DBpia

MPEG 오디오 심리음향모델의 비교

정회원 남 승 현*, 박 재 홍**

Comparison of Psychoacoustic Models for MPEG Audio

Seung Hyon Nam*, Jae Hong Park** *Regular Members*

요 약

본 연구에서는 MPEG-1에서 사용되고 있는 두 가지 심리음향모델(모델 1과 모델2)을 비교하였다. 이들은 서로 다른 매스킹 패턴과 매스킹 인덱스 그리고 매스킹 레벨의 산출 과정을 거친다. 모의 실험결과 모델 2는 중고 주파수 영역에 존재하는 순음에 대한 순음지수를 비교적 낮게 판별하는 반면 모델 1은 정확하게 순음과 잡음을 판별한다. 그러나 전반적인 양자화/코딩 결과를 비교하면 모델 2가 모델 1보다 우수한 것으로 드러났다. 그 이유는 첫째, 모델 1이 SMR(Signal-to-Mask Ratio) 산출에 최악상태를 가정했기 때문이며 둘째 모델 2가 한 프레임의 전후반에 각각 한번씩 SMR을 산출하고 그 최대값을 사용하는 반면 모델 1은 한 프레임에서 한번만 계산하기 때문이다. 따라서 모델 2가 모델 1보다 훨씬 정확하며 신호의 변화에 민감하게 적응한다고 볼 수 있다. 고정 비트율에서는 동적 비트 할당 방식을 사용하기 때문에 두 모델이 거의 유사한 음질을 보장한다고 볼 수 있으나, 가변 비트율을 사용할 경우 모델 2는 모델 1에 비해 비트율에서 20-30 kbps/ch의 추가적인 이득을 제공한다.

ABSTRACT

In this study, we investigate two psychoacoustic models, Pyschoacoustic Model 1 and Model 2, employed in the MPEG-1 audio. These two models use different masking patterns, masking indices, and computational procedures. Simulation results show that Model 2 estimates the tonality index relatively low for actual tones in mid and high frequency ranges while Model 1 gives tone/noise decision correctly. Model 2, however, outperforms Model 1 in the overall performance including quantization/coding. The main reasons are: firstly, Model 1 uses worst case assumptions in calculating the SMR. Secondly, Model 2 calculates the SMR twice for the first and second half of a frame while Model 1 calculates ones. It turns out that Model 2 adapts to the input audio signal more precisely than Model 1. Two models provide nearly the same audio quality in fixed bitrates due to dynamic bit allocation. In variable bitreates, however, Model 2 may provide additional 20~30 kbps/ch gain in the bitrate over Model 1.

I. 서 론

지난 10여년 동안 MPEG을 중심으로 오디오 신호 압축기술은 급격히 발전했다[1]. 이 발전은 사람의 청각 특성을 이용하여 잡음을 효율적으로 은폐시키는 심리음향모델에 크게 힘입은 것이다[2]. 현재 많이 사용되고 있는 모델은 MPEG에서 채택되었던 심리음향모델 1과 2이다[3,4]. 심리음향모델 1은 MUSICAM의 개발 과정에서 비롯된 것으로[5] MPEG-1,-2의 Layer I과 II에서 주로 활용되어 왔으며 모델 2는 ATT사의 PAC(Perceptual Audio Coder)의 개발 과정에서 비롯된 것으로[6] MPEG의 Layer III에 적용되었고 이후 MPEG-2 AAC와 MPEG-4에서도 적용되고 있다[1]. 그림 1은 이러한 오디오 부호화기의 일반적인 구조를 보여준다.

* 배재대학교 전자공학과 : 정회원, ** 한국전자통신연구원 위성방송연구소 방송기술연구부 : 정회원
논문번호 : 97450-1210, 접수일자 : 1997년 12월 10일

www.dbpia.co.kr

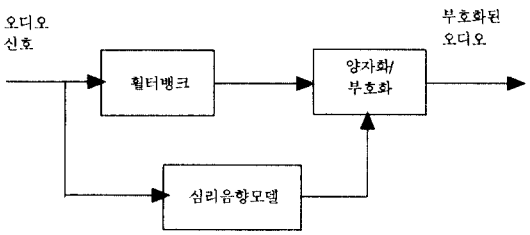

그림 1. 심리음향을 이용한 오디오 부호화기의 일반적인 구조
Fig. 1 A General Structure of Audio Coders Employing Psychoacoustic Characteristics

이들 두 심리음향모델은 각각 다른 매스킹 패턴, 매스킹 인덱스, 순음/잡음 판별법등을 사용하고 있으며 매스킹 레벨의 산출 과정도 다르다. 또 산출된 매스킹 레벨은 양자화에 직접 적용하지는 않으며 대신 매스킹 레벨로부터 SMR(Signal-to-Mask Ratio)을 산출하고 이를 이용하여 신호를 양자화한다. 따라서 산출된 매스킹 레벨이 거의 유사하더라도 결과적인 SMR이 판이하게 다를 수 있는데 이는 오디오 신호의 비정적인 부분을 처리하는 방식의 차이에서 비롯된다. 일반적인 변환 코더는 오디오 신호의 정적인 부분에서 블럭 크기를 대략 43ms까지 설정하지만 비정적인 부분에서는 거의 3ms 정도의 시간 해상도를 요구하기도 한다. 이러한 신호의 급작스런 변화에 대처하기 위해 오디오 코더는 시간-주파수 해상도의 적절한 타협을 필요로 한다. MPEG-1,-2의 Layer I과 II에서는 32개의 균등 분할된 서브밴드필터를 사용함으로써 시간 해상도를 충분히 고려한 반면 저주파수 영역에서의 주파수 해상도를 어느 정도 희생했다. 시간-주파수 해상도의 타협을 좀더 효율적으로 이루기 위한 방안으로 동적 윈도우 스윗칭 방식이 도입되어졌다.

이러한 시간-주파수 해상도의 문제는 비단 양자화 자체에만 국한되어 있는 것이 아니라 당연히 양자화에 필요한 정보를 제공하는 심리음향모델에서도 고려되어져야 한다. 본 논문에서는 MPEG에서 사용된 두 심리음향모델을 비교하여 각각의 장단점을 분석하고 그에 따른 활용을 모색하고자 한다.

II. 심리음향모델

심리음향모델 1[3]

심리음향모델 1의 SMR 산출 과정은 대략 다음과 같다.

1) FFT 변환
2) 각 서브밴드에서의 음압(SPL)산출
3) 절대음압 고려
4) 순음/잡음 성분 판별
5) 주요 매스커만을 걸름
6) 각 매스커로부터 매스킹 레벨 산출
7) 전역 매스킹 레벨 산출
8) 각 서브밴드별 최소 매스킹 레벨 산출
9) 각 서브밴드별 SMR 산출

순음 성분은 일정 스펙트럼 영역 내의 피크로 설정되는데 이 영역은 주파수가 높아질수록 넓어진다. 일단 순음이 판명되면 순음의 음압을 결정하고 순음을 찾는데 사용된 조사범위 내의 모든 스펙트럼은 0으로 처리된 후 잡음 결정 단계로 넘겨진다. 순음이 제거된 이 스펙트럼은 임계대역별로 구분되어 각 임계대역별로 하나의 잡음 매스커를 산출한다. 각 매스커로부터 매스킹 레벨을 산출할 때 사용되는 매스킹 패턴은 최악상태를 가정한 모델로 그림 2에 있는 바와 같다. 또 적용되는 매스킹 인덱스는 매스커가 순음인 경우

$$av_{tm} = 1.525 + 0.275 * z(j) + 4.5$$

잡음인 경우

$$av_{nm} = 1.525 + 0.175 * z(j) + 0.5$$

로 주어진다. 여기서 $z(j)$는 $LT_{\min}(n)$으로 결정된다. 또 n번째 서브밴드의 SMR은 매스커의 위치이다. 순음 매스킹 인덱스가 잡음 매스킹 인덱스보다 크다는 것은 순음 매스커의 매스킹 능력이 떨어진다는 것을 의미한다. 또한 매스킹 인덱스가 주파수에 따라 증가하는 것을 볼 수 있다. 매스킹 레벨은 n번째 서브밴드내의 최소 매스킹 레벨

$$SMR_{sb}(n) = L_{sb}(n) - LT_{\min}(n) \text{ dB}$$

에 의해 구해진다. 여기서 $L_{sb}(n)$는 n번째 서브밴드에서의 오디오 신호의 음압으로

$$L_{sb}(n) = \max[X_{spl}, 20 * \log_{10}(scf_{\max}(n) * 32768) - 10]$$

이다. 여기서 주목해야 할 것은 SMR이 최악 상태를 고려한 값이란 점이다. 즉 각 서브밴드에서 최소 매스

www.dbpia.co.kr

킹 레벨을 선택했고 반면에 신호의 음압으로는 최대값을 사용한 점이다. 신호에 attack이 있는 경우 SMR은 각 서브밴드 프레임마다 많은 차이를 보일 수 있다.

심리음향모델 2[3]

심리음향모델 2의 SMR 산출과정은 대략 다음과 같다.

1) 컴플렉스 FFT 변환
2) 예측불가 지수 산출
3) 1/3 임계대역별로 에너지 및 에너지로 가중화된 예측불가 지수 산출
4) 에너지와 매스킹 패턴의 콘볼루션
5) 예측불가 지수를 순음지수 α, $(0<\alpha<1)$로 변환
6) 매스킹 인덱스 계산 : $a_i = \alpha * TMN + (1-\alpha) * MNT$ dB
   NMT(Noise Masking Tone) = 잡음 매스킹 인덱스
   TMN(Tone Masking Noise) = 순음 매스킹 인덱스
7) 절대음압을 고려하여 매스킹 레벨 fthr 산출
8) 각 서브밴드별 SMR 산출 : $SMR = 10 * \log_{10}(\frac{epart(n)}{npart(n)})$

여기서 $epart(n)$와 $npart(n)$는 각각 n번째 서브밴드 내의 에너지와 매스킹 레벨이다.

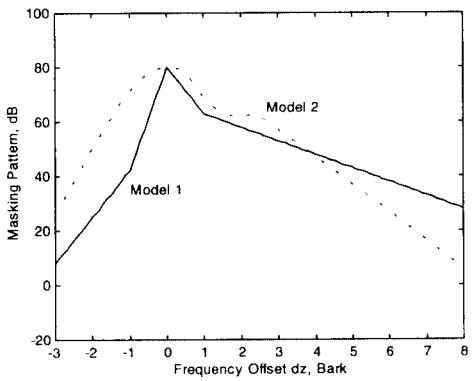

그림 2. 매스킹 패턴
Fig. 2 Masking Patterns

그림 2는 모델 2에서 사용된 매스킹 패턴을 보여준다. 그림에서 볼 수 있듯 모델 2의 매스킹 패턴은 끝부분이 둥글고 왼쪽 기울기가 모델 1 보다 완만하기 때문에 패턴의 왼쪽 부분에서 높은 매스킹 패턴을 보여준다. 한편 매스커의 설정을 보면 모델 1은 매스커를 순음과 임계대역의 잡음으로 규정하고 각각의 매스킹 인덱스를 적용하고 있으나, 모델 2에서는 1/3 임계대역별로 하나의 매스커를 규정하며 매스커의 순음/잡음 판별을 순음지수로 규정하고 매스킹 인덱스는 순음과 잡음 인덱스를 선형보간법을 사용하는 점이 다르다. 따라서 모델 2는 경우에 따라 모델 1보다 3배 가량 많은 매스커를 사용한다고 볼 수 있다. 잡음 매스킹 인덱스는 5.5 dB 순음 매스커의 경우 $z(j)$에 매스커가 있을 때 max(24.5, (14.5 + z(j))로 주어진다. 따라서 저주파수 영역에서 24.5 dB이며 주파수가 증가함에 따라 40.3 dB까지 증가한다. 이는 순음지수가 0.3 일때 고주파수 영역에서 모델 1의 순음 매스킹 인덱스보다 약 3 dB 작은 값이다. 심리음향모델 2에서는 SMR을 산출할 때 저주파수 영역에서는 최소 매스킹 레벨을 사용하고 중고주파수 영역에서는 평균 매스킹 레벨을 사용하며 신호의 에너지로는 서브밴드 내의 평균값을 적용한다. 이러한 SMR 산출은 프레임의 전후반에 각각 한번씩 실행되며 그 중 최대값이 선택된다.

III. 시뮬레이션 결과

먼저 MPEG-1 Layer II 알고리즘을 이용하여 두 심리음향모델로부터 산출되는 매스킹 레벨을 조사하였다. 그림 3은 사용된 매스킹 패턴의 차이에서 비롯된 매스킹 레벨의 차이를 보여준다. 즉 모델 2는 모델 1 보다 완만한 왼쪽 기울기를 갖기 때문에 5 kHz 부근에서 모델 2의 매스킹 레벨이 모델 1 보다 높게 산출되었음을 알 수 있다. 반면에 10 kHz 이상의 고주파수 영역에서는 모델 2가 모델 1 보다 낮은 매스킹 레벨을 산출함을 볼 수 있다.

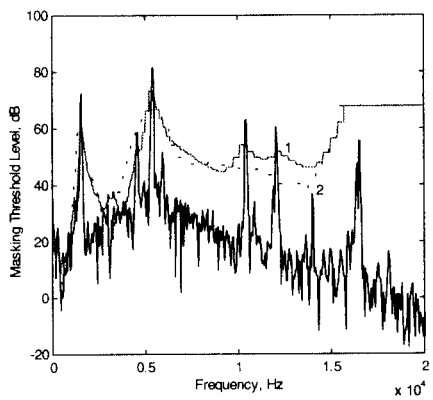

그림 3. 모델 1과 2의 매스킹 레벨 비교
Fig. 3 Comparison of Masking Levels from Model 1 and 2

www.dbpia.co.kr

고주파수 영역에서의 매스킹 레벨 차이는 그림 4에 나타난 바와 같이 모델 1과 2의 순음/잡음 판별의 차이에서 비롯된다. 모델 1은 순음/잡음을 0/1로 판별하는데 비해 모델 2는 0에서 1까지의 순음지수를 이용하는데 그림 4에서 볼 수 있듯 저주파수 영역에서는 비교적 정확한 순음지수를 산출하지만 고주파수 영역에서의 순음지수 판별이 부정확한 것을 볼 수 있다. 따라서 모델 2는 3 kHz 이내의 저주파수 영역에서는 비교적 정확한 매스킹 레벨을 산출할 수 있었으나 주파수가 높아짐에 따라 매스킹 레벨이 다소간 부정확하게 된다. 특히 10~15 kHz 영역에서 매우 낮은 매스킹 레벨을 제공하는 것을 볼 수 있다. 한편 4~6 kHz의 영역에서 모델 2가 모델 1 보다 높은 매스킹 레벨을 산출하는 이유는 그림 3에서 보여진 바와 같이 매스킹 패턴의 왼쪽 기울기의 차이 때문이다. 즉 모델 1이 모델 2보다 좀더 보수적인 매스킹 패턴을 사용하는데 기인한 것이다.

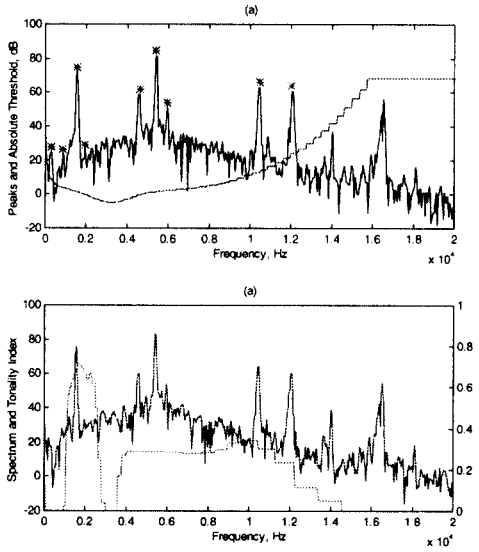

그림 4. 모델 1(a)과 2(b)의 순음/잡음 판별
Fig. 4 Tone/Noise Discrimination of Model 1 and 2

그러나 실제로 양자화에 적용되는 것은 매스킹 레벨이 아니라 SMR이므로 매스킹 레벨과 신호의 레벨로부터 어떻게 SMR을 산출하는가 역시 매우 중요하다. 따라서 최종적인 SMR을 이용하여 양자화를 했을 때 요구되는 비트율을 산출해 보았다. 그림 5는 심리음향모델 1과 2를 사용한 경우 시간에 따른 비트율의 변화를 보여준다. 비트율은 입력되는 오디오 신호에 따라 많이 변화한다. 모델 1의 경우 비트율은 대부분의 시간에서 128 kbps이하이고 평균 84 kbps이다. 모델 2의 경우 평균 비트율은 55 kbps로 모델 1 보다 훨씬 낮은 비트율을 요구한다. 그림 6에서 관찰할 수 있는 또 다른 특징은 모델 1이 모델 2 보다 훨씬 급한 비트율의 변화를 보여준다는 점이다. 이러한 현상은 그림 7의 "Castanets" 오디오 신호의 경우에서 극명하게 드러난다. 이러한 현상의 주된 이유는 모델 1과 2가 SMR을 계산하는 방법에서 많이 다르기 때문이다. 앞에서 살펴 본 바와 같이 모델 1은 SMR을 계산할 때 신호의 에너지로 서브밴드의 최대 스펙트럼 성분을 사용하고 매스킹 레벨로는 해당 서브밴드 내의 최소값을 적용하는 매우 보수적인 최악상태를 가정한 접근 방식을 사용하기 때문에 매우 큰 SMR을 요구하게 되는 것이고 자연히 비트율이 커지게 된다. 그러나 모델 2에서는 신호의 에너지와 매스킹 레벨을 각 서브밴드 내에서 평균하여 사용하기 때문에 상대적으로 작은 SMR과 비트율을 산출할 수 있는 것이다. 실제로 사람의 귀는 한 임계대역에서의 평균적인 크기에 더 잘 반응하는 것으로 알려져있다[2]. 더구나 모델 2는 한 오디오 프레임을 전반과 후반의 두 지역으로 나누어 각각 SMR을 구한 다음 각 대역별로 두

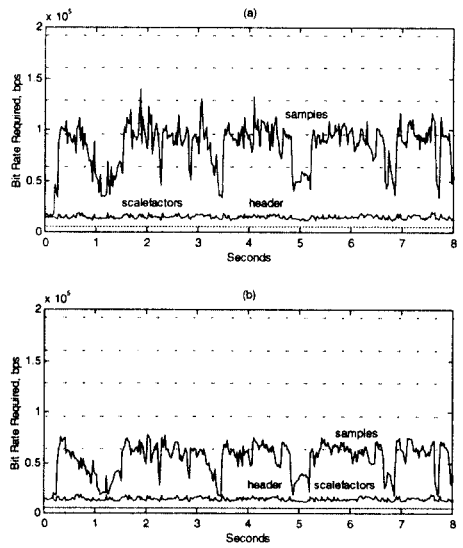

그림 5. 요구되는 bit rate의 변화 ("Male Speech") (a) 모델 1 (b) 모델 2
Fig. 5 Variation of Required Bit Rates ("Male Speech") (a) Model 1 (b) Model 2

www.dbpia.co.kr

값의 최대값을 사용하는 방식을 취하기 때문에 attack에 대해 매우 효과적인 비트율을 산출하게 된다. 이에 반해 모델 1은 무조건 한 오디오 프레임에 한번의 SMR을 산출하기 때문에 신호의 attack을 효과적으로 처리하지 못하고 있다. 물론 같은 계산을 두 번씩 반복함으로써 야기되는 계산량의 문제가 있으나 계산량의 문제는 인코더의 복잡도는 디코더의 복잡도에 비해서는 대부분의 경우 심각한 문제가 되지 않는다는 사실을 상기할 필요가 있다.

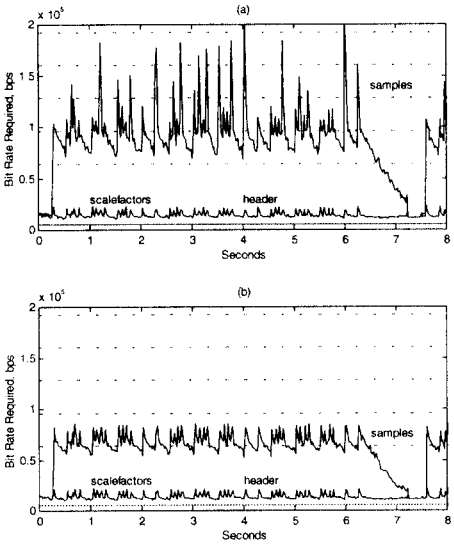

그림 6. 요구되는 bit rate의 변화 ("Castanets") (a) 모델 1 (b) 모델 2

Fig. 6 Variation of Required Bit Rates ("Castanets") (a) Model 1 (b) Model 2

그림 7은 "Ornette Coleman" 오디오 항목에 대한 비트율의 변화를 보여준다. 이 그림에서 4~5초의 기간은 소리가 줄어들어 거의 사라지는 부분에 해당한다. 모델 1은 이 시간 동안에도 128 kbps 정도의 비트율을 거의 일정하게 요구하고 있지만 모델 2는 128 kbps에서 64 kbps 정도의 아주 낮은 비트율로 낮아지는 것을 볼 수 있다. 즉 모델 2는 오디오 신호의 입력 변화에 보다 민감하게 반응하여 효율적으로 비트율을 산출하는 것을 알 수 있다. 이 시간 동안에 요구된 비트율이 얼마나 정확한 값인가를 알아보기 위해 4.5초 시점에서 각 모델에서 64 kbps의 비트율로 인코딩/디코딩한 결과 얻어진 매스킹 레벨과 잡음 수

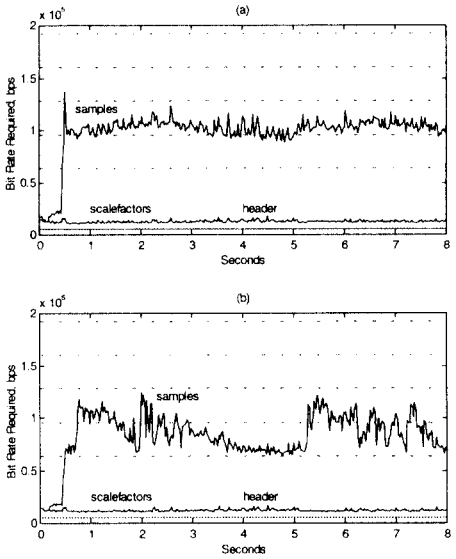

그림 7. 요구되는 bit rate의 변화 ("Ornette Coleman") (a) 모델 1 (b) 모델 2

Fig. 7 Variation of Required Bit Rates ("Ornette Coleman") (a) Model 1 (b) Model 2

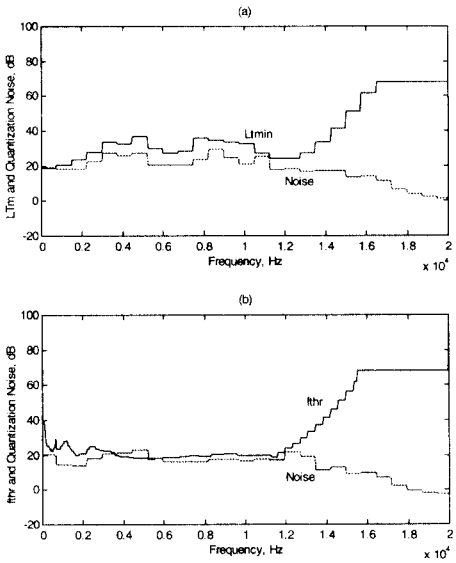

그림 8. 양자화 잡음과 매스킹 레벨의 비교 ("Ornette Colemen" 4.5초 시점, 64kbps): (a) 모델 1 (b) 모델 2

Fig. 8 Comparison of Quantization Noise and Masking Levels (frame at 4.5 sec of "Ornette Colemen", 64 kbps): (a) Model 1 (b) Model 2

www.dbpia.co.kr

준을 비교해보았다. 여기서 잡음 수준은 원래의 신호와 인코딩/디코딩된 신호의 차이로부터 얻어진 결과이다. 모델 1의 경우 요구되는 비트율이 128 kbps이므로 64 kbps의 비트율에서는 잡음 수준이 매스킹 레벨 보다 훨씬 클 것을 예상 할 수 있다. 그러나 실제로 그림 8은 잡음 수준이 매스킹 레벨 보다 현저히 낮은 것을 알 수 있다. 이것은 모델 1에서 예상한 비트율이 과도하게 산출된 것임을 입증하는 것이다. 이에 비해 모델 2는 64 kbps보다 약간 높은 비트율을 요구하고 있으며 결과로 얻어진 매스킹 레벨과 잡음 수준은 요구된 비트율과 거의 일치함을 보여준다. 따라서 모델 2가 모델 1 보다 훨씬 적응적이고 정확한 비트율을 산출하는 것을 알 수 있다.

그림 8에서 얻을 수 있는 또 하나의 사실은 모델 2의 경우 비트율과 매스킹 레벨, 그리고 잡음 수준의 일치도가 매우 높기 때문에 이를 객관적인 음질의 지표로 활용할 수 있다는 점이다. 그러나 모델 2를 사용하는 경우 잡음 수준와 매스킹 레벨의 차이는 음질의 지표로 활용될 수 없다. 표 1은 9개의 오디오 항목에 대해 매스킹 레벨을 넘는 평균적인 서브밴드와 MNR (Mask-to-Noise Ratio)를 산출한 결과이다. 표에서 볼 수 있듯 모델 2의 경우 MNR과 매스킹 레벨을 넘는 평균 서브밴드의 수가 각 비트율에 대해 비교적 일치하는 것을 알 수 있다. 매스킹 레벨을 넘는 평균 서브밴드의 수와 MNR의 관계를 살펴보면 대략 10 dB의 MNR이면 사람의 모든 잡음이 효과적으로 은폐될 수 있음을 보여준다. 그러나 모델 1의 경우는 이러한 관계가 일정하게 드러나지 않으며 이 사실은 앞에서 언급한 바와 같이 오디오 신호에 attack이 심하게 나타나는 "Castanets" 항목과 그렇지 않은 "Ornette Coleman"을 비교하면 자명하다. 한편 모델 2의 경우 거의 대부분의 오디오 항목에서 96 kbps 일 때 거의 대부분의 서브밴드에서 잡음이 매스킹 레벨 보다 낮다는 사실을 알 수 있다.

마지막으로 9개 오디오 항목에 대해 평균 비트율을 산출한 결과한 결과가 표 1에 나타나있다. 모델 1을 사용한 MPEG-1 Layer II 알고리즘은 128 kbps/ch에서 투명한 음질을 보장하는 것으로 알려졌으나 실제 요구되는 평균 비트율은 그보다 훨씬 낮다는 것을 알 수 있다. 앞에서 살펴 본 바와 같이 모델 2가 모델 1에 비해 낮은 비트율을 요구함을 알 수 있으며 그 차이는 "Glokenspiel"이나 "Bass Synth"를 제외한 거의 모든 항목에서 약 20~30 kbps 정도가 된다. 이러한 특성을 활용하면 기존의 MPEG-1 Layer II를 가변적인 비트율로 개조하여 평균 비트율을 훨씬 낮출 수 있다. 이러한 이득은 오디오 디코더의 개조없이 인코더만을 약간 개조함으로써 얻을 수 있다. 최근 Philips사는 모델 1을 이용한 가변 비트율 오디오 코더를 제안한 바 있다[7]. 그러나 만약 모델 2를 사용한다면 Philips사의 가변 오디오 코더 보다 약 20-30 kbps/ch 더 낮은 비트율을 갖는 코더를 예상할 수 있을 것이다. 여기에 인텐서티 모드를 사용하면 비트율은 더욱 낮아진다. 따라서 대부분의 스테레오 오디오 항목들이 약 64 kbps/ch 이내에서 처리될 수 있으리라 기대된다.

표 1. 평균 MNR(dB)과 매스킹 되지 않은 평균 서브밴드 수 (괄호 안의 값)

Table 1. Average MNR(dB) and Average Number of Unmasked Subbands (Numbers in Parentheses)

	모델 1			모델 2		
Bitrate[kbps/ch]	128	96	64	128	96	64
Suzanne Vega	10.7 (0.2)	3.1 (3.0)	-4.0 (15.5)	19.3 (0.0)	10.9 (0.0)	2.5 (3.1)
Glockenspiel	26.9 (0.0)	19.3 (0.1)	11.8 (1.1)	32.3 (0.0)	24.1 (0.0)	14.9 (0.8)
Fireworks	15.8 (0.3)	7.4 (1.0)	-2.0 (11.0)	25.7 (0.0)	17.1 (0.0)	6.8 (0.5)
Ornette Coleman	8.9 (0.0)	1.0 (6.4)	-5.6 (17.9)	14.3 (0.0)	5.8 (1.8)	-2.0 (10.1)
Bass Synth	16.0 (0.0)	12.0 (0.0)	7.5 (0.0)	27.6 (0.0)	23.8 (0.0)	17.4 (0.0)
Castanets	11.6 (1.4)	4.0 (3.0)	-3.2 (13.0)	20.1 (0.0)	11.9 (0.0)	3.2 (2.5)
Male Speech (English)	11.8 (0.0)	4.6 (2.2)	-2.4 (11.4)	20.5 (0.0)	13.1 (0.0)	5.1 (1.0)
Bass Guitar	9.8 (1.3)	1.9 (6.8)	-6.0 (14.7)	20.8 (0.0)	12.1 (0.0)	3.4 (3.5)
Trumpet(Haydn)	14.9 (0.0)	6.4 (0.8)	-2.1 (11.7)	23.6 (0.0)	15.0 (0.0)	5.4 (2.0)
평균	14.0 (0.4)	6.6 (2.6)	-0.7 (8.5)	22.7 (0.0)	14.9 (0.2)	6.3 (2.6)

표 2. 요구되는 평균 비트율 [kbps/ch]

Table 2. Required Average Bit Rates [kbps/ch]

오디오 항목	모델 1	모델 2
Suzanne Vega	97	68
Glokenspiel	49	40
Fireworks	82	55
Ornette Coleman	102	83
Bass Synth	28	24
Castanets	90	63
Male Speech(English)	84	55
Bass Guitar	102	67
Trumpet	85	61
평균	80	57

www.dbpia.co.kr

IV. 결 론

본 연구에서는 MPEG-1에서 사용하고 있는 두 가지 심리음향모델을 비교하였다. 이들은 서로 다른 매스킹 패턴과 매스킹 인덱스 그리고 매스킹 레벨의 산출 과정을 거친다. 모델 2는 중고주파수 영역에 존재하는 순음에 대한 순음지수를 0.3 정도로 비교적 낮게 판별하는 반면 모델 1은 정확하게 순음과 잡음을 판별한다. 그러나 전반적인 양자화/코딩 결과를 비교하면 모델 2가 모델 1보다 우수한 것으로 드러났다. 그 이유는 첫째, 모델 1이 SMR 산출에 최악상태를 가정했기 때문이며 둘째 모델 2가 한 프레임에서 전후반 2번의 SMR을 산출하고 그 최대값을 사용하는 반면 모델 1은 한 프레임에서 한번만 계산하기 때문이다. 따라서 모델 2가 모델 1보다 훨씬 정확하며 신호의 변화에 민감하게 적응한다고 볼 수 있다. 고정된 비트율을 사용할 경우 모델 1이 요구하는 비트를 충족시켜주지 못 한다고 해서 곧 음질이 저하되는 것은 아니다. 따라서 128 kbps 이상의 고정된 높은 비트율에서는 특정 서브밴드의 SMR은 최종적인 음질에 크게 영향을 미치지는 않는다. 이것이 MPEG의 Layer II가 모델 1을 사용하여 코딩했을 때 요구되는 비트율이 비록 128 kbps를 초과할지라도 거의 모든 오디오 신호에 대해 투명한 음질을 얻는 주된 이유이다. 그러나 가변 비트율을 사용할 경우 모델 2는 모델 1에 비해 20-30 kbps의 추가적인 이득을 제공한다.

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. W. Zwicker and H. Fastl, Psychoacoustics, Berlin: Springer-Verlag, 1990.
3. ISO/IEC 11172-3, 1993.
4. ISO/IEC 13818-3, 1994.
5. R. N. J. Velhuis, "Bit Rates in Audio Source Coding", IEEE Trans. on Selected Areas in Commun., vol. 10, No., 1, pp. 86-96, Jan., 1992.
6. D. Johnston, "Transform Coding of Audio Signals using Perceptual Noise Criteria", IEEE J. on Selected Areas in Comm., vol.6, No.2, pp.314-323, Feb. 1988.
7. 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.

남 승 현(Seung Hyon Nam) 정회원

1980년 2월 : 서강대학교 전자공학과(학사)

1987년 8월 : The Univ. of Alabama in Hunstiville, 전기 및 컴퓨터공학과(석사)

1992년 12월 : Texas A&M University 전기공학과(박사)

1979년 12월 ~ 1985년 6월 : 국방과학연구소 연구원

1993년 3월 : 배재대학교

현재 : 전자공학과 부교수

<연구분야> 음성 및 오디오 부호화, 적응훨터, 이동통신

박 재 홍(Jae Hong Park) 정회원

1978년 2월 : 서울대학교 공업교육과(학사)

1980년 2월 : 서울대학교 전자공학과(석사)

1995년 2월 : 서울대학교 전자공학과(박사)

1979년 10월 ~ 1985년 6월 : 국방과학연구소 연구원

1985년 7월 : 한국전자통신연구원

현재 : 책임연구원, 방송기술연구부장

<연구분야> 이동통신, 위성통신 및 전파방송

www.dbpia.co.kr