MPEG 오디오 부호화 방법의 성능 향상

*신종인, 김기수, 윤대희, 차일환
연세대학교 전자공학과 음향 음성 및 신호처리 연구실

Improved MPEG-Audio Coding Method

*Jong-In Shin, Ki-Soo Kim, Dae-Hee Youn, Il-Whan Cha
ASSP Lab. Dept. of Electronic Eng. Yonsei University

요 약

ISO/MPEG에서는 스테레오 신호만을 부호화할 수 있는 MPEG-1 오디오 부호화 방법을 5.1 채널(L, R, C, LS, RS, LFE)의 다채널 신호로 확장한 MPEG-2 오디오 방법을 제안하였다. 압축해야 될 신호가 증가하면서 MPEG에서는 채널 내의 부호화 방법으로는 MPEG-1에서 제안된 방법을 사용하고, 부가적으로 채널 간의 부호화 방법을 이용하여 MPEG-1과 호환이 가능하도록 하는 부호화 방법을 다방면에 걸쳐서 연구하여 표준화 작업을 진행하고 있다[1][2].

본 논문에서는 MPEG 오디오 부호화 방법을 두가지 측면에서 효율적으로 향상시키는 방법을 제안하고자 한다. 첫번째는 MPEG에서 제안한 오디오 부호화 알고리듬을 개선하여 음질과 비트율에 있어 향상시키는 것으로 각 서브밴드의 비트 할당 방법과 시간 영역에서의 마스킹 효과 등을 사용한 심리음향 모델 등의 개선 방법이 제안되었다. 두번째 방법은 부호화기의 계산량을 감소시키는 방법으로 심리음향 모델이나 비트 할당시의 계산 과정에 있어 반복적인 과정은 시간 영역에서의 중복성을 이용하여 계산량에 대한 향상율 얻을 수 있었다.

1 장. 서 론

국제 표준화 가구 산하의 동화상 전문가 그룹(ISO/MPEG)에서는 동화상과 고품위의 오디오를 1.5 Mbit/s 급의 저장 매체인 CD에 압축, 저장할 수 있는 MPEG-1 표준안을 1991년에 확정하였다[1]. MPEG-1 표준안은 멀티미디어, Video-CD, Video On Demand 등에 활발히 응용되고 있다. 그러나 HDTV, DTV, DAB와 같은 방송 매체에 적용할 경우 영상의 질이 떨어질 뿐 아니라 오디오에 있어서도 다채널, 음성 다중 등의 많은 부가 서비스를 필요로 하므로 새로운 표준안인 MPEG-2에 대한 연구와 함께 표준화가 진행되어 왔다[2]. 특히, HDTV의 경우는 5채널 이상의 많은 채널을 필요로 하므로 CD와 같은 음질을 얻기 위해선 3.5 Mbit/s 이상의 전송율이 필요하다. 따라서 고음질을 유지하면서 큰 압축율을 얻는 기술의 개발은 필수적이다. MPEG-2 표준안은 위와 같은 조건을 만족하면서 약 384 kbit/s의 전송율로 다채널 오디오를 압축할 수 있는 표준안이다. 여기에 사용된 오디오 압축 알고리듬은 MPEG-1과 같이 청각 특성을 이용한 서브밴드 부호화 방법으로 주관적인 음질의 손상없이 큰 압축율을 얻을 수 있다. 이와 같은 고음질 오디오 부호화 기술은 공통적으로 사람의 청각 특성과 기존의 데이타 압축 기법이 결합된 형태를 갖는다[3][4][5][6]. 오디오 신호는 광범위한 음원(Source)을 갖고 있기 때문에 음성 부호화와 같은 음원 발생 모델을 적용할 수 없다. 따라서 수신원인 귀의 청각 특성을 이용하여 중복성을 제거하여야 하는데 여기에 적용된 주된 특성은 마스킹 현상이다. 마스킹 현상은 음압이 다른 두 음이 존재할 때 음압이 큰 음이 작은 음을 들리지 않게 하는 현상을 말한다[7].

본 논문에서는 MPEG 오디오 부호화 방법에서 중요한 역할을 하는 심리음향 모델과 비트 할당 방법을 개선하는 방법을 제안하였다. 심리음향 모델에서는 주파수 영역 마스킹 만을 사용하는 MPEG 방법과 달리 시간 영역 마스킹을 포함시켜 더욱 정확한 마스킹 값을 얻을 수 있었다. 또한, 시간적으로 안정한 신호의 반복적 계산량을 줄이기 위해 새로운 방법을 사용하였다. 비트 할당 방법에서는 Time Differential Bit Allocation Infoation(TDBSI)을 사용하여 비트 할당 정보에 사용되는 비트를 줄였다. 또한 각 서브밴드의 비트 할당 계산에서 반복적인 과정을 Signal to Mask Ratio (SMR)로 부터 직접 구하는 방법을 사용하여 계산량에 대한 향상을 얻을 수 있었다.

본 연구는 한국 방송 공사(KBS)의 연구비 지원으로 이루어졌습니다.

2 장. MPEG 오디오 압축 알고리듬

MPEG의 오디오 압축 알고리듬은 오디오 신호의 두가지 중복성을 제거하는 방법으로 설명될 수 있다. 먼저, 통계적인 중복성을 제거하기 위해 신호를 32개의 필터뱅크를 통과시켜 각 대역의 크기에 따라 비트 할당을 달리하는 서브밴드 부호화방법을 사용한다. 이 신호를 더욱 압축하기 위해 마스킹 현상과 임계 대역, 절대 가청 한계 등의 청각 특성을 이용하여 지각적인 중복성을 제거한다. 위와 같은 두가지 방법을 사용하여 입력 오디오 신호는 주관적인 음질을 떨어뜨리지 않고 큰 압축율을 가져온다. 그림 1은 MPEG 오디오 부호화기의 블럭 다이어그램을 나타낸다.

2-1. 서브밴드 부호화

서브밴드 부호화는 신호를 대역 통과 필터를 통과시켜 여러 주파수 대역으로 나누어 부호화하는 방식으로 각 대역에 할당되는 비트 수를 신호의 크기에 따라 달리하여 신호대 잡음비를 향상할 수 있었다. 입력단에서 N 개의 대역 통과 필터로 구성된 분석 필터뱅크(analysis filterbank)에 입력 신호를 통과시킨다. 이때 부호화하는 샘플의 수가 동일해지도록 대역 통과 필터의 수만큼 간축(decimation)하여 표본화율을 낮추는 과정을 갖는다. 각 대역 통과 신호는 Adaptive PCM(APCM) 등을 사용해서 부호화하고 부가 정보와 다중화하여 전송한다[8][9].

서브밴드 부호화에서 가장 중요한 필터 뱅크는 QMF 를 비롯하여 여러 방식이 개발되었는데 최근에는 임계 대역과 유사한 모양을 갖도록 비선형적 크기를 갖는 필터뱅크를 설계하기도 한다. MPEG에서는 32개의 동일 크기를 갖는 가중 중첩 가산(Weighted Overlap-Add) 방법의 Single Side Band(SSB) 필터 뱅크를 사용하였다[1]. 가중 중첩 가산 방법의 필터뱅크는 블럭 단위로 데이타를 처리하여 효율적인 계산이 가능하다. 서브밴드 분석에 사용되는 필터는 512-tap 저역 통과 필터가 기본이 되며 행렬에 의해서 주파수 천이되어 32개의 동일 크기 서브밴드가 된다.

2-2. 심리음향 모델

심리음향 모델은 원음이 존재할 때, 양자화 잡음이 마스킹되어 들리지 않게되는 최대값을 각 주파수 대역에서 얻는 방법이다. 따라서 서브밴드 부호화나 변환 부호화에서 심리음향 모델을 사용하면 각 밴드에서 원음에 의해 마스킹되어 들을 수 없는 최대의 잡음 레벨을 결정할 수 있다. 이 잡음 레벨, 즉 마스킹 임계값을 사용해서 각 밴드의 실제 양자화기를 결정하는 비트 할당을 할 수 있다. 양자화 잡음의 크기는 각 밴드에 할당되는 비트의 수에 의해 결정되므로 비트 할당을 적절하게 하면 주관적인 왜곡을 최소화할 수 있다.

심리음향 모델은 큰 압축율을 필요로 하고 부호화기의 시스템이 복잡해도 무관한 경우에는 지연 시간도 크고 계산량도 많지만 정확한 모델이 사용되고 낮은 압축율에서 응용되는 간단한 시스템에서는 마스킹 곡선을 대략적으로 모델링하는 방법이 사용된다. 예를 들어 HDTV와 같이 부호화기가 대량으로 필요하지 않는 방송 시스템에서는 정교한 모델을 사용하여 압축비와 음질을 높히고, 부호화기까지 필요로 하는 DCC, MD 등에서는 압축비는 낮지만 간단한 모델을 사용한다.

MPEG에서는 두가지의 심리음향 모델을 제공하는데 두 모델을 응용 분야에 따라 적절하게 선택해서 사용할 수 있다. 첫번째 방법은 FFT 스펙트럼을 순음과 잡음 성분으로 나누어 각 성분에 의한 마스킹 임계값을 구한 후, 절대 가청 한계를 고려하여 마스킹 임계값을 구하는 비교적 간단한 방식으로 계층 1, 2에 사용한다. 그 과정은 그림 2와 같다. 두번째 방법은 FFT 스펙트럼을 청신경의 여기 모델인 스프레딩 함수와 콘볼루션하여 마스킹 임계값을 구하는 방식으로 계산량이 많지만 더 정교한 결과를 얻을 수 있으며 계층 3에 사용된다[1].

2-3. 비트 할당과 양자화

서브밴드 샘플값을 정규화시키기 위한 scalefactor는 매 프레임(12샘플)마다 찾아진다. 찾아진 scalefactor는 그 자체가 부호화되는 것이 아니라 그 값에 해당하는 scalefactor 표의 인덱스가 부호화된다. 또한 시간 영역에서 비슷한 값을 갖는 경우가 많으므로 인접된 scalefactor는 scalefactor 선택 정보(scfsi : scalefactor select information)를 이용하여 사용되는 비트수를 줄이는 과정을 거친다.

서브밴드 부호화에서 가장 중요한 비트 할당에 사용되는 기본 법칙은 한 프레임에 사용 가능한 비트를 넘지 않으면서 그 프레임 전체의 Mask to Noise Ratio(MNR)를 최대로하여 주관적 음질을 향상시키는 것이다. 한 샘플당 사용 가능한 비트는 서브밴드에 따라 0 비트에서 16, 14, 6, 3 비트까지 사용하며 1 비트는 신호대 잡음비의 증가가 없으므로 사용하지 않는다. 위와 같은 값을 갖는 이유는 계층 2에서 고주파 대역의 비트 할당 정보에 사용되는 비트수를 줄이기 때문이다. 즉, 통계적 특성을 이용하여 고주파 대역에 할당되는 비트수는 저주파 대역에 할당되는 비트수보다 현저히 작으므로 미리 정의하여 사용 비트를 줄일 수 있다. 각 서브밴드에 대해서 MNR은 SNR에서 SMR을 빼서 얻어지고 SNR은 사용되는 비트 수에 의해 결정된다.

$$MNR(dB) = SNR(dB) - SMR(dB)$$

처음에 0 비트가 모든 밴드에 할당된 후 다음 과정을 반복 수행한다.

1) 최소 MNR을 갖는 서브밴드를 찾는다.
2) 그 서브밴드에 비트를 할당하여 SNR을 향상시킨다.
3) 이 서브밴드의 새로운 MNR을 계산한다.
4) 남아 있는 비트 수를 계산한다.

위의 반복 과정은 남아있는 비트가 한 번 수행할 때 쓰여지는 비트수보다 작지 않을 때까지 수행한다.

서브밴드 샘플을 양자화하기 위해서 0 근처의 작은 값에 대한 오차를 줄일수 있는 mid_tread uniform 양자화기가 사용된다. 각 서브밴드 샘플들은 scalefactor에 의해 나누어져 정규화된 후 양자화된다. 할당된 비트수가 작은 서브밴드의 경우 3개의 연속된 샘플을 묶어서 부호화하여 비트를 더 줄일 수 있다.

3 장. MPEG 오디오 압축 방법의 성능 향상

MPEG 오디오 부호화 방법은 프레임 내의 중복성 즉, 주파수 영역에서의 마스킹 방법과 서브밴드 부호화 만을 이용한 방법이다. 본 연구에서는 시간 영역에서의 중복성을 MPEG 방법에 결합하여 더욱 성능을 향상시킬 수 있도록 두가지 방향에서 연구를 진행하였다. 먼저, 심리음향 모델과 비트 할당 방법에서 시간 영역 중복성을 이용한 음질 향상과 비트를 효율적으로 사용하는 방법에 대해서 연구하였다. 두번째는 부호화기의 계산량을 줄이기 위해 시간 영역에서 오디오 신호의 안정 상태(Stationary State)를 이용하는 방법을 연구하였다.

3-1. 제안된 심리음향 모델

3-1-1. 시간 영역 마스킹(Temporal Masking)

그림 3은 시간과 주파수 영역에서 순음에 의한 마스킹 곡선을 나타낸다[10]. 그림을 보면 주파수 영역에서와 같이 시간 영역에서도 마스킹 현상이 나타나는 것을 알 수 있다. 시간 영역 마스킹에는 전마스킹(pre_masking)과 후마스킹(post_masking)으로 나누어지는데 전마스킹의 경우 지속 시간이 짧고 마스킹의 크기도 작으므로 이용하기 어렵다. 그러나 후마스킹의 경우는 지속 시간이 약 100 ~ 200 msec 정도로 길고 상당히 큰 값을 갖기 때문에 심리음향에서 중요한 역할을 한다.

후마스킹의 크기는 세개의 변수, 즉 마스커(masker)의 주파수, 지속 시간(duration), 지연 시간(delay)에 따라 결정된다. 마스커의 주파수에 따른 변화는 저주파일 경우 큰 값을 갖고 고주파일 때는 매우 작은 값을 갖는다. 약 1 kHz 이내의 신호에 대해서는 큰 값을 갖는다. 지속 시간의 경우는 마스커의 지속 시간이 길수록 더욱 큰 마스킹 효과를 갖으며 200 msec 이상의 지속 시간을 갖는 경우는 동일한 값을 갖는다[7]. 지연 시간은 마스커와 마스키(maskee) 간의 시간 차이로 약 100 msec 이내일 때 더욱 효과적이다.

MPEG의 계층 2에서 한 프레임의 길이는 약 20 ~ 30 msec로 지연 시간이 고정 값을 갖으므로 주파수와 지속 시간을 변수로 사용하여 후마스킹 성분을 이용할 수 있다. 심리음향 모델1에서 시간 영역 마스킹을 이용하기 위해 이전 1 프레임의 신호만을 이용한 심리음향 모델을 사용함으로써 좀 더 정확한 결과를 얻을 수 있었다.

3-1-2. 고속 알고리듬

음성 및 오디오 신호는 시간 영역의 작은 구간(약 20 ~ 100 msec)에서는 안정하다고 가정하며 대부분의 경우에 실제 분석을 할 때도 이와 같은 가정을 이용한다[11]. 실제 오디오 신호의 경우 안정 구간이 대부분을 차지하며 천이 구간은 타악기나 효과음 등에서 부분적으로 나타나는 경우가 많으므로 심리음향 모델에 사용되는 반복적인 계산을 줄일 수 있다.

다음은 본 논문에서 제안한 고속 심리음향 모델이다. 버퍼에 이전 프레임의 FFT 스펙트럼에서 절대 가청 한계값보다 큰 값을 저장한다. 현재의 FFT 스펙트럼과의 차이값의 제곱을 계산한다. 이때 미리 정해진 임계값을 넘으면, 즉, 천이 구간에서는 MPEG의 심리음향 모델이 적용된다. 임계값보다 작은 안정 구간에서는 전 프레임의 마스킹 임계값을 사용하여 현재 프레임의 마스킹 임계값을 얻는다. 각 서브밴드에서의 음압 변화값(SPL variance)을 계산한다. 이 값을 가지고 전 프레임의 마스킹 임계값에서 새로운 마스킹 임계값을 얻고 각 서브밴드에서의 SMR을 구한다. 이때, 최종적인 결과인 각 서브밴드에서의 비트 할당 결과는 원래의 방법과 거의 변화를 주지 않으면서 계산량은 크게 줄일 수 있다. 실제 부호화기에서 심리음향 모델이 차지하는 계산량은 전체 계산량의 약 60 % 정도로 위와 같은 방법을 사용했을 때 30 ~ 50 % 정도의 계산량 이득을 가져올 수 있다.

그림 4에 고속 심리음향 알고리듬의 흐름도를 나타내었다. 실제 응용시에는 PC 상에서의 오디오 데이타 압축과 같이 고정 계산량을 필요로 하지 않는 시스템 등에서 큰 효과를 나타낼 수 있다. 일정한 계산량을 필요로 하는 실시간 시스템에서는 몇 프레임을 하나의 패킷으로 묶어 부호

화하거나 버퍼를 두어 사용하면 일정한 계산량을 제공할 수 있다.

3-2. 비트 할당 방법

본 논문에서는 MPEG에서 제안한 비트 할당 방법을 좀 더 효율적으로 개선해서 음질과 계산량 면에서 최적화가 가능하도록하는 알고리듬을 제안하였다.

MPEG에서는 최소 MNR 값을 갖는 서브밴드를 찾아 비트를 할당한 후 다시 MNR을 계산하여 반복적으로 서브밴드에 비트를 할당하는 방법을 사용하였다. 이 방법은 반복 계산을 필요로 하므로 효과적이지 못하다. 각 서브밴드에 직접 모든 비트를 할당하는 방법을 사용하기 위해 심리음향 결과로부터 얻어진 SMR 값으로부터 지각적으로 동일한 복원음을 얻기 위해 필요한 비트(total_net_bit)를 다음식에 의해 계산한다.

$$total\_\text{net}\_bit = \sum_{n=0}^{31} \text{net}\_bit_n$$

$$\text{SMR} > 0\ \text{dB}: \quad \text{net\_bit}_n = \text{SMR(dB)}/7 + 1$$

$$\text{SMR} < 0\ \text{dB}: \quad \text{net\_bit}_n = 0$$

사용 가능한 비트가 total_net_bit 값보다 클 경우 각 밴드에 비트를 할당해주고 남는 비트를 저주파 대역 밴드로부터 차례로 할당해 준다. 만약 비트가 모자라는 경우는 각 서브밴드의 중요도에 따라 비트를 줄여준다. 중요도를 나타내는 가중 함수는 각 서브밴드 당 임계 대역의 수로써 얻어진다.

비트 할당 후 각 서브밴드의 비트 할당 정보를 복호화단에 전송해주기 위해 필요한 비트는 계층 1에서 프레임당 128 비트가 필요하고 계층 2에서는 약 90 비트 가량이 필요하다. 여기에 사용되는 비트를 줄이기 위해 scfsi에 사용된 것과 같이 이전 프레임과 비트 할당값이 같은 경우 전 프레임 값을 사용하고 다른 경우는 그 차이 값만을 전송한다. 이와 같은 방법을 사용하면 비트 할당 정보에 사용되는 비트는 프레임당 약 40 ~ 70 비트만으로도 충분하므로 각 서브밴드에 더 많은 비트를 할당하여 음질을 향상시킬 수 있다. 표 1은 이와 같은 방법을 사용한 복원음과 MPEG 방법을 사용하여 동일하게 96 kbit/s에서 부호화한 복원음의 음질 평가 결과이다. 평가 데이타는 4개의 유럽 방송 연맹의 음질 평가 자료(EBU/SQAM) 데이타와 영화 효과음을 사용하였고 음질 평가 방법은 마스킹 값과 잡음 레벨과의 평균 거리를 나타내는 MNR을 사용하였다. 결과를 살펴보면 96 kbit/s 일 때는 약간 좋은 결과를 갖는 것을 알 수 있다. 실제로 계층 2의 경우 96 kbit/s 이내의 전송율에서 응용되지 않지만 더 낮은 비트율에서는 제안된 방법이 보다 뛰어난 음질을 갖는다.

4 장. 결 론

본 논문에서는 차세대의 디지탈 오디오 압축 분야에서 표준안으로 채택된 MPEG 오디오 압축 알고리듬의 성능 향상에 대해 연구하였다. MPEG 오디오 압축 방법은 심리음향 모델과 서브밴드 부호화가 결합된 형태로 압축율을 높히면서 복원음의 음질이 중요하므로 많은 계산량을 요구한다.

오디오 부호화 방법의 성능 향상을 위해서 심리음향 모델과 비트 할당 방법의 개선하였다. 계산량과 비트율을 줄일 수 있는 방법으로 심리음향 모델에서는 시간 영역 마스킹 효과를 포함하여 더 정확한 마스킹 임계값을 얻었다. 또한 심리음향 모델에서 계산량을 줄이기 위해 안정한 신호에 대해서는 이전 프레임의 마스킹 임계값과 음압 레벨의 차를 사용하여 반복적인 계산량을 줄일 수 있었다. 비트 할당 방법에서는 TDBSI를 사용하여 각 서브밴드의 비트 할당 정보에 사용되는 비트를 절약하여 서브밴드 샘플에 할당, 전체적인 음질을 향상시킬 수 있었다. 반복 계산을 줄이기 위해 프레임 내에서의 net_bit를 계산하여 비트 할당을 함으로써 계산량을 줄였다.

음질 평가를 위해 기존의 방법과 제안된 방법을 96 kbit/s의 압축율에서 MNR을 측정하였다. 5개의 오디오 시료에 대한 결과에서 제안된 방법이 향상된 결과를 나타내었다. 또한 96 kbit/s 이내의 낮은 비트율에서는 더욱 큰 음질 향상을 가져올 수 있었다.

참 고 문 헌

[1] ISO/IEC JTC1/SC29/WG11 No. 71 "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to 1.5 Mbit/s - CD-11172-3(Part 3. MPEG-Audio)"

[2] ISO/IEC JTC1/SC29/WG11 No. 703 "Generic Coding of Moving Pictures and Associated Audio - CD-13818-3(Part 3. MPEG-Audio)"

[3] K. Brandenburg, "OCF - a new coding algorithm for high quality sound signals." *Proc. ICASSP* pp. 141-144, 1987

[4] Y. F. Dehery, *et al.* "A MUSICAM source codec for digital audio broadcasting and storage." *Proc. ICASSP* pp.3605-3608, 1991

[5] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria." *IEEE J. Selected Areas Comm.* pp. 314-323, 1988

[6] 김기수, 윤대희 "고음질 오디오 부호화" 제10회 음성통신 및 신호처리 워크샵 1993, pp. 233-237

[7] E. Zwicker, *Psychoacoustics*. Springer-Verlag, New York, 1982

[8] R. E. Crochire, S. A. Webber, J. L. Flanagan, "Digital Coding of Speech in Subbands," *Bell Syst. Tech. J.*, vol. 55, pp.1069-1085, Oct. 1976

[9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms : Principles and Applications to Speech and Video*. Englewood Cliffs, NJ:Prentice Hall, 1984

[10] J. G. Beerends and J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation" J. Audio Eng. Soc., Vol. 40, 1992 Dec.

[11] S. Furui, M. M. Sondhi, *Advances in speech signal processing*. Marcel Dekker, 1991

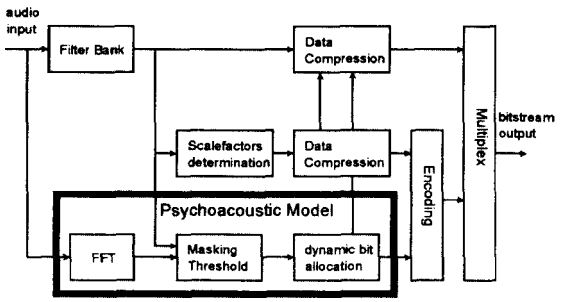

그림 1. MPEG 오디오 부호화기의 블럭 다이어그램

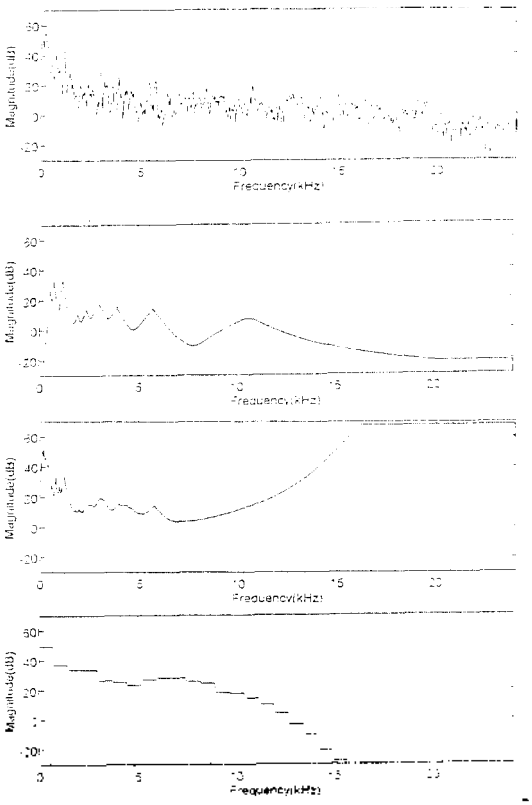

그림 2. 심리음향 모델의 결과
a) FFT 스펙트럼 b) 개별 마스킹 임계값
c) 전체 마스킹 임계값 d) 신호대 마스크비

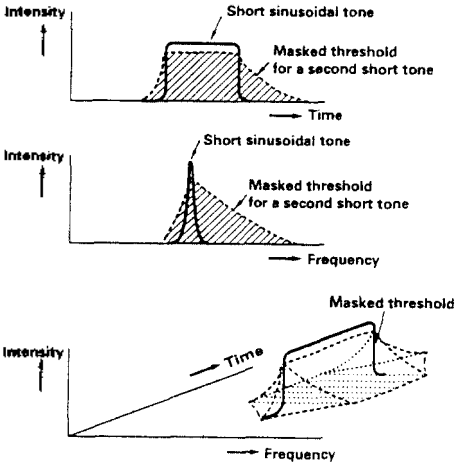

그림 3. 순음에 의한 마스킹 곡선[10]

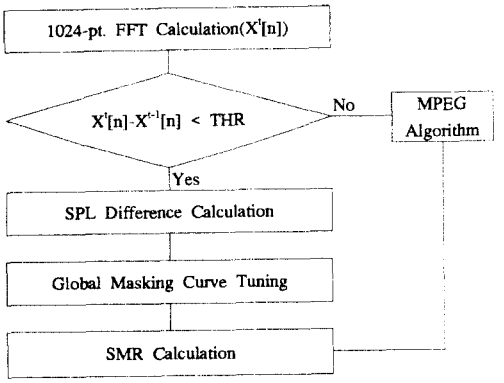

그림 4. 고속 심리음향 모델 알고리듬

표 1. 음질 평가 결과(MNR)

Method / Data	MPEG(dB)	Improved MPEG(dB)
Soprano	6.77	7.34
Violin	2.60	5.53
Trumpet	8.09	7.00
Opera	4.93	6.75
Movie Effect (Terminator II)	4.93	6.92