

반향 신호를 이용한 오디오 워터마킹에 관한 연구

Study for Audio Watermarking Using Echo Signal

오현오, 김현욱, 윤대희, 차일환

연세대학교 전기·전자공학과 음향·음성·신호처리 연구실

E-mail: oho@cyclon.yonsei.ac.kr

요 약

본 논문에서는 고음질 오디오 신호에 임의로 삽입된 반향(Echo) 신호가 음질에 미치는 영향을 조사하고, 이를 이용한 오디오 워터마킹 기법에 대해 다룬다. 일반적으로 오디오 신호에 반향을 첨가하게 되면 음색이 더욱 풍부해지는 효과를 얻을 수 있지만, 이 때 삽입된 반향신호의 시간 지연과 크기가 충분히 작을 경우에는 심리 음향 모델의 시간영역 마스킹 효과에 의해 지각되지 않을 수도 있다. 한편 오디오 신호의 구간별로 임의의 삽입된 반향의 시간지연을 검출할 수 있다면, 이를 이용한 정보 감춤(data hiding) 및 워터마킹 기법에 활용할 수 있다. 반향 신호를 이용하여 원 신호에 정보를 삽입하게 되면 가우시안 잡음이나 PN 시퀀스를 이용하는 경우처럼 오디오 신호에 이질적인 잡음을 첨가하지 않기 때문에 청감 특성상 유리하며, 오디오 신호 고유의 통계적 특성을 유지할 수 있는 장점이 있다. 그러나 반향의 첨가가 음질의 왜곡은 초래하지 않으면서 정보의 검출이 가능하도록 하기 위해서는 원 신호의 특성에 따른 반향 첨가 기술이 요구된다.

1. 서 론

인터넷의 확산과 함께 MP3와 같은 고음질의 디지털 오디오가 불법적으로 유통되면서 이를 막고 저작권을 보호하기 위한 방법이 필요하게 되었다. 디지털 워터마킹은 이러한 문제를 해결할 수 있는 최후의 방어 수단으로 인식되고 있는 새로운 연구 분야이다. 디지털 오디오 워터마킹이란 들리지 않는 정보를 오디오 신호 안에 은닉시키는 것을 말한다.

방송, 저장 및 인터넷 전송되는 오디오 콘텐츠의 저작권 보호를 위한 워터마킹은 다음과 같은 특징을 지니고 있어야 한다 [1]

- 지각되지 않아야 한다.

- 원음에 손상이 가지 않고서는 제거되지 않아야 한다.
- D/A-A/D 변환, 복사, 편집, 압축 등의 신호처리 과정에서도 변질되지 않도록 강인해야 한다.
- 원 신호 없이 복호화가 가능해야 한다.
- 실시간 처리가 가능해야 한다.

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 숨길 수 있는 영역도 상대적으로 부족하다. 그러므로 영상 신호의 워터마킹에 사용되던 기술을 바로 오디오 신호에 적용하게 되면 음질이 손상되기 쉬우며 강인성도 떨어질 수 있다.

반향(Echo)을 이용한 워터마킹 기법에서는 오디오 신호에 인위적인 반향을 첨가함으로써 정보를 삽입하게 된다. 적절하게 삽입된 반향은 사람에게 왜곡으로 인지되지 않으며 오히려 음색을 풍부하게 해주는 효과를 기대할 수 있기 때문에 오디오 워터마킹에 사용하기에 유리한 방법이다. 특히 삽입하는 반향의 크기와 시간 지연을 적절히 조절하면 원음과 청각적으로 전혀 구별되지 않는 반향 신호를 삽입할 수 있다. 또한 이렇게 삽입된 워터마킹 정보는 원 신호 없이도 쉽게 복호화가 가능하다는 특성을 갖는다. 본 논문에서는 청각적으로 원음과 구별되지 않는 반향을 삽입하는 워터마킹 방법을 제안하고 성능을 평가한다.

2. 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 필터를 통과하는 것으로 모델링 할 수 있다. 이 때 반향 특성을 결정하는 요소는 초기 진폭(initial amplitude), 감쇄율(decay rate), 오프셋(offset), 델타(delta) 등이 있다. 초기 진폭은 첫번째 반향

이 값은 크기가 감쇄율이란 두 번째 반향 신호부터 적용되는 초기 진폭으로부터의 상대적인 크기의 비율을 나타낸다. 오프셋은 원래 신호로부터 첫번째 반향까지의 시간지연을 의미하고 이후의 감쇄되는 반향 신호 사이의 간격은 델타로 정의한다.

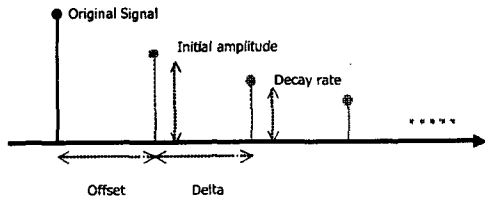


그림 1. 반향 신호의 충격 응답

2.1 부호화

오디오 신호에 반향이 삽입되는 과정은 원 신호가 그림 1과 같은 형태의 충격 응답 함수를 통과하는 것으로 해석할 수 있다. 반향을 워터마킹 정보의 삽입에 사용하기 위해서는 우선 이진수 '0' 인 경우와 이진수 '1' 인 경우에 각각 다른 오프셋을 갖는 반향 충격 응답 함수를 생성한다. 오디오 신호를 프레임별로 나누어 각 프레임마다 한 비트씩의 정보를 할당한 뒤, 그에 해당하는 반향 충격 응답 함수를 통과시킴으로써 워터마킹 정보를 은닉할 수 있다. 이때, 각 프레임 사이의 반향 오프셋의 급격한 변화는 시간 축에서 불연속점을 생성하여 음질에 영향을 미칠 수 있기 때문에 전후 프레임이 적절히 섞여 이어지는 접침 구간을 둘 필요가 있다[2].

2.2 복호화

반향을 이용한 워터마킹에서 삽입된 정보는 두 개의 다른 시간 지연(오프셋)을 갖는 반향 신호이다. 그러므로 워터마킹 정보를 복호화하는 것은 곧 각 프레임에서의 반향 신호의 오프셋 값을 찾아내는 것이다. 오디오 신호에 삽입된 반향의 시간 지연을 검출할 수 있는 가장 좋은 방법은 켈스트럼(Cepstrum)의 자기상관(Autocorrelation)을 구하는 것이다.

일반적인 오디오 신호에 대한 켈스트럼은 0에 가까운 부근에 많은 에너지가 몰려있는 형태로 나타나며 반향 충격 응답 함수의 켈스트럼은 오프셋 및 그의 배수에 해당하는 위치에 지수적으로 감쇄하는 주기적인 펄스가 존재하는 이산 지수(Discrete-time Exponential) 함수와 같은 꼴로 나타난다. 반향이 첨가되어 있는 오디오 신호의 켈스트럼은 이 두 켈스트럼의 합인 형태로 나타나므로 오프셋의 배수 위치에 주기적으로 피크가 나타나는 파형을 갖게 된다. 따라서 켈스트럼 신호에 자기상관함수(autocorrelation)를 취하게 되면 반향이 존재하는 신호의 경우 오프셋 위치에서 상대적으로 큰 피크를 검출할 수 있어서 이를 통해 워터마킹 정보를 복호화할 수 있다. 켈스트럼에 대한 자기상관함수 연산을 '오토켈스트럼(Autocepstrum)' 이라고 정의하며 자기상관과 파워스펙트럼 사이의 관계에 의해 다음 식과 같이 간단히 계산될 수 있다.

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

3. 최적의 반향 신호 설계

3.1 반향 신호의 분석

워터마킹에 이용되는 반향은 오디오 신호의 왜곡을 최소화하기 위해 하나의 펄스만을 가진, 즉 오프셋 위치에서의 첫번째 펄스만 존재하고 감쇄율이 0인 반향 커널을 사용할 수 있다. 이는 그림 2에서 처럼 한쪽 벽면만 존재하는 환경에서 소리가 발생할 때 마이크로폰으로 수음하는 경우와 생각할 수 있다. 이때 반향 경로를 모델링한 시간 축에서의 충격 응답과 이의 주파수 응답은 그림 3과 같은 형태로 나타난다.

그림 3의 충격 응답에서 두 펄스 사이의 간격이 멀어질수록 주파수 응답에서의 변화율이 높아지게 된다. 이렇게 삽입된 반향 신호는 초기 진폭이 같아도 오프셋의 크기에 따라 다른 청취상의 특성을 일으킨다. 오프셋이 200msec 이상으로 큰 경우에는 흔히 메아리라고 하는 경우와 같은 자연스런 울림 현상이 나타나고, 10msec 이하로 작은 경우는 반향은 생기지 않고 주파수 응답의 영향에 의한 음색(timbre)의 변화로 인지되며 이를 착색효과(coloration)라 한다. 같은 충격 응답에 대해 오프셋의 차이로 이렇게 다른 효과가 나타나는 이유는 HAS의 인지 특성에서 기인한다.

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[3]. 이때 각 임계 대역의 신호를 받아들이는 필터들은 메모리를 가지고 있어서 일정 시간만큼의 입력 신호에 대해 한 번에 처리하게 되며 각 대역내에서는 상세한 주파수 형태보다는 에너지의 합으로 소리를 구별하는 특성을 갖는다[4]. 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 인지하지 못하기 때문에 메아리는 인지하지 못한다. 반면 긴 오프셋을 갖는 충격 응답의 경우는 귀가 갖는 필터의 메모리 길이보다 펄스 사이의 간격이 길기 때문에 두 펄스에 의한 주파수 응답 특성은 나타나지 않고 대신 시간 영역에서 지연된 신호를 다시 듣게 되는 자연스런 메아리가 인지된다. 그러나 초기 진폭이 충분히 작으면 착색효과나 메아리효과 모두 귀의 분해능력의 한계로 인해 인지되지 않는다. 또한 착색효과와 메아리효과 모두 음색이 풍부해지고 공명이 증가하는 방향으로 변화하기 때문에 그 차이를 인지한다 해도 왜곡으로 판단하기 어렵다.

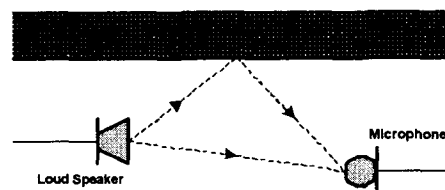


그림 2. 단 반향(Single Echo)이 발생하는 환경.

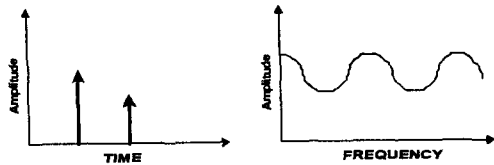


그림 3. 단 반향의 충격 응답과 주파수 응답.

3.2 순시 마스크

심리음향학에서 연구되는 순시 마스크(Temporal Masking)이란 시간영역에서 레벨이 큰 신호가 발생한 직후 혹은 직전에 존재하는 작은 신호는 귀에 들리지 않는 현상을 말한다[1]. 그림 4는 순시 마스크의 예를 보이는 시간 축에서의 곡선이다. 마스크(masker)가 존재하는 시간 직후와 직전에 소리가 들리지 않는 음영구간이 생기는데 특히 마스크 이후에 생기는 후-마스크(post-masking)의 영향이 더 크다. 마스크 현상의 원인이 되는 마스크의 음압과 지속시간에 따라 후-마스크이 미치는 범위는 50~200msec 까지 달라지는 것으로 알려져 있다. 또한 같은 주파수 대역에 존재하는 신호가 다른 대역의 신호보다 순시 마스크에 의해 더 많은 영향을 받는다. 반향을 위해 첨가되는 신호는 원 신호와 정확히 같은 주파수 분포를 갖는 신호이므로 순시 마스크에 의해 가장 잘 마스크가 되는 신호라고 볼 수 있다. 그러므로 초기 진폭과 오프셋을 후-마스크 임계치 아래에 위치하도록 조절하여 삽입된 반향은 순시 마스크에 의해 귀에 들리지 않게 된다.

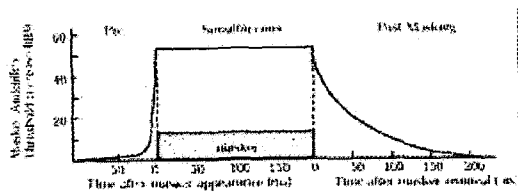


그림 4. 순시 마스크(temporal masking).

3.3 제안된 반향 신호 설계

앞 절에서 설명된 반향 신호에 관한 심리음향학적 해석을 종합해보면, 반향 신호의 삽입은 오프셋에 따라 음색의 변화를 일으키거나 자연스런 메아리로 인지된다. 순시 마스크 효과의 이득을 보기 위해서는 오프셋이 작을수록 유리하지만, 너무 작게 되면 음색 변화가 둔탁함을 느끼게 하는 왜곡으로 인지되어 좋지 않다. 또한, 음색 변화가 인지되지 않을 정도의 반향을 삽입하기 위해서는 초기 진폭이 작아져야 하는데, 초기 진폭이 작으면 그만큼 검출에 어려움이 발생한다. 그러므로 순시 마스크와 착색효과 영향, 그리고 검출에 요구되는 최소의 크기를 복합적으로 고려하여 반향 펄스의 오프셋과 초기 진폭을 설정해야 한다.

반향 펄스가 하나만 존재하는 경우를 단 반향(Single Echo)이라고 할 때, 여러 개의 반향 펄스를 사용하는 경우는 다중

반향(Multiple Echo)이 된다. 반향 펄스를 여러 개 사용하고 대신 각 펄스의 크기를 단 반향보다 작게 만들면, 음질 왜곡은 줄이면서도 검출 성능은 향상시킬 수 있다[6]. 그러나 청취 실험 결과에 따르면 하나의 반향을 사용하는 경우보다 좋은 검출 성능을 얻을 수 있을 만큼의 다중 반향을 삽입하게 되면, 단순한 음색의 변화가 아닌 음질의 왜곡이 발생하였다. 이는 각각의 반향 펄스가 합쳐진 형태로 귀에 인지되기 때문에 발생하는 문제로 해석된다.

본 논문에서는 음의 반향을 포함한 형태의 새로운 다중 반향 워터마킹 삽입 방법을 제안한다. 음의 반향이란 반향 펄스의 크기가 음수인, 즉 위상이 반전된 형태의 반향을 말한다. 그림 5의 (a)와 (b)는 단 반향일 때의 충격 응답과 주파수 응답을 각각 나타내고 있으며, (c)와 (d)는 본 논문에서 제안하는 양의 반향과 음의 반향이 섞여있는 다중 반향의 충격 응답과 주파수 응답을 나타낸다. 제안한 방법으로 다중 반향을 삽입하게 되면, 기존의 양의 반향만을 갖는 다중 반향의 경우와 같은 검출 특성을 얻으면서도 음질의 왜곡은 인지되지 않는다. 오토레스트림에 의한 반향 피크 검출 시에는 반향의 부호가 영향을 주지 않는 것으로 검증되었다.

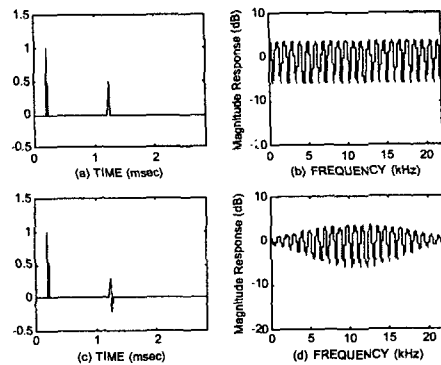


그림 5. 반향 충격 응답과 주파수 응답.

4. 실험 및 결과

제안된 반향 삽입 방법의 성능을 평가하기 위해 여러가지 반향 삽입 방법과 비교하여 주관적 음질 평가와 복호화 성능 평가를 수행하였다. 비교를 위해 사용된 방법은 단 반향 A (음색 변화가 생기지 않는 수준에서 초기 진폭을 최대화한 방법), 단 반향 B (약간의 음색 변화를 감수하고 검출 성능을 높인 방법), 다중 반향(제안된 방법과 같은 수준의 검출 성능을 나타내는 양의 다중 반향 삽입)이다.

4.1 주관적 음질 평가

음질 평가 방법은 고음질 오디오에 대한 주관적 음질 평가에 주로 사용되는 Double blind triple stimulus with hidden reference 방법(ITU-R의 권고안 BS.1116)을 응용하였다[4]. 평가에 사용된 음악은 클래식, 팝, 재즈의 다양한 장르에서

다양한 악기 편성을 가지는 6 개의 곡을 선택하였다. 음질 평가에 참여한 평가자는 음향 및 음성 신호처리를 전공하고 있는 10 명의 대학원생이다. 네 가지 평가 대상이 되는 워터마킹 방법들에 대해 10 명의 평가자로부터 얻은 점수를 ITU-R 의 5 단계 열화 척도로 나타낸 최종적인 평가 결과를 표 1 에 정리하였다.

반향을 이용한 워터마킹 삽입은 기본적으로 음질에 심각한 왜곡을 가하지 않지만, 표의 결과에서 보듯이 큰 크기를 갖는 양의 다중 반향을 삽입할 경우, 귀에 약간 거슬린다는 의미의 'Slightly Annoying' 평점을 얻는다. 단 반향을 이용한 경우에도 초기 진폭이 커질 경우, 단 반향 B 의 방법에서처럼 어느 한계를 넘어서면 음질이 열화되는 것으로 나타났다. 본 논문에서 제안한 방법에 경우 양의 반향만을 사용한 다중 반향의 경우와는 다르게 원음과 구별되지 않는 음질을 제공한다 는 결과를 얻을 수 있었다.

4.2 복호화 성능 평가

복호화 과정은 2 장에서 언급한 대로 반향 오프셋 위치에서의 오토캡스트럼의 피크를 검출하여 프레임 별로 '0' 과 '1' 의 정보를 얻어내는 것이다. 따라서 복호화 성능을 평가하기 위해서는 각 방법에 의해 부호화된 반향이 형성하는 오토캡스트럼의 피크가 얼마나 크게 나타나는지를 확인함으로써 알 수 있다.

원 신호에 대해 얻은 m 번째 프레임의 n 번째 오토캡스트럼 계수를 $\tilde{x}_m(n)$ 이라고 하면, 오프셋 δ_k 에서의 오토캡스트럼 계수는 $\tilde{x}_m(\delta_k)$ 이다. 마찬가지로 반향이 삽입된 신호의 오토캡스트럼을 $\tilde{y}_m(n)$ 이라고 할 때 δ_k 에서의 오토캡스트럼 계수는 $\tilde{y}_m(\delta_k)$ 이라 할 수 있다. 반향이 삽입된 신호에 대한 오토캡스트럼의 상대적인 *peakness* P_m 를 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 오토캡스트럼 계수가 반향 삽입 이전에 비해 얼마나 증가하였는지를 나타낸다. 따라서 이 값이 크면 해당 프레임에서 그만큼 피크 검출이 용이하며 신호가 변형되었을 경우에도 살아남을 확률이 높아짐을 의미하는 것이므로 복호화 성능을 평가할 수 있는 척도로 사용할 수 있다.

표 1 에 제시된 수치는 각 방법에서의 오토캡스트럼 *peakness* P_m 의 평균값을 나타낸다. 다중 반향을 사용하는 경우가 단 반향을 사용하는 경우보다 높은 *peakness* 를 나타내는 것을 볼 수 있으며, 특히 제안한 음의 반향을 사용하는 다중 반향 삽입방법에서 가장 높은 평균 *peakness* 를 얻는 것으로 나타났다. 실제 워터마크 정보를 삽입한 후 복호화했을 때, 네 가지 방법 모두 오차없이 정보를 얻어낼 수 있었다. 그러나 mp3 부호화와 같은 다양한 신호처리적 공격에 의해 오디오 신호

가 변형된 이후 복호화를 할 경우에는 *peakness* 가 높은 방법일수록 낮은 오차율을 보이는 것으로 나타났다.

5. 결 론

본 논문에서는 오디오 신호에 들리지 않는 정보를 은닉하는 디지털 오디오 워터마킹 방법 중 반향을 이용한 워터마킹에 대한 연구를 통해 성능을 최대화할 수 있는 반향 신호 삽입 방법을 제안하고 이를 이용한 워터마킹 시스템의 성능을 평가하였다. 반향을 이용한 워터마킹 방법에서는 음색에 영향을 주지 않으면서 복호화 성능을 최대화 시키는 최적의 반향 신호 설계가 중요하며 실험에 사용된 네 가지 반향 삽입 방법 중 제안한 방법이 음질에 변화를 영향을 주지 않으면서도 단 반향을 삽입하는 경우보다 두 배 이상 좋은 복호화 성능을 보이는 것으로 나타났다.

6. 참고문헌

- [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] E. Zwicker, *Psychoacoustics*. Springer-Verlag, New York, 1982.
- [4] G. Studebaker and I. Hochberg, *Acoustical Factors Affecting Hearing Aid Performance*, Allyn and Bacon, 1993.
- [5] 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.
- [6] C. Xu, and *et al.*, "Applications of Digital Watermarking Technology in Audio Signals," *J. Audio Eng. Soc.*, Vol. 47, No. 10, 1999 Oct.

표 1. 음질 및 복호화 성능 평가 결과.

삽입 방법	Subjective Quality	peakness
단 반향 A	Imperceptible	0.2535
단 반향 B	Perceptible, but not Annoying	0.4531
다중 반향	Slightly Annoying	0.5141
제안된 반향	Imperceptible	0.5214