

심리음향 모델을 이용한 오디오 워터마킹

이희숙^o 이우선
창원대학교 전자계산학과
lhs@sarim.changwon.ac.kr

Audio Watermark Using Psychoacoustic Model

Hee-Suk Lee^o Woo-Sun Lee
Dept. of Computer Science, Changwon National University

요 약

본 논문은 오디오의 masking 특성을 적용한 심리음향 모델을 이용하여 오디오의 고음질을 보장 하면서 잡음과 압축 등의 공격에 강한 오디오 워터마킹 방법을 제안한다. 제안하는 워터마킹 방법은 심리음향 모델에 의해 생산되는 masking thresholds와 원신호의 power spectral density의 각 주파수별 차이 에너지를 이용하여 시간도메인에서 워터마크를 삽입하는 방법으로 오디오의 품질을 유지할 수 있다. 워터마크로는 자기상관성이 강한 PN-시퀀스를 이용하여 강한 워터마킹을 구현한다. 그리고 PN-시퀀스와 같은 이진 시퀀스 워터마크의 검출을 위한 유사도 측정식을 제안한다.

1. 서 론

최근 오디오 디지털 워터마킹분야에서는 인간의 청각 시스템(Human Auditory system)의 특성 중 masking을 이용한 정보의 은폐 연구들이 이루어지고 있다[1][2][3]. masking이란 어떤 음이 다른 음의 영향으로 회피하게 또는 감지되지 않는 HAS의 특성을 말한다. 주파수별 마스킹되는 에너지를 이용하여 워터마크를 삽입함으로써 오디오의 품질을 보장할 수 있다.

L. Boney[2]와 T. Furun[3]는 심리음향모델의 masking thresholds를 구하고 masking thresholds에 근사화한 필터로 워터마크 신호를 필터링한 후, 시간 도메인에서 필터링된 워터마크를 비례요인과 곱하여 원신호에 더하는 방법들을 제안하였다. 이러한 방법들은 masking되는 에너지를 고려하였으나 더해진 워터마크 값들로 masking thresholds를 초과할 수 있으므로 왜곡이 예상될 수 있다.

본 논문에서는 masking thresholds와 원신호의 power spectral density와의 최대값을 취하여 시간도메인으로 역변환하고, 이와 원신호와와의 차이값을 워터마크 삽입에너지로 이용하는 방법을 제안한다. masking되는 에너지를 이용하여 워터마크를 삽입함으로써 오디오의 왜손을 줄일 수 있다. 제안한 방법에서는 워터마크로 PN-시퀀스는 사용하였다. PN-시퀀스는 강한 자기상관특성으로 여러 공격후의 워터마크검출과 오류검출등에 강인함을 보인다. 그리고 대부분의 워터마킹 연구에서 워터마킹 추출에 사용되는 Cross-Correlation은 이진코드의 유사도를 측정하는데 적합하지 않으므로 이진 시퀀스인 PN-시퀀스를 사용한 본 연구의 방법에 적합한 유사도 측정식을 제안하고 이를 이용한 워터마크 검출 결과를 보인다. 이 연구에서는 T. Patinter[4]가 서술한 MPEG-1 Psychoacoustic model I을 적용하여 masking thresholds를 구하였다.

2. 심리음향 모델

MPEG-1의 Psychoacoustic model I은 다음의 과정에 따라 특정 신호 에너지에 의해 masking되는 주파수영역의 에너지 임계곡선(masking threshold curve)을 계산한다.

① 정규화와 Power spectral density(PSD)를 구한다.

입력신호 $s(n)$ 을 식(1)과 같이 정규화한 후, 식(2)에 의해 신

호 $x(n)$ 의 Power spectral density $P(k)$ 를 구한다.

$$x(n) = \frac{s(n)}{N(2^{b-1})} \quad (1)$$

$$P(k) = PN + 10 \log_{10} \left| \sum_{n=0}^N w(n)x(n)e^{j\frac{2\pi kn}{N}} \right|^2 \quad 0 \leq k \leq \frac{N}{2} \quad (2)$$

여기서 N 은 FFT 길이이고 b 는 샘플당 비트 수, $w(n)$ 은 window함수이다. 그리고 PN 은 PSD를 정규화하는 상수로 90.302 dB를 적용한다.

② tonal과 noise masker의 정의

음색을 결정하는 성분인 tonal중 masker의 특성을 가지는 큰 에너지의 주파수 신호를 식(3)에 의해 찾고, 식(4)에 의해 k 주파수의 tonal masker의 에너지 $P_{TM}(k)$ 를 계산한다.

$$S_T = \left\{ \begin{array}{l} P(k) > P(k \pm 1), \\ P(k) > P(k \pm \Delta_k) + 7dB \end{array} \right\}$$
$$\Delta_k \in \left\{ \begin{array}{ll} 2 & 2 < k < 63 \quad (0.17 \sim 5.5kHz) \\ [2.3] & 63 \leq k \leq 127 \quad (5.5 \sim 11kHz) \\ [2.6] & 127 \leq k \leq 256 \quad (11 \sim 20kHz) \end{array} \right. \quad (3)$$

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

tonal masker의 에너지가 구해지면 noise중 masker로 작용하는 에너지 $P_{NM}(k)$ 를 각 critical band마다 식(5)에 의해 계산한다. 이 때 k 는 critical band 내의 기하평균값으로 정해진다.

$$P_{NM}(k) = 10 \log_{10} \sum_j 10^{0.1P(j)} \quad (dB), \quad \forall P(j) \notin \{P_{TM}(k, k \pm 1, k \pm \Delta_k)\} \quad (5)$$

③ masker의 소거와 재구성

앞의 과정에서 구해진 masker중 절대가칭임계치보다 적은 값은 소거한다. 그리고 0.5 bark 내에 두 개의 masker가 있다면 sliding 0.5 bark-wide window를 이용하여 재구성한다.

④ 개별 마스킹 임계치의 계산

재구성된 masker에 의해 주변의 주파수영역에 마스킹되는 임계치를 각각 계산한다. 이는 오디오특성의 spread function 효과를 적용한 것이다. 식(6)은 j 주파수의 tonal masker에 의해 마스킹되는 i 주파수의 마스킹 에너지 $T_{TM}(i, j)$ 를 구하는 것이다. 식(7)은 j 주파수의 noise masker에 의해 i 주파수의 마스킹 에너지 $T_{NM}(i, j)$ 를 구하는 식이다.

$$T_{TM}(i, j) = P_{TM}(j) - 0.275z(j) + SF(i, j) - 6.025 \quad (dB) \quad (6)$$

$$T_{NM}(i,j) = P_{NM}(j) - 0.175z(j) + SF(i,j) - 2.025 \quad (dB) \quad (7)$$

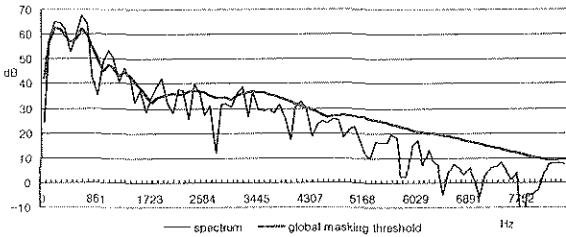
여기서 $SF(i,j)$ 는 spread function으로 j 의 masker 에너지가 주변 주파수 i 에 퍼지는 마스킹 에너지 값을 구하는 함수이다. [4] $z(j)$ 는 j 주파수의 Bark frequency를 말한다.

㉑ 전체 마스킹 임계치의 계산

개별 마스킹 임계치를 이용하여 식(8)에서는 각 주파수에 대한 전체 마스킹 임계치를 계산한다.

$$T_q(i) = 10 \log_{10} \left(\sum_{m=1}^L 10^{0.1T_{NM}(i,m)} + \sum_{m=1}^M 10^{0.1T_{NM}(i,m)} \right) \quad (8)$$

여기서 T_q 는 각 주파수별 절대가청임계치이고 L, M 은 각각 tonal masker의 수와 noise masker의 수이다.



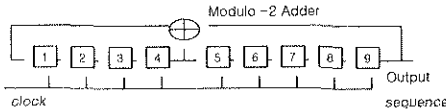
(그림 1) 오디오 신호의 PSD와 masking threshold 곡선

이상의 과정에 의해 주파수 도메인의 masking thresholds를 구한다. (그림 1)은 클래식 기타 연주곡의 임의의 512샘플신호에 대한 power spectral density와 masking thresholds를 보인다.

3. 워터마크의 생성과 삽입

3.1 PN-시퀀스 워터마크 생성

PN-시퀀스는 N 길이의 feedback shift register(그림 2)에 의해 생성되는 $2^N - 1$ 길이의 구간 잡음형 이진 시퀀스로 $2^N - 1$ 의 유일한 코드들을 생산할 수 있고 강한 자기상관성을 가진다. 그리고 모든 시퀀스들은 각각 1의 개수가 0의 개수보다 1개 많은 특성이 있으므로 모든 PN-시퀀스는 균일한 에너지를 가진다.



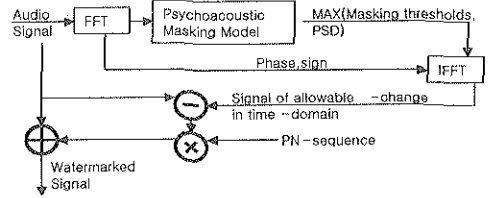
(그림 2) N=9의 Feedback shift register

강한 자기 상관성과 시퀀스들의 균일한 에너지, 저차 증명을 위한 유일한 코드들 다량 생성할 수 있기 때문에 많은 워터마크 방법에 사용된다.

3.2 워터마크 삽입과정

(그림 3)는 제안하는 워터마크 삽입 과정을 보인다. 먼저 2절에서 설명한 과정에 의해 오디오 신호의 masking thresholds를 구한다. 구해진 masking thresholds와 식(2)에 의해 구한 원신호의 PSD중 각 주파수별 최대값을 구한다. 이는 (그림 1)의 두 곡선의 최대값들을 연결한 곡선을 의미한다. 이를 역변환(IFFT)하여 시간도메인의 신호를 구한다. 이 신호는 시간도메인에서 원신호의 음질 훼손없이 변화가능한 신호를 의미한다.

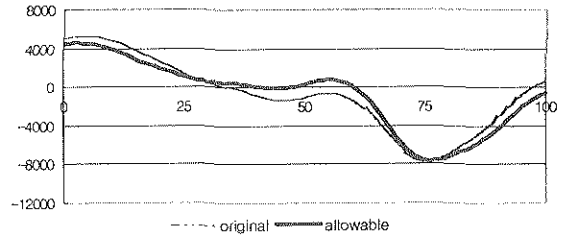
여기서, masking thresholds와 PSD는 신호의 진폭 스펙트럼만으로 계산되었다. IFFT후 원신호와와의 차이값을 정확히 구하기 위해서는 위상과 부호가 복원되어야 한다. 위의 변화가능한 신호의 위상과 부호를 보완하기 위해 원음의 FFT의 출력값인 위상과 부호를 적용하여 역변환한다.



(그림 3) 워터마크 삽입과정

(그림 4)는 원신호와 위에서 서술한 PSD, masking thresholds를 이용하여 역변환한 변경가능(allowable-change)한 신호를 보인다. 본 연구에서는 두 곡선간의 오차를 워터마크 삽입의 적절한 에너지로 사용한다.

이 과정에서 고려한 사항으로, 워터마크의 공격중 오디오의 품질훼손없이 쉽게 워터마크를 훼손할 수 있는 공격이 고주파를 제거하는 필터링이다. 따라서 필터링에 강한 워터마킹을 위하여 우리는 6.5kHz 이상의 고주파 영역은 워터마크 삽입에 사용하지 않았다. 즉 masking thresholds의 고주파영역은 최대값을 구하는 과정에서 적용되지 않았다.



(그림 4) 원 오디오신호(original)와 변경가능한 신호

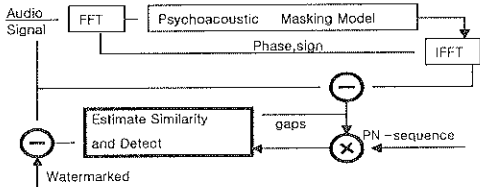
시간도메인에서의 변화가능한 신호가 생산되면 이를 원 오디오 신호와 감산하여 음질의 훼손없이 각 샘플에 워터마크가 들어갈 수 있는 적절한 진폭을 얻을 수 있다. 이렇게 생산된 차이값들과 워터마크인 PN-시퀀스를 배열곱셈하여 원신호의 모든 샘플에 더한다. 즉, PN-시퀀스의 1인 값에 대응되는 원신호의 샘플은 차이값이 더해지고 0인 값에 대응하는 원신호의 샘플에는 아무런 변화가 없다.

이 때 고려한 사항으로, PSD를 생성하는 과정에서 log연산에 의해 오차가 발생하므로, 음질의 더 나은 보장을 위해 삽입값을 정적비율로 줄여 삽입한다. 그리고 PN-시퀀스는 잡음형 시퀀스로 이를 그대로 적용할 경우 고주파성분이 많은 워터마크가 수행되고 저역통과필터의 공격에 쉽게 훼손될 수 있다. 따라서 이 연구에서는 PN-시퀀스의 각 비트를 3번 반복한 형태로 확장하여 워터마크에 적용한다. 즉 샘플링주파수가 44.1kHz 일 경우 PN-시퀀스의 가능한 최고주파수는 7.4kHz이다

이러한 과정으로 워터마크로 인한 음질의 손실을 최대한 줄이고, 최대한의 변경가능한 진폭 정보를 이용하여 시간 도메인에서 워터마크를 삽입한다.

3.3. 워터마크 추출과정

워터마크의 추출과정은 (그림 5)와 같다. 원 오디오 신호에서 심리음향모형을 이용하여 시간 도메인의 변경가능한 신호를 구하는 과정까지는 워터마크삽입과정과 같다. 추출과정의 설명을 위한 몇가지 정의로, gap과 w, w'를 정의한다. gap은 변경가능한 신호와 원신호간의 차이로 식(9)에 의해 정의한다. 그리고 w(i)는 gap과 PN시퀀스간의 배열곱셈(표기: ".*")한 시퀀스로 식(10)에 의해 정의하고 이는 원신호에 더해질 워터마크값을 의미한다. w'는 워터마크된 신호에서 추출한 워터마크 값으로 식(11)로 정의한다.



(그림 5) 워터마크 추출 과정

$$gap(i) = s(i)_{original} - s'(i)_{allowable} \quad (9)$$

$$w(i) = gap(i) * PNsequence(i) \quad (10)$$

$$w'(i) = s(i)_{watermarked} - s(i)_{original} \quad (11)$$

워터마크 추출을 위해 위의 w 와 w' 의 유사도측정을 수행한다. 그런데 워터마크연구에서 주로 사용되는 유사도 측정식은 식(12)와 유사한 형태이다.

$$Similarity = \frac{\sum w w'}{\sum w^2} \quad (12)$$

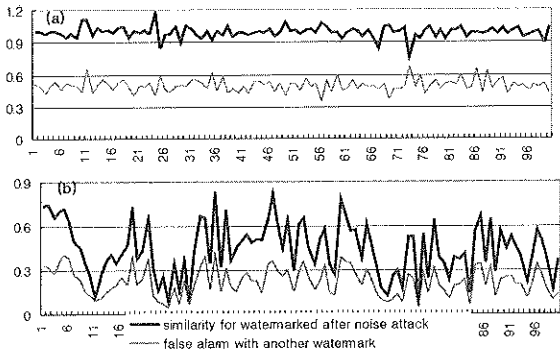
이 식은 워터마크로 이진 시퀀스를 이용하는 워터마크방법에는 적합하지 않다. 이진값 1에 대해서만 유사도를 측정하게 되기 때문이다. 따라서 이진 시퀀스를 워터마크로 사용하는 본 연구에서는 w 와 w' 를 다음의 식(13)과 같이 바꾸어 1과 0에 대한 모든 유사도를 측정하는 식(14)를 제안한다.

$$w(i)_{0-inverse} = \begin{cases} w(i) & | PNsequence(i)=1 \\ gap(i)-w(i) & | PNsequence(i)=0 \end{cases} \quad (13)$$

$$w'(i)_{0-inverse} = \begin{cases} w'(i) & | PNsequence(i)=1 \\ gap(i)-w'(i) & | PNsequence(i)=0 \end{cases} \quad (14)$$

$$Similarity = \frac{\sum w_{0-inverse} w'_{0-inverse}}{\sum w_{0-inverse}^2}$$

제안한 식(14)의 성능을 기존의 식(12)의 유사도 측정결과와 비교하여 보인다.(그림 6)



(그림 6) 제안한 유사도측정식(a)과 기존식(b)과의 결과비교
위의 결과는 워터마크된 오디오신호에 36dB의 잡음을 추가한 후, 삽입한 워터마크와 삽입하지 않은 다른 워터마크와의 유사도를 각각 신호구간별로 측정하여 두 식의 판별력을 비교하였다. 세로축은 유사도이고 가로축은 오디오 신호의 각 구간을 뜻한다. 따라서 제안한 유사도 측정식이 이진 시퀀스 워터마크에 대한 워터마크 추출 판별에 효과적임을 알 수 있다.

이상에서 서술한 과정으로 유사도를 측정하여 적정 임계치 내에 있을 경우 해당 오디오 신호구간에 해당 PN-시퀀스가 삽입된 것으로 판단한다. 실험에서는 PN-시퀀스 길이의 세 배에 해당하는 구간마다 유사도를 측정, 적정 임계치이상인 구간수와 전체구간과의 비율로 워터마크의 삽입을 판단하였다.

4. 공격에 대한 워터마크의 강인성 실험

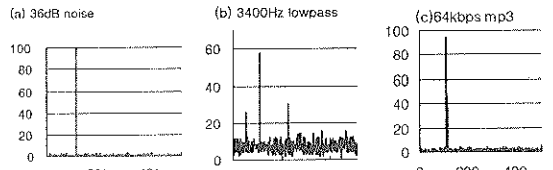
제안한 방법의 실험에는 16bit, 44.1kHz의 오디오 신호를 사

용하였다. 오디오 음질변화의 판단을 쉽게 하기 위해 각각 클래식 기타, 피아노연주를 녹음한 오디오 신호를 사용하였다. 심리음향모델의 masking thresholds 계산시에는 Hann window를 적용하였고 32 샘플중첩의 512 길이의 FFT를 적용하여 해상도를 높였다. 그리고 511길이의 PN-시퀀스를 3배 확장하여 워터마크로 사용하였다. 워터마크 추출에서는 511×3 샘플구간마다 유사도 측정을 하였다.

강인성 실험을 위하여 워터마크된 신호에 잡음공격, 압축을 달린 MP3 coder/decoder 변환, 제한대역통과를 각각 가한 후 워터마크 추출율을 확인하였다. <표 1>에서 보이는 결과값은 임계치 이상의 워터마크 유사도를 보인 구간의 전체구간에 대한 비율을 보인다. 그리고 다른 PN-시퀀스들에 대한 오류검출에 대한 실험을 하였다.(그림 7) shift-register에 십진수로 113을 초기화하여 PN-시퀀스를 생산하고 이를 워터마크한 오디오 신호에 워터마크 훼손이 심한 공격을 각각 가한 후, 모든 생성가능한 워터마크, 즉 1~511의 십진수를 초기치로 생산된 PN-시퀀스들의 워터마크 검출결과를 보인다. 가로축은 PN-시퀀스 생산시의 초기치이고 세로축은 전체구간에 대한 추출구간의 비율이다. 적용된 113의 워터마크 검출만이 높은 추출율을 보임을 알 수 있다. SDMI등의 워터마크 표준협회의 측정기준인 50%의 검지율을 기준할 때 이 워터마크방법은 효과적임을 알 수 있다.

<표 1> 공격에 의한 워터마크 추출 구간 비율

attack	Noise		MP3 변환(압축율)			Lowpass		
	36dB(SNR)	64kbps	96kbps	3400Hz	4500Hz	5500Hz		
검출 구간비율%	100	0.97	0.98	0.58	0.85	0.98		



(그림 7) 511개의 PN시퀀스에 대한 워터마크추출구간비율

5. 결론

본 논문에서는 심리음향모델을 이용하여 음질을 보장하면서 최대한 변화가능한 진폭을 확보하여 PN-시퀀스를 삽입하는 워터마크 방법을 제안, 구현한다. 그리고 이진 시퀀스 워터마크에 대한 유사도 측정식을 제안하였다.

이 방법은 음질에 훼손이 최소화된 범위내에서 많은 워터마크의 정보를 추가함으로써 잡음과 압축변환, 제한대역통과에 대해 강인함을 보인다. 향후 다중 워터마크를 위한 워터마크 삽입공간의 분배와 샘플링률의 변화등의 공격에 대한 연구가 필요하다.

참고 문헌

[1] M. D. Swanson, B. Zhu, A. H. Tewfik, "Current state of the art, challenges and future directions for audio watermarking," ICMCS, Vol. 1 pp.19-24 1999.
 [2] L. Boney, A. H. Tewfik, K. N. Hamdy, "Digital Watermarks for Audio Signals," ICMCS pp. 473-480, 1996.
 [3] T. furon, N. Moreau, P. Duhamet, "Audio public key watermarking technique," Proc. Int. Conf. Acoust., Speech, Signal Processing, June 2000.
 [4] T. Painter, A. Spanias, "Perceptual Coding of Digital Audio," Portions published in Proc. of IEEE, April 2000.
 [5] B. Lincoln, "An Experimental High Fidelity Perceptual Audio Coder Project in MUS420 Win 97," <http://ccrma-www.stanford.edu/~bosse/proj/proj.html>