

# 반향 오디오 워터마킹의 성능 향상

오현오, 윤대희, 석종원\*, 홍진우\*

연세대학교 전기전자공학과 미디어·통신 신호처리 연구실  
\* 한국전자통신연구원 무선방송기술연구소 방송미디어부

## Enhancement of Echo Audio Watermarking

Hyen-O Oh, Dae Hee Youn, Jong-Won Seok\*, Jin-Woo Hong\*

MCSP Lab., Dept. of Electrical and Electronic Eng., Yonsei University  
\* Radio & Broadcasting Tech. Lab., ETRI  
(E-mail : oho@mcs.yonsei.ac.kr)

### 요 약

반향(Echo)을 이용한 워터마킹은 오디오 신호에 인위적인 반향을 첨가함으로써 정보를 삽입한다. 다른 오디오 워터마킹 방법과 마찬가지로 반향 오디오 워터마킹은 시간축 공격에 대해 강인하지 못한 단점을 가지고 있다. 특히, 오디오 신호의 피치를 보존하면서 재생 시간을 변형시키는 시간 스케일 변형(Time Scale Modification)에 대해서는 별도의 방어를 위한 알고리즘이 없을 경우 전혀 복호화가 이뤄지지 않는다. 본 논문에서는 반향 오디오 워터마킹의 성능 향상을 위해 시간 스케일 변형 공격에 대응하여, 변형된 정도를 검출하고 보상하여 복호화가 가능하도록 하는 새로운 알고리즘을 제안한다.

### 1. 서 론

디지털 워터마킹은 MP3와 같은 고음질의 디지털 오디오 콘텐츠의 불법적 유통을 막고 저작권을 보호하기 위해 사용되는 새로운 기술이다. 디지털 오디오 워터마킹이란 들리지 않는 정보를 지각적 왜곡 없이 오디오 신호 자체에 은닉시키는 방법으로, 특히, 저작권 보호를 위해 사용되는 워터마킹은 여러 가지 신호처리 과정을 거친 후에도 사라지지 않도록 강인하게 삽입되어야만 한다[1]. 지금까지 개발된 여러 워터마킹 기법들은 특히 시간 영역에서의 신호 처리 공격 가운데 하나인 시간 스케일 변형(Time Scale Modification)에 대해 매우 취약한 대응 결과를 나타낸다.

반향(Echo)을 이용한 워터마킹에서는 오디오 신호에 인위적인 반향을 삽입함으로써 정보를 삽입하게 된다. 적절하게 삽입된 반향은 사람에게 왜곡으로 인지되지 않으며 오히려 음색을 풍부하게 해주는 효과를 기대할 수 있기 때문에 오디오 워터마킹에 사용하기에 유리한 방법이다 [1][2][3]. 특히, 삽입하는 반향의 크기와 시간 지연을 적절히 조절하면 원음과 청각적으로 전혀 구별되지 않는 반향 신호를 삽입할 수 있다. 그리고, Oh *et al* [2][3]이 제안한 반향 커널을 이용하게 되면, 음질 왜곡 없이 매우 큰 에너지를 갖는 반향의 삽입이 가능하며, MP3 압축, 필터링 등의 공격에 강인한 워터마킹이 가능하다. 또한, 시간 영역에서의 공격 가운데, 선형 속도 변화(Linear Speed Change) 같은 경우, 변화된 시간 길이에 비례하여 삽입된 반향의

시간 지연이 변화한다는 성질을 이용하면, 적절한 보상에 의해 대응이 가능하다[2]. 그러나, 시간 스케일 변형에 대해서는 기존의 반향 오디오 워터마킹에서도 아직 뚜렷한 대응 방법이 제안되지 못하였으며, 공격에 의해 시간축 길이가 변화하고 이에 따라 프레임 동기를 상실함으로써 전혀 복호화가 이뤄지지 않는 문제를 가지고 있다.

본 논문에서는 반향을 이용한 오디오 워터마킹에서 시간 스케일 변형 공격에 대응하여, 변형된 정도를 검출하고 보상하여 복호화가 가능하도록 하는 새로운 알고리즘을 제안한다.

### 2. 시간 스케일 변형

고음질의 시간 스케일 변형은 노래방 기기의 키 변형, wavetable 기반의 음악 합성, 비디오 및 음성 녹음 장치의 고속 검색 등에 활용되고 있는 기술이다. 시간 스케일 변형이란 음성 및 오디오 신호가 가지고 있는 본래의 피치(pitch) 특성을 유지하면서 시간 길이를 다르게 하는 기술을 말한다.

시간 스케일 변형은 크게 시간 영역에서 변형하는 방법과 주파수 영역에서 변형하는 방법으로 나누어진다. 시간 영역에서 변형하는 방법은 주파수 영역에서 변형하는 방법에 비해 간단한 연산량만으로도 고음질의 시간 스케일 변형이 가능하기 때문에 널리 사용되고 있다. 대표적인 시간 영역의 알고리즘으로는 SOLA (Synchronized Overlap and Add) [2]가 있다. SOLA는 중첩 가산 방식을 기반으로 오디오 신호의 시간 변형을 수행하는데, 중첩될 구간이 시간 축에서 유사한 곳을 찾아 중첩 가산함으로써 음질을 보장하는 방식을 취하고 있다.

SOLA를 비롯한 시간 영역에서의 시간 스케일 변형 방법들은 중첩이 이루어지는 구간에서만 원 신호에 변형이 생기고, 나머지 구간에서는 원 신호와 동일한 신호를 갖는다. 또한, 시간 스케일 변형은 시간 축에서의 길이만을 변형할 뿐, 피치를 비롯한 원 신호의 주파수 특성을 그대로 보장하기 때문에, 반향 워터마킹에서 검출에 이용되는 반향의 피크 특성 등의 변수들도 변화하지 않는다. 다만, 시간축의 길이가 변화되어, 이로 인한 프레임의 위치의 변동이 워터마크의 복호화를 방해하는 요소로 작용하게 된다. 위

터마킹의 공격 대상이 되는 시간 스케일 변형은 +/- 4% 이내이며, 이 범위 안에서 임의로 변형되었을 때, 워터마크의 복호화가 가능하여야 한다.

그림 1은 시간 영역에서 10개의 구간으로 나누어진 오디오 신호 (a)가 시간 스케일 변형에 의해 길이가 9개로 짧아진 경우 (b)와 11개로 길어진 경우 (c)에 대한 예를 나타낸다. 원 신호 (a)의 5번째와 6번째 구간이 중첩되면서 시간을 줄이거나 늘리고 있는 경우이다. 그림 1의 (b), (c)를 살펴보면, 5번째 구간 이후는 시간 축 위치의 변형이 생기기 때문에, 이를 적절히 보상해주지 못할 경우, 워터마크의 복호화 시 프레임의 동기를 잃게 되어 올바른 값을 얻을 수 없다. 시간 스케일 변형에 의한 시간 축의 위치 변화 정도는 프레임이 진행될수록 커지기 때문에, 적절한 보상을 해주지 못하면, 오차는 계속해서 누적된다.

시간 스케일 변형에서 중첩되는 구간의 길이는 원 신호의 최소 주파수가 갖는 주기 이상이어야만 한다. 그러나, 중첩 구간이 길수록 변형 알고리즘의 연산량이 증가하므로, 가급적 작은 값을 사용하는 것이 일반적이다. 워터마킹에서 사용되는 프레임의 길이는 이 중첩 구간보다 충분히 커야만 프레임 손실 없이 복호화가 가능하다.

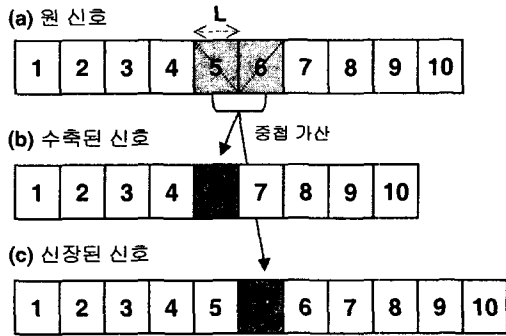


그림 1. 시간 스케일 변형에 의한 길이 변형의 예

### 3. 시간 스케일 변형에 강인한 반향 워터마킹

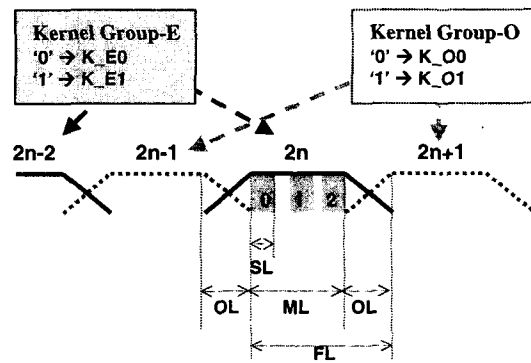
본 논문에서는 동기화를 위한 별도의 추가 신호 삽입 없이 워터마크 신호인 반향만을 이용하여 시간 변형된 오디오 신호로부터 프레임 동기를 찾아내는 알고리즘을 제안한다. 이를 위해, 부호화 과정에서는 짝수 프레임과 홀수 프레임에 다른 반향 커널을 삽입하는 듀얼 커널 방식을 적용하고, 복호화 과정에서는 프레임 동기 추정을 위해 프레임 내의 각기 다른 위치에서 반향의 피크를 검출하는 방법을 제안한다.

#### 3.1 부호화 과정

제안하는 동기화 방법의 적용을 위한 부호화 과정은 짝수 프레임과 홀수 프레임에 다른 반향 커널을 사용하여 삽입한다는 것을 제외하고는 기존의 반향 워터마킹 방식과 동일한 구조를 갖는다. 짝수 프레임과 홀수 프레임에 다른 반향 커널을 삽입하는 방식을 듀얼 커널이라 부르며, 이는

이전에 또 다른 반향 삽입에 의한 공격에 대한 대응을 위해 사용된 바 있는 방법이다 [2].

그림 2는 제안하는 부호화 및 복호화 과정에 사용되는 반향 워터마킹의 프레임 구조를 나타낸다. 반향 커널은 이진수 '0'을 나타내기 위한 커널 2개, 이진수 '1'을 나타내기 위한 커널 2개가 사용되므로 총 4개의 각기 다른 시간 지연을 갖는 커널이 이용된다. 워터마크 삽입 시에, 짝수 프레임에서는 커널 그룹 E의 반향 커널을 이용하고, 홀수 프레임에서는 커널 그룹 O의 반향 커널을 이용하여 워터마크를 삽입하게 된다. 예를 들어, 짝수 프레임에 이진수 '0'을 삽입하고자 할 경우에는 반향 커널  $K_{E0}$ 를 사용하고, 홀수 프레임에 이진수 '1'을 삽입하는 경우에는 반향 커널  $K_{O0}$ 를 사용한다. 프레임 길이 (FL)는 그림 1의 시간 스케일 변형을 위한 중첩 구간 길이 (L)보다 큰 값을 사용하며, 또한, 한 프레임 내에 3개의 서브프레임이 존재하도록 프레임 길이를 채택하여 워터마크를 삽입한다.



SL : 서브프레임 길이      OL : 중첩 구간 길이  
FL : 프레임 길이        ML : 비 중첩 구간 길이

그림 2. 시간 스케일 변형 검출을 위한 워터마킹 삽입 및 검출 프레임 구조

#### 3.2 복호화 과정

반향 워터마킹의 복호화 과정은 각 프레임별로 캡스트럼 분석을 수행하고, 반향 커널의 시간 지연에 대응되는 캡스트럼 피크의 위치에 따라 이진수 워터마크를 검출하는 순서로 이루어진다 [2]. 듀얼 커널을 이용한 경우, 동일한 이진수 값이 삽입되었어도 짝수 프레임과 홀수 프레임에서 피크가 나타나는 위치는 서로 다르다. 한 프레임을 3개의 서브프레임 (그림 2의 0,1,2)으로 나누고 각 서브프레임별로 독립된 피크 검출을 수행할 때, 시간 축이 변화하지 않고 정상적인 경우이면, 각 서브프레임에서의 결과는 같은 값을 나타낼 것이다. 즉, 해당하는 반향 커널의 시간 지연 위치에서는 모두 피크가 검출되어야 하고, 그렇지 않은 시간 지연 위치에서는 피크가 검출되지 않아야 한다. 그림 3의 (a)는 시간 변형이 발생하지 않아 프레임 동기가 맞았을 경우, 2n 번째 프레임의 각 서브프레임에서 캡스트럼 피크가 정상적으로 검출될 수 있음을 도식적으로 나타내고 있다. 그러나, 그림 3의 (b)의 경우처럼, 2n-1 번째 프레임에서 시간 스케일 변형에 의해 길이가 S만큼 단축될 경

우, 2n 번째 프레임의 시작점은 S 만큼 왼쪽으로 이동되게 될 것이다. 그 영향으로 2n 번째 프레임의 서브프레임 중  $i=2$  의 결과는 정상적인 피크가 검출되지 않을 가능성이 높아진다. 한편, 그림 3의 (c)에서처럼 2n-1 번째 프레임이 시간 스케일 변형에 의해 길어진 경우, 2n 번째 프레임의 시작점은 S' 만큼 오른쪽으로 이동하게 된다. 그리고, 그 결과  $i=0$  에서의 피크는 정상값 보다 낮은 값이 나타나고 오히려 2n-1 프레임의 신호의 영향으로 홀수 프레임의 반향 커널인 그룹 O의 커널 위치에서 피크가 나타날 수 있다.

결국, 복호화기에서는 짝/홀수 프레임에 따라 해당 커널의 시간 지연 위치에 펄스가 정상적으로 나타나는지 검출함으로써 시간 스케일 변형 공격 여부를 확인할 수 있다. 또한, 변형된 정도에 따라 프레임의 시작점을 보정함으로써, 이후 프레임들에서 동기를 회복할 수 있다.

표 1은 각 서브 프레임에 검출된 피크 값의 패턴과 시간 축 변화 사이의 상관 관계를 나타낸 논리표이다. 짝/홀수 프레임에 따라 결정되는 그룹 E 혹은 그룹 O의 커널 중 하나의 결과를 위의 논리표와 비교하여 적절한 시간축 동기를 보정한다. 이때, 해당 커널 그룹에 대한 상대 커널 그룹의 캡스트럼 결과를 함께 고려하면 보다 안정적이고 정확한 보상을 할 수 있다.

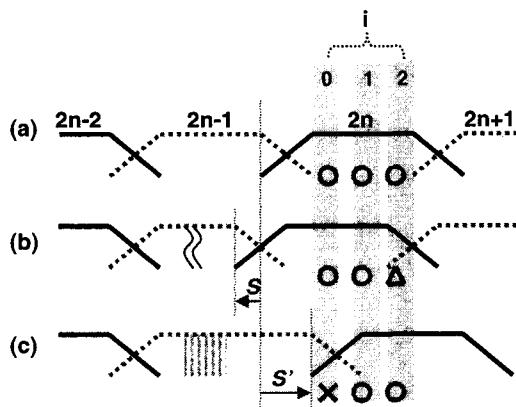


그림 3. 시간 스케일 변형에 의한 워터마킹 복호화 프레임 위치 변화

표 1. 서브프레임에서의 피크 값에 따른 시간축 변화 관계

SubFrame (i)	시간축 변화	SubFrame (i)	시간축 변화
0 1 2		0 1 2	
O O O	없음	X X X	FL 단축/신장
O O Δ	OL/2 단축	Δ O O	OL/2 신장
O O X	OL 단축	X O O	OL 신장
O Δ X	3*OL/2 단축	X Δ O	3*OL/2 신장
O X X	2*OL 단축	X X O	2*OL 신장

O : 정상 피크 검출, Δ : 정상 피크의 1/2 검출, X : 피크 검출 안됨.

그림 4는 각 서브 프레임에서의 캡스트럼을 이용하여 복호화기에서 생성해낸 동기 신호이다. 그림 4의 (a)는 커널 그룹 E의 시간 지연 위치에서의 캡스트럼 값 중 큰 값을 선택하여 나타낸 것으로, 짝수 프레임에서는 어떤 임계치 이상의 피크가 형성되고 홀수 프레임에서는 피크가 나타나지 않아야 하므로 프레임 동기 신호와 같은 결과가 나타난다. 그림 4의 (b)는 (a)의 캡스트럼 파형을 2개의 임계치에 의해 3 단계로 양자화하여 나타낸 것이고, 그림 4의 (c)는 이렇게 결정된 동기 신호의 오차를 나타내고 있다. 시간 스케일 변형 공격 이전에는 그림에서처럼 매우 안정적인 동기 신호를 얻을 수 있다.

그림 5는 워터마크가 삽입된 오디오 신호의 길이를 단축시키는 시간 스케일 공격이 가해졌을 때, 그림 4와 같은 방법으로 나타낸 프레임 동기 신호이다. 5 번째 프레임인  $M=15$  부근에서부터 프레임 동기가 어긋나  $m=45$  에서는 완전히 반전된 결과를 나타내고 있는 것을 관찰할 수 있다. 즉, 그림 3의 (b)와 같은 현상이 발생하여 동기 신호가 수축된 것과 같이 나타나게 된다. 제안하는 복호화기에서는 그림 5의 (c)와 같은 오차 신호에 대해 표 1의 논리표를 바탕으로 동기가 어긋나는 곳에서 적절한 프레임의 시작점을 보정하여 동기를 맞추게 된다.

그림 6은 제안하는 방법에 의해 프레임 시작점을 보정하여 동기를 복원한 결과이다. 그림 5와 비교하여 오차 신호가 크게 줄었으며, 특히, 한 프레임 이상 어긋나 동기 신호의 위상이 반전되는 경우는 발생하지 않는 것을 확인할 수 있다.

시간 스케일 변형에 의해 길이가 늘어나는 신장의 경우도 마찬가지로 방법에 의해 보정할 수 있다.

#### 4. 실험 결과

표 2는 반향 워터마킹된 오디오 신호에 서로 다른 시간 스케일 변형 공격이 가해질 경우에 나타나는 검출 오차이다. 실험은 특성이 서로 다른 특성을 갖는 여섯 종류의 음악을 각각 30 초씩 연결하여 만든 180.0 초, 7938014 샘플의 오디오 신호를 사용하였으며, 시간 스케일 변형은 +/- 4% 이내의 값을 사용하였다. 검출 오차는 임의로 생성된 이진수 비트 열을 워터마크 신호로 삽입하고 이를 검출할 때의 Bit Error Rate (BER)로 정리하였다. 시간 스케일 변형 공격 전의 경우(No TSM)는 제안하는 프레임 동기화 방법의 적용 전/후 모두 0.16%의 오차를 나타냈다. 그러나, 시간 스케일 변형이 발생한 경우, 제안하는 알고리즘을 사용하기 이전에는 모든 실험 대상에서 전혀 복호화가 이뤄지지 않았지만, 제안하는 방법을 사용한 이후에는 모두 2% 미만의 오차만을 나타내었다. 특히 시간 스케일을 재생 과정 중간에 임의로 변화시키는 경우('Random')에도 1.01%의 낮은 오차율을 보여, 제안하는 방법이 시간 스케일 변형에 대한 프레임 동기화를 잘 수행하고 있음을 확인할 수 있다.

표 2의 마지막 열에는 제안하는 방법에 의해 시간축 길이를 보정하는 과정에서 발생할 수 있는 길이 추정 오차를 나타낸다. 즉, 시간 스케일 변형에 의해 변화된 시간축 길이를 제안하는 방법이 얼마나 잘 추정해내는 지를 나타내는 척도이다.

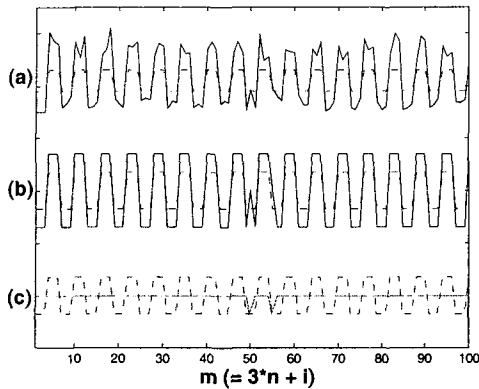


그림 4. 콤스트림 피크를 이용한 동기 신호의 생성 결과(시간 스케일 공격 전)

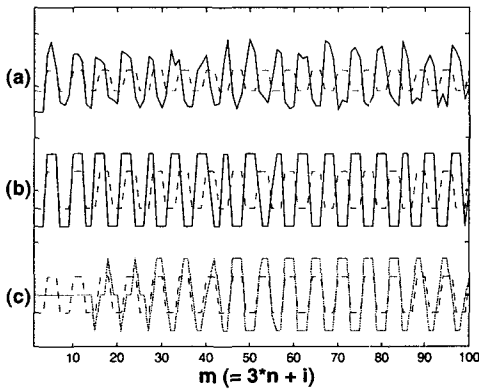


그림 5. 콤스트림 피크를 이용한 동기 신호의 생성 결과(시간 스케일 공격 -Shorten 후)

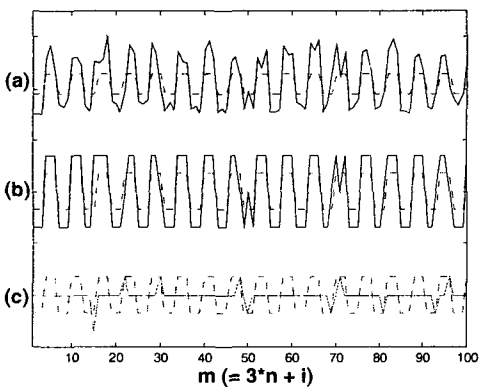


그림 6. 제안하는 방법에 의해 프레임 동기를 복원한 결과

표 2. 제안하는 방법 적용 전 후의 BER 비교

시간 스케일 변형 비율	BER (%)		동기 추정 오차 (+/-)
	적용 전	적용 후	
No TSM	0.16 %	0.16%	-
- 4%	X	0.31%	0.62 %
- 2%	X	0.23%	1.30 %
- 1%	X	0.23%	2.79 %
+ 1%	X	0.85%	0.62 %
+ 2%	X	0.85%	0.76 %
+ 4%	X	1.86%	3.57 %
Random	X	1.01%	0.57 %

## 5. 결론

본 논문에서는 기존의 워터마킹 방법에서 대응이 취약한 공격적인 시간 스케일 변형에 대한 해결책으로 반향 워터마킹의 반향 피크 위치를 이용한 프레임 동기화 방법을 제안하였다. 제안하는 방법은 기존의 반향 워터마킹과 비교하여 별도의 신호 삽입 없이 시간 스케일 변형된 정도를 검출하여 보상할 수 있기 때문에 음질 손상을 발생시키지 않는다. 시간 스케일 변형의 검출은 기존의 복호화 과정을 그대로 유지하면서 각 반향 커널 위치에서의 피크 값에 따라 간단한 논리 연산만으로 이루어지기 때문에, 추가적인 복호화기 연산량이 높지 않다. 실험을 통해 제안하는 방법 사용 전에는 전혀 복호화가 불가능하였으나, 제안하는 방법 사용 후에는 2% 미만의 매우 우수한 복호화 성능을 보이는 것을 확인하였다.

## 6. 참고문헌

- [1] W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding", *IBM System Journal*, Vol 35, Nos 3&4, 1996.
- [2] 오현오, 김현욱, 윤대회, 석종원, 홍진우, "강인한 오디오 워터마킹을 위한 새로운 반향 커널 설계," *한국음향학회지*, 제 20 권, 제 2 호, 2001년.
- [3] Hyen-O Oh, Jong-Won Seok, Jin-Woo Hong, Dae-Hee Youn, "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking," *Proc. of ICASSP 2001*, May 2001.
- [4] S. Roucos and A. Wilgus, "High Quality Time Scale Modification for Speech," *Proceedings of ICASSP 1985*, pp. 493-496, Mar. 1985.