J Inno Sci Eng,2022, 6(1):-122-137
http://doi.org/ 10.38088/jise.932549

Research Article

Comparative Analysis of Audio Steganography Methods

Funda ASLANTAS BEYIN [1*], Cemal HANILCI [2]

[1] *National Defense University, Army NCO Vocational HE School, Electronic Communication Technology, 10185, Balıkesir, Turkey.*

[2] *Bursa Technical University, Faculty of Engineering, Electrical and Electronics Engineering, 16330, Bursa, Turkey*

Abstract

Information security has been one of the most important issues of all time for both individuals and companies. Delivering data to the correct recipient is crucial for personal data protection, the privacy of personal life, and national security. To this end, different methods have been developed over the years to hide information from malicious individuals. Steganography is one of the most important information hiding methods that received great attention. In this study, five different audio steganography techniques (least significant bit, echo hiding, wavelet coding, spread spectrum, and cepstrum) are utilized and a comparison of these techniques is performed on Turkish audio recordings. To this end, hidden messages of various sizes were embedded into 20 audio recordings from 10 male and 10 female speakers using different embedding algorithms. Signal-to-noise ratios (SNR) computed between stego and cover audio files show that embedded message length and frame size are the main factors that determine the quality. In addition, it is observed that there is no perceptual difference between the cover and stego audio recordings. Hence, the human auditory system is unable to determine whether an audio recording is authentic or conveys a hidden message. Experimental results show that as the message length increases, the average SNR value decreases irrespective of the steganography technique, as expected. The well-known least significant bit (LSB) technique yields the highest average SNR value among the five steganography methods. The spectrographic comparison of the cover and stego audio recordings shows that hiding the secret message in an original audio signal highly affects the high-frequency region more than the low-frequency components.

Keywords: Audio Steganography Techniques, Stego Audio, Human Auditory System, Data Hiding, Secret Message.

.

Cite this paper as:
Beyin Aslantas, F. and Hanılcı, C. (2022). *Comparative Analysis of Audio Steganography Methods*.Journal of Innovative Science and Engineering.6(1):122-137

*Corresponding author: Funda Aslantas Beyin
E-mail: ???

Received Date:07/05/2021
Accepted Date:22/01/2022

1. Introduction

Information security is a very important topic in all areas of life. Steganography which is the subject of this study is a popular information hiding technique. Steganography, coming from the Greek words steganos, meaning roof or covered and graphia that means writing, is the art and science of hiding the fact that communication is taking place [1]. Increasing cyber-attacks with the advancement of technology and the widespread use of the internet is one of the main motivations of data hiding studies. It is clear that instead of encrypting the data (text, image, video, or audio) integrating the hidden message with the media file via different methods will make it more secure. However, the original media file is desired to preserve its originality and content while doing this, the media file should not be too far from its original state. Otherwise, the steganography process will not go beyond corrupting the original file. The media file in which the hidden message is embedded is referred to as the *cover* file and the resulting multimedia content after hiding the data is called the *stego* file. Therefore, using the audio signal as a cover provides a perspective that improves the quality of the stego signal. This is because each sample of an audio signal is usually represented by 16 bits (in standard pulse code modulation (PCM) audio waveform); therefore, using audio signal as the cover file yields a large number of cover bits. Due to the large number of cover bits, the inserted message bits do not cause considerable distortion in the signal. As the number of bits used in the cover file increases, the discernibility (by Human Auditory System - HAS) of the message hidden in the media and the distortion rate of the signal decreases which increases the signal-to-noise ratio (SNR) of the signal. If an image file is chosen as a cover file instead of an audio file, it is more likely to be noticed because the message embedding is performed on each pixel value. To eliminate this disadvantage, the image file should contain a large number of samples (pixels) like an audio file. This corresponds to an increase in resolution and data size. However, this is not necessary for audio steganography. In audio steganography, the message is embedded in an audio file. As with the cover file, the message to be hidden can also have different file types such as image, video, and sound. In other words, an audio file or an image file can be embedded into the cover audio file. Although the user on the sender side determines the data type in the message file, different parameters such as the corruption rate, embedding capacity, and size of the file play an important role in determining the message type. If a text message is selected as the message type, it will be more practical to convert this message to the binary sequence with the alphanumeric equivalents of the characters and integrate them into the cover file. In other message types, the message media file should be selected and sampled in accordance with the cover audio file. After pre-processing the message file, the embedding task is performed. Therefore, the process of embedding a text message in the cover audio file will be faster than other message types since pre-processing a text file is more practical than other file types.

The original audio file called the cover, the message to be hidden, the stego key required for the receiver to reach the message, and the stego sound file generated after embedding the secret message are the main elements of the steganography process. Stego key should be selected by taking the Kerckhoff's principle into account. According to Auguste Kerckhoffs, even if the key is known by others, this should not create any system vulnerability capacity. At the same time, Kerckhoffs argued that the key should have a changeable structure

and should be easy to use and transmit. [2] While the message embedded in the cover medium is perceptually indistinguishable by the listener, users who know the stego key and data embedding algorithm can access the secret message. This is because the message inside the stego file is not noticed by a listener, so the human sense system cannot detect changes in bits or pixels.

Steganography has always been an interesting topic for researchers and there exist many studies in the literature about its definition and history [1]. Few studies have been conducted due to the difficulty of applying audio steganography. Usually, text files or images are chosen as cover media files to hide messages. Dutta et al. [3] put forward in their study that the least significant bit (LSB) method has the highest quality rate in sound steganography. Cvejic and Seppänen [4] embedded four bits per sample into the audio signal to increase the embedding capacity and it was shown that the proposed method increased the embedding capacity without substantial deterioration in quality. Contrary to the five techniques used in [4], only ASDWT-TIA (Audio Steganography Discrete Wavelet Transform-Text in Audio) technique was used to hide the text message in the audio file. In [4], SNR and Squared Pearson Correlation Coefficient (SPCC) metrics were used together for performance evaluation and the performance of the method against possible common attacks was also evaluated. It was reported that the proposed method provides high embedding capacity. In [5], LSB coding, parity coding, echo hiding, spread spectrum, phase coding, tone insertion, and amplitude coding techniques were explained in theory, but a comparative performance analysis of these different methods was not reported. Although six different audio steganography techniques were discussed in [6], authors focused on safety concerns coming to the fore rather than the performance comparison of the methods. In the same study, the authors proposed a 2-level steganography technique (LSB and bitwise embedding) to increase the security. Bhavana et al. [7] performed steganography with MATLAB using audio files in PCM (wav) and mp3 file format.

Wakiyama et al. [8] examined only LSB among various steganographic methods. To increase embedding capacity, a new method called "*The variable low bit coding*" was proposed. In the experiments, the authors used audio files sampled at 22.05 kHz and SNR was used as the performance evaluation metric.
In [9], LSB, echo hiding, tone insertion, phase coding, cepstral domain, spread spectrum, and wavelet transform-based audio steganography techniques were compared according to SNR criterion. In [10], "Using XORing of LSB" technique, which offers a new look at the LSB method for data hiding, was proposed. The proposed method, unlike the classical LSB, first checks the parity of the samples instead of directly replacing the message bits and LSBs and then embedding them. In the same study, LSB, parity coding, phase coding, spread spectrum, and echo hiding were also included and an overview was presented. In the experiments, audio files with 2-8 seconds duration where each sample is represented with 16 bits were used.

Shah et al. [11] performed audio steganography by replacing the wavelet packet coefficients of the original cover audio signal with the proposed technique, called Adaptive Wavelet Packet Based Audio Steganography. It was shown that the noise generated when applying steganography with this technique is quite low and therefore they obtained a high SNR (dB) value. Asad et al. [12] proposed two methods to increase the

sensitivity of the LSB technique, which is frequently used in steganography and steganalysis. The first method is to randomize the number of message bits to be embedded whereas the second method randomly selects the signal sample to be changed in the cover sound signal. Gupta and Sharma [13] proposed a new method using the Discrete Wavelet Transform (DWT) concept and the LSB technique, again taking the security concerns into account. After converting the cover audio file into binary form, an image message was embedded into the cover audio file. Then Discrete Wavelet Transform (DWT) was applied to the new signal formed after the message was embedded. In this way, it was shown that the steganography process is safer. Bilal et al. [14] discussed the latest techniques of steganography and presented a comparison of their advantages and disadvantages. Kartheeswaran et al. [15] suggested Multi-Agent Based steganography, which applies an encryption technique to the message, and suggested that a more secure steganography could be made. Lindawati and Siburian [16] implemented steganography through LSB technique using an android smartphone. While applying the LSB technique, Rajput et al [17] proposed to use the most significant bit (MSB) of the cover audio signal. They argued that they increased the security in steganography by embedding two bits of the message signal into its LSB at once depending on the three MSB of the cover audio signal. Teotia and Srivastava [18] presented a hybrid algorithm that reduces errors in audio and video steganography. They compressed the data using the Huffman coding compression technique. It was shown that the proposed method reduces the data storage space while increasing the security level since the user on the receiving end cannot notice the compression process. Anwar et al [19] performed steganography using Lifting Wavelet Transform (LWT) and Dynamic Key technique. In the proposed technique, the secret message was encrypted with Advanced Encryption Standard (AES) and then it was embedded into the cover audio file with the LWT method. Sobin C.C. and Manikandan [20] mentioned the necessity of using a secure encryption scheme in the steganography process. It was suggested that the genetic algorithm reduces the distortions on the stego sound after the encrypted secret message is embedded in the selected bit-plane. Ying et al. proposed a different method from the classical steganography techniques in their work [21]. This method is Syndrom Trellis Code (STC) which is a near-optimal convolutional method for adaptive steganography. The Parity Check matrix and the custom adaptive parity matrix are designed to constrain the embedding location and change the designed distortion cost. They proposed a targeted intelligent optimization algorithm (named GOAS) that can adaptively generate the parity-check matrix in accordance with different audio covers. Experimental results showed that the proposed method outperforms state-of-the-art adaptive steganography with reduced insertion changes and improved sound quality while providing the capability against steganalysis. Rashmi proposed a method to increase the safety in steganography in his study. With this method, the message is encrypted before it is embedded in the audio file [22]. This encryption process reduces the perceptibility of the secret message by the human auditory system. RC4 encryption and 3DES encryption algorithms are used separately to encrypt original text. The cover object used in the proposed technique is an audio file. The sampling procedure is applied to an audio and then the appropriate bit is modified with cipher bit. The proposed algorithm highlights the use of cryptography reconciliated with proposed steganography technique by providing multilevel security to the confidential data. Ganvani et al. performed audio steganography with the modified LSB technique in their work [23]. They used RSA and ChaCha20 encryption techniques to increase data security with the

technique they developed. The encrypted messages were embedded in the audio file with the LSB technique and reached above-average quality values.

In this paper, we compare the most well-known and popular audio steganography techniques using 20 Turkish audio recordings sampled at 8 kHz. The utilized steganography methods are standard LSB method, Echo hiding, Wavelet coding, Spread Spectrum and Cepstrum based techniques. The main contribution of our study is to reveal the advantages and disadvantages of the popular audio steganography methods on Turkish audio cover files and to compare their performances using the same experimental setup. Thus, a unified performance evaluation comparison can be found in a single study.

2. Audio Steganography Methods

In steganography, embedding the hidden message into cover audio is generally performed on the bit-level. Hidden data, cover audio file, and stego key are all converted into binary form. Stego audio generated after embedding is obtained in pulse code modulation (PCM) format. The important issue after embedding the secret message in all steganographic techniques is that the receiver obtains the original file without any data loss. Although the process of extracting messages from the stego audio file for each technique varies based on the operation applied, the algorithm is basically the same.

In this study, different methods of audio steganography were applied to the same audio files and it was aimed to reveal the advantages of these methods over each other. The methods applied are least significant bit, echo hiding, wavelet coding, spread spectrum, and cepstrum.

2.1. Least Significant Bit

The least significant bit method (LSB) is a well-known and popular technique used in audio steganography. It is based on the principle that the bits of the message to be hidden with the last bit of each sample in the cover audio file are displaced in order. An example of the displacement principle of bits is shown in Figure 1. The number of bits to be changed can be set as the last two or the last three bits of the cover sound samples, but this operation degrades the quality of the stego audio. The factor that determines the embedding capacity in the LSB coding method is the sampling frequency. Sampling frequency determines the number of samples produced per second. As the sampling frequency increases, the number of samples in the signal increases. Thus, increasing the number of samples increases the embedding capacity, but increasing the sampling frequency without taking the frequency components of the signal into account will cause distortions in the signal.

In LSB, the channel capacity is 1 kilobit per kilohertz [3]. If the length of the message embedded in LSB increases, it is seen in the results that the SNR value decreases. As the length of the message to be embedded increases, the distortion of the signal increases, so the perceptual transparency of the signal decreases [4].

a 16-bit sample of cover audio	1 0 1 1 0 0 0 0 1 0 0 1 1 1 0 0
message bits	1 1 0 1 0 1 0 1 1 0
a 16-bit sample of stego audio	1 0 1 1 0 0 0 0 1 0 0 1 1 1 0 1

Figure 1. Obtaining 16-bit stego audio sample from LSB method from a 16-bit sample of cover sound.

If the text message is chosen to be long enough to use all the remaining bits, except the first 40 bytes in the audio file, the embedding capacity will be 100%. The sound created by replacing the message bits with the least significant bits in the cover sound file is stego audio. Stego audio is again of the PCM file type and the file size is the same as the original audio file. Since only the last bit of the sample is replaced by the message bit in the LSB method, the number of bits in the sample does not matter, so the LSB method is not affected by the frame size.,

2.2. Echo Hiding

The echo data hiding method uses the inability of the human ear to detect short-term echoes on the sound file [24]. Bender et al. [10], presented the idea of adding echo to the original audio file. In the echo hiding technique, the message to be hidden is generated as an echo in a separate signal and then the echo is embedded in the audio file. To add the echo to the audio file, the audio file is first divided into short frames with the specified frame size. In Figure 2, an example of the original signal divided into equal parts is given. The maximum number of bits that can be embedded in these parts is the same as the frame size. Echoes created after dividing into equal parts are added to the cover sound.

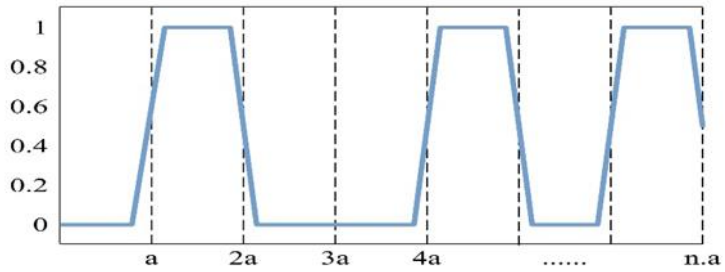

Figure 2. An example signal divided by equal time intervals.

The echo signal is obtained by the convolution of the original audio signal and the echo kernel. If the bit to be hidden is one, the binary one kernel is needed, whereas if the bit to be hidden is zero, the binary zero kernel is used.
The echo kernel is mathematically defined as,

$$h(t) = \delta(t) + \alpha\delta(\mathrm{t} - \Delta\mathrm{t}) \quad (1)$$

$\delta(t)$ is the unit impulse function, $\Delta(\mathrm{t})$ represents delay time and α is echo amplitude. In figure 2, echoed part 1 is created by using binary one kernel and binary one mixer in the parts where one bit will be added from the fragmented parts of the original signal. The same is true when the zero bit is added as an echo and in this case, an echoed part 2 is created using the binary zero kernel and the binary zero mixer. The flowchart of the method is shown in, Figure 3. Stego audio is obtained by combining the created echoed part 1 and echoed part 2.

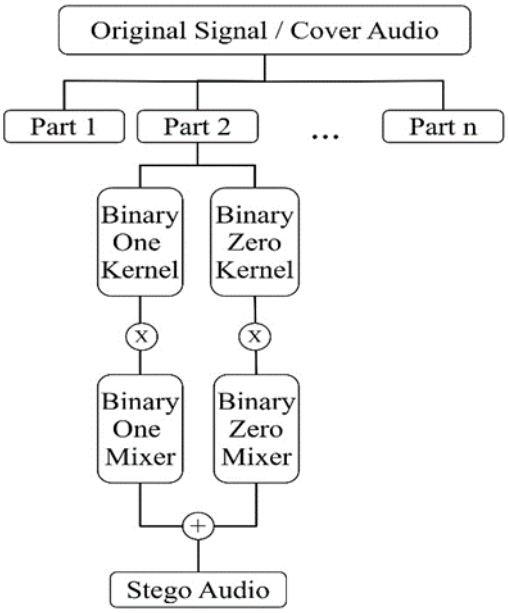

Figure 3. Stego signal generation with echo kernels and mixers.

There are three different echo kernels that allow echo data hiding in different ways. These are called single echo, bipolar echo and bipolar backward-forward echo. In this study, steganography was performed using the single echo hiding method [25]. Given an audio frame $x(t)$, the resulting echoed audio signal $y(t)$ is obtained.

$$y(t) = x(t) * \mathrm{h(t)} \tag{2}$$

$$y(t) = x(t) + \alpha \mathrm{x(t - \Delta t)}$$

This method allows high data transmission speed in parallel with the spread spectrum method [1]. Besides the data transmission speed, the quality of stego audio is also important. It is clear that stego voices with low SNR value at high data transmission speed will be meaningless in data hiding.

In order to successfully store the information, three parameters of the echo are changed, such as amplitude, decline rate, deviation from the original signal (delay time). All three parameters are set below the sensing threshold of the human ear. Accordingly, thanks to the parameters set to be lower than the threshold level, stego audio becomes difficult to distinguish from the original sound.

2.3. Wavelet Coding

Steganography can be performed by directly changing the samples of the signal without any conversion or by using the conversion coefficients in the frequency domain with appropriate domain transformations. However, direct changes on the signal samples may make the information to be hidden more vulnerable to pirates. One of the methods developed by taking the security concerns into account. Different methods have been proposed to eliminate these concerns. For example, mathematician Alfréd Haar first proposed the wavelet method in his 1909 thesis [6]. Then, in 1988, Ingrid Daubechies proposed a new wavelet transform method to create the compact support orthogonal wave [5]. In fact, wavelet transform is similar to Fourier Transform in terms of function definition. The most important point that highlights the wavelet transform from the Fourier Transform is that the wavelet transform generates the frequency

content of the signal depending on time [26]. In Figure 4, the basic structure diagram of the 3-level wavelet transform is given. In the figure, cA represents approximate coefficients while cD represents detail coefficients.

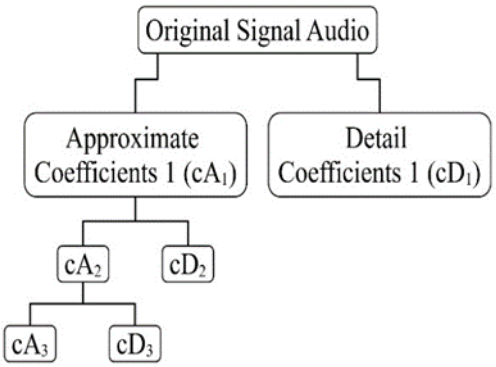

Figure 4. The basic structure diagram of the 3-level wavelet transform.

There are two different ways to apply the wavelet transform. These are discrete wavelet transform and continuous wavelet transform [7]. The main purpose of the discrete wavelet transform is to discretize the signal and sample each wavelet discretely [9]. One-dimensional discrete wavelet decomposition for a signal $f(t)$ is given in (3) [27].

$$W_f(j,k) = \int_{-\infty}^{\infty} f(t).\psi_{j,k}{}^*(t)dt\ ,\psi_{j,k}(t) = a_0{}^{-\frac{j}{2}}.\psi(a_0{}^{-j}t - b_0k) \tag{3}$$

$\psi(t)$ represents the main wavelet while j is the level of wavelet decomposition. a_0 and b_0 are constants, k is defined as the time shift factor and ψ * is the complex conjugate. In general, the form given in (4) when a_0 and b_0 values are set to 2 and 1, respectively is used.

$$W_f(j,k) = \int_{-\infty}^{\infty} f(t).\psi_{j,k}{}^*(t)dt\ ,\psi_{j,k}(t) = 2^{-\frac{j}{2}}.\psi(2^{-j}t - k) \tag{4}$$

Transform coefficients are calculated after discretization is completed. Both the approximate and detail coefficients of the one-dimensional signal should be calculated to modify the message bits. After all the details and approximate coefficients are calculated, the total signal can be recovered by [28]

$$f_j(t) = \sum_k W_f(j,k)\,\psi_{j,k}{}^*(2^{-j}t - k) \tag{5}$$

The embedding process begins with a 3-level 1-D wavelet decomposition. Approximate coefficients at level 3 are used for embedding. Depending on the message bits (one or zero), while the approximate coefficients in the 3rd level are reconstructed, no change is made in the detail coefficients. It is aimed to have low noticeability of the message by not making changes on the detail coefficients. The new third-level approximate coefficients in which the message bits are embedded are combined with the detail coefficients. Stego audio is obtained by using the inverse discrete wavelet transform to the row matrix obtained [29].

2.4. Spread Spectrum

The spread spectrum technique is one of the most commonly used techniques in steganography. The aim of the spread spectrum is to spread the information to be hidden to the frequency spectrum of the cover audio signal as much as possible. In fact, the process of placing hidden information is similar to the LSB technique. The difference of the spread spectrum method from LSB is that the spread spectrum makes the hidden information embedded in the frequency spectrum of the cover signal.

In order to obtain maximum performance after steganography, the cover audio signal must be analyzed before embedding. The psychoacoustic model was used for this purpose. Thanks to this model, it can be determined which samples of sound are more suitable for embedding messages. The Psychoacoustic model allows embedding messages in parts of the signal that are dominant in the spectrum. In this way, it prevents the message bit from being embedded in the insignificant parts of the signal and allows the quality of the resulting stego to increase [30]. After analysing the original signal in this technique, unlike other techniques, the message signal is modulated with the pseudo-random sequence (PRS) and the secret message signal is created. The pseudorandom noise sequence is a sequence of bits zero and one. Although they are called random, they are generated according to an algorithm. For each frame size in the study, pseudorandom sequences consisting of the same number of bits in each frame were generated. During this module, the spread process of the message is realized. By combining the scaled cover signal and the modulated message signal, stego audio is obtained. Figure 5 shows the general structure of embedding of the spread spectrum technique. The scaling factor, x, can be used to reduce the noticeability of the secret message generated and to control the amplitude of the message signal [24].

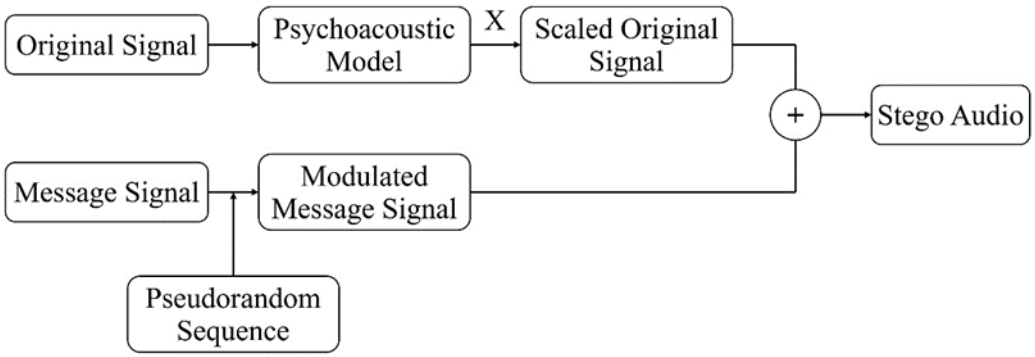

Figure 5. Secret message embedding procedure in spread spectrum.

After the spread spectrum technique is applied, the bandwidth of the stego signal is larger than the bandwidth of the original signal [10].

2.5. Cepstrum

This technique is also a data hiding technique in which embedding is applied over the frequency domain coefficients [20]. Cepstrum can actually be expressed as a result of a Fourier transform calculated with a logarithm [8]. For Cepstrum, the Fourier transform is applied to the cover signal first. The second operation is done by taking the logarithm of the Fourier outputs of the signal. After the logarithm process, the cepstrum is obtained with the inverse Fourier transform [32]. Before applying the inverse Fourier transform, the message signal is embedded in the cepstrum coefficients of the

cover signal [33]. In this method, the message bits are randomly placed in the cover sound to increase privacy and reduce noticeability. With this, the important components of the cover signal are not specifically selected for embedding.
In this process, any desired changes on the message signal can be performed before embedding it into cover audio. The cepstrum of a discrete time signal $x[n]$ is defined as [34]:

$$\hat{x}[n] = IDFT\ [\ \ln\{\ DFT\ (x[n])\ \}\ ] \tag{6}$$

Since the variance of the cepstrum coefficients is smaller, cepstrum coefficients are more suitable for embedding messages against possible threats.

3. Expermental Setup

Digital signals are used in almost all signal processing techniques. The digitized signals obtained by the sampling process with the determined sampling frequency become modifiable. For example, in the LSB method, replacing the last bits of the signal samples with the message signal bits is performed directly on the samples of the signal. Twenty Turkish audio recordings from ten male and ten female speakers are used in the experiment. The audio signals used in the experiments were selected from the TURTEL data set created by the National Research Institute of Electronics and Cryptology for Voice Communication and Coding Systems. The content of this database includes training records collected from 65 speakers and test records collected from 28 speakers. The ages of the speakers range from 17 to 55. The database consists of a total of 373 words and 15 sentences. The words have been chosen considering the distribution of trio-sounds in Turkish. The total number of samples in each audio signal used in the experiments varies from 20000 to 32000 according to the duration of the audio signals. The recordings were taken over three different types of phones. These are landline phones, mobile phones, and hands-free phones. After the sound information coming from the telephone line was recorded at 48 kHz, it was transferred to the computer with a sampling frequency of 8 kHz (Mono). The recording resolution of the audio files is 16 bits/sample. The audio signals are in the PCM format without any compression. In order to investigate the effect of the message length on the steganography techniques, the secret message length varying from 32 bits to 448 bits were used in the experiments.
Steganography methods that do not directly modify the samples of the cover signal first divides the signal into small segments. This process is necessary especially for methods that will use the frequency spectrum. This equal divide operation is a kind of segmentation. The total number of bits that can be embedded into each frame depends on the frame size. In the experiments, audio frames consisting of 256 samples are used. In the wavelet method, experiments were executed using the db4 wavelet transform function. In the part about the Wavelet transform of the study, firstly, the time-dependent function of the signal in the frequency domain was obtained with the approximate and detail coefficients using the Wavelet transform. The average signal-to-noise ratio (SNR) metric is used for the performance evaluation in the experiments. Given a pair of audio signals $x[n]$ and $y[n]$ where $x[n]$ represents the cover signal and $y[n]$ represents the stego signal obtained after embedding the message, the SNR (in dB) between $x[n]$ and $y[n]$ is computed by [4]

$$SNR = 10.\log_{10}\frac{\sum_n x^2(n)}{\sum_n [x(n)-y(n)]^2} \tag{7}$$

The SNR values computed between each cover audio file and its stego counterpart are then averaged over all recordings to obtain the average SNR value for the comparison.

4. Result and Discussion

The average SNR values obtained using different steganography techniques with varying message lengths are summarized in Fig. 6. From the figure, it is observed that as the message length increases, lower SNR values are obtained, in general, as expected. This is because increasing the message length eventually modifies the higher number of signal samples and therefore SNR value decreases. Among the five different steganography techniques, LSB method yields the highest SNR values. This is possibly because of the fact that, since the LSB method only modifies the last bit of each signal sample, it results in less distortion on the original signal.

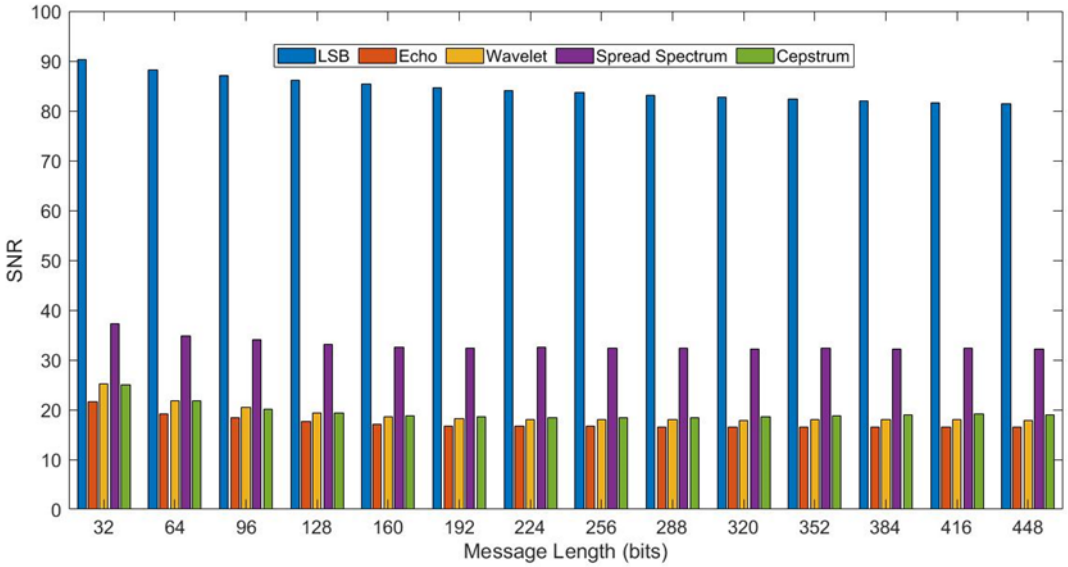

Figure 6. Comparison of the methods for SNR(dB) values at message lengths between 32-448 bits.

Although LSB technique is superior to other methods with a remarkable difference in table, the obtained SNR values are reasonable since it modifies only the least significant bit. When the message length is 32 bits, the average SNR value obtained in 20 audio files is 90 dB on average, in LSB. This value tends to decrease as the message length increases. When the message length increases from 32 bits to 448 bits, the SNR value reduces to around 80 dB. The second best method is the spread spectrum technique. This technique, which is realized by spreading the message bits over the frequency spectrum of the cover signal, is one of the important techniques for steganography. When the message length is 32 bits, An 37 dB average SNR value is obtained over all 20 audio recordings using the spread spectrum technique. Increasing the message length from 32 bits to 448 bits using the spread spectrum method yields approximately 13% reduction on the average SNR value where the average SNR reduces from 37 dB to 32 dB. Comparing the average SNR values obtained using the maximum and minimum message lengths, it is observed that the minimum performance gap between these two extreme points is obtained using the spread spectrum method. While using this technique, the cover sound signal is decomposed by obtaining the detail and approximate coefficients up to the third level. When the message length is 32 bits, the average SNR value obtained is around 24 dB. However, the average SNR value drops to around 19 dB when a 448-bit message is embedded in the cover audio file. Interestingly, the average SNR values obtained using

the Wavelet and Cepstrum techniques show similar trends with a negligible difference for each message length. While it has an average SNR of 21 dB in 32-bit message length, its SNR value in 448-bit message length is around 16 dB.

From the results reported in Figure 6, it can be observed that the steganography technique and the message length both have a considerable impact on the resulting SNR values. Although the data hiding method has more impact than the message length, these two parameters should be taken into account as a whole rather than considering each of them separately.

In Table 1, we summarize the average SNR values obtained using five steganography methods for six selected message lengths in order to provide a better comparison of different techniques. As we observed in Fig.6 as the message length increases the average SNR values decrease irrespective of steganography methods. The highest SNR values are obtained using the LSB technique.

Table 1. The average SNR values obtained using different audio steganography methods for some selected message length.

Message Length	Steganography Methods				
	LSB	Echo	Wavelet	S.Spectrum	Cepstrum
32	97.139	27.976	31.815	44.085	32.197
64	95.275	22.203	24.233	38.337	24.928
224	90.275	16.673	19.965	32.702	20.193
256	90.139	16.689	19.783	32.803	19.495
416	88.127	16.609	20.027	32.528	21.384
448	87.26	16.657	19.556	32.402	21.944

In Figure 7, spectrogram images of the same cover and stego signals (obtained using different techniques) are given for the comparison of the time-frequency graphs. The spectrogram images show that although there is no visible difference between the cover and stego signals, it is observed that steganography mostly affects the high-frequency region of the audio signals rather than low-frequency components.

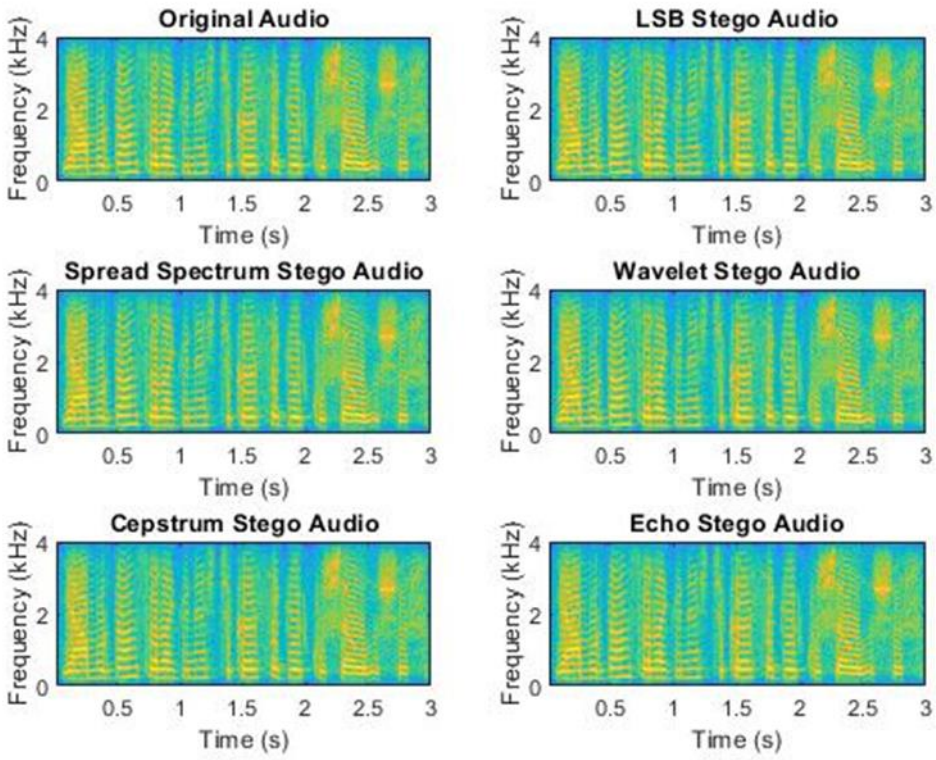

Figure 7. Spectrogram images of a selected cover and stego signals obtained using different techniques.

There are certain parameters to compare the superiority of steganography techniques with respect to each other. Although reliability is the first parameter among the others, it is not sufficient alone. Another feature sought while providing these parameters is how long a message can be embedded in the cover signal at once. The desired situation is to provide all of these parameters at the same time, but in practice, this is not possible. Although the LSB technique, which gives the highest quality value in comparison to other methods utilized in the experiments, is an easy technique to implement compared to the other four techniques, it is more vulnerable to possible attacks since direct changes are made on the signal samples.

The message length is the key parameter in the Spread Spectrum method. This is because in the Spread Spectrum technique, the cover signal is first divided into small frames (segments) and the total number of these frames determines the maximum number of message bits that can be embedded. Therefore, the length of the cover audio signal considerably affects the embedding capacity in the Spread Spectrum technique. However, the production of a PRS increases the security level of steganography. In the Wavelet Coding technique, the security level is high because the message hiding operation is performed on the frequency information of the signal. If message hiding is applied to the frequency components of the signal, the signal will become more protected against threats, but since it works at dominant frequencies, the possibility of increasing the visibility of the message by anyone should be considered. In the Cepstrum technique, after the Discrete Fourier Transform (DFT), logarithm and the Inverse Discrete Fourier Transform of the cover signal are taken, respectively, the cepstrum is obtained. Since the message bits are randomly embedded in the cepstrum coefficients, the security of the message is ensured. It is more difficult to apply than other steganographic methods. In the Echo Hiding technique, the embedding capacity is low, as the message signal is placed in the divided parts of the signal in the form of echo. If this capacity is exceeded, anyone listening to the audio signal may notice the

distortion in the sound, but if steganography is performed considering the embedding capacity and signal length, it is a highly reliable technique.

5. Conclusion

In this paper, we experimentally compared the performances of the five well-known audio steganography techniques on Turkish audio recordings for various hidden message lengths. Experimental results conducted on 20 Turkish audio recordings from 10 male and 10 female speakers showed that as the message length increases, the average SNR value decreases irrespective of the steganography technique, as expected. For example, while an SNR value of 90 dB was obtained when a message signal consisting of 32 bits is embedded the sing LSB technique, the SNR value reduces to 80 dB when a message signal composed of 448 bits is embedded. Similarly, SNR value reduces from approximately 38 dB to 34 dB when message length increases from 32 bits to 448 bits. Among the five audio steganography methods, thepopular and easy-to-implement LSB technique yields the highest average SNR value than the remaining four methods. This is reasonable because LSB only modifies the last bit of each audio signal sample thus resulting in stego signal indistinguishable from the original cover signal. The spread spectrum method which embeds the hidden message bits into the spectrum of the cover signal was found to be the second best method after the LSB technique. Interestingly Cepstrum and Wavelet techniques show similar performance and yield similar SNR values with a negligible difference for various hidden message lengths. Finally, the Echo hiding method which hides the hidden message bits with the echo kernel was found be inferior to the remaining four methods independent of the message length. Spectrographic analysis of the five different steganography methods showed that hiding a secret message to an audio signal affects the high frequency components rather than the low-frequency region. Further more detailed analysis of the five techniques utilized in this study can be conducted by steganalysis which aims at detecting whether an audio signal includes a hidden message or not which is left as future work.

References

[1] Warkentin M., Schmidt M., Bekkering E. (2007). Steganography and steganalysis, Intellectual Property Protection for Multimedia Information Technology, IGI Global, 374-380, 9781599047621.

[2] Guojie, H., Zhengjin, F., & Ruiling, M. (2003). Chosen ciphertext attack on chaos communication based on chaotic synchronization. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 50(2), 275-279.

[3] Dutta, P., Bhattacharyya, D., Kim, T. (2009). Data Hiding in Audio Signal: A Review, International Journal of Database Theory and Application, 2(2):1-8.

[4] Cvejic, N., Seppänen, T. (2003). Increasing the capacity of LSB-based audio steganography, IEEE Workshop on Multimedia Signal Processing, 336-338. doi: 10.1109/MMSP.2002.1203314.

[5] Misshra S., Yadav, V.K., Trivedi, M.C., Shrimali, T. (2018). Audio Steganography Techniques: A Survey, Advances in Computer and Computational Sciences, Advances in Intelligent Systems and Computing 581-589. doi: 10.1007/978-981-10-3773-3_56.

[6] Kaur, N., Behal, S. (2014). Audio Steganography Techniques-A Survey, Int. Journal of Engineering Research

and Applications 94-100. doi: 10.14445/22315381/IJETT-V11P276.

[7] Bhavana, D., Tej, D.R., Puneeth, M., Harini, A., Shankar, Y.M., Madhuri, T.S. (2015). Steganography Using Matlab, International Journal of Advanced Engineering Research and Science (IJAERS) 19-24. doi:10.22161/ijaers.

[8] Wakiyama, M., Hidaka, Y., Nozaki, K. (2010). 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing.

[9] Djebbar, F., Ayad, B., Meraim, K.A., Hamam, H. (2012). Comparative study of digital audio steganography techniques, EURASIP Journal on Audio, Speech, and Music Processing doi.org/10.1186/1687-4722-2012-25.

[10] Bender, W., Gruhl, D., Morimoto, N., Lu, A. (1996). Techniques for data hiding, IBM Systems Journal 313-336. doi: 10.1147/sj.353.0313.

[11] Shah, P., Choudhari, P., Sivaraman, S. (2008). Adaptive Wavelet Packet Based Audio Steganography using Data History, 2008 IEEE Region 10 and the Third international Conference on Industrial and Information Systems, doi: 10.1109/ICIINFS.2008.4798397.

[12] Asad, M., Gilani, J., Khalid, A. (2011). An enhanced least significant bit modification technique for audio steganography, International Conference on Computer Networks and Information Technology, doi: 10.1109/ICCNIT.2011.6020921.

[13] Gupta, N., Sharma, N. (2014). Dwt and Lsb Based Audio Steganography, 2014 International Conference on Reliability Optimization and Information Technology (ICROIT), doi: 10.1109/ICROIT.2014.6798368.

[14] Bilal, I., Roj, M.S., Kumar, R., Mishra, P.K. (2014). Recent advancement in audio steganography, 2014 International Conference on Parallel, Distributed and Grid Computing, doi: 10.1109/PDGC.2014.7030779.

[15] Kartheeswaran, T., Senthooran, V., Pemadasa, T.D.D.L. (2015). Multi agent based audio steganography, 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), doi: 10.1109/ICCIC.2015.7435706.

[16] Lindawati., Siburian, R. (2017). Steganography implementation on android smartphone using the LSB (least significant bit) to MP3 and WAV audio, 2017 3rd International Conference on Wireless and Telematics (ICWT), doi: 10.1109/ICWT.2017.8284161.

[17] Rajput, S.P., Adhiya K.P., Patnaik, G.K. (2017). An Efficient Audio Steganography Technique to Hide Text in Audio, 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), doi: 10.1109/ICCUBEA.2017.8463948.

[18] Teotia, S., Srivastava, P. (2018). 2018 International Conference on Communication and Signal Processing (ICCSP), doi: 10.1109/ICCSP.2018.8524182.

[19] Anwar, M., Sarosa, M., Rohadi, E. (2019). Audio Steganography Using Lifting Wavelet Transform and Dynamic Key, 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), doi: 10.1109/ICAIIT.2019.8834579.

[20] C. C. Sobin., Manikandan, V.M. (2019). A Secure Audio Steganography Scheme using Genetic Algorithm, 2019 Fifth International Conference on Image Information Processing (ICIIP), doi: 10.1109/ICIIP47207.2019.8985689.

[21] Ying, K., Wang, R., Lin, Y., & Yan, D. (2021). Adaptive Audio Steganography Based on Improved Syndrome-Trellis Codes. IEEE Access, 9, 11705-11715.

[22] Rashmi, N. M. (2020, January). Analysis of Audio Steganography combined with Cryptography for RC4 and 3DES Encryption. In 2020 Fourth International Conference on Inventive Systems and Control (ICISC) (pp. 210-214). IEEE.

[23] Ganwani, P., Gupta, L., Jain, C., Kulkarni, R., & Chaudhari, S. (2021, July). LSB Based Audio Steganography using RSA and ChaCha20 Encryption. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

[24] Yıldıray, Y. (2018) Sayısal imgelerde güvenli veri gizlemeye yönelik uygun maske belirleme, Master Thesis, Fırat Üniversitesi, Elazığ, Turkey, 50 p.

[25] Tokur, Y., Erçelebi, E. (2005). Spread Spectrum Audıo Watermarkıng Scheme Based On Psychoacoustıc Model, The International Conference on Electrical and Electronics Engineering(ELECO 2005), 2:143-147.

[26] Rabiner, L., & Schafer, R. (2010). Theory and applications of digital speech processing. Prentice Hall Press.

[27] Lin, Y., Abdulla, W. H., M. (2014). Audio Watermark: A Comprehensive Foundation Using MATLAB, Springer, 1-183.

[28] Haddadi, R., Abdelmounim, E., Belaguid, A. (2014). Discrete Wavelet Transform Based Algorithm for Recognition of QRS Complexes, World of Computer Science and Information Technology Journal, 4 (9), 127-132. doi:10.1109/ICMCS.2014.6911261.

[29] Hemalatha, S., Acharyaa, U. D., Renuka, A. (2015). Wavelet transform based steganography technique to hide audio signals in image, Procedia Computer Science, 47, 272-281. doi.org/10.1016/j.procs.2015.03.207.

[30] Chun-Lin, L. (2010). A tutorial of the wavelet transform. NTUEE, Taiwan.

[31] Akansu, A. N., Haddad, R. A., Haddad, P. A., & Haddad, P. R. (2001). Multiresolution signal decomposition: transforms, subbands, and wavelets. Academic press.

[32] Gopalan, K. (2005). Audio Steganography by Cepstrum Modification, IEE International Conference: Acoustic, Speech, and Signal Processing, 5:481-484.

[33] Tekeli, K., Asiyan, R. (2009). A Comparison of Echo Hiding Methods, The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 1:397-403,

[34] Retrieved from https://dergipark.org.tr/tr/pub/epstem/issue/31865/365048.

[35] Alp, H., Akıncı, T. Ç., Albora, M. (2008). Comparıson Of Fourier And Wavelet Transforms In Geophysical Applicatıons, *Journal Of Engineerıng Sciences,* 14 (1), 67-76.