

Audio Watermarking Using Wavelet Transform and Genetic Algorithm for Realizing High Tolerance to MP3 Compression

Shinichi Murata¹, Yasunari Yoshitomi², Hiroaki Ishii³

¹Panasonic Corporation, Kadoma, Osaka, Japan

²Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan

³School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo, Japan

E-mail: yoshitomi@kpu.ac.jp

Received April 12, 2011; revised June 8, 2011; accepted June 20, 2011

Abstract

Recently, several digital watermarking techniques have been proposed for hiding data in the frequency domain of audio signals to protect the copyrights. However, little attention has been given to the optimal position in the frequency domain for embedding watermarks. In general, there is a tradeoff between the quality of the watermarked audio and the tolerance of watermarks to signal processing methods, such as compression. In the present study, a watermarking method developed for a visual image by using a wavelet transform was applied to an audio clip. We also improved the performance of both the quality of the watermarked audio and the extraction of watermarks after compression by the MP3 technique. To accomplish this, we created a multipurpose optimization problem for deciding the positions of watermarks in the frequency domain and obtaining a near-optimum solution. The near-optimum solution is obtained by using a genetic algorithm. The experimental results show that the proposed method generates watermarked audios of good quality and high tolerance to MP3 compression. In addition, the security was improved by using the characteristic secret key to embed and extract the watermark information.

Keywords: Audio Watermarking, Genetic Algorithm, Optimization, Wavelet Transforms, Secret Key

1. Introduction

Recent progress in digital media and digital distribution systems, such as the Internet and cellular phones, has enabled us to easily access, copy, and modify digital content, such as electric documents, images, audio, and video. Under these circumstances, techniques to protect the copyrights of digital data and to prevent unauthorized duplication or tampering of these data are strongly desired.

Digital watermarking (DW) is a promising method for the copyright protection of digital data. Several studies have investigated audio DW [1-12]. Currently, digital audio clips distributed over the Internet or cellular phone systems are often modified by compression, which is one of the easiest and most effective ways to overcome DW without significantly deteriorating the quality of the audio. Two important properties in audio DW are the inaudibility of the distortion due to DW, and the robustness

against signal processing methods, such as compression. In addition to these properties, the data rate and the complexity of DW have attracted attention when discussing the performance of DW.

We developed a method in which 1) a digital watermark can be sufficiently extracted from watermarked audio, even after compression, and 2) the quality of the audio remains high after embedding the digital watermark. However, there generally is a trade-off relation between these two properties.

In the present study, we improved both the extraction of digital watermarks and the quality of the watermarked audio by developing a multipurpose optimization problem for deciding the positions of digital watermarks in the frequency domain and obtaining a near-optimum solution by using a discrete wavelet transform (DWT) and a genetic algorithm (GA) [13,14] for realizing high tolerance to compression by MP3, which is the most popular compression technique. The proposed method

enables us to embed digital watermarks in a near-optimum manner for each audio file. In addition, the security of the watermarked audio is improved by using a characteristic secret key to embed and extract digital watermarks.

2. Wavelet Transform

Original audio data $s_k^{(0)}$ is used as the level-0 wavelet decomposition coefficient sequence, where k denotes the element number in the data. The data is decomposed into the multi-resolution representation (MRR) and the coarsest approximation by repeatedly applying a DWT. The wavelet decomposition coefficient sequence $s_k^{(j)}$ at level j is decomposed into two wavelet decomposition coefficient sequences at level $j+1$ by (1) and (2):

$$s_k^{(j+1)} = \sum_n \overline{p_{n-2k}} s_n^{(j)}, \quad (1)$$

$$w_k^{(j+1)} = \sum_n \overline{q_{n-2k}} s_n^{(j)}, \quad (2)$$

where p_{n-2k} and q_{n-2k} denote the scaling and wavelet sequences, respectively, and $w_k^{(j+1)}$ denotes the development coefficient at level $j+1$. The development coefficients at level J are obtained by using (1) and (2) iteratively from $j=0$ to $j=J-1$. **Figure 1** shows the process of a multi-resolution analysis by DWT.

The signal is re-composed by using (3) repeatedly from $j=J-1$ to $j=0$.

$$s_n^{(j-1)} = \sum_k \left[p_{n-2k} s_k^{(j)} + q_{n-2k} w_k^{(j)} \right] \quad (3)$$

In the present study, we use the Daubechies wavelet for DWT. As a result, we obtain the following relation between p_{n-2k} and q_{n-2k} :

$$q_k = (-1)^k p_{1-k} \quad (4)$$

3. Wavelet Domain Digital Watermarking Based on Threshold-Variable Decision

It is known that the histogram of the wavelet coefficients of each domain of MRR sequences has a distribution that

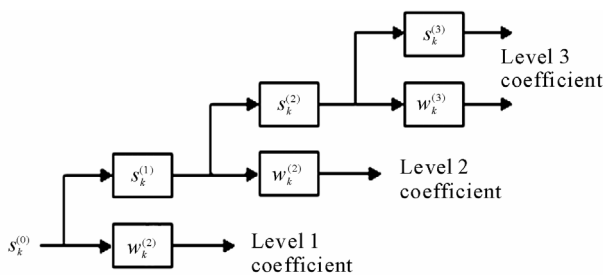


Figure 1. Multi-resolution analysis by the DWT.

is centered at approximately 0 when DWT is performed on a natural visual image [15]. For an audio clip, we also found the same phenomena. **Figure 2** shows an example of an audio histogram.

In the present research, the technique [15] for exploiting the above phenomena on a natural image for embedding a digital watermark on the wavelet coefficients of MRR sequences is applied to audio DW. The procedure is described below.

3.1. Embedment of Watermark Information

3.1.1. Setting of Parameters

For the watermarking of an audio clip, we obtain the histogram of the wavelet coefficients V at the selected level of MRR sequences. **Figure 3** shows a schematic diagram of the histogram of the wavelet coefficients V of an MRR sequence. As with the DW techniques for images [15,16], we set the following watermarking parameters:

The values of $Th(\text{minus})$ and $Th(\text{plus})$ (see **Figure 3**) are chosen such that the non-positive wavelet coefficients (S_m in total frequency) are equally divided into two groups by $Th(\text{minus})$, and the positive wavelet coefficients (S_p in total frequency) are equally divided into two groups by $Th(\text{plus})$. Next, the values of $T1$, $T2$, $T3$, and $T4$, which are the parameters for controlling the embedment strength, are chosen to satisfy the following conditions:

- 1) $T1 < Th(\text{minus}) < T2 < 0 < T3 < Th(\text{plus}) < T4$.
- 2) The value of S_{T1} , the number of wavelet coefficients in $(T1, Th(\text{minus}))$, is equal to S_{T2} , the number of wavelet coefficients in $[Th(\text{minus}), T2]$. In short, $S_{T1} = S_{T2}$.
- 3) The value of S_{T3} , the number of wavelet coefficients in $(T3, Th(\text{plus})]$, is equal to S_{T4} , the number of

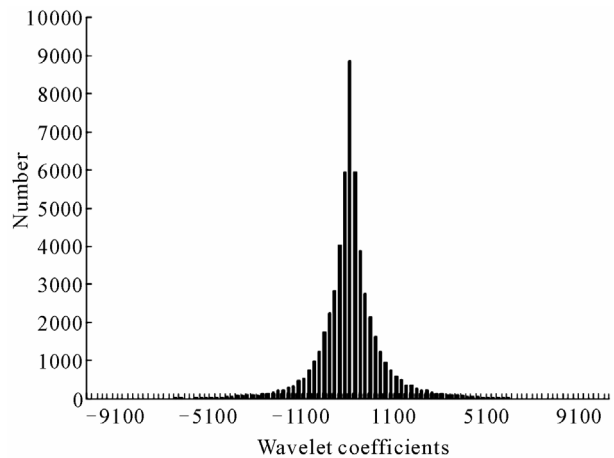


Figure 2. Histogram of the wavelet coefficients of an MRR sequence at level 3 (jazz).

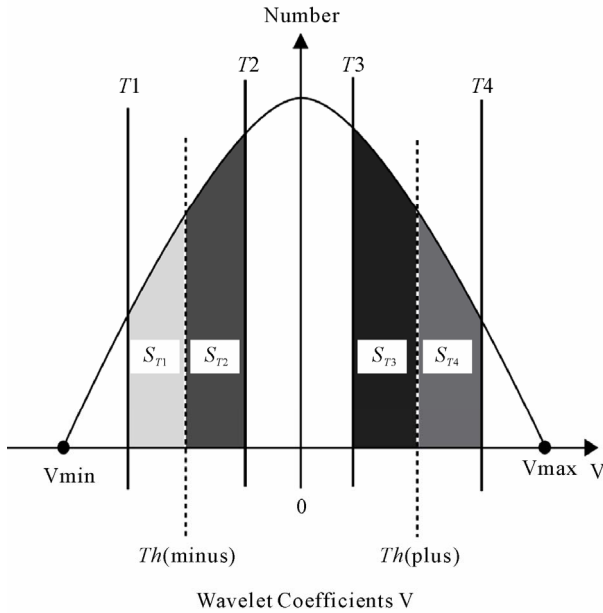


Figure 3. Schematic diagram of the histogram of MRR wavelet coefficients.

wavelet coefficients in $(Th(plus), T4)$. In short, $S_{T3} = S_{T4}$.

$$4) S_{T1}/S_m = S_{T3}/S_p.$$

In the present study, the values of both S_{T1}/S_m and S_{T3}/S_p are set to 0.2, which was determined experimentally.

3.1.2. Embedment of Watermark Information

The wavelet coefficients of MRR are rewritten according to the following rules in embedding digital watermark. Here, V_i denotes one of the wavelet coefficients.

- 1) In the case that bit W_i in watermark W is 0,
 - when $V_i < T2$, V_i is changed to $T2$.
 - when $V_i > T3$, V_i is changed to $T3$.
 - when $T2 \leq V_i \leq T3$, V_i is kept.
- 2) In the case that bit W_i in watermark W is 1,
 - when $T1 < V_i \leq 0$, V_i is changed to $T1$.
 - when $0 < V_i < T4$, V_i is changed to $T4$.
 - when $V_i \leq T1$ or $V_i \geq T4$, V_i is kept.

The wavelet coefficient V_i is set in the range of $T2 \leq V_i \leq T3$ when bit W_i in watermark W is 0, whereas the DWT coefficient V_i is set in the range of $V_i \leq T1$ or $V_i \geq T4$ when bit W_i in watermark W is 1. The frequency of the change of V_i toward the inside is expected to be approximately equal to the change toward the outside when the number of 0 bits is approximately the same as the number of 1 bits.

3.1.3. Generation of Watermarked Audio

The inverse DWT (IDWT) is performed to wavelet coefficients V'_i embedded with the watermark to obtain the

audio with the watermark.

3.2. Detection of the Watermark

3.2.1. Presumption of Parameters

The watermarked audio, which may be modified by signal processing methods, such as MP3 compression, is converted into wavelet coefficients. The wavelet coefficient in the region where the watermark information is embedded is denoted as V'' .

For the histogram of V'' , the two parameters $Th'(\text{minus})$ and $Th'(\text{plus})$, which correspond to $Th(\text{minus})$ and $Th(\text{plus})$ for the histogram of V before embedding the watermark, are obtained in the same manner as that for $Th(\text{minus})$ and $Th(\text{plus})$, mentioned in Section 3.1.1. $Th'(\text{minus})$ and $Th'(\text{plus})$ can be used as presumptive values for $Th(\text{minus})$ and $Th(\text{plus})$, respectively, because the distribution of the histogram of V'' is expected to be approximately the same as that of V before embedding the watermark.

The watermarked audio can undergo certain types of audio processing, including compression, such that the difference between the distribution of the histogram of V'' after audio processing and that of V before embedding the watermark is not negligible. In such a case, it may not be persuasive that $Th'(\text{minus})$ and $Th'(\text{plus})$ can be used as presumptive values for $Th(\text{minus})$ and $Th(\text{plus})$, respectively.

3.2.2. Detection of Watermark Information

When the wavelet coefficient V_i'' is in the range of $Th'(\text{minus}) \leq V_i'' \leq Th'(\text{plus})$, the corresponding bit W_i'' in measured watermark W'' is judged to be 0. When the DWT coefficient V_i'' is in the range of $V_i'' < Th'(\text{minus})$ or $V_i'' > Th'(\text{plus})$, the corresponding bit W_i'' in the measured watermark W'' is judged to be 1.

The detection rate $d(x)$ (%) is defined as the percentage of correspondence between the bit W_i in watermark W and the corresponding bit W_i'' in measured watermark W'' .

4. Use of the Secret Key

When we embed a digital watermark by using the partial problem described in Section 5, the watermark is produced by using a secret key $S(\gamma)$, which is composed of a row of γ integers randomly selected once or less per integer in the integer range from 1 to N , as shown in the example below ($\gamma = 400$), where γ is the number of bits of the watermark and N is the total number of wavelet coefficients that are candidates embedded with DW.

$$S(400) = (271, 72, 39, 990, 524, 88, \dots, 1011, 688, 312) \quad (5)$$

Each value and order of numbers in $S(\gamma)$ indicate the position of each bit of the digital watermark in the DW region. Here, the position in the region is expressed as a one-dimensional coordinate. For example, the first number, 271, and the second number, 72, in $S(400)$ mean that the first and second bits of the watermark are set for the wavelet coefficients at the coordinates of 271 and 72 in the DW region, respectively.

As described in Section 3.1.2, the selected wavelet coefficient in the target region is changed according to the value of each bit of the digital watermark, the secret key, and the shift value k of the coordinate, which is described in Section 5. The value of each bit of the digital watermark and the secret key decide an initial bit pattern in the positions in the DW region.

The DW positions of the wavelet coefficients decided by secret key $S(4)$ presented below in (6) are simply demonstrated in **Figure 4**.

$$S(4) = (271, 72, 39, 990) \quad (6)$$

The coordinate shift is performed by generating $S'(\gamma)$ such that the shift value is added to all values of the elements in $S(\gamma)$. For example, it is assumed that the shift value is 10. As a result,

$$S'(4) = \{281, 82, 49, 1000\} \quad (7)$$

The DW positions of the wavelet coefficients decided by secret key $S'(4)$ described in (7) are simply demonstrated in **Figure 5**.

5. Optimal Watermarking Problem

Because our approach of DW optimization is on the first and challenging stage, we formulate the problem in a simple way on the viewpoint of optimization problem. Therefore, in the present study, we formulate the optimization problem as minimization for distortion. We can also formulate the problem using the constraint of keeping distortion less than the masking threshold. Such more elaborate approach is our next target.

To minimize the error $e(\mathbf{x})$ caused by watermarking and to maximize the detection rate $d(\mathbf{x})$ (%) of digital watermark after compression by the MP3 technique under a restriction condition on $d(\mathbf{x})$, an optimum watermarking problem is formulated as follows:

$$P \text{ Minimize } e(\mathbf{x}), \quad (8)$$

$$\text{Maximize } d(\mathbf{x}), \quad (9)$$

$$\text{Subject to } d(\mathbf{x}) > a, \quad (10)$$

$$e(\mathbf{x}) = \sum_{i=1}^N (y_i - y'_i)^2, \quad (11)$$

Case WCV level 3

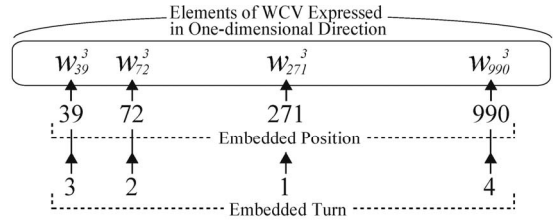


Figure 4. The DW positions of wavelet coefficients decided by secret key. $S(4) = (271, 72, 39, 990)$ in the case of DWT level 3.

Case WCV level 3 $k=10$

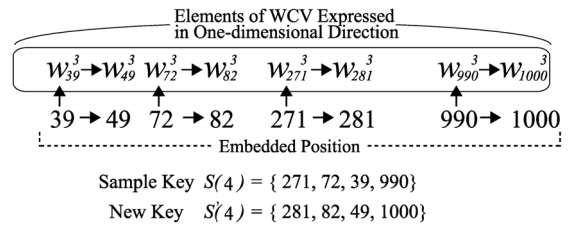


Figure 5. The DW positions of wavelet coefficients decided by secret key. $S'(4) = \{281, 82, 49, 1000\}$ obtained from $S(4) = (271, 72, 39, 990)$ by using the shift value $k = 10$ in the case of DWT level 3.

$$\sum_{i=1}^N x_i = R, \quad (12)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_N), \quad (13)$$

$$x_i \in \{0, 1\}, \quad (14)$$

where y_i, y'_i denote the values of the i -th sound data before and after embedding the digital watermark, respectively; N denotes the total number of wavelet coefficients at the DWT level selected for embedment; a, R are constants; and x_i is a 0 - 1 variable that decides the embedment of the watermark on the corresponding wavelet coefficient, where 1 denotes an embedment and 0 denotes a non-embedment.

When the number of wavelet coefficients that are possible targets for digital watermark embedment becomes larger, the solution space of P becomes larger, with the result that a search for an optimal or near-optimal solution is time-consuming or difficult. Accordingly, we define the partial problem P' as follows:

$$P' \text{ Minimize } e(s), \quad (15)$$

$$\text{Maximize } d(s), \quad (16)$$

$$\text{Subject to } d(s) > a, \quad (17)$$

$$e(s) = \sum_{i=1}^N (y_i - y'_i)^2, \quad (18)$$

$$x_{i+s} = c_i, \quad (19)$$

$$\sum_{i=1}^N c_i = R, \quad (20)$$

$$\mathbf{c} = (c_1, c_2, \dots, c_N), \quad (21)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_N), \quad (22)$$

$$c_i \in \{0, 1\}, \quad (23)$$

$$x_i \in \{0, 1\}, \quad (24)$$

where s is an integer variable ranging from 0 to $N-1$; c_i is a 0 - 1 constant that decides the digital watermark embedment on the corresponding wavelet coefficient, where 1 denotes an embedment and 0 denotes a non-embedment, and y_i, y'_i, N, a, R, x_i are the same as those described for P . We prepare a random initial pattern for $\mathbf{c} = (c_1, c_2, \dots, c_N)$ to solve P' .

For getting the detection rate $d(\mathbf{x})$, we use the watermark, the wavelet transformation level, and the (near-)optimum solution for P or P' as input data to the decoder for the proposed method.

In the present study, we use DWT. However other optimal watermarking problems can be formulated using other transforms as shown in our previous study [17,18] where discrete Fourier transform and a watermarking method proposed in the reported study [19] were used.

6. GA Approach

GA is one of the most acknowledged methods for near-optimization. We presume that P and P' might have many locally optimum solutions. According to our experiences, GA was fairly effective for searching the acceptable near-optimum solutions for many discrete optimization problems even when many locally optimum solutions could exist. We use GA in the present study because we have much more successful experiences on GA application to optimization problems than those on other methods. Other acknowledged techniques such as tabu search [20] for solving discrete optimization problems could be other options for solving P and P' .

The GA, which is based on biological evolution, has been applied for solving the optimization problem. The solution of the optimization problem is expressed as a genotype. In each generation, there is a population composed of several individuals identified by their genotypes. The basic idea of GA is that if the number of better individuals is increased by generation updating, an optimum or approximately optimum solution, as expressed by an individual, will eventually be obtained. Chromosome composed of genes is string specifying an individual. For the generation updating, crossover and mutation are performed. The crossover takes two parent strings and generates two offspring strings. The mutation changes se-

lected strings in a random way. In the references [13,14], the GA is explained in detail.

In this section, we explain our approach for obtaining near-optimum solutions for P and P' by using a GA.

6.1. Coding

The GA coding for Experiment 1, described below, is performed as follows.

A gene is expressed by a bit of value 0 or 1. Accordingly, each chromosome is composed of a row of bits. The total number of bits is m , in which n bits of the higher ranks are associated with the level of DWT and $(m-n)$ subordinate bits are associated with the shift value s , described for P' in the binary expression (Figure 6).

When an individual associated with a level that actually does not exist in a list of levels used for DWT is generated in the GA process, the individual is judged to have a fatal gene and is deleted, and a new individual is generated.

The GA coding for Experiment 2, described below, is performed as follows.

A gene is expressed by a bit of value 0 or 1 for P . Accordingly, each chromosome is composed of a row of bits. The total number of bits is k . Each bit is assigned to each DWT coefficient for possible embedment of the watermark (Figure 7). The value of 1 for a bit means that the corresponding DWT coefficient is selected as an object of watermarking, whereas 0 denotes non-embedment for the corresponding DWT coefficient.

A gene is also expressed by a bit of value 0 or 1 for P' . Accordingly, each chromosome is composed of a row of bits. The total number of bits is l . The chromosome expresses shift value s described in P' in the binary expression (Figure 8).

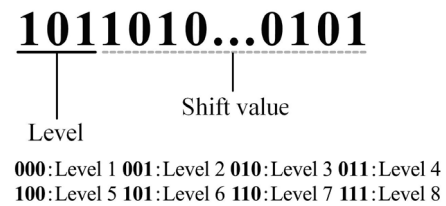


Figure 6. Schematic diagram of a chromosome structure in Experiment 1 for P' .

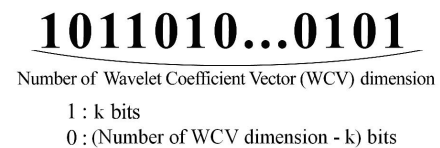


Figure 7. Schematic diagram of a chromosome structure in Experiment 2 for P .

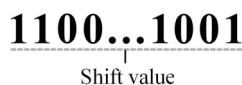


Figure 8. Schematic diagram of a chromosome structure in Experiment 2 for P' .

6.2. Strategy

For P , a one-point crossover and a mutation by the exchange(s) of pairs of 0 and 1 on a chromosome are used, while for P' a two-point crossover and a one-point mutation are used. The fitness function $f = (d - a + 1)/e$ is used for both P and P' , where d , a , and e were introduced in Section 5.

7. Numerical Experiments

In this section, we describe our computer experiments and the results for evaluating the performance of the proposed method.

7.1. Method

The experiment was performed in the following computational environment: the personal computer was a Dell Dimension DXC051 (CPU: Pentium IV 3.0 GHz; main memory: 1.0 GB); the OS was Microsoft Windows XP; the development language was Microsoft Visual C++ 6.0.

For DWT, we use Daubechies wavelets, which were successfully used in related research on DW techniques for images [16]. Moreover, a string composed of 400 randomly generated bits is used as the watermark information.

Five music audio files, composed of the first entry in five genre categories (classical, jazz, popular, rock, and hip-hop) in the research music database RWC [21], were copied from CDs onto the personal computer as WAVE files with the following specifications: 44.1 kHz, 16 bits, and monaural. For each music audio file selected from the database, one 10-sec clip of the music audio (hereafter referred to as the original music audio clip) was extracted starting at 1 minute from the beginning of the audio file and saved on a personal computer. The watermarked music audio clip was produced by embedding a digital watermark on the original audio clip by the proposed method.

In Experiment 1, where the near-optimum solutions for P' were obtained by using GA to evaluate the tolerance of watermarking to compression, MP3, AAC, and WMA compression systems were each used to compress the watermarked music audio clip to bitrates of 64, 96, and 128 kbps. The bitrate of 32 kbps was also used for MP3. Moreover, for P' , the fitness function value of the

near-optimum solutions obtained with GA was compared with the fitness function values of the feasible solutions at the initial generation.

In Experiment 2, the near-optimum solutions for P and P' were obtained by using GA, and the performances of those solutions were compared with respect to the calculation times for getting those solutions, the quality of the watermarked music audio clip, and the detection rate of the watermarks after compression by the MP3 technique. Moreover, for P and P' , the near-optimum solutions obtained by using GA were compared with the solutions produced by random generation of individual, neglecting the restrictions (10) for P and (17) for P' , with respect to the quality of the watermarked music audio clip, and the detection rate of the watermarks after compression by the MP3 technique.

7.2. Procedure

The procedure in the experiment is as follows.

Step 1:

First, an initial population consisting of several individuals is generated. In the process of generating an initial population having a given number of individuals, the individual that does not meet the restriction or that has a fatal gene is deleted as soon as it is produced, and a new individual is generated. If all individuals generated in 300 continuous trials do not meet the restriction or have at least one fatal gene, the procedure is terminated. When an initial population having a given number of feasible individuals is generated, go to Step 2.

Step 2:

The embedment of digital watermark according to the condition decided by each individual, the sound compression with the MP3 technique and the detection of digital watermark after MP3 decoding are performed, and then the fitness is calculated. When the generation is final, the procedure is terminated. Otherwise, go to Step 3.

Step 3:

The roulette strategy for selection, crossover, and mutation are performed. Go to Step 2.

The near-optimum solution, which is defined as the solution having the highest fitness through all generations, is obtained by repeating the process from Steps 2 to 3 until the given final generation.

7.3. Conditions

Table 1 shows the conditions of the GA strategy. In addition to the conditions shown in **Table 1**, the lower bound of the detection rate a , described by the restriction conditions (10) and (17), was set to 90. **Table 2** shows

Table 1. Conditions of the GA strategy.

Exp. No.	Problem	Population size	Generation loop	Crossover rate	Mutation rate
1	P'	30	100	0.6	0.1
2	P'	30	100	0.6	0.1
	P	50	200	0.6	0.2

Table 2. Conditions of chromosome structure in GA.

Exp. No.	Problem	Parameter values
1	P'	$m = 19, n = 3$
2	P'	$l = 16$
	P	$k = 400$

the conditions of the chromosome structure in GA (k, l, m, n are introduced in Section 6).

For Experiment 1, 32 and 64 kbps were used as the bitrates of the MP3 compression and DWT levels ranging from one to eight were selected as the search range. For Experiment 2, 96 kbps was used as the bitrate of the MP3 compression for level 3 of DWT, and 32 kbps was used for levels 4 to 6 of DWT.

For Experiment 1, we obtained the segmental-signal-to-quantization-noise ratio (hereafter referred to as SNR_{seg}), as defined by (25) and (26).

$$SNR_j = 10 \log_{10} \left\{ \frac{\sum_{r=1}^{N_j} y_{j,r}^2}{\sum_{r=1}^{N_j} (y_{j,r} - y'_{j,r})^2} \right\}, \quad (25)$$

$$SNR_{seg} = \frac{1}{N_f} \sum_{j=1}^{N_f} SNR_j, \quad (26)$$

where $y_{j,r}, y'_{j,r}$ denote the values of the r -th sound data of frame j before and after embedding the digital watermark, respectively, and N_j, N_f denote the number of sound data at frame j and the number of frames to be measured for SNR_{seg} , respectively. In the present study, we used 209 ms as the time length of one frame. When calculating SNR_{seg} , we excluded the frame with $SNR_j = \infty$, which means there is no change of all values of the sound data of frame j . An audio frame size of 209 ms is adopted for comparison with results obtained by our previous watermarking method [17,18], where the frame size was decided in the relation to the condition of watermarking.

7.4. Results and Discussions

7.4.1. Experiment 1

In this subsection, to examine the performance of the proposed method in Experiment 1, the fitness function

value of the near-optimum solution for the partial problem P' is first compared with that obtained from 30 feasible solutions generated at random for the partial problem P' . Next, the tolerances of watermark obtained by the proposed method to compression by MP3, AAC, and WMA are shown, and the time to obtain the near-optimum solution for the partial problem P' is shown to check the practicability of the proposed method.

Each DWT level obtained as an element composed of a chromosome of the near-optimum solution by GA was 4 for classical and jazz, 5 for rock and hiphop, and 6 for popular music for 32 kbps of the bitrate condition of MP3, while the DWT level for 64 kbps as the bitrate condition of MP3 was 3 for classical, jazz, and rock music, 4 for popular, and 5 for hiphop.

As shown in **Figures 9-13**, GA successfully found a good solution considered as the near-optimum solution for each case. **Table 3** shows the tolerance of watermark to the compression. **Table 4** shows the time to obtain a near-optimum solution. For the MP3 bitrate condition of 32 and 64 kbps in the process of GA, the average detection rate after compression by MP3 with the bitrate of 32 to 128 kbps was 98.39% and 93.88%, respectively, and the average time to obtain the near-optimum solution was 5.94×10^2 and 3.78×10^2 sec, respectively. As a condition for obtaining the high tolerance of watermark to MP3 compression, 32 kbps was better than 64 kbps. However, it took more time to obtain the near-optimum solution for the MP3 bitrate condition for 32 kbps than the time for 64 kbps. Moreover, for the MP3 bitrate condition of 32 kbps and 64 kbps in the process of GA, the average detection rate after compression by AAC with the bit rate of 64 to 128 kbps was 93.71% and 84.62%, respectively, and that by WMA with the bitrate of 64 to 128 kbps was 98.33% and 95.07%, respectively. As shown in **Table 5**, SNR_{seg} after embedding the watermark on the condition of the near-optimum solution obtained by the proposed method was 69.7 to 78.6 dB. These values suggest that noise due to the watermarking was difficult to perceive.

7.4.2. Experiment 2

In this subsection, to examine the performance of the proposed method in Experiment 2, the performances of the near-optimum solutions for the original problem P and the partial problem P' and those obtained from the solutions generated at random for the original problem P and the partial problem P' are compared with respect to the detection rates of the watermarks after MP3 compression and by the errors of the watermarking. Next, the times to obtain the near-optimum solution for the original problem P and the partial problem P' are shown to

check the practicability of the proposed method.

As shown in **Figures 14 to 18**, GA successfully found a good solution considered as the near-optimum solution in each case for the original problem P and the partial problem P' . Moreover, the near-optimum solution for

the original problem P had better performance as a condition for embedding watermarks than that for the partial problem P' . However, the time to obtain the near-optimum solution for the original problem P was approximately 2 to 200 times, compared with that for the partial

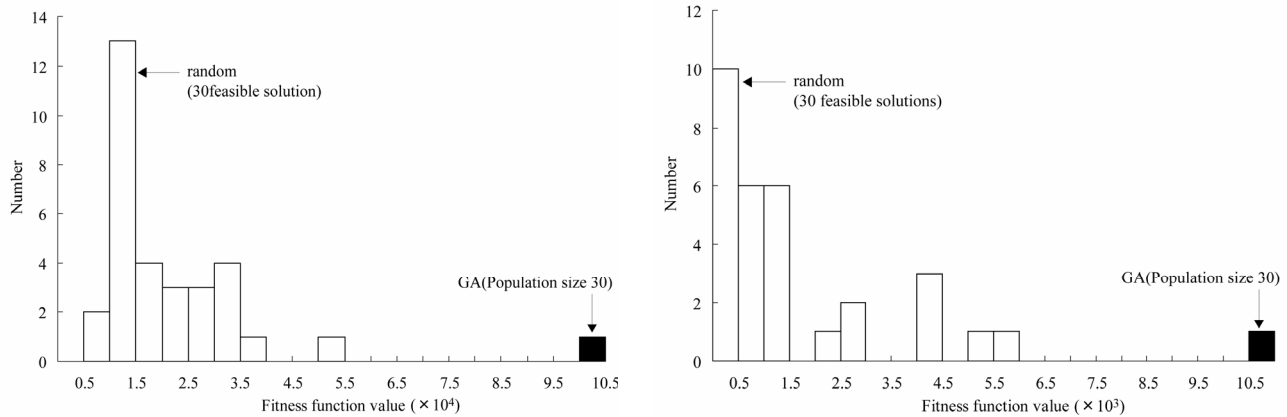


Figure 9. Comparison of fitness function value between the near-optimum solution and the 30 feasible solutions generated at random (Experiment 1, music file: classical). Left figure: MP3 bit rate; 32 kbps. Right figure: MP3 bit rate; 64 kbps.

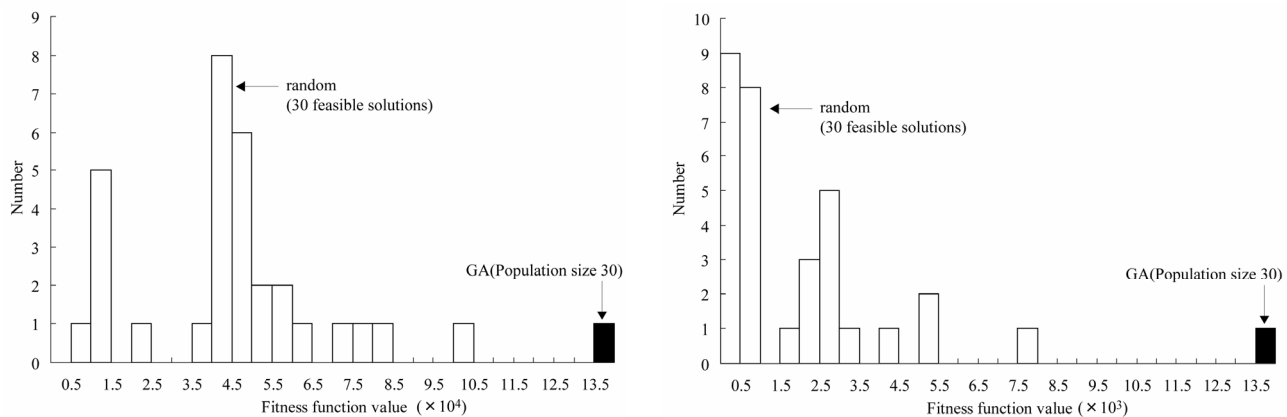


Figure 10. Comparison of fitness function value between the near-optimum solution and the 30 feasible solutions generated at random (Experiment 1, music file: jazz). Left figure: MP3 bit rate; 32 kbps. Right figure: MP3 bit rate; 64 kbps.

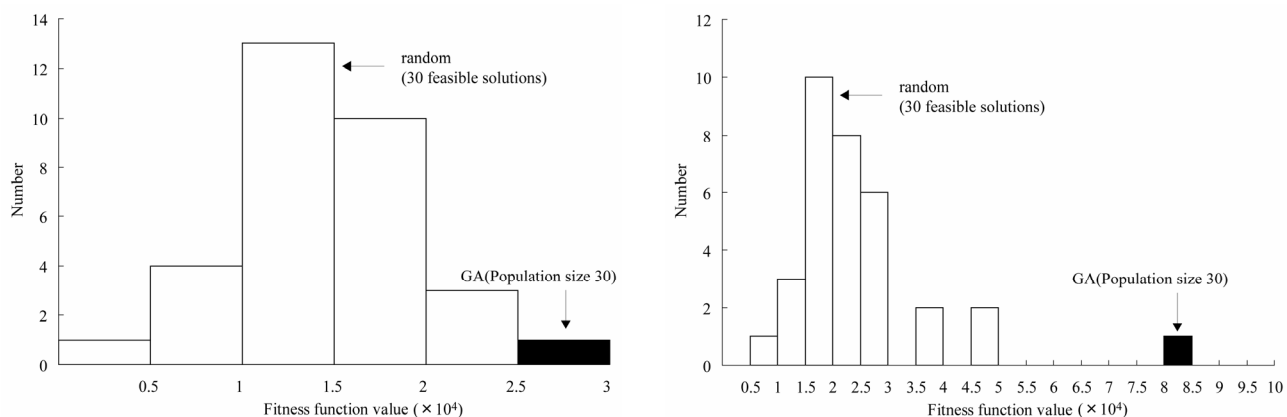


Figure 11. Comparison of fitness function value between the near-optimum solution and the 30 feasible solutions generated at random (Experiment 1, music file: popular). Left figure: MP3 bit rate; 32 kbps. Right figure: MP3 bit rate; 64 kbps.

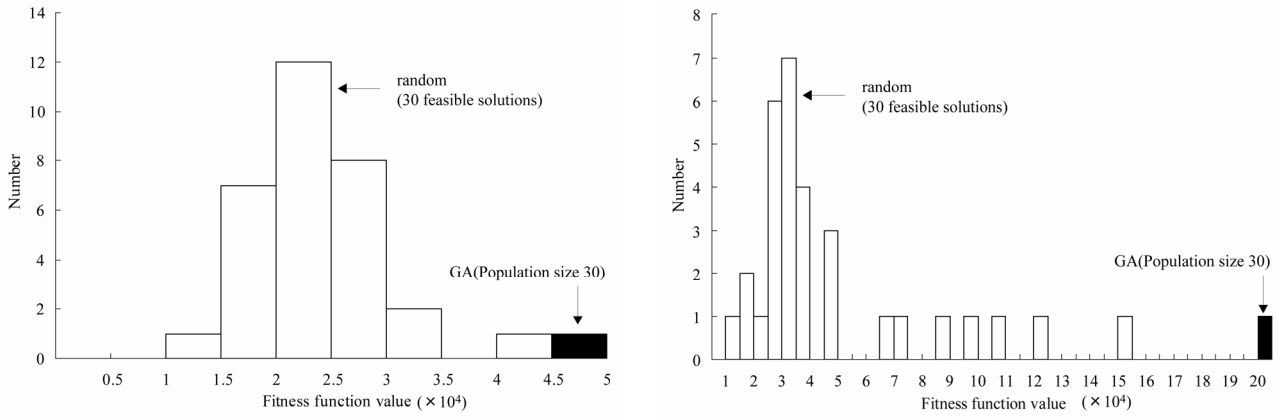


Figure 12. Comparison of fitness function value between the near-optimum solution and the 30 feasible solutions generated at random (Experiment 1, music file: rock). Left figure: MP3 bit rate; 32 kbps. Right figure: MP3 bit rate; 64 kbps.

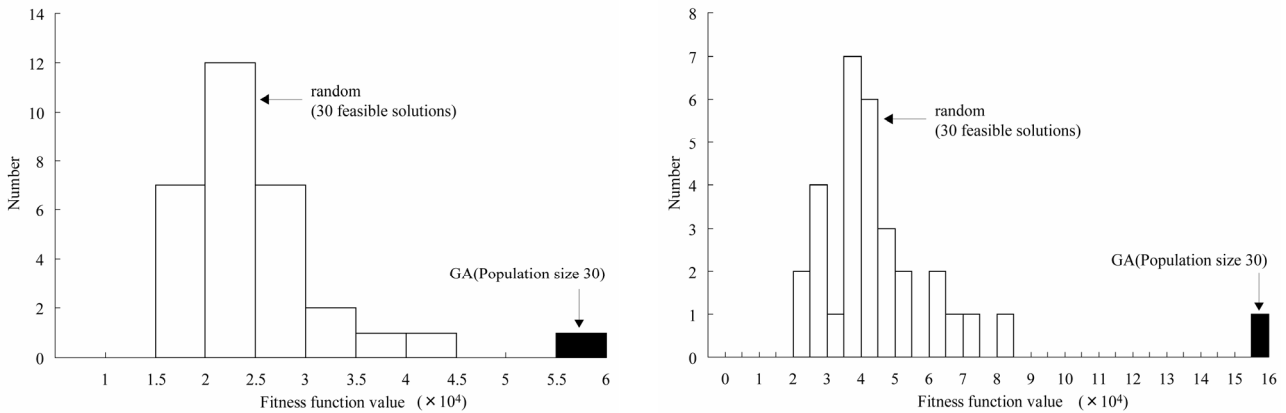


Figure 13. Comparison of fitness function value between the near-optimum solution and the 30 feasible solutions generated at random (Experiment 1, music file: hip-hop). Left figure: MP3 bit rate; 32 kbps. Right figure: MP3 bit rate; 64 kbps.

problem P' (Table 6). Practically, the watermarking decided by a near-optimum solution for the partial problem P' is recommended. In addition, an initial solution for the partial problem P' can be considered to be a secret key. Therefore, the partial problem P' has an advantage over the original problem P from the viewpoint of watermarking security.

7.4.3. Comparison with Another Technique

For making the technical level of the proposed method clear, we compared the results by the proposed method with those by our reported method [17,18]. In our reported study, another optimal watermarking problem was formulated using discrete Fourier transform and a watermarking method proposed in the reported study [19]. In our reported study, a string composed of 92 randomly generated bits was used as the watermark information, the same clips as used in the present study were used, and SNR_{seg} was measured using the same condition as that used in the present study. Table 7 shows the toler-

ance of watermark to the compression in using our reported method. Comparing Table 7 with Table 3, it is clear that the proposed method had higher tolerance to compression by each of MP3, AAC, and MWA than that by our reported study. As shown in Table 8, SNR_{seg} after embedding the watermark on the condition of the near-optimum solution obtained by our reported method was 36.8 to 52.2 dB. Although the amount of watermark information used in the present study was more than 4 times to that of our reported study, the proposed method realized lower noise than that by our reported method (Tables 5 and 8).

7.4.4. Discussion for Practical Setup

It takes much time to apply our technique to a song of 3 - 5 minutes as one clip. For practical usage of our approach, we will select some short clips of 10 second, for example. Then, we will use our approach to each short clip. The methodology for effective selections of short clips in a song is our next target.

Table 3. Watermarking tolerance to compression measured by detection rate (%) in Experiment 1.

(a) MP3 bitrate as GA condition: 32 kbps

Compression		Classical	Jazz	Popular	Rock	Hiphop
Method	Bitrate (kbps)					
MP3	128	99.75	99.75	99.5	99.25	99.25
	96	99.75	99.75	99.5	99.25	99.25
	64	99.25	99.25	98.25	97.25	98.25
	32	93.75	93.5	98.25	97.25	97.75
AAC	128	99	99	98.25	97	97.25
	96	89.75	91.15	95.25	94.5	93.25
	64	87.25	87	92.5	93	91.5
WMA	128	99.75	99.75	99.5	99.75	99.25
	96	98.75	97.75	99.5	98.75	99
	64	98.25	95.5	97.25	96	96.25
(%)						

(b) MP3 bitrate as GA condition: 64 kbps

Compression		Classical	Jazz	Popular	Rock	Hiphop
Method	Bitrate (kbps)					
MP3	128	100	100	99.25	100	98.25
	96	99.25	99.25	98.75	97.75	97.5
	64	95.5	95.5	95.25	92	96.25
	32	80.25	80.25	88.5	75.5	88.5
AAC	128	95.25	95.25	93.5	87.75	96.25
	96	80.25	80.25	87	74.75	91.5
	64	77.75	77.75	80.25	67.25	84.5
WMA	128	100	99	99.75	97.75	98.5
	96	97.75	93	96.25	92.75	94.75
	64	94.25	89	94	89	90.25
(%)						

Table 4. Time (sec.) to obtain a near-optimum solution in Experiment 1.

MP3 bitrate	32 kbps	64 kbps
Classical	4.02×10^2	3.98×10^2
Jazz	4.20×10^2	2.70×10^2
Popular	1.16×10^3	3.93×10^2
Rock	4.67×10^2	3.17×10^2
HipHop	5.19×10^2	5.12×10^2

Table 5. SNR_{seg} [dB] after embedding a watermark on the condition of near-optimum solution obtained by the proposed method in Experiment 1.

MP3 bitrate	32 kbps	64 kbps
Classical	70.1	73.2
Jazz	76.3	78.6
Popular	69.7	73.9
Rock	72.9	78.1
HipHop	73.6	71.1

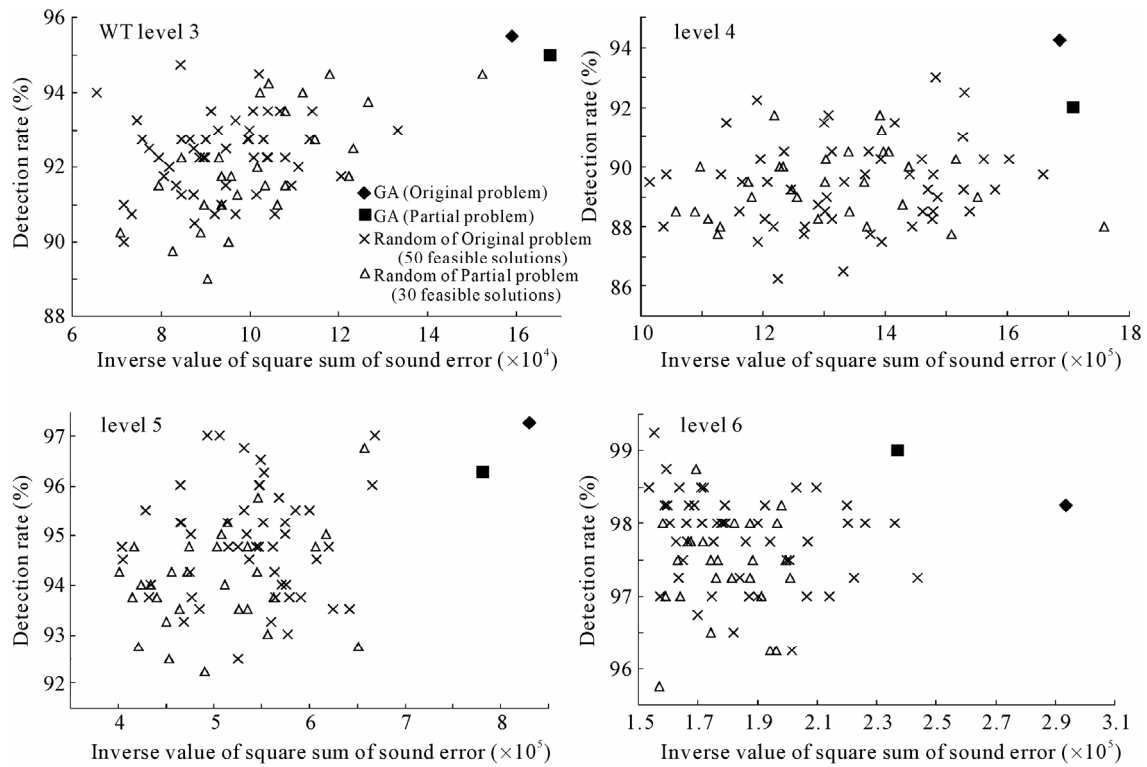


Figure 14. Comparison between the near-optimum solutions and the solutions generated at random (Experiment 2, music file: classical).

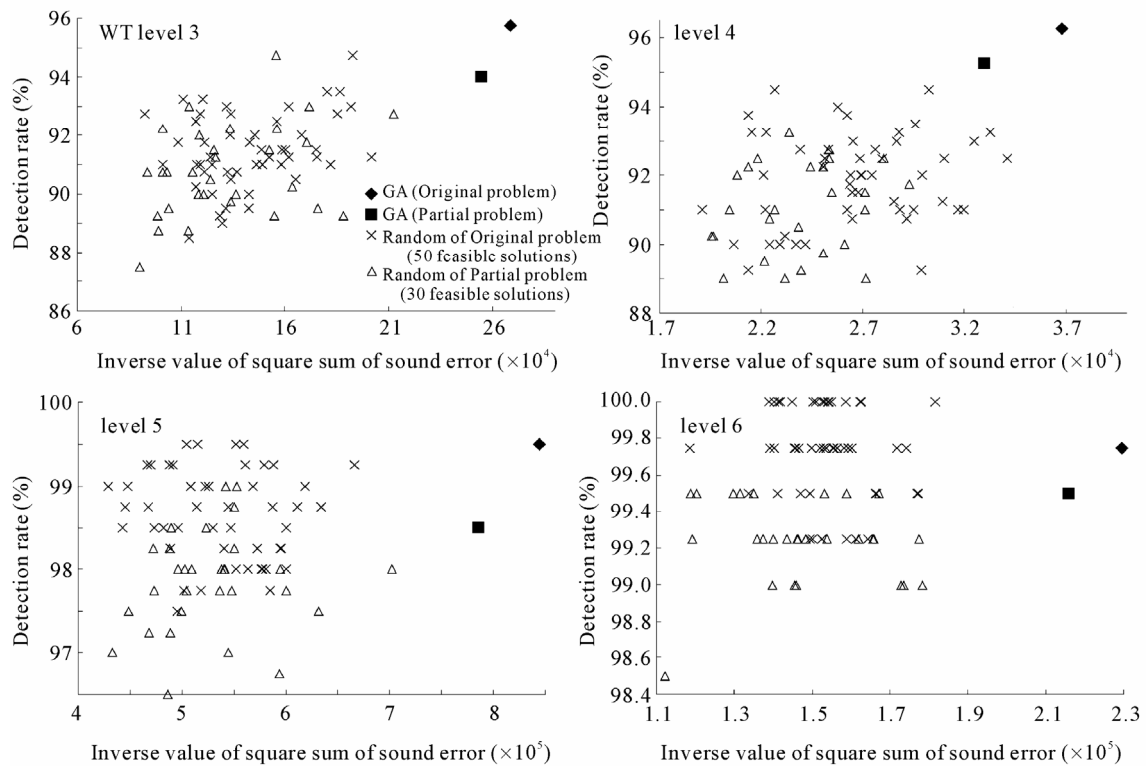


Figure 15. Comparison between the near-optimum solutions and the solutions generated at random (Experiment 2, music file: jazz).

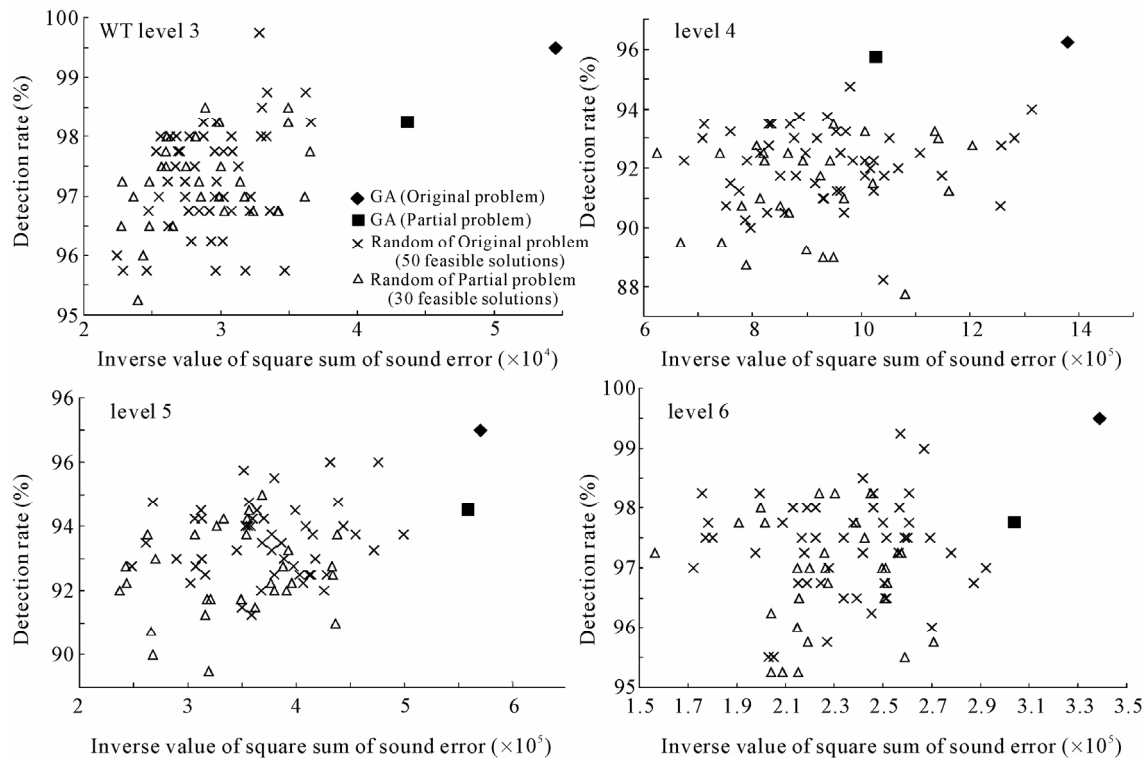


Figure 16. Comparison between the near-optimum solutions and the solutions generated at random (Experiment 2, music file: popular).

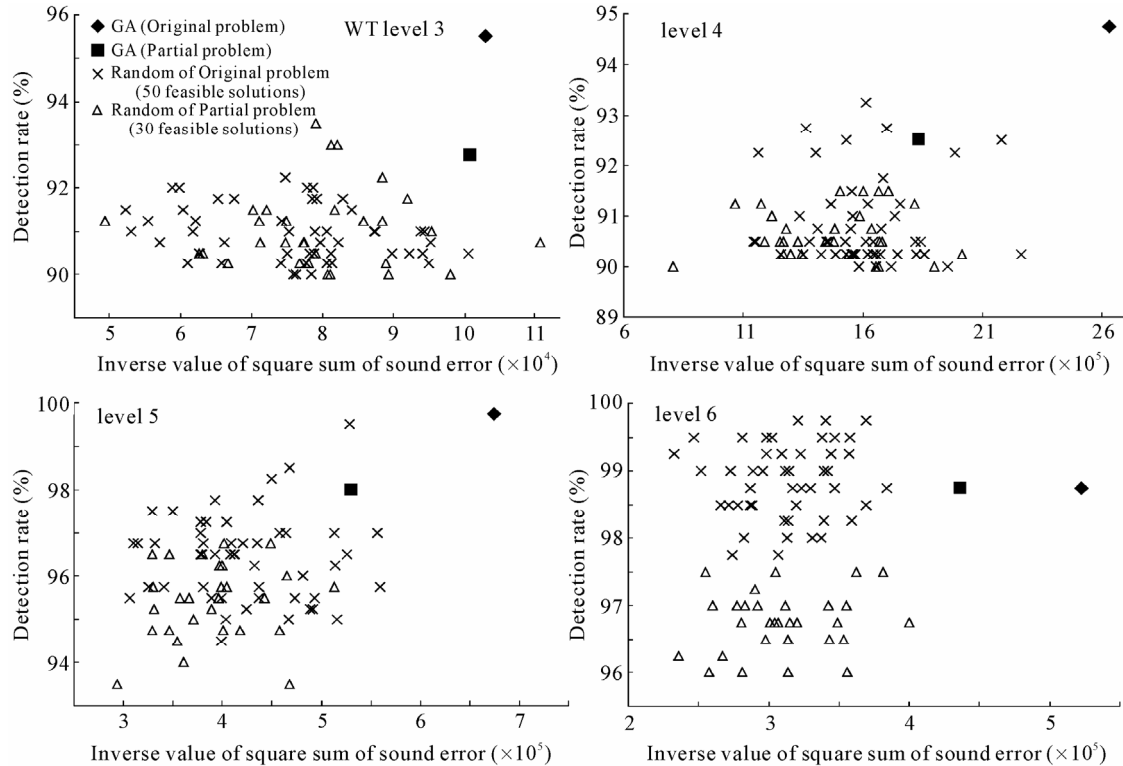


Figure 17. Comparison between the near-optimum solutions and the solutions generated at random (Experiment 2, music file: rock).

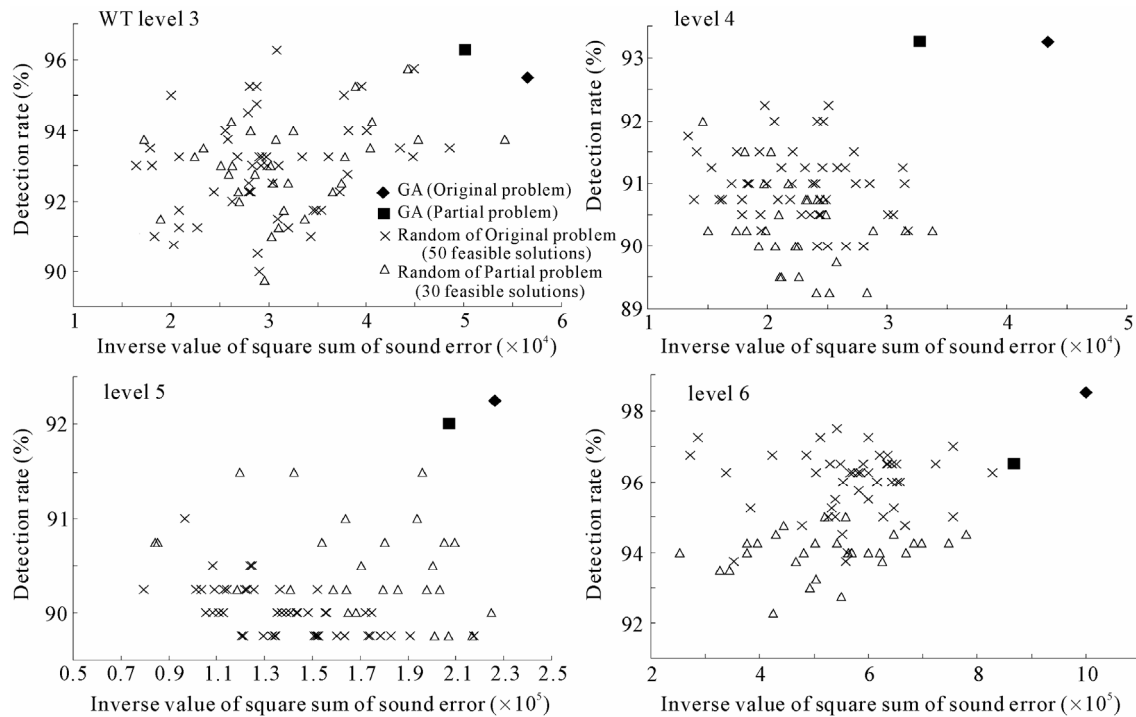


Figure 18. Comparison between the near-optimum solutions and the solutions generated at random (Experiment 2, music file: hip-hop).

It is difficult to model the compression attack in general. In this paper, the framework of watermark optimization is shown on the viewpoint of realizing good quality and high tolerance to MP3 compression. Because MP3 compression is one of the easiest attacks to watermark, we have selected it as an example. In addition, the high tolerance of watermark by the proposed method to each of AAC and WMA compression, which are representative audio compression techniques, is also shown in this paper.

8. Conclusions

A method for embedding digital watermark using DWT and GA to realize high tolerance to compression by MP3 is proposed. The proposed method enables us to embed digital watermark in a near-optimum manner for each music audio clip. Moreover, the near-optimum solution for the original problem P and partial problem P' , and the initial pattern of digital watermark for P' are used as secret keys in extracting the digital watermark. The experimental results show that the proposed method generates watermarked audio of good quality and high tolerance to MP3 compression.

9. Acknowledgements

The authors would like to thank Associate Professor H.

Okuhara of the Graduate School of Osaka University for his valuable advice on this research. The authors would also like to thank Associate Professor M. Tabuse of Kyoto Prefectural University for his valuable support of this research.

10. References

- [1] D. Kirovski and H. S. Malvar, "Spread-Spectrum Watermarking of Audio Signals," *IEEE Transactions on Signal Processing*, Vol. 51, No. 4, April 2003, pp. 1020-1033. [doi:10.1109/TSP.2003.809384](https://doi.org/10.1109/TSP.2003.809384)
- [2] I.-K. Yeo and H. J. Kim, "Modified Patchwork Algorithm: A Novel Audio Watermarking Scheme," *IEEE Transactions on Speech and Audio Processing*, Vol. 11, No. 4, July 2003, pp. 381-386. [doi:10.1109/TSA.2003.812145](https://doi.org/10.1109/TSA.2003.812145)
- [3] S. Wu, J. Huang, D. Huang and Y. Q. Shi, "Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission," *IEEE Transactions on Broadcasting*, Vol. 51, No. 1, March 2005, pp. 69-76. [doi:10.1109/TBC.2004.838265](https://doi.org/10.1109/TBC.2004.838265)
- [4] X. Y. Wang and H. Zhao, "A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT," *IEEE Transactions on Signal Processing*, Vol. 54, No. 12, December 2006, pp. 4835-4840. [doi:10.1109/TSP.2006.881258](https://doi.org/10.1109/TSP.2006.881258)
- [5] S. Xiang and J. Huang, "Histogram-Based Audio WaterMarking against Time-Scale Modification and Cropping Attacks," *IEEE Transactions on Multimedia*, Vol. 9,

- No. 7, November 2007, pp. 1357-1372.
[doi:10.1109/TMM.2007.906580](https://doi.org/10.1109/TMM.2007.906580)
- [6] S. Kirbiz, A. N. Lemma, M. U. Celik and S. Katzenbeisser, "Decode-Time Forensic Watermarking of AAC Bitstreams," *IEEE Transactions on Information Forensics and Security*, Vol. 2, No. 4, December 2007, pp. 683-696.
[doi:10.1109/TIFS.2007.908194](https://doi.org/10.1109/TIFS.2007.908194)
 - [7] D. J. Coumou and G. Sharma, "Insertion, Deletion Codes with Feature-Based Embedding: A New Paradigm for Watermark Synchronization with Applications to Speech Watermarking," *IEEE Transactions on Information Forensics and Security*, Vol. 3, No. 2, June 2008, pp. 153-165.
[doi:10.1109/TIFS.2008.920728](https://doi.org/10.1109/TIFS.2008.920728)
 - [8] S. Xiang, H. J. Kimb, and J. Huang, "Audio Watermarking Robust against Time-Scale Modification and MP3 Compression," *Signal Processing*, Vol. 88, No. 10, October 2008, pp. 2372-2387.
[doi:10.1016/j.sigpro.2008.03.019](https://doi.org/10.1016/j.sigpro.2008.03.019)
 - [9] X. Y. Wang, P. P. Niu and H. Y. Yang, "A Robust, Digital-Audio Watermarking Method," *IEEE Multimedia*, Vol. 16, No. 3, July 2009, pp. 60-69.
[doi:10.1109/MMUL.2009.44](https://doi.org/10.1109/MMUL.2009.44)
 - [10] N. K. Kalantari, M. A. Akhaee, S. M. Ahadi and H. Amindavar, "Robust Multiplicative Patchwork Method for Audio Watermarking," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 17, No. 6, August 2009, pp. 1133-1141.
 - [11] X. Y. Wang, P. P. Niu and H. Y. Yang, "A Robust Digital Audio Watermarking Based on Statistics Characteristics," *Pattern Recognition*, Vol. 42, No. 11, November 2009, pp. 3057-3064.
 - [12] K. Yamamoto and M. Iwakiri, "Real-Time Audio Watermarking Based on Characteristics of PCM in Digital Instrument," *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 1, No. 2, April 2010, pp. 59-71.
 - [13] D. Goldberg, "Genetic Algorithm in Search, Optimization, and Machine Learning," Addison-Wesley, Reading, Boston, 1989.
 - [14] J. H. Holland, "Adaptation in Natural and Artificial Systems," The University Michigan Press, Ann Arbor, 1975, and MIT Press, Cambridge, 1992.
 - [15] M. Shino, Y. Choi and K. Aizawa, "Wavelet Domain Digital Watermarking Based on Threshold-Variable Decision," *Technical Report of IEICE, DSP2000-86*, in Japanese, Vol. 100, No. 325, September 2000, pp. 29-34.
 - [16] D. Inoue and Y. Yoshitomi, "Watermarking Using Wavelet Transform and Genetic Algorithm for Realizing High Tolerance to Image Compression," *Journal of the Institute of Image Electronics Engineers of Japan*, Vol. 38, No. 2, March 2009, pp. 136-144.
 - [17] M. Tanaka and Y. Yoshitomi, "Digital Audio Watermarking Method with MP3 Tolerance Using Genetic Algorithm," *Proceedings of the 2006 IEICE General Conference*, Tokyo, 21 March 2006, p. 182.
 - [18] M. Tanaka and Y. Yoshitomi, "Digital Audio Watermarking Method with MP3 Tolerance Using Genetic Algorithm," *Proceedings of 11th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, Sendai, 15-17 September 2008, pp. 81-85.
 - [19] R. Tachibana, "Capacity Analysis of Audio Watermarking Based on Logarithmic Amplitude Modification against Additive Noise," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, in Japanese, Vol. J86-A, No. 11, November 2003, pp. 1197-1206.
 - [20] F. Glover, "Future Paths for Integer Programming and Links to Artificial Intelligence," *Computers and Operations Research*, Vol. 13, No. 5, May 1986, pp. 533-549.
[doi:10.1016/0305-0548\(86\)90048-1](https://doi.org/10.1016/0305-0548(86)90048-1)
 - [21] M. Goto, H. Hashiguchi, T. Nishimura and R. Oka, "RWC Music Database: Database of Copyright-Cleared Musical Pieces and Instrument Sounds for Research Purposes," *Transactions of IPSJ*, in Japanese, Vol. 45, No. 3, March 2004, pp. 728-738