

공격에 강인한 고품질 디지털 오디오 워터마킹 시스템에 관한 연구

김현욱, 오현오, 윤대희, 차일환

연세대학교 전기, 전자공학과 음향, 음성, 신호처리 연구실

Study for A Robust Digital Audio Watermarking System

Hyun-Wook Kim, Hyen-O Oh, Dae-Hee Youn, Il-Whan Cha

Dep. of Electrical and Electronic Engineering, Yonsei University

(E-mail : woogie@radar.yonsei.ac.kr)

요 약

디지털 워터마킹은 오디오, 이미지, 영상과 같은 고품질 디지털 멀티미디어 콘텐츠의 불법적인 복제와 유통을 막고 저작권을 보호하기 위한 최후의 방어 수단으로 인식되고 있는 신호처리 기술이다. 오디오 신호에 삽입되는 워터마크 신호는 들리지 않아야 하는 특성과 함께, 압축이나 필터링과 같은 신호처리 과정 후에도 추출이 가능한 강인성을 지녀야 한다. 본 연구에서는 현재 제안되고 있는 여러가지 오디오 워터마킹 방법 중에서 대역확산을 기반으로 한 PN 시퀀스 오디오 워터마킹 시스템에 대해서 다루었다. 삽입된 워터마크 신호가 들리지 않도록 하기 위해 인간의 청각 시스템에 기반한 심리음향모델을 적용하였으며, 효율적인 검출과 청각상의 특성을 고려한 여러 가지 실험을 수행하였다. 워터마크 정보의 검출에 있어서는 원 신호가 필요하지 않은 Blind Detection 방법을 제안하여 효율적인 워터마킹 시스템을 구현할 수 있도록 하였다.

1. 서 론

워터마킹(Watermarking)은 정보감춤(Information Hiding)의 한 종류로서 사용자가 원하는 데이터를 다른 사람은 알 수 없도록 원래 데이터의 내부에 숨기는 것을 말한다. 그리고 그렇게 숨기는 정보를 워터마크(Watermark)라고 한다.

아날로그 워터마크는 상당한 역사를 가지고 있으며 중세 시대 교회의 비밀문서나 군사용 문서를 전달하기 위한 용도로 사용되어왔다[3]. "워터마크"라는 말은 종이에 보이지 않는 잉크로 글씨를 쓴 비밀 편지에서 유래되었다. 워터마크의 또 다른 말인 그리스어 "Steganography"는 Stegano (Covered)와 Graphos (Write)의 합성어로서 "Covered Writing" 즉, "숨겨진 글"을 의미한다[3].

디지털 워터마킹의 기본 목적도 이와 동일하다. 즉, 디지털 매체 내부에 다른 사람은 알 수 없도록 비밀 데이터를 삽입하고 필요한 경우 그 데이터를 추출할 수 있도록 만드는 것이다. 다만 적용하려는 매체와 방법이 디지털화된 것이 차이점이다.

디지털 워터마킹의 연구는 이미지에 대해 먼저 시작되었는데 인터넷의 확산과 함께 MP3와 같은 디지털 음악 매체의 사용이 일반화되면서 오디오에 대한 저작권의 보호가 필요하게 되었으며 그 해결 방안으로서 오디오 워터마킹이 떠오르게 되었다. 오디오 워터마킹은 이미지 워터마킹의 방법을 확장하여 적용하는 연구로부터 시작되었으며 곧 이어 본격적인 디지털 오디오 워터마킹 방법에 대한 연구가 세계적으로 활발히 이루어지게 되었다. 특히 최근에는 SDMI(Secure Digital Music Initiative)라는 기구를 중심으로 오디오 워터마킹의 표준을 만들기 위한 노력이 진행 중이다[7].

지금까지 연구된 대표적인 워터마킹 방법으로는 대역확산을 이용하는 방법([1][2][4])과 위상 정보를 변형하는 방법([2]), 에코를 삽입하는 방법([2]) 등이 있다. 본 논문에서 다룬 방법은 현재 가장 많은 연구가 진행되고 있는 대역확산(Spread Spectrum)을 이용한 PN 시퀀스 워터마킹이다. 특히, 워터마크 삽입 과정에서 심리음향모델을 적용하여 PN 시퀀스를 변형해 줌으로써 들리지 않게 워터마크를 삽입하면서도 검출 성능을 크게 향상시킬 수 있음을 실험을 통하여 보여주었다.

2. 디지털 오디오 워터마킹

2.1 개요 및 목적

디지털 오디오 워터마킹은 오디오 신호 안에 저작권과 같은 부가 정보를 들리지 않도록 삽입하고 추출하는 것

이다. 여기서 신호 안에 삽입한다는 것은 헤더와 같은 데이터의 부가적인 부분에 삽입하는 것이 아니라 원 신호 자체를 변경하거나 특성을 부여함으로써 워터마크를 삽입한다는 것이다. 헤더에 넣은 정보는 다른 사람에게 의해 쉽게 노출되어 삭제나 변경이 가능하기 때문에 워터마크로써 사용될 수 없다.

이러한 워터마킹의 주요 목적은 오디오 콘텐츠에 대한 저작권을 보호하고 불법복제나 무분별한 유통을 막을 수 있는 장치를 마련하는 것이다. 따라서 삽입되는 워터마크는 주로 저작권 정보나 사용자에 대한 제어 정보 등이 된다. 이 외에도 콘텐츠의 내용에 따른 검색을 위한 색인과 같은 메타 데이터로서의 기능도 가능하다.

2.2 워터마킹 시스템

워터마킹 시스템은 삽입(Embedding)과 검출(Detecting) 두 과정으로 나눌 수 있는데 개발할 때에는 두 과정을 동시에 고려해야 한다. 오디오 워터마킹이 갖추어야 할 필수적인 요건은 다음과 같다[1][3][9].

- 들리지 않아야 한다. (비지각성)
- 일반적인 신호처리과정과 악의의 공격에도 견딜 수 있도록 강인해야 한다.
- 원래의 데이터 없이 검출이 가능해야 한다.

이 외에도 다른 사람이 쉽게 알아낼 수 없도록 하는 보안성을 지녀야 하고, 추출한 워터마크 정보가 믿을 수 있을 정도의 신뢰성을 가져야 한다. 그리고 구현할 때의 시스템의 복잡성도 고려해야 한다.

3. 대역확산을 이용한 오디오 워터마킹

오디오 워터마킹은 방법적으로 크게 두 가지로 나눌 수 있다. 첫번째는 부가 신호의 삽입에 의한 것이고, 두번째는 원래 신호의 변형에 의한 것이다. 두 방법 모두 워터마킹에 의한 소리의 왜곡을 가져오게 되는데 이러한 왜곡을 최소화 하면서도 신뢰할 만한 검출이 가능하도록 만드는 것이 워터마킹의 핵심적인 문제이다.

대역확산(Spread Spectrum)을 이용한 방법은 삽입에 의한 워터마킹의 일종으로 PN 시퀀스를 워터마크 신호로 삽입하는 방법이다. 삽입되는 PN 시퀀스는 오디오 신호의 관점에서 볼 때 단순한 부가잡음이므로 검출이 가능한 한도 내에서 그 에너지를 최소로 하여 들리지 않도록 삽입해야 한다. 하지만 너무 작은 에너지로 삽입하면 강인성을 잃게 되어 공격에 의해 쉽게 없어질 수가 있으므로 비지각성과 강인성 사이에서 삽입될 워터마크 신호의 에너지 레벨을 결정해 주어야 한다.

3.1 PN 시퀀스의 특성

PN 시퀀스는 CDMA 와 같은 대역확산 통신에서 잘 알려진 것으로 잡음과 유사한 특성을 갖도록 만들어진 시퀀스이다. 백색 잡음과 유사한 통계적 특성을 갖도록 인위적으로 만들어진 신호이지만 실제로는 주기를 갖는

결정 신호(deterministic signal)이다. PN 시퀀스의 일반적인 특성은 다음과 같다[8].

- 백색 잡음과 유사한 통계적 특성을 갖는다.
- 간섭 신호에 대한 영향이 적다.
- 좋은 자기 상관(Auto-Correlation) 특성을 갖는다.

이러한 특성 중에서 특히 좋은 자기 상관을 갖는 것 때문에 워터마크로써의 활용이 가능하다. 즉, PN 시퀀스는 작은 에너지로 삽입이 되더라도 자기 상관도가 다른 신호와의 상관도에 비해 상대적으로 매우 큰 값을 갖기 때문에 검출에 용이한 것이다.

PN 시퀀스의 생성은 Shift Register 구조를 이용하여 만들게 된다. PN 시퀀스의 주기는 그 생성 다항식의 차수에 의해 결정되므로 Shift Register 의 개수와 직결된다. 즉, 레지스터 수가 m 개일 때 그 주기 N 은, $N = 2^m - 1$ 이 된다. 예로 레지스터 9개의 생성기에서 만들어지는 PN의 주기는 $511(2^9-1)$ 이 된다. 워터마크로 사용한 PN은 그 자체가 키 값이 되므로 추출을 할 때 반드시 알고 있어야만 한다. 반대로 삽입시에 사용된 PN 시퀀스를 정확하게 알지 못하면 워터마크를 검출할 수가 없게 된다. PN 시퀀스의 이러한 성질로 인해 워터마크의 보안성을 높일 수가 있다.

3.2 워터마크 삽입 (Embedding)

워터마크의 삽입 과정은 그림 1 과 같다.

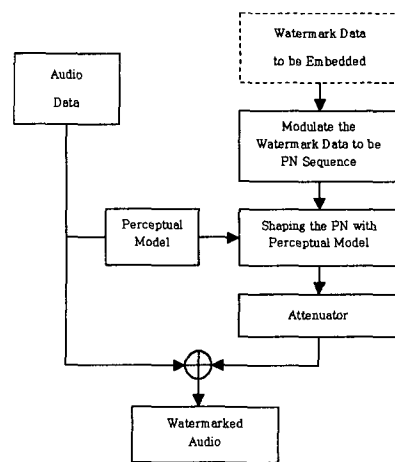


그림 1. 워터마크 삽입과정

삽입 과정을 살펴 보면, 먼저 입력 오디오를 프레임 단위로 나눈 후 각 프레임에 심리음향모형을 적용하여 마스킹 레벨을 결정한다. 그 다음 이 마스킹 곡선에 맞추어 삽입하려는 데이터에 의해 생성된 PN 시퀀스를 변형시키게 된다. 그러면 백색잡음처럼 평평한 스펙트럼을 갖는 PN 시퀀스가 그림 2와 같이 오디오 신호의 마

스킹 곡선을 따라 변형되고 결국에는 오디오 신호에 의해 마스킹 되어 들리지 않도록 만들어줄 수 있게 된다. 여기에 마지막으로 pre-echo 현상을 막기 위해 워터마크 신호의 에너지를 시간 영역에서 감쇄시킨 후 해당 프레임의 오디오 신호에 선형적으로 더해지게 된다. 그림 3은 시간 영역에서 오디오에 삽입된 워터마크의 일부를 나타낸 것이다.

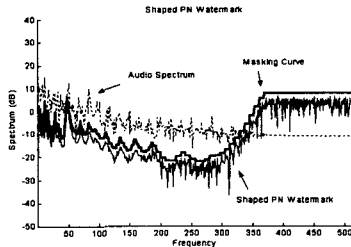


그림 2. 변형 삽입된 PN 워터마크 (주파수영역)

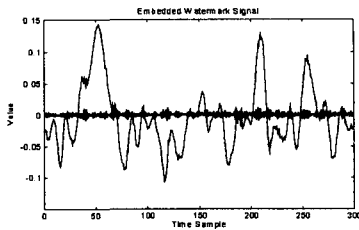


그림 3. 변형 삽입된 PN 워터마크 (시간영역)

3.3 심리음향모델 (Psychoacoustic Model)[5][6]

심리음향모델의 원리는 인간의 청각적 특성에 의한 마스킹 현상이다. 마스킹에는 동시마스킹과 순시마스킹이 있는데, 동시마스킹은 어떤 주파수 대역에 순음 또는 잡음이 존재할 때 그 대역 주변에 있는 작은 음압 레벨의 신호는 들리지 않는 현상이고, 순시마스킹은 시간의 차이를 두고 일어나는 두 신호의 시간 영역에서 나타나는 마스킹 현상이다[6]. 일반적으로 심리음향모델에서는 주파수 영역에서의 동시마스킹을 모델링하게 된다. 본 연구의 실험에서는 MPEG 오디오 부호화 표준의 심리음향모델 I을 사용하였다[5].

3.4 워터마크 검출 (Detecting)

워터마크의 검출은 PN의 특성을 이용한 상관 검출법을 사용하였다. 검출을 할 때에는 삽입시에 사용한 PN 시퀀스, 삽입한 오디오의 프레임과 삽입 동기점을 알아야만 한다. 즉, 이러한 정보가 워터마크 검출에 필요한 키값이 되는 것이다. 워터마크 검출에서 원 신호가 필요 없는 경우를 Blind Detection이라고 하는데 이렇게 원 신호 없이 워터마크를 추출할 수 있어야 워터마크로서의 활용성이 커지게 된다. 본 실험에서 제안한 워터마크의 검출 과정은 그림 4와 같다.

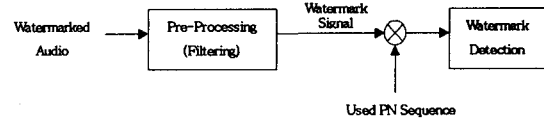


그림 4. 워터마크 검출과정

그림에서 보던 원 신호를 사용하지 않고 대신에 전처리 과정을 도입하여 워터마크가 삽입된 오디오 신호로부터 워터마크 신호만을 추출하고 있다. 그림 5는 전처리과정으로부터 추출된 워터마크 신호를 그린 것이다. 그런 다음 추출된 워터마크 신호와 삽입에 사용한 PN 시퀀스의 상관도를 취함으로써 워터마크 유무를 검출하게 된다. 그림 6은 한 프레임에 대해서 상관도를 취한 결과를 나타낸다.

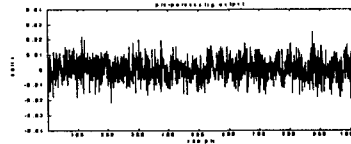


그림 5. 추출된 워터마크 신호

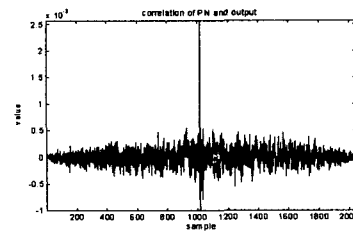


그림 6. 워터마크 신호와 사용한 PN 시퀀스의 상관도

4. 실험 및 결과

4.1 실험 내용

실험은 C1, C2, ..., C6로 명명한 보컬, 재즈, 클래식 등 다양한 장르의 곡에서 발췌한 6개의 테스트 음원을 가지고 수행하였다. 이들 각각에 대해 두 가지 워터마킹 방법 즉, "PN 직접 삽입 방법"과 "심리음향모델 적용 방법"으로 워터마크를 삽입하고 검출하여 결과를 관찰하였다. 이때 워터마크의 검출은 앞에서 제안한 검출기(그림 4)를 동일하게 사용하였다.

삽입되는 PN 시퀀스를 잡음이라고 보았을 때 오디오 신호의 에너지 대 삽입된 워터마크의 에너지 비는 다음과 같은 SNR로 볼 수가 있다. 즉,

$$SNR = 10 \log_{10} \left(\frac{\text{audio power}}{\text{watermark power}} \right) \text{가 된다.}$$

이 SNR 값에 따라 워터마크 신호의 지각도와 검출성능이 좌우되는데 본 실험에서는 비교를 위해 각각의 음원들과 삽입 방법에 대해서 20dB, 30dB, 40dB로 SNR을 변화시켜 실험을 수행하였다.

실험 결과 데이터로는 출력 오디오의 음질 평가와 검출 오류확률을 측정하였다. 우선 음질 평가는 원음과의 비교 청취를 통해서 잡음의 인지도를 측정하였는데 14명의 참가 인원에 대해서 1점부터 5점까지 점수를 주도록 하였다. 점수는 원음과 구별이 안되는 경우에는 5점을 주고 워터마크로 인한 잡음이 많이 들릴수록 낮은 점수를 주도록 하였다. 그리고 최종적으로 14명의 데이터에 대한 평균을 취하여 결과 값을 얻었다.

검출오류확률은 검출기로부터 나온 출력 비트와 삽입한 비트를 비교하여 발생한 오류 비트의 개수를 측정하였다. 워터마크 데이터의 한 비트는 삽입 과정에서 정의한 오디오의 한 프레임에 삽입된다. 실험에 사용한 여섯개 음원의 프레임 수는 각각 C1은 450, C2는 444, C3는 432, C4는 443, C5는 430, C6는 449 이었다. 이것을 이용하여 오류발생확률의 백분율을 구하였다.

4.2 실험 결과

실험의 결과는 표 1, 표 2와 같다.

직접 삽입	20dB		30dB		40dB	
	BER	MOS	BER	MOS	BER	MOS
C1	0	1.1	0	2.4	26.0	3.7
C2	0	1.2	1.4	3.1	19.1	4.3
C3	1.6	1.1	33.6	2.7	70.8	4.1
C4	0	1.3	1.6	2.4	39.3	3.8
C5	0	1.1	0	2.1	11.2	3.8
C6	0	1.1	0	1.9	2.9	3.3

표 1. PN 직접삽입 워터마킹 결과

변형 삽입	20dB		30dB		40dB	
	BER	MOS	BER	MOS	BER	MOS
C1	0	4.7	0	4.8	15.8	5.0
C2	0	4.8	1.6	5.0	19.1	5.0
C3	0.2	4.8	17.1	4.8	64.6	4.9
C4	0	4.8	0.2	4.9	14.7	5.0
C5	0	4.9	0	5.0	7.2	5.0
C6	0	4.7	0	4.9	1.8	5.0

표 2. 심리음향모델적용 워터마킹 결과

결과를 보면 SNR에 따른 검출 성능이 음원에 따라서 차이가 나지만 두 가지 삽입 방법에 대해서는 서로 크게 차이나지 않음을 알 수 있다. 하지만 음질 평가 결과에서는 “직접 삽입”의 경우 SNR 40dB 에서도 만족스러운 음질평가가 나오지 않았지만 “심리음향모델을 적용한 방법”에서는 SNR 20dB 에서 원 신호와 잘 구별이

안될 정도의 좋은 음질평가가 나왔다. 결국 심리음향모델을 사용한 방법이 사용하지 않은 것보다 20dB 이상의 검출 성능 향상을 가져올 수 있었다.

5. 결 론

오디오 워터마킹의 필수적인 사항은 워터마크 신호가 들리지 않으면서도 강인성을 가져야 한다는 것이다. 이를 위해서 워터마킹 시스템을 디자인할 때는 삽입(Embedding)과정과 검출(Detection) 과정을 모두 고려해 주어야 한다. PN 시퀀스와 같은 잡음 성분을 삽입하는 워터마킹의 경우 삽입과정에서 심리음향모델을 적용하면 검출성능은 유지하면서도 워터마크가 잡음으로 들리지 않게끔 만들어줄 수가 있다. 실험 결과를 보면 심리음향모델의 적용이 음질 면에서 20dB 이상의 향상을 가져왔으므로 결국 검출성능이 20dB 이상 높아진 것으로 볼 수 있다.

심리음향모델의 적용은 삽입과정에서 워터마킹 시스템의 성능을 향상시키는 것이라고 할 수 있는데 이와 함께 검출기의 구조를 개선하여 검출성능을 높임으로써 전체 워터마킹 시스템의 성능을 향상시킬 수 있을 것이다.

참 고 문 헌

- [1] L. Boney, A. Tewfik and K. Hamdy, "Digital Watermarks for Audio Signals," IEEE Proceedings of Multimedia '96, 1996.
- [2] W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding", IBM System Journal, Vol 35, Nos 3&4, 1996.
- [3] F. Petitcolas, R. Anderson, and M. Kuhn, "Information Hiding - A Survey," Proceedings of the IEEE, special issue on protection of multimedia content, vol. 87, No. 7, pp 1062-1078, July 1999.
- [4] I. J. Cox, J. Kilian, T. Leighton and T. Shanon, "A secure, robust watermark for multimedia", Proc. Workshop on Information Hiding, Univ. of Cambridge, U.K., May 30-June 1, 1996.
- [5] ISO/IEC JTC1/SC29/WG11 No.71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - CD 11172-3(Part3. MPEG-Audio)", 1992.
- [6] E. Zwicker, Psychoacoustics. Springer-Verlag, New York, 1982.
- [7] <http://www.sdmi.org/>
- [8] Bernard Sklar, Digital Communications Fundamentals and Applications, Prentice Hall, Inc., Englewood Cliffs, NJ, 1988.
- [9] M. Swanson, B. Zhu, and A. Tewfik, "Current state of the art, challenges and future directions for audio watermarking," 1999.