한국음향학회지 제25권 제2호 pp. 56-60 (2006)

ML 시험 기반 오디오 워터마크 디코더의 최적 변수추정

Optimal Parameter Estimation of the ML Test Based Audio Watermark Decoder

이 진 걸*
(Jingeol Lee*)

*배재대학교 전자공학과
(접수일자: 2005년 9월 13일; 채택일자: 2005년 10월 10일)

시간영역에서 오디오 신호는 일반화된 가우시안 분포를 갖는다는 사실에 기초하여 BER (Bit Error Rate)이 최소가 되는 ML (Maximum Likelihood) 시험 기반 오디오 워터마크 디코더의 최적 변수추정방법을 제안하였다. 제안한 방법이 기존의 변수추정방법이나 상관관계에 기초한 디코더 보다 성능이 우수함을 실험적으로 증명하였다.

핵심용어: 워터마크, ML, BER, 최적화

분류번호: 뉴미디어 분야 (13.4)

Based on the fact that audio signals in the time domain have the generalized Gaussian distribution, an optimal parameter estimation of the ML (maximum likelihood) test based audio watermark decoder, which leads to the minimal bit error rate, is proposed. Its superiority of performance over the existing estimation and the conventional correlation based decoder is demonstrated experimentally.

Keywords: Watermark, ML, BER, Optimization

ASK subject classification: New Media (13.4)

I. 서 론

디지털 기술이 발전함에 따라 정지영상, 동영상, 오디오와 같은 디지털 매체의 생산 및 전파가 용이해짐과 동시에 매체의 재생산과 같은 저작권침해 또한 쉽게 이루어지고 있다. 이러한 환경 속에서 워터마킹은 디지털 매체의 저작권보호의 수단으로 주목을 받아왔다[1]. 대부분의 워터마크 디코더는 시각적으로나 청각적으로 지각되지 않도록 형상화된 워터마크가 삽입된 신호와 워터마크와의 상관관계에 기초한다[1-5]. 통신이론에 기초한 워터마킹 모델의 관점에서 삽입되는 워터마크는 신호이고 원래 신호는 잡음이 되는데 이 잡음이 가우시안 분포일 때만이 상관관계가 최적임은 널리 알려져 있다[2, 6]. 이와 관련하여 Hernandez는 ML (Maximum Likelihood) 시험 기반 정지영상 워터마크 디코더를 제안하였는데 이는 영상의 DCT 계수들이 가우시안 분포 보다 일반화된 가우시안 분포에 더 잘 맞는다는 사실에 근거한다[2]. 그는 시각 모델에 의해 형상화된 워터마크가 삽입된 영상에 ML 시험을 적용하여 DCT 분포와 관련된 변수를 가진 ML 시험이 상관관계에 기초한 디코더 보다 BER (Bit Error Rate) 및 탐지성능이 우수하다고 결론지었다. 한편, 시간영역에서의 오디오 신호도 일반화된 가우시안 분포를 가지는 것으로 알려져 있다[7]. 따라서, 시간영역 오디오 워터마크 디코더에서 ML 시험이 기존의 상관관계 보다 성능이 우수할 것으로 기대된다. 본 논문에서는 BER이 최소가 되는 ML 시험 기반 오디오 워터마크 디코더의 최적 변수추정방법을 제안하였으며 제안된 방법이 기존의 변수추정방법이나 상관관계에 기초한 디코더 보다 성능이 우수함을 실험적으로 증명하였다.

책임저자: 이 진 걸 (jingeol@pcu.ac.kr)
302-735 대전광역시 서구 도마2동 439-6 배재대학교 전자공학과
(전화: 042-520-5707; 팩스: 042-520-5773)

II. ML 시험 기반 디코더

다음과 같이 오디오 신호는 각각 N 개의 샘플로 구성된 N_s개의 프레임 (frame)으로 나누어진다.

$$x_k(i)=x(kN+i) \quad i=0,\mathrm{K},N-1, \quad k=0,\mathrm{K},N_S-1 \tag{1}$$

한 프레임 내의 오디오 샘플들 $x_k(i)$은 다음의 식과 같은 일반화된 가우시안 분포를 가진다[2].

$$f_x(x)=Ae^{-|\beta x|^c} \tag{2}$$

A와 β는 다음과 같이 변수 c 와 표준편차 σ 의 함수이다.

$$\beta=\frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}, \quad A=\frac{\beta c}{2\Gamma(1/c)} \tag{3}$$

Γ는 감마 함수이다. 일반화된 가우시안 분포는 가우시안 분포와 라플라시안 분포를 포함하는데 이는 각각 $c=2$와 $c=1$에 해당한다.

프레임 별 워터마크 삽입 및 디코더에서의 ML 시험은 다음의 식과 같다.

$$y_k(i)=x_k(i)+b_k t_k(i)s_k(i) \tag{4}$$

$$r_k=\sum_i \frac{|y_k(i)+t_k(i)s_k(i)|^{c_k}-|y_k(i)-t_k(i)s_k(i)|^{c_k}}{\sigma_k^{c_k}} \tag{5}$$

$b_k \in \{-1,1\}$는 삽입되는 비트이고, $t_k(i)$는 마스킹 문턱값이며, $s_k(i)$는 그 값이 비밀 키 (key)에 따라 다른 유사 랜덤 수열이다. 비트의 디코드된 값은 다음과 같이 정해진다.

$$\hat{b}_k=sign(r_k) \tag{6}$$

수식 (5)에 의해 주어진 ML 시험은 $c_k=2$ 인 경우 분산에 의해 정규화된 상관관계가 된다.

III. 변수 c_k의 추정

수식 (5)의 c_k 값은 한 프레임 내의 오디오 샘플 평균과 샘플 분산을 매칭 (matching) 시켜 추정할 수 있다 [2, 8].

$$c_k=F^{-1}\left(\frac{E[|x_k(i)|]}{\sigma_k}\right) \tag{7}$$

$$F(c_k)=\frac{\Gamma(2/c_k)}{\sqrt{\Gamma(1/c_k)\Gamma(3/c_k)}} \tag{8}$$

Hernandez는 "변수 c_k 를 추정하는 제안된 방법이 디코딩이나 탐지 성능을 최적화하는 c_k 의 값을 구하는 것과 등가적이지는 않지만 이 방법이 비슷한 결과를 낼 것이다" 라고 서술하였다[2]. 그러나, 다음 장의 실험결과에서 보듯 기존의 추정치는 BER에서 상관관계 이상으로 성능향상을 이루지 못한다. 그러므로 본 논문에서는 BER이 최소가 되는 새로운 추정방법을 제안하였다.

디코더의 출력을 가우시안 랜덤 변수로 가정하여 디코더 출력의 신호대잡음비는 다음의 식으로 주어진다[2].

$$SNR=\frac{(E[r_k])^2}{Var[r_k]}=\frac{\left(\sum_i \frac{E[r(i)]}{\sigma_k^{c_k}}\right)^2}{\sum_i \frac{Var(r_i)}{\sigma_k^{2c_k}}} \tag{9}$$

$$E[r(i)]=\frac{1}{2}\left[|x_k(i)++2t_k(i)|^{c_k}+|x_k(i)-2t_k(i)|^{c_k}\right]-|x_k(i)|^{c_k} \tag{10}$$

$$Var(r(i))=\frac{1}{4}\left[|x_k(i)++2t_k(i)|^{c_k}-|x_k(i)-2t_k(i)|^{c_k}\right]^2 \tag{11}$$

BER의 값은 신호대잡음비가 높을수록 낮아진다는 사실에 기초하여 최적의 변수 c_k 추정치는 다음과 같이 구할 수 있다.

$$c_{opt}=\max_c \ SNR \tag{12}$$

제안된 추정방법은 BER에서 최고의 성능이 이루어지도록 디코더의 출력을 이용하는 반면 기존의 추정치는 오디오 샘플의 분포로부터 구해진다. Hernandez가 서술한 것처럼 이러한 분포로부터 구해진 추정치는 디코드의

BER에서 최고 성능이 되는 최적의 값이 되지 못한다. 반면, 최적화기법으로 수식화된 제안한 추정방법은 디코드의 출력에서 신호대잡음비가 최대가 되도록 함으로서 BER이 최소가 된다는 점에서 그 타당성이 있다.

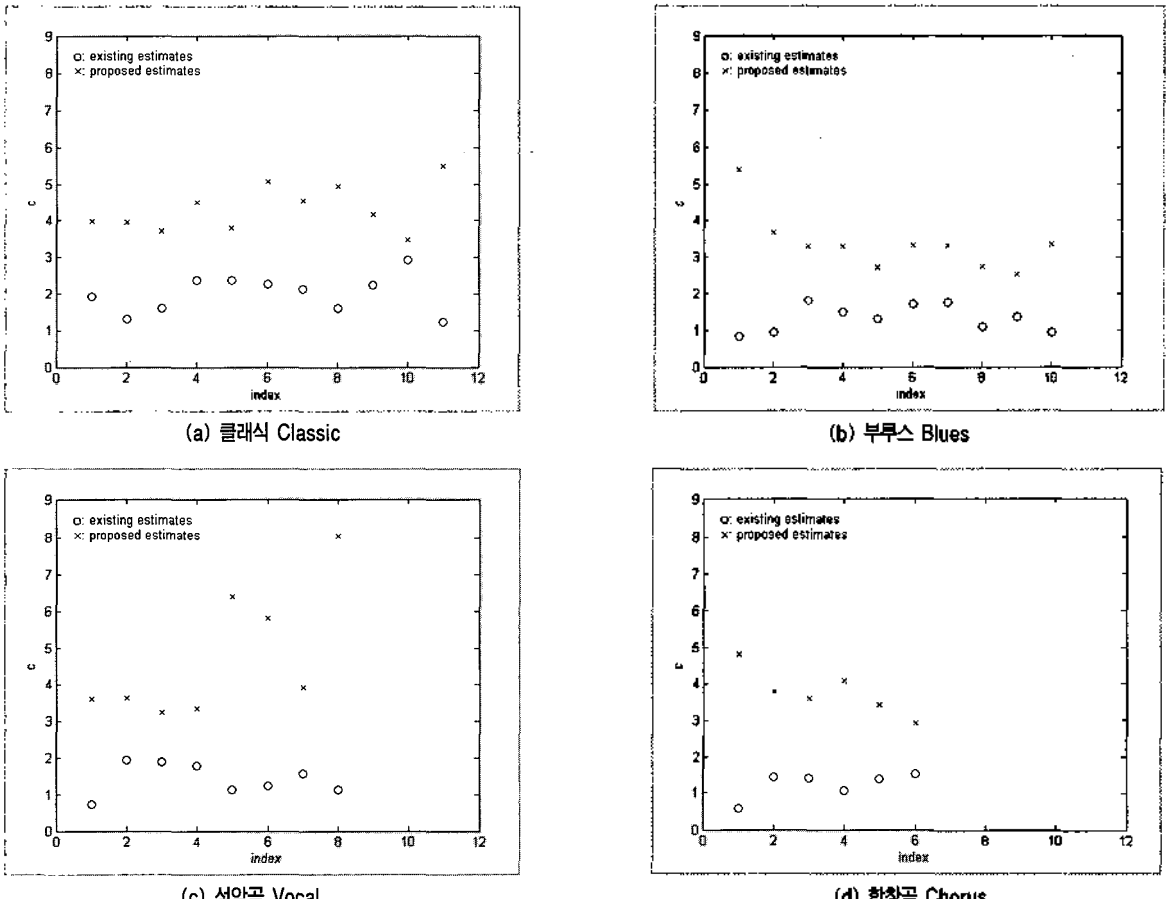
IV. 실험결과

오디오 신호에 대한 마스킹 문턱값은 심리음향모델을 이용하는 방법이 있으나[4] 계산시간이 많이 소요되므로 Bassia가 제안한 방법을 이용하여 다음과 같이 간단히 구하였다[3].

$$t_k(i) = 0.01|x_k(i)| \tag{13}$$

심리음향모델을 이용한 방법으로 지각되지 않는 한도 내에서 최대 크기의 워터마크를 설계할 수 있으나 이 방법은 결과적으로 Bassia가 제안한 방법과 같이 마스킹 문턱값이 신호의 크기에 따라 변화하고, 본 논문은 ML 시험 기반 디코더의 최적 변수추정을 다루므로 마스킹 문턱값은 연산이 간단한 Bassia가 제안한 방법을 채택하였다. 유사 램덤 수열은 -1과 1의 값을 취하고 $s_k(i) \in \{-1, 1\}$, 비트 값 $b_k \in \{-1,1\}$ 이 마스킹 문턱값과 램덤 수열에 곱해져 프레임에 삽입된다. 시험 오디오 (샘플링 주파수: 44100 Hz, 길이: 각각 약 30 초)로 클래식 (Tchaikovsky symphony No. 6), 부루스, 성악곡, 합창곡을 사용하였다. 프레임의 길이를 다소 길게 2^{17} 개의 샘플로 ($N = 2^{17}$, 약 3초) 한 경우, 시험 오디오에 대한 기존의 방법에 의한 변수 추정치 및 제안한 방법에 의한 추정치를 그림 1에 나타내었다. 그림에서와 같이 오디오 신호의 대부분의 프레임은 기존의 방법에 의한 변수 추정치가 1과 2사이에 존재하는데 이는 오디오 신호가 일반화된 가우시안 분포를 가짐을 나타낸다. 또한, 제안한 방법에 의한 추정치는 기존의 방법에 의한 추정치 보다 모든 프레임에서 더 큰 값을 가지며 두 추정치 간 상관관계는 없음을 알 수 있다.

프레임의 샘플 개수를 2^{10} 개에서 두 배씩 증가하여

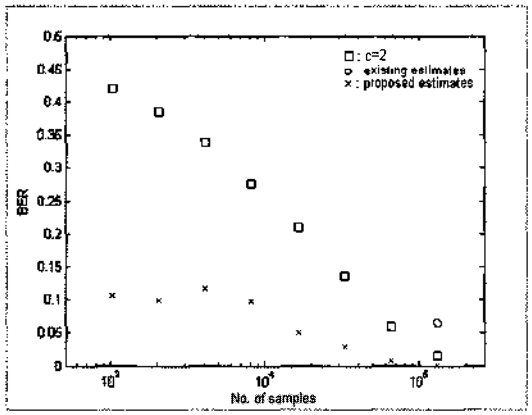


(a) 클래식 Classic (b) 부루스 Blues

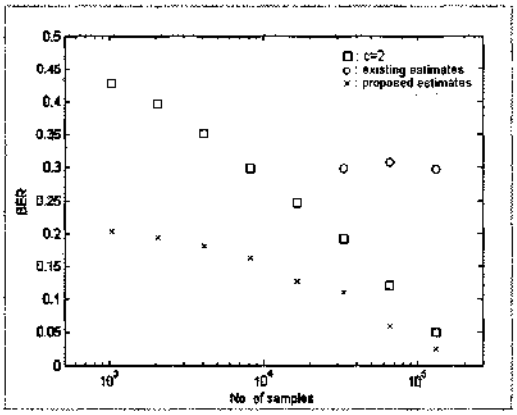
(c) 성악곡 Vocal (d) 합창곡 Chorus

그림 1. 변수추정치
Fig. 1 Parameter estimates.

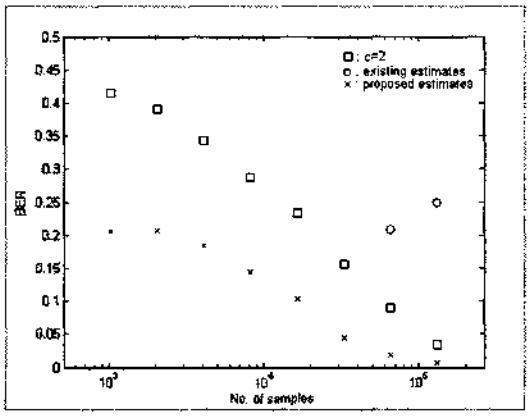
2^{17} 개까지 변화시키면서 측정한 BER을 그림 2에 나타내었다. 그림에서 측정값(BER)은 $b_k=-1$와 $b_k=1$에 대해 프레임의 샘플 개수가 2^{10}개인 경우 1개, 2^{11}개인 경우 2개, …, 2^{17}개인 경우 128개의 램덤 수열을 사용하여 구하였다. 따라서, 클래식, 부루스, 성악곡, 합창곡에 대한 BER은 오류 비트의 수를 각각 2816, 2560, 2048, 1536으로 나눈 값이다. 측정결과로부터 제안한 방법에 의한 추정치가 모든 오디오에 대해 최고의 성능을 내는 반면, 기존의 추정치로는 성능향상을 이루지 못하며 오히려 상관관계에 해당하는 $c_k=2$의 일정한 값을 사용하는 경우보다 성능이 떨어짐을 알 수 있다. 제안한 방법에 의한 BER은 기존의 상관관계에 의한 방법보다 모든 프레임 크기에 대해 반 이하로 낮음을 알 수 있다. 또한, 프레임의 크기가 상대적으로 작은 경우 기존의 추정치에 의한 BER 측정값이 누락되어 있는데 이는 이 경우 수식 (7)의 해를 구할 수 없는 것에 기인한다. 따라서, 제안한 방법은 기존의 방법에 비해 계산적으로도 더 안정화되어 있음을 알 수 있다.



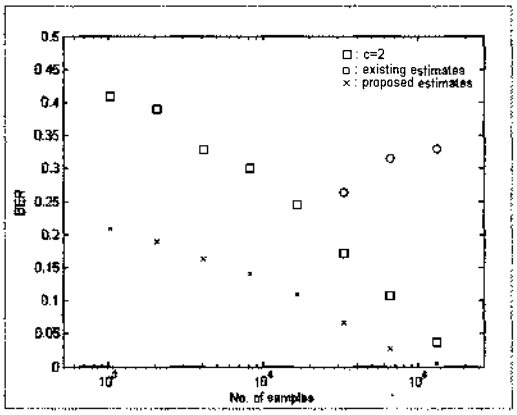
(a) 클래식 Classic



(b) 부루스 Blues



(c) 성악곡 Vocal



(d) 합창곡 Chorus

그림 2. BER 측정
Fig. 2 BER Measurements.

V. 결 론

시간영역에서 오디오 신호는 일반화된 가우시안 분포를 갖는다는 사실에 기초하여 BER이 최소가 되도록 ML 시험 기반 오디오 워터마크 디코더의 출력을 이용하는 새로운 변수 추정방법을 제안하였으며 제안한 방법의 우수성을 실험적으로 증명하였다. 실험결과에서 제안한 방법이 모든 시험 오디오에 대해 기존의 변수추정방법이나 상관관계에 기초한 방법 보다 BER 측정값이 더 낮음을 보였으며 계산적으로도 기존의 추정방법보다 더 안정화되어 있음을 보였다.

참 고 문 헌

1. Christine I. Podilchuk, and Edward J. Delp, "Digital Watermarkings, Algorithms and Applications," IEEE Signal Processing Magazine, 33 46, July, 2001
2. Juan R. Hernandez, Martin Amado, and Fernando Perez-Gonzalez, "DCT-Domain Watermarking Techniques for

Still Images: Detector Performance Analysis and a New Structure," IEEE Transactions on Image Processing, **9** (1), 55-68, January, 2000.
3. Paraskevi Bassia, Ioannis Pitas, and Nikos Nikolaidis, "Robust Audio Watermarking in the Time Domain," IEEE Transactions on Multimedia, **3** (2), 232 241 June, 2001
4. Mitchell D. Swanson, Bin Zhu, Ahmed H. Tewfik, and Laurence Boney, "Robust Audio Watermarking using Perceptual Masking," Signal Processing, 66, 337 355, 1998
5. 우동훈, 정의필, "특정 주파수계수를 이용한 오디오 워터마킹," 한국음향학회지, **22** (4), 269 275, 2003
6. Ingemar J. Cox, Matthew L. Miller, and Jeffrey A. Bloom, "Digital Watermarking," Morgan Kaufmann Publishers, San Francisco, 2002
7. Te-Won Lee, *Independent Component Analysis* (Kluwer Academic Publishers, Boston, 1998)
8. Stephane G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, **11** (7), 674-693, July, 1989

저자 약력

• **이 진 걸 (Jingeol Lee)**

1981년 2월: 고려대학교 전자공학과 졸업 (공학사)
1985년 8월: 고려대학교 대학원 전자공학과 졸업 (공학석사)
1994년 12월: 미국 University of Florida 전기공학과 졸업 (공학박사)
1982년 9월~1990년 12월: 국방과학연구소, 연구원
1995년 6월~1997년 2월: 삼성전자주식회사, 수석연구원
1997년 3월~현재: 배재대학교 전자공학과, 부교수
※주관심분야: 워터마킹, 악기음향학