International Journal of Innovative Computing, Information and Control
Volume **6**, Number **3(B)**, March **2010**

 ISSN 1349-4198
pp. 1–IHMSP08-13

METHOD OF DIGITAL-AUDIO WATERMARKING BASED ON COCHLEAR DELAY CHARACTERISTICS

Masashi Unoki and Daiki Hamada

School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
{ unoki; hamada }@jaiat.ac.jp

Received December 2008; revised June 2009

Abstract. *This paper proposes a state-of-the-art method of digital-audio watermarking, based on the properties of the human cochlear. It is based on the concept of embedding inaudible watermarks (e.g., copyright data) into an original sound by controlling the phase characteristics of the sound in relation to the characteristics of cochlear delay. The method involves two processes, i.e., embedding and detecting data. The data-embedding process was carried out by applying two types of IIR all-pass filters with cochlear delays, and then selecting the filtered signal from these according to the watermarks. The data-detection process involved estimating the group delay of the filter from the phase difference between the original and watermarked sounds to detect the embedded data. We designed optimal group delays for the filters. We experimentally evaluated the proposed method by carrying out subjective detection tests, two objective detection tests (PEAQ and LSD), bit-detection tests, and three objective tests for robustness (against down sampling, amplitude compression, and mp3-compression). We also comparatively evaluated the proposed method with four other methods (LSB, DSS, ECHO, and PPM). The results revealed that subjects could not detect the embedded data in any of the watermarked sounds we used, and that the proposed approach could precisely and robustly detect the embedded data from the watermarked sounds.*

Keywords: Digital-audio watermarking, Copyright protection, Cochlear delay characteristics, IIR all-pass filter, Group delay

1. **Introduction.** Almost all music content is now available in the form of digital-audio data that can be downloaded from the Internet due to the popularity of PCs and the copious number of high-speed large-capacity networks. This content may be treated with various methods including the use of digital-audio composers/editors and music-distribution services. It is very useful and convenient for users to be able to manipulate digital content on their PCs and this is one of the main advantages of digital-audio content. However, there are serious social issues involved in protecting the copyright of all digital content by preventing it from being illegally copied and distributed over the Internet. Since demands to protect the copyright of digital-audio content have greatly increased in recent years, complete copyright protection is currently a very important topic in this research field.

The most straightforward and commonly used techniques of protecting the copyright of digital content have been based on encryption [1]. Although these have been effective, encryption makes the original sounds inaudible and the severe restrictions that are imposed on legal users are extremely inconvenient. Another frequently used method has been inserting the copyright information in the header of the musical content. Although this allows users to enjoy listening to the content, the embedded copyright information

can easily be removed by deleting the header of the digital content. Therefore, these methods have not sufficiently protected the copyrights of digital-audio data.

Another approach has been to use information-hiding techniques for digitalized multimedia content such as audio, images, movies, and characters (e.g., [2, 3, 4, 5]). An especially useful approach in these techniques has been to use audio watermarking methods to protect the copyrights of digital-audio content. Many methods have been proposed (e.g., [1, 6, 7, 8]). Their aim has been to embed digital codes for the copyright information in the digital-audio content, which are inaudible to users. Since the embedded data cannot be detected by users, they cannot illegally manipulate the watermarked data to remove the copyright information. To provide a useful and reliable form of copyright protection, audio-watermarking methods must satisfy three requirements:

(a) **inaudibility** (inaudible to humans with no sound distortion caused by embedding),
(b) **confidentiality** (secure and accurate concealment of embedded data), and
(c) **robustness** (not affected when subjected to techniques such as data compression).

The first requirement 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 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 is important to ensure the watermarking methods are tamper-proof to resist any manipulations by illegal users.

Typical methods of watermarking have been based 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., [6, 8]), embedding in the Adaptive Differential Pulse Code Modulation (ADPCM) quantizer proposed by Iwakiri and Matsui [9], concealment of information for G.711 speech [10], and the spread spectrum approach (e.g., direct spread spectrum (DSS)) proposed by Boney *et al.* [11]. 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 bit-replacement/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 three requirements, especially with regard to inaudibility. It is therefore difficult to embed inaudible watermarks into the amplitude information.

Gruhl *et al.* [12] have also proposed an echo-hiding approach and Takahashi *et al.* [13] have been proposed a time-spread echo-hiding scheme. The latter is an extension of the former to increase its robustness. These 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.

However, techniques of watermarking based on the characteristics of human-auditory perception such as various masking phenomena have been proposed. There is a method

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

Method	(a) Inaudi.	(b) Confid.	(c) Robust.	Weaknesses
LSB	$\circ$	$\circ$	$\times$	Not Robusted due to signal manipulation
DSS	$\times$	$\circ$	$\circ$	Distorted and poor sound quality
ECHO	$\circ$	$\times$	$\circ$	Easy to detect watermarks
PPM	$\circ^-$	$\circ$	$\circ^-$	Watermarks in pulsive sound audible

based on mpeg proposed by Nakayama *et al.* [14], one based on octave similarity proposed by Muramatsu and Arakawa [15], and that based on the effect of masking for amplitude modulation (AM) proposed by Nishimura [16], for example. These psychoacoustically embed watermarks into the amplitude of digital-audio signals, based on human-auditory perception. Although this generally seems to successfully produce inaudible watermarking, the sound quality of the watermarked signal is reduced when these averaged characteristics of masking are unconsciously perceived by users. Even if the watermarks are embedded into the amplitude in signals as well, humans can easily detect small differences in the amplitudes of the signals due to the fundamental properties of the human-auditory system. It is therefore difficult to embed inaudible watermarks into the amplitudes by using various masking phenomena.

Nishimura *et al.* proposed a method based on periodical phase modulation (PPM) [17, 18]. This was used to embed specific data into the digital-audio signal using PPM. This was also 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 [19, 20].

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.

This paper proposes a novel approach to protecting digital-audio content by using an inaudible method of watermarking based on the characteristics of cochlear delay. It is organized as follows. Section 2 describes cochlear-delay characteristics and psychoacoustical studies related to cochlear delay. It then explains the underlying concept and method of digital-audio watermarking based on cochlear-delay characteristics. Section 3 describes how the parameters of the IIR cochlear-delay filter were determined based on experiments and simulations. Section 4 presents the results of objective/subjective evaluations and assessments of the robustness of the proposed method. Section 5 summarizes the key

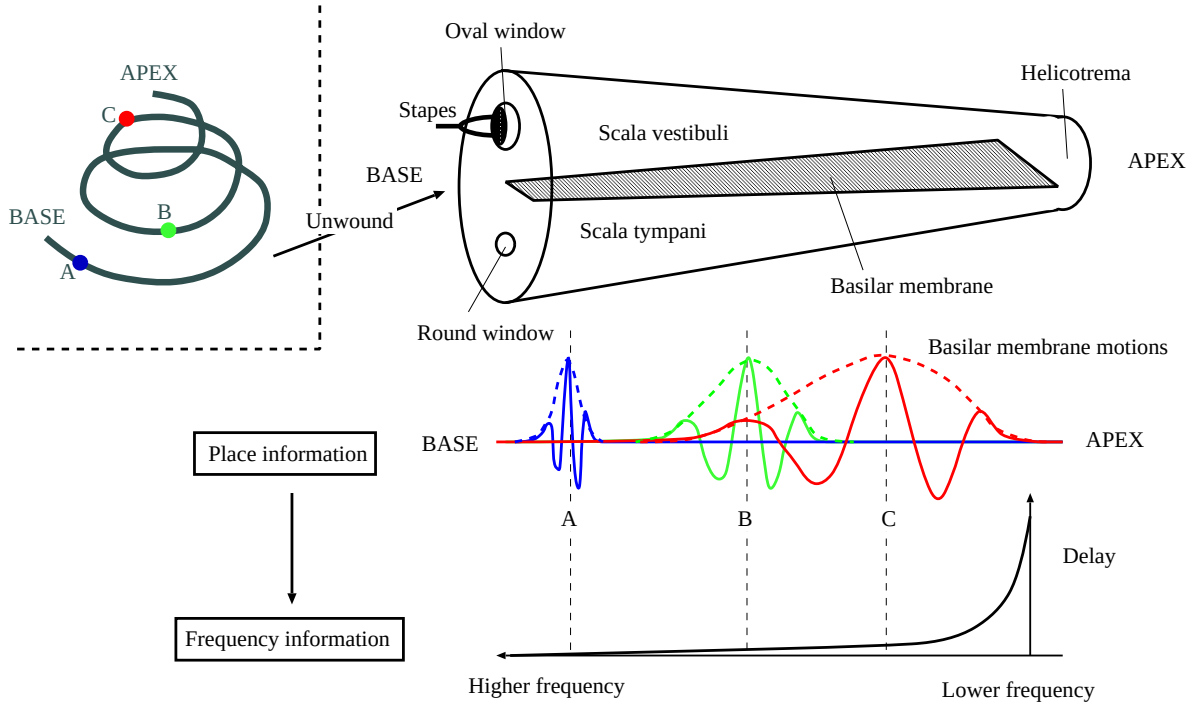

FIGURE 1. Schematic of cochlea with unwound spiral. Vertical dimension has been exaggerated relative to horizontal. Reissner's membrane and the scala media have not been illustrated.

findings with comparative evaluations and discusses improvements that are needed. Section 6 summarizes the proposed scheme for inaudible watermarking and briefly describes future work.

2. Watermarking Based on Cochlear-delay Characteristics.

2.1. Cochlear-delay characteristics. Figure 1 has a schematic of a cochlea. The cochlea is a fluid-filled cavity that is within the same compartment as the scalas vestibuli, media, and tympani and has a thin-tube that has been coiled up to save space, as shown in the top left of Figure 1. The whole structure of the cochlea forms a spiral like a snail's shell, which is not a straight tube. The cochlea can be represented by the shape at the top right of Figure 1 by stretching out this form and uncoiling the spiral (note that this cannot actually be done). The tube is divided along its length by two membranes, i.e., Reissner's membrane and the basilar membrane (BM), which create three fluid-filled compartments: the scalas vestibuli, media, and tympani. A traveling wave of sound enters the cochlea through an opening (the oval window) covered by a membrane. As the fluid in the cochlea is almost incompressible, if the oval window suddenly moves inward, due to pressure from the stapes, Reissner's membrane and the BM are pushed down, and the round window moves outward. It follows that the vibration of the stapes leads to vibration of the BM. As the properties of the BM vary continuously between these extremes along its length, each location on the BM has a particular frequency of sound, i.e., a characteristic frequency. These patterns of vibrations are called a "traveling wave", as shown in Figure 1. The characteristic motion of the traveling wave occurs because there is progressive phase delay from the base to the apex. That is, the vibration of the BM at the apex lags behind that at the base.

Based on the above properties of the human cochlea, let us consider that pulse-like sound is perceived by humans as synchronous. However, this is not synchronous on the

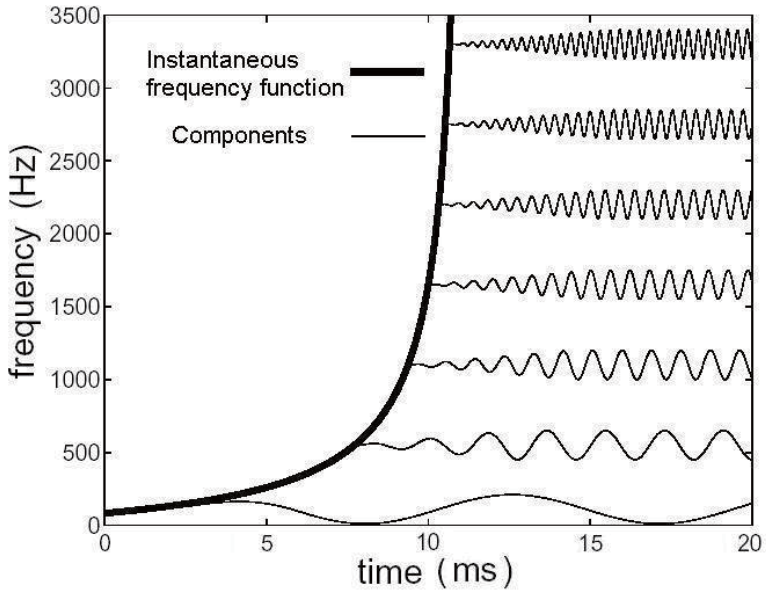

FIGURE 2. Compensation of delay pattern for simulated cochlear-delay characteristics following Dau *et al.* [21]. Bold line indicates compensated delay pattern. Waveforms of part of sinusoidal components follow this line. Adapted from Figure 1 in [22].

BM motion of the cochlea even if the sound components physically begin synchronously. The reason for this is as follows.

A transient sound wave progresses along the BM in the cochlea (from the basal to the apical side), passing through the outer ear as spectral modifications (air-pressure variations) and the middle ear as an impedance-matching transformer (mechanical vibrations). The BM motions are then converted into neural firings to be transmitted to the brain passing through the cochlear nucleus, superior olive, lateral lemniscus, inferior colliculus, medical geniculate, and auditory cortex (e.g., see Section 4 in [19]). Since the vibrations of the BM result in spatial separation of the frequency components of an acoustic signal, a pulse-like sound must be represented as white-like spectra (having all frequencies) throughout the entire frequencies. The low-frequency components of pulse-like sound require more time to reach the area of maximum displacement in the BM, near the apex of the cochlea, while the higher frequency components of the sound elicit a maximum closer to the base. The time course of pulse-like sound is, therefore, represented as asynchronous components in the BM. This time course is referred to as "cochlear delay" [21].

2.2. **Related studies on perception.** Aiba *et al.* [22, 23] investigated whether cochlear delay significantly affected the perceptual judgment of the synchronization of two sounds or not. They used three types of chirp sounds: (1) a pulse sound (intrinsic cochlear delay), (2) a compensatory delay chirp (i.e., group delay was compensated as zero in the BM), and (3) an enhanced delay chirp (i.e., group delay was longer than the previous one according to cochlear delay), by following the procedure that Dau *et al.* used [21].

The increasing chirp-frequency pattern originally calculated by Dau *et al.* was based on the one-dimensional linear cochlear model of de Boer [24]. The stiffness of an object is largely responsible for the propagation speed of a traveling wave. Therefore, the basic assumption underlying this increasing frequency pattern is that the physical stiffness of the human BM decreases exponentially along the cochlear partition from the base to the

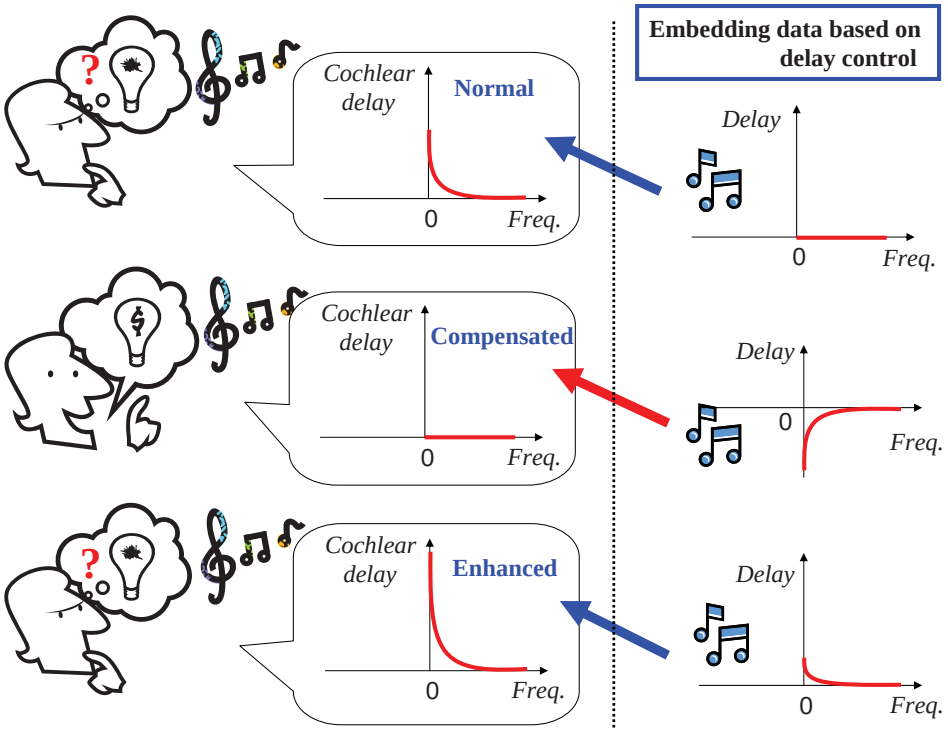

FIGURE 3. Schematic of main idea underlying audio watermarking based on cochlear-delay characteristics. Right panels show the physical delays for embedding. Left panels show the cochlear delay corresponding to each chirp signal.

apex. According to de Boer [24], the BM stiffness can approximately be represented as

$$f = \frac{1}{a}\left(\left[e^{-(\alpha/2)L}\left(1+\frac{t_0-t}{\beta}\right)\right]^{-2/\alpha c} - 1\right), \tag{1}$$

where $a = 0.006046$ (Hz^{-1}), $\tilde{c} = 1.67$ mm, and $c = \tilde{c}/\ln 10$. Here, $L = 34.85$ mm, $\alpha = 3$, and $\beta = 2/\alpha\sqrt{2\rho/h/c_0}$. The $c_0 = 10^{-9}$ dyne/cm^3, $\rho = 1$ g/cm^3, $h = 0.1$ cm, and $t_0 = 11.2$ ms. These parameter values were set according to the results of de Boer [24]. Figure 2 shows the increasing frequency pattern of the compensatory-delay condition that Aiba *et al.* used [22]. The bold line indicates compensatory delay as a function of frequency, and the enhanced delay of the waveforms of the seven sinusoids have been superimposed as examples.

Their experimental task was designed to estimate the threshold of judgment, i.e., how much time was required by subjects to detect the onset of asynchrony between sounds. The results revealed that the threshold of judgment for signal (2) was the highest (i.e., subjects needed a relatively long time lag to judge the synchrony for signal (2)). The threshold of judgment for signal (3) was almost the same as that for signal (1), which meant that the accuracy of determining synchrony did not improve even if cochlear delay was compensated for. Their results suggest that the auditory system cannot distinguish sound with enhanced delay and non-processing sound.

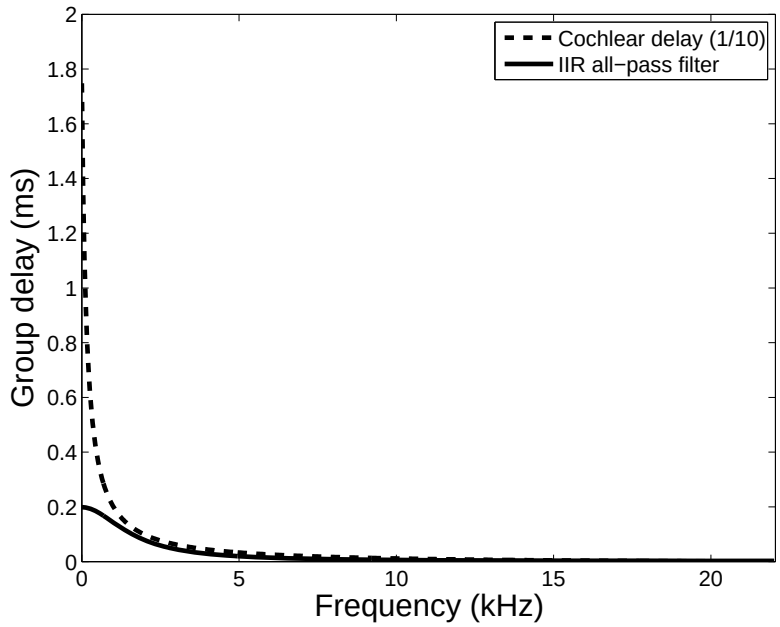

FIGURE 4. Cochlear delay and group-delay characteristics of IIR all-pass filter in Eq.(2). Cochlear delays are scaled by 1/10.

2.3. **Our concept for inaudible watermarking scheme.** Based on Aiba *et al.*'s results [22, 23], as outlined in Figure 3, 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 (copyright data) into an original signal. We thus investigated the feasibility of a method of inaudible embedding based on enhancing group delays (phase information) related to cochlear delays, and propose an approach to digital-audio watermarking based on the characteristics of cochlear delay.

The dashed line in Figure 4 plots the cochlear-delay characteristics described by Dau *et al.* [21], where the delay time was scaled by 1/10. As previously described, the delay time at a lower frequency is somewhat longer than that at a higher frequency, especially within the lower frequency range ($\leq$ 5 kHz). If this cochlear-delay 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 the respective group delays in the filter to those of the digital copyright data ("1" and "0").

We designed the following IIR all-pass filter to model the cochlear-delay characteristics:

$$H(z) = \frac{-b + z^{-1}}{1 - bz^{-1}}, \tag{2}$$

where b is the filter parameter. An IIR all-pass filter is usually used to control delays in which amplitude spectra are passed equally without any loss. Although a higher order IIR filter could be considered to incorporate the characteristics of cochlear delay, we used the simplest IIR filter in Eq.(2).

To determine the optimal value of filter parameter b in Eq.(2), we fitted the group delay characteristics of $H(z)$ to the cochlear-delay characteristics (scaled by 1/10 indicated by the dashed line in Figure 4) by utilizing the least mean squire (LMS) optimization. In

the fit, the group delay $\tau(\omega)$ can be obtained as:

$$\tau(\omega) = -\frac{d\arg(H(e^{j\omega}))}{d\omega}, \tag{3}$$

where $H(e^{j\omega}) = H(z)|_{z=e^{j\omega}}$. The solid line in Figure 4 plots the approximated cochlear delay, i.e., the group-delay characteristics of the IIR all-pass filter in Eq.(3). Here, the optimized value of b is 0.795.

We used two filters in Eq.(2), i.e., $H_0(z)$ and $H_1(z)$ to embed the copyright data ("0" and "1") based on the cochlear-delay characteristics (scaled by 1/10) in the original signal. The phase components of the original signal were enhanced by these filters. Here, the filter parameters, b_0 and b_1, have been defined as b for $H_0(z)$ and $H_1(z)$, respectively. We developed a method of digital-audio watermarking based on the cochlear-delay characteristics with these components.

2.4. **Implementation.** 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. Below, we describe how these processes were implemented.

2.4.1. *Data-embedding process.* Figure 5 has a block diagram of the data-embedding process. Watermarks were embedded as follows:

Step 1: Two IIR all-pass filters, $H_0(z)$ and $H_1(z)$, were designed using different values for b ($b_0 = 0.795$ and $b_1 = 0.865$) to enhance the cochlear delay. These values were determined by taking experimental conditions into consideration (see Section 4 for details).

Step 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. (4) and (5)).

Step 3: The embedded data, $s(k)$, were set to conform to the copyright data, e.g., "010010101100110" as shown in Figure 5.

Step 4: The intermediates, $w_0(n)$ or $w_1(n)$, were selected by stitching the embedded data $s(k)$ ("0" or "1"), and merging them with the watermarked signal, $y(n)$, in Eq.(6).

$$w_0(n) = -b_0 x(n) + x(n-1) + b_0 w_0(n-1), \tag{4}$$

$$w_1(n) = -b_1 x(n) + x(n-1) + b_1 w_1(n-1), \tag{5}$$

$$y(n) = \begin{cases} w_0(n), & s(k) = 0 \\ w_1(n), & s(k) = 1, \end{cases} \tag{6}$$

where $(k-1)\Delta W \leq n < k\Delta W$. Here, n is the sample index, k is the frame index, and $\Delta W = f_s/N_{\text{bit}}$ is the frame length (frame overlap is a half of one frame.). Also f_s is the sampling frequency of the original signal and N_{bit} is the bit rate per second (bps) for embedding the data.

A weighting ramped cosine function was used to avoid discontinuity with this method between the marked segments in the watermarked signal, $w_0(n)$ and $w_1(n)$.

2.4.2. *Data-detection process.* Figure 6 shows the flow for the data-detection process we used. Watermarks were detected as follows:

Step 1: We assume that both $x(n)$ and $y(n)$ are available with this watermarking method.

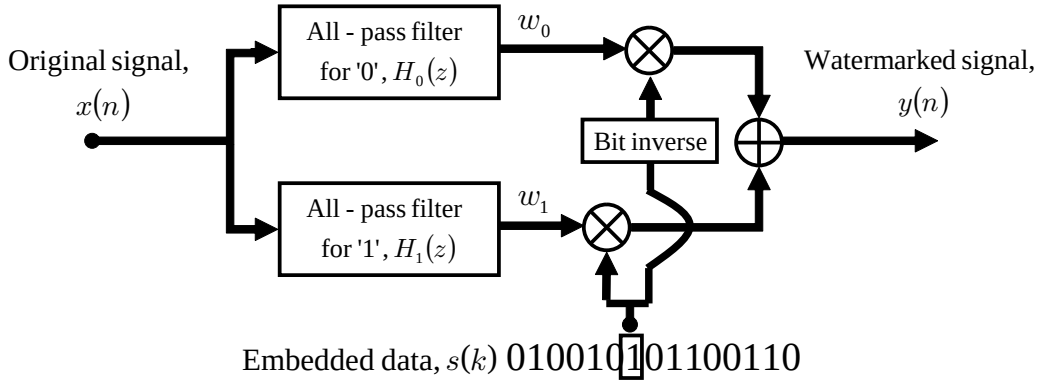

FIGURE 5. Block diagram of data embedding with the proposed watermarking method.

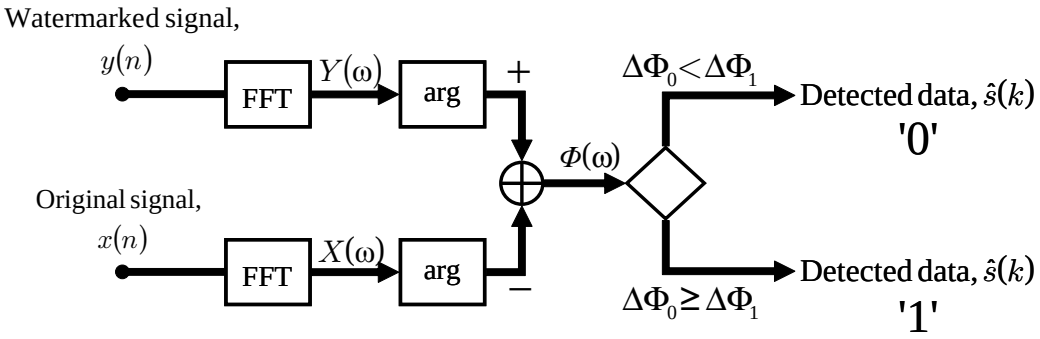

FIGURE 6. Block diagram of data detection with the proposed watermarking method.

Step 2: The original, $x(n)$, and the watermarked signal, $y(n)$, are decomposed to be overlapped segments using the same window function used in embedding the data.

Step 3: The phase difference $\phi(\omega)$ is calculated in each segment, using Eq.(7). FFT[·] is the fast Fourier transform (FFT).

Step 4: To estimate the group delay characteristics of $H_0(z)$ or $H_1(z)$ used in data embedding, the summed phase differences of $\phi(\omega)$ to the respective phase spectrum of the filter ($H_0(z)$ and $H_1(z)$), $\Delta\Phi_0$ and $\Delta\Phi_1$, are calculated as in Eqs. (8) and (9).

Step 5: The embedded data $\hat{s}(k)$ are detected using Eq.(10).

$$\phi(\omega_m) = \arg\left(\mathrm{FFT}\left[y(n)\right]\right) - \arg\left(\mathrm{FFT}\left[x(n)\right]\right), \tag{7}$$

$$\Delta\Phi_0 = \sum_m \left|\phi(\omega_m) - \arg(H_0(e^{j\omega_m}))\right|, \tag{8}$$

$$\Delta\Phi_1 = \sum_m \left|\phi(\omega_m) - \arg(H_1(e^{j\omega_m}))\right|, \tag{9}$$

$$\hat{s}(k) = \begin{cases} 0, & \Delta\Phi_0 < \Delta\Phi_1 \\ 1, & \text{otherwise} \end{cases} \tag{10}$$

2.5. **Examples.** Figure 7 shows examples of data embedded and detected with the proposed method as described in the previous subsection. Here, a piano recital (No. 59, "K.485") in the RWC music genre database [25] was used. First, 2 s of this data was

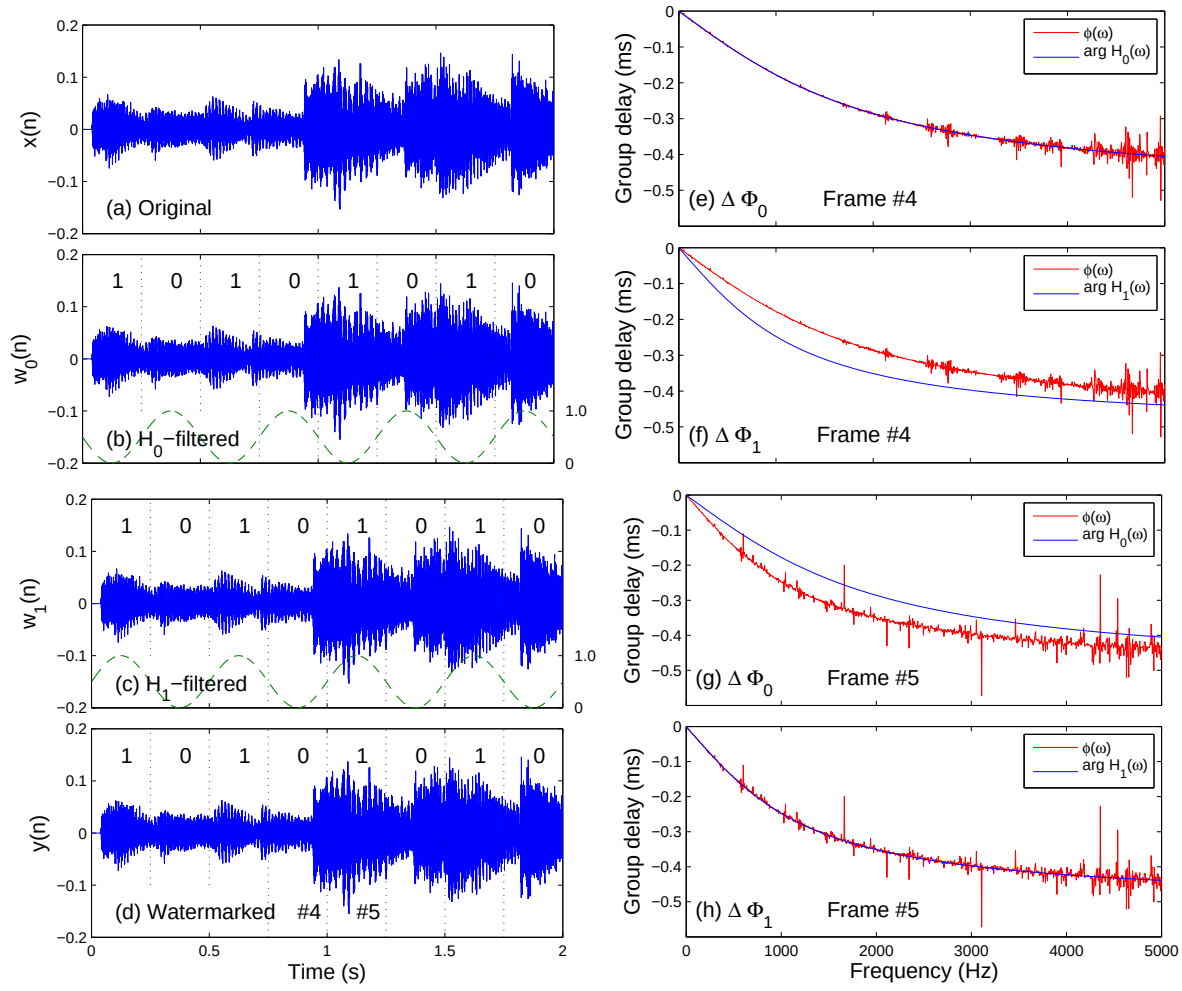

FIGURE 7. Examples of data embedding and data detection with proposed method: (a) original sound $x(n)$, (b) $H_0(z)$-filtered sound $w_0(n)$, (c) $H_1(z)$-filtered sound $w_1(n)$, and (d) watermarked sound $y(n)$. Panels (e) and (f) indicate matches of $\Delta\Phi_0$ and $\Delta\Phi_1$ in Eqs. (8) and (9) at frame #4. Panels (g) and (h) indicate the same results at frame #5.

used as the original sound, $x(n)$, as shown in Figure 7(a). A sampling frequency of 44.1 kHz was used. The conditions for these examples were as follows: digital codes, $s(k)$, "10101010" (corresponding to the first eight bits of "U0") were embedded. The bit-rate, N_{bit}, was set at 4 bps. Thus, the frame length, ΔW, was 250 ms and k indicates the frame number. Two all-pass filters, $H_0(z)$ for "0" and $H_1(z)$ for "1", were applied to obtain $w_0(n)$ and $w_1(n)$, as shown in Figures 7(b) and 7(c). Here, a weighting function was used to avoid discontinuity between the marked segments in the watermarked signal, such as "1" and "0" or "0" and "1". This function is plotted in Figures 7(b) and 7(c) by the dashed curves. Finally, watermarked signal $y(n)$ could be obtained by calculating the weighted sum of these filtered-signals, $w_0(n)$ and $w_1(n)$.

Although there might be no visual difference between $x(n)$ and $y(n)$, embedded information can be encoded as the phase or group-delay information of $y(n)$ in each segment. In data-detection, digital codes $\hat{s}(k)$ can easily be detected by using Eq.(10), as shown in Figures 7(e) and 7(f) or 7(g) and 7(h). In frame #4, $\hat{s}(4)$ can easily be determined to be "0" by using $\Delta\Phi_0 < \Delta\Phi_1$, while $\hat{s}(5)$ can be determined to be "1" by using $\Delta\Phi_0 > \Delta\Phi_1$ in frame #5.

3. **Parameter Settings for Proposed Method.** We proposed an inaudible watermarking scheme in the previous section based on cochlear-delay characteristics. To achieve this scheme in realistic watermarking systems, we have to determine the parameter values in detail, such as the number of cascaded IIR cochlear-delay filters and the filter parameter (b).

3.1. **Determination of number of cascaded IIR filters.** The $b = 0.795$ in Eq.(2) was obtained in the data-embedding process by using the method of optimization to be used as 1/10-cochlear delay. Generally, $H(z)$ can also be represented as

$$H(z) = \prod_{\ell=1}^{L} \frac{-b_\ell + z^{-1}}{1 - b_\ell z^{-1}}, \tag{11}$$

where all b_ℓs are the same as b (e.g., $b = 0.795$). Where $L = 10$, the group delay of this filter can almost be close to human cochlear delay as shown in Figure 2. However, the total delay in the human cochlear here will be twice the actual cochlear delay if the phase information in the signals is manipulated with the filter. It has not yet been confirmed how much the limitations with group delay will affect the requirement for inaudibility, i.e., the number of cascades in Eq.(11). To find the extent of the limitations with the number of cascades, we investigated the audibility threshold for embedding the data in the watermarked signals by using subjective and objective evaluations, in which $b_\ell = 0.795$ was fixed in Eq.(11).

Three stimuli were used in both the subjective and objective evaluations: (i) a pulse train with a period of 2 ms, (ii) a female vocal (No. 39, "Julia") and (iii) a piano recital (No. 59, "K.485") in the RWC music-genre database [25]. A sampling frequency of 44.1 kHz was used. These stimuli were filtered using the cascaded IIR all-pass filter in Eq.(11) as a function of the number of cascades k (from 1 to 20). The time inverse of the filter was used to obtain the in-causal response of the compensated stimulus to enable it to be compared with the compensatory chirp condition.

We carried out a subjective experiment (detection test) to evaluate the extent to which users could perceive the embedded data from the watermarked signals. Four young subjects with normal hearing participated in the experiments. The AXB-method with a two-alternative forced choice (2AFC) was used. Participants were required to determine whether "X" was the closest in sound to A or B. In this experiment, the signal duration was 3 s and the silence between two signals was 500 ms. Twenty trials were done under all conditions.

Figures 8 (a)–(c) show the averaged detection results for the three stimuli. The level of chance was 50%, so that 75% was chosen as the threshold for audibility. All detection rates were under the threshold when the number of cascades L ranged from 1 to 10 (delay units ranged from 0.1 to 2.0 for the enhanced chirp and ranged from -0.1 to -2.0 for the compensatory chirp). These results revealed that none of subjects could perceive the embedded data in the watermarked signals.

We also carried out an objective experiment (simulation) to evaluate the perceptual evaluation of audio quality (PEAQ) [26, 27] and log spectrum distortion (LSD) 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 [26] was used to assess the ODGs of the stimuli. The LSD was also used

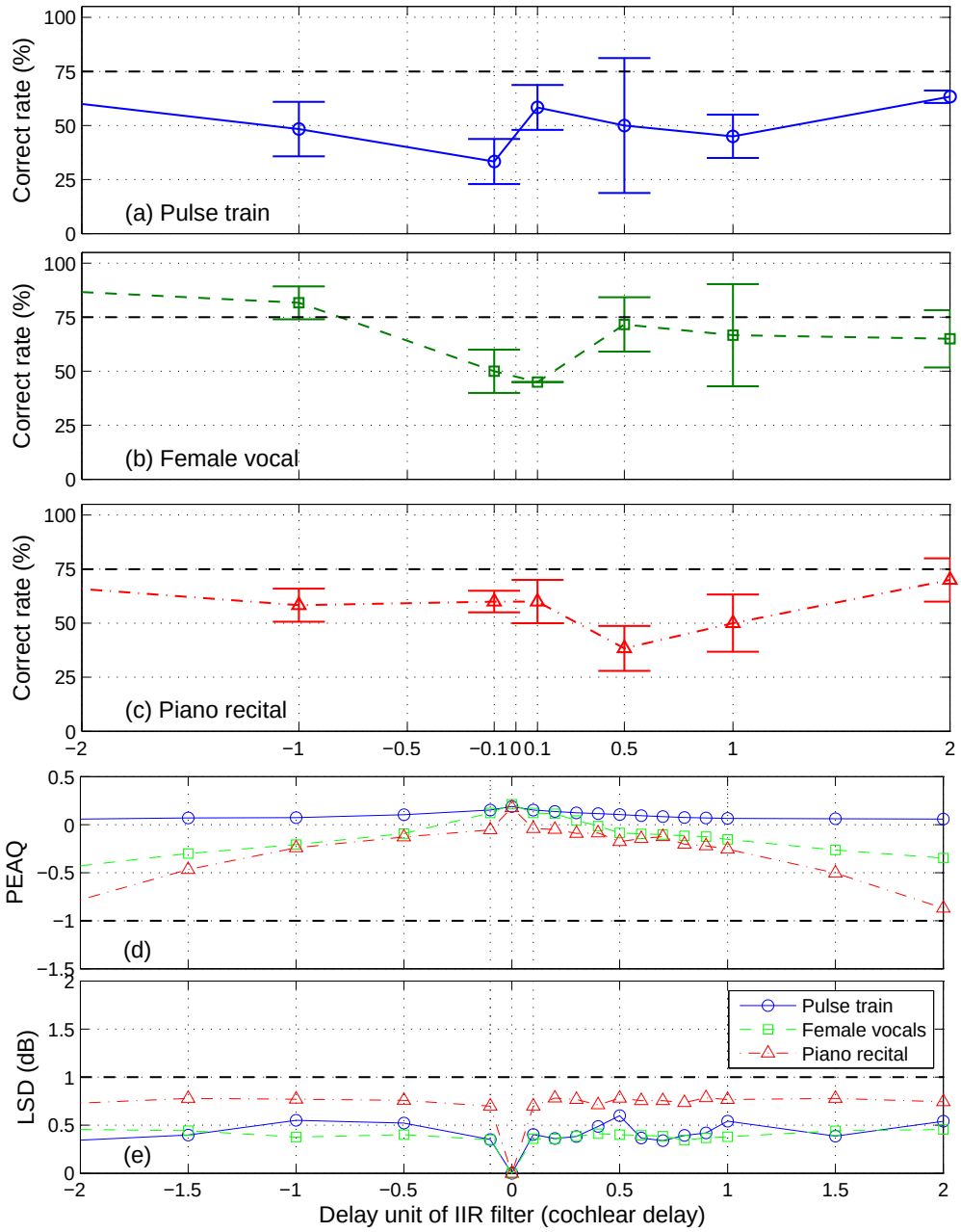

FIGURE 8. Results of subjective and objective evaluations for degree of detection of watermarked signals due to manipulation of cochlear delay. Subjective evaluations for ability to detect (correct rates) (a) pulse train, (b) female vocals, and (c) piano sounds. Objective evaluation of (d) ability to detect perceptual evaluation of audio quality (PEAQ) and (e) relative objective evaluation of log spectrum distortion (LSD) of these three sounds.

to assess distortion in the stimuli by using

$$\mathrm{LSD} = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\frac{|Y(\omega,m)|^2}{|X(\omega,m)|^2}, \qquad \text{(dB)}, \tag{12}$$

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

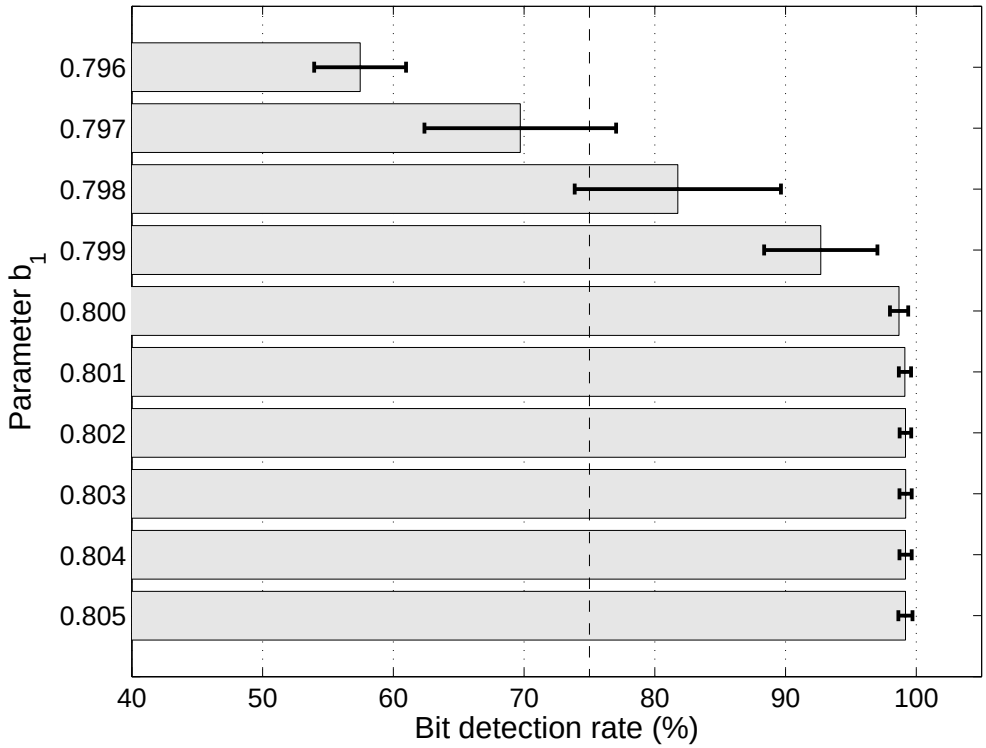

FIGURE 9. Relationship between bit-detection rate and parameter value of b_1, where $b_0 = 0.795$.

Figures 8 (d) and 8(e) show the PEAQ and LSD in decibels for the three stimuli. The numbers of cascades, L, we used in the evaluations, were 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, and 20 (the delay units corresponded to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, and 2.0) for the enhanced chirp condition, and 1, 5, 10, 15, and 20 (delay units corresponded to -0.1, -0.5, -1.0, -1.5, and -2.0) for the compensated chirp condition. All PEAQs for the three stimuli were over a grade of -1 and the LSDs were under 1 dB. Therefore, these results in Figures 8 (a)-(e) revealed that none of the four young subjects with normal hearing could detect the embedded data from the watermarked signals under these conditions. In particular, since the simplest filter should be used to make the method of watermarking inaudible to all users, 1 was chosen for L (only 1/10 cochlear-delay characteristics). Thus, 0.795 was also chosen for b in the research discussed in this paper.

3.2. **Determination of optimal parameter b.** If the values of b_0 and b_1 in the data-detection process used in the data-embedding process are too close to each other, the detection rate may be severely reduced. Therefore, we had to determine these values to ensure $\hat{s}(k)$ could be successfully detected from the watermarked signal $y(n)$ in all cases.

We thus carried out bit-detection tests with regard to the filter parameters, b_0 and b_1. The conditions in these tests were set to $b_1 = 0.796$, 0.797, 0.798, $\cdots$, and 0.805 while b_0 was fixed at 0.795 (optimal value). All the 102 tracks of the RWC music-genre database [25] were used as the original signals in the evaluation. The original track had a 44.1-kHz sampling frequency, was 16 bits, and had two channels (stereo). All these signals were watermarked under the above conditions and these were then tested to detect the embedded data from all the watermarked signals.

The bar chart in Figure 9 shows the detection rate of a bit in the watermarking signal, $\hat{s}(k)$, under the various conditions. The bar lengths and error bars correspond to the

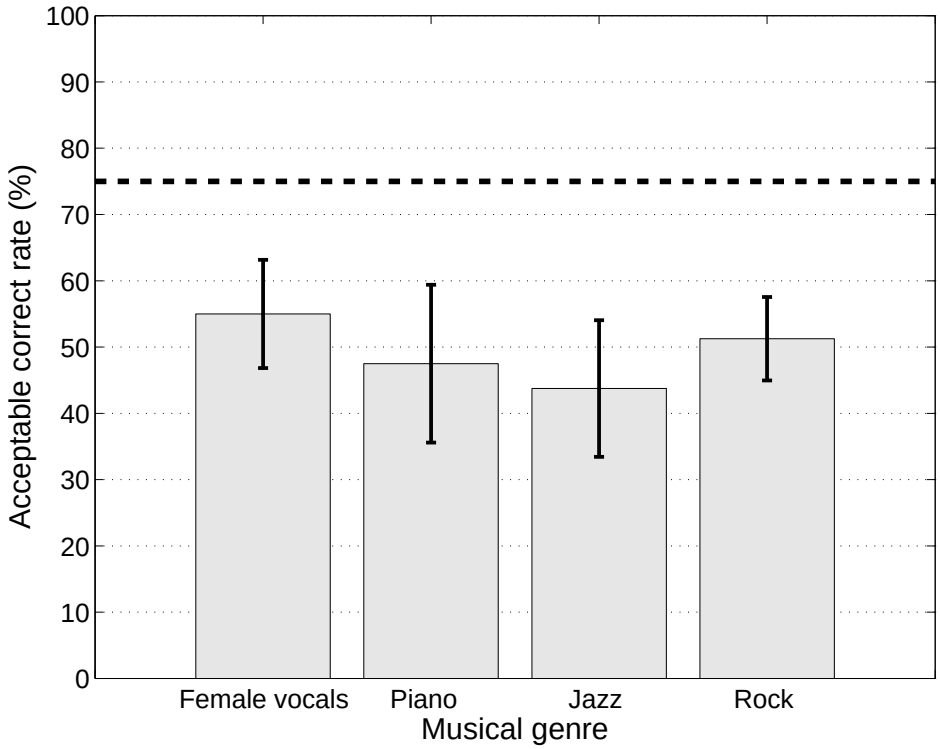

FIGURE 10. Results of subjective evaluations (detection tests) for proposed method. Four stimuli (Female vocals, Piano, Jazz, and Rock sounds) were used. A 2AFC procedure was used in these evaluations so that the chance levels were just 50%. The dashed line at 75% indicates the threshold for an acceptable detection rate for all four stimuli.

mean and standard deviation of the bit-detection rate under each condition. We found from these results that a difference of at least 0.003 between b_1 and b_0 was required for the data-detection process to obtain an acceptable detection rate (over 75%). From the results of preliminary experiments with regard to typical attacks such as those involving resampling, quantization, and data compression, a 0.07-difference was sufficient to detect data with acceptable detection of 75%. Thus, to ensure there was sufficient difference in this study, we used a parameter set of $b_0 = 0.795$ and $b_1 = 0.865$ with a difference of 0.07 rather than 0.003.

4. **Evaluation of Proposed Method.** To confirm the advantages of the proposed method with regard to inaudibility and to demonstrate its usefulness in regard to robustness for a digital-audio watermarking system, we experimentally evaluated it by carrying out subjective and objective detection tests and three objective tests for robustness. All the 102 tracks of the RWC music-genre database [25] were used as the original signals in the evaluations. The original track had a 44.1-kHz sampling frequency, was 16 bits, and had two channels (stereo). The STEP2001 [28] 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 $N_{\mathrm{bit}} = 4$ bps as this critical condition. The same watermarks with eight characters ("AIS-Lab.") were embedded into both the right and left (R-L) channels by using the proposed approach.

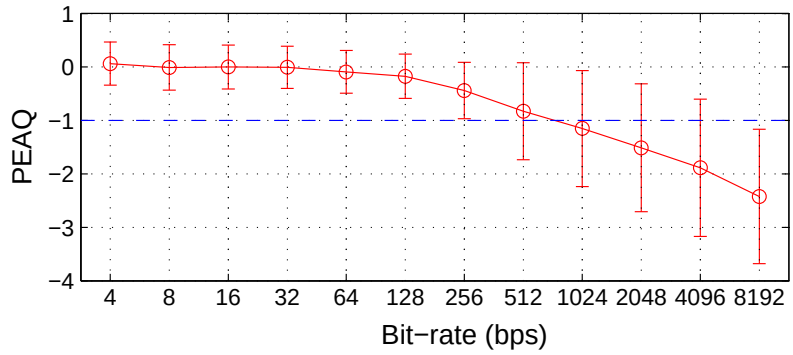

FIGURE 11. Results of perceptual evaluation of audio quality (PEAQ) for proposed method as function of bit-rate N_{bps} (bps).

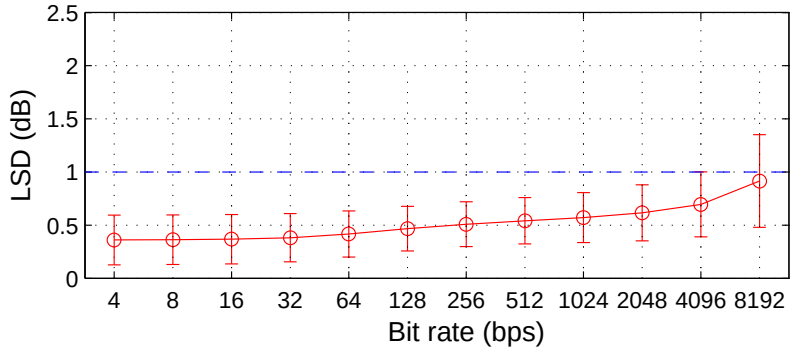

FIGURE 12. Results of objective evaluation (log spectrum distortion: LSD) for proposed method as function of bit-rate N_{bps} (bps).

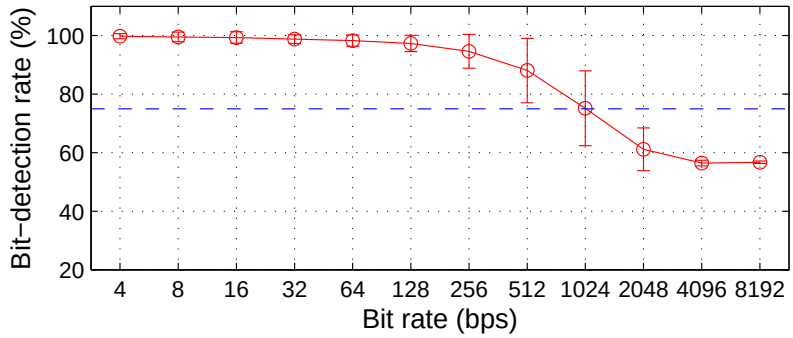

FIGURE 13. Results of bit-detection rate for proposed method as function of bit-rate N_{bps} (bps).

4.1. **Subjective evaluation.** We carried out a subjective experiment (detection test) to evaluate the extent to which users could perceive the embedded data from the watermarked signals. Four young subjects with normal hearing (the same subjects as in Section 4) participated in the experiments. The ABX-method (the ABABX method) with 2AFC was used. The signal duration for all the stimuli (A, B, and X) was 10 s and the silence between two stimuli was 500 ms. Participants were required to determine whether "X" was closest in sound to A or B.

Figure 10 is a bar chart that shows the averaged detection results for the four genres of music (female vocals (No. 39), piano recital (No. 59), jazz melody (No. 29), and rock

track (No. 9) in the database [25]). As the level of chance was 50%, 75% was chosen as the threshold for audibility. All detection rates were under this threshold. These results revealed that none of the four subjects could perceive the embedded data in the watermarked signals. Some demonstrations are available on our Web site [29].

4.2. **Objective evaluation.** We next carried out an objective experiment (PEAQ test) to evaluate the extent to which users could objectively perceive the embedded data from the watermarked signals. The PEAQs for these four stimuli (female vocal, piano, jazz, and rock pieces) in this experiment corresponded to 0.13, 0.005, 0.17, and 0.18. Therefore, these measurements ensured that none of the subjects could perceive the embedded data in the watermarked signals. Thus, based on the results, we evaluated the PEAQs of the embedding for all 102 tracks in the RWC music-genre database [25] as a function of N_{bit}. The N_{bit}s in this experiment were 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, and 8192. A threshold of -1 was chosen to evaluate the limitations in satisfying the inaudibility requirement (a) using the PEAQs in this experiment.

Figure 11 shows the averaged ODGs of the PEAQs we used in Section 4 for the watermarked signals. The circles indicate the averaged ODGs and the error bars indicate the standard deviations for these ODGs. The PEAQs were under the evaluation threshold (> -1) in which the N_{bit}s ranged from 4 to 512 bps while the PEAQs were gradually reduced as the N_{bit}s increased over 128 bps.

LSDs were used to evaluate the distortion in the watermarked signals under the same conditions, to enable it to be compared with the distortion in the PEAQs. Figure 12 shows the averaged LSDs in Eq.(12) for the watermarked signals. The circles and error bars indicate the averaged values and standard deviations of the LSDs. As lower LSDs (under 1 dB) generally indicated that the sound quality of the synthesized signal could be preserved, an evaluation threshold of 1 dB was chosen for the experiment discussed in this paper. We found that the LSDs increased as N_{bit}s increased and that they were under this evaluation threshold under all conditions.

Hence, these results ensured that the proposed method with $N_{\mathrm{bit}}\mathrm{s} \leq 256$ could be used to embed the watermarks into the original signals to satisfy the inaudibility requirement (a) described in Section 1. This means that the limitation with the bit rate in the proposed method was 256 bps in this case.

4.3. **Evaluation of detection.** 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 rates for all signals were evaluated as a function of the bit rate, N_{bit}. 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 less that those of the evaluated signals ($> 75\%$) in which the bit rates, N_{bit}s, ranged from 4 to 1024 bps. This ensured that the proposed method with $N_{\mathrm{bit}} = 1024$ bps could be used to detect the watermarks from the watermarked signals to satisfy the confidentiality requirement (b).

4.4. **Evaluation of robustness.** We carried out three final robustness tests to evaluate how well the proposed method could accurately and robustly detect embedded data from the watermarked audio signals. The same original signals (102 tracks) were used in these tests. Based on suggestions from STEP2001, the main manipulation conditions used were: (i) down sampling (44.1 kHz $\rightarrow$ 20 kHz, 16 kHz, and 8 kHz), (ii) amplitude manipulation (16 bits $\rightarrow$ 24-bit extension and 8-bit compression), and (iii) data compression (mp3: 128 kbps, 96 kbps, and 64-kbps mono). Here, 64-kbps mono indicates 128-kbps compression

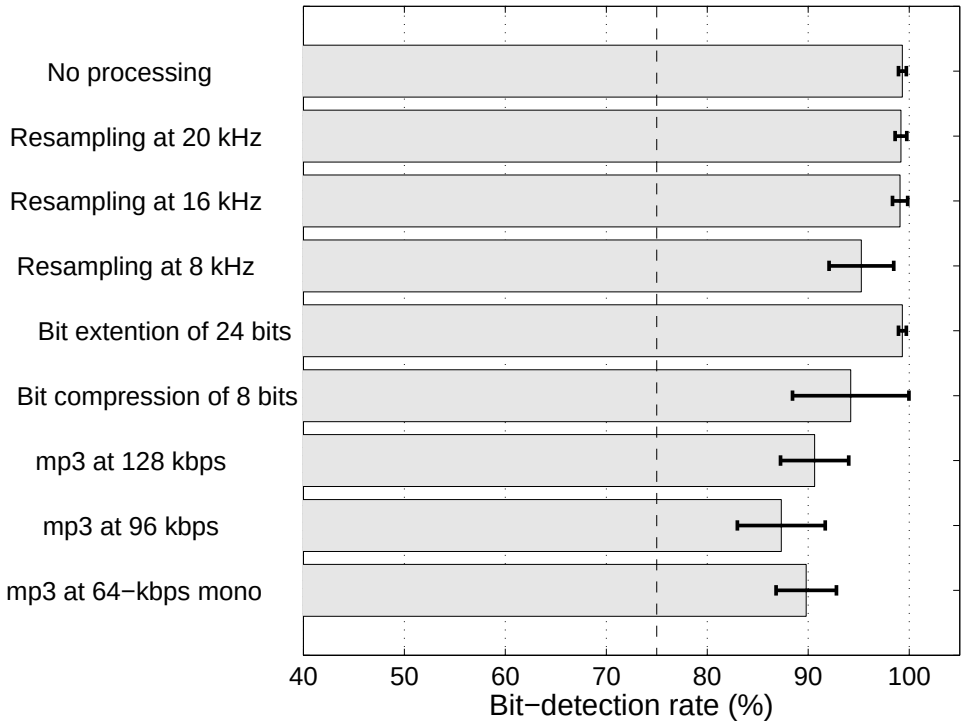

FIGURE 14. Results of objective evaluations (robustness tests) for proposed method. Three robustness-tests (down-sampling, quantization, and data compression) were carried out.

of monaural data converted from stereo data. We reproduced wav files under condition (iii) by converting them from mp3-compressed files to wav files after compression.

Figure 14 shows the results of the evaluations, where the bar lengths and error bars correspond to the means and standard deviations for the bit-detection rates. The bit-detection with the proposed method 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%. Limitations with bit-detection in the robustness tests were simultaneously investigated as a function of N_{bit}. N_{bit}s in which the limitations with bit-detection rate were just over 70% were 1024, 512, 512, 320, 512, 512, 192, 128, and 160 bps corresponding to no processing, resampling of 20 kHz, resampling of 16 kHz, resampling of 8 kHz, bit extension (24 bits), bit compression (8 bits), mp3 (128 kbps), mp3 (96 kbps), and mp3 (64-kbps mono).

Hence, these results indicate that our proposed approach could accurately and robustly watermark copyrighted data in original digital-audio content.

5. **Key Technology, Comparative Evaluations, and Discussion.** As listed in Table 1, we considered that a suitable scheme for watermarking would be processing based on phase information, such as ECHO and PPM. These methods are similar to the processing domain in the proposed approach. Figure 15 has a schematic of the key technology in these watermarking methods.

The echo-hiding approach controls echo-delays (T_0 and T_1) corresponding to digital codes ("0" and "1") in the watermarked signal, $y(n)$, by using an echo-impulse response, as seen in Figure 15(a). This impulse response consists of a direct path (relative amplitude of 1 and no delay) and a 1-st reflection (relative amplitude A and echo delay (T_0 or T_1)).

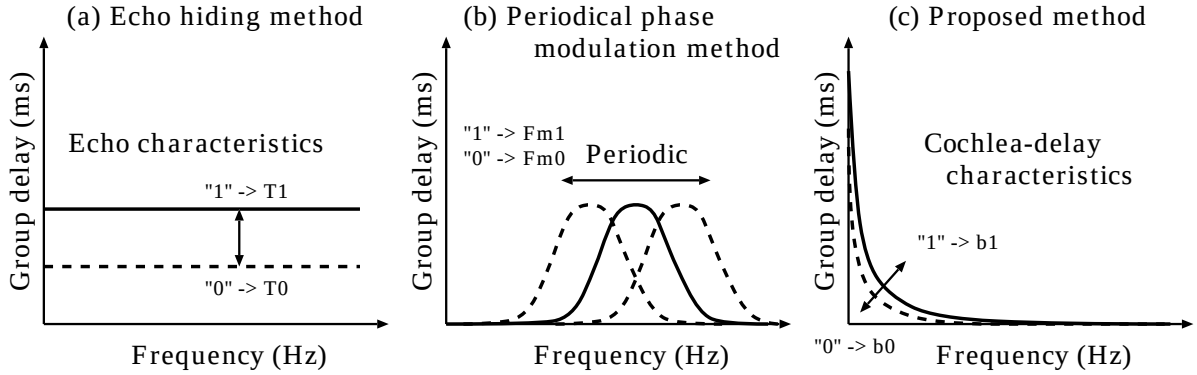

FIGURE 15. Schematic of key technology: (a) echo hiding, (b) periodical phase modulation, and (c) cochlear-delay characteristics.

Since all frequency components were delayed according to echo delay (T_0 or T_1), humans cannot perceive these echoes as different sounds if the delay time is not very long. However, these delays can very easily be detected by using auto-correlation or eliminated by using cepstrum-based processing. Therefore, we found that this technique lacked confidentiality (requirement (b)).

However, the PPM approach periodically controls certain group delays derived from phase modulation around a certain frequency range (from 8 to 20 kHz, determined by Nishimura and Suzuki [17]), as shown in Figure 15(b). Digital codes ("0" and "1") with this technique are embedded as periodic information (modulation frequency F_{m0} and F_{m1} in phase modulation) in the watermarked signal, $y(n)$. Based on their reports of psychoacoustical experiments, humans cannot perceive frequency components processed by these modulations if the modulation frequency is very low ($\leq$ 10 Hz). However, since pulse-like sounds such as the rapid onset of piano sounds have wide frequency components, especially around higher frequencies, this kind of phase modulation disrupts the phase spectra of components at higher frequencies and these modulated components (embedded information) 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)). These two drawbacks motivated us to consider the possibility of inaudible watermarking based on human auditory perception. As introduced in Section 2, cochlear-delay characteristics can be relied on as in Figure 15(c). Digital codes ("0" and "1") are embedded with the proposed method corresponding to the cochlear-delay characteristics (the dashed or solid curve: b_0 for $H_0(z)$ or b_1 for $H_1(z)$) in the watermarked signal, $y(n)$. Although the origin of the idea for inaudible watermarking with the proposed method (cochlear-delay characteristics) is different in PPM, manipulations of group delay are very similar with both methods, as shown in Figures 15(b) and 15(c). Based on the results of psychoacoustical studies [22, 23] and our experiments, humans cannot perceive these delays in the watermarked signal if the delay curve is plotted on the curves of cochlear-delay characteristics. From the results of evaluation in Section 4, we found that the proposed technique satisfied all three requirements of inaudibility, confidentiality, and robustness. These are significant advantages of the new technique.

We next carried out the same four tests (PEAQ, LSD, bit-detection, and robustness) to demonstrate the potential impact of the new approach to comparatively evaluate four typical watermarking methods (LSB, DSS, ECHO, and PPM) with the proposed method. The experimental conditions were the same as those in the evaluations described above. LSB, DSS, ECHO, and PPM were implemented simply by the authors according to the original ones [6, 8].

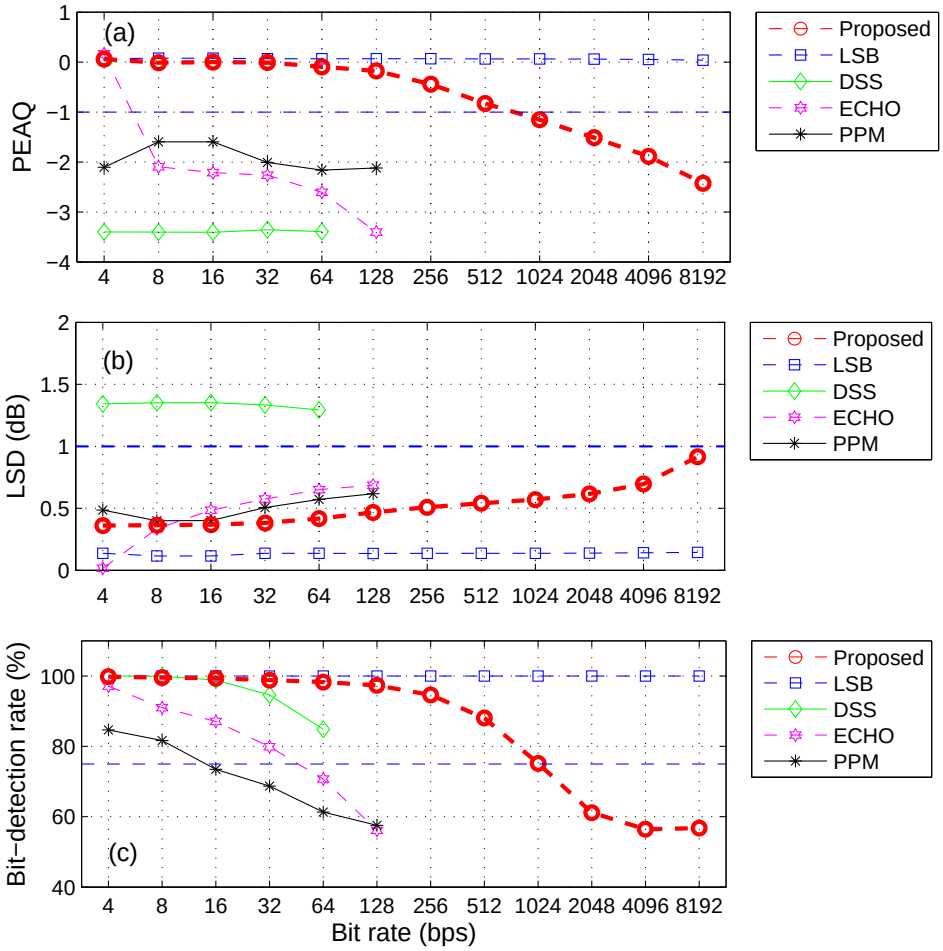

FIGURE 16. Results of comparative evaluations: (a) PEAQ, (b) LSD, and (c) bit-detection rate for the five methods (ours, LSB, DSS, ECHO, and PPM methods) as function of bit-rate N_{bps} (bps).

TABLE 2. Results of objective evaluations (robustness tests) for the five methods. Three robustness tests (down-sampling, quantization, and data compression) were carried out. N_{bit}s were fixed at 4 bps.

Modification	LSB	DSS	ECHO	PPM	Proposed
No processing	100.00	100.00	96.71	84.68	99.32
Resampling at 20 kHz	57.19	99.02	94.25	58.95	99.18
Resampling at 16 kHz	56.76	99.02	93.34	57.10	99.09
Resampling at 8 kHz	54.32	98.33	88.06	53.10	95.26
Bit extension of 24 bits	100.00	99.02	96.71	84.68	99.32
Bit compression of 8 bits	51.00	98.20	85.69	54.65	94.21
mp3 at 128 kbps	50.94	99.02	95.49	58.36	90.63
mp3 at 96 kbps	49.76	99.02	94.51	57.54	87.33
mp3 at 64-kbps mono	50.18	99.02	94.63	57.05	89.80

The tip rate and data rate in DSS were set to 4 and 8192. A carrier frequency of 0 Hz and a key of pseudo-random sequences 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. The relative amplitude of the echoes was set to $A = 0.6$. The F_{m1} and F_{m2} in PPM were set to 8 and 10 Hz. Here,

data detection with LSB, DSS, and ECHO were implemented as blind detection while data detection with PPM was implemented as non-blind detection (phase derivation by using the original signal, as proposed in [17]). Although, the bit-rates N_{bit}s used in these tests generally ranged from 4 to 8192, the N_{bit}s for the tests were restricted by limitations with the split of time-scaled frames in each method, i.e., 64 bps for DSS and 128 bps for ECHO and PPM.

Figure 16 plots the results of comparative evaluations (PEAQ, LSD, and bit-detection rate). Table 2 lists the results of the robustness tests. 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 robustness for watermarking although it could satisfy inaudibility and confidentiality requirements (a) and (b). We also found that DSS and ECHO could satisfy robustness (c), but DSS had a drawback with (a) inaudibility and ECHO with (b) confidentiality. Although we did not have the original code for PPM, it had slight problems with inaudibility with detecting the embedded watermarks. However, these may be able to be resolved if PPM is precisely tuned.

6. **Conclusion.** We proposed a novel method of inaudible digital-audio watermarking based on cochlear-delay characteristics. A subjective evaluation of the method revealed that subjects could not perceive embedded data in watermarked signals or detect any difference in the watermarked signals. Objective evaluations of the proposed approach indicated that the results of subjective evaluation were valid assessed by evaluating the PEAQ as well as the LSDs of the watermarked signals. These evaluations revealed that the limitation with embedded N_{bit} (bps) by using the new method was 128 bps (below the limitations of robustness tests for mp3). This is a sufficient number of bits per second for copyright protection as suggested by STEP2001 [28]. An evaluation of the method's robustness demonstrated that it could precisely and robustly detect embedded data such as those copyrighted with a watermarked audio signal, to protect them against various signal transformations such as resampling, amplitude compression, and data compression. We also comparatively evaluated the proposed method with the four other methods (LSB, DSS, ECHO, and PPM). These results suggest that our proposed approach could provide a useful way of protecting copyright.

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.* [30], (2) extend this method to more robust systems, and (3) reconsider the possibility of extending the limitations of embedding with the method so that it can satisfy all requirements of inaudibility, confidentiality, and robustness. Considerations with regard to a technique that can correct errors in coding and prevent realistic attacks that destroy watermarks in the embedding process of the proposed method can important issues that need to be resolved to develop a realistic digital-watermarking system to protect copyrights.

Acknowledgments. This work was supported by a Grant-in-Aid for Scientific Research (No. 21650035) made available by the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, Information hiding – A survey, *Proc. IEEE Special Issue on Protection of Multimedia Content*, vol.87, no.7, pp.1062-1078, 1999.

[2] C. C. Chang, T. Lu, Y. F. Chang, and C. T. Lee, Reversible data hiding schemes for deoxyribonucleic acid (DNA) medium, *Int. Journal of Innovative Computing, Information and Control*, vol.3, no.5, pp.1145-1160, 2007.

[3] T.-H. Chen, T.-H. Hung, G. Horng, and C.-M. Chang, Multiple watermarking based on visual secret sharing, *Int. International Journal of Innovative Computing, Information and Control*, vol.4, no.11, pp.3005-3026, 2008.

[4] C.-C. Chen and D.-S. Kao, DCT-based zero replacement reversible image watermarking approach, *Int. Journal of Innovative Computing, Information and Control*, vol.4, no.11, pp.3026-3036, 2008.

[5] C.-C. Lo, J.-S. Pan, and B.-Y. Liao, A HOS-based watermark detector, *Int. Journal of Innovative Computing, Information and Control*, vol.5, no.2, pp.293-300, 2009.

[6] N. Cvejic and T. Seppänen, *Digital Audio Watermarking Techniques and Technologies*, Idea Group Inc. (IGI), 2007.

[7] W. Bender, D. Gruhl, and N. Morimoto, Techniques for data hiding, *IBM Systems Journal*, vol.35, nos.3/4, pp.131-336, 1996.

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

[9] W. Iwakiri and K. Matsui, Embedding a text into audio codes under ADPCM quantizer, *J. IPSJ*, vol.38, no.10, pp.2053-2061, 1997.

[10] N. Aoki, A band extension technique for G.711 speech using steganography, *IEICE Trans. Commun.*, vol.E89-B, no.6, pp.1896-1898, 2006.

[11] L. Boney, H. H. Tewfik, and K. N. Hamdy, Digital watermarks for audio signals, *Proc. ICMCS*, pp.473-480, 1996.

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

[13] H. Takahashi, R. Nishimura, and Y. Suzuki, Time-spread echo digital audio watermarking tolerant of pitch shifting, *Acoust. Sci. & Tech.*, vol.26, no.6, pp.530-532, 2005.

[14] A. Nakayama, J. Lu, S. Nakamura, and K. Shikano, Digital watermarks for audio signal based on psychoacoustic masking model, *IEICE*, vol.J83-D-II, no.11, pp.2255-2263, 2000 (in Japanese with English abstract).

[15] I. Muramatsu and K. Arakawa, Digital watermark for audio signals based on octave similarity, *IEICE* vol.J87-A, no.6, pp.787-796, 2004 (in Japanese with English abstract).

[16] A. Nishimura, Audio watermarking based on sinusoidal amplitude modulation, *Proc. IEEE Int. Conf. Acoust. Speech, and Signal Processing*, vol.4, pp.797-800, 2006.

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

[18] 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.

[19] C. J. Plack, *The Sense of Hearing*, Lawrence Erlbaum Association, London, 2005.

[20] 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 (in Japanese with English abstract).

[21] 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.

[22] 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.

[23] E. Aiba, M. Tsuzaki, S. Tanaka, and M. Unoki, Judgment of perceptual synchrony between two pulses and its relation to the cochlear delays, *J. Psychological Research*, vol.50, no.4, pp.204-213, 2008.

[24] E. de Boer, Auditory physics, physical principles in hearing theory I., *Physics Report*, vol.62, pp.87-174, 1980.

[25] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, RWC music database: Music genre database and musical instrument sound database, *Proc. ISMIR 2003*, pp.229-230, 2003.

[26] P. Kabal, An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality, *TSP Lab. Technical Report*, Dept. Electrical & Computer Engineering, McGill University, 2002.

[27] Y. Lin and W. H. Abdulla, Perceptual evaluation of audio watermarking using objective quality measures, *Proc. Int. Conf. Acoust. Speech, and Signal Processing*, pp.1745-1748, 2008.

[28] 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.

[29] http://www.jaist.ac.jp/~unoki/WtrMrk2008-demo.files/frame.htm.
[30] 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.