



Frequency domain analysis of MFCC feature extraction in children's speech recognition system

Risanuri Hidayata^{1*}

¹Universitas Gadjah Mada

¹Yogyakarta, Indonesia

*Corresponding email: risanuri@ugm.ac.id

Received 10 November 2021, Revised 17 November 2021, Accepted 16 February 2022

Abstract —The research on speech recognition systems currently focuses on analyzing robust speech recognition systems. In this paper, we conducted an advanced analysis on one stage of the Mel Frequency Cepstral Coefficients (MFCC) process, the Fast Fourier Transform (FFT), in children's speech recognition system. The FFT analysis in the feature extraction process determined the effect of frequency value characteristics utilized in the FFT output on the noise disruption. The analysis method was designed into three scenarios based on the value of the employed FFT points. The difference between scenarios is based on the division of FF points which in this study are split manually on the MFCC algorithm. This study utilized children's speech data from the isolated TIDIGIT English digit corpus. The results showed that using a particular frequency portion following the scenario designed on MFCC affected the recognition system performance, which was relatively significant on the noisy speech data. The designed method in the scenario 3 (C1) version generated the highest accuracy, exceeded the accuracy of the conventional MFCC method. The average accuracy in the scenario 3 (C1) method increased by 1% more than all the tested noise types. Using various noise intensity values (SNR), the testing process indicates that scenario 3 (C1) generates a higher accuracy than conventional MFCC in all tested SNR values. It proves that the selection of specific frequency utilized in MFCC feature extraction significantly affects the recognition accuracy in a noisy speech.

Keywords – Fast fourier transform, Frequency domain, Mel frequency cepstral coefficients, Speech recognition

Copyright © 2022 JURNAL INFOTEL

All rights reserved.

I. INTRODUCTION

One example of technological development is the implementation of robotic technology. Robots invented as daily housework assistant is a promising robotic technology target. Robots are expected to be able to complete housework in natural and friendly manners [1]. One of the implementations of robotic technology widely used in the latest technology is the automatic speech recognition system [2]. Automatic speech recognition systems are often used in technology to facilitate human activities, such as security systems, electronic switches, language translators, and automatic device control. Many researchers consecutively carried out the development of speech recognition system technology. One of the main focuses of popular research today is the

development of a robust speech recognition system. Speech recognition by the human hearing system notably depends on the perception and speech signal input features [3]. When the speech signals are combined with noise, the recognition system becomes distracted. They struggle to identify the speech sounds. Therefore, the development of a noise-resistant recognition system continues to be carried out. The principle of an anti-noise speech recognition system is to eliminate noise from the speech signals and restore the original information signals [2]. However, noise is generally unpredictable; therefore, the system cannot accurately extract the original speech signal.

One of the important stages and the focus of research by many researchers in the speech recognition process is the feature extraction process. The features

signify the most distinct properties drawn from speech signals consisting of pauses, silence, and other information [4]. The Mel Frequency Cepstral Coefficients (MFCC) method is one of the most widely used methods to extract speech signal features. Some studies have been conducted by developing and analyzing the MFCC method for robust speech recognition systems. Several studies have been carried out using a Multitaper MFCC for a robust recognition system development [5], [6]. The developed method in the study focuses on the windowing process at MFCC. Research on robust speech recognition development using MFCC has also been done by analyzing the filterbank process [7]. The researchers compared the performance of triangular mel filter utilization with the Barc-scale and Gammatone filterbank. Research on a frequency domain analysis has been carried out [8]; however, it is limited to focusing on sampling frequency values. The researchers tested several sampling frequencies values (8 kHz, 16 kHz, 32 kHz, and 44.1 kHz). From several studies that have been carried out before, it is proven that the analysis results at the MFCC process stages also affect the recognition system performance. The MFCC method focuses on data analysis in the frequency domain. The speech signal is closely related to the frequency range. Certain frequency ranges are very susceptible to noise. Therefore, research on frequency domain analysis at the MFCC process stage is very important to do. In this paper, the researchers analyzed the frequency domain at one of the MFCC stages, namely, the Fast Fourier Transform (FFT). The FFT converts speech data in the time domain into data in the frequency domain. Human speech has various frequency value characteristics. There are differences in human speech and noise characteristics, one of them is in the frequency value. Therefore, an analysis is needed to determine the effect of the characteristic frequency values used in the FFT process on noise disturbance.

Various human speech recognition developments have been carried out with adult human speech as objects. However, the recognition accuracy will significantly decrease when the system is trained with an adult speaker and applied to recognize children's speech. The decrease in accuracy occurs due to differences in the acoustic and linguistic characteristics of children and adults. The distinguishing thing is mainly on the differences in articulatory control, morphological differences in vocal channel geometry, and a smoother ability to control segmental features [9]. Unfortunately, there is not much research on children's speech recognition systems found. Based on that, the needs of speech recognition development with children's speech objects are still essential. Some studies using children's speech were carried out by Rahman et al. [10] and Soe Naing et al. [7]. The researchers are working to develop a speech recognition system for Malay-speaking children using a small speech database. The ASR system developed

in the study can recognize words with 76% accuracy. In addition, Soe Naing et al. [7] analyzed the filterbank process on MFCC features extraction method in children's speech recognition systems. The utilized database was the TIDIGITS corpus with 101 children's speakers (50 male and 51 female). Analysis results in the filter bank process are proven to affect the speech recognition system performance.

In this paper, researchers develop previously researched children's speech recognition [7] by analyzing the FFT process on the MFCC feature extraction method. The FFT output analysis in the MFCC feature extraction method was carried out to determine the effects of frequency value characteristics utilized against noise disruption. The analysis method was designed in three scenarios based on the utilized FFT point value. FFT coefficient is divided into sections with different scenarios. The results were then applied to the classification process, and the accuracy was analyzed. The best result was reported on the FFT point scenario with the highest accuracy. The utilized speech data were boys' and girls' speech obtained from the TIDIGITS corpus.

II. RESEARCH METHODS

The utilized speech data were children's speech data in the form of eleven isolated words in English ("zero" to "nine" and "oh"). The speech data was extracted from the TIDIGITS speech corpus of several thousand continuous digit speeches (available from the Linguistic Data Consortium) [11]. The data were obtained from the TIDIGITS corpus and limited only to children's speech with 101 speakers with two iterations. Therefore, the total speech data utilized in this study was 2222 speeches (11 digits x 101 speakers x 2 times), and they were divided into training and test data. The authors then added the noise obtained from the AURORA database into the test data. The added noise consisted of various environments (subway, babble, car, street, and restaurant) and SNR (5dB, 10dB, 15dB, and 20dB). The TIDIGIT corpus provided the proportion of training and test data with details: 1122 training data and 1100 test data, then applied to all feature extraction methods (proposed and conventional methods) testing processes.

The conducted discussion focused on feature extraction in the speech recognition system. The speech recognition system has several essential stages in recognizing the speech signals, from the speech signals recording, initial processing, feature extraction, and classification [12]. The Mel Frequency Cepstral Coefficients (MFCC) method was used as a feature extraction method, while the Support Vector Machine (SVM) method was used for the classification.

A. Initial Process

Several stages were conducted at the initial processing to smooth the previously recorded speech signals. The stages in the initial processing included

the DC removal filtering process and speech signals data cut. DC bias is the average of the non-zero digital time samples. The unwanted DC bias on a signal can cause problems. The DC removal filtering was used to remove silent signals found at the beginning and the end of speech signals. In comparison, the data cut process was carried out to uniform the length of signal data utilized in the feature extraction and classification processes. It was essential to carry out because the data length value would affect the accuracy and computation time of the speech recognition system. Speech data on the TIDIGIT corpus had varying speech duration. The determination of the length of voice data was based on references to previous studies [13]. This study employed 4.540 data lengths with a sampling frequency of 20.000; therefore, the speech duration used in this study was 0.227 seconds.

B. Feature Extraction

The feature extraction process was carried out to retrieve special features on each speech signal. These features were used to distinguish between speech signals from each other. The employed feature extraction method was Mel Frequency Cepstral Coefficients (MFCC). The MFCC is one of the feature extraction methods prevalent in speech recognition systems [14]. Some key stages of the MFCC process can be seen in Figure 1.

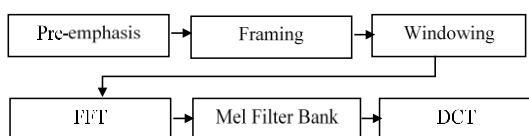


Fig.1. Processes in MFCC feature extraction method

The MFCC feature extraction process starts from pre-emphasis that compensates for speech signals at high frequency [15]. The mathematical equation for the pre-emphasis process can be seen in equation (1).

$$Y[n] = X[n] - 0.95 X[n - 1] \quad (1)$$

where $Y[n]$ is the output signal, and $X[n]$ is the input signal. After the pre-emphasis process, the speech signals will go through the Framing and Windowing processes. The framing process divides the signal into several overlapping parts to avoid the signal being interrupted due to the cutting process [16]. In this study, the cut speech data were divided into several frames with a data length of 512 for each frame. In addition, an overlap with a value of 100 on the front and back of each frame was arranged. The FFT process converted speech signals into frequency domain following the equation (2).

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N} \quad (2)$$

where X_n is the FFT output, X_k is the input signals, and N has a value of $0, 1, 2, \dots, N-1$. In this study, the utilized FFT point value was 256. The speech information signal was then processed in Mel filterbank. Mel filterbank is shaped like a bandpass filter with linear properties below 1000Hz frequency and logarithmic above 1000Hz [17]. The final process of the feature extraction is the Discrete Cosine Transform (DCT), following the equation (3).

$$\sum_{k=1}^N \log(Y(i)) * \cos[mx(k - 0.5)x\pi \div N] \quad (3)$$

where Y is the input data, N is the number of triangular bandpass filters, m has a value from 1 to L , and L is the number of Mel-scale cepstral coefficients. Thus, the number of MFCC coefficients used in this study was 13 coefficients referring to previous studies [12], [14], [18].

The employed classification method was the support vector machine (SVM). Support Vector Machine (SVM) is a supervised machine learning model. This model aims to calculate and create a hyperplane that classifies all training vectors. After creating a hyperplane, the next step was to determine the maximum margin between the data points and the hyperplane, referred to as a support vector [19]. The optimal hyperplane in the SVM classifier will result in a better classification process. The optimal hyperplane is the hyperplane that has the maximum margin value [20].

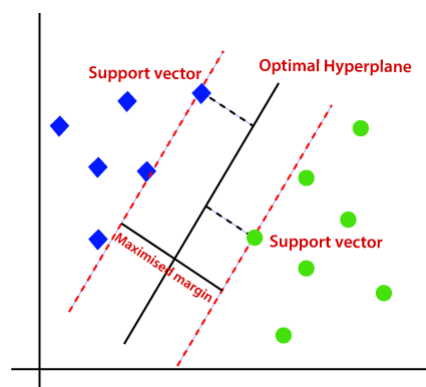


Fig.2. Linear SVM example [19]

C. Frequency Domain Analysis

One of the stages of the MFCC process is Fast Fourier Transform (FFT) which corresponds to equation (2). FFT is a fast algorithm development in implementing Discrete Fourier Transform (DFT). DFT computation time is too long and inefficient, so FFT can perform efficiency calculations [21]. The FFT processing results would produce speech data in a frequency domain. Each speech data processed in the FFT would be converted into 256 FFT point data representing the frequency magnitude. A greater the FFT point values indicated a greater the frequency values. An example of an FFT output display in Matlab

software for one of the speech data can be seen in Figure 3.

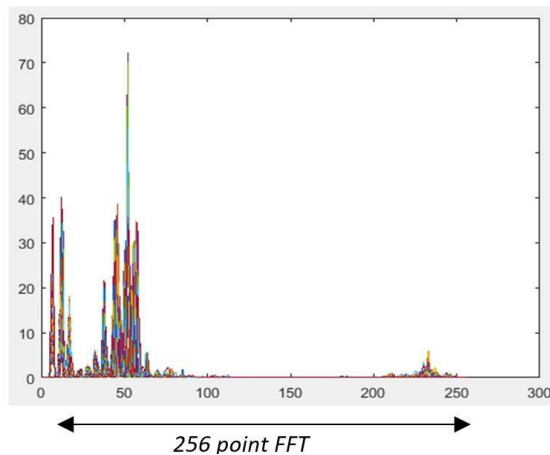


Fig.3. FFT process outputs on MFCC feature extraction

The analysis carried out in this study employed FFT points representing the speech data frequency. It was carried out by separating speech data (in the FFT points domain) containing essential and less significant information. Besides, an analysis was carried out by identifying vulnerable-to-noise speech data. Speech data yielded from the FFT output process would be identified; therefore, not all n point FFT were used. The method was divided into three scenarios to determine which part of the n point FFT was used. In each scenario, the frequency of the FFT process was divided into several ranges. First, it was done to separate FFT points that contain important information and FFT points which contain noise. Hence, later, it could be performed to eliminate the data in the FFT points containing noise to improve recognition performance.

a) First Scenario: In this scenario, the points were divided into four versions, i.e., Low Frequency 1 (LF1), Low Frequency 2 (LF2), High Frequency 1 (HF1), and High Frequency 2 (HF2). All of the utilized FFT data points in each version had equal size; therefore, all FFT data points were divided into four parts. Each version employed 64 FFT data points. The application of the FFT data point in the first scenario can be seen in Table 1.

Table 1. The FFT point data utilized in the first scenario

Version	FFT Points
LF1	1 – 64
LF2	65 – 128
HF1	129 – 192
HF2	193- 256

b) Second Scenario: In the second scenario, the points were divided into three versions, i.e., the Low Frequency (LF), Middle Frequency (MF), and High Frequency (HF). All of the utilized FFT data points in each version had equal size; therefore, all FFT

data points were divided into three parts. Each version employed 85 FFT data points; however, the HF version employed 86 FFT data points. The application of the FFT data point in the third scenario can be seen in Table 2.

Table 2. The FFT point data utilized in the second scenario

Version	FFT Points
LF	1 – 85
MF	86 – 170
HF	171 – 256

c) Third Scenario: The third scenario was designed by combining low and high frequencies. In this scenario, the method was divided into two versions, i.e., Combination 1 (C1) and Combination 2 (C2). The application of the FFT data point in the third scenario can be seen in Table 3.

Table 3. The FFT point data utilized in the third scenario

Version	FFT Points
C1	1 – 64 & 193 – 256
C2	1– 85 & 171 – 256

III. RESULT

In the first part, the recognition system is tested using the conventional MFCC method by including all the FFT coefficients. The testing was carried out using clean speech data and various 10dB noisy speech data (car, restaurant, street, subway, and babble). It was conducted to determine the proposed method's performance in various speech environments (clean and noisy). The test results can be seen in Table 4.

Table 4. Recognition accuracy (%) using the conventional MFCC method

Clean	Car	Rest	Street	Subway	Babble
91	65.27	65.18	65.63	65.36	65.72

Table 4 shows the accuracy of recognition results in clean speech and noisy speech in various environments. The employed feature extraction method was the conventional MFCC by using all frequencies at the FFT output. The recognition results on noisy speech data show a significant decrease in accuracy compared to clean speech data.

A. First Scenario

The system was tested using the designed method in several scenarios. In the first scenario, the FFT output is divided into four parts according to the value of the frequency range. The low frequency was divided into two parts, i.e., LF1 and LF2, as well as the high frequency, i.e., HF1 and HF2. The results in the first scenario are shown in Table 5.

Table 5. Recognition accuracy (%) in the first scenario

Noise	Recognition accuracy in each version (%)			
	LF1	LF2	HF1	HF2
Clean	91	90.54	90.45	91.09
Car	64.27	65.81	63.54	66.24
Rest	65.09	66.27	62.72	66.31
Street	65	65.54	64	65.9
Subway	64.9	65.9	63.27	66.53
Babble	65.27	66	64.09	66.45

B. Second Scenario

Subsequently, the system was tested using the second scenario. In the second scenario, the FFT output was divided into three parts according to its frequency magnitude, i.e., the low frequency (LF), middle frequency (MF), and high frequency (HF). The test results of the second scenario are shown in Table 6.

Table 6. Recognition accuracy (%) in the second scenario

Noise	Recognition accuracy in each version (%)		
	LF	MF	HF
Clean	90.9	90.72	91
Car	65.18	60.54	64.18
Rest	65.36	60.36	64.45
Street	65.45	60.45	64.27
Subway	64.63	59.9	63.9
Babble	65.63	60.9	64.27

C. Third Scenario

The third scenario was designed using a combination of low and high frequencies. The utilized frequency combinations were designed into two types (C1 and C2). The test results of the third scenario are shown in Table 7.

Table 7. Recognition accuracy (%) in the third scenario

Noise	C1	C2
Clean	91.27	90.90
Car	66.36	63.72
Rest	66.54	64.18
Street	66.81	64.54
Subway	66.54	64.54
Babble	66.54	64.36

IV. DISCUSSION

A comparison between the performances of all proposed scenarios with conventional MFCC is shown to determine the performance of all designed scenarios.

The recognition results on the proposed method are compared with the conventional MFCC method which has been popularly used by several researchers [12]-[14]. The best performances of each scenario were the HF2 version in the first scenario, LF in the second scenario, and C1 in the third scenario.

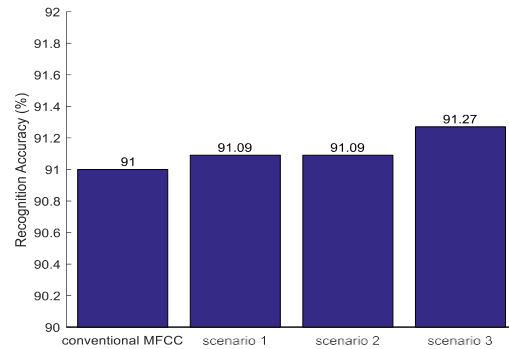


Fig.4. A comparison of recognition accuracy of all methods in the clean speech

Figure 4 compares the recognition accuracy of all methods in the clean speech data. The recognition accuracy of all scenarios shows almost the same results, with the most significant difference of only 0.27%. It shows that using specific frequencies according to scenarios designed in the MFCC feature extraction method does not significantly affect the recognition system performance on the clean speech data. However, as shown in Figure 5, there is a considerably significant accuracy difference in the noisy speech data.

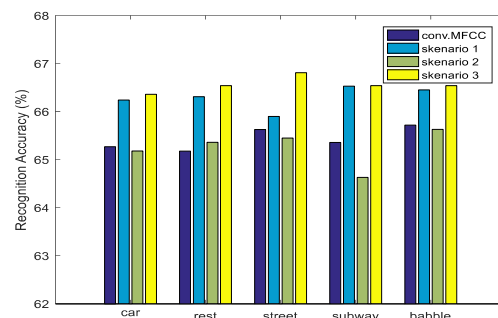


Fig.5. A comparison of recognition accuracy of all methods scenarios in the noisy speech

The recognition results in the noisy speech data of the designed method scenario 3 version C1 generates the highest accuracy, exceeding the accuracy of the conventional MFCC method. The average accuracy increase in the scenario 3 (C1) method was more than 1% in all noise types. Scenario 3 (C1) method employed a combination of data at low and high frequencies and ignored data in the middle frequency. For a total of 256 FFT points, the ignored data in scenario 3(C1) were the 65-191 data. It shows minimal information in the data, and it is also susceptible to noise. Hence, if the data is ignored in the feature extraction process, it will generate better accuracy. At

the same time, the data used in scenario 3(C1) were 1-64 and 193-256 FFT points. Therefore, the data was considered to contain much information from the speech data that would be recognized. The design of the FFT point application in scenario 3(C1) method can be seen more obviously in Figure 6.

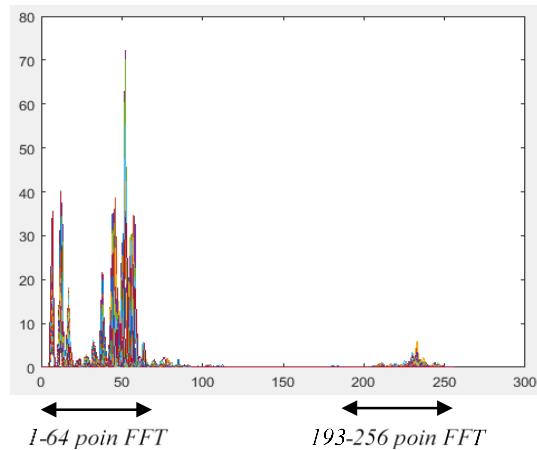


Fig.6. A design of some FFT points application in scenario 3 (C1) as the best method

A test on noisy speech with various SNR values was conducted to evaluate the method performance in scenario 3(C1). Figure 7 shows a recognition accuracy comparison between the conventional MFCC method with the scenario 3(C1) method. The conventional MFCC method applies all frequencies (256 FFT points). In contrast, the scenario 3 (C1) method ignores the middle frequency (65-191 FFT points). The results show that the scenario 3 (C1) method presents higher accuracy in all tested SNR values. It proves that the selection of specific frequency utilized in MFCC feature extraction significantly affects the recognition accuracy in a noisy speech.

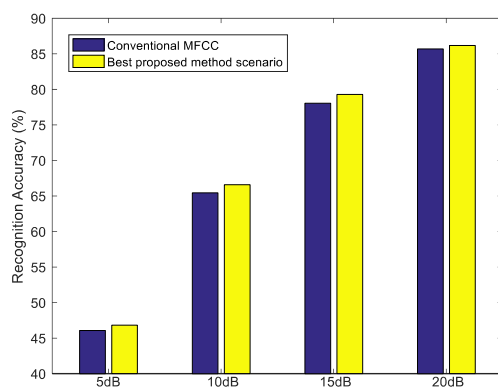


Fig.7. Accuracy comparison of conventional method and frequency modification method (scenario 3-C1 method)

V. CONCLUSION

The application of certain frequencies (FFT points) according to the scenario designed in MFCC does not

significantly affect the performance of the children's speech recognition system on clean voice data. However, there is a significant difference in the noise data. The designed method in scenario 3 version C1 generates the highest accuracy by 91.27% (clean speech), 66.36% (car noise), 66.54% (restaurant noise), 66.81% (street noise), 66.54% (subway noise), and 66.54% (babble noise). The proposed modification method is proven to increase the accuracy of the speed recognition system using the conventional MFCC method. The average accuracy increase in the scenario 3 (C1) method is more than 1% in all tested noise types (car, street, restaurant, subway, babble). Tests carried out using various values of noise intensity (5dB, 10dB, 15dB, 20dB), showed that scenario 3 (C1) method produced higher accuracy than the conventional MFCC method on all tested SNR values. The best performance of the proposed scenario 3 (C1) modification method shows that the results of the FFT process analysis on the MFCC feature extraction method have been successfully carried out. Scenario 3 (C1) is designed using a combination of FFT points at low and high frequencies. The results show that a combination of 1-64 points FFT and 193-256 points FFT can produce the best recognition performance. It shows that the noise is located at the FFT point in the middle (medium frequency). Therefore, eliminating the FFT points in the middle will improve the recognition system performance. It proves that the selection of specific frequency utilized in MFCC feature extraction significantly affects the recognition accuracy in a noisy speech. Recommendations for further research are frequency domain analysis for voice data of all ages and various genders. The MFCC method can be combined with a wavelet transform-based denoising method. The development method must be able to produce good performance on all human speech objects.

REFERENCES

- [1] P. N. Hung and T. Yoshimi, "Extracting Actions from Instruction Manual and Testing Their Execution in A Robotic Simulation," *ASEAN Engineering Journal*, Volume 6, No.1, pp. 47-58, 2016.
- [2] J. Liu and X. Xiang, "Review of the anti-noise method in the speech recognition technology," In: *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Siem Reap, pp. 1391-1394, Jun. 2017.
- [3] Md. S. Alam, M. S. A. Zilany, and E. Davies-Venn, "Effects of speech-shaped noise on consonant recognition in Malay," In: *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, pp. 22-25. 2017.
- [4] A. N. Chadha, M. A. Zaveri, and J. N. Sarvaiya, "Optimal feature extraction and selection techniques for speech processing: A review," In: *2016 International Conference on*

- Communication and Signal Processing (ICCSP)*, Melmaruvathur, Tamilnadu, India, pp. 1669–1673, 2016.
- [5] K. P. Bharath and R. Kumar M., "Multitaper Based MFCC Feature Extraction for Robust Speaker Recognition System," In: *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vellore, India, pp. 1–5, 2019.
- [6] K. V. Veena and D. Mathew, "Speaker identification and verification of noisy speech using multitaper MFCC and Gaussian Mixture models," In: *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Thrissur, India, pp. 1–4, 2015.
- [7] H. M. Soe Naing, Y. Miyanaga, R. Hidayat, and B. Winduratna, "Filterbank Analysis of MFCC Feature Extraction in Robust Children Speech Recognition," In: *2019 International Symposium on Multimedia and Communication Technology (ISMATC)*, Quezon City, Philippines, pp. 1–6, 2019.
- [8] O. L. Baroi, A. Niaz, Md. J. Islam, Md. S. A. Kabir, E. Islam, and Md. J. Rahimi, "Effects of Different Environmental noises and Sampling Frequencies on the performance of MFCC and PLP based Bangla Isolated Word Recognition System," In: *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh, pp. 1–6, 2019.
- [9] D. Giuliani and M. Gerosa, "Investigating recognition of children's speech," In: *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hongkong, China, pp. II–137, 2003.
- [10] F. D. Rahman, N. Mohamed, M. B. Mustafa, and S. S. Salim, "Automatic speech recognition system for Malay speaking children," In: *2014 Third ICT International Student Project Conference (ICT-ISPC)*, Nakhon Pathom, Thailand, pp. 79–82, 2014.
- [11] R. Leonard, A database for speaker-independent digit recognition, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'84), vol.9, pp.328-331, 1984. [32].
- [12] A. Winursito, R. Hidayat, A. Bejo, and M. N. Y. Utomo, "Feature Data Reduction of MFCC Using PCA and SVD in Speech Recognition System," In: *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, Kuala Lumpur, Malaysia, pp. 1–6, 2018.
- [13] R. Hidayat and A. Winursito, "Analysis of Amplitude Threshold on Speech Recognition System," *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020, pp. 449-453
- [14] R. Hidayat, A. Bejo, S. Sumaryono, and A. Winursito, "Denosing Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System," In: *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Bali, Indonesia, pp. 280–284, 2018.
- [15] N. Alcaraz Meseguer, "Speech Analysis for Automatic Speech Recognition," Department of Electronics and Telecommunications, Norwegian University of Science and Technology, Gjøvik, Norway, 2009.
- [16] Adiwijaya, M. N. Aulia, M. S. Mubarak, W. U. Novia, and F. Nhita, "A comparative study of MFCC-KNN and LPC-KNN for hijaiyyah letters pronunciation classification system," In: *2017 Fifth International Conference on Information and Communication Technology (ICOICT)*, Melaka, Malaysia, pp. 1–5, 2017.
- [17] A. Vijayan, B. M. Mathai, K. Valsalan, R. R. Johnson, L. R. Mathew, and K. Gopakumar, "Throat microphone speech recognition using MFCC," In: *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, Trivandrum, Kerala, India, pp. 392–395, 2017.
- [18] O. Eltiraiifi, E. Elbasheer, and M. Nawari, "A Comparative Study of MFCC and LPCC Features for Speech Activity Detection Using Deep Belief Network," in *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEE)*, Khartoum, pp. 1–5, 2018.
- [19] R. Y. Rumagit, G. Alexander, and I. F. Saputra, "Model Comparison in Speech Emotion Recognition for Indonesian Language," *Procedia Comput. Sci.*, vol. 179, pp. 789–797, 2021
- [20] M. Ameliasari, A. G. Putrada, and R. R. Pahlevi, "An Evaluation of SVM in Hand Gesture Detection Using IMU-Based Smartwatches for Smart Lighting Control," *J. INFOTEL*, vol. 13, no. 2, pp. 47–53, May 2021.
- [21] H. Heriyanto, T. Wahyuningrum, and G. F. Fitriana, "Classification of Javanese Script Hanacara Voice Using Mel Frequency Cepstral Coefficient MFCC and Selection of Dominant Weight Features," *J. INFOTEL*, vol. 13, no. 2, pp. 84–93, May 2021.