Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing Feature Selection, to Test the Shahada Recitation

Heriyanto[1], Dyah Ayu Irawati[2]

[1,2] Informatics Department, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

Abstract

Voice research for feature extraction using MFCC. Introduction with feature extraction as the first step to get features. Features need to be done further through feature selection. The feature selection in this research used the Dominant Weight feature for the Shahada voice, which produced frames and cepstral coefficients as the feature extraction. The cepstral coefficient was used from 0 to 23 or 24 cepstral coefficients. At the same time, the taken frame consisted of 0 to 10 frames or eleven frames. Voting as many as 300 samples of recorded voices were tested on 200 voices of both male and female voice recordings. The frequency used was 44.100 kHz 16-bit stereo. This research aimed to gain accuracy by selecting the right features on the frame using MFCC feature extraction and matching accuracy with frame feature selection using the Dominant Weight Normalization (NBD). The accuracy results obtained that the MFCC method with the selection of the 9th frame had a higher accuracy rate of 86% compared to other frames. The MFCC without feature selection had an average of 60%. The conclusion was that selecting the right features in the 9th frame impacted the accuracy of the voice of shahada recitation.

Keywords: Extraction, Feature, Accuracy, Cepstral, Coefficient

INTRODUCTION

Research for speech recognition in the process required MFCC feature extraction. MFCC feature extraction has been carried out in various fields, including literature, language, and reading speech. MFCC is a feature extraction that produces features or characteristics in the frame and cepstral coefficient parameters. Features are different from one another in the form of parameters. MFCC feature extraction can recognize more voice characters with non-linear voice signals while using Linear Predictive Code (LPC) only for linear ones. The introduction of feature extraction is the first step to get the feature parameter. Voice recognition using different feature selection methods will produce different outputs, so a feature selection method is needed, such as the Dominant Weight Normalization (NBD) feature. Feature extraction and appropriate feature selection methods are required for this research. In addition, the problem of the quality of the speech recognition system is also influenced by the frame length, overlap length, the number of filter banks, and the number of coefficients. Research on feature extraction and selecting the right features to improve the accuracy in checking the suitability of the shahada recitation aims to use MFCC feature extraction and design the proper feature selection method to be more accurate in checking the suitability of the shahada recitation.

Speech recognition research using MFCC feature extraction has been carried out in the field of Arabic by Chamidy (Chamidy, 2016), resulting in the extraction of Mel Frequency Cepstral Coefficients (MFCC) in the form of features to get the suitability value of Indonesian speakers against native speakers, classified using the Hidden Markov Model (HMM).

MFCC is applied in the Indonesian language field 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 Cepstral Coefficient (MFCC) and Hidden Markov Model (HMM), which can recognize phoneme segmentation in Indonesian. A similar study on phonemes was also conducted by Cahyarini et al. (Cahyarini et al., 2013), who were able to identify speech pauses between phonemes.

Language research on hijāʾī (Arabic alphabet) recognition by Bethaningtyas (Hertiana Bethaningtyas, 2017) uses 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 hijāʾī letters by Heriyanto (Heriyanto, 2015) used the average energy and wave deviation methods as a comparison. Meanwhile, another study associated with the hijāʾī phonemes was conducted by Subali et al. (Subali et al., 2015), using the LPC and DTW methods to produce the speakers' formant frequency in pronunciation; DTW has the advantage of autocorrelation. Another MFCC research by way of modification was carried out by (Leon, 2009) in the windowing section. Other studies have also modified MFCC to produce acoustic signal analysis with the stages of pre-emphasis, frame blocking, hamming windowing, Fast Fourier Transform, Mel Filterbank, Discrete Cosine Transform (DCT), Delta energy, and delta spectrum (Muda et al., 2010).

Speech recognition research using an artificial neural network (Sanjaya & Salleh, 2014) requires a long training time, while the Hidden Markov Model (HMM) method (Chamidy, 2016) produces robustness. Voice recognition with Dynamic Time Wrapping (DTW) (Miftahuddin & Hakim, 2017). Feature extraction and appropriate feature selection methods are needed in this research. In addition, the problem of the quality of the speech recognition system is also influenced by the frame length, overlap length, the number of filter banks, and the number of coefficients (Putra, 2011). Selection of the right features is needed in voice checking because it impacts the accuracy of voice checking (Heriyanto et al., 2018a).

Problem Statement

Based on the background description, the problem statement of this research is feature extraction and selection of the right features to increase the accuracy in checking the suitability of shahada recitation.

Research Objective

This research aims to use MFCC feature extraction and design the proper feature selection method to be more accurate in checking the suitability of shahada recitation.

Research Advantage

The benefit of this research is a contribution of science and technology in sound as an alternative method for selecting features for the accurate shahada recitation.

LITERATURE REVIEW

The voice recognition method using feature extraction is fundamental because feature extraction and feature selection influence the suitability and checking of pattern recognition. Feature extraction research includes Mel Frequency Cepstrum Coefficients (MFCC) and Linear Predictive Code (LPC) (Abriyono & Harjoko, 2013); both methods have weaknesses and strengths in feature extraction that produces features.

MFCC has weaknesses, including low frequency, environmental noise, sensitivity, almost similar sound patterns, and classification (Syafria et al., 2014). Meanwhile, MFCC has advantages such as capturing

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
Heriyanto, Dyah Ayu Irawati

important voice characteristics in recognition, capturing critical information in sound, producing minimal data without losing information, and replicating human auditory sounds (Manunggal, 2005). In addition, feature extraction using MFCC is widely used for speech recognition because it is more precise in various conditions (Chamidy, 2016) and is non-linear.

Linear feature extraction using Linear Predictive Code (LPC) has weaknesses, including noise, fluctuating speech frequencies, and classification (Irmawan et al., 2014). The advantage of LPC is in the form of autocorrelation (Abriyono & Harjoko, 2013; Thiang, 2005) because it is linear.

Feature extraction using MFCC and LPC has the same weaknesses, including noise, almost similar speech frequencies, frequently changing frequencies, and classification. The disadvantage of the two methods is also expressed by Abriyono and Harjoko (Abriyono & Harjoko, 2013) that feature extraction using MFCC and LPC is not suitable for recognizing vast numbers of sounds classification is needed.

Based on the weaknesses and strengths of the two methods, both feature extraction using MFCC and LPC, feature extraction using MFCC from an accuracy level of 65% is better than LPC with an accuracy rate of 52% (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 non-linear.

Voice recognition using other methods such as Dynamic Time Wrapping (DTW) results in calculating the distance between two-time series data (Putra et al., 2011). This method has the advantage of calculating the distance between two data vectors of different lengths or knowing the value of the smallest distance matching between the voices of novice speakers and expert speakers. According to Miftahuddin and Hakim (Miftahuddin & Hakim, 2017), DTW is an algorithm as a non-linear sequence alignment used to measure the similarity of a pattern in a data series area that varies with time and is more realistic.

DTW has a weakness in its accuracy, namely with highly variable results (Novianto & Yuliantari, 2017). Still, it equals the Hidden Markov Machine (HMM) accuracy level (Chamidy, 2016) of 80% to 85%. Meanwhile, according to Suyanto and Putra (Suyanto & Putra, 2014), the HMM method has weaknesses in terms of being less resistant or robust.

Another speech recognition method using Neural Network (NN) has advantages in learning systems, knowledge acquisition, classification, and generalization of a pattern (Sanjaya & Salleh, 2014). According to Martyna and Sudaryanto (Martyna & Sudaryanto, 2011), NN has a weakness in the training process that requires a long time with a large amount of data. The same statement by Aibinu et al. (Aibinu et al., 2011), identifying the utterances of numbers one to nine has a problem when the training process with extensive data requires a very long processing time.

Research Contribution

The research contribution lies in the Comparison method for selecting the right features to increase the accuracy of the shahada recitation suitability.

RESEARCH METHOD

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

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation
Heriyanto, Dyah Ayu Irawati

MFCC is a feature extraction that produces features or characteristics that distinguish one another in the form of cepstral coefficient parameters (Abriyono & Harjoko, 2013). Mel Frequency Cepstral Coefficient (MFCC) feature extraction converts sound waves into several parameters, such as the cepstral coefficient representing 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 stages: pre-emphasis, frame blocking, windowing, Fast Fourier Transform (FFT), Mel Frequency Wrapping (MFW), Discrete Cosine Transform (DCT), and cepstral liftering, which results in feature extraction in the form of parameters like features, namely frame and cepstral coefficient (Putra, 2011).
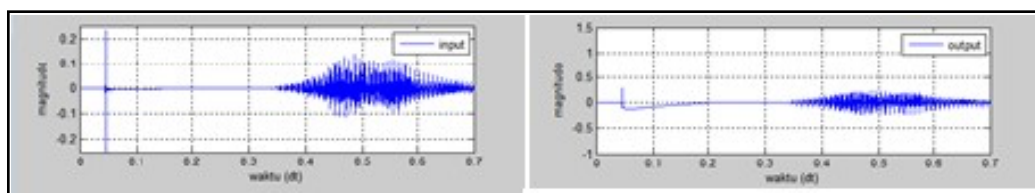
Pre-emphasis

According to Chitode (Chitode, 2010), pre-emphasis emphasizes high-frequency components, aligning at low and high frequencies. According to Tokunbo Ogunfunmi (Tokunbo Ogunfunmi, 2015), pre-emphasis reduces noise to improve Signal to Noise Ratio (SNR) and reduce unwanted noise. Pre-emphasis is a simple signal processing, basically a linear filter, and is still in the time domain (Holmes, 2003).

Pre-emphasis is an early stage in the MFCC process. This stage is done because the signal often experiences noise interference, so it is necessary to reduce noise or noise. Pre-emphasis is a straightforward way to reduce noise. Pre-emphasis aims that the baseband level on the high-frequency section still has good signal quality. The pre-emphasis process, according to Proakis and Manolakis (Proakis & Manolakis, 1996), is with the value α between 0 to 1 or between $0.9 \leq \alpha \leq 1.0$ 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 signal before pre-emphasis; the symbol n is the serial number of the signal, α is the pre-emphasis filter constant between 0.9 to 1.0, and s is the signal. As shown in Figure 1, the pre-emphasis process shows that part (a) is the input sound before the pre-emphasis, while Figure 1 part (b) is the output signal processing after the pre-emphasis.

(a)   (b)

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

The nth signal retrieval in pre-emphasis is carried out as long as one word or two words are recited, with a time of one to three seconds, with the value α of 0.92.

Frame blocking

The signal that has gone through the pre-emphasis process is followed by a frame blocking process blocked in a frame with N samples and shifted by M samples so that N=2M with M<N. Figure 2 shows an illustration of frame blocking (Abriyono & Harjoko, 2013). The width of the frame is denoted by N, while the shift width of each frame is as M. The overlap width is calculated as the N-M difference.

Frame blocking, according to Holmes (Holmes, 2003), analyzes the speech signal into frames represented by a single feature vector depicted in a spectrum of averaged time intervals. The time in the frame is taken on average between 20 to 40 milliseconds, according to Chamidy (Chamidy, 2016). Frames are taken as long as possible to get a good frequency resolution, while the shortest possible time is meant to get the best time domain. Calculation of the number of blocking frames uses equation (2)

$$f_l(n) = y(Ml + n).$$

(2)

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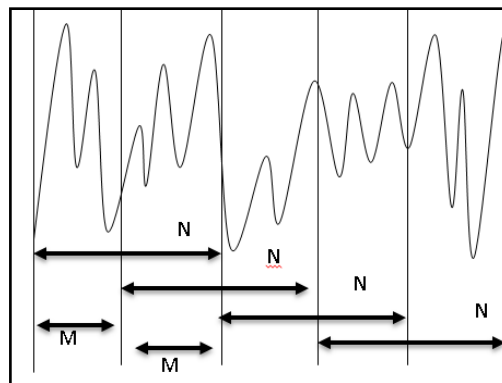

Figure 2 Frame blocking illustration (Abriyono & Harjoko, 2013)
Figure 2 shows that M is the voice signal's first frame in the formula, symbolized as fl, then M+M=N.

Windowing
According to Proakis and Manolakis (Proakis & Manolakis, 1996), windowing has a smoothing effect on the spectrum after going through the frame blocking process. Windowing aims to reduce the discontinuity effect at the edges of the frame generated by the frame blocking process. Windowing used is Rectangular Window, Hamming Window, and Hanning Window (Chamidy, 2016). Of the three windowing functions, researchers use Hanning windowing because it is smoother than the others (Putra, 2008). Representation of windowing function uses equation (3)

$$X(n) = f_l(n)w(n).$$

(3)

In this case, the function X(n) is the windowing signal, where fl 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, the Hanning windowing function uses equation (4)

$$w(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{M-1}\right)\right).$$

(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.
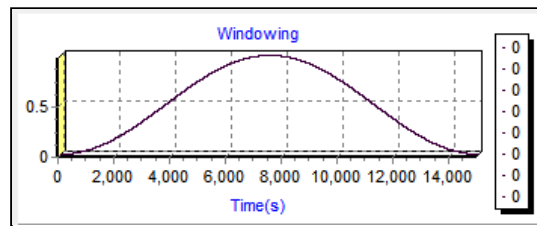

Figure 3 Windowing illustration (Proakis & Manolakis, 1996)

Fast Fourier Transform (FFT)

Fast Fourier Transform is developing the Discrete Fourier Transform (DFT) algorithm, which converts digital signals in the time domain to the frequency domain (Abriyono & Harjoko, 2013). This signal represents the signal decomposition in terms of a sinusoidal component. A sinusoid is a sinusoid of the same frequency but different amplitude and phase. FFT is an algorithm developed by Cooley and Turkey which processes signals from the time domain to the frequency domain.

According to Proakis and Manolakis (Proakis & Manolakis, 1996), the Fourier transform is one of several useful mathematical tools in analyzing and designing Linear Time-Invariant (LTI) systems and Fourier series.

FFT is a fast algorithm method to be able to implement Discrete Fourier Transform (DFT). According to Proakis and Manolakis (Proakis & Manolakis, 1996), DFT is a computational tool that plays a vital role in many digital signal processing applications, such as frequency analysis, power spectrum estimation, and linear filters. DFT computation time is too long and inefficient then FFT can perform calculation efficiency. Proakis and Manolakis (Proakis & Manolakis, 1996) stated that the FFT method is an efficient way to calculate DFT. Discrete Fourier Transform (DFT) using equation (5)

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

(5)

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

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

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

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
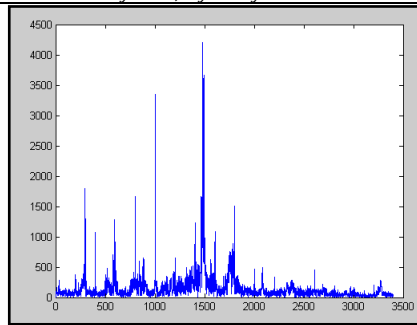*Heriyanto, Dyah Ayu Irawati*

Figure 4 FFT produces spectrum (Kumar, 2013)

Mel Frequency Wrapping (MFW)

Mel Frequency Wrapping (MFW) is a filter in the form of a filterbank to determine the energy size of a particular frequency band in the voice signal (Miftahuddin dan Hakim, 2017; Putra, 2011). MFW, according to Laha (Laha, 2007), converts the frequency into mel.

Meanwhile, according to Tshilidzi Marwala (Tshilidzi Marwala, 2012), MFW contains a filterbank spaced on a mel scale. Filterbank has a frequency response through a triangular path whose distance and a constant frequency interval determines magnitude. 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)

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

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

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

In this case, the mel uses a frequency with the mel scale and f as the 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.

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation
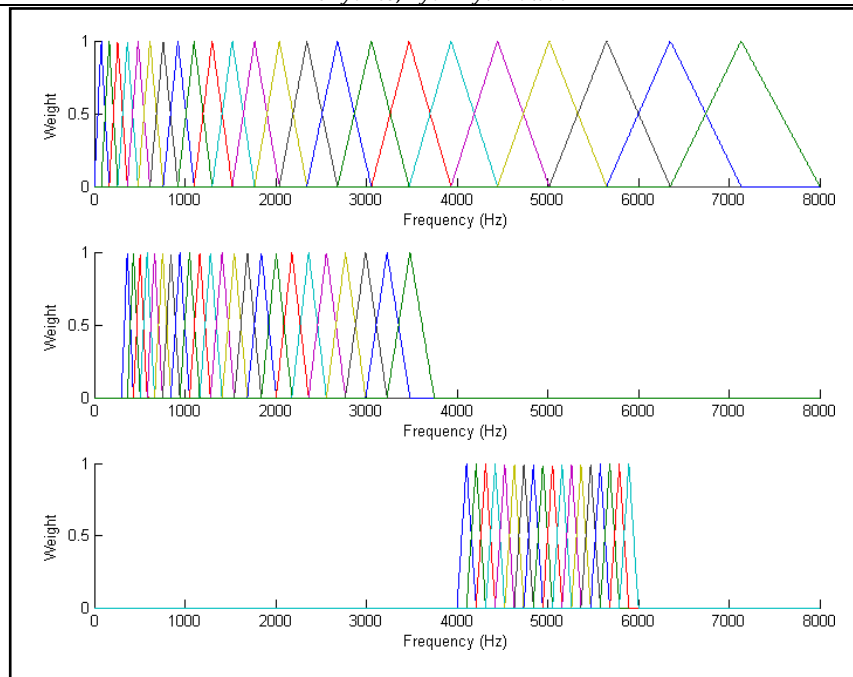*Heriyanto, Dyah Ayu Irawati*

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 the color; the weight, in this case, the db, is the amplitude.

Discrete Cosine Transform (DCT)

DCT, according to Smith (Smith, 2000), is a relative of the signal-decomposed Fourier transform into a cosine wave. DCT is also widely used to process image processing, such as JPEG files. The concept of DCT has similarities with the inverse Fourier transform. However, the result of DCT is approaching Principle Component Analysis (PCA). PCA is a classical static method that is widely used in data analysis and compression.

DCT is assumed to replace the inverse Fourier transform in the MFCC feature extraction process (Putra, 2011). Discrete Cosine Transforms (DCT) are members of the class of sinusoidal unit transformations (Britanak et al., 2007). DCT has the goal of producing a mel cepstrum to improve recognition quality. DCT uses equation (9).

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

In this case, Cm 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 cepstrum.

Cepstral Liftering

According to Proakis and Manolakis (Proakis & Manolakis, 1996), the cepstral coefficient is the Fourier series coefficient. The cepstrum defined by Rabiner and Schafer (Rabiner & Schafer, 2007) is the opposite of the Fourier transform. Cepstral liftering increases the accuracy used for pattern matching, both speaker and speech recognition (Putra, 2011). Cepstral coefficient uses equation (10)

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

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 cepstral liftering results in frames and cepstral coefficients, which are then processed into feature selection.

FRAME FEATURE SELECTION

Selection of the right features is by finding the correct number of frames by determining threshold, range, filtering, and eliminating the duplication, normalization and dominant weights.
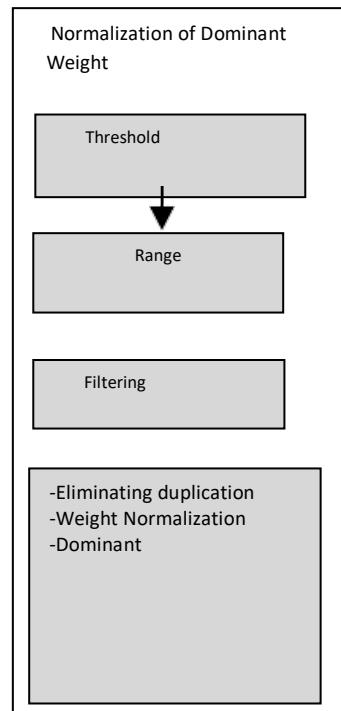
Figure 6 Frame Feature Selection

Threshold

Threshold determination is based on MFCC results in the form of cepstral coefficient (c)=w(k), and frame parameters formed produces 1st to 6th thresholds (parameters b1 to b6). The threshold uses equations (11) to (16)

$$b_1 = min((w(k)),$$

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

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

$$b_4 = average(w(k)),$$

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

$$b_6 = max\left(w(k)\right). \qquad (16)$$

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 specific threshold values that produce maximum accuracy. Next, the minimum and maximum thresholds are taken by dividing the threshold into two, namely min and max. Then min and max are also divided by two to get the median and average.

Coverage similarity

The coverage similarity is based on a predetermined threshold, and with the threshold (parameters b1 to b6), a rule is made with various conditions. The rule checks in the form of conditions from cepstral coefficients (c) with the rule if they meet, the weights become one, or the rule uses parameters r1 to r7.

Coverage Conformity Check
Coverage b_1 to b_6 by checking the condition is in each frame *(I = 0.1, …, F)* with r_1 to r_7 Filtering rules that accumulate with R_{ij} with j = 1, 2, 3, 5, 6, 7; Weight = 1
1. Condition r_1 is if *(b₁)* min = w(k) then R_{i1} = weight else if weight = 0
2. Condition r_2 is if *(b₁)* ≥ w(k) and (w(k) <*b₂*) then R_{i2} = weight elseif weight = 0
3. Condition r_3 is if *(b₂)* ≥ w(k) and (w(k) <*b₃*) then R_{i3} = weight elseif weight = 0
4. Condition r_4 is if *(b₃)* ≥ w(k) and (w(k) <*b₄*) then R_{i4} = weight elseif weight = 0
5. Condition r_5 is if *(b₄)* ≥ w(k) and (w(k) <*b₅*) then R_{i5} = weight elseif weight = 0
6. Condition r_6 is if *(b₅)* ≥ w(k) and (w(k) <*b₆*) then R_{i6} = weight elseif weight = 0
7. Condition r_7 is if *(b₆)* max = w(k) then R_{i7} = weight elseif weight = 0

Figure 7 Checking the conformity of the coverage and filtering

Figure7showscheckingthecepstralcoefficient*(c)*orw*(k)*withcoverageandfiltering.

Filtering

The filtering is 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 r1, r2, r3, r4, r5, r6, and r7; "if it meets the threshold or does not meet the cepstral coefficient (c) = w(k) according to the rules for the coverage of Figure 4.5, then it is given weights one in each frame and recitation"; so that they are summed or accumulated. The filtering is done to select by separating or selecting or filtering so that the filtering results become the parameter Pij in the feature table forming p1 to p7. The sum of the filtering results uses equations (17) and (18).

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

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

In this case, U is the total number of Gj. The symbol Gj is the result of the total of pij; overall, the symbol j is 1 to 7, while the frames formed in the form of i are the 0th frame to the Fth frame. The pij symbol is the result of filtering or weight accumulation. Symbol a = 1, 2, ..., 7.

Eliminating weight duplication

After going through the filtering process, there is a duplication of weights, so it is necessary to eliminate the weight duplication 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. Eliminate weight duplication by looking for similarities and then eliminated by weight = 0. Set that Qij=pij 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 weight duplication on the parameter Qij if it is equal to pij with the same check, is given a weight = 0. Weight duplication is removed, and then the total weights are calculated using equation (19).

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

(19)

In this case, the symbol Z is the calculation of the total weight Q. The weight Qij is the result of filtering the weights whose duplication is removed; j is 1 to 7, i is the 0th frame to the Fth frame.

Weight Normalization

Weight Normalization is an equalizing process, aligning the weights in a more balanced way so that they are more proportional. The result of weight normalization (Sj) is the parameter npf in the feature table. The weight normalization calculation uses equation (20)

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

(20)

In this case, S is the result of weight normalization. The symbol j with j = 2, 3, 5, and 6, while Z calculates the number of weights. The weight Qij is the result of weight filtering whose duplication has been removed. The symbol i is the frame. The taken frames are from i = 0 to F.

Sequential Multiplication

The sequential multiplication is performed on the feature table of the total weight and dominant weight. This multiplication is used since multiplication has the largest value (Marlina et al., 2017). The result of sequential multiplication is the parameter Pj. The sequential multiplication uses equation (21

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
Heriyanto, Dyah Ayu Irawati

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

In this case, Pj is the result of sequential multiplication, and the symbol Rij is the result of the filtering process coverage in the form of accumulated weights. The symbol j is 2, 3, 5, and 6. The symbol B is the dominant weight normalization. The symbol Z is the total weight Q calculation, while Q results from the weight filtering whose duplication has been removed. The symbol i is the frame from i = 0 to F.

Conformity of Pattern Uniformity (KKP)
Checking the recitation conformity by the KKP calculation to approach the results of the feature table whether it is appropriate or not. According to Yuwono and Antonio (Yuwono & Antonio, 2015), more complex signals are used with an average frequency. The KKP results using Kj calculations form parameters j = 2, 3, 5, and 6. The KPKPHcalculation uses equations (22) and (23)

$$Kj = \frac{P_jU}{\overline{Z}_j^2},$$

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

In this case, Pj is the result of sequential multiplication. The symbol U is the total of P2, P3, P5, and P6. The symbol K is the overall pattern conformity (KKP). The symbol Z is the total weight. The symbols for j are 2, 3, 5, and 6.
TEST

    a. **Test Calculation with Mean Percentage**
        This research uses the mean calculation in the form of a percentage. According to Putra (Darma Putra, 2011), the percentage is the correct mean calculation by means of the correct number divided by the total number of matches and multiplied by 100% and the mean percentage of errors. The mean percentage is right, and the mean percentage is wrong using equations (21) and (22)

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

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

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

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

    b. **Test and Indicator**
    This research is tested after checking the conformity of the shahadah recitation with the indicators of research achievement in percentages in Table 1. The tests carried out consist of two stages: testing the MFCC and selecting the correct features, and without selecting the features for the shahada recitation.

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
Heriyanto, Dyah Ayu Irawati

Table 1 Test and Indicator

No	Test	Objective	Indicator
1	MFCC	Finding the correct mean	50-75%
2	Selecting the correct feature		
	NBD	Finding the correct mean	85-90%
	Total frame	Finding the correct mean	85-90%

Researchers looked for the correct mean percentage of the research results based on the test and indicator. The first testing stage is carried out on selecting the right features on the frame to get high accuracy. The feature selection process obtains feature selection through the similarity of threshold, coverage, filtering, eliminating weight duplication, weight normalization, and dominant weights (Heriyanto et al., 2018b).

FINDING AND DISCUSSION

The results of testing the selection of the right features are carried out on the number of cepstral coefficients and the number of frames, while the Feature Extraction Results using MFCC produce frames and cepstral coefficients with eleven frames (11) and as many as 24 cepstral coefficients, can be seen in Figure 8.

```
PreProcessing:
Open File : Asyhaduallah01.wav Read
DataSinyal[Progress...] Framing
Data...
PreEmphasis ...
PreEmphasis ...
Windowing ... FFT ...
Cepstral ... SUCCESS ...
Amount coeffisien Cepstral :24
Amount Frame :11


Frame[0]Coef[0]       -3,0543164080676
Frame[0]Coef[1]       -4,34152405140291
Frame[1]Coef[11]      8,51988514472823
Frame[1]Coef[12]      -1,77962259814521
Frame[1]Coef[13]      -1,04648262663401
Frame[1]Coef[14]      -9,74338793951055
Frame[11]Coef[23]     2,53878980233908
```

Figure 8. Frame output and cepstral coefficient

Figure 8 shows the MFCC outputs in the form of twenty-four cepstral coefficients and frames.

The proposed feature selection has six stages: determining the same threshold, making the same coverage, filtering, eliminating weight duplication, weights normalization, and dominant weights (Heriyanto et al., 2018a)(Heriyanto, 2021). These stages are applied in the feature selection algorithm for Dominant Weight Normalization (NBD) to generate a feature table. The feature table is then checked for accuracy and finds the best frame features.

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
Heriyanto, Dyah Ayu Irawati
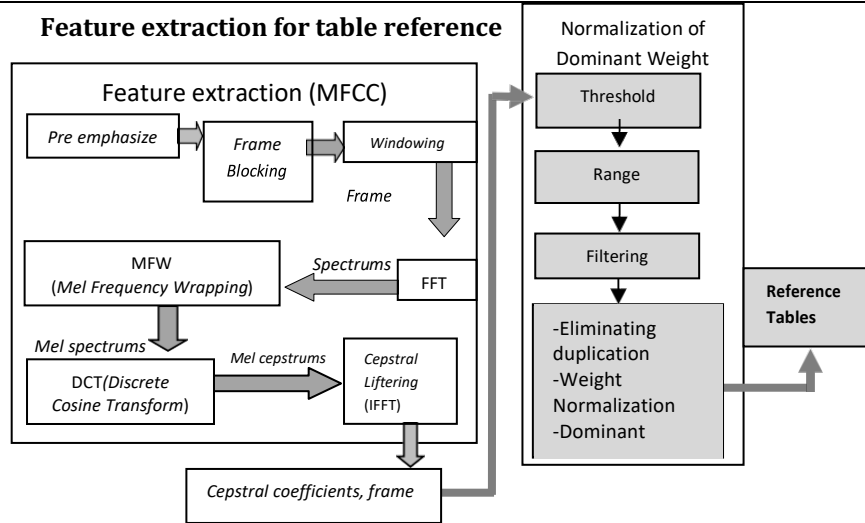
Figure 9. Feature Extraction and NBD Features Selection (Heriyanto et al., 2018a)(Heriyanto et al., 2021)

Figure 9 shows the voice feature extraction research using MFCC, which produces features in frame and cepstral coefficient parameters, and is processed to select NBD features starting from the threshold, coverage, filtering, eliminating weight duplication, weight normalization, to dominant weight.

Table 2. MFCC results of the Shahada recitation

File code	Frame	Recitation	Coefficient	Cepstral_coefficient_value
245	0	ašhadu ʾan lā	0	23.66
246	0	ašhadu ʾan lā	1	-9.09
247	0	ašhadu ʾan lā	2	20.83
253	0	ašhadu ʾan lā	8	56.93
266	0	ašhadu ʾan lā	21	3.36
267	0	ašhadu ʾan lā	22	4.06
268	0	ašhadu ʾan lā	23	-0.22

Table 2 shows the results of feature extraction using the MFCC of the Shahada recitation. The result of feature extraction is a frame consisting of eleven frames and twenty-four cepstral coefficients. In the next process, the frame and cepstral coefficient results are selected by selecting the feature by selecting the frame feature with a feature selection algorithm using Dominant Weight Normalization.

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation
Heriyanto, Dyah Ayu Irawati

Feature Selection Algorithm of Dominant Weight Normalization (NBD)
0. Start
1. Record recitation
2. *MFCC function with frame=f, cepstral coefficient (c)=**w(k)***
3. **Determining threshold 1 to threshold 6 *(b₁-b₆)***

   a. Create $\mathbf{b_1} = min(w(k))$,

   b. Create $\mathbf{b_2} = \dfrac{min(w(k))+(\frac{min(w(k))+max(w(k))}{2})}{2}$

   c. Create $\mathbf{b_3} = average(w(k))$,

   d. Create $\mathbf{b_4} = \dfrac{min(w(k))+max(w(k))}{2}$

   e. Create $\mathbf{b_5} = \dfrac{(\frac{min(w(k))+max(w(k))}{2})+max(w(k))}{2}$,

   f. Create $\mathbf{b_6} = max(w(k))$.

4. **Creating coverage**
   Coverage by checking conditions in each frame with cepstral coefficient *(c)=w(k)*
   Weight=1

   a. Rule₁ is if *(b₁)* min=w(k) then $\mathbf{p_{i1}}$ =weight elseif weight=0

   b. Rule₂ is if *(b₁)* ≥ = w(k) and (w(k) <*b₂*) then $\mathbf{p_{i2}}$ =weight elseif weight=0

   c. Rule₃ is if *(b₂)* ≥ = w(k) and (w(k) <*b₃*) then $\mathbf{p_{i3}}$ =weight elseif weight=0

   d. Rule₄ is if *(b₃)* ≥ = w(k) and (w(k) <*b₄*) then $\mathbf{p_{i4}}$ =weight elseif weight=0

   e. Rule₅ is if *(b₄)* ≥ = w(k) and (w(k) <*b₅*) then $\mathbf{p_{i5}}$ =weight elseif weight=0

   f. Rule₆ is if *(b₅)* ≥ = w(k) and (w(k) <*b₆*) then p_{i6}=weight elseif weight=0

   g. Rule₇ is if *(b₆)* max= w(k) then $\mathbf{p_{i7}}$ =weight elseif weight=0

5. *Filtering*
   Filtering results are $\mathbf{p_{i1}}$ to $\mathbf{p_{i7}}$

   $$G_j = \sum_{i=0}^{F} p_{ij}$$

   Calculate total $\mathbf{p_{i1}}$ to $\mathbf{p_{i7}}$ by

   $$U = \sum_{j=0}^{a} G_j$$

   Calculate total number of patterns by

6. **Eliminating weight duplication**
   Determine that $Q_{ij}=p_{ij}$ and find similarity eliminate 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 $Q_{i0}=p_{i2}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i3}$ then $Q_{ij}=0$, if $Q_{i2}=p_{i4}$ then $Q_{ij}=0$
   if $Q_{i0}=p_{i10}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i10}$ then $Q_{ij}=0$, if $Q_{i2}=p_{i10}$ then $Q_{ij}=0$

   g. Calculate Total by

   $$Z_j = \sum_{i=0}^{F} \xi_{ij}$$

   Weight Normalization
   Weight Normalization Calculation by

   $$S_j = \sum_{i=0}^{F} \frac{Q_{ij}}{Z_i}$$

   h. Dominant Weight
   Sort the largest value S_j in the feature table (npf₂, npf₃, npf₅, npf₆) into variable $\mathbf{B_j}$

   i. Save the table of features total calculation weight=Z,
   Weight normalization =S_j, total pattern=U
7. end

Figure 10. Feature Selection Algorithm of Dominant Weight Normalization (NBD) Algorithm (Heriyanto et al., 2021)

Figure 10 shows the NBD algorithm started from the first and second steps of collecting the recitation voice, and the third step performs the threshold process. Then, in the fourth step and so on, the coverage is made, and so is the filtering, eliminating weight duplication, weight normalization, and dominant weights.

Data Collection

The voice sampling is carried out as many as 300 recordings and tested on 200 recorded voices. The cepstral coefficient used starts from 0 to 23 as many as 24 cepstral coefficients and 0 to 10 frames. The frequency used is 44.100 kHz 16-bit stereo by recording both male and female voices.

Table 3. Male and Female voice recording data

NO	Sex	Recitation	Total Sampling
1	Male	ašhadu ʾan lā	30
2	Male	ilāha ʾilla -llāh	30
3	Male	wa-ʾašhadu ʾanna	30
4	Male	Muhammadar	30
5	Male	rasūlu -llāh	30
1	Female	ašhadu ʾan lā	30
2	Female	ilāha ʾilla -llāh	30
3	Female	wa-ʾašhadu ʾanna	30
4	Female	Muhammadar	30
5	Female	rasūlu -llāh	30
			300

Table 4. Test recording data

NO	Sex	Recitation	Total Sampling
1	Male	ašhadu ʾan lā	20
2	Male	ilāha ʾilla -llāh	20
3	Male	wa-ʾašhadu ʾanna	20
4	Male	Muhammadar	20
5	Male	rasūlu -llāh	20
1	Female	ašhadu ʾan lā	20
2	Female	ilāha ʾilla -llāh	20
3	Female	wa-ʾašhadu ʾanna	20
4	Female	Muhammadar	20
5	Female	rasūlu -llāh	20
			200

Tables 3 and 4 are voice sampling collections for each recitation voice of five males and females; 20 samples are taken, both male and female. The total number of voices recording data for feature extraction and testing is 500 voices.

DISCUSSION

Checking the conformity of the shahada recitation is carried out on the feature table by calculating the Conformity of Pattern Uniformity (KKP). The calculation of the results of the recitation conformity is expected to be close to the results in the feature table so that the percentage of recitation conformity increases.

Checking the feature table with a threshold, with the same coverage and filtering, sequential calculation, and KKP calculation, using the equations as in the algorithm. The recitation check is carried out to select the right reference and the right feature (Heriyanto et al., 2018a),(Heriyanto, 2021).
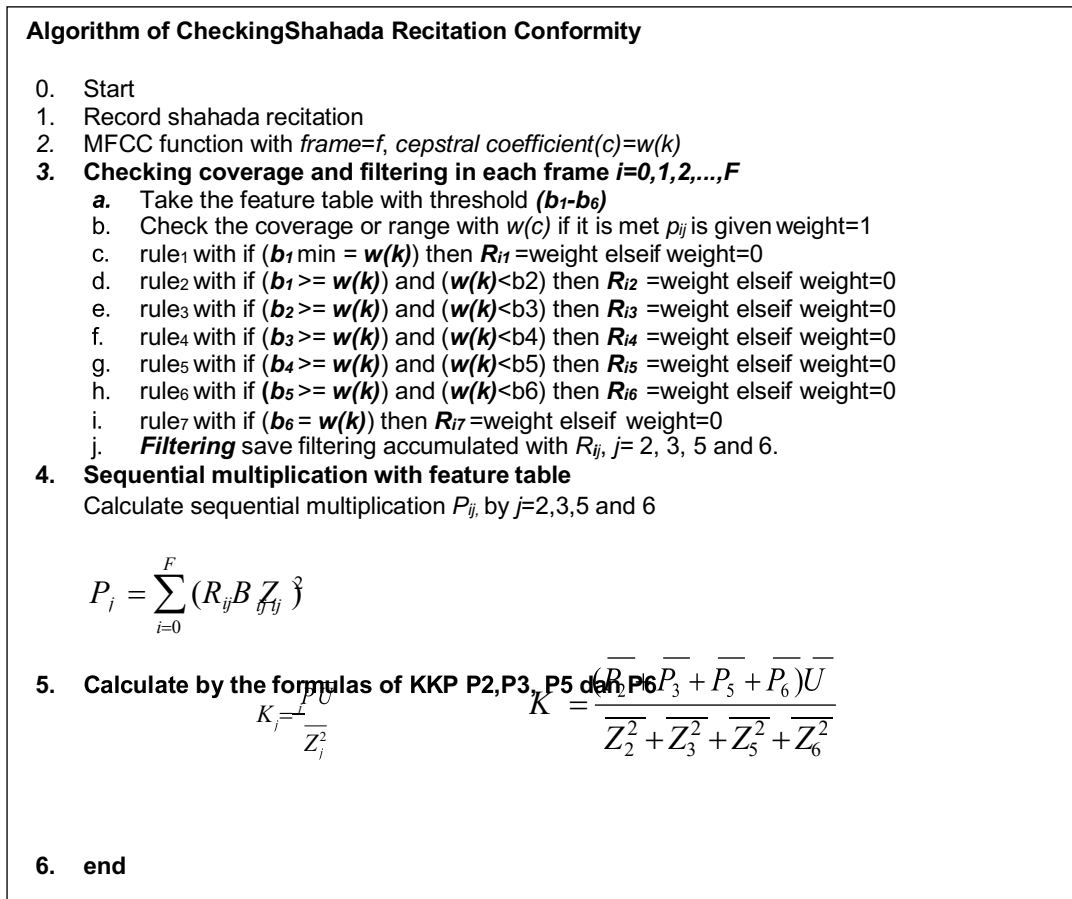

Algorithm of CheckingShahada Recitation Conformity

0. Start
1. Record shahada recitation
2. *MFCC function with frame=f, cepstral coefficient(c)=w(k)*
3. **Checking coverage and filtering in each frame i=0,1,2,...,F**
   a. *Take the feature table with threshold (b₁-b₆)*
   b. Check the coverage or range with $w(c)$ if it is met p_{ij} is given weight=1
   c. rule₁ with if (**b₁** min = **w(k)**) then R_{i1} =weight elseif weight=0
   d. rule₂ with if (**b₁** >= **w(k)**) and (**w(k)**<b2) then R_{i2} =weight elseif weight=0
   e. rule₃ with if (**b₂** >= **w(k)**) and (**w(k)**<b3) then R_{i3} =weight elseif weight=0
   f. rule₄ with if (**b₃** >= **w(k)**) and (**w(k)**<b4) then R_{i4} =weight elseif weight=0
   g. rule₅ with if (**b₄** >= **w(k)**) and (**w(k)**<b5) then R_{i5} =weight elseif weight=0
   h. rule₆ with if (**b₅** >= **w(k)**) and (**w(k)**<b6) then R_{i6} =weight elseif weight=0
   i. rule₇ with if (**b₆** = **w(k)**) then R_{i7} =weight elseif weight=0
   j. ***Filtering*** save filtering accumulated with R_{ij}, j= 2, 3, 5 and 6.
4. **Sequential multiplication with feature table**
   Calculate sequential multiplication P_{ij}, by j=2,3,5 and 6

$$P_j = \sum_{i=0}^{F} (R_{ij} B_{(j-1)j} Z_j^2)$$

5. **Calculate by the formulas of KKP P2,P3, P5 dan P6**

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

6. **end**

Figure 11. Algorithm of Checking Shahada Recitation Conformity (Heriyanto et al., 2018a)(Heriyanto, 2021)

Figure 11 shows the algorithm of checking the shahada recitation voice conformity by taking the feature results from the MFCC feature extraction in frame and cepstral coefficient parameters. The conformity checking algorithm checks the coverage, filters each frame, then sequential multiplication, and the last calculation of the Conformity of Pattern Uniformity (KKP). The results of the frame and cepstral coefficient can be seen in Table 5.

Table 5. Shahada Recitation Results of Frame Feature Selection

No	RECITATION	Sampling (L/P)	MFCC (%)	Frame (%)					
				0	1	2	3	4	5
1	ašhadu 'an lā	40	60	20	10	30	40	30	40
2	ilāha 'illa -llāh	40	70	30	40	50	70	60	70
3	wa-'ašhadu 'anna	40	60	40	50	70	60	80	50
4	Muhammadar	40	50	50	50	70	50	60	70
5	rasūlu -llāh	40	60	60	60	60	70	70	80

No	RECITATION	200 Sampling (L/P)	60 MFCC (%)	40 Frame (%) 6	42 7	56 8	58 9	60 10	62
1	ašhadu ʾan lā	40	60	30	40	50	70	50	
2	ilāha ʾilla -llāh	40	70	50	40	70	90	70	
3	wa-ʾašhadu ʾanna	40	60	70	50	80	100	90	
4	Muhammadar	40	50	60	60	80	90	90	
5	rasūlu -llāh	40	60	70	80	80	80	80	
		200	60	56	54	72	86	76	

Based on Table 5, the selection of the best frame features on the 9th frame with a mean value of 86% better than the other frames. MFCC without feature selection has a mean of 60%.

CONCLUSION

This research uses MFCC feature extraction and frame feature selection with the correct Dominant Weight Normalization algorithm for the shahada recitation that produces frames and cepstral coefficients. The cepstral coefficient used is from 0 to 23 as many as 24 cepstral coefficients. The sampling of 500 voices is tested against 300 voices of reference recordings and 200 of test data recordings. The frequency used is 44.100 kHz 16-bit stereo. The accuracy results obtained that the MFCC method with the selection of the 9th frame feature has a higher accuracy rate of 86% compared to other frames. The MFCC without feature selection has an average of 60%.

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

Achmad Rizal, V. S. (2014). Aplikasi Pengolahan Sinyal Digital pada Analisis dan Pengenalan Suara Jantung dan Paru untuk Diagnosis Penyakit Jantung dan Paru Secara Otomatis.

Aibinu, A. M., Salami, M. J. E., Najeeb, A. R., Azeez, J. F., & Rajin, S. M. A. K. (2011). 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, 17–19. https://doi.org/10.1109/ICOM.2011.5937134

Bender, W., Gruhl, D., Morimoto, N., & Lu, A. (1996). Techniques for data hiding. IBM Systems Journal, 35(3.4), 313–336. https://doi.org/10.1147/sj.353.0313

Cahyarini, R., Yuhana, U. L., & Munif, A. (2013). Rancang Bangun Modul Pengenalan Suara Menggunakan Teknologi Kinect. Jurnal Teknik Pomits, 2(1), 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), 36–39.

Darma Putra, A. R. (2011). Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW. LONTAR KOMPUTER Biometrika, Universitas Udayana ISSN:2088-1541, 2(1), 8–21.

Davis, S. B., & 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), 357–366. https://doi.org/10.1109/TASSP.1980.1163420

Hassan, T., Wassim, A., & Bassem, M. (2007). Analysis and Implementation of an Automated Delimiter of " Quranic " Verses in Audio Files using Speech Recognition Techniques. Robust Speech Recognition and Understanding, June, 352–362. https://doi.org/10.5772/4759

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
Heriyanto, Dyah Ayu Irawati

Heriyanto. (2015). Analisa Deteksi Huruf Hijaiyah Melalui Voice Recognition Menggunakan Kombinasi Energy. Telematika, 12(01), 11–22.

Heriyanto. (2021). Good Morning to Good Night Greeting Classification Using Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction and Frame Feature Selection. Telematika, 18(1), 88–

105. https://doi.org/10.31515/telematika.v18i1.4495

Heriyanto, H.-, Hartati, S., & Putra, A. E. (2018a). Evaluation of Suitability of Voice Reading of Al-Qur'an Verses Based on Tajwid Using Mel Frequency Cepstral Coefficients (MFCC) and Normalization of Dominant Weight (NDW). Advances in Image and Video Processing, 6(2). https://doi.org/10.14738/aivp.62.4268

Heriyanto, H., Hartati, S., & Putra, A. E. (2018b). Ekstraksi Ciri Mel Frequency Cepstral Coefficient (Mfcc) Dan Rerata Coefficient Untuk Pengecekan Bacaan Al-Qur'an. Telematika, 15(2), 99. https://doi.org/10.31315/telematika.v15i2.3123

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

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), 49–52.

Hidayat, S., Hidayat, R., & 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, 246–251.

Holmes, J. H. and W. (2003). Speech Synthesis and Recognition, Second Edition. https://doi.org/10.1145/1185448.1185459

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

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

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, 835–840. https://doi.org/10.1109/NISS.2009.12

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.

Marlina, M., Saputra, W., Mulyadi, B., Hayati, B., & Jaroji. (2017). Aplikasi Sistem Pakar Diagnosis Penyakit ISPA Berbasis Speech Recognition Menggunakan Metode Naive Bayes. Teknologi Informasi Dan Komunikasi Digital Zone, Vol.8(x), No.1.

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

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

Novianto, D., & Yuliantari, R. V. (2017). Pengenalan Isyarat Tutur Vokal Bahasa Indonesia Menggunakan Metode Dynamic Time Wraping (Dtw) Berbasis Fungsi Jarak. 1, 1–4.

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

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

RSF Conference Series: Engineering and Technology
Vol. 1 (1), 335-354
**Comparison of Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction, With and Without Framing
Feature Selection, to Test the Shahada Recitation**
Heriyanto, Dyah Ayu Irawati

Rabiner, L. R., & Schafer, R. W. (2007). Introduction to digital speech processing. In Foundations and Trends in Signal Processing (Vol. 1, Issue 1). https://doi.org/10.1561/2000000001

Sanjaya, M. W. ., & 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), 1–19.

Sari, P. K., Priandana, K., & Buono, A. (2013). Perbandingan Sistem Perhitungan Suara Tepuk Tangan dengan Metode Berbasis Frekuensi dan Metode Berbasis Amplitudo Comparison of Applause Calculation Systems using Frequency- Based Method and Amplitude-Based Method. Jurnal Ilmu Komputer Agri-Informatika, 2 Nomor 1, 29–37.

Smith, S. W. (2000). Digital signal processing. In The Scientist and Engineer's Guide to Digital Signal Processing (Vol. 17, Issue 2). https://doi.org/10.1109/79.826412

Subali, M., Andriansyah, M., & 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, S60-72.

Suyanto, & Hartati, S. (2013). Design of Indonesian LVCSR using Combined Phoneme The Approaches of LVCSR. Icts, 191–196. https://doi.org/10.12962/p9772338185001.a33a33

Suyanto, S., & Putra, A. E. (2014). 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), 97–112. https://doi.org/10.5614/itbj.ict.res.appl.2014.8.2.2

Syafria, F., Buono, A., & 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), 19–24.

Tshilidzi Marwala. (2012). Condition Monitoring Using Computational Intelligence Methods. https://doi.org/10.1007/978-1-4471-2380-4

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

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

Yuwono, E. I., & Antonio, T. (2015). Studi Format Audio dan Teks Untuk Modul Speech to Text Studi Format Audio dan Teks Untuk Modul Speech to Text. Juisi, 01(01), 11–20.