Вісник ЛДУБЖД
Bulletin of Lviv State University of Life Safety
https://journal.ldubgd.edu.ua/index.php/Visnuk
ISSN 2078-4643 (print), ISSN 2708-1389 (online)

UDC 004.932: 004.056.55
DOI: 10.32447/20784643.30.2024.10

O. -S. I. Malets[1], O. O. Smotr[1], D. D. Peleshko[2], V. Pylypenko[1]
[1]Lviv State University of Life Safety, Lviv, Ukraine
[2]Ivan Franko National University of Lviv, Lviv, Ukraine
ORCID: https://orcid.org/0000-0003-2767-5019 – O. O. Smotr
https://orcid.org/0000-0003-4881-6933 – D. D. Peleshko
https://orcid.org/0009-0008-5957-4822 – V. Pylypenko
olgasmotr@gmail.com

MODERN METHODS AND PROBLEMS OF DIGITAL WATERMARKING OF AUDIO CONTENT

Introduction. Digitalization of information has greatly simplified the process of creating, distributing, and sharing audio content. However, the increased availability of digital media has led to an increase in unauthorized copying and distribution of audio files, which violates intellectual property rights. Currently, various methods are used to protect audio content, including file encryption, digital rights management (DRM), digital watermarking technologies, authentication methods, etc. Digital watermarks stand out among these technologies due to their ability to provide content tracking and allow for the identification of file origin. These methods helps you to identify the owner and protect the content even when it is further distributed. However, despite the achievements, a number of unresolved issues remains. The most significant among them are: ensuring the resistance of watermarks to various types of audio processing and high-quality hiding of watermarks, which should be invisible to listeners and preserve sound quality. The need for more reliable and versatile content protection methods remains a hot topic for research and development today.

Purpose. To investigate the current state of the digital watermarking of audio content, methods and technologies of the process of deep digital watermarking of audio content, to assess their level of quality, security and to outline promising areas of research in the field of digital watermarking of audio content.

Methods. To achieve this goal, a combination of qualitative and quantitative research methods was used. Namely, scientific research methods, in particular, a systematic review and analysis of scientific articles, patents, industry reports on practical implementations of existing digital watermarking methods, methods for analyzing the security and resilience of watermarking methods, in particular, the DCT-MLP-LSB method, benchmarking tools (AudioMarkBench) to conduct a comparative analysis of the resilience and invisibility of watermarking methods based on their overall effectiveness, a method of comparing results and a method of synthesizing results.

Results. Based on the conducted research, a comprehensive understanding of the current state of digital audio content marking is obtained, the main areas of application are considered, and promising areas of application of these technologies are highlighted. In particular, 7 main segments of digital watermarking for audio content have been identified: - authentication; - copyright; - broadcast verification; - audience measurement; - expertise; - data hiding; - communication. An analysis of the literature shows that by 2020, 80% of research in this area was conducted to watermark confidential information, and only 20% to process standard content. Today, research in these two segments is evenly distributed, and parity has almost been achieved. The rapid development of artificial intelligence algorithms causes a sharp increase in interest in watermarking standard content. The key aspects of the two main processes (embedding/extracting) of the watermarking system are investigated and schematized. The 4 key parameters that have the most significant impact on the performance of audio content watermarking systems are identified and considered: - security; - capacity; - reliability; - invisibility.

A study was conducted to evaluate the effectiveness of modern methods of digital audio content watermarking and identify gaps in existing research. Despite numerous studies in automatic sound classification, modern methods only partially reproduce human perception. As a result, these methods work well in some cases, but are practically unusable in others. We propose to replace the MLP (Multi-Layer Perceptron) model with the SVM (Support Vector Machine) model in the pipeline of the hybrid DCT-MLP-LSB algorithm in order to develop it. The main directions of promising research in the field of digital watermarking of audio content are identified.

Conclusion. The conclusions of the article demonstrate the relevance of the research topic and outline a segment of unresolved problems in the field of audio content watermarking.

Keywords: digital watermark, audio file, security, methods of audio content protection, digital watermark technologies, deep learning, neural network.

О. -С. І. Малець[1], О. О. Смотр[1], Д. Д. Пелешко[2], В. М. Пилипенко[1]
[1]Львівський державний університет безпеки життєдіяльності, Львів, Україна
[2]Львівський Національний університет імені Івана Франка, Львів, Україна

СУЧАСНІ МЕТОДИ ТА ПРОБЛЕМИ ЦИФРОВОГО ВОДЯНОГО МАРКУВАННЯ АУДІОКОНТЕНТУ

Проблема. Цифровізація інформації значно спростила процеси створення аудіоконтенту, його поширення та обміну ним. Проте зростання доступності цифрових медіа призвело до збільшення випадків несанкціонованого копіювання та розповсюдження аудіофайлів, що порушує права інтелектуальної власності. Наразі для захисту аудіоконтенту використовуються різні методи, зокрема шифрування файлів, управління цифровими правами (DRM), технології цифрових водяних знаків, методи аутентифікації тощо. Цифрові водяні знаки виділяються серед цих технологій завдяки здатності забезпечувати слідкування за контентом і дозволяти ідентифікацію походження файлів. Такі методи надають змогу визначати власника та захищати контент навіть при його подальшому розповсюдженні. Однак, незважаючи на досягнення, залишається ряд невирішених проблем. Найбільш вагомими серед них є: забезпечення стійкості водяних знаків до різних видів аудіообробки та якісне приховування водяних знаків, які повинні бути непомітними для слухачів і зберігати якість звуку. Потреба в більш надійних та універсальних методах захисту контенту залишається актуальною темою для наукових досліджень і розробок сьогодення.

Мета. Дослідити сучасний стан галузі цифрового нанесення водяних знаків на аудіоконтент, методи та технології процесу глибокого цифрового водяного маркування аудіоконтенту, оцінити їх рівень якості, безпеки та окреслити перспективні напрямки досліджень в царині цифрового водяного маркування аудіоконтенту.

Методи. Для досягнення цієї мети, в роботі була використана комбінація якісних і кількісних методів дослідження. А саме, методи наукового дослідження зокрема, систематичний огляд та аналіз наукових статей, патентів, галузевих звітів щодо практичних реалізацій існуючих методів цифрового нанесення водяних знаків, методи аналізу безпеки та стійкості методів нанесення водяних знаків, зокрема метод DCT-MLP-LSB, інструментарій бенчмаркінгу (AudioMarkBench), для проведення порівняльного аналізу стійкості та непомітності методів нанесення водяних знаків на основі їх загальної ефективності, метод співставлення результатів та метод синтезу результатів.

Основні результати дослідження. Базуючись на проведених дослідженнях отримано всебічне розуміння поточного стану цифрового маркування аудіоконтенту, Розглянуто основні сфери застосування та виділено перспективні напрямки застосування цих технологій. Зокрема виокремлено 7 основних сегментів використання цифрових водяних знаків для аудіоконтенту: - автентифікація; - авторське право; - верифікація трансляції; - вимірювання аудиторії; - експертиза, - приховування даних; - спілкування. Аналіз літературних джерел засвідчує, що до 2020 року 80% досліджень у цій галузі проводились з метою водяного маркування конфіденційної інформації, і лише 20% для обробки стандартного контенту. На сьогодні дослідження у цих двох сегментах розподілені рівномірно, практично досягнуто паритету. Таке різке збільшення інтересу до водяного маркування стандартного контенту, спричинене стрімким розвитком алгоритмів штучного інтелекту. Досліджено та схематично відображено ключові аспекти двох основних процесів (вбудовування / вилучення) системи водяних знаків. Виокремлено та розглянуто 4 ключові параметри, що мають найбільш вагомий вплив на продуктивність систем нанесення водяних знаків на аудіоконтент: - безпека; - ємність; надійність; - непомітність.

Проведено дослідження оцінки ефективності сучасних методів цифрового водяного маркування аудіоконтенту та виявлено прогалини в існуючих дослідженнях. Зокрема, незважаючи на численні дослідження в галузі автоматичної класифікації звуків, сучасні методи лише частково відтворюють людське сприйняття. Як наслідок, ці методи добре працюють в одних випадках, та практично непридатні для використання у інших. Ми пропонуємо в пайплаймі гібридного алгоритму DCT-MLP-LSB з метою його розвитку замінити моделі MLP (багатошаровий перцептрон, Multi-Layer Perceptron) на модель SVM (метод опорних векторів, Support Vector Machine). Визначено основні напрямки перспективних досліджень у царині цифрового водяного маркування аудіоконтенту.

Висновки. Висновки статті засвідчують актуальність теми дослідження та окреслюють сегмент невирішених проблем в царині водяного маркування аудіоконтенту.

Ключові слова: цифровий водяний знак, аудіофайл, безпека, методи захисту аудіоконтенту, технології цифрових водяних знаків, глибинне навчання, нейромережа

Introduction

Information is one of humanity's most treasured assets, representing knowledge in various forms. The emergence of digital technology has facilitated numerous affordable and user-friendly methods for sharing ideas and exchanging information. However, this rapid shift to digital formats has also led to a surge in unauthorized copying, which infringes upon

intellectual property rights. To address these challenges, embedding digital watermarks in content can provide crucial details for tracking usage and enforcing copyright.

Digital watermarking not only helps protect copyrighted materials but also serves as a versatile framework for integrating information into different types of data for diverse applications. In this paper, multiple digital watermarking applications were discussed, dividing them into categories of security and non-security. Additionally, we would go through two distinct watermarking techniques designed for handling standard content as well as sensitive information, such as political news, audio recordings, and confidential communications. Each technique is specifically crafted to address the needs of security-oriented and non-security contexts.

Terminology and key concepts.

The watermarking system is fundamentally divided into two main processes [1]: embedding and extraction. These processes typically employ a cryptographic key, which may either be a public or a private key. The watermark functions as a concealed signature embedded within the original digital content. The final output, known as watermarked documents, is produced by layering the original content with the hidden signature. The embedding process and extraction process are schematically illustrated below.

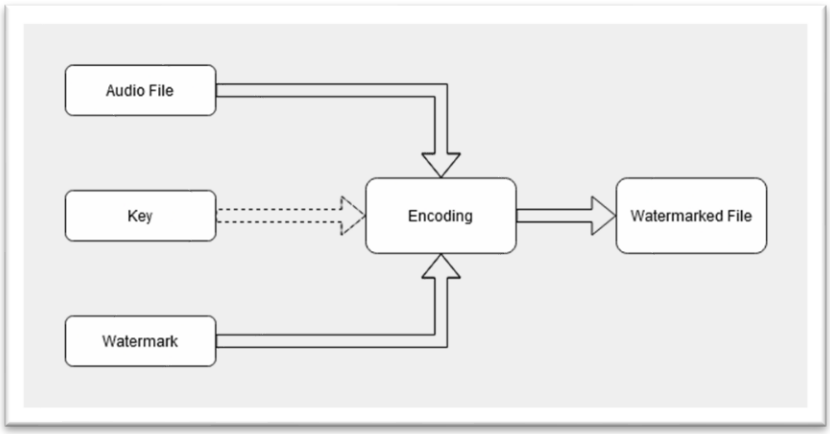

Figure 1 – Embedding process

Watermark, original digital content, and sometimes the keys were set as the inputs to the embedding process. One basic requirement to differentiate between watermarking techniques is the insertion domain [2]–[5]: insertion domain with no transformation, frequency domain, and multi-resolution domain.

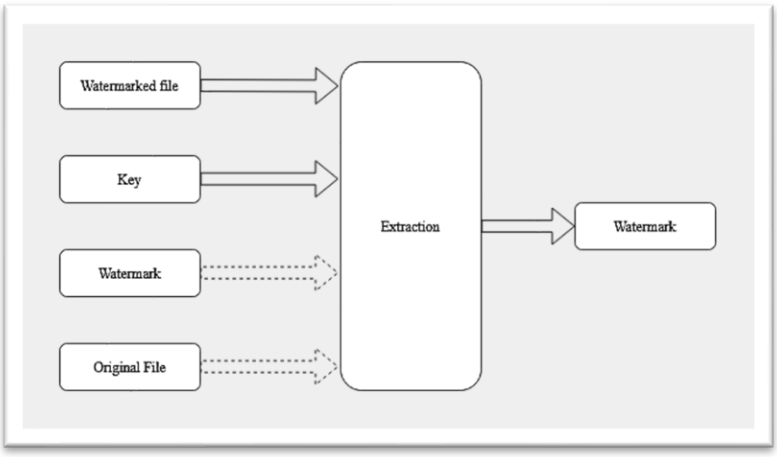

Figure 2 – Extraction process

If the original document is not needed for the watermark detection process, the watermarking scheme is classified as blind; if the original is required, it is considered non-blind.

The effectiveness of watermarking systems is determined by several key properties: - imperceptibility; - robustness; - security; - capacity.

Imperceptibility - this is a crucial criterion in digital watermarking. The problem is that, in an effort to achieve high imperceptibility, developers may reduce the amount of information that can be encoded (capacity) or use less complex methods that are easier to hide. This reduces the capacity and reliability of the sign, as the information can be lost during audio

compression or processing. While many watermarking techniques aim for imperceptibility, some methods intentionally embed perceptible watermarks.

Robustness - refers to the watermark's ability to withstand various attacks and modifications. Generally, robust watermarks are preferred, although there are instances where we may need to process a watermarked document to completely remove the signature. In some cases, a semi-fragile watermarking approach is desirable, allowing the watermark to survive certain alterations while being lost under others. To achieve high reliability, it is usually necessary to embed the sign in more robust parts of the signal, which often reduces the invisibility and can increase the capacity. However, increased robustness usually reduces invisibility, as a robust sign is easier to spot and potentially identify.

Capacity indicates the maximum amount of data that can be embedded within a digital document. A higher capacity is often important, as many applications require substantial payloads. The more information a watermark contains, the more opportunities there are for authentication and content tracking. However, increasing the capacity often makes the watermark more visible and increases the likelihood that it can be detected and removed. In this way, capacity conflicts with invisibility and security, as a visible mark is easier to find and remove, which reduces its security.

Security ensures that only authorized users can extract the watermark, which is essential for effective copyright protection. High security requires complex coding, which can be less invisible, and an increase in the amount of data in the mark, which reduces capacity and can affect invisibility. Thus, the desire for security often conflicts with the requirements for capacity and invisibility.

Analysis of recent research and publications

Historically, digital watermarking technologies have been mostly developed and applied to protect confidential and valuable information, especially in the security, intellectual property, military, and government projects. An analysis of the literature shows that until 2020, 80% of research in this area was conducted to watermark confidential information, and only 20% to process standard content. Since 2020, the situation has changed dramatically: with the development of digital technologies and the proliferation of digital media, watermarks are increasingly used to protect standard content - music, movies, podcasts - and to support intellectual property rights in the public domain. Today, research in these two segments is evenly distributed, with almost parity achieved.

This can be explained by the fact that the development of reliable, secure, and invisible watermarks required significant resources and sophisticated technologies, so such solutions were mostly relevant in industries where information protection is crucial (confidential information). The use of watermarks in public media and for protecting standard content began to develop later, when the technology fell in price and became more affordable for media companies, rights holders and content streaming platforms.

Let's take a closer look at the range of applications of digital watermarks for audio content:

1. Data Hiding [6]: This method embeds data in audio content to ensure that it remains hidden from unauthorized parties. It’s often used in contexts where secrecy is paramount, making it difficult for unauthorized users to detect or extract the hidden data.

2. Communication [7]: Primarily applied in military and intelligence contexts, steganography is used to send covert messages without detection. By embedding these messages within audio signals, covert communications are made more secure and less likely to be intercepted.

3. Ownership [8]: Enables lawful embedding of ownership credentials within digital content, preventing unlawful alteration and asserting ownership. It acts as proof of authenticity, allowing owners to establish their rights over the content if challenged.

4. Authentication [9]: Audio watermarking that confirms authenticity by embedding a hidden signature that verifies the content's originality. If tampering occurs, the altered watermark reveals the modification, thus helping to detect any interference and ensuring the audio’s integrity.

5. **Forensics** [10]: This method enhances the owner’s ability to identify and respond to unauthorized use of their content. It is used not only to collect evidence against violations but also to support adherence to usage agreements between the content owner and users.

6. **Broadcast Verification** [11]: Broadcast verification involves embedding a watermark in digital content, such as radio ads, to track its broadcast details. This ensures advertisers can confirm that their ads are played the agreed number of times and on the correct stations. For example, watermarks help verify the station, frequency, and time of each broadcast, supporting accurate compliance with contractual obligations.

7. Audience Measurement [12]: In audience measurement, watermarking embeds a unique identifier in digital content, enabling broadcasters to track viewership across channels. Audiometers in households detect this embedded data, collecting information such as channel, exposure time, and program details. This data is then transferred to a central database, allowing for accurate, consistent daily reporting of audience metrics across multiple platforms.

These applications make watermarking essential for managing digital rights, and enabling secure and verifiable communication across various fields.

With the significant development of artificial intelligence algorithms and their availability, more and more researchers propose to use the capabilities of neural networks in audio content watermarking systems [13-24]. For example, Pengcheng Li, Xulong Zhang, Jing Xiao, and Jianzong Wang in their study [13] are working on developing a double-embedding watermarking model for efficient watermarking and improving attack resistance by studying the effect of the attack level on the inverted neural network during reliability training. However, the problem of how to provide greater resilience to new types of attacks and maintain high audio quality under different processing conditions remains. Paper [14] describe the use of a deep neural network and perceptual losses to embed watermarks, taking into account psychoacoustic effects. In [15], the authors use the characteristics of the human auditory system (HAS) and neural networks in the field of discrete cosine transform (DCT) for imperceptible watermarking. The issue of achieving a balance between the imperceptibility and robustness of watermarks at different levels of audio compression and processing remains open.

Paper [16] provides an overview of various watermarking methods using deep neural networks, including a new taxonomy and examples of methods. However, the lack of a unified methodology for evaluating the effectiveness of different watermarking methods makes standardization essential.

Research [17] proposes to use of the convolutional neural networks for increasing the resistance of watermarks to various attacks. Paper [18] presents a countermeasure learning method to improve watermarking resilience to attacks. It should be noted that these methods will not work with new, unpredictable types of attacks.

In [19], Arjon Das and Xin Zhong presented a watermarking scheme for copyright protection by embedding watermarks in audio content based on deep learning. The authors failed to ensure high watermark robustness without losing audio quality and reducing computational complexity.

Most of the existing deep watermarking algorithms use an encoder-decoder architecture, which is redundant. In [20], David Megías proposes an audio watermarking/extraction algorithm based on adversarial perturbation, AAW. It adds tiny, imperceptible perturbations to the underlying audio and extracts the watermark using a pre-trained decoder. However, this system remains unstable to signal changes and requires optimization of recurrent neural networks for different types of audio content.

Jiaming Hu1, Boon-Han Lim and others in [22] propose a hybrid approach that combines different types of neural networks to improve the performance of watermarking in audio content.

All of these studies show significant progress in the use of neural networks for audio content watermarking, but there are still issues that require further study and optimization.

Methods

A number of studies suggest using human auditory perception to improve automatic classification systems. The authors present a model of auditory memory and psychoacoustic features, combining them with deep learning, in particular deep neural networks (DNNs), to classify audio content. The classified information is then embedded in the audio using a watermarking technique.

Sound signals in indexing applications vary widely, from music to speech, requiring different analysis techniques based on the nature of the audio. For instance, while speech analysis focuses on phonemes that last milliseconds, music genres require longer durations for accurate classification. Despite extensive research on automatic sound classification, current methods only partially replicate human perception. As a result, these methods work well in some cases but fall short in others.

Systems for automatic classification of sound into categories such as music, speech, gender, and emotion typically embed a watermark using the DCT (Discrete Cosine Transform) [24], MLP (Multi-Layer Perceptron) [25] and LSB (Least Significant Bit) [26].

The system usually works in the following stages:

1. Audio preprocessing: audio extraction and feature extraction.

2. Transformation using DCT: extracting important components of the audio signal.

3. Watermark embedding using LSB: embedding the watermark in low-value bits while maintaining invisibility.

4. Analysis and classification using MLP: A neural network processes features to categorize audio.

For a better understanding of the DCT -MLP - LSB method, we have displayed the general stages of the watermark embedding and audio classification process in Figure 3.

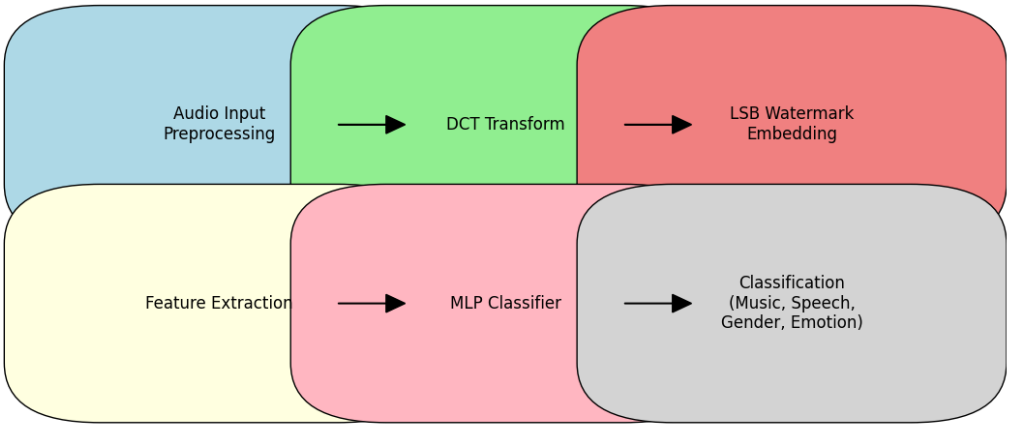

Figure 3 – Steps of the DCT -MLP - LSB method.
Source: author's development.

The experimental results [23] demonstrate the system's effectiveness in both public datasets and watermark robustness. For non-security applications, robustness to intentional attacks isn't critical, but resistance to signal processing like compression is necessary. The watermark, in this case, holds metadata such as artist information or registration location, aiding in signal indexing.

In summary, the technique not only improves audio classification but also embeds essential metadata for efficient audio management. However, it is worth noting that it would be advisable to take into account such categories as age and national identity when classifying audio, as they also have a significant impact on the rhythm, timbre and modulation of the voice.

Given that the effectiveness of the audio content watermarking system largely depends on the nature of the sound, a detailed study of the key stages of the audio content characterization and classification process was conducted.

Audio Feature Extraction. Feature extraction is crucial for machine learning and pattern recognition tasks. To achieve better results, extracted features should be tailored to the specific application. Typically, audio files are divided into overlapping windows, and descriptors are calculated for each frame. Then, statistics are gathered over longer, mid-term windows. This creates two levels of processing: short-term and mid-term.

The purpose of feature extraction is to identify a set of characteristics from the dataset, often as a means to reduce data complexity. Directly processing the original audio signal is challenging, so this step is necessary to reduce data while maintaining accuracy. Selecting the appropriate set of features for a particular application is critical to ensure high performance and reliable classification outcomes.

Short-term Analysis. In short-term, or frame-based, audio processing [27], the audio signal is divided into overlapping frames of 10–50ms duration, during which the signal is assumed to be stationary. Descriptors are extracted and computed within each frame. To avoid discontinuities at the frame boundaries, windowing techniques like the Hamming window are applied.

Once windowing is complete, features are calculated per frame. These features can be categorized into time-domain and frequency-domain (spectral) features.

1. *Temporal Audio Features*: These are derived directly from the raw audio samples. Key features include:

- Short-term energy: Measures the energy in each frame.
- Energy entropy: Captures the variability of energy distribution.
- Zero-crossing rate: Counts how often the signal changes sign, providing insight into signal complexity.

These time-domain features provide a straightforward way to analyze audio signals and are part of technique’s feature extraction phase.

Spectral Audio Features: Combining time-domain features with spectral features enhances the accuracy of audio analysis. Spectral features are computed using Discrete Fourier Transform (DFT) coefficients. Prominent spectral features include:

- Spectral flux: Measures the rate of change in the spectrum.
- Spectral centroid: Represents the "center of mass" of the spectrum, often linked to the perceived pitch.
- Spectral roll-off: Indicates the frequency below which a certain percentage of the spectral energy is contained.

– Mel-Frequency Cepstral Coefficients (MFCCs) [28]: Capture the shape of the audio spectrum in a way that reflects human hearing.

– Chroma vector: Represents the energy distribution of pitch classes.

– Relative Spectral Analysis-Perceptual Linear Prediction (Rasta PLP) [29]: A feature extraction method that improves robustness to noise and channel variations.

These combined features are critical for effective audio classification, ensuring both time and frequency information are captured during analysis.

Mid-term Analysis. Following the short-term or frame-based analysis, mid-term statistics are calculated on longer windows of approximately 1 to 10 seconds. While frame-based processing is often favored in speech analysis, longer-term windows are necessary to capture the semantic meaning of the audio signal. In mid-term analysis, the audio is divided into clips, each comprising several consecutive frames. For each clip, key statistics such as mean value, standard deviation, skewness, and kurtosis are computed from the extracted short-term features.

These statistics are initially computed separately and then combined using feature fusion to improve the overall representation of the audio signal.

Deep Learning-Based Audio Classification. This system employs Deep Neural Networks (DNNs) for audio classification, using a Multilayer Perceptron (MLP) architecture. In this configuration content characterization handles three key tasks: music vs. speech discrimination, speaker gender recognition, and speech emotion identification. The model utilizes categorical cross-entropy as its loss function and softmax as the activation function in the final dense layer.

Watermarking Technique for Audio Content Characterization. The watermarking technique used, DCT-MLP-LSB, is detailed in Figure 4. This method classifies and embeds information about the audio content (e.g., music, speech, speaker gender) into the signal. First, the audio file is divided into fixed-length segments, each analyzed and classified. The retrieved content information is encoded into a binary vector, where, for instance, "0" could represent music and "1" speech. To improve robustness, a Hamming encoder (8,12) is applied.

The audio signal is further divided into 512-sample blocks, each transformed using Discrete Cosine Transform (DCT). The middle-frequency band is selected for embedding, and the Least Significant Bit (LSB) in this band is replaced by the watermark. The process is repeated for each block along the stream, ensuring that each watermarked audio segment carries its classification data (e.g., music, speech, speaker emotion).

During extraction, the process is reversed. The audio is divided, DCT is applied, and the LSB of each block is analyzed to reconstruct the watermark, which is then decoded to reveal the classification of the audio content.

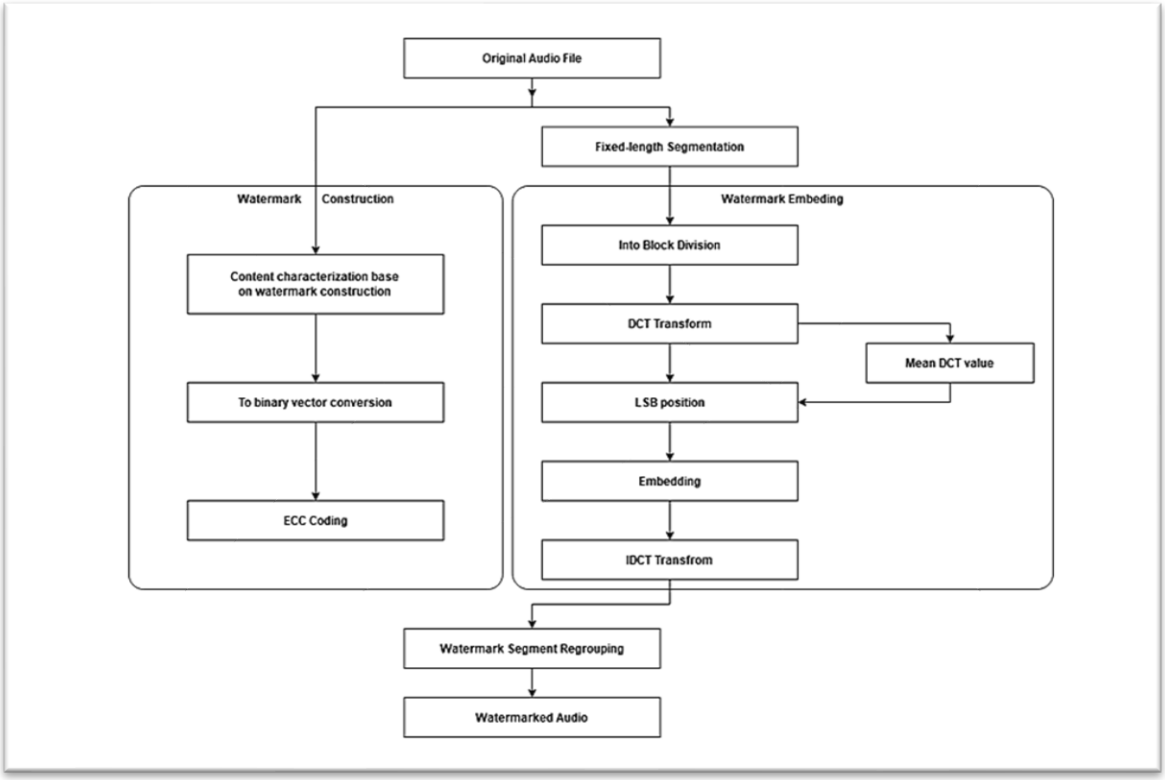

Figure 4 – Watermarking scheme of audio signal

Source:author's development (based on the analysis and generalization of the researched methods)

In our opinion, the most comprehensive and well-documented experimental studies on digital audio watermarking methods to date are those conducted by Hongbin Liu, Moyang Guo, Zhengyuan Jiang, Lun Wang, and Neil Zhenqiang Gong [33], Guangyu Chen, Yu Wu, Shujie Liu, Tao Liu, Xiaoyong Du, and Furu Wei [34], as well as Charfeddine, M., Mezghani, E., Masmoudi, S., and Ben Amar, C. [23]. These three studies cover key aspects of watermarking, ranging from the first systematic approach to testing (StirMark) to modern performance evaluation methods (AudioMarkBench) and applied uses of the technology. All three works emphasize the importance of balancing security, robustness, capacity, and imperceptibility, as well as the need for unified testing approaches and the development of algorithms capable of withstanding complex modern attacks.

Study [32] introduces StirMarkBenchmark, a set of tests designed to evaluate the robustness of audio watermarks against various attacks. The authors thoroughly examined the effects of standard attacks, such as compression, cropping, noise addition, playback speed changes, and resampling. The study aimed to assess the reliability of watermarks under real-world conditions. The primary contribution of this work lies in the creation of a systematic testing approach, which has since become a foundation for many subsequent studies in digital watermarking. The study highlighted the critical need to design algorithms that remain resilient even after aggressive audio processing.

In study [33], the authors propose a novel platform, AudioMarkBench, for assessing the robustness of audio watermarking algorithms. The focus is on modern attacks, including those enabled by advancements in machine learning, such as generative noise algorithms. The researchers compared popular watermarking methods, including DCT, LSB, and wavelet transformation, based on metrics such as robustness, imperceptibility, and capacity. The study evaluates three state-of-the-art watermarking methods against 15 types of perturbations under no-box, black-box, and white-box conditions. Testing was conducted using 18 NVIDIA RTX-6000 GPUs, each with 24 GB of memory, and the full suite of experiments required approximately 430 GPU-hours to complete.

Special recognition is due to the authors of this study for making the AudioMarkBench tool publicly available, thereby contributing to the standardization of watermark evaluation. Additionally, the authors released the dataset "audiomarkdata_20k.csv," used in the AudioMarkBench testing. We had the opportunity to analyze this dataset, and the results of the analysis are presented in Figure 5.

	name	age	gender	language
count	20000	20000	20000	20000
unique	20000	4	2	25
top	audiomarkdata_eu_21689570.mp3	teens	female	Basque
freq	1	5000	10000	800

Figure 5 – Characteristics of the "audiomarkdata_20k.csv" dataset.
Source: author's development

The dataset "audiomarkdata_20k.csv" contains a total of 20,000 audio recordings representing four age groups of various genders, recorded in 25 different languages. Each age group is represented by 5,000 recordings, and each language is represented by 800 recordings.

The findings highlighted in the study emphasize that while many modern algorithms perform well across individual metrics, they rarely achieve a balance among all key parameters.

In study [23], contemporary methods for digital audio watermarking and their applications, such as content protection, audience measurement, and copyright management, were analyzed. The primary focus is on balancing security, robustness, imperceptibility, and capacity—critical requirements for watermarking systems. The authors also examined the role of modern algorithms, including hybrid approaches (e.g., DCT+LSB) and artificial intelligence-based methods. The study provides a detailed review of common attacks, such as compression and alterations of time-frequency characteristics, and analyzes their impact on the performance of watermarking systems.

Results.

Through the combinatorial application of qualitative and quantitative research methods, result comparison techniques, and result synthesis, our study has achieved a comprehensive understanding of the current state of digital watermarking for audio

content. Key application areas were analyzed, and promising directions for the use of these technologies were identified. Specifically, seven primary segments for the application of digital watermarks in audio content were outlined: authentication, copyright protection, broadcast verification, audience measurement, forensic analysis, data hiding, and communication.

A review of the literature indicates that, as of 2020, 80% of studies in this field focused on watermarking confidential information, while only 20% addressed standard content. Today, research in these two segments is evenly distributed, reaching near parity. This sharp increase in interest in watermarking standard content is driven by the rapid development of artificial intelligence algorithms.

We have examined and schematically represented the key aspects of the two main processes (embedding and extraction) in watermarking systems. Four critical parameters that significantly influence the performance of audio watermarking systems were identified and analyzed: security, capacity, robustness, and imperceptibility.

The efficiency of modern methods for digital watermarking of audio content was evaluated, and gaps in existing research were revealed. Despite numerous studies in the field of automatic sound classification, contemporary methods only partially replicate human perception. Consequently, we identified limitations in existing methods, particularly regarding the incomplete representation of human auditory perception in classification algorithms and the challenge of balancing core characteristics (security, capacity, robustness, imperceptibility). These limitations highlight the need for further research.

In order to develop the considered DCT-MLP-LSB algorithm, we propose to replace the MLP (Multi-Layer Perceptron) model in the DCT-MLP-LSB algorithm pipeline with the SVM (Support Vector Machine) model.

The study's findings demonstrate that the flexibility of methods and their adaptability to various scenarios are critical for their widespread adoption.

Discussion:

The first scheme focuses on security applications, particularly for copyright protection. By embedding a signature in the mid-frequency range of an audio frame, combined with neural network (NN) techniques for insertion and detection, it enhances both security and robustness. The integration of frequency perceptual masking (HPM) ensures the watermark remains hidden without noticeable degradation in audio quality. The experimental results demonstrated that this scheme provides improved robustness compared to traditional NN-based and other watermarking techniques, meeting the standards set by the IFPI for imperceptibility and security.

The second proposed scheme addresses non-security applications, focusing on audio content characterization. Here, the watermark contains information about the audio's characteristics (e.g., speech segments, speaker emotion, etc.), allowing users to navigate the content based on specific criteria. By combining a deep learning framework with the Discrete Cosine Transform (DCT), the scheme enables efficient classification and watermarking with high performance.

As a result of the study, we can state that in the process of watermarking/extracting audio signals using the methods described above, a number of risks may arise related to the deterioration of key parameters of audio content watermarking systems. First of all, this is the risk of changing the audio quality, i.e. the risk associated with perceptibility. After all, replacing the least significant bits in the mid-frequency band can degrade the sound quality, especially if the audio signal is compressed (for example, MP3) or has a high dynamic complexity. Another risk is related to perceptibility - the risk of detecting a watermark. After all, if changes become noticeable to the listener, it can reduce the value of the audio and draw attention to the watermark.

The second group of risks is related to robustness. First of all, this is the risk of lossy compression, which can damage the mid-frequency band, leading to the loss of part or all of the watermark. Another is the risk of noise. Adding noise (either background or due to processing) can change the LSB and make the watermark unrecognizable. The third risk of the robustness group is cropping or resampling. Changing the length of the audio file or its sampling rate can disrupt the distribution of blocks and make it difficult to extract the watermark.

The third group of risks includes risks related to capacity. These are the risks of limiting/overloading information. Due to the small number of bits in the LSB of the mid-frequency band, the capacity of the watermark is limited, which may not allow storing a sufficient amount of classification data. If the information content of the watermark exceeds the available capacity, this may affect the stability and accuracy of extraction.

The fourth group of risks is security risks. First of all, this is insufficient data protection: If classification data is not encrypted before embedding, attackers can access and manipulate the watermark. This is an attack vulnerability.

The LSB method, which is often used today and is simple and quite effective, is vulnerable due to its predictability and susceptibility to signal manipulation, such as noise contamination, compression, or sample rate conversion. It is less

secure and can be easily removed or modified if an attacker knows how it is implemented. Therefore, it is more appropriate to use hybrid methods, such as DCT + LSB or DWT + LSB**.** Such approaches combine the advantages of different methods, increasing the robustness and security of the watermark or methods that use artificial intelligence (e.g., neural networks to generate or recognize watermarks). However, these methods are more difficult to reverse engineer. In addition, there is always a risk of neural network error when using artificial intelligence tools. A machine learning classification model (MLP) may produce incorrect results due to insufficient training or incorrectly selected features. If the classification depends on the characteristics of the mid-frequency band, embedding a watermark can affect the classification result.

We also believe that in order to develop the algorithm under consideration, it would be advisable to replace the MLP (Multi-Layer Perceptron) model in the DCT-MLP-LSB algorithm pipeline with the SVM (Support Vector Machine) model. In our opinion, replacing MLP with SVM is justified by the fact that the audio signal may contain significant fluctuations (changes in the time or frequency domain, such as noise, compression, clipping, etc.) SVM is better at dealing with such changes because SVM focuses on key features, ignoring secondary fluctuations, and is focused on finding the best possible data separation in a multidimensional space. In addition, compared to neural networks, SVMs often demonstrate more stable results with a limited amount of training data. All in all, this will increase the algorithm's resistance to attacks and improve the accuracy of watermark extraction.

After analyzing the key properties of an audio watermarking system, we can observe that balancing these characteristics to achieve system efficiency is a challenging task. The nature of the interdependencies among these parameters is illustrated in Figure 6. Efforts to ensure security often conflict with the requirements for capacity and imperceptibility.

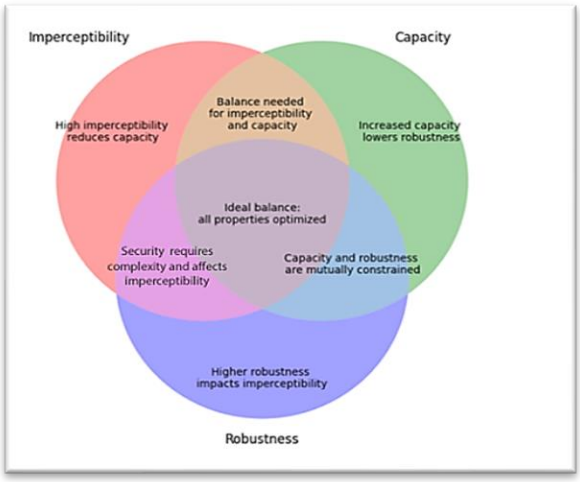

Figure 6 – Interdependencies in digital watermarking properties.
Source: author's development.

To avoid the above risks, it is possible to optimize the invisibility by using perceptual models to select the least significant bits in the range that do not affect the sound quality. In order to increase the reliability of the audio content watermarking system includes an error correction code (e.g., Reed-Solomon or Hamming) to increase resistance to distortion. The watermark is encrypted using cryptographic encoding to protect the content from attacks. It is also necessary to use dynamic algorithms that take into account the specifics of the audio signal. This will ensure the adaptability of the system. In addition, thorough testing is required on a regular basis. The system is tested on a variety of audio files and scenarios, including noise, compression, and cropping.

Therefore, we can state that the biggest challenges and problems in the field of audio content watermarking are

- ensuring that watermarks remain intact after various signal processing operations such as compression, filtering, and resampling, which is a significant challenge;
- watermarks need to be invisible to listeners to avoid degradation of sound quality, which is a delicate task to strike a balance;
- an ongoing challenge to increase the amount of data that can be embedded without compromising reliability and transparency;
- developing methods that can embed and detect watermarks in real time for live audio streams;
- protecting watermarks from deliberate attacks aimed at removing or altering them without degrading the audio quality.

Conclusions:

The research has resulted in a comprehensive understanding of the current state of digital audio

content labeling. In recent years, there has been a significant breakthrough in the development of digital watermarking methods and technologies due to the rapid development of artificial intelligence algorithms. Until 2020, 80% of research in this area was conducted for the purpose of watermarking confidential information, and only 20% for processing standard content. Today, research in these two segments is evenly distributed, and parity has almost been achieved.

Despite the significant progress, this area requires further in-depth research. Research on the segments of the watermarking system related to solving the following problems is promising: - system resistance to attacks; - transparency for perception; - data capacity and transmission speed; - real-time processing of audio content: - security.

Modern methods of automatic sound classification only partially reproduce human perception. As a result, we have identified the limitations of existing methods, in particular, the incomplete reproduction of human perception of sound in classification algorithms, the difficulty of ensuring a balance between the main characteristics (security, capacity, stability, invisibility). This, in turn, requires new research.

For the development of the considered DCT-MLP-LSB algorithm, we propose to replace the MLP (Multi-Layer Perceptron) model with the SVM (Support Vector Machine) model in the pipeline of the hybrid DCT-MLP-LSB algorithm.

To minimize the risks associated with the use of digital watermarks, it is recommended to optimize the imperceptibility by applying perceptual models to select the least significant bits in the ranges that do not affect the sound quality; increase the reliability of the system by using error correction codes (e.g., Reed-Solomon or Hamming) to ensure resistance to distortion; encrypt watermark using cryptographic encoding to protect against potential attacks; - use of adaptive algorithms that take into account the specifics of the audio signal, which increases the flexibility of the system; regularly test the system on a variety of audio files and scenarios, including the impact of noise, compression and cropping.

It is necessary to develop recommendations for improving digital watermarking algorithms, in particular by integrating modern artificial intelligence approaches, improving signal processing models, and developing hybrid algorithms.

References

1. Lu, Z., Guo, S. Introduction. In Elsevier eBooks, 2017, pp. 1–68. doi.org/10.1016/b978-0-12-812006-4.00001-2.

2. Charfeddine, M., El'arbi, M., Koubaa, M., Ben, A. C. DCT based blind audio watermarking scheme. Proc. Int. Conf. Signal Process. Multimedia Appl. (SIGMAP), 2010, pp. 139-144.

3. Bhat, V., Sen, G. I. D. A. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain. Digit. Signal Process., 2010, vol. 20, no. 6, pp. 426-436.

4. Charfeddine, M., El'arbi, M., Ben Amar, C. A blind audio watermarking scheme based on neural network and psychoacoustic model with error correcting code in wavelet domain. Proc. 3rd Int. Symp. Commun. Control Signal Process., 2008, pp. 1138-1143.

5. Charfeddine, M., Masmoudi, S., Bellaaj, M., Ben Amar, C. Un schéma aveugle de tatouage audio numérique opérant sur les bits les moins significatifs dans le domaine fréquentiel utilisant un code correcteur d'erreurs. Proc. 6émes Ateliers de Traitement et Analyse de l'Inf. Méthodes et Appl. (TAIMA), 2009, pp. 371-377.

6. Megías, D., Mazurczyk, W., Kuribayashi, M. Data hiding and its applications: digital watermarking and steganography. Applied Sciences, 2021, vol. 11, no. 22, p. 10928.

7. Data Hiding with Deep Learning: A Survey Unifying Digital Watermarking and Steganography. (n.d.).

8. Zhao, H., Shen, D. An audio watermarking algorithm for audio authentication. 2010 IEEE International Conference on Information Theory and Information Security, Beijing, China, 2010, pp. 807-809. https://doi.org/10.1109/ICITIS.2010.5689694.

9. Al-yaman, M. S., Al-taee, M. A., Shahrour, A. T., Al-husseini, I. A. Biometric based audio ownership verification using discrete wavelet transform and SVD techniques. Eighth International Multi-Conference on Systems, Signals & Devices, Sousse, Tunisia, 2011, pp. 1-5. https://doi.org/10.1109/SSD.2011.5993562.

10. He, J., Zhu, P., Liu, Z., Cao, Y. A Novel Digital Audio Encryption and Forensics Watermarking Scheme. IEEE Access, 2024, vol. 12, pp. 103565-103582. doi.org/10.1109/ACCESS.2024.3434576.

11. Petrovic, R. Digital watermarks for audio integrity verification. TELSIKS 2005 - 2005 uth International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services, Nis, Serbia, 2005, pp. 215-220 vol. 1. https://doi.org/10.1109/TELSKS.2005.1572095.

12. Sensio, E. Médiamétrie - Watermarking - a technological innovation for television audience measurement. (n.d.).

13. Li, Pengcheng, Zhang, Xulong, Xiao, Jing, Wang, Jianzong. IDEAW: Robust Neural Audio Watermarking with Invertible Dual-Embedding. 2024. https://doi.org/10.48550/arXiv.2409.19627.

14. Moritz, M., Olán, T., Virtanen, T. Noise-to-Mask Ratio Loss for Deep Neural Network Based

Audio Watermarking. 2024 IEEE 5th International Symposium on the Internet of Sounds (IS2), Erlangen, Germany, 2024, pp. 1-6. https://doi.org/10.1109/IS262782.2024.10704132.

15. Tsai, H.H., Cheng, J.S., Yu, P.T. Audio Watermarking Based on HAS and Neural Networks in DCT Domain. EURASIP J. Adv. Signal Process., 2003, 764030.doi.org/10.1155/S1110865703208027.

16. Li, Y., Wang, H., Barni, M. A survey of deep neural network watermarking techniques. Neurocomputing, 2021, vol. 461, pp. 171-193. https://doi.org/10.48550/arXiv.2103.09274.

17. Liu, X., Li, X., Niu, X. et al. Robust audio watermarking algorithm resisting cropping based on SIFT transform. Multimed Tools Appl, 2024, vol. 83, pp. 40657–40676.
https://doi.org/10.1007/s11042-023-16827-3.

18. Pourhashemi, S.M., Mosleh, M., Erfani, Y. A novel audio watermarking scheme using ensemble-based watermark detector and discrete wavelet transform. Neural Comput & Applic, 2021, vol. 33, pp. 6161–6181.doi.org/10.1007/s00521-020-05389-2.

19. Das, A., Zhong, X. A Deep Learning-based Audio-in-Image Watermarking Scheme. 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 2021, pp. 1-5.
https://doi.org/10.1109/VCIP53242.2021.9675375.

20. Megías, D. Neural Network-Based Audio Watermarking for Copyright Protection. Journal of Digital Rights Management, 2023, no. 1, pp. 50-65.

21. Peleshko, D., Rak, T., Peleshko, M., Izonin, I., Batyuk, D. Two-frames image superresolution based on the aggregate divergence matrix. 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 2016, pp. 235-238. doi.org/10.1109/DSMP.2016.7583548.

22. Hu, J., Lim, B.H. Hybrid Neural Network Approach for Audio Watermarking. Journal of Artificial Intelligence Research, 2021, no. 5, pp. 300-315.

23. Furu Wei Charfeddine, M., Mezghani, E., Masmoudi, S., Ben Amar, C., Alhumyani, H. Audio Watermarking for Security and Non-Security Applications. IEEE Access, 2022, vol. 10, pp. 12654-12677. doi.org/10.1109/ACCESS.2022.3145950.

24. Zhou, J., Chen, P. Generalized Discrete Cosine Transform. 2009 Pacific-Asia Conference on Circuits, Communications and Systems, Chengdu, China, 2009, pp. 449-452. doi.org/10.1109/PACCS.2009.62.

25. Alsmadi, M.K., Omar, K.B., Noah, S.A., Almarashdah, I. Performance Comparison of Multi-layer Perceptron (Back Propagation, Delta Rule and Perceptron) algorithms in Neural Networks. 2009 IEEE International Advance Computing Conference, Patiala, India, 2009, pp. 296-299.
https://doi.org/10.1109/IADCC.2009.4809024.

26. Kurdi, M.M., Elzein, I.A., Zeki, A.M. Least Significant Bit (LSB) and Random Right Circular Shift (RRCF) in digital watermarking. 2016 12th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 2016, pp. 111-116.https://doi.org/10.1109/ICENCO.2016.7856454.

27. Gupta, A., Yilmaz, A. Social network inference in videos. In Elsevier eBooks, 2018, pp. 395–424.

28. Ramirez, A.D.P., de la Rosa Vargas, J.I., Valdez, R.R., Becerra, A. A comparative between Mel Frequency Cepstral Coefficients (MFCC) and Inverse Mel Frequency Cepstral Coefficients (IMFCC) features for an Automatic Bird Species Recognition System. 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Gudalajara, Mexico, 2018, pp. 1-4.
https://doi.org/10.1109/LA-CCI.2018.8625230.

29. Zulkifly, M. A. A., Yahya, N. Relative spectral-perceptual linear prediction (RASTA-PLP) speech signals analysis using singular value decomposition (SVD). 2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation (ROMA), Kuala Lumpur, Malaysia, 2017, pp. 1-5.
https://doi.org/10.1109/ROMA.2017.8231833.

30. Scheirer, E., Slaney, M. Construction and evaluation of a robust multifeature speech/music discriminator. Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 1997, vol. 2, pp. 1331-1334.

31. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., Weiss, B. A database of German emotional speech. Interspeech, 2005, vol. 5, pp. 1517-1520.

32. Steinebach, M., et al. StirMark benchmark: audio watermarking attacks. Proceedings International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, 2001, pp. 49-54.
https://doi.org/10.1109/ITCC.2001.918764.

33. Liu, H., Guo, M., Jiang, Z., Wang, L., & Gong, N. Z. AudioMarkBench: Benchmarking Robustness of Audio Watermarking. arXiv preprint arXiv:2406.06979, 2024.

34. Chen, G., Wu, Y., Liu, S., Liu, T., Du, X., Wei, F. Wavmark: Watermarking for audio generation. arXiv preprint arXiv:2308.12770, 2023.

Науково-методична стаття.
Надійшла до редакції 20.11.2024.
Прийнято до публікації 18.12.2024.