LOG-SCALING WATERMARK DETECTION IN DIGITAL AUDIO WATERMARKING

Byeong-Seob KO, Ryouichi NISHIMURA, Yôiti SUZUKI

RIEC/GSIS, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, JAPAN
E-mail: kobs, ryou, yoh@ais.riec.tohoku.ac.jp

ABSTRACT

A pitch-scaling modification of a watermaked signal is a severe attack on audio watermarking techniques. In particular, watermarking techniques using a pseudo-noise (PN) sequence as a secret key are fragile in the face of this modification because of the desynchronization between the original and embedded PN sequences. We, therefore, propose a log-scaling watermark detection method in this paper to cope with this problem. This method is based on a general logarithm operation in the frequency or quefrency axis in the decoding process. This method is also advantageous in that no other modification of the original embedding process is necessary. The validity of the log-scaling watermark detection method herein proposed is evaluated by employing the time-spread echo method as an example of such a watermarking technique.

1. INTRODUCTION

Digital watermarking is the process to imperceptibly embed secondary data, i.e., watermarks, into digital media contents for the purpose of identification, annotation, and copyright protection. Hence, the embedded watermarks have to survive after exposure to attacks, i.e., manipulations and signal processing, by editing software on a personal computer.

Typical attacks include digital audio compression (such as MP3, AAC, etc.), digital-to-analog and analog-to-digital conversion, sampling frequency conversion, time-scaling, and pitch-scaling for digital audio contents [1]. Of these attacks, pitch-scaling is regarded as one of the severest attacks [2]. In particular, the techniques which use a pseudo-noise (PN) sequence as a secret key in the frequency or quefrency domain, such as those in the previous studies [3, 4, 5] are very vulnerable to pitch-scaling because of the desynchronization between the original and embedded PN sequences.

Some techniques to cope with this problem have been proposed [6, 7]. However, they showed unsatisfactory robustness against pitch-scaling. Moreover, they had some drawbacks such as very low bit rates as well as increased complexity of the embedding and decoding processes. Like others, the time-spread echo method, which has been proposed and deveploed by the authors, is quite vulnerable to even a very small amount of pitch-scaling, although it is sufficiently robust against the other kinds of attacks [8, 9].

To improve the robustness against pitch-scaling, we propose a log-scaling watermark detection method without modifying the original embedding process. This method is based on a logarithm operation to recover synchronization between the orignal and embedded PN sequences. Evaluation of the validity of this log-scaling method was examined by employing it in the decoding process of the time-spread echo method.

2. PITCH-SCALING

The term "pitch-scaling" of an audio signal commonly indicates that the pitch of a signal is scaled by a scale factor κ [10, 11, 12]. This operation is simply formulated as follows:

$$f' = \kappa \cdot f, \tag{1}$$

where κ (> 0) is the scale factor, f is the original frequency, and f' is the κ-scaled frequency. If $\kappa > 1$, the pitch becomes higher and if $\kappa < 1$, it becomes lower. Figure 1 shows an example of pitch-scaling in the frequency domain.

Since cepstrum analysis is carried out by taking the inverse Fourier transform of the logarithm of Eq. 1, the effects of pitch-scaling as shown in Fig. 1 appear in the quefrency domain as follows:

$$\tau' \;=\; \beta \cdot \tau, \tag{2}$$

where β ($= 1/\kappa$) is the pitch-scaling factor in the quefrency domain, τ is the original quefrency, and τ' is the

This study was supported by a Grant-in Aid for Development of Innovative Technologies (Millennium Project, 12107) by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

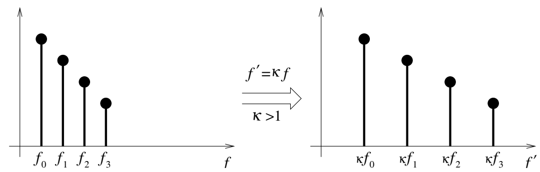


Fig. 1. Example of pitch-scaling in the frequency domain

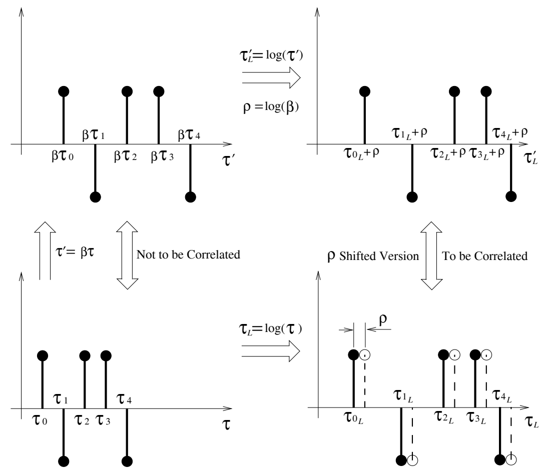


Fig. 2. Schematic explanation of the decoding process based on log-scaling so as to be robust against pitch-scaling

scaled quefrency.

Accordingly, from Eqs. (1) and (2), if the amount of pitch-scaling were known, the pitch-scaled secret key, i.e., a PN sequence, would be simply descaled by $\tau = (1/\beta) \cdot \tau'$. However, the amount is generally unknown and very difficult to estimate without the original signal. Consequently, a method to recover the sychronization is required for blind detection.

3. LOG-SCALING METHOD

A logarithm is widely exploited to convert the operation of multiplication into addition. Similarly, it converts the scaling (multiplying) process into the shifting (adding) process in the logarithm axis. That is, taking the logarithm of Eq. (2), the equation is rewritten as follows:

$$\log(\tau') = \log(\tau) + \log(\beta), \tag{3}$$

where $\log(\cdot)$ is the common logarithm. In Eq. (3), we see that scaling by β is converted into shifting by $\log(\beta)$. This logarithm operation for the frequency or quefrency axis is referred to as "log-scaling" in this paper.

Figure 2 shows a schematic explanation of how log-scaling works in watermark detection from a pitch-scaled signal. If the original version, i.e., a signal without pitch-scaling, and the pitch-scaled version are both converted by log-scaling, the patterns of the two log-scaled versions are brought into accord though the pitch-scaled verion is shifted according to the amount of pitch-scaling. Thus, correlation between the original and pitch-scaled version can be expected regardless of whether there is pitch-scaling or not.

However, the log-scaling in Eq. (3) cannot be directly implemented because discrete-time signal processing is basically used in the embedding and decoding process. That is, assuming that τ is either 1100 or 1101 sample points, for example, $\log(\tau)$ is then about 3.04139 or 3.04178 sample points, respectively, if the logarithm of base 10 is used. These values, therefore, cannot be directly used because the point of each sample must be an integer.

To cope with this problem, the log-scaling in Eq. (3) is firstly scaled by a scale factor γ to obtain discrete sample points as follows:

$$\tau'_L = \lfloor \gamma \cdot \log(\tau') \rfloor, \tag{4}$$

where $\lfloor \cdot \rfloor$ rounds the element to the nearest integer towards minus infinity and γ is a log-scaling factor. Applying Eq. (4) to the preceding examples, 1100 and 1101 sample points of τ are scaled into 3041 and 3041 sample points, respectively, when $\gamma = 1000$, and 30413 and 30417 sample points, respectively, when $\gamma = 10000$. These examples indicate that the larger γ is, the better the log-scaling result. However, the number of samples dramatically increases when a large γ is applied, it might bring about many problems such as long processing time and a huge amount of necessary memory. Hence, γ should be carefully selected. To obtain the complete log-scaled version of the signal, linear interpolation is performed to obtain the samples between the log-scaled samples.

4. DECODING PROCESS BASED ON LOG-SCALING FOR THE TIME-SPREAD ECHO METHOD

In the previous section, the theoretical basis of the log-scaling method was presented to realize robustness against pitch-scaling. In this section, the log-scaling watermark detection method is employed in the decoding process of the time-spread echo method to evaluate the validity in an actual application. The watermarked signal is assumed to be exposed to pitch-scaling.

Figure 3 shows the embedding and decoding process of the time-spread echo method. As shown in Fig. 3, a PN sequence is used in the embedding and

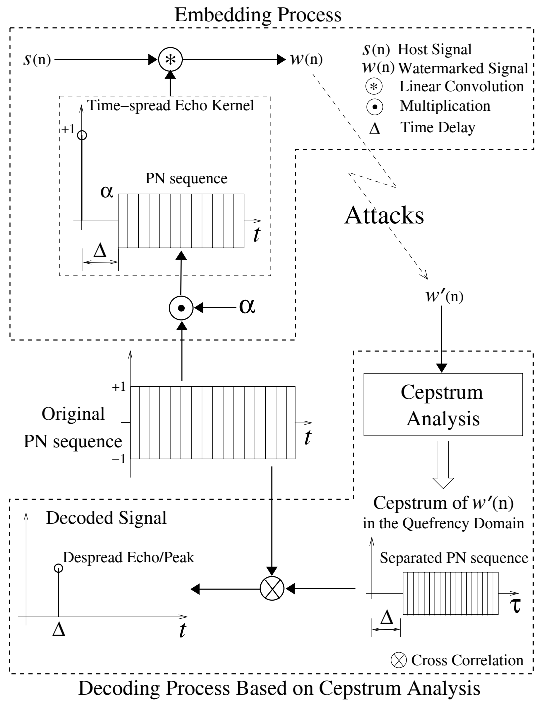


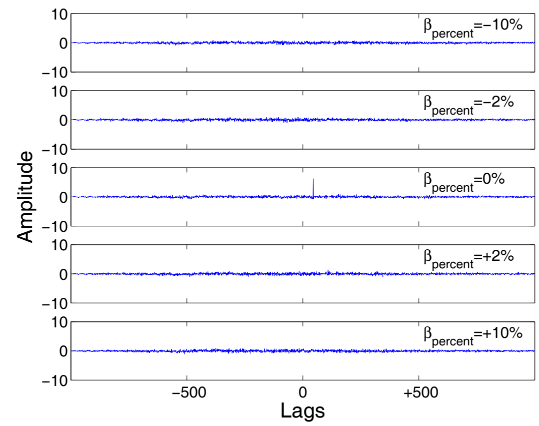
Fig. 3. Embedding and decoding process of the time-spread echo method

decoding processes, and the embedded PN sequence is separated from the watermarked signal by cepstrum analysis in the quefrency domain. In the decoding process of Fig. 3, if the watermarked signal is exposed to pitch-scaling, correlation between the orignal and separated PN sequence cannot be performed because the pattern of the separated PN sequence is changed by the pitch-scaling. Hence, the proposed method can potentially provide a very useful watermark detection method for the time-spread echo method.

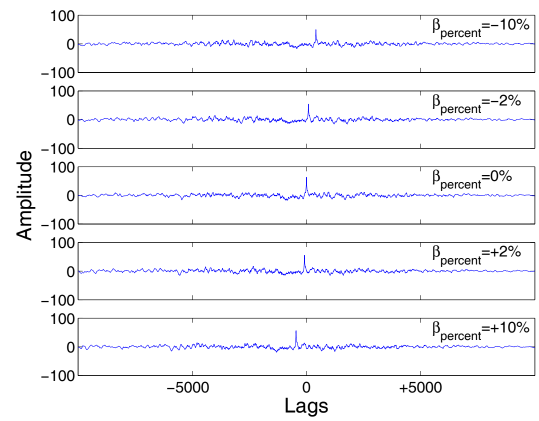
Before employing the log-scaling method, we insert $\Delta - 1$ zeros in front of the original PN sequence to eliminate the effect of the time delay Δ. In this process, note that the delay time Δ is assumed to be known though the amount of pitch-scaling is unknown. Since $\Delta - 1$ zeros are inserted in front of the original PN sequence, the despread echo, i.e., peak, in the decoded signal appears at $t = 0$ when there is no pitch-scaling. On the contrary, if there is pitch-scaling, the peak is shifted according to the amount of pitch-scaling.

Figure 4 shows examples of decoded signals for several amounts of pitch-scaling. In this figure, $\beta_{percent}$ is the percentage of pitch-scaling where $\beta_{percent} = 0\%$ corresponds to $\beta = 1$ in Eq. (2).

Figure 4(a) shows the case where log-scaling in not used and the Fig. 4(b) shows that when log-scaling is used in the decoding process. In Fig. 4(a), a distinct



(a) Not using log-scaling



(b) Using log-scaling

Fig. 4. Examples of decoded signals for several amounts of pitch-scaling: (a) not using log-scaling where $\Delta = 45$ sample points and (b) using log-scaling where $\gamma = 10000$

peak appears at Δ only when $\beta_{percent} = 0\%$. On the contrary, Fig. 4(b) clearly shows that distinct peaks appear regardless of $\beta_{percent}$ and that the location of each peak shifts corresponding to the amount of pitch-scaling.

The amount of peak shifting, $\varepsilon_{\pm}$, is given by the following equation:

$$\varepsilon_{\pm} = \gamma \cdot \log(\beta), \tag{5}$$

where γ is the log-scaling factor and β is the pitch-scaling factor. Here, ε_{+} (< 0) means the case of pitch-scaling towards a higher pitch and ε_{-} (> 0) is for pitch-scaling towards a lower pitch. From Eq. (5), the pitch-scaling factor β is estimated as follows:

$$\beta = 10^{\varepsilon_{\pm}/\gamma}. \tag{6}$$

Figure 5, which is plotted by overlapping the results of

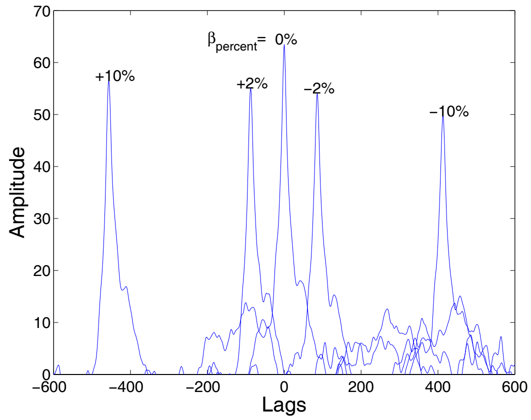


Fig. 5. Actual example of the shifts of the peak according to the amount of pitch-scaling. This figure was plotted by magnifying the results of Fig. 4(b)

Fig. 4(b) with magnification, shows an actual example of the shifts of the peak according to the amount of pitch-scaling. Table 1 shows the actual and calculated amounts of shifting, $\varepsilon_{\pm}$, and the pitch-scaling factor β. The actual $\varepsilon_{\pm}$ was measured from the results of Fig. 5, and the actual β was calculated by substituting the actual $\varepsilon_{\pm}$ into Eq. (6). The calculated $\varepsilon_{\pm}$ was, on the other hand, computed with Eq. (5) based on the given β. Table 1 shows that the actual and calculated values show a good coincidence. These results from Fig. 5 and Table 1 clearly show that the amount of the peak shift exactly corresponds to the amount of pitch-scaling.

Table 1. Actual and calculated $\varepsilon_{\pm}$ and β

$\beta_{percent}$	Actual		Calculated	
	$\varepsilon_{\pm}$	β	$\varepsilon_{\pm}$	β
-10	413	1.10	413.9	1.10
-2	86	1.02	86.0	1.02
0	0	1	0	1
+2	-87	0.98	-87.7	0.98
+10	-457	0.90	-457.5	0.90

5. CONCLUSIONS

In this paper, we have proposed a new digital watermarking method based on log-scaling of quefrency in the decoding process for robust detection against pitch-scaling. Evaluation of the validity of the proposed method was examined by employing it in the decoding process of the time-spread echo method. Results show that the proposed method realizes robust extraction of pitch-scaled watermarks. Moreover, the proposed method is also advantageous in that log-scaling is applicable without any modification of the original embedding process. The concept of the log-scaling proposed in this paper can be extended to the decoding and embedding process of other watermarking techniques to realize robustness against pitch-scaling.

6. REFERENCES

[1] STEP2001, [Online] www.trl.ibm.com/projects/RightsManagement/datahiding/dhstep_e.htm

[2] Michael Arnold, *et al.*, *Techniques and Applications of Digital Watermarking and Content Protection*, Artech House, 2003.

[3] I. J. Cox, *et al.*, "Secure Spread Spectrum Watermarking for Multimedia," *NEC Research Institute, Technical Report*, 95-10.

[4] W. Bender, *et al.*, "Techniques for data hiding," *IBM Systems Journal*, Vol. 35, Nos 3&4, 1996.

[5] Sang-Kwang Lee *et al.*, "Digital Audio Watermarking in the Cepstrum Domain," *IEEE Trans. on Consumer Electronics*, Vol. 46, No. 3, AUG 2000.

[6] R. Tachibana, *et al.*, "An Audio Watermarking Method Using a Two-Dimensional Pseudo-Random Array," *Signal Processing*, 82, pp.1455-1469, 2002.

[7] D. Kirovski, *et al.*, "Robust Convert Communication Over a Public Audio Channel Using Spread Spectrum," *Information Hiding:4th International Workshop, Vol. 2139 of Lecture Notes in Computer Science, Portland, OR: Springer-Verlog*, pp.354-368, APR 2001.

[8] Byeong-Seob KO, *et al.*, "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking," *ICASSP 2002*, Orlando Florida, USA, MAY 13-17, 2002.

[9] Byeong-Seob KO, *et al.*, "Time-Spread Echo Method for Digital Audio Watermarking," *IEEE Trans. on Multimedia*, Accepted.

[10] Björn Carlson, *et al.*, "Real Time Pitch Scaling," Project Group Blue, MAY 28, 2001.

[11] John Garas, *et al.*, "Time/Pitch Scaling Using the Constant-Q Phase Vocoder," Proceedings of STW's 1998 workshops CSSP98 and SAFE98, pp.173-176, 1998.

[12] Mark Kahrs, *et al.*, *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer Academic Publishers, 1998.