Power System Protection and Control
ISSN:1674-3415

SECURE TEXT HIDING IN AUDIO STEGANOGRAPHY USING SHUFFLED ARNOLD CAT MAP BASED ENCRYPTION AND ADAPTIVE STEP SIZE QUANTIZATION INDEX MODULATION

Shyla Nagarajegowda[1], Kalimuthu Krishnan[2]

[1]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai.
shylaashok@gmail.com
[2]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai.
kalimutk@srmist.edu.in
Corresponding author: Shyla Nagarajegowda

Received: 26 September 2025 **Revised**:19 October 2025 **Accepted**: 22 November 2025

ABSTRACT:

Audio steganography offers promising solution by enabling embedding of secret messages in audio files. However, achieving high security, robustness against attacks, and minimal distortion of host audio signals remains significant challenge. This research addresses challenges of imperceptibility in hiding sensitive frequency regions of messages ensuring robustness against audio processing and encrypts message for added security. Shuffled Arnold Cat Map (SACM) and Adaptive Step Size Quantization Index Modulation (ASSQIM) are proposed to encrypt secret data and hide text in audio steganography. ASSQIM embeds secret messages by adjusting step size based on frequency analysis. Lower step size is used in sensitive frequency region to manage imperceptibility, whereas larger step size is applied in nonsensitive regions to enhance robustness against attacks. Hidden message is encrypted using SACM from cover audio, which enhances security under different attacks. Experimental analysis shows SACM-ASSQIM obtains high Normalized Correlation (NC) of 0.9999 compared to QIM with reduction in Log Spectral Distance (LSD) from 0.998 to 0.944 representing significance on enhanced robustness and security with audio-text. Additionally, SACM- ASSQIM obtains high Peak-to-Signal-Noise Ratio (PSNR) of 29.4726 dB in high-pass filter attack and 98.3547 dB on TIMIT and GTZAN datasets when determining with audio-image compared to Linear Predictive Analysis (LPA).

Keywords: Adaptive Quantization Index Modulation, Audio steganography, Embedding, Hiding text, Shuffled Arnold Cat Map.

INTRODUCTION

Digital steganography is the art and science of utilizing human perception redundancy to embed secret messages into a digital cover, such as audio, images, and video. As digital communication expands across platforms, securing confidential data without modifying perceptible content becomes a critical challenge. Among images and videos, audio steganography is preferred because audio signals have high redundancy and capacity to embed secret messages. The human auditory system exhibits lower sensitivity to minor signal variations, which enables effective embedding of hidden messages within audio data [2-5]. Because human hearing is more sensitive than vision, slight embedding introduces noise into audio, which limits the availability of embedding locations in digital media, making steganography more challenging [6-8]. As more than 90% of digital media shared online consists of music, it is crucial to explore information hiding in audio [9-11]. Imperceptibility is vital for securing embedded data against unauthorized access during transmissions. Steganography provides a strategic technique for addressing threats that enhance security protocols and protect sensitive information [12] [13]. In audio steganography, the main components are the original audio and secret messages, which together act as inputs for the embedding process. The resulting stego audio acts as both input and output for the overall steganographic process [14] [15]. Eventually, the extraction process retrieves the hidden secret message while restoring the original audio from stego audio [16-18].

DOI: 10.46121/pspc.53.4.18

Four primary facets are considered when evaluating data-hiding techniques: security, hiding capacity, robustness, and imperceptibility [19] [20]. Hiding capacity refers to the maximum size of a secret message embedded within an audio signal without significantly affecting the quality. Security not only ensures the imperceptibility of secret messages but also makes them undetectable. Imperceptibility indicates that the stego and cover audio files are indistinguishable [21]. Robustness represents the capacity to recover a secret message effectively with minimal errors as well as the capability of a stego file to withstand various attacks [22]. The effectiveness and hiding payload of the secret message are two key considerations for any strong steganography system [23-25]. Researchers have emphasized the importance of preserving the visual quality of stego images to closely match the cover image. Any noticeable distortion raises suspicion and increases the risk of attackers, a factor that depends on user requirements, and the type of steganography used [26]–[29]. This research goal is to design an efficient data-hiding mechanism utilizing the steganography method to ensure integrity and confidentiality. Additionally, it aims to enhance stego audio security by using a method that ensures no noticeable differences in the image when analyzed with a visual attack tool [30-32].

1.1 Problem Statement

Accomplishing high security, robustness against attacks, and minimal distortion remains a challenge in audio steganography because the human auditory system is highly sensitive, which makes even small modifications to audio signals noticeable. Existing methods, such as Least Significant Bit (LSB) and LSB-Binaries of Message Size Encoding (LSB-BMSE), are vulnerable to noise attacks owing to easily identified embedding patterns, whereas Linear Predictive Analysis (LPA) suffers from low robustness because speech parameters are distorted easily. Likewise, QIM suffers from a fixed quantization step size, where larger steps enhance robustness but cause audible distortion in sensitive frequency bands, and a smaller step size conserves imperceptibility but affects resistance to attacks which makes the design of a reliable and secure audio steganography challenging.

1.2 Objective

To address above stated issue, the Shuffled Arnold Cat Map (SACM) for encryption and Adaptive Quantization Index Modulation (ASSQIM) for effective data embedding is proposed with less distortion. SACM enhances security by introducing complex and nonlinear transformations with the integration of pixel shuffling, divide-and-rotate operations, and an integer-based diffusion mechanism. Hence, this transformation breaks spatial correlations, makes them resistant to statistical attacks, and increases security. ASSQIM is applied to embed the data into the audio signal by adjusting the step size depending on the frequency analysis across the sensitive and non-sensitive frequency regions. A larger step size is used in non-sensitive regions (below 2000 Hz and above 5000 Hz) to increase robustness against attacks, and a lower step size is employed in sensitive frequency bands (2000–5000 Hz), which reduces distortion and maintains imperceptibility. Therefore, this adaptability makes the model balance between embedding and imperceptibility, which makes it challenging to determine hidden information.

1.3 Contribution

The main contribution of this research is as follows.

- FFT is used to convert an audio signal into a frequency domain depending on the sampling rate, which provides an accurate representation of the signal's frequency components in the audio. FFT ensures compatibility with different sampling rates, which enhances adaptability and robustness in steganography.
- In traditional ACM, the Divide & divide-and-rRotate and shuffling sample processes are involved in establishing a chaotic encryption scheme during the confusion process that provides resistance against noise effects and data loss.
- The Discrete Cosine Transform (DCT) enhances robustness by operating in the frequency domain, where hidden data are affected less by noise. DCT separates low- and high-frequency components, effectively making it easier to extract embedded data with less distortion. Its compatibility with primary audio processing results in a reliable recovery even after transformations.
- In audio steganography, ASSQIM is applied to embed data by dynamically adjusting the quantization step size based on audio signal characteristics. This reduces distortion and ensures the imperceptibility of hidden data, which maintains audio quality.
- The primary aim of SACM-ASSQIM is to determine audio as cover data and text as secret data. Additionally, SACM-ASSQIM was analyzed using audio as cover data and images, as well as audio as secret data.

DOI: 10.46121/pspc.53.4.18

The rest of section is structured as follows: Section 2 presents a literature review of existing methods. Section 3 presents the research motivation and novelty, and Section 4 discusses the proposed methodology. Section 5 describes the metrics utilized for evaluating steganography and Section 6 analyzes the experimental results. Finally, the conclusion of research paper are presented in Section 7.

LITERATURE SURVEY

Related work on audio steganography has been analyzed using various methodologies, advantages, and limitations. A detailed analysis presents the trade-off between the existing methods and provides insights for establishing secure and efficient audio steganography systems. Lang Chen et al. [33] introduced an audio steganography model to generate superior steganographic cover audio to embed messages. The training model involved three components: the generator, discriminator, and training-based steganalysis. Subsequently, the Least Significant Bit Matching (LSBM) method was applied to embed the secret message into the cover audio, which helped obtain stego audio. Next, adversarial training was established between the three components to generate steganographic cover audio, which ensured effective message embedding. However, LSBM was vulnerable to statistical attacks because the embedding process changes the predictable noise patterns, making it easier for attackers to detect hidden data. Lang Chen et al. [34] suggested imperceptible audio steganography using a psychoacoustic model for audio steganography. Initially, perturbation was applied to the stego audio to construct a noisy stego, which was performed for misclassification. The perturbation was optimized in an adversarial process that provided optimal performance and guaranteed the undetectability and imperceptibility of the stego audio. A two-stage optimization method was used to reduce the loss function using gradient back propagation. Nevertheless, a small modification to the audio was detected by forensic tools in the proposed method, which threatened the confidentiality of hidden messages. Rajeev Kumar and Jainath Yadav [35] presented an LPA for speech watermarking between vowel and nonvowel frames. Numerous vowel segment frames were utilized in the LPA to estimate the predicted values. LPA helps embed the minimum value acquired from the grayscale watermark image. Hence, vowel and non-vowel approaches provide more security than state-of-the-art methods. However, LPA has lower robustness because it relies on predicting speech signal parameters, which were easily altered by noise, leading to reduced security. Kasetty Praveen Kumar and Aniruddha Kanhe [36] established a framelet transform and Singular Value Decomposition (SVD) for audio steganography. The scaling parameter was optimized to maximize robustness by considering the Perceptual Evaluation of Speech Quality (PESQ) for the speech signal, while the Objective Difference Grade (ODG) score was used for the music signal to ensure imperceptibility. The embedding was applied to low-pass framelet transform coefficients via SVD with optimized scaling parameters. The established method provides better robustness against primary signal processing attacks. Nevertheless, the established approach was vulnerable to noise and signal distortion because it is based on signal decomposition into frequency and amplitude components, which were easily interrupted by external interference. Mahmoud M. Huwaida et al.T. Elshoush [37] developed LSB-BMSE for audio steganography. Initially, the secret message was compressed using Huffman coding, and then the Advanced Encryption Standard-128 (AES-128) was applied to encrypt the data. The audio cover was split into blocks based on the secret message size. After hiding its random sample size, the developed approach employs a BMSE to embed a secret message that provides high security. However, LSB-BMSE was vulnerable to noise because modifying LSB represents the smallest portion of the audio signal sample's binary value, leading to noticeable distortions. Huwaida T. Elshoush and Mahmoud M. Mahmoud et al. [38] presented an LSB Piecewise Linear Chaotic Map (LSBPWLCM) for embedding secret data in random samples in audio steganography. Random samples were generated using PWLCM to provide security in a one-time pad as an input key. The presented approach provides dual protection by combining steganography with a highly secure one-time pad, which achieves better imperceptibility and higher capacity. Nevertheless, the presented approach was sensitive to key initialization, such as low entropy in the key selection and insufficient randomness in the key generation process, where improper key generation leads to predictable patterns and less security. Yu Tang et al. [39] suggested SilentTrig for speaker identification based on a backdoor attack. The suggested approach employs an optimized steganographic network for embedding triggers and establishes a two-stage adversarial optimization process. This ensures that poisoned samples remain acoustically indistinguishable, which enhances both attack effectiveness and imperceptibility. The suggested approach provides robust mapping among the imperceptible trigger and victim models. However, SilentTrig struggled with robustness against backdoor attacks because it was based on a trigger pattern exploited by attackers. Kaiyu Ying et al. [40] established a Generation-Optimal Allocation Strategy (GOAS) for generating a parity-check matrix based on different cover audio. The adaptive Syndrome-Trellis Code (STC) minimizes the embedding modification of the cover and increased audio quality by ensuring resistance against steganalysis. The established adaptive STC refined the structure of the unique parity-check matrix and attained a better safety-

DOI: 10.46121/pspc.53.4.18

defensive performance in steganalysis. Nevertheless, the GOAS leads to high distortion in audio quality owing to the aggressive allocation of data for concealment that compromises the perceptual quality of audio.

Hamza Kheddar and David Megias [41] introduced Steganography-based Interpolation and Auto-Encoding (SIAE) to embed secret data. The primary aim of the introduced approach was to securely transmit secret hidden-speech data within the cover image. SIAE embeds steganograms in four interpolated and Quantized Line Spectral Pair (QLSP) vectors. To reduce the modifications in the cover speech, SIAE employed a 1D autoencoder to compress the payload. Then, the secret data was expanded effectively at the receiver side to its original size in the decoding stage, which led to minimized steganographic quality loss and enhanced undetectability. However, SIAE suffers from limited capacity in terms of embedding large volumes of information without significant degradation, which introduces artifacts and compromises embedded data. Bekkar Laskar and Merouane Bouzid [42] developed a QIM for audio steganography in low bitrate speech streams. For codebook division, the developed approach employs Irving's stable room-mate issue method and performs a key-based mechanism that enhances security. Minimizing detectable cover changes was considered to strengthen resilience against the advanced steganalysis method across the embedding rate range. Nevertheless, QIM limits the capacity for secure embedding owing to the reduced bitrate, which restricts the available space for embedding secret information, making it easier for attackers to extract the secret information. Jinglei Wang and Kaixi Wang [43] presented a novel audio steganography method based on segmentation of the background and foreground of digital audio. To enhance the embedding capacity, an adaptive threshold was determined for the cover audio to segment the background and foreground areas. Thus, more bits were embedded into the sampling points. Moreover, a random selection method was used to select the sampling points to be embedded, which enhances security. The results indicate that the presented method has a greater embedding capacity and less impact on the security and imperceptibility. However, the presented method suffers from reduced robustness because segmentation errors lead to poor embedding of hidden data retrieval.

Table 1 shows a detailed summarization of existing methods in audio steganography which was categorized based on embedding domains. The table represents substitution-based, transform-based, predictive, coding, hybrid, and quantization based methods along with publication years and embedding domains. While these methods contribute to message hiding in various domains such as frequency, time, and decomposition models but suffer from vulnerability to noise, minimized robustness, high distortions, limited capacity for secure embedding, and less security. To solve this issue, this research proposes SACM- ASSQIM for encryption and text hiding in audio steganography. SACM provides high security by introducing complex and non-linear transformation that contains integration of pixel shuffling, divide-and-rotate operations, and integer-based diffusion mechanism. This approach ensures that encrypted data remains highly unpredictable, whereas ASSQIM reduces audio distortion and dynamically adjusts step size based on sensitive and non-sensitive index regions optimising embedding capacity. Overall, the proposed approach enhances robustness, security, and audio quality effectively in this research.

Table 1. Summarization of existing methods

Category	Methods	Embedding domain	References and Year
Substitution based	LSBM	Time domain	[33], 2021
	LSB-BMSE	Time domain	[37], 2022
	LSBPWLCM	Time domain + Chaotic map	[38], 2023
Transform-based	SVD	Decomposition model	[36], 2023
Predictive model	LPA	Time domian	[35], 2022
Coding-based	STC	Time domain	[40], 2021
Hybrid model	SIAE	Line spectral pair domain	[41], 2022
Quantization-based	QIM	Frequency domain	[42], 2024

2. Research Motivation and Novelty

1) **Research Motivation:** This research is motivated by the need for a highly secure and resilient steganographic method that hides messages in sensitive frequency regions. The proposed method addresses this problem by encrypting the message before embedding, which enhances both robustness and security. This method ensures that hidden data remain undetectable to unauthorized users, which makes it ideal for secure communication in sensitive applications to exchange the data confidentially.

DOI: 10.46121/pspc.53.4.18

Power System Protection and Control
ISSN:1674-3415

2) **Research Novelty:** This work introduces novel contributions in the field of audio steganography using the SACM-ASSQIM method to enhance security against attacks. 1) Frequency domain representation: FFT is applied to the audio signal to analyze the audio in the frequency domain, which provides a frequency vector based on the sampling rate and audio file. 2) Encryption: SACM achieves a complex and nonlinear transformation with a combination of the divide and rotation processes, pixel shuffling, confusion, and integer-based diffusion mechanism. The confusion process applies numerous Arnold transformations along with rotation angles to ensure that the encrypted output is highly unpredictable. This transformation renders spatial correlations resistant to attacks and enhances security. 3) Embed messages into the frequency domain: DCT is used to embed information into the low-frequency components of the audio signal, which achieves minimal perceptual distortion and robustness. 4) Embedding: ASSQIM dynamically adjusts the step size depending on the frequency analysis, particularly distinguishing between sensitive and non-sensitive frequency regions. In this method, lower step sizes are used in sensitive frequency regions (2000–5000 Hz) to minimize distortion, whereas larger step sizes (below 2000 Hz and above 5000 Hz) are employed in non-sensitive regions to enhance robustness against attacks. 5) De-embedding and decryption: De-embedding involves applying a DCT to stego audio that extracts message bits from the DCT depending on the step size thresholds. Then, the extracted binary data are converted back to text and decrypted using SACM, which ensures integrity and confidentiality.

PROPOSED METHODOLOGY

This research proposes SACM-ASSQIM to encrypt secret data from cover audio. First, a cover audio file is acquired and a secret message is established for embedding. To increase security, the message is encrypted using SACM, and it is converted into an 8-bit binary format. Subsequently, FFT is applied to the audio signal, which evaluates the audio in the frequency domain by generating a frequency vector based on the sampling rate. Subsequently, the DCT is used for frequency-domain audio signals to enable message embedding. ASSQIM is performed where a smaller step size is utilized for sensitive regions to minimize distortion, whereas a larger step size is employed in nonsensitive regions, which enhances robustness against attacks and noise. Finally, inverse progress is generated on the receiver side to reconstruct the secret data from the original audio. Figures 1 and 2 show the transmitter and receiver processes for SACM-ASSQIM.

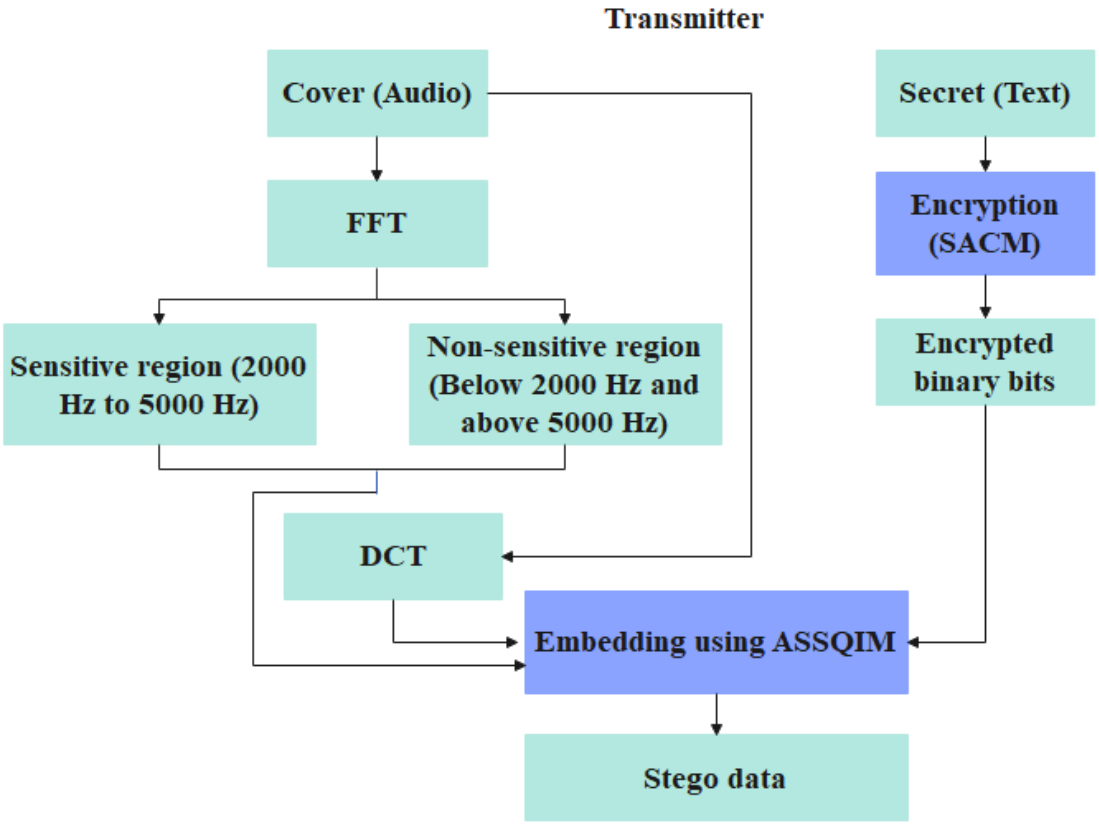


Figure 1. Encryption and embedded process for audio steganography at transmitter

DOI: 10.46121/pspc.53.4.18

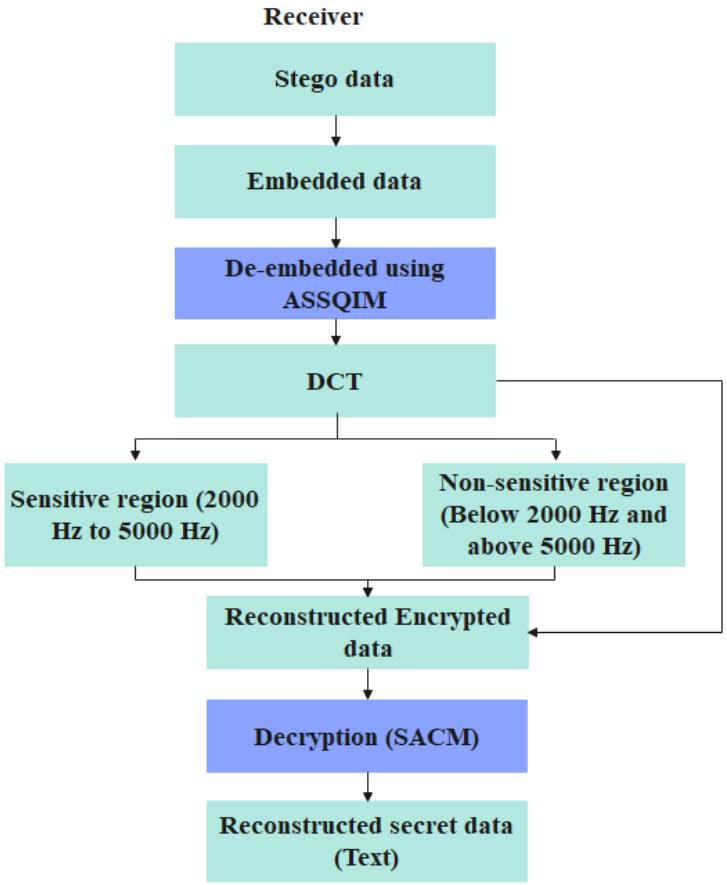


Figure 2. De-embed and decryption process for audio steganography at receiver

The workflow of the proposed methodology is discussed below in detail

Step 1: Initially, the cover audio is acquired from the TIMIT and GTZAN datasets, and the secret message is organized for embedding. The message is encrypted using SACM to improve security. Subsequently, the encrypted message is converted into an 8-bit binary format, and an 8-bit delimiter is added to represent the end of the hidden message.

Step 2: To determine the audio in the frequency domain, FFT is used on the audio signal, which generates a frequency vector that depends on the audio file length and sampling rate.

Step 3: The frequency spectrum is split into sensitive and non-sensitive regions, with frequencies between 2000 Hz and 5000 HZ are considered as sensitive because the human ear has the ability to distinguish sounds within this range. Frequencies below 2000 Hz and above 5000 Hz are categorized as nonsensitive regions. Logical masks are formed to distinguish these regions, which ensures that embedding in sensitive areas decreases to preserve audio quality.

Step 4: Subsequently, the DCT is applied to the frequency-domain audio signal to ease message embedding. ASSQIM is used where smaller step sizes are employed in sensitive regions to minimize perceptual distortions and larger step sizes are used in non-sensitive regions to enhance robustness against attacks and noise.

Step 5: According to the specified step sizes, the DCT coefficients are quantized and then binary message bits are embedded by slightly adjusting quantized value based on whether the bit indicates a “0” or “1.” Therefore, adaptive embedding provides robustness and imperceptibility.

DOI: 10.46121/pspc.53.4.18

Step 6: After embedding, an inverse DCT is performed to modify the coefficients for reconstructing the audio signal. The reconstructed audio contained a hidden message that is saved for further processing.

Step 7: The stego audio is read and transformed back into the frequency domain to extract the hidden message using the DCT. Subsequently, the embedded bits are retrieved from the DCT coefficients using predefined step-size thresholds. During embedding, the delimiter is appended to determine the endpoint of the hidden message.

Step 8: The extracted binary data is converted back to its textual form, and then SACM's inverse transformation is applied to decrypt the message, which ensures that the retrieved text matches the original secret message.

Step 9: The performance of the proposed method is analyzed by utilizing the SNR and Peak Signal Noise Ratio (PSNR) metrics to calculate the quality of the modified audio. Region-wise evaluations are performed to evaluate the impact of embedding on non-sensitive and sensitive frequency bands. Moreover, the robustness is analyzed by simulating attacks on the addition of Gaussian noise and low-pass filtering with PSNR and SNR, which are recalculated for each scenario.

Step 10: Finally, the extracted message is compared with the original message by utilizing similarity metrics, such as cosine. This metric ensures that the extracted message retains a high degree of similarity to the original message, thereby validating the efficiency of the embedding and extraction processes.

2.1 Cover data

The TIMIT and GTZAN datasets are used to establish robustness, capacity parameters, and imperceptibility to determine model performance in steganography. TIMIT [44] dataset includes 400 speakers with a sampling frequency of 16 kHz. All speakers stated ten phrases out of a total of 4000, and for evaluation purposes, 400 speech signals ranging from 2 to 4s are considered. In the GTZAN [45] dataset, uncompressed audio files are utilized as cover audio with 661,500 samples in durations ranging from 1 to 30s. Figure 3 shows the sample cover audio.

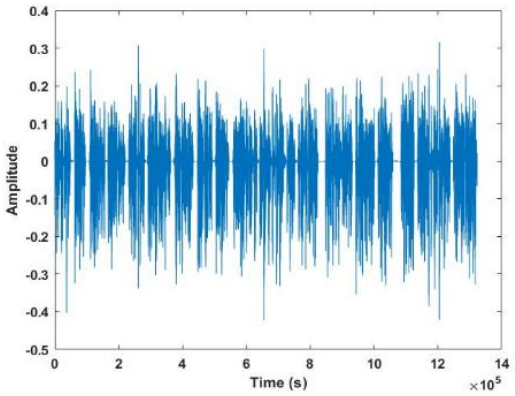


Figure 3. Sample cover audio

2.2 Transform model

After gathering the cover audio, FFT [46] is used to convert a time-domain audio signal into a frequency domain that enables the precise embedding of the secret data in less perceptible frequency components. FFT helps determine the frequency components of an audio file that contains secret information embedded in specific frequency bands without significantly affecting the audio quality. To split an audio signal into components below 200 Hz and above 5000 Hz, FFT is a better choice than the DCT. Because the FFT provides both phase and magnitude information, precise frequency filtering is enabled. Standard filtering techniques such as band-pass, high-pass, and low-pass filtering rely on FFT for frequency-domain processing. However, DCT primarily focuses on energy compaction used in compression formats, such as Moving Picture 3 (MP3) and Joint Photographic Expert Group (JPEG), which do not retain phase information. Moreover, DCT is not ideal for frequency-based filtering, especially when a precise frequency cutoff is required. Hence, FFT is used to ensure the robustness, imperceptibility, and effective utilization of the audio's spectral properties for secure communication. The FFT output provided the phase and magnitude of each frequency component. Later, the frequency vector is generated based on the sampling rate and length of audio that assists in identifying frequency bands, such as sensitive and non-sensitive region indices. The embedding process considers the sensitivity of the human auditory system to

DOI: 10.46121/pspc.53.4.18

Power System Protection and Control
ISSN:1674-3415

specific frequency ranges. Hence, the frequency spectrum is divided into sensitive and non-sensitive regions, whereas frequencies between 2000 Hz and 5000 Hz are considered sensitive because of the heightened sensitivity of the human ear in this range. Frequencies below 2000 Hz and above 5000 Hz are non-sensitive regions preferred for embedding, as modifications in these regions are less likely to introduce perceptual artifacts. This step size allocation balances robustness and imperceptibility, ensuring high-quality audio steganography without compromising security. Logical masks are applied to these regions, ensuring that embedding in sensitive regions is minimized to preserve audio quality. Figure 4 shows the frequency spectrum of the audio signal for the sensitive and non-sensitive regions, and Table 2 explains its purpose.

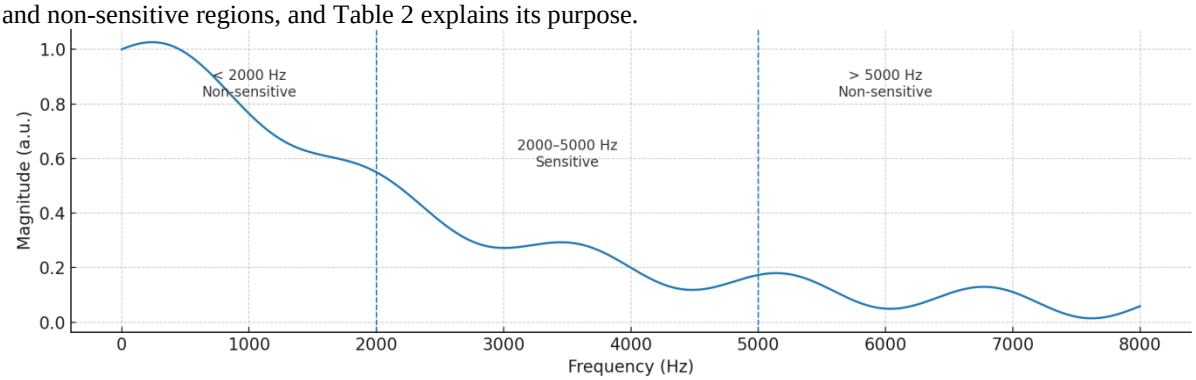


Figure 4. Frequency spectrum of audio signal showing sensitive and non-sensitive regions depending on human auditory system's perceptual characteristics

Table 2. Description of sensitive and non-sensitive regions with corresponding human auditory system sensitivity and purpose

Region	Frequency range (Hz)	Human Auditory system	Purpose
Non-sensitive	Below 2000 and above 5000	Lower	This range is categorized as non-sensitive due to human auditory system is less perceptive to smaller distortions. This enable larger embedding step size which strengthen hidden data and make more resistance to attacks
Sensitive	Between 2000 to 5000	Highest	In sensitive region, human ear is highly responsive to even slight modifications, so smaller embedding step size is applied. This reduce distortion which ensures hidden data does not establish audible artifacts and preserves imperceptibility as well as manage audio quality

Subsequently, the secret text undergoes an encryption process to enhance security following the transformation process and before embedding. This dual-layered encryption method and selective frequency embedding provided robust and secure steganographic communication.

2.3 Secret data

After the transformation model, the secret data based on the text are encrypted. Consider that the input secret text is "Hello, this is a secret message!" encrypted using SACM. For example, each character in the secret text is converted to the American Standard Code for Information Interchange (ASCII) values, as shown in Figure 5.

Original text	H	e	l	l	o	,		t	h	i	s
ASCII value of Original text	72	101	108	108	111	44	32	116	104	105	115
Original text		i	s		a		s	e	c	r	e
ASCII value of Original text	32	105	115	32	97	32	115	101	99	114	101
Original text	t		m	e	s	s	a	g	e	!	-
ASCII value of Original text	116	32	109	101	115	115	97	103	101	33	-

Figure 5. Original text and ASCII values of the original text

DOI: 10.46121/pspc.53.4.18

2.4 Encryption using Shuffled Arnold Cat Map (SACM)

This section discusses the encryption method used to secure secret data before embedding using the SACM confusion and diffusion process. SACM is used for encryption because of its ability to provide strong security in audio steganography via chaotic transformation. Traditional ACM suffers from periodicity, which means that after a fixed number of iterations, the transformation repeats, making it reversible. SACM solves this issue by creating an additional chaotic process that increases encryption strength and unpredictability. It includes shuffling and pixel rotation, which results in enhanced robustness of the encrypted data. Hence, this process provides that even minor changes in the input lead to a significant transformation in the output and enable high resistance to attacks. Initially, the ASCII values are converted into a square matrix by applying SACM, after which the matrix is filled row by row with ASCII values utilizing padding with zeros, while the cells are unused. The number of added zero paddings varies depending on the size of the secret input message. The matrix representation of ASCII values means that unused spaces are filled with zero padding, which helps prevent patterns and compromise security. Next, the SACM transformation is applied for a particular number of iterations (i=5). After every iteration, the scrambled matrix demonstrates the encrypted data. Finally, the encrypted text is extracted from the matrix by removing the padded cells and converting them back into characters. The Arnold map generated in the confusion process undergoes multiple rounds to obtain satisfactory outcomes, which leads to a longer execution time. Therefore, the Divide & Rotate and shuffling pixel processes are incorporated in a traditional Arnold map [47], which is termed SACM, for one round to establish a chaotic encryption scheme. The confusion process is performed using SACM to offer resistance to noise effects and data loss. Finally, the integration of integer value manipulation and pixel scrambling is generated using the original Arnold map, which enhances the resistance and sensitivity to differential attacks. The experimental evaluations indicate that five iterations of SACM provide an optimal balance because it provides a high security level without significant computational overhead. More iterations resulted in an enhanced execution time without significantly increasing the encryption strength. Figure 6 shows the original converted matrix of the text.

72	32	105	101	109	101
101	116	115	99	101	33
108	104	32	114	115	0
108	105	97	101	115	0
111	115	32	116	97	0
44	32	115	32	103	0

Figure 6. Original converted matrix of text

2.4.1 Secret Key and Pixel adding

The secret key is the foundation of the encryption process, and ensures security and sensitivity to small changes. It generates Arnold's map parameters and pixel manipulation during the diffusion step. The secret key = (a, b, u_i, d_i) is generated with a size of 2^{122} where (a, b) represents the parameter of Arnold's map generation integer type with $a = [1, 2^{24}]$ and $b = [1, 2^{18}]$. The $(u_i, d_i)(i = 1, 2, 3, 4, 5)$ indicates the sequence of initial values for the forward and backward values where $(u_i, d_i) == [1, 2^8]$. Adding random pixels establishes additional randomness and ensures that the encryption output differs even from identical inputs, which enhances security. The two types of integer arrays Ac and Ar are added randomly with sizes of M and $N + 1$ for a text of size (M, N). The Ar is the first row and Ac is added as the first column. The Ac and Ar arrays produce additional randomness in the encryption scheme by adjusting the pixel positions in both column and row orientations. By embedding these arrays in the first stage, the scrambling sequence becomes highly unpredictable, which increases the difficulty in the decryption stage without the correct key. After the secret key and adding pixels, a confusion process is performed to disrupt the spatial relationships among pixels and increase the unpredictability of the encryption.

2.4.2 Confusion process

After the secret key and adding a pixel, the encryption scheme shuffles the pixels and breaks the correlation between the adjacent pixels. The encryption method utilizes a scrambling process on the added pixels and generates SACM in a confusion process to identify the position of the new pixel. The confusion process rearranges the pixels to break spatial relationships and makes the text unrecognizable. Divide, & rotation, and pixel shuffling

DOI: 10.46121/pspc.53.4.18

are incorporated into the confusion process. The generated Arnold Map is divided into four sub-matrices $R_i(i = 1, 2, 3, 4)$ and each sub-matrix R_i is split into four fragments $F_{ij}(j = 1, 2, 3, 4)$. Next, each F_{ij} undergoes rotation by K_j degree, where K_j= (180, -90, 90, 180) increases the scrambling process by introducing multi-directional transformations. These varied rotation angles ensured that each divided fragment matrix underwent distant rotations, maximized disarray, and removed direct inversion possibilities. This results in a higher security and enhanced resistance to differential and statistical attacks. Moreover, scrambling provides security by rearranging the pixel positions, which disrupts the spatial relationships within the secret data. Hence, it is nearly impossible for attackers to reconstruct the original data without a decryption key. Next, the rotated fragments are concatenated in various positions compared to the sub-matrix to acquire R_i'. Another rotation of R_i' of K_i degree is followed, and finally, the rotated submatrix of numerous orders is concatenated. The pixels of the generated Arnold map are shuffled using the pixel values of the rotated Arnold map to evaluate the new coordinates of the pixels. Once the pixel positions are randomized effectively via the confusion process, the diffusion process ensures encryption robustness by propagating the modifications of pixels across the entire matrix.

2.4.3 Diffusion process

The encryption system effectively diffuses any changes in the pixel values after the confusion process. The encryption scheme utilizes a diffusion process that relies on two rounds of shuffling and integer-value manipulation. Equation (1) is applied to each pixel, starting with the first and proceeding forward until the final one. Subsequently, the pixels are shuffled using a generated Arnold map. Finally, equation (2) is employed to start from the final pixel and go backward. Therefore, the forward and backward mechanisms enable the spread of the modification to any pixel value via the entire text.

$$D_i' = \begin{cases} D_i + u_i \ mod\ 256, & i \in [1,5] \\ D_i + D_{i-5} \ mod\ 256, & i > 5 \end{cases} \tag{1}$$

$$D_i'' = \begin{cases} D_i' + d_i' \ mod\ 256, & j \in [MN-5, MN] \\ j' = MN - i + 1 & \\ D_j' + D_{j+5}'' \ mod\ 256, & j < MN - 5 \end{cases} \tag{2}$$

Where MN represents the number of rows and columns, j indicates the pixel index being processed during backward diffusion, j'denotes reverse index mapping, D_j' illustrates the pixel value from forward diffusion at position j, D_{j+5}'' refers to the pixel value from backward diffusion, 5 pixels ahead of j, and $mod\ 256$ ensures that the pixel values remain within the range [0, 255]. Because the secret data are in text form, it is first converted into ASCII values, with each character indicating an 8-bit range of 0 to 255. The "mod 256" provides that transformed values remain within this range, which manages compatibility and prevents overflow with standard text encoding. This ensures that, after encryption and diffusion, all values are mapped correctly back with the associated characters during decryption. The 8-bit delimiter ([00000000]) is appended to signify hidden message completion. SACM is used for both encryption and decryption processes. In the embedding pipeline, SACM encrypts secret messages before embedding, whereas in decryption, it reconstructs the original text from the extracted message, which significantly increases security. Figure 7 shows the encrypted matrix. Following encryption, the transformed text is embedded using DCT, which leverages the frequency domain of the audio signal to achieve imperceptible and robust steganography.

72	103	32	108	115	115
0	116	105	0	99	32
97	97	108	101	105	44
101	104	33	101	32	0
32	101	109	115	111	115
116	101	32	115	0	114

Figure 7. Encrypted matrix

DOI: 10.46121/pspc.53.4.18

Power System Protection and Control
ISSN:1674-3415

2.5 Embedding Message into the Frequency Domain using DCT

After encryption, the message in the frequency domain is embedded using a DCT [48], making it less perceptible to human hearing and ensuring minimal distortion of the audio signal. To embed information in the low-frequency components of the audio signal and convert the audio signal to a low-frequency signal, DCT is more effective than the FFT. DCT focuses most of the signal energy on low-frequency components, which is beneficial for embedding information with less perceptual distortion. Unlike FFT, DCT manages real numbers (no phase), simplifying the embedding process. DCT is primarily employed in audio compression to make the embedded information more robust to processing and compression. However, FFT provides complex numbers with phase information, allowing for more accurate embedding, and is prone to introducing audible artifacts. However, FFT does not focus on the energy at low frequencies as effectively as DCT. Therefore, the DCT is used for low-frequency embedding tasks because of its simplicity and energy compaction. Embedding at higher frequencies exploits redundancy, which reduces its impact on audio quality. The DCT ensures secure and imperceptible embedding by preserving the integrity of the carrier signal. The converted binary bits are grouped into segments that are embedded in the selected DCT coefficients of the block. The DCT is an important transformation utilized in image steganography. For consistency, it decomposes the audio signal into low (D_{DC}) and high (D_{AC}) frequencies. The mathematical formula for 2-D DCT is expressed in Equation (3), and Figure 8 represents a frequency domain signal.

$$D(s,r) = \delta(s)\delta(r)\sum_{n_1=0}^{M-1}\sum_{n_2=0}^{M-1} S(n_1,n_2)cos[\frac{(2n_1+1)s\pi}{2M}]cos\,[\frac{(2n_2+1)r\pi}{2M}] \quad (3)$$

Where $S(n_1,n_2)$ indicates the pixel value at (n_1,n_2) position, $\delta(s)$ and $\delta(r)$ determine the normalized coefficients, which are calculated using Equation (4).

$$\delta(s)\delta(r) = \begin{cases} \sqrt{\frac{1}{N}}, & s,r = 0 \\ \sqrt{\frac{2}{N}}, & s = 1,2,3,\dots,M-1 \\ & r = 1,2,3,\dots,M-1 \end{cases} \quad (4)$$

After the DCT embedding process, ASSQIM is applied to further refine the embedded data, which improved the robustness of steganography. Adapting quantization indices based on embedded information provides security and imperceptibility to the hidden data.

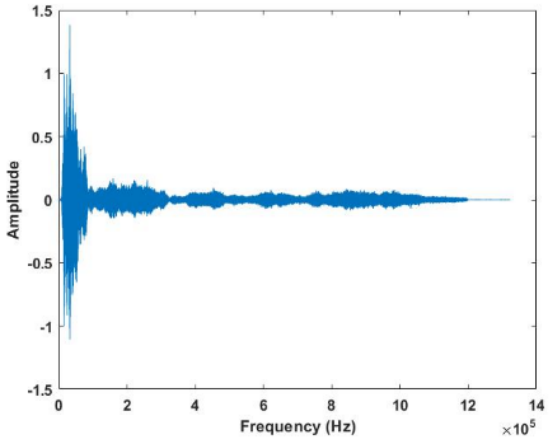


Figure 8. Frequency domain signal

2.6 Embed data using ASSQIM

ASSQIM enhances the DCT embedding by dynamically adjusting the quantization step depending on the frequency sensitivity. This integration allows high-frequency components that are less perceptible to the human auditory system to transmit more hidden information, while maintaining imperceptibility. The sensitive region index uses a smaller step size of 0.01, corresponding to low-frequency components, because the human ear is highly perceptive in this range. This reduces distortion, preserves audio quality, and ensures that the hidden data remain imperceptible. The nonsensitive region index employs a larger step size of 0.1, corresponding to high-frequency components, where the human ear is less sensitive. This enhances the robustness against attacks and noise, which is termed ASSQIM. During quantization, the DCT coefficients are applied based on the step size. To embed the message bits, the quantized value of a coefficient is adjusted slightly to represent a binary 0 or 1. This method ensures a balance between robustness and embedding capacity, which makes it ideal for secure data-

DOI: 10.46121/pspc.53.4.18

hiding. The process involves two primary stages: embedding and detection. In the embedding stage, the following phases are executed.

Lattice construction and Labelling: A pair of nested lattices A_f and $A_c \subset A_f$ are assumed with generators G_f and G_c. In labelling, message space M is represented as $= z_\alpha^N = \{0, \ldots, \alpha - 1\}^N$. Consider a host signal vector s for hiding an information vector; $m_i \in M$ s known as labelling, is utilized. The mathematical formula for labelling is expressed in Equation (5).

$$d_i = \mathcal{L}(m_i) \triangleq G_f . \emptyset(m_i) \tag{5}$$

Where $\emptyset: Z_\alpha^N \to R^N$ represents a natural mapping function.

Quantization: The encoder of QIM quantizes the host signal s to the closest lattice point in A_i by applying Equation (6).

$$s_w = Q_{\wedge_i}(s) = Q_{\wedge_c}(s - d_i) + d_i \tag{6}$$

Where the index i in $\wedge_i$ represents closet $\wedge_i$ is associates to message m_i i.e., it is utilized for transmitting the message m_i. The payload of the embedding process is α^Nresults in a code rate per dimension of $R = log\alpha$. In the detection stage, the watermark is embedded from the watermarked content to determine its integrity and presence. The detection step of QIM is represented as follows. Consider the vector of the received signal $y = x + n$ where n indicates the perturbation noise. QIM executes the subsequent de-quantization step to extract the closest representative using Equation (7):

$$j = \underset{i \,\epsilon\, \{1, 2, \ldots, \alpha^N\}}{argmin} dist(y, \wedge_i) \tag{7}$$

Where $dist(y, \wedge_i) \triangleq min_{\rightarrowtail \epsilon \wedge_i} ||y - \rightarrowtail||$. If the noise is sufficiently small such that $Q_{\wedge_f}(n) = 0$, later the embedded message $\hat{m}$ acquired from delabeling is correct. De-quantization provides accurate extraction because it retrieves the closest quantized value, thereby ensuring that the embedded bit is recovered correctly. If the perturbation noise remains at an acceptable threshold, the extracted message retains high fidelity. The estimated message is represented by Equation (8).

$$\hat{m} = G_f^{-1} . d_j \; mod \; \alpha \tag{8}$$

The embedded secret data are fed into the inverse DCT to reconstruct the modified audio. The primary aim is to modify the specific DCT coefficients based on the information to be embedded. Clockwise embedding is chosen because it follows a structured traversal pattern that maintains spatial coherence by evenly distributing data across the frequency domain, which ensures that embedded data remain less susceptible to attacks such as low-pass filtering. The threshold for the ASSQIM step size is adaptive, based on the frequency distribution components of the host audio signal. This allows step sizes to adjust dynamically which minimize distortion while enhancing robustness and Figure 9 shows a stego audio_text.

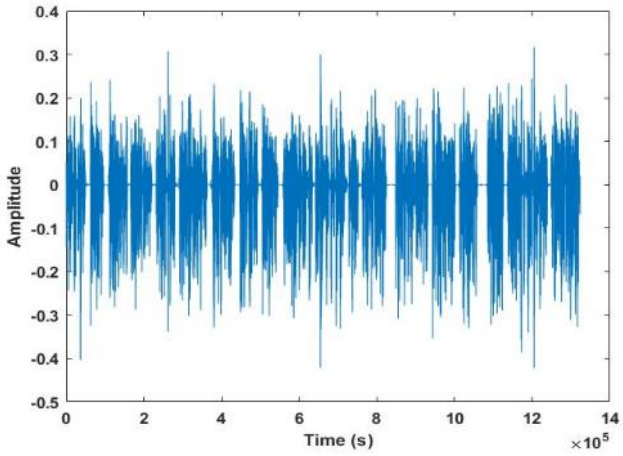


Figure 9. Stego audio_text

2.7 De-embedding

The de-embedding process is similar to the embedding process and involves extracting hidden information from stego data. To extract the hidden message, the modified audio file is read, and subsequently, DCT is applied to the stego data. The message bits are extracted from the DCT coefficients using step size thresholds. The extraction process is accomplished based on the de-embedding process, and the coefficients are adjusted to determine whether it represents a 0 or 1. The delimiter is subsequently used to identify the end of the hidden message.

DOI: 10.46121/pspc.53.4.18

Finally, the binary message is converted back into text and decrypted using SACM, which restores the hidden message. This process ensures secure embedding and decryption of the secret data, which preserves both confidentiality and data integrity. Algorithm 1 shows the encryption and decryption steps using SACM.

Algorithm 1:
Encryption Steps
Step 1: Input Text
i) Input is: 'Hello, this is a secret message!'
Step 2: Convert Text to ASCII
ii) Each character is converted to an ASCII value; for example, H -> 72, e -> 101, l -> 108.
Step 3: Make a Square Matrix
iii) Using SACM, a data is arranged into a square matrix
iv) For unused cells, the matrix is filled row by row with ASCII values by applying padding with zeros.
Step 4: Display Original Matrix
v) A matrix is obtainable to visualize ASCII value organization
Step 5: Apply SACM Transformation
vi) where a and b represent transformation constants, and n indicates the matrix size (rows and columns).
vii) Apply transformation iteratively for a number of iterations (iterations = 5).
Step 6: Resulting Encrypted Matrix
viii) After each iteration, scrambled matrix represents encrypted data
Step 7: Convert Back to ASCII
ix) Extract encrypted text from a matrix by eliminating padded cells and adapting it back to characters using char ().
Decryption Step
Step 8: Apply Inverse SACM
x) Accomplish inverse transformation for the same number of iterations
Step 9: Convert Decrypted Matrix Back to Text
xi) The original text is extracted by reading the matrix row by row to the original text length.

Final Output
Encrypted Text: Encrypted version of the text is represented by a scrambled character sequence
Decrypted Text: After applying inverse SACM, decrypted text matches the original text.
Pseudo code for SACM- ASSQIM
Input: Secret text and output: decrypted text
Encryption process
Input: Secret text
Convert each character of secret text to ASCII values
ASCII values are organized into a square matrix M, where M represents a 2D array that stores ASCII values.
Pad unused cells with zeros if needed
Show original matrix M
Apply SACM transformation:
a. Set transformation constants
b. Set number of iterations n, where n denotes predefined number of transformations for scrambling data
c. For $i = 1$ to n
    For each element in M:
    Apply transformation process
Store resulting encrypting matrix M_{enc}
Transform encrypted matrix back to representation of ASCII
Output: Encrypted text
Decryption process
Input: Encrypted text
Transform encrypted text back to ASCII values and form matrix M_{enc}
Apply inverse SACM transformation:
a. For $i = 1$ to n:
        For each element in M_{enc}:
        Apply inverse transformation process:

DOI: 10.46121/pspc.53.4.18

Extract ASCII values row by row
Transform ASCII values back to characters
If decrypted text matches secret text:
    Output: Decryption successful
Else:
    Output: Decryption failed
End

METRICS UTILIZED FOR EVALUATING STEGANOGRAPHY

The evaluation metrics for steganography are justified through robustness (attacks), imperceptibility (using Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE), Signal-to-Noise Ratio (SNR), and Log Spectral Distance (LSD)), integrity of secret information (via Normalized Correlation (NC)), steganalysis (Manhattan distance), and capacity analysis (utilizing capacity and embedding rate). These parameters are employed to assess the quality and effectiveness of the proposed method, which are discussed in detail below.

5.1 Imperceptibility

It is exactitude among original cover audio and stego audio as well as between reconstructed and secret message. SNR measures distortion in imperceptibility between the input and output. Thus, the SNR determines the quality of the output signal after the embedding process in decibels (dB), using Equation (9). PSNR is measures in dB, and the maximum SNR ratio of an audio using equation (10). After embedding the secret data, MSE measures the distortion introduced into the cover audio in Equation (11), whereas LSD is used to determine the difference between the spectral properties of the cover and stego signals.

$$SNR = 10log_{10}\frac{\sum_{j=1}^{N}[s_j']^2}{\sum_{j=1}^{N}[s_j'-s_j]^2} \quad (9)$$

$$PSNR = 10log_{10}\frac{S_J^2}{E} \quad (10)$$

$$MSE = \frac{1}{m}\sum_{\tau=1}^{m}(\lambda_\tau - \widehat{\lambda_\tau})^2 \quad (11)$$

where m represents the number of data points, λ_τ indicates the actual values, $\hat{\lambda}_\tau$ denotes the mean values, s_j illustrates the cover speech signal, M and N determine the size of the host and watermarked signals that contain the same length.

5.2 Integrity of secret information

During decoding, a secret message is extracted, and the NC coefficient is employed to compute secret message integrity. NC evaluates the difference between the extracted and original secret messages, which is formulated in Equation (12):

$$NC = \frac{\sum_{i=1}^{N} X_i \times X_i'}{\sqrt{\sum_{i=1}^{N} X_i^2} \times \sqrt{\sum_{i=1}^{N} X_i'^2}} \quad (12)$$

Where X_i and X_i' indicate the original and extracted secret messages, respectively, N denotes the secret message length. The NC value is between 0 and 1; when it is 1, the extracted secret message is the same as the original message.

5.3 Steganalysis

The Manhattan distance is utilized to evaluate the distance between two points in a grid-like path. Initially, histograms of the cover and stego audio coefficients are drawn, and then the Manhattan distance of the two histograms is computed using equation (13).

$$d_{manh} = \sqrt{\sum_i^d |C_i - O_i|} \quad (13)$$

where d represents the number of histogram bars, C and O determine the probability densities of the coefficient histogram of the cover and stego audio, respectively.

5.4 Capacity, Embedding rate, Histogram error rate

Capacity is defined as the number of bits embedded in the watermark divided by the number of units in the cover medium, using Equation (14):

$$Capacity = \frac{Number\ of\ bits\ embedded\ in\ watermark}{Number\ of\ units\ in\ cover\ medium} \quad (14)$$

DOI: 10.46121/pspc.53.4.18

The embedding rate determines the number of secret messages embedded per byte of the cover audio. Notably, the embedding rate of the proposed method is directly proportional to the sampling rate of the cover audio. It is measured in bps, representing the number of secret message bits embedded in each cover audio sample using Equation (15).

$$Embedding\ rate = \frac{Number\ of\ bits\ embedded\ in\ watermark}{Number\ of\ units\ in\ cover\ medium} \times 100 \quad (15)$$

The histogram error rate measures the difference between the original audio cover and the stego audio using Equation (16). A lower histogram error rate represents a higher imperceptibility and better resistance to statistical attacks.

$$Histogram\ error\ rate = \frac{\sum_{i=1}^{N}(His_c - His_s)^2}{\sum_{i=1}^{N} His_c^2} \quad (16)$$

Where His_c and His_s represents histograms of cover and stego audios.

Experimental Results

The proposed SACM-ASSQIM is simulated using MATLAB R2020b with a system requirement of 128 GB RAM, an i5 Intel processor, and a Windows 10 operating system. The primary aim of SACM-ASSQIM is to determine with audio as the cover data and text as the secret data. Additionally, SACM-ASSQIM is analyzed with audio as cover data and image and audio as secret data. The comparative analysis outcomes are evaluated based on audio as cover data using the TIMID and GTZAN datasets, and text and image as secret data.

Performance Analysis

Table 3 provides a performance analysis of different embedding techniques. Existing techniques such as the Least Significant Bit (LSB), Discrete Cosine Transform (DCT), Singular Value Decomposition (SVD), and QIM are compared with ASSQIM techniques. Compared to these techniques, ASSQIM obtains a better PSNR of 59.061 dB, 54.498 dB, and 94.09 dB due to its ability to adaptively modify the quantization step size depending on audio signal characteristics. This indicates that the embedding process introduces less distortion, thereby preserving audio signal quality. Based on the host signal characteristics, ASSQIM decreases distortion by adaptively adjusting the quantization step size. Unlike static quantization, which employs a fixed step size, ASSQIM selectively adjusts the step sizes to balance robustness and imperceptibility. Moreover, ASSQIM prevents quantization in sensitive regions and minimizes the noise amplification. Hence, this adaptability increases its performance in terms of PSNR when maintaining steganographic hidden data integrity. Figure 10 shows standard MATLAB images as secret data embedded into cover data, called stego audio images, and Figure 11 represents sample audio as secret data embedded into cover data called stego audio_ audio.

Table 3. Performance analysis of different embedding techniques

Embedding methods	SNR (dB)	PSNR (dB)	MSE	LSD	NC	Embedding rate (bps)	Capacity value (bps)
Audio as cover data and text as secret data							
LSB	39.897	54.398	1.99E-06	1.387	0.9912	0.0150	0.00015
DCT	28.384	43.298	9.92E-06	1.276	0.9945	0.0170	0.00017
SVD	24.974	41.208	3.50E-05	1.462	0.9968	0.0185	0.00018
QIM	40.198	56.208	1.49E-07	0.998	0.9975	0.0190	0.00019
ASSQIM	41.978	59.061	1.24E-07	0.944	0.998	0.0200	0.0002
Audio as cover data and image as secret data							
LSB	34.522	53.124	4.49E-07	4.102	0.9978	1.9542	0.0193
DCT	33.201	49.214	4.80E-08	4.554	0.9986	1.7310	0.0217
SVD	32.105	48.4224	5.89E-07	3.912	0.9967	1.6845	0.0186
QIM	35.201	52.423	5.96E-07	3.7059	0.9989	2.1143	0.0229
ASSQIM	37.415	54.498	3.55E-07	3.829	0.9999	2.4774	0.0248
Audio as cover data and audio as secret data							
LSB	45.327	61.204	8.42E-10	0.054	0.9852	65.0	0.65
DCT	50.128	68.402	9.31E-10	0.041	0.9901	70.0	0.70
SVD	47.845	66.110	11.98E-10	0.045	0.9875	72	0.72
QIM	52.763	71.885	9.94E-10	0.0325	0.9937	80	0.80

DOI: 10.46121/pspc.53.4.18

ASSQIM	78.153	94.09	7.67E-11	0.023	1	100	1

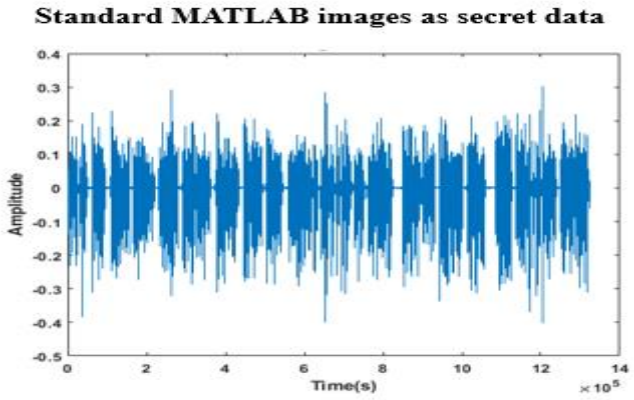


Figure 10. Standard MATLAB images as secret data embedded into cover data called stego audio image.

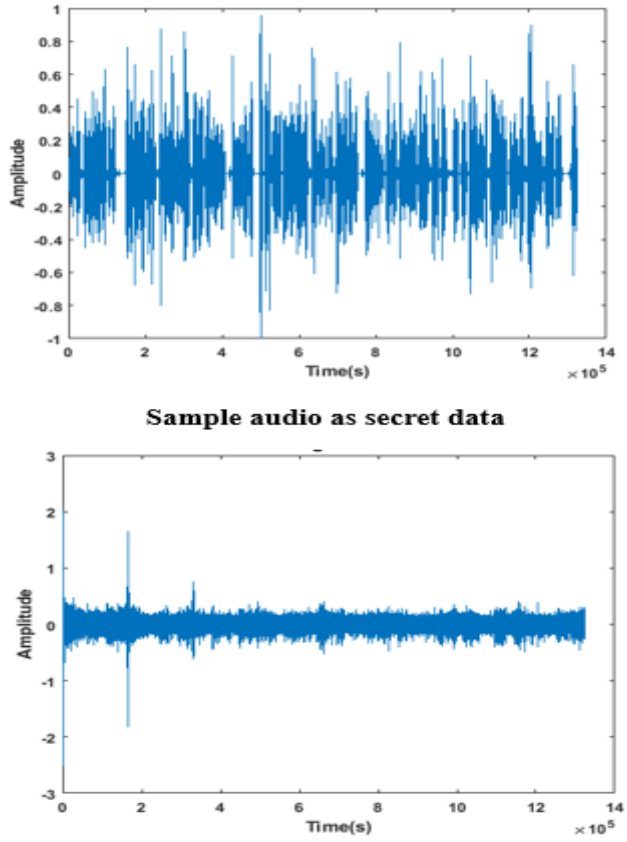


Figure 11. Sample audio as secret data embedded into cover data called stego audio_ audio

DOI: 10.46121/pspc.53.4.18

Table 4 presents the performance analysis for different step sizes for the sensitive regions. Different step sizes (0.1, 0.07, 0.05, 0.03, and 0.01 are considered to evaluate the model performance. Compared to different step sizes, the 0.01 obtains a PSNR of 59.140 dB, 34.832 dB, and 91.475 dB are obtained because it produces a finer quantization level that leads to minimal distortions. In the mid-frequency range (2000–5000 Hz), the human auditory system is more sensitive to distortions because of its natural frequency response. These frequencies are vital for speech perception, making even minor alterations more noticeable. By employing a smaller step size, the embedding process conserves the fidelity of the original signal more effectively to minimize the impact. A larger step size creates coarser quantization, which leads to more noticeable artifacts and a lower PSNR. Therefore, a 0.01 step size produces higher signal quality and better imperceptibility in this region index.

Table 4. Different step size analysis for sensitive region

Different step size	SNR (dB)	PSNR (dB)	MSE	LSD	NC
Audio as cover data and text as secret data					
0.1	27.895	45.142	2.89E-06	1.625	0.9798
0.07	31.182	48.124	1.54E-06	1.430	0.9995
0.05	33.322	51.175	7.21E-07	1.300	0.9990
0.03	37.836	55.627	2.59E-07	1.096	0.9997
0.01	41.560	59.140	1.15E-07	0.961	0.9999
Audio as cover data and image as secret data					
0.1	37.446	54.309	3.50E-07	3.596	0.9857
0.07	20.558	37.421	5.71E-05	6.552	0.9956
0.05	23.411	40.274	4.87E-06	5.895	0.9977
0.03	27.855	44.718	3.18E-06	5.043	0.9991
0.01	42.968	59.832	3.10E-05	2.997	0.9999
Audio as cover data and audio as secret data					
0.1	78.992	85.067	7.68E-11	0.025	0.9785
0.07	75.804	82.551	9.43E-11	0.028	0.9987
0.05	76.985	83.318	8.98E-11	0.027	0.9992
0.03	77.462	84.022	8.40E-11	0.026	0.9996
0.01	80.218	91.475	1.052E-10	0.016	0.9995

Table 5 shows the different step sizes for the non-sensitive region index. The 0.1 step size achieves a better PSNR of 59.088 dB compared to 0.05, 0.03, 0.02, and 0.01 due to the balance between the signal distortion and embedding capacity. In non-sensitive regions, this is less problematic because the human auditory system is less responsive to changes outside the 2000-5000 Hz range; therefore, using a larger step size of 0.1 allows more efficient embedding with minimal perceptual distortion. This avoids underutilization of the embedding capacity and establishes less quantization noise relative to the region's sensitivity, which enhances the PSNR. Moreover, smaller step sizes lead to an underutilized capacity and slightly coarser quantization effects in low-perception zones, which leads to relatively lower PSNR values. Hence, 0.1 obtains an optimal balance of capacity and imperceptibility in non-sensitive regions results in a higher PSNR.

Table 5. Different step size analysis for non-sensitive region

Different step size	SNR (dB)	PSNR (dB)	MSE	LSD	NC
Audio as cover data and text as secret data					
0.05	28.081	45.874	3.15E-06	1.599	0.9987
0.03	31.090	48.172	1.52E-06	1.431	0.9956
0.02	34.107	51.050	7.85E-07	1.282	0.9977
0.01	38.547	55.490	2.83E-07	1.088	0.9991
0.1	42.668	59.088	1.26E-07	0.953	0.9999
Audio as cover data and image as secret data					
0.05	20.556	37.629	1.72E-05	6.965	0.9956

DOI: 10.46121/pspc.53.4.18

0.03	23.383	40.456	9.00E-06	6.315	0.9977
0.02	27.782	44.855	3.27E-06	5.446	0.9991
0.01	17.888	34.96	3.19E-05	7.534	0.9919
0.1	37.404	54.476	1.56E-07	3.919	0.9999
Audio as cover data and audio as secret data					
0.05	79.650	93.845	8.89E-11	0.015	0.9995
0.03	80.781	94.412	8.12E-11	0.014	0.9998
0.02	81.423	94.802	7.93E-11	0.014	0.9999
0.01	78.110	93.218	9.62E-11	0.016	0.9990
0.1	82.218	95.019	6.77E-11	0.013	1

Table 6 presents a performance analysis of the smaller and larger step sizes for the ASSQIM. Compared to different step size analyses, (0.01, 0.1) obtains a better PSNR of 59.061 dB, 54.047 dB, and 94.09 dB due to its ability to finely tune the quantization step size depending on the characteristics of the audio signal. A smaller step size reduces distortion and preserves perceptual quality, leading to a lower MSE and higher PSNR. Moreover, it adaptively ensures that embedding creates minimal changes that maintain audio reliability in nonsensitive regions. The combination of step sizes of (0.01, 0.1) ensures perceptually sensitive regions represent fewer distortions, whereas non-sensitive regions make robust embedding. This balanced result achieves optimal security and imperceptibility in steganography.

Table 6. Analysis of smaller step size and larger step size for ASSQIM

Different Step size (Sensitive, non-sensitive)	SNR (dB)	PSNR (dB)	MSE	LSD	NC
Audio as cover data and text as secret data					
(0.3,0.05)	27.989	45.071	3.11E-06	1.598	0.9992
(0.1,0.03)	32.43	49.513	1.12E-06	1.360	0.9997
(0.07,0.02)	34.016	51.098	7.77E-07	1.282	0.9998
(0.05,0.01)	35.956	53.039	4.97E-07	1.192	0.9998
(0.01,0.1)	41.978	59.061	1.24E-07	0.944	0.9999
Audio as cover data and image as secret data					
(0.3,0.05)	10.422	27.515	2.18E-06	9.434	0.9566
(0.1,0.03)	17.586	34.679	3.08E-05	7.508	0.9913
(0.07,0.02)	20.116	37.21	1.72E-05	6.940	0.9951
(0.05,0.01)	22.958	40.052	8.94E-06	6.285	0.9974
(0.01,0.1)	36.954	54.047	1.56E-07	3.829	0.9999
Audio as cover data and audio as secret data					
(0.3,0.05)	42.37	58.21	1.94E-09	0.087	0.975
(0.1,0.03)	58.92	75.06	3.50E-10	0.058	0.990
(0.07,0.02)	64.33	81.70	1.72E-10	0.045	0.995
(0.05,0.01)	70.18	88.12	1.03E-10	0.033	0.998
(0.01,0.1)	78.153	94.09	0.67E-11	0.023	1

Robustness Analysis with an Attack

Robustness refers to the ability to retrieve a secret message successfully, with or without minimal errors. To test the reliability and robustness of the proposed method, four primary attacks, namely Gaussian filtering, low-pass filter, random noise, and time scale modification, are conducted. The attacks are briefly discussed below.

1. Gaussian filtering attack:

Gaussian filtering is used to smooth or blur signals by minimizing the high-frequency components. It is primarily applied to eliminate noise; however, in steganography, it tests how well hidden data withstands signal smoothing. To check the robustness, hidden messages are evaluated which remains intact or reliably recovered with minimal distortion after filtering.

DOI: 10.46121/pspc.53.4.18

2. Low pass filter attack:
A low-pass filter allows low-frequency signals to pass while attenuating high-frequency components, which are commonly used to eliminate noise or compress signals. A low-pass filter assesses the system's ability to recover secret messages post-filtering, which represents tolerance to frequency-domain manipulation.

3. Random noise attack:
This attack involves adding unpredictable random noise to the stego signal to simulate the transmission distortion. It evaluates the accuracy of the embedded message extracted from a noisy signal.

4. Time-Scale Modification attack (speeding up audio):
It alters the speed of audio playback without modifying its pitch. In a speeding-up time-scale modification attack, audio is played faster than the original speed, thus minimizing its duration. This results in incomplete or inaccurate extraction of hidden messages.

Table 7 determines the performances of the different attacks. The performances of Gaussian filtering, low-pass filtering, random noise, and time scale modification (speeding up audio) are compared without noise. The proposed SACM achieves better performance even in the presence of noise owing to its ability to preserve the original audio information that prevents degradation caused by noise. The SACM effectively prevents the degradation of audio quality by ensuring a randomized embedding structure. This minimizes the impact of noise and filtering attacks by avoiding localized distortions, leading to higher resilience.

Table 7. Performance analysis of different attacks

Different attacks	SNR (dB)	PSNR (dB)	MSE	LSD	NC
Audio as cover data and text as secret data					
Gaussian filtering	12.924	30.007	9.10E-05	25.982	0.9754
Low pass filter	11.928	17.984	9.27E-05	0.004	0.1542
Random noise	13.976	31.208	8.28E-05	16.086	0.97285
Time scale modification (Speeding up audio)	38.209	54.109	2.04E-07	1.254	0.085
Without noise	41.978	59.061	1.24E-07	0.944	0.999
Audio as cover data and image as secret data					
Gaussian filtering	12.922	29.991	9.99E-05	28.468	0.9754
Low pass filter	2.2357	13.847	8.23E-05	31.173	0.1542
Random noise	12.477	29.801	6.45E-04	29.701	0.9728
Time scale modification (Speeding up audio)	30.522	14.802	5.28E-06	76.242	0.0047
Without noise	37.415	54.498	3.5507e-07	3.8293	0.9999
Audio as cover data and audio as secret data					
Gaussian filtering	16.99	32.93	2.36E-06	7.068	0.9901
Low pass filter	3.865	12.072	3.58E-05	45.134	0.2702
Random noise	22.63	38.712	3.32E-05	5.130	0.9972
Time scale modification (Speeding up audio)	2.417	13.657	5.36E-05	8.845	0.0310
Without noise	78.153	94.09	7.67E-11	0.0232	1

Steganalysis Test
Table 8 presents the performance analysis of the stegnalysis results for the proposed method using the Manhattan distance, which quantifies the difference between the statistical evaluation of cover and stego audio signals. Lower values indicate higher imperceptibility and resistance. While text is utilized as secret data, the Manhattan distance remains relatively low, ranging from 0.1180 to 0.2377 across bit rates of 64 to 192 kbps, which represents minimal distortion. Similarly, embedding images led to a slightly higher distance of 0.1500 to 0.2603. Moreover, when audio is employed as a secret message, the distance significantly increases to approximately 0.8680 across all bits, demonstrating that embedding audio introduces more noticeable changes. Overall, the proposed method represents strong resistance to steganalysis for text, image, and audio payloads.

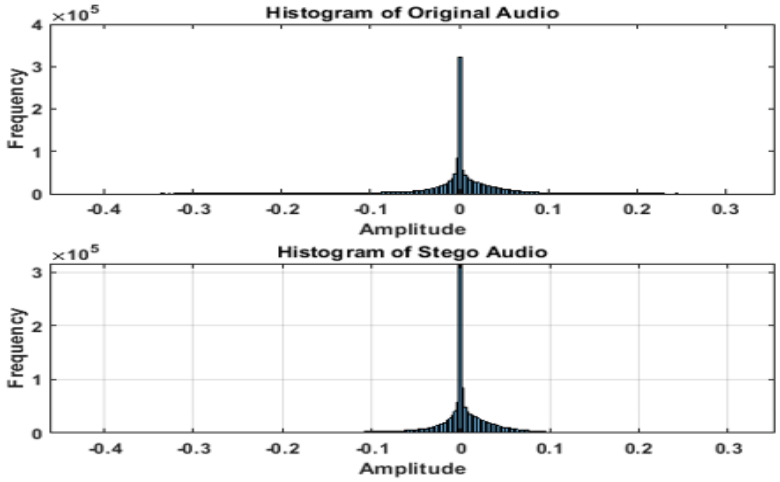
DOI: 10.46121/pspc.53.4.18

Power System Protection and Control
ISSN:1674-3415

Table 8. Performance analysis steganalysis results for proposed method

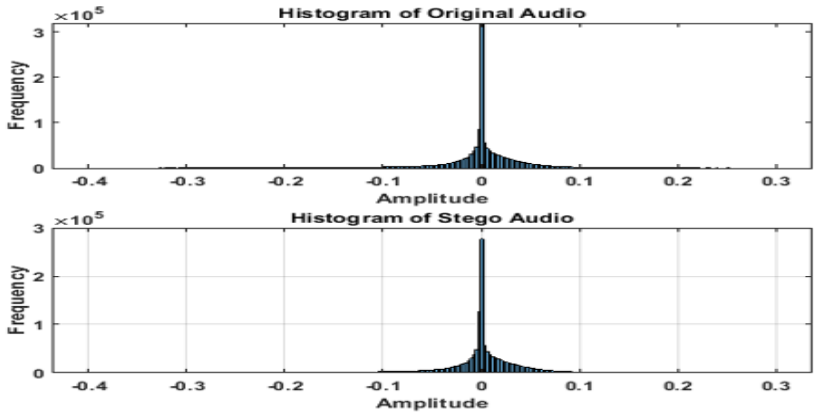
Bit rate(kbps)	Manhattan Distance
Audio as cover data and text as secret data	
64kbps	0.1180
128kbps	0.2373
192kbps	0.2377
Audio as cover data and image as secret data	
64kbps	0.1500
128kbps	0.2594
192kbps	0.2603
Audio as cover data and audio as secret data	
64kbps	0.8680
128kbps	0.8684
192kbps	0.8686

Figures 12, 13, and 14 present a graphical comparison of histograms between the original audio and the corresponding stego audio text, stego audio image, and stego audio audio. This analysis contained cover audio, which demonstrated that the amplitude distributions before and after embedding remained nearly the same. This minimal variation indicates that the embedding process does not introduce statistical artifacts. As a result, the histogram error rate remained close to zero across all tested audio clips. These findings confirm that the proposed method is highly resistant and ensures that hidden data remains undetectable.



Histogram error rate = 0.001451

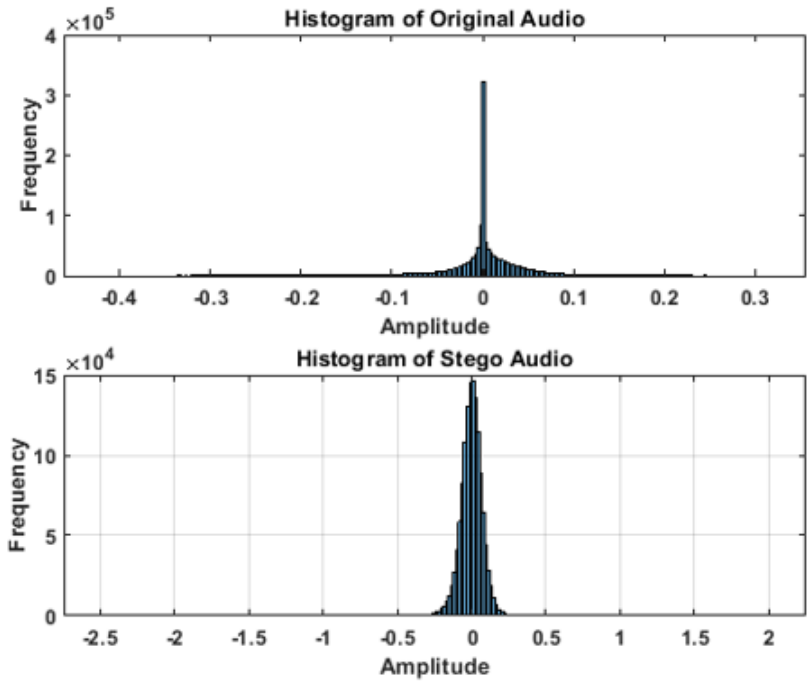
Figure 12. Histogram error rates for cover audio and their corresponding stego audio text



Histogram error rate = 0.026571

Figure 13. Histogram error rates for cover audio and their corresponding stego audio image

DOI: 10.46121/pspc.53.4.18



Histogram error rate = 0.745120

Figure 14. Histogram error rates for cover audio and their corresponding stego audio_audio

Table 9 shows the evaluation of the ablation study for the individual components. Compared with individual components, SACM-ASSQIM achieves a high PSNR because it effectively balances robustness and imperceptibility via adaptive embedding. Traditional QIM introduces distortion in sensitive frequency bands because of fixed step sizes, whereas ASSQIM adjusts the quantization step based on frequency regions, which minimizes the reduction in perceptual distortion. Unlike ACM and its variants, the proposed SACM-ASSQIM integrates encryption with adaptivity, which allows secret data to be effectively hidden. Consequently, the proposed method conserves more of the original audio structure, which leads to a high PSNR.

Table 9. Evaluation of ablation study for individual components

Methods	SNR (dB)	PSNR (dB)	MSE	LSD	NC	Embedding rate (bps)	Capacity value (bps)
Audio as cover data and text as secret data							
ACM	31.795	49.581	1.39E-06	1.584	0.8791	0.0187	0.00011
SACM	32.394	50.798	1.35E-06	1.648	0.8935	0.0156	0.00015
QIM	40.198	56.208	1.49E-07	0.998	0.9975	0.0190	0.00019
ASSQIM	34.058	53.086	1.33E-05	1.579	0.9958	0.0167	0.00005
ACM-QIM	34.123	56.499	1.31E-04	1.635	0.9369	0.0038	0.0004
ACM-ASSQIM	35.397	57.362	1.28E-06	0.968	0.9657	0.0187	0.00019
SACM-QIM	38.589	58.059	1.27E-06	1.254	0.9789	0.0196	0.00019
SACM-ASSQIM	41.978	59.061	1.24E-07	0.944	0.998	0.0200	0.0002
Audio as cover data and image as secret data							
ACM	29.485	41.452	3.98E-05	4.567	0.8978	1.2568	0.0158
SACM	30.079	43.289	4.36E-05	4.168	0.9235	1.3546	0.0189
QIM	35.201	52.423	4.96E-07	3.705	0.9989	2.1143	0.0229
ASSQIM	31.487	45.078	3.89E-07	3.967	0.9785	1.8964	0.0028
ACM-QIM	32.498	46.389	3.84E-06	4.065	0.9634	2.1455	0.0048

DOI: 10.46121/pspc.53.4.18

ACM-ASSQIM	34.596	47.158	3.95E-06	4.125	0.9578	2.3790	0.0155
SACM-QIM	35.489	49.825	3.65E-06	4.659	0.9237	2.1548	0.0205
SACM-ASSQIM	37.415	54.498	3.55E-07	3.829	0.9999	2.4774	0.0248
Audio as cover data and audio as secret data							
ACM	49.367	67.457	8.57E-09	0.058	0.8946	78	0.75
SACM	51.897	69.485	8.04E-10	0.057	0.9458	79	0.79
QIM	52.763	71.885	9.94E-10	0.032	0.9937	80	0.80
ASSQIM	55.697	70.486	10.58E-09	0.158	0.8792	84	0.82
ACM-QIM	56.489	71.246	11.87E-09	0.069	0.9875	86	0.86
ACM-ASSQIM	60.487	75.681	8.78E-09	0.439	0.9754	94	0.89
SACM-QIM	61.796	79.126	9.78E-08	0.056	0.9836	95	0.92
SACM-ASSQIM	78.153	94.09	7.6741e-11	0.023	1	100	1

6.2 Comparative Analysis

Tables 10, 11, and 12 illustrate a comparative analysis of the existing methods utilizing the TIMIT and GTZAN datasets. Existing methods such as LPA [35], LSB-BMSE [37], LSB PWLCM [38], and novel audio steganography with adaptive threshold [43] are compared with the proposed SACM. The proposed SACM-ASSQIM obtains high PSNR of 29.4726 dB and 98.3547 dB because of its ability to effectively generate secret data in the cover image. The TIMID dataset represents a diverse set of speech samples, making it appropriate for determining the robustness of audio steganography techniques across various accents and speakers. SACM randomizes the embedded data position by evenly spreading distortions, which prevents visible artifacts in the cover image and renders it less perceptible. Moreover, the SACM iterative process enables precise control embedding and reduces interference with the original data.

Table 9. Comparative analysis of average results for different attacks using the TIMIT dataset

Methods	Attacks	PSNR (dB)	MSE	NC
LPA [35]	No attack	47.9287	0.000019	1.0000
	High pass filter (50 Hz)	26.8257	0.000028	0.6291
	Low pass filter (4 kHz)	26.1576	0.000030	0.6082
Proposed SACM-ASSQIM	No attack	52.3975	0.000012	1.0000
	Low pass filter (4kHz)	29.6475	0.000019	0.7548
	High pass filter (50 Hz)	29.4726	0.000025	0.6598

Table 10. Comparative analysis of existing methods using TIMID dataset

Methods	Average MSE	Average SNR
Novel audio steganography with adaptive threshold [43]	1.41×10^{-9}	53.73
Proposed SACM- ASSQIM	1.12×10^{-9}	59.12

Table 11. Comparative analysis of existing methods using GTZAN dataset

Methods	PSNR (dB)	MSE	SNR (dB)
LSB-BMSE [37]	N/A	0.01995	99.98
LSB PWLCM [38]	94.5671	1.40089E-09	90.6455
Proposed SACM-ASSQIM	98.3547	0.00487	99.99

6.3 Research Implications

Enhancing security in audio steganography through the use of SACM-ASSQIM has several important implications. 1) Enhancing security and robustness in audio steganography: By embedding encrypted messages in sensitive frequency regions, the proposed method significantly strengthens resistance against unauthorized access. The integration of encryption and adaptive quantization ensures that even if an attacker extracts hidden data, it remains inaccessible without proper decryption keys. This enhancement improves the security of covert

DOI: 10.46121/pspc.53.4.18

communication and makes it suitable for several applications. 2) Improving imperceptibility and audio quality preservation: Conventional steganographic methods include audio quality that makes hidden messages detectable. The SACM-ASSQIM optimizes message embedding by modifying the DCT coefficients to ensure that audio distortions remain below the human perceptual threshold. The experimental results confirm that the proposed method manages high imperceptibility while embedding more information than existing methods, making it ideal for secure communication. 3) Increasing resistance and steganalysis attacks: Numerous existing audio steganography methods are vulnerable to steganalysis, in which adversaries analyze statistical patterns to detect hidden messages. However, the proposed method leverages encrypted and adaptive quantization to address such attacks and provides covert communication. In summary, this research has significant implications for enhancing the security, robustness, and imperceptibility of audio steganography, and contributes to broader efforts in secure data transmission across diverse fields.

DISCUSSION

The proposed SACM-ASSQIM significantly enhances audio steganography by combining the encryption and adaptive embedding processes. The results show that SACM enhances security by creating an unpredictable transformation, whereas ASSQIM optimizes the imperceptibility through dynamic quantization. Compared to existing methods, such as LPA and LSB-BMSE, the proposed SACM-ASSQIM achieves superior robustness against attacks while maintaining high audio quality. The SACM-ASSQIM achieves a better performance of 29.4726 dB PSNR in high-pass filter attacks and 98.3547 dB using the TIMIT and GTZAN datasets compared to existing methods owing to its adaptability and robustness. The capacity of the proposed SACM-ASSQIM is 22,050 bits/s. Because SACM effectively distributes secret data across the audio cover in a randomized manner, it provides hidden embedded data and enhances security. This randomization prevents attackers from extracting hidden data easily. Moreover, SACM presents resistance to attacks by managing the integrity of hidden data. Based on the frequency analysis, ASSQIM adjusts the step size dynamically, especially distinguishing between sensitive and non-sensitive regions. A lower step size utilizing sensitive frequency bands (2000–5000 Hz) reduces distortions and handles imperceptibility. Furthermore, larger step sizes are used in non-sensitive regions, which increases robustness against attacks. Therefore, the combination of robustness, imperceptibility, and security makes SACM-ASSQIM effective for audio steganography.

CONCLUSION

This research proposes SACM-ASSQIM an advanced audio steganography technique, to hide the text in cover audio with the help of an encryption method. SACM is used to encrypt data by making the data more complex, which makes it harder to identify and enhance security. The nature of the chaotic map ensures that small changes in input lead to significant changes in the output, which strengthens the encryption process. The SACM includes additional shuffling and pixel rotation, which leads to enhanced robustness in the encrypted data. This process ensures that even minor changes in the input lead to significant transformations in the output that provide high resistance to attacks. ASSQIM is performed based on the frequency component distribution in the host audio signal, which allows the step size to be dynamically adjusted to decrease distortions while enhancing robustness. When compared to the existing methods such as LPA and LSB-BMSE, the proposed SACM-ASSQIM obtains a high PSNR of 29.4726 dB PSNR in a high-pass filter attack and 98.3547 dB using TIMIT and GTZAN datasets, respectively. In the future, hybrid encryption methods such as Advanced Encryption Standard (AES) or homomorphic encryption will be considered to enhance security and improve real-world applicability in secure audio communication and digital watermarking.

Author Contributions
N. Shyla: Methodology; Software; Conceptualization; Validation; Resources; Write-up of the original draft.
Kalimuthu Krishnan: Project administration; Data curation; Resources; Formal analysis; Investigation; Review & editing of the original draft.

All authors have read and approved the final manuscript

Declarations
Funding: This research received no external funding.
Conflict of Interest: The authors declare that they have no conflict of interest.

DOI: 10.46121/pspc.53.4.18

Ethics Approval: I/We declare that the work submitted for publication is original, previously unpublished in English or any other language(s), and is not under consideration for publication elsewhere.

Consent to participate / Informed Consent: Not Applicable.

Consent for publication: I certify that all the authors have approved the paper for release and are in agreement with its content.

Data Availability: The datasets generated during and/or analysed during the current study are available in the TIMID and GTZAN datasets [https://paperswithcode.com/dataset/timit] and [https://paperswithcode.com/dataset/gtzan].

REFERENCES

1. P. Zhuo, D. Yan, K. Ying, R. Wang, L. Dong, Audio steganography cover enhancement via reinforcement learning, SIViP 18 (2024) 1007–1013. https://doi.org/10.1007/s11760-023-02819-1.
2. X. Li, L. Chen, J. Lai, Z. Fu, S. Liu, GAN-based image steganography by exploiting transform domain knowledge with deep networks, Multimedia Systems 30 (2024) 224. https://doi.org/10.1007/s00530-024-01427-4.
3. X. Zhang, C. Li, L. Tian, Advanced audio coding steganography algorithm with distortion minimization model based on audio beat, Computers and Electrical Engineering 106 (2023) 108580. https://doi.org/10.1016/j.compeleceng.2023.108580
4. G. Suresh, G. Bhuvaneswari, G. Manikandan, P. Shanthakumar, Chronological bald eagle optimization based deep learning for image watermarking, Expert Systems with Applications 238 (2024) 121545. https://doi.org/10.1016/j.eswa.2023.121545
5. J. Cai, F. Xiao, K. Zhang, X. Gao, Adaptive region assisted GAN for image steganography, Multimedia Systems 31 (2025) 203. https://doi.org/10.1007/s00530-025-01785-7.
6. D.T. Firdaus, N.J. De La Croix, T. Ahmad, AudioSecure: An open-source code to secure data using interpolation and multi-layering techniques within audio covers, Software Impacts 22 (2024) 100707. https://doi.org/10.1016/j.simpa.2024.100707.
7. J. Peng, Y. Liao, S. Tang, Audio steganalysis using multi-scale feature fusion-based attention neural network, IET Communications 19 (2025) e12806. https://doi.org/10.1049/cmu2.12806.
8. W. Su, J. Ni, X. Hu, B. Li, Efficient Audio Steganography Using Generalized Audio Intrinsic Energy With Micro-Amplitude Modification Suppression, IEEE Trans.Inform.Forensic Secur. 19 (2024) 6559–6572. https://doi.org/10.1109/TIFS.2024.3417268
9. H.A. Rehman, U.I. Bajwa, R.H. Raza, S. Alfarhood, M. Safran, F. Zhang, Leveraging coverless image steganography to hide secret information by generating anime characters using GAN, Expert Systems with Applications 248 (2024) 123420. https://doi.org/10.1016/j.eswa.2024.123420
10. Y. Samudra, T. Ahmad, Segmentation embedding method with modified interpolation for increasing the capacity of adaptable and reversible audio data hiding, Journal of King Saud University - Computer and Information Sciences 35 (2023) 101636. https://doi.org/10.1016/j.jksuci.2023.101636

DOI: 10.46121/pspc.53.4.18

11. S. Norouzi Larki, M. Mosleh, M. Kheyrandish, Quantum Audio Steganalysis Based on Quantum Fourier Transform and Deutsch–Jozsa Algorithm, Circuits Syst Signal Process 42 (2023) 2235–2258. https://doi.org/10.1007/s00034-022-02208-y.
12. I.K. Bhat, F. Qadir, M. Neshat, A.H. Gandomi, Exploring Cellular Automata Learning: An Innovative Approach for Secure and Imperceptible Digital Image Watermarking, IEEE Access 12 (2024) 159748–159774. https://doi.org/10.1109/ACCESS.2024.3428362
13. E. Gul, A.N. Toprak, Contourlet and discrete cosine transform based quality guaranteed robust image watermarking method using artificial bee colony algorithm, Expert Systems with Applications 212 (2023) 118730. https://doi.org/10.1016/j.eswa.2022.118730.
14. W. Li, S. Wu, B. Li, W. Tang, X. Zhang, Payload-Independent Direct Cost Learning for Image Steganography, IEEE Trans. Circuits Syst. Video Technol. 34 (2024) 1970–1975. https://doi.org/10.1109/TCSVT.2023.3294291
15. R. Huang, C. Lian, Z. Dai, Z. Li, Z. Ma, A Novel Hybrid Image Synthesis-Mapping Framework for Steganography Without Embedding, IEEE Access 11 (2023) 113176–113188. https://doi.org/10.1109/ACCESS.2023.3324050
16. M.A. Nasr, W. El-Shafai, E.-S.M. El-Rabaie, A.S. El-Fishawy, H.M. El-Hoseny, F.E. Abd El-Samie, N. Abdel-Salam, A robust audio steganography technique based on image encryption using different chaotic maps, Sci Rep 14 (2024) 22054. https://doi.org/10.1038/s41598-024-70940-3.
17. A.A. Krishnan, Y. Ramesh, U. Urs, M. Arakeri, Audio-in-Image Steganography Using Analysis and Resynthesis Sound Spectrograph, IEEE Access 13 (2025) 75184–75193. https://doi.org/10.1109/ACCESS.2025.3563781.
18. Zhang, S., Tian, B., Gao, Y., Dai, M. and Yang, W., 2025. BirdsSong: A stylized generative audio steganography. Computers and Electrical Engineering, 123, p.110112.
19. M.A. Hameed, M. Hassaballah, T. Qiao, IS-DGM: an improved steganography method based on a deep generative model and hyper logistic map encryption via social media networks, Multimedia Systems 30 (2024) 129. https://doi.org/10.1007/s00530-024-01332-w.
20. D.T. Firdaus, N.J.D.L. Croix, T. Ahmad, D. Mukanyiligira, L. Sibomana, Steganographic model to conceal the secret data in audio files utilizing a fourfold paradigm: Interpolation, multi-layering, optimized sample space, and smoothing, Journal of Safety Science and Resilience 6 (2025) 138–149. https://doi.org/10.1016/j.jnlssr.2024.09.004.
21. B. Chen, Y. Nie, J. Yang, Toward high imperceptibility deep JPEG steganography based on sparse adversarial attack, Journal of Visual Communication and Image Representation 97 (2023) 103977. https://doi.org/10.1016/j.jvcir.2023.103977.
22. X. Li, X. Li, Y. Zhao, H. Cho, Dual-domain joint optimization for universal JPEG steganography, Journal of Visual Communication and Image Representation 101 (2024) 104151. https://doi.org/10.1016/j.jvcir.2024.104151.
23. K. Manjunath, G.N.K. Ramaiah, M.N.G. Prasad, Optimal secure XOR encryption with dynamic key for effective audio steganography, Int J Speech Technol 26 (2023) 589–598. https://doi.org/10.1007/s10772-021-09945-6.
24. S. Paul, D. Mishra, Hiding images within audio using deep generative model, Multimed Tools Appl 82 (2023) 5049–5072. https://doi.org/10.1007/s11042-022-13034-4.

DOI: 10.46121/pspc.53.4.18

25. J. Luo, P. He, J. Liu, H. Wang, C. Wu, S. Zhou, Reversible adversarial steganography for security enhancement, Journal of Visual Communication and Image Representation 97 (2023) 103935. https://doi.org/10.1016/j.jvcir.2023.103935.
26. G.V. Kiran, K. Vidhya, Novel multi-media steganography model using meta-heuristic and deep learning assisted adaptive lifting wavelet transform, Journal of Statistical Computation and Simulation 93 (2023) 3126–3155. https://doi.org/10.1080/00949655.2023.2218522.
27. M.A. Nasr, W. El-Shafai, N. Abdel-Salam, E.-S.M. El-Rabaie, A.S. El-Fishawy, F.E.A. El-Samie, Efficient information hiding in medical optical images based on piecewise linear chaotic maps, J Opt 52 (2023) 1852–1866. https://doi.org/10.1007/s12596-023-01128-7.
28. A. Martín, A. Hernández, M. Alazab, J. Jung, D. Camacho, Evolving Generative Adversarial Networks to improve image steganography, Expert Systems with Applications 222 (2023) 119841. https://doi.org/10.1016/j.eswa.2023.119841.
29. L. Yu, S. Weng, M. Chen, Y. Wei, RCDD: Contrastive domain discrepancy with reliable steganalysis labeling for cover source mismatch, Expert Systems with Applications 237 (2024) 121543. https://doi.org/10.1016/j.eswa.2023.121543.
30. F. Guo, S. Sun, S. Weng, L. Yu, J. He, A two-stream-network based steganalysis network: TSNet, Expert Systems with Applications 255 (2024) 124796. https://doi.org/10.1016/j.eswa.2024.124796.
31. Y. Shen, C. Tang, Z. Fan, T. Wu, Z. Lei, Blind watermarking scheme for medical and non-medical images copyright protection using the QZ algorithm, Expert Systems with Applications 241 (2024) 122547. https://doi.org/10.1016/j.eswa.2023.122547.
32. M. Alanzy, R. Alomrani, B. Alqarni, S. Almutairi, Image Steganography Using LSB and Hybrid Encryption Algorithms, Applied Sciences 13 (2023) 11771. https://doi.org/10.3390/app132111771.
33. L. Chen, R. Wang, D. Yan, J. Wang, Learning to Generate Steganographic Cover for Audio Steganography Using GAN, IEEE Access 9 (2021) 88098–88107. https://doi.org/10.1109/ACCESS.2021.3090445.
34. L. Chen, R. Wang, L. Dong, D. Yan, Imperceptible adversarial audio steganography based on psychoacoustic model, Multimed Tools Appl 82 (2023) 26451–26463. https://doi.org/10.1007/s11042-023-14772-9.
35. R. Kumar, J. Yadav, Vowel and non-vowel frame segmentation based digital speech watermarking technique using LPA method, Journal of Information Security and Applications 68 (2022) 103218. https://doi.org/10.1016/j.jisa.2022.103218.
36. K.P. Kumar, A. Kanhe, An Adaptive Embedding Approach for High Imperceptible and Robust Audio Watermarking Using Framelet Transform and SVD, Circuits Syst Signal Process 42 (2023) 5684–5713. https://doi.org/10.1007/s00034-023-02382-7.
37. M.M. Mahmoud, H.T. Elshoush, Enhancing LSB Using Binary Message Size Encoding for High Capacity, Transparent and Secure Audio Steganography–An Innovative Approach, IEEE Access 10 (2022) 29954–29971. https://doi.org/10.1109/ACCESS.2022.3155146.
38. H.T. Elshoush, M.M. Mahmoud, Ameliorating LSB Using Piecewise Linear Chaotic Map and One-Time Pad for Superlative Capacity, Imperceptibility and Secure Audio Steganography, IEEE Access 11 (2023) 33354–33380. https://doi.org/10.1109/ACCESS.2023.3259902.

DOI: 10.46121/pspc.53.4.18

39. Y. Tang, L. Sun, X. Xu, SilentTrig: An imperceptible backdoor attack against speaker identification with hidden triggers, Pattern Recognition Letters 177 (2024) 103–109. https://doi.org/10.1016/j.patrec.2023.12.002.
40. K. Ying, R. Wang, Y. Lin, D. Yan, Adaptive Audio Steganography Based on Improved Syndrome-Trellis Codes, IEEE Access 9 (2021) 11705–11715. https://doi.org/10.1109/ACCESS.2021.3050004
41. H. Kheddar, D. Megías, High capacity speech steganography for the G723.1 coder based on quantised line spectral pairs interpolation and CNN auto-encoding, Appl Intell 52 (2022) 9441–9459. https://doi.org/10.1007/s10489-021-02938-7.
42. B. Laskar, M. Bouzid, Enhancing secure communication: a QIM-based steganography approach for G.722.2 speech streams with Stable Roommate Index Division, Multimed Tools Appl 84 (2024) 13295–13313. https://doi.org/10.1007/s11042-024-19496-y.
43. J. Wang, K. Wang, A novel audio steganography based on the segmentation of the foreground and background of audio, Computers and Electrical Engineering 123 (2025) 110026. https://doi.org/10.1016/j.compeleceng.2024.110026.
44. TIMID dataset link: https://paperswithcode.com/dataset/timit (Accessed on 08 November 2024)
45. GTZAN dataset link: https://paperswithcode.com/dataset/gtzan (Accessed on 08 November 2024)
46. R.T. Leon, P.C. Sherrell, A. Šutka, A.V. Ellis, Decoupling piezoelectric and triboelectric signals from PENGs using the fast fourier transform, Nano Energy 110 (2023) 108445. https://doi.org/10.1016/j.nanoen.2023.108445
47. M. Turan, E. Gökçay, H. Tora, An unrestricted Arnold's cat map transformation, Multimed Tools Appl 83 (2024) 70921–70935. https://doi.org/10.1007/s11042-024-18411-9.
48. W. Alomoush, O.A. Khashan, A. Alrosan, H.H. Attar, A. Almomani, F. Alhosban, S.N. Makhadmeh, Digital image watermarking using discrete cosine transformation based linear modulation, J Cloud Comp 12 (2023) 96. https://doi.org/10.1186/s13677-023-00468-w.

DOI: 10.46121/pspc.53.4.18