DATA HIDING IN AUDIO SIGNALS USING WAVELET TRANSFORM WITH ENHANCED SECURITY

Deepthi S.[1] ,Renuka A.[2] and Hemalatha S.[3]

[1]Department of Computer Science and Engineering,Manipal Institute of Technology
deepthi.s@manipal.edu
[2]Department of Computer Science and Engineering,Manipal Institute of Technology
renuka.prabhu@manipal.edu
[3]Department of Computer Science and Engineering,Manipal Institute of Technology
hema.shama@manipal.edu

ABSTRACT

Rapid increase in data transmission over internet results in emphasis on information security. Audio steganography is used for secure transmission of secret data with audio signal as the carrier. In the proposed method, cover audio file is transformed from space domain to wavelet domain using lifting scheme, leading to secure data hiding. Text message is encrypted using dynamic encryption algorithm. Cipher text is then hidden in wavelet coefficients of cover audio signal. Signal to Noise Ratio (SNR) and Squared Pearson Correlation Coefficient (SPCC) values are computed to judge the quality of the stego audio signal. Results show that stego audio signal is perceptually indistinguishable from the cover audio signal. Stego audio signal is robust even in presence of external noise. Proposed method provides secure and least error data extraction.

KEYWORDS

Audio steganography, Lifting wavelet transform, Mean square Error (MSE), SNR & SPCC

1. INTRODUCTION

During transmission of confidential information, malicious user can illegally copy, modify or destroy the information being conveyed on internet. As a result, information security becomes a vital issue. Certain degree of security can be achieved by using cryptographic techniques. But, the resultant cipher text, which appears to be gibberish, may attract attackers more than the normal text. Also cipher text is easy to be detected and also many techniques already exist to crack it. Steganography is also information security technique which hides secret information within a normal carrier media, such as digital image, audio, video, etc. [1]. In Steganography, the secret message is cloaked to hide the existence and is made "invisible", thus concealing the fact that message is sent altogether, only the sender and the authorized recipient can detect the presence of secret information. With Steganography, people can send messages without anyone having

Natarajan Meghanathan et al. (Eds) : ITCSE, ICDIP, ICAIT - 2013
pp. 137–146, 2013. © CS & IT-CSCP 2013 DOI : 10.5121/csit.2013.3914

knowledge of the existence of the communication. Steganography can be a solution which makes it possible to send news and information without being censored and without the fear of the messages being intercepted and traced back. Steganography becomes more important as more people join cyberspace revolution [2]. In audio steganography, the weak point of Human Auditory System (HAS) is used to hide information in the audio. Because the human auditory system has more accurate than Human Visual System (HVS), audio steganography is more challenging than image steganography. Audio file used to hide the secret message is called as cover audio and once secret data is embedded into cover audio, resultant audio is called as stego audio. The three important parameters in designing steganography method are perceptual transparency, hiding capacity and robustness. The hidden information is imperceptible if a listener is unable to single out between the cover- and the stego-audio signal. Hiding capacity refers to the amount of obscured data (in bits) within a cover audio signal. The robustness criteria are assessed through the survival of concealed data against noise and manipulations of the audio signal [3]. Three prominent data embedding approaches have been investigated, namely hiding in temporal domain, in Transform domains and in coded domain. Out of these wavelet transform provides more security and robustness than the other approaches.

1.1 Wavelet transformation of audio signal

The wavelet transform (WT) has gained widespread acceptance in signal processing and image compression. Wavelet transform is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet [4]. A wavelet is a waveform of effectively limited duration that has an average value of zero. For signals; identity of the signal is given by the low-frequency component. The high-frequency content only imparts savour or nuance. In human voice, if high-frequency components are removed, the voice sounds different, but still it can be understood. If low frequency components are removed, signal sounds gabble. On applying wavelet transformations on audio signal, approximation and detail components of audio can be obtained. The approximations are low-frequency components of the signal and details are high-frequency components. The first level detail coefficients have less importance in comparison with detail coefficients of next levels and approximation coefficients because of their low energy level. Figure 1 shows the decomposition of audio signal on wavelet transform.

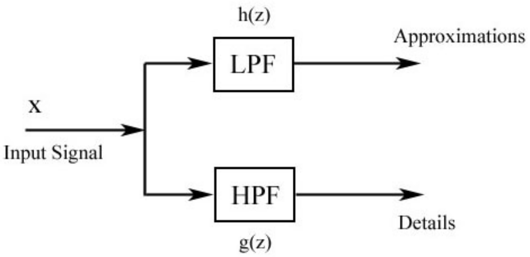


Figure 1. One stage signal decomposition

1.2 Lifting Wavelet Transform (LWT)

The lifting scheme is a technique for both designing wavelets and performing the Discrete Wavelet Transform (DWT) [5].The problem with DWT is that when applying it on an integer signal, the resulted coefficients are not integers. If the algorithm needs to access the binary value of the resultant coefficients, then conversion of coefficients from floating to binary will require to scale and then convert them to a binary. To solve this problem, lifting scheme can be used to produce integer to integer wavelets. In this, the resultant coefficients are integers for all integer signals. This eliminates the need for scaling the coefficients and converting them to binary representation. The main sources of errors arise during this conversion such as rounding errors and out of range errors, do not occur.

Section 2 will brief on some of related works which has been done on audio steganography. Section 3 will explain the methodology. Section 4 contains experimental results and analysis. The Final section will brief on conclusion.

2. LITERATURE SURVEY

In recent years, several researchers have concentrated on developing algorithms for hiding data in an audio signal. Jisna Antony et al [6] discuss about different audio steganographic techniques available in different domain. Lots of work is done in all domains. Baritha and Venkataramani [7] propose a new dictionary based text compression technique. Dictionary based compression bits are hidden into the LSB bit of audio signals. In this secret text is hidden using an identifier. Identifier along with length and width are hidden inside audio. Identifier indicates whether there is secret text hidden or not. This paper is implemented in temporal domain. Ahmad Delforouzi and Mohammad Pooyan [8] proposes an algorithm which embeds secret data in temporal domain. In this algorithm first embedding threshold in the time domain is estimated. Then this threshold is used for data concealment in the time domain. Drawback of audio steganography in temporal domain is even though it is easy to hide data; it lacks security as well as has less hiding capacity compared to hiding in wavelet domain. Dora and Juan [9] proposes a new scheme of data hiding which takes advantage of the masking property of the Human Auditory System (HAS) to hide a secret (speech) signal into a host (speech) signal. In embedding process, wavelet coefficients of the secret signal are sorted and embedded in the wavelet coefficients of the host signal. And their original positions are used as key. Delay is inserted in each cycle to achieve synchronization. This approach consumes more time; retrieved secret signal is not same as the original because there is error in reconstruction of host signal. Also as there is need to store the positions of frames in stego signal, it reduces the hiding capacity of the host signal. Yongfeng Huang et al [10], proposed an algorithm which performs data embedding while pitch period prediction is conducted. Embedding the secret data is done during low bit-rate speech encoding. Drawback of this technique is that stego audio has been detected in steganalysis. Parul Shah et.al, [11] developed an algorithm where modification of host audio is done by imposing a constraint which forces the modified value to be in the same range as its neighbourhood. In this paper, host signal is decomposed using wavelet packet up to third level, then selected band coefficients are sampled and then converted to 2D. This 2D matrix is then divided to 2X2 non overlapping blocks used to embed covert data using pseudorandom sequence. Secret data is embedded into host based on trend mapping. Sajad Shirali-Shahreza and M.T. Manzuri-Shalmani [12] developed an audio steganography algorithm to hide text which uses lifting scheme to create perfect reconstruction. In this secret data is stored based on the details coefficient value. To calculate the number of bits to hold data in a coefficient with value 'c', the biggest power of 2 named 'p' which is smaller than 'c' i.e., $2p\leq c <2p+1$ is found out. The number of bits used to hide in this coefficient is p – OBH where Original Bit to Hold (OBH) is a constant which shows how many bits of the original signal is kept unchanged so that stego audio is imperceptible and how many bits of the signal are replaced with the data. Ahmad Delforouzi [13] describes an algorithm where LWT is applied on host audio signal. Host audio signal is decomposed to fifth level and sub bands are used to hide the secret data using the threshold calculated. Drawback of this algorithm is that threshold value calculation.

3. PROPOSED METHOD

This section discusses the algorithm used to hide encrypted text in cover audio signal. Algorithm has two phases – embedding and extraction. In embedding phase, encrypted text is hidden inside the cover audio signal. It should be made sure that there should not be any distortion in the cover audio by hiding the secret data. In extraction phase, the secret text is retrieved from the stego-audio. In this algorithm, audio samples are transformed into wavelet domain. Secret data here is text, which is encrypted using dynamic encryption algorithm. These transformed values of text are then hidden in LSB's of detail coefficients.

3.1 Embedding Phase

Step 1: Audio Processing

Read the cover audio file. Audio samples are stored in a vector and are signed floating point values. When an audio signal is transformed to another domain, then changed back to time domain, the resulting signal is not necessarily integer. In order to get integer coefficients from audio samples, audio samples must be converted to integers. This conversion is performed here.

Step 2: Apply LWT based on lifting scheme

This algorithm uses integer to integer transformation which is implemented using lifting wavelet transformation (LWT). LWT uses Lifting Scheme (LS). In LS, among the various wavelets available, appropriate wavelet is chosen. As integer coefficients are required, 'int2int' transformation has to be specified. Based on the LS, apply the LWT to cover audio to get detail and approximation coefficients, CD and CA respectively. Convert CD to binary.

Step 3: Calculate number of bits to be replaced.

The number of bits used to hold data is calculated using the logic explained by Sajad Shirali-Shahreza and M.T. Manzuri-Shalmani [12]. This algorithm chooses dynamic approach to find the bits to hold the secret text. Detail coefficients are selected to hold the secret text. Number of bits of CD to be replaced (NBR) is based on the fact that if the coefficient value is more, then changing more bits will not cause major difference in the signal. So, more secret data bits are hidden in bigger coefficients and fewer in smaller coefficients.

Step 4: Read text file and encrypt it.

Read the text file. Find the size of the text to be hidden. The text is encrypted by subtracting ASCII value of each character by message size. Cipher text is then converted to binary string. Reason behind implementing simple encryption technique using message size is that there is no need to hide the encryption key in cover audio. This allows more data to be hidden and also provides security without reducing the hiding capacity.

Step 5: Embed the encrypted text

This step is sub divided into two parts: Hiding the size of the text and hiding the actual text. "Text" is also referred as "message" in rest of the paper.

- Hide message size

It is necessary to embed the secret message size into the cover audio because during extraction of text from stego-audio, receiver should know how many bits have to be extracted from stego-audio. Also, receiver has to decrypt the secret text based on message size. First 16 replaceable LSB's are reserved to store the message size, based on NBR calculated for each CD value. Message size bits, starting from MSB, are stored in LSB's of CD's.

- Hide the actual message

Remaining replaceable bits of each CD are used to store the encrypted secret message. Message bits starting from MSB, are stored in LSB's of each CD.

Step 6: Reconstruction of stego-audio signal.
After embedding the secret message into CD; using CA and modified CD, stego- audio signal is reconstructed by applying inverse LWT. This stego-audio sounds same as the cover audio.
Figure 2 shows the embedding process.

3.2 Extraction Phase

Step 1: Audio Processing
Read the stego audio file. Then convert the audio samples into integers. This step is same as step 1 in the embedding phase.

Step 2: Apply LWT based on lifting scheme
Select the same lifting scheme which is used in the embedding phase. Based on this LS, apply the LWT to cover audio to get detail and approximation coefficients, CD and CA respectively. Convert CD to binary.

Step 3: Calculate number of bits to be replaced
This is exactly same as step 3 in the embedding phase. Use same OBH value that is used in the embedding phase.

Step 4: Extract the hidden message
This step is sub divided into two parts: retrieve the size of the message and retrieve the message.

- Retrieve message size
Based on NBR calculated for each CD in step 3, 16 bits are retrieved from LSB's of CD's to obtain the message size.
- Retrieve the message
Encrypted message bits starting from MSB, are retrieved from LSB's of remaining CD's using the message size and store it in a buffer.

Step 5: Decryption of message and writing it into a file
After retrieving, the encrypted secret message bits are converted to decimal. Resultant message bits are then decrypted using extracted message size and converted to character. This is again written to an output text file. Figure 3 depicts the extraction phase.

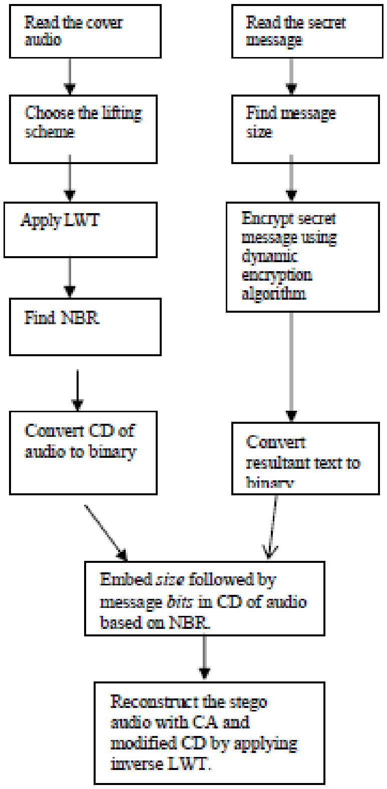

Read the cover audio
Choose the lifting scheme
Apply LWT
Find NBR
Convert CD of audio to binary
Read the secret message
Find message size
Encrypt secret message using dynamic encryption algorithm
Convert resultant text to binary
Embed size followed by message bits in CD of audio based on NBR.
Reconstruct the stego audio with CA and modified CD by applying inverse LWT.

Figure 2. Embedding Phase

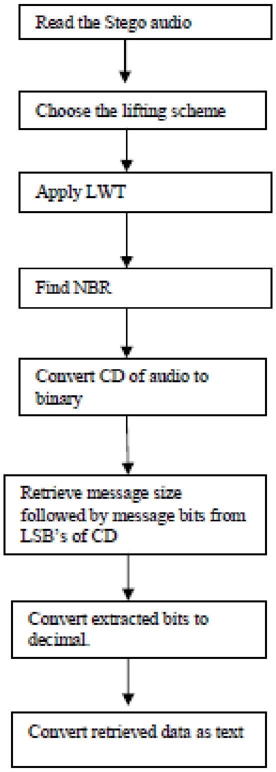

Read the Stego audio
Choose the lifting scheme
Apply LWT
Find NBR
Convert CD of audio to binary
Retrieve message size followed by message bits from LSB's of CD
Convert extracted bits to decimal.
Convert retrieved data as text

Figure 3. Extraction Phase

4. Experimental Results

This section focuses on the experimental results. Results for different audio files with different amount of data hidden are shown. Quality of the stego-audio is analyzed using MSE and SNR. MSE serves as an important parameter in gauging the performance of the steganographic system. Suppose that x = {xi | i = 1, 2. . . N} and y = {yi | i = 1, 2. . . N} are two finite-length, discrete signals, for e.g., visual images or audio signals. Then MSE between the signals is given by equation (4).

$$MSE(x,y) = \frac{1}{N}\sum_{i=1}^{N}(xi - yi)^2 \qquad (4)$$

where,

N is the number of signal samples.
xi is the value of the ith sample in x.
yi is the value of the ith sample in y.

SNR is a term that refers to the measurement of the level of an audio signal as compared to the level of noise that is present in that signal. It is expressed in decibels (dB). A larger SNR value indicates a better quality. It is given by equation (5).

$$\text{SNR} = 10^{\log_{10}\left(\frac{\frac{1}{N}\sum_{i=0}^{N} xi^2}{MSE}\right)} \qquad (5)$$

Recommended SNR for audio signal is above 30dB.

Another metric based on correlation of samples is Squared Pearson Correlation Coefficient (SPCC). The higher the SPCC, the better is the quality of the output signal. Its range is between 0 and 1. It is given by equation (6).

$$SPCC = \left[\sum \frac{(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}\right]^2 \qquad (6)$$

where x, y, $\bar{x}$ and $\bar{y}$ are the cover signal, stego signal, average of the cover signal and average of the stego signal, respectively.

The algorithm is implemented in MATLAB 11 on an Intel core 2 Duo CPU at 2.00 GHz with 2.00 GB RAM. Four audio files Two.wav with 276347 samples, Woody2.wav with 37620 samples, b.wav with 77175 samples and 1.wav with 69860 samples are considered as cover audio signals. Four text files Test1.txt consisting of 22 characters, Test2.txt consisting of 2739 characters, Test3.txt consisting of 4328 characters, Test4.txt consisting of 5 characters and Test5.txt with 5737 characters are considered as secret messages to test the algorithm. Table 1 shows MSE, SNR and SPCC values of various audio signals when different amount of secret message is hidden with OBH equal to 1. It is observed from Table 1 that SNR decreases and MSE increases, as the hiding capacity is increased. Maximum hiding capacity of any cover audio signal depends upon the sample values. Wavelet used is “db2”. White Gaussian noise with different SNR values is added to stego audio signal. Secret data is able to be retrieved without any errors. This is used to check the robustness of the algorithm. Experiment is conducted with other wavelets as well; there is no significance change in the results.

TABLE 1. MSE, SNR and SPCC values for different hiding capacities

Cover Audio	**Text file**	**Number of characters**	**MSE**	**SNR(dB)**	**SPCC**
Two.wav	Test1.txt	22	0.0069	63.79	0.9432
Woody2.wav	Test1.txt	22	0.0083	63.36	0.9385
Two.wav	Test2.txt	2739	0.9975	42.20	0.9058
Woody2.wav	Test2.txt	2739	2.51	38.57	0.9001
Two.wav	Test3.txt	4328	1.5432	40.30	0.9039
Woody2.wav	Test5.txt	5737	5.2778	35.34	0.8982
b.wav	Test4.txt	5	0.0057	67.56	0.9835
b.wav	Test1.txt	22	0.0086	66.79	0.9832
1.wav	Test4.txt	5	0.0016	73.35	0.9895
1.wav	Test1.wav	22	0.0086	65.79	0.9751

Experiment is also conducted with different OBH values. Results of embedding Test1.txt are tabulated in Table 2.

TABLE 2. MSE and SNR(dB) values with different OBH

Audio	**OBH=2**			**OBH=4**		
	MSE	**SNR**	**SPCC**	**MSE**	**SNR**	**SPCC**
Two.wav	0.0035	66.71	0.9851	0.0021	67.90	0.9897
Woody2.wav	0.0056	65.08	0.9765	0.0043	66.25	0.9855

From Table 2, it is clear that if more number of bits are replaced, hiding capacity increases but MSE increases and also audio quality is degraded.

Figures 4, 6 and 8 shows the cover, stego and stego with noise audio with encrypted text being hidden in woody2.wav, respectively. It can be observed that significant changes are not perceptible. Figure 5 shows the secret message to be hidden. It shows the original secret message and encrypted message. Figure 7 shows the output of extraction phase, it shows the received encrypted message and decrypted message from stego-audio.

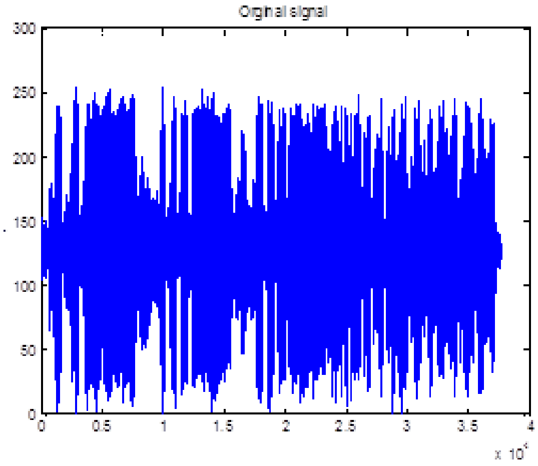
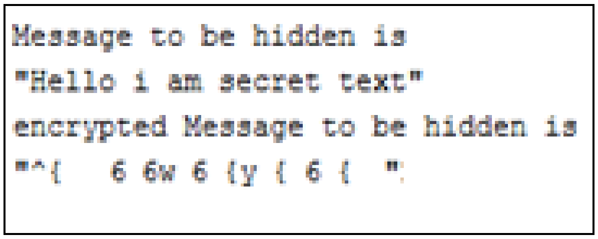


Figure 4. cover audio – woody2.wav



```
Message to be hidden is
"Hello i am secret text"
encrypted Message to be hidden is
"^{   6 6w 6 {y { 6 {  ".
```

Figure 5. Test1.txt embedded in Two.wav

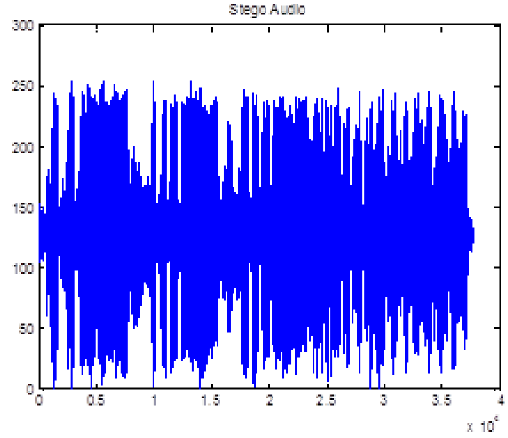
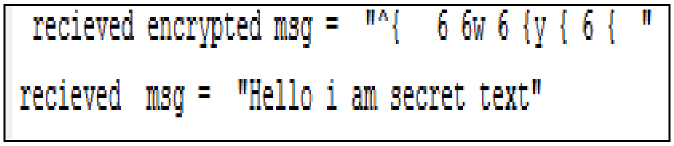


Figure 6 stego-audio



```
 recieved encrypted msg =  "^{   6 6w 6 {y { 6 {  "
recieved  msg =  "Hello i am secret text"
```

Figure 7. Retrieved secret text

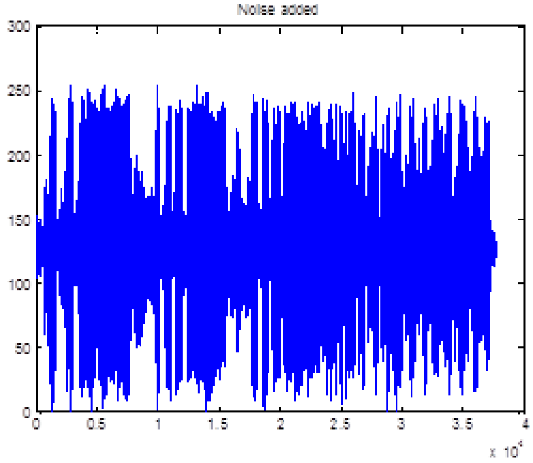


Figure 8. Stego-audio with white Gaussian noise

Subjective tests for audio quality evaluation are also performed. Five listeners were presented with a set of audio clips containing six songs, two original and two stego and two stego audio added with white Gaussian noise, in a random order. For most of the cases, listener could not differentiate between the different between original and stego audio, i.e., noise was inaudible. These results show the proposed method does not degrade the audio quality for almost all the cases.

5. CONCLUSION

Objective of the paper is to hide encrypted text in cover audio using lifting wavelet transform. Based on the values of coefficients, number of bits used to hold secret data is chosen. In the proposed method, text is encrypted based on the message size and then hidden in cover audio. Results are computed and observed. This algorithm yields zero error extraction, good SNR and SPCC. Similar technique is used by Sajad Shirali-Shahreza and M.T. Manzuri-Shalmani [12] without encryption. There SPCC is not calculated, which is a good metric to test the audio quality based correlation. In the proposed method approximately same values of SNR and MSE are obtained as in [12] even with encryption and noise added. As the audio samples for even 30 secs audio file is in lakhs, processing it and hiding text and extracting it takes lot of time. This drawback can be eliminated by implementing in parallel using GPU's.

REFERENCES

[1] K. Ramani et al, "Steganography using BPCS to the integer wavelet transformed image", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.7, 2007, pp. 293- 302

[2] Fatiha Djebbar et al," Comparative study of digital audio steganography techniques", EURASIP Journal on Audio, Speech, and Music Processing 2012,pp no. 1192-1203

[3] Abbas Cheddad, "Digital image steganography: Survey and analysis of current methods",Signal Processing,Vol 90,Issue 3,March 2010,pp. 727-752

[4] Michael Weeks, "Digital Signal Processing Using MATLAB and Wavelets", Pearson publications, ISBN – 81-297-0272-X.

[5] Elham Ghasemi, Jamshid Shanbehzadeh and Bahram ZahirAzami, "A Steganographic method based on Integer Wavelet Transform and Genetic Algorithm", IEEE conference 2011.

[6] Jisna Antony and Sobin C," Audio Steganography in Wavelet Domain – A Survey", International Journal of Computer Applications, Volume 52, No.13, 2012, pp. 33-37

[7] M.Baritha Begum and Y.Venkataramani, "LSB Based Audio Steganography Based on Text Compression", International Conference on Communication Technology and System Design, 2011,pp. 703-710

[8] Ahmad Delforouzi and Mohammad Pooyan, "Adaptive and Efficient Audio Data Hiding Method in Temporal Domain", IEEE ICICS, 2009.

[9] Dora M. Ballesteros L and Juan M. Moreno A," Real-time, speech-in-speech hiding scheme based on least significant bit substitution and adaptive key", Computers and Electrical Engineering Vol 39, Elsevier, 2013,pp. 1192-1203

[10] Yongfeng Huang, Chenghao Liu and Shanyu Tang," Steganography Integration into a Low-Bit Rate Speech Codec", IEEE transactions on information forensics and security, vol. 7, no. 6, 2012

[11] P. Shah, P. Choudhari, and S. Sivaraman, "Adaptive wavelet packet based audio steganography using data history", IEEE Region 10 and the Third international Conference on Industrial and Information Systems, ICIIS, IEEE, 2008.

[12] S. Shahreza and M. Shalmani, "High capacity error free wavelet domain speech steganography", IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2008.

[13] Mohammad Pooyan, Ahmad Delforouzi, "Adaptive Digital Audio Steganography based on Integer Wavelet Transform", Circuits, Systems, Signal Process, 2008.