논문 2010-47SP-5-6

객체 오디오 부호화 표준 SAOC 기술 및 응용

(Object Audio Coding Standard SAOC Technology and Application)

오 현 오*, 정 양 원**

(Hyen-O Oh and Yang-Won Jung)

요 약

객체 기반 오디오 부호화 기술은 다양한 응용 분야를 기대할 수 있는 차세대 오디오 기술로써 관심이 높다. 최근 MPEG에서는 SAOC (Spatial Audio Object Coding)라는 압축 효율이 우수한 Parametric 객체 부호화 방법을 표준화하였다. 본 논문에서는 SAOC를 중심으로 Parametric 객체 오디오 부호화의 기술을 소개하고, 이를 실제 적용하기 위한 고려사항들에 대해 다룬다.

Abstract

Object-based audio coding technology has been interested with its expectation to apply in wide areas. Recently, ISO/IEC MPEG has standardized a parametric object audio coding method, the SAOC (Spatial Audio Object Coding). This paper introduces parametric object audio coding techniques with special focus on the MPEG SAOC and also describes several issues and solutions that should be considered for a success in its application.

Keywords: SAOC, MPEG, Parametric Coding, Audio Object, Immersive Multimedia

I. 서 론

객체 기반의 신호처리 및 부호화 기술의 실현은 오디오뿐 아니라 영상을 포함한 모든 멀티미디어 응용분야에 있어서, 미래의 기술로 관심이 높다. 영화나 TV 컨텐츠의 장면을 구성하는 각 객체에 대해 사용자 임의의 추출, 확대/축소, 이동 등의 효과를 입히는 것이 가능하다면, Interactive 컨텐츠 개발, 직관적인 인터페이스 구현, 현장감 높은 A/V 체험, Tele-presence 의 실현 등 다양한 응용 분야에서의 활용을 기대할 수 있다. 여기서 말하는 객체(Object)란 오디오 신호인 경우 음악을 구성하는 각 악기 등을 의미한다. MPEG에서는 일찌감치 1998년부터 MPEG-4 표준화의 화두를 객체 기반 부호화로 가져가고 이를 위해 필요한 요소 기술들을 찾고자 하였다[1]. 그러나 카메라로 찍은 영상이나 마이크를 통해 취득한 오디오 등 자연의 컨텐츠로부터 이를 구성하는 각 객체 신호를 효과적으로 생성하는 것에서부터 어려움이 있었기 때문에 MPEG-4의 본래 취지는 사실상 많이 퇴색되고 특히 A/V Codec의 경우, MPEG-2를 성공적으로 잇기 위한 추가적인 압축 부호화 기술 표준화의 방향으로 진화하였다. MPEG-4 오디오의 경우도 표준화 초기에 자연 오디오, 음성, 합성 오디오 등 입력된 각 객체 신호를 그 특성에 따라 분류하고, 각각 최적의 방법으로 부호화하여 전송하고, 이를 복호화단에서 객체 별로 분리할 수 있는 오디오 부호화 기술을 표준화하고자 하였다[2]. 또한, 이와 같이 전송된 신호는 복호화 과정에서 사용자 임의의 방법으로 컨텐츠를 재구성할 수 있어야 하는데, 이를 위한 장면 기술자로써 MPEG-4 BIFS (Binary Format for Scene)와 같은 기술을 표준화하였다[3~4]. 그러나 전술한 바와 같이 객체별 컨텐츠 획득의 어려움과 전송의 복잡성, 이를 재생

* 정회원, LG전자
(LG Electronics)
** 정회원, 인텔렉추얼 벤처스 코리아
(Intellectural Ventures Korea)
접수일자: 2010년7월5일, 수정완료일: 2010년8월5일

하기 위한 복호화기의 복잡성 (객체 수 만큼의 독립적인 복호화를 병렬 처리한 후, 다시 장면을 합성하는 과정을 거쳐야 함) 등으로, MPEG-4 본래 취지에 맞는 실체를 구현하지 못한 실정이다.

한편, MP3 (MPEG-1/2 Layer III)[5]의 시장 성공을 뒤이어, MPEG-2/4 AAC (Advanced Audio Coding)[6~7]가 성공적으로 표준화 되고 시장에 널리 사용되면서, MPEG Audio 표준화는 고품질 오디오 신호의 효과적인 압축을 향해 진화해 왔다. 그와 같은 진화의 연장선에서 SBR (Spectral Bandwidth Replication)[8] 및 PS (Parametric Stereo)[9] 등의 기술이 개발되었으며, 이를 AAC와 연동할 수 있도록 표준화한 HE AAC (High Efficiency AAC) V1/V2 가MPEG-4 Audio에 포함되었다[10~11]. AAC가 약 64~128 kbps / stereo의 비트율에서 고품질을 보장하는 부호화 방법인데 반해, HE AAC V1 (AAC+SBR)은 32~64 kbps / stereo, 그리고 HE AAC V2 (AAC+SBR+PS)는 24~32 kbps / stereo 일 때 적절한 음질을 보장한다. 더불어 SBR과 PS 기술은 그 이전의 오디오 부호화의 패러다임이던 심리음향모델에 기반하고MDCT Filterbank를 사용하는 Perceptual transform coding을 Parametric coding으로 전환하는 계기를 제공하였다.

Parametric coding 패러다임은 이후 MPEG Surround 표준화로 이어진다[12]. MPEG Surround는 모노 혹은 스테레오 (다운믹스) 신호에 약간의 부가 정보 (Side Information, 약 10kbps 정도)로써 공간 파라미터 (Spatial Parameter)를 함께 전송하여, 수신단에서 5.1 채널 등 다채널 신호를 복호화 얻을 수 있도록 하는 기술이다[13]. 이때, 흔히 다운믹스 신호라고 칭하는 모노 혹은 스테레오 신호는 기존의 모노/스테레오 복호화만을 지원하는 단말(예를 들어 MP3 player)에서 복호화가 가능하고, MPEG Surround 기능을 가진 새로운 단말에서는 다운믹스 신호와 함께 부가 정보를 이용하여 다채널 신호를 생성할 수 있다. 따라서 기존 단말과의 하향 호환성 (Backward Compatible)이 있으면서도 다채널 서라운드 신호 재생이라는 추가적인 기능을 제공해준다.

MPEG Surround 표준화 이후, MPEG에서는 자연스럽게 Parametric coding의 개념을 객체 부호화에도 적용할 방안을 모색하게 되었다. MPEG Surround의 기본 구조 안에서 입력신호를 각 채널 신호 대신 각 객체 신호로 치환한다면, 결국 수신단에서 객체 별 신호를 채

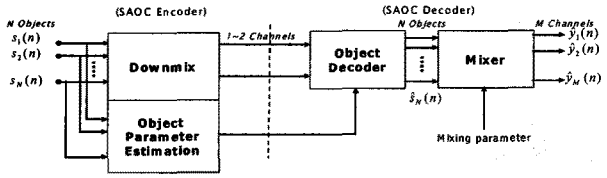


그림 1. Parametric 객체 오디오 부호화기구조도
Fig. 1. Block diagram of Parametric Object-Oriented Audio Coding.

널 신호대신 얻을 수 있을 것이란 점에 착안하여, 표준을 확장할 방안을 고려하게 된다. 그림 1은 이와 같이 MPEG Surround로부터 진화하여 Parametric 객체 오디오 부호화를 실현할 수 있도록 변형된 개념의 객체 오디오 부호화기 구조도이다. 다운믹스 신호와 이에 포함된 객체 정보를 이용하여 수신단에서 각 객체를 추출할 수 있다고 가정하면, 사용자가 원하는 대로 (Mixing Parameter) 오디오 장면을 재구성한 출력신호를 얻어낼 수 있다. 이는 MPEG-4에서 표방하던 객체 지향 부호화의 틀에 잘 부합하면서 현재의 기술 수준 및 범주 안에서 구현이 가능한 표준 기술이 만들어 질 수 있음을 시사하였다. 이와 같은 동기로 표준화된 기술이ISO/IEC 23003-2, MPEG-D SAOC (Spatial Audio Object Coding)이다[14~15]. 한편MPEG Surround의 이론적 근거를 제시했던 유명한 논문이기도 한 C. Faller 의 BCC (Binaural Cue Coding)에서는, Flexible rendering 이라는 개념으로 SAOC와 같은 Parametric 객체 부호화가 가능할 수 있음을 제시한 바 있다[16~17].

SAOC는 2007년 1월에 CfP (Call for Proposal)[18]가 발표되면서 본격적인 표준화의 궤도에 올라 만 3년만인 2010년 2월에 FDIS (Final Draft International Standard)를 완성하며 표준화 일정을 마쳤다[15]. CfP에서는 SAOC의 가능한 응용분야로 Interactive Music Remix, Teleconference, Gaming/ Rich Media 를 제시하고 있다. 아울러, 특히 3DTV의 활성화와 함께 실감방송에 대한 관심이 고조되고 있는 시점에서, SAOC의 객체 별 제어 가능성과 기존 방송과의 하향 호환성 (Backward Compatibility) 제공은 차세대 방송 표준으로의 활용에도 관심을 높이고 있다.

한편 MPEG SAOC 가 그 이전의 다른 오디오 부호화 표준과 구별되는 또 다른 개념적 차이는 복호화를 위해서는 송신단에서 전송하는 비트열 이외에도 수신단에서 사용자의 입력이 필수적으로 요구되며, 이를 다루고 처리하는 기술이 표준화의 범주 안에 진입했다는 의

의가 있다.

본 논문에서는 차세대 오디오 기술로써, 특히 새로운 형태인 객체 기반의 차세대 오디오 부호화 방법인 MPEG SAOC 표준 기술을 소개하고, 특히 SAOC를 주요 응용 분야에 적용 시 고려해야 할 몇 가지 기술적 사안 및 이에 대한 해결 방안을 제시하고자 한다.

Ⅱ. 객체 오디오 기술의 응용 분야

1. 음악Remix

SAOC의 가장 큰 응용 분야로 예상되는 것이 바로 Interactive Music Remix이다. 일반적인 음악 컨텐츠는 음악을 구성하는 각각의 악기를 개별적으로 녹음(각각을 트랙이라고도 함)한 후, 이것을 믹싱 단계에서 프로듀서의 의도에 따라 적절히 조합하여 만들어진다(mastering). SAOC는 이렇게 만들어진 다운믹스 신호에 약간의 부가 정보를 더하여, 수신단의 사용자가 마치 자신이 프로듀서가 된 것처럼 각각의 오브젝트에 대해 독립적인 제어, 즉 리믹스 (Remix) 기능을 제공한다. 예를 들어, 특정 오브젝트의 크기를 줄이거나 키울 수 있고, 특정 사운드 스테이지 안에서 오브젝트의 공간적인 위치를 변경하는 등의 제어가 가능해진다. 극단적으로는 특정 오브젝트의 신호를 제거하는 것도 생각할 수 있다 (보컬 오브젝트를 제거하여 노래방 반주 즉 karaoke scene을 연출할 수 있다). 또한, MPEG Surround와 마찬가지로 SAOC는 기존 오디오 부호화 포맷과 하향 호환성을 갖고 있기 때문에, SAOC 부가정보가 포함된 다운믹스 오디오 신호는 기존의 일반적인 오디오 재생기를 통해 원 오디오 신호의 재생이 가능하다.

2. 차세대 방송

종래의 단방향성을 가지는방송에서 시청자의 Interactivity혹은 자유도를 제공하려는 시도들이 있어 왔는데, 예를 들어 어학 학습을 위해 방송에서 한국어와 외국어의 음성을 선택적으로 청취하거나, 스포츠 중계에서 관중 함성 등의 현장음과 캐스터/ 해설의 음성 크기의 비율을 사용자의 취향에 맞게 조절한다거나, 야간 영화 시청 시 화려한 효과음과 배경음은 줄이고, 대사만을 주로 재생할 수 있도록 하는 등의 필요성과 응용 분야가 제시되었다. 방송의 경우 각 객체 신호에 대해 독립적인 채널로써 부호화하여 전송하여 위와 같은

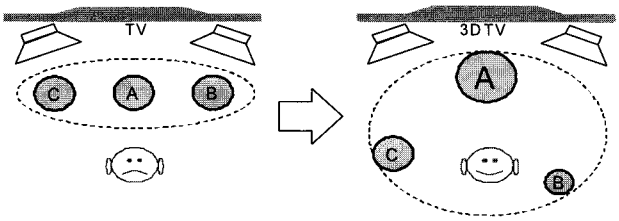


그림 2. SAOC를 이용한 실감 방송 서비스
Fig. 2. Example of Immersive Broadcasting Service Application.

응용을 구현할 경우 필요한 대역폭이 전송되는 객체 수에 비례하여 증가하며, 수신단에서는 각각의 채널을 복호화한 후 믹싱하여야 최종적인 오디오 장면을 구성할 수 있기 때문에 과도하게 비싼 솔루션이다. SAOC를 이용하면 증가되는 대역폭이 기존 오디오 신호 비트율의 100~150% 정도면 가능하고, 하방 호환성을 보장하기 때문에, 기존 방송 인프라의 부가 서비스로 도입이 용이하며, 차세대 방송 서비스로의 다양한 활용을 기대할 수 있다. 최근 3D TV의 유행은 객체 기반 컨텐츠의 필요성을 더욱 크게 하는데, 장면을 구성하는 객체 별로 임의의 3차원 공간에 rendering이 가능하여, 사용자 취향에 따른 입체감 제어를 보다 잘 구현할 수 있다.그림 2는 3DTV와 같은 실감 방송 서비스에서의 SAOC 사용 예를 나타낸다. 대사에 해당하는 객체 A는 더 강조하여 중앙에서 재생되도록 하는 동시에, 서라운드 효과가 바람직한 객체 C와 D에 대해서는 3D rendering을 실현하여 보다 현장감 있는 사운드를 제공할 수 있음을 나타낸다.

3. 원격 회의

SAOC 표준화 과정에서 음악 Remix와 함께 주요 응용 분야로 기대가 높았던 것이 바로 원격 회의(Teleconference)이다. 보통 모노 (혹은 최대 스테레오)로 전송되는 원격 회의 환경의 경우, 원격의 참가자들의 음성이 모두 동일한 물리적 위치 (모노 스피커의 위치)에서 출력되기 때문에, 현재 발화하는 참가자를 식별하기에 어렵고, 화자간 중첩에 의해 대화 내용의 이해에 어려움이 있다. 이러한 문제를 극복하고 원격 회의에서 실제감, 현장감을 높이기 위한 연구가 활발히 진행되고 있는데, SAOC는 이와 같은 실감 원격 회의 환경에 활용하기 적합한 부호화 기술이다. SAOC를 이용할 경우, 기존의 Teleconference 망과 호환되는 종래의 모노 전송 채널을 이용하더라도 각각의 원격 참가자

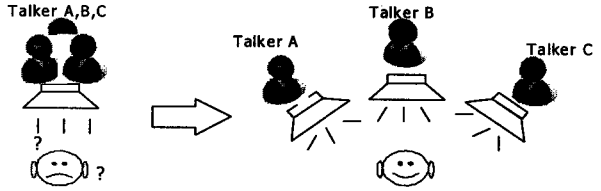


그림 3. SAOC를 이용한 원격 회의
Fig. 3. Example of Immersive Teleconference Application.

의 음성을 가상의 회의실 공간상에 독립된 물리적 위치에서 출력하는 것이 가능하고, 이를 통하여 음성 명료도를 증가시키고 발화하는 화자를 명확하게 식별하는 것이 가능하다. 특히, 화상 회의의 경우, 화면상의 원격 참가자의 위치와 음성 재생 위치를 공간적으로 일치시킴으로써 실제감을 높이는데 활용될 수 있다. Teleconferencing에서 더 나아가 원격지에 존재하는 사람이 동일 공간에 존재하는 것과 같은 상황을 연출하는 Telepresence에도 SAOC를 활용할 수 있다.그림 3은 원격회의 상의 SAOC 적용 시나리오를 나타낸다.

4. 게임/ Rich Media

Gaming 혹은 Rich Media의 경우도 앞선 사례와 마찬가지로 SAOC의 활용 가능성이 높은 응용 분야이다. 게임의 가상 공간 속 player간의 대화나 효과음을 각각의 오디오 객체로 표현하여 SAOC로 전송하게 되면, 전송 효율에서 매우 유리한 장점을 가진다.

Ⅲ. MPEG SAOC 표준 기술

본 장에서는 MPEG SAOC 부호화 및 복호화기 구조를 중심으로 주요 기술에 대해 간략히 설명한다. 보다 상세한 기술별 설명은 표준 문서 및 다른 논문을 참고할 수 있다[15, 19~20].

1. 부호화 과정

그림 1에서처럼 SAOC 부호화기는 N개의 객체 신호 $s_i(n)$을 입력으로 하여 다운믹스 신호 $x_j(n)$를 생성하며, 이 과정은 다음 식으로 설명할 수 있다.

$$\mathbf{X}=\mathbf{DS}=\begin{pmatrix} d_{11} & \cdots & d_{1N} \\ d_{21} & \cdots & d_{2N} \end{pmatrix}\begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix}=\begin{pmatrix} x_L \\ x_R \end{pmatrix} \quad (1)$$

여기서 $d_{j,i}$는 i 번째 오브젝트 신호가 j 번째 다운믹스 채널에 포함되는 정도를 나타내는 다운믹스 게인이며, $\mathbf{D}$는 이러한 다운믹스 게인들로 구성된 Downmix Matrix이다. 부호화기의 출력 신호는 다운믹스 신호인 $x_j(n)$와 추출된 객체 파라미터로 구성되는 SAOC bitstream이다. 출력된 다운믹스 신호는 기존의 MP3/AAC와 같은 오디오 부호화 방법에 의해 부호화하여 전송될 수 있다.

SAOC의 비트열을 구성하는 객체 파라미터에는 다운믹스 신호에 포함된 객체 신호의 레벨값에 대응하는 OLD (Object Level Difference)와 OLD 생성에 사용되는 정규화 값인NRG (absolute object eNeRGy), 각 오브젝트 간의 상관 관계에 관한 값인 IOC (Inter-Object Cross-correlation), 그리고, 객체신호가 다운믹스에 포함된 정도를 표현하기 위한 DMG(Downmix Gain), DCLD (Downmix Channel Level Difference) 가 있다. 이 가운데, OLD, NRG, IOC는 MPEG Surround와 동일한 28개의 파라미터 밴드 단위로 값을 가지며, DMG, DCLD는 전 밴드에 대해 하나의 값으로 추출된다[12]. 각각의 객체 파리미터의 추출은 다음과 같은 식에 의거하여 이루어진다.

$$OLD_i(pb)=10\log 10\left(\frac{\sum_n\sum_{m\in pb} s_i^{n.m}s_i^{n.m*}}{NRG(pb)}\right) \quad (2)$$

$$NRG(pb)=\max_k\left(\sum_n\sum_{m\in pb} s_k^{n.m}s_k^{n.m*}\right) \quad (3)$$

$$IOC_{ij}(pb)=\mathrm{Re}\left\{\frac{\sum_n\sum_{m\in pb} s_i^{n.m}s_j^{n.m*}}{\sqrt{\sum_n\sum_{m\in pb} s_i^{n.m}s_i^{n.m*}\sum_n\sum_{m\in pb} s_j^{n.m}s_j^{n.m*}}}\right\} \quad (4)$$

$$DMG_i=10\log_{10}(D_{1,i}^2+D_{2,i}^2+\varepsilon) \quad (5)$$

$$DCLD_i=20\log_{10}\left(\frac{D_{1,i}}{D_{2,i}}\right) \quad (6)$$

여기서 pb 는 파라미터 밴드를 의미한다.

추출된 파라미터들은 각 파라미터가 가지는 특성과 범위에 맞게 적합한 양자화 과정을 거친 후, MPEG

Surround와 유사한 Differential 및 Huffman 부호화 과정을 거쳐 비트열로 생성된다. SAOC 비트열은 객체당 약 3kbps 정도의 비트율을 갖는다.

2. 복호화 과정

SAOC 는 다채널 재생 지원을 위해 MPEG Surround를 결합한 독특한 구조를 가지고 있다. MPEG Surround와 파라미터의 생성 원리 및 전송 방법이 유사한 점을 고려하여 복호화기에서는 SAOC로 전송된 비트열을 MPEG Surround 비트열 형태로 transcoding 함으로써, 연결되어 있는MPEG Surround 복호화기를 이용하면, 사용자가 원하는 오디오 scene을 다채널로 재생할 수 있다. 그러나 MPEG Surround만을rendering engine으로 이용할 경우, 객체의 panning에 제약 사항이 발생할 수 있는데 이 문제를 해결하기 위해 Downmix processor가 도입되었으며, Downmix processor를 이용하게 되면, MPEG Surround를 사용하지 않고도 stereo 신호까지는 재생이 가능하다[21]. Downmix processor와 MPEG Surround 두 개의 랜더링 방법을 가지게 된 SAOC는 복호화기의 선택한 필요한 출력 채널 수에 따라SAOC만으로 최종 출력 신호를 생성하는 Decoder Mode 와, MPEG Surround가 최종 오디오 신호를 출력하는 Transcoder Mode를 선택적으로 사용할 수 있다. 이러한 Operation Mode는 출력 채널의 수에 따라 아래 표 1과 같이 결정된다.

서론에서 언급한 것과 같이 복호화기의 입력으로 Rendering Matrix (사용자가 원하는 오디오 scene)을 갖는다는 점은 SAOC 가 여타의 오디오 부호화 방법과 다른 차별점이라 할 수 있다. Rendering Matrix 는 각 오디오 객체를 원하는 출력 채널에 매핑시키는 것에 관

표 1. SAOC 동작 모드
Table 1. SAOC Operation Mode.

SAOC module mode	Output signal config.	# of output channels	SAOC module output	MPS decoder required
Decoder	Mono / Stereo / Binaural	1 or 2	PCM output	No
Trans coder	Multi-channel	> 2	MPS bitstream, Downmix signal	Yes

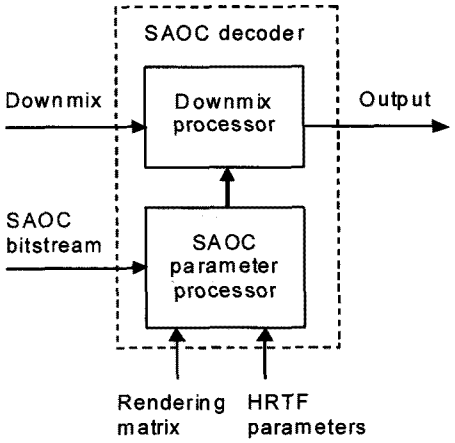


그림 4. SAOC 복호화기 (Decoder Mode)
Fig. 4. SAOC Decoder (Decoder Mode).

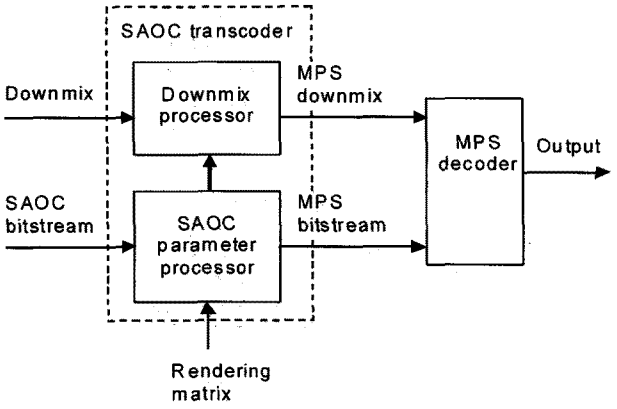


그림 5. SAOC 복호화기 (Transcoder Mode)
Fig. 5. SAOC Decoder (Transcoder Mode).

한 것으로, 다음과 같은 수식으로 정의될 수 있다.

$$\mathbf{M} = \begin{pmatrix} m_{1,Lf} & \cdots & m_{N,Lf} \\ m_{1,Rf} & \cdots & m_{N,Rf} \\ m_{1,Rf} & \cdots & m_{N,Rf} \\ m_{1,Lfe} & \cdots & m_{N,Lfe} \\ m_{1,Ls} & \cdots & m_{N,Ls} \\ m_{1,Rs} & \cdots & m_{N,Rs} \end{pmatrix} \tag{7}$$

여기서 $m_{i,ch}$ 는 i 번째 객체를 ch (Lf, Rf, … Rs)번째 출력 채널에 할당하는 게인을 나타내는 값이다.

SAOC가 Transcoder Mode로 동작할 때의 구조는 그림 5와 같다. Decoder Mode와 비교하면, 파라미터 프로세서가 MPEG Surround bitstream을 출력하고, Downmix processor가 최종 복호화된 오디오 신호가 아닌 MPEG Surround 입력이 되는 다운믹스 신호를 출력한다는 점이 크게 다르다.

3. Enhanced Mode

음악 Remix 응용 사례에서, Karaoke (보컬 object를 완전히 제거) 혹은 Solo (보컬을 제외한 나머지 object를 모두 제거) 는 중요한 기능이 될 수 있는데, SAOC의 parametric한 부가 정보만을 이용하여 이와 같은 극단적인 객체 제어를 하는 경우 성능 열화가 크다. 즉, 완전히 제거를 하는 경우 남아있는 오디오 신호의 왜곡이 매우 심하거나, 왜곡을 최소화하기 위해 제어 범위를 일정 수준에서 제한하는 것이 바람직하다. 이를 극복하기 위해 극단적인 제어를 요구하는 소수의 객체에 대해서는 기본 부가 정보 이외에 객체 신호 자체를 AAC를 이용한waveform 코덱으로 부호화한 데이터(residual)를 부가적으로 전송하고 이를 활용하여 SAOC 복호화의 성능을 향상시키는 모드를 개발하였다. 이렇게 residual을 포함한 오브젝트를 EAO (Enhanced Audio Object)라고 하며, 그림6은 EAO를 포함한 SAOC decoder/ transcoder의 구조를 나타낸다. EAO를 포함한 object는 일반적인 SAOC와는 다른 decoding 과정을 거치게 되는데, 그림에서 Residual processor는 이를 수행하는 block이다. 또한 Residual processor에 의한 처리 과정 이외에 일반적인 오브젝트 파라미터를 이용한 처리과정은 SAOC downmix pre-processor에서 수행되며, EAO와 일반적인 오브젝트에 대한 처리 결과는 object combiner에 의해 합성되어 최종적인 출력신호를 얻을 수 있다.

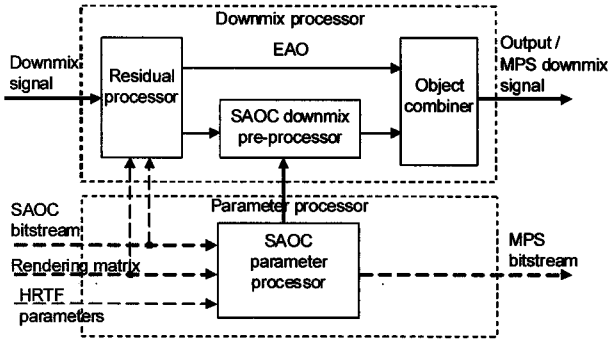


그림 6. Enhanced Audio Object (Residual)이 사용된 경우의 decoder 구조

Fig. 6. Decoder with Enhanced Audio Object (Residual).

Ⅳ. Rendering 관련 기술

앞서 언급된 것과 같이 SAOC는 지금까지의 오디오 부호화 방법들과는 다르게 복호화단에서 사용자로부터 Rendering Matrix를 입력 받아야 정상적인 복호화가 가능하다. 이렇게 표준화의 범위에 들어오게 된 Rendering Matrix의 사용에 있어서, 다양한 고려 사항이 발생하였다. 본 장에서는 관련한 몇 가지 이슈와 이의 해결을 위한 기술들을 다룬다.

1. 사용자 인터페이스

SAOC는 식 (7)과 같이 각 객체 신호와 출력 채널 신호간의 관계로 정의된 Rendering Matrix를 통해 사용자 입력을 이해한다. 그러나 음악 Remix를 비롯하여 실제 응용에서 이 값은 사용자에게 직관적이지 않다. 그림 7은 스테레오 음악 Remix 응용에서 사용될 수 있는 전형적인 사용자 인터페이스를 예시한다. 스튜디오의 믹싱 콘솔과 유사하게 각 객체별 게인을 제어하는 Level Fader와 각 객체의 스테레오 panning을 제어하는 Panning knob을 가지고 있다. 이와 같은 사용자 인터페이스로부터 입력되는 제어 정보 L_i (i번째 객체의 Level)와 θ_i (i번째 객체의 panning)를 Rendering Matrix의 계수 값 $m_{i,ch}$로 변환하는 과정이 필요한데, 이에 대한 SAOC의 가이드는 다음 식과 같다.

$$m_{i,ch_k} = 10^{0.05L_i}\sqrt{\frac{g_{i,ch_k}}{1+g_{i,ch_k}}}, \tag{8}$$

$$m_{i,ch_{k+1}} = 10^{0.05L_i}\sqrt{\frac{1}{1+g_{i,ch_k}}} \tag{9}$$

여기서 g_{i,ch_k}는 이웃하는 두 스피커 채널 간의 게인 비율로써 panning 정보 θ_i로부터 얻을 수 있다. 스테레오 출력인 경우 좌/우 두 채널 스피커에 각각 해당된다.

한편 실제 음악 레코딩 스튜디오에서 볼 수 있는 대

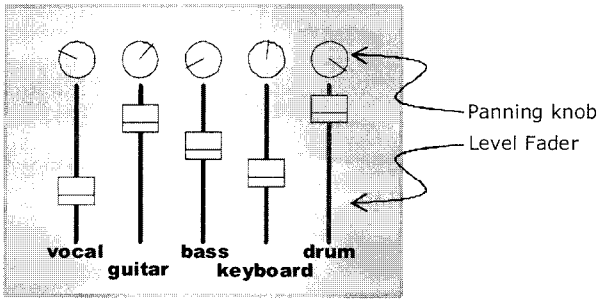


그림 7. 음악 Remix의 전형적 사용자 인터페이스

Fig. 7. Typical User Interface in Music Remix.

부분의 객체 신호 (멀티트랙)는 자체가 스테레오 객체이다. 즉, 드럼, 보컬 등 하나의 음원에 해당하는 객체가 좌, 우 2채널로 구성되어 있는 경우이다. SAOC에서는 이를 두 개의 다른 객체로 받아들이며, 다만 객체의 Metadata 에 수록된 정보 등을 참조하여 두 객체가 하나의 음원에 대한 좌, 우 트랙임을 알 수 있다. (좌 채널에 해당하는 객체는 스테레오 다운믹스 상에도 좌 채널에만 실릴 것이므로, DCLD가 +150dB로 나타날 것이다.) 사용자 인터페이스 관점에서는 이와 같은 스테레오 객체도 하나의 Level Fader 및 Panning Knob을 제어하는 것이 바람직할 것이다. 이 경우 하나의 사용자 입력이 두 개의 SAOC 객체에 대한 Rendering Matrix 계수를 제어하는 결론에 이르게 된다. 이를 수식적으로 나타내면 다음과 같다. 식에서 i번째, j번째 객체가 스테레오 객체의 좌/우 트랙을 각각 나타낸다고 가정한다.

$$m_{i,ch_k} = 10^{0.05L_i}\sqrt{\frac{g_{i,ch_k}}{1+g_{i,ch_k}}}, \tag{10}$$

$$m_{i,ch_n} = 0, \tag{11}$$

$$m_{j,ch_k} = 0, \tag{12}$$

$$m_{j,ch_n} = 10^{0.05L_i}\sqrt{\frac{1}{1+g_{i,ch_k}}}. \tag{13}$$

또한 그림7과 같은 인터페이스를 사용자에게 제공하기 위해서는 현재 컨텐츠에 대한 객체의 개수와 함께 해당 객체의 속성 (모노/스테레오 여부, 객체의 이름)을 알아야 한다. SAOC 표준에서는 이를 위한 Metadata를 표현하는 방법을 정의하고 있다.

Rendering Matrix 를 입력 받기 전에는 SAOC가 정상적인 복호화가 불가능하므로, SAOC 복호화 과정의 초기 상태를 위해서는 Rendering Matrix의 초기값을 정의해줄 필요가 있다. 가장 바람직한 초기값은 DMG와 DCLD로 얻을 수 있는 Downmix Matrix 일 것이다. 즉, 다운믹스 신호 그대로 출력하는 초기 재생 상태로 만들기 위해 Downmix Matrix를 Rendering Matrix 초기값으로 입력한다.

2. Preset

SAOC에서는 Predefined rendering information이라는 이름으로 Rendering Matrix를 전송 및 저장하기 위한 Preset 기술을 제공한다. Preset은 음악Remix에서는 컨텐츠 제공자가 직접 몇 개의 추천하는 mixing mode (예를 들어, Karaoke, Club mix, Acoustic mix, 등)를 하나의 비트열로 표현할 수 있는 강력한 기능이고, 방송에서는 방송국에서 무색 무취의 단순하게 마스터링된 신호의 송출이 아니라, Silver mode (대사/음성 중심의 mix), Cinema mode (효과음 극대화한 mix) 등 다양한 모드를 시청자 취향에 따라 선택할 수 있는 기능을 제공할 수 있다. 원격 회의에서는 회의 참여자들을 가상 공간에 자동으로 위치시키기 위한 정보의 전송에 활용 가능하다. 또한, 사용자가 본인만의 최적의 장면/모드로 마스터링한 Preset 정보를 제3자에게 전송하는 형태로도 활용이 가능한데, 이는 컨텐츠 자체의 불법적인 유통 및 공유 없이 Peer간에 자신만의 Remix 버전, Rendering scene을 효과적으로 공유할 수 있는 수단이다[22]. 이를 위해 SAOC에서는 Preset을 SAOC 비트열 내뿐만 아니라 별도의 독립된 파일 및 비트열 형태로 저장, 전송할 수 있도록 표준화 하였다.

Preset은 또한 하나의 SAOC 컨텐츠 내에서 고정된 값을 갖는 정적 Preset과 시간에 따라 가변할 수 있는 동적 Preset으로 구별할 수 있다. 음악 및 방송 컨텐츠에서 각 객체에 대해 보다 적극적인 실감 음장을 생성하기 위해서는 프레임에 따라 다른 장면을 구성할 수 있는 동적 Preset이 바람직하다.Video Conferencing에서 회의 참가자가 움직이며 발언을 하는 상황도 동적 Preset으로 잘 묘사할 수 있다.

3. DCU

Parametric 부호화의 한계를 고려하면 SAOC에서 객체 별 극단적인 rendering을 제약하는 것이 바람직한 경우가 있다. EAO가 없는 객체에 대해 Karaoke처럼 극단적인 감쇄를 실시할 경우, 왜곡이 심해져 의도한 효과를 기대할 수 없을 것이다. 이는 시장에서 SAOC를 이용해 컨텐츠 사업을 하려는 입장에서 볼 때, 품질에 대한 사용자의 심각한 불만 요소가 될 수 있으며, 따라서, SAOC의 시장 활성화에 제약 요인이 될 수 있다. 또한 음악 프로듀서나 영화 제작자 등 컨텐츠 원 저작자 입장에서 SAOC에 의해 컨텐츠가 지나치게 훼손되는 상황을 원치 않을 수 있다.

SAOC에서는 이처럼 음질을 왜곡할 만큼 극단적인 rendering에 대해서 음질이 보장되는 영역까지만 rendering이 될 수 있도록 자동적으로 완충 역할을 제

공할 수 있는 DCU (Distortion Control Unit) 기술을 표준에 포함하였다.

그림 8은 DCU가 적용된 SAOC 복호화기의 구조도를 나타낸다. DCU를 동작을 제어하기 위한 약간의 부가정보가 SAOC 비트열에 포함되어 전송되면, DCU는 Downmix Matrix (식 1)와 Rendering Matrix(그림에서 $\mathbf{M}_{ren}^{l,m}$)의 관계를 이용하여, 적당한 (왜곡이 없을 것이라고 판단되는) 범위 안에서 rendering이 이뤄질 수 있도록 Modified Rendering Matrix, $\mathbf{M}_{ren,lim}^{l,m}$를 생성한다. 계산된 Modified Rendering Matrix가 이후의 SAOC 복호화 과정에 적용됨으로써, SAOC는 사용자의 의도적 혹은 실수에 의한 극단적 요구로부터 음질이 보장되는 영역 안에서의 rendering처리를 보장할 수 있다. 다음 식은 $\mathbf{M}_{ren}^{l,m}$으로부터 $\mathbf{M}_{ren,lim}^{l,m}$을 계산해내는 DCU의 처리 과정을 나타낸다.

$$\mathbf{M}_{ren,lim}^{l,m} = (1-g_{DCU})\mathbf{M}_{ren}^{l,m} + g_{DCU}\mathbf{M}_{ren,tar}^{l,m} \tag{14}$$

여기서 $g_{DCU} \in [0,1]$는 SAOC 비트열로 전송되는 DCU의 강도를 제어하는 값으로, 0이면 DCU가 작동하지 않음을 의미하고, 1인 경우 극단적인 통제를 수행됨을 의미한다. 또한 $\mathbf{M}_{ren,tar}^{l,m}$는

$$\mathbf{M}_{ren,tar}^{l,m} = \sqrt{\mathbf{N}_{BE}^{l}}\mathbf{D}^{l} \tag{15}$$

과 같이 정의되는데, 즉, Downmix Matrix $\mathbf{D}^{l}$를 계산된 크기로 변형하여 생성한 DCU의 목표 값이 된다.

한편, DCU에 의해 Rendering Matrix가 변형된 경우, 이를 사용자에게 통지할 수 있는 인터페이스를 가지는 것이 바람직하다. Remix의 경우 그림 7에 예시된 것과 같은 Mixer GUI를 입력장치로 사용할 수 있는데, $\mathbf{M}_{ren,lim}^{l,m}$를 Mixer GUI로 출력하여 사용하게 되면, 사용자가 제어한 rendering과 실제 SAOC 처리되어 재생되는 소리가 불일치하는 문제를 해결할 수 있다. 즉, DCU를 사용하면서도 WYSIWYG (What You See Is What You Get)을 보존할 수 있다.

일반적으로 Rendering의 정도와 왜곡의 정도 간의 관계는 객체 간의 유사성에 따라 다르다. 즉, 바이올린과 첼로 합주로부터 바이올린을 추출하는 경우와 바이올린과 타악기가 섞인 신호에서 바이올린을 추출하는 난이도는 다를 수 있다. 또한, EAO가 포함된 SAOC에서는 EAO에 대한 제어 시 왜곡과 다른 객체에 대한 제어 시 왜곡 정도가 다를 수 있다. 현재 SAOC에 적용된 DCU는 이와 같이 객체 별 특성 및 객체간의 관계는 고려하지 않는다. 이와 같은 점을 고려하기 위해서는 개선된 형태의DCU를 연구해볼 필요가 있다.

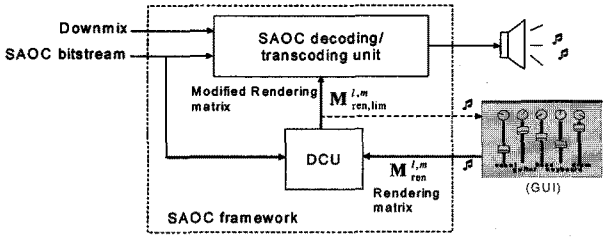


그림 8. SAOC의Distortion Control Unit 적용
Fig. 8. Distortion Control Unit (DCU) in SAOC framework.

V. 음질 평가

오디오 부호화 기술의 가장 중요한 성능 평가 항목인 음질 평가와 관련해서도 SAOC의 표준화 과정은 몇 가지 기술적 이슈를 제기했다. 수신단에서 제공되는 값인 Rendering Matrix에 의한 객체 별 제어의 특성으로 SAOC에 대한 음질 평가는 원신호로부터의 상대적 왜곡에 대한 평가뿐 아니라, 의도한 Rendering이 얼마나 정확하게 이뤄졌는지에 대한 평가를 함께 고려할 수 있는 새로운 음질 평가 방법의 필요성이 제기 되었다[23]. 또한, 평가의 Reference가 되는 원음이 부호화기의 입력으로 사용되는 신호와는 다른 점, DCU 사용 시 Rendering의 약화에 따라 평가 신호의 loudness가 달라진 점 등도 이슈가 되었다. 그러나 표준화 과정에서는 이와 같은 이슈를 음질 평가자의 주관에 맡기고 기존과 마찬가지로 MUSHRA (MUltiple Stimulus Hidden Reference and Anchor) 평가 방법을 사용하였다[24]. SAOC에 대한 표준화가 완료된 후, Verification Report를 위해 상용 부호화 시스템과 성능을 비교하는 공식 음질 평가가 진행 중이며, 현재까지 정리된 주요 결과는 그림 9 및 10과 같다[24~25].

각 응용 시나리오 별로 기존의 부호화 기술과 비교하는 주관적 음질 평가를 수행하였으며, 각 평가 별로 Fraunhofer IIS, Dolby, Philips, LG전자, ETRI 등에서 30여명의 전문가가 참여하였다. 그림 9의 결과는 음악 Remix 경우를 염두한 음질평가 결과로써, 결과에 나타난 각 System의 의미는 다음과 같다.

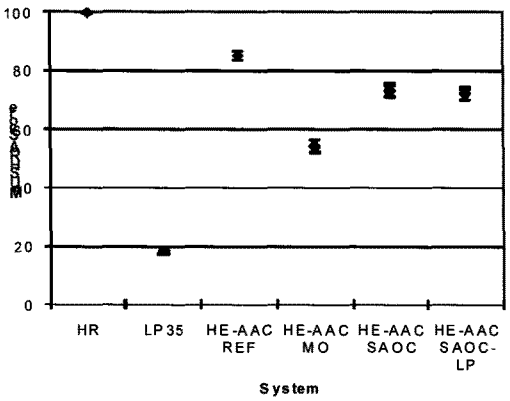


그림 9. SAOC Verification 평가 결과 (1)
- 음악 Remix[26]

Fig. 9. SAOC Verification Test Result (1)
- Remix Application[26].

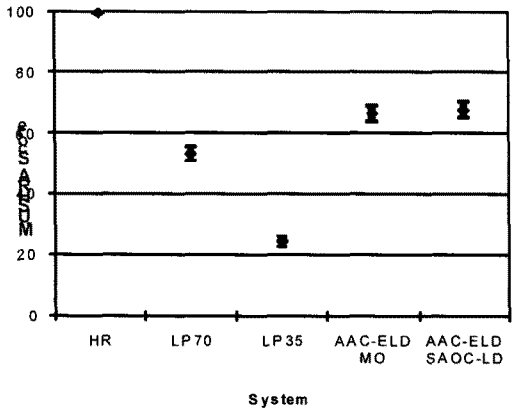


그림 10. SAOC Verification 평가 결과 (2)
- 다채널 원격회의[26]

Fig. 10. SAOC Verification Test Result (2)
- Multichannel Teleconferencing[26].

- HR: Hidden Reference (각 객체 원음 신호를 Rendering Matrix에 따라 Mix한 이상적 원음신호)
- LP35: HR 신호를 3.5kHz 저역 통과한 신호
- HE-AAC REF: HR를 64kbps 의 HE-AAC로 부호화한 신호 (다운믹스 코덱 자체의 왜곡 척도 확인 목적, 실제 응용에서 얻을 수 없는 신호)
- HE-AAC MO: 총 64kbps의 비트율 안에서 각 객체를 독립적으로 부호화하여 전송할 때, 이를 이용하여 생성할 수 있는 신호
- HE-AAC SAOC: 총 64kbps 비트율로 SAOC 부호화된 신호를 통해 생성한 신호
- HE-AAC SAOC-LP: HE-AAC SAOC와 동일한 비트열을 Low Power Mode로 복호화한 신호

그림 9의 결과로부터 HE-AAC SAOC 및 HE-AAC SAOC-LP를 HE-AAC MO 와 비교함으로써, 동일한 비트율(64kbps)에서 기존 기술대비 MUSHRA 절대 점수 약 20점 정도의 성능향상을 기대할 수 있음을 알 수 있다. 한편 그림 10은 원격회의 응용의 경우를 위한 음질 평가 결과이다. 여기서 사용된 System에 대한 설명은 다음과 같다.

- HR: Hidden Reference (각 객체 원음 신호를 Rendering Matrix에 따라 Mix한 이상적 원음신호)
- LP70: HR를 7.0kHz 저역 통과한 신호
- LP35: HR를 3.5kHz 저역 통과한 신호
- AAC-ELD MO: 총 94kbps 의 비트율 안에서 각 객체(4개의 음성신호)를 독립적으로 부호화하여 전송할 때, 이를 이용해 생성한 신호
- AAC-ELD SAOC-LD: 총 40kbps의 비트율로 SAOC-LD (Low Delay ModE)로 부호화된 신호를 통해 생성한 신호

그림 10의 결과는 SAOC-LD와 AAC-ELD (다운믹스 코덱)를 이용하게 되면 2배 이상 비트를 절약한 40kbps에서 SAOC를 이용하여 다채널 rendering된 원격회의를 실현할 수 있음을 시사한다.

VI. 결 론

최근 MPEG 에서 표준화를 완료한 객체 오디오 부호화 기술인 SAOC에 대해 살펴보았다. SAOC는 MPEG 오디오의 기술적 유행이 되고 있는 Parametric 부호화 방법을 이용하며, 특히 MPEG-4의 핵심 가치이던 객체 기반 부호화를 실현한 유일한 오디오 부호화 표준이라는데 큰 의의를 지닌다. 또한, 하향 호환성을 제공은 다양한 응용 분야에 활용을 기대할 수 있다. 기존의 오디오 표준과 같이 부호화 및 복호화 방법만을 정의한 것이 아니라, 오디오 장면의 재구성 과정과 이를 위한 Rendering Matrix및 관련 기술 Tool들을 표준 범주에 포함한 점, MPEG Surround 기술과의 연동을 통한 다채널 지원 구조 제공 등은 표준화 자체로도 기존과 다른 요소를 고려한 점에서 큰 의의가 있다고 하겠다. 한편, 표준화 과정에서는 음질평가 방법 및 결과에 대해서도 많은 이슈가 제기되었는데, 이에 대한 고민은 향후 중요한 연구주제가 될 수 있다. Parametric 부호화의

특성상 극단적 제어 성능에서의 한계가 있지만, 매우 낮은 비트율로 부호화 가능하고, 이동 기기에서 실시간 구현 가능한 연산 복잡성을 가진 점을 고려할 때, 가까운 미래에 다양한 응용 분야에서의 활용이 기대되는 차세대 오디오 기술이다.

참 고 문 헌

[1] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N4668, "MPEG-4 Overview," Jeju, Mar. 2002.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N2431, "MPEG Audio FAQ Version 9," Atlantic City, Oct. 1998.

[3] ISO/IEC Int. Std. 14496-11:2005, Information technology - Coding of audio-visual objects - Part 11: Scene description and application engine, 2006.

[4] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N7608, "MPEG-4 BIFS white paper," Nice, Oct. 2005.

[5] ISO/IEC Int. Std. 11172-3:1993, Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s - Part 3: Audio,, 1993.

[6] ISO/IEC Int. Std. 13818-7:1997, Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC), 1997.

[7] ISO/IEC Int. Std. 14496-3:1999, Information technology - Coding of audio-visual objects - Part 3: Audio, 1999.

[8] M. Dietz, L. Liljeryd, K. Kjorling, O. Kunz, "Spectral Band Replication, a Novel Approach in Audio Coding," in AES 112th convention, preprint 5553, Apr., 2002.

[9] E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegard, "Low Complexity Parametric Stereo Coding," in AES 116th convention, preprint 6073, May, 2004.

[10] ISO/IEC Int. Std. 14496-3, Amd.1:2003, 2003.

[11] ISO/IEC Int. Std. 14496-3, Amd.2:2004, 2004.

[12] ISO/IEC Int. Std. 23003-1:2007, MPEG audio technologies - Part 1: MPEG Surround, 2007.

[13] J. Herre and et al, "MPEG Surround-The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding, J. Audio Eng. Soc., Vol. 56, No. 11, Nov., 2008.

[14] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. M13632, "From Channel-Oriented to Object-Oriented Spatial Audio Coding," Klagenfurt, July 2006.

[15] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N11207, "ISO/IEC FDIS 23003-2: 2010, Spatial Audio Object Coding," Kyoto, Jan. 2010.

[16] F. Baumgarte and C. Faller, "Binaural Cue Coding - Part I: Psychoacoustic fundamentals and design principles," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003.

[17] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003.

[18] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N8853, "Final Call for Proposals on Spatial Audio Object Coding," Morocco, January 2007.

[19] J. Engdegard and et al, "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding," in AES 124th convention, preprint 7377, May, 2008.

[20] 정양원, 오현오, "오디오 객체 부호화 표준 - MPEG SAOC," 한국음향학회지, 제28권, 제7호, 630-639쪽, 2009년 10월

[21] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. M14159, "Comments on Draft Call for Proposals on Spatial Audio Object Coding," Marrakech, Jan., 2007.

[22] Yang-Won Jung and Hyen-O Oh, "Personalized Music Service Based on Parametric Object Oriented Spatial Audio Coding," in Proc. of The 34th AES International Conference, Aug. 2008.

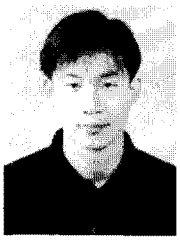
[23] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. M14735, "Comments on SAOC evaluation," Lausanne, Jul., 2007.

[24] RECOMMENDATION ITU-R BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems", 2003.

[25] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N11298, "Workplan for SAOC Verification Test," Dresden, Apr., 2010.

[26] ISO/IEC JTC1/SC29/WG11 (MPEG) Doc. N11507, "Draft SAOC Verification Test Report," Geneva, Jul., 2010.

저 자 소 개



오 현 오(정회원)-교신저자

1996년 연세대학교 전자공학과 학사.

1998년 연세대학교 전자공학과 석사.

2002년 연세대학교 전기전자공학과 박사.

2002년~현재 LG전자 Digital TV연구소 책임연구원

<주관심 분야: 오디오 신호처리, 심리음향, 오디오/음성 코덱 표준화>

정 양 원(정회원)

1998년 연세대학교 전자공학과 학사.

2000년 연세대학교 전기컴퓨터공학과 석사.

2005년 연세대학교 전기전자공학과 박사.

2005년~2009년 LG전자 Digital TV연구소 책임연구원

2009년~현재 인텔렉추얼 벤처스 코리아 이사

<주관심분야: 오디오 신호처리, 오디오/음성 코덱 표준화>