

# 간략화된 심리음향모델을 이용한 MPEG Audio Layer-III 부호화기

송창준, 오현오, 박영철\*, 윤대희

연세대학교 전기전자공학과 미디어·통신 신호처리 연구실

\*연세대학교 신호처리연구센터

## MPEG Audio Layer-III Encoder Using Approximated Psychoacoustic Model

Chang-Jun Song, Hyen-O Oh, Young-Cheol Park\*, and Dae-Hee Youn

MCSP Lab., Dept. of Electrical and Electronic Eng., Yonsei University

\*Center for Signal Processing Research, Yonsei University

E-mail: cooljun@cyclon.yonsei.ac.kr

### 요약

MPEG Audio Layer-III(MP3) 알고리즘은 복호화기에 비해 부호화기가 월등히 많은 연산량을 가지고 있는 비대칭 구조를 가지고 있다. MP3 부호화기의 대부분의 연산량은 복잡한 초월함수 연산이 포함되는 심리음향모델과 반복 루프 과정을 수행하는 비선형 양자화와 비트 할당 과정이 차지한다.

본 논문에서는 MP3 부호화기의 실시간 구현을 위한 알고리즘 레벨의 최적화를 수행하였다. MP3 부호화기의 연산량을 줄이기 위해 심리음향모델을 간략화하고 반복 루프의 회수를 최소화할 수 있는 방법을 제안하였다. 프레임당 한 그래놀의 심리음향모델 정보를 계산하여 한 프레임 내에서의 심리음향모델 정보를 추정함으로써 연산량을 45% 이상 감소시켰다. 또한 외부 반복 루프의 반복 회수를 줄이기 위하여 외부 반복 루프의 반복에 따른 스케일 팩터(Scale Factor) 및 양자화 스텝의 증가 패턴을 관찰하고 최적화된 스케일 팩터 증가 방법을 제안하였다.

제안된 고속화 방법은 주관적 음질 평가를 통해 성능을 검증하였다.

### 1. 서론

인터넷의 보급과 더불어 MP3가 널리 사용되고 MP3 재생기의 보급률 또한 증가하였다. 이로 인해 휴대용 MP3 부호화기의 요구가 대두되고 있고, DSP를 사용한 실시간 구현이 필요하게 되었다. MP3는 CD 수준의 음질을 유지하면서 약 12:1의 압축 성능을 갖는 지각적 오디오 부호화 방식이다. MPEG Audio 부호화 방식은 복호화기보다 부호화기가 훨씬 복잡한 비대칭 구조를 가지

고 있고, 그 중 Layer III 부호화기의 경우 Layer III 복호화기의 3배 이상, Layer I 부호화기의 7.5배 이상의 복잡도를 가진다[1].

MP3 부호화기를 실시간 구현하기 위해선 연산량이 가장 큰 문제가 된다. 복호화에 비해 많은 연산량을 가지는 부호화기는 연산량을 줄일 경우 음질 저하의 문제가 수반된다. 따라서 고음질을 유지하면서 연산량을 줄일 수 있는 알고리즘이 필요하다[2][6].

본 논문에서는 MP3 부호화기의 실시간 구현을 위해 고음질을 유지하면서 연산량을 줄일 수 있도록 심리음향모델의 적용을 간략화 하고 외부 반복 루프의 반복 회수를 줄이는 최적화 방법을 제안한다.

표1.MPEG 부호화기/복호화기의 복잡도

Layer	Complexity	
	Encoder	Decoder
I	1.5-3	1.0
II	2-4	1.25
III	>7.5	2.5

### 2. MPEG AUDIO Layer-III 부호화기

#### 2.1. 부호화 과정

MP3 부호화 과정은 그림 1과 같다. 입력 오디오 신호는 통계적인 중복성을 제거하기 위하여 입력 신호를 32개의 가중 중첩 가산 방식 등간격 필터뱅크에 통과시켜 서브밴드 샘플로 변환하고 변환된 서브밴드 샘플은 다시 MDCT를 수행함으로써 주파수 해상도를 높이는 하이브리드 변환 부호화 방식을 사용한다. 이와 동시에 FFT를 사용하는 심리음향모델을 수행하여 마스킹 임계

치(Masking Threshold) 및 지각 엔트로피(Perceptual Entropy), 블록 형태 등을 결정하게 되고, 이는 다시 반복 루프의 입력으로 들어간다[2][8].

반복 루프는 양자화 잡음을 최소화 하기 위한 외부 루프와 부호화된 비트를 최소화 하는 내부 루프로 나눌 수 있다. 외부 반복 루프에서는 내부 반복 루프에서 수행된 양자화 과정으로 인한 양자화 잡음이 심리음향모델이 요구하는 잡음 변형 모양에 최대한 충족되도록 스케일 팩터 대역을 증폭시키는 일을 수행한다. 모든 스케일 팩터 대역이 정해진 한계만큼 증폭되거나, 양자화 잡음이 허용 한계치 이하가 될 때까지 외부 반복 루프는 반복을 계속 한다. ISO/IEC 11172-3[2]에서는 외부 반복 루프가 한번 수행될 때마다 스케일 팩터를 1씩 증가시킨다. 내부 반복 루프에서는 현재의 프레임에서 사용 가능한 비트수를 최대한 이용하여 부호화하도록 양자화 스텝 크기를 결정하고 허프만 부호화를 수행한다. 이를 위해서 양자화 스텝 크기의 초기치를 선택하고 스펙트럼 데이터가 실제로 양자화되며 이때 부호화된 비트수도 계산한다. 만약 부호화된 비트수가 사용 가능한 비트수를 초과하면 양자화 스텝 크기를 한단계 증가시켜서 부호화된 비트가 사용 가능한 비트 이내가 될 때까지 위의 과정을 반복한다[2][8].

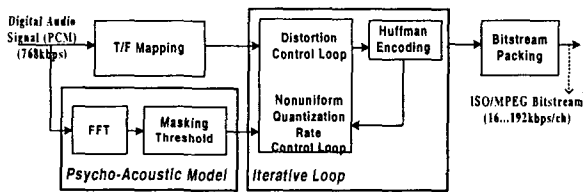


그림. 1. MPEG Audio Layer-III 부호화 과정

## 2.2. 부호화기의 복잡도

· 일반적인 MP3 부호화기의 예측된 연산량은 표2와 같다[3]. ISO/IEC 11172-3을 만족하는 MP3 부호화기의 전체 연산량은 190 MIPS정도로 실시간 구현을 위해서는 연산량을 감소 시킬 필요가 있다. 이중 대부분의 연산량을 차지하는 과정이 심리음향모델-II와 중첩된 반복 루프이다[4].

심리음향모델-II의 경우 FFT, 에너지 계산, 마스킹 임계치 계산 등의 과정을 수행하는 부분으로 연산량의 40% 이상을 차지한다[3]. 특히, 심리음향모델은 DSP 칩에서의 구현에 적합하지 않은 많은 로그 연산과 지수 연산 등을 포함하고 있기 때문에 부호화기의 실시간 구현에 있어서 가장 어려운 장애 요소가 된다[4]. 비트율이 충분히 클 경우에는 심리음향모델을 사용하지 않아도 고음질의 MP3를 얻을 수 있지만 2채널의 160kbps 비트율 이하에서는 심리음향모델을 전혀 사용하지 않을 경우 음질 저하의 문제가 발생한다[3][7]. 따라서 실시간 구현을 위해서는 심리음향모델을 사용하되, 기존의 심리음향모델보다 연산량을 줄일 수 있는 방법이 필요하게 된다.

주어진 비트율에서 최적의 음질을 제공하기 위해 어

떤 스케일 대역을 증폭해야 하는지를 결정하는 방법 중의 하나가 중첩된 반복 루프 과정이다. 심리음향모델의 결과로 얻어지는 마스킹 임계치가 가장 중요한 결정 척도가 되고, 잡음 변형을 위해 개발된 방법 중 가장 우수한 음질을 제공하는 것이 두 개의 중첩된 루프를 사용한 방법이다. 이와 같은 이중 반복 루프 과정은 최상의 음질을 제공하지만 많은 연산량을 요구하고 따라서 외부 반복 루프의 회수가 줄어들면 상대적으로 전체 반복 루프의 회수가 줄어들어 연산량 측면에서 큰 이득을 가져올 수 있다.

표2. MP3 오디오 부호화기의 예측 연산량

MP3 Encoder	MIPS	%
T/F Mapping	25	13.2
PAM-II	90	47.4
Iteration Loop	70	36.8
Etc	5	2.6
<b>Total</b>	<b>190</b>	<b>100</b>

## 3. 부호화기 최적화

### 3.1 심리음향모델의 간략화

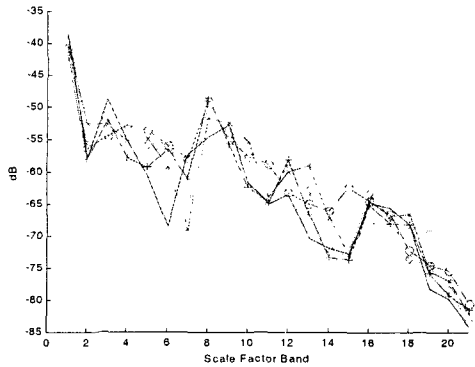
심리음향모델은 MP3부호화기의 음질 및 연산량에 큰 영향을 미치는 부분이다[3]. 심리음향모델을 사용하지 않을 경우 마스킹 임계값은 절대가청한계가 되며, 외부 반복 루프는 양자화 잡음이 절대가청한계 이하가 될 때까지 스케일 팩터를 계속 증가 시키게 된다. 하지만 대부분의 경우 절대가청한계 이하로의 잡음 변형이 이루어지지 못하고 스케일 팩터가 한계치 만큼 증가하여 외부 반복 루프를 빠져나가게 된다. 이는 외부 반복 루프의 반복 회수를 증가시킨다. 또한, 심리음향모델을 수행하지 않을 경우 지각 엔트로피를 구할 수 없고 이로 인해 단구간 윈도우 사용 그래늘을 검출 할 수 없으므로 프리에코 등의 문제로 인한 음질 저하문제가 발생한다[3][8].

실시간 구현을 위해 연산량 측면에서 심리음향모델-II를 간략화 하였다. 귀의 청각 특성 중 하나인 그림 2는 연속하는 6개의 그래늘에 대한 마스킹 임계치 곡선을 나타낸다. 연속하는 그래늘에서의 신호는 유사성을 가지고 있고, 입력신호를 사용하여 구해지는 마스킹 곡선 역시 유사하다. 이 유사성을 이용하면 프레임 내의 첫번째 그래늘에서 구한 마스킹 곡선을 두번째 그래늘에서도 사용할 수 있다. 즉, 심리음향모델을 첫번째 그래늘에서만 연산함으로써 연산량을 줄일 수 있다. 이러한 연속 신호의 유사성을 이용한 방법은 비트 할당 등에도 사용되고 있다[5][6].

실시간 동작을 보장하기 위해 DSP 구현에서는 최대 연산량이 중요한 문제가 된다. 따라서 최대 연산량을 줄이기 위하여 심리음향모델을 두 그래늘에 분산시킨다. 이때, 단구간 윈도우를 사용하는 그래늘이 두번째 그래늘이 될 경우, 단구간 윈도우를 적용할 수 없게 된다는 단점이 있다. 이를 보완하기 위하여 첫번째 그래늘이 단구간 윈도우를 사용할 경우 두 그래늘 모두에서 심리음

향모델을 수행하여 심리음향모델 간략화로 인한 음질 손실을 줄일 수 있다. 단구간 윈도우의 사용 빈도는 5% 이내로 음질의 손상을 감안하고 장구간 윈도우로 고정하여 수행할 경우 최대 인산량을 줄일 수 있다[3].

제안한 방법은 심리음향모델을 사용하지 않는 경우보다 외부 반복 루프의 반복 회수를 줄일 수 있으며, 음질을 향상시킬 수 있다. 또한 심리음향모델을 사용하는 기존의 MP3 부호화기에 비해 큰 음질저하 없이 심리음향모델-II의 연산량을 40%이상 줄일 수 있다.



○ : 1, X : 2, + : 3, △ : 4, □ : 5, 실선 : 6

그림 2. 연속된 6개의 그래늘에서의 마스킹 임계치 곡선 (Masking Threshold Curve)

### 3.2 반복 루프의 최적화

고음질을 유지하면서 연산량을 감소시킬 수 있는 방법으로 외부 반복 루프의 반복 회수를 줄이는 방법을 제안하였다. 외부 반복 루프는, 양자화 잡음이 마스킹 임계치 이하가 되는 조건을 만족시킬 경우, 스케일 팩터를 증가시킬 필요가 없다. 내부 반복 루프의 경우, 부호화된 비트가 사용 가능한 비트 이내이기만 하면, 양자화 스텝을 증가시킬 필요가 없다. 양자화 스텝이 커지면, 양자화 잡음이 커지고, 이는 외부 반복 루프에서 스케일 팩터를 증가시키게 된다. 이러한 과정이 반복되면서 외부 반복 루프는 양자화 잡음을 줄이기 위하여 반복 회수가 계속 증가하면서 스케일 팩터가 커지게 되고, 내부 반복 루프는 부호화된 비트를 줄이고자 반복을 계속하면서 양자화를 증가시키며 양자화 잡음 또한 증가시키게 된다.

그림 3에 의하면 스케일 팩터가 1만큼 증가함으로 인해 감소된 양자화 잡음은, 양자화 스텝이 1만큼 증가함으로 인해 증가하는 양자화 잡음과 같다. 즉 같은 그래늘, 같은 스케일 팩터 대역 내에서의 양자화 잡음은 다음과 같은 관계를 가진다.

$$\text{양자화 잡음} \propto S, S: \text{양자화 스텝-스케일 팩터} \dots (1)$$

따라서 양자화 스텝이 1만큼 증가함으로 인한 양자화 잡음 증가를 해결하기 위해, 이전 내부 반복 루프에 의해 결정된 양자화 스텝보다 현재 양자화 스텝이 클 경우 스케일 팩터를 양자화 스텝 증가 이상으로 증가시킨

다.

한편 스케일 팩터의 증가는 내부 반복 루프에서의 비트가 증가하는 문제가 발생한다. 스케일 팩터 증가에 따른 비트 증가는 다음과 같이 나타낼 수 있다.

$$\text{Scf 1 증가 } \beta \approx \frac{\text{sfb\_width}}{\text{total\_width}} * \alpha \text{ (bits)} \dots (2)$$

$$\text{Scf 2 증가 } 2\beta \approx \frac{\text{sfb\_width}}{\text{total\_width}} * \alpha * 2 \text{ (bits)} \dots (3)$$

$\alpha$  : 양자화 스텝 1 증가시 전체 그래늘에서의 감소 비트수

$\beta$  : 스케일 팩터 1 증가시 해당 스케일 팩터 대역에서의 증가 비트수

Scf : 스케일 팩터, Sfb : 스케일 팩터 대역

그래늘 전체에서의 비트를 고려할 때 스케일 팩터를 2만큼 증가 시킨 경우 증가하는 비트는 양자화 스텝이 1 감소할때 증가하는 비트에 비해 크지 않다. 즉 스케일 팩터를 2만큼 증가시킴으로 인해 증가하는 비트를 줄이고자 양자화 스텝을 증가 시키더라도 스텝 1의 증가면 대부분의 경우 충분하다. 따라서 양자화 잡음이 마스킹 임계치 보다 작은 특정 스케일 팩터 대역에서의 스케일 팩터를 2만큼 증가시킴으로 인해 감소한 양자화 잡음은 양자화 스텝 증가로 인해 증가한 양자화 잡음의 영향을 받지 않는다. 이는 그림 4와 5에서 볼 수 있듯이 전체적인 스케일 팩터의 최대, 평균 값이 작아지고, 상대적으로 양자화 스텝 역시 작아진다. 또한 전체 외부 반복 루프의 반복 회수 역시 줄어든다.

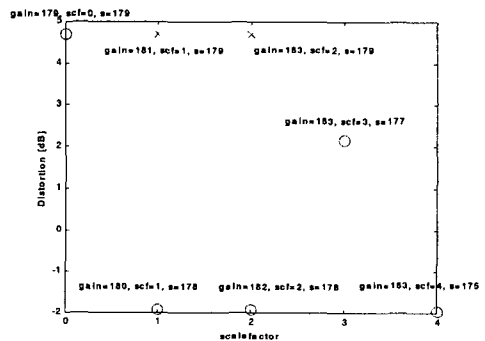


그림 3. 양자화 스텝과 스케일 팩터의 차에 따른 양자화 잡음

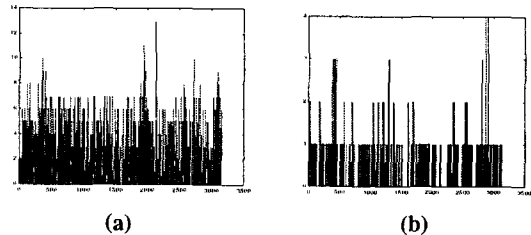


그림 4. 반복 루프 수행 후 최종 결정된 스케일 팩터  
(a) ISO MPEG Audio 부호화기  
(b) 제안된 스케일 팩터 증가 방법을 적용한 부호화기

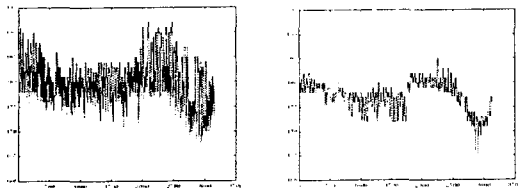


그림 5. 반복 루프 수행 후 최종 결정된 양자화 스텝과 스케일 팩터 대역 0에서의 스케일 팩터와의 차

- (a) ISO MPEG Audio 부호화기  
(b) 제안된 스케일 팩터 증가 방법을 적용한 부호화기

#### 4. 실험 및 결과

2채널 128kbps, 44.1kHz 일때, 스케일 팩터 증가 스텝 조절 및 심리음향모델의 간략화를 적용하여 연산량의 감소를 측정하고 SNR을 살펴보았다. 실험에 쓰인 입력 PCM은 15초 길이의 팝, 재즈, 가요, 클래식, 피아노 솔로 등의 5가지 장르, 10가지 입력 PCM신호를 사용하여 평균을 구하였다.

표3에서 알수 있듯이 외부 반복 루프의 경우 평균 반복 회수가 6.8645회에서 2.3918회로 65%이상 감소하였고, 심리음향모델-II의 경우 50%정도의 연산량이 감소하여, 전체 연산량의 47%정도의 연산량이 줄어들었다.

입력 PCM 신호와 부호화/복호화 과정을 거친 뒤의 PCM 신호간의 SNR을 구한 결과 0.02dB 정도 낮아진다. 표5는 비트율에 따른 부호화기의 신호도 평가 결과이다. 표6의 SNR과 주관적 음질 평가 결과 제안한 알고리즘을 적용한 경우, 기존 부호화기의 신호와 거의 구분되지 않는 좋은 음질을 제공한다는 결과를 얻을 수 있었다.

표3. 비트율에 따른 외부 반복 루프의 평균 반복 회수

	외부 반복 루프 평균 반복 회수		
비트율(kbps)	128	96	64
기존 부호화기	6.8645	7.8892	7.7893
제안된 스케일 팩터 증가 적용	2.3918	2.3597	2.5117

표4. 최적화 과정 수행 후의 부호화기의 연산량 (MIPS) 및 각 블록별 연산량 감소, 전체 연산량 감소

MP3 Encoder	MIPS	연산량 감소(%)
T/F Mapping	25	-
PAM-II	47	48
Iteration Loop	24.4	65
Etc	5	-
Total	101.4	47

표5. 비트율에 따른 기존 부호화기(A)와 제안한 부호화기(B)의 신호도 평가

비트율 (kbps)	A - B 신호도 평가결과		
	A	B	구분안됨
128	2	2	11
96	1	2	12
64	1	0	14

표6. 2채널 128kbps 비트율에서의 SNR 비교

	SNR (dB)
Full Complexity Encoder	17.8874
제안된 스케일 팩터 증가 적용 .. A	18.0137
심리음향모델 간략화 적용 .. B	17.7515
A+B	17.8646

#### 5. 결론

본 논문에서는 2채널 128kbps 비트율 MPEG Audio Layer-III 부호화기의 고음질을 유지하면서 연산량을 줄일 수 있는 알고리즘 최적화 방법을 제안하였다. 모의실험을 통하여 부호화 과정 전체의 연산량이 45% 이상 감소하는 것을 확인하였고, SNR 및 주관적 음질테스트를 통하여 음질 손실이 없음을 확인하였다.

향후, 연산량을 더욱 감소 시키기 위해 비트 할당을 하는 내부 루프를 더욱 빨리 수행할 수 있는 알고리즘을 적용할 수 있고, 심리음향모델-II를 알고리즘 레벨에서 더욱 간략화 할 수 있다[3][5][6].

#### 6. 참고문헌

- [1] <http://www.licensing.philips.com/mpeg/main.html>
- [2] ISO/IEC JTC1/SC29/WG11 No.71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - CD 11172-3(Part3. MPEG-Audio)"
- [3] Hyen-O Oh, Joon-Seok Kim, Chang-Jun Song, Young-Cheol Park and Dae-Hee Youn "Low Power MPEG/AUDIO Encoders Using Simplified Psychoacoustic Model and Fast Bit Allocation," in *Proc. of ICCE*, pp. 320 -321,2001.
- [4] Johannes Hilpert, Michael Braun, Manfred Lutzky, Stefan Geyersberger and Rainer Buchta, "Implementing ISO/MPEG-2 Advanced Audio Coding in Realtime on a Fixed Point DSP," presented at the 105<sup>th</sup> AES Convention, San Francisco 1998,preprint 4822.
- [5] Kai-Tat,Fung,Yui-Lam,Chan and Wan-Chi Siu, "A fast bit allocation algorithm for MPEG audio encoder," in *Proc. International Symposium on Intelligent Multimedia, Video and Speech Processing*, 2001.
- [6] Hyen-O Oh, Sung-Youn Kim, Dae-Hee Youn and Il-Whan Cha, "New Implementation Techniques of a Real-Time MPEG-2 Audio Encoding System," in *Proc. ICASSP*, pp. 1594-1597, Mar. 1999.
- [7] J.D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," *IEEE J. Selected Areas in Commun.*, vol. 6, pp.314-323, Feb.1988
- [8] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for coding of High-Quality Digital Audio," *J. Audio Eng. Soc.*, vol. 42, pp.780-792, Oct. 1994