

시간 영역에서의 효율적인 오디오 워터마크 추출

강혜원, 정성환
창원대학교 컴퓨터공학과
e-mail: treebystream@hanmail.net

An Efficient Audio Watermark Extraction in Time Domain

Hea-Won Kang, Sung-Hwan Jung
Dept. of Computer Engineering, Changwon National University

요 약

본 논문은 기존의 오디오 워터마크 추출과정을 수정함으로써 원 신호의 영향을 크게 줄이고 추출 과정에서 효율성을 높이는 방법을 제안한다. 즉, 추출과정에서 전 처리 과정으로 원 신호와 유사한 신호를 제거하기 위해 간단한 저주파 필터의 사용과, 워터마크의 반복삽입을 통해 추출의 효율성을 향상시킨다. 몇가지 실제 오디오 데이터를 이용하여 실험한 결과, 약 79%의 워터마크 검출을 개선을 얻었다.

1. 서론

최근 인터넷의 발달로 인해 오디오 콘텐츠의 불법 복제와 유포가 확산되고 있으므로, 이를 억지하고 저작권을 보호하기 위한 오디오 워터마크(audio watermarking)에 관한 연구가 활발히 이루어지고 있다. 오디오 워터마크는 삽입 후의 신호가 삽입 전의 신호와 인지적으로 차이점이 없어야 하는 비지각성을 요구한다. 또한 추출할 때 여러 공격에 살아남아서 정확한 추출을 할 수 있는 강인성을 요구하는데, 비지각성과 강인성은 서로 상반되는 특징으로 워터마크는 두 가지 모두를 만족시켜야 한다[1].

현재 주파수 영역과 시간영역에서 적용되고 있는 워터마크 알고리즘은 비지각성을 만족시키기 위하여 잡음의 형태로 오디오에 삽입되는 경우가 많다. 이 경우, 상대적으로 삽입된 워터마크 신호에 비하여 원 신호의 에너지가 훨씬 더 강하므로, 추출과정에서의 워터마크와 워터마크가 삽입된 신호와의 상관관계를 현저하게 떨어뜨리는 원인이 되어 정확한 워터마크의 추출이 어렵다[2].

본 논문은 이러한 문제점을 해결하기 위하여 Bassia가 제안한 방법[3] 토대로 수정함으로써 원 신호의 영향을 크게 줄이고, 추출과정에서 효율성을 높이는 방법을 제안한다. 즉, 추출과정에서 전 처리 과정으로 원 신호와 유사한 신호를 제거하기 위해 간단한 저주파 필터의 사용과, 워터마크의 반복삽입을 통해 추출의 효율성을 향상시켰다.

2. 시간영역에서의 오디오 워터마크 시스템

2.1 삽입과정

오디오 신호는 L 개의 샘플을 가진 N_s 개의 프레임으로 분해된다. k 번째 프레임에서의 오디오 샘플을 $x_k(i)$ 로 정의하면,

$$x_k(i) = x(kL + i) \quad \begin{matrix} i = 1, 2, \dots, L \\ k = 0, 1, \dots, N_s - 1 \end{matrix} \quad (1)$$

이다. 워터마크 $w(i)$ 는 -1과 1로 랜덤하게 구성된

길이 L 의 워터마크이다.

$$w(i) \in -1, 1, \quad i = 1, 2, \dots, L \quad (2)$$

삽입된 워터마크를 잘 인지하지 못하게 하려면 원 신호의 에너지가 강한 곳은 강하게, 약한 곳은 약하게 들어가야 한다[2]. 그러므로 이 때 k 번째 프레임에 삽입되는 신호 의존적(signal dependent) 워터마크 $w'_k(i)$ 는 다음 식 (3)과 같다.

$$w'_k(i) = \alpha |x_k(i)| \oplus w(i) \quad i = 1, 2, \dots, L$$

$$k = 0, 1, \dots, N_s - 1 \quad (3)$$

여기서 α 는 상수이고, \oplus 는 덧셈, 곱하기, 승(power) 중의 하나를 의미한다.

$$y_k(i) = x_k(i) + w'_k(i) \quad (4)$$

최종적으로 k 번째 프레임에서 워터마크가 삽입된 오디오 신호 $y_k(i)$ 는 식 (4)이다.

2.2 추출과정

다음은 Bassia가 제안한 워터마크 추출방법이다 [3]. 추출대상인 신호 y 를 길이가 L 인 N_s 개의 프레임으로 나눈다. 워터마크의 삽입여부는 k 번째 프레임에서의 신호 $y_k(i)$ 와 워터마크 패턴 $w(i)$ 와 상관관계 계산을 토대로 한다.

$$S_k(n) = \frac{1}{L} \sum_{i=1}^L y_k((i+n) \bmod L) w(i) \quad (5)$$

여기서 $k=0, 1, \dots, N_s-1$ 이다.

그리고 식(4),(5)를 이용하여 식(6)을 얻는다.

$$S_k(n) = \frac{1}{L} \sum_{i \in A} x_k((i+n) \bmod L) w(i) + \frac{1}{L} \sum_{i \in B} w'_k((i+n) \bmod L) w(i) \quad (6)$$

$\sum_{i \in A} w(i) = \Delta w$ 이라 하면, $w(i)$ 는 합이 0인 부분집합 A와 나머지 부분인 합이 Δw 인 부분집합 B로 나눌 수 있다. 식 (6)의 첫 번째 항을 집합 A와 집합 B를 사용하여 식 (7)과 같이 표현할 수 있다.

$$S_k(n) = \frac{1}{L} \sum_{i \in A} x_k((i+n) \bmod L) w(i) + \frac{1}{L} \sum_{i \in B} x_k((i+n) \bmod L) w(i) + \frac{1}{L} \sum_{i \in B} w'_k((i+n) \bmod L) w(i) \quad (7)$$

위의 첫 번째 항을 $T_{1,k}(n)$ 로, 두 번째 항을 $T_{2,k}(n)$ 로, 세 번째 항을 $T_{3,k}(n)$ 로 정의한다. Bassia는 첫 번째 항에서 집합 A의 평균이 0이므로 $T_{1,k}(n)$ 을 0으로 가정하고,

$$S_k(n) = T_{2,k}(n) + T_{3,k}(n) \quad (8)$$

이라고 한다.

따라서 만일 워터마크가 삽입되지 않은 경우에는 $T_{3,k}(n)$ 가 0이므로, $S_k(n) = T_{2,k}(n)$ 가 된다. 그러므로 다음 식 (9)를 이용하여 워터마크를 추출한다.

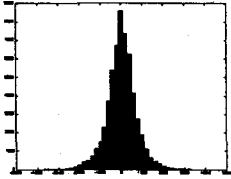
$$r_k(n) = \frac{S_k(n) - T_{2,k}(n)}{T_{3,k}(n)} \quad (9)$$

즉, $r_k(n)$ 이 1에 가까우면 워터마크가 있는 것으로, 0에 가까우면 워터마크가 삽입되지 않은 것으로 판단한다.

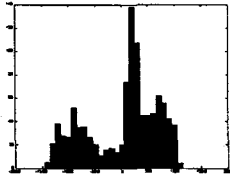
2.3 기존방법의 문제점

Bassia는 상관관계를 구할 때 식 (7)에서 첫 번째 항인 $T_{1,k}(n)$ 가 거의 0이라는 전제 하에서 처리하였다. 그러나, $T_{1,k}(n)$ 이 0이 되려면 프레임의 길이 L 이 충분히 크며, 원 신호가 대칭구조 혹은 가우시안 분포를 가져야 하는데 실제 워터마크가 삽입된 오디오 신호는 이런 전제를 만족시키지 못한다. 일반적으로 통계에서 모집단의 크기가 커질수록 분포는 이상적인 가우시안 분포에 다가가므로, 원 신호 전체의 히스토그램은 [그림 1]과 같이 가우시안 혹은 대칭구조에 가깝다. 그러나 식 (9)는 각각의 프레임에 적용되는데, L 개의 샘플로 이루어진 각 프레임의 히스토그램은 실제로는 [그림 2]와 같이 대칭구조를 이루지 못하는 경우가 대부분이다.

이러한 단점을 보완하고 실험의 신뢰성을 높이기 위해 H. Malvar 등은 워터마크를 삽입하기 전에 원 신호가 평균이 0인 가우시안 분포가 되도록 전 처리를 한다[4]. 그러나 이러한 처리는 비가역적인 오디오 신호의 변화를 가져올 뿐만 아니라, 오디오 자체의 음질을 손상시킬 수 있다. 또한 일반적으로 삽입되는 워터마크 신호에 비해서 원 신호의 상대적인



[그림 1] 오디오 신호 전체의 히스토그램



[그림 2] 20번째 프레임의 히스토그램

에너지가 훨씬 더 강하다.

각 프레임 신호의 히스토그램이 비대칭이라는 것과 원 신호가 워터마크에 비해 상대적으로 강하다는 두 가지 이유로 기존의 상관관계 식 (9)를 아무런 전 처리 없이 사용할 경우, 원 신호의 영향이 커져서 워터마크 추출이 어렵다.

본 연구에서는 이런 원 신호의 영향을 최소화하기 위해 전 처리 과정으로 간단한 평균 필터를 이용하여 얻은 신호와 원 신호의 차를 이용하여 상관관계를 구하고, 워터마크의 반복 삽입을 통해 삽입된 워터마크 추출의 정확성을 높이는 방법을 제안한다.

3. 제안한 워터마킹 시스템

3.1 삽입과정

워터마크 생성은 2장에서 설명한 방법과 동일하나, 서로의 상관관계가 영인 두 개의 워터마크 패턴 $w_{-1}(i)$, $w_1(i)$ 를 만든다. -1과 1로 이루어진 데이터 시퀀스가 각 프레임마다 삽입되는데, 각 프레임에 삽입되는 데이터 시퀀스를 j 라 정의한다. j 는 원 신호에 따라 식 (10)과 같이 변형된다.

$$w_{j,i}(i) = \alpha x_k(i) \oplus w_j(i), \quad i=1, 2, \dots, L$$

$$k=1, 2, \dots, N_s, \quad j=-1, 1 \quad (10)$$

워터마크인 데이터 시퀀스가 오디오 원 신호에 삽입되면 식 (11)과 같다.

$$y_{j,k}(i) = x_k(i) + w'_{j,k}(i) \quad (11)$$

만일 삽입될 데이터 시퀀스의 길이가 M 이고, 전체 프레임의 개수가 N_s 이면, 삽입되는 데이터 시퀀스의 반복회수 N 은 식 (12)와 같이 될 수 있다.

$$N = \frac{N_s}{M} \quad (12)$$

삽입될 데이터 시퀀스의 반복회수 N 이 클수록 워터마크의 효율적인 추출이 가능하다. 전체 오디오 프레임의 개수 N_s 가 제한되어 있으므로 데이터 시퀀스의 길이 M 이 클수록 반복회수 N 은 줄어들게 된다.

3.2 추출과정

본 시스템은 전 처리 과정으로 원 신호의 영향을 줄이기 위하여 저주파 필터를 사용하였다. 저주파 필터로는 간단한 평균 필터를 이용하여 워터마크 삽입된 신호 y 로부터 평균필터를 이용하여 신호 y' 를 얻는다. 그리고 식 (13)에 의해서 신호 y 와 대부분의 에너지를 차지하는 y' 의 차를 구한다.

$$z = y - y' \quad (13)$$

신호 z 를 길이 L 의 프레임으로 분해한다.

$$z_k(i) = z(kL+i) \quad i=1, 2, \dots, L \quad (14)$$

k 번째 프레임에서 j 는 -1과 1, 모두에 대한 $z_k(i)$ 와 $w_j(i)$ 에 대한 상관관계 $S_{j,k}(n)$ 구한다.

$$S_{j,k}(n) = \frac{1}{L} \sum_{i=1}^L z_k((i+n) \bmod L) w_j(i)$$

$$j = \forall -1, 1 \quad (15)$$

$S_{-1,k}(n)$ 와 $S_{1,k}(n)$ 중에서 더 큰 값을 S_k 로 정한다. 그리고 실제로 프레임 길이 L 이 작은 경우에서 식 (9)에서 $T_{2,k}$ 는 작은 값이 되므로

$r_{j,k}(n) \approx \frac{S_k}{T_{3,k}}$ 로 근사할 수 있다. 따라서 k 번째 프레임의 삽입 데이터는 다음 식 (16)과 같이 정할 수가 있다.

$$\begin{aligned} &\text{if } r_{-1,k} > r_{1,k}, \text{ Data} = -1 \\ &\text{if } r_{-1,k} < r_{1,k}, \text{ Data} = 1 \end{aligned} \quad (16)$$

길이 M 인 삽입된 데이터 시퀀스에서 i 번째 데이터를 $Data(i)$ 로 정의하자. 삽입된 데이터 시퀀스가 N 번 반복되므로, p 번째 반복될 때의 데이터는 $Data_p(i)$ 로 정의한다. 여기서 $p=1, 2, \dots, N$, $i=1, 2, \dots, M$ 이다.

$$Q(i) = \frac{1}{N} \sum_{j=1}^N Data_j(i) \quad (17)$$

$Q(i)$ 는 삽입된 데이터 시퀀스의 평균값이다.

식 (18)에 의하여 최종 추출될 데이터 시퀀스를 결정한다.

$$\begin{aligned} \text{if } Q(i) > Th^+, \quad D(i) &= 1 \\ \text{if } Q(i) < Th^-, \quad D(i) &= -1 \end{aligned} \quad (18)$$

여기서 Th^+ 와 Th^- 는 적절한 임계치(threshold)이고, $D(i)$ 는 추출된 데이터이다.

4. 실험 및 토의

본 연구에서는 제안된 방법을 실험하기 위해 <표 1>과 같은 실험 데이터를 비롯하여 여러 가지 오디오 데이터를 사용하여 실험하였다.

<표 1> 실험에 사용된 오디오 데이터

No.	창작자	곡명	특징	파일형태
1	비발디	사계 중 봄	클래식	wave
2	creed	you	락	wave
3	박진영	너의 뒤에서	팝	wave
4	쇼팽	죽혼환상곡	클래식	wave
5	생상스	동물의 사육제	클래식	wave
6	라흐마니노프	첼로소나타	클래식	wave

본 실험에서는 프레임 길이 $L=1024$, 삽입된 데이터 시퀀스의 길이 $M=8$, 데이터 시퀀스가 반복 삽입된 회수 $N=12$ 로 하였다. Th^+ 는 0.5, Th^- 는 -0.5로 정하였다.

<표 2> 기존의 방법과 제안한 방법의 검출률 비교 1

데이터 시퀀스	No.1				No.2				No.3			
	기존		제안		기존		제안		기존		제안	
	Q	D	Q	D	Q	D	Q	D	Q	D	Q	D
-1	0.5	1	-0.67	-1	0.17	?	-1	-1	0	?	-0.5	-1
-1	0	?	-1	-1	-0.50	-1	-1	-1	0.33	?	-0.83	-1
-1	0	?	0.67	1	0.33	?	0.83	1	-0.67	-1	0.33	?
-1	0	?	-1	-1	-0.33	?	-0.83	-1	-0.17	-1	-0.67	-1
1	-0.17	?	0.5	1	0	?	1	1	0.33	?	1	1
1	-0.17	?	0.83	1	0	?	0.67	1	-0.33	?	0.5	1
-1	0	?	-0.83	-1	-0.33	?	-1	-1	0.17	?	-0.83	-1
-1	0.5	1	-0.83	-1	-0.17	?	-0.83	-1	0.17	?	-0.5	-1
검출률%	0		100		12.5		100		0		87.5	

<표 2>는 <표 1>에 있는 실험 오디오 데이터 중에서 대표적인 클래식, 락, 팝 음악을 사용하여 제안된

방법과 기존의 방법을 비교한 결과이다. 표에서 “?”는 임계치를 사용하여 판독할 수 없는 데이터를 의미한다.

<표 3>은 기존방법과 제안한 방법의 검출율을 비교한 것이다.

<표 3> 기존의 방법과 제안한 방법의 검출률 비교 2

	No.	1	2	3	4	5	6	평균
검출율	기존	0	12.5	0	0	37.5	50	16.7
%	제안	100	100	87.5	87.5	100	100	95.8

표에서 필터링을 하지 않는 기존의 방법은 검출율의 평균이 약 17%로 낮게 나타났다. 그러나 제안된 방법에서는 검출율의 평균이 약 96%로 기존의 방법보다 약 79%가 개선되었다.

5. 결론

본 연구는 시간 영역 상에서의 기존 Bassia 방법의 추출방법을 수정해서 효율적인 워터마크 신호의 추출 방법에 관한 것이다. 본 논문에서는 전처리로 간단한 평균 필터를 사용하여 추출과정에서의 원 신호의 영향력을 감소시켰고, 워터마크의 반복삽입을 통해 추출의 효율성을 향상시켰다.

비발디 등의 대표적인 오디오 데이터들로 실험한 결과 기존 방법에 비해 검출율이 약 87% 증가가 있었다. 차후 연구로서는 실험 데이터의 확장과 강인성을 보완한 추출 알고리즘을 연구할 계획이다.

참고문헌

[1] I. J. Cox, J. Kilian, T. Leight and T. Shmoon, "A Secure, Robust Watermark for Multimedia," in Workshop on Information Hiding, Newton Intitute, Univ. of Cambridge, pp.147-158, May 1996.
 [2] S. Jung, J. Seok, and J. Hong, "An Improved Detection Technique for Spread Spectrum Audio Watermarking with a Spectral Envelope Filter," in ETRI Journal, Vol.25, No.1, pp.52-54, February 2003.
 [3] P. Bassia, I. Pitas and N. Nikolaidis, "Robust Audio Watermarking in the Time Domain," in IEEE Transactions on Multimedia, Vol.3, No.2, June 2001
 [4] D. Kirovski and H. Malvar, "Spread Spectrum Watermarking of Audio Signals," in IEEE Transactions on Signal Processing, Vol.51, No.4, pp.1020-1033, April 2003.