# 토널 특성을 이용한 브라인드 오디오 워터마킹

이희숙[†] · 이우선[††]


## 요 약

이 논문에서는 토널 특성을 이용한 브라인드 오디오 워터마킹을 제안한다. 먼저 기존의 심리음향연구를 통해 토널의 인지영향에 대해 살펴보고, 토널 성분이 여러 신호처리 후 변동측면에서 매우 안정적인 특성을 가짐을 다른 워터마크에 이용되는 특성들과 비교하여 보였다. 이를 기반으로 토널 마스커를 구성하는 주파수 신호들의 관계를 이용한 브라인드 오디오 워터마킹(blind audio watermarking) 기법을 제안하였다. 이 기법이 적용된 오디오에 대한 SDG(Subjective Diff-Grades) 음질평가에서 평균 SDG 0.27의 결과를 얻었고 이는 비지각성 면에서 토널의 인지 영향을 이용한 워터마킹이 유용하다고 볼 수 있다. 또한 time shift를 제외한 여러 신호처리 후의 워터마크 추출 결과는 98%이상으로 제안한 워터마킹의 강인성을 보였다. Time shift처리에 대해서는 시간 축 상에서 최적의 위치를 찾아 추출하는 새로운 방법을 적용하여 추출율 90%의 결과를 얻었다.


# A Blind Audio Watermarking using the Tonal Characteristic

Hee-Suk Lee[†] and Woo-Sun Lee[††]


## ABSTRACT

In this paper, we propose a blind audio watermarking using the tonal characteristic. First, we explain the perceptional effect of tonal on the existed researches and show the experimental result that tonal characteristic is more stable than other characteristics used in previous watermarking studies against several signal processing. On the base of the result, we propose the blind audio watermarking using the relation among the signals on the frequency domain which compose a tonal masker. To evaluate the sound quality of our watermarked audios, we used the SDG(Subjective Diff-Grades) and got the average SDG 0.27. This result says the watermarking using the perceptional effect of tonal is available from the viewpoint of non-perception. And we detected the watermark bits from the watermarked audios which were changed by several signal processing and the detection ratios with exception of the time shift processing were over 98%. About the time shift processing, we applied the new method that searched the most proper position on the time domain and then detected the watermark bits by the ratio of 90%.


**Key words:** audio watermarking, psychoacoustic model, tonal

## 1. 서 론

오디오 워터마킹은 오디오 데이터 내에 음질의 손상을 최소화하는 방법으로 부가적인 정보를 내장하여 전송하는 기술이다. 오디오 워터마킹의 응용분야로는 일반적인 부가정보의 전송, 방송시스템의 상업적 목적을 위한 특정정보의 전송, 그리고 지적소유권 보호를 위한 정보의 내장 등을 들 수 있다[1]. 저작권 보호를 위한 적용의 경우, 오디오 워터마킹 기법의 여러 평가기준 중 다음 조건을 우선적으로 만족할 수 있어야 한다.

1. 워터마킹 후, 오디오의 음질이 보장되어야 한다(inaudibility).

2. 일반적인 오디오 처리나 고의적인 공격에 대해 워터마크 정보를 유지해야 한다(robustness).

접수일 : 2002년 11월 6일, 완료일 : 2003년 2월 6일

[†] 정회원, 창원대학교 컴퓨터공학과

[††] 정회원, 창원대학교 컴퓨터공학과 교수

초기의 오디오 워터마킹 연구는 LSB삽입기법, 패치워크(patchwork) 기법 등이 제안되었으나 여러 신호처리 후 워터마크의 손실이 매우 크다는 단점을 보였다. 이후 여러 처리와 공격에 강인하기 위하여 신호의 변화에 비해 상대적으로 덜 지각되는 특성을 이용하여 강인한 워터마크를 내장하는 워터마킹 기법이 제안되었으며 그 예로 echo hiding, phase coding, 심리음향모델(psychoacoustic model)을 이용한 spread spectrum 삽입 기법 등이 있다[2-4]. 작은 세그먼트마다 인지 한계이하의 시간지연 echo를 삽입함으로써 정보를 내장하는 echo hiding은 음질과 압축, 노이즈, 필터링 등의 여러 처리에 대한 강인성을 보였다[2,4]. 그러나 echo의 추출에 의한 워터마크의 노출과 echo 제거가 가능하다는 단점이 있다. phase coding은 짧은 구간내의 위상의 변화를 이웃 구간과의 급격한 차이만 없다면 인지하지 못한다는 특성을 이용한 기법이다. 이 기법은 음질면에서 매우 효과적인 워터마킹 기법이다. 그러나 MPEG 코딩과 같은 신호처리에 위상정보가 쉽게 변하는 단점을 가진다[5]. Psychoacoustic model을 이용한 spread spectrum 기법은 워터마크를 noise-like 형태로 변환한 후 마스킹 효과를 고려하여 각 주파수별로 마스킹되는 에너지이하의 크기로 내장하는 방법이다[3,6,7]. 참고문헌[3]은 워터마크 추출방법으로 원음 공개 워터마크 추출(non-blind watermark detection)을 제안하였다. 저작권의 보호를 전제로 하는 여러 응용에서는 비공개 워터마크 추출(blind watermarking detection)을 허용한다[1]. 참고문헌[3,6,7]은 워터마크 추출에 spread spectrum의 자기상관함수를 이용하므로, 자기상관함수를 적용하기 위하여 시간동기화가 정확히 선행되어야 하는 제약이 있다.

이 논문에서는 음질과 강인성을 고려한 워터마킹을 위하여 토널(tonal) 특성을 제안한다. 음을 구별하는 요소들 중 하나로 주기성을 강하게 보이는 신호들을 의미하는 토널은 여러 신호처리에 매우 안정적인 특성을 보인다. 오디오 상품의 불법적인 유통을 막기 위한 오디오 워터마킹 기술의 평가에서 음의 가치를 떨어뜨리는 공격은 평가사항이 아니다[8,9]. 토널은 음 인지에서 중요한 역할을 하므로, 토널의 변화를 일으키는 신호처리는 음질의 손상을 준다고 가정할 수 있다. 뒤집어 표현하자면 오디오 워터마킹의 평가사항인 음질을 어느 정도 유지하는 여러 처리들에 대해 토널의 변화는 매우 적다는 가정이 가능하다. 이 논문에서는 각 신호처리 후의 토널의 안정성을 실험으로 보이고 이를 이용한 워터마킹 기법을 제안한다.

이 논문의 구성은 다음과 같다. 2장에서 토널의 특성을 설명하고 여러 처리 후 토널의 안정성을 다른 워터마킹 특성과 비교하여 보인다. 3장에서는 이 논문에서 제안하는 토널 특성을 이용한 브라인드 오디오 워터마킹 기법을 제안하다. 4장에서는 제안한 기법의 성능평가결과를 논의하고 5장에서 결론을 맺는다.

## 2. 토널의 특성

토널 성분은 음을 구성하는 요소 중 뚜렷한 크기의 에너지를 가지는 주파수 영역의 요소를 말한다. 토널 성분들의 분포는 음의 인지를 결정짓는 성분들 중 하나이다. 심리음향모델에서 토널은 마스커(masker)로 작용하여 음의 인지 요인으로 작용한다[10].

이 연구에서는 MPEG-1 심리음향모델 1에 기술된 방법에 의해 토널 성분을 분석하였다[10]. 토널 성분 추출은 다음의 식(1)(2)(3)으로 추출된다. N은 한 프레임의 샘플 수이고, $w(n)$은 윈도우 함수이다. $P(k)$는 로그 크기의 파워 스펙트럼(log-scaled power spectrum)이다. 한 프레임내의 토널의 집합 $S_t$는 식(3)의 조건에 의해 결정된다. 즉 토널이란 한 바크(bark) 범위 내에 이웃 주파수들보다 7dB이상 큰 에너지를 가지는 성분을 의미한다. 식(3)에 의해 결정된 각 토널들은 음 인지에 영향을 주는 토널 마스커(tonal masker)들로 구성되고 토널 마스커의 크기는 식(4)와 같이 계산된다. 토널 마스커는 스프레딩 함수(spreading function)에 의해 주변 주파수 중 상대적으로 작은 에너지의 신호를 마스크(mask)한다.

식(4)는 식(5)로 풀이될 수 있고, 이를 해석하면 토널 마스커는 파워 스펙트럼 상에서의 토널 에너지와 인접한 두 주파수의 에너지들의 합을 로그크기로 계산한 것임을 알 수 있다. 즉 토널 마스커를 구성하기 위해 짧은 프레임내에서 큰 에너지를 가지는 토널과 이웃 주파수 신호의 진폭 제곱이 동일한 비중으로 누적됨을 알 수 있다. 따라서 토널에 대해 식(3)의

$$F(k)=\sum_{n=0}^{N-1} w(n)s(n)\, e^{-j\frac{2\pi kn}{N}},\ 0\le k\le N \qquad (1)$$

$$P(k)=10\ \log_{10}|F(k)|^2(dB),0\le k\le\frac{N}{2} \quad (2)$$

$$S_t=\left\{P(k)\left|\begin{array}{l}P(k)>P(k\pm1)\\P(k)>P(k\pm\triangle k)\pm7dB\end{array}\right.\right\}$$

$$\triangle k\in\begin{cases}2 & 2<k<63 \quad (0.17\sim5.5kHz)\\ [2..3] & 63\le k\le127 \quad (5.5\sim11kHz)\\ [2..6] & 127\le k\le256 \quad (11\sim20kHz)\end{cases} \quad (3)$$

$$P_{TM}(k)=10\ \log_{10}\sum_{j=-1}^{1}10^{0.1P(k+j)}(dB) \quad (4)$$

$$P_{TM}(k)=10\ \log_{10}\sum_{j=-1}^{1}|F(k+j)|^2(dB) \quad (5)$$

조건을 유지하면서 토널과 인접주파수의 에너지를 변화시킬 경우, 같은 토널 마스커를 생산할 수 있음을 알 수 있다.

그림 1(a)는 식(2)에 의해 한 오디오 프레임의 로그 크기의 파워 스펙트럼을 구한 결과이고 그림 1(b)는 식(3)(4)에 의해 구해진 토널 마스커를 덧붙인 결과이다.

이 연구에서 제안하는 워터마킹 기법은 토널 마스커를 구성하는 토널의 이웃 주파수 신호간의 진폭을 변조하여 워터마크 정보를 저장한다. 이러한 제안은 변동률 분석실험을 통하여 토널 마스커의 구성신호들이 여러 오디오 처리에 대해 거의 변하지 않는다는 사실에서 기인한다.

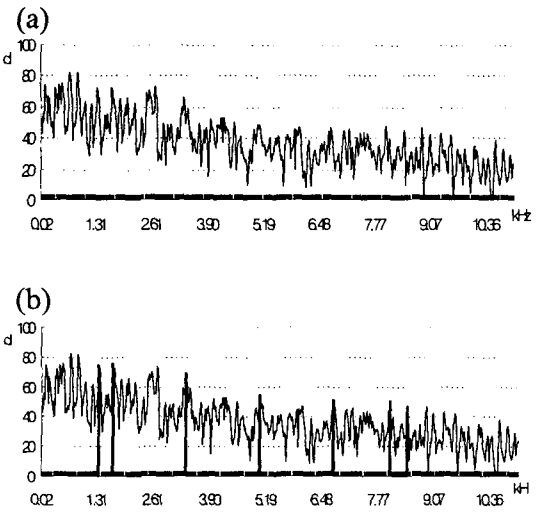



그림 1. (a). 한 오디오 신호 부분의 파워 스펙트럼(고주파영역 생략)
(b). 그림 1(a)에 심리음향모델의 토널 마스커를 덧붙임

이 절에서는 여러 처리 후에 토널과 이웃 주파수 신호의 에너지의 부동적인 특성을 분석하였다. 이 분석의 비교를 위해 기존의 워터마킹 기법[2,3]에서 사용된 위상 특성과 스프레드 스펙트럼을 조정하는 데 이용하는 각 주파수 신호들의 에너지 특성을 비교대상으로 선택하였다. 그림 2의 과정을 통해 원 오디오의 이 세 특성과 여러 신호 처리를 적용한 후의 각각의 특성들과의 변동률을 계산하여 분석하였다. 각 특성들은 정보 내장에 주로 사용되는 비교적 안정적인 저,중주파 대역(6.5kHz이하)의 것으로 분석하였다. 그리고 동등한 조건에서의 세 특성의 변동율 비교를 위해 위상과 진폭은 프레임 당 평균 tonal의 개수만큼 랜덤하게 선택하여 분석하였다. 토널 주파수의 에너지와 랜덤하게 선택된 주파수의 에너지는 식(6)에 의해, 변동율은 식(7)에 의해 계산하였다.

$$E_{freq}=RE^2_{freq}+Im^2_{freq} \quad (6)$$

$$\triangle c=\frac{x'-x}{x} \quad (7)$$

x는 원 오디오의 특성이고 x'는 각 신호처리 후 변화된 오디오의 특성이다. 적용된 신호처리는 다음과 같다.

a. 36dB의 가우시안 잡음 추가
b. 100Hz~6kHz 대역 통과 필터링
c. MPEG-1 layer III의 64kps 압축
d. 96kHz sampling rate 변환
e. 50% 피드백 100ms 지연의 Echo 변환
f. Equalization
10-band 이퀄라이저를 이용한 전형적인 방법
Hz : 31 62 125 250 500 1k 2k 4k 8k 16k
gain dB:-6 +6 -6 +6 -6 +6 -6 +6 -6 +6
g. Windows Media Audio 변환

이 분석실험에는 비발디의 <사계-봄-allegro> (44.1kHz, 16bit wav file)에서 각각 50,000개의 특성

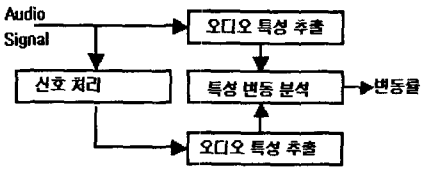



그림 2. 신호처리에 대한 오디오특성의 변동율 분석과정

값들을 추출하여 통계처리하였다. 표 1은 각 특성들의 변동율을 10% 절사평균 $\overline{\mu}_{.10}$과 표준 편차 $\overline{\sigma}_{.10}$으로 구한 결과이다. 하나의 관찰치가 평균에 영향을 줄 만큼의 이상치를 가질 경우, 평균은 전반적인 특성을 반영하기 어려우므로, 이러한 경우에는 중위수(median)에서 양끝의 관찰치 10%를 제거한 절사평균을 선택하는 것이 적절하다[11]. 이 실험에서는 랜덤하게 선택되어진 특성 x값이 매우 작을 경우, 실제 변동의 영향은 적더라도 식(7)에 의해 그 변동률은 매우 크게 나타나는 이상치를 보인다. 이 실험은 신호처리에 대한 각 특성값들이 전반적으로 얼마나 변동하는지를 비교하기 위함이므로 세 특성의 변동률에 대해 절사평균을 적용하였다. 그림 3은 각 신호

표 1. 각 오디오 특성들의 변동율 분석

| 신호처리 | $\overline{\mu}_{.10}$ | | | $\overline{\sigma}_{.10}$ | | |
|---|---|---|---|---|---|---|
| | tonal E | random E | random P | tonal E | random E | random P |
| a. | 0.0055 | 0.2370 | -0.1567 | 0.0009 | 0.0217 | 0.0047 |
| b. | -0.0001 | -0.0205 | -0.0078 | 0.0000 | 0.0002 | 0.0003 |
| c. | -0.0327 | 0.0283 | -0.2241 | 0.0016 | 0.0097 | 0.0069 |
| d. | 0.0000 | 0.0004 | -0.2112 | 0.0000 | 0.0003 | 0.0089 |
| e. | 0.0370 | 0.0722 | -0.1059 | 0.0023 | 0.0053 | 0.0046 |
| f. | 0.7261 | 0.6832 | -0.0084 | 0.0067 | 0.0069 | 0.0012 |
| g. | 0.0107 | 0.0768 | -0.1864 | 0.0012 | 0.0091 | 0.0058 |

tonal E:토널의 에너지, random E:랜덤하게 선택된 주파수신호의 에너지, random P:랜덤하게 선택된 주파수신호의 위상.

처리 적용 후에 특성들의 변동율을 히스토그램으로

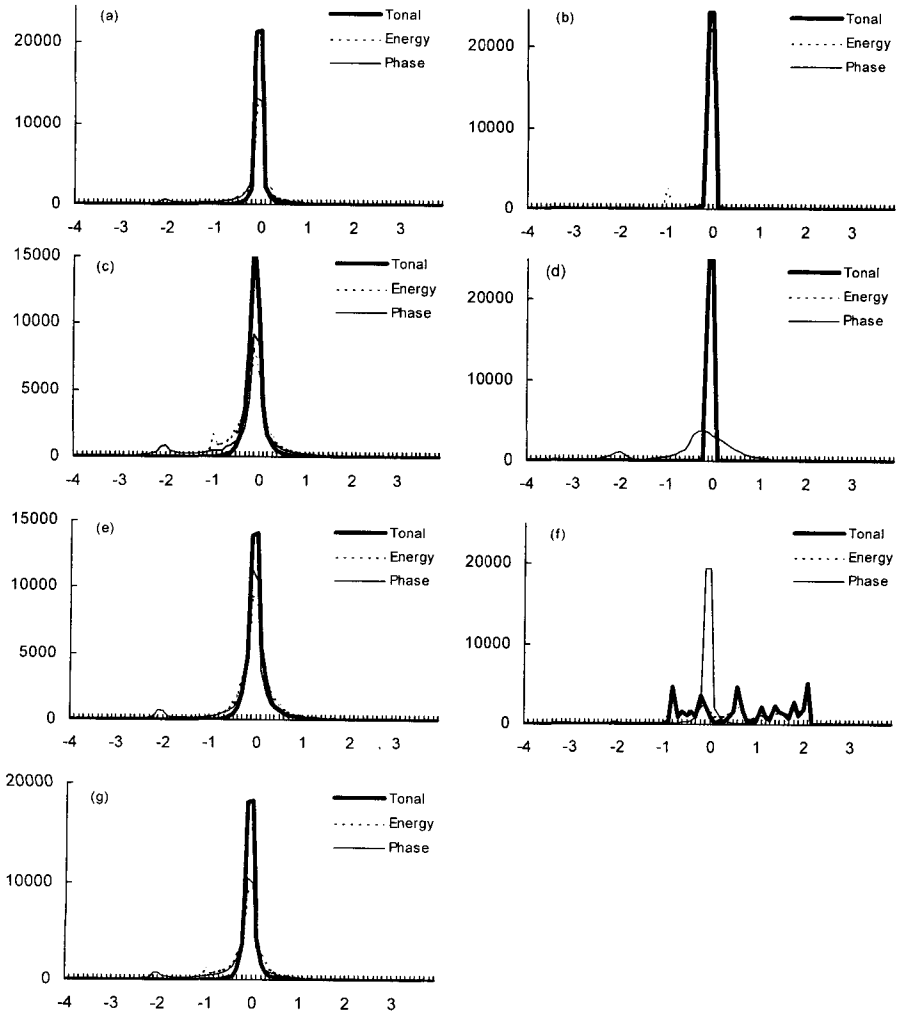



그림 3. 각 신호처리(a~g) 후의 특성들의 변동율 분포도 (세로축:발생빈도, 가로축:변동율 $\triangle C$)

보인다.

위의 결과에 따르면 위상 특성은 압축과, 샘플링 변화 그리고 echo처리에 상대적으로 많이 변화함을 알 수 있다. 랜덤하게 선택된 주파수 신호의 에너지 특성은 잡음과 equalization 처리, 에코처리에 비교적 많이 변화하였고 토널 에너지는 equalization 처리를 제외하고는 비교적 안정적인 결과를 보인다. 이 변동율 분석결과로 판단할 때 토널 에너지가 워터마킹에 사용되는 다른 두 특성에 비해 여러 신호처리에 매우 안정적임을 알 수 있다. 토널 마스커를 구성하는 토널의 이웃 주파수 에너지들도 또한 토널 에너지와 유사한 결과를 보였다. 이러한 결과를 바탕으로 우리는 토널 특성을 이용한 워터마킹을 제안한다.

## 3. 제안하는 워터마킹 방법

우리는 기존의 심리음향연구에서 정의된 토널 마스커 구성성분의 인지영향과 여러 신호 처리에 강인한 토널의 특성을 바탕으로 토널과 이웃 주파수들과의 관계를 조작하여 정보를 삽입하고 원음의 요구없이 이 정보를 추출하는 브라인드 워터마킹을 제안한다. 그림 4는 이 연구에서 워터마크 정보 비트 삽입에 이용한 토널과 이웃 주파수 에너지간의 관계이다. 이 관계는 식 (8)로 구체화 할 수 있으면 삽입할 워터마크 정보의 각 비트에 따라 토널의 이웃 주파수간의 차이를 조작한다. 식 (9)는 식(8)에서 적용한 limit 결정 수식이다. $E_t$는 토널 주파수의 에너지이고 $E_l$, $E_r$은 토널의 이웃 주파수의 에너지이다. $\alpha$는 음질과 워터마킹의 견고함 간의 trade off에 의해 결정되는 값이다.

이러한 관계 조작을 이용한 정보 내장이 유지될 수 있는지 확인하기 위해서는 토널의 안정성이외에 이웃 주파수와의 관계 또한 안정적이어야 한다. 즉 식(8)이 유지되기 위해서는 식(10)의 특성이 변동율이 적어야 한다. 따라서 식(10)과 같은 세 주파수신호의 관계가 여러 신호처리에 대해 통계적으로 변하지 않는지 분석하였다. 표 2는 식(10)의 관계 특성에 대한 변동율 분석 결과를 2절의 토널 에너지의 것과 비교하여 보인다. 결과는 토널 에너지의 변동률과 유사하였고, 특히 equalization 처리에서는 토널 에너지에 비해 매우 안정적이다. equalization 처리는 같은 주파수 영역에 같은 비율의 증가 또는 감소를 적용하므로 식(10)의 관계는 매우 적은 변동율을 보인다. 이러한 분석에 따라 제안한 방법이 여러 신호처리 후에도 워터마크 정보가 안정적으로 유지될 수 있음을 확인하였다.

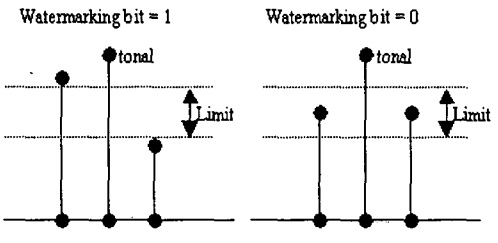



그림 4. 워터마크 삽입에 이용되는 주파수 에너지간의 관계

$$\begin{cases} R(E_l, E_r) > Limit, & \text{if } Bit = 1 \\ R(E_l, E_r) < Limit, & \text{if } Bit = 0 \end{cases} \quad \text{where } R(E_l, E_r) = |E_l - E_r| \tag{8}$$

$$Limit = \left(E_t - \left(\frac{E_l + E_r}{2}\right)\right)\alpha \tag{9}$$

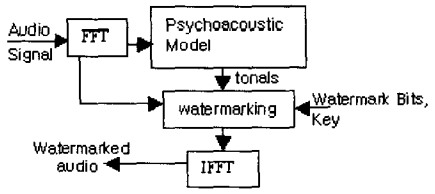
$$x = \left(E_t - \left(\frac{E_l + E_r}{2}\right)\right) / \ |E_l - E_r| \tag{10}$$

그림 5은 제안하는 워터마킹 삽입, 추출 방법의 개략도이다. 워터마크 삽입은 구체적으로 아래의 과정을 거친다.

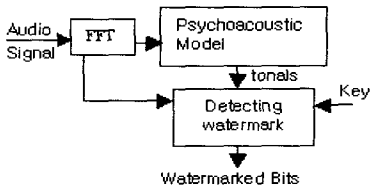
1. n개의 샘플들 단위로 프레임을 구성, 식(1)에 따라 FFT 변환을 한다.
2. 식(2)(3)에 따라 한 프레임내의 토널 집합을 구한다.
3. 각 프레임당 워터마크 삽입에 적용할 하나의 토널을 저, 중주파 대역에서 주어진 키를 이용하여 임의의 규칙에 따라 선택한다.
4. 한 프레임에 내장될 정보비트에 따라 위의 식

표 2. 토널 에너지 특성의 변동률과 토널과 이웃 주파수의 식 (10)의 특성값의 변동률 비교

| 신호 처리 | $\bar{\mu}_{.10}$ | | $\bar{\sigma}_{.10}$ | |
|---|---|---|---|---|
| | tonal E | 식(10)의 특성 | tonal E | 식(10)의 특성 |
| a. | 0.0055 | 0.0068 | 0.0009 | 0.0029 |
| b. | -0.0001 | -0.0003 | 0.0000 | 0.0001 |
| c. | -0.0327 | 0.0369 | 0.0016 | 0.0071 |
| d. | 0.0000 | 0.0000 | 0.0000 | 0.0002 |
| e. | 0.0370 | 0.0280 | 0.0023 | 0.0056 |
| f. | 0.7261 | 0.0018 | 0.0067 | 0.0017 |
| g. | 0.0107 | 0.0212 | 0.0012 | 0.0051 |




(a) 워터마크 삽입 과정




(b) 워터마크 추출 과정

그림 5. 제안하는 오디오 워터마킹 과정

(8)이 유지되도록 토널 이웃 주파수의 에너지를 조작한다. 이때 각각의 위상과 식(4)의 토널 마스커 에너지는 유지한다.

5. 이를 다시 IFFT 변환하여 오디오로 복구한다.

이 삽입과정에서 한 프레임의 선택된 토널 에너지가 매우 낮을 경우, 여러 신호처리에 의해 토널의 관계정보가 변경될 수 있다. 이를 보완하기 위하여 워터마크 정보를 M번 반복 내장하였다. 워터마크 추출시에는 삽입과정의 1,2,3의 처리를 통하여 워터마크된 토널을 찾고 식(8)의 조건을 역 적용하여 숨겨진 비트값은 판단한다. 이상의 방법으로 토널 성분을 이용한 브라인드 오디오 워터마킹 기법을 제안한다.

## 4. 실험 결과

다음은 위의 제안한 방법에 대해 실험을 통하여 추출율 측정과 음질 측정면에서의 워터마킹의 효용성을 보인다. 비발디의 <사계> 중 <봄>을 약 1분 단위의 크기로 8조각 나누어 실험하였다. 실험에 쓰인 소유권 정보는 72bit로 가정하였고 주파수 분석구간은 2048 sample로 64sample을 중첩하였다. 이 실험에서는 원음을 요구하는 워터마킹 방법에 비해 상대적으로 추출율이 떨어지는 브라인드 워터마킹의 추출율을 높이기 위해 72개의 워터마크 비트들을 10회 반복 내장하였다. 추출 시 반복 추출한 10개의 정보를 누적하여 내장된 워터마크 정보를 판단하였다. 이 실험에서는 72bit를 10회 반복 내장하는데 약 32초의 오디오 데이터가 필요하다.

표 3은 실험 오디오 조각 파일 중 한 파일에 대한 반복회수에 따른 워터마크 추출율을 보인다. 워터마크되지 않은 원 오디오와 워터마크된 오디오, 워터마크 후 각각의 처리를 적용한 오디오에 대해 실험하였다. 워터마크가 적용되지 않은 원 오디오에서는 추출 횟수를 거듭할수록 워터마크의 추출율이 0.5로 고정되었다. 워터마크가 내장된 오디오와 워터마크 후 각 처리를 적용한 오디오들은 10회의 반복추출에 100%의 정확한 워터마크 추출율을 보였다.

이러한 신호 처리들 외에 time shift 처리 후의 워터마킹 추출을 실험하였다. 위의 정해진 프레임간격마다 워터마크를 추출하는 방법으로는 time shift후의 워터마크 추출이 어렵다. 따라서 time shift에 대한 적합한 워터마크 추출방법이 필요하다. 즉 2차적인 추출 방법이 필요하다.

Time shift처리는 오디오를 선형적으로 늘이거나 줄이지 않는다. 상용 소프트웨어 cool edit를 이용하여 +4%의 time shift 처리한 후, 변경된 오디오를 살펴보면 약 0.5초 이내의 한 지점에서 약 800샘플이 복사되어 바로 다음 위치에 끼워넣어진 예를 볼 수

표 3. 한 오디오 조각에 대한 제안하는 워터마킹 기법의 반복 추출에 따른 워터마크 추출율

| 처리 내용 | 반복 추출 회수 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 원 오디오 | 0.51 | 0.35 | 0.54 | 0.31 | 0.51 | 0.38 | 0.49 | 0.44 | 0.50 | 0.50 |
| 워터마크된 오디오 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 36dB 가우시안 노이즈 삽입후 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 100Hz~6kHz 대역통과 필터링 후 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 64kbps MP3 변환후 | 0.88 | 0.83 | 0.96 | 0.97 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 |
| 96kHz sampling 변환 후 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 50% 피드백 100ms 지연 echo 변환후 | 0.86 | 0.79 | 0.96 | 0.93 | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | 1.00 |
| equalization 처리 후 | 0.93 | 0.86 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 48kbps WMA 변환 후 | 0.94 | 0.96 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

있다. 이러한 큰 변화에도 원음과 인지적으로는 거의 차이가 없었다. 이러한 신호처리는 워터마킹 공격의 매우 강력한 도구가 될 수 있다.

이 논문에서는 SDMI가 제안하는 ±4%의 time shift신호처리[8]에 대해 새로운 워터마크 추출방법을 제안하였다. 간단히 설명하면 time shift 공격과 같이 워터마크 추출도 시간축을 따라 시프트하면서 최적의 위치를 찾아가는 방법이다. 자세한 과정들은 다음과 같고 여기서 C는 5프레임 단위 추출의 시작 위치이고 T는 5프레임 크기의 8%를 의미한다.

1. 먼저 5프레임 단위 추출을 위한 시작위치 C를 정한다.
2. C 에서 T의 위치로 이동한다.
3. 이동한 위치에서 5프레임 크기의 샘플들을 읽어 각 프레임에서 식(8)의 관계값 $R(E_l, E_r)$를 계산한다.
4. 이를 내장된 것으로 예상되는5개의 워터마크 값과 비교한다. 이때 0과 1비트의 일치하는 개수를 판단하는 것이 아니고 과정 3의 계산값들이 식(8)의 조건식을 만족하기 위한 비율을 계산하여 5개의 비율을 합한다.
5. 다시 50 샘플만큼 이동하여 3,4의 과정을 처리한다.
6. 5의 과정을 C에서 T의 위치에 이동할 때까지 반복한다.
7. 계속 반복된 처리동안 4의 과정에서 합산된 비율값중 최대의 값을 찾고 이 값을 산출한 위치를 최적의 위치로 판단하고 5개의 워터마크 비트를 추출한다. 그리고 이 위치에서 5프레임 크기만큼 이동하여 다음 5프레임 단위 추출의 시작위치 C를 정한다.
8. 2에서 7의 과정을 반복한다.

이 실험과정에서 선택된 임의의 값들은 cool edit의 time shift 영향을 분석한 근거 하에 좋은 추출결과를 보이는 값들로 선택하였다. 위의 제안한 과정으로 추출한 결과, 약 90%의 워터마크 정보를 추출할 수 있었다. 이 방법은 동기화를 위한 부가적인 정보 없이 추출이 가능하나 워터마크 비트 추출을 위해 상당한 계산량이 요구되는 단점이 있다.

표 4는 8개의 오디오 조각파일에 대한 10회 반복 추출의 워터마크 추출 결과이다. 먼저 정해진 프레임 간격으로 추출하여 정적치 이상의 추출이 되지 않을 경우 time shift 공격으로 의심하여 2차적인 방법으로 위의 설명된 방법을 적용하였다. Echo와 time shift처리를 제외한 다른 신호처리에 대해 100%의 정확한 추출율을 보였다. 이러한 결과에 따라 토널 특성을 이용한 워터마킹의 강인성을 보였다.

표 4. 제안한 워터마킹 기법의 전체 추출율

| 처리 내용 | 추출율 |
|---|---|
| 워터마크된 오디오 | 1.000 |
| 36dB 가우시안 노이즈 삽입후 | 1.000 |
| 100Hz~6kHz 대역통과 필터링 후 | 1.000 |
| 64kbps MP3 변환후 | 1.000 |
| 96kHz sampling 변환 후 | 1.000 |
| 50% 피드백 100ms지연 echo변환후 | 0.986 |
| equalization 처리 후 | 1.000 |
| 48kbps WMA 변환 후 | 1.000 |
| +4% time shift 처리 후 | 0.901 |
| -4% time shift 처리 후 | 0.908 |

다음으로 제안하는 워터마킹 방법에 대한 음질평가를 위해 주관적인 청취 평가를 하였다. 20명의 일반 청취자를 대상으로 헤드폰을 이용한 환경에서 실험하였다. 실험자는 원음을 미리 알지 못하며, 두 음의 비교를 위해 여러 회 반복청취가 가능하다. 실험 오디오는 위의 실험에 적용한 오디오 파일을 10초단위로 끊어 12개의 원 오디오와 워터마크된 오디오 쌍간의 차이를 세밀히 분석 가능하도록 하였다. 실험결과 표 5는 12개의 오디오 파일에 대한 SDG(Subjective Diff-Grades) scale의 음질 평가 결과이다[12]. 주관적 척도 SDG 0.0은 두 오디오 간의 차이를 알아차릴 수 없음을 의미하고 척도 1.0은 알 수는 있으나 성가시지 않음을 의미한다. 전체 SDG평균이 0.27로 대부분이 알아차릴 수 없다고 평가하였다. 240개의 응답 중 176개의 응답이 0.0이었고 나머지 64개의 응답이 1.0이었다. 토널의 인지영향을 이용한 워터마킹 방법이므로 비지각성 면에서 좋은 결과를 보였다.

표 5. 제안한 워터마킹 기법의 SDG scale 음질 평가 결과

| 오디오 파일 | 파일1 | 파일2 | 파일3 | 파일4 | 파일5 | 파일6 |
|---|---|---|---|---|---|---|
| $\mu$(SDG) | -0.30 | -0.30 | -0.10 | -0.15 | -0.25 | -0.30 |
| $\sigma$ | 0.47 | 0.47 | 0.31 | 0.37 | 0.44 | 0.47 |
| 오디오 파일 | 파일7 | 파일8 | 파일9 | 파일10 | 파일11 | 파일12 |
| $\mu$(SDG) | -0.35 | -0.25 | -0.45 | -0.25 | -0.20 | -0.30 |
| $\sigma$ | 0.49 | 0.44 | 0.51 | 0.44 | 0.41 | 0.47 |

## 5. 결 론

지금까지 우리는 강인하고 원음에 가까운 오디오 워터마킹을 위한 토널 특성을 제안하였다. 그리고 이를 이용한 브라인드 오디오 워터마킹 방법을 제안하였다.

이 논문에서는 강인한 워터마킹을 위한 음향 특성으로서의 판단을 위해, 여러 신호처리 후의 토널과 이웃 주파수 에너지의 안정성에 대해 분석하였다. 이러한 토널의 특성를 바탕으로 토널 마스커를 이루는 주파수 신호들간의 에너지 관계를 이용하여 워터마킹 기법을 제안하였다. 이 방법에 대한 여러 신호처리 후의 워터마크 추출 실험결과, echo와 time shift 처리를 제외하고는 100%의 추출율을 보였다. time shift처리에서도 90%의 추출율로 워터마킹 방법으로서 유용함을 보였다. 음질 평가 또한 0.27 SDG의 결과로 오디오 워터마킹 응용에 적용할 만한 결과를 보였다. 이 후 Time shift 처리에 대한 좀더 개선된 워터마크 추출방법이 필요하다고 사료된다.

## 참 고 문 헌


[1] C.Neubauer, and J.Herre, "Advanced watermarking and its applications," 109th convention of AES, Sep, 2000.

[2] W.Bender, D.Gruhl, N.Morimoto, and A.Lu, "Techniques for data hiding," IBM Systems journal. Vol.35, 1996.

[3] M.D.Swanson, B.Zhu, A.H.Tewfik, and L.Boney, "Robust audio watermarking using perceptual masking," Signal Processing 66 pp.337-355, 1998.

[4] H.O.Oh, J.W.Seok, J.W.Hong, and D.H.Youn, "New echo embedding technique for robust and imperceptible audio watermarking," ICASSP 2001, vol 3, pp.1341-1344, 2001.

[5] J.D.Gordy and L.Bruton, "Performance evaluation of digital audio watermarking algorithms," IEEE MWSCAS 2000.

[6] D.Kirovski and H.Malvar, "Robust spread-spectrum audio watermarking," ICASSP 2001, vol. 3, pp.1345-1348, 2001.

[7] P.bassia, I.Pitas, and N.Nikolaidis, "Robust audio watermarking in the time domain," IEEE Transactions on multimedia, Vol.3, pp. 232-241, June 2001.

[8] SDMI, FRWG000224-01, "Call for proposals for phase II screening technology version 1.0," Feb. 2000.

[9] M.Wu, S.A.Craver, E.W.Felten, and B.Liu, "Analysis of attacks on SDMI audio watermarks," ICASSP 2001, May 2001.

[10] T.Painter, and A.Spanias, "Perceptual coding of digital audio," Portions published in Proc. IEEE, April, 2000.

[11] 이 순, "통계학," pp.515-549, 법문사, 2000.

[12] M.Arnold, "Audio watermarking : features, applications, and algorithms," ICME2000, vol. 2, NY, Jul. 2000.

**이 희 숙**

1990년 2월 창원대학교 전자계산학과 학사
1992년 2월 경북대학교 컴퓨터공학과 석사
2001년 8월 창원대학교 컴퓨터공학과 박사과정 수료

관심분야 : 오디오 신호처리, 디지털 컨텐츠 보호, 심리음향
E-mail: lhs@changwon.ac.kr

**이 우 선**

1981년 3월 고려대학교 전자공학과 학사
1983년 3월 고려대학교 대학원 전자공학과 석사
1988년 8월 고려대학교 대학원 전자공학과 박사
1987년 3월~현재 창원대학교 컴퓨터공학과 교수

관심분야 : 신호처리
E-mail: wslee@sarim.changwon.ac.kr

교 신 저 자

이 희 숙 641-773 경남 창원시 사림동 9번지 창원대학교 컴퓨터 공학과