CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter serves to establish a theoretical background regarding the topic of audio steganography. Subsequently, this literature will lead to the development of this research proposed algorithm theoretical framework.

This chapter begins to discuss audio steganography by discussing its basic framework, characteristics technique. This is followed by the related works in audio steganography. Next the research gap and critical analysis are presented leading to the development of theoretical work. Figure 2.1 presents the literature map to guide the reader throughout this chapter.

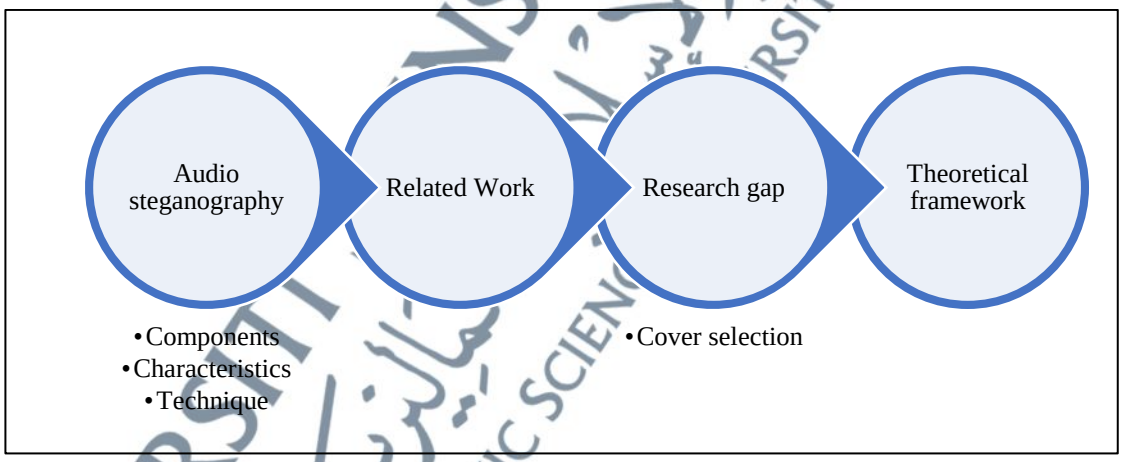


Figure 2.1: Literature Map

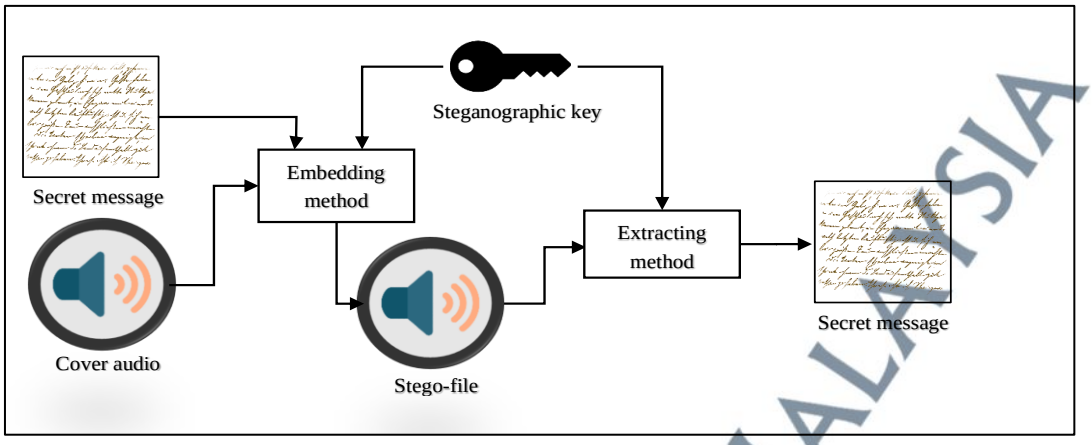
2.2 Audio Steganography

Steganography is both art and science that conceals the fact that communication has been established and will occur (Almalki & Mohammed, 2022; Linga Murthy et al., 2022; Maheswari et al., 2022). It is one of the strategies for preventing data breaches

during data transmission (Vimal, 2014). It conceals hidden messages within a proper multimedia carrier, such as an text, image, audio, or video, referred to as a cover file (Boulesnane et al., 2021; Kuznetsov et al., 2021; Wu et al., 2020). Steganography based on images and videos exploits the Human Visual System's (HVS) inability to recognize luminance differences higher than 1 in 240 in continuous grey levels or 1 in 30 in random patterns (Banjarnahor et al., 2019; Karnataka, 2020). However, audio-based steganography is used to influence the Human Auditory System (HAS) (L. Chen et al., 2021; Choudhury, 2018; Ganwani et al., 2021). Nonetheless, HAS is more sensitive to noise than HVS (Banjarnahor et al., 2019; Moorthy & Venkataraman, 2021). It is a difficult method to master because the human ear can detect even minor changes in audio. However, this approach can hold much information while only slightly changing the amplitude (Sharma & Ravinder, 2015).

2.2.1 Audio Steganography Components

There are three main components in the audio steganography which are cover audio used, secret message and the embedding method itself (Ahmed A. Alsabhany, Ali, et al., 2020). The most commonly used file format for the cover audio is .wav, AAC and .mp3 (K. Chen et al., 2020; Ren et al., 2023; Yang et al., 2020). Often, some embedding methods require a steganographic key to improve the security of their methods. Figure 2.2 illustrates how the components of audio steganography fit in audio steganography model.



Source Abdulrazzaq et al. (2020)

Figure 2.2: Basic Audio Steganography Model

Based on Figure 2.2, to generate the stego-file, this embedding method needs cover audio, secret message and steganographic key. After it completes its own processes, it produces stego-file. Then the extracting method takes the stego-file to run its own process to retrieve back the secret message.

2.2.2 Audio Steganography Performance Characteristics and Their Evaluation Metrics

Any audio steganography technique must satisfy five essential characteristics of capacity, imperceptibility, robustness and security (Bhowal et al., 2017). There are two different types of evaluation conducted to evaluate these five characteristics which are objective evaluations and subjective evaluations. This literature focuses on objective evaluations as subjective evaluations tend to produce slightly different results depending on the person who was tested on.

2.2.2.1 Capacity

Capacity is defined as secret message size ratio to the cover file size, or it can be defined as the embedding rate (ER), which refers to the maximum message size per second of audio. The main aim of the capacity evaluation is to measure the maximum size of a secret message that can be embedded. One of the metrics which can be used to measure capacity is embedding rate, ER. ER can be calculated using the following Equation 2.1:

$$ER = BPDU * DUPS \quad (2.1)$$

where *BPDU* denotes the number of embedded bits per data unit, and *DUPS* denotes the number of data units per second.

However, ER was not always can be used in the capacity evaluation because some audio steganography methods compared with the proposed method have non-sequential embedding behaviour. The capacity of the steganography method which has non-sequential embedding behaviour is impossible to measure accurately using the embedding rate due to some embedding rules and mapping, which avoided some samples. Instead of using ER, it would be better to utilise the maximum secret message that could be successfully inserted in each file as the metric for the evaluation.

2.2.2.2 Imperceptibility

Imperceptibility is the degree of ability to embed hidden code without the degree of similarity and error between the original cover and the stego-file (Zumchak, 2016). The main aim of the imperceptibility evaluation is to measure the difference in degradation between the original cover and stego-file. The most common evaluation metric for evaluating the imperceptibility of audio steganography is the signal-to-noise ratio (SNR), which has been used in more than 90 audio steganography methods

(Ahmed A. Alsabhany, Ali, et al., 2020). The SNR measures the degradation level in signal quality which is measured in decibel units (dB) and calculated using Equation 2.2 (Al-Hooti et al., 2018).

$$SNR = 10 * log_{10} \frac{\sum_1^n x(i)^2}{\sum_{i=1}^n (x(i)-y(i))^2} \quad (2.2)$$

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

Other objective evaluation metrics such as mean squared error (MSE) and peak signal-to-noise ratio (PSNR), have also been implemented in the literature to provide another evaluation for assessing the performance of the existing method. Although identical patterns are predicted to be formed, these evaluation metrics were nonetheless employed to verify that the SNR result is complemented and validated. It may also be utilised to discover any anomalies in the result generated by the SNR assessment, since it is known that the pattern generated is similar. MSE computes the sum of squared errors between the stego-file and the original cover audio. It can be calculated using Equation 2.3 (El-Khamy et al., 2018):

$$MSE = \frac{1}{n} \sum_1^n (x(i) - y(i))^2 \quad (2.3)$$

where x, y, n and i carry the same definition as those used in the *SNR* equation.

PSNR assesses the maximum SNR and can be calculated using Equation 2.4 (El-Khamy et al., 2018):

$$PSNR = 10 * log_{10} \frac{R^2}{MSE} \quad (2.4)$$

where R is the peak signal value in the original cover audio.

2.2.2.3 Robustness

Robustness is the ability of a secret message to withstand attack which try to destroy the stego-file (Srivastava & Rafiq, 2012; Vimal, 2014; Zumchak, 2016). The main aim of the robustness evaluation is to measure the resistance of the stego-file against these attacks. The most common attack used for evaluating the robustness of audio steganography is Additive White Gaussian Noise (AWGN). AWGN is an attack that introduces white Gaussian noise to the stego-file in order to disrupt its signal. Another common attack is compression attack. It compresses the stego-file and decompresses back the stego-file which will affect the signal of the stego-file. Several other attacks which also disrupt the signal are scaling, resampling, resizing, cropping, amplification, and filtration. The degree of the disruption caused by this attack commonly be measured by bit error rate (BER). It is an objective measurement to calculate the error from the secret message retrieval and measures the ratio of the number of embedded message bits that lead to an error in the retrieval process to the total secret message size. *BER* is calculated using equation 2.5:

$$BER = \frac{number\ or\ retrieval\ error}{total\ number\ of\ message\ bits} * 100 \quad (2.5)$$

BER was used in the experiments to highlight the secret message embedding and retrieval accuracy. A BER rate of 45% and above indicates a retrieval failure. It suggests that the message retrieval is capped based on error probability, which is 50%, as there are only two values in binary form, 0 and 1 (Ahmed A Alsabhany, 2019).

2.2.2.4 Security

Security represents the ability to prevent a secret message from being illegally accessed (Bhowal et al., 2017; Mary Jenifer et al., 2018). The main aim of security to

measure the defence level of the stego-file against these attacks. There are various category attacks such as unauthorised removal, unauthorised embedding and unauthorised detection. Example of attacks under unauthorised removal category are masking attack, eliminating attack and collusion attack. Example of attack under unauthorised embedding are forged embedding to overwrite the original stego file. Lastly, example of attacks under unauthorised detection such as information leakage attacks and secret message reading attacks. Usually, after these attacks are conducted on the stego-file, and accuracy of these attack are measured to determine the security level through the percentage of the successful attack out of total attack launch.

2.2.2.5 Dynamic Security

In addition to these four characteristics, there is also a new characteristic that needs to be considered, namely dynamic security, which was recently derived from security characteristics by Alsabhany (2019). Dynamic security refers to the ability of the embedding technique to appropriately distribute the secret data throughout the cover audio in any situation, preventing a visual or statistical attack. A visual or statistical attack is an attack that uses visual representation or some statistical data in order to distinguish the altered and unaltered parts of a stego-file in different situations. This research categorises situations as underloading, standard-loading, and overloading. Underloading is described as the situation where the audio steganography cannot distribute the secret message throughout the cover used due to the small size of the secret message. Standard-loading is the situation where the audio steganography manages to distribute the secret message throughout the cover audio. Lastly, the overloading is described as the situation where the audio steganography does not

manage to embed all the secret message due to the huge size of the secret message, which that cover cannot contain.

The real threat that dynamic security aims to minimise is when the steganalysis approach is unable to reliably distinguish embedded signals from clear signals from the underloading and standard-loading situations. Figure 2.3 displays the secret message distributions between the visual attack on the stego-file on the underloading situation.

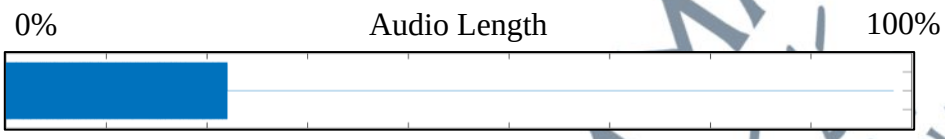


Figure 2.3: Secret Message Distribution in Underloading Situation

Figure 2.3 shows that the secret message can only be distributed to a part of the cover audio. This situation makes it obvious that there are differences between the unaltered cover and altered cover, which give the attacker information that the message length is that many percentages compared to the audio sample used. Next, Figure 2.4 displays the secret message distributions between the visual attack on the stego-file in a standard-loading situation.

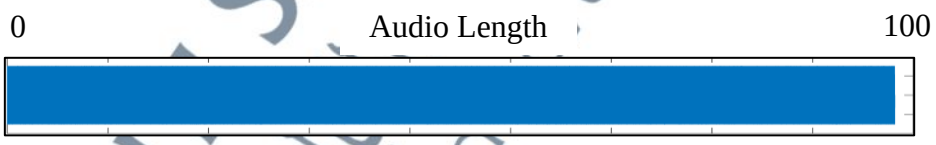


Figure 2.4: Secret Message Distribution in standard-loading Situation

Figure 2.4 shows that the secret message distribution is managed to be distributed throughout the cover. A well-distributed secret message in a standard-loading situation makes it harder for the attacker to determine the actual message length. Lastly, Figure 2.5 displays the secret message distributions between the visual attack on the stego-file on the overloading situation.

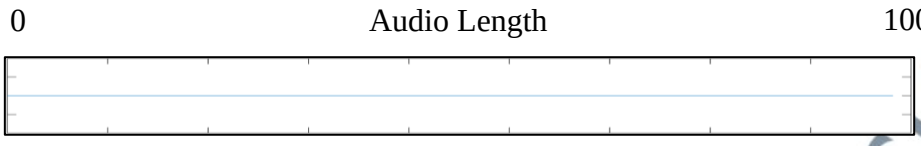


Figure 2.5: Secret Message Distribution in Overloading Situation

Figure 2.5 displays that the secret message cannot be distributed at all. In this situation, there is no distribution of secret messages due to the embedding failure, which makes the sender fail to communicate.

The main aim of dynamic security evaluation is to measure the defence level of the stego-file against attacks. There are two common attacks which are seg-SNR spike visual test and difference signal visual test. Both of these attacks require examiner to evaluate manually the visual produced.

Seg-SNR spike visual test compares the cover audio used and their respective stego-files regarding their segmental-SNR. The seg-SNR, the cover and the stego-file were divided into equal sizes with 50% overlap and the SNR of each segment was calculated. The outputs of these SNR values were saved into an array, divided into several distinct groups. The frame size is set to 256 with 50% overlap following Alsabhany et al. (2019) and ten (10) clustering groups were used. The length of segments is typically set between 0.015 to 0.020 seconds for speech (Özer et al. , 2003). Given that the speech audio has a 44.1kHz sample rate, the audio sample per segment is estimated to be more than 600, which is higher than the segment size used by Alsabhany et al. (2019). Then, each cluster's average was calculated to represent the SNR in a certain segment.

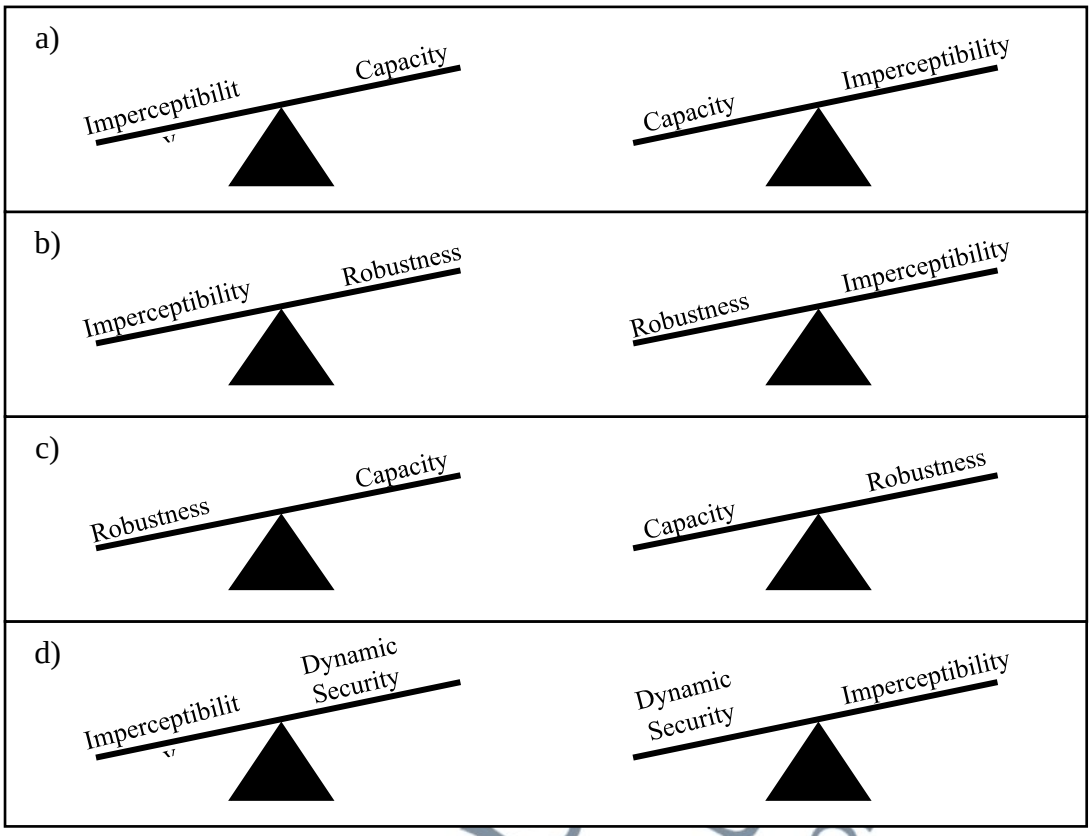
Difference signal visual test was introduced by Ballesteros and Moreno (2012). It analyses the difference between cover and stego-files using a wavelet representation of each signal. It focuses on the error distribution in an underloading situation. The

difference in the signals is determined concerning the independent capacity of each method.

2.2.3 Audio Steganography Characteristics Trade-offs

It is almost impossible to achieve excellent performance in all five characteristics simultaneously due to the following trade-offs: 1) trade-off between capacity and imperceptibility; 2) trade-off between imperceptibility and robustness; 3) trade-off between robustness and capacity, and 4) trade-off between imperceptibility and dynamic security. Figure 2.6 displays all the characteristics with their trade-off relationship.

Based on Figure 2.6, the first trade-off, capacity–imperceptibility, states that as message size increases, transparency (as shown by degraded cover quality) decreases (Kaliappan Gopalan & Shi, 2010). Next, imperceptibility-robustness, demonstrates that even when capacity is limited, keeping a high level of robustness leads to a significant loss in imperceptibility quality, as demonstrated in the method proposed by Kaur et al. (2017). Furthermore, robustness–capacity, suggests that as message size rises, the embedded message's ability to withstand modification attack decreases (Bender & Gruhl, 1996). Finally, dynamic security-imperceptibility suggests that a minimum size of the secret message is needed depending on the embedding techniques themselves to ensure that the message can be distributed through the cover audio. Until the embedding technique achieves the minimum size required, it always becomes a trade-off between these two, as the capacity needs to be increased to achieve the minimum size. While increasing the capacity, the imperceptibility is reduced.



Source: Alsabhany et al. (2019)

Figure 2.6: Characteristics of Trade-off Relationships

2.2.4 Audio Steganography Techniques

Audio steganography can be implemented either in time domain or in transform domain (Kasetty & Kanhe, 2020; Narula et al., 2020). Several existing audio steganography techniques commonly implemented under these domains are LSB modification, phase coding, echo hiding, parity coding, spread spectrum, and wavelet domain (Divyashree et al., 2020; Mostafa et al., 2019).

LSB modification: this technique alters the audio signal by embedding the secret message into the low significant bit (Adhiyaksa et al., 2022; Dutta et al., 2023). It is easy to implement and can be combined with other techniques (Altinbaç & Yalman, 2021).

Parity coding: this technique decomposes the original signal into distinct sample sections and conceals the secret message in the sample region's parity bits. If the secret message matches the check bit, proceed. Otherwise, reverse the LSB bit of a sample in this region.

Echo hiding: the hidden data is disguised as a brief echo embedded within the original signal in this technique. The essence of echo is an extra resonance to the original signal. To avoid echo interference, stego-file must maintain the same statistical and perceptual characteristics as the original signal after echo addition. The approach modifies three parameters of the echo signal to render it inaudible: decay rate, offset, and beginning amplitude.

Phase Coding: this technique can embed the secret message bit into the phase shift in the appropriate spectrum of digital signal by altering the phase of the original audio segment since the human auditory system cannot distinguish the phase change in the signal simply as the noise in the signal. It uses discrete fourier transform (DFT), which is a transformation algorithm for the audio signal. At the receiver, we must know the length of the segment before we can apply the DFT to get the phase to extract the secret message.

Spread spectrum: this technique is comparable to the system implemented by LSB technique, which distributes the bits of messages on audio files at random. It distributes the secret message over the frequency spectrum of the audio file using a code which is independent of the real signal. Therefore, the final signal takes up more bandwidth than is necessary, which will also introduce random noise to the audio and cause data loss

Wavelet domain: this technique is based on discrete wavelet transform. The signal can be divided into high frequency and low frequency parts, and the low frequency part can be decomposed into high frequency and low frequency parts again. The number of

signal decomposition is mainly dictated by the application and length of the original signal. By integrating with wavelet energy, masking effect, adaptive and LSB, we can embed secret information into discrete wavelet coefficients.

All these techniques have different approaches in order to improve certain aspects of these characteristics. After considering the most common audio steganography techniques, this research is restricted to the Least Significant Bit (LSB) modification because of its high capacity and imperceptibility (Tan et al., 2019). As the name implies, LSB embedding is used to embed binary data by modifying the audio frame's least significant bit. It is straightforward to implement and generally has a high capacity and imperceptibility. It also famous for it simplicity (Cui et al., 2020). However, it lacks robustness, since minor format changes caused by file conversion, compression, or an Additive White Gaussian Noise (AWGN) steganalysis attack can readily corrupt the concealed data.

2.2.5 Cover Selection

One of the important elements in steganography is the cover itself. As discussed in Figure 2.2, the sender must send the stego-file created throughout the unsecure channel. The stego-file, which the sender sends, has a better chance of being undetected by the attacker if the stego-file has better characteristics which would avoid the suspicion of the attacker. Therefore, the steganography algorithm can create a better stego-file with high quality by choosing a good cover.

Cover selection has three main elements: 1) types of steganography cover, 2) comparative criterion, and 3) searching method. The first element is the types of steganography cover. Common cover types used in steganography are image, video, and audio. Next, the comparative criterion is the rule used to compare between the

available covers to select the most suitable one. These rules are based on steganography characteristics such as robustness and imperceptibility. Lastly, the third element is the searching method. In this research, a searching method is defined as a method used to retrieve the solution by calculating the search space of a problem domain, which is the trade-off between characteristics of audio steganography.

Commonly, the searching method is divided into two major categories: brute-force search and heuristic search. All these categories have different approaches in order to find the solution. However, this research focuses on the heuristic which is relatively faster compared to the brute-force search considering cover selection deals with arrays of audio samples, binary secret messages and other lengthy parameters during repeated embedding processes which is a time-consuming problem (Russell & Norvig, 2010). There are several existing heuristic search method such as Non-dominated Sorting Genetic Algorithm (NSGA) (Siinivas & Deb, 1994), Non-dominated Sorting Genetic Algorithm 2 (NSGA-II) (Deb et al., 2002), Strength Pareto Evolutionary Algorithm (SPEA) (Zitzler & Thiele, 1999) and Strength Pareto Evolutionary Algorithm 2 (SPEA2) (Zitzler et al., 2001). As old methods, SPEA2 and NSGA2 have already supplanted NSGA and SPEA, respectively, they are no longer considered. There is no obvious winner between NSGA-II and SPEA2, since their employment depends on the problem that was sought to be addressed. However, NSGA-II offers advantages over SPEA2 in noisy environments regardless of the quantity of noise present in the problems (Bui et al., 2004). Thus, NSGA-II is chosen to be implemented as searching method in this research.

When creating a new cover selection technique, the researchers always consider two characteristics: accuracy and time consumption. The accuracy refers to the ability to determine the best cover for the steganography algorithm which produces the best

stego-file possible reliably and precisely. On the other hand, time consumption refers to the ability to find a cover for the steganography implementation swiftly. However, there are trade-offs among them. In order to find cover accurately, a huge amount of time must be invested to search the entire pool of cover. Some of the cover selection methods that prefer accuracy such as Molato and Gerardo (2018) and Rashid (2020). On the other hand, the selection process cannot undergo the entire available cover to determine the best cover for the steganography implementation. For example, Hajduk (2017) and Shah and Bichkar (2020) select the appropriate cover instead of the best cover to be used to speed up the selection process. This research focuses on the accuracy characteristic. The next decision in determining the cover selection class is based on this characteristic.

2.3 Approach in LSB Technique

Common approaches for embedding secret information in cover audio based on LSB technique are:

1) Low Bit Embedding

   Low bit embedding is an approach that embeds secret information in the audio sample's low-bit (s). First, the audio sample and the secret message were converted to binary format. Then the audio sample's least significant bit(s) was substituted with the secret message one by one until the entire secret message was successfully embedded. This approach embedded one to eight (8) bits per sample (*bps*) from low bit audio samples, with an embedding potential of up to 50%. Several methods in this approach were analysed, and the summary of this analysis is presented in Appendix 1.

2) Variable Low Bit Embedding

Variable bit embedding is a more advanced variation of the low-bit embedding method that provides greater dynamic security. Some previous methods that fall under this approach use fixed audio sample amplitudes. In contrast, the others use dynamic audio sample amplitudes to determine the audio sample and bit embedded. Several strategies in this approach have been analysed, and the summary of this analysis is provided in Appendix 2.

3) Sample Selection Embedding

Several rules were used as an indicator to select the sample used for embedding the secret information. Sample selection has better dynamic security than low-bit embedding as it skips several samples, providing a better distribution of secret information. Several strategies in this approach have been analysed. The summary of this analysis is provided in Appendix 3.

4) Bit Selection Embedding

Similar to sample selection, several criteria were used as an indicator to select the bit used instead of selecting the sample audio for embedding the secret information. However, bit selection has a lower performance in terms of dynamic security than sample selection as it embeds consecutively into the audio sample array. The summary of this analysis is provided in Appendix 4.

5) Parity Coding-based Embedding

The parity coding-based embedding approach does not embed actual secret information inside the audio sample. Instead, parity coding-based embedding uses parameter values such as secret message, audio sample or secret keys to determine the parity of the audio sample before deciding to flip its least significant bit (LSB). As a result, it provides low dynamic security as it embeds

the secret information in sequential audio samples. Appendix 5 summarise the analysis of several methods under this approach.

6) XOR-based Embedding

The XOR-based embedding method is comparable to the parity coding-based embedding method in that neither approach embeds actual secret information into the audio sample. Instead, the XOR-based embedding approach conducted an XOR operation on the parameters, such as the binary value at 1st and 2nd place LSB. It also provides low dynamic security as it embeds secret information in sequential audio samples. Appendix 6 summarise the analysis of several methods under this approach.

7) Averaging the Amplitude-based Embedding

The last approach is embedding based on averaging the amplitude. The amount of embedded *bps* depends on whether it is above, at, or below the average. This approach also provides better dynamic security than the one that embeds the secret information into audio samples sequentially, as it only selects the audio samples with higher amplitude than the average amplitude. Appendix 7 summarises the analysis of several methods under this approach.

All these approaches are able to improve certain aspects of these characteristics. After considering the most common approaches, this research is focus on the low-bit embedding approach because of it's high embedding capacity compared to other approaches. However, as mentioned above, it lacks dynamic security, hence a new design needs to be considered in order to maintain the distribution of the secret message to achieve high dynamic security.

2.4 Related Works

Various methods were proposed to improve the characteristics of audio steganography. Two common elements are used to improve the capacity characteristics, which are: data compression and embedding more bits per sample. Vimal & Alex, (2014) implement Huffman encoding to improve the capacity characteristic of the method through data compression. On the other hand, Indrayani (2020), M. A. Ahmed et al. (2010) and Cvejic & Seppanen (2002) embeds the secret message up to 8 *bps* to improve their methods capacity. There are also methods that combined both elements to improve the capacity such as Alsabhany et al. (2018) and Alsabhany (2020). Based on existing methods, all of them provide better capacity performance than the original LSB proposed by Bender & Gruhl (1996). However, it is crucial to highlight that the best capacity performance may be attained by adopting data compression and increasing *bps* elements simultaneously. It is worth to noticed that, although implementing compression element can increase capacity, it requires additional computer calculations compared to raising the *bps*. Besides, increasing *bps* increases capacity more than data compression does. In general, the performance of LSB approaches differs based on the element used. However, suppose all the approaches standardise the elements used. In that case, the basic low-bit embedding approach has the highest capacity as it sequentially embeds the secret message in all the samples without any condition. In contrast, the others have some conditions for the sample used for the embedding process. Although the bit selection approach appears to have similar capacity performance based on the definition given above and the standardised elements used, it has a limit. Consider that the maximum embedding is 8 *bps* on the 16-bit length audio sample. In order to select the bit at the lower significant bits of the audio sample (which is the maximum of 8 bits from 16 bits), the bit selection approaches cannot embed at 8

bits as it still needs to do the selection process on which bit(s) can be used. Hence the maximum number of bits available is 7 bits out of 8 available bits from the low significant bits of the audio sample. Compared to sequential low bit embedding, which does not have any conditions and not skipping any samples, it managed to have 1 extra *bps* for higher capacity performance.

Next, various researcher improve the imperceptibility of LSB technique by implementing bit adjustments elements, embedding at lower LSB, or combining both. Bender & Gruhl (1996) embeds at lower LSB which only use least significant bit to embed 1 bit per sample. On the other hand, Cvejic, N., & Seppänen, (2005), Zamani et al. (2009), Park (2013), Hosny et al. (2019) and Mostafa et al. (2019) adjust the bits after the embedding process to reduce the error created. Cvejic, N., & Seppänen, (2005) introduces bit adjustment based on several conditions to reduce the error whle Hosny et al. (2019) implemented Simulated Annealing Algorithm and adjusting the audio bit to reduce the error. Zamani et al. (2009), Park (2013) and Hosny et al. (2019) take the advantage of genetic algorithm to find the best way to adjust the modified audio bit to reduce the error made. In general, the performance of LSB approaches differs based on the element used. However, suppose all the approaches standardize the elements used. In that case, the approaches that do not embed the actual data, such as, XOR-based embedding and parity coding-based embedding have the highest imperceptibility. However, it is not always true in the worst condition, where all the bits must be flipped. Therefore, the method with the next best imperceptibility performance is one which skips several samples based on its conditions, such as variable low-bit embedding and sample selection embedding.

Furthermore, to improve the robustness of the LSB technique most of the researcher embed at higher level of LSB. Hosny et al. (2019) embeds as high as 3 or 4

LSB place while Cvejic, N., & Seppänen, (2005) embed as high as 4 LSB place. On the other hand, Indrayani (2020) and Ahmed et al. (2010) embeds as high as possible which is at 8th place of LSB. It is crucial to highlight that the highest robustness can be achieved only by embedding at the highest bit possible. For example, with the usage of 16-bit audio samples, the highest possible LSB that can be embedded is at 8^{th} LSB which is why most of the researchers who target to improve the robustness embed as high as at 8^{th} LSB. In general, the performance of LSB approaches produces similar robustness if all the approaches embedded at the same level.

Next, to improve security focusing on the unauthorised detection such as message retrieval, the existing methods implements several elements such as embedding at higher level of LSB (M. A. Ahmed et al. 2010), introducing random element (Alsabhany et al., 2018; Park, 2013), using encryption (Sharma and Ravinder, 2015; Abdullah et al., 2015; Teotia and Srivastava, 2018) or using encryption and introducing random elements simultaneously (Vimal and Alex, 2014). Randomness can be introduced into the embedding technique of an audio steganography algorithm. For example, the cover audio parameter is used to implement randomness in algorithms such as those introduced by Vimal and Alex (Vimal & Alex, 2014), Satish Bhalshankar (2015), TVS et al. (2015), Meligy et al. (2015), and Alsabhany et al. (2018). However, using the cover audio parameters could expose the stego-file to the signal analysis. Besides, embedding techniques that use a genetic algorithm, such as the algorithm introduced by Park (2013), do not always produce the best-constructed cover audio possible, which further causes the inconsistency issue. Ali et al. (2018) introduces the embedding technique that uses a chaotic map to introduce randomness and needs to rely on the secure key management algorithm to work effectively.

Lastly, there are no previous LSB method addresses directly on the low dynamic security in traditional LSB. Currently, only Alsabhany et al. (2019) tackles dynamic security directly. However, this method is using phase coding technique. In general, existing methods under LSB technique such as Cvejic, N., & Seppänen, (2005), Zamani et al. (2009), Park (2013), Hosny et al. (2019) and Mostafa et al. (2019) has low dynamic security because of poor distribution of secret information throughout the cover audio used. However there are also some other LSB methods that are able to achieve manageable dynamic security such as method proposed by Alsabhany et al. (2018) and Alsabhany (2020). These methods have high secret message distribution compared to the other methods however, error differences between embedded segments are sometimes high due to the different embedding levels.

Based on existing method, bit embedded per sample manipulates the capacity, imperceptibility and robustness. After considering the most common audio steganography method proposed in the past and their focused characteristics, this research does not restrict fix number of bit embedded per sample which allow to manipulate more freely the characteristics that need to be improved. In addition of that, this research focuses on chaotic map as a mechanism to improve the security of the audio steganography method as it does not rely on the cover audio parameters which can be possibly retrieved by statistical attack. Lastly, although there are no LSB method yet that address dynamic security directly, however, several methods which have good distribution of secret message during the embedding process. As these methods achieve manageable dynamic security level, hence, this research focuses on improving the dynamic security through distribution of secret message.

2.5 Research Gap

Analysis on review paper (Alsabhany et al., 2020) and existing audio steganography methods presented in Section 2.4 was conducted to find the research gap. Several observations were made based on this analysis.

From audio steganography method development perspective, most works focused in increasing the characteristics through the development of the method itself.

Existing audio steganography methods emphasise the methods' formulation and design, such as the sample or bit used for the embedding, thereby ignoring the cover audio selection used to embed the secret message. As the quality of the stego-file relies heavily on the method itself, the sender is expected to be smart and meticulous when using any steganography technique and needs to understand about the results of the insertion of the secret data (Hasanah et al., 2018). The sender who does not know about audio steganography tends to choose any cover audio without considering the important parameters posed by the audio, contributing to either enhancement or cutback of audio steganography characteristics. This may lead to an improper selection of cover audio.

Compared to image steganography and video steganography, they consider the cover selection during the development of these methods which increased the method's performance. To support the claim on no audio steganography method that implement cover selection process, a further analysis on the 29 cover selection articles (Appendix 8 - Appendix 11) was undertaken, focusing on the type of cover used, the following graph in Figure 2.7 was plotted and observation was made:

Based on Figure 2.7, most of the research used images as a cover type for their cover selection-based steganography research, with only one research were using video as a cover type. None is using audio as a cover type to improve the steganography method.

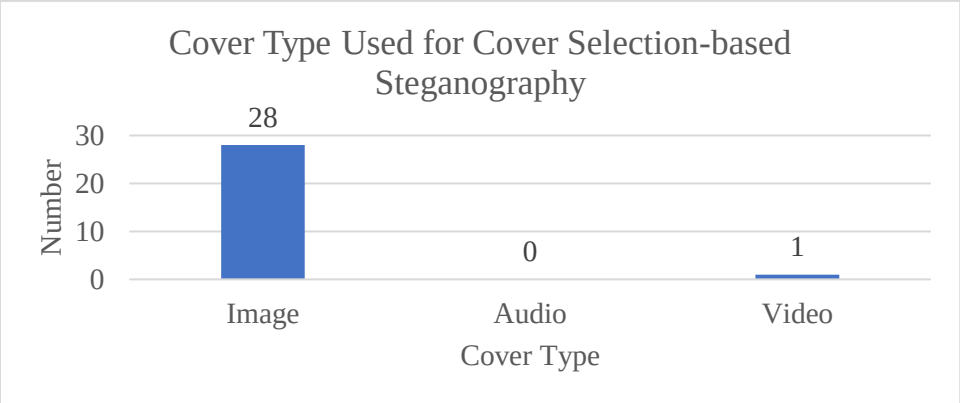


Figure 2.7: Cover Type Used for Cover-Selection-based Steganography

2.6 Cover Selection Critical Analysis

In addition to the fact that there is no cover selection for the audio proposed previously, there are also no reported criteria or categories to differentiate and emphasise the major variations between current cover selection approaches. These limitations catalyse the development of a new category. The disadvantages mentioned above are avoided in the proposed classification, and the details of cover selection in each method reflected by the behaviour are prioritised to classify and categorise these methods.

Although the three previously mentioned cover selection elements provide an excellent foundation for comprehending the literature, certain classification criteria cannot be constructed using the underlying characteristics for the following reasons:

1. The classification of techniques according to their searching method and steganography carrier are insufficient to show the variations between them. These characteristics, however, produce useful data solely for literature.
2. While the evaluation metric used for comparative criterion classification produces comprehensive classifications, it cannot highlight the most frequently used cover selection concepts. Additionally, the number of

categories in this categorization may be unrealistic due to minor discrepancies across the approaches.

To elaborate further, consider the following example. Both methods, A and B, operate over the same steganography carrier and use brute force for the searching method. However, the accuracy of A is lower than that of B, whereas the time consumption of A is lower than that of B. Indeed, many current approaches have identical values for these two characteristics. The introduction of a new category was motivated by the necessity to categorise and compare cover selection methods.

As a result, the searching method is viewed as a coarse-grained and all-purpose classification criterion. However, certain techniques may require numerous selection behaviours. As a result, the most prominent idea (MPI) is introduced, defined as the method's most prominent cover selection behaviour.

After extracting the MPI, the techniques are categorized according to their MPI similarity. A classification based on MPI can give a better level of segregation than existing approaches, hence boosting critical review. Additionally, this classification facilitates the mapping of the literature and the display of the most frequently followed directions in cover selection. In the previous example, the classification based on MPI gives a more comprehensive explanation for the performance discrepancies between existing methods. Method A follows a subregion-based selection approach where the selection is based on a certain area of cover used and partial secret message. On the other hand, method B implements a cover-quality-based selection approach where the selection is based on the cover quality. Hence, it is understandable why the method A is faster compared to the method B as it only computes and processes certain region which fasten the selection process compared to method B that computes the cover as a whole. It is also the reason why method B is more accurate compared to method A. This

approach will be explained further along with other proposed classification in Section 2.6.2.

2.6.1 Cover Selection Evaluation Metric

The quality of the cover selected is reflected by the metric when measuring the performance of characteristic individually. However, since this metric only reflects one characteristic, new metric needs to be adapted for evaluating these characteristics simultaneously as the trade-off also needs to be considered. Approach when handling multi-objective problem (MOP) can be taken and adapted in order to measure the trade-off. This research only measures imperceptibility, robustness and dynamic security. Although the capacity is not measured directly, any cover that cannot be used for embedding is rejected earlier. The best solution, *Best_Sol* can be calculated by normalizing the three objective function values and converting them into uniform and dimensionless scales. The highest value of *Best_Sol* indicates that it has the best quality since all the objective functions try to maximise the objective. The *Best_Sol* can be calculated using Equation 2.6 until Equation 2.12.

$$Best\_Sol_{dynamic\_security} = \frac{dynamic\_security_{sol} - dynamic\_security_{min}}{dynamic\_security_{max} - dynamic\_security_{min}} \quad (2.6)$$

$$Best\_Sol_{imperceptibility} = \frac{imperceptibility_{sol} - imperceptibility_{min}}{imperceptibility_{max} - imperceptibility_{min}} \quad (2.7)$$

$$Best\_Sol_{robustness} = \frac{dynamic\_security_{sol} - dynamic\_security_{min}}{dynamic\_security_{max} - dynamic\_security_{min}} \quad (2.8)$$

$$ratio_{dynamic\_security} = (Best\_Sol_{dynamic\_security} * 0.33) \quad (2.9)$$

$$ratio_{imperceptibility} = (Best\_Sol_{imperceptibility} * 0.33)$$
(2.10)

$$ratio_{robustness} = (Best\_Sol_{robustness} * 0.34)$$
(2.11)

$$Best_{Sol} = ratio_{robustness} + ratio_{imperceptibility} + ratio_{dynamic\_security}$$
(2.12)

Another approach of evaluating the selection made is by calculating the number of selected covers which is non dominated by other solution. The number of Pareto solutions (NPS) is the metric used to measure the number of solutions which has the highest characteristic at certain point without being dominated by other characteristic at that point.

2.6.2 Classification of Cover Selection

This section reviews the categorised steganography cover selection literature using MPI. Figure 2.8 shows the method distribution for each cover selection behaviour. Based on Figure 2.8, the cover quality-based selection has the highest percentage while subregion-based selection has the lowest percentage. Each cover's selection behaviour and the associated approaches are then extensively analysed and described. It is observed that the highest percentage of the methods falls under cover-quality-based selection followed by batch-based selection. Finally, similar small percentage falls under sub-region-based and similarity-based selection.

To enable effective comparison of steganography cover selection methods, each category is defined by key features that capture its essential aspects. These include the searching method, comparative criterion, and steganography carrier, which are fundamental components of any cover selection approach. Additionally, supportive techniques are highlighted to showcase any additional techniques that are incorporated

with the cover selection method, while limitations are also emphasized as a disadvantage feature. A special feature is included to validate and support the classification process based on the new classification, and to capture the differences that exist among methods of the same group. Overall, this classification system provides a comprehensive understanding of each class's performance, simplifying the cover selection process for this research.

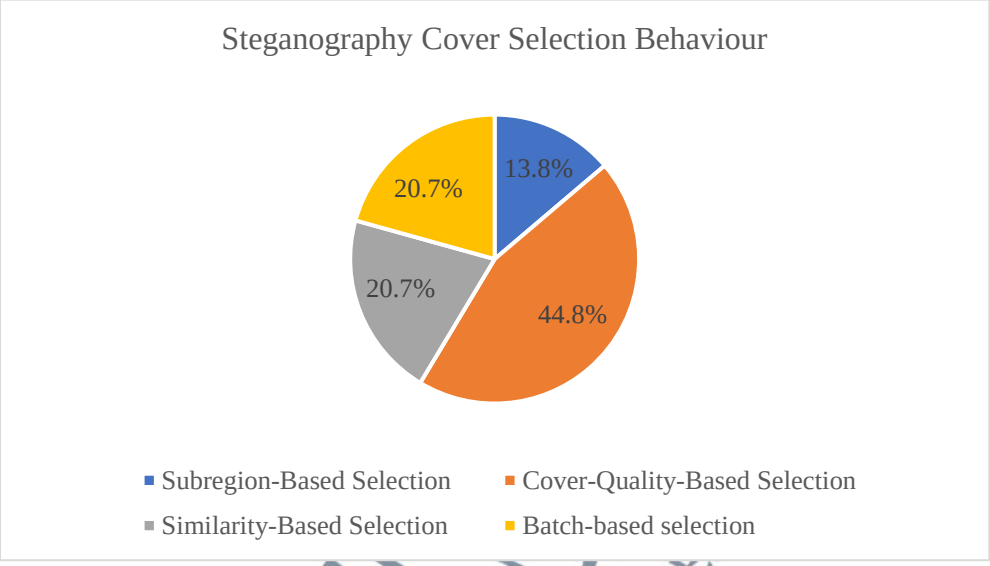
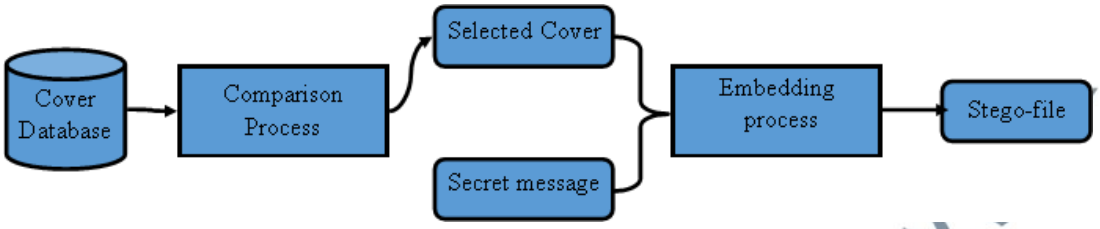


Figure 2.8: Steganography Cover Selection Behaviour and the Method Distribution

2.6.2.1 Cover-Quality-Based Selection

The method under this category selects the cover used for steganography based on the quality of the cover. The quality of the cover can be anything that defines the cover itself, such as contrast, image texture, and correlation parameter. The main concept of the cover-quality-based selection is illustrated in Figure 2.9. This pattern of cover selection has been observed in the methods summarised in Appendix 8.



Source: Hamid et al. (2021)

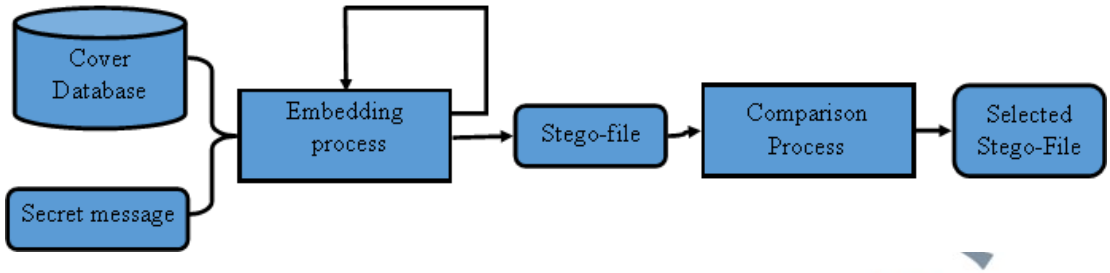
Figure 2.9: General Flow Cover-Quality-Based Selection

Based on Figure 2.9, the comparison process normally is conducted before the embedding process happens. The most general advantage of this method of cover selection behaviour is that it achieves high accuracy in determining the quality of stego-file produced. However, the main disadvantage is that it generally has high time consumption.

2.6.2.2 Similarity-Based Selection

The method under this category selects the cover used for steganography based on the similarity between the binary series of secret messages and the binary series of selected samples for embedding. The similarity metric can be anything that defines the similarity itself, such as SNR, number of flipping bits, and number of matching bits. The main concept of similarity-based selection is illustrated in Figure 2.10. Based on Figure 2.10, usually, the comparison process is conducted after the embedding process has happened. This pattern of cover selection has been observed in the methods summarised in Appendix 9.

The most general advantage of this method of cover selection behaviour is that it achieves high accuracy in determining the quality of the stego-file produced. However, the main disadvantages are high time consumption due to the need to conduct the embedding process for each cover in the database.

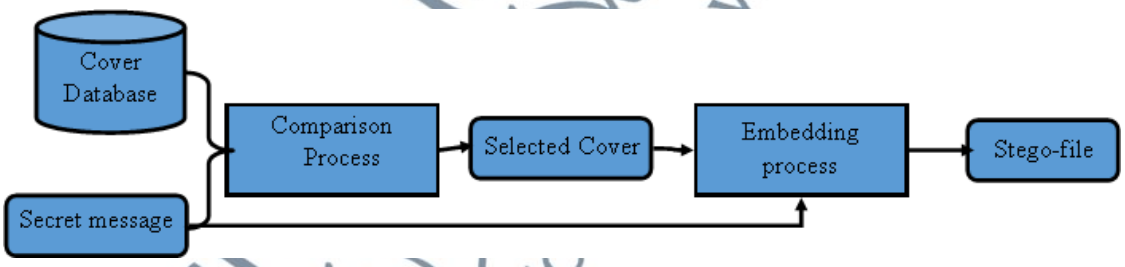


Source: Rashid (2020)

Figure 2.10: General Flow Similarity-Based Selection

2.6.2.3 Subregion-Based Selection

The method under this category selects the cover used for steganography based on a certain region of cover and some part of the secret message. The comparison metric is usually a fast-computed metric such as comparing the percentage number of the bits '1' between a partial secret message and a partial cover sample. The main concept of the subregion is illustrated in Figure 2.11. Based on Figure 2.11, typically, the comparison procedure occurs prior to the embedding step. This selection pattern has been observed in the methods summarised in Appendix 10.



Source: Wang et al. (2019)

Figure 2.11: General Flow Subregion-Based Selection

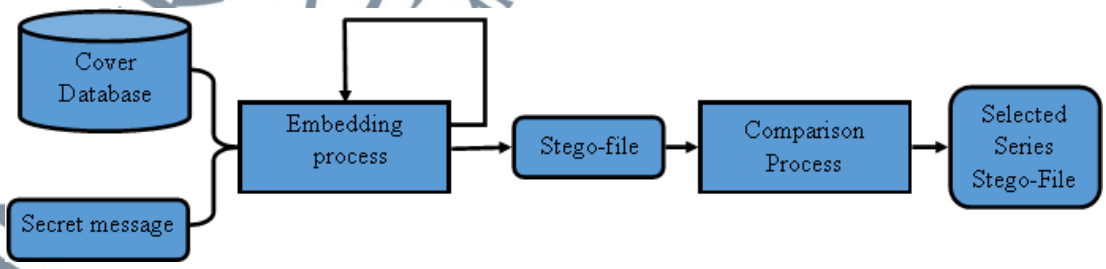
The most general advantage of this method of cover selection behaviour is that it achieves low time consumption because the method considers only certain areas in the cover and the secret message itself. However, the primary disadvantage is that it is less accurate than other methods since it does not consider total cover and an overall secret message, which results in somewhat different real results than expected results.

2.6.2.4 Batch-Based Selection

The method under this category selects the cover used for steganography based on the quality of the series of stego-file. Normally, it uses the same metric as similarity-based cover selection. However, instead of selecting only one stego-file to send through an unsecured channel, it selects a series of stego-files to send over. In addition, it has additional rules, such as finding the series of stego-files with the least amount of cover used and the lowest mean distortion. The main concept of batch-based selection is illustrated in Figure 2.12. Based on Figure 2.12, typically, the comparison procedure occurs after the embedding step. This embedding pattern has been observed in the methods summarised in Appendix 11.

The most general advantages of this method of cover selection behaviour are that it achieves high accuracy in determining the quality of the stego-file produced and improving the security against pooled steganalysis. However, the main disadvantage is that it has the highest time consumption in general due to the need to conduct the selection in two layers: individual layer and series layer.

All these selection approaches search for the most optimised cover based on the criteria they specify. However, this research focuses on the cover-quality-based as it focuses on the accuracy characteristic, secret message and the cover quality itself. The key research gaps and their discussion are presented in the next section.



Source: Xin & Jiaojiao (2018)

Figure 2.12: General Flow Batch-Based Selection

2.6.2.5 Summary of Cover Selection Based on Cover Selection Characteristics

All the methods in these classes achieve feasible accuracy and time consumption characteristics performances. However, each class has its suitability depending on the situation. For example, the cover-quality-based selection class works best when the secret message is known to the method without constant adjustment. In contrast, the similarity-based selection class works best when the secret message is unknown. Hence the consideration of the cover and the secret message is needed. Next, the subregion-based selection class is suitable when there is a time constraint in executing the selection process. Lastly, the batch-based selection class is suitable for the batch-based steganography. Based on these situations, this research selects a similarity-based cover selection class as a basis for the cover selection method proposed. As mentioned early in Section 2.2.5, this research focus on the accuracy of cover selection, the similarity-based cover selection class is the most suitable class due to current setup which the secret message is not known until the user insert it, ensuring the selection accuracy is not in jeopardy.

2.7 Theoretical Framework

The main objective in all steganography methods is to hide information by improving the steganography characteristics as much as possible whether capacity, imperceptibility, robustness, security, dynamic security, or combination of them. There are two main basic theoretical frameworks used in this research. The first one is through developing steganography method (Alsabhany et al., 2019; Bender & Gruhl, 1996) and the second one is through the input inserted such as secret message, cover file used and steganographic keys (Sajedi & Jamzad, 2008; Zhong et al., 2021). Figure 2.13 shows an illustration of theoretical frameworks of audio steganography.

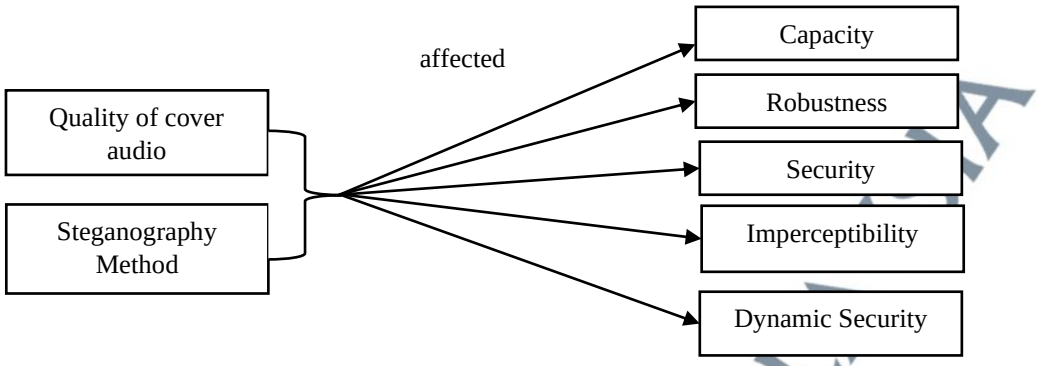


Figure 2.13: Theoretical Model of Audio Steganography

As previously stated, the main problem is improper cover selection which led to the low dynamic security. On the other hand, the current state of art of audio steganography method also has low dynamic security. Based on this problem, both theoretical frameworks relevant to this research as the improvement can be made by improving the quality of cover method and steganography method to improve the whole steganography system.

2.8 Summary

This chapter focuses on the literature review of the basics of audio steganography in terms of its components, performance characteristics, characteristic trade-offs, technique and cover selection. Furthermore, approaches in LSB are discussed such as low bit embedding, variable low bit embedding, sample selection embedding, bit selection embedding, XOR embedding, parity coding embedding and averaging the amplitude-based embedding. Related works and research gap were also discussed in this chapter. Besides, the cover selection critical analysis was presented which included newly proposed cover selection classification. The issues and solutions discussed previously are summarised in Table 2.1.

Table 2.1: Issues and Solutions for Proposed Method Design

Issues	Solutions
Low dynamic security	High distribution of secret message.
Capacity and dynamic security trade-off	Increasing bit embedded per sample and fully utilised the audio samples.
Direct message retrieval attack	Chaotic map
No cover selection method for audio	a selection rule or metric should be formulated to suitably evaluate the audio.
Audio steganography has trade-off which need to be considered.	all characteristics must be evaluated simultaneously to ensure the selection process does not neglect any characteristic
Evaluate dynamic security precisely	new precise evaluation metric.

Lastly, the theoretical framework is presented and becomes the basis of this research. The next chapter describes the research methodology.