

토널 마스크 특성을 이용한 오디오 워터마킹

An Audio Watermarking Method Using the Attribute of the Tonal Masker

이 희 숙*, 이 우 선*
(Hee-Suk Lee*, Woo-Sun Lee*)

*창원대학교 컴퓨터공학과

(접수일자: 2003년 3월 19일; 수정일자: 2003년 6월 10일; 채택일자: 2003년 6월 27일)

본 논문에서는 토널 마스크 특성을 이용한 오디오 워터마킹 기법을 제안한다. 먼저 오디오 워터마킹 특성으로서의 토널 마스크 특성을 분석한다. 기존의 심리음향 연구들에 의하여 토널 마스크를 구성하는 주파수 신호들에 대해 비지각적인 변조가 가능하다. 그리고 여러 신호처리에 대해 처리 전, 후의 톤과 좌, 우 주파수 신호의 에너지 관계를 비교하면 매우 적은 변화를 보인다. 이러한 토널 마스크 특성을 이용한 워터마킹 기법을 제안한다. 제안하는 기법은 톤의 좌, 우 주파수의 에너지 차이를 변조하여 하나의 워터마크 비트를 내장하고, 원음없이 내장시의 비밀 키를 이용하여 변조된 토널 마스크를 찾은 후 내장된 워터마크 비트를 판단하는 기법이다. 제안한 기법은 잡음 삽입과 압축, 대역 통과 필터링, 샘플링 변환, 반향 처리, 주파수 등화의 공격 후, 클래식과 팝에 대해 각각 평균 오류율 0.11%, 1.26%를 보였다. 그리고 주관적인 음질 평가를 통하여 평균 SDG (Subjective Diff-Grades) -0.31의 결과를 얻었다.

핵심용어: 오디오, 워터마킹, 토널, 마스크, 심리음향

주요분야: 음향 신호처리 분야 (1, 2)

In this paper, we propose an audio watermarking method using the attribute of tonal masker. First, the attribute of tonal masker as an audio watermarking attribute is analyzed. According to existing researches, it is possible to be imperceptible modulation for the energies of the frequencies that compose a tonal masker. And when the relation between the tone energy and the left or right frequency energy after various signal processing is compared with the one before the processing, very few changes are showed. We propose an audio watermarking method using these attributes of tonal masker. A watermark bit is embedded by the modulation of the difference between the two neighboring frequency energies of a tone. In the detection, the modulated the tonal masker is searched using the key used in the embedding without original audio and the embedded watermark bit is detected. After each attack of noise insertion, band-pass filtering, re-sampling, compression, echo transform, and equalization, the detection error ratios of the proposed method were average 0.11%, 1.26% for Classics and Pops. And the SDG (Subjective Diff-Grades) scale evaluation of the sound quality of the watermarked audio result in the average SDG -0.31.

Keywords: Audio, Watermarking, Tonal, Masker, Psychoacoustics

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

I. 서론

지금까지 여러 특성들을 이용한 오디오 워터마킹이 연구되어졌다. 초기의 오디오 워터마킹 연구는 LSB 삽입 기법, 패치워크 (patchwork) 기법 등이 제안되었으나 이 방법들은 여러 신호처리 후 워터마크의 손실이 매우 크

다는 단점을 보였다. 이후 여러 처리와 공격에 강인하기 위하여 신호의 변화에 비해 상대적으로 덜 지각되는 특성을 이용하여 더욱 큰 크기의 워터마크를 내장하는 기법이 제안되었다. 반향 내장 (echo hiding)[1,2], 위상 코딩 (phase coding)[1], 심리음향모델 (psychoacoustic model) 을 이용한 확산 대역 (spread spectrum) 신호 삽입 기법 [3-6] 등이 그 대표적인 예이다. 반향 내장 기법은 오디오의 짧은 구간마다 인지 한계 이하의 시간 지연된 반향을 삽입함으로써 워터마크를 내장하는 기법으로 높은 수

준의 비지각성을 보였고 압축, 잡음 삽입, 대역 통과 필터링 등의 여러 처리에 대해 좋은 강인성을 보였다[1,2]. 그러나 반향의 추출에 의한 워터마크의 노출과 반향 제거가 가능하다는 단점이 있다. 위상 코딩은 이웃 구간과의 급격한 차이만 없다면 짧은 구간 내의 위상의 변화를 인지하지 못한다는 청각특성을 이용한 기법이다. 이 기법은 음질 면에서 매우 효과적인 워터마킹 기법이다. 그러나 MPEG 코딩과 같은 신호처리에 위상정보가 쉽게 변하는 단점을 가진다[7]. 심리음향모델을 이용한 확산대역신호 삽입기법은 워터마크를 잡음과 유사한 형태로 변환한 후 각 주파수별로 마스킹되는 에너지 이하의 크기로 조정하여 내장하는 기법이다[3-6]. 이 기법은 워터마크 추출에 자기상관함수를 이용하므로 매우 정확한 동기화가 선행되어야 하는 단점이 있다. 그리고 원 오디오없이 워터마크를 추출하는 기법들에서 저역통과 필터링, MP3압축, 재샘플링 등의 신호처리에 대해 워터마크의 손실이 크다 [5,6].

본 연구에서는 워터마크된 오디오의 음질왜곡을 줄이고 더욱 강인한 워터마크의 내장을 위해 토널 마스커 특성을 이용한 오디오 워터마킹 기법을 제안한다.

토널 성분 (tonal component)이란 일반적으로 순음이 라 정의할 수 있으나 본 논문에서는 MPEG 심리음향모델 [8,9]에서 서술한 좁은 의미의 것으로 정의하고 이를 톤 (tone)이라 부른다. 이 모델에서 톤은 11.6 ms~23.2 ms 시간 구간의 주파수 신호 에너지들 중에서 각 바크 (bark) 를 대표하는 2개 이하의 국부적 최대치 (local maximum) 를 말한다. 톤은 좌,우의 인접 주파수 신호와 함께 토널 마스커로 작용하여 주변의 상대적으로 작은 에너지 신호들을 마스킹한다. MPEG 코딩에서 토널 마스커들은 마스킹 임계치 (masking threshold)를 결정짓는 중요한 요소이며 스펙트럼의 피크들 중 일부를 구성한다. 이러한 사실들을 토널 마스커가 음의 인지적 측면에서 중요한 역할을 함을 의미한다. 인지적 측면에서 중요하지 않은 요소들에 내장된 워터마크는 쉽게 제거될 수 있으므로 강인한 워터마킹을 위해서는 인지적으로 중요한 요소에 내장하여야 한다[10]. 따라서 토널 마스커 특성은 이러한 조건에 적합하다. 그리고 토널 마스커는 기존의 심리음향 연구와 MPEG 심리음향모델 I의 토널 마스커 구성식을 통하여 비지각적 변조가 가능하다.

본 논문에서는 기존의 연구를 통해 토널 마스커 구성 신호들의 비지각적 변조가 가능한지 분석한다. 그리고 이러한 변조가 여러가지 신호처리를 거친 후에도 유지될 수 있는지를 분석한다. 이러한 분석들을 통해 토널 마스

커 특성이 비지각적이고 강인한 워터마크 내장 특성으로 적합한지를 확인한다.

그리고 이러한 토널 마스커 특성을 이용한 블라인드 (blind) 오디오 워터마킹 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 오디오 워터마킹 특성으로서의 토널 마스커 특성을 분석한다. 3장에서는 토널 마스커 특성을 이용한 오디오 워터마킹 기법을 제안한다. 4장에서는 제안한 워터마킹 기법의 성능을 평가하고 5장에서 결론을 맺는다.

II. 정보내장을 위한 토널 마스커의 특성

이 장에서는 토널 마스커 특성이 오디오 워터마킹 특성으로서 비지각성과 강인성을 가지는지 분석한다.

먼저 본 연구에 적용된 MPEG 심리음향모델 [8,9]에서의 토널 마스커를 살펴본다. 식 (1)~(4)는 토널 마스커의 음압레벨 (SPL: Sound Pressure Level)을 구하는 계산식이다[8,9].

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

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

$$S_t = \left\{ P(k) \left| \begin{array}{l} P(k) > P(k \pm 1), \\ P(k) > P(k \pm \Delta k) \pm 7 \text{ dB} \end{array} \right. \right\}$$

$$\Delta k \in \left\{ \begin{array}{ll} 2 & 2 < k < 63 \quad (0.086 \sim 0.271 \text{ kHz}) \\ [2..3] & 63 \leq k < 127 \quad (0.271 \sim 5.469 \text{ kHz}) \\ [2..6] & 127 \leq k < 255 \quad (5.469 \sim 10.981 \text{ kHz}) \\ [2..12] & 255 \leq k \leq 500 \quad (10.981 \sim 21.533 \text{ kHz}) \end{array} \right. \tag{3}$$

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

N 은 한 프레임의 샘플 수이고, $w(n)$ 은 윈도우 함수, $P(k)$ 는 k 번째 주파수 신호의 음압 레벨이다. 한 프레임 내의 톤의 집합 S_t 는 식 (3)의 조건에 의해 결정된다. 식 (3)에서 톤이란 1~0.5 바크 (bark) 대역 내에서 이웃 주파수들보다 7 dB 이상 큰 에너지를 가지는 주파수 신호를 의미한다. 식 (3)에 의해 결정된 톤들에 대해 식 (4)는 토널 마스커의 음압레벨을 계산한다. 토널 마스커의 음압레벨은 마스킹 임계치를 구하는데 적용되어, 주변 주파수 중 상대적으로 작은 에너지의 신호를 마스킹 (mask)하는지 판단할 수 있다. 그림 1은 식 (2)에 의해 구해진 한

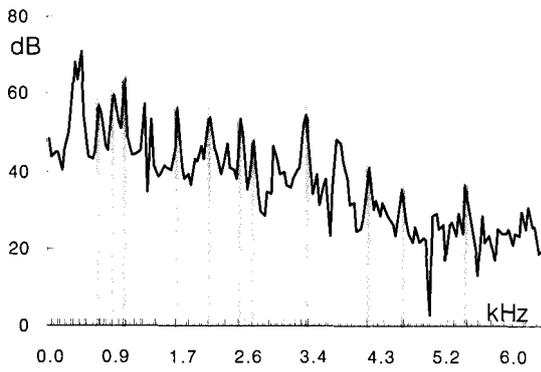


그림 1. 파워 스펙트럼과 토널 마스크 (고주파대역 생략)
 Fig. 1. Power spectrum and tonal maskers (omission of high frequency bands).

오디오 프레임의 파워 스펙트럼과 식 (3), (4)에 의해 구해진 토널 마스크를 보인다.

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

토널 마스크 음압레벨을 계산하는 식 (4)는 식 (5)로 풀이될 수 있고, 이를 해석하면 토널 마스크는 파워 스펙트럼 (power spectrum) 상에서의 톤 에너지¹⁾와 안접한 두 주파수 신호의 에너지들의 합을 로그 크기로 계산한 것임을 알 수 있다. 즉 토널 마스크의 음압레벨은 한 오디오 프레임 내에서 톤과 좌,우 주파수 신호의 진폭 제곱이 동일한 비중으로 누적됨을 알 수 있다. 따라서 톤에 대해 식 (3)의 조건을 유지하면서 톤과 좌,우 주파수 신호의 에너지를 변화시키더라도 같은 토널 마스크를 생산할 수 있다.

아래의 변조를 가정해 보자.

[변조 예] 1024 크기의 프레임에서 1 kHz에 존재하는 톤에 대해, 톤의 왼쪽 주파수 신호의 에너지를 만큼 줄이고 오른쪽 신호의 에너지를 만큼 증가시킨다. 이 변조는 식 (4)에 의해 같은 토널 마스크의 음압 레벨을 유지할 수 있다. 이 때 톤의 좌,우 주파수 신호의 간격은 0.086 kHz로 1 kHz 주파수 영역에서 약 0.5 바크의 대역폭이다.

이러한 변조가 인간의 청각시스템 (HAS)에서 같은 음으로 인지되는지 기존의 연구를 살펴본다. 기존의 심리음향 연구결과 중의 하나인 그림 2와 그림 3[11]을 통해 이 변조는 청각시스템에 의해 같은 음 크기 (loudness)로 인지됨을 알 수 있다. 그림 2는 1 kHz를 중심으로 대역통과하는 균등 잡음 (uniform noise)에 대해 대역폭과 음압레

벨을 바꾸면서 실험하여, 같은 음 크기로 인지되는 대역폭과 음압레벨의 관계를 보인 것이다. [11]에서는 어느 정도의 좁은 대역을 통과하는 균등 잡음에 대해 통과한 주파수 신호의 음압레벨의 합이 같으면 같은 음 크기로 인지된다고 설명한다. 그림 3은 1 kHz 중심-대역통과 균등잡음에 대해 대역을 통과하는 신호들의 음압 레벨의 합을 20, 40, 60, 80의 상수로 각각 맞추고, 대역폭을 넓히면서 몇 dB의 1 kHz 톤 (정현파 신호)과 같은 음 크기로 인지되는지 실험한 결과이다. 4개의 선에 표시된 음압레벨은 대역통과 균등잡음의 음압레벨의 합이다. 이 그림에서 0.16 kHz 대역폭 이하의 경우, 같은 음압레벨의 1 kHz 톤과 같은 음 크기로 인지됨을 알 수 있다. 여기서 0.16 kHz는 1 kHz 주파수 영역의 1 바크 대역폭에 해당하며 인간의 청각이 바크 단위로 음 크기의 인지 기능을 달리함을 알 수 있다.

그림 2와 그림 3[11]은 대역통과 균등잡음과 하나의 톤

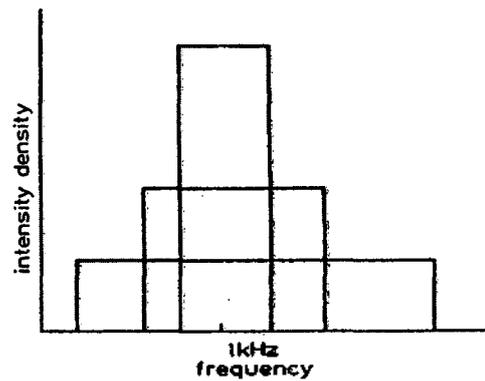


그림 2. 같은 인지적 음 크기를 출력하는 대역통과 잡음의 여러 대역폭에 따른 음압레벨
 Fig. 2. Sound pressure level of band-pass noises for different bandwidths producing the same perceptible loudness.

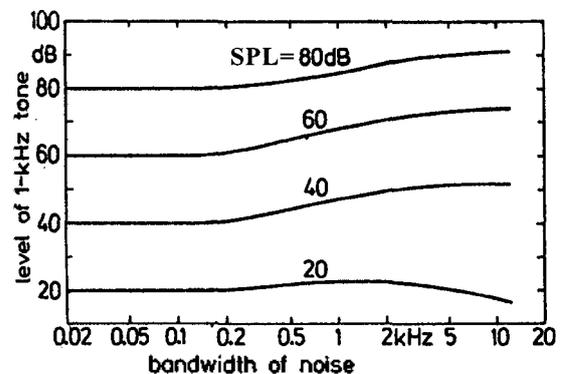


그림 3. 1 kHz 중심의 대역통과 잡음의 대역폭과 같은 인지적 음 크기를 출력하는 1 kHz 톤의 음압레벨
 Fig. 3. Sound pressure level of 1 kHz tone judged as loud as a band-pass noise at a center frequency of 1 kHz (Parameter is the sum of SPL of the band-pass noise).

1) 본 논문에서 에너지는 진폭의 제곱을 의미한다.

에 대한 음 크기 인지 관계를 보이지만 이 결과들을 고려해 볼 때 위의 <변조 예>에서 가정한 톤의 좌,우 주파수 신호의 에너지 변조는 0.086 kHz 대역 내에서 음압레벨의 합이 그대로 유지되는 변조이므로 같은 음 크기로 인지됨을 확인할 수 있다.

그리고 바크 대역은 청각시스템의 주파수 분해능력을 반영하여 24개의 영역으로 나눈 것이다. 위의 <변조 예>에서 중심에 큰 에너지의 톤을 두고 0.5 바크 대역 내의 좌,우 주파수 신호의 에너지 증감 정도는 인지적으로 음높이 (pitch)의 변화를 구별하기가 어렵다. 이러한 이유에서 <변조 예>와 같은 약 0.5 바크 대역 내의 변조는 비지각적인 워터마크 내장으로 이용할 수 있다.

다음은 위의 <변조 예>가 여러 신호처리 후에도 어느 정도의 변화를 가지는지 살펴본다. 만일 변화가 매우 적다면 워터마크 내장을 위해 변조하였을 경우 강인한 워터마크 내장이 가능함을 의미한다. 본 논문에서는 토널 마스크 구성 신호들의 변동률을 분석하기 위해 다음의 특성들을 실험하였다.

- (a) 톤의 에너지
- (b) 톤의 좌,우 주파수의 에너지
- (c) 톤의 좌,우 주파수에너지 / 톤의 에너지

그리고 이 실험에 적용된 신호 처리들[12]은 다음과 같다.

1. 36 dB의 가우시안 잡음 추가
2. 100 Hz ~ 6 kHz 대역 통과 필터링
3. MPEG-1 layer III의 64 kps 압축
4. 96 kHz 샘플링 변환 (re-sampling)
5. 50% 피드백 100 ms 지연의 반향 처리
6. 주파수 등화 (Equalization)

Hz: 31 62 125 250 500 1k 2k 4k 8k 16k
 gain dB: -6 +6 -6 +6 -6 +6 -6 +6 -6 +6

변동률 분석실험은 신호처리 전과 후의 (a), (b), (c) 특성들을 추출하여 식 (6)에 따라 계산하였다. x는 신호처리 전의 특성이고 x'는 신호처리 후의 변화된 특성이다. 이 분석실험은 1024 샘플 크기 프레임으로 주파수 분석하였고 각 특성을 50,000개씩 추출하여 변동률을 계산하였다. 모든 특성은 워터마킹에 주로 사용되는 비교적 안정적인 저,중주파 영역 (6.6 kHz 이하)의 것으로 추출하였다. 신호처리에 대한 각 특성들의 전반적인 변동률을 판단하기 위해 중간값 (median)으로부터 양끝의 5%를 각각 제거한 10% 절사평균 $\bar{\mu}_{.10}$ 과 편차 $\bar{\sigma}_{.10}$ 를 계산하여 표 1에서 보인다[13].

표 1. 신호처리에 대한 토널 마스크 특성의 변동률
 Table 1. Change ratio of tonal masker attributes on various signal processing.

Signal processing	$\bar{\mu}_{.10}$			$\bar{\sigma}_{.10}$		
	(a)	(b)	(c)	(a)	(b)	(c)
1	0.0008	0.0022	0.0012	0.0003	0.0007	0.0006
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	-0.0369	-0.0234	0.0135	0.0012	0.0026	0.0022
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0270	0.0313	0.0045	0.0022	0.0027	0.0015
6	0.1814	0.1874	0.0048	0.0088	0.0090	0.0007

$$\Delta c = \frac{x' - x}{x} \tag{6}$$

이 실험결과에서 (a), (b)는 주파수 등화를 제외하고 최대 -0.037 ~ + 0.0313의 변동률을 보인다. 주파수 등화는 인접한 주파수 영역의 신호 에너지들을 같은 비율로 증가시키거나 감소시키는 처리이다. 따라서 톤과 좌,우 주파수 에너지 관계 (c)는 주파수 등화를 포함한 여러 신호처리에 0.0135 이하의 적은 변동률을 보인다. 위의 결과로 볼 때 톤과 좌,우 주파수 에너지 관계를 이용하여 워터마크를 내장한다면 내장된 워터마크 또한 이러한 정도의 변화를 가질 것이다.

그러므로 위의 <변조 예>에서 변조 에너지 크기 Δe 를 이 변동률을 보상할 만큼 크게 적용한다면 이 변조를 이용하여 내장된 정보를 유지할 수 있다. 따라서 토널 마스크 구성 신호들의 관계 특성은 강인한 워터마킹 내장 특성으로 이용할 수 있다.

이상으로 워터마킹 특성으로서의 토널 마스크 특성의 비지각성과 강인성을 검토하였다.

다음은 토널 마스크 특성에 대하여 정보내장을 위한 용량적인 측면을 검토하였다. 여러 장르의 오디오를 대상으로 톤의 분포를 분석하였다. 표 2는 실험에 이용된 오디오 데이터로, No.1~5는 고음과 저음의 악기들이 다양하게 연주되는 클래식 음악들이고 No.6, 7은 남성 보컬의 락 (rock), No.8은 힙합 (hip hop), No.9, 10은 재즈 (jazz) 장르의 음악들이다. 표 3은 각 오디오의 모든 프레임에 대해 톤의 수를 추출하여 프레임당 톤의 수를 평균한 결과이다. 그리고 그림 4는 실험 오디오 No.2에서 선택한 연속된 50개의 오디오 프레임에서의 톤의 대역별 분포를 보인다. 실험오디오에서 톤은 전체 바크 영역에 고루 분포함을 알 수 있다. 이상의 결과들에 의해 토널 마스크 특성은 프레임당 평균 10개의 분포를 가지며, 정보내장을 할 경우 묵음 또는 톤이 없는 잡음 영역을 제외하고는

표 2. 실험에 이용된 오디오 데이터
Table 2. Audio data using experiments.

No	Artist Title
1	Beethoven-Symphony No. 5 c-Moll/c-minor op.67 Allegro con brio
2	Vivaldi-Spring E major Allegro
3	Vivaldi-Spring E major Danza pastorale Allegro
4	Schubert-Symphony No 8 In B Minor Andante Con Moto
5	Mozart-Horn Concerto No 2 E flat major (K 418) Andante
6	Sting-When We Dance
7	Sting-English Man In New York
8	Eminem-The Real Slim Shady
9	Stan Getz-Sometimes I'm happy
10	Sarah Vaughan-I cried for you

표 3. 프레임 당 톤의 평균 수
Table 3. average number of tone per frame.

Audio No	1	2	3	4	5	6	7	8	9	10
Avg. number	9.44	12.06	13.56	9.93	11.00	7.39	8.77	9.52	7.2	11.78

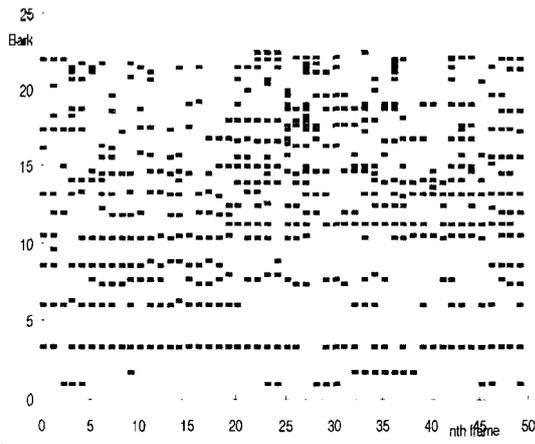


그림 4. 연속 프레임에서의 토널 마스크의 분포
Fig. 4. Distribution of tonal maskers at the successive frames.

정보를 내장할 만한 용량을 가짐을 알 수 있다.

III. 제안하는 오디오 워터마킹

이 장에서는 톤과 톤의 좌,우 주파수들간의 관계를 변조하여 정보를 삽입하고 원음의 요구없이 워터마크를 추

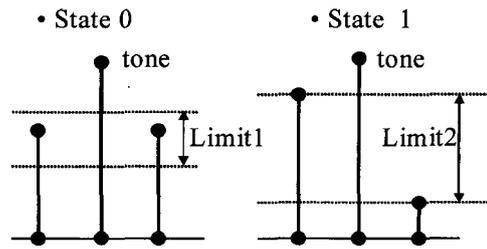


그림 5. 워터마크 내장에 적용되는 이진 상태
Fig. 5. Binary state to be applied in the watermark embedding.

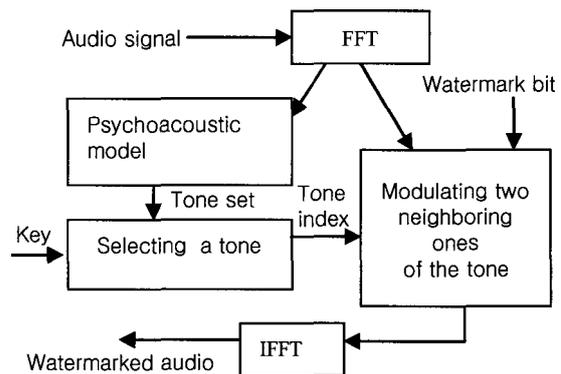


그림 6. 한 프레임의 워터마크 내장 과정
Fig. 6. Watermark embedding at a frame.

출하는 워터마킹 기법을 제안한다. 그림 5는 이진 워터마크 비트 내장에 이용할 톤과 좌,우 주파수 에너지간의 관계이고 이는 식 (7)로 정의된다. E_l 는 톤 주파수의 에너지이고 E_r , E_c 은 톤의 좌,우 인접주파수의 에너지들이다. α , β 는 음질과 워터마킹의 강인함 간의 trade off에 의해 결정되는 값이다. 그림 4의 Limit1, Limit2는 식 (7)에서 $L_1=Limit1/2$, $L_2=Limit2/2$ 로 각각 표기한다.

$$\begin{cases} R(E_l, E_r) \leq L_1, & \text{if state 0} \\ R(E_l, E_r) \geq L_2, & \text{if state 1} \end{cases}$$

where $R(E_l, E_r) = |E_l - E_r|/2$,
 $L_1 = (E_l - E_{avg})\alpha$, $L_2 = (E_l - E_{avg})\beta$,
 $E_{avg} = (E_l + E_r)/2$ (7)

그림 6은 한 프레임에 하나의 워터마크 비트를 내장하는 과정의 개략도이고 자세한 과정은 다음과 같다.

1. n샘플 단위의 프레임 구성, 식 (1)에 따라 FFT 변환을 한다.
2. 식 (2), (3)에 따라 한 프레임 내의 톤 집합을 구한다.
3. 키를 이용하여 저, 중 주파수 영역에서 무작위로 주파수를 선택한다. 선택된 주파수로부터 키에 의해 선택된 방향으로 톤을 검색하여 처음 만나는 톤을 선택한다.

4. 선택된 톤의 좌,우 주파수 에너지를 식 (8)에 의해 변조한다. WM은 한 프레임에 삽입할 워터마크 비트이다. 식 (8)은 토널 마스크의 에너지를 유지하면서 톤의 좌,우 주파수 에너지를 최소의 변화로 변조하기 위하여 좌,우의 크기 관계에 따라 변조한다. 이때 위상은 원래 값을 유지하여야 한다.
5. 이를 다시 IFFT 변환하여 오디오로 복구한다.

$$\left\{ \begin{array}{l} \left(\begin{array}{l} E_l = E_{avg} + L_1 \\ E_r = E_{avg} - L_1 \end{array} \right), \text{ if } E_l \geq E_r \text{ and } R(E_l, E_r) > L_1 \\ \left(\begin{array}{l} E_l = E_{avg} - L_1 \\ E_r = E_{avg} + L_1 \end{array} \right), \text{ if } E_l < E_r \text{ and } R(E_l, E_r) > L_1 \\ \hspace{10em} \text{if } WM = 0 \\ \left(\begin{array}{l} E_l = E_{avg} + L_2 \\ E_r = E_{avg} - L_2 \end{array} \right), \text{ if } E_l \geq E_r \text{ and } R(E_l, E_r) < L_2 \\ \left(\begin{array}{l} E_l = E_{avg} - L_2 \\ E_r = E_{avg} + L_2 \end{array} \right), \text{ if } E_l < E_r \text{ and } R(E_l, E_r) < L_2 \\ \hspace{10em} \text{if } WM = 1 \end{array} \right. \quad (8)$$

목음과 잡음 영역의 경우 워터마크가 내장되지 않는다. 이러한 문제점을 보완하기 위하여 워터마크 비트들을 M번 반복하여 내장한다.

그림 7은 한 프레임에서 내장된 하나의 워터마크 비트를 추출하는 개략도이다. 내장시의 키를 이용하여 다시 같은 톤을 선택할 수 있고 선택된 톤의 좌,우 주파수 에너지들의 관계를 판단하여 0과 1상태의 근접함을 -1.0과 1.0 사이의 값으로 추출한다. 이 값들을 반복하여 추출한 후, 이를 통계 처리하여 내장된 워터마크 시퀀스를 알아낸다.

위의 그림 6, 7은 제안하는 워터마크 내장과정과 추출 과정을 단순하게 설명하고자 하였다. 제안한 기법은 심

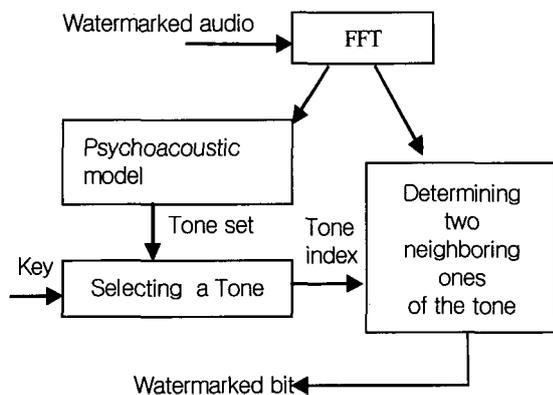


그림 7. 한 프레임의 워터마크 추출 과정
Fig. 7. Watermark detection at a frame.

리음향모델의 계산과정 중 톤의 집합을 구하는 계산이 요구된다. 이 과정의 계산량을 줄이기 위해서는 톤의 집합을 미리 구하지 않고 키에 의해 주파수와 검색방향을 결정한 후, 해당 영역에 대해서 하나의 톤을 구하는 절차로 바꿀 수 있다.

IV. 제안한 워터마킹 방법에 대한 실험 평가

이 절에서는 토널 마스크 특성을 이용한 워터마킹 방법의 성능을 평가한다. 성능 평가를 위한 기본적인 실험 환경은 다음과 같다. 표 2의 오디오를 실험 데이터로 적용하였고 오디오 프레임은 2048 샘플로, 64 샘플을 중첩하여 구성하였다. 2048 샘플의 프레임으로 FFT를 할 경우, 톤의 좌,우 주파수간의 간격은 0.043 kHz로 음질의 변화를 더욱 줄일 수 있다. 식 (7)의 α, β 는 각각 0.4, 0.05로 적용하였고 워터마크 내장을 위해 톤을 선택하는 주파수 영역은 0.86 khz ~ 5.2 kHz 사이로 제한하였다.

이러한 실험환경에 의해 워터마크된 오디오의 SNR (Signal-to-Noise Ratio)은 표 4에서 보인다. 그러나 SNR에 의해 음질을 판단할 수 없으므로, 주관적 음질 평가 실험을 통하여 제안한 기법의 비지각성 평가를 하였다. ITU-R의 "the double-blind triple stimulus with hidden reference" 방법[14]으로 45명의 일반 청취자를 대상으로 실험하였다. 클래식 경우 평균 SDG (Subjective Diff-Grades) -0.285, 팝의 경우 평균 SDG -0.340의 평가를 받았다. SDG 0.0~-1.0의 값은 변화된 음을 지각할 수는 있으나 거슬리지 않는다는 의미로 제안한 기법의 워터마크된 오디오는 워터마킹 응용분야에서 실용할만한 음질로 평가할 수 있다.

제안한 기법의 강인성 평가는 위의 실험환경에서 워터마크 시퀀스를 72 bit로 가정하고 30초 구간마다 추출한 워터마크의 오류율을 분석하였다. 먼저 30초 구간에서 반복 추출한 워터마크 비트들을 통계 처리하여 72비트의 워터마크 시퀀스를 결정하고, 추출 오류율을 구한다. 이를 전 오디오 구간에서 반복하여 각 실험 데이터마다 평균 오류율을 구하였다. 표 5는 워터마크된 오디오와 워터

표 4. 제안한 워터마킹 기법의 SNR
Table 4. SNR of the proposed method.

Audio No	1	2	3	4	5	6	7	8	9	10
SNR (dB)	31.7	26.4	26.5	31.1	28.9	32.8	30.3	33.5	36.5	27.0

표 5. 제안한 워터마킹 기법의 평균 오류율
Table 5. Average error ratio of the proposed watermark detection.

Signal processing	N0.1	N0.2	N0.3	N0.4	N0.5	N0.6	N0.7	N0.8	N0.9	N0.10	Classics	Pops
None	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Noise adding	0.0000	0.0000	0.0000	0.0046	0.0000	0.0013	0.0000	0.0015	0.0174	0.0046	0.0009	0.0050
BP filtering	0.0000	0.0000	0.0000	0.0000	0.0000	0.0038	0.0017	0.0062	0.0000	0.0000	0.0000	0.0023
MP3 compression	0.0040	0.0000	0.0000	0.0020	0.0023	0.0543	0.0608	0.0556	0.0521	0.0602	0.0017	0.0566
Re-sampling	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0046	0.0000	0.0009
Echo transform	0.0050	0.0020	0.0017	0.0026	0.0046	0.0051	0.0000	0.0108	0.0208	0.0093	0.0032	0.0092
Equalization	0.0030	0.0000	0.0000	0.0060	0.0023	0.0126	0.0035	0.0077	0.0347	0.0138	0.0023	0.0145
Average	0.0017	0.0003	0.0002	0.0022	0.0013	0.0110	0.0094	0.0117	0.0179	0.0132	0.0011	0.0126

다킹 후 각 신호처리를 거친 오디오에 대한 평균 오류율을 보인다. 이때 적용한 신호 처리들은 2장에서 적용한 처리들과 같다.

표 5의 결과로 제안한 기법은 클래식(No.1~5)에 대한 적용이 팝(No.6~10)에 대한 적용에 비해 강인성 면에서 효과적임을 알 수 있다. 특히 팝은 압축 처리에 비교적 높은 오류율을 보인다. 일반적으로 클래식은 한 프레임의 스펙트럼 상에서 톤과 다른 주파수 신호들의 구별이 드드러지게 유지되는 반면에 팝은 톤이 아닌 주파수 신호들이 비교적 높은 음압을 가지고 저, 중주파 영역에 분포함으로 톤과의 구별이 모호한 예들이 많다.

그리고 위의 평가 실험 이외에 제안한 기법의 성능을 분석하였다.

워터마크 추출과정의 전처리 문제인 프레임의 동기화 [5]에 대해 제안한 기법의 워터마크 추출율을 실험하였다. 제안한 기법은 100 샘플의 프레임 동기 오차를 가질 경우 평균 91% 이상의 워터마크 시퀀스를 추출할 수 있다.

그러나 시간늘금변형 (time scaling: pitch-invariant time scaling +/- 4%) 처리[12]에 대해서는 제안한 기법의 워터마크 추출이 불가능하다. 이 처리는 오디오의 작은 비율의 증감이 필요할 때 가장 많이 이용되는 처리로 피치불변 (pitch-invariant)을 유지하기 위해 작은 크기의 샘플들을 불규칙적으로 복사, 또는 삭제한다. 그림 8은 시간늘금변형의 처리에 의한 오디오의 비동기를 분석한 결과이다. 상용 오디오 편집 소프트웨어로 +4% 피치불변 시간늘금변형 처리를 하였을 때, 1초 구간 (44,100 샘플) 의 2, 3개의 위치에서 바로 앞의 샘플들의 복사가 발생하여 급격한 증가를 하는 것을 볼 수 있다. 이러한 처리에도 오디오의 음질은 원래의 것과 구별이 어렵다. 따라서 이러한 처리들에 대한 프레임 동기화 알고리즘 개발이 필요하다.

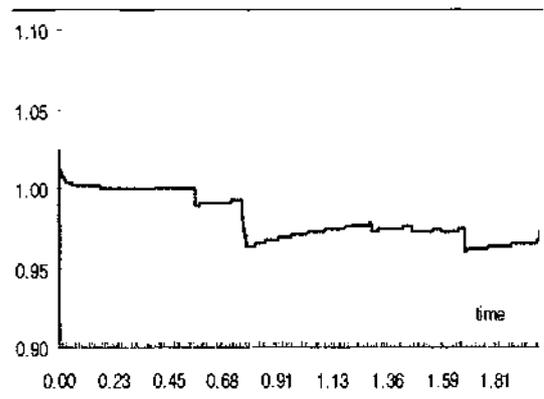


그림 8. +4% time shifting 후의 시간축 변형 (세로축: 신호처리 전 한 신호의 위치/처리 후 같은 신호의 위치)
Fig. 8. Time scaling modification after +4% time shifting (axis of ordinates: Position of a signal before the processing / Position of the same signal after the processing).

V. 결론

본 논문에서는 오디오 워터마킹 특성으로서 MPEG 심리음향모델 I의 토널 마스크 특성을 제안하였고 토널 마스크 구성 신호들의 에너지 변화가 비지각적임을 기존의 연구를 통하여 분석하였다. 그리고 톤과 톤의 좌,우 주파수 신호의 에너지 관계가 여러 신호처리 후에도 적은 변동률을 보임을 실험에 의해 분석함으로써 이 에너지 관계의 변조를 이용하여 강인한 워터마킹을 할 수 있음을 확인하였다.

이러한 토널 마스크 특성을 이용한 오디오 워터마킹 기법을 제안하고 그 성능을 평가하였다. 비지각성 평가 결과, 클래식과 팝에 대해 각각 평균 SDG -0.285, -0.340의 결과를 보였고 강인성 평가 실험에서는 잡음 추가, 대역 통과 필터링, MP3압축, 재샘플링, 반향 처리, 주파수

등화의 신호 처리를 각각 적용한 후 워터마크 추출 오류율을 계산하였다. 클래식과 팝에 대해 각각 평균 오류율 0.11%, 1.26%의 결과를 보였다.

제안한 기법은 프레임의 크기와 토널 마스크의 변조 비율 α , β 를 조정하면서 비지각성과 강인성간의 수준을 조절할 수 있으므로 워터마킹 응용분야에 따라 융통성 있는 적용이 가능할 것이다.

참고 문헌

1. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems journal*, 35, 1996.
2. 오현오, 김현욱, 윤대희, 석종원, 홍진우, "강인한 오디오 워터마킹을 위한 새로운 변형 커널 설계," *한국음향학회지*, 20 (2), 66-76, 2001.
3. M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Signal Processing*, 66, 337-355, 1998.
4. D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking," *ICASSP 2001*, 3, 1345-1348, 2001.
5. P. bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on multimedia*, 3, 232-241, June 2001.
6. J. Seok, J. Hong, and J. Kim, "A novel audio watermarking algorithm for copyright protection of digital audio," *ETRI Journal*, 24 (3), Jun, 2002.
7. J. D. Gordy and L. Bruton, "Performance evaluation of digital audio watermarking algorithms," *IEEE MWSCAS 2000*.
8. T. Painter, and A. Spanias, "Perceptual coding of digital audio," *Portions published in Proc. IEEE*, April, 2000.

9. ISO/IEC IS11172-3, 1993.
10. H. Oh, *Robust Audio Watermarking Using Imperceptible Echo Embedding Technique*, Ph. D thesis, Yonsei Univ, Korea, 2001.
11. E. Zwicker, *Psychoacoustics-Facts and Models*, 208-213, Springer-Velag, 1982.
12. SDMI, FRWG000224-01, "Call for proposals for phase II screening technology version 1.0," Feb. 2000.
13. 이 순, "통계학," 법문사, 515-549, 2000.
14. ITU-R Rec. BS,1116, 1994.
15. 정사라, 홍진우, "시간축 변형을 고려한 디지털 오디오의 계층적 워터마크," *한국음향학회지*, 21 (8), 719-726, 2002.

저자 약력

● 이 희 숙 (Hee-Suk Lee)



1990년 2월: 창원대학교 전기계산학과 학사
 1992년 2월: 경북대학교 컴퓨터공학과 석사
 2001년 8월: 창원대학교 컴퓨터공학과 박사과정 수료
 * 주관심분야: 오디오 신호처리, 심리음향

● 이 우 선 (Woo-Sun Lee)



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