Study of Linear Discriminant Analysis to Identify Baby Cry Based on DWT and MFCC

Ledya Novamizanti¹², Anggunmeka Luhur Prasasti¹, Bangun Satria Utama¹
¹School of Electrical Engineering, Telkom University
Jl. Telekomunikasi Terusan Buah Batu Bandung 40257 Indonesia
²ledyaldn@telkomuniversity.ac.id

Abstract. Baby crying is a common behavior among babies and is a means of verbal communication for babies who express their needs and desires. A baby's cry identification system is needed because it makes it easy for adults to find out the meaning of a baby's cry. This study proposes a system for classifying infant crying sounds using Linear Discriminant Analysis (LDA) with Discrete Wavelet Transform (DWT) and Mel-frequency Cepstral Coefficient (MFCC) as a feature extraction method. Based on experiments, the system can identify the sound of crying babies grouped into 5 (five) classes, namely discomfort, hunger, colds, burp, and drowsiness. The system achieves an accuracy of 94% and an average computing time of 1.5506 seconds. The performance of Linear Discriminant Analysis (LDA) outperformed Principal Component Analysis (LDA) in the identification of crying babies.

1. Introduction

In newborns, communication is done through crying, movement or certain cues. Infants over the age of 9 months will develop the ability to mimic the sounds of adults [1]. After the baby is born, the baby cry has a function of responses such as changes in temperature, pain, hunger, or discomfort [2]. The research carried out by Dunstan Baby Language (DBL) explains that the sound of a baby cry has meanings. For example, "HEH" means feeling uncomfortable, usually occurs when a baby wants her diaper or clothes to be replaced because she feels hot and dirty. Other examples, there are "NEH" that means hungry, "EAIRH" that means the baby is sick and depressed in the abdomen or caught cold, "EH" that means to belch, and "OWH" that means tired and sleepy [3].

Related research conducted by [1], identified baby cry using MFCC and the Linear Predictive Coding classification. The number of baby cry data is 25 sound samples through which there are 5 classes of data identified such as uncomfortableness, hunger, cold, belching and drowsiness. It reported achieving an accuracy of 76%. Welly et al. [4] using MFCC and K-Nearest Neighbor classification. The number of their baby cry data is 139 samples through which there are 4 classes of data identified such as 'Eairh', 'Heh', 'Neh', 'Owh'. They reported yielding an accuracy of 96.57%. Another study is also carried out by Bhagatpatil and Sardar [5] using LFCC and VQ codebook classification. The number of baby crying data is 120 samples. The system is divided into 5 classes: discomfort, hunger, cold, belching and drowsiness, and reach an accuracy of 92.45%. Bhagatpatil also tested MFCC performance and the code book method. Their system yielded an accuracy of 83.75%. Meanwhile, [6] used MFCC and Reservoir Network (RN). The number of their baby cry data is 132 samples through which there are 5 classes of data identified such as uncomfortableness, hunger, cold, belching and drowsiness, with an accuracy of 94.5%. Another study was reported by [7]. They use DWT and extreme as their method.
The MFCC method is a feature extraction method using a mel calculation that can equate with the human voice. Hence, MFCC will easily recognize the sound of crying babies by extracting traits through the mel scale [6-8]. Meanwhile, DWT is a feature extraction method of decomposing a signal that separates between a low signal and a high signal. This is very useful and makes it efficient when a signal can be separated according to its frequency and can distinguish the desired frequency signal [9-10]. The LDA is a classification method to discriminate data according to linear analysis. LDA maximizes the covariance matrix between classes and minimizes the covariance matrix with classes. More class members spread and ultimately, can increase the success of the introduction [11].

2. Methodology

2.1. Discrete Wavelet Transform (DWT)
DWT is a signal decomposition at a frequency, and a sub-band wavelet component is produced by decreasing the decomposition level. Implementation of DWT can be done by passing the signal into two DWT filters, i.e. Low Pass Filter (LPF) and High Pass Filter (HPF) [10]. DWT serves as pre-processing aim for decomposing the sound signals so that the signals can conform to required specifications. The general formula of LDA is represented by the following equation [12-14],

\[ x[n] = \sum_{k} (y_{\text{high}}[k] \cdot g[-n + 2k]) + (y_{\text{low}}[k] \cdot h[-n + 2k]) \]  

(1)

2.2. MFCC
MFCC is a signal feature extraction method with a Mel frequency scale which is known to approach the human hearing response of 20-20,000 Hz [6]. MFCC adheres to the workings of the human ear where the human ear is a linear filter at low frequencies and works logarithmically at high frequencies. Before entering the MFCC process, a pre-processing process will be carried out aiming for incoming sound signals by the required specifications. The pre-processing process is DC removal and pre-emphasize filtering [15-16].

2.3. Linear Discriminant Analysis (LDA)
LDA is a method with a combination of arithmetic operations and statistics that perform separate statistical tests for each object. The general formula of LDA is represented as follows [17],

\[ SW = \frac{1}{2} \sum_{s=1}^{S} \frac{1}{T_s} \sum_{x=1}^{T_s} (X_s + \bar{\Theta})(X_s + \bar{\Theta})^T \]  

(2)

The SW covariance matrix is minimized, and SB is maximized, then eigenvalue and equation are searched. The formula is defined as follows.

\[ SB = SW \cdot V \]  

(3)

\[ cov = SB \cdot SW \]  

(4)

3. Proposed Method
This system is divided into three stages of the process, namely the acquisition of baby crying sound, feature extraction with DWT and MFCC, and classification with LDA. The system identifies the sound of crying babies into 5 classes, namely discomfort, hunger, colds, belching, and drowsiness. Figure 1 shows the block diagram of system design—total records 200 sound of crying babies, consisting of 150 training data and test data 50 [12]. Figure 2 is an example of a baby crying signal with five classes.
3.1. Pre Processing
Pre-processing is conducted so that the incoming sound matches the required specifications. DC removal is done by reducing the value of each sampled sound with the average value of the data. The function of pre-emphasis filtering is to balance or stabilize the signal spectrum. Figure 3 shows the block diagram of Pre-Processing.

3.2. Feature Extraction
The filtering process is done until it finds the desired level. Feature extraction with Mel-frequency Cepstral Coefficient (MFCC) is performed to find out the special traits that are in the sound signal taken.
3.3. Classification
The classification process aims to classify data based on class. Feature vectors that have been obtained through the feature extraction process are classified using LDA according to their baby’s cry identification class. The type of class, which is uncomfortable, hungry, catching cold, burping and sleepy. Figure 5 displays the flowchart of classification using LDA.

4. Result and Discussion
4.1. MFCC coefficients and DWT level decomposition
This scenario evaluates the effect of changing the number of MFCC coefficients and the DWT level decomposition on accuracy. The accuracy parameter is calculated based on the amount of data classified correctly divided by the total data. Figure 6 displays the effect of the number of coefficients and the level of DWT decomposition on accuracy. The system shows 100% accuracy when the MFCC coefficient value is 20, and the DWT Level 1. The more MFCC coefficient numbers are tested, the more features are produced for testing by the system. The MFCC coefficient is greater the degree of impact of decomposition on system accuracy. This occurs because an indication that the level of additional decomposition causes information to appear more detailed. More information can contribute
to better model performance, but we suspect it can reduce system stability. In the end, we decided that the baby cry detection system used level 1 DWT.

![Figure 6](image)

**Figure 6.** The effect of MFCC coefficient and DWT level on accuracy

### 4.2. Frame Numbers

This scenario observes the effect of the number of frames on system accuracy. Figure 7 displays the effect of the number of frames on system accuracy.

![Figure 7](image)

**Figure 7.** Effect of frame number on accuracy

The best results are achieved when the number of frames is 256, with 100% system accuracy. Signal cuts per frame into smaller parts, making the signal diameter more stable. It was observed that the number of frames tested resulted in fluctuating system accuracy. The results of our system show the number of frames 256 to be the best barometer.

### 5. Conclusion

In this study, the proposed baby cry classification system uses the DWT, MFCC and LDA methods. The baby's crying is separated into 5 classes, namely discomfort, hungry, catching a cold, burp and drowsiness. This system has an accuracy of 94% and a computing time of around 1.5506 seconds. The
best parameters obtained during training are DWT level 1, DB 8, C 20, and the number of frames is 256. The more coefficient numbers are used, the more features are produced from the crying baby. When compared with previous studies, with the same feature method (MFCC) but different classification methods using PCA [12], this LDA method has better system accuracy. When using LFCC feature extraction, and the KNN classification method [18], which only reaches 90% accuracy, this system still has better accuracy.

6. References
[1] Dewi I A, Adriana Z and Lestari A A 2018 Identifikasi Suara Tangisan Bayi menggunakan Metode LPC dan Euclidean Distance Jurnal ELKOMIKA 6 153
[2] Brazelton T B 1962 Crying in infancy Pediatrics 29 579.
[3] P. Dunstan, “Open Up and Discover Your Baby’s Language,” 2012. [Online]. Available: www.babytaal.nl/media/PDF/ComprehensiveBooklet(2).pdf. [Accessed: 24-July-2020]
[4] Welly S L, Chastine F, Umi Y L 2016 Rancang Bangun Aplikasi Pendeteksi Suara Tangisan Bayi Thesis (Surabaya: Institut Teknologi Sepuluh Nopember)
[5] Bhagatpatil V V, Sardar V M 2015, An Automatic Infant’s Cry Detection Using Linear Frequency Cepstrum Coefficients (LFCC) International Journal of Scientific & Engineering Research 3 29
[6] Renanyi M D, Buono A, Kusuma W A 2013 Infant Cries Identification By Using Codebook As Feature Matching and MFCC as Feature Extraction Thesis (Bogor: Institut Pertanian Bogor)
[7] Putra D K, Tritosamoro I I and Atmaja R D 2017 Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (MFCC) Dan Gaussian Mixture Model (GMM) e-Proceeding of Engineering 4 1766
[8] Ikhsan I, Novamizanti L, Ramatryana I N A 2014 Automatic musical genre classification of audio using Hidden Markov Model 2nd International Conference on Information and Communication Technology (Indonesia: IEEE)
[9] Stavros N 2015 Audio Pattern Recognition of Baby Crying Sound Events Journal of the Audio Engineering Society 63 358
[10] Chaiwachiragompol A, Suwanna N 2016 The Features Extraction of Infants Cries by Using Discrete Wavelet Transform Techniques Procedia Computer Science 86 285
[11] Borade S N, Deshmukh R R and Ramu S 2016 Face recognition using fusion of PCA and LDA: Borda count approach Proceedings of the 2016 24th Mediterranean Conference on Control and Automation (Greece: IEEE) 1164
[12] Prasasti A L, Novamizanti L, Razik M I 2019 Identification of baby cry with Discrete Wavelet Transform, Mel Frequency Cepstral Coefficient and Principal Component Analysis Journal of Physics: Conference Series 1367 012061
[13] Duzenli T and Ozkurt N 2011 Comparison OF Wavelet Based Feature Extraction Methods for Speech/Music Discrimination National Conference on Electrical, Electronics and Computer Engineering (Turkey: IEEE)
[14] Arabjamaloei R, et al. 2011 Exact Lithologic Boundary Detection Based on Wavelet Transform Analysis and Real-Time Investigation of Facies Discontinuities Using Drilling Data Petroleum Science and Technology 29 569
[15] Tiwari V 2010 MFCC and its applications in speaker recognition International Journal on Emerging Technologies 1 19
[16] Permama I S, Indrawaty Y and Zulkarnain A 2018 Implementasi metode MFCC dan DTW untuk jenis suara pria dan wanita Mind Journal 4 49
[17] Wang, Y. and Y. Chen, 2017. A new feature extraction algorithm based on fisher linear discriminant analysis. Proceedings of the 2017 3rd International Conference on Control, Automation and Robotics (Japan: IEEE) 414
[18] Dewi S P, Prasasti A L, Irawan B 2019 Analysis of LFCC Feature Extraction in Baby Crying Classification using KNN International Conference on Internet of Things and Intelligence System (IoTais) (Indonesia: IEEE)