

# 주파수 영역에서 주요 피크에 QIM을 적용한 오디오 워터마킹

## Audio Watermarking Using Quantization Index Modulation on Significant Peaks in Frequency Domain

강 중 순, 조 상 진\*  
(Jungsun Kang, Sangjin Cho\*)

울산과학기술대학교 전기전자학부, \*울산대학교 전기공학부

(접수일자: 2011년 6월 22일; 수정일자: 2011년 7월 14일; 채택일자: 2011년 7월 25일)

본 논문에서는 주파수 영역에서의 주요 피크를 선정하고, 주파수 크기 응답에 따라 각기 다른 스텝 사이즈를 사용하는 QIM 오디오 워터마킹을 제안한다. 오디오 신호에 대해 사각 윈도우로 오버랩 없이  $L$  샘플을 취하고 영교차율을 통해 해당 프레임에 워터마크 삽입 가능 여부를 판단 후 푸리에 변환을 통해 주파수 영역에서의 크기 응답을 구한다. 프레임 별 주파수 크기 응답의 최대값에 따라 스텝 사이즈를 결정하고 스펙트럼에서 주요 피크  $n$  개를 검출하여 주변 샘플  $w$  개를 포함한  $n \times (w+1)$  개 샘플에 대해 QIM을 이용하여 워터마크를 삽입한다. 워터마크 검출은 유클리디안 거리 (Euclidean distance)를 이용한 블라인드 검출 방식으로 수행된다. 오디오 워터마크 벤치마킹을 통한 강인성 검증에서 대부분 강인한 특성을 보였다.

**핵심용어:** 오디오 워터마킹, 주요 피크, 주파수 영역, QIM, 워터마크 삽입, 블라인드 검출, 양자화 스텝 사이즈

**투고분야:** 뉴미디어 분야 (13.4)

This paper describes an audio watermarking using Quantization Index Modulation (QIM) on significant peaks in frequency domain. The audio signal is broken up into  $L$  samples length frames with non-overlapping and rectangular window. The zero-crossing rate of each frame is calculated for decision whether it is proper to be watermarked or not. If the frame is legitimate, frequency magnitude response is computed by discrete Fourier transform. For the QIM, we set the quantization step size based on maximum value of frequency magnitude response and select  $n$  significant peaks with  $w$  samples around them in frequency domain, totally  $n \times (w+1)$  samples. Finally, watermark embedding is performed. Decoder extract watermarks based on Euclidean distance, that is a blind detection. The proposed method is robust against many attacks of watermark benchmark.

**Keywords:** Audio watermarking, Significant peak, Frequency domain, Quantization index modulation, Watermark embedding, Blind detection, Quantization step size

**ASK subject classification:** New Media (13.4)

### I. 서론

최근 인터넷 사용이 급증함에 따라 온라인에서의 오디오, 비디오, 영상과 문서를 포함한 수많은 디지털 멀티미디어 콘텐츠 배포도 급속히 증가하고 있다. MP3 파일이나 VOD 서비스 같이 인터넷을 통해 제공되는 멀티미디어 콘텐츠의 대표적인 저작권 보호 방법으로 DRM (Digital Rights Management)이 있는데, 이는 인증된 사용자만이

콘텐츠에 접근, 사용할 수 있도록 하는 소프트웨어로 MP3에 주로 적용해 왔다. 하지만 이 DRM 음원은 서비스 제공업자간이나 각 디바이스 간 상호 운용성의 문제로 사용자에게는 많은 불편을 주어 DRM-free라는 서비스가 개발되었는데, 일정 금액을 지불한 사용자는 DRM이 없는 음원을 무제한으로 다운로드할 수 있도록 하는 것이다. 현재 애플의 아이튠즈와 마이스페이스, 아마존 등에서 이 서비스를 제공하고 있다. DRM-free 음원은 사용자들 자신이 구매한 음원을 마음대로 사용하도록 해 적법 구매를 유도하는 것에 어느 정도 성공한 것처럼 보이나 아직은 이러한 서비스가 불법 다운로드를 대량으로 양산

할 수 있다는 우려를 완전히 불식시키지는 못하고 있다. 따라서 DRM-free 환경에서 콘텐츠의 저작권을 보호하기 위한 대안 기술로 디지털 워터마킹 기술이 대두되고 있다 [1].

디지털 워터마킹은 저작권에 대한 정보를 디지털 멀티미디어 콘텐츠에 워터마크의 형태로 삽입하는 것으로, 이 워터마크 신호는 콘텐츠로부터 쉽게 검출할 수 있어야 하고, 여러 신호처리 공격에 강인해야하며, 정확하게 저작권을 나타낼 수 있어야한다 [2]. 워터마크의 삽입은 Host-interference nonrejecting 방법과 Host-interference rejecting 방법으로 분류한다. 전자는 원신호에 워터마크를 추가적으로 더해주는 방법으로 대역 확산 (spread spectrum) [3], 패치워크 (patchwork) [4] 등이 해당되며, 후자는 워터마크에 따라 원신호를 수정하는 방식으로 LBM (low bit modulation) [5]과 QIM (quantization index modulation) [6] 등이 있다. 이 중 QIM을 이용한 워터마킹은 대역 확산 방식이나 LBM 방식보다 강인하고, 유클리디안 거리 (Euclidean distance)를 이용하는 간단한 블라인드 검출 방식을 사용한다는 장점이 있다 [6]. 하지만 이 방법은 워터마크 삽입 시 사용한 양자화기의 스텝 사이즈를 검출 시에도 동일하게 사용하여야 하므로, 만약 이 정보가 노출되었을 경우 누구라도 워터마크를 삭제할 수 있다는 단점이 있다 [7]. 최근 이러한 문제를 해결하기 위하여 워터마크 삽입 영역 (domain)을 다양화 하고 적응적으로 스텝 사이즈를 결정하는 연구가 다수 발표되었다 [7-10]. 특히 [8]에서는 HAS (human auditory system)을 고려하여 주파수를 바크 (Bark) 스케일로 변환, 두 개의 대역으로 분할한 후 각각의 스텝 사이즈를 설정하였다. 또한 신호의 스펙트럼을 양자화 하는 과정에서 발생할 수 있는 오류를 줄이기 위해 임계값 (스텝 사이즈/5)을 설정하고 이보다 큰 요소들에 대해서만 워터마크를 확산 방식으로 삽입하였다. 신호처리에 강인한 워터마킹을 위해서는 워터마크를 신호의 주요 부분에 삽입함으로써 워터마크의 제거를 시도하였을 경우, 심각한 음질 저하가 발생하도록 설계하는 것이 바람직하는데 [1,8]의 경우는 신호의 주요 부분에 선택적으로 삽입하지 않고, 양자화 오류가 발생하지 않도록 설정한 임계값을 기준으로 삽입하였다. 본 논문에서는 프레임 별 주파수 크기 응답의 최대값에 따라 스텝 사이즈를 결정하고 오디오 신호의 스펙트럼에서 에너지가 큰 주요 피크  $n$ 개를 검출하여 주변 샘플  $w$ 개를 포함한  $n \times (w+1)$ 개 샘플에 대해 QIM을 이용하여 워터마크를 삽입하는 오디오 워터마킹을 제안한다.

## II. 배경 이론 및 제안한 알고리즘

### 2.1. QIM을 이용한 워터마킹

QIM은 우선 인덱스 변조에 의해 정보를 삽입하고, 동일한 인덱스를 갖는 양자화기를 이용하여 원신호를 양자화 하는 것을 의미한다 [6]. 즉, QIM을 이용한 워터마킹을 위해서는 정보 삽입을 위한 디터 변조 (dither modulation)와 양자화기 설계가 선행되어야 한다. 만약,  $N$ 개의 샘플로 구성된 원신호 벡터  $\mathbf{x}$ 에  $L$ 샘플마다 1개의 메시지 비트를 워터마크로 삽입한다고 하면, 식 (1)과 같은 디터 벡터 (dither vector),  $\mathbf{d}$ 가 정의된다.

$$d[k,1] = \begin{cases} d[k,0] + \Delta_k/2, & d[k,0] < 0 \\ d[k,0] - \Delta_k/2, & d[k,0] \geq 0 \end{cases} \quad (1)$$

여기서,  $k = 1, 2, \dots, N/L$ 로서 삽입될 워터마크 인덱스이고,  $d[k,0]$ 는  $[-\Delta_k/2, \Delta_k/2]$ 에 균일하게 분포하는 유사 랜덤 시퀀스 (pseudo random sequence)이며  $\Delta_k$ 는 균일 양자화기의 스텝 사이즈이다 [6].  $d[k,0]$ 와  $d[k,1]$ 은 각각 메시지 비트 0과 1을 삽입하기 위해 사용하는 디터 벡터이다. 원신호는 디터 벡터와 같이 양자화 되는데 삽입을 위한 함수는 식 (2)로 정의한다.

$$\mathbf{s}(\mathbf{x}; m) = \mathbf{q}(\mathbf{x} + \mathbf{d}(m)) - \mathbf{d}(m) \quad (2)$$

여기서  $\mathbf{q}(\cdot)$ 은 스텝 사이즈  $\Delta_k$ 를 갖는 균일 스칼라 양자화기 (uniform scalar quantizer)이다. 이렇게 워터마크가 삽입된 신호  $\mathbf{s}(\mathbf{x}; m)$ 가 다양한 통신 채널을 통과한 후 (또는 공격을 받은 후)의 신호를  $\mathbf{y}$  라고 하면, 디코더에서는 식 (3)과 같은 유클리디안 거리 (Euclidean distance)를 이용하여 메시지 비트를 검출한다.

$$\hat{m}(\mathbf{y}) = \underset{m}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{s}(\mathbf{y}; m)\| \quad (3)$$

### 2.2. 제안한 알고리즘

본 논문에서는 오디오 신호에 대해 사각 윈도우 (rectangular window)로 오버랩 (overlap) 없이  $L$ 샘플을 취하고, 식 (4)의 영교차율 (zero-crossing rate)을 통해 해당 프레임에 워터마크 삽입 가능 여부를 판단한다.

$$zcr = \frac{1}{L} \sum_{i=1}^{L-1} |\operatorname{sign}(x_i) - \operatorname{sign}(x_{i-1})| \quad (4)$$

여기서  $\operatorname{sign}(\cdot)$ 은 샘플의 부호를 반환하는 함수이다.

영교차율은 대상 오디오 신호에 대한 노이즈 프로파일로부터 결정하는데, 노이즈 프로파일은 대상 오디오 신호의 평균 에너지 값을 보다 작은 에너지를 갖는 프레임을 대상으로 선정한다. 만약 해당 프레임이 워터마크를 삽입할 경우 음질의 열화를 발생시킬 수 있는 무음 구간 (silence)으로 판단되면 다음 프레임으로 넘어가고, 그렇지 않으면 푸리에 변환을 통해 주파수 영역에서의 크기 응답을 구한다. QIM을 이용한 워터마킹을 위해 스텝 사이즈는 식 (5)와 같이 결정한다.

$$\Delta_k = \lfloor \frac{\max(|\mathbf{X}_k|)}{\Delta_T} \rfloor \quad (5)$$

여기서  $\mathbf{X}_k$ 는  $k$ 번째 프레임의 주파수 크기 응답이고,  $\Delta_T$ 은 신호 전체에 대한 고정된 스텝 사이즈이다. 이는 프레임별 주파수 크기 응답에 적응적으로 변하여 음질 열화를 최소화하고 공격에 강인한 스텝 사이즈를 제공할 수 있다. 프레임별 스텝 사이즈가 결정되고 나면 스펙트럼에서 주요 피크  $n$ 개를 검출하여 주변 샘플  $w$ 개를 포함한  $n \times (w + 1)$ 개 샘플에 대해 식 (2)를 이용한 워터마크 삽입을 수행한 후 역푸리에 변환으로 워터마크가 삽입된 신호를 시간영역으로 복원한다. 워터마크 검출은 식 (3)을 이용한 블라인드 검출 방식으로 수행된다.

### III. 실험 결과 및 토의

본 논문에서는 워터마킹 대상 오디오 파일로 폭넓은

스펙트럼을 보이는 클래식 음악 두 종류와 팝, 기타, 피아노 연주곡 그리고 음성신호를 사용하였다. 이들 모두 44.1 kHz 샘플링에 16비트로 양자화된 음원이다. 시뮬레이션은 MATLAB을 이용하였고, 오디오 파일을 읽어들일 때 최대값을 1로 정규화 (normalization)함을 감안하여 스텝 사이즈  $\Delta_T$ 는 10으로 설정하였고 프레임 길이  $L$ 은 2048 샘플로 하였다. 영교차율을 이용한 프레임 선택에는 대상 오디오 파일의 유사 무음 구간을 선택하여 노이즈 프로파일을 만들어 임계값을 결정하였다. 주요 피크의 개수는 최적의 성능을 보인 15~20개 사이에서 선택하였고, 주변 샘플 수는 2개로 고정하였다. 강인성 검증을 위하여 표 1과 같은 StirMark [11], SMDI [12], STEP2000 [13] 벤치마크의 시간 및 주파수 영역에서의 공격을 가하여 BER (bit-error-ratio)로 정량화하였고, 시간영역에서의 QIM 기반 워터마킹 방법과 비교하였다. 시간영역에서의 방법에서는 스텝사이즈를 음질의 열화가 발생하지 않도록 실험을 통해 0.01로 설정하였다.

제안한 워터마킹 방법은 대부분의 공격에 약 5% 이하의 BER을 보여 강인성을 확인할 수 있었고 시간영역에서의 QIM 워터마킹 방식보다 월등히 우수함을 확인할 수 있었다. 하지만 몇몇 공격에는 취약성을 보였는데 이에 대한 분석 결과는 다음과 같다.

Down sampling 공격은 원신호의 길이를 변화시키므로 프레임 기반의 워터마킹에 매우 강한 특성을 보인다. 하지만, 만약 디코더에서 원신호에 대한 샘플링 정보를 알고 있다면 재샘플링 (resampling)이 해결책이 될 수 있

표 1. 강인성 검사를 위한 공격의 종류  
Table 1. Attacks for robustness test.

| 공격 종류                 | 파라미터                                | 벤치마킹     |
|-----------------------|-------------------------------------|----------|
| Amplitude Compression | Quantization bit : 16 bit to 8 bit  | STEP2000 |
| BPF                   | Cut-off freq. : 100 Hz, 6 kHz       | SDMI     |
| Codecs                | MPEG-1 Audio Layer3                 | SDMI     |
| Crop                  | Cut and replace time : 2s           | -        |
| Down Sampling         | Sampling freq. : 44.1 kHz to 16 kHz | STEP2000 |
| Echo                  | Delay: 100 ms, Feedback Coeff.: 0.5 | SDMI     |
| FFT_Invert            | FFT size : 1024 samples             | StirMark |
| FFT_RealReverse       | FFT size : 1024 samples             | StirMark |
| Invert                | -                                   | StirMark |
| LPF                   | Cut-off freq. : 10 kHz              | StirMark |
| LSBZero               | -                                   | StirMark |
| Noise Addition        | White noise, S/N : -40 dB           | STEP2000 |
| Normalize             | Maximum value                       | StirMark |
| Nothing               | -                                   | StirMark |
| Resampling            | Sampling freq. : 48 kHz to 44.1 kHz | SDMI     |

표 2. 강인성 검사 결과 비교  
Table 2. Comparison of robustness test.

| 공격 종류                 | Proposed |          |       |       |        |       | Time     |          |       |       |        |       |
|-----------------------|----------|----------|-------|-------|--------|-------|----------|----------|-------|-------|--------|-------|
|                       | Classic1 | Classic2 | Pop   | Voice | Guitar | Piano | Classic1 | Classic2 | Pop   | Voice | Guitar | Piano |
| Amplitude Compression | 0        | 0        | 0     | 0     | 0      | 0     | 0.92     | 0        | 0     | 1.86  | 0      | 0     |
| BPF                   | 0.25     | 1.10     | 5.15  | 5.17  | 2.24   | 0     | 48.85    | 49.51    | 49.57 | 48.11 | 50.24  | 47.88 |
| Codecs                | 2.11     | 10       | 2.0   | 10.34 | 0      | 0     | 41.38    | 45.67    | 48.73 | 35.84 | 52.19  | 52.81 |
| Crop                  | 0.15     | 0        | 0     | 0     | 0      | 0     | 0        | 0        | 0     | 0     | 0      | 0     |
| Down Sampling         | 49.09    | 41.11    | 42.26 | 34.48 | 33.70  | 34.12 | 45.90    | 46.15    | 24.78 | 17.92 | 2.92   | 13.38 |
| Echo                  | 24.67    | 17.77    | 12.37 | 12.09 | 16.85  | 7.69  | 48.79    | 46.15    | 51.68 | 50.94 | 51.21  | 51.40 |
| FFT_Invert            | 0        | 0        | 0     | 0     | 0      | 0     | 50.42    | 45.19    | 50.42 | 43.39 | 50.24  | 52.81 |
| FFT_RealReverse       | 29.16    | 30       | 23.71 | 24.13 | 19.10  | 21.15 | 11.26    | 16.34    | 11.76 | 14.15 | 19.02  | 9.85  |
| Invert                | 0        | 0        | 0     | 0     | 0      | 0     | 50.01    | 48.07    | 54.20 | 61.31 | 56.09  | 43.66 |
| LPF                   | 0        | 0        | 0     | 0     | 0      | 0     | 5.82     | 27.40    | 13.02 | 7.54  | 4.87   | 20.42 |
| LSBZero               | 0        | 0        | 0     | 0     | 0      | 0     | 0        | 0        | 0     | 0     | 0      | 0     |
| Noise Addition        | 0        | 0        | 0     | 0     | 0      | 0     | 0        | 0        | 0     | 0     | 0      | 0     |
| Normalize             | 0        | 0        | 0     | 0     | 0      | 0     | 40.88    | 35.09    | 49.15 | 27.35 | 51.21  | 38.02 |
| Nothing               | 0        | 0        | 0     | 0     | 0      | 0     | 0        | 0        | 0     | 0     | 0      | 0     |
| Resampling            | 0        | 0        | 0     | 0     | 0      | 0     | 0        | 0        | 0     | 0     | 0      | 0     |

다. 이 경우 표 2의 Resampling 공격에 대한 BER에서 알 수 있듯이 100 % 워터마크를 검출할 수 있다.

Echo 공격은 신호를 중첩시켜 원신호의 스펙트럼에 큰 변화를 준다. 이는 주파수 크기 응답에 워터마크를 삽입하는 방법에 대해 매우 강한 특성이 있다. 제안한 알고리즘에서 주요 피크의 개수와 주변 샘플의 개수를 증가시킬수록 BER이 개선되었으나 결국 표 2의 결과에 수렴하였다.

FFT RealReverse 공격은 푸리에 변환 후 실수부의 순서를 역으로 바꾸는 것으로, 실제 DC성분은 다른 값으로 대체되고 허수부분은 한 샘플씩 이동된다. 그 결과 주파수 크기 응답이 달라져 주요 피크가 바뀌게 되었고, BER이 높게 나왔다. 그러나 이 공격은 가역적 (invertible)이어서 공격에 대한 정보가 디코더에 제공된다면, Down sampling 공격과 마찬가지로 BER이 0 %가 될 수 있다.

워터마킹에 대한 공격의 목적이 원 신호를 훼손시키지 않고 삽입된 워터마크만 제거하는 것이라면, 본 논문에서 제안한 알고리즘은 이러한 공격에 매우 강한 특성을 보임을 알 수 있다. 본 논문에서 사용한 모든 오디오 파일과 설명은 “<http://signal.ulsan.ac.kr/doc/qimwm/>”에서 확인할 수 있다.

#### IV. 결론 및 향후과제

본 논문에서는 프레임 별 주파수 크기 응답의 최대값에 따라 스텝 사이즈를 결정하고 오디오 신호의 스펙트럼에서 에너지가 큰 주요 피크와 주변 샘플에 대해 QIM을 이용하여 워터마크를 삽입하는 오디오 워터마킹을 제안하였다. StriMark, SDMI, STEP2000 벤치마킹에 사용되는 공격 중 주로 사용되는 15가지를 선택하여 강인성 검증한 결과, 스펙트럼에 크게 영향을 주는 몇몇 공격을 제외한 나머지 공격에 대해서는 매우 강한 결과를 보였다. 세 가지 벤치마크 프로그램에서 중복되는 공격을 제외하면 대략 50여종이 되는데, 향후 다른 모든 공격에 대해서도 실험이 수행되어야 할 것이다. 또한 제안한 알고리즘은 프레임 기반의 워터마킹이므로 프레임 간 동기화 (synchronization) 문제도 고려해야 할 것이며, 이를 바탕으로 제안한 알고리즘을 개선해 나가야 할 것이다.

#### 감사의 글

이 논문은 2010년 울산과학대학 교내학술연구비 지원에 의해 수행됨 (This work was support by the 2010 Research Fund of Ulsan College).

## 참고 문헌

1. MarkAny, "마크애니 오디오 워터마킹 솔루션", <http://www.markany.com>
2. Editors: J.-S. Pan, H.-C. Huang, and L. C. Jain, *Intelligent Watermarking Technique*, World Scientific, River Edge, New Jersey, 2004.
3. I. J. Cox, J. Killizn, F. T. Leighton, and T. Shamoan, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, pp. 1673-1687, 1997.
4. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Sys. J.*, vol. 35, no. 3, pp. 313-336, 1996.
5. M. D. Swanson, B. Zhu, and A. H. Tewfik, "Data hiding for video-in-video," *Proc. IEEE Intl. Conf. Image Process.*, Piscataway, NJ, vol. 2, pp. 676-679, 1997.
6. B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Information Theory*, vol. 47, no. 4, pp. 1423-1443, May, 2001.
7. 이윤호, 이광우, 김승주, 원동호, 양형규, "QIM 워터마킹 방식에서의 양자화 구간 간격 최적화에 관한 연구", *정보보호학회논문지*, 16권, 1호, 45-53쪽, 2006.
8. N. Khademi, M. A. Akhaee, S. M. Ahadi, M. Moradi, and A. Kashi, "Audio watermarking based on quantization index modulation in the frequency domain," *Proc. 2007 IEEE Intl. Conf. Signal Process. and Comm.*, pp. 1127-1130, Dubai, United Arab Emirates, 2007.
9. K. Sonoda and O. Takizawa, "Audio watermarking using QIM on wavelet packet coefficients," *Proc. 2009 IEEE Intl. Conf. Intelligent Inform. Hiding and Multimedia Signal Process.*, pp. 72-75, 2009.
10. A. Sarkar and B. S. Manjunath, "Double embedding in the quantization index modulation framework," *Proc. 2009 IEEE Intl. Conf. Image Process.*, pp. 3653-3656, 2009.
11. A. Lang, *Stirmark benchmark for audio*, <http://www.witi.cs.uni-magdeburg.de/~alang/smba.php>, 2007.
12. Y. Lin and W. H. Abdulla, "Audio watermarking for copyright protection," Technical Report, School of Engineering Report No. 650, University of Auckland, Auckland, New Zealand, 2007.
13. The Japaneses Society for Rights of Author, Composers and Publishers (JASRAC), Nomura Research Institute, Ltd. (NRI), "Announcement of Evaluation Test Results for "STEP 2000", *International Evaluation Project for Digital Watermark Technology for Music*," <http://www.jasrac.or.jp/watermark/ehoukoku.htm>, 2000.

---

## 저자 약력

---

### •강 중 순 (Jungsun Kang)



1983년: 울산대학교 전자공학과 졸업 (공학사)  
 1989년: 고려대학교 전자공학과 졸업 (공학석사)  
 2009년: 몽골국립과학기술대학 졸업 (공학박사)  
 1991년 ~ 현재: 울산과학기술대학교 전기전자학부 교수  
 ※ 주관심 분야: 무선통신, 이동통신, 멀티미디어 신호처리

### •조 상 진 (Sangjin Cho)

한국음향학회지 제23권 제7호 참조