

가변 대역분할 웨이블릿필터를 이용한 오디오 부호화기

Audio Coder Using Variable Subband Wavelet Filter

김 준 성*, 강 현 철**, 변 윤 식*

(Jun Sung Kim*, Hyun Chul Kang*, Youn Shik Byun*)

요 약

본 논문에서는 입력신호의 시변특성에 따라 분석 필터의 대역을 가변 시키는 필터뱅크의 구조를 제안한다. 제안된 필터뱅크는 일반적으로 32개의 균일한 대역으로 나누어 임계대역의 표현을 적절히 표현하지 못하는 Polyphase 필터의 단점을 극복하면서 시스템 설계에 높은 계산량을 요구하는 QMF-tree 필터의 단점을 보완한다. 본 연구에서 분할 대역은 4개에서 26개의 대역으로 가변하고, 웨이블릿 필터중 Daubechies 필터를 사용했다. 제안된 구조의 부호화기는 128kbps에서 MPEG-1 오디오와 비슷한 수준의 CD 음질을 유지하며, 연산량 비교결과는 PolyPhase filter를 이용한 MPEG보다 부호화, 복호화 과정을 합쳐 다양한 전송률과 음원에서 평균 19%의 감소를 얻었다.

ABSTRACT

In this paper, a new variable wavelet packet decomposition audio coder, based on the time varying characteristic of the audio signals, is proposed. 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 a sampling frequency of 44.1 kHz and reduces complexity by 19% for various bit-rates and sources with encoding and decoding process.

I. 서 론

본 논문에서는 입력신호의 시변특성의 하나인 영교차 기준(zero-crossing rate)에 기인하여 오디오 신호의 갑작스런 변화에 대해서는 최대 26대역의 분석 필터 뱅크를 사용하고, 준-안정신호(quasi static signals)에 대해서는 최소 4대역의 분석 필터뱅크를 사용하여 계산량과 음질의 향상을 동시에 달성할 수 없는 몇 개의 조건을 최적화 하였다.

가장 일반적으로 알려진 ISO/MPEG 오디오의 경우 Layer I,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 지각적 중복성을 제거하기 위해 사람의 청각특성을 모델링한 심리음향모델의 임계대역을 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 대처하

여 제안된 필터뱅크의 구조는 입력신호의 시변 특성에 근거하여 갑작스런 신호의 변화와 정상신호에 대해 필터뱅크의 분석 깊이(depth)를 가변시켜 위와 같은 단점들을 보완하였다. 이러한 부호화기의 구조는 MPEG-2나 AC-3 같은 멀티채널의 부호화에서 각채널의 신호특성에 기인하여 허용되는 잡음의 최대 한계를 수용하면서, 최소의 연산량을 요구하는 가장 효율적인 방법을 제공한다.

II. 웨이블릿과 필터뱅크

2.1 QMF-Tree filter 구조의 웨이블릿 변환

웨이브릿 확장과 변환에 관한 기본식은 식(1)과 같다.

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

$$f(t) = \sum_{j,k} \sqrt{2^j} b(j, k) w(2^j t - k)$$

여기서, $w(t)$ 는 웨이블릿 함수이다. 웨이블릿 $w(t)$ 는 스케일 인자 2에 의해서 스케일 되고 k에 의해서 이동(shift)된다. 즉, 고주파 신호에 대해서는 짧은 웨이블릿을 사용하고 저주파에서는 긴 웨이블릿을 사용한다. 웨이블릿 변환시 각 스케일은 이전의 스케일에서 2로 다운샘플

* 인천대학교 전자공학과

** 인천대학교 정보통신공학과

접수일자 : 1998년 5월 2일

링(down sampling)을 하기 때문에 이전 스케일의 1/2이 된다. 이 웨이블릿 변환은 입력신호가 저역통과필터와 고역통과필터를 통과한 뒤 다운샘플링이 이루어진다. 이러한 과정을 반복하여 다해상도를 가지는 필터뱅크를 설계할 수 있다[1, 2, 3, 4]. 그리고 이는 분석필터뱅크의 깊이(depth)가 변하더라도, 한 프레임의 입력 개수는 변함이 없이 유지된다는 뜻이다. 이러한 웨이블릿 변환된 신호는 합성 필터뱅크를 통과하여 완전복원(Perfect Reconstruction)이 이루어져야 한다[5]. 분석필터와 합성필터의 관계는 다음과 같다.

$$h(l) = (-1)^l g(p-1-l) \quad (2)$$

$$\overline{g(l)} = g(p-1-l)$$

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

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

웨이블릿은 길이가 6인 Daubechies 웨이블릿을 이용하였다. 본 연구에서 사용한 비균일한 웨이블릿 필터뱅크는 인간의 청각특성을 나타내는 임계대역과 유사한 웨이블릿 패킷변환을 이용하여 26개의 대역으로 나누었다[9]. 그림 1에 설계된 웨이블릿 패킷 필터뱅크를 나타내고 있다.

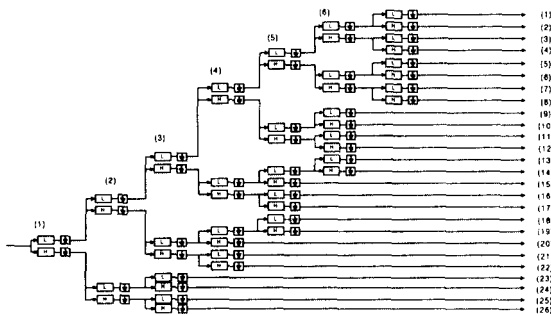


Fig. 1 Wavelet packet filter bank.

III. 제안된 가변대역 필터뱅크의 구조 및 양자화

3.1 가변 대역 필터뱅크의 구조

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

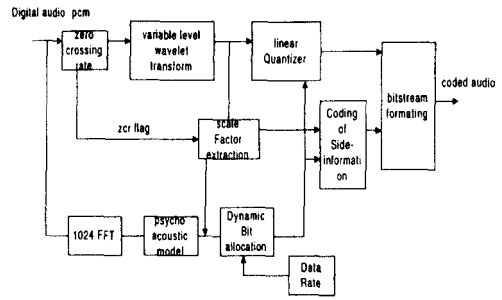


Fig. 2 The structure of encoder.

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. 즉, 시간축을 기준으로 교차되는 수를 카운트하여 그 값을 저장한다. 정확한 스위칭 판별을 위하여 입력되는 신호의 저주파 부분의 신호를 변 높은 주파수 성분만을 가지고 카운트하였으며, 식(4)에 나타나 있다.

$$ZCR = \sum_{i=0}^{N-1} |sgn(x(i)) - sgn(x(i+1))| * (1/2) \quad (4)$$

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

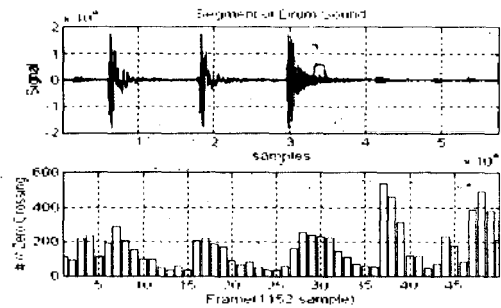


Fig. 3 The zero crossing rate of drum.

그림 1에서 보면 26 부대역 분석필터뱅크는 1152개의 한 프레임에 대해 표 1과 같은 샘플들이 위치하게 된다.

표 1. 한 프레임에 대한 서브밴드당 샘플수

band	sample num	band	sample num
1	9	14	18
2	9	15	36
3	9	16	36
4	9	17	36
5	9	18	36
6	9	19	36
7	9	20	72
8	9	21	72
9	18	22	72
10	18	23	144
11	18	24	144
12	18	25	144
13	18	26	144

이에 준하여 영교차율이 120이상인 구간에서는 전체 26대역의 분석필터뱅크를 사용하고 그 이하는 실험에 의해 표 2와 같이 정의하였다.

표 2. 영교차율에 따른 분석필터뱅크

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

양자화는 분석 필터의 출력을 하나의 블록으로 하고, 그 블록단위로 스케일링을 위한 최대값 산출, 다이내믹 비트 할당에 의한 양자화 레벨 결정, 그리고 스텝사이즈의 결정 순으로 이루어진다[6][7]. 본 연구에서 사용된 양자화기는 변환된 신호의 계수들간에 상관성이 높지 않은 점을 고려하여, 스칼라 양자화기를 사용하였으며, 26개의 대역에 대한 비트할당 정보를 이용하여, 다음과 같은 간단한 비트 조작으로 가변대역의 비트 할당을 적용시켰으며, 그림 4에 나타나 있다.

즉, 신호의 비정상 정도(Non stationary measure)를 영교차율에 의해 미리 구한 후, 이러한 정보를 바탕으로 신호처리에 요구되어지는 주파수 해상도를 고음질을 유지할 수 있을 만큼의 최소값을 갖게 하며, 이는 최대의

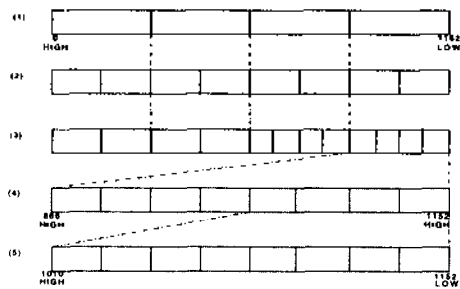


Fig. 4 26, 22, 17, 8, 4 Subband structure.

부대역 크기를 만들므로, 양자화에서 각 서브밴드당 필요로 하는 비트 할당, 스케일링 등의 과정을 신호의 정상성을 유지하는 한도에서 각각 재 사용한다는 것으로 볼 수 있다. 그러한 결과는 많은 계산량의 감소와 처리전력의 낭비를 막는 장점들이 있다.

그림 5는 한번의 QMF를 이용한 대역분할과 합성과정을 보여준다[8][9]. 이러한 과정이 대역분할 모듈마다 사용되

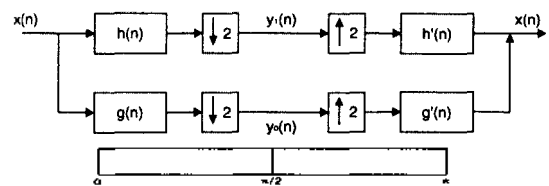


Fig. 5 Subband coding.

며, 또한 분석필터를 통과한 신호는 인간의 청취할 수 없는 최대의 감음허용, 즉 손실부호화(loss coding)를 이용하여 전송률을 줄이는데, 이러한 과정을 양자화기에서 각 대역마다의 원신호의 스펙트럼과 심리음향모델과 스펙트

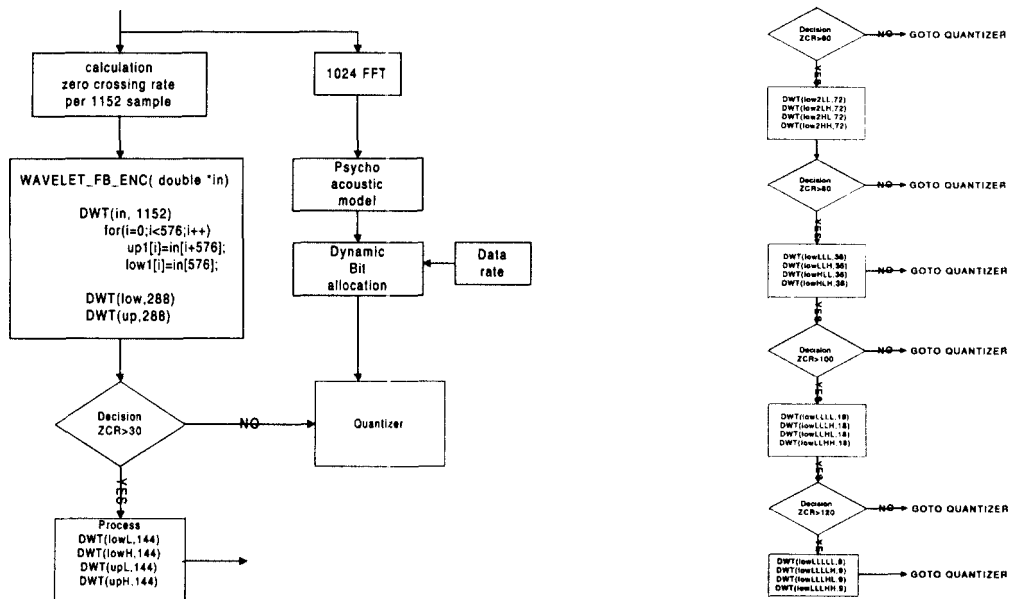


Fig. 6 Subband coding control scheme.

럽 정보로부터 최적의 비트할당율을 제공받으며, 각각의 계수들은 양자화 잡음을 최소화하기 위해 스케일링된다. 이러한 과정은 오디오 코더의 전체 연산량의 반 이상을 차지한다.

제안된 부호화기는 이러한 26개의 비트할당 정보와 스케일링 정보를 가지고 있으며, 신호의 정상도를 여러 계산함으로써, 신호의 특성에 따라 목음구간이나 변화율이 작은 구간에서 작게 나누어진 부대역의 비트할당 정보들의 평균을 큰부대역의 비트할당 정보로 사용한다. 이는 요구되어지는 주파수 해상도와 시간축 해상도를 신호의 특성을 이용한 것으로, 고음질을 유지하며, 계산량이 낮은 코덱(codec)이 설계된다. 이러한 분석모듈 설계구성이 그림 6에 나타나 있다.

IV. 모의실험 결과

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

4.1 오디오 코덱의 연산량

본 연구에서 제안된 코덱(codec)은 CD수준의 음질을 유지하면서, 그에 요구되어지는 계산량을 최소로 갖는다. 이는 신호의 특성과 전송률에 따라 큰 차이를 보인다.

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

4.1.1 비트할당

그림 8의 (a)는 MPEG의 64kbps에서 50프레임에 대한 비트할당과정이고, (b)는 한 프레임 그리고 (C)는 제안된 가변 부호화기의 3 band 대역분할사 한 프레임에 대한 비트할당 과정이다.

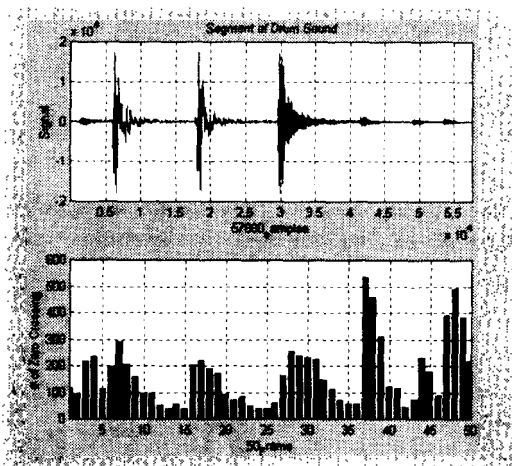


Fig. 7 (a)source signal, (b)zero crossing rate.

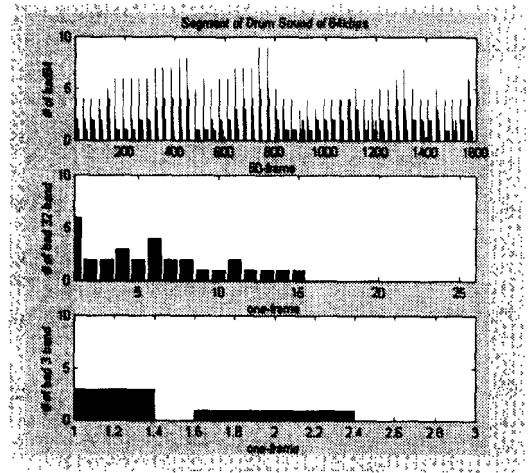


Fig. 8 MPEG bit allocation data, (a)(b)32band, (c)3band.

drum signal의 경우는 시간에 따라 갑작스런 변화가 일어나는 신호구간과 신호성분이 거의 없거나 설대에너지가 작아 들리지 않는 구간, 변화가 상대적으로 없는 구간등 다양하게 구성되어 있다(그림7(a)). 주파수 평면으로 변환된 신호에 대하여 심리음향모델에 기인한 비트할당은 시간평면에서의 정보가 잘 나타나 있지않다. 그림 8(c) band1, band2는 각각 그림 8(b)의 band1~8, band9~18을 포함하며, 각각 포함된 밴드의 비트할당의 평균을 이용한다. 이러한 주파수 평면 분석은 시간평면에서 변화율이 작은 구간이나, 목음구간에서는 음질저하에 영향을 미치지 않으며, 큰 부대역의 사용은 작은 부대역보다 요구되어지는 부가정보(side-information)의 오버헤드(over-head)가 작으므로 부호화과정의 주 정보(main information)에 할당 bit-rate을 증가시킬수 있어, 변환부호화(transform coding)의 성능을 향상시킨다 [11].

제안된 필터뱅크는 입력신호의 특성에 따라 계산량이 변화한다. 다음은 한 예로써, drum신호의 26부대역 필터뱅크 사용시 요구되어지는 연산량과 4부대역 필터뱅크 사용시의 연산량을 비교하여 보았다. 이는 1채널 중에서도 한 프레임에 해당하는(1152 sample) 것이며, 연산량을 계산하는 단위는 operation(시스템 클럭의 한 사이클을 필요로 하는 연산량)을 사용하고, 곱셈기 1개와 덧셈기 1개를 각각 지원하는 시스템을 기준으로 연산량을 계산하였다. 빠른 곱셈기가 1사이클만에 곱셈을 계산하므로 덧셈과 동일한 연산량을 가진다고 충분히 가정할 수 있다. 넓은 부대역을 사용함으로써 얻어지는 연산량 감소의 이득은 양자화기와 필터뱅크의 연산에서 나타난다.

4.1.2 양자화기

그림 8에서 (c)의 경우 band1은 (b)의 band1부터 band8 까지를 포함한다.

그림 8의 (b)의 band1에서, band8까지는 각각 $2^6=64$ level, $2^2=4$ level, $2^2=4$ level, $2^3=8$ level, $2^2=4$ level, $2^1=16$ level, $2^2=4$ level, $2^1=1$ level을 양자화기의 레벨로 가지며, 각각의 대역에는 36 sample (1152/32)을 가

지고 있다. 이때 요구되어지는 비교연산의 계산량은 다음과 같으며, 비교연산 역시 1사이클이 소요된다고 가정한다.

$$C = S' \times B'$$

$$= [36\text{샘플} \times (\text{band1의 비트할당} + \text{band2의 비트할당} + \dots + \text{band8의 비트할당})]$$

$$= [36 \times (64 + 4 + 4 + 8 + 4 + 16 + 4 + 1) = 3780[\text{ops}]]$$

한편 그림 8(c)의경우는 다음과 같다.

$$C = (36\text{샘플} \times 8\text{band} \times \text{할당비트평균}) = (36 \times 8 \times 8) = 2304[\text{ops}]$$

이는 전송률이 높을수록 감소량이 기하급수적으로 감소함을 또한 보여준다.

큰 부대역을 사용함으로써 음질이 저하되는 경우가 있는데, 이는 영교차율(0~600)이 120이하인 구간(목음구간)이나, 프레임 내에서 변화가 적은 구간)에 적용되므로 극복되어 질 수 있다.

4.1.3 필터뱅크의 분석 합성과정

양자화기에서 비교 연산량이 크게 감소하듯이 분석 및 합성과정에서는 곱셈과 덧셈 연산량이 감소한다. 덧셈기와 곱셈기의 연산량 계산방법은 다음과 같다.

$$\text{곱셈기}[\text{ops}] = C(\text{웨이블릿 계수의 차수}) \times N(\text{한 프레임의 샘플 개수}) \times L(\text{분석레벨}) \times 2(\text{분석} + \text{합성과정})$$

$$\text{덧셈기}[\text{ops}] = C(\text{웨이블릿 계수의 차수}) \times N(\text{한 프레임의 샘플 개수}) \times L(\text{분석레벨}) \times 2(\text{분석} + \text{합성과정})$$

표 3. 한 프레임에 대한 서브밴드당 샘플수

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

4.2 오디오 신호의 음질평가 결과

한 채널의 신호에 대해 다른 전송률에 따른 음질평가로 비전문 오디오 청취자 10명이 참가하였다.

표 3. MOS 테스트 결과

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 384kbps 일 때 소요시간을 100(%)으로 하여 상대적인 비교를 하였으며 이것은 부호화과정과 복호화 과정을 모두 포함하는 시간으로 전송률이 높을수록 감소율은 증가하였다.

표 4. 여러 음원의 평균 계산량 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. 결 론

본 연구에서는 넓은 동적 영역을 가지는 오디오신호의 시변특성을 분석하여, CD 수준의 음질을 유지하면서, 필요 계산량을 최소로 하는 오디오 코덱을 설계하였다. 제안된 시스템은 QMF 필터뱅크의 구조를 가지며, 입력신호의 시변특성에 기인하여, 서브밴드대역의 수, 즉, 블럭의 크기를 달리하였다. 128kbps에선 CD 수준의 음질을 유지하며, 계산량에 있어서는, MPEG 오디오 코덱에 평균 19% 감소시키는 효과를 보았으며, 신호의 종류에 따라 많은 차이를 보였다. 이는 MPEG-2나 AC-3같은 5채널을 전송함에 따라 요구되어지는 많은 계산량과 전송률을 각 채널의 신호특성에 기인하여, 독립적으로 최소의 계산량과 전송률을 요구하는 코덱(codec)의 설계로, 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.

참 고 문 헌

- G.Strang and T.Nguyens, "Wavelet and Filter Banks", *Wellesley-Cambridge Press*, Wellesley, MA, 1996.
- P.P. Vaidyanathan, "Multirate System and Filter Bank", *Prentice Hall*, 1993.
- S. Shlien, "Guide to MPEG-1 Audio Standard", *IEEE Transactions on Broadcasting*, Vol., 40, No. 4, DEC. 1994.
- 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.
- C. Sidney Burrus, Ramch A. Gopinath Haitao Guo "Wavelets and Wavelet transforms" *Prentice Hall* 1998.
- Michel Misti, Y Misiti, G Oppenheim, J-MI Poggi, "Wavelet Toolbox", *The Math Works* 1996.
- K. Sayood, "Introduction to Data compression", *Morgan Kaufmann Publishers*, 1996.
- 배석모, 김도형, 정재호, "웨이브렛 변환을 이용한 오디오 코덱", *한국음향학회지* 제16권4호, 1997.

9. 신유철, 김준성, 강현철, 장병진, 변윤식, "웨이브렛 변환과 MDCT의 스위칭 기법을 이용한 오디오 부호화". *신호처리 학술대회* pp.725-728 Sept.1997.
10. O. Riouland M. Vetterli, "Wavelets and Signal Processing", *IEEE SP Magazine*, pp.14-38, Oct.1991.
11. N.S Javant et al., "Digital Coding of Waveforms". *Englewood Cliffs, NJ: Prentice Hall* 1984.

▲김 준 성(Junsung Kim)



1997년 2월: 시립 인천대학교 전자공학과 졸업 (학사)
 1997년 3월~현재: 시립 인천대학교 전자공학과 (석사과정)
 ※주관심분야: 멀티미디어, 디지털 오디오 코딩

▲강 현 철(Hyunchul Kang)



현재: 시립 인천대학교 정보통신공학과 부교수
 한국음향학회지 제17권 1호 참조
 ※주관심분야: 영상처리, 패턴인식, 회상통신

▲변 윤 식(Youn-Shik Byun)



현재: 시립 인천대학교 전자공학과 부교수
 한국음향학회지 제17권 1호 참조
 ※주관심분야: 신호처리 및 그 응용, 디지털 통신