The Journal of Engineering
Paichai University, Korea
Vol.2, No.1 : 57 ~ 62, 1997

가변 전송율 MPEG 오디오

남 승 현
배재대학교 컴퓨터전자정보통신공학부

Variable Bitrate MPEG Audio

Seung Hyon Nam
Division of Computer, Electronics, and Information Engineering, Paichai University

MPEG-1에서 사용하고 있는 두가지 심리음향모델은 서로 다른 매스킹 패턴과 매스킹 인덱스 그리고 매스킹 레벨의 산출 과정을 거친다. 결과적으로 모델 1은 우수한 순음/잡음 판별로 인하여 정확한 매스킹 레벨을 산출하지만 SMR 산출에 worst case를 가정하고 오디오 신호의 동적인 상태를 무시하기 때문에 모델 2보다 저하된 성능을 보여주는 것으로 밝혀졌다. 본 연구에서는 고정 전송율로 설계된 MPEG-1 오디오를 가변 전송율로 변환하였을 때 심리음향모델 1과 2의 성능이 어떻게 나타나는지에 대해 알아보았다. 모의 실험 결과 모델 2는 모델 1에 비해 dual channel 모드에서 평균 30 kbps, joint stereo 모드에서 평균 20 kbps 정도 이득을 제공하는 것으로 나타났다. 일반적으로 joint stereo 모드는 dual channel 모드에 비해 많은 이득을 제공하는 것으로 알려져 있는데, 이러한 현상은 attack이 많은 오디오 신호의 경우 모델 1에서 더 심하게 나타남을 알 수 있다. 이는 모델 1이 pre-echo를 줄이기위해 각 채널에서 과도하게 SMR을 산출하기 때문이다.

Two psychoacoustic models used in MPEG-1 employ different masking patterns, different masking indexes, and different computational procedures. As a result, Model 1 is inferior to Model 2 due to its worst case approach in computing the SMR even though it determines tonality and masking levels accurately. In this study, we investigate the performances of psychoacoustic models when we modify the MPEG-1 audio coder for variable bitrates. Simulation results show that Model 2 has a gain of 30 kbps in the dual channel mode and 20 kbps in the joint stereo mode. It is generally known that the joint stereo mode has a gain in bitrate compare to the dual channel mode. For signals with frequent attacks, this gain becomes larger in Model 1 than in Model 2. This is due to the fact that Model 1 uses the worst case approach in computing the SMR to reduce pre-echo

Key words : Audio, Coder, Psychoacoustic, MPEG, Masking

I. 서 론

80년대 초 디지털 오디오 CD가 시장에 나온 이후, 지난 10여년간 오디오 코딩 알고리즘은 90년초에 표준화된 MPEG-1 오디오에서 MPEG-2 오디오를 거쳐 현재 표준화가 진행 중인 MPEG-4 오디오에 이르기까지 많은 발전을 거듭해왔다. MPEG 오디오 코딩 알고리즘은 디지털 위성방송으로부터 인터넷을 통한 음악의 배포에 이르기까지 폭 넓게 우리의 일상에 응용되고 있다[1]. MPEG의 활용은 조만간 실시될 국내 DAB(Digital Audio Broadcasting)와 인터넷 사용의 확장으로 더욱 가속화될 전망이다.

MPEG-1 오디오는 사람의 청각 특성인 매스킹 현상을 이용하여 오디오 신호를 양자화할 때 필연적으로 발생하는 양자화 잡음이 들리지 않도록 은폐함으로써 오디오 데이터를 6분의 일로 압축했다[2]. MPEG-1 오디오는 이후 다중채널(5+1

채널)이 가능하며 MPEG-1 오디오와 호환이 가능한 MPEG-2 오디오로 발전되었다[3]. 그러나 최근 MPEG-1 오디오와 호환되지 않는 MPEG-2 NBC(Nonbackward Compatible Coder) 오디오 표준이 마련되었고 이는 MPEG-2 AAC(Advanced Audio Code)로 정식 개명되기에 이르렀다. 이러한 일련의 MPEG 오디오 표준들은 일반적인 오디오 신호의 고품질 압축을 목표로 하고 있으며 주로 방송이나 저장 매체에 응용되고 있다[1].

MPEG-4 오디오는 기존의 오디오 표준과는 달리 빠르게 발전하며 팽창해가는 멀티미디어 산업 전반에 활용하기 위한 표준으로 설정되었다. 현재의 멀티미디어 산업은 이동통신, 상호 작용적인 컴퓨터 산업, 다양한 오디오 데이터의 활용 등으로 그 성격을 규정할 수 있다. MPEG-4 오디오의 표준화는 이들 각 영역의 급속한 발전과 상호간의 통합적인 진화를 염두에 두고 시작되었다. MPEG-4 오디오가 지녀야 하는 특성들은 따라서 높은 압축효율, 다양한 접근성(Accessibility), 상호 작용성(Interactivity), 그리고 새로운 기술 개발을 포용하기 위한 확장성과 유연성 등이다.

이러한 오디오 부호화 기술의 발전에도 불구하고 이미 널리 사용되고 있는 MPEG-1 오디오 알고리즘을 약간 개조하여 성능의 향상을 꾀하는 것은 매우 의미있는 일이다[4]. MPEG-1 Layer I과 II는 원래 고정 전송율로 개발되었기 때문에 오디오 신호의 특성이 다양하게 변화하는 점을 충분히 충족시키지 못하는 단점이 있고 자연히 요구되는 전송율은 과도하게 높게 마련이다. 이러한 점을 보완하기 위해 Layer III에서는 비트 저수지(bit reservoir) 방식을 사용하여 평균적인 전송율을 일정하게 유지하면서도 오디오 신호의 특성에 따라 비트를 효율적으로 사용한다.

본 논문은 MPEG-1 Layer II에서 인코더를 가변전송율이 가능하도록 개조하고 여기에 기존의 심리음향모델 1과 2를 각각 사용한 결과 요구되는 전송율과 실제 사용되는 전공율을 비교함으로써 효율적인 가변 전송율 MPEG-1 오디오 부호화를 제안하고자 한다. 이 방식은 디코더의 개조를 요구하지 않으므로 기존의 MPEG-1 오디오 디코더가 사용되는 곳에는 아무런 제약없이 사용할 수 있는 장점이 있다. 특히 network 환경이나 저장매체 등에아주 효율적으로 사용될 수 있다. 먼저 MPEG-1의 심리음향모델 1, 2와 비트 할당 방식에 대하여 알아보고 가변 전송율이 가능하도록 인코더를 개조하는 방안에 대해 살펴본 다음 시뮬레이션 결과를 검토하기로 한다.

II. MPEG-1 오디오 Layer II

2.1 MPEG-1 오디오 Layer II의 양자화 방식

MPEG-1 오디오 Layer II는 1152 샘플로 구성된 한 프레임의 입력 신호를 36개의 샘플로 구성된 서브밴드로 구분한다. 따라서 한 프레임은 48 kHz 샘플링 주파수를 기준으로 할 때 약 24 msec에 해당한다. 이 시간은 오디오 신호가 일반적으로 정적이라고 가정되는 시간이다. 각 서브밴드로 분리된 신호는 12개 샘플 블럭으로 다시 나누어져서 각 서브 블럭의 최대값을 scalefactor로 설정한다. 따라서 각 서브밴드마다 3개의 scalefactor, scf_1, scf_2, scf_3를 갖는다. 한편 심리음향모델 1은 주파수 영역에서 각 서브밴드의 매스킹 레벨 M_k를 산출한다. 이제 서브밴드를 양자화하는데 소요되는 비트 수를 b_k라하면 양자화기의 스텝의 크기는 $\triangle_k = 2^{-b_k}$가 된다. 만약 양자화기의 잡음이 uniform한 분포를 갖는다고 가정하면 잡음의 세기는 $\triangle^2/12$가 될 것이다. 이제 양자화 잡음이 사람의 귀에 들리지 않도록 하기 위해서는 양자화 잡음이 매스킹 레벨 아래에 위치하도록 양자화기의 스텝 수를 설정하여야 한다. 즉,

$$(2^{-b_k} scf_{k)}^2/12 < M_k$$

이 만족되어야 한다. 이 조건으로부터 서브밴드 샘플들을 부호화하는데 소요되는 비트 수는

$$b_k = \frac{1}{2} \log_2 \frac{scf_k^2}{3M_k}$$

로 결정된다. 따라서 1152개의 모든 서브밴드 샘플을 양자화하는데 소요되는 비트 수는

$$B_{req} = 36 \sum_{k=0}^{31} b_k + \text{추가정보}$$

가 된다. 여기서 추가 정보란 scalefactor, 비트 할당 정보, 헤더 정보 등을 부호화하는데 소요되는 비트 수를 의미한다. 스테레오 신호의 경우 왼쪽 오른쪽 채널에 대해 부호화가 독립적으로 이루어지기 때문에 채널 별로 각각 요구되는 비트 수를 산출하여야 한다. 그러나 joint

stereo 모드에서는 각 채널 간의 상관성을 이용하여 양자화가 이루어지기 때문에 요구되는 비트 산출 방식이 달라지게 된다. 실제로 MPEG-1 오디오에서는 신호의 세기와 매스킹 레벨로부터 SMR(Signal-to-Masking Ratio)를 산출하고 이미 계산된 각 양자화기의 SNR 값과 비교하여 SMR 보다 큰 SNR을 갖은 양자화기 중 가장 작은 것을 선택하는 방식을 택하고 있다.

Joint stereo 모드는 채널 간의 상관성을 활용하여 부호화의 효율성을 높이고자 한 방법이다. 스테레오 오디오 신호의 고주파수 영역은 신호의 크기와 방향성 성분으로 구분할 수 있다. 일반적으로 사람은 2 kHz 이상의 주파수 영역에서 fine structure 보다는 스펙트럼의 포곡선에 의해 스테레오 이미지를 느낀다는 사실이 알려져 있다. 따라서 joint stereo 모드에서는 각각의 채널 신호를 독립적으로 부호화하는 대신 그 평균값을 부호화하며 scalefactor 만을 달리 사용하게 된다. 결과 fine structure는 다소간 상실하지만 전반적인 스펙트럼의 포곡선은 그대로 유지하게 된다. MPEG-1 Layer II에서는 joint stereo 모드를 가용한 비트 수에 따라 4, 8, 12, 또는16번째 서브밴드부터 적용할 수 있도록 되어있다. 만약 비트수가 충분하다면 해당 서브밴드 대역에서 joint stereo 모드 자체가 불필요할 것이다. 이제 이러한 두 가지 모드에서 가변 전송율을 사용하기 위해 인코더를 개조할 때 고려되어야 하는 사항을 살펴보기로 하자.

2.2 가변 전송율 인코더에 고려할 사항

MPEG-1 Layer II는 고정 전송율로 설계되었기 때문에 전송율에 따라 가능한 모드들이 표 1과 같이 설정되어 있다. 한편 샘플링 주파수와 채널 당 요구되는 전송율에 따라 사용 이미 비트 할당이 고정되어 있다. 이 설정은 채널 당 전송율이 낮은 경우 오디오 신호의 대역폭을 제한한 것이다. 표 2는 샘플링 주파수가 48 kHz인 경우 비트 할당표로 채널 당 전송율이 32, 48 kbps일 때 주파수 대역폭은 약 5.25 kHz 정도로 제한됨을 보여준다. 만약 샘플을 양자화 할 때 요구되는 양자화기의 스텝 수가 3, 5, 9이면 효율을 높이기 위해 3개의 샘플들을 하나의 group로 묶어 양자화하지만 그렇지 않은 경우에는 각각의 샘플을 독립적으로 양자화한다.

인코더의 전송율을 R, 샘플링 주파수를 F_s, 프레임 당 샘플 수를 S 라고 하면 각 프레임 당 가용한 비트 수는 $cb = RS/F_s$로 주어진다. 이 가용 비트 중에서 실제로 오디오 샘플들을 부호화하는데 사용되는 비트는 adb = cb - (bhrd + bcrc + bbal + banc) 로 주어진다. 여기서 bhdr은 헤더에 필요한 32 비트, bcrc는 error check에 필요한 16 비트, banc는 부수 데이터이고 bbal은 비트할당 정보에 필요한 비트 수를 의미한다. 이 중 bcrc와 banc는 선택적으로 포함될 수 있는 요소이며 bbal은 서브밴드별로 사용되는 양자화기를 표시하는 것으로 양자화기 데이블에 따라 대역폭에 따라 값이 달라진다. 이제 각 서브밴드별로 매스킹 레벨이 산출되면 이 값을 이용하여 샘플들을 양자화할 때 요구되는 비트 수를 산출하고 이로부터 전송율을 산출한 다음 해당 모드에서 이 전송율이 가능한가를 점검한다. 다음 비트 할당 데이블을 점검하여 대역폭의 제한이 있는지 없는지를 점검한다. 만약 대역폭의 제한이 부적절하면 다른 비트 할당표를 사용해야하며 자연히 전송율은 다음 단계로 상향 조정된다. 이 과정을 거치면 모든 서브밴드에서 잡음이 매스킹 레벨보다 작도록 양자화기가 선택될 수 있다. 그럼에도 불구하고 서브밴드 잡음이 매스킹되지 않는 경우가 발생할 수 있는데 이는 입력 오디오 신호의 scalefactor 중 최대값을 신호의 에너지로 사용하였기 때문이다.

Joint stereo 모드의 경우 가변 전송율의 사용은 이보다 복잡하다. 고정 전송율에서 joint stereo 모드는 주어진 비트 수로 잡음을 은폐하기 위해 joint stereo가 적용되는 서브밴드를 찾았지만 가변 전송율에서는 목표로하는 전송율이 가변적이기 때문에 이 방법은 효율적이지 못하다. 따라서 먼저 목표로 하는 전송율을 효과적으로 설정한 다음 이를 기준으로 joint stereo 모드를 적용할 서브밴드를 설정하는 방안을 사용해야 한다[4]. 목표 전송율은 일반적으로 joint stereo 모드가 전송율에서 20~30 kbps의 이득을 제공한다는 사실을 이용하여 먼저 dual channel으로 코딩하는 경우 요구되는 전송율을 산출한 다음 이로부터 적절한 값의 오프셋을 주어 목표 전송율을 구한다.

III. 시뮬레이션 결과

MPEG-1 Layer II 인코더 알고리즘은 앞에서 언급한 바와 같이 가변 전송율이 가능하도록 개

조되었다. 가변 전송율에 대한 시뮬레이션은 심리음향모델 1과 2를 사용하여 dual channel 모드와 joint stereo 모드에 대해 이루어졌다. MPEG-1에서는 표 1에 보여진 바와 같이 몇가지 단계의 전송율만을 사용하기 때문에 이 전송율 중에서 가장 낮은 전송율을 선택하여 부호화하게 된다.

Dual channel 모드의 실험 결과는 표 2와 3에 보여진 바와 같다. 여기서, 가변 전송율의 경우 요구되는 전송율은 양자화기의 SNR과 심리음향모델로부터 산출된 SMR을 이용하여 직접 산출한 결과이며 부호화된 전송율은 요구되는 SMR을 사용하여 양자화했을 때 모든 서브밴드에서 잡음이 매스킹 레벨 아래에 위치하도록 재설정한 결과이다. 이 때 각 오디오 항목의 전 구간에 걸쳐 산출된 전송율 중 최대값을 고정 전송율로 설정하였다. 표 2와 3은 각각 심리음향모델 1과 2를 사용하여 가변 전송율로 부호화한 결과이다. 두 심리음향모델을 비교해보면 고정된 전송율을 사용하는 경우 "Ornette Coleman", "Glockenspiel" 그리고 "Bass Synth" 을 제외한 거의 모든 오디오 항목에서 모델 2가 모델 1에서 요구되는 전송율의 약 반정도로 부호화될 수 있음을 보여준다. "Ornette Coleman"은 앞서 지적한 대로 고주파수 영역에서의 순음/잡음 판별의 문제라고 여겨지지만, 나머지 두 오디오 항목은 신호의 attack의 횟수가 아주 적거나 오디오 신호의 전력 자체가 매우 작은 경우에 해당한다. 따라서 모델 2가 고정 전송율에서 모델 1에 비해 많은 이득을 제공함을 알 수 있다. 가변 전송율로 사용할 경우 두 모델은 공통적으로 고정 전송율에 비해 약 반정도의 전송율로 부호화될 수 있음을 알 수 있다. 모델 2를 사용할 경우 평균적인 이득은 약 30 kbps이지만 "Glockenspiel"과 "Fireworks"와 같은 특수한 경우를 제외하면 약 40 kbps 정도 차이 나는 것을 알 수 있다. 즉 오디오 신호에 attack이 많을수록 차이가 많이 난다. 따라서 모델 2의 적응성이 가변 전송율을 사용한 경우에도 많은 영향을 미치는 것을 알 수 있다. 가변 전송율을 사용하는 경우 두 심리음향모델에서 공통적으로 확인되는 사실은 요구되는 전송율과 실제 부호화된 전송율이 약 20 kbps 정도 차이난다는 사실이다. 이는 표 1에서 알 수 있듯 전송율의 각 단계들이 약 32 kbps 정도씩 차이나는데다, 앞에서 언급한 바와 같이 전송율을 가변적으로 결정하는데 전송율에 따른 대역폭의 제한을 고려했기 때문이다.

표 2 심리음향모델 1의 경우 요구되는 전송율 (Dual channel 모드)

오디오 항목	가변전송율		고 정 전송율
	요구되는 전송율	부호화된 전송율	
Suzanne Vega	183	200	384
Glockenspiel	74	119	384
Fireworks	132	146	384
Ornette Coleman	195	212	256
Bass Synth	26	65	112
Castanets	175	191	384
Male Speech(English)	145	165	384
Bass Guitar	190	203	384
Trumpet	158	176	320
평 균	127.8	147.7	299.2

표 3 심리음향모델 2의 경우 요구되는 전송율 (Dual channel 모드)

오디오 항목	가변전송율		고 정 전송율
	요구되는 전송율	부호화된 전송율	
Suzanne Vega	130	147	192
Glockenspiel	73	119	192
Fireworks	104	117	160
Ornette Coleman	159	177	320
Bass Synth	28	65	112
Castanets	122	140	192
Male Speech(English)	101	119	160
Bass Guitar	128	144	192
Trumpet	117	134	224
평 균	96.2	116.2	174.4

이제 joint stereo 모드에 대한 결과를 살펴보자 먼저 표 3과 4를 비교하면 고정 전송율을 이용한 경우 joint stereo 모드를 사용함으로써 얻는 전송율의 이득은 모델 1의 경우 약 60 kbps, 모델 2의 경우 약 40 kbps 정도가 됨을 알 수 있다. 오디오 항목 별로 살펴보면 채널 간의 상관성이 비교적 높은 오디오 항목 등에서 많은 이득이 있다는 사실을 알 수 있다. 이제 joint stereo 모드에서 가변 전송율을 적용한 결과를 살펴보면 모델 1의 경우 전송율을 평균 약 반 정도로 낮출 수 있음을 보여준다. 이것은 dual channel 모드의 경우와 같다. 모델 2의 경우는 약 40 kbps 정도의 이득을 보여준다. 가변 전송율을 사용하는 경우 모델 1과 2를 비교하면 전송

율의 차이는 약 10 kbps 정도로 줄어드는 것을 알 수 있다. 이것은 dual channel 모드의 경우와 비교하면 차이가 상당히 줄어든 것으로 모델 1의 단점이 joint stereo 모드에서 가변 전송율을 이용하여 사용할 때 많이 개선된다고 볼 수 있다.

표 4 심리음향모델 1의 경우 요구되는 전송율 (Joint stereo 모드)

오디오 항목	가변전송율		고정
	요구되는 전송율	부호화된 전송율	전송율
Suzanne Vega	132	142	224
Glockenspiel	69	114	384
Fireworks	108	117	320
Ornette Coleman	142	151	192
Bass Synth	26	65	96
Castanets	129	142	384
Male Speech(English)	104	119	224
Bass Guitar	140	152	384
Trumpet	113	126	192
평 균	96.3	112.8	240

표 3 심리음향모델 2의 경우 요구되는 전송율 (Joint stereo 모드)

오디오 항목	가변전송율		고정
	요구되는 전송율	부호화된 전송율	전송율
Suzanne Vega	106	114	128
Glockenspiel	69	113	160
Fireworks	100	109	128
Ornette Coleman	125	134	256
Bass Synth	28	64	112
Castanets	101	112	160
Male Speech(English)	189	102	128
Bass Guitar	105	114	160
Trumpet	101	112	160
평 균	127.8	97.4	139.2

끝으로, 그림 1과 2는 dual channel 모드와 joint stereo 모드의 경우 요구되는 전송율의 변화를 보여준다. 두 심리음향모델 공통적으로 joint stereo 모드를 사용하면 전송율의 이득이 있으나 모델 1의 경우가 모델 2에 비해 많은 이득을 얻는 것을 명확하게 보여준다. 이는 상대적으로 모델 1이 pre-echo를 피하기 위해 과도하게 SMR을 산출하였기 때문이다.

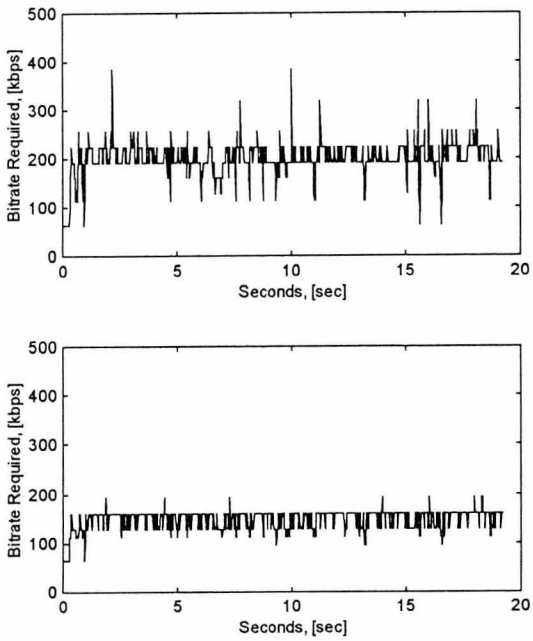

그림 1. 전송율의 변화 ("Suzanne Vega", dual channel 모드) (위) 모델 1 (아래) 모델 2

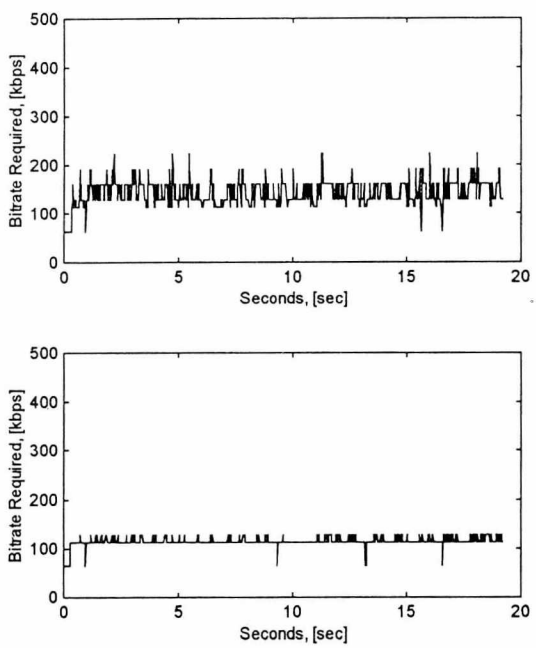

그림 2. 전송율의 변화 ("Suzanne Vega", joint stereo 모드) (위) 모델 1 (아래) 모델 2

IV. 결 론

본 연구에서는 고정 전송율로 설계된 MPEG-1 오디오를 가변 전송율로 변환하였을 때 심리음향 모델 1과 2의 성능이 어떻게 나타나는지에 대해 알아보았다. 모델 1은 모델 2에 비해 비교적 정확한 순음/잡음 판별과 그에 따른 정확한 매스킹 레벨 산출을 이루어내지만 SMR 산출에서 pre-echo를 피하기 위해 worst case 접근 방식을 사용함으로써 모델 2 보다 훨씬 큰 전송율을 요구한다. MPEG-1 오디오를 가변 전송율로 개조하여 사용할 경우 이러한 모델 간의 차이가 큰 의미를 갖는다. 모의 실험 결과 모델 2는 모델 1에 비해 dual channel 모드에서 평균 30 kbps, joint stereo 모드에서 평균 20 kbps 정도 이득을 제공하는 것으로 나타났다. 일반적으로 joint stereo 모드는 dual channel 모드에 비해 많은 이득을 제공하는 것으로 알려져있지만 attack이 많은 오디오 신호에서는 모델 1이 모델 2 보다 많은 차이를 낸다는 사실을 알 수 있었다. 이는 모델 1이 pre-echo를 줄이기위해 각 채널에서 과도하게 SMR을 산출하기 때문이다.

참 고 문 헌

1. K. Brandenburg, "Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding", JAES, vol. 45, No. 12, pp. 4-21, Jan/Feb 1997.
2. ISO/IEC 11172-3, 1993.
3. ISO/IEC 13818-3, 1994.
4. W. Oomen, F. de Bont, L.M. Kerkhof, "Variable Bit Rate Coding for MPEG-1 Audio, Layer I and II", AES 98th Convention, Paris, Feb. 25, 1995.