

MPEG-H 3D Audio의 배경 및 진행과정

MPEG 표준화 현황

오디오 코덱 기술은 음성/오디오 신호를 압축전송하기 위한 방식으로 개발되어 왔다. 한정된 전송 대역폭을 활용하여 음성 및 오디오 신호를 전송하기 위해서는 원본 신호의 압축이 필요했으며, 음질의 손상을 최소화하는 방안에 대해서 깊은 연구가 진행되었고, 이를 오디오 압축 기술에 적용하여 효과적인 오디오 데이터를 전송할 수 있게 되었다. 이러한 방식으로 개발된 기술들이 MP3 혹은 AAC가 되겠다. 그러나 코덱 기술의 진화와 더불어 전송 기술 및 환경도 개선됨에 따라, 모노/스테레오와 같은 통상적인 오디오 포맷에 대한 전송 문제는 더 이상 큰 기술적 이슈를 제기할 수 없었으며, 충분한 대역폭에서 기존의 기술로도 충분히 고품질의 오디오 서비스를 실현할 수 있게 되었다.

오디오는 비디오와 달리, 고해상도를 지원한다 하더라도 데이터양이 크게 증가하지 않을뿐더러 음질 개선 여부를 쉽게 판단하기 힘들다. 따라서 새로운 압축방식에 대한 요구사항이 비디오에 비하여 낮을 수밖에 없다. 이러한 한계를 극복하고 차별화된 새로운 오디오 서비스를 추구하기 위해서 다양한 시도들이 진행되어 왔다. 일차적으로 차별화된 오디오 서비스를 제공하기 위하여 고려된 사항은 멀티채널 오디오 서비스이다. 이는 기본적인 모노/스테레오 신호를 벗어나, 5.1/7.1 등 수평면에서 표현될 수 있는 오디오 장면에 대해서 오디오 채널신호로 표현하고 이를 전송하는 다채널 오디오 압축 기술이다. Dolby 사의 5.1 surround 기술이나, MPEG의 MPEG Surround 기술이 이에 해당되는 기술로, 2000년대 초반에 활발히 연구 개발되었으며 현재 상용 솔루션이 가정 혹은 극장환경에 활용되고 있다. 이러한 서라운드 기술은 사용자 단말환경에 따라 선택적으로 5.1 서라운드 오디오 신호를 재생할 수 있으며 오디오 장면의 표현도 전방에 국한되지 않고 후방까지 표현할 수 있는 기술이다.

Codec	Bitrate / ch	Feature
CD	1.4Mbps / stereo	
MP3	192Kbps / stereo	Compression
AAC	128Kbps / stereo	
HE-AAC v1	96Kbps / stereo	
HE-AAC v2	64Kbps / stereo	
MPEG-S	128Kbps / 5.1	Compression + Functionality
USAC	8~96Kbps / stereo	
MPEG-H 3D Audio	256k ~ 1.2Mbps / 24	

표 1. 오디오 코덱 종류

이후 차별화된 오디오 서비스를 위하여 객체기반 오디오 서비스를 지원하기 위한 오디오 코덱 기술이 개발되었다. 오디오 객체 신호는 단순히 오디오 신호를 듣는 것이 아니라, 사용자의 취향 혹은 요구에 맞추어 오디오 장면을 달리할 수 있는 기능을 제공하기 위하여 오디오 신호가 분리되어 제어 가능한 포맷으로 전달되는 오디오 신호이다. 쉽게 생각해 볼 수 있는 서비스는, 음성 등을 분리하여 오디오 객체로 전송한다면 드라마 및 영화 시청 중에 배우의 음성을 사용자 가 조절하여 들을 수 있는 서비스나, 스포츠 중계에서 아나운서 혹은

해설가의 목소리를 따로 볼륨 조절하여 시청이 가능한 서비스를 생각할 수 있겠다. 이러한 객체기반 오디오 서비스를 지원하기 위한 코덱 기술로 MPEG의 SAOC(Spatial Audio Object Coding) 기술이 2010년도에 표준화가 완료되어 공개되었다.

또 다른 기능적 요구사항으로, MPEG에서는 음성과 음악신호를 동일한 코덱으로 압축 전송할 수 있는 기술에 대한 개발 필요성이 야기되었다. 이는 기존의 오디오 코덱들이 음악에 대해서는 높은 성능을 나타내나, 음성에 대해서는 상대적으로 떨어지는 음질 성능을 나타냄으로써 하나의 단일 오디오 코덱으로 음성과 음악을 모두 우수한 품질로 압축 전송할 수 있는 기술이 존재하지 않았기 때문이다. MPEG USAC(Unified Speech and Audio Coding) 기술은 이를 충족시키기 위해 개발된 기술로 2012년도에 표준화를 완료하였으며, 실제로 음성/음악에 대해서 기존의 음성 코덱 및 오디오 코덱 대비 뛰어난 성능을 나타내는 기술로 검증되었다.

이렇게 MPEG에서는 2012년도까지 다채널 및 오디오 객체에 대한 기능적인 부분, 음성/음악에 통합 고품질을 제공하는 압축 성능 등을 충족시킬 수 있는 모든 기술들을 표준화하였다. 더 이상 오디오 압축 기술을 위하여 개발할 수 있는 새로운 기술들이 포함된 상태였으며, 새로운 표준화를 이끌고 갈 아이템과 동력을 찾기 힘들어 보였다. 그러던 와중에, NHK에서는 새로운 기술 개발에 대한 필요성을 MPEG에 개진하였으며, 이를 표준화 아이템으로 선정하고 진행하기를 희망하였다. 이는 NHK가 당시 개발 중이었던 SHV(Super Hi-Vision)에 적용하기 위한 오디오 시스템의 형상을 새롭게 제안하였으며, 이를 위하여 새로운 오디오 코덱 기술이 필요함을 MPEG 오디오 그룹에 의견을 개진하였다. NHK가 제안하는 오디오 시스템의 형상은 22.2 채널 오디오 시스템으로 상층, 중층 및 하층의 3개의 계층으로 22개의 스피커들을 3차원 공간상에 배치된 형상이다. 특히 대화면 디스플레이가 위치하는 전방에는 스피커들을 촘촘히 배치하여(총 11채널) 대화면 UHD 영상과 공간상의 동일한 위치에 음원들이 재현되도록 하였다. 또한 2개의 서브우퍼 채널을 이용하여 좌우가 분리된 풍부한 저음역대를 재생할 수 있다.

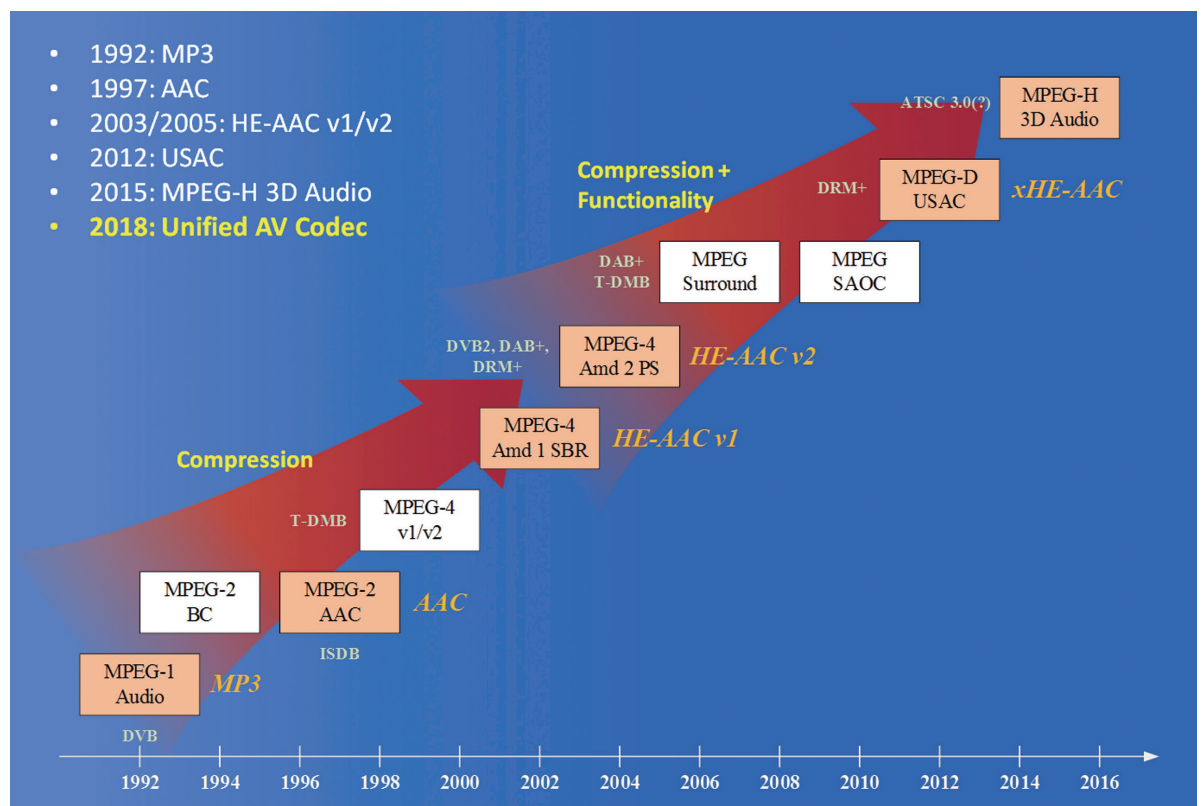


그림 1. MPEG Audio 표준화 동향

이러한 오디오 시스템의 형상은 기존의 다채널 오디오 포맷으로 널리 활용되던 5.1/7.1 등과 차별화된 것으로 상층과 하층을 두어 평면상의 오디오 장면 표현뿐만 아니라 상/하 전방위에 대한 오디오 장면 표현이 가능한 포맷이다. MPEG에서는 이러한 요구사항이 자극제가 되어 새로운 오디오 코덱 표준화 아이템을 발굴하기 위하여 세계 각 기관의 정보를 받아들이며 구체적인 요구사항을 도출하였다. 도출된 요구사항의 궁

극적인 목표는 이머시브 오디오(impressive audio) 코덱 기술 개발로, 언제 어디서나 상황에 맞는 최상의 이머시브 오디오 서비스를 제공하고 자 한다는 것이다. 이를 MPEG-H 3DA (3D Audio) 명칭 하에 표준화를 추진하였다. 이상 살펴본 MPEG 오디오 표준화 활동내용은 [그림 1]과 같이 정리하였다. 정리된 [그림 1]에서 알 수 있듯이, 최근 오디오 코덱 기술은 압축뿐만 아니라 다양한 기능을 제공하기 위한 형태로 발전되어 왔으며, 현재 MPEG-H 3DA가 압축과 기능 측면에서 가장 최신의 오디오 코덱이라 할 수 있겠다.

MPEG-H 3DA 기술 개요

그럼 본격적으로 MPEG-H 3DA에 대해서 설명하기로 하겠다. MPEG-H 3DA는 최상의 이머시브 오디오 서비스를 달성하기 위하여 입력신호의 규격을 다음과 같이 세 가지 형태로 지원할 수 있도록 정의하였다.

고차 다채널 오디오(High-order Multichannel Audio) : NHK 22.2채널과 같이 상층 채널(Hight Channel)을 포함하여 스피커들이 3차원으로 배치되는 다채널 오디오 재생 시스템을 지원하여야 한다. 따라서 고차 다채널 오디오 신호는 전방위 음향장면을 표현할 수 있으며, 적어도 9.1 이상의 채널로 구성되어 있다.

객체기반 오디오(Object-based Audio) : 다채널 오디오 배경음과 분리되어 있는 오디오 객체 음원들을 전송하고 단말에서 제어할 수 있어야 한다. 오디오 객체는 자유롭게 음향 장면을 구성할 수 있으며, 사용자와 상호작용 서비스를 가능하게 한다.

장면기반 오디오 신호(HOA-based Audio) : 음향 장면기반 오디오 신호를 전송할 수 있어야 한다. 음향장면기반의 오디오 신호는 특정 오디오 획득 장비를 통해 녹취될 수 있으며, 오디오 채널 신호와 장면을 표현하는 메타데이터로 구성된다. MPEG에서 장면기반 오디오 신호를 획득하기 위한 수단으로 HOA(High Order Ambisonics) 포맷을 채택하였으며, HOA 포맷은 오디오 채널 신호와 채널 신호를 제배치하기 위한 메타데이터 정보가 필요하다.

MPEG-H 3DA 인코더는 상기 언급한 세 가지 포맷에 대해서 압축 전송이 가능하여야겠다. 인코더의 주요 기능은 고차 오디오 채널 신호, 다채널 오디오 신호, HOA 채널신호 등 기본적으로 고차 오디오 채널 신호를 압축하여 전송하는 기능을 담당한다. 기존의 스테레오 및 5.1 채널보다 다루는 입력 채널수가 고차이므로, MPEG-H 3DA 인코더의 압축 성능도 배제할 수 없다. MPEG-H 3DA는 압축 모듈로 USAC 기술을 채택하였다. 상기에서 살펴본 바와 같이, USAC은 MPEG-H 3DA 이전에 표준화된 최신 코어코덱 기술로 음성/음악에서 고른 성능을 보이며 압축 효율 측면에서도 매우 뛰어난 성능을 보였다.

MPEG-H 3DA는 새로운 압축 모듈을 개발하고 탑재하는 대신 USAC 기술을 활용하되, 고차 채널신호 압축효율을 높이기 위하여 MPEG Surround 기술을 접목하였다. MPEG Surround 기술도 이미 표준화가 완료된 기술로 MPEG-H 3DA에 USAC과 더불어 활용된 기술이다. 마찬가지로, 객체 오디오 신호와 장면 오디오 신호를 인코딩하고 디코딩하기 위하여 각각에 상응하는 코딩 모듈이 필요하다. 앞서 살펴본 바와 같이 오디오 객체신호는 MPEG SAOC 기술을 활용할 수 있다. HOA 기반 장면 오디오 정보는 MPEG-H 3DA에 새롭게 소개된 포맷으로 HOA를 위한 인코딩/디코딩 모듈이 필요하다. 종전의 MPEG 오디오 표준기술로 HOA 기반 오디오 채널 신호를 복원할 수 없었으므로 새로운 HOA 인코딩/디코딩 모듈이 필요하다. HOA는 채널 신호와 메타데이터로 표현되는 포맷으로 채널신호는 USAC으로 인코딩/디코딩하고 추가된 HOA 모듈에서 전송된 메타데이터로 장면정보를 복원하는 기능을 수행한다.

이상에서 MPEG-H 3DA 주요 특징을 살펴본 바를 정리하면, 입력 오디오 포맷을 지원하기 위하여 해당하는 MPEG 오디오 표준기술을 활용하였으며, HOA 메타데이터 처리를 위한 HOA 메타데이터 처리 모듈만이 새롭게 개발되어 포함된 것이다. 즉, 압축하고 전송하는 측면에서의 MPEG-H 3DA 기술은 최대한 기존 MPEG 기술을 활용한 것으로 상호 간에 유연한 동작을 위하여 하나의 코덱 구조에 적절하게 조합되어 구현되었다. [그림 2]는 MPEG-H 3DA 디코더 구조로 소개한 주요 모듈이 포함되어 있는 것을 확인할 수 있다.

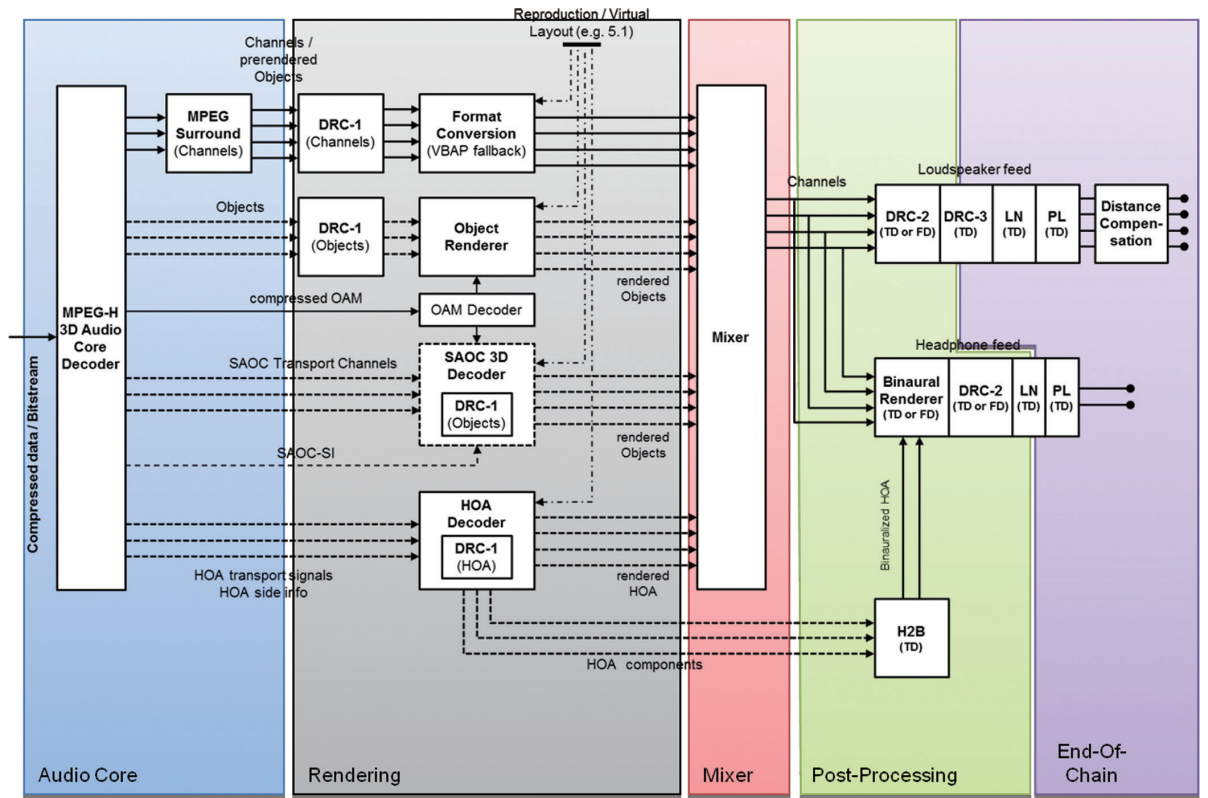


그림 2. MPEG-H 3D Audio 디코더 구조도

그렇다면, 사용자 환경에 따른 최상의 이머시브 오디오를 재생하기 위해서 추가로 요구되는 기능에 대해서 살펴볼 필요가 있겠다. 예를 들어 22.2채널의 콘텐츠를 인코딩/디코딩 과정을 거쳐 사용자 단말에 전달되었다 하더라도, 상황에 따라서 사용자는 스테레오 혹은 5.1채널로 22.2채널 콘텐츠를 소비하기를 희망한다고 했을 때, MPEG-H 3DA는 이를 충족시킬 수 있는 기능을 제공하여야 한다. 이를 위해 필요한 것이 후처리 기술로 3DA rendering 기술이다. 3DA rendering 기술은 콘텐츠 형태와 사용자 단말환경에 무관하게 최상의 이머시브 환경을 제공하기 위한 기술로, [그림 2]에서 format converter가 주되게 본 기능을 수행한다. format converter는 입력채널 레이아웃과 재생 출력 레이아웃 간의 변환장치로 수동적(passive) 믹싱과정을 통해 출력 채널을 생성하는 대신에, 원 콘텐츠의 효과를 최대한 반영하기 위한 능동적(active) 믹싱과정을 지원한다. 예를 들어, 22.2채널을 단순히 다운믹싱하여 스테레오 채널 신호를 생성할 수 있겠으나, 채널 간의 상관관계, 위치정보들을 고려하여 다운믹싱을 수행하면 보다 원 콘텐츠 음향장면에 가까운 스테레오 신호를 얻을 수 있다.

만일 사용자 환경이 헤드폰 환경을 선호하게 된다면 MPEG-H 3DA는 원 콘텐츠의 효과를 최대한 반영할 수 있도록 헤드폰을 위한 스테레오 신호를 생성할 수 있어야 한다. 이를 위해 MPEG-H 3DA는 바이노럴 렌더링 모듈을 탑재하고 있다. 바이노럴 렌더링 모듈은 디코딩된 초 다채널 신호를 스테레오 신호로 변환하는 과정에서 공간상의 스피커 위치에서 발생하는 전달함수를 적용하여 스테레오 신호를 생성하는 모듈이다. 스피커 위치에서 발생하는 전달함수를 BRIR(Binaural Room Impulse Response)라고 하며, 채널 수 만큼의 BRIIR이 존재하며 이를 활용하여 바이노럴 스테레오 신호를 생성한다. 원리는 단순히 디코딩된 채널 신호와 해당하는 BRIR을 컨볼루션(convolution)하는 것이다. 그러나 채널수가 증가함에 따라 컨볼루션 연산량이 급격하게 증가할 수밖에 없으며, MPEG-H 3DA에서는 이를 효과적으로 최적화된 연산량으로 처리될 수 있도록 모듈이 설계되었다. 따라서 추가적인 바이노럴 렌더링 모듈을 활용하기보다는 MPEG-H 3DA의 바이노럴 렌더링을 활용하는 것이 복잡도를 최소화하는 측면에서 유리하다.

마지막으로, MPEG-H 3DA에서 지원하는 기능으로 DRC(Dynamic Range Control) 모듈이 있다. MPEG-H 3DA는 기본적으로 다양한 포맷의 오디오 신호를 처리하므로, 상호 간에 발생하는 오디오 신호의 레벨차가 존재할 수 있다. DRC의 첫 번째 요구되는 기능은 상호 포맷 간의 레벨차를 맞추어 주어야 한다. 따라서 프로그램별로 다양한 형태의 오디오 포맷이 존재하더라도 최종 출력레벨의 차이를 느낄 수 없도록 DRC가 오디오 포맷별로 출력 레벨을 조절할 수 있어야 한다. 이는 MPEG-H 3DA의 입력 포맷으로 발생하는 요구사항이며, 기본적으로 DRC 기능은 출력 신호의 레벨을 자동적으로 일정레벨로 맞추어 줄 수 있어야 한다. 예를 들어, 방송콘텐츠별로 상이한 오디오 레벨로 제작되었다 하더라도, 사용자 단말에서는 사용자가 볼륨조절을 개별적으로 조절하지 않아도 동일한 라우드니스(loudness)를 가질 수 있도록 DRC가 처리할 수 있어야 한다. 이는 최근 들어 많은 이슈가 되어왔던 사항으로, 사용자가 채널별로 혹은 콘텐츠별로 인지하는 라우드니스 레벨 차이가 상이하여 시청 시, 오디오 볼륨을 상시 조절해야 하는 번거로움을 배제하기 위한 기술이다. MPEG-H 3DA는 코덱 단에서 원천적으로 DRC를 통해 라우드니스 제어하여 출력신호를 제공함으로써, 사용자의 볼륨조절의 번거로움을 최소화시킬 수 있다.

MPEG-H 3DA 프로파일

MPEG 오디오 표준화 기술은 활용분야에 따라 채널 수, 비트레이트 및 기능 등이 상이하게 다를 수 있다. 예를 들어 고품질 방송 서비스를 위해서는 표준에서 제시하는 모든 기능을 동원하여 최상의 음질 및 효과를 제공하기를 희망하겠으나, 모바일 환경에서는 한정된 채널 수와 적절한 음질 및 낮은 복잡도를 갖는 형상을 요구할 수 있다. 이를 위하여, MPEG 오디오 표준화 기술은 프로파일을 정의한다. 각 프로파일은 서비스 목적에 따라 선택적으로 복잡도를 제공할 수 있고 기능 및 지원 가능한 채널수를 명시하고 있다. MPEG-H 3DA는 3개의 프로파일을 가지고 있으며 이는 아래와 같다.

Main Profile : MPEG-H 3DA 표준화는 Phase 1과 Phase 2로 표준화가 진행되었다. Phase 1의 경우 상대적으로 높은 비트레이트에서 높은 음질 및 기능을 제공하기 위한 것으로 2015년 2월에 완료되었다. Main Profile은 Phase 1에서 개발된 툴들에 대한 프로파일이다. 복잡도의 제약사항을 두지 않고, 고품질 및 다기능을 지원하기 위하여 최대한 필요한 툴들을 모두 포함시키고 있으며, 채널수 및 객체수도 최대한 제약을 두지 않고 있다. 그러나 성능에 주안점을 두고 있다 하더라도 높은 복잡도로 인하여 활용도가 떨어질 수 있다는 우려가 있었다. Main Profile에 포함된 툴들은 앞서 열거한 USAC, SAOC, MPEG Surround, HOA 등으로 기존 표준기술들을 그대로 포함시킨 super-set의 프로파일이다.

Low-Complexity(LC) Profile : MPEG-H 3DA가 고음장감을 제공하고 다기능을 지원함으로써 새로운 오디오 코덱 형상을 제시한 것은 사실이나, Main Profile은 실제 서비스 단말에 지원하기 위해서는 높은 복잡도로 인하여 다소 부담이 되는 것은 사실이다. 이를 해결하기 위하여 Phase 2 기술 개발기간동안에 저 복잡도를 갖는 툴, 혹은 기존 툴을 저 복잡도로 단순화한 기술들이 제안되었다. LC Profile은 이러한 툴들을 조합하여 최선의 저 복잡도를 갖는 프로파일을 제안하는 것이다. 다행스러운 것은 LC Profile이 저 복잡도를 갖는다 하더라도 음질 성능 및 기본적인 기능에는 Main Profile과 크게 다르지 않다는 것이다.

High-level Main Profile: High-level Main Profile은 Main Profile과 LC Profile의 superset Profile이다. LC Profile이 저 복잡도를 위한 Profile이기는 하나, 새롭게 추가된 툴들로 인하여 Main Profile에 기술되지 않은 툴들이 발생하였다. 따라서 MPEG-H 3DA의 모든 툴들이 포함된 Profile이 기본적으로 표준에서는 제공하여야 하기에 High-level Main Profile을 새롭게 정의하여 모든 툴들을 포함시켰다.

차세대 방송 표준화 현황

차세대 방송 표준으로 추진 중인 UHD 방송 규격은 국내외에서 활발하게 진행 중에 있다. 가장 대표적인 방송 표준으로 북미의 방송 표준을 제정하는 기구인 ATSC(Advanced Television Systems Committee)에서는 ATSC 3.0이라는 프로젝트를 2012년도부터 진행해왔으며, UHD 방송표준을 제정하는 것을 목적으로 하고 있다. ATSC 3.0 표준의 핵심은 미디어 및 전송에 대한 후보 표준을 선정하는 것이다. 비디오의 경우, UHD 영상을 위한 압축 방식으로 MPEG의 HEVC(High Efficiency Video Coding) 기술이 채택되는 것에 대하여 이견이 없었으나, 오디오

의 경우 이해관계자들의 참여한 대립으로 인하여 오디오 코덱 표준 선정의 어려움이 있었다. 이해관계에 놓인 가장 대표적인 그룹으로 AC-4(Advanced Coding)를 지지하는 Dolby사와 MPEG-H 3DA를 지지하는 MPEG Audio Alliance 그룹이었다. Dolby는 다년간 방송 시스템에 대한 노하우를 바탕으로 UHD 방송을 위한 AC-4 기술을 제안하였으며, MPEG은 HEVC와 마찬가지로 차세대 오디오 코덱 기술로 추진 중인 MPEG-H 3DA 기술을 후보 표준기술로 제안하였다. 최종적으로 2015년 11월 회의에서 AC-4 기술을 북미의 방송 표준 기술로 활용할 것으로 확정하였으나, MPEG-H 3DA 기술도 ATSC 3.0 후보 표준코덱 기술로 포함시켰다. 즉, 두 개의 오디오 표준 코덱을 포함하고 있으나, 북미에서는 AC-4를 활용함을 의미하는 것으로 사실상 북미 UHD 방송 표준에 활용될 오디오 코덱 기술은 AC-4로 선정된 것이다.

국내에서도 2015년 UHD 방송을 위한 700MHz 주파수 대역을 각 방송사가 활용할 수 있도록 할당 받음으로써, 본격적인 UHD 방송을 위한 작업에 착수하였다. ATSC와 마찬가지로 비디오 코덱 선정은 HEVC로 무난하게 결정되었으나, 오디오의 경우 AC-4와 MPEG-H 3DA 기술이 다시한 번 경합을 벌이게 되었다. 최종적으로 국내에서는 이머시브 오디오 서비스를 실현하기 위한 렌더링 기능을 포함한 MPEG-H 3DA 기술을 잠정 표준 RM(Reference Model)으로 선정하고 표준 초안을 작성하였다. 현재 이를 국내 표준화 단체인 한국정보통신기술협회에 상정하여 표준화를 진행 중에 있으며, 2016년 7월 이전에 표준안 작업을 마무리할 계획이다.

맺음말

본 글을 통하여 UHDTV와 같은 차세대 실감방송을 위하여 MPEG에서 표준화가 진행 중인 MPEG-H 3D Audio의 배경 및 진행과정을 살펴본다. 아직도 대부분의 방송 프로그램은 스테레오 포맷으로 제작되고 있으며, 5.1채널 방송 프로그램을 위한 제작환경이나 전문 음향엔지니어가 부족한 실정이다. 이러한 상황에서 5.1채널보다 훨씬 많은 채널들로 구성되는 22.2채널과 같은 고차 다채널 오디오 프로그램을 제작할 수 있는 환경이 보급되기까지는 많은 시간이 소요될 것으로 예상된다. 또한 일반 가정에 이렇게 많은 수의 스피커들을 설치하도록 유도하는 것도 쉬운 일이 아닐 것이다. 이러한 현실에서 과연 이렇게 많은 채널들로 구성되는 오디오 서비스가 보급되기 어렵다고 생각할 수 있지만, 전통적으로 새로운 오디오 서비스는 영화로부터 시작된 역사를 되짚어 보면 비관적이지만은 않다. Dolby ATMOS, AURO 3D, 소닉티어 14.2채널/30.2채널과 같은 다채널 영화음향 기술들이 주목받고 있으며, 시장에서도 긍정적인 반응을 받고 있다. 영화에서부터 이러한 고차 다채널 오디오로 구성된 콘텐츠들이 제작되기 시작하면 방송 등과 같은 다양한 분야에서도 활용될 것이다. 따라서 MPEG-H 3D Audio 표준이 실제 방송에 활용되는 시점이 우리가 생각하는 것보다 훨씬 빨라질 수도 있다.

본 원고는 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.B0101-16-0295, 초고 품질 콘텐츠 지원 UHD 실감방송/디지털시네마/사이니지 융합서비스 기술 개발) 