RESEARCH OF IMPROVED ECHO DATA HIDING: AUDIO WATERMARKING BASED ON REVERBERATION

Gui-jun Nian [1,2] *, Shu-xun Wang* [1] *, Yun-lu Ge* [1]

(1.Information Dept., Jilin University, Changchun 130025, China; 2.Physics Dept., Jilin University, Changchun 130021, China)

niangj@jlu.edu.cn

Abstract

Reverberation is one of the most heavily used effects in music. A new watermarking method based on reverberation theory is proposed in this paper. The embed process same as the process that adds reverberation to recordings in music studio to make it sound quite natural like in a real concert hall. The watermark are embedded by convoluting the original audio signal and a room impulse response that obtained by image method[4]*. The parameters for calculating the room impulse response are used as secret key that must be used in the decoding process to ensure security. When decoding, homomorphic deconvolution method is used to separate the kernel function from the cepstrum of the watermarked audio signal, and then take cross-correlation with the delay sequence of room impulse response to get watermark by searching the peak.*

Keywords: audio watermarking, reverberation, image method

1. Introduction

Echo hiding embeds data into audio signal by introducing an echo in time domain, which use the masking characteristic of HAS. But the drawback of lenient detection and low detection ratio restricts application. To get better trade-off between robustness and imperceptibility, several useful encoding processes for echo hiding have been proposed. *Xu et al* [1] proposed a multiple echo technique. Instead of embedding one large echo into an audio segment, four smaller echoes with different offsets were chosen to reduce the coloration. *Hyen 0 Oht,et*[2] proposed echo kernel comprises multiple echoes by both positive and negative pulses but with different offsets in the kernel. More transparent response obtained than the conventional positive single or multiple echo methods. Byeong-Seob Ko et al [3] proposed a time-spread echo as an alternative to a single echo or a multi-echo in echo hiding techniques. By spreading an echo using pseudonoise (PN) sequences to make it difficult to decode the embedded data without the PN sequence used in the embedding process. The security is improved. However, the amplitude and time-slot of the time-spread echo kernel is fixed.

Reverberation is different from echo. Echo generally implies a distinct, delayed version of a sound. With reverb, each delayed sound wave arrives in such a short period of time that we do not perceive each reflection as a copy of the original sound. Although a multiple echo [2] or time-spread echo [3] can add a similar effect, there is one very important feature that a simple delay unit will not produce the rate of arriving reflections changes over time, whereas the echo can only simulate reflections with a fixed time interval between them. In reverb, for a short period after the direct sound, there is generally a set of well defined and directional reflections that are directly related to the shape and size of the room, as well as the position of the source and listener in the room.

Add appropriate reverberation to music can make it sound much more natural and majestic like in real concert hall. Based on this concept, a watermarking method use the reverberation theory was proposed in this paper.

In this paper, Section 2 describes the reverberation and room impulse response that used to embed watermark. In section 3, the process of watermark embedding and detection are presented. The experimental result and evaluation are given in Section 4. Finally, conclusions are drawn in Section 5.

2. Reverberation and Room impulse

Reverberation is the result of the many reflections of a sound that occur in a real room. We actually hear reverb every day, and are so accustomed to hearing reverberation without any specifically sense.

Reverberation is probably one of the most heavily used effects in music. It has been developed a variety of ways to synthetically add reverberation to the recordings. Virtually, adding reverberation is to take linear convolution of the room impulse response (IR) and sound source.

There are various schemes for calculating a room impulse response. We use an image method [4] to obtain a room impulse response that determined by the shape, size and surface materials coefficient of the room, as well as the position of the source and listener in the room. It is a one-dimensional discrete function of time that can be described by equation

$$h(n) = a_1 * \delta(n - n_1) + a_2 * \delta(n - n_2) + \cdots + a_L * \delta(n - a_L) \quad (1)$$

where L is the length of room impulse response $h(n)$, a_i represent the magnitude of the i-th reflection sound, n_i

is the delay time of the *i*-th reflection sound. It will allow us to use it as an IR (impulse response) for simulating reverb. Experiments demonstrates that the impulse responses $h(t)$ are obviously dissimilar in the same room parameters but the position of the source and listener is different. We just use the difference to embed watermark.

3. Embedding and detection

3.1 Embedding Process

The principle for the watermark embedding is that according as the music style to select the virtual room parameters such as shape, size, and surface materials to make them resemble to that of the real concert hall which adapt to listen that style music. It would ensure the best artistic effect of the watermarked audio. Locate the sound source position (s_x, s_y , s_z)，select two listener position (l_{x1}, l_{y1} , l_{z1}) and (l_{x2}, l_{y2} , l_{z2})，calculate two room impulse response $h_1(n)$ and $h_2(n)$

use the method in section 2, then use the two room impulse responses to construct two kernel functions $f_1(n)$ and $f_2(n)$.

$$f_1(n) = \delta(n) + h_1(n) \qquad (2.a)$$

$$f_2(n) = \delta(n) + h_2(n) \qquad (2.b)$$

Where δ(n) is the Dirac Delta function. The watermarked signal is obtained by taking a linear convolution of the host audio signal and the kernel function. It is same as the process of adding artificial reverberation to recordings in music studio.

Therefore, the watermarked signal is denoted as

$$x_i^w(n) = x_i(n) * f(n) \qquad (3)$$

Where $x_i^w(n)$ is the watermarked frame, $x_i(n)$ is the host signal, $f(n)$ is the kernel function, it is $f_1(n)$ for embedding a binary "1" and $f_2(n)$ for a "0". The symbol ∗ is a linear convolution.

In fact, the kernel function $f_1(n)$ and $f_2(n)$ are two one-dimensional discrete time sequence determined by the shape, size and surface materials coefficient of the room, as well as the position of the source and listener in the room. So, these parameters are secret key *k*. The watermark couldn't be decoded without it.

3.2 Decoding Process

Fig.1 shows the block diagram of detection process in this paper.

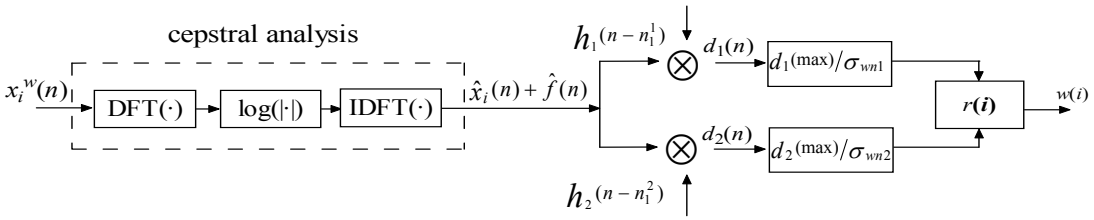


Fig.1. Block diagram of decoding process

Because the data was embedded by taking convolution, the first decoding step is deconvolution. Here, we adopt homomorphic processing with cepstral analysis to separate signal and kernel function. Based on the definition of cepstrum by Bogert[5], two convolution signal becomes summation in cepstrum domain. In other words, the cepstrum of the watermarked signal $\hat{x}_i^w(n)$ can be denoted as

$$\hat{x}_i^w(n) = \hat{x}_i(n) + \hat{f}(n) \qquad (4)$$

where $\hat{x}_i(n)$ and $\hat{f}(n)$ are the cepstrum of $x_i(n)$ and $f(n)$ respectively. The system function $f(n)$ be expanded by using the Dirac delta function $\delta(n)$, then the $\hat{f}(n)$ is represented as

$$\hat{f}(n) = \sum_{l=1}^{L}\sum_{k=1}^{N}(-1)^{k+1}\frac{a_l^k}{k}\delta(n - k * n_l) \qquad (5)$$

where *L* is the length of room impulse response $h(n)$, a_i is the magnitude of each Dirac delta function $\delta(n\text{-}n_i)$.

It is a pulse sequences. In (4), the cepstrum of a watermarked signal is the summation of the real cepstrum of a host signal $\hat{x}_i(n)$ and that of a kernel function $\hat{f}(n)$. In order to detect the embedded data, take cross-correlation between equation (4) and $h_1(n-n_1^1)$, $h_2(n-n_1^2)$ respectively. $h_1(n-n_1^1)$ and $h_2(n-n_1^2)$ are expanded by using the Dirac delta function $\delta(n)$ as follows:

$$h_1(n-n_1^1) = a_1 * \delta(n_1^1) + a_2 * \delta(n_2^1) + \cdots + a_L * \delta(n_L^1) \qquad (6.a)$$

$$h_2(n-n_1^2) = a_1 * \delta(n_1^2) + a_2 * \delta(n_2^2) + \cdots + a_L * \delta(n_L^2) \qquad (6.b)$$

n_i^1 and n_i^2 is the delay time of first pulse in room impulse response $h_1(n)$ and $h_2(n)$ respectively. Consequently, we get two decoding signal $d_1(n)$ and $d_2(n)$

$$d_1(n) = \hat{x}_i^w(n) \otimes h_1(n-n_1^1) = n_{x_1}(n) + [\sum_{l=1}^{L}\sum_{k\in A}\frac{a_l^k}{k}\delta(-n+k*n_l)] \otimes h_1(n-n_1^1)$$

$$+[\sum_{l=1}^{L}\sum_{k\in B}\frac{a_l^k}{k}\delta(n-k*n_l)] \otimes h_1(n-n_1^1) \qquad (7.a)$$

$$d_2(n) = \hat{x}_i^w(n) \otimes h_2(n-n_1^2) = n_{x_2}(n) + [\sum_{l=1}^{L}\sum_{k\in A}\frac{a_l^k}{k}\delta(-n+k*n_l)] \otimes h_2(n-n_1^2)$$

$$+[\sum_{l=1}^{L}\sum_{k\in B}\frac{a_l^k}{k}\delta(n-k*n_l)] \otimes h_2(n-n_1^2) \qquad (7.b)$$

where $\otimes$ is an operator of the linear cross-correlation , $n_{x_1}(n) = \hat{x}_i(n) \otimes h_1(n-n_1^1)$ and $n_{x_2}(n) = \hat{x}_i(n) \otimes h_2(n-n_1^2)$ corresponds to the effect of the host signal as noise. The third term in (7) is employed to detect the embedded watermarks because we are interested in only $n>0$ in the detection process and other two terms have small contributions. By substituting (6) into (7), it should be noted that a distinct peak should be seen in detection signal $d_1(n)$ at n_1 and no distinct peak in $d_2(n)$ if the watermark signal were a binary "1".

Fig.2(a). shows the decoding signal $d_1(n)$ and $d_2(n)$ for one frame. There is a distinct peak in $d_1(n)$ while no distinct peak in $d_2(n)$. σ_{wn1} is the standard deviation of $d_1(n)$ except for the value at n_1 due to the existence of a strong peak. corresponds to the noise by a host signal, $n_{x_1}(n) = \hat{x}_i(n) \otimes h_1(n-n_1^1)$. Smaller σ_{wn} is able to provide a clearer peak corresponds to the embedded information. In order to detect the watermark correctly, we construct the ratio $r(i)$

$$r(i) = \frac{d_1(\max)}{\sigma_{wn1}} \Big/ \frac{d_2(\max)}{\sigma_{wn2}} \qquad (8)$$

where $d_1(\max)$ and $d_2(\max)$ are the maximum of the decoding signal $d_1(n)$ and $d_2(n)$, σ_{wn2} is the standard deviation of $d_2(n)$ except for the biggest value. Thus, the numerator present would be much bigger than the denominator if there were a strong peak in decoded signal $d_1(n)$ while no clear peak in $d_2(n)$, i.e. the watermarked frame were embedded by the kernel function $f_1(n)$. When decoding, we compare the ratio $r(i)$ with a threshold T. A "1" is decoded if the ratio $r(i)>T$. A "0" is decoded if the reverse is true. Fig.2.(b) shows the ratio curve for several watermarked frames. The ratio $r(i)$ of embedded "1" and "0" is not overlapped, thus the watermark can be decoded correctly.

4. Experimental Result and Evaluation

The distance of the two listeners in the image method that produce the room impulse response $dl = \sqrt{(l_{x1}-l_{x2})^2 + (l_{y1}-l_{y2})^2 + (l_{z1}-l_{z2})^2}$ determines the watermarking system imperceptible and robustness performance. To evaluate the relationship between the distance *dl* and the performance, 100 bits watermark were embedded at a rate of 7 bps.

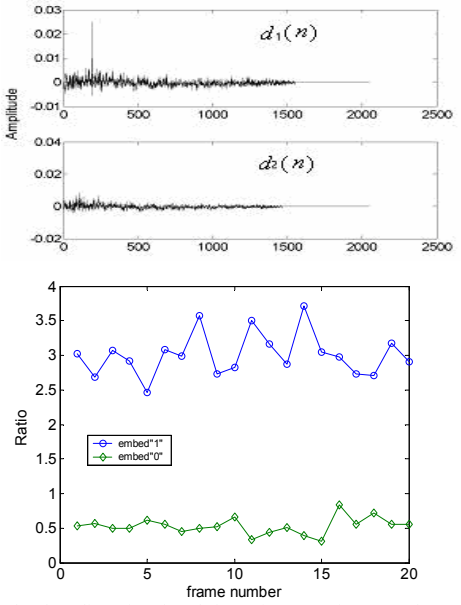


Fig.2. the decoding signal and the ratio curve for several watermarked frames.

Subjective evaluation of watermarked audio was done via ABX testing method [6], 4 different music signals including a violin sonata by Berg, a symphony by Beethoven, a popular song with impulsive sounds in the background and a piano sonata by Beethoven were selected. They are denoted by s01, s02, s03 and s04 respectively. All signals are PCM format (44.1kHz, 16bits/sample, mono). 13 listeners were involved in the experiments. Some of them have some background in music. In Fig3, we see that correct response rate improves as the distance *dl* becomes smaller. If only the distance smaller than 15 the imperceptibility is perfect.

We tested robustness of the proposed method according to the SDMI (Secured Digital Music Initiative) Phase-Ⅱ robustness test procedure [7]. The audio editing and attacking tools adopted in experiment are Cool Edit Pro 2.0, Power MP3 WMA converter v1.14 and Stirmark for Audio v0.2. Experiments results have shown that the algorithm is robust against most attacks including MP3, LP filtering, Noise addition, Echo addition, Resample, Equalization, Pitch-shift, Random cropping, Jittering and TSM(time-scale modification). The error bit rates are less than 4%. The error bit rate is the rates at which the error 100-bit message is detected. In the Experiments the distance *dl* is 8.378 of which its imperceptibility is perfect. Figure 4 is the the ratio curve before and after several representative attacks.

Then, we test the relationship of the distance *dl* and robustness. From the Figure4 (e) we can see that if only the *dl* longer than 5 the robustness is perfect. The error bit rate value is the average error bit rate of several attacks .

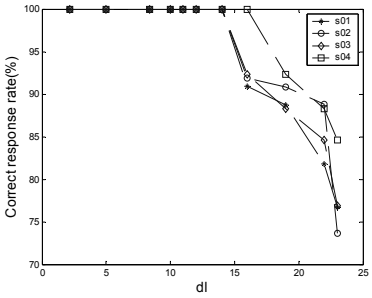


Figure3. Correct response rate according to the distance *dl*

5. Conclusion

In this paper, we proposed a new embedding technique based on reverberation theory for robust and imperceptible audio watermarking. The embedding scheme same as the process that adds reverberation to recordings in music studio, not only to make it more natural sound quality but also get perfect imperceptibility. The image method used to get the room impulse response to embed watermark. The parameters such as virtual room shape, size, surface materials and sound source position used secret key to decode the watermark. The experiments of fidelity and robustness test demonstrated that the proposed method provide better imperceptibility and robustness.

6. References

[1]C. Xu, and ***et al.,*** "Applications of Digital Watermarking Technology in Audio Signals," *J. Audio Eng. Soc.,* Vol. 47,No. 10, 1999 Oct: 165-178.

[2] Hyen O Oh, Jong Won Seok, Jin Woo Hong and Dae Hee Youn. New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking. Proceeding of the IEEEinternational conference on Acoustics, Speech and Signal processing, May 7～11, 2001,3:1341～1344

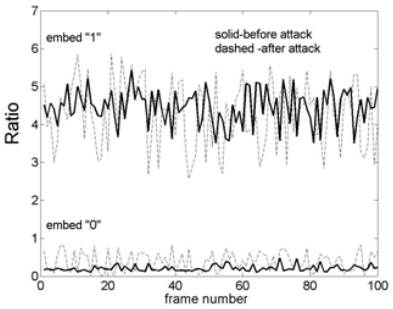
[3] B.-s. Ko, R. Nishimura, and Y. Suzuki, " Time-Spread Echo Method for Digital Audio Watermarking" IEEE Transaction on multimedia,vol.7.No.2,ApRil 2005:212-221.

[4] Jont Allen and David Berkley. Image Method for Efficiently Simulating Small Room Acoustics. Journal of the Acoustic Society of America, April 1979: 912-915.

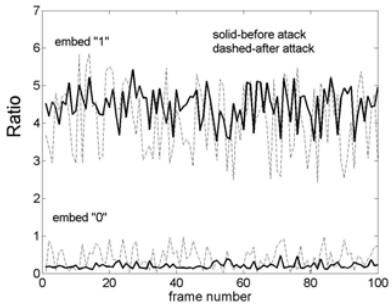
[5] Mouyan Zhou "Deconvolution and Signal Recovery" National defence industry publishing company, 2001:97-101.

[6]B.C.J. Moore, An Introduction to the Psychology of hearing, 4thed. New York: Acdemic,1997.

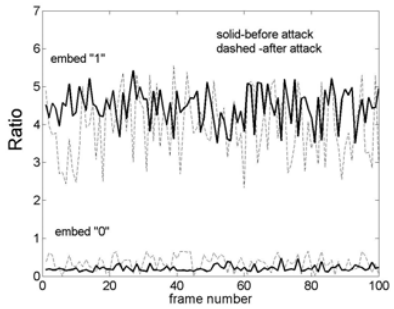
[7] SDMI Phase II Screening Technology Ver 1. (2000). [Online].Available:http://www.sdmi.org/download/frwg00022401-ph2_cfpv1.0.pdf



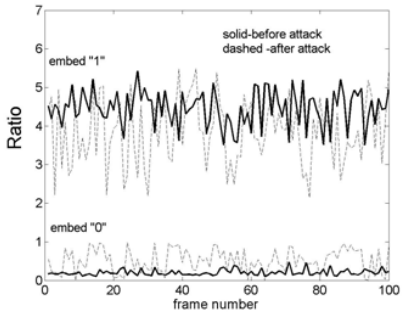
(a) TSM+4%



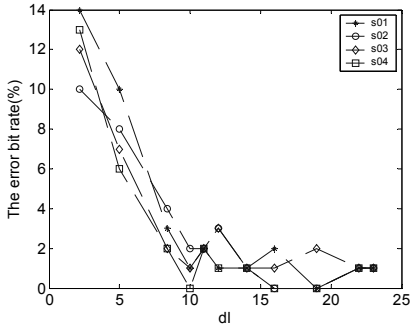
(b) MP3 coding/decoding at 64kbps



(c) resample (44100Hz-22050Hz-44100Hz)



(d) jittering (1/100)



(e) the relationship of the distance *dl* and robustness

Fig 4. the ratio curve before and after several representative attacks