Prosiding Seminar Nasional Teknologi Informasi dan Kedirgantaraan : Peran Teknologi untuk Revitalisasi Bandara dan Transportasi Udara, Yogyakarta, 10 Desember 2019
SENATIK 2019, Vol. V, ISBN 978-602-52742-1-3
DOI: 10.28989/senatik.v5i0.326

IMPLEMENTATION OF MEL FREQUENCY CEPSTRAL COEFFICIENT AND DYNAMIC TIME WARPING FOR BIRD SOUND CLASSIFICATION

Hari Prapcoyo[1], Bertha Pratama Adhita Putra[2], Rifki Indra Perwira[3]
Dept. of Informatics Engineering, Faculty of Industrial Engineering
UPN "Veteran" Yogyakarta
Jl. Babarsari 2 Tambakbayan Yogyakarta
Email : [1]hari.prapcoyo@upnyk.ac.id, [2]pasienandroid@gmail.com, [3]rifki@upnyk.ac.id

Abstract

Lovebird (Agapornis) is a type of bird that has become the belle of new pet birds lately. The interest of the hobbyist in this one song is because Lovebird has a unique chirp. For beginner lovebird fans, the lack of knowledge and experience about lovebird birds results in various cases of fraud in choosing a quality lovebird. They were disappointed expensive lovebirds that had been purchased but did not match what was expected. Lovebird chirping voice recognition can be learned and recognized through the learning process of speaker recognition, which is part of voice recognition. Speaker recognition captures the frequency of the lovebird's voice, then compares it with the sound frequency of the existing training data. The sound frequency and the long duration of chirping of lovebird birds will be extracted through the Mel-Frequency Cepstral Coefficient (MFCC) method. Information in the form of Mel Frequency Cepstrum Coefficients from input data and training data is then compared to the Dynamic Time Warping method. The methodology used in this study uses the grapple method. The results of this study were obtained an accuracy value of sound validation by 80%. It is hoped that with the capabilities of this system, it can help bird chirping lovers know the sound quality of lovebird birds that are good, moderate, and less. Also, it can help the jury of birds chirping, so that it can be used as an accurate standard in classifying lovebird sounds.

Keywords: *Lovebird, Voice Recognition, MFCC, DTW.*

1. Introduction

Lovebird (Agapornis) is a type of bird that has become the belle of new pet birds lately. The interest of the hobbyist in this one song because lovebird has a unique chirp [1]. Therefore, there is no doubt that the existence of a lovebird which continues to creep up and is increasingly taken into account even though not immediately skyrocketed when contested. Its popularity as a contest bird can be said to be more stable compared to other birds, such as green cucumbers, stone magpie, and pitcher pots in the contest arena, which immediately appears horrendous and exorbitant prices [6].

For beginner lovebird fans, the lack of knowledge and experience about lovebird birds results in a variety of fraud cases in choosing a quality lovebird that has a long blistering sound, while birds that have long blistering are not necessarily good quality. They were disappointed, expensive lovebirds that had been purchased but did not match what was expected [6]. Also, for fans who have a busy life, they can not monitor the development of their lovebird at any time.

Previous studies discussing speaker recognition technology using MFCC and DTW methods, namely research conducted by [2] with the title Lovebird Voice Classification with the Mel Frequency Cepstral Coefficient (MFCC) Method and Fuzzy Logic. This study applies the Mel Frequency Cepstral Coefficients (MFCC) method as a characteristic to distinguish the quality of lovebird chirping and the Fuzzy Logic algorithm for the classification of lovebird chirping. The results of this study were obtained an accuracy value of 91.67% with a computing time of 190,229 seconds. As for some suggestions for further research is to change the method of feature extraction and classification to increase the value of accuracy. Also, it is hoped that the system can then be developed on an android application and can be used in real-time [2].

Research conducted by [13] with the title Implementation of the MFCC and DTW Methods for the Introduction of Male and Female Voice Types. MFCC (Mel - Frequency Cepstral Coefficient) and DTW (Dynamic Time Warping) are a method of processing sounds, in this study sound processing is conducted aimed at the introduction of male and female voice types. Determination of the type of male or female voice is usually done in the determination of the choir group. MFCC is a method for feature extraction; in addition to MFCC, another method is used, namely, DTW, which is a method of matching training sounds and test sounds, this technique is useful for calculating the distance between two data with different patterns and calculating the value and distance from the data. In this study, a system was built that could recognize male and female voice types; male voice types were divided into Tenor, Bariton, and Bass, while in women were divided into Sopran, Mezzosopran, and Alto. The research results obtained are for the level of accuracy in women with alto type of voice obtained a percentage of 80%, for the level of accuracy of the type of mezzo-soprano voice obtained 90%, for the level of accuracy of the type of soprano voice obtained 80%. Then in the type of male voice, for the type of bass sound obtained an accuracy rate of 80%, for the type of baritone sound obtained an accuracy rate of 70%, and for the type of tenor, sound obtained an accuracy rate of 60% [13].

Research conducted by [14] with the title Speaker Recognition Using MFCC and DTW Algorithms. Speaker recognition is the ability of a machine or program to recognize or ascertain the identity of a speaker based on the characteristics of his voice. There are two types of text in speaker recognition, namely text-dependent and text-independent. Several studies have verified the sound using Dynamic Time Warping (DTW) has got good results. Likewise, voice identification research with Mel Frequency Cepstral Coefficients (MFCC) and GMM algorithm. Most of these studies use English, Indian, Persian, Tamil, and Indonesian. Speech data generally take sound samples, say a word or several words, and are text-dependent. Therefore, a speech recognition study will be conducted using the MFCC method and the DTW algorithm with independent text types and tells several 30-second sentences. Data obtained from the results of the recording itself 20 speakers. In three different room conditions. Testing is known by using the Maximum Likelihood Classification method. The test results show the ability of the system to verify speakers by 70% for dimensions one and two, and the ability to identify speakers by 15% for dimensions one and two [14].

As for several things that distinguish this study from previous research, namely, no one has used the Dynamic Time Warping (DTW) method to classify lovebird sounds after extraction using the Mel Frequency Cepstral Coefficient (MFCC) method. In this study, the system can be used to record lovebird bird sounds directly, after getting a lovebird bird voice recording file then directly processed to get a classification of the quality of the lovebird sounds. Also, the performance of the system in assessing and classifying the sound quality of lovebird birds in this study is different. The main contributions of this paper are:

1) Propose MFCC and DTW for the classification of bird sounds.
2) Test the accuracy of the birds sound classification system using MFCC and DTW
3) Help chirping mania and bird judges in choosing quality birds.

Based on these problems, this research will create applications that can help to classify the sounds of lovebird birds using speaker recognition technology with the MFCC and DTW methods. The assessment is done based on the sound of the chirping produced.

2. Research Method

Data collection research methods consist of observational studies, interviews, and related methods used in this study.

2.1 Observations

Observations were made with several fellow lovebird hobbyists scattered throughout Indonesia and focused in Solo, Central Java. The observations were made by observing lovebird birds that have good, moderate, and poor sound quality. Observations were made starting from December 2018 - August 2019 in several places, including in Solo, Central Java, Yogyakarta, and in some fellow lovebird hobbyists.

2.2 Interview

This interview is a data collection process that is carried out with a question and answer process. The results of the observation through observation were strengthened by interviews with several sources related to the sound quality of lovebird birds, including with fellow lovebird hobbyists.

2.3 *Speaker Recognition*

Speaker recognition can be defined as the process of selecting speakers (speakers) that have characteristics that are close to the same sound as the input sound. The input sound is extracted in its characteristics to be compared with some existing reference models of speakers, and it is sought which ones are close to the same to be decided as the input voice speaker earlier. This process compares the 1: N speaker reference model and generally consists of three main parts, as in Figure 1.

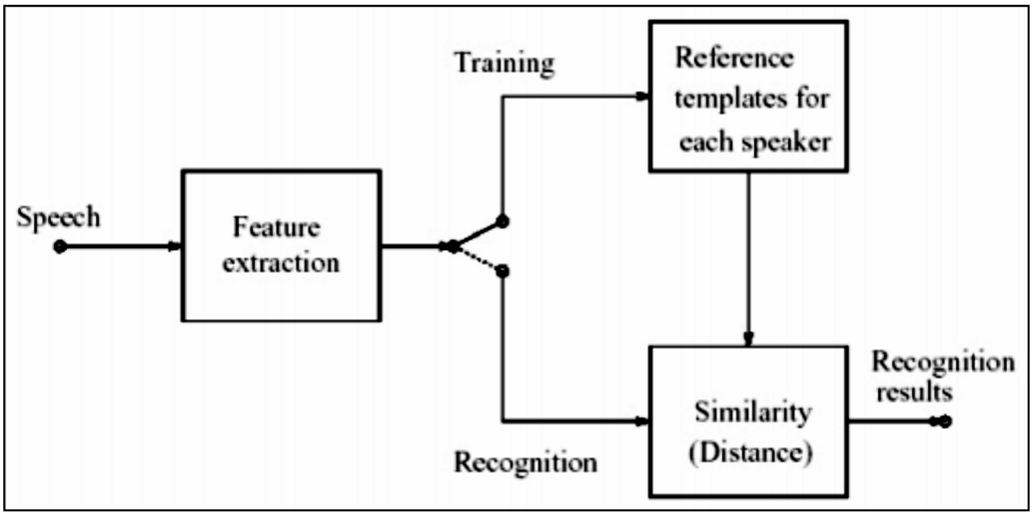


Figure 1. Reference model speaker

The first part is Feature extraction; here, the input voice signal is extracted by the signal characteristics using the feature extraction method, the Mel-Frequency Cepstral Coefficient (MFCC) method.

The second part is the Training phase section, in this section the characteristics of the input sound signal was in the form of a series of acoustic vectors, classified and then stored as a speaker reference model, in classifying the speaker reference model this can use the Dynamic Time Warping (DTW) method.

The third part is the Recognition or Testing phase, looking for the similarity of sound features by measuring the distortion between some existing speaker model references and the input voice signal that will be recognized. The closest or smallest distance is decided as the speaker of the input voice signal. When applied to speaker verification, it measures the distance only between the input signal and the recognized speaker reference model [12]

2.4 *Mel-frequency Cepstral Coefficient (MFCC)*

MFCC is one method used to perform feature extraction on a signal. This method is a standard method for character extraction, where this method adopts the workings of human hearing. Sound signals will be filtered linearly for frequencies below 1000 Hz and logarithmic for frequencies above 1000 Hz [2]. The process of the MFCC method can be seen in Figure 2.

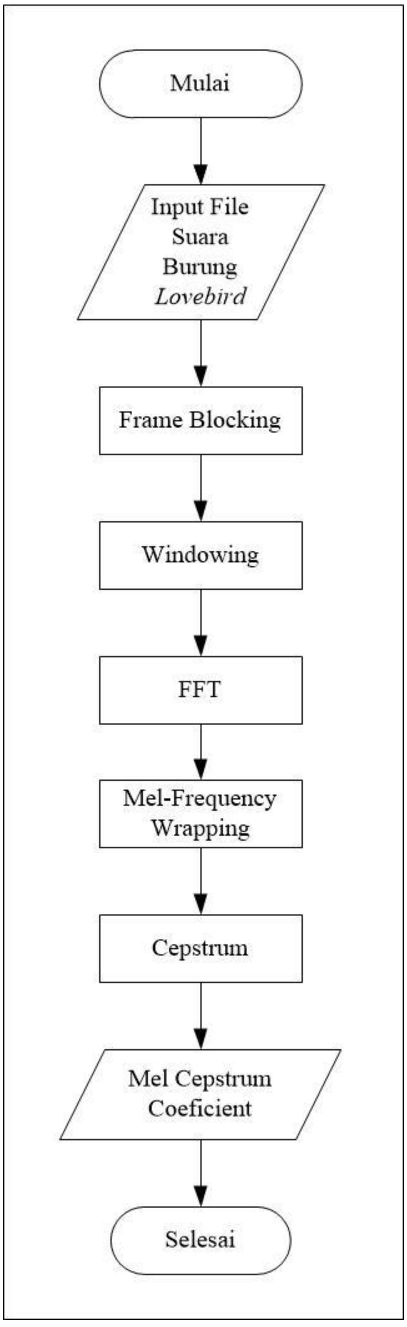
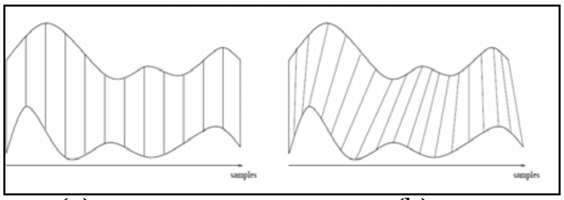


Figure 2. Proses *MFCC*

The feature extraction process is the process of extracting audio data from an incoming sound file to produce a coefficient value. The process starts with reading the input sound file. The incoming sound file is then processed by frame blocking. Frame blocking is the process of dividing the sound into several frames, and one frame consists of several samples. Then each frame produced is windowed to minimize signal discontinuity at the beginning and end of each frame. The next process is FFT (Fast Fourier Transform), which changes each sample frame from the time domain to the frequency domain. The next process is frequency wrapping, where each FFT value is multiplied by the corresponding filter gain, and the results are summed. The final step in the MFCC process is to change the spectrum of time so that it produces MFCC (Mel Frequency Cepstrum Coefficients [12].

2.5 Dynamic Time Warping (DTW)

DTW (DynamicTime Warping) is a method for calculating the distance between two-time series data. The advantage of DTW over other distance methods is that it can calculate the distance of two data vectors of different lengths. The distance of DTW between two vectors is calculated from the optimal warping path of the two vectors. Illustration of matching with the DTW method is shown in Figure 3 below.



(a) (b)

Figure 3. The DTW matching *sequence* a and b

Of the several techniques used to calculate DTW, one of the most reliable is the dynamic programming method. The distance of DTW can be calculated by the formula [3]:

$$D(U,V) = \gamma(m, n)$$

$$\gamma(m, n) = d_{base}(u_i, v_j) + \min \begin{Bmatrix} \gamma(i-1, j) \\ \gamma(i-1, j-1) \\ \gamma(i, j-1) \end{Bmatrix}$$

3. Result and Discussion

The results of this study are to make a lovebird bird sound classification application using speaker recognition technology with MFCC and DTW methods. The results and discussion are presented based on the interface in the application accompanied by pictures from each interface.

3.1 The Lovebird Sound Classification

This page functions to make the lovebird bird sound classification process. There are three menus on the menu bar, namely home, process, and database. Lovebird bird, sound classification process, begins when the user inputs the name of the owner and the name of the bird, after that the user can input the existing sound file or record directly the sound of lovebird birds for 150 seconds. After inputting the owner's name, the bird's name, and the lovebird's voice recording file, the user presses the process button. The application will start processing the sound or recording file that was previously inputted. The first process is the application will divide the sound file into five parts, where each part of the sound file is 30 seconds long. Files that have been shared will be saved into an existing folder. Once saved, the sound file is extracted using the MFCC function, which will then generate the MFCC coefficient from the input data. After getting the MFCC coefficient from the input data, the next step is to extract the training data to get the MFCC coefficient from the training data. After getting the results of the MFCC coefficient from the input data and training data, then it is compared to find the closest distance between the MFCC coefficient from the input data and training data with the DTW function. After completion, the application will display lovebird bird voice data and classes from the audio training data, which have the closest distance in the process table. Then the results are displayed on the interface of the lovebird sound classification process on the right in the result grub box. The interface of the lovebird bird sound classification process can be seen in Figure 4.

Figure 4. Page of lovebirds voice classification

In Figure 4, there is a process table that contains no, time, class, similar file, distance, and path. This table explains the results of bird sound files that have been processed, where each column has a different function. One of them is a class; the column explains the parameters of the lovebird sound class, which is a good, medium, and less class. In a similar column file, it explains the similarity between the audio data input and the existing training data. In the distance column, explain the distance of the MFCC coefficient between the audio data input and the training data that has previously been extracted using the MFCC function, and the final result of the outgoing distance is calculated using the DTW function.

3.2 *Data Training*

The employee database page is the interface used to display all training data that is registered in the database. Users can access the training database page by accessing training actions on the database menu. On this page, the menu bar contains a home, process, and database. The page title is at the top of the table. When the application displays the training database page, all registered training data is displayed in a table. The columns in the table consist of the data-id, no, name, class, and path. There are three buttons located next to the table. The play button functions to play training data sound files. Change button functions to change files. The plus button functions to access the voice data added interface. The training database page interface can be seen in Figure 5.

Figure 5. Data training *database*

3.3 Process Result

The results database page is an interface that is used to display all results data that are registered in a database that has previously been processed. Users can access the results database page by accessing the result's actions in the database menu. On this page, the menu bar contains a home, process, and database. The page title is at the top of the table. When the application displays the results database page, all listed result data is displayed in the results table. The columns in the table consist of the data-id, no, date, time, owner's name, bird's name, class, conclusion, and path. There is a button located below the table. The play button functions to play the result data sound file. The results database page interface can be seen in Figure 6.

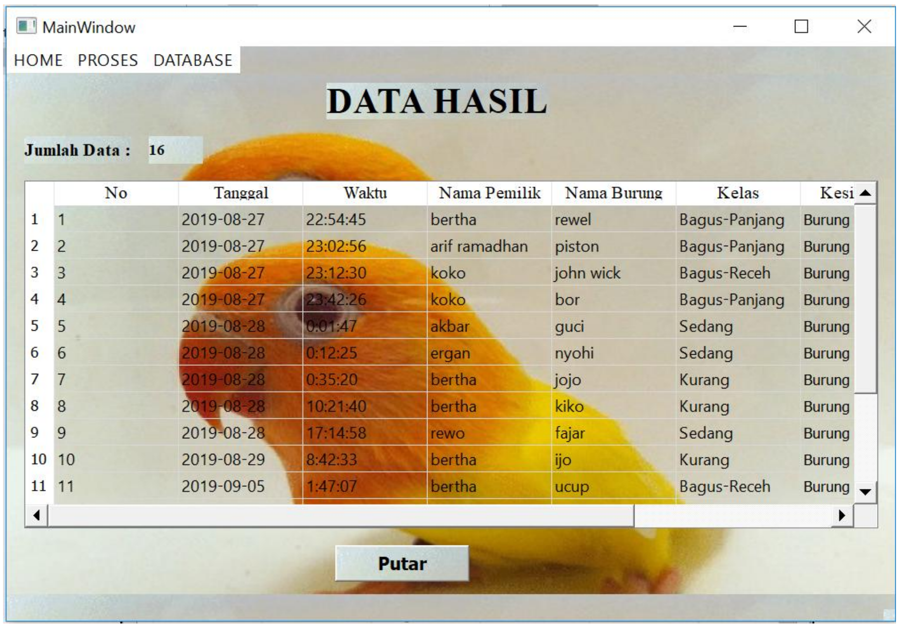


Figure 6. Process result

Based on the analysis, design, and research results, this paper produces an application that has a degree of accuracy in classifying the quality of bird sounds compared to previous studies [2] and [6].

Another difference is in terms of system performance in classifying lovebird quality. Research conducted by Hanesia [2] and Sejati [6] focuses on the number of breaks or the number of pauses that occur when lovebirds emit their chirping in a time of observation. The less the number of pauses, the better the sound quality of the lovebird, conversely the more pauses performed by the lovebird when singing, the sound quality will be worse because it will sound dotted [6]. The results of research by Sejati [6] are getting an accuracy rate of 92.16% with a computing time of 0.1886 seconds and a good MOS category while the results of research by Hanesia [2] obtained an accuracy value of 91.67% with a computing time of 190,229 seconds [2].

However, in this study, the level of system speed in classifying lovebird bird sounds has the disadvantage of relatively long computation time for 150 seconds recording requires 454 seconds of computation time but has been equipped with direct bird sound recording features.

4. Conclusion

Based on the results of the analysis, design, and discussion that have been done before, it can produce sound classification applications to determine the quality of lovebird birds with the method of MFCC and dynamic time warping. The conclusions that can be drawn from this study include:

1. The results of this study can assess the quality of lovebird birds based on their chirping sound classification.
2. The results of this study can reduce fraud cases in choosing quality lovebird birds
3. The Mel-Frequency Cepstral Coefficient and Dynamic Time Warping method can extract and classify lovebird sounds based on the sound quality with the highest accuracy level of 80% on the sound validity test conducted.

Reference

[1] Ariawan, J., W, E. T. B., Alfahmi, N., Stmik, D., Sarana, B., Stmik, M., & Sarana, B. (2016). Sistem Pakar Menentukan Gen Anakan pada Lovebird, *6*(2).

[2] Hanesia, W. I. (2015). Klasifikasi Suara Lovebird Dengan Metode Mel Frequency Cepstral Coefficient (MFCC) Dan Fuzzy Logic Warble Of Lovebird Classification Using Mel Frequency Cepstral Coefficient (MFCC) and Fuzzy Logic, *2*(2), 2968–2976.

[3] Mfcc, M., & Dtw, D. (2011). Verifikasi Biometrika Suara Menggunakan, *2*(1), 8–21.

[4] Permana, I. S., & Nurhasanah, Y. I. (2018). Dan Wanita, *3*(1), 49–63.

[5] Rasyid, M. F. (2018). Aplikasi Pengenalan Penutur Pada Identifikasi Suara Penelepon Menggunakan Mel-Frequency Cepstral Coefficient Dan Vector Quantization (Studi Kasus : Layanan Hotline Universitas Pembangunan Nasional “Veteran” Yogyakarta).

[6] Sejati, T. G., Rizal, A., Gozali, A. A., Prodi, S., Telekomunikasi, T., Elektro, F. T., & Telkom, U. (2015). Klasifikasi Suara Burung Lovebird Dengan Algoritma Fuzzy Logic Classification Of Lovebird Sound ’ S With Fuzzy Logic Algorithm, *2*(2), 3171–3178.

[7] Studi, P., Informatika, T., & Inggris, B. (2012). Speaker Recognition Menggunakan MFCC dan Algoritma DTW, (x), 1–11.

[8] Distance, E., & Citra, P. (n.d.). Pengenalan jenis burung lovebird dengan menggunakan content based image retrival berbasis color histogram, 1–11.

[9] Effendi, Z., Erlina, T., & Aisuwarya, R. (2015). Pengenalan Suara Menggunakan Metode MFCC (Mel Frequency Cepstrum Coefficients) dan DTW (Dynamic Time

Warping) untuk Sistem Penguncian Pintu ISBN : 979-26-0280-1 ISBN : 979-26-0280-1, 239–243.

[10] Elektro, J. T., Brawijaya, U., & Mustofa, A. (2007). Sistem Pengenalan Penutur dengan Metode Mel-frequency Wrapping, *7*(2), 88–96.

[11] Prabandaru, B. (2016a). Ekstraksi Ciri Suara Burung Berdasarkan Domain Waktu dan Frekuensi.

[12] Rasyid, M. F. (2018). Aplikasi Pengenalan Penutur Pada Identifikasi Suara Penelepon Menggunakan Mel-Frequency Cepstral Coefficient Dan Vector Quantization (Studi Kasus : Layanan Hotline Universitas Pembangunan Nasional "Veteran"

[13] Permana, I. S., & Nurhasanah, Y. I. (2018). Dan Wanita, *3*(1), 49–63.

[14] Studi, P., Informatika, T., & Inggris, B. (2012). Speaker Recognition Menggunakan MFCC dan Algoritma DTW, (x), 1–11.