

디코더 구현 관점에서 본 MPEG-H TV Audio System 소개

*곽경철 양진영 배성용

한국정보통신기술협회

*kc.kwak@tta.or.kr

Introduction of MPEG-H TV Audio System from the Perspective of Decoder Implementation

*Kyungchul Kwak Jinyoung Yang Sungyong Bae

Telecommunications Technology Association(TTA)

요약

최근 다양한 국제 방송표준에서 차세대 오디오 코덱의 하나로 MPEG-H 3DA(3D Audio)가 채택되었으며, 이를 활용한 몰입형 오디오 서비스들이 개발되고 있다. 이러한 몰입형 서비스를 원활히 제공하기 위해서는 표준에 정의된 기술을 구현한 제품간의 상호호환성 검증이 필수적으로 추진되어야 하며, 이를 위해 개발된 MPEG-H TV Audio System 인증 프로그램에 대해 대상 제품과 시험서비스의 구조에 대해 설명하고 있다.

1. 서론

북미 ATSC를 비롯하여, 국내 TTA 및 유럽의 ETSI에서는 차세대 TV 서비스를 위한 오디오 기술의 하나로 MPEG-H 3DA(3D Audio)를 채택하였다[1][2][3]. 채택된 MPEG-H 3DA의 주요한 특징은 방송 신호 내 오디오 객체들에 대한 프리셋(Preset) 제공을 통해 몰입형(immersive) 사운드를 포함한 다양한 새로운 오디오 서비스 기능을 지원한다는 점이다.

객체 오디오를 통해서 시청자 또는 청취자는 청취환경과 개인적 선호도를 고려하여 개인화된 최적의 설정을 실행할 수 있으며, 헤드폰을 통해 몰입감을 극대화할 수 있다.

국내에서는 세계 최초로 MPEG-H 3DA를 이용한 지상파 UHD 방송서비스를 2017년 5월부터 제공하고 있지만, MPEG-H 3DA의 다양한 몰입형 서비스는 제공되지 못하고 단순 오디오 디코더 역할만 수행하고 있는 실정이다.

이에 대한 원인은 방송 제작환경의 한계로 몰입형 서비스 제공을 위한 다양한 메타데이터 처리와 음원 획득 과정에서의 한계도 분명 존재하지만 MPEG-H 3DA 디코더를 포함한 오디오 재생 장비인 TV, 사운드 바 등에서 이를 처리하지 못한 것도 분명한 원인이다. 또한, 개별적인 제품은 표준을 따른 다 하더라도 표준에 정의된 기술을 구현하는 방식의 차이로 제품 간 상호호환성을 보장하지 못할 가능성이 충분이 높다.

이로인해 시청자에게 좋은 콘텐츠 경험을 제공할 수 있는 MPEG-H 3DA 기술이 오히려 서비스의 활성화를 가로막는 장애 요인이 될 수 있으며, 이를 해결할 필요성이 제품의 개발과 서비스 개시 단계에서 제기되었다.

본 논문에서는 이러한 표준 구현에 대한 문제점을 해결하기 위해 개발된 MPEG-H TV Audio System[4][5]에 대한 소개를 통해 제품

개발자들에게 표준적합성 평가프로그램의 필요성과 내용을 설명하고자 한다.

2. MPEG-H 3DA 특징

차세대 음향 시스템은 몰입형 사운드와 3D 음향 재현을 위해 수직 차원의 음향 이미지 확장을 통한 서라운드 구현을 주된 특징으로 갖는다. 몰입형 사운드의 구현은 통상적인 채널 기반의 다채널 오디오뿐만 아니라, 오디오 오브젝트(Object)와 장면(Scene) 기반 오디오를 통해 이뤄진다.

MPEG-H 3DA를 포함한 차세대 오디오 시스템은 전통적인 5.1 채널 또는 7.1 채널의 입체음향에 천장 스피커 구성을 추가하여 3차원 입체 음향을 제공할 수 있으며, 통상적으로 4개의 스피커가 배치된다. MPEG 표준에서는 스피커 구성의 모호함을 방지하기 위해 MPEG CICP(Coding-Independent Code Points)를 표준화하였으며, 5.1+4H(Height channel), 7.1+4H, 22.2와 같은 다양한 스피커 구성을 지원하고 있다[6].

오브젝트 기반 오디오는 게임과 멀티미디어 시스템에서 3차원 공간 상에서 음원을 재현하는데 사용되어 왔다. 또한, 다양한 청취환경과 청취자의 이동 등을 고려하면, 이러한 오브젝트 오디오에 대한 제어는 청취자의 몰입감을 증대할 수 있는 좋은 요소이다. MPEG-H 3DA 표준은 최대 128 채널과 128개의 오브젝트를 64개의 스피커에 매핑할 수 있도록 정의하고 있다.

장면 기반 오디오의 기본 철학은 가능한 정확하게 음압레벨에서 음원을 재현하는 것이며, HOA(High Order Ambisonics)를 이용하여 음압을 구형 고조파 기저함수의 파형 방정식으로 모델링하여 이를 구현한다. 이를 위해서는 Eigen-마이크와 같은 음장 획득이 가능한 특수

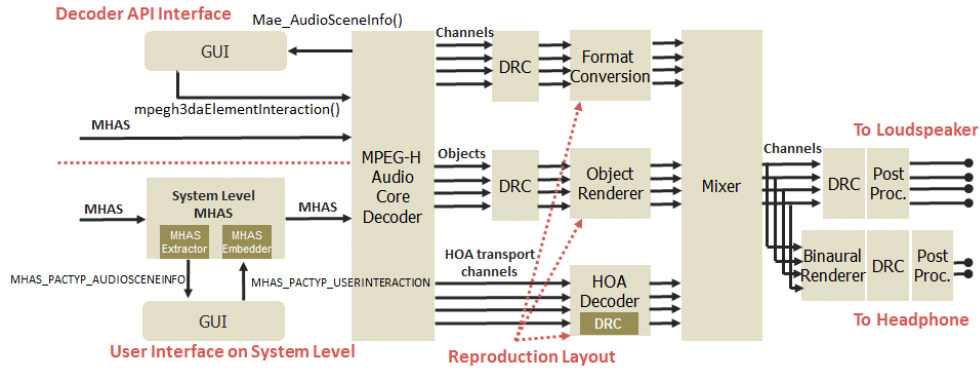


그림 1. MPEG-H 3DA 디코더 Architecture

마이크가 음원 획득 단계에서 사용되어야 한다.

또 다른 MPEG-H 3DA의 특징은 청취자의 환경 및 상황에 따른 개별적인 청취환경의 변경을 통한 개인화가 가능하다는 점이다. 이는 오디오 콘텐츠와 함께 관련 메타데이터를 통해서 가능하며, 프리셋에는 방송사에서 기본적으로 제공하는 프리셋 설정과 사용자가 변경 가능한 오디오 설정의 한계가 포함되어 있다.

3. 글로벌 표준에 정의된 MPEG-H 3DA 디코더 구조

MPEG-H 3DA가 채택된 표준은 북미의 A/342 Part3, 한국의 TTA/KO-07.0127/R1, 유럽의 ETSI TS 101 154가 있다[1][2][3]. 이들 표준에서 정의된 기술을 수신하여 처리할 MPEG-H 3DA의 디코더 구조는 동일하며, 모두 ISO/IEC 23008-3을 참조하여 그림 1과 같이 정리할 수 있다[7].

사용자와의 상호작용을 위해서 디코더 API를 이용하는 경우에는, 입력 MHAS(MPEG-H Audio Stream) 내 Core 디코더에서 Mae_AudioSceneInfo() 프레임에 이용하여 오디오 장면 정보를 디코더의 GUI(Graphic User Interface)로 전달하며, 사용자의 선호도가 반영된 결과는 mpeg3daElementInteraction() 프레임에 통해 Core 디코더로 전달된다. 그리고, 시스템 레벨에서 사용자와 상호작용이 이뤄지는 경우에는 Core 디코더 전 단계에서 MHAS Extractor를 통해 오디오 장면 정보를 추출하고, 시스템 레벨 GUI 처리 후 사용자의 선호도 정보를 받아 MHAS embedder에서 이를 반영하여, Core 디코더로 보낸다.

이후 Core 디코더에서 채널 기반, 객체 기반, HOA 기반 오디오를 분리하여, 청취 환경 및 콘텐츠에 따른 출력 음량 조절을 수행하는 DRC(Dynamic Range Control)를 수행한다. 이후, 출력 스피커 구성에 따른 출력 포맷 변환, 객체 오디오에 대한 재현, HOA 디코더 결과를 최종 믹싱하여 후처리 후 스피커 또는 헤드폰을 통해 오디오를 출력하는 구조이다.

4. MPEG-H TV Audio System 개요

앞서 소개한 MPEG-H 3DA를 표준 오디오 코덱으로 정의한 표준들에서 공통적으로 정의한 MPEG-H 3DA 프로파일은 Low Complexity 프로파일이며, 표 1의 저복잡도 프로파일의 레벨 3까지 지원하도록 되어 있다. 이는 디코더 구현의 복잡도와 MPEG-H 3DA를

이용한 청취자에게 만족도를 고려한 수준이다.

이러한 공통적인 기술 규격은 MPEG-H 3DA 디코더를 구현한 제품인 TV, 사운드 바, 셋톱박스, AVR(AV Receiver) 등에 공통적으로 적용된다. 따라서 이러한 공통적용된 기술 규격의 구현을 확인하고 장비간의 상호호환성의 보장은 나아가 서비스의 보급을 지원하는 후방 지원으로 간주할 수 있다.

이를 위해, MPEG-H 3DA에 대해 원천기술을 보유한 독일의 Fraunhofer IIS에서는 디코더 제품에 대한 상호호환성과 기본 구현 확인을 위한 “MPEG-H TV Audio System Trademark Program”을 개발하였다[5]. 현재, 해당 로고 프로그램에 대한 인증등록 비용은 무료이며, 해당 로고 프로그램의 장점은 다음과 같다. 우선, 사용자 입장에서는 MPEG-H 3DA 지원 제품에 대한 쉬운 구분이 가능하며, 인증 제품이 시장에 유통됨에 따라 방송사업자 입장에서는 MPEG-H 3DA 표준에 따른 다양한 실감음향 서비스를 제공할 수 있고, 제품 개발사 입장에서는 개발된 제품에 대한 검증에 대한 시간과 비용을 절약할 수 있다.

로고프로그램에서의 인증 대상장비는 TV, 셋톱박스, AVR, 사운드 바이며, 표준 제품의 구조는 다음 그림 2와 같이 정리할 수 있다. TV나 셋톱박스와 같은 장비는 사운드 바 또는 AVR과 같은 오디오 처리 전문 장비로 오디오 스트림을 전달하도록 이용이 가능하다. 이 경우, HDMI 인터페이스를 이용하여 사용자의 상호작용 처리를 거친 MHAS 스트림을 TV나 셋톱박스에서 사운드 바 또는 AVR로 전달하도록 되어 있다. 이러한 제품의 구성은 어디까지나 표준 외 구현의 영역이므로 산업계에서의 로고 프로그램을 통한 공통된 사용자 경험을 제공해 주는 것이 매우 중요하다.

현재 로고프로그램에서는 인증을 받는 과정은 그림 3과 같이 셀프 테스트와 공인시험소를 통한 인증시험의 두 단계로 나뉘어져 있다. TTA는 2017년 11월 프라운호퍼로부터 공인시험소 자격을 획득하고 시험 대상장비 확대에 따른 시험규격 검증을 진행하고 있다.

표 1. MPEG-H 3DA Low Complexity 프로파일 특징

Levels for LC profile	1	2	3	4	5
Max Sampling Rate, kHz		48			96
Max Core Codec Channels	10	18	32		56
Max Decoded Core Codec Channels	5	9	16		28
Max Loudspeaker Output	2	8	12		24
Example Loudspeaker Conf.	2.0	7.1	7.1.4	22.2	22.2
Max Decoded Objects	5	9	16	28	28

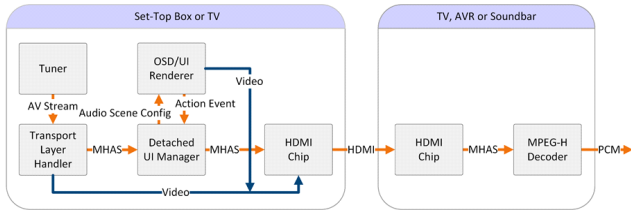


그림 2. MPEG-H TV Audio System 인증 대상 장비 구조

셀프테스트와 인증시험소를 이용하여 제품인증을 받는 경우 모두 사전에 MPEG-H 3DA 디코더에 대한 인증이 선행되어야 한다. 이는 개발된 디코더/렌더러 라이브러리에 대한 검증으로 셀프테스트를 통해 시험결과를 Fraunhofer측에 전달하여 인정받도록 진행된다.

이후, 개발된 제품에 대해 셀프테스트 또는 공인시험소를 통해 인증시험을 추진할 수 있으며, 셀프테스트의 경우 필요시 제조사 심사과정(audit)을 거치게 된다.

5. 결론

본 논문에서는 MPEG-H 3DA의 주요한 특징으로 수직채널의 확장과 함께 객체 오디오 등을 활용한 몰입형 사운드 재현을 소개하였으며, 북미, 국내, 유럽 등 글로벌 표준에서 적용된 공통된 MPEG-H 3DA 디코더 구조를 확인하였다. 이를 바탕으로 제품간 상호호환성 보장을 목적으로 개발된 MPEG-H TV Audio System의 범위와 절차를 소개하였다. 본 인증프로그램은, 앞으로 서비스될 몰입형 오디오 서비스의 활성화에 크게 기여할 것으로 예상된다.

참고문헌

[1] ATSC, *A/342 Part 3, MPEG-H System*, A/342-3:2017, March 2017
 [2] TTA, *Transmission and Reception for Terrestrial UHD TV Broadcasting Service*, TTA.KO-07.0127/R1, June 2016
 [3] ETSI, *Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream*, ETSI TS 101 154 v.2.3.1, Feb. 2017
 [4] Robert L. Bleit, et. al., "Development of the MPEG-H TV Audio System for ATSC 3.0", *IEEE Trans. on Broad. Vol. 63. No.1*, March 2017
 [5] <https://www.mpeg-h.com>
 [6] ISO/IEC, *High Efficiency Coding and Media Delivery in Heterogeneous Environments-Part 3: 3DA*, ISO/IEC 23008-3:2015, 2015
 [7] ISO/IEC, *MPEG Systems Technologies-Part 8: Coding-Independent Code Points*, ISO/IEC 23001-8:2016, 2016

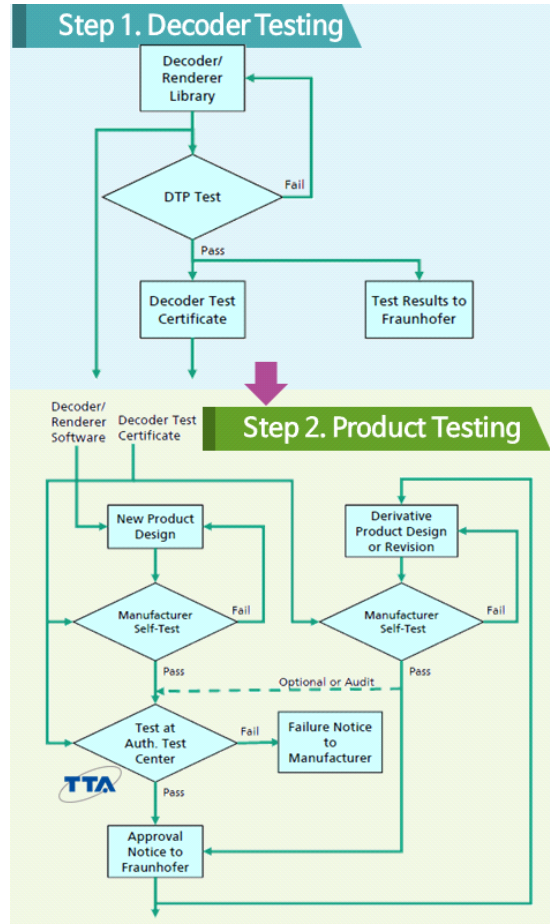


그림 3. MPEG-H TV Audio System 인증 절차