

主題

오디오 데이터 부호화 기술

삼성종합기술원 전문연구원 김 상 욱

차 례

- I. 서 론
- II. 심리음향 모델
- III. 심리음향 모델을 이용한 오디오 데이터 압축 방식
- IV. 오디오 데이터 압축 방식 : MPEG
- V. 결 론

I. 서 론

오디오 신호의 저장매체는 아날로그 오디오에서 사용되던 Tape, LP에서, 디지털 오디오에서 사용하는 DAT, CD로 변화되었고, 그 후 디지털 데이터 압축 기술의 발전에 따라, DCC, MiniDisc로 변화되었다. 아날로그 오디오에서 디지털 오디오로 바뀔 때, 열화등에 의해 발생하는 잡음에 대해 강해졌고, 높은 음질, 넓은 대역폭의 오디오 신호 제공 등의 장점은 생겼으나, 아날로그 오디오 신호를 48kHz 또는 44.1kHz로 표본화하여 디지털 데이터인 PCM 데이터를 구하게 됨에 따라서, 데이터 양이 많아지는 점이 있다. 최근 이러한 데이터들의 표현에 사용되는 비트 수를 줄여주기 위한 데이터 압축 방식들에 대한 연구⁽¹⁾⁻⁽⁴⁾가 활발히 진행되고 있는데, 데이터를 압축해 줌으로써 같은 용량의 기록저장매체에 보다 많은 양의 데이터를 저장할 수가 있고, 또 정해진 전송률의 채널을 이용해 전달할 때 정보의 전송에 사용되는 시간이 줄어들게 됨으로써 전송선로의 보다 효과적인

사용이 가능하게 한다.

오디오 압축 기술들은 크게 국제 표준 기술과 시장에서의 표준 기술로 구분될 수 있다.

현재, 국제 표준 기술로는 모노/스테레오 처리 기술인 ISO/IEC 11172-3, HDTV등의 다채널 오디오 처리용 ISO/IEC 13818-3, ISO/IEC 13818-7이 있고, 1998년 12월에 표준 제정을 목표로 하는 ISO/IEC 14496-3이 있다. 시장에서의 표준 기술로는 모노 또는 스테레오 처리 기술로는 Philips의 DCC용 PASC, 일본 Sony의 MiniDisc용 ATRAC 기술이 있고, 다채널 오디오 처리 기술로는 미국 Dolby Lab.의 AC-3가 있다.

본 논문에서는 오디오 신호를 압축할 때 사용되는 심리 음향 모델에 관련해 2장에서 다루고, 3장에서는 심리음향을 이용한 오디오 데이터 압축 방식들을 보인 뒤, 4장에서는 대표적인 오디오 압축 표준 방식인 MPEG 오디오에 대해서 이야기 하고자 한다.

II. 심리음향 모델

심리음향 모델^{(5),(7)}은 최근 개발되는 오디오 데이터 압축 방식들의 핵심 기술이다. 처리에 활용되는 심리음향 모델의 예는 ISO/IEC 11172-3⁽¹⁰⁾에 나와 있는데, 하나는 MUSICAM에서 사용된 방식이고, 다른 하나는 ASPEC에서 사용된 방식이다. 각 방식은 사용된 압축 방식의 독자적인 특징을 고려해 사용되었기 때문에, 여기서는 공통 개념에 대해 서술하기로 한다. 심리음향 모델에서는 어떤 주파수 성분이 존재할 때 근처의 낮은 레벨의 주파수 성분이 느껴지지 않도록 만드는 마스킹 특성과 어느 정도 이하의 음압을 갖는 주파수 성분을 인지하지 못하게 되는 최소 가청 한계 특성을 계산해 이용한다. 마스킹 이라는 것은 예로 지하철을 타기 위해 기다리고 있을 때, 정상시의 목소리 크기로 이야기를 할 경우 들리던 신호가 지하철이 바로 옆을 지나갈 때에는 소음에 묻혀서 같은 크기로 이야기를 할 때 들리지 않는 경우에 대한 것으로, 이런 현상을 지하철의 소리에 의해 대화

하는 소리가 마스킹 되었다고 한다. 그림 1에서는 주파수 B의 신호가 주파수 A의 신호에 의해 마스킹 된 것을 보인다.

심리음향 모델을 이용하기 위해서는 존재하는 프레임의 주파수 영역 신호 레벨들로부터, 각 신호에 의해 마스킹 되는 주파수 영역의 레벨을 표시한 마스킹 곡선을 계산해 낸다. 그 마스킹 곡선과 최소 가청 한계 곡선의 값들 중 큰 값이 그 프레임에서 허용하는 잡음 레벨의 상한을 표시한다. 서브 대역들의 양자화 결과에 의해 생기는 양자화 잡음을 모두 허용하는 잡음 레벨 아래에 위치시키면, 잡음이 검출되지 않게 양자화 처리를 할 수가 있다.

III. 심리음향 모델을 이용한 오디오 데이터 압축 방식

오디오 데이터의 압축시 인간의 심리음향 모델을 고려해준 방식들의 예는 표 1에서 들었다.

그림 1. 최소 가청 한계 값과 마스킹의 일 예.

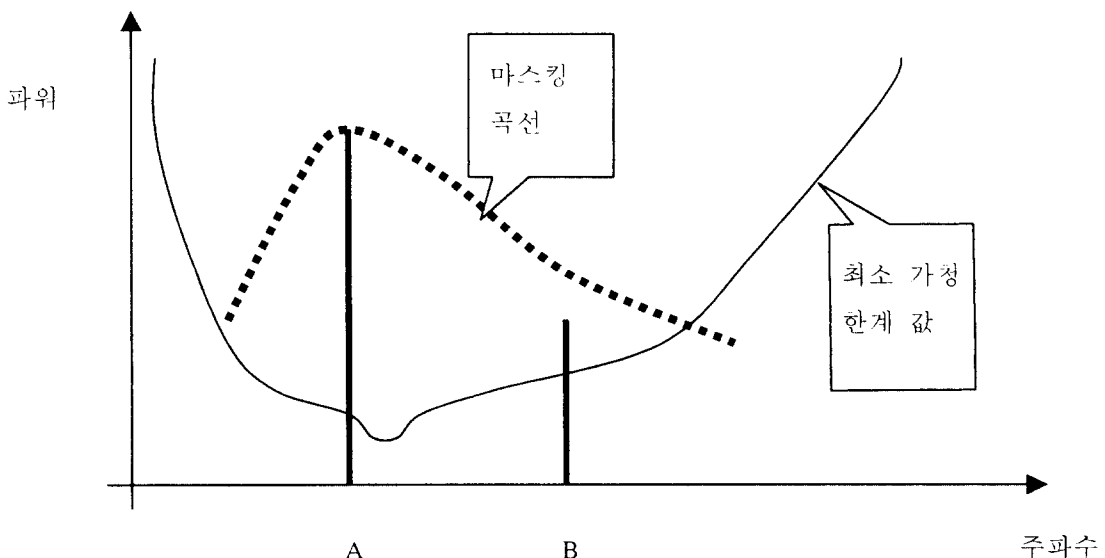


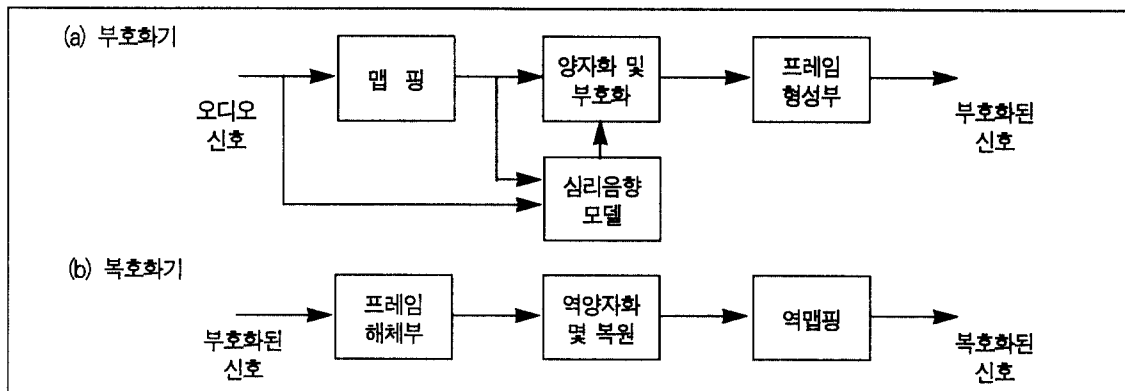
표1. 심리음향 모델을 이용한 오디오 데이터 압축 방식

압축 방식	내 용
ISO/IEC 11172-3	MPEG-1이라 불림. 모노, 스테레오 신호 처리용. 계층 I, II, III으로 구성. 계층 III은 MP3 이라고도 불림.
ISO/IEC 13818-3	MPEG-2 BC임. 다채널 오디오 처리용. 11172-3 과 양립성 제공.
ISO/IEC 13818-7	MPEG-2 NBC 또는 MPEG-2 AAC라 불림. 13818-3의 음질 문제 개선한 다채널 오디오 압축 방식
ISO/IEC 14496-3	MPEG-4 오디오라고 불림 음성부터 오디오에 이르는 다양한 신호의 효과적인 처리를 목적.
MP3	ISO/IEC 11172-3 (MPEG-1)의 계층 III에 대한 기술임.
SoundVQ	NTT사의 Twin-VQ 기술로 14496-3에 TwinVQ 기술이 일부 포함됨
A2b	13818-7의 LC 프로파일에서 저작권 보호 기술 첨가한 것임.
PASC	DCC에서 사용되는 오디오 압축 기술
ATRAC	MiniDisc에서 사용되는 오디오 압축 기술
AC-3	미국방식 DVD에서 사용되는 다채널 오디오 압축 기술.

인간의 심리음향 모델을 이용한 오디오 데이터 압축 기술들은 인간의 소리에 대한 인식 성질들 가운데 공간적인(spectral), 시간적인(temporal) 마스킹 효과를 처리에 고려해 주기 위하여 입력신호를 인간이 느끼는 단위 주파수 대역들로 나누어 주기 위한 맵핑 과정을 갖는다. 맵핑 과정에서는 Pseudo-QMF (Quadrature Mirror Filters) 또는 MDCT (Modified DCT)들이 많이 사용된다. 심리음향 모델에서는 입력신호를 FFT해서, 단위 주파수 대역 신호

의 중요도에 따라서 맵핑된 결과 값들의 표현에 사용될 각 대역의 비트 수들을 계산해 할당해 주는 처리를 한다. 할당된 비트 수에 따라서 맵핑된 신호는 양자화 및 부호화 단계에서, 선형 또는 비선형 양자화 처리 후, 방식에 따라서는 엔트로피 부호화를 거친다. 프레임 형성부에서는 입력신호를 부호화 처리한 결과를 비트스트림으로 만들어 준다. 복호화때에는 역순에 의해 신호를 처리한다. 그림 2에 심리음향 모델을 이용한 오디오 부호화기, 복호화기의 예를 보였다.

그림 2. 심리음향 모델을 이용한 오디오 부호화기, 복호화기의 일 예.



IV. 오디오 데이터 압축 방식 : MPEG

ISO(국제표준위원회)와 IEC(국제전기학회)가 설립한 공동기술협의회 JTC1(Joint Technical Committee 1) 산하의 SC(Sub-Committee) 가운데 SC29에는 영상 압축에 관련된 여러 WG(Working Group)들이 있다. 이 가운데, WG 11은 동화상 압축에 관련된 그룹으로 일명 MPEG이라 한다. MPEG은 1988년 5월 캐나다 오타와에서 제1회 회합을 시작한 뒤, 1992년 11월 MPEG-1 오디오, 1994년 11월에는 MPEG-2 BC 오디오 부분, 1997년 4월에는 MPEG-2 NBC오디오에 대한 국제 표준을 제정하였다. 현재는 1998년 12월 MPEG-4 오디오에 대한 1차 IS가 제정될 예정이고, 2000년 2월에는 MPEG-4에 대한 2차 IS가 제정될 예정이다.^{(4), (8)~(11)}

1. MPEG-1 오디오 (ISO/IEC 11172-3)

MPEG-1 오디오는 32, 44.1, 48 kHz로 표본화된 PCM 오디오 신호를 모노 채널인 경우 32에서 192kbps, 스테레오 채널인 경우 64에서 384kbps로 부호화할 수 있는 표준 방식으로, 동작 모드들은 모노 채널의 경우와 스테레오, 조인트 스테레오, 듀얼 채널들이 있다.

MPEG-1 오디오⁽¹⁰⁾ 부호화 방식에 대한 기술 제안은 1989년에 받았고, 14개의 제안들이 제출되었다. 제출된 제안들은 알고리즘들의 유사성에 의해 ASPEC, ATAC, MUSICAM 과 SB/ADPCM의 4개 그룹으로 나누어졌고 청취 실험이 수행되었다. 방식들 가운데, ASPEC은 64kbps/channel에서 좋은 음질을 보였고, MUSICAM은 구현상의 복잡도와 복호화 처리시의 지연에 있어서 장점을 가지고 있어, MPEG 표준안 위원회는 ASPEC과 MUSICAM들로 오디오 부호화 표준안 초안을 만들기로 하였고, 각 방식들에 근거해 3개 계층의 부호화 알고리즘을 정의하게 되었다.

계층 I과 II에는 MUSICAM에 근거한 방식으로, 맵핑 부분에 Pseudo-QMF를 사용한다. 32개 대역으

로 나누어진 신호들은 각 대역 상호간의 신호들의 파워에 의한 마스킹과 절대 마스킹 값들을 고려해 사용 가능한 비트 수로부터 비트들을 할당 받는다. 할당된 비트 수들에 의해 선형 양자화 처리가 양자화 및 부호화부에서 행해진다. 양자화 처리가 끝난 비트들은 계층 I과 II가 갖는 비트스트림 양식에 맞추어 프레임 형성부에서 비트스트림을 형성한다. 계층 II의 경우에는 단위 시간별로 나누어진 연속된 프레임간의 정보를 표현하기 위해서 사용된 부가정보의 중복성을 줄여주어 보다 처리에 효율성을 기한 것이 특징이다. 그림 3과 4에서 계층 I, II의 프레임 형식을 보인다.

그림 3. ISO/IEC 11172-3 계층 I의 프레임 형식.

Header	Alloc	Scale factors	Samples	Anc

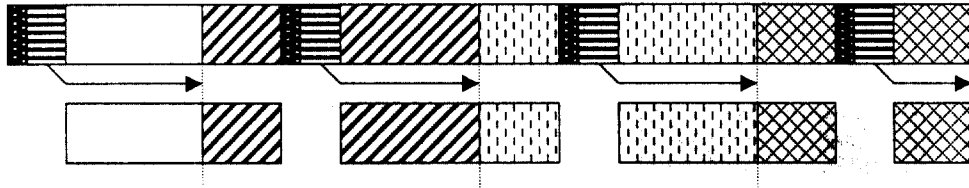
그림 4. ISO/IEC 11172-3 계층 II의 프레임 형식.

Header	Alloc	scfsi	Scale factors	Samples	Anc

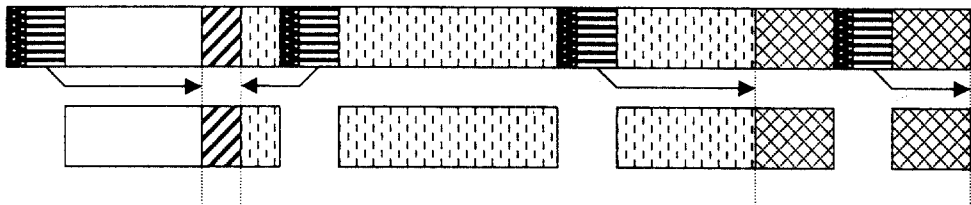
계층 III은 ASPEC에 근거한 방식으로 현재 MP3로 더 많이 알려진 기술이다. 계층 III은 맵핑 부분에서 Pseudo-QMF에 의한 계산결과에 MDCT 연산을 추가해 주파수 영역의 해상도를 높여 준다. MDCT 연산을 하는데 있어서, 입력신호의 특성을 고려해, 처리시 사용되는 윈도우의 크기를 변형시켜 줌으로서, 시간과 주파수 영역에서의 해상도를 조절해주어 계층 I과 II에서는 고려되지 않았던 프리에코에 대한 처리를 해주고, MDCT 결과 발생하는 알리아싱에 대한 보상을 해준다. 심리음향 모델에서는 맵핑 결과값들과 입력 값을 이용해 허용 가능한 각 단위 대역에서의 오차 값을 구해 주고, 양자화 및 부호화부에서는 각 대역에서 비선형 양자화 처리 후 허프만 부호화 했을 때 발생하는 양자화 잡음이 주어진 문턱치 값 아래가 되도록 처리를 하는 잡음 할당 처리를 한다. 계층 III은 각 처리 단위 시간에 사용되는 비트 수들을 버퍼제어에 의해서 달리 갖음으로써 신호의 특성에 따라서 비트들을 보다 효과적으로 사용 할 수 있도록 하였다.

그림 5. ISO/IEC 11172-3 계층 Ⅱ의 데이터 형식.

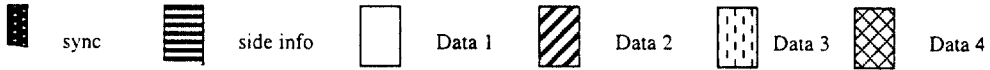
(a) 일반적인 경우의 예



(b) 주정보에서 최대, 최소의 비트가 사용되는 경우의 일 예.



(b)



여기서, 그림5의 (b)는 데이터 2의 표현에 최소의 비트들을 사용하고, 데이터 3의 표현에 최대의 비트 수를 사용하는 경우에 대한 일 예를 보인다.

2. MPEG-2 오디오

(ISO/IEC 13818-3, ISO/IEC 13818-7)

ISO/IEC 13818-3⁶⁹⁾은 ISO/IEC 11172-3을 기본 골격으로 하면서, 11172-3과의 양립성을 제공한다. 13818-3은 현장감 있는 음악의 재생을 위한 3/2채널 오디오 데이터의 압축 방식으로, 낮은 표본화 주파수(16, 22.05, 24kHz)에서는 중간급 음질을 지원하고, 8개국 언어까지 동시에 다중언어방송이 가능하다.

13818-3의 특징들 가운데 BC(Backward Compatibility)는 11172-3 오디오 복호화기가 13818-3 다채널 데이터 가운데 스테레오 정보를 정확히 복호 처리 할 수 있어야 함을 말한다. 그림 6에 그 비트스트림의 예를 보인다. MPEG-2 다채널 오디오 부호화 시스템은 비트율이 384 kbps일 때, 16 비트 PCM신호의 음질과 유사한 것을 제공하고자 했

으나, MPEG-2 BC들의 청취실험 결과, ITU/R이 요구하는 방송의 질에 적합한 점수를 받지 못했기 때문에 그림 7에서와 같이 확장 부분을 사용해 사용 가능한 비트 수를 늘려 음질을 보완시켜 주기로 하였고, 한편으로 11172-3과 양립성을 제공하지 않는 MPEG-2 확장에 대한 표준을 만들기로 하였다.

MPEG-2 확장 기술은 달리 MPEG-2 오디오 NBC(Non- Backward Compatibility), 또는 MPEG-2 AAC(Advanced Audio coding) 이라고도 불리우며 1995년 3월 기술 제안을 받기 시작해 1997년 4월에 국제표준으로 제정되었다(ISO/IEC 13818-7). ISO/IEC 13818-7⁶⁹⁾의 프로파일 가운데 LC (Low Complexity) 프로파일은 현재 A2b 라는 이름으로 인터넷에서 쉽게 찾아 볼 수가 있다. A2b는 AT&T사에서 ISO/IEC 13818-7의 LC 프로파일 에 암호화 기술을 사용하여 저작권에 대한 대책을 세운 것으로 음질의 저하가 없는 상태에서 최고 20배 까지 압축이 가능하게 한다.

그림 8에 13818-7 부호화 복호화 처리시의 블록도를 보인다.

그림 6. ISO/IEC 13818-3 계층의 다채널 확장된 구조가 ISO/IEC 11172-3 계층 II에 후방 양립성을 갖는 것을 보이는 예.

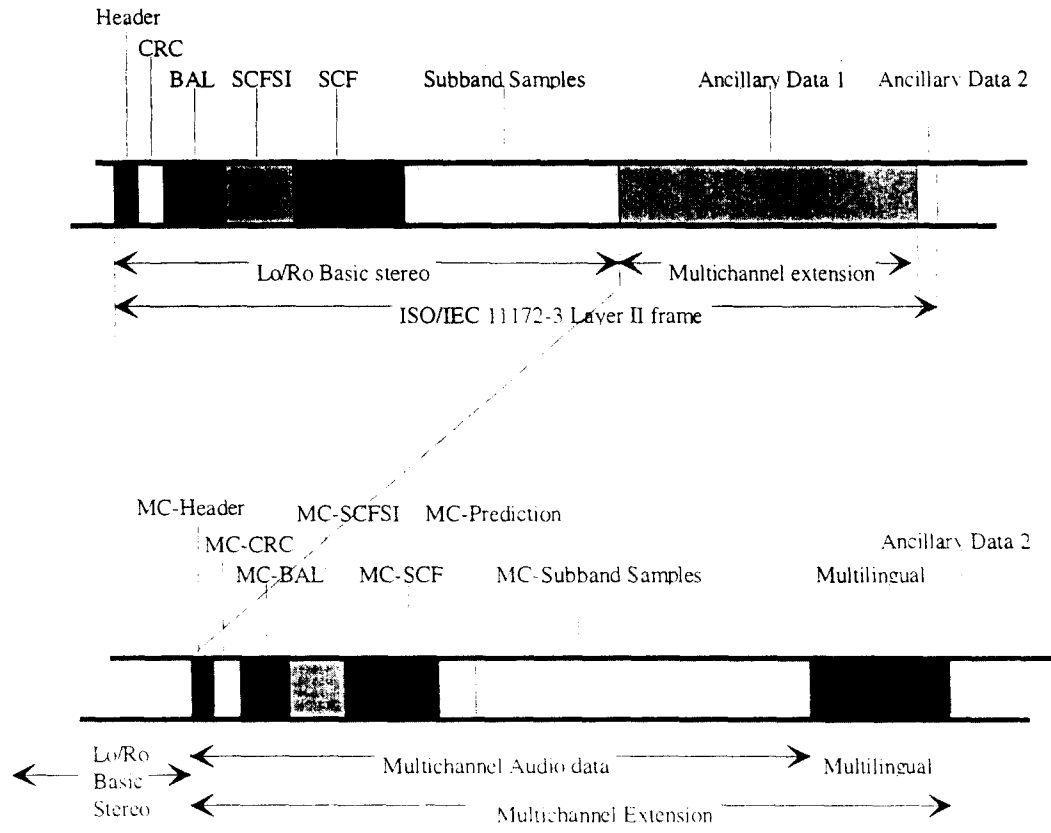


그림 7. ISO/IEC 13818-3 계층 II가 ISO/IEC 11172-3과 호환인 부분과 확장된 비트스트림을 사용하는 경우에 대한 예.

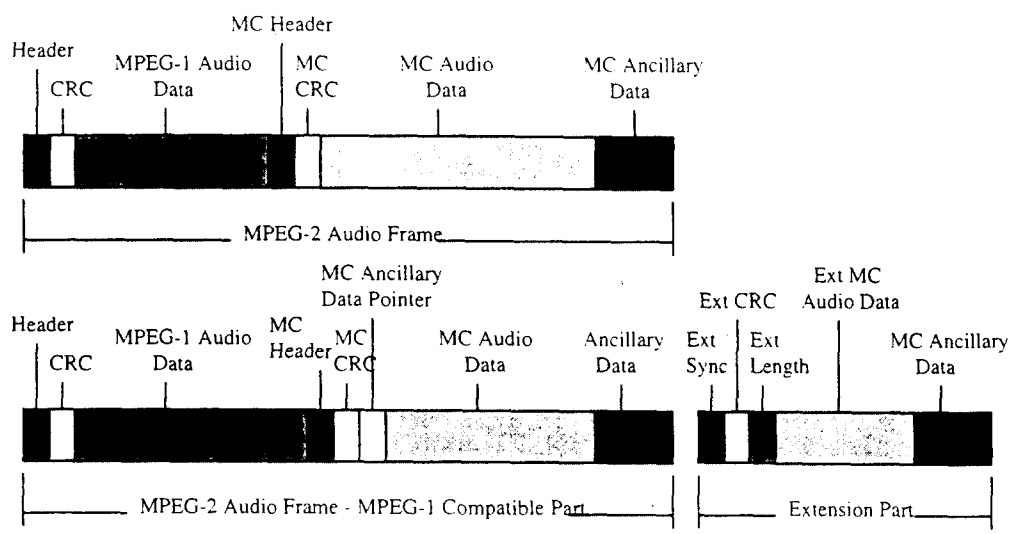
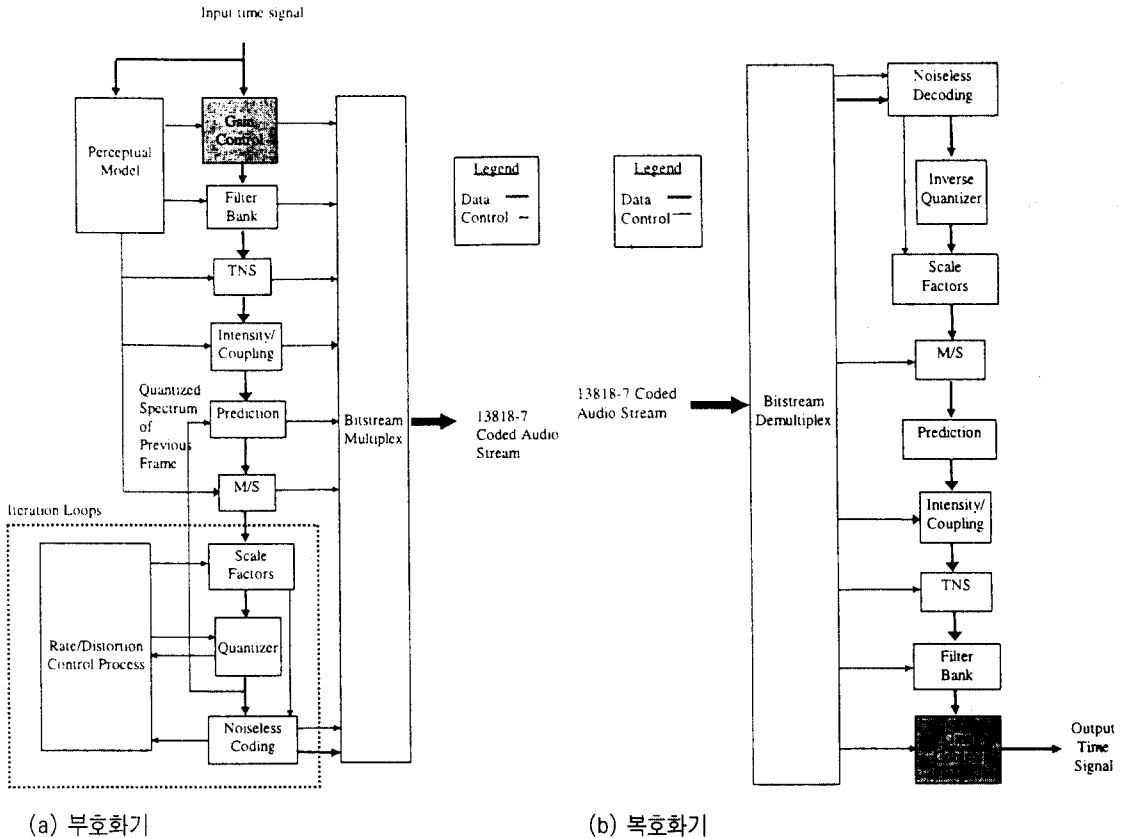


그림 8. ISO/IEC 13818-7의 블록도



3. MPEG-4 오디오 (ISO/IEC 14496-3)

MPEG-4 오디오에서는 개선된 코딩 능력, 내용(object 또는 content)에 근거한 해상도 조절 능력, 내용에 근거한 연산 기능 및 비트스트림 편집 기능, 오류 신호에 대한 대책, 내용에 근거한 멀티미디어 데이터 입수 장치, 동시 다중 데이터 스트림의 부호화 기능, 하이브리드 데이터 처리 방식(자연 데이터와 합성 데이터), 초저비트율에서의 개선된 랜덤 액세스 기능들을 효과적으로 제공하는 방식을 표준으로 만들기 위한 일이 진행되고 있다. 위의 기능들을 만족하는 기술들로 1998년 12월에 MPEG-4 IS(International Standard) (ISO/IEC 14496-3)의 1차 버전이 완성될 예정이고, 그 추가 기술에 대

한 CD를 1998년 12월에 만든 뒤, 2000년 2월에는 추가 기술에 대한 IS가 제정될 예정이다. MPEG-4 오디오 표준 제정에 있어서는 국내에서는 ETRI와 삼성종합기술원이 기술 제안 등을 통한 활발한 참여를 하고 있다.

MPEG-4 오디오⁽¹¹⁾에서 대상으로 하는 비트율과 그 대상 기술들은 그림 9에서 보인다. MPEG-4는 MIDI 데이터, CELP 데이터, 여러 상태의 음질을 제공하는 음성 신호, 여러 상태의 음질을 제공하는 오디오 신호들을 다룰 수 있는 방식들을 표준의 일부로 채택하였다. MPEG-4 오디오의 대표적인 경우들 가운데 하나인 음성신호와 오디오 신호를 처리하기 위해 결합되어 있는 경우에 대한 구조의 일 예는 그림 10과 같다.

그림 9. MPEG-4 오디오에서 대상으로 하는 비트율과 대상 기술

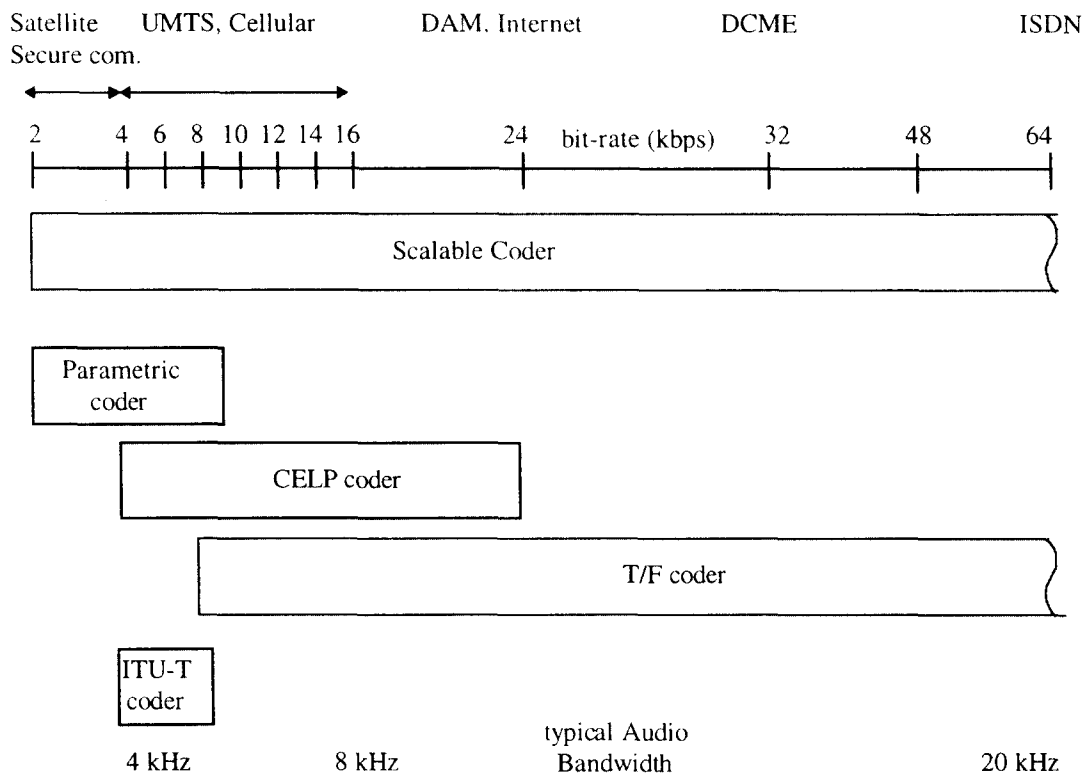
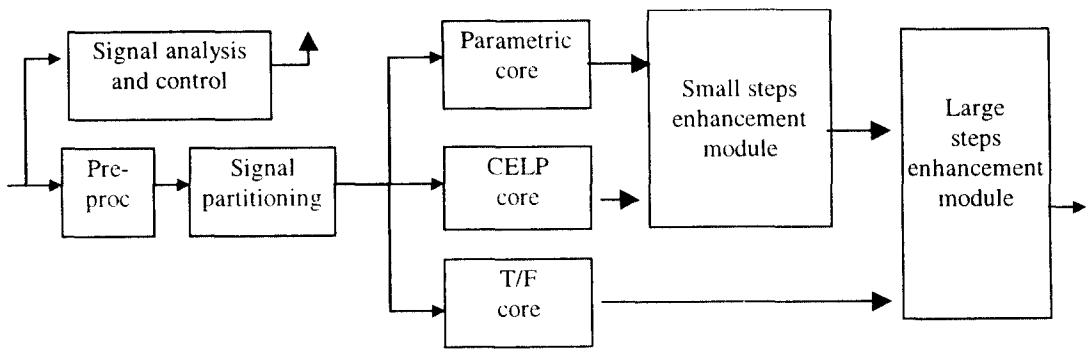


그림 10. 오디오 신호와 음성 신호의 처리를 위한 구조의 일 예



인터넷에서 접할 수 있는 기술들 가운데, 아마하사의 사운드VQ는 ISO/IEC 14496-3에 있는 Twin-VQ 기술의 변형이다. MPEG-4 오디오가 MPEG-1, MPEG-2와 다른 점들 가운데 하나로는 해상도의

조절이 가능한 비트스트림을 제공한다는 것이다. 해상도의 조절이 가능한 점은 사용되는 채널의 비트 전송률이 바뀌는 경우에도 바뀐 전송률에 맞추어서 사용이 가능한 특징을 제공한다.

V. 결 론

컴퓨터 기술 및 디지털 통신 기술, 디지털 신호 처리기들의 발달에 힘입은 디지털 오디오 데이터 압축/복원 방식들은 같은 용량의 기록 저장 매체에 보다 많은 정보들을 저장할 수 있고, 데이터의 전송에 드는 시간을 줄여주는 기능을 갖는다. 본 논문에서는 이러한 디지털 오디오 데이터 압축/복원 방식의 일반적인 구조와 사용 기술들을 다루어 주었고, 인간의 음향 심리를 이용한 대표적인 오디오 데이터 압축 기술인 ISO/IEC JTC1/SC29/WG11(MPEG)에서 국제 표준 기술로 제정되어 온 ISO/IEC 11172-3, 13818-3, 13818-7, 14496-3에 대해서 간단히 소개를 하였다.

※ 참 고 자 료

- (1) A. Gersho, "Advances in Speech and Audio Compression," Proc. of IEEE, vol. 82, no. 6, pp. 900-918, June 1994.
- (2) P. Noll, "Wideband Speech and Audio Coding," IEEE Commun. Mag., vol. 31, no. 11, pp. 34-44, Nov. 1993.
- (3) N. S. Jayant, J. Johnston, and R. Sofranek, "Signal Compression based on models of human perception," Proc. IEEE, vol. 81, no. 10, pp. 1385-1422, Oct. 1993.
- (4) K. Brandenburg, and G. Stoll, "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio," Presented at 92nd Audio Eng. Soc. Conv., Preprint 3336, Mar. 1992.
- (5) T. D. Rossing, The Science of Sound (2nd ed.), Addison-Wesley, Reading: Massachusetts, 1990.
- (6) J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, The MIT Press: Cambridge, Massachusetts, 1983.
- (7) J. Tobias (ed.), Foundations of Modern

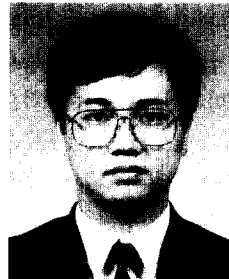
Auditory Theory, New York: Academic, 1970.

(8) ISO/IEC 13818-7 IS, "MPEG-2 AAC: Advanced Audio Coding" April, 1997.

(9) ISO/IEC 13818-3 IS, "Information Technology- Generic Coding of Moving Pictures and Associated Audio: Audio," November, 1994.

(10) ISO/IEC 11172-3 IS, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBit/s: Audio," 1992.

(11) ISO/IEC 14496-3 DIS, "Information Technology- Generic Coding of Audio-Visual Objects: Audio," October, 1998.



김 상 욱

1967년 출생.

1989년 연세대 전기공학과 졸업 (공학사)

1991년 한국과학기술원 전기및전자공학과 졸업 (공학석사)

1991년 3월 ~ 현재 삼성종합기술원 전문연구원

주 관심 분야 : 디지털 오디오 데이터 압축,
삼차원 오디오 처리