

다채널 오디오 코딩을 위한 MPEG Surround-1부: 기본 구조

MPEG Surround for Multi-Channel Audio Coding-Part 1: Basic Structure

방 희 석*
(Hee-Suk Pang*)

*세종대학교 전자공학과
(접수일자: 2009년 8월 31일; 채택일자: 2009년 9월 17일)

최근 국제 표준으로 등록된 MPEG의 다채널 오디오 코딩 기술인 MPEG Surround에 대해 소개한다. 이 오디오 표준은 멀티채널 신호를 다운믹스하여 모노나 스테레오 신호를 생성하고 동시에 공간 파라미터를 추출하는 방법을 이용하여 부호화 과정을 수행한다. 부호화 과정에서는 다운믹스 신호와 공간 파라미터들을 사용하여 다채널 신호를 생성하는데, 다운믹스 신호는 AAC나 MP3 등 일반 오디오 코딩 방법에 의해 부호화하며 공간 파라미터들도 매우 낮은 정보량을 사용하기 때문에 전체적으로 매우 적은 비트율을 사용하여 고음질의 신호를 생성할 수 있다는 장점이 있다. 또한, 공간 파라미터들을 무시할 경우 휴대용 오디오 재생 기기에서 다운믹스 신호를 재생할 수 있으므로 기존 코딩 방법에 대해 역방향 호환성을 가진다. 본 논문의 1부에서는 MPEG Surround의 기본 구조에 대해 소개하고, 이어지는 2부에서는 헤드폰이나 이어폰을 이용하여 가상의 5.1 채널을 지원하는 바이노럴 모드를 포함해 다양한 모드와 툴들에 대해 소개한다. 여러 회사와 기관들에 의해 수행된 청취 평가의 결과도 함께 보이고 있다.

핵심용어: 다채널 오디오 코딩, MPEG Surround, 다운믹스, 공간 파라미터

투고분야: 뉴미디어 분야 (13.3)

An overview of the recently finalized multi-channel audio coding standard MPEG Surround is provided. This audio coding standard downmixes multi-channel signals to mono or stereo signals and, simultaneously, extracts spatial parameters for its encoding process. In its decoding process, it reconstructs multi-channel signals based on the downmix signals and spatial parameters. Since the downmix signals are coded in conventional audio coding format such as AAC and MP3 and the spatial parameters require a small amount of information, MPEG Surround guarantees high sound quality multi-channel audio at low bit rates. Besides, it is backward-compatible to conventional audio coding techniques because the downmix signals can be played on portable audio devices ignoring the spatial parameter information. In this paper, Part 1 presents an overview of the basic structure of MPEG Surround and Part 2 describes various modes and tools including the binaural mode which supports the virtual 5.1-channel playback via headphones or earphones. The listening test results by various companies and organizations are also presented.

Keywords: Multi-Channel Audio Coding, MPEG Surround, Downmix, Spatial Parameters

ASK subject classification: New Media (13.3)

I. 서론

스테레오 오디오가 소개된 지 약 반세기가 지난 현재 좀 더 자연스러운 오디오 재현을 위해서 다채널 오디오의

사용이 증가하고 있다. 초기에 다채널 오디오는 DVD와 같은 영화 관련 산업에서 사용이 시작되었으며, 최근에는 DVD-audio와 같은 오디오 전용 콘텐츠도 사용이 되고 있다. 다채널 오디오의 경우 2가지의 코딩 방법이 주로 사용되어 왔다. 첫째로 각각의 채널을 부호화 및 부호화하는 방법이다 [1]. 이 방법은 DVD나 DTV 등에서 사용되고 있으며, 음질은 좋지만 비트율이 매우 높다는 문제

책임저자: 방 희 석 (hspang@sejong.ac.kr)
134-747 서울시 광진구 군자동 세종대학교
전자정보공학대학 전자공학과
(전화: 02-3408-3727; 팩스: 02-3408-4329)

점이 있다. 예를 들어, 5.1채널의 경우 스테레오 채널에 비해 비트율이 약 2.5배 높게 된다. 두 번째 방법은 다채널 오디오를 모노나 스테레오로 다운믹스한 후 부호화하고 이를 복호화한 후 다시 다채널로 재생하는 것이다. 이 방법의 예로는 Dolby Prologic II [2]나 Logic7 [3] 등이 있는데, 이들의 경우 비트율은 기존의 스테레오 코딩 방법과 동일하지만 음질이 좋지 않다는 단점이 있다. 한편, 기존의 재생 매체나 방송 환경에서는 스테레오 포맷이 콘텐츠의 대부분을 차지하고 있으므로, 이를 다채널 포맷으로 변환할 경우 기존 사용자들의 기기에서 재생이 되지 않는 문제가 발생한다. 따라서, 새로운 다채널 오디오 코딩 기술은 낮은 비트율과 높은 음질, 기존의 스테레오 재생 시스템과의 역방향 호환성 (Backward-Compatibility) 등을 모두 지원하는 것이 요구된다.

MPEG에서는 스테레오 채널에 대해 비트율을 낮추면서도 음질을 보장하기 위해 M/S (Mid/Side) 기법이나 IS (Intensity Stereo) 기법을 사용하여 왔다 [1]. 이를 더욱 발전시킨 기술이 HE-AAC V2 (High-Efficiency Advanced Audio Coding Version 2)에서 사용하는 PS (Parametric Stereo) 기술인데, 이는 부호화부에서 스테레오 채널을 모노 채널로 다운믹스하고 동시에 스테레오 파라미터 (Stereo Parameter)를 추출한 후 복호화부에서 이들 모노 다운믹스 신호와 스테레오 파라미터를 이용하여 다시 스테레오 신호로 재생을 한다 [4]. 따라서, 이와 같은 기본 원리를 이용하면 다채널 오디오 또한 다운믹스 신호와 약간의 부가 정보로 부호화하고 이를 다시 복호화하여 다채널 신호를 생성할 수 있다. 이와 같은 원리를 이용하여 Faller와 Baumgarte는 BCC (Binaural Cue Coding)를 제안한 바 있다 [5][6].

이와 같이 새로운 다채널 오디오 코딩 기술에 대한 요구가 증가되면서 MPEG에서는 2003년에 SAC (Spatial Audio Coding)라는 가칭 아래 다양한 요구조건 및 이를

만족하는 기술에 대한 조사를 시작하였으며 [7][8], 이를 바탕으로 2004년에 CFP (Call for Proposal)을 공표하여 새로운 기술에 대한 제안을 받게 되었다 [9]. CFP에 응한 다양한 제안 기술들에 대해 청취 평가가 수행되어졌고 [10], 이 기술들 중 가장 우수한 2개 기술을 합쳐서 2005년에 RMO (Reference Model 0)이 완성되었다 [11]. RMO은 다채널 오디오 신호를 모노나 스테레오 신호로 다운믹스 하고 동시에 공간 파라미터 (Spatial Parameter)를 추출하여 부호화한 후, 복호화 시 이들을 이용해 다시 다채널 신호를 복원하는 것을 기본 구조로 하고 있다. 본 논문의 1부에서는 이와 같은 MPEG Surround의 기본 구조에 대해 소개한다. RMO 결정 이후에 추가적인 표준화 진행 과정을 통해 다양한 기술들이 추가 및 보완되었는데, 이들 중 스테레오 휴대 기기에서 가상 5.1채널을 재생하는 바이노럴 모드 (Binaural Mode), 공간 파라미터 없이 스테레오에서 5.1채널로 확장하는 Enhanced Matrix 모드, 다양한 채널 수에 대한 지원, 연산량을 낮추면서도 음질은 유사하게 유지하는 저전력 (Low Power) 버전 등을 대표적인 예로 들 수 있다. 본 논문의 2부에서 이와 같은 다양한 모드 및 툴들에 대해 소개한다. 한편, SAC는 표준화 과정 중에 기술명을 MPEG Surround로 개정하였으며, 2007년에 국제 표준으로 등록되었다 [12].

본 논문에서는 MPEG Surround 기술에 대한 대략적인 개요를 설명하고 있다. MPEG Surround에 대한 추가적인 개요는 [13][14]를, 구체적인 기술에 대한 심도 있는 리뷰는 [15-17]을 추천한다.

II. MPEG Surround의 기본 구조

MPEG Surround의 부호화 및 복호화 구조는 그림 1과 같다 [15]. MPEG Surround 부호화기는 다채널 오디오

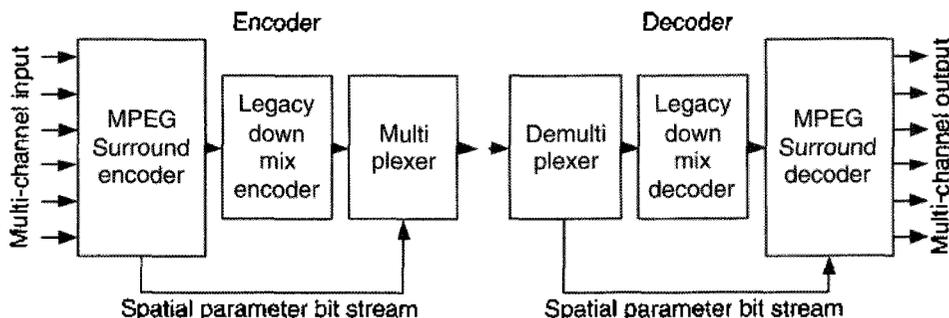


그림 1. MPEG Surround의 부호화와 복호화 과정
Fig. 1. MPEG Surround encoding and decoding processes.

신호를 다운믹스하고 동시에 공간 파라미터들을 추출한다. 다운믹스 신호는 MP3나 AAC와 같은 기존의 오디오 코딩 (Legacy Audio Coding) 방법에 의해 부호화되며, 공간 파라미터들도 부호화된다. 복호화부에서는 우선 다운믹스 신호가 복호화되고 공간 파라미터들과 함께 다채널 신호를 생성한다. MP3나 AAC 등의 오디오 코딩 방법들은 일반적으로 부가 정보를 비트열에 포함할 수 있는 방법을 제공하므로, MPEG Surround 비트열을 부가 정보에 포함되어 저장 혹은 전송할 수 있다. 따라서, MPEG Surround가 지원되는 오디오 재생 기기에서는 다채널 혹은 다운믹스 신호를 선택적으로 재생할 수 있고, 기존의 오디오 재생 기기에서는 부가 정보로 삽입된 MPEG Surround 정보를 무시하고 다운믹스 신호만을 재생할 수 있다. 다음에서는 MPEG Surround 표준 [12]과 Reference S/W [18]를 기준으로 하여 부호화부와 복호화부를 설명한다.

2.1. 부호화부

MPEG Surround 부호화기의 기본 구조는 그림 2에 나타나 있다 [15]. 우선 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) 과정을 거쳐서

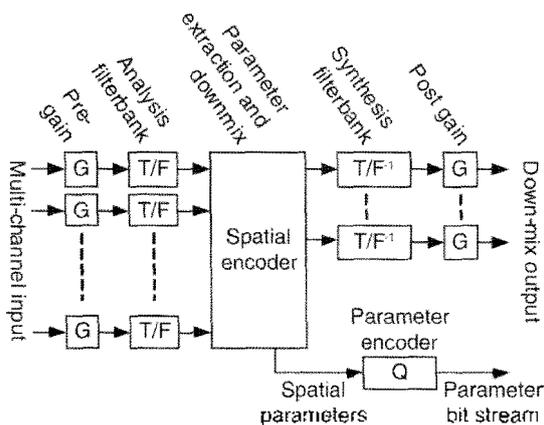


그림 2. MPEG Surround의 부호화부 구조
Fig. 2. MPEG Surround encoding structure.

시간 영역의 다운믹스 신호로 생성된다. 일반적으로 다운믹스 신호는 비트율을 낮추기 위해서 MP3나 AAC와 같은 오디오 코딩을 추가로 수행하게 된다. 다운믹스 과정은 일반적으로 클리핑 문제를 발생시킬 수 있기 때문에 다운믹스 신호의 레벨을 조정하기 위해서 게인 조정 (Gain Control)이 사용될 수 있다. 다음에서 각각의 부분을 설명한다.

1) QMF 및 Nyquist 필터 분석

MPEG Surround의 QMF 및 Nyquist 필터 분석은 HE-AAC V2에서 사용하는 Parametric Stereo 기술의 그것과 동일하며 그림 3에서 보이고 있다 [4]. 우선 QMF 분석은 32 kHz, 44.1 kHz, 48 kHz 등의 샘플링 주파수를 사용할 때 64개의 필터뱅크로 구성되며, 특정 샘플링 주파수에 대해 각 필터뱅크 대역폭은 동일하다. MPEG Surround의 한 프레임의 길이를 2048 샘플이라고 할 때, 이는 한 프레임이 64개의 QMF 서브밴드에 대해 각각 32개의 타임 슬롯 (Time Slot)을 가짐을 의미한다.

심리음향학에 따르면 인간의 청각 특성은 고주파보다는 저주파에서 더욱 높은 해상도를 요구한다. 이는 임계대역 (Critical Band)의 대역폭이 저주파에서는 거의 상수에 가깝다가 고주파로 가면서 주파수 대역폭과 비례함으로도 알 수 있다 [19]. 따라서, MPEG Surround에서는

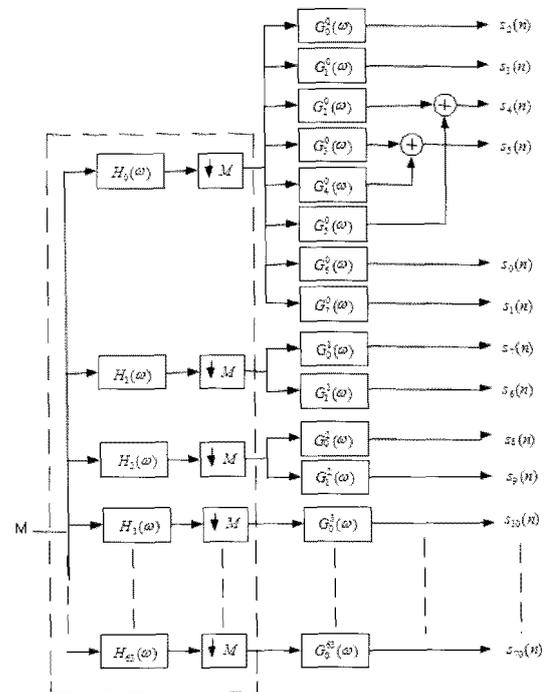


그림 3. QMF와 Nyquist 필터 분석
Fig. 3. QMF and Nyquist Filter Analysis.

64개의 QMF 출력 신호에 대해 Nyquist 필터 분석을 통해 저주파 성분의 주파수 해상도를 더욱 높이게 된다. 즉, 첫 번째 QMF 출력 신호는 6개의 주파수 대역으로, 두 번째와 세 번째 QMF 출력 신호는 각각 2개의 주파수 대역으로 다시 분할이 된다. 네 번째 이상의 QMF 출력 신호는 특별히 주파수 해상도를 더 높일 필요가 없기 때문에, 저주파 영역에 적용되는 필터에 의한 시간 지연을 맞추기 위해 시간 지연 필터 (Delay Filter)만 통과하게 된다. 최종적으로 71 주파수 대역 신호가 출력되며, 이들의 영역을 Hybrid 서브밴드 영역으로 정의한다.

MPEG Surround에서는 32 kHz보다 낮은 샘플링 주파수를 사용할 때는 32개의 QMF를 사용하며, 48 kHz보다 높은 샘플링 주파수를 사용할 때는 128개의 QMF를 사용한다. 이와 관련한 자세한 내용은 [12]를 참조하기 바라며, 이후로 특별한 언급이 없는 한 64개 대역의 QMF 구조를 기준으로 한다.

2) 다운믹스 신호 및 공간 파라미터 생성

MPEG Surround의 대표적인 트리 구조 (Tree Structure) 는 5-1-5, 5-2-5, 7-1-7, 7-2-7, 7-5-7 구조 등을 들 수 있다. 예를 들어, 5-2-5 구조는 부호화부에서 5.1 채널 신호가 스테레오 신호로 다운믹스되고 복호화부에서 다시 5.1 채널 신호로 생성되는 것을 의미한다. 그림 4에서 5-2-5 구조와 5-1-51 구조를 보이고 있는데, 여기에

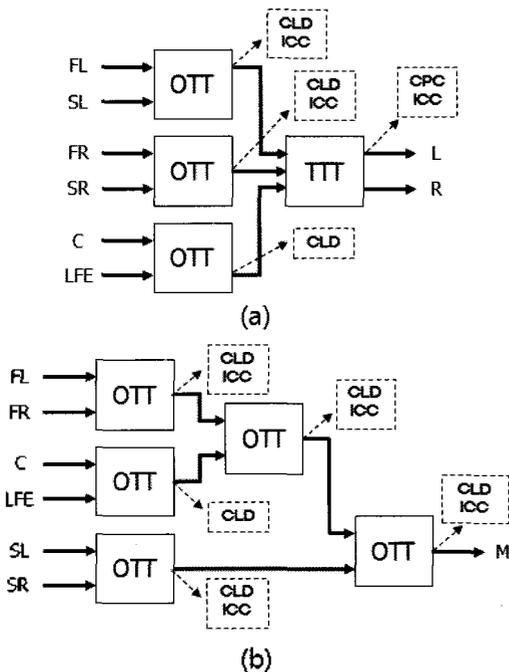


그림 4. MPEG Surround의 부호화부 구조
 (a) 5-1-2 구조 (b) 5-1-51 구조
 Fig. 4. MPEG Surround encoding structure.
 (a) 5-1-2 tree structure (b) 5-1-51 tree structure

서 FL, FR, SL, SR, C, LFE, L, R, M는 각각 Front Left, Front Right, Surround Left, Surround Right, Center, Low Frequency Enhancement, Left, Right, Mono 채널을 의미한다. 5-2-5 구조에서 우선 OTT (One-To-Two) 부호화부를 이용해 FL과 SL를, FR과 SR를, C와 LFE를 합해서 각각 하나의 다운믹스 신호를 생성하고 동시에 공간 파라미터를 추출한다. 또한, 이 세 신호는 TTT (Two-To-Three) 부호화부에 입력되어 L, R 다운믹스 신호로 출력되고 동시에 공간 파라미터가 추출된다. TTT 부호화부 및 복호화부의 경우 L, R, C 등 센터 채널이 들어간 3채널 신호에 대해 처리할 때 주로 사용된다. OTT 부호화부와 TTT 부호화부는 각각 R-OTT (Reverse OTT) 또는 TTO (Two-To-One), R-TTT (Reverse TTT) 또는 TTT (Three-To-Two)의 용어가 더 정확한 표현이지만, 복호화부와 연관성을 고려하여 그림에서는 OTT와 TTT라는 용어를 그대로 사용하였다.

5-2-5 구조와 달리 5-1-5 구조에서는 2가지의 다운믹스 방법이 존재한다. 그림 4(b)에서 보이는 것은 5-1-51 구조로써, FL과 FR, C와 LFE, RL과 SR로부터 각각 다운믹스 신호를 생성하고, 이들을 다시 순차적으로 다운믹스하는 방식이다. 또 다른 하나는 5-1-52 구조로써, FL과 SL, FR과 SR, C와 LFE를 각각 다운믹스한 후 이들을 다시 순차적으로 다운믹스하는 방식이다. 5-1-5 구조에서는 TTT 부호화부를 사용하지 않고 OTT 부호화부만 사용한다.

이상의 구조들 이외에도 MPEG Surround는 임의의 트리 데이터 (Arbitrary Tree Data)를 이용해 임의의 채널 수를 지원할 수 있다. 현재의 표준에서는 27개 출력 채널에 대해 정의하고 있으며, 5개의 준비 채널을 포함하여 총 32개의 채널을 지원할 수 있다 [12]. 이들 채널 중에는 수평뿐 아니라 수직 각도를 가지는 채널들도 정의되어 있다.

(1) 다운믹스 신호 생성

다양한 입력 다채널 신호 중 가장 많이 사용되는 5.1 채널 오디오 신호에 대해서 다운믹스 신호는 모노와 스테레오 중 하나를 선택할 수 있으며, 스테레오 다운믹스 신호는

$$\begin{aligned}
 x_L &= \frac{1}{a_d} (x_{FL} + \frac{1}{\sqrt{2}} x_c + \frac{1}{a_s} x_{SL} + \frac{1}{\sqrt{2} a_{LFE}} x_{LFE}), \\
 x_R &= \frac{1}{a_d} (x_{FR} + \frac{1}{\sqrt{2}} x_c + \frac{1}{a_s} x_{SR} + \frac{1}{\sqrt{2} a_{LFE}} x_{LFE})
 \end{aligned} \tag{1}$$

과 같이 생성되며, 모노 신호는

$$x_m = \frac{1}{a_d}(x_{FL} + x_{FR} + x_c + \frac{1}{a_s}x_{SL} + \frac{1}{a_s}x_{SR} + \frac{1}{a_{LFE}}x_{LFE}) \quad (2)$$

과 같이 생성된다. 이 중 a_s 와 a_{LFE} 는 서라운드 게인 (Surround Gain)과 LFE 게인 (LFE Gain)을 의미하고, 각각 $2^{k/4}$ ($k=0\sim4$)와 $10^{k/2}$ ($k=0\sim4$) 중 하나의 값을 선택할 수 있다 [12]. 또한, a_d 는 다운믹스 신호 전체의 게인을 조정하는 게인 (Downmix Gain)을 의미하며 $2^{k/4}$ ($k=0\sim6, 8$)의 값 중 하나를 가질 수 있다 [12].

위와 같은 다운믹스 과정은 Hybrid 서브밴드 영역에서 수행되는데, 이는 일반적으로 신호의 중첩이나 상쇄에 의해 신호의 왜곡을 초래할 수 있다. 따라서, 서브밴드 영역에서 이에 대한 왜곡을 처리하는 후처리 과정을 추가 할 경우 부호화기의 성능을 높일 수도 있다.

(2) 공간 파라미터 추출

MPEG Surround에서 기본적인 공간 파라미터는 CLD (Channel Level Difference), ICC (Inter-Channel Correlation), CPC (Channel Prediction Coefficient) 등 3가지를 사용한다. 이들의 추출은 Hybrid 서브밴드 영역에서 이루어지는데, Nyquist 필터의 출력 신호는 71개의 주파수 대역으로 나누어지기 때문에 모든 밴드별로 공간 파라미터를 추출할 경우 비트율이 높아지게 된다. 따라서, 28, 20, 14, 10, 7, 5, 4 중 하나의 값을 선택하고 이를 파라미터 밴드 (Parameter Band) 수로 사용한다 [12]. 즉, 71개의 주파수 대역이 아닌 파라미터 밴드 수만큼 공간 파라미터가 추출됨을 의미하며, 이에 의해 높은 음질이 필요 할 때는 높은 파라미터 밴드 값을 사용하고 혹은 낮은 비트율이 필요할 때는 낮은 파라미터 밴드 값을 사용하는 등 비트율에 따라서 성능을 조절할 수 있음을 의미한다.

공간 파라미터 중 CLD는 두 채널 사이의 레벨 차이를 의미하며, 부호화부에서 두 채널의 신호로부터 한 채널의 신호를 생성할 때 추출 및 복호화부에서 한 채널의 신호로부터 두 채널의 신호를 생성할 때 사용된다. Hybrid 서브밴드 신호 $X_{i,k}(m)$ 에 대해 채널, 시간, 서브밴드 인덱스를 각각 i, m, k 라 하고 b 번째 파라미터 밴드의 시작 서브밴드 인덱스를 k_b 라고 할 때, b 번째 파라미터 밴드의 CLD는 다음과 같이 정의된다.

$$CLD_b = 10 \log_{10} \left(\frac{\sigma_{x_{i,b}}^2}{\sigma_{x_{j,b}}^2} \right) \quad (3)$$

단, $\sigma_{x_{i,b}}$ 는 다음과 같이 정의되며, 여기에서 *는 켈레 복소수 (Complex Conjugate)를 의미한다.

$$\sigma_{x_{i,b}}^2 = \sum_m \sum_{k=k_b}^{k_{b+1}-1} X_{i,k}(m) X_{i,k}^*(m) \quad (4)$$

공간 파라미터 중 ICC는 두 채널 신호 사이의 상관관계 계수의 의미를 가지며, b 번째 파라미터 밴드의 ICC는 다음과 같이 정의된다.

$$ICC_b = Re \left\{ \frac{\sum_m \sum_{k=k_b}^{k_{b+1}-1} X_{i,k}(m) X_{j,k}^*(m)}{\sigma_{x_{i,b}} \sigma_{x_{j,b}}} \right\} \quad (5)$$

공간 파라미터 중 CPC는 다음과 같이 L, R 다운믹스 신호로부터 L, R, C 채널 신호를 생성하기 위해서 사용된다. CPC는 그 특성 상 CLD와 유사한 측면이 있으며 TTT 복호화부 및 부호화부에서만 사용된다. CPC 행렬 M_{CPC} 에 대한 자세한 설명은 [12]를 참조하기 바란다.

$$\begin{bmatrix} X_L \\ X_R \\ X_C \end{bmatrix} = M_{CPC} \begin{bmatrix} X_{L, Downmix} \\ X_{R, Downmix} \end{bmatrix} \quad (6)$$

이 외에도 부호화부에서는 공간 파라미터들이 적용될 시간 정보를 추출하여 공간 파라미터 세트와 함께 전송한다. 예를 들어, 안정적인 신호 (Stable Signal)의 경우 프레임 당 1 세트 이하의 공간 파라미터들과 이들이 적용될 시간 정보가 전송될 수 있다. 변동이 심한 신호 (Transient Signal)의 경우 프레임 당 최대 8개의 공간 파라미터 세트와 관련 시간 정보가 전송될 수 있다. MPEG Surround와 유사한 원리를 사용하는 PS 기술에서는 채널 신호 사이의 시간 차이 혹은 위상각 차이에 해당되는 정보들을 선택적으로 사용하였지만, MPEG Surround에서는 사용하지 않는다.

(3) 비트열 구성

MPEG Surround의 비트열은 전체 헤더와 프레임 데이터로 구성된다. 전체 헤더는 샘플링 주파수, 프레임 당 샘플 수, 파라미터 밴드 수, 트리 구조 등 한번 정의되면

변하지 않는 환경 설정 값들로 구성되어 있다. 이 데이터는 파일 콘텐츠 등 저장 매체에 대해서는 처음 부분에 한 번만 정의되면 되고, 방송 등 스트리밍 서비스 시에는 적절한 시간마다 한 번씩 비트열에 삽입되어야 복호화를 수행할 수 있다.

프레임 데이터는 매 프레임 변하는 정보로 구성되어 있으며, 공간 파라미터들이 적용될 Hybrid 서브밴드 영역의 시간 정보, OTT 및 TTT 데이터, 그 외 다양한 부가 정보 및 추가 톨 관련 데이터로 구성된다. 공간 파라미터들은 양자화되며 그 인덱스가 PCM 코딩, 시간 차동 코딩 (Time Differential Coding), 주파수 차동 코딩 (Frequency Differential Coding), 파일럿 코딩 (Pilot-Based Coding) 등 4가지 방법 중 하나로 코딩된다. 이 값들은 비트율을 최소화하기 위해서 허프만 코딩에 의해 변환된 후 비트열에 저장된다.

3) Nyquist 및 QMF 필터 합성

MPEG Surround의 Nyquist 필터 및 QMF 합성 또한 HE-AAC V2에서 사용하는 Parametric Stereo 기술의 그것과 동일하며 다음의 그림 5에서 보이고 있다 [4]. 우선 처음 6개의 Hybrid 서브밴드 신호를 합해서 첫 번째 QMF의 입력으로 사용하고, 다음으로 각각 2개의 Hybrid 서브밴드 신호를 합해서 두 번째와 세 번째의 QMF 입력으로

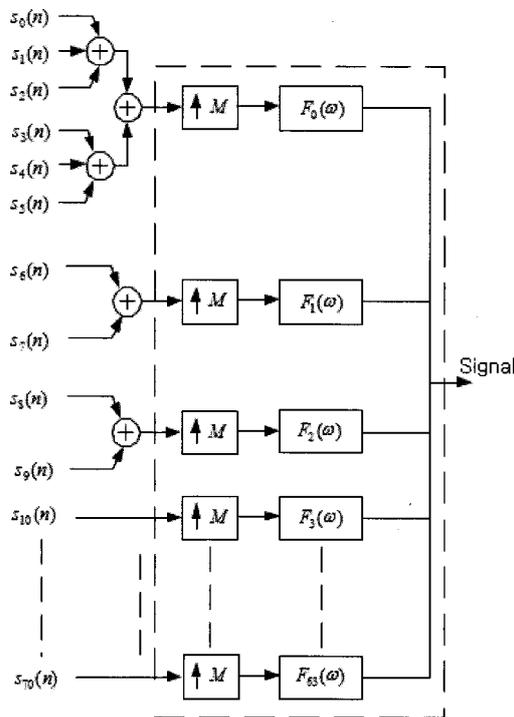


그림 5. Nyquist 필터와 QMF 합성
Fig. 5. Nyquist Filter and QMF Synthesis.

사용한다. 그 외의 Hybrid 서브밴드 신호는 바로 QMF의 입력으로 사용한다. 이 신호들에 대해 QMF 합성 과정을 거치면 최종 결과로 시간 영역의 다운믹스 신호가 합성된다. QMF 분석, Nyquist 필터 분석, Nyquist 필터 합성, QMF 합성으로 이루어진 전체 과정의 충격 응답 (Impulse Response)은 거의 임펄스 (Impulse)에 가까우며, 이는 전체 과정이 완전 복원 (Perfect Reconstruction)에 매우 가깝다는 것을 의미한다.

2.2. 복호화부

표준 복호화기의 기본 구조는 그림 6에 보이고 있다 [15]. 우선 MP3나 AAC 등에 의해 다운믹스 신호가 먼저 복호화되고, 이의 결과인 PCM 샘플들은 QMF 및 Nyquist 필터 분석을 순차적으로 수행하여 Hybrid 서브밴드 영역으로 변환된다. 예외적으로 다운믹스 코덱이 HE-AAC이고 32 kHz, 44.1 kHz, 48 kHz 중 하나의 샘플링 주파수를 사용할 경우 이는 MPEG Surround와 동일한 64개의 QMF 구조를 가지기 때문에 [4][12], 이 둘 사이는 QMF 서브밴드 영역에서 바로 연결이 될 수도 있고 시간 영역에서 연결이 될 수도 있다 [12]. Hybrid 서브밴드 영역의 다운믹스 신호는 복호화된 공간 파라미터가 적용되어 다채널 신호로 변환이 되고, 마지막으로 Nyquist 필터 및 QMF 합성을 거쳐 시간축의 다채널 신호가 출력된다. 다음에서 각각의 부분을 설명한다.

1) QMF 및 Nyquist 필터 분석 및 합성

복호화기의 QMF 및 Nyquist 필터 분석 및 합성 과정은 부호화기의 과정들과 동일하다.

2) 다채널 신호 합성

MPEG Surround의 대표적인 트리 구조에서는 부호화부에서 설명한 것과 같이 5.1 채널 혹은 7.1 채널 신호가

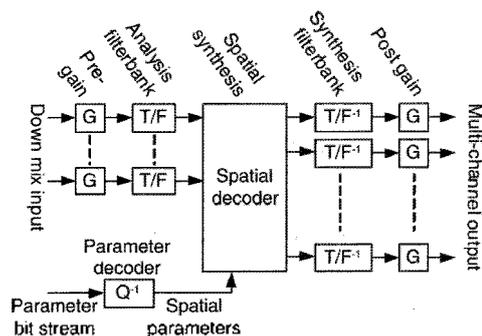


그림 6. MPEG Surround의 복호화부 구조
Fig. 6. MPEG Surround decoding structure.

출력된다. 그림 7에서 5-2-5 구조와 5-1-51 구조에 대한 복호화부 구조를 보이고 있는데, 전자의 경우는 TTT 복호화기가 사용된 후에 OTT 복호화기가 사용되고, 후자의 경우는 OTT 복호화기만 사용된다. 임의의 채널 수에 대해서는 대표적인 트리 구조로 복호화한 후에 비트열에 추가된 정보를 이용하여 후행렬 (Post-Matrix)을 구현하는 방식으로 복호화할 수 있다. 예를 들어, 스테레오 다운믹스 신호로부터 우선 5.1 채널 신호를 복호화하고 이에 다시 후행렬을 적용하여 채널 수를 다시 늘리는 과정을 들 수 있다. 다음에서 공간 파라미터가 다운믹스 신호에 적용되는 과정 및 각각의 기본적인 복호화 과정에 대해 설명한다.

(1) 공간 파라미터의 적용

MPEG Surround의 비트열 중 가장 기본적인 데이터는 공간 파라미터라고 할 수 있다. 이들 데이터는 전체 헤더에 포함된 파라미터 밴드의 수, 프레임 데이터에 포함된 프레임 당 공간 파라미터의 수, 공간 파라미터가 적용될 시간 정보, 공간 파라미터 데이터 등으로 구성이 된다. 이 데이터들은 그림 8과 같이 적용이 되는데, 즉 비트열에 Hybrid 서브밴드 영역의 Time Slot 단위의 시간 정보를 이용해 공간 파라미터를 적용하게 된다. 이와 같은 갱신 (Refresh) 지점 이외의 구간에 대해서는 기본적으로

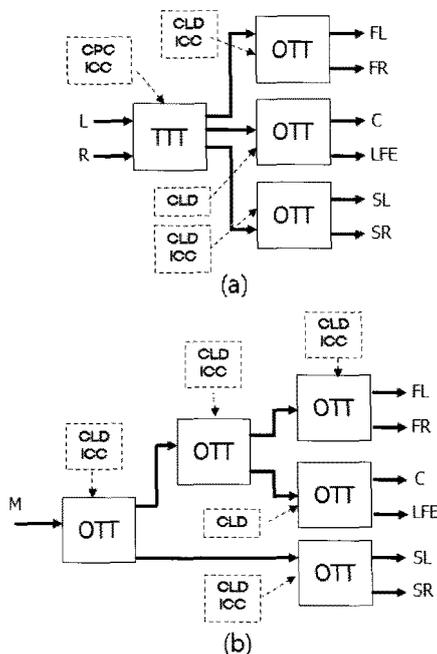


그림 7. MPEG Surround의 복호화부 구조
 (a) 5-1-2 구조 (b) 5-1-51 구조
 Fig. 7. MPEG Surround decoding structure.
 (a) 5-1-2 tree structure (b) 5-1-51 tree structure

선형 보간을 해서 서브밴드 샘플들에 적용될 공간 파라미터의 값을 정하게 된다. 공간 파라미터는 파라미터 밴드 단위로 적용이 되므로, 같은 파라미터 밴드 내의 Hybrid 서브밴드 샘플들은 같은 공간 파라미터 값이 적용된다.

(2) 행렬 처리

다운믹스 신호는 기본적으로 전행렬 (Pre-Matrix, M1), 믹스 행렬 (Mix-Matrix, M2), 후행렬 (Post-Matrix, M3) 등 3가지의 행렬 처리 과정을 거친다. M1은 다운믹스 신호를 역상관기 (Decorrelator)의 입력으로 어떻게 넣어야 할지를 정의하며, 따라서 다운믹스 신호의 채널 수와 역상관기의 수에 행렬의 크기가 결정된다. M1은 다시 세 부분으로 나눌 수 있으며, 첫째는 CLD나 CPC를 이용하여 개인 처리를 하는 부분이고, 둘째는 ADG (Arbitrary Down-mix Gain)이나 잔차 신호 코딩 (Residual Coding)을 적용하는 부분, 셋째는 행렬 호환성 (Matrix Compatibility)을 처리하는 부분이다. 이들 중 둘째 및 셋째와 관련된 내용은 이 논문의 2부에서 설명한다.

M1의 출력 신호는 역상관기를 통해 역상관 신호 (Decorrelated Signal)로 변환되는데, M2는 원신호 (Direct Signal)와 역상관 신호를 적절히 믹싱하는 역할을 수행한다. 이때, 비트열의 CLD, ICC 등의 값이 믹싱을 위한 정보로 사용된다. M3는 MPEG Surround의 대표적인 트리 구조에 의해 출력된 다채널 신호에 대해 추가적인 채널 확장을 하는 부분으로써, 임의의 트리 구조 부분의 복호화 과정을 수행한다.

(3) 역상관기

역상관기는 M1과 M2의 사이에서 다운믹스 신호에 적용되며, 일종의 서브밴드 영역의 잔향 필터로써 입력과 상관관계가 없도록 출력을 생성한다. 그림 9에서 Hybrid

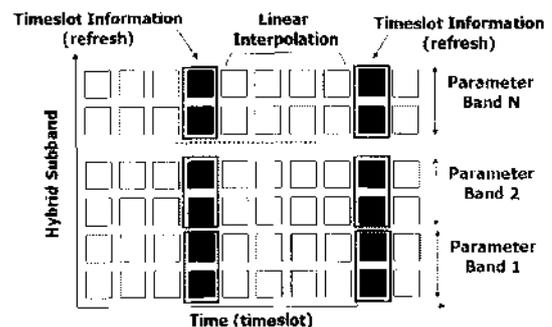


그림 8. 공간 파라미터의 적용
 Fig. 8. Application of spatial parameters.

서브밴드 영역에서 동작하는 역상관기를 보이는데 [15], 시간 지연, 격자 올패스 필터 (Lattice Allpass Filter), 에너지 보정부 등 3가지 부분으로 구성된다. 이들 파라미터는 역상관기들마다 다른 값으로 사용되며, 따라서 다수의 역상관기들이 사용될 경우라도 각각의 출력 신호들이 상호 무상관성 (Mutually Uncorrelated)을 가지도록 한다. 또한, 특정 신호에 대해 잔향 성분을 느끼지 않도록 에너지 보정부에서 모든 밴드에 대해 출력 신호의 레벨이 입력 신호의 레벨과 같아지도록 조정한다.

(4) 서브밴드 시간 처리

역상관기에서 출력된 역상관 신호는 원신호에 비해 시간 영역에서의 포락선 (Temporal Envelope)이 달라지기 때문에, 이를 원신호와 합쳤을 때 포락선의 왜곡을 초래하여 음질에 영향을 줄 수 있다. 이 현상은 변동이 많은 신호에 대해 더욱 심해지며, 따라서 신호의 고주파 부분이 큰 역할을 한다. 서브밴드 시간 처리 (Subband Domain Temporal Processing, STP)는 선택적으로 사용되어 역상관 신호의 시간 포락선이 원신호의 그것과 동일하게 되도록 역상관 신호를 수정하는 역할을 수행한다. 그림 10에서 STP의 구조를 보이고 있는데 [16], STP가 활성화 되면 역상관 신호의 고주파 성분은 원신호와 에너지 포락선의 비로 계산되는 성형 팩터 (Shaping Factor)에 의해 포락선이 수정된 후 최종적으로 원신호와 합쳐진다.

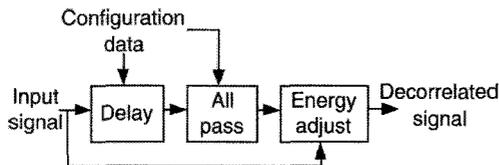


그림 9. 역상관기의 구조
Fig. 9. Structure of a decorrelator.

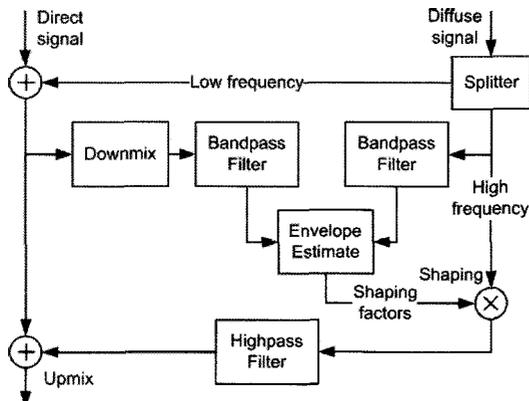


그림 10. STP의 구조
Fig. 10. Structure of STP.

III. 성능 평가

MPEG Surround의 성능 평가는 MPEG에서의 표준화 과정 중 다양한 참여 기관들의 청취 평가를 이용해 광범위하게 수행된 바 있다. 특히, 표준화 과정 중에는 표 1에서 보이고 있는 11개의 5.1 채널 테스트 아이템을 주로 이용하였으며, 이를 이용한 다양한 기고문들의 청취 평가 결과를 [20]에서 정리하여 공표하였다. 또한, 표준화가 거의 완료된 시점에서는 테스트 아이템들을 표 2에서 보이는 10개의 5.1 채널 아이템들로 모두 교체하고 MPEG Surround의 중요 기술들에 대해서 청취 평가를 수행하여 성능 평가 보고서를 공표하였다 [21]. 모든 테스트 아이템들은 44.1kHz의 샘플링 주파수를 사용하며, LFE 채널에 신호가 있는 경우 카테고리에서 따로 표시하였다. 본 논문의 1부에서는 MPEG Surround의 성능 평가 결과를 효과적으로 보이기 위해서 [15]와 [21]의 결과 중 일부를 이용하여 성능을 보인다. 이어지는 본 논문의 2부에서는

표 1. MPEG Surround 표준화를 위한 테스트 아이템
Table 1. Test items for MPEG Surround standardization.

이름	카테고리
BBC applause	Pathological/ambience
ARL applause	Pathological/ambience
Chostakovitch	Music
Fountain music	Pathological/ambience
Glock	Pathological
Indie2	Movie sound
Jackson1	Music
Pops	Music
Poulenc	Music
Rock concert	Music
Stomp	Music (with LFE)

표 2. MPEG Surround 성능 평가 보고서를 위한 테스트 아이템
Table 2. Test items for the MPEG Surround verification test.

이름	카테고리
Station_Atmo_6ch	Ambience
tennis	Ambience
tower1	Jazz
bonobo_1	Movies Drama
elliott2	Movies Drama
ravel1c	Orchestra
violin2	Orchestra
Lavilliers_14	Pop (Lavilliers, with LFE)
thalheim4	Pop
SantaCruz_09a	Pop (SantaCruz)

[20]과 [21]의 결과를 이용하여 다양한 모드 및 틀들에 대한 성능 평가 결과를 보인다. 1, 2부 모두에서 특별한 언급이 없을 경우 다운믹스 신호는 스테레오 신호를 사용하며, 2부의 바이노럴 모드를 제외하면 청취 평가는 기본적으로 5.1 채널 재생을 이용한다.

MPEG에서의 MPEG Surround 청취 평가 방법은 ITU-R의 표준인 MUSHRA 방법에 근거한다 [22]. MUSHRA에서는 은닉 원음 (Reference), 원음이 저역통과필터를 거친 앵커음 (Anchor), 함께 평가할 대상들을 상호 비교하여 점수를 주는 방식을 취하고 있으며, 이들은 매번 무작위 순서로 배열된다. 점수는 대상음들에 대해 각각 0점에서 100점까지 중 하나의 값으로 선택하며, 원음의 경우 비교 대상음들 사이에 은닉되는 것과는 별도로 따로 들을 수 있다. 비교 대상음들 사이에 은닉된 원음을 판별하지 못한 청취자의 결과는 후스크리닝 (Post-Screening) 과정을 통하여 제외할 수 있다. MPEG에서는 MUSHRA 테스트의 결과 분석을 위해 최소 8인 이상의 유효한 청취자 결과를 필요로 한다. 청취 평가 결과는 평균값과 함께 95% 신뢰구간을 보이며, 신뢰구간이 겹치지 않을 경우 유의미한 차이가 있다고 판단한다.

우선 MPEG Surround의 성능을 보이기 위해 [15]의 청취 평가 결과를 보인다. 표 3에서 비교 대상의 코덱들에 대해 보이고 있는데, AAC와 MPEG Surround를 사용한 것 (AAC+MPS), MP3와 MPEG Surround를 사용한 것 (MP3+MPS), Fraunhofer의 자체 기술인 MP3 Surround (MP3Surround) [23], AAC와 Dolby Prologic II를 사용한 것 (AAC+DPLII) [2], AAC로 다채널 코딩을 적용한 것 (AAC-MC) 등 5개를 사용하고 있다. 모든 코덱의 경우 다운믹스 코덱과 부가 정보의 비트율의 합이 160 kbps가 되도록 하였다. 그림 11과 부가 8명 그림 11과 부에 대해, 표 1의 테스트 아이템을 사용하여 위에서 설명한 5개의 비교 대상음과 은닉 원음 (Ref), 원음을 3.5kHz 저역통과 필터에 통과시킨 앵커음 (BW35)의 결과를 비교하여 보이고 있다 [15]. 결과에 의하면 MPEG Surround는 다운믹스

표 3. 그림 11을 위한 테스트용 코덱
Table 3. Test Codecs for Fig. 11.

코덱	전체 비트율 (kbps) (다운믹스코덱+부가정보)
AAC Stereo + MPS	160(128+32)
MP3 + MPS	160(149+11)
MP3 Surround	160(144+16)
AAC Stereo + DPLII	160(160+0)
AAC Multi-Channel	160(160+0)

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 32 kbps만을 사용하기 때문에 음질이 매우 낮게 나왔음을 알 수 있다.

다음으로 MPEG Surround 성능 평가 보고서 중 다운믹스 코덱으로 HE-AAC를 선택한 경우에 대한 청취 평가 결과를 보인다 [21]. 이 결과가 중요한 이유는 MPEG Surround의 경우 낮은 비트율에서 사용될 확률이 높는데, HE-AAC의 경우 낮은 비트율에서 고음질의 성능을 보장한다는 측면에서 중요한 다운믹스 코덱이기 때문이다. 표 4에서 테스트용 코덱을 보이고 있는데, HE-AAC와 MPEG Surround를 사용한 경우 (HE-AAC_MPS)와 HE-AAC를 다채널 코딩한 경우 (HE-AAC_MC)를 높은 비트율 (160 kbps)과 낮은 비트율 (64 kbps)에 대해 각각 사용하고 있다. 또한,

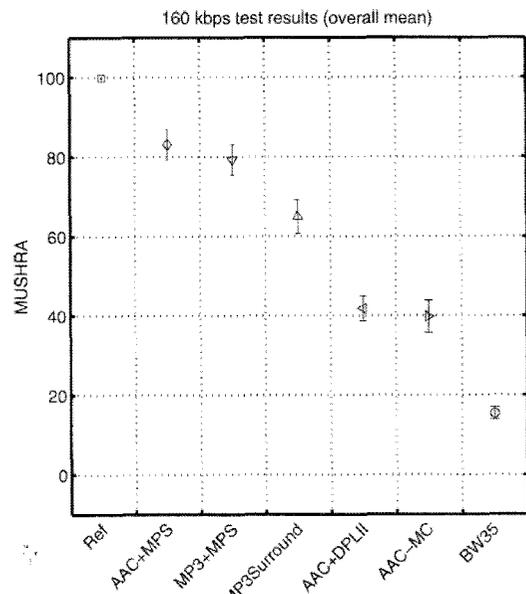


그림 11. MPEG Surround의 청취 평가 결과
Fig. 11. Listening test results of MPEG Surround.

표 4. 그림 12를 위한 테스트용 코덱
Table 4. Test Codecs for Fig. 12.

코덱	전체 비트율 (kbps)
HE-AAC_MPS_A	160
HE-AAC_MC_A	160
HE-AAC_MPS_B	64
HE-AAC_MC_B	64
L2_MPS	256
L2_DPL2	256

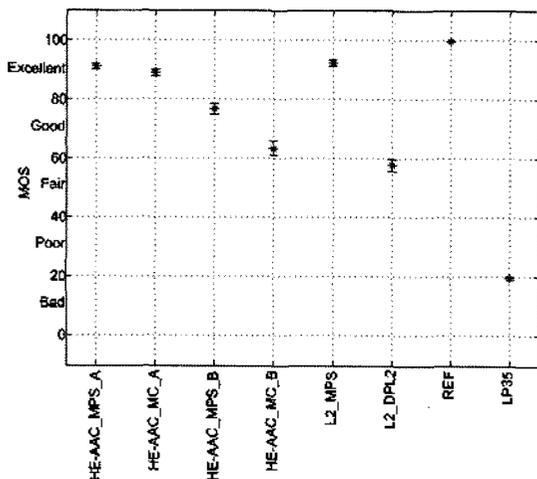


그림 12. HE-AAC + MPEG Surround의 청취 평가 결과
Fig. 12. Listening test results of HE-AAC + MPEG Surround.

다운믹스 코덱의 비교를 위해 MPEG Layer2와 MPEG Surround를 사용한 것 (L2_MPS), 다운믹스 코덱과 다채널 코덱을 함께 비교하기 위해 MPEG Layer 2와 Dolby Prologic II를 같이 사용한 것 (L2_DPL2)도 함께 사용하고 있다. 그림 12에서는 41인의 청취자에 대해 표 2의 테스트 아이템을 이용하여 앞에서 설명한 6개의 코덱의 결과를 은닉 원음 (REF) 및 앵커음 (LP35)의 결과와 함께 보이고 있다. 이에 따르면 160 kbps의 비트율로 HE-AAC와 MPEG Surround를 사용한 것은 동일 비트율로 HE-AAC 다채널 코딩한 것보다 약간 더 우수하며, 256 kbps의 비트율로 MPEG Layer2와 MPEG Surround를 사용한 것에 거의 근접한 성능을 보인다. 또한, 64 kbps의 비트율로 HE-AAC와 MPEG Surround를 사용한 것은 동일 비트율로 HE-AAC 다채널 코딩한 것보다 성능이 확연히 우수하다. 이상에서의 결과를 종합하면, MPEG Surround는 저비트율에서 다른 기술 대비 성능이 더욱 우수하다는 것을 알 수 있다. MPEG Layer2와 Dolby Prologic II를 사용한 것은 256 kbps의 높은 비트율에도 불구하고 성능이 좋지 않다.

IV. 결론

수년간의 MPEG Audio 서브그룹의 활동으로써, 낮은 비트율에서 높은 음질을 보장하며 기존의 스테레오 코딩 기술과 역방향 호환성을 가지는 다채널 오디오 코딩 기술인 MPEG Surround의 표준화가 완료되었다. 본 논문의 1부에서는 MPEG Surround의 부호화기 및 복호화기에 대한 기본 구조 및 성능을 소개하였다. MPEG Surround

는 다채널 신호를 생성하기 위한 부가 정보를 매우 적은 비트율을 사용하여 전송하며, 따라서 기존의 스테레오 신호와 거의 동일한 비트율을 사용하여 다채널 신호를 압축 및 복원할 수 있다. 예를 들어, HE-AAC를 다운믹스 코덱으로 사용할 경우 64 kbps의 비트율로 만족스러운 수준의 5.1 채널 오디오를 재생할 수 있다. MPEG Surround는 추후 DMB (Digital Multimedia Broadcasting)이나 DAB (Digital Audio Broadcasting) 등의 방송 시스템에서 유용한 다채널 오디오 기술로 사용될 것이 예상된다. 이어지는 본 논문의 2부에서는 MPEG Surround의 다양한 모드 및 볼륨과 이들의 성능에 대하여 소개한다.

감사의 글

본 논문은 지식경제부 IT 산업원천 사업의 지원을 받았습니다. MPEG Surround 표준화 과정에 같이 참여하였던 LG 전자 DM 연구소의 김동수, 임재현, 윤성용, 이현국 연구원과 DTV 연구소의 오현오, 정양원 연구원께 감사의 말씀을 드립니다.

참고 문헌

1. M. Bosi, "High-quality multichannel audio coding: trends and challenges," *J. Audio Eng. Soc.*, vol. 48, no. 6, pp. 588-595, 2000.
2. *Dolby Prologic II*, http://www.dolby.com/consumer/technology/prologic_ii.html, 2009.
3. *Logic 7*, <http://www.lexicon.com/logic7/>, 2009.
4. ISO/IEC 14496-3: 2005, *Information technology - Coding of audio-visual objects - Part 3: Audio*, 2005.
5. F. Baumgarte and C. Faller, "Binaural cue coding-part I: psychoacoustic fundamentals and design principles," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 509-519, 2003.
6. C. Faller and F. Baumgarte, "Binaural cue coding-part II: schemes and applications," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 520-531, 2003.
7. ISO/IEC JTC 1/SC 29/WG 11, *Investigation on Spatial Audio Coding*, N6023, Oct, 2003.
8. ISO/IEC JTC 1/SC 29/WG 11, *Call for Information on Spatial Audio Coding*, N6150, Dec, 2003.
9. ISO/IEC JTC 1/SC 29/WG 11, *Call for Proposals on Spatial Audio Coding*, N6455, Mar, 2004.
10. ISO/IEC JTC 1/SC 29/WG 11, *Report on Spatial Audio Coding RMO Selection Tests*, N6813, Oct, 2004.
11. ISO/IEC JTC 1/SC 29/WG 11, *Text of Working Draft for Spatial Audio Coding*, N7136, Apr, 2005.
12. ISO/IEC 23003-1, *Information technology - MPEG audio technologies - Part 1: MPEG Surround*, 2007.

13. S. Quackenbush and J. Herre, "MPEG Surround," *IEEE Multi-media*, vol. 12, no. 4, pp. 18–23, 2005.
14. J. Hilpert and S. Disch, "The MPEG Surround audio coding standard," *IEEE Signal Proc. Mag.*, vol. 26, no. 1, pp. 148–152, 2009.
15. J. Breebaart, G. Hotho, K. Koppens, E. Schuijers, W. Oomen, and S. van der Par, "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.
16. J. Herre, K. Kjørling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Roden, W. Oomen, K. Linzmeier, and K. S. Chong, "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.
17. 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.
18. ISO/IEC JTC 1/SC 29/WG 11, *ISO/IEC 23003-1:2007/FDAM 2, MPEG Surround Reference Software*, N9635, Jan. 2008.
19. E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer, Berlin, 1999.
20. ISO/IEC JTC 1/SC 29/WG 11, *Performance of MPEG Surround Technology*, N7950, Jan. 2006.
21. ISO/IEC JTC 1/SC 29/WG 11, *Report on MPEG Surround Verification Test*, N8851, Jan. 2007.
22. ITU-R BS 1534-1, *Method for the Subjective Assessment of Intermediate Quality Levels of Coding Systems*, 2003.
23. *MP3 Surround*, <http://www.mp3surround.com>, 2009.

저자 약력

• 방희석 (Hee-Suk Pang)



1994년 : 서울대학교 전자공학과 (학사)
 1996년 : 서울대학교 전자공학과 (석사)
 2001년 : 서울대학교 전기컴퓨터공학부 (박사)
 2001년~2008년 : LG전자 DM연구소 책임연구원
 2008년~현재 : 세종대학교 전자공학과 조교수
 ※주관심 분야 : 오디오/음성 코딩, 오디오 신호처리,
 음악 신호 분석