Identification of baby cry with Discrete Wavelet Transform, Mel Frequency Cepstral Coefficient and Principal Component Analysis

Anggunmeka Luhur Prasasti¹, Ledya Novamizanti², Muhammad Ichwannul Razik³

School of Electrical Engineering, Telkom University
Jl. Telekomunikasi Terusan Buah Batu Bandung 40257 Indonesia

E-mail: ¹anggunmeka@telkomuniversity.ac.id, ²ledyaln@telkomuniversity.ac.id, ³muhammadichwannulrazik@gmail.com

Abstract - A baby’s cry is a sign of a baby that shows the feelings and desires of a baby. However, many people misinterpret the baby’s cry so that less precise handling happens often. Based on the research from Dunstan Baby Language (DBL), there are 5 types of languages used by infants namely NEH (Hungry), HEH (Discomfort), EAIRH (Lower Wind/Gas), EH (Burp), and OWH (Tired). This research designed the baby’s voice-based speech processing sound identification system. The baby’s cry is recorded by using the audio record feature on the smartphone then get the extraction feature by using Discrete Wavelet Transform (DWT). DC removal and pre-emphasis stages are needed as pre-process. The sound signal is performed extraction feature by Mel Frequency Cepstral Coefficient (MFCC) and Principal Component Analysis (PCA) methods. The result of the feature extraction will be classified with Euclidean distance to measure the resemblance of the extraction value of each cry by calculating the difference in distance 2 matrices of features. From 150 training data and 50 test data, this system can identify the sound of crying babies on 5 conditions of a baby crying, namely Neh (Hungry), Heh (Discomfort), Eairh (Lower Wind/Gas), Eh (Burp), and Owh (Tired). The best parameter is obtained at a frame size of 1024 data per frame, MFCC feature coefficient of 32, DWT at level 1, and DB 2. This system can detect the baby crying sound with the best accuracy of 90% and computing time 0.5542 seconds.

1. Introduction
When a new baby is born into the world, the first thing that the baby is doing is crying. The baby’s cry is ordinary and indicates the baby is born safely and normally. By crying, the baby can communicate with both parents because the cry of newborn babies is one form of communication from the baby. However, there is often a lack of proper handling for new parents who have babies. From the results of the study, there are 5 languages used by the babies, namely NEH (Hungry), HEH (Discomfort), EAIRH (Lower Wind/Gas), EH (Burp Me), and OWH (Tired) [1]. Dunstan Baby Language (DBL) can classify the baby crying sound at the age of 0-3 months based on his condition. The results of Dunstan Baby Language (DBL) Explain that infants aged 2-6 months and do not depend on the type of race and language in this world crying with a uniform initial word on each condition, e.g. hungry, in the wind, belching, sleepy and uncomfortable [1].
Research on identifying the cry of this baby has been done by some researchers before. In the year 2018, research by Irma Amelia Dewi, Adriana Zulkarnain and Ayu Aprilia conducted a test of 25 sound sample sounds with 5 sounds of tears per condition and got a 76% accuracy result using Linear Predictive Coding (LPC) methods and Euclidean Distance [2]. In 2016, research by Welly Setiawan, Chastine Fatichah and Umi Laili tested as many as 139 sound samples with 5 cries of tears grouped into 4 classes and got 79.95% accuracy result using Mel-Frequency method Cepstrum Coefficient (MFCC) and K-Nearest Neighbor (KNN) [3]. In 2016, research by Anyawee Chaiwachiragompol and Nattawoot Suwannata tested and gained an accuracy of 98% using DWT and ELM methods [4]. In 2013, research by Medhanita Dewi tested as many as 175 sound samples with 140 training data and 35 test data and received a 94% accuracy result using codebook and MFCC [5]. In 2010, research by Rachmad Ariyadi, Mauridhi Hary, Nana Ramadijanti and Bima Sena tested as many as 39 sound samples and received an 89.7% accuracy result using the Neural Network [6] method. In 2014, the study by Krittakom Srijiranon and Narissara Eiamkanitchat identified the baby’s cry using the Neuro-Fuzzy and Perceptual Linear Prediction methods to recognize the cry of the baby [7]. The latest one, in 2019 Sita Purnama Dewi dkk has studied about baby crying analysis using MFCC and LFCC using different classification methods show that LFCC has better accuracy in woman speech recognition [8].

Mel Frequency Cepstral Coefficient (MFCC) is regarded as a feature vector and obtained by applying a Discrete Wavelet Transform (DWT) decomposition to the accumulated speech signal. The feature extraction technique as well as Principal Component Analysis (PCA) is regarded as a feature vector and obtained by applying DWT decomposition on the accumulated speech signal [9] [10]. DWT is not only used in 1D signal processing, but also in 2D signal processing and it has good performance in feature extraction [11]. Euclidean distance is used as a scoring classification algorithm to calculate the difference in distance between two matrices to measure the similarity of extraction value in tears. In this research, the performance analysis was conducted by calculating the accuracy and timing of computation.

2. Research Method

This research basically divided into 3 (three) things namely feature extraction using DWT, MFCC and PCA method. Classification using Euclidean Distance method.

2.1. Dunstan Baby Language

Crying is the primary way for a newborn to communicate. The communication is expected in order for both parents to understand what is happening in the baby. The cry will be a response to the need for hunger, pain, lack of a baby and change in temperature.

Dunstan Baby Language (DBL) is one of the techniques for the baby’s crying classification, which is about vocal reflexes associated with babies as a sound signal. The DBL states that there are five infant tears classifications throughout the culture and linguistic group, which are used by infants prior to language acquisition. The cry will be a response to the need for hunger, pain, lack of a baby and change in temperature. From the results of the study, there are 5 languages used by the babies NEH (Hungry), HEH (Discomfort), EAIRH (Lower Wind/Gas), EH (Burp), and OWH (Tired).

2.2. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) has characteristics of high time resolution, low-frequency resolution and low time resolution for low frequencies. DWT has the function to present the signal in several frequency components through signals on the filter and signal decomposition, which aims to obtain a signal composition at high and low frequencies [12]. This wavelet decomposition process misses the original audio signal through 2 filters namely High Pass Filter (HPF) and Low Pass Filter (LPF). The analysis was conducted on the results of the HPF to analyze high frequencies and the LPF was used for low frequencies [13]. The equation of the signal at level 1 at DWT is as follows:

$$x[n] = \sum_{k=-\infty}^{\infty} (y_{high}[k].g[-n+2k]) + (y_{low}[k].h[-n+2k])$$ (1)
2.3. **Daubechies Wavelet**

Daubechies's wavelet transformation is defined in the same way as the Haar wavelet transformation. By calculating the average and the difference that goes through a scalar product with a scaling signal and wavelets. The difference between this is how these scaling signals and wavelets are defined [13] [14].

2.4. **Mel Frequency Cepstral Coefficient**

Mel-Frequency Cepstral Coefficients (MFCC) is the most popular feature extraction technique used in voice recognition based on a frequency domain using the Mel scale based on the human ear scale. The MFCC is considered a much more accurate feature of the frequency domain than the time domain feature [15],[16],[17]. Before the MFCC process, then do the preprocessing process. Preprocessing is a sound signal process there is still a lot of noise in it or the initial data preparation process before it is extracted. Preprocessing includes 2 phases, namely:

- DC removal is the process of eliminating a DC signal or a unidirectional signal that will be a frequency signal that occurs alternating signal or AC. DC removal serves to calculate the average of the sampled sound and reduce each sample with value the average. The following formulas are used:

\[ y[n] = x[n] - \bar{x}, \quad 0 \leq n \leq N - 1 \]  

with \( y[n] \) is a DC Removal result signal sample, \( x [n] \) is the original signal sample, the \( \bar{x} \) is the average value of the original signal sample, and \( N \) is the signal length.

- Pre-Emphasis Filtering was conducted using the first order of Finite Impulse Response (FIR), High Pass Filter (HPF) which is based on the relationship between input and output in the time domain. The tool used can be a filter with a fixed or adaptive coefficient, where the efficiency is adjusted against the time depending on the auto-correlation value of the speech signal file used:

\[ y[n] = s[n] - a.s[n - 1], 0.9 \leq a \leq 1 \]  

with \( y[n] \) is a pre-emphasis signal, \( s [n] \) is the signal before pre-emphasis and the value \( a \) used is 0.97.

- Frame blocking is the process of cutting sound signals into parts so as to facilitate the calculation and analysis of sound. Voice signals are processed on a short-term basis. The frame systems used are 128, 256, 512, and 1024 data.

\[ Frame = \frac{M}{T_s} \]  

where \( M \) is duration (in ms) and \( T_s \) is number of frame.

- Windowing is done on each frame with the aim of minimizing the discontinuity between two frames which the start and end parts have constraints. The type of Hamming used is the Hamming window which has the equation of function:

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

with \( w(n) \) is the window function.

- Fast Fourier Transform serves to change the sound signal from the time domain to a frequency domain. That is, the sound is stored digitally in the form of sound spectrum waves and resulted in a wave detection of discrete-form domain \( A \).
with \( F(k) \) is an FFT result signal, \( n \) is the sample index, \( N \) is the number of samples, and \( K \) is 0, 1, 2, ..., \((N - 1)\).

- Mel-Frequency Wrapping is the process of changing the frequency scale to a mail scale. The Mel-Frequency scale is a linear frequency below 1 kHz and logarithmic above 1 kHz. The Mel scale is obtained with the equation:

\[
\text{Mel}(f) = \frac{2595 \times \log_{10}(1 + \frac{f}{700})}{s/2}
\]

with \( \text{Mel}(f) \) is the Mel-Frequency scale, and \( f \) is a linear frequency (Hz).

- Discrete Cosine Transform is the process to change the spectrum to Cepstrum. The result of the DCT poses is Mel-Frequency Cepstrum Coefficient (MFCC). The following formulas are used:

\[
C_n = \sum_{k=1}^{N} (\log S_k) \times \cos \left[ n \left( k - \frac{1}{2} \right) \frac{\pi}{N} \right], \quad n = 1, 2, ..., k
\]

with \( S_k \) is the output of the filterbank process at index \( K \), and \( K \) is the expected number of coefficients.

**Figure 1.** Sample of Signal Baby Cry
• Cepstral Lifting is used to obtain information from the spoken voice signals. Cepstral lifting calculation can be done by implementing the function windowing:

\[ w(n) = \left( 1 + \frac{L}{2} \sin \left( \frac{n\pi}{L} \right) \right) \]

with \( L \) is the cepstral amount of coefficients, and \( N \) is the index of cepstral coefficients.

Based on Figure 1, showing sample of signal baby cry in each class. With A is heh (Discomfort), B is Neh (Hungry), C is Eairh (Lower Wind/gas), D is eh (Burp), and E is Owh (Tired).

2.5. Principal Component Analysis
Principal Component Analysis (PCA) is a method used by reducing dimensions and for feature extraction. Finding a set of orthogonal vector bases that maximally captures the connectedness between dimensions in the original data is the goal of PCA [18]. Finding patterns on data and compressing data by reducing the number of dimensions without having to lose a lot of information is one of the advantages of PCA. The PCA algorithm is as follows [18], [19], [20]:

1. Count the mean of the MFCC feature, then the mean deductible from each value data

\[ \text{mean} = \frac{\sum_{i=1}^{n} x_i}{n} \]  

with \( x_i \) is feature value from MFCC and \( n \) is number of feature data.

2. Calculate the normalization of metrics by using equation as follows:

\[ A = T - \text{mean} \]

with \( T \) is threshold which is used.

3. Calculate the covariant matrix value with the following equation:

\[ L = A^T \times A \]

with \( A^T \) is transpose of matrix normalization.

4. Next, sort from large to small eigenvalue (D) and Eigenvector (V) based on eigenvalue sequence. Then, calculate the matrix value of eigenface with the equation as follows:

\[ \text{Eigenface} = A \times \text{Eigenvector} \]

5. Calculating the result of the sound as follows:

\[ \text{Sound} = \text{Eigenface}^T \times A \]

Sound value is got from transpose of eigenface matrix is multiplied with normalization matrix.

2.6. Euclidean Distance
Euclidean distance as a classification algorithm assessment to calculate the difference in distance between two matrices to measure the resemblance of the extraction value is defined as follows:

\[ D_{\text{min}}(ts, tr) = \sqrt{\sum_{i=1}^{n} (tr_i - ts_i)^2} \]

with \( tr_i \) is the training data, \( ts_i \) is test data, and \( D_{\text{min}} \) is the minimum distance.
3. System Models
In the design and implementation of this final task study, the system is divided into 4 main parts, namely the acquisition of data, preprocessing, extraction of features and classification. The stage was shown in Figure 2 of the system design.

In Figure 2, explains bring the system designed using the MATLAB software as a data processor. The initial stage of the system is data acquisition or crying babies. The data used is 150 training data and 50 test data. After the retrieval process of the baby’s cry, the voice of the Crying baby will enter the DWT extraction process to obtain low frequency (cA) signal and high-frequency signal (CD). The signal Data is then extracted again by the MFCC to obtain the MFCC (cA) and MFCC (CD) characteristic coefficients. Then the signal data is extracted back by the PCA to obtain a coefficient of features to be classified. The Data obtained is then classified using the Euclidean distance method.

![Figure 2. System Flowchart](image)

![Figure 3. Flowchart of the (a) Training Process and (b) Test Process](image)
In Figure 3 is a flowchart of the design flow of the baby cries system, test data, and training data will be extracted features with the Discrete Wavelet Transform (DWT), Mel-Frequency Cepstral Coefficient (MFCC) and Principal Component Analysis (PCA) methods. The result of the feature extraction will be classified with Euclidean distance to calculate the difference in distance between the two matrix sound traits.

3.1 Extraction Feature DWT

DWT feature extraction aims to determine the sound signal features by sharing to Subband highpass filter (HPF) and lowpass filter (LPF), the input signal is passed on the Highpass filter to then analyzed the high frequency, then the input signal is analyzed on the lowpass filter for low-frequency analysis. There are simple system models in Figure 4:

![Figure 4. DWT Feature Extraction Block Diagram](image)

The extraction process of the DWT feature is shown in Figure 4. In this process, the initial signal will be unraveled to the specified level. After that, the signal will be grouped into low signal and high signal. The end result will result in a normal encoding that corresponds to the decomposition level. Then the next process is preprocessing.

3.2 Pre Processing

Preprocessing is a sound signal process there is still a lot of noise in it or the initial data preparation process before it is extracted. There are simple system models in Figure 5:

![Figure 5. Pre-Processing Phase Block Diagram](image)

In Figure 5, this is a preprocessing process. This process includes 2 steps, namely DC Removal and Pre-emphasis. Then the next process is MFCC.

3.3 Extraction Feature MFCC

Mel-Frequency Cepstral Coefficients (MFCC) is the most popular feature extraction technique used in voice recognition based on a frequency domain using the Mel scale based on the human ear scale. MFCC is considered a much more accurate feature of the frequency domain than the time domain features. There are simple system models in Figure 6:

![Figure 6. MFCC Feature Extraction Block Diagram](image)

In Figure 6, is a block of extraction diagram of MFCC feature. After the preprocessing process, the signal will be done by the MFCC process. Then the next process is PCA.

3.4 Principal Component Analysis
To analyze data, Principal Component Analysis is an excellent method. Finding a set of orthogonal vector bases that maximally captures the connectedness between dimensions on the original data is the goal of the PCA. Finding patterns on data and compressing data by reducing the number of dimensions without having to lose a lot of information is one of the advantages of PCA. There are simple system models in Figure 7:

![Figure 7. PCA Process](image)

Figure 7 shows a flowchart on the PCA process. After the extraction by MFCC, the next process will be feature extractions by PCA.

3.5 *Euclidean Distance*

The next stage is a classification process aimed at grouping by classification. The vector characteristics that have been acquired through the process will be classified according to their baby class crying. This class is hungry, discomfort, lower wind/gas, burp, and tired.

After creating the training database, the next step is to test the test data taken from the folder defined as test data. The Euclidean Distance classification process is done by testing the test data that will be compared with training sound of baby cry data that has been formed in the training database.

4. Results and Discussion

Testing using the best parameters, namely parameters with the number of MFCC 32 coefficients, DWT level 1, DB 2, frame number 1024 with 150 train data and 50 test data.

4.1 *Effect of Number Frames*

The signal will be segmented into the frame and the number of tested frames are 128, 256, 512, and 1024. The parameters used are 150 training data, DB2, feature coefficient 32, Euclidean threshold 1.9395 and frequency sampling 44100Hz. From the test, the results obtained are as shown in Figure 8.

Based on Figure 8, obtained the best accuracy result of 100% with time 0.0971 seconds. The more the number of frames then the training will be more detailed as well as and the fewer number of frames tested then the computing time will be longer because the data testing is getting more detailed.

4.2 *Effect of the number MFCC coefficients*

The number of MFCC feature coefficients that will be tested on the system are 8, 16, 24, 32, 40 and 48. The parameters used are 150 training data, DB2, 1024 frames, Euclidean threshold 1.9395 and frequency sampling 44100Hz. From the test, the results obtained are as shown in Figure 9.

Based on Figure 9, obtained the best accuracy result of 100% with time 0.0971 seconds. The more the coefficient of the feature and the more features the system tests, the less the coefficient of the computation, the longer the compute time during testing.
4.3 Effect of DWT decomposition levels

DWT decomposition level that will be tested on the system is level 1, and level 2. The parameters used are 150 training data, DB2, 1024 frames, coefficient feature 32, Euclidean threshold 1.9395 and frequency sampling 44100Hz. From the test, the results obtained are as shown in Figure 10.

Based on Figure 10, obtained using DWT Level 1 obtained the best accuracy result of 100% with time 0.0971 seconds. As well as acquired at DWT Level 2 obtain the best accuracy results with 93.33% accuracy and calculating the time 0.0698 seconds. The system could only get to the DWT level at one level. This is because of the increased level of DWT accuracy and computational time is decreasing and does not meet to proceed to the next level. Given the training data and test data using unclear sound samples, the duration of the voting time ranges from a range of 1 second to 30 seconds.

4.4 Effect of Daubechies Wavelet levels

Daubechies Wavelet to be tested on systems is DB1, DB2, DB4, DB6, DB8, and DB10. The parameters used are 150 training data, 1024 frames, coefficient feature 32, Euclidean threshold 1.9395 and frequency sampling 44100Hz. From the test, the results obtained are as shown in Figure 11.
Based on Figure 11, obtained the best accuracy result of 100% with computing time of 0.0971 seconds. DB's effect on testing is that smaller DB is the more detailed the data is researched and the computation time is getting faster.

4.5 Effect of training and testing data

Testing the impact of the amount of training data and test data on the accuracy and timing of system computing. Testing with variations of the amount of training and test data with the best parameter of 1024 frames, features of coefficient 32, DB2, DWT level 1, Euclidean 1.9450 and 44100Hz frequency sampling. From the test, the results obtained are as shown in Table 1.

| Table 1. Total Train Data and Data Test |  |
|----------------------------|----------------|
| **Train Data** | **Test Data** | **Accuracy (%)** | **Computing Time (s)** |
| 100 | 25 | 68% | 0.2608 |
| 100 | 50 | 74% | 0.1241 |
| 100 | 75 | 69.33% | 0.3033 |
| 125 | 25 | 84% | 0.2923 |
| 125 | 50 | 86% | 0.1217 |
| 125 | 75 | 78.67% | 0.5257 |
| 150 | 25 | 88% | 0.1465 |
| 150 | 50 | 90% | 0.1647 |
| 150 | 75 | 80% | 0.5646 |

Based on Table 1, the results were passed best accuracy with 150 training data on sound samples and test data 50 sound samples with 90% accuracy and 0.1647 computing time. Figure 12 shows the class accuracy results.
4.6 Effect of noise

Testing on the effect of dB noise influence and noise type on compute accuracy testing and system time. Noise will be provided to voice signals with large-0dB, -10dB, -20dB, -30dB, and -40dB for impact on motor, music, and sound noise. The parameter used are 1024 frames, features of coefficient 32, dB2, DWT level 1, Euclidean 1.9450 and 44100Hz frequency sampling. The results obtained are as shown in Figure 13.

![Figure 13. Effect of dB type Noise](image)

Based on Figure 13, obtained the test results when the sound was given to-80dB noise with 52.23% accuracy and compute the time of 0,1620 seconds. It explains that the system can identify noise below-80dB. From the test, the results obtained are as shown in Figure 14.

![Figure 14. Effect of Noise-10dB on Noise type](image)

Based on Figure 14, obtained test results when noise-10dB with sound type motor with 60% accuracy and compute time 0.2434 seconds. This explains that noise affects system accuracy. The greater the noise level, the greater the accuracy increase.

5. Conclusion

In this study, it has been designed a system to identify the baby’s cry using the method DWT, MFCC, and PCA with the algorithm Euclidean distance. The system uses 150 training data and the number of test data is 50. The system can identify the baby's cry on 5 conditions of a baby crying Neh (Hungry), Heh (Discomfort), Eairh (Lower Wind/Gas), Eh (Burp), and Owh (Tired). The best Parameter is obtained at a frame size of 1024 data per frame, MFCC feature coefficient of 32, DWT at level 1, DB 2, and Euclidean distance value is 1.9450. The system can detect the baby's crying sound with the best accuracy of 90% and computational time 0.5542 seconds. For the analysis, we can conclude that the more number of frames used, the longer computation time. The more number of coefficients is used, the more features of the baby's crying voice then it takes longer computation time. When DWT level is higher than the accuracy gets lower. Then DWT Level 1 is used as the best parameter. Number of training data and testing data affect the system accuracy. When the baby sounds have much noise, the best parameter is below-80dB and it gives 60% accuracy.
6. References

[1] P. Dunstan, “Open Up and Discover Your Baby’s Language,” 2012. [Online]. Available: www.babytaal.nl/media/PDF/ComprehensiveBooklet(2).pdf. [Accessed: 10-Sep-2018].

[2] I. A. Dewi, A. Zulkarnain, and A. A. Lestari, “Identifikasi Suara Tangisan Bayi menggunakan Metode LPC dan Euclidean Distance,” *Elkomika*, vol. 6, no. 1, pp. 153–164, 2018.

[3] W. S. Limantoro, 2016, C. Fatichah, and L. Yuhana, “Rancang Bangun Aplikasi Pendeteksi Suara Tangison Bayi,” Institut Teknologi Sepuluh November.

[4] A. Chaiwachiragompol and N. Suwannata, 2016, “The Features Extraction of Infants Cries by Using Discrete Wavelet Transform Techniques,” *Procedia Comput. Sci.*, vol. 86, no. March, pp. 285–288.

[5] M. D. Renanti, A. Buone, and Wi. A. Kusuma, 2013, “Infant Cries Identification By Using Codebook As Feature Matching And MFCC As Feature Extraction,” *JATIT*, vol. 56, no. 2.

[6] N. Ramadjianti, B. Sena, and B. Dewantara, 2016, “Pengenalan Rasa Lapar Melalui Suara Tangis Bayi Umur 0-9 Bulan Dengan Menggunakan Neural Network,” Politeknik Elektronika Negeri Surabaya.

[7] K. Sririjanon and N. Eiamkanitchat, 2014, “Application of Neuro-fuzzy Approaches To Recognition And Classification of Infant Cry,” *IEEE*.

[8] Dewi, S. P., Prasasti, A. L., & Irawan, B. (2019). The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods. *The 2019 IEEE International Conference on Signals and Systems (ICSigSys)* (pp. 19-24). Bandung: IEEE.

[9] M.A.Anusuya and S.K.Katti, June, 2014, “Mel Frequency Discrete Wavelet Coefficients for Kannada Speech Recognition using PCA,” *ACEEE*.

[10] G. C. N. Neha Baranwal, Ganesh Jaiswal, 2014, “A Speech Recognition Technique Using MFCC with DWT In Isolated Hindi Words,” Indian Institute of Information Technology, Allahabad.

[11] Kurniawan, A., Purboyo, T. W., & Prasasti, A. L. (2017). Implementation of Image Compression Using Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). *International Journal of Applied Engineering Research*, 12 (23), 13951-13958.

[12] Ha. Ali, N. Ahmad, X. Zhou, K. Iqbal, and S. M. Ali, 2012, “DWT Features Performance Analysis for Automatic Speech Recognition of Urdu,” *Springer Plus*, vol. 3, no. February, p. 1.

[13] M. Murugappan, N. Q. Idayu Baharuddin, and J. S, 2012, “DWT and MFCC Based Human Emotional Speech Classification Using LDA,” University Malaysia Perlis.

[14] Y. Meyer, “Daubechies wavelets,” 1999. [Online]. Available: http://dsp-book.narod.ru/PWSA/8276_02.pdf. [Accessed: 24-Sep-2018].

[15] I. S. Permua and Y. I. Nurhasanah, 2018, “Implementasi Metode MFCC Dan DWT Untuk Pengenalan Jenis Suara Pria Dan Wanita,” *MINd*, vol. 3, no. 1, pp. 49–63.

[16] J. Givary, R.P. Astuti, L. Novamizanti, 2017, Analysis effect of discrete wavelet transform in multi carrier code division multiple access, IEEE Asia Pacific Conference on Wireless and Mobile 2016 (APWiMob).

[17] S. Anggora Wicaksono, Sukmawati NE, Satriyo Adhy, 2014, “Aplikasi Speech Recognition Bahasa Indonesia Dengan Metode MFCC Dan LVQ Untuk Pengendalian Gerak Robot,” Universitas Diponegoro.

[18] Z. Ge, S. R. Sharma, and M. J. T. Smith, 2016, “PCA/LDA Approach for Text-Independent Speaker Recognition,” Purdue.

[19] L. I. Smith, “A Tutorial on Principal Components Analysis Introduction,” 2002. [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf. [Accessed: 24-Sep-2018].

[20] R. D. Putra, A. L. Prasasti, and T. Waluyo, 2018, “Analysis Of Retinex Algorithm On Digital Image From CCTV Camera For Face Recognition,” *JATIT*, vol. 96, no. 23, pp. 7942–7962.