

그룹 양자화를 이용한 오디오 워터마킹 기술 개발*

신승원, 박창목, 김종원, 최종욱

(주)마크애니, 마크애니 연구소

Development of Audio Watermarking Technique using Group Quantization

Seungwon Shin, Changmok Park, Jongweon Kim, and Jonguk Choi

MarkAny Inc., MarkAny Research Institute,
Email: {swshin, cmpark, jwkim, juchoi}@markany.com

요약

In this paper, we propose a watermarking technique that it is possible to winnow illegal contents from scattered contents on the internet. The identification is performed using an embedded unique content ID by the watermarking technique.

The proposed watermarking technique accepts A/D-D/A conversion and a lot of lossy compression such as MP3, AAC, WMA and Real Audio. Watermark robustness is enabled using group quantization, selecting watermark inserting point, and error correction code. Test results show that the correct extraction is about 90% and SNR is above 50~60dB. The above figures means that the proposed technique is able to extract encoded information at least one more times per audio and that it is very difficult to discriminate between a watermarked audio and a original audio.

1. 서론

인터넷을 통한 디지털 저작물의 유통 시장이 활성화되기 위해서는 저작물의 불법 복제, 전송, 배포 등을 방지하고 모니터링할 수 있는 기술이 절대적으로 요구된다. 국내외에서 멀티미디어의 디지털화에 대한 논란은 끊이지 않고 있으며 이런 논란의 중심에는 기술적으로 불법 행위를 근절시킬 수 있는 암호화 기술과 워터마킹 기술이 있고, 다른 한편으로는 소비자의 이익을 보장하면서 법으로 불법 행위를 막을 수 있는 법 제정에 노력을 기울이고 있다.

암호화 기술은 원천적으로 불법 복제된 저작물의 재생을 불가능하도록 하기 때문에 불법 복제 자체의 행위를 막거나 억제시키는 역할을 할 수 있지만, D/A 단계에서 이미지 캡처링, 오디오 녹음 등으로 해킹이 가능하기 때문에 완벽한 대책이 되지 못한다. 따라서 이러한 단점을 보완하기 위해서 워터마킹 기술이 함께 이용되는데, 워터마킹 기술은 저작물 자체에 정보를 숨겨두

기 때문에 불법 복사된 저작물의 유통 경로를 추적하고 불법 저작물을 식별할 수 있다[1].

2. 선행 연구

지금까지 국내외의 논문을 통해서 알려진 오디오 워터마킹 기술을 주요 응용 기술로 구분하면 Spread Spectrum을 이용한 방법[2][3][4], 반향을 워터마크로 삽입하는 방법[5][6][7], 디지털 필터를 이용한 방법[8], 복제 신호를 이용한 방법[9][10] 등 매우 다양한 연구가 진행되어 왔다.

Spread Spectrum을 이용한 연구는[2][3][4] 워터마킹 분야에서 가장 먼저 활용된 응용 기술로 PN-sequence를 심리음향 모델(psycho-acoustic model) 또는 선형 예측 모델(linear prediction mode)을 이용한 필터로 마스킹하여 삽입함으로써 견고성과 비가청성을 만족시키고, 추출은 상호 상관함수(cross-correlation)을 이용한다. 또, 시간영역에서 삽입하는 방법보다는 FFT와 같은 주파수 변환 장치를 통해서 주파수 영역에서 삽입하는 것이 일반화되어 있다.

Spread Spectrum은 일반 손실 압축을 비롯한 여러 공격에 강인하나 시간축을 변형시키는 공격에는 취약하고, 비가청성을 높이기 위한 심리음향 모델을 적용하는 데 많은 연산량을 요구한다는 단점이 있다.

반향 삽입법[5][6][7]은 매우 짧은 시간 간격을 두고 발생하는 반향을 사람의 귀로 식별하기 어렵다는 점에 착안한 워터마킹 기술로서, Bender에 의해서 처음 제안되었다. 이 기술은 오디오 신호에 대한 이해와 분석이 매우 절실하게 요구되는 기술로 압축에 대한 검출율이 높고, 구현이 쉬우나, 캡스트럼 분석(cepstrum analysis)을 통해서 워터마크의 위치를 쉽게 찾아낼 수 있기 때문에 악의적인 해킹에 노출되기 쉽다.

디지털 필터링을 이용한 방법[8]은 오디오 신호의 특정 주파수 신호를 주파수 차단 필터로 제거하고, 특정 주파수의 성분이 감쇄된 위치를 확인하여 은닉된 정보를 찾는다. 특히, 이 방법은 linear speed change, time scaling에 영향을 받지 않으나 많은 정보를 삽입하기가 어렵고, 반향 삽입법과 마찬가지로 주파수 분석을 통해

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

서 워터마크 정보의 삽입 위치를 판단하기 때문에 해킹에 대한 우려가 높다.

마지막으로 복제 신호 (Replica signal) 컨트롤을 이용한 워터마킹 기술[9][10]은 오디오 신호와 모든 면에서 동일한 특성을 갖고 있으나, 주파수의 위치 또는 위상 지연 등의 일부 특징만을 변경한 복제 신호를 워터마크로 이용한다. 원 오디오 신호로부터 워터마크를 생성한다는 점에서 손실 압축에 견고하지만, 삽입되는 워터마크의 강도가 강해지면 소리가 심하게 변형되는 문제점이 있다. 다만, 복제 신호를 생성하는 매개 변수가 매우 다양하기 때문에 해킹에 대한 안전성이 높다.

현재 알려진 워터마킹 기술들은 대부분 위에서 언급된 기술을 기반으로 해서 개선 알고리즘의 연구가 주류를 이루고 있다.

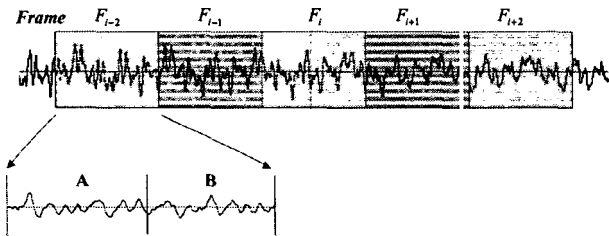
3. 그룹 양자화를 이용한 워터마킹 기술

앞서 살펴본 선행 연구를 보면, 공격에 대한 강인성 (robustness), 비가청성 (inaudibility), 해킹에 대한 안전성 (security) 등을 고루 갖춘 기술을 개발하기 위한 노력들이 진행되고 있음을 알 수 있다.

본 논문에서는 위에서 제안된 기술에 비해서 매우 손쉬운 방법으로 이 조건들을 효과적으로 만족할 만한 워터마킹 알고리즘을 제안한다.

1.1 기본 아이디어

본 논문에서 제안하는 워터마킹 방법은 다음과 같다. <그림 1>과 같이 오디오 신호를 일정한 크기의 프레임으로 나눈다. 프레임의 길이는 100ms이하의 크기로 한다. 프레임을 구분한 다음에 각 프레임에 1비트의 정보를 삽입한다.



<그림 1> 워터마킹 삽입을 위한 프레임 구분

- (1) 임의의 프레임 F_i 를 동일한 길이로 구분하여 각각 "A", "B"라고 하자.
- (2) A와 B의 오디오 신호의 제곱의 합을 수식 1과 같이 구한다.

$$\text{수식 1: } S_A = \sum_{t=i-1/2}^{t=i-1/2} s^2(t), S_B = \sum_{t=i-1/2}^{t=i-1/2} s^2(t)$$

- (3) 다음에, S_A 와 S_B 를 이용해서 이 오디오 신호의 특성값을 수식 2와 같이 구한다.

$$\text{수식 2: } F = \frac{S_A - S_B}{S_A + S_B}$$

- (4) 이때, 오디오 특성값 F 를 변경하여 비트 정보를 삽입하는데, 만약 비트 정보 "0"을 삽입하고자 할 때에는 집합 Q_0 의 원소중에서 가장 가까운 값으로 F 를 양자화 하고, 삽입하고자 하는 비트 정보가 "1"인 경우에는 집합 Q_1 의 원소중에서 가장 가까운 값으로 F 를 양자화 한다.

$$\text{수식 3: } Q_0 = [-0.7, -0.3, 0.1, 0.5, 0.9] \\ Q_1 = [-0.9, -0.5, -0.1, 0.3, 0.7]$$

- (5) 이때, 오디오 특성값 F 를 양자화 하기 위해서는 수식 4와 같이 삽입 강도 g 를 구하여 A 영역에는 더해주고, B 영역은 빼주면 된다.

$$\text{수식 4: } \begin{cases} A: s'(t) = s(t) + gs(t) \\ B: s'(t) = s(t) - gs(t) \end{cases}$$

- (6) 단계 (5)에서 삽입 강도 g 는 다음과 같은 절차에 의해서 구해질 수 있다. 수정된 오디오 특성값을 F' , 양자화 대상이 되는 값을 Q 라고 하면 수식 4를 수식 2에 대치하여 수식 5와 같이 표현된다. 이를 풀면 최종적으로 삽입강도 g 같은 간단한 식으로 구해진다.

수식 5:

$$F' = \frac{S'_A - S'_B}{S'_A + S'_B} = \frac{\sum (s_1(t) + gs_1(t))^2 - \sum (s_2(t) - gs_2(t))^2}{\sum (s_1(t) + gs_1(t))^2 + \sum (s_2(t) - gs_2(t))^2} = Q$$

수식 6:

$$\frac{(\sum s_1^2(t) + 2g \sum s_1^2 + g^2 \sum s_1^2) - (\sum s_2^2(t) - 2g \sum s_2^2 + g^2 \sum s_2^2)}{(\sum s_1^2(t) + 2g \sum s_1^2 + g^2 \sum s_1^2) + (\sum s_2^2(t) - 2g \sum s_2^2 + g^2 \sum s_2^2)} \\ \approx \frac{\sum s_1^2(t) - \sum s_2^2(t) + 2g \sum s_1^2 + 2g \sum s_2^2}{\sum s_1^2(t) - \sum s_2^2(t) + 2g \sum s_1^2 + 2g \sum s_2^2} = Q$$

수식 7:

$$\sum s_1^2(t) - \sum s_2^2(t) + 2g \sum s_1^2 + 2g \sum s_2^2 \\ = Q (\sum s_1^2(t) - \sum s_2^2(t) + 2g \sum s_1^2 + 2g \sum s_2^2)$$

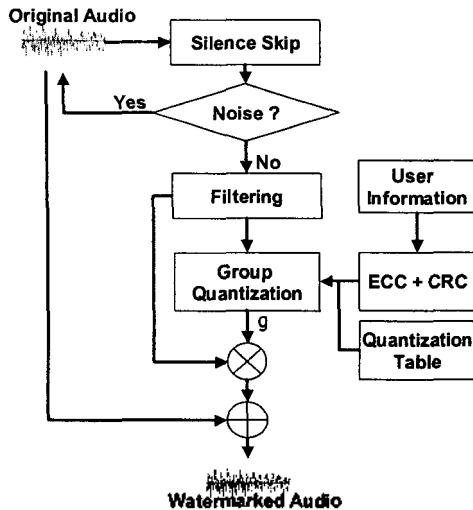
이때, $F_n = \sum s_1^2(t) - \sum s_2^2(t)$, $F_d = \sum s_1^2 + \sum s_2^2$ 라고 하면,

$$F_n + 2gF_d = QF_d + 2gF_nQ$$

$$\text{수식 8: } g = \frac{1}{2} \frac{Q - F}{1 - QF}$$

1.2 워터마킹 삽입 및 추출 방법

삽입을 위한 주요 알고리즘은 2장의 1.1과 같으며, 세부적인 절차는 <그림 2>와 같다.



<그림 2> 워터마킹 삽입 절차도

1차적으로 원본 오디오 (original signal)로부터 정보 삽입을 시작하기 전에 무음, 또는 노이즈 신호는 그대로 저장을 하고, 실제 음성 또는 음향 신호가 발생하는 부분을 찾는다. 이 과정을 Silence skip이라고 본 논문에서는 칭하고 있다. 이 절차는 signal segmentation 분야에서 연구되고 있으므로 본 논문에서는 설명을 하지 않는다.

두번째 단계로 입력된 오디오 신호를 필터링한다. 그룹 양자화를 오디오 신호 전체를 대상으로 처리하게 되면 노이즈가 발생할 우려가 높기 때문에 2-4kHz대역의 신호만을 추출하여 이용한다.

세번째 단계로 필터링된 신호를 앞서 기본 이론에서 설명한 방법에 따라 양자화하기 위한 오디오 특성값 F 와 삽입 강도 g 를 구하여 수식 4와 같이 곱하여 더해주면 워터마크가 삽입된 오디오 (watermarked audio)가 생성된다.

위 절차를 오디오 신호의 모든 프레임에 걸쳐서 반복하면 된다. 프레임에 1비트의 정보를 삽입한다면, 64비트의 정보를 삽입하기 위해서는 CRC 16비트를 포함한 80비트를 터보코드(turbo code)로 코딩하게 되므로 246비트를 삽입해야 한다. 본 논문의 실험에서는 프레임의 크기를 80ms로 이용하였기 때문에 약 20초 정도의 오디오 신호에 64비트 정보를 삽입할 수 있다.

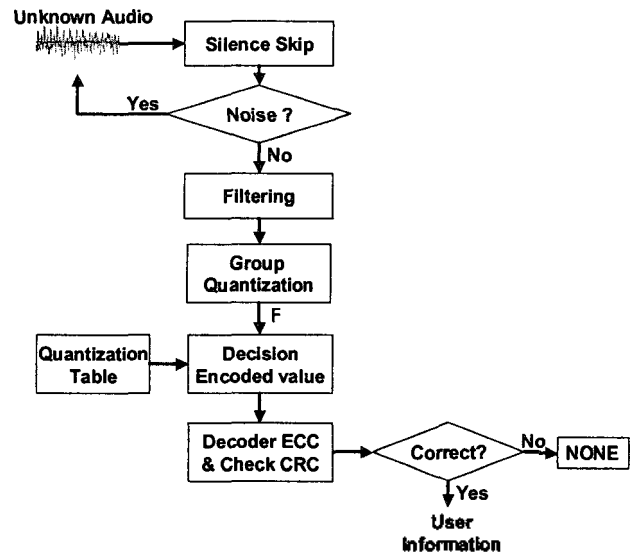
만약, 더 짧은 신호에 많은 정보를 삽입해야 한다면, 필터링의 대상 대역을 변경하여 같은 프레임에 2-3비트를 삽입할 수 있다.

추출 과정은 삽입 과정과 매우 유사하다. <그림 3>은 추출 과정을 도식화 한 것인데, 임의의 오디오 신호에 삽입된 정보를 추출하고자 한다면 삽입과정과 마찬가지로 무음과 노이즈 부분은 아무 처리없이 여과시키고, 실제 신호가 발생하는 부분부터 삽입 절차에 사용된 필터 대역을 필터링하여 오디오 특성값 F 를 구한다.

오디오 특성값 F 를 구한 다음에는 이 특성값이 그룹 양자화 테이블 Q_0, Q_1 중 어떤 집합의 원소와 더 근사한 값을 갖는지 판단해서 그 값을 터보코드의 디코더로 보내면 오류 정정 코드에서는 실제 삽입된 정보 64비트와 CRC 16비트를 출력하게 된다.

마지막으로, 추출된 신호가 정확한지 CRC를 체크해

서 참이면 삽입된 정보를 출력하고, 그렇지 않으면 "NONE"이라는 문구를 출력하게 된다.



<그림 3> 워터마킹 추출 절차도

이 외에 보다 정확한 워터마크 검출을 위해서는 정보의 시작을 나타내는 동기 신호를 삽입함으로써 정보가 시작되는 부분을 정확히 찾아서 추출하는 것이 좋다. 본 논문에서는 기본적인 알고리즘을 설명하는 것에 초점을 두고 있기 때문에 기본 알고리즘 외에 동기 신호를 이용한 정확성 향상은 논의하지 않는다.

4. 실험 및 평가

제안된 알고리즘을 이용해서 64비트를 5곡의 음악에 반복해서 삽입한 다음에 검출실험을 하였다. <표 1>은 여러가지 외부의 공격에 대한 검출 결과이다.

<표 1> 실험 결과

Signal Processing		01	02	03	04	05	Total	Ratio
Embedding Num		10	10	6	16	8	50	
Common	No Processing	10	10	6	16	8	50	100%
	Channel Conversion	7	9	6	16	7	45	90%
	Down Sampling	9	10	3	16	7	45	90%
Time Scaling	10%	9	10	6	16	7	48	96%
	-10%	9	5	6	16	7	43	86%
Pitch Shift	10%	8	9	6	15	5	43	86%
	-10%	7	0	6	0	4	17	34%
Compression	MP3 64kbps	9	10	6	16	7	48	96%
	AAC 64kbps	9	10	6	16	5	46	92%
	ATRAC3 66kbps	9	10	6	16	6	47	94%
	RealAudio 44kbps	9	9	5	15	6	44	88%
Total		95	92	62	158	69		
Ratio		86%	84%	94%	90%	78%		

<표 1>를 보면 64비트를 삽입한 총 횟수는 50번이고, 64비트를 정확히 검출한 횟수는 대략 40회 이상인 것으로 나타났다. 다만, pitch shift 공격에 대해서만 검출률이 30%대로 저조하다. 이는 삽입 대역이 2-4kHz로 이용한 다음에 공격에 의해서 삽입 대역이 변경됨으로써 발생한 것으로 분석된다.

그리고, MP3, AAC, ATRAC3, Real Audio 등 강도 높은 압축률에서도 매우 신뢰할 만한 검출률을 보이는 것을 볼 수 있는데, 제안된 기술이 압축공격에 매우 강인한 특성을 갖고 있음을 알 수 있다.

오디오의 음질에 대한 평가는 정성적인 평가를 골든 이어를 통해서 원본과 워터마크된 오디오의 식별 가능성을 측정해야 하나, 본 논문에서는 SNR를 이용해서 단편적인 비가청성을 평가하였다.

<표 2>에서 볼 수 있듯이, 50-60dB의 SNR를 보이고 있는데, 이 수치는 매우 민감한 청취력을 갖고 있는 사람도 쉽게 구별할 수 없을 정도로 원본과 워터마크된 음악에 큰 차이가 없음을 나타낸다.

<표 2> 워터마크된 오디오의 SNR

Test Audio	01	02	03	04	05	Mean
SNR (dB)	49	62	56	64	63	58.8

5. 결론 및 향후 연구과제

본 논문에서는 그룹 양자화를 이용한 오디오 워터마킹 기술을 제안하였다. 매우 간략화된 처리 과정으로 초당 약 10비트의 정보를 삽입할 수 있으며, ECC와 CRC를 병행하여 이용함으로써 저작권 정보와 기타 사용자 정보의 검출결과에 대한 신뢰성이 매우 높음을 실험을 통해서 증명하였다.

향후 제안된 기술을 실제 시스템에 적용하기 위해서는 본 논문에서 간략히 언급한 바와 같이 오디오 신호의 무음 부분에 대한 처리와 동기 신호를 삽입하여 동기 신호를 검출하여 가급적 프레임의 위치를 삽입 위치와 동일하게 추출 위치를 설정해 주는 것이 더욱 높은 검출율을 보장하는 방법으로 생각된다. 이와 같은 추가적인 기술이 보완되면 실제 시스템에 활용할 만큼 높은 성능을 발휘할 수 있을 것으로 판단된다.

6. 참고문헌

- [1] C. I., Podilchuk, E.j. Delp, "Digital watermarking: algorithms and applications", IEEE Signal Processing Magazine, Volume: 18 Issue: 4, July 2001 pp.33-46
- [2] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain", IEEE Transactions on Multimedia, Vol. 3, June 2001, pp. 232-241.
- [3] D. Kirovski and H. Malvar, "Robust spread-spectrum audio watermarking", 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 2001, pp. 1345-1348.
- [4] L. Boney, A.H. Tewfik, and K.N. Hamdy, "Digital watermarks for audio signals", in International Conference on Multimedia Computing and Systems, pp. 473-480, IEEE, Hiroshima, Japan, 17-23 Jun. 1996.
- [5] D. Gruhl, A. Lu, W. Bender, "Echo hiding", in Proc. Information Hiding Workshop, University of Cambridge, U.K., 1996, pp. 295-315.
- [6] S. W. Foo, T. H. Yeo, and D. Y. Hwang, "An Adaptive Audio Watermarking System", Electrical and Electronic Technology, 2001. TENCON. Proceedings of IEEE Region 10 International Conference on, Volume: 2, 2001 pp.509-513
- [7] H. O. Oh, J. W. Seok, J. W. Hong and D. H. Youn, "New echo embedding technique for robust and

imperceptible audio watermarking", Acoustics, Speech, and Signal Processing, 2001. Proceedings. 2001 IEEE International Conference on, Volume: 3, 2001, pp.1341-1344

- [8] 신승원, 김종원, 최종욱, "디지털 필터를 이용한 오디오 워터마킹", 한국정보보호학회 종합학술발표회 논문집, Vol. 11. No.1 pp.464-468.
- [9] R. Petrovic, "Audio signal watermarking based on replica modulation", Telecommunications in Modern Satellite, Cable and Broadcasting Service, 2001. TELSIKS 2001. 5th International Conference on, Volume: 1, 2001 pp. 227-234.
- [10] 신승원, 김종원, 최종욱, "피치 스케일링을 이용한 강인한 오디오 워터마킹 알고리즘", 한국정보처리학회 2002 춘계학술발표논문집, Vol. 9, No. 1, pp. 209-212, 2002. 4. 12-13.