Optimally Locating for Hiding Information in Audio Signal

Premalatha P
Amrita Vishwa Vidyapeetham
Coimbatore

Amritha P P
Amrita Vishwa Vidyapeetham
Coimbatore

ABSTRACT
Steganography provides security and privacy of information on open environment systems. Audio steganography plays a vital role in hiding information by exploiting the human ear perceptibility. In this paper, the harmonic component that are imperceptible to the human auditory system are manipulated using Fast Fourier Transform to hide data within the samples. The decoder samples the modified song and extracts the hidden message with the key, using an error correcting code to fix any bits altered by the channel.

General Terms
Signal processing, Multimedia Security.

Keywords
Human Auditory System, Audio Steganography, Fast Fourier Transform, Harmonics.

1. INTRODUCTION
With the widespread deployment of public networks like the Internet and frequent use of digital media in different applications, the field of information hiding [1] got a new lease of life. Steganography in today's computer era is considered a sub-discipline of data communication security domain. It is the act of covert communications, which means that only the sender, Alice, and receiver, Bob, are aware of the secret communication. To accomplish this, the secret message is hidden within benign-looking communications known as cover texts or cover Works. To an adversary, Eve, it is clear that Alice and Bob are communicating, but the combined cover text and hidden message, referred to as a stego text or stego Work, appears to be innocuous (i.e., Eve is unaware that the innocuous content hides a message).The main requirement of steganography is undetectability, which, loosely defined, means that no algorithm exists that can determine whether a Work contains a hidden message[2]. Given the source of cover Works, Alice and Bob need to construct the embedding and extraction functions. Fundamentally, an embedding function can be based on three different principles, namely:

1. The cover Works are preexisting and the embedder does *not* modify the cover Works. This is referred to as steganography by cover lookup.

2. The cover Works are generated based on the hidden message and the embedder does *not* modify the cover Works. This is referred to as cover synthesis.

3. The cover Works are preexisting and the embedder modifies the cover Works. This is referred to as steganography by cover modification.

Steganography by cover modification which is used in the proposed technique, describes methods where Alice alters an existing cover Work to create a stego Work that conveys the desired message. This approach is both the most common and the most advanced. The type of changes introduced by the embedder, together with the location of these changes within the cover Work, have a major influence on how inconspicuous the embedded message will be. Intuitively, changes of large magnitude will be more obvious than changes of smaller magnitude. Consequently, most steganographic schemes try to modify the cover Work as little as possible. Alice and Bob may wish to adopt a more sophisticated key management and periodically change the key according to some pre-agreed protocol. For example, the message may be communicated using a session key that is different for each cover and communicated in the cover itself [2].Modern techniques of steganography exploit the characteristics of digital media by utilizing them as carriers (covers) to hold hidden information. Cover scan be of different types including image, audio, video, text, and IP datagram. The availability and the popularity of audio files make them eligible to carry hidden information. Data hiding in audio files as shown in figure 1, is especially challenging because of the sensitivity of the Human Auditory System (HAS). For example, loud sounds tend to mask out quiet sounds. Additionally, there are some common environmental distortions, to the point that they would be ignored by listeners in most cases. These properties have led us to explore the utilization of audio signals as carriers to hide data. There are three main digital audio formats typically in use. Sample Quantization which is a 16-bit linear sampling architecture used by popular audio formats such as (.WAV and. AIFF). Temporal Sampling Rate uses selectable frequencies (in the KHz) to sample the audio[3]. Perceptual Sampling changes the statistics of the audio drastically by encoding only the parts the listener perceives, thus maintaining the sound but changing the signal (eg.MP3). In this paper, Sample quantization audio format (.wav) is used for experimental results.

The proposed method in this paper can be used in all the four transmission medium introduced by W. Bender [4].

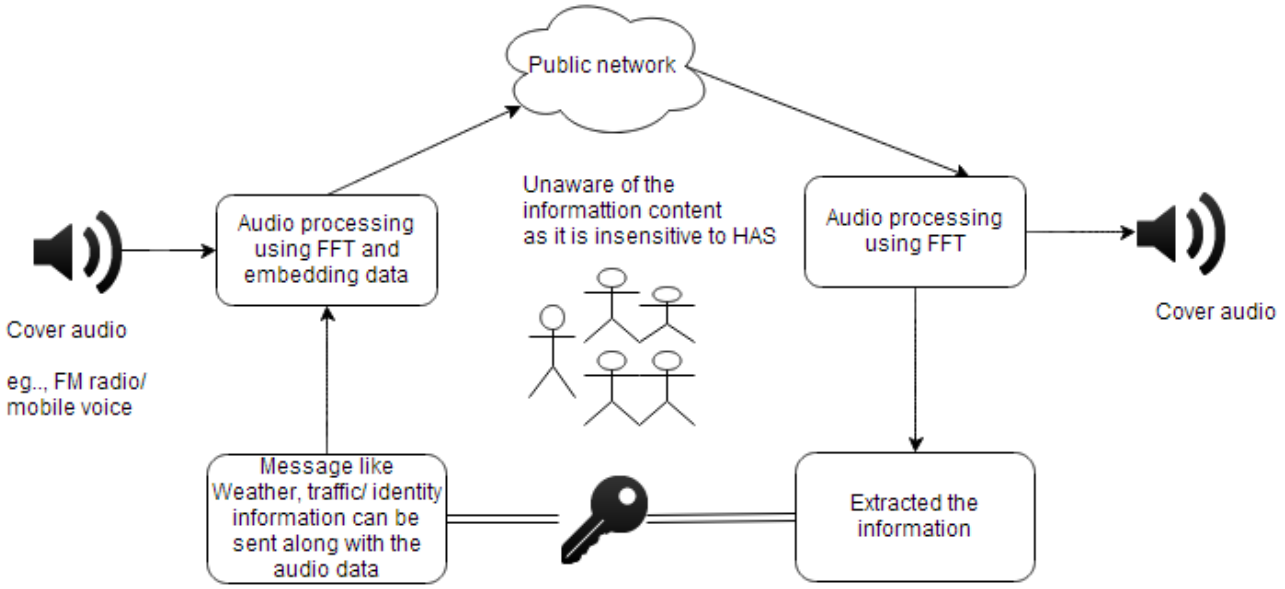

Figure 1: Audio steganography workflow

This paper is organised as follows. Section 2 explains the existing methodology. Section 3 describes the areas that the proposed system can be applied. Section4 gives description about human perception of sound. Section5 discusses the proposed algorithm. Section 6 explains its implementation result and its discussion. Section 8 summarizes the paper and provide future works.

2. EXISTING METHODOLOGY

In Audio steganography, information can be hidden in three domains namely temporal domain, frequency/wavelet domains and coded domain. All the domain strength and weakness are briefly explained in the paper [5,6].

2.1 Hiding in Temporal domain

2.1.1 LSB coding method

This method is one of the earliest methods used for information hiding. Traditionally, it is based on embedding each bit from the message in the least significant bit of the cover audio in a deterministic way (see Figure 2). Thus, for a 16 kHz sampled audio, 16 kbps of data are hidden. The LSB method allows high embedding capacity for data and is relatively easy to implement or to combine with other hiding techniques. However, this technique is characterized by low robustness to noise addition which reduces its security performance since it becomes vulnerable even to simple attacks. Filtration, amplification, noise addition and lossy compression of the stego-audio will very likely destroy the data [5].

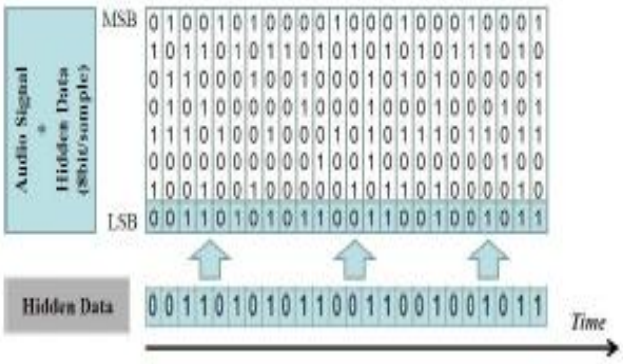

Figure 2: LSB encoding

In LSB coding method, the number of used LSBs during LSB coding increases or, equivalently, depth of the modified LSB layer becomes larger, probability of making the embedded message statistically detectable increases and perceptual transparency of stego objects is decreased.

2.1.2 Echo hiding

Artificial echo are used to hide the embedded data .If the echo delay between the original source audio and the echo decrease it becomes harder for the human ear to distinguish between the two signals until eventually a created carrier sound's echo is just heard as extra resonance. Data are hidden by manipulating three parameters of the echo signal: the initial amplitude, the offset (delay) and the decay rate so that the echo is not audible (Figure 3). The amplitude and the decay rates could be set to values under the audible threshold of the human ear. Data could thus be hidden without being perceptible[5].

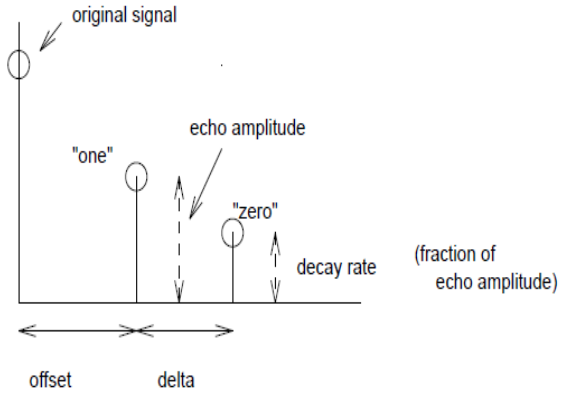

Figure 3: Echo data hiding

However, the drawback of this method is the limitation of induced echo signal size which restrict its related application domains. Hence, the limited amount of works investigate the application of this method.

2.1.3 Hiding in silence intervals

A simple and effective embedding method has been used to exploit silence intervals in speech signal. Changes in silence intervals can lead to false data extraction. To overcome this shortcoming, suggested to slightly amplify speech interval samples and reduce the silence interval samples. Thus, silence sample intervals will not be interpreted as speech samples and vice-versa. The first and last interval added to the speech during MP3 coding is simply ignored in data hiding and retrieval[5].

2.2 Hiding in Transform domain

To achieve the inaudibility, these methods exploit the frequency masking effect of the HAS directly by explicitly modifying only masked regions or indirectly by altering slightly the audio signals samples[5].

2.2.1 Spread spectrum

Spread spectrum systems encode data as a binary sequence which sounds like noise but which can be recognised by a receiver with the correct key. A conventional direct sequence spread spectrum (DSSS) technique was applied to hide confidential information in MP3 and WAV signals[5].

2.2.2 Discrete Wavelet Transform

Audio steganography based on Discrete Wavelet Transform (DWT) is described in [6,11]. Data is hidden in the LSBs of the wavelet coefficients of the audio signals. Even though data hiding in wavelet domain procures high embedding rate, data extraction at the receiver side might not be accurate [5].

2.2.3 Tone Insertion

Tone insertion techniques rely on the inaudibility of lower power tones in the presence of significantly higher ones. Embedding data by inserting inaudible tones in cover audio signals is presented in [8].Tone insertion method can resist to attacks such as low-pass filtering and bit truncation[5].

2.2.4 Phase coding

Phase coding relies on the fact that the phase components of sound are not as perceptible to the human ear as noise is. Rather than introducing perturbations, the technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to-perceived noise ratio(see Figure 4)[5].

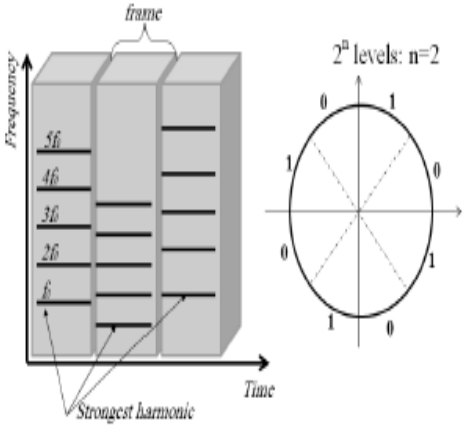

Figure 4: Phase coding for strongest harmonics

2.2.5 Amplitude coding

In Amplitude coding [5], HAS characteristics depend more on the frequency values as it is more sensitive to amplitude components. Following this principle, authors in [9] proposed a steganographic algorithm that embeds high-capacity data in the magnitude speech spectrum while ensuring the hidden-data security and controlling the distortion of the cover-medium. The hidden data (payload) could be of any type such as: encrypted data, compressed data, groups of data (LPC, MP3, AMR, CELP, parameters of speech recognition, etc). The proposed algorithm is based on finding secure spectral embedding-areas in a wideband magnitude speech spectrum using a frequency mask defined at 13 dB below the original signal spectrum. The embedding locations and hiding capacity in magnitude components are defined according to a tolerated distortion level defined in the magnitude spectrum. Since the frequency components within the range of 7 kHz to 8 kHz contribute minimally to wideband speech intelligibility, [10] proposed a method to hide data in this range by completely replacing the frequencies 7-8 kHz by the message to be hidden. The method realizes high hiding capacity without degrading the speech quality.

2.3 Coded domain

When considering data hiding for real time communications, voice encoders such as: AMR, ACELP and SILK at their respective encoding rate are employed. When passing through one of the encoders, the transmitted audio signal is coded according to the encoder rate then decompressed at the decoder end. Thus, the data signal at the receiver side is not exactly the same as it was at the sender side, which affects the hidden data-retrieval correctness and therefore makes these techniques very challenging. There are two such techniques, namely in-encoder and post-encoder techniques[5].

In all the existing methodology perceptibility of human ear is checked. In our method we include the concept of amplitude coding where only higher frequency values are concentrated. we also make use of the sensitivity of the human ear directly and find those frequencies and low magnitude values that are not sensitive to ear to embed the message. Thus optimally locating the sample points to hide the information in a signal.

3. APPLICATION

The proposed steganographic technique can be used to embed meta-data in audio. It can play a part in adding identification of the parties involved during telephone or video conference conversation. Audio steganography can also be used to embed a unique token into music. This can be used for a Digital Rights Management (DRM) scheme that would help limit piracy. An useful application relates to automatic audit of advertisements played on the radio or TV channels. The audio/video clip of the advertisement is marked and a suitable monitoring device is placed in front of the radio/TV that extracts its serial/identification number. In this way it is able to keep a log of the time and frequency of an advertisement played on a specific channel. To minimize the difference between the cover and the stego-medium recent steganography techniques utilize natural limitations in human auditory system[13].

4. THE EAR AS A FREQUENCY SPECTRUM ANALYSER

Sound waves reaching the ear are the vibrations of air particles. But not all vibrations are perceived by the human ear. The human ear has an enormous range of response, both in frequency and intensity which will be discussed in the following section.

4.1 The place principle

When exposed to a high frequency signal, the basilar membrane resonates where it is stiff, resulting in the excitation of nerve cells close to the oval window. Likewise, low frequency sounds excite nerve cells at the far end of the basilar membrane. This makes specific fibers in the cochlear nerve respond to specific frequencies. This organization is called the place principle, and is preserved throughout the auditory pathway into the brain. For this reason, frequencies of between1,000 and 6,000Hz , the range to which the human ear is most sensitive, the intensity range is from about 0db to 120db[11].

4.2 Range of human hearing

The range of human hearing is generally considered to be 20 Hz to 20kHz, but it is far more sensitive to sounds between 1 kHz and 4 kHz as shown in figure 1. For example, listeners can detect sounds as low as 0 dB SPL at 3 kHz, but require 40 dB SPL at 100 hertz (an amplitude increase of 100).Listeners can tell that two tones are different if their frequencies differ by more than about 0.3% at 3 kHz[11]. Hence, the minimum magnitude values and the higher frequency values in the audio signal can be modified to encode the secret message such that the only the decoder can detect the message at the receiver side but the human cannot. Diagrammatic representation of the range of hearing is shown in figure 5 [11].

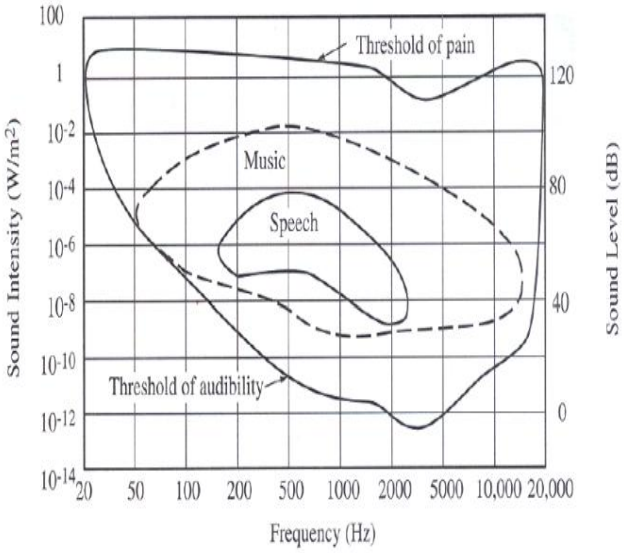

Figure 5: Range of hearing

5. PROPOSED TECHNIQUE

5.1 Information hiding algorithm

Input: Original audio signal, text message (plaintext format).
Output: Modified audio signal with embedded message.
Method: Embedding a message in the selected frequency areas by making use of the human perceptibility as follows

Step 1: Find all frequency components using Fast Fourier Transform (FFT).

Step 2: Find the magnitude range such a way that our human ear not able to distinguish the difference after changing those harmonic components as explained in section 4.

Step 3: Within the range use any two amplitude value to distinguish message bit 0 and the message bit 1 as within that amplitude range any change to the magnitude value of the harmonics can't be distinguished by our ear. These amplitude values& position where embedding data starts, are chosen as a key and it is shared between the communication parties through secret channel.

Step 4: Use the error correction technique here repetition code, though there is a noise/disturbance in the channel the message can be extracted.

Step 5: Apply Inverse Fast Fourier Transform (IFFT) and send the stego audio signal to the channel.

5.2 Information extracting algorithm

Input: Stego audio signal, key.
Output: Text message.

Step 1: Find all frequency components of the received audio signal using Fast Fourier Transform (FFT).

Step 2: From the position mentioned in the key, search for the amplitude values in the key, if the magnitude value of any frequency component equals one of the amplitudes decode as a corresponding bit either zero or one and the other vice versa.

Step 3: There is a possibility that another amplitude value is likely to be obtained due to the presence of noise in the channel while decoding the first message block. The error can be neglected as the message is repeated entirely in the selected harmonic components.

6. EXPERIMENTAL RESULTS

For experimental observation, a strip of 2.068 sec music bit has been taken. Details of audio signal shown in table 1.

6.1 Original song

The graphical representation of the original song, considering sampled values of *x* (*n*, 1) [mono type] is shown in the figure 6. In figure the sample time vector(x axis) is plotted against loudness/amplitude values(y axis).

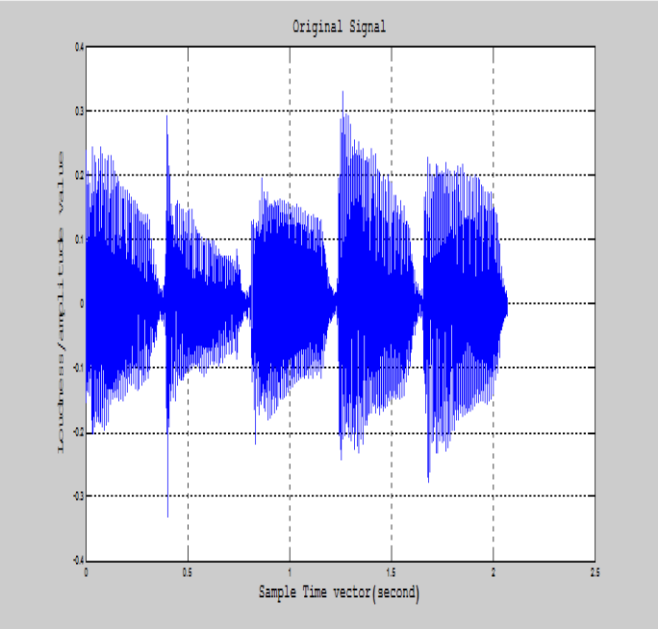

Figure 6: Plot of an Original audio signal in time domain

The absolute magnitude 0.06 occurs at 393Hz which is the fundamental frequency. In figure 7 frequency(x-axis) versus magnitude(y axis) is shown. The maximum peak in the figure below is the fundamental frequency corresponding to magnitude 0.06.

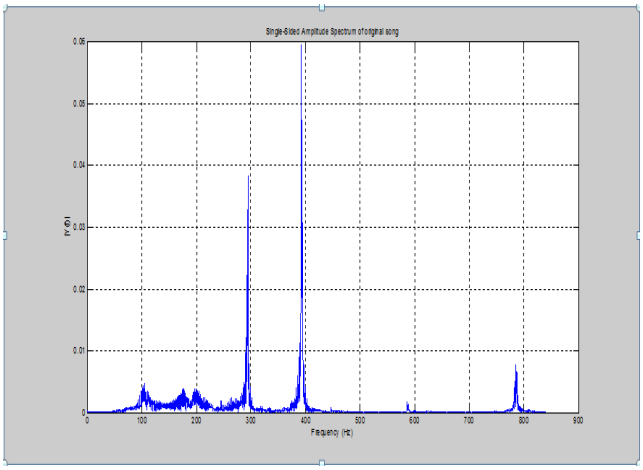

Figure 7: Plot of an Original audio signal in frequency domain

. Table 1: Parameters of audio signal

Parameters	Values
Duration	2.068 sec
Sampling Frequency	44100 samples/sec
Length of the signal	2.068*44100 = 91279 samples
Peak to peak sound intensity	0.06 watts/m^2

6.2 Stego audio signal

The harmonic magnitude between the range $-8*10^{-9} < F(\omega) > 10^{-6}$ is used for embedding the message such a way it is not affecting the sensitivity of HAS. The magnitude value chosen within threshold range for a message bit 1 is $-5*10^{-7}$ and for message bit 0 is 0 itself. After embedding the data, the stego signal plotted is shown in figure 8. This figure is similar to the plot of original signal shown in figure 6.

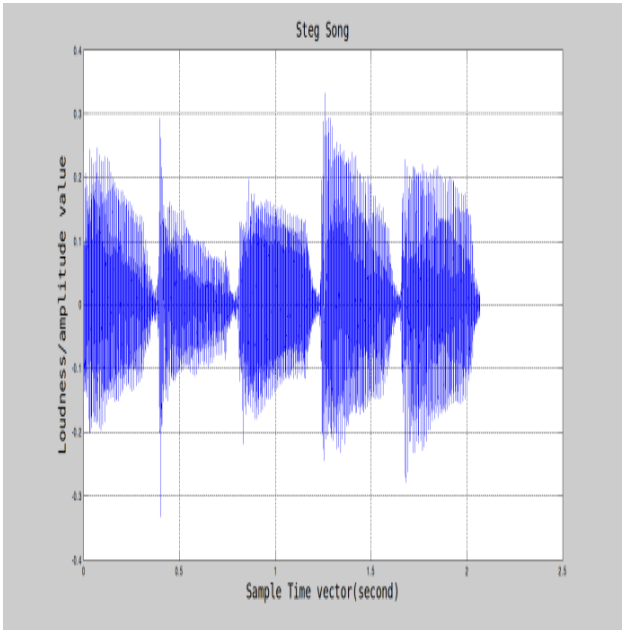

Figure 8: Plot of Stego audio signal travel in the channel not known to human

So this method proved that there is no change in the amplitude of signal even after embedding.

6.3 Performance metrics

Below table shows the measures used to estimate the efficiency of this proposed method.

Table 2: Performance Metrics

Parameters	Values
Mean Squared Error (MSE)	2.0521e-008
Signal to Noise ratio (SNR)	177.0183dB
Peak Signal to Noise Ratio (PSNR)	154.9255dB
Payload	17218 among 91279 samples (18.86%)

Thus, for 2.068 sec audio signal the minimum payload capacity is of 18.86% .So if it is 3 minute song then there can be excess payload capacity.

7. CONCLUSION AND FUTURE WORK

In this paper, an algorithm is proposed by analyzing the magnitude of frequency components not sensitive to human ear and found optimal positions to hide information. An error occurred during extracting information can be avoided using repetition code. This work can be extended in stereophonic sound. The proposed system can also be implemented in a DSP device which will inject data into an audio signal such that when the music/conversation occurs, a receiver device can detect the data through the sound waves but a human cannot.

8. ACKNOWLEDGMENTS

We express our gratitude to Dr. Krishnan, Head of Engineering research and Senior Professor, Electrical sciences, Amrita Vishwa Vidyapeetham, Coimbatore, for his guidance

9. REFERENCES

[1] Anderson R. J. (ed.), Information Hiding: First International Workshop Proceedings, Lecture Notes in Computer Science, Vol. 1174, Springer-Verlag, 1996.

[2] Cox, et al, “Digital watermarking and Steganography", USA: Massachusetts, 2008.

[3] Bret Dunbar, “A detailed look at Steganographic Techniques and their use in an Open-Systems Environment”, Information Security Reading Room, SANS Institute 2002.

[4] 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.

[5] F.Djebbar, B. Ayad, K. Abed-Meraim and H. Hamam, 2012. Comparative study on digital audio steganographic techniques.

[6] F. Djebbar, B. Ayad , K. Abed-Meraim and H. Hamam , "A view on latest audio steganography", 7th IEEE International Conference on Innovations in Information Technology, Abu Dhabi, UAE, 2011.

[7] N. Cvejic, T. Seppanen, “A wavelet domain LSB insertion algorithm for high capacity audio steganography”, Proc. 10th IEEE Digital Signal Processing Workshop and 2nd Signal Processing Education Workshop, pp. 5355, 1316 October 2002.

[8] K. Gopalan, et al, “Covert Speech Communication Via Cover Speech By Tone Insertion”, Proceeding of IEEE Aerospace Conference, Big sky, MT, March 2003.

[9] F. Djebbar, B. Ayad, K. Abed-Meraim and H. Habib, “Unified phase and magnitude speech spectra data hidingalgorithm”, Journal of Security and Communication Networks, John Wiley and Sons, Ltd, April, 2012.

[10] D. Guerchi, H. Harmain, T. Rabie, and E. Mohamed, “Speech secrecy: An FFT-based approach”, International Journal of Mathematics and Computer Science, vol.3, no.2, pp.1-19, 2008.

[11] Steven W. Smith ,"Digital Signal Processing: A Practical Guide for Engineers and Scientist", USA: Massachusetts, 2003.

[12] Mondal, Uttam Kr., Mandal, J.K., “Audio signal authentication through secret embedded self-harmonic component", International Journal of Security, Privacy and Trust Management (IJSPTM), Vol. 1, No 3/4, August 2012.

[13] Pal S. K., Saxena P. K. and Muttoo S. K., Smart Steganographic Applications, Proceedings of the Pacific Rim Workshop on Digital Steganography, STEG’02, Japan, July 2002, pp. 11-19.