논문96-1-1-10

시간-주파수 구조에 근거한 지각적 오디오 부호화기

김 기 수*, 서 호 선*, 이 준 용**, 윤 대 희*

A Pereceptual Audio Coder Based on Temporal-Spectral Structure

Ki-Soo Kim*, Ho-Sorn Suh*, Joon-Yong Lee**, and Dae-Hee Youn*

요 약

일반적으로 고음질 오디오 부호화 방법은 전통적인 데이타 압축 기법과 인간의 청각 모델을 결합한 구조를 갖고 있다. 고음질 오디오 부호화에 사용되는 주요한 청각 특성은 주파수 영역에서의 마스킹 현상이므로 서브밴드 부호화나 변환 부호화와 같은 주파수 영역 방법들이 널리 사용된다[1][2]. 그러나 지금까지의 고음질 오디오 부호화에서 시간 영역 마스킹과 시간 영역 중복성을 제거하는 방법은 적용되지 않았다.

본 논문에서 제안한 오디오 데이타 압축 방법은 시간 및 주파수 영역에서 통계적, 지각적 중복성을 제거한다. 주파수 영역으로 변환된 오디오 신호는 6프레임으로 구성된 패킷으로 나뉘어 진다. 한 패킷은 1536 샘플(256*6)로 되어 있으며 패킷 내에서의 중복성은 시간 및 주파수 영역에서 존재한다. 각 패킷에서 두 중복성이 동시에 제거되어 진다. 심리음향 모델에 있어서도 세밀한 주파수 마스킹과 함께 시간 영역 마스킹을 고려하여 보다 정확한 결과를 얻을 수 있도록 향상되었다. 양자화를 위해서 각 패킷은 비선형적인 임계 대역과 시간적인 청각 특성을 반영할 수 있도록 설계된 부블럭으로 분할되었다. 따라서 낮은 비트율에서 고음질의 복원음을 얻을 수 있었다.

Abstract

In general, the high quality audio coding(HQAC) has the structure of the convertional data compression techniques combined with moodels of human perception. The primary auditory characteristic applied to HQAC is the masking effect in the spectral domain. Therefore spectral techniques such as the subband coding or the transform coding are widely used[1][2]. However no effort has yet been made to apply the temporal masking effect and temporal redundancy removing method in HQAC.

The audio data compression method proposed in this paper eliminates statistical and perceptual redundancies in both temporal and spectral domain. Transformed audio signal is divided into packets, which consist of 6 frames. A packet contains 1536 samples(256×6) and redundancies in packet reside in both temporal and spectral domain. Both redundancies are elminated at the same time in each packet. The psychoacoustic model has been improved to give more delicate results by taking into account temporal masking as well as fine spectral masking. For quantization, each packet is divided into subblocks designed to have an analogy with the nonlinear critical bands and to reflect the temporal auditory characteristics. Consequently, high quality of reconstructed audio is conserved at low bit-rates.

*연세대학교 전자공학과
Dept. of Electronic Engineering, Yonsei Univ.
**KBS 기술연구소
※본 논문은 한국방송공사의 연구비지원으로 이루어졌습니다.

Ⅰ. 서 론

디지탈 오디오는 80년대에 접어들면서 CD나 DAT와 같은 대용량 저장 매체의 개발과 함께 오디오 기기의 표준이 되었다. 그러나 많은 양의 디지탈 오디오 데이타는 방송, 통신 등의 응용에 장애가 되었다. 이에 80년대 후반부터 세계 각국의 여러 연구소에서는 CD 수준의 디지탈 오디오 데이터를 지각적인 음질을 떨어뜨리지 않고 압축하는 기술을 연구,개발하였다[1][2][3][4].

1988년에는 국제 표준화 기구 내의 동화상 전문가 그룹(ISO/MPEG)이 창설되었다. 여기에서 오디오 압축 기법의 표준화 작업에 착수하여 동화상과 CD 수준의 디지탈 오디오를 1.5Mbit/s급의 디지탈 저장 매체에 압축, 저장할 수 있는 MPEG-1 표준안을 91년 확정하였다. 그러나 HDTV와 같은 방송 매체에 적용할 겨우 화질이 떨어질 뿐 아니라 오디오에 있어서도 다채널, 음성 다중 등의 많은 부가 서비스를 필요로 하므로 새로운 표준안인 MPEG-2에 대한 표준안을 지난 94년 11월 확정하였다[5].

고음질 오디오 부호화 기술은 공통적으로 청각 특성을 기존의 데이타 압축 기법과 결합한 형태를 갖는다. 오디오 신호는 광범위한 음원(Source)을 갖고 있을 뿐만 아니라 고음질을 필요로 하기 때문에 음성 부호화와 같은 음원 발생 모델을 적용할 수 없다. 따라서 수신원인 귀의 청각 특성을 이용하여 중복성을 제거하여야 하는데 여기에 주로 적용된 특성은 마스킹(masking) 현상이다. 마스킹 현상은 음압이 다른 두 음이 존재할 때 음압이 큰 음이 작은 음을 들리지 않게 하는 현상으로 주파수 영역에서 큰 값을 갖는다. 따라서 고음질 오디오 부호화에는 필터뱅크를 사용한 서브밴드 부호화나 FFT, DCT 등을 사용하는 변환 부호화 방식이 사용된다[6].

본 논문에서 제안한 오디오 부호화 방법은 오디오 신호의 통계적, 지각적 중복성을 시간 영역과 주파수 영역에서 동시에 제거하여 낮은 비트율에 있어서도 높은 음질을 갖도록 구성되어 있다. 이를 위해 MDCT(Modified Discrete Cosine Transform)를 통해 주파수 변환된 오디오 신호는 6 프레임으로 구성되는 패킷으로 나누어진다. 1536 샘플(256×6=35msec)로 구성된 패킷은 시간 및 주파수 영역에서 동시에 중복성을 갖는다. 따라서 패킷 내에서는 시간 및 주파수 영역 중복성을 동시에 제거할 수 있다. 심리음향 모델에 있어서도 주파수 영역 마스킹과 함께 시간 영역 마스킹을 고려하여 더욱 정교한 결과를 얻도록 구성되었을 뿐 아니라 신호의 특성에 따라 계산량을 줄일 수 있도록 구성하였다. 양자화를 위해 각 패킷은 다시 부블럭으로 나뉘어지는데 주파수 영역에서 임계대역과 유사한 구조를 갖도록 비선형적으로 분할하였고 시간적 청각 특성까지 고려될 수 있도록 설계되었다. 비트 할당에 있어서는 신호대마스크비(SMR : Signal to Mask Ratio)와 각 부블럭의 특성에 맞도록 설계하였다. 양자화기는 부블럭의 특성에 따라 달리하였다.

Ⅱ. 음의 청각적/통계적 특성

음의 인식에 있어서 청각 구조는 매우 비선형적인 특성을 갖는다. 그림 1은 이와 같이 주파수, 크기 성분이 지각적인 단위로 변환되어 인식되는 구조를 보여준다[7]. 귀에서의 주파수 인식은 비선형적인 로그함수와 유사한 임계 대역에 따라서 음을 인식하며 그 단위는 bark이다. 따라서 저주파 대역에서 높은 해상도를 가지며 고주파 대역에서는 낮은 해상도를 가지므로 스펙트럼의 미세 구조보다는 대략적인 포락선만을 인식하게 된다. 또한 음의 크기 인식에 있어서도 절대 가청 한계가 존재하여 조용한 환경에서도 저주파 대역과 고주파 대역에서는 어느 정도의 음압 레벨 이상일 경우만 음을 인식한다. 2~5kHz 정도의 중간 주파수 대역은 외이의 공명 주파수와 일치하는 대역으로 가청 한계가 낮고 민감한 청각 특성을 갖는다. 이와 같은 지각적 음압 레벨을 sone이라고 한다[6].

위의 두가지 특성 외에 오디오 데이터 압축에 필수적인 청각 특성으로 마스킹 현상이 있다. 마스킹 현상은 인접 대역에 두 음이 존재할 때 음압이 큰 음(masker)이 작은 음을 차폐시키는 현상을 가리킨다. 마스킹은 발생 영역에 따라 시간 마스킹(temporal masking)과 주파수 마스킹(simultaneous masking)으로 나뉘어진다. 시간 영역 마스킹에서는 마스커가 뒤따라오는 신호를 차폐시키는 후마스킹(postmasking)이 큰 값을 갖는다. 주파수 영역에서도 저주파에 존재하는 마스커가 고주파의 신호를 쉽게 차폐시키는 특성을 갖는다. 그림 2는 1kHz에 존재하는 비순음(nontonal component)에 의한 시간 및 주파수 영역 마스킹 곡선

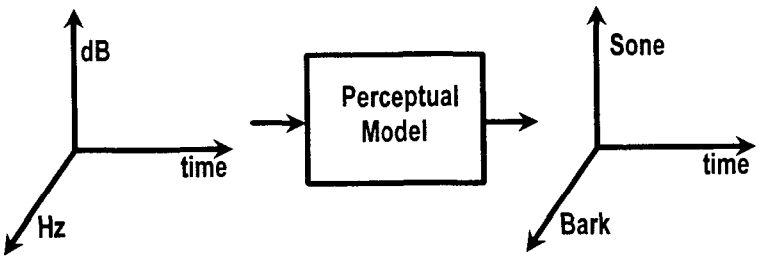


그림 1. 음의 지각적 인식(internal representation)

을 대략적으로 나타내고 있다[9].

위와 같은 현상을 이용하여 고음질의 오디오 부호화기에서는 양자화시에 발생하는 잡음의 형태를 변형하여 신호에 의해 완전히 마스킹되도록 설계한다. 비트율이 충분할 경우에는 양자화 잡음이 마스킹 곡선의 아래에 존재하게 되고, 지각적으로 동일한 복원음을 얻을 수 있다고 가정할 수 있다. 그러나 실제 부호화 알고리듬에 적용하기 위해서는 각 주파수 대역에 따라서 임계 대역이 다르고 그에 따른 마스킹 특성도 저주파 대역과 고주파 대역에서 각각 다른 특성을 갖기 때문에 정확한 모델을 찾는데는 어려움을 갖는다. 또한 인간의 청각 특성에 있어서도 저주파 대역과 고주파 대역의 특성은 매우 다르기 때문에 실제 부호화기의 설계에 있어서도 이와 같은 특성을 반영해야 한다.

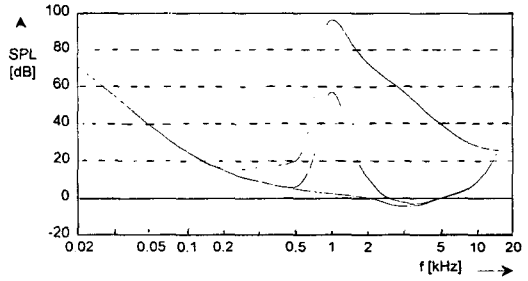


그림 2. 1kHz에서의 비순음 마스킹 곡선

각 주파수 대역에서 오디오 신호의 통계적 특성과 귀의 지각적 특성을 정리하면 다음과 같다.

1) 저주파 대역(LF region) : 대부분의 오디오 신호는 이 대역에서 매우 큰 음압 레벨을 갖는다. 또한 저주파 대역의 신호는 비교적 안정하여 급격한 변화가 적기 때문에 시간적인 중복성이 있음을 알 수 있다. 지각적으로는 이 대역에서 주파수 해상도를 결정하는 임계 대역 값이 매우 작게 나타나므로 주파수 선택성이 좋으며 가청 한계도 매우 낮다. 주파수 영역 마스킹 값이 작게 나타나는 반면 큰 시간 마스킹 값을 갖는다. 따라서 지각적으로도 시간 영역 중복성이 크게 나타난다.

2) 중간 주파수 대역(MF region) : 이 대역에서는 저주파 대역에서와 마찬가지로 오디오 신호의 음압 레벨이 크고 지각적으로도 중요한 역할을 하는 부분이다. 오디오 신호의 통계적 중복성은 저주파 대역과 고주파 대역의 중간 값을 갖는다. 지각적인 면에서는 가청 한계가 가장 낮은 대역이므로 음의 인식에 있어서도 매우 중요한 역할을 한다.

3) 고주파 대역(HF region) : 오디오 신호의 음압 레벨은 매우 작지만 음색과 깊이 등의 인식에 있어서는 매우 중요한 역할을 하는 대역이다. 주파수 선택성이 떨어지며 가청 한계도 높고 주파수 영역 마스킹이 크게 일어난다. 세밀한 주파수 스펙트럼보다는 대략적인 포락선만을 인지할 수 있다. 따라서 주파수 영역에서 큰 중복성을 갖는다.

Ⅲ. 제안된 오디오 부호화기

그림 3은 본 논문에서 제안한 부호화기의 구조를 보여준다. 오디오 데이타는 MDCT에 의해 주파수 축으로 변환되고 6 프레임이 하나의 패킷을 구성한다. 각 패킷은 시간 및 주파수 영역에서 청각 특성에 적합하도록 부블럭으로 분할되어 진다. 심리음향 모델에서는 시간 및 주파수 마스킹을 이용해서 각 부블럭의 마스킹 임계값을 계산하고 이 값은 각 부블럭의 중복성과 함께 비트 할당의 기본이 된다. SMR을 이용하여 초기 비트 할당값을 계산한 후 각 부블럭의 중복성과 사용 가능한 비트의 양에 따라 비트 할당값을 조정하게 된다. 양자화기는 각 주파수 대역의 특성에 맞도록 저주파 및 중간 주파수 대역과 고주파 대역을 다르게 설계하였다.

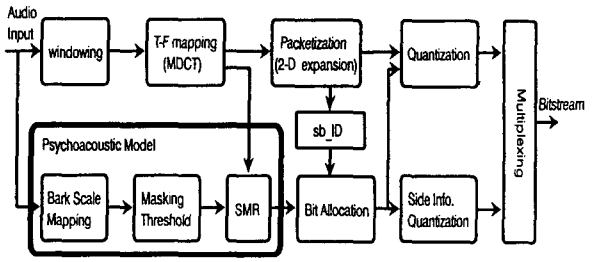


그림 3. 제안된 오디오 부호화기의 블럭 다이어그램

1. 주파수 변환

청각 특성을 이용한 오디오 부호화기에 있어서 시간 영역의 신호를 주파수 영역으로 변환하여 부호화한다. 이는 청각 기관이 음을 인식하는 특성이 시간 영역 보다는 주파수 영역에서 더 잘 표현되기 때문이다. 본 논문에서는 512 point MDCT를 사용하여 시간 영역의 신호를 주파수 대역으로 변환하였다. 여기에 사용된 윈도우는 Dolby AC-3에서와 같은 Fielder Window를 사용하였으며 50% 중첩하여 시간 영역 에이리어징을 제거하였다. 6 프레임을 하나의 패킷으로 구성하여 시간 및 주파수 영역 중복성을 제거할 수 있도록 구성하였다. 다음 식은 MDCT와 IMDCT 식을 나타낸다.

Forward MDCT :

$$M(k)=-\frac{2}{N}\sum_{n=0}^{N-1}\left\{x(n)\cos\left(2\pi\frac{(2n+1)(2k+1)}{4N}+\pi\frac{(2k+1)}{4}\right)\right\} \quad (1)$$

Inverse MDCT :

$$x(n)=-2\sum_{n=0}^{N-1}\left\{X(k)\cos\left(2\pi\frac{(2n+1)(2k+1)}{4N}+\pi\frac{(2k+1)}{4}\right)\right\} \quad (2)$$

2. 패킷화

본 논문에서는 오디오 신호의 시간 및 주파수 특성을 효과적으로 반영하기 위해서 패킷을 구성하였다. 패킷은 6개의 시간 영역 프레임으로 구성되며 각 프레임은 주파수 영역에서 256 샘플의 DCT 계수들로 이루어 진다. 각 패킷은 다시 60개의 시간 및 주파수에서의 부블럭으로 분할된다. 그림 4는 각 부블럭으로 나뉘어진 패킷의 구조를 보여준다. 패킷을 구성하였을 때 발생하는 장점은 고정 비트율 부호화기에 적합하다는 것이다. 지금까지 시간 영역 중복성의 제거를 고려한 부호화기는 오디오 신호를 천이 구간과 안정구간으로 나누어 안정 구간에서 발생한 시간 이득을 버퍼 조정을 이용한 비트율 조정 과정이 필요하게 된다. 본 논문에서 제안한 부호화기는 한 패킷 내에서 동일한 비트율을 제공해 줄 수 있으며 심리음향 모델의 계산량을 줄이는데 있어서도 적합하다고 할 수 있다.

저주파 대역은 약 0~2.6kHz까지의 주파수 대역으로 10밴드로 분할된다. 따라서 각 밴드는 대역폭이 260Hz 정도의 높은 주파수 해상도를 갖는다. 반면에 시간축에 있어서는 6 프레임을 하나의 단위로 처리하므로 시간 해상도는 35msec가 된다.

중간 주파수 대역은 2.6~7.7kHz까지의 대역으로 2장에서 설명한 바와 같이 귀의 민감도가 매우 높아 가청 한계가 매우 낮고 시간 및 주파수 해상도는 중간 정도를 갖는다. 따라서 주파수 대역에서 520Hz의 대역폭을 갖는 10개의 밴드로 나뉘어 지며, 시간 영역에서 2개의 밴드로 분할하여 20개의 부블럭으로 나누어진다.

고주파 대역은 7.7~18.1kHz까지의 주파수 대역으로 주파수 해상도가 낮은 반면 시간 해상도가 높은 대역이다. 본 논문에서는 각 프레임마다 10개의 주파수 대역으로 분할하여 5.8msec의 시간 해상도를 제공하는 30개의 부블럭으로 나뉘어진다. 18.1kHz 이상의 초고주파 대역은 음에 대한 민감도가 매우 낮고 신호의 크기도 상당히 작은 대역이다. 실제 청신경에 있어서도 해상도가 떨어지므로 특별한 경우를 제외하고는 고려하지 않는다. 표 1은 각 대역의 특성을 보여준다.

3. 심리음향 모델

심리음향 모델은 고음질 오디오 부호화기에 있어서 가장 핵심적인 부분으로 많은 계산량을 필요로 한다. 본 눈문에서는 MPEG

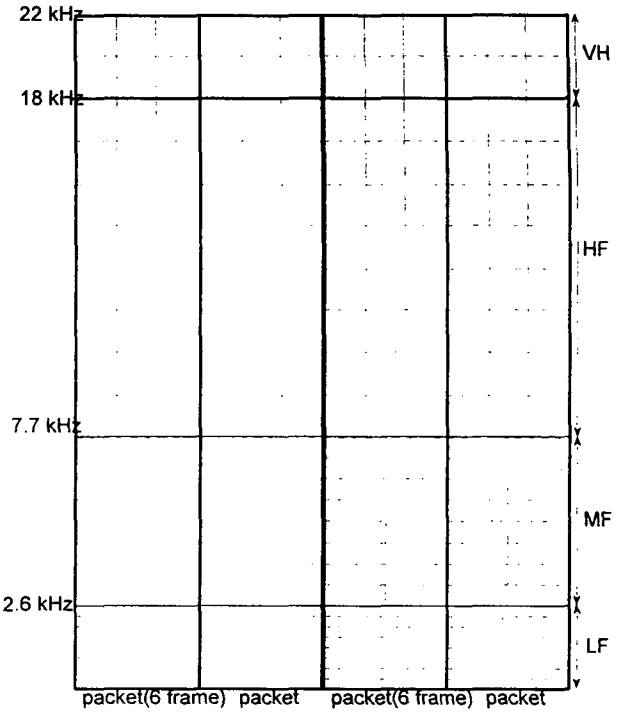


그림 4. 패킷 구조

에서 제안한 심리음향 모델-1을 기반으로 시간영역 마스킹과 함께 계산량을 고려한 새로운 심리음향 모델을 구성하였다.

시간영역 마스킹은 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 100~200msec 정도로 길고 상당히 큰 값을 갖기때문에 지각적으로 중요한 역할을 한다. 후마스킹의 지속 시간은 마스커의 지속 시간이 길수록 더욱 큰 마스킹 효과를 갖으며 200msec이상의 지속 시간을 갖는 경우는 거의 동일한 값을 갖는다[7]. 지연 시간은 마스커와 마스키(maskee) 간의 시간 차이로 약 100msec 이내일 때 더욱 효과적이다. 본 연구에서 한 프레임의 길이는 5.6msec, 패킷의 길이는 약 20~30msec로 충분히

표 1. 대역 특성

properties / band	Subblock bin (t*f)	Frequency resolution	Time resolution	Critical Band number(CB/SB)
LF (0~2584Hz)	6*3	260Hz	34.8msec	14(1.4)
MF (~7752Hz)	3*6	520Hz	17.4msec	7(0.7)
HF (~18088Hz)	2*12	1040Hz	11.6msec	3.5(0.35)

작은 고정된 지연 시간을 갖으므로 주파수와 지속 시간을 변수로 사용하여 후마스킹 성분을 효과적으로 이용할 수 있다.

시간 영역 마스킹은 주파수, 음압 레벨, 지연 시간 등의 여러가지 변수들에 의해 영향을 받으므로 정확한 값을 얻기는 매우 어렵다. 그러나 실제 각 프레임 간의 시간 차인 지연시간이 길고 음압레벨에 의한 차이는 매우 작으므로 본 논문에서는 가장 큰 영향을 받는 값인 주파수만을 고려하여 시간 영역 마스킹 값을 얻었다[6]. 프레임간의 지연시간에 의한 음압 에너지 감쇄율은 간단히 다음 식으로 표시할 수 있다[7].

$$f(t)=e^{-t/\tau} \tag{3}$$

위 식에서 t값은 프레임 간의 지연 시간이며 τ는 각 주파수에의 시간 상수이다. t는 약 5.8msec이고 τ는 저주파 대역에서는 300~30msec 그 외에 대역에서는 거의 일정한 값을 가지므로 25msec로 고정된 값을 사용하였다[7][8]. 위 식에 의해 계산된 음압레벨이 다음 프레임에 전파되어진다. 비트 할당에 필요한 마스킹 곡선을 얻기 위해서 각 부블럭으로 분할된 신호의 스펙트럼 성분으로부터 순음과 비순음으로 분류한 후 각각의 개별 마스킹 임계값을 구한다. 이 값과 절대 가청 한계를 고려하여 최종적인 주파수 마스킹 값을 얻었다.

심리음향 모델의 계산량은 전체 부호화기를 위한 계산량의 약 50%를 차지할 정도로 많은 계산량을 필요로 한다. 본 연구에서는 패킷 당 6번의 계산량을 필요로 하는 심리음향 모델의 계산량을 효과적으로 줄일 수 있는 방법을 사용했다. 6 프레임의 FFT 스펙트럼 값을 각 부블럭 별로 분할한 후 스펙트럼의 시간 및 주파수 특성을 비교하여 심리음향 모델을 적용하였다. 이와 같은 방법을 사용할 경우 FFT에서는 동일한 계산이 필요하게 되지만 그 외에 부분에 있어서는 MPEG의 심리음향 모델-1에 비해서 6 : 1 정도로 계산량을 현저히 줄일 수 있었다.

4. 비트 할당

비트 할당은 부가 정보에 사용되는 비트를 제외한 사용 가능 비티를 각 부블럭에 대해서 지각적으로 최적이 되도록 할당하는 것을 의미한다. 그 과정은 심리음향 모델의 결과인 마스킹 임계값고 각 부블럭에서의 신호 크기, 부블럭의 중복성(sb-ID)을 고려하여 각 부블럭에 할당될 비트를 결정한다. 이때 각 부블럭에서의 최종적인 마스크대잡음비(MNR : Mast to Noise Ratio) 값이 0dB를 넘었을 경우 주관적으로 동일한 복원음을 얻을 수 있다고 가정한다. 비트 할당 과정은 다음과 같다.

먼저 심리음향 모델의 결과인 마스킹 임계값과 각 부블럭의 신호 크기인 비인 SMR을 구한다. MNR 값이 0dB를 넘었을 경우 주관적으로 동일한 복원음을 얻을 수 있다고 가정한다. 비트 할당 과정은 다음과 같다.

먼저 심리음향 모델의 결과인 마스킹 임계값과 각 부블럭의 신호 크기의 비인 SMR을 구한다. MNR 값이 0dB 이상이 되기 위한 최소 비트를 계산한다. 이 값에서 각 부블럭의 중복성을 고려하여 각 블럭에 할당된 비트를 줄이게 된다. 최종적으로 남은 비트를 고려하여 각 부블럭에 비트를 가감해주게 된다. 각 부블럭에 필요한 비트를 할당하기 위해 심리음향 결과로부터 얻어진 SMR 값으로부터 지각적으로 동일한 복원음을 얻기 위해 필요한 총 비트(total-bit)는 다음식에 의해 계산한다.

$$total-bit=\sum_{n=1}^{60} sbblk-bit_n \tag{4}$$

여기서 $sbblk-bit_n=\frac{SMR}{6}-18*\mathrm{sb}-ID_n$ for n=1…30

$sbblk-bit_n=\frac{SMR}{6}$ for n=31…60

위 식에서 *sb-ID*는 각 부블럭의 중복성을 나타내는 값으로 시간 및 주파수 영역에서의 스펙트럼 평탄도를 나타내는 값이 된다. 이 값은 각 부블럭의 dynamic range를 결정하는 것으로 시간-주파수 영역에서 차등 이익을 가리킨다. 이와같이 각 부블럭에 먼저 비트를 할당하고 사용 가능한 비트가 total-bit 값보다 클 경우 남는 비트를 저주파 대역밴드로부터 차례로 할당해 준다. 만약 비트가 모자라는 경우는 각 서브밴드의 중요도에 따라 비트를 줄여준다. 중요도를 나타내는 가중 함수는 초기 비트 할당이 끝난 후의 각 부블럭 MNR 값이 된다.

본 눈문에서는 각 부블럭을 청각적인 가중 함수를 두어 먼저 모든 비트를 할당하고 후에 변형하는 방법을 사용하였다. 위와 같은 방법은 비트 할당에 사용되는 반복적인 계산도 어느 정도 줄일 수 있는 장점이 있다. MPEG에서는 최소 MNR 값을 갖는 부블럭을 찾아 비트를 할당한 후 다시 MNR을 계산하여 반복적으로 서브밴드에 비트를 할당하는 방법을 사용하였다. 이 방법은 반복 계산을 필요로 하므로 많은 계산량을 필요로 한다.

5. 양자화

일반적으로 짧은 구간의 오디오 신호, 즉 그 구간 내에서 어느 정도 안정하다고 가정할 수 있는 신호를 주파수 변환한 DCT 계수들은 시간 및 주파수 중복을 의미하는 인접 계수간의 유사도를 이용하여 보다 효율적인 압축을 행하였다. 이러한 방법은 MPEG의 오디오 2계층에서는 부가 정보 부호화(scalefactor coding)에 시간 영역 중복성 제거를 위해 사용되어지며 Dolby AC-3에서는 지수 부호화에서 주파수 영역 중복성 제거를 위해 사용되고 있다.

저주파 대역과 중간 주파수 대역에서는 시간-주파수 영역에서의 상관 관계에 따라 각 부블럭에 사용될 비트가 달라진다. 각 부블럭의 평균 에너지(seed) 값에서 각 계수와의 차를 계산한 후 이 값들의 음압 레벨이 원 신호의 음압 레벨에 비해 이득이 있을 경우 차등 부호화를 사용한다. 이 때 평균 에너지만을 고정 비트로 양자화하고 나머지 계수는 인접한 계수와의 차를 부호화한다. 본 연구에서는 먼저 seed만을 부호화 및 복호화하여 양자화 오차가 다른 계수들로 전파되는 것을 방지하였다. 또한 seed 계수는 양자화 오차와는 연관이 없으므로 4비트 정도의 적은 비트로 부호화해도 충분한 효과를 가져올 수 있다.

고주파 대역의 신호는 시간 및 주파수 영역에 있어서의 중복성

이 매우 작으므로 신호자체의 중복성을 이용하는 위와 같은 방법이 큰 효과를 얻지 못한다. 그러나 청각적인 면에서 보면 음압 레벨도 작고 절대 한계값이 크기 때문에 중요도 면에서는 떨어진다. 또한 임계 대역폭이 넓기 때문에 주파수 해상도가 떨어져 음의 인식에 있어서도 스펙트럼의 세밀한 구조보다는 전체적인 형태에 의해 좌우되는 경우가 대부분이다. 따라서 주파수 영역에서 지각적인 중복성을 갖게된다. 각 부블럭 내의 에너지를 보존하면서 할당된 비트에 따라 주파수 스펙트럼의 형태를 결정하는 방법을 사용하였다. 고주파 대역의 부블럭에서는 심리음향 모델의 결과인 SMR값에 따라 비트 할당을 달리하였다. 또한 비트가 할당되지 않은 부블럭에도 신호의 에너지(seed)만을 전송한 후 백색 잡음을 재생하여 심리적인 음질 향상을 고려하였다.

부가적으로 전송되어야 하는 정보에는 각 부블럭에서의 비트 할당값과 scalefactor가 있다. 비트 할당 값은 각 대역의 특성에 맞는 비트 할당 표를 사용하여 부가적으로 비트를 감축하였다. 또한 부가 정보에 들어가는 비트 할당, scalefactor 등에 있어서도 시간적인 중복성이 존재한다. 간단한 DPCM을 사용하여 부가 정보에 사용되는 비트를 줄였다.

Ⅳ. 음질 평가

복원된 음질의 정확한 평가는 고음질 오디오 부호화기의 성능 평가와 직결되는 매우 중요한 영역이다. 음질 평가 방법은 청취자가 직접 들어서 평가하는 방법과 원음과 복원음과의 왜곡 정도를 측정하는 방법이 있다[6]. 가장 정확한 음질 평가방법은 청취자가 직접 들어서 평가하는 MOS가 있다. 그러나 MOS평가의 경우 청취자의 구성, 청음 환경 등에 큰 영향을 받으므로 실험실 환경에서 정확한 음질 평가를 위한 환경을 마련하기는 매우 어렵다. 특히, 청각 특성을 이용한 오디오 부호화기의 경우는 복원음과 원음과의 구별이 어려우므로 청취자가 전문가로 구성되지 않는 경우에는 무의미한 결과가 얻어질 수 있다.

따라서 본 연구에서는 원음과 복원음의 오차를 이용한 평가방법만을 이용하였다. 이러한 객관적 평가 방법에는 SNR, SNRseg, Spectral distance measure, MNR, Masking flag 등이 있다. 본 논문에서는 주관적 방법과의 상관도가 비교적 높은 segmental SNR과 청각 특성을 이용한 고음질 오디오의 평가에 많이 사용되는 MNR을 이용하여 음질을 평가하였다. 그러나 MNR에 있어서는 사용된 심리음향 모델이 같아야 한다는 전제가 되어져야 한다. 본 연구에서는 MPEG의 심리음향 모델-1을 기반으로 시간 영역 마스킹 등의 일부만 변경하였으므로 비교적 정확한 음질을 반영해 줄 수 있을 것이다. 음질 평가에 사용된 데이타는 EBU의 SQAM(Sound Quality Assessment Materials)과 음악 CD로부터 얻었으며 비트율은 96kbps로 고정하여 MPEG Layer-2 방법과 비교하였다. segmental SNR의 결과는 표 2와 같다. 대부분의 데이타에서 MPEG과 비슷한 결과를 얻을 수 있었다. 그러나 심리적인 음질과 상과 관계가 높은 MNR에 있어서는 MPEG Layer-2와 비교하여 약간 향상된 성능을 나타내었다. 이는 각 부블럭의 중복성이 효과적으로 제거되었음을 반영해 줄 수 있다.

표 2. 구간 신호대잡음비(96kb/s)

	speech (Female)	Pop	Wind Ensemble	violin
MPEG layer-2	36.5	24.1	27.0	32.3
Proposed	35.4	23.1	29.5	34.1

표 3. 마스크대잡음비(96kb/s)

	speech (Female)	Pop	Wind Ensemble	violin
MPEG layer-2	32.2	7.1	20.4	19.3
Proposed	34.3	8.9	23.0	21.5

Ⅴ. 결론

본 논문에서는 고음질 오디오 부호화 방법에서 시간 영역 중복성을 함께 제거하기 위해 6개의 프레임을 패킷으로 구성하여 청각 특성에 따라 분할하고 이러한 부블럭 내에서 시간 및 주파수 중복성을 제거할 수 있는 새로운 오디오 부호화 방법을 제안하였다. 또한 고음질 오디오 부호화기의 성능에 중요한 영향을 끼치는 심리음향 모델에서 시간 영역 마스킹을 고려하였다. 그리고 마지막으로 많은 계산량을 필요로 하는 심리음향 모델과 비트 할당 과정에 있어서도 신호의 특성을 반영하여 계산량을 줄이는 방법을 제안하였다. 제안된 방법은 동일한 비트율에서 구간 신호대잡음비와 마스크대잡음비에 있어서 MPEG Layer-2와 비슷하거나 약간 향상된 결과를 얻을 수 있었다.

참 고 문 헌

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

[2] N. Jaynat, J. Johnston, and R. Safrnek, "Signal Compression Based on Models of Human Perception," *Proc. IEEE* Oct. 1993.

[3] M. Paraskevas and J. Mourjopolos, "Results of a differential perceptual audio coding technique."

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

[5] ISO-IEC JTC1/SC29/WG11 "Coding of Moving Pictures and Associate Audio for Digital Storage Media at up to about 1.5 Mbps-CD 11172(Part-3, MPEG-Audio)," Nov. 1991.

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

[7] J. G. Beerends, J. A. Stemerdink, "A Perceptual Audio

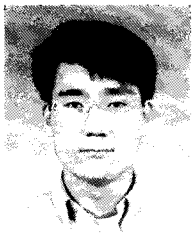
Quality Measure Based on a Psychoacoustic Sound Representation" *J. Audio Eng. Soc.*, Dec. 1992.

[8] W. Jesteadt, S. Bacon, J. Lehman, "Forward masking as a function of frequency, masker level, and signal delay," *J. Acoust. Soc. Am.*, Apr. 1982.

[9] R. N. Veldhuis, "Bit rates in audio source coding," *IEEE J. Selected Areas Comm.*, pp. 86-96, 1992.

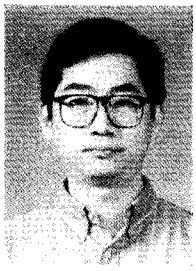
[10] M. Iwadare, *et al.* "A 128 kbit/s hi-fi audio codec based on adaptive transform coding with adaptive block size MDCT." *IEEE J. Selected Areas Comm.*, pp. 138-144, 1992.

저 자 소 개



金 基 秀
1969년 1월 19일생.
1991년 2월 연세대학교 전자공학과 졸업.
1993년 2월 연세대학교 대학원 전자공학과 졸업(공학석사).
1993년 3월 ~ 현재 연세대학교 전자공학과 박사과정

徐 皓 善
1957년 8월 17일생.
1981년 2월 연세대학교 전자공학과 졸업.
1983년 2월 연세대학교 대학원 전자공학과 졸업(공학석사).
1993년 2월 연세대학교 전자공학과 졸업(공학박사).
1993년 9월 ~ 1994년 9월 한국과학기술원Post doc.
현재 연세대학교 신호처리연구센타 연구원



異 準 容
1962년 3월 2일생.
1989년 2월 연세대학교 전자공학과 졸업.
1991년 2월 KAIST 전기 및 전자공학과 졸업(공학석사).
1991년 3월 ~ 현재 한국 방송공사 기술연구소 연구원

尹 大 熙
1951년 5월 25일생.
1977년 2월 연세대학교 전자공학과 졸업.
1979년 8월 Kansas State Univ. 공학석사
1982년 8월 Kansas State Univ. 공학박사 (Dept. of Electrical Eng.)
1982년 8월 ~ 1985년 6월 Univ.of lowa Assistant Professor
1985년 9월 ~ 현재 연세대학교 전자공학과 교수