Performance Analysis of DWT-Based Echo Hiding Audio Steganography under Imperceptibility and Robustness Evaluation

Jeffrey Bram Pattipeilohy[1], Mardi Hardjianto[2]*, Blasius Dala Nai[3]
Master of Computer Science [1], [2],[3]
Budi Luhur University
Jakarta, Indonesia
2211600370@student.budiluhur.ac.id[1], mardi.hardjianto@budiluhur.ac.id[2],
2311600825@student.budiluhur.ac.id[3]

Abstract—Audio steganography conceals information within audio signals while preserving message confidentiality and minimizing perceptual distortion. However, balancing audio quality, extraction accuracy, and robustness remains challenging, particularly in conventional time-domain approaches. This study proposes a hybrid audio steganography method that integrates Echo Hiding with the Discrete Wavelet Transform (DWT) by embedding secret information into the detail coefficients (cD) of the transformed audio signal. The novelty of this work lies in implementing Echo Hiding in DWT detail coefficients and systematically analyzing embedding parameters to evaluate the trade-off between imperceptibility, extraction accuracy, and robustness under selected signal-processing conditions. Experiments were conducted using six WAV audio files and two secret messages. Performance was evaluated using Signal-to-Noise Ratio (SNR), Bit Error Rate (BER), and Mean Opinion Score (MOS). The selected parameters, X0 = 60, X1 = 260, and α = 0.7, achieved error-free extraction under clean conditions, with BER = 0 and SNR values ranging from 15.96 dB to 49.92 dB. Subjective evaluation by 10 respondents produced an average MOS of 4.8. Robustness evaluation showed low BER under mild Gaussian noise, while MP3 compression at 128 kbps produced average BER values of 0.0561 and 0.0477 for Message 1 and Message 2, respectively. These findings indicate a practical balance between audio quality and extraction accuracy under the evaluated conditions.

Keywords*: *Steganography, Echo Hiding, Discrete Wavelet Transform, Bit Error Rate, Signal-to-Noise Ratio

I. INTRODUCTION

In the digital age, the exchange of sensitive and confidential information has become an integral part of global communication. Data transmission over the internet necessitates a high degree of confidentiality and security [1], [2]. Security efforts also include protecting data from unauthorized access by irresponsible parties [3], [4]. Cryptography and information concealing are the two types of digital information security solutions. In order to guarantee information security, these technologies are necessary for data storage and transmission. [5]. Cryptography is the key to various security systems, protecting the confidentiality of sensitive information [6]. The encryption prepare makes data garbled to anybody without the suitable decoding key, guaranteeing that as it were authorized people can get to the data. The advent of the computer age has made cryptography an important tool for data protection in digital technology.

Besides cryptography, another important technique in security systems is information hiding. It works by embedding information into a medium. In recent years, data covering up has gotten to be progressively critical for secure capacity and transmission. One of the most common methods of hiding digital information is steganography. Steganography is an alternative approach that aims to hide the existence of a message to avoid raising suspicion among third parties [7]. Media that can be utilized in steganography are pictures, sound and video. Image-based steganography is the most widely used medium because of its ease of implementation [8]. However, audio media offers a higher level of imperceptibility, despite its high complexity. Digital audio media also has considerable redundancy [9]. This makes audio media an ideal cover object for inserting confidential data [10].

There are several methods developed for steganography, such as Least Significant Bit (LSB), Phase Coding, Discrete Wavelet Transform (DWT), Spread Spectrum, Discrete Cosine Transform (DCT), and combinations of several methods. The most widely used method is LSB because it offers high capacity and a simple implementation [11], [12]. However, the LSB method is not resistant to signal-processing attacks, such as compression, noise addition and filtering [13].

One technique known for its high undetectability and resilience is Echo Hiding [14]. This method inserts data by adding artificial echoes into the audio signal. The data is encoded through variations in parameters such as echo delay, amplitude, and decay rate [15]. Although the echo is masked

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

below the threshold of human hearing and is relatively difficult to detect, it still has limitations in terms of resilience to time-domain attacks and signal processing [16].

On the other hand, the Discrete Wavelet Transform (DWT) is frequently utilized in signal processing due to its ability to represent signals in both the time and frequency domains simultaneously [17]. DWT allows the separation of audio signals into several frequency subbands, enabling the data embedding process to be carried out on components that are more stable and less sensitive to interference [18]. Various studies have shown that transform-based steganography, especially DWT, has better resilience than time-domain methods [19], [20], [21].

Recent studies have explored various audio steganography techniques, including LSB, Echo Hiding, and transform-domain approaches. These studies have mainly focused on improving embedding capacity, imperceptibility, message recovery, or robustness against signal-processing attacks. A comparison of representative studies is presented in TABLE I.

TABLE I. COMPARISON OF AUDIO STEGANOGRAPHY METHODS IN PREVIOUS STUDIES

Reference	Method	Main Contribution	Remaining Issue
Imelda, Hardjianto	M-Bit LSB	Increased embedding capacity up to four times while maintaining acceptable audio quality for $m \leq 4$	Higher embedding levels ($m > 4$) introduce audible noise despite SNR remaining above 30 dB
Carpentieri et al.	Echo Hiding	Improved Echo Hiding implementation using frame skipping, stereo embedding, and AES encryption while maintaining imperceptible echoes	Further improvement in echo distribution is still required
Rafiee & Fakhredanesh	Echo Hiding + Spread Spectrum	Improved message recovery and security through modified cepstral autocorrelation and Spread Spectrum integration	Embedding capacity can still be further improved while maintaining recovery performance
Lahiri	Echo Hiding + Wavelet	Applied Echo Hiding in the wavelet domain using a pseudorandom sequence	Limited analysis of embedding parameters and wavelet coefficient selection

As shown in TABLE I, previous studies have addressed different aspects of audio steganography, including embedding capacity, imperceptibility, message recovery, and security improvement. Although LSB-based techniques have a high embedding capacity and are easy to deploy, they are typically susceptible to signal-processing processes. Echo Hiding provides better imperceptibility by embedding information in artificial echo patterns, but the reliability of its extraction may be affected by changes in the audio signal. Meanwhile, DWT-based approaches provide time–frequency localization and enable embedding in selected frequency subbands, but previous studies have not sufficiently examined how echo parameters affect extraction accuracy and audio quality when Echo Hiding is applied directly to DWT detail coefficients.

Based on these limitations, the research gap addressed in this study is the lack of systematic analysis of Echo Hiding parameters in the DWT detail-coefficient domain, particularly regarding the trade-off among imperceptibility, extraction accuracy, and robustness under selected signal-processing conditions. Therefore, this study proposes a DWT-based Echo Hiding audio steganography method in which secret information is embedded into the detail coefficients (cD) of the transformed audio signal. The proposed approach is designed to evaluate the effects of delay parameters and echo amplitude on Bit Error Rate (BER), Signal-to-Noise Ratio (SNR), and message recovery performance.

The main contribution of this study is the imple mentation and performance evaluation of Echo Hiding in DWT detail coefficients, supported by systematic parameter analysis using X0, X1, and α. This study also evaluates the proposed method under clean conditions, a baseline method comparison, a Gaussian noise attack, and an MP3 compression attack. The findings are expected to provide insight into the practical balance between imperceptibility, extraction accuracy, and robustness in DWT-based audio steganography systems.

II. METHODOLOGY

This research aims to assess the effectiveness of an audio steganography technique that combines the Discrete Wavelet Transform (DWT) and Echo Hiding. This combination is designed to embed secret messages into digital audio signals while maintaining perceptual audio quality and supporting reliable message extraction under the evaluated conditions. The research steps used to illustrate the general sequence of activities are shown in Fig. 1.

A. Audio Dataset and Secret Messages

The dataset used in this research consists of audio as a cover medium and a secret message to be inserted. The audio used is in .wav format with a mono channel configuration, 16-bit resolution, and a frame rate of 44100 Hz. The Nyquist-Shannon sampling theorem states that to reproduce a frequency accurately, the sampling rate must be at least twice that frequency. The human ear's ability to capture sound frequencies is from 20 Hz to 20 kHz [22], so the most suitable sample rate is 44.1 kHz. The 16-bit bit depth was chosen because audio CDs originally employed 16 bits, which translates to a dynamic range (dB) of 96 dB. A dynamic range of 96 dB is more than sufficient to achieve good recording quality. The .wav file is uncompressed and provides direct

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

access to amplitude values for signal processing using the Discrete Wavelet Transform (DWT) and Echo Hiding methods. Meanwhile, the secret message is a text (string) converted to binary using a coding standard such as ASCII, then inserted into the audio signal.

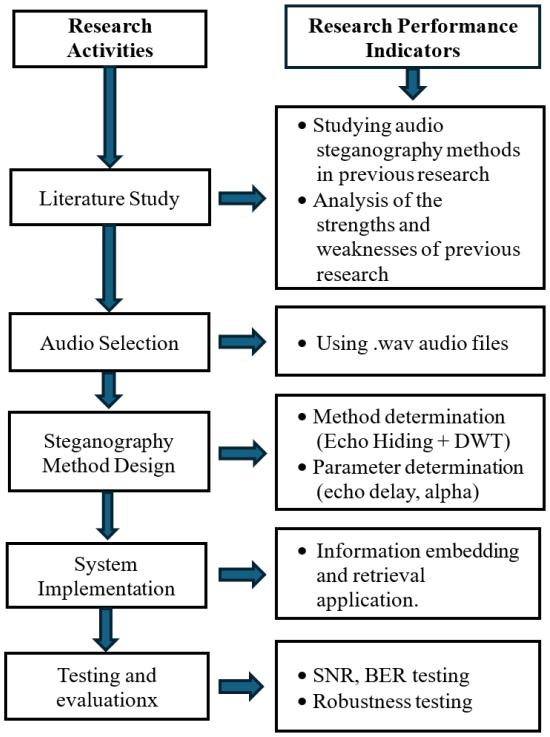

Fig. 1. Research Steps

B. Echo Hiding and DWT Algorithm

The message embedding methods chosen in this research are Echo Hiding and DWT. The DWT method was chosen as the initial stage because of its ability to represent audio signals in both the time and frequency domains. The transform-domain approach is often used for embedding secret messages because it can help maintain imperceptibility and improve resistance to certain signal-processing operations [23].

Furthermore, transform-domain embedding can help maintain the quality of stego audio. Echo Hiding was chosen as the embedding method because it represents secret bits as artificial echo-delay patterns that remain perceptually unobtrusive. The selection of the Echo Hiding and DWT algorithms in this research is based on the need to evaluate the trade-off among audio quality, extraction accuracy, and robustness under the selected testing conditions.

The Daubechies wavelet (db4) was selected for its good balance between time and frequency localization, making it suitable for non-stationary audio signals. The decomposition level was chosen to ensure sufficient separation between frequency subbands while preserving the quality of signal reconstruction. Higher decomposition levels were avoided due to potential loss of temporal resolution and increased reconstruction distortion.

C. Message Embedding Process

In the initial stage, the cover audio file is read and represented as an amplitude series in the time domain. The audio data is normalized into the range -1 to 1. The low-frequency and high-frequency components of the signal are then separated using the Discrete Wavelet Transform (DWT) technique. Approximation coefficients (cA) are used to represent low-frequency components, and detail coefficients (cD) are used to represent high-frequency components. The low-frequency components are represented as approximation coefficients (cA), while the high-frequency components are represented as detail coefficients (cD). The calculation for finding the cA and cD values is shown in (1).

$$cA(k) = \sum_{n=0}^{L-1} h[n] \bullet x[2k-n]$$
$$cD(k) = \sum_{n=0}^{L-1} g[n] \bullet x[2k-n] \quad (1)$$

Where
x[n] = original audio signal (amplitude)
h[n] = low-pass filter
g[n] = high-pass filter
L = filter length (db4 = 8)
k = DWT result index
cA(k) = approximation coefficient
cD(k) = detail coefficient

Secret message embedding is performed on the detail coefficient (cD). The human auditory system, which is less sensitive to little variations in high-frequency components, is the basis for the choice of cD. Small modifications to this component can improve imperceptibility, as high-frequency changes are generally less noticeable to human hearing. The secret message is transformed into binary before being embedded. Each message bit is embedded by adding an echo component of its own with a small amplitude. This embedding process is shown in (2.

$$cD'(k) = cD(k) + \alpha \bullet cD(k-d) \quad (2)$$

Where
cD'(k) = detail coefficient after being embedded at index k.
cD(k) = original detail coefficient at index k
cD(k – d) = echo of detail coefficient
α = echo amplitude factor
k = detail coefficient index
d = delay parameter (d = X_0 for bit 0, d = X_1 for bit 1)

The values of the echo amplitude (α) and delay parameters (X_0 and X_1) were determined through a

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

preliminary sensitivity analysis. A range of parameter values was evaluated to observe the trade-off between imperceptibility, measured by SNR, and extraction accuracy, measured by BER. The selected parameters were then used in the main experiment to achieve a practical balance between audio quality and message extraction reliability. Some cover audio files may contain initial regions with zero-valued detail coefficients. This condition can affect the embedding process because an echo component generated from a zero-valued delayed coefficient will also have a value of zero and will not effectively modify cD'. Therefore, the embedding process is initiated after a sequence of consecutive non-zero detail coefficients is found, ensuring that the embedded echo pattern can be formed and later detected during extraction.

The detail coefficients, once embedded with the message bits, then return the audio signal to the time domain using the Inverse Discrete Wavelet Transform (IDWT). The result of this process is stego audio that is perceptually almost identical to the original audio, but contains hidden information. This stego audio must be extracted to obtain the message by detecting the delay echo pattern in the wavelet domain.

D. Message Extraction Process

The message extraction process aims to recover the secret message embedded in the audio file. Message extraction is carried out using a non-blind approach. Parameters such as the frame length, the delay values for bits 0 and 1, and the message bit length are known. Consequently, the extraction process requires prior knowledge of these parameters, which limits the applicability of the proposed method in fully blind communication scenarios. In the initial stage, the stego audio file is read and converted into a sequence of time-domain amplitude values. The audio data is normalized to the range -1 to 1 to maintain stability in the signal processing. Next, the signal is transformed into the frequency domain using a DWT, producing the approximation coefficient (cA) and the detail coefficients (cD). The embedding process is performed on the detail coefficients, so the extraction process also focuses on the detail coefficients.

The detail coefficients are divided into frames of predetermined length. One bit of the secret message is assumed to be contained in each frame. The beginning of a frame is determined by checking for a sequence of consecutive nonzero detail coefficients. To improve the stability of the spectral analysis and reduce the impact of discontinuities at frame boundaries, each frame is multiplied by a Hann Window. The purpose of this step is to suppress spectral leakage that can interfere with echo detection in the next stage. Next, the cepstrum is calculated for the signal of that frame. Cepstrum analysis is useful for detecting echoes in the signal by identifying peaks in the frequency domain associated with the echo delay used during embedding.

The next step is to find the peak within the specified quefrency range based on the echo delay values representing bit 0 and bit 1. The highest value in the cepstrum area indicates the location of the strongest echo delay. The identified peak location is then compared with two delay values, namely the delay indicating bit 0 and the delay indicating bit 1. If the peak position is closer to the delay bit 0, the frame is classified as bit 0; if it is closer to the delay bit 1, it is classified as bit 1. The classification result is then stored as a sequence of bits. This series of bits is collected and divided into eight-bit groups to form one byte. Each byte is converted into an ASCII character so that the original text message can be obtained.

E. Performance Testing and Evaluation

The research results were evaluated to assess the performance of the proposed steganography method with respect to two main aspects: imperceptibility and robustness. The degree of similarity between the original audio (cover audio) and the audio embedded with a secret message (stego audio) is measured by imperceptibility. Measurements were made using the Signal-to-Noise Ratio (SNR) by comparing the audio signal energy before and after the embedding process. The SNR value was calculated using (3).

$$SNR = 10\ log_{10}\left(\frac{\sum_{i=1}^{N} x(i)^2}{\sum_{i=1}^{N}(x(i)-y(i))^2}\right) \quad (3)$$

Where
SNR = Signal to Noise Ratio
x(i) = original audio signal
y(i) = embedded audio signal
N = total number of samples

Robustness testing is also performed to evaluate how resilient the secret message is to processes or interference that may occur in the stego audio. At this stage, the stego audio is reprocessed to extract the hidden secret message. The extraction results are then compared with the original message to calculate the Bit Error Rate (BER). The BER value is calculated using (4).

$$BER = \frac{N_e}{N_t} \quad (4)$$

Where
BER = Bit Error Rate
Ne = number of bits that are in error during extraction
Nt = total number of embedded message bits

The proposed method mainly consists of DWT/IDWT transformations, echo embedding, and cepstrum-based extraction. The computational cost is dominated by the FFT-based cepstrum computation, resulting in an overall complexity of approximately O(N log N). This indicates that

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

the proposed method remains computationally feasible for the evaluated audio dataset.

Robustness was evaluated by applying selected signal-processing attacks to the stego audio before extraction. In this study, additive white Gaussian noise (AWGN) and MP3 compression at 128 kbps were used as attack scenarios. The extracted message after each attack was compared with the original message using BER. Filtering and resampling attacks were not included in the current evaluation and are considered as future work.

III. Results and Discussion

A. Cover Audio

The cover audio files used in this research consist of six .wav song files. The songs in the cover audio are both vocal and instrumental. The audio data used is shown in Table II.

Table II. Cover Audio

#	File Name (.wav)	File Size (Byte)	Total Sample	Total Detail Coefficient (cD)	Capacity (Character)
1.	In_Flight	15.537.438	7768656	3884331	118
2.	Too_Many_Times	17.766.134	8883001	4441504	135
3.	BeautifulLiar	19.341.148	9670500	4835253	147
4.	Chinese-beat	18.081.836	9040896	4520451	137
5.	Smooth-AC	16.422.956	8211456	4105731	125
6.	Ultimatum	24.016.940	12008448	6004227	183

The total sample is obtained by reading the number of amplitude samples in each audio file. Next, the total detail coefficients (cD) are calculated through decomposition using the DWT at level 1 and are approximately half of the total number of samples. The message capacity is calculated based on the total frames that can be formed from the cD coefficients. Each frame consists of 4096 samples, each representing one bit of the message, so the total number of bits that can be embedded is ⌊total cD / 4096⌋. One character consists of 8 bits, so the capacity in character units is ⌊total cD / 4096⌋ / 8. As the number of cD coefficients increases, the capacity to embed secret messages increases.

B. Message Embedding Process

The message embedding process begins by reading the .wav audio file used as the cover audio. The resulting amplitude data is normalized to the range -1 to 1. A fragment of the normalized amplitude data is shown in Fig. 2.

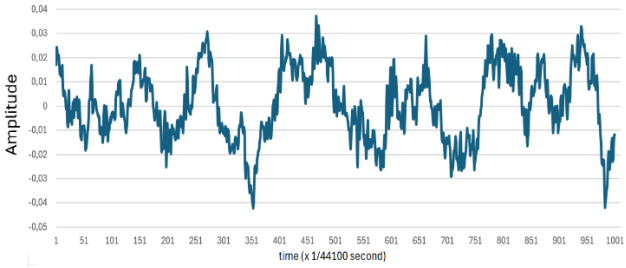

Fig. 2. Normalized .WAV File Amplitude Data Snippet

The cA and cD components are then obtained by applying the Discrete Wavelet Transform (DWT) with the Daubechies 4 (db4) wavelet to the audio signal. The secret message is inserted into the cD component because its sensitivity is lower than that of human hearing. A fragment of the cD component is shown in Fig. *3*.

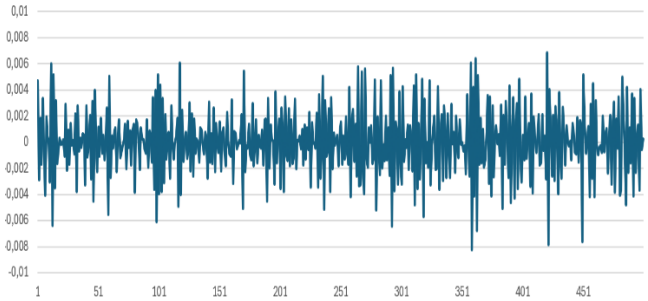

Fig. 3. Detail Coefficient (cD) Fragment

The secret message used is a string. This message is first converted into binary, then embedded into frames of detail coefficients. Each frame represents one bit of the message. The length of each frame is 4096 samples. Based on the results of repeated experiments, the delay values X0 = 60 for bit 0 and X1 = 260 for bit 1 were chosen. The large delay difference produces a clear *quefrency* peak, simplifying the extraction process. The gain parameter value (alpha) used in the embedding process is set at 0.7. This parameter controls the strength of the embedded echo signal. An alpha value of 0.7 was chosen based on experimental results to determine the minimum threshold for successful message extraction. Experiments with alpha <= 0.6 failed to retrieve messages, while alpha >= 0.7 succeeded. One of the secret messages used to embed text is the sentence "The quick brown fox jumps over the lazy dog," in which all the letters are used. After embedding the secret message into the detail coefficient, the shape of the detail coefficient is shown in Fig. *4*. After all messages are inserted into the detail coefficients (cD), the next step is to return the signal to the time domain so that it can be represented as an audio signal.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

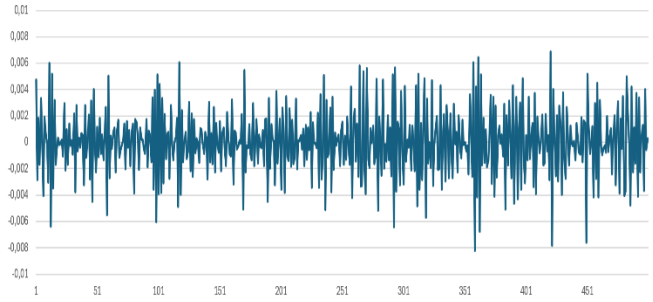

Fig. 4. Detail Coefficient that has embedded secret message

C. Message Extraction Process

The audio signal in .wav format is read and represented as digital amplitude values. Next, a DWT is applied to obtain cA and cD. The message extraction process is carried out by analyzing the embedded audio signal in the detail coefficient domain (cD). Each cD frame is processed with the Hann Window, and its cepstrum is computed to obtain a representation in the quefrency domain. The presence of echo is indicated by a peak at a certain delay. If the peak appears around the X_0 position, it represents bit '0', while if it appears around X_1, it represents bit '1'. Thus, the secret message can be reconstructed from the audio signal. An example of the cepstrum calculation results is shown in Fig. 5. In the test results, the main peak is detected near the X-axis position ≈ 261. This position indicates the presence of an echo component with a specific delay embedded in the signal. This position is closer to X_1 than X_0, so it can be concluded that the frame represents bit '1'.

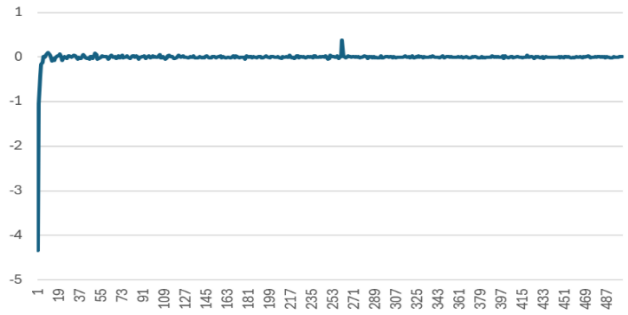

Fig. 5. Cepstrum Calculation Results

D. Performance Evaluation and Robustness Testing

The performance evaluation was conducted to assess imperceptibility, extraction accuracy, and robustness of the proposed method. Imperceptibility was measured using SNR, and extraction accuracy using BER. The evaluation includes clean-condition testing, alpha parameter analysis, baseline comparison, Gaussian noise attack, and MP3 compression attack.

Testing was conducted to evaluate the performance of the developed steganography method. Six WAV audio files (see Table II) were used as cover audio. Two secret messages were used for embedding, as shown in Table III.

TABLE III. MESSAGE FOR EMBEDDING

#	Message
Message 1	The quick brown fox jumps over the lazy dog
Message 2	String with ASCII codes ranging from 0 to as many as the cover audio can hold.

Each message will be embedded in each cover audio file. After embedding, the Signal-to-Noise Ratio (SNR) value is calculated to measure the imperceptibility of the stego audio to the original audio. The results of this calculation are used to analyze the extent to which the secret message embedding affects audio quality and to ensure that the message can be embedded without significantly reducing it. The test findings are displayed in Table IV.

TABLE IV. SNR, BER VALUE OF STEGO AUDIO WITH PARAMETERS X_0=60, X_1=260, ALPHA=0.7

Experiment	Cover Audio	Message 1		Message 2	
		SNR (dB)	BER	SNR (dB)	BER
1.	Cover Audio-1	47.11	0.00	39.69	0.00
2.	Cover Audio-2	32.35	0.00	28.36	0.00
3.	Cover Audio-3	24.24	0.00	19.92	0.00
4.	Cover Audio-4	20.96	0.00	15.96	0.00
5.	Cover Audio-5	49.92	0.00	45.45	0.00
6.	Cover Audio-6	48.73	0.00	42.65	0.00

The SNR values generated by embedding the two messages ranged from 15.96 dB to 49.92 dB. Although the lowest SNR was 15.96 dB, subjective testing using the Mean Opinion Score (MOS) method with 10 respondents yielded an average of 4.8, indicating that the audio quality was still in the good category. This condition indicates that the message embedding did not significantly interfere with the listener's perception. In addition, the insertion of the detail coefficient (cD) and structured echo characteristics caused the signal changes to be perceived as non-disturbing noise.

Testing by embedding message 1 into cover audio-4 and cover audio-5 produces significantly different SNR values. The SNR value for audio cover-4 is 20.96 dB, while for audio cover-5 it is 49.92 dB. This can occur because the SNR is greatly influenced by the characteristics of the audio cover, not just the algorithm. Signals with loud audio, or in other words, large amplitude, will produce high SNR values. Conversely, low-amplitude audio signals will yield low SNR values. Therefore, the SNR value is also influenced by the strength of the signal from the audio cover.

In addition to low amplitudes, small SNR values can also be affected by the α (alpha) parameter. The larger the α value, the stronger the echo produced, thereby increasing the difference between the original and stego signals. This results in a lower SNR. Conversely, small α values tend to produce higher SNR values, as the changes occurring in the signal are relatively smaller. However, increasing the α value will

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

strengthen the echo, thereby increasing the success rate of secret message retrieval.

At $\alpha = 0.7$, all messages can be extracted correctly (BER = 0). This indicates that the echo strength is sufficient to form a clear delay pattern. However, when α is lowered to 0.6, extraction errors begin to occur (BER > 0). This is caused by the reduced echo amplitude, which makes it less dominant over the original signal, especially in the low-signal region. This indicates that decreasing α improves audio quality but reduces the system's robustness. The SNR and BER values with $\alpha = 0.6$ are shown in Table V.

TABLE V. SNR, BER VALUE OF STEGO AUDIO WITH PARAMETERS X_0=60, X_1=260, ALPHA=0.6

Experiment	Cover Audio	Message 1		Message 2	
		SNR (dB)	BER	SNR (dB)	BER
1.	Cover Audio-1	49.34	0.00	39.7	0.00
2.	Cover Audio-2	34.60	0.00	28.4	0.01
3.	Cover Audio-3	26.48	0.00	19.9	0.00
4.	Cover Audio-4	23.22	0.00	16.0	0.00
5.	Cover Audio-5	52.03	0.02	45.5	0.01
6.	Cover Audio-6	50.91	0.00	42.7	0.00

In Cover Audio-5, the BER value for message 1 is 0.02. This indicates a failure to retrieve the embedded message. TABLE VI shows the differences between message 1 and the extracted message. The failure occurs when the space character between the words "the" and "lazy" is retrieved. The resulting character is a null character.

TABLE VI. DIFFERENCE BETWEEN ORIGINAL MESSAGE AND EXTRACTED MESSAGE WITH ALPHA=0.6

Original Message	The quick brown fox jumps over the lazy dog
Extracted Message	The quick brown fox jumps over thelazy dog

To further validate the effectiveness of the proposed method, a comparative evaluation was conducted using conventional audio steganography techniques under the same experimental conditions. The comparison includes Least Significant Bit (LSB), standalone Echo Hiding, DWT-based LSB steganography, and the proposed DWT + Echo Hiding method using the same cover audio files, secret message, and evaluation metrics. The comparison results in terms of Signal-to-Noise Ratio (SNR) and Bit Error Rate (BER) are presented in TABLE VII.

TABLE VII. PERFORMANCE COMPARISON OF AUDIO STEGANOGRAPHY METHODS

Method	Average SNR (dB)	Average BER
LSB	107.6763	0.000000
Echo Hiding	6.8433	0.000706
DWT + LSB	65.2592	0.102843
DWT + Echo Hiding	34.6117	0.000000

As shown in TABLE VII, LSB achieved the highest average SNR and zero BER under clean extraction conditions. However, the proposed DWT + Echo Hiding method also achieved zero BER while maintaining a moderate average SNR. This indicates that the proposed method does not outperform LSB in terms of imperceptibility under non-attack conditions. Nevertheless, the proposed method provides a different balance by combining echo-based embedding with wavelet-domain coefficient modification. Therefore, the contribution of this study should be interpreted as a balanced implementation of Echo Hiding in DWT detail coefficients rather than a universal superiority over all conventional methods.

To further evaluate the robustness of the proposed method, the stego audio generated using the selected embedding parameters ($X_0 = 60$, $X_1 = 260$, and $\alpha = 0.7$) was subjected to additive white Gaussian noise (AWGN). Different values of the noise standard deviation (σ) were used to simulate channel distortion prior to message extraction. The resulting Bit Error Rate (BER) values under each noise condition are presented in TABLE VIII.

TABLE VIII. BER OF THE PROPOSED METHOD UNDER GAUSSIAN NOISE ATTACK

Stego Audio	Standard Deviation (σ)				
	0.00001	0.00005	0.00010	0.00050	0.00100
Stego Audio-1	0.00000	0.01483	0.03072	0.17373	0.26801
Stego Audio-2	0.00318	0.00318	0.01271	0.03920	0.04661
Stego Audio-3	0.00000	0.00000	0.00000	0.01271	0.02966
Stego Audio-4	0.00530	0.00847	0.00845	0.01271	0.02436
Stego Audio-5	0.02966	0.12923	0.25212	0.40466	0.40466
Stego Audio-6	0.00212	0.01377	0.03708	0.24788	0.32309

As shown in TABLE VIII, the BER generally increased as the noise standard deviation (σ) increased, indicating that higher levels of additive white Gaussian noise (AWGN) progressively degraded the embedded echo pattern and reduced the accuracy of cepstrum-based delay detection. At low noise levels ($\sigma \leq 0.0001$), most cover audio files still achieved low BER values, indicating that the proposed method is tolerant of mild channel distortion. However, as the noise intensity increased to $\sigma = 0.0005$ and $\sigma = 0.001$, the BER increased noticeably, indicating that excessive noise interferes with the embedded echo signal and reduces the reliability of message extraction.

To evaluate the robustness of the proposed method against lossy compression, all stego audio files generated from six cover audio files and two secret messages were compressed to MP3 at 128 kbps, then converted back to WAV before extraction. This test was conducted to examine whether the embedded echo pattern in the DWT detail coefficients could still be detected after the compression–decompression process. The BER values obtained under the MP3 compression attack are presented in TABLE IX.

TABLE IX. BER OF THE PROPOSED METHOD UNDER MP3 COMPRESSION ATTACK

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

Cover Audio	BER	
	Message 1	Message 2
Cover Audio-1	0.0731	0.0678
Cover Audio-2	0.0381	0.0350
Cover Audio-3	0.0350	0.0286
Cover Audio-4	0.0403	0.0297
Cover Audio-5	0.1335	0.1186
Cover Audio-6	0.0169	0.0064
Average	**0.0561**	**0.0477**

As shown in TABLE IX, MP3 compression introduced extraction errors in all tested stego audio files. The average BER values were 0.0561 for Message 1 and 0.0477 for Message 2. The lowest BER was observed with Cover Audio-6, whereas Cover Audio-5 produced the highest BER for both messages. These results indicate that the proposed method has partial tolerance against MP3 compression, but it cannot guarantee error-free extraction after lossy compression. The variation in BER values also shows that compression robustness is influenced by the characteristics of the cover audio.

IV. CONCLUSION

This research proposed an audio steganography method that integrates Echo Hiding into the detail coefficients of the Discrete Wavelet Transform (DWT). The proposed method was evaluated using six WAV audio files and two secret messages. The experimental results show that the selected parameters, X0 = 60, X1 = 260, and α = 0.7, achieved error-free message extraction under clean conditions, with BER = 0 for all tested audio files and messages. The obtained SNR values ranged from 15.96 dB to 49.92 dB, while the subjective evaluation produced an average MOS of 4.8, indicating that the stego audio quality remained perceptually acceptable.

The results confirm that parameter selection significantly affects the balance between imperceptibility and extraction accuracy. Reducing α to 0.6 increased the SNR values in several cases, but extraction errors began to occur, showing that weaker echo embedding may reduce distortion while also reducing extraction reliability. The robustness evaluation under AWGN and MP3 compression shows that the proposed method has partial tolerance against selected signal-processing attacks, but it cannot guarantee error-free extraction under distorted conditions. In particular, MP3 compression at 128 kbps produced average BER values of 0.0561 for Message 1 and 0.0477 for Message 2.

Overall, the proposed method provides a practical balance between audio quality and extraction accuracy under the evaluated conditions. However, the robustness evaluation is still limited to Gaussian noise and MP3 compression attacks. Future work should include filtering and resampling attacks, larger, more diverse audio datasets, statistical analysis, and comparative benchmarking under identical attack conditions. Further improvements may also consider adaptive synchronization, stronger echo coding, and error-correction mechanisms to improve message recovery under stronger signal distortions.

REFERENCES

[1] O. Evsutin, A. Melman, and R. Meshcheryakov, 'Digital steganography and watermarking for digital images: A review of current research directions', *IEEE Access*, vol. 8, pp. 166589–166611, 2020, doi: 10.1109/ACCESS.2020.3022779.

[2] M. B. Yel and M. K. M. Nasution, 'KEAMANAN INFORMASI DATA PRIBADI PADA MEDIA SOSIAL', *Jurnal Informatika Kaputama*, vol. 6, no. 1, pp. 92–101, Jan. 2022.

[3] F. Hazzaa, A. M. Shabut, N. H. M. Ali, and M. Cirstea, 'Security Scheme Enhancement for Voice over Wireless Networks', *Journal of Information Security and Applications*, vol. 58, p. 102798, May 2021, doi: 10.1016/j.jisa.2021.102798.

[4] S. S. Radke and D. S. Mishra, 'Review of Image Security approaches: Concepts, Issues, Challenges and Applications', *Journal of University of Shanghai for Science and Technology*, vol. 23, pp. 1047–1054, Jun. 2021.

[5] Musa. M. Yahaya and A. Ajibola, 'Cryptosystem for Secure Data Transmission using Advance Encryption Standard (AES) and Steganography', *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 317–322, Dec. 2019, doi: 10.32628/cseit195659.

[6] A. K. Agrahari, M. Sheth, and N. Praveen, 'Comprehensive Survey on Image Stegnography Using LSB With AES In-depth study of various requirements in the process of embedding, encrypting and extracting of text data', 2018. [Online]. Available: http://www.ripublication.com

[7] R. K. Dewi and R. Munir, 'PERBANDINGAN BERBAGAI METODE STEGANOGRAFI PADA CITRA DIGITAL', *Jurnal Informatika Polinema*, vol. 9, no. 3, pp. 289–230, May 2023, doi: 10.33795/jip.v9i3.1336.

[8] F. Yanti and K. Budayawan, 'Jurnal Vocational Teknik Elektronika dan Informatika', *Jurnal Vocational Teknik Elektronika dan Informatika*, vol. 11, no. 1, Mar. 2023, [Online]. Available: http://ejournal.unp.ac.id/index.php/voteknika/index

[9] A. Dwi Hendrata and A. Prihanto, 'Analisis Kualitas Suara Stego Audio Penyisipan Informasi Tersembunyi dengan Metode Least Significant Bit', *Journal of Informatics and Computer Science*, vol. 02, 2021, [Online]. Available: https://www2.cs.uic.edu/~i101/SoundFiles/.

[10] N. Kaur and S. Behal, 'Audio Steganography Techniques-A Survey', 2014. [Online]. Available: www.ijera.com

[11] S. A. Br Sibarani, A. Munthe, R. Purba, and A. A. Lubis, 'Pengamanan Citra Digital Menggunakan Kriptografi DNA dan Modified LSB', *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1233–1242, Dec. 2024, doi: 10.25126/jtiik.2024117666.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :

[12] L. F. Adhimah, I. Nurhafiyah, A. A. Muntahar, F. Kristiaji, and D. Mustofa, 'IMPLEMENTASI APLIKASI STEGANOGRAFI BERBASIS WEB MENGGUNAKAN ALGORITMA LSB DAN BPCS', *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 2, pp. 100–108, 2023.

[13] S. Rahman *et al.*, 'A novel and efficient digital image steganography technique using least significant bit substitution', Nature Research, Dec. 2025. doi: 10.1038/s41598-024-83147-3.

[14] C. E. Pangaribuan, O. K. Surbakti, F. Sihotang, Y. Tamba, Y. D. Ndruru, and M. Hasugian, 'Analysis Of The Utilization Of Echo Hiding Algorithm In Hiding Secret Messages', *Journal Majelis Paspama*, vol. 2, no. 2, pp. 102–107, 2024.

[15] S. Mishra, V. K. Yadav, M. C. Trivedi, and T. Shrimali, 'Audio steganography techniques: A survey', in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2018, pp. 581–589. doi: 10.1007/978-981-10-3773-3_56.

[16] S. Wen, H. Tan, J. Li, and T. Hu, 'A robust audio watermarking method based on dual-encoder U-Net and short-time Fourier transform', *Cyber Security and Applications*, vol. 4, Jan. 2026, doi: 10.1016/j.csa.2026.100129.

[17] I. Putu *et al.*, 'Perancangan Sistem Steganografi Berbasis Transformasi Wavelet Diskrit Terintegrasi Algoritma Rijndael dan QR-Code', *JNATIA*, vol. 2, no. 4, 2024.

[18] R. F. Damanik, R. Ardhia Pramesthi, G. Budiman, and S. Saidah, *Audio Watermarking Berbasiskan DWT-DCT-SVD-CPT Menggunakan QIM dan SS Audio Watermarking Based on DWT-DCT-SVD-CPT Using QIM and SS*. 2019.

[19] G. A. Simanungkalit, D. S. Tarigan, and D. R. Simangunsong, 'Discrete Wavelet Transform (DWT) Based Steganography Implementation', 2023. [Online]. Available: https://jurnal.seaninstitute.or.id/index.php/jutiρ13

[20] C. Umam and M. Muslih, 'Enkripsi Data Teks Dengan AES dan Steganografi DWT', *InComTech : Jurnal Telekomunikasi dan Komputer*, vol. 13, no. 1, p. 28, Apr. 2023, doi: 10.22441/incomtech.v13i1.15059.

[21] A. Syahdilan and M. A. Prawira, 'IMPLEMENTASI ALGORITMA DISCRETE WAVELET TRANSFORM UNTUK MENAMBAHKAN INVISIBLE WATERMARKING PADA CITRA DIGITAL', 2024.

[22] H. Rafiee and M. Fakhredanesh, 'Presenting a Method for Improving Echo Hiding', *Journal of Computer and Knowledge Engineering*, vol. 2, no. 1, 2019, doi: 10.22067/cke.v2i1.74388.

[23] S. Lahiri, 'Audio Steganography using Echo Hiding in Wavelet Domain with Pseudorandom Sequence', *Int. J. Comput. Appl.*, vol. 140, no. 2, pp. 16–22, Apr. 2016, doi: 10.5120/ijca2016909223.

[24] N. Faradiba, 'Frekuensi dan Intensitas Suara yang Bisa Didengar Manusia', *https://www.kompas.com/sains/read/2022/06/18/123100923/frekuensi-dan-intensitas-suara-yang-bisa-didengar-manusia*, Jun. 18, 2022.

[25] M. Hamdani and G. N. Samosir, 'IMPLEMENTASI STEGANOGRAFI UNTUK KEAMANAN PENGIRIMAN CITRA DIGITAL MENGGUNAKAN METODA DCT (DISCRETE COSINE TRANSFORM)', *Jurnal Penelitian dan Pengkajian Elektro*, vol. XX, no. 2, p. 42, Apr. 2018.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.xx.xx, Copyright ©2020
Submitted : , Revised : , Accepted : , Published :