

An Authentication Method for Digital Audio Using a Discrete Wavelet Transform

Yasunari Yoshitomi, Taro Asada, Yohei Kinugawa, Masayoshi Tabuse

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

E-mail: yoshitomi@kpu.ac.jp

Received November 11, 2010; revised February 15, 2011; accepted February 26, 2011

Abstract

Recently, several digital watermarking techniques have been proposed for hiding data in the frequency domain of audio files in order to protect their copyrights. In general, there is a tradeoff between the quality of watermarked audio and the tolerance of watermarks to signal processing methods, such as compression. In previous research, we simultaneously improved the performance of both by developing a multipurpose optimization problem for deciding the positions of watermarks in the frequency domain of audio data and obtaining a near-optimum solution to the problem. This solution was obtained using a wavelet transform and a genetic algorithm. However, obtaining the near-optimum solution was very time consuming. To overcome this issue essentially, we have developed an authentication method for digital audio using a discrete wavelet transform. In contrast to digital watermarking, no additional information is inserted into the original audio by the proposed method, and the audio is authenticated using features extracted by the wavelet transform and characteristic coding in the proposed method. Accordingly, one can always use copyright-protected original audio. The experimental results show that the method has high tolerance of authentication to all types of MP3, AAC, and WMA compression. In addition, the processing time of the method is acceptable for everyday use.

Keywords: Authentication, Audio, Copyright Protection, Tolerance to Compression, Wavelet Transforms

1. Introduction

Recent progress in digital media technology and distribution systems, such as the Internet and cellular phones, has enabled consumers to easily access, copy, and modify digital content, such as electric documents, images, audio, and video. Therefore, techniques to protect the copyrights for digital data and prevent unauthorized duplication or tampering are urgently needed.

Digital watermarking (DW) is a promising method of copyright protection for digital data. Several studies have investigated audio DW [1-12]. Two important properties of audio DW are inaudibility of DW-introduced distortion, and robustness to signal processing methods, such as compression. In addition to these properties, the data rate and complexity of the DW have attracted attention when discussing the performance of a DW.

We have attempted to develop a method in which 1) the DW can be sufficiently extracted from the watermarked audio, even after compression, and 2) the quality of the audio remains high after embedding the DW. How-

ever, there is generally a tradeoff between these two properties. Therefore, we focus on this tradeoff and attempt to overcome this critical difficulty by optimizing the positions of the DW in the frequency domain. Recently, digital audio distributed over the Internet and cellular phone systems is often modified by compression, which is one of the easiest and most effective ways to defeat a DW without significantly deteriorating the quality of the audio.

In previous research, we simultaneously improved both the extraction performance of the DW and the quality of the DW-contained audio by developing a multipurpose optimization problem for deciding the positions of the DW in the frequency domain and obtaining a near-optimum solution for the problem using a discrete wavelet transform (DWT) and a genetic algorithm (GA) for realizing high tolerance to MP3 compression, which is the most popular compression technique [13,14]. Our method enabled us to embed the DW in an almost optimal manner within any digital audio. However, obtaining the near-optimum solution was very time consuming. In the

present study, to overcome this issue essentially, we have developed an authentication method for digital audio to protect the copyrights. In contrast to the DW, no additional information is inserted into the original audio by the proposed method, and the digital audio is authenticated using features extracted using the DWT and characteristic coding of the proposed method. This paper presents an analysis of the performance of the method.

2. Wavelet Transform

The original audio data $s_k^{(0)}$, which is used as the level-0 wavelet decomposition coefficient sequence, where k denotes the element number in the data, are decomposed into the multi-resolution representation (MRR) and the coarsest approximation by repeatedly applying the 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 using (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_k and q_k 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 using (1) and (2) iteratively from $j=0$ to $j=J-1$. **Figure 1** shows the process of multi-resolution analysis by DWT.

In the present study, we use the Daubechies wavelet for the DWT, according to the references [14,15]. As a result, we obtain the following relation between p_k and q_k :

$$q_k = (-1)^k p_{1-k} \quad (3)$$

We select the Daubechies wavelet because we compared the results by the proposed method with those by our retorted method [13,14], where the Daubechies wavelet was used for the DW.

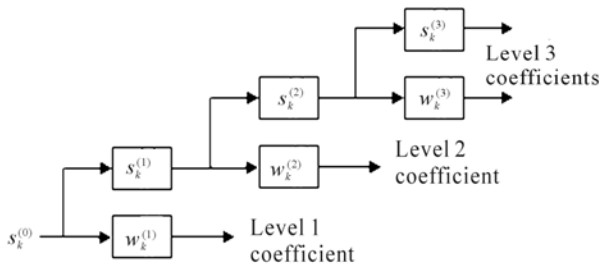


Figure 1. Multi-resolution analysis by the DWT.

3. Proposed Authentication Algorithm

It is known that the histogram of the wavelet coefficients of each domain of the MRR sequences has a distribution which is centered at approximately 0 when the DWT is performed on audio data, as shown in **Figure 2** [13]. In the present study, the above phenomenon is exploited for the authentication of the audio signal. The procedure is described below.

3.1. Selective Coding

3.1.1. Setting of Parameters

For the coding of the audio, we obtain the histogram of the wavelet coefficients V at the selected level of an MRR sequence (see **Figure 3**). Like the DW techniques for images [15,16] and digital audio [13,14], we set the following coding parameters:

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

$$1) T1 < Th(minus) < T2 < 0 < T3 < Th(plus) < T4.$$

2) The value of S_{T1} , the number of wavelet coefficients in $(T1, Th(minus))$, is equal to S_{T2} , the number of wavelet coefficients in $[Th(minus), T2)$. In short, $S_{T1} = S_{T2}$.

3) The value of S_{T3} , the number of wavelet coefficients in $(T3, Th(plus)]$, is equal to S_{T4} , the number of wavelet coefficients in $(Th(plus), T4)$. In short, $S_{T3} = S_{T4}$.

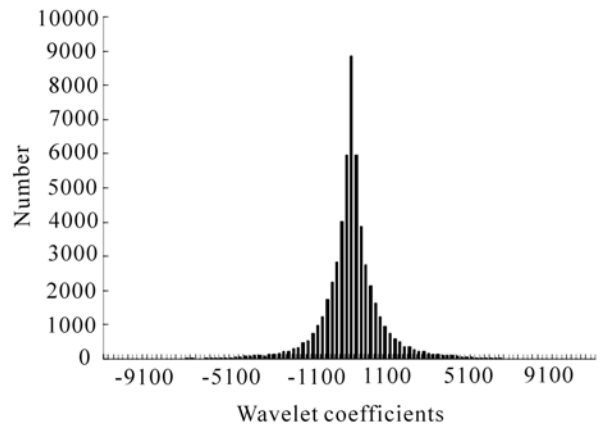


Figure 2. Histogram of the wavelet coefficients of an MRR sequence at level 3 (jazz) [13].

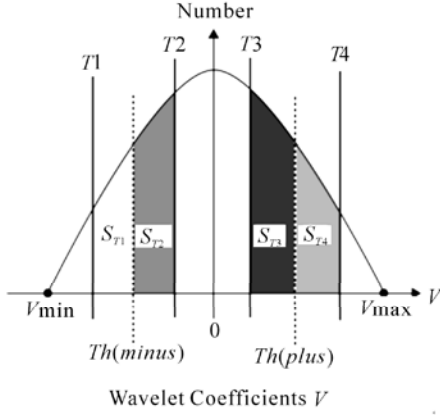


Figure 3. Schematic diagram of the histogram of MRR wavelet coefficients.

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

In the present study, the values of both S_{T1}/S_m and S_{T3}/S_p are set to 0.3, which was determined experimentally.

3.1.2. Domain Segmentation in a Wavelet Coefficient Histogram

In preparation of a coding for authentication, the procedure separates the wavelet coefficients V of an MRR sequence into five sets (hereinafter referred to as A, B, C, D, and E), as shown in **Figure 4**, using the following criteria:

- $A = \{V \mid V \in V^{SC}, V \leq T1\},$
- $B = \{V \mid V \in V^{SC}, T1 < V < T2\},$
- $C = \{V \mid V \in V^{SC}, T2 \leq V \leq T3\},$
- $D = \{V \mid V \in V^{SC}, T3 < V < T4\},$
- $E = \{V \mid V \in V^{SC}, T4 \leq V\},$

where V^{SC} is the set of wavelet coefficients in the original audio file.

3.1.3. Selective Coding Algorithm

The wavelet coefficients of an MRR sequence are coded according to the following rules, in which V_i denotes one of wavelet coefficients:

When $V_i \in C$, flag f_i is set to be 1, and bit b_i is set to be 0.

When $V_i \in (A \cup E)$, flag f_i is set to be 1, and bit b_i is set to be 1.

When $V_i \in (B \cup D)$, flag f_i is set to be 0, and bit b_i is set to be 0.5.

For the authentication of the digital audio, we use a

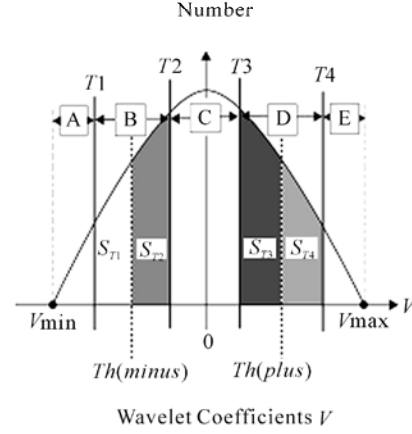


Figure 4. Five sets (A, B, C, D, and E) are described by a histogram of wavelet coefficients V of an MRR sequence for the assignment of a bit.

code C (hereinafter referred to as an original code), which is the sequence of b_i defined above. For the coding and authentication, we assign a sequence number and a flag for each wavelet coefficient. The flag $f_i = 1$ for a V_i means that the V_i is assigned a bit ($b_i = 0$ or 1) for a coding. The flag $f_i = 0$ for a V_i provides that the V_i is not assigned a bit of 0 or 1: b_i is externally set to be 0.5 as an arbitrary constant and the value of b_i does not influence the performance of the proposed method described in Section 3.2. The exclusion of all V_i belonging to the sets B and D, where the magnitude of the V_i are intermediate, from the objects for coding is a novel feature of the present study.

3.2. Authentication

3.2.1. Setting of Parameters

We authenticate not only an original digital audio file but also a signal-processed version. Compression, one example of signal processing, is often applied to digital audio for the purposes of distribution via the Internet or for saving on a computer. Through the same procedure as described in Section 3.1, we applied the DWT to digital audio and obtained a histogram of wavelet coefficients V' at the same level of the DWT as that of the coding for the original audio file, which is described in Section 3.1. Then, we set the authentication parameters as follows:

The values of $Th'(minus)$ and $Th'(plus)$ (see **Figure 5**) are chosen such that the non-positive wavelet coefficients (S'_m in total frequency) are equally divided into two groups by $Th'(minus)$, and the positive wavelet coefficients (S'_p in total frequency) are equally divided into two groups by $Th'(plus)$. Next, the values of $T1'$, $T2'$, $T3'$, and $T4'$, the parameters for controlling the authentication precision, 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 wavelet coefficients in $(Th'(\text{plus}), T4')$. 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 be 0.3, the same as the settings used for the coding for the original audio file, which is described in Section 3.1.

3.2.2. Domain Segmentation in a Wavelet Coefficient Histogram

In the preparation of a coding for authentication, the procedure separates the wavelet coefficients V' of an MRR sequence into three sets (hereinafter referred to as F, G, and H), as shown in Figure 5, using the following criteria:

- $F = \{V' | V' \in V'^{AC}, V' < Th'(\text{minus})\}$,
- $G = \{V' | V' \in V'^{AC}, Th'(\text{minus}) \leq V' \leq Th'(\text{plus})\}$,
- $H = \{V' | V' \in V'^{AC}, Th'(\text{plus}) < V'\}$,

where V'^{AC} is the set of wavelet coefficients of the a target audio file for making the code for authentication.

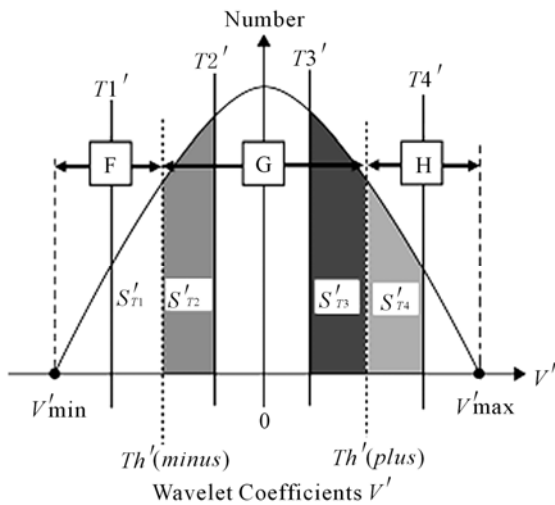


Figure 5. Three sets (F, G, and H), indicated on the histogram, of MRR wavelet coefficients used for the authentication.

3.2.3. Authentication Algorithm

The wavelet coefficients of an MRR sequence are coded according to the following rules, in which V'_i denotes one of wavelet coefficients:

When $f_i = 1$ and $V'_i \in G$, bit b'_i is set to be 0.

When $f_i = 1$ and $V'_i \in (F \cup H)$, bit b'_i is set to be 1.

When $f_i = 0$, bit b'_i is set to be 0.5.

When $f_i = 0$, b'_i is externally set to be 0.5 as an arbitrary constant and the value of b_i does not influence the performance of the proposed method described below.

For the authentication of the digital audio, we use the code C' (hereinafter referred to as an authentication code), which is the sequence of b'_i defined above. The authentication ratio AR (%) is defined by the following:

$$AR = \frac{100 \sum_{i=1}^N f_i (1 - |b'_i - b_i|)}{\sum_{i=1}^N f_i} \quad (4)$$

where N is the number of wavelet coefficients assigned flags in the coding for the original audio file, which is described in Section 3.1. According to (4), the values of neither b_i nor b'_i influence the value of AR in the case that $f_i = 0$, which provides that the corresponding V'_i is not assigned a bit of 0 or 1 for the coding for the original audio file.

For using the proposed method, we need store flags f_i and an original code C of each audio file whose copy right we should protect. In calculating (4) for the authentication of an original audio file, we do not use an original audio file but the flags f_i and the original code C of the original audio file.

4. Experiment

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

4.1. Method

The experiment was performed in the following computational environment: the personal computer was a Dell Dimension 8300 (CPU: Pentium IV 3.2 GHz; main memory: 2.0 GB); the OS was Microsoft Windows XP; the development language was Microsoft Visual C++ 6.0.

Five music audio files, which were composed of the first entries in the five genre categories: classical, jazz, popular, rock, and hiphop in the music database RWC for research purpose [17], were copied from CDs onto a 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-second clip of music audio (hereinafter referred to as an original music audio clip) was extracted starting at 1 minute from the beginning of the audio file and saved on a personal computer. In addition, for each of the five music audio files mentioned above, several 10-second audio clips were extracted by shifting the start-time 1 second at a time from the beginning of the audio file and were saved on a personal computer for use in evaluating the performance of the proposed method. For the purpose of evaluating the tolerance of authentication to compression, MP3, AAC, and WMA compression systems were each used to compress the original music audio clip to bit rates of 64, 96, and 128 kbps. The process of the experiment was as follows: obtain the code of the original WAVE file, compress the file by MP3, AAC, and WMA, and then convert the compressed files into the WAVE files used for the authentication.

For the DWT, we use Daubechies wavelets. Level 8 was chosen as the standard for the DWT based on an analysis of preliminary experiments, and this level was used for most of the experiments. The influence of the level of the DWT on the authentication ratio was also analyzed as part of the experiments.

Instead of the Dell Dimension 8300 (CPU: Pentium IV 3.2 GHz; main memory: 2.0 GB), Dell Dimension DXC051 (CPU: Pentium IV 3.0 GHz, memory: 1.0 GB) is used only for the comparison with the reported study [13,14].

4.2. Results and Discussion

4.2.1. Authentication Process

Table 1 illustrates the process of authentication of audio clips. The jumps in the wavelet coefficient number, such as from 573 to 578, indicate that the intervening wavelet coefficients belong to either the set B or D, which are out of assignment of a bit to 0 or 1 for the coding for the original music audio clip. The authentication ratio AR defined by (4) was $(6/7) \times 100$ in the case of **Table 1**, where the bits of the music audio clip after MP3 compression were equal to those of the original music except for wavelet coefficient number 579.

4.2.2. Robustness to Compression

Whenever we applied the proposed method to the five original music audio clips, the authentication ratio was 100%. When we applied it to several music clips compressed by MP3, AAC, and WMA, the authentication ratio was at or near 100% (**Table 2**).

4.2.3. Authentication Ratios for Other

Non-Signal-Processed Music Audio Clips

The purpose of authentication is to protect the copyright

Table 1. Authentication process (hiphop).

Wavelet coeffi- cients No.	Original		After MP3 compression (128kbps)		Bit correspondence
	set	bit	set	bit	
572	C	0	G	0	Yes
579	C	0	F	1	No
580	C	0	G	0	Yes
584	C	0	G	0	Yes
588	E	1	H	1	Yes
589	C	0	G	0	Yes
590	C	0	G	0	Yes

Table 2. Authentication tolerance to compression (%).

Method	Compression		Classical	Jazz	Popular	Rock	Hiphop
	Bit rate (kbps)						
MP3	128	99.86	99.86	99.86	100	99.57	
	96	99.86	100	99.86	100	99.86	
	64	99.72	99.72	99.86	100	98.28	
AAC	128	100	100	100	100	97.55	
	96	99.86	100	99.86	100	100	
	64	100	100	99.14	99.72	99.00	
WMA	128	100	100	100	100	100	
	96	100	100	100	100	100	
	64	100	99.86	100	100	100	
(%)							

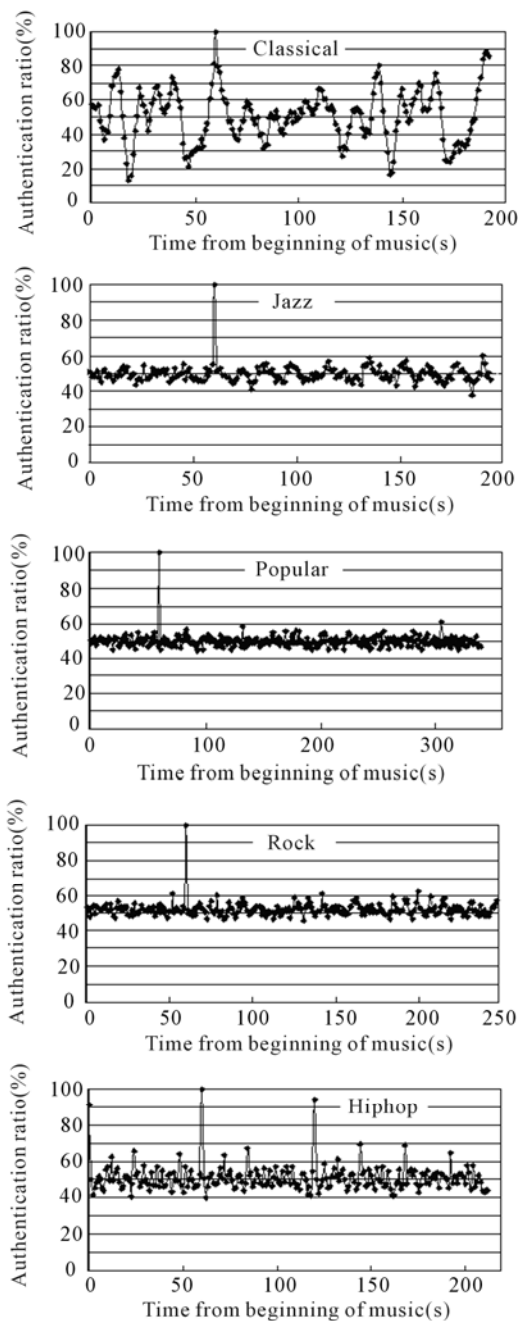
on audio data. When the music audio file targeted for being authenticated was different from that used for making the code of the original music audio clip, the authentication ratios AR defined by (4) were about 50% (more precisely, they fell in the range 44.09 to 55.62%), which was about half of the authentications ratios when authenticating the same clip as the original music audio clip (100% in all cases in this experiment; see **Table 3**). An authentication ratio of 50% corresponds to the value in the case that randomly generated bits are used for b_i and/or b'_i in (4). Accordingly, the proposed method distinguishes an original music audio clip from each of the other four used in this experiment.

Using the original code obtained from the original music audio clip, which was the 10-second clip extracted staring at 1 minute from the beginning of each of the five music audio files, we calculated the authentication ratio to the 10-second clips extracted by shifting the start-time for the clip 1 second at a time. For each of the original music audio clips, the authentication ratio was 100% when an original code was used as the authentication code (**Figure 6**). In **Figure 6**, the point 60 seconds on the horizontal axis corresponds to the case that the original code is used as the authentication code. Not including these cases, the authentication ratio for jazz, popular, and rock music audio fell mostly in the 40 to 60% range. In contrast, the authentication ratio for classical and hiphop varied according to the start time. Not including the case of using the original code for the authentication code, the highest authentication ratio was 93.95%, which was ob-

Table 3. Authentication ratio (%) in all combinations of original and authentication.

		Original				
		Classical	Jazz	Popular	Rock	Hiphop
Authenti- cation	Classical	100	44.22	47.98	52.89	53.61
	Jazz	55.62	100	49.86	51.01	46.54
	Popular	44.09	53.89	100	49.57	52.88
	Rock	47.69	50.86	49.42	100	48.13
	Hiphop	48.41	50.43	47.55	51.44	100

(%)

**Figure 6. Authentication ratios using clips shifted 1 second at a time for each of the five selected audio clips.**

served for hiphop. Accordingly, the threshold of the authentication ratio for judging authentication of an original music audio clip should be about 95%. As the authentication ratios to music clips extracted from the music audio files, from which the original music audio clip were obtained, stayed under 95% (again, excluding the cases of using the identical clip), we conclude that the probability of getting an authentication ratio above 95% would be small if we applied the proposed method to other music selected from the database. In other words, we propose that music audio be judged as authenticated when the file gives an authentication ratio of 95% or higher for a certain clip taken from a music audio file. When we used 95% as a threshold for the authentication ratio, both the false negative and positive rates for the authentication of the music audio clip were zero in the both cases shown in **Table 3** and **Figure 6**.

4.2.4. Influence of DWT Level on Authentication Ratio

All authentication ratios described above were obtained using a DWT at level 8. The tolerances of the authentication ratio to signal processing by MP3, AAC, and WMA at DWT levels of 2 to 8 with bit rates of 128, 96, and 64 kbps are shown for each bit rate in **Tables 4-6**, respectively. The authentication ratio does not noticeably change at bit rates of from 64 to 128 kbps. The authentication ratio tends to be slightly higher with increases in the DWT level of the original coding, which is the same as that of the authentication coding. For DWT levels of 7 or 8, the authentication ratio exceeds 95% for all settings of MP3, AAC, and WMA compression tested. The lowest authentication ratio, 94.57%, occurred for DWT level 6 applied to the hiphop audio clip compressed by AAC with a bit rate of 128 kbps. The number of data of the original music audio clip, which is treated as the amount of data at DWT level 0, was 441,000. The number of wavelet coefficients of MRR sequences was reduced by half for an increase of DWT level by one, meaning that the number of 0 or 1 bits in both the original and the authentication coding was also reduced by half.

4.2.5. Comparison with Watermarking

There is generally a tradeoff between 1) the tolerance of the DW to signal processing, such as compression, and 2) the quality of the music audio after embedding the DW. In other words, to improve the first property tends to cause a deterioration of the second property. We had overcome this critical difficulty of the DW by optimizing the positions of the DW in the frequency domain [13,14], [18-20]. However, it took much time to get the condition for embedding the DW by the reported method.

Figure 7 shows the relationship between the quality of music audio and the detection rate of the DW after MP3 compression, using the jazz clip as the original music au-

Table 4. Authentication ratio (%) of music audio compressed by MP3, AAC, and WMA at DWT levels of 2 to 8 with a bit rate of 128 kbps.

(1) Classical			
DWT level	Signal processing		
	MP3	AAC	WMA
2	99.07	99.41	99.99
3	99.67	99.69	100
4	99.99	99.9	100
5	99.98	99.98	100
6	100	100	100
7	100	100	100
8	99.86	100	100
(2) Jazz			
DWT level	Signal processing		
	MP3	AAC	WMA
2	98.63	99.33	99.95
3	99.64	99.8	100
4	99.97	99.99	100
5	100	100	99.98
6	100	100	100
7	100	100	99.93
8	99.86	100	100
(3) Popular			
DWT level	Signal processing		
	MP3	AAC	WMA
2	92.78	95.08	99.95
3	95.05	98.5	100
4	97.37	99.7	99.99
5	98.53	99.98	100
6	99.71	100	100
7	100	100	100
8	99.86	100	100
(4) Rock			
DWT level	Signal processing		
	MP3	AAC	WMA
2	94.2	95.88	99.81
3	96.72	98.79	99.98
4	98.89	99.85	100
5	99.64	99.98	100
6	100	100	99.96
7	100	100	100
8	100	100	100
(5) Hiphop			
DWT level	Signal processing		
	MP3	AAC	WMA
2	95.65	61.31	99.6
3	96.27	72.89	99.69
4	96.85	84.3	99.69
5	98.19	91.64	99.89
6	99.67	94.57	100
7	99.93	96.67	100
8	99.57	97.55	100

(%)

Table 5. Authentication ratio (%) of music audio compressed by MP3, AAC, and WMA at DWT levels of 2 to 8 with a bit rate of 96 kbps.

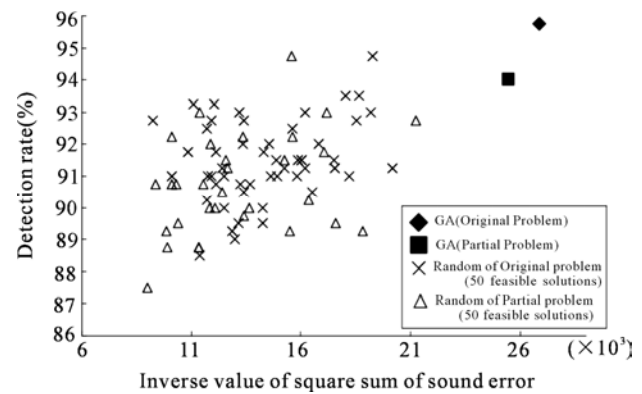
(1) Classical			
DWT level	Signal processing		
	MP3	AAC	WMA
2	95.95	98.76	99.96
3	96.92	99.08	99.99
4	97.89	99.48	100
5	98.53	99.64	100
6	99.53	99.78	100
7	99.71	100	100
8	99.71	99.86	100
(2) Jazz			
DWT level	Signal processing		
	MP3	AAC	WMA
2	96.48	98.62	99.97
3	98.42	99.39	100
4	99.79	99.95	100
5	100	99.98	99.98
6	100	100	100
7	100	100	99.93
8	100	99.86	100
(3) Popular			
DWT level	Signal processing		
	MP3	AAC	WMA
2	85.61	89.7	99.62
3	94.74	94.7	99.9
4	97.43	98.4	99.85
5	98.79	99.42	99.98
6	99.89	100	100
7	99.86	100	100
8	99.86	100	100
(4) Rock			
DWT level	Signal processing		
	MP3	AAC	WMA
2	89.79	92.56	99.33
3	95.37	96.3	99.78
4	98.58	98.94	99.98
5	99.69	99.8	100
6	99.96	99.89	99.96
7	100	100	100
8	100	100	100
(5) Hiphop			
DWT level	Signal processing		
	MP3	AAC	WMA
2	92.76	94.86	99.46
3	94.13	96.16	99.54
4	95.38	97.43	99.6
5	97.01	98.84	99.84
6	98.91	99.67	100
7	99.57	99.93	100
8	99.86	100	100

(%)

Table 6. Authentication ratio (%) of music audio compressed by MP3, AAC, and WMA at DWT levels of 2 to 8 with a bit rate of 64 kbps.

(1) Classical			
DWT level	Signal processing		
	MP3	AAC	WMA
2	98.18	98.08	99.88
3	98.87	98.51	99.96
4	99.54	99	99.98
5	99.91	98.73	100
6	99.96	99.13	100
7	99.93	99.42	100
8	99.86	100	100
(2) Jazz			
DWT level	Signal processing		
	MP3	AAC	WMA
2	92.41	97.23	99.85
3	95.45	98.45	100
4	98.39	99.54	100
5	99.93	99.95	99.98
6	100	99.89	100
7	100	99.86	99.93
8	99.71	100	100
(3) Popular			
DWT level	Signal processing		
	MP3	AAC	WMA
2	73.87	81.86	96.73
3	87.94	90.43	99.44
4	95.21	96.02	99.42
5	98.26	97.77	99.84
6	99.78	98.8	100
7	100	99.57	100
8	99.86	99.14	100
(4) Rock			
DWT level	Signal processing		
	MP3	AAC	WMA
2	82.67	88.43	96.12
3	91.55	93.67	99.14
4	97.09	97.56	99.69
5	99.4	98.91	99.95
6	99.93	99.17	99.96
7	100	99.35	100
8	100	99.71	100
(5) Hip-hop			
DWT level	Signal processing		
	MP3	AAC	WMA
2	82.13	92.95	97.86
3	86	94.35	98.85
4	89.31	95.82	98.78
5	92.62	96.7	99.31
6	97.43	97.9	99.86
7	98.48	98.26	99.86
8	98.27	98.99	99.86

(%)

**Figure 7. Relationship between sound quality after embedding the DW and detection rate of the DW [13].**

dio clip and 96-kbps MP3 compression [13]. The same original music audio clip was also used in the present experiment. In order to achieve a high detection rate of the DW and high quality of the original music audio clip after embedding the DW, the reported method using a genetic algorithm was effective, as shown in **Figure 7**. In the present study, the authentication ratio for the same original music audio clip as that used for getting the results of **Figure 7** was 100%, and a deterioration in the quality of the original music audio clip did not occur, which corresponds to an infinite value on the horizontal axis shown in **Figure 7**.

Moreover, it took 2.41×10^4 to 3.20×10^4 s and 1.59×10^2 to 1.85×10^3 s (with the personal computer referred to as PC2), respectively, to embed the DW using as the formula of the optimization problem the original problem and the partial problem (which had a much smaller search space) [14], while it took 2.05×10^{-1} to 2.10×10^{-1} s (with the personal computer referred to as PC2) and 2.03×10^{-1} to 2.19×10^{-1} s (with the personal computer referred to as PC1) for one coding for an original music audio clip in the present study (**Table 7**). In the reported study [13,14], the experiment was performed in the following computation environment: the personal computer was a Dell Dimension DXC051 (CPU: Pentium IV 3.0 GHz; main memory: 1.0 GB), which is referred to as PC2 in **Table 7**; the OS was Microsoft Windows XP; the development language was Microsoft Visual C++ 6.0. The average time for one coding for an original music audio clip was less than 10^{-5} times that to embed the DW using as the formula of the optimization problem the original problem, and less than 10^{-3} times that to embed the DW using as the formula of the optimization problem the partial problem in the reported study. In addition, no deterioration in quality of the original music audio clip ever occurred using the proposed method. These two factors strongly suggest that the proposed method is far superior to the reported method.

Table 7. Comparison of time(s) to obtain a coding of the proposed method or to embed the DW using as the formula of the optimization problem the original problem and the partial problem of the reported study [14].

	Coding		DW	
			Original problem	Partial problem
	PC1	PC2	PC2	PC2
Classical	2.19×10^{-1}	2.05×10^{-1}	3.20×10^4	1.04×10^3
Jazz	2.03×10^{-1}	2.06×10^{-1}	2.52×10^4	5.12×10^2
Popular	2.19×10^{-1}	2.10×10^{-1}	2.41×10^4	1.85×10^3
Rock	2.19×10^{-1}	2.08×10^{-1}	2.44×10^4	2.02×10^2
HipHop	2.19×10^{-1}	2.08×10^{-1}	2.78×10^4	1.59×10^2
Average	2.16×10^{-1}	2.07×10^{-1}	2.67×10^4	7.53×10^2

PC1: Dell Dimension 8300 (CPU: Pentium IV 3.2 GHz; main memory: 2.0 GB); PC2: Dell Dimension DXC051 (CPU: Pentium IV 3.0GHz; main memory: 1.0 GB).

5. Conclusions

We have developed an authentication method for music audio using a DWT. When we applied this method to five original music audio clips, the authentication ratio was 100%. Moreover, for music audio data compressed by MP3, AAC, or WMA, the authentication ratio was always at or near 100%. We used flags for distinguishing the wavelet coefficients used for storing a 0 or 1 bit of the original and authentication coding from other coefficients. The method never deteriorated the quality of the original music audio because it does not change it. When a level 8 DWT was used, which was the standard in this experiment, the mean time for the coding for the original music audio clip was 2.16×10^{-1} s and that for the authentication was 2.22×10^{-1} s for a 10-second original music audio clip. We propose that a music audio file should be judged to be authenticated when the file gives a 95% or higher authentication ratio for a certain clip taken from the music audio file.

For using the proposed method, we need to store in a data base 1) flags used for selective coding, and 2) an original code of each audio file whose copy right we should protect. In calculating the authentication ratio for the authentication of an original audio file, we do not need an original audio file but 1) the flags, and 2) the original code of the original audio file.

6. References

- [1] D. Kirovski and H. S. Malvar, "Spread-Spectrum Watermarking of Audio Signals," *IEEE Transactions on Signal Processing*, Vol. 51, No. 4, 2003, pp. 1020-1033. [doi:10.1109/TSP.2003.809384](https://doi.org/10.1109/TSP.2003.809384)
- [2] 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, 2003, pp. 381-386.
- [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, 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, 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, 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, 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, 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, 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, 2009, pp. 1133-1141.
- [11] X. Y. Wang, P. P. Niub and H. Y. Yangb, "A Robust Digital Audio Watermarking Based on Statistics Characteristics," *Pattern Recognition*, Vol. 42, No. 11, 2009, pp. 3057-3064. [doi:10.1016/j.patcog.2009.01.015](https://doi.org/10.1016/j.patcog.2009.01.015)
- [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, 2010, pp. 59-71.
- [13] S. Murata, Y. Yoshitomi and H. Ishii, "Optimization of Embedding Position in an Audio Watermarking Method Using Wavelet Transform," *Autumn Research Presentation Forums of ORSJ*, Japanese, October 2007, pp. 210-211.
- [14] S. Murata, "Optimization of Embedding Position in an Audio Watermarking Method Using Wavelet Transform," Master's Thesis, Osaka University, Suita, Japanese, 2006, pp. 53.

- [15] D. Inoue and Y. Yoshitomi, "Watermarking Using Wavelet Transform and Genetic Algorithm for Realizing High Tolerance to Image Compression," *Journal of the IIEEJ*, Vol. 38, No. 2, March 2009, pp. 136-144.
- [16] M. Shino, Y. Choi and K. Aizawa, "Wavelet Domain Digital Watermarking Based on Threshold-Variable Decision," Technical Report of IEICE, DSP2000-86, Japanese, Vol. 100, No. 325, September 2000, pp. 29-34.
- [17] 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*, Japanese, Vol. 45, No. 3, March 2004, pp. 728-738.
- [18] M. Tanaka and Y. Yoshitomi, "Optimization Problem for Embedding Position in an Audio Watermarking Based on Logarithmic Amplitude Modification for Realizing High Tolerance to MP3 Compression," *Autumn Research Presentation Forums of ORSJ*, Japanese, September 2006, pp. 70-71.
- [19] 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, September 2008, pp. 81-85.
- [20] 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*, Japanese, Vol. J86-A, No. 11, November 2003, pp. 1197-1206.