Comparative Study of Digital Audio Steganography Techniques

Fatiha Djebbar*[1] and Beghdad Ayad[2]and Karim Abed Meraim[3]and Habib Hamam[4]

[1]UAE university, UAE
[2]Al ain university, Al Ain, UAE
[3]Telecom ParisTech, Paris, France
[4]Faculty of Engineering Université de Moncton, Moncton, NB, Canada

Email: Fatiha Djebbar*- fdjebbar@uaeu.ac.ae;

*Corresponding author

Abstract

The rapid spread in digital data usage in many real life applications have urged new and effective ways to ensure their security. Efficient secrecy can be achieved, at least in part, by implementing steganograhy techniques. Novel and versatile audio steganographic methods have been proposed. The goal of steganographic systems is to obtain secure and robust way to conceal high rate of secret data. We focus in this paper on digital audio steganography, which has emerged as a prominent source of data hiding across novel telecommunication technologies such as covered voice-over-IP, audio conferencing, etc. The multitude of steganographic criteria has led to a great diversity in these system design techniques. In this paper, we review current digital audio steganographic techniques and we evaluate their performance based on robustness, security and hiding capacity indicators. Another contribution of this paper is the provision of a robustness-based classification of steganographic models depending on their occurrence in the embedding process. A survey of major trends of audio steganography applications is also discussed in this paper.

1 Introduction

The growing use of Internet among public masses and the abundant availability of public and private digital data has driven industry professionals and researchers to pay a particular attention to data protection. Currently, three main methods are being used: cryptography, watermarking, and steganography. Cryptography techniques are based on rendering the content of a message garbled to unauthorized people. In watermarking, data are hidden to convey some information about the cover medium such as ownership and copyright. Even though cryptography and watermarking techniques are salient for reinforcing data security, a heightened interest in exploring better or complementary new techniques has been the focus of much ongoing research. Figure 1 exhibits the differences and the similarities between steganography, watermarking and cryptography. The terminology used for steganography blocks was imposed for the first time at the first international conference on information hiding [1].

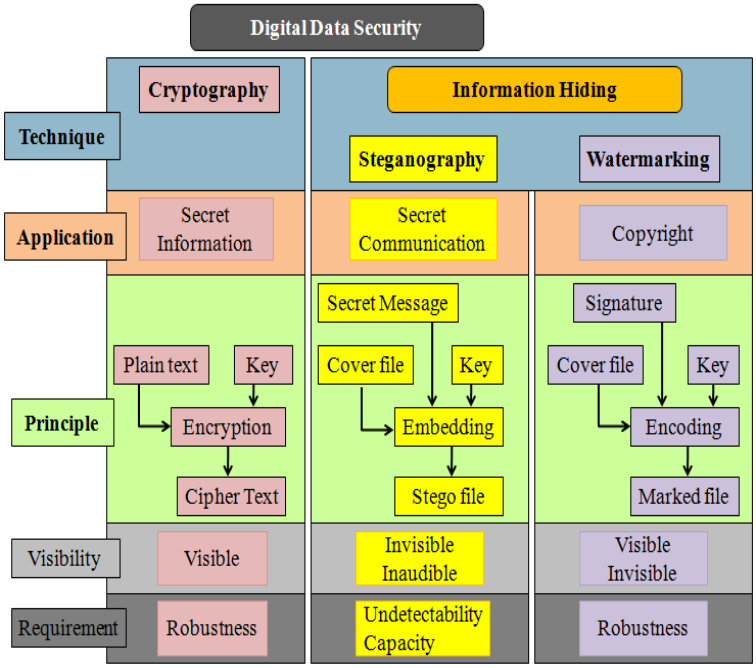


Figure 1: Digital data security disciplines.

The primary goal of steganography is to reliably send hidden information secretly, not merely to obscure its presence. Steganography in today's computer era is considered a sub-discipline of data communication security domain. Lately, new directions based on steganographic approaches started to emerge to ensure data secrecy. Rather than as a substitute to existing solutions, these approaches could achieve better data secrecy if combined with conventional security techniques. Modern techniques of steganography exploit the

characteristics of digital media by utilizing them as carriers (covers) to hold hidden information. Covers can be of different types including image, audio, video, text, and IP datagram. An example of audio steganography is depicted in Figure 2, where the cover file in use is a digital audio file. The sender embeds data of any type in a digital cover file using a key to produce a stego-file, in such a way that an observer cannot detect the existence of the hidden message [2]. At the other end, the receiver processes the received stego-file to extract the hidden message. An obvious application of such steganographic system is a covert communication using innocuous cover audio signal, such as telephone or video conference conversations.
To minimize the difference between the cover- and the stego-medium, recent steganography techniques utilize natural limitations in human auditory and visual perceptions. Image and video based steganography rely on the limited human visual system to notice luminance variation at levels greater than 1 in 240 across uniform grey levels, or 1 in 30 across random patterns [2]. However, audio-based steganography exploits the masking effect property of the Human Auditory System (HAS) [3] as explained later in this paper.
Various features influence the quality of audio steganographic methods. The importance and the impact of each feature depend on the application and the transmission environment. The most important properties include robustness to noise, to compression and to signal manipulation, as well as the security and the hiding-capacity of hidden data. The robustness requirement is tightly coupled with the application, and is also the most challenging requirement to fulfill in a steganographic system when traded with data hiding-capacity. Generally, the robustness and the capacity hardly coexist in the same steganographic system due to tradeoffs imbalance between these two criteria where increased robustness levels result in decreasing data hiding capacity [2].
In this work, several works in audio steganography are discussed as well as a thorough investigation of the use of audio files as a cover medium for secret communications. The present review paper builds on our previous work [4], however, our contributions are as follows:

- We survey latest audio steganographic methods and reveal their strengths and weaknesses.
- We propose a classification of the reviewed audio steganographic techniques relative to their occurrence in voice encoders.
- We compare steganographic methods based on selected robustness criteria.
- We evaluate the performance of the reviewed steganographic techniques.

The remainder of this paper is organized as follows: Section 2 presents the motivations related to the use of

audio signals as carriers as well selecting some performance criteria used to assess hidden data tolerance to common signal manipulations. Section 3 presents reviewed steganography methods. However, Section 4 proposes a classification of existing audio steganographic techniques based on their occurrence instances in voice encoders. Evaluation and possible applications are presented in Section 5 and 6. Finally, conclusions and future work are presented in Section 7.

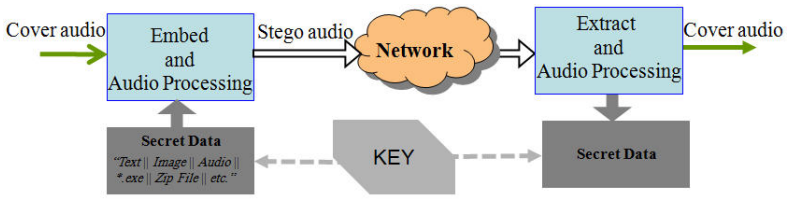


Figure 2: Audio steganography workflow.

2 Motivation and Background

2.1 Audio File as a Cover

The particular importance of hiding data in audio files results from the prevailing presence of audio signals as information vectors in our human society. Prudent steganography practice assumes that the cover utilized to hide messages should not raise any suspicion to opponents. In fact, the availability and the popularity of audio files make them eligible to carry hidden information. In addition, most steganalysis efforts are more directed towards digital images leaving audio steganalysis relatively unexplored. Data hiding in audio files is especially challenging because of the sensitivity of the HAS. However, HAS still tolerates common alterations in small differential ranges. 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 researchers to explore the utilization of audio signals as carriers to hide data [4–9]. The alterations of audio signals for data embedding purposes may affect the quality of these signals. Assessing the tradeoffs between these alterations and the induced quality is discussed next.

2.2 Comparison Criteria

Various parameters influence the quality of audio steganographic systems. Besides, the amount of the hidden data and its imperceptibility level, robustness against removal or destruction of embedded data remains the most critical property in a steganographic system. The robustness criteria are assessed

through the survival of concealed data to noise, compression and manipulations of the audio signal (e.g., filtering, re-sampling, re-quantization). In this section, we discuss some selected comparison criteria between the cover- and the stego-signals. We only focus on those methods' properties that have been evaluated and verified in the reviewed techniques. These properties are listed as follows:

- Hiding rate: Measured in bps and refers to the amount of concealed data (in bits) within a cover audio signal, and correctly extracted.
- Imperceptibility: This concept is based on the properties of the HAS which is measured through perceptual evaluation of speech quality ($PESQ$) [1]. The hidden information is imperceptible if a listener is unable to distinguish between the cover- and the stego-audio signal. The PESQ test produces a value ranging from 4.5 to 1. A $PESQ$ value of 4.5 means that the measured speech has no distortion, it is exactly the same as the original. A value of 1 indicates the severest degradation. Another measure which is widely used is the level of distortion in audio signals and it is captured through $SegSNR$ [2](i.e., Signal to Noise Ratio) [10]. It is important that the embedding process occurs without a significant degradation or loss of perceptual quality of the cover signal.
- Amplification: This criterion results in increasing the magnitude of the audio signal which could alter the hidden data if a malicious attack is intended.
- Filtering: Maliciously removes the hidden data by cutting-off selected part of the spectrum.
- Re-quantization: This parameter modifies the original quantization of the audio signal. For example, a 16 bits audio signal is quantized to 8 bits and back to 16 bits in an attempt to destroy the hidden data.
- Re-sampling: Similarly to the above operation, this parameter triggers the sampling frequency of the audio signal to another one, i.e., wideband audio signal sampled at 16 kHz to 8 kHz and back to 16 kHz.
- Noise addition: Adding noise to the audio signal in an attempt to destroy the hidden data, i.e., WGN (White Gaussian Noise).
- Encoding/Decoding: This operation reduces the amount of data by removing redundant or unnecessary information. Thus, a hidden message can be completely destroyed. This is also true if the audio file is converted into another format. MP3 compression, for example, changes a wave file to an MP3 file before it reaches the receiver.
- Transcoding: It is the process of decoding the audio signal with a decoder that is different than the one used in the encoding operation.

3 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. A summary evaluation of these techniques based on the selected comparison criteria is presented in Table 1, Table 2 and Table 3.

[1]Standard ITU-T P862.2
[2]Segmental SNR

3.1 Hiding in Temporal Domain

The majority of temporal domain methods employ low-bit encoding techniques, which we describe next. Other candidate techniques that fall under temporal domain category are also presented in the subsequent sections.

3.1.1 Low-bit Encoding

Also known as LSB (Least Significant Bit), this method is one of the earliest methods used for information hiding [2]. 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 3). 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. Furthermore, since data are embedded in a very deterministic way, an attacker can easily uncover the message by just removing the entire LSB plane. In [11], a simple LSB strategy has been applied to embed a voice message in a wireless communication. While this method achieves the imperceptibility at high embedding rate, the security and robustness of hidden data are easily compromised. In an attempt to augment the hiding capacity while minimizing the error on the stego audio, [12] adopted a minimum error-replacement method while embedding four bits per sample. The embedding error is then diffused on the next four samples.

Figure 3: LSB in 8 bits per sample signal is overwritten by one bit of the hidden data.

To improve the robustness of LSB method against distortion and noise addition, [13–15] have increased the depth of the embedding layer from 4th to 6th and to 8th LSB layers without affecting the perceptual transparency of the stego audio signal. In [13, 14], only bits at the sixth position of each 16 bits sample of the original host signal are replaced with bits from the message. To minimize the embedding error, the other bits can be flipped in order to have a new sample that is closer to the original one. For example, if the original sample value was 4 which is represented in binary by "0100", and the bit to be hidden into the

4th LSB layer is 1, instead of having the value 12='1100' produced by the conventional LSB algorithm, the proposed algorithm produces a sample that has value 3= '0011', which is much closer to the original sample value (i.e., 4). On the other hand, [15] has shifted the LSB embedding to the eighth layer and has avoided hiding in silent periods or near silent points of the host signal. The occurrence of embedding instances in the eighth bit will slightly increase the robustness of this method compared to the conventional LSB methods. However, the hiding capacity decreases since some of the samples have to be left unaltered to preserve the audio perceptual quality of the signal. In addition, the easiness of the hidden message retrieval is still one of the major drawback of the LSB and its variants, if the hidden bits at the sixth or the eighth position are maliciously revealed out of the stego audio signal.

3.1.2 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 [16] (Figure 4). 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 restrict its related application domains. Hence, the limited amount of works which investigate the application of this method.

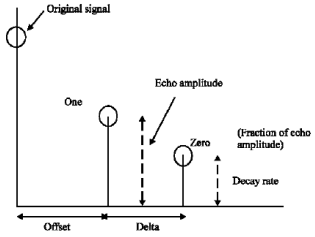


Figure 4: Echo data hiding adjustable parameters [16].

Due to the low embedding rate and security, and to the best of our knowledge, no audio steganography system based on echo hiding has been presented in recent research works. Moreover, only few techniques have been proposed, even for audio watermarking. To improve the watermark system robustness against

common linear signal processing, an echo hiding-time spread technique has been proposed in [17]. Compared to the conventional echo-hiding system, this proposed method spreads the watermark bits throughout the whole signal and it recover them based on the correlation amount at the receiver. The presented system is cepstral content based in which the original signal cepstral portion of error is removed at the decoder which leads to a better detection rate.

3.1.3 Hiding in Silence Intervals

In [18], 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 $0 < x < 2^{nbits}$, and $nbits$ is the number of bits needed to represent a value from the message to hide. For the extraction process x is evaluated as $mod(NewIntervalLength, 2^{nbits})$. For example, if we want to hide the value 6 in a silence interval with length=109, we 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, we 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. Changes in silence intervals length will lead to false data extraction. To overcome this shortcoming, [19] 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 are simply ignored in data hiding and retrieval.

3.1.4 Strengths and Weaknesses of Temporal Domain Methods

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 [13–15], but at a very low hiding capacity. At present, only few time domain hiding techniques have been developed. An evaluation of steganographic systems based on these techniques is shown in Table 1. The presence of (✓) sign denotes that the property is validated while (-) indicates the inverse or the information is unavailable.

Table 1: Temporal Domain: Methods Comparison

Method properties	Conventional LSB	LSB's variants	Silence intervals
imperceptibility	✓ [11]	✓ [13, 15]	✓ [18, 19]
WGN addition	-	✓ [15]	✓ [19]
Compression	-	-	✓ [19]

3.2 Hiding in Transform Domain

The human auditory system has certain peculiarities that must be exploited for hiding data effectively. The "masking effect" phenomenon masks weaker frequencies near stronger resonant ones [20, 21]. Several methods in the transform domain have been proposed in the literature as described next. To achieve the inaudibility, these methods exploit the frequency masking effect of the HAS directly by explicitly modifying only masked regions [7, 24, 25, 27] or indirectly [29, 36] by altering slightly the audio signals samples.

3.2.1 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 [22], 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 [23], conventional direct sequence spread spectrum (DSSS) technique was applied to hide confidential information in MP3 and WAV signals. However, to control stego-audio distortion, [24, 25] have proposed an embedding method where data are hidden under a frequency mask. In [24], 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, [25] 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.

3.2.2 Discrete Wavelet Transform

Audio steganography based on Discrete Wavelet Transform (DWT) is described in [26]. Data are hidden in the LSBs of the wavelet coefficients of the audio signals. To improve the imperceptibility of embedded data, [27] employed a hearing threshold when embedding data in the integer wavelet coefficients, while [28] 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.

3.2.3 *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 [29, 30]. To embed one bit in an audio frame, this research suggests a pair of tones which is generated at two chosen frequencies f_0 and f_1. The power level of the two masked frequencies (p_{f0} and p_{f1}) is set to a known ratio of the general power of each audio frame p_i where: $i = 1, ...n$ and n is the frame number as shown in Figure 5. 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 p_i for each frame is computed as well as the power p_{f0} and p_{f1} for the chosen frequencies f_0 and f_1. If the ratio, $\frac{p_i}{p_{f0}} > \frac{p_i}{p_{f1}}$, then the hidden bit is '0', otherwise it is '1'.

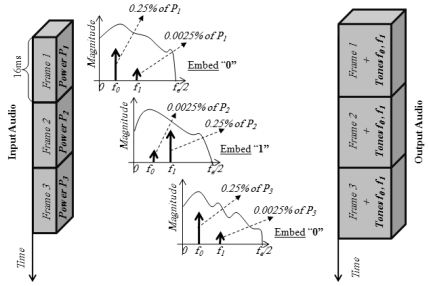


Figure 5: Data embedding by inserting tones at two distinct frequencies

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 to overcome these drawbacks by varying four or more pairs of frequencies in a keyed order.

3.2.4 *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 [31]. It is worth mentioning that among data hiding techniques, phase coding tolerates better signal distortion [2]. Authors in [31] 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 6. 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'.

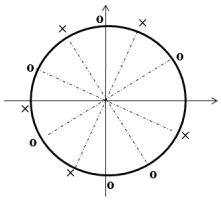


Figure 6: Phase quantization [31].

For greater embedding capacity, [32] has applied QIM on the phase of the strongest harmonic with a step size of $\pi/2^n$ (Figure 7). 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 [2]. Consequently, an intruder can also introduce imperceptible frequency modulation and eventually destroy the used phase quantization scheme.

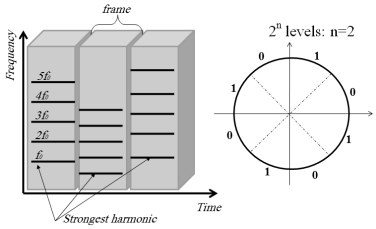


Figure 7: Phase encoding for strongest harmonics.

3.2.5 Amplitude Coding

The HAS characteristics depend more on the frequency values as it is more sensitive to amplitude components. Following this principle, authors in [7] 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, [33] 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.

3.2.6 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 [34]. 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 [35], the cepstrums of two selected frequencies f_1 and f_2 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 [36] to use the latter algorithm and embed data with different arbitrary frequency components at each frame.

3.2.7 Allpass Digital Filters

Using allpass digital filters (APFs), authors in [37] embed data in selected subbands using distinct patterns of APF. The proposed scheme is robust against: noise addition, random chopping, re-quantization and re-sampling. To further increase the robustness of this hiding scheme, a set of n_{th} order APFs were used in [38]. The value of n is an even positive integer and pole locations may be chosen in a variety of ways. data are embedded in selected APF parameters and retrieved using the power spectrum to estimate APF pole locations.

3.2.8 Strengths and Weaknesses

It has been proven that hiding in frequency domain rather than time domain will give better results in terms of signal to noise ratio [2]. 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. A great number of transform domain have been presented in the last decade and to a certain extent, these techniques have succeeded in realizing the security and the robustness of hidden data against simple audio signal manipulations such as amplification, filtration or re-sampling as shown in Table 2.

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.

Table 2: Transform Domain: Criteria comparison

Method properties	**Tone insertion**	**Phase coding**	**Amplitude coding**	**Cepstral Domain**	**SS**	**APFs**	**DWT**
imperceptibility	✓ [30]	✓ [31, 32]	✓ [33]	✓ [35]	✓ [24, 25]	✓ [37, 38]	✓ [27, 28]
Amplification	-	✓ [32]	-	✓ [36]	-	-	-
Noise addition	-	-	-	✓ [35]	✓ [24]	✓ [37, 38]	-
Low pass filtering	✓ [30]	-	-	✓ [35]	-	✓ [37, 38]	-
Requantization	-	✓ [31, 32]	-	-	-	✓ [37, 38]	-
Re-sampling	-	-	-	-	-	✓ [37, 38]	-
Compression	-	✓ [31]	-	✓ [35, 36]	-	✓ [37, 38]	-

3.3 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.

3.3.1 In-Encoder Techniques

A research work where embedded data survives audio codec, compression, reverberations and background noises is presented in [39]. 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 [40]. 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 [20]. 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 8.

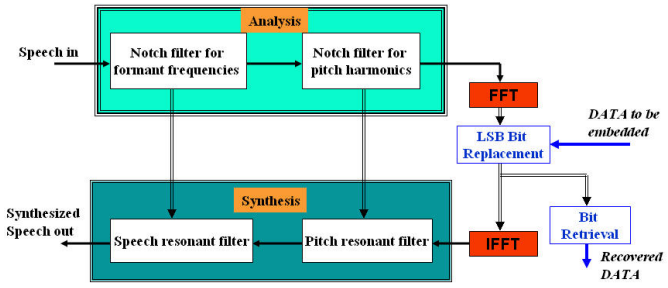


Figure 8: Embedding Data in the LSB of the prediction residual.

3.3.2 *Post-Encoder Techniques*

An alternative to in-encoder techniques is the post-encoder (or in-stream) techniques. To survive audio encoders, authors in [41] 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 [42]. 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 [43], where audio sample amplitudes are amplified with a pre-defined level 'i'. The audio signal samples with absolute amplitudes vary from 0 to i are utilized in the hiding process. For a greater hiding capacity, [44] 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.

3.3.3 *Strengthes and Weaknesses*

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.

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. Table 3 summarizes coded domain techniques based on selected robustness criteria.

Table 3: Codecs based techniques: Criteria's comparison

Method properties	**In-Encoder**	**Post-Encoder**
Imperceptibility	✓ [20, 39]	✓ [41]
Noise addition	✓ [39]	✓ [41]
Decoding/Encoding	✓ [39, 40]	✓ [42]

4 Classification of Audio Steganography Methods

Robustness, security and hiding capacity are the three major performance criteria that revolve around the existing steganography methods. To categorize and evaluate the above-discussed methods considering these criteria, the transmission environment and the application in use are considered. Covert communication for example requires high level of robustness due to the passage of data by one of the existing coders that can heavily affect the integrity of the transmitted data. The encoder process reduces the amount of data in the audio signal by eliminating redundant or unnecessary data. Resisting the encoder/decoder processes is hard to satisfy and when fulfilled it is usually done at the cost of the hiding capacity. Thus, we choose to study the behavior of the reviewed steganography methods with respect to their occurrence in the coders as shown in Figure 9. The security aspect of each method is evaluated by a third party effort cost to retrieve the embedded data. Three distinct embedding groups are used when designing data-in-audio steganograhic system [41], which we explain next.

4.1 Pre-Encoder Embedding

The pre-encoder methods apply to time and frequency domains where data embedding occurs before the encoding process. A greater part of the methods belonging to pre-encoder embedding class does not guarantee the integrity of the hidden data over the network. Noise addition in its different forms (e.g., WGN) and high-data rate compression induced by one of the encoding processes such us ACELP or G.729, will likely affect the integrity of embedded data. In other methods, embedded data resists only to few audio manipulations such as resizing, re-sampling, filtering etc, and they only tolerate noise addition or

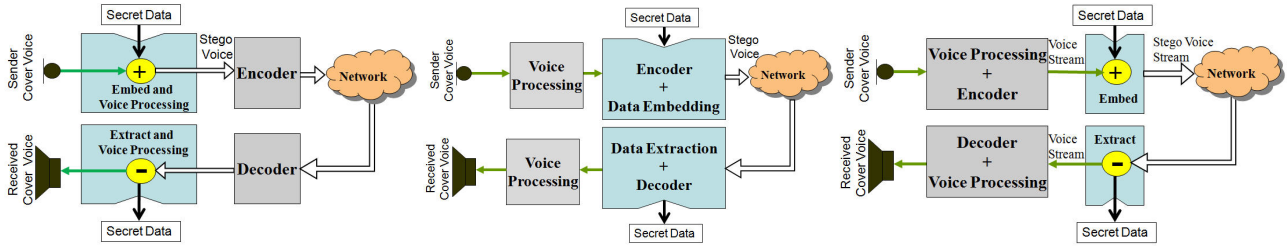
data compression at very low rate. High embedding data rate can be achieved with methods designed for noise-free environments.

4.2 In-Encoder Embedding

The robustness of embedded data are the main advantage of this approach. This approach is based on data-embedding operation within the codebook of the codecs. The transmitted information is hidden in the codebook parameter after a re-quantization operation. Thus, each audio signal parameter has a double significance: embedded-data value and audio codebook parameter. One of the drawbacks of this method arises when the encoded parameters traverse a network such as GSM that have for example a voice decoder/encoder in the Radio Access Network (BST, BSC, TRAU) and/or in the Core Network (MSC). In this configuration, hidden data values will be modified. These modifications might also happen when a voice enhancement algorithm is enabled in the Radio Access Network and/or in the Core Network.

4.3 Post-Encoder Embedding

In this approach, data are embedded in the bitstream resulting from the encoding process and extracted before traversing the decoder side. Since the bitstream is more sensitive to modifications than the original audio signal, the hiding capacity should be kept small to avoid embedded data perceptibility. Furthermore, transcoding can modify embedded data values and therefore could alter the integrity of the steganographic system. However, one of the positive sides of these methods is the correctness of data retrieval. Hidden message-extraction is done with no loss in tandem-free operations since it is not affected by the encoding process. A general scheme of the three steganography approaches is illustrated in Figure 9.



(a) Pre-encoder embedding (b) In-encoder embedding (c) Post-encoder embedding

Figure 9: General audio steganography approaches

To sum up strengths and weaknesses of the reviewed techniques, Table 4 focuses on factors such as security against hostile channel attacks, robustness or larger hiding capacity depending on the application and the

channel transmission conditions.

Table 4: General recapitulation

Hiding domain	**Methods**	**Embedding techniques**	**Advantages**	**Drawbacks**	**Hiding rate**
Temporal domain	Low bit encoding	LSB of each sample in the audio is replaced by one bit of hidden information	Simple and easy way of hiding Information with high bit rate	Easy to extract and to destroy	16kbps
	Echo hiding	Embeds data by introducing echo in the cover signal	Resilient to lossy data compression algorithms	Low security and capacity	50bps
	Silence intervals	Uses the number of samples in silence interval to represent hidden data	Resilient to lossy data compression algorithms	Low capacity	64bps
Transform Domain	Magnitude spectrum	Use frequency bands to hide data	Longer message to hide and less likely to be affected by errors during transmission	Low robustness to simple audio manipulations	20Kbps
	Tone insertion	insertion of inaudible tones at selected frequencies	Imperceptibility and concealment of embedded data	Lack of transparency and security	250bps
	Phase spectrum	Modulate the phase of the cover signal	Robust against signal processing manipulation and data retrieval needs the original signal	Low capacity	333bps
	Spread spectrum	Spread the data over all signal frequencies	Provide better robustness	Vulnerable to time scale modification	20 bps
	Cepstral domain	Altering the cepstral coefficients for embedding data	Robust against signal processing operations	Perceptible signal distortions and low robustness	54bps
	Wavelet	Altering wavelet coefficients for embedding data	Provide high embedding capacity	lossy data retrieval	70kbps
Codecs domain	Codebook modification	Altering codebook parameters	Robust	Low embedding rate	2kbps
	Bitstream hiding	LSB is applied on the bitstream resulting from the encoder process	Robust	Low embedding rate	1.6kps

5 Audio Steganography Evaluation

To evaluate the performance of the reviewed techniques, the imperceptibility and the detectability rate of hidden data are assessed. Next, imperceptibility evaluation of selected temporal, transform and coded domain steganography tools and methods is discussed.

5.1 Imperceptibility Evaluation

The criteria segmental signal-to-noise ratio $SegSNR$ which represents the average of the SNRs of all modified audio signal frames and the $PESQ$ measure are used. The value of SegSNR indicates the distortion amount induced by the embedded data in the cover audio signal $s_c(m, n)$. In audio signals for example, an SNR below 20 dB, generally denotes a noisy audio signal, while an SNR of 30 dB and above

indicates that the audio signal quality is preserved. SNR value is given by the following equation:

$$SNR_{dB} = 10\log_{10}\left(\frac{\sum_{n=1}^{N}|s_c(m,n)|^2}{\sum_{n=1}^{N}|s_c(m,n)-s_s(m,n)|^2}\right) \quad (1)$$

$s_s(m,n)$ is the stego-audio signal where: $m = 1,...M$ and $n = 1,...N$. M is the number of frames in milliseconds (ms) and N is the number of samples in each frame. The SNR (dB) values and payload (kbps) are used to evaluate the methods. For that purpose, we use online available audio steganography software in [45–50]. We used a total of forty male and female 16 bits WAV format audio (speech and music) signals. The speech files are sampled at 16 kHz while music at 44.1 kHz. The duration of audio files varies between 4 to 10 s length, spoken in English by different male and female talkers. Our results (i.e., SNR and hiding rate) are recorded in Table 5. The noise level induced by the embedding operation in each software is depicted in Figure.10.

Hiding in speech, speech pauses or music audio signals as shown in Figures (10a), (10b), (10c) and in Table 5 indicates that Steganos software induces more noise, where H4PGP shows better performance in terms of SNR and hiding capacity. However, the other softwares behave almost alike. In addition, our results show that music signals are better hosts to hide data in terms of imperceptibility and capacity.

Software	Payload	SNR	PESQ
Invisible Secrets	7.8	58.1	4.499
Hide4PGP	7.8	53.5	4.500
s-tools	7.8	68.5	4.499
Steganos	7.8	13	3.517

(a) Hiding in speech embedding

Software	SNR
Invisible Secrets	41.9
Hide4PGP	42
s-tools	44.4
Steganos	-3

(b) Hiding in speech pauses

Software	Payload	SNR
Invisible Secrets	21	64.8
Hide4PGP	21	67.93
s-tools	21	67.9
Steganos	21	19.64

(c) Hiding in music

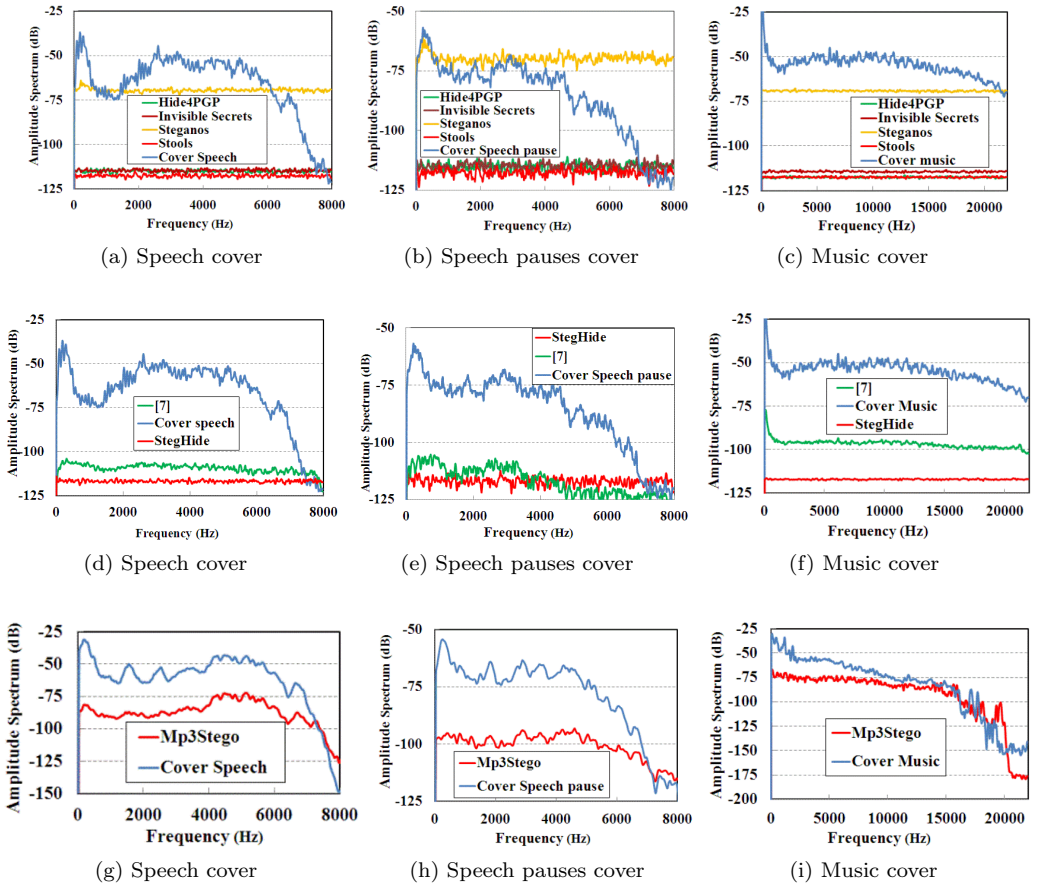
Audio type	Method	Payload	SNR	PESQ
Music	StegHide	21	67.8	-
	Mp3Stego	0.78	30.2	-
Speech	StegHide	5.86	60.5	4.499
	Mp3Stego	0.076	36	2.54
Speech pause	StegHide	5.86	44	-
	Mp3Stego	0.076	31.6	-

(d) Hiding in transform and coded domains

Table 5: Payload versus SNR in temporal domain (Table (5a), (5b) and (5c)) approaches depicted by each software tool appearing in [45–48] and in transform and coded domains (Table (5d) methods appearing respectively in [49,50].

To control the distortion induced by the embedding process, most audio steganography methods based on

transform domain use a perceptual model to determine the permissible amount of data embedding without distorting the audio signal. Previous investigations evaluated frequency domain method are reported in Figure 10. Related results are reported in Table (5d). In a more challenging environment, such as real time applications, encoded domain methods ensure robustness against compression. A similar performance investigation reports the results shown in Table (5d) and in Figures (10g), (10h) and (10i). Our results show that while using the same embedding capacity in temporal and frequency domains, stego signals generated in the frequency domain are less distinguishable than the ones produced by hiding data in the temporal domain.



(a) Speech cover (b) Speech pauses cover (c) Music cover

(d) Speech cover (e) Speech pauses cover (f) Music cover

(g) Speech cover (h) Speech pauses cover (i) Music cover

Figure 10: Noise level induced in speech (Figure 10a, Figure 10g) speech pause (Figure 10b, Figure 10h) and music (Figure 10c, Figure 10i) audio signal covers by data embedding using temporal (Stools, Stegnos and Hide4PGP), transform (Steghide and [7]) and encoded (Mp3Stego) domains steganographic tools

5.2 Evaluation by Steganalysis

Steganalysis is the science of detecting the presence of hidden messages. To investigate the delectability rates of steganographic algorithms presented in the above section, we use a reference audio steganalysis method presented in [51]. The selected reference method was applied successfully in detecting the presence of hidden messages in high capacity LSBs-based steganography algorithms. It allows the enhancement of the signal discontinuities due to the noise generated by the hidden data [51]. The method is based on extracting Mel-cepstrum coefficients (or features) from the second order derivative of audio signals. A support vector machine (SVM) with RBF kernel [52] is then applied to the features to distinguish between cover- and stego-audio signals. For each studied steganographic tool and algorithm, two datasets are produced: training and testing. Each dataset contains 350 stego and cover WAV audio signals of 10 s length. All signals are sampled at 44.1-kHz and quantized at 16-bits. Each training and testing dataset contains 175 positive (stego) and 175 negative (cover) audio signals. We used on-line audio files from different types such as speech signals in different languages (English, Chinese, Japanese, French, and Arabic) and music (classic, jazz, rock, blues). All stego-audio signals are generated by hiding data from different types: text, image, audio signals, video and executable files. To make a fair comparison between all assessed algorithms [47–49], the cover-signals were embedded with the same capacity of data. More precisely, S-Tools's with hiding ratio of 50% is used as a reference hiding capacity for the candidate steganographic algorithms and tools. The performance of each steganographic algorithm is measured through the levels by which the system can distinguish between the stego and the cover-audio signals (Table 7a). In order to analyze the obtained results, we first present the contingency table (see Table 6).

Table 6: The contingency table

	Stego-signal	**Cover-signal**
Stego classified	True positives (tp)	False negatives (fn)
Cover classified	False positives (fp)	True negatives (tn)

The entries of the contingency table are described as follows:

- *tp*: stego-audio classified as stego-audio signal
- *tn*: cover-audio classified as cover-audio signal
- *fn*: stego-audio classified as cover-audio signal
- *fp*: cover-audio classified as stego-audio signal

In subsequent formula, *all* represents the number of positive and negative audio signals. The value of the information reported in Table 6 is used to calculate the following measure:

$$Accuracy(AC) = \frac{tp + tn}{all} \tag{2}$$

Following the preparation of the training and testing datasets, we used the SVM library tool available at *http://www.csie.ntu.edu.tw/∼cjlin/libsvm* to discriminate between the cover- and the stego-audio signals. The results of the comparative study are reported in Table 7a. The accuracy of each studied tool is measured by the accuracy (AC). The values presented in Table 7a are the percentages of the stego-audio signals correctly classified. Higher score values are interpreted as high-detection rates. Consequently, frequency-domain steganography technique described in Steghide tool shows a performance improvement over time domain techniques (Stools and Hide4PGP). These results are consistent with our finding in the imperceptibility evaluation presented in the previous section.

Hiding methods	**AC**
Stools	0.73
Steghide	0.68
Hide4PGP	0.85

(a) Dataset of 350 music and speeches audio signals

Hiding methods	**Audio signal**	**AC**
Stools	Music	0.69
	Speech	0.77
Steghide	Music	0.63
	Speech	0.72
Hide4PGP	Music	0.79
	Speech	0.88

(b) Two separate datasets of 175 speeches and 175 music audio signals

Table 7: Overall steganalysis study results for data in audio (Table 7a), in speech signals only and in music only (Table 7b) depicted by each software tool appearing in [47–49].

In Table 7b, further investigation is done to put more emphasis on the behavior of the tested algorithms when music- and speech-audio signals are used separately to convey hidden data. The results show that hiding in music is less detectable than speech audio signals. In fact, the reference steganalysis method uses features extracted from high frequencies (lower in energy) to discriminate between cover- and stego-signals. Therefore, it allows to intensify the signal discontinuities due to the noise generated by data embedding. As the number of low-energy frequency components in music audio signals is smaller than that in speech audio-signals, the detection rate is expected to be lower.

6 Applications and Trends

A various range of audio steganographic applications have been successfully developed. Audio Steganography techniques can be applied for covert communications using unclassified channels without

additional demand for bandwidth or simply for storing data. In general, three application types for audio steganography techniques are distinguished and can be categorized as discussed next.

6.1 Secret Communication

To maintain patients' medical records secrecy, [53] proposed to telemedicine users, a multilevel-access control audio steganography system for securing transmission of medical images. The system embeds medical images in audio files that are sent to different recipients such as doctors in-charge of the corresponding patient. For more security, only intended receivers have the knowledge of a key that will be used to extract the medical images. To exploit the expanding use of audio multimedia messaging (MMS) among mobile phone users, [54] presented an alternative way for hidden communications, where data are hidden in text messages (SMS) or in MMS. However, in [55], a real time application that hides text in image and then disseminates it in MMS is presented. The system is created on a pair of Nokia 3110c handsets in Java 2 platform, micro edition (J2ME). The system makes use of the 4 last bits of a snapshot image taken by the camera phone to embed the message and then send it using a carrier medium such as MMS or Bluetooth. A preestablished key between the sender and the receiver is used to open the image and read the message. The general principle of MMS use in audio steganography is shown in Figure 11.

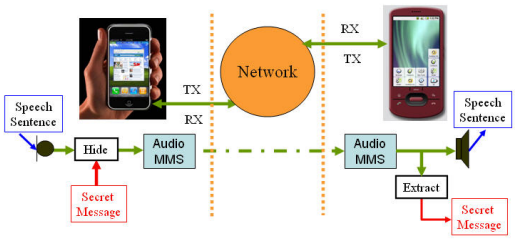


Figure 11: Audio steganography in MMS.

6.2 Improved Communication

In order to improve the intelligibility and the perceived quality of telephone speech (PSTN), [56, 57] proposed a data hiding technique to extend the PSTN channel bandwidth. Since human voice occupies 8 kHz or more in bandwidth, wideband speech (which lies in an interval of 50 Hz to 7 kHz) provides a higher intelligibility compared to narrowband speech (where the only information that could be transmitted is in the frequency band of 200 Hz to 3.5 KHz). Wideband speech is divided into three subbands: lower band (LB) 50-200, narrowband 0.2-3.5 and upper band (UB) 3.5-7 kHz. The characteristics (magnitude frequencies and their locations) of LB and UP are embedded in the narrowband part of the speech based

on a perceptual masking principle. While this hidden signal is not audible to the human ear, PSTN channel utilizes normal narrowband speech, but at the receiver side the embedded sub-bands are extracted. Thus, the speech takes the form of a wideband speech with higher intelligibility and better quality. Improved communication was also a target for steganographic systems where hidden data are sent over acoustic channels as described in Figure 12. In [58, 59], data are pushed into live music or ambient sounds and transmitted over an acoustic channel. The transmitter in this case is a speaker, and the receiver is a microphone which are already present in numerous devices and environments. The developed technique was applied in a simple navigation system, where acoustic data are embedded into background music to indicate the location of the receiver.

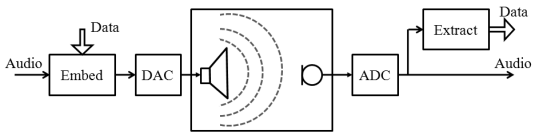


Figure 12: Embedded data transmission over acoustic channels.

6.3 Data storage

Given the possibility to hide more than 16 Kbps in a wide-band audio file with a conventional LSB encoding method, digital information can be reliably stored in audio steganographic systems. Another application for data storage could be seen in subtitled movies. Actors speech, film music, background sounds could be used to embed the text needed for translation. In this case, bandwidth is substantially reduced.

7 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 strengthes 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. A comparison as well as a performance evaluation (i.e., imperceptibility and steganalysis) for the reviewed techniques have been also presented. This study showed that 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. From our point of view, the diversity and large number of existing audio steganography techniques expand application possibilities. 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.

References

1. Ross J. Anderson, editor. Information hiding: 1st international workshop, volume 1174 of Lecture Notes in Computer Science, Isaac Newton Institute, Springer-Verlag, Berlin, Germany, May 1996.
2. 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.
3. E. Zwicker and H. Fastl, Psychoacoustics, Springer Verlag, Berlin, 1990.
4. F.Djebbar, B. Ayad, K. Abed-Meraim and H. Hamam, "A view on latest audio steganography", 7th IEEE Internationl Conference on Innovations in Information Technology, Abu Dhabi, UAE, 2011.
5. Mehdi Fallahpour and David Megias, "High capacity audio watermarking using FFT amplitude interpolation", IEICE Electron. Express, Vol. 6, No. 14, pp.1057-1063, 2009.
6. F. Djebbar, D. Guerchi, K. Abed-Maraim and H. Hamam, "Text-in speech spectrum steganography", ISSPA Mai 2010, Malaysia, 2010.
7. F. Djebbar, B. Ayad, K. Abed-Meraim and 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.
8. F. Djebbar, K. Abed-Maraim, D. Guerchi, and H. Hamam, "Energy based text-in speech spectrum hiding using speech mask properties", ICSRA Mai 2010, China, 2010.
9. F. Djebbar, H. Hamam, K. Abed-Meraim and D. Guerchi, "Controlled distortion for high capacity data-in-speech spectrum steganography", International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IEEE-IIHMSP), ISBN: 978-0-7695-4222-5, 212-215, 2010.
10. Y. Hu, P. Loizou, "Evaluation of objective quality measures for speech enhancement", IEEE Transactions on Speech and Audio Processing, 16(1), 229-238, 2008.
11. K. Gopalan, "Audio steganography using bit modification", Proceedings of International Conference on Multimedia and Expo, Vol. 1, pp.629-632, 6-9 July 2003.
12. N. Cvejic, T. Seppiinen, "Increasing the capacity of LSB-based audio steganography", IEEE Workshop on Multimedia Signal processing, pp. 336 -338, 2002.
13. 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, pp. 533, 2004.
14. N. Cvejic, and T. Seppanen, "Reduced distortion bit-modification for LSB audio steganography", Journal of Universal Computer Science, vol. 11, no.1, pp. 56-65, January 2005.
15. Mohamed A. Ahmed, Laiha Mat Kiah, B.B. Zaidan and A.A. Zaidan, "A Novel Embedding Method to Increase Capacity and Robustness of Low-bit Encoding Audio Steganography Technique Using Noise Gate Software Logic Algorithm", Journal of Applied Sciences, vol. 10, pp. 59-64, 2010.
16. D. Gruhl and W. Bender, "Echo hiding", Proceeding of Inforomation Hiding Workshop, pp. 295315, 1996.

17. Erfani, Y. and Siahpoush, S, "Robust audio watermarking using improved TS echo hiding", Digital Signal Processing, vol. 19, pp.809-814, September 2009.

18. S. Shirali-Shahreza and 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

19. S. Shirali-Shahreza and M. Shirali-Shahreza, "Real-time and MPEG-1 layer III compression resistant steganography in speech", The Institution of Engineering and Technology Information Security, IET Inf. Secur., vol. 4, no. 1, pp. 17, 2010.

20. G.S.Kang, T.M.Moran, D.A.Heide, "Hiding Information Under Speech", Naval Research Laboratory, Washington, DC 20375-5320, NRL/FR/5550–05-10,126, 2005.

21. B. Paillard, P. Mabilleau, S. Morissette, J. Soumagne, "PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals", journal of Audio Engeneering Society, vol. 40, pp 21-31, February 1992.

22. Khan, K. "Cryptology and the origins of spread spectrum", IEEE Spectrum 21, pp. 70-80, 1984.

23. S. Hernandez-Garay, R. Vazquez-Medina, L. Nino de Rivera, V. Ponomaryov, "Steganographic communication channel using audio signals", 12th International Conference on Mathematical Methods in Electromagnetic Theory, (MMET), pp. 427 - 429, 2 July 2008.

24. H. Matsuka, "Spread spectrum audio steganography using sub-band phase shifting", In IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP06), pp. 36, Pasadena, CA, USA, December 2006.

25. X. Li, H.H. Yu, "Transparent and robust audio data hiding in subband domain", Proceedings of the Fourth IEEE International Conference on Multimedia and Expo, (ICME 2000), New York, NY, pp. 397400, 2000.

26. 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.

27. Mohammad Pooyan, Ahmed Delforouzi, "Adaptive Digital Audio Steganography Based on Integer Wavelet Transform", Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2007), vol. 2 pp. 283 - 286, 2007.

28. S. Shirali-Shahreza and M. Shirali-Shahreza, "High capacity error free wavelet domain speech steganography", Proc. 33rd Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp. 17291732, 30 March 2008.

29. K. Gopalan, et al, "Covert Speech Communication Via Cover Speech By Tone Insertion", Proceeding of IEEE Aerospace Conference, Big Sky, MT, March 2003.

30. K. Gopalan and S. Wenndt, "Audio Steganography for Covert Data Transmission by Imperceptible Tone Insertion", WOC 2004, Banff, Canada July 8 10, 2004.

31. Gang. L, A.N. Akansu, M. Ramkumar, "MP3 resistant oblivious steganography", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, Vol. 3, pp.1365-1368, 7-11 May 2001.

32. X. Dong, M. Bocko, Z. Ignjatovic, "Data hiding via phase manipulation of audio signals", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 377-380, 17-21 May 2004.

33. 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.

34. X. Li and H.H. Yu, "Transparent and robust audio data hiding in cepstrum domain", Proc. IEEE International Conference on Multimedia and Expo, (ICME 2000), New York, NY, 2000.

35. K. Gopalan, "Audio Steganography by Cepstrum Modification", Proc. of the IEEE 2005 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), Philadelphia, March 2005.

36. K. Gopalan, "A unified audio and image steganography by spectrum modification", IEEE International Conference on Industrial Technology (ICIT), pp.1-5, 10-13 Feb. 2009.

37. R. Ansari, H. Malik, and A. Khokhar, "Data-hiding in audio using frequency-selective phase alteration", IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '04), pp. 389-392, Montreal, Quebec, Canada, May 2004.

38. H. M. A. Malik, R. Ansari, and A. A. Khokhar, "Robust Data Hiding in Audio Using Allpass Filters", IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 4, pp. 1296 - 1304, May 2007.

39. A. Nishimura, "Data hiding for audio signals that are robust with respect to air transmission and a speech codec", IIH-MSP'08, pp. 601-604, 15-17 Aug 2008.

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

41. B. Geiser, P. Vary, "High rate data hiding in ACELP speech codecs", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008), pp. 4005 - 4008, 4 April 2008.

42. Naofumi Aoki, "A Technique of Lossless Steganography for G.711 Telephony Speech", International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2008), pp. 608-611, 2008.

43. Naofumi Aoki, "A Semi-Lossless Steganography Technique for G.711 Telephony Speech", International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2010), pp. 534-537, 2010.

44. Y. F. Huang, S. Tang, J. Yuan, "Steganography in Inactive Frames of VoIP Streams Encoded by Source Codec", IEEE Transactions on Information Forensics and Security 6(2): 296-306, 2011.

45. Invisible secrets, http://www.invisiblesecrets.com/
46. Steganos Security Suite 7, http://www.steganos.com
47. Stools Version 4.0, http://info.umuc.edu/its/online_lab/ifsm459/s-tools4/
48. Hide4PGP, http://www.heinz-repp.onlinehome.de/Hide4PGP.htm
49. Steghide, http://steghide.sourceforge.net/
50. Mp3Stego, http://www.petitcolas.net/fabien/steganography/mp3stego/
51. Qingzhong Liu, Andrew H. Sung, Mengyu Qiao, "Temporal derivative-based spectrum and mel-cepstrum audio steganalysis", IEEE Transactions on Information Forensics and Security, vol. 4, no. 3, pp. 359-368, 2009.
52. Cristianini N, Shawe-Taylor J, "An introduction to Support Vector Machines Cambridge University Press", 2000.
53. J. Nafeesa Begum ; K. Kumar ; Dr. V. Sumathy, "Design And Implementation Of Multilevel Access Control In Medical Image Transmission Using Symmetric Polynomial Based Audio Steganography", International Journal of Computer Science and Information Security, Vol 7, pp. 139-146, 2010
54. M. Shirali-Shahreza, "Steganography in MMS", in Multitopic Conference, INMIC 2007. IEEE International, pp. 1-4, 2007
55. Paik, Michael, "Blacknoise: Low-fi Lightweight Steganography in Service of Free Speech", NYU. pp. 1-11, January 2010.
56. P. Vary and B. Geiser, "Steganographic wideband telephony using narrowband speech codecs", in Conference Record of Asilomar Conference on Signals, Systems, and Computers, Grove, CA, USA, Nov 2007.
57. S. Chen, H. Leung, H. Ding, "Telephony Speech Enhancement by Data Hiding", Instrumentation and Measurement, IEEE Transactions on Volume 56, Issue 1, pp. 63-74, Feb. 2007.
58. N. Lazic and P. Aarabi, "Communication over an Acoustic Channel Using Data Hiding Techniques", IEEE Transactions on Multimedia, Vol. 8, No. 5, October 2006
59. Po-Wei Chen, Chun-Hsiang Huang, Yun-Chung Shen, Ja-Ling Wu, "Pushing information over acoustic channels", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1421-1424, 2009.