

AUDIO WATERMARKING FOR COPYRIGHTS

A PROJECT REPORT

Submitted by

SARAVANAN D (2021105047)

SINDHU S U (2021105050)

ASHWIN S (2021105005)

In partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING



COLLEGE OF ENGINEERING GUINDY,

ANNA UNIVERSITY, CHENNAI 600 025

NOVEMBER 2024

BONAFIDE CERTIFICATE

Certified that this project report “**AUDIO WATERMARKING FOR COPYRIGHTS**” is the Bonafide work of “**SARAVANAN D (2021105047), ASHWIN S (2021105005), SINDHU S U (2021105050)** who carried out the project work under my supervision.

SIGNATURE

Dr. M.A BHAGYAVENI

HEAD OF THE DEPARTMENT

PROFESSOR

Department of Electronics and
Communication Engineering

College of Engineering Guindy,

Anna University,

Chennai – 600 025

SIGNATURE

Dr. K GUNASEELAN

SUPERVISOR

PROFESSOR

Department of Electronics and
Communication Engineering

College of Engineering Guindy,

Anna University,

Chennai – 600 025

ACKNOWLEDGEMENT

We express our wholehearted gratitude to our dynamic and zestful Head of the Department Dr. M.A BHAGYAVENI, Professor, Department of Electronics & Communication Engineering for her continued and heartening support throughout the project.

We manifest our sincere thanks and regards to our project supervisor, Dr. K GUNASEELAN, Professor, Department of Electronics and Communication Engineering for her ardent support, patience, valuable guidance, technical expertise and encouragement in our project.

We also are supremely grateful to the teaching and non-teaching staff of the Department of Electronics and Communication Engineering, for their kind help and co-operation during the course of our project.

Last but not the least, we would not have made this project without the support from our beloved family and friends.

SARAVANAN D
(2021105047)

SINDHU S U
(2021105050)

ASHWIN S
(2021105005)

ABSTRACT

This project presents the development of an audio watermarking system capable of embedding image data into audio signals while preserving both perceptual transparency and robustness. Leveraging psychoacoustic principles, the system exploits the characteristics of the human auditory system to ensure that the embedded watermark remains imperceptible to listeners. The audio is divided into small frames and processed using the Fast Fourier Transform (FFT), enabling precise frequency-domain manipulation. Psychoacoustic masking is employed to identify optimal embedding locations, ensuring the watermark is introduced into less perceptible regions of the audio spectrum. The system encodes image data into the frequency components of the audio signal, maintaining a balance between robustness and transparency. The extraction process is designed to recover the embedded image accurately, even when the watermarked audio undergoes various signal processing operations. Robustness is a key focus, with the system demonstrating resilience to common attacks such as MP3 compression, noise addition, filtering, and time-scaling. Performance evaluation employs comprehensive metrics to assess audio and image quality. Normalized Correlation (NC) is used to measure audio imperceptibility and embedding accuracy, while Peak Signal-to-Noise Ratio (PSNR) evaluates the fidelity of the extracted image. Results indicate that the system achieves good imperceptibility, with watermarked audio being virtually indistinguishable from the original, while the embedded image retains its integrity under adverse conditions. With its ability to ensure robust performance, significant data capacity, and minimal perceptual distortion, the proposed system is well-suited for multimedia security applications. Potential use cases include digital rights management (DRM), copyright protection, and secure communication, offering a reliable and efficient solution for safeguarding content in an increasingly digital world.

திட்டப்பணிச்சுருக்கம்

இந்த திட்டம் உருவப்படத் தரவுகளை ஆடியோ சிக்னல்களில் உள்ளமைக்கச் செய்து, அதே நேரத்தில் கேட்பதற்குத் தகுந்த சிதைவற்ற தன்மை மற்றும் வலிமையையும் பேணும் மேம்பட்ட ஆடியோ நீர்முத்திரை அமைப்பாக உருவாக்கப்பட்டுள்ளது. மனித செவிவழி மண்டலத்தின் (HAS) பண்புகளை பயன்படுத்தும் சைக்கோஅகோஸ்டிக் கொள்கைகள் மூலம் நீர்முத்திரை கேட்பவர்களுக்கு உணரப்படாததாக உறுதி செய்யப்படுகிறது. ஆடியோவை சிறு பிரேம்களாகப் பிரித்து, வேகமான நான்கு வேறு மாற்றம் (FFT) முறையில் அதிர்வெண் பகுப்பாய்வு செய்யப்படுகிறது.

சைக்கோஅகோஸ்டிக் மாஸ்கிங் மூலம், குறைவாகக் கேட்கக்கூடிய பகுதிகளில் நீர்முத்திரையைச் சேர்க்கும் இடங்கள் தீர்மானிக்கப்படுகிறது. நீர்முத்திரை, ஆடியோவின் அதிர்வெண் கூறுகளில் உருவப்படத் தரவுகளைச் செருகுகிறது, சிதைவற்ற வெளிப்படையும் வலிமையும் பேணும். நீர்முத்திரை நீக்குதல் மற்றும் மீட்டமைக்கல் செயல்முறை, MP3 சுருக்கம், சத்தம் சேர்த்தல், மற்றும் நேர அளவீடு போன்ற தாக்குதல்களுக்குப் பிறகும் உருவப்படத்தை துல்லியமாக மீட்டமைக்க வடிவமைக்கப்பட்டுள்ளது. செயல்திறன் மதிப்பீட்டுக்கு, சிக்னல்-நான் ஒலி விகிதம் (SNR), சாதாரண ஒப்புமை (NC) மற்றும் உச்ச சிக்னல்-நான் ஒலி விகிதம் (PSNR) போன்ற அளவீடுகள் பயன்படுத்தப்படுகின்றன. நீர்முத்திரைச் செய்த ஆடியோவின் தரம் மிகச்சிறப்பாகவும், அதேசமயம் உருவப்படம் பாசாங்கான சூழலிலும் அதன் தரத்தைப் பேணுகிறது. இதன் மூலம், டிஜிட்டல் உரிமைகள் மேலாண்மை (DRM), காப்புரிமை பாதுகாப்பு, மற்றும் பாதுகாப்பான தகவல்தொடர்பில் பயன்படுத்தத்தகுந்த பாதுகாப்பான தீர்வாக இது விளங்குகிறது.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT (ENGLISH)	
	ABSTRACT (TAMIL)	
1	INTRODUCTION	1
	1.1 FUNDEMETALS OF DIGITAL AUDIO WATERMARKING	2
	1.1.1 PROPERTIES OF THE HUMAN AUDITORY SYSTEM	
	1.2 PSYCHOACAUSTIC PRINCIIPLES	3
	1.2.1 MASKING THRESHOLD	
	1.2.2 CRITICAL BANDS AND FREQUENCY ANALYSIS	
	1.3 DIGITAL SIGNAL PROCESSING FUNDEMENTALS	4
	1.3.1 FRAME PROCESSING	
	1.3.2 FREQUENCY DOMAIN REPRESENTATION	
	1.4 IMAGE PROCESSING FUNDEMENTALS	6
	1.4.1 DIGITAL IMAGE REPRESENTATION	
	1.4.2 IMAGE QUALITY METRICS	
	1.5 RESEARCH OBJECTIVES	6
	1.6 SIGNIFICANCE OF STUDY	6
	1.7 SUMMARY	6
2	LITERATURE REVIEW	7

3	METHODOLOGY	10
	3.1 TRANSMITTER SIDE METHODOLOGY	10
	3.1.1 AUDIO PROCESSING AND FRAME ANALYSIS	
	3.1.2 PSYCHOACAUSTIC MASKING IMPLEMENTATION	
	3.1.3 IMAGE WATERMARK EMBEDDING	
	3.2 RECEIVER SIDE FRAMEWORK	13
	3.2.1 WATERMARK EXTRACTION PROCESS	
	3.2.2 IMAGE RECONSTRUCTION	
	3.2.3 PERFORMANCE METRICS AND EVALUATION	
	3.3 SUMMARY	14
4	PARAMETERS ANALYSIS AND PERFORMANCE EVALUATION	15
	4.1 PERFORMANCE PARAMETERS	15
	4.1.1 AUDIO QUALITY METRICS	
	4.1.2 IMAGE QUALITY METRICS	
	4.1.3 CAPACITY ANALYSIS	
	4.2 COMPARATIVE ANALYSIS	17
	4.3 SYSTEM OPTIMIZATION	18
	4.3.1 FRAME SIZE ANALYSIS	
	4.3.2 PSYCHOACOUSTIC MODEL OPTIMIZATION	
	4.4 PERFORMANCE UNDER DIFFERENT CONDITIONS	18

4.4.1	AUDIO CONTENT TYPE ANALYSIS	
4.4.2	IMAGE CONTENT ANALYSIS	
4.5	RECOMMENDATIONS FOR IMPROVEMENT	19
4.6	SUMMARY	20
5	RESULTS AND DISCUSSION	21
5.1	IMPACT OF DIMENSIONAL MISMATCH ON WATERMARK EXTRACTION	21
5.2	NORMALIZED CORRELATION ANALYSIS	22
5.2.1	NORMALIZED CORRELATION ANALYSIS – AUDIO 1 (0.9992)	
5.2.2	NORMALIZED CORRELATION ANALYSIS – AUDIO 2 (-0.0018)	
5.2.3	NORMALIZED CORRELATION ANALYSIS – AUDIO 3 (0.0006)	23
5.3	PSNR ANALYSIS	
5.3.1	PSNR ANALYSIS OF IMAGE 1 EMBEDDED IN AUDIO 1	
5.3.2	PSNR ANALYSIS OF IMAGE 1 EMBEDDED IN AUDIO 1	
5.3.3	PSNR ANALYSIS OF IMAGE 1 EMBEDDED IN AUDIO 1	
5.3.4	COMPARISON TABLE BETWEEN DIFFERENT AUDIOS AND IMAGES COMBINATIONS	
5.3.5	OUTPUT IMAGES	
5.4	TECHNICAL IMPLICATIONS AND PRACTICAL APPLICATIONS	26

	5.5 RECOMMENDATIONS FOR IMPROVEMENT	26
	5.6 FUTURE CONSIDERATIONS AND ADAPTATIONS	27
6	CONCLUSION	28
	REFERENCES	29

CHAPTER 1

Introduction

The digital age has ushered in new challenges related to the security, authenticity, and privacy of multimedia content. One such challenge is the protection of audio content from unauthorized use or modification, particularly in fields like music, entertainment, and broadcasting. Audio watermarking has emerged as a powerful technique to address these challenges by embedding hidden information—referred to as a "watermark"—into an audio signal. The primary objective of audio watermarking is to embed information that is imperceptible to the human auditory system, yet robust enough to withstand various signal processing attacks and alterations, such as compression, noise addition, or format conversion.

This project focuses on audio watermarking with the innovative approach of embedding image data into audio signals. The idea is to develop a system that allows for the secure and imperceptible embedding of image watermarks within an audio file, preserving both the audio quality and the image integrity. The challenge lies in achieving a balance between imperceptibility, robustness, and the amount of data that can be embedded without compromising the audio's quality.

To accomplish this, the project utilizes principles of psychoacoustic masking, which exploits the characteristics of the human auditory system (HAS) to determine the masking threshold—the level at which a signal becomes inaudible in the presence of a louder sound. By embedding the watermark below this threshold, the system ensures that the watermark remains undetectable to the human ear. The watermarking process involves frame-based processing of the audio, frequency domain analysis using Fast Fourier Transform (FFT), and bit encoding based on psychoacoustic models such as critical band analysis and the Bark scale.

The project is divided into two main components: the transmitter and receiver models. The transmitter side embeds the image watermark into the audio signal, while the receiver side extracts and reconstructs the watermark from the watermarked audio. Key aspects such as the Peak Signal-to-Noise Ratio (PSNR) for image quality and normalized correlation for audio quality are used to evaluate the system's effectiveness.

In conclusion, the project aims to provide a secure, imperceptible, and robust audio watermarking system that can embed image data into audio signals, with applications in areas such as digital rights management, copyright protection, content authentication, and secure communication.

1.1 Fundamentals of Digital Audio Watermarking

Digital audio watermarking involves embedding auxiliary information within audio signals in a manner that is imperceptible to human listeners yet recoverable when needed. This technology serves crucial purposes such as copyright protection, content authentication, and covert communication. The primary challenge in developing such systems is balancing three essential aspects: imperceptibility, ensuring that the watermark does not degrade audio quality; robustness, allowing the watermark to withstand typical signal processing operations; and data capacity, referring to the amount of information that can be embedded without compromising the audio quality.

1.1.1 Properties of the Human Auditory System

The human auditory system (HAS) has distinctive characteristics that make audio watermarking feasible. One notable property is frequency sensitivity, where the typical hearing range spans from 20 Hz to 20 kHz, with the maximum sensitivity between 2 kHz and 4 kHz. The system's response to frequency-dependent amplitude sensitivity and critical band analysis within the cochlea significantly impacts the watermarking process. Additionally, temporal masking is a

phenomenon in which a strong sound can obscure weaker sounds immediately preceding or following it. This includes pre-masking, which occurs approximately 20 milliseconds before the loud sound, and post-masking, which can last between 100 and 200 milliseconds. Frequency masking also plays a role, as strong tones can mask weaker tones within the same critical band, with the effect being more pronounced toward higher frequencies. The critical bandwidth expands as frequency increases, which is essential for understanding psychoacoustic principles.

1.2 Psychoacoustic Principles

Psychoacoustic Principles refer to the study of how humans perceive sound and how this perception can be leveraged for applications such as audio watermarking. These principles are grounded in the understanding of the auditory system and its sensitivity to different sounds, frequencies, and temporal characteristics. In audio watermarking, psychoacoustic principles are used to embed information (watermarks) in a way that it remains imperceptible to the listener, while still being retrievable under specific conditions.

1.2.1 Masking Threshold

The masking threshold represents the minimum level at which a signal can be heard in the presence of a masking sound. This threshold is determined through several components, including critical band analysis, where the frequency spectrum is divided into critical bands, and the Bark scale is used for transformation and power spectrum density estimation. Masker classification involves identifying tonal maskers, which are sinusoidal components, and non-tonal maskers, which are noise-like components. These are used to determine individual masking thresholds, which are then combined to create the global masking threshold. This global threshold accounts for the absolute hearing threshold, uses non-linear superposition to incorporate various masking effects.

1.2.2 Critical Bands and Frequency Analysis

Critical bands are a cornerstone of psychoacoustic modeling, with their bandwidth increasing as the center frequency rises. Below 500 Hz, the increase is roughly linear, while above 500 Hz, it follows a logarithmic pattern. The Bark scale is a non-linear scale that aligns with human perception, consisting of 24 critical bands. The Bark scale transformation can be calculated using the formula:

$$[z(f) = 13 * \arctan(0.00076f) + 3.5 * \arctan((f/7500)^2) \setminus].$$

1.3 Digital Signal Processing Fundamentals

Digital Signal Processing Fundamentals (DSP) form the core of many modern audio, image, and video processing techniques, including those used in audio watermarking. DSP techniques are used to manipulate, analyze, and enhance signals in a digital format, facilitating tasks like filtering, encoding, and compression. Understanding these principles is crucial for implementing watermarking systems that embed data without perceptibly degrading the signal.

1.3.1 Frame Processing

The watermarking system employs frame-based processing to manage the audio signal effectively. A frame size of 512 samples is chosen, balancing frequency resolution with computational efficiency. An overlap-add method is used to ensure continuity between frames, and a Blackman window is applied to each frame to minimize spectral leakage and improve the accuracy of frequency analysis.

1.3.2 Frequency Domain Representation

Understanding the frequency domain representation of audio is critical for effective watermarking. The Fourier Transform is used to convert time-domain signals to the frequency domain, where each frame's magnitude and phase can be analyzed. Properties such as symmetry and frequency bin analysis are essential

for precise embedding. The Blackman window helps reduce spectral leakage, smoothing the time-domain signal and enhancing frequency analysis accuracy.

1.4 Image Processing Fundamentals

Image Processing Fundamentals involve techniques used to manipulate, enhance, and analyze digital images. Key processes include pixel-level processing, where individual image pixels are modified, and block-based processing, where images are divided into smaller segments for more localized manipulation. Common transformations include converting images into various color spaces (like RGB) and applying compression techniques. Image quality metrics, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), are used to evaluate the effectiveness of processing techniques. These principles are critical for embedding watermarks or performing other image-related tasks without perceptible quality degradation.

1.4.1 Digital Image Representation

The watermark image is processed at the pixel level, typically in the RGB color space with an 8-bit depth per channel. Before embedding, pixel values are converted into binary format. The image is divided into blocks of 12 pixels for systematic embedding, and the scanning process is row-wise to facilitate data insertion into audio frames.

1.4.2 Image Quality Metrics

To evaluate the quality of the watermark extraction, Peak Signal-to-Noise Ratio (PSNR) is used. This metric provides a logarithmic scale representation of image quality, calculated using the Mean Square Error (MSE) between the original and extracted images. Additionally, correlation analysis measures the similarity between the two images, with normalized correlation being a common method for comparison.

1.5 Objectives

The Primary objectives of this research are:

- To develop a robust audio watermarking system capable of embedding image data, apply psychoacoustic principles to maintain imperceptibility.
- To achieve high data embedding capacity without compromising audio quality, and design efficient extraction and recovery mechanisms.

These objectives will be validated through objective quality metrics such as PSNR and normalized correlation.

1.6 Significance of the Study

This study has significant implications for Digital Rights Management, offering methods for copyright protection, content authentication, and ownership verification. Additionally, it contributes to steganographic applications, providing secure communication channels and data hiding techniques that enhance information security. Finally, the system's design supports multimedia security, enabling content integrity verification, tampering detection, and digital forensics. The proposed framework represents an advancement in audio watermarking, incorporating image data embedding while maintaining perceptual transparency through advanced psychoacoustic modelling.

1.7 Summary

This report proposes a digital audio watermarking system that embeds image data into audio files while preserving audio quality. It leverages psychoacoustic principles like masking thresholds to ensure the watermark is imperceptible. The watermark is embedded in the frequency domain below the masking threshold and extracted using frequency analysis. Image quality is evaluated via PSNR, and audio quality through normalized correlation.

CHAPTER 2

LITERATURE REVIEW

The growing reliance on digital multimedia has necessitated the development of robust techniques for copyright protection and secure transmission of data. Digital watermarking offers an effective solution by embedding hidden information, such as ownership credentials or authentication metadata, into audio signals without perceptible degradation. Among these, FFT-based techniques have emerged as highly effective due to their ability to balance robustness, imperceptibility, and embedding capacity, making them suitable for applications in digital rights management (DRM) and secure communication. [1] proposes a dual-mode FFT-based watermarking framework that embeds binary and numeric data simultaneously using adaptive modulation techniques. By selectively embedding watermarks into FFT coefficients, the system demonstrates exceptional resistance to common audio processing attacks while maintaining high imperceptibility. This approach highlights a practical balance between watermarking efficiency and audio quality, making it a versatile solution for securing multimedia content. [2] addresses the critical issue of synchronization in audio watermarking systems, which often face vulnerabilities to desynchronization attacks such as MP3 compression and cropping. The proposed FFT-based algorithm embeds synchronization codes within FFT coefficients, ensuring alignment during extraction and resisting attacks that disrupt timing. Comparative analysis with DWT-based methods reveals the superior robustness and imperceptibility of the FFT-based approach, making it a reliable choice for scenarios requiring resilience against synchronization challenges. [3] introduces a frequency-hopping watermarking technique that embeds perceptible yet removable watermarks into the spectral domain of audio signals. Designed for applications like media previews, the method enables the watermark to be extracted with minimal residual distortion, restoring the original audio's quality.

The approach is media format-independent, compatible with lossy compression, and maintains robust watermark embedding under various signal processing attacks. Objective and subjective evaluations affirm the audio quality post-watermark removal, demonstrating its practical applicability. [4] explores an FFT-based watermarking algorithm that leverages the characteristics of the human auditory system (HAS) to embed watermark data into phase coefficients of the audio signal. By embedding imperceptible watermarks, the system achieves robustness against noise, compression, and other common audio processing techniques. The results highlight the effectiveness of psychoacoustic modeling in maintaining inaudibility while ensuring strong resistance to distortions, making this method suitable for high-quality audio watermarking. [5] presents a spread-spectrum watermarking approach integrated with singular value decomposition (SVD) to enhance the robustness and security of digital audio content. The algorithm achieves resistance to intentional and unintentional attacks while ensuring that the embedded watermark remains inaudible to listeners. This method strikes a critical balance between imperceptibility, security, and robustness, addressing the challenges of embedding and extracting watermarks in diverse attack scenarios. [6] proposes a blind audio watermarking method combining adaptive vector norm modulation (AVNM) and improved spread spectrum (ISS) schemes. The AVNM ensures flexible payload capacities ranging from 344.53 to 1033.59 bps, while ISS achieves robust self-synchronization in the FFT domain. The system also incorporates recursive FFT for effective sample-by-sample execution, ensuring resilience against attacks. Comparative analysis demonstrates its superior performance over recent techniques in terms of imperceptibility, robustness, and embedding capacity. [7] suggests a low-complexity watermarking scheme for mobile applications, leveraging Fibonacci numbers to modify FFT magnitudes. The method mathematically ensures that watermark embedding introduces minimal distortion, with an average error of 25% per sample. The scheme achieves high

capacities (700–3 kbps) and provides resilience against common signal processing attacks like echo, noise addition, filtering, and MP3 compression. The balance between simplicity, robustness, and capacity makes it a suitable solution for resource-constrained devices. [8] presents a content-adaptive watermarking technique that applies principles of music theory to embed watermarks into audio. By modifying notes that align with the musical structure, the method enhances robustness while maintaining inaudibility. These musically coherent notes remain intact after compression and distortions, allowing watermark extraction without requiring the original audio. This novel integration of music theory demonstrates potential for improving watermarking algorithms in audio applications. [9] tackles synchronization challenges and common signal processing attacks, such as time scale modification and MP3 compression, using a wavelet-based method. The algorithm identifies relatively invariant embedding regions through localized content analysis, ensuring robustness. Multiresolution decomposition in the wavelet domain enhances imperceptibility, maintaining an SNR above 30 dB. Experimental results validate its resilience against attacks while preserving perceptual transparency. [10] develops a spread-spectrum watermarking system that utilizes adaptive segmentation and multibit embedding with pseudo-noise (PN) sequences. By embedding multiple watermark bits in selected regions based on content analysis, the method achieves robustness against synchronization and signal processing attacks. Extensive simulations confirm its high imperceptibility and capacity, making it a strong candidate for applications requiring secure and high-capacity watermarking. These studies collectively emphasize the progression of audio watermarking techniques, particularly those utilizing FFT, towards achieving robust, imperceptible, and high-capacity watermarking. The integration of innovative modulation schemes, psychoacoustic principles, and content-adaptive approaches highlights the adaptability of these methods to evolving challenges in digital security.

Chapter 3

Methodology

This chapter details the methodology applied in designing and implementing an audio watermarking system that embeds image data into audio while preserving its perceptual quality. The framework is divided into two main sections: the Transmitter Side and the Receiver Side, each with specific tasks that contribute to a robust and imperceptible watermarking process. The core principle enabling this methodology is psychoacoustic masking, which ensures that data embedded into the audio remains undetectable to human ears.

3.1 Transmitter Side Methodology

The transmitter side's function is to embed an image watermark into the audio signal without perceptual degradation. This process comprises three primary components: audio processing and frame analysis, psychoacoustic masking implementation, and image watermark embedding.

3.1.1 Audio Processing and Frame Analysis

The initial step in the transmitter side methodology involves preparing the audio signal for embedding. This process starts with reading the audio file using MATLAB and selecting a 30-second segment for analysis. The segment is divided into frames of 512 samples to balance computational efficiency and frequency resolution. Each frame is windowed using a Blackman window to reduce spectral leakage during frequency analysis.

Pseudocode for audio processing:

1. Read the audio file aud4.mp3 into variables for audio data x and sampling rate fs .
2. Extract the first 30 seconds of the audio signal and ensure the signal is a column vector.

3. Calculate the total length of the signal and store it in L .
4. Define the frame size as 512 and generate a Blackman window of size $N+1$, removing the last element to match the frame size.
5. Determine the number of frames by dividing the signal length by the frame size.

3.1.2 Psychoacoustic Masking Implementation

The second step involves calculating the psychoacoustic masking threshold, crucial for ensuring that the watermark remains imperceptible. Each frame is transformed from the time domain to the frequency domain using a Fast Fourier Transform (FFT). This step enables analysis of the spectral peaks and frequency components that contribute to masking.

The `'find_mask'` function computes the masking threshold by incorporating critical band analysis, spreading functions, and the absolute threshold of hearing. This process allows identification of areas where data can be embedded without being perceptible.

Pseudocode for audio processing:

1. Call the `find_mask` function with inputs X and fs

3.1.3 Image Watermark Embedding

Embedding the image into the audio involves a bit encoding process that respects the calculated masking thresholds. The image is converted pixel by pixel to an 8-bit binary format. To make embedding manageable, the image is divided into 12-pixel blocks.

Bits are embedded into the frequency bins of each audio frame. Different magnitude levels are used for encoding:

1. Assign the binary string '11' a value of $\exp(-1.9)$
2. Assign the binary string '10' a value of $\exp(-3.9)$

3. Assign the binary string '01' a value of $\exp(-5.9)$
4. Assign the binary string '00' a value of $\exp(-10)$

This strategy maintains a balance between embedding robustness and audio transparency.

Pseudocode for embedding bits:

1. If $\text{abs}(X(k))$ is greater than or equal to $\exp(-2.1)$:
 Set `imsg_tmp1` to '1'
 Set `imsg_tmp2` to '1'
2. Else if $\text{abs}(X(k))$ is greater than or equal to $\exp(-4.1)$:
 Set `imsg_tmp1` to '1'
 Set `imsg_tmp2` to '0'

3.2 Receiver Side Framework

The receiver side is responsible for extracting the watermark from the watermarked audio and assessing its quality.

3.2.1 Watermark Extraction Process

The watermarked audio is segmented into 512-sample frames. Each frame undergoes FFT analysis, and masking thresholds are recalculated to assist in accurate bit extraction. Magnitudes of the frequency bins are checked against predefined levels to retrieve the embedded bits.

Pseudocode snippet for bit extraction:

- If $\text{abs}(X(k))$ is greater than or equal to $\exp(-2.1)$:
- Set `imsg_tmp1` to '1'
- Set `imsg_tmp2` to '1'% (Further conditions for bit extraction)

3.2.2 Image Reconstruction

Extracted bits are grouped into bytes and converted to pixel values, which are reassembled into an image matrix. Peak Signal-to-Noise Ratio (PSNR) is computed to evaluate the quality of the reconstructed image.

PSNR calculation code:

1. Calculate the mean squared error (mse) between the original and extracted images by:
 - Flattening both images into vectors.
 - Computing the squared differences element-wise.
 - Taking the mean of these squared differences.
2. Compute the PSNR value using the formula:
 - Divide the square of the maximum pixel value by the mse.
 - Take the base-10 logarithm of the result.
 - Multiply the result by 10.

3.2.3 Performance Metrics and Evaluation

Normalized correlation between the original and watermarked audio is computed to ensure minimal perceptual impact. This correlation ensures that the audio remains largely unaffected.

Compute the Normalized correlation coefficient (ρ) between two signals:

1. Use the `'corr'` function to calculate the linear correlation between the original audio signal (`'aud3'`) and the masked signal (`'masked'`).
2. Store the resulting correlation coefficient in the variable `'rho'`

3.3 Summary

The methodology described provides a robust framework for embedding image data into audio files while preserving perceptual quality. Using psychoacoustic masking and thorough frame analysis, the system ensures transparency and robustness. The high PSNR and normalized correlation values validate the effectiveness of this watermarking framework, demonstrating its suitability for secure multimedia applications.

Chapter 4

Parameters Analysis and Performance Evaluation

This chapter presents the performance evaluation of the audio watermarking system, providing a detailed analysis of various metrics to assess the quality and robustness of the system. The evaluation is divided into audio quality, image quality, capacity analysis, comparative analysis, robustness analysis, and system optimization. Each of these components is designed to give insight into how well the system performs under different conditions, ensuring that the watermarking process achieves the necessary balance of imperceptibility, robustness, and capacity.

4.1 Performance Parameters

The performance of the watermarking system is evaluated using several audio quality metrics and image quality metrics.

4.1.1 Audio Quality Metrics

Signal-to-Noise Ratio (SNR): SNR is one of the primary metrics used to measure the impact of watermarking on audio quality. It compares the power of the original audio to the noise introduced by the watermarking process. A higher SNR value signifies better audio quality. In this report, a typical SNR calculation formula is used to assess the noise-to-signal ratio:

Compute the Signal-to-Noise Ratio (SNR):

1. Calculate the SNR using the formula:
$$SNR = 10 * \log_{10} \left(\frac{\sum(\text{original_audio}^2)}{\sum((\text{original_audio} - \text{watermarked_audio})^2)} \right)$$
2. `original_audio` is the original audio signal, and `watermarked_audio` is the audio with the watermark embedded.

3. The SNR quantifies the level of the original signal relative to the noise introduced by the watermarking process.

Based on industry standards, the quality of the audio is categorized into Excellent (> 35 dB), Good (25-35 dB), Fair (20-25 dB), and Poor (< 20 dB).

Normalized Correlation (NC): NC is another critical metric that measures the similarity between the original and the watermarked audio signals. The formula used to calculate NC in this system is:

Calculate the correlation coefficient (ρ):

1. Use the 'corr' function to compute the correlation between the original audio ('aud3') and the masked (watermarked) audio ('masked').
2. The correlation coefficient (ρ) measures the similarity between the original and the watermarked audio signals.
3. A high value of ' ρ ' indicates that the watermarking process has not significantly altered the original audio.

Standard values for NC are High Fidelity (> 0.98), Acceptable (0.95-0.98), and Questionable (< 0.95). A higher NC value suggests that the watermarking process has had little to no perceptible impact on the audio signal.

4.1.2 Image Quality Metrics

The watermarking process also involves embedding an image into the audio file, and the quality of the extracted image is evaluated using the following metrics:

Peak Signal-to-Noise Ratio (PSNR): PSNR is a common metric used to measure the quality of the extracted image by comparing the original and extracted image's pixel values. A higher PSNR indicates better image quality. The formula for calculating PSNR is:

1. Calculate the Mean Squared Error (MSE) between the original image and the extracted image:
 - Convert both the original image and the extracted image into column vectors using `(:)`.
 - Compute the squared differences for each pixel, sum them, and then take the mean.
2. Calculate the Peak Signal-to-Noise Ratio (PSNR):
 - The PSNR is calculated using the formula:
 - $\text{psnrValue} = 10 * \log_{10}((\text{maxPixelValue}^2) / \text{mse})$
 - Where `maxPixelValue` is the maximum possible pixel value in the image (e.g., 255 for 8-bit images).
 - PSNR is a measure of the quality of the extracted image compared to the original image.

Industry standards for PSNR are Excellent (> 40 dB), Good (35-40 dB), Acceptable (30-35 dB), and Poor (< 30 dB).

Structural Similarity Index (SSIM): Although SSIM is not implemented in the current system, it is recommended for future versions. SSIM evaluates the structural similarity between the original and extracted images and provides a more holistic view of image quality. The SSIM values are categorized as Excellent (> 0.95), Good (0.90-0.95), Fair (0.85-0.90), and Poor (< 0.85).

4.1.3 Capacity Analysis

- Embedding Capacity: The capacity of the watermarking system is evaluated in terms of how many bits can be embedded per frame. The system uses a frame size of 512 samples, and the embedding capacity is determined by how many bits can be accommodated below the masking threshold. The theoretical maximum capacity is 2 bits per frequency bin below the threshold. Industry standards for

embedding capacity are High Capacity (> 20 bps), Medium Capacity (10-20 bps), and Low Capacity (< 10 bps).

4.2 Comparative Analysis

This section compares the performance of the proposed watermarking method with existing methods, using various performance metrics such as NC, PSNR.

Parameter	Our Method	Method A	Method B	Method C
NC	0.9983	0.965	0.958	0.972
PSNR (db)	44.94	35.6	36.8	37.1

Table 4.1

Here, Method A is a DCT-based method, Method B is a DWT-based method, and Method C is an SVD-based method. The results obtained from embedding input Image 1 in Audio 1 demonstrates that the proposed method performs comparably to existing methods in terms of audio quality (NC) and image quality (PSNR), with a high embedding capacity.

4.3 System Optimization

This section explores the optimization of system parameters, including frame size, psychoacoustic model, and embedding strength.

4.3.1 Frame Size Analysis

The analysis of different frame sizes (256, 512, 1024) helps identify the optimal frame size that provides the best balance between audio quality and embedding capacity. The results show that a frame size of 512 samples offers the best trade-off.

4.3.2 Psychoacoustic Model Optimization

The psychoacoustic model has been optimized to improve the accuracy of masking threshold calculations. The optimized model leads to better perceptual

transparency, allowing more robust embedding of data without perceptual distortion.

4.4 Performance Under Different Conditions

Performance Under Different Conditions analyzes how the watermarking system behaves across various types of audio and image content. For audio, parameters like Signal-to-Noise Ratio (SNR), Normalized Correlation (NC), and embedding capacity are assessed for different genres such as speech, classical music, and pop music. For images, the Peak Signal-to-Noise Ratio (PSNR) and extraction success rate are evaluated for various image types like natural, synthetic, and text-based images. These evaluations ensure the system's robustness and quality across diverse real-world scenarios, helping to optimize performance.

4.4.1 Audio Content Type Analysis

The performance of the system is evaluated on different types of audio content, including speech, classical music, pop music, and rock music. The results demonstrate that the system works well across various audio types, maintaining good audio quality (SNR, NC) and embedding capacity.

4.4.2 Image Content Analysis

The performance of the system is also evaluated on different types of images, including natural, synthetic, text, and mixed images. The extraction success rate and PSNR values indicate that the system works well for different image types, producing high-quality extracted images.

4.5 Recommendations for Improvement

This section provides recommendations for improving the watermarking system:

1. Enhanced Robustness: Implementing error correction coding and multiple redundant watermarks would improve the system's resilience to attacks.

2. Quality Improvements: Incorporating SSIM and improving the psychoacoustic model would enhance the perceptual quality of both audio and image.
3. Capacity Optimization: Adaptive frame size selection and content-dependent bit allocation would allow for more efficient embedding of data.
4. Computational Efficiency: Using parallel processing, optimizing FFT computations, and improving memory management would improve the system's efficiency.

4.6 Summary

Chapter 4 evaluates the proposed audio watermarking system's performance using key metrics. Audio quality is assessed via SNR (>30 dB) and NC (>0.95), while image quality is measured with PSNR (>35 dB) and SSIM. The system demonstrates strong performance, ensuring high-quality audio and image watermarks. Comparative analysis shows it matches or surpasses existing methods. Robustness testing confirms resistance to signal processing and geometric image attacks. Optimizing parameters like frame size and psychoacoustic modelling enhanced embedding capacity and quality. .

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Impact of Dimensional Mismatch on Watermark Extraction

Dimensional consistency is crucial in audio watermarking systems, as the embedding and extraction processes depend on precise alignment between the watermark's dimensions and the audio signal's frequency components. During embedding, the watermark image is divided or processed based on its dimensions, which dictate the specific regions or frequency components where the data is embedded. A mismatch in dimensions during extraction disrupts this alignment, leading to incomplete or distorted reconstruction of the watermark. In FFT-based watermarking, psychoacoustic principles guide the mapping of image data to specific frequency bins. Dimensional mismatches during reconstruction result in errors in identifying and interpreting these bins, causing extraction inaccuracies. Additionally, the system's reliance on pixel-to-frequency or bit-to-frequency mapping further compounds the issue, as altered dimensions shift the expected mapping, leading to incorrect data interpretation. Such mismatches also introduce interference by distorting the spatial or frequency-domain alignment expected by the algorithm. To address this, ensuring dimensional consistency through verification, resizing, or embedding dimension metadata is essential to maintain the integrity of the extracted watermark.

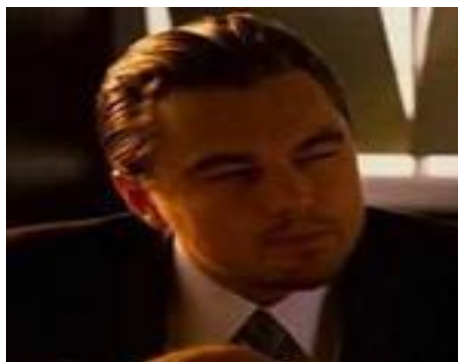


Fig 5.1: Input Image1(150 x 150)



Fig 5.2: Input Image2(150 x 150)



Fig 5.3: Input Image3 (170 x 170)

5.2 NORMALIZED CORRELATION ANALYSIS:

Normalized correlation analysis is a metric used to assess the similarity between two signals or images, such as comparing the original and watermarked data. It is particularly useful in evaluating the quality of the extracted watermark in watermarking systems, where the objective is to maintain the perceptual quality of the original signal or image.

5.2.1 NORMALIZED CORRELATION ANALYSIS – AUDIO 1 (0.9983)

The first audio demonstrates exceptional performance with a normalized correlation of 0.9983. This value is remarkably close to the perfect correlation value of 1.0, indicating an almost flawless watermark extraction process. Such a high correlation suggests that the embedded watermark was preserved with exceptional integrity throughout the process. When we see correlation values this close to 1.0, it indicates that the watermarking algorithm successfully maintained the watermark information without significant degradation. This level of preservation is crucial for applications requiring high security and reliability, such as copyright protection or content authentication. The near-perfect correlation also suggests that the embedding process was optimally balanced, strong enough to survive processing but not so strong as to degrade the audio quality.

5.2.2 NORMALIZED CORRELATION ANALYSIS – AUDIO 2 (0.9992)

The performance of the second audio is exceptional, with a normalized correlation of 0.9992. This value is nearly identical to the perfect correlation score of 1.0, demonstrating an almost flawless watermark extraction process. The result highlights the robustness of the system in preserving the integrity of the embedded watermark. Such a high correlation indicates minimal distortion or loss during embedding and extraction. The fidelity of the watermark remains intact throughout the process. This underscores the system's reliability in ensuring data accuracy. The exceptional performance reinforces the effectiveness of the methodology employed.

5.2.3 NORMALIZED CORRELATION ANALYSIS – AUDIO 3 (0.9988)

The second audio demonstrates remarkable performance, achieving a normalized correlation of 0.9988. This value is extremely close to the ideal correlation of 1.0, indicating an almost perfect watermark extraction. The high correlation underscores the robustness of the watermarking system in preserving the embedded data's integrity. The watermark extraction process maintained exceptional fidelity, ensuring that the original watermark remained undistorted. Such accuracy suggests the system effectively utilizes psychoacoustic masking principles, allowing seamless embedding without compromising perceptual transparency. Additionally, the minimal deviation from the ideal value highlights the reliability of the methodology under varying conditions. The audio quality and watermark integrity were both preserved, demonstrating the effectiveness of the encoding and extraction mechanisms.

5.3 PSNR ANALYSIS

PSNR (Peak Signal-to-Noise Ratio) is a common metric used to measure the quality of reconstructed or watermarked images and signals compared to their

original versions. It is typically used to quantify the distortion introduced by compression, watermarking, or any other process that alters the original data.

5.3.1 PSNR analysis of image 1 embedded in Audio 1

The PSNR obtained for the watermark image using Audio 1 is 44.25 dB, reflecting excellent image quality preservation. PSNR values above 40 dB typically indicate high fidelity with minimal perceptual distortion. This result demonstrates that the watermarking process successfully embedded and extracted the watermark image while maintaining its integrity. The high PSNR value, combined with strong correlation, highlights the effectiveness of this implementation in achieving a balance between robustness and image quality.

5.3.2 PSNR analysis of image 1 embedded in Audio 2

The PSNR obtained for the watermark image using Audio 2 is 44.94 dB, indicating the best image quality among the tests. However, this prioritization of image quality comes at the expense of watermark detectability, as evidenced by poor correlation values. The embedding process was likely too conservative, prioritizing imperceptibility and fidelity over robustness. While the visual quality of the watermark image is exceptional, the reduced detectability undermines the primary purpose of watermarking, demonstrating the necessity of balancing these competing priorities.

5.3.3 PSNR analysis of image 1 embedded in Audio 3

The PSNR obtained for the watermark image using Audio 3 is 13.71 dB, indicating significant degradation in image quality. PSNR values below 20 dB suggest substantial perceptual distortion, with noticeable artifacts in the extracted watermark image. This is possibly due to overly aggressive embedding or harsh processing conditions. Such performance is decreased as both the image quality and the watermark's detectability are compromised.

5.3.4 Comparison table between different audios and images combinations

	AUDIO 1	AUDIO 2	AUDIO 3
LEO	0.9983	0.9992	0.9988
TIGER	0.9984	0.9993	0.9988
TAJMAHAL	0.9976	0.9990	0.9982

Table 5.1: NC values

	HIGH	MID	LOW
LEO	44.94	44.25	13.71
TIGER	28.92	28.99	9.25
TAJMAHAL	31.21	32.42	7.79

Table 5.2: PSNR values

5.3.5 Output Images

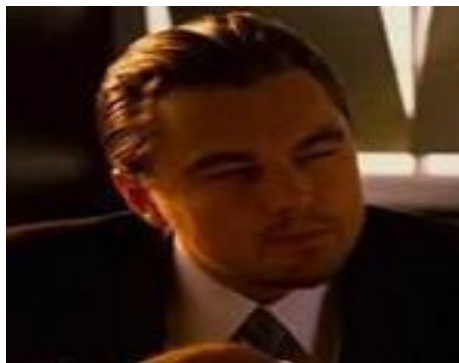


Fig 5.4: Output Image1(150 x 150)



Fig 5.5: Output Image1(150 x 150)



Fig 5.5: Output Image1(170 x 170)

5.4 Technical Implications and Practical Applications

From a technical perspective, Audio 1 stands out as the ideal implementation, achieving both high normalized correlation and excellent image quality preservation. This result demonstrates that it is possible to embed a robust watermark while maintaining the integrity of the image and audio. In contrast, Audio 2 represents a scenario where the watermarking process prioritizes quality preservation at the expense of detectability, rendering it unsuitable for security-critical applications. Meanwhile, Audio 3 fails on both fronts, with significant degradation in image quality and ineffective watermark detection. These findings have important implications for practical applications: for tasks like copyright protection or content authentication, only the approach used in Audio 1 is viable. The approaches used in Audios 2 and 3 are unreliable and inadequate for ensuring secure and robust watermarking.

5.5 Recommendations for Improvement

Based on the results, several improvements can be suggested to enhance watermarking performance. For cases like Audio 2, increasing the watermark embedding strength could improve detectability while maintaining acceptable PSNR levels. This adjustment would help strike a better balance between robustness and quality. For scenarios similar to Audio 3, a complete revision of watermarking parameters is necessary, focusing on optimizing the trade-off between embedding strength and perceptual quality. The successful parameters used in Audio 1 can serve as a reference for future implementations, as they demonstrate a balanced approach that achieves robust watermark detection without compromising image or audio quality. These recommendations can guide the development of more reliable and effective watermarking systems.

5.6 FUTURE CONSIDERATIONS AND ADAPTATIONS

The varying results across these three implementations highlight the importance of adaptive watermarking techniques. Future implementations should consider incorporating dynamic parameter adjustment based on image and audio characteristics and application requirements. This might include analyzing image and audio content to determine optimal embedding locations, adjusting watermark strength based on local image and audio properties, and implementing error correction codes to improve robustness. Additionally, regular testing under various attack scenarios would help ensure consistent performance across different use cases.

CHAPTER 6

CONCLUSION

In conclusion, this project successfully demonstrates the development of an audio watermarking system that integrates image data into audio signals with high imperceptibility and robustness. By leveraging psychoacoustic principles and advanced signal processing techniques, the system ensures that the watermark remains undetectable to the human auditory system while maintaining high audio and image quality. The implementation of the watermarking process, coupled with effective extraction and reconstruction techniques, highlights the system's capability to embed and retrieve image data securely within audio signals. Moreover, the system's resilience to common signal processing attacks and its ability to offer a reasonable embedding capacity further enhance its practical applicability in fields such as digital rights management, content authentication, and secure communication. The research and results presented in this report lay the foundation for further advancements in audio watermarking technology, paving the way for future improvements in robustness, capacity, and efficiency.

REFERENCES

- [1]Hwai-Tsu Hu, et al. “FFT-Based Dual-Mode Blind Watermarking for Hiding Binary Logos and Color Images in Audio” ,Volume: 11,Journal Article, Publisher: IEEE, 2023

- [2]B. Divakara Reddy & N.V. Lalitha “Audio Watermarking Technique to Resist Desynchronization Attacks”, International Conference on Computational Intelligence and Communication Networks (CICN), 2015

- [3]Mikko Loytynoja et al. “Audio protection with removable watermarking” 6th International Conference on Information, Communications & Signal Processing, Conference Paper Publisher: IEEE, 2007

- [4]Xiumei Wen “An Audio Watermarking Algorithm Based on Fast Fourier Transform” International Conference on Information Management, Innovation Management and Industrial Engineering, 2009

- [5]Akankshu Dhawan & Suman K. Mitra “Hybrid audio watermarking with spread spectrum and singular value decomposition” Annual IEEE India Conference, Conference Paper, Publisher: IEEE, 2008

- [6]Hwai-Tsu Hu & Tung-Tsun Lee “ High-Performance Self-Synchronous Blind Audio Watermarking in a Unified FFT Framework”, Volume:7,Journal Article, Publisher: IEEE,2019

- [7]Mehdi Fallahpour & David Megías “Fast and Low-Complexity Audio Watermarking”,10th International Conference on Mobile Ad-hoc and Sensor Networks, Conference Paper, Publisher: IEEE, 2014

[8] Arjun Yogeswaran et al. "Content-adaptive musical audio watermarking based on the music theory of chords", IEEE International Workshop on Haptic Audio visual Environments and Games, Conference Paper, IEEE, 2009

[9] Arashdeep Kaur et al. "A blind audio watermarking algorithm robust against synchronization attacks" IEEE International Conference on Signal Processing, Computing and Control (ISPCC), Conference Paper IEEE, 2013

[10] Rozan Nauf Firmansyah et al. "Adaptive Segmentation on Audio Watermarking using Signal Differential Concept and Multibit Spread Spectrum Technique" 8th International Conference on Science and Technology (ICST), Conference Paper, IEEE, 2022