論文2001-38CI-2-5

MPEG-2 AAC Encoder의 심리음향 모델 최적화
(An Optimization on the Psychoacoustic Model for MPEG-2 AAC Encoder)

朴 鍾 泰*, 文 奎 聲*, 李 康 鉉*
(Jong Tae Park, Kyu Sung Moon, and Kang Hyeon Rhee)

요 약

최근 멀티미디어 분야 중에서 가장 중요한 기술 중의 하나가 압축이다. 오디오 파일들은 인터넷을 중심으로 급속히 전파되어가고 있으며, 그 중에서 가장 유명한 것이 MP-3(MPEG-1 Layer3)인데, MP-3는 128Kbps에서 CD음질을 얻을 수 있지만 64Kbps 이하에서는 음질이 급속히 떨어진다. 반면에 MPEG-2 AAC(Advanced Audio Coding)는 MPEG-1과 호환성을 무시하지만 MP-3보다 1.4배의 높은 압축 율을 갖으며, 최대 7.1채널과 96KHz의 샘플 율을 갖는다. 본 논문에서는 MPEG-2 AAC 인코더 부분에서 막대한 연산 량을 갖는 심리음향 모델을 최적화하여 AAC 인코딩 연산 량을 감소시키며 처리속도를 증가하는 알고리즘을 제안한다. 심리음향 모델 최적화 응용 프로그램은 C++언어를 이용하여 구현하였으며, 실험결과 심리음향 모델은 SMR(Signal to Masking Ratio)을 위하여 44.1KHz의 샘플 율을 갖고 2048포인트의 FFT(Fast Fourier Transform)연산을 수행하며, 인코더 블록의 제어를 위하여 서브밴드 필터에 각각의 엔트로피 값들이 입력된다. 제안된 심리음향 모델은 비 예측성 값의 최적화로 인하여 빠른 속도로 수행되었다. 또한 비 예측성 값을 순음지수로 변환 시, 고 주파수 영역의 순음지수 값의 최적화로 연산처리 속도가 증가하였다.

Abstract

Currently, the compression is one of the most important technology in multimedia society. Audio files are rapidly propagated throughout internet. Among them, the most famous one is MP-3(MPEG-1 Layer3) which can obtain CD tone from 128Kbps, but tone quality is abruptly down below 64Kbps. MPEG-Ⅱ AAC(Advanced Audio Coding) is not compatible with MPEG-Ⅰ, but it has high compression of 1.4 times than MP-3, has max. 7.1 and 96KHz sampling rate. In this paper, we propose an algorithm that decreased the capacity of AAC encoding computation but increased the processing speed by optimizing psychoacoustic model which has enormous amount of computation in MPEG-2 AAC encoder. The optimized psychoacoustic model algorithm was implemented by C++ language. The experiment shows that the psychoacoustic model carries out FFT(Fast Fourier Transform) computation of 3048 point with 44.1KHz sampling rate for SMR(Signal to Masking Ratio), and each entropy value is inputted to the subband filters for the control of encoder block. The proposed psychoacoustic model is operated with high speed because of optimization of unpredictable value. Also, when we transform unpredictable value into a tonality index, the speed of operation process is increased by a tonality index optimized in high frequency range.

* 正會員, 朝鮮大學校 電子情報工科大學 電子情報通信工學部
(School of Elec. & Info.-Comm. Eng., Elec.-Info. College, Chosun University)

※ 본 논문을 면밀한 검토와 조언을 해주신 보이지 않는 심사위원들과 CAD툴을 지원해 준 IDEC에 감사드리며, 본 논문은 1999년도 조선대학교 연구비 지원에 의해 작성되었습니다.
接受日字: 2000年 11月1日, 수정완료일 :2001年 3月6日

Ⅰ. 서 론

정보 사회의 지속적인 발전에서 네트워크 통신의 발전에 따른 멀티미디어 시스템의 표준이 MPEG (Moving Picture Expert Group)이다. 표준화 기구 ISO/IEC JTC1/SC29/WC11과 MPEG은 1988년에 설립되어 디지털 비디오와 오디오 코딩의 낮은 데이터 율에 대한 규격화 계획을 발표했다. MPEG은 1992년 첫 번째로 오디오의 사양을 IS 11172-3에 명시하였다.[1] 두 번째로 MPEG-1 오디오와 하위 호환되는 MPEG-2 BC(Backward Compatible)와 MPEG-1의 낮은 샘플링 주파수에 대한 오디오 코딩 표준안을 IS 13818-3에 정의했다. MPEG-2 AAC는 IS 13818-7에서 표준화되었고, 하위 호환되지 않으며 이를 AAC라 명명하였다. 이것은MPEG-1에서 필요로 하는 고음질 오디오 즉, 고음질 멀티 채널 표준을 정립시켰고 또한 ITU-R에서 Full-Bandwidths Channel 오디오 신호와 구별할 수 없는 신호로 인정받았다.[2, 3]

오디오 신호는 2채널인 경우 48KHz, 16bit/sample을 이용하면 1,536Kbps의 데이터가 생성되며, 이를 한 채널 당 128Kbps 6:1로 압축하여 전송하며 이러한 경우 거의 CD 수준의 음질을 얻을 수 있다.[4]

MPEG 오디오 부분에서 가장 널리 알려진 것이 MP-3(MPEG-1 Layer-3)이다. MP-3의 경우 인터넷 기반으로 네티즌을 중심으로 급속히 전파되었다. 반면에, MPEG-2 AAC는 MPEG-1과 호환성을 무시하는 반면, 차후 버전인 MPEG-4, MPEG-7 오디오 포맷과 모두 호환성을 갖기 때문에 차후 대체효과가 크다. 또한 AAC는 MP-3(1/12)에 비해 높은 압축 율(최대 1/20)과 고음질을 채택한 뛰어난 심리음향코딩 방식이다.[5]

MP-3와 AAC의 데이터 구조를 비교하면, MP-3의 데이터 구조가 고정적인 반면 AAC의 데이터는 유동적이다. MP-3는 곡 전체의 정보를 담는 헤더 뒤에 데이터가 프레임 단위로 저장된다. 이 프레임의 크기는 고정되어 있어 압축 율이 높은 부분에서도 쓸모 없는 용량을 차지한다. 이에 반해 AAC의 프레임은 가변 구조로 압축 율에 따라 크기가 변하므로 전체 파일의 용량이 훨씬 줄어든다. 실제로 MP-3 파일과 비교하면 최대 30%까지 용량을 줄일 수 있다. 두 번째로 고음질의 장점은 MP-3에 비해 AAC는 TNS(Temporal Noise Shaping)와 Prediction이라는 두 가지 기법을 통해 음질을 향상한다.[6,7]

고음질과 고 압축 율을 갖는 AAC 오디오지만, 가장 큰 문제점으로 대두되는 것이 인코딩 시간이 너무 길다는 것이다. 인코딩 시간은 심리음향 모델과 직접적인 연관성을 갖고 있다. 그래서 심리음향 모델의 최적화 및 하드웨어 설계가 큰 문제로 대두되었다

본 논문에서는 이를 해결하기 위하여 2채널 방식의 AAC 인코더의 심리음향모델을 최적화 및 하드웨어를 설계한다. 제안된 알고리즘은 NMT(Noise Masking Tone)를 6dB로 설정하고, 심리음향모델 14단계에서 비 예측성 값을 최적화시켜서 SNR(Signal to Masking Ratio)의 효율을 향상시킴으로 써 전체 마스킹 효율을 개선시킨다.

본 논문의 구성은 Ⅱ장에서 심리음향의 이론적 배경을 설명하고, Ⅲ장에서는 심리음향 모델의 최적화 알고리즘을 제안한다. 그리고 Ⅳ장에서 제안된 알고리즘의 실험과 결과 고찰을 통하여 Ⅴ장에서 결론을 맺는다.

Ⅱ. 심리음향의 이론적 배경

인간의 귀가 인지할 수 있는 신호의 주파수는 20Hz부터 20KHz까지라고 알려져 있다. 이때 음성의 경우 40~7,000Hz의 주파수를 가지고 있다. 오디오 신호는 40~15,000Hz정도로 표현되며 Dynamic Range는 20~95dB이다. 사람의 귀는 어떤 특정 주파수에 따라 듣는 특성이 달라지며, 신호의 크기인 Dynamic Range가 일정 수준 이상에 이를 때까지 그 소리를 듣지 못하는데 이를 절대 가청 주파수(Absolute Threshold)라고 한다. 이 주파수는 낮은 영역과 높은 영역에서 큰 값을 갖고 중간 대역(1KHz~5KHz)에서 낮은 값을 갖는다. 이것은 중간 대역에 음성의 중요한 정보가 들어있다는 것과 일치한다. 청각 시스템은 15KHz 이상의 고주파는 소리의 크기가 커야 들을 수 있고 100Hz 이하의 저주파 역시 일정 한도의 소리 크기가 보장되어야만 들을 수 있다. 이러한 사람의 귀의 특성을 이용한 오디오 압축 방식을 심리음향 압축 방식이라 한다.[8]

MPEG-2 AAC 인코더에서 가장 많은 연산을 수행하는 심리음향 모델 블록은 전체 연산 량의 50%를 차지한다. 심리음향 모델 방식에는 모델 1과 모델 2를 사용하는데, 모델 1은 정확하게 순음과 잡음을 판별하

는 특성이 있고, 모델 2는 중고주파 영역에 존재하는 순음에 대한 순음지수를 비교적 낮게 판별하는 특성이 있다. 그리고 복잡도 면에서도 모델 1보다 모델 2가 높으나 전반적인 양자화 코딩 결과를 비교해 보면 모델 2가 더 뛰어나다. 그러므로 MP-3, AAC 모두 모델 2 방식을 사용하지만 MP-3보다 AAC에서 복잡도가 높다

1. MPEG-2 ACC프로필

MPEG-2 AAC Profile은 크게 3가지로 나뉜다. 주 프로화일(Main Profile)은 최대 7.1채널을 제공하며, 이보다 간단한 LC(Low Complexity) profile은 2채널로 구성되어 휴대용 오디오에 적합하게 되어 있으며, 이 두 가지 프로필은 최대 8KHz-96KHz까지의 샘플링 주파수를 사용한다. SSR(Scalable Sampling Profile) 프로필은 통신환경에서 낮은 비트 율로 네트워 전송이 가능하도록 6KHz-20KHz까지의 비교적 낮은 샘플링 주파수를 사용한다. MPEG-2 AAC는 시스템의 사용목적에 맞게 부호화/복호화 할 수 있게 프로필 주파수를 사용한다.

그림 1은 각 프로필간의 상호작용을 나타낸다

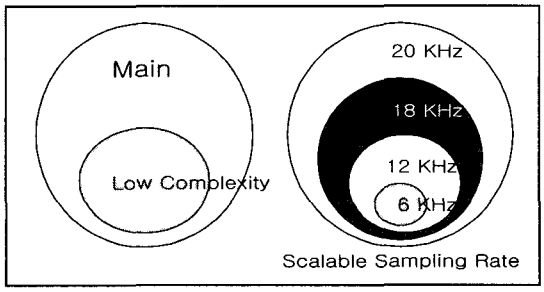

그림 1. AAC 프로화일의 상호작용
Fig. 1. Interoperability of AAC profile.

즉 프로필은 컴퓨팅 자원의 제한이 없는 환경에서 최고의 압축 율과 음질을 갖도록 되어 있고, LC 프로필은 컴퓨팅 자원이 제한적인 경우에 적합하도록 prediction과 Gain control 툴을 사용하지 않으며 TNS필터의 차수도 제한되어 있다. SSR 프로필은 최소의 비트 율을 사용하며 통신환경에서 낮은 비트 율로 통신 가능하도록 Gain control 툴을 사용하여 오디오 대역을 강제적으로 제한하며, prediction과 coupling channel을 지원하지 않는다.[9,10]

2. 마스킹 효과

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 분석 과정에서 인간의 청각 기관이 갖는 해상력의 한계에 의해 마스킹 효과가 발생한다.[11] 마스킹 효과가 일어나는 주파수 폭을 임계적 밴드(critical band)라 한다. 그림 2의 임계적 밴드 특성을 보면 밴드 폭은 1KHz 이상의 주파수에서는 주파수에 거의 비례하고, 1KHz 미만의 주파수에서는 100Hz의 폭으로 거의 일정하다.

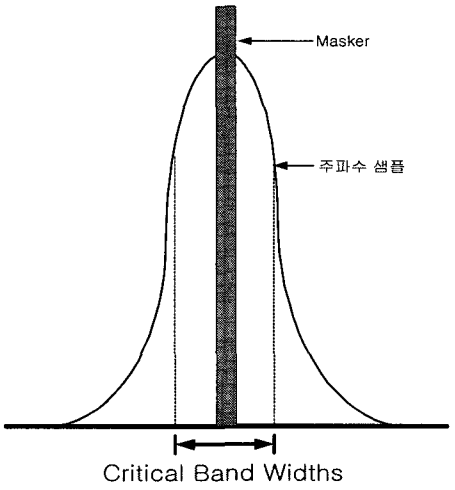

그림 2. 임계적 밴드
Fig. 2. Critical band.

그림 3은 가청한계곡선과 마스킹의 관계로써 마스킹의 주체를 마스커(Masker), 마스킹의 객체를 마스키(Maskee)라 한다. 마스커와 마스키의 시간적인 차이가 짧으면 마스킹 효과가 크게 일어나는데 이러한 효과를 시간적인 마스킹(Temporal Masking)이라 한다.[12]

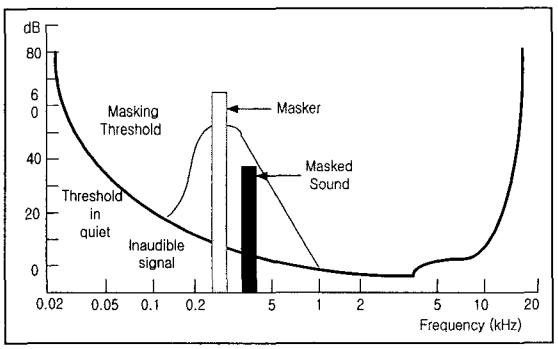

그림 3. 최소한의 가청한계와 마스킹 곡선
Fig. 3. Minimun threshoid and masking curve.

심리음향 모델에서 최종적으로 얻어지는 값은 SMR(Signal to Masking Ratio)이다. 그림 4에서 보면 최소한의 가청 한계 기준 이하를 NMR (Noise to Masking Ratio) 기준치 이상을 SMR(Signal to Masking Ratio)이라 한다. SMR 산출 식은 식 (1)과 같다.

$$SMR = SNR - NMR \quad (1)$$

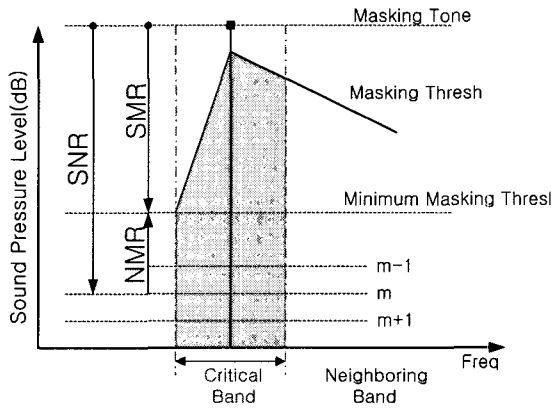

그림 4. 마스킹의 구조적 표현
Fig. 4. The structural presentation of the masking.

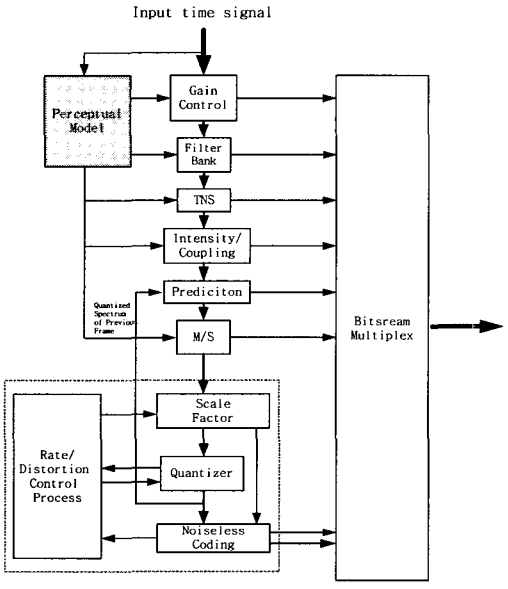

그림 5. MPEG-2 AAC 인코더
Fig. 5. MPEG-II AAC encoder.

3. AAC Encoder와 확산 함수

그림 5는 MPEG-2 AAC인코더의 전체 블록을 나타낸다. 데이터의 흐름은 위에서 아래로, 제어선의 흐름은 좌에서 우로 이동한다.[13]

AAC 인코더의 심리음향 모델 블록에서 필수 적으로 사용되는 함수가 확산 함수(Spreading Function)이다. 확산 함수가 기본 구조는 다음과 같다.

확산 함수에는 임시적으로 선언되는 3개의 변수가 정의되어 있다. *tmpx*는 시간적인 변수를 나타내고, *tmpy*와 식 (2) 와 같고, *tmpz*는 식 (3)과 같다.

```
if j >= I
   tmpx = 3.0(j-I)
else
   tmpx = 1.5(j-I)
```

tmpx = Temporary 변수.
I = 확산 신호의 Bark 값.
j = 확산 밴드의 Bark 값.

$$tmpy = 15.811389 + 7.5(tmpx + 0.474) - 17.5(1.0 + (tmpx + 0.474)^2)^{0.5} \quad (2)$$

$$tmpz = 8 \cdot ((tmpx - 0.5)^2 - 2(tmpx - 0.5), 0) \quad (3)$$

식 (2), (3)에서 정의된 변수를 이용하여 다음과 같이 확산함수를 구한다.

```
if (tmpy ≤ -100) then
   sprdngf(i, j) = 0
else
   sprdngf(i, j) = 10^((tmpz + tmpy)/10)
```

Ⅲ. 심리음향 모델의 최적화

AAC의 심리음향 모델은 MPEG-1의 심리음향 모델 2와 동일한 구조를 갖고있으나, 파라미터와 다른 추가 사항들이 첨부되어있다. 심리음향 모델은 신호에너지에 의해 마스킹 되는 최대 왜곡에너지를 계산하는데, 이 에너지가 드레숄드이고 입력신호를 주파수와 위상으로 분석한다. 입력된 웨이브 파일에서 데이터를 받아들여 심리음향 모델 블록을 거쳐 출력이 되는데 출력은 Long블록과 Short블록으로 구분된다.

출력의 Long, Short 안의 데이터 값들은 채널 대역폭, 채널 수, 에너지 밀도로 구성되어있다. 출력 값들은 Gain Control, 필터 뱅크, TNS, Intensity/Coupling,

M/S 모듈에서 제어역할을 수행한다. 심리음향 모델의 세부 블록을 보면 그림 6과 같다. 이와 같은 구조를 갖는 심리음향 모델 블록은 전체 14단계의 연산과정을 거친다. 첫째 단의 FFT블록에서 방대한 연산이 수행되고, 비 예측성 값을 측정한 이후에 심리음향 엔트로피 값을 구한다. 심리음향 엔트로피 값을 기준으로 하여 윈도우 형태를 결정한다. 그리고 입력버퍼에서 바로 출력버퍼로 지연보상을 해주는 블록은 MDCT 필터뱅크에서 수행한다

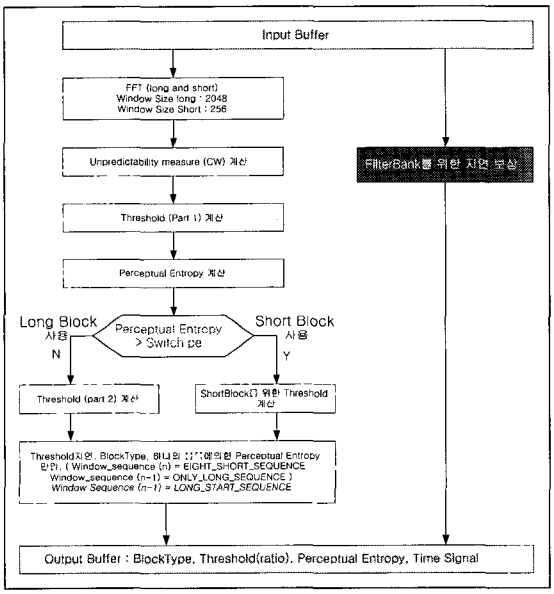

그림 6. 심리음향 모델의 블록도
Fig. 6. Block diagram of psychoacoustic model.

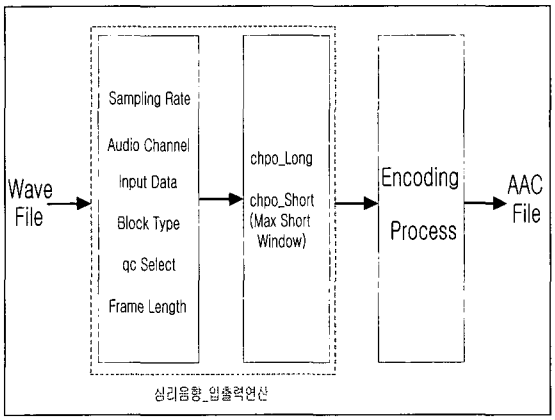

그림 7. 제안된 심리음향 모델의 최적화.
Fig. 7. Optimization of the psychoacoustic model proposed.

본 논문에서는 압축 율이 뛰어나고 고음질의 주 프로파일 방식을 사용하였다. 그리고 제안한 심리음향의 구조적 표현은 그림 7과 같다. 입력으로 웨이브 파일을 받아서 최소한의 가청한계 이하는 잘라내는 방식으로 압축을 하는데, 실제 연산을 수행하는 블록은 전체 인코딩 연산 중 점선으로 표시된 부분에서 수행한다.

■ 심리음향 모델 14단계

본 논문에서 디지털 오디오파일은 제안된 <심리음향 모델의 최적화> 14단계에서 연산을 수행하고 인코더 블록의 제어신호로 입력된다. [Step 1]의 *iblen*은 프로세서에서 쉬프트 값을 나타내고, 사용되는 윈도우는 Hann 윈도우를 사용한다.

심리음향 모델의 최적화는 [Step 4]의 비 예측성 부분이다. 예측성 값은 두 개의 드레숄드 계산과정에서 연산된다. 이후의 비 예측성 값 연산으로 Short FFT는 Short 블록을 위해 사용되는데 예측성 값이 저장상태라면 스펙트럼을 0.4로 초기화 시킨다. 비 예측성 값을 순음지수로 변환 시, 순음지수의 범위는 (0≤순음지수≤1)로써 고 주파수 영역에 존재하는 순음지수 값을 최적화하여 연산처리 속도를 높인다.

■ 심리음향 모델의 최적화

[Step 1] 입력신호 샘플의 재구성 ($2 \cdot iblen$)

iblen : Threshold연산 프로세스를 위한 쉬프트 값.

[Step 2] 입력신호의 복잡한 스펙트럼 계산

$$sw(i)=s(i)\cdot(0.5-0.5\cdot\cos\pi\cdot\frac{(i+0.5)}{iblen}) \quad (4)$$

$s(i)$: Hann 윈도우 창

[Step 3] 예측성값 계산

$$r_{pred}(w)=2.0\cdot r(t-1)-r(t-2) \quad (5)$$

$$f_{pred}(w)=2.0\cdot f(t-1)-f(t-2) \quad (6)$$

t : 최근 블록의 수.
t-1 : 이전 블록들의 데이터.
t-2 : Threshold계산 블록 전의 데이터.

[Step 4] 비 예측성 값($C(w)$) 계산.

$$c(w)=(((r(w)\cdot\cos(f(w))-r_{pred}(w)\cdot\cos(f_{pred}(w)))^2+\left[\frac{(r(w)\cdot\sin(f(w))-r_{pred}(w)\cdot\sin(f_{pred}(w)))^2)^{0.5}}{(r(w)+abs(r_{pred}(w))}\right] \quad (7)$$

식 (7)은 Short FFT와 함께 각 Short Block을 위해서 사용된다

[Step 5] Threshold 영역 안에서 에너지와 비 예측성 값 계산.

$$e(b)=e(b)+r(w)^2 \quad (8)$$

$e(b)$: 각 영역 안에서의 에너지

$$c(b)=c(b)+r(w)^2\cdot c(w) \quad (9)$$

[Step 6] 에너지와 비예측성값의 콘볼루션.

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

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

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

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

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

$$rnorm(b)=\frac{1}{tmp(b)} \quad (10)$$

[Step 7] Tonality 인덱스 변환.

$$tb(b)=-0.299-0.43\log_e(cb(b)) \quad (11)$$

$$0<=tb(b)<1$$

[Step 8] 각 분할영역의 SNR 계산.

NMT(Noise Masking Tone)

NMT(b) = 6dB

$$SNR(b)=tb(b)\cdot TMN(b)+(1-tb(b))\cdot NMT(b) \quad (12)$$

[Step 9] 에너지 밀도 {bc(b)} 계산.

$$bc(b)=10^{\frac{-SNR(b)}{10}} \quad (13)$$

[Step 10] Energy Threshold {nb(b)} 계산.

$$nb(b)=en(b)\cdot bc(b) \quad (14)$$

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

[Step 11] Pre_Echo Control과 Threshold

Pre_Echo를 피하기 위해서 Pre_Echo Control은 Short, Long FFT를 계산 한다.

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

Short블록을 위해서는 '1'로 설정되고, Long블록을 위해서는 '2'로 설정 된다.

[Step 12] PE(Perceptual Entropy) 계산.

$$PE=\frac{e(b)}{nb(b)} \quad (15)$$

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

[Step 13] Block Type 결정.

인코더의 코딩은 Long, Short블록의 형태를 결정한다.

[Step 14] swb안에서의 1/SMR 계산.

swb : 분할코더의 스케일 팩터 밴드

$$epart(n)=\sum_{W_{low}}^{W_{high}}r(w)^2 \quad (17)$$

스펙트럼의 Threshold를 위한 연산은

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

FFT레벨의 스케일 펙터 안의 잡음레벨 $npart$는

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

$epart(n)$와 $npart(n)$을 이용하여 SMR을 구한다.

$$SMR(n)=\frac{epart(n)}{npart(n)} \quad (18)$$

Ⅳ. 실험결과 및 고찰

본 논문은 MPEG-2 AAC 인코더의 속도 향상을 위하여 심리음향 모델의 최적화를 제한하였다. 심리음향 모델 블록을 C++언어를 사용하여 응용프로그램을 구현하였다. 그림 8은 실험에 사용된 "Castanets" 원 오디오 신호로써 44.1KHz, Mono, 16bit이다.

그림 9는 제안된 알고리즘을 사용하여 구현된 응용프로그램으로써 원 오디오 파일에 관한 데이터를 Short과 Long 블록으로 나누어 표 1의 3가지 데이터 값을 출력한다. 실험에 사용된 오디오 파일은 572 Kbyte의 "Castanets" 신호로써 Pentium-Ⅲ 650 MHz, 256M 베모리 환경에서 실험하였다.

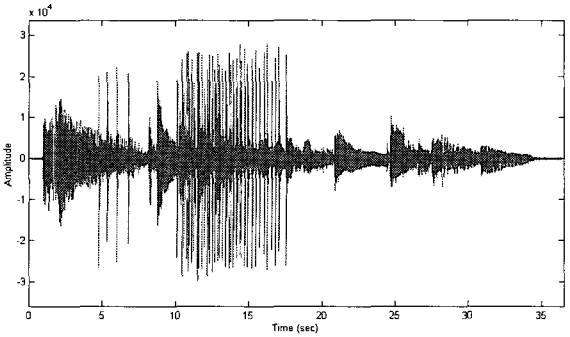

그림 8. 원 "Castanets" 신호
Fig. 8. Original "Castanets" signal.

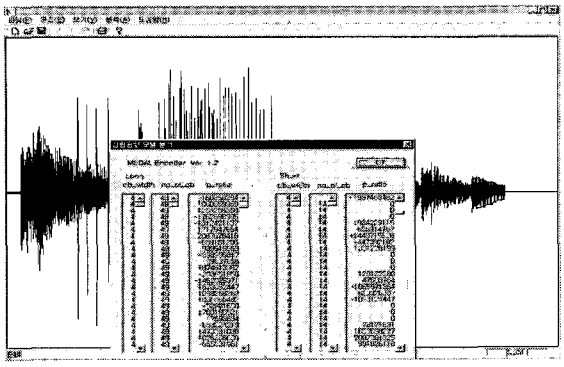

그림 9. 심리음향 모델 응용프로그램
Fig. 9. Application program of psychoacoustic model.

표 1은 심리음향 모델에서 "Castanets" 신호를 입력으로 받아 연산을 수행하여 출력된 에너지 밀도, 채널 수, 채널 대역폭을 나타낸다. 채널 대역폭은 44.1KHz, Short블록과 Long블록을 위한 FFT연산의 파라미터 값이고 4로써 일정한 출력 값을 갖는다. 채널 수는 Short과 Long 블록의 형태에 따라 다른 출력 값을 갖는다.

표 1. 심리음향 모델의 출력 값
Table 1. The results of psychoacoustic model.

윈도우범위 / 출력값	Short	Long
채널 대역폭	4	4
채널의 수	14	49
에너지 밀도	연속적인 출력	이산적인 출력

연산 속도에 가장 큰 영향을 미치는 에너지 밀도로써 Hann 윈도우 영향을 받아 하나의 Long블록은 8개의 Short블록과 같다. Short인 경우 윈도우 형태가 8개로 분할된 상태이기 때문에 연속적인 출력을 갖고, Long의 경우 하나의 윈도우 형태로 5개씩 이산적으로 출력된다.

표 2. 심리음향 모델 비교표
Table 2. Comparison table of the conventional and the proposed psychoacoustic model.

모델 / 점유율	기존 심리음향 모델	제안된 심리음향 모델
CPU 점유율	70%	62%
메모리 점유율	10%	8%
처리 속도	22 Sec.	16 Sec.

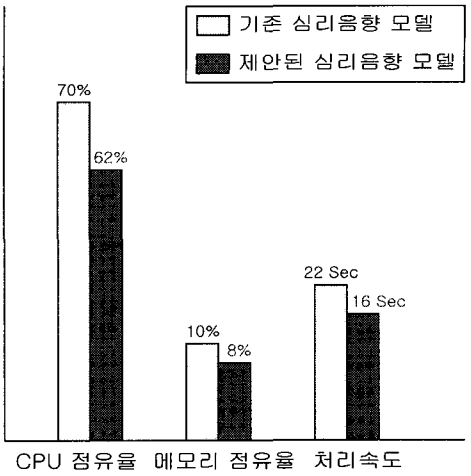

그림 10. 심리음향 모델의 비교 그래프
Fig. 10. Comparison graph of the conventional and the proposed psychoacoustic model.

표 2는 기존 방식의 심리음향 모델(그림 6)과 제안된 방식의 모델(그림 7)을 수행한 비교 결과 값이고, 비교 그래프는 그림 10과 같다.

심리음향 모델에서 가장 중요한 출력 값은 에너지 밀도로써 Gain control, 필터뱅크, TNS, Intensity/coupling, M/S 블록에서 제어역할을 수행한다. 그림 11, 12는 심리음향모델 최적화 과정을 거친 에너지 밀도의 결과 값이다. 표 1과 같이 Short인 경우 주기적으로 연속적인 출력을 갖고, Long인 경우 이산 적으로 출력된다.

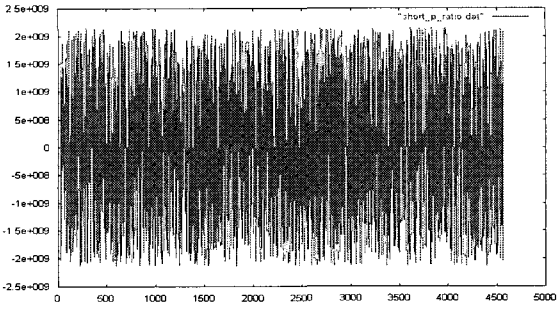

그림 11. Short 출력 값
Fig. 11. The result of Short output.

본 논문의 심리음향 모델은 한 프레임에서 전 후반 2회의 SMR을 산출해서 최대 값을 사용한다. 비 예측성값을 순음지수로 변환 시, 순음지수의 범위는 0~1구간이지만, 본 논문에서는 순음지수 스펙트럼을 0.4로

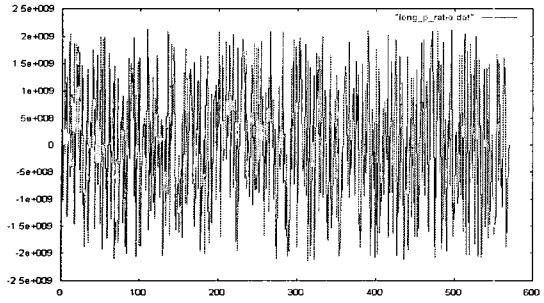

그림 12. Long 출력 값
Fig. 12. The result of Long output.

초기화시켜서 고 주파수 영역에 존재하는 순음 지수 값이 최적화되어 연산처리 속도가 높아졌다.

Ⅵ. 결 론

MPEG-2 AAC 인코더 부분에서 절반 가량의 연산량을 차지하는 부분이 심리음향 모델이다. 인코딩 속도를 향상시키기 위해서는 심리음향 모델의 속도를 향상시키는 것이 필수적이다. 본 논문에서는 고음질과 압축율이 높은 심리음향 모델의 비 예측성 값에서 표준안을 만족하는 범위 (0≤순음지수≤1)를 0.4로 초기화시키고 순음 지수 값을 최적화하였다.

따라서 심리음향 모델 14단계와 마스킹 현상 및 임계적 밴드 등 심리음향 모델의 최적화로 AAC 오디오 파일의 채널 대역폭, 채널 수, 에너지 밀도의 출력 값 계산의 연산시간을 단축하였으며, 심리음향 엔트로피에서 블록형태의 결정 시 처리속도도 증가하였다.

심리음향 블록의 선 마스킹 (Pre Masking)은 5ms 이내에서 발생하였고, 후 마스킹(Post Masking)은 50~300ms로 본 논문의 심리음향 모델은 DIS 13818-7 표준안을 만족한다.

본 논문의 결과는 추후, 심리음향 모델의 연산 속도를 높이기 위해 FFT 블록이나 심리음향 블록을 H/W로 설계 시, AAC 인코더의 ASIC 구현에서 인코딩 처리속도를 증가하는데 응용할 수 있다.

참 고 문 헌

[1] 구대성, 김병규, 김윤홍, 신광규, 이강현 : "MPEG 오디오를 위한 분산 연산방식의 합성 필터뱅크 설계" 대한전자공학회 IMMC'99, pp. 21~25, 1999

[2] 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.

[3] J. Princen, A.Johnson, A.Bradley, "Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Alias ing Cancellation," Proc. of the ICASSP, pp. 2161-2164, 1987.

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

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

[6] Mark Kahrs, Karlheinz Brandenburg, "APPLICATIONS OF DIGITAL SIGNAL PROCESSING TO AUDIO AND ACOUSTICS," Kluwer Academic publishers, 1988.

[7] 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 Oct. 1991.

[8] M. Bosi, "Overview of MPEG Audio : Current and Future Standards for Low-Bit-Rate Audio Coding," J. AES, Vol. 45, No 1/2, pp. 4-21, Jan/Feb. 1997.

[9] M. Bosi, "ISO/IEC MPEG-2 Advanced Audio Coding," J. AES, Vol. 45, No. 10, pp. 789-814, Oct. 1997.

[10] D. Pan, "A Tutorial on MPEG/Audio Compression," IEEE Trans. on Multimedia, Vol. 2, No. 2, pp. 60-74, 1995.

[11] ATSC, "Digital Audio Compression Standard (AC-3)," Dec. 1995.

[12] ALEXANDER D. POULARIKAS, "The Transforms and Applications Hand Book," CRC and IEEE Press, 1996.

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

저 자 소 개

朴 鍾 泰(正會員)

1970년 3월 22일생. 1996년 2월 25일: 조선대학교 공과대학 전자공학과 공학사. 1998년 2월 25일 : 조선대학교 전자공학과 공학석사. 1998년 3월 1일~현재 : 조선대학교 대학원 박사과정. 1998년 3월~현재 : 목포과학대학 겸임교수 및 동강대학 강사. 관심분야: 멀티미디어 비트스트림 처리 및 VLSI 구조설계

李 康 鉉(正會員)

1953년 7월 12일생. 1977년 2월 : 조선대학교 전자공학과 공학사. 1981년 8월 : 조선대학교 대학원 전자공학과 공학석사. 1991년 2월 : 아주대학교 대학원 전자공학과 공학박사. 1999년 5월 : 멀티미디어 기술사. 1991, 1994년 : 미 스탠포드대 CRC 협동연구원. 1996년 12월부터 1997년 3월: 호주 시드니대학 및 울릉공대학 객원 교수. 1999년 5월부터 현재: 멀티미디어 기술사협회 이사. 2001년도 : IEEE 호남세션 부의장. 2001년도 : 한국인터넷 방송/TV 학회 부회장. 2001년도 : 대한전자공학회 멀티미디어 연구회 전문위원장. 1977년부터 현재: 조선대학교 전자정보통신공학부 교수. 관심분야: VLSI 설계, 멀티미디어 ASIC 설계 및 디지틀 콘텐츠, Biochip 설계

文 奎 聲(正會員)

1949년 7월 3일생. 1971년 2월 25일 : 조선대학교 공과대학 전기공학과 공학사. 1986년 8월 25일 : 조선대학교 전자공학과 공학석사. 1976년 3월 1일~현재 : 숭신공고 재직 및 송원대학 강사. 1999년 9월~현재 : 조선대학교 대학원 박사과정. 관심분야: 멀티미디어 신호처리 및 바이오공학