MPEG-II AAC Encoder의 Perceptual Model에 관한 연구

°구대성, 김정태, 이강현
조선대학교 전자정보통신공학부, 멀티미디어 ASIC설계 실험실
Tel : (062) 230-7066 / Fax : (062) 233-1120
E-mail : {orion, space, khrhee}@vlsi.chosun.ac.kr

A study on the Perceptual Model for MPEG II AAC Encoder

Dae-Sung KU, Jung-Tae KIM, Kang-Hyeon RHEE
School of Electronics, Information and Communication Eng.,
Multimedia ASIC Design Lab., Chosun University

Abstract

Currently, the most important technology is the compression methods in the multimedia society. Audio files are rapidly propagated through internet. MP-3 is offered to CD tone quality in 128Kbps, but 64Kbps below tone quality is abruptly down and high bitrate. on the other hand, MPEG-II AAC (Advanced Audio Coding) is not compatible with MPEG-I, but AAC has a high compression ratio 1.4 better than MP-3. Especially, AAC has max. 7.1 channel and 96KHz sampling rate. In this paper, the perceptual model is dealt with 44.1KHz sampling rate for SMR(Signal to Masking Ratio)

I. 서 론

MPEG-II AAC(Advanced Audio Coding)는 뛰어난 압축율과 고음질의 특성을 갖는 오디오로써, 인간의 청각 특성을 이용한 압축방식으로 마스킹 효과를 이용한다. 심리음향모델 방식에는 모델 1과 모델 2가 있는데, 모델 1은 정확하게 순음과 잡음을 판별하는 특성이 있고, 모델 2는 중고주파수 영역에 존재하는 순음에 대한 순음지수를 비교적 낮게 판별하는 특성이 있다.[1] 그리고 복잡도 면에서도 모델 1보다 모델 2가 높으나, 전반적인 양자화코딩 결과를 비교해 보면 모델 2가 더 뛰어나다. 그러므로 MP3에서도 모델 2방식을 사용하고, AAC 역시 모델 2 방식을 사용한다. Encoder 블록에서 가장 많은 연산을 수행하는 블록은 심리음향 모델과 필터뱅크 블록이다.[2] 본 논문에서는 인간의 청각특성을 이용한 압축방식으로써 가청한계 주파수 이외는 잘라내는 방식을 사용한다. 전반적인 AAC Encoder블록은 그림 1과 같다.

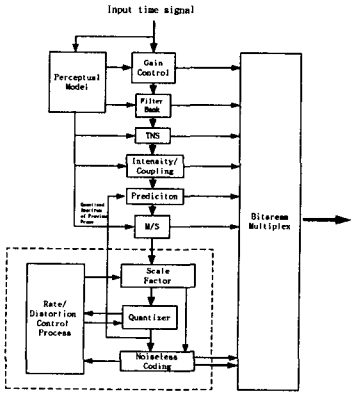

그림 1. MPEG-II AAC Encoder

MPEG-II AAC는 다채널 고음질 오디오 표준으로 AAC 오디오 표준은 5채널이고, 오디오 신호 전 대역에 대해 320Kbit/Sec 데이터 레이트에서 원음과 식별이 불가능하다. AAC의 Profile은 크게 3가지로 나뉜다.

Main profile은 최대 7.1채널을 제공하며, 이보다 간단한 LC(Low Complexity) profile은 2채널로 구성되어 휴대용 오디오에 적합하게 되어 있으며, 이 두 가지 profile은 최대 8KHz-96KHz까지의 샘플링 주파수를 사용한다. SSR(Scalable Sampling Profile) profile은 통신환경에서 낮은 비트율로 네트웍 전송이 가능하도록 6KHz-20KHz까지의 비교적 낮은 샘플링 주파수를 사용한다. MPEG-II AAC는 시스템의 사용목적에 맞게 부호화/복호화 할 수 있게 profile을 사용한다. Main profile은 컴퓨팅 자원의 제한이 없는 환경에서 최고의 압축율과 음질을 갖도록 되어 있고, LC profile은 컴퓨팅 자원이 제한적인 경우에 적합하도록 prediction과 Gain control tool을 사용하지 않으며 TNS필터의 차수도 제한되어 있다. SSR profile은 최소의 비트율을 사용하며 통신환경에서 낮은 비트율로 통신 가능하도록 Gain control 툴을 사용하여 오디오 대역을 강제적으로 제한하며, prediction과 coupling channel을 지원하지 않는다.[3][4]

II. Masking

심리음향의 압축 원리는 청각 특성의 마스킹 현상에 바탕을 두고 있다. 마스킹효과란 특정 신호에 의해서 다른 신호가 가려지는 현상을 말한다. 인간의 청력기관은 입력음성을 각기 다른 특성을 갖는 수많은 필터뱅크에 의해 주파수 분석을 수행한다. 이때 주파수 분석과정에서 인간의 청각 기관이 갖는 해상력의 한계에 의해 마스킹 효과가 발생한다. 마스킹 효과가 일어나는 주파수 폭을 크리티컬 밴드(critical band)라 한다.

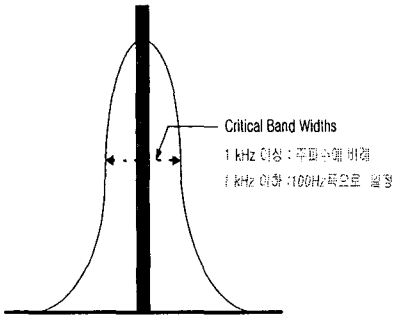

그림 2. 크리티컬 밴드

그림 2의 크리티컬 밴드 폭은 1kHz 이상의 주파수에서는 주파수에 거의 비례하고, 1kHz 이하의 주파수에서는 100Hz의 폭으로 거의 일정하다. 그림 3은 가청한계 곡선과 마스킹의 관계로써, 마스킹의 주체를 마스커, 마스킹의 객체를 마스키라 한다. 마스커와 마스키의 시간적인 차이가 짧으면 마스킹효과가 크게 일어나는데 이러한 효과를 템포럴 마스킹(temporal masking)이라 한다.[5][6]

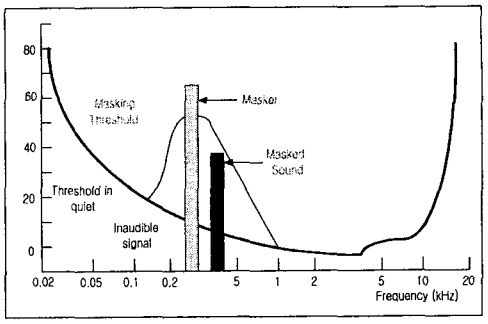

그림 3. 최소한의 가청한계와 마스킹 곡선

III. 심리음향모델

인코딩 연산에서 입력 Wav파일이 심리음향 연산블록을 거쳐 각각의 블록들을 제어하는데 이 과정을 보면 그림 4와 같다.

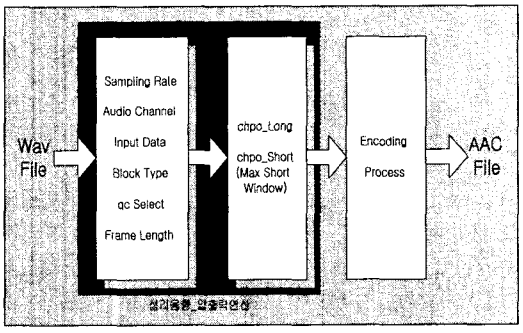

그림 4. 인코딩 연산 과정

AAC의 심리음향 모델은 MPEG-1의 Psychoacoustic model II와 동일한 구조이며 파라미터와 다른 추가사항들이 첨부되어있다. 심리음향모델은 신호에너지에 의해 마스킹되는 최대왜곡에너지를 계산하는데 이 에너지를 Threshold라하고 입력신호를 주파수와 위상으로 분석한다. 입력된 wav파일에서 데이터를 받아들여 심리음향 모델 블록을 거쳐 출력이 되는데 출력은 long블록과 short블록으로 구분된다. 출력의 Long, Short 안의 데이터 값들은 p_ratio, cb_width, no_of_cb로 구성되어있다. 출력 값들은 Gain Control, Filter Bank, TNS, Intensity/Coupling, M/S 모듈에서 제어역할을 수행한다. 심리음향모델의 세부 블록을 보면 그림 5와 같다. 그림 5와 같은 구조를 갖는 심리음향모델 블록은 전체 14단계의 연산과정을 거친다.

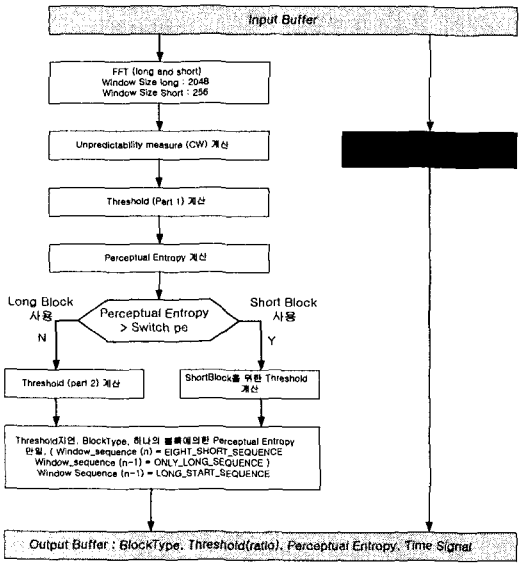

그림 5. 심리음향 모델 블록

3.1 심리음향 모델 14단계.

① 입력신호 샘플의 재구성 (2 · *iblen*)

iblen : Threshold연산 프로세스를 위한 쉬프트 값

② 입력신호의 복잡한 스펙트럼 계산

$sw(i) = s(i) \cdot (0.5 - 0.5 \cdot \cos((\pi \cdot (i+0.5))/iblen)$

$s(i)$: Hann 윈도우 창

③ Predicted 값 계산

$r_{pred}(w) = 2.0 \cdot r(t-1) - r(t-2)$

$f_{pred}(w) = 2.0 \cdot f(t-1) - f(t-2)$

t : 최근 블록의 수.

t-1 : 이전 블록들의 데이터.

t-2 : Threshold계산 블록 전의 데이터.

④ 비 예측성 측정값 계산.

$$c(w) = (((r(w) \cdot \cos(f(w)) - r_{pred}(w) \cdot \cos(f_{pred}(w)))^2 + \left[\frac{(r(w) \cdot \sin(f(w)) - r_{pred}(w) \cdot \sin(f_{pred}(w)))^2)^{0.5}}{(r(w) + abs(r_{pred}(w))}\right]$$

위 식은 Short FFT와 함께 각 Short Block을 위해서 사용된다.

⑤ Threshold 영역 안에서 에너지와 비예측 값 계산.

$e(b) = e(b) + r(w)^2$

$e(b)$: 각 영역 안에서의 에너지

$c(b) = c(b) + r(w)^2 \cdot c(w)$

⑥ 에너지와 비 예측 값의 콘볼루션.

$ecb(b) = ecb(b) + e(bb) \cdot sprdngf(bval(bb), bval(b))$

$ct(b) = ct(b) + c(bb) \cdot sprdgnf(bval(bb), bval(b))$

$$cb(b) = \frac{ct(b)}{ecb(b)}$$

$en(b) = ecb(b) \cdot rnorm(b)$

$tmp(b) = tmp(b) + sprdngf(bval(bb), bval(b))$

$$rnorm(b) = \frac{1}{tmp(b)}$$

⑦ Tonality 인덱스 변환.

$tb(b) = -0.299 - 0.43 \log_e(cb(b))$

$0 \le tb(b) < 1$

⑧ 각 분할영역의 SNR 계산.

NMT(Noise Masking Tone)

NMT(b) = 6dB

$SNR(b) = tb(b) \cdot TMN(b) + (1 - tb(b)) \cdot NMT(b)$

⑨ Power Ratio{bc(b)} 계산.

$$bc(b) = 10^{\frac{-SNR(b)}{10}}$$

⑩ Energy Thrshold{nb(b)} 계산.

$nb(b) = en(b) \cdot bc(b)$

$nb(b)$: M/S 모듈 안에서 사용하고, Xthr의 'X'= [R,L,M,S]와 동일하다.

⑪ Pre_Echo Control과 Threshold

Pre_Echo를 피하기 위해서 Pre_Echo Control은 Short, Long FFT를 계산한다.

$nb(b) = \max(qsthr(b), \min, (nb(b), nb_{l(b)} \cdot rpelev))$

Short블록을 위해서는 '1'로 설정되고, Long블록을 위해서는 '2'로 설정된다.

⑫ PE(Perceptual Entropy) 계산.

$$PE = \frac{e(b)}{nb(b)}$$

$$PE = PE - (W_{high}(b) - W_{low}(b)) \cdot \log 10 \frac{nb(b)}{(e(b)+1)}$$

⑬ Block Type 결정.

엔코더의 코딩은 Pseudo Code이기 때문에 인코딩 과정을 위해 Long, Short블록 형태를 결정해야한다.

⑭ swb안에서의 1/SMR 계산.

swb : 분할코더의 스케일 팩터 밴드.

$$epart(n) = \sum_{W_{low}}^{W_{high}} r(w)^2$$

스펙트럼의 Threshold를 위한 연산은

$$thr(W_{low}(b), .. W_{high}(b)) = \frac{nb(b)}{(W_{high}(b) + - W_{low}(b))}$$

FFT레벨의 스케일 팩터 안의 잡음레벨 $npart$는

$$npart(n) = \min(thr(W_{low}(n)), \ldots, thr(W_{high}(n)) \cdot (W_{high(n)+1} - W_{low}(n))$$

$epart(n)$과 $npart(n)$을 이용하여 SMR을 구할 수

있다.

$$SMR(n) = \frac{epart(n)}{npart(n)}$$

본 논문에서는 입력 wav파일을 심리음향 모델 14단계에서 연산을 수행하고 인코더 블록의 제어신호로 입력되어진다.

Ⅳ. 실험방법 및 고찰

본 논문의 실험에서는 심리음향모델 블록을 C언어로 구현하였고, 사용된 툴은 VC++ v6.0, Matlab v5.3, GNU plot v3.7을 사용하였다. 그림 6은 오리지널 오디오 신호를 나타낸다.

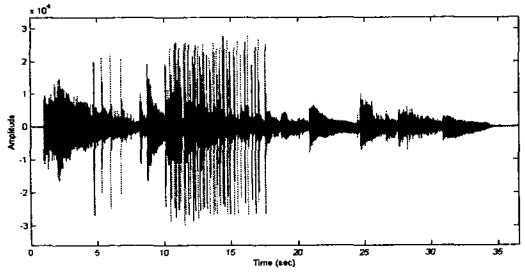

그림 6. "Castanets" 신호

그림 6의 오리지널 'Castanets" 신호를 44.1kHz, mono, 16bits를 이용하여 실험하였다. 그림 7,8은 심리음향모델 과정을 거친 P_ratio의 결과 값이다. 그리고 no_of_cb는 short인 경우 14, Long인 경우 49이고 cb_width는 Short인 경우와 Long인 경우 모두 4로 동일하다.

Ⅴ. 결 론

본 논문에서는 최고의 음질과 압축을 자랑하는 MPEG-II AAC의 심리음향 모델의 동작과 구성에 관하여 알아보았다. 인코더 부분에서 절반 가량의 연산을 차지하는 부분이 심리음향모델이다.

표준안에 철저히 만족하고, 심리음향모델 14단계와 마스킹 현상 및 크리티컬 밴드 등 여러가지 심리음향 특성을 이용하여 Wav파일을 입력으로 만족할만한 결과를 얻었다. 추후과제로서는, 심리음향모델의 연산속도를 높이기 위해 H/W로 설계한다면, AAC Encoder를 FPGA로 구현 시 규모와 속도면에서 최적화를 시킬 수 있으리라 생각한다.

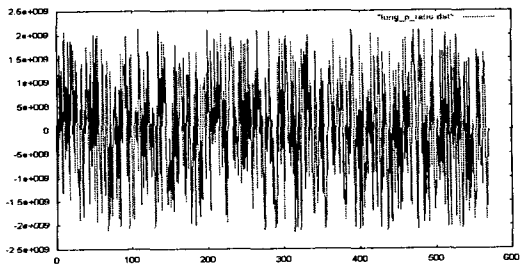

그림 7. Long 출력 값

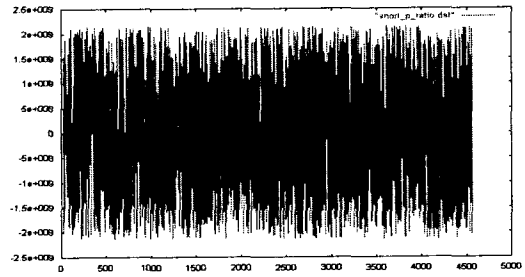

그림 8. Short 출력 값

참 고 문 헌

[1] ISO/IEC 11172-3 Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s Part 3 : Audio.

[2] KEN C. POHLMANN "Principle of Digital Audio", Fourth Edition , McGrow-Hill Book Co, 1999

[3] ISO/IEC 13818-7 Information technology-Generic coding of moving pictures and associated audio information part 7 : Advanced Audio Coding (AAC).

[4] ISO/IEC 14496-3 Information Technology very low Bitrate Audio-Visual Coding Part 3 : Audio

[5] Mark Kahrs, Karlheinz Brandenburg, "APPLICATIONS OF DIGITAL SIGNAL PROCESSING TO AUDIO AND ACOUSTICS", 1998 by Kluwer Academic Publishers.

[6] ITU-R Document TG10-2/3-E only, "Basic Audio Quality Requirements for Digital Audio Bit-Rate Reduction Systems for Broadcast Emission and Primary Distribution", 28 Octorber 1991.