MUSICAL INSTRUMENT RECOGNITION USING MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) AND LEARNING VECTOR QUANTIZATION (LVQ) IN CLASSIFICATION SOUND SOURCES ON MUSIC INSTRUMENT

Sofiyanudin[1], Irfan Maliki[2]

[1,2] Universitas Komputer Indonesia
Jl. Dipatiukur No 112-116 Bandung. 40132
E-mail : sofiyanudin.first@gmail.com[1], irfanmaliki007@gmail.com[2]

ABSTRACT

Musical Instrument Recognition is the process used to recognize the sound of a musical instrument, in this study the sounds of musical instruments known, namely the sound of musical instruments based on the source of the sound. Based on the source of the sound of the resulting instrument is divided into five parts, among others, Aerophone, Elektrophone, Chordofone, Idiophone, a Membranophone. Because the results of the previous research to voice recognition with LVQ method and MFCC produces a good accuracy. Then on this research will be done with voice recognition a similar method to classify the source sound on the instrument. As for the goal of the research is to find out the perfomansi at the MFCC and LVQ method as the introduction of sound in classifying the source sound on the instrument.

Kata Kunci: *Musical Instrument Recognition*, Klasifikasi, LVQ, MFCC, Bunyi alat musik

1. INTRODUCTION

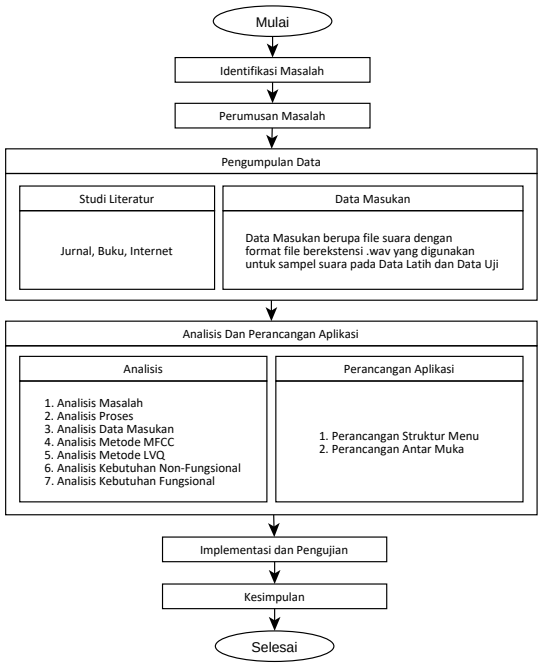
1.1 Problem

To be able to classify the sound of musical instruments, then the need for voice recognition, there are the methods of extraction and classification methods to be able to recognize the vote. Previous research has been done on the voice recognition with the method of classification of Learning Vector Quantization (LVQ) and Mel frequency cepstral method of extraction coefficients (MFCC) in some cases, such as the identification of lung sounds with accuracy 87.83% [2], multi language identification with 90% accuracy [3], the application for the transcription of voice into text with accuracy of 98.57% [4].

Because the results of the previous research to voice recognition with LVQ method and MFCC produces a good accuracy. Then on this research will be done with voice recognition a similar method to classify the source sound on the instrument. As for the goal of the research is to find out the perfomansi at the MFCC and LVQ method as the introduction of sound in classifying the source sound on the instrument.

1.2 Research methodology

Quantitative research methods can be interpreted as research methods based on positivism philosophy, used to examine the population or a particular sample, data collection using research instruments, quantitative data analysis / statistics, with the aim to test the hypothesis that has been determined [21].



1.3 Issue and Purpose

The purpose of this research is to the implementation of sound extraction method of MFCC and LVQ as voice recognition and the aim of this research is to know the perfomansi LVQ in classifying musical instruments based on the source of the sound.

1.4 Sound Sources

Based on the classification of musical instruments is divided into three sections, which are: based upon how to play musical instruments, based

upon the source of the sound of musical instruments, based upon the functions of the instrument [1].

1) *Aerophone is a musical instrument whose source of sound comes from the air that vibrates by being blown or in a pump. Examples: flute, flute, recorder, tube, melodeon, clarinet, saxophone and horn are played. The coordinates are pumped*

2) *Idiophone is a musical instrument whose source of sound comes from a metal or wooden rod being struck or the source of the sound comes from the instrument itself. Example: bellira, calung, angklung, kulintang, gamelan device*

.

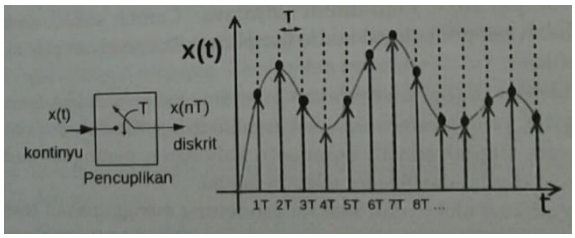
3) *Elektrophone is a musical instrument whose source of sound comes from a series of electronics contained in the tool. The current technological developments make this tool bias sound from various types of musical instruments. Examples: keyboard, Organ electric, Electric guitar, electric bass and others.*

4) *Membranophone is a musical instrument whose source of sound is sourced from the vibration on a thin membrane made of leather or plastic that sounds by being hit. Example: tambourine, tambourine, drum set, ketipung, bongo, konga, timpani and others.*

5) *Cordophone is a musical instrument whose source of sound comes from a series of strings that vibrate Because picked, swiped or pressed.*
   a. *In quotation: guitar, harp mandolin, ukulele, banjo, siter, kecapi bass and others.*
   b. *In friction: violin, viola, double bass, rebab*
   c. *On tap: acoustic piano [1]*

1.5 Analog And Digital Signals

Continuous and disctrit signals. The amplitude of the continuous signal is defined for all t times, while the amplitude of the disctrit signal is defined for a given t value. The disctrict signal is obtained by sampling a continuous signal at certain times. Sampling is done periodically with period T (Figure 2.1). Continuous signal sampling x (t) is done by disconnecting the switch with period T. to produce discrete signal x (nT).



Continuous signals can be expressed by x (t) and discrete signals represented by x (nT) or abbreviated x [n], with n integers and T sampling periods. The amplitude of the discrete signal x (nT) is undefined in T seconds, 2T, 3T and so on.
An analog signal is a continuous signal that has continuous amplitude within a certain range. In other words, these analog signals have the characteristic of time and continuous amplitudes.
Digital signals are discrete signals that have a discrete (limited) amplitude range. In other words, these digital signals have discrete timing and amplitudes. Discrete time is obtained by a process called quantization [5].

1.6 Process Analysis

To identify the problems that exist in classifying the sound source. The description of the process analysis as follows:

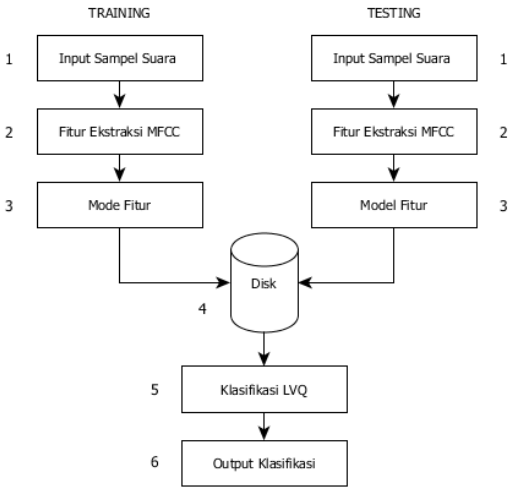


Figure 1. Process Analysis
The following is an explanation of the flow drawing process:

The following is an explanation of the flow drawing process:

1) The input stage of the sound sample data with the file is berektensi. Wav.
2) Phase perform feature extraction from sound samples using MFCC method. The result of the extraction feature is the value of the vector to be included in the feature model, the process flow can be seen in Figure 2.1 The flow of MFCC processes.
3) The Feature Model is a model performed at the coding stage, while the model contains

information from each sound sample consisting of filename, sample name, cardinal extraction, and acoustic vector.

4) From Model Feature saved to file with extension (.ser) on Internal Storage.
5) Classification process by LVQ method to know the sound source sound sample that existed in Feature Model.
6) 6.Reference of sound source class with Aerofon class, Electrophone, Idiofon, Kordofon, and Membranofon.

2. CONTENT

2.1 Analysis Dataset

The dataset is used as the parameter used in Train Data and Test Data. The input data is processed from the sound file data by extension (.wav). The value taken is a sample of sound results from processing analog data to digital data.

The type of sound used as a sample calculation is the Aerofon class with sample rate 8000, mono chanel and duration of 1 second. Results of processing analog to digital data can be seen in Figure 1 using Audacity.

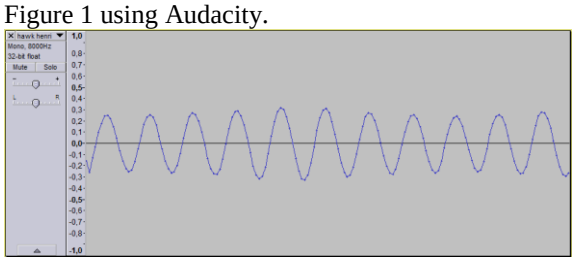


Figure 2. Dataset

2.2 MFCC

MFCC method used in extracting features of sound, the result of extraction is an acoustic vector value which is used as a parameter in the classification algorithm. In performing feature extraction using MFCC method consists of several stages, among others, Pre-Emphasis, Framing, Windowing, FFT (Fast Fourier Transform), Mel Filter Bank, DCT (Discrete Cosine Transform) [1] [2]..

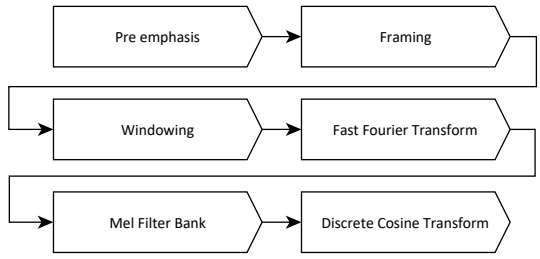


Figure 3. Process of MFCC

2.2.1 Pre-Emphasis

In this step will be process the signal through a filter that emphasizes the higher frequency. [5]. In the previous study α = 0.95 [5] The equation of pre-emphasis is in equation (1):

$$Sa[n] = s[n] - \alpha . s[n-1] \quad (1)$$

2.2.2 Framing

The process of segmenting the sound samples obtained from analog to digital conversion (ADC) into smaller frames with lengths in the range of 20 to 40 ms. The purpose of framing is to avoid signal discontinuity, Because signal discontinuity can cause extraction of wrong parameters during analysis. To avoid signal discontinuity on two consecutive frames, each frame overlaps each other [5]. Previous research has been done 20ms framing, overlaping 10ms [9].

The process of framing can be seen in Figure 2 [9]:

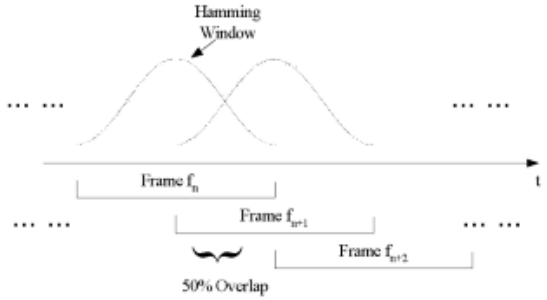


Figure 2. Process of Framing

As the equation of framing is in equation (2), (3, (4):

$$Ns = sample\ rate \times durasi\ framing \quad (2)$$

$$No = sample\ rate \times durasi\ overlapping \quad (3)$$

$$Nf = \frac{sample\ rate}{Ns} \quad (4)$$

2.2.3 Windowing

In the next step in processing MFCC is to do windowing, each individual frame will minimize signal discontinuity at the beginning and end of the frame. The concept is to minimize the spectral deviation by using a window by streaming the signal to zero at the beginning and end of each frame [11]. The equation of windowing is in equation (5), (6):

$$Wham[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{\mathrm{Ns} - 1}\right) \quad (5)$$

where $0 \leq n \leq Ns - 1$

$$Sb[n] = Sa[n] \times Wham[n] \quad (6)$$

where $0 \leq n \leq Ns - 1$

2.2.4 FFT

In the next step is FFT, changing each frame of sample N from time domain into frequency domain. FFT is a fast algorithm for implementing Discrete Fourier Transform (DFT) [11]. The equation of FFT is in equation (7), (8):

$$F[k] = \sum_{n=0}^{Ns-1} \left(\mathrm{Sb[n]} \cos\frac{2\pi nk}{Ns}\right) - j \sum_{n=0}^{Ns-1} \left(\mathrm{Sb[n]} \sin\frac{2\pi nk}{Ns}\right)$$

where $k = 0,1,2, \ldots, Ns - 1$

(7)

$$Sc[n] = \left|[R^2 + I^2]^{1/2}\right| \quad (8)$$

2.2.5 Mel Filter Bank

Psychophysical studies have shown that the human perception of the frequency of sound signals does not follow a linear scale. So for each note with the actual frequency f measured in Hz, the subjective tone is measured on a scale called the 'mel' scale. The frequency scale of mel is a linear frequency range below 1000 Hz and a logarithmic distance above 1000 Hz. [11]. In previous research the number of Filter Bank 30, Low mel frequency 130Hz, High mel frequency 6800Hz [12]. The equation of the bank's filter is in equations (9), (10), (11), (12), (13)

$$mel(f) = 2595\, log_{10}\left(1 + \frac{f}{700}\right) \quad (9)$$

$$mel^{-1}(f) = 700 \times \left(10^{\frac{f}{2595}} - 1\right) \quad (10)$$

$$f_{b(i)} = \left(\frac{Ns}{Fs}\right) \times\ mel^{-1}\left(mel(Nlow) + i\frac{mel(Nhigh) - mel(Nlow)}{Nfbank - 1}\right) \quad (11)$$

$$h(i,k) = \begin{cases} 0, & untuk\ k < f_{b(i-1)} \\ \dfrac{k - f_{b(i-1)}}{f_{b(i)} - f_{b(i-1)}}, & untuk\ f_{b(i-1)} \le k < f_{b(i)} \\ \dfrac{f_{b(i+1)} - k}{f_{b(i+1)} - f_{b(i-1)}}, & untuk\ f_{(i)} \le k < f_{b(i+1)} \\ 0, & untuk\ k > f_{b(i+1)} \end{cases} \quad (12)$$

$$Sd[i] = \sum_{k=0}^{Ns-1} |Sc[i]\,|^2 \times\ h(i,k) \quad (13)$$

where $i = 1,2 \ldots Nfbank$

2.2.6 DCT

This is the process for converting the Mel spectrum log into a time domain using Discrete Cosine Transform (DCT). The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficients is called an acoustic vector. Therefore, each input utterance is transformed into an acoustic vector sequence [5]. The DCT equation is in equation (14):

$$Se[n] = \sqrt{\frac{2}{Nbank}} \sum_{k=0}^{Nbank-1} \log(Sd[k+1]) \times \cos\left[n\left(\frac{2k-1}{2}\right)\frac{\pi}{Nfbank}\right]$$

Where $n = 0,1,2 \ldots Nk - 1$

2.3 LVQ

LVQ is a neural calculation algorithm with computational intelligence, using competitive learning to classify. LVQ studies the feature vector of characteristic extraction produced by the MFCC method.

In previous research, LVQ test has been done with 98.57% accuracy [2], with epoh 90 paramter value, learning rate (a) 0.007 and decreasing learning rate (decrease a) 0.977 [2], then this research will be done with the same parameter . The algorithm on learning rate among others [2]:

1. Initial parameters:
   $a, penurunan\ a, epoh$

   where:
   a = *Learning Rate*
   $penurunan\ a$ = penurunan *Learning Rate*
   $epoh$= Maksimum epoh
2. Do the calculations of steps 3 and 4 as long as the loop is smaller than the maximum epoh
   Callculation X_i from W_i with equation *Euclidean Distance* as follows:

$$C_{xw} = \sqrt{\sum_{i=1}^{n}(X_i - W_i)^2 + \cdots + (X_n - W_n)^2} \quad (2.1)$$

   Where:
   X_i = Value Data Training
   W_i = Value Initial Weight
3. Once the value is obtained C_{xw} from $X_i$from value W_i, Then get a new weight value with the following equation:

If $T = C_{xw}$ then

$$W_i(baru) = W_i(lama) + \alpha(X_i - W_i(lama)) \quad (2.16)$$

If $T \neq C_{xw}$ then

$$W_i(baru) = W_i(lama) - \alpha(X_i - W_i(lama)) \quad (2.17)$$

Where:
T = Target of C_{xw} With the smallest value
Stage looking for new weight, after getting value X_i from W_i, Then the next change α with the equation as follows:

$$\alpha = a - penurunan\ a \times a \quad (2.2)$$

4. Do the calculations X_i from W_i of steps 3 and 4 as long as the loop is smaller than the maximum epoh
5. After performing the calculation up to the maximum epoh, then we get new weighted data that will be used as weight data for testing LVQ method.

2.4 Perfomansi

Perfomansi test conducted to determine the performance and also measure the accuracy of the classification method is LVQ method. In this test is divided into 3 testing process, while the testing process is done as follows:

1) Perfomansi test by using training test.
2) Perfomansi test by using supplied test.

3) Perfomansi test using k-fold cross validation method.
4) Perfomansi test by using supplied test with noise

2.4.1 Perfomansi Training Test

In the test of this performance, for training data and test data are used with the same dataset. This test is performed with a dataset of 750 and 5 classes, each class of 150 datasets for test data and training data. Confusion Matrix results from this test can be seen in table 1.

Tabel 1 Confusion Matrix Training Test

Class	A	E	I	K	M	Accuracy
A	145	0	0	5	0	96.67%
E	0	144	0	0	6	96%
I	0	0	149	0	1	99.33%
K	2	0	0	127	21	84.66%
M	0	1	0	3	146	97.33%
Total Accuracy						94.80%

The calculation for test accuracy in the training test as follows:

Aerofon $= \frac{145}{150} \times 100 = 96.67\%$

Elektrofon $= \frac{144}{150} \times 100 = 96\%$

Idiofon $= \frac{149}{150} \times 100 = 96.33\%$

Kordofon $= \frac{127}{150} \times 100 = 84.66\%$

Membranofon $= \frac{146}{150} \times 100 = 97.33\%$

Total Accuracy $= \frac{145 + 144 + 149 + 127 + 146}{750} \times 100 = 94.80\%$

2.4.2 Perfomansi Supplied Test

In the test of this performance, for test data use different dataset from train data. The training data used 750 datasets with each class of 150 datasets and 250 dataset test data with each class of 50 datasets. Confusion Matrix results of this test can be seen in table 2.

Tabel 2 Confusion Matrix Supplied Test

Kelas	A	E	I	K	M	Accuracy
A	43	0	0	7	0	86%
E	0	38	0	0	12	76%
I	0	0	50	0	0	100%
K	0	0	0	42	8	84%
M	0	0	0	0	50	100%
Total Akurasi						89.20%

The calculation for test accuracy on supplied test as follows:

Aerofon $= \frac{43}{50} \times 100 = 86\%$

Elektrofon $= \frac{38}{50} \times 100 = 76\%$

Idiofon $= \frac{50}{50} \times 100 = 100\%$

Kordofon $= \frac{42}{50} \times 100 = 84\%$

Membranofon $= \frac{50}{50} \times 100 = 100\%$

Total Accuracy $= \frac{43 + 38 + 50 + 42 + 50}{250} \times 100 = 89.20\%$

2.4.3 Perfomansi k-fold cross validation

Perfomansi test is performed using k-fold cross validation method with k = 5, where the dataset is divided into 5, with dataset 750 that is A1 = 150, A2 = 150, A3 = 150, A4 = 150, A5 = 150. Testing is done as much Value k, as for the test performed as follows:

1. P1 = A1 as Data Testing and A2, A3, A4, A5 as Data Training.
2. P2 = A2 as Data Testing and A1, A3, A4, A5 as Data Training.
3. P3 = A3 as Data Testing and A1, A2, A4, A5 as Data Training.
4. P4 = A4 as Data Testing and A1, A2, A3, A5 as Data Training.
5. P5 = A5 as Data Testing and A1, A2, A3, A4 as Data Training.

Tabel 3 Confusion Matrix K-Fold Cross Validation P1

Class	A	E	I	K	M	Accuracy
A	30	0	0	0	0	100%
E	0	30	0	0	0	100%
I	0	0	30	0	0	100%
K	0	0	0	26	4	86.67%
M	0	0	2	4	24	80%
Total Accuracy						93.33%

The calculation for test accuracy on k-fold cross validation P1 as follows:

Aerofon $= \frac{30}{30} \times 100 = 100\%$

Elektrofon $= \frac{30}{30} \times 100 = 100\%$

Idiofon $= \frac{30}{30} \times 100 = 100\%$

Kordofon $= \frac{26}{30} \times 100 = 86.67\%$

Membranofon $= \frac{24}{30} \times 100 = 80\%$

Total Accuracy $= \frac{30 + 30 + 30 + 26 + 24}{150} \times 100 = 93.33\%$

Tabel 4 Confusion Matrix K-Fold Cross Validation P2

Class	A	E	I	K	M	Accuracy
A	30	0	0	0	0	100%
E	0	27	0	0	3	90%
I	0	0	30	0	0	100%
K	1	0	0	28	1	93.33%
M	0	0	0	1	29	96.67%
Total Accuracy						96%

The calculation for test accuracy on k-fold cross validation P2 as follows:

Aerofon $= \frac{30}{30} \times 100 = 100\%$

Elektrofon $= \frac{27}{30} \times 100 = 90\%$

Idiofon $= \frac{30}{30} \times 100 = 100\%$

Kordofon $= \frac{28}{30} \times 100 = 93.33\%$

Membranofon $= \frac{29}{30} \times 100 = 96.67\%$

Total Accuracy $= \frac{30 + 27 + 30 + 28 + 29}{150} \times 100 = 96\%$

Tabel 5 Confusion Matrix K-Fold Cross Validation P3

Class	A	E	I	K	M	Accuracy
A	30	0	0	0	0	100%
E	0	27	0	0	3	90%
I	0	0	29	0	1	96.67%
K	1	0	0	29	0	96.67%
M	0	0	0	2	28	93.33%
Total Accuracy						95.33%

The calculation for test accuracy on k-fold cross validation P3 as follows:

Aerofon $= \frac{30}{30} \times 100 = 100\%$

Elektrofon $= \frac{27}{30} \times 100 = 90\%$

Idiofon $= \frac{29}{30} \times 100 = 96.67\%$

Kordofon $= \frac{29}{30} \times 100 = 96.67\%$

Membranofon $= \frac{28}{30} \times 100 = 93.33\%$

Total Accuracy $= \frac{30 + 27 + 29 + 29 + 28}{150} \times 100 = 95.33\%$

Tabel 6 Confusion Matrix K-Fold Cross Validation P4

Class	A	E	I	K	M	Accuracy
A	30	0	0	0	0	100%
E	0	30	0	0	0	100%
I	0	0	30	0	0	100%
K	0	0	0	25	5	83.33%
M	0	0	0	0	30	100%
Total Accuracy						96.67%

The calculation for test accuracy on k-fold cross validation P4 as follows:

Aerofon $= \frac{30}{30} \times 100 = 100\%$

Elektrofon $= \frac{30}{30} \times 100 = 100\%$

Idiofon $= \frac{30}{30} \times 100 = 100\%$

Kordofon $= \frac{25}{30} \times 100 = 83.33\%$

Membranofon $= \frac{28}{30} \times 100 = 100\%$

Total Accuracy $= \frac{30 + 30 + 30 + 25 + 30}{150} \times 100 = 96.67\%$

Tabel 7 Confusion Matrix K-Fold Cross Validation P5

Class	A	E	I	K	M	Accuracy
A	22	0	0	8	0	73.33%
E	0	30	0	0	0	100%
I	0	0	29	0	1	96.7%
K	1	0	0	23	6	76.67%
M	0	0	0	0	30	100%
Total Accuracy						89.33%

The calculation for test accuracy on k-fold cross validation P5 as follows:

Aerofon $= \frac{22}{30} \times 100 = 73.33\%$

Elektrofon $= \frac{30}{30} \times 100 = 100\%$

Idiofon $= \frac{29}{30} \times 100 = 96.7\%$

Kordofon $= \frac{23}{30} \times 100 = 76.67\%$

Membranofon $= \frac{30}{30} \times 100 = 100\%$

Total Accuracy $= \frac{22 + 30 + 29 + 23 + 30}{150} \times 100 = 89.33\%$

Tabel 8 Confusion Matrix total average accuracy K-Fold Cross Validation

Class	A	E	I	K	M	Accuracy
A	142	0	0	0	0	94.67%
E	0	144	0	0	0	96%
I	0	0	148	0	0	98.67%
K	0	0	0	131	0	87.33%
M	0	0	0	0	141	94%
Total Average Accuracy						94.13%

The calculation for test accuracy on k-fold cross validation as follows:

Aerofon $= \frac{142}{150} \times 100 = 94.67\%$

Elektrofon $= \frac{144}{150} \times 100 = 96\%$

Idiofon $= \frac{148}{150} \times 100 = 98.67\%$

Kordofon $= \frac{131}{150} \times 100 = 87.33\%$

Membranofon $= \frac{141}{150} \times 100 = 94\%$

Total aerage Accuracy $= \frac{142 + 144 + 148 + 131 + 141}{750} \times 100 = 94.13\%$

2.4.4 Perfomansi Supplied Test with Noise

Kelas	A	E	I	K	M	Akurasi
A	0	6	0	0	44	0%
E	0	44	0	0	6	88%
I	0	0	27	0	23	54%
K	2	0	0	5	45	10%
M	0	0	0	0	50	100%
Total Accuracy						50.40%

The calculation for test accuracy on supplied test with noise as follows:

Aerofon $= \frac{0}{50} \times 100 = 0\%$

Elektrofon $= \frac{44}{50} \times 100 = 88\%$

Idiofon $= \frac{27}{50} \times 100 = 54\%$

Kordofon $= \frac{5}{50} \times 100 = 10\%$

Membranofon $= \frac{50}{50} \times 100 = 100\%$

Total Accuracy $= \frac{0 + 44 + 27 + 5 + 50}{250} \times 100 = 50.40\%$

3. PENUTUP

Based on the tests, the MFCC and LVQ methods can classify the sound source of the instrument with an accuracy of 94.80% of the test training test, 89.20% of the supplied test, 94.13% of the k-fold cross validation test, 50.40% of the supplied test Test with noise. It can be concluded that MFCC and LVQ methods can be implemented in classifying sound sources on musical instruments with good accuracy.

DAFTAR PUSTAKA

[1] R. Hidayatullah, M.Pd. dan Hasymkan, S.Sn., M.Sn, Dasar-Dasar Musik, Bandar Lampung: arttex, 2016.

[2] F. Syafria, A. Buono dan B. P. Silalahi, "A Comparison of Backpropagation and LVQ : a case study of lung sound recognition," *IEEE,* 2014.

[3] R. Haldar dan P. K. Mishra, "Learning Vector Quantization (LVQ) Neural Network Approach for Multilingual Speech Recognition," *International Research Journal of Engineering and Technology (IRJET),* vol. 3, no. 5, 2016.

[4] L. N. Sari, "Penerapan learning vector quantization (LVQ) dan ekstraksi ciri menggunakan mel-frequency cepstrum coefficents (MFCC) untuk transkripsi suara ke teks," *Institut Pertanian Bogor,* 2014.

[5] E. D. Widianto, Sistem Digital. Analisis, Desain dan Implementasi, 2014: Graha Ilmu.

[6] S.-H. Chen dan Y.-R. Luo, "Speaker Verification Using MFCC and Support Vector Machine," *IMECS,* vol. 1, 2009.

[7] K. K. Kumbhare dan K. V. Warkar, "Secure E-Commerce Transaction using Noisy Password with Voiceprint and OTP," *International Journal on Recent and Innovation Trends in Computing and Communication,* vol. 4, no. 8, 2016.

[8] "About the Android Open Source Project," [Online]. Available: https://source.android.com. [Diakses 27 Maret 2017].

[9] "Codenames, Tags, and Build Numbers," [Online]. Available: https://source.android.com/source/build-numbers. [Diakses 2017 Maret 2017].

[10] "Mengenal Android Studio," [Online]. Available: https://developer.android.com/studio/intro/index.html?hl=id. [Diakses 27 Maret 2017].

[11] Rosa A. S dan M. Shalahuddin, Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek), Bandung: Informatika Bandung, 2016.

[12] L. Muda, M. Begam dan I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," *JOURNAL OF COMPUTING,* vol. 2, no. 3, 2010.

[13] C. G. K. Leon , "Robust Computer Voice Recognition Using Improved MFCC Algorithm," *IEEE International Conference on New Trends in Information and Service Science,* 2009.

[14] F.-Y. Leu dan G.-L. Lin , "An MFCC-based Speaker Identification System," *IEEE 31st International Conference on Advanced Information Networking and Applications,* 2017.

[15] K. Chakraborty , A. Talele dan S. Upadhya , "Voice Recognition Using MFCC Algorithm," *International Journal of Innovative Research in Advanced Engineering (IJIRAE) ,* vol. 1, no. 10, 2014.

[16] S. K. Kopparapu dan M. Laxminarayana , "Choice of mel filter bank in computing MFCC of Resampled Speech," *International Conference on Information Science, Signal Processing and their Applications ISSPA* , 2010.

[17] W. HAN, C.-F. CHAN, C.-S. CHOY dan K.-P. PUN, "An Efficient MFCC Extraction Method in Speech Recognition," *IEEE ISCAS,* 2006.

[18] R. S. Pressman, Software Enginerring A Practitioner's Approach, McGraw Hill, 2012.

[19] R. Payam Refaeilzadeh, T. Lei dan L. Huan , Cross-Validation, Arizona State University, Tempe, AZ, USA , 2009.

[20] Sugiyono, Metode Penelitian Kuantitatif Kualitatif dan R&D, Bandung: Alfabeta, 2014.

[21] A. Eronen dan A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features," *IEEE,* 2000.

[22] L. Muda, M. Begam dan I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," *Journal Of Computing*, vol. 2, no. 3, 2010.