International Journal of Research in Advent Technology, Vol.4, No.4, April 2016
E-ISSN: 2321-9637
Available online at www.ijrat.org

Review on Speech and Audio Steganography Techniques

Pratibha A. Nandane[1], Prof. Mrs.P.P.Belagali[2]
Department of Electronics Engineering[1] , Department of Electronics And Telecommunication Engineering[2].
P.G.Student[1] ,, Professor[2]
Email: pratibha.n07@gmail.com[1] , ppbelagali@rediffmail.com[2]

Abstract- Audio and speech steganography is the method of hiding secret information in an audio or speech file. The audio file in which the secret information is hidden is known as audio cover. The sender embeds the secret message in the audio cover file using a key to produce a stego-file. At the receiver end, the receiver processes the received stego-file and extracts the hidden message. In this paper study of different types of audio steganographic methods are done. This paper explores the different method of data hiding.

Index Terms- Embedding; Data hiding; Speech Steganography; Stego Signal.

1. INTRODUCTION

In this era of emerging technologies, electronic communication has become an integral and significant part of everyone's life because it is simpler, faster and more secure. Steganography is the technology of secret communication via a digital cover media such as image, audio or video files. Embedding secret messages into speech is known as speech Steganography.The ultimate goal of a speech steganography is to conceal the presence of secret message embedded in the cover media. Speech steganography is a powerful tool which increases security in data transferring and archiving.

In speech steganography, the speech signal is called as cover signal. The secret message data is embedded into speech signal and form a new signal called as stego signal. This stego signal looks same as cover signal. At the receiver side, the secret message is extracted from this stego signal using extraction method. Speech steganography is concerned with hiding information in a cover (host) audio signal in an imperceptible way. Hidden information from the stego, or data-embedded audio signal, is retrieved using a key similar to the one that was employed during the hiding phase. The objective of this paper is to come up with at technique hiding the presence of secret message. Steganography is the art of secret communication. Its purpose is to hide the presence of communication, as opposed to cryptography, which aims to make communication unintelligible to those who don't possess the right keys. The different techniques are presented. [1].

2. REVIEW OF AUDIO STEGANOGRAPHY METHODS

Based on the reviewed methods in this paper, three prominent data embedding approaches have been investigated, namely hiding in temporal domain, in frequency/wavelet domains and in coded domain.

2.1. Hiding in temporal domain

2.1.1. Low-bit encoding

Also known as LSB (Least Significant Bit), this method is one of the earliest methods used for information hiding [1]. 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 1). 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.

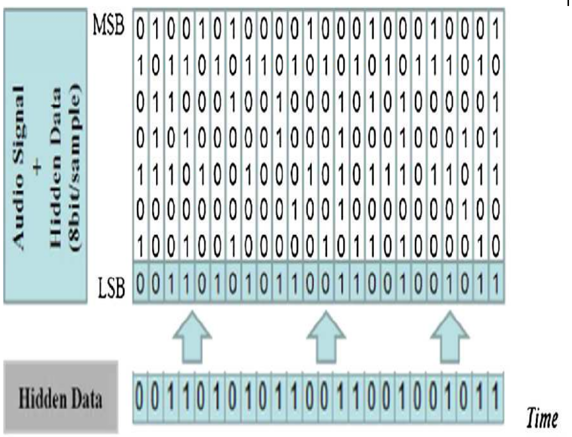

Fig. 1. LSB in 8 bits per sample signal is overwritten by one bit of the hidden data.

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. Furthermore, since data are embedded in a very deterministic way, an attacker can easily uncover the message by just removing the entire LSB plane.

2.1.2. Echo hiding

Echo hiding method embeds data into audio signals by introducing a short echo to the host signal. The nature of the echo is a resonance added to the host audio. Therefore, the problem of the HAS sensitivity to the additive noise is avoided. After the echo has been added, the stego signal retains the same statistical and perceptual characteristics. 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 [2] (Figure 2). For a delay up to 1 ms between the original signal and the echo, the effect is indistinguishable.

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

2.1.3. Hiding in silence intervals

In [3], a simple and effective embedding method has been used to exploit silence intervals in speech signal. Initially, the silence intervals of the speech and their respective lengths (the number of samples in a silence interval) are determined. These values are decreased by a value x where 0 < x < 2nbits, and n bits is the number of bits needed to represent a value from the message to hide. For the extraction process x is evaluated as mod (New Interval Length, 2nbits).

For example, if we want to hide the value 6 in a silence interval with length=109, then remove 7 samples from this interval which makes the new interval length 102 samples. To extract the hidden data from this silent interval in the stego-signal, and compute mod (102, 8) = 6. Small silence intervals are left unchanged since they usually occur in continuous sentences and changing them might affect the quality of the speech. This method has a good perceptual transparency but obviously it is sensitive to compression.

Strength and weakness of temporal domain

LSB method is simple and easy way of hiding information with high bit rate. Echo hiding and silence intervals method is resilient to lossy data compression algorithm.

Although robustness and security are not the main characteristics of temporal domain steganographic methods, conventional LSB technique and its variants provide an easy and simple way to hide data. Tolerance to noise addition at low levels and some robustness criteria have been achieved with LSB variants' methods [4-6], but at a very low hiding capacity. At present, only few time domain hiding techniques have been developed.

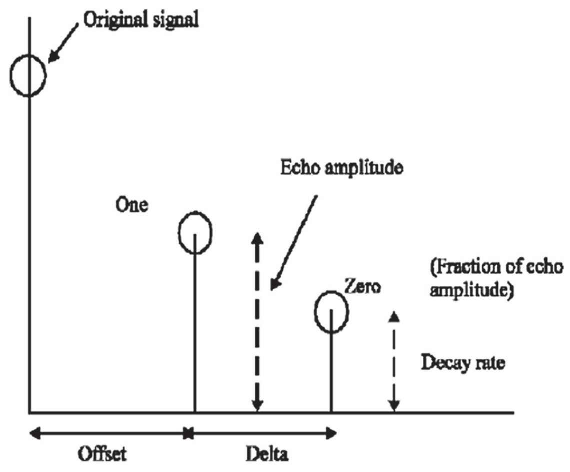

Fig. 2. Echo data hiding adjustable parameters

2.2. Hiding in transform domain

2.2.1. Spread spectrum

Spread spectrum technique spreads hidden data through the frequency spectrum. Spread spectrum (SS) is a concept developed in data communications to ensure a proper recovery of a signal sent over a noisy channel by producing redundant copies of the data signal. Basically, data are multiplied by an M-sequence code known to both sender and receiver [7], and then hidden in the cover audio. Thus, if noise corrupts some values, there will still be copies of each value left to recover the hidden message.

In [8], conventional direct sequence spread spectrum (DSSS) technique was applied to hide confidential information in MP3 and WAV signals. However, to control stego-audio distortion, [9,10] have proposed an embedding method where data are hidden under a frequency mask. In [9], spread spectrum is combined to phase shifting in order to increase the robustness of transmitted data against additive noise and to allow easy detection of the hidden data. For a better hiding rate, [10] used SS technique in the sub-band domain. Appropriately chosen sub-band coefficients were selected to address robustness and resolve synchronization uncertainty at the decoder.

2.2.2. Discrete wavelet transform

Audio steganography based on Discrete Wavelet Transform (DWT) is described in [11]. Data are hidden in the LSBs of the wavelet coefficients of the audio signals. To improve the imperceptibility of embedded data, [12] employed a hearing threshold when embedding data in the integer wavelet coefficients, while [13] avoided data hiding in silent parts of the audio signal. Even though data hiding in wavelet domain procures high embedding rate, data extraction at the receiver side might not be accurate.

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 [14, 15]. To embed one bit in an audio frame, this research suggests a pair of tones which is generated at two chosen frequencies f0 and f1. The power level of the two masked frequencies (pf0 and pf1) is set to a known ratio of the general power of each audio frame pi where: i = 1; :::n and n is the frame number as shown in Figure 3. By inserting tones at known frequencies and at low power level, concealed embedding and correct data extraction are achieved. To detect the tones and thus the hidden information from the stego-audio frames, the power pi for each frame is computed as well as the power pf0 and pf1 for the chosen frequencies f0 and f1. If the ratio, pi pf0 > pi pf1, then the hidden bit is ’0’, otherwise it is ’1’.

Tone insertion method can resist to attacks such as low-pass filtering and bit truncation. In addition to low embedding capacity, embedded data could be maliciously extracted since inserted tones are easy to detect. The authors suggest overcoming these drawbacks by varying four or more pairs of frequencies in a keyed order.

2.2.4. Phase Coding

Phase coding exploits HAS insensitivity to relative phase of different spectral components. It is based on replacing selected phase components from the original audio signal spectrum with hidden data. However, to ensure inaudibility, phase components modification should be kept small [16]. It is worth mentioning that among data hiding techniques, phase coding tolerates better signal distortion [1]. Authors in [16] have inserted data in phase components using an independent multi-band phase modulation. In this approach, imperceptible phase modifications are achieved using controlled phase alteration of the host audio as shown in Figure 4. Quantization index modulation (QIM) method is applied on phase components, where phase value of a frequency bin is replaced by the nearest o point to hide ’0’ or x point to hide ’1’. For greater embedding capacity, [17] has applied QIM on the phase of the strongest harmonic with a step size of =2n (Figure 5). Robustness to MP3 encoder with BER (Bit Error Rate) value near zero was also achieved. Despite the fact that phase quantization is robust to perceptual audio compression, HAS is not very sensitive to phase distortion [1]. Consequently, an intruder can also introduce imperceptible frequency modulation and eventually destroy the used phase quantization scheme.

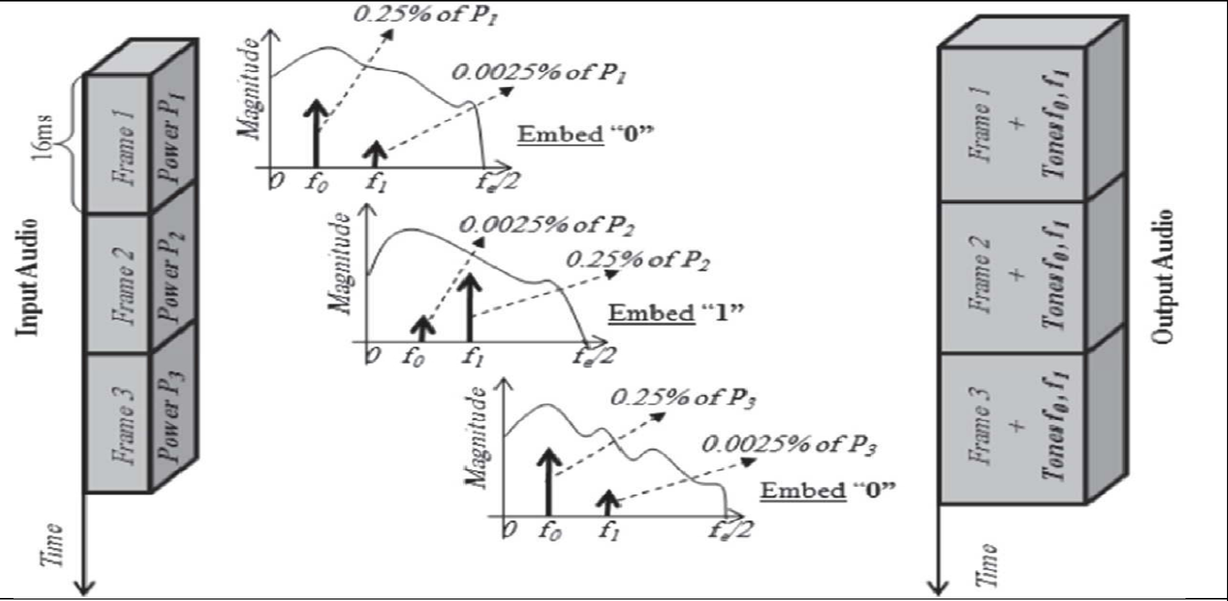

Fig. 3. Data embedding by inserting tones at two distinct frequencies

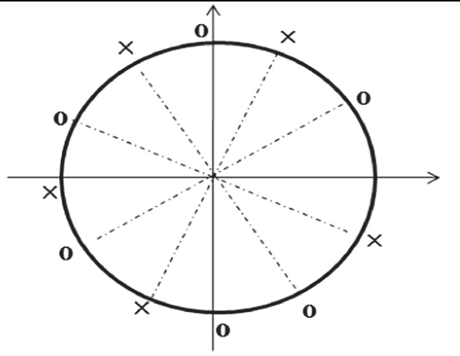

Fig. 4. Phase quantization

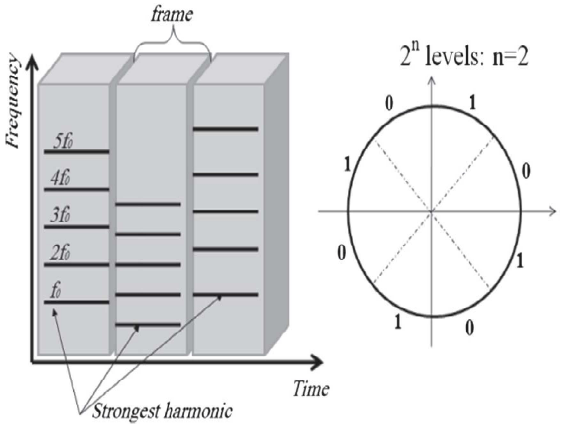

Fig. 5. Phase encoding for strongest harmonics.

2.2.5. Amplitude Coding

The HAS characteristics depend more on the frequency values as it is more sensitive to amplitude components. Following this principle, authors in [18] propose 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, [17] proposed a method to 11 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.

2.2.6. Cepstral Domain

Known also as log spectral domain, data in this method is embedded in the cepstrum coefficients which tolerate most common signal processing attacks. In addition, cepstrum alteration at frequencies that are in the perceptually masked regions of the majority of cover audio frames ensures inaudibility of the resulting stego audio frames. Employing cepstral domain modification is proposed in [19].

The cover signal is first transformed into cepstral domain then data are embedded in selected cepstrum coefficient by applying statistical mean manipulations. In this method, an embedding rate of 20 to 40 bps is achieved while guarantying robustness to common signal attacks. In [20], the cepstrums of two selected frequencies f1 and f2 in each energetic frame are modified slightly to embed bit ’1’ or ’0’. For more security of the embedded data, the author of the previous research, suggested later in [21] to use the latter algorithm and embed data with different arbitrary frequency components at each frame.

Strengths and Weaknesses of transform domain

It has been proven that hiding in frequency domain rather than time domain will give better results in terms of signal to noise ratio [1]. Indeed, audio steganography techniques in the transform domain benefit from the frequency masking effect. Most of data hiding algorithms based on transform domain use a perceptual model to determine the permissible amount of embedded data to avoid stego signal distortion. Although hidden data robustness against simple audio signal manipulation is the main characteristic of transform domain techniques, embedded data will unlikely survive noisy transmission environment or data compression induced by one of the encoding processes such us: ACELP, G.729, etc.

Magnitude spectrum has Longer message to hide and less likely to be affected by errors during transmission. But it is Low robustness to simple audio manipulations. Tone method has Imperceptibility and concealment of embedded data. But has lack of transparency and security. Phase spectrum is robust against signal processing manipulation and data retrieval needs the original signal. But it has Low capacity. Spread spectrum provides better robustness. But it is Vulnerable to time scale modification. Cepstral domain method is robust against signal processing operations but perceptible signal distortions and low robustness. Wavelet method provides high embedding capacity but it is lossy data retrieval.

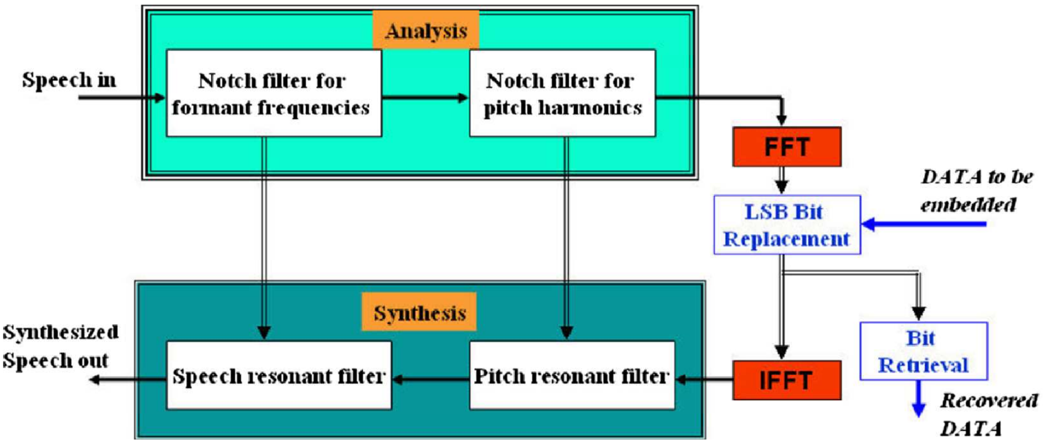

Fig. 6. Embedding Data in the LSB of the prediction residual

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. We distinguish two such techniques, namely in-encoder and post-encoder techniques, which we discuss thoroughly next.

2.3.1. In-Encoder Techniques

A research work where embedded data survives audio codec, compression, reverberations and background noises is presented in [22]. The technique hides data into speech and music signals of various types using subband amplitude modulation. Embedding data in the LPC vocoder was further proposed in [23]. The authors used an auto-correlation based pitch tracking algorithm to perform a voiced/unvoiced segmentation. They replaced the linear prediction residual in the unvoiced segments by a data sequence.

Once the residual’s power is matched, this substitution does not lead to perceptual degradation. The signal is conceived using the unmodified LPC filter coefficients. Linear prediction analysis of the received signal is used to decode hidden data. The technique offers a reliable hiding rate of 2kbps. Exploiting the LSB technique to hide data in the audio codecs is described in [24]. This technique embeds data in the LSB of the Fourier transform in the prediction residual of the host audio signal. An LPC filter is used to automatically shape the spectrum of LSB noise. Consequently, the noise generated by data hiding is substantially less audible in this system as depicted in Figure 6.

2.3.2. Post-Encoder Techniques

An alternative to in-encoder techniques is the post-encoder (or in-stream) techniques. To survive audio encoders, authors in [25] have embedded data in the bitstream of an ACELP codec. This technique hides data jointly with the analysis-by-synthesis codebook search. The authors applied the concept on the AMR encoder at a rate of 12.2 kbit/s and were able to hide 2 kbit/s of data in the bitstream. The quality of the stego speech is evaluated in terms of signal to noise ratio at 20.3 dB. A lossless steganography technique for G.711-PCMU telephony encoder has been proposed in [26]. Data in this case is represented by folded binary code which codes each sample with a value between -127 and 127 including -0 and +0. One bit is embedded in 8-bits sample which absolute amplitude is zero.

Depending on the number of samples with absolute amplitudes of 0, a potential hiding rate ranging from 24 to 400 bps is obtained. To increase the hiding capacity, the same authors have introduced a semi-lossless technique for G.711-PCMU [27], where audio sample amplitudes are amplified with a pre-defined level ’i’. The audio signals samples with absolute amplitudes vary from 0 to i are utilized in the hiding process. For a greater hiding capacity, [28] suggested to embed data in the inactive frames of low bit rate audio streams (i.e., 6.3 kbps) encoded by G.723.1 source codec.

Strengths and Weaknesses of coded domain

Robustness and security of embedded data are the main advantages of in-encoder approaches. Hidden data survives noise addition and audio codecs such as ACELP, AMR or LPC. Some of the coded domain methods have achieved a considerably high hiding capacity comparing to the used codecs rate. Since hidden data are not affected by the encoding process, data-extraction correctness is fulfilled in tandem-free operation. Codebook modification and bitstream hiding method is robust.

Despite their robustness, hidden data integrity in in-encoder audio steganography techniques could be compromised if a voice encoder/decoder (transcoding) exists in the network. Furthermore, hidden data could be also subject to transformation if a voice enhancement algorithm such as echo or noise reduction is deployed in the network. Since bitstream is more sensitive to modifications than the original audio signal, the hiding capacity should be kept small to avoid embedded data perceptibility. Coded domain techniques are well suited for real-time applications. But these methods has low embedding rate.

3. CONCLUSION

In order to provide better protection to digital data content, new steganography techniques have been investigated in recent researcher works. The availability and popularity of digital audio signals have made them an appealing choice to convey secret information. Audio steganography techniques address issues related to the need to secure and preserve the integrity of data hidden in voice communications in particular. In this work, a comparative study of the current-state-of-the-art literature in digital audio steganography techniques and approaches is presented. In an attempt to reveal their capabilities in ensuring secure communications, we discussed their strengths and weaknesses.

Also, a differentiation between the reviewed techniques based on the intended applications has been highlighted. Thus, while temporal domain techniques, in general, aim to maximize the hiding capacity, transform domain methods exploit the masking properties in order to make the noise generated by embedded data imperceptible. On the other side, encoded domain methods strive to ensure the integrity of hidden data against challenging environment such as real time applications. To better estimate the robustness of the presented techniques, a classification based on their occurrence in the voice encoder is given. The frequency domain is preferred over the temporal domain and music signals are better covers for data hiding in terms of capacity, imperceptibility and undetectability. The advantage on using one technique over another one depends on the application constraints in use and its requirement for hiding capacity, embedded data security level and encountered attacks resistance.

ACKNOWLEDGEMENTS

The authors would like to thanks to the earlier work regarding different Steganography methods that contribute the work made in this paper. All work done in this paper will surely help to the researchers for future work on Steganography methods.

REFERENCES

[1] Rakhi, Suresh Gawande, A Review On Steganography Methods, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 10, October 2013.

[2] W Bender, D Gruhl, N Morimoto, A Lu, Techniques for Data Hiding. IBM Syst. J. 35(3 and 4), 313–336 (1996).

[3] D Gruhl, W Bender, Echo hiding, Proceeding of the 1st Information Hiding Workshop, Lecture Notes in Computer Science, (Isaac Newton Institute, England, 1996), pp. 295–315.

[4] S Shirali-Shahreza, M Shirali-Shahreza, Steganography in Silence Intervals of Speech, proceedings of the Fourth IEEE International Conference on Intelligent Information Hiding and Multimedia Signal (IIH-MSP 2008). (Harbin, China, August 15-17, 2008), pp. 605–607.

[5] N Cvejic, T Seppanen, Increasing Robustness of, LSB Audio Steganography Using a Novel Embedding Method, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC04). vol. 2, (Washington, DC, USA, 2004), pp. 533.

[6] N Cvejic, T Seppanen, Reduced distortion bit-modification for LSB audio steganography., J. Universal Comput. Sci. 11(1), 56–65 (2005).

[7] MA Ahmed, LM Kiah, BB Zaidan, AA Zaidan, A Novel Embedding Method to Increase Capacity and Robustness of Low-bit Encoding Audio Steganography Technique Using Noise Gate Software Logic Algorithm., J. Appl. Sci. 10, 59–64 (2010).

[8] K Khan, Cryptology and the origins of spread spectrum, IEEE Spectrum. 21, 70–80 (1984).

[9] S Hernandez-Garay, R Vazquez-Medina, LN de Rivera, V Ponomaryov, Steganographic communication channel using audio signals, 12th International Conference on Mathematical Methods in Electromagnetic Theory, (MMET). (Odesa, Ukraine, 2 July 2008), pp. 427–429.

[10] H Matsuka, Spread spectrum audio steganography using sub-band phase shifting, in IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP’06) (Pasadena, CA, USA, December 2006), pp. 3–6.

[11] X Li, HH Yu, Transparent and robust audio data hiding in sub band domain, Proceedings of the Fourth IEEE International Conference on Multimedia and Expo, (ICME 2000), (New York, USA, 2000), pp. 397–400.

[12] 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, (Georgia, USA, 13-16 October, 2002), pp. 53–55.

[13] M Pooyan, A Delforouzi, Adaptive Digital Audio Steganography Based on Integer Wavelet Transform, Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2007). vol. 2, (Splendor Kaohsiung, Taiwan, 2007), pp. 283–28.

[14] S Shirali-Shahreza, M Shirali-Shahreza, High capacity error free wavelet domain speech steganography, Proc. 33rd Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008). (Las Vegas, Nevada, USA, 30 March 2008), pp. 1729–1732.

[15] K Gopalan, et al, Covert Speech Communication Via Cover Speech By Tone Insertion, Proceeding of IEEE Aerospace Conference, (Big Sky, Montana, March 2003).

[16] K Gopalan, S Wenndt, Audio Steganography for Covert Data Transmission by Imperceptible Tone Insertion, WOC 2004, (Banff, Canada, July 8–10, 2004).

[17] L Gang, AN Akansu, M Ramkumar, MP3 resistant oblivious steganography, Proceedings of, IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 3, (Salt Lake City, UT. 7-11 May 2001), pp. 1365–1368.

[18] X Dong, M Bocko, Z Ignjatovic, Data hiding via phase manipulation of audio signals, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’04). vol. 5, (Montreal, Quebec, Canada, 17-21 May 2004), pp. 377–380.

[19] F Djebbar, B Ayad, K Abed-Meraim, H Habib, Unified phase and magnitude speech spectra data hiding algorithm, Accepted in Journal of Security and Communication Networks. (John Wiley and Sons, Ltd, 4 April, 2012).

[20] X Li, HH Yu, Transparent and robust audio data hiding in cepstrum domain, Proc. IEEE International Conference on Multimedia and Expo, (ICME 2000), (New York, USA, 2000)

[21] K Gopalan, Audio Steganography by Cepstrum Modification, Proc. of the IEEE 2005 International Conference on, Acoustics, Speech, and Signal Processing (ICASSP’05), (Philadelphia, USA, March 2005).

[22] K Gopalan, A unified audio and image steganography by spectrum modification, IEEE International Conference on Industrial Technology (ICIT’09), (Gippsland, Australia, 10-13 Feb 2009), pp. 1–5.

[23] A Nishimura, Data hiding for audio signals those are robust with respect to air transmission and a speech codec, IIH-MSP’08. (Harbin, China, 15-17 Aug 2008), pp. 601–604.

[24] K Hofbauer, G Kubin, High-rate data embedding in unvoiced speech, in, Proc. Int. Conf. Spoken Language Processing (INTERSPEECH), (Pittsburgh, PY, USA, September 2006), pp. 241–244

[25] GS Kang, TM Moran, DA Heide, Hiding Information Under Speech, Naval Research Laboratory, (Washington, DC NRL/FR/5550–05-10, 126, 2005), 20375-5320.

[26] B Geiser, P Vary, High rate data hiding in, ACELP speech codecs, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’08), (Las Vegas, USA, 4 April 2008), pp. 4005–4008.

[27] N Aoki, A Technique of Lossless Steganography for G.711 Telephony Speech, International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP’08), (Harbin, China, 2008), pp. 608–611.

[28] N Aoki, A Semi-Lossless Steganography Technique for G.711 Telephony Speech, International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2010), (Darmstadt, Germany, 2010), pp. 534–537.

[29] YF Huang, S Tang, J Yuan, Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec, IEEE Trans. Inf. Forensics and Security. 6(2), 296–306 (2011).