

시간축 변형을 고려한 디지털 오디오의 계층적 워터마크

Robust Layered Watermarking of Digital Audio for Possible Timing Changes

정사라*, 홍진우*
(Sa-Rah Jung*, Jin-Woo Hong*)

*한국전자통신연구원 무선방송연구소 방송미디어부

(접수일자: 2002년 7월 26일; 수정일자: 2002년 10월 5일; 채택일자: 2002년 10월 28일)

본 논문에서는 디지털 오디오 신호에 부가 정보를 삽입하는 기술로써 계층적 워터마크를 사용하여 시간축 변형을 감지할 수 있고, 필요에 따라 검출 복잡도를 조절할 수 있는 기법을 제안한다. 1계층 워터마크는 오디오 신호의 시간축 변형을 감지할 수 있도록 오디오 신호의 반향을 이용하고, 2계층 워터마크는 1계층 워터마크된 오디오 신호를 기준으로 대역 확산 기법을 이용하여 저작권 정보 등의 요구량이 많은 부가 정보를 삽입한다. 이 때, 2계층 워터마크는 프레임의 동기 확보를 위한 동기 수열, 다른 하나는 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.

핵심용어: 워터마크, 워터마킹, 오디오, 반향, 대역확산, 계층, 시간축 변형, 검출

투고분야: 음향 신호처리 분야 (1,2)

In this paper, we present a layered watermarking technique for digital audio data that is capable of detecting timing change and adapting complexity in detection. The proposed watermarking uses echo hiding as the first layer, which enables the detector to estimate linear speed change. The spread spectrum watermark is then inserted in the second layer which includes additional information like copyright data. We use two kinds of sequences in the second layer, one of which is for synchronization and the other is for data. The results of previous layer are used to make estimate of timing change in the next layer. The detector in the presented method can select detecting range from the first layer to the first layer, second pre-layer, or second main-layer due to the required system specification. Experimental results show that the proposed watermarking technique is robust to several processing attacks including timing change.

Keywords: Watermark, Watermarking, Digital audio, Echo, Spread spectrum, Layer, Timing change, Detection

ASK subject classification: Acoustic signal processing (1,2)

I. 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 신호에 지각적으로 인식되지 않을 만큼 작은 변화를 주면서, 의도적이거나 또는 비의도적인 여러 가지 신호 처리에도 살아남아야 하므로 본질적으로 비인 지성과 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 기술이란 협대역의 강한 신호를 광대역의

책임저자: 정사라 (sjung@etri.re.kr)
335-350 대전광역시 유성구 가정동 161
한국전자통신연구원 방송미디어부
(전화: 042-860-1558; 팩스: 042-860-5479)

약한 신호로 대체하여 전송한 후 수신단에서 복원하여 데이터를 검출하는 기술이다. 광대역 신호를 모르는 제 3자에 대해서는 잡음과 같아서 쉽게 발견되지 않으며, 주파수 대역이 넓게 퍼져 있어서 협대역 신호 처리나 잡음에 별로 영향을 받지 않는 장점이 있다. 따라서 대역 확산 기술은 주파수 동화(equalize), 고음질(고화질) 압축, 잡음 첨가 등 일반적인 신호 처리에 강인하며, 특히 약한 전력의 신호로 전송할 수 있는 점은 워터마크로서 좋은 특성이 된다. 그러나 이러한 대역 확산 워터마크를 검출하기 위해서는 신호 프레임의 동기 확보가 필수적이기 때문에 시간축 변형(기하학적 변형)과 같이 전송된 신호의 주기가 달라질 때에는 살아남지 못한다. 오디오 신호는 방송 시간을 맞추거나 조절하기 위해서 시간 폭을 변형할 때가 있는데 이러한 변화를 고려하지 않은 단순한 시스템에서는 대역 확산 워터마크를 제대로 검출해 낼 수 없다.

본 논문에서는 이러한 시간축 변형에 대응하기 위해 반향 워터마크와 pn (pseudo noise) 대역 확산 워터마크를 결합한 계층적 워터마크 알고리즘을 제안한다. 1계층 워터마크로 반향을 삽입하여 검출기에서 워터마크의 유, 무 정보와 선형 속도 변화량(linear speed change)을 추정하고, 2계층 워터마크로 pn 수열을 이용하여 정보량이 많은 데이터를 삽입한다. 이 때, 데이터 삽입을 위한 데이터용 pn 수열 외에 프레임 동기를 위한 동기용 pn 수열을 두어서 검출부에서 동기 문제를 용이하게 해결할 수 있다. 1계층에서 추정된 선형 속도 변화량을 이용하여 2계층 전단계에서는 프레임 동기를 확보하고, 본단계에서는 앞 계층들에서 추정된 시간축 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 기존 워터마크 시스템의 문제점 및 본 논문에서 제안한 해결 방안에 대해 간단히 기술하고, 3장에서 제안한 계층적 오디오 워터마크의 삽입 시스템과 검출 시스템을 설명한다. 4장에서는 이 시스템에 대해 실험을 수행하여 결과를 분석하고, 5장에서 결론을 맺는다.

II. 기존방식의 문제점 및 해결 방안 제안

대역 확산을 이용한 워터마크의 기본적인 개념은 전송할 데이터를 pn 수열로 변조한 후 pn 수열을 삽입하고, 검출부에서는 입력 신호에 대해 pn 수열 정합 필터를 통과

시켜 그 값으로 워터마크 데이터를 추출하는 것이다. 즉 삽입된 수열과 정합 필터의 수열이 같고, 동기가 정확하게 일치해야 올바른 검출값을 얻을 수 있다. 수열이 같더라도 동기가 불일치 할 경우, 워터마크의 검출은 이루어지지 않는다. 동기가 어긋난 경우에 대비해 [4], [5]에서는 각각 특정 해더 수열을 이용하거나, 두 수열의 부호를 이용하여 프레임의 동기를 확보하는 방법을 제안하였다. 그런데 이 두 방법은 모두 프레임의 동기를 찾기 위해 다음 (1)식과 같은 샘플별 교차상관(cross correlation) 값의 계산을 필요로 한다. 여기서 $x(n)$ 은 검출부의 수열, $w(n)$ 은 입력 오디오 신호, N 은 프레임당 샘플수에 해당한다. $w^*(n)$ 은 $w(n)$ 의 공액복소수의 의미가 있으나 여기서는 $w(n)$ 이 실수이므로 사실상 $w(n)$ 과 동일하다.

$$r_{xw}(l) = \sum_{n=0}^{N-1-l} x(n)w^*(n-l) \quad (1)$$

이 때, $r_{xw}(0)$ 를 구하기 위한 입력 오디오 신호 $w(n)$ 의 샘플 범위는 0에서 $N-1$ 까지이고, $r_{xw}(N-1)$ 에 대해서는 $-(N-1)$ 에서 0까지의 입력 샘플이 필요하다. 즉, 교차상관값을 구하기 위해서는 입력 오디오 신호가 샘플씩 쉬프트되어야 하고, 한 프레임에 대한 교차상관값을 구하기 위해서는 2프레임의 입력 신호가 요구된다. 다시 말해서 한 프레임의 샘플 교차 상관값을 얻기까지는 한 프레임에 해당하는 지연시간이 발생하며, 필요한 곱셈 계산량은 N^2 이 된다. 본 논문에서는 프레임 교차상관값의 계산을 효율적으로 처리하기 위해 동기용 pn 수열을 사용한다. 동기용 pn 수열은 비트 정보에 상관없이 일정하게 삽입되므로 수식 (1)의 교차상관을 수식 (2)의 회전식(circular) 교차상관으로 얻을 수 있다.

$$\begin{aligned} \widetilde{r}_{xw}(l) &= \sum_{n=0}^{N-1} x(n)w^*((n-l))_N \\ \xleftrightarrow{DFT} \widetilde{R}_{xw}(k) &= X(k)W^*(k) \end{aligned} \quad (2)$$

$(n)_N$ 은 n 을 N 으로 나눈 나머지이며, $\widetilde{R}_{xw}(k)$, $X(k)$, $W(k)$ 는 각각 $\widetilde{r}_{xw}(n)$, $x(n)$, $w(n)$ 의 푸리에 변환한 값이다. 수식 (2)에 의하면, $x(n)$, $w(n)$ 을 각각 푸리에 변환한 후 곱하고, 그 값을 다시 역 푸리에 변환하여 한 프레임 전체의 교차상관값을 얻을 수 있다. N 샘플 푸리에 변환에 필요한 곱셈 계산량이 $N \cdot \log_2 N$ 이므로, 이 경우 곱셈 계산량은 $N \cdot (3 \log_2 N + 1)$ 로서 전송한 경우에 비해 상당한 정도로 줄일 수 있으며 지연시간도 발생하지 않는다. 그런데 이와 같이 수열의 동기를 확보하는

방식은 수열의 주기가 변화되지 않은 것을 전제로 한다. 즉, 오디오 신호에 선형 속도 변화가 있어서 수열의 주기 자체가 변할 경우에는 대처할 수가 없다.

오디오 신호의 선형 속도 변화는 영상의 크기 변화(resize)에 해당되는 것으로 오디오 신호 전체에 대해서 길이를 변화시키는 처리 과정이다. 예를 들어 (-1, 1)인 수열이 2배의 선형 속도 변화를 겪으면, (-1, -1, 1, 1)이 되면서 크기가 2배가 된다. 오디오 신호에 반향을 삽입하고 검출기에서 그 반향의 위치 변화를 관측하면, 속도 변화량이 측정된다[6, 7]. 예를 들어 10 ms 위치에 반향을 삽입한 신호에 10% 선형 속도 변화가 가해질 경우 검출시 반향 위치는 11 ms로 나타난다. 그런데 시간축 변형에는 이렇게 수열의 주기가 변화하는 선형 속도 변화 외에 피치를 변화시키지 않는 시간 늘임 변형이 있다. 이것은 오디오 신호의 주파수 고유 특성을 그대로 유지하면서 주기적으로 일정 시간이 증첩 또는 소실되는 형태로 나타난다. 예를 들어 -1이 5번, 1이 5번 발생하는 (-1, -1, -1, -1, -1, 1, 1, 1, 1, 1)과 같은 수열이 있을 때, 10% 시간 늘임 변형이 가해지면, 중간에 한 개 숫자가 소실되면서 -1이 4번, 1이 5번인 (-1, -1, -1, -1, 1, 1, 1, 1, 1)의 형태 등으로 나타나는 것이다. 따라서 수열의 주기는 변하지 않지만, 동기의 위치가 달라지게 된다. 이 경우는 신호가 균일하게 변화를 겪는 것이 아니라 중간에 일정 프레임이 소실된 것이므로 반향으로는 감지하기 힘들고, 오히려 대역 확산 기법의 pn 수열의 교차 상관값을 여러 프레임에 걸쳐 구한 후 그 최대값의 위치 분포로 동기 위치를 찾을 수 있을 것이다.

본 논문에서는 대역확산을 이용한 워터마킹 방식에서 시간축 변형에 대처하기 위한 방안으로 2계층으로 워터마크를 삽입하고 검출하였다. 먼저 1계층 워터마크로 반향을 삽입하였는데, 이것은 검출시에 선형 속도 변화량을 측정함으로써 워터마크의 유, 무 검출과 함께 pn 수열의 주기를 보정하기 위한 것이다. 다음 2계층으로는 동기용 pn 수열과 데이터용 pn 수열이 삽입되었다. 동기용 pn 수열은 검출시 2계층 전단계 과정에서 교차상관값으로 프레임의 동기를 확보하기 위한 것이다. 이것은 또한 워터마크의 유, 무 검출에도 사용될 수 있다. 마지막으로 데이터용 pn 수열은 1계층, 2계층 전단계에서 확보한 동기 위치를 기준으로 삽입된 데이터의 내용을 검출한다.

III. 제안한 계층적 워터마크

3.1. 워터마크 삽입부

그림 1에 제안한 워터마크의 삽입 시스템을 나타내었

다. 먼저 입력 디지털 오디오 신호에 1계층 워터마크로 반향을 삽입한다. 즉, 원래의 입력 신호를 $s(n)$ 이라 할 때 1계층 워터마크된 신호 $w_1(n)$ 는 다음 식으로 주어진다.

$$w_1(n) = s(n) + \alpha \cdot s(n - \sigma) \tag{3}$$

여기서 α 는 크기 조정 계수, σ 는 반향의 지연 시간에 해당되는 것으로 비인지성이 유지되는 한도로 적절하게 조정한다. 보통 α 는 0.3~0.8이고, σ 는 수 내지 수십 밀리초의 범위이다. 1계층 워터마크된 신호에 대해서 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 오디오 프레임을 AR (Auto Regressive) 모델한 후, 선형 예측 계수 (linear predictive coefficient)를 구하여 이용하였다. 여기서 선형 예측 계수는 식 (4)로 얻어진 계수 $\{a_k | k=1, 2, \dots, p\}$ 이며, p 는 선형 차수에 해당한다.

$$w_1(n) \approx \sum_{k=1}^p a_k \cdot w_1(n-k) \tag{4}$$

(4)의 수식으로 얻어진 선형 예측 계수로 pn 수열 $x(n)$ 을 필터링하여 $\tilde{x}(n)$ 을 얻는 과정은 다음과 같다.

$$\tilde{x}(n) = \gamma \cdot x(n) + \sum_{k=1}^p a_k \cdot \tilde{x}(n-k) \tag{5}$$

$n=0, 1, \dots, N-1$

여기서 γ 는 크기 조정 계수이며, N 은 프레임 길이에 해당한다. 이 과정으로 데이터용 pn 수열 $x_d(n)$, 동기용

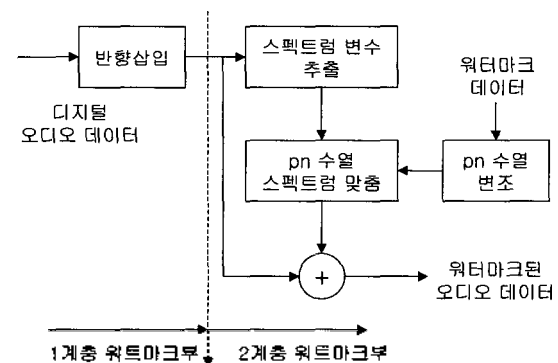


그림 1. 계층적 오디오 워터마크 삽입 시스템
Fig. 1. Embedding system of the proposed layered audio watermark.

pn 수열 $x_s(n)$ 을 필터링하여 각각 $\tilde{x}_d(n)$, $\tilde{x}_s(n)$ 을 얻는다. 선형 예측 필터링에 대한 설명은 3.2절의 워터마크 검출 부분에서 좀 더 보충하기로 한다. 삽입하려는 워터마크 데이터의 비트 정보 '1', '0'에 따라 데이터용 pn 수열 $\tilde{x}_d(n)$ 부호의 양, 음을 조절하고, 동기용 pn 수열 $\tilde{x}_s(n)$ 은 데이터에 관계없이 항상 같은 부호로 삽입된다. 따라서 2계층 워터마크된 최종 출력 신호 $w_2(n)$ 은 다음과 같다.

$$w_2(n) = w_1(n) + \tilde{x}_s(n) + \tilde{x}_d(n), \quad data = 1,$$

$$w_2(n) = w_1(n) + \tilde{x}_s(n) - \tilde{x}_d(n), \quad data = 0 \quad (6)$$

3.2. 워터마크 검출

가. 오디오 신호 전처리

그림 2에 제안한 워터마크의 검출 시스템을 나타내었다. 일반적으로 삽입된 pn 수열의 크기가 오디오 신호에 비해 상당히 작으므로 검출 성능을 향상시키기 위해 여러 가지 전처리 필터링을 수행하는데, 본 시스템에서는 [3]에서 제안된 역 선형 예측 필터 (inverse linear predictive filter)를 이용하였다. 즉, 검출부의 입력 오디오 신호 $w(n)$ 을 AR 모델로 선형 예측 계수를 구한 후, 그 계수로 신호를 필터링하여 오디오 신호의 영향을 크게 줄였다. $w(n)$ 과 잔차 신호 $\tilde{w}(n)$ 의 관계는 다음 식과 같다.

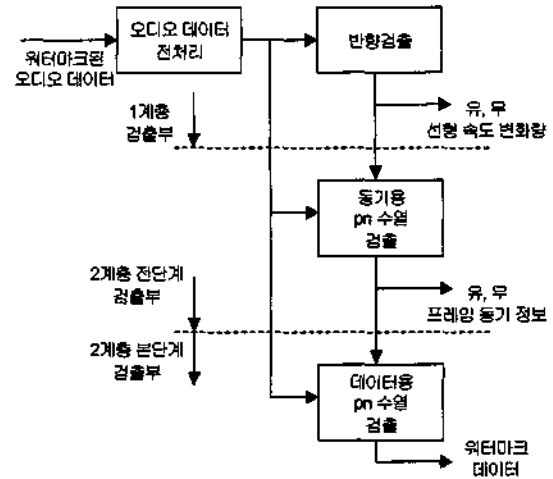


그림 2. 계층적 오디오 워터마크 검출 시스템
Fig. 2. Detecting system of the proposed layered audio watermark.

$$\tilde{w}(n) = w(n) - \sum_{k=1}^p \tilde{a}_k \cdot w(n-k) \quad (7)$$

여기서 선형 예측 계수 $\{\tilde{a}_k | k=1, 2, \dots, p\}$ 는 식 (4)에서 $w(n)$ 을 대입하여 얻은 값이다. 역 선형 예측 필터는 입력 오디오 신호를 백색화 (whitening) 시키므로 필터 출력 후 워터마크 데이터의 상대적인 에너지비를 증가시키는 효과를 볼 수 있다. 특히 본 시스템에서는 pn 수열이 삽입부에서 선형 예측 필터링되고, 검출부에서는 역 선형 예

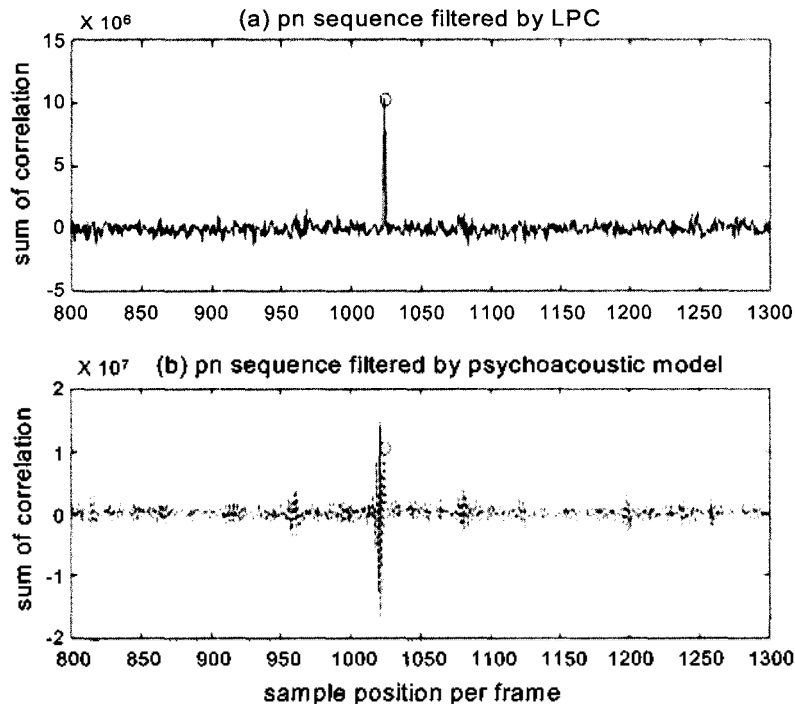


그림 3. 교차상관값의 비교: (a) 제안된 워터마크 시스템, (b) 심리음향모델로 처리하여 삽입한 경우
Fig. 3. Comparison of correlation property: (a) proposed watermarking, (b) watermarking using psychoacoustic model.

를 필터링되므로 신호 처리 공격이 없다면, 거의 원형으로 복구되어 검출부의 교차 상관 (cross correlation)을 향상시키는 효과까지 볼 수 있다.

pn 수열이 겪는 처리 과정이 검출 성능에 미치는 효과를 확인하기 위해서 제안된 워터마킹 시스템과 pn 수열을 심리음향 모델로 처리하여 삽입한 경우를 비교하였다. 일반적으로 심리음향 필터링을 할 경우 저대역통과 (Low-pass) 필터링이나 압축 부호화 과정에 대비해서 고대역 부분을 미리 저지 필터링하는데, 여기서는 11.05 kHz대의 스펙트럼 값으로 그 이후 대역의 스펙트럼 값을 고정하였다. 그림 3은 이 두 방식으로 워터마크된 오디오 신호를 검출부에서 전처리한 후, 원래의 pn 수열과 교차상관을 구하여 그 값을 위치별로 더한 결과이다. 이 때 더한 프레임의 개수는 1280이며, 사용한 AR필터의 차수 p 는 8이다. 두 경우 SNR은 각각 제안된 삽입 방식이 25.32 dB, 심리음향 필터링을 이용한 방식이 24.35 dB로 평균 잡음 전력을 거의 같게 하였다. 그림 3을 관측하면, 제안된 방식에서는 정확히 동기 위치에서 상관값의 피크가 발생하고, 이것이 뚜렷이 구분되는데 비해 삽입할 때 심리음향 필터링한 경우, 동기 위치 전후로 상관값이 크게 나타나서 피크가 두드러지지 않는다. 즉, 제안된 워터마킹 시스템에서 pn 수열을 삽입-검출할 때, 처리 과정이 상호 보상되어 검출시 pn 수열의 교차 상관 특성을 잘 이용할 수 있음을 확인할 수 있다.

나. 데이터 검출

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

1계층 검출이 완료되면, 선택적으로 2계층 전단계 검출이 시작되는데, 1계층 반향 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 신호 (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.

$$\max_l \{ \tilde{r}_{xw}(l) : l \in \{0, 1, \dots, N-1\} \} \quad (8)$$

$$\text{단, } \tilde{r}_{xw}(l) = \sum_{n=0}^{N-1} \tilde{x}_s(n) \tilde{w}^*((n-l)_N)$$

여기서 $\tilde{r}_{xw}(l)$ 은 2장에서 설명한 회전식 교차 상관값이고, N 은 1계층 검출부에서 보정된 프레임당 샘플 수, $\tilde{x}_s(n)$ 은 길이 보정된 동기용 pn 수열, $\tilde{w}(n)$ 은 전처리된 오디오 신호이다.

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 신호) 데이터용 pn 수열의 교차 상관을 구한다. 교차 상관의 부호로 '1', '0'의 비트 정보를 결정한다. 이 과정을 수식으로 표현하면 아래와 같다.

$$\frac{c_{sw}}{\sqrt{E_x} \sqrt{E_w}} \geq 0, \quad \text{data} = 1, \quad (9)$$

$$\text{otherwise} \quad \text{data} = 0$$

단, $c_{sw} = \sum_{n=0}^{N-1} \tilde{x}_d(n) \tilde{w}^*(n-l)$, $E_x = \sum_{n=0}^{N-1} x^2(n)$. 이 때, $\tilde{x}_d(n)$ 은 길이 보정된 데이터용 pn 수열이며, l 은 (8)에서 구한 동기 위치이다.

IV. 실험 결과 및 분석

제안된 워터마킹 알고리즘의 성능을 검증하기 위해 강인성 테스트를 실시하였다. 실험에서 사용한 오디오 신호는 표본주파수 44.1 kHz, 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. 편의상, 2계층 본단계까지 검출하는 것을 원칙으로 하되, 30초씩 2계층 전단계까지의 결과를 분석한 후 그것을 바탕으로 2계층 본단계를 검출하고, 다음 30초의 데이터를 검출하는 형태로 이루어졌다. 실험 환경은 펜티엄III 500 MHz, 512M램 PC의 MATLAB 윈도우 환경이며, 30초 오디오 신호에 대해 삽입 소요시간은 9초, 2계층 전단계까지의 검출 시간은 12초, 2계층 본단계 검출 시간은 7초로 총 검출 시간은 19초가 소요되었다.

먼저, 오디오 신호에 가해진 신호 처리가 워터마크 검출에 미치는 영향을 보기 위해서 워터마크를 비교적 강하게 삽입하여 (SNR = 13.59 dB) 2계층 전단계의 결과를

살펴보았다. 2계층 전단계는 동기 확보가 목적이므로 프레임 단위로 피크 위치를 구한 후 30초 후 샘플 위치 별로 피크 발생의 빈도수를 정리하였다. 예컨대 워터마크가 되어 있고 아무런 공격이 없으면, 교차 상관값의 피크는 수열의 동기가 일치하는 샘플값에서 발생하며, 이 피크가 발생하는 위치는 매 프레임마다 같을 것이다. 즉 워터마크된 오디오 신호의 프레임 수가 1000이라고 할 때, 첫 번째 샘플 위치가 동기 위치이면 샘플값 1에서 피크 발생 빈도가 거의 1000에 가깝고 그 외의 위치에서는 매우 빈도가 낮을 것이다. 한편 워터마크가 없거나 워터마크가 되어 있더라도 오디오 신호에 시간 폭으로 심한 변형이 있을 경우에는 수열의 동기 위치가 불분명해진다. 따라서 피크의 발생은 수열 주기 내의 모든 샘플값에 대해 낮

은 빈도로 나타날 것을 예상할 수 있다. 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 결과 즉, 샘플 위치 별 피크 발생 빈도의 결과는 각각 그림 4~7와 같다.

그림 4의 워터마크가 없는 데이터에 대해서는 피크 발생 위치가 무작위이므로 각 샘플 위치당 5번 내외의 빈도수로 골고루 분포한다. 반면, 워터마크가 있을 때는 그림 5에서 확인 할 수 있듯이 피크는 한 위치에 거의 일정하게 나타난다. 따라서 이러한 피크 발생 빈도로써 워터마크 유, 무를 분명하게 판단할 수 있으며 동기의 위치도 확인할 수 있다.

공격이 없을 때인 그림 5와 10% 선형 속도 변화가 가해

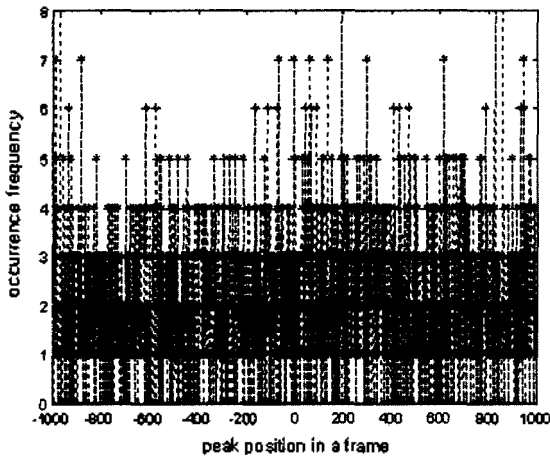


그림 4. 워터마크가 없을 경우 샘플 위치당 피크 발생 빈도
Fig. 4. The frequency of peak occurrence at each sample position without watermark.

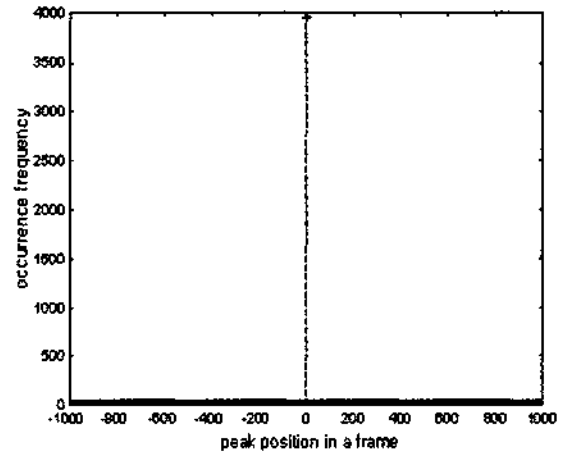


그림 5. 워터마크된 데이터의 샘플 위치당 피크 발생 빈도
Fig. 5. The frequency of peak occurrence at each sample position with watermark.

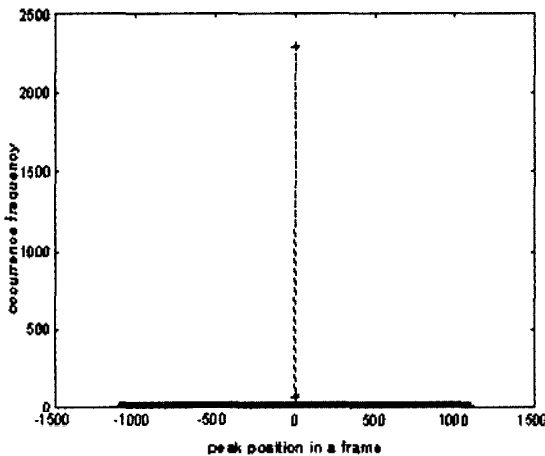


그림 6. 워터마크된 데이터에 10% 선형 속도 변화가 가해진 경우 샘플 위치당 피크 발생 빈도
Fig. 6. The frequency of peak occurrence at each sample position with +10% linearly speed changed watermark.

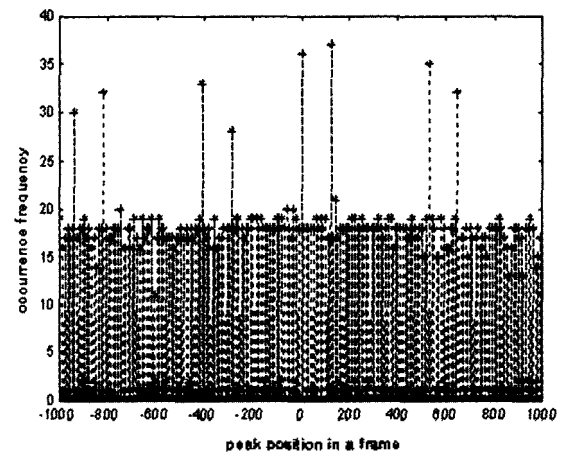


그림 7. 워터마크된 데이터에 4% 시간 눈금 변형이 가해진 경우 샘플 위치당 피크 발생 빈도
Fig. 7. The frequency of peak occurrence at each sample position with +4% time scaling modified watermark.

진 그림 6을 비교하면, 거의 비슷한 형태인 것을 관측할 수 있다. 다만 피크가 발생한 샘플 위치의 범위가 (-1000 1000)에서 (-1100, 1100)으로 10% 증가한 것이 다른 점이다. 이것은 선형 속도 변화에 의해 pn 수열의 주기가 10% 늘어났기 때문이다. 한편 시간 눈금 변형이 일어난 그림 7을 관찰하면 언뜻 보기에는 워터마크가 없는 그림 4의 경우와 비슷하게 보이지만, 자세히 관측하면 조금 다른 것을 확인할 수 있다. 그림 4는 모든 샘플 위치에서 발생 빈도수가 5번 내외로 골고루 분포하는 것에 비해서 그림 7은 5개 이내의 분포를 갖는 것과 10개~30개 정도의 빈도수를 갖는 것이 구별되어 있다. 두 경우의 차이는 프레임별로 피크 위치를 나타낸 그림 8에서 더욱 확연해진다. 즉 시간 눈금 변형이 일어난 경우 피크 위치가 주기적으로 변하는 것을 확인할 수 있는데, 이것은 시간 눈금 변형이 오디오 신호를 균일하게 변화시키는 것이 아니라 일정한 주기로 프레임을 중첩, 소실시키는 형태로 일어나기 때문이다. 즉 프레임이 중첩 또는 소실될 때 동기 위치는 이동하며, 이후에 다시 프레임의 중첩, 소실이 발생할 때까지 동기 위치가 유지되므로 동기 위치를 추적하면 그림 8의 (b)와 같은 결과를 얻게 된다. 반면 워터마크가 없는 그림 8의 (a)에서는 주기성이 없이 무작위로 나타나게 된다.

그림 4~7의 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 빈도를 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 13.59 dB일 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[8] 컨피덴셜 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 17.11 dB일 때는 변형을 감지하고 동기 위치를 변화시키며 데이터를 검출할 수 있었지만, SNR이 23.13 dB일 때는 워터마크 강도가 약하기 때문에 동기 위치 주변으로 피크가 분산되어서 2계층 전단계 검출 결과를 분석하는 과정에서 시간축 변형을 감지하지 못하고 큰 변형을 겪고 피크가 무작위로 분포된 (5)번으로 판정되었다.

표 1의 실험 결과 본 논문에서 제안한 계층적 워터마크 시스템이 신호 처리에 대해서 비교적 강인한 것을 확인할

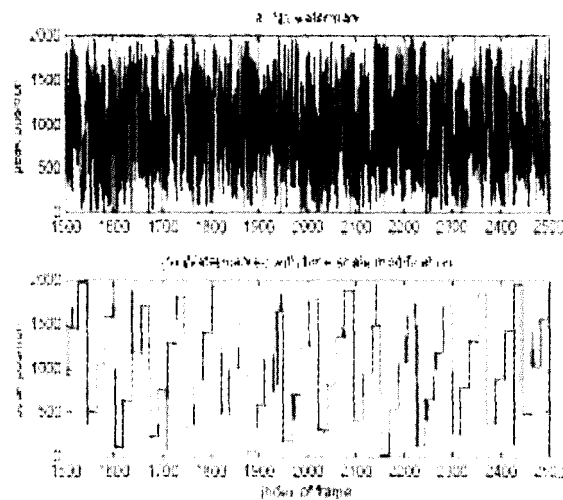


그림 8. 프레임별 피크 위치 비교: (a) 워터마크가 없을 때, (b) 4% 시간 눈금 변형이 가해진 경우
Fig. 8. Comparison of peak position at each frame: (a) without watermark, (b) +4%time scaling modified watermark.

표 1. 다양한 신호 처리에 따른 BER 검출 성능
Table 1. Bit error rate (BER) performance of proposed watermarking about various signal processing attacks.

신호처리	BER (SNR=17.11dB)	BER (SNR=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	0.0	1.892e-002
중간대역 필터	4.288e-003	8.172e-002
선형 속도 변화	5.058e-004	2.807e-002
시간 눈금 변형	변형 감지	변형 감지 불가
반향 첨가	8.323e-003	7.238e-002

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 시간 눈금 변형이 일어난 경우, 피크의 분포 때문에 아예 검출하지 못하는 결과가 발생할 수 있다. 이렇게 커버 오디오 신호의 영향으로 피크가 그 전후 주변으로 퍼져서 분포하게 될 경우를 대비하여 피크 히스토그램에 대한 보다 정교한 해석이 필요하다. 마지막으로 SNR이 17.11 dB일 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 신호의 특성에 따라 적응적으로 워터마크의 전력 이득을 변화시킨다면, BER 성능과 음질면에서 좀 더 향상된 결과를 기대할 수 있을 것이다.

V. 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 신호에 가해진 선형 속도 변화를 추정할 수 있고, 그 검출 결과를 바탕으로 부가 정보량이 많은 2계층 pm 수열 워터마크를 선택적으로 검출할 수 있다. 2계층 워터마크는 프레임의 동기 확보를 위해 동기용 pm 수열과 데이터용 pm 수열을 분리하여 검출부에서 데이터에 가해진 변형을 추정하는데 용이하게 하였다. 본 논문에서 제안한 계층적 오디오 워터마크 시스템은 오디오 신호에 가해진 시간축 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

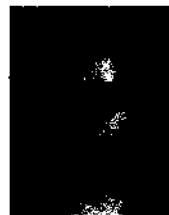
본 논문은 정보통신부 지원 디지털 방송 콘텐츠 보호 및 관리기술 개발 과제의 결과로서 정보통신부 관계자에게 감사의 글을 전합니다.

참고 문헌

1. F. Petitcolas, R. Anderson, and M. Khun, "Information hiding a survey," *Proc. IEEE*, 1062-1078, July 1999.
2. L. Boney, A. H. Tewlik, and K. N. Hamdy, "Digital watermarks for audio signals," *IEEE Proc. Multimedia*, 473-480, 1996.
3. J. W. Seok and J. W. Hong, "Audio watermarking for copyright protection of digital audio data," *Electronics Letters*, 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 학술발표대회논문집 20 (1), 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, 293-315, 1996.
7. 오현오, 김현욱, 윤대희, 석종원, 홍진우, "강인한 오디오 워터마크를 위한 새로운 반향 커널 설계," 한국음향학회지 20 (2), 68-76, 2001.
8. <http://www.sdmi.org>

저자 약력

● 정 사 라 (Sa-Rah Jung)



1998년 2월: 한국과학기술원 전기 및 전자공학과 (공학사)
 2000년 2월: 한국과학기술원 전기 및 전자공학과 (공학석사)
 2000년 12월 ~ 현재: 한국전자통신연구원 방송미디어부 연구원
 * 주관심분야: 오디오 신호처리, 디지털 콘텐츠 보호 및 관리, 통신 신호처리

● 홍 진 우 (Jin-Woo Hong)



1978년 3월 ~ 1982년 2월: 광운대학교 응용전자공학과 졸업 (공학사)
 1982년 3월 ~ 1984년 2월: 광운대학교 대학원 전자공학과 졸업 (공학석사)
 1990년 3월 ~ 1993년 8월: 광운대학교 대학원 전자계산기공학과 졸업 (공학박사)
 1998년 9월 ~ 1999년 2월: 독일 프라운호퍼연구소 (파견연구원)
 1984년 3월 ~ 현재: 한국전자통신연구원 방송콘텐츠 연구팀장 (책임연구원)

2000년 1월 ~ 현재: 한국음향학회 홍보이사, 뉴미디어융합 학술분과위원장, 한국방송공학회 편집위원
 1993년 1월 ~ 현재: 정보통신부주노하연구단 방송기술위원회 위원
 2001년 6월 ~ 현재: SEDICA 운영위원
 * 주관심분야: 오디오 신호처리 및 부호화, 디지털 콘텐츠 보호 및 관리, 디지털 오디오 방송