Title	Study on suitable-architecture of IIR all-pass filter for digital-audio watermarking technique based on cochlear-delay characteristics
Author(s)	KOSUGI, Toshizo; HANIU, Atsushi; MIYAUCHI, Ryota; UNOKI, Masashi; AKAGI, Masato
Citation	2011 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'11): 135-138
Issue Date	2011-03-01
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9977
Rights	This material is posted here with permission of the Research Institute of Signal Processing Japan. Toshizo KOSUGI, Atsushi HANIU, Ryota MIYAUCHI, Masashi UNOKI, and Masato AKAGI, 2011 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'11), 2011, pp.135-138.
Description	

2011 International Workshop on Nonlinear Circuits,
Communication and Signal Processing
NCSP'11, Tianjin, China, March 1-3, 2011

Study on suitable-architecture of IIR all-pass filter for digital-audio watermarking technique based on cochlear-delay characteristics

Toshizo KOSUGI, Atsushi HANIU, Ryota MIYAUCHI, Masashi UNOKI, and Masato AKAGI

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
TEL/FAX:+81-761-51-1391/+81-761-51-1149
Email:{s0910023, a-haniu, ryota, unoki, akagi}@jaist.ac.jp

Abstract

We investigated embedding limitations with our proposed method of audio watermarking. This method was based on the concept of embedding inaudible watermarks into an original sound by controlling its phase characteristics in relation to cochlear delay. We improved the original method by designing a composite architecture for cochlear-delay filters. We evaluated the methods to investigate the embedding limitations by carrying out four objective experiments, i.e., with PEAQ, LSD, bit-detection, and robustness tests. The results indicated the embedding limitation with the composite architecture in the best case was 256 bps, while the embedding limitations with the parallel and cascade architectures were 192 and 128 bps, respectively.

1. Introduction

Digital-audio watermarking has recently been focused on as a state-of-the-art technique enabling copyright to be protected. This has aimed to embed codes to protect the inseparable and inaudible copyright codes separable by users, and to detect embedded codes from watermarked signals [1, 2].

Watermarking methods must satisfy three requirements to provide a useful and reliable form of copyright protection: (a) **inaudibility**, (b) **confidentiality**, and (c) **robustness**. Although several methods (such as LSB [2], DSS [3], ECHO [4], and PPM [5]) have been proposed, these methods have suffered from serious drawbacks in either of the three requirements, especially in (a) inaudibility and (c) robustness due to embedding or reduced security [2].

As the first step toward solving the problems with regard to requirements (a) and (c), a method of audio watermarking based on cochlear delay has been proposed by Unoki & Hamada [6] (a base architecture). Imabeppu *et al.* investigated embedding limitations with their proposed approaches by carrying out four objective experiments [7]. We then improved the proposed method by designing parallel and cascade architectures for cochlear-delay filters [8]. As a result, our proposed architectures made it possible to increase embedding limitations from those with the base architecture.

This paper proposes a composite architecture by reasonably incorporating parallel and cascade architectures to further improve embedding limitations with our proposed approach. We used objective evaluations to systematically investigate and confirm the advantages of the proposed approach.

2. Composite architecture

A cochlear-delay filter is designed as the following 1st-order IIR all-pass filter to model cochlear delay characteristics [6]:

$$H(z) = \frac{-b + z^{-1}}{1 - bz^{-1}}, \qquad 0 < b < 1. \tag{1}$$

An IIR all-pass filter is usually used to control delays in which amplitude spectra are passed equally without any loss. Here, 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 $\tau(\omega)$ is fitted to the cochlear delay (scaled by 1/10 as indicated by the dashed line Fig. 2). Here, this architecture is referred to as a base architecture.

Imabeppu *et al.* improved the previous approach to improve embedding limitations with the method by using a parallel architecture [7]. Based on the expression of N-bits, it is also possible to control $M(= 2^N)$-th cochlear delays using the parallel architecture. However, M-th cochlear delays must not be beyond the cochlear delay, which was scaled by 1/10. We have improved our previous approach to improve embedding limitations with the method by using a cascade architecture [8]. Based on the expression of L-bits, it is also possible to control $R(= 2^L)$-th cochlear delays using the cascade architecture. However, inaudibility is affected increasing the number of R-th cochlear delays. In addition, the value of parameter b must be from 0 to 1. Thus, we propose a composite architecture by reasonably incorporating parallel and cascade architectures. Based on the expression of $N \cdot L$-bits,

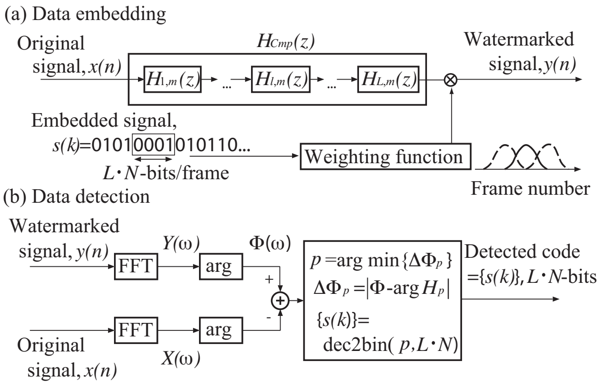

Figure 1: Block diagram of composite architecture.

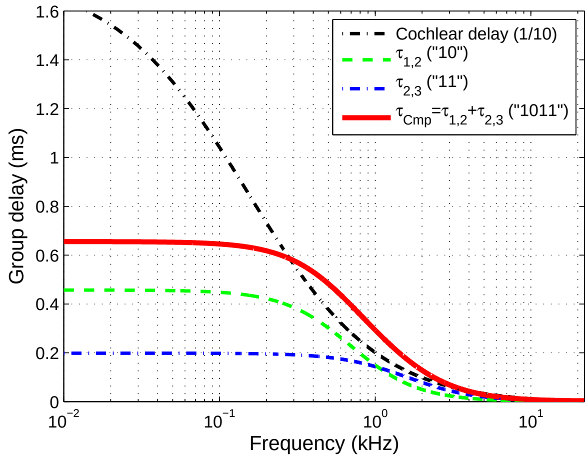

Figure 2: Cochlear-delay and group-delay characteristics of composite architecture ($L = 2$ and $N = 2$).

it is also possible to control $U(= 2^{N \cdot L})$-th cochlear delays using the composite architecture.

2.1. Data embedding process

Figure 1(a) has a block diagram of the data-embedding process. We designed the composite architecture for the cochlear delay filter $H_{\rm Cmp}(z)$ as follows:

$$H_{\rm Cmp}(z) \;:=\; \prod_{\ell=1}^{L} H_{\ell,m}(z) = \prod_{\ell=1}^{L} \frac{-b_{\ell,m} + z^{-1}}{1 - b_{\ell,m} z^{-1}} \quad (3)$$

where $\ell = 1, 2, \cdots, L$ and $m = 0, 1, \cdots, M-1$. Here, the group delay, $\tau_{\rm Cmp}(\omega)$, can be obtained as:

$$\tau_{\rm Cmp}(\omega) \;=\; \sum_{\ell=1}^{L} \tau_{\ell,m}(\omega) \quad (4)$$

$$\tau_{\ell,m}(\omega) \;=\; -\frac{d \arg(H_{\ell,m}(e^{j\omega}))}{d\omega} \quad (5)$$

For example, the group delays in the composite architecture with $N = 2$ and $L = 2$ are represented as 16-types of $\tau_{Cmp}(\omega)$ in Eq. 5. Therefore, the composite architecture can embed 4-bits per frame into the original signal. Figure 2 plots the group-delay characteristics of the cochlear delay filter in the composite architecture.

2.2. Data detection process

Figure 1(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 overlapping 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). (4) The summed phase differences of $\phi(\omega)$ to the respective phase spectrum of the filters, ($\Delta\Phi_p$), are calculated as in Eq. (7) to estimate the group delay characteristics of ($H_{\rm Cmp}(z) = H_{\rm p}(z)$) used for embedding the data. (5) The embedded data, $\hat{s}(k)$, are detected using the p-th cochlear filter.

$$\phi(\omega_q) \;=\; \arg Y(\omega_q) - \arg X(\omega_q) \quad (6)$$

$$\Delta\Phi_p \;=\; \sum_q |\phi(\omega_q) - \arg(H_{\rm p}(e^{j\omega_q}))| \quad (7)$$

3. Evaluations

We evaluated the improved methods (the parallel ($N = 1, 2, 3,$ and 4), the cascaded ($L = 1, 2, 3,$ and 4), and the composite architectures ($L = 2$ and $N = 2$)) by carrying out four objective experiments, i.e., with perceptual evaluation of sound quality (PEAQ) [10], Log spectrum distortion (LSD), bit-detection, and robustness tests to investigate the extent of embedding limitations with the improved methods.

3.1. Objective evaluations

All 102 tracks in the RWC music-genre database [9] were used in these evaluations. The original tracks had a 44.1-kHz sampling frequency, 16-bit quantization, and two-channels (stereo). Here, the unit of fps represents frames per sec. The same watermarks with eight letters (“AIS-Lab.”) were embedded into both channels by using the proposed methods. The frame-rates in these experiments were $4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096,$ and 8192 fps.

We carried out the first objective experiment (PEAQ) to evaluate the sound quality of the watermarked signals. PEAQs were used to output the objective difference grades (ODGs). The ODGs were graded on a five-point scale as 0 (imperceptible), -1 (perceptible but not annoying), -2 (slightly annoying), -3 (annoying), and -4 (very annoying). An evaluation threshold of -1 was chosen to evaluate inaudibility in this experiment. Figures 3(a), 4(a), and 5(a) plot the

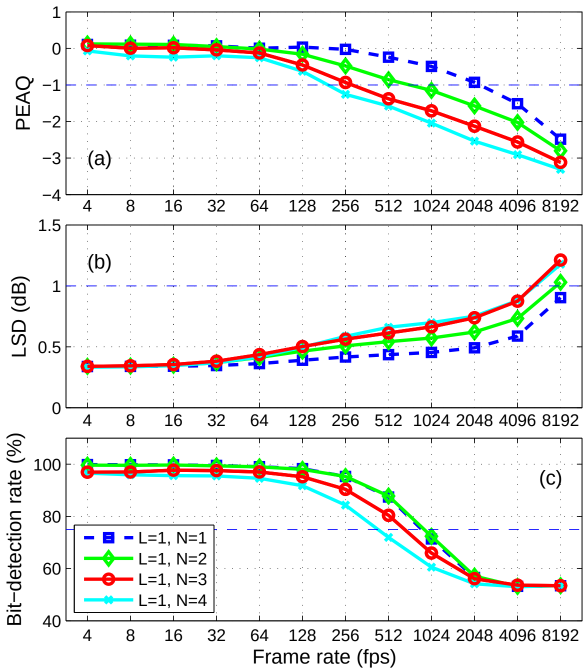

Figure 3: Evaluations of parallel architectures ($N = 1, 2, 3,$ and 4): (a) PEAQ, (b) LSD, and (c) bit-detection rate.

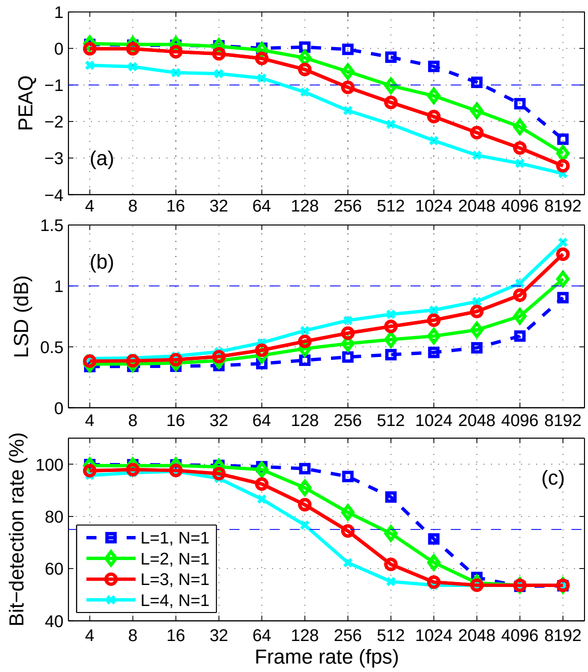

Figure 4: Evaluations of cascade architectures ($L = 1, 2, 3,$ and 4): (a) PEAQ, (b) LSD, and (c) bit-detection rate.

averaged ODGs of the PEAQs. Figures 3(a) and 4(a) show that the ODGs of the PEAQs have decreased with an increase in the number of cochlear-delay filters.The PEAQs in a composite architecture ($L = 2$ and $N = 2$) were under the evaluation threshold (> -1) in which the frame rates ranged from 4 to 64 fps.

We carried out the second objective experiment (LSD measures) to evaluate the sound quality of the watermarked signals. Figures 3(b), 4(b), and 5(b) plot the averaged LSD for the watermarked signals. The LSDs in the cascade architectures ($L = 1, 2, 3,$ and 4) were under the evaluation threshold (< 1 dB) in which the frame rates ranged from 4 to 2048 fps, while the LSDs in the parallel architectures ($N = 1, 2, 3,$ and 4) were under the threshold in which the frame rates ranged from 4 to 4096 fps. The LSDs in a composite architecture ($L = 2$ and $N = 2$) were under the evaluation threshold (< 1 dB) in which the frame rates ranged from 4 to 1024 fps.

We carried out a bit-detection test in the third objective experiment to evaluate how much embedded data could be detected from the watermarked audio signals. An evaluation threshold of 75% was chosen as the embedding limitation to evaluate the bit-detection rate in this experiment. Figures 3(c), 4(c), and 5(c) plot the averaged bit-detection rate for the watermarked signals. The detection rates in the parallel architectures ($N = 1, 2, 3,$ and 4) were over the evaluation threshold ($> 75\%$) in which the frame rates ranged from 4 to 256 fps. The detection rates in the cascade architectures ($L = 1, 2, 3,$ and 4) were over the evaluation threshold ($> 75\%$) in which the frame rates ranged from 4 to 128 fps. The detection rates in the composite architecture ($L = 2$ and $N = 2$) were over the evaluation threshold ($> 75\%$) in which the frame rates ranged from 4 to 256 fps.

The results revealed that the most optimal parallel, cascade, and composite architectures corresponded to (N, L)=$(2, 1)$, $(1, 2)$, and $(2, 2)$, where the maximum detection rate for the composite architecture was 64 fps when it represented 4-bits per frame. Therefore, the embedding limitation with the composite architecture was 256 ($= 64$ fps $\times 4$) bps.

3.2. Evaluation of robustness

We carried out three types of robustness tests in the fourth experiment to evaluate how well the methods could accurately and robustly detect embedded data from the watermarked-audio signals. The manipulation conditions we used were: (i) down sampling (44.1 kHz $\to$ 20, 16, and 8 kHz), (ii) amplitude manipulation (16 bits $\to$ 24-bit extension and 8-bit compression), and (iii) data compression (mp3: 128 kbps, 96 kbps, and 64 kbps-mono).

Table 1 lists the results of evaluations for the base, parallel, cascade, and composite architectures. It summarizes the maximum fps over the evaluation threshold ($> 75\%$) of bit detection. The maximum detection rate with all architectures decreased when the signals were compressed by mp3 with 96 kbps. Here, the maximum detection rates were 32 and 64 fps with the parallel ($N = 4$) and composite architecture ($L = 2$ and $N = 2$), respectively. The bit detection rate with the cascade architecture ($L = 4$) did not exceed the evaluation

Table 1: Results of robustness tests on embedding limitations (frame per sec (fps)).

Modification	Base	Parallel Architecture			Cascade Architecture			Composite Arc.
	$L=1, N=1$	$L=1, N=2$	$L=1, N=3$	$L=1, N=4$	$L=2, N=1$	$L=3, N=1$	$L=4, N=1$	$L=2, N=2$
Non-process	512	512	512	256	256	256	128	256
DS 20 kHz	256	256	256	128	128	128	64	128
DS 16 kHz	256	256	256	128	128	128	64	128
DS 8 kHz	128	128	128	64	128	64	64	64
BC 24 bits	256	256	256	128	128	128	128	128
BC 8 bits	256	256	256	128	128	128	64	64
mp3 128 kbps	128	128	128	64	128	64	32	64
mp3 96 kbps	64	64	**64**	32	**64**	32	—	**64**
mp3 64 kbps	128	128	64	64	64	64	32	64

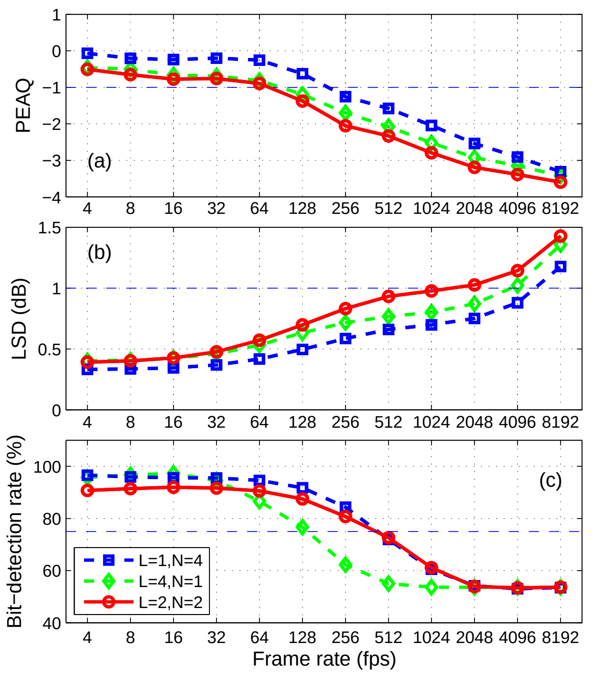

Figure 5: Evaluations of composite architectures ($(L, N) = (1, 4)$, $(4, 1)$, and $(2, 2)$): (a) PEAQ, (b) LSD, and (c) bit-detection rate.

threshold at any frame-rate.

4. Conclusions

We investigated how the proposed approach could be implemented to produce an efficient architecture to further improve embedding limitations with our proposed approach. We carried out objective evaluations and robustness tests on composite architectures including base, parallel, and cascade architectures. The results of objective evaluations revealed that embedding limitations with the parallel ($L = 1$ and $N = 2$) and cascade architecture ($L = 2$ and $N = 2$) were 1024 ($= 512$ fps $\times 2$) bps. The results of robustness tests revealed that embedding limitations with the composite architecture ($L = 2$ and $N = 2$) was 256 ($= 64$ fps $\times 4$) bps. Both results revealed that the composite architecture ($L = 2$ and $N = 2$) was the optimal architecture for the proposed approach.Therefore, the embedding limitations were improved to architecture of the cochlear delay filter (L and N) by optimal choice.

Acknowledgments This work was supported by a Grant-in-Aid for Challenging Exploratory Research (No. 21650035) made available by the JSPS.

References

[1] STEP2001. "News release, Final selection of technology toward the global spread of digital audio watermarks," http://www.jasrac.or.jp/ejhp/release/2001/062 9.html.

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

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

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

[5] Nishimura, R., and Suzuki, Y., "Audio watermark based on periodical phase shift." *J. Acoust. Soc. Jpn.*, Vol. 60, No. 5, pp.269–272, 2004.

[6] Unoki, M., and Hamada, D., "Method of digital–audio watermarking based on cochlear delay characteristics," *International Journal of Innovative Computing Information and Control*, **6**(3), 1325–1346, 2008.

[7] Imabeppu, K., Hamada, D., and Unoki, M., "Embedding limitations with audio-watermarking method based on cochlear delay characteristics," *Proc. IIHMSP2009*, 82–85, 2009.

[8] Unoki, M., Kosugi, T., Haniu, A., and Miyauchi, R., "Design of IIR all-pass filter based on cochlear delay to reduce embedding limitations," *Proc. IIHMSP2010*, 2010.

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

[10] Kabal, P., "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," *TSP Lab. Tech. Rep., Dept. Elec. & Comp. Eng.*, McGil Univ. 2002.