

# 최신 MPEG 오디오 코딩 기술

방 희 석\*

## 1. 서 론

전통적으로 MPEG 오디오에서는 MPEG1/2 Layer3 (MP3), Advanced Audio Coding (AAC), High-Efficiency AAC (HE-AAC), HE-AAC v.2 등 지각 부호화에 근거한 오디오 압축 기술에 중점을 두어왔으나, 최근에는 이와 같은 지각 부호화에서 벗어나 다양한 분야의 오디오 코딩 기술들을 표준화하고 있다. 특히, 최근의 중요 기술들은 무손실 및 스케일러블 무손실 압축 기술, 다채널 오디오 코딩 및 객체 기반 오디오 코딩 기술, 저비트율에서의 오디오 및 음성을 위한 범용 코딩 기술 등 크게 보아 3가지 정도로 구분할 수 있다.

이 중 첫 번째 기술에서 무손실 압축 기술에 해당되는 것은 Audio Lossless Coding (ALS)[1], 그리고 스케일러블 무손실 압축 기술에 해당되는 것은 Scalable Lossless Coding (SLS)[2]를 들 수 있다. ALS의 경우는 무손실 오디오 압축을 목적으로 하고 있으며, 그 대상은 임의의 샘플링 주파수, 32bit까지의 비트 해상도, 2<sup>16</sup>개의 채널까지를 포함한다[3]. 따라서, 44.1kHz의 샘플링 주파수와

16bit 해상도를 가지는 CD 수준의 음질에서부터 96kHz/192kHz의 샘플링 주파수와 24bit의 해상도를 가지는 고음질 오디오를 모두 지원한다. SLS의 경우에는 AAC로 구성되는 코어 레이어 (core layer)와 잔류신호 (residual signal)로 구성되는 향상 레이어 (enhancement layer)로 비트열이 구성된다. 따라서, 이는 비트율에 따라서 AAC 수준의 음질에서부터 무손실 음질까지를 스케일러블한 방식으로 지원할 수 있음을 의미한다[4].

최신 MPEG 오디오 기술 중 둘째는 다채널 오디오 코딩 및 객체 기반 오디오 기술로써, 전자는 MPEG Surround[5], 후자는 Spatial Audio Object Coding (SAOC)[6]을 들 수 있다. MPEG Surround는 5.1 채널 등 다채널 신호를 모노 혹은 스테레오 신호로 다운믹스하고 동시에 공간 파라미터들을 추출하여 전송하고, 이 다운믹스 신호와 공간 파라미터들로부터 다채널 혹은 바이노럴 (binaural) 신호를 생성하는 기술이다[7,8]. MPEG Surround가 채널 기반의 압축 기술이라면, SAOC는 오디오 객체 (audio object) 기반의 압축 기술이다. 즉, SAOC에서는 오디오 객체들이 다운믹스되고 동시에 객체 파라미터들 (object parameters)이 추출되어 전송되고, 이들 신호를 이용하여 각 객체들을 생성 후 일종의 믹싱 과정을 통해서 최종적인 채널 신호들을 만들게 된다. 따라서, 이 믹싱 과정에서 사용자가 객체들에 대

※ 교신저자(Corresponding Author): 방희석, 주소: 서울시 광진구 군자동 98 세종대학교 전자공학과(143-747), 전화: 02)3408-3727, FAX: 02)3408-4329, E-mail: hspang@sejong.ac.kr

\* 세종대학교 전자공학과 조교수

※ 본 논문은 지식경제부 IT 산업원천 사업의 지원을 받았습니다.

해 여러 가지 기능별로 조절할 수 있는 입력이 가능한 것이 가장 큰 특징이라고 할 수 있다[9].

마지막으로 저비트율에서 음성과 오디오를 모두 지원하는 기술은 Unified Speech and Audio Coding (USAC)을 들 수 있다[10,11]. 이는 MPEG 오디오에서 HE-AAC v.2 이후 오디오 뿐 아니라 음성에도 사용 가능한 범용 오디오 코덱의 개발에 착수해서 현재 표준화가 진행 중인 기술이다[10]. 또한, USAC에서는 HE-AAC v.2 및 AMR-WB+의 기술들을 조합 및 개선하여 추가적인 성능 향상을 수행하였기 때문에, 음성 및 오디오에 모두 적용 가능하다는 범용성 이외에 기존의 기술들에 비해 더욱 낮은 압축률 및 좋은 음질을 지원할 수 있다는 장점이 있다.

본 논문에서는 이상에서 소개한 ALS, SLS, MPEG Surround, SAOC, USAC 등 5가지의 최신 MPEG 오디오 기술에 대해서 그 구조와 성능 등을 간단히 소개하고자 한다.

## 2. ALS

그림 1에서 ALS의 부호화기 및 복호화기의 기본 구조를 보이고 있다[3]. 우선 부호화기에서 입력 오디오 데이터는 프레임으로 나뉘게 되고, 블록 길이 스위칭 (block length switching)을 위해서 프레임 내에서도 다시 블록으로 나뉘게 된다. 각각의 블록에서는 적응 예측(adaptive prediction)을 이용해 예측 잔류신호(prediction residual)가 계산되는데, 이 과정에서 단기간 예측(short-term prediction)과 장기간 예측(long-term prediction)은 같이 사용될 수 있다. 채널간의 리던던시는 조인트 채널 코딩(joint channel coding)으로 제거될 수 있는데, 이는 각 채널 신호의 차 등을 이용하거나 다채널 코딩을 이용하게 된다. 이후 남아있는 예측 잔류신호는 Rice code 혹은 Block Gilbert-Moore Code (BGMC) 등의 방법을 이용하여 엔트로피 부호화된다. 복호화기에서는 부호

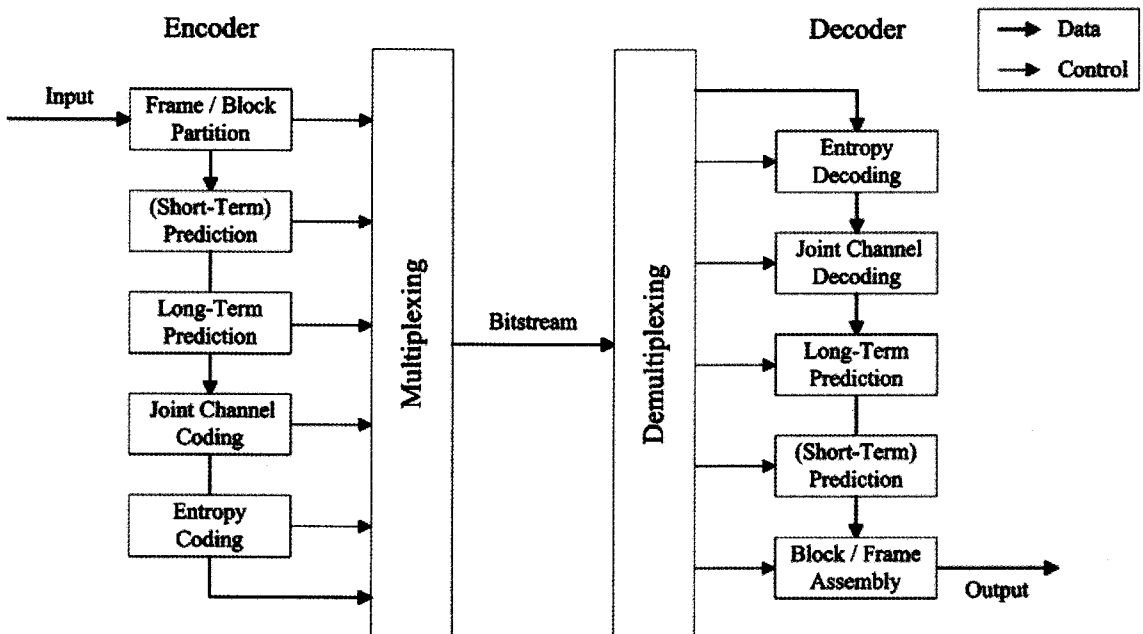


그림 1. ALS의 부호화/복호화 구조(3)

표 1. ALS의 압축률(%) [3]

Format (stereo)	FLAC (best)	ALS (low)	ALS (medium)	ALS (max)
48kHz/16bit	48.6	46.5	45.3	44.6
48kHz/24bit	68.4	63.9	63.2	62.7
96kHz/24bit	56.7	47.4	46.3	46.1
192kHz/24bit	45.3	38.4	37.6	37.5
Average	54.8	49.1	48.1	47.7

화기에서의 과정이 역순으로 진행되며, 최종적으로는 원본과 동일한 신호를 생성한다.

표 1에서는 ALS에 대해 다음과 같이 정의된 압축률 C를 보이고 있다[3].

$$C = \frac{\text{Compressed File Size}}{\text{Original File Size}} \times 100\% \quad (1)$$

이 결과는 MPEG4 무손실 코딩을 위한 표준 스테레오 오디오 시퀀스로부터 계산된 것이다. 이에 따르면 최대 압축률의 MPEG4 ALS는 FLAC[12]에 비해 최대 10% 내외의 압축률 개선

을 보임을 알 수 있다. 또한, 48kHz, 16bit 스테레오 샘플에 대해서 1.7GHz 펜티엄 기반 PC로 MPEG4 ALS 복호화를 수행하였을 경우, 최소 압축률 옵션에서는 약 1.6%, 최대 압축률 옵션에서는 약 18.2%의 CPU 점유율을 필요로 한다[3]. 이상에서의 MPEG4 ALS에 대해 좀 더 자세한 내용은 [1,3,13,14]를 참고하기 바란다.

### 3. SLS

SLS는 기존의 지각 오디오 코딩 기술(AAC)을 스케일러블한 방식으로 고해상도 오디오 신호의 무손실 코딩으로 확장한 것이라고 할 수 있다. 그림 2에서 SLS 부호화기 및 복호화기의 구조를 보이고 있는데[4], 우선 위쪽의 부호화부에서 입력 신호는 AAC 부호화기로 부호화된 후 SLS 알고리즘으로 입력되며, SLS 부호화부에서는 원본 신호와 AAC 신호의 차이를 이용하여 AAC 수준의 음질에서부터 무손실 음질까지를 스케일러블하

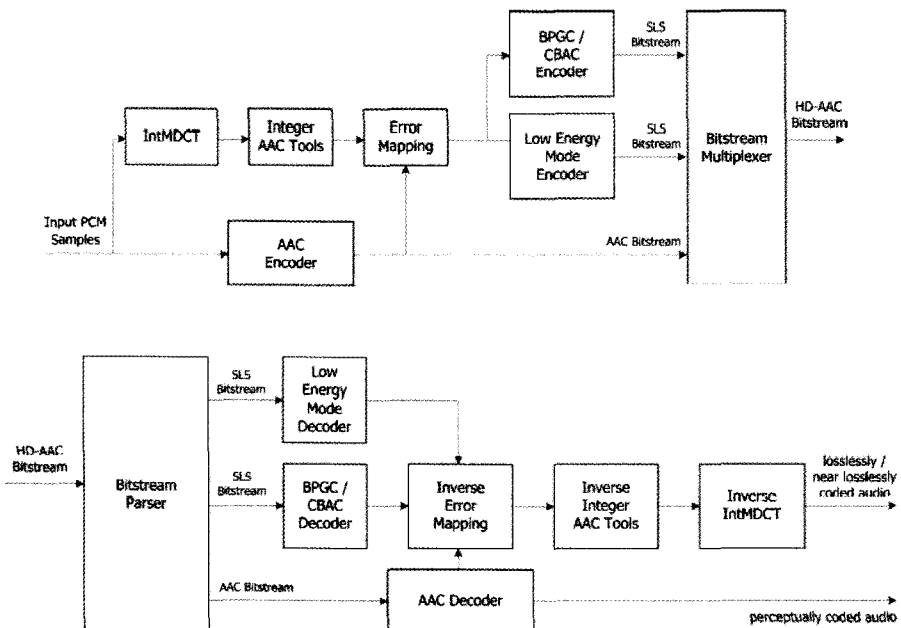


그림 2. SLS의 부호화/복호화 구조(4)

게 조정할 수 있다. 즉, 전체 비트열은 코어 레이어와 향상 레이어로 구성되며, 전체 비트율에 따라서 향상 레이어가 어느 수준의 음질을 지원하게 될지 결정된다.

SLS는 lifting scheme 혹은 ladder network을 사용하여 MDCT의 가역 정수 구현을 가능하게 하는 정수 MDCT (IntMDCT)에 기반하고 있다 [15]. 즉, 원본 신호를 IntMDCT 변환한 후 AAC에서의 MDCT 계수들을 빼면 잔류신호 (residual signals)를 얻을 수 있는데, 이 잔류신호들에 대해 향상 레이어에서 부호화한다. 부호화시에는 bit-plane 코딩을 모든 스케일 팩터 밴드에서 MSB에서 시작하여 LSB를 향해 진행하게 된다. 이러한 방식으로 bit-plane 코딩 프로세스는 양자화 잡음의 전반적인 스펙트럼 형태를 보존하게 되는데, 이는 AAC의 노이즈 세이핑 프로세스의 결과이며 따라서 AAC의 지각 모델의 장점을 활용하게 된다. 이 잔류신호들에 대해 Bit-Plane Golomb Code (BPGC) 코딩과 Context-Based Arithmetic Coding (CBAC)을 같이 사용하는 것을 기본으로 하며, 컨텐트에 따라서는 low-energy-mode 코딩을 사용하기도 한다[4].

SLS 복호화부에서는 그림 2의 아래쪽과 같은 구조로 신호를 복원한다[4]. 즉, AAC로부터 MDCT 계수를 계산한 후 향상 레이어의 정보로부터 잔류 신호의 MDCT를 계산하고 이 둘을 합한다. 그 후

추가적인 틀들을 거친 후 역 IntMDCT를 수행하면 최종적으로 시간영역 신호를 출력한다.

SLS는 이상에서의 AAC와 SLS의 조합 구조 이외에도 독립적으로 동작하는 stand-alone 모드를 포함한다. SLS는 향상 레이어의 비트율에 따라서 음질이 결정되는데, 즉 향상 레이어에 비트가 할당되지 않을 경우 코어 레이어의 AAC 수준의 음질이 지원되고 향상 레이어에 비트가 할당됨에 따라서 음질이 점점 향상되어 최종적으로 무손실 압축을 달성하면 원본 수준의 음질이 지원된다. 표 2에서 AAC(128kbps)+SLS 모드 및 SLS stand-alone 모드에 대한 무손실 압축 시의 압축률 및 평균 비트율을 보이고 있다[4]. 압축률은

$$CR = \frac{\text{Original File Size}}{\text{Compressed File Size}} \quad (2)$$

로 정의되며, ALS에서의 식 (1)과는 역의 형태이다. 실험에 사용된 샘플들은 MPEG4 무손실 코딩을 위한 표준 스테레오 오디오 시퀀스이다.

연산량과 관련해서는 AAC+SLS의 형태의 경우 최종 출력 샘플당 약 300 cycle/sample 이하의 32bit 연산을 필요로 한다[4]. SLS와 관련된 자세한 내용을 위해서는 [2,4,16]을 참고하기 바란다.

#### 4. MPEG Surround

HE-AAC v.2에서 사용된 Parametric Stereo 틀의 경우 모노 다운믹스(downmix) 신호와 스테

표 2. SLS의 압축률 및 비트율[4]

	SLS+AAC @ 128 kbps/Stereo (AAC @ 48 kHz sampling rafe)		SLS Stand-Alone	
	Compression Ratio	Average Bit Rate (kbps)	Compression Ratio	Average Bit Rate (kbps)
48 kHz/16 bit	2.09	735	2.20	698
48 kHz/24 bit	1.55	1490	1.58	1454
96 kHz/24 bit	2.09	2201	2.13	2160
192 kHz/24 bit	2.60	3543	2.63	3509
Overall	2.08	1992	2.12	1955

레오 파라미터들(stereo parameters)을 이용하여 스테레오 신호를 합성하는 기술이다[17]. 따라서, MPEG 오디오에서는 이를 확장하여 5.1채널 등의 다채널 신호의 압축에 활용하고자 하는 표준화를 진행하였고, 초기에 이 기술을 Spatial Audio Coding (SAC)로 명명하고 표준화를 진행하던 과정 중에 MPEG Surround란 이름으로 개정하였다.

MPEG Surround의 기본 구조는 그림 3에 표시되고 있다[8]. 우선 부호화부에서 다채널 오디오 신호는 QMF 분석(Quadrature Mirror Filter analysis)과 Nyquist 필터 분석(Nyquist filter analysis)을 순차적으로 수행하여 hybrid 서브밴드 영역(hybrid subband domain)으로 변환된다. Hybrid 서브밴드 영역에서는 두 가지 작업, 즉 다채널 신호로부터 공간 파라미터를 추출하는 것과 다운믹스 신호를 생성하는 것이 수행된다. 마지막으로 공간 파라미터들은 양자화 및 부호화하여 비트열로 생성되고, 다운믹스 신호는 Nyquist 필터 합성(Nyquist filter synthesis)과 QMF 합성(QMF synthesis)를 통해 시간 영역의 다운믹스 신호로 생성된다. 일반적으로 다운믹스 신호는 비트율을 더욱 낮추기 위해서 MP3나 AAC와 같은 오디오 코딩을 추가로 수행하게 된다.

부호화부에서는 우선 MP3나 AAC 등에 의해

부호화되어 있는 다운믹스 신호가 먼저 복호화되고, 이의 결과인 PCM 샘플들은 QMF 및 Nyquist 필터 분석을 순차적으로 수행하여 hybrid 서브밴드 영역으로 변환된다. 예외적으로 다운믹스 코덱이 HE-AAC이고 32kHz, 44.1kHz, 48kHz 중 하나의 샘플링 주파수를 사용할 경우 MPEG Surround와 동일한 64개의 QMF 구조를 가지기 때문에, 이 둘 사이는 QMF 서브밴드 영역에서 바로 연결이 될 수도 있고 시간 영역에서 연결이 될 수도 있다[5]. Hybrid 서브밴드 영역의 다운믹스 신호는 복호화된 공간 파라미터가 적용되어 다채널 신호로 변환이 되고, 마지막으로 Nyquist 필터 및 QMF 합성을 거쳐 시간축의 다채널 신호가 출력된다.

MPEG Surround에서는 이상의 다채널 복호 모드 이외에도 가상 5.1 채널을 지원하여 기존의 오디오 재생 기기에서 헤드폰이나 이어폰을 이용한 재생이 가능하게 하는 바이노럴 모드(binaural mode) 및 이미 3D 정보가 포함되어 있는 신호에 대한 다채널 합성 방법을 지원하는 3D 스테레오 모드(3D stereo mode)를 지원한다. 또한, 부가 정보 없이 스테레오 채널로부터 5.1 채널을 합성하는 enhanced matrix 모드, 아티스틱 다운믹스 신호에 적용되는 Arbitrary Downmix Gains (ADGs)

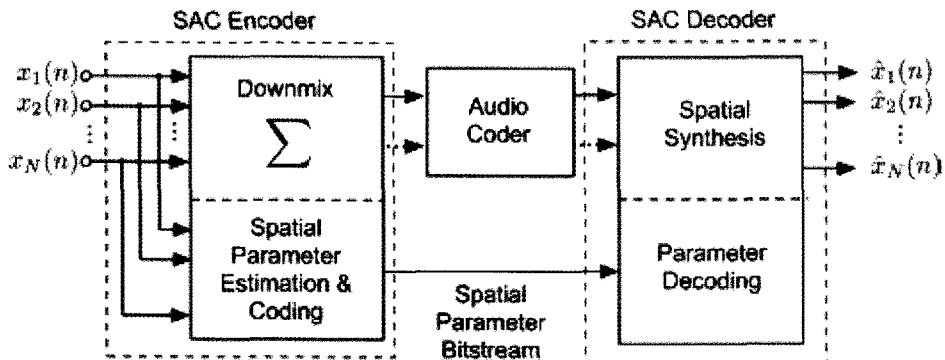


그림 3. MPEG Surround의 부호화/복호화 구조[8]

틀, MPEG Surround가 아닌 기술에 의해 생성된 다운믹스 신호에 적용되는 행렬 호환성(matrix compatibility) 틀, 고비트율을 사용하여 고음질을 지원할 수 있는 잔차 신호 코딩(residual coding) 틀, 박수 소리 등 특수한 소리에 대한 음질을 향상시키는 GES (Guided Envelope Shaping) 틀 등이 지원된다.

MPEG Surround의 성능을 청취 평가를 이용하여 그림 4에서 보이고 있는데[7], 이의 청취 평가 환경에 관한 자세한 내용은 [7]을 참고하기 바란다. 이 결과는 AAC와 MPEG Surround를 사용한 것(AAC+MPS), MP3와 MPEG Surround를 사용한 것(MP3+MPS), Fraunhofer의 자체 기술인 MP3 Surround (MP3Surround) [18], AAC와 Dolby Prologic II를 사용한 것(AAC+DPLII) [19], AAC로 다채널 코딩을 적용한 것(AAC-MC) 등 5개를 비교하고 있다. 이 때, 모든 코덱의 경우 다운믹스 코덱과 부가 정보의 비트율의 합이 160 kbps가 되도록 하였다. 또한, 청취 평가는 MUSHRA 테스트[20]에 의해 진행되었으므로 위에서 설명

한 5개의 비교 대상음과 함께 은닉 원음(Ref), 원음을 3.5kHz 저역통과필터에 통과시킨 앵커음(BW35)의 결과를 같이 보이고 있다. 결과에 따르면 MPEG Surround는 다운믹스 코덱의 종류와 상관없이 다른 기술들과 95% 신뢰 구간이 겹치지 않는 등 월등한 음질을 보임을 알 수 있다. 특히, MUSHRA 테스트에서는 약간의 차이에 대해 많은 점수 차이를 부여하는 경향이 있기 때문에, 원음에 대한 MPEG Surround의 음질 차이는 실제로는 매우 적다. AAC 멀티 채널의 경우 채널당 약 32kbps만을 사용하기 때문에 음질이 매우 낮게 나왔음을 알 수 있다.

MPEG Surround의 경우 다채널 신호를 생성하기 때문에 복잡도가 높은 편이다. 따라서, 일반 모드 대신에 저전력 모드 (low power mode)를 선택하면 약간의 음질을 희생하고 연산량을 50% 정도 낮출 수 있도록 하고 있다[5]. MPEG Surround에 대한 자세한 내용은 [5,7,8,21-23]을 참고하기 바란다.

### 5. SAOC

MPEG Surround의 표준화가 완료되어 가면서 MPEG 오디오에서는 객체 기반 오디오 코딩(object-based audio coding)에 대한 논의를 시작하였다. MPEG 오디오에서는 2007년 1월에 SAOC에 대한 Call for Proposal (CfP) 문서가 나왔고[24], 계속 표준화를 진행하여 최근에 FDIS 문서가 나와서[6] 조만간 표준화 작업을 완료할 예정이다.

본 절에서는 [25]의 내용에 기반하여 SAOC에 대해 설명한다. 그림 5에서 SAOC의 객체 기반 오디오 코딩 개념도를 보이고 있다[25]. 우선 부호화부에서는 오디오 객체(audio object)들이 입력된 후 각각의 객체들을 다운믹스하여 다운믹스

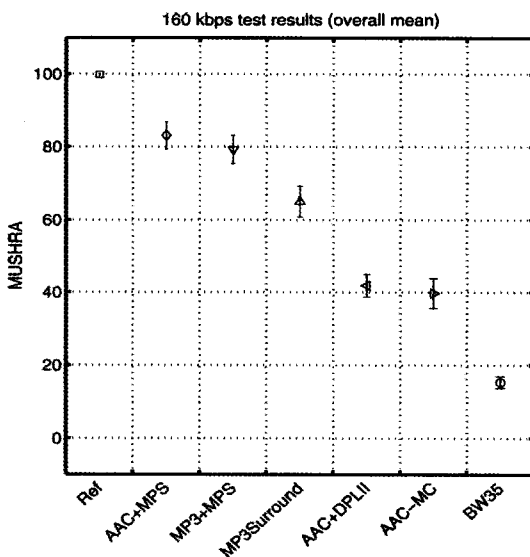


그림 4. MPEG Surround의 청취 평가 결과(7)

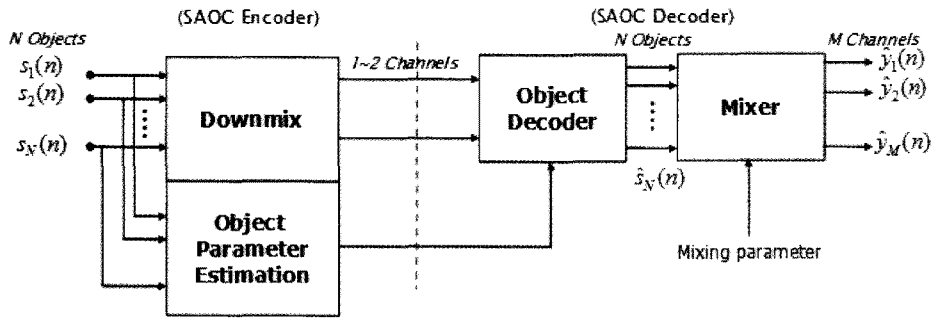


그림 5. SAOC의 객체 기반 오디오 코딩 개념도(25)

신호를 생성하고, 동시에 각 객체들로부터 객체 파라미터 (object parameter)들을 추출한다. 이들은 부호화되어 비트열을 구성하고 전송되며, 복호화부에서는 이들을 이용하여 각각의 오디오 객체들을 합성하게 된다. 이 오디오 객체들을 최종적으로 사용자가 듣기 위해서는 채널 신호로 구성되어야 하므로 믹서에서의 믹싱 작업을 통해 객체 신호들은 채널 신호로 변환된다. 객체 기반 오디오 코딩이 다채널 오디오 코딩과 다른 점은 여러가지가 있겠지만, 우선 객체 기반 오디오 코딩의 입력 신호는 음악을 구성하는 각각의 악기음, 음성, 혹은 이들의 조합으로 이루어진 신호 등의 오디오 객체 (audio object)이며, 이는 다채널 오디오 코딩에서 입력신호가 이미 객체들이 믹싱된 채널 신호라는 것과는 구별된다. 또한, 다채널 오디오 코딩에서는 출력 신호를 바로 채널별로 재생함에 반해, 객체 기반 오디오 코딩에서는 믹싱이라는 과정이 추가로 존재하며, 이 과정 중에서 사용자가 렌더링 행렬(rendering matrix)라는 기술을 사용하여 믹싱 파라미터 등을 조절하거나 입력할 수 있다. 한편, 이 렌더링 행렬과 관련하여 사용자 입력이 없거나 잘못된 값이 부여된 경우 복호화 자체가 불가능할 수도 있으므로, SAOC에서는 프리셋 신택스(preset syntax)를 이용하여 콘텐츠 제공자 혹은 다른 사용자가 기본적인 렌더링

신(rendering scene)을 제공할 수 있다. 또한, 과도한 특정 객체의 증폭/감쇄 등 극단적인 렌더링의 요구는 음질을 크게 저하시키는 문제를 발생시키게 되는데, 이에 대한 보호를 위해 렌더링 행렬의 범위를 제한하기 위한 가이드 정보를 전송하는 방법에 대해서도 논의가 되고 있다.

SAOC 복호화부에 대해서 그 사용 시나리오에 따라서 2가지의 복호화기 구조가 존재한다. 즉, 출력 신호가 모노, 스테레오, 혹은 3D 재생을 위한 바이노럴 신호일 경우 그림 6의 오른쪽처럼 다운믹스 프로세서(downmix processor)를 이용한 SAOC 디코더 처리 모드(SAOC decoder processing mode)를 사용한다. 이는 모노 혹은 스테레오, 바이노럴 신호 출력을 위해서 구조나 형태를 최소화 및 최적화한 형태로 볼 수 있다. 한편, 사용자가 의도하는 출력 신호가 다채널 신호일 경우, SAOC는 그림 6의 왼쪽과 같이 SAOC 트랜스코더 처리 모드(SAOC transcoder processing mode)로 동작하며, 이 모드에서의 출력은 SAOC에 의해 생성된 다운믹스 신호와 MPEG Surround 비트열이 된다. 따라서, 이 출력은 다시 MPEG Surround 복호화기에 연결되어 추가적인 복호화 작업을 거치게 되는데, 이는 MPEG Surround가 현존하는 MPEG 다채널 오디오 코딩 기술 중 비트율 대비 최상의 음질을 지원하는 기

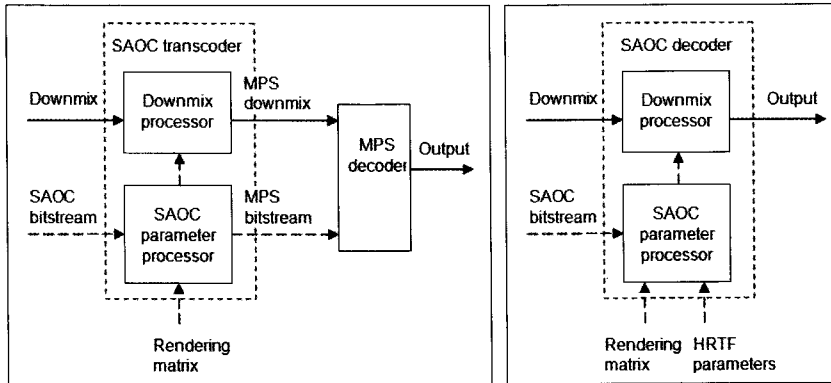


그림 6. SAOC 복호화기의 트랜스코더 처리 모드 (왼쪽) 및 디코더 처리 모드 (오른쪽)(6)

술이기 때문이다. 한편, 하나의 시스템 안에서 SAOC 트랜스코더와 MPEG Surround가 구현되는 경우는 굳이 MPEG Surround 비트열을 생성하지 않고 양자화되지 않은 파라미터 레벨에서 SAOC-MPEG Surround 연결이 가능하다.

SAOC의 가장 큰 응용 분야로 예상되는 것이 바로 인터랙티브 리믹스 (interactive re-mix)이다. 즉, 프로듀서의 의도에 따라 오디오 객체들이 적절히 믹싱된 음악 신호에 대해 사용자가 마치 자신이 프로듀서가 된 것처럼 각각의 오디오 객체에 대해 독립적인 제어, 즉 리믹스 기능을 제공한다. 예를 들어, 특정 오디오 객체의 크기를 줄이거나 키울 수 있고, 특정 사운드 스테이지 안에서 오브젝트의 공간적인 위치를 변경하는 등의 제어가 가능해진다. 극단적으로는 보컬 객체를 제거하여 가라오케에 활용하는 등 특정 오브젝트의 신호를 제거하는 것도 가능하다. 단, SAOC의 일반 기능으로는 특정 오브젝트를 완전히 제거하는 것은 어렵기 때문에 잔류신호 코딩 틀을 이용하여 잔류신호를 비트열에 삽입하는 방법을 이용하여 이를 지원하게 된다. 이 인터랙티브 리믹스의 개념을 방송에 적용하는 것도 가능하다. 예를 들어, 어학 학습용 방송에서 한국어와 외국어의 음성을 선택적으로 청취하거나, 스포츠 중계에서 관중의

소리와 캐스터의 음성 크기의 비율을 취향에 맞게 조절하는 것, 영화/드라마 등에서 효과음 등 배경 신호의 과도한 음량으로 배우의 대사가 묻히는 경우가 많이 있는데 이때 시청자에게 사운드 신(sound scene) 구성의 자유를 제공함으로써 이러한 문제를 해결할 수 있다. 특히 방송의 경우 기존의 방식과 같이 각 트랙에 대해 독립적인 오디오 부호화를 통해 전송할 경우 필요한 대역폭이 전송되는 트랙 수에 비례하여 증가하기 때문에 상업적으로 적합하지 않지만, SAOC를 이용하면 증가되는 대역폭이 기존 오디오 신호 대역의 100~150% 정도면 가능하기 때문에 전송 효율 면에서 큰 이득이 있다. 또한, SAOC는 객체 파라미터를 무시할 경우 다운믹스 신호만으로 재생이 가능한 하방 호환성(backward compatibility)을 보장하기 때문에, 기존 방송 인프라의 부가 서비스로 도입이 용이하여 차세대 방송 서비스로의 다양한 활용을 기대할 수 있다.

인터랙티브 리믹스와 함께 SAOC의 주요 응용 분야로 예상되는 곳이 바로 원격 회의(teleconference)이다. 보통 모노 혹은 최대 스테레오로 전송되는 원격 회의 환경의 경우 하나의 스피커를 통해 원격의 참가자들의 음성이 모두 동일한 물리적 위치에서 출력되기 때문에, 현재 발화하는 참



가자를 식별하기에 어렵고 화자 간 중첩에 의해 대화 내용의 이해에 어려움이 있다. 이러한 문제를 극복하고 원격 회의 실제감과 현장감을 높이기 위한 연구가 활발히 진행되고 있는데, SAOC는 이와 같은 환경에 활용하기 좋은 부호화 기술이다. SAOC를 이용할 경우 종래의 모노 전송 채널을 이용하더라도 각각의 원격 참가자의 음성을 가상의 회의실 공간상의 독립된 물리적 위치에서 출력하는 것이 가능하고, 이를 통해 음성 명료도를 증가시키고 발화하는 화자를 명확하게 식별하는 것이 가능하게 된다. 특히, 화상 회의(video conferencing)의 경우, 화면상의 원격 참가자의 위치와 음성 재생 위치를 공간적으로 일치시킴으로써 청취자의 실제감을 높이는데 활용될 수 있다.

마지막으로 게임 혹은 rich media의 경우도 앞선 사례와 마찬가지로 SAOC의 활용 가능성이 높은 응용 분야이다. 게임의 가상 공간속 플레이어나 유저 간의 대화나 효과음을 각각의 오디오 객체로 표현하여 SAOC로 전송하게 되면 전송 효율에서 매우 유리한 장점을 가진다. SAOC에 대한 자세한 내용은 [6,9,25,26]을 참고하기 바란다.

## 6. USAC

각각 인간의 음성 생성 모델과 청각 모델을 기반으로 사용하는 음성 부호화 기술과 오디오 부호화 기술은 독자적으로 기술 및 서비스의 발전을 이루어 왔다. 하지만 최근 방송과 통신이 융합하는 방향으로 기술이 발전하면서 더 이상 음성과 오디오 신호를 별도의 콘텐츠로 분리하는 것이 어렵게 되었다. 따라서 하나의 부호화기로 음성과 오디오 신호 모두를 처리할 수 있는 범용 오디오/음성 부호화 기술이 필요하게 되었고, 이에 따라 MPEG 오디오에서는 음성과 오디오 신호를 통합적으로 부호화하는 MPEG-D USAC 기술의 표준

화를 2007년 10월 CfP[27]를 시작으로 2010년 1월에 Working Draft 6 (WD6)[10]를 승인하는 등 표준화 작업을 현재 진행하고 있다. 본 장에서는 [28]의 내용을 기반으로 USAC에 대해 설명한다.

USAC 부호화기는 그림 7의 왼쪽과 같이 스테레오 부호화를 위한 MPEG Surround 틀, 고주파 대역 부호화를 위한 enhanced Spectral Band Replication (eSBR) 틀, 입력신호의 특성을 분석하여 저대역 부호화 방법을 선택하기 위한 신호 분류기(signal classifier), 입력 신호의 특성에 따라 저대역 신호를 부호화하기 위한 Algebraic Code Excited Linear Prediction (ACELP), Transform Coded eXcitation (TCX) 및 AAC 틀, 그리고 서로 다른 저대역 부호화 방식 사이를 왜곡 없이 연결하기 위한 천이 기술 등으로 이루어진다. 따라서, USAC는 HE-AAC v.2, MPEG Surround, AMR-WB+ 등의 기술을 조합하고 각각의 세부 기술들에 대해 더욱 향상시킨 것으로 볼 수 있다.

USAC 부호화부의 동작을 좀 더 자세히 살펴보면, 우선 입력신호가 스테레오인 경우에 대해 MPEG Surround의 Two-To-One (TTO) 블록을 이용하여 다운믹스 신호 생성 및 공간 파라미터들을 추출하고, 생성된 모노 신호에 대해 eSBR 틀에서 고주파 대역 부호화를 위한 파라미터를 추출한다. eSBR은 HE-AAC 혹은 HE-AAC v.2의 SBR 기술에 저 비트율에서 하모닉을 이용하는 기술을 추가한 향상된 SBR 기술이다. 이제 저대역 모노 신호에 대해서는 신호 분류기를 이용하여 그 특성을 분석하는데, 신호가 음악 특성이 강한 경우 AAC 기반으로, 음성 특성이 강한 경우 ACELP 혹은 TCX 기반으로 부호화를 진행한다. AAC 기반 부호화 기술의 경우 엔트로피 부호화를 위해 허프만 부호화가 아닌 컨텍스트 기반의 산술 부호화 (context-based arithmetic coding)

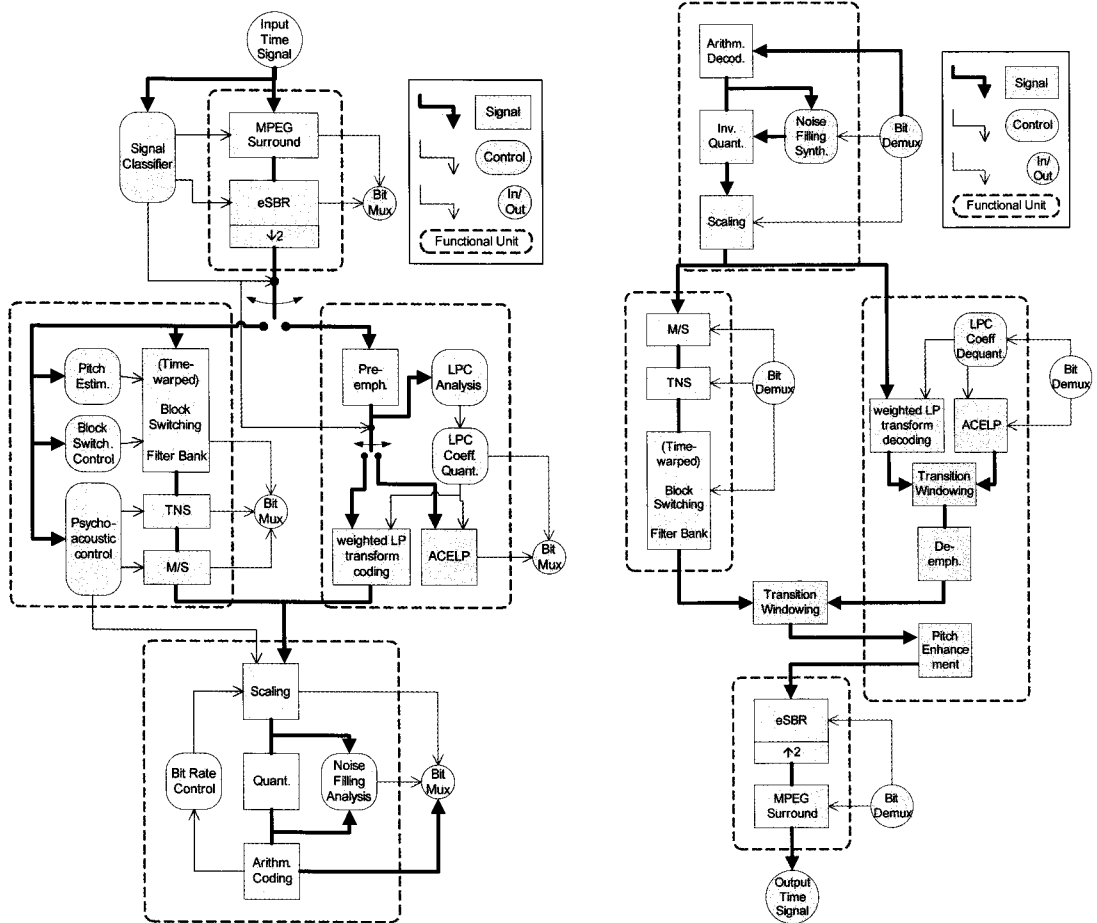


그림 7. USAC 부호화기 (왼쪽) 및 복호화기 (오른쪽) 구조(10)

를 사용하는 것, ACELP 혹은 TCX와 연결될 경우를 위한 새로운 윈도우를 추가한 것, time-warping 필터 뱅크를 이용하여 시간 영역에서 서로 다른 해상도를 갖는 주파수 분해 기능을 제공하는 것 등이 특징이다. ACELP와 TCX 기반 기술의 경우 둘 중 SNR의 이득이 더욱 높은 것을 사용하는 방식을 선택하여 부호화를 진행하게 되며, Linear Prediction Coefficients (LPC)의 양자화 방식, TCX에서의 Fast Fourier Transform (FFT) 대신에 Modified Discrete Cosine Transform (MDCT) 사용 및 이에 따른 다양한 천이 윈도우 사용, TCX의 엔트로피 부호화 방식

이 상기에서 기술한 AAC 방식의 산술 부호화 기술을 사용하는 것 등이 기존의 AMR-WB+ 기술에서 더욱 향상된 것으로 볼 수 있다.

그림 7의 오른쪽은 USAC의 복호화기 구조인데, 여기에서는 부호화기의 역순으로 복호화 과정을 진행한다. 즉, 입력 비트스트림은 AAC, ACELP, TCX 기술 중 해당 기술을 이용하여 복호화를 진행하고, 추후 모노인 경우는 eSBR 틀, 스테레오인 경우는 eSBR 및 MPEG Surround 틀을 거쳐서 신호를 출력하게 된다. USAC에서는 AAC, ACELP, TCX 등 3가지 저대역 부호화 기술을 사용하기 때문에, 복호화기에서의 각각의 등

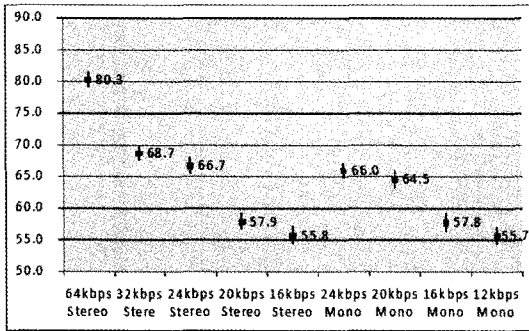


그림 8. USAC RM0 성능 평가 결과[28]

작모드 사이의 천이를 왜곡 없이 연결하기 위해 특별히 설계한 윈도우를 이용한다.

그림 8에서 CfP에 의한 제출 기술 중 최고의 음질을 보인 Reference Model 0 (RM0)의 음질 평가 결과를 보인다[28]. 이를 위한 아이템은 음성, 음악, 음성 및 음악 혼합신호 각각 4개씩 사용하여 전체 12개의 아이템을 이용하였으며, 이와 관련한 자세한 실험 환경은 [28]을 참고하기 바란다. 결과에서 알 수 있듯이 저 비트율인 12kbps에서는 55점대의 평가 점수를 얻었고, 고 비트율인 64kbps에서는 80점 정도의 평가 점수를 얻었다. USAC RM0 기술은 현존하는 저비트율 오디오-음성 코딩 기술들 중 최고 기술이라 할 수 있는 HE-AAC v.2 및 AMR-WB+에 비해 모든 경우에 대해 95% 신뢰도 수준에서 더 낮거나 혹은 동등한 수준이다[28].

USAC는 현재 표준화 과정을 계속 진행하고 있으므로, 최종 표준 기술은 현재보다 향상된 음질을 지원할 것으로 기대된다. USAC에 대한 자세한 내용은 [10,28-30]을 참고하기 바란다.

## 7. 결 론

본 논문에서는 ALS, SLS, MPEG Surround, SAOC, USAC 등 5가지의 최신 MPEG 오디오

코딩 기술에 대해 소개하였다. 각 기술의 세부적인 기술 내용은 각 장에서 기술한 참고 문헌을 참조하기 바란다.

## 참 고 문 헌

- [ 1 ] ISO/IEC 14496-3:2005/Amd.2:2006, "Information technology - Coding of audio-visual objects - Part 3: Audio, Amendment 2: Audio Lossless Coding (ALS), new audio profiles and BSAC extensions," International Standards Organization, Geneva, Switzerland, 2006.
- [ 2 ] ISO/IEC 14496-3:2005/Amd.3:2006, "Information technology - Coding of audio-visual objects - Part 3: Audio, Amendment 3: Scalable Lossless Coding (SLS)," International Standards Organization, Geneva, Switzerland, 2006.
- [ 3 ] T. Liebchen, "MPEG-4 ALS - The Standard for Lossless Audio Coding," J. Acoust. Soc. Kor., Vol.28, No.7, pp. 618-629, 2009.
- [ 4 ] R. Geiger et al., "ISO/IEC MPEG-4 High-Definition Scalable Advanced Audio Coding," J. Audio Eng. Soc., Vol.55, No. 1/2, pp. 27-43, 2007.
- [ 5 ] ISO/IEC 23003-1, "Information technology - MPEG audio technologies - Part 1: MPEG Surround," International Standards Organization, Geneva, Switzerland, 2007.
- [ 6 ] ISO/IEC FDIS 23003-2:2010(E), "Information technology - MPEG audio technologies - Part 2: Spatial Audio Object Coding (SAOC)," Mar. 2010.
- [ 7 ] J. Breebaart et al., "Background, Concept,

- and Architecture for the Recent MPEG Surround Standard on Multichannel Audio Compression,” J. Audio Eng. Soc., vol. 55, no. 5, pp. 331-352, 2007.
- [8] J. Herre et al., “MPEG Surround - The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding,” J. Audio Eng. Soc., Vol.56, No.11, pp.932-955, 2008.
- [9] J. Engdegard et al., “Spatial Audio object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding,” presented at the 124th AES convention, paper no. 7377, 2008.
- [10] ISO/IEC JTC1/SC29/WG11, “WD6 of USAC,” N11213, Jan. 2010.
- [11] M. Neuendorf et al., “Unified Speech and Audio Coding Scheme for High Quality at Low Bitrates,” in Proc. IEEE ICASSP2009, pp.1-4, 2009.
- [12] FLAC - Free Lossless Audio Codec, <http://flac.sourceforge.net>.
- [13] T. Liebchen et al., “The MPEG-4 Audio Lossless Coding (ALS) Standard - Technology and Applications,” presented at the 119th AES convention, paper no. 6589, 2005.
- [14] T. Liebchen and Y. Reznik, “MPEG-4 ALS: An Emerging Standard for Lossless Audio Coding,” in Proc. Data Compression Conference, pp.439-448, 2004.
- [15] R. Geiger et al., “Audio Coding Based on Integer Transforms,” presented at the 111th AES Convention, paper no. 5471, 2001.
- [16] R. Geiger et al., “MPEG-4 SLS - Lossless and Near-Lossless Audio Coding Based on AAC,” in Proc. International Symposium on Communications, Control and Signal Processing, 2006.
- [17] H. Purnhagen, “Low Complexity Parametric Stereo Coding in MPEG-4,” in Proc. DAFX'04, pp. 163-178, 2004.
- [18] MP3 Surround, <http://www.mp3surround.com>.
- [19] Logic7, <http://www.lexicon.com/logic7/>.
- [20] ITU-R BS 1534-1, Method for the Subjective Assessment of Intermediate Quality Levels of Coding Systems, 2003.
- [21] G. Hotho, L. F. Villemoes, and J. Breebaart, “A Backward-Compatible Multichannel Audio Codec,” IEEE Trans. Audio, Speech, Lang. Process., Vol.16, No.1, pp.83-93, 2008.
- [22] 방희석, “다채널 오디오 코딩을 위한 MPEG Surround-1부: 기본 구조,” 한국음향학회지, 28권, 7호, pp.599-609, 2009.
- [23] 방희석, “다채널 오디오 코딩을 위한 MPEG Surround-2부: 다양한 모드 및 틀들,” 한국음향학회지, 28권, 7호, pp.610-617, 2009.
- [24] ISO/IEC JTC1/SC29/WG11, “Final Call for Proposals on Spatial Audio Object Coding,” N8853, 2007.
- [25] 정양원, 오현오, “오디오 객체 부호화 표준 - MPEG SAOC,” 한국음향학회지, 28권, 7호, pp. 630-639, 2009.
- [26] 서정일, 강경옥, “파라메트릭 다객체 오디오 부호화 기술,” 전자공학회지, 36권, 4호, pp. 47-55, 2009.
- [27] ISO/IEC JTC1/SC29/WG11, “Call for

Proposals on Unified Speech and Audio Coding,” N9519, 2007

- [28] 이태진, 강경옥, 김환우, “MPEG-D USAC: 통합 음성 오디오 부호화 기술,” 한국음향학회지, 28권, 7호, pp. 589-598, 2009.
- [29] 박호중, 박영철, “MPEG 오디오 표준 기술: USAC,” 방송공학회지, 14권, 2호, pp. 63-74, 2009.
- [30] M. Neuendorf et al., “A Novel Scheme for Low Bitrate Unified Speech and Audio Coding - MPEG RM0,” presented at the 126th AES convention, paper no. 7713, 2009.



방 희 석

- 1990년~1994년 서울대학교 전자공학과 학사
  - 1994년~1996년 서울대학교 전자공학과 석사
  - 1996년~2001년 서울대학교 전기컴퓨터공학부 박사
  - 2001년~2005년 LG전자 DM연구소 선임연구원
  - 2005년~2008년 LG전자 DM연구소 책임연구원
  - 2008년~현재 세종대학교 전자공학과 조교수
  - 관심분야 : 오디오 코딩, 오디오 신호처리, 음악 신호 처리
- 
-