

Sinusoidal Function Pattern 을 이용한 오디오 워터마킹의 시퀀스 생성 및 Threshold 설정 방안

김 태 훈*, 김 형 중**
강원대학교 제어계측공학과

Sequence generation and measuring threshold of audio watermarking using sinusoidal function pattern

Tae Hoon Kim*, Hyoung Joong Kim**
Department of Control & Instrumentation Engineering
Kangwon National University
hoon@multimedia.kangwon.ac.kr khj@kangwon.ac.kr

Abstract

본 논문에서는 정현파를 이용한 spread-spectrum watermarking 에서 비가청성을 높이기 위한 방법과 효율적인 threshold 설정 방법을 제안한다. 제안하는 방법에서는 spread-spectrum 기법을 사용할 때 계산량이 많이 요구되는 심리음향모델 계산을 피하면서도 가청잡음을 줄이기 위한 방법을 제시한다. 또한 outlier 를 이용하여 워터마크 검출에서 적절한 threshold 설정방법을 제안한다.

이용하여 워터마킹 오디오 신호의 비가청성을 더욱 높일 수 있는 방법을 제시한다. 그리고 spread spectrum 기법에서 워터마크 여부를 판단하는 threshold 를 이용하여 효율적으로 설정하는 방법을 제안한다.

2 장에서는 spread spectrum 및 정현파 패턴을 이용한 워터마킹에 대해 알아보고, 3 장에서는 주파수 분석 및 시간 축에서의 에너지 분석에 의한 음질 향상 방안을 제안하고, 4 장에서는 outlier 를 이용한 Threshold 의 설정방법을 제안하고 5 장에서 결론을 맺는다.

I. 서론

지금까지 다양한 audio watermarking 기법들이 제안되고 발전되어 왔는데 대표적 기법은 spread spectrum [1-2, 4], echo hiding [1, 3], replica modulation [5] 등이 있다.

그 가운데 spread spectrum 기법은 이미지 및 오디오 워터마킹에서 가장 널리 사용되는 방법이다. 이 기법은 통신이론의 대역확산 기법을 이용한 것으로 임의의 PN(Pseudo Noise)시퀀스로 워터마크 신호를 대역확산시켜 오디오 신호에 더하는 방법을 취하고 있다. 삽입한 워터마크 시퀀스와 동일한 시퀀스를 이용하여 상관계수를 구하여 워터마크를 검출한다. 하지만 spread spectrum 기법은 비가청성을 만족시키기 위한 연산과정이 복잡하고 많은 정보를 삽입하기가 어렵기 때문에 실제 시스템에 적용하는데 여러 가지 문제점이 있다.

Liu [4] 등은 spread spectrum 기법에서 임의의 PN 시퀀스 대신 정현파 함수를 워터마킹 시퀀스로 사용하는 방법을 제안했다 [4].

이 논문에서 그는 매우 좁은 주파수 영역을 선택하고 그 주파수를 이용하여 여러 개의 정현파를 생성하며, 그 신호를 워터마킹 시퀀스로 사용한다. 이때 각각의 정현파들은 pseudo-random sequence 를 이용 위상이 변조되어 직교성을 유지할 수 있게 된다. 이 방법으로 생성된 시퀀스들은 특정 주파수 만을 사용하기 때문에 오리지널 신호의 주파수 분석에 의한 워터마킹 시퀀스의 강도 조절이 용이하다는 장점을 가지고 있다.

본 논문에서는 위의 연구를 토대로 실제 시스템에 적용함에 있어 정현파를 사용할 때 얻을 수 있는 장점을

II. 대역 확산 오디오 워터마킹

2.1 Spread Spectrum Watermarking

대역 확산 기법을 이용한 워터마킹은 생성된 PN 시퀀스 패턴을 원본신호에 삽입하고 원본 신호와 PN 시퀀스 패턴간의 상관관계를 이용하여 삽입된 비트를 찾아내게 된다. LFSR(Linear Feedback Shift Register)을 이용하여 생성된 시퀀스들 간에는 아주 좋은 상관관계 성질을 가진다. 그리고 LFSR의 결과 값 중 0을 -1로 치환한 후 생성된 시퀀스들 간의 상관관계수 특성은 다음과 같다.

$$\frac{1}{N} \sum_{i=0}^{N-1} w_a(i)w_b(i) = \begin{cases} 1 & \text{if } a = b \\ -1/N & \text{if } a \neq b \end{cases} \quad (1)$$

여기서 w_a 와 w_b 는 각각 서로 다른 LFSR의 초기 값에 의해 생성된 pseudo-random sequence 이다. 그리고 w_k ($0 < k < M$)로 표현되는 워터마크 시퀀스의 삽입은 다음과 같다.

$$I_i' = I_i + \alpha_i \cdot w_k(i) \quad (2)$$

여기서 α_i 는 워터마크의 강인성, 비가청성을 고려하여 결정 되어야 한다. 워터마크 시퀀스와 워터마킹 신호의 상관계수 q 는 다음과 같이 표현된다.

$$q = \frac{1}{N} \sum_{i=0}^{N-1} I' w_k(i) \quad (3)$$

여기서 I'_i 는 위터마킹 신호를 말하며 N 은 사용된 샘플의 수를 말한다.

2.2 정현파를 이용한 위터마크 패턴

위터마크 패턴에 사용될 정현파 함수의 직교성은 다음과 같이 표현된다.

$$\begin{aligned} & \frac{1}{N_b} \sum_{k=0}^{N_b-1} 2\sin(2\pi kn/N_b) \sin(2\pi km/N_b) \\ &= \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m \end{cases} \end{aligned} \quad (4)$$

여기서 N_b 는 신호의 샘플 사이즈를 말하며 n 과 m 은 사용된 주파수를 말한다. N 개의 정현파를 이용해 주파수 f_0 에서 $f_0 + N$ 범위에서 Pseudo-random Sequence를 생성한다면 위터마킹 시퀀스 w_k 는 다음과 같이 생성된다.

$$P_k = \sum_{n=0}^{N-1} m_n 2\sin[2\pi i(f_0 + n)/N_b] \quad (5)$$

여기서 m_n 은 LFSR에서 생성된 +1 또는 -1의 값을 가지며 이것을 이용하여 위상을 변화시킨다. 그리고 생성된 시퀀스들 간의 상관계수는 다음과 같다.

$$\frac{1}{N} \sum_{i=0}^{N-1} P_a(i) P_b(i) = \begin{cases} 1 & \text{if } a = b \\ -1/N & \text{if } a \neq b \end{cases} \quad (6)$$

여기서 P_a 와 P_b 는 각각의 위터마크 패턴을 말한다. 결국 PN 시퀀스에 의해 위상 변조된 정현파 위터마크 패턴은 PN 시퀀스에서 생성된 위터마크 시퀀스와 동일한 성질을 가지고 있음이 이론적으로 증명 되었다.

III. 위터마크 시퀀스 생성 방법

그런데 spread spectrum을 이용한 위터마킹은 여러 가지 장점에도 불구하고 실제 사용함에 있어 비가청성을 만족시키기 위해 여러 가지 복잡한 연산이 필요하기 때문에 문제점 또한 가지고 있었다. 이 장에서는 정현파 패턴을 이용할 때 주파수 및 시간 축 상의 오리지널 신호의 특성을 빠르게 분석하고 비가청성을 높일 수 있는 방법을 제시한다.

3.1 주파수 분석을 통한 위터마크 시퀀스 생성 방법

정현파 함수를 사용하면 특정 주파수만을 사용하기 때문에 스펙트럼 분석이 용이하다. 주파수 분석을 통한 위터마크 시퀀스 w 생성은 다음과 같이 표현된다.

$$S(n) \leq JND(FFT(I), f_0 + n) \quad (7)$$

여기서 I 는 오리지널 신호이며 JND(Just Noticeable Difference)는 특정 주파수에서 그 주파수 성분을 증가시켜 증가 전과 후의 차이를 명확히 귀로 구분할 수 있는 최소값을 말한다. JND는 그림 1과 같이 나타낼 수 있다. JND를 알면 그 범위 내에서 신호성분을 증가시켜도 귀로 식별이 불가능하기 때문에 비가청잡음을 삽입할 수 있으며, 검출효율을 높이기 위해 강도를 더 높일 수 있는 여지를 확보할 수 있다.

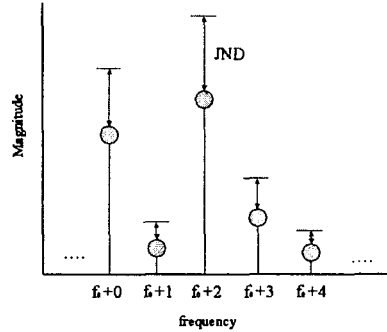


그림 1. 원본 신호의 주파수 크기에 따른 JND

그리고, 각각의 오리지널 신호에서 얻어진 $S(n)$ 을 이용하여 각각의 정현파 함수를 생성할 때 서로 다른 알파 값을 적용할 수 있다. 또한 식 (5)는 다음과 같이 다시 쓸 수 있다.

$$P_k' = \sum_{n=0}^{N-1} m_n 2\sin[2\pi i(f_0 + n)/N_b] \cdot S(n) \quad (8)$$

여기서 각각의 정현파 함수 패턴들은 서로 더해지기 전 미리 분석된 S (강도)에 의해 크기가 변하게 된다. 결국 생성된 w_k' 시퀀스는 원본 신호의 주파수 성분 크기에 따라 강도 조절이 가능하며 음질 향상에 좋은 성능을 가질 수 있다. 그림 2는 식 (8)을 적용하여 생성된 정현파 함수 패턴에 대한 주파수 변환 후의 모양을 보여준다.

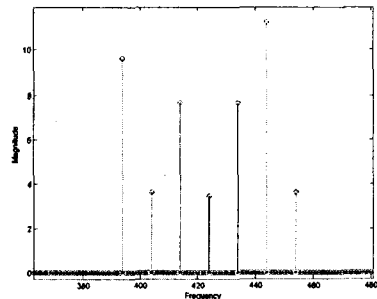


그림 2. 위터마크 시퀀스의 주파수 스펙트럼

그림에서와 같이 원본 신호에 따라 에너지가 큰 주파수에는 큰 정현파가 더해지고 작은 곳에서는 작은 양의 정현파가 더해지므로 위터마킹된 음악 신호에서 잡음을 줄일 수 있다.

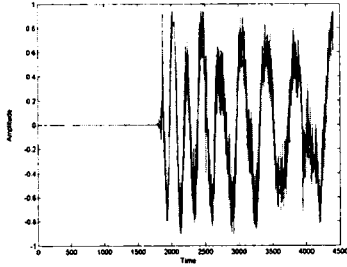


그림 3. 잡음이 생성되는 경우의 예

3.2 시간 축 에너지 분석을 통한 워터마크 시퀀스 생성 방법

3.1의 방법을 이용하여 상당한 음질 향상을 얻을 수 있었다. 하지만 주파수 분석만을 통하여 알파 값을 적용하게 되면 시간 축 상의 에너지 변화를 반영 할 수 없기 때문에 빠르게 변하는 신호에는 대처 할 수 없었다. 그 예로 다음과 같은 신호를 들 수 있다. 그림 3 에서와 같이 에너지가 적은 부분에서 갑자기 큰 에너지로 변하게 되면 에너지가 상대적으로 적은 부분에서는 오리지널 신호의 주파수에 따라 생성된 워터마크 시퀀스라 할지라도 소리가 들리게 된다. 그래서 다음과 같은 방법을 이용하여 시간축 에너지를 측정하고 그것에 따라 알파값을 다시 설정한다.

$$S_k = \sum_{i=0}^{N-1} \left(\frac{1}{n} \sum_{j=i}^{i+n-1} |I(j)| \right) \quad (9)$$

여기서 I 는 오리지널 신호를 말하며 n 은 사용하고자 하는 구간을 말한다. 식 (9)를 이용하여 원하는 구간들의 평균을 구할 수 있다. 그리고 식 (10)을 이용하여 각각의 구간 평균 중 최대값을 +1로 정하고 그보다 작은 값들을 스케일링 한다.

$$S_k' = \sum_{i=0}^{n-1} S_k(i) \frac{1}{\text{MAX}(S_k)} \quad (10)$$

여기서 S 는 각 구간들의 스케일링 값을 말하며 전체 에너지의 크기에 관계없이 비율로만 나타낸다. 그림 4는 식 (10)이 적용된 S' 의 값을 보여준다.

식 (10)에서 설정된 scale 값을 이용하여 다음과 같이 워터마크 시퀀스를 변환할 수 있다.

$$P_k'' = \sum_{i=0}^{N-1} \sum_{j=i}^{i+n-1} P_k'(j) S_k'(i) \quad (11)$$

여기서 P 는 방법 1에서 생성된 시퀀스이며 S 는 식(10)에서 결정된 scale 값이다. 결국 식(11)에서 생성된 최종 워터마크 시퀀스는 그림 5와 같이 오리지널 신호와 매우 유사한 형태를 가지며 비가청성을 더욱 높일 수 있다. 여기서 그림 6에서와 같이 워터마크 패턴의 스케일 변화로 인해 선택된 주파수 주변에 원하지 않는 적은 양의 주파수가 생성된다. 하지만 생성된 주파수들은 선택된 주파수에 비해 상당히 작은 양이므로 마스킹 효과에 의해 충분히 무시될 수 있는 부분이다. 또한 워터마크 복호기에서는 시간축 상에서의 유사도만을 이용하므로 상관계수에도 영향을 미치지 않는다.

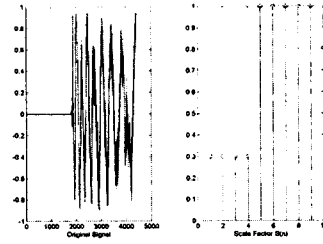


그림 4. 에너지 크기에 따른 비율 설정

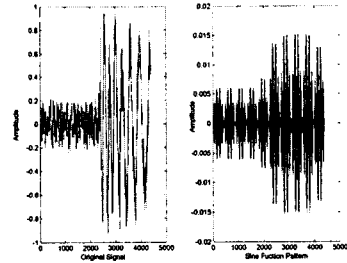


그림 5. 스케일 값에 따른 워터마크 시퀀스

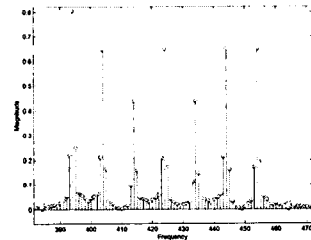


그림 6. 스케일 변환 후 워터마크의 스펙트럼

IV. Outlier를 이용한 Threshold의 설정

Outlier란 통계적으로 데이터를 분석할 때 표본이 되는 데이터들 보다 아주 크거나 작은 값을 말한다. 그래서 표본 값들 중 outlier로 판정된 값을 제외하고 통계적 데이터 분석을 수행하기도 한다.

Spread Spectrum 워터마킹에서는 워터마크를 판별하기 위해 상관관계를 계산하여 피크가 되는 지점의 값을 임계값(Threshold)과 비교하게 된다. 만약 피크가 되는 지점의 값이 미리 약속된 threshold 보다 크다면 워터마크가 존재하는 것으로 판단하고 작다면 존재하지 않는 것으로 판단하게 된다. 하지만 실제 신호에 있어서 임의의 계산 없이 고정된 임계값을 사용하여 워터마크를 판단하기에는 무리가 있다. 그래서 본 논문에서는 통계학에서 사용되는 outlier를 이용하여 threshold를 결정하고 워터마크 존재 여부를 판단하게 된다.

일반적으로 z-score가 3 이상인 데이터, 즉 $|z| \geq 3$ 을 만족하는 데이터를 outlier로 판정한다. 하지만 본 논문에서는 실험적으로 z-score를 계산하여 적당한 outlier를 설정하여 threshold를 정하였다. z-score 계산에 필요한 표준편차(standard deviation)는 다음과 같이

계산된다.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (12)$$

여기서 \bar{x} 는 평균을 나타내며 x_i 는 각각의 샘플을 나타낸다. 그리고 n 은 총 샘플 수를 말한다. z-score 는 다음과 같이 계산된다.

$$z = \frac{|x_i - \bar{x}|}{\sigma} \quad (13)$$

그림 7 에서 51 번째 값이 다른 값들에 비해 상당히 큰 피크로 나타나게 된다. 이 값을 위의 식(13)을 적용하여 구하게 되면 σ (표준편차) = 10.1948 이고 $z=8.3588$ 로 상당히 큰 값을 가짐을 알 수 있다. 반면에 그림 8 에서와 같이 서로 다른 시퀀스들 간의 상관관계를 구한 후 최대값을 outlier 로 가정하고 z-score 를 구했을 경우에는 σ (표준편차) = 5.9564 이고 $z=2.4634$ 로 상당히 적은 값을 가지게 된다. 즉 워터마킹 알고리즘에서 위의 z-score 를 이용한 outlier 를 적용 함으로서 가변적인 상관관계 결과 및 피크를 가질 때 효율적으로 고정적인 threshold 를 결정 할 수 있다.

그림 9 에서는 1000 개의 임의의 오디오 샘플 블럭을 사용하여 3 장에서 제안된 방법으로 생성된 워터마크 시퀀스들 삽입 했을 때 correlation 의 z-score, 그리고 삽입되지 않았을 경우의 z-score 를 보여주고 있다. 그림 9 에서와 같이 삽입된 경우 z 값이 대략 10 부근에 나타나고 삽입되지 않은 경우에는 대략 4.5 정도에 나타나게 되고 threshold 는 대략 7.5 정도로 설정이 가능하다. 결국 일반적인 오디오 신호에서도 correlation 크기에 관계없이 outlier 를 이용하여 threshold 를 효율적으로 결정 할 수 있음을 실험적으로 보여 주었다.

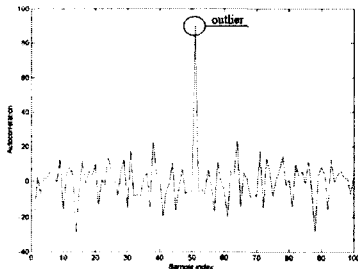


그림 7. 동일한 신호에서의 상관관계 및 outlier

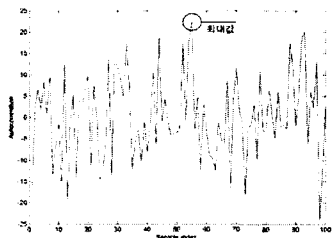


그림 8. 서로 다른 신호에서의 상관관계

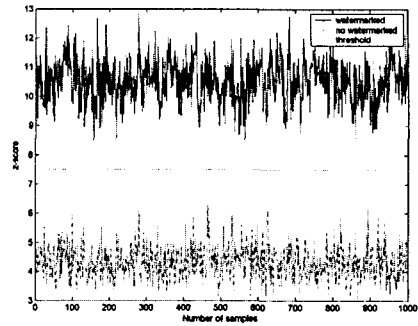


그림 9. z-score 에 따른 threshold 의 설정

V. 결론

본 논문에서는 PN 시퀀스를 이용한 spread spectrum 워터마킹의 단점 중의 하나인 심리음향모델 계산과 같은 복잡한 연산을 줄이고 정현파 패턴을 이용한 효율적인 오디오 워터마킹 기법을 제시했다. 심리음향모델을 적용하지 않고도 비가청잡음을 생성할 수 있는 장점들을 이용하여 시퀀스의 강도를 빠르게 조절하여 잡음을 줄일 수 있고 검출효율도 향상시킬 수 있는 방법을 제안했다. 그리고 outlier 를 이용하여 correlation 크기에 관계 없이 threshold 를 효율적으로 설정 할 수 있는 방법을 제안하였다. 또한 제안된 방법을 사용하여 비가청성을 높였을 경우에도 참고 문헌 [4]에서의 동일한 결과를 얻을 수 있었으며 더욱 구체화된 구현 방법을 제안했다.

참고문헌

- [1]. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," IBM System Journal, vol. 35, nos.3&4, pp. 313-336, 1996.
- [2]. I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Transactions on Image Processing, vol. 6, no. 12, pp. 1673-1687, 1997
- [3]. D. Gruhl, A. Lu, and W. Bender, "Echo Hiding," Information Hiding Workshop, University of Cambridge, U.K, pp. 295-315, 1996.
- [4]. Z. Liu, Y. Kobayashi, S. Sawato, and A. Inoue, "A Robust Audio Watermarking Method Using Sine Function Pattern Based on Pseudo-random Sequences," Pacific Rim Workshop on Digital Steganography, pp. 167-173, 2002.
- [5]. R. Petrovic, "Audio Signal Watermarking Based on Replica Modulation," International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service, vol. 1, pp. 227-234, 2001.
- [6]. V. Barnett, and T. Lewis, Outliers in Statistical Data, 3rd Ed, John Wiley & Sons, New York, 1994.