LETTER Special Section on Enriched Multimedia

Sound Quality Evaluation for Audio Watermarking Based on Phase Shift Keying Using BCH Code

Harumi MURATA[†a)], **Akio OGIHARA**[††], *Members*, *and* **Masaki UESAKA**[†††], *Nonmember*

SUMMARY Yajima et al. proposed a method based on amplitude and phase coding of audio signals. This method has relatively high sound quality because human auditory property is considered for embedding. However, in this method, the tolerance to attacks tends to be weak. Hence, we propose a high-tolerance watermarking method using BCH code which is one of error correcting code. This paper evaluates whether our method preserves the sound quality while ensuring high tolerance.

key words: *audio watermarking, BCH code, phase shift keying*

1. Introduction

Recently, copyright infringement has become a social problem such that the illegal reproduction is distributed on the Internet. Hence, audio watermarking methods, which embed proprietary data into digital audio data, have attracted attention as prevention techniques against copyright infringement [1]–[4].

For audio signals, Yajima et al. proposed a method based on amplitude and phase coding [2]. This method considers octave similarity to preserve high sound quality of the stego signal. Octave similarity means that the notes one octave above or below the root note are acoustically perceived as similar to the root note. If a frequency spectrum which is octave relationship (twice, four times, eight times, . . .) with a large-amplitude frequency spectrum is modified for embedding, the modification effects are not easily perceived by the characteristic component [3]. However, it is hard to say that this method has sufficiently tolerant to attacks.

Hence, in this paper, we propose a high-tolerance watermarking method using BCH code which is one of error correcting code. We also evaluate whether our method preserve the sound quality while ensuring high tolerance. The validity of the proposed method is confirmed by comparison with an existing method [1]. This existing method is highly tolerant to various attacks including MP3 compression, and has high sound quality. To compare the sound quality between the two methods, we apply perceptual evaluation of audio quality (PEAQ) [5] as the objective evaluation and ABX and ABC/HR audio comparison as the subjective evaluation.

2. Conventional Audio Watermarking Based on Low-Frequency Amplitude Modification

This chaper introduces the time-domain audio watermarking method based on low-frequency amplitude modification [1]. In this paper, we regard this method as conventional method.

The host signal $x(n)$ is divided into consecutive L-length group of samples (GOSs). Each GOS contains three non-overlapping sections (sec1, sec2, and sec3), and the lengths of these three sections are L_1, L_2 and L_3, respectively. Hence, $L = L_1 + L_2 + L_3$.

The average of absolute amplitudes (AOAAs) are calculated from the three sections as follows.

$$E_{i1} = \frac{1}{L_1} \cdot \sum_{n=0}^{L_1-1} |x(L \cdot i + n)| \tag{1}$$

$$E_{i2} = \frac{1}{L_2} \cdot \sum_{n=L_1}^{L_1+L_2-1} |x(L \cdot i + n)| \tag{2}$$

$$E_{i3} = \frac{1}{L_3} \cdot \sum_{n=L_1+L_2}^{L_1+L_2+L_3-1} |x(L \cdot i + n)| \tag{3}$$

where i is the GOS index; $i = 0, 1, 2, \dots$

E_{i1}, E_{i2} and E_{i3} are sorted in descending order, and they are renamed as $E_{\max}$, E_{mid} and $E_{\min}$, respectively. The differences of them are calculated by Eqs. (4) and (5).

$$A = E_{\max} - E_{\mathrm{mid}} \tag{4}$$

$$B = E_{\mathrm{mid}} - E_{\min}. \tag{5}$$

The relationship $A < B$ is called state "0" and it means watermark bit "0." Similarly, the relationship $A \geq B$ is called state "1" and it means watermark bit "1." Hence, one binary bit can be embedded in one GOS by modifying the host signal.

- To embed watermark bit "1"
  If ($A - B \geq$ Thd1), then no operation is performed. Else increase $E_{\max}$ and decrease E_{mid} by the same amount in order that the above condition is satisfied.
- To embed watermark bit "0"
  If ($B - A \geq$ Thd1), then no operation is performed. Else increase E_{mid} and decrease $E_{\min}$ by the same amount in

---

Manuscript received March 31, 2014.
Manuscript revised August 14, 2014.

[†] The author is with the Department of Information Engineering, School of Engineering, Chukyo University, Toyota-shi, 470–0393 Japan.

[††] The author is with the Department of Informatics, Faculty of Engineering, Kinki University, Higashihiroshima-shi, 739–2116 Japan.

[†††] The author is with the Department of Computer Science and Intelligent Systems, Graduate School of Engineering, Osaka Prefecture University, Sakai-shi, 599–8531 Japan.

a) E-mail: murata_h@sist.chukyo-u.ac.jp

DOI: 10.1587/transinf.2014MUL0003

order that the above condition is satisfied.

$E_{\max}$, E_{mid} or $E_{\min}$ are increased (decreased) by amplifying (attenuating) the amplitude of the host signal according to the above embedding conditions. The threshold Thd1 is calculated by Eq. (6)

$$\mathrm{Thd1} = (E_{\max} + 2E_{\mathrm{mid}} + E_{\min}) \cdot d_{\mathrm{c}} \tag{6}$$

where d_{c} is a parameter that adjusts the threshold.

In the extracting process, as a same manner as the embedding process, the AOAA of each section in each GOS is calculated. Comparing A and B, the retrieved bit is “1” if $A \geq B$ and “0” if $A < B$. This process is repeated for every GOS to extract the entire embedded bits.

3. Proposed Audio Watermarking Based on Phase Shift Keying Using BCH Code

This chapter presents our proposed audio watermarking method.

3.1 Embedding and Extracting Watermarks

Watermarks are embedded and extracted based on a method proposed in [2]. The host signal $x(n)$ is divided into segments, each containing N samples. Discrete Fourier transform (DFT) is applied to each segment of $x(n)$. $X_i(k)$ $(k = 0, 1, \ldots, N-1)$ corresponding to the i-th segment $x_i(n)$ is given by Eq. (7).

$$X_i = X_{i\mathrm{Re}}(k) + jX_{i\mathrm{Im}}(k) \;\; (k = 0, 1, \ldots, N-1). \tag{7}$$

Generally, an audio signal is characterized by a frequency component with a large amplitude $A_i(k)$. Hence, we search the frequency k_{i0} with maximum amplitude from a range of frequencies $[a, b]$, as shown in Fig. 1, and embed a watermark at frequency $2k_{i0}$, which has an octave relationship with k_{i0}.

Moreover, Eqs. (8) and (9) are processed using a phase parameter d_{p}[rad] $(0 \leq d_{\mathrm{p}} \leq \pi/4)$ by which the strength of operating a phase characteristic is determined. The phase parameter d_{p} improves the noise tolerance close to $\pi/4$. The following description of the embedding process is limited to the first quadrant for simplicity. The second, third, and fourth quadrants are processed in the same manner.

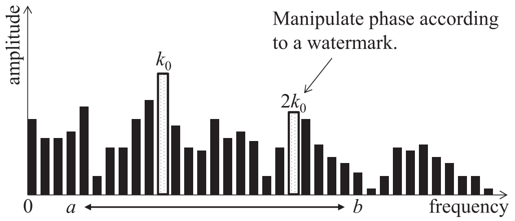

Fig. 1 Selection of the frequency to embed a watermark.

$$X'_{i\mathrm{Re}}(2k_{i0}) = \begin{cases} \mathrm{sign}(X_{i\mathrm{Re}}(2k_{i0}))A_i(2k_{i0})\cos(\pi/4 - d_{\mathrm{p}}) \\ \quad \text{if } (\pi/4 - d_{\mathrm{p}}) \leq \phi_i(2k_{i0}) \leq \pi/4 \\ \mathrm{sign}(X_{i\mathrm{Re}}(2k_{i0}))A_i(2k_{i0})\cos(\pi/4 + d_{\mathrm{p}}) \\ \quad \text{if } \pi/4 \leq \phi_i(2k_{i0}) \leq (\pi/4 + d_{\mathrm{p}}) \\ X_{i\mathrm{Re}}(2k_{i0}) \text{ else.} \end{cases} \tag{8}$$

$$X'_{i\mathrm{Im}}(2k_{i0}) = \begin{cases} \mathrm{sign}(X_{i\mathrm{Im}}(2k_{i0}))A_i(2k_{i0})\sin(\pi/4 - d_{\mathrm{p}}) \\ \quad \text{if } (\pi/4 - d_{\mathrm{p}}) \leq \phi_i(2k_{i0}) \leq \pi/4 \\ \mathrm{sign}(X_{i\mathrm{Im}}(2k_{i0}))A_i(2k_{i0})\sin(\pi/4 + d_{\mathrm{p}}) \\ \quad \text{if } \pi/4 \leq \phi_i(2k_{i0}) \leq (\pi/4 + d_{\mathrm{p}}) \\ X_{i\mathrm{Im}}(2k_{i0}) \text{ else.} \end{cases} \tag{9}$$

where, the sign(s) function outputs the sign, and is given by Eq. (10).

$$\mathrm{sign}(s) = \begin{cases} 1 & \text{if } s \geq 0 \\ -1 & \text{if } s < 0. \end{cases} \tag{10}$$

Following the rules of Fig. 2 and Eqs. (11)–(14), watermark m_i is embedded in each segment. An example of an embedded watermark bit “0” is shown in Fig. 3. In this case, the watermark is embedded by Eqs. (13) and (14).

- In case of $m_i = 1$

$$\hat{X}_{i\mathrm{Re}}(2k_{i0}) = \begin{cases} \mathrm{sign}(X'_{i\mathrm{Re}}(2k_{i0}))|X'_{i\mathrm{Im}}(2k_{i0})| \\ \text{if } \phi'_i(2k_{i0}) \geq \pi/4 \\ X'_{i\mathrm{Re}}(2k_{i0}) \text{ else.} \end{cases} \tag{11}$$

$$\hat{X}_{i\mathrm{Im}}(2k_{i0}) = \begin{cases} \mathrm{sign}(X'_{i\mathrm{Im}}(2k_{i0}))|X'_{i\mathrm{Re}}(2k_{i0})| \\ \text{if } \phi'_i(2k_{i0}) \geq \pi/4 \\ X'_{i\mathrm{Im}}(2k_{i0}) \text{ else.} \end{cases} \tag{12}$$

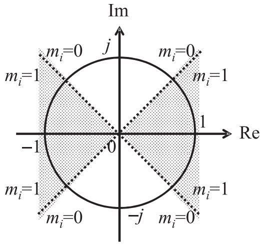

Fig. 2 Division of unit circle by phase.

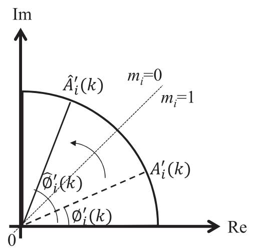

Fig. 3 Embedding watermark bit “0.”

- In case of $m_i = 0$

$$\hat{X}_{i\mathrm{Re}}(2k_{i0}) = \begin{cases} \mathrm{sign}(X'_{i\mathrm{Re}}(2k_{i0}))|X'_{i\mathrm{Im}}(2k_{i0})| \\ \text{if } \phi'_i(2k_{i0}) < \pi/4 \\ X'_{i\mathrm{Re}}(2k_{i0}) \text{ else.} \end{cases} \tag{13}$$

$$\hat{X}_{i\mathrm{Im}}(2k_{i0}) = \begin{cases} \mathrm{sign}(X'_{i\mathrm{Im}}(2k_{i0}))|X'_{i\mathrm{Re}}(2k_{i0})| \\ \text{if } \phi'_i(2k_{i0}) < \pi/4 \\ X'_{i\mathrm{Im}}(2k_{i0}) \text{ else.} \end{cases} \tag{14}$$

In the extracting process, as a same manner as the embedding process, the stego signal is divided into segments of N samples. DFT is applied to each segment to obtain $\hat{X}_i(2k_{i0})$. Watermark $\hat{m}_i$ is then extracted by Eq. (15).

$$\hat{m}_i = \begin{cases} 1 & \text{if } \hat{\phi}_i(2k_{i0}) < \pi/4 \\ 0 & \text{else.} \end{cases} \tag{15}$$

3.2 Retention of Maximum Amplitude Component

Amplitudes below the masking level are removed by MP3 compression. If a watermark is embedded at a frequency $2k_{i0}$ whose amplitude is below the masking level, that watermark cannot be correctly extracted. Hence, the tolerance to MP3 compression is improved by considering a masking curve [6], and the difference between the amplitude of $2k_{i0}$ and masking level is calculated by Eq. (16).

$$\mathrm{MSK[dB]} = A_{i\mathrm{dB}}(2k_{i0}) - \mathrm{LT_{min}}(q_{2k_{i0}}) \tag{16}$$

where $q_{2k_{i0}}$ is the number of the frequency band including $2k_{i0}$, and $\mathrm{LT_{min}}(q)$ [dB] is the minimum masking level of each band q ($q = 1, 2, \ldots, 32$).

If $\mathrm{MSK} < 0$, a component of frequency $2k_{i0}$ might be removed by masking when the audio signal is compressed and decompressed by MP3. Therefore, the tolerance to MP3 compression is improved by amplifying the amplitude of $2k_{i0}$ according to Eq. (17).

$$A_i(2k_{i0}) \leftarrow \begin{cases} A_i(2k_{i0}) \cdot 10^{-\mathrm{MSK}/20} & \text{if } -5 < \mathrm{MSK} < 0 \\ A_i(2k_{i0}) & \text{else.} \end{cases} \tag{17}$$

If the amplitude of $2k_{i0}$ is below the masking level, Eq. (17) amplifies it up to the masking level. Generally, music data include notes from multiple musical instruments. The amplitude of frequency $2k_{i0}$ is not necessarily large simply because it is harmonic to the frequency with maximum amplitude k_{i0}. If the amplification is excessively high, it perceived as noise. Hence, in this paper, Eq. (17) is applied only to frequency components that are amplified by less than 5 dB.

Furthermore, the second largest amplitude in the i-th segment, $A_i(k_{i1})$, may be larger than the maximum amplitude $A_i(k_{i0})$ if the stego signal is compressed and decompressed by MP3. To avoid this problem, the maximum amplitude $A_i(k_{i0})$ is amplified by Eq. (18).

$$A_i(k_{i0}) \leftarrow \begin{cases} A_i(k_{i1}) + \mathrm{thd}(i) & \text{if } A_i(k_{i0}) - A_i(k_{i1}) < \mathrm{thd}(i) \\ A_i(k_{i0}) & \text{else} \end{cases} \tag{18}$$

where $\mathrm{thd}(i)$ quantifies the amplitude modification in the i-th segment and is selected by the following process.

Step 1. The host signal and degraded signal resulting from MP3 compression and decompression of the host signal are divided into N-sample segments, respectively. Set $i \leftarrow 1$.

Step 2. The maximum amplitude $A_i(k_{i0})$ of the host signal and maximum amplitude $\tilde{A}_i(k_{i0})$ of the degraded signal are calculated in the i-th segment. If the frequency of $A_i(k_{i0})$ differs from that of $\tilde{A}_i(k_{i0})$, go to Step 3. Otherwise, $i \leftarrow i + 1$. If i is not the final segment, repeat Step 2, otherwise go to Step 4.

Step 3. Let temp1 be the difference between $A_i(k_{i0})$ and the second largest amplitude $A_i(k_{i1})$ of the host signal, and temp2 be the difference between $\tilde{A}_i(k_{i0})$ and the second largest amplitude $\tilde{A}_i(k_{i1})$ of the degraded signal. The sum temp1+temp2 is recorded as temp(i). temp(i) estimates the minimum difference between the maximum amplitude and the amplitudes of other frequencies. It is intended to leave the frequency of maximum amplitude unchanged after MP3 compression. $i \leftarrow i + 1$. If i is not the final segment, go to Step 2, otherwise go to Step 4.

Step 4. All temp(i) candidates are arranged in ascending order, and the sorted segments are indexed by $i'(t)$. 60% of the candidate values are modified as the saving range. $t \leftarrow 1$.

Step 5. For all temp($i'(t)$) within the saving range, the maximum amplitude of the $i'(t)$-th segment is modified by Eq. (18), and its signal is compressed and decompressed by MP3. In both host and degraded signals, frequencies of the maximum amplitude are calculated respectively. If frequencies of the host and degraded signal are equal, temp($i'(t)$) becomes the threshold thd($i'(t)$) in the $i'(t)$-th segment. If the frequencies are unequal, proceed to Step 6. Otherwise, $t \leftarrow t + 1$. If $i'(t)$ is within the saving range, repeat Step 5, otherwise go to Step 9.

Step 6. $u \leftarrow 1$.

Step 7. If the frequency of maximum amplitude cannot be correctly detected in the $i'(t)$-th segment, the current temp($i'(t)$) is updated to temp($i'(t + u)$). If $i'(t + u)$ is within the saving range, return to Step 5. Otherwise, go to Step 8.

Step 8. Using temp($i'(t + u)$), modify the maximum amplitude in the $i'(t)$-th and $i'(t + u)$-th segments by Eq. (18) and compress and decompress their corresponding signals by MP3.

Step 8-1. If the frequency of maximum amplitude of the degraded signal equals to that of the host signal in both the $i'(t)$-th and $i'(t + u)$-th segments, calculate the signal-to-noise ratio (SNR) after modifying the $i'(t)$-th segment and again after modifying the $i'(t+u)$-th segment. Save the segment yielding the better SNR, and return to Step 5. Otherwise, go to Step 8-2.

Step 8-2. If the frequency of maximum amplitude of the degraded signal equals to that of the host signal in the $i'(t)$-th or $i'(t+u)$-th segments, save the segment that the frequency of the degraded signal equals to that of the host signal, and return to Step 5. Otherwise, go to Step 8-3.

Step 8-3. If the frequency of maximum amplitude of the degraded signal does not equal to that of the host signal in either the $i'(t)$-th or $i'(t+u)$-th segments, set $u \leftarrow u+1$, and return to Step 7.

Step 9. The algorithm terminates once all processes are complete.

3.3 Introduction of BCH Code

In [2], it is hard to say that the tolerance to attacks are enough. Hence, the proposed method attempts to improve the attack tolerance by introducing BCH code which is one of error correcting code.

The BCH code is a representative cyclic code compatible with various error correcting requirements and code lengths and is relatively easy to encode and decode. Denoting the code length and number of information bits by l and p, respectively, we refer to such a code as a BCH(l, p). The BCH encodes p consecutive bits of an embedded bit sequence. BCH code with $2t$ roots can correct up to t errors. In this paper, decoding is performed by a Euclidean method.

The proposed method embeds payloads and synchronization codes as watermarks. Payloads are encoded by BCH(31,11) and BCH(15,5) codes, and their inclusion is preceded by synchronization codes. The BCH(31,11) and BCH(15,5) codes can correct errors up to 5 bits and 3 bits, respectively.

4. Experiments

In order to confirm the sound quality of proposed method, we examined objective and subjective evaluations of the sound quality based on evaluation criteria for audio information hiding technologies [7]. Furthermore, we validated the proposed method in comparison with an established method [1]. This method is highly tolerant to various attacks including MP3 compression, and has high sound quality. For testing, we used 8 music data selected from "SQAM recordings for subjective test," and 12 music data selected from "RWC music database: music genre," 60 seconds duration, at a 44.1 kHz sampling rate, with stereo channel. In the conventional method, 90-bit payloads and synchronization code per 15 seconds were embedded into each music data. The GOS length L was 7350, the lengths of three sections L_1, L_2 and L_3 were equal, and d_c was 0.05. In the proposed method, 263-bit BCH-encoded payloads and synchronization code per 15 seconds were embedded into each music data as shown in Fig. 4. In both methods, the synchronization code was 63-bit M-sequence. In the proposed method, other parameters were: segment size $N = 1024$, $d_p = \pi/8$, and frequency analysis range $[a, b] = [5, 140]$, corresponds to the frequency range 200 – 6000 Hz.

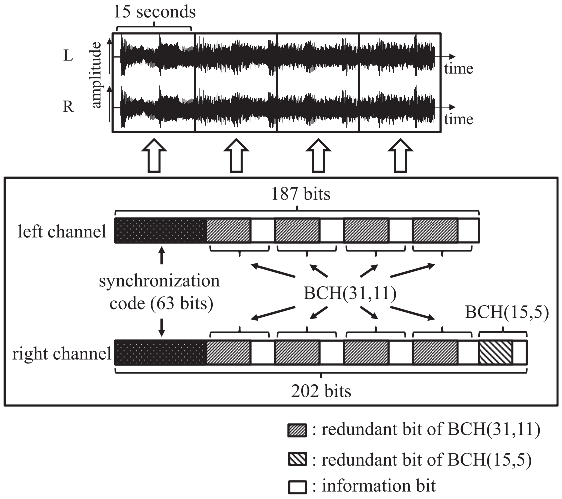

Fig. 4 Payloads using BCH code.

4.1 Tolerance to MP3 Compression

For the sound quality evaluation, the tolerance to MP3 compression of the proposed method was nearly equal to that of the conventional method. The bit error rate (BER) of watermarks was defined as the follows:

$$\mathrm{BER} = \frac{\text{number of error bits}}{180 \text{ bits}} \cdot 100\ [\%] \tag{19}$$

where the denominator represents the number of watermark bits per 30 seconds interval [7]. The payloads were extracted from consecutive 45 seconds of stego data from which the initial sample is randomly chosen in the initial 15 seconds. BER is defined as the number of mismatched bits between the embedded and extracted payloads relative to the 180 bits that are embedded into 15 to 45 seconds of the stego data.

(a) and (b) of Table 1 show the BER results after MP3 compression in the conventional and proposed methods, respectively. We observe that the BERs of both methods are very similar for all but two of the music data. For these exceptions, the embedding strength d_c of the conventional method was changed in the preliminary experiment, but the BER of the proposed method could not be equal to that of the conventional method.

(c), (d), and (e) of Table 1 show the BER results after MP3 compression in [2], retaining the maximum amplitude component, and using only BCH code, respectively. We find that tolerance to MP3 compression was drastically improved using only BCH code in 17 music data. However, retaining the maximum amplitude component significantly

Table 1 BER [%] after MP3 compression: (a) in conventional method [1], (b) in proposed method, (c) in [2], (d) in case of retention of maximum amplitude component, (e) in case of only using BCH code.

Music No.	BER [%]				
	(a)	(b)	(c)	(d)	(e)
SQAM Track 27	0	3.89	36.88	11.60	33.33
SQAM Track 32	4.44	3.89	31.94	6.46	30.00
SQAM Track 35	0.56	0	3.80	2.09	1.11
SQAM Track 40	0.56	1.11	5.89	5.70	1.11
SQAM Track 65	0.56	0	1.52	1.71	0
SQAM Track 66	0.56	0	4.94	5.13	0
SQAM Track 69	2.78	0	2.85	2.85	0
SQAM Track 70	0.56	0	1.71	1.52	0
RWC-MDB-G-2001 No.1	0	0	1.71	1.52	0
RWC-MDB-G-2001 No.7	0	0	1.71	2.09	0
RWC-MDB-G-2001 No.13	0	0	3.42	2.47	0
RWC-MDB-G-2001 No.28	0	0	2.47	2.28	0
RWC-MDB-G-2001 No.37	0	0	1.33	1.71	0
RWC-MDB-G-2001 No.49	0	0	1.52	1.33	0
RWC-MDB-G-2001 No.54	0.56	0	4.37	3.80	0
RWC-MDB-G-2001 No.57	0.56	0	2.09	1.71	0
RWC-MDB-G-2001 No.64	0.56	0	3.61	2.66	0
RWC-MDB-G-2001 No.85	0.56	0	1.52	1.33	0
RWC-MDB-G-2001 No.91	0.56	0	0.76	1.14	0
RWC-MDB-G-2001 No.100	0	0	2.66	2.47	0
average	0.64	0.44	5.84	3.08	3.28

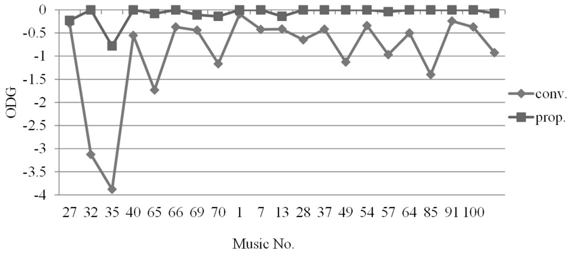

Fig. 5 ODG of conventional and proposed methods.

Table 2 Criterion for ABC/HR audio comparison.

score	criterion for ABC/HR audio comparison
5	imperceptible, transparent
4	perceptible but not annoying
3	slightly annoying
2	annoying
1	very annoying

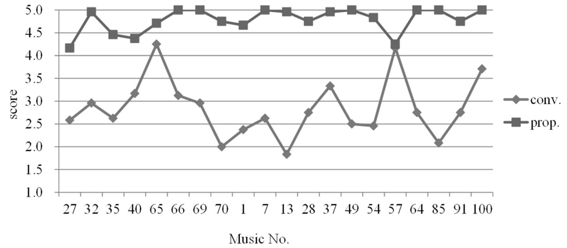

Fig. 6 The averaged score of 13 test subjects.

reduced the BER in 3 music data. BCH coding further improved the tolerance in these music data. It is confirmed that the tolerance is effectively increased by both processes.

4.2 Evaluation of Sound Quality

First, we evaluated the objective sound quality by PEAQ [5]. PEAQ uses some features of both host and stego signals and represents the quality comparison result as objective difference grade (ODG). The ODG ranges from 0 to −4, with higher values indicating greater watermark transparency. Figure 5 compares the ODG values between the conventional and proposed methods. For all music data, the ODG values were higher in the proposed method than in the conventional method. According to [7], the ODG values should exceed −2.5; the ODG values of the proposed method consistently exceeded −1. Hence, the proposed method is considered to have high objective sound quality.

Next, we evaluated the subjective sound quality using ABX and ABC/HR audio comparison. We used a digital audio processor ONKYO SE-U55SXII and a stereo headphone SONY MDR-CD-900ST. Nine males and 4 females (aged in the twenties) were test subjects. In the ABX test, the host data A or stego data B were randomly played to the test subjects as audio X. Subjects were requested to determine whether X was A or B. To ensure a statistically reliable result, each subject performed each test 8 times. In the ABC/HR audio comparison, test subjects score evaluation sound from 1 to 5 about the comparison between original sound and evaluation sound. The evaluation criteria are listed in Table 2. If the number of correct answer was more than 7 times in the ABX test, we judged that the subject can perceive a difference between the host data and the stego data in this experiment. If the stego data could not be differentiated from the host data in the ABX test, the score of ABC/HR audio comparison was automatically 5. Otherwise, the sound quality of each stego data was evaluated according to Table 2. Figure 6 shows the average ABC/HR audio comparison scores of the 13 test subjects. The proposed method was rated more highly than the conventional method in all music data. Furthermore, the average score of the 20 music data of proposed method was 1.8 points higher in the proposed method than in the conventional method.

The objective and subjective evaluations of sound quality confirmed the high sound quality of the proposed method.

5. Conclusion

In this paper, we proposed an audio watermarking method based on phase shift keying using BCH code. By virtue of the BCH code, this method improves the sound quality of the stego signal in consideration of human auditory property and improves the tolerance to attacks. The proposed method ensures approximately the same tolerance to MP3 compression as the conventional method while having a superior sound quality. In this paper, we evaluated the tolerance only for MP3 compression. In future works, we will propose and develop a method that tolerates a broader range of attacks.

Acknowledgments

This work was supported by Grants-in-Aid for Scientific Research (KAKENHI) 23500224, 26870681, and 26330214.

References

[1] W.N. Lie and L.C. Chang, “Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification,” IEEE Trans. Multimedia, vol.8, no.1, pp.46–59, 2006.
[2] N. Yajima and K. Oishi, “Digital watermarks for audio signals based on amplitude and phase coding,” IEICE Technical report, CAS2004-91, 2005.
[3] I. Muramatsu and K. Arakawa, “Digital watermark for audio signals based on octave similarity,” IEICE Trans. Fundamentals (Japanese Edition), vol.J87-A, no.6, pp.787–796, June 2004.
[4] A. Ogihara, M. Uesaka, S. Hayashi, and H. Murata, “A sound quality improve method for phase shift keying based audio watermarking considering masking curve,” Proc. 27th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC2012), 2012.
[5] ITU-R Rec. Bs.1387, “Method for objective measurements of perceived audio quality,” 2001.
[6] “Information technology-Coding of moving pictures and associated audio for storage media at up to about 1.5 Mbit/s - Part 3 : Audio,” ISO/IEC 1172-3, pp.96–107, 1993.
[7] http://www.ieice.org/iss/emm/ihc/audio/audio2013v2.pdf, Accessed Oct. 1, 2013.