

피치 및 시간 스케일링에 강인한 오디오 워터마킹 기법*

박창목, 변영배, 김종원, 최종욱
(주) 마크애니 부설 연구소

An Audio watermarking method robust against time- and frequency- scaling

Changmok Park, Youngbae Byun, Jongweon Kim, and Jonguk Choi
MarkAny Inc., MarkAny Research Institute
Email: {cmpark, byun, jwkin, juchoi}@markany.com

요약

본 연구에서는 주파수 영역에서의 확산 스펙트럼 방식을 이용한 오디오 워터마킹 기법을 사용하고 있다. 워터마크 삽입은 오디오 신호를 MCLT(Modulated Complex Lapped Transform)로 분석한 후, 특정 주파수 영역의 진폭에 삽입되며 추출은 상관도를 이용하여 추출하게 된다. 워터마크 삽입은 44.1 kHz의 음악에 80 bits의 정보가 4초 단위로 반복적으로 삽입되며, 추출에서는 무작위로 추출된 8초 분량의 오디오 신호로부터 80 bits 비트 열과의 상관도를 계산하여 선정된 문턱 값을 초과하게 되면 워터마크가 존재하는 것으로 판단하게 된다. 피치 스케일링에 대응하기 위하여 120개 정도의 탐색을 수행하며, 시간 스케일링에 대응하기 위하여 상관도의 지역 최대 점을 추출하고, 이러한 지역 최대 점들로부터 추출된 비트 열과 실제 비트 열과의 상관도를 계산하게 된다. 그러나 추출된 비트 열은 삽입 에러와 삭제 에러를 가질 수 있기 때문에 이러한 비트 열과의 최대 상관도를 구하기 위하여 본 연구에서는 동적계획법에 의한 최대 상관도 추출 알고리즘을 제시한다. 제안된 방법은 피치 및 시간 스케일링 변환 뿐만 아니라, 오디오 압축에도 견고함을 보인다.

1. 서론

디지털 워터마크는 네트워크상에서 사용 가능한 상태로 널리 분포, 유통될 수 있는 멀티미디어 데이터 및 출판물과 같이 지적 재산권 보호 대상 성격을 지니는 자료(intellectual copyright property)에 대해 원 데이터에 권리자 및 인증과 같은 추가적인 정보를 삽입하여 데이터에 대한 지적재산을 보호하기 위한 기법이다. 즉, Watermarking 저작권 보호(Copyright Protection) 기술의 일종으로서 영상, 비디오, 오디오 등의 멀티미디어 데이터에 소유주만이 아는 신호를 사람의 육안이나 귀로는 구별할 수 없게 삽입하고 불법 복제 유통 시 이를 추출/검지하여 저작권을 입증할 수 있다. 따라서 영상, 음향, 비디오 등의 자료에 저작자 고유의 신호를 넣고, 후에 사용자들에 의해 데이터가 불법적으로 복제 유통되었을 때 원 소유주의 신호를 추출/검지하는 기술을 의미한다. 이는 CP(Contents

Provider)서버로부터 사용자에게 디지털 콘텐츠 데이터를 제3자가 알아볼 수 없도록 암호화하는 Front-End 기술과는 달리 저작권을 확증해주는 Back-End 기술이다.

2. 기존 연구

워터마크 삽입 알고리즘은 하위 비트 조작기법(Low-bit coding), 위상 부호화 기법(Phase coding), 확산 스펙트럼 기법(Spread Spectrum), 반향 은폐 기법(Echo hiding)으로 분류할 수 있으며, 데이터 삽입을 위한 Transform 방법은 푸리에 변환, Hartley 변환, 웨이블릿 변환 등을 사용한다. 적합한 변환의 선택은 삽입된 데이터의 견고성(robustness), 비가청성(inaudibility)등에 큰 영향을 주는 것으로 알려지고 있다.

2.1 하위 비트조작기법(Low-bit coding)

하위 비트 조작기법은 가장 단순한 오디오 워터마크 삽입 방법으로 각 sampling 지점에 이진정보를 추가하고 삭제함으로써 해서 워터마크를 삽입하는 방법이다. 전송률에 따라서 그 데이터 양이 조절될 수 있다. 예를 들어, 44.1kHz의 sampling 지점은 모두 44100개이다. 따라서 44100개의 비트 정보가 들어갈 수 있는 것이다. 이 방법의 장점은 많이 데이터를 삽입할 수 있는 것이다. 가장 큰 단점은 워터마크 데이터가 너무 약하다는 것이다. 데이터에 noise의 발생이나, re-sampling과 같은 간단한 변화나 조작에 의해서 쉽게 그 워터마크 데이터가 손실된다는 것이다. 하위 비트 조작기법은 주로 시간영역(Time domain)에서 워터마크의 삽입과 추출이 이루어진다. 시간영역에서 워터마크 삽입추출에 관한 알고리즘은 의사난수를 워터마크로 사용하는 방법과 이미지데이터를 워터마크로 사용하는 방법이 있다. 의사난수 방법과 이미지를 워터마크로 사용하는 방법은 하위 비트 조작기법 임에도 불구하고, re-sampling과 noise의 발생에 비교적 강인함(robustness)을 나타낸다[4].

2.2 위상 부호화 기법(Phase coding)

오디오 데이터는 위상(phase)과 크기(magnitude)의

* 본 논문은 국가지정 연구실 사업(과제번호: 2000N-NL-01-C-286)의 지원으로 수행되었음

두 가지로 구성되어 있다. 사람의 귀는 크기에는 민감하지만, 위상의 변화에는 둔감하다. 위상 부호화 기법은 이런 이론을 이용하여 워터마크 정보를 위상에 삽입하는 방법이다. 이 방법은 SNR(Signal to Noise Ratio)을 줄일 수 있는 가장 효과적인 방법이지만 오디오 압축에 취약하다. [4].

2.3 반향 은폐 기법(Echo hiding)

워터마크는 반향의 새가지 성분 초기 진폭(initial amplitude), 쇠퇴율(decay rate), 갈라짐(offset)에 의해서 숨겨진다. 사람은 어떤 시점에서 원본 오디오 신호와 반향 신호를 구분하지 못한다. 그러나 이 방법은 원 신호에 의해서 나타나는 반향 효과로 워터마크 추출이 까다로운 단점이 있으며, 제 3자가 위해 쉽게 감지 할 수 있어 보안에 문제가 있다. [6].

2.3 확산 스펙트럼(Spread Spectrum)기법

확산 스펙트럼 기법은 워터마크의 강인성 때문에 워터마크 기법 중 많이 연구되어지고 있는 분야로서, 시간영역 신호 혹은 주파수 스펙트럼에 조작을 가함으로써 전 대역에 워터마크를 분산하여 삽입하는 방법이다. 일반적으로 삽입되는 데이터는 특정 값에 의해서 무작위로 생성된 의사난수비트열을 사용한다. 기존의 방법에서는 주파수 스펙트럼을 구하기 위하여 FFT(Fast Fourier Transform), MDCT(Modified Discrete Transform)등이 사용되어져 왔다.

m 을 원래 신호의 스펙트럼 Magnitude(dB)라고 할 때 워터마크된 Magnitude는 $y = m + w$ 로 표현할 수 있다. 워터마크 w 는 $-\Delta$ 혹은 $+\Delta$ 을 가지는 N 개의 w_i 로 구성된다. 워터마크된 스펙트럼과 워터마크 비트 열과의 상관도는 <식 1>과 같이 구해질 수 있다[2].

$$Corr = y \cdot m = (m + w) \cdot w = m \cdot w + N\Delta^2 \quad (1)$$

이러한 상관도는 워터마크가 삽입된 부분에서 상대적으로 큰 값을 가지게 될 것이다.

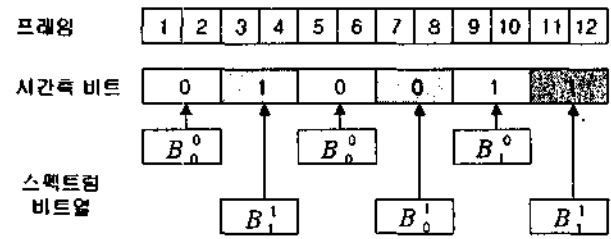
3. 워터마크 삽입

본 연구에서는 워터마크 삽입 및 추출 방법으로서 확산 스펙트럼 기법을 사용하였다. 오디오 신호의 주파수 변환을 위하여 MCLT(Modulated Complex Lapped Transform)을 사용하였다[1]. MCLT는 블록 변환 방법 중 블록 왜곡이 현저히 적으며, 시간 축 이동에 불변한 성질을 가지고 있다. 본 연구에서는 44.1 kHz 오디오 신호를 1024 블록 사이즈의 MCLT로 변환하여 3kHz ~ 8kHz의 스펙트럼 Magnitude에 워터마크를 삽입하였다. 이러한 대역은 워터마크의 비 가청성 및 압축동에 견고성을 고려하여 선택되었다.

스펙트럼에 삽입되는 비트 열은 <식 2>와 같은 4가지 형태를 가지게 된다.

$$B_0^0 = -B_1^0, \quad B_0^1 = -B_1^1 \quad (2)$$

시간 축상에 삽입되는 워터마크 비트 열은 <그림 1> 같은 형태로 삽입되며, 44.1 kHz의 음악에 80 bits의 정보가 4초 단위로 반복적으로 삽입 된다.



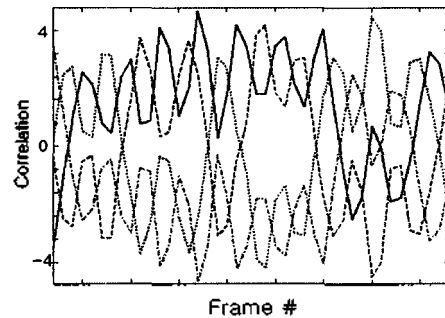
<그림 1> 워터마크 삽입 구조

4. 워터마크 추출

4초 단위로 80bits가 반복적으로 삽입되었기 때문에 무작위로 추출된 8초 정도의 오디오 신호를 이용하여 워터마크의 유무를 판단할 수 있다. 이러한 시간 축상에서 몇 개의 프레임별로 삽입된 비트 정보를 추출하기 위하여 3개 프레임의 스펙트럼 이동 평균을 이용하여 프레임별 상관 신호를 구하면, <그림 2>와 같이 비트가 삽입된 프레임 영역에서 지역 최대 점을 형성하게 된다. 이러한 지역 최대 점을 이용하여 80개의 비트 열을 추출하게 되므로 시간 축 스케일링에 의해 영향을 받지 않게 된다. 그러나 이러한 지역 최대 점은 <그림 2>와 같이 4 가지의 스펙트럼 비트 열을 이용한 4가지 상관도 신호에 분포되어 있으므로 이를 구분하기 위한 전처리 과정이 필요하다.

4.1 상관도 신호의 전처리

4 가지 스펙트럼 비트열을 이용하여 4 가지 형태를 가지는 상관도 신호가 <그림 2>와 같이 구해진다.



<그림 2> 4 가지 비트열을 이용한 상관도 신호

<그림 2>와 같은 상관도 신호가 구해지면 첫번째로 지역 최대 점을 가려내고, 이러한 지역 최대 점은 4 개의 상관도 신호의 동일한 프레임에 존재할 수 없으므로 동일한 프레임에 존재하는 지역 최대 점들 중 가장 큰 값만 사용하게 된다.

4.2 동적 계획법을 이용한 비트 상관도 계산

상관도 신호의 전처리 과정 후에 시간상 순차적으로 지역 최대 점이 나타나는 형태에 따라 추출 비트를 -1

혹은 1로 정의할 수 있다. 그러나 추출된 비트는 삽입 혹은 삭제에러를 포함할 수 있으며, 무작위로 추출된 신호이기 때문에, 워터마크 비트의 이동 정보도 알 수 없다. 다만 8 초의 오디오 신호를 이용했기 때문에 80bits의 정보는 특정 구간에 존재하게 된다. 본 연구에서는 이러한 환경에서 두개의 비트열의 상관도를 최대화하는 알고리즘을 제시하였다. 임의의 오디오 신호에서 추출된 길이 M 인 비트 정보를 U_i 라 하고, 미리 알고 있는 길이 N 인 비트 정보를 T_j 라 하면, 둘 간의 최대 상관도는 <식 3>과 같은 회귀식으로 계산할 수 있다.

$$D_{i,j} = \max(D_{i-1,j-1} + T_{i-1,j-1}, D_{i-1,j} + T_{i-1,j}, D_{i,j-1} + T_{i,j-1}) + d_{i,j} \quad (3)$$

$D_{i,j}$ 는 i, j 까지의 최대 상관도, $d_{i,j}$ 는 i, j 에서 상관도. 즉 $U_i \cdot T_j$ 를 나타내며, $T_{i-1,j-1}$ 는 $i-1, j-1$ 에서 i, j 로의 전이비용을 나타낸다. 전이비용은 <식 4>로 정의된다.

$$\begin{aligned} T_{i-1,j-1} &= 0 \\ T_{i,j-1} &= -1 - d_{i,j} \\ T_{i-1,j} &= -d_{i,j}, \text{ for } j = N-1 \\ T_{i,j} &= -1 - d_{i,j}, \text{ for } j \neq N-1 \end{aligned} \quad (4)$$

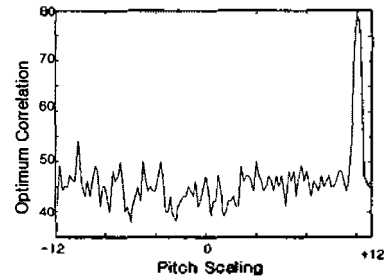
알고리즘의 초기 조건은 <식 5>와 같다.

$$\begin{aligned} D_{i,0} &= d_{i,0}, \quad i = 0, \dots, M \\ D_{0,j} &= -1, \quad j = 1, \dots, N \end{aligned} \quad (5)$$

결국 최대 상관도는 $D_{M,N}$ 을 구함으로써 얻어질 수 있다.

4.3 피치 스케일 탐색

피치 스케일은 스펙트럼 Magnitude의 이동 현상을 발생시키며, 10% 피치를 올릴 경우 100Hz의 Magnitude가 110Hz로 이동하고 1000Hz의 Magnitude가 1100Hz로 이동하게 된다. 반대로 10% 피치를 내릴 경우 100Hz의 Magnitude가 90Hz로 이동하고 1000Hz의 Magnitude가 900Hz로 이동하게 된다. 본 연구에서는 -12% ~ 12%의 피치 스케일을 고려하기 위하여 125번의 상관도 계산을 수행하게 된다. 즉 각 피치 스케일이 발생되었을 때의 새로운 주파수 인덱스를 이용하여 스펙트럼상에서의 상관도를 계산하고, 다시 시간축 상에서 추출된 비트 정보와 T_j 와의 최적 상관도를 구하게 된다. <그림 3>은 10% 피치를 올렸을 경우 125번의 탐색 결과 상관도의 변화를 나타낸다. 10% 부근에서 높은 최적 상관도를 나타냄을 알 수 있다.



<그림 3> 피치 스케일 탐색에 의한 상관도 변화

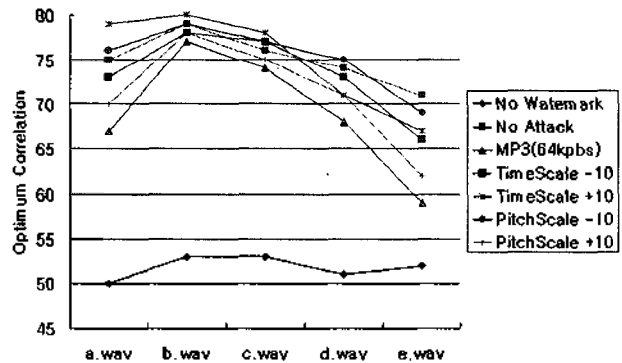
5. 실험

본 실험에서는 팝, 클래식 등 5개의 곡을 선정하였으며, 삽입에 사용된 Δ 는 3dB를 사용하였고 워터마크 삽입에 의한 SNR 결과는 <표 1>과 같다.

	a.wav	b.wav	c.wav	d.wav	e.wav
SNR(dB)	29.	24	23	37	24

dB 스케일로 워터마크를 삽입하였기 때문에 <표 1>에서 보는 바와 같은 조용한 음악(d.wav)인 경우 상대적으로 높은 SNR을 보이고 있으며, 비교적 사끄럽고 복잡한 음악에서는 낮은 SNR을 보인다.

<그림 3>은 무작위로 추출된 오디오 신호 일부를 이용하여 최적 상관도를 계산한 것이다. 8초 길이의 오디오 신호에서 워터마크를 감지하는 시간은 Pentium 600MHz PC에서 약 3.3초이며, 실험에 사용된 소프트웨어는 CoolEdit Pro를 사용하였다.



6. 결론

본 연구에서는 시간 축 이동에 불변한 MCLT를 사용하여 주파수 스펙트럼 상에서 워터마크를 삽입하는 방법을 사용하고 있다. 이러한 시간 축 이동 불변 성질은 워터마크의 주요 공격종인 Time Scaling Attack에 강인한 방법을 제공하고 있다. 그러나 비트 추출에 있어서 비동기화, 삽입 혹은 삭제 에러 때문에 상관도를 정확히 계산하기 어렵다. 본 연구는 동적계획법을 이용하여 최적 상관도를 계산함으로써 이러한 문제를 해결하고 있다. 또한 피치 변환에 대응할 수 있도록 탐색을 수행하여 최적 상관도를 계산하게 된다.

향후에는 앞에서 언급한 전처리 과정을 개선할 필요가 있다. 이러한 전처리 과정은 워터마크를 약하게 삽입하여도 정확한 비트를 추출할 수 있게 할 수 있으리라 사려된다.

7. 참고문헌

- [1] H. S. Malvar, "A modulated complex lapped transform and its application to audio process," Proc. ICASSP, pp 1421-1424, 1999
- [2] Kirovski and H. S. Malvar, "Robust spread-spectrum audio watermarking," IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001.
- [3] 지적재산권 보호를 위한 정보은닉 기술 및 표준화 연구, 한국 전산원, 2000
- [4] Changsheng Xu, Jiankang Wu, Qibin Sun, and Kai xin, "Applications of Digital Watermarking Technology in Audio Signals," J. Audio eng. Soc. Vol 47, No.10, 1999
- [5] L. R. Rabiner/R. W. Schafer, "Digital Processing of Speech Signals," Prentic-Hall, 1978
- [6] W. Bender/D. Gruhl, "Techniques for data hiding," IBM Systems Journal vol 35, nos 3&4, 1996
- [7] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal Shamooh, "Secure spread spectrum watermarking for images, audio and vidio," ICIP'96, vol III, 1996
- [8] Stefan Katzenbeisser, Fabien A. P. Petitcolas, "Information Hiding Techniques for Steganography and Digital Watermarking," Artech House, 2000