JAIST Repository
https://dspace.jaist.ac.jp/

Title	A study on audio watermarking method based on the cochlear delay characteristics
Author(s)	Hamada, Daiki; Unoki, Masashi
Citation	Journal of Signal Processing, 12(4): 315-318
Issue Date	2008-07
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/7753
Rights	Copyright (C) 2008 信号処理学会. Daiki Hamada and Masashi Unoki, Journal of Signal Processing, 12(4), 2008, 315-318.
Description	

Japan Advanced Institute of Science and Technology

A study on audio watermarking method based on the cochlear delay characteristics

Daiki Hamada and Masashi Unoki

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

Abstract

We propose a novel audio watermarking method based on the cochlear delay characteristics. This method consists of two processes: data embedding and data detection. The data embedding process was carried out by applying two types of all-pass filters with the cochlear delays, and then selecting the filtered signal from them according to the watermarks. The data detection process involved by 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 three objective tests of robustness (against down sampling, amplitude compression, and mp3-compression). Results showed that the proposed method precisely and robustly detected the inaudible embedded data from the watermarked sounds.

1. Introduction

Almost all music content is now available in the form of digital audio data downloadable from the Internets. It is very useful and convenient for users to be able to manipulate digital content on their PC 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 illegal copying and distribution. Complete copyright protection is required for the industries involved and this is a very important research topic.

The most straightforward techniques for protecting the copyright of digital content are based on encryption methods. These methods are effective, but encryption makes the original sounds inaudible and the strong restrictions for legal users. Another frequently used method is to insert the copyright information in the header of the content. This method allows users to enjoy listening to the content, but the embedded copyright can easily be removed by deleting the header of the digital content. Therefore, these methods do not provide sufficient copyright protection for digital audio data.

Another approach is to use watermarking methods to protect the copyright of the content. Several methods have been proposed. Their aim is to embed digital codes for the copyright in digital audio content that are inaudible to users. Since the embedded data cannot be detected by users, they cannot illegally attack the watermarked data to remove the copyright. To provide a useful and reliable form of copyright protection, watermarking methods must satisfy the following requirements: (a) **inaudibility** (inaudible to humans and no sound distortion caused by embedding), (b) **confidentiality** (secure and accurate detection of embedded data), and (c) **robustness** (not affected when subjected to techniques such as data compression). The first requirement is most important because the watermarking method must not affect the sound quality of the original audio. If the sound quality is degraded, the original content may lose its commercial value.

Typical watermarking methods are based on characteristics of human auditory perception. For example, Gruhl *et al.* [1] proposed a method based on echo hiding and Nishimura [2] proposed a method based on the effect of masking for amplitude modulation (AM). These methods can embed watermarks such as copyright into digital audio signals and detect the embedded data from the watermarked signals. However, they do not completely satisfy the three requirements listed above. In the echo-hiding method [1], the embedded data can be easily detected by the simplest methods. In the AM method [2], the sound quality of the watermarked signal is reduced when the modulation index of the AM is too large.

Nishimura *et al.* proposed a method based on periodical phase modulation [3]. This method embeds inaudible specific data into the digital audio signal using a periodical phase modulation. However, the phase modulation disturbs the phase spectra of components in higher frequencies randomly so that these modulated components may be able to be detected by humans in the watermarked pulse-like sounds. This is because humans can perceive rapid phase-variations related to long and rapid group delays in sounds [4].

In this paper, we propose a novel approach to protecting digital audio content by using an inaudible watermarking method based on the characteristics of cochlear delay.

2. Cochlear delay characteristics and related study

A pulse-like sound is heard as a synchronous sound. However, it is not synchronous on the basilar membrane (BM) motion of the cochlea even if the sound components physically begin synchronously. The reason is as follows. A transient sound wave progresses along the BM in the cochlea (from the basal side to the apical side), passing through the outer and middle ear. Since the mechanical vibrations in 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. Low-frequency components of a pulse-like sound require more time to reach the area of maximum displacement in the BM, near the apical side, while higher frequency components of the sound elicit a maximum closer to the basal side. The time course of a pulse-like sound is, therefore, represented as asynchronous components in the BM. This time course is referred to as "cochlear delay" [5].

Aiba *et al.* [6] investigated whether the cochlear delay significantly affects the perceptual judgment of the synchronization of two sounds or not. They used three types of chirp sounds: (a) a pulse sound (intrinsic cochlear delay), (b) a compensatory delay chirp (i.e., group delay was compensated to be zero in the BM), and (c) an enhanced delay chirp (i.e., group delay was longer than the previous one according to the cochlear delay). Their experimental task was designed to estimate the threshold of judgment; that is, how much time was required for subjects to detect onset asynchrony between sounds. The result showed that the threshold of judgment for signal (b) was highest. The threshold of judgment for signal (c) was almost the same as that for signal (a), which meant that the accuracy of judging the synchrony did not improve even if the cochlear delay was compensated for. Their results suggest that the auditory system cannot distinguish an enhanced delay sound and a non processing sound.

We considered that these characteristics could be used for effectively embedding inaudible watermarks into an original signal. We thus investigated the feasibility of an inaudible embedding method based on enhancing group delays related to the cochlear delays, and proposed an audio watermarking method based on the characteristics of cochlear delay.

3. Proposed watermarking method

3.1. Our concept

The dashed line in Fig. 1 shows the cochlear delay characteristics described by Dau *et al.* [5], where the delay time was scaled by $1/10$. As described above, the delay time in a lower frequency is somewhat longer than that in a higher frequency, especially within the lower frequency range (≤ 5 kHz). If this cochlear delay characteristic can be modeled as a phase characteristic of a digital filter, an audio watermarking method based on the cochlear characteristics could be established by controlling the respective group delays in the filter to the digital copyright data ("1" and "0").

In this paper, we designed the following IIR all-pass filter to model the cochlear delay characteristics:

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

where b is the filter parameters. The IIR all-pass filter is used to control delays in which amplitude spectra are passed

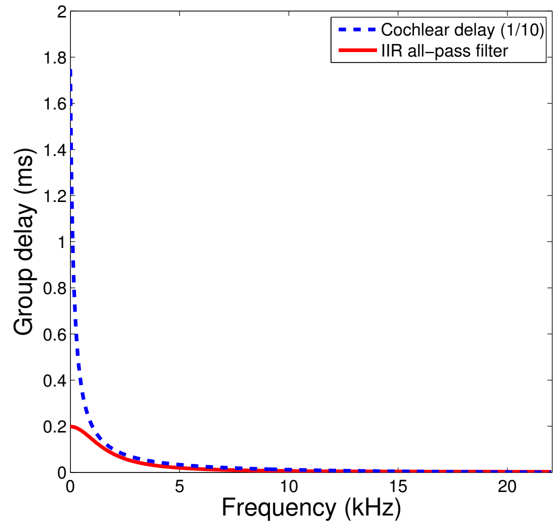

Figure 1: Cochlear delay characteristics.

equally without any loss. Although the higher order IIR filter can be considered to incorporate the cochlear delays, we used the IIR filter in Eq. (1) as the simplest filter.

To determine the optimal value of the filter parameter b in Eq. (1), we fitted the group delay characteristics of $H(z)$ to the cochlear delay characteristics (scaled by $1/10$ shown by the dashed line in Fig. 1) by utilizing the LMS optimization. In the fit, the group delay $\tau(\omega)$ can be obtained as:

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

where $H(e^{j\omega}) = H(z)|_{z=e^{j\omega}}$. The solid line in Fig. 1 shows the approximated cochlear delay, i.e., the group delay characteristics of the IIR all-pass filter in Eq. (2).

We used two filters in Eq. (1), $H_0(z)$ 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, are defined as b for $H_0(z)$ and $H_1(z)$, respectively. With these components, we developed an audio watermarking method based on the cochlear delay characteristics.

Our proposed method consists of two processes: a data embedding process and a data detection process. Below, we describe how these processes were implemented.

3.2. Data embedding process

Figure 2 shows 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 of b to enhance the cochlear delay. These values are determined in Sec. 3.4.

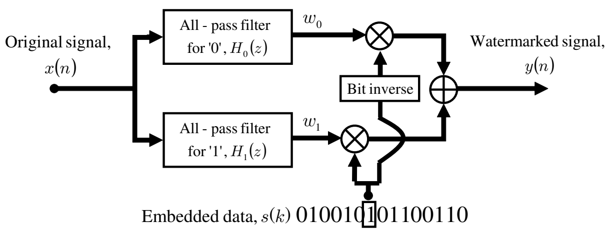

Figure 2: Block diagram of data embedding.

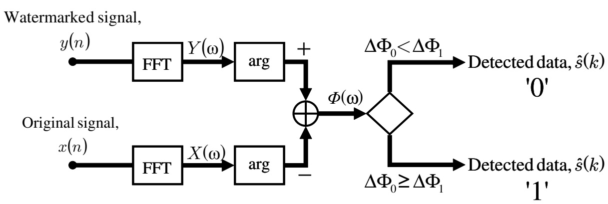

Figure 3: Block diagram of data detection.

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 of these systems (Eqs. (3) and (4)).

Step 3: The embedded data, $s(k)$, was set by the copyright data, e.g., "010010101100110" as shown in Fig. 2.

Step 4: The intermediate, $w_0(n)$ or $w_1(n)$ was selected by stitching the embedded data $s(k)$ ("0" or "1"), and merging them with the watermarked signal, $y(n)$ in Eq. (5).

$$w_0(n) = -b_0x(n) + x(n-1) + b_0w_0(n-1) \quad (3)$$
$$w_1(n) = -b_1x(n) + x(n-1) + b_1w_1(n-1) \quad (4)$$
$$y(n) = \begin{cases} w_0(n), & s(k) = 0 \\ w_1(n), & s(k) = 1 \end{cases} \quad (5)$$

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

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

3.3. Data detection process

Figure 3 shows the data detection process used. Watermarks were detected as follows:

Step 1: We assume that both $x(n)$ and $y(n)$ are available in this 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 the data embedding.

Step 3: The phase difference $\phi(\omega)$ is calculated in each segment, using Eq.(6). $\mathrm{FFT}[\cdot]$ is the fast Fourier transform (FFT).

Step 4: To estimate the group delay characteristics of $H_0(z)$ or $H_1(z)$ used in the 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. (7) and (8).

Step 5: The embedded data $\hat{s}(k)$ is detected using Eq. (9).

$$\phi(\omega_m) = \arg\left(\mathrm{FFT}\left[y(n)\right]\right) - \arg\left(\mathrm{FFT}\left[x(n)\right]\right) \quad (6)$$
$$\Delta\Phi_0 = \sum_m \left|\phi(\omega_m) - \arg(H_0(e^{j\omega_m}))\right| \quad (7)$$
$$\Delta\Phi_1 = \sum_m \left|\phi(\omega_m) - \arg(H_1(e^{j\omega_m}))\right| \quad (8)$$
$$\hat{s}(k) = \begin{cases} 0, & \Delta\Phi_0 < \Delta\Phi_1 \\ 1, & \text{otherwise} \end{cases} \quad (9)$$

3.4. Determination of optimal b

In the data detection process, if the values of b_0 and b_1 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 every case.

We thus carried out bit detection tests with regard to the filter parameters (b_0 and b_1). In these tests, the conditions were set to be $b_1 = 0.796, 0.797, 0.798, \cdots, 0.805$ while b_0 was fixed at 0.795 (optimized value in Fig. 1). Two-strong attack conditions, 8 kHz resampling and 8 bits amplitude compression, were also used.

All of the 102 tracks of the RWC music genre database [7] were used as the original signals in the tests. The original track has a 44.1 kHz sampling frequency, 16 bits and two channels (stereo). STEP2001 [8] suggested that 72 bits per 30 seconds is required to ensure a reasonable bit detection rate in an audio watermarking method. Thus, in this paper, we used 4 bps as this critical condition. The same watermarks with 8 characters ("AIS-Lab.") were embedded into both R-L channels using the proposed method.

Figure 4 shows the detection rate of a bit in the watermarking signal $\hat{s}(k)$ in various conditions. The bar length and error-bar indicate the mean and standard deviation of the bit detection rate at each condition, respectively. From these results, we found that a difference of at least 0.07 between b_1 and b_0 was required for the data detection process to obtain an acceptable detection rate (over 75%). Thus, in this study, to ensure a sufficient difference, we used a parameter set of $b_0 = 0.795$ and $b_1 = 0.865$ with difference of 0.07.

4. Evaluation

We carried out three robustness-tests to evaluate how well the proposed method could accurately and robustly detect embedded data from the watermarked audio signals. The same

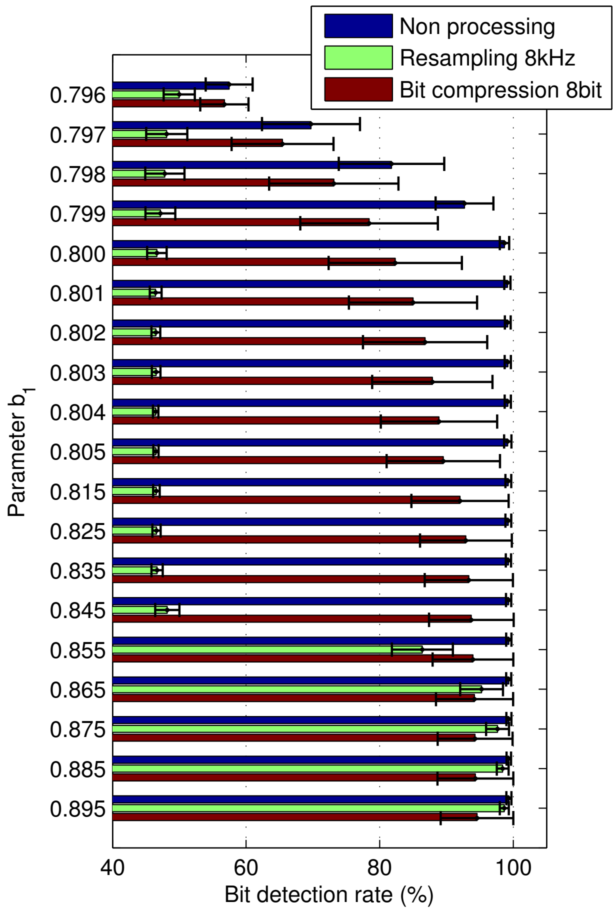

Figure 4: Relationship between bit detection rate and b.

original signals (102 tracks) were used in these tests. Based on suggestions from STEP2001, the main attacking conditions used were: (i) down sampling (44.1 kHz $\rightarrow$ 20 kHz, 16 kHz, and 8 kHz), (ii) amplitude manipulation (16 bits $\rightarrow$ 24 bits extension and 8 bits compression), and (iii) data compression (mp3: 128 kbps, 96 kbps, and 64 kbps-mono). Here, 64 kbps-mono indicates 128 kbps compression of monaural data converted from stereo data. In the condition (iii), we reproduced wav files by converting from mp3-compressed files to wav files, after compression.

Figure 5 shows the results of the evaluation. In this figure, the bar length and error-bar indicate the mean and standard deviation of the bit detection rates, respectively. The bit detection rate of the proposed method was 99.3% in the case of no attack (default case). In contrast, the bit detection rates in the strong-attack conditions (down sampling from 44.1 kHz to 8 kHz, amplitude compression from 16 bits to 8 bits, and data compression of 96 kbps) were 96.7%, 94.1%, and 87.3%, respectively. The results indicate that our proposed method can accurately and robustly watermark copyright data in original digital audio content.

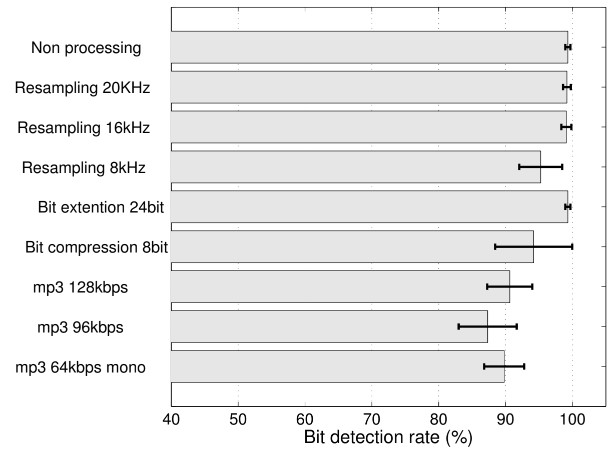

Figure 5: Results of objective evaluation (robustness tests).

5. Conclusion

In this paper, we proposed a novel audio-watermarking method based on the cochlear delay characteristics. We designed optimal group delays for the filters, based on the bit detection tests. An evaluation of the method's robustness showed that it could precisely and robustly detect embedded data such as copyright from watermarked audio signal, against various signal transformations such as resampling, amplitude compression, and data compression. These results suggest that our proposed method could provide a useful way of protecting copyright.

References

[1] Gruhl, D., Lu, A. and Bender, W., "Echo Hiding," *Proc. Information Hiding 1st Workshop*, 295–315, 1996.

[2] Nishimura, A., "Audio Watermarking Based on Sinusoidal Amplitude Modulation," *Proc. ICASSP*, **4**, 797–800, 2006.

[3] Nishimura, R., Suzuki, Y., "Audio watermark based on periodical phase shift," *J. Acoust. Soc. Jpn.*, **60**(5), 269–272, 2004.

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

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

[6] Aiba, E., Tsuzaki, M., Tanaka, S. and Unoki, M., "Judgment of perceptual synchrony between two pulses and its relation to the cochlear delays," *Proc. of the 23rd Annual Meeting of the International Society for Psychophysics*, 211–214, Dec. 2007.

[7] Goto, M., Hashiguchi, H., T., Nishimura, and Oka, R., "RWC Music Database: Music Genre Database and Musical Instrument Sound Database," *Proc. ISMIR 2003*, 229–230, Oct. 2003.

[8] 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.