Steganography in Audio Signals using Variable Bit Replacement Method in DCT Domain

Vikash Ramesh,
Department of Electronics and Communication,
Amrita VishwaVidyapeetham,
Coimbatore, India

Kaushik Narayanan,
Department of Electronics and Communication,
Amrita VishwaVidyapeetham,
Coimbatore, India

PremalathaPandian,
Department of Electronics and Communication,
Amrita VishwaVidyapeetham,
Coimbatore, India

***Abstract*—Steganography is the method of concealing a message or image within a media, be it audio, image or video. The purpose is the same as that of cryptography, which is to secure data. But the difference is that in steganography, the sole existence of such a secret message is known only to the sender and receiver of that message. Audio steganography is such a technique of hiding a message within an audio file (known as the cover audio). Lately, lot of novel and versatile Audio Steganographic methods have been proposed. A good audio steganographic technique not only intends at embedding data in such a way that the changes made in the cover audio are indiscernible but also at the efficient extraction of the data. The least-significant-bit (LSB) based approach is popular among the steganographic algorithms. Our aim here is creating an audio steganography technique by changing the LSBs of the Discrete-Cosine-Transform (DCT) coefficients of the Cover Audio. It is different from the conventional LSB technique in DCT domain because, here in proposed method data is embedded by spreading it all over the layers of the LSBs of the DCT coefficients. The technique has proven to be robust than the conventional LSB technique and to the channel induced noise as well, in terms of performance evaluation. The same conclusion is drawn from the experimental results obtained by using the DCT domain features and the Support Vector Machine (SVM) as the classifier. Besides, check on the integrity of the message and imperceptibility of the changes is also performed.**

***Index Terms*—Audio Steganography, LSB based steganography, DCT, Embedding Capacity, SNR.**

I. INTRODUCTION

Due to the advancement in communication, there has been a rapid growth in digital data usage. Efficient secrecy can be achieved by implementing cryptography, watermarking, or steganography techniques [1]. Cryptography techniques are based on rendering the content of a message garbled to unauthorized people. In watermarking, data is hidden to convey some information about the cover medium such as ownership and copyright, whereas steganography is a process of embedding secret messages in a cover signal to avoid illegal detection [2]. Steganography differs from cryptography in terms of message visibility. It hides secret messages totally compared to cryptography where the secret message is still visible [3].

Steganography is mostly used in secret communication like military and government communications. Often it requires relatively high payloads when compared to watermarking. The major requirements that should be satisfied for good steganography algorithms include perceptual transparency, payload or capacity and robustness [4]. High capacity is considered as an important aspect for steganography when compared to watermarking. For watermarking, robustness should be a dominant factor. Improvement for one of the mentioned requirements will tend to degrade the other performances as they are contradictory according to the magic triangle [5].In recently years many techniques have been developed for information hiding [6, 7, 8], and most of these techniques used either image and video media but rarely use audio signal as a cover signal especially in high rate of data embedding, most likely due to Human Auditory System (HAS) which is more sensitive compared to the Human Visual System (HVS) [8]. But still with help of robust algorithms it is possible to exploit the 'holes' in the HAS.

Hiding information in an audio file requires following elements (as shown in Figure.1) [9]:

- The cover media that will hold the hidden data
- The secret message, may be plain text, cipher text or any type of data
- The stego function and its inverse
- An optional stego-key or password may be used to hide and unhide the message.

II. RELATED WORK

Steganography can be done both in the spatial and in the frequency domain. The least significant-bit (LSB) based techniques are very popular for steganography in spatial domain. But even though conventional LSB technique and its variants provide an easy and simple way to hide data robustness and security are not the main characteristics of

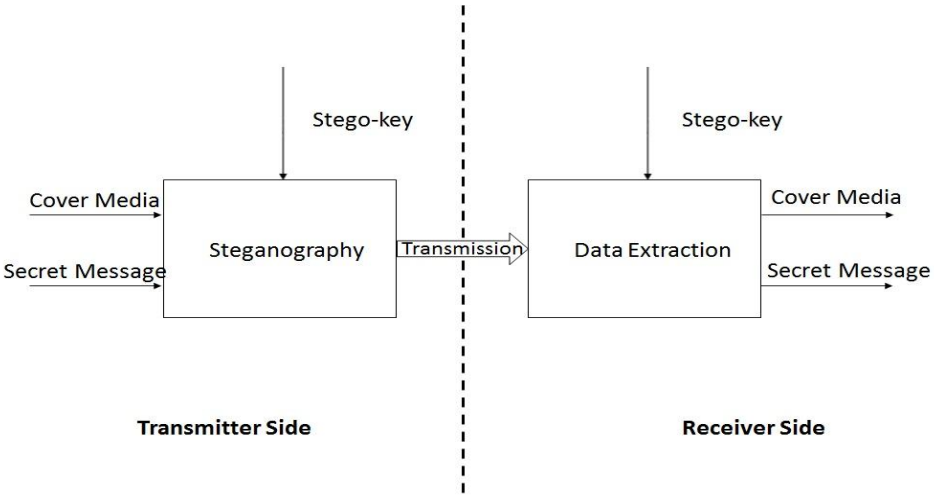

Fig. 1 The Steganographic operation

temporal domain steganographic methods [10]. Tolerance to noise addition at low levels and some robustness criteria havebeen achieved with LSB variants' methods [11-12], but at a very low hiding capacity. Hence embedding of data in the frequency domain was preferred. According to [10] transform domain techniques provide better robustness and high embedding capacity.

In the frequency domain, embedding in the DCT domain is a very popular technique. The Discrete Cosine Transform (DCT) decomposes a signal into two components, high and low frequency components. Most power of the input signal is concentrated in low frequency component called DC signal, while little power exists in the high frequency component or known as an AC signal.

The modification in the AC component little effect on the reconstructed signal, However modification in the DC component or low frequency component may affect significantly the reconstructed signal. Therefore using AC components as a cover for information embedded process enable high payload and an acceptable quality, when it is used in the steganography [8]. However, information embedding in AC component can affect its robustness as it is possible to remove a secret message by signal processing for example an attacker may reset the AC coefficients.

III. DISCRETE COSINE TRANSFORM

The discrete cosine transform is the spectral transformation, which has the properties of Discrete Fourier Transformation (DFT) [13]. Usage of DCT is popular than DFT because unlike the latter, DCT uses only cosine functions of various wave numbers as basic functions and operates on real valued signals and spectral coefficients. The reconstruction of original signalfrom its DCT coefficients is termed as inverse discrete cosine transform (IDCT).

Some other advantages are the properties of DCT like de-correlation, energy compaction, reparability, symmetry and orthogonality [14]. DCT packs the energy of the signal into the low frequency regions which provides an option of reducing the size of the signal without reducing the quality of the signal.

Since we are working on audio signals here, we focus on the DCT of a 1-D sequence, especially DCT-II and its respective inverse, IDCT-II.Following are the definitions of the DCT-IIand IDCT-II [15]:

A. *DCT-II*

$$y(k) = w(k)\sum_{k=1}^{N} x(k)\cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right), \quad k=1,2\ldots N$$

B. *IDCT-II*

$$x(k) = \sum_{k=1}^{N} w(k)\, y(k)\, cos\left(\frac{\pi(2n-1)}{2N}\right), \quad k=1,2\ldots N$$

where*w(k)* used both in DCT and IDCT is defined as

$$w(k) = \begin{cases} \frac{1}{\sqrt{N}} & k = 1 \\ \sqrt{\frac{2}{N}} & , \ 2 \le k \le N \end{cases}$$

IV. COMPRESSION AND ENCODING

A. *Compression using Huffman Coding:*

The secret message can be of any length, depending on the user's necessity. In order to achieve high embedding capacity, the message is to be first compressed using an adaptive Huffman Code. Here in the proposed method, the probabilities of every character in the secret message is found out with respect to the secret message and the probability table (known as the Huffman dictionary) is used to compress the message, using Huffman encoder.

B. *Repetition Code:*

The compressed message has to be embedded in the DCT coefficients and then retrieved at the receiver's side using the IDCT. Though DCT is a lossless transform, the

floating point representation of the DCT coefficients would induce some loss into the data that would pose a problem to the perfect extraction. Thus this situation demands the secret message to be encoded for efficient recovery at the receiver's side. The repetition code increases redundancy and decreases the probability of error. The extraction at the receiver's side is done by using the Majority Logic Decoding, i.e., the bit with highest frequency is considered as the message bit.

V. PROPOSED METHOD

The developed technique is based on LSB steganography, a substitution steganography that replaces the least significant bit of the DCT coefficients of the cover audio.

The whole process is mainly of two stages: The sender's side and the receiver's side:

A. Sender's Side:

a) Compression of the secret message:

Here the secret message is taken as an input from the user, and the corresponding Huffman Dictionary is calculated dynamically, using which the message is compressed using Huffman Coding.

b) Usage of repetition code:

The compressed message along with the Huffman dictionary is encoded using a (5,1) repetition code.

c) Finding the DCT coefficients of the Cover Audio:

In this step the Cover Audio is first mapped onto the DCT domain, using the formula for DCT in Section III. It is then converted to its corresponding binary form, after multiplying it with a suitable scaling factor (say 2^{15}).

d) Embedding of the data:

The spread factor and bit position are taken as inputs from the user in this step. The compressed and encoded secret message is then spread along layers of the LSBs of the DCT coefficients in a sine wave pattern. The bit position defines the amplitude of the sine wave pattern. The spread factor defines the frequency of variation of LSBs changed. The message is then embedded by changing the LSBs depending on the spread factor and relative to the length of the message (See Figure.2). In the end a parity-bit is added to facilitate the integrity check on the message.

e) Transmission:

The resulting binary sequence is then converted to the decimal form.

The IDCT-II of decimal data in the previous stage is taken to get the stego file, which is transmitted to the receiver.

f) Key sequence:

The spread factor and the bit position is XORed with a predefined sequence and is transmitted to the receiver through a covert channel as a 'key sequence'. Refer Figure.3 for the process on the sender's side.

B. Receiver's Side:

a) *Extraction of the data:*

The DCT-II of the stego file received at the receiver's end is taken. It is again scaled and converted to binary in the same way as in the sender's side. Using both the covertly received key sequence and majority logic decodingthe Huffman dictionary and compressed message is extracted. The message is then decoded to its original form using the Huffman dictionary.

The SNR is found out for varying embedding capacities in different audio files. A parity – bit check is also done on the decoded message to check its integrity. Also, the DCT features were extracted from the stego files and used to steganalyse using LIBSVM.

VI. EVALUATION METRICS

A. Signal-to-Noise Ratio (SNR):

The SNR is very sensitive to the time alignment of the original and distorted audio signal [16]. The value of SNR indicates the distortion amount induced by the embedded data in the cover audio signal [10]. The SNR is measured as

$$SNR = 10\ log_{10}\left(\frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N}(x(i)-y(i))^2}\right)$$

where $x(i)$ is cover audio and $y(i)$ is the stego audio.

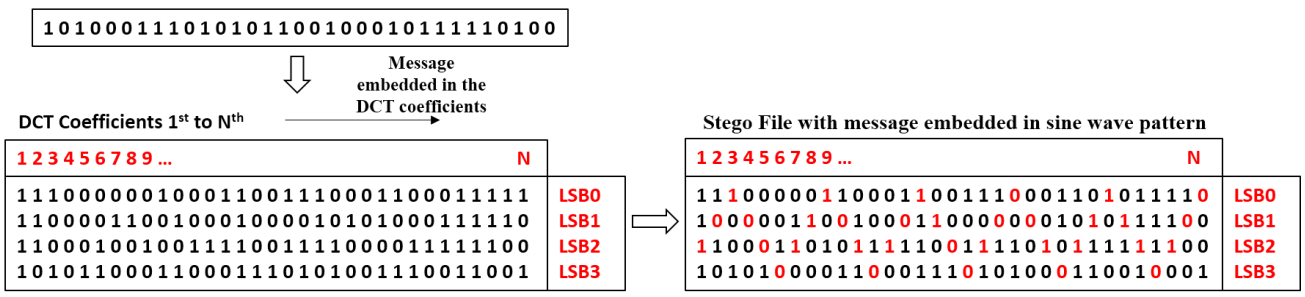

Fig. 2 Embedding Process

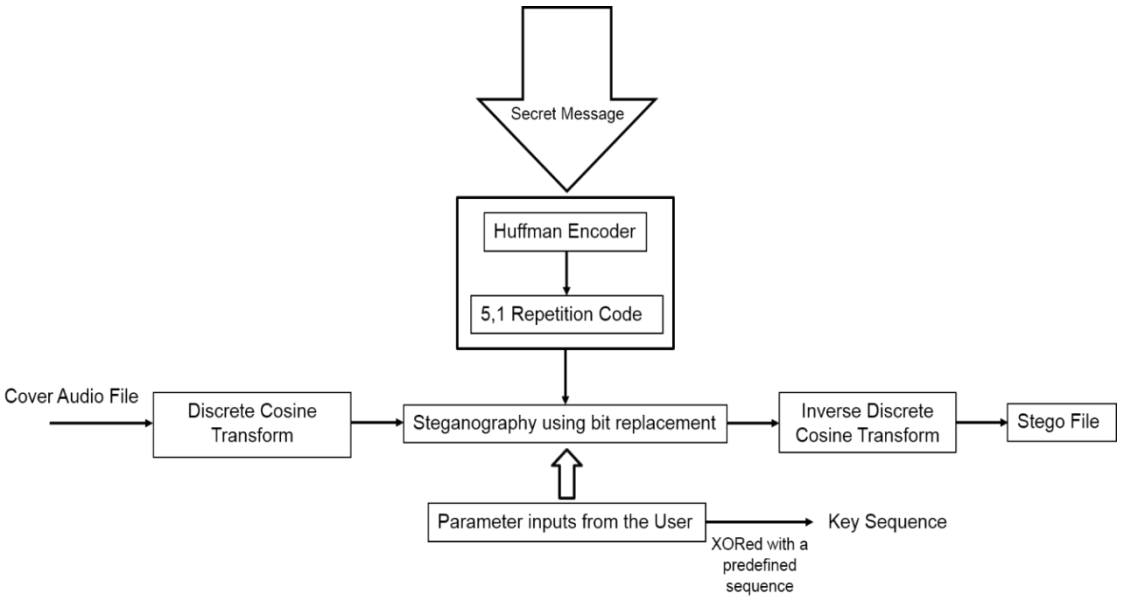

Fig. 3 Sender's Side

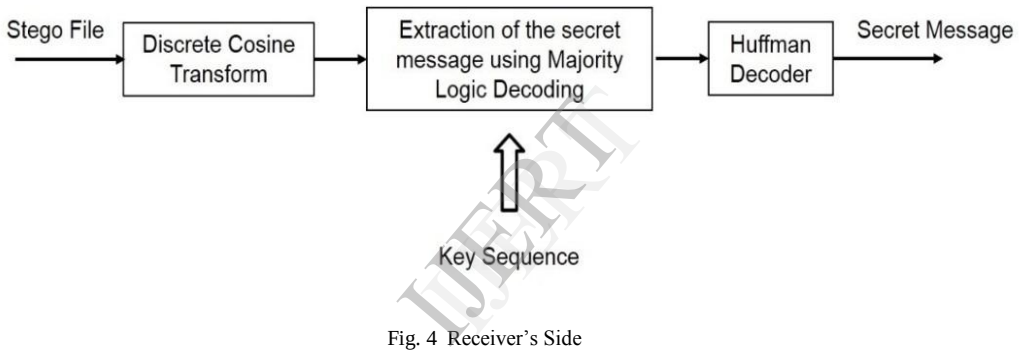

Fig. 4 Receiver's Side

B. *Probability of Error (P_E):*

For every steganographic method, stego files using a range of different embedding capacities were created, and trained a separate classifier to detect each of them. Before classification, all cover-stego pairs were divided into 80% for training and 20% testing, respectively. The minimal average error under equal prior probabilities is given by

$$P_E = minP_{FA}\frac{(P_{FA} + P_{MD}(P_{FA}))}{2}$$

where P_{FA} is the false alarm rate and P_{MD} is the missed detection rate. Lesser the P_E, better is the steganography algorithm.

C. *Accuracy Of Detection:*

This metric is the accuracy of detection of the SVM classifier, i.e. the accuracy with which it detects stego files. Hence lesser is the accuracy, more robust is the steganography in terms of imperceptibility.

VII. DCT FEATURES

Here in our paper we have tried the 1-D incorporation of the DCT steganalysis features analyzed in [17, 18]. We limit our analysis only to three main features off all of those studied in [17, 18], namely – Global histogram, Local histogram and Markov horizontal features.

i. *Global and Local Histograms:*

The simplest first order statistic of DCT coefficients is their histogram. Suppose the stego file is represented with a DCT coefficient array $d_k(i)$ i= 1,…,m (m is the length of the frame chosen), k = 1, …, B. The symbol $d_k(i)$ denotes the (i)-th DCT coefficient in the k-th block (there are total of B blocks).

Thus the first functional or the Global histogram is the histogram **H** of all m x k DCT coefficients

$$\mathbf{H} = (H_L, \ldots, H_R),$$

where$L = \min_{k,i} d_k(i)$ and $R = \max_{k,i} d_k(i)$.

However, those schemes only preserve the global histogram and not necessarily histograms of individual DCT modes (local histograms). Thus, we add individual histograms for low frequency DCT modes to our set of functionals. For a fixed DCT mode (i), let, h_r^i , r = L, …,R, denote the individual histogram of values dk(i), k = 1, …, B. We only use histograms of low frequency DCT coefficients (say, modes [2-4]) because histograms of coefficients from medium and higher frequencies are usually statistically unimportant due to the small number of non-zero coefficients [17].

ii. Markov horizontal features:

The Markov feature set models the differences between absolute values of neighboring DCT coefficients as a Markov process. The feature calculation starts by forming the horizontal array $F_h(u)$ of absolute values of DCT coefficients of the audio. The 1-D incorporation of the Markov horizontal features referred in [18] is:

$$M_h = \frac{\sum_{u=1}^{S_u-1} \delta(F_h(u) = i, F_h(u+1) = j)}{\sum_{u=1}^{S_u-1} \delta(F_h(u) = i)},$$

whereS_u is the length of the audio file and δ = 1 if and only if its argument(s) are satisfied i.e., $(i,j) \in \mathbf{A}$, where A is a definite set of integers. In our proposed method we have taken **A**= [-6, +6].

VIII. EXPERIMENTAL EVALUATION

250 cover audio files of .WAV format each ranging between 5 to 7 second was taken. Each audio file was subjected to four types of steganographic algorithms – LSB[1], SLSB[2], DCTLSB[3], DCTSLSB[4]

To find the SNR of each audio file for different embedding capacities, 5 files – c, c1, l, s1, and s2 were selected and their SNR was calculated for embedding capacities of 25%, 50%, 75%, and100%. The values of SNR with respect to varying embedding capacities were tabulated for each audio file subjected to different steganographic algorithms as shown in Table I.

From the total data set of 250 files, 200 files were used to train the LIBSVM, and the remaining 50 files were used as the test files to predict and classify using the LIBSVM. A total set of 136 feature – sets were taken for the purpose of analysis.The performance measures - Accuracy of Detection and Probability of Error were found out as in Table II.

[1] Changing LSB in Time Domain
[2] Changing LSB in Time Domain in sine wave pattern
[3] Changing LSB in DCT domain
[4] Proposed method by changing LSB in DCT domain in sine wave pattern

TABLE I. SNR (IN DB) FOR VARYING EMBEDDING CAPACITIES

Audio File (.wav)	Embedding Capacity = 25%			
	LSB	SLSB	DCTLSB	DCTSLSB
c	170.7969	116.1443	155.1791	116.0791
c1	175.0825	120.5241	159.7979	120.3259
l	169.5309	114.9561	154.3582	114.7729
s1	176.3382	121.5952	161.5996	121.6297
s2	183.0878	128.3370	162.4775	128.2016

Audio File (.wav)	Embedding Capacity = 50%			
	LSB	SLSB	DCTLSB	DCTSLSB
c	165.2484	110.5716	153.4207	110.5038
c1	169.5140	114.8876	157.1472	114.7433
l	163.9076	109.1776	152.6969	109.4110
s1	170.7238	116.1404	158.8070	116.0233
s2	177.5150	122.7871	160.7939	122.7022

Audio File (.wav)	Embedding Capacity = 75%			
	LSB	SLSB	DCTLSB	DCTSLSB
c	161.6891	106.9848	152.4618	106.9777
c1	165.9311	111.2898	157.1380	111.2264
l	160.3332	105.6780	151.5000	105.7855
s1	167.1616	112.8217	156.9387	112.4516
s2	173.9274	119.2011	159.7952	119.2170

Audio File (.wav)	Embedding Capacity = 100%			
	LSB	SLSB	DCTLSB	DCTSLSB
c	159.0944	104.5320	151.5535	104.3840
c1	163.2958	108.7317	156.2396	108.6597
l	157.6975	103.2243	150.8143	103.1387
s1	166.3490	112.0197	154.0402	111.6424
s2	171.3135	116.6709	157.8580	116.5901

TABLE II. PERFORMANCE EVALUATION FOR VARIOUS EMBEDDING CAPACITIES USING LIBSVM

Method	Embedding Capacity (%)	Accuracy of Detection (%)	P_E
LSB	25	60	0.480
	50	64	0.486
	75	64	0.474
	100	66	0.468
SLSB	25	50	0.480
	50	52	0.474
	75	58	0.474
	100	62	0.462

DCTLSB	25	46	0.280
	50	42	0.280
	75	38	0.276
	100	38	0.284
DCTSLSB	25	28	0.276
	50	30	0.272
	75	30	0.284
	100	36	0.268

CONCLUSION

Upon hard scrutiny of the results of the experimental analysis, it can be concluded that the proposed method (DCTSLSB) is good when compared to LSB, SLSB, and DCTLSB. This can be concluded from the following checks:

- The proposed method has the least Probability of Error (P_E) when compared to other methods taken into consideration. Also, it has the lowest Accuracy of Detection (see Table II). This means it is robust in terms of detection when compared to other methods. Thus the proposed method has a very good performance in terms of security and imperceptibility towards detection.

- The integrity check on the secret message, using a parity bit, was conclusive enough to establish the fact that the proposed method is reliable enough in all the embedding capacities.

- However on seeing Table I, it can be seen that there is a trade-off been done in SNR, which is very good in the LSB method. However as SNR in the case of the proposed method is still above 100 dB in the proposed method it remains efficient in terms of imperceptibility to HAS.

Thus the proposed method is good than the other methods in terms of security of the information, reliability and imperceptibility to the HAS.

ACKNOWLEDGMENT

At the very outset, we would like to thank the Almighty who gave us the wisdom and knowledge to complete this dissertation.

We express our gratitude to our guide, Ms. Premalatha. P, Assistant Professor, Department of Electronics and Communication, Amrita VishwaVidyapeetham , Coimbatore, for her valuable suggestion and timely feedback during the course of this dissertation.

REFERENCES

[1] Walter Bender, Daniel Gruhl, Norishige Morimoto, Anthony Lu, “Techniques for Data Hiding”, IBM Systems Journal, vol.35, no. 3 and 4, pp. 313-336, 1996.

[2] N. Provos and P. Honeyman, “Hide and seek: An introduction to steganography,” IEEE Security and Privacy Magazine, Vol. 1, No.3, June 2003, pp. 32–44.

[3] H. Wang, and S. Wang, “Cyber warfare: Steganography vs.Steganalysis,” Communications of the ACM magazine, Vol. 47,No.10, October 2004, pp. 76-82.

[4] Y. Wang, and P.Moulin, “Perfectly Secure Steganography: Capacity, Error Exponents, and Code Constructions,” Information Theory IEEE Transactions, Vol. 54, No. 6, Jun 2008, pp. 2706 – 2722.

[5] N. Cvejic,“Algorithms for Audio Watermarking and Steganography,” MSc. thesis, Department of Electrical and Information Engineering, Information Processing Laboratory, University of Oulu, Finland Oulu, Finland, 2004.

[6] K. Bailey, and K. Curran, “An Evaluation of image based Steganography methods,” Journal of multimedia Tools and Applications, Vol. 30, No. 1, July, 2006, pp. 55-88.

[7] Meghanathan, Natarajan, and LopamudraNayak. "Steganalysis Algorithms For Detecting The Hidden Information In Image, Audio And Video Cover Media," International Journal of Network Security & Its Applications 2.1 (2010), Vol.2, No.1, January 2010, pp. 43-55.

[8] S. Shahreza and M. Shalmani, “High capacity error free wavelet Domain Speech Steganography,” IEEE International conference on acoustics, speech, and signal processing, March 31 -April 4, 2008, pp. 1729 – 1732.

[9] Adhiya, K. P., and Swati A. Patil. "Hiding Text in Audio Using LSB Based Steganography," Information & Knowledge Management (2224-896 X) 2.3 (2012), Vol 2, No.3, 2012.

[10] Djebbar, Fatiha, et al. "Comparative study of digital audio steganography techniques," EURASIP Journal on Audio, Speech, and Music Processing 2012.1 (2012): 1-16.

[11] N. Cvejic, and T. Seppanen, "Reduced distortion bit-modification for LSB audio steganography," Journal of Universal Computer Science, vol. 11, no. 1, pp. 56-65, January 2005.

[12] Ahmed, Mohamed A. Miss Laiha Mat Kiah, BB Zaidan and AA Zaidan, “A Novel Embedding Method to Increase Capacity and Robustness of Low-bit Encoding Audio Steganography Technique Using Noise Gate Software Logic Algorithm,” Journal of Applied Sciences 10 (2010): 59-64.

[13] A. B. Watson, “Image Compression Using the Discrete Cosine Transform”, Mathematical Journal , vol. 4(1), pp. 81-88, 1994.

[14] S. Ali Khayam, “The Discrete Cosine Transform (DCT): Theory and Application,” Information Theory and Coding, Seminar 1 – The Discrete Cosine Transform: Theory and Application, March 10, 2003.

[15] Ahmed, N., Natarajan, T., Rao, K. R. (January 1974), "Discrete Cosine Transform," IEEE Transactions on Computers C–23 (1): 90–93.

[16] Ozer, Hamza, et al. "Steganalysis of audio based on audio quality metrics," Electronic Imaging 2003. International Society for Optics and Photonics, 2003.

[17] Fridrich, Jessica. "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," Information Hiding. Springer Berlin Heidelberg, 2005.

[18] Pevny, Tomas, and Jessica Fridrich. "Merging Markov and DCT features for multi-class JPEG steganalysis," Electronic Imaging 2007. International Society for Optics and Photonics, 2007.