

# 웨이블릿 변환을 이용한 디지털 오디오 무손실 압축기법

박세일, 박세형, 신재호  
동국대학교 전자공학과

## A lossless Audio Coding method using wavelet transform

Se-Yil Park, Se-Hyoung Park, Jaeho Shin  
Department of Electronics, Dongguk University  
E-mail: wildred@dgu.edu

### 요 약 문

오늘날 오디오 포맷의 대부분은 마스킹 효과를 이용한 손실 압축 기법(lossy coding)이다. 그러나 고음질을 요구하는 매니아가 있고 DVD등 차세대 멀티미디어 포맷으로 무손실 오디오 압축기법이 요구되고 있고 이를 위해 일반적인 무손실 데이터 압축방법인 PkZip 등으로 압축할 경우 압축비가 약 1.07:1 밖에 되지 않으므로 적절하지 못하여, 디지털 오디오 데이터를 위한 무손실 압축 알고리즘의 연구가 필요하게 되었다. 본 논문에서는 프레임링(Framing), 상호상관제거(intrachannel decorrelation), 엔트로피 코딩(Entropy coding)의 세가지 단계로 구성되는 일반적인 오디오 무손실 압축 기법을 효율적으로 압축하기 위해 wavelet transform을 적용한 알고리즘을 제안한다.

### I. 서론

디지털 오디오 압축은 손실 압축(lossy compression)과 무손실 압축(lossless compression)으로 나눌 수 있다. 오늘날 오디오 포맷의 대부분은 마스킹 효과를 이용한 심리 음향 모델 등을 적용한 MPEG-I의 Layer 1, 2, 3, MPEG-II의 AAC, AC-3(Dolby. 社)등 손실 압축 기법(lossy coding)이다. 손실 압축 기법은 12:1의 우수한 압축율과 CD에 가까운 음질 등의 장점을 갖고 있어 많은 연구가 이루어진 반면 디지털 오디오 데이터를 위

한 무손실 압축 기법은 손실 압축 기법들에 비해 상대적으로 많은 연구가 이루어지지 못했다. 그러나 음질 저하가 없는 고음질 오디오 시스템을 원하는 매니아를 위해서 DVD등 차세대 멀티미디어 포맷으로 무손실 압축기법이 요구되고 있고 이를 위한 일반 무손실 데이터 압축 기법에 의한 압축율은 저조 하기 때문에 오디오 신호의 특성을 이용한 오디오 무손실 압축 기법이 연구가 필요하게 되었다. 본 논문에서는 샘플링 주파수 등 각종 파라미터들을 포함하는 프레임(framing)부분, 중복성(redundancy)을 제거하는 상호상관제거(Intrachannel decorrelation)부분, 엔트로피 코딩(Entropy coding)부분의 세가지 단계로 구성되는 일반적인 오디오 무손실 압축 기법을 효율적으로 압축하기 위해 입력신호를 웨이블릿 변환(wavelet transform)으로 대역별로 분리하면 decorrelation을 효율적으로 처리할 수 있어 압축율을 향상시킬 수 있다. 이 변환 실수 신호를 정수로 양자화 하여 intra-channel decorrelation 과정을 거친 후 마지막으로 엔트로피 코딩 과정에서 Rice coding을 이용하여 부호화 하는 알고리즘을 제안한다.



그림 1. 일반적인 무손실 압축 알고리즘 블록도

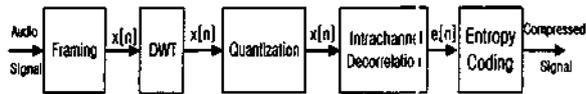


그림 2. 제안된 무손실 압축 부호화 알고리즘 블록도

## II. 무손실 압축 기법 제안

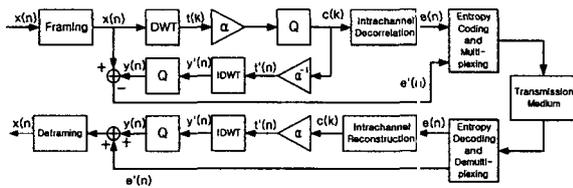


그림 3. 전체적인 encoding, decoding 블록도

### 1. 프레임링(Framing)

오디오 프레임은 프레임 헤더가 제일 앞에 붙고 그 헤더가 반복되어 이어진다. 헤더는 싱크 코드와 함께 시작하고 스트림을 플레이 하는 디코더를 위해 필요한 최소 파라미터 정보를 담고 있다. sample rate 와 bit per sample 등의 정보가 있다.

### 2. 이산 웨이블릿 변환과 양자화

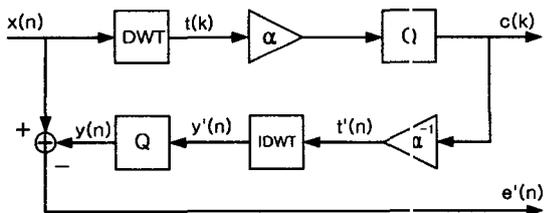


그림 4. 웨이블릿 변환과 양자화 과정

본 논문에서는 웨이블릿 변환 후 변환된 계수를 예측 모델(prediction model)을 적용하는 과정을 적용하게 되는데, 이러한 과정은[2]에서와 같이 (그림 3)과 같은 구조를 가지게 된다. 이때, 변환기를 사용하게 되면, 양자화시 발생하는 에러를 따로 부호화 하여야 한다. 입력 신호를 변환한 후 이 값을 스케일링을 하게 된다. 다음 이를 양자화하게 되는데, 양자화를 하게 될 경우 발생하는 에러에 대한 고려가 필요하다. 이에 대해서는 다음절에서 자세히 설명한다. 이산 웨이블릿변환은 웨이블릿 필터(wavelet filters)를 이용하여 입력신호를 필터링하고, 이를 다운 샘플링하는 과정으로 이루어져 있다.

웨이블릿변환의 장점은 여러 가지 있으나, 본 논문에서는 에너지 집중(Energy Compaction)현상을 이용하였다. DFT는 신호의 전체 에너지가 전체에 분산되는 반면, DCT와 DWT는 신호의 에너지가 저주파에 집중되는 특성을 가지고 있다. 이러한 특성은 신호의 압축과정에서 유용하게 사용될 수 있다. 이는 웨이블릿 필터의 주파수 응답을 보면 알 수 있다. 그러나 웨이블릿 계수가 길어지면 좋은 분해능을 갖는 반면 연산량이 늘어나는 단점을 가지게 된다. 따라서 적절한 선택이 중요하다. DWT를 수행하고 난 결과는 일종의 각 대역필터로 필터링 된 후, 이를 다운 샘플링한 과정이 된다. 본 논문에서는 정수형 입력신호를 웨이블릿 변환한 후, 웨이블릿 변화계수를 Prediction을 수행하였다. 이는 각각의 서브밴드마다  $\sigma_B^2$ 의 분산이 달라지기 때문에 양자화 Prediction의 결과가 다르게 나타날 수 있다. 저주파의 경우는 분산은 크지만, 신호의 변화는 빠르지 않은 반면, 고주파의 경우는 분산은 작지만, 신호의 변화가 빠르게 된다. 따라서 각각의 특성에 따라 적합한 모델을 사용할 경우 coding gain을 높일 수 있다.

본 논문에서는 무손실 오디오 코딩의 변환 부호화(Transform coding)와 reconstruction을 위한 적합한 양자화 방법[4]에서 제시한 구조에 DWT으로 8개의 대역으로 나누고 스텝 사이즈  $\Delta=1$  로 양자화 하고  $\alpha=1$  로 scaling하였다.

예측 모델에서 대역별로 상관제거(decorrelation)하기 위해서는 이산 웨이블릿 변환된 실수 값을 정수 값  $c(k)$ 으로 양자화 해야 된다. 양자화 에러값  $e'(n)$ 은 엔트로피 코드를 거쳐 부호화 되고 이 값은 디코딩 할 때 더함으로서 보상되기 때문에 손실되지 않는다.

### 4. 예측 모델(Prediction Model)

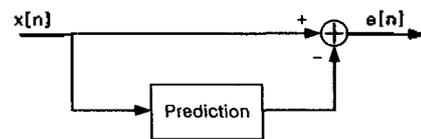


그림 5. 상호상관제거 block을 위한 예측 모델

웨이블릿 변환과 양자화 과정에서 출력된 정수값  $c(k)$ 가 상호상관제거 입력 값으로 들어온다. 상호상관제거는 각 프레임 내에 샘플들의 중복성(redundancy)를 제거하는 과정을 (1)식과 (그림 6)으로 나타낼 수 있다. 웨이블릿 변환으로 8개의 대역으로 나뉘어 각각 처리된다.

상호상관제거의 에러  $e[n]$ 은  $x[n]$ 보다 작다. 에러값

을 디지털로 표현하게 되면 더 작은 비트로 표현할 수 있게 된다. 이는 식(1)과 같이 표현 될 수 있다.

$$e[n] = x[n] - Q\left\{\sum_{k=1}^M \hat{a}_k x[n-k] - \sum_{k=1}^M \hat{b}_k e[n-k]\right\}, \quad (1)$$

$Q(\cdot)$ 는  $x[n]$ 의 워드 길이(word length)의 양자화를 의미 한다. (그림 6)에서  $\hat{A}(z)$  과  $\hat{B}(z)$ 는 아래의 feedforward 와 feedback으로 나타낸 z-transform 다항식이다.

$$\sum_{k=1}^M \hat{a}_k x[n-k] ; \text{feedforward.}$$

$$\sum_{k=1}^M \hat{b}_k e[n-k] ; \text{feedback.}$$

$\hat{B}(z) = 0$  이라면 피드백(feedback)이 생략되고 예측 에러 필터(prediction error filter)는 FIR 구조를 갖을 수 있다. 따라서  $\hat{A}(z)$ 는 FIR 구조이다. 만약  $\hat{B}(z) \neq 0$ 라면 예측기(predictor)는 IIR 구조가 된다.

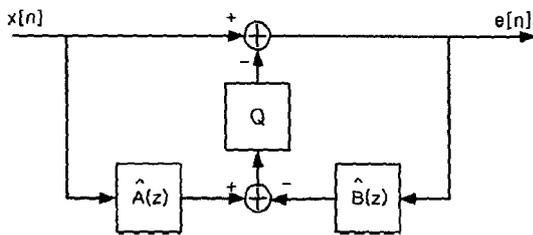


그림 6. 일반적인 prediction 구조

reconstruction은 (2)식과 (그림 7)와 같이 나타낼 수 있다.

$$x[n] = e[n] + Q\left\{\sum_{k=1}^M \hat{a}_k x[n-k] - \sum_{k=1}^M \hat{b}_k e[n-k]\right\}, \quad (2)$$

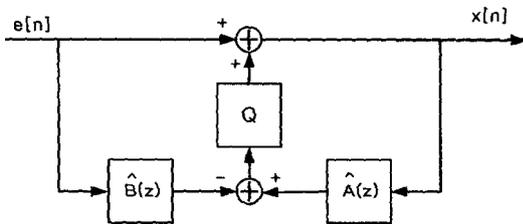


그림 7. 일반적인 reconstruction 구조

예측 필터(prediction filter)의 계수들  $\hat{A}(z)$ 는 mean-squared prediction error값을 최소화 하도록 결정된다. 양자화기(quantizer)가 없고  $\hat{B}(z) = 0$ 인 FIR

predictor라면 예측 계수들은 간단한 선형 방정식들[6] 계산으로 얻어 낼 수 있다. reconstruction에도 같은 계수들이 사용된다.

예측 에러 신호(prediction error signal) 계산에 있어 단순화가 중요하다. (앞에서 나타낸 알고리즘은 예측 계수까지 계산 할 때 전체적인 시스템의 계산량이 많아진다). 따라서 연산은 정수 계수들만을 사용한 아주 간단한 FIR 적용 예측 방법(adaptive prediction method)를 이용한다.

Shorten[3]에서 제안한 근사화 방법은 다음과 같다. 정수 계수들을 갖는 네 개의 간단한 FIR predictor들이 있다.(3)식

$$\begin{cases} \hat{x}_0[n] = 0 \\ \hat{x}_1[n] = x[n-1] \\ \hat{x}_2[n] = 2x[n-1] - x[n-2] \\ \hat{x}_3[n] = 3x[n-1] - 3x[n-2] + x[n-3]. \end{cases} \quad (3)$$

다항식은  $e_p[n] = x[n] - \hat{x}_p[n]$ 과 같이 근사화 된다.

$$\begin{cases} e_0[n] = x[n] \\ e_1[n] = e_0[n] - e_0[n-1] \\ e_2[n] = e_1[n] - e_1[n-1] \\ e_3[n] = e_2[n] - e_2[n-1]. \end{cases} \quad (4)$$

각 프레임 샘플마다 네 개의 예측 잉여값들(prediction residuals)  $e_0[n], e_1[n], e_2[n], e_3[n]$ 이 계산된다.

## 5. Entropy Coding

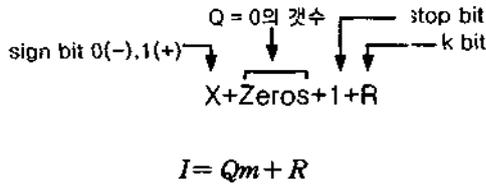
Entropy coding은 예측 에러값(Prediction error)인  $e[n]$ 에서 중복성(redundancy)를 제거하는 것이다.  $e[n]$ 값들은 저장매체 또는 전송매체 형태로 bit stream으로 변환되어야 한다. 라이스 코딩은 prediction error값을 부호화 하는데 가장 간단하고 효과적인 방법이다. 본 논문은 [2]에서 제시한 방법으로 골롬코드(Golombcode)의 일종인라이스 코딩(Rice coding)을 이용한다.

대부분의 무손실 오디오 코더(lossless audio coder)들은 라이스 코딩을 이용한다. (Lossless Transform Audio Compression(LTAC)[4], Shorten[3], etc). 여기서 는 골롬 코드의 파라미터  $m$ , 라이스 코드는 프레임 단위로 상수 값으로 파라미터  $k$ 을 정의 할 수 있다.

$$k = \log_2(\log_2(2)E(|e[n]|)) \approx \log_2 E(|e[n]|), \quad (3)$$

$E()$  : 수학적 기대치(즉, 평균값),

$e[n]$ : prediction error의 절대값의 평균값.



$I$ : 입력값 ( $e[n]$ ).

$Q$ : 단일 값(몫).

$R$ : 나머지.

$m$ : Golomb 파라미터. ( $2^k$ )

Table 1.  $e[n]$ 을 Rice coding한 예.

입력값	파라미터	Rice code 값				
		$m$ ( $2^k$ )	sign bit	$Q$	stop bit	$R$
10	2	4	1	00	1	10
-17	2	4	0	0000	1	01
28	2	4	1	0000000	1	00
19	2	4	1	000000	1	01

## VII. 비교 평가 및 결과

본 논문에서는 제안한 무손실압축 부호화기의 성능을 평가하기 위하여 SQAM[7]에 제고되는 데이터를 이용하였다. 각각의 파일은 약20에서 30초 사이의 44.1kHz의 샘플링 주파수를 가지고, 16비트의 해상도를 갖는 선형양자화된 모노포닉 신호음악 데이터이다.

본 논문은 [3]에서 의해서 제안된 기법과 이외에 많은 기법 중 LTAC와의 데이터를 비교하여 보았다.

## VIII. 결론

LTAC 방법의 압축율과 DWT를 이용한 무손실 압축율을 비교하였다. (Table 2) LTAC에 비교할 때 약 1.74배 가량 성능 차이가 있음을 확인 하였다. 본 논문에서는 DWT를 이용하여 대역을 나누는 Shorten[3]의 방법을 그대로 수행 하였으나 DWT에 적합한 예측 모델(Prediction Model)에 대한 연구가 더 필요하다.

Table 2. LTAC와의 compress ratio 비교

Track	Index	Time	LTAC의 compress ratio	DWT 방법의 compress ratio
47	01	0:30	2.485	1.446
07	01	0:35	4.096	2.061
35	01	0:35	3.157	1.515
35	02	0:24	2.551	1.191
40	01	0:23	2.265	1.235
23	02	0:31	3.076	1.627
48	01	0:28	2.239	1.509
44	01	0:28	2.517	1.560
49	01	0:23	2.090	1.316
51	01	0:21	1.887	1.245
53	01	0:21	2.100	1.319
50	01	0:22	2.124	1.351
52	01	0:24	2.135	1.336
54	01	0:21	2.143	1.320
21	02	0:23	3.087	1.643
10	02	0:37	2.720	1.664
평균값			2.542	1.459

## 참고문헌

- [1] M. Hans, R.W. Schafer "Lossless compression of digital audio" IEEE Signal Processing Magazine, Volume: 18 Issue: 4, July 2001 Page(s): 21 -32.
- [2] Peter Fenwick "Punctured Elias Codes for variable-length coding of the integers" Tech. Rep 137 ISSN 1173-3500 December 5, 1996 Department of Computer Science, The University of Auckland, New Zealand.
- [3] T. Robinson, "SHORTEN: Simple lossless and near-lossless waveform compression," Cambridge Univ. Eng. Dept., Cambridge, UK, Tech. Rep
- [4] M. Purat, T. Liebchen, and P. Noll, "Lossless transform coding of audio signals," in Proc. 102nd AES Conv., Munich, Germany, 1997, preprint, LTAC (Lossless Transform Audio Compression)  
<http://www-ft.ee.tu-berlin.de/~liebchen/ltac.html>
- [5] N. Jayant and P.Noll, Digital Coding of Waveforms, Principles and Applications to Speech and Video. Englewood Cliffs, NJ: Prentice-Hall, 1984
- [6] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [7] SQAM(Sound Quality Assessment Material), CD # 422 204-2. European Broadcasting Union, Geneva, 1988.