Quadcopter Control Using Speech Recognition

H Malik, S Darma* and S Soekirno*
Department of Physics, Faculty of Mathematics and Natural Sciences, University of Indonesia, Kampus UI Depok, Depok 16424, Indonesia

E-mail : suryadarma@sci.ui.ac.id, santoso.s@sci.ui.ac.id

Abstract. This research reported a comparison from a success rate of speech recognition systems that used two types of databases they were existing databases and new databases, that were implemented into quadcopter as motion control. Speech recognition system was using Mel frequency cepstral coefficient method (MFCC) as feature extraction that was trained using recursive neural network method (RNN). MFCC method was one of the feature extraction methods that most used for speech recognition. This method has a success rate of 80% - 95%. Existing database was used to measure the success rate of RNN method. The new database was created using Indonesian language and then the success rate was compared with results from an existing database. Sound input from the microphone was processed on a DSP module with MFCC method to get the characteristic values. Then, the characteristic values were trained using the RNN which result was a command. The command became a control input to the single board computer (SBC) which result was the movement of the quadcopter. On SBC, we used robot operating system (ROS) as the kernel (Operating System).

1. Introduction
Speech recognition is an algorithm that makes the machine can recognize the voice spoken by humans. Research on speech recognition had been started since the 1930s. But the development began to look significant around the 1960s.

In speech recognition there are two major stages, they are feature extraction and classification. Each stage has its own methods. Some researchers have tried several methods such as cross-correlation, Mel frequency cepstral coefficient, hidden Markov model, artificial neural network and so on. However, for a relatively crowded environment, the success rate of the speech recognition system was relatively decreased [1]. Therefore, some researchers such as Yanmin Qian et al. in his paper proposed ways to improve speech recognition for environments with relatively more noise with the Deep Neural Network (DNN) method [1]. In addition, various methods of feature extraction stage in speech recognition have various degrees of success rate [2]. The Mel Frequency Cepstral Coefficient (MFCC) method had an error rate of 7.21-17.8%, while with Mel Filter-Bank (FBANK) had an error rate of 1.35-6.82% [2].

The methods used to create speech recognition systems had various types. Such as used a combination of Mel frequency cepstral coefficient (MFCC) as feature extraction with Dynamic Time Wrapping (DTW) as shifting features and K-Nearest Neighbor (KNN) as a classification [3], combined method between Mel frequency cepstral coefficient (MFCC) as feature extraction with Vector quantization (VQ) and Gaussian Mixture Models (GMM) [4], combined method between Mel
frequency cepstral coefficient (MFCC) as feature extraction with Linear Discriminant Analysis (LDA) as a classification [5], and the method of combination of Discrete wavelet Transform (DWT) as feature extraction with Linear Discriminant Analysis (LDA) as a classification [5]. Each method had its own success rate, as in [3] with a success rate of 98.4%, and at [4] with a success rate of 94.31%.

In this research, the authors designed a speech recognition system as a quadcopter control tool using Mel Frequency Cepstral Coefficient (MFCC) method as feature extraction and Recursive Neural Network (RNN) as classification that is expected can produce better results, and then made a comparison from two types of databases, they are existing databases with new databases.

2. The Proposed Method

MFCC, as described by Davis and Mermelstein, was as a profitable approach to speech recognition. MFCC is one of the most common and familiar algorithms, used in signal applications for feature extraction. The following steps were used in this method which was summarized in Figure 1. An explanation of the MFCC already existed in the paper [3] [4] [5].

![Figure 1. MFCC Process Stages [3][4][5]](image)

In this research, the quadcopter was controlled using voice commands. The sound signal was captured by the microphone which was processed using the DSP module that had been available on the laptop by using the software.

Then the characteristic value that was obtained from MFCC was trained using Recursive Neural Network [1] [6] [7]. In this research, Recursive Neural Network type that used was Recurrent Neural Network [6].

After the training process, then we did the testing process to see the results when inputted voice command which output was worth close to the target value of each word. The output value of the test result was the input condition for moving the quadcopter. ‘naik’ command represented ‘take off’, ‘bawah’ command represented ‘landing’, ‘maju’ command represented ‘positive pitch’, and ‘balik’ command represented ‘positive yaw’. Here it was a flowchart of the created system.

![Figure 2. Speech Recognition System Flow Chart](image)
The quadcopter movement was controlled by using the PID controller [8] created by removing the Integral component into a PD controller. In this research, the PD controller served to adjust the amount of distance that was taken by the quadcopter. The PD controller was a controller that the output signal was directly proportional to the error signal and error rate. The equation for the PD controller was,

\[ PD = k_p e_p + k_d e_d \]  

where \( k_p \) was a proportional constant, while \( k_d \) was a derivative constant, \( e_p \) was an error of a proportional component, and \( e_d \) was an error of a derivative component.

3. Evaluation and Experimental Result

In this research, the sound database was needed from every word of the command used. We used an existing database and new database then compared both of them. The new database that created then was trained using the Recurrent Neural Network with the target value ‘1’ for ‘maju’, ‘2’ for ‘mundur’, and ‘3’ for ‘kanan’. After the database was trained, the calibration data testing process did ten times for each word. From the testing process we obtained the following results.

**Table 1. Success rate of characterization test**

| Command | Maju | Mundur | Kanan | Success Rate (%) |
|---------|------|--------|-------|------------------|
| Maju    | 4    | 6      | 0     | 40               |
| Mundur  | 5    | 5      | 0     | 50               |
| Kanan   | 8    | 2      | 0     | 0                |

After calibration data processing, then the data was tested twenty times with four types of words that were ‘maju’, ‘mundur’, ‘naik’, and ‘balik’. After that, we compared existing databases with new databases that were created in this research, and the result was shown in Figure 3.

**Figure 3. (a).** The success rate of each command used existing databases for 20 experiments. **(b).** The success rate of each command used new databases for 20 experiments

**Table 2. Fine tuning data results**

| No | Kp | kd | Target Distance (cm) | Reached Distance (cm) | Time (second) |
|----|----|----|----------------------|-----------------------|---------------|
| 1  | 0  | 1  | 200                  | 245                   | 2.844         |
| 2  | 0.02 | 0.9 | 200                  | 235                   | 3.32          |
The quadcopter movement was controlled by using the PID controller. So, we had to calibrate the controller to find $k_p$ and $k_d$. Here, it was the results of calibration data to find $k_p$ and $k_d$ values that had precision distance and the least travel time.

The $k_p$ and $k_d$ values that yield the best value were $k_p = 0.14$ and $k_d = 0.3$. From the value of $k_p$ and $k_d$, then tested PID controller with three tests. Here it was the results of PID testing.

| No | Target Distance (cm) | Reached Distance (cm) | Time (second) |
|----|----------------------|-----------------------|---------------|
| 1  | 200                  | 200                   | 2.766         |
| 2  | 200                  | 194                   | 3.263         |
| 3  | 200                  | 190                   | 3.345         |

Table 3. PID test results

4. Conclusion and Future Work
Based on the test and discussion data, we could conclude that, Speech Recognition System created could already recognize the spoken word, the success rate for each word used existing databases was 100% except for the word ‘balik’ by 95%. While, the success rate for each word used new databases was 45% for ‘naik’, 75% for ‘bawah’, 25% for ‘balik’, and 55% for ‘maju’. From that, we could conclude that a new databases still need more improvement. Based on table 2, the value of $k_p$ and $k_d$ that produced the best value was when $k_p = 0.14$ and $k_d = 0.3$ with distance reached = 200 cm and time = 2.766 seconds.

In making speech recognition system there was still much to be improved. Therefore, there were some suggestions that could be did for the development of the next system such as, making databases had to be from the more varied responders in order to increase the quality of the databases created, the sound of the quad-spin propeller motion could increase the disturbance for the system made, so it should be added filter to resolve this, as well as PID controller that made could still be refined again for the value of its response to be better.

5. Acknowledgment
This research was funded by Hibah PITTA UI, under contract number 679/UN.2.R3.1/HKP.05.00/2017

References
[1] Y Qian, T Tan, and D Yu 2016 Neural Network Based Multi-Factor Aware Joint Training for Robust Speech Recognition vol 24, no 12, pp 2231–2240 (IEEE/ACM Trans. Audio Speech Lang. Process)
[2] H B Sailor and H A Patil 2016 Novel Unsupervised Auditory Filterbank Learning Using Convolutional RBM for Speech Recognition no 99, pp 2341–2353 (IEEE/ACM Trans. Audio Speech Lang. Process)
[3] M A Imtiaz 2016 Isolated Word Automatic Speech Recognition (ASR) System using MFCC, DTW & KNN pp 106–110 (Asia Pacific Conference on Multimedia and Broadcasting (APMediaCast))

[4] U G Patil 2016 Automatic Speech Recognition of Isolated Words in Hindi Language using MFCC pp 433–438 (International Conference on Computing, Analytics and Security Trends (CAST))

[5] H Ali, X Zhou, and S Tie Comparison of MFCC and DWT Features for Automatic Speech Recognition of Urdu pp. 154–158

[6] X Chen, X Liu, Y Wang, M J F Gales, and P C Woodland 2016 Efficient Training and Evaluation of Recurrent Neural Network Language Models for Automatic Speech Recognition vol 24, no 11, pp 2146–2157 (IEEE/ACM Trans. Audio Speech Lang. Process)

[7] R Setiono, B Baesens, and C Mues 2008 Recursive neural network rule extraction for data with mixed attributes vol 19, no 2, pp 299–307 (IEEE Trans. Neural Networks)

[8] B. Yasmina and A. Zoubir 2017 Dynamic Modeling, Simulation and PID Controller of Unmanned Aerial Vehicle UAV no. Intech, pp. 1–6