Enhanced Secret Communication Using Multilevel Audio Steganography

K. Bhowal[1], D. Chanda(Sarkar)[2], S. Biswas[3], P.P. Sarkar[4]
[1, 2, 3, 4] *DETS, University of Kalyani, Kolkata, India.*

ABSTRACT

The paper presents multilevel audio Steganography, which describes a new model for hidden communication in secret communication technology. In multilevel Steganography, at least two embedding methods are used in such a way that the second method will use the first method as a carrier. This approach has several potential benefits in hidden communication. It can be used to increase the level of security while transmitting the confidential information over public channels or internet. It can also be used to provide two or more information hiding solutions simultaneously. Another important advantage is that the lower-level embedding / extracting method and upper-level embedding / extracting method are interrelated in terms of functionality and this makes the hidden communication harder to detect. If the cover object is decoded by any adversary, he/she only obtains a decoy message or a partial message. The performance of the proposed scheme in terms of imperceptibility, capacity & security is measured through different experiments and results are included in this paper.

Keywords: *Audio Steganography, Multilevel Steganography, Secret Communication, Information security, Imperceptibility, Embedding Capacity, Discrete Wavelet Transform*

I. INTRODUCTION

The main aim of Steganography is to hide secret information in digital cover. The modification of the cover caused by embedding secret information remains invisible to the third party observer. This is possible by designing a suitable embedding algorithm and choosing an appropriate cover. That means, there will be no significant difference between original cover and modified cover. Thus, secret information not only are hidden inside the cover, but the fact of the secret information transmission is also hidden. Each Steganographic method may be characterized by following requirements. First, undetectability is defined as the inability of detecting secret information inside the cover. In fact, the distortion of the embedded cover convinces the adversary to analyze the statistical properties of the cover and compare them to the distinctive properties of that cover. So, imperceptibility or inaudibility is directly proportional to the undetectability. Second, embedding capacity is defined as an amount of secret information can be transmitted using a particular algorithm per unit of time. Third, Steganographic cost, which defines the amount of distortion of the cover caused due to the secret information embedding procedure. The Steganographic cost depends on the cover used as a carrier and embedding algorithm.

For each Steganographic method, there is always a trade-off between maximizing embedding capacity and remaining secret information undetected. Therefore, a certain level of tuning between embedding capacity and undetectability is required. As long as embedding and extracting algorithm remains secret to the adversary, it can be used to transmit confidential information freely. But, if the both algorithms are known to the adversary, anyone may be able to extract the secret information. The problem may be solved by using the encryption algorithm. The encryption algorithm AES may be used to encrypt secret information before embedding process. So, in this case, extracted information will not be readable. Still, there is a problem with this approach. Because, the encryption key and the encrypted information are transmitted using the same embedding method. Thus, the encryption key and the secret information both will be discovered on detection. On the other hand, embedding capacity may be compromised due to embedding of key in cover. The Multilevel Steganography was originally proposed by Al-Najjar for picture steganography in [1]. The basic idea in this paper was to hide a decoy image into LSB positions of the cover and the original secret information is embedded into the LSB positions of the decoy image.

Confidential information hiding takes normally two general methods: encryption and steganography [2, 3, 4]. In encryption, information will hide using an encoding method that only authorized persons with the proper key can decode it. On the other hand, Steganography hides information such a way that there is hidden information is not specious to the regular observer. The secret raw information can be inserted directly or some transformation can be applied to it before the hiding process. Normally, transformations include encryption, compression, transformation or a combination of digital transformation techniques. An example is the hiding of a text-object into another text-object [5]. A method of hiding plain text into an audio signal is proposed in [4]. Another method is proposed in [4], where audio is hidden in an image object. In [6], audio is hidden in an image after performing encryption and compression. The hiding an image into another image is proposed in [5, 7] and hiding an image into a video is proposed in [5, 8].

The information hiding techniques are developed for the protection of medical information in [9]. This paper suggests a multiple-layer data hiding technique in spatial domain. It utilizes a reduced difference expansion method to embed the bitstream in the least significant bits (LSBs) of the expanded differences. By using the reduced difference expansion method, a large amount of data is embedded in a medical image whose quality can also be maintained. Moreover, the original image can be restored after extracting the hidden data from the stego-image.

In this paper, multilevel audio steganography is proposed to address the above stated problems. The proposed approach extends the concept of steganography to use it in more general purpose.

1.1 Classification based on the area of application:

Multilevel steganography can be categorized based on the requirement of its application. Embedding capacity is the basic requirement in some of the applications where imperceptibility may be compromised in a certain level. On the other hand, imperceptibility is the main requirement in some of the applications where embedding capacity may be compromised in a certain level. So, the multilevel steganography may be classified as like below:

i) Single message multiple covers - multilevel steganography denoted as **TYPE-I**
   A message is embedded in multiple covers using several embedding functions to increase the level of security of the system. This approach provides better imperceptibility, but embedding capacity may be compromised in most of the cases.

ii) Single cover multiple messages - multilevel steganography denoted as **TYPE-II**
   Multiple messages are embedded in a single digital cover using several related embedding functions to increase the embedding capacity of the system. This approach provides better embedding capacity, but imperceptibility may be compromised in most of the cases.

II. PROPOSED METHOD

In this section, TYPE-I and TYPE-II types of multilevel steganography models are discussed.

2.1 Single message multiple covers - multilevel steganography model (TYPE-I)

Suppose, the Message is denoted as M, the Covers are denoted as C_i, the Intermediate Covers or stego-covers are denoted as I_i. Here, the value of i depends on the level of steganography is to be performed.

The message M is passed through the transformation T_i. The transformations may include compression, encryption or a transforms like Discrete Cosine Transform (DCT), Fourier Transform (FT) or Discrete Wavelet Transform (DWT), etc. Sometimes a combination of techniques may be used as required by the particular application.

Message embedding and extracting operations are performed by the embedding and extracting function pairs embed() and extract() and denoted by f and f' respectively. The message embedding function may vary to improve the steganography attributes like imperceptibility, capacity, and robustness.

T_i = I means no transformation is applied. In the blind system, hidden information is extracted without using cover C_i at the receiving end.

The TYPE-I multilevel steganography model (for i = 3) is presented in Figure 1 and Figure 2.

At the sender end, in phase 1, secret message M is embedded in cover object C_1 using transformation T_1 and embedding function f_1 and stego-object I_1 is generated. In the next phase, stego-object I_1 is hidden in another new cover object C_2 using transformation T_2 and embedding function f_2 and stego-object I_2 is generated. This process is continued as per the requirement of the application. There are three level of embedding process is shown in Figure 1.

At the receiver end, according to the above 2 level embedding process, in phase 1, stego-object I_1 is generated from stego-object I_2 and applying T_2' transformation and f_2' embedding function. In the next phase, secret message M is generated from stego-object I_1 by applying T_1' transformation and f_1' embedding function. There are three level of extraction process shown in Figure 2.

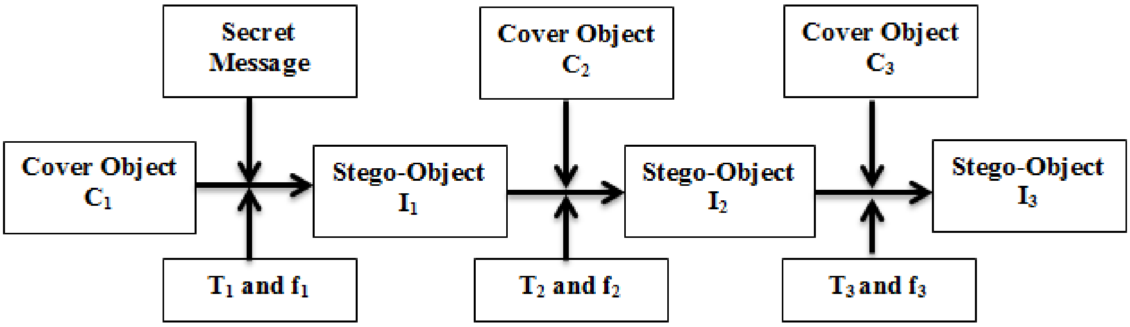

Figure .1: TYPE-I Multilevel Steganography model at the sender end for Level = 3

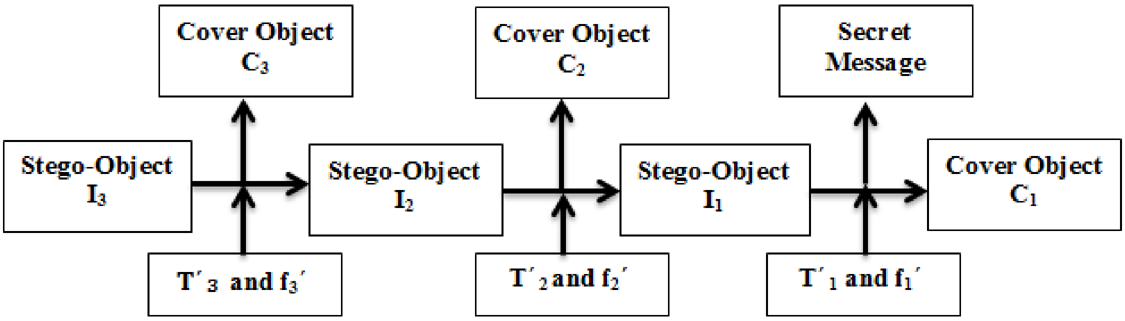

Figure 2: TYPE-I Multilevel Steganography model at the receiver end for Level = 3

2.1.1 Example of TYPE-I: 2 level steganography

Secret Message Embedding Process:

Level-1: for i = 1

The cover is a grayscale image (C_1) and Message is a text message (M). Here, transmission T_1 is an encryption process. That means, secret message is encrypted using some standard encryption algorithm. The encrypted secret message bits are embedded at the 2^{nd} LSB position of each pixel value of the cover image. The embedding function f_1 is defined as f_1(mbit) = C_1.LSB(2) and f_1 is used to generate Intermediate cover or stego-cover I_1.

Leve-2: for i = 2

In this step, the cover is an audio signal (C_2) and Intermediate Cover or stego-cover is I_1. I_1 is generated in the previous step and it is an embedded image. The image is converted to a bit stream and each bit is embedded at the 1^{st} LSB position of each audio sample of the audio signal. Here, transformation T_2 = I and the embedding function f_2 is defined as f_2(ibit) = C_2.LSB(1) and f_2 is used to generate Intermediate cover or stego-cover I_2.

Secret Message Extraction Process:

Leve-2: for i = 2

The Intermediate Cover (I_2) is an embedded audio signal and the embedded image bits are extracted from the 1^{st} LSB position of each audio sample. Here, transformation T_2’ = I and the extracting function f_2’ is defined as f_2’(ibit) = I_2.LSB(1) and f_2’is used to generate Intermediate cover or stego-cover I_1.

Level-1: for i = 1

The Intermediate Cover (I_1) is an embedded image and the message bits are extracted from the 2^{nd} LSB position of each pixel value of the embedded image. Here, transmission T_1’ is a decryption process of the corresponding encryption algorithm used during embedding process. The extraction function f_1’ is defined as f_1’(mbit) = I_1.LSB(2) and f_1’is used to generate secret message M.

2.2 Single cover multiple messages - multilevel steganography model (TYPE-II):

Suppose, the Cover is denoted as C, the Messages are denoted as M_i, the Intermediate Covers or stego-covers are denoted as I_i. Here, the value of i depends on the level of steganography is to be performed.

The messages M_i are passed through the transformation T_i as like previous section. The TYPE-II multilevel steganography model (for i = 3) is presented in Figure 3 and Figure 4.

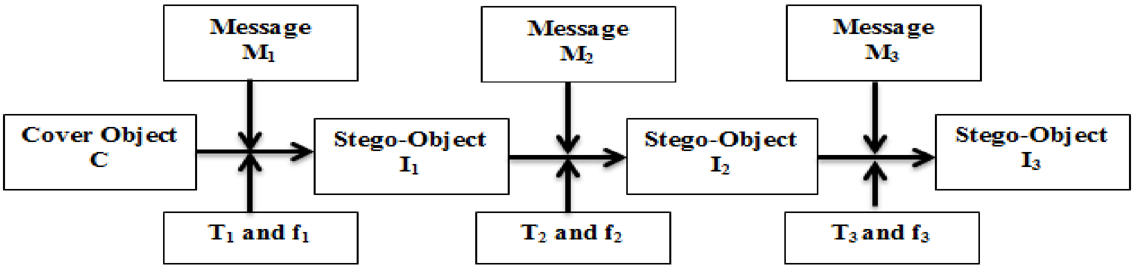

Figure 3: TYPE-II Multilevel Steganography model at the sender end for Level = 3

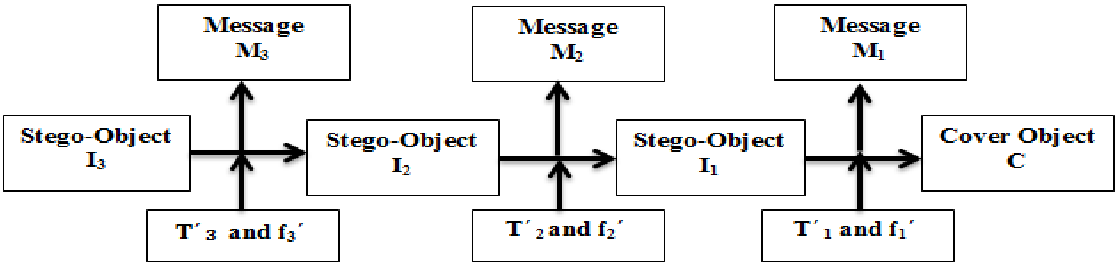

Figure 4: TYPE-II Multilevel Steganography model at the receiver end for Level = 3

At the sender end, in phase 1, secret message M_1 is embedded in cover object C using transformation T_1 and embedding function f_1 and stego-object I_1 is generated. In the next phase, another message M_2 is hidden in stego-object I_1 using transformation T_2 and embedding function f_2 and stego-object I_2 is generated. This process is continued as per the requirement of the application. There are three level of embedding process is shown in Figure 3.

At the receiver end, according to the above 2 level embedding process, in phase 1, message M_2 is generated from stego-object I_2 and applying T_2' transformation and f_2' embedding function. In the next phase, secret message M_1 is generated from stego-object I_1 by applying T_1' transformation and f_1' embedding function. There are three level of extraction process shown in Figure 4.

2.2.1 Example of TYPE-II: 2 level steganography

Secret Message Embedding Process:

Level-1: for i = 1

The cover is an audio clip (C) and the two secret messages are M_1 and M_2. Here, transmission T_1 is Discrete Wavelet Transform (DWT) and Inverse DWT (IDWT) of audio signal. The M_1 message bits are embedded at the 2nd LSB position of each DWT coefficient of the audio signal. The embedding function f_1 is defined as f_1(mbit) = C.LSB(2) and f_1 and IDWT are used to generate Intermediate cover or stego-cover I_1.

Leve-2: for i = 2

In this step, the cover is an Intermediate Cover or stego-cover (I_1). I_1 is generated in the previous step and it is an embedded audio signal. The M_2 message bits are embedded at the 1st LSB position of each audio sample of the audio signal. Here, transformation T_2 = I and the embedding function f_2 is defined as f_2(mbit) = C.LSB(1) and f_2 is used to generate Intermediate cover or stego-cover I_2.

Secret Message Extracting Process:

Leve-2: for i = 2

The Intermediate Cover (I_2) is an embedded audio signal and the M_2 message bits are extracted from the 1st LSB position of each audio sample. Here, transformation T_2' = I and the extracting function f_2' is defined as f_2'(abit) = I_2.LSB(1) and f_2'is used to generate Intermediate cover or stego-cover I_1.

Level-1: for i = 1

The Intermediate Cover (I_1) is an embedded audio and the message bits are extracted from the 2nd LSB position of each DWT coefficient of the embedded audio signal. Here, transmission T_1' is DWT and IDWT of the embedded audio signal. The extraction function f_1' is defined as f_1'(abit) = I_1.LSB(2) and f_1'is used to generate secret message M_1.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Proposed algorithm has been tested on 10 audio sequences from different music styles (classic, jazz, country, pop, rock, etc.). All the Clips are 44.1 kHz sampled mono audio files, represented by 16 bits per sample, and length of the clips ranged from 10 to 20 seconds. An image and the all audio clips are used to test TYPE-I type of algorithm and all the audio clips are used to test TYPE-II algorithm.

3.1 Imperceptibility Test

The main basic requirement is the imperceptibility in most of the applications. That means, after hiding secret messages in audio signals, quality of the embedded audio signals should remain same as original audio signals. The Subjective Difference Grade (SDG), Objective Difference Grade (ODG) and Signal-to-Noise Ratio (SNR) is used to measure the imperceptibility of the proposed method. The SDG and ODG listening tests use the 5-grade scale shown in Table 1.

Table 1: Subjective and objective grades for audio quality measurement

Audio quality	**Subjective difference grade (SDG)**	**Objective difference grade (ODG)**
Imperceptible	5	0.0
Perceptible, but not annoying	4	-1.0
Slightly annoying	3	-2.0
Annoying	2	-3.0
Very annoying	1	-4.0

3.1.1 Objective Quality Measurements

Objective Difference Grade (ODG) is a suitable measurement of audio quality, since it is assumed to provide a precise model of the Subjective Difference Grade (SDG) results that may be obtained by listening tests of a group of expert listeners. In this work, the ODG measurements of different audio clips are provided using the advanced ITU-R BS.1387 standard [10] and calculated using the Opera software [11] which is implemented by maintaining ITU-R BS.1387 standard. ODG values for TYPE-I and TYPE-II approaches are reported in Table 2 and Table 3 respectively for different types of audio signals.

3.1.2 Subjective Quality Evaluation

Subjective quality measurements [12, 13] have been performed to evaluate the imperceptibility of our proposed data hiding scheme. The ten participants were nominated for these subjective listening tests, five of them were experts in music and the rest of the five was general listeners. All of the participants are presented with the original and the embedded digital audio signals and were asked to report any difference between them, using five-points SDG as given in Table 1. The output of the subjective tests is an average of the quality ratings called a Mean Opinion Score (MOS). SDG values for TYPE-I and TYPE-II approaches are reported in Table 2 and Table 3 respectively for different types of audio signals.

3.1.3 Signal-to-Noise Ratio (SNR) Measurement

The signal-to-noise ratio (SNR) value is used to make the difference between the original and embedded audio signal [14]. Normally, if the SNR value is higher than 50 dB, then the secret data which are hidden in the audio signal are imperceptible to the human auditory system. The SNR values are calculated using equation no. (1) for different embedded audio signals. The original audio signal is denoted x(i), i = 1 to N while the stego audio signal is denoted as y(i), i = 1 to N.

$$\mathrm{SNR} = 10 \log_{10} \frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N} \big(x(i) - y(i)\big)^2} \;\ldots\ldots(1)$$

The ODG, SDG, and SNR values are evaluated for different audio signals. Here, 2 level multilevel steganography is performed for TYPE-I and TYPE-II models. The results are presented in Table 2 and Table 3 of TYPE-I and TYPE-II models respectively. For simplicity, 10 audio clips are denoted as A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9 and A_{10}.

Table 2: ODG, SDG & SNR values for different audio clips (TYPE-I, Level = 2)

Audio Types	**Objective Difference Grade (ODG)**	**Subjective Difference Grade (SDG)**	**Signal-to-Noise Ratio (SNR(dB))**
A_1	-0.52	4.9	90.15
A_2	-0.72	4.8	89.31
A_3	-0.70	4.8	88.93
A_4	-0.49	4.9	90.41
A_5	-0.50	4.9	90.22
A_6	-0.51	4.9	90.16
A_7	-0.68	4.8	89.25
A_8	-0.69	4.8	89.36
A_9	-0.52	4.9	90.19
A_{10}	-0.73	4.8	89.14

Table 3: ODG, SDG & SNR values for different audio clips (TYPE-II, Level = 2)

Audio Types	Objective Difference Grade (ODG)	Subjective Difference Grade (SDG)	Signal-to-Noise Ratio (SNR(dB))
A_1	-0.51	5.0	92.25
A_2	-0.63	4.9	91.41
A_3	-0.61	4.9	91.63
A_4	-0.52	5.0	92.43
A_5	-0.49	5.0	92.54
A_6	-0.50	5.0	92.35
A_7	-0.64	4.9	91.55
A_8	-0.59	4.9	91.57
A_9	-0.49	5.0	92.36
A_{10}	-0.53	4.9	91.78

3.2 Embedding Capacity Analysis:

One of the basic requirements of the secret communication using steganography is increasing the embedding capacity by keeping the imperceptibility in a desired level. In the proposed system, if TYPE-II approach is followed, multiple messages may be embedded in a single cover object by designing appropriate transforms and embedding functions.

3.3 Security Analysis:

Security is another very important requirement of hidden communication using steganography. In TYPE-I approach, a message is hidden in a cover object and that stego-cover object is hidden in another cover object, and so on. This approach increases the level of security of the system. Again, number of level is used during the embedding process in multilevel steganography is very important information at the receiving end. That means, security may be increased by varying the number of levels during embedding process. Along with this, any of the encryption algorithms may be used at a transformation phase of the system to increase the security of the system.

3.4 Comparative Study:

In this section, a comparative study is performed with the very recent works on audio steganography as well as audio watermarking proposed by different authors. Actually, impartial comparison is very difficult, because every approach have its own characteristics and also designed to fulfill certain basic requirement. Anyway, most of the algorithm have some common characteristics like embedding capacity, imperceptibility etc. Here, comparisons are performed based on embedding capacity and imperceptibility (SNR & ODG) of the system and reported in Table 4.

Table 4: Comparative studies among different works

Algorithm	Capacity (bps)	SNR (dB)	ODG
[15]	2	42.8 to 44.4	-1.66 <ODG<-1.88
[16]	4.3	29.5	Not reported
[17]	3k	30.55	-0.6
[18]	2k-6k	Not reported	-0.6 < ODG <-1.7
[19]	11 k	30	-0.7
[20]	64	30-45	-1< ODG
[21]	4-512	Not reported	-1 < ODG
[22]	8	Not reported	-3 < ODG < -1
Proposed	44100	92.54	-0.49 < ODG < -0.64

IV. CONCLUSION

In this work, two multilevel steganography models are proposed. Normally, requirement of data hiding application are varied from application to application. The proposed models are designed such a way that the customization may be done as per the requirement of a particular application. That means, number of embedding and extracting levels, number of messages to be hidden, and number of cover objects to be used etc. are customizable. The suggested model enhances the security level of the steganography technique. The stego-object usually does not seem suspicion, since it looks similar to the original object for the general observer. An adversary may be satisfied with the decoy as the hidden message and may not use additional tools to look further. The authorized receivers have information about the hidden message, as well as the information required to extract the message. Hence, it can be concluded that the proposed models enhanced potentially more security to information hiding.

REFERENCES

[1] A. J. Al-Najjar. The Decoy: Multi-Level Digital Multimedia Steganography Model. In Proc. of 12th WSEAS International Conference on Communications, Heraklion, Greece, July 23-25, 2008.

[2] R. J. Anderson, F. A. P. Petitcolas. On the Limits of the Steganography. IEEE Journal of Selected Areas in Communications. 16: 474-481, 1998.

[3] D. Artz. Digital Steganography: Hiding Data within Data. IEEE Internet Computing. pages 75-80, May-June 2001.

[4] D. Vitaliev. Digital Security and Privacy for Human Rights Defenders. The International Foundation for Human Right Defenders. Pages 77-81, Feb. 2007.

[5] F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn. Information Hiding - A Survey. Proceedings of the IEEE, special issue on protection of multimedia content, 87:1062-1078, July 1999.

[6] A. J. Al-Najjar, A. K. Alvi, S. U. Idrees, A. M. Al-Manea. Hiding Encrypted Speech Using Steganography. MIV'07, WSEAS 2007, pages 275-281, Sept. 15-17, Beijing, China, 2007.

[7] L. M. Marvel. Image Steganography for Hidden Communication, Ph.D. Dissertation, University of Delaware, Spring 1999.

[8] M. Solanki. Multimedia Data Hiding: From Fundamental Issues to Practical Techniques, Ph.D. Dissertation, University of California, Santa Barbara, US, December 2005.

[9] D. C. Lou, M. C. Hu, J. L. Liu. Multiple layer data hiding scheme for medical images. Computer Standards and Interfaces 31: 329–335, 2009.

[10] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten. PEAQ - The ITU standard for objective measurement of perceived audio quality. J. AES. 48: 3–29, 2000.

[11] OPTICOM OPERA software site, [Online]. Available: http://www.opticom.de/products/opera.html

[12] M. Unoki, K. Imabeppu, D. Hamada, A. Haniu, and R. Miyauchi. Embedding limitations with digital-audio watermarking method based on cochlear delay characteristics. J. Inf. Hiding Multimedia Signal Process. 2: 1–23, 2011.

[13] S. Wang and M. Unoki. Speech watermarking method based on formant tuning. IEICE Trans. Inf. Syst. E98-D: 29–37, 2015.

[14] S. R. Quackenbush, T. P. Barnwell III, and M. A. Clements. Objective Measures of Speech Quality. Prentice Hall, Englewood Cliffs, 1988.

[15] S. Xiang, H. J. Kim, and J. Huang. Audio watermarking robust against time-scale modification and mp3 compression. Signal Process. 88: 2372–2387, Oct. 2008.

[16] M. Mansour and A. Tewfik. Data embedding in audio using time-scale modification. IEEE Trans. Speech Audio Process. 13: 432–440, May 2005.

[17] M. Fallahpour and D. Megías. High capacity audio watermarking using fft amplitude interpolation. IEICE Electron. Express, 6:1057–1063, 2009.

[18] M. Fallahpour and D. Megías. High capacity method for real-time audio data hiding using the fft transform. in Advances in Information Security and Its Application. Berlin, Germany: Springer-Verlag, pages 91–97, 2009.

[19] M. Fallahpour and D. Megías. High capacity audio watermarking using the high frequency band of the wavelet domain. in Multimedia Tools and Applications. New York, NY, USA: Springer, vol. 52, pages 485–498, 2011.

[20] X. Kang, R. Yang, and J. Huang. Geometric invariant audio watermarking based on an LCM feature. IEEE Trans. Multimedia, 13: 181–190, Apr. 2011.

[21] M. Unoki and D. Hamada. Method of digital-audio watermarking based on cochlear delay characteristics. Int. J. Innovat. Comput., Inf. Control, 6:1325–1346, Mar. 2010.

[22] A. Nishimura. Audio data hiding that is robust with respect to aerial transmission and speech codecs. Int. J. Innovat. Comput., Inf. Control, 6: 1389–1400, Mar. 2010.