

논문 2004-41SP-6-29

MPEG 심리음향모델의 고속 구현을 위한 효율적 FFT 연산

(An Efficient Computation of FFT for MPEG/Audio Psycho-Acoustic Model)

송 건 호*, 이 근 섭**, 박 영 철***, 윤 대 희**

(Gun-Ho Song, Keun-Sup Lee, Young-Cheol Park, and Dae-Hee Youn)

요 약

본 논문은 MPEG 오디오 부호화 과정 중 심리음향모델(PAM: Psychoacoustic Model)에 사용되는 FFT를 효율적으로 계산할 수 있는 방법을 제안한다. MPEG 오디오 부호화 과정 중 심리음향 모델 연산은 많은 연산량을 차지하며, 부호화의 실시간 및 저 전력 구현을 위해서는 이의 최적화가 요구된다. 최적화를 위하여 제안하는 알고리즘은 심리음향모델에서 사용하는 입력 신호에 대한 1024-샘플 FFT 대신에 서브밴드과정에서 나온 각각의 32개의 밴드에 대하여 32-샘플 FFT를 함으로써 1024-샘플 FFT의 효과를 얻는데 있다. 이때 서브밴드과정을 거친 신호에 대한 FFT에는 에일리어징이 발생하며 이를 줄이기 위하여 버터플라이 구조의 에일리어징 제거블록을 제안한다. 제안한 알고리즘으로 심리음향모델을 구현할 경우 SMR이 약 1dB이내의 오차를 가지며, 기존 알고리즘에 비해 절반의 연산량으로 심리음향 모델을 위한 FFT 스펙트럼을 얻을 수 있다.

Abstract

In this paper, an efficient algorithm for computing in the MPEG/audio Layer III (MP3) encoder is proposed. The proposed algorithm performs a full-band 1024-point FFT by computing 32-point FFT's of 32 subband outputs. To reduce the aliasing caused by the analysis filter bank, an aliasing cancellation butterfly is developed. A major benefit of the proposed algorithm is the computational saving. By using the proposed algorithm, it is possible to save 40~50% of computations for FFT, which results in about 20% reduction of the PAM-2 complexity.

Keywords: 심리음향모델, 오디오 부호화, PAM, MP3

I. 서 론

MPEG/AUDIO layer-III(이하MP3)^[1]는 휴대용 오디오 재생기, 인터넷 오디오 등에 널리 사용되고 있는 고품질 오디오 부호화 기술이다. MP3는 인간의 청각 특성을 고려한 심리음향모델을 사용하여 지각되지 않는 신호를 제거하고, 주파수 영역에서의 신호처리를 통하여 약 12:1의 압축 후에도 원음과 거의 구별되지 않는 고품질의 오디오 신호를 제공한다^{[2][3]}.

이와 같은 장점으로 인해 최근 많은 휴대용 MP3 재생기가 상품화되고 있으나, 실시간으로 부호화가 가능한 휴대용 MP3 부호화기는 상용화된 예가 드물다. 이는 저전력 DSP를 이용하여 휴대용 MP3 부호화기를 실시간 구현하기에는 연산량 부담이 크기 때문이다.

일반적인 MP3 부호화 과정을 살펴보면 그림 1과 같다.

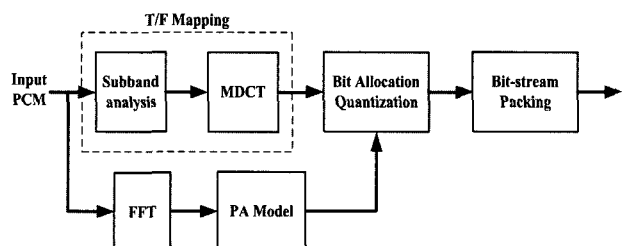


그림 1. 일반적인 MP3 부호화 과정

Fig 1. General MP3 Encoding Block Diagram.

* 정회원, LG 전자기술원 모바일 멀티미디어 연구소 (Mobile Multimedia Lab. LG Electronics Inc.)

** 정회원, 연세대학교 전기 전자공학과 (Dept. of Electrical & Electronic Eng., Yonsei Univ.)

*** 정회원, 연세대학교 정보기술학부 (Division of information technology, Yonsei Univ.)

접수일자: 2003년9월27일, 수정완료일: 2004년11월4일

우선 입력된 PCM 신호는 인간의 주관적인 심리청각특성(Psychoacoustics)을 이용하여 부호화하기 위해 주파수 영역으로 변환된다. 이때 이에 적용되는 신호의 청각적 특성은 심리음향모델을 병행함으로써 얻어낸다. 주파수 영역으로 변환된 신호는 동시에 얻어진 심리음향모델의 마스킹 임계치(Masking threshold)값을 이용하여 각각에 효율적인 비트가 할당되고 MP3 표준 비트열로 구성된다^[1]. 부호화 과정의 연산량을 살펴볼 때 심리음향모델과 비트할당 과정이 전체 연산량의 80%이상을 차지하므로, 실시간 MP3 부호화를 위해서는 이 두 과정의 연산량을 줄이는 것이 필요하다^[4].

MPEG 오디오 부호화에서 사용되는 심리음향모델은 모델 I과 모델 II가 있으며^[1], 두 모델 모두 FFT 스펙트럼으로부터 부호화에 사용될 마스킹 임계치를 얻는다. 모델 I의 경우는 입력 신호의 파워 스펙트럼(FFT의 크기 정보)만을 가지고 임계치를 얻고, 모델 II에서는 FFT의 크기와 위상 정보를 모두 사용하여 임계치를 얻는다. 이때 FFT가 차지하는 연산량은 모델 I의 경우 약 50%, 모델 II의 경우 약 40%로 비중이 매우 높다. 따라서 FFT 스펙트럼의 정보를 효율적으로 구할 수 있는 방법이 있다면, 연산량의 많은 이득을 기대할 수 있다.

MPEG 오디오 부호화 과정에서는 심리음향모델을 위한 FFT 변환 이외에 부호화를 위한 T/F 변환 과정으로 서브밴드 필터뱅크 및 MDCT를 수행한다^[1]. T/F 변환을 통해 얻어진 스펙트럼은 FFT 스펙트럼과 일치하지는 않지만, 유사한 특성을 갖고 있다. 그러나 T/F 변환은 이상적인 대역통과 필터를 사용할 수 없기 때문에 서브밴드 간의 중첩에 의한 에일리어징이 발생한다^{[5][6]}. 따라서 서브밴드 필터뱅크의 결과를 심리음향모델에 직접 사용하게 되면, 많은 왜곡이 발생하여 정확한 마스킹 임계치를 얻을 수 없다.

본 논문에서는 심리음향모델에 사용되는 FFT 스펙

트럼을 서브밴드 필터뱅크의 결과를 이용하여 구함으로써, FFT 스펙트럼을 기존 방법^[1]에 비해 절반의 연산량으로 얻을 수 있는 방법을 제안한다.

II. 서브밴드 분석과 FFT

MP3의 T/F변환 과정은 그림 2와 같이 입력된 PCM 신호를 서브밴드별로 분석하는 과정과 각 서브밴드신호에 대하여 MDCT를 수행하는 과정으로 구성된 하이브리드 구조를 갖는다. 입력된 PCM 신호는 32개의 대역통과필터를 통과 한 후 간축(decimation)되고 각 밴드별로 MDCT 변환이 수행된다^{[1]-[3]}.

각 밴드별로 N:1로 간축되어 나온 신호를 (T/N)-샘플 FFT를 하면 각 bin이 나타내는 주파수의 성분은 간축되지 않은 신호의 T-샘플 FFT에서와 같은 주파수 분해능을 갖는다^[7]. 이 점에 착안하여 각 서브밴드 필터뱅크의 출력을 (T/N)-샘플 FFT한 후 이들을 밴드 순으로 정렬하면 T-샘플 FFT와 유사한 형태의 스펙트럼을 얻을 수 있다. 이 과정을 그림 3에 나타내었다.

그림 3은 입력신호에 대한 서브밴드분석을 한 후 각 밴드별로 나온 신호에 대하여 각각 FFT를 하여 정렬함으로써 입력신호를 직접 FFT하여 얻은 기존 방법에서의 스펙트럼과 유사한 값이 나오게 한 것을 보여준다.

이때 서브밴드 필터뱅크를 통과한 신호에 대한 FFT와 시간 영역 신호에 대한 FFT는 크기 스펙트럼은 유사한 형태를 갖지만 위상 스펙트럼에 대하여는 유사성이 없다.

MP2의 경우에는 단순히 FFT의 크기 정보를 이용하여 심리음향모델을 구현하므로 위상의 스펙트럼의 비유사성이 구현상 문제가 되지 않지만 MP3에서는 심리음향 모델을 FFT 크기와 위상 스펙트럼 모두를 이용하여 심리음향모델을 구현하므로 위상스펙트럼의 비유사

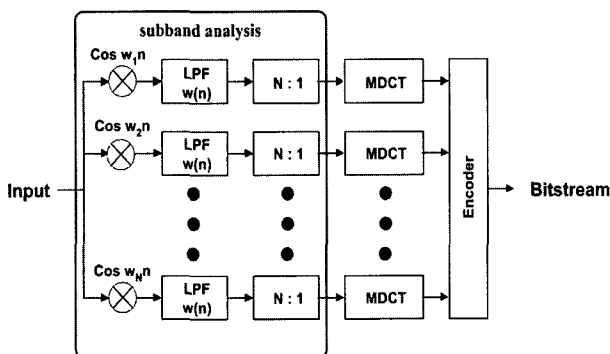


그림 2. 시간-주파수 변환과정
Fig 2. Time-Frequency Transform.

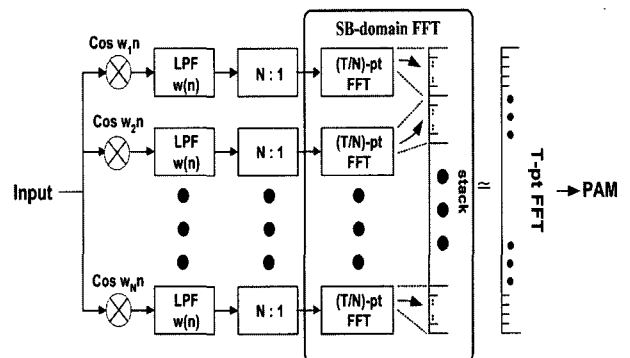


그림 3. 서브밴드 샘플을 이용한 FFT 연산과정
Fig 3. FFT Using Subband Analysis.

성이 구현상 문제가 될 수 있다.

그러나 MP3의 심리음향모델을 살펴보면 FFT 위상 스펙트럼은 순음지수(tonality)의 기준이 되는 비예측성(unpredictability; c_w)을 구하는데 쓰이게 된다. 비예측성은 이전 두 프레임에서 구한 FFT 스펙트럼을 이용하여 현재의 스펙트럼을 예측하고 현재의 실제 FFT 스펙트럼과 비교하여 얼마나 예측이 잘되었는가를 나타냄으로써 순음(tonal)적인 성분을 판단하는 기준이 된다^[1]. 그러므로 각각 3개의 프레임에 스펙트럼 성분이 모두 같은 위상으로 대역통과필터를 통과한다고 본다면 모두 같은 위상으로 틀어져 있으므로 비예측성을 계산하는 과정에는 상관없게 된다. 따라서 FFT의 위상스펙트럼의 비유사성 또한 MP3에서도 문제가 되지 않는다.

서브밴드 필터에 입력되는 PCM신호는 서브밴드 필터를 통과하기 때문에 필터의 특성상 생기는 지연을 고려해야 한다. MP3의 서브밴드 필터는 512-tap의 FIR 필터를 기반으로 설계되어 있다. 따라서 각 서브밴드를 통과한 신호의 FFT 스펙트럼은 256샘플 지연된 PCM 입력에 대한 FFT 스펙트럼에 대응된다.

서브밴드 필터뱅크는 원형(Proto-type) FIR 필터의 주파수 변조를 통해 얻어졌기 때문에 주파수 영역에서 인접 밴드 사이에 중첩이 존재한다. 따라서 인접 밴드의 신호가 에일리어징 신호로써 각 밴드에 영향을 미치게 되며, 이는 원래의 스펙트럼에 대한 왜곡으로 작용하게 된다^{[5][8]}.

그림 4는 32서브밴드 출력을 32-샘플 FFT를 하여 재구성한 결과를 전체밴드신호를 1024-샘플 FFT하여 얻은 결과와 비교한 것이다. 보는 바와 같이 한 주파수에 의한 신호가 해당 서브밴드 이외에 인접밴드에도 영

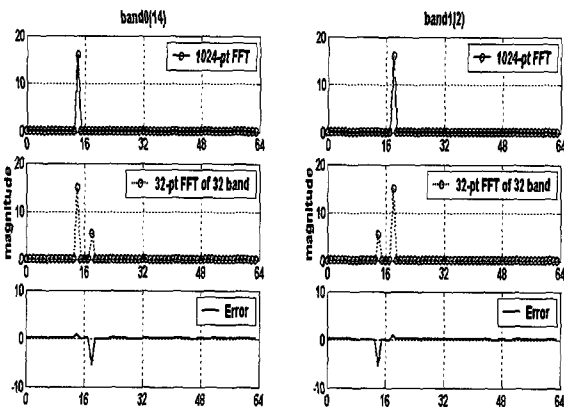


그림 4. 단일주파수의 정현파에 의한 감쇄된 성분과 왜곡 성분

Fig 4. Attenuate and Alias Component by Pure Tone.

향을 미치는 것을 관찰할 수 있다. 이때 발생하는 에일리어징은 복호화단에서 수행되는 합성 필터뱅크에 의해 상쇄되기 때문에 부호화 관점에서는 문제가 되지 않지만, 심리음향모델을 위한 스펙트럼을 얻는 관점에서는 제거되어야 하는 왜곡이 된다.

III. 에어리어징 제거

1. 버터플라이 구조의 에일리어징 제거기

2절에서 서브밴드를 통과한 신호를 이용하여 FFT를 할 경우 왜곡이 발생하는 것을 보았다. 왜곡의 원인을 살펴보면 서브밴드 분석을 하는 과정에서 간축에 의한 중첩에 의해 생기는 것으로 볼 수 있다^[9].

그림 5에서 밴드0을 살펴볼 때 표현가능 한 주파수 범위가 간축에 의해 1/N으로 감소된데 비해 밴드를 통과한 신호는 Nyquist주파수보다 높아서 그 잔상에 의해 왜곡이 생기게 된다^[10]. 마찬가지로 나머지 모든 밴드가 같은 이유에서 중첩되며 왜곡이 생기게 된다.

이미 왜곡 성분을 포함하고 있는 주파수의 하나의 bin만을 가지고 원 신호에서 왜곡성분을 분리하여 제거할 순 없다. 하지만 대역통과 필터가 원형(prototype) 저역통과필터를 이용해 만든 것이므로 각 밴드마다 그 형태가 같다는 점을 이용한다면 그림 6과 같은 구조에서 왜곡을 줄일 수 있다.

그림 4에서 본 것과 같이, (n-1)번째 밴드에 속한 bin의 주파수 성분은 서브밴드 분석블록을 거치면서 (n-1)번째 밴드에서는 감쇄된 형태로, n번째 밴드에서는 왜

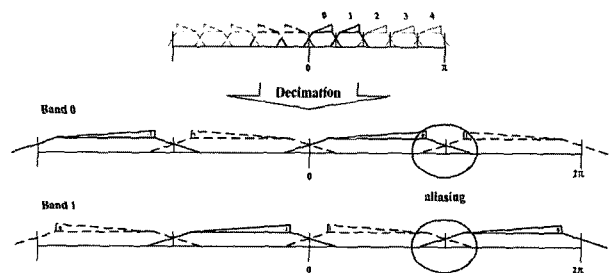


그림 5. 5개로 구성된 서브밴드 필터와 간축의 예
Fig 5. Operation of 5-Subband Filter Decimation.

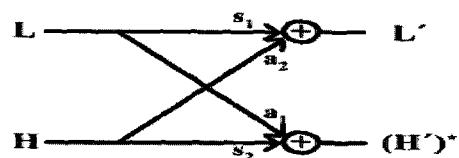


그림 6. butterfly 구조에 의한 왜곡 성분 추정
Fig 6. Estimate to Alias Component From Butterfly Structure.

곡성분으로 작용한다. 이와 대칭적으로 n번째 밴드에 속한 bin의 주파수 성분은 서브밴드 분석블록을 거치면서 n번째 밴드에서는 감쇄된 형태로, (n-1)번째 밴드에서는 왜곡성분으로 나타나게 된다. 따라서 한 주파수 bin은 대칭적인 구조를 가지는 두 개의 주파수 성분으로부터 만들어 진다고 볼 수 있다.

$$L \cdot s_1 + H \cdot a_2 = L', H \cdot s_2 + H \cdot a_1 = (H')^* \quad (1)$$

식(1)에서 s_1, s_2 는 감쇄상수, a_1, a_2 는 왜곡상수를 나타낸다.

다시 말해, 간축에 의해 $band_{n-1}(j)$ 에서 나온 값은 $L, band_n(k)$ 에서 나온 값을 H' 라 하면 이 값들은 다음과 같이 표현할 수 있다($1 \leq n \leq 31, 9 \leq j \leq 15, 1 \leq k \leq 7, n, j, k$ 는 정수). L 는 $band_{n-1}$ 에 해당하는 값 L 의 감쇄와 $band_n$ 의 값 H 에 의해 생기는 왜곡성분의 합으로 표현할 수 있다. 마찬가지로 H' 는 $band_{n-1}$ 에 해당하는 L 에 의해 생기는 왜곡과 $band_n$ 의 값 H 의 감쇄 성분의 합으로 표현할 수 있다. 여기에 적용되는 계수 값은 필터의 특성으로 구할 수 있는데, $s_1(s_2)$ 는 $L(H)$ 를 $L'(H')$ 로 감쇄 시키는 값이고, $a_1(a_2)$ 는 $L(H)$ 을 $H'(L')$ 만큼의 a_2 왜곡성분으로 더해지는 값이다.

식(1)의 두 방정식을 풀면 식(2)와 같이 전체샘플에 대한 FFT의 값에 해당하는 신호를 얻을 수 있다.

$$H = \frac{1}{a_1 a_2 - s_1 s_2} (a_1 \cdot L' - s_1 \cdot (H')^*)$$

$$L = \frac{1}{a_1 a_2 - s_1 s_2} (a_2 \cdot (H')^* - s_2 \cdot L') \quad (2)$$

2. 중간 값의 에일리어징 제거

서브밴드 필터의 주파수 응답을 살펴보면 그림 7과 같이 그 대칭적인 구조가 정확히 밴드와 밴드의 사이 값이 아닌 높은 밴드 쪽의 1번째 bin을 중심으로 나타남을 볼 수 있다.

따라서 $band_{n-1}(j)$ 와 $band_n(k)$ 로 버터플라이를 적용하여 왜곡성분을 제거 할 때 $band_n(0)$ 은 적용할 수 없다.

각 밴드별로 32-샘플 FFT를 한 후, 그 결과 값들을 이용하여 1024-샘플 FFT와 유사한 스펙트럼을 만들 때 각 밴드별로 16개의 스펙트럼만을 이용하며,

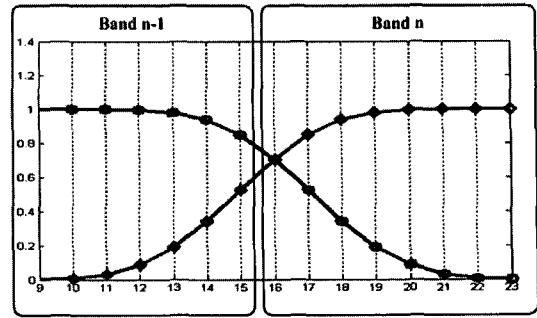


그림 7. 인접한 Band에 Bandpass Filter의 주파수 응답
Fig 7. Frequency response of adjacent two bands.

$band_{n-1}(16)$ 에 해당하는 주파수 성분은 사용하지 않는다. 이를 그림 7에서 살펴보면 이 성분이 이웃 밴드인 $band_n(0)$ 에 해당하는 값과 같은 성분임을 알 수 있다.

이 성분은 (n-1)번째 밴드와 n번째 밴드의 주파수 응답이 겹치는 부분이고, 또한 간축될 때에도 다른 주파수 성분에 영향을 받지 않는다. 오직 그 주파수 성분이 감쇄되고 그 잔상과 더해져서 실수 값을 갖게 됨을 알 수 있다. 이를 그림 8과 같이 나타낼 수 있다.

$band_{n-1}(16)$ 과 $band_n(0)$ 의 bin은 (n-1)번째 서브밴드와 n번째 서브밴드의 필터를 거쳤다고 보면 그 크기는 같으나 서브밴드 필터의 위상성분에 의해 두 bin간에 위상이 틀어져 있음을 알 수 있다. MPEG에서 적용되는 서브밴드 필터는 2절에서 언급한 바와 같이 저역통과 필터를 코사인변조(cosine modulation)를 통해 만든 것이고 그 인접한 밴드간의 주파수 응답을 보면 90도씩 틀어져 있음을 볼 수 있다.

이를 생각하면 그림 8에서와 같이 하나의 주파수 성분이 똑같은 크기를 갖고 90도 틀어진 형태로 필터링된 성분 (A, B)으로 나타나게 된다고 볼 수 있고, 그 성분들은 각각의 잔상(A^*, B^*)과 더해져서 값의 허수 부분이 없어지고 실제로 관찰 되는 부분은 실수부분만 남아 나타나게 된다(A', B'). 이 값이 (n-1)번째 밴드의 $band_{n-1}(16)$ 과 n번째 밴드의 $band_n(0)$ 으로 나타나게 된다.

따라서, 두 실수 값, $band_{n-1}(16)$ 과 $band_n(0)$ 을 가지고 원래 신호(X)의 크기를 구해 보면

$$X = \sqrt{\left(\frac{A'}{2}\right)^2 + \left(\frac{B'}{2}\right)^2} \times \sqrt{2} \quad (3)$$

$$(X = Bin'_{(16 \times n)}, A' = band_{n-1}(16), B' = band_n(0))$$

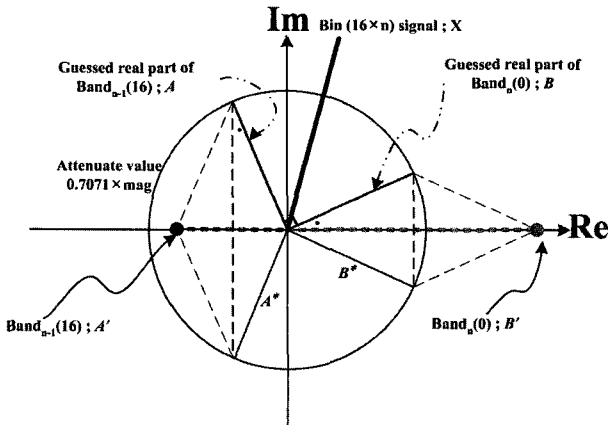


그림 8. $band_n(0)$ 과 $band_{n-1}(16)$ 으로 나타나는 $16 \times n$ 번째 Bin
 Fig 8. Estimate magnitude spectrum from $band_n(0)$ and $band_{n-1}(16)$.

과 같이 구할 수 있다. 이때 추정된 값은 이미 실수화 되어 있는 값으로 구하므로 단지 그 크기만을 추정할 수 있다.

IV. 심리음향 모델에서의 적용

이 장에선 오디오 부호화 과정 중 FFT의 결과가 심리음향모델II를 사용하는 MP3 부호화기에서 어떻게 적용되는지를 살펴본다.

MP3에서는 FFT의 결과를 이용하여 비예측성(unpredictability, $c_w \in [0, 1]$)을 구하게 된다. c_w 는 0에 가까운 값일수록 예측률이 높음을 뜻한다^[1]. c_w 값은 1024-샘플 FFT 결과의 저역주파수 bin 6개와 별도의 256-샘플 FFT의 고역주파수를 이용하여 구하게 된다. 1024-샘플 FFT만을 제안한 방법으로 대체할 경우 하나의 그레놀(Granule)의 c_w 값 중 저역주파수 6개 bin만이 바뀌게 된다. 그러나 256-샘플 FFT 또한 대체할 경우 고역 주파수 bin까지 포함하여 바뀌게 된다. c_w 값은 식(4)에 의해서 계산되는데 다음과 같다.

$$c_w = \frac{\sqrt{(r_u \cdot \cos f_w - \hat{r}_u \cdot \cos \hat{f}_w)^2 + (r_w \cdot \sin f_w - \hat{r}_w \cdot \sin \hat{f}_w)^2}}{r_w + abs(\hat{r}_w)} \quad (4)$$

r_w : 크기 \hat{r}_w : 예측한 크기

f_w : 위상 \hat{f}_w : 예측한 위상

이 c_w 값은 e_b , c_b 값에서 c_b 값을 구하게 되는 경우에

사용된다. e_b , c_b 는 각각 파티션(partition)별로 에너지를 더한 것과 그 에너지에 비예측성을 곱한 형태로, 다음과 같다.

$$e_b = \sum_{w=low}^{high} r_w^2 \quad (5)$$

$$c_b = \sum_{w=low}^{high} r_w^2 \cdot c_w \quad (6)$$

위 식에서 c_w 는 주파수 ω 에서의 비예측성을, r_w 는 주파수 ω 에 해당하는 FFT의 크기값을 나타낸다.

식(6)에서와 같이 c_w 값이 심리음향 모델에 반영되는 비중은 FFT의 절대값 크기와 비례하게 각 서브밴드별로 더해지게 되므로 r_w 가 아주 작은 경우에 c_w 의 에러는 큰 문제가 되지 않는다. 여기서 사용되는 r_w 는 롱블록 FFT의 결과를 이용하게 된다.

심리음향모델을 수행하고 나오는 최종 값은 SMR (Signal to Mask Ratio)이다. 이는 신호에 대하여 마스킹 되어지는 값이 어느 정도인지를 나타내는 값으로써 매 그레놀마다 21개의 값으로 나타내게 된다^[1].

심리음향모델에서 블록을 마치고 나온 이 SMR값은 MP3에서 반복루프(Iteration Loop)를 실행할 때 크기인 자대역(Scalefactor Band)에서 허용 가능한 왜곡치(Allowed Distortion)를 나타내는 $Xmin$ 의 값을 측정할 때 사용되게 된다.

$$Xmin = ratio(sb) \times en(sb) / bw(sb) \quad (7)$$

식 (7)에서 보는바와 같이 $Xmin$ 의 값은 심리음향모델에서 나온 SMR, 즉 $ratio$ 값과 심리음향모델과는 상관없는 MDCT의 결과에서 나온 en 과 bw 의 곱으로 나타내게 되므로 결국 $Xmin$ 의 오차는 SMR값의 오차와 같게 된다.

앞서 언급한 것과 같이 MP3 부호화기에서는 저주파와 고주파에서의 분해능을 변화시키기 위해 1024-샘플 FFT와, 256-샘플 FFT를 동시에 수행한다. 두 가지 모두에 대해 제안한 FFT연산 방법을 적용할 수 있다. 1024-샘플 FFT는 32-서브밴드 출력을 32-샘플 FFT를 이용하고, 256-샘플 FFT는 32band의 서브밴드 출력을 8-샘플 FFT를 이용하여 유사한 결과를 구할 수 있다.

8-샘플 FFT의 경우에는 하나의 bin에 대하여 에일리어징 제거 버터플라이를 적용하게 되며, 분해능이 낮아 에일리어징제거 블록에 적용할 수 있는 bin은 각

밴드당 하나의 bin밖에 적용할 수 없다.

V. 실험 및 결과

1. 성능평가

이 장에선 제안된 알고리즘의 성능을 평가해 본다. 우선 기존의 방법과 제안한 방법의 FFT 스펙트럼을 비교하여 그 오차를 살펴보았다. 입력신호를 0dB FS의 백색잡음의 랜덤신호를 사용하고, 제안한 알고리즘을 적용하였다. 그림 9-(a)는 롱블록의 경우 밴드 경계부분에서 에일리어징 제거 알고리즘을 적용하지 않았을 경우에 비해 알고리즘을 적용할 경우 약 15dB 정도의 오차가 줄어들음을 보여준다. 이 때 기존의 심리음향모델에 사용되는 FFT신호에 대한 SNR은 약 24dB이다. 한편 그림 9-(b)와 같이 숏블록의 경우에선, 에어리어징 제거 알고리즘을 적용하였을 경우가 그렇지 않았을 경우에 비해 약 4dB정도 오차가 줄어들음을 보여준다. 이때의 SNR은 약 13dB이다.

다음은 이 경계부분의 에러가 심리음향모델에 어떤 영향을 주는지를 살펴보았다. 0dB FS 2kHz 사인파에 -26dB FS noise신호가 첨가된 신호를 이용하여 테스트 하였다. 2kHz는 1024-샘플 FFT에서 46번째와 47번째 bin 사이에 해당하며, 이는 서브밴드2와 3의 경계부분이다, 또한 심리음향모델II의 중간과정에 사용되는

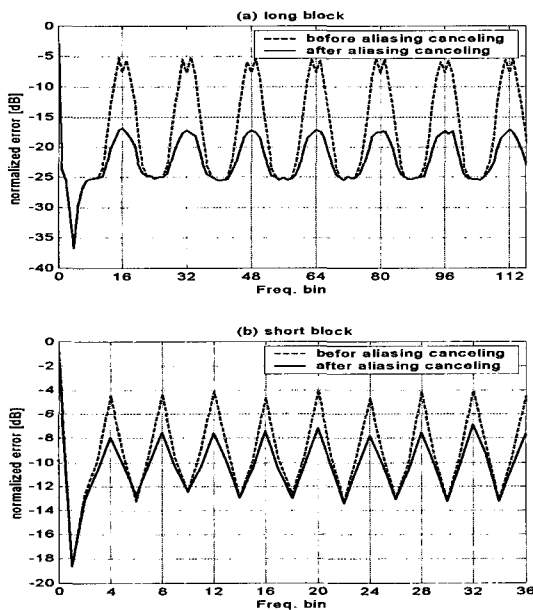


그림 9. (a)1024FFT와 32x32FFT의 크기의 오차
(b)256FFT와 32x8FFT의 크기의 오차
Fig 9. (a) error between 1024pt-FFT and 32x32pt-FFT
(b) error between 256pt-FFT and 32x8pt-FFT.

partition과 Critical Band(CB)에서 살펴볼 경우, 30번째 partition, 12번째 CB에 해당하는 성분이다^[1]. 실험은 1024-샘플 FFT만을 제안한 방법으로 대체한 경우와 1024-샘플 FFT와 256-샘플 FFT를 모두들 제안한 방법으로 대체한 두 가지 경우에 대하여 살펴보았다. 우선 1024-샘플 FFT만을 대체한 경우에 있어서 FFT의 결과가 가장 먼저 쓰이는 c_w 값을 살펴보면, 롱블록 FFT만을 제안한 방법으로 구현하였을 경우 앞 절에서 언급한 것과 같이 c_w 값은 앞에 6개에서 작은 오차가 생

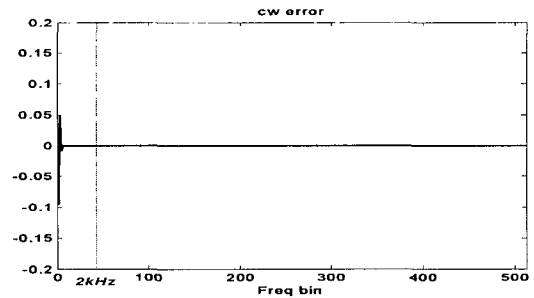


그림 10. 롱블록 FFT를 제안한 방법으로 대체하였을 때 c_w 오차
Fig 10. c_w error due to using the proposed long block FFT.

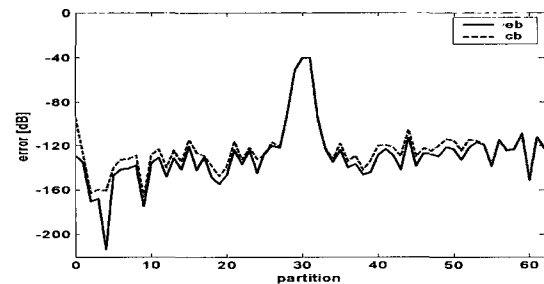


그림 11. 롱블록 FFT를 제안한 방법으로 대체하였을 때 e_b, c_b 오차
Fig 11. e_b, c_b error due to using the proposed long block FFT.

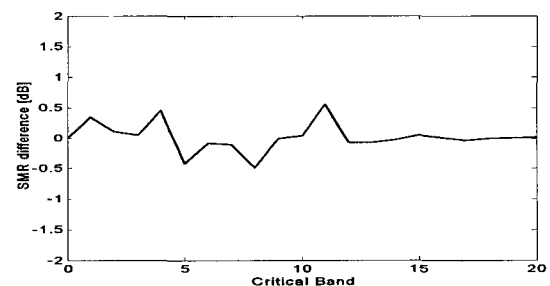


그림 12. 롱블록 FFT를 제안한 방법으로 대체하였을 때 SMR의 오차
Fig 12. SMR error due to using the proposed long block FFT.

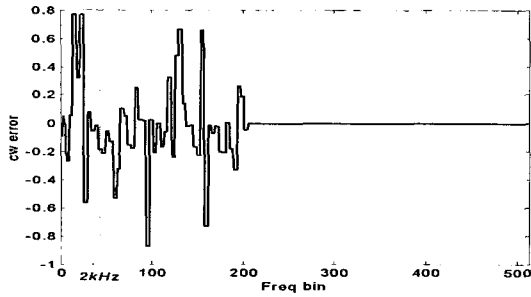


그림 13. 롱블록, 숏블록 FFT모두 제안한 방법으로 대체 하였을 때 c_w 오차

Fig 13. c_w error due to using the proposed long & short block FFT.

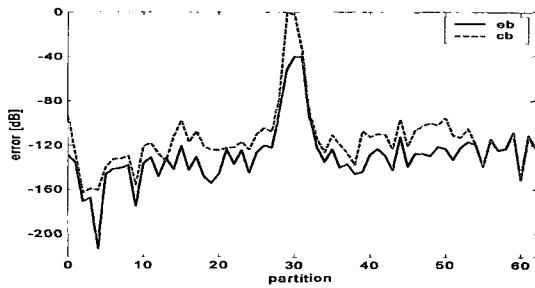


그림 14. 롱블록, 숏블록 FFT모두 제안한 방법으로 대체 하였을 때 e_b, c_b 오차

Fig 14. e_b, c_b error due to using the proposed long & short block FFT.

기고, 나머지 c_w 값은 일치하게 된다[그림 10]. 2kHz에 의한 partition에서의 에너지 오차는 e_b 로 표현이 되며 이 값은 그림 11에서 보는바와 같이 -40dB정도로 작다. c_w 의 값이 제안한 방법과 기존의 방법에서 매우 유사하므로, 이 오차는 e_b 의 경우도 비슷하게 되고, 결국 심리 음향모델의 최종결과인 CB에서의 SMR값의 dB 오차는 그림 12에서 보는바와 같이 약 1dB이내로 들어오게 된다. SMR의 1dB는 사람의 귀가 인지하는데 별 무리가 없다^[12].

그러나 256-샘플 FFT의 경우와 1024-샘플 FFT 모두에 제안한 알고리즘을 적용할 경우, 256-샘플 FFT에서 4개의 bin마다하나에서 중간 값을 이용하여 절대값의 크기를 예측하므로 phase값을 가지고 있지 않고, 분해능이 떨어져 에일리어징 성분이 크게 제거되지 않으므로 c_w 값이 틀려지게 된다. 따라서 그림 13에서 보는 바와 같이 오차가 발생하게 된다. 오차가 발생을 하더라도 2kHz이외의 부분에서는 그 에너지가 매우 작으므로 c_w 값의 에러는 큰 의미가 없다. 그러나 그림 13의 결과는 2kHz 성분에서도 오차가 발생함을 보여주며, 이

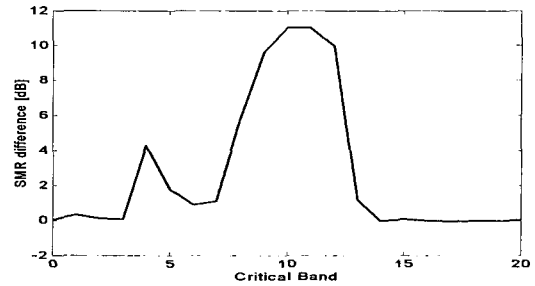


그림 15. 롱블록, 숏블록 FFT모두 제안한방법으로 대체 하였을 때 SMR 오차

Fig 15. SMR error due to using the proposed long & short block FFT.

표 1. 제안한 알고리즘과 기존의 알고리즘의 MOS test 결과

Table 1. MOS test of proposed Algorithm vs conventional Algorithm.

● 1024FFT만을 32x32FFT로 대체한 경우

	기존>제안	기존<제안	총합
샘플개수	4	6	10
차이점수 (기존방법-제안방법)	0.12	-0.11	-0.016

● 1024FFT를 32x32FFT로, 256FFT를 32x8FFT로 대체한 경우

	기존>제안	기존<제안	총합
샘플개수	7	3	10
차이점수 (기존방법-제안방법)	0.21	-0.10	0.12

c_w 값의 오차로 인해 그림 14에서 보는 것과 같이 e_b 가 -40dB의 작은 오차를 가지고 있더라도 c_b 는 0dB에 가까운 큰 오차를 가지게 된다. 이 오차로 인해 심리음향 모델의 최종 결과물인 CB에서 SMR이 그림 15와 같이 약 12dB정도의 값의 차이를 가지게 된다.

주관적인 음질 평가를 통하여 SMR의 오차가 어느 정도 지각되는지를 실험 하였다. 실험방법은 오디오 샘플 10개를 10명의 오디오를 전공하는 사람에 대하여 권고안(ITU-R BS.1116)^[11]에서 제시한 방법으로 하였다. 음악샘플마다 음질에 따라 1부터 5까지의 점수를 줄 수 있으며, 0.1 단위로 점수를 부여할 수 있다. 5점이 가장 좋은 음질을 나타낸다. 표 1은 제안한 방법과 기존 방법의 상대적인 점수 차이를 보여주고 있다.

표 1과 같이 10개의 음원에 대하여 각각의 MOS (means opinion score) test 결과 값을 평균을 내었다. 실험 결과 롱 블록에 대해서만 제안한 방법의 알고리즘을 적용하였을 경우에는 기존의 알고리즘으로 부호화를 하였을 때와 지각적인 측면에서는 별 차이가 없음을 볼 수 있다. 그러나 롱 블록과 숏 블록 모두를 제안한 방법으로 제안하였을 경우에는 몇몇 테스트 음원에 대하여

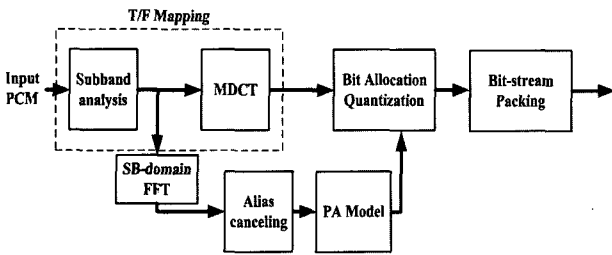


그림 16. 제안한 MP3부호화 과정

Fig 16. Proposed encoding flow with alias canceling block.

표 2. 1024-샘플과 32band 32-샘플 FFT의 연산량 비교

Table 2. Complexity of Fully FFT vs split FFT in 32 band.

		Multiplication [(N/2)·log ₂ N]	Addition [N·log ₂ N]	Total
Full-band	1024-ptFFT	5120	10240	15360
	256-pt FFT	1024	2048	3072
32 band split	32x32-pt FFT	2560	5120	7680
	32x8-pt FFT	384	768	1152

표 3. 에일리어징 제거 블록의 연산량

Table 3. Complexity of alias canceling block (k is the number of butterfly).

	k	Multiplication [(N-1)·k·4]	Addition [(N-1)·k·2]	Total
Alias canceling block	Long	7	868	434
	Short	1	124	62

음질차이를 느낄 수 있었으나, 그 정도가 심하지 않았다.

2. 연산량 측정

FFT의 연산량은 $(N/2) \cdot \log_2 N$ 개의 복소수 곱셈과 $N \cdot \log_2 N$ 개의 복소수 덧셈으로 구하게 된다^[10]. 표 2에 심리음향모델에 적용되는 FFT의 연산량을 기존의 방법과 제안한 방법에 대하여 비교하였다. 또한 제안한 알고리즘은 에일리어징 제거 블록의 연산량을 고려해야 한다. 에일리어징 제거 블록의 연산량을 표 3에 나타내었다. 표 3에 나타난 만큼의 연산량이 증가하게 되므로, 제안한 방법의 연산량은 기존방법에 비해 약 40% 감소된다.

그러나 실제 오디오 부호화에 있어 신호가 0이 되는 높은 밴드에 대해서는 부호화를 하지 않는다. 최소한 이 밴드는 3개 이상이므로 제안한 방법에서 해당 밴드에 대한 FFT연산을 선택적으로 수행할 수 있다. 따라서 실제 연산량의 감소량을 생각하면 약 50%의 이상의 감소를 얻을 수 있다.

표 4. 에일리어징 제거 버터플라이 상수

Table 4. The coefficients of aliasing cancellation butterfly.

● Long Block

Butterfly Index	s ₁	s ₂
1	-0.40524492355344 + j0.747096630507	0.07804031144576 + j0.846333520255
2	-0.28368023190658 + j0.895467057872	0.82287298302789 + j0.453004371337
3	-0.11420709536532 + j0.975169682319	0.87431794961783 - j0.446729940248
4	0.07328717778620 + j0.993545083585	0.12205566902257 - j0.988740346981
5	0.26066664379403 + j0.964897779698	-0.75272821567163 - j0.657554255843
6	0.43861392630603 + j0.898675306632	-0.96700693033489 + j0.254752308976
7	0.60062479080509 + j0.799555025009	-0.33122746780005 + j0.943571315185
	a ₁	a ₂
1	0.46311227716308 + j0.251206344982	0.52463442525880 - j0.048379137943
2	0.32692734465429 + j0.103574064933	0.16539192616833 - j0.300428644921
3	0.18867679886826 + j0.022099662134	-0.08643571117338 - j0.169166865620
4	0.08633818813966 - j0.006368309143	-0.08592245108201 - j0.010606375159
5	0.02972457697376 - j0.008030367818	-0.02025924065964 + j0.023186831531
6	0.00694725049148 - j0.003390729077	0.00196533601642 + j0.007475920120
7	0.00084557446098 - j0.000636636178	0.00099741609622 + j0.000350576724

● Short Block

Butterfly Index	s ₁	s ₂
1	0.07328749080502 + j0.99354505189413	0.1219515596210 - j0.98875206374376
	a ₁	a ₂
1	0.08633889957720 - j0.00636739482922	-0.08592247296036 + j0.1059612842576

VI. 결 론

본 논문에서는 서브밴드 분석 필터뱅크를 통과한 서브밴드 샘플들에 대하여 각각의 서브밴드별로 32-샘플 FFT를 수행하고 에일리어징 제거 알고리즘을 적용함으로써 1024-샘플 FFT와 유사한 스펙트럼을 얻는 방법을 제안하였다. 이는 기존의 1024-샘플 FFT 방법에 비해 연산량 측면에서 효율적이며, MP3 부호화기의 실시간 구현에 있어서 큰 문제가 되는 심리음향모델을 연산량을 크게 줄일 수 있다. 또한 심리음향의 음질보다 연산량 측면을 고려해야 할 경우 256-샘플 FFT 도 8-샘플 FFT를 이용하여 대체할 수 있다.

제안하는 방법은 FFT 스펙트럼의 위상 성분을 사용

하지 않는 심리음향모델 I이나 위상성분을 사용하는 심리음향모델 II에 모두 적용할 수 있으며, 이때 심리음향 모델 전체의 연산량은 기존에 비해 각각 25%, 20% 감소 효과를 얻을 수 있다.

참 고 문 헌

[1] ISO/IEC IS 11172-3, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s-Audio Part," Nov. 1992.
 [2] P. Noll, "MPEG Digital Audio Coding," Signal Processing Magazine, IEEE, Vol: 14 Issue: 5, Sept. 1997.
 [3] K. Brandenburg, G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," J. Audio Eng. Soc., Vol.42, No. 10, Oct. 1994.
 [4] Hyen-O Oh, J. S. Kim, C. H. Song, Y. C. Park, and D. H. Yoon, "Low power MPEG/AUDIO encoders using simplified psychoacoustic model and fast bit allocation", Consumer Electronics, IEEE Transactions on , Vol: 47 Issue: 3 , Aug. 2001 Page(s): 613-621
 [5] B. Edler, "Aliasing Reduction in Subbands of Cascaded Filter Banks with Decimation", Electronics Letters, Vol.28, No.12, pp 1104-1105, June 1992.

[6] Chi-Min Liu, Wen-Chieh Lee, "The Design of A Hybrid Filter Bank for The Psychoacoustic Model in ISO/MPEG Phases 1, 2 Audio Encoder", IEEE Transactions on Consumer Electronics, Vol.43, No.3, Aug. 1997.
 [7] P.P. Vaidyanathan, "MULTIRATE SYSTEMS AND FILTER BANKS," Prentice Hall Inc., 1993.
 [8] Davis Pan, "A Tutorial on MPEG/Audio Compression," IEEE Multimedia , Vol. 2 Issue. 2, pp 60-74, Summer 1995.
 [9] J. H. Rothweiler, "Polyphase Quadrature Filters A New Subband Coding Technique," in Proc. Int. Conf. IEEE ASSP. pp1280-1283
 [10] John G. Proakis, D. G. Manolakis, "DIGITAL SIGNAL PROCESSING," Prentice Hall Inc., 1996.
 [11] Rec. ITU-R BS.1116-1, "METHODS FOR THE SUBJECTIVE ASSESSMENT OF SMALL IMPAIRMENTS IN AUDIO SYSTEMS INCLUDING MULTICHANNEL SOUND SYSTEMS," 1994-1997.
 [12] Christian Neubauer, "A Compatible Family of Bitstream Watermarking Schemes for MPEG-Audio", AES 110th Convention, Amsterdam, May 2001.

저 자 소 개



송 건 호(정회원)
 2002년 한양대학교 전자컴퓨터 공학부 학사졸업.
 2004년 연세대학교 전기전자 공학과 석사졸업.
 2004년 현재 LG전자 전자기술원 근무

<주관심분야: 신호처리, 오디오 부호화>



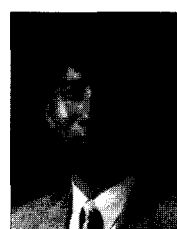
이 근 섭(정회원)
 1997년 연세대학교 전자공학과 학사졸업.
 1999년 연세대학교 전기전자 공학과 석사졸업.
 2004년부터 현재 연세대학교 전기 전자공학과 박사과정

<주관심분야: 디지털 신호처리, 오디오 신호처리>



박 영 철(정회원)
 1986년 연세대학교 전자공학과 학사졸업.
 1988년 연세대학교 전기전자 공학과 석사졸업.
 1993년 연세대학교 전기전자 공학과 박사졸업.

2004년 현재 연세대학교 정보기술학부 교수
 <주관심분야: 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리>



윤 대 희(정회원)
 1977년 연세대학교 전자공학과 학사졸업.
 1979년 Kansas State University 석사졸업.
 1982년 Kansas State University 박사졸업.

2004년 현재 연세대학교 전기전자공학과 정교수
 <주관심분야: 적응신호처리, 음성신호처리, 오디오 신호처리>

