Article

Wavelet-Domain Information-Hiding Technology with High-Quality Audio Signals on MEMS Sensors

Ming Zhao [1], Shuo-Tsung Chen [2],* and Shu-Yi Tu [3],*

1   School of Computer Science, Yangtze University, Jingzhou 434025, China
2   Department of Applied Mathematics, Tunghai University, Taichung City 407224, Taiwan
3   Department of Mathematics, University of Michigan, Flint, MI 48502, USA
*   Correspondence: shough34@yahoo.com.tw (S.-T.C.); sytu@umich.edu (S.-Y.T.)

Abstract: Due to the rapid development of sensor technology and the popularity of the Internet, not only has the amount of digital information transmission skyrocketed, but also its acquisition and dissemination has become easier. The study mainly investigates audio security issues with data compression for private data transmission on the Internet or MEMS (micro-electro-mechanical systems) audio sensor digital microphones. Imperceptibility, embedding capacity, and robustness are three main requirements for audio information-hiding techniques. To achieve the three main requirements, this study proposes a high-quality audio information-hiding technology in the wavelet domain. Due to the fact that wavelet domain provides a useful and robust platform for audio information hiding, this study applies multi-coefficients of discrete wavelet transform (DWT) to hide information. By considering a good, imperceptible concealment, we combine signal-to-noise ratio (SNR) with quantization embedding for these coefficients in a mathematical model. Moreover, amplitude-thresholding compression technology is combined in this model. Finally, the matrix-type Lagrange principle plays an essential role in solving the model so as to reduce the carrying capacity of network transmission while protecting personal copyright or private information. Based on the experimental results, we nearly maintained the original quality of the embedded audio by optimization of signal-to-noise ratio (SNR). Moreover, the proposed method has good robustness against common attacks.

Keywords: sensor; digital information; MEMS; DWT; optimization; compression

1. Introduction

The uses of digital information transmission in Internet applications, artificial intelligence, and data sensing [1–10] are more frequent. In many cases, without permission from the legal owner, the digital information is often stolen, copied, or even turned into profit by criminal individuals. In general, an audio information-hiding technique should possess three properties: make the piece of hidden information imperceptible in the embedded audio, provide a signal-to-noise ratio (SNR) of 20 dB or more, and maintain the embedding capacity of at least 20 bps (bits per second) [11,12]. Moreover, hidden information is resistant to most attacks, which include re-sampling, MP3 compression, filtering, amplitude modification, time scaling, and so on [12–15].

Audio information-hiding techniques are classified according to their domain. These algorithms are categorized as time-domain techniques and transform-domain techniques. Discrete wavelet transform (DWT) is one practical transform domain for hiding audio information [13–27]. In the literature, several earlier procedures embedded watermarks into DWT low-frequency coefficients using the quantization-based technique so that they could obtain adaptive performance [15,20,24]. Chen et al. [15,24] proposed an optimization quantization approach to fixed-weighting DWT coefficients to gain high-quality modified audio and high robustness against many common attacks. Li et al. [27] proposed a new

audio watermarking technique. They performed the norm ratio on fixed-scaling DWT coefficient quantization without considering the signal compression in the implementation. In addition, the quality of modified audio worsens with weighting variation, and hidden information is inadequately robust to time-scaling attacks.

This study proposes an optimization model to integrate optimization-based signal steganography [24] with threshold-based compression in the wavelet domain. Firstly, we utilized binary digits to store data and represent information. We modified the signal-to-noise ratio (SNR) and the amplitude-quantization rules in the wavelet domain as performance index and constraints. At the same time, to reduce the amount of the embedded audio data signal, we also employed threshold-based compression technology in the constrictions. Then, we obtained an optimization model that enhanced the audio quality in the information-hiding and compression processes. Secondly, the optimization model was solved by the matrix-type Lagrange principle and graphic illustration Accordingly, we performed information hiding and data compression on each audio signal for private information transmission on the Internet or MEMS (micro-electro-mechanical systems) audio sensor digital microphones. On the other end of the transmission process, the hidden information was extracted smoothly without either the original audio or the recovery of the compressed audio signal adopting a cubic spline. To demonstrate the quality of the proposed performance, we measured the appropriate threshold ε and embedding strength Q in our experiment. The proposed algorithm reduces the amount of carried network transmission but preserves the original audio signals and protects personal privacy.

The rest of this study is as follows: Section 2 presents the proposed method and introduces the embedding technique, optimization model, and compressions. The illustrations of the optimization model, presentation of the recovery method, and the extraction technique are in Section 3. Section 4 contains discissions of experimental results, and some remarks and conclusions are in Section 5.

2. Proposed Method

This section introduces the embedding technique, the extraction technique, and the compression of the proposed method. Figure 1 shows the block diagram of the proposed algorithm; further detailed introduction will appear in Sections 2.1 and 2.2.
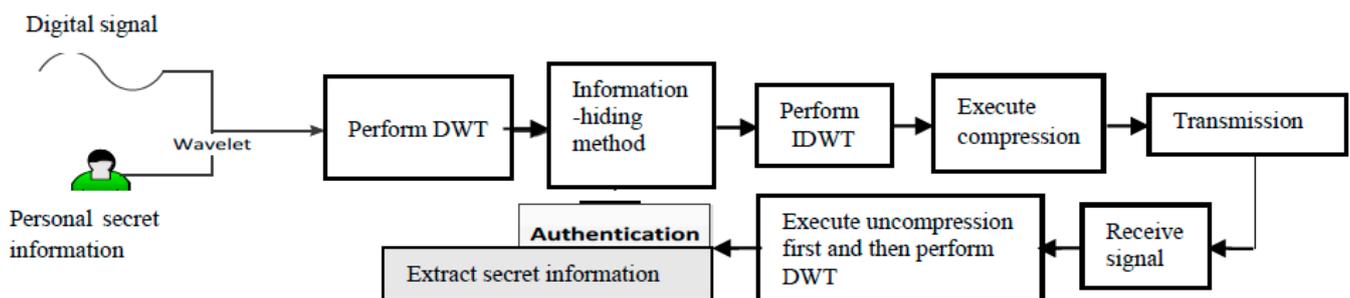
Figure 1. The block diagram of the proposed algorithm.

2.1. Embedding Technique

To embed the private information into the lowest DWT coefficients, we implemented DWT using the single prototype function $\psi(x)$. This function is regulated by a scaling parameter and a shift parameter [28,29]. The discrete normalized scaling and wavelet basis function was defined as

$$\varphi_{i,n}(t) = 2^{\frac{i}{2}} h_i(2^i t - n) \tag{1}$$

$$\psi_{i,n}(t) = 2^{\frac{i}{2}} g_i(2^i t - n) \tag{2}$$

where i and n are the dilation and translation parameters, and h_i and g_i denote low-pass and high-pass filters. Orthogonal wavelet basis functions provide simple calculation of

coefficient expansion and easily express audio signals $S(t) \in L^2(\mathrm{R})$ as a series expansion of orthogonal scaling functions and wavelets. Throughout this study, we used the host digital audio signal $S(n)$, $n \in \mathrm{N}$, to denote samples of the original audio signal at the nth sample time, and each piece of audio signal was cut into segments on which DWT was performed. As a result, the signal-to-noise ratio (SNR)

$SNR = -10 \log_{10} \left(\|\widetilde{S}(n) - S(n)\|_2{}^2 / \|S(n)\|_2{}^2 \right)$ can be rewritten as

$$SNR = -10 \log_{10} \left(\|\widetilde{X}_n - X_n\|_2{}^2 / \|X_n\|_2{}^2 \right) \tag{3}$$

where $\widetilde{S}(n)$ is the modified digital audio signal and the vector form $\widetilde{X}_n = \begin{bmatrix} |\widetilde{x}_1| & |\widetilde{x}_2| & \cdots & |\widetilde{x}_n| \end{bmatrix}^T$ consists of the n unknown absolute values of DWT coefficients with respect to the original DWT coefficient vector $X_n = \begin{bmatrix} |x_1| & |x_2| & \cdots & |x_n| \end{bmatrix}^T$ in each segment.

For convenience, the secret information is usually stored as a binary sequence. To embed the binary bit "$1 \in \mathrm{B}$" or "$0 \in \mathrm{B}$" as shown in Figure 1, we performed DWT and then determined n unknown values of DWT coefficients, $\widetilde{x}_1, \widetilde{x}_2, \cdots, \widetilde{x}_n$. Accordingly, an optimization-based model for embedding the binary bit was proposed as follows.

We determined the vector $\widetilde{X}_n$ such that the $SNR = -10 \log_{10} \left(\|\widetilde{X}_n - X_n\|_2{}^2 / \|X_n\|_2{}^2 \right)$ is maximized. Due to the fact that all logarithmic functions are one-to-one, that is, for all x and y in the domain of logarithmic function, if $\log_{10} x = \log_{10} y$, then $x = y$. We defined a performance index of the form $\|\widetilde{X}_n - X_n\|^2 / \|X_n\|^2$ so that the binary sequence with binary bit "$1 \in \mathrm{B}$" or "$0 \in \mathrm{B}$" can be embedded by the proposed optimization model described below:

- If the bit "$1 \in \mathrm{B}$" is embedded into X_n, then $\sum\limits_{i=1}^{n} |x_i|$ is quantized by

$$minimize \frac{\left(\widetilde{X}_n - X_n\right)^T \left(\widetilde{X}_n - X_n\right)}{X_n{}^T X_n} \tag{4a}$$

$$subjected\ to\ (a)\ A\widetilde{X}_n = \left\lfloor \frac{\sum\limits_{i=1}^{n} |x_i|}{Q} \right\rfloor Q + \frac{3}{4}Q \tag{4b}$$

$$\text{(b) Compression constraint;} \tag{4c}$$

- If the bit "$0 \in \mathrm{B}$" is embedded into X_n, then $\sum\limits_{i=1}^{n} |x_i|$ is quantized by

$$minimize \frac{\left(\widetilde{X}_n - X_n\right)^T \left(\widetilde{X}_n - X_n\right)}{X_n{}^T X_n} \tag{5a}$$

$$subjected\ to\ (a)\ A\widetilde{X}_n = \left\lfloor \frac{\sum\limits_{i=1}^{n} |x_i|}{Q} \right\rfloor Q + \frac{1}{4}Q \tag{5b}$$

$$\text{(b) Compression constraint;} \tag{5c}$$

where $\lfloor \rfloor$ is the floor function; Q is the quantization size or embedding strength which is adopted as the secret key K; the compression constraint is described in Equations (6)–(8) in Section 2.2.

2.2. Compression Constraint

Solving the optimization models (4) and (5), the watermarked audio signal $\overline{S}$ with optimal SNR is obtained after applying the IDWT. To reduce the amount of data when trans-

mitting on the Internet, we compressed the embedded audio signal $\bar{s}$ using the threshold compression method formulated as follows:

$$\hat{s}_0 = \bar{s}_0, \hat{s}_N = \bar{s}_N \tag{6}$$

$$\hat{s}_i = \begin{cases} \phi & \text{if} |\bar{s}_{i-1} - \bar{s}_{i+1}| < \varepsilon \\ \bar{s}_i & \text{otherwise} \end{cases}, i = \{1, \cdots N - 1\} \tag{7}$$

where ε represents the threshold.

To recover the signal $\{\bar{s}_i\}_{i=0}^{N}$ from the compressed signal $\{\hat{s}_i\}_{i=0}^{N}$, we used the cubic function, which is formulated as $f_i(t) = \alpha_i + \beta_i(t - t_i) + \gamma_i(t - t_i)^2 + \eta_i(t - t_i)^3$. We found the N cloud-gauge line collection of functions $\{f_i(t) | i = 1, \ldots, N\}$ to describe the entire set of data, where $f_i(t)$ must satisfy

$$f_i(t_i) = \hat{s}_i = f_{i-1}(t_i), f'_i(t_i) = f'_{i-1}(t_i), f''_i(t_i) = f''_{i-1}(t_i), f''_1(t) = f''_N(t) = 0 \tag{8}$$

To ensure the recovery quality, we adjusted the compression threshold ε while considering Q to better fit the optimal SNR.

3. Proposed Optimization Solution in Embedding and Extraction Method

In this section, we solve the optimization problem described in models (4) and (5) in two steps. Since the optimization problems (4) and (5) are similar, we first solve (4) and then apply the optimal solution to (5) using the same method.

3.1. First Step in Finding the Optimal Solution

Applying Theorems A1 and A2 introduced in Appendix A, the Lagrange multiplier λ is utilized to combine (4a) and (4b) into a function F without any constraints,

$$F(\widetilde{X}_n, \lambda) = \frac{(\widetilde{X}_n - X_n)^T (\widetilde{X}_n - X_n)}{X_n{}^T X_n} + \lambda \left\{ A \widetilde{X}_n - u_1 \right\}, \tag{9}$$

where setting $\left\lfloor \dfrac{\sum_{i=1}^{n} |x_i|}{Q} \right\rfloor Q + \frac{3}{4}Q = u_1$. The necessary conditions for minimizing $F(\widetilde{X}_n, \lambda)$ are

$$\frac{\partial F}{\partial \widetilde{X}_n} = 2(\widetilde{X}_n - X_n) + A^T \lambda X_n{}^T X_n = 0 \tag{10a}$$

$$\frac{\partial F}{\partial \lambda} = A \widetilde{X}_n - u_1 = 0 \tag{10b}$$

Multiplying (10a) by A to observe that

$$2(A \widetilde{X}_n - A X_n) + A A^T \lambda X_n{}^T X_n = 0 \tag{11}$$

Since $A \widetilde{X}_n = u_1$, Equation (11) can be rewritten as

$$u_1 - A X_n + \frac{1}{2} A A^T \lambda = 0 \tag{12}$$

Hence, the optimal solution of λ is

$$\lambda^* = 2(A A^T)^{-1}[A X_n - u_1] \tag{13}$$

Moreover, by substituting (13) into (10a), the optimal DWT coefficients are

$$\begin{aligned} \widetilde{X}_n{}^* &= X_n - \frac{1}{2} A^T \lambda^* \\ &= X_n - A^T (A A^T)^{-1}[A X_n - u_1] \end{aligned} \tag{14}$$

where the superscript $*$ denotes the optimal result with respect to the corresponding variable.

3.2. Audio Recovery and Information Extraction

To extract the hidden confidential data, we first recover the signal $\{\bar{s}_i\}_{i=0}^{N}$ from the compressed signal $\{\hat{s}_i\}_{i=0}^{N}$ using the cubic function, which is formulated as $f_i(t) = \alpha_i + \beta_i(t - t_i) + \gamma_i(t - t_i)^2 + \eta_i(t - t_i)^3$. We found that the N cloud gauge line collection of functions $\{f_i(t) | i = 1, \ldots, N\}$ to describe the entire set of data, where $f_i(t)$ must satisfy $f_i(t_i) = \hat{s}_i = f_{i-1}(t_i), f'_i(t_i) = f'_{i-1}(t_i), f''_i(t_i) = f''_{i-1}(t_i), f''_1(t) = f''_N(t) = 0$.

Next, we extract the hidden information from the DWT coefficients $\{\bar{c}_i\}_{i=0}^{N}$ of the recovered audio signal $\{\bar{s}_i\}_{i=0}^{N}$ according to the following steps:

Split the test audio into segments and perform DWT on each segment. If $\widetilde{X}^*_n = \{|\widetilde{x}^*_1|, |\widetilde{x}^*_2|, \cdots, |\widetilde{x}^*_n|\}$ presents n consecutive DWT lowest-frequency coefficients, the binary sequence is extracted from $\widetilde{X}^*_n$ by the following proposed extraction technique:

- If

$$\sum_{i=1}^{n} |\widetilde{x}_i^*| - \left\lfloor \frac{\sum_{i=1}^{n} |\widetilde{x}_i^*|}{Q} \right\rfloor Q \geq \frac{Q}{2}, \tag{15a}$$

then the extracted value is 1.

- If

$$\sum_{i=1}^{n} |\widetilde{x}_i^*| - \left\lfloor \frac{\sum_{i=1}^{n} |\widetilde{x}_i^*|}{Q} \right\rfloor Q < \frac{Q}{2}, \tag{15b}$$

then the extracted value is 0.

Finally, the hidden information is recovered from the binary sequence. In addition, to closely monitor the accuracy of the extracted private data, its ratio of bit errors (BER) is measured to check if an attack occured. The BER is usually expressed as a percentage and can be formulated as BER $= (B_{error} / B_{total}) \times 100\%$, where B_{error} and B_{total} denote the numbers of error binary bits and total binary bits during a tested period.

3.3. Application Scenarios of Our Proposal

The model and techniques proposed in this study combine information hiding and data compression of audio signals. During network transmission, if the amount of data (including the hidden private information) is large and the network speed is slow, the compression ratio can be increased to save transmission time; on the contrary, if the amount of data is small or the network speed is fast, one can just perform information hiding without performing data compression to improve the accuracy of the data transmission. This is the biggest difference between the proposed method and other methods.

4. Experimental Results

This section presents experimental results from testing the proposed algorithm. Without loss of generality, we investigate various forms of audio signals, such as love songs, symphonies, and dance and folkloric music. Ten songs per audio were averaged to evaluate the performance of the proposed method. These mono-type signals achieved a sampling rate of 44.1 kHz, which means there were 512,000 samples in each piece of selected information. They all came with a bit depth of 16 bits and 11.6 s in length. In the embedding procedure, each audio signal with 512,000 samples was initially cut into four segments of equal length; an 8-level discrete wavelet transform was performed on each evenly cut piece. This process ensured each piece of data had the total number of lowest-frequency

coefficients $512000/(4 \cdot 2^8) = 2000$. The values for Q are 13,000 and 26,000 for $n = 2$ and 4, respectively. To show a better comparison, we also implemented the two methods listed in references [24,27]; the experimental results are shown to compare with our algorithm.

4.1. Embedding Capacity and Averaged SNR

As listed in Table 1, the embedding capacities for $n = 2$ and 4 are 1000 and 500 bits, which satisfy the IFPI requirement—providing at least 20 bps (200 bits/10 s) embedding capacity. However, if the group size is greater than 16, this requirement is violated. Since we aim to present an optimization model for DWT multi-coefficients in this study, the resulting SNR of the proposed method clearly shows our SNR is much better than those SNRs using the methods in [24,27].

Table 1. Embedding capacity and SNR.

	Number of Consecutive Coefficients in DWT Level 8	Embedding Capacity (bits/11.6 s)	Averaged SNR (dB)			
			Dance	Love Song	Folklore	Symphony
Reference [24]	$n = 2$	1000	35.8	33.4	27.9	26.3
	$n = 4$	500	37.7	33.5	28.6	26.2
Reference [27]	$n = 2$	1000	24.3	25.4	23.2	22.3
	$n = 4$	500	24.1	26.0	23.6	22.9
Proposed Method	$n = 2$	1000	38.3	35.6	28.7	27.5
	$n = 4$	500	37.1	41.3	34.5	33.2

4.2. Robustness Measurement

We used five types of common attacks: re-sampling, low-pass filtering, amplitude scaling, time scaling, and MP3 compression, to evaluate how robust the proposed algorithm is. The performance quality is measured according to the averaged BER and its standard deviation (SD). A detailed discussion is illustrated below.

(1) Re-sampling: In the re-sampling process, the sampling rate of an audio signal can be increased (up-sample) or decreased (down-sample) in three stages: (i) down-sample, (ii) interpolation, and (iii) up-sample. We down-sampled the sampling rate of embedded audios from 44.1 kHz to 22.05 kHz, then up-sampled them from 22.05 kHz back to 44.1 kHz with a linear interpolation filter. A similar approach allowed the sampling rates to change from 44.1 kHz to 11.025 kHz and 8 kHz and regain the original rate of 44.1 kHz. Table 2 shows the BER of testing re-sampling on audio signals. One can see that when the re-sampling rate is 8 kHz, the proposed embedding method has lower BER than those from implementations in [24,27]. In those cases when the re-sampling rates are 22.05 kHz and 11.025 kHz, the proposed method shows comparable robustness.

(2) Low-pass filtering: Table 3 presents the BER while testing low-pass filters with cutoff frequencies of 3 kHz and 5 kHz. The BER results show that models in [24,27] have slightly higher robustness. Since both references [24,27] also adopted quantization-based embedding technique, the BER evaluation of the proposed method gives extremely similar results to theirs during the process of low-pass filtering.

(3) MPEG Audio Layer-3 (MP3) compression: Table 4 shows the BER from testing MP3 compression with different bit rates on the embedded audio data. The BER values reflect that the proposed model has similar robustness to that in references [24,27].

(4) Amplitude scaling: Since the amplitude-scaling attack usually results in saturation, in this study, we selected four distinct values for the amplitude-scaling factor: 0.5, 0.8, 1.1, and 1.2. The experimental results in Table 5 confirm that the proposed algorithm is much more robust than the methods in references [24,27].

(5) Time scaling: Table 6 lists the BER from testing time-scaling attacks with a ±2% and ±5% range. The BER results show that our method has comparable robustness to those in references [24,27].

Based on the experimental outcomes and aforementioned discussions, the proposed method generally achieves high SNR and is almost zero-error against the amplitude-scaling attacks. However, it shows slightly lower robustness against low-pass filtering attacks and poor robustness against time-scaling attacks.

Table 2. BER of Testing Re-sampling.

Audio Type			Dance			Folklore			Love Song			Symphony		
Re-Sampling Rate (kHz)			22.05	11.025	8	22.05	11.025	8	22.05	11.025	8	22.05	11.025	8
Reference [24]	$n=2$	mean	8.32	13.31	14.01	0.74	4.22	4.36	5.74	3.08	2.39	0.78	4.52	4.76
		SD	0.40	0.43	0.41	0.23	0.28	0.26	0.16	0.15	0.16	0.16	0.26	0.28
	$n=4$	mean	2.36	8.01	8.01	0.20	1.24	1.26	0.72	1.05	1.06	0.32	1.29	1.29
		SD	0.25	0.38	0.36	0.18	0.21	0.19	0.10	0.12	0.12	0.13	0.21	0.21
Reference [27]	$n=2$	mean	9.14	15.26	15.31	0.74	4.22	4.36	5.74	3.08	2.39	0.78	4.52	4.76
		SD	0.41	0.43	0.42	0.19	0.24	0.27	0.17	0.14	0.13	0.14	0.27	0.25
	$n=4$	mean	2.17	8.03	8.04	0.23	1.21	1.31	0.62	1.21	1.02	0.35	1.27	1.28
		SD	0.21	0.39	0.37	0.15	0.20	0.19	0.11	0.12	0.11	0.13	0.22	0.21
Proposed method	$n=2$	mean	8.25	14.42	0.87	0.82	4.65	4.35	4.87	3.29	0.68	0.66	1.28	1.28
		SD	0.39	0.4	0.16	0.18	0.22	0.21	0.15	0.16	0.09	0.14	0.21	0.21
	$n=4$	mean	2.1	8.16	0.26	0.23	1.42	1.25	1.34	1.45	0.57	0.53	1.24	1.23
		SD	0.23	0.38	0.14	0.16	0.19	0.15	0.13	0.11	0.08	0.12	0.20	0.19

Table 3. BER of Testing Low-pass Filtering.

Audio Type			Love Song		Symphony		Dance		Folklore	
Cutoff Frequency			3 kHz	5 kHz	3 kHz	5 kHz	3 kHz	5 kHz	3 kHz	5 kHz
Reference [24]	$n=2$	mean	24.18	25.82	27.58	8.68	33.62	21.52	33.62	15.72
		SD	0.28	0.22	0.29	0.20	0.35	0.29	0.34	0.21
	$n=4$	mean	23.82	23.48	27.55	8.41	33.25	21.28	33.02	13.84
		SD	0.27	0.19	0.28	0.20	0.36	0.27	0.34	0.19
Reference [27]	$n=2$	mean	26.18	25.82	27.53	8.68	33.62	21.52	33.62	15.72
		SD	0.29	0.21	0.28	0.19	0.35	0.27	0.33	0.19
	$n=4$	mean	25.82	24.81	27.54	8.41	33.02	20.87	33.02	11.84
		SD	0.29	0.21	0.25	0.21	0.34	0.28	0.33	0.17
Proposed method	$n=2$	mean	22.84	23.63	27.85	8.38	32.28	20.03	31.82	13.32
		SD	0.27	0.19	0.27	0.18	0.35	0.26	0.29	0.15
	$n=4$	mean	21.42	23.63	27.54	8.25	30.39	20.02	32.50	13.15
		SD	0.25	0.18	0.24	0.19	0.33	0.25	0.30	0.16

Table 4. BER of Testing MP3 compression.

Audio Type			Love Song				Symphony				Dance				Folklore			
Bit Rate (kbps)			**128**	**112**	**96**	**80**	**128**	**112**	**96**	**80**	**128**	**112**	**96**	**80**	**128**	**112**	**96**	**80**
Reference [24]	n = 2	mean	0.16	1.38	2.09	2.72	0.35	1.45	2.44	3.17	0.74	2.11	2.12	3.02	0.36	1.48	2.42	3.12
		SD	0.11	0.11	0.13	0.15	0.12	0.13	0.14	0.17	0.15	0.18	0.18	0.21	0.12	0.15	0.14	0.16
	n = 4	mean	0.09	0.11	1.41	2.53	0.14	0.15	2.29	3.84	0.11	0.15	1.02	3.0	0.15	0.15	2.40	3.93
		SD	0.10	0.09	0.12	0.15	0.10	0.10	0.13	0.16	0.12	0.13	0.17	0.23	0.11	0.10	0.13	0.17
Reference [27]	n = 2	mean	0.15	1.32	2.13	2.73	0.27	1.45	2.44	3.17	0.75	2.13	2.16	3.02	0.36	1.46	2.43	3.14
		SD	0.11	0.13	0.13	0.14	0.13	0.14	0.15	0.15	0.14	0.18	0.19	0.20	0.13	0.13	0.14	0.15
	n = 4	mean	0.09	0.11	1.42	2.53	0.14	0.17	2.32	3.74	0.11	0.16	1.02	3.0	0.15	0.13	2.40	3.02
		SD	0.10	0.10	0.11	0.15	0.12	0.09	0.13	0.15	0.13	0.15	0.15	0.20	0.12	0.08	0.13	0.15
Proposed method	n = 2	mean	0.75	2.67	2.91	3.31	0.18	0.15	2.29	3.92	0.83	2.46	2.54	2.62	0.45	2.13	2.64	3.25
		SD	0.13	0.14	0.14	0.15	0.14	0.08	0.12	0.15	0.16	0.18	0.19	0.19	0.14	0.12	0.12	0.16
	n = 4	mean	0.69	2.23	2.24	2.28	0.17	0.12	1.93	2.09	0.15	0.13	2.48	2.49	0.39	1.94	1.95	1.94
		SD	0.12	0.14	0.13	0.13	0.13	0.06	0.09	0.12	0.15	0.15	0.16	0.16	0.13	0.13	0.09	0.10

Table 5. BER of Testing Amplitude Scaling.

Audio Type		Love Song				Symphony				Dance				Folklore			
Amplitude Modification Factor		**0.5**	**0.8**	**1.1**	**1.2**	**0.5**	**0.8**	**1.1**	**1.2**	**0.5**	**0.8**	**1.1**	**1.2**	**0.5**	**0.8**	**1.1**	**1.2**
Reference [24]	n = 2	47.25	45.55	41.40	43.85	48.00	38.72	23.63	24.54	43.12	41.40	40.15	40.84	45.90	43.52	42.54	42.86
	n = 4	43.82	40.63	40.84	41.25	45.22	32.04	23.15	23.56	42.33	41.02	39.56	40.16	42.52	41.86	41.35	41.24
Reference [27]	n = 2	40.02	32.15	31.18	33.65	38.06	31.22	28.13	28.55	38.92	31.41	32.10	34.24	39.82	33.12	32.74	32.62
	n = 4	38.22	30.63	30.84	31.25	35.22	32.04	23.15	23.56	40.02	31.11	30.51	30.46	32.42	26.81	24.75	24.26
Proposed method	n = 2	2.03	1.15	1.08	1.13	1.65	0.97	1.43	1.45	2.85	1.76	1.85	2.06	1.67	1.31	0.93	1.32
	n = 4	0.97	0.86	0.84	0.92	1.14	0.88	0.92	0.98	2.04	1.56	0.98	1.93	1.05	0.86	0.83	0.85

Table 6. BER of Testing Time Scaling.

Audio Type			Love Song				Symphony				Dance				Folklore			
Time-Scaling (%)			**−5**	**−2**	**2**	**5**	**−5**	**−2**	**2**	**5**	**−5**	**−2**	**2**	**5**	**−5**	**−2**	**2**	**5**
Reference [24]	n = 2	mean	47.11	42.91	43.67	46.32	42.74	37.82	46.42	46.19	45.18	40.21	46.58	47.98	43.18	39.12	46.35	47.43
		SD	0.10	0.09	0.08	0.09	0.11	0.11	0.09	0.10	0.12	0.12	0.11	0.13	0.12	0.12	0.10	0.12
	n = 4	mean	47.04	40.23	45.11	46.58	43.11	36.64	46.24	46.86	44.37	39.91	44.92	47.98	43.03	38.62	46.53	47.54
		SD	0.08	0.07	0.07	0.08	0.09	0.10	0.09	0.10	0.12	0.11	0.11	0.12	0.13	0.12	0.09	0.10
Reference [27]	n = 2	mean	48.24	45.03	41.13	42.62	42.24	40.73	43.62	45.21	46.29	42.07	44.98	45.18	44.15	40.22	45.39	47.37
		SD	0.09	0.08	0.07	0.08	0.09	0.10	0.08	0.09	0.13	0.13	0.12	0.14	0.13	0.12	0.11	0.11
	n = 4	mean	46.12	41.25	44.01	44.52	42.01	38.34	45.27	45.89	45.27	40.91	45.02	46.13	42.53	39.24	45.63	45.58
		SD	0.08	0.08	0.08	0.07	0.10	0.09	0.09	0.09	0.12	0.13	0.11	0.14	0.13	0.11	0.10	0.09
Proposed method	n = 2	mean	47.23	42.05	43.53	45.15	42.32	37.64	45.18	46.21	45.35	40.42	46.24	46.47	43.18	38.93	46.41	47.13
		SD	0.07	0.07	0.06	0.07	0.08	0.09	0.09	0.10	0.11	0.10	0.10	0.13	0.12	0.11	0.08	0.09
	n = 4	mean	46.43	40.08	44.37	46.54	43.06	36.83	46.32	46.25	44.14	39.65	44.78	47.95	42.25	38.26	46.42	46.37
		SD	0.07	0.05	0.05	0.06	0.08	0.09	0.07	0.08	0.11	0.11	0.10	0.12	0.10	0.10	0.08	0.08

4.3. Compression Measurement

The purpose of compression is to have maximal compression ratio (CR) under maximal SNR, where CR is defined by

$$CR = \frac{\text{Data size before compression}}{\text{Data size after compression}}$$

The threshold ε and the embedding strength Q directly affect compression ratio (CR) and SNR, respectively. As shown in Figure 2, it can be said that the larger the threshold and the embedding strength, the larger the compression ratio CR, that is, the better the compression effect, but the worse the SNR before and after compression.

(a)

(b)

(c)

Figure 2. *Cont.*

(**d**)

(**e**)
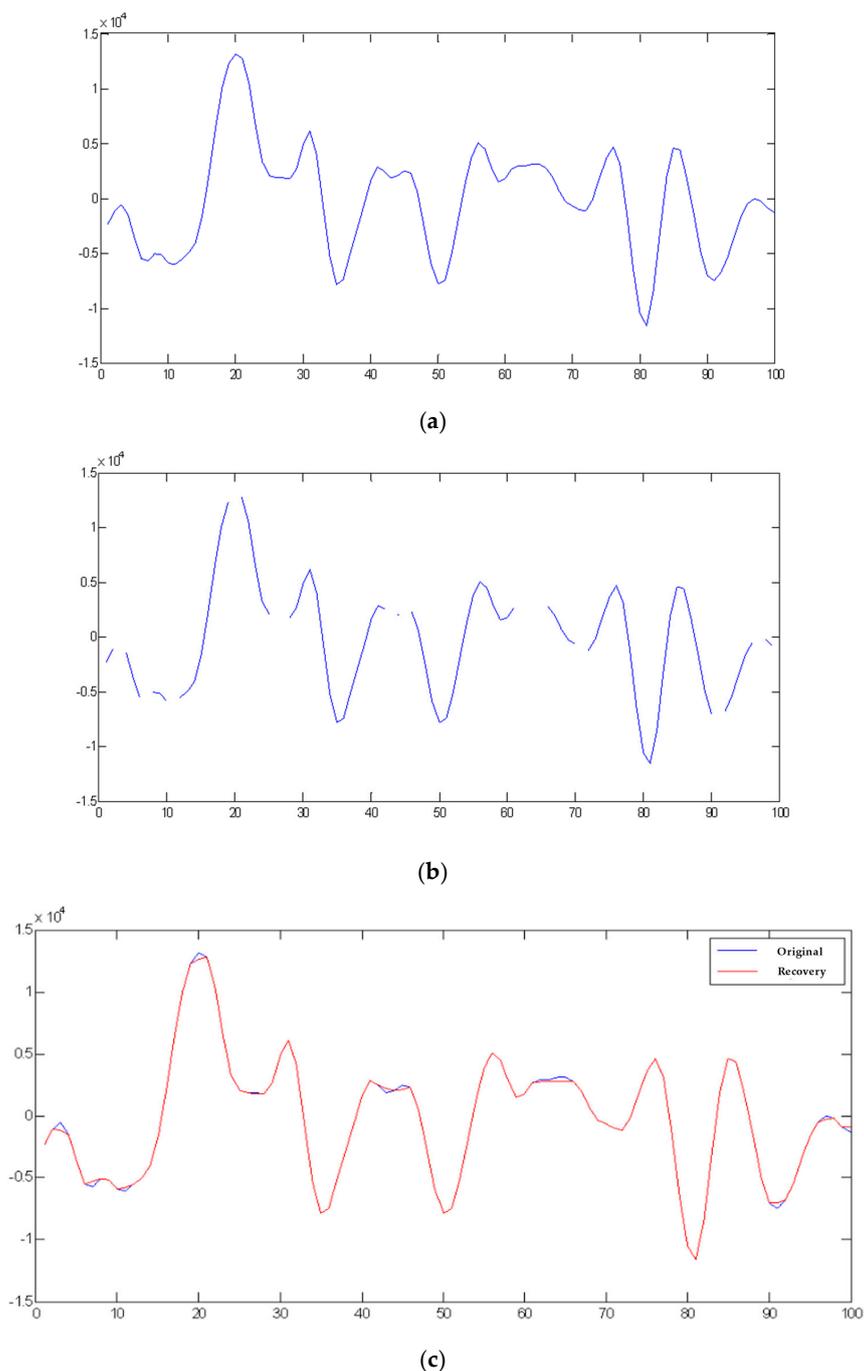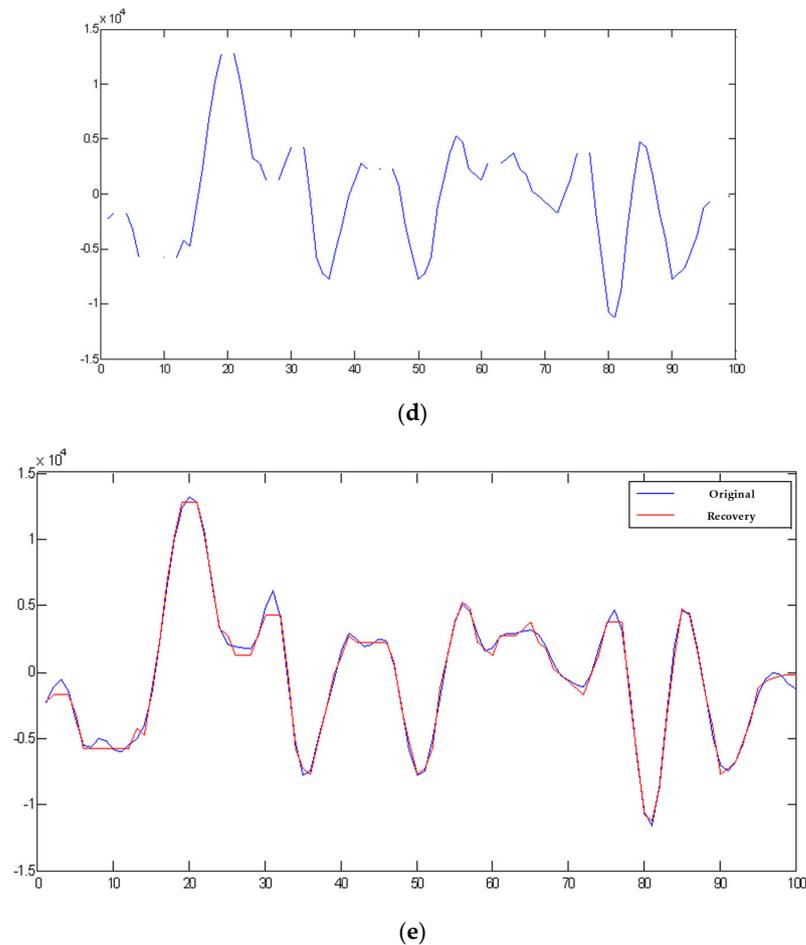
Figure 2. Comparison among the original audio, compressed audio, and decompressed audio in 1 and 100 audio samples with a threshold value of 500 and with/without embedding private information. (**a**) Original audio. (**b**) Compressed audio with threshold value of 500. (**c**) Recovering the compressed audio in (**b**). (**d**) Compressed audio with threshold value of 500 and embedding private information of embedding strength $Q = 1000$. (**e**) Recovering the compressed audio in (**d**).

Data in Table 7 show that the CR and SNR obtained without embedding information (denoted by N) vary under distinct threshold values. Such a result is consistent with the fact that when CR increases, SNR worsens. While signals are embedded, the result of the investigation of the relationship between the two is also in Table 7. From the experimental outcomes, we observed two noteworthy findings. First, with the same threshold value, if the embedding strength is higher, the overall audio values vary less. Though SNR seems worse, it demonstrates a more powerful embedding strength and a better compression effect. Second, in cases where threshold values change, we see that if the threshold value is greater than the embedding strength, the CR value becomes higher. That is to say, the compression effect enhancesand the relative decompression effect worsens, but the total effectiveness remains almost unchanged.

Moreover, to better understand the relationship between CR and SNR, we used the graph in Figure 3 to find appropriate values of threshold ε and embedding strength Q. We changed the threshold ε to obtain the relationship between CR and SNR using different markers by keeping green and blue fixed to include all the Q values obtained in Table 7. By doing this, we made sure the effect between CR and SNR remained optimized.

Table 7. Relationship between CR and SNR with and without embedding private information (*N*).

Threshold ε	*Q*	CR	SNR before Decompression
0.1	1	1.0016	36.2503
	100	1.0173	38.9726
	500	1.0905	31.7549
	1000	1.1900	28.0046
	2048	1.4187	23.0792
	4096	1.8970	17.4999
10	1	1.0028	35.2693
	100	1.0173	38.9726
	500	1.0905	31.7549
	1000	1.1900	28.0046
	2048	1.4187	23.0792
	4096	1.8970	17.4999
100	1	1.0314	34.1682
	100	1.0173	38.9726
	500	1.0905	31.7549
	1000	1.1900	28.0046
	2048	1.4187	23.0792
	4096	1.8970	17.4999
500	1	1.1953	25.4036
	100	1.1415	29.6514
	500	1.0905	31.7549
	1000	1.1900	28.0046
	2048	1.4187	23.0792
	4096	1.8970	17.4999
1000	1	1.4308	22.1784
	100	1.4115	25.8566
	500	1.3092	27.0426
	1000	1.1900	28.0046
	2048	1.4187	23.0792
	4096	1.8970	17.4999
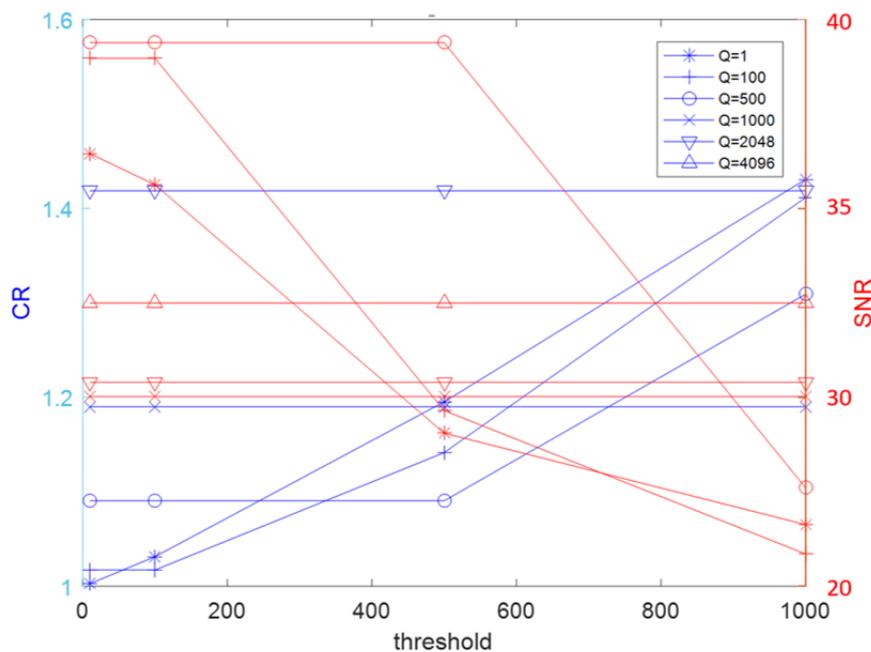
Figure 3. Changing the threshold ε to obtain the relationship between CR and SNR using different markers by keeping green and blue fixed to include all the *Q* values in Table 7.

5. Conclusions

This study proposes a method to seek the integration between the information-hiding process and data compression for five types of commonly seen audio signals. Under the proposed model, simulation results demonstrated that each piece of hidden audio signal attains high SNR and showed strong robustness. SNRs of most hidden audios were more than 35, and some were even higher than 40. On the other hand, most BERs were as low as 5% or less. In addition, we obtained the relationship between CR and SNR with embedded private information and observed two critical outcomes. First, with a fixed threshold value, a high embedding strength makes the differences between the overall audio values smaller. Such an algorithm shows better embedding strength and enhanced compression effects but reflects worse SNR values. Second, when playing with distinct threshold values, we found that if the threshold value ε is set higher than the embedding strength Q, the CR value drops. That means the compression effect becomes better and the relative decompression effect worsens, but the total effectiveness remains almost unchanged.

Author Contributions: Conceptualization: S.-T.C.; methodology: S.-T.C.; software: S.-T.C.; validation: S.-Y.T. and M.Z.; Writing—review and editing: S.-T.C. and S.-Y.T. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Appendix A

Theorem A1. *Let A be a matrix of size $n \times n$. If $\overline{X}$ and X are $n \times 1$ column vectors, then the following statement holds* [30,31]:

$$\frac{\partial AX}{\partial X} = A,$$

$$\frac{\partial (X - \overline{X})^T (X - \overline{X})}{\partial X} = 2(X - \overline{X}).$$

Theorem A2. *Suppose that g is a continuously differentiable function of X on a subset of the domain of a function f. If X_0 minimizes (or maximizes) f(X) subjected to the constraint g(X) = 0, then $\nabla f(X_0)$ and $\nabla g(X_0)$ are parallel and one is a constant multiple of another. That is, if $\nabla g(X_0) \neq 0$, then there exists a non-zero scalar λ such that*

$$\nabla f(X_0) = \lambda \nabla g(X_0).$$

Based on Theorem A2, if an augmented function is defined as

$$H(X, \lambda) = f(X) + \lambda g(X),$$

then an optimal solution of the optimization problem is to compute the extreme of the unconstraint function $H(X, \lambda)$. The necessary conditions to ensure the existence of the extreme of H are [30,31]

$$\frac{\partial H}{\partial \lambda} = 0, \frac{\partial H}{\partial X} = 0$$

References

1. Chang, C.-L.; Chang, C.-Y.; Tang, Z.-Y.; Chen, S.-T. High-Efficiency Automatic Recharging Mechanism for Cleaning Robot Using Multi-Sensor. *Sensors* **2018**, *18*, 11. [CrossRef] [PubMed]
2. Chang, C.-L.; Chen, C.-J.; Lee, H.-T.; Chang, C.-Y.; Chen, S.-T. Bounding the Sensing Data Collection Time with Ring-based Routing for Industrial Wireless Sensor Networks. *J. Internet Technol.* **2020**, *21*, 673–680.
3. Zuo, Z.; Liu, L.; Zhang, L.; Fang, Y. Indoor Positioning Based on Bluetooth Low-Energy Beacons Adopting Graph Optimization. *Sensors* **2018**, *18*, 3736. [CrossRef]
4. Chang, C.-L.; Chen, S.-T.; Chang, C.-Y.; Jhou, Y.-C. The Application of Machine Learning in Air Hockey Interactive Control System. *Sensors* **2020**, *18*, 7233. [CrossRef]
5. Lin, S.-J.; Chen, S.-T. Enhance the perception of easy-to-fall and apply the Internet of Things to fall prediction and protection. *J. Healthc. Commun.* **2020**, *5*, 52.
6. Zhang, X.; Zhang, S.; Huai, S. Low-Power Indoor Positioning Algorithm Based on iBeacon Network. *Hindawi Complex.* **2021**, *2021*, 8475339. [CrossRef]
7. Zhou, C.; Yuan, J.; Liu, H.; Qiu, J. Bluetooth indoor positioning based on RSSI and Kalman filter. *Wirel. Pers. Commun.* **2017**, *96*, 4115–4130. [CrossRef]
8. Song, W.; Lee, H.M.; Lee, S.H.; Choi, M.H.; Hong, M. Implementation of android application for indoor positioning system with estimote BLE beacons. *J. Internet Technol.* **2018**, *19*, 871–878.
9. Cui, L.; Yang, S.; Chen, F.; Ming, Z.; Lu, N.; Qin, J. A survey on application of machine learning for Internet of Things. *Int. J. Mach. Learn. Cyber.* **2018**, *9*, 1399–1417. [CrossRef]
10. Baldini, G.; Dimc, F.; Kamnik, R.; Steri, G.; Giuliani, R.; Gentile, C. Identification of mobile phones using the built-in magnetometers stimulated by motion patterns. *Sensors* **2017**, *17*, 783. [CrossRef]
11. IFPI (International Federation of the Phonographic Industry). Available online: http://www.ifpi.org (accessed on 10 January 2021).
12. Katzenbeisser, S.; Petitcolas, F.A.P. (Eds.) *Information Hiding Techniques for Steganography and Digital Watermarking*; Artech House, Inc.: Norwood, MA, USA, 2000.
13. Al-Haj, A.; Mohammad, A.A.; Bata, L. DWT-based audio watermarking. *Int. Arab. J. Inf. Technol.* **2011**, *8*, 326–333.
14. Xiang, S. Robust audio watermarking against the D/A and A/D conversions. *EURASIP J. Adv. Signal Processing* **2011**, *3*, 29. [CrossRef]
15. Chen, S.-T.; Wu, G.-D.; Huang, H.-N. Wavelet-Domain Audio Watermarking Scheme Using Optimization-Based Quantization. *IET Signal Processing* **2010**, *4*, 720–727. [CrossRef]
16. Noriega, R.M.; Nakano, M.; Kurkoski, B.; Yamaguchi, K. High Payload Audio Watermarking: Toward Channel Characterization of MP3 Compression. *J. Inf. Hiding Multimed. Signal Process.* **2011**, *2*, 91–107.
17. Mishra, J.; Patil, M.V.; Chitode, J.S. An Effective Audio Watermarking using DWT-SVD. *Int. J. Comput. Appl.* **2013**, *70*, 6–11. [CrossRef]
18. Zhao, M.; Pan, J.-S.; Chen, S.-T. Entropy-Based Audio Watermarking via the Point of View on the Compact Particle Swarm Optimization. *J. Internet Technol.* **2015**, *16*, 485–495.
19. Darabkh, A.K. Imperceptible and Robust DWT-SVD-Based Digital Audio Watermarking Algorithm. *J. Softw. Eng. Appl.* **2014**, *7*, 859–871. [CrossRef]
20. Chen, S.-T.; Guo, Y.-J.; Huang, H.-N.; Kung, W.-M.; Tseng, K.-K.; Tu, S.-Y. Hiding Patients Confidential Data in the ECG Signal via a Transform-Domain Quantization Scheme. *J. Med. Syst.* **2014**, *38*, 54. [CrossRef]
21. Zear, A.; Singh, A.K.; Kumar, P. A proposed secure multiple watermarking technique based on DWT, DCT and SVD for application in medicine. *Multimed. Tools Appl.* **2016**, *77*, 4863–4882. [CrossRef]
22. Wu, Q.; Wu, M. A Novel Robust Audio Watermarking Algorithm by Modifying the Average Amplitude in Transform Domain. *Appl. Sci.* **2018**, *8*, 723. [CrossRef]
23. Karajeh, H.; Khatib, T.; Rajab, L.; Maqableh, M. A robust digital audio watermarking scheme based on DWT and Schur decomposition. *Multimed. Tools Appl.* **2019**, *78*, 18395–18418. [CrossRef]
24. Chen, S.-T.; Huang, H.-N. Optimization-Based Audio Watermarking with Integrated Quantization Embedding. *Multimed. Tools Appl.* **2016**, *75*, 4735–4751. [CrossRef]
25. Shankar, T.; Yamuna, G. Optimization Based Audio Watermarking using Discrete Wavelet Transform and Singular Value Decomposition. *Int. J. Electron. Electr. Comput. Syst.* **2017**, *6*, 375–379.
26. Dhar, P.K.; Shimamura, T. Blind Audio Watermarking in Transform Domain Based on Singular Value Decomposition and Exponential-Log Operations. *Radio Eng.* **2017**, *26*, 552–561. [CrossRef]
27. Li, J.-F.; Wang, H.-X.; Wu, T.; Sun, X.-M.; Qian, Q. Norm ratio-based audio watermarking scheme in DWT domain. *Multimed. Tools Appl.* **2018**, *77*, 14481–14497. [CrossRef]
28. Mallat, S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intel.* **1989**, *11*, 674–693. [CrossRef]
29. Burrus, C.S.; Gopinath, R.A.; Gao, H. *Introduction to Wavelet Theory and Its Application*; Prentice-Hall: Hoboken, NJ, USA, 1998.
30. Lewis, F.L. *Optimal Control*; John Wiley and Sons: New York, NY, USA, 1986.
31. Bartle, R.G. *The Elements of Real Analysis*, 2nd ed.; Wiley: New York, NY, USA, 1976.