MJoSHT
Malaysian Journal of Science, Health & Technology
[mjosht.usim.edu.my]

USIM

Article

Identification of Key Parameters for Cover Audio Generation in Imperceptible Audio Steganography

Muhammad Harith Noor Azam[1,2], Farida Ridzuan[1,2], A H Azni[1,2], Sakinah Ali Pitchay[1,2] and Nur Hafiza Zakaria[1,2]

[1]*Faculty of Science and Technology, Universiti Sains Islam Malaysia, Nilai 71800, Negeri Sembilan, Malaysia.*

[2]*CyberSecurity and Systems Research Unit, Faculty of Science and Technology, Universiti Sains Islam Malaysia, Nilai 71800, Negeri Sembilan, Malaysia.*

Correspondence should be addressed to:

Muhammad Harith Noor Azam; m.harith@usim.edu.my

Article Info
Article history:
Received:5 September 2025
Accepted: 12 February 2026
Published:15 Mac 2026

Academic Editor:
Nurdiana Azizan

Malaysian Journal of Science, Health & Technology

MJoSHT2025, Volume 11, Special Issue on the 5th International Conference on Recent Advancements in Science and Technology (ICoRAST 2025): Responsible Artificial Intelligence – Advancing Science and Technology for Humanity
eISSN: 2601-0003

https://doi.org/10.33102/mjosht.540

***Abstract*— The security of data during communication is an ever-growing challenge, as outdated algorithms face increasing vulnerability. Audio steganography offers a promising alternative by embedding data within audio signals in a manner that is imperceptible to human listeners. Recent advancements in cover selection as well as cover generation techniques have improved the quality and suitability of audio for embedding, yet the parameters influencing imperceptibility remain underexplored. This research aims to identify and evaluate key audio parameters which are energy, the total and value distribution of silent samples. These parameters can be assessed prior to the embedding process which offer a potential advantage over existing metrics, typically require post-embedding analysis to accurately assess imperceptibility. The methodology of this research consists of five main stages: parameter identification, dataset preparation, parameter extraction, embedding, and evaluation. Signal-to-Noise Ratio (SNR) was employed as the primary metric to assess perceptual transparency, with 20% data capacity embedded in each audio of total 220 audio. The first experiment revealed a positive correlation as high as r = 0.706 between energy and SNR, indicating that higher-energy audio benefit from auditory masking, resulting in improved imperceptibility. The second experiment showed a negative correlation, r = –0.787 between total silent sample and SNR, confirming that embedding in silent sample increases the risk of perceptible artifacts. Silent sample that has higher value exhibited superior SNR values compared to those with extremely lower value. These findings highlight the importance of pre-embedding analysis in identifying suitable samples for embedding. By identifying and evaluating these influential parameters, this research contributes to the development of intelligent cover generation strategies that enhance the transparency of audio steganography systems. The insights presented offer a foundation for future design, optimization, and real-world application of robust data-hiding techniques.**

***Keywords*— Audio Steganography, Cover Generation, Audio Parameters**

I. INTRODUCTION

Audio steganography is a branch of information hiding that involves concealing secret data within audio signals in such a way that the presence of the hidden information remains imperceptible to unintended listeners [1]. This technique leverages the characteristics of the Human Auditory System (HAS), which is less sensitive to minor alterations in audio signals, thereby enabling covert communication between the sender and the intended recipient [2]. Common audio formats used in steganographic applications include uncompressed waveform audio files (.wav) [3] and compressed formats such as MPEG-1 Audio Layer III (.mp3) [4], each offering distinct advantages in terms of data capacity and processing complexity.

The effectiveness of any audio steganography method is typically evaluated based on four fundamental characteristics: capacity, imperceptibility, robustness and security. Capacity refers to the amount of secret data that can be embedded within the cover audio [5]. Imperceptibility ensures that the embedded data does not produce audible distortions detectable by human listeners [6]. Robustness measures the resilience of the hidden data against various attacks aimed at corrupting or removing the secret message from the stego-file [7]. Lastly, security represents the ability to prevent a secret message from being illegally accessed or a stego-file from being detected [8]. These characteristics are interdependent, and three primary trade-offs are commonly observed among them: 1) capacity-imperceptibility, 2) imperceptibility-robustness, and 3) robustness-capacity [7]. Achieving an optimal balance among these characteristics is challenging, as improvements in one often lead to compromises in the others.

Audio steganography can be implemented either in the time domain or in the transform domain [9], [10]. Several existing audio steganography techniques commonly implemented under these domains are LSB modification, phase coding, echo hiding, parity coding, spread spectrum, and wavelet domain [11]. All these techniques have different approaches to improving certain aspects of their characteristics.

To enhance the core characteristics of steganography, researchers generally adopt two main approaches. The first approach involves improving the embedding strategy by proposing new techniques such as variations of phase coding [12], [13] or Least Significant Bit (LSB) [14]. This includes the use of cost functions to adaptively determine optimal embedding locations based on the properties of the cover audio.

The second approach shifts the task of cover selection from manual to automated systems. This can involve selecting an optimal cover from a database tailored to the embedding strategy [15] or generating a cover from scratch based on specific parameters aligned with the secret message [16]. These approaches may be implemented independently or in combination, depending on the research objectives.

Each approach offers distinct advantages. The first strengthens the technical foundation of steganography by refining embedding mechanisms, while the second reduces human error by ensuring optimal cover selection. Effective implementation of either or both approaches is reflected in improved imperceptibility, robustness, capacity and security of the steganographic system.

While numerous parameters have been proposed in previous research to support cover generation or cover selection, most of these efforts have been concentrated within the domain of image steganography [17]. In contrast, research involving audio steganography has largely relied on parameters derived after the embedding process, such as Signal to Noise Ratio (SNR) and sample difference before and after embedding [3]. This post-embedding dependency introduces inefficiencies, particularly when evaluating multiple embedding configurations. Therefore, there is a clear need to identify and utilise audio-specific parameters that can be assessed prior to the embedding process. Such an approach would enable early-stage evaluation of cover suitability, reduce computational complexity, and improve the overall efficiency of audio steganography systems.

The remainder of this paper is outlined as follows. Section II presents the literature review. Section III describes the research methodology. Section IV discusses the evaluation results and findings. Section V concludes the paper and highlights directions for future research.

II. LITERATURE REVIEW

Audio steganography is a technique used to conceal a secret message within a cover audio, such that the presence of the hidden data remains undetectable to unintended recipients. The standard audio steganography model, illustrated in Figure 1 [17], represents the conventional approach where the user manually selects a cover audio file and provides the secret message. The system then embeds the message into the chosen cover and produces the stego-file. This model is widely adopted due to its simplicity and direct workflow. However, it places the burden of cover selection on the user, which may lead to suboptimal choices that compromise its characteristics, such as either one or more [18].

To address these limitations, the cover selection model introduces an automated mechanism for identifying the most appropriate cover file from a predefined database [19]. As shown in Figure 2 [17], the system begins by requesting the secret message from the user. It then evaluates available cover files using specific comparison criteria, such as maximum embedding size, as well as SNR, to select the optimal candidate for embedding. One real-world application of cover selection is selecting a set of cover files from the fewest possible files. Since extensive calculations are needed to determine each audio's capacity to ensure that the minimum number of images can conceal the maximum size of the secret message, a cover selection process is designed to speed up the calculation process, hence providing optimised selection in a short amount of time instead of relying on the manual calculation by the human [20].

Expanding further, the cover generation model, depicted in Figure 3 [17], eliminates the need for a pre-existing cover database altogether. Instead, the system dynamically generates a cover audio file tailored to the characteristics of the secret message and the embedding strategy. After receiving the message from the user, the system synthesises a suitable cover, embeds the message, and produces the stego-file. This model significantly reduces user involvement and enhances adaptability, making it particularly advantageous for real-time applications or environments where preparing a large cover database is impractical. This approach improves the overall performance of the steganographic process by generating covers that meet optimal characteristics.

Together, these models reflect the evolving strategies in audio steganography from manual workflows to intelligent automation, highlighting the importance of cover selection and generation in achieving secure and imperceptible communication.

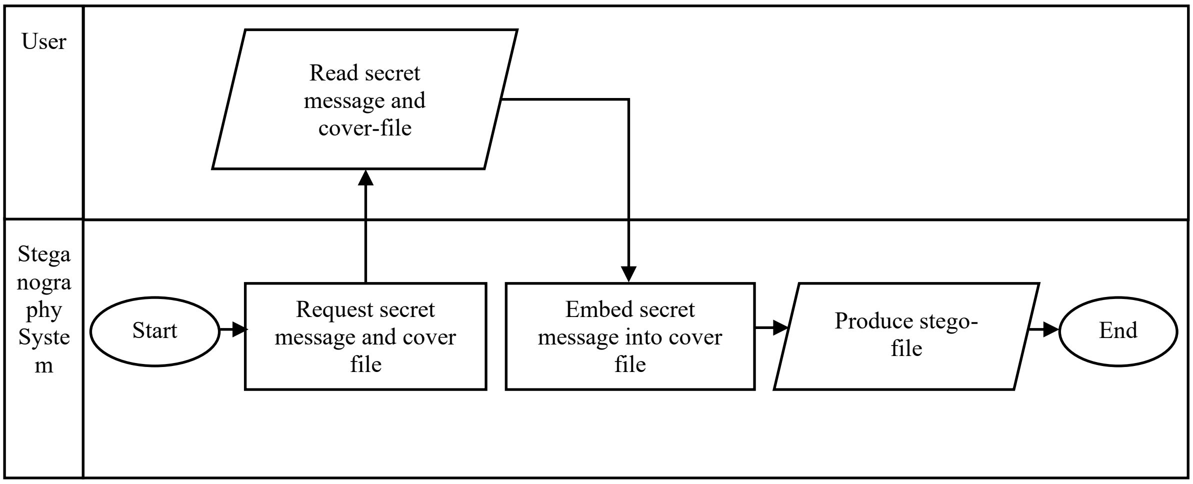
User
Read secret message and cover-file
Steganography System
Start
Request secret message and cover file
Embed secret message into cover file
Produce stego-file
End

Figure 1. The state-of-the-art steganography model in the form of an activity diagram [17].

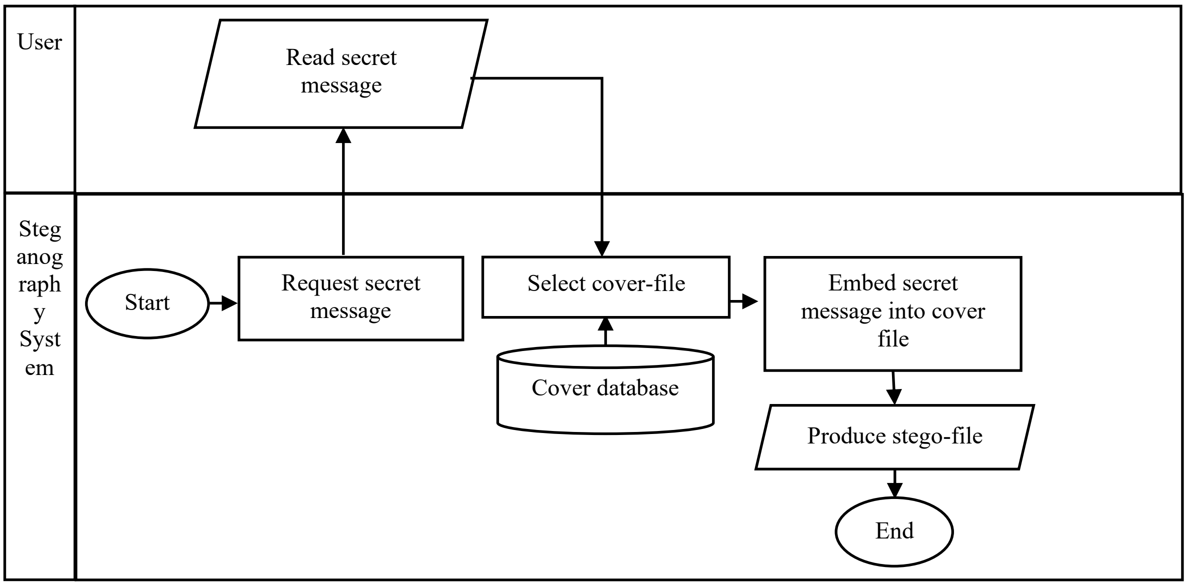

Figure 1. The steganography model involving the cover selection method in the form of an activity diagram [17].

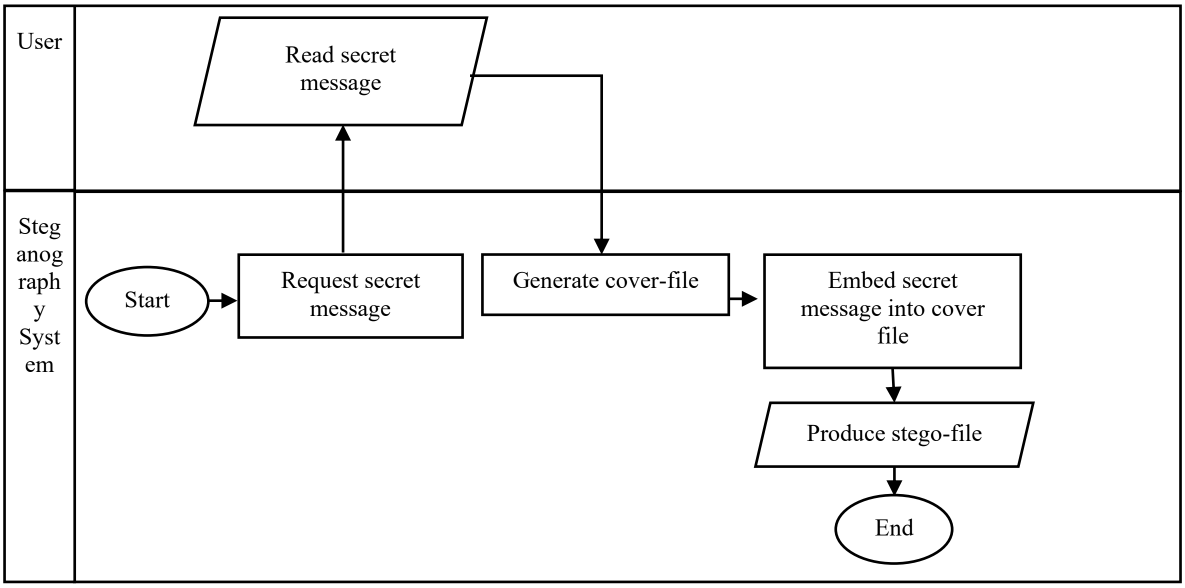

Figure 3. The steganography model involving the cover generation method in the form of an activity diagram [17].

A. Parameters Relevant to Audio Cover and Audio Generation

The effectiveness of audio steganography is highly dependent on the characteristics of the cover audio used for embedding. Whether the cover is selected from an existing database or generated dynamically, several key parameters must be considered to ensure optimal performance in terms of imperceptibility, robustness, embedding capacity and security.

1) Capacity:

Several key parameters influence the capacity characteristics, which are sampling rate, bit depth, duration of the audio file, and channel configuration.

Firstly, the sampling rate plays a crucial role. Higher sampling rates, such as 44.1 kHz, yield more data points per second, thereby enabling finer granularity in the embedding process. This increased resolution can potentially expand the available space for hidden information [21].

Similarly, bit depth is a significant factor; greater bit depths, such as 32 bits, allow for a more precise representation of audio samples. This precision can be leveraged to embed data within the least significant bits, minimising perceptual distortion [22].

The duration of the audio file also directly affects embedding capacity. Longer durations naturally provide more audio samples based on the sample rate for embedding [21].

Channel configuration further influences the embedding potential. Stereo or multi-channel audio formats offer additional pathways for data distribution compared to mono audio [22].

Collectively, these parameters determine the payload capacity of the cover audio. They are possibly evaluated prior to embedding.

2) Robustness and Security:

Security in audio steganography is often closely linked to robustness, as both characteristics are influenced by similar underlying parameters. Key factors such as the choice of audio format, sample entropy, and amplitude-related parameters, including stability and scaling, play a significant role in enhancing the robustness and, consequently, the security of the steganographic system.

The audio format is a critical determinant of robustness [23]. Uncompressed formats such as WAV are generally preferred due to their high fidelity and predictable sample structure. These formats preserve the integrity of embedded data even after basic signal processing, making them suitable for techniques that require precise bit-level manipulation, such as Least Significant Bit (LSB) embedding combined with error correction codes. In contrast, compressed formats like MP3 utilise perceptual coding and quantisation, which can introduce artifacts and distortions that may compromise the embedded message.

Entropy, which measures the randomness or unpredictability of audio samples, also plays a vital role [24]. Low-entropy regions are more stable and less susceptible to distortion during operations such as compression or filtering, making them ideal for embedding error-sensitive data. Embedding in these regions increases the likelihood that the hidden message will remain intact after processing. Additionally, entropy analysis can guide the placement of redundant message copies in stable zones, improving the chances of successful recovery even if parts of the audio are corrupted. Conversely, high entropy enhances security by increasing randomness, which is beneficial in resisting steganalysis attacks that attempt to detect patterns in the embedded data.

Amplitude-related parameters carry significant weight in determining both robustness and security. Amplitude stability refers to the consistency of an audio signal's energy distribution over time. Regions with high amplitude stability where the signal maintains a relatively constant energy level are less affected by dynamic range compression or gain adjustments. These regions are ideal for embedding redundant copies of the hidden message, as they help preserve data integrity during playback or transmission. In contrast, low amplitude stability increases vulnerability to distortion, reducing the reliability of the embedded message.

In the context of amplitude scaling, it has been demonstrated that scaling factors can remain invariant under amplitude scaling attacks when properly adjusted [25]. This implies that even if the audio signal is amplified or attenuated, the embedded data can still be reliably extracted. Therefore, selecting and tuning appropriate scaling factors enhances the robustness of the steganographic system against such transformations. Collectively, these parameters determine the security and the robustness of the cover audio. They are possibly evaluated prior to embedding.

3) *Imperceptibility:*

Several key parameters influence the imperceptibility characteristics, which are SNR, Peak Signal to Noise Ratio (PSNR) and Czernikowski Distance.

One of the most widely used metrics is the SNR [3], which quantifies the ratio between the original audio signal and the noise introduced during the embedding process. A higher SNR value indicates that the embedded data has minimal impact on the audio quality, thereby enhancing the transparency of the stego-file. SNR serves as a general indicator of the system's ability to conceal data without compromising auditory fidelity. It can be calculated using Equation 1.

$$SNR = 10 * log_{10} \frac{\sum_1^n x(i)^2}{\sum_{i=1}^n (x(i)-y(i))^2} \qquad (1)$$

where x and y the original cover audio and stego-file sample respectively, while i denote samples index, while n is denoted as the total audio sample.

Complementing SNR is the PSNR [26], which focuses on the maximum deviation between the original and stego-file signals. PSNR is particularly effective in identifying localised distortions that may not significantly affect the overall SNR but could still be perceptible to listeners. Higher PSNR values are typically associated with better imperceptibility and are useful in fine-grained evaluations of embedding effects. It can be calculated using Equation 2.

$$PSNR = 10 * log_{10} \frac{R^2}{MSE} \qquad (2)$$

where R is the peak signal value in the original cover audio and MSE is defined in Equation 3.

$$MSE = \frac{1}{n} \sum_1^n (x(i) - y(i))^2 \qquad (3)$$

where x, y, n and i carry the same definition as those in the Equation 1.

In addition to these mathematical metrics, perceptual models such as the Czernikowski Distance offer a more nuanced assessment of imperceptibility [27]. This metric evaluates the dissimilarity between the original and stego-file signals based on psychoacoustic principles, incorporating aspects of human auditory perception. Unlike SNR and PSNR, which rely solely on sample values differences, Czernikowski Distance accounts for perceptual masking and sensitivity, making it a more accurate indicator of auditory transparency. Lower values of this metric suggest a higher degree of similarity and, consequently, better imperceptibility.

Although these parameters are commonly used in the selection or generation of cover audio, they can only be accurately measured after the embedding process has taken place. This limitation may delay the audio selection or generation process and introduce additional computational complexity into the steganographic method. Therefore, this research aims to identify and utilise audio-related parameters those associated with imperceptibility that can be evaluated prior to the embedding process. By doing so, the system can streamline the cover audio generation or selection phase.

III. RESEARCH METHODOLOGY

This research modifies and extends a structured methodology [26] to investigate imperceptibility in audio steganography by analysing audio-related parameters that can be measured prior to the embedding process. The methodology process flow is shown in Figure 4.

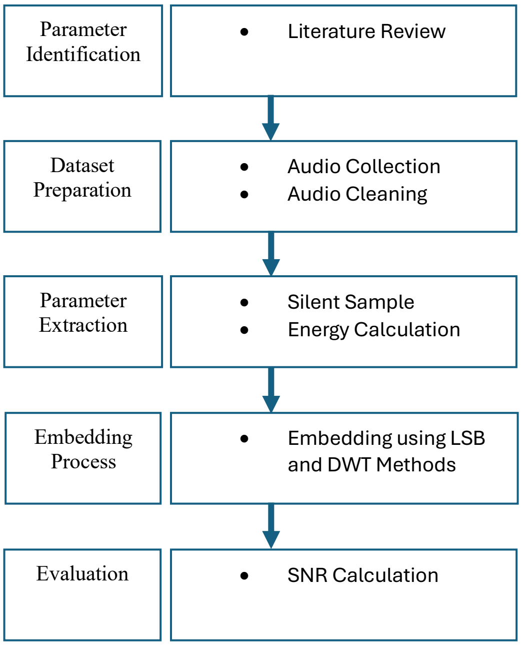

Figure 4. Research methodology.

The methodology consists of five main stages: parameter identification, dataset preparation, parameter extraction, embedding, and evaluation. Additional details for each step are provided in the subsection below:

A. *Parameter Identification*

The initial stage of this research involves identifying key audio parameters that influence imperceptibility in steganographic embedding. This was accomplished through an extensive literature review focusing on cover generation and selection techniques within the broader steganography domain.

Due to the limited availability of research specifically addressing audio steganography in this context, the review was extended to include literature from image and video steganography.

In image steganography, energy is frequently used as a metric to assess imperceptibility, with higher-energy samples offering greater masking potential for hidden data [28]. Drawing from this concept, energy was selected as a candidate parameter for audio steganography, where it similarly reflects the intensity of the signal and its potential to conceal embedded modifications through auditory masking.

Additionally, the review identified a second parameter relevant to audio which is silent samples. These are audio samples characterized by minimal or near-zero values, indicating low sound. Silent samples are analogous to low-texture regions in images [29], which are more sensitive to alterations. In audio, these low-value samples are perceptually vulnerable, and modifications within them are more likely to be detected by listeners. Therefore, analysing the total count and distribution of silent samples provides insight into regions where embedding may compromise imperceptibility.

By identifying energy, total and distribution of silent sample as key parameters, this research establishes a foundation for evaluating and optimizing cover audio selection and generation. These parameters are further extracted and analysed in subsequent stages to assess their impact on the perceptual transparency of stego-file.

B. Dataset Preparation

The audio dataset used in this research was sourced from Freesound.org, a collaborative database of Creative Commons licensed audio samples. All selected audio files were downloaded in WAV format, mono channel, and 16-bit resolution to ensure consistency and compatibility with the embedding technique. The dataset includes both speech and music samples to ensure generalisability across different audio types. The audio samples were then cleaned and trimmed to a uniform duration of exactly one second. This preprocessing step ensures that all samples are standardised for parameter analysis and embedding.

C. Parameters Extraction

Two key parameters were extracted from each audio sample to support imperceptibility analysis in audio steganography, which are total silent audio sample and signal energy. These parameters were selected based on their relevance to perceptual transparency and their ability to be measured prior to the embedding process.

The silent audio sample count refers to the number of samples within the values range of 0 to 4095. These samples represent inaudible part of audio. Due to the sensitivity of the human auditory system, even minor modifications in these segments can produce perceptible artifacts such as hissing or distortion. Therefore, analysing the presence and value distribution of silent samples is essential for identifying regions that may pose a higher risk of detection when used for embedding.

In parallel, the energy of each audio sample was calculated to assess its dynamic characteristics. Sample energy reflects the overall strength or intensity of the audio waveform and is defined mathematically as in Equation 4.

$$E = \sum_{1}^{n} x(i)^2 \qquad (4)$$

where$x$, n and i carry the same definition as those represented in Equation 1.

Higher energy values indicate louder or more active regions in the audio, which are generally more suitable for embedding due to the auditory masking effect, which is a phenomenon where stronger sounds mask weaker ones, making small modifications less perceptible. Thus, embedding in high-energy regions tends to improve imperceptibility, as changes are less likely to be detected by human listeners.

D. Embedding Process

The data embedding process was conducted using two primary techniques: the Least Significant Bit (LSB) method and the Discrete Wavelet Transform (DWT).

The LSB method is a spatial (or time) domain technique that operates by directly substituting the least significant bit of each audio sample with a bit from the secret message. Because this modification occurs at the lowest level of binary representation within the sample, the original amplitude of the audio signal is altered only by a negligible margin, preserving the integrity of the carrier signal. For this study, the LSB algorithm was implemented using Python to handle bitwise operations on raw pulse-code modulation (PCM) data.

In contrast, the DWT method is a transform-domain technique that shifts the embedding process from raw samples to frequency coefficients. The audio signal is passed through a series of filters specifically a high-pass and a low-pass filter to decompose the signal into two distinct components which are 1) approximation coefficients that represents the low-frequency and stable parts of the signal, and 2) detail coefficients that represents the high-frequency and transient parts. The secret data is embedded into these coefficients rather than the time-domain samples. After the data is integrated into the chosen sub-bands, an Inverse Discrete Wavelet Transform (IDWT) is applied to reconstruct the stego-file back into the time domain. This dual-layer approach allows for data hiding within the mathematical structure of the frequency spectrum.

Both methodologies were implemented in a Python environment to ensure a consistent framework for the evaluation of their specific embedding mechanics.

E. Evaluation

The SNR was employed as the primary objective metric to evaluate the imperceptibility of the stego-file. SNR measures the ratio between the original audio signal and the noise introduced during the embedding process. A higher SNR value indicates better transparency and lower perceptual distortion, thereby validating the effectiveness of the selected audio parameters and the embedding strategy.

To assess the impact of pre-embedding audio characteristics on imperceptibility, two experiments were conducted using SNR as the evaluation metric. The first experiment examined the impact of audio energy on imperceptibility, analysing how variations in signal energy influence the perceptual quality of the stego-file. The second experiment focused on the impact of a silent audio sample count, investigating how the presence of low-value samples affects the detectability of embedded data. These evaluations offer empirical observations about how different audio parameters, measured prior to embedding, contribute to the overall transparency of the steganographic system.

IV. EVALUATION AND DISCUSSION

This section presents and discusses the results obtained from the evaluation of imperceptibility in audio steganography using the SNR as the primary metric. The evaluation focuses on two key audio parameters, which are 1) energy and 2) silent audio sample. These parameters were analysed to determine their influence on the perceptual transparency of the stego-file. 20% of the capacity was embedded into the audio in both experiments.

A. Impact of Energy on Imperceptibility

The objective of this experiment was to investigate the relationship between the energy level of an audio signal and the imperceptibility of embedded data. Analysis was performed on a total of 220 audio samples, comprising both music and speech types. For the LSB embedding method, host signal energy values ranged from approximately 3.68×10^{8} to 1.22×10^{13} while the corresponding SNR values ranged from 49.23 dB to 94.46 dB. The mean SNR across all samples was 79.35 dB, with a standard deviation of 9.58. The scatter plot in Figure 5 illustrates the relationship between energy and SNR for both LSB and DWT techniques.

The analysis reveals a strong positive correlation between energy and SNR for both methods, with Pearson correlation coefficients of 0.706 for LSB and 0.696 for DWT. This suggests that audio samples with higher energy consistently produce higher SNR values, indicating superior objective imperceptibility. This trend strongly supports the hypothesis that louder or more dynamic audio segments provide a better 'masking' environment, effectively concealing the modifications introduced during the embedding process.

However, a comparison between methods shows that LSB maintains a significantly higher average SNR (79.35 dB) compared to DWT (47.11 dB), even though both benefit similarly from high-energy host signals. It is understandable due to the nature of embedding itself where LSB directly manipulates the least significant bits of audio sample while DWT alters frequency coefficients that, when inverted, can result in wider changes to the audio sample.

Some outliers remain present specifically, very low-energy samples in the DWT method yielded SNR values as low as 17.04 dB, whereas the LSB method stayed above 49 dB. These variations highlight that while energy is a primary predictor of quality, the choice of embedding algorithm significantly determines the baseline transparency.

In conclusion, the results demonstrate that energy is a critical predictor of imperceptibility in audio steganography. By analyzing energy levels prior to embedding, researchers can identify high-energy masking regions that minimize the noise-to-signal ratio, thereby enhancing the overall transparency and security of the stego-file.

B. Impact of Silent Sample on Imperceptibility

The objective of this experiment was to investigate the relationship between the quantity of silent samples and the imperceptibility of the stego-file. Similar with first experiment, analysis was performed on a total of 220 audio samples, comprising both music and speech types. Silent samples are defined as those with values ranging from 0 to 4095. The total count of silent samples in each audio file was calculated and compared against the corresponding SNR values.

The dataset includes samples with total silent counts ranging from 6,002 to 44,100. The mean silent sample count was approximately 34,633, with a standard deviation of 9,926.8. For the LSB method, the SNR values ranged from 49.23 dB to 94.46 dB, with a mean of 79.35 dB with standard deviation of 9.58. Figure 6 illustrates the relationship between total silent samples and SNR for both LSB and DWT techniques.

The analysis reveals a strong negative correlation between the total silent sample count and SNR, with Pearson correlation coefficients of –0.770 for LSB and –0.787 for DWT. This indicates that audio samples with a higher density of silent samples consistently produce lower SNR values, confirming a reduction in objective imperceptibility. These findings align with psychoacoustic principles, as the human auditory system is more sensitive to modifications in quiet regions. When embedding occurs in low amplitude (silent samples), the noise introduced is proportionally much larger relative to the local signal power, leading to a more significant degradation of the SNR compared to embedding in high-energy regions. Furthermore, the data shows that as the number of silent samples approaches the total frame size (44,100 samples), the SNR reaches its lowest points for both methods. This suggests that the distribution of sample values is a critical factor; audio files dominated by extremely low-value samples provide poor masking for steganographic data. In contrast, files with fewer silent samples (closer to the 6,000 range) maintain significantly higher transparency, with SNR values frequently exceeding 85 dB in the LSB method.

In conclusion, the results confirm that the total count of silent samples is a highly reliable predictor of imperceptibility. A high concentration of silent samples significantly increases the noise-to-signal ratio, thereby degrading the quality of the stego-file. These insights suggest that pre-embedding analysis of silent sample density can be used to steer data toward more 'active' audio regions, effectively maximizing the perceptual transparency and objective quality of the stego-file.

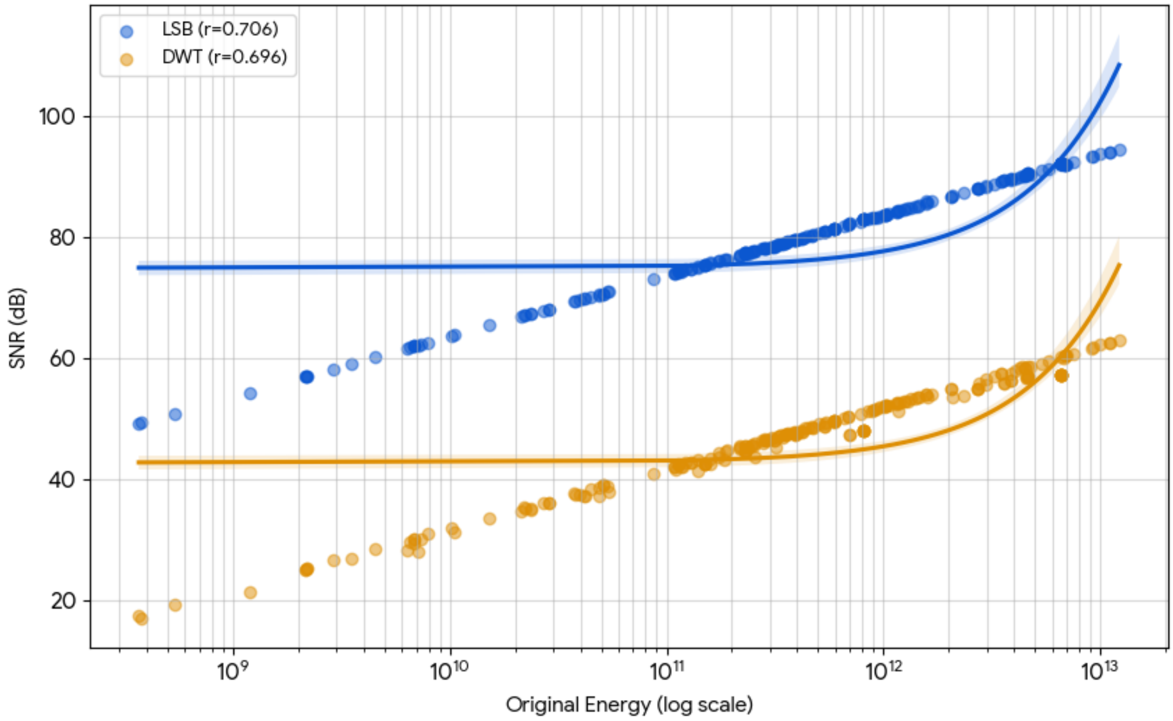

Figure 5. Relationship between energy and SNR (LSB vs DWT).

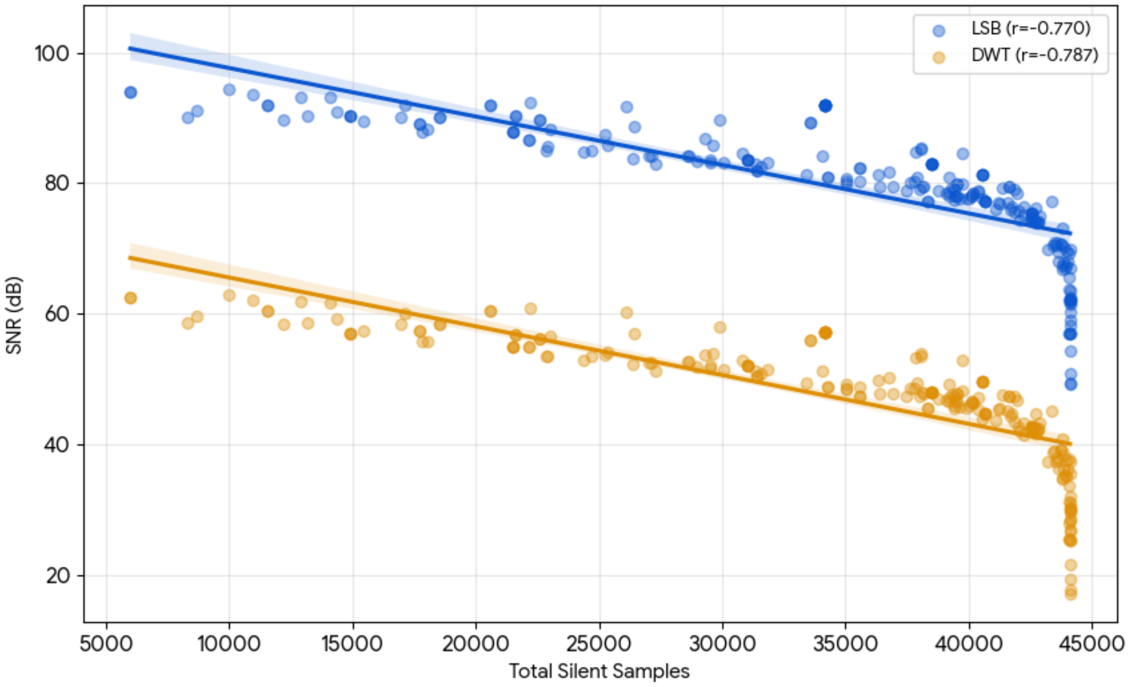

Figure 6. Scatter plot relationship between silent sample and SNR.

V. CONCLUSIONS

This research has presented a systematic investigation into audio-related parameters that influence imperceptibility in audio steganography, with a particular focus on parameters that can be measured prior to the embedding process. Through two key experiments, it shows that imperceptibility parameters can be evaluated prior to the embedding process.

The first experiment revealed a positive correlation between audio energy and SNR, indicating that higher-energy signals tend to mask embedded data more effectively, thereby enhancing imperceptibility. The second experiment indicated that the total and value distribution of silent samples play a critical role in perceptual quality. Specifically, embedding in extremely low amplitude samples increases the risk of perceptible artifacts, while targeting higher-values samples within the silent range improves transparency.

These findings support the development of pre-embedding analysis techniques that can guide the selection of suitable cover audio, reducing computational complexity and improving the overall efficiency of the steganographic process.

Future work will focus on the development of an intelligent cover audio generation framework. This system will synthesise or select audio content based on optimised energy levels and silent sample distributions to ensure high imperceptibility and robustness. By integrating these pre-evaluated parameters into the generation process, the framework aims to automate and enhance the security of audio steganography applications, particularly in dynamic or real-time environments.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

ACKNOWLEDGMENT

This research was funded by the Ministry of Higher Education (MOHE) Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2024/ICT02/USIM/02/1). The authors would like to express their gratitude to Universiti Sains Islam Malaysia (USIM) and MOHE for the support and facilities provided.

REFERENCE

[1] B. E. Carvajal-Gámez, M. A. Castillo-Martínez, L. A. Castañeda-Briones, F. J. Gallegos-Funes, and M. A. Díaz-Casco, "Audio Steganalysis Estimation with the Goertzel Algorithm," Applied Sciences (Switzerland), vol. 14, no. 14, Jul. 2024, doi: 10.3390/app14146000.

[2] J. Jezdimirovic, N. Pekez, and J. Kovacevic, "Security enhancement of LSB-based audio steganography method," 2023 IEEE Zooming Innovation in Consumer Technologies Conference, ZINC 2023, pp. 77–82, 2023, doi: 10.1109/ZINC58345.2023.10174020.

[3] M. H. Noor Azam, F. Ridzuan, and M. N. S. Mohd 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, doi: https://doi.org/10.32890/jict2023.22.2.5

[4] Y. Ren, D. Liu, C. Liu, Q. Xiong, J. Fu, and L. Wang, "A Universal Audio Steganalysis Scheme Based on Multiscale Spectrograms and DeepResNet," IEEE Trans. Dependable Secure Comput., vol. 20, no. 1, pp. 665–679, 2023, doi: 10.1109/TDSC.2022.3141121.

[5] D. R. I. M. Setiadi, "Improved payload capacity in LSB image steganography uses dilated hybrid edge detection," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 2, pp. 104–114, 2022, doi: 10.1016/j.jksuci.2019.12.007.

[6] S. Rustad, D. R. I. M. Setiadi, A. Syukur, and P. N. Andono, "Inverted LSB image steganography using adaptive pattern to improve imperceptibility," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6, pp. 3559–3568, 2022, doi: 10.1016/j.jksuci.2020.12.017.

[7] M. H. Noor Azam, F. Ridzuan, M. N. S. Mohd Sayuti, A. H. Azni, S. Ali Pitchay, and N. H. Mohd Alwi, "A Method of Cover Audio Selection for Embedding Based on Various Criteria," in ITM Web of Conferences, vol. 63, 2024, p. 01001. doi: 10.1051/itmconf/20246301001.

[8] M. H. N. Azam, F. Ridzuan, M. Norazizi Sham Mohd Sayuti, A. H. , Azni, N. H. , Zakaria, and B. Harjito, "Improving Dynamic Security of the Least Significant Bit using Block-based Chaotic Multi-Level LSB (BCM-LSB)," in International Conference on Cyber Security, Privacy in Communication Networks, Springer Nature Singapore, 2024.

[9] M. A. Nasr et al., "A robust audio steganography technique based on image encryption using different chaotic maps," Sci. Rep., vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-70940-3.

[10] J. Wang and K. Wang, "A novel audio steganography based on the segmentation of the foreground and background of audio," Computers and Electrical Engineering, vol. 123, Apr. 2025, doi: 10.1016/j.compeleceng.2024.110026.

[11] F. ASLANTAŞ and C. HANİLÇİ, "Comparative Analysis Of Audio Steganography Methods," Journal of Innovative Science and Engineering (JISE), Jan. 2022, doi: 10.38088/jise.932549.

[12] A. A. Alsabhany, "The Adaptive Multi-Level Phase Coding Method in Audio Steganography for Confidential Communication," Universiti Sains Islam Malaysia, 2019.

[13] 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, Jan. 2024, doi: 10.59324/ejtas.2024.2(1).55.

[14] M. M. Mahmoud and H. T. Elshoush, "Enhancing LSB Using Binary Message Size Encoding for High Capacity, Transparent and Secure Audio Steganography-An Innovative Approach," IEEE Access, vol. 10, pp. 29954–29971, 2022, doi: 10.1109/ACCESS.2022.3155146.

[15] S. Nazari, "Cover Selection Steganography Via Run Length Matrix and Human Visual System," Journal of Information Systems and Telecommunication, vol. 1, no. 2, 2013.

[16] Y. Hu, Z. Yang, H. Cao, and Y. Huang, "Multi-modal Steganography Based on Semantic Relevancy," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Melbourne, VIC, Australia: Springer, 2021, pp. 3–14. doi: 10.1007/978-3-030-69449-4_1.

[17] M. H. Noor Azam, F. Ridzuan, M. N. S. Mohd Sayuti, A. H. Azni, N. H. Zakaria, and V. Potdar, "A systematic review on cover selection methods for steganography: Trend analysis, novel classification and analysis of the elements," May 01, 2025, Elsevier Ireland Ltd. doi: 10.1016/j.cosrev.2025.100726.

[18] P. D. Shah and R. S. Bichkar, "Genetic Algorithm Based Approach to Select Suitable Cover Image for Image Steganography," in 2020 International Conference for Emerging Technology, INCET 2020, 2020, pp. 1–5. doi: 10.1109/INCET49848.2020.9154032.

[19] V. Hajduk and D. Levický, "Cover Selection Steganography with Intra-Image Scanning," 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA), pp. 1–4, 2018.

[20] Z. Wang, X. Zhang, and Z. Yin, "Joint Cover-Selection and Payload-Allocation by Steganographic Distortion Optimization," IEEE Signal Process. Lett., vol. 25, no. 10, pp. 1530–1534, 2018, doi: 10.1109/LSP.2018.2865888.

[21] M. H. Noor Azam, "Enhancement of Cover Selection-based Audio Steganography (CAS) using Block-based Chaotic Multi-level LSB (BCM-LSB) for Balanced Performance," Universiti Sains Islam Malaysia, 2023.

[22] B. J. Bhatkalkar, R. V. Arjunan, A. Alok, R. Sanghavi, D. Cenitta, and R. Kamath, "A Novel Method for Sample Selection in Audio Stenographic Systems," in 6th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 664–672. doi: 10.1109/ICECA55336.2022.10009492.

[23] X. Dong, M. F. Bodo, and Z. Ignjcrtovic, "Robustness Analysis of Digital Audio Steganographic Method based on Phase Manipulation," in Proceedings 7th International Conference on Signal Processing, IEEE, 2004, pp. 2375–2378.

[24] H. Miao, L. Huang, Y. Shen, X. Lu, and Z. Chen, "Steganalysis of compressed speech based on markov and entropy," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2014, pp. 63–76. doi: 10.1007/978-3-662-43886-2_5.

[25] S. T. Chen, T. W. Huang, and C. T. Yang, "High-SNR steganography for digital audio signal in the wavelet domain," Multimed. Tools Appl.,

vol. 80, no. 6, pp. 9597–9614, Mar. 2021, doi: 10.1007/s11042-020-09980-6.

[26] M. H. N. Azam, F. Ridzuan, and M. N. S. M. Sayuti, “A New Method to Estimate Peak Signal to Noise Ratio for Least Significant Bit Modification Audio Steganography,” Pertanika J. Sci. Technol., vol. 30, no. 1, pp. 497–511, 2022, doi: 10.47836/pjst.30.1.27.

[27] I. Avcibaş, “Audio Steganalysis With Content-Independent Distortion Measures,” IEEE Signal Process. Lett., vol. 13, no. 2, pp. 92–95, 2006, doi: 10.1109/LSP.2005.862152.

[28] R. Shmueli, D. Mishra, T. Shmueli, and O. Hadar, “A novel technique for image steganography based on maximum energy seam,” Multimed. Tools Appl., vol. 83, no. 28, pp. 70907–70920, Aug. 2024, doi: 10.1007/s11042-024-18476-6.

[29] [T. Wu, X. Hu, and C. Liu, “Security-oriented steganographic payload allocation for multi-remote sensing images,” Sci. Rep., vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-55474-y.