웨이브릿과 마스킹 효과를 이용한 디지털 오디오 워터마킹

황원영[1], 강환일[1], 한승수[1], 김갑일[2], 강환수[3]
명지대학교 정보제어공학과[1], 전기공학과[2], 동양공업전문대학 전자상거래과[3]
전화 : (031) 330-6757 / 팩스 : (031) 335-6478

A Digital Audio Watermark Using Wavelet Transform and Masking Effect

Won Young Hwang[1], Hwan Il Kang[1], Seung Soo Han[1], Kab Il Kim[2], Hwan Soo Kang[3]
[12]Mjongji University, Division of Electrical Information Control Engineering
San 38-2, Namdong, Yongin Kyunggido 449-728, South Korea
[3]Dongyang Technical College, Department of Electronic Commerce 62-160
Gochuk-Dong, Kuro-Ku, South Korea
E-mail : 1978polo@hanmail.net, {hwan, shan, kkl}@mju.ac.kr, hskang@dongyang.ac.kr

Abstract

In this paper, we propose a new digital audio watermarking technique with the wavelet transform. The watermark is embedded by eliminating unnecessary information of audio signal based on human auditory system (HAS). This algorithm is an audio watermarking method, which does not require any original audio information in watermark extraction process. In this paper, the masking effect is used for audio watermarking, that is, post-temporal masking effect. We construct the window with the synchronization signal and we extract the best frame in the window by using the zero-crossing rate (ZCR) and the energy of the audio signal. The watermark may be extracted by using the correlation of the watermark signal and the portion of the frame. Experimental results show good robustness against MPEG1-layer3 compression and other common signal processing manipulations. All the attacks are made after the D/A/D conversion.

I. Introduction

With the rapid spread of computer networks and the further progress of digital contents (audio, image, video, document [1] etc.), security and legal issues of copyright protection have become important. Digital watermark method is one promising technique for effectively protecting the copyright of digital contents. The important properties of the embedded watermark are the quality of the contents having embedded watermark data, the robustness of the watermark against modification of the contents, resistance to intentional removal of or tampering with the watermark, and the reliability of extracted watermark data. In case of audio watermark, the standard work is in process by SDMI (secure digital music initiative). Compared to the video signal, the audio signal can be represented as the small number of samples per time. It represents that there is no enough room for the watermark information, compared to the video signal. Another problem in the audio watermarking is that the HAS (Human Auditory System) is more sensitive than the HVS (Human Visual System). Many researchers have developed the audio watermarking methods. In echo hiding, the echo signal is embedded into the original audio signal [2]. Some algorithms use the patchwork method [3], and the spread spectrum is one of the most general audio watermarking method [4]. The spread spectrum technique is designed to encode a stream of information by spreading the encoded data across as much of the frequency spectrum as possible. This allows the signal survival excellent, even if there are interferences on some frequencies. But it has a fatal weakness in the asynchronous attack. A solution to overcome such a problem is one of the most difficult matters in audio watermarking method. In the algorithm proposed in this paper, the degradation of sound quality can be minimized using some characteristics of HAS and the wavelet transform can be obtained by utilizing zero-crossing rate (ZCR) and the energy of audio signal. Technologies of the audio watermarking have using the masking effect of the human auditory system [5] [6]. The masking effect of the HAS is secure of high quality in technologies of audio watermarking. The robustness of this algorithm was tested by applying several attacks such as clipping, quantization, time stretch, MPEG layer3, WMA [7], VQF [8] compression. All the attacks are made after the D/A/D conversion.

II. Proposed Audio Watermark Algorithm

The algorithm proposed in this paper uses the HAS and the psychoacoustics model [9]. This algorithm embeds the watermark into the audio signal by DWT, ZCR-energy analysis, and the post-temporal masking effect. The watermark embedding region was selected by using the time domain post-masking effect. When extracting the embedded watermark, the first process is the synchronization, which means finding the watermarked region by searching the post-masking region..

II. 1 Back ground

II. 1. 1 Masking effect

The masking effect is related to the limitation of the certain sound according to the noise and distortion. The audio masking is the effect by which the faint but audible sound becomes inaudible in the presence of another louder audible sound. The masking effect is consisted of the temporal masking in the time domain and simultaneous masking in the frequency domain. The simultaneous masking is divided into two, the tone-masking-noise and the noise-masking-tone. The tone-masking-noise is a phenomenon that the noise became masking in center of critical band. The temporal masking refers to both pre- and post- temporal masking. The pre-masking effect make weaker signals inaudible before the stronger masker make turned on and the post-masking effect make weaker signals inaudible after the stronger masker make turned off. In this paper, the watermark is embedded using the post-temporal masking .

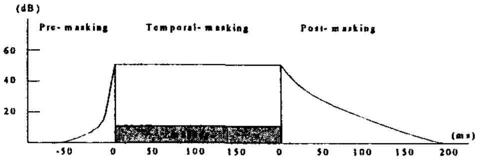

그림 1 순시적 마스킹

Fig. 1 Post-temporal masking

II. 1. 2 Zero-Crossing Rate and Energy analysis

By applying the appropriate size of the window to the audio signal in the time domain, the audio signal is divided into several frames and the ZCR and energy of each frame are obtained. These two factors are used for the start-end point detection of the watermark. In this paper, these two factors are used to search regions to embed the watermark. The frames with the high-energy with the low-ZCR are selected as the post-temporal masking regions. This region is selected as a good point to embed the watermark.

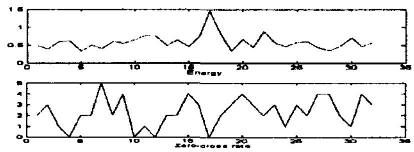

그림 2 에너지와 ZCR 분석

Fig. 2 Energy and ZCR analysis

II. 1. 3 Discrete wavelet transform (DWT)

The Wavelet Transform (WT) is a technique for analyzing signals. It was developed as an alternative to the short time Fourier Transform (STFT) to overcome problems related to its frequency and time resolution properties. More specifically, unlike the STFT that provides uniform time resolution for all frequencies the DWT provides high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies. In that respect, it is similar to the human ear which exhibits similar time-frequency resolution characteristics. The Discrete Wavelet Transform (DWT) is a special case of the WT that provides a compact representation of a signal in time and frequency that can be computed efficiently.

The DWT is defined by the following equation:

$$W(j,k) = \sum_{j} \sum_{k} x(k) 2^{-j/2} \varphi(2^{-j} n - k) \qquad (1)$$

Where $\varphi(t)$ is a time function with finite energy and fast decay called the mother wavelet. The DWT analysis can be performed using a fast, pyramidal algorithm related to multirate filterbanks [10]. As a multirate filterbank the DWT can be viewed as a constant Q filterbank with octave spacing between the centers of the filters. Each subband contains half the samples of the neighboring higher frequency subband. In the pyramidal algorithm the signal is analyzed at different frequency bands with different resolution by decomposing the signal into a coarse approximation and detail information. The coarse approximation is then further decomposed using the same wavelet decomposition step. This is achieved by successive highpass and lowpass filtering of the time domain signal and is defined by the following equations:

$$y_{high}[k] = \sum_n x[n]g[2k-n]$$
$$y_{low}[k] = \sum_n x[n]h[2k-n] \quad (2)$$

Where $y_{high}[k]$, $y_{low}[k]$ are the outputs of the highpass (g) and lowpass (h) filters, respectively after subsampling by 2. Because of the downsampling the number of resulting wavelet coefficients is exactly the same as the number of input points. [11] [12] In this paper, watermark is embedded using DWT.

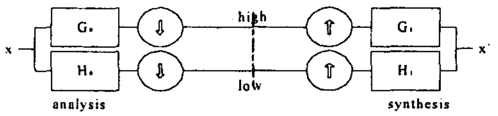

그림 3 1차 DWT의 분해와 합성

Fig. 3 One-level analysis and synthesis system

II. 2 Watermark embedding and extractions

STEP1. The window is divided into the length 1024 bits from the synchronization extraction point. (32 by 32)(Fig.4)

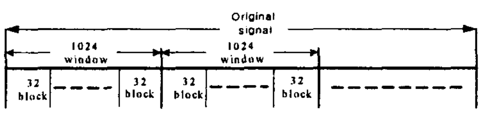

그림 4 워터마크 삽입을 위한 window 선택

Fig. 4 Window select to watermark embed

STEP2. We obtain the energy and ZCR from each frame.

STEP3. We embed the part of watermark in the frame of the high-energy with the low-ZCR. (Co)

STEP4. We embed process of the additive watermark in the wavelet domain by multiplying watermark signal (Ws) with embedding strength.

STEP5. The new signal (Con) is made through inverse DWT.

STEP6. The extract process of watermark is performed from similarity of Con and watermark signal. (Fig.5)

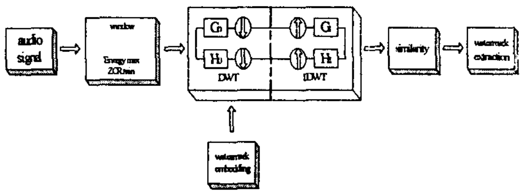

그림 5 워터마크 삽입과 추출 과정

Fig. 5 Watermark embedding and extractions procedure

III. Experimental Results

The experiments were performed according to the requirements of SDMI [13]. The audio signals under test were 16 bits mono signal sampled at 44.1 kHz. All the experiments are performed using the analog signal passed through the speaker but not using the digital signal. The audio samples include Ballade (Audio1), Rock (Audio2), Metal (Audio3), and Classic (Audio4) music. To test the robustness of the propose watermarking algorithm against various types of attacks. All the attacks are made after the D/A/D conversion.

Table 1 shows the results of the no manipulation, clipping, quantization, MPEG1 layer, WMA, VQF compression, time stretch and echo addition attacks. Watermark extraction rata is more than 90%. But with the VQF compression process, the watermark extraction rata is 70%.

표 1 여러 가지 공격에 대한 워터마크 추출율

Table 1 Watermark extraction results for the attacks

	Audio1	Audio2	Audio3	Audio4	Extraction rate
No manipulation	10	10	10	10	100%
Clipping	9	10	9	8	90%
Quantization	9	10	10	9	95%
MPEG1 layer 3	9	10	9	8	90%
Time stretch (+10%)	8	10	10	8	90%
Time stretch (-10%)	8	9	9	10	90%
WMA	9	10	10	9	95%
VQF	5	9	8	6	70%
Echo addition (3%)	9	10	9	10	95%

Table 2 shows the results of the performance comparison between the proposed watermarking algorithm and the spread spectrum method, which is proposed by Cox. Et. Al. Spread spectrum method spreads data in frequency domain using by the discrete cosine transform (DCT). An obvious weakness appears in spread spectrum method with the asynchronous attacks (quantization, MPEG1 layer, WMA and VQF). Comparatively, the proposed algorithm show robust results in asynchronous attacks.

표 2 제안한 알고리즘과 스프레드 스펙트럼 기법을 비교
Table 2 Proposed algorithm vs. spread spectrum method

	proposed algorithm extraction rata	Spread spectrum method rata
No manipulation	100%	100%
Clipping	90%	90%
Quantization	95%	15%
MPEG1 layer 3	90%	15%
Time stretch (+10%)	90%	90%
Time stretch (-10%)	90%	90%
WMA	95%	20%
VQF	70%	10%
Echo addition (3%)	95%	90%

Ⅵ. Conclusions

In this paper, we proposed a new HAS and DWT based algorithm for the audio watermarking. This algorithm utilizes the ZCR and the energy analysis, and the post-temporal masking effect for embedding the watermark. The watermark may be detected by using the correlation of the watermark signal and the portion of the frame. By applying several attacks proposed by SDMI, the robustness of the proposed algorithm is tested, and the experiments shows good performance in the synchronization and the other audio signal manipulation with minimal audio quality degradation. Also, the proposed algorithm is shown to make good performance in D/A/D conversion.

Acknowledgements

This work was supported in part by the Myongji University in the framework of the project during the sabbatical year of the second author.

Reference

[1] Hwan Il Kang, Jong Uk Choi, Kab Il Kim, "A vector watermarking using the generalized square mask", Information Technology: Coding and Computing, 2001. Proceedings. International Conference on, Apr 2001 pp.234 -236

[2] Say Wei Foo, Theng Hee Yeo, and Dong Yan Huang "An adaptive audio watermarking system,"Electrical and Electronic Technology, 2001. TENCON. Proceedings of IEEE Region 10 International Conference, vol.2, (2001) 509-513

[3] H. Kii J. Onishi, and S. Ozawa, "The digital watermarking method by using both patchwork and DCT", Multimedia Computing and System, 1999 IEEE International Conference on, vol.1 (1999) 895-899

[4] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Transaction on Image Processing, vol. 6, pp.1673-1687, Dec. 1997.

[5] Yong-Hun Kim, Seung-Soo Han, Hwan-Il Kang, Kab-Il Kim 'A Digital Watermarking Using Two Masking Effects'Third IEEE Pacific Rim Conference on Multimedia, -PCM2002. Hsinchu, Taiwan, Dec. 2002. pp.655-662

[6] Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik, Laurence Boney 'Robust audio watermarking using perceptual masking'Signal Processing 66 (1998) 337-355

[7] http://www.vqf.com

[8] http://www.microsoft.com

[9] ISO/IEC IS 11172, Information technology - coding of moving pictures and associated audio for digital storage up to about 1.5Mbits/s

[10] S.G Mallat "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation" IEEE. Transactions on Pattern Analysis and Machine Intelligence, Vol.11, 1989, 674-693

[11] I.Daubechies "Orthonormal Bases of Compactly Supported Wavelets" Communications on Pure and Applied Math. Vol.41 1988, 909-996

[12] Tzanetakis, G., Essl, G., Cook, P.R., "Audio Analysis using the Discrete Wavelet Transform," In Proc. WSES Int. Conf. Acoustics and Music: Theory and Applications (AMTA 2001) Skiathos, Greece, 2001.

[13] http://www.julienstern.org