韓國컴퓨터情報學會 論文誌
第15卷 第2號, 2010. 2.

2010-15-2-3-3

Robust Audio Watermarking in Frequency Domain for Copyright Protection

Pranab Kumar Dhar*, Jong-Myon Kim**

저작권 보호를 위한 주파수 영역에서의 강인한 오디오 워터마킹

프라납 쿠마르 다르*, 김 종 면**

Abstract

Digital watermarking has drawn extensive attention for protecting digital contents from unauthorized copying. This paper proposes a new watermarking scheme in frequency domain for copyright protection of digital audio. In our proposed watermarking system, the original audio is segmented into non-overlapping frames. Watermarks are then embedded into the selected prominent peaks in the magnitude spectrum of each frame. Watermarks are extracted by performing the inverse operation of watermark embedding process. Simulation results indicate that the proposed scheme is robust against various kinds of attacks such as noise addition, cropping, resampling, re-quantization, MP3 compression, and low pass filtering. Our proposed watermarking system outperforms Cox's method in terms of imperceptibility, while keeping comparable robustness with the Cox's method. Our proposed system achieves SNR (signal-to-noise ratio) values ranging from 20 dB to 28 dB. This is in contrast to Cox's method which achieves SNR values ranging from only 14 dB to 23 dB.

요 약

디지털 워터마킹은 불법 복제로부터 디지털 콘텐츠를 보호하기 위해 광범위하게 주목을 받아왔다. 본 논문은 디지털 오디오의 저작권 보호를 위해 주파수 영역에서의 새로운 워터마킹 구조를 제안한다. 제안하는 워터마킹 시스템에서는 디지털 오디오가 중첩되지 않는 프레임들로 분리된다. 분리된 각 프레임의 크기 대역에서 선택된 최고치에 워터마크가 삽입된다. 모의실험 결과, 제안하는 방법은 노이즈 추가, 잘라내기, 재배열, 양자화, MP3 압축, 저역통과 필터 등과 같은 공격에서 강인성을 보인다. 제안한 방법의 이러한 결과는 잘 알려진 Cox방법과 비교하여 유사한 강인성을 보이지만, SNR 측면에서는 Cox방법보다 우수한 성능을 보였다. 제안한 방법은 20dB에서 28dB의 SNR을 보인반면, Cox방법은 단지 14dB에서 23dB의 성능을 보였다.

▸ Keyword : 저작권보호(Copyright Protection), 디지털 워터마킹(Digital Watermarking), 사운드 콘텐츠 (Sound Content)

• 제1저자 : 프라납 쿠마르 다르 교신저자 : 김종면
• 투고일 : 2009. 10. 12, 심사일 : 2009. 11. 26, 게재확정일 : 2010. 02. 22.
* 울산대학교 컴퓨터정보통신공학부 석사과정 **울산대학교 컴퓨터정보통신공학부 교수
※ 이 논문은 2009년도 울산대학교 교내 연구비 지원으로 연구되었음

Ⅰ. Introduction

Recent years have seen a rapid growth in the availability of multimedia content in digital form. A major problem faced by content providers and owners is protection of their material. Digital audio watermarking has drawn extensive attention for copyright protection of audio data [1][2]. A digital audio watermarking is a process of embedding watermarks into audio signal to show authenticity and ownership. Audio watermarking should meet the following requirements :(a) Imperceptibility: the digital watermark should not affect the quality of original audio signal after it is watermarked; (b) Robustness: the embedded watermark data should not be removed or eliminated by unauthorized distributors using common signal processing operations and attacks; (c) Capacity: capacity refers to the numbers of bits that can be embedded into the audio signal within a unit of time; (d) Security: security implies that the watermark can only be detectable by the authorized person. All these requirements are often contradictory with each other. However, it should satisfy the important properties such as imperceptibility and robustness.

In this paper, we propose a new watermarking scheme in frequency domain for audio copyright protection. The watermarks are embedded into the selected prominent peaks of the magnitude spectrum of each non-overlapping frame. Experimental results indicate that the proposed watermarking system provides similar robustness as Cox's method [3] against several kinds of attacks such as noise addition, cropping, resampling, re-quantization, MP3 compression, and low pass filtering. However, our proposed watermarking system outperforms Cox's method in terms of imperceptibility. It achieves SNR values ranging from 20 dB to 28 dB. This is in contrast to Cox's method which achieves SNR values ranging from only 14 dB to 23 dB.

The rest of this paper is organized as follows. Section II provides a brief description of related research including Cox's method. Section III introduces our proposed watermarking system including watermark embedding process and watermark detection process. Section IV compares the performance of our proposed scheme with Cox's method in terms of imperceptibility and robustness. Section V concludes this paper.

Ⅱ. Related Research

A significant number of watermarking techniques have been reported in recent years in order to create robust and imperceptible audio watermarks. Lie et al. [4] propose a method of embedding watermarks into audio signals in the time domain. The proposed algorithm exploits differential average-of-absolute-amplitude relations within each group of audio samples to represent one-bit information. It also utilizes the low-frequency amplitude modification technique to scale the amplitudes in selected sections of samples so that the time domain waveform envelope can be almost preserved. In [5], authors propose a blind audio watermarking system which embeds watermarks into audio signal in time domain. The strength of the audio signal modifications is limited by the necessity to produce an output signal for watermark detection. The watermark signal is generated using a key, and watermark insertion depends on the amplitude and frequency of audio signal that minimizes the audibility of the watermark signal. Ling et al. [6] introduce a watermarking scheme based on nonuniform discrete Fourier transform (NDFT), in which the frequency points of embedding watermark are selected by the secret key. Zeng et al. [7] describe a blind watermarking system which embed watermarks into DCT coefficients by utilizing quantization index modulation technique. In [8], the authors propose a watermarking system which embed synchronization signals in time domain to resist against several attacks. Pooyan et at. [9] introduce a audio watermarking system which embed watermarks in wavelet domain. The watermarked data is then encrypted and combined with a synchronization code and embedded into low frequency coefficients of the sound in wavelet domain. The magnitude of quantization step and embedding strength is adaptively determined according to the characteristics of human auditory system. Wang et al. [10] proposes a blind audio watermarking scheme using adaptive quantization against synchronization attack. In addition, the multiresolution characteristics of discrete wavelet transform (DWT) and the energy compression characteristics of discrete cosine transform (DCT) are combined in this scheme to improve the transparency of digital watermark. Watermark is then embedded into low frequency components by using adaptive quantization according to human auditory system. In [11], authors propose a watermarking system in cepstrum domain in which a pseudo-random sequence is used as a watermark. The watermark is then weighted in the cepstrum domain according to the distribution of cepstral coefficients and the frequency masking characteristics of human auditory system. Liu et al. [12] propose a blind watermarking system which takes the advantages of the attack-invariant feature of the cepstrum domain and the error-correction capability of BCH code to increase the robustness as well as imperceptibility of audio watermarking.

In Cox's method [3] watermarks are embedded into the higher m DCT coefficient of the whole sound excluding the DC component by the following formula:

$$v_i' = v_i(1+\alpha x_i) \quad (1)$$

where, m is the length of the watermark sequence, vi is a magnitude coefficient into which a watermark is embedded, xi is a watermark to be inserted into vi, α is a scaling factor, and vi′ is an adjusted magnitude coefficient. The watermark sequence is extracted by performing the inverse operation of (1) represented by the following formula:

$$x_i^* = (\frac{v_i^*}{v_i} - 1)/\alpha \quad (2)$$

Cox's method provides good results in terms of robustness. However, this method cannot achieve good imperceptability in terms of SNR because it embeds watermark into higher DCT components of the sound which sometimes effects the quality of the sound. To overcome this problem, we propose a new watermarking system which embed watermarks into the selected prominent peaks of the magnitude spectrum of each non-overlapping frame. This provides better results than Cox's method in SNR aspect for watermarked audio signals, while keeping comparable robustness with Cox's method against several kinds of attacks.

III. Proposed Watermarking System

A watermark must not be embedded into insignificant region of the audio signal because many common signal processing attacks affect these components. For example, if a watermark is embedded in the high frequency spectrum of an audio signal, low-pass filtering can easily eliminate the watermark. Thus, it is very important to find the significant regions and embed a watermark into those regions without distortion of the original audio signal.

In this section, we give an overview of our basic watermarking scheme which consists of watermark embedded process and watermark detection process. In this implementation, a watermark consists of a sequence of real numbers X= {x1, x2, x3,..., xn}. We create a watermark where each value of xi is chosen independently according to N(0,1) where N(μ, σ2) denotes a normal distribution with mean μ and variance σ2.

3.1 Watermark embedding process

The proposed watermark embedding process is shown in Figure 1. The embedding process is implemented in the following seven steps:

1. The original audio is first segmented into non-overlapping frames.
2. Calculate the magnitude and phase spectrum of each frame using fast Fourier transform (FFT).
3. Find the n most prominent peaks **V**= $\{v_1, v_2, v_3,..., v_n\}$ from magnitude spectrum using a peak detection algorithm.
4. Place watermarks into the n prominent peaks of magnitude spectrum to obtain watermarked peaks **V′**= $\{v_1', v_2', v_3',..., v_n'\}$. This ensures that the watermark is located at the most significant perceptual components of the audio. When we insert the watermark X into V to obtain V′, we specify a scaling parameter α, which determines the extent to which X alters V, shown in the equation (1) [3]:
5. Insert back the n modified peaks into the magnitude spectrum of each non-overlapping frame.

6. Take an inverse FFT of the complex spectrum to calculate the watermarked frame.
7. Finally concatenates all watermarked frames to calculate the watermarked audio signal.

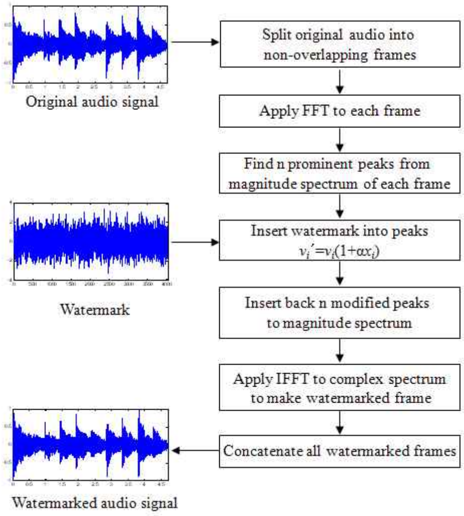

Fig 1. Watermark embedding process
그림 1. 워터마크 임베딩 프로세스

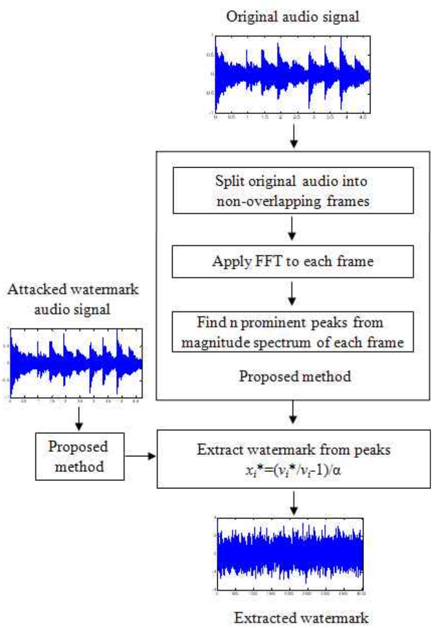

Fig 2. Watermark detection process
그림 2. 워터마크 검출 프로세스

3.2 Watermark detection process

The proposed watermark detection process is shown in Figure 2. The detection process is implemented in the following three steps:

1. Calculate the FFT of the attacked watermark audio frame.
2. Extract n peaks from the magnitude spectrum which are located at the same position in the embedding process above.
3. Extract the watermark sequence by performing the inverse operation of (1) represented by the equation (2).

IV. Simulation Results

In this section, we evaluate the performance of our watermarking system for four different types of 16 bit mono audio signals sampled at 44.1 kHz: (a) 'Let it be' written by Beatles; (b) 'the beginning of the Symphony No. 5 in c minor, Op. 67', written by Ludwig van Beethoven; (c) an instrumental song 'Hey Jude' played with a Korean traditional musical instrument called Gayageum; (d) a human voice providing a direction of TOEIC listening test. Each audio file contains 206,336 samples (duration 4.679 sec). By considering frame size of 512 samples, we have 403 frames for each audio. From each frame we detect 10 peaks to embed watermark. Thus, the length of the watermark sequence is 10×403=4,030.

In order to evaluate the performance of the proposed watermarking scheme, the correlation coefficient between the original watermark X and the extracted watermark X* is calculated by the following similarity SIM(X, X*) formula:

$$SIM(X, X^*) = \frac{X \cdot X^*}{\sqrt{X^* \cdot X^*}} \qquad (3)$$

It is highly unlikely that X* will be identical to X. To decide whether X and X* match, we determine whether the SIM(X, X*) > T, where T is a detection threshold. In this study, the selected detection threshold (T) value is 6 [3].

Figure 3 shows a qualitative evaluation of the original audio with a watermarked audio in which the watermarks are imperceptible using the proposed system.

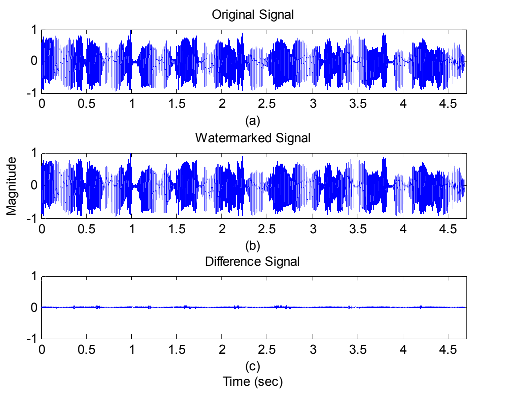

Fig 3. Imperceptibility of the watermarked audio using the proposed method: (a) original human voice sound, (b) watermarked human voice sound (c) difference between original and watermarked human voice sound
그림 3. 제안한 방법을 사용하여 워터마크가 삽입된 오디오의 불변성: (a) 오리지널 인간 음성 사운드, (b) 워터마크가 삽입된 인간 음성 사운드, (c) 오리지널 음성과 워터마크가 삽입된 음성의 차이

In order to evaluate the quality of watermarked signal, the following signal-to-noise ratio (SNR) equation is used:

$$SNR = 10\log_{10}\frac{\sum_{n=1}^{N} S^2(n)}{\sum_{n=1}^{N}\left[S(n) - S^*(n)\right]^2} \quad (4)$$

where S(n) and S*(n) are original audio signal and watermarked audio signal respectively. After embedding watermark, the SNR of all selected audio signals using the proposed method are above 20 dB which ensures the imperceptibility of our proposed scheme.

Figure 4 shows the peak detection of first frame of human voice sound. In our proposed system, watermarks are embedded into the selected prominent peaks of the magnitude spectrum of each frame which provides high robustness against different kinds of attacks as well as good SNR values for watermarked audio signals.

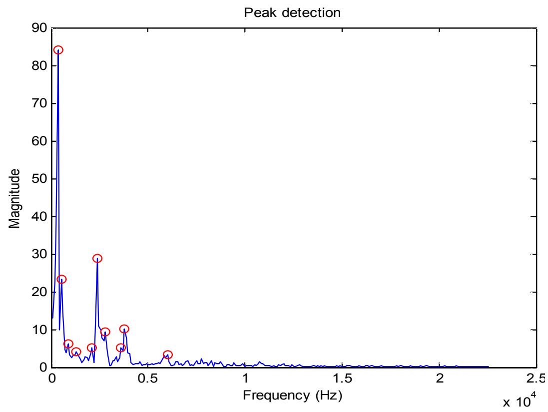

Fig 4. Peak detection in frequency spectrum
그림 4. 주파수 크기 대역에서 최고점 검출

In Cox's method, on the other hand, watermarks are embedded into the higher DCT coefficients of the whole sound excluding the DC component. Figure 5 shows the higher 4030 DCT coefficients of the human voice sound in which watermarks are embedded.

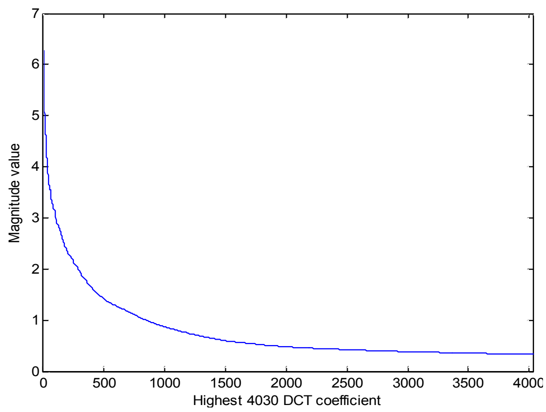

Fig 5. DCT coefficients for human voice sound
그림 5. 인간 음성 사운드에서 DCT 계수

Table 1 shows the SNR comparison between the proposed watermarking scheme and Cox's method for different values of α. Our proposed scheme achieves SNR values ranging from 20 dB to 28 dB for different watermarked sounds. This is in contrast to Cox's method which achieves SNR values ranging from only 14 to 23. In other words, our proposed watermarking system provides 6 dB higher SNR values than Cox's method for different watermarked sounds. Thus, our proposed watermarking system outperforms Cox's method in terms of imperceptibility.

Table 1. SNR comparison between the proposed and Cox's methods
표 1. 제안한 방법과 Cox방법의 SNR 비교

Types of Signal	Cox	Proposed	Cox	Proposed	Cox	Proposed
	α=0.1		α=0.2		α=0.3	
Let it be	23.561	25.654	17.586	23.176	14.067	20.202
Symphony No 5	23.772	27.395	17.611	23.652	14.089	20.649
Hey Jude	23.741	27.218	17.668	23.843	14.146	20.776
Human Voice	23.786	28.259	17.968	24.163	14.446	20.779

4.1 Imperceptibility Test

Informal listening using head set reveals that the watermark embedding into the original audio using the proposed watermarking system does not affect the quality of the sound, which ensures imperceptibility of the embedded watermark.

Table 2. Watermark detection results of the proposed scheme and Cox's method without attacks
표 2. 공격이 없는 경우 제안한 방법과 Cox방법의 워터마크 검출 결과

Types of signals	SIM	
	Proposed	Cox
Let it be	64.6345	64.9933
Symphony No 5	64.4371	64.9933
Hey Jude	64.5267	64.9932
Human Voice	64.2483	64.9933

4.2 Robustness Test

Table 2 shows the performance comparison in terms of similarity between the proposed scheme and Cox's method when no attack is applied to four different types of watermarked audio signals. The proposed method is comparable with Cox's method in terms of similarity.

Figures 6 and 7 show the response of the watermark detector to 100 randomly generated watermarks where correct watermark is at the 500th position when no attack is applied to the watermarked human voice sound for α=0.1 using the proposed and Cox's methods respectively.

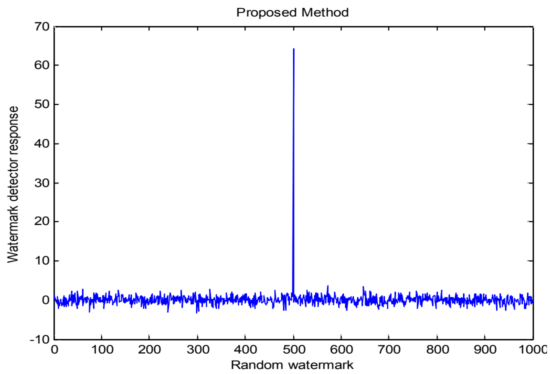

Fig 6. Watermark detector response using the proposed method
그림 6. 제안하는 방법을 사용한 워터마크 검출 응답

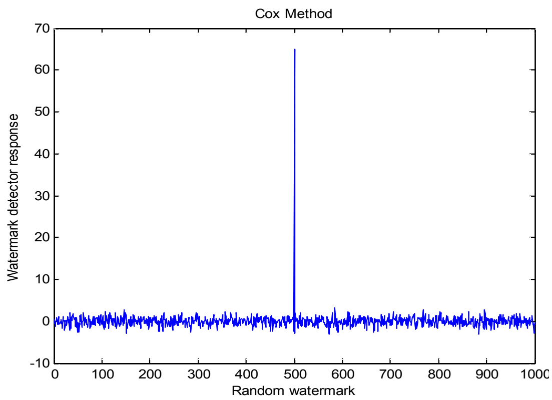

Fig 7. Watermark detector response using Cox's method
그림 7. Cox방법을 사용한 워터마크 검출 응답

In order to test the robustness of our proposed scheme, six different types of attacks, summarized in Table 3, were performed to the watermarked audio signal.

Table 3. Attacks used in this study for the watermarked sound
표 3. 워터마크 사운드를 위해 본 연구에서 사용된 공격

Attacks	Description
Noise addition	Additive white Gaussian noise (AWGN) is added with the watermarked audio signal.
Cropping	We removed 10% samples from the beginning of the watermarked signal and then replaced these samples by the original signal.

Resampling	The watermarked signal originally sampled at 44.1 kHz is resampled at 22.050 kHz, and then restored by sampling again at 44.1 kHz.
Re-quantization	The 16 bit watermarked audio signal is quantized down to 8 bits/sample and again re-quantized back to 16 bits/sample.
MP3 Compression	MPEG-1 layer 3 compression with 128 kbps is applied to the watermarked signal.
Low-pass filtering	The low-pass filter used in this study is a second order Butterworth filter with cut-off frequency 10 kHz.

Figure 8 shows the response of the watermark detector to 1000 randomly generated watermarks where correct watermark is at the 500th position against Gaussian noise attack for α=0.1 using the proposed method.

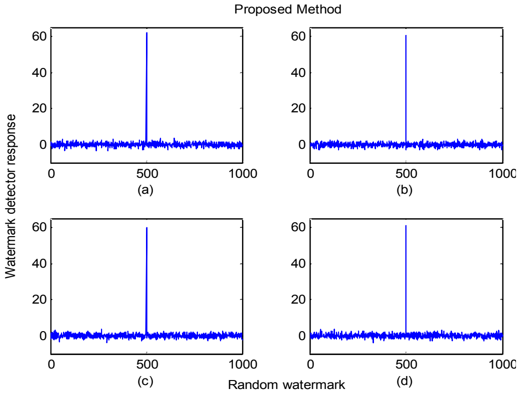

Fig 8. Watermark detector response against Gaussian noise attack using the proposed method: (a) Let it be, (b) Symphony No. 5, (c) Hey Jude, (d) Human Voice
그림 8. 제안하는 방법을 사용하여 가우시안 노이즈 공격에 대한 워터마크 검출 응답: (a) Let it be, (b) Symphony No. 5, (c) Hey Jude, (d) Human Voice

Tables 4, 5, 6 and 7 show the performance comparison in terms of robustness between the proposed scheme and Cox's method for different values of α against several kinds of attacks applied to four different types of watermarked audio signal 'Let it be', 'Symphony No 5', 'Hey Jude', and 'human voice' respectively. We observe that when α increases, similarity results of the proposed scheme and Cox's method also increase.

Table 4. Similarity results of the proposed scheme and Cox's method against different attacks for the audio signal 'Let it be'
표 4. 제안한 방법과 Cox방법의 공격에 대한 유사성 결과

Types of attack	Cox	Proposed	Cox	Proposed	Cox	Proposed
	α=0.1		α=0.2		α=0.3	
Noise addition	62.195	61.913	62.726	62.325	62.937	62.538
Cropping	58.729	55.652	59.237	57.824	60.462	59.372
Resampling	64.989	62.555	64.990	63.543	64.992	63.846
Re-quantization	64.772	62.286	64.875	63.725	64.923	63.948
MP3 compression	63.524	58.735	63.897	60.247	63.927	62.357
Lowpass filtering	61.514	58.511	61.625	60.725	61.826	61.436

Table 5. Similarity results of the proposed scheme and Cox's method against different attacks for the audio signal 'Symphony No 5'
표 5. 제안한 방법과 Cox방법의 공격에 대한 유사성 결과

Types of attack	Cox	Proposed	Cox	Proposed	Cox	Proposed
	α=0.1		α=0.2		α=0.3	
Noise addition	62.314	60.481	62.723	61.317	62.942	61.424
Cropping	57.961	59.493	58.627	60.527	59.266	61.478
Resampling	64.993	63.790	64.993	63.815	64.993	63.936
Re-quantization	64.993	63.921	64.993	63.957	64.993	63.983
MP3 compression	60.512	57.352	61.726	59.528	62.268	60.674
Lowpass filtering	59.623	57.272	60.216	59.632	60.583	60.215

Table 6. Similarity results of the proposed scheme and Cox's method against different attacks for the audio signal 'Hey Jude'
표 6. 제안한 방법과 Cox방법의 공격에 대한 유사성 결과

Types of attack	Cox	Proposed	Cox	Proposed	Cox	Proposed
	α=0.1		α=0.2		α=0.3	
Noise addition	59.619	59.805	60.347	60.528	61.413	61.725
Cropping	57.287	58.737	58.376	59.419	59.269	60.348

Resampling	64.983	61.977	64.993	62.576	64.993	62.865
Re-quantization	62.182	59.103	62.746	60.214	62.876	60.795
MP3 compression	63.492	58.413	63.872	60.528	63.945	61.213
Lowpass filtering	60.736	59.898	60.917	60.637	61.527	61.357

Table 7. Similarity results of the proposed scheme and Cox's method against different attacks for the audio signal 'Human voice'
표 7. 제안한 방법과 Cox방법의 공격에 대한 유사성 결과

Types of attack	Cox	Proposed	Cox	Proposed	Cox	Proposed
	α=0.1		α=0.2		α=0.3	
Noise addition	63.626	61.018	63.728	62.218	63.923	62.735
Cropping	59.760	56.468	60.246	58.583	61.372	59.672
Resampling	61.642	60.238	62.357	61.567	62.893	61.784
Re-quantization	63.933	61.072	64.215	62.136	64.637	62.538
MP3 compression	61.506	57.682	61.927	59.258	62.535	60.385
Lowpass filtering	60.137	56.516	60.728	58.327	60.927	58.727

Overall, our proposed watermarking system outperforms Cox's method in terms of SNR values while keeping comparable robustness with Cox's method against several attacks such as noise addition, cropping, resampling, re-quantization, MP3 compression, and low-pass filtering.

V. Conclusion

In this paper, we have presented a new watermarking system in frequency domain for audio copyright protection. Watermarks are embedded into the selected prominent peaks of the magnitude spectrum of each frame. Experimental results show that our proposed watermarking scheme shows better results than Cox's method in terms of imperceptibility while keeping comparable robustness with Cox's method against several kinds of attacks such as noise addition, cropping, resampling, re-quantization, MP3 compression, and low pass filtering. Our proposed method achieves SNR values ranging from 20 dB to 28 dB for different watermarked sounds. This is in contrast to Cox's method which achieves SNR values ranging from only 14 dB to 23 dB. These results demonstrate that our proposed watermarking scheme is a suitable candidate for audio copyright protection.

Reference

[1] 용승림, "신뢰기관 비참여의 핑거프린팅 기법에 관한 연구," 한국컴퓨터정보학회, 제 14권 7호, 81-88쪽, 2009년.

[2] 윤석민, 김인식, "프로젝트 난이도와 규모에 따른 소프트웨어 개발 프로세스," 한국컴퓨터정보학회, 제 14권, 제 6호, 33-40쪽, 2009년.

[3] I. Cox, J. Killian, F. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Transactions on Image Processing, Vol. 6, No. 12, pp. 1673-1687, Dec. 1997.

[4] W. N. Lie and L. C. Chang, "Robust and High-Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification," IEEE Transaction on Multimedia, Vol. 8, No. 1, pp. 46-59, Feb. 2006.

[5] P. Bassia, I. Pitas and N. Nikolaidis "Robust Audio Watermarking in the Time domain," IEEE Transaction on Multimedia, Vol. 3, No. 2, pp. 232-241, June 2001.

[6] L. Xie, J. Zhang and H. He, "Robust Audio Watermarking Scheme Based on Nonuniform Discrete Fourier Transform," in Proceedings of IEEE International Conference on Engineering of Intelligent System, pp. 1-5, April 2006.

[7] G. Zeng and Z. Qiu, "Audio Watermarking in DCT: Embedding Strategy and Algorithm," in Proceedings of 9th International Conference on Signal Processing (ICSP'08), pp. 2193-2196, 2008.

[8] J. Huang, Y. Wang, and Y. Q. Shi, "A Blind Audio Watermarking Algorithm with Self-Synchronization," in Proceedings of IEEE International Symposium on Circuits and Systems, (ISCAS'02), Vol. 3, pp. 627 - 630, 2002.

[9] M. Pooyan and A. Delforouzi, "Adaptive and Robust Audio watermarking in Wavelet Domain," in Proceedings of International Conference on International

Information Hiding and Multimedia Signal Processing (IIH-MSP 2007), Vol. 2, pp. 287-290, 2007.
[10] X. Y. Wang, H. Zhao, "A novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT," IEEE Transaction on Signal Processing, vol. 54, no. 12, pp 4835-4840, Dec. 2006.
[11] S. K. Lee and Y. S. Ho, "Digital audio watermarking in the cepstrum domain," IEEE Transactions on Consumer Electronics, Vol. 46, No. 3, pp. 744-750, Aug. 2000.
[12] S. C. Liu and S. D. Lin, "BCH code based robust audio watermarking in the cepstrum domain," Journal of Information Science and Engineering, Vol. 22, pp. 535-543, 2006.

저 자 소 개

프라납 쿠마르 다르

2004 : CSE, Chittagong University of Engineering and Technology, Chittagong, Bangladesh, Lecturer
2005-2008 : CSE, Chittagong University of Engineering and Technology, Chittagong, Bangladesh, Lecturer
2008 - 현재 : 울산대학교 컴퓨터정보통신공학부 석사과정
관심분야 : 사운드 합성, 디지털 워터마킹, 멀티미디어 압축 등

김 종 면

1995 : 명지대학교 전기공학사
2000 : University of Florida ECE 석사
2005 : Georgia Institute of Technology ECE 박사
2005 - 2007 : 삼성종합기술원 전문연구원
2007 - 현재 : 울산대학교 컴퓨터정보통신공학부 교수
관심분야 : 프로세서 설계, 임베디드 시스템, SoC, 컴퓨터 구조, 병렬처리