

디지털 오디오에서 신호처리 기술의 응용

김 기 수 · 윤 대 희

(연세대학교 전자공학과)

■ 차 례 ■

I. 서 론

II. 심리음향 모델

III. 고음질 오디오 부호화기

IV. 고음질 오디오 부호화기의 연구 현황

V. 결 론

I. 서 론

1877년 T. Edison이 축음기를 발명한 후 1980년대에 CD(Compact Disk)와 DAT(Digital Audio Tape)의 등장으로 오디오 기술은 새로운 도약을 하였다. 즉 기존의 아날로그 오디오의 한계를 뛰어넘은 고음질의 디지털 오디오를 감상할 수 있게 된 것이다. 또한 MIDI(Musical Instrument Digital Interface)와 디지털 피아노를 비롯한 디지털 악기의 보급으로 주위에서 쉽게 디지털 오디오에 접할 수 있게 되었다. 디지털 신호는 아날로그 신호에 비해 반 영구적이고 신뢰성이 크며 복사할 때 원음에 비해 음질이 떨어지지 않는 장점을 가지고 있다. 그러나 CD나 DAT와 같이 20 kHz 정도의 대역에서 16 비트 PCM(Pulse Code Modulation) 방식으로 음을 부호화하면 약 700 kbit/s의 많은 정보량이 필요하다. 따라서 CD 정도의 음질을 필요로 하는 HDTV(High Definition Television)나 멀티미디어(multimedia)의 오디오 신호로 사용하는데 어려움이 있다.

80년대에 들어 사람의 청각 특성과 디지털 신호 처리 기술을 이용한 새로운 오디오 신호 압축 기술(HQAC: High Quality Audio Coding)이 연구되어 졌다[1][2][3][4]. 이 방법은 심리음향 모델을 사용해서 얻어진 마스킹 곡선을 이용하여 원음이 양자화 잡

음을 완전히 마스킹(masking)하도록 비트 할당을 한다. 마스킹 현상을 이용하기 위해서는 서브 밴드 부호화(subband coding)나 변환 부호화(transform coding)와 같은 주파수 영역에서 부호화하는 기술이 필요하다. 약 64~192 kbit/s의 저장 또는 전송용에서도 원 신호와 지각적으로 동일한(perceptually transparent) 복원 신호를 얻을 수 있다. 따라서 본고에서는 이러한 오디오 신호 압축 방법에 대해 알아보고 최근의 연구 현황을 살펴본다.

II. 심리음향 모델(psychoacoustic model)

사람의 귀는 시간 영역의 오디오 신호를 필터 बैं크와 같이 주파수 영역의 각 대역으로 나누어서 음을 인식하고, 그 대역은 임계 대역(critical band)이라 불리워진다. 임계 대역을 실제 측정할 때는 두 순음의 선택성에 따라 나눌 수 있는데 임계 대역 내에 두 순음이 존재할 때는 같은 청신경을 자극하므로 구별할 수 없다는 성질을 이용하여 측정한다. 임계 대역은 대략 24~26개로 나누어지며 표 1은 위의 방법으로 측정된 임계 대역을 나타낸다. 임계 대역은 낮은 주파수에서 세밀하고 높은 주파수로 갈수록 넓어지는 로그 함수적인 특성을 갖는다는 사실을 보여준다.

음의 인식은 조용한 환경에서도 어떠한 임계값을

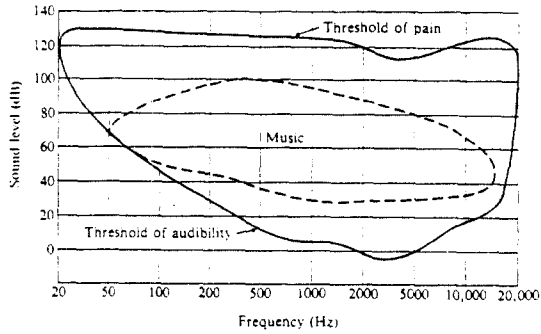


그림 1. 절대 가청 한계(threshold of audibility)

표 1. 일계 대역

Rate [Barks]	f_l [Hz]	f_h [Hz]	f_c [Hz]
1			100
2	100	100	200
3	100	200	300
4	100	300	400
5	110	400	510
6	120	510	630
7	140	630	770
8	150	770	920
9	160	920	1080
10	190	1080	1270
11	210	1270	1480
12	240	1480	1720
13	280	1720	2000
14	320	2000	2320
15	380	2320	2700
16	450	2700	3150
17	550	3150	3700
18	700	3700	4400
19	900	4400	5300
20	1100	5300	6400
21	1300	6400	7700
22	1800	7700	9500
23	2500	9500	12000
24	3500	12000	15500

f_l : low band boundaries

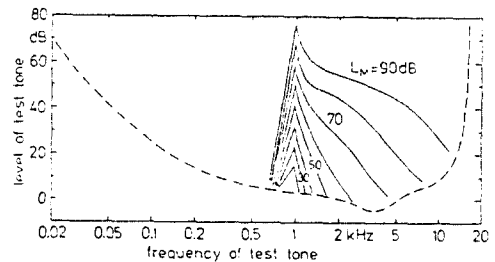
f_h : high band boundaries

넘는 순간에서부터 이루어 지는데 이 값을 절대 가청 한계(threshold of audibility)라고 한다. 이 값은 그림 1 과 같이 주파수에 따라 달라진다. 그림에서 볼 수 있

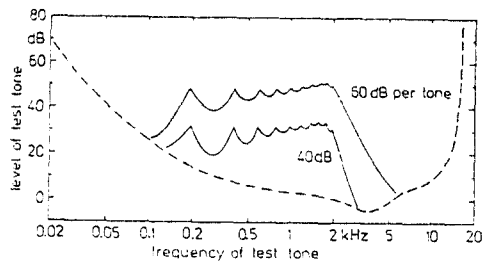
듯이 저주파와 고주파 대역에서는 큰 값을 갖고 약 2 kHz에서 5kHz의 중간 주파수 대역에서는 아주 작은 값을 갖는 것을 알 수 있다. 이것은 외이의 공명 주파수가 약 3kHz이고 이 부분에 음성의 대부분의 정보가 모여 있는 것과 일치한다.

청각 특성을 이용한 고품질 오디오 부호화기에서 사용되는 가장 중요한 청각 특성은 마스킹 현상(masking effect)이다. 마스킹이란 큰 에너지를 갖는 음이 존재할 때 작은 에너지를 갖는 음이 들리지 않는 현상을 가리킨다. 예를 들면, 어떤 주파수 대역에 순음이 존재하거나 잡음이 존재할 때 주위 대역의 작은 신호들이 들리지 않게 된다. 따라서 절대 가청 한계와 신호의 마스킹 곡선을 이용하면 양자화 잡음이 들리지 않는 전체적인 마스킹 곡선을 얻게 된다.

마스킹 곡선을 구하기 위해서 특정 주파수 대역에 일정한 음압 레벨(SPL : Sound Pressure Level)을 갖는 순음이나 잡음을 발생시킨 후 각 주파수 영역에서 나타날 순음을 발생시켜 들리게 되는 순간의 음압 레벨을 연결한 곡선이다. 그림 2 (a)는 1kHz에 순음이 존재할 때 각 레벨에 따른 마스킹 곡선을 나타낸다. 일반적으로 마스킹 곡선은 마스크(masker)의 음압 레벨이 40dB를 넘을 때에는 고주파 쪽에서 더 넓은 마스킹



(a)



(b)

그림 2. 마스킹 현상

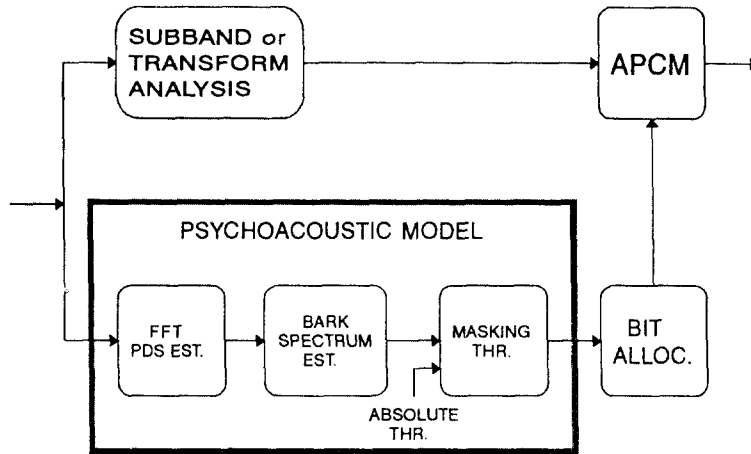


그림 3. HQAC의 기본 구조

곡선을 가짐을 알 수 있다. 그림 2 (b)는 200Hz에서 2 kHz까지의 10개의 순음들에 의해 만들어진 마스크링 곡선을 나타낸다. 그림 2 (b)에서와 같이 마스크가 여러 개 존재할 때의 마스크링 곡선을 각각의 순음에 대한 마스크링 곡선이 서로 더해졌을 때와 비슷함을 알 수 있다[5].

Ⅲ. 고품질 오디오 부호화기

HQAC는 음의 지각 특성을 이용하여 들을 수 없는 부분을 부호화하지 않거나 잡음이 마스크링 되도록 각 대역의 비트를 할당하게 된다. 따라서 객관적인 SNR보다는 주관적인 SNR을 고려하여 부호화하므로 실제 SNR이 약 20dB 정도의 낮은 비트율에서도 음에 따라서는 원음과 동일한 복원음을 얻을 수 있다[6]. 고품질 오디오 부호화기의 기본 구조는 그림 3과 같다.

그림 3에서 알 수 있듯이 HQAC의 부호화 방법은 서브밴드 부호화와 변환 부호화로 나누어 질 수 있다. 변환 부호화의 경우 각 계수들을 가지고 모든 변환 계수에 가중 벡터를 곱해서 마스크링 임계값을 구한다. 변환 부호화에서 블럭 크기를 너무 작게 하면 블럭과 블럭간의 잡음(block edge effect)이 생기며 너무 클 때는 프리에코 현상(pre-echo effect)이 생기므로 적당적으로 블럭 크기를 조절해야 한다[4]. 서브밴드 부호화의 경우에는 일반적으로 더 세밀한 주파수 분석을 위해 부가적인 FFT(Fast Fourier Transform)를 통

해서 스펙트럼을 구한다. FFT 스펙트럼은 임계 대역의 단위인 Bark로 변환된 후 마스크링 임계값을 구한다.

이와 같이 구한 마스크링 임계값을 가지고 각 변환 계수나 서브밴드에 비트할당을 하게 된다. HQAC에 사용되는 필터 뱅크는 다음과 같은 조건을 만족해야 한다[6].

- 1) 분석 합성 시에 완전 복원(perfect reconstruction) 조건을 만족해야 한다. 이 조건을 부호화할 때 발생하는 모든 잡음이 양자화 잡음임을 의미하므로 잡리음형 모델에 맞게 양자화 잡음을 제어할 수 있다.
- 2) 통계적으로 효과적인 압축 효율을 얻으려면 많은 수의 필터 뱅크가 필요하다. 그러나 뱅크가 너무 많아지면 시스템이 복잡해지므로 적당한 뱅크 수를 선택해야 한다. 일반적으로 한 대역의 크기는 귀의 최소 임계 대역과 같은 크기로 맞추어준다.
- 3) 주파수 영역에서의 잡음을 시간 영역으로 전파되므로 시간 영역에서 블럭의 크기는 충분히 작아야 한다. 블럭이 너무 클 때 시간 영역으로 전파된 잡음을 프리에코라 한다.

그림 4에 간단한 잡리음형 모델을 사용해서 마스크링 임계값을 구하는 과정을 나타냈다. 이 방법은 MPEG (Motion Picture Expert Group)에서 제안한 MUSICAM (Masking pattern Adapted Universal Subband Integrated Coding And Multiplexing) 방식에 사용된 잡리음형 모델 1로서 FFT 스펙트럼을 가지고 순음 성분과 임계 대역 내의 잡음 성분으로 나누어 각각에 대해

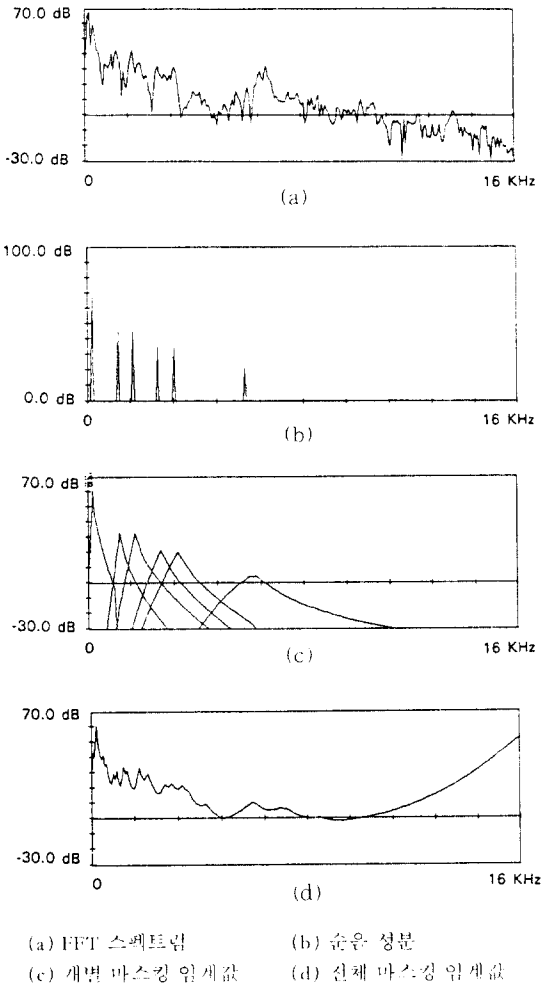


그림 4. 심리음향 모델의 예

개별 임계값을 구한다. 그리고 절대 가청 한계를 고려하여 전체 마스크 임계값을 구한다[7].

IV. 고음질 오디오 부호화기의 연구 현황

현재 청각 특성을 이용한 고음질 오디오 부호화 시스템에는 개발이 완료된 소니사의 MD(Mini Disk), 필립스사의 DCC(Digital Compact Cassette)가 있고 그 외에 ISO/MPEG(Motion Picture Expert Group of International Standard Organization)에서 제안한 MUSICAM과 PEXM(Perceptual transform coding), OCF(Optimum Coding in Frequency domain) 등이 있다[6][8]

[9]. 다음에 각각에 대해서 간략하게 알아본다.

1. MUSICAM

MPEG에서는 영상 부호화를 포함하여 약 1.5 Mbit/s에서 CD 정도의 음질을 얻을 수 있는 고품질 오디오 부호화 방식으로 Philips, IRT, CCETT 등에 의해 개발된 MUSICAM 방식을 채택하였다[10][11]. MUSICAM 방식은 복잡도와 성능에 따라 세개의 계층으로 나뉘어 있는데 각각에서 64~192kbit/s의 비트율을 가지고 고품질을 제공한다.

MUSICAM 시스템의 기본 구조는 그림 5와 같다. 부호화기에서는 32밴드의 필터뱅크를 사용하여 서브밴드 부호화를 하며 부가적인 FFT를 통해 얻어진 마스크 임계값을 얻는다. 각 서브밴드에서 최소 마스크 임계값을 구하여 비트할당을 한다. 32개의 서브밴드 신호는 부가 정보와 같이 압축된 비트열을 구성한다[2][12].

부호화기는 심리음향 모델이 사용되지 않으므로 간단히 구성될 수 있다. 비트열을 풀고 난 후 합성 필터뱅크를 통과시켜 복원된 PCM 샘플을 얻는다.

2. MD

소니(Sony Corp)사에서 MD는 기존 CD의 절반 크기인 64mm의 크기에 CD의 단점인 재생 전용이라는 문제점을 광자기 디스크(magneto-optical disk)를 사용하여 해결하였다. 또한 충격 방지 기능까지 갖추고 있으며 약 74분간 녹음과 재생이 가능하다. 표본화 주파수는 44.1kHz이고 16bit PCM 방식으로 음을 부호화한다.

사용된 데이터 압축 방법은 Atrac(Adaptive transfer acoustic coding)으로 약 5:1의 압축율을 얻을 수 있다. 20msec 정도의 신호(1024 샘플)를 블록으로 한 변환 부호화의 계수들을 사용해서 마스크 임계값을 구한다.

3 DCC

필립스(Philips Consumer Electronics)사에서 개발한 DCC는 기존의 아날로그 카세트 테이프와 호환 가능하다는 장점을 가지고 기존 LP 레코드를 CD가 대체해 가고 있듯이 아날로그 카세트 테이프를 대체할 것으로 기대된다. 표본화 주파수는 32kHz, 44.1kHz, 48kHz 이고 약 108dB(18bit)의 dynamic range를 제공한다. 표 2에 MD와 DCC를 비교하였다.

PASC(Precision Adaptive Subband Coding)라는 고

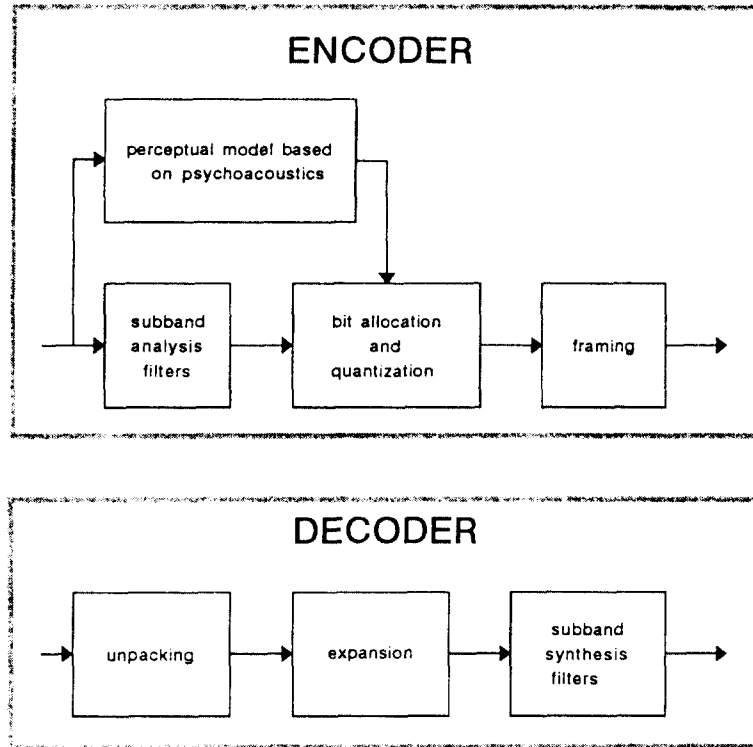


그림 5. MUSICAM 시스템의 기본 구조

음질 오디오 부호화 방법을 사용하여 약 4:1의 압축율을 얻는다. PASC는 절대 가장 한계와 각각의 서브밴드 신호를 이용한 마스킹 임계값을 더하여 전체 마스킹 임계값을 구한다. 부가적인 FFT를 하지 않고 시

프랜트 신호만 가지고 마스킹 임계값을 구하므로 정확한 마스킹 임계값을 얻을 수 없으므로 큰 압축율을 얻을 수 없다. 약 192kb/s에서 고음질의 오디오를 제공할 수 있다[13].

표 2. MD와 DCC의 비교

항 목	MD(Mini Disk)	DCC (Digital Compact Cassette)
표본화 주파수	44.1 kHz	32, 44.1, 48 kHz
dynamic range	96 dB (16 bit)	108 dB (18 bit)
압축 방법	Atrac (Adaptive transfer acoustic coding)	PASC (Precision Adaptive Subband Coding)
압축비	5:1	4:1
매체 크기	2.5 inch disk	3.78 mm wide tape
기록 시간	74분	90분
저장매체	magneto-optical disk	chrome tape
호환성	CD와 호환 불가능	카세트 테이프와 호환

4. PFXM and OCF

Johnston이 제안한 PFXM은 FFT를, Brandenburg가 제안한 OCF는 MDCT(Modified Discrete Cosine Transform)를 사용한 HQAC이다[1][3]. 변환 계수를 사용하여 마스크 임계값을 얻는다. 양자화는 모두 허프만 부호화기(huffman coder)를 사용하며 PFXM의 경우 스테레오 신호간의 중복(redundancy)을 이용하여 부가적으로 오디오 신호를 압축할 수 있다. 약 2.5~3 bit/sample에서 고품질을 제공한다.

IV. 결 론

지금까지 청각 특성을 이용한 고품질 오디오 압축 방법과 연구 현황에 대해 알아보았다. 현재 디지털 신호처리 기술과 저장 매체의 발달로 점점 기존의 아날로그 방식에서 디지털 방식의 오디오로 바뀌어 가고 있다. LP를 대신해서 CD나 MD가, 아날로그 카세트 테이프를 대신해서 DAT, DCC가 미래의 표준으로 대체 될 것은 명확하다.

국내에서도 HQAC에 대해서 많은 연구가 진행되고 있고 이미 금성사에서는 DCC를 개발하였다고 한다. HQAC는 DBS(Direct Broadcasting Satellite)를 통한 DAR(Digital Audio Radio), HDTV의 오디오부, 오디오 테이타베이스 등 광범위한 응용 영역을 갖는다. 따라서 더욱 압축율을 높이기 위한 새로운 방법들이 개발될 것이다. 이 때 좀 더 압축율을 높이기 위해서 계산면이나 성능면에서 향상된 심리음향 모델의 개발, 새로운 비트 할당 방법과 양자화 방법의 연구 등이 이루어져야 하겠다.

참 고 문 헌

1. K. Brandenburg, "OCF - a new coding algorithm for high quality sound signals." *Proc. ICASSP* pp. 141-144, 1987.
2. Y. F. Dehery, et al. "A MUSICAM source codec for digital audio broadcasting and storage." *Proc. ICASSP* pp. 3605-3608, 1991.
3. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria." *IEEE J. Selected Areas Comm.* pp. 314-323, 1988.
4. 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.
5. E. Zwicker, *Psychoacoustics*. Springer-Verlag, New York, 1982.
6. S. Furui, M. M. Sondhi, *Advances in speech signal processing*. Marcel Dekker, 1991.
7. ISO-IEC JTCl / SC29 / WG11 MPEG-Audio 1991. Nov.
8. R. K. Jurgen, "Consumer electronics" *IEEE Spectrum*, 1992 Jan.
9. N. S. Javant, "High Quality Coding of Telephone Speech and Wideband Audio." *Proc. ICC* pp. 927-931, 1990.
10. R. N. Veldhuis, "Bit rates in audio source coding." *IEEE J. Selected Areas Comm.* pp. 86-96, 1992.
11. 이문호, 이광재 지, 멀티미디어 / 하이퍼미디어의 부호화, 내영사.
12. G. Stoll, and Y. F. Dehery, "High quality audio bit-rate reduction system family for different applications" *Proc. ICC* pp. 937-941, 1990.
13. R. N. Veldhuis, et al. "Subband coding of digital audio signals without loss of quality." *Proc. ICASSP*, pp. 2009-2012, 1989.

 김 기 수

- 1969년 1월 18日生
- 1991년 2월 : 연세대학교 전자공학과
- 1991년 3월 ~ 현재 : 연세대학교 전자공학과 대학원 석사과정 재학중


 윤 대 희

- 1951년 5월 25日生
- 1977년 2월 : 연세대학교 전자공학과 졸업(工學士)
- 1975년 5월 : Department of Electrical Engineering, Kansas State University(M.S.)
- 1982년 5월 : Department of Electrical Engineering, Kansas State University(Ph. D)
- 1982년 8월 ~ 1985년 6월 : Department of Electrical and Computer Engineering, University of Iowa助教授
- 1985년 9월 ~ 現在 : 연세대학교 전자공학과 助教授