

## MPEG Surround 멀티채널 오디오 부호화 기술

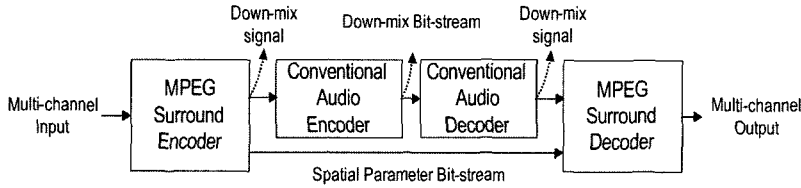
김광기, 한민수(한국정보통신대학교), 서정일, 백승권, 장대영(한국전자통신연구원)

### 요약

MPEG Surround는 멀티채널 오디오 신호를 모노 혹은 스테레오로 다운믹스한 신호와 멀티채널 신호가 다운믹스되는 과정에서 생성되는 공간 파라미터를 이용하여 효율적으로 부호화하는 기술이다. 기존의 오디오 코덱들이 심리음향모델에 기반한 인간의 청각특성을 이용하는 데 반해, MPEG Surround는 오디오 신호의 공간상의 이미지를 효과적으로 표현하는데 초점을 맞추고 있다. MPEG Surround의 다운믹스 신호는 AAC와 같은 기존의 오디오 압축기술을 이용하므로 역호환성을 제공하며, 멀티채널 복원을 위한 공간 파라미터는 채널간 크기 차이와 상관도 등으로 구성되므로 매우 낮은 비트율로 표현할 수 있다. 따라서 호환성과 고압축율을 가지는 MPEG Surround는 현재 급속하게 확대되고 있는 멀티채널 오디오 재생 환경과 다양한 멀티채널 콘텐츠에 대한 수요의 증가로 인해 멀티채널 방송 및 콘텐츠 서비스에 이용될 것으로 예상된다.

### 1. 서론

최근 들어 보다 향상된 멀티미디어 서비스를 위하여 멀티채널 오디오 처리 기술이 새롭게 요구되고 있으며 멀티채널 오디오는 기존의 오디오 부호화기를 이용하여 부호화 및 전송 시, 상당한 대역폭을 필요로 하기 때문에 이를 표현하는 방법에 있어서 새로운 전략의 오디오 부호화 기술이 요구된다. 멀티채널 오디오 부호화 기술에 대한 연구는 1990년대 중반부터 MPEG을 중심으로 이루어졌다. DVD 등의 멀티채널 소스의 폭발적인 증가와, 사용자들의 멀티채널에 대한 욕구증가와, 통신환경을 통한 서비스의 필요성 증가가 멀티채널 오디오 기술에 대한 연구를 더욱 부추기게 하였다. 이러한 이유로 MPEG-2 BC(backward compatible), MPEG-2/4 AAC(advanced audio coding) 등과 같은 멀티채널 오디오 부호화기술이 표준화되어 상업적으로 큰 실효를 거두었다. 아울러 최근에는 BCC(binaural cue coding)과 같은 혁신적인 공간오디오 부호화(spatial audio coding) 기법에 대한 연구가 활발히 진행되고 있는데<sup>1)</sup>, 이를 통해서 보다 적은 정보량을 이용하여 현실감 있는 멀티채



〈그림 1〉 일반적인 MPEG Surround 부호화 방법

널 오디오 데이터의 전달을 목적으로 하고 있다.

MPEG에서는 2002년도부터 MPEG Surround란 이름으로 공간 오디오 부호화 기술에 대한 표준화가 진행 중이다<sup>2)</sup>. MPEG Surround는 M개의 오디오 신호를 N개의(M>N) 오디오 신호와 사람이 음원의 위치를 판단하는 공간 파라미터(spatial parameter)로 구성되는 부가정보(side information)를 이용하여 나타내는 파라메트릭 멀티채널 오디오 압축 기술이다. MPEG Surround 부호화기는 멀티채널 오디오 신호를 모노 또는 스테레오 채널로 다운믹스(down-mix)한 후 기존의 MPEG-4 오디오 도구(MPEG-4 AAC, MPEG-4 HE-AAC 등)로 압축하고, 멀티채널 오디오 신호로부터 공간 파라미터를 추출하여 다운믹스 오디오 신호와 다중화한다. MPEG Surround 복호화기는 역다중화기를 이용하여 다운믹스 오디오 신호와 공간단서를 분리하고, 다운믹스 오디오 신호에 공간단서를 적용하여 멀티채널 오디오 신호를 합성한다.

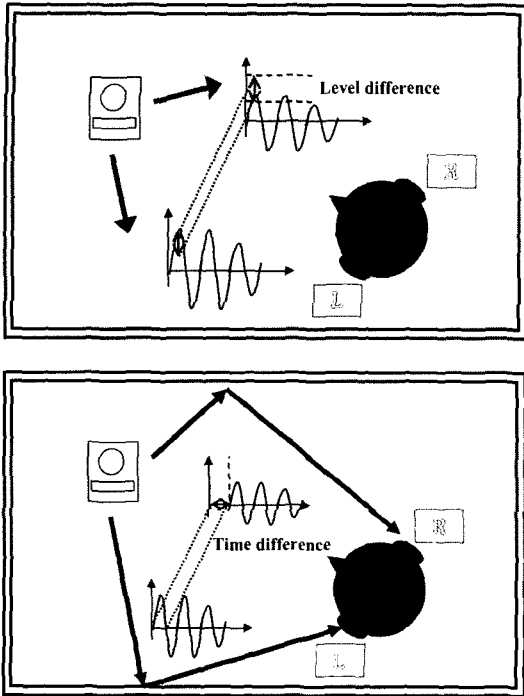
본 논문에서는 MPEG Surround의 부호화/복호화 과정과 이 과정에서 사용되는 다양한 오디오 코딩 기법에 대하여 소개하도록 하며, 본 논문의 구성은 다음과 같다. II장에서 MPEG Surround에서 사용하는 기본 개념에 대해서 설명하며, III장과 IV장에서는 MPEG Surround의 부호화기와 복호화기에 대해서 설명한다. V장에서는 현재까지 표준화가 완료된 MPEG Surround

RM0(reference model 0) 코덱의 성능을 살펴본 후 VI장에서 결론을 맺는다.

## II. MPEG Surround 기본 개념

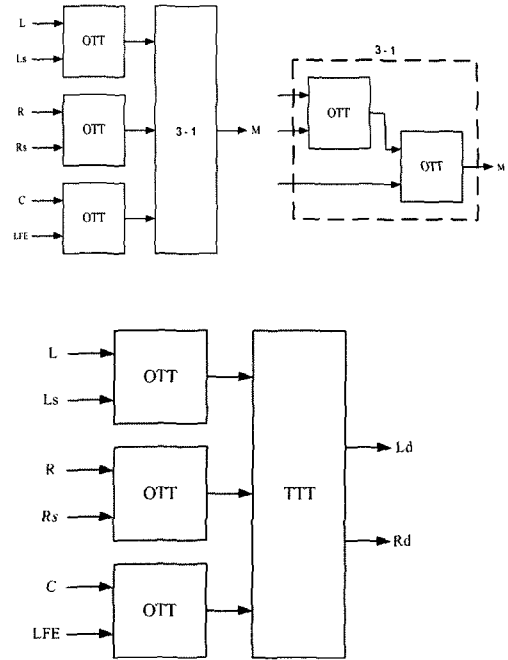
MPEG Surround는 멀티채널 오디오 신호를 모노 혹은 스테레오로 다운믹스한 신호와 멀티채널 신호가 다운믹스되는 과정에서 생성되는 공간 파라미터를 이용하여 효율적으로 부호화하는 기술이다. MPEG Surround는 일반적인 오디오 부호화 기술과 유사하게 인간의 청각적 특성에 기반을 하지만 접근 방식에 있어서는 차이가 있다. 즉, 멀티채널 오디오 신호가 재생될 때 오디오의 공간 이미지(spatial image)가 생성되므로 이를 표현하는 공간 파라미터를 이용하여 멀티채널 오디오 신호를 효과적으로 부호화하는 접근 방식을 따른다. 따라서 일반적인 오디오 부호화 기술이 심리음향 모델(psychoacoustic model)에 기반한 부호화 전략을 따르는 것과 달리 MPEG Surround는 다운믹스 신호와 공간 파라미터를 이용하여 공간상의 오디오 이미지를 효과적으로 표현하는 부호화 전략을 따른다. 그림 1은 일반적인 MPEG Surround 부호화 방법을 나타낸다.

일반적으로 오디오 신호의 공간상의 이미지는 신호가 자유공간 상에서 사람의 두 귀로 전달되는 경로의 차이로 인해 발생하는 두 귀에서의 신호 크기 차, ILD(interaural level difference)와 도



〈그림 2〉 두 귀 사이의 신호 크기 차와 신호 도달 시간 차

달 시간 차, ITD(interaural time difference)에 의해서 모델링 될 수 있다<sup>4)</sup>. 예를 들어 왼쪽 귀에 가까운 곳에서 신호가 발생한다고 하였을 때, 왼쪽 귀로 들어오는 신호 크기가 오른쪽 귀의 신호 크기보다 더 크고, 신호의 도달 시간도 왼쪽 귀가 더 빠르게 되어 신호의 위치가 왼쪽에 있다고 인지하게 된다. 그림 2는 두 귀사이의 ILD와 ITD를 나타낸다. 이러한 ILD와 ITD를 이용한 공간상의 오디오 이미지의 모델링과 달리 MPEG Surround에서는 채널간 신호 크기 차이인 CLD(channel level difference)와 채널간 상관도인 ICC(inter-channel correlation) 등과 같은 공간 파라미터를 이용하여 공간상의 오디오 이미지를 표현한다. 즉 ILD와 ITD가 사람의 양이 효과에 의한 공간상의 오디오 이미지의 표현이라면 공간 파라미터는 각각의 채널이 가지는 신호 자체에 의해서 공

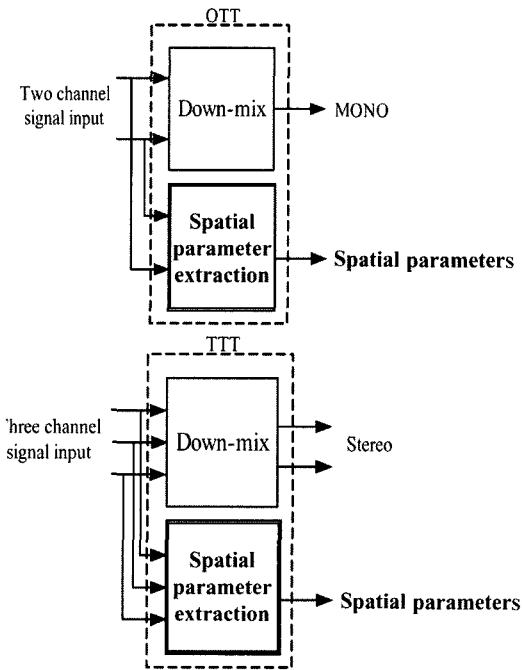


〈그림 3〉 파라메트릭 스테레오 코딩에 기반한 MPEG Surround 부호화기 구조

간상의 오디오 이미지를 표현한다는 차이가 있지만, 두 표현 모두 궁극적으로 공간상의 오디오 이미지를 표현한다는 점에서는 같다고 할 수 있다.

MPEG Surround에서는 그림 3에서 보는 바와 같이 MPEG-4 파라메트릭 스테레오 부호화(parametric stereo coding)에 기반한 구조를 이용하여 멀티채널 신호를 효과적으로 부호화하며<sup>4)</sup>, 다운믹스 신호수에 따라서 515 모드와 525 모드로 나뉘어진다.

즉 그림 3처럼 OTT(one to two)와 TTT(two to three)로 대표되는 다운믹스와 공간 파라미터 추출 모듈을 통해 멀티채널 신호를 모노 또는 스테레오 다운믹스 신호와 공간 파라미터로 표현을 하게 되고, 복호화 단계에서는 그림 3의 역의 방향으로 다운믹스 신호와 공간 파라미터를 이용한 합성과정을 거침으로써 멀티채널 신호를 복원하



〈그림 4〉 부호화기의 OTT와 TTT 모듈

게 되는 것이다<sup>5)</sup>. 그림 4는 MPEG Surround의 부호화기에서 사용되는 OTT와 TTT 모듈을 나타낸다. 그림 4에서 보는 바와 같이 OTT 모듈은 두 개의 입력 신호를 이용하여 모노 다운믹스 신호와 공간 파라미터를 추출하고 TTT 모듈은 세 개의 입력 신호를 이용하여 스테레오 다운믹스 신호와 공간 파라미터를 추출한다. 복호화기에서의 OTT와 TTT 모듈은 부호화기의 역으로 수행되며, 다운믹스 신호와 공간 파라미터를 이용하여 두 개 또는 세 개의 신호를 생성하게 된다.

### III. MPEG Surround Encoder

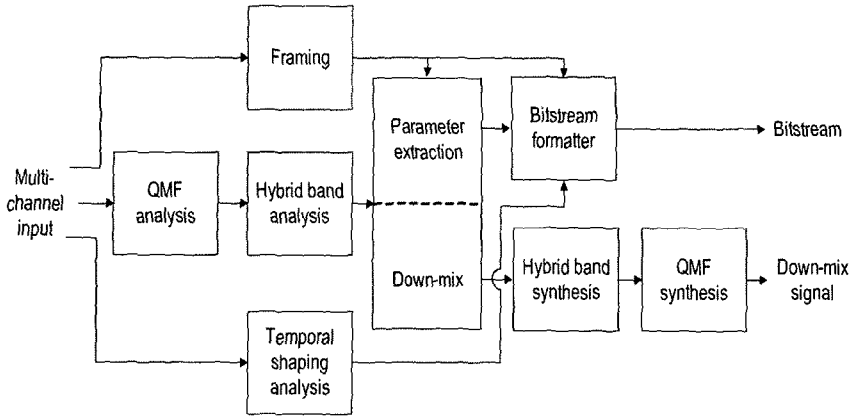
#### 1. 개요

MPEG Surround 인코더의 주요 기능은 멀티채널 입력 신호를 역호환성(backward compatibility)

를 가지는 모노 또는 스테레오 다운믹스 신호와 공간 파라미터를 이용하여 효율적으로 표현하는 것이다. 이러한 기능을 위해 MPEG Surround에서는 두 개의 신호를 하나로 묶어서 처리하는 OTT 모듈과 세 개의 신호를 두 개로 묶어서 처리하는 TTT 모듈을 사용한다. OTT와 TTT 모듈을 거치며 입력 신호는 모노 혹은 스테레오로 다운믹스되고, 다운믹스 신호를 다시 원래의 멀티채널 신호로 복원시키기 위한 CLD, ICC 등의 공간 파라미터가 추출된다. 또한 과도(transient) 성분에 따른 효율적인 부호화를 위해 과도 성분의 검출을 통해 추출되는 공간 파라미터 양을 조절하는 framing, 임펄스성 성분을 많이 포함하는 동적 신호의 음질 향상을 위해 TS 분석(temporal shaping analysis)를 수행한다. MPEG Surround 인코더는 앞에서 언급했던 바와 같이 멀티채널 신호를 묶는 방법에 따라서 스테레오 다운믹스 기반의 525 모드와 모노 다운믹스 기반의 515 모드가 있으며 일반적인 부호화기의 구조는 그림 5와 같다.

#### 2. Time to Frequency Transformation

MPEG Surround에서는 시간 영역의 신호를 주파수 영역의 서브 밴드로 분할하는 방법으로써 낮은 복잡도로 이를 수행하기 위하여 프레임 길이가 2048인 입력 신호에 대해서 이를 32개의 타임슬롯(time slot, 1 타임슬롯 = 64 샘플)으로 나눠서 버퍼링 과정을 통해서 하나의 타임슬롯씩 이동(shifting)하며 구성되는 640개의 버퍼 신호에 대해서 64 밴드 QMF(quadrature mirror filter)를 적용한다. 즉 하나의 프레임에 대해서 64개의 QMF 밴드 신호가 32개의 타임슬롯 개수만큼 생성된다. 이러한 QMF는 SBR(Spectral



(그림 5) MPEG Surround 부호화기

Band Replication)과 같은 통과도의 호환성도 제공함으로써 보다 효율적인 부호화를 유도할 수 있다. QMF를 거친 각 서브밴드는 Nyquist 필터뱅크를 이용하여 균등분할구조인 서브밴드로 나누고 이를 사람의 청각 시스템의 주파수 분해능과 유사하게 재구성하기 위하여 저주파 QMF 밴드를 세분화 하는 Hybrid 밴드 분석을 거치게 되며, MPEG Surround에서는 71개 또는 105개의 Hybrid 밴드를 사용하고 있다. 이렇게 주파수 서브 밴드로 분할된 신호는 공간 파라미터 추출을 위한 파라미터 밴드 신호로 변환(mapping)되는데 MPEG Surround에서는 Hybrid 밴드와 파라미터 밴드간의 변환표를 이용하며, 일반적으로 각각의 타임슬롯에서의 Hybrid 밴드를 저주파에서부터 고주파까지 여러 개씩 묶어서 더함으로써 28개의 파라미터 밴드 신호를 생성한다. 생성된 파라미터 밴드 신호는 위 과정의 역변환을 통해서 다시 시간 영역의 신호로 복원이 된다.

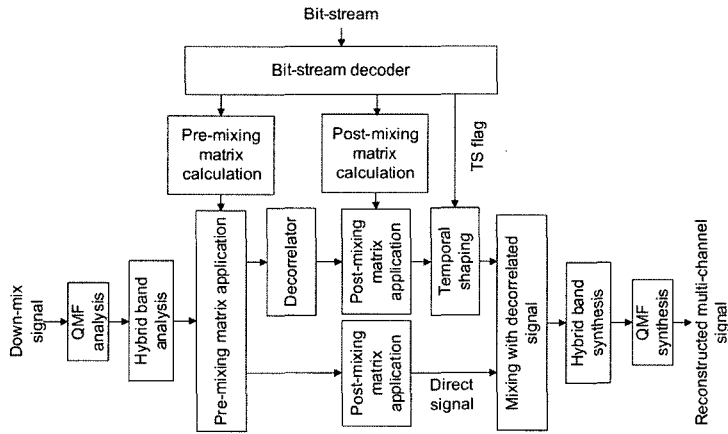
### 3. Framing

Framing은 과도 성분에 따른 효율적인 부호화

를 위하여 하나의 프레임 내에서 프레임의 분할을 다양하게 하는 방법이다. Framing이 사용되지 않는 경우에 하나의 프레임 전체에 대해서 공간 파라미터를 추출하는데 반해, framing을 사용하면 프레임 내에서 과도 성분이 없을 경우는 framing을 사용하지 않는 경우와 마찬가지로 프레임 전체에서 하나의 공간 파라미터를 추출하고, 과도 성분이 있을 경우 과도 성분을 기준으로 프레임을 2개의 서브 프레임으로 나눠서 각각의 서브 프레임에 대해서 공간 파라미터를 추출한다.

### 4. Temporal shaping analysis

MPEG Surround에서는 다운믹스 과정에서 손실되는 음향환경에 대한 보상을 위해 복호화기에서 디코리레이터(decorrelator)를 사용하고 있으며 일반적인 디코리레이터처럼 이전 신호를 지연시켜 현재의 신호에 더해주는 처리를 해준다. 이로 인해 박수소리와 같은 임펄스성 성분을 많이 가지고 있어 시간영역에서의 높은 해상도를 필요로 하는 신호는 지연된 신호에 의해 시



〈그림 6〉 MPEG Surround 복호화기

간영역에서의 해상도가 저하되어 복원되는 신호의 음질 열화가 심하다. 이를 해결하기 위해 복호화기에서 적용하는 것이 TS이며, TS는 부호화기에서 각각의 멀티채널 입력 신호 별로 생성되어 전송되는 TS 플래그(flag)에 따라서 복호화기에서 적용여부가 결정된다. 따라서 부호화기에서는 입력 신호에 대해서 TS 분석을 수행하며, TS 분석은 임펄스성 신호가 주파수 영역으로 변환 시 평탄하다(flat)는 특성을 이용하여 스펙트럼 평탄도 측정(spectral flatness measure, SFM)을 통해서 수행한다<sup>6)</sup>.

### 5. Parameter extraction

MPEG Surround에서 사용하는 공간 파라미터는 CLD, ICC, CPC(channel prediction coefficient)와 잔차 신호(residual signal)이 있으며 각각의 공간 파라미터는 부호화기의 모드에 따라서 선택적으로 추출된다. 공간 파라미터는 앞에서 설명한 28개의 파라미터 밴드에 대해서 추출하며, 크게 515 모드와 525 모드에 따라서 추출되는 파라미터에 차이가 있다. 먼저 515 모

드에서는 그림 3에서 보는 바와 같이 5개의 OTT 모듈을 사용하며, 각각의 모듈에서 CLD와 ICC를 추출한다. OTT 모듈에서 추출되는 파라미터를 1개로 보았을 때 총 5개의 CLD와 4개의 ICC가 추출된다. 여기에서 C(center)와 LFE(low frequency enhancement)에 해당하는 OTT 박스에서는 CLD만 추출하며, LFE가 저주파 대역의 신호만을 갖고 있으므로 앞에서부터 2개의 파라미터 밴드에 대해서만 CLD를 추출한다. 525 모드에서는 3개의 OTT 모듈과 1개의 TTT 모듈을 사용하며, 일반적으로 추출되는 파라미터는 OTT 모듈의 3개의 CLD와 2개의 ICC, TTT 모듈 에너지 모드의 CLD 2개가 있다. TTT 모듈의 예측(prediction) 모드에서는 2개의 CPC와 1개의 ICC를 추출하며, 이와 더불어 잔차 신호를 추출한다.

## IV. MPEG Surround Decoder

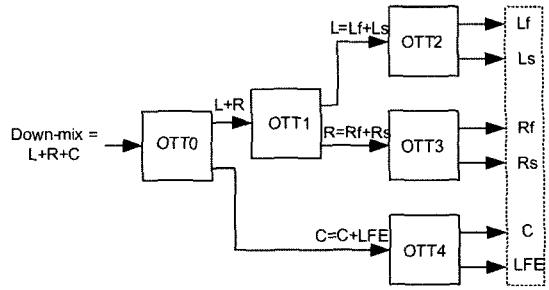
### 1. 개요

MPEG Surround 복호화기는 그림 6에서 보

는 바와 같다. 복호화기의 입력은 공간 파라미터를 포함하고 있는 비트스트림과 멀티채널 신호가 모노 혹은 스테레오로 다운믹스된 신호이다. 이러한 복호화기 입력은 MPEG Surround의 복호화 과정이 매트릭스 연산을 통해 이루어 지므로 매트릭스로의 변환 과정을 거친다. 즉 그림에서 보는 바와 같이 다운믹스 신호를 Hybrid 분석을 통해 시간 영역에서 주파수 영역으로 변환하고 전송된 비트스트림으로부터 읽어온 공간 파라미터를 이용하여 전 믹싱 매트릭스(pre-mixing matrix)와 후 믹싱 매트릭스(post-mixing matrix)를 생성하여 이들 사이의 매트릭스 연산을 수행한다. 이러한 매트릭스 연산을 통해 멀티채널 오디오 신호를 복원하게 되며, 디코리레이션(decorrelation)과 TS 등과 같은 부가적인 오디오 처리 기법을 이용하여 고음질의 출력 신호를 생성하게 된다.

## 2. Pre-mixing matrix

전 믹싱 매트릭스는 멀티채널 신호의 생성을 위한 전 처리 매트릭스로서 그림 6에서 보는 바와 같이 직접신호(direct signal)과 디코리레이터에 입력되는 신호의 이득(gain)을 나타내는 매트릭스이다<sup>6)</sup>. 전 믹싱 매트릭스의 요소 중 직접신호의 이득은 그림 3의 부호화기의 역구조를 갖는 복호화기를 고려했을 때, 다운믹스가 최초 입력되는 OTT 모듈 또는 TTT 모듈 자체가 갖는 이득은 1이므로 직접신호의 이득은 1이 된다. 또한 디코리레이터가 각각의 OTT 또는 TTT 모듈의 다운믹스 과정에서 소멸되는 음향환경을 다시 복원하기 위해서 적용되므로 OTT나 TTT 모듈에서 출력되는 다운믹스 신호의 이득과 최종 다운믹스 신호의 이득간의 비율이 디코리레



(그림 7) 515 모드의 전 믹싱 매트릭스의 요소

이터에 입력되는 신호의 이득이 된다. 그림 7은 515 모드에서의 전 믹싱 매트릭스의 요소들의 이득을 나타낸 것으로 실제 각각의 OTT 모듈에서의 직접신호와 디코리레이터에 입력되는 신호의 이득은 최초 입력되는 다운믹스 신호로 나누어 정규화된 값을 갖는다.

## 3. Post-mixing matrix

후 믹싱 매트릭스는 직접신호 또는 디코리레이터를 거친 신호가 멀티채널 신호로 확장되는 업믹싱(up-mixing) 과정을 나타내는 매트릭스이다<sup>7)</sup>. 515 모드를 예로 들어 설명하면 먼저 전 믹싱 매트릭스의 직접신호는 그림과 같이 다섯 개의 OTT 모듈을 거치며 각각의 모듈에서 CLD와 ICC를 이용하여 멀티채널 직접 신호를 생성하며, 다음 식은 Lf의 직접신호의 생성 예이다.

$$L_f = \frac{L+R}{L+R+C} \times \frac{L}{L+R} \times \frac{L_f}{L_f+L_s} \times \text{down-mix}(=L+R+C) \quad (1)$$

이 수식의 분수 각각은 Lf가 생성되기 위해 거치는 OTT 모듈의 CLD와 ICC를 이용해서 계산되는 이득을 나타내며 그림 7에서 보는 바와 같이 Lf는 OTT0, OTT1, OTT2에 의해서 생성된다. 다음으로 디코리레이터를 거친 신호가 멀

터채널 신호로 확장되는 것을 OTT1의 디코리레이터를 예로 들어 설명하면 OTT1에서 소멸된 음향환경은 Lf, Ls, Rf, Rs의 원신호가 가지고 있던 것이다. 따라서 OTT1에 적용되는 디코리레이터를 거친 신호는 OTT2와 OTT3의 입력으로 들어가게 되고, 각각의 OTT 모듈을 통해 디코리레이션된 Lf, Ls, Rf, Rs신호를 생성하게 된다. 여기에서 OTT1의 디코리레이터를 거친 신호는 OTT2와 OTT3에 적용되는 디코리레이터를 다시 거치지 않는다. 즉 각각의 OTT 모듈에서 소멸되는 음향환경은 각각의 OTT 모듈의 디코리레이터를 통해 한번만 보상을 해주게 된다. 따라서 최종적으로 생성되는 신호는 각 채널별로 직접신호와 직접 신호를 생성하기 위해서 거치는 OTT 모듈 각각의 디코리레이터로부터 생성되는 디코리레이션된 신호들의 합으로 이루어진다. Lf를 예로 들면 OTT0, OTT1, OTT2를 거치며 생성되는 직접신호, OTT0의 디코리레이터를 거친 신호가 OTT1과 OTT2를 거치고 나서 생성된 디코리레이션된 신호, OTT1의 디코리레이터를 거친 신호가 OTT2를 거치고 나서 생성되는 디코리레이션된 신호, OTT2의 디코리레이터를 거쳐서 생성되는 디코리레이션된 신호의 합으로 최종 출력되는 Lf 신호가 생성된다.

#### 4. Decorrelator

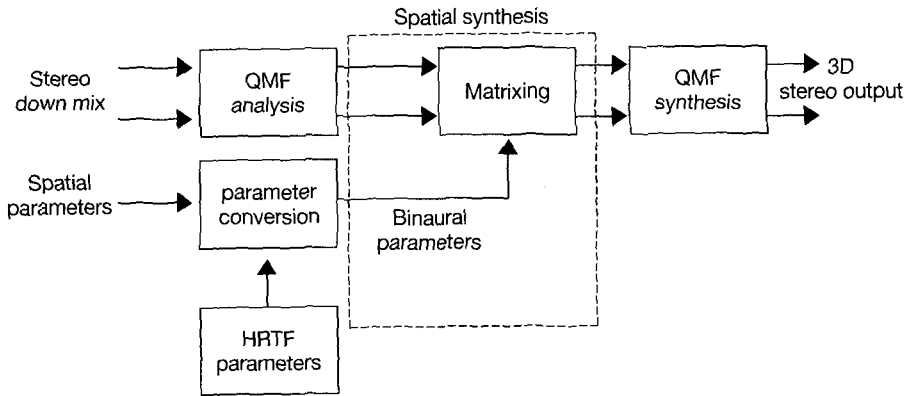
Decorrelator는 멀티채널 신호가 다운믹스되는 과정에서 잃게 되는 음향 정보에 대해서 보상을 해주는 기능을 갖으며, 음향 환경에 의해서 생성되는 신호는 원신호가 음향환경의 여러 경로를 거쳐 지연되어 들어오는 것이라는 개념을 이용하여 현재의 신호를 생성할 때 이전 신호의 지연된

신호를 더해줌으로써 손실된 음향환경을 복원해주는 것이다<sup>7)</sup>. MPEG Surround에서 사용하는 디코리레이터는 IIR(infinite impulse response) 필터를 이용하여 구현된다. Hybrid 밴드를 4개의 주파수 대역으로 나눠서 저주파 대역은 지연시간을 길게 주고, 고주파 대역은 지연시간을 짧게 주어 주파수 대역 별 특성을 반영하며, 또한 각각의 Hybrid 밴드마다 프랙셔널 지연(fractional delay)를 주어 지연시간이 일정하지 않고 계속 변화하도록 하여 디코리레이터의 성능을 향상 시킨다<sup>8)</sup>. MPEG Surround는 515 모드나 525 모드에 따라서 다운믹스과정을 갖는 OTT와 TTT 모듈의 수가 다르기 때문에 사용되는 디코리레이터의 개수가 다르다. 이런 이유로 디코리레이터 10개를 미리 생성하게 되고, 실제 복호화 과정에서 필요한 개수만큼 디코리레이터를 사용한다. 즉 각각의 OTT나 TTT 모듈에서 사용되는 디코리레이터는 모두 다른 것들을 사용한다.

#### 5. Temporal shaping

TS는 박수소리와 같이 임펄스 성분을 많이 갖는 신호를 복원할 때 디코리레이션된 신호에 적용하며, 디코리레이션된 신호의 고주파 대역의 포락선이 직접 신호의 포락선을 따라가도록 하여 임펄스 성분으로 인한 음질의 열화를 최소화하도록 한다<sup>9)</sup>. TS를 적용하는 주파수 밴드는 13~71까지의 Hybrid 밴드이며, 12 이하의 밴드에 대해서는 TS를 적용하지 않고 디코리레이션된 신호를 그대로 직접 신호에 더해준다. 즉 TS는 고주파 영역의 디코리레이션된 신호가 시간 영역으로 변환된 신호의 에너지와 직접 신호의 시간 영역에서의 에너지 간의 비율을 고려하여 디코리레이션된 신호의 포락선이 직접 신호의





〈그림 8〉 MPEG Surround 바이노럴 복호화기

포락선을 따라가도록 조절하는 것이다. TS의 결과 신호는 나머지 신호와 더하여 최종 멀티채널 출력신호를 생성하게 된다.

운믹스 신호와의 매트릭스 연산을 수행하여 바이노럴 스테레오 신호를 출력하게 되는 것이다.

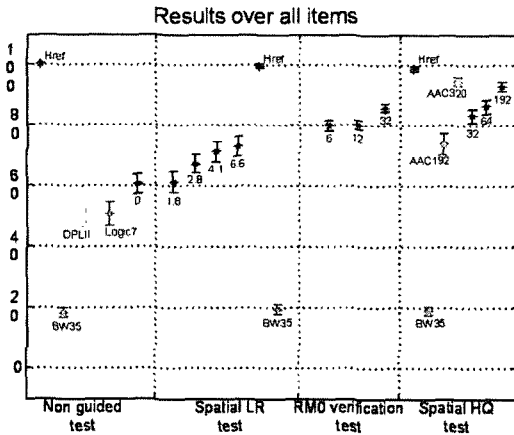
## V. MPEG Surround 성능

### 6. MPEG Surround binaural decoder

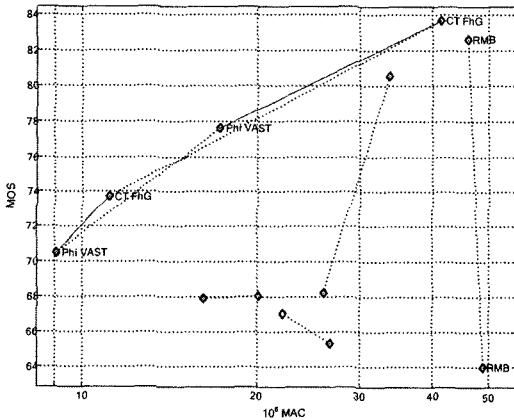
MPEG Surround에서는 스테레오 신호에 머리전달함수를 적용하여 멀티채널 오디오 신호의 효과를 느낄 수 있도록 하는 기능을 갖는 바이노럴 복호화기를 제공하고 있다. 즉 핸드폰 등과 같이 스테레오 오디오 출력만을 가진 단말이나 헤드셋을 이용하는 환경에서는 멀티채널 효과를 느낄 수 없으므로 이를 위하여 머리전달함수를 이용한 멀티채널 효과를 부여하는 기능을 갖는 바이노럴 복호화기를 제안하고 이를 MPEG Surround에 포함시켰다<sup>10)</sup>. 그림 8은 스테레오 다운믹스 신호를 이용하여 바이노럴 스테레오 신호를 출력하는 과정을 나타낸다. 그림에서 보는 바와 같이 전송된 공간 파라미터를 사전에 모델링을 통해 구해 놓은 머리전달 함수 파라미터를 이용하여 바이노럴 파라미터로 변환을 수행한후 QMF 밴드로 변환된 스테레오 다

MPEG 오디오 서브 그룹에서는 MPEG Surround의 성능 평가를 위하여 ITU-R에서 권고하는 MUSHRA(Multi-Stimulus test with Hidden Reference and Anchor)<sup>10)</sup> 방법을 이용하였다. 이는 중간 정도의 음질을 가지는 오디오 시스템의 성능을 평가하기 위하여 고안된 평가 방법으로써 숨겨진 기준 신호(hidden reference)와 저역 필터로 음질을 열화 시킨 앵커(anchor) 신호들을 포함한 실험 시스템을 구성하여 주관적으로 음질을 평가하는 방법이다.

그림 9는 MPEG Surround의 전반적인 음질 평가 결과이다<sup>10)</sup>. 그림에서 보듯이, 행렬화 기반의 멀티채널 재생방법인 Dolby Prologic II와 Logic7과의 비교 및 다양한 비트율의 AAC 코덱과의 비교를 통하여 MPEG Surround의 음질 성능을 살펴볼 수 있다. 여기서 MPEG Surround의 다운믹스 신호에 대한 비트율은 128 kbps이



〈그림 9〉 MPEG Surround 코덱의 성능



〈그림 10〉 MPEG Surround 바이노럴 디코더 성능 (quality)과 복잡도(complexity) 비교

며 그래프 하단에 있는 숫자는 부가정보의 비트율을 나타낸다. 기본적으로 MPEG Surround는 기존 행렬화 기반의 코덱보다 고음질을 제공할 수 있음을 확인할 수 있다. 또한 MPEG Surround의 부가 정보의 비트율에 따라 저 비트율(LR; Low-rate) 모드, 기본 모드(RM0 verification), 고효율(HQ; High-quality) 모드 등으로 나뉘어 질 수 있으며, 평가 결과, 부가정보의 비트율이 증가함에 따라 점진적인 음질향상이 관측되었다. 즉, MPEG Surround는 기존 저 비트율 코덱의

성능으로부터 끊김 없는(seamless) 점진적 음질 향상을 기대할 수 있으며 비트율이 증가할수록 원음과 가까운 손실 없는(lossless) 음질을 제공할 수 있으며 이를 기반으로 확장성(scalability)을 제공할 수 있다.

그림 10은 MPEG Surround 바이노럴 디코더 RM0(reference model 0)로 선정된 CT/FHG 모델과 Phi/VAST 모델의 음질(quality)로 표현된 음질과 복잡도(complexity)를 나타낸 그림이다<sup>[11]</sup>. 그림에서와 같이 CT/FHG 모델이 가장 우수한 음질을 제공하는 반면 Phi/VAST 모델은 낮은 복잡도를 제공하면서도 적절한 음질을 제공하고 있다. 그림에서 같은 모델이면서 우수한 음질을 제공하는 것은 스테레오 다운믹스 신호일 경우이며, RMB는 MPEG Surround RM0를 이용했을 경우를 나타낸다.

## VI. 결 론

MPEG에서 표준화가 진행되어 현재 거의 완료된 MPEG Surround는 멀티채널 오디오 신호를 효율적으로 압축하면서 기존의 스테레오 오디오 부호화기와의 호환성을 갖는 유용한 압축 도구이다. MPEG Surround는 오디오 신호의 공간상의 이미지를 효과적으로 표현하는데 초점을 맞추고 있으며, 이를 위하여 CLD와 ICC같은 공간 파라미터를 이용한다. MPEG Surround는 높은 압축율과 호환성을 제공하므로 DAB와 DMB등과 같은 스테레오 오디오 방송 시스템을 멀티채널 오디오 방송 시스템으로의 확장이 가능하며, 최근에 기능이 추가된 바이노럴 부호화기를 이용하여 스테레오 재생환경에서도 가상의 멀티채널 환경과 같은 느낌을 갖도록 해줄 수 있다. 이러한 MPEG Surround는 현재 급속하게

확대되고 있는 멀티채널 오디오 재생 환경과 다양한 멀티채널 콘텐츠에 대한 수요의 증가로 인해 멀티채널 방송 및 콘텐츠 서비스에 이용될 것으로 예상된다.

### 참고문헌

[1] C. Faller and F. Baumgarte, "Binaural Cue Coding: A Novel and Efficient Representation of Spatial Audio," in Proc. ICASSP 2002, Orlando, FL, May 2002.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N6455, "Call for Proposals on Spatial Audio Coding," Munich, Mar. 2004.F.

[3] C. Faller, "Parametric coding of spatial audio," Proc. 7th Int. Conf. on Digital Audio Effects, Naples, Italy, Oct. 2004, pp. 151-156

[4] ISO/IEC JTC1/SC29/WG11 (MPEG), "CT/Philips Contribution to CFP on Spatial Audio Coding," Document M11001, Redmond, July 2004.

[5] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N7947, "ISO/IEC 23003-1: 2006/FCD, MPEG Surround," Bangkok, Jan. 2006.

[6] J. Herre, "Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)," 101st AES Convention, Los Angeles, Nov 1996, Preprint 4384.

[7] Jonas Engdegard, Heiko. Purnhagen, Jonas Roden, Lars Liljeryd, "Synthetic Ambience in Parametric Stereo Coding," 116th AES Convention, Berlin 2004.

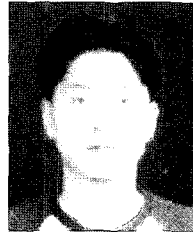
[8] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N8177, "Study on Text of ISO/IEC 23003-1, MPEG Surround," Montreux, April. 2006.

[9] ITU-R Recommendation BS. 1534-1, "Method for the subjective assessment of intermediate sound quality (MUSHRA)," International Telecommunications Union, Geneva, Switzerland, 2001.

[10] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N7701, "Performance of MPEG Surround Technology," Nice, 2005.

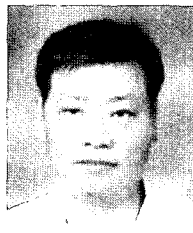
[11] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N8178, "MPEG Surround Binaural/3D-Stereo test Report," Montreux, April, 2006.

### 저자소개



김 광 기

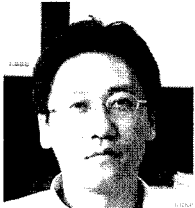
2002년 한국항공대학교 전자공학과 졸업(학사)  
 2004년 한국정보통신대학교 공학부 졸업(석사)  
 2006년-현 재 한국정보통신대학교 공학부 박사  
 과정 재학  
 주관심분야 speech and audio coding



한 민 수

1979년 서울대학교 전기공학과 졸업(학사)  
 1981년 서울대학교 전기공학과 졸업(석사)  
 1989년 University of Florida 졸업(박사)  
 1982년-1985년 한국표준과학연구원 연구원  
 1990년-1997년 한국전자통신연구원 책임연구원  
 1998년-현 재 한국정보통신대학교 교수  
 2004년-현 재 한국정보통신대학교 디지털미디어  
 연구소 소장  
 주관심분야 speech and audio coding

저자소개



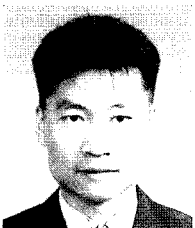
서 정 일

1994년 경북대학교 전자공학과 졸업(학사)  
 1996년 경북대학교 전자공학과 졸업(석사)  
 2005년 경북대학교 전자공학과 졸업(박사)  
 1998년-2000년 LG 반도체 주임연구원  
 2000년-현 재 한국전자통신연구원 선임연구원  
 주관심분야 음향신호처리, 음성신호처리, 객체기반 3차원 오디오 신호처리, MPEG, DMB 등



백 승 권

1999년 한국항공대학교 항공전자공학과 졸업(학사)  
 2001년 한국정보통신대학교 전자공학과 졸업(석사)  
 2005년 한국정보통신대학교 전자공학과 졸업(박사)  
 2005년-현 재 한국전자통신연구원 선임연구원  
 주관심분야 오디오 신호처리, 음원분리, 객체기반 3차원 오디오 신호처리, MPEG, DMB 등



장 대 영

1999년 부경대학교 전자공학과 졸업(학사)  
 2000년 배재대학교 컴퓨터공학과 졸업(석사)  
 1991년-현 재 한국전자통신연구원 책임연구원  
 주관심분야 오디오 신호처리, 음원분리, 객체기반 3차원 오디오 신호처리, MPEG, DMB 등

용 어 해 설

택소노미

Taxonomy, 分類學 [관리운동]

가나다, ... ABC, ...와 같이 표준화되고 체계적으로 분류된 전통적인 분류학(taxonomy) 기반의 분류체계.

트리형의 위계적 구조로서 이미 결정된 체계를 가지고 있다는 특징이 있다. 그리스어로 '분류하다'라는 'tassein'과 '법, 과학'이라는 'nomos'의 합성어로 사람들에 의해 이해되는 관계를 기준으로 분류되는 폭소노미(folksonomy)에 대비되는 용어이다.

적재 가능 커널 모듈

Loadable Kernel Module, LKM [컴퓨터]

커널 재컴파일 없이 작동중인 운영체제 커널에 실행 기능을 모듈로 추가하기 위한 기법. LKM은 현재 리눅스, 솔라리스, FreeBSD를 포함한 많은 운영체제에서 사용되고 있다. 트로이목마에 감염된 모듈이 LKM에 있으면 시스템을 리부팅 한다해도 부팅 과정에서 다른 커널 모듈들과 마찬가지로 감염된 LKM 모듈을 다시 적재하게 된다. 따라서 LKM 사용상의 이점이 널리 인정된다 하더라도 악의적인 침입자들에 의해 악용할 수도 있기 때문에 기능 자체를 반대하는 사람들도 있다.