SECURE AUDIO STEGANOGRAPHY USING LIFTING WAVELET TRANSFORM

DEEPTHI S
Manipal Institute of Technology, Manipal – 576104, India
deepthi.s@manipal.edu

RENUKA A, HEMALATHA S AND JYOTHI UPADHYA K
Manipal Institute of Technology, Manipal-576104, India
{renuka.prabhu, hema.shama, jyothi.k} @manipal.edu

ABSTRACT: Due to rapid increase in digital data transmission over internet, information security becomes a vital issue. Audio steganography is one of the ways used for information security over the internet. In audio steganography, the secret data is transmitted with audio signal as the carrier. In this paper, an algorithm for hiding image data in an audio signal is proposed. The cover audio signal is transformed from time domain to wavelet domain using lifting wavelet transformation. Image is encrypted using Arnold transformation. Transformed image is then hidden in wavelet coefficients of cover audio. The quality of the resulting stego audio signal is tested by computing Squared Pearson Correlation Coefficient (SPCC) and SNR values. And the extracted image quality is tested by computing Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM) and Universal Image Quality Index (UIQI). Results show that stego audio is perceptually indistinguishable from the cover audio. Cover audio is robust even in presence of external noise. It provides high capacity, secure data extraction.

KEYWORDS: Audio, steganography, arnold transformation, Lifting wavelet transform, SNR, SPCC, PSNR, SSIM, UIQI

INTRODUCTION

Steganography is one of the information security techniques. Information security is the practice of defending information from illegal access, use, disclosure, disruption, modification, perusal, recording or destruction [1]. The rapid blowouts in digital data usage in many real life applications have urged new and effective ways to ensure their security. Mainly there are three ways of securing data during transmission - Cryptography, Steganography and Digital watermarking. Efficient confidentiality can be achieved, at least in slice, by implementing steganography techniques. The goal of steganographic systems is to obtain secure and robust way to conceal high rate of secret data. The primary goal of steganography is to reliably send hidden information secretly, not merely to obscure its presence. Audio steganography is a technique where secret data is hidden in an innocuous audio file and is sent over internet. This prevents the transmission of secret data to be unseen by unauthorized people. Modern techniques of steganography exploit the characteristics of digital media by utilizing them as carriers (covers) to hold hidden information. Cover medium can be of different types including image, audio, video and text. To minimize the difference between the cover- and the stego-medium, recent steganography techniques utilize natural limitations in human auditory and visual perception system. Audio steganography exploits the masking effect property of the Human Auditory System (HAS).

Wavelet transformation of audio

For many signals; the low-frequency content is the most important part and it gives the signal its identity [2]. On the other hand, the high-frequency content imparts flavor or nuance. In wavelet analysis, approximations and details are often dealt with. The approximations are the low-frequency components of the signal. The details are the high-frequency components. Proposed method uses detail coefficients of the audio signal for hiding secret image. Discrete Wavelet Transform (DWT) is used for digital signals. The problem which wavelet domain steganography using DWT, has is that when applying them on an integer signal, the resulted coefficients are not integer. To solve this problem, lifting scheme can be used to produce Int2Int wavelets. Secret image is hidden in these wavelet coefficients. Fig. 1 shows the decomposition of audio signal on wavelet transform.

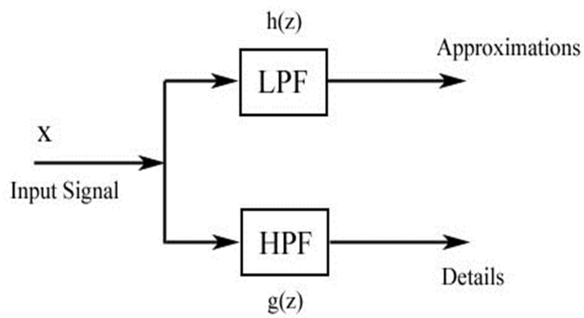


Figure 1. One stage signal decomposition

Lifting Wavelet Transform (LWT)

The lifting scheme is a technique for both designing wavelets and performing the discrete wavelet transform. If the steganographic methods need to process the binary value of the resultant coefficients, then coefficients have to be converted from floating to binary. This conversion requires first scaling and then converting into 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. This prevents errors such as rounding errors and out of range errors. These errors are usually encountered with DWT.

Arnold Transformation

Arnold's Cat Map is a transformation that is applied to the secret image. The pixels of the image appear to be randomized, but when the transformation is applied in quick succession, the original image will recur [3]. The transformation is as follows:

- Shear the secret image in the both the x- and y-directions by a factor of 1.
- Wrap the resultant image back onto itself i.e. treat it like a torus (doughnut shape).

Section II will brief on some of related works which has been done on audio steganography. Section III will explain the methodology. Chapter IV contains experimental results and its analysis.

LITERATURE SURVEY

In recent years, several researchers have concentrated on developing algorithms for hiding data in an audio signal. Fatiha Djebbar et.al, [4] focus on different steganographic techniques used, evaluate their performance and analyze them. Rully Adrian Santosa and Paul Bao [5] propose an audio steganographic scheme based on wavelet audio-to-image transform. In the scheme, the host audio signal is transformed into image, the covert data are embedded in the image by an image steganographic scheme and finally, the image is transformed back into audio signal. Weakness of this scheme is bit error i.e. Secret data is not exactly same as original data which was embedded.
Xi Lua et.al, [6] proposes a method where in digital audio is transformed into a 2D image called sound map. Each window of sound map is encrypted with Arnold transformation. Then the resultant is re-encrypted using double random phase encoding technique. Sateesh Gudla et.al, [7] proposed a new Least Significant Bit (LSB) replacement steganographic algorithm for both audio and video based on Diffie-Hellman key exchange. The proposed algorithm takes the secret data and conceals them in LSB pixels of the cover frames of video and the selected frames of audio in order to retain the quality of audio and video.

Audio steganography in wavelet domain gives better stego audio and quality of extracted data is also quite good compared to that in other domains. SajadShirali-Shahreza and M.T. Manzuri-Shalmani [8] 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 detail 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., $2^p \leq c < 2^{p+1}$ 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 the stego audio is imperceptible.

Ahmad Delforouzi [9] 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. Yongfeng et al.,[10] propose a new algorithm by integrating information hiding into the process of speech encoding, in low bit-rate VoIP audio streams. Data embedding is conducted during low bit-rate speech encoding, while pitch period prediction is carried. Siwar Rekik et.al, [11] describes an secure method in which Discrete Wavelet Transformation is applied to the audio signal and using low pass spectral properties a secret audio signal is hidden in high frequency components of cover audio. Disadvantage of this method is issue of preserving the secret speech after decoding the compressed stego audio.

PROPOSED METHOD

This section discusses the algorithm used to hide randomized image in cover audio. Resultant audio when secret image is hidden in cover audio is called stego-audio. Algorithm has two phases – embedding and extraction. In embedding phase, secret image is hidden inside the cover audio. 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 data is retrieved from the stego-audio. In this algorithm, audio samples are transformed into wavelet domain. Secret data here is image, which is scrambled using Arnold transformation. These transformed values of image are then hidden in LSB's of detail coefficients.

Embedding Phase

Step 1: Image Pre-processing: Read grayscale image to be hidden. Choose the iteration value, based on it; apply Arnold transformation on image matrix. Resultant image is scrambled image. Rearrange the resultant matrix into vector and convert this vector into binary. Size of vector is stored.

Step 2: Audio Processing: Read the cover audio file. Audio samples are stored in a vector and are signed floating point values. In order to get integer coefficients from audio samples, audio samples must be converted to integers. This conversion is performed here. Negative values are handled using flags.

Step 3: Apply LWT based on lifting scheme: Integer to integer (int2int) transformation can be implemented using Lifting Wavelet Transformation (LWT). LWT uses lifting scheme (LS). In LS appropriate wavelet is chosen and '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 4: Calculate number of bits to be replaced: This algorithm chooses dynamic approach to find the bits to hold the secret scrambled image. The number of bits used to hold data is calculated using the concept explained by Sajad Shirali-Shahreza and M.T. Manzuri-Shalmani [8]. Detail coefficients are selected to hold the secret image.

Step 5: Hiding procedure: This step is sub divided into two parts: Hiding the size of the image and hiding the image.

- Hide image size: It is necessary to embed the secret image vector size into the cover audio because during extraction of image from stego-audio, receiver need to know how many bits has to be extracted from stego- audio. First 16 replaceable LSB's are reserved to store the image size, based on NBR calculated for each CD value. Image size bits, starting from MSB, are stored in LSB's of CD's.
- Hide the image: Remaining replaceable bits of each CD are used to store the secret image. Image bits starting from MSB, are stored in LSB's of each CD.

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

Fig. 2 depicts the embedding phase.

Extraction Phase

Step 1: Audio Processing: Read the stego audio file. Audio samples are stored in a vector and are signed floating point values. In order to get integer coefficients from audio samples, audio samples must be converted to integers. This conversion is performed here. This step is same as in embedding.

Step 2: Apply LWT based on lifting scheme: Select the same lifting scheme which is used during embedding. 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 in embedding phase (Step 4).

Step 4: Retrieving procedure: This step is sub divided into two parts: retrieve image size and retrieve secret image.

- Retrieve image vector size: Based on NBR calculated for each CD value using image size bits starting from MSB, are retrieved from LSB's of CD's. This step loops until 16 bits are retrieved.
- Retrieve the actual image: Scrambled image bits starting from MSB, are retrieved from LSB's of remaining CD's and store it in a buffer, using extracted image size. After retrieving, scrambled image bits are converted to decimal and then rearranged as matrix.

Step 5: Inverse Arnold transformation is applied on encrypted image matrix to retrieve the secret image. Fig. 3 depicts the extraction phase.

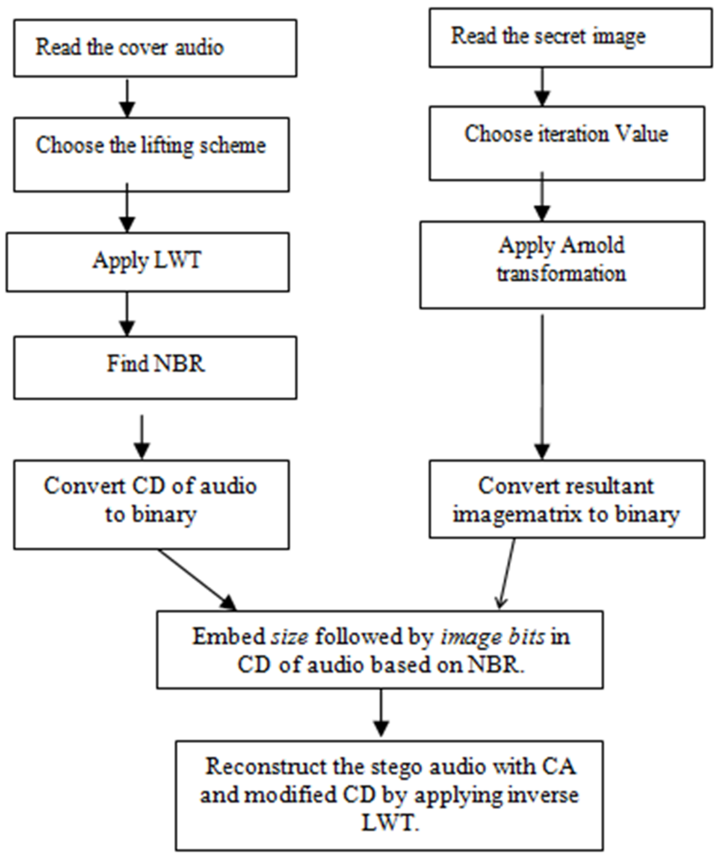


Figure 2. Embedding Phase

Experimantal Results

The algorithm is tested using two audio signals, two.wav and woody2.wav. Duration of two.wav is 6 seconds and that of woody2.wav is 3 seconds. Sampling rate of woody2.wav is 11025 samples per second and that of two.wav is 1024 samples per second. 276347 audio samples are present in two.wav and 37620 samples are present in woody2.wav. Here the maximum hiding capacity does not depend on size of the audio rather it depends on wavelet coefficients values of cover

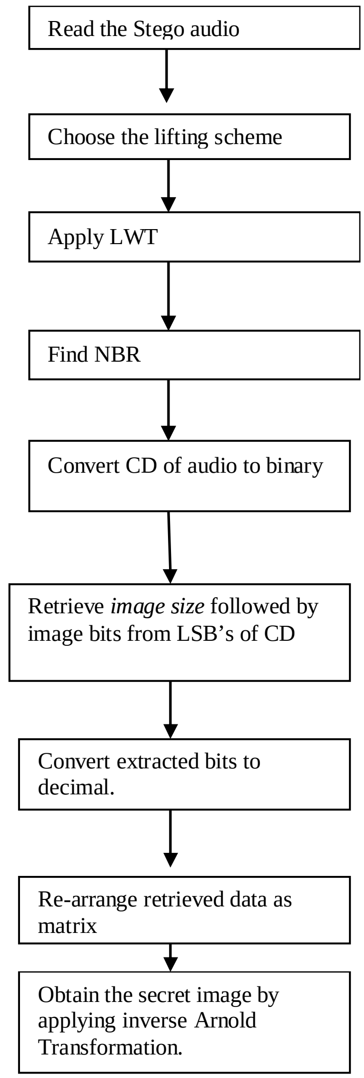


Figure 3. Extraction Phase

which in turn depends on audio sampling rate. Wavelet used is "db2". Images of different formats and of size 64 X 64 pixels are hidden in cover audio signals. Quality of stego-audio is analysed using Signal to Noise Ratio (SNR) and Squared Pearson Correlation Coefficient (SPCC).

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. The Mean Square Error (MSE) between the signals is given by (1).

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

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. SNR specifications are important sound level measurements used in describing the capabilities and qualities of many electronic sound components. The measurement is usually expressed in decibels (or dB). A larger SNR value means a better quality [12].

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

Recommended SNR for audio signal is above 30db.
SPCC measures the level of similarity between the input and output signals. The higher the SPCC, the better is the quality of the output signal. Its range is between 0 and 1. It is given by (3).

$$SPCC = \left[\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}\right]^2 \quad (3)$$

where x, y, $\bar{x}$ and $\bar{y}$ are the input signal, output signal, average of the input signal and average of the output signal, respectively.
Table I shows SNR and SPCC values for different stego audio signals. It can be observed that as the hiding capacity is increased, SNR is decreased. SPCC is greater than 0.9, which is a good measure. These experiments are done with OBH=1.
The extracted image quality is tested by computing the Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM) and Universal Image Quality Index (UIQI).
The most widely used parameter to measure the quality of an image is PSNR. It is expressed in decibels (dB). It is given by (4).

$$PSNR = 10 * \log_{10}\left(\frac{MAX^2}{MSE}\right). \quad (4)$$

MAX is the maximum value of pixels (255 for gray scale images). MSE is the mean square error between the original and stego images. Greater PSNR values indicate better quality.
SSIM and UIQI are two Human Visual System (HVS) based metrics [13]. SSIM is an objective image quality metric and is superior to traditional measures such as MSE and PSNR. PSNR estimates the perceived errors, whereas SSIM considers image degradation as perceived change in structural information. It is given by (5). UIQI is given by (6).

$$SSIM = \frac{(2\times\bar{x}\times\bar{y}+C1)(2\times\sigma_{xy}+C2)}{({\sigma_x}^2+{\sigma_y}^2+C2)\times(\bar{x}^2+\bar{y}^2+C1)} \quad (5)$$

$$\text{UIQI} = \frac{4 \times \sigma_{xy} \times \bar{x} \times \bar{y}}{(\sigma_x{}^2 + \sigma_y{}^2) \times (\bar{x}^2 + \bar{y}^2)} \quad (6)$$

where $\bar{x}, \bar{y}, \sigma_x{}^2, \sigma_y{}^2 \, and \; \sigma_{xy}$ are given as

$$\bar{x} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j))$$

$$\bar{y} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (y(i,j))$$

$$\sigma_x{}^2 = \frac{1}{M \times N - 1} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j) - \bar{x})^2$$

$$\sigma_y{}^2 = \frac{1}{M \times N - 1} \sum_{i=1}^{M} \sum_{j=1}^{N} (y(i,j) - \bar{y})^2$$

$$\sigma_{xy} = \frac{1}{M \times N - 1} \sum_{i=1}^{M} \sum_{j=1}^{N} ((x(i,j) - \bar{x})\,(y(i,j) - \bar{y}))$$

$C1 = (k_1 L)^2$, and $C2 = (k_2 L)^2$ are two constants used to avoid null denominator. L is the dynamic range of the pixel values (typically this is $2^{\#\text{bits per pixel} -1}$). $k_1 = 0.01$ and $k_2 = 0.03$ by default. The dynamic range of SSIM is between -1 and 1. Maximum value of 1 will be obtained for identical images.

In this algorithm the secret images are extracted without any error. That means there is no difference between the secret and extracted images. So an infinite value of PSNR and SSIM, UIQI values of 1 are obtained.

White Gaussian noise with different SNR values were added to stego – audio and observed the output. Secret data was able to be retrieved without any errors. This is used to check the robustness of the algorithm. External noise of different levels can be added to the stego – audio.

Table 1. SNR and SPCC values of stego audio signal

Audio	Image (64 X 64)	SNR(dB)	SPCC
Two.wav	Emo.png	39.20	0.98
Woody2.wav	Emo.png	35.11	0.95
Two.wav	Butterfly.jpg	39.09	0.97
woody2.wav	Butterfly.jpg	34.5	0.92
Two.wav	Lena.bmp	40	0.985
woody2.wav	Lena.bmp	39	0.98

Figures 4-6 show the cover, stego and stego with noise audio with lena image being hidden in two.wav, respectively. It can be observed that significant changes are not perceptible.

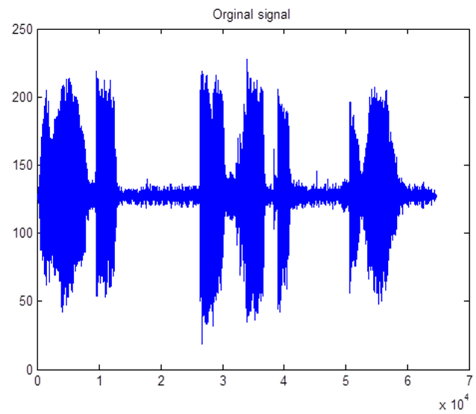


Figure 4. Cover audio – two.wav

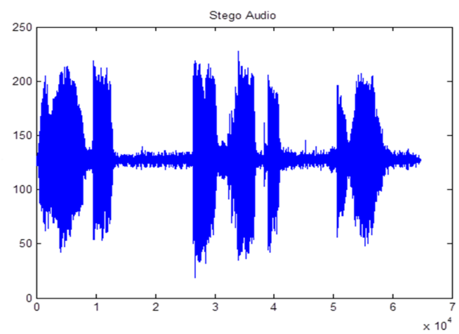


Figure 5. Stego-audio

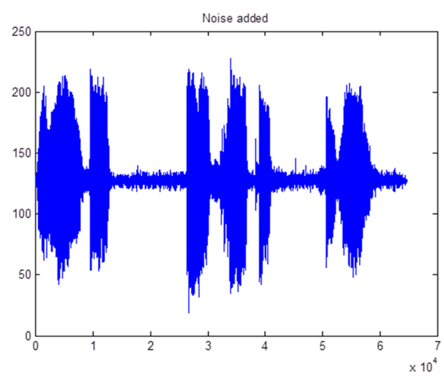
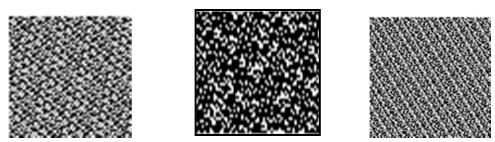


Figure 6. Stego-audio with white Gaussian noise

Fig. 7 shows the secret images used for hiding. Fig. 8 shows the encrypted images. Fig. 9 shows the extracted images from two.wav.

(a) (b) (c)

Figure 7 Secret Images (a) Lena (b) Emo (c) Butterfly



(a) (b) (c)

Figure 8. Encrypted Secret Images (a) Lena (b) Emo (c) Butterfly

(a) (b) (c)

Figure 9. Extracted Secret Images (a) Lena (b) Emo (c) Butterfly

As seen in Fig. 9, the extracted secret images are exactly same as the original secret images. Subjective tests for audio quality evaluation were 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 arandom order. For most of the cases, listener could not differentiate 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. Experiment was conducted with other wavelets as well; there is no significance change in the results.

CONCLUSION

Objective of the paper is to hide image in cover audio using lifting wavelet transform. Based on the values of coefficients, number of bits used to hold secret data is chosen. In proposed method, image is scrambled using Arnold transform and then this transformed image is hidden in cover audio. Results are computed and observed. Arnold Transformation used induces security to some extent. This algorithm yields zero error extraction, good SNR and SPCC. Infinite PSNR value and unity SSIM and UIQI values indicate that the secret image is retrieved without any loss. Similar technique is used by Sajad Shirali-Shahreza and M.T. Manzuri-Shalmani [8] but with text data. There SPCC is not calculated, which is a good metric to test the audio quality based on correlation.

In the proposed method approximately same values of SNR is obtained as in [8] even with encryption and noise added when the secret information is an image. As the audio samples for even 30 sec audio file is in lakhs, processing it and hiding image and extracting it takes lot of time. This drawback can be eliminated by implementing whole project in parallel using GPU's.

REFERENCES

Jisna Antony and Sobin C, "Audio Steganography in Wavelet Domain – A Survey", *International Journal of Computer Applications*, Volume 52– No.13, August 2012.

Michael Weeks, "Digital Signal Processing Using MATLAB and Wavelets", 2007

Arnold Transformations, www.jasondavies.com/catmap/

Fatiha Djebbar and Beghda Ayad, "Comparative Study of Digital Audio Steganography Techniques", *EURASIP Journal on Audio, Speech, and Music Processing*, 2012

Rully Adrian Santosa and Paul Bao, "Audio-to-image wavelet transform based audio steganography", *IEEE Int. Symp.*, June 2005.

Xi lua, "Digital audio information hiding based on Arnold transformation and double random-phase encoding technique", Optik 123, 2012,

Sateesh Gudla, Suchitra Reyya, Aswini Kotyada, Aditya Sangam ,"key based least significant bit (lsb) insertion for audio and video steganography", *International Journal of Computer science and Engineering Research and Development (IJCSERD)*, Volume 3, Number 1, Jan-March 2013, pp: 60-69

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

Mohammad Pooyan, Ahmad Delforouzi," LSB-based Audio Steganography Method Based on Lifting Wavelet Transform", *IEEE International Symposium on Signal Processing and Information Technology* 2007.

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

Siwar Rekik, Driss Guerchi, Sid-Ahmed Selouani and Habib Hamam, "Speech steganography using wavelet and Fourier transforms.",EURASIP Journal on Audio, Speech, and Music Processing, 2012.

Baiying Lei, Ing Yann Soon, Feng Zhou, Zhen Li, Haijun Lei, "A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition", Signal Processing, Vol. 92, September 2012, pages 1985-2001.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh and Eero P. Simoncelli, "Image Quality Assessment:From Error Visibility to Structure Similarity", IEEE Transactions on image processing, Vol. 13, No. 4, 2004.