특집

파라메트릭 다객체 오디오 부호화 기술

서정일·강경옥 (한국전자통신연구원 방통융합미디어연구부)

I. 서 론

2000년대 초 MPEG(Moving Picture Expert Group)에서 표준화가 완료된 AAC(Advanced Audio Coding)가 심리음향 모델을 이용하여 오디오 신호내의 지각적 잉여성분을 제거하는 지각 오디오 부호화(Perceptual Audio Coding) 기술의 이론적 한계에 도달하게 되었다. AAC는 64kbps 내외에서 CD(Compact Disc)와 구별할 수 없는(indistinguishable) 음질을 제공할 수 있으나, 64kbps이하의 제한된 대역폭에서 고품질을 음질을 제공해야하는 이동형 방송환경에서는 적절하지 않았으므로 지각 오디오 부호화 기술의 기술적 한계를 뛰어넘기 위한 새로운 오디오 부호화 기법에 대한 연구가 활발히 이루어졌으며 이의 결과로 파라메트릭 오디오 부호화 기술이 개발되었다.

초기의 파라메트릭 오디오 부호화 기술은 오디오 신호를 구성하는 음원들을 모델링하여 소수의 파라미터로 표현함으로써 비트율을 극도로 절약하는 것을 목표로 하였으며, MPEG에서는 HILN(Harmonic and Individual Lines and Noise)과 SSC(SinuSoidal Coding)이란 이름으로 표준화 되었다. 그러나 제한된 파라미터만으로 오디오 신호를 표현함으로 인하여 음질의 열화가 심하였기 때문에 오디오 방송과 같은 상용서비스로의 적용에는 적절하지 못하였다.

2000년대 초반부터는 오디오 신호 자체를 제한된 파라미터로 표현하고자 하는 전방위적 접근방법에서 벗어나 오디오 신호의 일부 대역을 특징적인 파라미터로 표현하거나, 스테레오나 멀티채널 오디오 신호를 공간인지(spatial cue) 파라미터로 표현하는 기술에 대한 연구가 시작되었다.

SBR(Spectral Band Replication)은 사람이 고주파수 신호에 대해서는 세밀한 스펙트럼 성분을 인지하는 것이 아니라 신호 레벨의 괘적(envelop)에 의지하여 인지한다는 특성을 이용하여 입력된 오디오 신호의 저주파수 대역은 기존의 지각 오디오 부호화 기술을 이용하여 부호화하고 고주파수 대역은 저주파수 대역의 스펙트럼을 복사한 후 원신호 스펙트럼의 괘적과 특이성분(강한 하모닉 성분 등)을 모사해 줌으로써 부호화에 필요한 데이터량을 획기적으로 감소시키는 기술이다. SBR은 AAC와 결합된 HE-AAC(High Efficiency AAC), aacPlus 등

의 이름으로 상용화 되었으며, MP3, BSAC(Bit Sliced Arithmetic Coding) 등과 같은 오디오 코덱과의 결합도 시도되고 있다.

스테레오나 멀티채널 오디오 신호가 공간상에서 인지되는 특성 파라미터를 이용하여, 스테레오나 멀티채널 신호를 모노나 스테레오 신호로 다운믹스(downmix)하여 일반적인 오디오 코덱으로 부호화하고 스테레오나 멀티채널 오디오 신호가 가지고 있는 공간적인 특성을 제한된 개수의 파라미터로 표현하는 기술이 제안되었다. MPEG에서는 스테레오 오디오 신호를 위한 PS(Parametric Stereo)와 멀티채널 오디오 신호를 위한 MPS(MPEG Surround)가 표준화 되었으며 수차례에 걸친 실험과 청취평가를 통하여 표준화된 기술들이 상용서비스에도 적합함을 확인하였다.

또한, MUSIC2.0 음반과 같이 오디오 콘텐츠를 구성함에 있어서 목표하는 채널(스테레오 또는 멀티채널)로 믹싱되기 이전인 오디오 객체 신호들을 직접 이용함으로써 청취자에게 다양한 기능을 제공하는 객체기반 오디오 서비스가 등장하고, 원거리회의(tele-conference) 서비스와 같이 다수의 화자들로 구성된 음원들을 독립적으로 처리하면서도 비트율을 감소시키고자 하는 연구가 진행되었다. 이는 MPS와 같은 파라메트릭 멀티채널 오디오 부호화 기법의 입력 신호를 멀티채널에서 멀티객체로 변환된 것으로 볼 수 있으므로 MPS와 유사한 접근방법으로 해결하고자 하는 노력이 MPEG에서 진행되고 있으며 SAOC(Spatial Audio Object Coding)이란 이름으로 표준화가 진행 중이다.

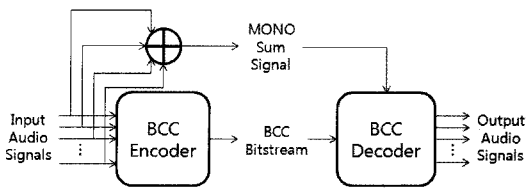
본 논문의 구성은 다음과 같다. 2장에서는 공간파라미터를 이용하여 멀티채널 및 멀티객체 오디오 신호를 압축하는 공간 오디오 부호화 기술에 대해 설명하고, 3장에서는 현재 표준화가 진행 중인 SAOC 기술에 대해서 소개하고, 4장에서 결론을 맺는다.

II. 공간 오디오 부호화 기술

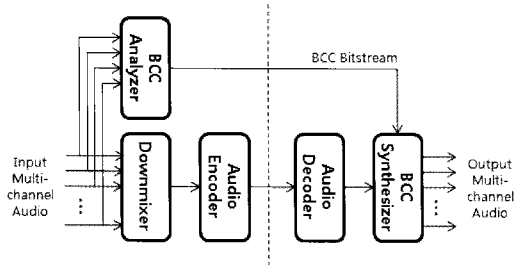
멀티채널 또는 멀티객체 신호를 모노나 스테레오 다운믹스 신호와 공간 파라미터로 압축하여 표현하는 공간 오디오 부호화(Spatial Audio Coding, SAC) 기술은 Faller에 의하여 BCC (Binaural Cue Coding)이란 이름으로 2000년대 초 소개되었다.

<그림 1>에서와 같이 BCC 인코더는 입력된 오디오 신호들의 합신호(mono)와 저비트율의 BCC 비트스트림으로 변환하는 동작을 수행한다. BCC 디코더는 인코더의 역과정으로써 합신호와 BCC 비트스트림을 이용하여 원신호를 복원하는 과정을 수행한다.

BCC는 입력되는 오디오 신호의 특성에 따라서 Flexible Rendering을 위한 Type I과 멀티채널 오디오 신호의 부호화를 위한 Type II로 구분된다. BCC Type II는 <그림 2>와 같이 BCC 인코더의 입력 신호로써 5.1채널과 같은 멀티채널 오디오 신호를 입력받아 모노 합(mono downmix) 신호와 BCC 비트스트림을 생성한다. BCC 비트스트림은 주파수 대역별로 얻어진 각 채널간의 크기차이, 시간차이, 상관도로 구성되



〈그림 1〉 BCC 인코더 및 디코더



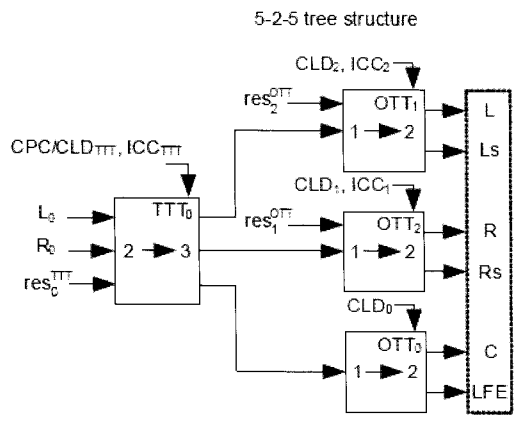
〈그림 2〉 BCC Type II: 멀티채널 오디오 인코더 및 디코더

며 이와 같은 BCC 파라미터를 예측 및 적용하기 위하여 CFB(Cochlear Filter Bank)를 이용한다.

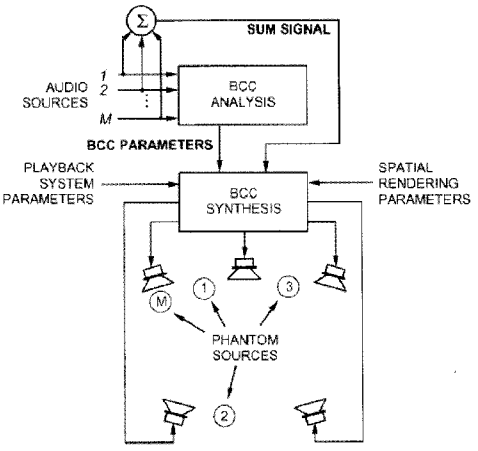
MPS는 BCC Type II를 기반으로 표준화 과정을 거쳐 성능이 개선된 것으로써 인코더와 디코더의 개념적인 동작과정은 BCC와 동일하지만 멀티채널 오디오 신호를 해석 및 합성을 위한 기본 블록으로 OTT(One-To-Two)와 TTT(Two-To-Three)의 조합을 이용한다는 차이점이 있다. <그림 3>은 5.1채널 오디오 신호를 스테레오 다운믹스 신호로부터 합성할 때(5-2-5 Configuration) OTT와 TTT 블록들이 어떻게 조합되는가에 대한 예이다. 물론, BCC의 성능 개선을 위하여 Hybrid QMF(Quadrature Mirror Filter) 필터뱅크, STP(Subband Domain Temporal Processing), Decorrelator, Residual Coding 등 다양한 툴들이 적용되었다.

<그림 3>에서 CLD(Channel Level Difference), CPC(Channel Prediction Coefficient), ICC(Inter Channel Correlation)은 MPS 비트스트림을 구성하는 공간 파라미터들이며 res는 추가적인 음질 개선을 위해 제공되는 잔차 신호(residual signal)를 의미한다.

Flexible Rendering으로 표현되는 BCC TypeII은 <그림 4>와 같이 BCC 인코더는 음원신호(audio source)들을 입력받아 합신호와 이들로부터 예측된 BCC 파라미터를 출력하고, BCC 디코더는 합신호와 BCC 파라미터로부터 음원신호를 복원하고 재생 시스템 환경에 맞게 음원들을 가상 음향공간에 렌더링하는 과정을 수행한다. 물론, 복원된 음원을 가상 음향공간에 렌더링하기 위해서는 음원의 위치정보를 추가로 입력받아야 하며, 이러한 위치정보는 인코더로부터 주



〈그림 3〉 5-2-5 조합에서의 5.1채널 신호 합성을 위한 OTT및 TTT 구성 블록도



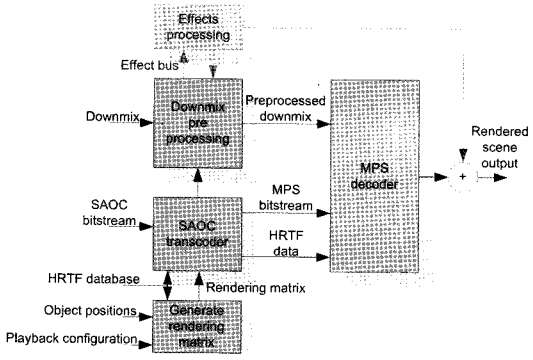
〈그림 4〉 Flexible Rendering을 위한 BCC에서의 오디오 신호 압축 및 복원과정

어지거나 디코더에서 청취자의 선택에 따라 결정될 수 있다. 따라서 BCC Type I은 객체기반 음반, 원격회의시스템 등과 같이 많은 음원들을 직접 압축하여 전송하고 청취자의 취향이나 기호대로 재생할 수 있는 대화형 오디오 서비스를 낮은 비트율에서 제공할 수 있다.

BCC Type II를 기본으로 하여 멀티채널 오디오 부호화를 위한 MPS를 표준화한 MPEG에서는 BCC Type I이 객체기반 오디오 서비스, 게임, 원격회의 등과 같은 응용에 적절한 것으로 판단하여 SAOC란 이름으로 표준화를 진행하여 2009년말 국제표준으로 공표될 예정이다. 이어지는 3장에서 SAOC 기술에 대해서 상세히 소개하고자 한다.

III. Spatial Audio Object Coding

2006년말 MPS의 표준화가 완료되어갈 즈음에 BCC Type I을 이용한 멀티객체 오디오 부호화 기술에 대한 표준화가 독일의 프라운호퍼 연구소에 의해 제안되었으며 ETRI, LG전자 등과

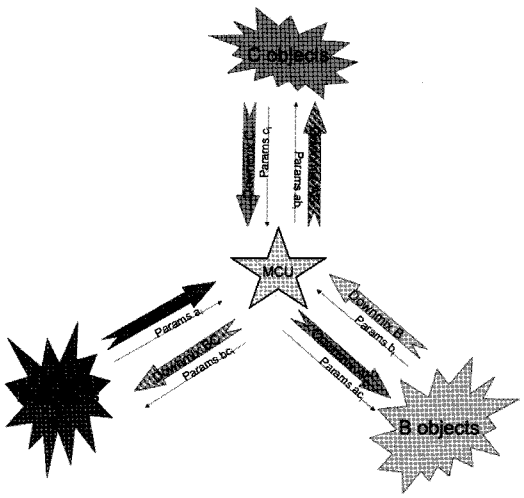


〈그림 5〉 SAOC 트랜스코더와 MPS 디코더의 결합구조. SAOC 표준화 초기에는 핵심 표준화 대상(normative part)은 다운믹스 전처리 과정과 SAOC 트랜스코더로 한정함

같은 여러 기관들의 공동작업을 통해 CFP(Call for Proposal)가 2007년 1월 발표되었다. 표준화 초기에 SAOC가 목표로 한 것은 MPS를 기본 디코더로 이용하고 SAOC 비트스트림을 MPS 비트스트림으로 변환하는 Transcoder를 표준화하고자 하였다<그림 5 참조>.

1. Application Scenarios

SAOC의 응용분야는 음원신호를 직접 효율적으로 부호화하는 것을 전제로 한 객체기반 대화형 오디오(interactive re-mix) 서비스 및 게임/리치미디어 분야와 음원으로 원격회의시스템의 음성을 다루는 Teleconferencing/telecommunication 이었다. Interactive re-mix와 Game/rich media 서비스에서는 입력된 음원신호들을 공간 오디오 부호화 기술을 이용하여 낮은 비트율로 압축하는 것을 목표로 하였으며, Teleconferencing

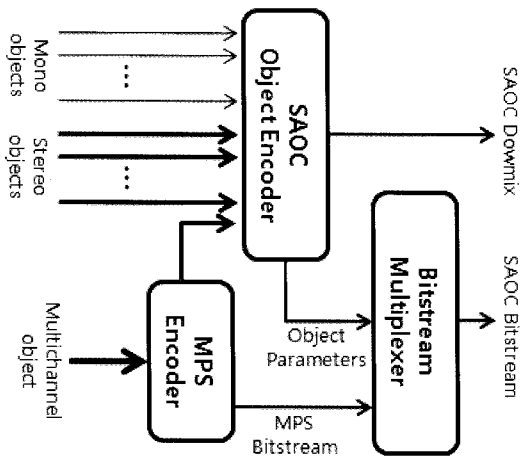


〈그림 6〉 원격회의 시스템에서 SAOC 비트스트림을 이용한 음성신호 결합 및 배분과정. MCU(Multipoint Control Unit)는 여러 사이트에서 전달된 음성객체 신호들을 결합하여 다른 사이트들로 분배하는 과정을 담당한다

서비스에서는 <그림 6>과 같이 원격회의에 참여하는 음성신호들을 효율적으로 결합 및 배분하는 것을 목표로 하였다.

2. SAOC Encoder

<그림 7>은 SAOC 인코더의 구조를 나타낸 블록도이다. SAOC 인코더는 입력 신호로써 모노, 스테레오 또는 멀티채널 오디오 객체를 입력받을 수 있다. 그러나 오디오 객체 파라미터를 추출하는 과정을 간략화하기 위하여 스테레오 오디오 객체는 모노 오디오 객체가 두 개 존재하는 것으로 처리하며, 멀티채널 오디오 객체의 경우에는 MPS 인코더를 통과한 스테레오 다운믹스 신호를 입력 오디오 객체로 처리하고 MPS 비트스트림을 객체 파리미터와 다중화하여 SAOC 비트스트림을 구성한다. 멀티채널 오디오 객체는 SAOC 트랜스코더나 디코더에 의해서 생성되는 음향장면에서 멀티채널 배경음을 구성할 뿐 청취자에 의하여 제어되지 않는 것을 가정하고 있기 때문이다.



〈그림 7〉 SAOC 인코더 블록도

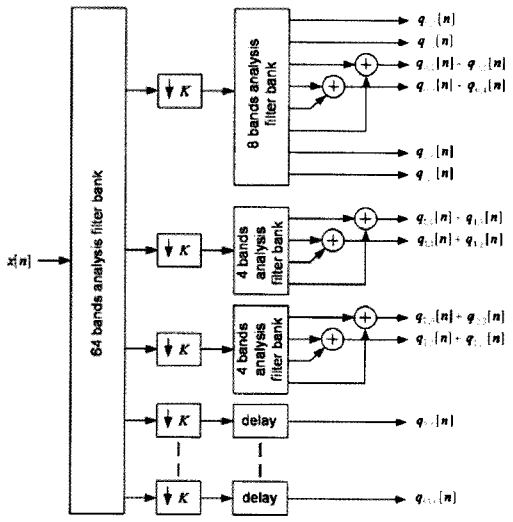
SAOC 인코더 과정 중 해석되는 객체 파라미터는 아래와 같다.

- NRG(absolute object eNeRGy): 가장 큰 에너지를 가진 객체의 에너지 값
- OLD(Object Level Difference): 해당하는 객체와 가장 큰 에너지를 가진 객체와의 에너지비
- IOC(Inter-Object Correlation): 객체들 간의 유사도
- DMG(DownMix Gain): 객체들을 다운믹스 할 때 취해지는 에너지 스케일값
- DCLD(Downmix Channel Level Difference): 스테레오 다운믹스 신호의 각 채널에 대한 객체들의 레벨 차이값
- PDG(Post Downmix Gain): SAOC 인코더에서 생성된 다운믹스 신호와 외부에서 입력된 다운믹스 신호사이의 레벨 차이값

3. T/F Transform

멀티객체 신호들로부터 객체 파라미터를 예측하고 복원하는 과정은 주파수 대역별로 수행되며 입력된 시간축 신호를 주파수축으로 변환하기 위한 툴로써는 MPS와 동일한 Hybrid QMF를 이용한다.

낮은 연산량과 시간축 필터링을 이용할 수 있다는 장점을 지닌 QMF는 입력되는 시간축 신호를 고속 연산을 위해서 2N밴드로 균등하게 분배한다. MPS와 SAOC에서는 64밴드의 QMF 필터를 이용하게 되는데 고주파수 대역에서는 각 밴드의 분해능이 사람이 주파수를 분해하는 능력인 critical band보다 우수하기 때문에 문제가 되지 않으나 저주파수 대역에서는 하나의 QMF



〈그림 8〉 Hybrid QMF 해석필터 블록도

밴드가 여러 개의 critical band에 맵핑됨으로 인해 적절한 객체 파라미터 해석이 불가능하다. 이러한 QMF 변환의 단점을 보완하기 위하여 저주파수 밴드를 다시 낮은 차수의 QMF 변환을 적용함으로써 critical band와 유사한 주파수 분해능을 가지도록 한 것이 Hybrid QMF 변환이다.

<그림 8>은 MPS와 SAOC에서 사용하는 71밴드 Hybrid QMF 필터의 블록도이다. 실제 객체 파라미터의 해석은 71개의 hybrid 밴드를 4개에서 28개의 파라미터 밴드로 묶어서 처리하게 된다. 또한, 파라미터 기반의 오디오 부호화 기법에서 발생할 수 있는 위상의 불일치를 보정하기 위하여 Hybrid QMF는 실수와 허수 부분을 분리하여 처리하게 되며, 연산량의 감소를 위하여 Low-Power SAOC 디코더에서는 일부 대역에 대해서는 실수 Hybrid QMF만을 수행한다.

또한, SBR에서도 동일한 64밴드 QMF 변환을 이용하므로 다운믹스 신호가 SBR로 처리되었을 경우에는 QMF 영역의 다운믹스 신호를 직접 입력받아 처리하는 것도 가능하다. 단, 다운믹스 신호의 결합형태에 따라서 T/F 변환에 의한 지연(delay)시간이 달라지므로 이에 대한 보상이 필요하다.

4. SAOC Bitstream

SAOC 인코더를 통하여 부호화된 SAOC 비트스트림은 디코딩에 필요한 모든 정보를 포함하며 주요 syntactic element는 아래와 같다.

- SAOCSpecificConfig(): SAOC 디코더를 초기화하기 위한 헤더
- SAOCFrame(): Huffman 부호화된 SAOC 파라미터를 저장하는 프레임 데이터
- SAOCExtensionConfig(): residual 신호, 프리셋정보와 같이 부가적으로 추가될 수 있는 데이터를 전송하기 위한 container에 대한 헤더
- ObjectMetaData(): 객체에 대한 메타데이터 정보
- PresetMatrixData(): 사전에 정의된 렌더링 메트릭스(Preset) 정보
- ResidualData(): 특정 객체의 음질을 보상하기 위한 residual signal 정보

상기 element들 가운데 PresetMatrixData()에는 SAOC 디코더가 복원된 객체 신호들을 어떠한 비율로 믹싱하여 출력신호를 생성하는가에 대한 정보인 렌더링 메트릭스를 담고 있으며 이를 프리셋(preset)이라 명명하였다.

5. SAOC Operation Mode

SAOC 비트스트림을 해석하여 출력 신호를

〈표 1〉 SAOC의 동작모드

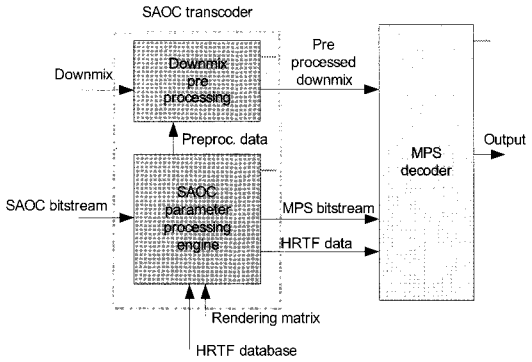
	SAOC Decoder	SAOC Transcoder
출력신호형태	mono/stereo/ binaural	multichannel
채널수	1 or 2	〉 2
SAOC 모듈의 출력신호	PCM output	MPS bitstream, downmix signal
MPS 디코더	NO	YES

생성할 때 출력되는 오디오 신호의 채널 수에 따라서 SAOC는 디코더 형태를 <표 1>과 같이 구분하여 정의한다.

SAOC 디코더는 일반적인 오디오 디코더와 같이 다운믹스 신호와 SAOC 비트스트림을 입력받아 디코딩된 오디오 신호를 출력하지만, SAOC 트랜스코더는 SAOC 비트스트림을 MPS 비트스트림으로 변환(transcoding)하여 조정된 다운믹스 신호와 함께 MPS 디코더로 전달함으로써 최종 멀티채널 오디오 출력신호는 MPS 디코더가 생성하게 한다. 위와같이 처리한 이유는 SAOC의 멀티채널 오디오 출력을 가정하지 않았으며 멀티채널 복원은 MPS 디코더에 의해 처리되도록 결정하였기 때문이다.

6. SAOC Transcoder Mode

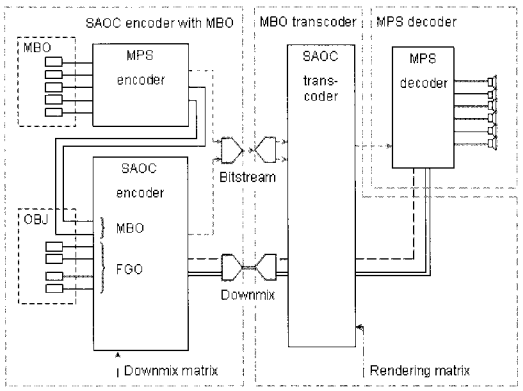
<그림 9>에서와 같이 SAOC 트랜스코더는 다운믹스 신호를 조정하는 다운믹스 전처리부와 SAOC 파라미터 처리 엔진부로 구성된다. 다운믹스 전처리부에서는 입력된 다운믹스 신호에서 특정 객체를 삭제하거나 MPS 디코더에서 불가능한 위치로 객체를 이동시키기 위하여 좌우 채널신호를 변경하는 등의 전처리 과정을 수행한다. SAOC 파라미터 처리 엔진부에서는 입력된 SAOC 비트스트림을 MPS 비트스트림으로 변환하는 과정을 수행하며, 이를 위하여 외부나



〈그림 9〉 SAOC Transcoder 블록도

SAOC 비트스트림으로부터 전달받은 렌더링 메트릭스를 이용한다. <그림 9>에서 HRTF(Head Related Transfer Function) 데이터가 트랜스코딩 과정에서 변경되는 것으로 되어있지만 바이노럴 출력 모드는 SAOC 디코더 모드에서 처리되는 것으로 표준이 변경되어 SAOC 트랜스코더 모드에서는 필요하지 않게 되었다.

멀티채널 배경음 객체(MBO: Multichannel Background Object)가 MPS 인코더로 부호화되어 SAOC 비트스트림으로 입력되었을 경우에는 MBO가 부호화된 MPS 비트스트림이 MPS 디코더로 직접 입력되게 되며, 객체 파라미터들

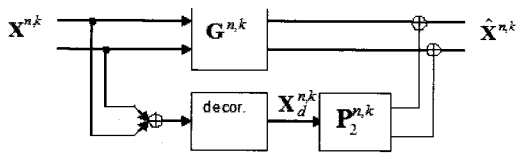


〈그림 10〉 MBO 트랜스코딩 처리 과정도

로부터 변환된 MPS 비트스트림도 함께 MPS 디코더 입력되어 멀티채널 출력신호를 생성한다. 또한, MBO 객체의 재생 유무와 객체신호의 재생 유무에 따라서 MBO 비트스트림만이 전송되거나 제거되기도 한다.

7. SAOC Decoder Mode

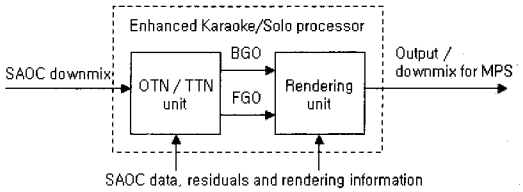
앞서 설명한바와 같이 SAOC 디코더는 모노 또는 스테레오 다운믹스 신호를 입력받아 모노, 스테레오, 바이노럴 스테레오 신호를 출력한다. 바이노럴 스테레오 신호는 다운믹스 전처리부와 MPS 바이노럴 디코더가 결합된 형태를 의미한다. <그림 11>은 스테레오 다운믹스 신호가 입력되고 바이노럴 스테레오 신호를 출력하는 과정을 나타낸 그림이다. G 매트릭스는 객체 파라미터와 렌더링 메트릭스에 의해 결정되는 upmix 매트릭스이며 P 매트릭스는 G 매트릭스와 유사한 기능을 수행하지만 decorrelation된 다운믹스 신호를 입력 받는다는 차이가 있다.



<그림 11> 스테레오 다운믹스 신호에 대한 바이노럴 디코더 처리 구조

8. Enhanced Karaoke/Solo Mode

실제로는 SAOC 인코더에서 다운믹스 과정을 통해 합쳐진 신호에서 객체 파라미터만으로 원래의 객체 신호들을 완벽하게 복원하는 것은 불가능하다. 따라서 SAOC에서는 12dB 정도로 객체의 레벨을 조절하는 것을 목표로 하였으며, 객체



<그림 12> Enhanced Karaoke/Solo 모드 디코더의 처리구조

신호를 모두 복원하는 과정은 고려하지 않았었다. 그러나 가라오케와 같이 특정 객체를 완벽하게 삭제하거나 특정 객체만을 재생하는(Solo) 응용분야가 존재하기 때문에, 이와같은 상황에서도 만족할만한 음질을 제공하기 위하여 SAOC에서는 Enhanced Karaoke/ Solo 모드를 제공한다.

특정 객체 신호를 완벽하게 제거하거나 복원하기 위해서 원음과 객체 파라미터로 복원되는 음 사이의 잔차신호(residual signal)을 MPS 표준에서 사용하던 잔차신호 부호화 기법을 이용하여 추정하여 이용한다. SAOC는 MPS의 기본 처리 블록인 OTT와 TTT를 사용하지 않지만 MPS의 잔차신호 부호화를 재사용하기 위해서 이용한다. Enhanced Karaoke/Solo 모드의 처리과정은 <그림 12>와 같으며, BGO (BackGround Object)는 객체 파라미터 만으로 처리되는 객체를 뜻하며 FGO(ForeGround Object)는 잔차신호를 이용하여 특별히 관리되는 객체를 의미한다.

Ⅳ. 결 론

지각 오디오 부호화 기술의 기술적 한계를 극복하기 위하여 파라메트릭 오디오 부호화 기술이 연구 및 개발되고 있으며, MPEG에서 표준화된 Parametric Stereo, MPEG Surround가 압축율에서나 음질에서 만족스러운 성능을 보임에 따라 DMB(Digital Multimedia Broadcasting),

DAB+(Digital Audio Broadcasting plus)등과 같은 이동형 방송에서 스테레오나 멀티채널 오디오를 위한 코덱으로 선정되었다.

또한 MUSIC 2.0과 같은 객체기반 음반서비스의 출현과 Teleconferencing 시스템에서의 효율적인 부호화를 위해 파라메트릭 부호화 기술을 이용하는 SAOC의 표준화가 거의 완료단계이다. SAOC는 새로운 개념의 대화형 오디오 서비스를 기존 오디오 서비스와 유사한 대역폭으로 제공하는 것이 가능하므로 빠른 시기에 가라오케, AOD(Audio on Demand) 등과 같은 응용분야에 적용될 것으로 예상된다.

참고문헌

[1] ISO/IEC 14496-3:2005, MPEG-4 coding of audio-visual object, Part 3: Audio, 2005.

[2] ISO/IEC 23003-1:2007, MPEG-D MPEG audio technologies, Part 1: MPEG Surround, 2007.

[3] ISO/IEC FCD 23003-2, MPEG-D MPEG audio technologies, Part 2: Spatial Audio Object Coding (SAOC), 2009.

[4] Christof Faller and Frank Baumgarte, "Binaural Cue Coding Applied to Audio Compression with Flexible Rendering," 113th AES Convention, Oct., 2002.

저자소개

서 정 일

1994년 2월 경북대학교 전자공학과 학사
1996년 2월 경북대학교 전자공학과 석사
2005년 8월 경북대학교 전자공학과 박사
1998년 3월 ~ 2000년 10월 LG반도체 주임연구원
2000년 11월 ~ 현재 한국전자통신연구원 선임연구원

주관심 분야 : 오디오 부호화, 다채널 음장재현 시스템, 3차원 오디오, 디지털 방송 시스템, 객체기반 오디오 시스템

강 경 옥

1985년 2월 부산대학교 물리학과 학사
1988년 2월 부산대학교 대학원 물리학과 석사
2004년 2월 한국항공대학교 대학원 항공전자공학과 박사
2006년 4월 ~ 12월 영국 Southampton 대학 방문연구원
1991년 2월 ~ 현재 한국전자통신연구원 팀장

주관심 분야 : 음향 신호처리, 3차원 오디오, 디지털 방송 시스템, 객체기반 오디오 시스템