

Chapter 2

Audio Watermarking

2.1 Introduction

Audio watermarking is a well-known technique of hiding data through audio signals. It is also known as audio steganography and has received a wide consideration in the last few years. So far, several techniques for audio watermarking have been discussed in literature by considering different applications and development positions. Perceptual properties of human auditory system (HAS) help to hide multiple sequences of audio through a transferred signal. However, all watermarking techniques face to a problem: a high robustness does not come with a high watermark data rate when the perceptual transparency parameter is considered as fixed. Furthermore, selection of a suitable domain, cover, and considering the problems associated with data-hidden techniques must be considered for designing the path to achieve a data-hidden purpose.

The remainder of this chapter is organized as follows: Transmission channel for audio watermarking is discussed. Different audio watermarking attacks are explained. Various audio watermarking techniques are compared.

2.2 Transmission Channel

A signal travels from different transmission environment during its journey from transmitter to receiver. As schematically illustrated in Fig. 2.1 [1], there are four classes of transmission environments. They are digital, resampling, analog, and on the air environments.

A signal passes through a digital end-to-end environment that is the way from which a digital file is copied from a machine to another one, with no further modifications and the same sampling at the encoder and decoder. For these reasons,

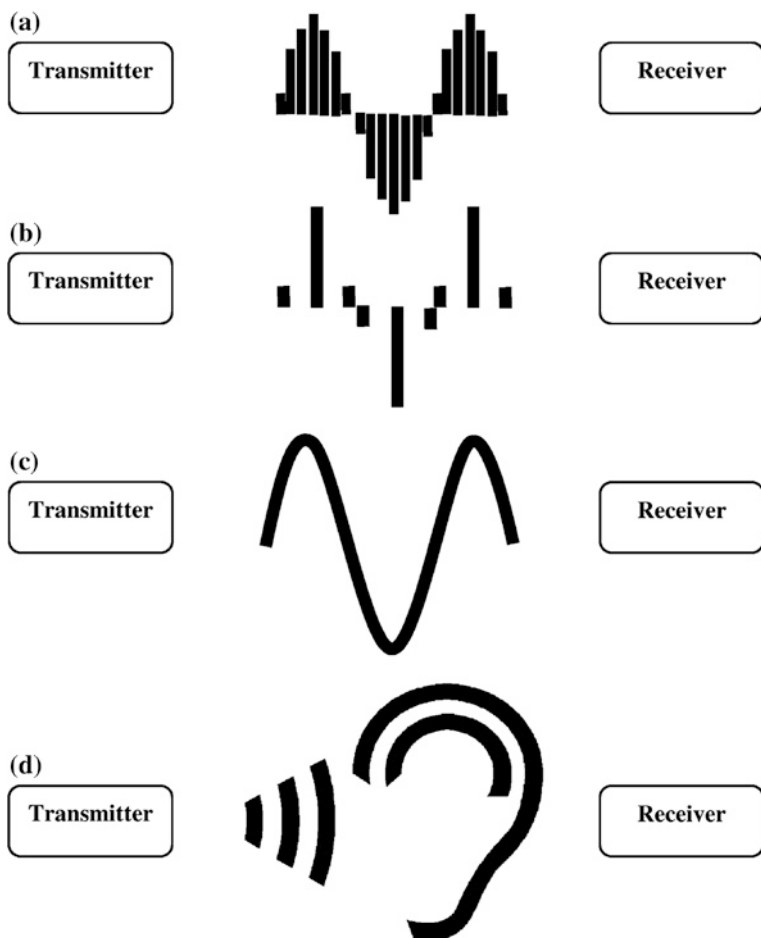


Fig. 2.1 Various transmission channels including: **a** digital, **b** resample **c** analog, and **d** over the air

the least data hidden can be applied in this class. Resampling is the second class of environment for a signal. The sampling rate for a signal during resampling is not necessary the same as its first sampling rate and temporal characteristics of the signal are subject to some modifications. Nevertheless, the signal remains in digital form throughout its way and almost the magnitude and phase of the signal remains intact. When a signal is played in analog environment, its phase is generally preserved. But some of its features do not hold their initial values, e.g., absolute signal magnitude, sample quantization, and temporal sampling rate. There is a final class on environment that is met when a signal is played on the air and is resampled with a microphone. The fact is that the signal can be modified in a nonlinear manner in terms of phase, amplitude, and frequency components

(e.g., echoes). Due to different impacts of transmission environments on the characteristics of a signal and data-hiding method, it is necessary to consider all the possible environments that a signal may pass.

2.3 Audio Watermarking Techniques

Generally, many audio watermarking techniques have been developed. The well-known methods of audio watermarking based on the limitations of perceptual properties of HAS are including simple least significant bits (LSB) scheme or low-bit encoding, phase coding, spread spectrum, patchwork coding, echo coding, and noise gate technique.

A pathway for watermarking especially for the famous patchwork algorithm was proposed in [2]. His method improves the performance of the original patchwork algorithm. Another method called as modified patchwork algorithm (MPA) [3] enhanced the power of Arnold's algorithm and improved its performance in terms of robustness and inaudibility. A mathematical formulation has also been presented that aids to advance the robustness.

Spread-spectrum technology has been utilized in audio watermarking in [1] which was originally introduced in [4]. Another method based on the spread-spectrum technology in [5] is a multiple echo technique that replaces a large echo into the host audio signal with multiple echoes with different offsets. Next method is the positive and negative echo-hiding scheme [6]. Each echo contains positive and negative echoes at adjacent locations. In the low-frequency band, the response of positive and negative echoes forms a smooth shape that is resulted by similar inversed shape of a negative echo with that of a positive echo. When positive and negative echoes are employed, the quality of the host audio is not obviously depreciated by embedding multiple echoes.

Backward and forward kernels are employed in an echo-hiding scheme presented by Kim and Choi [7]. They theatrically provided some results showing that the robustness of echo-hiding scheme improves by using backward and forward kernels. They showed that when the embedded echoes are symmetric, for an echo position associated with a cepstrum coefficient, the amplitude in backward and forward kernels is higher than when using the backward kernel.

Time-spread echo kernel is then proposed by Ko et al. [8]. A pseudo-noise sequence acts as a secret key that spreads out an echo as numerous little echoes in a time region. This secret key is then applied for extraction of the embedded data of the watermarked signal. The usage of the pseudo-noise sequence is essential, because the extraction process of a watermarked audio signal becomes very tough with no secret key.

In this part, the available audio watermarking techniques are divided into the three major categories. Three categories for audio watermarking are summarized in Fig. 2.2 which are based on prominent domains for embedding data in an audio signal: temporal, frequency and coded domains. In the reminder of this chapter, each method is summarized and their advantages and disadvantages are discussed.

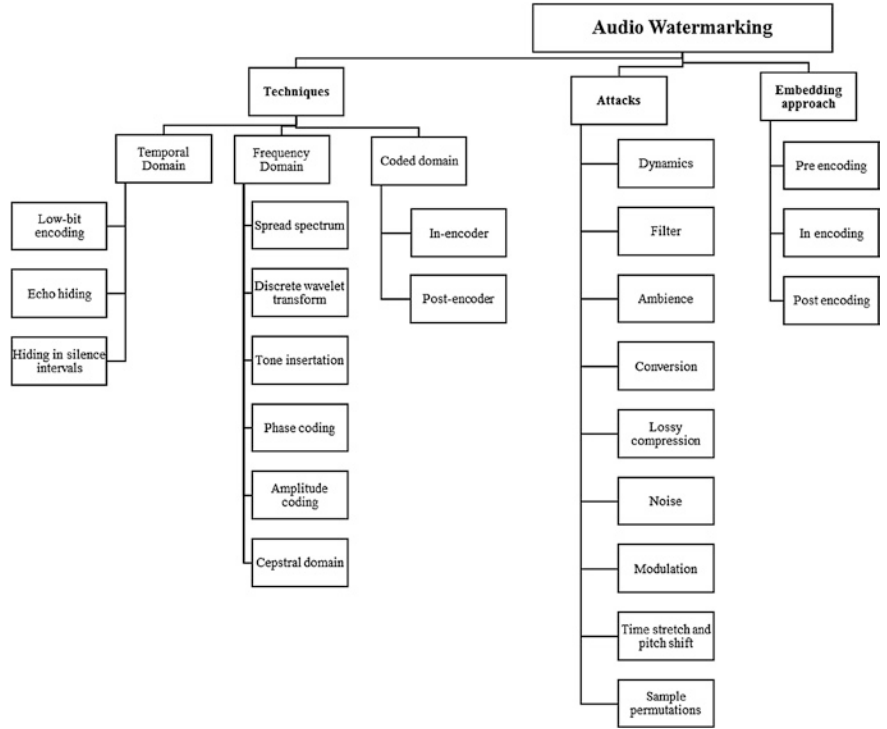


Fig. 2.2 Audio watermarking techniques

2.3.1 Temporal Domain

Audio watermarking techniques based on temporal domain are summarized in this section. Famous techniques for temporal domains are including low-bit encoding, echo hiding, and hiding in silence interval. In the following, each technique is fully discussed in detail.

2.3.1.1 Low-Bit Encoding

The most applied method for data hiding is called as low-bit encoding or least significant bit (LSB) [9]. Basically, the least significant bit of the cover audio is utilized for embedding each bit from the message. For example, 8 kbps data are hidden in a signal with 8 kHz sampled audio which has 8 bits per sample. This method is relatively simple and has a high capacity for hiding data. The robustness of this method is increased when it is combined with other watermarking methods. Nevertheless, the low-bit encoding method is sensitive to noises, which reduces the security and robustness. The position of hidden data in the watermarked

signal is known which makes this method vulnerable to attacks and an attacker via elimination of entire LSB plane can easily discover a message or destroy the watermark.

Basic LSB has been performed for transmission of an audio signal on a wireless network in [10]. The results verified that the method reduces the robustness and security at high rate of embedding data, but it does not harm the imperceptibility of final signal.

A method for embedding four bits per sample was presented in [11] that enhanced the hiding capacity. This method reduces the impact of error on the watermarked audio signal by defusing the embedding error on the next four samples. The depth of embedding layer of data increased from 4 layers to 6 and 8 LSB layers with no significant effect on the imperceptibility of the audio signal [12]. The results showed that the methods with higher embedding layer enhanced the robustness of previous method when noise addition and distortion occurs.

In [13], bits of the message are replaced with the bits at the sixth position of each 16-bit sample of the original audio signal. An approach for reducing the embedding error is replacing the message bits in such a way that the resulted bit sequence becomes closer to the original one. For this purpose, other bits are permitted to be flipped for increasing the closeness of bit sequence to the original one. As an instance, if four bits "0100" (value 4) are used for embedding data and bit "1" must be embedded in the bit sequence, it is suggested to select the bit sequence of "0011" (that is value 3) instead of having "1100" (that is value 12). The reason is that value 3 is closer to value 4 and the result is a lower embedding error rate.

The other approach in [12] suggests an eight layer for LSB embedding. In order to enhance imperceptibility of watermarked signal, the approach avoids hiding data in silent periods of the original signal. Due to assigning 8 bits for LSB embedding, the hiding capacity of the result becomes lower than the previous methods. However, it improves the robustness. The major disadvantage of embedding data in 6th or 8th position of LSB is the difficulty to reveal the original audio signal especially when the bits are shifted or flipped to enhance the embedding error rate.

2.3.1.2 Echo Hiding

An audio effect is known as echo which repeats some parts of the sound by creating delay inside the audio signal. In order to hide an echo, echo-hiding method generates a short echo by using a resonance and adds the echo to the original audio signal. The addition of the short echo is not recognizable by HAS; therefore, this method is not sensitive to noise addition. Other perceptual and statistical properties of original signal are kept in resulted signal.

Three parameters of the echo signal are the candidates for hiding the data. They include the initial amplitude, the delay (or offset), and the decay rate. The data can be successfully hidden in the audio signal if their values are managed to keep

the imperceptibility of audio signal [14]. For this reason, the values of amplitude and decay rates should be set below the audible threshold of HAS. As an example, when the time difference between the original signal and the echo stays below 1 ms, there is no annoying effect on the audibility of the signal.

Due to the induced size of echo signal, low embedding rate, and security, there are few systems and applications that practically developed this method. To the best of our knowledge, there is no real system that uses echo hiding in audio watermarking which cannot provide sufficient data for evaluation. An echo-hiding-time spread technique has been introduced to resolve the low robustness of echo-hiding technique in facing with common linear signals [15]. This method spreads the watermark bits all over the original signal and the destination recovers them by using the correlation amount. As a result of being a cepstral content-based method, the cepstral portion of error is detached and the detection rate at the decoder gets higher.

2.3.1.3 Hiding in Silence Intervals

Another candidate for embedding data is silence intervals in speech signal. A simple approach for hiding in silence intervals is proposed [16]. Consider n as the number of required bits for denoting a value from the message to hide. The silence intervals in audio signal should be detected and measured in terms of the number of samples in a silence interval. These values are decremented by x , $0 < x < 2n$ bits, where $x = \text{mod}(\text{new_interval_length}, 2n)$. As an instance, consider that the value 6 is hidden in a silence interval with length 109. Taken 7 samples out from the interval, 102 samples are remained in the new interval. The value x is computed as $x = \text{mod}(102, 8) = 6$. The short length of silence intervals that commonly seen in continuous parts of normal audios is omitted from the portions for hiding data. The perceptual transparency of this method is acceptable, but compression of signal misleads the data extraction process. As a solution for this problem, an approach is presented in [17] which separates the silence intervals from audio intervals so that they are not interpreted as one another. Thus, it reduces the samples in silence intervals and slightly augments the samples of the audio interval. The first and last interval added to the audio during MP3 coding is simply ignored in data hiding and retrieval.

As a general conclusion, conventional LSB approach is simpler than other methods; however, its capacity for hiding data is low. Moreover, it is resilient to noise additions and shows higher robustness in comparison with its variants [12, 13]. The main difficulty is a few number of applications that use time domain techniques.

2.3.2 Frequency Domain

Main idea behind using the frequency domain (or transform domain) for hidden data is the limitation of HAS when frequency of an audio signal fluctuates very rigid. The “masking effect” phenomenon enables the HAS to mask weaker

frequency near stronger resonant frequencies [18]. It provides a time duration that can be utilized for embedding data. The data hidden in this space is not perceptible by HAS. Watermark methods in frequency domain directly manipulate the masking effect of HAS by explicit modification of masked regions or indirectly by slight change of the samples of the audio signals.

2.3.2.1 Spread Spectrum

By spreading data in the frequency domain, spread spectrum (SS) technique ensures an appropriate recovery of the watermarked data when communicated over a noise-prone channel. SS utilizes redundancy of data for degrading the error rate of data hiding. An M-sequence of code handles the data and is embedded in the cover audio. This sequence is known to sender and receiver and if some parts of these values are modified by noise, recovery of data is feasible by using other copies [19]. The SS technique was developed in MP3 and WAV signals for the purpose of hiding confidential information in the form of conventional direct-sequence spread spectrum (DSSS) technique [20].

A frequency mask was suggested for embedding the data in a watermarked audio signal [21]. When a phase-shifting approach is combined to SS, the result is a watermarked signal with a higher level of noise resistance and robustness. As discussed in [21], the detection of hidden data is simple in the new method, but the rate of hiding data is low. As a solution, sub-band domain is chosen to provide better robustness and improving the decoder's synchronization uncertainty which require to select proper coefficients in sub-band domain [22].

2.3.2.2 Discrete Wavelet Transform

Discrete wavelet transform (DWT) is multi-scale and multi-resolution technique to decompose signal to different time-frequency components. A watermarking method is proposed by DWT which hides data in LSB of the wavelet coefficients [23]. The imperceptibility of hidden data is low in DWT. Whenever the integer wavelet coefficients are available, a hearing threshold is useful to improve the audio inaudibility as presented in [24]. If a DWT watermarking technique evades embedding data in silent parts, hidden data does not annoy the audience [25]. DWT provides a high rate of data hiding; nevertheless, the procedure for data extraction at the receiver is not always accurate.

2.3.2.3 Tone Insertion

HAS does not detect audio signals when lower power tones are located near very high tones. Tone insertion benefits this HAS feature for data hiding. The method to embed inaudible tones in cover signal was introduced in [26]. Given that one bit

is planned to be hided in an audio frame, two frequencies of f_0 and f_1 are selected and a pair of tones is created in this area. Each frequency has a masked frequency, e.g., pf_0 for f_0 and pf_1 for f_1 . Considering there are n frames and the power of each frame is denoted by pi where $i = 1, \dots, n$. The value of each masked frequency is set to a predefined value that is the ratio of the general power of each audio frame pi . A correct data extraction from watermarked data is obtained when tones are inserted at known frequencies and at low power level.

Procedure of detection of the hidden data from the inserted tones is performed by computing the power of each frame, pi , including the power of pf_0 for f_0 and pf_1 for f_1 . If the ratio $pi / pf_0 > pi / pf_1$, then the hidden bit is assumed as "0"; otherwise, it is considered as "1." Thus, the hidden data is extracted. As perceived, the data-hiding capacity of tone insertion method is low. Some attacks can be tolerated by tone insertion method, e.g., low-pass filtering and bit truncation; nonetheless, the attackers can simply detect the tones and extract the hidden data. Similar to LSB, this problem can be resolved by varying four or more pairs of frequencies in a keyed order.

2.3.2.4 Phase Coding

Another limitation of HAS is its inability to detect the relative phase of different spectral components. It is the basis of interchanging hidden data with some particular components of the original audio signal. This method is called as phase coding and works well on the condition that changes in phase components are retained small [27]. Phase coding tolerates noises better than all other above-mentioned methods [1, 28].

An independent multi-band phase modulation is utilized for phase coding [27]. In phase modulation method, phase alteration of the original audio signal is controlled to obtain imperceptibility of phase modifications. Phase components are determined by quantization index modulation (QIM). Then, the nearest "o" and "x" points are replaced with phase values of frequency bin to hide "0" and "1," respectively. Therefore, phase coding achieves a higher robustness when perceptual audio compression is applied [1].

QIM was widely been used that improves the capacity of data hiding of phase coding by replacing the strongest harmonic with step size of $\pi/2n$ [29]. Phase coding has zero value of bit error rate (BER) when MP3 encoder is applied that demonstrates the high robustness of this method.

As HAS is not sensitive to phase changes, an attacker simply can replace his/her data with the real hidden data. S/he can apply frequency modulation in an inaudible way and modify the phase quantization scheme.

2.3.2.5 Amplitude Coding

The sensitivity of HAS is high for frequency and amplitude components. Therefore, it is possible to embed hidden data in the magnitude audio spectrum.

The capacity of hiding data is high by using this method as presented in [28] and the tolerance of the method regarding noise distortion and its security in facing with different attacks is high. Hiding different types of data is feasible by using this method. Encrypted data, compressed data, and groups of data (LPC, MP3, AMR, CELP, parameters of speech recognition, etc.) can be hidden by using amplitude coding.

Initially, some spectrum areas for secure embedding data are found in the wide-band magnitude audio spectrum. For this purpose, an area below 13 dB of the original signal spectrum is taken into account and a frequency mask is defined in this area. In regard to the magnitude spectrum, a distortion level that is resilient to noise distortion is considered. Then, candidate locations and the capacity for hiding data can be determined.

For 7 to 8 kHz frequencies, the effect on the wideband speech is minimum [30]. Therefore, this area is a good space for hiding data with not compromising the inaudibility of watermarked signal. For this purpose, the entire range between 7 and 8 kHz can be filled with hidden data.

2.3.2.6 Cepstral Domain

Cepstrum coefficients provide spaces for watermarking. This method is resilient to well-known attacks in signal processing and is also known as log-spectral domain. It locates the hidden data in the portions of frequencies that are inaudible by HAS and obtains a high capacity of hiding data, between 20 and 40 bps [31]. Initially, the domain of original audio signal is modified to cepstral domain. Statistical mean function helps to choose some cepstrum coefficients that are later altered by hidden data. As the masked regions of the majority of cover audio frames are utilized for data hiding, the imperceptibility of watermarking is relatively high in cepstral domain.

The robustness of this method was improved by considering high energetic frames and replacing cepstrum of two selected frequencies F_u and f_2 by bit “1” or “0” [32]. The security and robustness of this method was later improved by considering different arbitrary frequency components at each frame [33]. Distinct types of all-pass digital filters (APF) choose sub-bands that are suitable for embedding hidden data. A hiding method based on APF improves the robustness of watermarked audio signal facing with addition of noise, random chopping, e-quantization, and resampling [34]. Given n as an even positive integer, the robustness can be further improved by applying a set of n -order APFs as is in [35]. Pole locations of an APF are calculated from the power spectrum by several approaches. Finally, the data is hidden in some chosen APF parameters.

According to calculations, all the above-mentioned techniques have higher resilience against noise additions in frequency domain (or transform domain) [28]. Almost all data-hiding methods in transform domain benefits the perceptual models of HAS, especially frequency masking effect, to improve the data-hiding capacity as long as signal distortion can be tolerated. Most of the watermarking

methods in transform domain is tolerating simple noise distortions including amplification, filtration, or resampling. However, the probability of them to tolerate noisy transmission environment or data compression in ACELP and G.729 is low.

2.3.3 Coded Domain

In real-time communications, coded domain is favorable. Despite the benefits of transform domain in comparison with time domain, it does not act well when real-time applications and voice encoders under particular encoding rates, e.g., AMR, ACELP, and SILK, are employed. An encoder codes the audio signal while it is transferring through communication channels and at the end, a decoder is responsible for decoding the coded data. As the encoder and decoder have their own rates, a decoded signal might slightly differ the original signal. Therefore, the procedure for data extraction and retrieval is complicated in coded domain. Furthermore, the correctness of the extracted data is a challenge itself.

2.3.3.1 In-Encoder Techniques

A coded technique called as in-encoder technique was introduced that can successfully tolerate noise distortion, audio codec, compression, and reverberations [36]. Different types of audio signals including music and speech were evaluated for embedding watermarked data when sub-band amplitude modulations have been used.

A pitch-tracking algorithm based on autocorrelation performed voiced/unvoiced segmentation in [37] based on the LPC vocoder. A data sequence was embedded in the unvoiced segments by alteration of the linear prediction residual. This method does not affect the audibility of the watermarked signal if the residual's power is matched. Capacity of a reliable data hiding is up to 2 kbps. Hidden data is replaced with the unmodified coefficients of the LPC filter, and for decoding the embedded data, a linear prediction analysis on the transmitted audio signal is performed.

A coded technique that hides the data in the audio codecs and in the LSB of the Fourier transform was proposed in [18]. This technique embeds data in the LSB of the Fourier transform of the prediction residual of the host audio signal. This technique does not guarantee inaudibility of watermarked data and its imperceptibility is considered as low. It automatically shapes the spectrum of LSB noise when an LPC filter is employed; thus, the watermarked data has a less impact on audibility of the audio signal.

2.3.3.2 Post-encoder Techniques

The watermark can be embedded in the coded domain by the post-encoder (or in-stream) techniques. A post-encoder technique was developed on an AMR encoder at a rate of 12.2 Kbit/s and in the bitstream of an ACELP codec [38]. It works together with the analysis-by-synthesis codebook search and the results showed that it hides 2 Kbit/s of data in the bitstream and obtains a noise ratio of 20.3 dB. A lossless post-encoder technique was developed that works on G.711-PCMU telephony encoder [39]. Data is presented in the form of folded binary code. The value of each sample varies between -127 and $+127$ (consists of values -0 and $+0$). For every 8-bit sample with absolute amplitude of zero, one bit is hidden. Thus, the capacity of hidden data varies between 24 and 400 bps. As a solution for improving capacity of hidden data for G.711-PCMU, a semi-lossless approach was proposed in [40]. A predefined level, denoted as “ i ,” amplifies the sample’s amplitudes. Hereafter, the samples with absolute amplitude between 0 and i are applied for embedding data. For increasing the capacity of watermarked data, in [41] the inactive frames in low-bit-rate audio stream (i.e., 6.3 kbps) were used for encoding a G.723.1 source codec.

In general, coded domain techniques are well suited for real-time applications. Watermarking techniques especially in-encoder approaches benefits from a high robustness and security. While capacity of hidden data is higher than the codec data in some techniques; due to high sensitivity of bitstream to modifications, it is held small to limit the perceptibility. Although ACELP, AMR, or LPC audio codecs and noise additions are tolerable by coded domain techniques, the integrity of hidden data cannot be promised where transcoding (i.e., a voice encoder/decoder) is available in the networks. A voice enhancement method that is applied for reducing the noise or echo can modify the hidden data, as well. However, the procedure of data extraction in tandem-free operation guarantees that hidden data remains intact during encoding the data encoding.

2.4 Embedding Approach

In covert communication, data is transferred through multiple encoders/decoders. An encoder reduces the size of transmitted data by removing the redundant or unused data. Thus, each coder influences the integrity of data, while the robustness of covert communications requires a high integrity of watermarked data. Although, there are some ways to ensure data integrity in encoder/decoder, it imposes negative impacts on hiding capacity of data. There are three levels for embedding a data-in-audio watermark system [38]. Figure 2.3 summarizes the aforementioned methods for audio steganography according to the occurrence rate. The evaluation of security requires a third-party effort cost to retrieve the hidden data. Each level has some benefits and weaknesses that are discussed as follows.

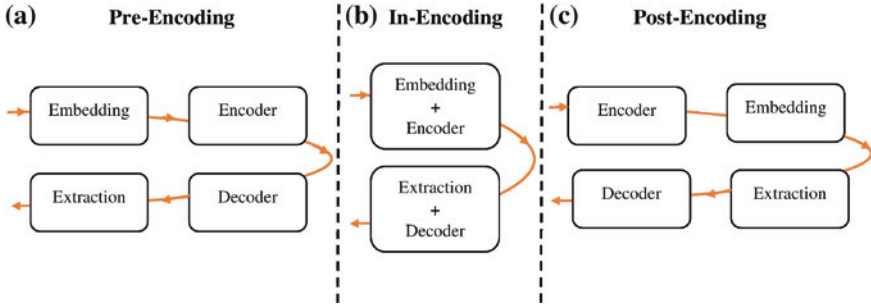


Fig. 2.3 Different approaches for embedding the watermark

2.4.1 Embedding Before Encoding (*Pre-encoding*)

Prior to encoding process, the data is embedded in time and frequency domain. This level is known as pre-encoder embedding. The integrity of data, during transmission over network, is not guaranteed in this level because high degree of data compression in encoders (e.g., in ACELP or G.729) and addition of noise (in any form, e.g., WGN) can compromise the integrity of data. On the other hand, there are some methods that allow a low degree of modifications on the audio signal including resizing, resampling, filtering. Therefore, they are resilient to low degree of noise addition or data compression. Only noise-free environments provide a space for high rate of data hidden.

2.4.2 Embedding During Encoding (*in-Encoder*)

This data embedding level provides a robust data hiding. For this purpose, a codebook of codecs is necessary. The codebook keeps the information of transmitted data once the requantization operation is performed. As a result, for every parameter of audio signal, two important values of embedded-data and codebook parameters are kept. When value of embedded data is manipulated for any reason, this method faces to a severe problem for data extraction. It can occur when the data passes through a voice encoder/decoder in a radio access network (BST, BSC, TRAU) and/or in the core network (MSC) in a GSM network. Similar modifications occur when a voice enhancement algorithm is developed in a radio access network and/or in the core network.

2.4.3 *Embedding After Encoding (Post-encoder)*

This level of embedding data acts on bitstreams rather than the original audio signal. Data is hidden in a bitstream once it passes the encoder and before entering the decoder. Thus, value of data and the integrity of watermarked audio signal are vulnerable to undesirable modifications. Bitstreams are naturally more sensitive to alteration than audio signals and data integrity should be kept small to avoid imperceptibility of audio signal. Nevertheless, post-encoder embedding ensures the correctness of data once it is extracted in tandem-free operations and the message is retrieved in a lossless way.

2.5 Audio Attacks

As shown in Fig. 2.2, there are many attacks that can degrade the watermark data and as a consequence decrease the robustness of the audio watermarking techniques. Some of the attacks have already discussed in the literature for still images and some have been particularly mentioned for audio watermarking. In this section, the impact of each attack according to the audibility of hidden data by HAS is measured and the most effective attacks on audio signals are highlighted. Some of the attacks mostly occur in real environments. Suppose an audio signal is prepared to be broadcast on a radio channel. Based on the audience confidence and quality parameters of the radio channel, the audio material is normalized and compressed to fit the necessary level of loudness for transmission. Then, the quality of signal is optimized by equalization; undesired parts are demised or dehisced; useful frequencies are kept and unnecessary ones are omitted by filters.

In some applications, the robustness of watermarked audio signal should be high, e.g., in commercial radio transmission or copyright protection of music. In both examples, the watermark technique should not allow the signal to be destroyed or manipulated by attackers and if an attack occurs, it should not allow the attacker to misuse or reuse the signal. A well-known attack in this situation is lossy compression in MP3 at high rate of compressions. In addition to individual attacks, some attacks act in the form of groups. The group of attacks is also taken into account for performance evaluation of watermarking techniques. Main group attacks are including dynamics, filter, ambience, conversion, loss comparison, noise, modulation, time stretch (pitch shift), and sample permutation.

2.5.1 *Dynamics*

This group of attacks influences the loudness profile of an audio file. Some attacks including increasing or decreasing are simple and considered as the basic attacks.

Some attacks perform nonlinear functions including compression, expansion, and limiting. Thus, they are complicated. In another category, frequency range or a part of that is modified by frequency-dependent algorithm.

2.5.1.1 Compressor

When it is desired to decrease the strength of a signal in terms of its range, a compressor can be utilized. It can increase the overall loudness of a signal by degrading the peaks below a particular value with no distortions. Given a fast and inaudible attack that changes all signals louder than -50 dB by a small amount. It has the following properties: Attack time 1 ms, release time 500 ms, output gain 0 dB, threshold -50 dB, and ratio 1:1.1.

2.5.1.2 Denoiser

In some cases, it is essential to find a way for noise removal from the signal. Denoiser acts as a gate. It passes the eligible parts of the signal and blocks the noises. A denoiser needs a value to be used for detection of a noise. A basic denoiser simply considers loudness of signal as a noise, prior that a proper value of the loudness should be set. Here, the setting is assumed as -80 and -60 dB. Indeed, for detection of complicated noises, other techniques, e.g., DE clickers, and advanced tools are required.

2.5.2 Filter

Filters modify a spectrum by passing desired values and omitting undesired parts of the signal. Various filters have been introduced in signal processing. The basic filters are the high-pass filter and the low-pass filter. As equalizers increase or decrease some particular parts of spectrum, they can be counted as filters.

High-pass filter eliminates all frequencies below a particular value, here 50 Hz.

Low-pass filter eliminates all frequencies above a particular value, here 15 kHz.

Equalizer subtracts the frequency by a particular value, here by 48 db. The used bandwidth was frequency/10.000. Three versions of this attack have been tested using a range from 31 Hz to 16 kHz: 10 frequencies with the distance of 1 octave, 20 frequencies with the distance of 1/2 octave, and 30 frequencies with the distance of 1/3 octave.

L/R-splitting is an equalizer effect that increases the supposed stereo image. It works on two channels. In one channel, the frequency shares are reduced and are increased in the other channel. 20 frequency channels divide the spectrum. For each and every second, the value of frequency on the left channel is subtracted by dB and is increased by this value on the right radio channel. Finally, the volume of both channels is normalized to cover the volume changes.

2.5.3 Ambience

Consider an audio signal broadcasting in a room. In order to simulate this condition, reverb and delay parameters assist this group. By assigning various values to each parameter, many different qualities of effects are achieved.

Delay: The original signal is duplicated, and by the addition of the copy to the original audio signal, a wide space is simulated. Here, the volume of the delayed signal is 10 % of the original one and the delay duration is 400 ms.

Reverb: For simulation of rooms or building, reverb is utilized. Although it is similar to delay, it is shorter in delay time and reflections.

2.5.4 Conversion

Depending on the application and tools, the formats of audio material are modified, e.g., to play a mono-audio material on an stereo device, data is duplicated. The sampling rate of devices has been changed from 32 to 48 kHz and now even 96 kHz or sample size changes from 16 to 24 bit and vice versa.

Resampling: Sometimes for adaptation of devices, an audio signal is resampled by a different sampling frequency from the initial one, e.g., in CD production an audio signal is downsampled from 48 to 44.1 kHz. Resampling is similar to low-pass filter when a reduction to the highest possible frequency performed, e.g., a change from 44.1 to 29.4 kHz.

Inversion: inversion changes the sign of the samples, but the changes are imperceptible. For a comprehensive evaluation of watermarking technique, this test is also taken into account.

2.5.5 Loss Compression

Some compression algorithms work based on psychoacoustic effects of audio signal. They reduce the size of the compressed data to 10 or less times of the original data size.

2.5.6 Noise

So far, several attacks have been discussed. The result of most of the attacks is a noise. As already discussed, different sources of noise are known. Hardware components are the most effective sources of noise in audio signals. There is another attack that adds noise to terminate the watermark.

Random noise: This noise is made by addition of random numbers to the samples of an audio signal. Random numbers are limited to a particular percentage of the original audio signal. It can be considered up to 0.91 % of the original sample value on the condition that it does not compromise the quality of signal.

2.5.7 Modulation

Modulation effect can be considered as attacks, but they usually do not happen in postproduction. Software for processing audio signals can include modulation attacks. They are as follows:

Chorus: Sounds from multiple resources in the form of a modulated echo is added to the original audio signal. The delay time and strength and number of voices are different. Here, 5 voices, 30 mms max. delay, 1.2 Hz delay rate, 10 % feedback, 60 ms voice spread, 5 db vibrato depth, 2 Hz vibrato rate, 100 % dry out (unchanged signal), and 5 % wet out (effect signal) are taken into account.

Flanger: when a delayed signal is added to the original signal, flanger is generated. The delay is short and the length changes constantly.

Enhancer: An audio signal becomes more brilliant or excited if the amount of high frequencies is increased. To simulate the effect of enhancer (or exciter), sound forge is applied and medium setting is used. Detailed information about the parameters is not provided by the program.

2.5.8 Time Stretch and Pitch Shift

Time stretch and pitch shifts help to fine-tuning or fitting audio into time windows by changing the length of the audio signal with no changes in the pitch or vice versa.

Pitch Shifter: A complicated algorithm for editing audio signals is pitch shifter. This algorithm changes the base frequency of the signal with no modifications in the speed. So far, multiple pitch shifter algorithms have been presented in the literature. Selection of proper algorithm depends on the expected quality of the signal. The sound forge increases the pitch by 5 cent, and this is 480th of an octave.

Time Stretch: Time stretch prolongs or shortens the duration of an audio signal with no modification on the pitch. Here, a sound that forges with a length of 98 % of the original duration is considered.

2.5.9 Sample Permutations

An uncommon way to attack watermarks hidden in audio files is sample permutation. This group consists of algorithms that permute or drop samples and are not applicable in normal environments.

Table 2.1 Comparison among various audio watermarking techniques

Watermarking domain	Technique	Description	Benefits	Drawback	Capacity
Temporal domain	Low-bit encoding	<ul style="list-style-type: none">– The most applied method for data hiding– The simplest method for data hiding into data structures, e.g., data of audio in image file or data of image in audio file– Replaces LSB plane of each sampling point with hidden data– Higher embedding layer enhanced the robustness when noise addition and distortion occurs.	Simple to develop and high bit rate	Low security, sensitive to attacks, easy to intrude	16 kbps
		<ul style="list-style-type: none">– Embedding data in a short echo– Echo is generated by a resonance– Data hiding can be applied by three parameters of an echo signal: initial amplitude, the delay (or offset), and the decay rate.– The values of amplitude and decay rates should be set below the audible threshold of HAS– Two echoes with different offsets are utilized for embedded data: the binary datum “one” and the other to represent the binary datum “zero.”	Lossy data compression is tolerated	Low security, low hidden data capacity	50 bps
Transform domain	Silence intervals	<ul style="list-style-type: none">– Embeds the hidden data in silence intervals of a speech audio signal– Instead of time domain, frequency domain is utilized– It is more resilient to noises in comparison with time domain	Lossy data compression is tolerated	Low hidden data capacity	64 bps
	Magnitude spectrum		More resilient to noise addition during communications, higher rate of data hiding	Low robustness to simple audio manipulations	20 Kbps

(continued)

Table 2.1 (continued)

Watermarking domain	Technique	Description	Benefits	Drawback	Capacity
	Tone insertion	<ul style="list-style-type: none">– Embeds inaudible tones in cover signal– A correct data extraction from watermarked data is obtained when tones are insert at known frequencies and at low power level– The data-hiding capacity of tone insertion method is low.– Some attacks can be tolerated by tone insertion method, e.g., low-pass filtering and bit truncation; nonetheless, the attackers can simply detect the tones and extract the hidden data.– Security can be upgraded by varying four or more pairs of frequencies in a keyed order.	Inaudibility of hidden data	<ul style="list-style-type: none">– Low transparency– Low security	250 bps
	Phase spectrum	<ul style="list-style-type: none">– Hides data in a reference phase– Replaces the phase of original audio signal with a reference phase– Phase of subsequent segments is adjusted in order to preserve the relative phase between segments– Works well if changes in phase components are retained small.– Tolerates noises well	Robust against signal processing manipulation and data retrieval needs the original signal	Low rate of data hiding	333 bps
	Spread spectrum	<ul style="list-style-type: none">– Spreads hidden data in frequency domain– By spreading the encoded data, encodes stream of information on as much of the frequency as possible– Utilizes redundancy of data for degrading the error rate of data hiding– If interference on some– Frequencies is existed, the signal reception is permitted	high robustness	Vulnerable to time Scale modification	20 bps

(continued)

Table 2.1 (continued)

Watermarking domain	Technique	Description	Benefits	Drawback	Capacity
Coded domain	Cepstral domain	<ul style="list-style-type: none">– Data is replaced with cepstral coefficients– Locates the hidden data in the portions of frequencies that are inaudible by HAS– Obtains a high capacity of hiding data– APF improves the robustness of watermarked audio signal facing with addition of noise, random chopping, e-quantization, and resampling.	Robust against signal processing operations	Perceptible signal distortions and low robustness	54 bps
	Wavelet	<ul style="list-style-type: none">– Data is replaced with the coefficients of wavelet– Hides data in LSB of the wavelet coefficients– The imperceptibility of hidden data is low in DWT– Whenever the integer wavelet coefficients are available, a hearing threshold is useful to improve the audio inaudibility	High rate of data hiding	Inaccurate data extraction at the receiver	70 kbps
	Codebook modification	<ul style="list-style-type: none">– Requires a codebook– Codebook parameters are modified to hide data	High robustness	Low capacity of hidden data	2 kbps
	Bitstream hiding	<ul style="list-style-type: none">– Generates a bitstream by encoding– LSB is applied on the bitstream– Data is hidden in a bitstream– Bitstreams are naturally more sensitive to alteration than audio signals	High robustness	Low capacity of hidden data	1.6 kps

Zero-Cross Inserts: This attack finds value 0 in the samples and replaces them with 20 zeros. The result is a small pause in the signal. The pause length is minimum 1 s.

Copy Samples: this attack randomly selects some samples and duplicates throughout the signal. Therefore, the signal becomes longer than the original length. Here, the signal was repeated 20 times in 0.5 s.

2.6 Comparison Among Different Audio Watermarking Methods

In order to compare and classify the audio watermarking methods, some criteria must be chosen and defined. Based on the literature, major criteria for analysis and comparison of watermarking methods are considered as robustness, security, and hiding capacity (payload). Other parameters including the transmission environment and the application influence the evaluation criteria. For that, they should be considered for performance evaluation of every watermarking technique.

In an application where multiple levels of coding and decoding are planned, evaluation of a criterion like robustness is not possible without considering the environment constraints. Table 2.1 demonstrates general watermarking domains by taken into account the major techniques in each domain (the main idea is got from [28]). The details of each technique along with benefits, drawbacks, and obtained capacity of watermarking are brought in the table as well.

References

1. Bender, W., et al. 1996. Techniques for data hiding. *IBM Systems Journal*. 35(3.4): 313–336.
2. Arnold, M. 2000. Audio watermarking: features, applications, and algorithms. In *IEEE International Conference on Multimedia and Expo (II)*. Citeseer.
3. Yeo, I.-K., and H.J. Kim. 2003. Modified patchwork algorithm: A novel audio watermarking scheme. *IEEE Transactions on Speech and Audio Processing* 11(4): 381–386.
4. Cox, I.J., et al. 1997. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing* 6(12): 1673–1687.
5. Xu, C., et al. 1999. Applications of digital watermarking technology in audio signals. *Journal of the Audio Engineering Society* 47(10): 805–812.
6. Oh, H.O., et al. 2001. New echo embedding technique for robust and imperceptible audio watermarking. In *2001 IEEE international conference on acoustics, speech, and signal processing, 2001. Proceedings. (ICASSP'01)*. IEEE.
7. Kim, H.J., and Y.H. Choi. 2003. A novel echo-hiding scheme with backward and forward kernels. *IEEE Transactions on Circuits and Systems for Video Technology* 13(8): 885–889.
8. Ko, B.-S., R. Nishimura, and Y. Suzuki. 2005. Time-spread echo method for digital audio watermarking. *IEEE Transactions on Multimedia* 7(2): 212–221.
9. Chowdhury, R., et al. 2016. A view on LSB based audio steganography.
10. Gopalan, K. 2003. Audio steganography using bit modification. In *ICME'03. Proceedings. 2003 International Conference on Multimedia and expo, 2003*. IEEE.

11. Cvejic, N., and T. Seppanen. 2002. Increasing the capacity of LSB-based audio steganography. In *2002 IEEE workshop on multimedia signal processing*. IEEE.
12. Ahmed, M.A., et al. 2010. 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* 10(1): 59–64.
13. Cvejic, N., and T. Seppanen. 2004. Reduced distortion bit-modification for LSB audio steganography. In *2004 7th international conference on signal processing, 2004. Proceedings. ICSP'04*. IEEE.
14. Gruhl, D., A. Lu, and W. Bender. 1996. Echo hiding. In *Information Hiding*. Springer.
15. Erfani, Y., and S. Siahpoush. 2009. Robust audio watermarking using improved TS echo hiding. *Digital Signal Processing* 19(5): 809–814.
16. Shirali-Shahreza, S., and M. Shirali-Shahreza. 2008. Steganography in silence intervals of speech. In *International conference on intelligent information hiding and multimedia signal processing*. IEEE.
17. Shirali-Shahreza, M.H., and S. Shirali-Shahreza. 2010. Real-time and MPEG-I layer III compression resistant steganography in speech. *Information Security, IET* 4(1): 1–7.
18. Kang, G.S., T.M. Moran, and D.A. Heide. 2005. *Hiding information under speech*. DTIC Document.
19. Li, R., S. Xu, and H. Yang. 2016. Spread spectrum audio watermarking based on perceptual characteristic aware extraction. *IET Signal Processing*.
20. Kirovski, D., and H.S. Malvar. 2003. Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing* 51(4): 1020–1033.
21. Matsuoka, H. 2006. Spread spectrum audio steganography using sub-band phase shifting. In *International conference on intelligent information hiding and multimedia signal processing, 2006. IHH-MSP'06*. IEEE.
22. Li, X., and H.H. Yu. 2000. Transparent and robust audio data hiding in subband domain. In *International conference on information technology: coding and computing, 2000. Proceedings*. IEEE.
23. Cvejic, N., and T. Seppänen. 2002. A wavelet domain LSB insertion algorithm for high capacity audio steganography. In *Proceedings of 2002 IEEE 10th digital signal processing workshop, 2002 and the 2nd signal processing education workshop*. IEEE.
24. Delforouzi, A., and M. Pooyan. 2008. Adaptive digital audio steganography based on integer wavelet transform. *Circuits, Systems and Signal Processing* 27(2): 247–259.
25. Shirali-Shahreza, S., and M. Manzuri-Shalmani. 2008. High capacity error free wavelet domain speech steganography. In *IEEE international conference on acoustics, speech and signal processing, 2008. ICASSP 2008*. IEEE.
26. Gopalan, K., and S. Wenndt. 2004. Audio steganography for covert data transmission by imperceptible tone insertion. In *Proceedings of the IASTED international conference on communication systems and applications (CSA 2004)*, Banff, Canada.
27. Ngo, N.M., and M. Unoki. 2016. Method of audio watermarking based on adaptive phase modulation. *IEICE transactions on information and systems* 99(1): 92–101.
28. Djebbar, F., et al. 2012. Comparative study of digital audio steganography techniques. *EURASIP Journal on Audio, Speech, and Music Processing* 2012(1): 1–16.
29. Dong, X., M.F. Bocko, and Z. Ignjatovic. 2004. Data hiding via phase manipulation of audio signals. In *IEEE international conference on acoustics, speech, and signal processing, 2004. Proceedings.(ICASSP'04)*. IEEE.
30. Guerchi, D., et al. 2008. Speech secrecy: an FFT-based approach. *International Journal of Mathematics and Computer Science* 3(2): 1–19.
31. Li, X., and H.H. Yu. 2000. Transparent and robust audio data hiding in cepstrum domain. In *2000 IEEE international conference on multimedia and expo, 2000. ICME 2000*. IEEE.
32. Gopalan, K. 2005. Audio steganography by cepstrum modification. In *IEEE international conference on acoustics, speech, and signal processing, 2005. Proceedings.(ICASSP'05)*. 2005. IEEE.

33. Gopalan, K. 2009. A unified audio and image steganography by spectrum modification. In *IEEE international conference on industrial technology, 2009. ICIT 2009*. IEEE.
34. Ansari, R., H. Malik, and A. Khokhar. 2004. Data-hiding in audio using frequency-selective phase alteration. In *IEEE international conference on acoustics, speech, and signal processing, 2004. Proceedings. (ICASSP'04)*. IEEE.
35. Malik, H., R. Ansari, and A.A. Khokhar. 2007. Robust data hiding in audio using allpass filters. *IEEE Transactions on Audio, Speech, and Language Processing* 15(4): 1296–1304.
36. Nishimura, A. 2008. Data hiding for audio signals that are robust with respect to air transmission and a speech codec. In *IIHMSP'08 international conference on intelligent information hiding and multimedia signal processing, 2008*. IEEE.
37. Hofbauer, K., and G. Kubin. 2006. High-rate data embedding in unvoiced speech. In *INTERSPEECH*.
38. Geiser, B., and P. Vary. 2008. High rate data hiding in ACELP speech codecs. In *IEEE international conference on acoustics, speech and signal processing, 2008. ICASSP 2008*. IEEE.
39. Aoki, N. 2008. A technique of lossless steganography for G. 711 telephony speech. In *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE.
40. Aoki, N. 2010. A semi-lossless steganography technique for G. 711 telephony speech. In *2010 sixth international conference on intelligent information hiding and multimedia signal processing (IIH-MSP)*. IEEE.
41. Huang, Y.F., S. Tang, and J. Yuan. 2011. Steganography in inactive frames of VoIP streams encoded by source codec. *IEEE Transactions on Information Forensics and Security* 6(2): 296–306.

Digital Watermarking

Techniques and Trends

Nematollahi, M.A.; Vorakulpipat, C.; Rosales, H.G.

2017, XXV, 203 p. 42 illus., 13 illus. in color., Hardcover

ISBN: 978-981-10-2094-0