Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 4, No. 3, August 2019, Pp. 233-240

Pattern recognition bird sounds based on their type using discrete cosine transform and gaussian methods

Hendro Nugroho[*1], Wahyu Widodo[2], Andy Rachman[3]
Institut Teknologi Adhi Tama Surabaya, Indonesia[1,2,3]

Article Info
Keywords:
Pattern Recognition, DCT, Gaussian, Bray Curtis

Article history:
Received 16 March, 2019
Revised 16 April, 2019
Accepted 25 June, 2019
Published 30 July 2019

Cite:
Nugroho, H., Widodo, W., & Rachman, A. (2019). Pattern Recognition Bird Sounds Based on Their Type Using Discreate Cosine Transform (DCT) and Gaussian Methods. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 4(3). doi:http://dx.doi.org/10.22219/kinetik.v4i3.791

* Corresponding author.
Hendro Nugroho
E-mail address:
dosh3ndro@itats.ac.id

Abstract
To know the type of bird, most people know from the shape of bird species and the sound of birds. In this study, it identified the pattern of bird sounds. The bird sounds studied were Canary Trills, Vulture and Crow birds. In the introduction of the type of bird sound pattern in this study using the Discrete Cosine Transform (DCT) method and Gaussian value. The researcher conducted several steps to get the sound model of birds, among others, namely (1) bird sound input in the form of WAV file, (2) Hamming Windowing, (3) Discrete Fourier Transform/Fast Fourier Transform, (4) Mel Bank Filter, (5) DCT, and (6) Value Gaussian. The output obtained is in the form of vector values and represented in graphical form. The results obtained in the study of pattern recognition of bird sound types get the results of observations in the same bird sound duration and frequency of the same, then the same pattern is obtained in the same bird as evidenced by calculating the closest distance value with Bray Curtis method. For the same duration of time and the length of the frequency that is not the same; it found that the pattern of bird sounds is not the same.

1. Introduction

Along with the development of speech recognition technology which is often called Speech Recognition or also called sound pattern recognition is very developed rapidly. In designation before the use of Speech Recognition is used for the technique of identifying someone based on sound [1]. The recognition of sound patterns is an automatic conversion spoken in the form of lip movements that can be read by the system [2]. The introduction of models in extraction forms widely used in sound pattern recognition was carried out by previous researchers [3]. To further develop Speech Recognition, this study uses bird sounds to find out information on bird species. The use of Speech Recognition is to recognize the digital signal patterns of bird sounds that matched with other types of bird sounds. Bird's voice digital signals in the form of sound waves converted into a set of digit digits extracted to get a specified pattern that can identify the bird species.

For this study, the extraction of bird sound features using Discrete Cosine Transform, where previously the birds sound signal input was carried out by the Hamming Windowing process, Discrete Fourier Transform (DFT) / Fast Fourier Transform (FFT), Mel Filter Bank Processing, and finally DCT. The process of this research is almost the same as the research using the MFCC (Mel-Frequency spectrum Coefficient) method [3]. MFCC is a speech recognition method that has the highest working principle of human [4]. However, the research process lies in not using the frame blocking process because this study calculates all the whole sound of birds without dividing the sound process into several frames and one frame from several samples. The researcher used the DCT process to return the frequency domain signal to the time domain. The results of this process produce spectrum [3]. The results obtained from the extraction of DCT features are in the form of an acoustic vector sequence.

The result of the next stage is the sequence of bird sound acoustic vectors carried out in the process of calculating the Gaussian value. Gaussian values and standard deviations are used as values to determine the pattern of bird sounds. The researcher uses sound patterns from the Gaussian values as training data or testing data.

2. Research Method

The use of the DCT method in this study, carried out several processes or stages of sound pattern recognition starting from Hamming Windowing, DFT / FFT, filtering the bank and DCT filters. To find out the process of swallowing the type of bird sound pattern can be seen in Figure 1.

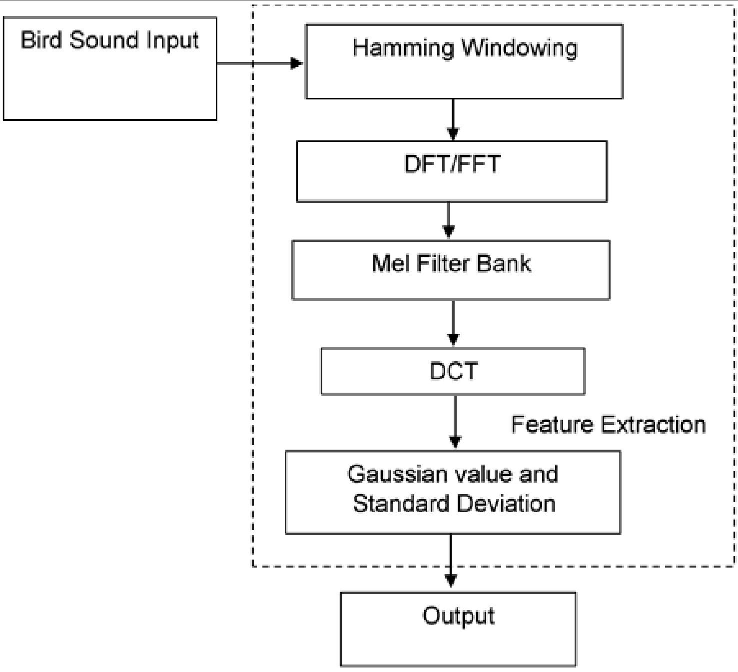

Figure 1. Block Diagram of Bird Type Sound Pattern Recognition

In the Block Diagram of Bird Type Sound Pattern Recognition Figure 1, there is a feature extraction process to find out the acoustic model which can find out the sound pattern in the form of vector values or input sound pattern graphics. These sound bird patterns will be observed based on the type of bird.

2.1. Hamming Windowing

Using Hamming Windowing as a form of window that is blocked in the extraction chain feature and integrates all the closest frequency lines. The definition of Window as W (n), n = 0,1,2, ... N-1, the result of signal Windowing is shown in Equation 1 [5].

$$Y[n] = X[n] * W[n] \tag{1}$$

The results of Hamming Windowing produce level that is not too high (approximately -43 dB), and noise that is not too large. Use of the Hamming Windowing equation in Equation 2 [5].

$$W(n) = 0.54 - 0.46 \cos\left[\frac{2\pi n}{N-1}\right] \tag{2}$$

2.2 Discrete Fourier Transform (DFT) / Fast Fourier Transform (FFT)

The researcher performs the DFT calculation process by analyzing the frequency of the discrete time signal x (n) in the form of vector values [6], it is necessary to do the DFT calculation process. The researcher formulated the use of DFT as shown in Equation 3.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi n\left[\frac{k}{N}\right]}, 0 \leq k \leq N-1 \tag{3}$$

From Equation 3, Sound signal formation in the time domain uses frequency. The use of DFT in the form of FFT analysis on the voice data signal into the time domain becomes a data spectrum in the frequency domain [7].

The use of DFT calculations directly in computing is very long because the calculation requires N2 multiplication of complex numbers. To overcome this problem by using the FFT method. FFT has a complexity of X (N Log2 N), so the use of FFT saves calculation time, because there is a calculation of Log2 N, so that N does not occur the square of two.

2.3 Mel Filter Bank

Mel Filter Bank is used to convert sound signals in the frequency domain to the frequency domain mel and shows several energy quantities in the frequency range available in each filter mel. The approach in calculating mel is in frequency f (Hz) in Equation 4 [8].

$$F(Mel) = [2595 * ln[1 + \frac{f}{700}) \quad (4)$$

For calculating filter banks use convolution representation in performing filter signals. The use of convolution by multiplying the signal spectrum and the filter bank coefficient value. The used of filterbank in Equation 5.

$$Y[i] = \sum_{j=1}^{Ns} S[j]H_i\,[j] \quad (5)$$

2.4 Discrete Cosine Transform (DCT)

From the results of Mel Filter bank, a power spectrum is reprocessed using DCT to obtain an acoustic vector value [9]. The DCT equation in Equation 6.

$$C_n = \sum_{K=1}^{K} (LogS_K)cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right]; n = 1,2,..K \quad (6)$$

The researcher removes the DCT coefficient value to zero even though it indicates the value of the frame signal [8]. Elimination of zero costs because previous studies conducted are not reliable on speech recognition [10].

2.5 Gaussian

This study uses the standard deviation (σ^2) and mean µ with the gaussian equation with the aim of getting the normal distribution value in the DCT vector Equation 7 [11].

$$fx(x) = \frac{1}{\sqrt{2\pi\sigma}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty \quad (7)$$

In Equation 7, the Gaussian value can show too random and uneven values, so the Gaussian value makes it easier for researchers to know the characteristics of bird sound signals used for research.

2.6 Bray Curtis

In a study using the Bray Curtis method to find out the similarity of data from the type of bird sounds. Bray Curtis' calculation is not the closest distance calculation, because it does not fulfill the value of the triangle equation vector. Therefore, Bray Curtis called Bray Curtis for data inequality/similarity Equation 8 [12].

$$j(v_1, v_2) = \frac{\sum_{k=1}^{N} |v_1(k) - v_2(k)|}{\sum_{k=1}^{N} v_1(k) + \sum_{k=1}^{N} v_2(k)} \quad (8)$$

3. Results and Discussion

The data used in the study were bird sounds consisting of bird species such as Canary Thrills, Vulture, and Crow. Sound data of this type of bird with a duration of time from 1 to 2 second (s).

3.1 Birds Sound Input Data

The researcher used the sound of bird types WAV file format as input data. The WAV file format belongs to Microsoft from the RIFF specification section which is used to store multimedia files. The file has two parts, namely the header and the chunk data. The WAV file section consists of 3 parts, namely the main chunk, chunk format, and chunk data.

The value obtained from the WAV file in discrete form is a series of numbers that represent the amplitude in the time domain. In the header section, there is information about the value of the sample rate, number of channels, and bits per sample. Table 1 shows the initial report on bird type sound data for initial details.

Table 1. Example of Bird Sound Input Data

Voice Data	Information
File Name	Bird007
Format	WAVE
Bird Species	Canary Trills
Chunk ID	RIFF
Chunk Size	27796
Sample Rate	8000
Byre Rate	32000
Bits per Sample	16
Data Length	27760

The value of the voice signal in the discrete form represented in the form of a signal graph made from the Delphi programming language. The results of the chart in Figure 2. In Figure 2 the signal is divided into two, namely the left signal and the right signal.

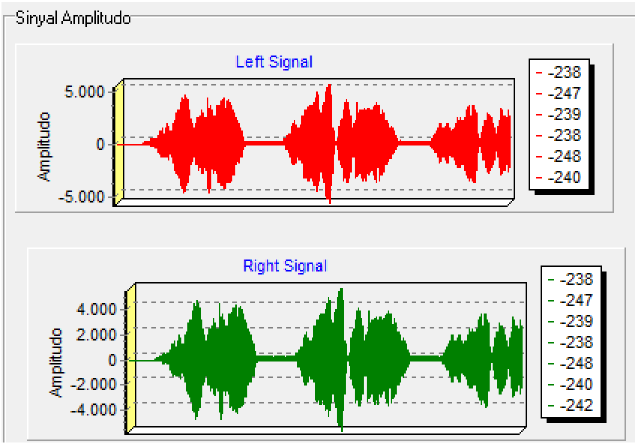

Figure 2. Bird Sound Amplitude Signal (bird007)

3.2 Feature Extraction

The birds sound signal is carried out by the Hamming Windowing process before the DFT / FFT process is carried out. The DFT / FFT process is used to analyze sound signals with a discrete value; the results obtained to show in Figure 3.

After the results of DFT / FFT, the value is converted using the Mel Filter Bank method. This process produces a time-frequency domain signal into the Mel frequency domain using Equation 4 to create a Mel filter. The results obtained show some energy quantities in the range of frequencies that exist in each Mel filter. In Equation 5 for calculating Filter banks use convolution representation in performing filter signals. The results to show in Figure 4 which represents in graphical form.

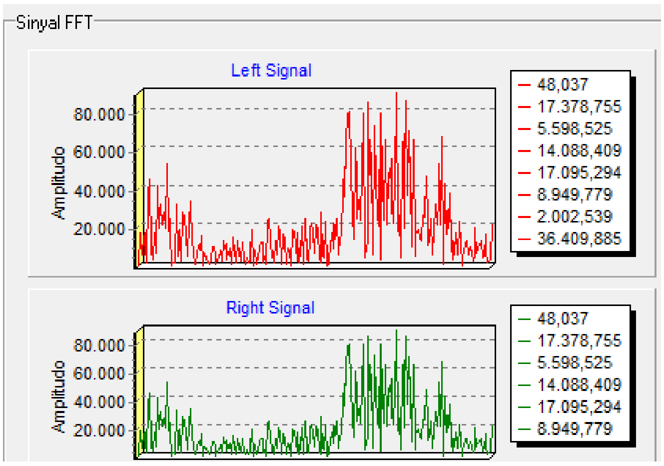

Figure 3. Bird Sound Graph In The Form of DFT/FFT

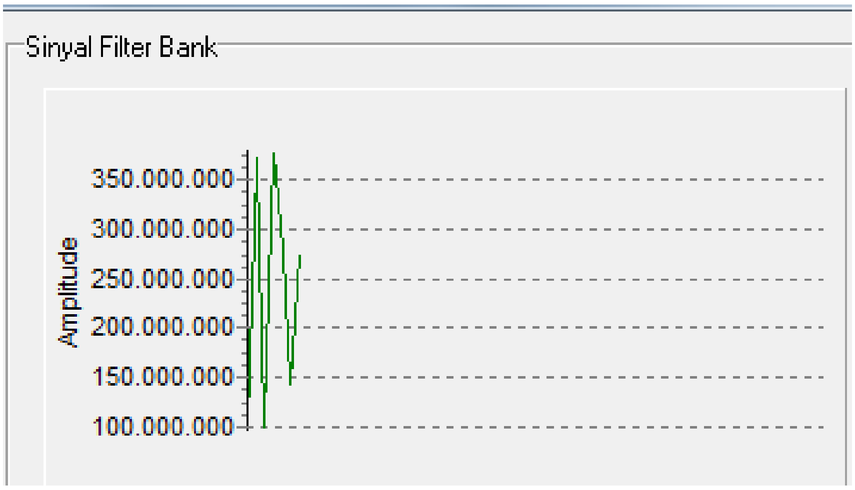

Figure 4. Signals for Mel Filter Bank

After the Mel Filter Bank process, the calculation is done using the DCT method. The DCT method is used to convert frequency domain signals to acoustic vector values which to show in Table 2 And to describe the results of the DCT in graphical form to show in Figure 5.

Table 2. Sample Result of DCT Canary Thrills Bird Sounds

Voice Data	Frame						
	1	2	3	4	5	6	7
Bird007	0,0062	0,0065	0,0062	0,0066	0,0065	0,0063	0,0065
Bird008	0,0059	0,0064	0,0064	0,0067	0,0068	0,0056	0,0056

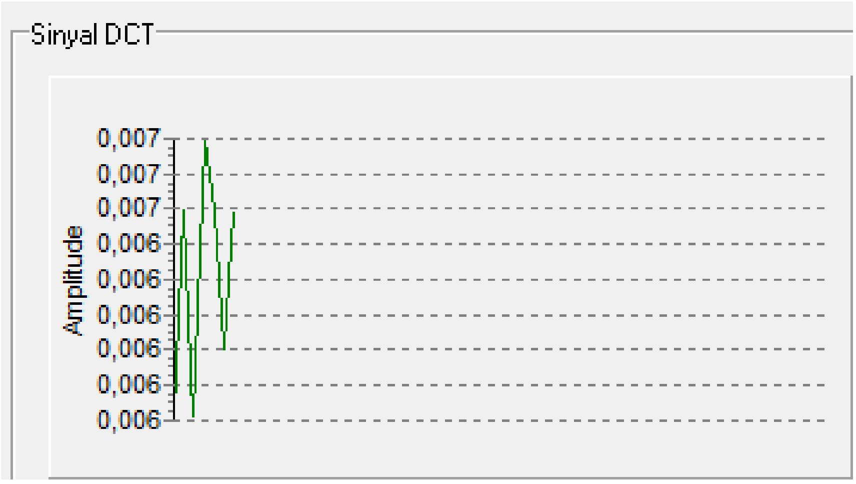

Figure 5. Graph DCT

To get a proper analysis result, using equation seven the normalized Gaussian value can find out the characteristics of the DCT value results that have values that are too random and the distribution value is not evenly distributed. The results of the Gaussian value to show in Table 3. And for results that are easy to analyze can be represented in graphical form, the graph results to display in Figure 6.

Table 3. Example of The Results of The Canary Thrills Bird Sound Gaussian Data

Voice Data	Frame						
	1	2	3	4	5	6	7
Bird007	8,9	8,6	8,9	8,5	8,6	8,8	7,9
Bird008	8,9	8,4	8,7	8,4	8,2	8,7	7,9

Cite: Nugroho, H., Widodo, W., & Rachman, A. (2019). Pattern Recognition Bird Sounds Based on Their Type Using Discreate Cosine Transform (DCT) and Gaussian Methods. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 4(3). doi:http://dx.doi.org/10.22219/kinetik.v4i3.791

3.3 Discussion

The researcher used the Bay Curtis distance calculation method with Equation 8 to find out the similarities in the pattern recognition of bird sounds. For observations of the introduction of bird sound type patterns, there are three examples of experimental results. Calculation of the closest distance for the opening of bird sound type patterns using the Gaussian mean value and the average value of the Standard Deviation as the vector value. From the calculation of Bray Curtis distance can be seen in Table 4 for the type of bird sound with the same or similar pattern. And graphs for sound patterns with the same can be seen in Figure 6.

Table 4. Example of The Same Bird Sound Pattern Data

Voice Data	Information
Bird007	Canary Thrills
Bird008	Canary Thrills
Gaussian Average Bird007	8,6
Gaussian Average Bird008	8,5
SD 007	3,1
SD 008	3
Time 007	1 s
Time 008	1 s
Frame Length Bird007	7
Frame Length Bird008	7
The Closest Distance Value	0,01

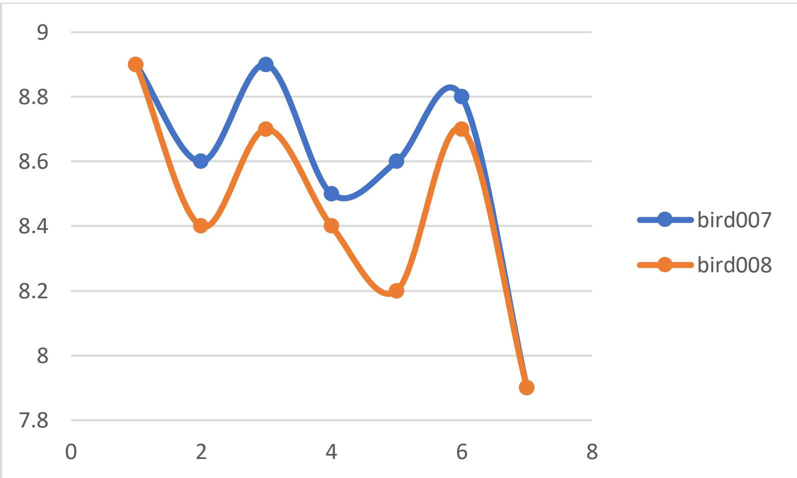

Figure 6. Graph of The Canary Thrills Sound Pattern (Gaussian Value)

The observations for the sound data of bird species that are not the same are the types of Vulture birds. The kind of sound of the Vulture bird in this study is the same as the type of Crow bird sound. From the observation of the sound patterns of the Vulture birds that were the same as Crow's, it was due to Vulture's and Crow's voice recording time being the same as 1 s. As for other observations on the frame length of the DCT result, there are 3 (three). Examples of research results on the types of Vulture and Crow birds to show in Table 5, Table 6, and Table 7.

Table 5. Example of Bird003 Bird Sound Data

Voice Data	Information
Bird Name	Vulture
Gaussian Average	7,9
SD Average	1,8
Frame Length	22
Time	2 s

Table 6. Example of Bird004 Bird Sound Data

Voice Data	Information
Bird Name	Vulture
Gaussian Average	7,2
SD Average	3,6
Frame Length	3
Time	1 s

Table 7. Example of Bird005 Bird Sound Data

Voice Data	Information
Bird Name	Crow
Gaussian Average	7,2
SD Average	3,6
Frame Length	3
Time	1 s

To calculate the closest distance using the Bray Curtis method on the type of Vulture and Crow birds that are closest to Vulture (Bird004) and Crow (Bird005) sound files, with a value of 0, while comparing the Vulture (Bird003) bird sound type with Vulture (Bird004) with the closest amount of 0.12. The graph pattern of Vulture (Bird003) and Vulture (Bird004) birds to show in Figure 7.

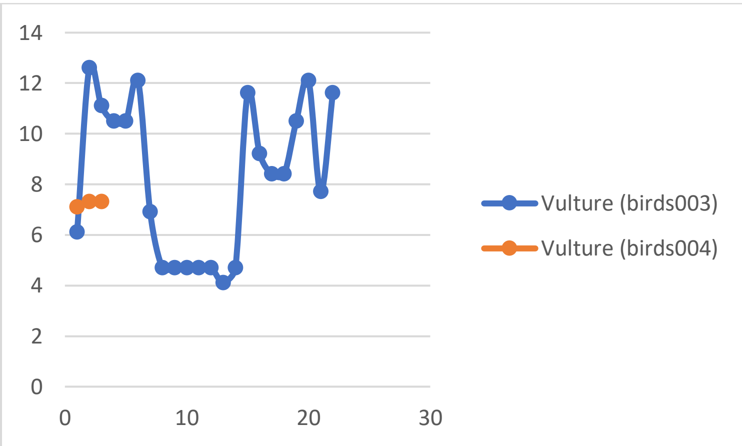

Figure 7. Vulture (Bird003) and Vulture (Bird004) Bird Sound Pattern Charts

4. Conclusion

From the results of the research obtained to find out the pattern recognition of bird sounds, can be summarized as follows. For the same pattern recognition of bird sounds caused by the same frame length and time or sound duration, in Canary Trills bird data with the closest distance value of Bray Curtis 0.01. For sound patterns that are not the same due to frame length and time or duration that are not the same. Suggestions expected in this study are to obtain a good pattern of bird sound recognition so that the period of sound time is expected to be the same so that for the addition of the method it is recommended to use frame blocking in the MFCC method model to be able to eliminate low-frequency values.

5. Notation

The example of mathematics described with the following description:

n : the number of data
N : the amount of data/frame.
μ : the average value of data (Mean).
F : Frequency.
Y[n] : Windowing
W[n] : Hamming Windowing.
X[k] : DFT
k : the number of data
f(mel) : Mel Frequency

Y[i] : FilterBank
C_n : DCT
$F_x(x)$: Gaussian
σ : Standar Deviasi
Π : pi
V_i : Vector value
J(v1,v2): Distance value Bray Curtis

References

[1] B. C. K. Dong-III Kim, “Speech Recognition using Hidden Markov Models in Embedded Platform,” *Indian J. Sci. Technol.*, Vol. 8, No. 34, 2015. http://dx.doi.org/10.17485/ijst%2F2015%2Fv8i34%2F85039

[2] S. Ananthi and P. Dhanalakshmi, “Speech Recognition System and Isolated Word Recognition Based on Hidden Markov Model (HMM) for Hearing Impaired,” *Int. J. Comput. Appl.*, Vol. 73, No. 20, Pp. 30–34, 2013.

[3] *Mahdi Shaneh and Azizollah Taheri, “Voice Command Recognition System Based on MFCC and VQ Algorithms,” World Acad. Sci. Eng. …*, Vol. 2, No. 3491, Pp. 501–505, 2009.

[4] S. Berhaningtyas Hertiana, Muh Khaerul Amri S.P, “Pengenalan Huruf Hijayyah Berbasis Pengolahan Sinyal Suara dengan Metode Mel Cepstrum Frequency Cepstrum Coefficient (MFCC),” Momentum, Vol. 13, No. 2, Pp. 49–52, 2017.

[5] Y. R. Prayogi and J. L. Buliali, “Identifikasi Parameter Optimal Gaussian Mixture Model Pada Identifikasi Pembicara di Lingkungan Berderau Menggunakan Residu Deteksi Endpoint,” Vol. 13, No. 2, Pp. 198–206, 2015. http://dx.doi.org/10.12962/j24068535.v13i2.a489

[6] M. Vyas, “A Gaussian Mixture Model Based Speech Recognition System Using Matlab,” S*ignal Image Process. An Int. J.*, Vol. 4, No. 4, Pp. 109–118, 2013.

[7] D. K. Putra, I. Iwut, and R. D. Atmaja, “Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (MFCC) Dan Gaussian Mixture Model (GMM),” *eProceedings Eng.*, Vol. 4, No. 2, Pp. 1766–1772, 2017.

[8] P. Upadhyaya, O. Farooq, M. R. Abidi, and P. Varshney, “Comparative Study of Visual Feature for Bimodal Hindi Speech Recognition,” *Arch. Acoust.*, Vol. 40, No. 4, Pp. 609–619, 2015.

[9] X. Cheng and Q. Duan, “Speech Emotion Recognition Using Gaussian Mixture Model,” Pp. 1222–1225, 2012. https://dx.doi.org/10.2991/iccasm.2012.311

[10] Suherdiansyah Fajar, “Klasifikasi Gerak Bibir Berdasarkan Pola Suara Menggunakan Metode Mel-Frequency Cepstrum Coefficients (MFCC) dan Hidden Markov Model (HMM) untuk Mengenal Kata Sederhana Indonesia,” Sekolah Tinggi Teknik Surabaya, 2019.

[11] M. I. Ribeiro, “Gaussian Probability Density Functions: Properties And Error Characterization,” *Inst. Super. Tcnico, Lisboa, Port. Tech. Rep,* No. February, Pp. 1049–1, 2004.

[12] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, “Experiments of Distance Measurements in a Foliage Plant Retrieval System,” *Int. J. Signal Process. Image Process. Pattern Recognit*., Vol. 5, No. 2, Pp. 47–60, 2012.