Audio Steganography: A Comprehensive Survey of Techniques and Emerging Challenges

Suhaib Ali*, Talaat Wahbi†, Nabeil Eltayieb *, Rashad Elhabob *
*Faculty of Computer Science and Information Technology, Karary University, Khartoum 12345, Sudan
†School of Computer and Technology, City University, Ulster Lusail 200197, Qatar
suhaipalig@gmail.com

Abstract

Recently, information-hiding techniques have become important due to the widespread use of multimedia and the Internet. Actually, information hiding falls into two large categories: steganography, which is the art and science of secure communication, and watermarking, which is used to protect copyright and data ownership. Audio has some unique features that have motivated many researchers and developers to use it as an information-hiding medium. The audio, referred to as cover audio, is used to conceal a secret message, making it unnoticeable to the observer due to steganography techniques. This paper reviews audio steganography techniques and classifies them scientifically. The strengths and limitations of each steganography technique are presented in this work, moreover, it shows how the researchers treat the weaknesses of each method. The paper highlights an audio steganography emerging challenges such as advanced hiding methods and real-time steganography systems which have special constraints discussed in this paper. Audio steganography can utilize cover audio in various domains, which we have compared in our work, along with a comparison of steganography techniques. Subjective and objective quality measurements of audio steganography are depicted in this paper

Key words-*Audio steganography; Real-Time steganography; Spacial domain; Frequency domain; PSNR*

I. Introduction

Data protection guarantees the confidentiality, integrity, and accessibility of information while concurrently shielding it from unauthorized access, corruption, theft, or loss. Information hiding represents a data security approach that extends beyond cryptographic methods. This technique can be categorized into two primary types: watermarking and steganography [1]. Watermarking serves to safeguard the copyright and ownership rights of the data [2]. In contrast, steganography is employed to maintain the confidentiality and integrity of the information [3]. Today, it plays a crucial role in data protection in modern society, where multimedia and the Internet are widely used. Even though the cryptographic system includes a variety of invulnerable algorithms to keep the data confidential, it faces suspicion. The suspicion of the intruders might again lead them to steal the encryption key through different ways, which might involve a social engineering [4], or threatened blackmail. The primary objective of steganography is to eliminate any suspicion among observers regarding the presence of a concealed message [5]. Many systems combine steganography with cryptography to enhance security and performance by first encrypting the secret data and then concealing the encrypted text using steganography [6]. A steganography system can use different mediums such as text, images, audio, video, and other formats for hiding the secret message. Audio is considered to be a potent medium to hide secret messages due to its high redundancy and high data transmission rates [7]. Audio faces an added challenge due to the extraordinary features of the human auditory system. HAS system has a wide dynamic frequency range and it is ultra sensitive [7], [8]. Figure (1) illustrates the block diagram of the audio steganography system. The sender and receiver have already agreed on key information such as embedding and extraction techniques, stego-key. The secret message may be text, image, video, or audio. The secret key works as tool used with the algorithm to conceal the data [9]. The embedding method is employed to embed the secret message in the cover audio using the stego-key. The generated stego-audio contains both the cover audio and the hidden secret message, which is transmitted over an unsecured channel to the receiver. The receiver extracts the hidden secret message from the generated stego-audio using the extraction technique based on the stego-key.

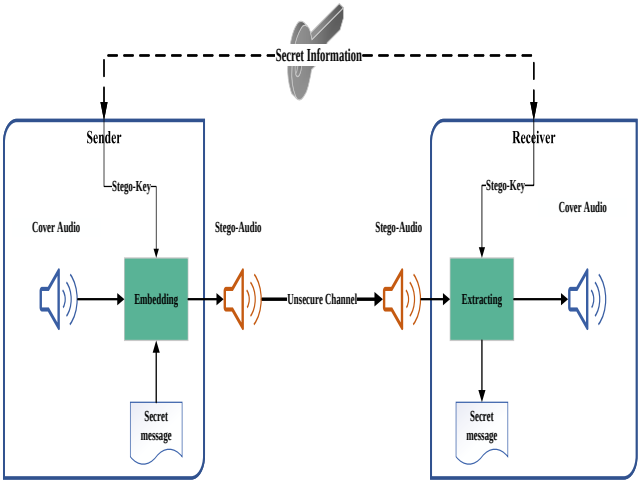


Fig. 1: Audio Steganography Block Digram.

The performance of audio steganography is evaluated based on three parameters: capacity, transparency, and robustness [10] as illustrated in Figure(2), the Steganography triangle portrays the trade-offs between these parameters, an efficient steganographic system should aim to minimize these trade-

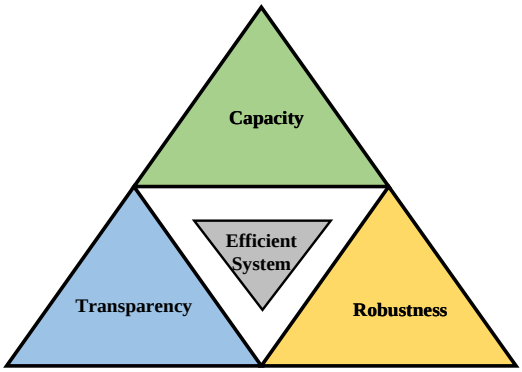


Fig. 2: Steganography Triangle.

offs. Neglecting any feature entirely could result in an unbalanced and impractical audio steganography method [11]. Capacity is the size ratio of the secret file to the cover file and is calculated as a percentage. The quantity of bits conveyed per time unit can signify capacity and is quantified in bits per second (bps) [1]. Transparency refers to the extent of resemblance and discrepancies between the original audio and the stego-audio [12], which is evaluated through audio quality metrics as discussed in section 6. Conversely, robustness denotes the resilience of stego-audio in the face of both intentional and credential attacks, as well as its ability to recover secret messages with minimal error. The ethical test attacks on stego-audio and stego-analysis act as pointers to robustness. In watermarking applications, the parameter of importance is robustness whereas for steganography applications, transparency and hiding capacity are more important [13], [14].

Although numerous review papers on audio steganography have been published, many suffer from significant limitations. Most notably, they often overlook the specific weaknesses of individual techniques and fail to discuss how researchers have addressed these challenges. This lack of critical analysis hinders a deeper understanding of the practical limitations of each method and the effectiveness of the proposed solutions. * These gaps underscore the need for a more detailed, deep-looking analytical review—one that not only categorizes existing approaches but also rigorously examines their limitations and the strategies devised to overcome them.

A. Our contributions

Most review papers on audio steganography primarily focus on techniques for hiding secret data. In contrast, our work aims to provide a comprehensive overview of all aspects of audio steganography, The main contributions are summarized as follows.

1) This paper presents a comprehensive review of audio steganography techniques, highlighting techniques and their strengths and weaknesses.
2) Unlike prior studies, our work systematically identifies the shortcomings of each technique and critically analyzes the strategies researchers have employed to overcome them.
3) Emerging challenges in steganography, such as advanced hiding techniques and real-time implementation, are presented in this paper.
4) We conduct a comparison of different audio steganography domains and techniques, emphasizing the strengths and weaknesses of each method.
5) Additionally, this paper evaluates both subjective and objective audio quality metrics.

B. Organization

We have structured the paper as follows: Section 2 presents an overview of related surveys and identifies existing research gaps. Section 3 classifies various audio steganography techniques. Sections 4 and 5 explore emerging challenges in audio steganography and provide a comparative analysis of the techniques, respectively. Section 6 discusses both objective and subjective audio quality metrics, and the paper concludes with Section 7.

II. Related Surveys and Research Gaps

The foundational work by Bender *et al.* [7] introduced key concepts in data hiding, including spatial and transform-domain techniques, but lacked quantitative metrics and security analysis. Later surveys sought to fill this gap but still fall short in examining vulnerabilities and mitigation strategies.

Dutta *et al.* [6] offer a broad classification of classical audio steganography methods (LSB, echo, phase, spread spectrum), comparing them using SNR and payload metrics. However, they do not link specific design choices to security weaknesses or discuss countermeasures. Alyousuf *et al.* [9] compare spatial and transform-domain methods qualitatively, but their review lacks metric standardization and robustness evaluation.

Rakhra *et al.* [5] examine hybrid steganography-cryptography techniques across multiple media. Their broad scope results in limited audio-focused insights and no unified evaluation. Duarte *et al.* [4] focus on social engineering in steganography contexts, contributing a sociotechnical perspective, yet offering no technical analysis of audio-based vulnerabilities.

Driss *et al.* [16] present a comprehensive survey of steganography in the context of IoT systems. While it covers both classical and modern approaches across multiple media, its treatment of audio techniques is limited. It also lacks detailed vulnerability analysis and discussion of mitigation trends.

These studies collectively reveal persistent gaps: **(G1)** lack of weakness-centered evaluation, **(G2)** absence of metric standardization across classical and hybrid methods, and **(G3)** limited tracking of how recent works address prior limitations. Our survey closes these gaps by (i) mapping vulnerabilities to design elements, (ii) unifying evaluation criteria, and (iii) tracing mitigation strategies chronologically.

Taken together, these surveys expose three recurring gaps: **(G1)** absence of a weakness-centric evaluation framework, **(G2)** lack of metric uniformity spanning classical and hybrid (crypto-assisted) methods, and **(G3)** limited tracking of how newer studies remediate earlier weaknesses. The present

TABLE I: Comparison of representative audio-steganography surveys

Survey	Year	Techniques covered	Metrics analysed	Stated limitations	Gaps our work addresses
Bender *et al.* [7]	1995	Spatial & transform fundamentals	None (conceptual)	No quantitative eval.; no security view	G1, G2
Djebbar *et al.* [15]	2012	LSB, phase, echo, transform	Capacity, robustness (qual.)	Outdated; no DL methods; limited metrics	G1, G2
Dutta *et al.* [6]	2020	LSB, phase, echo, spread-spectrum	Capacity, SNR	No attack analysis; no mitigation tracking	G1, G3
Alyousuf *et al.* [9]	2020	Spatial vs. transform methods	Qualitative only	Lacks unified metrics; limited robustness study	G2, G3
Rakhra *et al.* [5]	2021	Steganography + cryptography (multi-media)	None	Audio treatment shallow; no cross-metric study	G1, G2
Duarte *et al.* [4]	2021	Social-engineering attack vectors	None	Non-technical; no method-level discussion	G1
Reyers [10]	2023	Tool-focused comparison (LSB, transform, DL)	Capacity, SNR, BER	Small public datasets; tool-centric; no attack study	G1, G2
Driss *et al.* [16]	2025	IoT stego (multi-media incl. audio)	Security, scalability (qual.)	Audio coverage brief; no weakness mapping	G1, G3

survey closes these gaps by (i) mapping design decisions to known attack surfaces, (ii) standardising evaluation metrics and datasets, and (iii) chronologically tracing mitigation strategies to reveal research trajectories.

In summary, while previous surveys have offered foundational classifications and broad overviews of audio steganography, they often lack a critical, vulnerability-focused perspective. Table(I) presents a comparative analysis of related works, highlighting the limitations of each method. These recurring shortcomings reveal three key research gaps—(G1), (G2), and (G3)—which this survey addresses by mapping vulnerabilities to design decisions, standardizing evaluation metrics, and tracing mitigation efforts over time to clarify the evolution of the field.

III. Classification

Audio steganography techniques can be categorized into four primary domains: spatial, transform, frequency, and codec. Each domain encompasses unique methods for embedding hidden information, offering specific advantages and limitations. This classification serves as a guide to select the most suitable technique, based on the requirements of the particular information-hiding application, such as robustness, capacity, or imperceptibility.

A. Spatial domain

1) Least Significant Bit (LSB): The LSB process involves replacing the least significant bit in each byte of the digitized cover audio with the corresponding bit from the secret message as shown in Figure 3. Classic LSB technique is inaudible, simple, fast, and easy to implement [17], but easy to detect by statistical analysis [15] making it less Robust against attacks. To improve confidentiality, LSB can be enhanced by avoiding sequential modification of each byte in the digitized cover audio. Instead, a random hopping key can be used to determine where the secret message bits will be embedded [18][19]. Multi-level LSBs and error-correcting codes increase cover audio payload and robustness respectively [20][21] [22].The Hybrid LSB technique, combined with encryption or algorithms such as chaotic maps, offers an efficient and secure

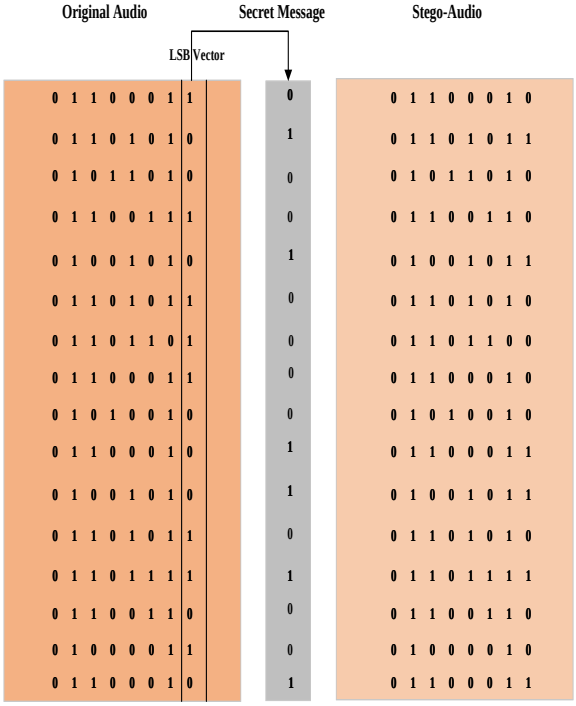


Fig. 3: Classic LSB.

approach to steganography.[23][24]. Dynamic and adaptive LSB methods are used to balance the performance of the audio steganography parameters [25][22][26].

2) Parity Coding: Parity coding involves modifying the parity of specific sets of bits within the audio signal to embed secret information. Parity coding is an improvement of the LSB technique, which operates on a group of samples instead of individual samples. The cover audio is divided into segments, each containing a predefined number of samples as shown in Figure 22. Calculate the parity for each segment and each bit of the secret message is embedded by adjusting the parity of the segment (evenness or oddness). If the current parity matches the bit to be embedded, no change is necessary.

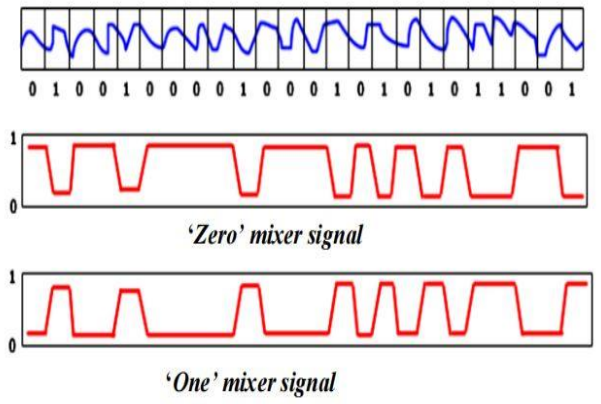


Fig. 4: Echo Hiding.

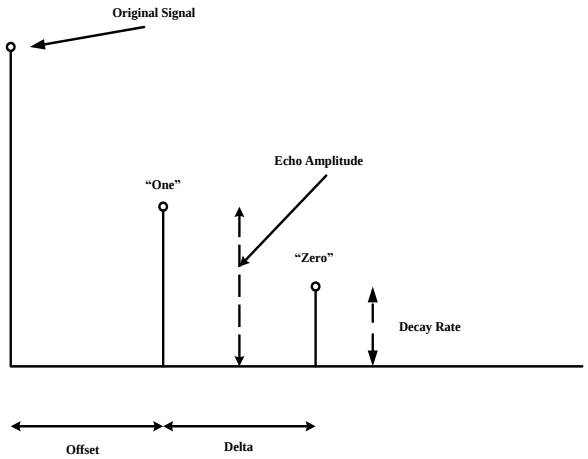


Fig. 5: Adjustable parameters.

Otherwise, the least significant bit of single sample in the segment is flipped to change the parity. Parity coding enhances robustness and transparency using random position hiding as shown in [27] , but its data hiding capacity is lower than that of classic LSB. As a result, it is commonly used in watermarking [22] and data integrity applications, but is rarely applied in audio steganography [28].

3) Echo hiding: Echo hiding leverages the relatively small differential range limitation of the Human Auditory System (HAS) [8], which means HAS cannot perceive absolute phases affected by common environmental distortions. It embeds digital identification tags into an audio signal with minimal noticeable degradation to the original signal [29]. Echo hiding is an improvement over adding random noise to audio, which can destroy hidden data during lossy compression. It is inaudible and robust against compression and filtering, making it well-suited for watermarking and steganography. Figure(4) illustrate echo hiding and Figure(5) illustrates its adjustable parameters [29]. There are two key thresholds: the audible threshold and the HAS's discriminatory ability. The echo amplitude and decay rate must remain below the audible threshold, while the offset and delta must meet the second threshold. A main drawback of echo hiding is low payload capacity, which can be addressed using quantum reversible circuits [30], due to their capability to handle more complex transformations while maintaining reversibility.

B. Transform domain

1) Discrete Cosine Transform (DCT): (DCT) is a spectral transformation based on the cosine function and shares several properties with the Discrete Fourier Transform (DFT), including de-correlation, energy compaction, separability, symmetry, and orthogonality [31]. It is widely used in various multimedia applications, particularly in compression and watermarking, due to its ability to concentrate most of the information into a small number of coefficients. The audio signal $x[n]$ is transformed using DCT transformation as shown in Equation(1). The inverse of DCT (IDCT) generate stego-audio $\tilde{X}[k]$ wich is calculate by add hidden secret data $\Delta[k]$ to the DCT transform

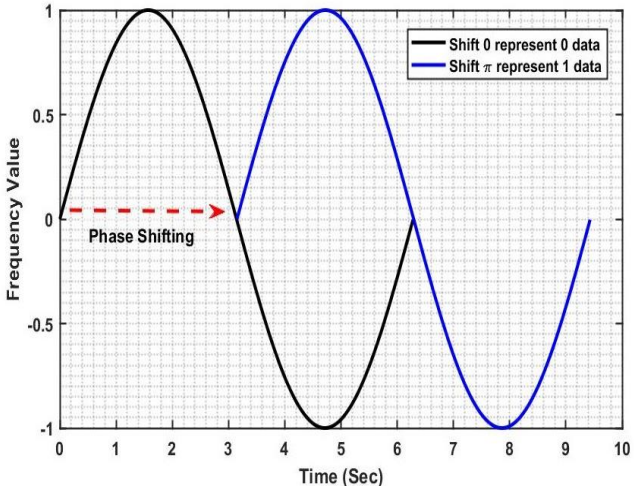


Fig. 6: Phase Shifting.

as shown in Equation(2). DCT-based methods is often embedded by altering the least significant bits (LSB) of the DCT coefficients [32], [33] or through coefficient comparison within the DCT domain [34]. DCT is applied to the voiced parts of the speech signal, making it highly imperceptible and robust by modifying the DCT magnitude to hide secret data [35], [36]. DCT can introduce artifact problems [37], which can be mitigated through adaptive embedding by selecting frequency components that are less sensitive to human perception.

$$X[k] = \alpha_k \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] \tag{1}$$

where

$$\alpha_k = \begin{cases} \sqrt{\frac{1}{N}} & \text{if } k = 0, \\ \sqrt{\frac{2}{N}} & \text{if } k \neq 0. \end{cases}$$

$$\tilde{X}[k] = X[k] + \Delta[k] \tag{2}$$

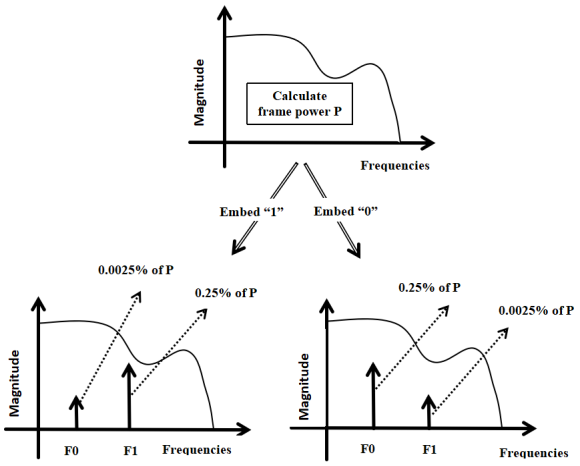


Fig. 7: Tone Insertion.

2) Discrete Wavelet Transform (DWT): (DWT) is one of the most powerful transform techniques provides a time-frequency representation of the signal simultaneously, making it particularly effective for steganography. Equation(3) illustrate the DWT coefficient which is modifying to hide secret data and then re-scaled to produce stego-audio. The receiver applies the inverse DWT (IDWT) to extract the hidden data.[38] presents an audio steganography method based on sample comparison in the DWT domain. Selected coefficients of a segment are compared with predetermined threshold values to achieve a high payload capacity. Further advancements in stego-audio capacity are explored in [39], which combines LSB techniques with direct message replacement. Their approach aims to hide secret data while achieving an average hiding capacity across various music and speech signals, all while maintaining suitable quality. The DWT technique addresses the trade-off between imperceptibility and robustness, as discussed in [40], using a quantization-based optimization model. The study aims to maximize the signal-to-noise ratio (SNR) and reduce the bit error rate (BER) by optimizing DWT coefficients and scaling factors.

$$x(t) = \sum_j \sum_k \mathrm{W}_{j,k}\phi_{j,k}(t) + \sum_j \sum_k \mathrm{D}_{j,k}\psi_{j,k}(t) \quad (3)$$

where $\mathrm{W}_{j,k}$ are the approximation coefficients, $\mathrm{D}_{j,k}$ are the detail coefficients, and $\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ are the scaling and wavelet functions.

C. Frequency domain

1) Phase Coding: Phase coding is utilized to embed secret data in the phase spectrum of the cover audio. Phase coding exploits HAS insensitivity to the relative phase of different spectral components [15]. Short-Time Fourier Transform (STFT) can be used to modify certain components of cover audio. Equation (4,5) describe the signal representation of audio using STFT and the embedding process respectively. Adjusting the phase of cover audio segment Φ to 0 or π implicitly hides 0 or 1 as shown in Figure(5). Phase coding is robust against noise, interference, and vulnerable to compression but has a low payload. Multi-phase angles (0,π/2,π,3π/2) can increase capacity using four different phase angles leading to hiding two bits in the audio segment known as Quadrature Phase Shift Keying (QPSK). Also, the Adaptive Multi-Level Phase Coding (AMPC) method with balanced performance can maximize capacity while maintaining imperceptibility[41], [42].

$$X(t,f) = |X(t,f)|e^{j\phi(t,f)} \quad (4)$$

Where $X(t,f)$ is the STFT of the audio signal at time t and frequency f, $X(t,f)$ is the magnitude of the signal in the frequency domain, and $\phi(t,f)$ is the phase of the signal. After embedding the message:

$$\tilde{X}(t,f) = |X(t,f)|e^{j(\phi(t,f)+\Delta\phi(t,f))} \quad (5)$$

where $\tilde{X}(t,f)$is the phase-modified STFT after embedding the secret message, and $\Delta\phi(t,f)$ is the phase shift used to encode the secret message.

2) Tone Insertion: The tone insertion technique injects tone with weak power and masks it with a louder sound, either simultaneously or in a temporal manner. The injected frequency must reside within the critical band of the louder Sounds to become inaudible. The tone insertion is powered to hide data imperceptibly with high robustness based on HAS limitation of masking weaker sound [8]. The tone insertion technique can use two frequencies, F0 and F1, with alternating tone intensity levels (0.025 of P) and (0.0025 of P) where P is the cover segment power as shown in Figure(6). This technique is inaudible to HAS and robust against noisy and cropped conditions [43], [44] but has low capacity. Studies treated the low capacity limitation of the tone insertion by increasing either inserted tone or power levels[39] [45]. Inserting eight tones into the audio segment to hide four bits may increase distortion and be audible. carefully choose the frequencies of the inserted tone to avoid collision with existing frequencies which may increase the error bit.

3) Spread Spectrum (SS): Spread Spectrum (SS) audio steganography is a sophisticated technique that embeds secret information by dispersing it across a wide frequency range, making it less detectable and resistant to interference. It takes advantage of HAS limitations by masking a weaker sound that contains secret data. A masking threshold is calculated using a psycho-acoustic model to determine how much secret information can be embedded in an audio signal without being perceptible to the human ear [46]. Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS) are used in the Spread Spectrum technique in analog and digital signals respectively. FHSS uses hopping to spread secure data across different frequencies while DSSS combines data signals directly with higher bit rate sequences. Spread Spectrum can use deep learning to get optimal spectrum for hiding secret data [47] and can integrated with other techniques and theories [48], [49].

D. Codec domain

Coder-decoder (codec) domain or compression domain converts voice signals into digital data packets for transmission

and then decoding it at the destination to reduce bandwidth and latency. The Codec domain ensures the robustness of hidden data and leverages the HAS perceptual characteristics. It's compatible with Streaming and Telecommunication Systems such as Skype, WhatsApp, and Google Talk [50]. Codec domain steganography utilized before audio codec (Pre-Codec) [51] or after the audio has been coded (Post-Codec) [52]. Pre-codec has a higher capacity compared with post-codec which is more suitable for real-time streaming systems. Lossy speech codec methods such as G.729 and G.723.1 are widely adopted in steganography [53], [54], [51]. LPC (Linear Predictive Coding) [55] and CELP (Code-Excited Linear Prediction) [56] are speech Coders standards, which are official NATO standards that are widely used in voice and speech processing [57]. Deep learning expands the codec domain with Speech Recognition to generate a steganography System based on an audio adversarial model [58].

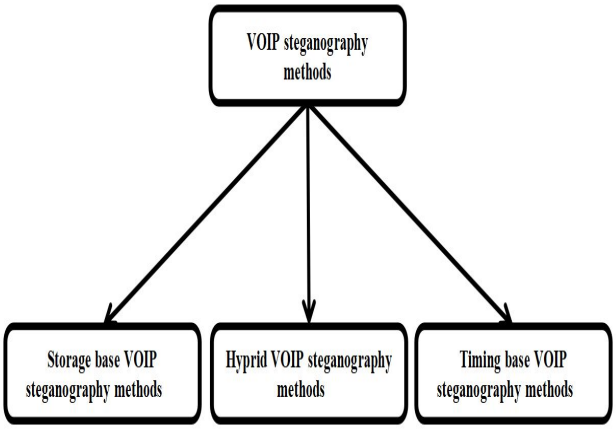


Fig. 8: Classification of VIOP Steganography.

IV. Steganography Emerging Challenges

There are specific areas that focus on the critical obstacles and challenges currently facing the field of steganography, which significantly impact the design and performance of steganographic systems. This section addresses the challenges of audio steganography, including advanced hiding methods, real-time implementation, differences between music and speech in steganography, stego-analysis techniques, and the application requirements of steganographic systems.

A. Advanced Methods

Traditional steganography techniques focus on resolving isolated issues within steganography and need expert knowledge of the embedding process [59], while advanced methods offer a comprehensive approach that enhances multiple elements simultaneously without adverse effects. Advanced methods such as **fractal coding, Genetic algorithm, and machine learning** are used to improve all key factors of the steganography system. **fractal coding** technique was introduced by Benoit Mandelbrot in 1977 [60] as a lossy compression and started to gain attention in the field of audio steganography with efficient applications [1], [61]. **Genetic Algorithm (GA)** optimize audio steganography system through either determining the optimum position to hide data [62], [63], or determining the optimum cover audio for steganography [64]. **Deep learning** provides tools that enhance the primary parameters of steganographic systems besides adaptability. Generative Adversarial Networks (GANs) [65], [66], Conventional Neural Networks (CNNs) [67], [68], and Recurrent Neural Networks (RNNs)[69], [70]

B. Real-Time Steganography

Real-time increases security in steganographic systems due to the processing time constraint by eavesdroppers and the hidden secret data may be invalid. A combination of Audio steganography with network steganography achieves a higher capacity and good robustness [71], which can be named Audio Stream or real-time Steganography. Real-time steganography systems have additional perimeters such as latency, data loss, and unordered receiving data. Therefore, these systems have limited time for embedding data, requiring faster, simpler, and synchronized methods [72]. Real-time steganography systems can utilize Radio Frequency (RF) communication systems or Voice Over Internet Protocol (VOIP) which can be classified based on storage, time, or hybrid [71] as shown in Figure (8). Real-time steganography systems based on modifying the VOIP packages are adopted [73], [74], while other steganography systems Intentionally delay the packets which are hiding secret data based on LACK (Lost Audio Packets Steganographic Method) [75], [76]

C. Speech and music

Speech and Music possess unique strengths and characteristics that enhance the data-hiding process shown in TABLE(II). Consequently, a steganography algorithm designed for speech may not be suitable for music, and vice versa. Music range of 20 Hz to 20,000 Hz (20 kHz) which is the HAS frequency range. Speech typically falls within the human vocal frequency range, spanning from 300 to 3400 Hz used for standard telephone systems according to ITU-T-Rec p.861 [77]. The fundamental speech frequencies typically range from 120 Hz for men to 210 Hz for women [78] which is conducted based on standard deviation analysis of speech across various sex and languages. Music often contains redundant information due to structured temporal patterns and harmonies containing repetitive themes or motifs. Many intervals and unvoiced frames in speech inspire researchers to increase the capacity of hidden data [35], [79], but using intervals may lead to incorrect data extraction. The capacity and security of speech steganography can be treated using two security levels based on Counterlet Transform (CT) and dufling oscillator [80].

D. Stego-analysis

Steganography seeks to hide information within seemingly harmless media, whereas steganalysis is dedicated to detecting and revealing these concealed messages [81]. Steganalysis is essential across multiple fields, serving to extract stego-messages, detect hidden suspicious content, and assess the

TABLE II: Speech and Music Steganography Comparison

Aspect	Speech	Music
Amplitude	20 dB- 110 dB	30 dB- 120 dB
Frequency Range	human vocal range 300 Hz-3400 Hz	HAS range 200 Hz-2000 Hz
Redundancy	High redundancy	Lower redundancy
Content Pattern	Unexpected notion	Repetitive music notes
Data Capacity	Higher	Lower
Transparency	more sensitive to distortion	less sensitive to distortion
Detection and Robustness	More robust against detection	More susceptible to detection

robustness of current steganography techniques.Stego-analysis is categorized into two types [82]: targeted stego-analysis, which applies to known steganography techniques, and universal stego-analysis, which addresses unknown techniques. Universal stego-analysis can further be divided into blind and un-blind methods. Stego-analysis can be classified based on the tools and methods used into two categories: signal processing and deep learning. Audio transformation coefficients, including DCT, DWT, and the Goertzel algorithm, can estimate the presence of hidden information[83], while deep learning tools, such as Support Vector Machines (SVMs), are used to detect secret messages [84].

E. Multi-Level Audio Steganography

Al-Najjar was the first to introduce the concept of multi-level steganography (MLS), utilizing images as cover media [85]. MLS leverages an upper-level steganographic technique as a carrier to develop and integrate subordinate, lower-level methods, creating a hierarchical framework that combines multiple techniques. MLS is employed to enhance key steganographic parameters, including capacity, transparency, and robustness, while making the embedding process more adaptive. MLS can be dynamically implemented through hybrid multi-steganography techniques, such as multi-level Least Significant Bit (LSB) methods, to maximize data capacity [86]. Alternatively, it can combine various steganography types, offering users the flexibility to select their preferred multimedia cover for concealing the real content [87]. The challenges of MLS include balancing trade-offs between steganographic features and managing the increased complexity compared to single-layer techniques. The method proposed by [88] addresses the balance between capacity and security by integrating Least Significant Bit (LSB) and Phase Coding techniques within a multi-layered framework.

F. Application Requirements

Implementing steganography requires several critical elements for security and efficiency. Audio steganography depends on the cover audio's characteristics, such as file size, sample rate, and sampling frequency. The cover size must be large enough to conceal secret data without causing distortion. Audio is digitized through samples (e.g., 8-bit, 16-bit, 24-bit), where bit depth impacts the quality of the digital waveform [89]. The sample rate (FS), measured in Hertz (Hz), indicates how many samples are captured per second and must be at least twice the highest frequency in the audio signal to prevent aliasing, as per the Shannon sampling theorem [90]. Both sampling and sample rate directly proportional to payload capacity and imperceptibility. Sophisticated systems select cover audio based on characteristics suited to the hidden data. Requirements also vary by embedding techniques, which may have specific conditions regarding the cover medium characteristics. To protect sensitive data, encryption is applied before steganography for added security. It's also essential that the chosen methods maintain a balance between security and imperceptibility to effectively conceal the embedded information without compromising the quality of the carrier medium.

V. Audio Steganography Comparison

Different representations or analyses of an audio signal (audio domains) significantly influence the choice of steganography methods. Each domain captures distinct characteristics of the audio signal and presents specific strengths and weaknesses for steganography, as shown in TABLE (III). The spatial domain is the simplest and has the lowest computational cost compared to other domains, but it offers low confidentiality for hidden data due to its higher detectability and vulnerability to destruction. TABLE (IV) compares the different audio steganography techniques mentioned in the paper and highlights their capabilities and limitations.

VI. Audio quality metrics

The performance of audio steganographic systems primarily focuses on the quality of stego-audio, payload capacity, and robustness. Among these, the quality of stego-audio is the most critical factor, as poor quality could defeat the purpose of steganography by arousing the observer's suspicion. Stego-audio quality can be assessed using both subjective and objective methods. Although the International Telecommunication Union's ITU-R BS.1387 standard outlines the principles and characteristics of objective perceptual audio quality measurement methods in general[59], section 6.2 focuses on commonly used objective audio quality metrics in steganographic systems. On the other hand, subjective measurements involve evaluating the audio based on individual human perception, which can be more accurate in detecting subtle artifacts or distortions that algorithms might miss. However, subjective assessments are time-consuming and require the participation of multiple individuals.

A. Subjective Listening Testing

Subjective audio quality evaluation measures have different types of listening testing Mean Opinion Score (MOS), Double-Blind Tests, and ABX Testing. Using the MOS listener scale the quality of stego-audio is in the range(of 1(Bad)-5(Exeleant)), meanwhile, in double-blind tests listeners give feedback on the audio's quality after listening to the original and stego-audio and they do not know both. On the other hand, ABX Testing Listeners compare two audio samples (A and B) with a reference (X) to determine which one matches the reference in quality.

TABLE III: Audio Steganography Domains Comparison

Audio Domain	Description	Strengths	Weaknesses	Techniques
Spatial domain	Embedding secret data into the audio by modifying the audio's sample values	Simple High capacity Low cost	Low security Low robustness	LSB Parity coding Echo hiding
Transform domain	Modifies the audio coefficients signal through a transform formula	High security Robust against compression and filtering	Low capacity High cost	DCT DWT
Frequency domain	Hide the secret data within the frequency components of the audio file	High transparency High security Robust against attacks	Low capacity Complex	Phase coding Tone insertion Spread spectrum
Codec domain	Embeding secret data into the compression bit stream of audio file	Compatible with standard decoders Resilient to compression attacks	Limited capacity Quality degradation Complex	Pre-codec Post-codec

B. Objective Measurements

1) MSE: Mean Squared Error (MSE) is used to quantify the mathematical difference between the values of the original audio and the stego-audio, thereby determining the level of distortion introduced Equation(6). MSE ranges from 0 to 1, where 0 indicates no distortion, and 1 signifies significant distortion in the stego-audio.The root-mean-squared error (RMSE) and mean absolute error (MAE) are evaluations of MSE, with RMSE being more suitable for normally distributed errors [91].

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{6}$$

where x_i represents the original audio signal values, $\hat{x}_i$ represents the stego-audio values, and N is the total number of samples in the audio signal.

2) SNR and PSNR: Signal-to-Noise Ratio (SNR) and Peak Signal-to-Noise Ratio (PSNR) are metrics used to assess the degradation of audio quality after modifying the original audio. Both SNR and PSNR are calculated as the ratio between the power of the original audio and the power of the noise introduced by the modifications. These values are expressed as percentages, where higher values indicate lower distortion and better audio quality, while lower values indicate higher distortion. PSNR is an extension of SNR that specifically focuses on the peak error between the original and modified audio. The ratio of the size between the secret message and the host audio significantly affects both SNR and PSNR, as some parts of the host audio remain unchanged compared to the original[60]. Equations(7,8) are illustrate that the SNR and PSNR depends on the signal and noise power. Methods for pre-estimating PSNR in stego-audio are employed to facilitate early evaluation of imperceptibility and reduce the trade-off between imperceptibility and robustness [92], [93].

$$\text{SNR} = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{7}$$

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right) \tag{8}$$

where
P_{signal} is the power of the signal
P_{noise} is the power of the noise.
MAX is the maximum possible value of the signal
MSE is the mean squared error between the original and the reconstructed signal.

$$r = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2 \sum_{i=1}^{N}(y_i - \mu_y)^2}} \tag{9}$$

where x_i and y_i are the samples of the original and stego-audio, μ_x and μ_y are their mean values, and N is the total number of samples.

3) SSIM: The Structural Similarity Index (SSIM) was originally developed to evaluate image quality [94], measuring similarity based on luminance, contrast, and structure. It was later extended to Multi-Scale SSIM for greater flexibility [95]. Classic SSIM has proven to be effective for audio quality prediction [96], [97], and has been adapted for evaluating audio quality as Audio-Visual SSIM (AVSSIM) and Audio-Visual MS-SSIM (AVMSSSIM) [98]. Equation(10) illustrates the mathematical comparison between an audio signal and a reference signal using SSIM, evaluating aspects such as loudness, frequency structure, and amplitude variations for each audio frame. SSIM values range from -1 to 1, where 1 indicates perfect similarity, 0 denotes no similarity, and negative values signify significant differences or phase inversion.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{10}$$

where x and y are original and stego-audio segment respectively, $\mu_x\mu_y$ are the mean values of the segments, σ_x^2 and σ_y^2 are the variances of the segments, σ_{xy} is the covariance between them, and C_1 and C_2 are small constants.

4) PEAQ: The Perceptual Evaluation of Audio Quality (PEAQ) was employed to assess the quality of received audio in [99] and has been standardized by the International Telecommunication Union (ITU) [100]. PEAQ is an advanced objective method that simulates the human auditory system to predict perceived audio quality. It uses a complex psychoacoustic model to compare original and cover audio, producing a score known as the Objective Difference Grade (ODG, Equation 10), which ranges from -4 (very annoying) to 0 (imperceptible difference).

$$\text{ODG} = f\left(\text{Features}(x_{\text{orig}}) - \text{Features}(x_{\text{proc}})\right) \tag{11}$$

TABLE IV: Comparison of Audio Steganography Techniques.

Ref	Year	Technique	Methodology	Capabilities	Limitations
[25]	2021	LSB	Reducing the hamming distance between the original and modified signals of the Quantization Index Modulation (QIM)	Introduces fewer modifications to the original audio signal Balance between security and robustness	Limited resistance to all Attacks Limited scalability for real-time applications due to complexity.
[26]	2020	LSB	The K-means clustering algorithm is used to group the audio signal data into clusters, select the location of hiding secret data	Enhancing robustness and imperceptibility Adaptive method High embedding capacity	Limited testing with varied speech types Vulnerability to advanced attacks
[27]	2015	Parity coding	Combines spatial and frequency techniques to improve the performance of data hiding	Better balance between robustness and transparency High robustness against noise addition, compression, and filtering High capacity	High computational complexity Audio quality degradation
[29]	1996	Echo hiding	Hides information in an audio signal by adding echoes to the sound	High imperceptibility Satisfied degree of robustness	Sensitivity to echo parameters Limited Capacity
[32]	2024	DCT	Used LSB to embed secret data into the DCT coefficient of audio signal	Enhancing security High imperceptibility	Complex implementation Limited capacity
[40]	2021	DWT	Manipulate Wavelet transform to hide secret data	High imperceptibilities Flexible in hiding capacity Comprehensive evaluation	Complex implementation Vulnerability to advanced detection methods
[42]	2004	Phase coding	Manipulate phase of the audio signal with the secret data after implement Fourier transformation to the audio	High imperceptibility Robustness against compression and filtering	Complex implementation Limited Data Capacity
[44]	2004	Tone insertion	generated tones belong to the secret message (tone for every bit) and insert it in the cover audio with scaled power	High Imperceptibility High robustness Flexible application	Limited capacity Complexity of implementation
[45]	2017	Tone insertion	generate tone for every 2 bits of the secret message and insert it in the cover audio with scaled power	High imperceptibility Increase capacity	Implementation complexity
[49]	2019	Spread spectrum	The authors introduce optimization strategies that balance capacity, robustness, and security using spread spectrum	Enhanced security by spreading hidden data over a wide frequency range Chaos theory complicates detection for attackers	Complex implementation Potential for reduced capacity
[51]	2022	Pre-codec	Utilizes LSP interpolation to achieve high embedding capacity in speech signals specifically for the G.723.1 codec based on (CNN)	High capacity Improves the robustness Balance between security and capacity	Codec specificity (G.723.1) not commonly used Implementation complexity
[52]	2022	Post-codec	replacing the least significant bits of audio samples with hidden data	Simple and effective cause of LSB High imperceptibility Ease of implementation	Limited capacity Susceptible to detection
[1]	2018	Advance method	Utilizes fractal coding in combination with chaotic maps to encode data in the temporal domain	High capacity cause of fractal coding Enhancing security cause of randomness of chaotic maps	Complex implementation Sensitivity to audio quality
[64]	2023	Advance method	Leverages a multi-objective evolutionary algorithm that selects an optimal cover audio based on specific fitness functions	Optimized cover selection through multi-objective evolutionary algorithm (MOEA) Improved performance metrics Robustness against attacks	Complexity implementation Dependence on input data quality Potential computational overhead
[65]	2019	Advance method	Implements GANs to generate realistic steganographic audio by learning the distribution of natural audio signals	High-quality imperceptible cause of the capabilities of GANs Adaptability to various types of audio signals	Complex implementation Extensive training data for effective GAN performance is rarely available

where x_{orig} is the original audio signal, x_{proc} is stego-audio signal, and the function f represents the psychoacoustic model used to compute the difference between the two signals' perceptual features.

5) Correlation Coefficient: The Correlation Coefficient evaluates audio quality by comparing the test audio with a reference audio. It relies on statistical relationships between the two audio signals and provides a measure of similarity. The result is scaled in the range of -1 to 1: 1 indicates perfect positive correlation in terms of shape and timing, 0 denotes no correlation, and -1 means the audio signal is a mirror image of the reference. This method also assesses temporal and cross-correlation to cover specific aspects of the audio signals. Equation(9) describes the calculation of the Correlation Coefficient.

VII. Conclusion

This paper has offered a comprehensive review of audio steganography techniques, analyzing their strengths and limitations, and exploring how researchers have tackled these

challenges. It also highlights emerging issues, such as the development of more advanced hiding methods and the need for real-time steganography. Furthermore, the comparison of various audio steganography domains and techniques revealed important distinctions in their performance. Both subjective and objective audio quality metrics are crucial in evaluating the effectiveness of steganographic systems, ensuring that hidden data remains undetectable while maintaining high audio quality.

As the field evolves, addressing these challenges with innovative solutions is essential. Future research should focus on improving real-time steganography capabilities, bolstering resistance to steganalysis, and refining quality metrics to achieve an optimal balance between security and audio imperceptibility.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the staff of Karary University for their continuous support and encouragement throughout the development of this research work.

REFERENCES

[1] A. H. Ali, L. E. George, A. Zaidan, and M. R. Mokhtar, "High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain," *Multimedia Tools and Applications*, vol. 77, no. 23, pp. 31487–31516, 2018.

[2] H. Baisuni, D. Djulaeka, and M. A. Sajjad, "Legal protection for unauthorized copying of songs on digital platforms through audio watermarking method," *JUSTISI*, vol. 10, no. 3, pp. 547–564, 2024.

[3] F. Hemeida, W. Alexan, and S. Mamdouh, "A comparative study of audio steganography schemes," *International Journal of Computing and Digital Systems*, vol. 10, pp. 555–562, 2021.

[4] N. Duarte, N. Coelho, and T. Guarda, "Social engineering: The art of attacks," in *International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability*, pp. 474–483, Springer, 2021.

[5] M. Rakhra, R. Kumar, H. Walia, *et al.*, "A review on data hiding using steganography and cryptography," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 1–4, IEEE, 2021.

[6] H. Dutta, R. K. Das, S. Nandi, and S. M. Prasanna, "An overview of digital audio steganography," *IETE Technical Review*, vol. 37, no. 6, pp. 632–650, 2020.

[7] W. R. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," in *Storage and Retrieval for Image and Video Databases III*, vol. 2420, pp. 164–173, SPIE, 1995.

[8] M. N. Kunchur, "The human auditory system and audio," *Applied Acoustics*, vol. 211, p. 109507, 2023.

[9] F. Q. A. Alyousuf, R. Din, and A. J. Qasim, "Analysis review on spatial and transform domain technique in digital steganography," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 573–581, 2020.

[10] P. M. Reyers, "A comparative analysis of audio steganography methods and tools," B.S. thesis, University of Twente, 2023.

[11] K. Gopalan, "Audio steganography for information hiding and covert communication–a tutorial," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0242–0243, IEEE, 2018.

[12] A. A. Alsabhany, F. Ridzuan, and A. Azni, "The adaptive multi-level phase coding method in audio steganography," *IEEE Access*, vol. 7, pp. 129291–129306, 2019.

[13] D. M. Ballesteros L and J. M. Moreno A, "Highly transparent steganography model of speech signals using efficient wavelet masking," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9141–9149, 2012.

[14] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding-a survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, 2002.

[15] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative study of digital audio steganography techniques," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, no. 1, p. 25, 2012.

[16] M. Driss, L. Berriche, S. B. Atitallah, and S. Rekik, "Steganography in iot: A comprehensive survey on approaches, challenges, and future directions," *IEEE Access*, 2025.

[17] A. Singh and H. Singh, "An improved lsb based image steganography technique for rgb images," in *2015 IEEE International Conference on electrical, computer and communication technologies (ICECCT)*, pp. 1–4, IEEE, 2015.

[18] K. Bhowal, A. J. Pal, G. S. Tomar, and P. Sarkar, "Audio steganography using ga," in *2010 International Conference on Computational Intelligence and Communication Networks*, pp. 449–453, IEEE, 2010.

[19] U. Ehsan Ali, E. Ali, M. Sohrawordi, and M. N. Sultan, "A lsb based image steganography using random pixel and bit selection for high payload," 2021.

[20] N. Cvejic and T. Seppanen, "Increasing the capacity of lsb-based audio steganography," in *2002 IEEE Workshop on Multimedia Signal Processing.*, pp. 336–338, IEEE, 2002.

[21] C. Parthasarathy and S. Srivatsa, "Increased robustness of lsb audio steganography by reduced distortion lsb coding," *Journal of Theoretical and Applied Information Technology*, vol. 7, no. 1, pp. 080–086, 2005.

[22] R. Martinez-Noriega, H. Kang, B. Kurkoski, K. Yamaguchi, K. Kobayashi, and M. Nakano, "Increasing robustness of audio watermarking dm using athc codes," in *Proceedings of the Mexican conference on informatics security*, 2006.

[23] B. Harjito, B. Sulistyarso, and E. Suryani, "Audio steganography using two lsb modification and rsa for security data transmission," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 10, no. 2-4, pp. 107–111, 2018.

[24] I. Q. Abduljaleel and A. H. Khaleel, "Hiding text in speech signal using k-means, lsb techniques and chaotic maps.," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 10, no. 6, 2020.

[25] X. Sun, K. Wang, and S. Li, "Audio steganography with less modification to the optimal matching cnv-qim path with the minimal hamming distance expected value to a secret," *Multimedia Systems*, vol. 27, no. 3, pp. 341–352, 2021.

[26] Z. Wu, R. Li, and C. Li, "Adaptive speech information hiding method based on k-means," *IEEE Access*, vol. 8, pp. 23308–23316, 2020.

[27] G. P. TVS and S. Varadarajan, "A novel hybrid audio steganography for imperceptible data hiding," in *2015 International Conference on Communications and Signal Processing (ICCSP)*, pp. 0634–0638, IEEE, 2015.

[28] M. Tayel, A. Gamal, and H. Shawky, "A proposed implementation method of an audio steganography technique," in *2016 18th international conference on advanced communication technology (ICACT)*, pp. 180–184, IEEE, 2016.

[29] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *International Workshop on Information Hiding*, pp. 295–315, Springer, 1996.

[30] M. Velayatipour, M. Mosleh, M. Y. Nejad, and M. Kheyrandish, "Quantum reversible circuits for audio watermarking based on echo hiding technique," *Quantum Information Processing*, vol. 21, no. 9, p. 316, 2022.

[31] S. A. Khayam, "The discrete cosine transform (dct): theory and application," *Michigan State University*, vol. 114, no. 1, p. 31, 2003.

[32] P. Venugopala, S. Raghavendra, B. Ashwini, *et al.*, "Crypticcare: A strategic approach to telemedicine security using lsb and dct steganography for enhancing the patient data protection," *IEEE Access*, 2024.

[33] B. A. Dharani, B. Yashaswini, G. R. S. Reddy, and S. M. Rajagopal, "Multimodal steganography: a comparative analysis of lsb and dct methods for image and audio data concealment," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pp. 1–5, IEEE, 2024.

[34] M. P. Jain and P. Trivedi, "Effective audio steganography by using coefficient comparison in dct domain," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 8, 2013.

[35] A. Kanhe and G. Aghila, "Dct based audio steganography in voiced and un-voiced frames," in *Proceedings of the international conference on informatics and analytics*, pp. 1–4, 2016.

[36] A. Kanhe and G. Aghila, "A dct–svd-based speech steganography in voiced frames," *Circuits, systems, and signal processing*, vol. 37, no. 11, pp. 5049–5068, 2018.

[37] S. Hemalatha, U. D. Acharya, and A. Renuka, "Wavelet transform based steganography technique to hide audio signals in image," *Procedia Computer Science*, vol. 47, pp. 272–281, 2015.

[38] S. Gupta and N. Dhanda, "Audio steganography using discrete wavelet transformation (dwt) & discrete cosine transformation (dct)," *IOSR Journal of Computer Engineering*, vol. 17, no. 2, pp. 32–44, 2015.

[39] M. Sheikhan, K. Asadollahi, and R. Shahnazi, "Improvement of embedding capacity and quality of dwt-based audio steganography systems," *World Applied Sciences Journal*, vol. 13, no. 3, pp. 507–516, 2011.

[40] S.-T. Chen, T.-W. Huang, and C.-T. Yang, "High-snr steganography for digital audio signal in the wavelet domain," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 9597–9614, 2021.

[41] L. Gang, A. N. Akansu, and M. Ramkumar, "Mp3 resistant oblivious steganography," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, vol. 3, pp. 1365–1368, IEEE, 2001.

[42] X. Dong, M. F. Bocko, and Z. Ignjatovic, "Data hiding via phase manipulation of audio signals," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. V–377, IEEE, 2004.

[43] K. Gopalan, S. Wenndt, A. Noga, D. Haddad, and S. Adams, "Covert speech communication via cover speech by tone insertion," in *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*, vol. 4, pp. 4_1647–4_1653, IEEE, 2003.

[44] K. Gopalan and S. Wenndt, "Audio steganography for covert data transmission by imperceptible tone insertion," in *Proc. The IASTED International Conference on Communication Systems And Applications (CSA 2004), Banff, Canada*, 2004.

[45] M. H. Sayed, "Audio steganography using tone insertion technique,"

[46] B. A. Patil and V. A. Chakkarwar, "Review of an improved audio steganographic technique over lsb through random based approach," *IOSR J Comput Eng*, vol. 9, no. 1, pp. 30–34, 2013.

[47] Z. Kexin, "Audio steganalysis of spread spectrum hiding based on statistical moment," in *2010 2nd International Conference on Signal Processing Systems*, vol. 3, pp. V3–381, IEEE, 2010.

[48] R. Kaur, A. Thakur, H. S. Saini, and R. Kumar, "Enhanced steganographic method preserving base quality of information using lsb, parity and spread spectrum technique," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, pp. 148–152, IEEE, 2015.

[49] R. Tanwar, K. Singh, M. Zamani, A. Verma, and P. Kumar, "An optimized approach for secure data transmission using spread spectrum audio steganography, chaos theory, and social impact theory optimizer," *Journal of computer networks and communications*, vol. 2019, no. 1, p. 5124364, 2019.

[50] B. Q. A. Ali, H. I. Shahadi, M. S. Kod, and H. R. Farhan, "Covert voip communication based on audio steganography," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 821–830, 2022.

[51] H. Kheddar and D. Megias, "High capacity speech steganography for the g723. 1 coder based on quantised line spectral pairs interpolation and cnn auto-encoding," *Applied Intelligence*, vol. 52, no. 8, pp. 9441–9459, 2022.

[52] E. A. Alsolami, "Audio steganography method using least significant bit (lsb) encoding technique," *Journal Of Theoretical And Applied Information Technology*, vol. 100, p. 12, 2022.

[53] C. Zhang, S. Jiang, and Z. Chen, "Spm: estimating payload locations of qim-based steganography in low-bit-rate compressed speeches," *Multimedia Tools and Applications*, vol. 83, no. 37, pp. 85227–85252, 2024.

[54] F. Li, B. Li, L. Liu, Y. Feng, and L. Peng, "A steganography method for g. 729a speech coding," *International Journal of Performability Engineering*, vol. 16, no. 5, p. 784, 2020.

[55] J. N. Hasoon and S. N. Al-Saad, "Speech hiding based on quantization of lpc parameters,"

[56] N. Tahilramani and N. Bhatt, "Information hiding in proposed 10.6 kbps cs-acelp based speech codec using quantization index modulation," *International Journal of Speech Technology*, vol. 25, no. 1, pp. 219–230, 2022.

[57] S. Talati, P. Etezadifar, M. H. Ahangar, and M. Molazade, "Investigation of steganography methods in audio standard coders: Lpc, celp, melp," *Majlesi Journal of Telecommunication Devices*, vol. 12, no. 1, pp. 7–15, 2023.

[58] H. Tan, C. Liu, Y. Lyu, X. Zhang, D. Zhang, and Z. Gu, "Audio steganography with speech recognition system," in *2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC)*, pp. 256–263, IEEE, 2021.

[59] Z. Wang, O. Byrnes, H. Wang, R. Sun, C. Ma, H. Chen, Q. Wu, and M. Xue, "Data hiding with deep learning: A survey unifying digital watermarking and steganography," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 6, pp. 2985–2999, 2023.

[60] M. Nelson and J.-L. Gailly, "The data compression book 2nd edition," *M & T Books, New York, NY*, 1995.

[61] A. H. Ali, L. E. George, and M. R. Mokhtar, "An adaptive high capacity model for secure audio communication based on fractal coding and uniform coefficient modulation," *Circuits, Systems, and Signal Processing*, vol. 39, no. 10, pp. 5198–5225, 2020.

[62] B. A. Mitras and N. F. Hassan, "Using hybird genetic algorithm in audio steganography," *Iraqi Journal of Statistical Sciences*, vol. 13, no. 25, pp. 150–164, 2013.

[63] A. Q. Raheema, "Adopt an optimal location using a genetic algorithm for audio steganography," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 9, no. 4, pp. 1070–1082, 2021.

[64] M. H. N. Azam, F. H. M. Ridzuan, and M. N. S. M. Sayuti, "Optimized cover selection for audio steganography using multi-objective evolutionary algorithm," *Journal of Information and Communication Technology*, vol. 22, no. 2, pp. 255–282, 2023.

[65] D. Ye, S. Jiang, and J. Huang, "Heard more than heard: An audio steganography method based on gan," *arXiv preprint arXiv:1907.04986*, 2019.

[66] L. Chen, R. Wang, D. Yan, and J. Wang, "Learning to generate steganographic cover for audio steganography using gan," *IEEE Access*, vol. 9, pp. 88098–88107, 2021.

[67] R. Zhang, H. Dong, Z. Yang, W. Ying, and J. Liu, "A cnn based visual audio steganography model," in *International conference on adaptive and intelligent systems*, pp. 431–442, Springer, 2022.

[68] J. Wang, R. Wang, L. Dong, and D. Yan, "Robust, imperceptible and end-to-end audio steganography based on cnn," in *International conference on security and privacy in digital economy*, pp. 427–442, Springer, 2020.

[69] Z. Yang, X. Du, Y. Tan, Y. Huang, and Y.-J. Zhang, "Aag-stega: Automatic audio generation-based steganography," *arXiv preprint arXiv:1809.03463*, 2018.

[70] Z. Lin, Y. Huang, and J. Wang, "Rnn-sm: Fast steganalysis of voip streams using recurrent neural network," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1854–1868, 2018.

[71] N. MV and M. GR, "A survey on audio stream steganography techniques," in *Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021)*, 2021.

[72] R. Shrivastava, M. Singh, K. S. S. R. Teja, *et al.*, "A real-time implementation for the speech steganography using short-time fourier transformior secured mobile communication," in *Journal of Physics: Conference Series*, vol. 2089, p. 012066, IOP Publishing, 2021.

[73] S. Deepikaa and R. Saravanan, "Coverless voip steganography using hash and hash," *Cybern. Inf. Technol*, vol. 20, pp. 102–115, 2020.

[74] P. Bedi and A. Dua, "Network steganography using the overflow field of timestamp option in an ipv4 packet," *Procedia Computer Science*, vol. 171, pp. 1810–1818, 2020.

[75] W. Mazurczyk, "Lost audio packets steganography: the first practical evaluation," *Security and Communication Networks*, vol. 5, no. 12, pp. 1394–1403, 2012.

[76] H. Moodi and A. R. Naghsh-Nilchi, "A new hybrid method for voip stream steganography," *Journal of Computing and Security*, vol. 3, no. 3, pp. 175–182, 2016.

[77] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes, "Peaq-the itu standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, 2000.

[78] H. Traunmüller and A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults," *Unpublished manuscript*, vol. 11, 1995.

[79] S. Shirali-Shahreza and M. Shirali-Shahreza, "Steganography in silence intervals of speech," in *2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 605–607, IEEE, 2008.

[80] A. S. Hameed, "A high secure speech transmission using audio steganography and duffing oscillator," *Wireless Personal Communications*, vol. 120, no. 1, pp. 499–513, 2021.

[81] D. A. Shehab and M. J. Alhaddad, "Comprehensive survey of multimedia steganalysis: Techniques, evaluations, and trends in future research," *Symmetry*, vol. 14, no. 1, p. 117, 2022.

[82] N. R. K. Kesa, "Steganography a data hiding technique," 2018.

[83] B. E. Carvajal-Gamez, M. A. Castillo-Martinez, L. A. Castaneda-Briones, F. J. Gallegos-Funes, and M. A. Diaz-Casco, "Audio steganalysis estimation with the goertzel algorithm," *Applied Sciences*, vol. 14, no. 14, p. 6000, 2024.

[84] A. Singhal and P. Bedi, "Blind quantitative steganalysis using svd features," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 369–374, IEEE, 2018.

[85] A. J. Al-Najjar, "The decoy: multi-level digital multimedia steganography model," in *Proceedings of the 12th WSEAS international conference on Communications*, pp. 445–450, 2008.
[86] A. A. Alsabhany, F. Ridzuan, and A. Azni, "The progressive multilevel embedding method for audio steganography," in *Journal of physics: conference series*, vol. 1551, p. 012011, IOP Publishing, 2020.
[87] E. Satish, N. Sreenivasa, E. Naresh, P. R. Naidu, and A. Ramachandra, "Multimedia multilevel security by integrating steganography and cryptography techniques," in *ITM Web of Conferences*, vol. 57, p. 01012, EDP Sciences, 2023.
[88] M. H. Sayed and T. M. Wahbi, "Information security for audio steganography using a phase coding method," *European Journal of Theoretical and Applied Sciences*, vol. 2, no. 1, pp. 634–647, 2024.
[89] C. Sujatha, *Vibration, Acoustics and Strain Measurement*. Springer, 2023.
[90] M. Dodson, "Shannon's sampling theorem," *Current science*, vol. 63, no. 5, pp. 253–260, 1992.
[91] T. O. Hodson, "Root mean square error (rmse) or mean absolute error (mae): When to use them or not," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–10, 2022.
[92] Z. Mazdak, B. A. M. Azizah, M. A. Shahidan, and S. C. Saman, "Mazdak technique for psnr estimation in audio steganography," *Applied Mechanics and Materials*, vol. 229, pp. 2798–2803, 2012.
[93] M. H. N. Azam, F. Ridzuan, and M. Sayuti, "A new method to estimate peak signal to noise ratio for least significant bit modification audio steganography.," *Pertanika Journal of Science & Technology*, vol. 30, no. 1, 2022.
[94] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
[95] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The thrity-seventh asilomar conference on signals, systems & computers, 2003*, vol. 2, pp. 1398–1402, Ieee, 2003.
[96] S. Kandadai, J. Hardin, and C. D. Creusere, "Audio quality assessment using the mean structural similarity measure," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 221–224, IEEE, 2008.
[97] C. Sloan, N. Harte, D. Kelly, A. C. Kokaram, and A. Hines, "Objective assessment of perceptual audio quality using visqolaudio," *IEEE Transactions on Broadcasting*, vol. 63, no. 4, pp. 693–705, 2017.
[98] X. Min, G. Zhai, J. Zhou, M. C. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Transactions on Image Processing*, vol. 29, pp. 6054–6068, 2020.
[99] P. Kabal, "Perceptual evaluation of audio quality (peaq)," 2004.
[100] S. Kämpf, J. Liebetrau, S. Schneider, and T. Sporer, "Standardization of peaq-mc: Extension of itu-r bs. 1387-1 to multichannel audio," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, Audio Engineering Society, 2010.



Suhaib Ali received the B.S. degree in 2008 from the Faculty of Computer Science and Information Technology, University of Karary, Khartoum, Sudan. In 2014, he received the M.S. degree from the Faculty of Computer Science and Information Technology, Sudan University of Science and Technology (SUST), Khartoum, Sudan. He is currently pursuing the Ph.D. degree at the University of Karary. His research interests include information hiding, with a current focus on audio steganography.



Nabeil Eltayieb is an Assistant Professor in the Faculty of Computer Science and Information Technology, University of Karary, Khartoum, Sudan. He is currently pursuing a postdoctoral degree at the University of Electronic Science and Technology of China, Chengdu, China. His research interests include applied cryptography, cloud computing, and secure data sharing.



Rashad Elhabob is an Assistant Professor at the College of Computer Science and Information Technology, University of Karary, Khartoum, Sudan. He received his Ph.D. in Cryptography from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019. He also holds a Master's degree in Information Security from the University of Khartoum and a B.Sc. in Computer Science from the College of Engineering, Karary University. Recently, he completed a postdoctoral fellowship at UESTC. His research interests include applied cryptography, information security, and cloud computing. He has published more than 35 papers in reputable international journals and conferences.

Talaat M. Wahbi is an Associate Professor and former Dean of the Faculty of Computer Science and Information Technology at Sudan University of Science and Technology (SUST), Sudan. He is currently affiliated with the Faculty of Computer Science at City University, Ulster, Qatar. His research interests include distributed systems, information security, and advanced computing technologies.