

디지털 오디오의 계층적 워터마크

정사라, 홍진우
한국전자통신연구원 방송미디어연구부
대전광역시 유성구 가정동 161번지

The Layered Digital Audio Watermark

Sarah Jung and Jin Woo Hong
Electronics and Telecommunications Research Institute (ETRI)
E-mail: sjung@etri.re.kr

요약

본 논문에서는 디지털 오디오 데이터에 부가 정보를 삽입하는 기술로써 계층적 워터마크를 사용하여 시간 변형을 감지할 수 있고, 필요에 따라 검출 복잡도를 조절할 수 있는 기법을 제안한다. 1계층 워터마크는 오디오 데이터의 시간축 변형을 감지할 수 있도록 오디오 데이터의 반향을 이용하고, 2계층 워터마크는 1계층 워터마크된 오디오 데이터를 기준으로 대역 확산 기법을 이용하여 저작권 정보 등의 요구량이 많은 부가 정보를 삽입한다. 이 때, 2계층 워터마크는 프레임의 동기 확보를 위한 동기 수열, 부가 정보 삽입을 위한 데이터 수열, 두 개를 이용한다. 검출기에서는 시스템의 요구 사항에 따라 1계층, 2계층 전단계, 2계층 본단계 등의 계층적 순서로 검출할 수 있으며, 각 계층은 데이터에 가해진 변형 정도를 추정하여 다음 계층의 검출단에 정보를 제공한다. 여러 가지 실험 결과를 통하여 제안한 방식이 다양한 신호 처리에 강인함을 보였다.

1. 서론

최근 인터넷이나 무선 네트워크를 통한 멀티미디어 전송 서비스의 이용이 급격히 늘어나면서 디지털 데이터의 불법 복제, 무단 이용 등의 문제가 심각하게 대두되었다. 이것을 해결할 하나의 대안으로 워터마크에 대한 연구가 활발하게 진행되고 있는데, 여기서 워터마크란 디지털 멀티미디어 데이터에 삽입된 인지할 수 없는 정보를 의미한다[1].

워터마크는 본 데이터에 지각적으로 인식되지 않을 만큼 작은 변화를 주면서, 의도적이거나 또는 비의도적인 여러 가지 신호 처리에도 살아 남아야 하므로 본질적으로 비인지성과 강인성의 두 요구가 상충되는 특성을 갖는다. 대역 확산 기법(spread spectrum)은 이러한 특성에 잘 부합하는 기술로 많은 연구가 이 기법을 이용하였다[2]-[5].

대역 확산 기술을 이용한 워터마크는 크기 변형, 주파수 등화(equalize), 고음질(고화질) 압축, 잡음 첨가 등 대부분의 신호 처리에 대해서는 매우 강인하지만, 시간 변형(기하학적 변형)에는 살아 남지 못한다. 오

디오 신호는 방송 시간을 맞추거나 조절하기 위해서 시간 변형을 할 때가 있는데 이러한 변화를 고려하지 않은 단순한 시스템에서는 대역 확산 워터마크를 제대로 검출해 낼 수 없다. 오디오 신호의 시간 변형 형태로는 크게 선형 속도 변화(linear speed change)와 피치를 보존하는 시간 눈금 변형(time scale modification), 두 가지를 고려해 볼 수 있다. 선형 속도 변화는 오디오 신호에 반향을 삽입하고, 검출기에서 그 반향의 위치 변화를 관측함으로써 속도 변화량을 측정할 수 있다[6]. 그러나 피치를 변화시키지 않는 시간 눈금 변형은 오디오 신호의 주파수 고유 특성을 그대로 유지하면서 주기적으로 일정 구간이 중첩 또는 소실되는 형태로 일어나기 때문에 반향으로는 감지하기 어렵다. [4], [5]에서는 이와 같이 동기가 어긋난 경우에 대비해 각각 특정 헤더 수열을 이용하거나, 두 수열의 부호를 이용하여 프레임의 동기를 확보하는 방법을 제안하였다. 그러나, 이 두 방식은 모두 선형 속도 변화가 일어났을 경우에는 대처할 수 없으므로 다른 보정 방법이 필요하다.

본 논문에서는 시간 변형에 대응하기 위해 반향 워터마크와 pn(pseudo noise) 수열 워터마크를 결합한 계층적 워터마크 알고리즘을 제안한다. 1계층 워터마크로 반향을 삽입하여 검출기에서 워터마크의 유, 무 정보와 선형 속도 변화량을 추정하고, 2계층 워터마크로 pn 수열을 이용하여 정보량이 많은 데이터를 삽입한다. 이 때, 데이터 삽입을 위한 데이터용 pn 수열 외에 프레임 동기 확보를 위한 동기용 pn 수열을 두어서 검출부에서 동기문제를 용이하게 해결할 수 있다. 1계층에서 추정된 선형 속도 변화량을 이용하여, 2계층 전단계에서는 프레임 동기를 확보하고, 본단계에서는 앞 계층들에서 추정된 시간 변형 정보를 이용하여 삽입된 부가 정보를 추출한다. 검출기에 요구된 복잡도, 필요한 정보 정도 등에 따라 1계층, 2계층 전단계, 2계층 본단계 등으로 검출 계층을 조절할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서 제안한 계층적 오디오 워터마크의 삽입 시스템과 검출 시스템을 설명한다. 3장에서는 이 시스템에 대해 실험을 수행하여 결과를 분석하고, 4장에서 결론을 맺는다.

2. 제안한 계층적 오디오 워터마크

2.1 워터마크 삽입

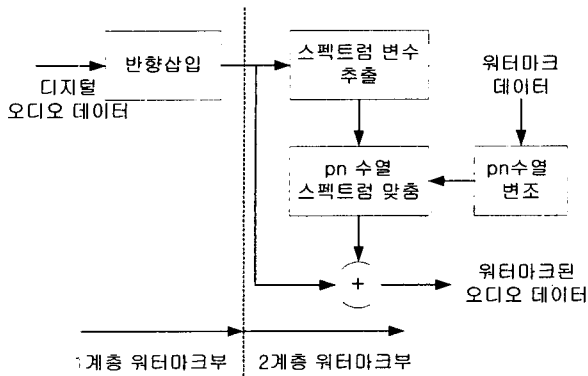


그림 1. 계층적 오디오 워터마크 삽입 시스템

그림 1에 제안한 워터마크의 삽입 시스템을 나타내었다. 먼저 입력 디지털 오디오 데이터에 1계층 워터마크로 반향을 삽입한 후 그 데이터의 스펙트럼 변수를 추출한다. 여기서 스펙트럼 변수란, 삽입할 pn 수열이 비인지성을 유지하는 한도에서 최대의 전력으로 삽입될 수 있도록 입력 오디오의 스펙트럼 특성을 추출하는 것을 의미한다. 스펙트럼 변수는 [2]에서와 같이 고음질 오디오 압축 부호화에 쓰이는 심리 음향 모델을 이용하거나, 다양한 방식의 신호 처리를 이용할 수 있는데, 본 시스템에서는 검출부의 전처리 과정을 고려하여, 선형 예측 계수(linear predictive coefficient)를 이용하였다. 이에 대한 자세한 설명은 2.2 절의 워터마크 검출 부분에서 하기로 한다. 삽입하려는 워터마크 데이터의 비트 정보 '1', '0'에 따라 데이터용 pn 수열의 부호를 반대로 변조하고, 동기용 pn 수열은 데이터에 관계없이 항상 같은 부호로 삽입된다.

2.2 워터마크 검출

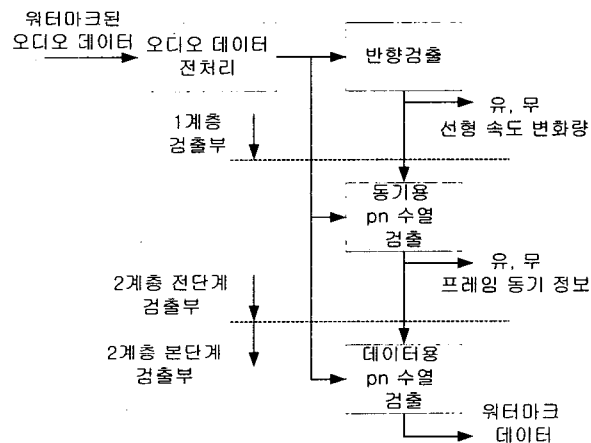


그림 2. 계층적 오디오 워터마크 검출 시스템

그림 2에 제안한 워터마크의 검출 시스템을 나타내었다. 일반적으로 삽입된 pn 수열의 크기가 커버 오디오 데이터에 비해 상당히 작으므로 검출 성능을 향상시키기 위해 여러 가지 전처리 필터링을 수행하는

데, 본 시스템에서는 [3]에서 제안된 역 선형 예측 필터(inverse linear predictive filter)를 이용하였다. 역 선형 예측 필터는 입력 오디오 데이터를 백색화(whitening)시키므로 필터 출력 후 워터마크 데이터의 상대적인 에너지비를 증가시키는 효과를 볼 수 있다. 특히 본 시스템에서는 그림 3과 같이 pn 수열이 삽입부에서 선형 예측 필터링되고, 검출부에서는 역 선형 예측 필터링되므로 신호 처리 공격이 없다면, 거의 원형으로 복구되어 검출부의 교차 상관(cross correlation)을 향상시키는 효과까지 볼 수 있다.

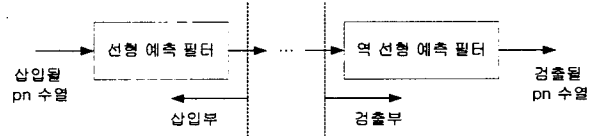


그림 3. 제안된 워터마크 시스템에서 pn 수열이 겪는 신호처리 과정

심각한 음질 왜곡이 없으면서 발생할 수 있는 선형 속도 변화는 $\pm 10\%$ 정도이다. 따라서 이 범위 내에서 반향의 위치 변화를 관측함으로써 입력 오디오 데이터가 겪은 선형 속도 변화를 추정할 수 있다.

1계층 검출이 완료되면, 선택적으로 2계층 전단계 검출이 시작되는데, 1계층 검출 결과 선형 속도가 변화되었을 경우 그 변화량만큼 검출부의 pn 수열의 길이를 변화시킨다. 프레임 동기 확보를 위해 샘플 단위로 검출부의 동기용 pn 수열과 입력 오디오 데이터(전처리된 오디오)의 교차 상관을 구하여, 프레임 내에서 교차 상관 값이 가장 큰 샘플 위치가 그 프레임의 동기 위치가 된다. 이 과정을 수식으로 표현하면, 아래와 같다.

$$\max_l \{r_{xy}(l) : l \in \{0, 1, \dots, N-1\}\} \quad (1)$$

이 때, N 은 프레임당 샘플 수이며,

$$r_{xy} = \sum_{n=0}^{N-1} x(n)y^*(n-1)$$

$x(n)$: 동기용 pn 수열,

$y(n)$: 전처리된 오디오 데이터

이다. 그런데, 제안된 시스템에서 동기용 pn 수열은 부호가 바뀌지 않고, 일정하게 삽입되었으므로 다음의 DFT (Discrete Fourier Transform) 특성을 이용하여 프레임 임별로 원형 상관(circular correlation)을 구하고 이 값을 교차 상관값으로 이용할 수 있다.

$$\tilde{r}_{xy}(l) \xrightarrow{DFT, N} \tilde{R}_{xy} = X(k)Y^*(k) \quad (2)$$

$$\text{이 때, } \tilde{r}_{xy} = \sum_{n=0}^{N-1} x(n)y^*((n-1)_N)$$

$$x(n) \xrightarrow{DFT, N} X(k),$$

$$y(n) \xrightarrow{DFT, N} Y(k)$$

이고, $(n)_N$ 은 n 의 N 에 대한 모듈로 값이다. 샘플 단위로 시간 영역에서 교차 상관을 구하려면, $N \times N$ 번의 곱셈이 필요하다. 본 시스템에서와 같이 상관값을 구하기 위해 원형 상관을 이용하면, DFT-샘플별 곱셈-IDFT(Inverse Discrete Fourier Transform)의 과정을

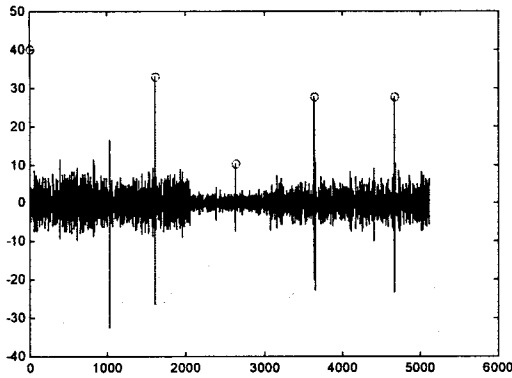


그림 4. 워터마크된 오디오 데이터에 시간 눈금 변형이 있을 때, 데이터 중첩(소실) 부분에서 검출단 pn 수열과의 교차 상관 특성

거치므로, $N \cdot (2 \cdot \log_2 N + 1)$ 의 곱셈이 필요하다. 그러므로 일반적으로 프레임당 1000샘플 이상의 길이를 갖는다고 할 때 계산량 면에서 훨씬 효율적이 된다. 그림 4에 시간 눈금 변형된 워터마크 오디오 데이터의 데이터 중첩(소실) 부분에서 동기용 pn 수열과의 교차 상관 특성을 나타내었다. 데이터가 오버랩 되면서 피크의 주기는 일정하고, 동기 위치만 변화되는 것을 볼 수 있다.

1계층, 2계층 전단계 검출이 완료되고 선택적으로 2계층 본단계 워터마크 검출이 이루어진다. 1계층 검출에서 추정된 선형 속도 변화량에 따라 데이터용 pn 수열을 변화시킨 후, 2계층 전단계에서 확보한 프레임 동기 정보를 이용하여 입력 오디오와(전처리된 오디오 데이터) 데이터용 pn 수열의 교차 상관을 구한다. 교차 상관의 부호로 '1', '0'의 비트 정보를 결정한다. 이 과정을 수식으로 표현하면, 아래와 같다.

$$c_{xy} = \frac{\sum_{n=0}^{N-1} z(n)y^*(n-l)}{\sqrt{\sum_{n=0}^{N-1} z^2(n)} \sqrt{\sum_{n=0}^{N-1} y^2(n-l)}} \geq 0? \quad (3)$$

이 때, $z(n)$ 은 데이터용 pn 수열이며, l 은 (1)에서 구한 동기 위치이다.

3. 실험 결과 및 분석

편의상, 2계층 본단계까지 검출하는 것을 원칙으로 하되, 30초씩 2계층 전단계까지의 결과를 분석한 후 그것을 바탕으로 2계층 본단계를 검출하고, 다음 30초의 데이터를 검출하는 형태로 이루어졌다. 실험에서 사용한 오디오 데이터는 표본주파수 44.1kHz, 샘플당 16비트 크기의 어쿠스틱 재즈, 작은 소리의 피아노 연주, 바이올린 연주, 드럼과 키보드가 포함된 재즈 연주, 아카펠라, 현악 클래식 6곡에 대해서 각각 30초씩 편집하여 만든 3분 길이의 곡이다. 프레임당 샘플 수는 2000개로 하였고, pn 수열은 무작위적으로 발생시킨 난수를 이용하였다.

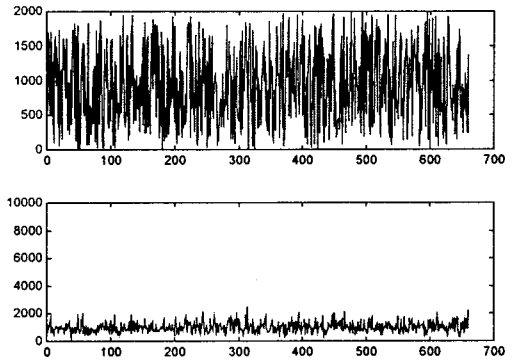


그림 5. 워터마크가 없을 경우 2 계층 전단계의 실험 결과: 프레임당 피크 위치(위), 피크 크기(아래)

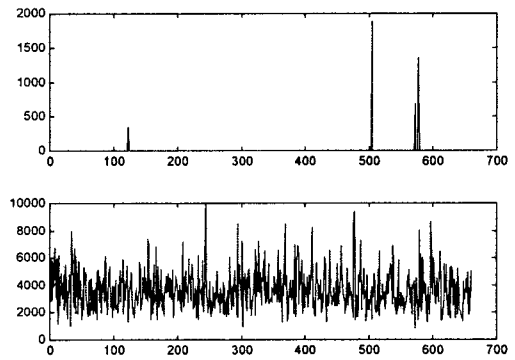


그림 6. 워터마크된 데이터에 공격이 없는 경우 2 계층 전단계의 실험 결과: 프레임당 피크 위치(위), 피크 크기(아래)

먼저, 오디오 데이터에 가해진 신호 처리가 워터마크 검출에 미치는 영향을 보기 위해서 워터마크를 비교적 강하게 삽입하여(SNR = 13.59dB) 2계층 전단계의 결과를 살펴보았다. 워터마크를 하지 않은 원곡, 워터마크 데이터에 공격이 없는 경우, 10% 선형 속도 변화가 가해진 경우, 4% 시간 눈금 변형이 가해진 경우에 대해서 2계층 전단계의 실험 결과는 각각 그림 5~8과 같다. 워터마크가 없는 데이터의 경우 1단계에서 '워터마크 없음'의 결과를 얻고 그 이후 계층의 검출을 중단하기 때문에 비교를 위해서 그 결과를 무시하고 2계층 전단계를 수행하도록 하였다.

워터마크가 있을 때, 피크는 한 위치에 거의 일정하게 나타나고 그 크기도 2000이상으로 크다. 그림 6,에서는 샘플값 1에서, 그림 7에서는 샘플값 2200이 동기 위치에 해당된다. 반면, 그림 5와 같이 워터마크가 없는 데이터에 대해서는 피크 위치가 무작위적이며, 그 크기도 2000이하로 작아서 워터마크 유, 무의 경우를 분명하게 판단할 수 있음을 알 수 있다. 한편, 시간 눈금 변형이 일어난 그림 8을 관찰하면, 피크 위치가 주기적으로 변하는 것을 확인할 수 있다. 이것은

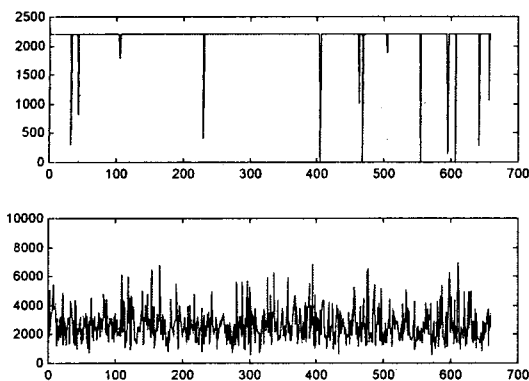


그림 7. 워터마크된 데이터에 10% 선형 속도 변화가 가해진 경우 2계층 전단계의 실험 결과(다); 프레임당 피크 위치(위), 피크 크기(아래)

그림 4에서와 같이 시간 눈금 변형이 주기적으로 일정 구간 반복되거나 소실되는 형태로 이루어지기 때문이다.

그림 5-8의 실험 결과를 바탕으로 2계층 전단계 검출 결과에서 데이터 종류를 판정하는 다음의 과정을 유도할 수 있다. 이것은 1계층 검출 결과에서 '워터마크 있음'으로 판정받고 2계층이 진행된 것을 전제로 하며, 변형 신호 처리는 분석 구간에 대해 고르게 일어난 것을 가정한다.

- (1) 피크 위치에 대한 히스토그램을 구한다.
- (2) M개 이상의 도수를 갖는 피크 위치를 찾는다.
- (3) (2)에서 찾은 피크 위치가 N개 이상이면, 이 데이터는 시간 눈금 변형이 일어난 것으로 판정한다.
- (4) (2)에서 찾은 피크의 위치가 하나 이상 N개 미만이면, 가장 큰 도수를 갖는 위치를 전체 데이터의 프레임 동기 위치로 판정한다.
- (5) (2)에서 찾은 피크의 위치가 0일 때에는 신호에 큰 변형이 있을 것으로 추정한다.

M, N은 pn 수열을 삽입하는 프레임의 길이와 전체 곡 길이에 따라서 달라질 수 있는데, 여기서는 편의상 10으로 두었다.

SNR이 13.59dB일 때는 음질에 잡음감이 있으므로 워터마크의 강도를 줄여서 즉, SNR을 증가시켜서 BER(bit error rate) 검출 성능을 조사하고, 그 결과를 표 1에 나타내었다. 여기서 가해진 신호 처리는 대부분 SDMI[7] 컨퍼넬 2.1버전의 강인성 테스트를 따른 것인데, 고음질 오디오 압축은 편의상 MPEG-2 AAC 코덱을 사용하였다. SNR이 높아질수록 즉, 삽입한 워터마크의 강도가 약해지면, BER 검출 성능이 떨어진다. 특히 시간 눈금 변형이 일어난 경우 SNR이 17.11dB일 때는 변형을 감지하고 동기 위치를 변화시키며 데이터를 검출할 수 있었지만, SNR이 23.13dB일 때는 워터마크 강도가 약하기 때문에, 동기 위치 주변으로 피크가 분산되어서 2계층 전단계 검출 결과를 분석하는 과정에서 큰 변형을 겪고 피크가 무작위로 분포된 (5)번으로 판정되었다.

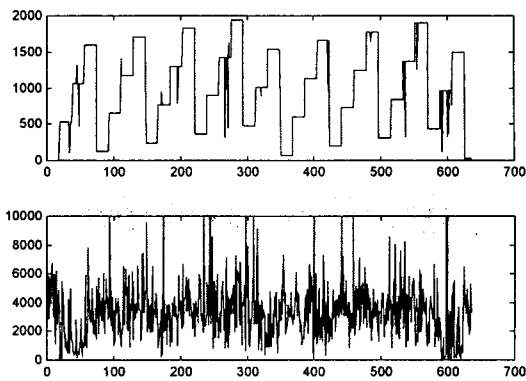


그림 8. 워터마크된 데이터에 4% 시간 눈금 변형이 가해진 경우 2계층 전단계의 실험 결과(다); 프레임당 피크 위치(위), 피크 크기(아래)

신호처리	BER (17.11dB)	BER (23.13dB)
없을 때	5.044e-004	1.135e-002
MPEG-2 AAC 압축 (main profile, 128kbps)	2.522e-003	8.197e-002
크기 압축	7.566e-004	1.488e-002
주파수 등화	0	1.892e-002
중간대역 필터	4.288e-003	8.172e-002
선형 속도 변화	5.058e-004	2.807e-002
시간 눈금 변형	O	X
반향 첨가	8.323e-003	7.238e-002

표 1. 다양한 신호 처리에 따른 BER 검출 성능

표 1의 실험 결과 본 논문에서 제안한 계층적 워터마크 시스템이 신호 처리에 대해서 비교적 강인한 것을 확인할 수 있으나, 몇 가지 보정 작업이 필요하다. 첫째, 시간 눈금 변형이 일어난 경우 그것을 감지하고, 프레임 동기를 찾는다 하더라도 프레임이 중첩되거나 소실되므로 여기서부터는 워터마크 데이터의 비트 동기가 어긋나게 된다(프레임 동기가 아님). 따라서 이러한 부분을 고려하여 삽입, 검출하는 알고리즘이 필요하다. 둘째, 표 1의 SNR이 23.13dB일 때의 결과와 같이 워터마크의 강도를 줄일 경우 전체적으로 BER 검출 성능이 떨어지고, 특히 시간 눈금 변형이 일어난 경우, 피크의 분포 때문에 아예 검출하지 못하는 결과가 발생할 수 있다. 이렇게 커버 오디오 데이터의 영향으로 피크가 그 전후 주변으로 퍼져서 분포하게 될 경우를 피크 히스토그램에 대해 보다 정교한 해석이 필요하다. 마지막으로, SNR이 17.11dB일 때 BER 검출 성능은 매우 좋지만, 음질면에서 음악적 내용이 갑자기 변화하거나 음압 레벨이 작은 부분에서는 잡음감이 발생하는 문제가 있다. 이것은 제한한 시스템에서 워터마크를 삽입할 때, 전 프레임에 대해

이득값을 일정하게 했기 때문이다. 그러므로 적절한 범위에서 커버 오디오 데이터의 특성에 따라 적응적으로 워터마크의 전력 이득을 변화시킨다면, BER 성능과 음질면에서 좀 더 향상된 결과를 기대할 수 있을 것이다.

[7] <http://www.sdmi.org>

4. 결론

본 논문에서는 반향과 대역 확산 기법을 결합한 계층적 오디오 워터마크 시스템을 제안하고 다양한 신호처리에 대하여 검출 성능을 제시하였다. 1계층 반향 워터마크로 워터마크의 유, 무 정보와 오디오 데이터에 가해진 선형 속도 변화를 추정할 수 있고, 그 검출 결과를 바탕으로 부가 정보량이 많은 2계층 pn 수열 워터마크를 선택적으로 검출할 수 있다. 2계층 워터마크는 프레임의 동기 확보를 위해 동기용 pn 수열과 데이터용 pn 수열을 분리하여 검출부에서 데이터에 가해진 변형을 추정하는데 용이하게 하였다. 본 논문에서 제안한 계층적 오디오 워터마크 시스템은 오디오 데이터에 가해진 시간 변형을 추정하여, 워터마크 데이터의 검출을 강인하게 할 뿐 아니라 시스템의 요구 사항에 따라 선택적으로 계층을 검출하여 복잡도를 조절할 수 있는 융통성을 가진다. 추후 워터마크를 삽입할 때 적응적으로 전력 이득을 변화시키고, 계층별 검출 결과에 대한 해석을 보강한다면, 보다 정교하고, 강인한 워터마크 시스템이 될 것으로 기대된다.

감사의 글

본 논문은 정보통신부 지원 “디지털 방송 콘텐츠 보호 및 관리기술 개발” 과제의 결과로서 정보통신부 관계자에게 감사의 글을 전합니다.

참고 문헌

- [1] F. Petitcolas, R. Anderson, and M. Khun, “Information hiding - a survey.” *Proc. IEEE*, July 1999, pp. 1062-1078.
- [2] L. Boney, A. H. Tewfik, and K. N. Hamdy, “Digital watermarks for audio signals,” *IEEE Proc. Multimedia* pp. 473-480, 1996.
- [3] J. W. Seok and J. W. Hong, “Audio watermarking for copyright protection of digital audio data,” *Electronics Letters*, pp. 60-61, Jan. 2001.
- [4] B. A. Garcia, “Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory,” *In 107th AES convention*, New York, AES. Preprint 5073, Sep. 1999.
- [5] 정사라, 석종원, 홍진우, “시간 변형에 강인한 오디오 워터마크 알고리즘,” *한국음향학회* 제 20 권 제 1 호, pp. 879-882, 2001.
- [6] D. Gruhl, W. Bender, and A. Lu, “Echo hiding,” in *Information hiding: 1st Int. Workshop (Lecture Notes in Computer Science)*, vo. 1174, R. J. Anderson, Eds. Berlin, Germany: Springer-Verlag, pp. 293-315, 1996.