UDC 004.67 **doi: 10.26906/SUNZ.2024.4.132**

Roman Rastegayev[1], Vitalii Martovytskyi[1], Natalia Bolohova[1], Bohdan Filonenko[2], Oleksandr Chechui[2]

[1] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
[2] Ivan Kozhedub Kharkiv National University of the Air Force, Kharkiv, Ukraine

REVIEW OF METHODS FOR EMBEDDING DIGITAL WATERMARKS FOR AUDIO FILE PROTECTION

Abstract. The article presents an analysis of modern approaches to audio information protection using digital watermarks. It discusses various watermark embedding methods, including those based on the time, frequency, and time-frequency domains. Special attention is given to the characteristics of watermark robustness and imperceptibility, which are critical for ensuring high sound quality and reliable protection against attacks. Methods based on transformations, such as the discrete cosine transform (DCT), as well as adaptive approaches that take into account the properties of audio files, are analyzed. The article also provides an overview of criteria for evaluating the effectiveness of watermarking methods, such as signal-to-noise ratio (SNR) and detection probability. The conclusions of the study emphasize the need for careful selection of methods to achieve an optimal balance between protection, sound quality, and resistance to manipulation.

Keywords: steganography, digital watermark, information protection, audio information protection.

Introduction

The rapid development of information and communication technologies (ICT) and their convergence has led to a dramatic increase in the volume of digital content that is created, stored, distributed, and used across various fields. The term "content" in a broad sense refers to any digital information, such as audio, video, graphics, animation, images, text, or any combinations of these types. This digital content can be easily accessed, copied, quickly distributed, and widely used without quality loss, unlike the situation with earlier analog media, such as audio cassettes and VHS tapes. However, these advantages of digital media formats over analog ones turn into disadvantages in terms of copyright management, as the ability for unlimited copying without loss of authenticity has led to significant financial losses for copyright holders [1].

To reduce financial losses from unauthorized copying, content owners most often turn to cryptography, which is one of the most widely used methods of digital content protection. When cryptographic methods are used, the content is encrypted before being provided to the consumer, and then the decryption key is given only to those who have purchased legal copies of the content. However, cryptographic methods do not offer a reliable solution for combating content piracy. For example, a pirate can legally purchase the encrypted content and then use the decryption key to illegally produce and distribute copies of the content. In other words, once the content is decrypted, it no longer has any further protection.

Thus, there is an urgent need for an alternative or complement to cryptographic methods for protecting audio content. To address the issues faced by cryptography, watermarking has been proposed, as it has the potential to offer greater reliability. Watermarks can protect digital content during its normal use, as copyright information is embedded in the content in such a way that it cannot be removed. This unique feature of watermarks makes them one of the most promising methods for digital content protection, which has been a motivating factor for most research in the past two decades.

Problem statement and its connection to important scientific or practical tasks

Digital Watermarks (DW) in audio files are one of the modern technologies for copyright protection, increasingly applied in the music industry, media, and digital products. The main problem lies in ensuring reliable copyright protection while preserving the quality of audio content and maintaining resistance to various types of attacks.

n order to better understand the process of embedding watermarks in digital audio and subsequently identify some unresolved issues in current implementations, a comprehensive literature review was conducted. The algorithms examined are diverse and therefore divided into different categories, such as time-domain-based algorithms, transform-domain-based algorithms, and hybrid algorithms, according to the methodology employed by each. The advantages and disadvantages of key algorithms in each category are reviewed based on the following criteria:

– performance in terms of imperceptibility, robustness, capacity, and computational complexity: Each algorithm is evaluated based on how well it ensures that the watermark remains undetectable to the human ear (imperceptibility), its ability to withstand attacks (robustness), the amount of information the watermark can carry (capacity), and the computational resources required (complexity).

– reliability of the results presented for each algorithm: This involves examining the testing methodology used, such as the types of tests conducted (e.g., resistance to attacks, quality checks) and whether the results are reproducible and applicable to real-world scenarios.

– determining if the embedded digital watermark can be removed: This focuses on the algorithm's ability to resist removal attempts, such as through manipulation or degradation of the audio file.

– determining if the algorithms incorporate additional processes to bypass the trade-off between imperceptibility and robustness: Some algorithms may implement extra techniques or processes to balance the

need for the watermark to be both imperceptible and resilient against tampering, ensuring that neither aspect is compromised.

This comprehensive analysis aims to uncover gaps and strengths in existing approaches and point towards improvements in digital watermarking techniques for audio files.

The reason for choosing the aforementioned criteria is that they are critical factors for evaluating the effectiveness of watermarking algorithms in practical applications. Furthermore, these criteria are useful in determining whether further research on the reviewed algorithms is warranted.

Quantitative Evaluation of the Performance of Digital Watermarking Methods for Audio Files

In order for digital watermarks (DW) to effectively perform their function, it is necessary to evaluate their efficiency using quantitative methods. Such an evaluation includes analyzing various aspects, such as the imperceptibility of the watermark, its robustness against attacks, and its ability to be accurately extracted.

Let's consider the main criteria for the quantitative evaluation of the effectiveness of audio watermarks. These criteria include:

– imperceptibility (Integration): The watermark must be imperceptible to the listener, meaning it should not degrade the quality of the audio.

– robustness: The watermark should remain unchanged and recognizable even after applying various types of processing, such as compression, editing, or noise attacks.

– extractability: It is important that the watermark can be reliably extracted and recognized using specialized algorithms.

– computational efficiency: The process of embedding and extracting the watermark should be efficient in terms of computational resources and time.

The discussed criteria are key factors that influence the evaluation of watermarking algorithms and determine their practical value. Let's analyze each of these aspects, present methods for their quantitative assessment, and discuss the testing results of various watermarking algorithms using real audio files as examples.

The goal is to provide a comprehensive overview of the methods for quantitatively assessing the effectiveness of audio watermarks and to conduct a thorough evaluation of digital watermarking methods. This will contribute to the further development of copyright protection technologies and enhance their reliability in the digital environment. In general, there are three approaches to evaluating the perceptual quality of audio:

– subjective assessment through human listening tests;

– objective assessment using signal-oriented methods, such as Signal-to-Noise Ratio (SNR);

– objective assessment that incorporates a model of the human auditory system (HAS), such as Perceptual Evaluation of Audio Quality (PEAQ).

Subjective assessment can be conducted in several ways. One approach involves using the ABX test. Each test consists of the original audio file A, the watermarked audio file B, and an unknown audio file X, which can be either A or B. The listener is asked to determine whether X is A or B. A high level of correct identification indicates that the watermark is noticeable, while approximately 50% correct identification suggests that the watermark is imperceptible, as the identification resembles random guessing.

In addition, the Mean Opinion Score (MOS) can be used to evaluate the subjective quality of listening to watermarked content. The MOS is a scale that quantifies listener perceptions, where higher scores indicate better quality. The MOS rating scale is presented in Table 1.

However, subjective evaluation based on human listening tests is time-consuming, and the results may be inconsistent among different listeners. This inconsistency arises because the auditory abilities of different listeners vary depending on factors such as age, exposure to loud sounds throughout life, and even personal musical preferences. Additionally, some listeners may be trained expert listeners. Therefore, it can sometimes be challenging to fairly compare different subjective assessment results, and it is preferable to have a more objective evaluation based on specific signal characteristics.

Table 1 – **Systematization of Attacks on Watermarks**

MOS	5	4	3	2	1
Description	Excellent	Good	Satisfactory	Poor	Very Poor

The Signal-to-Noise Ratio (SNR) is widely used as an objective measure of sound quality. It is easy to interpret, straightforward to apply, and signal-oriented. According to the recommendation of the International Federation of the Phonographic Industry (IFPI), when the SNR exceeds 20 decibels (dB), audio watermarks will be considered imperceptible. SNR can be formulated as follows:

$$SNR = 10\log_{10}\frac{\sum_n s^2(n)}{\sum_n [s(n)-s'(n)]^2}, \qquad (1)$$

where *s(n)* is the time-domain original signal, and *s'(n)* is the time-domain watermarked signal.

Since equation (1) equally weights all errors in the time domain without considering the energy that varies over time and distortions that change over time, an improved estimate can be obtained by calculating the SNR for short frames and averaging the results. The frame measure, referred to as "Segmental Signal-to-Noise Ratio" (SNRseg), is defined as follows:

$$SNR_{seg} = \frac{1}{M}\sum_{j=1}^{M} 10\log_{10}\left[\sum_{n=N*(j-1)+1}^{N*j}\frac{s^2(n)}{[s(n)-s'(n)]^2}\right], \qquad (2)$$

where *M* is the number of frames, and *N* is the frame size.

Problems with SNRseg arise when including silent frames, as they can lead to large negative values for SNRseg. This issue can be addressed by setting a low threshold and replacing all frames with SNR_{seg} values below this threshold with the threshold level

(a reasonable threshold is 0 dB). On the other hand, frames with SNRseg values above 35 dB are not perceived by listeners as significantly different but still affect the resulting SNRseg. The upper threshold (typically 35 dB) can be used to cap any unusually high SNRseg values to this upper limit.

A low SNR or SNRseg clearly indicates that the distortions introduced by watermarks are audible; however, a high SNR or SNRseg is not sufficient to claim that the watermark is imperceptible, as this measure does not take into account any model of the Human Auditory System (HAS). Based on numerous experiments, the behavior of HAS has been thoroughly investigated by many researchers. These studies have made significant progress in defining the characteristics of HAS. Some terms have been proposed, such as "absolute threshold of hearing," "simultaneous masking," and "temporal masking."

The "absolute threshold of hearing" characterizes the amount of energy required for a pure tone to be recognized by a listener in a silent environment. "Masking" is the phenomenon where one sound becomes inaudible due to the presence of another sound. This phenomenon can occur in the frequency domain, known as "simultaneous masking," or in the time domain, referred to as "temporal masking." To more accurately reflect human perception, it is preferable to have an objective assessment that incorporates one of the HAS models.

PEAQ is one such objective assessment method. It has been defined as a recommended standard in BS.1387. The result of PEAQ is the Objective Difference Grade (ODG). It classifies the perceptual differences between the original audio signal and the watermarked audio signal. The ODG values range from [-4, 0], as shown in Table 2, where 0 means that both signals are perceived as identical, and -4 indicates that the differences between them are "very annoying." Thus, the closer the ODG value is to zero, the greater the likelihood that the signals are perceived as identical.

Table 2 – **Description of the ODG Indicator**

ODG	0	-1	-2	-3	-4
Description	Imperceptible	Perceptible but not annoying	Slightly annoying	Annoying	Very annoying

The correlation between PEAQ and subjective listening tests has been investigated. It was found that the correlation coefficients are 0.837 and 0.851 for the basic and extended versions of PEAQ, respectively. Undoubtedly, PEAQ cannot fully replace subjective listening tests, but it is a widely accepted objective measure of sound quality in the industry and is extensively used to evaluate the imperceptibility of watermarking algorithms. The accuracy of a watermarking algorithm is defined as the accuracy of detecting a watermark without the influence of any attack. It can be measured by the bit error rate (BER) [17], which is defined by the formulas:

$$BER(W_1, W_2) = \frac{\sum_{i=1}^{N} W_1(i) \oplus W_2(i)}{N} \qquad (3)$$

where W_1 and W_2 denote the original watermark bit sequence and the detected watermark bit sequence, respectively, N represents the number of bits, and iii denotes the bit index. In this article, "accuracy" is used to evaluate performance, as it is more straightforward. It is defined as follows:

$$\Pr ecision(W_1, W_2) = \frac{N - \sum_{i=1}^{N} W_1(i) \oplus W_2(i)}{N} = 1 - BER(W_1, W_2). \qquad (4)$$

The value of each variable is the same as in equation (3). If N audio signals are used in the experiment, the average accuracy, denoted as Precision mean, is calculated using formula (5), where \(i \) is the signal index.

$$Precision_{mean} = \frac{\sum_{i=1}^{N} Precision_i}{N}. \qquad (5)$$

The robustness of the watermarking algorithm is defined as the accuracy of watermark detection after attacks. It can also be measured using Bit Error Rate (BER). Similarly, accuracy is used to assess robustness.

Generally, to enhance robustness, a "repetition" process is included in the scheme, where the same sequence of watermark bits is repeated. On the detection side, the "mode" operation is used to identify the watermark bit sequence. In statistics, the "mode" operation is used to find the most frequently occurring data in a given dataset. For example, in the dataset {0, 1, 0, 1, 1}, the mode is "1," as it occurs one more time than "0." The procedure for incorporating the "repetition" process into the watermarking scheme can be formalized as follows:

1. Generate the watermark bit sequence B_w for the signal to be watermarked.

2. At the embedding stage, insert B_w into the signal, say d times.

3. At the detection stage, the detected bit sequence Be is divided into d groups: B_{e1}, B_{e2}, …, B_{ed}. The i-th bit of the detected watermark bit sequence B'_w is determined as the mode of the bit set { $B_{e1,i}$, $B_{e2,i}$, …, $B_{ed,i}$}.

The capacity of the watermark can be measured as the number of bits per second (bps). Assuming the duration of the audio recording is k seconds and the number of bits in the embedded watermark is n, the bandwidth is given by $n\backslash k$ bps.

Computational efficiency can be assessed as the processing time required for embedding and detecting watermarks. This depends on the implementation platform.

Overview of Audio Watermarking Algorithms

The variety of available algorithms can be categorized based on the methodology they employ. The vast majority of audio watermarking algorithms fall into three main categories:

1. Time-Domain Algorithms.
2. Frequency-Domain Algorithms.
3. Hybrid Algorithms.

Watermarking algorithms belonging to each of these three categories will be discussed in detail further.

Time-domain algorithms literally embed the watermark into the time domain. They are simple to implement. Many time-domain algorithms have been developed [19-21]. However, algorithms in this category are less robust against attacks, and statistical methods are often employed to enhance their robustness [22].

Let's consider two main algorithms in this category: the Least Significant Bit (LSB) algorithm and the Echo Hiding algorithm.

LSB (Least Significant Bit) is one of the earliest methods for embedding watermarks in audio as well as other types of digital content. The standard approach involves embedding watermark bits by altering the values of certain samples in the digital audio. The watermark bits are detected by comparing the modified sample values with the original sample values.

The primary advantage of this algorithm is its ability to achieve exceptionally high capacity. However, its main drawback is its extremely low robustness, as random signal alterations can destroy the watermark. It is very unlikely that the embedded watermark bits will survive DAC (Digital-to-Analog Converter) and subsequent ADC (Analog-to-Digital Converter) processes. Additionally, altering the quantization values introduces low-power additive white Gaussian noise, which makes this algorithm less transparent to perception, as listeners are very sensitive to this type of noise.

A significant improvement to the standard LSB algorithm was proposed in the paper [22]. The main idea is that after embedding watermark bits by manipulating a single bit of a 16-bit WAV sample, all other 15 bits of the sample can also be altered in such a way that the difference between the original sample value and the manipulated sample value is minimized. As a result, this leads to reduced distortions.

For example, if the original 16-bit sample value is "0000000000001000" in binary format, and the watermark bit to be embedded is "0," suppose the watermark bit is embedded in the fourth least significant bit of the sample. Instead of creating the value "0000000000000000" in binary, as would be expected in the standard algorithm, the improved algorithm also flips the first three bits of the sample, creating the value "0000000000000111" in binary. Consequently, the difference between the original sample and the manipulated sample is only 1 in decimal notation, which is the closest possible to the original sample value. Thus, the distortions introduced are minimized.

Echo hiding embeds watermark bits by introducing an "echo signal." An echo is a reflection of sound that reaches the listener some time after the direct sound [24]. Four parameters of the echo are used: the initial amplitude, the decay rate of the echo signal's amplitude, the "unit" shift (delay time to the original signal), and the "zero" shift. As the shift between the original and the echo signal decreases, the two signals merge. At a certain point, the human ear hears not the original signal and the echo but one mixed signal. It is challenging to determine the exact moment when this occurs, as it depends on the quality of the original recording, the type of sound being echoed, and the listener. The algorithm employs two different kernels: the "unit" kernel, which is used to generate the echo signal with a "unit" shift corresponding to a binary "1," and the "zero" kernel, which is used to generate the echo signal with a "zero" shift corresponding to a binary "0" [25].

Transformation-based algorithms typically embed watermark bits by utilizing the properties of data in the post-transformation representation. Popular transformations include the Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), and Discrete Wavelet Transform (DWT) [26-27]. Some methods, such as Quantization Index Modulation (QIM), Singular Value Decomposition (SVD), and interpolation, are often used to manipulate data to embed watermark bits in the post-transformation representation. Many watermarking algorithms fall into this category, as the embedded watermark bits are more resistant to attacks.

The HAS model is typically used to minimize the perceptual distortions introduced during watermark embedding. However, there is a trade-off involved, as embedding watermark bits into perceptually significant components is more robust but less transparent to perception. On the other hand, embedding watermark bits into less perceptually significant components is less robust but more transparent to perception. Additionally, using the HAS model increases computational time, limiting the applicability of these algorithms in time-critical applications. Typical algorithms in this category will be discussed further. The FFT was developed as a fast version of the Discrete Fourier Transform (DFT). The DFT is a well-known and powerful computational tool for performing frequency analysis of discrete time signals. It takes a discrete signal in the time domain and transforms it into a discrete frequency domain. Numerous watermarking algorithms have been proposed that are based on manipulating the components contained in the FFT spectrum. Most algorithms manipulate the magnitudes of the FFT components and enhance robustness against typical audio compression systems by incorporating the HAS model.

The scheme proposed in [27] selects a set of frequencies by comparing the FFT spectrum of the original signal with that of the corresponding compressed decompressed signal. Watermark bits are embedded at those frequencies that have similar magnitudes in both spectra. However, this selection leads to perturbations in the output signal at the most significant frequencies, which is undesirable from a perceptual transparency perspective. The scheme proposed in [28] introduces some randomness into the frequency selection process, allowing for improved transparency at the cost of some robustness. All of these schemes are not blind, meaning that the spectrum of the output signal is required to detect the embedded watermark bits.

The algorithm proposed in [29] embeds watermark bits based on spline interpolation of data obtained from the FFT transformation. The embedding process is illustrated in Fig. 1. As shown in the figure, FFT analysis is applied to each frame (i.e., short segment) of the output signal to obtain the magnitudes of the odd bits. Then, the interpolated values of the even segments are obtained through spline interpolation of the odd segment

values. The watermark bits are embedded by manipulating these spline-interpolated values of the even segments. Finally, the watermarked signal is reconstructed using the inverse FFT.

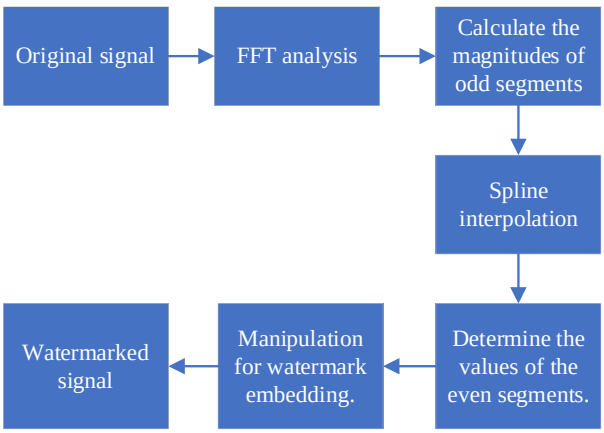

Fig. 1. Process of embedding in the algorithm proposed in [29]

As seen in Fig. 2, FFT analysis is applied to the watermarked signal to obtain the values of the odd and even segments based on each frame. Spline interpolation is then used to derive the interpolated values of the even segments. These interpolated even bit values are compared with the even bit values obtained via FFT to detect the watermark bits. The process of watermark detection is presented in Fig. 2.

This algorithm achieves a high bitrate of about 3000 bits per second and is resilient to most attacks. The average ODG score is -0.5, which is acceptable. The computational efficiency of this algorithm is high, as it only involves interpolation, FFT, and inverse FFT processes. A drawback of this algorithm is that the embedded watermark bits can be easily removed since the embedding positions are known. Additionally, since this algorithm is based solely on comparing values that can be easily disrupted, it will be vulnerable to certain attacks. Finally, since the test was based on only five songs from a single album, the assessment was limited.

This algorithm achieves a high bitrate of approximately 3000 bits per second and is resilient to most attacks. The average ODG score is -0.5, which is acceptable. The computational efficiency of this algorithm is high, as it only involves interpolation, FFT, and inverse FFT processes. A drawback of this algorithm is that the embedded watermark bits can be easily removed since the embedding positions are known. Additionally, because this algorithm is based solely on comparing values that can be easily disrupted, it will be vulnerable to certain attacks. Finally, since the test was conducted based on only five songs from a single album, the assessment was limited.

Hybrid algorithms are new algorithms, such as the Chirp coding algorithm [30], the "patchwork encryption" algorithm [31], and the SVD-based algorithm [32], that cannot be easily classified into either of the two aforementioned categories. The primary reason for highlighting this category is to emphasize their novelty. An example of one of these algorithms will be examined in detail below.

Fig. 2. The watermark detection process proposed in the article [29]

In [33], a fragile watermarking algorithm was proposed. The embedding process can be described as follows: first, a 7-level wavelet decomposition of the signal is performed to obtain 7 levels of "detail" coefficients. To measure the global effect of watermarking on the signal, the "approximation" coefficients at the 7th level are also used. Thus, a total of 8 decomposition vectors are generated. The reason for using the "detail" coefficients is that they are highly sensitive to attacks such as lossy compression and audio trimming. Then, the percentage of energy of each vector is calculated relative to the total energy of the 8 vectors. These percentages are rounded to the nearest whole number and converted into a binary stream that will be used as the watermark bit sequence.

Next, a Chirp function is created. This function is then multiplied by a new signal formed based on the binary sequence and scaled by a predetermined scaling factor to obtain the Chirp code. This Chirp code is added to the original signal to create the watermarked signal. To make the watermark inaudible, the generated Chirp code has very low frequency and amplitude.

On the detection side, the same Chirp function used during the embedding process is applied to the watermarked signal. This allows the watermark bits to be recovered. Subsequently, it can be verified whether

the signal has been tampered with by comparing the restored sequence of watermark bits with a potentially altered binary stream that can be generated directly from the watermarked signal.

In terms of this algorithm, the embedded watermarks are difficult to remove from the host signal, as both the starting and ending frequencies of the Chirp function are determined at the user's discretion, and its position in the data stream can be varied through shifting. All these parameters collectively form a private key. Listening tests have shown no perceptual difference between the original signal and the watermarked signal. The recovery of the Chirp code is uniquely robust even in cases of very low SNR, making this algorithm easily adaptable as a robust watermarking technique.

Thus, various popular algorithms developed for watermarking audio files have been discussed. To compare all these algorithms, imperceptibility is evaluated uniformly using the Mean Opinion Score (MOS), and the Objective Difference Grade (ODG) can be directly correlated with the MOS score, as shown in Table 3.

Tables 3-5 summarize the results of typical algorithms that have been reviewed. The Chirp-coding-based audio watermarking algorithm is not included in these tables because it was originally developed as a fragile watermark. However, it has significant potential for development as a robust watermarking algorithm. Table 4 presents four main characteristics of each watermarking algorithm: imperceptibility, assessed by MOS; robustness; bitrate; and computational efficiency. Table 5 lists some other characteristics of the algorithms.

Table 3 – **Relationship Between MOS and ODG**

ODG	0	-1	-2	-3	-4
MOS	5	4	3	2	1

Table 4 – **Four Main Characteristics of Each Typical Audio Watermarking Algorithm**

	LSB	**Echo hiding**	**FFT**
Imperceptibility	5	5	4.5
Robustness	Low	Low	Low
Capacity	44100	n/a	3000
Efficiency	High	n/a	High

Table 5 – **Other characteristics of each audio watermarking algorithm**

	LSB	**Echo hiding**	**FFT**
Invisibility.	+	-	+
Additional processing	-	-	-
Extractability	Easy	Difficult	Easy
Reliability	-	+	+

Tables 4 and 5 show that different algorithms have different strengths and weaknesses.

Conclusions

The article discussed different approaches to protecting audio files using digital watermarks. There are a large number of methods for applying digital watermarks, which can be classified according to various criteria: by the type of signal, the method of embedding, the place of embedding, resistance to attacks and other parameters. The most important characteristics of watermarks are their resistance to attacks (in particular, to changes and manipulations of audio files) and invisibility to the listener. This means that the watermark should not affect the sound quality, but should be strong enough to withstand various types of audio processing.

Methods that use the frequency or time-frequency domain for watermarking are often more resistant to attacks and changes compared to methods that work in the time domain. Transform-based methods such as discrete cosine transform (DCT) and others are widely used. These techniques allow watermarks to be embedded in a way that makes them less vulnerable to attack. Adaptive methods that take into account the properties of the audio file when embedding the watermark can significantly improve both the robustness and imperceptibility of watermarks.

Various metrics such as signal-to-noise ratio (SNR), detection probability, and others are used to evaluate the performance of watermarking methods. Experimental studies show that the combined use of different methods can provide better results.

The choice of a particular method depends on the specific requirements and conditions of use, such as the level of protection required, the acceptable changes in sound quality, and the expected types of attacks.

REFERENCES

1. Corporate Kaijus Clash: 15 Famous Copyright Infringement Cases [Електронний ресурс]. – Режим доступу: https://www.abounaja.com/blogs/copyright-infringement-cases
2. Zhou, N.R., Hou, W.M.X., Wen, R.H. et al. Imperceptible digital watermarking scheme in multiple transform domains. Multimed Tools Appl 77, 30251–30267 (2018). https://doi.org/10.1007/s11042-018-6128-9
3. Hosny, Khalid M., and Mohamed M. Darwish. "Invariant image watermarking using accurate polar harmonic transforms." Computers & Electrical Engineering 62 (2017): 429-447. https://doi.org/10.1016/j.compeleceng.2017.05.015
4. Tao, H., Chongmin, L., Zain, J. M., & Abdalla, A. N. (2014). Robust image watermarking theories and techniques: A review. Journal of applied research and technology, 12(1), 122-138.
5. Yuan, X. C., & Li, M. (2018). Local multi-watermarking method based on robust and adaptive feature extraction. Signal Processing, 149, 103-117. https://doi.org/10.1016/j.sigpro.2018.03.007
6. Yamni, M., Karmouni, H., Sayyouri, M., & Qjidaa, H. (2022). Efficient watermarking algorithm for digital audio/speech signal. Digital Signal Processing, 120, 103251. https://doi.org/10.1016/j.dsp.2021.103251
7. M. Torcoli, T. Kastner and J. Herre, "Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1530-1541, 2021, doi: 10.1109/TASLP.2021.3069302.
8. P. M. Delgado and J. Herre, "A Data-Driven Cognitive Salience Model for Objective Perceptual Audio Quality Assessment," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 986-990, doi: 10.1109/ICASSP43922.2022.9747064.

9. Greenspun, Philip, and Leigh Klotz. "Audio analysis VI: testing audio cables." Computer Music Journal 12.1 (1988): 58-64.
10. Streijl, R.C., Winkler, S. & Hands, D.S. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. Multimedia Systems 22, 213–227 (2016). https://doi.org/10.1007/s00530-014-0446-1
11. J. Zhang, K. Tan, J. Zhao, H. Wu and Y. Zhang, "A Practical SNR-Guided Rate Adaptation," IEEE INFOCOM 2008 - The 27th Conference on Computer Communications, Phoenix, AZ, USA, 2008, pp. 2083-2091, doi: 10.1109/INFOCOM.2008.274.
12. Deller Jr, John R., John G. Proakis, and John H. Hansen. Discrete time processing of speech signals. Prentice Hall PTR, 1993.
13. Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: A timeline of development. International Journal of Audiology, 46(9), 460–478. https://doi.org/10.1080/14992020701383019
14. Vecchi, A. O., Varnet, L., Carney, L. H., Dau, T., Bruce, I. C., Verhulst, S., & Majdak, P. (2022). A comparative study of eight human auditory models of monaural processing. Acta Acustica, 6, 17. https://doi.org/10.1051/aacus/2022008
15. Becerra Martinez, H.; Hines, A.; Farias, M.C.Q. Perceptual Quality of Audio-Visual Content with Common Video and Audio Degradations. Appl. Sci. 2021, 11, 5813. https://doi.org/10.3390/app11135813
16. Kabal, Peter. "An examination and interpretation of ITU-R BS. 1387: Perceptual evaluation of audio quality." TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University (2002): 1-89.
17. Juliy, Boiko; Andriy, Mokrytsky; Ilya, Pyatin. ДОСЛІДЖЕННЯ КІЛ СИНХРОНІЗАЦІЇ ЦИФРОВИХ СИСТЕМ ЗВ'ЯЗКУ. Хмельницького національного університету, 2022, 113. DOI 10.31891/2307-5732-2022-313-5-113-121
18. Griffiths, Dawn. Head first statistics. O'Reilly Germany, 2008.
19. Y. Xiong and Z. X. Ming, "Covert Communication Audio Watermarking Algorithm Based on LSB," International Conference on Communication Technology, ICCT 06, pp. 1-4, 2006.
20. B. S. Ko, R. Nishimura, and Y. Suzuki, "Time-spread Echo Method for Digital Audio Watermarking," IEEE Transactions on Multimedia, vol.7(2), pp. 212-221, 2005.
21. H. O. Oh, J. W. Seok, J. W. Hong, and D. H. Youn, "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking," Proc. ICASSP 2001, pp. 1341-1344, 2001.
22. Wang, Xiang-Yang, Pan-Pan Niu, and Hong-Ying Yang. "A robust digital audio watermarking based on statistics characteristics." Pattern recognition 42.11 (2009): 3057-3064. https://doi.org/10.1016/j.patcog.2009.01.015
23. N. Cvejic and T. Seppänen, "Increasing robustness of LSB audio steganography by reduced distortion LSB coding," Journal of University Computer Science, vol. 11(1), pp. 56-65, 2005.
24. N. Rashmi, "Analysis of Audio Steganography combined with Cryptography for RC4 and 3DES Encryption," 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2020, pp. 210-214, doi: 10.1109/ICISC47916.2020.9171142.
25. Gruhl, Daniel, Anthony Lu, and Walter Bender. "Echo hiding." Information Hiding: First International Workshop Cambridge, UK, May 30–June 1, 1996 Proceedings 1. Springer Berlin Heidelberg, 1996.
26. Huang, HN., Chen, ST., Lin, MS. et al. Optimization-Based Embedding for Wavelet-Domain Audio Watermarking. J Sign Process Syst 80, 197–208 (2015). https://doi.org/10.1007/s11265-013-0863-y
27. Masmoudi, S., Charfeddine, M. & Ben Amar, C. A Semi-Fragile Digital Audio Watermarking Scheme for MP3-Encoded Signals Using Huffman Data. Circuits Syst Signal Process 39, 3019–3034 (2020). https://doi.org/10.1007/s00034-019-01299-4
28. D. Megías, J. H. Joancomartí, and J. Minguillón, "Total Disclosure of the Embedding and Detection Algorithms for a Secure Digital Watermarking Scheme for Audio," ICICS 2005, pp. 427-440, 2005.
29. M. Fallahpour and D. Megas, "High capacity audio watermarking using FFT amplitude interpolation," IEICE Electronics Express, vol. 6 (14), pp. 1057–1063, 2009.
30. C. Cai, Z. Chen, J. Luo, H. Pu, M. Hu and R. Zheng, "Boosting Chirp Signal Based Aerial Acoustic Communication Under Dynamic Channel Conditions," in IEEE Transactions on Mobile Computing, vol. 21, no. 9, pp. 3110-3121, 1 Sept. 2022, doi: 10.1109/TMC.2021.3051665
31. He, J., Liu, Z., Lin, K. et al. A novel audio watermarking algorithm robust against recapturing attacks. Multimed Tools Appl 82, 18599–18616 (2023). https://doi.org/10.1007/s11042-022-14197-w
32. Abdelwahab, Khaled M., et al. "Efficient SVD-based audio watermarking technique in FRT domain." Multimedia Tools and Applications 79 (2020): 5617-5648.
33. J.Blackledge, "Digital Watermarking and Self-Authentication using Chirp Coding," ISAST Transactions on Electronics and Signal Processing, ISSN 1797-2329, vol. 1 (1), pp. 61 – 71, 2007.

Received (Надійшла) 13.09.2024
Accepted for publication (Прийнята до друку) 20.11.2024

Огляд методів нанесення цифрових водяних знаків для захисту аудіофайлів

Р. І. Растєгаєв, В. О. Мартовицький, Н. М Бологова, Б. В. Філоненко, О. В. Чечуй

Анотація. У статті викладено аналіз сучасних підходів до захисту аудіоінформації за допомогою цифрових водяних знаків. Розглянуто різні методи вбудовування водяних знаків, включаючи ті, що базуються на часовій, частотній та часово-частотній областях. Особлива увага приділена характеристикам стійкості та непомітності водяних знаків, які є критичними для забезпечення високої якості звуку та надійного захисту від атак. Проаналізовано методи на основі перетворень, такі як дискретне косинусне перетворення (DCT) також адаптивні підходи, що враховують властивості аудіофайлів. Стаття також містить огляд критеріїв оцінки ефективності методів водяних знаків, таких як співвідношення сигнал/шум (SNR) та ймовірність виявлення. Висновки дослідження підкреслюють необхідність ретельного вибору методів для досягнення оптимального балансу між захистом, якістю звуку та стійкістю до маніпуляцій.

Ключові слова: стеганографія, цифровий водяний знак, захист інформації, аудіоінформація.