JURNAL INFOTEL
Informatics - Telecommunication - Electronics
Website Jurnal : http://ejournal.st3telkom.ac.id/index.php/infotel
ISSN : 2085-3688; e-ISSN : 2460-0997

Classification of Javanese Script Hanacara Voice Using Mel Frequency Cepstral Coefficient MFCC and Selection of Dominant Weight Features

Heriyanto[1*], Tenia Wahyuningrum[2], Gita Fadila Fitriana[3]
[1]Universitas “Veteran” Yogyakarta
[2,3]Institut Teknologi Telkom Purwokerto
[1]02 Babarsari Street, Yogyakarta, Indonesia
[2,3]128 D.I Panjaitan Street, Purwokerto, Banyumas, Indonesia
*Corresponding email: heriyanto@upnyk.ac.id

Received 21 April 2021, Revised 18 May 2021, Accepted 30 May 2021

Abstract — This study investigates the sound of Hanacaraka in Javanese to select the best frame feature in checking the reading sound. Selection of the right frame feature is needed in speech recognition because certain frames have accuracy at their dominant weight, so it is necessary to match frames with the best accuracy. Common and widely used feature extraction models include the Mel Frequency Cepstral Coefficient (MFCC). The MFCC method has an accuracy of 50% to 60%. This research uses MFCC and the selection of Dominant Weight features for the Javanese language script sound Hanacaraka which produces a frame and cepstral coefficient as feature extraction. The use of the cepstral coefficient ranges from 0 to 23 or as many as 24 cepstral coefficients. In comparison, the captured frame consists of 0 to 10 frames or consists of eleven frames. A sound sampling of 300 recorded voice sampling was tested on 300 voice recordings of both male and female voice recordings. The frequency used is 44,100 kHz 16-bit stereo. The accuracy results show that the MFCC method with the ninth frame selection has a higher accuracy rate of 86% than other frames.

Keywords – feature extraction, accuracy, cepstral, coefficient

I. INTRODUCTION

Speech recognition in languages has been widely used. Several models have been developed with a language approach in grammar or other methods with feature extraction and speech recognition system matching. Voice recognition through vocabulary, both linking and stopping the reading, has flexible rules. This rule becomes flexible because it depends on the length of breath and the rhythm of the reader's reading. The rules listed for shaping this language model are becoming more numerous and very complex. Common and widely used feature extraction models include the Mel Frequency Cepstral Coefficient (MFCC). Research on the sound of the Javanese language Hanacaraka features the right features to choose the best frame feature in checking the sound of the Javanese language script reading. Selection of the right frame feature is needed in speech recognition because certain frames have accuracy at their dominant weight, so it is necessary to match frames with the best accuracy. Before the feature selection stage is carried out, the first feature extraction is carried out with MFCC. The general and widely used feature extraction method uses the Mel Frequency Cepstral Coefficient (MFCC).

A. Feature Extraction Method

The voice recognition method that uses feature extraction is significant because the feature extraction results significantly affect the match results and pattern recognition checks. Research that uses feature extraction methods includes Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Code (LPC) [1]. Both methods have weaknesses and advantages in feature extraction that produces features. The Problem of extraction feature with the MFCC method has an accuracy of 50% to 60%. In comparison, the Linear Predicted Code (LPC) is only 45% to 50%, so the non-linear MFCC method has higher accuracy than the linear approach method with the LPC method.

MFCC has weaknesses, including low frequency, environmental noise, sensitivity, almost similar sound patterns, and classification [2]. Meanwhile, MFCC has advantages, including capturing voice characteristics that are important in recognition, capturing critical information in voice, producing minimal data without losing information, and replicating the sound of human hearing [3]. In addition, feature extraction using MFCC is widely used for speech recognition because it is more precise in various conditions [4]. The feature extraction method using Linear Predictive Code (LPC) has weaknesses, including noise, changing speech frequency, and classification [5]. This method has the advantage of autocorrelation [1], [6].

The research of sound feature extraction using both MFCC and LPC has the same weaknesses, including noise, almost similar speech frequencies, frequently changing frequencies, and classification. The weakness of these two methods was also revealed by [1] that feature extraction using MFCC and LPC is not suitable for recognizing huge numbers of sounds, so classification is needed.

Based on the weaknesses and strengths of the two methods, both feature extraction using MFCC and or LPC, the researchers prefer feature extraction using MFCC because the level of accuracy is better than LPC [1][7][8]. MFCC feature extraction was between 58-75% [7]. In addition, the LPC method, research by [9] is more suitable for linear computations, whereas the human voice is essentially non-linear.

Another study related to language regarding hijaiyah letter recognition by Bethaningtyas [13] used MFCC by comparing 3, 6, 9, and 12 channels of the training data model and the deviation values. Another study related to hijaiyah letters by Heriyanto [14] used the method of average energy and waved deviation as a comparison. Meanwhile, other research related to the hijaiyah letter phoneme by Subali et al., [15] using the LPC and DTW methods resulted in the formant frequency. Speakers in pronunciation and DTW have the advantage of autocorrelation.

Another MFCC research by modifying was carried out by [16] in the windowing section. Other research also by modifying MFCC resulted in 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 [17].

B. *Matching Speech Recognition*

Matching research on speech recognition using different methods produces different outputs, including through neural networks [18], Hidden Markov Model (HMM) [4], speech recognition with Dynamic Time Wrapping (DTW) [20].

Speech recognition research using the feature extraction method MFCC has been widely carried out in all fields, including applied language. Research on speech recognition in the field of Arabic by [4] states that the extraction of Mel Frequency Cepstral Coefficients (MFCC) in the form of a feature to get the conformity value of Indonesian speakers to native speakers is classified using matching with the Hidden Markov Model (HMM).

MFCC is applied in other language fields, including Indonesian, by identifying speech signals into vocabulary, resulting in Phoneme and Syllable Models and segmentation [10]. Similar research was conducted by [11] using the Mel Frequency Cepstral Coefficient (MFCC) and Hidden Markov Model (HMM), which can recognize phoneme segmentation in Indonesian. Similar research on phonemes was also carried out by Cahyarini [12], who identified speech pauses between phonemes.

Voice recognition using the DTW method was carried out to calculate the distance between two-time series data [20]. This method has the advantage of calculating the distance between two data vectors of different lengths or knowing the value of the smallest matching distance between the voices of novice speakers and expert speakers [9].

DTW is an algorithm as a non-linear sequence alignment used to measure the similarity of a pattern in a time-variable and more realistic data series area for matching. DTW has a weakness in terms of accuracy, namely with wildly varied results [21], and still equals the accuracy level of HMM [4]. Meanwhile, the use of the HMM method, based on [11], has a weakness in terms of less resistance or robustness.

Another speech recognition method using Neural Network (NN) has advantages in learning systems, knowledge acquisition, classification, and generalization patterns [18]. According to [22], NN has a weakness in the training process, which requires a long time with a large amount of data. The same statement by [7] to identify the number one to nine utterances has a problem when the training process with massive data requires a very long processing time.

II. RESEARCH METHODS

A. *Mel Frequency Cepstral Coefficients (MFCC)*

The research method is divided into two major parts, MFCC feature extraction and Normalization of Dominant Weight. An explanation of the steps in each section can be seen in Fig.1.

Figure 1 shows the research of sound feature extraction using MFCC. It produces features in frame and cepstral coefficient parameters. It is processed into dominant weight normalization feature selection starting from the threshold, range, filtering, eliminating weight duplication, weight loss, and weight loss normalization to dominant weight.

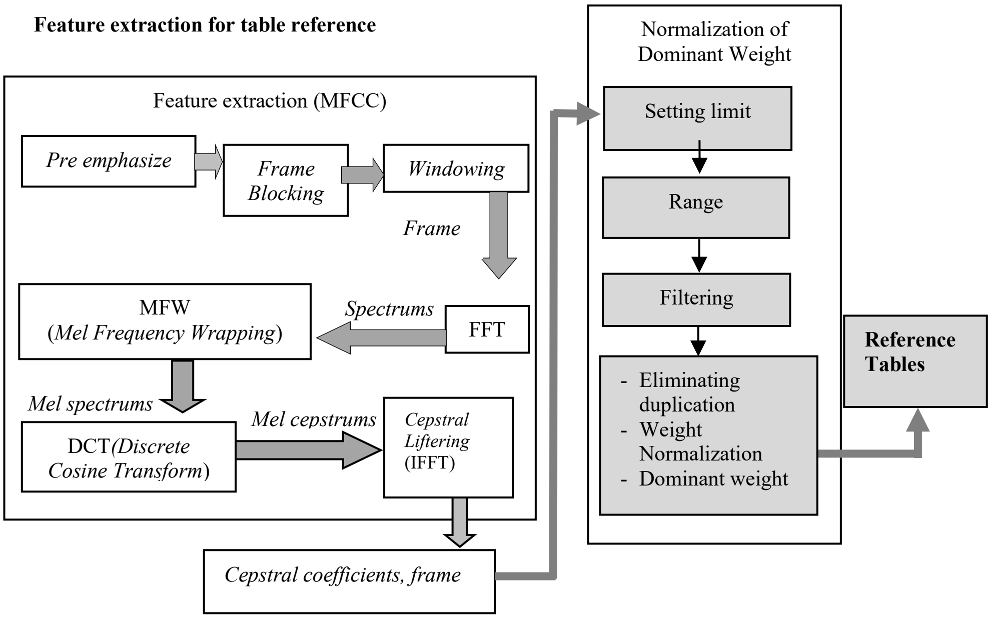

Feature extraction for table reference
Feature extraction (MFCC)
Pre emphasize
Frame Blocking
Windowing
Frame
MFW (Mel Frequency Wrapping)
Spectrums
FFT
Mel spectrums
Mel cepstrums
DCT(Discrete Cosine Transform)
Cepstral Liftering (IFFT)
Cepstral coefficients, frame
Normalization of Dominant Weight
Setting limit
Range
Filtering
- Eliminating duplication
- Weight Normalization
- Dominant weight
Reference Tables

Fig.1. Feature Extraction and Selection of Dominant Weight Normalization [31]

The MFCC method was first introduced by Davis and Mermelstein around 1980. MFCC is a method that is quite good in speech recognition in the field of speech recognition [23]. MFCC is the feature extraction that is most widely used in speaker recognition and speech recognition.

MFCC is a feature extraction that produces features or features that differentiate one another in the cepstral coefficient parameter [1]. Feature extraction of the Mel Frequency Cepstral Coefficient (MFCC) converts sound waves into several parameters, such as the cepstral coefficient representing audio files [4]. In addition, MFCC produces vector features that convert voice signals into several vectors for speech feature recognition [20].

MFCC has stages, namely pre-emphasis, frame blocking, windowing, Fast Fourier Transform (FFT), Mel Frequency Wrapping (MFW), Discrete Cosine Transform (DCT), and cepstral liftering, which produce parameters like features, namely the frame and cepstral coefficient [20].

B. Pre-emphasis

According to Tokunbo [24], pre-emphasis is an early-stage process and a very simple way to do it. The signal often experiences noise or noise interference to improve the Signal to Noise Ratio (SNR). Pre-emphasis has the aim that the high-frequency part still has good signal quality and is still in the realm of time [25]. Pre-emphasis according to [26] with α values between 0 to 1 or between $0.9 \leq \alpha \leq 1.0$ using (1)

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

In this case, the $y\ (n)$ symbol is the signal of the pre-emphasis result. In contrast, $s\ (n)$ is the signal symbol before pre-emphasis, the n symbol is the serial number of the signal, α is the pre-emphasis filter constant between 0.9-1.0, and s are signals. Taking the nth signal on pre-emphasis is carried out along the reading of one syllable with a time of one to two seconds.

C. Frame Blocking

The frame blocking process is blocked in a frame with N samples and shifted by M samples so that $N = 2M$ with $M < N$. Figure 1 shows an illustration of frame blocking [1]. The width of the frames is denoted by N, while the width of the shift for each frame is as M. The overlap width is calculated by the $N - M$ difference.

The average shooting time is between 20-40 milliseconds [4]. Frames are taken as long as possible to get a good frequency resolution, while the shortest possible time is intended to get the best time domain. Calculation of the number of frame blocking using equation (2)

$$f_l(n) = y(M_l) + n \qquad (2)$$

In this case, symbol $f_l(n)$ is the result of frame blocking, symbol n is $0,1,\ldots N-1$. The symbol N represents the number of samples, M is the frame length, l is $0,1,\ldots\ L-1$. The symbol L represents all signals, and y is the result of pre-emphasis.

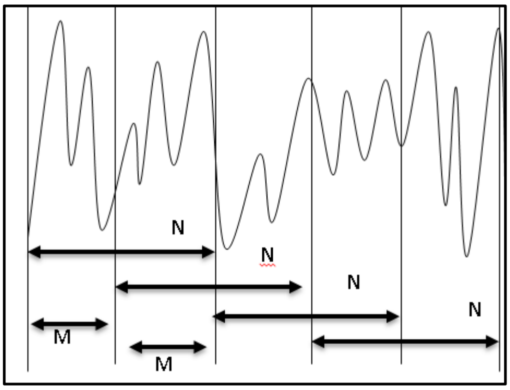


Fig. 2. Frame blocking illustration [1]

Figure 2 shows M is the first frame of the sound signal in the formula symbolized by f_l then $M + M = N$.

D. *Windowing*

Windowing aims to reduce discontinuation effects at the edges of the frame generated by the frame blocking process. Windowing used is the Rectangular Window, Hamming Window, and Hanning Window [4]. The researcher uses Hanning windowing of the three windowing functions because it is smoother than the others [27]. Representation of the windowing function using (3)

$$X(n) = f_l(n)w(n) \qquad (3)$$

In this case, the function $X\ (n)$ is the windowing signal, where f_l is the result of frame blocking, where n is $0,1,\dots,N-1$. The symbol N is the number of samples in each frame, and $w\ (n)$ is the window function. While the Hanning windowing function uses (4)

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

In this case, $w\ (n)$ is the window function using hanning, where n is $0,1,\dots,M-1$, M is the length of the frame.

E. *Fast Fourier Transform (FFT)*

Fast Fourier Transform is developing a fast algorithm to implement Discrete Fourier Transform (DFT), which converts digital signals in the time domain to the frequency domain [1] algorithm developed by Cooley and Turkey.

DFT computation time is too long and inefficient then FFT can perform efficiency calculations. Proakis and Manolakis [26] stated that the FFT method is efficient in calculating DFT. Discrete Fourier Transform (DFT) using (5).

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

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

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

In this case, the function T (m) is the result of the math Fast Fourier Transform calculation, the symbol $X\ (n)$ is the result of the nth windowing calculation. The n symbol is the signal serial number. The m symbol is the index of the frequency $(1,2,\dots\ N)$.

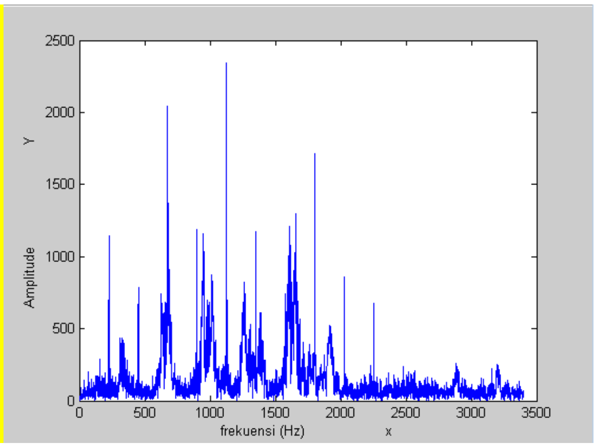


Fig 3. Fast Fourier Transform (FFT)

Figure 3 shows the results of the FFT processing. FFT has a frequency domain and generates a spectrum.

F. *Mel Frequency Wrapping (MFW)*

Mel Frequency Wrapping (MFW) is a filter in the form of a filter bank to determine the energy size of a particular frequency band in sound signals [19][20] MFW, according to Laha [28], converts the frequency into Mel Frequency Wrapping (MFW).

Filterbank has a frequency response via a triangular-shaped path whose distance and constant frequency intervals determine size. The output process obtained from the filter is known as the mel spectrum using (7).

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

In this case, the symbol Y [i] results from the calculation of MFW it-i, where G is the total magnitude spectrum (GN). Then 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 filterbank. The approach used is in the form of MFW using (8).

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

In this case, MFW uses a frequency with the MFW scale, f as the frequency. MFW produces a mel spectrum. MFW frequency scale is a linear frequency scale at frequencies below 1,000 Hz and is a logarithmic scale at frequencies above 1,000 Hz [20].

G. Discrete Cosine Transform (DCT)

DCT, according to Smith [29], is a relative of the Fourier transform that decomposes the signal to the cosine wave. DCT has been widely used in sound and image processing, for example, JPEG or BMP files. The concept of DCT is similar to the inverse Fourier transform, and DCT is close to the Principal Component Analysis (PCA) method, which is a classical static that is widely used in data analysis and compression.

DCT can be assumed to replace the inverse Fourier transform in the MFCC feature extraction process [20]. Discrete Cosine Transforms (DCT) is a member of the sinusoidal unit transformation class [30]. DCT aims to produce spectrum mel to improve the quality of recognition. DCT uses (9).

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

In this case, C_m is the coefficient, where $Y_{[i]}$ is the output of the filter bank process on the index, r is the number of coefficients, and K is the expected number of coefficients. The DCT process produces spectrum mel.

H. Cepstral Liftering

Cepstral liftering increases the accuracy of pattern matching, speech recognition, and speech recognition [20]. The cepstral coefficient uses equation (10)

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

In this case, the symbol $w(k)$ is the window function of the cepstral features, C is the cepstral coefficients, the k symbol is the index of the cepstral coefficients. The cepstral liftering processing results in the form of frames and cepstral coefficients then processed to feature selection. The feature selection is described in Chapter V of the feature selection model.

I. Selection of Dominant Weight Normalization Feature

The MFCC feature extraction result is the frame and cepstral coefficient, which significantly influences speech matching and recognition. The selection of features is carried out using the dominant weight normalization model. The model has six stages, namely determining the threshold, making the range, filtering, eliminating the duplication of weights, normalizing the weight and dominant weight resulting in a table of features [31].

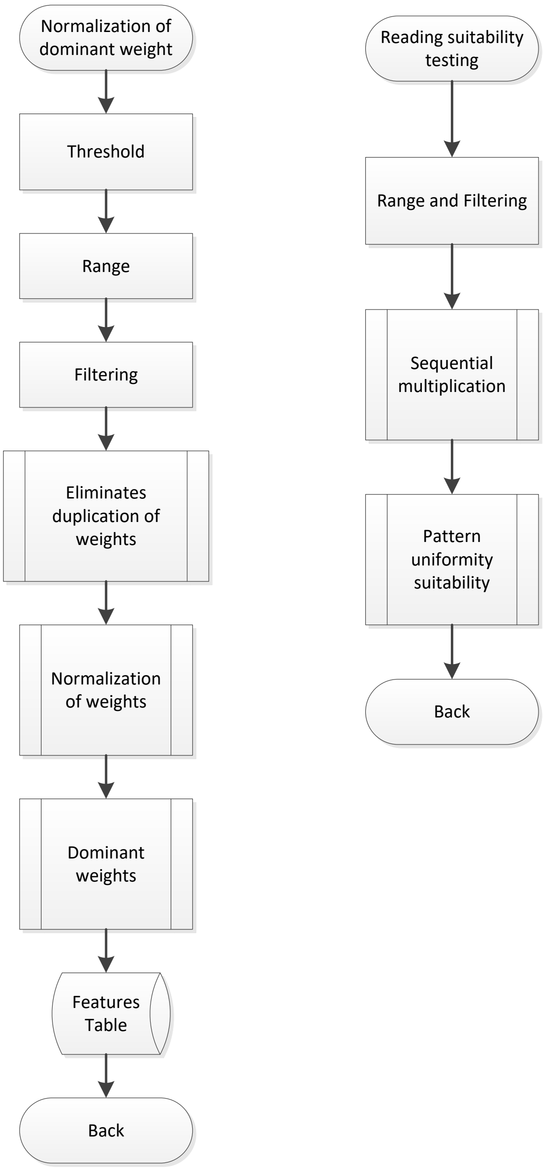

Normalization of dominant weight
Threshold
Range
Filtering
Eliminates duplication of weights
Normalization of weights
Dominant weights
Features Table
Back
Reading suitability testing
Range and Filtering
Sequential multiplication
Pattern uniformity suitability
Back

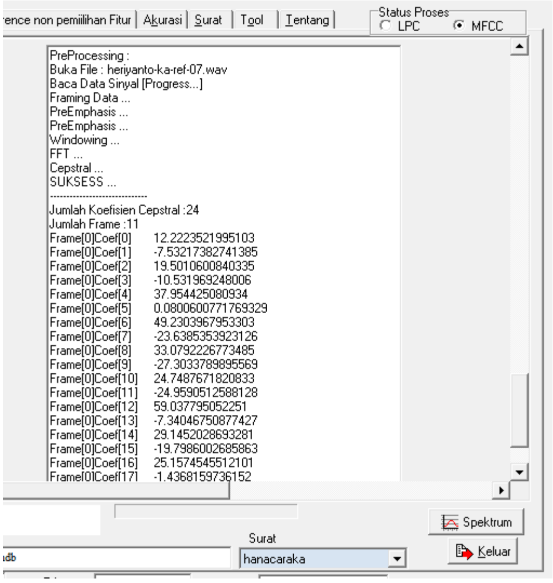
Fig. 4. Normalization of Dominant Weight And Suitability testing

Normalization of dominant weights and conformity testing. Figure 4 shows the feature selection carried out starting with MFCC feature extraction. The results of the feature extraction are then matched by testing the dominant weight normalization algorithm.

III. RESULTS

The test results on selecting the right features were carried out on the number of cepstral coefficients and number of frames. In contrast, the Characteristic Extraction Results using MFCC produced frame and cepstral coefficients with eleven framed and 24 cepstral coefficients. It can be seen in Fig.5 and Fig.6.

Figure 5 shows the results of the MFCC output in the form of a cepstral coefficient of twenty-four cepstral coefficient and a frame.

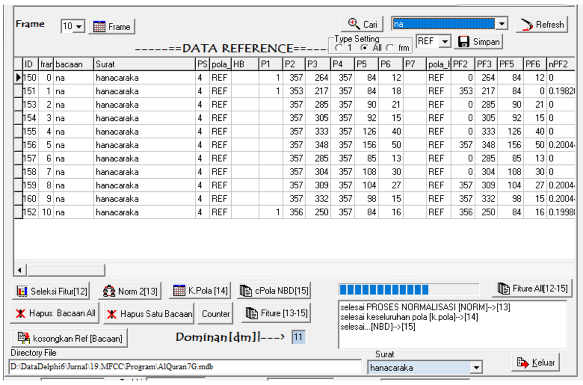


```
PreProcessing :
Buka File : heriyanto-ka-ref-07.wav
Baca Data Sinyal [Progress...]
Framing Data ...
PreEmphasis ...
PreEmphasis ...
Windowing ...
FFT ...
Cepstral ...
SUKSESS ...
-----------------------------
Jumlah Koefisien Cepstral :24
Jumlah Frame :11
Frame[0]Coef[0]     12.2223521995103
Frame[0]Coef[1]     -7.53217382741385
Frame[0]Coef[2]     19.5010600840335
Frame[0]Coef[3]     -10.531969248006
Frame[0]Coef[4]     37.954425080934
Frame[0]Coef[5]     0.0800600771769329
Frame[0]Coef[6]     49.2303967953303
Frame[0]Coef[7]     -23.6385353923126
Frame[0]Coef[8]     33.0792226773485
Frame[0]Coef[9]     -27.3033789895569
Frame[0]Coef[10]    24.7487671820833
Frame[0]Coef[11]    -24.9590512588128
Frame[0]Coef[12]    59.037795052251
Frame[0]Coef[13]    -7.34046750877427
Frame[0]Coef[14]    29.1452028693281
Frame[0]Coef[15]    -19.7986002685863
Frame[0]Coef[16]    25.1574545512101
Frame[0]Coef[17]    -1.4368159736152
```

Surat: hanacaraka

Fig. 5. *Output Frame and Cepstral Coefficient*

Figure 6 shows the results of the MFCC greeting "hanacaraka" consisting of eleven frames and twenty-four cepstral coefficients.



Frame: 10

na

Type Setting ○ 1 ● All ○ frm

REF

-----==DATA REFERENCE==----

ID	fran	bacaan	Surat	PS	pola_	HB	P1	P2	P3	P4	P5	P6	P7	pola_I	PF2	PF3	PF5	PF6	nPF2
150	0	na	hanacaraka	4	REF		1	357	264	357	84	12		REF	0	264	84	12	0
151	1	na	hanacaraka	4	REF		1	353	217	357	84	18		REF	353	217	84	0	0.1982
153	2	na	hanacaraka	4	REF			357	285	357	90	21		REF	0	285	90	21	0
154	3	na	hanacaraka	4	REF			357	305	357	92	15		REF	0	305	92	15	0
155	4	na	hanacaraka	4	REF			357	333	357	126	40		REF	0	333	126	40	0
156	5	na	hanacaraka	4	REF			357	348	357	156	50		REF	357	348	156	50	0.2004
157	6	na	hanacaraka	4	REF			357	285	357	85	13		REF	0	285	85	13	0
158	7	na	hanacaraka	4	REF			357	304	357	108	30		REF	0	304	108	30	0
159	8	na	hanacaraka	4	REF			357	309	357	104	27		REF	357	309	104	27	0.2004
160	9	na	hanacaraka	4	REF			357	332	357	98	15		REF	357	332	98	15	0.2004
152	10	na	hanacaraka	4	REF		1	356	250	357	84	16		REF	356	250	84	16	0.1998

```
selesai PROSES NORMALISASI [NORM]-->[13]
selesai keseluruhan pola [k.pola]-->[14]
selesai...[NBD]-->[15]
```

Dominan[dm]]---> 11

Directory File: D:\DataDelphi6\Jurnal\19.MFCC\Program\AlQuran7G.mdb

Surat: hanacaraka

Fig.6. *The Results of the MFCC Frame and Cepstral Coefficients*

A. *Feature Selection*

The proposed feature selection has six stages, namely determining the same threshold, creating the same range, filtering, eliminating duplication of weights, normalizing the weight and dominant weight [31] [32]. All these steps are applied in the dominant of weight normalization feature selection algorithm to produce a feature table.

Table 1. Results of MFCC Greeting Hanacaraka

File code	Frame to	Reading	Cofficient to	Value Coefficient Cepstral
245	0	ha	0	65.2
246	1	ha	1	56.3
247	2	ha	2	12.2
252	3	ha	3	-12.3
253	0	na	0	16.93
254	1	na	1	21.75
255	2	na	2	32.24
256	3	na	3	-3.84
266	0	ca	0	13.36
267	1	ca	1	4.06
268	2	ca	2	-0.22
269	3	ca	3	19.86
270	0	ra	0	-17.17
271	1	ra	1	17.35
272	2	ra	2	23.66
273	3	ra	3	-9.09
274	0	ka	0	20.83
275	1	ka	1	23.30
276	2	ka	2	12.13
277	3	ka	3	23.12

Table 1 shows the results of feature extraction using the MFCC saying "ha, na, ca, ra, ka. The result of feature extraction is a frame consisting of eleven frames and a twenty-four cepstral coefficient. The next process of the frame and cepstral coefficient results is carried out by selecting features by selecting frame features with feature selection algorithms using Dominant Weight Normalization.

Figure 6 shows the dominant weight normalization algorithm, which starts from the first and second steps of taking speech sounds; the third step is carrying out the threshold process. Then, in the fourth step and so on, coverage, filtering, eliminating duplication of weights, normalization of dominant weights and weights are carried out.

B. Data Collection

Sound sampling was carried out as many as 300 recorded voices and tested on 200 recorded voices. The cepstral coefficient, ranging from 0 to 23, is 24 cepstral coefficient and 0 to 10 frames. The frequency used is 44,100 kHz 16bit stereo by recording male voices and female voices.

Table 2. Voice recording data training

Number	Gender	Voice	The number of sampling
1	Male	Ha	30
2	Male	Na	30
3	Male	Ca	30
4	Male	Ra	30
5	Male	Ka	30
1	Female	Ha	30
2	Female	Na	30
3	Female	Ca	30
4	Female	Ra	30
5	Female	Ka	30
			300

```
Dominant Weight Normalization Feature Selection Algorithm pseudocode
0.  Start
1.  Take voice, say
2.  Mfcc Functiion with frame = f, cepstra coefficient
      (c)=w(k)
3.  Determine threshold 1 until threshold 6 (b1-b6)
        a.  Create b1 $= min(\text{w(k)})$,
        b.  Create b2 $= \frac{\min(w(k))+\left(\frac{\min(w(k))+\max(w(k))}{2}\right)}{2}$
        c.  Create b3 $= rata - rata(w(k))$,
        d.  Create b4 $= \frac{\min(w(k))+\max(w(k))}{2}$
        e.  Create b5 $= \frac{\left(\frac{\min(w(k))+\max(w(k))}{2}\right)+\max(w(k))}{2}$,
        f.  Create b6 $= max\ (w(k))$.
4.   Make the range
 range by checking the conditions in each frame cepstral coefficient(c)=w(k)
            Weight=1
  a.Rule1 is if (b1)  min=w(k) then pi1 =weight elseif weight=0
  b.Rule2 is if (b1) ≥ = w(k) and (w(k) <b2) then pi2 = weight elseif bobot=0
  c.Rule3 is if (b2) ≥ = w(k) and (w(k) <b3) then pi3 = weight elseif bobot=0
  d Rule4 is if (b3) ≥ = w(k) and (w(k) <b4) then pi4 = weight elseif bobot=0
  e.Rule5 is if (b4) ≥ = w(k) and (w(k) <b5) then pi5 = weight elseif bobot=0
  f.Rule6 is if (b5) ≥ = w(k) and (w(k) <b6 ) then pi6= weight elseif bobot=0
  g Rule7 is if (b6) max= w(k) then pi7 = weight elseif weight =0
5.  Filtering
      The filtering results are in the form of pi1 until pi7
      Count the number pi1 until pi7 with $G_j = \sum_{i=0}^{F} p_{ij}$
      Count the total number of patterns with $U = \sum_{j=0}^{a} G_j$
6.  Eliminates duplication of weights
Specify that Qij=pij and looking for similarities eliminating duplications
      if Qi0=pi1 then Qij=0, if Qi1=pi2 then Qjj=0, if Qi2=pi3 then Qij=0
      if Qi0=pi2 then Qij=0, if Qi1=pi3 then Qjj=0, if Qi2=pi4 then Qij=0
      if Qi0=pi10 then Qij=0, if Qi1=pi10 then Qjj=0, if Qi2=pi10 then Qij=0
g.    Calculation of the amount with $Z_j = \sum_{i=0}^{F} Q_{ij}$
      Normalized Weights
      Calculation of Normalized Weight with $S_j = \sum_{i=0}^{F} \frac{Q_{ij}}{Z_j}$
h.  Dominant Weight
      Sort the largest value Sj in the feature table (npf2, npf3, npf5, npf6) into variabel Bj
 i.   Save the table of features calculation of total
        weightsSimpan tabel fitur = Z,
        Normalized weights =Sj, the whole pattern=U
7. Done
```

Fig. 6. Dominant Weight Normalization Feature Selection Algorithm Pseudocode [31][32]

Table 2 and Table 3 sampling of votes for each voice of speech reading 37 samplings, both male and female. The total number of votes was 296 votes.

Table 3. Record Test Data

Number	Gender	Voice	The number of sampling
1	Male	Ha	20
2	Male	Na	20
3	Male	Ca	20
4	Male	Ra	20
5	Male	Ka	20
1	Female	Ha	20
2	Female	Na	20
3	Female	Ca	20
4	Female	Ra	20
5	Female	Ka	20
			200

IV. DISCUSSION

Checking the suitability of Javanese hanacaraka words is carried out on the feature table by calculating the Uniformity of Pattern Conformity. The word suitability results calculation is expected to be close to the results in the table of features so that the percentage of speech suitability increases.

The existing feature table threshold is checked with similarity range and filtering, sequential calculations, and Uniformity of Pattern Conformity calculations using equations as in the algorithm. Speech checking is performed to select the correct reference and select the right features [31] [32].

Figure 7 shows the algorithm for checking the sound suitability of Javanese script speech by taking the feature results from MFCC feature extraction in the form of frame parameters and cepstral coefficient. Algorithm for conformity checking through range checking, filtering each frame, then sequential multiplication, and the final calculation of Pattern Uniformity Conformity. The results of the frame and cepstral coefficient can be seen in Table 4.

Based on Table 4, selecting the best frame features on the 9th frame with an average value of 85% better than other frames.

V. CONCLUSION

This study uses MFCC feature extraction and frame feature selection with the appropriate Dominant Weight Normalization algorithm for reading Javanese script language Hanacaraka, producing frame and cepstral coefficient. The cepstral coefficient ranging from 0 to 23 is 24 cepstral coefficients. A sampling of sounds as many as 296 recorded voices tested against 200 recorded voices. The frequency used is 44,100 kHz 16-bit stereo. The accuracy results show that the MFCC method with the 9th frame feature selection has a higher accuracy rate of 86% than other frames. The frame feature with the 9th because the feature does not minimalize and not maximum. This frame medium or intermediate.

Table 4. Hanacaraka Speech Results

No	Saying	Sampling	MFCC	Frame (%)										
				0	1	2	3	4	5	6	7	8	9	10
1	Ha	40	60	20	10	30	40	30	40	30	40	50	70	50
2	Na	40	90	30	40	50	70	60	70	50	40	70	90	70
3	Ca	40	90	40	50	70	60	80	50	70	50	80	100	90
4	Ra	40	80	50	50	70	50	60	70	60	60	80	90	90
5	Ka	40	80	60	60	60	70	70	80	70	80	80	80	80
		200	80	40	42	56	58	60	62	56	54	72	86	76

```
Hanacaraka Speech Conformity Checking Algorithm
0. Start
1. Take the sound of the greeting greeting
2. MFCC function with frames =f, cepstral coefficient(c)=w(k)
3. Checking the range and filtering of each frame i=0,1,2,...,F
    a. Take a table of features for which there is a threshold ($b_1$-$b_6$)
    b. Check the range or range with w(c) if it is met $p_{ij}$ given a weight =1
    c. rule$_1$ with if ($b_1$ min = w(k)) then $R_{i1}$ =weight elseif weight=0
    d. rule$_2$ with if ($b_1$ >= w(k)) and (w(k)<b2) then $R_{i2}$ =weight elseif weight=0
    e. rule$_3$ with if ($b_2$ >= w(k)) and (w(k)<b3) then $R_{i3}$ =weight elseif weight=0
    f. rule$_4$ with if ($b_3$ >= w(k)) and (w(k)<b4) then $R_{i4}$ =bobot elseif bobot=0
    g. rule$_5$ with if ($b_4$ >= w(k)) and (w(k)<b5) then $R_{i5}$ =weight elseif weight=0
    h. rule$_6$ with if ($b_5$ >= w(k)) and (w(k)<b6) then $R_{i6}$ =weight elseif weight=0
    i. rule$_7$ with if ($b_6$ = w(k)) then $R_{i7}$ =weight elseif weight=0
    j. Filtering save accumulated filtering with$R_{ij}$, j= 2, 3, 5 and 6.
4. Sequential multiplication with feature table
Calculate sequential multuplicatiion $P_{ij,}$ dengan j=2,3,5 dan 6
    $P_j = \sum_{i=0}^{F} (R_{ij} B_{ij} Z_{ij}^2)$
5. Calculate with  formula KKP P2,P3, P5 and P6
    $K_j = \frac{\overline{P_j}\,\overline{U}}{\overline{Z_j^2}}$        $K = \frac{(\overline{P_2} + \overline{P_3} + \overline{P_5} + \overline{P_6})\overline{U}}{\overline{Z_2^2} + \overline{Z_3^2} + \overline{Z_5^2} + \overline{Z_6^2}}$
6. Done
```

Fig 7. Algorithm for Checking The Suitability of Greetings [31][32]

APPENDIX

APPENDIX 1

No	**Saying**	**Sampling**	**MFCC**	**Frame**										
				0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	Ha	40	60	20	10	30	40	30	40	30	40	50	70	50
2	Na	40	90	30	40	50	70	60	70	50	40	70	90	70
3	Ca	40	90	40	50	70	60	80	50	70	50	80	100	90
4	Ra	40	80	50	50	70	50	60	70	60	60	80	90	90
5	Ka	40	80	60	60	60	70	70	80	70	80	80	80	80
		200	80	40	42	56	58	60	62	56	54	72	86	76

REFERENCES

[1] Abriyono and A. Harjoko, “Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST,” *Indones. J. Comput. Cybern. Syst.*, vol. 6, no. 2, pp. 23–34, 2012.

[2] F. Syafria, A. Buono, and B. P. Silalahi, “Pengenalan Suara Paru-Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier”, J. Ilmu Komput Agri-Inf, vol. 3, no. 1, pp. 27-36, Jan. 2017.

[3] H. S. Manunggal, “Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara Dengan Menggunakan Analisa MFCC Feature Extraction.,” *Tugas Akhir Sarj. pada Jur. Tek. Inform. Fak. Teknol. Ind. Univ. Kristen Petra Surabaya*, 2005.

[4] T. Chamidy, “Metode Mel Frequency Cepstral Coeffisients (MFCC) Pada klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada Penutur Indonesia,” *Matics*, vol. 8, no. 1, pp. 36–39, 2016.

[5] Irmawan, H. Hikmarika, D. W. Sari, and M. C. Tammimi, “Pengenalan Kata dengan Metode Linear Predictive Coding dan Jaringan Syaraf Tiruan Pada Mobile Robot,” in CITEE 2014, Yogyakarta, 2014.

[6] Thiang, H. Saputra “Sistem Pengenalan Kata dengan Menggunakan Linear Predictive Coding dan Nearest Neighbor Classifier,” *Jurnal Teknik Elektro*, vol. 5, no.1, pp. 19–24, Maret, 2005.

[7] A. M. Aibinu, M. J. E. Salami, A. R. Najeeb, J. F. Azeez, and S. M. A. K. Rajin, “Evaluating the effect of voice activity detection in isolated Yoruba word recognition system,” *in 2011 4th Int. Conf. Mechatronics Integr. Eng. Ind. Soc. Dev. ICOM’11 - Conf. Proc.*, no. May, pp. 17–19, 2011, doi: 10.1109/ICOM.2011.5937134.

[8] S. Hidayat, R. Hidayat, and T. B. Adji, “Sistem Pengenal Tutur Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM,” *in Conf. Inf. Technol. Electr. Eng.*, no. September, pp. 246–251, 2015.

[9] S. M. Widodo, E. Siswanto, and O. Sudjana, “Penerapan Metode Mel Frequency Ceptral Coefficient dan Learning Vector Quantization untuk Text-Dependent Speaker Identification,” *Jurnal Telematika*, vol. 11, no. 1, pp. 15–20, 2016.

[10] Suyanto and S. Hartati, “Design of Indonesian LVCSR using Combined Phoneme The Approaches of LVCSR,” *Icts*, pp. 191–196, 2013.

[11] S. Suyanto and A. E. Putra, “Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation,” *J. ICT Res. Appl.*, vol. 8, no. 2, pp. 97–112, 2014.

[12] R. Cahyarini, U. L. Yuhana, and A. Munif, “Rancang Bangun Modul Pengenalan Suara Menggunakan Teknologi Kinect,” *J. Tek. Pomits*, vol. 2, no. 1, pp. 1–5, 2013.

[13] M. K. A. S. . dan S. Hertiana Bethaningtyas, “Pengenalan Huruf Hijayyah Berbasis Pengolahan Sinyal Suara dengan Metode MFCC,” *Momentum,* vol.13, no. 2, pp. 49–52, Oktober,2017.

[14] Heriyanto, “Analisa Deteksi Huruf Hijaiyah Melalui Voice Recognition Menggunakan Kombinasi Energy,” *Telematika*, vol. 12, no. 01, pp. 11–22, 2015.

[15] M. Subali, M. Andriansyah, and C. Sinambela, “Analisis Frekuensi Dasar dan Frekuensi Formant dari Fonem Huruh Hijaiyah Untuk Pengucapan Makhraj Dengan Metode DTW,” *in Pros. PESAT (Psikologi, Ekon. Sastra, Arsit. &Teknik Sipil), Depok, 2015, pp. -60-72.*

[16] C. Goh and K. Leon, "Robust Computer Voice Recognition Using Improved MFCC Algorithm," in 2009 *International Conference on New Trends in Information and Service Science*, 2009, pp. 835-840.

[17] L. Muda, M. Begam, and I. Elamvazuthi, “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques,” J. Computing, vol. 2, no. 3, pp. 138–143, 2010, [Online]. Available: http://arxiv.org/abs/1003.4083.

[18] M. W. . Sanjaya and Z. Salleh, “Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Adaptive Neuro-Fuzzy Inferense System (Anfis) Sebagai Kontrol Lampu Otomatis,” *Al-HAZEN J. Phys.*, vol. 1, no. 1, pp. 1–19, 2014.

[19] Y. Miftahuddin and M. R. Hakim, “Coefficient Dan Dynamic Time Warping Untuk Pengenalan Nada Pada Alat Musik Bellyra,” pp. 120–127, 2017.

[20] A. R. Darma Putra, “Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW,” *LONTAR*

Komput. Biometrika, vol. 2, no. 1, pp. 8–21, 2011.

[21] D. Novianto and R. V. Yuliantari, "Pengenalan Isyarat Tutur Vokal Bahasa Indonesia Menggunakan Metode Dynamic Time Wraping (Dtw) Berbasis Fungsi Jarak," Journal of Electrical Engineering, Computer and Information Technology, vol.1, no. 1, pp. 53–57, 2017.

[22] S. Martyna and S. Sudaryanto, "Penerapan Metode Particle Swarm Optimization pada Artificial Neural Network Backpropagation untuk Peramalan Penjualan Furniture pada CV. Octo Agung," Skripsi, Fakultas Ilmu Komputer, UDINUS, 2015.

[23] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoust.*, vol. 28, no. 4, pp. 357–366, 1980.

[24] R. T. M. (Sim) narasimha Tokunbo Ogunfunmi, *Speech and Audio Processing and Recognition*, no. part 1. springer, 2015.

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

[26] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, algorithms, and applications*. 1996.

[27] A. E. Putra, "Frekuensi Cuplik pada FFT," *Tan Li, Process. Digit. Signal*, vol. 1, 2008.

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

[29] S. W. Smith, *Digital signal processing*, vol. 17, no. 2. 2000.

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

[31] Heriyanto, S. Hartati, and A. E. Putra, "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)," *Adv. Image Video Process.*, vol. 6, no. 2, pp. 16-35, 2018.

[32] Heriyanto, "Good Morning to Good Night Greeting Classification Using Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction and Frame Feature Selection," *Telematika: Jurnal Informatika dan Teknologi Informasi*,vol. 18, no. 1, pp. 88–105, 2021.