

고차 다채널 실감 오디오 서비스를 위한 MPEG 3D Audio 표준화 동향

서정일*, 강경옥*, 정대권**

*한국전자통신연구원, **한국항공대학교

seoji@etri.re.kr, kokang@etri.re.kr, dgjeong@kau.ac.kr

Overview of MPEG 3D Audio Standard Activities for High-Order Multichannel Realistic Audio Service

Jeongil Seo*, Kyeongok Kang*, Dae-Gwon Jeong**

*Electronics and Telecommunications Research Institute (ETRI)

**Korea Aerospace University

요 약

본 논문에서는 최근 MPEG 오디오 서브그룹에서 활발히 논의 중인 3D Audio 표준화 동향에 대해서 소개하고, 관련한 국내외 기관들의 기술개발 현황에 대해서 알아본다. MPEG 3D Audio는 NHK 22.2 채널 방송과 같은 실감 오디오 서비스를 고다채널(High-Order Multichannel)로 특징짓고, 이러한 서비스를 위한 다채널 오디오 부호화 및 복호화 기술과 다양한 출력채널 환경에 적용할 수 있는 렌더링(rendering) 기술을 표준화 대상으로 규정하고 있다.

1. 서론

최근 MPEG에서는 NHK SHV(Super Hi-Vision) [1]과 같은 초고해상도(4K/8K) 대화면 디스플레이(100인치 이상)와 고차 다채널(22.2 채널 등)을 이용하는 실감 몰입형(immersive) 멀티미디어 서비스를 위한 표준화를 MPEG-H란 프로젝트 명으로 진행 중이다. MPEG-H의 part 2는 초고해상도 비디오를 위한 새로운 부호화 기술로서 ITU-T와 공동 설립한 JCTVC(Joint Collaboration Team Video Coding)팀에서 HEVC(High Efficiency Video Coding)란 이름으로 표준화가 진행 중이며, 2012년 2월 CD(Committee Draft)가 발간되었다[2]. MPEG-H의 part 1은 대용량의 비디오 및 오디오 비트스트림을 효율적으로 다중화 또는 패키징하기 위한 시스템 규격인 MMT(MPEG Media Transport)로 할당될 예정이며, MMT는 2012년 7월 CD 발간을 목표로 기술적 논의가 한창 진행 중이다[3]. 오디오 신호에 대한 표준은 part 3에 할당될 예정이며 세 개의 파트 중 가장 늦게 표준화에 대한 논의가 3D Audio란 이름으로 시작되어 2012년 7월 CFP(Call For Proposal)가 공지될 예정이다.

본 논문에서는 2012년 4월 100차 MPEG 회의까지 논의된 결과를 바탕으로 3D Audio 표준화 활동 전반에 대한 개요와 핵심 표준화 대상기술에 대한 소개를 담고자 한다[4][5].

2. MPEG-H 3D Audio

MPEG과 ITU-T가 공동으로 설립한 JCTVC에서 HEVC

표준화가 시작된 2010년 하반기에 HEVC가 목표로 하는 대화면 고화질 비디오와 부합하는 오디오 서비스에 대한 기술적 검토가 시작되었으며, NHK의 22.2 채널 오디오 시스템에 대한 소개와 함께 3D Audio란 이름의 AhG(Adhoc Group)이 Audio 서브그룹에 설립되었다. 당시 NHK에서는 자사의 22.2 채널 오디오 시스템의 국제표준 반영을 위하여 SMPTE, ITU-R, ATSC 등과 같은 다양한 국제표준화 단체에 새로운 오디오 표준에 대한 필요성을 제시하던 시점이었으며, 디지털 시네마 분야에서는 DTS, Dolby 등에서 기존 5.1/7.1 채널 서라운드 오디오 시스템을 획기적으로 개선할 수 있는 3차원 오디오 포맷에 대한 연구 및 기술개발이 활발히 진행되고 있었다. 또한, 국내에서는 차방포럼을 중심으로 UHDTV를 위한 비디오 및 오디오 신호규격을 제정하기 위한 연구가 진행되고 있었다.

상기와 같은 국내외 기관들의 차세대 오디오 서비스에 대한 연구/개발/표준화 활동과 보조를 맞추어 3D Audio 표준화에 대한 활발한 논의가 벌어졌으며 2012년 7월로 예정된 CFP 발표를 기점으로 본격적인 표준화가 시작될 것이다. 본 절에서는 2012년 4월 MPEG 회의까지 논의된 3D Audio 표준화 결과를 요약하고, 표준화 대상이 될 핵심 기술들을 도출하고자 한다.

2.1 Application Scenario

3D Audio가 목표로 하는 응용분야들은 아래와 같다.

- Home Theater: HEVC가 목표로 하는 대화면 고해상도 디스플레이에 부합할 수 있도록 함께 제공되는 오디오는 고음질을 가지면서도 안정적인 넓은 음장영역을 표현할 수 있어야 한다. 특히 비디오 이미지와 오디오 이미지가

일치될 수 있도록 정확한 음성재현 성능 및 음상의 고저감을 표현할 수 있어야 한다.

- Personal 3DTV: 22.2 채널과 같은 다채널로 대표되는 홈씨어터 뿐만 아니라 동일한 오디오 콘텐츠를 테블릿 PC 와 같은 소형 개인단말에서 헤드폰이나 스피커 어레이를 이용하여 원음장과 유사한 음질과 음장감을 제공할 수 있어야 한다.
- TV for Smart Phone: 헤드폰을 이용하여 3 차원 오디오 콘텐츠를 원음장과 유사한 음질과 음장감으로 재생할 수 있어야 한다.
- Multichannel Audio Program: 오디오 CD 와 같이 오디오 단독으로 다채널 재생환경을 이용하여 실감있고 몰입감있는 오디오 서비스를 제공하여야 한다.
- Telepresence: 원격회의 시스템, 가상현실 시스템 등과 같이 네트워크를 이용하여 원격에서 가상공간을 재현할 때 오디오도 3 차원 공간상에서 재현되어야 한다.

2.2 Requirements

2.1 절에 나열된 응용분야에 3D Audio 표준이 사용되기 위해서 만족해야 할 요구사항들이 논의되고 있다. 3D Audio 에서는 요구사항들을 주요(primary) 요구사항, 부가(secondary) 요구사항, 바람직한 특징(desirable feature) 세가지로 구분하여 정의하고 있다. 본 논문에서 제시되는 요구사항은 2012년 4 월까지 논의된 결과이며, 7 월 회의때 확정될 예정이므로 변경되거나 삭제될 수 있으며 새로운 요구사항이 추가될 가능성도 있다.

먼저 표준화에서 중요하게 다루어질 주요 요구사항(Primary Requirements)은 다음과 같다.

- High quality: 22.2 채널과 같은 고다채널 환경에서 높은 음질(perceptual transparent quality)을 제공하여야 한다.
- Localization and envelopment: 거리감을 포함하여 렌더링 되는 음상의 정확한 표현과 풍부한 음장 포위감(envelopment)을 제공하여야 한다.
- Rendering on setups with fewer loudspeakers: 입력 오디오 신호의 채널 수보다 적은 수의 채널로 재생/렌더링될 때에도 최대한 가능한 음질과 음장감을 제공하여야 한다.
- Flexible loudspeaker placement: 입력 오디오 신호의 채널배치와 다른 위치에 스피커가 위치할 때에도 최대한의 음질과 음장감을 제공하여야 한다.
- Audio/visual alignment & consistency: 비디오와 오디오 객체의 재현되는 공간상의 위치가 동일하여야 한다.
- Latency: 생방송과 같은 환경을 지원하기 위해 인코더/디코더에 의한 지연이 적어야 한다.
- Audio program inputs to the submitted encoding systems: 인코더로 입력되는 오디오 신호의 형태는 채널신호, 오디오 객체신호, High Order Ambisonics (HOA)와 같은 오디오 장면(Audio Scene) 신호가 될 수 있다.
- Transcoding for low bandwidth devices: 저대역(낮은 음질과 공간해상도) 단말을 위하여 오디오 비트스트림으로 부터 저대역 오디오 신호를 적절한

복잡도로 추출할 수 있어야 한다.

RM0(Reference Model 0) 선정단계에서는 고려하지 않았지만 다음으로 중요하게 다루길 요구사항(Secondary Requirements)은 다음과 같다.

- Efficiency for decoding on different setups: 입력 오디오 신호와 상이한 오디오 출력환경에서도 효율적으로 변환하여 적절한 음질을 제공하여야 한다.
- Computational complexity: 여러가지 응용환경에서 적절한 복잡도를 가져야 한다.
- Issues of backward compatibility: 기존 5.1 채널 재생환경에서도 변환하여 재생할 수 있어야 한다.
- Interactivity: 오디오 객체를 제어하는 것과 같이 대화형 오디오 서비스에서의 interactivity 를 제공할 수 있어야 한다.

마지막으로 3D Audio 코덱의 장점으로 부각될 수 있는 특징(Desirable Features)들은 아래와 같다.

- Compressed representation that allows extraction of lower-bitrate lower quality representations: 비트스트림 레벨에서 낮은 비트율과 낮은 음질의 신호를 손쉽게 추출할 수 있어야 한다.
- Low latency and low computational complexity: 양방향 통신환경이나 Tele-presence 와 같은 응용환경을 위하여 적은 지연과 낮은 복잡도를 가져야 한다.

2.3 표준화대상 기술분야

오는 7 월 공지될 CFP 에 응대하여 여러 기관에서 제출된 기술 제안서의 성능을 평가하기 위한 절차와 기준을 논의 중에 있으며, 현재까지 합의된 개괄적인 시스템 구성도는 그림 1 과 같다.

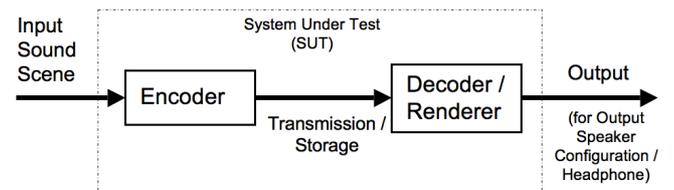


그림 1. MPEG 3D Audio 인코더/디코더/렌더러 시스템 구성도

그림 1 에서 정의된 용어들에 대한 정의는 아래와 같다.

- SUT(System Under Test): 3D Audio 콘텐츠를 특정 비트율에서 인코딩/디코딩/렌더링 하는 시스템을 의미하며, CFP 에 대응하여 제안된 개별적인 시스템들을 지칭한다.
 - ISC(Input Sound Scene): SUT 로 입력되는 오디오 신호를 의미하며 채널신호, 오디오 객체신호, 오디오 장면신호가 될 수 있다. 또한 세가지 신호들이 조합된 형태로도 입력이 가능하다.
 - OSC(Output Speaker Configuration): SUT 가 출력하는 오디오 신호를 의미하며 ISC 와 상이한 형태도 가능하다.
- 상기와 같은 형태로 오디오 코덱이 구성된다면 2.2 절에서 제시한 요구사항을 만족시키기 위해서는 아래와 같은 기능블록

들이 필요하다고 예측할 수 있다.

- 개별채널 신호 또는 객체신호를 고효율/고음질로 압축하기 위한 모노 또는 스테레오 코어코덱: 전통적으로 MPEG에서는 MPEG 표준기술을 활용하는 것을 권고하기 때문에 MPEG-4 HE-AAC v2 나 MPEG-D USAC이 활용될 가능성이 높다.
- 스케일러블 채널구조: 방송환경과 같이 제한된 대역폭에서 다양한 포맷(예: 22.2, 10.2, 5.1)을 지원하여야 할 경우 스케일러블 채널형태로 비트스트림을 구성할 수 있으며, 이로부터 발생할 수 있는 음질저하를 상쇄시킬 수 있는 기술이 필요하다.
- 채널신호와 객체신호를 모두 제어할 수 있는 Hybrid 코덱 구조: 객체신호만으로 제공되는 오디오 서비스는 일반적이지 않지만 객체신호가 채널신호와 공존하는 형태로의 서비스는 DTS와 Dolby에서 제안하고 있다. 따라서, 채널신호와 객체신호를 모두 제어할 수 있는 Hybrid 구조가 3D Audio 표준에 채택될 가능성이 높다.
- 다양한 출력채널(OSC) 환경과 스피커의 배치환경에 적용적인 렌더러 기술: 렌더러가 표준의 범주에 포함되는 것에 대해서 공감대가 형성되고 있으며, 코덱자체 보다는 렌더러가 3D Audio 표준에서 중요한 부분을 차지할 것으로 예측된다. 또한, 디코더와 렌더러가 분리된 형태가 아닌 결합된 형태가 될 것으로 예측되며, MPEG Surround 나 SAOC에서 표준화된 다양한 채널변환 툴들이 3D Audio 표준에서도 채택될 가능성이 높다.

3. 결론

MPEG에서 최근 표준화가 논의 중인 3D Audio는 기존의 MPEG 오디오 표준이 오디오 신호에 대한 압축 및 부호화에 초점이 맞추어진 것에 비하여 복원된 멀티채널 신호로부터 단말의 재생환경에 적절하도록 변형 또는 렌더링하는 방식에 초점이 맞추어질 전망이다. 이러한 렌더링방식은 표준에서 정하기 보다는 제조사나 사용자가 적절한 렌더링 기술을 선택하여 적용하는 후처리 기술의 범주에 있으므로, 표준화의 대상이 될 수 있는지에 대한 논란도 있다. 그러나 표준이 단말의 모든 오디오 재생환경에 맞출 수는 없지만 범용적인 렌더링 도구를 제공할 수 있다면 시장에서 받아들여 질 가능성도 높다고 할 수 있다.

비디오에 비해 3차원 오디오에 대한 출력포맷이 단일 규격으로 정해지기 어려운 현실을 감안할때 3D Audio의 표준화 접근방법이 설득력을 가진다고 볼 수 있으며, 표준이 완료되고 5.1 채널 이상의 3차원 오디오 재현환경이 부각될 시점을 위한 준비단계로서 상기 표준화는 큰 의미를 가진다고 할 수 있다.

감사의 글

본 논문은 2011년도 정부(한국방송통신위원회) 재원으로 방송통신기술개발사업의 지원을 받아 수행된 연구임(과제번호: KCA-2011-11921-02001).

참고문헌

- [1] K. Hamasaki, S. Komiyama, K. Hiyama, H. Okubo, "5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System," 2005 NAB BEC Proceedings, April 2005.
- [2] ISO/IEC JTC1/SC29/WG11, N12556, "Text of ISO/IEC CD 23008-2 High Efficiency Video Coding," Feb. 2012.
- [3] ISO/IEC JTC1/SC29/WG11, N12690, "Working Draft of MPEG Media Transport," April 2012.
- [4] ISO/IEC JTC1/SC29/WG11, N12811, "Draft Call for Proposals for 3D Audio," April 2012.
- [5] ISO/IEC JTC1/SC29/WG11, N12610, "Draft Use Cases, Requirements and Evaluation Procedures for 3D Audio," Feb. 2012.