Democratic Republic of Algeria and Popular

Ministry of higher education and scientific research

KASDI Merbah University – Ouargla

Faculty of New Information and Communication Technologies

Department of Computer Science and Information Technology

Thesis

Presented for the LMD Master's Diploma in
Network administration and security

Theme

Hybrid Blind Audio Watermarking Approach Based On DWT-SVD For Copyrights Protection

---

Presented by:
Benhammia Mohammed Mohcene
Zeghabi Oussama

Supervised by:
Dr. Euschi saleh

University year: 2023/2024

Acknowledgement

First of all, we would like to thank the Almighty God, who has given us the strength, patience, and determination to carry out this work.

We would like to thank our dear parents for their unlimited encouragement and support, God bless them.

We are grateful to Dr. Euschi saleh, our supervisor, for his availability, valuable advice, and encouragement, as well as our friends in the department.

We also extend our gratitude to the members of the jury for their interest in our research, for agreeing to review our work, and for enriching it with their suggestions.

ABSTRACT

Nowadays where the digital world is getting bigger, the distribution and sharing of audio content have become incredibly easy, leading to increased concerns over unauthorized copying and distribution. Protecting copyrights in the realm of audio has thus become a crucial issue. Here comes Audio watermarking as a solution to this problem, offering a method to embed information directly into audio files in a way that is imperceptible to human listeners but detectable by machines. The problem is Achieving a method robust enough to withstand common audio attacks and capable of embedding a sufficient amount of data without compromising the audio quality and imperceptibility is really hard and challenging. that's why in this project we used a hybrid watermarking approach that combines Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) to enhance the robustness and imperceptibility of the watermark. By decomposing the audio signal with DWT which can let us work in both domains time domain and frequency domain, and then embedding the watermark using SVD, we ensure minimal impact on audio quality while making the watermark resistant to common audio attacks and offer enough capacity. this method is directed to the field of copyrights protection. and we have measured all these factors using common methods like Signal-to-Noise Ratio (SNR) for imperceptibility and Bit Error Rate (BER) for robustness.

Keywords: Audio watermarking, watermarking domains, DWT, SVD, SNR, BER.

ملخص

في الوقت الحالي، حيث يتسع العالم الرقمي بشكل متزايد، أصبحت عملية توزيع ومشاركة المحتوى الصوتي سهلة بشكل لا يصدق، مما أدى إلى زيادة المخاوف بشأن النسخ والتوزيع غير المصرح بهما. لذلك، أصبحت حماية حقوق الطبع والنشر في مجال الصوت مسألة حاسمة. هنا يأتي دور العلامة المائية الصوتية كحل لهذه المشكلة، حيث تقدم طريقة لإدراج المعلومات مباشرة في الملفات الصوتية بطريقة غير محسوسة للمستمعين البشريين ولكن يمكن اكتشافها بواسطة الآلات. تكمن المشكلة في تحقيق طريقة قوية بما يكفي لتحمل الهجمات الصوتية الشائعة وقادرة على إدراج كمية كافية من البيانات دون المساس بجودة الصوت، وهو أمر صعب للغاية ومليء بالتحديات. لهذا السبب، استخدمنا في هذا المشروع نهجًا هجينًا للعلامات المائية يجمع بين تحويل المويجات المتقطع (DWT) و القيمة المنفردة (SVD) لتعزيز متانة وعدم محسوسية العلامة المائية. من خلال تحليل الإشارة الصوتية باستخدام (DWT) ، الذي يتيح لنا العمل في كل من المجال الزمني والمجال الترددي، ومن ثم إدراج العلامة المائية باستخدام (SVD)، نضمن تأثيرًا ضئيلًا على جودة الصوت مع جعل العلامة المائية مقاومة للهجمات الصوتية الشائعة وتوفير سعة كافية. هذه الطريقة موجهة لحماية حقوق الطبع والنشر. وقد قمنا بقياس كل هذه العوامل باستخدام طرق شائعة مثل نسبة الإشارة إلى الضوضاء (SNR) لعدم المحسوسية ومعدل خطأ البت (BER) للمتانة.

الكلمات المفتاحية : العلامة المائية الصوتية، مجالات العلامة المائية، DWT، SVD، SNR، BER.

Résumé

De nos jours, où le monde numérique prend de plus en plus d'ampleur, la distribution et le partage de contenu audio sont devenus incroyablement faciles, entraînant une augmentation des préoccupations concernant la copie et la distribution non autorisées. La protection des droits d'auteur dans le domaine de l'audio est donc devenue une question cruciale. C'est là qu'intervient le tatouage audio comme solution à ce problème, offrant une méthode pour intégrer des informations directement dans le fichier audio de manière imperceptible pour les auditeurs humains mais détectable par les machines. Le problème est de parvenir à une méthode suffisamment robuste pour résister aux attaques audio courantes et capable d'intégrer une quantité suffisante de données sans compromettre la qualité audio et l'imperceptibilité, ce qui est vraiment difficile et exigeant. C'est pourquoi, dans ce projet, nous avons utilisé une approche de tatouage hybride qui combine la transformation en ondelettes discrètes (DWT) et la décomposition en valeurs singulières (SVD) pour améliorer la robustesse et l'imperceptibilité du tatouage. En décomposant le signal audio avec DWT, qui nous permet de travailler à la fois dans le domaine temporel et le domaine fréquentiel, puis en intégrant le tatouage en utilisant SVD, nous assurons un impact minimal sur la qualité audio tout en rendant le tatouage résistant aux attaques audio courantes et en offrant une capacité suffisante. Cette méthode est destinée au domaine de la protection des droits d'auteur. Nous avons mesuré tous ces facteurs en utilisant des méthodes courantes telles que le rapport signal sur bruit (SNR) pour l'imperceptibilité et le taux d'erreur binaire (BER) pour la robustesse.

Mots-clés : Tatouage audio, domaines de tatouage, DWT, SVD, SNR, BER

Table of Contents

Acknowledgement 2

ABSTRACT 3

ملخص 3

Résumé 4

General Introduction 10

Chapter 1 Digital Watermarking and Digital Audio Fundamentals 12

1.1 Introduction 13

1.2 Information hiding techniques 13

1.2.1 Cryptography 13

1.2.2 Steganography 14

1.2.3 Watermarking 15

1.2.4 An overview of Digital Audio Watermarking 15

1.3 Digital Audio Background 16

1.4.1 The nature of sound 17

1.4.2 Sound wave representation 17

1.5 Sound Sampling 18

1.5.1 Pulse Code Modulation (PCM) 19

1.5.3 Bit Depth 20

1.5.4 Playback 20

2.6 Quantization 20

1.7 Compression 21

1.8 Audio File Format 22

2.8.1 Audio formats without compression 22

1.8.2 Audio formats with compression 23

1.9 Conclusion 24

Chapter 2 Audio Watermarking 25

3.1 Introduction to Audio Watermarking 26

3.2 General schema of digital watermarking 26

3.2.1 Watermark Generation 26
3.2.2 Embedding the Watermark 26
3.2.3 Watermark extraction 26
3.3 The main properties of a digital audio watermarking system 27
3.3.1 Imperceptibility 27
3.3.2 Robustness 28
3.3.2 Capacity 28
3.3.4 Computational Cost and Complexity 28
3.4 Classification of digital audio watermarking systems 28
3.4.1 According to Watermarking Domain 29
3.4.2 According to Robustness 30
3.4.3 According to embedding methods 31
3.4.4 According to Extraction Methods 32
3.4.5 According to application fields 32
3.5 Existing Methods in Spatial Domain 33
3.5.1 Least Significant Bit (LSB) Technique 33
3.5.2 Echo Hiding 34
3.6 Existing Methods in Transform Domain 35
3.7 Performance Evaluation Factors 40
3.7.2 Signal-to-Noise Ratio (SNR) 41
3.7.3 Bit Error Rate (BER) 41
3.7 Different Types of Attacks 42
3.7.1 Noise Addition 42
3.7.2 Filtering 42
3.7.3 Compression 42
3.7.4 Amplitude Scaling 42
3.7.5 Pitch Shifting 42
3.8 Conclusion 43
Chapter 3 Experiments and Results Discussion 44
4.1 Introduction 45
4.2 Environment 45

4.2.1 Python 45
4.2.2 Jupyter 46
4.2.3 PyQt5 46
4.3 Audio Watermark Algorithm Using DWT-SVD 47
4.3.1 Embedding Method 47
4.3.2 Extracting Method 49
4.4 Results 49
4.4.1 The Interface 49
4.4.2 Capacity Test 51
4.4.3 Imperceptibility Test 51
4.4.3 Robustness Test 52
4.5 Conclusion 55
General Conclusion and Perspectives 56
References 58

List of Figures

FIGURE 1 HIDING A SECRET MESSAGE IN A STEGANOGRAPHY SCHEME 14
FIGURE 2 DYNAMIC MICROPHONE 17
FIGURE 3 SOUND WAVE REPRESENTATION 18
FIGURE 4 SOUND SAMPLING PROCESS 19
FIGURE 5 QUANTIZATION PROCESS 21
FIGURE 6 CLASSIFICATION OF DIGITAL AUDIO WATERMARKING SYSTEMS 29
FIGURE 7 EXAMPLE OF LEAST SIGNIFICANT BIT (LSB) TECHNIQUE 34
FIGURE 8 ECHO-HIDING WATERMARKING SCHEME 35
FIGURE 9 DCT COEFFICIENTS WITH SIZE OF 8 × 8 36
FIGURE 10 ONE-LEVEL DWT SIGNAL DECOMPOSITION AND RECONSTRUCTION.. 37
FIGURE 11 LEVEL 3 DISCRETE WAVELET TRANSFORM 38
FIGURE 12 HAAR WAVELET TRANSFORM THREE-LEVELS-OF-DECOMPOSITION-IN-DWT 39
FIGURE 13 DIAGONAL MATRIX EXAMPLE 40
FIGURE 14 PYTHON LOGO 45
FIGURE 15 JUPYTER LOGO 46
FIGURE 16 PYQT5 LOGO 46
FIGURE 17 THE WATERMARK EMBEDDING PROCESS 48
FIGURE 18 THE INTERFACE IN EMBEDDING PROCESS 50
FIGURE 19 THE INTERFACE IN EXTRACTION PROCESS 50

List of Tables

TABLE 1 AUDIO FORMATS WITHOUT COMPRESSION 23
TABLE 2 AUDIO FORMATS WITH COMPRESSION 23
TABLE 3 SNR TEST RESULTS 51
TABLE 4 COMPARISON OF PROPOSED ALGORITHM SNR WITH PREVIOUS WORK. 52
TABLE 5 WATERMARK DETECTION RESULTS AGAINST COMMON ATTACKS (MUSIC AUDIO) 53
TABLE 6 WATERMARK DETECTION RESULTS AGAINST COMMON ATTACKS (SPEETCH AUDIO) 54

Abbreviations List

Abbreviations	Designations
HAS	Human Auditory System
HVS	Human Visual System
DCT	Discrete Cosine Transform
DWT	Discrete wavelet Transform
SVD	Singular Value Decomposition
LSB	Least Significant Bit
SNR	Signal to Noise Ratio
BER	Bit Error Rate
NC	Normalized Correlation
WAV	Waveform
MP3	MPEG-1 audio Layer 3
PCM	Pulse Code Modulation

General Introduction

The digital revolution, with the explosion of communication networks and the ever-growing enthusiasm of the general public for new information technologies, has led to an increased circulation of documents and digital works (images, videos, texts, sounds, etc.). However, this remarkable revolution has not occurred without raising concerns about the protection of copyrights.

The primary technology employed by owners to safeguard their works is cryptography. Cryptography is undoubtedly the most widespread method for protecting digital content, meeting user security needs such as confidentiality, integrity, and identification. However, once digital documents are decrypted, they lose their protection.

Thus, there is a pressing need for an alternative or complement to cryptography, a technology that allows content to remain protected even after decryption. Digital watermarking was introduced to address this need. This technique involves invisibly embedding a mark into a digital document, which can later be extracted from a suspected medium to verify its authenticity.

In recent years, the field of digital watermarking has gone through a significant development, with numerous methods proposed. The majority of efforts have focused on watermarking images and videos, with less focus on audio. Indeed, designing an audio watermarking algorithm presents greater challenges compared to images and videos due to the high sensitivity of the human auditory system (HAS) in contrast to the human visual system (HVS). However, as digital audio documents continue to see increased usage, the demand for an optimal and efficient solution has become indispensable.

Audio watermarking algorithms need to balance imperceptibility, robustness, and capacity. Imperceptibility ensures that the watermark does not degrade the audio quality. Robustness guarantees that the watermark can withstand various attacks. Capacity refers to the amount of information that can be embedded within the audio signal without compromising its quality or robustness.

One promising approach to audio watermarking, particularly for copyright protection, is the combination of Discrete Wavelet Transform (DWT) and Singular Value Decomposition

(SVD). DWT offers a multi-resolution representation of the audio signal, making it possible to embed the watermark in different frequency bands, which can enhance robustness. On the other hand, SVD provides a stable and effective ways of embedding the watermark by manipulating the singular values of the transformed audio signal.

The thesis is structured into three chapters:

Chapter 1 provides a comprehensive introduction to the core concepts of digital watermarking and digital audio. It covers the fundamentals of digital watermarking, including information hiding techniques and explores key concepts in digital audio.

Chapter 2 focuses on audio watermarking, detailing audio watermarking properties and methods in time domain, frequency domain and hybrid domain.

Chapter 3 covers the experimental setup, showcasing the results of implementing the hybrid dwt-svd proposed method, evaluating its performance in terms of robustness and imperceptibility, and comparing it with existing approaches.

Chapter 1 Digital Watermarking and Digital Audio Fundamentals

Chapter 1 Digital Watermarking Digital Audio Fundamentals

1.1 Introduction

Digital watermarking involves embedding a hidden, imperceptible signal or information within an audio, image, or video file, providing a way to verify authenticity and protect copyrights. Digital audio watermarking specifically targets audio files, embedding data that can later be extracted to identify the owner or verify the content's integrity. This technique is crucial in combatting unauthorized distribution and piracy. Understanding digital audio fundamentals, such as sampling, quantization, and compression, is essential for effectively implementing and analyzing audio watermarking methods, as they directly impact the feasibility and robustness of the watermarking process.

1.2 Information hiding techniques

Information hiding refers to the practice of concealing data within other data to protect it from unauthorized access or detection. It involves various techniques aimed at embedding or encoding information into carrier objects while maintaining their original functionality and appearance [1].

The main concepts in information hiding are **Steganography**, **Watermarking** and **Cryptography**. These concepts are essential components of information hiding and play critical roles in ensuring the security and integrity of digital information in various applications.

1.2.1 Cryptography

Cryptography is the practice of securing communication and data by converting plaintext into ciphertext using cryptographic algorithms and keys [2].

The primary goals of cryptography are confidentiality, integrity, authentication, and non-repudiation. Confidentiality ensures that only authorized parties can access the information, while integrity guarantees that the data remains unaltered during transmission. Authentication verifies the identities of communicating parties, and non-repudiation prevents individuals from denying their actions or commitments.

Cryptography operates through the use of cryptographic primitives such as encryption, decryption, hashing, and digital signatures. Encryption converts plaintext into ciphertext, making it unreadable without the corresponding decryption key. Hashing generates a fixed-size output, or hash value, from input data, commonly used for data integrity verification. Digital signatures provide a way to authenticate the origin and integrity of a message.

Modern cryptography distinguishes between symmetric and asymmetric encryption. Symmetric encryption uses a single key for both encryption and decryption, while asymmetric encryption employs a pair of keys: a public key for encryption and a private key for decryption. This duality enables secure communication between parties who have never met or shared a secret key.

1.2.2 Steganography

Steganography is the practice of concealing secret information within an ordinary, non-secret file or message to avoid detection. It involves embedding the hidden data into the cover medium in such a way that it remains undetectable to anyone except the intended recipient, as we can see in (Figure 1).

Unlike cryptography, which focuses on making the message unintelligible to unauthorized parties, steganography aims to hide the very existence of the message. This can be achieved by imperceptibly altering the cover medium's properties, such as modifying pixel values in images or least significant bits in audio files. Steganography finds applications in various fields, including digital forensics, data authentication, and covert communication. Understanding steganographic techniques is crucial for both detecting hidden messages and securely communicating sensitive information [3].

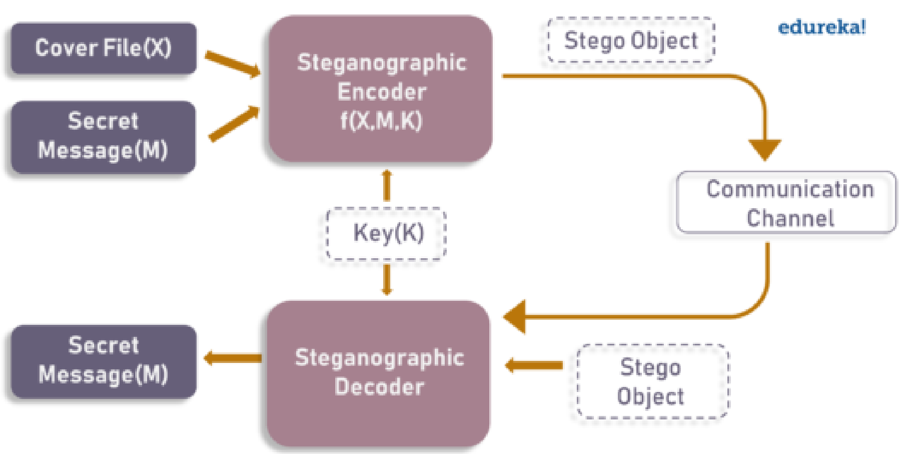


Figure 1 Hiding a secret message in a steganography scheme

1.2.3 Watermarking

Watermarking, tracing its roots back to the concept of traditional watermarks used in paper manufacturing, watermarks were patterns or logos embedded into paper during its production process, serving as a form of authentication and identification of the paper's origin or quality. It has evolved significantly in the digital era. which appears to have been first used by Tominaga and Komatsu in 1988. But it was in 1993 that Tirkel presented two watermarking techniques to hide data in images and used the term "digital watermarking" [4].

Watermarking is a technique used to embed imperceptible digital information, typically in multimedia content such as images, audio, or video, to assert ownership, authenticity, or to provide additional metadata. This process involves modifying the content in a way that the embedded information is robust against common signal processing operations and attacks, yet remains undetectable or minimally perceptible to the human senses.

Today, digital watermarking plays a crucial role in digital rights management (DRM), and content authentication across various industries including publishing, and digital distribution platforms. Its historical journey from traditional paper watermarks to modern digital techniques underscores the ongoing efforts to safeguard intellectual property rights and ensure the integrity of digital content in an increasingly interconnected world.

1.2.4 An overview of Digital Audio Watermarking

Digital audio watermarking is a technique used to embed imperceptible data into audio signals for various purposes such as copyright protection, content authentication, and ownership verification. This process involves altering the audio signal in a way that the embedded watermark remains hidden to human perception while being robust against common signal processing operations and attacks.

There are several approaches to digital audio watermarking, including time domain, and transform domain methods. Time domain techniques involve directly modifying the audio samples, while frequency domain methods operate on transformed representations such as the Fourier or wavelet domain.

Digital audio watermarking has applications in industries such as music, broadcasting, and digital media distribution. also finds significant applications in telemedicine. In the realm of telemedicine, audio data often contains sensitive patient information, and ensuring its authenticity and integrity is important. It enables copyright owners to track their audio content and its usage across different platforms. Additionally, it aids in detecting unauthorized copying, tampering, or distribution of audio files [5].

Researchers continue to advance digital audio watermarking techniques to improve their robustness, imperceptibility, and resistance against attacks. This ongoing development is essential for ensuring the integrity and security of audio content.

1.3 Digital Audio Background

The background of digital audio traces back to the mid-20th century when the concept of converting analog audio signals into digital form emerged. Analog audio, which represents sound waves as continuous electrical signals, had been the primary means of recording, transmitting, and reproducing sound for decades. However, analog audio suffered from limitations such as signal degradation over distance, susceptibility to noise, and challenges in storage and manipulation [6].

The shift towards digital audio began with the invention of pulse code modulation (PCM) in the 1930s, a technique that converts analog signals into digital data by sampling the signal at regular intervals and quantizing each sample into binary code. PCM laid the foundation for the digitization of audio signals, enabling them to be stored, transmitted, and processed in more efficient and reliable manner.

One of the pioneering developments in digital audio was the creation of the Compact Disc (CD) in the late 1970s and early 1980s. CDs utilized PCM encoding to store audio data as a series of pits and lands on a reflective surface, which could be read by a laser beam. This marked a significant leap forward in audio technology, offering improved sound quality, durability, and portability compared to analog formats like vinyl records and cassette tapes [7].

Throughout the 1980s and 1990s, digital audio technology continued to advance rapidly. The development of audio compression algorithms such as MPEG Audio Layer III (MP3) and Advanced Audio Coding (AAC) allowed for the efficient storage and transmission of audio data over digital networks. These compression techniques enabled the proliferation of digital music formats and the rise of online music distribution platforms.

1.4 Sound features

1.4.1 The nature of sound

When a sound is made by a clap or someone's voice the air molecules are disturbed as each molecule collides with its nearest neighbors a wave of alternating compressions and decompressions moves away from the source of the sound.

To capture this sound, we can use a microphone. A dynamic microphone contains a magnet surrounded by a coil of copper wire, and attached to the coil is a thin membrane, a diaphragm (Figure 2). When a sound wave hits the diaphragm, the coil moves in relation to the magnet, and this induces an electrical current in the wire. This current can then be recorded somehow, and the sound played back later.

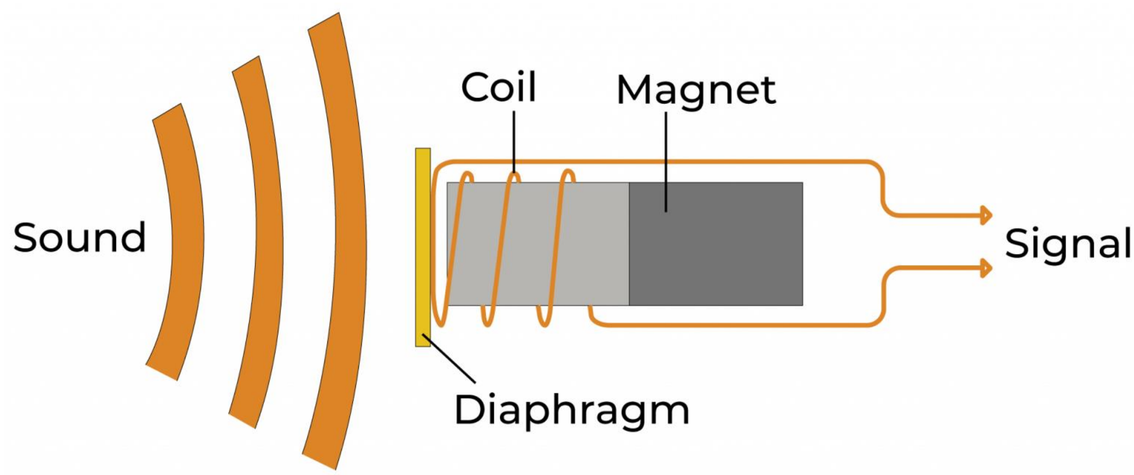


Figure 2 dynamic microphone

1.4.2 Sound wave representation

To describe a sinusoid, three elements must be specified: the frequency, (specified in Hertz, where 1 Hertz (Hz) = 1 cycle/s). the amplitude, and the phase, which corresponds to the part of the cycle by which the wave has moved forward relative to a fixed point in time.

When the wave travels a complete cycle, the phase changes by 360°, which is equivalent to 2π radians. For the continuous sinusoids, the phase is only important when considering the relationship between two or more different waves. The time required for a cycle (The entire waveform) is called the period [8].

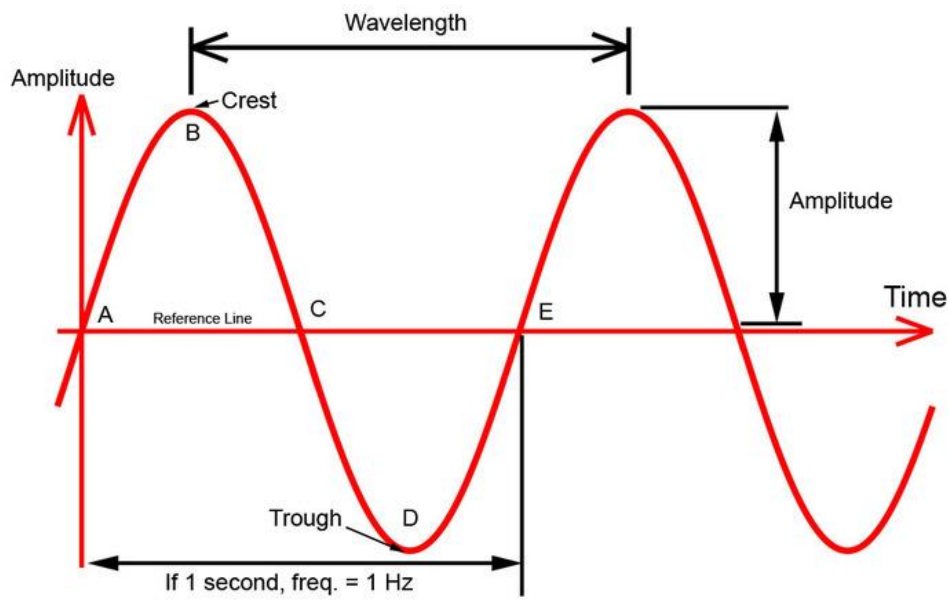


Figure 3 Sound wave representation

This sine wave in (Figure 3) is just an approximation of what might be produced by a guitar string. The maximum height of this graph is related to the number of molecules displaced by the original disturbance that created the sound. Or, to put it another way, the maximum height of this graph reflects the amount of compression at that point in time. The higher the level of compression, the louder the sound. this is the so-called amplitude. The distance between two adjacent peaks(crest), two adjacent troughs, or to be more precise, the length of a single cycle, is known as the wavelength. The shorter the wavelength, the higher the frequency of the sound, and hence the higher the pitch that you hear [8].

1.5 Sound Sampling

Sound sampling is the process of capturing and recording audio signals at discrete points in time (Figure 4). This process involves converting continuous analog sound waves into a digital format that can be stored, manipulated, and reproduced by electronic devices such as computers and audio equipment [9].

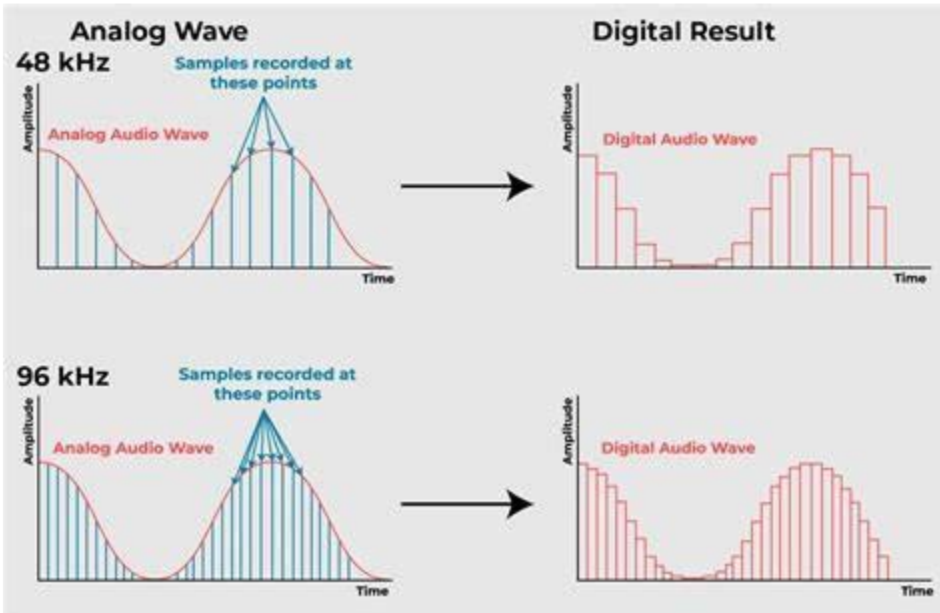


Figure 4 Sound sampling process

Here is a brief overview of the sound sampling process:

1.5.1 Pulse Code Modulation (PCM)

PCM is a technique that converts analog signals into digital data without compression, by sampling the signal at regular intervals and quantizing each sample into binary code. this digital representation of the analog signal can be easily stored, transmitted and processed in more efficient and reliable manner.

1.5.2 Sampling Rate

The sampling rate, measured in Hertz (Hz), determines how often the analog signal is sampled per second. Common sampling rates include 44.1 kHz used in CDs, 48 kHz used in digital audio workstations and DVDs, and higher rates such as 96 kHz or 192 kHz for high-resolution audio [9].

1.5.3 Bit Depth

The bit depth determines the resolution of each sample and affects the dynamic range of the digital audio signal. Common bit depths include 16-bit (CD quality) and 24-bit (high-resolution audio). Higher bit depths allow for more precise representation of the original analog signal, reducing quantization noise and increasing dynamic range.

1.5.4 Playback

To reproduce the original sound, the digital audio data is converted back into analog signals through a digital-to-analog converter (DAC). These analog signals can then be amplified and played through speakers or headphones.

2.6 Quantization

To represent an analog signal in the digital domain, we sample the signal at discrete points in time and ignore the rest. Similarly, each point on the amplitude scale can have a theoretically infinite resolution with an unending number of digits, so we really need to draw a line to represent the maximum resolution that we are willing to maintain. You can think about it as the accuracy of measurement. If the resolution is really small, the accuracy of the measurement drops as well. And if you have a large enough resolution, the measurements become super accurate. But you have to deal with the technical task of having to measure with such high accuracy consistently and fast enough before the next sampling interval arrives. Another consideration with high resolution is the storage capacity needed to represent and save all the sampled points and the bandwidth required to transmit large amounts of data. like any discipline of engineering, finding the right resolution is a balancing act [10].

The size of the sampling interval along the amplitude axis, or the y-axis in this case, determines the maximum dynamic range that the digital signal can represent. Dynamic range is the range between the highest and the lowest amplitude moments in the sound. The resolution only impacts the amount of noise present in the digitized signal.

Let me demonstrate this with a simple example. Let's take an arbitrary signal which is sampled along the time axis. And let's say that we split the amplitude scale into 8 discrete levels. At every sampled point, our measurements have to stick to one of these discrete levels. The eight levels are regarded as the resolution of the digitization process. This process of mapping the analog signal values to a limited range of discrete values like this is called quantization. You can think about quantization as a latch which either pulls or pushes a sample value to the nearest discrete measure. As you can see, with eight discrete points, the resolution is quite poor. This is in fact a 3-bit resolution. The difference between the original analog signal value and the discrete digital value that the signal is constrained to. this represents the quantization error within our digital signals when compared to the original analog signal.

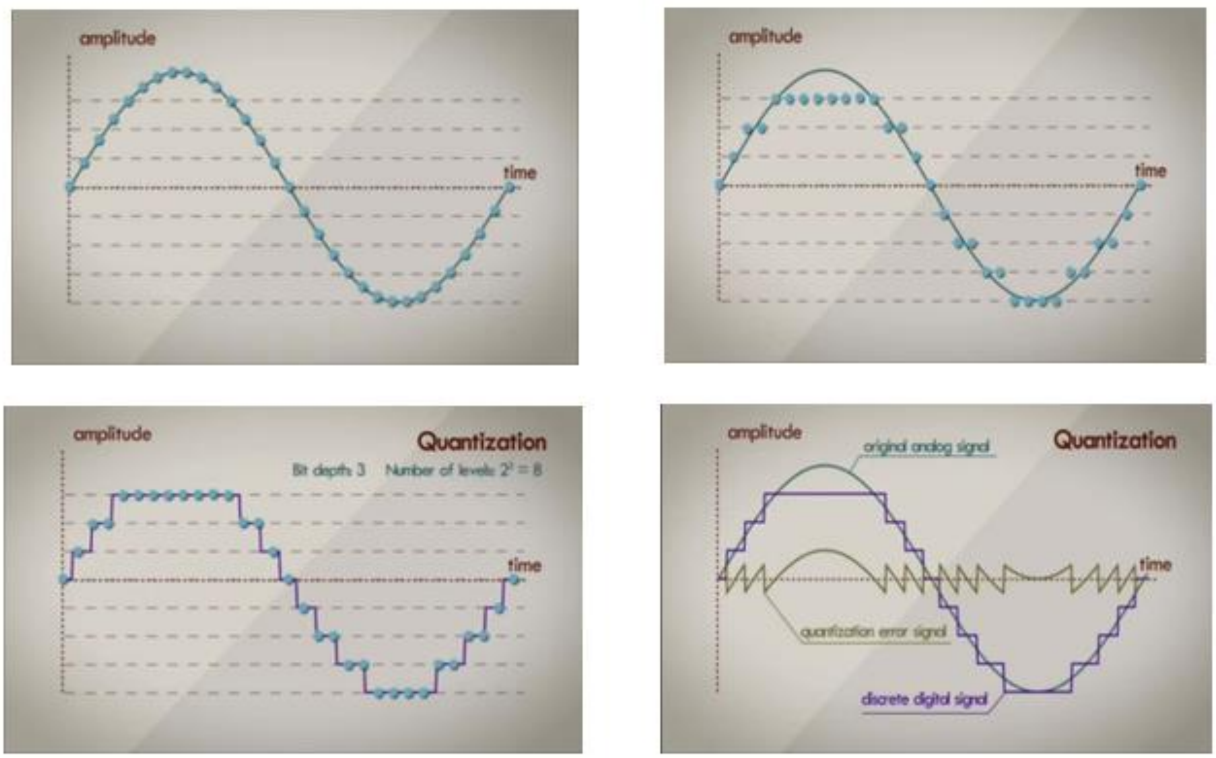


Figure 5 Quantization process

1.7 Compression

A digitized audio signal is stored on different devices like hard drives, CDs, and DVDs. But because of the rich info they carry, these files tend to be pretty big.

Digital sound became popular thanks to CDs, which use a simple way of encoding. A CD can hold about 74 minutes of music at 44100 samples per second, with each sample being 16 bits. That's roughly 650 MB of space. This CD breakthrough really boosted the audio scene. But sending such detailed sound is tough, needing high bitrates like 705.6 Kbits/s for mono and 1.411 M bits/s for stereo. Even with fast internet, it is still a challenge.

But these high bitrates mean high costs, for storage or transferring data. So, compressing sound aims to make files much smaller while keeping the original quality intact. Since compression means losing data, the trick is knowing what to cut without messing up the signal.

These techniques are already used a lot in movies and radio, through cable or satellite, shaping how media spreads nowadays.

1.8 Audio File Format

An audio file format is essen²tially a standardized way of storing digital audio data. It is like a blueprint that tells devices and software how to read and interpret the sound information. Think of it as a specific type of container for an audio data, with its own rules about how that data is organized and encoded. Common audio file formats include MP3, WAV, AAC, and FLAC. Each format has its own characteristics, such as compression methods, sound quality, and compatibility with different devices and software. [11]

MP3 is a widely used format known for its good balance between sound quality and file size, making it ideal for streaming and sharing music online. WAV, on the other hand, is an uncompressed format that preserves all the original audio data, making it great for professional audio production but resulting in larger file sizes.

Choosing the right audio file format depends on factors like your intended use, desired sound quality, and compatibility with your devices and software.

2.8.1 Audio formats without compression

Audio formats without compression are file formats used to store digital audio data in uncompressed state. These formats retain all original audio information without applying any data reduction techniques, ensuring maximum fidelity and quality preservation. Uncompressed audio formats typically encode audio data using methods such as Pulse Code Modulation (PCM), like WAV (Waveform Audio File Format) and AIFF (Audio Interchange File Format).

Table 1 Audio formats without compression

Form	**Discerption**
RIFF (Resource Interchange File Format)	RIFF is a flexible and extensible file format primarily used for storing multimedia data. It serves as a container format, enabling the encapsulation of various types of data, including audio, video, images, and more.
WAV (Waveform Audio File Format)	WAV, often referred to as WAVE, is an audio file format standard developed jointly by IBM and Microsoft in 1991. WAV files can be mono or stereo and are identified by the (.wav) file extension.
AIFF (Audio Interchange File Format)	AIFF is a sound storage format commonly used on Apple Mac OS computers. It is equivalent to the WAV format in the Windows. Files created in AIFF format typically have the .aif or .aiff file extension. Data is encoded in PCM format without compression, ensuring high-quality audio preservation.

1.8.2 Audio formats with compression

Audio formats with compression are file formats used to store digital audio data by employing algorithms to reduce file size while attempting to maintain acceptable audio quality. compressed audio formats sacrifice some degree of fidelity in exchange for reduced storage requirements and bandwidth usage. Examples of compressed audio formats include MP3 (MPEG Audio Layer-3), AAC (Advanced Audio Coding), and OGG Vorbis.

Table 2 Audio formats with compression

Form	**Discerption**

AC3 (Audio Coding 3)	AC3, is a digital audio compression algorithm developed by Dolby Laboratories in 1987. It is commonly used for encoding multichannel audio tracks in formats such as DVDs, Blu-ray discs. It achieves compression by removing redundant or less audible information from the audio signal. AC3 is suitable for encoding surround sound formats such as 5.1-channel audio commonly used in home theater systems [12].
MP3 (MPEG-1/2 Audio Layer 3)	MP3, is a digital audio encoding format developed by the Moving Picture Experts Group (MPEG). It uses perceptual audio coding to significantly reduce file size while preserving high sound quality enabling convenient storage and transmission over the internet. MP3 achieves compression by discarding audio data that's less perceptible to the human ear, making it an efficient format for distributing music and other audio content [13].
Ogg Vorbis	Ogg Vorbis is an open-source audio compression format developed by the Xiph.Org Foundation. It uses lossy compression to reduce file sizes without significant loss of audio quality. Ogg Vorbis files typically have the ".ogg" extension and are playable on various software and hardware platforms. [14]

1.9 Conclusion

In this chapter, we explored information hiding techniques: Steganography, Cryptography, and Watermarking, with a focus on audio watermarking. Steganography conceals data within other media to prevent detection, Cryptography encrypts data to ensure its security and integrity, and Watermarking embeds identifiable patterns to protect intellectual property.

we covered essential aspects of digital audio, including sound features, the nature of sound, sampling, compression, and various audio formats. This foundation in digital watermarking and digital audio is crucial for understanding the advanced techniques of audio watermarking, which will be discussed in the following chapters.

Chapter 2 Audio Watermarking

Chapter 2 : Audio Watermarking

3.1 Introduction to Audio Watermarking

the increase of multimedia content, particularly audio, has led to unprecedented challenges in protecting intellectual property rights and ensuring content integrity. Audio watermarking emerges as a crucial technique in addressing these challenges, by embedding additional data, into an audio file without perceptibly altering its auditory quality. Unlike traditional watermarks, which are visible marks on physical media, audio watermarks are imperceptible to the human ear. These embedded watermarks serve as unique identifiers or signatures, facilitating tasks such as copyright protection, content authentication, ownership verification, and digital rights management. This watermark can be a binary logo, a random sequence of noise, a grayscale image, or a string message. And it is inserted into the original signal by the encoder.

3.2 General schema of digital watermarking

3.2.1 Watermark Generation

The watermark data is generated. This may include information such as copyright details, ownership information, or any other data to be embedded into the audio. It can be an image or string and then it is transformed into bits to be embedded on the audio later.

3.2.2 Embedding the Watermark

The watermark data is embedded into the audio signal using a watermarking algorithm. This algorithm has to perform in a specific domain like **time domain** or **frequency domain.** After selecting a domain, the next step is modifying the selected features of the audio signal according to the watermark data and any additional parameters such as the embedding key. in a way that minimizes perceptual changes while ensuring the embedded data can be extracted later.

3.2.3 Watermark extraction

Watermark extraction is the process of retrieving the embedded watermark from a watermarked audio signal. It involves these general steps:

1. **Detection Algorithm**: A detection algorithm is applied to the watermarked signal to detect the presence of the watermark. This algorithm analyzes the features of the

watermarked audio signal and compares them against the expected characteristics of the embedded watermark.

2. **Thresholding:** In some cases, we have to apply a thresholding mechanism to the detection results to determine whether the watermark is present or not. it helps to differentiate between the watermark signal and any remaining noise in the watermarked audio.
3. **Watermark Extraction:** If the presence of the watermark is detected above the threshold, the extraction process retrieves the embedded watermark from the watermarked audio signal. This involves reversing the embedding process applied during watermarking to recover the original watermark data.

3.3 The main properties of a digital audio watermarking system

The most important properties of digital audio watermarking techniques are Imperceptibility, robustness, capacity, security, reliability of detection, and complexity. Optimizing these properties is mutually competitive and cannot be achieved simultaneously, a reasonable compromise is necessary. The International Federation of the Phonographic Industry (IFPI) outlines the requirements for an effective digital audio watermarking system, known as the properties of digital audio watermarking systems [15].

3.3.1 Imperceptibility

The watermarking algorithm must insert the mark without altering the perceptual quality of the original signal, ensuring that the added mark does not degrade the content's quality.

Imperceptibility is often evaluated using both objective and subjective measures. Subjective measures involve conducting listening tests with human participants to assess whether they can detect the presence of the watermark. objective mathematical measures like Signal-to-Noise Ratio (SNR). SNR measures the sensitivity of the embedded signal compared to background noise. The International Federation of the Phonographic Industry (IFPI) mandates an SNR of over 20 dB for any audio watermarking algorithm [16].

3.3.2 Robustness

Robustness refers to the ability of the watermark inserted into the original audio signal to survive various processing operations and attacks. The watermark must withstand general signal processing, filtering, compression, noise addition, and malicious attacks. Depending on the application, the watermarking technique can support different levels of robustness against changes to the watermarked signal caused by malicious attacks.

If the watermark is intended for copyright protection, it must be robust against any manipulation. However, if the watermark aims for content authentication, the robustness of the watermark must be controlled: it may tolerate certain manipulations of the digital document but must be insensitive to common treatments such as filtering and compression [17].

3.3.2 Capacity

Capacity in audio watermarking refers to the number of bits of information that can be embedded within the audio signal without significantly degrading its quality or perceptibility. Capacity is measured in terms of the number of bits inserted per unit of time, bits per second (bps) [18].

Capacity is a critical consideration in audio watermarking applications, as it determines the amount and type of information that can be reliably embedded within the audio signal for purposes such as copyright protection, content authentication, and metadata embedding.

3.3.4 Computational Cost and Complexity

The execution of the watermarking algorithm is a hard and complex task. So, in order to reduce the computational cost, the watermarking method should be less complex. Especially in environments such as mobile systems and in real-time applications. While In copyright protection applications time is not a crucial factor, we focus more on efficiency [17].

3.4 Classification of digital audio watermarking systems

Audio watermarking systems can be classified based on various criteria, including their watermarking domain, embedding methods, robustness, extraction methods, and application domains [19]. Here is a classification based on these factors:

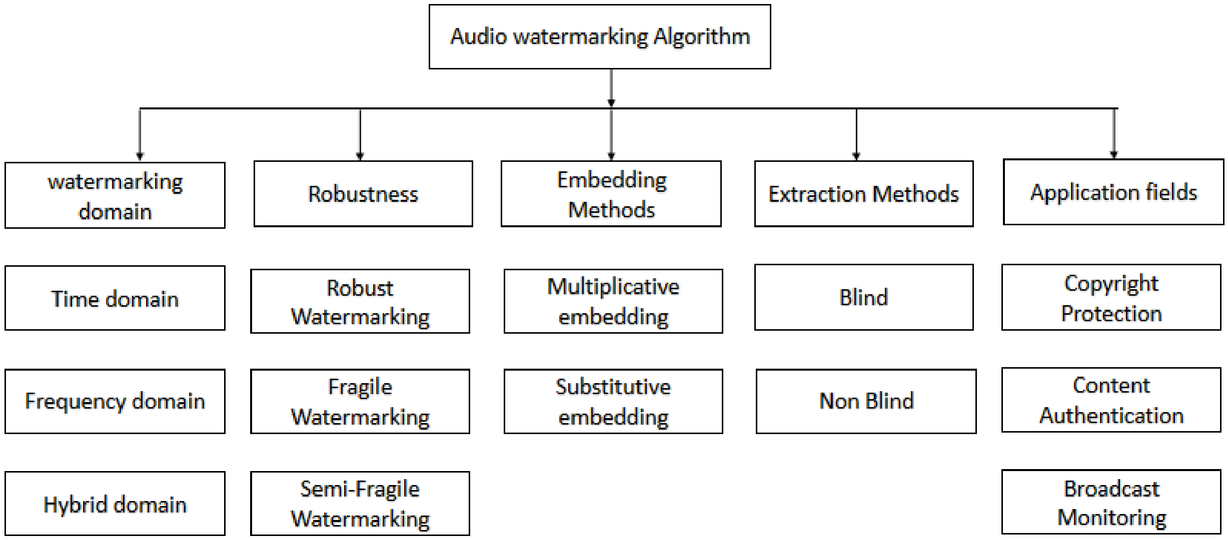


Figure 6 Classification of digital audio watermarking systems

3.4.1 According to Watermarking Domain

Classifying audio watermarking methods based on the watermarking domain provides insights into different approaches used in this field. These techniques can be categorized into three main domains. Where we can insert the watermark in **Time domain** (special domain) or, **Frequency domain** (Transform domain), or when we use both of them so it is called **Hybrid**. Each domain offers its own set of advantages and challenges, and the choice of domain depends on factors such as the application requirements, robustness against attacks, and computational complexity.

3.4.1.1 Time Domain Watermarking

Time domain watermarking operates directly on the audio samples. It embeds the watermark by modifying the amplitude or phase of the audio signal itself. Common techniques used in time domain watermarking include, LSB (Least Significant bit), Echo Hiding, Phase Coding. These techniques offer advantages such as simplicity, low computational complexity, and that makes it suitable for reel time applications like authentication. But it is limited in terms of Robustness and Capacity [20].

3.4.1.2 Frequency Domain Watermarking

Frequency domain watermarking techniques operate on transformed representations of the audio signal, such as the Fourier or wavelet transform. By converting the audio signal into the frequency domain, watermark embedding can be performed by manipulating the coefficients of the transformed signal. Some of these transformation techniques are: DCT, DFT, DWT. These techniques offer several advantages, including robustness against common signal processing operations and attacks compared to time domain techniques [21]. However, frequency domain watermarking may be more computationally complex due to transformation methods. Achieving a balance between robustness and capacity can be challenging.

3.4.1.3 Hybrid Watermarking

Hybrid watermarking combines multiple watermarking techniques or domains where you can insert the watermark in both time domain and Frequency Domain. By integrating different watermarking approaches, hybrid watermarking offers a flexible and adaptive approach to audio watermarking, allowing for customization and optimization based on specific application requirements and aims to achieve enhanced robustness, capacity, and perceptual quality. But this can be challenging and complex. One of the common hybrid-watermarking approaches is SVD-DWT [22].

3.4.2 According to Robustness

Audio watermarking techniques can be classified based on their robustness into three main categories:

3.4.2.1 Fragile Watermarking

Fragile watermarking techniques are designed to detect any modifications or tampering in the audio signal. These techniques are highly sensitive to even minor changes in the signal and are primarily used for ensuring the integrity and authenticity of the audio content [23].

3.4.2.2 Semi-Fragile Watermarking

Semi-fragile watermarking techniques are mor tolerant for some processing operations. but sensitive to malicious attacks or significant alterations in the audio signal. These techniques aim to balance robustness and sensitivity by incorporating features that are resistant to common signal processing operations while remaining detectable under intentional attacks [23].

3.4.2.3 Robust Watermarking

Robust watermarking techniques are designed to withstand a wide range of signal processing operations, and intentional attacks such as compression, filtering, noise addition, and format conversion while maintaining the integrity and detectability of the watermark. It is widely used in applications such as copyright protection, prevent unauthorized copying, and digital rights management (DRM).

3.4.3 According to embedding methods

Watermarking algorithms can be classified according to their embedding methods. so, there are typically two main methods of embedding the watermark: multiplicative (additive) and substitutive.

3.4.3.1 Multiplicative Watermarking

Multiplicative embedding involves modifying the original audio signal by multiplying it with a modulation signal that carries the watermark information. This modulation changes the audio signal's properties, encoding the watermark information without directly adding new components. Multiplicative embedding techniques may adjust the amplitude, phase, or frequency characteristics of the audio signal to encode the watermark [24].

3.4.3.2 Substitutive Watermarking

Substitutive watermarking is a simple technique used to embed watermarks into audio signals by directly replacing or modifying certain components of the original audio data. Unlike additive or multiplicative watermarking, substitutive watermarking selects a specific part and alters these parts of the audio signal to encode the watermark. After that in the extraction process the location of the replaced bit is known [25].

3.4.4 According to Extraction Methods

Watermarking algorithms can also be classified based on their extraction methods, which refer to the techniques used to recover the embedded watermark from the watermarked signal. Here are the main two extraction methods.

3.4.4.1 Non-Blind Extraction

Non-blind extraction methods require information about the original signal watermark, during the extraction. Non-blind extraction algorithms offer a higher reliability but the requirement of original signal can make it available and accessible to unauthorized sides. which may not always be practical in real-world scenarios [26].

3.4.4.2 Blind Extraction

Blind extraction methods do not require any information about the original signal during extraction. Blind extraction methods find applications in various domains. But achieving robust and reliable blind extraction requires a careful and complex design of detection algorithms, which may be a challenging task [27].

3.4.5 According to application fields

By classifying watermarking algorithms according to these fields, we can create and customize solutions that are specifically addressed to the needs of each domain. The main application fields are in Authentication Verification, Broadcast Monitoring and Digital Rights Management (DRM) or copyrights.

3.4.5.1 copyright protection

Algorithms in this category focus on protecting intellectual property rights, such as copyrighted material, against unauthorized duplication and distribution. By inserting metadata identifying copyright holders in the audio. So, it must be robust to various forms of manipulation and attacks.

3.4.5.2 Content Authentication

authentication protection involves embedding imperceptible digital watermarks into audio signals to verify their authenticity and integrity, ensuring they have not been altered. In this case fragile watermarking is the best choice because by tolerating some signal processing operations we can tell if the original signal has been modified or not [28].

3.4.5.3 Broadcast monitoring

. In broadcast monitoring, audio watermarking is used for Tracking the distribution of watermarked audio content across different broadcast channels, including radio stations, TV networks, streaming platforms, and online services. Monitoring systems can record the broadcast schedules, frequencies, and transmission times of each watermarked audio segment for tracking and analysis purposes [28].

3.5 Existing Methods in Spatial Domain

Audio watermarking in the spatial domain involves altering the sample values of the audio signal in the time-domain representation. Typically, these methods are easy to implement and do not require the original audio signal for watermark extraction (known as blind watermarking). They provide a good balance between payload capacity and transparency. However, their main disadvantage is their lack of robustness against various attacks. Spatial domain audio watermarking techniques are mainly classified into substitutive and additive embedding methods [29].

3.5.1 Least Significant Bit (LSB) Technique

The Least Significant Bit (LSB) method is one of the easiest techniques used in audio watermarking. In the digital audio, each sample is represented by a binary value. The LSB of each audio sample is chosen to embed the watermark. Since it is the least significant bit, any changes made to it are less likely to be noticeable in the audio signal. This process is done by replacing the original LSB with the watermark bits while keeping the remaining bits unchanged [30].

To extract the watermark from the watermarked audio, the LSB of each sample must be known and examined. The watermark bits are reconstructed from the LSBs of the audio samples according to a predetermined extraction algorithm.

LSB-based watermarks are often less robust to various signal processing operations and attacks compared to other watermarking techniques. Common operations like compression, filtering, and resampling can degrade or even completely remove the watermark.

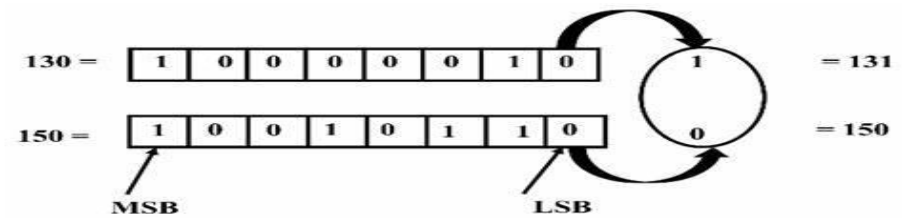


Figure 7 Example of Least Significant Bit (LSB) Technique

3.5.2 Echo Hiding

Echo hiding is a technique used in audio watermarking where data is embedded into the audio signal by introducing brief echoes. These echoes create resonances within the host audio, effectively avoiding the issue of the Human Auditory System's (HAS) sensitivity to added noise. This method is an additive and blind watermarking approach primarily tailored for audio signals [31].

In echo hiding, each echo serves as a carrier for one bit of data. To encode multiple bits, the original audio signal is divided into smaller segments. Within each segment, one bit of the watermark is concealed by introducing a tailored echo. This approach enables the seamless integration of digital information into the audio signal without perceptibly altering its overall characteristics.

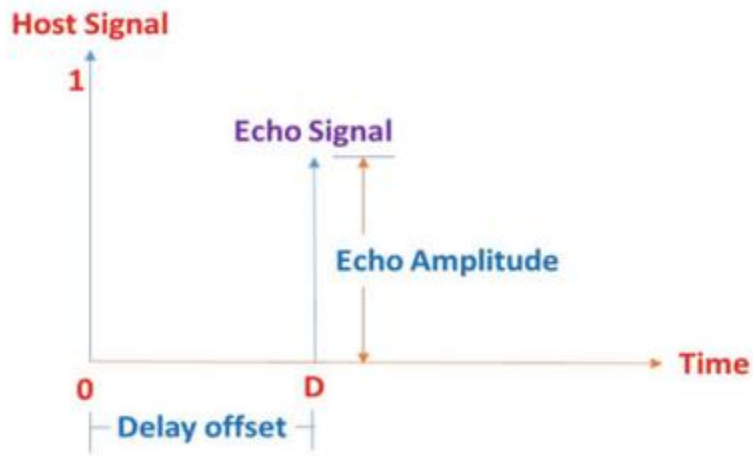


Figure 8 Echo-hiding watermarking scheme

3.6 Existing Methods in Transform Domain

Transform domain techniques involve converting the time-domain audio signal into another domain, such as the frequency or the time-frequency domain. It provides better robustness compared to spatial domain audio watermarking.

To convert from time-domain to frequency-domain we use mathematical transformations like the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT). This transformation allows for better manipulation and control over the signal's properties and coefficients, facilitating robust and imperceptible watermark embedding [32].

3.6.1 DCT

The Discrete Cosine Transform (DCT) is a mathematical technique used in signal processing to transform a signal or image from its spatial domain to the frequency domain. The audio signal is divided into smaller segments or frames, Then the DCT is applied to each segment.

After transforming from time-domain to the frequency domain the watermark is embedded into the coefficients of the DCT. This can be done by modifying certain coefficients to represent the watermark data. The modification should be imperceptible to human ears to maintain audio quality.

To use the two-dimensional Discrete-Cosine-Transform, first the audio signal is converted to two-dimensional matrix as you can see in (Figure 9), then DCT is applied to each block. the upper left corner of the matrix represents the lowest frequency coefficients, while the lower right corner represents the highest frequency coefficients. We chose the mid-band frequency coefficients to insert the watermark because it is less affect to attacks and provides more robustness compared to other coefficients. Thise coefficients are modified by scaling factor according to bit value of watermark. [33]

16	11	10	26	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figure 9 DCT coefficients with size of 8 × 8

3.6.2 DWT

The Discrete Wavelet Transform is a mathematical operation commonly used in signal processing and image compression. The DWT converts a signal into its wavelet representation.

Wavelets are small, oscillating functions that are localized in both time and frequency domains. Unlike sinusoidal functions in Fourier analysis which are infinite in extent, wavelets are finite [34].

DWT provides a multi-resolution analysis, which means it decomposes the signal into different frequency bands at different resolutions using a series of high-pass and low-pass filters. These filters separate the signal into approximation (low-frequency) and detail (high-frequency) components. As we can see in (Figure 11) below where the part (a) is one-level DWT signal decomposition and (b) one-level DWT signal reconstruction [35].

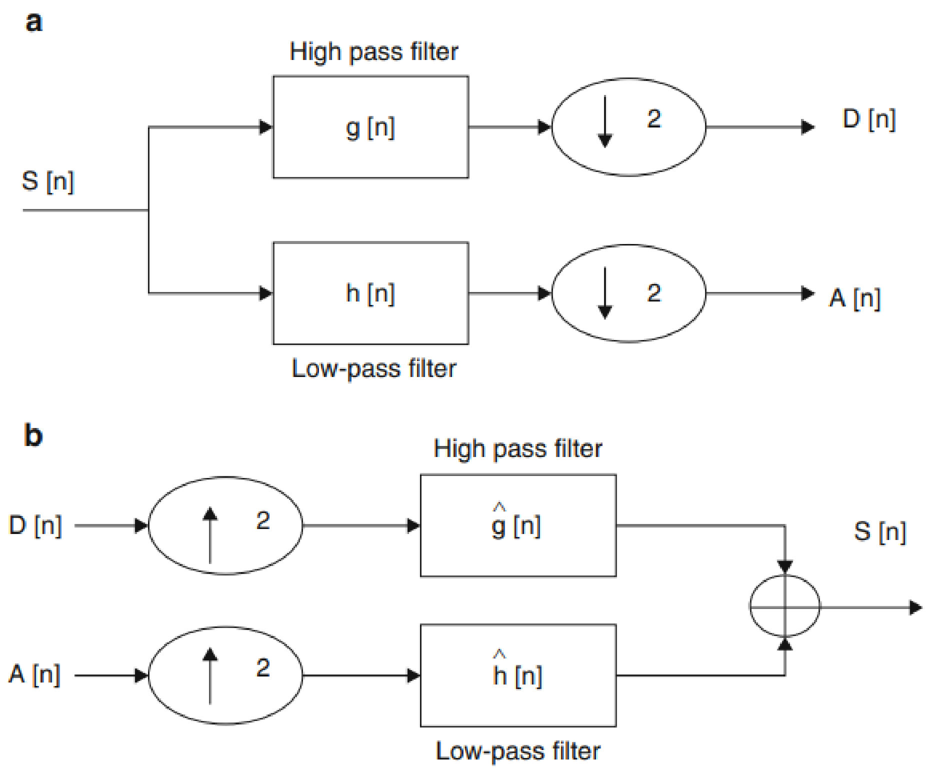


Figure 10 One-level DWT signal decomposition and reconstruction

The approximation coefficients (CA) capture the low-frequency components of the signal. These components represent the overall trend or coarse structure of the signal. It contains information about the lower-frequency content of the signal, which tends to represent global or slowly varying features. But in the other hand the detail coefficients (CD) represent the high-frequency components or fine details of the signal. These components capture localized changes or rapid variations in the signal. The decomposition process is typically represented like a tree structure as we can see in Figure 12, with each level of decomposition corresponding to a "degree" or "level" in the tree. The number of degrees or levels of decomposition in DWT depends on factors such as the desired resolution, the characteristics of the signal, and the application requirements.

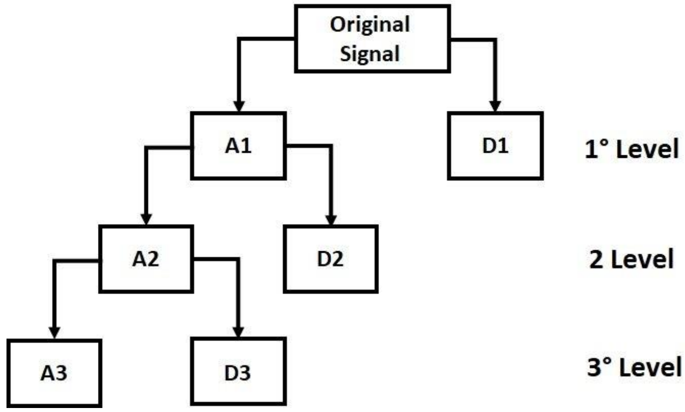


Figure 11 level 3 Discrete Wavelet Transform

Most of the time we chose the approximation coefficients for inserting watermarks. Because they contain a low-frequency information and are more robust against common signal processing operations like compression or noise addition compared to the detail coefficients. However, the thing that detail coefficients can provide is a higher capacity because it can tolerate stronger modifications without significant perceptual degradation.

There is a balance between the robustness and the capacity in finding out which level of DWT coefficients should be controlled. Controlling a higher level of coefficients gives a higher robustness, however, a lower capacity more computational cost. In the other hand Controlling a lower level of coefficients gives higher capacity yet a lower robustness [36].

One of the least difficult wavelets transform functions is the Haar wavelet transform (HWT). It is an orthogonal wavelet filter because The Haar wavelet functions are orthogonal to each other, meaning they are uncorrelated and preserve energy during transformation. The process of Haar wavelet transform is by decomposing the signal into approximation coefficients and detail coefficients, Then After filtering, the signal is down sampled by a factor of two, which reduces its size by half. This step is crucial for extracting different frequency components effectively. The more we repeat this process the more we get finer detail coefficients representing higher-frequency components of the signal. From a level one wavelet decomposition we obtain four groups or sub-bands as you can see in Figure *12*. where the (LL) band is Approximate Band, vertical band (LH), the horizontal band (HL), and detail band (HH). The LL sub-band contains the low-frequency components and we find higher-frequency components in LH, HL, HH sub-bands so we use them to insert the watermark [37]. For the extraction process we apply the IDWT to get the watermark.

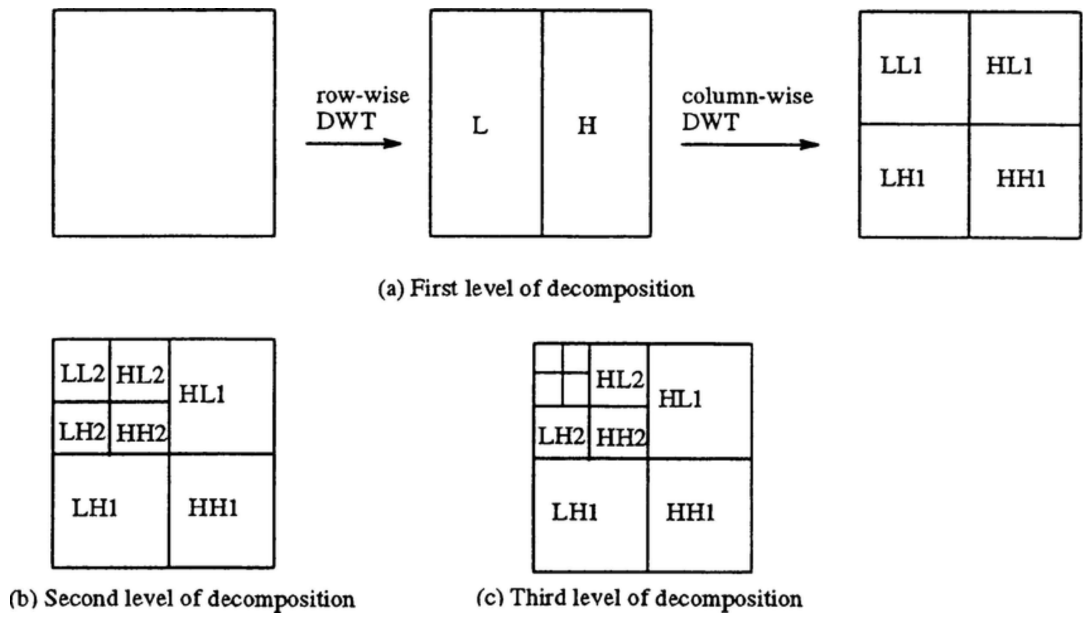


Figure 12 Haar wavelet transform Three-levels-of-decomposition-in-DWT

3.6.3 SVD (Singular Value Decomposition)

Singular Value Decomposition (SVD) is a powerful mathematical technique use to decompose real or complex matrix into three constituent matrices, facilitating various analytical and computational tasks and can be very useful in signal processing. SVD is orthogonal transformation, which can diagonalize matrix, and it expresses a matrix A as the product of three matrices [38]. A = USV^T where:

- U is an orthogonal matrix containing the left singular vectors
- S is a diagonal matrix containing the singular values in decreasing order
- V^T is the transpose of an orthogonal matrix containing the right singular vectors.

An orthogonal matrix represents a transformation that doesn't distort the shape of vectors. The key property of an orthogonal matrix is that when you multiply it by a vector, the length of the vector doesn't change. It is like taking a rigid object and moving it around without stretching or squeezing it. In an orthogonal matrix, the columns (or rows) are orthogonal to each other This property ensures that when you multiply the matrix by its transpose, you get the identity matrix, indicating that the transformation and its inverse cancel each other out.

A diagonal matrix is a square matrix where all off-diagonal elements are zero, meaning that all non-zero elements lie on the main diagonal from the top left to the bottom right representing the singular values. Diagonal matrices are often represented as in the Figure 13.

$$\operatorname{diag} x = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_n \end{bmatrix}$$

Figure 13 diagonal matrix example

In the context of audio watermarking, SVD plays a crucial role in embedding and detecting watermarks while ensuring robustness and imperceptibility. During embedding, the singular values of specific frequency bands or time segments obtained through SVD can be modified to embed the watermark while minimizing perceptual distortion. During detection, the extracted singular values are compared with the original watermarked singular values to detect the presence of the watermark. [39]

3.7 Performance Evaluation Factors

performance evaluation factors play a crucial role in evaluating the perceptual transparency, robustness, and quality of watermarking algorithms. Here are some key factors used for performance evaluation.

3.7.1 Robustness

Robustness measures the ability of a watermarking algorithm to withstand various signal processing operations and attacks without significant degradation of the embedded watermark.by using evaluation metrics that measures the resistance to common attacks such as compression, noise addition, filtering.

3.7.2 Signal-to-Noise Ratio (SNR)

Signal-to-Noise Ratio (SNR) is a commonly used metric in audio signal processing, that measures the ratio of the signal power to the noise power in the watermarked audio signal. The equation is as shown in the EQ (3.1) [40]. A higher SNR values indicate that the watermark is more robust and less affected by noise or other disturbances, while lower SNR values indicate that the watermark may be more susceptible to degradation or distortion. that gives us an idea about the perceptual transparency of the watermark to the human auditory system [41].

$$SNR(db) = 10\ log\ 10\ \frac{\sum_n A_n^2}{\sum_n (A_n - A/_n)} \qquad \textbf{(3.1)}$$

3.7.3 Bit Error Rate (BER)

Bit Error Rate (BER) measures the robustness by comparing the extracted watermark bits with the original watermark data. So (BER) is the ratio of the number of incorrectly detected bits to the total number of embedded bits. It is typically expressed as a percentage as you can see in the EQ (3.2). A lower BER value indicates better performance, as it signifies fewer errors in detecting the watermark. While, a higher BER value indicates a higher rate of errors in detecting the watermark, which may lead to inaccuracies or false detections.

$$BER = \frac{Number\ of\ incorrect\ bits}{Total\ number\ of\ bits} \qquad \textbf{(3.2)}$$

3.7 Different Types of Attacks

3.7.1 Noise Addition

Noise addition refers to the process of introducing additional noise to the audio signal in an attempt to damage the embedded watermark. This attack aims to make the watermark less detectable by watermark extraction algorithms. To mitigate the noise addition attack, watermarking algorithms must be robust, which ensures that the watermark remains detectable even in the presence of added noise.

3.7.2 Filtering

Filtering is a common attack on audio watermarking where filters are applied to the audio signal to degrade or remove the embedded watermark. Filters can be designed to specifically target the frequency ranges where the watermark is embedded. The basic high pass and low pass filters or Adaptive Filtering can be used to achieve these types of attacks [42].

3.7.3 Compression

Using lossy compression algorithms (e.g., MP3, AAC) to compress the audio signal, which can degrade the watermark and make it more challenging to detect.

3.7.4 Amplitude Scaling

Attackers adjust the amplitude (volume) of the audio signal, either increasing or decreasing it. This modification can affect the watermark because watermark embedding often relies on specific amplitude levels for detection.

3.7.5 Pitch Shifting

Pitch shifting changes the frequencies of the audio signal while preserving its duration. This alteration can damage the watermark's integrity because watermark embedding often relies on specific frequency components or modulation techniques for detection.

3.8 Conclusion

In this chapter, we get into the core subject of audio watermarking, covering its essential components and properties. We began with a general schema of digital watermarking. Then we explored the main properties of a digital audio watermarking system, highlighting the critical aspects such as robustness, imperceptibility, and capacity.

Then we classified the audio watermarking system based on different criteria, including watermarking domain, embedding methods, robustness, extraction methods, and application domains. And we discussed each one in details.

After that We examined different audio watermarking domains and methods, including LSB, echo hiding, DCT, DWT, and SVD, highlighting their unique advantages and applications. then, we addressed various types of attacks on audio watermarking systems, to clarify the need for Performance Evaluation Factors and resilient watermarking techniques.

This comprehensive overview of audio watermarking principles and practices sets the stage for our proposed hybrid DWT-SVD approach, aimed at enhancing the robustness and imperceptibility of audio watermarks for copyright protection.

Chapter 3 Experiments and Results Discussion

Chapter 3: Experiments and Results Discussion

4.1 Introduction

After discussing all these topics in previous chapters, now in this chapter we are going to discuss our main project. Based on the various methods available for digital audio watermarking and our researches, we chose a non-blind hybrid audio watermarking approach which is basically a combination of DWT and SVD methods. In a way to improve the imperceptibility and robustness of our audio watermarking algorithm. This algorithm is more suitable for the copyrights protecting field. The results and details of embedding method and extraction method we used is explained in this chapter.

4.2 Environment

4.2.1 Python

Python, with its extensive libraries and features and the ease of use, has become an indispensable tool in the field of digital signal processing, including audio watermarking. One of the primary advantages of using Python for audio watermarking is the availability of specialized libraries such as NumPy and SciPy, PyWavelets that provide robust tools for numerical computations and signal processing, enabling efficient manipulation of audio signals. And matplotlib library for visualizations.

Figure 14 Python logo

4.2.2 Jupyter

Jupyter Notebooks, an open-source web application, have changes the way data scientists, researchers, and developers interact with and present their code, especially in fields like audio watermarking. Combining the power of Python with an interactive, user-friendly interface, Jupyter Notebooks offer a dynamic environment for developing, testing, and demonstrating audio watermarking techniques.

Figure 15 jupyter logo

4.2.3 PyQt5

PyQt5 is a set of Python bindings for Qt libraries, which are widely used for developing graphical user interfaces (GUIs). It is a powerful tool that allows developers to create cross-platform applications with a native and modern look.

Figure 16 PyQt5 logo

4.3 Audio Watermark Algorithm Using DWT-SVD

4.3.1 Embedding Method

First, we have to generate the watermark. if it is a string we convert it to binary sequence, but if it is an image we need to process the image and change it from two-dimension to one-dimension and then convert it to binary sequence. After that we have to make sure that the length of the audio is sufficient for the image size. Suppose A is the audio data so, A length should be equal or bigger than the watermark bits. The watermark embedding process is illustrated in Figure *17*.

Here are the embedding steps in detail:

Step 1 Audio signal Segmentation: the audio signal is divided into segments or frames with fixed sizes. suppose A is the original audio signal, A can be represented as follows A = {A1, A2, A3 … Ai} where A is audio data and Ai the segment.

Step 2 2-level DWT: 2-level DWT is applied on each segment, so we can get the detail component (high frequency) and approximation component (low frequency) of the audio. Which are represented as cA_i2、cD_i2、 cDi1.

Step 3 Selecting low frequency coefficients: we choose approximation components to embed the watermark, because embedding in low frequency components gives us better imperceptibility then working on high frequency.

Step 4 SVD: we apply SVD on each approximation component cA_i2 so we can get diagonal matrix S. this matrix is where we are going to embed the watermark.

Step 5 Watermark Embedding bit: the way of embedding one bit of the watermark is by modifying only the value of $S_{(1,1)}$ of the matrix S. In a blind watermarking algorithm, we need to embed the watermark in a way that allows extraction without requiring the original audio signal. To achieve this, we calculate a factor based on $S_{(1,1)}$ and $S_{(2,2)}$ as follows:

$$\mathbf{Factor} = \mathbf{S_{(1,1)}} \ / \ (\ \mathbf{S_{(2,2)}} * \boldsymbol{\alpha}\ ) \qquad \mathbf{(4.1)}$$

We adjust this factor according to the watermark bit:

- If the watermark bit is 1, the factor should be odd. If it is even, we add 1 to make it odd.

- If the watermark bit is 0, the factor should be even. If it is odd, we add 1 to make it even.

After calculating the right factor The method of changing $S_{(1,1)}$ is:

$$S_{(1,1)} = S_{(1,1)} + \beta * (S_{(2,2)} * \alpha * \text{factor} - S_{(1,1)}) \qquad (4.2)$$

Or

$$S_{(1,1)} = S_{(1,1)} + \beta * (S_{(2,2)} * \alpha * (\text{factor} + 1) - S_{(1,1)}) \qquad (4.3)$$

α: Watermarking Strength Parameter **β:** Control Parameter

Step 6 Inverse SVD: apply inverse SVD on the modified matrix S to get the new approximation components cA_i2'.

Step 7 inverse DWT: to get the new watermarked segments we apply the inverse DWT on the new approximate component cA_i2', and the detail components cD_i2, cDi1. After that we combine all the segments together to struct the final watermarked audio signal.

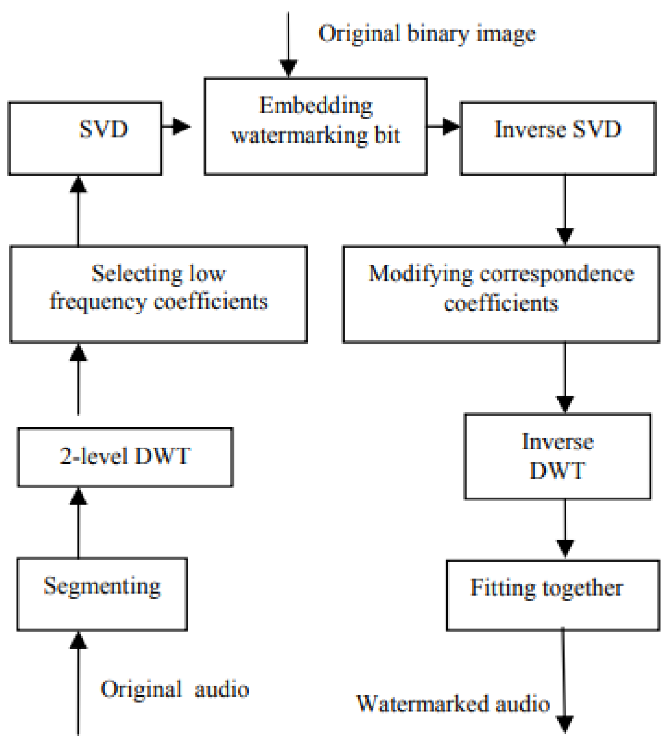


Figure 17 The watermark embedding process

4.3.2 Extracting Method

The extraction process does not require the original host audio signal. It is almost the reverse of the embedding process. Here are the steps:

Step 1 Audio signal Segmentation: the audio signal is divided into segments or frames with the same segment sizes of embedding process.

Step 2 2-Level DWT: 2-level DWT is applied on each segment, to get the approximation components (low frequency) of the audio where we inserted the watermark.

Step 3 SVD: In order to retrieve watermark bits, we need to get singular matrix by singular value decomposition for the approximation component.

Step 4 Extracting watermark bit: extracting the watermark bit is based on calculating the factor with the same chosen singular values $S_{(1,1)}$ and $S_{(2,2)}$ of the audio. Then take out a value from the diagonal matrix S.

Step 5: after extracting all the values of watermark image from the diagonal matrix, we turn the one-dimensional vector of embedded image into two-dimensional matrix to get the image.

4.4 Results

In this section, we present the results and experimentations of our proposed hybrid DWT-SVD audio watermarking approach for copyright protection. We will describe the interface and the experimental setup. Various metrics will be employed to assess the performance of the system under different conditions and types of attacks. Through a series of imperceptibility, capacity and robustness tests, we aim to demonstrate how our approach meets the critical requirements of a robust audio watermarking system, ensuring the protection of copyrights in digital audio content.

4.4.1 The Interface

Our application interface is designed to be user-friendly and efficient. Users can easily add their audio files through a brows button. Once the audio is uploaded, they have the option to embed a watermark, which can be either an image or text, into the audio file ensuring quick and effective embedding as you can see in (Figure 18).

In the extraction part, users can select various attack types to test the robustness and imperceptibility of the watermark. The application then displays the results of SNR, BER, and NC along with the extracted watermark (Figure 19). providing feedback on the watermark's robustness under different conditions.

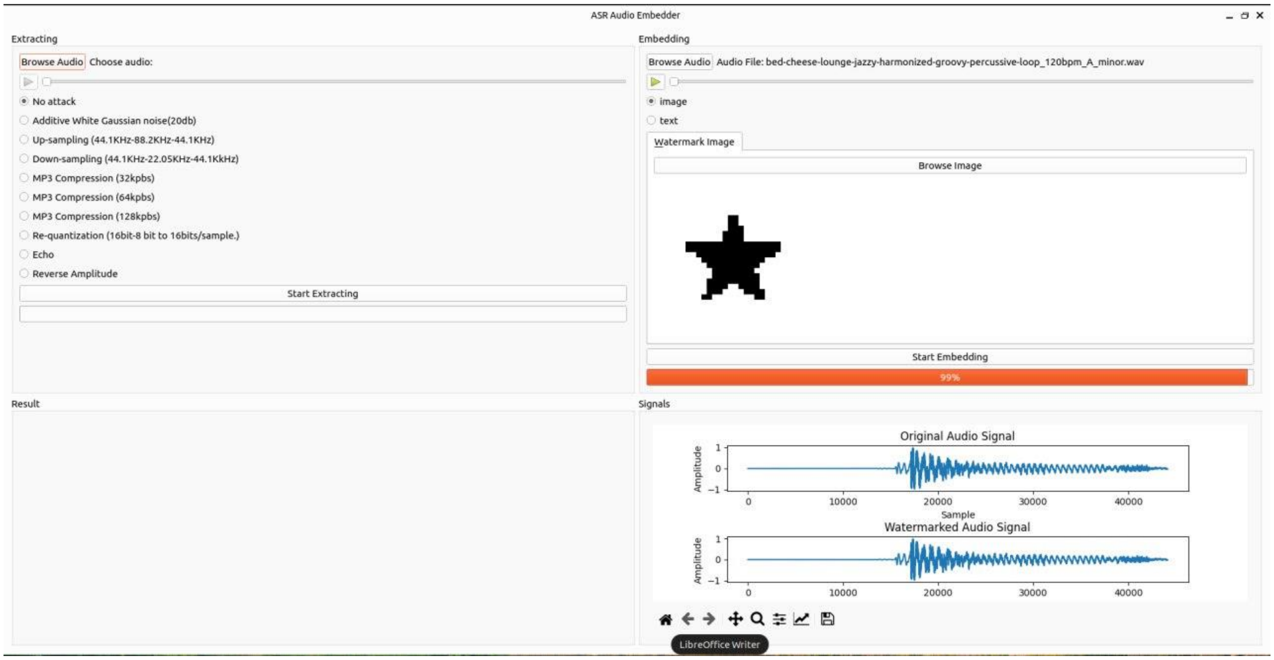


Figure 18 the interface in embedding process

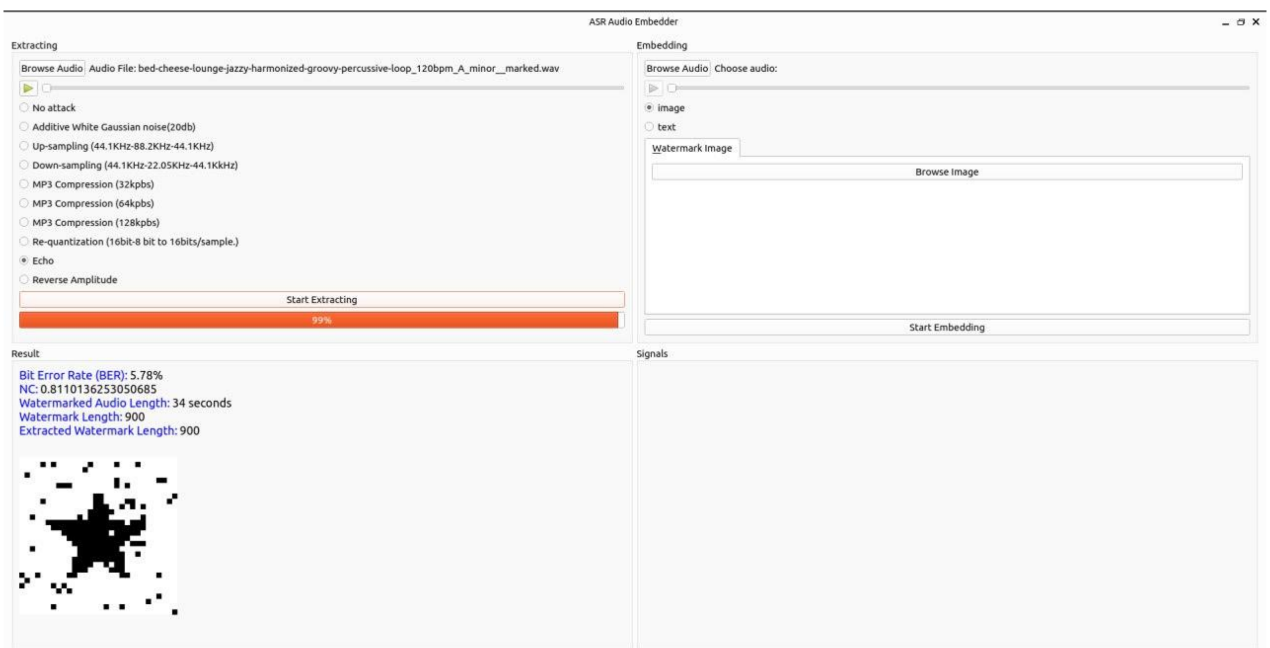


Figure 19 the interface in extraction process

4.4.2 Capacity Test

Capacity in audio watermarking refers to the amount of data that can be embedded within an audio signal without compromising imperceptibility and robustness. Finding the right capacity is a balancing act and it depends on the priorities of the algorithm. So, in order to get the best Imperceptibility and Robustness we did this calculation: 44100 is the number of samples in 1 second, and we embedded 1 bit in each 900 sample so the final result is: C =44100/900=49 bit/s. This result is Suitable for embedding simple watermarks.

4.4.3 Imperceptibility Test

Imperceptibility is a critical attribute in audio watermarking, as it ensures that the embedded watermark does not affect the quality of the audio.

To evaluate the imperceptibility of the watermarked audio we are using Signal-to-Noise Ratio (**SNR**). which is a commonly used metric in audio signal processing, that measures the ratio of the signal power to the noise power in the watermarked audio signal. The results of SNR test are shown in the table below

Table 3 .

Table 3 SNR test results

Audio name	Audio Type	SNR
S_audio1	Speech	30.66
S_audio2	Speech	27.89
M_audio1	music	30.02
M_audio2	music	29.13

Table 4 Comparison of proposed algorithm SNR with previous work

Author	Method	SNR
Bender et al. [31]	Echo	21.47
Al-Haj et al. [34]	DWT	28.55
CAI Yong-Mei. [43]	Blind DWT-SVD	20.7
Proposed	Blind DWT-SVD	29

The results that we get after SNR test were between 27 dB and 30 dB. It is considered as good imperceptibility. compared to existing algorithms our algorithm is better than the most and its better in security because it is a blind dwt-svd.

The results changed based on various factors like audio type, watermark length and audio length.

4.4.3 Robustness Test

In this section, we will evaluate how well our hybrid DWT-SVD audio watermarking algorithm withstands various common attacks that attempt to degrade or remove the watermark. This involves subjecting the watermarked audio to different types of manipulations and evaluate the ability of the algorithm to detect and extract the watermark accurately.

We employed two key metrics for this evaluation Bit Error Rate (BER) and Normalized Correlation (NC).

- **BER**: This metric quantifies the number of bits in the extracted watermark that differ from the original watermark. A lower BER indicates higher robustness.
- **NC**: This metric measures the similarity between the original watermark and the extracted watermark. An NC value close to 1 indicates a high degree of similarity.

We tested two different types of audio music and speech The results are shown in the two tables below (Table 5,Table 6).

Table 5 WATERMARK DETECTION RESULTS AGAINST COMMON ATTACKS (Music Audio)

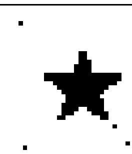
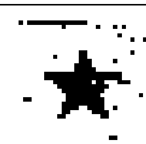
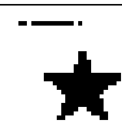
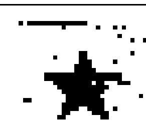
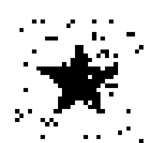
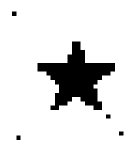
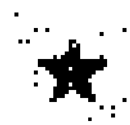
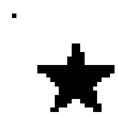
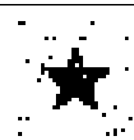
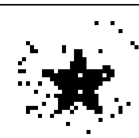
Attack Type	**Results**		
	BER	NC	Watermark image
Up-sampling	0.00%	1	
Down-sampling	0.00%	1	
MP3 Compression (128kpbs)	0.11%	0.9956	
MP3 Compression (64kpbs)	1.11%	0.9577	
MP3 Compression (32kpbs)	5.22%	0.8257	
Reverse amplitude	1.44%	0.9466	
Additive White Gaussian noise	4.44%	0.8456	
Echo hiding	5.78%	0.8110	

Table 6 WATERMARK DETECTION RESULTS AGAINST COMMON ATTACKS (SPEETCH Audio)

Attack Type	**Results**		
	BER	NC	Watermark image
Up-sampling	0.00%	1	
Down-sampling	0.00%	1	
MP3 Compression (128kpbs)	0.00%	1	
MP3 Compression (64kpbs)	0.44%	0.9828	
MP3 Compression (32kpbs)	2.44%	0.9101	
Reverse amplitude	0.11%	0.9956	
Additive White Gaussian noise	3.00%	0.8921	
Echo hiding	5.33%	0.8227	

the general ranges of BER results, categorized from poor to excellent:

Poor Robustness (BER > 10%)

Fair Robustness (5% < BER ≤ 10%)

Good Robustness (1% < BER ≤ 5%)

Very Good Robustness (0.1% < BER ≤ 1%)

Excellent Robustness (BER ≤ 0.1%)

By comparing the ranges against the results, our proposed blind DWT-SVD audio watermarking algorithm did good performance against various attack scenarios. For normal attacks such as Up-sampling, compression, the algorithm maintained a low Bit Error Rate (BER) and high Normalized Correlation (NC), demonstrating reliable watermark extraction.

Under advanced attacks the algorithm performed well, maintaining an acceptable BER and NC, indicating that the watermark remained detectable and verifiable even under severe conditions. Compared to several existing methods, our approach consistently showed a good performance, the only challenging part is enhancing the capacity while maintaining imperceptibility and robustness but considering that the algorithm is blind it is acceptable. These results confirm the robustness and effectiveness of our DWT-SVD algorithm, making it a strong candidate for copyrights audio watermarking applications.

4.5 Conclusion

In this chapter, we explained and evaluated the performance of our blind audio watermarking algorithm based on DWT-SVD through a series of experiments focusing on imperceptibility, robustness, and capacity.

General Conclusion and Perspectives

The need for robust audio watermarking has become increasingly critical in today's digital age, where unauthorized distribution and piracy of audio content are out of hand. Protecting copyrights, ensuring content authenticity, and managing digital rights is the most important for creators, distributors, and consumers alike. This project was taken to address these pressing needs by developing an advanced audio watermarking solution capable of embedding information within audio signals in a manner that is imperceptible to listeners yet detectable by specialized algorithms.

In this project, we developed and evaluated a blind hybrid DWT-SVD audio watermarking algorithm aimed at providing robust copyright protection. Our primary goal was to create an audio watermarking solution that strikes an optimal balance between imperceptibility, robustness, and capacity. We chose to embed the watermark in the low-frequency coefficients obtained through the Discrete Wavelet Transform (DWT) because these coefficients are less susceptible to common audio processing and attacks, thereby enhancing the robustness of the watermark. The Singular Value Decomposition (SVD) was then applied to these coefficients to further improve resilience. To evaluate our results, we used several metrics including Signal-to-Noise Ratio (SNR) to measure audio quality, Bit Error Rate (BER) to assess the accuracy of the extracted watermark, and Normalized Correlation (NC) to determine the similarity between the original and extracted watermarks.

Throughout the development process, we encountered several challenges. Balancing the imperceptibility and robustness and capacity was a hard task, requiring carful adjustment of embedding parameters to ensure the watermark's detectability without degrading audio quality. Handling diverse audio content necessitated adaptable strategies to different audio characteristics. Additionally, ensuring robustness against advanced and evolving attacks remains an ongoing challenge.

Looking forward, we see significant potential for further enhancing our watermarking algorithm through the integration of artificial intelligence (AI). AI and machine learning techniques can dynamically optimize the embedding and extraction processes, achieving a better balance between capacity, imperceptibility, and robustness. By analyzing audio characteristics in real-time, AI can adjust embedding parameters to maximize watermark strength without compromising audio quality. This adaptability can also improve the algorithm's resilience to diverse and evolving attacks, ensuring the watermark remains detectable and intact. Through these advancements, AI integration promises to significantly enhance the overall performance and effectiveness of our audio watermarking solution.

References

[1] I. J. Cox, M. L. Miller, A. B. Jeffrey, J. Fridrich and T. Kalker, Digital Watermarking and Steganography., San Francisco, USA.: Morgan Kaufmann Publishers., 2008.

[2] L. C. P. VaddeSeethaRama Rao, "MATLAB Implementation of Audio Steganography for Secure Data Transmission," *International Journal of Scientific Engineering and Technology Research,* vol. 6, no. 31, pp. 6069-6079, 2017.

[3] F. Cayre., Cryptographie, stéganographie et tatouage: des secrets partagés., 2008.

[4] R. v. Schyndel, A. Tirkel and C.F. Osborne, "A digital watermark," in *Proceedings of 1st International Conference on Image Processing,* Austin, TX, USA, November 1994.

[5] G. Widmer, D. Rocchesso, V. Välimäki, C. Erkut, F. Gouyon and D. Pressnitzer, "Sound and Music Computing: Research Trends and Some Key Issues," *Journal of New Music Research,* vol. 36, pp. 169-184, 2007.

[6] Morton and David, Sound recording : the life story of a technology, Baltimore, Md. : Johns Hopkins University Press, 1964.

[7] Pohlmann and K. C, The compact disc handbook, Madison, WI : A-R Editions, 1992.

[8] B. C. J. Moore, The nature of sound and the structure and function of the auditory , in An introduction to the psychology of hearing, San Diego, Calif. : Academic Press, 1997.

[9] K. c. Pohlmann, Principles of digital audio, New York : McGraw-Hill, 2005.

[10] U. Zölzer, Digital Audio Signal Processing, John Wiley & Sons, 2022.

[11] "Format de fichier audio," Wikipédia, [Online]. Available: https://fr.wikipedia.org/wiki/Format_de_fichier_audio. [Accessed 2/ 5/ 2024].

[12] "GENERAL knowledg of AC3 Formate," Apowersoft, [Online]. Available: https://www.apowersoft.com/what-is-ac3-format.html. [Accessed 10 5 2024].

[13] J. Watkinson, The MPEG handbook : MPEG-1, MPEG-2, MPEG-4, New York: Oxford ; Burlington, MA : Elsevier/Focal Press, 2004.

[14] "Ogg Vorbis," xiph.org, [Online]. Available: https://xiph.org/vorbis/. [Accessed 10/ 5/ 2024].

[15] N. Lalitha, R. C. Srinivasa and P. J. Sree, "A REVIEW OF DIGITAL AUDIO WATERMARKING SCHEMES," *Journal of critical reviews,* vol. 7, p. 870–880.

[16] Nguyen, Philippe and S. Baudry, "Le tatouage de données audiovisuelles." Les Cahiers du numérique", Lavoisier, 2003.

[17] P. Singh and R S Chadha, "A Survey of Digital Watermarking Techniques, Applications and Attacks," *International Journal of Engineering and Innovative Technology,* vol. 2, no. 9, 2013.

[18] C. S. WOO, "Digital image watermarking methods for copyright protection and authentication," 2002.

[19] N. Singh, M. Jain and S. Sharma, "A Survey of Digital Watermarking Techniques," *International Journal of Modern Communication Technologies & Research,* vol. 1, no. 6, 2013.

[20] M.Arnold, S.Wolthusen and M. Schmucker, Techniques and applications of digital watermarking and content protection, Artech House Publishers, 2003.

[21] I. Cox, M. Miller and J. Bloom, Digital watermarking, SanFrancisco, USA: Morgan Kaufmann Publishers, 2002.

[22] K. Loukhaoukha, "Tatouage numérique des images dans le domaine des on delettes basé sur la décomposition en valeurs singulières et l'optimisation multi objective. PhD thesis,Université Laval .," 2010.

[23] H. Xiong-Hua, "A complex cepstrum fragile audio watermarking algo rithm based on quantization," in *3rd International Conference on Genetic and Evolutionary Computing (WGEC),* Guilin, China, 2009.

[24] F. Y. Shih, Digital watermarking and steganography: fundamentals and techniques., CRC Press, 2017, p. 39–4.

[25] K. Adhiya and S. Patil, "Hiding text in audio using LSB based steganography," *Inf Knowl Manag,* 2012.

[26] S. Katzenbeisser and F. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, 2000.

[27] W. Kong, B. Yang, D. Wu, D. X. Wu and X. Niu, "svd based blind video watermarking," in *International Conference on Innovative Computing, Information and Control*, China, 2006..

[28] C. Nedeljko and S. Tapio, Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, IGI Global, 2007.

[29] S. Borra, R. Thanki and N. Dey, Digital Image Watermarking Theoretical and Computational Advances, USA: CRC Press, 2018.

[30] A. KP and P. SA, "Hiding text in audio using LSB based steganography.," *Information and Knowledge Management,* vol. 2, p. 8–15, 2012.

[31] D. Gruhl, A. Lu and W. Bender, "Echo hiding," *Proceeding of Inforomation Hiding Workshop,* p. 295–315, 1996.

[32] C. Podilchuk and E. Delp, "Digital watermarking: algorithms and applications," *IEEE Signal Process,* vol. 18, no. 4, p. 33–46, 2001.

[33] S. D. Lin, S.-C. Shie and J. Guo, "Improving the robustness of DCT-based image watermarking against JPEG compression," *Computer Standards & Interfaces,* vol. 32, p. 54–60, 2010.

[34] A.-H. A, M. AA and B. L, "DWT-based audio watermarking," *Int Arab J Inf Technol,* vol. 8, no. 3, p. 326–333, 2011.

[35] M. Begum and M. S. Uddin, "Digital Image Watermarking Techniques: A Review," 2020.

[36] C. JS and P. M, "Improved technique for audio watermarking based on discrete wavelet transform.," *Int J Eng Adv Technol,* vol. 2, no. 5, 2013.

[37] H. Tao, L. Chongmin, J. M. Zain and A. N. Abdalla, "Robust Image Watermarking Theories and Techniques Review," *J. Appl. Res. Technol,* vol. 12, 2014.

[38] B. K, S. I. V. and A. Das, "A new audio watermarking scheme based on singular value decomposition and quantization 30," p. 915–927, 2011.

[39] W. J, H. R and T. J, "A robust audio watermarking scheme based on reduced singular value decomposition and distortion removal.," *J Signal Process,* vol. 91, no. 8, p. 1693–1708, 2011.

[40] D. KA, "Imperceptible and robust DWT-SVD-based digital audio watermarking algorithm.," *Journal of Software Engineering and Applications,* pp. 859-871, 2014.

[41] B. JG and S. JA, "A perceptual audio quality measurement based on a psychoacoustic sound representation.," *J Audio Eng Soc,* p. 963–972, 1992.

[42] Tewari, T. K., V. Saxena and J. P. Gupta, "Audio watermarking: Current state of art and future objectives.," *International Journal of Digital Content Technology and its Applicat,* vol. 5, no. 7, 2011.

[43] C. Yong-mei, G. Wen-qiang and D. Hai-yang, "An Audio Blind Watermarking Scheme Based on DWT-SVD," *JOURNAL OF SOFTWARE,* vol. 8, no. 7, 2013.