A Systematic Review of Highly Transparent Steganographic Methods for the Digital Audio

Jerzy Pejaś and Łukasz Cierocki(✉)

Faculty of Computer Science and Information Technology, Department of Software Engineering and Cybersecurity, West Pomeranian University of Technology, 49 Żołnierska Street, 71-210 Szczecin, Poland
jpejas@wi.zut.edu.pl, lukasz.cierocki@zut.edu.pl

Abstract. Audio steganography is a rapidly growing aspect of broad information protection. This paper presents an overview of steganographic methods using audio as a medium. As an additional aspect during the review, an effort was made to focus on the transparency requirement of the considered methods used in the steganography process. Previous literature reviews have not focused on a single aspect of method evaluation in sufficient depth. Data for the review were collected from papers published between 2018 and 2022 and gathered from three source databases, i.e. Web of Science, IEEE, and ACM, resulting in a total of 32 entries. The obtained methods were classified according to one of the approaches previously proposed in the literature. A systematic comparative analysis of the retrieved methods has been done, comparing their capacity, robustness, and transparency, with particular emphasis on transparency. In addition, a few of the most promising methods were selected and thoroughly analyzed for transparency behavior.

Keywords: Audio steganography · Transparency · Literature review · Impreceptibility

1 Introduction

We refer to a steganographic system as three successive processes. The first involves embedding confidential information in a carrier, the second is sending this crafted message through a public channel, and the third is recovering the previously embedded message. Each method of audio steganography includes a way of inserting/extracting a secret message in/from an audio signal.

Along with cryptography, steganography is one of the most widely used ways to protect information. Cryptography hides the meaning of information, while steganography hides the very fact of its existence [6]. Such crafted media containing confidential information can be sent to the recipient through a public channel, as it does not present any value to a person not authorized to read it.

K. Saeed and J. Dvorský (Eds.): CISIM 2022, LNCS 13293, pp. 63–77, 2022.
https://doi.org/10.1007/978-3-031-10539-5_5

We can evaluate each steganographic method in terms of three basic requirements: transparency, capacity, and robustness. These requirements can be represented as a triangle like in Fig. 1, where each requirement lies on one of the vertices. Steganographic methods are characterized by the fact that a change in one parameter does not leave the others unaffected. Thus, an increase in capacity is associated with a change in transparency and robustness, and an improvement in transparency will not remain without effect on the capacity [25].

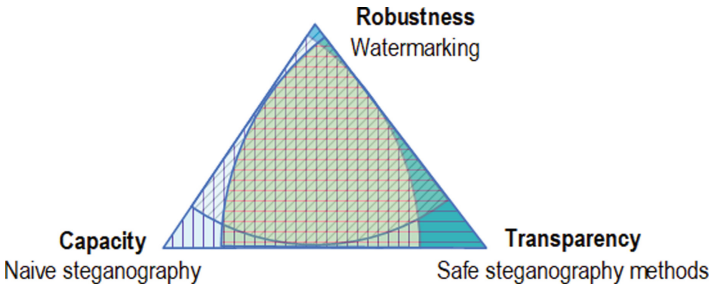

Fig. 1. Audio steganography triangle of requirements

2 Transparency

This paper focuses on the concept of transparency in the context of steganographic methods. The use of audio raises many challenges and research issues; however, it also represents a research gap. As a digital audio signal is a carrier used, the knowledge of limitations and subtle features of the sense of hearing as such allows creating more and more interesting methods using this particular container. The human auditory sense (HAS) is an extremely complex mechanism consisting of many anatomical structures and closely related processes [12].

2.1 Human Auditory System

Hearing perception can be described as a process that begins when sound waves travel through the air to the auricle, then through the external auditory canal to the tympanic membrane. Under the influence of the air vibrations, the eardrum moves the malleus adjacent to it. Vibrations from the malleus are transmitted to the incus and stapes and travel through the auditory tube to the inner ear, where they are converted into nerve impulses that travel through the auditory nerve to the hearing centers in the cerebral cortex [12].

As we can read in [36] and in [12], the "typical" human hearing range is 20 Hz to 20 kHz, with the highest sensitivity in the 1 kHz to 3 kHz range, perfectly matching the frequency range of human speech, which is 500 Hz to 3 kHz.

Chen, et al. write in their paper [16], that International Federation of the Phonographic Industry (IFPI) has its own set of requirements with respect to

method transparency. The method should be characterized by, inaudibility of the embedded message, offer an SNR of more than 20 dB. Furthermore, it is necessary that the capacity is at least 20 bps and that the embedded information is protected against typical stegoanalytic attacks e.g. re-sampling, re-quantization, compression attack etc.

2.2 Measures

As mentioned in the previous paragraph, transparency can be defined by a number of measures. We use two sets of measures to evaluate the transparency of the methods: objective tests, and subjective tests. The most commonly used are objective measures like MSE, PSNR, SNR and PRD, while less frequently used are PESQ, SDG and ODG.

PSNR (Peak to Signal Noise Ratio) is used to evaluate the stego audio quality compared to the cover audio and is expressed in decibels (dB). A higher PSNR value means higher audio quality. The formula gives PSNR:

$$PSNR = 10 \times log\left(\left(\frac{255^2}{MSE}\right)\right) \quad \text{where} \quad MSE = \frac{1}{M \times N}\sum_{0}^{M-1}\sum_{0}^{N-1}\|c - s\|^2 \tag{1}$$

where M, N are the width and height of the signal and c, s are the carrier and stego audio, respectively [4].

In audio specifications, SNR tells us the difference in the maximum volume we can get from the device's own noise. At low SNR and higher volumes this unwanted noise can be heard. A high SNR value is especially desirable for music with high dynamics such as classical or electronic music. SNR is given by formula:

$$SNR = 10 \times log_{10}\left(\frac{\sum_{i=1}^{N} c_i^2}{\sum_{i=1}^{N}(c_i - s_i)^2}\right) \tag{2}$$

where N are the number of signal samples and c, s are the carrier and stego audio, respectively [4].

PRD is a measure that determines the Percentage mean square Root of the Difference between two signals and takes values from 0 to 1 [39]. It is given by the formula,

$$PRD = \sqrt{\left(\frac{\sum_{i=1}^{N}(c_i - s_i)^2}{\sum_{i=1}^{N} c_i^2}\right)} \tag{3}$$

where N are the number of signal samples and c, s are the carrier and stego audio, respectively [4].

An interesting measure of transparency is the Pearson and Kendal correlation also called the normalized correlation (NC). This measure takes values from -1 to 1 with 1 indicating full cross-correlation of the signals and 0 indicating no such correlation and -1 indicating full negative correlation [41]. We calculate this correlation by the formula:

$$\rho(c,s) = \frac{\sum_{i=1}^{N}(c_i - \overline{c})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(c_i - \overline{c})^2}\sqrt{\sum_{i=1}^{N}(s_i - \overline{s})^2}} \tag{4}$$

Among other objective measures, the following PEAQ (Perceptual Evaluation of Audio Quality) [1] and PESQ (Perceptual Evaluation of Speech Quality) [2] standards developed by ITU are worth noting. These are standardized algorithms for objective evaluation of sound quality. The result of the PEAQ algorithm is an objective difference grade (ODG) measured on a 5-point scale from 0 (inaudible), −1 (audible but not annoying), −2 (mildly annoying), −3 (annoying) and −4 (very annoying). The generally accepted standard is for steganographic algorithms to achieve values in the range 0 a −1 [42].

2.3 Related Works

In the field of steganography using digital audio, at least a dozen valuable review articles have been created over the years [6,13,17–19].

One of the most recent is an article [6] where the authors conducted a systematic review of the literature, along with a proposed categorization of the methods, recognition of the key features of each method, and a detailed description of the measures and data sets used. Particularly valuable seems to be the methodological description of the approach to the analysis performed, in which the individual steps of the review are listed along with the keywords used.

In an interesting paper [19] authors gave an overview of the methods for audio and speech. Proposals were offered to classify the methods based on the type of embedding operation performed. Particularly interesting are methods that are based on the principals of human auditory sense (HAS).

A paper [18] where the authors also performed a systematic literature review, with an emphasis on grouping methods by information embedding domain, also deserves mention. In addition, the authors revealed the advantages and disadvantages of the described steganographic methods and made a classification based on the robustness of the method. A performance evaluation has also been made.

Almost every review article tries to bring up how to categorize steganographic methods. Some focus on the medium used, others try to assign methods by the type of embedding operation performed, and others by the domain in which the embedding operations are performed.

Compared to existing review articles, the classification proposed in the [6] article allows an unambiguous distinction between the embedding methods used, thus avoiding the problem of overlapping or low-level segregation of these methods. In particular, this approach allows to clearly distinguish classification of codec-based methods for which an additional coded domain has been introduced [18].

This paper uses during the analysis the categorization method proposed in the article [6]. Based on [6] we can distinguish 8 classes of methods that are applicable to audio steganography. Each of these classes is based on the key idea of information embedding.

In the following, the article is organized as follows. Section 3 describes the motivations and basic information about the steganographic process. In turn, Sect. 4 describes the methodology for conducting this review. Section 5 presents the results of the literature review performed. The final Sect. 6 describes a summary with conclusions and an outline of future work.

3 Motivation

Typically, research in the field of steganography addresses one of three aspects, viz: the domain of embedding, the type of carrier used, and the method of embedding. The first aspect defines the signal domain in which the information is embedded. The most commonly used domains are time, frequency, and wavelets. It is worth mentioning that each domain has different properties regarding the basic requirements of the method. The second aspect involves the type of media used. In steganography involving audio signals, digital lossless audio formats WAV, AAC, FLAC or the lossy compressed MP3 format are most commonly used as a carrier. Rarely, formats beyond this set are used, but there are known examples of VOiP being used as a carrier. The third aspect and, we believe, the most important is the method of embedding information.

There are many ways to embed information in a signal. It usually comes down to manipulating the signal at the bit level (LSB), or changing the values of the coefficients of the different types of transforms. There are also embedding methods that take advantage of changes in codewords used in the process of, for example, audio compression. There have been at least a dozen review articles trying to summarize and classify methods in audio steganography, but none of them has focused on the key requirement of transparency of a given method.

This paper extends the previous reviews with an important aspect such as transparency. During the literature overview, more attention was also paid to codecs-based methods. Due to their practical suitability for real-time transmission of audio information, it was proposed to assign these methods to the coded domain.

Due to this requirement, the data embedded in the audio signal can affect the quality of the signal and can be captured by humans because of the sensitivity of their HAS system. Hence, finding the trade-off between changes resulting from data embedding and signal quality has significant practical importance and is the main focus of the analysis presented later in this paper. For the purposes of the analysis, we propose a systematic division of methods according to the domain in which a given method works and the embedding method used. This allows for a better understanding of the extremely subtle differences between the proposed approaches.

4 Methodology

The methodology of the following literature review includes three steps: information gathering, identification of basic method features, and comparative analysis. On the basis of comparative analysis methods were objectively selected that

have a set of characteristics, allowing them to be called the best and their tests were carried out using a uniform set of data. Methods were taken from three databases: Web of Science, IEEE Explore, ACM based on a unified query performed through the advanced search tool.

The executed query included three parts. The first defined, the main idea of the search, the second, defined the embedding domain and the third, where the feature of the methods is described, which was emphasized in this review. The study was based on articles published between 2018 and 2022. Figure 2 shows the results of data collection and filtering process.

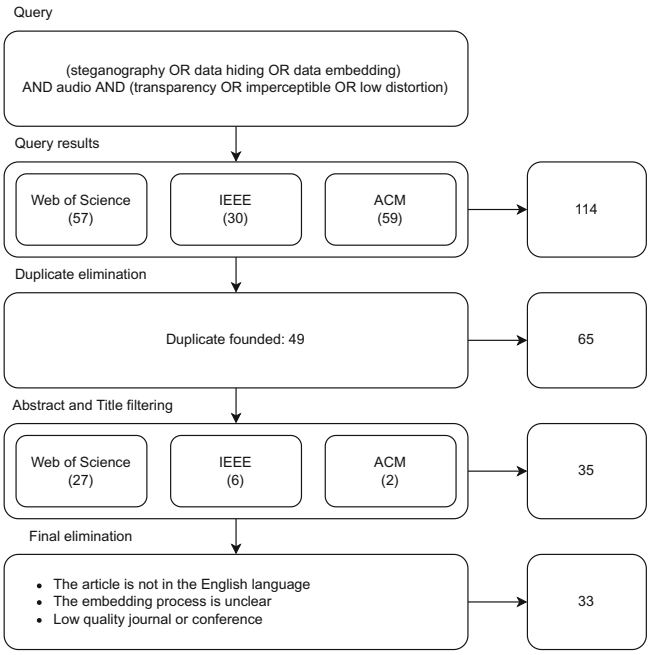

Fig. 2. Literature review methodology and filtering process

In the step where each method was analyzed, an effort was made to extract the main idea behind each method and to determine details such as the domain of embedding (DoE), the method of data injection (MODI) to be performed, the type of carrier, the secret message type, supporting techniques, and evaluation metrics. Finally, each method was classified into one of the categories that describe in general terms how the method works.

5 Review Results

The resulting literature search yielded 30 methods that were analyzed. The types of methods, the domain of embedding (DoE), and the main mode of embedding (MODI) were defined. In addition, parameters such as capacity, transparency, and robustness are taken into account. The type of media used was verified and the embedded message was characterized. Various supporting techniques and metrics used for each method were characterized.

Method types are determined by their general characteristics related to how the secret message is embedded in a cover file. The proposal of such types are described in the paper [6]. In our analysis, however, we focus primarily on those steganography methods that accord higher priority to transparency and capacity compared to robustness. Hence, among the 9 groups of methods defined in the paper [6] the first column of Table 1 presents only the 6 most prominent group of methods of embedding secret messages, whose main objective is to improve transparency followed by capacity, while keeping robustness at a desired level.

1. **Linear or sequential embedding.** The methods in this group are based on sequential, or linear access to data and then performing embedding operations. Methods that just use sequential data access are among the most commonly developed methods. Methods in this category have one significant feature - they often have high embedding rates. Speaking about the context of disadvantages, it is necessary to point out that this kind of methods is characterized by almost complete lack of resistance to typical attacks in which the signal is processed such as compression or filtering attack. Often the embedded data can succumb to simple stegoanalytic attacks.
2. **Selective embedding.** Selective methods stand, so to speak, in opposition to the linear methods described in the previous paragraph. In this type of methods, a selection of an appropriate, usually the least carrier-altering fragment is made in order to perform an embedding operation on it. The biggest advantage of this category of methods is to increase the security of the method by developing a nonlinear way to access the data during the embedding. This allows the development of methods with higher resistance to stegoanalytic attacks, but results in losses on the capacity side. Methods in this class also have high PSNR and SNR, but this may be due to the lower frequency of embedding.
3. **Frequency Masking and Amplitude Thresholding.** In this group of methods, deposition most often occurs in connection with the use of different acoustic properties of the carrier signal, moreover, different properties of the human sense of hearing are used. The biggest difference between the methods in this group and the selective methods is due to the fact that in the selective methods, the criterion for site selection is not based on a purely acoustic fact. The obvious advantage of these methods is that they have relatively high safety, resulting from the selection of sites where human hearing is less effective. However, as we can read in [6], it is necessary to test these methods against statistical tests.

4. **Error minimization embedding.** Minimization methods use techniques to minimize interference between the stegoobject and the carrier signal. This approach has the advantage of reducing the error, which directly improves the security and robustness of the stegoanalytic tests. In the context of capacity, we can talk about high variability and it depends on the method. Further research is needed on the possibility of increasing resistance to transformational type attacks like filtering or compression.
5. **Pattern - matching embedding.** The way this group of methods works is based on looking for patterns of the embedded message in the carrier signal. The patterns are in binary form. Obvious advantage of the methods used is relatively high security and transparency resulting not from the fact of embedding itself but rather appropriate "description" of carrier. This category of methods contains also a number of disadvantages, with computational and time complexity at the top.
6. **Phase coding.** The methods in this group are based on modifying the phase values of the frequency components. This is because the human sense of hearing is not immune to changes occurring in the phase domain. The biggest advantage of this type of methods is the resistance to various types of transformation attacks. This type of methods has the best balance between the three requirements to steganographic methods presented on Fig. 1.
7. **Spread spectrum.** Spectrum spreading methods, were originally developed to improve transmission quality in wireless media. The biggest advantage of this type of approach is the increased resistance to data loss during transmission, but these methods cause a lot of perceptual interference.
8. **Tone insertion.** This group of methods targets only music, as it uses specific musical elements such as drum sounds, or percussion sounds, tempo for embedding. Methods are known that take advantage of these properties and embed morse code into subtle changes in the tempo of a music track.
9. **Others.** This category includes methods that cannot be clearly assigned to any of the above categories. Very often they are single methods with a specific DoE. One such method is one that uses the encoder as its domain of operation and modification of codewords (MODI) and is shown in Table 1.

The second column of the Table 1 shows the method reference, and the publication year of the article. The third, on the other hand, describes the domain of embedding (DoE). We can distinguish between the time domain, labeled T, the frequency domain, F, and the wavelet domain, labeled W. Furthermore, some method was found whose embedding domain is changing encoding parameters, and this method is labeled as C.

The fourth column contains data about the embedding operation - MODI. For the methods emerged in this review, we can distinguish 5 modification methods:

1. LSB substitution (LSB)
2. coefficient modification (CM)
3. sample digit modification (SDM)

4. spectrum addition (SA)
5. Code word modification (CWM).

The next three columns of the table contain information about the basic parameters of each method - capacity, transparency, and robustness. Although most methods have variable capacitance we tried to evaluate the relative capacitance by averaging the values and assigning them to one of 3 groups. The group L (Low), has an average capacity between 0–250 bps, the group M (Medium) - between 250 and 750 bps, while the group H (High) has more like 750 bps.

The transparency of a given method can be expressed by a number of measures, i.e. PSNR, SNR, PESQ and others. For the purpose of this review, we defined 4 ranges of transparency values: UA (Unacceptable) when SNR is less than 20 dB, L (Low) for SNR between 20 and 40 dB, M (Medium) for 40–60 dB, and H (High) for values greater than 60 dB. The other measures, also are not without influence on the final evaluation.

The issue of method robustness is almost always problematic. Nearly 50% of methods have not been tested for robustness to typical transformational or statistical attacks. In the case of robustness, the rule of thumb is that if the method contains a robustness rating and the meaning of the message is preserved after the attack, it is assigned a rating of G (Good) and if the meaning of the message is lost, it is assigned a rating of W (Weak)

The next two columns of Table 1 contain information about the medium and the message being embedded. Within the media information, the format of the audio file used is determined. Within the context of the embedded message, it is determined whether the embedded message is audio, video or text. If the type of message being embedded is not explicitly specified, it is assumed to be bitstream.

The last two columns provide information on the various techniques supporting the method such as cryptography, compression, or scrambling, and make specifications of the methods used to evaluate the method.

According to Table 1 it is worth noting that most methods operating in the frequency domain have a high or medium level of transparency, with the time domain being dominated by methods having a low to medium level.

Figure 3 shows the ratio between the different types of methods. The largest percentage of methods are based on sequential or linear access to data, or on selective selection based on a specific access scheme. To a lesser extent, methods based on error minimization, phase coding and spectrum spreading are used. It seems interesting that sequential and selective methods appear overly frequently. Thus, we should suppose that minimization, phase coding, and spread spectrum methods seem to be interesting research directions. It has been observed that these methods have interesting properties in terms of transparency preservation with a satisfactory level of robustness.

Figure 4 presents the percentage of each embedding domain used. Significantly, the frequency domain and wavelet domain dominate among the methods found. This allows us to conclude that in methods that in their essence are supposed to provide a high transparency factor, it is these two domains that provide it.

Table 1. Review results

Method type	Ref., Year	DoE	MODI	Capacity	Transparency	Robustness	Carrier type	Message type	Supportive techniques	Evaluation features
Linear or sequential embedding	[8], 2021	T	SDM	L	H	N/A	WAV	Image	Encryption	PSNR SNR NC HC
	[16], 2021	W	CM	L	L	G	N/A	Bitstream	Scrambling	SNR BER BPS
	[3], 2020	W	CM	L	M	G	WAV	Audio	Chaotic maps scrambling	MSE SNR HC BPS NC PSNR SDG
	[44], 2020	T	SDM	L	L	N/A	N/A	Bitstream	None	ODG, SNR
	[29], 2019	W	LSB	M	H	G	N/A	Image	None	SNR SDG ODG BER, NC
	[23], 2019	W	CM	H	L	G	N/A	Audio	None	SNR, SDG, PSNR, HC
	[4], 2018	T	LSB	H	H	N/A	N/A	Audio	Chaotic maps fractal coding	SNR SDG HC PSNR
	[11], 2018	W	LSB	H	L	N/A	WAV	Bitstream	None	SNR
	[37], 2018	F	CM	M	H	N/A	WAV	Bitstream	Noise reduction	MSE, SNR, PSNR, MOS
	[30], 2018	W	CM	M	M	G	WAV	Image	None	MOS SNR NC BER
	[40], 2018	T	LSB	L	M	N/A	N/A	Bitstream	Compression encryption	PSNR MSE
	[26], 2018	C	CWM	H	H	N/A	WAV	Bitstream	N/A	PESQ
Selective embedding	[31], 2022	C	LSB	M	H	G	WAV	AMR	SIAE	PESQ, Test error rate
	[24], 2021	F	CM	M	H	G	WAV	Image	Encryption	PSNR SNR NC HC
	[34], 2021	F	CM	M	M	G	WAV	Image	Fuzzy logic, SVD	SNR BPS BER NC
	[45], 2020	F	CM	H	M	N/A	WAV	Bitstream	Scrambling	ODG BPS
	[33], 2020	T	SDM	L	M	G	MP3	Bitstream	None	ODG NC
	[35], 2019	T	SDM	M		N/A	WAV	Text	Encryption	SNR
	[32], 2019	W	CM	M	L	G	WAV	Audio, WAV	None	MSE, SNR, UACI
	[22], 2019	F	CM	N/A	M	N/A	WAV, MP3	Bitstream	None	PSNR, ODG, PEAQ
	[43], 2019	C	CWM	H	M	N/A	WAV, MP3	Bitstream	None	ODG, HC
	[27], 2018	F	CM	H	H	G	WAV	Bitstream	SVD	SNR, SDG, BPS
	[28], 2018	F	CM	H	H	G	N/A	Image	None	SNR, BER, SDG
	[9], 2018	T	LSB	M	H	N/A	WAV	Bitstream	Chaotic maps, Encryption	SNR
	[21], 2018	T	LSB	M	H	G	WAV	Image	Chaotic maps	PSNR, MSE, BER, SSIM
Error minimizing	[15], 2020	F	SA	L	L	N/A	N/A	Bitstream	None	BPS
	[38], 2019	C	CM	L	N/A	G	AMR	Bitstream	PDM-AFS pulse model	PESQ, Test Error Rate
Phase coding	[5], 2019	F	CM	M	M	NA	NA	Image	SVD	BER, SNR, ODG
	[7], 2019	F	SDM	H	L	M	WAV	Bitstream	Encryption	BER, SNR, SegSNR, Time
Spread spectrum	[14], 2019	F	CM	L	L	NA	NA	Image	SVD	BER SNR ODG
	[10], 2019	W	CM	L	L	G	MP3, WAV	Bitstream	None	SNR, ODG, BPS
Frequency masking and amplitude thresholding	[42], 2020	F	SA	L	L	W	N/A	Bitstream	None	PEAQ, ODG, SNR time
	[20], 2019	F	SA	M	M	NA	FLAC	Bitstream	None	BER

As shown in the Fig. 5, the most commonly used media type is WAV format. However, there are methods that use a hybrid approach, i.e., applicable to several media formats such as WAV, FLAC, MP3, AAC. Figure 6, on other hand, shows that over 50% of methods do not have a clearly defined type of embedded message narrowing down to simply specifying the message as a bitstream.

The metrics used to assess the quality of message embedding were also looked at carefully during the literature analysis conducted. As has been shown in Fig. 7, the most typical metric is the SNR metric, which was used in 21 articles analyzed. Next, ODG, PSNR metrics were used about 10 times. The other measures SDG, NC, MSE, MOS and PEAQ were not very popular.

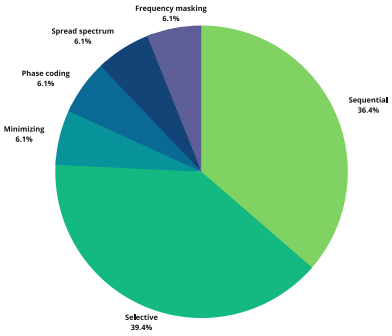

Fig. 3. Method types

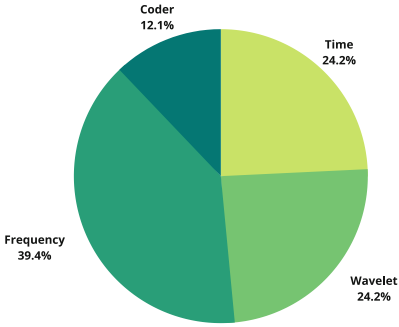

Fig. 4. Methods domains

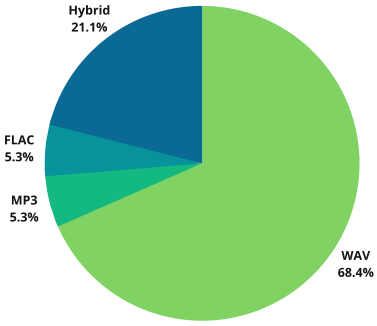

Fig. 5. Carrier types

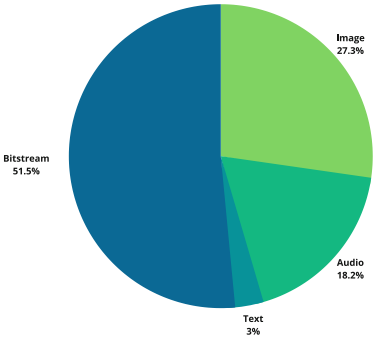

Fig. 6. Messages types

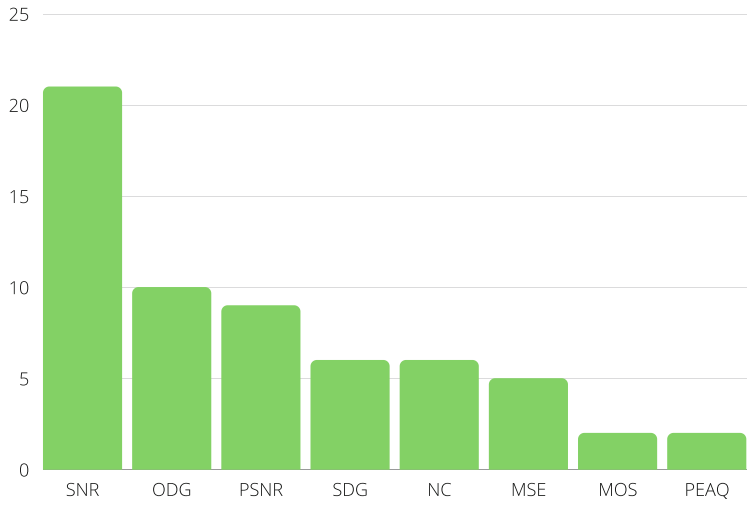

Fig. 7. Measures bar chart

6 Conclusions and Further Works

Along with cryptography, audio steganography is one of the main methods of hiding information. This paper presents an overview of steganographic methods with special emphasis on the transparency parameter. The data for this review was extracted from 3 large databases of scientific articles with unified query. Previous literature reviews have ambiguously described the various parameters of steganographic methods. In addition, this review systematically reviews metrics, for evaluating the transparency of given methods. Note the small number of publications in the categories of non-sequential and non-selective methods (compare Fig. 3). More attention to this category of methods is needed in further research. As noted, most methods operating in the frequency domain, have a higher degree of transparency than methods operating in the time domain. In the context of robustness, it is worth noting that if a method has it tested, it has good robustness.

References

1. BS.1387: Method for objective measurements of perceived audio quality. https://www.itu.int/rec/R-REC-BS.1387/en
2. P.862: Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. https://www.itu.int/rec/T-REC-P.862
3. Ali, A.H., George, L.E., Mokhtar, M.R.: An adaptive high capacity model for secure audio communication based on fractal coding and uniform coefficient modulation. Circ. Syst. Signal Process. **39**(10), 5198–5225 (2020). https://doi.org/10.1007/s00034-020-01409-7
4. Ali, A.H., George, L.E., Zaidan, A.A., Mokhtar, M.R.: High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain. Multimed. Tools Appl. **77**(23), 31487–31516 (2018). https://doi.org/10.1007/s11042-018-6213-0
5. Allwinnaldo, Budiman, G., Novamizanti, L., Alief, R.N., Ansori, M.R.R.: QIM-based audio watermarking using polar-based singular value in DCT domain. In: 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 216–221 (2019). https://doi.org/10.1109/ICITISEE48480.2019.9003921
6. AlSabhany, A.A., Ali, A.H., Ridzuan, F., Azni, A., Mokhtar, M.R.: Digital audio steganography: systematic review, classification, and analysis of the current state of the art. Comput. Sci. Rev. **38**, 100316 (2020). https://doi.org/10.1016/j.cosrev.2020.100316, https://linkinghub.elsevier.com/retrieve/pii/S1574013720304160
7. Alsabhany, A.A., Ridzuan, F., Azni, A.H.: The adaptive multi-level phase coding method in audio steganography. IEEE Access **7**, 129291–129306 (2019). https://doi.org/10.1109/ACCESS.2019.2940640
8. Altinbaş, A.E., Yalman, Y.: Bit Reduction based audio steganography algorithm. In: 2021 6th International Conference on Computer Science and Engineering (UBMK), pp. 703–706 (2021). https://doi.org/10.1109/UBMK52708.2021.9558943

9. Alwahbani, S.M.H., Elshoush, H.T.I.: Chaos-based audio steganography and cryptography using LSB method and one-time pad. In: Bi, Y., Kapoor, S., Bhatia, R. (eds.) IntelliSys 2016. LNNS, vol. 16, pp. 755–768. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-56991-8_54
10. Attari, A.A., Shirazi, A.A.B.: Robust and transparent audio watermarking based on spread spectrum in wavelet domain. In: 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), pp. 366–370 (2019). https://doi.org/10.1109/JEEIT.2019.8717415
11. Avci, D., Tuncer, T., Avci, E.: A new information hiding method for audio signals. In: 2018 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1–4 (2018). https://doi.org/10.1109/ISDFS.2018.8355361
12. Ballou, G.: Handbook for Sound Engineers. Taylor & Francis (2013)
13. Bilal, I., Kumar, R., Roj, M.S., Mishra, P.K.: Recent advancement in audio steganography. In: 2014 International Conference on Parallel, Distributed and Grid Computing, pp. 402–405 (2014). https://doi.org/10.1109/PDGC.2014.7030779
14. Budiman, G., Suksmono, A.B., Danudirdjo, D.: FFT-based data hiding on audio in LWT-domain using spread spectrum technique. Elektron. Elektrotech. **26**(3), 20–27 (2020). https://doi.org/10.5755/j01.eie.26.3.23950, https://eejournal.ktu.lt/index.php/elt/article/view/23950
15. Chen, K., Zhou, H., Li, W., Yang, K., Zhang, W., Yu, N.: Derivative-based steganographic distortion and its non-additive extensions for audio. IEEE Trans. Circ. Syst. Video Technol. **30**(7), 2027–2032 (2019). https://doi.org/10.1109/TCSVT.2019.2918511
16. Chen, S.T., Huang, T.W., Yang, C.T.: High-SNR steganography for digital audio signal in the wavelet domain. Multimed. Tools Appl. **80**(6), 9597–9614 (2021). https://doi.org/10.1007/s11042-020-09980-6
17. Dastoor, S.K.: Comparative analysis of Steganographic algorithms intacting the information in the speech signal for enhancing the message security in next generation mobile devices. In: 2011 World Congress on Information and Communication Technologies, pp. 279–284 (2011). https://doi.org/10.1109/WICT.2011.6141258
18. Djebbar, F., Ayad, B., Meraim, K.A., Hamam, H.: Comparative study of digital audio steganography techniques. **2012**(1), 25 (2012). https://doi.org/10.1186/1687-4722-2012-25, https://asmp-eurasipjournals.springeropen.com/articles/10.1186/1687-4722-2012-25
19. Dutta, H., Das, R.K., Nandi, S., Prasanna, S.R.M.: An overview of digital audio steganography. **37**(6), 632–650 (2020). https://doi.org/10.1080/02564602.2019.1699454, https://www.tandfonline.com/doi/full/10.1080/02564602.2019.1699454
20. Eichelberger, M., Tanner, S., Voirol, G., Wattenhofer, R.: Imperceptible audio communication. In: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 680–684 (2019). https://doi.org/10.1109/ICASSP.2019.8682262
21. El-Khamy, S.E., Korany, N.O., El-Sherif, M.H.: Chaos-based image hiding scheme between silent intervals of high quality audio signals using feature extraction and image bits spreading. In: 2018 35th National Radio Science Conference (NRSC), pp. 266–273 (2018). https://doi.org/10.1109/NRSC.2018.8354372
22. Garcia-Hernandez, J.J.: On a key-based secured audio data-hiding scheme robust to volumetric attack with entropy-based embedding. Entropy **21**(10), 996 (2019). https://doi.org/10.3390/e21100996, https://www.mdpi.com/1099-4300/21/10/996

23. Gupta, A., Kaur, A., Dutta, M.K., Schimmel, J.: Perceptually transparent & robust audio watermarking algorithm using multi resolution decomposition & Cordic QR decomposition. In: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), pp. 313–317 (2019). https://doi.org/10.1109/TSP.2019.8768894
24. Hameed, A.S.: A high secure speech transmission using audio steganography and duffing oscillator. Wirel. Pers. Commun. **120**(1), 499–513 (2021). https://doi.org/10.1007/s11277-021-08470-8
25. Hassaballah, M.: Digital Media Steganography: Principles, Algorithms, Advances (2020). https://doi.org/10.1016/C2018-0-04865-3
26. He, J., Chen, J., Xiao, S., Huang, X., Tang, S.: A novel AMR-WB speech steganography based on diameter-neighbor codebook partition. **2018**, e7080673 (2018). https://doi.org/10.1155/2018/7080673, https://www.hindawi.com/journals/scn/2018/7080673/
27. Kanhe, A., Aghila, G.: A DCT-SVD-based speech steganography in voiced frames. Circ. Syst. Signal Process. **37**(11), 5049–5068 (2018). https://doi.org/10.1007/s00034-018-0805-9
28. Kanhe, A., Gnanasekaran, A.: Robust image-in-audio watermarking technique based on DCT-SVD transform. EURASIP J. Audio Speech Music Process. **2018**(1), 16 (2018). https://doi.org/10.1186/s13636-018-0139-3
29. Karajeh, H., Khatib, T., Rajab, L., Maqableh, M.: A robust digital audio watermarking scheme based on DWT and Schur decomposition. Multimed. Tools Appl. **78**(13), 18395–18418 (2019). https://doi.org/10.1007/s11042-019-7214-3
30. Kaur, A., Dutta, M.K.: High embedding capacity and robust audio watermarking for secure transmission using tamper detection. Etri J. **40**(1), 133–145 (2018). https://doi.org/10.4218/etrij.2017-0092, https://onlinelibrary.wiley.com/doi/abs/10.4218/etrij.2017-0092
31. Kheddar, H., Megías, D.: High capacity speech steganography for the G723.1 coder based on quantised line spectral pairs interpolation and CNN auto-encoding (2022). https://doi.org/10.1007/s10489-021-02938-7
32. Liao, M., Dong, X., Chen, J., Zeng, D.: An audio steganography based on Twi-DWT and audio-extremum features. In: 2019 Chinese Control Conference (CCC), pp. 8882–8888 (2019). https://doi.org/10.23919/ChiCC.2019.8866035
33. Masmoudi, S., Charfeddine, M., Ben Amar, C.: A semi-fragile digital audio watermarking scheme for MP3-encoded signals using Huffman data. Circ. Syst. Signal Process. **39**(6), 3019–3034 (2020). https://doi.org/10.1007/s00034-019-01299-4
34. Mosleh, M., Setayeshi, S., Barekatain, B., Mosleh, M.: A novel audio watermarking scheme based on fuzzy inference system in DCT domain. Multimed. Tools Appl. **80**(13), 20423–20447 (2021). https://doi.org/10.1007/s11042-021-10686-6
35. Mostafa, R.M., Mohamed, M.H., Sewsey, A.A.: A hybrid system for securing data communication. In: 2019 15th International Computer Engineering Conference (ICENCO), pp. 56–61 (2019). https://doi.org/10.1109/ICENCO48310.2019.9027464
36. Noll, P.: Wideband speech and audio coding. IEEE Commun. Mag. **31**(11), 34–44 (1993). https://doi.org/10.1109/35.256878
37. Pal, D., Ghoshal, N.: Secured and imperceptible data transmission through digital audio signal with reduced internal noise. Wirel. Pers. Commun. **100**(2), 505–518 (2018). https://doi.org/10.1007/s11277-017-5095-1
38. Ren, Y., Wu, H., Wang, L.: An AMR adaptive steganography algorithm based on minimizing distortion. Multimed. Tools Appl. **77**(10), 12095–12110 (2018). https://doi.org/10.1007/s11042-017-4860-1

39. Renza, D., Ballesteros L., D.M., Lemus, C.: Authenticity verification of audio signals based on fragile watermarking for audio forensics. Expert Syst. Appl. **91**, 211–222 (2018). https://doi.org/10.1016/j.eswa.2017.09.003, https://www.sciencedirect.com/science/article/pii/S0957417417305997
40. Teotia, S., Srivastava, P.: Enhancing audio and video steganography technique using hybrid algorithm. In: 2018 International Conference on Communication and Signal Processing (ICCSP), pp. 1059–1063 (2018). https://doi.org/10.1109/ICCSP.2018.8524182
41. Torcoli, M., Kastner, T., Herre, J.: Objective measures of perceptual audio quality reviewed: an evaluation of their application domain dependence. IEEE/ACM Trans. Audio Speech Lang. Process. **29**, 1530–1541 (2021). https://doi.org/10.1109/TASLP.2021.3069302
42. Wang, S., Yuan, W., Unoki, M.: Multi-subspace echo hiding based on time-frequency similarities of audio signals. IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 2349–2363 (2020). https://doi.org/10.1109/TASLP.2020.3013785
43. Yi, X., Yang, K., Zhao, X., Wang, Y., Yu, H.: AHCM: adaptive Huffman code mapping for audio steganography based on psychoacoustic model. IEEE Trans. Inf. Forensics Secur. **14**(8), 2217–2231 (2019). https://doi.org/10.1109/TIFS.2019.2895200
44. Yu, H., Wang, R., Dong, L., Yan, D., Gong, Y., Lin, Y.: A high-capacity reversible data hiding scheme using dual-channel audio. IEEE Access **8**, 162271–162278 (2020). https://doi.org/10.1109/ACCESS.2020.3015851
45. Zhang, Z., Yi, X., Zhao, X.: An AAC steganography scheme for adaptive embedding with distortion minimization model. Multimed. Tools Appl. **79**(37), 27777–27790 (2020). https://doi.org/10.1007/s11042-020-09344-0