

Hidden Information Detection in Speech Signal Using MFCC and LDA

Alifia Putri Anantha*, Philip Tobianto Daely[†], Jae Min Lee[‡], and Dong-Seong Kim[§]

^{*†‡§}Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea

[†]Department of Information Technology, Institut Teknologi Telkom Surabaya, Surabaya, Indonesia

*ananthafia@kumoh.ac.kr, [†]philip.daely@kumoh.ac.kr, [‡]ljmpaul@kumoh.ac.kr, [§]dskim@kumoh.ac.kr

Abstract—This paper proposes a system that can detect an existence of hidden information on speech files by analyzing the values or characteristics of the speech signal. In order to know the characteristics itself, Mel-Frequency Cepstral Coefficient (MFCC) method is used to extract the features of the speech signals and Linear Discriminant Analysis (LDA) method is used to do the features selection. For the classification process, Support Vector Machine (SVM) is used. The final output of this system is a condition that states the speech signal is original, contains hidden message, or contains noise.

Index Terms—Hidden information detection, LDA, MFCC, speech signal, steganalysis, steganography

I. INTRODUCTION

Nowadays an exchange of information is very easy and fast. Behind the easiness of that data exchange, people can insert or hide secret messages into a media so that the existence of these messages could not be known by many parties. This method is called steganography [1]. The problem faced is that the secret message inserted is not always good. There are also secret messages that are bad (damaging) which usually contains messages with the intention of crime. Misuse of steganography techniques makes it difficult for civil or military authorities to find out messages containing illegal substances.

In its implementation, one of the media that most often found in the community is speech signals such as voice notes. In addition, the difference between the original speech signal and the one that has been inserted secret message is not easily known. Therefore, the possibility of confidential messages distribution through this speech signal audio is high. To increase security and prevent any dangerous messages, an effort to identify the existence of secret messages hidden in the media is required. The technique is called steganalysis [2].

Based on those problems, this paper proposes a steganalysis system to detect speech files with .wav format using Mel-Frequency Cepstral Coefficient (MFCC) and Linear Discriminant Analysis (LDA) methods. As a complement, Support Vector Machine (SVM) is used in the classification process. MFCC and LDA is used to find out the value distribution of characteristics of the speech signal detected. By knowing these values, the differences between original speech signal file and the one which inserted hidden information, or even just inserted noise can be known.

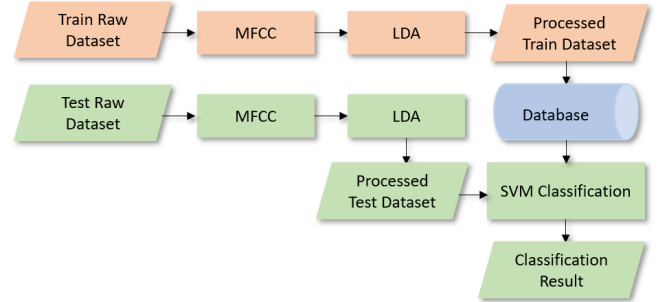


Fig. 1: Diagram block of system model design.

II. SYSTEM MODEL

The steganalysis system designed in this study consisted of two stages, namely training and testing stages. The system model is shown in Fig. 1. The explanation of each stage in the system configuration can be described as follows:

- 1) The speech signal file is extracted using MFCC, where FFT transformation takes place which functions as a signal converter from the time domain to the frequency domain. After that, an inverse FFT transformation takes place to convert the signal from the frequency domain to the time domain again.
- 2) The output from MFCC then through the feature selection process using LDA method. The goal is to optimize discriminant values and speed up the testing process.
- 3) After features extraction and selection, classification process using SVM method is carried out in order to maximally separate several sets of data between original class, noise, and stego.
- 4) If the classification results have been obtained, then a decision can be made and the steganalysis process is complete.

III. PROBLEM FORMULATION

MFCC is a reflection of the actual cepstral of the short-term signal which has been windowed, generated from Fast Fourier Transform (FFT) of the signal. FFT is used on a windowing signal to transform it into frequency domain. The Discrete Fourier Transform (DFT) over a discrete signal $x(n)$ of N samples, transforms each frame of N samples into the

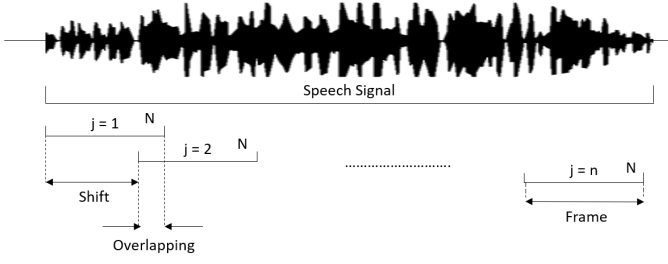


Fig. 2: Frame blocking of MFCC.

frequency domain. The FFT is the fast algorithm to enforce DFT, described as

$$X_k = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, k = 0, 1, 2, 3, \dots, N-1. \quad (1)$$

The outcome is often referred to as spectrum after this phase.

The steps for conducting MFCC are frame blocking and Mel-frequency wrapping. Frame blocking is the process of cutting a sound signal into several frame [3]. Fig. 2 shows the audio signal that is divided into several frames and overlap with each other. Mel-Frequency wrapping is usually done using a bank filter. The correlation between speech frequency and Mel-scale can be defined as equation below [4].

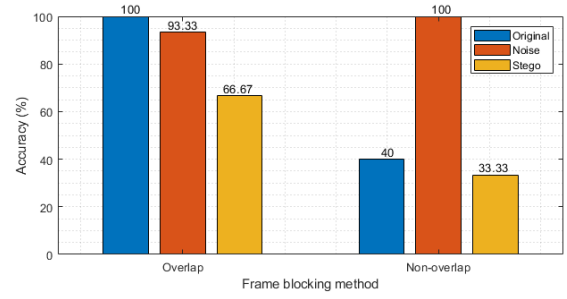
$$mel(f) = 2595 * \log \frac{1 + f(Hz)}{700}, \quad (2)$$

where $mel(f)$ is the Mel frequency scale and f is the actual frequency. MFCC utilizes Mel-scale filter banks, where higher frequency filters provide more bandwidth than lower filter frequencies.

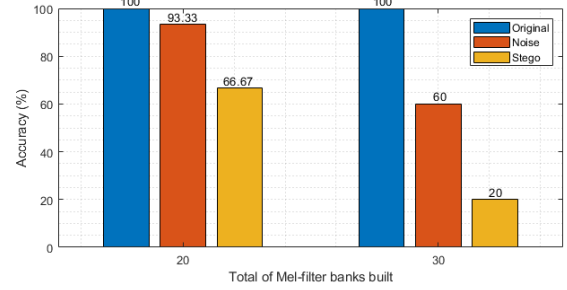
IV. PERFORMANCE EVALUATION

The total number of speech dataset used in this test are 35; 5 original data, 15 data which have been inserted noise, and 15 data that have been inserted hidden messages using steganography technique. SVM is done through two stages, first classifying audio test into original or non-original, and second classifying audio test which is classified as non-original into stego or noise. The test is carried out to determine accuracy based on the presence or absence of overlap at the frame blocking stage and accuracy based on the number of Mel-scale filter banks that want to be built at the Mel-frequency wrapping stage.

From Fig. 3a, it can be seen the effect of overlapping in the frame blocking process on system accuracy. In audio tests, steganalysis using the overlap process has a higher total accuracy which is 82.86% than those that do not use the overlap process which only achieves 62.86%. This proves that overlapping in samples can reduce the risk of loss of traits. Based on theory, the number of Mel-filters bank that are usually used is in the range of 20-40. From Fig. 3b we can see that the total accuracy values are 82.86% for 20 Mel-filter banks and 48.57% for 30 Mel-filter banks. It concludes that the greater number of Mel-filter banks were built on the Mel-frequency wrapping stage, the lower the accuracy obtained.



(a) Effect of overlapping processes on accuracy



(b) Effect of number of bank filters on accuracy

Fig. 3: Simulation results.

V. CONCLUSION

This paper proposes steganalysis using MFCC and LDA on speech signal. Based on the results, overlap the signal at frame blocking stage can result a better accuracy because it minimizes the risk of characteristics loss when cutting audio. From the overall test, the best system performance was obtained when the audio frames are overlapping with each other and the number of Mel-filters bank built are 20. For future work, we will consider to compare our proposed approach with the other existing schemes and try other classification methods besides SVM.

ACKNOWLEDGEMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2020-2020-0-01612) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation)

REFERENCES

- [1] N. R. K. Kesa, "Steganography a data hiding technique," 2018.
- [2] K. Karampidis, E. Kavallieratou, and G. Papadourakis, "A review of image steganalysis techniques for digital forensics," *Journal of information security and applications*, vol. 40, pp. 217–235, 2018.
- [3] M. Likitha, S. R. R. Gupta, K. Hasitha, and A. U. Raju, "Speech based human emotion recognition using mfcc," in *2017 international conference on wireless communications, signal processing and networking (WiSP-NET)*. IEEE, 2017, pp. 2257–2260.
- [4] A. Nosan and S. Sitjongsatoporn, "Speech recognition approach using descend-delta-mean and mfcc algorithm," in *2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. IEEE, 2019, pp. 381–384.