
   Sort the largest value of S_j in feature table (npf$_2$, npf$_3$, npf$_5$, npf$_6$) into variable B_j
9. Save the table of features for calculating the number of weight=*Z*,
   Weight normalization =S_j, the whole pattern=*U*
10. Finish

Figure 7. Feature selection algorithm with DWN (Heriyanto et al., 2021)

Tests for selecting the right features are carried out to find the number of cepstral coefficients with the best accuracy starting from components (c_0 to c_{23}).

Table 3. The results of testing the number of *cepstral coefficient* of surah Al-Fatihah

No	Surah Al-Fatihah	Tajwid	c_5	c_{10}	c_{15}	c_{20}	c_{23}
1	Alaihim	-	71.4	50	57.1	57.1	57.1
2	a-lamin	Mad	57.1	100	100	100	100
3	Alhamdulillah	Mad	85.7	100	100	100	100
4	an-am-ta	-	57.1	100	100	100	100
5	arr-rohman	Mad	71.4	80	85.7	85.7	85.7
6	bi-alai-him	-	71.4	100	100	100	100
7	Bismillah	Mad	100	100	85.7	85.7	85.7
8	ghoi-ril-magh-du	Mad	85.7	100	71.4	100	100
9	hirob-bil	-	71.4	70	85.7	85.7	85.7
10	hirr-rohman	Mad	57.1	100	71.4	71.4	71.4
11	ih-dinasy	-	71.4	50	85.7	85.7	85.7
12	ii-yaa	mad	42.8	100	100	100	100
13	kanak-budu	-	100	87	100	100	100
14	kanas-tain	-	57.1	100	85.7	85.7	85.7
15	la-zii-na	mad	71.4	100	85.7	85.7	85.7
16	maa-lik-ki	mad	42.8	90	71.4	71.4	85.7
17	Mustaqim	mad	71.4	100	100	100	100
18	nirr-rohim	mad	71.4	100	100	100	100
19	syi-ro-tol	mad	14.2	70	85.7	71.4	71.4
20	wa-ii-ya	mad	28.5	100	85.7	85.7	85.7
21	walad-dhoo-lin	mad	42.8	60	85.7	100	100
22	yau-midd-din	mad	100	80	100	100	100
	Rata-rata		65.5	88.04	88.3	89.6	90.2

Table 2 shows the results of the Al-Fatihah testing on the number of cepstral coefficients with components c_5 to c_{23} affecting its accuracy which turned out to be increased at c_{23} to 90.2 better and higher than the other cepstral coefficients.

Checking the Suitability of Qur'an Recitation

Checking the suitability of Qur'an recitation is carried out on the feature table with the calculation of KKP. The calculation of the recitation suitability results is expected to be close to the results in the feature table so that the percentage of recitation suitability increases.

Checks are carried out on feature table that has a threshold with the same range and filtering, sequential calculation and KKP calculation. Checking the recitation is carried out on the selection of the right reference and the selection of the right features.

Recitation Suitability Checking Algorithm

0. Start
1. Take the recording of the speaker
2. MFCC Function with *frame=f, cepstral coefficient(c)=w(k)*
3. **Range checking and filtering of each frame *i=0,1,2,...,F***
   Take a feature table that has a threshold ***(b_1-b_6)***
   Check the range with *w(c)* if *pij* I is fulfilled given a weight =1
   $rule_1$ with if (***b_1*** min = ***w(k)***) then ***R_{i1}*** =weight elseif weight=0
   $rule_2$ with if (***b_1*** >= ***w(k)***) and (***w(k)***<b2) then ***R_{i2}*** =weight elseif weight=0
   $rule_3$ with if (***b_2*** >= ***w(k)***) and (***w(k)***<b3) then ***R_{i3}*** =weight elseif weight=0
   $rule_4$ with if (***b_3*** >= ***w(k)***) and (***w(k)***<b4) then ***R_{i4}*** =weight elseif weight=0
   $rule_5$ with if (***b_4*** >= ***w(k)***) and (***w(k)***<b5) then ***R_{i5}*** =weight elseif weight=0
   $rule_6$ with if **(*b_5*** >= ***w(k)***) and (***w(k)***<b6) then ***R_{i6}*** =weight elseif weight=0
   $rule_7$ with if (***b_6*** = ***w(k)***) then ***R_{i7}*** =weight elseif weight=0
   Filtering store accumulated filtering with R_{ij}, *j*= 2, 3, 5 and 6.
4. **Sequential multiplication with feature table**
   Calculate sequential multiplication of $P_{ij,}$ with *j*=2,3,5 and 6
   $$P_j = \sum_{i=0}^{F}(R_{ij} B_{ij} Z_{ij}^2)$$
5. **Calculate with formula KKP P2,P3, P5 and P6**
   $$K_j = \frac{\overline{P_j}\,\overline{U}}{\overline{Z_j^2}} \qquad K = \frac{(\overline{P_2} + \overline{P_3} + \overline{P_5} + \overline{P_6})\overline{U}}{\overline{Z_2^2} + \overline{Z_3^2} + \overline{Z_5^2} + \overline{Z_6^2}}$$
6. **Finish**

Figure 8. Algorithm for checking the recitation suitability (Heriyanto et al., 2021)

Figure 8 shows the algorithm for checking the suitability of recitation with the feature table starting with the voice of the reader's recitation in the first step, then feature extraction using MFCC produces frame components and cepstral coefficients in the second step. The third step, the reader's recitation checking algorithm begins with MFCC feature extraction which produces frame components and cepstral coefficients.

The results of the study can be seen in the process stages in Figure 9 showing the proposed feature selection stage, checking and testing the suitability of Qur'an recitation. Feature selection is carried out to obtain a feature table starting from MFCC feature extraction, recitation check to recitation test. Checking the suitability of Qur'an recitation is used for testing with the selection of appropriate references and features. Checking the correct reference by repeating the recitation more than once, selecting the most appropriate one with the same range, filtering, sequential multiplication, and KKP calculations.

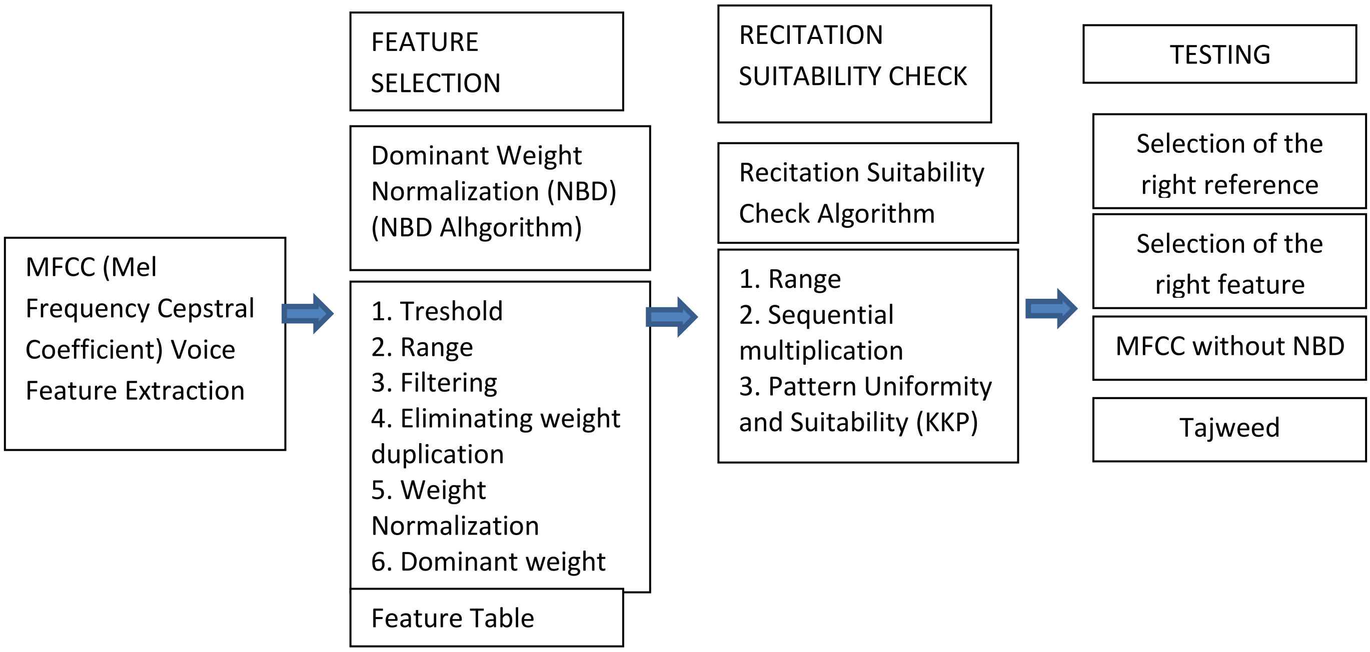


Figure 9. Proposed feature selection, checking and testing

Tests on the selection of the right features are carried out on the number of cepstral coefficients and the number of frames, while the MFCC test without NBD is to compare the recitation accuracy with the proposed NBD model. Other tests were also carried out on the law of reading to look for patterns of suitability of recitation with *tajweed*.

Table 4. Comparison of the results of the right reference of surah Al-Baqarah

No	Words	Tajwid	Expert Opinion	Sofware Accuracy	Distance
1	alif-lam-mim	Gunnah	100	100	0.00
2	all-lazi-na	Mad	100	100	0.00
3	djaa-likall	Mad	66.6	66.6	0.00
4	Ghoibi	-	66.6	66.6	0.00
5	hudal-lil	Idgom Bilagunnah	100	100	0.00
6	khiro-tihum	-	66.6	100	33.3
7	ki-tabulla	Mad	100	100	0.00
8	mink-kobliq	Ikhfa'	33.3	100	66.6
9	mut-taqim	-	100	100	0.00
10	Nabill	-	100	100	0.00
11	nabi-maa	Mad	66.6	66.6	0.00
12	Nasholaa	Mad	66.6	66.6	0.00
13	roi-ba-fih	Mad	66.6	66.6	0.00
14	rojak-na-hum	Mad	100	100	0.00
15	tawamimm-ma	Mad	100	100	0.00
16	unk-jila-ilaika	Ikhfa'	100	100	0.00
17	wabill-a	Mad	100	66.6	-33.3
18	wal-laji-na	Mad	100	100	0.00
19	wama-unk-jila	Ikhfa'	100	100	0.00
20	wayukii-mu	Mad	100	100	0.00

21	yu-kinun	Mad	66.6	66.6	0.00
22	yuk-minu	Mad	66.6	66.6	0.00
23	yunk-fikun	Ikhfa'	100	66.6	-33.3
24	Average		85.5	86.9	1.45

Table 4 is a comparison of the model results with expert opinions regarding reference recitation that are repeated more than once, then the most appropriate one is selected. The selection of the right reference from the proposed model is carried out by testing references until the accuracy is close to expert opinion.

Testing with a comparison of an expert with software close to a difference of 1.45% shows the closeness of accuracy.

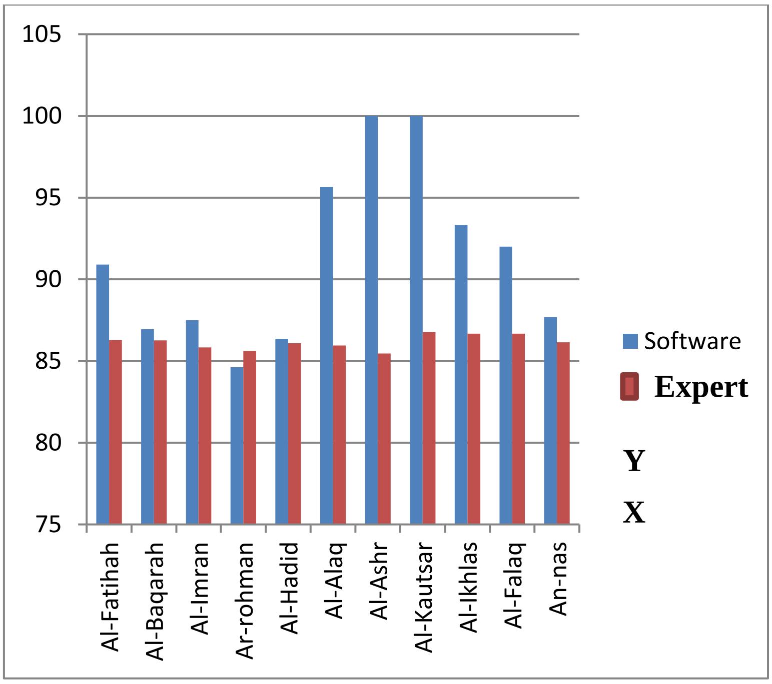


Figure 10. Results of the right reference comparison

Figure 10. shows the results of the comparison of the eleven surah of software accuracy with the opinion of an expert having closeness to the average difference between the accuracy of the software in checking the suitability of Qur'an recitation according to recitation of tajwid by 91.4% while the average accuracy of experts in checking the suitability of Qur'an recitation according to tajwid is 86. %. This happens because the results of feature extraction are then carried out with feature selection with NBD in the feature table, and there is a similarity in the threshold Table 5.17, range, filtering and is a unique and different feature in the feature table in the form of dominant weight.

Testing the Ease of Qur'an Recitation

Tests on tajweed for one rule of tajweed with different or highly varied voices produce different findings from the results of KKP parameters P_2, P_3, P_5, and P_6.

Table 5. Testing of 5 tajweed rules

			K_j							K
No	Surah	Tajweed rules	K_1	K_2	K_3	K_4	K_5	K_6	K_7	PG
1	Albaqarah	Gunnah		2	1		0	0	0	
	Al-Imran	Gunnah			3					
	Al-Alaq	Gunnah		1	2		1			
	Al-Ashr	Gunnah			1					
	Al-Kautsar	Gunnah		1						
	Al-Falaq	Gunnah								1
	An-nas	Gunnah			4					
	Jumlah	∑K1 sampai ∑K7	0	4	11	0	1	0	0	1
2	Al-Imran	Idgom Bigunnah			1			1		
	Jumlah	∑K1 sampai ∑K7	0	0	1	0	0	1	0	0
3	Albaqarah	Idgom Bilagunnah			1					
	Al-Imran	Idgom Bilagunnah			1					
	Ar-rahman	Idgom Bilagunnah			1					
	Al-Alaq	Idgom Bilagunnah			2					
	Al-Ikhlas	Idgom Bilagunnah								1
	Jumlah	∑K1 sampai ∑K7	0	0	5	0	0	0	0	1
4	Alfatehah	Idhar								1
	Al-Hadid	Idhar								1
	Jumlah	∑K1 sampai ∑K7	0	0	0	0	0	0	0	2
5	Albaqarah	Ikhfa'			3					1
	Ar-rahman	Ikhfa'			1					
	Al-Hadid	ikhfa'			1					
	Al-Alaq	Ikhfa'		1	1		2			1
	Al-Ashr	Ikhfa'			1					
	Al-Falaq	Ikhfa'		1	1					
	An-nas	Ikhfa'		1						
	Total	∑K1 to ∑K7	0	3	8	0	2	0	0	2

The ease of Qur'an recitation is applied to the TPA Nurul Huda Plus Purbayan Kotagede, Yogyakarta.

Table 12 shows that gunnah rule pronounced with different areas, namely K_j or K_2, K_3, K_5, K_6, and K or PG. Testing the tajweed rules on the sound of gunnah with different pronunciations in different letters also produces different or unequal patterns. Tajweed tests do not always have the same and different patterns

The tajweed test, for example *ikhfa*, is in the K_2 and K_3 areas. Tajweed rule of *mad* is also in K_2, K_3, and K_5 areas. Likewise, the tajweed rule for *qalqalah* is also in the K_2, K_3 and K_5 areas. This is due to one law with very different sound variants so that they do not find the same pattern.

Table 6. Testing of 3 tajweed rules

			K_j							K
No	Surah	Tajweed rules	K_1	K_2	K_3	K_4	K_5	K_6	K_7	PG
6	Al-Alaq	Iqlab			1					
	Total	∑K1 to ∑K7	0	0	1	0	0	0	0	0
7	Alfatehah	Mad		4	7	0	3	1	0	
	Albaqarah	Mad		1	7	0	1	0	0	0
	Al-Imran	Mad		1	6		1	1		
	Ar-rahman	Mad		2	6		2			1
	Al-Hadid	Mad			6					
	Al-Ashr	Mad		2			2			
	Al-Kautsar	Mad		1						1
	Al-Ikhlas	Mad			2					
	Al-Falaq	Mad			4					
	An-nas	Mad		1	2		1			
	Total	∑K1 to ∑K7	0	12	40	0	10	2	0	2
8	Al-Imran	Qalqalah			1					
	Ar-rahman	Qalqalah					1			
	Al-Alaq	Qalqalah			9		1			
	Al-Ashr	Qalqalah					1			
	Al-Kautsar	Qalqalah					1			
	Al-Ikhlas	Qalqalah		2	3		1			
	Al-Falaq	Qalqalah		1	1					
	Total	∑K1 to ∑K7	0	3	14	0	5	0	0	0

Table 13 shows that the K_2, K_3, K_5, K_6 areas are spread out on the K_j and K counts with PG being dominant in all tajweed rules. Table 6.9 and Table 6.10 show that all the reading laws do not determine the same pattern because the sounds in the tajweed rules are all different and highly varied.

The KKP test of the law of reading has different results, one tajweed rule with very different and varied sounds, as well as difficulties in determining K_2, K_3, K_5 and K_6.. The test for tajweed rules concludes that one tajweed rule contains many and different variations of sounds, so it is very difficult for different letters to be matched with certain tajweed rules. The problem or difficulty is concluded because a pronunciation of the rule is compared with different readings and variants so as to produce different patterns. Research on tajweed produces the distribution of right and wrong that can determine easy learning from tajweed of Qur'an recitation.

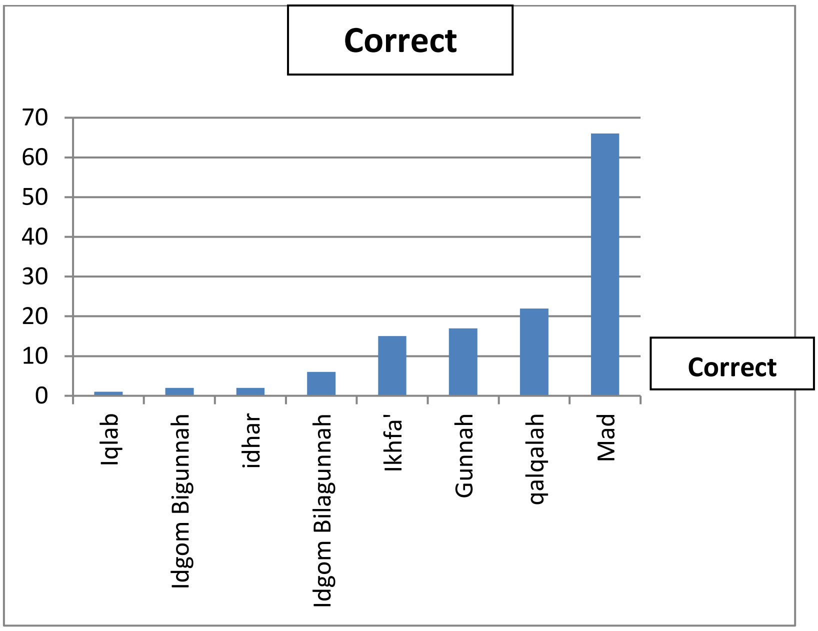


Figure 11 The easiest to the hardest tajweed rules

Figure 11 shows that the tajweed rules that are easy to teach start from 1. Mad, 2. Qalqalah, 3. Gunnah, 4. Ikhfah, 5. Idghom bilagunnah, 6. Idhar, 7. Idghom bigunnah, and 8. Iqlab.

CONCLUSION

The conclusion of this dissertation research is as follows:

1. Feature extraction has been successfully carried out, the selection of references and the selection of the right features can improve checking the suitability of Qur'an recitation with references above 85%.
2. The order of tajweed from the easiest to the most difficult one is as follows: 1. Mad, 2. Qalqalah, 3. Gunnah, 4. Ikhfah, 5. Idghom bilagunnah, 6. Idhar, 7. Idghom bigunnah, and 8. Iqlab

REFERENCE

Abriyono, A., & Harjoko, A. (2013). Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *7*(1), 23–34. https://doi.org/10.22146/ijccs.2149

Aibinu, A.M., Salami, M.J.E., Najeeb, A.R., Azeez, J.F. dan Rajin, S.M.A.K., 2011b. Evaluating the effect of voice activity detection in isolated Yoruba word recognition system. *2011 4th International Conference on Mechatronics: Integrated Engineering for Industrial and Societal Development, ICOM'11 - Conference Proceedings*, (May), hal.17–19.

Barkoni, 2017. Tajwid. *Pakar Tajwid*.

Bodruzzaman, M., Kuah, K., Jamil, T., Wang, C. dan Li, X., 1993. System Using Artificial Neural Network. , hal.1–3.

Bustami, Fadlisyah, I. mauliza, 2017. Sistem Pendeteksi Kesalahan Dalam Membaca Alquran Ayat 1-5 Menggunakan Metode Viterbi Bustami,. *TECHSI : Jurnal Penelitian Teknik Informatika Vol 9 NO. 1 hal 1-15*, Vol 9 No.1, hal.1–15. Available at: http://ojs.unimal.ac.id/index.php/techsi/article/view/205.

Cahyarini, R., Yuhana, U.L. dan Munif, A., 2013. Rancang Bangun Modul Pengenalan Suara Menggunakan Teknologi Kinect. *Jurnal Teknik Pomits*, 2(1), hal.1–5.

Chamidy, T., 2016. Metode Mel Frequency Cepstral Coeffisients (MFCC) Pada klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada Penutur Indonesia. *Matics*, 8(1), hal.36–39. Available at: http://ejournal.uin-malang.ac.id/index.php/saintek/article/view/3482.

Chen, M.L., Changchien, S.K., Zhang, X.M. dan Yang, H.C., 2011. The design of voice recognition controller via grey relational analysis. *Proceedings 2011 International Conference on System Science and Engineering, ICSSE 2011*, (June), hal.477–481.

Chen, X., Huang, J., Wang, Y. dan Tao, C., 2012. Incremental Feedback Learning Methods For Voice Recognition Based On DTW. , hal.1011–1016.

Chitode, D.J.., 2010. 01.Communication Theory.

Davis, S.B. dan Mermelstein, P., 1980. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), hal.357–366.

E.F Codd, J., Ritonga, P. dan Reply, L.A., 2015. Pengertian Normalisasi Database Dan Bentuk-. , hal.3–5.

Erma, A., 2004. Articulatory Class Based Spectral Envelope Representation for. , hal.1647–1650.

Ferdinando, H., 2010. *Dasar-Dasar Sinyal dan Sistem*, ANDI.

Heriyanto, 2015. Analisa Deteksi Huruf Hijaiyah Melalui Voice Recognition Menggunakan Kombinasi Energy. *Telematika*, 12(1), hal.11–22.

Hertiana Bethaningtyas, M.K.A.S.. dan S., 2017. Pengenalan Huruf Hijayyah Berbasis Pengolahan Sinyal Suara dengan Metode MFCC. *Momentum, Vol.13, No. 2 Oktober 2017 ISSN 2406-9329 Hal 49-52 Fakultas Teknik-Universitas Wahid Hasyim Semarange-*, 13(2), hal.49–52.

Hidayat, S., Hidayat, R. dan Adji, T.B., 2015. Sistem Pengenal Tutur Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM. *Conference on Information Technology and Electrical Engineering (CITEE)*, (September), hal.246–251.

Holmes, J.H. and W., 2003. *Speech Synthesis and Recognition, Second Edition*,

Irmawan, Hikmarika, H., Sari, D.W. dan Tammimi, M.C., 2014. Pengenalan Kata dengan Metode Linear Predictive Coding dan Jaringan Syaraf Tiruan Pada Mobile Robot. , (December). Available at:

https://www.researchgate.net/publication/288369279_Pengenalan_Kata_dengan_Metode_Linear_Predictive_Coding_dan_Jaringan_Syaraf_Tiruan_Pada_Mobile_Robot.

Heriyanto, H., Wahyuningrum, T., & Fitriana, G. F. (2021). Classification of Javanese Script Hanacara Voice Using Mel Frequency Cepstral Coefficient MFCC and Selection of Dominant Weight Features. *Jurnal Infotel*, *13*(2), 84–93. https://doi.org/10.20895/infotel.v13i2.657

Kumar, A. A. (2013). Digital signal processing. In *Published by Asoke K. Ghosh, PHI Learning Private Limited, M-97, Connaught Circus, New Delhi-110001 and Printed by Rajkamal Electric Press, Plot No. 2, Phase IV, HSIDC, Kundli-131028, Sonepat, Haryana* (Vol. 23, Issue 4). Prentice-Hall of India Pvt.Ltd. https://doi.org/10.1109/TASSP.1975.1162707

Proakis, J. G., & Manolakis, D. G. (1996). Digital Signal Processing: Principles, algorithms, and applications. In *Digital Signal Processing: Principles, algorithms, and applications*.

Laha, D., 2007. *Handbook of Computational Intelligence in Manufacturing and Production Manajemen*,

Leon, C.G.K., 2009. Robust computer voice recognition using improved MFCC algorithm. *Proceedings - 2009 International Conference on New Trends in Information and Service Science, NISS 2009*, hal.835–840.

Manunggal, H.S., 2005. Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara Dengan Menggunakan Analisa MFCC Feature Extraction. *Tugas Akhir Sarjana pada Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Kristen Petra Surabaya*.

Martyna, S. dan Sudaryanto, S., 2011. Penerapan Metode Particle Swarm Optimization pada Artificial Neural Network Backpropagation untuk Peramalan Penjualan Furniture pada CV. Octo Agung. , hal.1–9.

Meana, H.P. dan Meana, H.P., 2007. *Advances in Audio and Speech Signal Processing: Technologies and Applications*,

Miftahuddin, Y. dan Hakim, M.R., 2017. Coefficient Dan Dynamic Time Warping Untuk Pengenalan Nada Pada Alat Musik Bellyra. , hal.120–127.

Muda, L., Begam, M. dan Elamvazuthi, I., 2010. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. , 2(3), hal.138–143. Available at: http://arxiv.org/abs/1003.4083.

Novianto, D. dan Yuliantari, R.V., 2017. Pengenalan Isyarat Tutur Vokal Bahasa Indonesia Menggunakan Metode Dynamic Time Wraping (Dtw) Berbasis Fungsi Jarak. , (1), hal.1–4. Available at: http://jurnal.untidar.ac.id/index.php/thetaomega/article/view/456/363.

Proakis, J.G. dan Manolakis, D.G., 1996. *Digital Signal Processing: Principles, algorithms, and applications*, Available at: https://engineering.purdue.edu/~ee538/DSP_Text_3rdEdition.pdf.

Putra, A.E., 2008. Frekuensi Cuplik pada FFT. *Tan Li, Processing, Digital Signal*, 1.

Putra, D. dan Adi, R., 2011. Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW. *Biometrika, Universitas Udayana*, 2(1), hal.8–21.

Rabiner, L.R. dan Schafer, R.W., 2007. *Introduction to digital speech processing*,

Sanjaya, M.W.. dan Salleh, Z., 2014. Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Adaptive Neuro-Fuzzy Inferense System (Anfis) Sebagai Kontrol Lampu Otomatis. *Al-HAZEN Jurnal of Physics*, 1(1), hal.1–19.

Shah, S.A.A., Asar, A. ul dan Shah, S.W., 2007. Interactive Voice Response with Pattern

Recognition Based on Artificial Neural Network Approach. *2007 International Conference on Emerging Technologies*, hal.249–252. Available at: http://ieeexplore.ieee.org/document/4516352/.

Smith, S.W., 2000. *Digital signal processing*, Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=826412.

Subali, M., Andriansyah, M. dan Sinambela, C., 2015. Analisis Frekuensi Dasar Dan Frekuensi Formant Dari Fonem Huruh Hijaiyah Untuk Pengucapan Makhraj Dengan Metode DTW. *Prosiding PESAT (Psikologi, Ekonomi, Sastra, Arsitektur &Teknik Sipil) Vol. 6, Oktober 2015 ISSN: 1858-2559 Universitas Gunadarma - Depok - 20-21 Oktober 2015 Hal S-60-72*, 6, hal.S60-72.

Suyanto, S. dan Putra, A.E., 2014a. Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation. *Journal of ICT Research and Applications*, 8(2), hal.97–112. Available at: http://journal.itb.ac.id/index.php?li=article_detail&id=1804.

Suyanto, S. dan Putra, A.E., 2014b. Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation. *Journal of ICT Research and Applications*, 8(2), hal.97–112. Available at: http://journal.itb.ac.id/index.php?li=article_detail&id=1804.

Suyanto dan Hartati, S., 2013. Design of Indonesian LVCSR using Combined Phoneme The Approaches of LVCSR. *Icts*, hal.191–196.

Syafria, F., Buono, A. dan Silalahi, B.I.B.P., 2014. Pengenalan Suara Paru - Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier. , 3.

Thiang, H.S., 2005. Sistem Pengenalan Kata dengan Menggunakan Linear Predictive Coding dan Nearest Neighbor Classifier. *Universitas Kristen Petra*, 5(September), hal.19–24.

Tobin, P., 2007. *PSpice for Digital Signal Processing*, Available at: http://www.morganclaypool.com/doi/abs/10.2200/S00073ED1V01Y200612DCS011.

Tokunbo Ogunfunmi, R.T.M. (Sim) narasimha, 2015. *Speech and Audio Processing and Recognition*, springer.

Tomasouw, B.P. dan Irawan, M.I., 2012. Multiclass Twin Bounded Support Vector Machine Untuk Pengenalan Ucapan. *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, Fakultas MIPA, Universitas Negeri Yogyakarta,*, 2(2004), hal.1–10.

Tshilidzi Marwala, 2012. *Condition Monitoring Using Computational Intelligence Methods*,

Vinay K. Ingle, J.G.P., 2012. *Digital Signal Processing using MATLAB*,

Vladimir Britanak, Patrick C.Yip, K.R.R., 2007. *Discrete Cosine and Sine Transform*,

Widodo, S.M., Siswanto, E. dan Sudjana, O., 2016. Penerapan Metode Mel Frequency Ceptral Coefficient dan Learning Vector Quantization untuk Text-Dependent Speaker Identification. , 11(1), hal.15–20.

Zarkasyi I, 1995. Pelajaran Tajwid. *Gontor Ponorogo*, Trimurti P, hal.Hal 1-3.