

# 디지털 방송을 위한 오디오 부호화 기술 동향

## Audio Coding Technologies for Digital Broadcasting

□ 최승호\*, 김흥국\*\*, 석용호\*\*\* / \*서울산업대학교, \*\*광주과학기술원, \*\*\*(주)엠큐브웍스

### I. 서론

최근 들어 유무선 디지털 송수신 기술의 발전과 초고속 인터넷 보급의 일반화 등에 힘입어 우리 사회는 디지털 멀티미디어 정보화 사회로의 급격한 변화를 경험하고 있다. 이러한 환경에서 멀티미디어의 큰 축을 담당하고 있는 오디오의 품질에 대한 사용자의 기대는 날로 높아지고 있다. 그래서 각종 오디오 시스템은 고품질이며 실감형의 음질을 보장하여야 하며, 이를 위해 넓은 주파수 대역이면서 다채널의 신호를 처리할 수 있어야 한다. 하지만 실시간으로 고품질의 디지털 멀티미디어/오디오 방송을 서비스하기 위해서는 전송 채널의 제한을 극복하여야 하며, 이를 위해 고품질 오디오 부호화 기술에 대한 관심이 그 어느 때보다 고조되고 있다.

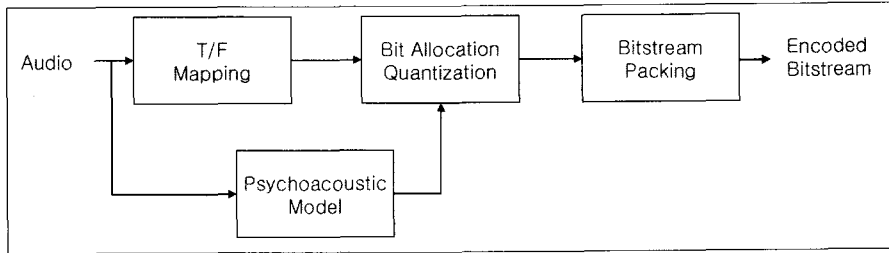
본 논문은 디지털 방송을 위한 오디오 부호화 기술에 대한 이해를 돕기 위한 것으로서 우선 제2장에

서 오디오 부호화 기술 즉, 심리음향 모델, 비트 할당 및 양자화 등의 원리를 소개한다. 제3장에서는 MPEG-2/4 AAC, AC-3 등의 표준 오디오 부호화기들을 소개하고 제4장에서는 오디오 품질 평가 방법에 대해서 기술한다. 그리고 제 5장에서는 국내외 각종 디지털 방송 시스템에서의 오디오 부호화기들의 적용 예와 기술 개발 동향 및 전망 등을 기술한다. 마지막으로 제6장에서 결론을 맺는다.

### II. 오디오 신호 부호화의 원리

#### 1. 오디오 부호화기 알고리즘의 일반적인 구조

오늘날 광범위하게 사용되는 오디오 부호화기는 대체적으로 <그림 1>과 같은 구조를 가지고 있다 [1]. 오디오 부호화기 알고리즘은 일반적으로 시간-



〈그림 1〉 일반적인 오디오 부호화기 기본 구조

주파수 매핑 (Time/Frequency mapping, T/F mapping), 심리음향 (psychoacoustic) 모델, 비트 할당 및 양자화 (bit allocation and quantization), 비트스트림 패킹 (bitstream packing) 등의 4가지 블록으로 구성된다. 주요 블록에 사용되는 기술에 대한 세부 내용은 다음과 같다.

### 2. 시간-주파수 분석

오디오 신호는 높은 압축률을 얻기 위해 시간 영역의 신호를 주파수 영역의 신호로 변환하게 된다. 이때 입력된 오디오 신호는 보통 2~50 ms 단위로 나누어 분석된다. 인간의 청각적 특성을 고려하여 분석된 결과는 심리 음향 모델의 적용을 받아 양자화되어 압축된다. 그리고 시간-주파수 분석에서 대표적으로 고려하여야 하는 것에는 해상도 트레이드 오프 (resolution tradeoff)와 프리에코 왜곡 (pre-echo distortion)이 있다.

첫째, 해상도 트레이드오프는 적당한 주파수 영역 분석 기법을 요구한다. 시간 영역의 신호를 주파수 영역으로 변환할 경우, 시간 영역의 해상도를 높이면 주파수 영역의 해상도가 감소하게 되며, 그 반대의 경우도 성립한다. 짧은 시간에 넓은 주파수 대역 분포를 갖고 있는 신호는 시간 영역의 해상도를 높이고 비교적 긴 시간에 좁은 주파수 대역 분포를

갖고 있는 신호는 주파수 영역의 해상도를 높여서 분석하는 것이 효율적이다. 하지만 실제 우리가 인지하는 오디오 신호는 위의 두 가지 경우가 혼합된 형태가 대부분이며 이에 대한 분석 방법으로 DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform), MDCT (Modified DCT) 등의 변환 코딩 (transform coding)과 다상 불균일 필터뱅크 (polyphase non-uniform filterbank), 웨이블릿 (wavelet) 등의 대역분할 코딩 (subband coding)으로 나뉜다 [2].

둘째, 프리에코 왜곡은 낮은 에너지의 영역 바로 뒤에서 갑작스럽게 발생하는 소리를 주파수 영역으로 변환할 경우 발생하게 된다. 프리에코가 발생하는 부분에 대해서 시간-주파수 변환의 범위를 작게 하여 프리에코 현상을 제어할 수 있다.

### 3. 심리음향 모델

심리음향이란 오디오 코딩에서 이용되는 청각적 특성을 말한다. 즉, 인간은 외부의 소리를 인지할 때 상황에 따라 특정 소리를 듣지 못하는 특성을 가지고 있다. 이를 이용하여, 오디오 압축 시에 이러한 소리들을 제거한 손실압축 (lossy compression)을 함으로써 높은 압축률을 얻는다. 이러한 방법은 높은 압축률에도 불구하고 복원된 오디오 신호는 원래 오디오

신호와 차이를 느끼지 못하는 장점을 가진다. 오디오 코딩 알고리즘에서 이용하는 중요한 두 가지 심리음향 특성은 최소 가청 한계 (threshold of hearing)와 마스킹 효과 (masking effect)가 있다 [3].

### 1) 최소 가청 한계

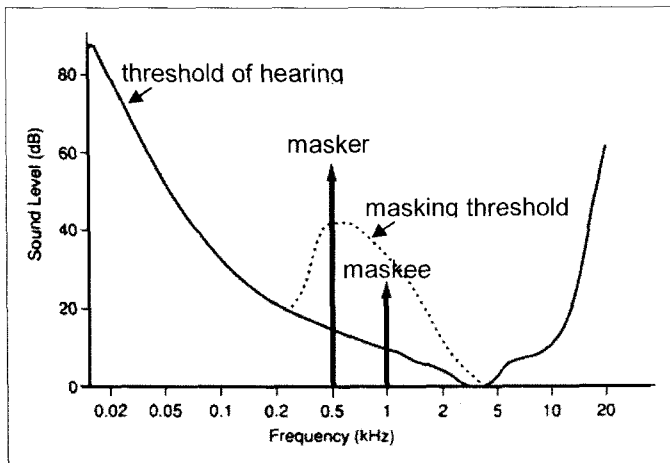
최소 가청 한계란 인간이 인지할 수 있는 소리의 최소 레벨을 뜻한다. 최소 가청한계보다 소리의 레벨이 작으면 인간은 그 소리를 들을 수 없게 되고, 통증한계 (threshold of pain)보다 높은 레벨의 소리는 듣기에 어려움이 생긴다. 따라서 오디오 코딩 시에 입력되는 신호의 레벨을 측정하여 가청범위에 있는 신호만을 선택하여 효율적이고 높은 압축을 실현할 수 있게 된다.

### 2) 마스킹 효과

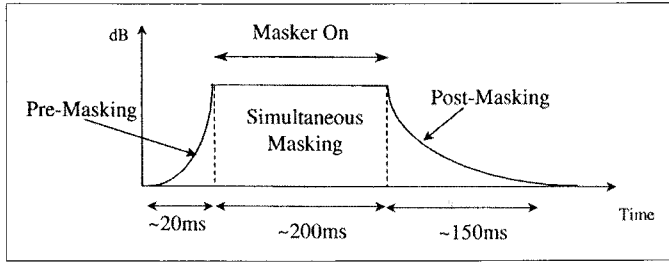
마스킹 (mask)란 덮어 가리는 것을 뜻하는 말로, 심리음향에서 마스킹 효과는 어떤 소리가 특정한 소리에 의해 들리지 않게 되는 현상을 일컫는다. 오디오 신호는 서로 다른 크기를 갖는 수많은 주파수 신

호들로 이루어져 있다. 어떤 주파수의 소리가 들리게 되면 그 주파수와 인접한 다른 주파수는 이미 들려진 주파수의 소리가 상대적으로 클 경우에 인지를 못하게 되는 경우가 발생한다. 이를 주파수 마스킹 (frequency masking)이라고 하며 마스킹되는 주파수 성분을 제거함으로써 오디오 데이터의 양을 줄일 수 있다. 인접 주파수를 제거할 것인지 제거하지 않을 것인지는 마스킹 임계값 (masking threshold)에 의해 결정된다. 이때, 레벨이 큰 신호를 마스커 (masker)라 하고, 마스커에 의해 마스킹되는 소리를 마스크 (maskee)라 한다. <그림 2>는 주파수 마스킹의 예로서 1 kHz의 톤 (tone) 신호 (makee)는 0.5 kHz의 톤 신호 (masker)에 의해 마스킹됨을 보여준다.

시간 마스킹 (temporal masking)은 어떤 소리가 사람 귀에 들려질 때 그 소리 전후로 짧은 시간 동안은 다른 소리에 대해서 사람이 인지하지 못하는 특성을 나타낸다. 예를 들어, 여러 악기가 동시에 연주되는 음악의 경우, 소리 레벨이 큰 한 악기에 의해 레벨이 작은 악기들의 소리가 들리지 않는



<그림 2> 주파수 마스킹 효과 예시



〈그림 3〉 시간 마스킹 효과 예시

현상이 발생한다. 〈그림 3〉은 시간 마스킹 효과가 미치는 시간간격의 예를 보여준다. 즉, 어떤 큰 소리는 이전의 20 ms 전에 들리진 소리와 이후의 최대 150 ms 이내에 들려지는 소리를 마스킹할 수 있다.

### 3. 비트 할당 및 양자화

비트 할당 및 양자화 과정은 시간-주파수 매핑을 통해 들어온 신호를 심리음향 모델에서 얻은 정보를 이용하여 비트할당 (bit allocation)을 하는 과정과 할당된 비트에 맞게 주어진 주파수 영역의 신호를 양자화하는 과정으로 나뉜다.

각 대역에서 마스킹 레벨보다 낮은 소리는 인간이 들을 수 없기 때문에 마스킹 레벨보다 낮은 신호에 대해서는 비트를 할당하지 않는다. 따라서 양자화 과정을 통해 양자화 잡음 (quantization noise)이 필연적으로 삽입되지만, 마스킹 효과를 이용하여 양자화 잡음의 레벨을 마스킹 레벨보다 낮게 조절한다. 이렇게 마스킹 레벨보다 낮게 잡음 레벨이 존재하게 되면 잡음을 느끼지 못하면서 비트를 줄일 수 있게 된다.

기본적인 양자화 기법은 스칼라 양자화 (scalar quantization), 벡터 양자화 (vector quantization),

파라메트릭 양자화 (parametric quantization)가 있다. 스칼라 양자화는 균일 (uniform), 차분 (differential), 적응 (adaptive) 양자화 등으로 나눌 수 있다. 스칼라 양자화의 경우 하나의 샘플을 하나의 심볼 (index)로 보내는 반면에 벡터 양자화는 여러 샘플을 하나의 심볼로 보낸다. 이 심볼은 코드북 (codebook)의 하나의 벡터 (codeword)를 가리키게 된다. 벡터 양자화는 압축효율은 뛰어나지만, 원 신호가 아닌 코드북에 있는 신호로 복호화되므로 음질 열화가 생길 수 있다. 대표적인 벡터 양자화로는 MPEG-4 TwinVQ가 있다[4]. 파라메트릭 양자화 기법은 샘플 정보를 보내는 것이 아니라 합성에 필요한 샘플의 특징을 추출하여 양자화하는 방법이다. 복호화기는 이 샘플의 특징정보를 이용하여 다시 음을 만들어서 복원한다. MPEG-4 AAC에서 사용되는 PNS (Perceptual Noise Substitution)가 그 예이다 [5].

## III. 표준 오디오 부호화기 소개

### 1. 개요

방송과 통신용 오디오 부호화기의 표준화는 주로

ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) 산하 MPEG (Moving Picture Experts Group) Audio 분과에서 진행되어 왔으며, 1992년 MPEG-1 Layer 1/2/3에서부터 최근 2006년 MPEG-4 ALS (Audio Lossless Coding)에 이르기까지 채널당 24~96 kbit/s의 비트율을 가지는 오디오 부호화기들이 표준화되어 왔다. 본 절에서는 현재까지 표준화된 대표적인 오디오 부호화기들의 기술적인 특징을 소개하기로 한다.

## 2. MPEG-1

MPEG-1은 VHS (Video Home System) 테이프 수준의 동영상과 오디오를 디지털 저장매체에 최대 1.5 Mbit/s로 압축하여 저장할 수 있도록 설계되었다 [6]. 이중 오디오 압축 알고리즘은 스테레오 CD 음질의 오디오 데이터를 압축했을 때 최대 비트레이트가 128 kbit/s (채널당 64 kbit/s)가 되도록 정해졌다. MPEG-1의 계층은 Layer 1, 2, 3으로 나뉘게 된다. Layer 1, 2는 비슷한 알고리즘으로 되어 있고, Layer 3의 경우 MDCT와 허프만 부호화 (Huffman coding) 등의 알고리즘이 추가되었다. 이를 통해 12:1의 압축률을 얻고 있다. 이중에서도

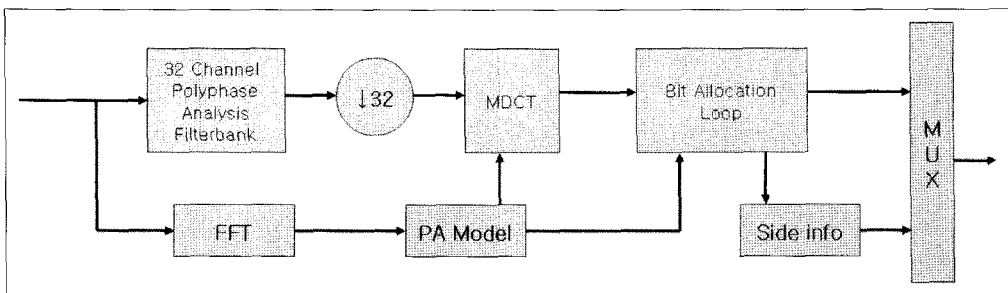
MP3는 1991년 MPEG에서 표준화된 고압축 디지털 오디오 포맷으로서 오디오 품질이 192 kbit/s의 압축률에서 원음과 투명성 (transparency)을 가지고 128 kbit/s에서는 어느 정도 만족할 만한 오디오 품질을 가진다. MP3 부호화기의 구조를 살펴보면 <그림 4>와 같다.

## 3. MPEG-2

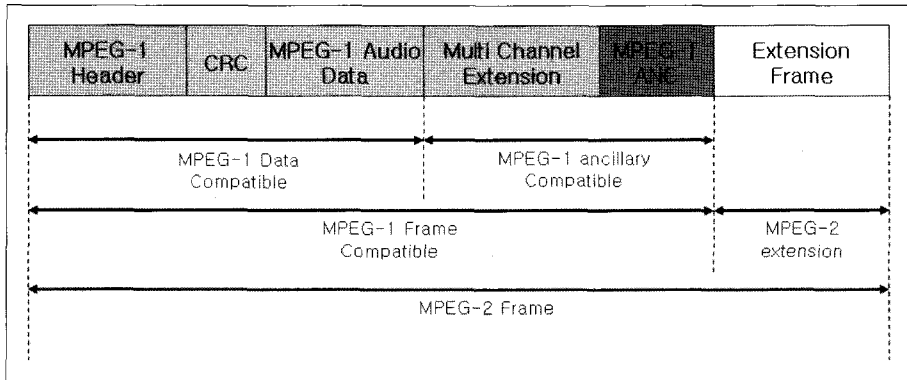
MPEG-2는 크게 MPEG-1과 호환성 (backward compatibility, BC)을 가지는 부호화기 (MPEG-2 BC)와 MPEG-1과 호환이 되지 않는 즉, 비호환성 (non-backward compatibility)를 가지는 부호화기 (MPEG-2 NBC)로 나뉜다.

### 1) MPEG-2 BC

호환성을 가지는 부호화기는 또 채널 수를 스테레오에서 다채널로 확장한 MC (Multi Channel) 부호화기와 MPEG-1에서의 샘플링 주파수의 반에서 동작하도록 확장한 LSF (Low Sampling Frequency) 부호화기로 나뉜다. MC와 LSF는 MPEG-1과 호환성을 가진다. MC는 부가적으로 다국어 (multilingual) 서비스가 가능하며 최대 7개 국의 부가 음성정보를 전송할 수 있다. MPEG-2



<그림 4> MP3 부호화기의 구조



〈그림 5〉 MPEG-2 MC 비트스트림 구조

MC의 비트스트림 구조를 살펴보면 〈그림 5〉와 같다[7].

MC의 경우 MPEG-1과의 호환을 위해 스테레오에 관한 부호화는 MPEG-1과 동일하게 하고 다른 채널에 대해 부가적으로 부호화한다. LSF의 경우 샘플링 주파수를 제외하고는 Layer 1, 2, 3 모두 MPEG-1과 동일한 알고리즘을 사용한다.

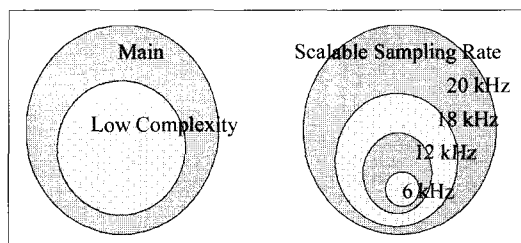
## 2) MPEG-2 NBC

앞절에서 설명한 MPEG-2 BC 오디오 부호화기는 MPEG-1과의 호환을 고려하며 개발되었기 때문에 새로운 코딩 기법 사용에 많은 제약을 받게 되었고, 압축률의 큰 향상을 가지고 오지 못했다. 이러한 문제를 해결하기 위해서 MPEG은 MPEG-1과 호환성을 고려하지 않은 MPEG-2 NBC인 AAC (Advanced Audio Coding)를 개발하게 되었다. Dolby, Fraunhofer, AT&T, Sony, Nokia 등 기존 MP3, AC-3, ATRAC (Adaptive Transform Acoustic Coding) 등 부호화기 개발에 많은 기여를 한 회사들이 개발에 참여하였다. AAC의 주요한 특징으로는 다음과 같다 [8].

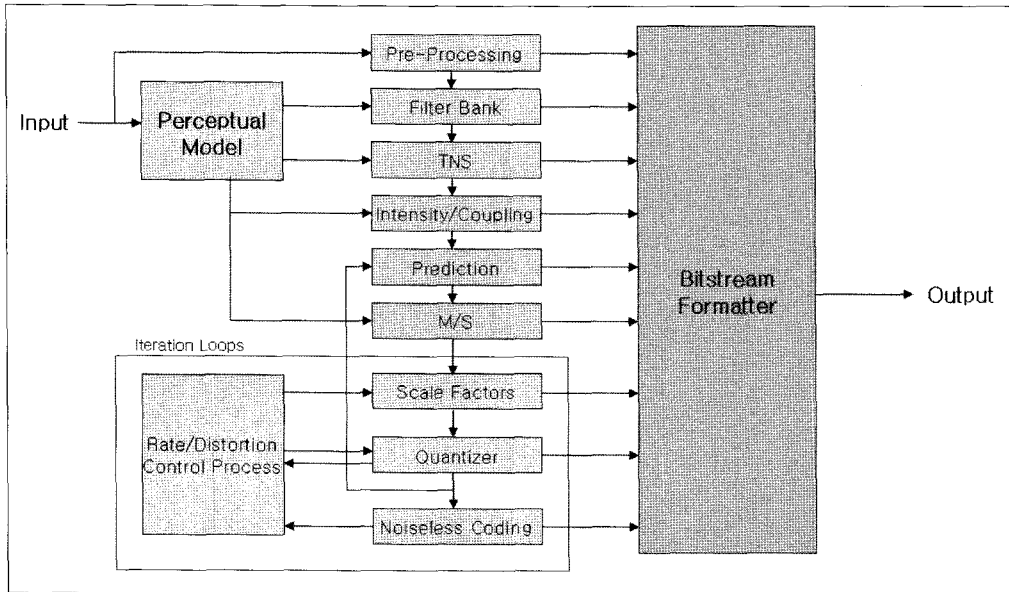


- 8~96 kHz 샘플링 주파수 지원.
- 멀티채널/다국어: 최대 8.1 채널, 7개국 부가 음성정보 전송.
- 320 kbit/s (5.1 채널)에서 원음과 구별할 수 없는 고품질 제공.
- 256/2048 MDCT, 심리음향 모델, 허프만 코딩 등 뿐만 아니라 TNS (Temporal Noise Shaping), prediction 등 MPEG-1과 다른 압축 도구 (tool)를 사용.

MPEG-1 이나 MPEG-2에서 Layer라는 개념을 쓴 것과 유사하게 AAC는 프로파일 (profile)이라는 계층이 있다. AAC는 〈그림 6〉과 같이 AAC-Main, AAC-LC (Low Complexity), AAC-SSR (Scalable Sampling Rate)의 3개의 프로파일로 되어 있으며,



〈그림 6〉 AAC 프로파일 구조



〈그림 7〉 AAC부호화기의 구조

활용분야에 따라 프로파일을 선택하여 사용할 수 있다.

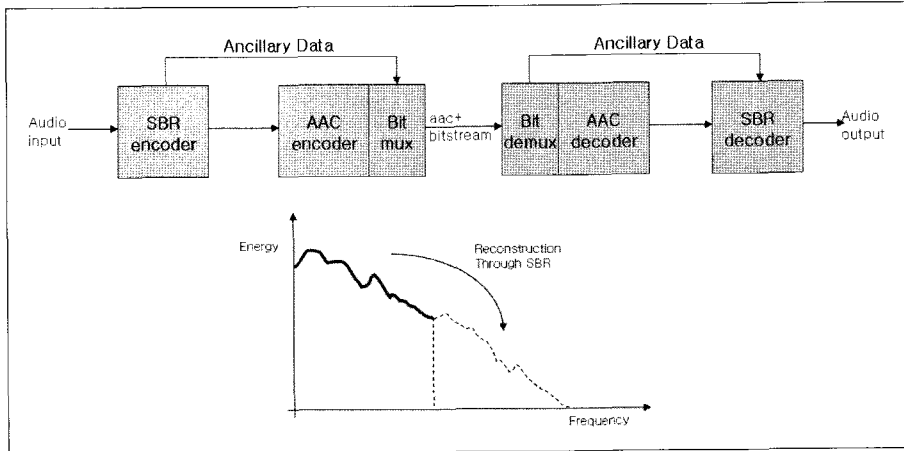
AAC-Main 프로파일은 AAC 표준안에서 제안하는 압축 도구 중에 AAC SSR 프로파일에서 사용되는 이득 조절 (gain control)을 제외한 모든 도구를 사용한다. AAC-LC 프로파일의 경우, AAC-Main 프로파일에서 prediction 부분이 제외되고, TNS의 계수를 12개로 제한하는 등, 복잡도가 높은 도구를 사용하지 않는다. AAC-LC 프로파일은 AAC-Main 프로파일에 비해서 음질의 차이는 많지 않지만 복잡도에서 많은 이득을 얻을 수 있기 때문에 가장 널리 사용되고, MPEG-4 AAC, HE-AAC (AAC+), HE-AAC v2 (AAC+ v2) 등 상위 버전의 기초가 된다. AAC-SSR 프로파일은 AAC-LC 프로파일에 하이브리드 필터뱅크 (IPQF + divided IMDCT)를 사용한 것이 특징이고 다른 프로파일에서는 사용되지 않는 이득 조절이 사용된다.

AAC의 오디오 품질은, MP3가 192 kbit/s에서 원음과 투명성을 가지고, MPEG-1 Layer 2가 256 kbit/s에서 투명성을 가지는 것에 비해, 128 kbit/s에서 투명성을 가지게 된다. 따라서 AAC는 MP3에 비해 33%, MPEG-1 Layer 2에 비해 50%의 부호화 효율을 갖는다. AAC 부호화기의 구조는 〈그림 7〉과 같다.

#### 4. HE-AAC와 HE-AAC v2

##### 1) HE-AAC

HE-AAC (High Efficiency AAC)는 MPEG-4 AAC (MPEG-2 AAC-LC 프로파일 + LTP (Long Term Prediction) + PNS)에 SBR (Spectral Band Replication)을 추가한 부호화기로서 AAC+라고도 불린다. PNS는 특정 밴드가 잡음에 가까울 경우, 노이즈 검출 (noise detection) 후 이득



〈 그림 8〉 SBR의 동작 구조 및 동작원리

(gain) 정보만 전송하고 복호화기는 이를 이용하여 노이즈를 만드는 기법이다. 그리고 노이즈 대역에 대해 비트 할당을 하지 않고 이득정보만 압축하기 때문에 압축률을 향상시킨다. LTP는 AAC의 예측(prediction) 틀이 높은 복잡도를 가지고 주파수 영역에서 예측을 수행하는 문제를 해결하기 위해 상대적으로 복잡도가 낮은 시간영역에서의 예측을 하는 기법이다.

SBR은 고주파 대역의 신호를 압축하여 전송하지 않고, 저주파 대역의 신호와 고주파에서 추출한 정보를 가지고 복호화기에서 고주파 대역을 생성하는 기술이다. SBR을 적용한 구조 및 동작원리는 〈그림 8〉과 같다.

오디오 입력 신호는 SBR 부호화기를 통해 고주파 성분이 사라진 신호와 부가 데이터로 바뀌게 된다. 고주파 성분이 사라진 신호는 AAC 부호화 과정을 거치고, 부가 데이터는 AAC 비트스트림과 함께 저장된다. SBR 복호화기는 부가 데이터와 AAC 복호화기에서 재생된 오디오 출력 신호를 이용해 고주파 성분을 복원하게 된다. SBR은 모든 오디오 부

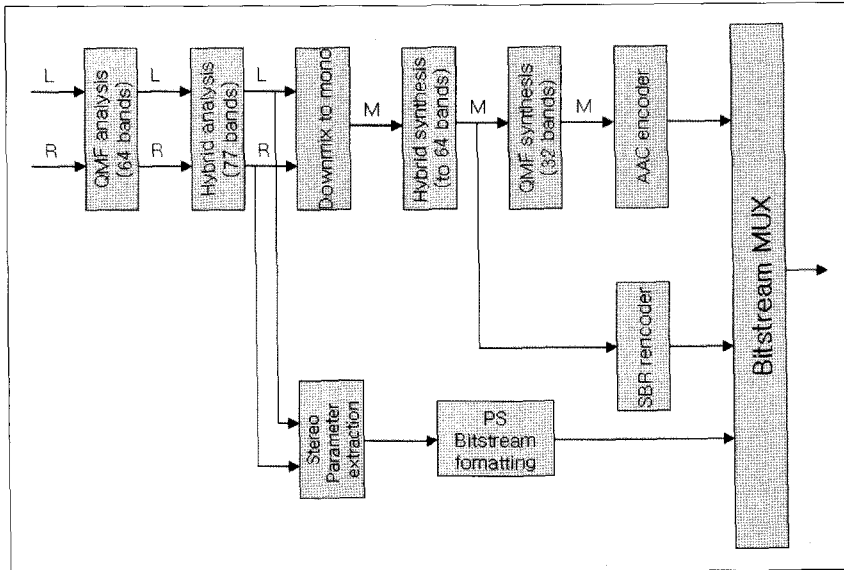
호화기에 적용이 가능하고, MP3PRO (MP3 + SBR)가 그 예이다[9]. HE-AAC는 96 kbit/s에서 원음과의 투명성을 가지며, 48 kbit/s 정도에서도 만족스러운 음질을 가진다. 하지만 여전히 복잡도가 높다는 단점이 있다.

## 2) HE-AAC v2

HE-AAC v2는 HE-AAC에 Coding Technologies사와 Philips사에서 개발한 PS (Parametric Stereo)를 추가한 부호화기다. PS를 적용한 구조는 〈그림 9〉와 같다[10].

PS는 우선 스테레오 채널을 모노 채널로 다운믹싱(downmixing)하고 다운믹싱을 통해 얻은 부가 정보를 비트스트림 멀티플렉서(Bitstream MUX)에 보내어 스테레오 채널의 오디오 입력 신호를 직접 부호화하는 것보다 압축률을 높이는 기술이다. 5.1채널 오디오의 경우 128 kbit/s의 압축률에서도 만족할만한 음질을 제공한다. 스테레오 오디오에 대해서는 32 kbit/s 이상에서 원음과 투명성을 가지게 되고, 24 kbit/s 이상에서 MP3 128 kbit/s와 유





〈그림 9〉 Parametric Stereo의 구조

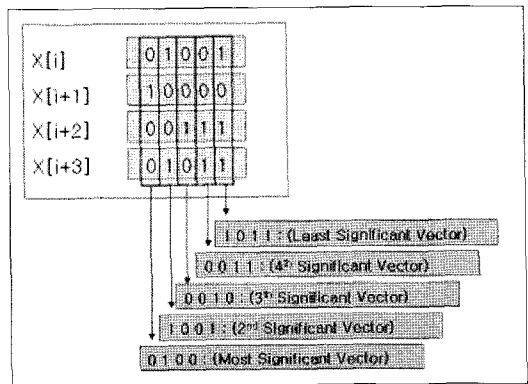
사한 음질을 가지게 된다.

### 5. BSAC

가변비트율을 가지는 오디오 부호화기 기술로는 대표적으로 MPEG에서 표준화한 BSAC (Bit-Sliced Arithmetic Coding)이 있다 [11]. BSAC은 MPEG-4 version 2에서 작은 단위 가변성 (small step scalability)을 가지는 서비스에 사용하기 위한 목적으로 표준화되었다. 큰 단위 가변성 (large step scalability)는 여러 부호화기를 필요로 하고 8 kbit/s 이상의 가변 step size를 가지는데 반해, BSAC은 그 자체만으로 1 kbit/s의 step size를 가지는 장점이 있다. BSAC은 〈그림 7〉에서 보여준 MPEG-4 AAC의 Quantizer와 Coding 부분을 대체하여 사용된다 [12].

또한 큰 단위 가변성을 시간영역에서 가변적인 (scalable) 부호화 방법으로 실현하는 것과 대조적으

로, BSAC은 스펙트럼 계수(spectral coefficients)를 처리하는 주파수 영역의 가변 부호화 방법이다. BSAC의 기본적인 개념은 〈그림 10〉과 같다.



〈그림 10〉 BSAC의 기본적인 개념

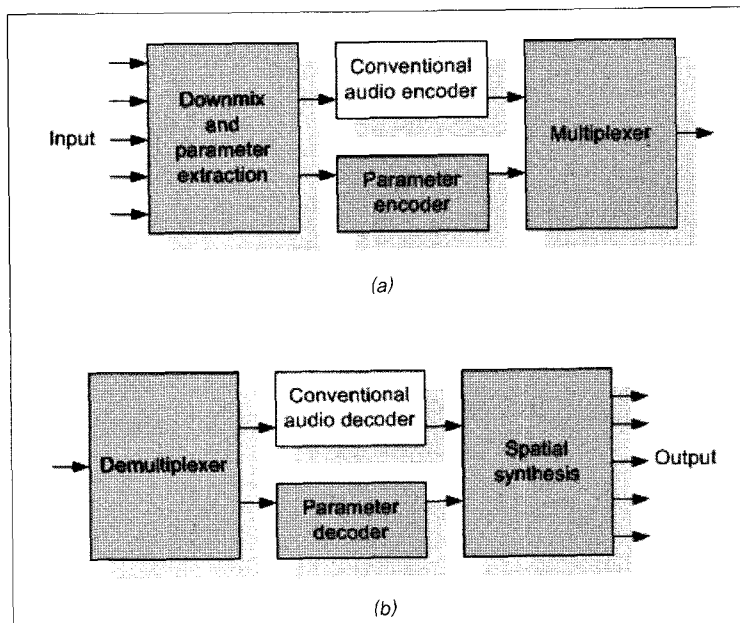
기존 MPEG-4와 다르게 BSAC에서는 양자화된 스펙트럼 데이터 (spectral data)와 크기정보 (scale factor)에 대해 부호화할 때, AAC에서 허프만 부호

화를 사용하는 것과 달리 산술 부호화 (arithmetic coding)을 사용한다. 일반적으로 MSB (Most Significant Bit)의 정보가 LSB (Least Significant Bit)의 정보보다 음질에 많은 영향을 미치기 때문에 환경에 따라 LSB를 전송하지 않더라도 음질은 큰 손상을 받지 않는다. 따라서 BSAC에서는 <그림 10>에서 보는 것과 같이 LSB로 이루어진 최저 중요 벡터 (Least Significant Vector)를 지우거나 전송하지 않음으로써 작은 단위 가변성을 제공한다. 64 kbit/s로 라디오, 96 kbit/s로 CD급의 음질을 제공하는 MPEG-4 BSAC은 채널당 16 kbit/s에서 64 kbit/s까지의 가변 압축률을 가지게 된다

### 6. MPEG Surround

전통적으로 영화 산업에 의해서 발전된 멀티채널 오디오 기술은 최근 들어 DVD (Digital Versatile/

Video Disc)나 SA-CD (Super Audio CD) 등에서의 멀티채널 콘텐츠의 증가, 멀티 채널 오디오 재생 장비의 저변 확대, 사용자의 욕구 증가 등으로 인해 연구개발이 활성화되었다. MPEG Surround는 2002년부터 MPEG에서 표준화가 진행중인 공간 오디오 부호화 (Spatial Audio Coding) 기술이다 [13]. 이 기술은 기존의 스테레오 오디오 부호화기와의 호환성을 유지하면서 채널당 4 kbit/s 이내의 낮은 비트율의 부가 정보를 추가하여 오디오 신호의 음장감을 제공하기 위한 멀티채널 오디오 부호화 기법이다. MPEG Surround의 구조는 <그림 11>과 같이 간략히 표현할 수 있다. 부호화기에서 멀티채널 신호는 다운믹싱되어 기존의 오디오 압축기술로 부호화된다. 또한 멀티채널 신호로부터 채널간 크기 차이, 채널간 유사도, 채널 예측 계수 및 예측 오차 등의 공간 파라미터를 추출 (parameter extraction)하고 부호화한다. 부호화된 각 비트스트



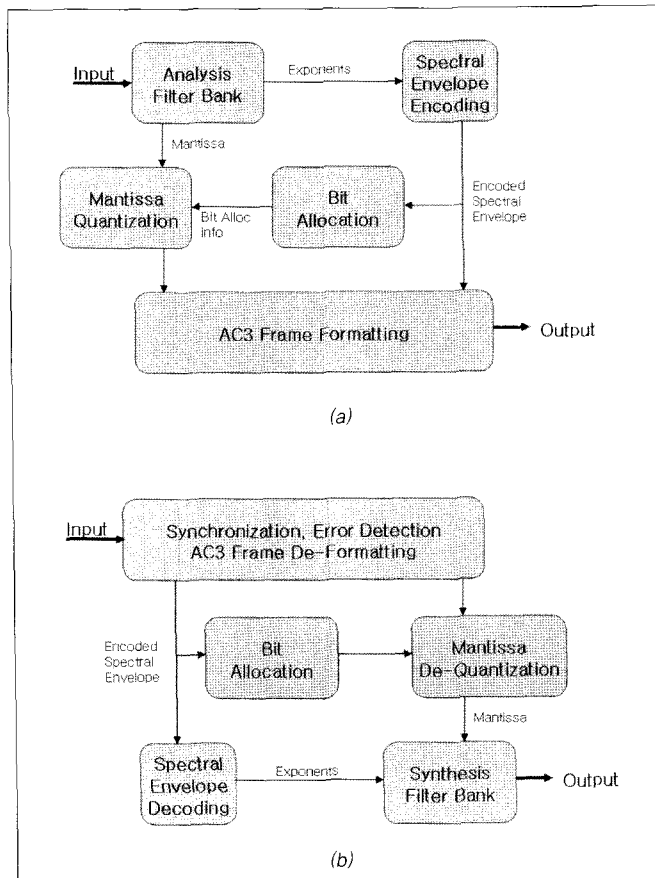
<그림 11> MPEG Surround의 기본적인 개념 (a) 부호화기 (b) 복호화기

림은 멀티플렉서를 통해 합쳐진다. 복호화기에서는 부호화기의 역 과정을 거쳐 멀티채널 신호를 합성한다.

### 7. 기타 오디오 부호화기

대표적인 비 MPEG 계열의 오디오 부호화기로 Dolby Digital (AC-3)이 있다. AC-3는 채널당 32~128 kbit/s의 비트율을 지원하고, 복호화기에서 1~5.1채널의 다운믹싱을 지원한다. 256/512 MDCT의 변환 필터뱅크를 사용하고 지수/가수 부호화

(exponent/mantissa encoding)를 한다. 비트할당을 위해 심리음향 모델을 사용하지만, hybrid backward/forward adaptive bit allocation 방식을 사용한다. AC-3의 구조는 <그림 12>와 같다[14]. 그림과 같이 필터뱅크를 통해 얻은 지수는 비트할당을 위한 정보를 생성하고 가수는 비트할당 정보에 의해 양자화 되는 구조를 가지고 있다. 다른 부호화기의 경우 필터뱅크로 얻은 값을 양자화하는 데 반해 AC-3는 가수와 지수로 나누어 각각 양자화한다. 또한 MP3가 lookahead 방식으로 블록 스위칭(block switching)을 하는 반면, AC-3는 시간영역에서 고



<그림 12> AC-3의 구조 (a) 부호화기 (b) 복호화기

역통과필터를 통해서 블록 스위칭을 한다.

## IV. 오디오 품질 평가

네트워크 환경의 발전에 따라 오디오 스트림을 무선 혹은 IP 네트워크를 통해 전송하는 일이 많아지면서 오디오 품질의 측정에 많은 관심이 증대되고 있다. ITU-T에서는 음성 압축 방법에 대한 평가의 방식으로 대표적인 객관적 음질 평가 방법인 PSQM (Perceptual Speech Quality Measure)를 표준화하였다. 그러나, PSQM은 실시간에서 발생하는 패킷 손실 (packet loss) 문제를 처리하지 못하였다. 따라서 PSQM의 개선된 알고리즘으로 PESQ (Perceptual Evaluation of Speech Quality) [15]가 표준화되었으며 현재도 널리 이용되고 있다. PESQ를 비롯한 알고리즘의 단점은 수신측 종단에서의 결과 호출을 위해 송신단에서 압축 전송에 사용된 원음성 혹은 오디오 신호가 필요하다는 데 있다. 최근 들어 ITU-T에서는 이러한 문제를 극복하기 위하여 PESQ에 비해 정확도는 떨어지지만 원음성이나 오디오가 없는 경우에 유용하게 사용할 수 있는 P.563 [16]이라는 방식을 표준화하였다.

오디오 품질 평가와 관련하여 ITU-R에서는 1994년에 이르러 객관적 지각 오디오 음질 평가의 표준화에 대한 요구와 필요성을 느끼고 표준화 작업에 들어갔다. OPTICOM, Fraunhofer Institute, German Telecom, University of Berlin, IRT Munich, KPN, CCETT 등 많은 기업 및 연구소에서 표준화 작업에 참여하였고, 기존에 제안되었던 DIX (Disturbance Index), NMR (Noise-Masking Ratio), PAQM (Perceptual Audio Quality Measure), PERCEVAL (Perceptual Evaluation),

POM (Perceptual Objective Measure) 등을 토대로 1998년에 PEAQ라 부르는 ITU-R BS.1387 [17]을 표준화하였다. 하지만 PEAQ는 64 kbit/s 이하로 압축된 오디오 음질평가 및 음성 대역만이 존재하는 오디오 부호화기에서 적절치 못한 성능을 보이고 있다. 또한 기존 PEAQ는 방송에서 발생하는 패킷 손실 환경이 고려되지 않은 단점이 있다. 즉, 향후 패킷 손실에 대해 기존의 PEAQ가 처리할 수 있도록 수정 및 개선된 알고리즘이 필요하다.

## V. 디지털 방송을 위한 오디오 부호화 기술 개발 동향 및 전망

### 1. 개요

현재 국내외적으로 기존 아날로그 방송을 디지털 방송으로 전환하는 추세이다. 이러한 디지털 방송으로의 전환 시 필연적으로 오디오 부호화기를 사용하게 된다. 비교적 초기부터 시행되어 온 DAB (Digital Audio Broadcasting)로 부터 현재 핫이슈로 떠오르고 있는 T-DMB (Terrestrial Digital Multimedia Broadcasting)/S-DMB (Satellite DMB), DVB-T/H (Digital Video Broadcasting-Terrestrial/Handheld) 및 IPTV (Internet Protocol Television)까지 모두 이러한 오디오 부호화기가 적용된다. 본 절에서는 각종 디지털 방송 규격에 적용되는 오디오 부호화기에 대해서 알아본다.

### 2. 디지털 방송을 위한 오디오 부호화기

국내 지상파 디지털 방송의 규격은 오랜 논쟁 끝

에 미국 방식 (Advanced Television Systems Committee, ATSC)으로 결정되었다. ATSC의 오디오 규격은 Dolby 사의 AC-3이며 기본적으로 5.1 채널을 지원한다는 장점이 있다. 이에 반해 유럽식 지상파 디지털 방송인 DVB-T의 경우 주로 AAC를 채택하고 있다.

DAB 서비스는 주로 유럽에서 채택되었으며 1995년 세계 최초로 영국에서 시작되었다. 이러한 DAB에 이용되는 오디오 부호화기는 MUSICAM (MPEG-1 Audio Layer II)이다 [18]. MUSICAM은 최근 다른 방송에 주로 채택되는 AAC 계열에 비해서 압축율이 좋지 못하다. 따라서 주파수 효율 및 품질 개선을 위해서 HE-AAC v2를 도입한 2세대 DAB, 즉 DAB+가 추진되고 있다 [19].

국내 휴대형 디지털 방송은 위성 DMB (S-DMB)와 지상파 DMB (T-DMB)로 구분된다. S-DMB와 T-DMB는 MPEG2 및 MPEG4 오디오 부호화기를 기반으로 하지만 서로 상이한 부호화기를 채택한다. 즉, S-DMB의 경우 AAC를 기반으로 하고 이를 SBR로 확장한 HE-AAC를 이용한다. 그러나 T-DMB의 경우 BSAC를 이용한다. T-DMB는 태생적으로 DAB를 확장한 서비스이므로 DAB의 디지털

라디오 서비스를 기본적으로 지원한다. 따라서 T-DMB의 경우 BSAC과 더불어 DAB의 기본 오디오 부호화기인 MUSICAM 역시 지원하게 된다.

유럽의 휴대형 디지털 방송은 크게 Nokia 사 측에서 제안한 DVB-H 진영과 Qualcomm사에서 주도하고 있는 MediaFLO (Forward Link Only) [20] 및 우리나라에서 개발한 T-DMB의 3가지 방식이 경쟁하고 있다. DVB-H는 HE-AAC 및 HE-AAC v2의 AAC 계열과 Microsoft 사의 WMA (Windows Media Audio)를 모두 채택하고 있다. Qualcomm사의 MediaFLO는 HE-AAC 및 HE-AAC v2의 AAC 계열 오디오 부호화기를 채택하고 있다. T-DMB는 국내에서 채택된 BSAC과 함께 HE-AAC 및 HE-AAC v2를 모두 채택하고 있다. 미국의 경우 DVB-H 및 MediaFLO 서비스가 추진되고 있으며 양쪽 모두 HE-AAC 및 HE-AAC v2를 채택하였다. 중국의 경우 아직 휴대형 디지털 방송 규격이 완전히 확정되지 않았으며 여러 가지 규격에 대한 시험 서비스가 진행되고 있는 실정이다. 특이한 점은 북경, 상하이 및 광둥성의 경우 한국의 T-DMB에 기반한 지상파 DMB 방송이 서비스되고 있지만 2008년 북경올림픽 이후 규격이 어떻게 결정될지

〈표 1〉 휴대형 디지털 방송에서의 오디오 부호화기

휴대형 디지털 방송	오디오 부호화기	비 고
T-DMB	BSAC (한국, 중국) HE-AAC, HE-AAC v2 (유럽, 중국) MUSICAM (한국, 유럽, 중국)	MUSICAM의 경우, T-DMB의 기반인 DAB 서비스용임
S-DMB	HE-AAC (한국, 일본)	
DAB	MUSICAM (유럽)	
DAB+	MUSICAM, HE-AAC v2	방송 추진중
DVB-H	HE-AAC, HE-AAC v2, WMA	유럽 및 미국
MediaFLO	HE-AAC, HE-AAC v2	유럽 및 미국
ISDB-T	HE-AAC (일본)	

는 미지수이다. 이러한 T-DMB 기반 지상파 DMB의 경우 BSAC 및 HE-AAC v2가 모두 사용된다. 일본의 경우 ISDB-T one-seg가 지상파 DMB에 해당된다. 이러한 ISDB-T (Terrestrial Integrated Services Digital Broadcasting) one-seg는 HE-AAC를 채택하였다. 또한 위성 DMB의 경우 우리나라보다 먼저 서비스가 시작되었으나 상용화 과정에서 거의 성공하지 못하였다. 이러한 위성 DMB는 HE-AAC를 채택하였다. <표 1>은 각 휴대형 디지털 방송에 채택된 오디오 부호화기를 나타낸다.

### 3. IPTV 서비스를 위한 오디오 부호화기

IPTV는 차세대 방송통신 융합의 대표적 사례이다. DMB와의 차이점은 무선이 아니라 유선 서비스를 전제로 한다는 점이다. 이러한 IPTV 역시 고품위 오디오 부호화기가 필수적으로 요구된다. DMB와 유사하게 각 서비스 제공자에 따라 오디오 부호화기가 달라진다. 국내 IPTV 서비스 제공자인 KT, 하나로텔레콤, LG 테이콤 등은 AAC 및 AC-3를 채택하고 다음 사 등은 WMT (Window Media Technology)를 채택하고 있다.

### 4. 향후 발전 전망

지금까지의 디지털 방송에 적용된 오디오 부호화기는 낮은 비트율에서의 높은 음질 위주로 채택되었으며, 스테레오 오디오 부호화기가 대부분이었다. 그러나 앞으로의 오디오 부호화기는 오류 강인성, 다채널, 현장감 등의 품질 위주로 발전하게 될 전망이다.

국내 지상파 DMB에서 채택하고 있는 BSAC의 경우, 비트 오류에 강인한 가변 부호화기이다. 그렇

지만 아직 HE-AAC 등에서 사용되는 SBR을 적용한 응용 예가 없다는 단점이 있다. 따라서 BSAC에 SBR을 추가하는 연구가 진행되고 있다 [21]. 그리고 기존 스테레오 오디오를 탈피한 5.1 채널 등의 다채널 오디오 역시 시도되고 있으며 기존 MPEG 오디오 규격에 이미 반영되어 있다. 고품질 디지털 방송의 경우 이러한 다채널 오디오를 채택하고 있으며 소비자의 요구에 따라서 다채널 오디오를 채택하는 비율이 점차 늘어날 것이다.

## VI. 결 론

본 논문은 디지털 방송을 위한 오디오 부호화 기술 동향에 대한 것으로서 첫째, 심리음향 모델, 양자화 및 비트 할당 등 오디오 부호화 기술의 원리를 소개하였고, 둘째로 널리 사용되는 MPEG-1, MPEG-2 LSF, MPEG-2 AAC, MPEG-4 AAC, HE-AAC, BSAC, AC-3와 같은 표준 오디오 부호화기들의 알고리즘 및 특성을 기술하였다. 그리고 오디오 품질 평가 방식의 기술 동향을 살펴보았고 마지막으로 디지털 방송을 위한 오디오 부호화기 기술의 개발 동향과 전망을 기술하였다.

향후에는 네트워크의 상태에 적응적인 QoS 기법 연구 등에 활용될 수 있는 오디오의 품질 평가 방식에 대한 연구가 필요할 것이다. 또한 최근 정보통신 기술 및 초고속통신망이 급속도로 발전하고 있고 기존의 음성, 오디오, 영상, 데이터 정보가 점점 통합되어가는 추세에 발맞추어 다양한 멀티미디어 서비스가 제공되는 부호화기의 융합도 연구되어야 할 중요한 분야로 판단된다.

● 참고 문헌 ●

- [1] M. Bosi and R. E. Goldberg, Introduction to digital audio coding and standards, Kluwer Academic Publishers: Norwell, MA, 2002.
- [2] T. Painter and A. Spanias, "Perceptual coding of digital audio," Proceedings of the IEEE, vol. 88, no. 4, pp.451-513, Apr. 2000.
- [3] E. Zwicker and H. Fastl, Psychoacoustics: Facts and models, Springer, 1999.
- [4] N. Iwakami, T. Moriya, and S. Miki, "High-quality audio coding at less than 64 kb/s by using transform-domain weighted interleaved vector quantization (TwinVQ)," in Proc. ICASSP, pp. 3095-3098, May 1995.
- [5] D. Schulz, "Improving audio codecs by noise substitution," J. Audio Eng. Soc., vol. 44, pp. 593-598, July/Aug. 1996.
- [6] ISO/IEC 11172-3, Information technology - Coding of moving pictures and associated, audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio, Aug. 1993.
- [7] ISO/IEC 13818-3, Information technology - Generic coding of moving pictures and associated audio information - Part 3: Audio, Dec. 1997.
- [8] ISO/IEC 13818-7, Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC), Dec. 1997.
- [9] T. Ziegler, A. Ehret, P. Ekstrand, and M. Lutzky, "Enhancing MP3 with SBR: Features and compatibilities of the next MP3PRO algorithm," presented at the 112nd AES Convention, Preprint 5560, May 2002.
- [10] 3GPP TS 26.401, Enhanced aacPlus general audio codec: General Description, Dec. 2004.
- [11] S. H. Park and Y. B. Kim, "Multi-layered bit-sliced bit rate scalable audio coding," presented at the 103rd AES Convention, Preprint 4520, Oct. 1997.
- [12] F. Pereira and T. Ebrahimi, The MPEG-4 Book, Prentice Hall, 2002.
- [13] ISO/IEC JTC1/SC29/WG11 (MPEG) Document N6455, Call for Proposals on Spatial Audio Coding, Munich, Mar. 2004.
- [14] C. Todd, et al., "AC-3: Flexible perceptual coding for audio transmission and storage," presented at the 96th AES Convention, Preprint 3796, Feb. 1994.
- [15] ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, Feb. 2001.
- [16] ITU-T Recommendation P.563, Single ended method for objective speech quality assessment in narrow-band telephony applications, Apr. 2004.
- [17] ITU-R Recommendation BS.1387, Method for objective measurements of perceived audio quality, Dec. 1998.
- [18] ETSI EN 300 401 v1.3.3, Radio Broadcasting Systems: Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers, May 2001.
- [19] ETSI TS 102 563 v1.1.1, Digital Audio Broadcasting (DAB): Transport of Advanced Audio Coding (AAC) audio, Feb. 2007.
- [20] <http://www.qualcomm.com/media/lo/index.html>.
- [21] M. Kim, E. Oh, and J. Kim, "High quality scalable audio codec," in Proc. SPIE Multimedia Systems and Application X, vol. 6777, pp. 67770E-1-11, Sept. 2007.

## 필자소개



### 최승호

- 1991년 : 한양대학교 전자공학과 (공학사)
- 1993년 : 한국과학기술원 전기및전자공학과 (공학석사)
- 1999년 : 한국과학기술원 전기및전자공학과 (공학박사)
- 1996년 ~ 2002년 : 삼성종합기술원 전문연구원
- 2002년 ~ 현재 : 서울산업대학교 전자정보공학과 조교수
- 주관심분야: 음성 및 오디오 코딩, 멀티미디어신호처리, HCI



### 김홍국

- 1988년 : 서울대학교 제어계측공학과 (공학사)
- 1990년 : 한국과학기술원 전기및전자공학과 (공학석사)
- 1994년 : 한국과학기술원 전기및전자공학과 (공학박사)
- 1990년 ~ 1998년 : 삼성종합기술원 전문연구원
- 1998년 ~ 1998년 : ㈜엠엠씨테크놀로지 선임연구원
- 1998년 ~ 2003년 : AT&T Labs-Research, Senior Member Technical Staff
- 2003년 ~ 현재 : 광주과학기술원 정보통신공학과 부교수
- 주관심분야 : 멀티미디어신호처리, 실시간 User Interface, 음성 및 오디오 코딩



### 석용호

- 1992년 : 한양대학교 전자공학과 (공학사)
- 1994년 : 한국과학기술원 전기및전자공학과 (공학석사)
- 2000년 : 한국과학기술원 전기및전자공학과 (공학박사)
- 2006년 ~ 현재 : ㈜엠큐브웍스
- 주관심분야 : 차세대 휴대형 멀티미디어, 오디오 코딩