

# 강인한 오디오 워터마킹을 위한 새로운 반향 커널 설계

## New Echo Embedding Technique for Robust Audio Watermarking

오 현 오\*, 김 현 옥\*, 윤 대 회\*, 석 종 원\*\*, 홍 진 우\*\*

(Hyen-O Oh\*, Hyun-Wook Kim\*, Dae-Hee Youn\*, Jong-Won Seok\*\*, Jin-Woo Hong\*\*)

\*연세대학교 전기전자공학과 음향, 음성, 신호처리 연구실,

\*\*한국전자통신연구원 무선, 방송 기술연구소

(접수일자: 2000년 10월 14일; 수정일자: 2001년 1월 5일; 채택일자: 2001년 2월 7일)

반향을 이용한 워터마킹에서는 오디오 신호에 청각적으로 지각되지 않는 작은 크기의 반향을 첨가함으로써 워터마크 정보를 은닉한다. 그러나 기존의 방법으로 반향을 삽입하게 되면 음질의 왜곡은 초래하지 않으면서 동시에 외부 공격에 대한 강인성을 보장하기에는 어려운 한계를 가지고 있다. 본 논문은 반향 (Echo) 신호에 대한 심리음향학적 분석을 통해 비지각성을 유지하면서 강인성을 높일 수 있는 새로운 반향 커널 설계 방법을 제안하고 이를 이용한 오디오 워터마킹 시스템의 성능을 평가한다. 제안한 반향 커널은 인간의 반향 지각 특성을 반영한 주파수 응답의 왜곡을 최소화하는 방향으로 설계되었기 때문에 기존의 방법에 비해 월등히 큰 에너지의 반향을 삽입할 수 있으며 따라서 공격에 강인하고 신뢰성 높은 복호화가 가능하다. 또한 몇 가지 신호처리 공격에 대한 분석을 통해 강인성을 보다 향상시키기 위한 기술들을 제안하고 적용하였으며 주관적 음질 평가 및 복호화 성능 평가를 통해 제안된 시스템의 성능을 검증하였다.

핵심용어: 오디오 워터마킹, 반향, 심리음향

투고분야: 음향 신호처리 분야 (1.1), 음악음향 및 음향심리 분야 (8.3, 8.7)

Conventional echo watermarking techniques often exhibit inherent trade-offs between imperceptibility and robustness. In this paper, a new echo embedding technique is proposed. The proposed method enables one to embed high energy echoes while the host audio quality is not deteriorated, so that it is robust to common signal processing modifications and resistant to tampering. It is possible due to echo kernels that are designed based on psychoacoustic analyses. In addition, we propose some novel techniques to improve robustness against signal processing attacks. Subjective and objective evaluations confirmed that the proposed method could improve the robustness without perceptible distortion.

**Key words:** Audio watermarking, Echo, Psychoacoustics

**Subject classification:** Acoustic signal processing (1.1), Musical acoustics and psychoacoustics (8.3, 8.7)

### I. 서 론

인터넷의 확산과 함께 MP3와 같은 고음질의 디지털 오디오 콘텐츠가 불법적으로 유통되면서 저작권 보호의 필요성이 급격히 대두되고 있다. 오디오 워터마킹은 이에 대한 해결책으로 인식되어 많은 관심을 갖는 새로운 연구 분야이다. 오디오 워터마킹이란 디지털 콘텐츠에 저작권 혹은 기타의 정보를 플러지 않도록 삽입하는 기술을 말한다.

오디오 워터마킹이 저작권 보호의 수단으로 사용되기 위해서는 다음과 같은 특징을 지니고 있어야 한다[1].

- ① 지각되지 않아야 한다.
- ② D/A-A/D 변환, 복사, 편집, 압축 등의 신호처리 과정에서 변질되지 않도록 강인해야 한다.
- ③ 의도적 공격에 대한 강인성을 지녀야 한다.
- ④ 원 신호 없이 복호화가 가능해야 한다.
- ⑤ 실시간 처리가 가능해야 한다.

이러한 특징 가운데 가장 중요한 워터마킹 개념의 본질은 워터마크가 삽입된 오디오 신호가 원음과 구별되지 않아야 한다는 점이다. 초기의 워터마킹 알고리즘들은 신호처리적인 공격이나 의도적인 공격을 고려하지 않았기 때문에 여기에 초점을 맞추어 개발되었다[1][2][3]. 그러나 워터마크가 저작권 보호의 목적으로 그 중요성이 높아지면서 일반적인 신호처리 과정에 대한 강인성 (Robustness)도 매우 중요한 요구사항이 되었다. 워터마킹 알고리즘의 강인성을 높이기 위해서는 보다 큰 에너지의 워터마크가

책임저자: 오현오 (oho@cyclon.yonsei.ac.kr)  
120-749 서울시 서대문구 신촌동 134 연세대학교 전기전자공학과  
음향, 음성, 신호처리연구실  
(전화: 02-2123-2863; 팩스: 02-312-4584)

삽입되어야 하며 특히 MP3와 같은 고음질 압축 과정 후에도 워터마크가 사라지지 않도록 하기 위해서는 지각적으로 중요한 부분에 워터마크가 은닉되어야 한다[4]. 그러나 이것은 지각되지 않아야 한다는 워터마크의 본질적 특징과 서로 상충되는 요구 조건이다. 결국 현재의 모든 오디오 워터마킹 방법에 있어서 비지각성 (Inaudibility)과 강인성 사이는 서로 대립적인 관계에 놓여있다. 즉, 비지각성을 유지하기 위해서는 강인성이 떨어지고, 강인성을 높이기 위해서는 그만큼 음질의 열화를 감수해야만 한다. 따라서 이 두 성질간의 만족할만한 타협점을 얻는 방법이 디지털 오디오 워터마킹 시스템 성능의 중요한 지표가 된다.

일반적으로 사람은 시각 정보보다 청각 정보에 보다 민감하기 때문에 오디오 신호의 손상이 영상 신호의 손상보다 더 쉽게 지각된다. 또한 오디오 신호는 영상 신호에 비해 데이터의 양이 적기 때문에 정보를 감출 수 있는 영역도 상대적으로 부족하다. 그러므로 영상 신호의 워터마킹 기술을 바로 오디오 신호에 적용할 경우 음질이 손상되기 쉬우며 강인성도 떨어질 수 있다. 따라서 오디오 신호 고유의 특성을 이용한 워터마킹 방법의 개발이 필요하다. 지금까지 연구된 대표적인 오디오 워터마킹 방법으로는 LSB 부호화 방법, 위상 정보를 변형하는 방법, 대역 확산을 이용한 방법, 반향 (Echo) 신호를 이용하는 방법 등이 있다[2][3][5]. LSB 부호화[3]는 양자화된 오디오 샘플의 최하위 비트들을 변형하여 원하는 정보를 삽입하는 방법인데, 삽입과 검출 과정이 간단하고 음질 왜곡이 적은 장점이 있지만, 강인성에 대한 고려가 전혀 이루어지지 않았기 때문에 저작권 보호를 위한 워터마킹으로는 사용하기 어렵다. 위상을 이용한 방법[3]에서는 사람의 청각 시스템이 오디오 신호의 절대 위상을 잘 구별하지 못한다는 특성을 이용하여, 초기 위상 정보를 변형하여 원하는 정보를 삽입하고 이를 검출한다. 그러나 이 방법은 음질을 보장하기 위해서 삽입할 수 있는 정보량을 제한해야 하며, 역시 압축과 같은 신호처리 과정에서 위상 정보가 쉽게 변형되기 때문에 공격에 대한 강인성이 떨어진다. 대역 확산을 이용한 방법[3][5]은 이미지 워터마킹에서부터 널리 사용되어 오디오 워터마킹에서도 현재 가장 많이 연구되고 있는 방법으로, 은닉하고자 하는 이진 데이터 열을 PN (Pseudo Noise) 시퀀스로 대역 확산하여 오디오 신호에 삽입한다. 삽입된 워터마크는 PN 시퀀스의 높은 자기 상관 성질을 이용하여 검출할 수 있으며 간섭에 강한 특징을 갖는다. 또한, 심리 음향 모델을 이용하여 PN 시퀀스를 변형하면 지각되지 않는 큰 에너지의 워터마크 삽입이 가능하여 강인성을 높일 수 있다 [6][7][8]. 그러나 대역 확산을 이용한 워터마킹은 원 신호 없이 검출하기 위한 검출 알고리즘의 개발에 어려움이 따르며, 선형 속도 변화 (Linear Speed Change)와 같은 시간 축 길이 변화형의 공격에 약한 치명적 단점을 가지고 있다. 만일 고음질 오디오 부호화 기술의 발전으로 지금보다 효율적인 압축 방법이 등장한다면, PN 시퀀스를 이용하여 삽입된 워터마크 신호는 결국 처리 과정에서

제거될 수도 있다. 또한, 고음질을 추구하는 청취자들은 비록 삽입된 PN 시퀀스를 듣지 못한다고 하여도 오디오 신호에 잡음을 첨가하는 것 자체를 선호하지 않는 경향이 있다.

반향 (Echo)을 이용한 워터마킹 기법에서는 귀에 들리지 않을 만큼 작은 반향을 오디오 신호에 첨가함으로써 워터마크 정보를 삽입하게 된다[3]. 즉 오디오 신호를 프레임 별로 나누고 각 프레임에 삽입하고자 하는 이진 워터마크에 따라 다른 반향을 삽입함으로써 부호화하고, 복호화 과정에서는 각 프레임 별로 삽입된 반향의 오프셋 (시간 지연)을 검출 함으로써 이진 정보를 복호화하게 된다. 이 경우 첨가되는 신호는 잡음이 아니라 원 신호와 같은 특성을 갖고 있는 오디오 신호 자체이기 때문에 사람에게 왜곡으로 인지되지 않으며 오히려 음색을 풍부하게 해주는 효과를 기대할 수 있기 때문에 고음질 오디오 워터마킹에 사용하기 적합하다. 또한 이렇게 삽입된 워터마크는 원 신호 없이도 복호화가 가능하다는 특성을 갖는다. 그러나 기존의 반향을 이용한 워터마킹에서도 워터마크의 강인성을 향상시키기 위하여 충분히 큰 반향 신호를 삽입하게 되면 오디오 신호에 왜곡을 초래하게 되며 결국 비지각성과 강인성을 동시에 만족시키기 어려운 문제점을 가지고 있다. 따라서 반향 기법을 워터마킹에 사용할 수 있기 위해서는 지각되지 않는 큰 반향 신호를 삽입할 수 있는 기술의 개발이 필요하다.

본 논문에서는 반향에 대한 심리음향학적 분석을 기반으로 원음은 손상시키지 않으면서 기존의 방법보다 강인성을 크게 향상시킨 새로운 반향 삽입 방법을 제안한다. 제안한 워터마킹 알고리즘은 음질 왜곡 없이 상대적으로 큰 에너지의 반향을 삽입할 수 있기 때문에 검출 과정의 신뢰성과 공격에 대한 강인성의 향상을 기대할 수 있다. 이를 위해 본 논문에서는 먼저 반향 신호에 대한 심리음향학적 분석을 수행하였다. 인간의 반향 신호 지각 특성을 반영하여 반향 삽입이 음질에 미치는 영향을 주파수 영역에서 분석한다. 지각 특성을 반영한 주파수 응답 왜곡이 최소화되는 방향으로 반향을 설계할 수 있는 방법을 제시하고 이를 통해 비지각성은 유지하면서 기존의 방법보다 두 배 큰 에너지로 반향을 삽입할 수 있는 새로운 반향 커널 (반향 삽입에 사용되는 시스템 충격 응답 함수, [3])을 제안한다. 또한 신호처리 공격이 오디오 워터마크에 미치는 영향에 대한 분석을 통해 이에 대응하여 강인성을 향상시킬 수 있는 몇 가지 새로운 기술도 제안하였다. 제안된 시스템에 대한 성능 검증을 위하여 기존의 반향 삽입 방법과 비교하여 주관적 음질 평가를 수행하였으며, 반향 커널의 강인성을 정량화 할 수 있는 평가 방법을 제안하여 성능을 평가하였다

## II. 반향을 이용한 워터마킹

반향은 적절한 흡음계수를 갖는 벽면으로 둘러 쌓인 공간에서 수음할 때 직접음 이외의 반사음들의 복합적인 작용에 의해 형성되며 그림 1과 같은 충격 응답을 갖는

시스템을 통과하는 것으로 모델링 할 수 있다. 이 때 충격 응답 함수를 반향 커널이라 부르며 커널의 특성을 결정하는 요소는 초기 진폭 (Initial amplitude), 감쇄율 (Decay rate), 오프셋 (Offset), 델타 (Delta) 등이 있다[3]. 초기 진폭은 첫번째 반향이 갖는 크기이고 감쇄율이란 두 번째 반향 신호부터 적용되는 초기 진폭으로부터의 상대적인 크기의 비율을 나타낸다. 오프셋은 원래 신호로부터 첫번째 반향까지의 시간 지연을 의미하고 이후의 감쇄되는 반향 신호 사이의 간격은 델타로 정의한다

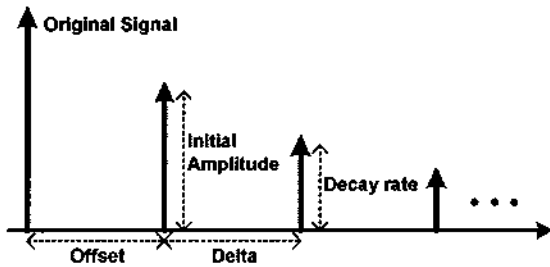


그림 1. 반향 신호의 충격 응답  
Fig. 1. Impulse response function of echo signal.

2.1. 부호화

그림 2는 반향 위터마킹의 부호화 과정을 나타내는 블록도이다. 반향을 위터마킹 정보의 삽입에 사용하기 위해서는 우선 이진수 '0'인 경우와 이진수 '1'인 경우에 각각 다른 오프셋을 갖는 반향 커널을 생성해 둔다. 입력 오디오 신호를 프레임별로 나누고 각 프레임에 할당된 위터마크 정보에 따라 반향 커널을 결정하여 오디오 신호와 컨볼루션(Convolution)하여 반향을 삽입할 수 있다. 이때, 각 프레임 사이의 반향 오프셋의 급격한 변화는 시간 축에서 불연속점을 생성하여 음질에 영향을 미칠 수 있기 때문에 전후 프레임이 적절히 겹쳐서 이어지는 중첩 구간이 있어야 한다[3].

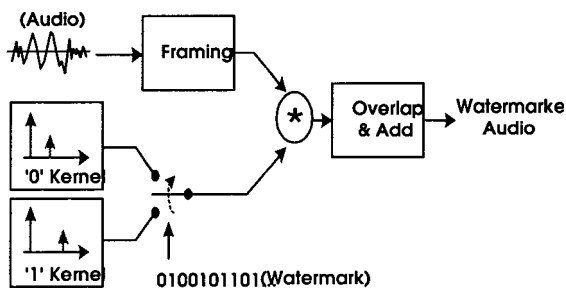


그림 2. 부호화 과정  
Fig. 2. Encoding Process.

2.2. 복호화

그림 3은 반향 위터마킹의 복호화 과정을 나타내는 블록도이다. 반향을 이용한 위터마킹에서 삽입된 정보는 두 개의 다른 오프셋을 갖는 반향 신호이므로 위터마킹 정보를 복호화하는 것은 곧 각 프레임에서의 반향 신호의 오프셋 값을 찾아내는 것이다. 이는 자기 상관 (autocorrelation) 방법에 의해 분해할 수 없을 정도로 작은 시간 지연을 갖는 신호에서 그 지연된 정도를 찾아내는 문제라 생각할 수 있다. 이 문제를 해결할 수 있는 좋은 방법은 캡스트럼 (Cepstrum)을 이용하는 것이다[3][9]. 반향이 삽입된 오디오 신호는 오프셋의 배수 위치에 피크가 나타나는 캡스트럼을 갖는다. 따라서 캡스트럼의 자기 상관을 구하여 그 진폭으로부터 삽입된 반향 오프셋을 찾아낼 수 있다. 캡스트럼의 자기상관함수 연산을 '오토캡스트럼 (Autocepstrum)' 이라고 정의하며 자기 상관과 파워 스펙트럼 사이의 관계에 의해 입력신호  $x_m(n)$ 에 대해 다음 식과 같이 계산된다[3].

$$\tilde{x}_m(n) = F^{-1}((\log(F(x_m(n))))^2)$$

식에서  $F$ 는 short-time Fourier Transform을,  $F^{-1}$ 는 그 역 변환을 의미하며  $m$ 은 프레임을 나타내는 변수이다.



그림 3. 복호화 과정  
Fig. 3. Decoding Process.

III. 반향의 심리음향학적 분석

오디오 신호에서의 반향은 같은 물리적 현상으로부터 발생한 것임에도 불구하고 사람의 귀에는 두 가지 다른 효과를 나타내며 지각된다. 한 가지는 흔히 메아리라고 일컫는 반향 (협의의 반향) 효과이고 다른 하나는 착색 (coloration) 효과이다. 이 현상은 귀의 응답 특성을 고려하여 단 반향 (하나의 반향 펄스만을 갖는 반향)에 대한 충격 응답과 주파수 응답을 해석함으로써 설명될 수 있다.

단 반향은 그림 4와 같이 한쪽 벽면만 존재하는 환경에서 소리가 발생할 때 마이크로폰으로 수음하는 경우에 발생한다. 이때 반향 경로를 모델링한 시간 축에서의 충격 응답과 이의 주파수 응답은 그림 5와 같은 형태로 나타난다. 마이크로폰까지 도달하는 과정은 두 개의 경로를 가지고 있기 때문에 충격 응답은 두 개의 펄스만을 갖는 형태로 나타나고, 이에 대한 주파수 응답은 전 주파수 대역에서 그림과 같이 변화하는 형태를 나타내게 된다. 이때 두 펄스 사이의 간격이 멀어질수록 주파수 응답이 나타내는 리플 (ripple)의 변화율은 높아지게 된다. 만약 두

펄스 사이의 간격 (오프셋)이 50ms 정도로 충분한 길이를 갖게 되면 명확한 반향이 들리고, 2ms 정도로 짧은 경우에는 반향 효과는 발생하지 않고 원음의 음색이 변화된 형태로 들리게 되는 착색 효과가 발생하게 된다[10]. 이러한 차이는 귀의 신호 전달 특성에 의해 결정된다.

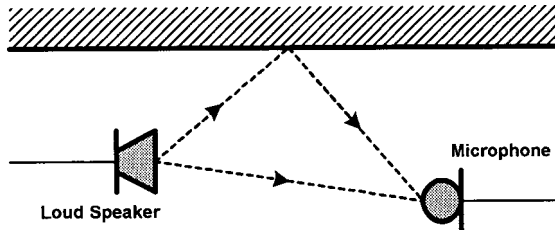


그림 4. 단 반향이 발생하는 환경  
Fig. 4. Sound reflected from a single wall results in 'single echo'.

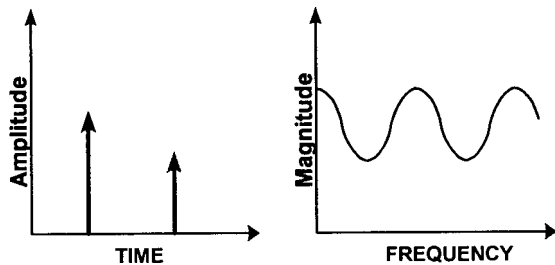


그림 5. 단 반향의 충격 응답과 주파수 응답.  
Fig. 5. Impulse Response and frequency response of single echo.

사람의 귀는 소리를 전기적 신호로 변환하여 뇌로 전달하는 역할을 담당하는 달팽이관의 작용 특성으로 인해 임계 대역 (critical band)이라고 알려진 필터 बैं크의 형태로 소리를 인지하게 된다[11]. 이때 각 임계 대역의 신호를 받아들이는 필터들은 5-20ms 정도의 길이를 갖는 메모리를 가지고 있어서 일정 시간만큼의 입력 신호에 대해 한 번에 처리하게 되며 각 대역내에서는 상세한 주파수 형태보다는 에너지의 합으로 소리를 구별하는 특성을 갖는다. 반향 신호를 나타내는 펄스 사이의 간격이 좁아서 인간 귀가 갖는 필터의 메모리 안에 들어오는 경우는 이 두 펄스의 상호 작용에 의한 주파수 응답이 음색에 영향을 주지만 지연된 시간 간격을 분해하지 못하기 때문에 반향으로 들리지는 않는다. 반면 20ms 이상의 긴 오프셋을 갖는 충격 응답의 경우는 귀가 갖는 필터의 메모리 길이보다 펄스 사이의 간격이 길기 때문에 주파수 응답에 의한 영향은 발생하지 않고 대신 시간 영역에서 지연된 신호를 다시 듣게 되는 자연스런 반향이 인지된다.

심리음향학에서 연구되는 순시 마스킹 (Temporal Masking)에 의하면 시간 영역에서 레벨이 큰 신호가 발생한 직후 혹은 직전에 존재하는 작은 신호는 귀에 들리지 않는다 [11]. 이때 마스킹되는 영역은 마스커의 음압, 지속시간 등에 따라 다르며 최대 200ms까지 지수적으로 감소하는

것으로 알려져 있다. 반향을 위해 첨가되는 신호는 원 신호와 정확히 같은 주파수 분포를 갖는 신호이므로 순시 마스킹에 의해 가장 잘 마스킹이 되는 신호라고 볼 수 있다. 순시 마스킹의 이득을 보기 위해서는 반향 커널의 오프셋은 작을수록 좋으며 따라서 착색 효과를 일으키는 영역에 존재해야 한다. 그러나 너무 작은 오프셋을 사용하는 경우는 음색 변화에 의해 소리가 둔탁해지는 현상이 나타나기 때문에 좋지 않으며, 복호화 과정에서 검출하는데 어려움이 따를 수 있다. 따라서 최적의 오프셋을 찾는 과정은 여러 종류의 음원에 대해 많은 실험을 통해 결정되어야 한다

그림 6은 본 논문에서 단 반향에 대해 수 차례의 주관적 음질 평가 실험을 통해 얻은 반향 커널의 오프셋과 초기 진폭에 따른 음질 상관 관계를 나타내는 개념도이다. 음질 평가 실험은 6.1절에 설명된 방법과 같이 실시되었으며, 첫 번째 평가는 방송 음향을 전공하는 22명의 20대 남성 대학생을 대상으로 수행하였고, 두 번째 평가는 오디오 및 음성 신호처리를 전공하는 대학원생 10명을 대상으로 수행하였다. 마지막으로 오디오 및 음악 평론가 3명에 의한 최종 평가를 거쳐 결과를 검증하였다. 이렇게 얻은 *offset\_min*과 *offset\_max*의 실험치는 각각 0.9 msec, 3.4msec이며 *Thr\_a*는 0.3이하이고 *Thr\_b*는 0.45이다. 즉, 음원의 특성에 따라 반향의 지각 정도에는 약간의 차이가 있지만, 0.9~3.4msec의 오프셋을 갖고 초기 진폭이 0.3이하인 커널에 의해 삽입된 반향은 그 오프셋에 관계없이 청취자의 귀에 지각되지 않는다는 실험 결과를 의미한다. 이 결과를 바탕으로, 워터마킹 시스템에서 이진수에 대응하는 두 개의 커널만을 사용하지 않고 제안하는 범위 내에 존재하는 오프셋을 갖는 여러 종류의 커널을 사용한다면 한 프레임에 여러 비트를 삽입하여 비트율을 높이거나 하나의 이진 정보에 대해 오프셋이 다른 커널들을 사용함으로써 삽입된 워터마크의 암호화를 도모할 수 있다.

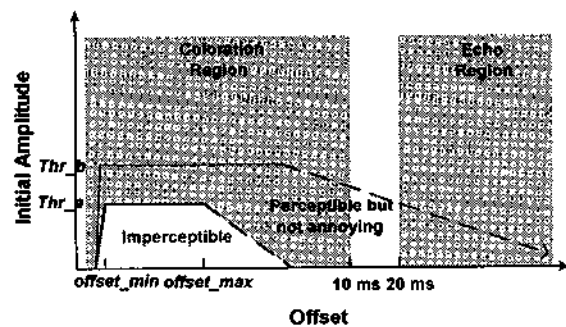


그림 6. 실험을 통해 얻은 반향 커널 파라미터와 음질 상관 관계도

Fig. 6. Relationship between echo kernel parameters and quality.

### IV. 새로운 반향 커널 설계

#### 4.1. 임계 대역을 주파수 응답 분석

위터마킹에 사용되는 반향 커널은 착색 효과를 나타내는 영역에 존재하기 때문에, 반향 삽입에 의한 오디오 신호의 왜곡을 분석하기 위해서는 반향 커널의 주파수 응답을 이용하는 것이 적합하다.

그림 7의 (a)는 '양(+)'의 단 반향 (positive single echo)을 나타내는 커널이고 그림 8은 이에 대한 주파수 크기 응답을 나타낸다. 이때 '양(+)'은 사용된 커널의 초기 진폭의 부호를 의미한다. 그림 8의 주파수 응답의 가로축은 사람의 귀가 소리를 인지하는 주파수 분해능에 근거한 임계 대역율 (critical band rate)을 이용하여 나타내고 있다.

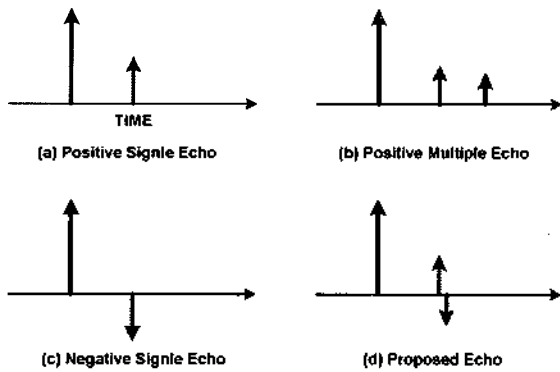


그림 7. 여러 가지 반향 커널의 충격 응답  
Fig. 7. Frequency response of positive single echo.

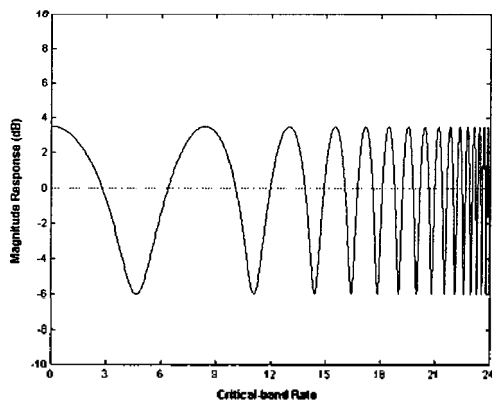


그림 8. 양의 단 반향 커널의 주파수 응답  
Fig. 8. Frequency response of positive single echo.

임계 대역율을 사용하여 주파수 응답을 나타내면 선형 주파수축에 의해 나타내는 것보다 착색 효과가 음질에 영향을 미치는 왜곡 정도를 보다 잘 묘사할 수 있다. 한 임계 대역내에서는 응답 곡선이 심하게 변화하여도 그것을

분해할만한 주파수 해상력을 갖지 못하고 그 평균적인 크기에 의해 소리를 인지하게 되므로 그림에서 높은 주파수 영역보다는 낮은 주파수 영역에서의 응답 특성이 보다 중요하다. 양의 단 반향에 대한 응답은 낮은 주파수 대역에서 각 임계 대역의 이득이 상대적으로 크게 변화하고 있기 때문에 이 대역에서의 음색 변화, 즉 왜곡이 크다. 특히 0~2bark, 즉 0~200Hz에 해당하는 대역은 응답 이득이 0dB 이상이므로 저음부의 증폭 효과를 가져오며 결국 양의 단 반향을 사용하면 음색을 풍부하게 하고 공명을 증가시키는 듯한 효과를 일으키게 된다.

반향 펄스가 하나만 존재하는 경우를 단 반향이라고 할 때, 여러 개의 반향 펄스를 사용하는 경우는 다중 반향 (multiple echo)이라 부른다. C. X. [12]는 크기가 작은 여러 개의 반향 신호를 사용하여 검출 성능은 하나의 큰 반향을 사용하는 경우와 같은 수준을 유지하면서 제 3자에게 반향 삽입 위치를 노출시키지 않을 수 있는 다중 반향에 의한 위터마킹 방법을 제안하였다. 그러나 다중 반향을 사용하더라도 각 반향 펄스 크기의 합에 비례하여 음색 변화가 발생하기 때문에 단 반향을 사용하는 경우보다 더 큰 반향 신호를 삽입할 수는 없다. 그림 7의 (b)는 두 개의 양의 반향 펄스가 존재하는 다중 반향 커널의 한 예를 나타내며 그림 9는 이와 같은 반향 커널에 대한 주파수 응답을 임계 대역율에 의해 나타내고 있다. 임계 대역 0~12bark까지는 대략 양의 단 반향 커널의 주파수 응답과 유사한 형태를 가지며 23bark 부근에서 리플이 가장 작다가 다시 커지는 오목한 형태의 응답 특성을 관찰할 수 있다. 음질과 많은 상관성을 갖는 낮은 주파수 영역의 응답이 단 반향의 응답과 유사하기 때문에 같은 크기의 다중 반향을 사용하는 것은 단 반향과 비슷한 정도의 음질 왜곡을 일으킨다. 또한 청취 실험 결과 지적될 수 있는 정도로 반향의 크기를 증가시키면 이때의 음색은 단 반향의 경우처럼 듣기 좋은 방향으로 변화하는 것이 아니라 약간 거슬리는 음질 왜곡으로 느껴지는 것으로 나타났다.

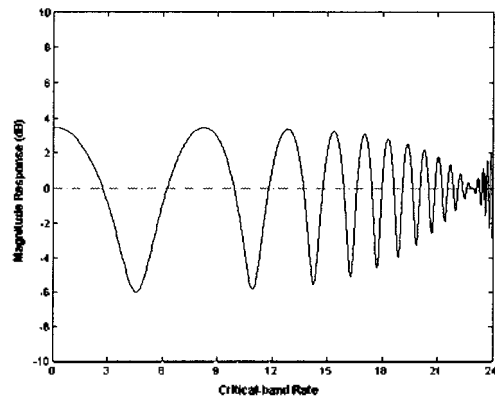


그림 9. 양의 다중 반향 커널의 주파수 응답  
Fig. 9. Frequency response of positive multiple echo.

4.2. 음의 반향을 이용한 새로운 반향 커널 제안

반향을 이용한 워터마킹에서 비직각성과 강인성을 결정짓는 중요한 요소는 커널의 오프셋과 초기 진폭이다. 오프셋은 그림 6에서 나타낸 범위 안에 존재하도록 위치시키면 지각되지 않는 반향을 삽입할 수 있으며, 복호화 성능도 이 영역 안에서 크게 달라지지 않는다. 강인성의 향상을 위해서는 결국 초기 진폭을 가급적 크게 하여 삽입하여야 하는데, 이는 워터마크의 비직각성 유지와 정면으로 상충되게 된다. 또한 큰 반향은 의도적으로 워터마크를 제거하려는 공격자에게 쉽게 노출되는 단점도 가지고 있다.

본 논문에서는 기존의 반향과 같이 초기 진폭이 항상 양의 값을 갖는 경우와는 달리 음의 값을 갖는 반향을 고려하여 임계 대역울축에서의 주파수 응답이 편평한 전달 특성을 얻을 수 있는 새로운 반향 커널을 제안한다. 먼저 음의 반향의 전달 특성을 알아보기 위해 그림 7의 (c)와 같은 커널을 설계하였다. 이때의 주파수 응답은 그림 10과 같다. 오프셋과 초기 진폭의 절대값은 그림 8의 양의 단 반향 커널에서와 같은 값을 갖는 경우이다. 음의 반향의 주파수 응답은 양의 반향일 때와 같이 전대역에서 주기적인 리플이 있는 형태로 나타나지만 주파수 0Hz (DC)에서의 이득이 양의 반향일 때와는 반대로 최소값에서 출발한다. 그림을 보면 양의 반향일 때 0~2bark (200Hz 미만)의 낮은 주파수 대역의 응답이 0dB 이상으로 크기를 증폭시키는 효과를 일으키는 반면 음의 반향일 때는 이 대역의 크기를 감소시키고 있는 것을 관찰할 수 있다. 청취 실험 결과 낮은 주파수 영역에서의 반전된 주파수 응답 특성으로 인해 양의 반향을 사용하면 저음부가 자연스럽게 증폭되어 소리에 풍부한 느낌을 주고 음의 반향을 사용하면 저음부가 짝여 소리를 가볍게 하는 것으로 나타났다. 일반적으로 사람들은 풍부한 음색을 더 선호하기 때문에 같은 크기의 반향이라면 음의 반향보다 양의 반향을 삽입한 소리를 더 선호하는 경향이 있다. 그러나 그림 6의 *Thr-a*이하로 초기 진폭을 작게 하는 경우는 음의 반향도 양의 반향과 마찬가지로 귀에 인지되지 않는 투명한 워터마크를 삽입할 수 있다. 또한, 음의 반향을 사용하는 경우에도 양의 반향인 경우와 마찬가지로 오토겟스트림 방법에 의해 오프셋 위치에서 피크를 검출할 수 있다. 즉, 복호화 관점에서 볼 때 반향 펄스의 부호는 성능에 영향을 미치지 않으며 오직 초기 진폭의 절대값에 비해 해서 검출 성능이 좋아진다. 이때, 오토겟스트림의 출력은 모두 양수로 나타나기 때문에 복호화 시에는 삽입된 반향 커널의 부호를 알 수 없다. 이는 삽입된 반향을 알아내려는 의도적 공격자에게 커널의 부호를 노출시키지 않을 수 있기 때문에 좋은 특성이다.

그림 8의 양의 반향에 대한 주파수 응답과 그림 10의 음의 반향에 대한 주파수 응답은 0dB를 기준으로 대칭적인 응답 특성을 가지고 있기 때문에, 오디오 신호가 이 두 커널을 동시에 거치게 되면 각 주파수별 이득이 서로 상쇄되어 원음과 완전히 같은 신호를 얻을 수 있을 것이다.

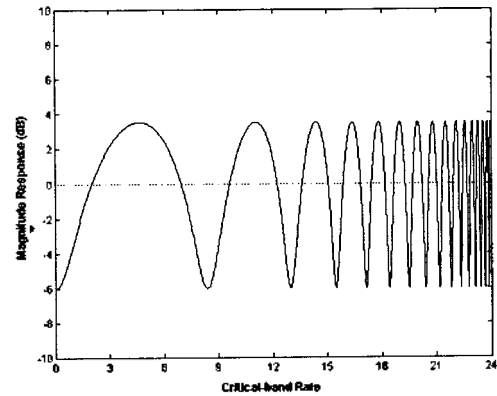


그림 10. 음의 단 반향 커널의 주파수 응답  
Fig. 10. Frequency response of negative single echo.

그러나 이렇게 하는 것은 결국 아무런 반향도 삽입하지 않은 것과 같다. 즉, 반향 커널에서 양의 펄스와 음의 펄스의 오프셋이 같아지므로 상쇄되어 초기 진폭이 0이 되어 결국 삽입되는 워터마크의 에너지가 0이 된다. 그러나 양의 펄스와 음의 펄스의 오프셋이 근소한 차이를 두고 연속된 경우에는 서로 상쇄되지 않기 때문에 워터마크로써 검출이 가능하고, 이때의 주파수 응답은 서로 정확히 대칭되지는 않지만, 많은 부분의 리플이 반대 위상을 가질 수 있게 된다. 이는 인접한 주파수를 갖는 두 개의 정현파가 시간 축에서 맥놀이 (beat) 현상을 일으키는 것과 비슷한 이치로, 주파수 축에서의 리플이 서로 상쇄되는 구간이 생기게 되는 것이다. 본 논문에서는 이러한 현상 즉, 인근 오프셋을 갖는 양의 반향과 음의 반향의 조합에 의해 특히 낮은 주파수 영역에서 리플이 상쇄되어 편평한 특성을 갖는 새로운 반향 커널을 제안한다. 그림 7의 (d)는 본 논문에서 제안하는 음의 반향을 포함한 새로운 다중 반향 커널의 충격 응답을 예시하며, 그림 11은 이에 대한 주파수 응답을 임계 대역을 축으로 나타내고 있다.

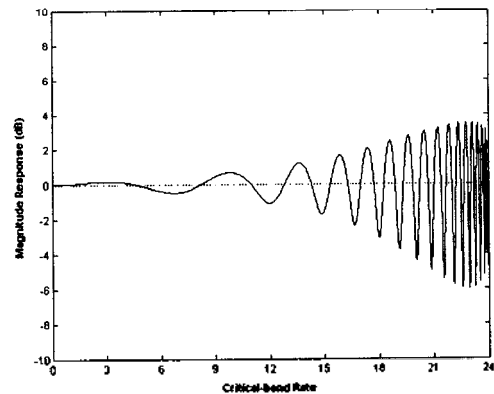


그림 11. 제안된 반향 커널의 주파수 응답  
Fig. 11. Frequency response of proposed echo kernel.

그림 11을 살펴보면 양의 반향에 대한 응답과 음의 반향에 대한 응답이 서로 상쇄효과를 일으키고 있는 0-2bark 까지의 대역은 거의 0dB에 가까운 투명한 전달 특성을 나타내며 이후의 임계 대역에서도 점차 증가하기는 하지만, 그림 8의 양의 반향의 경우와 비교하면 많은 임계 대역에서 월등히 낮은 리플이 나타나는 것을 확인할 수 있다. 이러한 응답 특성으로 인해 제안하는 반향 커널을 사용하면 같은 에너지를 갖는 단 반향을 사용하는 경우보다 음색의 변화, 즉 음질의 왜곡을 현저하게 줄일 수 있다. 이로 인해 단 반향 혹은 다중 반향의 경우에 비해 더 큰 에너지를 갖는 반향을 삽입하고도 지각되지 않는 신호를 생성할 수 있으며 결국 왜곡을 증가시키지 않으면서 강인성을 향상시킬 수 있게 된다.

## V. 강인성 향상을 위한 알고리즘

오디오 워터마킹이 저작권 보호와 같은 목적으로 활용될 수 있기 위해서는 오디오 신호가 일반적으로 겪을 수 있는 여러 가지 신호처리 과정 이후에도 항구적으로 복호화가 가능하여야 한다. 본 장에서는 반향 워터마킹 시스템의 성능에 영향을 미칠 수 있는 몇 가지 신호처리 과정들에 대한 문제점을 분석하고 이를 해결할 수 있는 알고리즘을 제안한다.

### 5.1. 듀얼 커널

워터마킹된 오디오 신호에 반향이 첨가될 경우, 반향의 시간 지연이 유연히 워터마킹 시스템의 프레임 길이와 일치하게 되면 복호화 성능을 떨어뜨리는 요소로 작용할 수 있다. 즉, 이전 프레임의 신호가 반향 효과에 의해서 현재 프레임에 간섭을 일으켜 검출 과정의 오차로 작용하게 되는 것이다. 이는 통신 채널 상에서 발생하는 문제 중 하나인 ISI (Inter Symbol Interference)와 유사한 유형의 간섭으로 해석할 수 있다.

이와 같은 문제의 해결을 위해 본 논문에서는 이웃 프레임들 사이에 서로 다른 반향 오프셋을 사용하여 검출 위치를 직교화하는 듀얼 커널 (Dual Kernel) 방식을 제안하고 적용하였다. 즉, 이전 프레임과 현재 프레임에 사용하는 반향 커널의 오프셋을 다르게 함으로써 이전 프레임 신호가 현재 프레임으로 스며들어 간섭을 일으키더라도 이 신호가 현재 프레임의 복호화를 위한 커널과 직교적인 위치에 존재하여 방해받지 않고 복호화할 수 있게 된다. 듀얼 커널 방식을 사용하면 결과적으로 네 개의 서로 다른 오프셋을 갖는 반향 커널이 필요하게 된다.

### 5.2. 시작점 동기화

반향 워터마킹은 삽입되는 신호가 오디오 신호 자체가 기 때문에 MP3 부호화와 같은 고음질 압축 과정에 의한 성능 저하가 다른 워터마킹 방식에 비해 적은 편이다. 그러나, MP3 압축 과정이나 D/A, A/D, 필터링과 같은 신호처리 과정이 워터마킹에 미치는 또 다른 문제 중의 하나는 이런 처리 과정이 갖는 시간 지연으로 인해, 출력으로

나오는 오디오 신호는 처리 전과는 다른 시간 축 인덱스를 갖게 된다는 것이다. 즉, 워터마크 부호화 과정에서 사용된 프레임 시작점과 신호처리 과정을 거친 다음의 신호를 복호화할 때 사용할 프레임의 시작점이 일치하지 않을 수 있으며 이는 워터마크 검출에 치명적인 오류를 일으킬 수 있다.

본 논문에서는 오디오 신호의 시간 축에서의 크기 정보를 이용하여 간단하면서도 안정적으로 시작점 위치를 동기화할 수 있는 방법을 제안한다. 입력 오디오 신호가 미리 설정해둔 임계값 이상으로 일정구간 들어오면 그 시점부터 부호화를 시작하고, 복호화 과정에서도 같은 방법으로 시작점을 찾아냄으로써 부호화 과정과 같은 시간 인덱스에서 프레임을 시작할 수 있도록 동기를 맞추는 방법이다. 제안한 방법은 압축, 잡음 첨가, 필터링과 같은 신호처리 과정 이후에도 +/-50샘플 미만의 오차를 가지고 시작점을 찾을 수 있었으며, 복잡성이 매우 낮아 실시간 구현에 용이한 장점이 있다.

### 5.3. 파일럿 프레임

일반적으로 오디오 워터마킹 알고리즘이 가지고 있는 가장 큰 문제점으로 시간축 길이가 변화하는 신호처리 공격에 대한 변역성이 떨어진다는 것이 있다. 시간축 길이 변화의 한 예로 선형 속도 변화 (linear speed change)가 있는데, 가령 100초 길이의 오디오 신호가 90초에 재생되거나 110초에 재생되는 경우를 말한다. 이렇게 되면 프레임 길이도 이에 비례해서 달라지기 때문에 속도 변화된 정도를 찾아내고 보정해주지 않으면 프레임이 진행되면서 동기가 어긋나게 되어 검출이 전혀 불가능하게 된다. 반향 워터마킹의 경우는 선형 속도 변화에 의해 커널의 오프셋도 달라지게 되며 따라서 변화된 정도를 반영해주지 않으면 프레임 동기를 맞출 수 있다고 하더라도 영통한 값을 복호화하게 된다.

본 논문에서는 선형 속도 변화가 삽입된 반향의 오프셋을 변화시킨다는 점에 근거하여 파일럿 프레임을 이용하여 복호화 과정에서 선형 속도 변화된 정도를 검출해내고 이를 보정해줄 수 있는 알고리즘을 제안하고 적용하였다. 파일럿 프레임이란 워터마크 삽입에 사용되는 커널과는 다른 오프셋을 갖는 파일럿 커널에 의해 반향이 삽입된 프레임을 말하며, 부호화 과정에서 첫 번째 프레임을 비롯하여 일정 간격을 두고 별도로 삽입시키는 프레임이다. 복호화 과정에서는 파일럿 프레임의 오토코러스 트럼을 구해 그 피크의 위치와 알고 있는 파일럿 커널의 오프셋을 비교하여 속도 변화된 비율을 찾아낼 수 있다. 선형 속도 변화가 검출되면 그 비율에 맞게 프레임 크기와 각 커널들의 오프셋 값을 알맞게 보정한 후 이후 프레임에 대해 복호화를 수행하고 변화가 없는 경우는 이전에 사용한 값에 의해 복호화를 수행하게 된다. 선형 속도 변화를 올바르게 검출할 확률은 속도 변화가 발생하기 이전에 시스템이 갖는 복호화 성능과 일치하며 따라서 속도 변화 공격 이전에 100% 복호화가 가능한 경우라면 속도 변화를 100% 검출할 수 있다.

VI. 실험 및 결과

제안된 반향 커널을 이용한 워터마킹 기법의 성능을 평가하기 위해 단 반향 및 양의 다중 반향 방법과 비교하여 주관적 음질 평가와 복호화 성능 평가를 수행하였다. 비교 실험에 사용된 다섯 가지 반향 삽입 방법은 표 1에 정리하였다. 표에서 스케일팩터란 반향 펄스 크기의 총 합을 말하며,  $\alpha$ 는 방법 I에 의해 워터마크 삽입된 오디오 신호가 원 신호와 청각적으로 구별되지 않는 최대값이 되도록 실험적으로 찾아낸 값으로 약 0.3정도이다.

부호화 및 복호화에 사용되는 프레임 길이도 반향 워터마크의 음질 및 강인성과 중요한 상관을 갖는 변수이다. 프레임 길이가 길어지면 그만큼 복호화에 유리하며 오프셋이 변화되는 중첩 구간의 발생 빈도도 줄어들어 음질에 대한 영향도 그만큼 줄어들지만, 대신 삽입되는 비트율이 줄어드는 단점을 가지게 된다. 실험 결과 프레임 길이 46msec 이상, 중첩 구간이 23msec 이상이면 음질과 복호화 성능이 크게 달라지지 않는 것으로 나타났으며, 본 논문의 실험을 위해서는 프레임과 중첩 구간에 여유 두고 각각 92.9msec, 46.4msec의 값을 사용하였다. 이는 한 프레임에 1비트를 삽입할 경우 약 7.17bps 정도의 비트율을 제공하는 프레임 크기이다.

표 1. 평가에 사용된 반향 커널들  
Table 1. Echo kernels evaluation.

	커널의 종류	스케일 팩터
방법 I	양의 단 반향	$\alpha$
방법 II	음의 단 반향	$\alpha$
방법 III	양의 단 반향	$1.7\alpha$
방법 IV	양의 다중 반향	$2\alpha$
방법 V	제안된 반향	$2\alpha$

6.1. 주관적 음질 평가

음질 평가 방법은 고품질 오디오에 대한 주관적 음질 평가에 주로 사용되는 Double blind triple stimulus with hidden reference 방법 (ITU-R의 권고안 BS.1116)을 응용하였다[13]. 평가에 사용된 음악은 클래식, 팝, 재즈의 다양한 장르에서 다양한 악기 편성을 가지는 6개의 곡으로 선택하였다. 선택된 음원 목록은 표 2에 정리하였다. 음질 평가는 음향 및 음성 신호처리를 전공하고 있는 10명의 대학원생을 대상으로 수행하였다.

다섯 가지 평가 대상이 되는 방법들이 6가지 음원에 대해 얻은 점수를 Diffgrade의 평균으로 나타낸 결과는 표 3과 같다. Diffgrade란 평가음이 얻은 MOS 값에서 원음이 얻은 MOS 값을 뺀 차이로써 원음이 높은 점수를 얻는 경우에는 음수로 나타나며, 반대로 평가음이 높은 점수를 얻는 경우에는 양수로 나타난다. 즉, Diffgrade가 양수이면 평가자가 원음과 평가음을 잘못 구분했음을 의미한다. 표 3의 마지막 열에는 전체 60개 평가 데이터 (6 음원 x 10명의 평가자) 중에 평가음을 원음이라고 오인

한 것의 개수인 '# of Transparent Items' 수치를 나타내고 있다. 반향을 이용한 워터마킹은 기본적으로 음질에는 심각한 왜곡을 가하지 않고 음색의 변화만을 가져오기 때문에 본 실험에서 얻은 수치가 일반적인 음질 평가의 결과와는 의미가 다를 수 있다. 그러나 워터마킹은 부호화 후에도 원음과 구별되지 않아야 한다는 본연의 취지에 충실하기 위해서 여기서는 음색의 변화도 왜곡으로 간주하도록 평가하였다.

표 2. 음질 평가에 사용된 음원 목록  
Table 2. Test materials used in subjective quality assessment.

	Diffgrade	# of Transparent Items
방법 I	0.15	32
방법 II	-0.08	22
방법 III	-0.53	3
방법 IV	-1.41	0
방법 V	0.04	27

표 3. 주관적 음질 평가 결과  
Table 3. Subjective quality assessment results.

곡명	특성
Maiden Voyage-Leny Andrade	베이스 심벌, 여성 보컬이 두텁게 강조된 곡.
Played Twice-The Fred Hersch Trio	피아노, 드럼, 베이스트리오 곡
I Love Paris-Johnny Frigo	바이올린이 흥겨운 연주곡
Sweet Georgia Brown-Monty Alexander	대편성 연주곡 (드럼 2세트, 베이스 2기)
Grandma's Hand-Livingston Taylor	4명의 남성 아카펠라 곡
Flute Concerto in D- Vivaldi	플루티 강조된 협주곡

표의 결과를 살펴보면 제안된 반향 삽입 방법인 방법 V와 초기 진폭을 지각 임계치 아래로 하여 단 반향을 삽입한 방법 I에서만 Diffgrade가 양의 값을 나타내었다. 즉, 이 두 방법에서는 반향을 삽입한 후의 오디오 신호에 대해 원음과 거의 구별을 못하고 있다. 이때 '# of Transparent Items' 값을 보아도 절반 가량이 평가음을 원음이라고 판단한 것으로 나타났다. 음의 단 반향을 삽입하는 방법 II의 경우는 같은 크기의 양의 단 반향을 삽입하는 방법 I과 비교할 때 약간 떨어지는 점수를 얻었지만, '# of Transparent Items' 가 22개나 될 정도로 원음에 가까운 음질을 제공하고 있다. 방법 I의 1.7배에 해당하는 반향을 삽입한 방법 III에서는 Perceptible but NOT annoying의 정도인 -0.53 점을 얻어 쉽게 지각은 되는 수준의 음질을 나타내며 양의 다중 반향을 삽입하는 방법 IV에서는 귀에 약간 거슬리는 정도인 'Slightly Annoying'의 -1.41점으로 가장 음질이 좋지 않은 것으로 나타났다. 제안한 반향 삽입 방법(방법 V)은 방법 I에 비해 두 배나 큰 크기의 반향을 삽입하고도 음질면에서 원음과 구별되지 않는 좋은 결과를 얻고 있다.



6.2. 복호화 성능 평가

복호화 과정은 2장에서 언급한 대로 반향 오프셋 위치에서의 오토켄스트럼의 피크를 검출하여 프레임 별로 '0'과 '1'의 정보를 얻어내는 것이다. 반향 오디오 워터마킹 기법의 강인성은 삽입된 반향의 크기, 즉 초기 진폭(다중 반향의 경우는 각 초기 진폭의 합)에 비례하며, 이는 복호화 과정에서 오토켄스트럼의 피크의 상대적 크기로부터 확인할 수 있다. 본 논문에서는 제안하는 기법의 강인성을 평가하기 위해 반향 삽입 이전의 오토켄스트럼과 각 방법에 의해 반향 삽입된 이후의 오토켄스트럼을 통계적으로 비교하는 실험을 수행하였다.

원 신호에 대해 얻은  $m$ 번째 프레임의  $n$ 번째 오토켄스트럼 계수를  $\tilde{x}_m(n)$ 이라고 하면, 오프셋  $\delta_k$ 에서의 오토켄스트럼 계수는  $\tilde{x}_m(\delta_k)$ 이다. 마찬가지로 반향이 삽입된 신호의 오토켄스트럼을  $\tilde{y}_m(n)$ 이라고 할 때,  $\delta_k$ 에서의 오토켄스트럼 계수는  $\tilde{y}_m(\delta_k)$ 이라 할 수 있다. 그림 12는 반향 삽입 이전의 오프셋 위치에서의 오토켄스트럼 계수  $\tilde{x}_m(\delta_k)$ 과 각 방법별로 반향을 삽입한 후의 오토켄스트럼 계수  $\tilde{y}_m(\delta_k)$ 의 프레임 누적 히스토그램을 통해 실험적으로 얻은 확률 분포이다. 단, 방법 IV, V에서는 두 개 오프셋 위치에서의 오토켄스트럼 계수를 합한 값에 대한 결과이다.

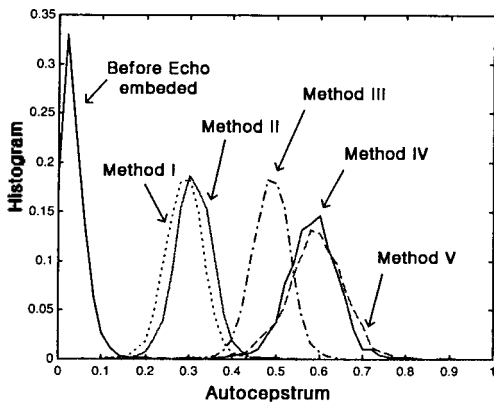


그림 12. 각 방법 별  $\tilde{y}_m(\delta_k)$ 의 실험적인 확률 분포  
Fig. 12. Empirical pdf of  $\tilde{y}_m(\delta_k)$  of each method.

반향 삽입 이전의 오토켄스트럼 값인  $\tilde{x}_m(\delta_k)$ 는 그림과 같이 0.0~0.17사이에서 분포한다. 반향이 삽입된 후에는 초기 진폭의 크기에 비례하는 평균을 갖는 정규 분포와 유사한 형태의 확률 분포 곡선이 나타난다. 확률 분포의 분산은 다중 반향을 사용하는 경우(방법 IV, V)가 단 반향을 사용하는 경우(방법 I, II, III)에 비해 더 크다. 평균은 피크가 그만큼 크게 나타남을 의미하므로 클수록 좋고, 분산은 일종의 검출 오차 분산으로 해석할 수 있으므로 작을수록 좋은 결과이다. 방법 I~V에 대해 실험적으로 얻은 평균과 분산은 표 4에 다시 정리하였다.

실험적으로 얻은 평균과 분산은 표 4에 다시 정리하였다.

표 4. 복호화 성능 평가 결과  
Table 4. Decoding performance test results.

	평균	분산	$p$
방법 I	0.2885	0.0019	0.2535
방법 II	0.3112	0.0019	0.2762
방법 III	0.4881	0.0019	0.4531
방법 IV	0.5843	0.0036	0.5141
방법 N	0.5916	0.0043	0.5214

다중 반향과 단 반향의 복호화 성능을 같은 척도로 비교하기 위해, 반향이 삽입된 신호에 대한 오토켄스트럼의 상대적인 피크 정도를 의미하는  $p_m$ 값을 다음과 같이 정의한다.

$$p_m = \sum_{k=1}^K (\tilde{y}_m(\delta_k) - \tilde{x}_m(\delta_k))$$

여기서  $K$ 는 삽입된 반향 펄스의 개수를 의미한다(즉, 단 반향인 경우는 1). 이 식의 의미는 반향이 삽입된 후에  $m$ 번째 프레임의 오프셋 위치에서의 오토켄스트럼 계수가 반향 삽입 이전에 비해 얼마나 증가하였는지를 나타낸다. 따라서 이 값이 크면 해당 프레임에서 그만큼 피크 검출이 용이하며 신호처리 공격 등에 의해 변형되었을 경우에도 살아남을 확률이 높아짐을 의미하는 것이므로 복호화의 성능 즉, 강인성을 평가할 수 있는 척도로 사용할 수 있다.

표 4의 마지막 열은 각 방법 별로 얻은  $p_m$ 의 평균값  $p$ 를 나타낸다. 결과를 살펴보면 반향 커널의 부호와 관계없이 삽입된 초기 진폭의 절대값에 비례하여  $p$ 값은 증가하며, 같은 크기일 때 음의 반향이 더 큰  $p$ 값을 나타내는 것으로 나타났다. 결과적으로 실험에 사용된 다섯 가지 방법 중에서 제안한 새로운 반향 커널을 이용한 방법 V가 가장 큰  $p$ 값을 갖는 것으로 나타나 강인성 측면에서 가장 우수함을 확인하였다

표 5는 방법 I과 제안한 방법(방법 V)의 커널을 이용하여 실제 워터마크가 삽입된 오디오 신호가 몇 가지 신호처리 공격에 의해 손상되었을 때 나타나는 복호화 오차를 BER로 정리한 것이다. 제안한 방법에 의한 경우는 5장에서 제안한 강인성 향상을 위한 기술들이 적용되었다. 각 신호처리 과정은 SDMI에서 제시하는 기준에 따랐다 [14]. 신호처리 공격을 거치기 전의 경우 방법 I은 0.32%의 오차를 보였지만 제안한 방법에서는 오차가 전혀 없이 100% 복호화가 가능한 것으로 나타났다. 실험에 사용된 모든 신호처리 공격 하에서도 제안한 방법이 방법 I에 비해 월등히 낮은 복호화 오차를 보이는 것을 알 수 있다. 특히 선형 속도 변화 공격의 경우는 파일럿 프레임을 사용하지 않는 방법 I에서는 최대 BER인 약 50%의 오차를

보이는 것으로 나타났으며, 제안한 방법에서는 속도 변화를 감지하고 보정함으로써 전혀 오차를 나타내지 않았다. 결과적으로 제안한 방법을 이용한 워터마킹 시스템은 위음과 구별되지 않는 음질을 제공하면서 방법 I과 비교할 때 매우 강인한 복호화 성능을 나타내는 것을 확인할 수 있다.

표 5. 신호처리 공격에 의한 손상 후 복호화 성능 평가 결과  
Table 5. Decoding performance results after signal processing attack.

신호처리 공격	BER (%)	
	방법 I	방법 V
공격이전	0.32	0.00
이퀄라이저	0.31	0.00
대역 통과 필터 (100-6kHz)	1.78	0.70
선형 속도 변환 (+/-10%)	49.98	0.00
MP3 압축 (56kbps/ch)	1.94	0.39
반향 삽입 (100ms, 0.5)	2.64	0.85

### VII. 결 론

반향을 이용한 워터마킹 방법에서는 음색 변화는 일으키지 않으면서 공격에 강인한 복호화 성능을 제공할 수 있는 반향 커널의 설계가 중요하다. 본 논문에서는 이를 위해 반향에 대한 심리 음향학적 분석을 통해 큰 에너지를 갖는 강인한 반향 커널을 새롭게 제안하고 이를 이용한 워터마킹 시스템의 성능을 평가하였다. 또한 반향 워터마킹 시스템의 추가적인 강인성 향상을 위하여 몇 가지 신호처리 공격에 대응할 수 있는 새로운 방법들을 제안하였다. 제안한 반향 커널은 기존의 반향 커널보다 2배 큰 반향을 음질 왜곡 없이 삽입할 수 있으며, 반향 펄스의 위치가 다중화되어 제3자에게 쉽게 노출되지 않는 장점을 갖는다. 또한 제3자가 오토캡스트럼을 이용하여 삽입된 워터마크 정보를 제거하려 할 경우 삽입된 커널의 부호는 알 수 없기 때문에 공격이 어렵다는 장점도 있다. 제안한 워터마킹 시스템은 주관적 음질 평가와 복호화 성능 평가 실험 결과 기존의 방법에 비해 매우 우수한 성능을 보이는 것으로 나타났다.

### 감사의 글

본 논문은 한국전자통신연구원이 지원한 방송 컨텐츠 보호를 위한 오디오 워터마킹 기술 개발 공동 연구의 일부로 수행된 것입니다.

### 참 고 문 헌

1. 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.  
 2. M. Swanson, M. Kobayashi and A. Tewfik, "Multimedia Data-Embedding and Watermarking Technologies (invited paper)", *Proc. of the IEEE*, vol. 86, No. 6, Jun., 1998.  
 3. W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding", *IBM System Journal*, vol 35, Nos 3&4, 1996.  
 4. I. Cox and M. Müller, "A review of watermarking and the importance of perceptual modeling", *Proc. of Electronic Image 97*, Feb., 1997.  
 5. I. J. Cox, J. Kilian, T. Leighton and T. Sharnoon, "A secure, robust watermark for multimedia", *Proc. Workshop on Information Hiding, Univ. of Cambridge, U.K.*, May 30-June 1, 1996.  
 6. L. Boney, A. Tewfik and K. Hamdy, "Digital Watermarks for Audio Signals", *IEEE Int. Conference on Multimedia Computing and Systems*, pp. 473-480, 1996.  
 7. M. Swanson, B. Zhu and A. Tewfik, "Current state of the art, challenges and future directions for audio watermarking", *IEEE Int. Conference on Multimedia Computing and Systems*, pp. 19-24, 1999.  
 8. 김현옥, 오현오, 윤대회, 차일환, "공격에 강인한 고품질 디지털 오디오 워터마킹 시스템에 관한 연구", 제13회 신호처리학술대회, vol. 13, No. 1, pp. 771-774, Sep., 2000.  
 9. A. Oppenheim and R. Shaffer, *Discrete-Time Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989.  
 10. G. Studebaker and I. Hochberg, *Acoustical Factors Affecting Hearing Aid Performance*, Allyn and Bacon, 1993.  
 11. E. Zwicker, *Psychoacoustics*. Springer-Verlag, New York, 1982.  
 12. C. Xu, and *et al.*, "Applications of Digital Watermarking Technology in Audio Signals", *J. Audio Eng. Soc.*, vol. 47, No. 10, Oct., 1999.  
 13. ITU-R Rec. BS.1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems", International Telecommunication Union, Geneva, Switzerland, 1994.  
 14. <http://www.sdmi.org/>

### ▲ 오 현 오 (Hyen-O Oh)



1996년 2월 : 연세대학교 전자공학과 (공학사)  
 1998년 2월 : 연세대학교 대학원 전자공학과 (공학석사)  
 1998년 3월 ~ 현재 : 연세대학교 대학원 전기전자공학과 박사과정  
 ※ 주관심분야 : 디지털 신호처리, 오디오 워터마킹, 오디오 부호화

## ▲ 김 현 욱 (Hyun-Wook Kim)



2000년 2월: 연세대학교 전자공학과  
졸업 (공학사)

2000년 3월~현재: 연세대학교 대학원  
전기전자공학과 석사과정

※ 주관심분야: 오디오 워터마킹, 오디오  
신호처리

## ▲ 석 증 원 (Jong-Won Seok)



1993년 2월: 경북대학교 전자공학과  
졸업 (공학사)

1995년 2월: 경북대학교 대학원 전자  
공학과 졸업 (공학석사)

1999년 8월: 경북대학교 대학원 전자  
공학과 졸업 (공학박사)

1999년 7월~현재: 한국전자통신연구원  
선업연구원

※ 주관심분야: 디지털 신호처리, 디지털 콘텐츠 보호

## ▲ 홍 진 우 (Jin-Woo Hong)



1982년 2월: 광운대학교 응용전자공학과  
졸업 (공학사)

1984년 2월: 광운대학교 대학원 전자  
공학과 졸업 (공학석사)

1993년 8월: 광운대학교 대학원 전자  
계산기공학과 졸업  
(공학박사)

1998년~1999년: 독일 프라운호퍼연구소 (교환연구원)

1984년 3월~현재: 한국전자통신연구원 음향기술연구팀장

2000년 1월~현재: 한국음향학회 홍보이사, 뉴미디어음향  
학술분과위원장

※ 주관심분야: 오디오 신호처리 및 부호화, 디지털 컨텐  
츠 보호, 디지털 오디오 방송

## ▲ 윤 대 희 (Dae-Hee Youn)

한국 음향학회지 제18권 제7호 참조