Received: November 6, 2018

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Reversible Data Hiding in Audio Based on Discrete Cosine Transform and Location Maps

Mohammed Hatem Ali Al-Hooti[1, 2]* **Tohari Ahmad[1]**

[1]*Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[2]*Faculty of Computer & Information Technology, Sana'a University, Sana'a, Yemen*
* Corresponding author's Email: moh_hat84@yahoo.com

Abstract: Transmitting information via public networks is prone to illegal attacks. Thus, data hiding is a suitable aspect that is useful to assure information security. Nowadays, multimedia tempering becomes the most variety problem that occurs on public networks. Audio steganography methods aim to hide secret data in an audio file called cover. Many methods in the time domain can hide large data (payload) with relatively low robustness. On the other hand, transformed based frequency algorithms can be more robust but the recovery of the cover audio is not guaranteed. In this paper, we investigate the time and frequency domains and tend to bridge gaps in the area of data hiding with the frequency transformation methods. In addition, we propose a new reversible audio data hiding based on Discrete Cosine Transform (DCT) and location maps. This kind of map helps to fully recover both the secret message and the cover file from the stego where the frequency sample values are increased or decreased by 10 in correspondence with the intended bit to be embedded. The experimental results prove that this method is able to hide high payload, and it is also robust since it guarantees 100% the reversibility of both the secret message and the cover audio file. The quality measured by Signal-to-noise ratio (SNR) is approximately 75 dB, Normalized Correlation (NC) value is 1.000, and Bit Error Rate (BER) is around 0.006. Additionally, this method is able to maintain the quality and invisibility, which may not be achieved by some previous research.

Keywords: Robust data protection, Audio, DCT, Reversible data hiding, Information security.

1. Introduction

Technology facilities have rapidly grown. This fast development assists in transferring all the information through any public network. Correspondingly, these transmitted data need to be secured in order to stand against any unauthorized manipulation. This issue may be solved by using certain information security techniques. Currently, data hiding topic implies as the most well-known and adopted solution that is used for securing the transmitted information [1]. The solution happens by hiding the data known as a secret message within another file called cover, which can be either a grayscale or color image, compressed or uncompressed audio, video, or text file [2-4].

Audio data hiding is extensively used due to the availability and reputation of the transmitted varied audio files e.g. sound-messages, songs, and dialogues. According to the facts that attackers commonly examine digital images, and the audio files may not be as easy as images to be analyzed, audio steganography is appropriate to protect many practical covert communications such as telephone conversations. It is supportive for many fields where almost all transmitted information is highly required to be secured, for example, banking and medical reporting. So, transferring medical reports including patients' information are commonly hidden within an audio file [5]. Banking data require audio data hiding because the secrecy of financial data needs to be highly maintained.

There are many measurement principles that are used to evaluate and ensure the performance of any audio data hiding algorithm such as capacity, imperceptibility, invisibility, complexity and

robustness. Each criterion may have its effect on the other parameters. For example, the better achievement in terms of capacity, the worse the performance of the imperceptibility whereas many researchers look for the balance which satisfies performance of some or all those parameters [6, 7].

Development of audio data hiding algorithms is based on two types of methods: reversible and irreversible. Reversible algorithms ensure the recovery of both the secret messages and the cover file from the stego file [8]. The state of the art of audio data hiding are mainly classified into two main groups based on their domains, namely time and frequency [7]. The time domain methods are capable to hide much payload. However, they are fragile and rarely consider recovering 100% of the cover audio file. On the other hand, frequency schemes produce robust audio data hiding methods but the capacity is not comparable to the time domain algorithms. Moreover, most frequency domain schemes do not pay attention to the recovery of the exact cover audio file.

In this paper, we propose a new method based on Discrete Cosine Transform (DCT) and location maps. The strength of this work over the other existing ones is summarized as follows:

- This work mainly focuses entirely on retrieving both the secret message and the cover audio file from the stego file. It is based on the original frequency sample values which are saved in location maps. This file is used as a key that can be shared between the sender and receiver.
- It is designed to stand against attacks such as re-sampling, re-quantization or MP3 conversion since the embedding occurs in the frequency domain.
- It is difficult to compromise the secret message since the extraction process is not possible to perform without using the location map file.
- This work is able to maintain the quality and invisibility compared to the work in [5]. This is done by modifying the frequency sample values based on the intended bit to be embedded. Thus, if the secret bit is 0 the sample value is added by the threshold value (-10), otherwise it is added by (+10).
- As a result, this work guarantees 100 % recovery of both the secret message and the original cover audio from the stego file. This full recovery is due to the use of the location map file.

The outlines of this paper are as follows. Basic theory and some state-of-the-art works are described in Section 2. Section 3 presents the proposed method. The performance of this work is analyzed in Section 4. Finally, the conclusions and future works are drawn in Section 5.

2. Basic theory

Based on the audio processing domain, the most commonly used audio steganography algorithms are classified into two groups as follows.

2.1 Audio data hiding based on time domain

This type of methods concerns to extract the audio samples and directly embeds the secret information into the sample values. The stego file is directly built. For example, authors of LSB methods [3, 8, 9] extract the quantized samples in binary and replace the least significant bits of each sample by a secret bit. These works, including our previous papers [2, 7], are well-known in terms of the simplicity of extracting the secret message.

The method presented in [8] reads the quantized samples based 16 bits in a decimal form and modifies each sample in order to make its remainder value to be similar to its corresponding secret bit. A location map is used as a guide to recover both the secret message and the cover file. Here, the embedding and the recovery operations are in the time domain.

2.2 Audio steganography based on frequency domain

The audio file is transformed into the frequency domain. This transformation happens by applying one of the most commonly used transformation algorithms such as Fast Fourier Transforms (FFT), Discrete Wavelet Transform (DWT), DCT [9, 10]. For instance, the authors in [1] proposed a method that uses Singular Value Decomposition (SVD) and DCT. This is done based on a synchronization procedure. Furthermore, this work achieves relatively good performance, concerning robustness and imperceptibility; while the reversibility of the cover is not considered.

The work in [5] transforms the audio file into the frequency domain using DWT. It compares the samples that compose the file into several sub-bands, and it changes the coefficients details by a threshold whose value depends on the expected the embedding secret bit. This work is able to stand against attacks, and have much payload with reasonable quality. However, it cannot guarantee full recovery of the cover file. Moreover, the processing time is also quite complex. The authors in [11] propose an algorithm that is also able to defend from attacks. It performs the hiding process by exploring DCT as the frames

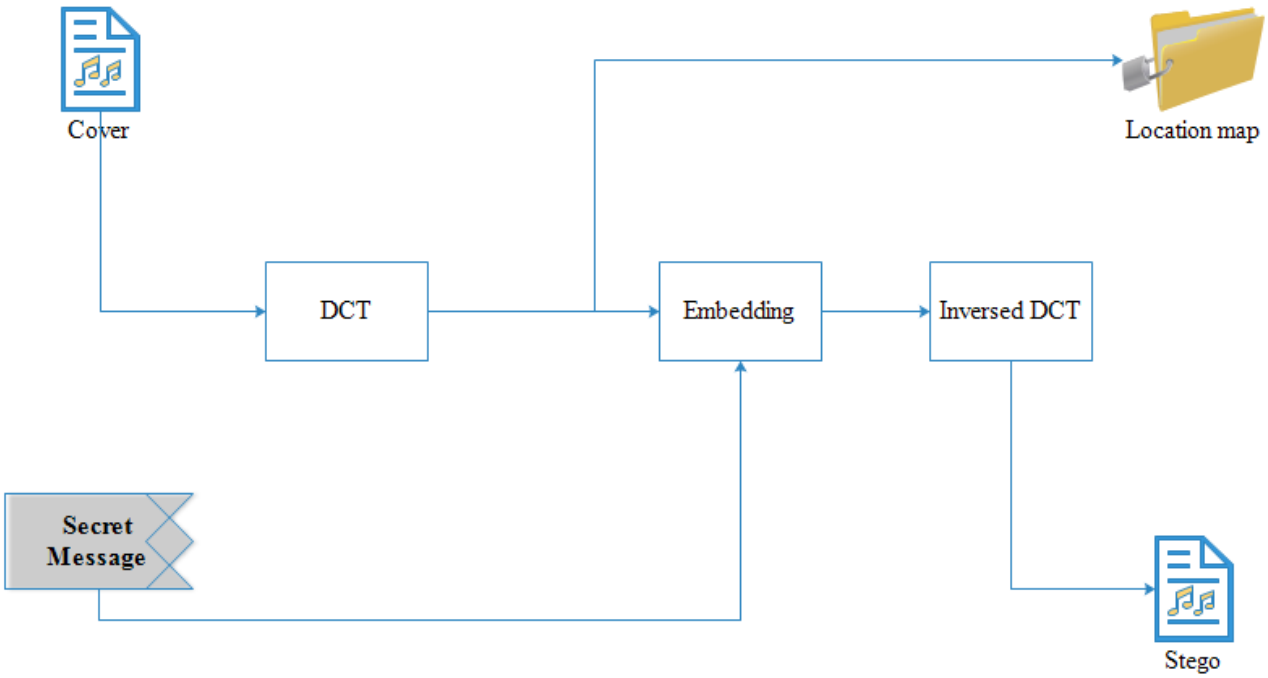

Figure. 1 The embedding procedure

basis. These frames have the same size, and they are not overlapping each other. This work produces reasonable quality, but the recovery of the audio cover is not guaranteed.

3. Proposed algorithm

Based on the methods in [5, 6, 8, 11], this paper proposes a new algorithm by exploring DCT technique and location map. This work mainly focuses on the recovery of both the secret message and the cover file while maintaining the capacity and quality. In this section, we have two main sub-sections that explicitly discuss and present the embedding and extraction operations.

3.1 The embedding

Targeting the robustness, this research considers the embedding process to be done in the frequency domain. The embedding operation is presented in the following steps which are generally demonstrated in Fig. 1.

- Read the samples of the cover audio file. These samples are assigned in a vector T.
- Apply DCT [6] to transform all cover samples T into the frequency values F as defined in Eq. (1). Here, m means the total number of samples.

$$F_n = q(n)\sum_{i=1}^{m} T(i) \times \cos\frac{\pi(2i-1)(n-1)}{2m}, \quad i = 1, \dots, m \tag{1}$$

$$q(n) = \begin{cases} \frac{1}{\sqrt{m}}, & n = 1 \\ \sqrt{\frac{2}{m}}, & 2 \le n \le m \end{cases}$$

- Assign the samples that are signed by F in a vector called location map which is represented using L as in Eq. (2). This helps to guarantee 100% retrieving both the secret message and the cover audio.

$$L = F \tag{2}$$

- Stream the secret message bits S.
- Determine a threshold value $h = 10$.
- Embed the secret bits based on Eq. (3). If the secret bit is 0, then the transformed sample is increased by the threshold value. Otherwise, the threshold value is multiplied by -1 and added to the sample value.

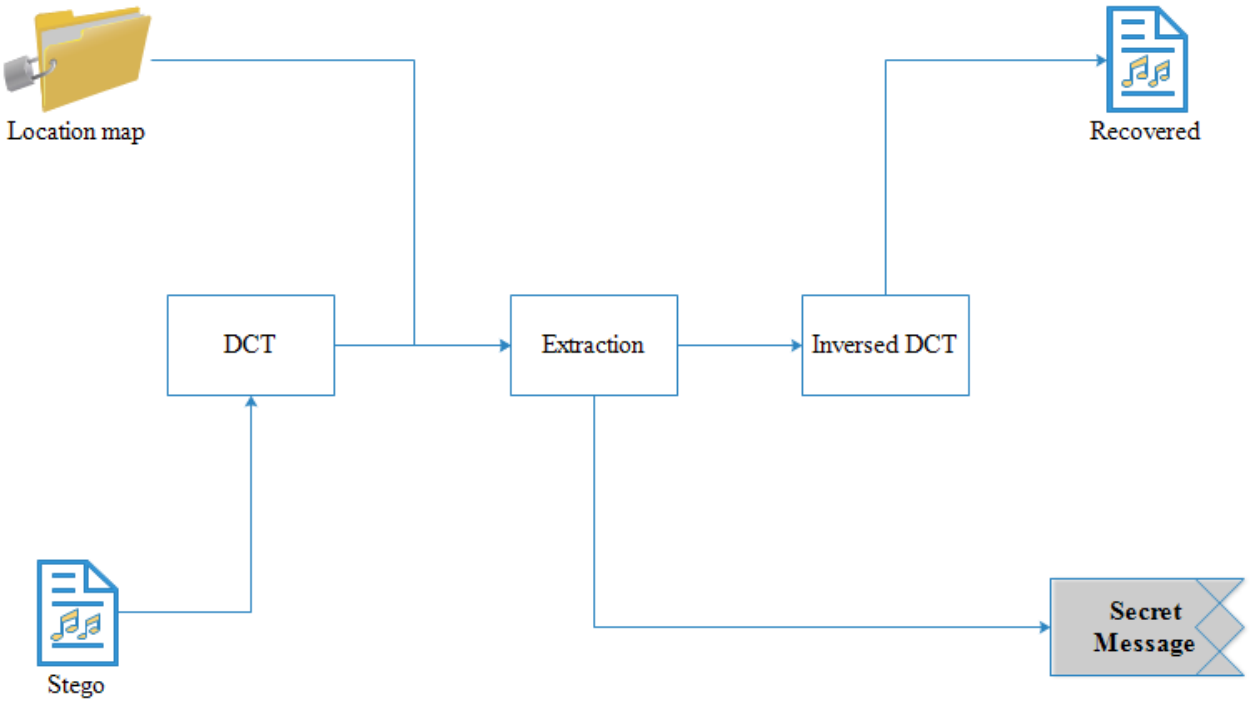

Figure. 2 The extraction procedure

$$F' = \begin{cases} F + (-h), & if\ S = 0 \\ F + (h), & if\ otherwise \end{cases} \quad (3)$$

where F' represents the stego sample in the frequency.

- Change back the stego sample value from the frequency $F'$to the time domain value T'. It is performed by using Inversed Discrete Cosine Transform (IDCT) as in Eq. (4).

$$T'(i) = \sum_{n=1}^{m} q(n) \times F'(n) \times \cos\frac{\pi(2i-1)(n-1)}{2m}, \quad i = 1, \dots, m \quad (4)$$

$$q(n) = \begin{cases} \frac{1}{\sqrt{m}}, & n = 1 \\ \sqrt{\frac{2}{m}}, & 2 \le n \le m \end{cases}$$

- Construct the stego audio and the location map file. For the security reason each file will be sent separately.

3.2 The extraction operation

In this part, the secret message is extracted from the stego audio. As in Fig. 2 the receiver has to obtain two files which are the stego audio and the location map. The second carries the original samples in the frequency domain. These values help to obtain the exact secret message and fully recover the cover audio.

- Read all 16 bit-depth stego audio samples T' in the time domain.
- Read the location map L.
- Use DCT function to transform the samples from time to the frequency domain F'.
- Extract the secret message bit using Eqs. (5) and (6). This process is done by obtaining the difference β between the transformed sample in the stego file and the frequency value that is saved in the location map. Based on it and Eq. (6), we have condition: if this difference is negative, then the secret bit is 0. Otherwise, it is 1.

$$\beta = F' - L \quad (5)$$

$$S' = \begin{cases} 0, & if\ \beta\ is\ negative \\ 1, & if\ otherwise \end{cases} \quad (6)$$

- Retrieve back the obtained original transformed sample values L that are saved in the location map. Convert them from frequency to the time domain T using IDCT.

- Build the recovered audio based on the obtained samples T.

4. The experimental results

In this section, the evaluation procedure of this work is conducted. This paper mainly considers evaluating the reversibility, robustness, capacity, and quality. Ten audio files in varied categories (e.g. speech, music, and animals) are chosen as covers [12-14]. All are read in a binary form that is converted into decimal values using unsigned integer function (Uint16). These values are transformed into the frequency domain. The first 40 samples are ignored since they cause high distortion. The time domain audio samples are transformed into the frequency form. The secret message is generated randomly in binary form using Randi function in Matlab 2017. Preferred data hiding algorithms are the ones that are able to produce a stego audio that is as much identical as possible to the original cover. Normalized Correlation (NC) [15, 16] and Bit Errors Rate (BER) [6, 11] are used for this kind of evaluation as presented in Eqs. (7) and (8).

$$\mu = \frac{\sum_{i=1}^{N}(\Upsilon_i \times \acute{\Upsilon}_i)}{\sqrt{\left(\sum_{i=1}^{N}(\Upsilon_i)^2\right)} \times \sqrt{\left(\sum_{i=1}^{N}(\acute{\Upsilon}_i)^2\right)}} \quad (7)$$

Here, μ represents normalized correlation, Υ_i means the original sample, and $\acute{\Upsilon}_i$ is the stego sample, and N represents the number of samples where μ value range is in between -1 and 1. The closer the value to 1, the better the performance.

$$BER = \frac{1}{N}\sum_{i=1}^{N}(S_i \oplus S_i') \quad (8)$$

The variable BER represents Bit Error Rate, and $\oplus$ means exclusive OR. The value of BER is in the interval [0, 1]. If the recovered values S_i' are exactly same as the original value S_i then BER is 0. In different words, the closer this value to 0 the better the performance, whereas the closer BER to the value 1 means there is no similarity. The evaluation is carried out by measuring some features as in the following points:

- The capacity: we count the number of bits that can be hidden within the cover. Since each sample hides a binary digit, the capacity can be counted using Eq. (9).

$$þ = \acute{C} \times N \quad (9)$$

The parameter þ means the capacity, and $\acute{C}$ is the capability or the number of bits that can be carried on each sample. The variable N represents the number of samples within the cover.

- Invisibility: this evaluates the difficulty that is faced to identify the noise that occurs in the stego file compared to the original one. Mostly, statistical evaluations are used in this type of measurement e.g. histogram analysis [17].
- The quality: it is measured using Signal-to-noise ratio (SNR), NC, and BER [11]. SNR [18] as provided in Eq. (10) always measures the correspondence between the cover and the stego audio files. The higher the value of SNR the bigger the similarity between them, and vice versa. The evaluation scheme BER exists in Eq. (8), it is also used to ensure the quality feature between the original cover audio and the stego one where S_i represents the original cover and S_i' is the stego. As aforementioned above, if the value is closer to 0, the superior the quality of the stego, and vice versa.

$$SNR = 10 \times log_{10}\frac{\sum_{i=1}^{N}(\Upsilon_i)^2}{\sum_{i=1}^{N}(\acute{\Upsilon}_i - \Upsilon_i)^2} \quad (10)$$

- Reversibility: it measures the similarity between the original cover audio and secret message with their corresponding audio and secret message which are recovered from the stego file.

This research is implemented in two scenarios. Based on them, the performance of this work can be indicated as in the following subsections.

4.1 The evaluation based on hiding 10 kb payload

In this scenario, each audio cover file is embedded with 10 kb of secret message. Then, the stego file is evaluated.

First, the invisibility of the stego file is evaluated using cover file No. 1, the signal of the same audio is plotted in both of its original and stego conditions. This plotted signal is shown in Fig. 3 where subfigure (a) is the original signal, and the subfigure (b) is the stego signal. This figure shows that the noise occurs on the stego file is difficult to be noticed by manual human visual.

Then, as shown in Table 1, the quality of the stego file is relatively high. For example, the value of SNR is between 73-97 dB which confirms a high performance in terms of the quality whereas NC cannot detect the small noise in the stego file. It is because the value of NC is 1.000. Moreover, our

work is classified as reversible since it is able to fully recover both the cover audio and the secret message from the stego file.

4.2 The evaluation based on embedding maximum capacity

In this scenario, the cover file is fully embedded by the secret binary digits. This payload is varied according to the number of samples in each file.

The performance of this work in regards to the capacity is presented in Table 2 where this method is able to hide high payload. Based on the evaluation by using Eq. (9), this method guarantees hiding 1 bit per sample.

Similar to the last scenario, invisibility of the obtained stego file is measured. This evaluation is done by plotting the cover file No. 1 where the signal of the same audio is plotted as original and its corresponding fully embedded stego signal. Both of them are shown in Fig. 3 where subfigure (a) shows the original signal, and subfigure (c) shows the stego signal. According to this figure, it is not likely to notice or differentiate which one is the stego signal.

Table 1. The quality of the stego file based on specified 10 kb of payload

No	SNR (dB)	NC	BER
1	90.00	1.000	0.013589
2	73.93	1.000	0.062605
3	90.14	1.000	0.013636
4	96.58	1.000	0.006226
5	94.90	1.000	0.0062727
6	94.64	1.000	0.0062653
7	90.96	1.000	0.013616
8	92.85	1.000	0.0088006
9	99.23	1.000	0.0050781
10	98.38	1.000	0.0051111

The imperceptibility is measured using SNR, NC, and BER. According to Table 2 it can be indicated that this work is capable to produce an imperceptible stego file since the value of SNR is still around 72 dB, NC is 1.000, and BER is close to 0.

After attacking the stego audio file e.g. resampling or MP3 conversion, this proposed research is able to obtain completely both the secret message and the cover file. MP3 conversion means

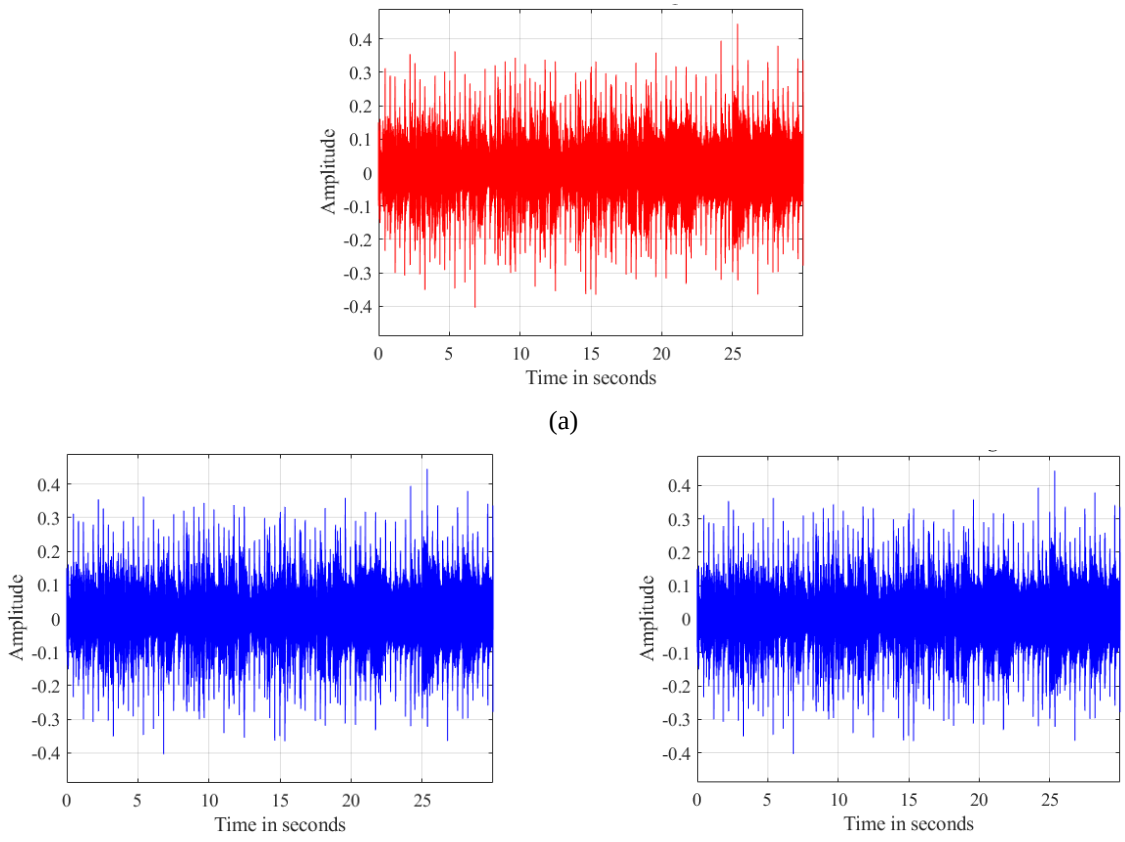

Figure. 3 Audio signal analysis and comparison: (a) original, (b) stego audio with 10 kb data, and (c) stego audio with the maximum capacity

Table 2. The quality of the stego file based on the maximum capacity

No	Payload (bits)	SNR (dB)	NC	BER
1	661500	72.13	1.000	0.012733
2	23113	70.40	1.000	0.067278
3	661500	72.27	1.000	0.011992
4	1801472	74.72	1.000	0.0081877
5	1764000	73.11	1.000	0.0074461
6	1764000	72.85	1.000	0.0074495
7	661500	73.11	1.000	0.01148
8	979360	74.16	1.000	0.0020738
9	5484382	73.16	1.000	0.0043323
10	4661839	72.97	1.000	0.0045613

converting the audio file from WAV to MP3 and retrieve it back to WAV.

This method is evaluated and compared with the previous work in [5]. This experiment is done in two scenarios by using the same secret message and similar cover audio files. Both of the proposed method and the work in [5] achieve the same amount of payload, which is 1 bit per sample.

Fig. 4 shows the performance of the comparison results in terms of the values of SNR after embedding the cover by 10 kb of secret message. It is noticeable that, the proposed method is better than the previous work [5].

Fig. 5 compares the imperceptibility results based on fully embedding the cover files with secret messages. Here, there is a decrease in SNR values. However, the proposed method can still maintain the quality, better than the method in [5].

Histogram analysis is conducted to measure and the invisibility as presented in Fig. 6. Here, subfigure (a) demonstrates the histogram of the original audio file (audio No.3), subfigure (b) demonstrates the stego signal using the proposed method, and

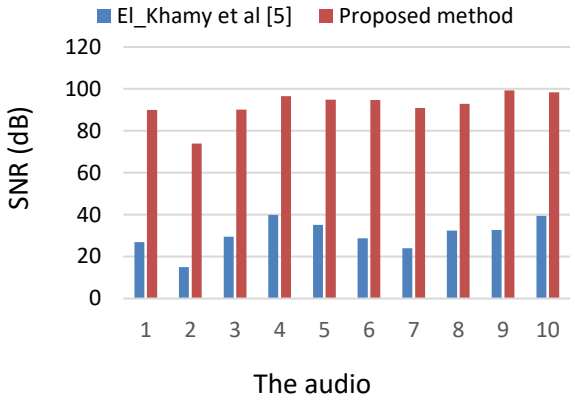

Figure. 4 Results comparison between the proposed method with the previous work after embedding 10 kilobits into each cover file

subfigure (c) is the stego signal using the previous work in [5]. The presented invisibility result proves that the proposed method highly maintains the invisibility of the stego signal compared to [5]. It is shown that the histogram of the stego file based on the proposed method looks similar to the original file.

The prosed method is able to fully recover the cover audio file from the stego one while the method in [5] does not pay attention of the recovery of the cover. However, its secrecy aspect is better than ours.

It is good to note that the previous work has tried to encrypt the secret message before the embedding process in order to be much secured. In contrast, we replaced encrypting the secret message by using the location map file. It is true that the use of the location map is not as secure as the use of the encryption. However, encryption algorithms consume much computational time.

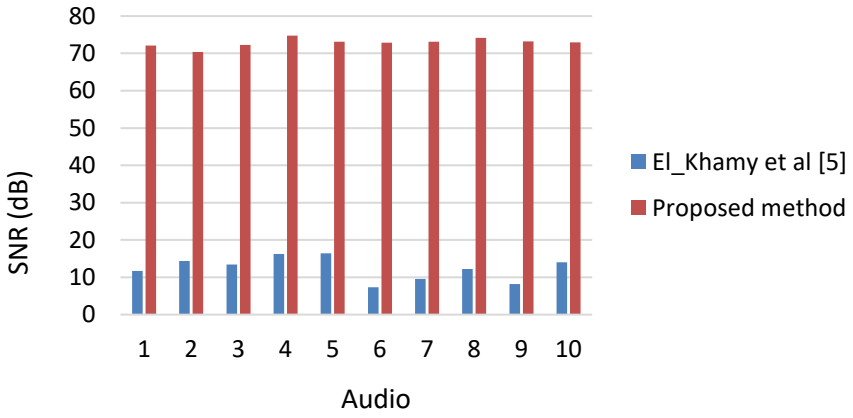

Figure. 5 Results comparison between the proposed method with the previous work based on fully embedded cover files

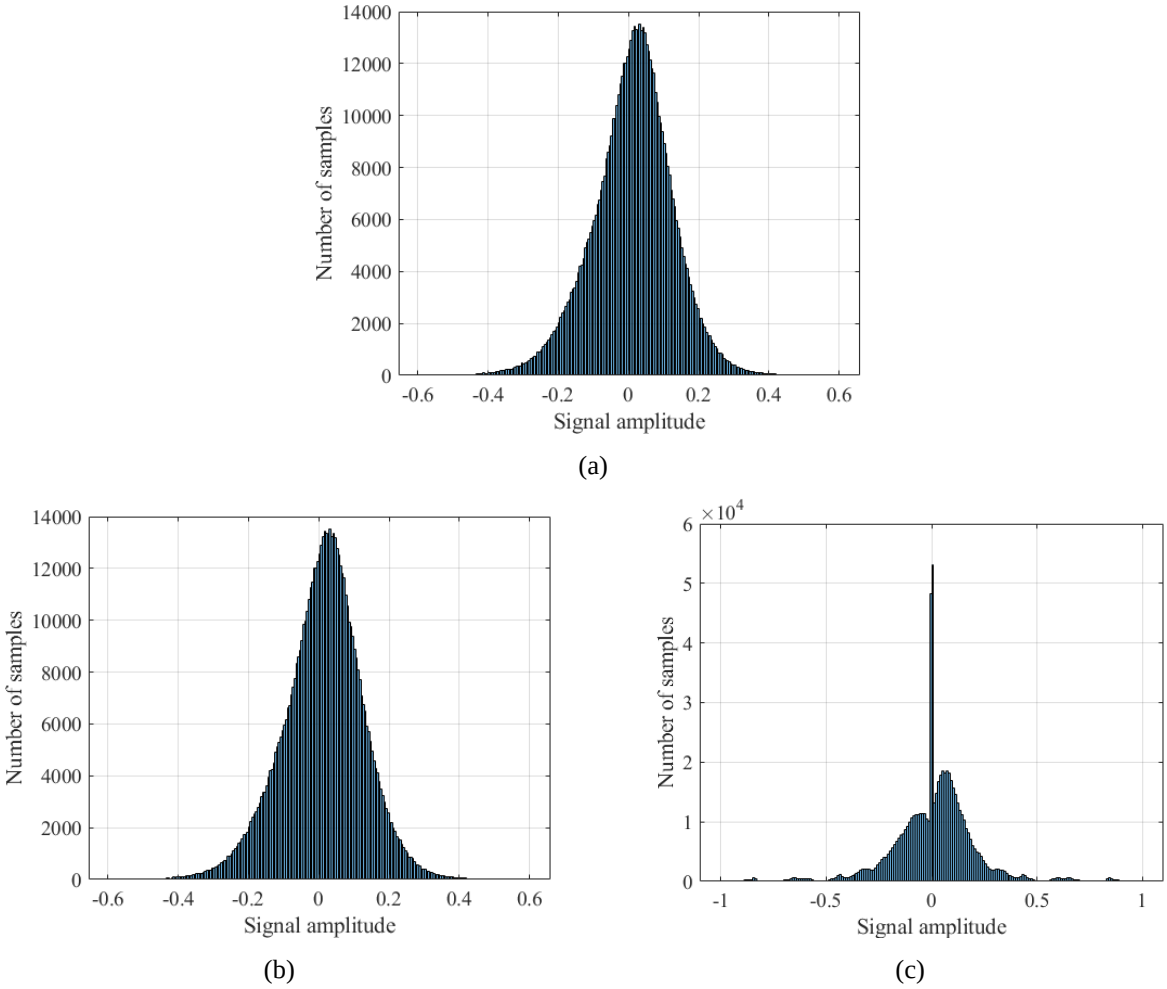

Figure. 6 Audio (No.3) histogram analysis and comparison: (a) original file, (b) stego file using the proposed method, and (c) stego file using El-Khamy et al [5]

5. Conclusion

In this paper, a reversible audio data hiding method is presented. This method uses DCT to transform the audio data into the frequency domain. We exploit the combination of the stego audio and the location map file in order to guarantee recovering the exact secret message and cover file. A threshold value is used to modify the frequency of the samples based on the value of the secret bit.

The results show that this method is able to completely recover both the secret message and the cover file. The capacity and quality are also high that this method is able to hide 1 binary digit per sample. The quality based on SNR is approximately 73 dB, NC is 1.000, and BER is around 0.011992. In addition, this method has same performance as the previous work in terms of the capacity. However, concerning invisibility, quality, and reversibility, this proposed method is superior.

In the future, we are planning to extend this method by exploiting the high quality stego in order to improve the capacity; in addition, the security is also a factor to further investigate. This improvement should still maintain the other features as much as possible.

Acknowledgments

This research is supported by Institut Teknologi Sepuluh Nopember, Indonesia.

Abbreviations

Acronym	Expended Form
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
IDCT	Inversed Discrete Cosine Transform
SVD	Singular Value Decomposition
SNR	Signal to Noise Ratio
NC	Normalized Correlation
BER	Bit Errors Rate

References

[1] B. Y. Lei, Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD–DCT", *Signal Processing*, Vol. 91, No. 8, pp. 1973-1984, 2011.

[2] M. H. A. Al-Huti, T. Ahmad, and S. Djanali, "Increasing the capacity of the secret data using DEpixels blocks and adjusted RDE-based on grayscale images", In: *Proc. of 2015 International Conference on Information, Communication Technology and System*, pp. 225-230, 2015.

[3] N. Kar, K. Mandal, and B. Bhattacharya, "Improved chaos-based video steganography using DNA alphabets", *ICT Express*, Vol. 4, No. 1, pp. 6-13, 2018.

[4] T. Venugopal and V. S. Kumar Reddy, "Image Watermarking Using Two Level Encryption Method Based on Chaotic Logistic Mapping and Rivest Shamir Adleman Algorithm", *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 6, pp. 271-281, 2018.

[5] S. E. El-Khamy, N. O. Korany, and M. H. El-Sherif, "A security enhanced robust audio steganography algorithm for image hiding using sample comparison in discrete wavelet transform domain and RSA encryption", *Multimedia Tools and Applications*, Vol. 76, No. 22, pp. 24091-24106, 2017.

[6] J. Panda, S. Choudhary, K. Nath, and S. Kumar, "Audio Zero Watermarking Scheme based on Sub Band Mean Energy Comparison using DWT-DCT", In: *Proc. of International Conference on Signal Processing and Communication*, pp. 352-357, 2016.

[7] M. Khalil and A. Adib, "Audio watermarking with high embedding capacity based on multiple access techniques", *Digital Signal Processing*, Vol. 34, pp. 116-125, 2014.

[8] M. H. A. Al-Hooti, S. Djanali, and T. Ahmad, "Audio Data Hiding Based on Sample Value Modification Using Modulus Function", *Journal of Information Processing Systems*, Vol. 12, No. 3, 2016.

[9] R. A. Alotaibi and L. A. Elrefaei, "Text-image Watermarking based on Integer Wavelet Transform (IWT) and Discrete Cosine Transform (DCT)", *Applied Computing and Informatics*, [In Press, Available online 2018].

[10] Y. Gangadhar, V. S. G. Akula, and C. R. Pakanati, "Image Adaptive Watermarking Using Feature Point Extraction Model", *International Journal of Intelligent Engineering and Systems*, Vol. 10, No. 1, pp. 95-103, 2017.

[11] H.-T. Hu and L.-Y. Hsu, "Robust, transparent and high-capacity audio watermarking in DCT domain", *Signal Processing*, Vol. 109, pp. 226-235, 2015.

[12] G. Tzanetakis, "MARSYAS Music Analysis and Synthesis for Audio Signal", [online, URL: http://marsyas.info/downloads/datasets.html], 2002.

[13] Illinois Speech and Language Engineering, [online, URL: http://www.isle.illinois.edu/].

[14] Music Technology Group, Universitat Pompeu Fabra, IRMAS: a dataset for instrument recognition in musical audio signals, [online, URL: https://www.upf.edu/web/mtg/irmas].

[15] F. Ciompi, C. Gatta, O. Pujol, O. Rodrıguez-Leor, J. M. Ferré, and P. Radeva, "Reconstruction and Analysis of Intravascular Ultrasound Sequences", *Recent Advances in Biomedical Signal Processing*, Vol. 223, No. 243, pp. 231-250, 2011.

[16] M. Hemis and B. Boudraa, "Digital Watermarking in Audio for Copyright Protection", In: *Proc. of International Conference on Advanced Computer Science and Information System*, pp. 189-193, 2014.

[17] A. Rashid and M. K. Rahim, "Critical Analysis of Steganography "An Art of Hidden Writing", *International Journal of Security and Its Applications*, Vol. 10, No. 3, pp. 259-282, 2016.

[18] M. H. A. Al-Hooti, T. Ahmad, and S. Djanali, "Developing Audio Data Hiding Scheme using Random Sample Bits with Logical Operators", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 13, No. 1, pp. 147-154, 2019.