

# 적응 웨이블릿 패킷을 이용한 오디오 부호화기와 심리음향 모델링

김준성\*, 이대중\*, 강현철\*\*, 변윤식\*

인천대학교 전자공학과\* 인천대학교 정보통신공학과\*\*

## Audio Coder Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling

Junsung Kim\*, Dae-Jung Lee\*, Hyunchul Kang\*\*, Youn-Shik Byun\*

Dept.of Electronic Eng., Univ. of Incheon\*

Dept.of Info. & Telecom. Eng.,Univ. of Incheon\*\*

E-mail : byun@lion.inchon.ac.kr

### Abstract

In this paper, a new variable wavelet packet decomposition audio coder, based on the time varying characteristic of the audio signals, is proposed and presents a technique to incorporate psychoacoustic models into an adaptive wavelet packet scheme. The proposed filterbank improves the defect of the polyphase filterbank that could not properly represent the critical band and the defect of QMF-tree filter that need high complexity to implement. The filterbank consists of varying number of subband from 4 to 26 bands and use Daubechies 6-order wavelet. The codec yields excellent quality at total bit rates of about 128kbps for monophonic CD-quality signals with an sampling frequency of 44.1kHz and reduces complexity by 19% for various bit-rates and sources with encoding and decoding process.

분석 필터뱅크를 사용하여 계산량과 음질의 향상을 동시에 달성할 수 없는 몇 개의 조건을 최적화 하였다. 가장 일반적으로 알려진 ISO/MPEG 오디오의 경우 Layer II에서는 Subband coding에 32개의 균일한 대역으로 나누는 Polyphase filter를 사용하여, QMF(Quadrature Mirror Filter)banks를 낮은 연산량으로 설계할 수 있게 하였다. 그러나 이러한 구조는 신호의 시각적 중복성을 제거하기 위해 사람의 청각특성을 모델링한 심리음향모델의 입력대역을 효과적으로 나타내지 못하며, 균일한 주파수 해상도는 갑작스런 오디오 신호에 대한 대처 능력이 떨어지는 반면, 정상신호에 대해서는 필요 없이 많은 대역분할로 처리전력과, 연산량의 낭비를 가져온다. 이에 대처하여 제안된 필터뱅크의 구조는 입력신호의 시변특성에 근거하여 필터뱅크의 분석깊이(depth)를 가변 시켜 위와 같은 단점들을 보완하였다.

### I. 서론

본 논문에서는 입력신호의 시변특성의 하나인 영교차기준(zero-crossing rate)에 기인하여 오디오 신호의 갑작스런 변화에 대해서는 최대 26대역의 분석 필터뱅크를 사용하고, 준-안정신호(quasi static signals)에 대해서는 최소 4대역의

### II. 웨이블릿과 필터뱅크

#### 2.1 QMF filter 구조의 웨이블릿 변환

웨이블릿 확장과 변환에 관한 기본 식은 식(1)과 같다.

$$b(j, k) = \int_{-\infty}^{\infty} f(t) \sqrt{2^j} w(2^j t - k) dt$$
$$f(t) = \sum_{j,k} \sqrt{2^j} b(j, k) w(2^j t - k) \quad \text{식(1)}$$

여기서,  $w(t)$ 는 웨이블릿 함수이다. 웨이블릿  $w(t)$ 는 스케일 인자 2에 의해서 스케일 되고,  $k$ 에 의해서 이동(shift)된다. 즉, 고주파 신호에 대해서는 짧은 웨이블릿을 사용하고 저주파에서는 긴 웨이블릿을 사용한다. 웨이블릿 변환시 각 스케일은 이전 스케일의 1/2이 되며, 입력신호가 저역통과 필터와 고역통과필터를 통과한 뒤 다운샘플링(down sampling)이 이루어진다. 이러한 과정을 반복하여 다해상도를 가지는 필터뱅크를 설계할 수 있다[1,2,3,4]. 이러한 웨이블릿 변환된 신호는 합성 필터뱅크를 통과하여 완전복원(Perfect Reconstruction)이 이루어져야 한다[5]. 분석필터와 합성필터의 관계는 다음과 같다.

$$h(l) = (-1)^l g(p-1-l)$$

$$\bar{g}(l) = g(p-1-l) \quad \text{식(2)}$$

여기서  $g(l)$ ,  $h(l)$ 은 각각 저역통과, 고역통과 필터이다. Daubechies 필터는 다음의 완전복원 조건인 Power Complementary 조건을 만족한다[4].

$$|H(e^{j\omega})|^2 + |G(e^{j\omega})|^2 = c \quad \text{식(3)}$$

웨이블릿은 길이가 6인 Daubechies 웨이블릿을 이용하였으며, 비균일한 웨이블릿 필터뱅크는 인간의 청각특성을 나타내는 임계대역과 유사하게 웨이블릿 패킷변환을 이용하여 26개의 대역으로 나누었다[9]. 그림1에 설계된 웨이블릿 패킷 필터뱅크를 나타내고 있다.

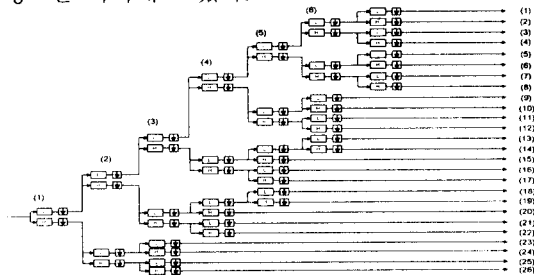


Fig 1. Wavelet packet filter bank

### 3.1 가변 대역 필터뱅크의 구조

제안된 부호화기의 구조를 그림2에 나타내었다. 프레임의 크기와 심리음향모델은 MPEG layer-2와 같게 각각 1152개와 심리음향모델-2를 사용하였다. 입력신호에 대한 분석필터뱅크의 길이는 영교차율에 따라 정해지고, zcr flag란 정보에 의해 양자화기 블록크기가 결정되며, 이에 따른 비트 할당과 스케일 인자의 크기도 가변 된다.

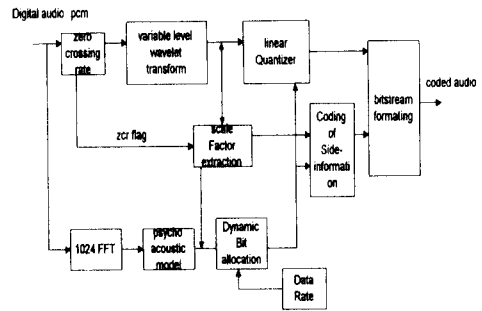


Fig 2. The structure of encoder

### 3.2 영교차율(Zero Crossing Rate)

일반적으로 비정상신호는 높은 주파수를 포함한다는 특성을 이용하였다. 입력되는 오디오 신호의 영교차수를 카운트(count)하여 그 수의 많고 적음에 따라 스위칭 하는 기준이다. 정확한 스위칭 판별을 위하여 입력되는 신호의 저주파수 부분의 신호를 높은 주파수 성분을 가지고 카운트하였으며, 식(4)에 나타나 있다.

$$ZCR = \sum_{i=0}^{N-1} |sgn(x(i)) - sgn(x(i+1))| * (1/2)$$

(  $sgn(x)=1(x>0)$ ,  $sgn(x)=-1(x<0)$  )    식(4)

그림 1의 26 부대역 분석필터뱅크를 1152개의 한 프레임에 준하여 영교차율이 120이상인 구간에서는 전체 26대역의 분석필터뱅크를 사용하고 그 이하는 실험에 의해 표-1과 같이 정의하였다.

표1. 영교차율에 따른 분석필터 बैं크

영교차율	분석 level	분석 대역수	Band당 샘플수
0 ~ 30	2	4	288 × 4
30 ~ 60	3	8	144 × 8
60 ~ 80	4	12	144 × 4, 72 × 8
80 ~ 100	5	17	144 × 4, 72 × 3, 36 × 10
100 ~ 120	6	22	144 × 4, 72 × 3, 36 × 5, 18 × 10

### 3.3 비트 할당 양자화

본 연구에서 사용된 양자화기는 변환된 신호의 계수들간에 상관성이 높지 않은 점을 고려하여, 스칼라 양자화기를 사용하였으며, 26개의 대역에 대한 비트 할당 정보를 이용하여, 가변대역의 비트 할당을 적용시켰으며, 그림4에 나타나 있다[6][7]. 즉, 신호의 비정상 정도(Non stationary measure)를 영교차율에 의해 미리 구한 후, 이

러한 정보를 바탕으로 신호처리에 요구되어지는 최소 주파수 해상도로 최대의 부대역 크기를 만든다.

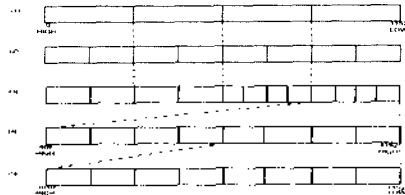


Fig 4. 26, 22, 17, 8, 4 Subband structure

### 3.4 심리음향 모델

심리음향을 사용하는 목적은 부호화기가 입력 오디오 신호를 분석하여 양자화 잡음을 줄이고, 전송율을 낮추는데 있다. 즉, NMR(noise to mask ratio)이 최소가 되는 비트할당 값을 구하는 것이며 식 (6)과 같이 유도된다[10].

$$L = 20 \log(p/p_0) [dB] \quad \text{식(6)}$$

$$LT_G(i) = 10 \log_{10} [ 10^{LT_m(i)/10} + \sum_{j=1}^m 10^{LT_{m(j,i)}/10} + \sum_{j=1}^n 10^{LT_{m(i,j)}/10} ] dB$$

$$SMR_{sb}(n) = L_{sb}(n) - LT_{min}(n) dB$$

여기서, L은 음압레벨, P0는 기준값, P는 오디오 신호의 음압이다. 두 번째 수식에 우항(right term)의 첫 항은 최소가청한계(in quiet), 두 번째 항은 순음(tonic)에 의한 레벨 그리고 마지막 항은 비순음(nontonal)의 음압 레벨이다. 그리고 i는 임계대역의 인덱스이다.

## IV. 모의실험 결과

본 연구에서는 신호의 특성에 따라 필터뱅크의 구조를 달리하는 부호화/복호화기와, MPEG 오디오 부호화/복호화기의 MOS(mean opinion score) 테스트를 통한 음질평가와 연산량을 비교하였다.

### 4.1 오디오 코덱의 연산량

가변대역에 따른 연산량 감소를 단적으로 보이기 위해 MPEG 32 부대역 부호화기와, 제안된 부호화기의 3 부대역 부호화기와의 연산량을 보인다.

#### 4.1.1 바드 할당

그림8의 (a)는 MPEG의 64kbps에서 50프레임에 대한 비트할당 과정이고, (b)는 한 프레임 3band

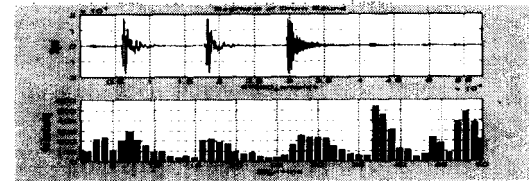


Fig 7. (a) source signal (b) zero crossing rate

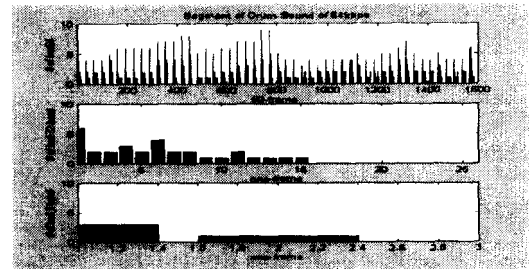


Fig 8. MPEG bit allocation data (a),(b),(c)

대역분할시 한 프레임에 대한 비트할당 과정이다. 그림 8(c) band1, band2는 각각 그림8(b)의 band 1~8, band9~18을 포함하며, 각각 포함된 밴드의 비트할당의 평균을 이용한다. 이러한 주파수 평면 분석은 시간 평면에서 변화율이 작은 구간이나, 특유구간에서 음질저하에 영향을 미치지 않으며, 큰 부대역의 사용은 작은 부대역보다 요구되어지는 부가정보(side-information)의 오버헤드(over-head)가 작으므로 부호화과정의 주 정보(main informaion)에 할당 bit-rate을 증가시킬 수 있어, 변환 부호화(transform coding)의 성능을 향상 시킨다[11]. 다음은 한 예로써, drum신호의 26부대역 필터뱅크 사용시 요구되어지는 연산량과 4부대역 필터뱅크 사용시의 연산량을 비교하여 보았다. 이는 1채널 중에서도 한 프레임에 해당하는(1152sample)것이며, 연산량을 계산하는 단위는 operation(시스템 클럭의 한 사이클을 필요로 하는 연산량)을 사용하고, 곱셈기 1개와 덧셈기 1개를 각각 지원하는 시스템을 기준으로 연산량을 계산하였다. 넓은 부대역을 사용함으로써 얻어지는 연산량 감소 이득은 양자화기와 필터뱅크의 연산에서 나타난다.

#### 4.1.2 양자화기

양자화기에서 입력 샘플당 요구되어지는 비교연산의 계산량은 다음과 같으며, 비교연산 역시 1사이클이 소요된다고 가정한다.

$$C = (\text{한band당샘플수} \times (\text{band당의 비트할당}))$$

한편 그림8(c)의 경우는 다음과 같다.

$$C = (\text{샘플수} \times 8\text{band} \times 8\text{band 할당비트평균})$$

### 4.1.3 필터뱅크의 분석 합성과정

덧셈, 곱셈기의 연산량 계산방법은 다음과 같다.

$$\begin{aligned} \text{곱셈기[ops]} &= C(\text{웨이블릿 계수의 차수}) \\ &\quad \times N(\text{한 프레임의 샘플개수}) \\ &\quad \times L(\text{분석레벨}) \times 2(\text{분석·합성과정}) \\ \text{덧셈기[ops]} &= C \times N \times L \times 2 \end{aligned}$$

표-2 분석레벨에 따른 연산량

분석 레벨	분할 대역	곱셈기 [ops]	덧셈기 [ops]	Total [ops]	연산량
7	26 대역	96,768	96,768	193,536	100%
6	22 대역	48,384	48,384	96,768	85.7%
5	17 대역	24,192	24,192	48,384	71%
4	12 대역	12,096	12,096	24,192	57.1%
3	8 대역	6,048	6,048	12,096	42.9%
2	4 대역	3,024	3,024	6,048	28.5%

### 4.2 오디오 신호의 음질평가 결과

표-3 MOS 테스트결과

오디오신호	128 kbps	64 kbps	32 kbps
pop song	4.32	3.21	2.83
drum	4.12	3.3	2.68
piano	4.23	3.54	3.06
violin	4.36	3.62	3.1
평균	4.25	3.41	2.91

### 4.3 코덱 전체 연산량 비교결과

MPEG-1 오디오와 제안된 시스템을 MPEG의 384kbps일 때 소요시간을 100(%)으로 하여 상대적인 비교를 하였다.

표-4 여러 음원의 평균 연산량 비교결과

전송률	MPEG	제안된 시스템	감소율(%)
384	100.00	63.33	36.77
128	38.33	23.33	15.00
64	35.00	21.66	13.34
32	33.33	20.00	13.33
평균			19.61

## V. 결론

본 연구에서는 넓은 동적 영역을 가지는 오디오 신호의 시변특성을 분석하여, CD수준의 음질을 유지하면서, 필요 연산량을 최소로 하는 오디오 코덱을 설계하였다. 제안된 시스템은 QMF 필터뱅크의 구조를 가지며, 입력신호의 시변특성에 기인하여, 서브밴드대역의 수, 즉, 블록의 크기를 달리하였다. 128kbps에선 CD수준의 음질을 유지하

며, 연산량에 있어서는 MPEG 오디오 코덱에 평균 19% 감소시키는 효과를 보았으며, 신호의 종류에 따라 많은 차이를 보였다. 이는 MPEG-2나 AC-3같은 다채널을 전송함에 따라 요구되어지는 많은 연산량과 전송률을 각 채널의 신호특성에 기인하여, 독립적으로 최소의 연산량과 전송률을 요구하는 코덱(codec)의 실제로, 특별한 프로세서 없이 소프트웨어로 실시간 멀티채널 코덱(codec)의 구현 가능성을 제시한다. 또한 가변대역에 적용되는 비트 할당 알고리즘의 최적화로 보다 나은 음질과 연산량의 감소가 기대 되어진다.

## 참고 문헌

- [1]. G.Strang and T.Nguyens, Wavelet and Filter Banks, Wellesley Cambridge Press, Wellesley, MA, 1996.
- [2]. P.P. Vaidyanathan, Multirate System and Filter Bank, Prentice Hall, 1993.
- [3]. S. Shien, Guide to MPEG-1 Audio Standard, IEEE Transactions on Broadcasting, Vol., 40, No. 4, DEC. 1994.
- [4]. K.N Hamdy, M.Ali and A.H.Tewfik, Low Bit Rate High Quality Audio Coding With Combined Harmonic And Wavelet Representations, IEEE ICASSP, pp 1045 - 1048, 1996.
- [5]. C.Sidney Burrus Ramesh A. Gopinath Haitao Guo Wavelets and Wavelet transforms Prentice Hall 1998.
- [6]. Michel Misti, Y Misiti, G Oppenheim, J-MI Poggi, Wavelet Toolbox, The Math Works 1996.
- [7]. K. Sayood, Introduction to Data compression, Morgan Kaufmann Publishers, 1996.
- [8]. 배석모, 김도형, 정재호, 웨이브릿 변환을 이용한 오디오 코딩, 한국 음향학회지 제16권 4호, 1997.
- [9]. 신유철, 김준성, 상현철, 상병건, 변윤식, 웨이브릿변환과 MDCT 스위칭 기법을 이용한 오디오 부호화, 신호처리 학술대회 1997.
- [10]. Zwicker, E., Fastl, H., Psychoacoustics Facts and Models, Springer-Verlag, 1990
- [11]. N.S Javant et al., Digital Coding of Waveforms. Englewood Cliffs, NJ: Prentice Hall 1984.