Computer Science Review 38 (2020) 100316

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

Digital audio steganography: Systematic review, classification, and analysis of the current state of the art

Ahmed A. AlSabhany [a,*], Ahmed Hussain Ali [b], Farida Ridzuan [a,c], A.H. Azni [a,c], Mohd Rosmadi Mokhtar [d]

[a] Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), Nilai 71800, Malaysia
[b] Ministry of Higher Education and Scientific Research, Studies Planning and Follow-up Directorate, Baghdad, Iraq
[c] CyberSecurity and Systems Research Unit, Islamic Science Institute (ISI), Universiti Sains Islam Malaysia (USIM), Nilai 71800, Malaysia
[d] Centre for CyberSecurity, Universiti Kebangsaan Malaysia UKM, Bangi, Malaysia

ARTICLE INFO

Article history:
Received 10 January 2020
Received in revised form 16 July 2020
Accepted 14 October 2020
Available online 7 November 2020

Keywords:
Audio steganography
Embedding
Data hiding
Classification
LSB substitution

ABSTRACT

Audio steganography is the process of hiding a message inside an audio container. This study aims to present a systematic review of audio steganography methods. Existing reviews of audio steganography suffer from a high overlap or a low segregation level among the methods reviewed due to poor classification criteria. These articles are also lacking in terms of depth of analysis and the number of articles reviewed. The data collection method includes surveying five databases, namely, Web of Science, IEEE Explore, ScienceDirect, Scopus, and Springer, resulting in a total of 134 articles. The review focuses on the embedding process of each method. The methods can be classified into several categories based on the most prominent idea in the embedding process; hence, a new classification is proposed. The proposed classification provides a scope for summarizing and understanding the most followed approaches in audio steganography. This study then critically reviews the strengths and weaknesses of each method on the basis of its embedding behavior. Other issues on the domain of embedding, carrier types, and evaluation environments are also discussed.

Contents

* Corresponding author.
E-mail address: ahmad88sabhany@gmail.com (A.A. AlSabhany).

https://doi.org/10.1016/j.cosrev.2020.100316
1574-0137/

1. Introduction

Steganography is usually classified alongside cryptography as methods for data security. Cryptography modifies a secret message into a meaningless and unreadable form to deny attackers from understanding it, whereas steganography confidentially conceals secret messages and their existence in an innocent carrier.

Perceptual transparency, hiding capacity, and robustness are the main requirements of steganography techniques used for evaluating the performance of any steganography approach (Fig. 1) [1]. Perceptual transparency is related to the similarity between the original cover and the stego files; it represents the degradation level of the cover files after hiding a secret message [2]. This characteristic is also related to the capability to avoid suspicions, as defined in [3]. Hiding capacity pertains to the percentage of message size to the cover file size; it can be denoted as the number of secret bits that can be embedded into the cover file [4]. Robustness indicates the resistance of the stego files against intentional and unintentional attacks [4,5]. However, three-way trade-off relationships exist among the three main requirements [5,6]. Increasing the performance in any requirement will decrease the performance of the other two requirements.

Many review articles have been published in the area of audio steganography. These articles can be classified into two categories; the first category classifies methods of audio steganography in general (e.g., such as [5,7–14]), whereas the second category is focused on certain types of method (e.g., [15–20]).

For the first category, a set of general approaches to audio steganography, such as the least significant bit (LSB) substitution, echo hiding, spread spectrum (SS), phase coding, and tone insertion was reviewed in [7–11]. The limitation in such review articles is that they do not cover all methods and ideas proposed in audio steganography.

The authors in [12] reviewed and proposed a classification, which was extended into many levels that highlighted multiple categories and types of audio steganography and watermarking methods. The classification included the domain of embedding (DOE) and some general approaches in audio steganography. Three groups of audio steganography methods were also presented on the basis of the main performance requirements. The multilevel classification suffers from a high overlap level and restrains certain methods of embedding in one domain.

The author in [14] reviewed audio steganography methods, in which detailed analysis was presented on the basis of the classification on the DOE. The article highlighted types of attack that counter audio steganography methods, the data hiding environment, and the most common criteria used in evaluating the proposed audio steganography and watermark techniques. However, the number of ideas covered in the review work was limited.

The authors in [5,13] conducted an in-depth analysis to classify and categorize the methods of audio steganography. In [13], methods were classified into three main categories, namely, temporal, transform, and encoder domains. The article also provided an in-depth analysis and evaluation of the reviewed methods. Meanwhile, [5] presented a comparative study to analyze and evaluate the performance of the reviewed methods. The classification of the methods was performed on the basis of the DOE; in each domain, the classification was branched out into subcategories. The article provided an in-depth discussion of the weaknesses and strengths of the reviewed methods and their influences on the main requirements of audio steganography. The main advantage of selecting the domain as a classification criterion is that it eliminates or reduces overlap. However, such classification can be general because it does not highlight the critical differences among methods. Both works highlighted the main differences by dividing and grouping the methods on the basis of the DOE [5,13]; however, this classification causes overlap issues. Some approaches were also not reviewed in both articles.

For the second category, review articles have included steganography or steganalysis or both. However, they have focused on one or several themes or aspects, such as methods, tools, media types, or specific file types. For example, a review article on audio steganography methods was performed on the basis of the concept of a genetic algorithm (GA) [16]. Another example that focused on steganalysis classifiers or steganalyzers was an in-depth comparative analysis over four important steganalyzers, namely, support vector machines (SVMs), Gaussian mixture models, deep belief networks, and recurrent neural networks on the basis of the comparison of detection accuracy [18]. The process was based on hiding secret information using three widespread software, namely, StegHide, Hide4PGP, and FreqSteg. Mel-frequency cepstral coefficients (MFCCs), were used for feature extraction, and the steganalyzers were used for classification. Another review that focused on audio steganography and steganalysis tools was conducted in [17]. Some commercial tools, such as Outguess, F5, Data Stash, S-tools, wbStego4, Stegdetect, and StegAlyzerSS, were reviewed in [17].

Other review articles have focused on several media types. For example, a review on steganography methods over various media types, such as text, image, audio, and video, including LSB coding, Spread Spectrum, and echo hiding was performed in [15], which highlighted several data hiding methods in the packet domain. In [17], the review was dedicated to steganography and steganalysis methods over image, audio, and HTML pages.

Furthermore, some review articles have highlighted a specific file type. For example, in [19], steganography methods in MP3 files were reviewed and classified into three main categories, namely, before, during, and after compression embedding. In addition, [20] focused on MP3 files; however, the review focused on the performance of a selected set of audio steganography methods in MP3. The performance evaluation included analyzing the steganalysis vulnerability, embedding capacity, and audio distortion. The review works have focused on several themes or aspects, by nature; however, they have not covered all methods of audio steganography.

In conclusion, the existing review articles suffer from overlapping or low segregation level among methods because of their classification process. Overlapping is a result of the high diversity in methods, whereas low segregation level is a result of using a general classification criterion that does not highlight the differences among the methods. Existing review articles are also lacking in terms of the limited number of the reviewed methods and ideas that do not fully cover or represent what has been proposed in this area. In addition to the limitations of the review articles discussed above, to the best of the authors' knowledge, no systematic review on digital audio steganography has been presented and published yet in this area.

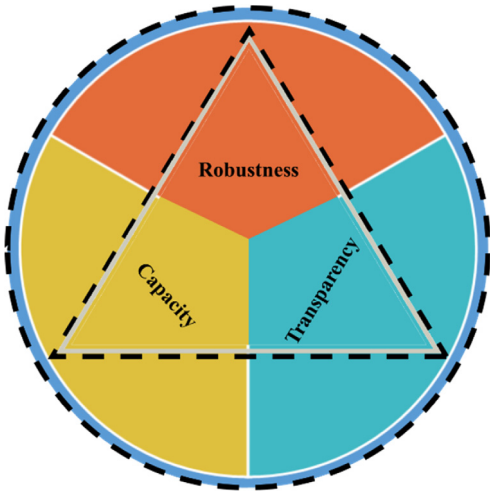


Fig. 1. Audio steganography trade-offs.

The present work aims to provide a systematic review of the methods in audio steganography. The methods can be classified into several categories based on the most prominent idea (MPI) in the embedding process; thus, a new classification is introduced. A comprehensive review of audio steganography papers that focus on the strengths and weaknesses of each method is then presented on the basis of the classifications.

The remainder of this paper is organized as follows. Section 2 presents the background and motivation. Section 3 presents the review method. The review results, which include the new classification and critical review of the methods in each classification, are discussed in Section 4. Section 5 presents additional findings on the DOE, carrier type, evaluation data, and evaluation metrics. In Section 6, the discussion is presented. A brief review of recent audio steganalysis methods is presented in Section 7. The limitations and future works are presented in Section 8, followed by the conclusions in Section 9.

2. Background and motivation

The process of audio steganography in all methods includes a way of inserting a message in an audio signal. The main concept of audio steganography is illustrated in Fig. 2.

The three main aspects where contribution in audio steganography are possible are the DOE, method of data injection (MODI), and carrier type. The first aspect is the domain, which is related to the final form of the cover data unit accessed for embedding. Time, Wavelet, and frequency domains are commonly used in audio steganography. The second aspect is the MODI. This method is the main embedding process, in which the message is inserted in the cover file. Among the MODIs are the LSB substitution, interpolation, quantization index modulation (QIM), syndrome-trellis codes (STCs), matrix embedding strategy (MES), and singular value decomposition (SVD) [6,21–24]. The last aspect is the carrier type, which is the type of cover used. In most cases, the carrier is a file that can be of WAV, MP3, or AMR type. However, in the case of voice over IP (VOIP) steganography, the carrier is a network packet or voice stream.

The sheer number of theories, approaches, and implementation makes it overwhelming to understand and summarize the methods in this area. Previous review works, such as those in [5,7–14], have their limitations in terms of poor classification, overlap, and depth of analysis. The present work extends the efforts of the existing reviews and aims to provide a better scope for understanding and differentiating the methods of audio steganography.

3. Methodology

The review methodology for this work involves data collection, method analysis, main behavior construction, and similarity-based classification.

The methods are collected from five reliable digital databases, namely, Web of Science, ScienceDirect, IEEE Explore, Scopus, and Springer, on the basis of a unified query that comprises three keywords, that is, “audio steganography” OR “audio embedding” OR “audio hiding”. The reviewed articles in the study are published between 2007 and 2017. Fig. 3 shows the results of data collection.

In the method analysis step, the methods are analyzed to extract the key idea and the operation details in each method, such as the DOE, MODI, carrier type, message type, supportive methods, main disadvantage, evaluation metrics, and experimental data. Then, in the behavior construction step, similar embedding methods are connected. Finally, all the methods are classified on the basis of the MPI to resolve the most common embedding behaviors in audio steganography.

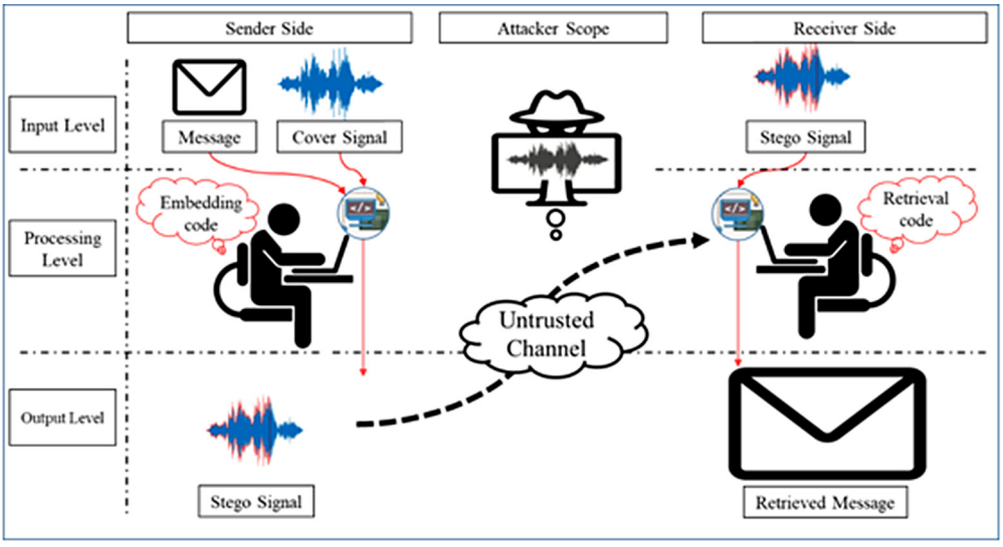


Fig. 2. Concept of audio steganography.

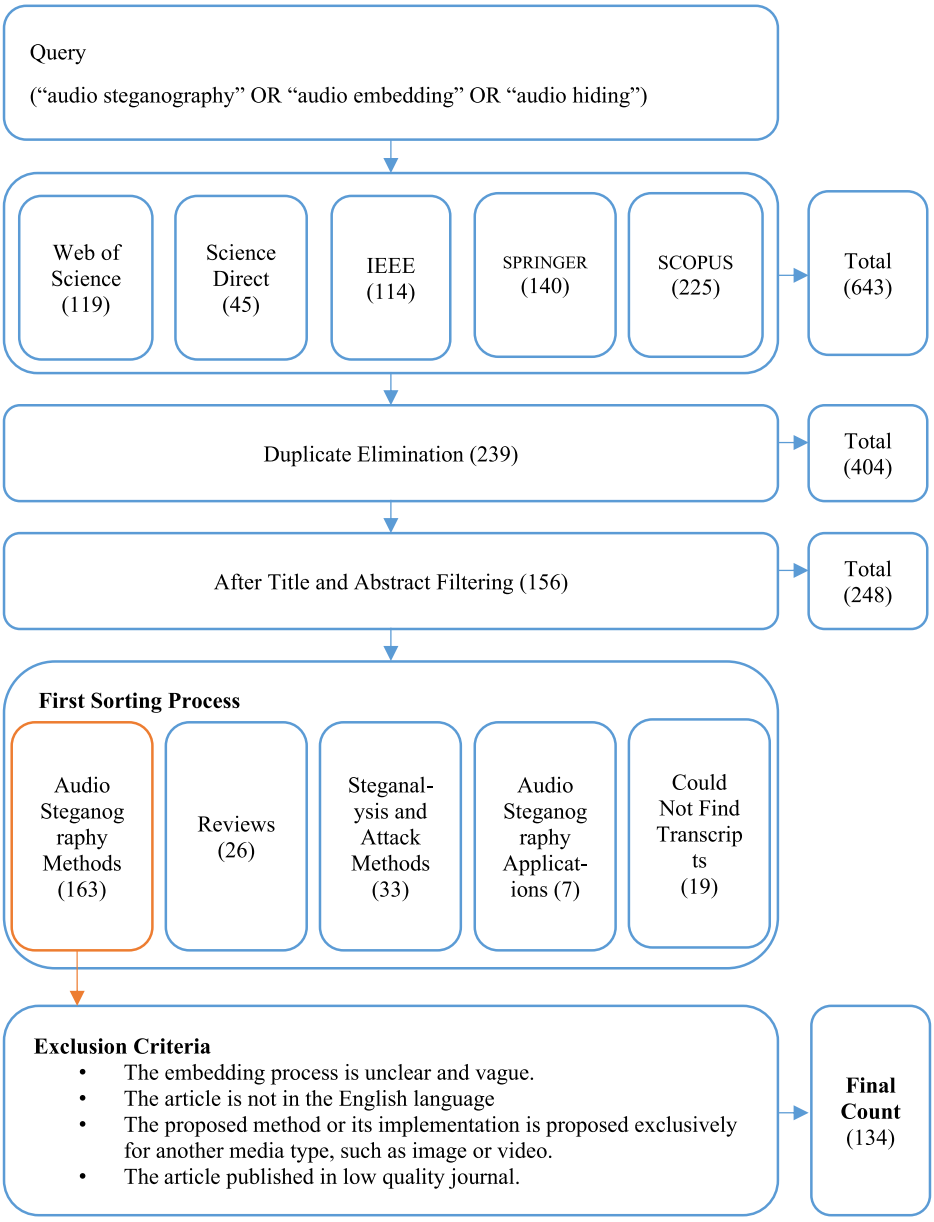


Fig. 3. Data collection and filtering process.

4. Review results

The extensive analysis of the methods indicate that audio steganography methods follow eight main embedding approaches or behaviors. Each of these behaviors is based on a theory or a key idea. The resultant behaviors are linear or sequential embedding, the selective-based embedding, frequency thresholding embedding, error minimizing embedding, pattern-matching embedding, phase coding, spread spectrum, and tone insertion. Fig. 4 shows the embedding behaviors and the ratio of each behavior. Each embedding behavior and the methods under each behavior are then analyzed and discussed thoroughly. The highest percentage of the methods are based on either selective-based, or sequential embedding. A good percentage of methods are found following frequency masking, amplitude thresholding, and error minimization embedding. A small percentage of the methods are found based on pattern-matching embedding, phase coding, SS, and tone insertion. Finally, some methods on were found following other behaviors.

Fig. 5 presents a tree diagram of the proposed classification and the methods/articles classified under each category.

Here, each section represents an embedding behavior group. In each group, a table that includes a number of general features is provided to capture the highlights and enable comparisons. Among the general features, four important features that must exist in any steganography method are the *DOE*, *MODI*, *cover type*, and *message type*. *Supportive techniques* are added to highlight any additional approach incorporated with the embedding method, such as data encryption, data compression, data scrambling, and key sharing methods. The *disadvantage* feature is also added to highlight the main limitation of the reviewed methods. In addition to these features, evaluation features, such as the *metrics* and *data* used, are included to highlight the most common evaluation features. To capture the inner differences among the methods of the same group, a special feature is introduced to distinguish each group. The special feature has two main goals, that is, to validate and support the classification process and to capture the inner differences among methods of the same group. In the tables of the

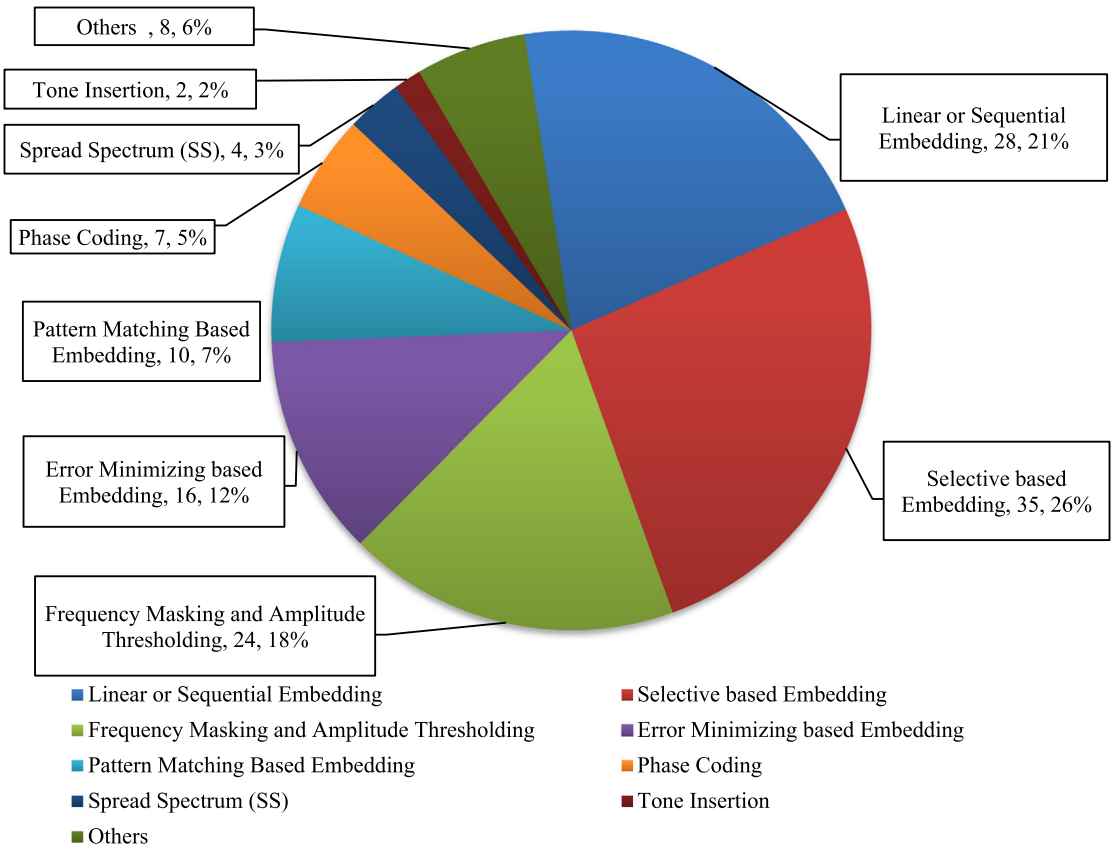


Fig. 4. Embedding behaviors and method distribution.

upcoming subsections, the special feature will be marked with an asterisk.

4.1. Linear or sequential embedding

Sequential embedding is based on sequentially accessing the cover data units and then performing the MODI. Methods in this group follow such behavior regardless of the utilized domain, MODI, or carrier type. The most important feature in this group is the sequential access of the host signal components for embedding. Fig. 6 shows the main concept of sequential embedding, in which the data units are accessed sequentially for embedding. This pattern of embedding is observed in the methods summarized in Table 1.

The most general advantage in the methods of this embedding behavior is that they achieve the highest embedding capacity and full embedding efficiency. Most of the data units are utilized for embedding. However, the main disadvantage of the methods of such behavior is their low robustness against all types of attack. In many methods of such behavior, the embedded message (encrypted or not) can be easily retrieved by simply concatenating the LSBs from stego audio samples, Wavelet coefficients, or frequency components. Another problem in some methods of this behavior is their low robustness against intentional and unintentional signal processing attacks, such as compression, low-pass filtering, and noise addition. This problem is due to the high fragility of a bit state in the LSB position against attacks. The detectability of such methods is also a critical issue. The main weaknesses regarding detectability are the capability of steganalysis methods to differentiate between two types of cover, namely, original and stego. This detection capability is improved if the cover audio contains silent periods because the added noise will be obvious and are thus easy to detect. Embedding capacity is increased by increasing the NBPDU. In many methods that utilize high NBPDU, high distortion is introduced to the signal. In addition to the high distortion and the lowered quality, such behavior increases the chances of message detection using statistical steganalysis methods, because the difference between embedded and original data units in the same signal is evident when the segments of the stego are compared with one another. Such a problem is increased when the signal contains silent intervals, such as in speech.

In Table 1, important embedding features are highlighted to capture the main differences and similarities in the methods of this behavior. The common factor in the methods in Table 1 is that they share a sequential embedding behavior. However, variables DOE and MODI are used in these methods. Furthermore, the implementation of supportive techniques for data encryption, digital signatures, data compression, and scrambling techniques is the focus of many of the approaches in this group.

4.2. Selective-based embedding

In this group, the opposite of sequential embedding is used. The methods under this category select the embedding locations on the basis of defined selection criteria, such as shared stego key, random sequence, or most significant bit (MSB) selection. Fig. 7 shows the main process in this embedding behavior, in which the data units are selected on the basis of the criteria. The most important shared feature among the methods under this group is the selection process. The selection process can be implemented on all data access domains and regardless of the MODIs. The most common domain in this group is the time domain due to its simplicity and customizability. The main feature of these methods is their goal to incorporate randomness with audio steganography to increase security. A common scenario in

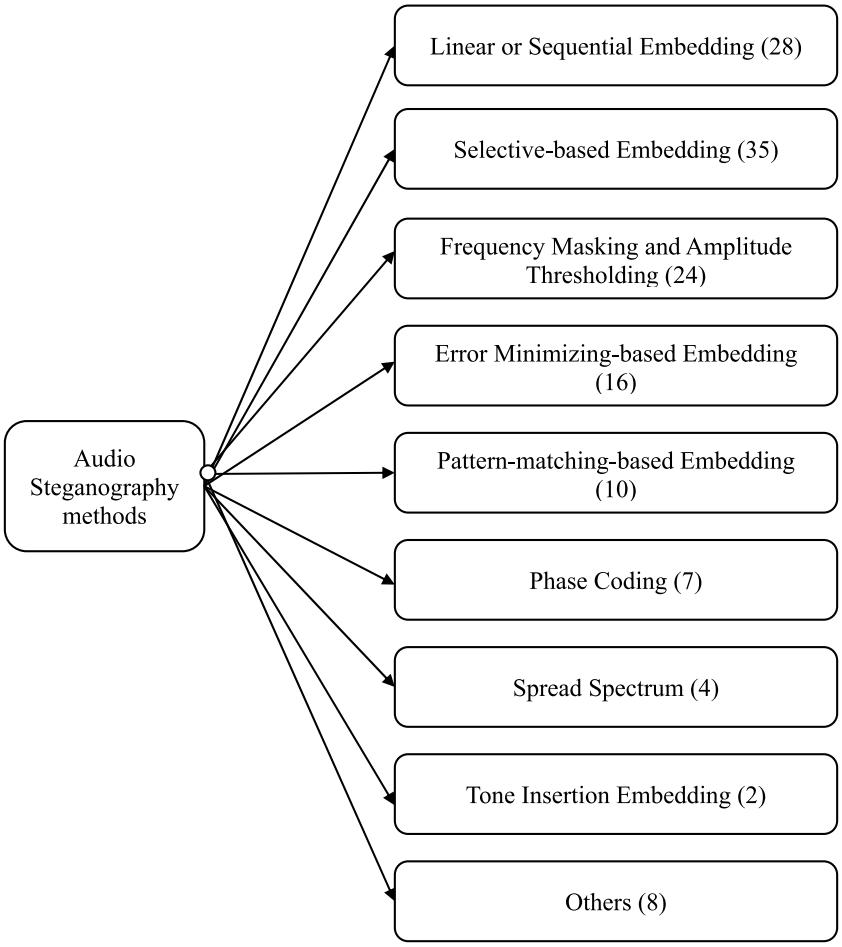


Fig. 5. Proposed classification.

many time-based methods is selecting the next sample to be used for embedding, which is similar to jumping in some samples and ignoring some. Such behavior randomly scatters the message data over the cover signal, which reduces the chances of detectability by statistical steganalysis methods. Consequently, the capacity and the embedding efficiency are reduced by the same rate. In some methods, the next embedding locations are determined based on the value of the current embedding data unit. Another disadvantage of such behavior is the fragility to errors, because the data retrieval at the receiver side is performed in a chained manner. Hence, the data retrieval of each bit depends on the correctness of the previously recovered bit; if one error exists at any point, then data recovery will fail. Some methods (e.g., [50–53]) operate by different approaches from the methods in the main scenario; they operate in a sequential manner, but with variable bit index selection. In these methods, low scatter ability is achieved. This pattern of embedding is observed in the methods summarized in Table 2.

The main advantage of the methods with such behavior is that they reduce or eliminate the vulnerability of message retrieval attacks. Selective embedding is more robust against message retrieval attacks than sequential embedding. However, such an advantage does not mean fully satisfying the robustness requirement. The robustness of such methods against message detection attacks is variable and is dependent on many factors, such as the severity of the change. However, such methods mostly obtain low robustness against message destruction attacks. Many methods target LSB locations or deeper locations with a low effect on the audio quality (Wavelet detail coefficients); consequently, the signal becomes vulnerable to signal manipulation attacks. The capacity of methods of such behavior is variable and is limited mainly by the ratio of data units excluded from being used in the embedding process. Methods of high exclusion rate usually obtain high SNR and PSNR values due to low embedding efficiency.

4.3. Frequency masking and amplitude thresholding

In this group, a pattern of similar ideas is observed and recorded. The main idea is constructed around conditioning or customizing the embedding in data units on the basis of an acoustical condition applied to the sound signal components. A number of keywords are usually incorporated with this embedding behavior; among them are amplitude threshold, energy level, silent or mute intervals, noisy periods, and inactive frames. The idea of embedding in nonsilent or noisy intervals of the audio signal has been implemented in all domains. However, in time- and Wavelet-domain methods, this idea has been referred to as amplitude thresholding. In many frequency-based methods, the idea has been referred to as frequency masking, because the human auditory system (HAS) cannot identify artifacts of the sound signal above or below a specific frequency range. Methods of this embedding behavior depend mainly on defining an acoustic feature that is used for conditioning the embedding process. The most common acoustic features are the amplitude threshold and energy level. Fig. 8 presents an overview of this embedding behavior alongside an example.

However, this idea has been implemented in a contradictory manner. Many researchers have designed methods that prevent embedding in silent or near-to-silent intervals of the audio

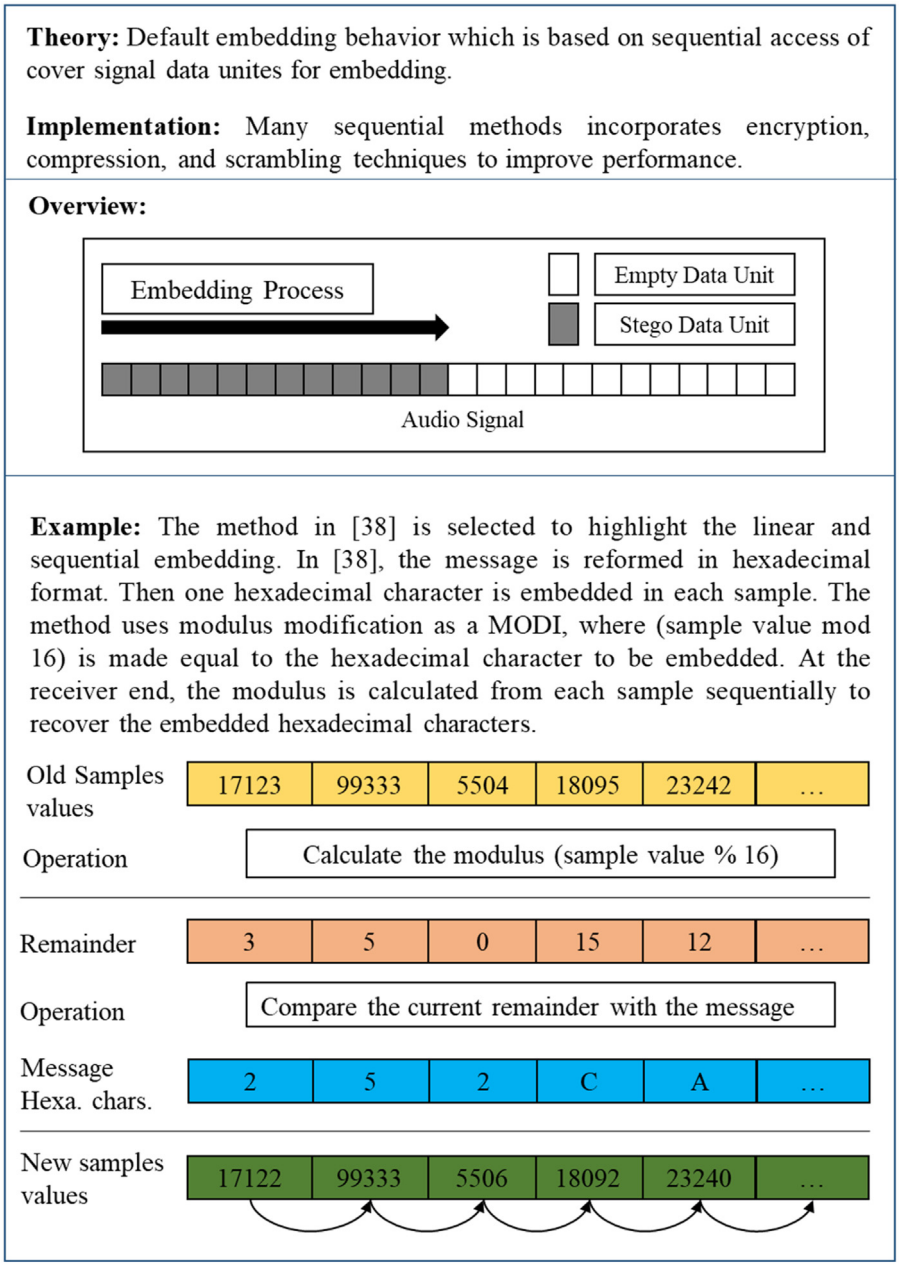


Fig. 6. Sequential or linear embedding.

to avoid or minimize the produced hissing sounds. Other researchers have designed methods to embed high payloads in the silent intervals of the audio. In all cases, the MPI in this group is the selection of embedding locations based on the acoustic condition. This group shares some similarities to the selective embedding behavior, given that both behaviors select the embedding locations. However, in selective embedding, the selection criteria are not related to any acoustic quality of the data units. Instead, the selection is related to the criteria that are aimed to incorporate randomness. The key difference and the common feature among the methods of this group is the existence of the acoustic condition. This pattern of embedding is observed in the methods summarized in Table 3.

This section features the acoustic embedding condition. Each method handles the acoustic condition in a different manner. For example, in adaptive embedding, many LSBs are allocated for data units with high values, whereas none-to-few LSBs are allocated for low-valued data units.

The major advantages of the methods of such behavior are the minimization, camouflage, or suppression of the introduced distortion in the sound signal. Such a behavior theoretically supposedly fools the human ears and therefore reduces the chance of detection by an active human warden. However, the detectability of such methods against statistical steganalysis methods requires further research, as demonstrated in [106]. The robustness of such behavior against signal manipulation attacks is variable and dependent on the individual method. Moreover, the audio quality represented by SNR and PSNR levels is variable and does not show any clear or significant pattern of improvement in quality. In terms of capacity, this behavior generally limits the capacity due to data unit exclusion or rejection by the selection process. Moreover, the embedding capacity is the least prioritized requirement in this group. However, this feature is a general attribute of this embedding behavior, and variable capacity levels exist among the methods of this behavior.

Table 1
Linear and sequential methods.

Ref., Year	DOE	MODI	Supportive techniques	Disadvantage	Number of bits per data unit (NBPDU)*	Carrier type	Message type	Evaluation features	
								Metrics	Data
[25], 2010	Time	Amplitude differencing, *a	N/A	Low robustness	Multiple/ variable	Two WAV files	Image	HC, TC, APS, SD, Kurtosis, ASC	WAV surfer website
[22], 2010	Time	LSB substitution	Zigzag scrambling	Weak security	1	WAV	Text	No Experiments	
[23], 2011	Discrete Wavelet transform (DWT)	Coefficient replacement	Noise addition	High distortion, low retrieval accuracy	Whole coefficient	WAV	Audio (LPC)	MOS	TIMIT
[24], 2011	Time	Modulus modification	Encryption	*b	Variable	WAV	N/A	HC, RMS, PSNR	Website
[26], 2012	Time	LSB substitution	XOR encryption	Low robustness	1	WAV	WAV	N/A	One file
[27], 2012	Time	LSB substitution	New dictionary compression	Low robustness	1	WAV	Text	SNR, BPC (compression rate)	CSFDG
[28], 2012	Time	Time	Simple LSB XOR	Weak security	1	WAV	N/A	MSE, PSNR, HC	N/A
[29], 2013	Time	LSB substitution	N/A	Low robustness	2	WAV	Text	SNR	CSF
[30], 2013	DWT	Adaptive sorting, scrambling, LSB substitution	Hardware implementation	Extremely high distortion	8 LSB, in CA; 6 LSB, in CD	WAV (speech)	WAV (speech)	SNR, SPCC	N/A
[31], 2013	Time	Revisable LSB substitution	Near neighbor error prediction and modification	Low robustness	1	WAV	N/A	SNR	Three CSF files
[32], 2013	FFT, DCT	Sum differencing and interpolation	Hashing and digital signatures	High distortion, *c	1	WAV	WAV	SNR, BER, distance, authenticity	CSF
[33], 2014	LWT	LSB substitution	Dynamic encryption key	High distortion	2×6 LSB in 1st CD; 4 LSB in 2nd CD	WAV	WAV	SNR, NC, HER, fourth first moments	CSF
[34], 2014	DWT	Spline interpolation	Dynamin watermark generation	Low capacity	1 bit in three coefficients	WAV	WAV	SNR, PESQ	CSF
[35], 2015	Time	LSB substitution	N/A	Low robustness, high distortion	Variables 4–7 LSB per sample	WAV	Text	SNR, PSNR, MSE, SLT	CSFMG
[36], 2015	MDCT MP3	LSB substitution	LAME 3.96.2 MP3 encoder	High distortion	1	MP3	N/A	BER, SNR, PSNR, ROC	CSFMG
[37], 2015	Time	LSB substitution	AES encryption	Low robustness	1	WAV	Text	SNR	NOIZEUS
[38], 2015	Time	Modulus modification	7-bit compressed ASCII code	*b	Variable	WAV	Text	SNR	N/A
[39], 2015	DWT	LSB substitution	Customized encryption, compression methods	Low security	Variable	WAV	Text	SNR, SPCC	CSF
[40], 2016	Time	LSB substitution	Encryption, sample rate	Low robustness	1	WAV	Text	SNR	CSF
[41], 2016	Time	LSB substitution	SHA-512	Low robustness	1	WAV	PDF	N/A	N/A
[42], 2016	Time	LSB substitution	RSA	Low robustness	1	WAV	Text	N/A	N/A
[43], 2016	DCT	N/A	Dynamic parameter setting, image steganography	Extremely slow and complicated series of encryption steps	N/A	WMV	Text	SNR, PSNR	N/A
[44], 2016	Time	Sample digit modification	N/A	Low robustness	1	WAV	WAV	SNR, PSNR	One file
[45], 2016	DWT	QIM	Orthogonal variable spreading factor	Reduced capacity	1	WAV	Text	PSNR, HER	*d
[46], 2017	Time	LSB substitution	Chaotic maps, XOR encryption	Low robustness, non-blind	N/A	WAV	Text	N/A	CSF
[47], 2017	Time	LSB substitution	N/A	Low robustness	2	WAV	Text	SNR, PSNR	CSFMG
[48], 2017	DWT	Threshold embedding	Chaotic maps, randomized order	High distortion	1	WAV	Image	MSE, PSNR, MSSIM	CSF
[49], 2017	ECG signal, DWT	LSB substitution	RSA, embedding location scrambling	Low robustness, high distortion	Variable; 5 or 6	ECG	Text	PSNR, PRD	N/A

*a, The method initially calculates the difference between the amplitude of two corresponding audio samples. In every two samples, the method embeds four message bits. Embedding is performed by replacing four message bits with the sample difference. The new difference is achieved by manipulating the sample value by addition or subtraction. The method aims to distribute the new difference value over the two samples evenly.
*b, The embedded data can be extracted easily by the adversary by identifying the modulus value in each sample.
*c, The embedding is performed in two parts. The first part embeds the digital signature using the spectrum of the FFT coefficients, and the second part embeds the message using DCT. In the first part, the signal is divided into frames, and FFT is performed. Each frame is further divided into two groups, A and B, by random selection. The sum of each group is calculated. To embed 0, the sum of Group A must be lower than that of Group B. To embed 1, the sum of Group A must be larger than that of Group B. To satisfy the conditions, all the coefficients in both groups are manipulated by addition and subtraction. In the second part, the same frame division process is performed, and the DCT coefficients for each frame are calculated. Each frame is further divided into two regions, A and B. The middle point in each frame is adjusted on the basis of a secret bit value. If the bit is 1, then the coefficient value is set based on the average of Region A. Otherwise, the value is set to the average of Region B. The main concern in such methods is to make each embedding independent from the other, such that they do not overwrite and corrupt each other. Orthogonality or separation can be performed on a spatial basis; for instance, the first embedding submethod initially embeds, and the other selects from where the first ends. Generally, each method must be aware of forming a pattern that can be noticed by an audio steganalysis method, especially if both methods form a shift in patterns.
*d, www.freesound.org

4.4. Error minimization-based embedding

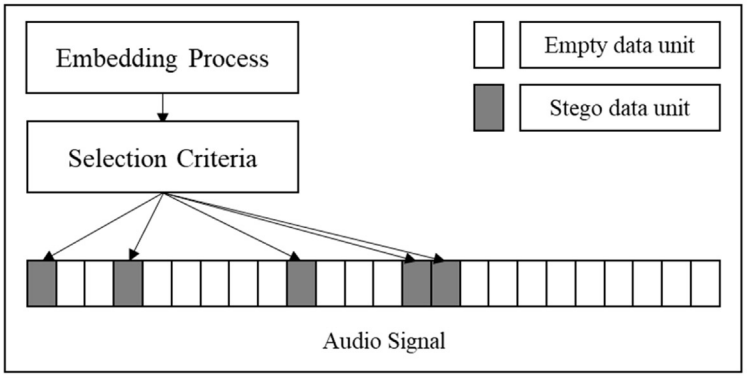
The methods under this group aim to reduce the effect of embedding by reducing the resultant error and therefore reduce the overall signal distortion. In this group, several methods, such as MES, STCs, and GA, are used. An overview of this embedding behavior alongside an example is illustrated in Fig. 9. This pattern of embedding is observed in the methods summarized in Table 4.

This method mainly aims to reduce the effect of embedding (error) and therefore obtains a high-quality audio signal that is extremely similar to the original cover file. Such behavior mainly falls under improving perceptual transparency. It also optimizes robustness against statistical steganalysis attacks. However, such behavior does not necessarily increase robustness against signal manipulation or destruction attacks. Consequently, low robustness against signal destruction attacks is a common

Theory: Scatter and randomize the embedding locations to increase the security against unauthorized retrieval and to reduce detectability by statistical steganalysis methods.

Implementation: This approach is usually implemented by considering embedding keys in the forms of secret vectors, MSB combinations, and hash functions.

Overview:



Example: The method in [56] is selected to highlight the selective based embedding. The embedding process is guided by the sample and bit index selection based on the MSB. The first three MSBs determine the sample and the bit index of the next message bits. Each MSB combination indicates a certain sample and bit index value.

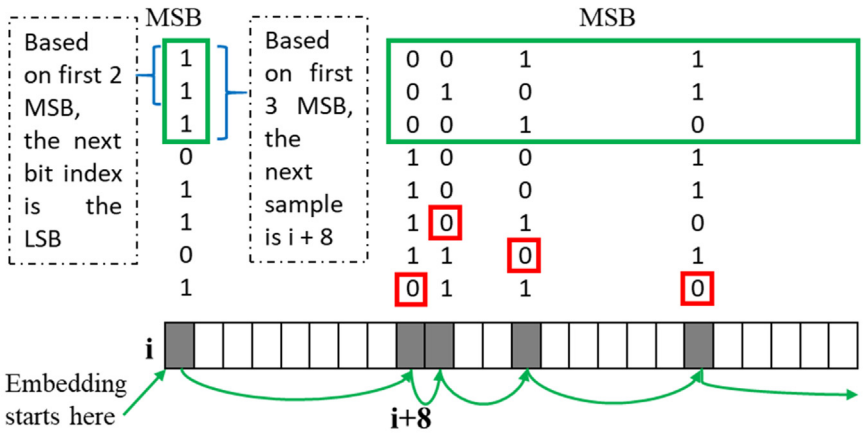


Fig. 7. Selective-based embedding.

disadvantage among the methods of such behavior, because most of the methods are implemented in the time domain and not because of a deficiency in the algorithm used. The capacity for methods of this behavior is variable and based individually on the method. Notably, MES and STC have been increasingly used in recent publications due to their performance qualities. The error minimization algorithm is a special feature in this group. Thus, the main algorithms, namely, MES, STC, and GA, are used in this group. However, many researchers have proposed similar versions of the GA with a similar approach.

4.5. Pattern-matching-based embedding

This group is formed after observing and recording a pattern of similar ideas. The idea is based on comparing a binary sequence from the message with another binary sequence from the host signal in an effort to find a match. In methods within this group, embedding is guided by the matching level between the two binary sequences. The method is implemented in two main domains, namely, time and DWT. The most important concept in this group is the binary comparison between message and host signals. In its core, the main idea aims to reduce the embedding error, similar to the methods in Section 4.4. However, in this group, such a goal can be achieved by taking advantage of the

Table 2
Selective embedding methods.

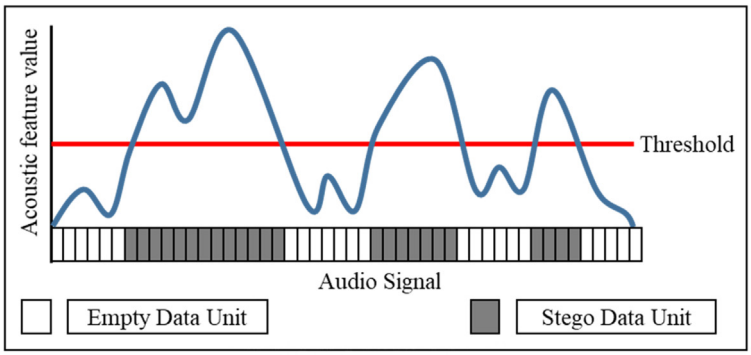
Ref., Year	DOE	MODI	Supportive techniques	Selection criteria*	Main disadvantage	Carrier type	Message type	Evaluation features	
								Metrics	Data
[54], 2008	WAVelet packet transform (WPT)	Pattern modification *a in WPT	N/A	Pseudo sequence	–	WAV	N/A	Signal-to-noise ratio (SNR), BER, MOS	CSF
[50], 2008	Time	LSB substitution	Variable bit index	MSB	Low robustness	WAV	WAV	SNR	MSFMG
[55], 2010	Time	LSB substitution	Variable bit index, XOR encryption	MSB	–	WAV	Text	SNR, BER	TIMIT
[4], 2011	Time	LSB substitution	AES encryption, variable bit index	MSB	Low robustness	WAV	Text	N/A	N/A
[56], 2012	Time	LSB substitution	Compression HE, AES encryption, variable bit index	MSB	Low robustness	WAV	Text	MSE, PSNR	N/A
[57], 2012	Time	LSB substitution	XOR encryption	MSB + dynamic value	Weak and short encryption key	WAV/MP3	N/A	N/A	N/A
[58], 2012	Time	LSB substitution	XOR encryption	Fibonacci sequence	Extremely low capacity	WAV	Text	SNR	N/A
[59], 2012	DWT	*b	LPC-to-LSF conversion	Fixed variable	Low robustness	WAV	WAV	SegSNR, PESQ	NOIZES
[51], 2013	Time	LSB substitution	Variable bit index	Hash function	Low robustness	WAV	Variable files	MSE, SNR,PSNR, MOS	N/A
[60], 2013	DCT	Coefficient modification	Segmentation	Second highest coefficient	Non-blind, low capacity	WAV	Image	N/A	CSF
[61], 2013	Time	LSB substitution	Encryption	MSB	Low robustness	WAV	Text	MOS, SNR, BER	CSFMG
[62], 2013	VOIP	LSB substitution	AES encryption	Fixed variable	Low robustness	VOIP	Text	MOS, PESQ	N/A
[63], 2013	Time	LSB using sample selection matrix	Arnold transform scrambling	N/A	Low capacity, low robustness	WAV	Image	ODG	CSF
[52], 2014	Time	LSB substitution	Variable bit index using MSB	MBS	Low robustness	WAV	N/A	No experiments	
[64], 2014	VOIP	LSB substitution	Key sharing, AES encryption	Fixed variable	Low robustness	VOIP	N/A	PESQ	CSF
[65], 2014	DWT	LSB substitution	N/A	Key	Low robustness	MP3	Image	PSNR	CSF
[66], 2014	DWT	Threshold embedding	XOR encryption	N/A	Low capacity, high signal distortion	WAV	Variable	SNR	CSF
[67], 2014	DCT	Threshold embedding	Compression differences comparison *e	Backpropagation neural network	Low capacity due to high elimination rate	WAV	Text and image	SNR, SDG	CSFMG
[2], 2014	Time	LSB substitution	Huffman encoding, RSA encryption, variable bit index	MSB	Low robustness	WAV	Text	N/A	One file
[68], 2015	FFT, Time	Component modification, LSB substitution	N/A	Secret key, matrix lower triangle	N/A	WAV	WAV, image	PSNR, MSE	CSF
[53], 2015	Time	LSB substitution	Variable bit index	MSB	Low robustness	WAV, AVI	Text	N/A	N/A
[69], 2015	Time	LSB substitution	Multiple sample modification, password	Local maxima and minima	Non-blind	WAV	text	N/A	N/A
[70], 2015	Time	LSB substitution	Data compression, data encryption	PRNG	Low robustness, low capacity	WAV	N/A	SNR, SLT	
[71], 2015	Time	LSB substitution	Parity coding, variable bit index	MSB equation	Low robustness	WAV	Text	SNR, PSNR, MSE, MOS	CSFMG
[72], 2015	Time	LSB substitution	AES encryption	Third-party key + MSB	Low robustness	WAV	Text	SNR, BER	CSF
[73], 2016	VOIP	LSB substitution	N/A	Number of hiding vectors	High complexity	VOIP	Text	PESQ, MFCC	N/A
[74], 2016	Time	LSB substitution	Compressed ASCII table	MSB equations	Low robustness	WAV	Text	PSNR	N/A
[75], 2016	Time	LSB substitution	Encryption, digital signature	Secret key	High complexity, low robustness	WAV	Multiple types	RMS, PSNR, SNR	Website Minitab
[76], 2016	Time	LSB substitution	Chaotic encryption, variable bit index	MSB	Low robustness	WAV	N/A	PSNR, MSE	One file
[77], 2016	Time	Parity bit	MD5	Stego key	Low robustness	WAV	Text	N/A	N/A
[78], 2017	Time	LSB substitution	Diffie–Hellman key transfer, AES	Stego key	Low robustness	WAV	Text	No experiments	
[79], 2017	VOIP	LSB substitution	Message digest, AES	Stego key	Low robustness	VOIP	N/A	PESQ, SNR, MWW	N/A
[80], 2017	FFT	Component modification	N/A	Synthesized pitches	Non-blind, high complexity	WAV	N/A	SNR, SNR,BER	N/A
[81], 2017	LWT	Threshold embedding	RSA, chaotic map techniques	PN sequence generator	Low capacity	WAV	Image	MSE, SNR, PSNR	CSFMG
[82], 2017	Time	Interpolation	PNG scrambling	Embedding condition	Low robustness	WAV	AMR	SNR, BER	CSF

*a, Embedding is performed by manipulating the coefficient values by addition or subtraction to construct a pattern that is related to a message segment.
*b, The method calculates the discrete Fourier transform for the detail vector and then modifies the magnitude components.
*c, Embedding is conducted by changing the sampling points on the matrix diagonal that represents the audio frame, matrix generation, and scrambling on the basis of a shared key.
*d, This algorithm presents a perfect example of why the embedding location should not be connected to or based on the message binary bits because the receiver will not be able to determine the location of the embedded data.
*e, The method compresses the WAV file to MP3 and then selects the blocks that obtain the minimal differences after compression.

Theory: Embedding in the noisy parts of the audio signal provides better hiding because the artifacts caused by embedding are masked by louder noises.

Implementation: acoustic features such as amplitude, magnitude, energy level are key to identify silent or quieter intervals in the signal. Then the embedding is made conditional based on the status of each data unit, frame of samples, or time interval.

Overview:



Example: The method in [92] is selected to highlight the amplitude thresholding embedding behavior. The method uses a fixed threshold to select the sample. If the sample is equal or above the appointed threshold, 8 LSBs will be substituted with message bits.

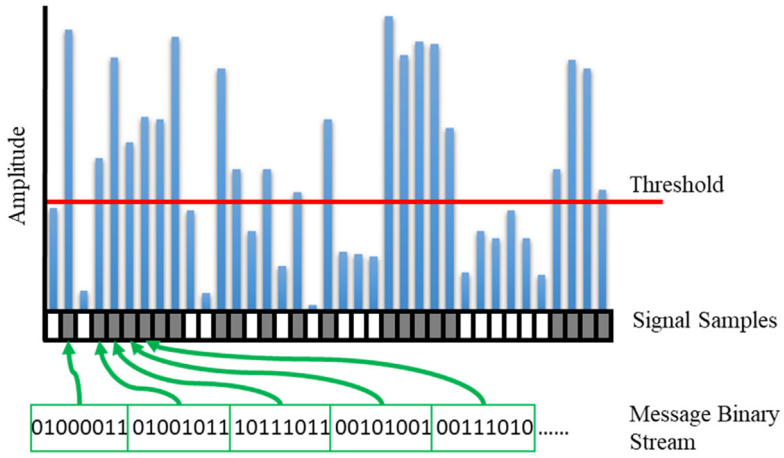


Fig. 8. Frequency masking and amplitude thresholding embedding.

existing binary combinations in the cover signal to recreate the secret message. An overview of this embedding behavior alongside an example are illustrated in Fig. 10. All methods in this group share two important steps, namely, matching and signaling. In the matching step, the algorithm searches for a match for a binary sequence of the message. Afterward, regardless of the result, the method signals the result of the search in shared locations of the audio to be later interpreted by the receiver. Conversely, the receiver can decide on how to reassemble the message correctly after reading the secret signal from the shared locations. This pattern of embedding is observed in the methods summarized in Table 5.

The main disadvantage in methods of such behavior is the high complexity of time and space because the majority of these methods are based on linear search algorithms that scan the cover file or parts of it. Another key issue in these methods is embedding efficiency because the goal of such behavior is to reduce the number of cover modifications. Obtaining a high embedding efficiency requires that the number of bits that are used for marking or signaling one data unit is less than the message segment that has been matched, that is, breaking the one-to-one linear ratio. Otherwise, the costly (time and space) matching process achieves only inefficient and, in many cases, insecure encryption purposes. Moreover, in all the methods that include cover segmentations, an idle case must exist to signal the no-match case. The embedding cost of the idle case will increment the total number of modifications for each time that the algorithm finds a mismatch, thereby further reducing the embedding efficiency and increasing the distortion level without embedding any real bits of the message. Another important issue

Table 3
Frequency masking and thresholding-based methods.

Ref., Year	DOE	MODI	Supportive techniques	Acoustic embedding condition*	Disadvantage	Carrier type	Message type	Evaluation features	
								Metrics	Data
[83], 2007	LWT	LSB substitution	N/A	Multiple thresholds, adaptive embedding, *b	Low robustness (sequential)	WAV	N/A	SNR, HC	CSF
[84], 2007	LWT	LSB substitution	N/A	One threshold, *a	Low robustness (sequential)	WAV	Binary string	SNR, BER, MOS	CSFMG
[85], 2008	LWT	LSB substitution	Data scrambling	Multiple hearing thresholds, empirical rule, adaptive embedding	Low robustness (sequential)	WAV	Binary string	SNR, BER, MOS	CSFMG
[86], 2008	FFT	Component differencing	N/A	Frequencies are selected from the most common frequencies (frequency masking)	Low capacity, high distortion	WAV	Random data	N/A	TIMIT
[87], 2009	FFT	Component differencing	N/A	Minimum threshold of sound pressure level, hearing threshold	Low capacity, high distortion	WAV	Random data	N/A	TIMIT
[88], 2009	Time	LSB substitution	Compression, encryption	Multiple hearing thresholds, empirical rule, adaptive embedding	Low robustness (sequential)	WAV	N/A	SNR, BER, Payload, MOS	CSFMG
[89], 2010	DWT	QIM	N/A	Frame selection is based on average energy to select mute periods; counter-propagation neural network	Low capacity	WAV	Image	NC, SNR	CSFMG
[90], 2010	FFT	*c	Harsh message compression	Selection is in the range of 18–22 KHz in the frequency level	Low capacity, high message degradation	MP3	Audio	ODG, PEAQ, ITU-R grade	N/A
[91], 2010	Time	LSB substitution	N/A	This paper includes two methods; the first one uses two fixed amplitude thresholds and embed 1 or 2 bits based on the sample value; the second method uses a dynamic threshold that is calculated from the average of the near samples	Low robustness	WAV	N/A	SNR	CSFMG
[92], 2010	Time	LSB substitution	8 NBPDU	Fixed amplitude threshold	Low robustness, high distortion	WAV	N/A	SNR, SDG, MOS	CSFMG
[93], 2011	PWT	LSB substitution	8 NBPDU	Fixed predefined hearing threshold is used to classify and select the appropriate subbands	Low robustness, *d	WAV	N/A	SNR, PAQM, MOS	CSFMG
[94], 2011	Time	LSB substitution	N/A	The method selects the shortest silent periods for embedding the watermark	Low capacity	WAV	N/A	SNR, ASNR	CSF
[3], 2012	DWT	LSB substitution, *e	5 NBPDU	The method hides secret data in high-frequency locations to achieve masking	Low message retrieval accuracy	WAV (speech)	WAV (speech)	Fourth first moments	N/A
[95], 2013	VOIP	LSB substitution	PCC encryption	The method analyzes audio frames and decides if the frame is active or inactive based on energy threshold; the method hides in inactive frames	Low robustness	PCM	N/A	N/A	N/A
[96], 2014	FFT	LSB substitution, *f	N/A	The method selects the first 16 frequencies for embedding that obtain a magnitude value above 15	Low capacity	WAV	N/A	ODG	N/A
[97], 2015	Time	LSB substitution	Variable bit index	Only samples with magnitudes of at least 3584 are considered for bit modification	Low robustness (sequential)	WAV	N/A	MBSD, PESQ	TIMIT
[98], 2015	Time‘	LSB substitution	Variable bit index, *g	The method considers only the sample above the amplitude threshold (3584) for embedding	Reduced capacity	WAV	N/A	SNR, BER	Greenflag database
[99], 2015	Time	LSB substitution	Chaotic scrambling, *h	The method aims to minimize embedding in long silent intervals, thereby avoiding hissing sounds; adaptive embedding	High complexity, high time and space overhead	WAV	N/A	MOS	CSFMG
[100], 2016	MP3-post	Direct substitution, *i	Encryption, scrambling	The method selects “count1” region in each frame for embedding because it contains intermediate frequency level and thus causes intermediate distortion	Low capacity, high distortion	MP3	N/A	PEAQ, ODG, NMR	CSFMG
[101], 2016	DCT	Coefficient modification, *j	N/	ZCC and STE are used to separate the voiced and unvoiced frames, such that if ZCC is small and STE is high, then the frames are voiced frame; otherwise, they are unvoiced	Low capacity	WAV	N/A	SNR, BER, SLT	NOIZEUS
[102], 2016	Time	LSB substitution	N/A	The method uses eight threshold intervals to allocate a variable number of bits (1–8) per sample; samples with lower threshold assigned with max payload 8 LSB, and vice versa	Low robustness	WAV	N/A	N/A	CSF

(*continued on next page*)

in methods of this behavior is the message type. If the message is a text type, then the matching algorithm must find the exact binary sequence, which makes this message type most sensitive for errors. By contrast, audio and image messages are more robust to errors due to high data redundancy. Ahmed et al. proposed a method for audio in audio hiding based on fractal coding, where the frame size, matching parameters encoding, and the error ratio are discussed in detail [133].

Table 3 (*continued*).

Ref., Year	DOE	MODI	Supportive techniques	Acoustic embedding condition*	Disadvantage	Carrier type	Message type	Evaluation features	
								Metrics	Data
[103], 2017	PWT	Coefficient modification, *k	N/A	The method calculates the energy of each subband and then selects which subbands are used for embedding based on the assigned energy threshold	High distortion	WAV	N/A	BER, SNR, PESQ	N/A
[104], 2017	FFT	Spectrum addition	N/A	Average energy; a region is selected, such that the amplitude–frequency is lower than the audibility level	High distortion	WAV	Speech	Cross-correlation	N/A
[105], 2017	Time	STC, *l	N/A	AAC compression difference is used to determine perceptually relevant and irrelevant samples	Low robustness	WAV	N/A	SNR	CSFMG

*a, If the coefficient value is extremely high or low, then a large number of bits are allocated for embedding. Conversely, a small number of bits are allocated in coefficients of midrange.
*b, In adaptive embedding, additional LSBs are allocated for data units with high values, and none-to-low number of LSBs are allocated for low-valued data units.
*c, The first and second methods are conducted using spectrum shifting and inserting and spectrum addition, respectively.
*d, In this method, all eight subbands are good candidates for embedding, and the method modifies a number of the subbands at each file. However, such an embedding method might lose orthogonality. Thus, writing in multiple concurrent subbands can cause the secret data to be erased and rewritten in the IDWT process. The BER test over message retrieval is vital in evaluating the accuracy of message recovery.
*e, The method sorts out the secret signal coefficients in such a manner that the maximum secret coefficient is paired with the maximum cover coefficient while keeping a record of the original order of the secret coefficients. After sorting, each signal consists of three main regions, namely, positive, null, and negative. Then, the positive coefficients of the secret signal will be located in the positions of the highest coefficients of the cover signal, whereas the negative coefficients of the secret signal will be located in the positions of the lowest coefficients of the cover signal. The cover coefficient is divided by the corresponding secret coefficient and multiplied by 100. This value will be embedded in the LSB. Finally, the method utilizes five LSBs to embed each coefficient of the secret signal. In the extraction process, the method recovers the hidden percentage from the last five LSBs. The secret coefficient is extracted by multiplying the recovered percentage by the total coefficient value and divided by 100.
*f, The embedding takes a variable of α=2.5 as an embedding key. It is performed by taking the modulus function over the absolute value of the magnitude and 5. If the bit to be embedded is 1, then abs (z) mod 5 is set above 2.5. In the case of 0 embeddings, then the abs (z) mod 5 is set below 2.5. In both cases, the maximum addition or subtraction is 2.5.
*g, The method to improve robustness proposes bit repetition embedding and binary voting at the receiver side.
*h, The method dynamically calculates multiple thresholds.
*i, Embedding is performed by replacing the Huffman code sign bit with the message bit.
*j, Embedding is performed by replacing the last m coefficient in case the frame is voiced or the $2m$ coefficients in case the frame is unvoiced. If the bit is 1, then the value is replaced with a constant. Otherwise, 0 is embedded as 0.
*k, In each frame, the four selected subbands are used to embed four secret bits. By manipulating the values of the four coefficients by addition or subtraction, 16 patterns of relationships are produced. An example of relationship pattern is a<b<c<d. Each pattern represents a distinct 4-bit stream, which is used to carry the message bits.
*l, The STC will be used to embed data in the perceptually irrelevant samples.

Table 4
Error minimization-based embedding.

Ref., Year	DOE	MODI	Supportive techniques	EM Algorithm*	Disadvantage	Carrier type	Message type	Evaluation features	
								Metrics	Data
[107], 2010	VOIP	MES	Variable matrix size	MES	Low robustness	VOIP	N/A	Embedding rate, embedding efficiency, average distortion,	N/A
[108], 2011	Time	LSB substitution	Deep layers	GA	Low robustness	WAV	N/A	PSNR	N/A
[109], 2011	Time	LSB substitution	Custom encryption method	*a	Low robustness	N/A	N/A	No experiments	
[110], 2011	Time	LSB substitution	Pixel value differencing, amplitude differencing	Error diffusion, *b	High distortion, low robustness	WAV	N/A	SNR, BER	N/A
[111], 2012	Time	LSB substitution	N/A	Error diffusion, *b	High distortion	WAV	Text	BER	CSFMG
[112], 2013	Time	LSB substitution	Random bit selection	*c	Low capacity	WAV	NA	No experiments	
[113], 2013	Time	LSB substitution	RSA for key transformation and encryption	GA	Low robustness	WAV	Text	SNR, SLT	CSFMG
[114], 2013	Time	LSB substitution	N/A	*d	Low robustness	WAV	N/A	No experiments	
[115], 2014	Time	LSB substitution	Prime sample number only	*e	Low robustness	WAV	Text	HC, SNR	N/A
[116], 2015	Time	LSB substitution	Lossless compression, prime sample number only	*e	Low robustness	WAV	Text	SNR	N/A
[117], 2014	DCT	Coefficient modification, *f	Reversible, non-blind	Selection by calculation	High distortion, non-blind	WAV	N/A	SegSNR, MOS, PESQ	N/A
[118], 2015	Time	LSB substitution	Deep layers	GA	Low robustness	WAV	N/A	PSNR, SLT	CSF
[119], 2016	Time	LSB substitution	Third and fourth LSBs	GA	Low robustness	WAV	N/A	No experiments	
[120], 2017	VOIP	Codeword modification, *g	N/A	MES	Low robustness	VOIP	N/A	MOS, PESQ, PSNR, HC, embedding efficiency	N/A
[121], 2017	In encoder	Coefficient modification, *h	Distortion function	STC	Low capacity	MP3	N/A	PEAQ, ODG	CSFMG
[122], 2017	Time	Codeword modification	Optimal probability of pulse function and pulse correlation function	STC	Low capacity	AMR	N/A	Test error rate, PESQ	CDB, SDB1, SDB2, SDB3

*a, The method is based on embedding in deep layers with shuffling and then attempts to minimize the errors by remodifying the other LSBs to produce a sample value that is similar to the original.
*b, In the error diffusion, the next four samples are modified with variable portions (percentage) of the error.
*c, The secret bit is embedded in a fixed position. Thereafter, all the other bits are remodified to produce a total byte value that is nearest to the original. The error value in this method is controlled by the deviation value. If the difference between the replaced and original bytes is larger than the deviation limit, then the method discards the new byte and uses the low bit to encode the message.
*d, The method embeds in the 3rd and 4th LSBs. After embedding, only the 2nd and 5th LSBs of each modified sample are manipulated if needed to minimize the resulted error.
*e, The method embeds the secret data in two LSBs (1st and 4th) of the cover samples before modifying the other LSBs to reduce the error.
*f, The method selects the blocks that scored the minimum error estimation for embedding. A table is created of zeros and ones, which represent the unmodified and modified coefficients, respectively. Embedding is directed by the table and performed by multiplying the coefficient value by 2, and the secret bit is then added. The table must be shared to enable message recovery.
*g, The embedding in this method is performed in the vector quantization (VQ) process of the LPC by modifying the codeword based on MES rules.
*h, After quantization, the code stream is converted to binary. Distortion is then calculated, and the message is embedded in the binary form of code using the STC. In the STC, the distortion function is based on finding an embedding path with minimum distortion.

Table 5
Pattern-matching-based methods.

Ref., Year	DOE	MODI	Supportive Methods	Matching Space/Process*	Signaling Process*	Disadvantage	Carrier type	Message type	Evaluation features	
									Metrics	Data
[123], 2011	PWT	LSB substitution	Cover segmentation, PWT	One subband of the PWT is selected for matching. The result of matching is the key that indicates the number of message bits matched in each coefficient	The key is encrypted and embedded in the shared locations	High time and space complexity, low robustness.	WAV	Audio, image, text; the experiments are conducted on images	SNR	CSFMG
[124], 2011	MDCT	Coefficient modification	N/A	Probabilistic global search Lausanne algorithm is used to find the best embedding location for each frame with the least error introduced	The message is embedded in the selected locations	High time and space complexity, long shared key.	WAV	N/A	BER, PSNR	TIMIT
[125], 2012	Time	LSB substitution	Amplitude thresholding	The first three MSBs are considered for matching to find the number of matched bits with the message sequence	The last three LSBs carry the result of matching	Low robustness, lack of encryption.	WAV	Text	SNR, MSE	CSF
[126], 2014	LWT	Coefficient replacement	Message encryption by XOR	After performing the LWT for three levels, the first and second detail vectors are merged to form the matching space	The third level detail vector is used to signal the matching results	Low robustness, high space, and time complexity	WAV	WAV, image, text	SNR, NC, HC	CSFMG
[127], 2014	Time	LSB substitution	N/A	A secret shared key defines the locations of bits of comparison in the cover file. If a match is found, then 1 is inserted at the LSB of the defined location; otherwise, 0 is inserted	LSB	Low capacity	WAV	Text	SNR	CSF
[128], 2015	Time	LSB substitution	N/A	The method searches for a matching sequence for the message segment using a sliding window	The LSB of the audio segment is marked when a match is found	High space and time complexity, low robustness	WAV	Text	Embedding error, SNR	N/A
[129], 2015	VOIP	Codeword modification	MES	The method compares the binary similarity among the original message, the encrypted message, and the cover segment	If the similarity level is above a certain threshold, then MES is used instead of bit replacement	High space and time complexity	VOIP	Text	ER, BCR, EBCR, MOS-LQO, PESQ	N/A
[130], 2016	Time	LSB substitution	N/A	The method selects a fixed pattern of the cover and attempts to match the message in the pattern	The index of the found match is embedded at the LSB	Low robustness, low capacity	WAV	Text	MSE, SNR	CSF
[131], 2016	IWT	Coefficient modification	N/A	The matching space is the approximate coefficient vector; the matching process produces a key that includes the indices of the selected coefficients	The key is embedded at the 3rd and 4th detail coefficients	Low message accuracy, high cover distortion	WAV	WAV	SNR, SPCC	"WAV file"-free database
[132], 2017	Time	LSB substitution	Fractal coding	The algorithm attempts to find a match for a variable number of samples (6 or 8) at a time	*a	Low message accuracy, low robustness.	WAV	WAV	SNR, HC, NC	"Marsyas"-free database website

*a, When the best match is found, the LSB of the cover is modified with a number of values that represent the location of the found segment, the similarity level, and other metadata.

4.6. Phase coding

Phase coding methods are based on manipulating or modifying the phase values of the frequency components. In the frequency domain (FFT), the coefficients are separated into two vectors, namely, magnitude and phase. Phase coding utilizes the low sensitivity of HAS to change in phase values [5]. Fig. 11 shows an overview of this embedding behavior alongside an example. The main advantage of phase coding methods is their high robustness against signal manipulation attacks [6,13,134]. The perceptual transparency of a phase coding method is based on the severity level of change and the embedding rate. This section features a group of phase coding methods. However, given the similarity level among the methods, the frame length, in samples or milliseconds, will be selected as a special feature to highlight the inner differences in this group with respect to the importance of the MODI. This pattern of embedding is observed in the methods summarized in Table 6.

Particularly, the methods in this group maintain the highest level of balance among the three main requirements. However, the embedding capacity is the least prioritized requirement. Moreover, the main advantage of methods of such behavior is the high robustness against two main types of attack, namely, signal manipulation and message retrieval. The robustness against detection attacks or steganalysis methods was discussed in [134]. However, the detection of the phase coding is based on the severity of modification, which was discussed and tested in-depth in [142].

4.7. SS

SS was originally proposed as a transmission method to increase robustness and ensure message delivery. For the same reasons, SS was adopted in audio steganography. In SS, each message bit is spread (repeated) n times based on the chip rate, which is the number of embedded bits for each message bit. The number of copies is determined by the chip rate or the number of bits embedded for each message bit. The process begins by changing the message from a binary to a $(1, -1)$ format. Thereafter, the modulation process is performed by multiplying each bit of the message by a pseudorandom sequence of $(1, -1)$ to spread each bit n times (n copies), where n is the length of the chip rate. Whenever the number of copies for each message bit increases, the robustness will increase (up to a limit), and the embedding capacity will be decreased by the same rate. Fig. 12 shows an overview of this embedding behavior. In most SS methods, the MODI is based on changing the sign of the data unit to signal the bit. As a consequence of such behavior, high distortion is introduced. This pattern of embedding is observed in the methods summarized in Table 7.

4.8. Tone insertion

The idea of the methods of this group is observed and recorded. The concept is based on adding foreign tones or manipulating original tones. This category differs from the category presented in a previous study [5]. Other methods in the frequency domain are named tone insertion metaphorically. However, methods under this group are based on physical tone insertion using audio mixing devices, which result in a high difference between the cover and stego signal and can result in serious quality degradation. Moreover, these methods operate only over music audios because they modify certain musical components.

In [147], a supraliminal channel method was proposed, such that the secret message was encoded in the semantic content of a cover object. In this method, the secret message was embedded in the audio cover as additional sounds, such as drum beats, hi-hats, or simple noises. Such sounds will not be perceived by the warden as a message but as normal parts of the audio recording instead.

In [148], a method called "StegIbiza" was proposed for hiding secret messages in streaming music service using tempo as a cover. The idea in this method is based on manipulating the beats per minute (BPM) in an audio music file. The method adapts the Morse code as a means to pre-encode the secret characters. The results of the Morse code encoding are a sequence of signs $(+, -)$. The method then adjusts the tempo scale by increasing or decreasing to embed the Morse codes. The receiver can realize the delicate shifts in the BPM and finally reconstruct the message in Morse code and finally to the ASCII characters by analyzing the same tempo.

4.9. Others

The methods featured in this section do not follow any of the aforementioned highlighted embedding behaviors. Moreover, the methods in this section follow a unique MODI. Therefore, the special feature, in this section, is the MODI because it represents the MPI in the method. Moreover, the supportive method field is removed for inactivity. These methods are summarized in Table 8.

5. Additional findings

In this section, two statistics of audio steganography methods are presented regarding the DOE and carrier type. Moreover, this section includes a review of the two main elements in the evaluation component in audio steganography, namely, dataset and metrics used. Sections 5.1 and 5.2 consider the DOE and the carrier type, respectively. Section 5.3 briefly reviews the datasets and the nature of audio files used in the evaluation part. Finally, Section 5.4 presents the most common objective and subjective metrics.

5.1. DOE

In this section, the reviewed methods are analyzed based on the DOE. In audio files, three main domains are highlighted, namely, time, Wavelet, and frequency. In addition to the three domains, the embedding process is based on two domains in some methods. Moreover, the audio stream in the VOIP is considered a separate domain. Fig. 13 shows the percentage of occurrences of each domain of the reviewed methods.

As shown in Fig. 13, the most common embedding domain is the time domain, which can be explained by its simplicity and customizability. Moreover, the Wavelet and frequency domains share a nearly similar number of methods, whereas a low number of methods are proposed in the VOIP and hybrid domains.

5.2. Carrier type

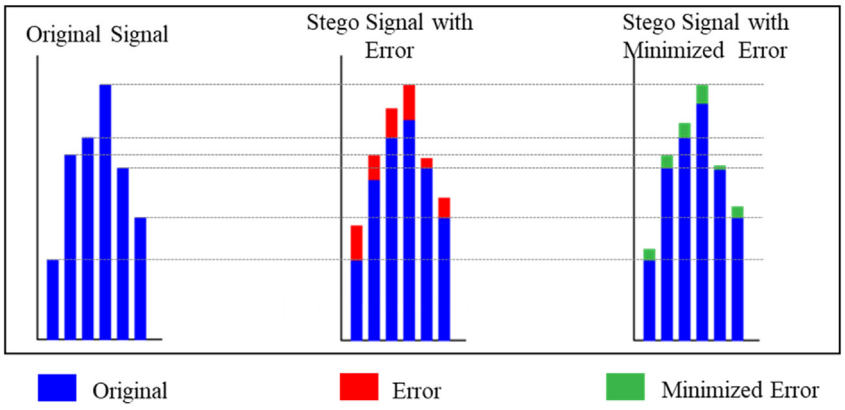
This subsection focuses on the most common types of audio cover files that are used in the literature. The analysis of the proposed methods shows that the most adopted audio types are WAV, audio codec of VOIP (audio stream), MP3, AMR, AU, and MIDI, as shown in Fig. 14.

The most common carrier type is WAV, which achieves 84% of the total methods, whereas the ratios for the VOIP and MP3 formats are also found. Moreover, few methods are used on AMR, AU, and MIDI files.

Theory: Minimize the error resulting from embedding.

Implementation: depending on the algorithm, coding theory based methods such as Matrix Embedding Strategy (MES) and Syndrome-Trilles Codes (STC) perform the minimizing processing before the embedding. While in Genetic Algorithm, the minimizing processing is carried out after embedding. MES and STC are more complicated algorithms and tend to achieve higher efficiency. For example, the MES is capable of hiding multiple message bits (2, 3, 4 and more) with one bit error.

Overview:



Example: The method in [118] is selected to highlight the error minimizing embedding behavior. The method uses a genetic algorithm to re-evaluate the modified sample to minimize the error. The example below illustrates the main concept of embedding using error minimizing method in [118]. The method initially embeds by LSB substitution, then the GA is carried out to optimize the embedding.

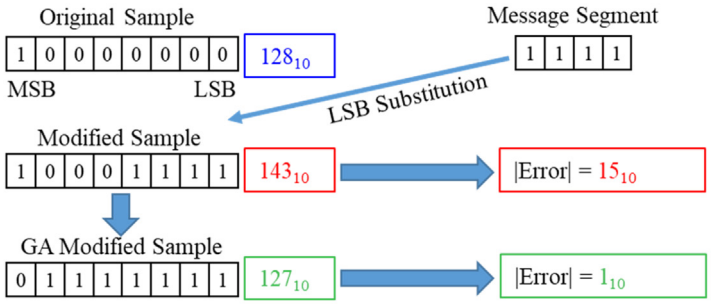


Fig. 9. Error minimizing embedding.

5.3. Dataset and selection of audio files

This subsection highlights how and from where the reviewed articles select the audio files for their experiments. Few of the articles have adopted well-defined and specific datasets for selecting the cover and secret audio files. By contrast, many of the articles have adopted generic or commercial websites or customized selected audio files, as shown in Fig. 15.

Most of the methods are evaluated based on custom selected audio files, which can imply a limitation with existing datasets (Fig. 15). This limitation can vary from being rare to find or inappropriate to be used for embedding evaluation for their content, lengths, and size of the dataset. Only 8% of the reviewed methods used a dataset for evaluation, whereas only 1% used audio tracks from websites for evaluation. Fig. 16 shows the datasets that were used in the reviewed articles.

Three types of sources are found in this paper, namely, dataset, website, and custom selected audio files. The common datasets are the TIMIT [157], NOIZEUS [101], GTZAN [158], TPS [159], and CORPORA [157]. The common websites are https://freesound.org, https://www.soundsnap.com/, http://www.WAVsurfer.com, and http://www.WAVsurfer.com/. However, the majority of the reviewed articles adopt custom selected audio files with specific attributes, such as bit per sample, channel, sample rate, and duration. Fig. 16 shows that TIMIT is the most common dataset.

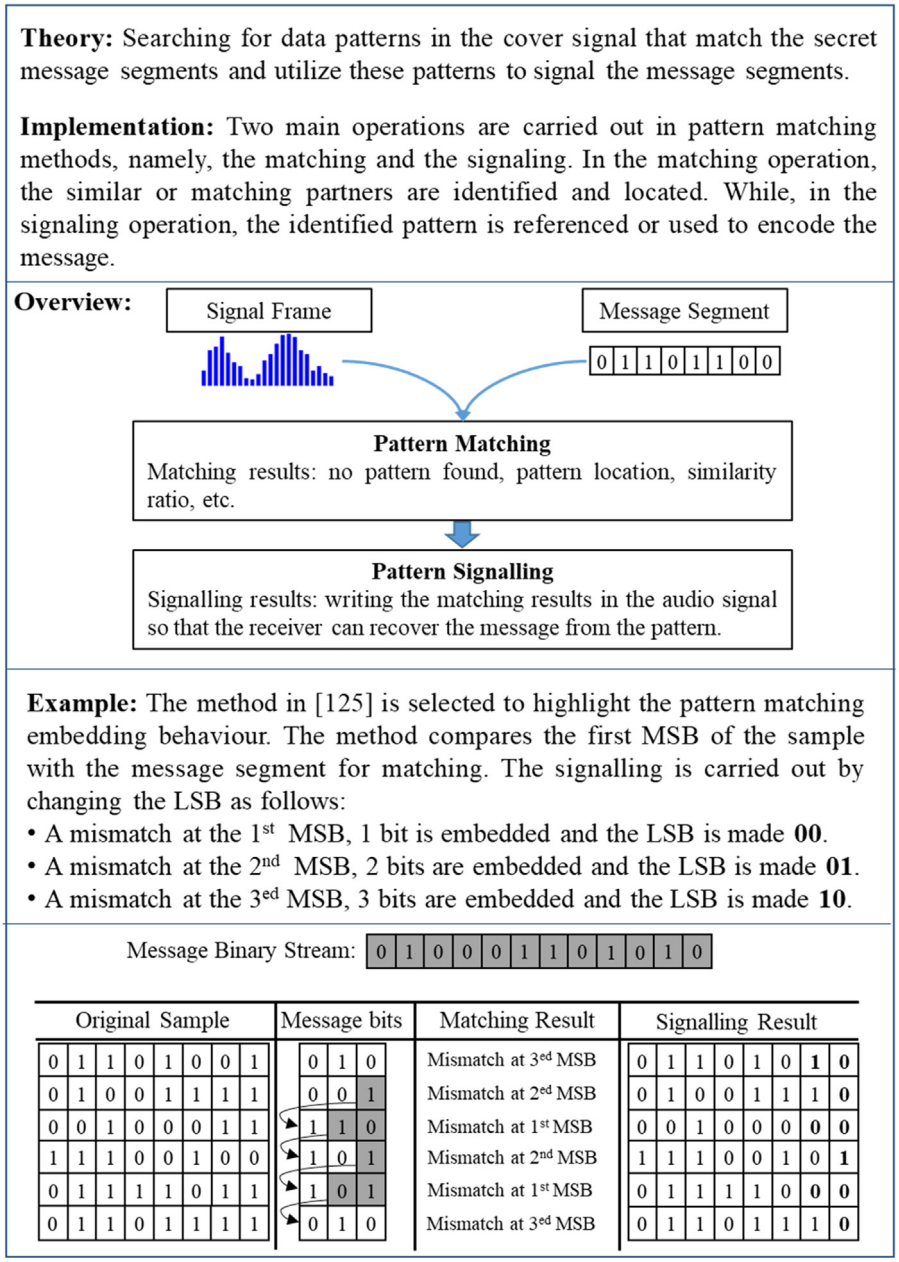


Fig. 10. Pattern matching embedding.

5.4. Evaluation metrics

This section highlights the most common metrics or measurements that are used to evaluate the performance of the proposed methods of the reviewed articles. These metrics are used to evaluate the main requirements of the data hiding in terms of transparency or imperceptibility, hiding capacity or payload, and robustness against different types of attack. Each metric is represented by the number of its occurrences in Fig. 17.

As shown in Fig. 17, the most common evaluation metric thus far is the SNR, which was used in almost 80 of the total reviewed methods. Moreover, other evaluation metrics, such as PSNR, BER, MSE, MOS, and PESQ share good popularity among the methods. Furthermore, the remaining metrics, including PEAQ–ODG, hiding capacity, NC, SPCC, MOS-LQO, SegSNR, fourth first moments, SDG, HER, MBSD, and PRD, are found with variable popularity.

6. Discussion

To understand the reviewed methods in this paper and the reasons behind such a level of diversity, discussing the three main requirements of audio steganography and existing trade-offs is vital. The performance improvement in each audio steganography method remains connected to approximately one of the three main requirements, namely, capacity, perceptual transparency, and robustness.

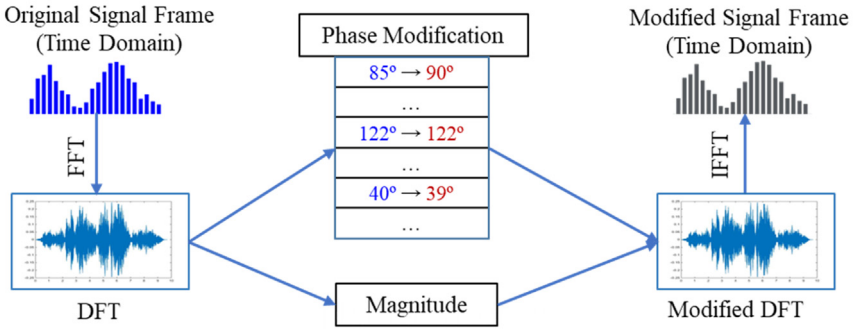
Capacity is related to a message size that can be embedded in an audio second or represented as the percentage of a message size to an audio second. Embedding capacity can be increased by many techniques, one of which is by simply increasing the NBPDU. Compression is another common backdoor technique.

Perceptual transparency is related to audio quality deterioration or represented by the difference between the original and stego audio. Notably, transparency is also connected to the secrecy level and capability to avoid suspicions and detection

Theory: modify the phase components of the Discrete Fourier Transform (DFT) to achieve transparent and robust embedding.

Implementation: after dividing the signal into frames, the first step is to convert the signal to the frequency domain, usually using the Fast Fourier Transform. Then the magnitude and the phase components are separated to access the phase bins. Based on the method in hand, the phase components are modified. Then the modified phase are rejoined with the original magnitude to achieve the modified DFT. Finally, the frame is converted back to time domain.

Overview:



Example: The method in [136] is selected to highlight the phase coding methods. The method uses differential phase where the difference between two adjacent phases hold the embedded bits. The method calculates the difference between two adjacent phases; if the bit to encode is 0, then the difference is made even; otherwise it is made odd.

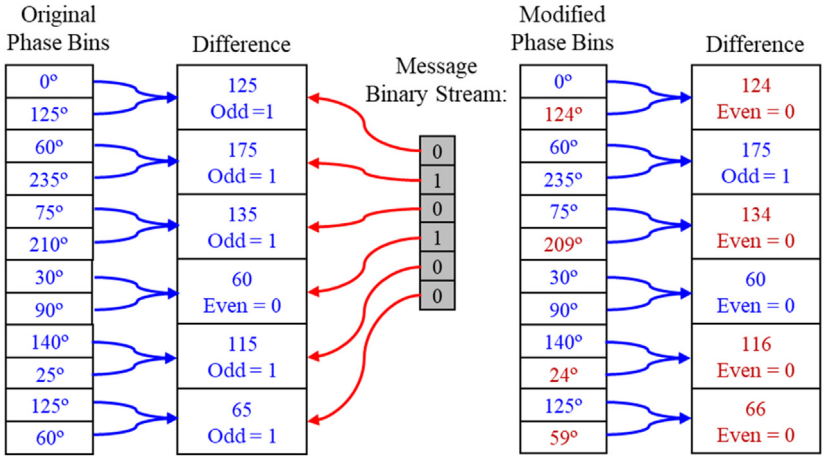


Fig. 11. Phase coding methods.

attacks [3]. The first trade-off exists between embedding capacity and perceptual transparency. Reducing the message size improves the quality of the stego file, whereas increasing embedding capacity causes a clear and considerable distortion.

Robustness is defined as the capability of steganography methods to resist intentional and unintentional attacks. Possible attacks are typically signal processing attacks that manipulate and deviate the signal to destroy the embedded secret messages. Among the signal processing attacks are compression, scaling, noise addition, resampling, resizing, cropping, amplification, and filtration. In the relationship between robustness and capacity, high capacity by nature increases the success rates of certain types of attack. However, as noted in other reviewed methods (Section 4.7), message bits are repeated several times to increase the chance of message survival. Thus, increasing the message survivability requires increasing the number of cover bits that are modified for each message bit. This contradiction is also reflected in the relationship with perceptual transparency. High levels of noise and distortion in the stego indicate an increment of detection chances. Simultaneously, low levels of distortion and high-quality stego file reduce the detection chances. However, the change that is caused by the embedding method must be extremely strong to survive the ferocity of message destruction attacks. Generally, such a strong change comes with the cost of high signal distortion. Rendering the secret message to be entirely undetectable can indicate its fragility against message destruction

Theory: Spreading the message bit *n* times over a frame of audio signal to increase the robustness. Spread Spectrum include two main methods, namely, Direct Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS).

Implementation: this method can be carried out in all embedding domains, and it include two unique parameters, namely, the chip rate which is number of copies for each message bit and the Pseudo Number (PN) sequence which is a randomly generated sequence to modulate the message bits.

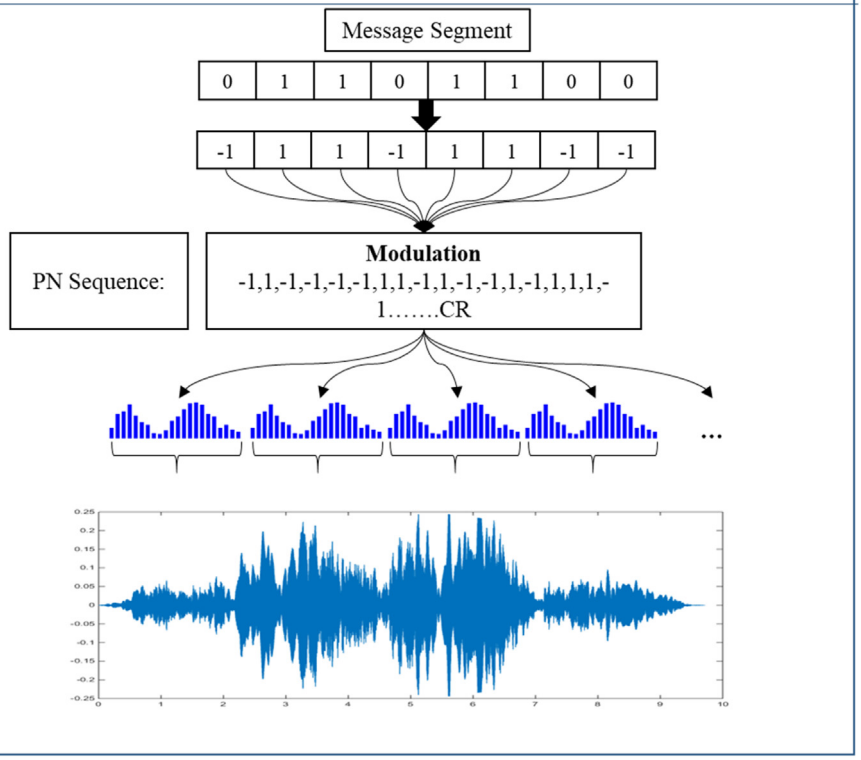


Fig. 12. SS.

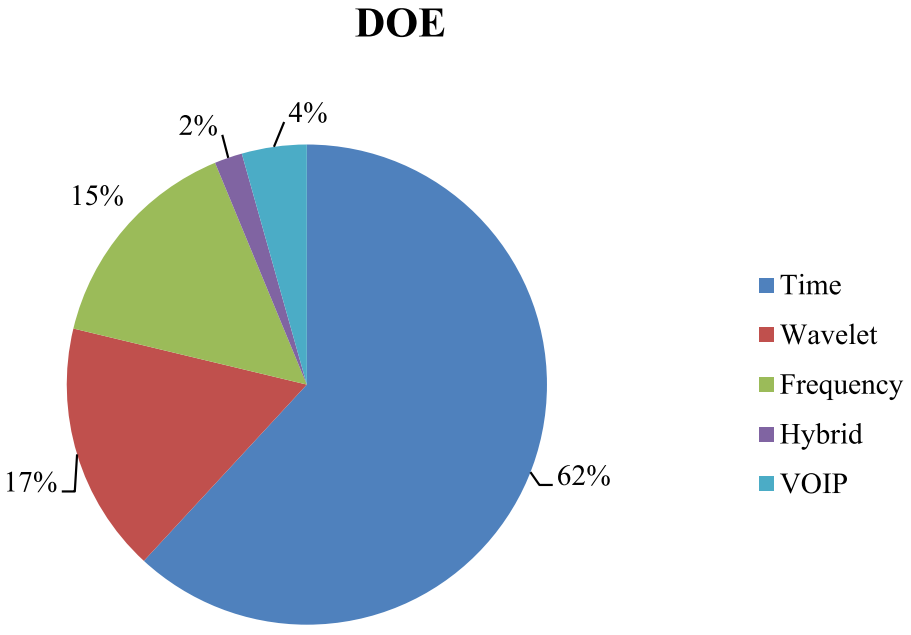


Fig. 13. DOE distribution.

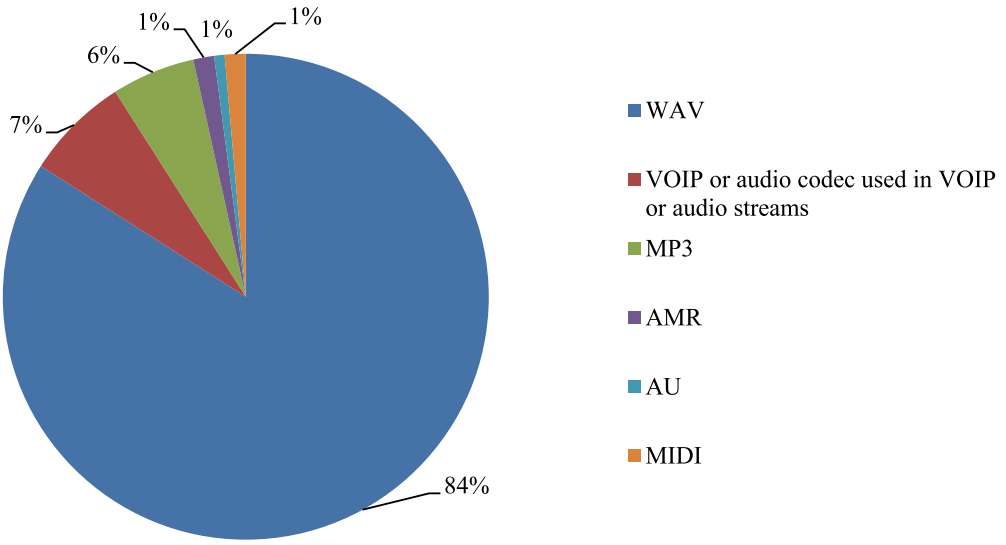


Fig. 14. Carrier type distribution.

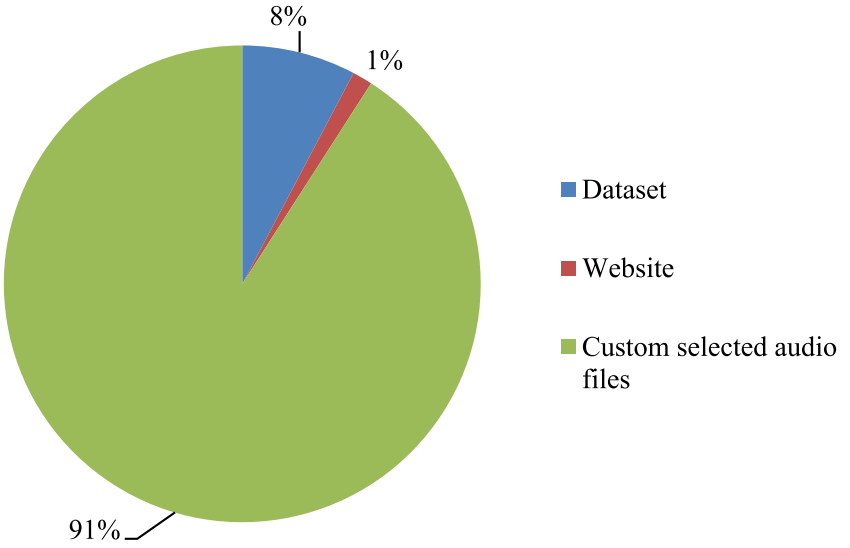


Fig. 15. Carrier type audio data.

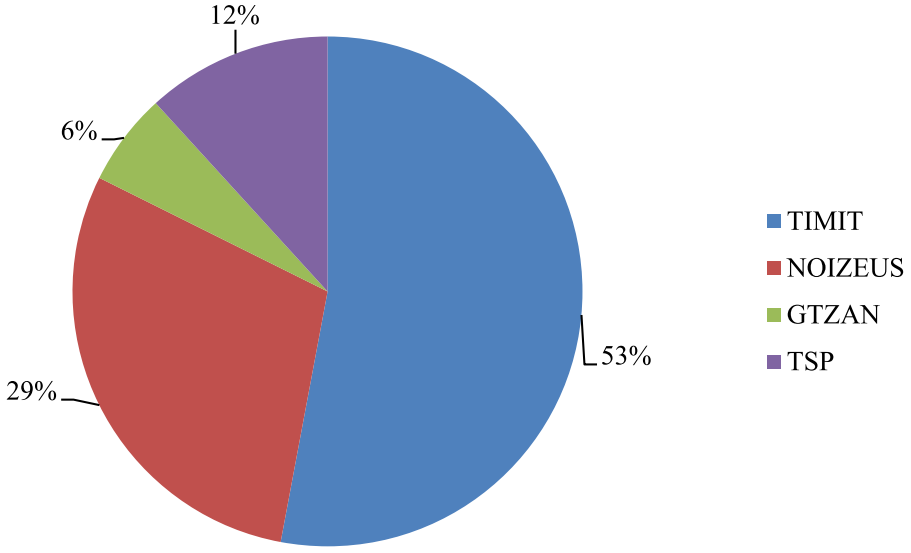


Fig. 16. Most common datasets.

Table 6
Phase coding methods.

Ref., Year	DOE	MODI	Supportive methods	Frame Length*	Disadvantage	Carrier type	Message type	Evaluation features	
								Metrics	Data
[135], 2007	FFT	Phase interpolation; the old phase value is replaced by the average of the nearest two phases (before and after) plus or minus a small value to mark the 0 and 1 cases	N/A	1024, 512, 256	Low capacity	WAV	N/A	RMSE	N/A
[136], 2011	FFT	Differential phase modification; the method calculates the difference between two adjacent phases; if the bit to encode is 0, then the difference is made even; otherwise, it is made odd	Integer rounding	10 ms	N/A	WAV	N/A	Correlation	N/A
[137], 2012	FFT	Double random-phase encoding technique; the method uses the coordinates of the 2D sound map as variables for encoding	Arnold transform for scrambling	256	High distortion	WAV	Image	SNR	N/A
[138], 2014	FFT	In the selected frequency bins, the binary representation of the phase value is modified with secret information	Phase bins selected by magnitude threshold	64 ms or 4 ms	High distortion	WAV	N/A	SNR, PSEQ	CSFMG
[139], 2014	FFT	$\pi/2$ is added to the old phase to signal the 0 cases, and $\pi/2$ is subtracted from the old phase to mark the 1 cases	*a	Variable	High distortion	AVI	N/A	PSNR	N/A
[140], 2015	FFT	The phase part is modified at a randomly selected bit position	SHA-1, PNRG	Two samples	High distortion	WAV	Text, image, audio	MSE, SNR, PSNR, MOS	N/A
[141], 2016	FFT, Time	The message is hidden in two methods, that is, LSB in the time domain and phase coding; the odd byte numbers are hidden using LSB, whereas even byte numbers are hidden using phase coding	N/A	N/A	Low robustness	N/A	Image	N/A	N/A

*a, The method takes a video file of type (.avi) as an input. After separating the images and the audio file, LSB substitution is performed to embed data in the image, whereas phase coding is conducted to embed data in the audio part.

Table 7
SS methods.

Ref., Year	DOE	MODI	Supportive Methods	Chip rate*	Disadvantage	Carrier type	Message type	Evaluation features	
								Metrics	Data
[143], 2011	FFT	Magnitude sign modification	N/A	10000	High distortion, extremely low capacity	WAV	8-byte text message	N/A	N/A
[144], 2012	DCT	Coefficient modification, *a	N/A	1024	High distortion	WAV	N/A	BER	CSFMG
[145], 2013	Time	Sample sign modification	N/A	50, 100, 120	High distortion	WAV	Text	SNR, SWR, BER	N/A
[146], 2015	Time, FFT	LSB substitution, parity coding, magnitude modification	N/A	N/A	High distortion	WAV	Text	BER, HC	N/A

*a, The DCT coefficients are divided in half if the message bit is 1. The first-half coefficient is divided by a constant; otherwise, it is a 0 case. The second half is divided by the same constant. High distortion and higher detection rates are the definite results of such embedding severity. However, the algorithm embeds watermark data in the audio signal, streams it through an air channel, records the signal, and reconstructs the watermark from the recorded signal. Therefore, the watermark must be harshly encoded (robust) in the audio to sustain the high error rate as a result of the noise addition.

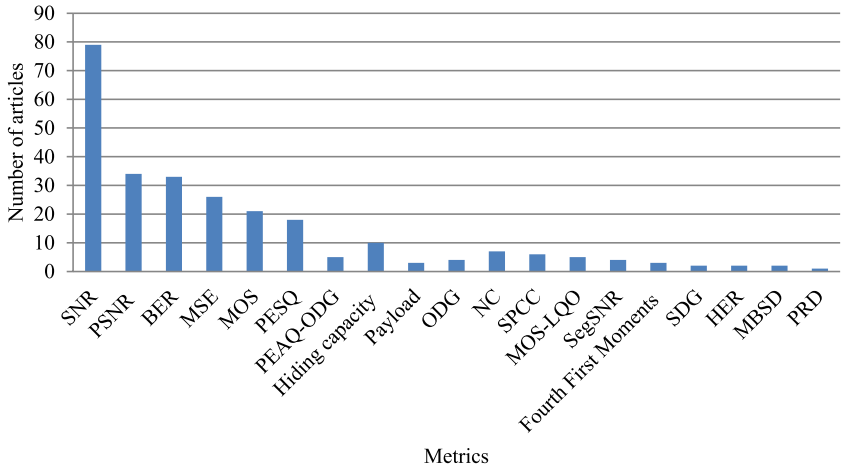


Fig. 17. Evaluation metric distribution.

Table 8
Other methods of audio steganography.

Ref., Year	DOE	MODI*	Disadvantage	Carrier type	Message type	Evaluation features	
						Metrics	Data
[149], 2010	VOIP RTP	The voice payload is modified with a secret message before transmission; certain packets with specified time and order are intentionally delayed to signal a piece of information to the receiver	N/A	VOIP	N/A	MOS	N/A
[150], 2011	MP3 in encoder	The method divides the Huffman tables into three groups of tables, namely, G-1, G0, and G1; if the selected table is from G1 and the secret bit is 0, then the method swaps the selected table to a table from the G0 group	Low capacity	MP3	N/A	HC, SNR, ODG	CSFMG
[151], 2012	MP3 in encoder	Embedding is performed by using the window switching mechanism in the psychoacoustic model; the method allocates one window status to embed the 1 case, whereas three window statuses are allocated to embed the 0 case	Low capacity	MP3	N/A	SNR, ER, variance, ODG	CSFMG
[152], 2012	DWT	The method calculates the detail coefficients of the DWT in the third level; the details are then transformed into an image. Then, the secret data are embedded in the obtained image based on VQ. Embedding is performed by varying the compression codes, namely, the search order coding for the 0 cases, and the original index values for the 1 cases	N/A	WAV	N/A	SNR, HC, bit error	CSFMG
[153], 2014	VOIP	The method begins by determining the originally used codec, which is achieved by reading the payload-type field in the RTP packet. Then, the method selects an appropriate codec that sustains a relatively high voice quality w.r.t to the original signal while occupying less payload space. The saved space will be used to embed the secret data	Low robustness	VOIP	N/A	HC, MOS, DSNR	TIMIT
[154], 2016	FFT (echo hiding)	The method aims to solve the robustness and transparency issues of echo hiding by introducing two pseudorandom sequences instead of fixed parameters; one to generate segments with different lengths, and the second sequence to generate a different echo bank for each segment using a different value of delays	N/A	WAV	N/A	BER	CSFMG
[155], 2016	VOIP	Audio steganography is achieved systematically, which is referred to as a microprotocol. Headers that represent the type of secret information are used to formalize the secret cover channel. In this procedure, the types of messages, such as REQ, RES, and DAT, are used. Embedding uses LSB versions	N/A	VOIP	Text	MSE, SNR, PSNR, MOS-LQO	TIMIT
[156], 2016	Time	The method is based on using the audio file as a reference to send an encrypted message without embedding any bit in the carrier. The result will be a sequence of locations where the message bits can be reassembled. Subsequently, this sequence is encrypted using a chaotic algorithm and then sent	N/A	WAV	Image, text	N/A	CSF

attacks. In addition, making the secret message difficult to destroy might reveal its cover and result in easy detection.

7. On audio steganalysis

As a countermeasure of steganography, steganalysis in audio is defined as the art and science of detecting hidden information patterns in audio covers [160]. The main challenge in audio steganalysis is to define discriminative features that indicate the hidden information existence. Audio steganalysis methods are classified as targeted and universal; targeted methods assume preliminary knowledge about embedding methods, such as steganalysis against phase coding [161] and echo coding [162], whereas universal methods aim to detect the message existence regardless of the embedding method [163]. Furthermore, universal steganalysis is divided into calibrated and noncalibrated methods. In the noncalibrated methods, only the cover signals are used to extract the crucial features, whereas in calibrated methods, the cover signal is compared with the estimated version(s) of the signal, such as a reference version, a re-embedded version, or a denoised version, to accomplish discrepancies as features. Another type of steganalysis is those specific to certain compressed audio formats, such as ACC [164] and MP3 [165]. In this section, selected methods are highlighted to shed light on the researchers' efforts on the countermeasure of audio steganography.

Ozer et al. proposed a detection method based on the characteristics of the denoised residuals of the audio file [166]. The goal was to uncover discrepancies among the cover signal and its denoised estimate. The method uses a large pool of audio quality metrics as statistical features, wherein each detection attempt only selected features that were used based on their effect on the classification process. One of the most well-known noncalibrated methods is using the MFCC and the Markov process, which claims the property that the previous sample does not influence the

future one if the current sample is known, which is referred to as Markov chain [167]. Better edge detection was achieved by calculating the second-order derivatives of the signal before calculating the MFCC [160].

Recently, Ghasemzadeh and Arjmandi proposed a method that addressed universal and targeted cases [168]. They argued that the first LSB is the most sensitive bit plan in audio steganography and can thus be generalized as a universal steganography method. The proposed method was based on the re-embedding calibration, where the cover signal was used to re-embed a random message in the LSB plane. Then, the reversed Mel-frequency scale aimed to uncover variation in the high-frequency regions. The energy of the R-MFCC was calculated for both versions and compared using statistical moments. Their results showed that their feature achieved better discriminative accuracy. Han et al. proposed a method based on linear predication, which is used originally for signal coding [169]. In this method, four main features are extracted, namely, linear prediction coefficients, linear prediction residual, linear prediction spectrum, and linear prediction cepstrum coefficients. The method was capable of highlighting significant differences between the cover and stego signals at an embedding rate of 6 KB per second in WAV file of CD quality (44,100 samples and 16-bit resolution). The method used SVM and the Gaussian radial basis function as a classifier. The study reported better detection accuracy than MFCC-based methods. In [170], a high pass filter was used to calculate the residual maps in cover and stego signals. Then, the resulted maps were compared using convolutional neural networks. After the training process, the method achieved high accuracy rates.

Notably, the latest or recent review on audio steganography was published by Ghasemzadeh and Kayvanrad in 2018 [163]. They highlighted the advancement that was achieved in these methods. However, comparing the performance of these methods is difficult because they used different datasets and training schemes. Moreover, high accuracy ratios are associated with significant and large payloads [169]. Recent studies have shown promising results, especially with the improvement of artificial classification methods, such as SVMs and neural networks. However, the detection of universal methods for highly transparent smaller payloads still remains a challenge.

8. Limitations and future works

This section highlights the limitations of the current work. The main aspect that can be covered further in this review is a comparison of the performance of the reviewed methods. This comparison can be conducted due to the inconsistency in evaluation methods and embedding environments in the reviewed methods. Such inconsistency is caused by many factors. Among these factors is diversity in cover file types, message sizes, and evaluation metrics. In other words, the lack of a unified dataset and a clear benchmark cause inconsistency in comparison. Such inconsistency may render any performance-based comparison inaccurate. As a solution, an effort has been made to review each method in a manner that highlights its performance based on the means of operation. Future work will focus on designing a multipurpose benchmark that covers the three main requirements of audio steganography. In addition, a unified evaluation environment that includes controlled parameters can be designed to enable a fair comparison.

9. Conclusions

Audio steganography is the process of hiding a message inside an audio cover file. In this paper, a review has been conducted on the methods of audio steganography. The methods have been collected in a systematic manner on five major digital databases using a unified query. The proposed review aims to shed light on the ideas and methods proposed in audio steganography. Existing reviews suffer from the high overlapping level or low level of segregation among the methods due to poor classification criteria. Furthermore, existing reviews have failed to cover the number of proposed methods and depth of analysis. After analyzing the reviewed methods, patterns of similarity are formed. These patterns are then observed and formulated to be used as classification criteria. This review proposes a new classification of audio steganography based on the MPI in the embedding methods. The method is found to be more accurate and detailed compared with the existing classification. The audio steganography methods are classified into selective-based embedding, linear or sequential embedding, frequency masking, amplitude thresholding methods, error minimization-based embedding, pattern-matching-based embedding, phase coding, SS, and tone insertion. Each classification has its advantages and disadvantages with regard to performance. However, the diversity of implementation and ideas produce different results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank USIM graduate Financial Assistance Scheme (PHD), UNIVERSITI SAINS ISLAM MALAYSIA, for the funding received under the Partial Graduate Research Assistantship (GRA) scheme. The authors also would like to thank the Ministry of Higher Education and Scientific Research, Studies Planning and Follow-up Directorate, Iraq and National University of Malaysia (UKM) for their support in the present work. The authors are also grateful to the anonymous reviewers for their constructive comments and valuable contributions.

References

[1] A. Kaur, M.K. Dutta, K.M. Soni, N. Taneja, Localized & self adaptive audio watermarking algorithm in the wavelet domain, J. Inf. Secur. Appl. 33 (2017) 1–15, http://dx.doi.org/10.1016/j.jisa.2016.12.003.

[2] J. Vimal, A.M. Alex, Audio steganography using dual randomness LSB method, in: 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies, ICCICCT 2014, 2014, pp. 941–944, http://dx.doi.org/10.1109/ICCICCT.2014.6993093.

[3] D.M. Ballesteros L, J.M. Moreno A, Highly transparent steganography model of speech signals using efficient wavelet masking, Expert Syst. Appl. 39 (10) (2012) 9141–9149, http://dx.doi.org/10.1016/j.eswa.2012.02.066.

[4] M. Asad, J. Gilani, A. Khalid, An enhanced least significant bit modification technique for audio steganography, in: Proceedings - International Conference on Computer Networks and Information Technology, 2011, pp. 143–147, http://dx.doi.org/10.1109/ICCNIT.2011.6020921.

[5] F. Djebbar, B. Ayad, K.A. Meraim, H. Hamam, Comparative study of digital audio steganography techniques, EURASIP J. Audio Speech Music Process. 2012 (1) (2012) 1–16, http://dx.doi.org/10.1186/1687-4722-2012-25.

[6] W. Bender, N. Morimoto, A. Lu, Techniques for data hiding, IBM Syst. J. 35 (3-4) (1996) 313–336.

[7] I. Bilal, R. Kumar, M.S. Roj, P.K. Mishra, Recent advancement in audio steganography, in: Proc. 2014 3rd Int. Conf. Parallel, Distrib. Grid Comput., PDGC 2014, 2015, pp. 402–405, http://dx.doi.org/10.1109/PDGC.2014.7030779.

[8] S.K. Dastoor, Comparative analysis of steganographic algorithms intacting the information in the speech signal for enhancing the message security in next generation mobile devices, in: Proc. 2011 World Congr. Inf. Commun. Technol., WICT 2011, 2011, pp. 279–284, http://dx.doi.org/10.1109/WICT.2011.6141258.

[9] M.H. Farouk, Steganography and security of speech signal, in: Application of Wavelets in Speech Processing, 2014, pp. 45–47.

[10] R. Tanwar, M. Bisla, Audio steganography, in: ICROIT 2014 - Proc. 2014 Int. Conf. Reliab. Optim. Inf. Technol., 2014, pp. 322–325, http://dx.doi.org/10.1109/ICROIT.2014.6798347.

[11] P.P. Balgurgi, S.K. Jagtap, Audio steganography used for secure data transmission, Adv. Intell. Syst. Comput. 174 (2013) 699–706, http://dx.doi.org/10.1007/978-81-322-0740-5_83.

[12] M. Zamani, A.B.A. Manaf, S.M. Abdullah, An overview on audio steganography techniques, Int. J. Digit. Content Technol. Appl. 6 (13) (2012) 107–122, http://dx.doi.org/10.4156/jdcta.vol6.issue13.13.

[13] F. Djebbar, B. Ayad, H. Hamam, K. Abed-Meraim, A view on latest audio steganography techniques, in: 2011 Int. Conf. Innov. Inf. Technol., IIT 2011, 2011, pp. 409–414, http://dx.doi.org/10.1109/INNOVATIONS.2011.5893859.

[14] M.L.M. Kiah, B.B. Zaidan, A.A. Zaidan, A.M. Ahmed, S.H. Al-bakri, A review of audio based steganography and digital watermarking, Int. J. Phys. Sci. 6 (16) (2011) 3837–3850, http://dx.doi.org/10.5897/IJPS11.577.

[15] A. Das, Steganography: Secret data hiding in multimedia, in: Signal Conditioning, Springer, Berlin Heidelberg, 2012, pp. 275–295.

[16] P. Johri, A. Kumar, Amba, Review paper on text and audio steganography using GA, in: International Conference on Computing, Communication and Automation, ICCCA 2015, 2015, pp. 190–192, http://dx.doi.org/10.1109/CCAA.2015.7148403.

[17] E. Abdelfattah, A. Mahmood, Steganography and Steganalysis: Current Status and Future Directions, vol. 151, Springer, New York, NY, 2013.

[18] C. Paulin, S.-A. Selouani, E. Hervet, A comparative study of audio/speech steganalysis techniques catherine, in: 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) A, 2017.

[19] M.S. Atoum, S. Ibrahimn, G. Sulong, A. Zeki, A. Abubakar, Exploring the challenges of MP3 audio steganography, in: Proceedings - 2013 International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2013, 2013, 2016, pp. 156–161, http://dx.doi.org/10.1109/ACSAT.2013.38.

[20] M.S. Atoum, A comparative study of combination with different LSB techniques in MP3 steganography, in: Information Science and Applications, Springer, Berlin, Heidelberg, 2015, pp. 551–560.

[21] K. Shafi, A. Sankaranarayanan, G. Prashanth, A. Mohan, A novel audio steganography scheme using amplitude differencing, in: Trendz Inf. Sci. Comput. (TISC2010), IEEE, 2010, pp. 163–167, http://dx.doi.org/10.1109/TISC.2010.5714631.

[22] D. Bhattacharyya, P. Dutta, M.O. Balitanas, T.H. Kim, P. Das, Hiding data in audio signal, in: Communications in Computer and Information Science, 77 CCIS, 2010, pp. 23–29.

[23] M. Khademi, M. Ali Tinati, Audio steganography by using of linear predictive coding analysis in the safe places of discrete wavelet transform domain, in: Electrical Engineering (ICEE), 2011 19th Iranian Conference on. IEEE, 2011, pp. 1–5.

[24] J.C. Collins, S.S. Agaian, System for non-disruptive high capacity indexed data embedding and recovery using multimedia signal covers, in: Mob. Multimedia/Image Process. Secur. Appl. 2011., vol. 8063, International Society for Optics and Photonics, 2011, pp. 1–11, http://dx.doi.org/10.1117/12.883740.

[25] K. Shafi, A. Sankaranarayanan, G. Prashanth, A. Mohan, A novel audio steganography scheme using amplitude differencing, in: Trendz in Information Sciences & Computing (TISC), 2010, 2010, pp. 163–167, http://dx.doi.org/10.1109/TISC.2010.5714631.

[26] K. Sakthisudhan, P. Prabhu, P. Thangaraj, C.M. Marimuthu, Dual steganography approach for secure data communication, Proc. Eng. 38 (2012) 412–417, http://dx.doi.org/10.1016/j.proeng.2012.06.051.

[27] M. Baritha Begum, Y. Venkataramani, LSB Based audio steganography based on text compression, Proc. Eng. 30 (2011) (2012) 702–710, http://dx.doi.org/10.1016/j.proeng.2012.01.917.

[28] P.P. Balgurgi, S.K. Jagtap, Intelligent processing: An approach of audio steganography, in: Proceedings - 2012 International Conference on Communication, Information and Computing Technology, ICCICT 2012, 2012, pp. 1–6, http://dx.doi.org/10.1109/ICCICT.2012.6398182.

[29] R.F. Olanrewaju, O. Khalifa, H.B. Abdul Rahman, Increasing the hiding capacity of low-bit encoding audio steganography using a novel embedding technique, World Appl. Sci. J. 21 (1) (2013) 79–83, Special Issue. http://dx.doi.org/10.5829/idosi.wasj.2013.21.mae.99926.

[30] D.M. Ballesteros L, J.M. Moreno A, Speech-in-speech hiding scheme based on least significant bit substitution and adaptive key, Comput. Electr. Eng. 39 (4) (2013) 1192–1203, http://dx.doi.org/10.1016/j.compeleceng.2013.02.006.

[31] Y. Huo, S. Xiang, S. Liu, X. Luo, Z. Bai, Reversible audio watermarking algorithm using non-causal prediction, Wuhan Univ. J. Nat. Sci. 18 (5) (2013) 455–460, http://dx.doi.org/10.1007/s11859-013-0956-2.

[32] Liehuang Zhu, D. Liu, L. Yu, Y. Xie, M. Wang, Content integrity and non-repudiation preserving audio- hiding scheme based on robust digital signature, Secur. Commun. Netw. 6 (11) (2013) 1331–1343.

[33] H.I. Shahadi, R. Jidin, W.H. Way, Lossless audio steganography based on lifting wavelet transform and dynamic stego key, Indian J. Sci. Technol. 7 (3) (2014) 323–334.

[34] J.Y. Li, Y.Q. Yang, B. Yang, An audio hiding algorithm based on spline interpolation and wavelet transform, Appl. Mech. Mater. 571–572 (2014) 39–43, http://dx.doi.org/10.4028/www.scientific.net/AMM.571-572.39.

[35] M. Bazyar, R. Sudirman, A new method to increase the capacity of audio steganography based on the LSB algorithm, J. Teknol. 74 (6) (2015) 49–53, http://dx.doi.org/10.11113/jt.v74.4667.

[36] M. Bazyar, R. Sudirman, A robust data embedding method for MPEG layer III audio steganography, Int. J. Secur. Appl. 9 (12) (2015) 317–328, http://dx.doi.org/10.14257/ijsia.2015.9.12.31.

[37] A. Kanhe, G. Aghila, C.Y.S. Kiran, C.H. Ramesh, G. Jadav, M.G. Raj, Robust audio steganography based on advanced encryption standards in temporal domain, in: 2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015, 2015, pp. 1449–1453, http://dx.doi.org/10.1109/ICACCI.2015.7275816.

[38] B. Datta, S. Tat, Robust high capacity audio steganography using modulo operator, in: Computer, Communication, Control and Information Technology (C3IT), 2015 Third International Conference on. IEEE, 2015.

[39] H. S, U.D. Acharya, R.A., D.S., J.U.K., Audio steganography in discrete wavelet transform domain, Int. J. Appl. Eng. Res. ISSN 10 (16) (2015) 973–4562.

[40] R.R. Devi, D. Pugazhenthi, Ideal sampling rate to reduce distortion in audio steganography, Proc. Comput. Sci. 85 (Cms) (2016) 418–424, http://dx.doi.org/10.1016/j.procs.2016.05.185.

[41] N. Taneja, P. Gupta, Implementation of dual security through DSA and audio steganography, in: Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015, 2016, pp. 1349–1352, http://dx.doi.org/10.1109/ICGCIoT.2015.7380676.

[42] A. Gambhir, S. Khara, Integrating RSA cryptography & audio steganography, in: Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2016, 2016, pp. 481–484, http://dx.doi.org/10.1109/CCAA.2016.7813767.

[43] F. Baig, M.F. Khan, S. Beg, T. Shah, K. Saleem, Onion steganography: a novel layering approach, Nonlinear Dynam. 84 (3) (2016) 1431–1446, http://dx.doi.org/10.1007/s11071-015-2580-5.

[44] C.K. Sahu, P.K. Sethy, A novel technique for embedding audio in audio to ensure secrecy, in: Int. Conf. Commun. Signal Process. ICCSP 2016, 2016, pp. 83–85, http://dx.doi.org/10.1109/ICCSP.2016.7754448.

[45] D. Renza, C. Lemus, D. M. B. L, Highly transparent and secure scheme for concealing text within audio, in: Iberoamerican Congress on Pattern Recognition, 2016, p. 10125, 27–35, http://dx.doi.org/10.1007/978-3-319-52277-7.

[46] M.C. Trivedi, S. Mishra, V.K. Yadav, Metamorphic cryptography using strength of chaotic sequence and XORing method, J. Intell. Fuzzy Syst. 32 (5) (2017) 3365–3375, http://dx.doi.org/10.3233/JIFS-169277.

[47] M. Parthasarathi, T. Shreekala, T. Nadu, T. Nadu, Secured data hiding in audio files using audio steganography algorithm, Int. J. Pure Appl. Math. 114 (7) (2017) 743–753.

[48] S.E. El-Khamy, N. Korany, M.H. El-Sherif, Robust image hiding in audio based on integer wavelet transform and chaotic maps hopping, in: Radio Science Conference (NRSC), 2017 34th National. IEEE, 2017, 2017, pp. 205–212.

[49] A. Devi, K.B. Shivakumar, Novel audio steganography technique for ECG signals in point of care systems (NASTPOCS), in: Proceedings - 2016 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2016, 2016, pp. 101–106, http://dx.doi.org/10.1109/CCEM.2016.026.

[50] X. Huang, R. Kawashima, N. Segawa, Y. Abe, Design and implementation of synchronized audio-to-audio steganography scheme, in: Proceedings - 2008 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2008, 2008, pp. 331–334, http://dx.doi.org/10.1109/IIH-MSP.2008.98.

[51] D. Pal, A. Goswami, N. Ghoshal, Lossless audio steganography in spatial domain (LASSD), in: Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), Springer, Berlin, Heidelberg, 2013, pp. 575–582, http://dx.doi.org/10.1007/978-3-642-35314-7.

[52] P. Pathak, A.K. Chattopadhyay, A. Nag, A new audio steganography scheme based on location selection with enhanced security, in: 1st International Conference on Automation, Control, Energy and Systems - 2014, ACES 2014, 2014, pp. 1–4, http://dx.doi.org/10.1109/ACES.2014.6807979.

[53] Y. Kakde, P. Gonnade, P. Dahiwale, Audio-video steganography, in: ICIIECS 2015-2015 IEEE International Conference on Innovations in Information, Embedded and Communication Systems, 2015, pp. 1–6, http://dx.doi.org/10.1109/ICIIECS.2015.7192885.

[54] P. Shah, P. Choudhari, S. Sivaraman, Adaptive wavelet packet based audio steganography using data history, in: IEEE Reg. 10 Colloq. 3rd Int. Conf. Ind. Inf. Syst. ICIIS, 2008, pp. 0–4, http://dx.doi.org/10.1109/ICIINFS.2008.4798397.

[55] K. Gopalan, Q. Shi, Audio steganography using bit modification - A tradeoff on perceptibility and data robustness for large payload audio embedding, in: Proceedings - International Conference on Computer Communications and Networks, ICCCN, 2010, http://dx.doi.org/10.1109/ICCCN.2010.5560097.

[56] M. Asad, J. Gilani, A. Khalid, Three layered model for audio steganography, in: Proc. - 2012 Int. Conf. Emerg. Technol., ICET 2012, 2012, pp. 270–275, http://dx.doi.org/10.1109/ICET.2012.6375438.

[57] S. Ghosh, D. De, D. Kandar, A double layered additive space sequenced audio steganography technique for mobile network, in: 2012 International Conference on Radar, Communication and Computing, ICRCC 2012, 2012, pp. 29–33, http://dx.doi.org/10.1109/ICRCC.2012.6450542.

[58] H. Kumar, Anuradha, Enhanced LSB technique for audio steganography, in: 2012 3rd International Conference on Computing, Communication and Networking Technologies, ICCCNT 2012, 2012, no. July, 26–29, 2012, http://dx.doi.org/10.1109/ICCCNT.2012.6395978.

[59] S. Rekik, D. Guerchi, S. Selouani, H. Hamam, Speech steganography using wavelet and fourier transforms, EURASIP J. Audio Speech Music Process. 1 (20) (2012) 1–14.

[60] X. Zhang, A novel algorithm for embedding watermarks into audio signal based on DCT, in: Proceedings of the International Conference on Information Engineering and Applications (IEA) 2012, 220, no. vol. 5, 2013, pp. 29–35, http://dx.doi.org/10.1007/978-1-4471-4844-9.

[61] N. Mukherjee, A. Bhattacharya, S.S. Bose, Evolutionary multibit grouping steganographic algorithm, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), in: LNCS, vol. 8303, 2013, pp. 285–296, http://dx.doi.org/10.1007/978-3-642-45204-8_22.

[62] Y. Jiang, L. Zhang, S. Tang, Z. Zhou, Real-time covert voip communications over smart grids by using AES-based audio steganography, in: Proc. - 2013 IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput. GreenCom-IThings-CPSCom 2013, 2013, pp. 2102–2107, http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.395.

[63] H.B. Dieu, N.X. Huy, Hiding data in audio using modified CPT scheme, in: 2013 International Conference on Soft Computing and Pattern Recognition, SoCPaR 2013, 2013, pp. 396–400, http://dx.doi.org/10.1109/SOCPAR.2013.7054098.

[64] S.Y. Tang, Y.J. Jiang, L.P. Zhang, Z.B. Zhou, Audio steganography with AES for real-time covert voice over internet protocol communications, Sci. China Inf. Sci. 57 (3) (2014) 1–14, http://dx.doi.org/10.1007/s11432-014-5063-2.

[65] N. Gupta, N. Sharma, Dwt and LSB based audio steganography, in: ICROIT 2014 - Proceedings of the 2014 International Conference on Reliability, Optimization and Information Technology, 2014, pp. 428–431, http://dx.doi.org/10.1109/ICROIT.2014.6798368.

[66] S.S. Verma, R. Gupta, G. Shrivastava, A novel technique for data hiding in audio carrier by using sample comparison in DWT domain, in: Proc. - 2014 4th Int. Conf. Commun. Syst. Netw. Technol., CSNT 2014, 2014, pp. 639–643, http://dx.doi.org/10.1109/CSNT.2014.134.

[67] M. Charfeddine, M. El'Arbi, C. Ben Amar, A new DCT audio watermarking scheme based on preliminary MP3 study, Multimedia Tools Appl. 70 (3) (2014) 1521–1557, http://dx.doi.org/10.1007/s11042-012-1167-0.

[68] I. Bilal, R. Kumar, Audio steganography using QR decomposition and fast fourier transform, Indian J. Sci. Technol. 8 (34) (2015) http://dx.doi.org/10.17485/ijst/2015/v8i34/69604.

[69] M. Punetha, R. Kumar, M. Bhattacharya, N. Jain, M. Gawande, Safe transmission of text files through a new audio steganography technique, in: Proceedings - 2014 2nd International Symposium on Computational and Business Intelligence, ISCBI 2014, 2015, pp. 58–62, http://dx.doi.org/10.1109/ISCBI.2014.20.

[70] N.C. Debnath, H. Abushama, C. Science, A Multilayered Scheme for Transparent Audio Data Hiding, in: 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), 2015, pp. 1–6.

[71] G.P. TVS, S. Varadarajan, A novel hybrid audio steganography for imperceptible data hiding, in: Communications and Signal Processing (ICCSP), 2015, pp. 634–638.

[72] V. Sharma, LSB modification based audio steganography using trusted third party key indexing method, in: 2015 Third International Conference On Image Information Processing (ICIIP), 2015, pp. 403–406.

[73] Y.F. Huang, S.Y. Tang, Covert voice over internet protocol communications based on spatial model, Sci. China Technol. Sci. 59 (1) (2016) 117–127, http://dx.doi.org/10.1007/s11431-015-5955-4.

[74] B. Datta, P.K. Pal, S.K. Bandyopadhyay, Multi-bit data hiding in randomly chosen LSB layers of an audio, in: International Conference on Information Technology, 2016, pp. 283–287, http://dx.doi.org/10.1109/ICIT.2016.36.

[75] M.J. Hussain, K.F. Rafat, Enhanced audio LSB steganography for secure communication, Int. J. Adv. Comput. Sci. Appl. 7 (1) (2016) 340–347.

[76] S.V. Jadhav, An efficient new audio steganography scheme based on location selection with enhanced security, Int. J. Eng. Sci. Comput. 5 (2) (2016) 334–337.

[77] P. Bhalde, Performance improvement: Audio steganography technique parity bit combined with cryptography, in: ACM International Conference Proceeding Series, 2016, http://dx.doi.org/10.1145/2905055.2905196.

[78] A.A. Alsabhany, F. Ridzuan, A.H. Azni, A hybrid method for data communication using encrypted audio steganography, Adv. Sci. Lett. 23 (5) (2017) http://dx.doi.org/10.1166/asl.2017.8946.

[79] Y. Jiang, S. Tang, An efficient and secure voip communication system with chaotic mapping and message digest, Multimed. Syst. (2017) 1–9, http://dx.doi.org/10.1007/s00530-017-0565-6.

[80] H.-J. Shiu, B.-S. Lin, C.-W. Cheng, C.-H. Huang, C.-L. Lei, High-capacity data-hiding scheme on synthesized pitches using amplitude enhancement—A new vision of non-blind audio steganography, Symmetry (Basel) 9 (6) (2017) 92, http://dx.doi.org/10.3390/sym9060092.

[81] S.E. El-Khamy, N.O. Korany, M.H. El-Sherif, A security enhanced robust audio steganography algorithm for image hiding using sample comparison in discrete wavelet transform domain and RSA encryption, Multimedia Tools Appl. 76 (22) (2017) 24091–24106, http://dx.doi.org/10.1007/s11042-016-4113-8.

[82] M. Zou, Z. Li, A wav-audio steganography algorithm based on amplitude modifying, in: Proc. - 2014 10th Int. Conf. Comput. Intell. Secur., CIS 2014, 2015, pp. 489–493, http://dx.doi.org/10.1109/CIS.2014.78.

[83] S.S. Shahreza, M.T.M. Shalmani, Adaptive wavelet domain audio steganography with high capacity and low error rate, in: 2007 International Conference on Information and Emerging Technologies, ICIET, 2007, pp. 25–29, http://dx.doi.org/10.1109/ICIET.2007.4381305.

[84] M. Pooyan, A. Delforouzi, LSB-Based audio steganography method based on lifting wavelet transform, in: ISSPIT 2007-2007 IEEE International Symposium on Signal Processing and Information Technology, 2007, pp. 600–603, http://dx.doi.org/10.1109/ISSPIT.2007.4458198.

[85] A. Delforouzi, M. Pooyan, Adaptive digital audio steganography based on integer wavelet transform, Circuits, Syst. Signal Process. 27 (2) (2008) 247–259, http://dx.doi.org/10.1007/s00034-008-9019-x.

[86] K. Gopalan, Audio steganography by modification of cepstrum at a pair of frequencies, in: International Conference on Signal Processing Proceedings, ICSP, 2008, pp. 2178–2181, http://dx.doi.org/10.1109/ICOSP.2008.4697579.

[87] K. Gopalan, A unified audio and image steganography by spectrum modification, in: Proc. IEEE Int. Conf. Ind. Technol, 2009, http://dx.doi.org/10.1109/ICIT.2009.4939516.

[88] A. Delforouzi, M. Pooyan, Adaptive and efficient audio data hiding method in temporal domain, in: ICICS 2009 - Conf. Proc. 7th Int. Conf. Information, Commun. Signal Process, 2009, pp. 0–3, http://dx.doi.org/10.1109/ICICS.2009.5397486.

[89] C.Y. Chang, H.J. Wang, W.C. Shen, Copyright-proving scheme for audio with counter-propagation neural networks, Digit. Signal Process. 20 (4) (2010) 1087–1101, http://dx.doi.org/10.1016/j.dsp.2009.12.001.

[90] D.E. Skopin, I.M.M. El-Emary, R.J. Rasras, R.S. Diab, Advanced algorithms in audio steganography for hiding human speech signal, in: Proceedings - 2nd IEEE International Conference on Advanced Computer Control, ICACC 2010, 3, 2010, pp. 29–32, http://dx.doi.org/10.1109/ICACC.2010.5486735.

[91] M. Wakiyama, Y. Hidaka, K. Nozaki, An audio steganography by a low-bit coding method with wave files, in: Proceedings - 2010 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP 2010, 2010, pp. 530–533, http://dx.doi.org/10.1109/IIHMSP.2010.135.

[92] M.A. Ahmed, M.L.M. Kiah, B.B. Zaidan, 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, J. Appl. Sci. 10 (1) (2010) 59–64.

[93] M. Sheikhan, K. Asadollahi, R. Shahnazi, Improvement of embedding capacity and quality of DWT-based audio steganography systems, World Appl. Sci. J. 13 (3) (2011) 507–516.

[94] T.S. Wu, H.Y. Lin, W.C. Hu, Y. Sen Chen, Audio watermarking scheme with dynamic adjustment in mute period, Expert Syst. Appl. 38 (6) (2011) 6787–6792, http://dx.doi.org/10.1016/j.eswa.2010.12.069.

[95] R. Roselinkiruba, R. Balakirshnan, Secure steganography in audio using inactive frames of voip streams, in: 2013 IEEE Conference on Information and Communication Technologies, ICT 2013, 2013, pp. 491–495, http://dx.doi.org/10.1109/CICT.2013.6558145.

[96] H.B. Dieu, An Improved Technique for Hiding Data in Audio, in: Digital Information and Communication Technology and it's Applications (DICTAP), 2014 Fourth International Conference on. IEEE, 2014, pp. 149–153.

[97] K. Gopalan, J. Fu, An imperceptible and robust audio steganography employing bit modification, in: Proc. IEEE Int. Conf. Ind. Technol. 2015-June, 2015, pp. 1635–1638, http://dx.doi.org/10.1109/ICIT.2015.7125331.

[98] K. Gopalan, A robust bit modification audio steganography for covert communication, in: The Eleventh Advanced International Conference on Telecommunications A, 2015, no. c, pp. 7–10.

[99] D.C. Kar, C.J. Mulkey, A multi-threshold based audio steganography scheme, J. Inf. Secur. Appl. 23 (2015) 54–67, http://dx.doi.org/10.1016/j.jisa.2015.02.001.

[100] M. Bazyar, R. Sudirman, A new data embedding method for mpeg layer iii audio steganography, J. Teknol. 78 (7–5) (2016) 65–73, http://dx.doi.org/10.11113/jt.v78.9452.
[101] A. Kanhe, G. Aghila, DCT Based audio steganography in voiced and unvoiced frames, in: Proc. Int. Conf. Informatics Anal. - ICIA-16, No. 1, 2016, pp. 1–4, http://dx.doi.org/10.1145/2980258.2980360.
[102] M. Shoaib, Z. Khan, D. Shehzad, T. Dag, A. Iqbal, N. Ul, Performance improvement of threshold based audio steganography using parallel computation, Int. J. Adv. Comput. Sci. Appl. 7 (10) (2016) 290–294, http://dx.doi.org/10.14569/IJACSA.2016.071039.
[103] H. Liu, J. Liu, R. Hu, X. Yan, S. Wan, Adaptive audio steganography scheme based on wavelet packet energy, in: 2017 IEEE 3rd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS), 2017, pp. 26–31, http://dx.doi.org/10.1109/BigDataSecurity.2017.20.
[104] M. Nathan, N. Parab, K.T. Talele, Audio steganography using spectrum manipulation, in: Technol. Syst. Manag. No. 2011, Springer, Berlin, Heidelberg, 2011, pp. 152–159, http://dx.doi.org/10.1007/978-3-642-20209-4_21.
[105] W. Luo, Y. Zhang, H. Li, Adaptive audio steganography based on advanced audio coding and syndrome-trellis coding, in: Digital Forensics and Watermarking: 16th International Workshop, IWDW 2017, Vol. 10431, 2017, pp. 177–186, http://dx.doi.org/10.1007/978-3-319-64185-0.
[106] E. Esen, A.A. Alatan, Comparison of forbidden zone data hiding and quantization index modulation, Digit. Signal Process. 22 (1) (2012) 181–189, http://dx.doi.org/10.1016/j.dsp.2011.10.005.
[107] L. Sun, et al., Dynamic matrix encoding strategy for voice-over-IP steganography, J. Cent. South Univ. Technol. 17 (6) (2010) 1285–1292, http://dx.doi.org/10.1007/s11771.
[108] M. Zamani, A.B.A. Manaf, H.R. Zeidanloo, S.S. Chaeikar, Genetic substitution-based audio steganography for high capacity applications, Internet Technol. Secur. Trans. 3 (1) (2011) 97–110.
[109] S. Anguraj, S.P. Shantharajah, R.A. Murugan, E. Balaji, R. Maneesh, S. Prasath, A fusion of A-B MAP cipher and ASET algorithms for the enhanced security and robustness in audio steganography, in: International Conference on Recent Trends in Information Technology, ICRTIT 2011, 2011, pp. 205–210, http://dx.doi.org/10.1109/ICRTIT.2011.5972254.
[110] R. Darsana, A. Vijayan, Audio steganography using modified LSB and PVD, Trends Netw. Commun. 197 (2011) 11–20.
[111] S. Roy, J. Parida, A.K. Singh, A.S. Sairam, Audio steganography using LSB encoding technique with increased capacity and bit error rate optimization, in: Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology - CCSEIT '12, No. October, 2012, pp. 372–376, http://dx.doi.org/10.1145/2393216.2393279.
[112] S. Banerjee, S. Roy, M.S. Chakraborty, S. Das, A variable higher bit approach to audio steganography, in: 2013 Int. Conf. Recent Trends Inf. Technol. ICRTIT 2013, Vol. 2, 2013, pp. 46–49, http://dx.doi.org/10.1109/ICRTIT.2013.6844178.
[113] K. Bhowal, D. Bhattacharyya, A. Jyoti Pal, T.H. Kim, A GA based audio steganography with enhanced security, Telecommun. Syst. 52 (4) (2013) 2197–2204, http://dx.doi.org/10.1007/s11235-011-9542-0.
[114] R. Tanwar, B. Sharma, S. Malhotra, A robust substitution technique to implement audio steganography, in: ICROIT 2014 - Proceedings of the 2014 International Conference on Reliability, Optimization and Information Technology, 2014, pp. 290–293, http://dx.doi.org/10.1109/ICROIT.2014.6798340.
[115] L.B.A. Rahim, S. Bhattacharjee, I.B.A. Aziz, An audio steganography technique to maximize data hiding capacity along with least modification of host, in: Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), Vol. 285, 2014, pp. 277–289, http://dx.doi.org/10.1007/978-981-4585-18-7.
[116] S. Bhattacharjee, L.B.A. Rahim, I.B.A. Aziz, Hiding of compressed bit stream into audio file to enhance the confidentiality and portability of a data transmission system, in: 2015 International Symposium on Mathematical Sciences and Computing Research, ISMSC 2015 - Proceedings, Vol. 2015, 2015, pp. 196–201, http://dx.doi.org/10.1109/ISMSC.2015.7594052.
[117] X. Huang, N. Ono, I. Echizen, A. Nishimura, Reversible audio information hiding based on integer DCT coefficients with adaptive hiding locations, Digit. Watermarking 8389 (2014) 376–389, http://dx.doi.org/10.1007/978-3-662-43886-2.
[118] M. Zamani, A.B.A. Manaf, Genetic algorithm for fragile audio watermarking, Telecommun. Syst. 59 (3) (2015) 291–304, http://dx.doi.org/10.1007/s11235-014-9936-x.
[119] M. Rana, F.B. Kunwar, A temporal domain audio steganography technique using genetic algorithm, in: Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016, pp. 2016, pp. 3141–3146.
[120] P. Liu, S. Li, H. Wang, Steganography integrated into linear predictive coding for low bit-rate speech codec, Multimed. Tools Appl. 76 (2) (2017) 2837–2859, http://dx.doi.org/10.1007/s11042-016-3257-x.
[121] K. Yang, X. Yi, X. Zhao, L. Zhou, Adaptive MP3 steganography using equal length entropy codes substitution, in: International Workshop on Digital Watermarking, Vol. 10431, 2017, pp. 202–216, http://dx.doi.org/10.1007/978-3-319-64185-0.
[122] Y. Ren, H. Wu, L. Wang, An AMR adaptive steganography algorithm based on minimizing distortion, Multimed. Tools Appl. 77 (10) (2017) 12095–12110, http://dx.doi.org/10.1007/s11042-017-4860-1.
[123] H.I. Shahadi, R. Jidin, High capacity and inaudibility audio steganography scheme, in: Proceedings of the 2011 7th International Conference on Information Assurance and Security, IAS 2011, 2011, pp. 104–109, http://dx.doi.org/10.1109/ISIAS.2011.6122803.
[124] E.T. Su, Robust data embedding based probabilistic global search in MDCT domain, in: International Conference on Informatics Engineering and Information Science, 2011, pp. 290–300.
[125] M. Srivastava, M.Q. Rafiq, A novel approach to secure communication using audio steganography, Adv. Mater. Res. 403–408 (2011) 963–969, http://dx.doi.org/10.4028/www.scientific.net/AMR.403-408.963.
[126] H.I. Shahadi, R. Jidin, W.H. Way, A novel and high capacity audio steganography algorithm based on adaptive data embedding positions, Res. J. Appl. Sci. Eng. Technol. 7 (11) (2014) 2311–2323, http://dx.doi.org/10.19026/rjaset.7.531.
[127] A.K. Mandal, M. Kaosar, M.O. Islam, M.D. Hossain, An approach for enhancing message security in audio steganography, in: 16th Int'L Conf. Computer and Information Technology, ICCIT 2013, 2014, pp. 383–388, http://dx.doi.org/10.1109/ICCITechn.2014.6997310.
[128] V. Van Tam, T. Duc-Tan, N.T. Thuy, P.T. Hanh, Embedding data into audio signal by combining sliding window technique with a novel marking bit method, Adv. Multimed. Ubiquitous Eng. 352 (June) (2015) 191–197, http://dx.doi.org/10.1007/978-3-662-47487-7.
[129] H. Tian, et al., Improved adaptive partial-matching steganography for Voice over IP, Comput. Commun. 70 (2015) 95–108, http://dx.doi.org/10.1016/j.comcom.2015.08.011.
[130] R. Chowdhury, D. Bhattacharyya, S.K. Bandyopadhyay, T.-H. Kim, A view on LSB based audio steganography, Int. J. Secur. Appl. 10 (2) (2016) 51–62, http://dx.doi.org/10.14257/ijsia.2016.10.2.05.
[131] S. Hemalatha, U.D. Acharya, A. Renuka, Audio data hiding technique using integer wavelet transform, Int. J. Electron. Secur. Digit. Forensics 8 (2) (2016) 131–147.
[132] A.H. Ali, M.R. Mokhtar, L.E. George, Enhancing the hiding capacity of audio steganography based on block mapping, J. Theor. Appl. Inf. Technol. 95 (7) (2017) 1441–1448.
[133] A.H. Ali, L.E. George, M.R. Mokhtar, An adaptive high capacity model for secure audio communication based on fractal coding and uniform coefficient modulation, circuits, syst, Signal Process. (2020) 1–28, http://dx.doi.org/10.1007/s00034-020-01409-7.
[134] F. Djebbar, B. Ayad, K. Abed-Meraim, H. Hamam, Unified phase and magnitude speech spectra data hiding algorithm, Secur. Commun. Networks 6 (8) (2013) 961–971.
[135] E. Rivas, Fourier phase domain steganography: Phase bin encoding via interpolation, mob. Multimedia/image process, Mil. Secur. Appl. 6579 (2007) 65790W, http://dx.doi.org/10.1117/12.719512.
[136] N. Parab, M. Nathan, K.T. Talele, Audio steganography using differential phase encoding, Technol. Syst. Manag. (2011) 146–151.
[137] X. Lu, Y. Cao, P. Lu, A. Zhai, Digital audio information hiding based on arnold transformation and double random-phase encoding technique, Optik (Stuttg) 123 (8) (2012) 697–702, http://dx.doi.org/10.1016/j.ijleo.2011.06.027.
[138] F. Djebbar, B. Ayad, Audio steganograpgy by phase modification, in: 8th International Conference on Emerging Security Information. Syst. Technol. 2014, pp. 31–35.
[139] S.K. Moon, R.D. Raut, Application of data hiding in audio-video using anti forensics technique for authentication and data security, in: 2014 IEEE International Advance Computing Conference (IACC), 2014, pp. 1110–1115, http://dx.doi.org/10.1109/IAdCC.2014.6779481.
[140] D. Pal, N. Ghoshal, Secured data transmission through audio signal (SDTAS), in: 9th Int. Conf. Ind. Inf. Syst. ICIIS 2014, 2015, http://dx.doi.org/10.1109/ICIINFS.2014.7036649.
[141] A. Fatnassi, H. Gharsellaoui, S. Bouamama, A new hybrid steganalysis based approach for embedding image in audio and image cover media, IFAC-PapersOnLine 49 (12) (2016) 1809–1814, http://dx.doi.org/10.1016/j.ifacol.2016.07.845.
[142] A.A. Alsabhany, F. Ridzuan, A.H. Azni, The adaptive multi-level phase coding method in audio steganography, IEEE Access 7 (2019) 129291–129306, http://dx.doi.org/10.1109/ACCESS.2019.2940640.
[143] R.M. Nugraha, Implementation of direct sequence spread spectrum steganography on audio data, in: 2011 International Conference on Electrical Engineering and Informatics (ICEEI), 2011, pp. 1–6.

[144] D. Chang, X. Zhang, Q. Liu, G. Gao, Y. Wu, Location based robust audio watermarking algorithm for social TV system, in: Adv. Multimed. Inf. Process. – PCM 2012, Vol. 7674, No. 61202470, 2012, pp. 726–738, http://dx.doi.org/10.1007/978-3-642-34778-8.
[145] S. Shokri, M. Ismail, N. Zainal, A. Shokri, Error probability in spread spectrum (SS) audio watermarking, in: International Conference on Space Science and Communication, IconSpace, No. July, 2013, pp. 169–173, http://dx.doi.org/10.1109/IconSpace.2013.6599457.
[146] R. Kaur, A. Thakur, H.S. Saini, R. Kumar, Enhanced steganographic method preserving base quality of information using LSB, parity and spread spectrum technique, in: Int. Conf. Adv. Comput. Commun. Technol. ACCT, 2015-April, 2015, pp. 148–152, http://dx.doi.org/10.1109/ACCT.2015.139.
[147] H. Crawford, J. Aycock, Supraliminal audio steganography: Audio files tricking audiophiles, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), in: LNCS, Vol. 5806, 2009, pp. 1–14, http://dx.doi.org/10.1007/978-3-642-04431-1_1.
[148] K. Szczypiorski, Stegibiza: New method for information hiding in club music, in: 2016 2nd International Conference on Frontiers of Signal Processing, ICFSP 2016, 2016, pp. 20–24, http://dx.doi.org/10.1109/ICFSP.2016.7802950.
[149] W. Mazurczyk, J. Lubacz, LACK-A voip steganographic method, Telecommun. Syst. 45 (2–3) (2010) 153–163, http://dx.doi.org/10.1007/s11235-009-9245-y.
[150] D. Yan, R. Wang, Huffman table swapping-based steganograpy for MP3 audio, Multimed. Tools Appl. 52 (2–3) (2011) 291–305, http://dx.doi.org/10.1007/s11042-009-0430-5.
[151] D. Yan, R. Wang, X. Yu, J. Zhu, Steganography for MP3 audio by exploiting the rule of window switching, Comput. Secur. 31 (5) (2012) 704–716, http://dx.doi.org/10.1016/j.cose.2012.04.006.
[152] W. Sun, R.J. Shen, F.X. Yu, Z.M. Lu, Data hiding in audio based on audio-to-image wavelet transform and vector quantization, in: Proceedings of the 2012 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2012, 2012, pp. 313–316, http://dx.doi.org/10.1109/IIH-MSP.2012.82.
[153] W. Mazurczyk, P. Szaga, K. Szczypiorski, Using transcoding for hidden communication in IP telephony, Multimed. Tools Appl. 70 (3) (2014) 2139–2165, http://dx.doi.org/10.1007/s11042-012-1224-8.
[154] H. Ghasemzadeh, M.H. Kayvanrad, Toward a robust and secure echo steganography method based on parameters hopping, in: 2015 Signal Process. Intell. Syst. Conf. Sp. 2015, No. November, 2016, pp. 143–147, http://dx.doi.org/10.1109/SPIS.2015.7422329.
[155] M. Naumann, S. Wendzel, W. Mazurczyk, J. Keller, Micro protocol engineering for unstructured carriers: on the embedding of steganographic control protocols into audio transmissions, Secur. Commun. Networks 9 (15) (2016) 2972–2985, http://dx.doi.org/10.1002/sec.1500.
[156] V.K. Yadav, M.C. Trivedi, A. Gupta, I. Hiding, Audio steganography using ZDT : Encryption using indexed based chaotic sequence, in: Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, 2016, http://dx.doi.org/10.1145/2905055.2905272.
[157] A. Janicki, W. Mazurczyk, K. Szczypiorski, Steganalysis of transcoding steganography, Ann. des Telecommun. Telecommun. 69 (7–8) (2014) 449–460, http://dx.doi.org/10.1007/s12243-013-0385-4.
[158] D.Y. Mohammed, P.J. Duncan, M.M. Al-Maathidi, F.F. Li, A system for semantic information extraction from mixed soundtracks deploying MARSYAS framework, in: Proceeding - 2015 IEEE International Conference on Industrial Informatics, INDIN 2015, No. July 2015, 2015, pp. 1084–1089, http://dx.doi.org/10.1109/INDIN.2015.7281886.
[159] TSP Lab - data, 2018, [Online]. Available: http://www-mmsp.ece.mcgill.ca/Documents/Data/. (Accessed 11 July 2018).
[160] Q. Liu, A.H. Sung, M. Qiao, Temporal derivative-based spectrum and mel-cepstrum audio steganalysis, IEEE Trans. Inf. Forensics Secur. 4 (3) (2009) 359–368, http://dx.doi.org/10.1109/TIFS.2009.2024718.
[161] Z. Wei, A. Haojun, H. Ruimin, A novel steganalysis algorithm of phase coding in audio signal, in: Proc. - ALPIT 2007 6th Int. Conf. Adv. Lang. Process. Web Inf. Technol., 2007, pp. 261–264, http://dx.doi.org/10.1109/ALPIT.2007.41.
[162] W. Zeng, H. Ai, R. Hu, An algorithm of echo steganalysis based on power cepstrum and pattern classification, in: ICALIP 2008-2008 Int. Conf. Audio, Lang. Image Process. Proc., 2008, pp. 1344–1348, http://dx.doi.org/10.1109/ICALIP.2008.4590036.
[163] H. Ghasemzadeh, M.H. Kayvanrad, Comprehensive review of audio steganalysis methods, IET Signal Process. 12 (6) (2018) 673–687, http://dx.doi.org/10.1049/iet-spr.2016.0651.
[164] Y. Ren, Q. Xiong, L. Wang, A steganalysis scheme for AAC audio based on MDCT difference between intra and inter frame, in: International Workshop on Digital Watermarking, 2017, pp. 217–231, http://dx.doi.org/10.1007/978-3-319-64185-0.
[165] Y. Wang, X. Yi, X. Zhao, MP3 steganalysis based on joint point-wise and block-wise correlations, Inf. Sci. (Ny) 512 (xxxx) (2020) 1118–1133, http://dx.doi.org/10.1016/j.ins.2019.10.037.
[166] H. Özer, B. Sankur, N. Memon, I. Avcibaş, Detection of audio covert channels using statistical footprints of hidden messages, Digit. Signal Process. A Rev. J. 16 (4) (2006) 389–401, http://dx.doi.org/10.1016/j.dsp.2005.12.001.
[167] Q. Liu, A.H. Sung, M. Qiao, Novel stream mining for audio steganalysis, No. December 2017, 2009, http://dx.doi.org/10.1145/1631272.1631288.
[168] H. Ghasemzadeh, M.K. Arjmandi, Universal audio steganalysis based on calibration and reversed frequency resolution of human auditory system, IET Signal Process. 11 (8) (2017) 916–922, http://dx.doi.org/10.1049/iet-spr.2016.0690.
[169] C. Han, R. Xue, R. Zhang, X. Wang, A new audio steganalysis method based on linear prediction, Multimed. Tools Appl. 77 (12) (2018) 15431–15455, http://dx.doi.org/10.1007/s11042-017-5123-x.
[170] Z. Zhang, X. Yi, X. Zhao, Improving audio steganalysis using deep residual networks, in: International Workshop on Digital Watermarking, 2019, pp. 57–70.