Journal of Information Hiding and Multimedia Signal Processing ©2011 ISSN 2073-4212
Ubiquitous International Volume **2**, Number **1**, January **2011**

Embedding Limitations with Digital-audio Watermarking Method Based on Cochlear Delay Characteristics

Masashi Unoki, Kuniaki Imabeppu, Daiki Hamada, Atsushi Haniu, and Ryota Miyauchi

School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292 Japan
{unoki, i-beppu, hamada, a-haniu, ryota}@jaist.ac.jp

Received June 2010; revised July 2010

Abstract. *We comparatively evaluated the proposed approach for inaudible audio watermarking with four typical methods (LSB, DSS, ECHO, and PPM) by carrying out objective (PEAQ and LSD) and subjective (inaudibility) evaluations, bit-detection test, and robustness tests (signal modifications and StirMark benchmark). The results of evaluations revealed that subjects could not detect the embedded data in any of the watermarked signals we used, and that the proposed approach could precisely and robustly detect the embedded data from the watermarked signals. We also investigated embedding limitations with our proposed method and improved the method by designing a parallel architecture for cochlear delay filters. We then evaluated our proposed and improved methods to investigate embedding limitations by carrying out five tests: LSD, PEAQ, bit-detection, and two robustness tests (signal modifications and StirMark benchmark). The results revealed that the methods could be used to inaudibly embed the watermarks into original signals and to accurately and robustly detect the embedded data from the watermarked signals. We also found that embedding limitations with the improved method ($M = 8$) amounted to 384 bps while that with our proposed method ($M = 2$) amounted to 128 bps.*
Keywords: Digital-audio watermarking, Cochlear delay characteristics, Inaudibility, Embedding limitations, Parallel architecture.

1. **Introduction.** Multimedia information hiding (MIH) techniques have aimed to help to preserve the values of multimedia information such as text, digital-audio, images, and video, help to hide imperceptible marks such as copyright notice into them, or even help to prevent their unauthorized copying. MIH techniques are, in general, composed of content protection of multimedia information such as watermarking and steganography that means hiding multimedia information in other multimedia information. Since it is possible to use MIH techniques together with cryptographic techniques, they are applicable for secure content authentication such as fingerprint.

Typical applications based on MIH techniques have recently been attracted as state-of-the-art techniques for copyright protection [1, 2] and these have been realized as digital watermarking methods. Their aim has been to embed digital codes for the copyright information in the multimedia contents, which are imperceptible to users. Since the embedded data cannot be detected by users, they cannot illegally manipulate the watermarked data to remove the copyright information. In particular, there have recently been serious social issues involved in protecting the copyright of all digital-audio content by preventing

it from being illegally copied and distributed on the Internet. Digital-audio watermarking has been focused on as a state-of-the-art technique enabling copyright protection, as

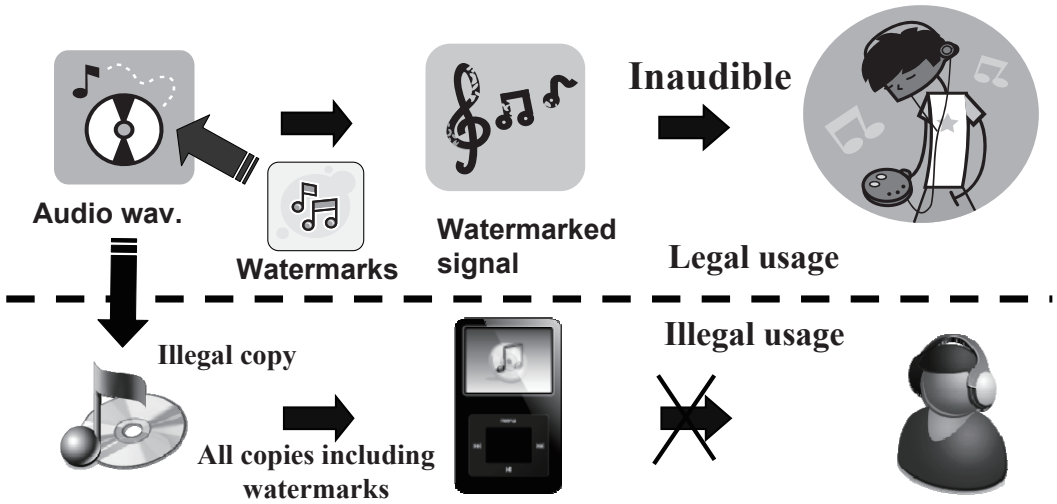

FIGURE 1. Schematic illustration of digital-audio watermarking.

shown in Fig. 1. This has aimed to embed codes to protect the copyright in audio content that are inaudible to and inseparable by users, and to detect embedded codes from watermarked signals [3]. However, in contrast with watermarking techniques for image/video contents, there seems to be no complete or successful method for digital audio contents in industrial applications. Although the reasons will be appeared in later, there are several issues that have to be resolved for realizing reasonable digital-audio watermarking.

In general, audio watermarking methods must satisfy three requirements to provide a useful and reliable form of copyright protection: (a) **inaudibility** (inaudible to humans with no sound distortion caused by the embedded data), (b) **confidentiality** (secure and undetectable concealment of embedded data), and (c) **robustness** (not affected when subjected to techniques such as data compression) [3, 4]. The first requirement (**inaudibility**) is the most important in the method of audio watermarking because this must not affect the sound quality of the original audio. If the sound quality of the original is degraded, the original content may lose its commercial value. The second requirement (**confidentiality**) is important to conceal watermarks to protect copyright, and it is important that users do not know whether the audio content contains watermarking or not. The last requirement (**robustness**) is important to ensure the watermarking methods are tamper-proof to resist any manipulations by illegal users.

Typical methods of watermarking have been based on signal manipulations in quantization /coding levels or in the amplitude (or amplitude spectrum). There are, for example, methods based on least significant bit (LSB) replacement in quantization (e.g., [3, 5]) and the spread spectrum approach (e.g., direct spread spectrum (DSS) proposed by Boney *et al.* [6]). These methods are used to directly embed watermarks such as copyright data into the quantization/coding levels or amplitude of digital-audio signals and detect the embedded data from the watermarked signals. Although methods of bitreplacement /manipulation such as LSB are relatively less audible than other conventional techniques of watermarking, these are not robust against various manipulations such as down-sampling/up-sampling or compression. Thus, these do not completely satisfy the three requirements, especially with regard to robustness. Spread spectrum methods such as DSS are relatively more robust than the others because watermarks are spread throughout whole frequencies that are preserved. However, this does not completely satisfy these

TABLE 1. Three requirements for digital-audio watermarking and weaknesses with typical watermarking methods. The "∘" and "×" indicate true and false as to whether inaudibility, confidentiality, and robustness requirements were satisfied or not. "∘⁻" means almost satisfied and occasionally with very slight problems.

Method	(a) Inaudi.	(b) Confid.	(c) Robust.	Weaknesses
LSB	∘	∘	×	Not Robusted due to signal manipulation
DSS	×	∘	∘	Distorted and poor sound quality
ECHO	∘	×	∘	Easy to detect watermarks
PPM	∘⁻	∘	∘⁻	Watermarks in pulsive sound audible
CD	∘	∘⁻	∘	—-

three requirements, especially with regard to inaudibility. It is therefore difficult to embed inaudible watermarks into the amplitude information.

Another typical methods of watermarking have been based phase spectrum (or group delay characteristics). There are, for example, an echo-hiding approach proposed by Gruhl *et al.* [8] and a method based on periodical phase modulation (PPM) proposed by Nishimura et al. [9, 10]. Echo-hiding approaches have been used to directly embed watermarks into the audio signals as time shifts. Thus, the two main advantages of using these approaches have been to embed watermarks into the original the signal with less distortion and at lower computational cost. Although they satisfy the inaudibility requirement, the former has a drawback in confidentiality because it is less secure (it is easy for anyone to detect the echo information) and neither method is as robust as the other established methods. PPM approach was based on aural capabilities in that PPM is relatively inaudible to humans. They found this phenomena when they conducted psychoacoustical experiments. However, as phase modulation randomly disrupts the phase spectra of components at higher frequencies, these modulated components (embedded data) may be able to be detected by humans in watermarked pulse-like sounds, especially around rapid onsets in musical sounds such as onsets in the piano. This is because humans can perceive rapid phase-variations related to long and rapid group delays in sounds [11, 12, 13, 14].

In summary, the typical watermarking methods used in LSB, DSS, ECHO, and PPM approaches could partially satisfy the three requirements. PPM, especially, was found to be the best of these methods. The features of these methods are listed in Table 1. These methods can be also categorized as watermarking processes in the amplitude or phase (time-delay) domains. The first two methods in Table 1 are in the amplitude domain, while the last two methods are in the phase domain. This table suggests us that it is very difficult to achieve inaudible watermarking that can satisfy all three requirements. The aim of our work was to find an inaudible watermarking scheme based on human auditory perception (without using amplitude manipulations or various masking phenomena) to satisfy the inaudibility, confidentiality, and robustness criteria.

To solve these problems, inaudible digital-audio watermarking has been based on the properties of the human cochlear, i.e., cochlear delay (CD) was proposed by the authors [15, 16]. Although this method has almost satisfied the three requirements, especially in (a) inaudibility and (c) robustness, it has not yet been investigated how effective this method is in embedding watermarks into digital-audio signals (see in Tab. 1). Therefore,

effectiveness and embedding limitations with the proposed method have not yet been discovered. In this paper, we comparatively evaluated our proposed approach against four other methods (LSB, DSS, ECHO, and PPM) by carrying out three objective evaluations, some subjective evaluations, and robustness tests. We also evaluated embedding limitations by carrying out objective and subjective experiments. We then improved the proposed method by using a parallel architecture for cochlear-delay (CD) filtering to further reduce their embedding limitations.

This paper proposes a novel approach for an inaudible method of watermarking based on CD characteristics to protecting digital-audio content by using a parallel architecture for CD filters. It is organized as follows. Section 2 explains the underlying concept and method of digital-audio watermarking based on CD characteristics. Section 3 describes how the method was implemented by using IIR CD filters. Section 4 presents the results of objective/subjective evaluations and assessments of the robustness of the proposed method to confirm effectiveness of the proposed method. Section 5 improves the proposed method by using a parallel architecture for CD filters to further reduce embedding limitations of them. Section 6 presents the results of objective evaluations and assessments of the robustness of the improved method to investigate embedding limitations with the proposed and improved methods. Section 7 summarizes the proposed scheme for inaudible watermarking and briefly describes future work.

2. **Concept of inaudible watermarking.** Cochlear delay (CD) is referred to as delay in the course of wave propagation in the basilar membrane (BM) [7]. Due to this, lower-frequency components require more time to reach the area of maximum displacement in the BM, near the apex, while higher frequency components elicit a maximum closer to the base. Aiba *et al.* [17, 18] studied whether cochlear delay significantly affected perceptual judgment of the synchronization of sounds. They used three types of chirp sounds: a pulse sound, a compensatory delay chirp, and an enhanced delay chirp. Their results suggest that the auditory system cannot distinguish between enhanced-delay and non-processing sounds.

Based on Aiba *et al.*'s results [17, 18], we found that it was very difficult for us to discriminate the enhanced delay chirp with the original (intrinsic sound) while it was very easy to discriminate the compensatory delay chirp with the original. We considered that these characteristics could be used to effectively embed inaudible watermarks into an original signal, and we therefore propose an audio-watermarking method based on CD characteristics. This method embeds watermarks by controlling the respective group delays in filters ($H_0(z)$ and $H_1(z)$) corresponding to the digital copyright codes ("0" and "1"). We designed the cochlear delay characteristics by using the following 1st-order IIR all-pass filter:

$$H_m(z) = \frac{-b_m + z^{-1}}{1 - b_m z^{-1}}, \qquad 0 < b_m < 1, \qquad m = 1, 0. \tag{1}$$

The group delay, $\tau_m(\omega)$, in Eq. (1) can be obtained as:

$$\tau_m(\omega) = -\frac{d\arg(H_m(e^{j\omega}))}{d\omega}, \tag{2}$$

where $H_m(e^{j\omega}) = H_m(z)|_{z=e^{j\omega}}$.

The group delay characteristics of $H_m(z)$ were fitted to the CD characteristics [18] (scaled by 1/10 as indicated by the dashed line in Fig. 2). The dashed line in Fig. 2 plots the CD characteristics described by Dau *et al.* [7], where the delay time was scaled by 1/10. The first two solid lines in Fig. 2 plot the group delays of the IIR all-pass filters in Eq. (2), i.e., $H_0(z)$ with $b_0 = 0.795$ and $H_1(z)$ with $b_1 = 0.865$ in Eq. (1). If

CD characteristic can be modeled as a phase characteristic of a digital filter, a method of audio watermarking based on cochlear characteristics could be established by controlling

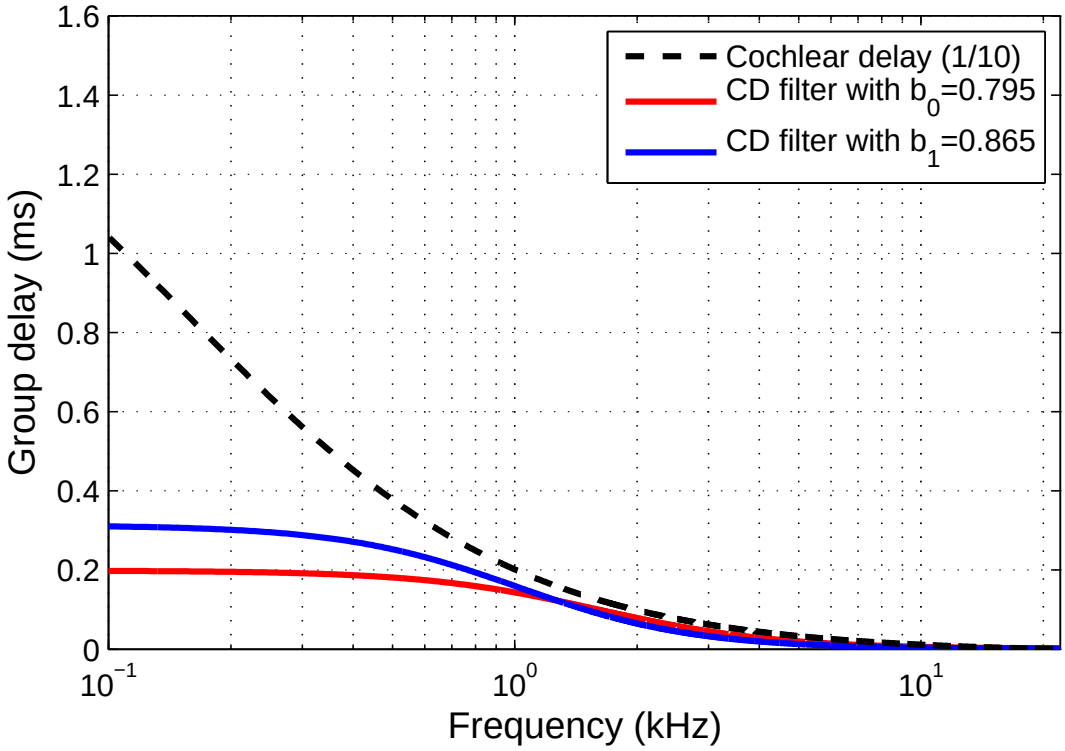

FIGURE 2. Cochlear-delay and group-delay characteristics of filter in Eq. (1).

the respective group delays in the filter to those of the digital copyright data ("1" and "0").

3. **Watermarking based on Cochlear-delay.** Our proposed method consists of two processes: a data-embedding and a data-detection process. A data-detection process should generally be accomplished as blind detection. Since our motivation was based on how inaudible watermarking could be attained, the data-detection process was achieved as non-blind detection in the first step. These are based on phase-shift-keying (PSK) techniques for digital signal modulation. Below, we describe how these processes were implemented.

3.1. **Data embedding process.** Figure 3(a) has a block diagram of the data-embedding process. Watermarks were embedded as follows: (1) Two IIR all-pass filters, $H_0(z)$ and $H_1(z)$, were designed using different values for $b_m (b_0 = 0.795$ and $b_1 = 0.865)$ to enhance the cochlear delay. These values were determined by taking experimental conditions into consideration. (2) The original signal, $x(n)$, was filtered in the parallel systems, $H_0(z)$ and $H_1(z)$, and intermediate signals, $w_0(n)$ and $w_1(n)$, were then obtained as the outputs for these systems (Eqs. (3) and (4)). (3) The embedded data, $s(k)$, were set to conform to the copyright data, e.g., "01010001010110..." as shown in Fig. 3(a). (4) The intermediates, $w_0(n)$ or $w_1(n)$, were selected by switching the embedded data $s(k)$ ("0" or "1"), and merging them with the watermarked signal, $y(n)$, in Eq. (5).

$$w_0(n) = -b_0 x(n) + x(n-1) + b_0 w_0(n-1), \tag{3}$$

$$w_1(n) = -b_1 x(n) + x(n-1) + b_1 w_1(n-1), \tag{4}$$

$$y(n) = \begin{cases} w_0(n), & s(k) = 0 \\ w_1(n), & s(k) = 1 \end{cases} \tag{5}$$

where $(k-1) \triangle W \leq n < k \triangle W$. Here, n is the sample index, k is the frame index, and

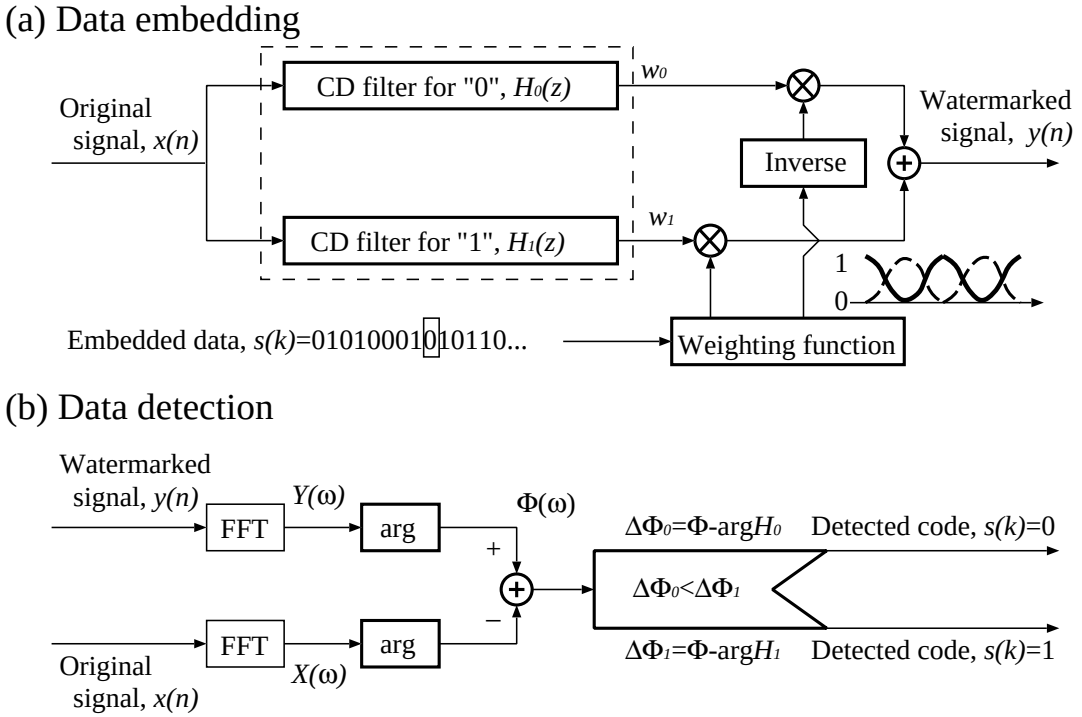

FIGURE 3. Block diagram for data embedding and data detection in the proposed method.

$\triangle W = f_s/N_{\text{bit}}$ is the frame length (the frame overlap is half a frame.). In addition, f_s is the sampling frequency of the original signal and N_{bit} is the bit rate per second (bps).

3.2. **Data detection process.** Figure 3(b) shows the flow for the data-detection process we used. Watermarks were detected as follows: (1) We assume that both $x(n)$ and $y(n)$ are available with this watermarking method. (2) The original, $x(n)$, and the watermarked signal, $y(n)$, are decomposed to become overlapped segments using the same window function used in embedding the data. (3) The phase difference, $\phi(\omega)$, is calculated in each segment, using Eq. (6). FFT[P] is the fast Fourier transform (FFT). (4) To estimate the group delay characteristics of $H_0(z)$ or $H_1(z)$ used in embedding the data, the summed phase differences of $\phi(\omega)$ to the respective phase spectrum of the filters ($H_0(z)$ and $H_1(z)$), $\triangle\Phi_0$ and $\triangle\Phi_1$ are calculated as in Eqs. (7) and (8). (5) The embedded data, $\hat{s}(k)$, are detected using Eq. (9).

$$\phi(\omega m) = \arg(FFT[y(n)]) - \arg(FFT[x(n)]), \tag{6}$$

$$\triangle\Phi_0 = \sum_m |\phi(\omega_m) - \arg(H_0(e^{j\omega m}))|, \tag{7}$$

$$\triangle\Phi_1 = \sum_m |\phi(\omega_m) - \arg(H_1(e^{j\omega m}))|, \tag{8}$$

$$\hat{s}(k) = \begin{cases} 0, & \triangle\Phi_0 < \triangle\Phi_1 \\ 1, & \text{otherwise} \end{cases} \tag{9}$$

3.3. **Key technology.** Figure 4 has a schematic of the key technology used in these watermarking methods. The echo-hiding approach controls echo-delay (T_0 and T_1) corresponding to digital codes ("0" and "1") in $y(n)$, using an echo-impulse response (relative amplitude A and echo delay (T_0 and T_1)), as seen in Fig. 4(a). Although humans cannot

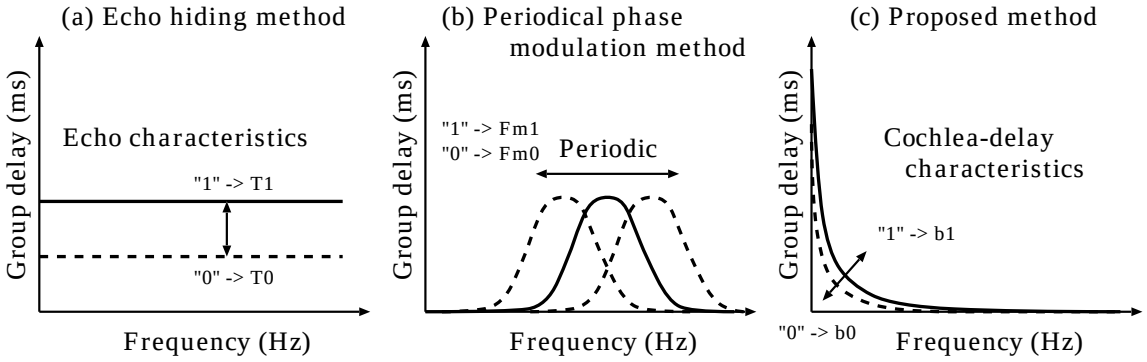

FIGURE 4. Schematic of key technology: (a) echo hiding, (b) periodical phase modulation, and (c) cochlear-delay characteristics.

perceive these echoes as different sounds if the delay time is not very long, these delays can very easily be detected by using auto-correlation. Therefore, we found that this technique lacked confidentiality (requirement (b)).

The PMM approach periodically controls certain group delays derived from phase modulation around a certain range (from 8 to 20 kHz) [9], as shown in Fig. 4(b). Digital codes with this technique are embedded as periodic information (F_{m0} and F_{m1} in phase modulation) in $y(n)$. However, since pulse-like sounds such as the rapid onset of sounds have wide frequency components, this kind of phase modulation disrupts the phase spectra of components at higher frequencies and these may be able to be detected by humans. Therefore, we discovered that this technique occasionally suffers from slight problems with regard to inaudibility (requirement (a)).

4. **Comparative evaluations of proposed method.** In this section, objective and subjective evaluations and robustness tests are carried out to reveal effectiveness of the proposed method. These evaluations and tests are also done for the other methods in comparison with the proposed method.

4.1. **Database and conditions.** All of the 102 tracks of the RWC music genre database [19] were used as the original signals in the evaluation. The original track has a sampling frequency of 44.1 kHz, 16 bits and two channels (stereo). The same watermarks with 8 characters ("AIS-lab.") were embedded into both R-L channels using the proposed method. The STEP2001 [4] suggested that 72 bits per 30 s was required to ensure a reasonable bit-detection rate with the method of audio watermarking. Thus, we used N_{bit} = 4 bps as this critical condition.

We comparatively evaluated our proposed method with four others (LSB, DSS, ECHO, and PPM) by carrying out two objective tests: Perceptual evaluation of sound quality (PEAQ) [20] and Log spectrum distortion (LSD), These measures were used to perceptually evaluate the digital-audio watermarking in Lin and Abdulla [21]. Bit-detection tests were also carried out. N_{bit} in these tests was fixed at 4 bps. The tip rate and data rate in DSS were set to 4 and 8192. A carrier frequency of 0 Hz and a key of a pseudo-random sequence of 1374 were used. The delay times for the echoes, T_0 and T_1 , were 2.3 and 3.4 ms with the ECHO method as shown in Fig. 4(a). The relative amplitude of the echoes was set to A = 0.6. The F_{m0} and F_{m1} in PPM were set to 8 and 10 Hz, as shown in

Fig. 4(b). Here, data detection with LSB, DSS, and ECHO were implemented as blind detection while data detection with PPM was implemented as non-blind detection.

All these signals were watermarked under the above conditions and these were then tested to detect the embedded data from all the watermarked signals.

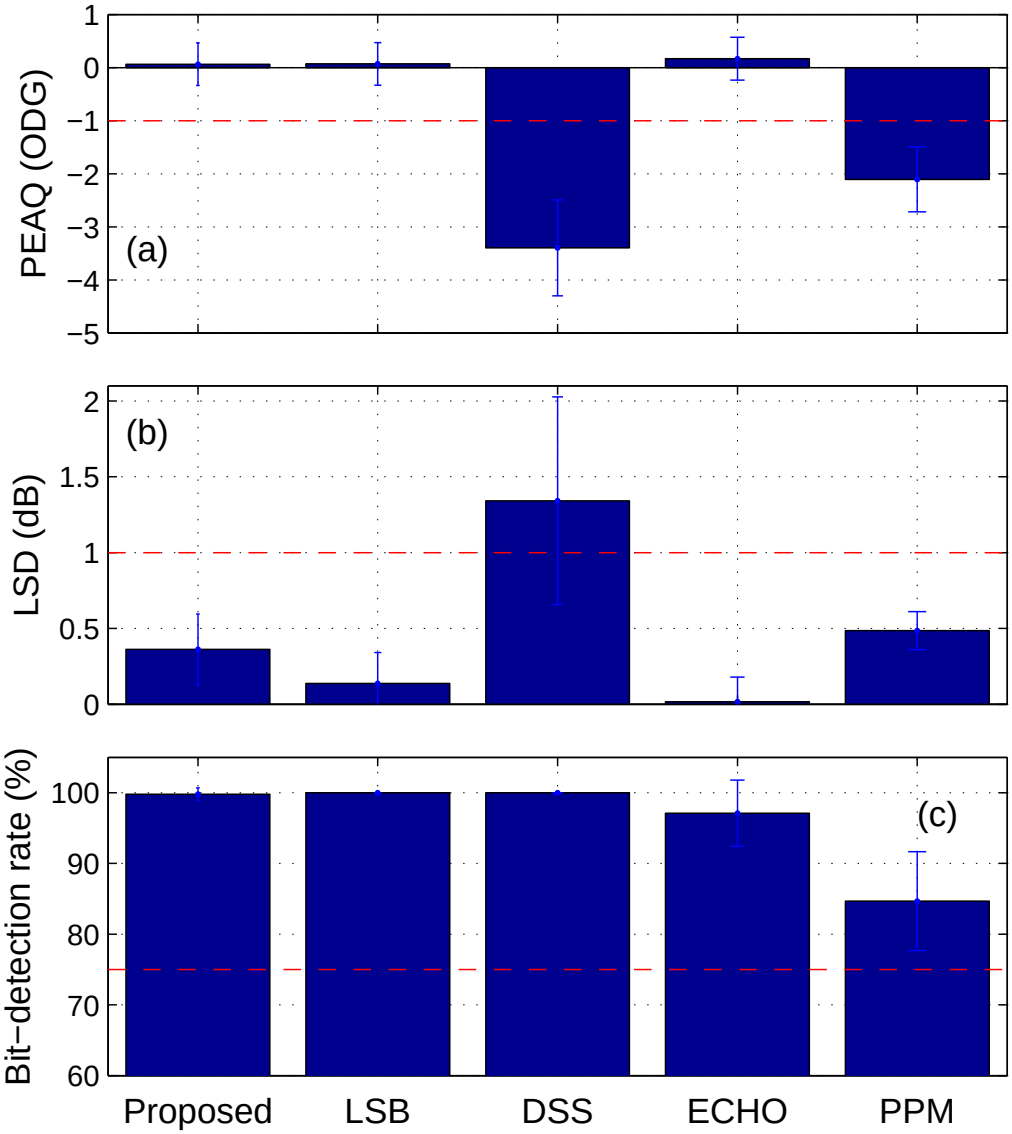

FIGURE 5. Results of evaluation for the proposed method: (a) PEAQ, (b) LSD, and (c) bit-detection rate.

4.2. **Objective evaluations.** We carried out an objective experiment (simulation) to evaluate the PEAQ measurements [20] between the original and the embedded signals. The PEAQ measurements, recommended by ITU-R BS.1387, were used to output the objective difference grade (ODG), which corresponded to the subjective difference grade (SDG) obtained from the procedure to evaluate subjective quality. The ODGs were graded as 0 (imperceptible), −1 (perceptible but not annoying), −2 (slightly annoying), −3 (annoying), and −4 (very annoying). The basic version of PEAQ [20] was used to assess the ODGs of the stimuli. A threshold of −1 was chosen as the embedding limitation to evaluate the PEAQs in this experiment.

Figure 5(a) shows the averaged ODGs of the PEAQs for the watermarked signals. The bars indicate the averaged ODGs and error bars indicate the standard deviations for

these ODGs. The PEAQs at the proposed, LSB, and ECHO-methods were under the evaluational threshold (> -1) in which the bit-rate was fixed 4 bps.

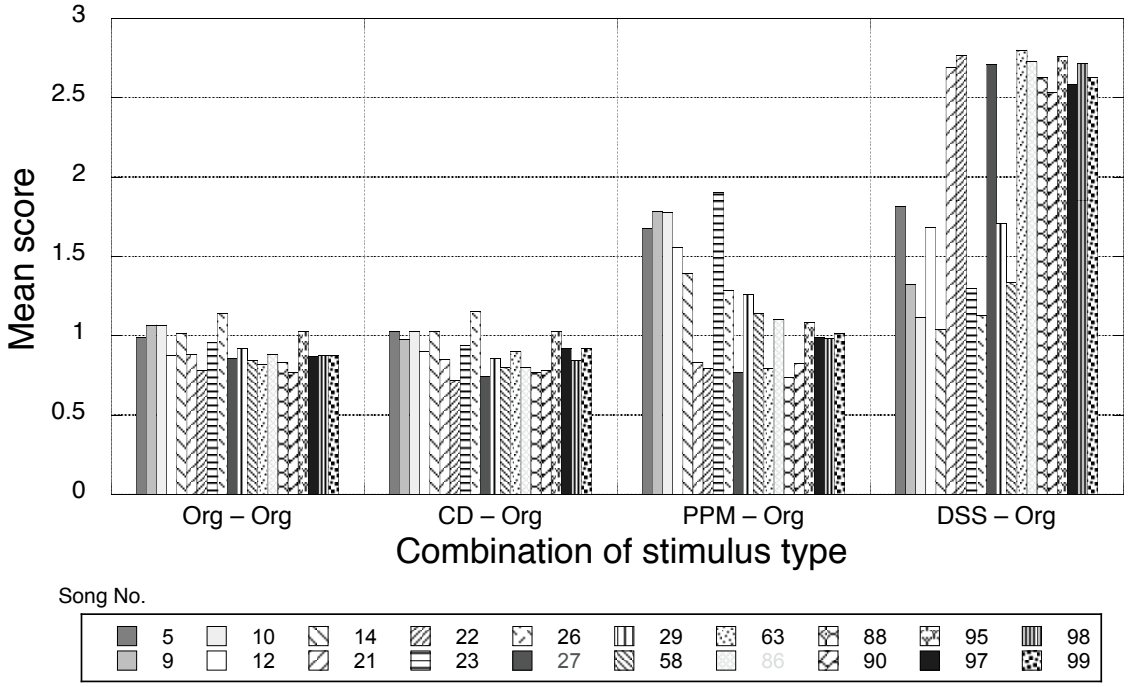

FIGURE 6. Results of subjective evaluations.

We also carried out LSD measurements to evaluate the sound quality of the watermarked signals.

$$\mathrm{LSD} = \frac{1}{K}\sum_{k=1}^{K} 10\log_{10}\frac{|Y(\omega,k)|^2}{|X(\omega,k)|^2}, \qquad (\mathrm{dB}), \tag{10}$$

where k is the frame index, K is the number of frames, and $X(\omega,k)$ and $Y(\omega,k)$ are the Fourier amplitude spectra for original signal $x(n)$ and watermarked signal $y(n)$ at the k-th frame. A frame length of 25 ms and 60% overlap (15 ms) were used in this research.

Figure 5(b) has the averaged LSD for the watermarked signals at 4 bps. The bars indicate the averaged LSD and the error bars indicate the standard deviations. These results ensure that the proposed method with N_{bit} of 4 could be used to embed the watermarks into the original signals to satisfy requirement (a). The LSDs in the proposed, LSB, ECHO, and PPM methods were under the evaluation threshold (1 dB).

We carried out a bit-detection test to evaluate how well the proposed method could accurately detect embedded data from the watermarked audio signals. The same original signals were used in this experiment. The bit-detection rates for all signals were evaluated as a function of the bit rate. A threshold of 75% was chosen as the limitation for embedding to evaluate the bit-detection rate in this experiment.

Figure 5(c) plots the averaged bit-detection rate of the watermarked signals. The detection rates were under the evaluation threshold ($> 75\%$) in which the bit rate is 4 bps. This ensured that the method could be used to detect the watermarks from the watermarked signals to satisfy requirement (b). Bit-detection rate in the other methods (DSS, LSB, ECHO, and PPM) were also under the evaluation threshold ($> 75\%$).

4.3. **Subjective evaluation.** To investigate inaudibility of a sound distortion caused by the embedded data based on CD, we conducted a subjective experiment. 20-tracks in the

TABLE 2. Results of robustness tests (bit-detection rate (%)).

Proc.	LSB	DSS	ECHO	PPM	Proposed
Non processing	100.0	100.0	96.71	84.68	99.32
Resampling 20k	57.19	99.02	94.25	58.95	99.18
Resampling 16k	56.76	99.02	93.34	57.10	99.09
Resampling 8k	54.32	98.33	88.06	53.10	95.26
Bit extension 24 bits	100.0	99.02	96.71	84.68	99.32
Bit compression 8 bits	51.00	98.20	85.69	54.65	94.21
mp3 128 kbps	50.94	99.02	95.49	58.36	90.63
mp3 96 kbps	49.76	99.02	94.51	57.54	87.33
mp3 64-kpbs mono	50.18	99.02	94.63	57.05	89.80

RWC music-genre database [7] were used in the subjective evaluation. The tracks were chosen according to the score of PEAQ (ODG) for all 102-tracks in the database.

The tracks of RWC-MDB-G-2001 No. 14, 5, 9, 23, 26, 10, 12, and 29 were objectively evaluated that a distortion caused by embedding was small (maximum and minimum values of ODGs at 4 bps were 0.18 and 0.15, respectively). The tracks of RWC-MDB-G-2001 No. 63, 58-2, 97, 99, 86, 95, 21, 90, 98, 27, and 22 were evaluated that the distortion was large (maximum and minimum values of ODGs were 0.16 and -0.27, respectively). The same watermarks with eight upper-case letters ("AIS-lab.") were embedded into L channel of the tracks by using the proposed method (CD), PPM, and DSS. The bit-rate, N_{bit} , was 4 bps.

Six naive paid volunteers took part in the experiment. In a trial, two tracks, which one was an original track (Org) and the other was the same original track (Org) or an embedded track (CD, PPM, or DSS) were sequentially presented to the participants. The participants task was to judge the similarity of the two tracks by a subjective scale consisted by following four scores: 0. completely the same, 1. probably the same, 2. probably different, and 3. completely different. Each participant performed 20 trials for 80 track-combinations (20 tracks ×4 combinations (Org-Org, Org-CD, Org-PPM, and Org-DSS)).

We calculated the mean scores of judgments for each participant (the mean scores of all participants showed in Fig. 6) and performed a two-way (20 tracks ×4 combinations) analysis of variance (ANOVA) on the mean scores of each participant (n = 6). The results of the ANOVA revealed a significant interaction between the two factors ($F_{57,285}$ = $17.4, p < .001$). Post hoc multiple comparison tests revealed that there were no significant differences among the mean scores of 20 tracks on the Org-Org and Org-CD combinations, whereas the main effect of tracks were significant on the Org-PPM and Org-DSS combinations. Furthermore, the differences between the mean scores of the Org-Org and Org-CD combinations on each tracks was not significant. These results indicate that the sound distortion caused by the embedded data based on CD is inaudible, and the inaudibility is not affected by characteristics of tracks. The same demonstrations that we used in subjective evaluations are available on our Web site [22].

4.4. **Evaluation of robustness.**

4.4.1. *Robustness test for signal modifications.* We carried out three types of robustness tests to evaluate how well the methods could accurately and robustly detect embedded data from the watermarked-audio signals. Based on suggestions from STEP2001 [4], the main manipulation conditions used were: (i) down sampling (44.1 kHz → 20, 16, and 8 kHz), (ii) amplitude manipulation (16 bits → 24-bit extension and 8-bit compression), and

TABLE 3. Content of each category.

Category	SMBA Attack
i) **Noise**	AddBrumm, AddDynNoise, AddFFTNoise, AddNoise, AddSinus, NoiseMax
ii) **Amplitude**	Amplify, Compressor, Normalizer1, Normalizer2
iii) **Bit**	BitChanger, LSBZero
iv) **Data**	CopySample, CutSample, Exchange, FlipSample, ReplaceSamples, ZeroCross, ZeroLength1, ZeroLength2, ZeroRemove
v) **Filtering**	BassBoost, ExtraStereo, FFT_HLPassQuick, RC_LowPass, RC_HighPass, Smooth1, Smooth2, State1, State2, VoiceRemove
vi) **Phase**	FFT_Invert, FFT_RealReverse, Invert
vii) **Echo**	Echo

(iii) data compression (mp3: 128 kbps, 96 kbps, and 64 kbps-mono). These conditions were the same as in Unoki and Hamada [15, 16].

Table 2 lists the results of evaluations for the proposed method (CD) and the other methods (DSS, LSB, ECHO, and PPM). The bit-detection with the proposed method (CD) was 99.3% where there was no manipulation (default case). In contrast, the bit-detection rates under the strong manipulation conditions (down sampling from 44.1 kHz to 8 kHz, amplitude compression from 16 bits to 8 bits, and data compression of 96 kbps) corresponded to 96.7%, 94.1%, and 87.3%. Hence, these results indicate that our proposed approach could accurately and robustly watermark copyrighted data in original digital-audio content. In addition, it was also found that LSB and PPM had a drawback in robustness for watermarking while DSS and ECHO could satisfy robustness requirement.

4.4.2. *StirMark benchmark test.* We finally carried other robustness tests by actual attacks to evaluate how well the methods could accurately and robustly detect embedded data from the watermarked-audio signals. The attacking tool employed in these robustness tests was StriMark Benchmark for Audio [23] version 1.3.2 (SMBA). 35 attacks of SMBA were used in these test. The parameter of each attack was a default value. We categorized 35 attacks as seven categories: (i) **Noise:** noise addition, (ii) **Amplitude:** amplitude operation, (iii) **Bit:** bit handling, (iv) **Data:** data substitution operation, (v) **Filtering:** filtering processing, (vi) **Phase:** phase manipulation, and (vii) **Echo:** reverberation process. Table 3 showed the content of each category. A competitor of CD method in these tests is DSS method which is the most robust method in the robustness test for signal modifications (Sec. 4.4.1).

Figure 7(a) shows the results of the benchmark tests of the CD method. The vertical axis is the attack category. The horizontal axis is the bit accuracy. The results indicate that bit-detection rates for (i) **Noise**, (ii) **Amplitude**, (iii) **Bit**, and (v) **Filtering** are 75% or more. These revealed that the CD method are robust against (i) Noise, (ii) **Amplitude**, (iii) **Bit**, and (v) **Filtering**. The results show that the bit-detection rates for (iv) **Data**, (vi) **Phase**, and (vii) **Echo** are less than 75%. The attacks of (iv) **Data**, (vi) **Phase** , (vii) **Echo** are signal processing that distorts the phase of the watermarked signal. Therefore, the CD method which embedded a watermark in phase domain is not robust to the attack of (iv) **Data**, (vi) **Phase**, (vii) **Echo**. Figure 7(b) shows the results of the benchmark tests of the DSS method. The results indicate that the DSS method is predictably robust against many attacks. The accuracy for (vi) **Phase** is, however, less

than 33%. These indicate that the DSS method is not robust to the attack of (vi) **Phase**.

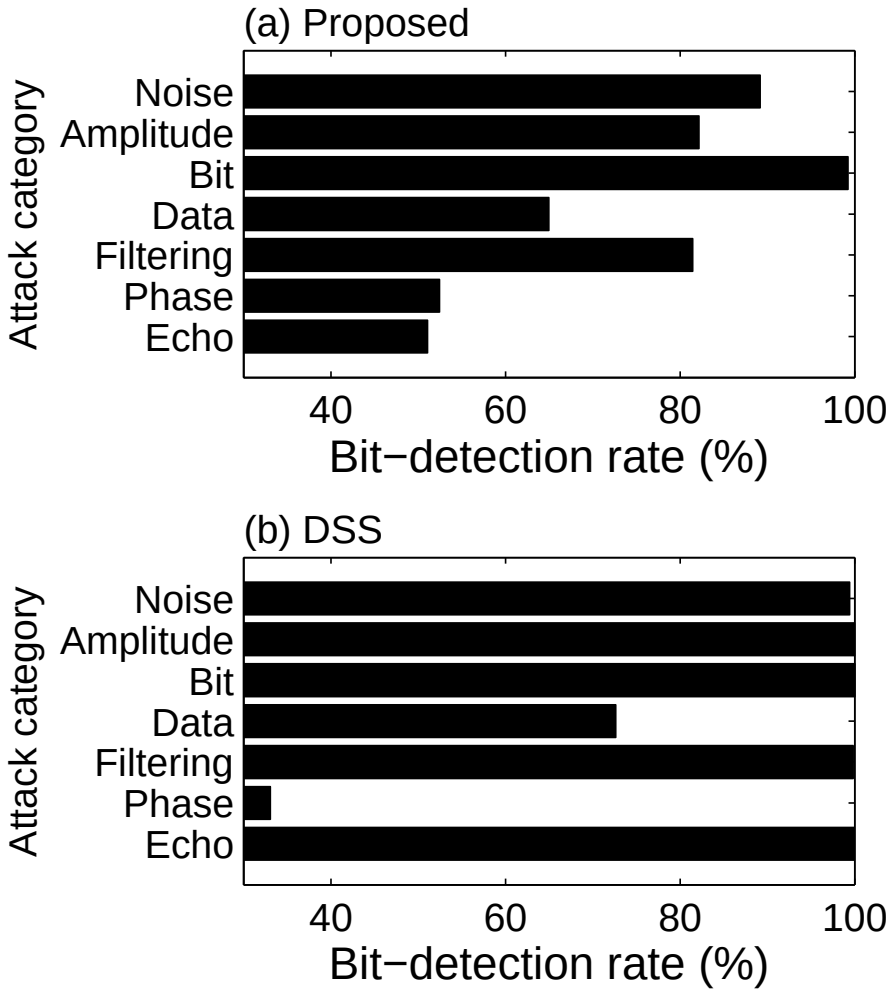

FIGURE 7. Results of the benchmark tests in (a) proposed method and (b) DSS.

4.5. **Discussion.** From the results of objective/subjective evaluations and robustness tests, the features of four typical methods we obtained were reconfirmed with these predicted features as listed in Table 1. We found that LSB had a drawback in robustness for watermarking although it could satisfy (a) inaudibility and (b) confidentiality requirements. We also found that DSS and ECHO could satisfy (c) robustness, but DSS had a drawback with (a) inaudibility and ECHO with (b) confidentiality. Although PPM, especially, was predicted to be the best of these methods, the present results indicated that it had slight problems with (a) inaudibility and (c) robustness. Since we did not have the original code for PPM, these may be able to be resolved if PPM is precisely tuned.

In summary, the typical watermarking methods used in LSB, DSS, ECHO, and PPM approaches could partially satisfy the three requirements (a)-(c). Table 1 suggests us that it is very difficult to achieve inaudible watermarking that can satisfy all three requirements, in particular, both requirements of (a) inaudibility and (c) robustness, simultaneously. In contrast, from the results of these evaluations, we found that the proposed technique adequately satisfied both requirements of (a) inaudibility and (c) robustness, simultaneously, and that the proposed method could partially satisfied another requirement of (b) confidentiality. Because, since we assumed that the data-detection process was achieved as non-blind detection in the first step, there are still remaining studies with regard to (b)

confidentiality, the realization of blind detection for watermarks and investigation with

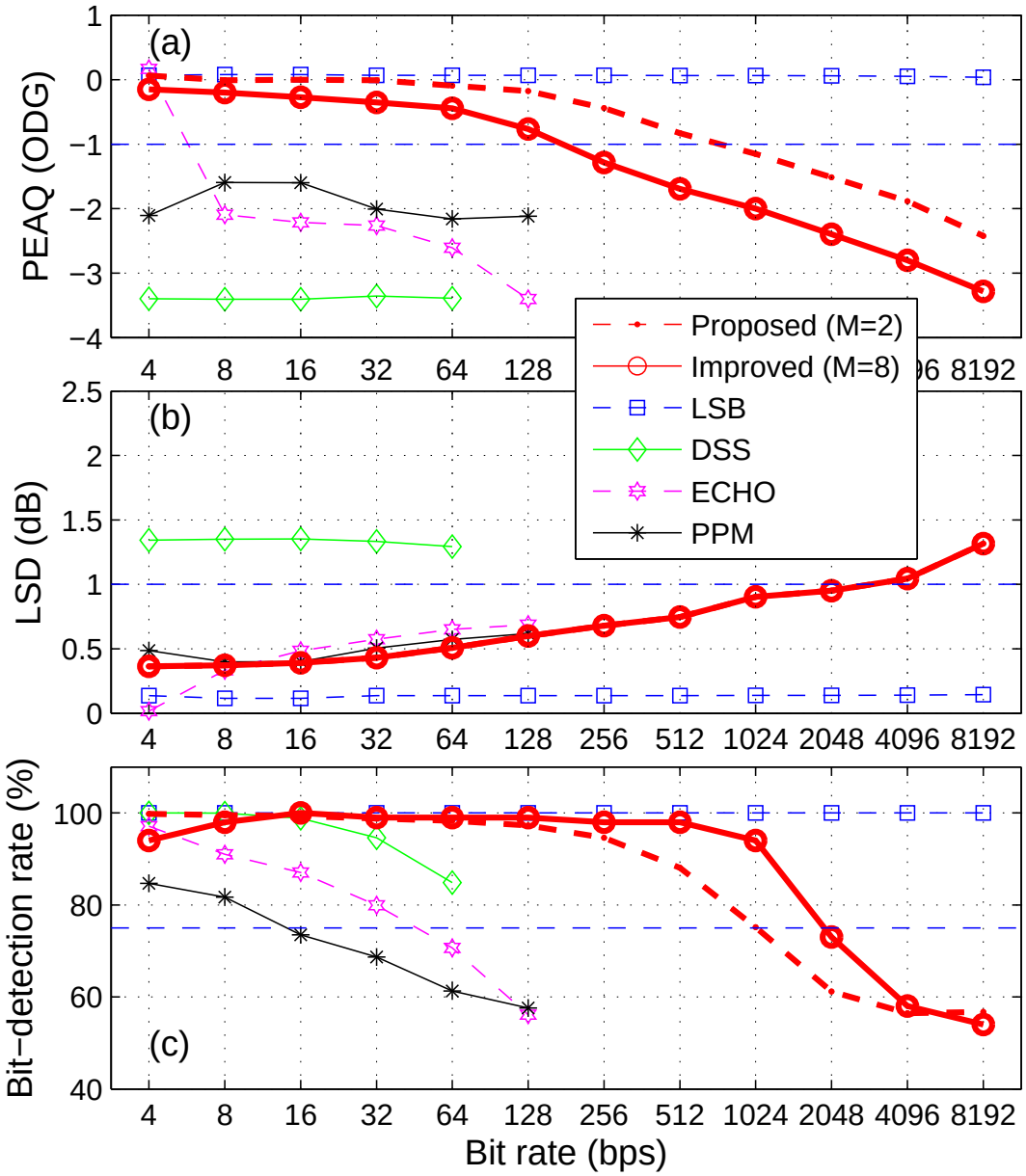

FIGURE 8. Results of evaluation for the proposed method: (a) PEAQ, (b) LSD, and (c) bit-detection rate.

regard to collusion attack. Although these are our next step in future works, it is regarded that the proposed approach can adequately satisfy all requirements by resolving the remaining issues. The results we obtained from all evaluations are significant advantages of the new technique and these results suggest that our proposed approach could provide a useful way of protecting copyright.

5. **Improved method.** In previous section, we comparatively evaluated the proposed approach for inaudible digital-audio watermarking with four other methods (LSB, DSS, ECHO, and PPM) by carrying out objective and subjective evaluations, bit-detection test, and robustness tests. These results revealed that the proposed method could adequately satisfied requirements (a) and (c). In this section, we then investigated how well this method can be used to embed watermarks into digital-audio signals, to clarify embedding limitations with the proposed method.

5.1. **Embedding limitations with the proposed method.** As the same in Sec. 4.2, we comparatively evaluated our proposed method with four others (LSB, DSS, ECHO, and PPM) by carrying out three tests: PEAQ, LSD, and bit-detection rate, in the cases of $N_{\mathrm{bit}}s$ were 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, and 8192 bps, to investigate embedding limitations with the proposed method.

Figure 8 plots the results obtained from the comparative evaluations. All plots and values were averaged for all stimuli. The thresholds for evaluation (PEAQ of −1, LSD of 1 dB, and bit-detection of 75%) were the same as those we used in Section 4. As listed in Table 1, we found that LSB had a drawback in (c) robustness for watermarking although it could satisfy inaudibility and confidentiality requirements (a) and (c) even if Nbit s increased from 4 to 8192 bps. Although embedding limitations with LSB method seems to be very high, these limitations will be definitely restricted by issue of (c) robustness. We also found that DSS and ECHO could satisfy robustness (c), but DSS had a drawback with (a) inaudibility and ECHO with (b) confidentiality. In particular, we found that the results of ECHO method, PEAQ and bit-detection rate, decreased as $N_{\mathrm{bps}}s$ increased, and that LSD of ECHO method increased as $N_{\mathrm{bps}}s$ increased. It is regarded that embedding limitations with ECHO method amounted to very low bit-rates. PPM had a reasonable in LSD measure except with PEAQ, however, these may be able to be resolved if PPM is precisely tuned.

In contrast, objective evaluations of the proposed approach indicated that PEAQs were under the evaluation threshold (> -1) in which the $N_{\mathrm{bit}}s$ ranged from 4 to 512 bps while the PEAQs were gradually reduced as the Nbit s increased over 128 bps. We also found that LSDs increased as Nbit s increased and that they were under this evaluation threshold (¡ 1 dB) under all conditions. In addition, we found that the bit detection rates were less than the evaluation threshold (75%) in which Nbit s ranged from 4 to 1024 bps. This ensured that the proposed method with N_{bit} = 1024 bps could be used to detect the watermarks from the watermarked signals. However, it was easily predicted that N_{bit} will be restricted by results of robustness tests.

These considerations predicted that embedding limitations with the proposed method amounted to around 512 bps and these limitations will be restricted by results of robustness tests. It was found that there is a trade-off between embedding limitations derived from requirements of (a) inaudibility and (c) robustness. Therefore, we have to reconsider the filter architecture for the CD filters in order to reduce embedding limitations with the proposed method.

5.2. **Parallel architecture.** We improved our proposed method to reduce embedding limitations with the method by using a parallel architecture for the first-order IIR filter (CD filter) in Eq. (1). In the proposed method, 1-bit expression ("0" and "1") was assigned at one-frame, as shown in Figs. 2 and 3. Based on the bit expression (L-bits) for $M = 2^L$ at each frame, it is possible to control M-CDs using the parallel architecture for $M-CD$ filters, as shown in Fig. 9. If signal distortion due to this style of embedding can be disregarded in requirement of (a) inaudibility and embedded data can be correctly detected in requirement of (c) robustness, embedding limitations with the improved method can be further reduced in comparison with those of the proposed method, as shown in Fig. 3.

The improved method consists of two processes: embedding and detecting data, as outlined in the flow diagrams in Fig. 10. For $L = 1$ (i.e., $M = 2$), these processes were the same as the processes in our proposed method.

M-CD filters ($H_0(z), H_1(z), ..., H_{M-1}(z), M = 2^L$) were used to embed watermarks into the audio signals in the data embedding process. The phase components of the original

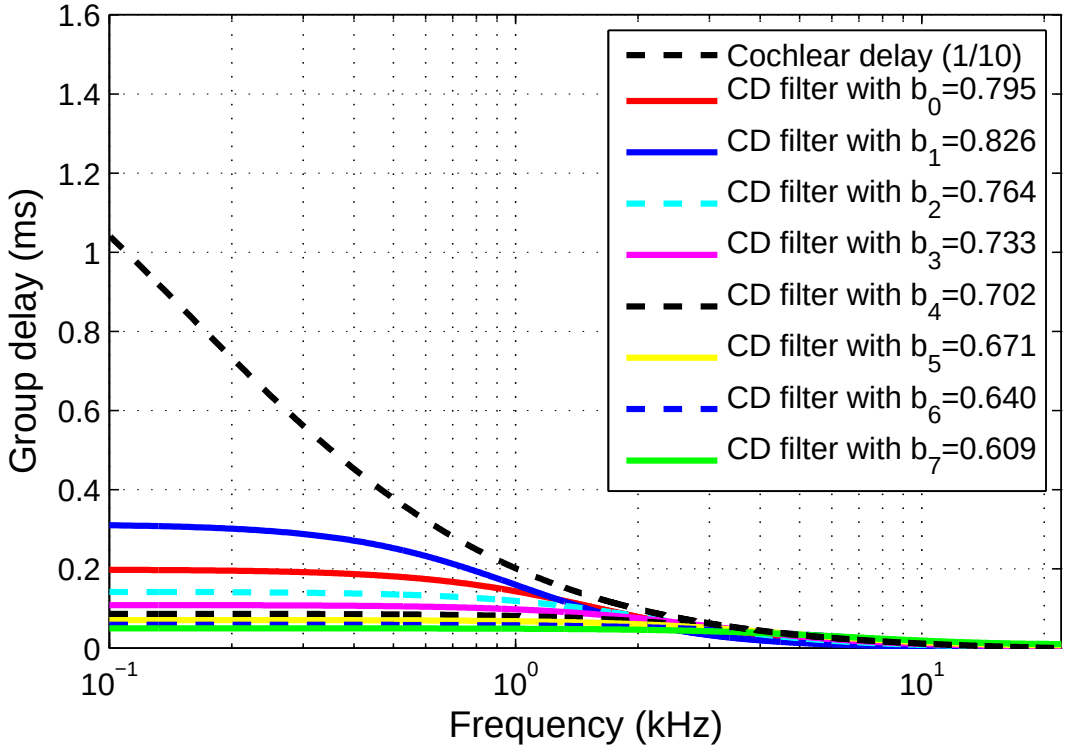

FIGURE 9. Cochlear-delay and group-delay characteristics in parallel architecture for CD filters.

signal were enhanced by these M-CD filters. For example, for $M = 2^3 = 8$ (3-bits expression: 000, 001, ..., 111), eight types ($M = 8$) of cochlear delays according to $b_0, b_1, ..., b_7$ were used. In this case, parameters of M-CD filters, $b_0, b_1, ..., b_7$, and corresponded CDs are drawn in Fig. 9.

The data detection process involves estimating the group delays (arg $H_0(z)$, arg $H_1(z)$, ..., arg $H_{M-1}(z)$) from the phase difference between the original and the watermarked sounds ($\phi(\omega)$) to the respective phase spectrum of the filter ($\triangle\phi_k = |\Phi(\omega) - \arg H_k(\omega)|$) to detect the embedded data. The selected filter number m corresponds to the bit expression (e.g., $m = 7$ and "111" for watermarks).

6. **Evaluations of embedded limitations.** We evaluated our proposed ($M = 2$) and improved methods ($M = 4$, 8, 16, and 32) by carrying out four objective experiments: PEAQ, LSD, bit-detection, and robustness tests, to investigate the extent of embedding limitations. All stimuli that were used in these evaluations were the same in Sec. 4.1. The bit-rates in these experiments were 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, and 8192 bps.

6.1. **PEAQ test.** We carried out a PEAQ test [20] to evaluate to what extent users could objectively perceive the embedded data from the watermarked signals. Figure 11 plots the averaged ODGs of the PEAQs for the watermarked signals for parallel filters of (a) $M = 2$, (b) $M = 4$, (c) $M = 8$, (d) $M = 16$, and (e) $M = 32$. The circles indicate the averaged ODGs and the error bars indicate the standard deviations for these ODGs. The PEAQs were under the evaluational threshold (> -1) in which the bit rate ranged from 4 to 128 bps while the PEAQs gradually reduced as the bit-rate increased over 256 bps. This upper limitation reduced as M increased from 2 to 32. The results ensured that the improved method at 128 bps and an M of 8 could be used to embed watermarks into the original signals to satisfy requirement (a), while our proposed method ($M = 2$) with 512

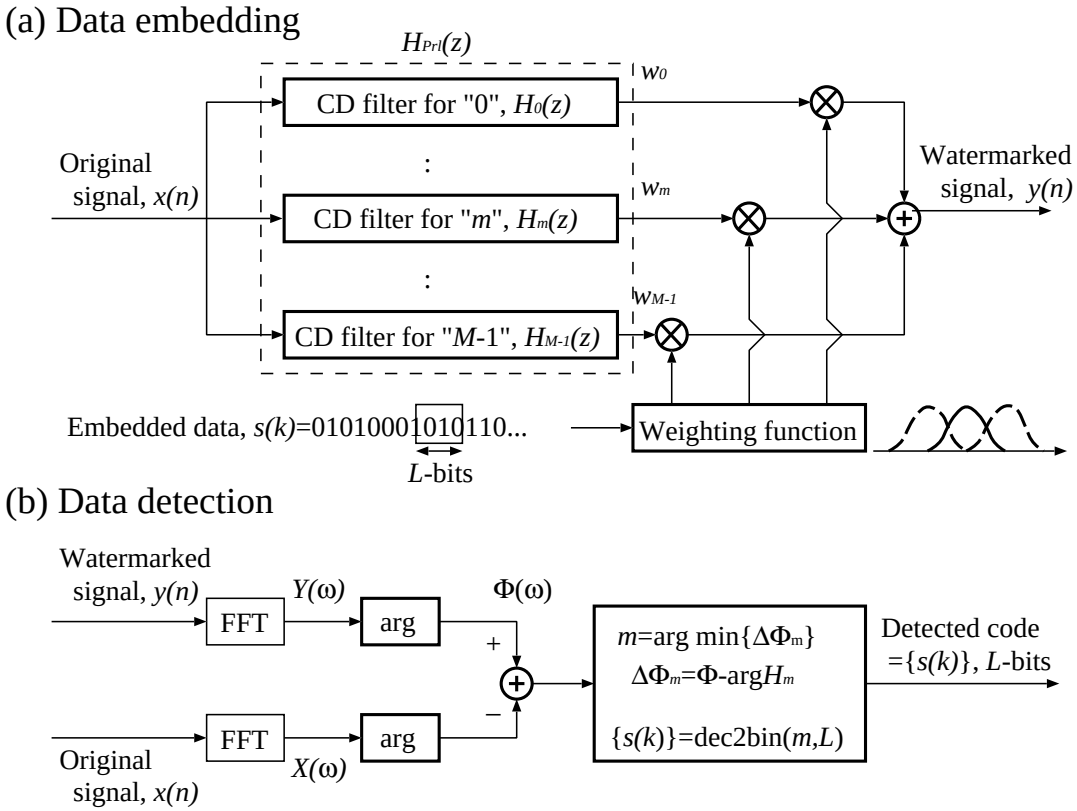

FIGURE 10. Block diagram for data embedding and data detection in parallel architecture for CD filters.

bps could be used to embed the watermarks into the original signals.

6.2. **LSD test.** We carried out an objective experiment (LSD measures) to evaluate the sound quality of the watermarked signals. Figure 12 has the averaged LSD for the watermarked signals. The circles indicate the averaged LSD and the error bars indicate the standard deviations. These results ensure that the proposed method with a bit rate of 4096 could be used to embed the watermarks into the original signals to satisfy requirement (a). The LSDs were under the evaluation threshold (1 dB) in which the bit rates ranged from 4 to about 2048 bps. This upper limitation reduced as M increased. The results ensured that the improved method with a bit rate of 2048 and an M of 8 could be used to embed the watermarks into the original signals.

6.3. **Bit-detection test.** We carried out a bit-detection test to evaluate how well the proposed and improved methods could accurately detect embedded data from the watermarked audio signals. The same original signals were used in this experiment. The bit-detection rates for all signals were evaluated as a function of the bit rate. A threshold of 75% was chosen as the limitation for embedding to evaluate the bit-detection rate in this experiment.

Figure 13 plots the averaged bit-detection rate of the watermarked signals. The detection rates were under the evaluation threshold ($> 75\%$) in which the bit rate ranged from 4 to 512 bps. This ensured that the improved method with 1024 bps and an M of 8 could be used to detect the watermarks from the watermarked signals to satisfy requirement (2), while our proposed method with 1024 bps could be used to detect the watermarks from the watermarked signals.

6.4. **Robustness tests.**

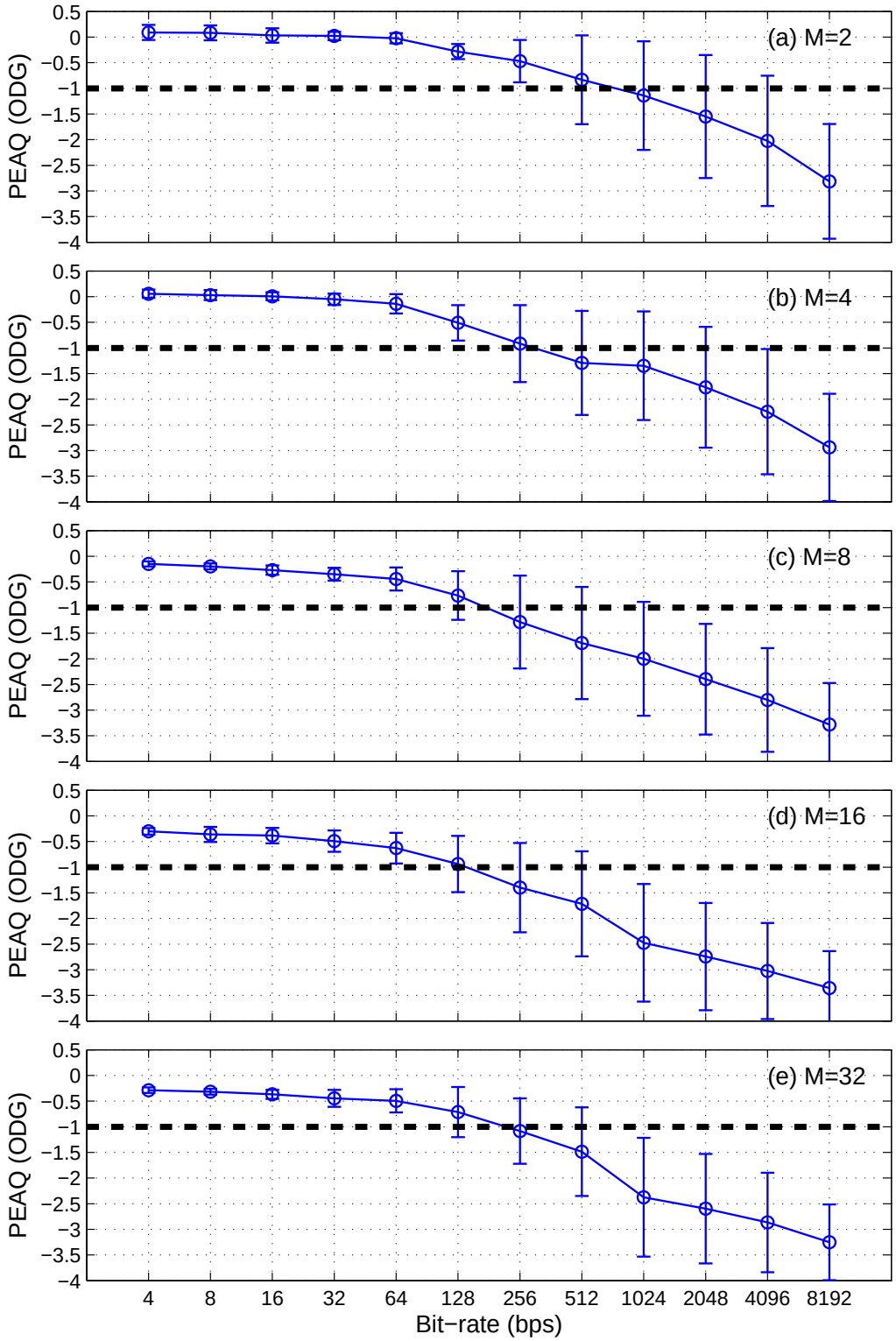

FIGURE 11. Results of PEAQ for (a) our previous method ($M = 2$) and improved method: (b) $M = 4$, (c) $M = 8$, (d) $M = 16$, and (e) $M = 32$.

6.4.1. *Robustness test for signal modification.* We next carried out three robustness tests to evaluate how well the methods could accurately and robustly detect embedded data from the watermarked-audio signals. As the same in Sec. 4.4, the main manipulation conditions used were: (i) down sampling (44.1 kHz → 20, 16, and 8 kHz), (ii) amplitude

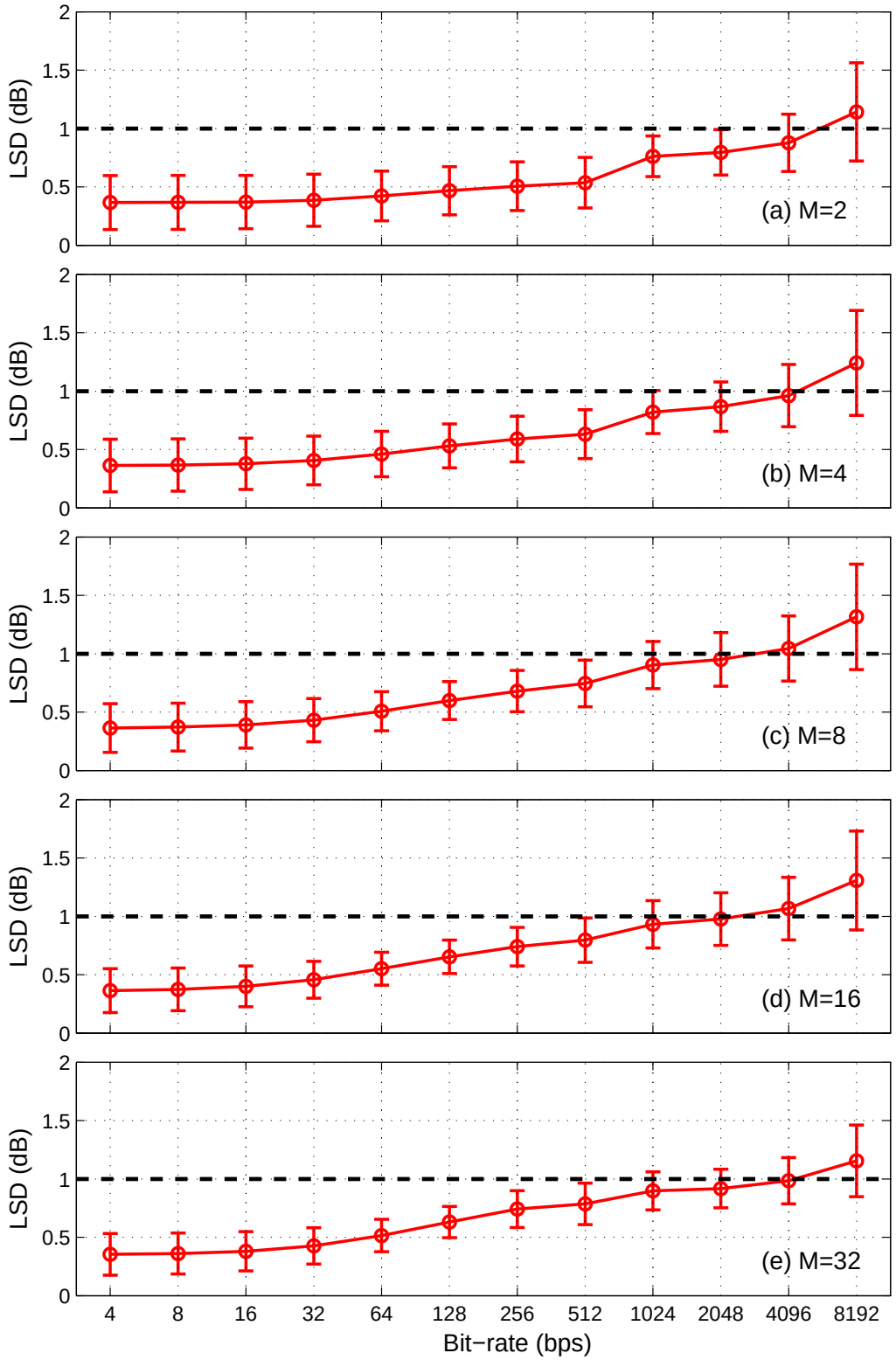

FIGURE 12. Results of LSD for (a) our previous method ($M = 2$) and improved method: (b) $M = 4$, (c) $M = 8$, (d) $M = 16$, and (e) $M = 32$.

manipulation (16 bits → 24-bit extension and 8-bit compression), and (iii) data compression (mp3: 128 kbps, 96 kbps, and 64 kbps-mono).

Table 4 lists the results of evaluations for the proposed and improved method. The "–" means that the detection rate was over the evaluation threshold. The bit detection was 1024 bps at an M of 8 where there was non-process. In contrast, embedding limitations

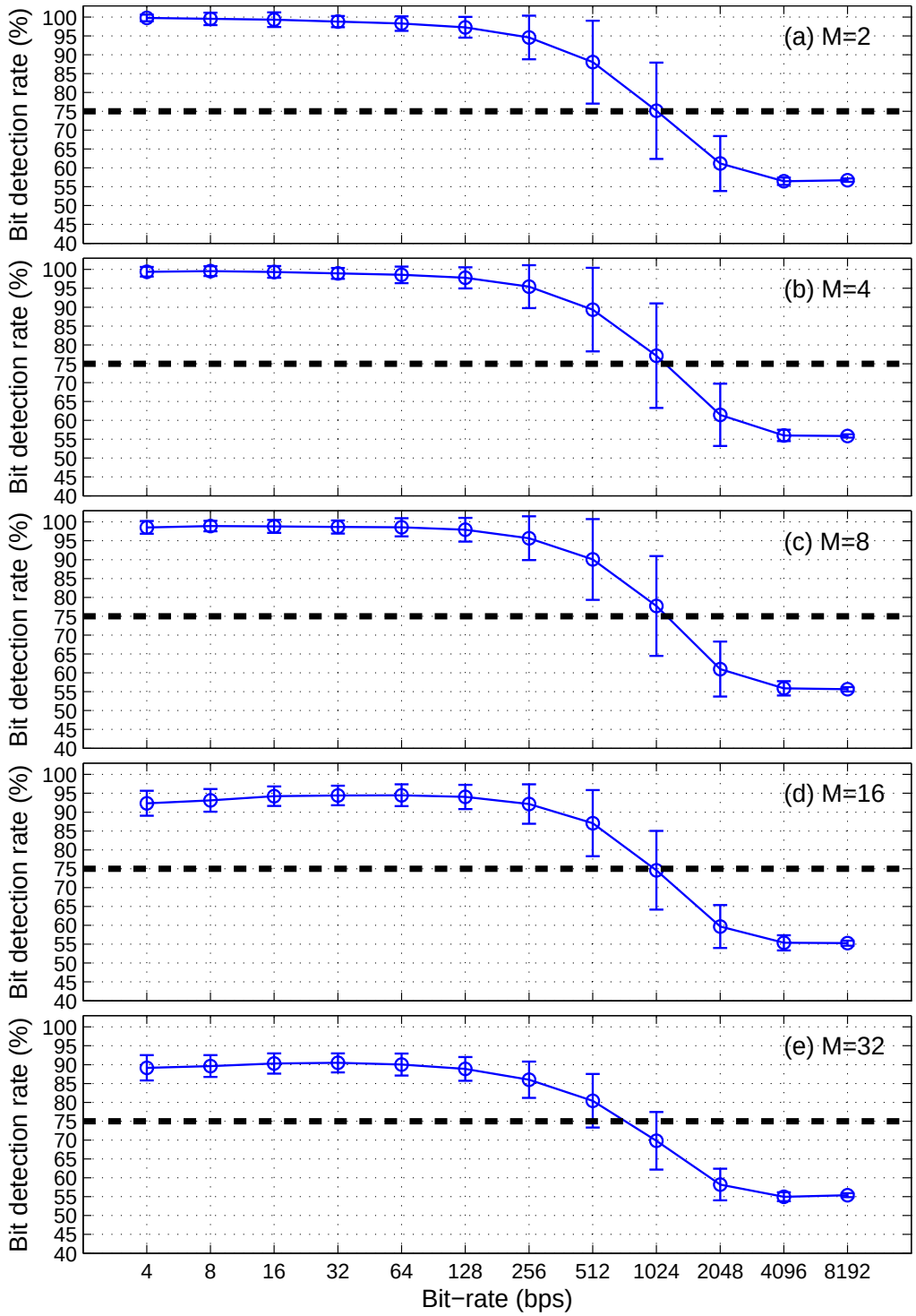

FIGURE 13. Results of bit-detection rate for for (a) our previous method (M = 2) and improved method: (b) = 4, (c) M = 8, (d) M = 16, and (e) M = 32.

for the bit-detection rate under strong manipulation conditions (mp3 96-kbps) was 128 bps at an M of 8. These results indicate that the improved approach could accurately and robustly watermark copyrighted data in original audio content.

6.4.2. *StirMark benchmark test.* We finally carried out robustness test for StirMark benchmark in order to clarify the robustness of the CD methods (M = 2, 4, 8, 16, and 32) against cracking watermark.

TABLE 4. Results of robustness tests for embedding limitations (bps).

Modification	$M = 2$	4	8	16	32
Non-process	1024	1024	**1024**	1024	512
DS 20 kHz	512	512	512	512	512
DS 16 kHz	512	512	512	512	512
DS 8 kHz	256	256	256	128	128
BC 24 bit	512	512	512	512	256
BC 8 bit	512	512	512	512	256
mp3 (128k)	128	128	128	64	—
mp3 (96k)	128	128	**128**	—	—
mp3 (64k)	128	128	128	—	—

Figure 14 shows the results of the StirMark benchmark tests of the CD methods. The vertical axis is the attack category. The horizontal axis is the bit accuracy. The results indicate that the bit-detection for (i) **Noise**, (ii) **Amplitude**, (iii) **Bit**, and (v) **Filtering** in $M = 2$, 4, 8, and 16 are 75% or more. The results also showed that the bid-detection rate for (iv) **Data**, (vi) **Phase**, and (vii) **Echo** in $M = 2$, 4, 8, and 16, and the bitdetection rate in $M = 32$ except for (iii) Bit are less than 75%. However, the results revealed that the bit-detection rate for the attacks of (iv) **Data**, (vi) **Phase**, and (vii) **Echo** are less than 75%. This is because these manipulations distort the phase of the watermarked signal.

In summary, these revealed that the CD methods are robust against (i) **Noise**, (ii) **Amplitude**, (iii) **Bit**, and (v) **Filtering** while these are, in general, not robust to the attacks of (iv) **Data**, (vi) **Phase** , and (vii) **Echo**. In addition, CD methods with $M = 2$, 4, and 8 can be regarded as reasonably robust to most of StirMark attacks.

6.5. **Discussion.** From the results of objective evaluations and robustness tests, embedding limitations with the proposed and improved methods were derived to satisfy all the requirements (a)-(c). Embedding limitations with the proposed method, derived from objective evaluations (PEAQ and LSD), bit-detection test, and robustness tests, were 512, 1024, and 128 bps, respectively. Hence, the overall embedding limitation with the proposed method was 128 bps.

In contrast, embedding limitations with the improved method were depended upon the number of CD filters in parallel architecture. From the results of robustness tests, the CD methods with $M = 2$, 4, and 8 can be regarded as reasonable. In the case of $M = 2$, the improved method was the same as the proposed method so that overall embedding limitation with the improved method with $M = 2$ was 128 bps. In the case of $M = 4$, the results demonstrated that the improved method at 128 bps could be used to embed watermarks into the original signals and to accurately and robustly detect the embedded data from the watermarked signals. This means that the overall embedding limitation with the improved method at M of 2 (L of 2) was 256 ($= 128 \times 2$) bps. As the same manner, the overall embedding limitation with the improved method at M of 8 (L of 3), hence, can be regarded as 384 ($= 128 \times 3$) bps. The improved method at M of 8 is the best in our current proposed approach. Results of comparative evaluations for the improved method ($M = 8$) with regard to PEAQ, LSD, and bit-detection rate are also shown in Fig. 8.

7. **Conclusions.** We comparatively evaluated the proposed approach with four typical methods (LSB, DSS, ECHO, and PPM). The results of subjective and objective evaluations revealed that the proposed method could be used to embed inaudible watermarks

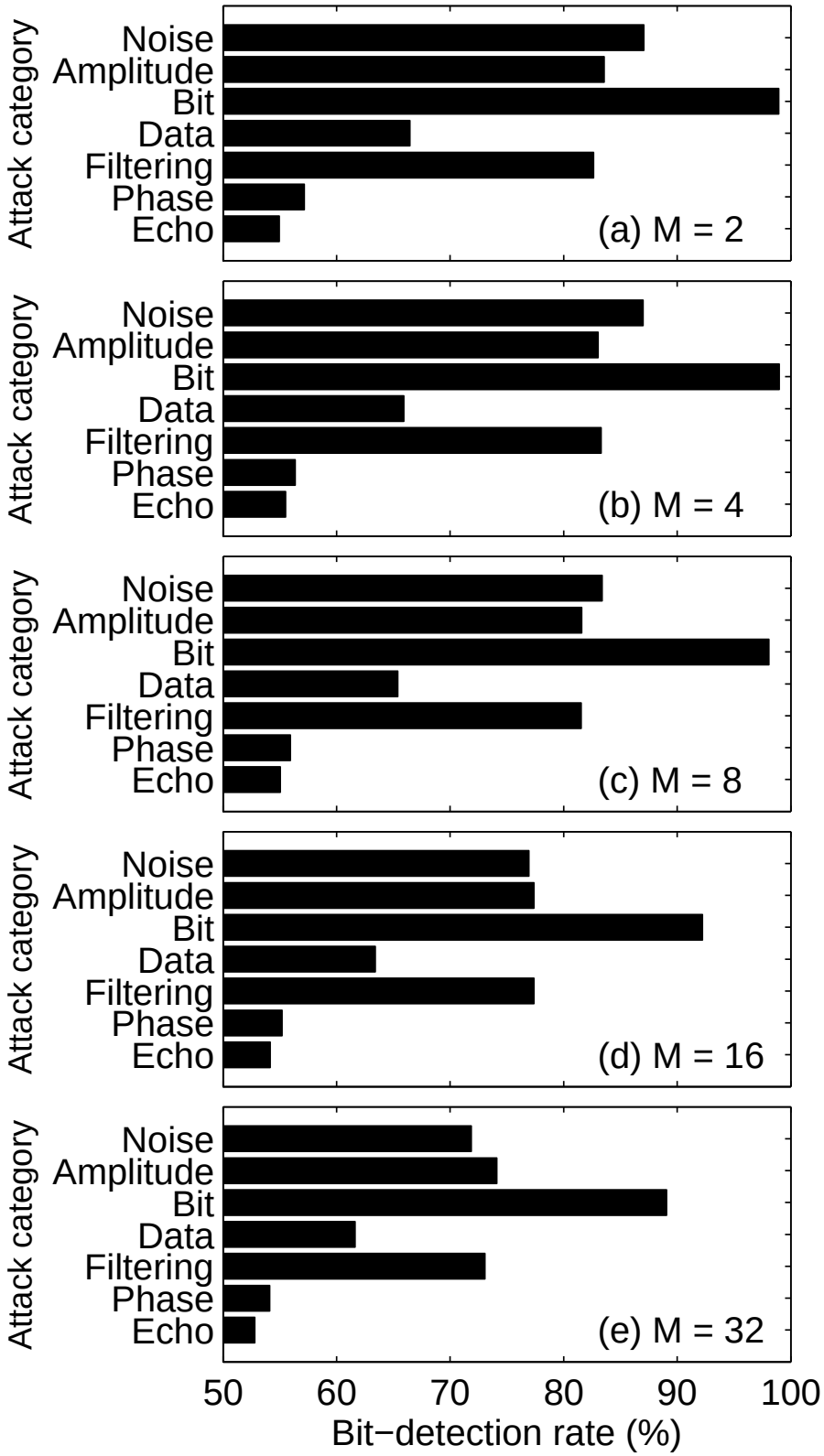

FIGURE 14. Results of the benchmark tests in improved method: (a) M = 2, (b) M = 4, (c) M = 8, (d) M = 16, and (e) M = 32.

into the original signals, and that subjects could not detect the embedded data in any of the watermarked signals we used. Our evaluations of robustness demonstrated that it could precisely and robustly detect embedded data such as those copyrighted with a

watermarked signal to protect them against various signal modifications. These comparative results suggest that our proposed approach could provide a useful way of protecting copyright.

We investigated embedding limitations with our proposed and improved methods of audio watermarking by carrying out five tests on LSD, PEAQ, bit-detection, and robustness tests (signal modifications and StirMark benchmark). To satisfy all the requirements (a)-(c), the results revealed that the improved method at 128 bps and an M of 8 could be used to embed watermarks into the original signals and to accurately and robustly detect the embedded data from the watermarked signals, while our proposed method at 128 bps and M of 2 could also be used. This also means that the best results were achieved with $M = 2^3$ CD filters and the embedding limitation with the improved method was 128 bps. Hence, the overall embedding limitation with the improved method was 384 ($= 128 \times 3$) bps, while that with our proposed method was 128 bps.

Our next step in future work, is to (1) consider the blind detection of embedded data from watermarked signals such as that in the study done by Sonoda *et al.* [24], and (2) investigate verification with regard to requirement of (b) confidentiality such as collusion attack.

8. **Acknowledgments.** This work was supported by a Grant-in-Aid for Challenging Exploratory Research (No. 21650035) made available by Japan Society for the Promotion of Science and Linking mechanism of research results to practical application made available by Japan Science and Technology Agency.

REFERENCES

[1] E. Isao, S. Yoiti, and X. Niu, Special issue on information hiding and multimedia signal processing, *International Journal of Innovative Computing, Information & Control*, vol. 6, no. 3(B), pp. 1207-1208, 2010.

[2] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn , Information hiding - a survey, *Proc. of IEEE special issue on protection of multimedia content*, vol. 87, no. 7, pp. 1062-1078, 1999.

[3] N. Cvejic and T. Seppänen, *Digital audio watermarking techniques and technologies*, IGI Global, Hershey, PA 2007.

[4] STEP2001, News release, final selection of technology toward the global spread of digital audio watermarks, Japanese Society for Rights of Authors, Composers and Publishers. http://www.jasrac.or.jp/ejhp/release/2001/0629.html.

[5] A. Nishimura, Information hiding in audio signals: Digital watermarking and steganography, *J. Acoust. Soc. Jpn.*, vol. 63, no. 11, pp. 660-667, 2007.

[6] L. Boney, A. H. Tewfik, and K. N. Hamdy, Digital watermarks for audio signals, *Proc. of International Conference on Multimedia Computing and Systems(ICMCS)*, pp. 473-480, 1996.

[7] T. Dau, O. Wegner, V. Mallert, and B. Kollmeier, Auditory brainstem responses (ABR) with optimized chirp signals compensating basilar membrane dispersion, *J. Acoust. Soc. Am.*, vol. 107, pp. 1530-1540, 2000.

[8] D. Gruhl, A. Lu, and W. Bender, Echo hiding, *Proc. of the 1st Information Hiding Workshop*, pp. 295-315, 1996.

[9] R. Nishimura and Y. Suzuki, Audio watermark based on periodical phase shift, J. Acoust. *Soc. Jpn.*, vol. 60, no. 5, pp. 269-272, 2004.

[10] A. Takahashi, R. Nishimura, and Y. Suzuki, Multiple watermarks for stereo audio signals using phase-modulation techniques, *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 806-815, 2005.

[11] C. J. Plack(eds), The sense of hearing, Lawrence Erlbaum Association, London, 2005.

[12] M. Akagi and K. Yasutake, Perception of time-related information: Influence of phase variation on timbre, *Technical report of IEICE.*, vol. 98, EA1998-19, pp. 15-22, 1998.

[13] K. Ozawa, Y. Suzuki, and T. Sone, Monaural phase effects on timbre of two-tone signals, *J. Acoust. Soc. Am.*, vol. 93, no. 2, pp. 1007-1011, 1993.

[14] K. K. Paliwal, and L. Alsteris, Usefulness of phase spectrum human speech perception, *Proc. of Eurospeech*, pp. 2117-2120, Geneva, 2003.

[15] M. Unoki, and D. Hamada, Audio watermarking method based on the cochlear delay characteristics, *Proc. of IIHMSP08, Harbin, China*, pp. 616-619, 2008.

[16] M. Unoki, and D. Hamada, Method of digital-audio watermarking based on cochlear delay characteristics , *International Journal of Innovative Computing, Information and Control*, vol. 6, no. 3(B), pp. 1325-1346, 2010.

[17] E. Aiba, and M. Tsuzaki, Perceptual judgement in synchronization of two complex tones: Relation to the cochlear delays, *Acoust. Sci. & Tech.*, vol. 28, no. 5, pp. 357-359, 2007.

[18] E. Aiba, M. Tsuzaki, S. Tanaka, and M. Unoki, Judgment of perceptual synchrony between two pulses and verification of its relation to cochlear delay by an auditory model, *Japan Psychological Research 2008*, vol. 50, no. 4, pp. 204-213, 2008.

[19] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, RWC music database: music genre database and musical instrument sound database, *Proc. of International Society for Music Information Retrieval (ISMIR2003)* , pp. 229-230, 2003.

[20] P. Kabal, An Examination and Interpretation of ITU-RBS.1387: Perceptual Evaluation of Audio Quality, *TSP Lab Technical Report*, Dept, Elect. Comp. Eng., Mc Gill University, Canada, 2002.

[21] Y. Lin, and W. H. Abdulla, Perceptual evaluation of audio watermarking using objective quality measure, *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP2008)*, pp. 1745-1748, 2008.

[22] http://www.jaist.ac.jp/ũnoki/02 demo/

[23] M. Steinebach, F. A. P. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, C. Seibel, N. Fatés, and Ferri, L. C. StirMark Benchmark: Audio watermarking attacks, *Proc. of Coding and Computing 2001*, pp. 49-54, 2001.

[24] K. Sonoda, R. Nishimura, and Y. Suzuki, Blind detection of watermarks embedded by periodical phase shifts, *Acoust. Sci. & Tech.*, vol. 25, no. 1, pp. 103-105, 2004.