MPEG-I Immersive Audio 표준화 및 기술 동향

Standardization of MPEG-I Immersive Audio and Related Technologies

장대영 (D.Y. Jang, dyjang@etri.re.kr) 미디어부호화연구실 책임연구원
강경옥 (K.O. Kang, kokang@etri.re.kr) 미디어부호화연구실 책임연구원
이용주 (Y.J. Lee, draball@etri.re.kr) 미디어부호화연구실 책임연구원
유재현 (J.H. Yoo, jh0079@etri.re.kr) 미디어부호화연구실 책임연구원
이태진 (T.J. Lee, tjlee@etri.re.kr) 미디어부호화연구실 책임연구원/실장

ABSTRACT

Immersive media, also known as spatial media, has become essential with the decrease in face-to-face activities in the COVID-19 pandemic era. Teleconference, metaverse, and digital twin have been developed with high expectations as immersive media services, and the demand for hyper-realistic media is increasing. Under these circumstances, MPEG-I Immersive Media is being standardized as a technologies of navigable virtual reality, which is expected to be launched in the first half of 2024, and the Audio Group is working to standardize the immersive audio technology. Following this trend, this article introduces the trend in MPEG-I immersive audio standardization. Further, it describes the features of the immersive audio rendering technology, focusing on the structure and function of the RM0 base technology, which was chosen after evaluating all the technologies proposed in the January 2022 "MPEG Audio Meeting."

KEYWORDS MPEG 오디오, 공간음향 렌더링, 공간음향 모델링, 몰입형 미디어

Ⅰ. 서론

2018년 개봉한 영화 "Ready Player One"에서는 주인공이 가상세계와 현실세계를 넘나들며 활약하는 모습을 보여주며, 최근 관심의 대상이 된 메타버스 서비스의 미래상을 보여주었다. 코로나19 팬데믹 사태로 인한 사회적 거리두기 정책과 맞물리며, 메타버스에서는 다양한 기관 및 조직의 공식 모임 또는 행사를 비대면으로 시행하는 사례가 급증하고 있고, 가상인간을 통한 광고, 연예인 및 아이돌 그룹이 탄생하며, 디지털, 모바일 환경에 익숙한 MZ 세대를 중심으로 미디어의 대변혁이 일어날 것을 조심스럽게 예측하도록 부추기고 있다.

기술적으로 보면, 가상세계에서 현실세계와 구별되지 않는 자연스러운 활동을 제공하기 위해서는 완전 몰입이 가능한 고품질 멀티모달 휴먼 인터페이스,

* DOI: https://doi.org/10.22648/ETRI.2022.J.370306
* 본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음[22ZH1200, 초실감 입체공간 미디어·콘텐츠 원천기술 연구].

즉 충실한 오감의 제공이 중요하다. 오감 중에서도 가장 중요한 감각 중 하나인 청각 기술에서 떠오르고 있는 Immersive Audio는 주어진 음향 공간에 완전 몰입함으로써 실제로 현장에 있는 듯한 실재감의 체험이 가능한 새로운 음향 솔루션이라고 할 수 있다.

Immersive Audio의 가장 중요한 특징은 휴먼 인터페이스 환경이 청취자의 Yaw, Pitch, Roll에 의한 머리의 회전을 포함하는 X, Y, Z 축의 자유로운 움직임을 추적하여 대응함으로써 6DoF(Degree of Freedom) 사용자 상호작용을 제공한다는 것이다. 6DoF의 음향 공간에서 몰입감/실재감을 제공하기 위해서는 시각 경험과 완벽히 일치되는 공간음향 경험이 중요한데, 이를 위해서는 음향에 의한 공간정보 인지 능력인 Echolocation이 발현되는 조건이라고 할 수 있는 음향적 Motion Parallax와 임의 공간 내에서의 청취자의 움직임에 따라 기대하게 되는 음향의 변화를 얼마나 잘 재현해 내는지가 중요한 성능 요인이 된다고 할 수 있다.

MPEG(Moving Picture Experts Group)에서는 이러한 추세에 따라 VR(Virtual Reality)/AR(Augmented Reality) 어플리케이션을 위한 몰입형 미디어 기술로서 MPEG-I Immersive Media 표준화를 추진하고 있다. MPEG 오디오 그룹(SC29/WG6)에서는 2024년 완료를 목표로 MPEG-I Immersive Audio 표준화를 진행하고 있으며, 그림 1과 같이 표준화 범위는 6DoF MPEG-I Immersive Audio 메타데이터 비트스트림과 실시간 렌더링 기술이 포함되어 있다[1]. 2022년 1월에 제안기술의 평가 결과를 바탕으로 RM0(Reference Model 0) 기술이 선정되었으며, 4월에는 Working Draft 문서 및 RM0 Reference SW를 발간할 예정이다[2].

본고에서는 MPEG-I Immersive Audio 표준화 현황과 기술 동향을 CfP(Call for Proposal) 평가 결과와 CfP에 제안된 기술들을 중심으로 설명하고자 한다. 이어지는 Ⅱ장에서는 MPEG-I Immersive Audio 기술 표준화 현황에 관하여 기술하고, Ⅲ장에서는 CfP 평가 결과와 그에 따른 RM0 개발 일정 및 계획을 요약하여 기술한다. Ⅳ장에서는 RM0 기술로 선정된 기술들을 중심으로 Immersive Audio의 기술 동향을 살펴보려고 하며, Ⅴ장에서는 Immersive Audio 서비스 전망과 함께 결론을 맺고자 한다.

Ⅱ. MPEG-I Immersive Audio

MPEG-I Immersive Audio 기술은 그림 1에서 나타낸 것과 같이 이전에 표준화가 완료된 MPEG-H 3D Audio 기술을 기본 코덱으로 사용하고 있고, VR/AR 어플리케이션의 6DoF 사용자 상호작용을 통한 사용자 경험을 제공하기 위한 MPEG-I 메타데이터를 전송하는 비트스트림과 MPEG-I Immersive Audio 렌더러를 표준화 범위에 포함하고 있다.

MPEG-I에서는 표준화를 위하여 6DoF 오디오의 요구사항을 정의하고 있는데, 요약하면 다음과 같다[3].

- 공간음향 재생: 청취자의 6DoF 움직임과 일치하는 사용자 경험 제공
- 비트스트림: 미디어 및 메타데이터의 효과적인 표현과 압축 제공
- 재생 방법: 헤드폰 및 멀티채널 스피커 재생
- 음원 모델: 지향성 및 볼륨 음원 제공

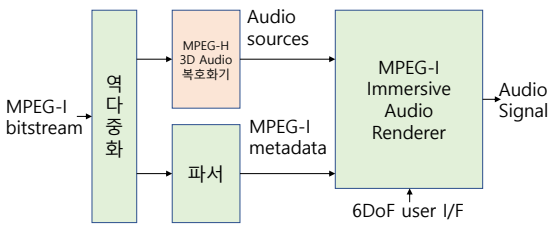


그림 1 MPEG-I Immersive Audio 표준화 범위(녹색)

- 공간음향 렌더링: 설득력 있는 실내 혹은 물리적 음향 현상 제공
- 장애물 효과: 방 구조 및 환경의 기하학적 장애물에 의한 투과, 회절 효과 제공
- 도플러 효과: 고속 이동 음원에 의한 피치 변화 효과 제공
- 사용자 음원: 로컬 및 원격 사용자의 현장음이 주어진 환경에 현장감 있게 렌더링될 것

6DoF 청취자 상호작용은 그림 2와 같이 머리의 회전과 신체의 움직임을 모두 추적하여 그에 맞는 공간의 음향 경험을 재현해 주는 기술로서, 기존의 제작 단계에서 완성된 멀티채널 기반의 콘텐츠를 일방적으로 소비하던 형태에서 직접 공간을 돌아다니며 물리적 공간과 상호작용하면서 실시간으로 변화되는 몰입형 음향 경험을 소비하는 형태로 변화된다.

이러한 특징에 의해 6DoF 오디오 렌더러는 공간음향 모델링과 렌더링을 동시에 수행하여야 하므로 일반적인 하드웨어 사양으로는 실시간 구현하는 것이 매우 어렵다. MPEG-I에서는 이러한 문제를 해결하기 위하여 콘텐츠 저작단계 및 모델링 단계에서 미리 결정할 수 있는 파라미터 생성을 인코더에서 수행하고, 디코더에서는 청취자의 움직임에 따라 실시간 렌더링에 필요한 처리만 수행하는 구조로 설계하였다[4]. 그림 3은 이러한 MPEG-I

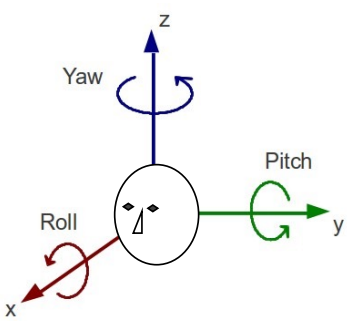


그림 2 6DoF 청취자 상호작용(회전/이동)

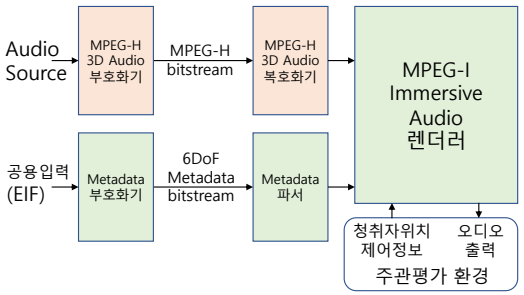


그림 3 MPEG-I Immersive Audio 기술의 기본 구조

Immersive Audio 기술의 기본 구조 및 주관 평가를 위한 인터페이스를 나타내고 있으며, 기존 표준 기술인 MPEG-H 3D Audio 코덱 기술과 MPEG-I 인코더 및 비트스트림과 실시간 렌더링 기술로 구분되어 있다. 인코더의 범위를 예측 가능한 범위로 한정하기 위하여 공용 입력 포맷으로 EIF(Encoder Input Format)를 규정하고 있으며, 렌더러는 실시간 주관평가를 위하여 외부 인터페이스를 규정하고 있다.

인코더의 EIF는 MPEG-I Immersive Audio 콘텐츠의 공간음향 장면 표현을 위하여 음원의 종류, 음원의 형상, 음원의 지향성 등 음원의 정보와 공간 구조 정보, 공간 재료 정보, 음향 환경 정보, 각 객체의 움직임 및 사용자 상호작용을 위한 갱신 정보 등을 포함하고 있다[5]. MPEG-I Immersive Audio 인코더는 EIF의 공간음향 장면 정보를 이용하여 공간음향의 렌더링에 필요한 메타데이터를 생성하며, 이 메타데이터는 비트스트림으로 전송되어 공간음향의 실시간 렌더링 처리에 사용된다. MPEG-I Immersive Audio 렌더러는 VR 헤드셋의 센서로부터 청취자의 움직임 및 머리 회전 정보를 입력받아 청취자의 현재 위치 및 머리 방향에 대응하는 공간음향을 재생하게 된다. MPEG-I Immersive Audio CfP 주관평가에서는 실시간 렌더러 성능 평가를 위하여 VR 영상 솔루션인 Unity와 연동하여 실시간 6DoF AV 사용

자 경험을 평가하게 되어 있다.

MPEG-I Immersive Audio 기술의 표준화를 위하여 2021년 1월에 Architecture and Requirements를, 4월에 CfP를 발행하였고, 같은 해 11월에 제출된 기술들을 평가하여 2022년 1월 WG6 미팅을 통하여 RM0 base 기술을 선정하였으며, 기술 통합을 통하여 2022년 4월 WG6 미팅을 통하여 RM0 기술을 개발하며, CE(Core Experiment)를 통하여 2023년 1월에 Committee Draft를 확정할 예정이다[2].

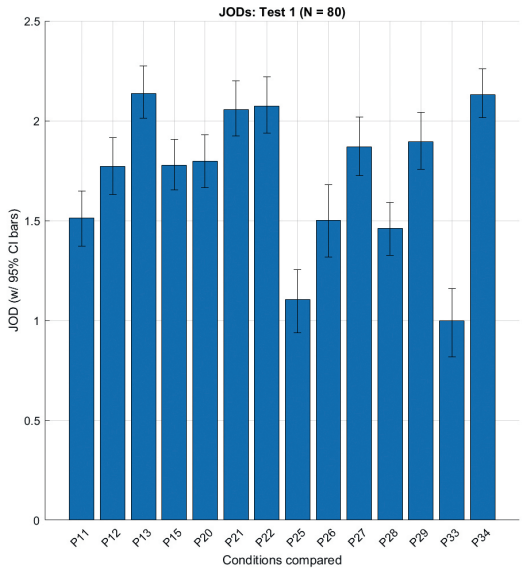
Ⅲ. CfP 평가 결과

CfP 주관평가는 다음의 세 가지 평가로 구성되어 있다.

- 평가 1: 객체, 채널, 3DoF HOA(High Order Ambisonic) 기반 VR
- 평가 2: 객체, 채널 기반 AR
- 평가 3: 객체, 채널, 6DoF HOA 기반 VR

평가 2는 특별히 AR 어플리케이션을 위하여 청취자 공간의 정보인 LSDF(Listener Space Description Format)가 EIF와 유사한 형태로 제공되어야 하며, 각 평가를 위하여 성능을 측정할 수 있는 속성을 정의하고 각 속성을 잘 평가할 수 있는 평가 콘텐츠를 제작하였다.

CfP에 대응하여 8개 기관에서 기술 제안을 하였는데, 각 제안 기관에서는 두 개씩의 렌더러를 제출하여 평가 1의 경우, 총 16개의 렌더러가 제출되었다. 이 중 한 기관에서 제출한 두 개의 렌더러는 평가 검증 단계에서 실시간 실행에 문제가 발생하여 제출이 철회되어 14개의 렌더러로 주관평가가 실시되었다. 주관평가에는 총 12개 기관이 참여하였으며, 평가 1에는 총 82명, 평가 2에는 총 39명, 그리고 평가 3에는 55명이 피험자로 참여하여 AB 비교를 통한 주관평가를 시행하였다. 결과 데이터는 post-screening을 통하여 동일한 렌더러를 구분하지 못하는 피험자의 데이터를 걸러내고 Thurstone V 분석을 사용하여 JOD(Just Objectionable Differences) 스케일을 산출하였다.



출처 Reproduced from [6], WG6 MPEG Audio Coding output document (N0119, Report on MPEG-I Immersive Audio Call for Proposals) ⓒ ISO/IEC [2022] – All rights reserved.

그림 4 평가 1의 주관평가 결과

이렇게 산출된 평가 1의 주관평가 결과는 그림 4와 같으며, 프라운호퍼, 에릭슨, 노키아 연합의 렌더러 P13이 최고 득점을 하여 RM0 base 기술로 선정되었다[6].

RM0 기술의 개발 방법에 따르면, 평가 1의 최고 득점 기술을 RM0 base 기술로 선정하고, 이를 기반으로 평가 2, 평가 3의 최고 득점 기술을 통합하며, 그 외에도 RM0 base 기술과 비교하여 복잡도 범주와 비트율 범주의 우수 기술을 선정하여 통합하는 것으로 되어 있다.

평가 2에서는 P21 렌더러가 최고 기술로 선정되고, 평가 3에서는 P12 렌더러가 최고 기술로 선정되

었는데, 두 렌더러 모두 프라운호퍼, 에릭슨, 노키아 연합이 제출한 렌더러로서 P13과 동일한 렌더러 구조에 기반하고 있다. 평가 1의 렌더러 중 복잡도 범주의 우수 기술은 기준에 적합한 렌더러가 없어 선정되지 않았으며, 비트율 범주의 우수 기술은 필립스, 돌비, 퀄컴 연합의 렌더러 P27이 선정되었다.

이렇게 선정된 각 평가의 최고 기술과 비트율 범주의 우수 기술을 통합하여 RM0 기술을 개발하는 것이 2022년 1월 SC29/WG6 회의에서 결정되었으며, 다음과 같은 절차로 통합하도록 하였다.

- 평가 1의 최고 기술이 RM0 base가 된다.
- 평가 2와 평가 3의 최고 기술의 독창적인 모듈을 RM0 기술과 통합한다.
- 평가 1의 비트율 범주의 우수 기술의 독창적인 모듈을 RM0 base 기술과 통합한다.

여기서 독창적 모듈이란 RM0 base 기술의 성능 및 기능을 향상시키거나 RM0 base 기술에 포함되어 있지 않은 모듈을 의미한다.

Ⅳ. RM0 선정 기술

MPEG-I Immersive Audio CfP에 대응하여 총 5개의 기술이 제출되었으며, RM0 base 기술로 선정된 프라운호퍼, 에릭슨, 노키아 그룹의 제안 기술[7]과 평가 1의 비트율 범주의 우수 기술인 필립스, 돌비, 퀄컴 그룹의 제안 기술[8]을 중심으로 MPEG-I Immersive Audio의 기술 동향을 소개하고자 한다.

1. RM0 base 선정 기술(P13)

프라운호퍼, 에릭슨, 노키아 그룹에서 제안한 기술은 CfP 주관평가에서 가장 높은 점수를 받아 RM0 base 기술로 선정되었으며, 객관적인 평가 지표인 복잡도 범주 및 비트율 범주에서는 다른 기술들에 비해 매우 열악한 성능을 보여주고 있다. 이 제안 기술이 주관평가에서 가장 높은 점수를 받기 위해 복잡도 범주 및 비트율 범주의 지표에 대해서는 희생하는 전략을 사용한 것으로 추정할 수 있다.

가. 인코더의 구조 및 기능

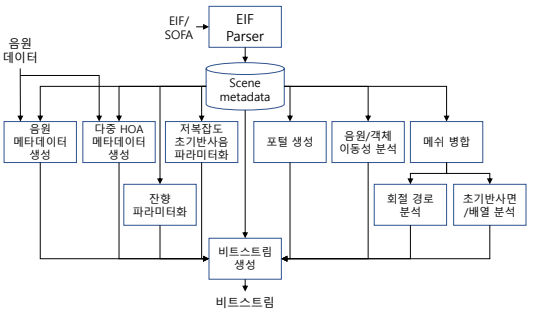
인코더의 대략적인 구조는 그림 5와 같으며, 주요 모듈별 기능에 대해 다음에서 간단히 기술한다.

1) EIF Parser

MPEG-I Immersive Audio 인코더의 공통 입력 포맷인 EIF 파일 및 SOFA(Spatial Oriented Format for Audio) 포맷의 지향성 정보를 입력하여 분석함으로써 콘텐츠의 장면 정보를 구성하는 요소들을 추출하는 모듈이다. 장면 정보를 구성하는 요소들은 공간의 기하학적인 구조 정보, 음원의 위치, 형상, 지향성 등 음원 정보, 재료 및 공간의 음향적 특성 정보, 그리고 움직임 정보를 포함하는 업데이트 정보 등이 있다.

2) 음원 메타데이터 생성

음원 신호의 특성 분석에 따른 렌더링 파라미터



출처 Reproduced from [7], WG6 MPEG Audio Coding contribution (M58913, Description of the MPEG-I Immersive Audio CfP submission of Ericsson, Fraunhofer IIS/AudioLabs and Nokia) © ISO/IEC [2022] – All rights reserved.

그림 5 RM0 base 인코더 구조

를 생성하는 모듈로서, 음향 신호 자체의 분석에 의한 거리감쇠 모델과 다중 음원을 가진 볼륨 음원의 렌더링 성능을 향상시키기 위한 확산 방사 특성을 포함하고 있다.

3) 다중 HOA 메타데이터 생성

다수의 HOA 음원을 동시에 사용하는 콘텐츠의 메타데이터로서 헤더와 정보 프레임과 공간 메타데이터 프레임으로 구분된다. 헤더에는 프레임의 길이, 정수값 변환을 위한 가중치를 가지며, 정보 프레임에는 위치 정보를 포함한다. 공간 메타데이터 프레임에는 HOA 음원 분석에 의한 DOA(Direction Of Arrival), DTR(Direct to Total energy Ratio), 그리고 총에너지를 포함한다.

4) 잔향 파라미터화

잔향 렌더링을 위한 파라미터를 생성하는 모듈로서 EIF에서 제공한 음향환경정보, 즉 RT60(Reverberation Time 60dB), 초기 지연, DDR(Diffuse to Direct Ratio)의 값에 기반하여 FDN(Feedback Delay Network)을 위한 감쇠 및 지연 파라미터, 그리고 음원 지향성에 따른 잔향의 방향성 필터 파라미터를 산출한다. 또한 EIF에서 음향환경정보를 제공하지 않는 공간에 대해서는 범용 음향환경정보를 산출하여 제공한다.

5) 저복잡도 초기반사 파라미터화

공간 분석에 의한 완전한 초기반사음 대신 초기반사음의 통계적 패턴 파라미터를 전송함으로써 렌더러에서 복잡한 연산을 거치지 않고 초기반사음을 근사하도록 한다.

6) 포털 생성

음향환경정보가 주어지는 방 사이 혹은 외부와의 통로를 통하여 전달되는 음향에 대한 파라미터로서, 음원이 있는 방 전체가 후기잔향을 포함하는 볼륨 음원이 되고, 일부 개방된 공간인 포털을 통하여 다른 공간의 청취자에게 전달되는 것을 모델링하는 모듈이다. 인코더에서는 EIF 기반의 기하학적 분석으로 모든 포털의 형상 정보를 생성한다.

7) 회절경로 분석

렌더러에서 회절효과를 효과적으로 구현하기 위한 회절경로를 미리 산출하는 모듈로서 정적인 음원에 대해서는 회절경로를 가능한 음원-청취자 쌍에 대해 미리 산출하고, 동적인 음원에 대해서는 음원에서 보이는 회절 에지를 추출하는 모듈이다.

8) 초기반사면 및 배열 분석

렌더러의 초기반사음 연산을 가속화시키기 위한 파라미터, 즉 각 음원의 초기반사를 일으키는 벽면과 반사되는 경로의 순서를 복셀로 표현되는 모든 음원 및 청취자 위치 쌍에 대해 미리 추출하는 모듈이다.

9) 비트스트림 생성

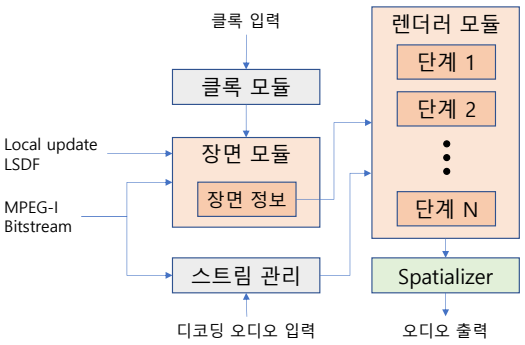
이상 인코더의 각 모듈에서 생성된 파라미터 집합과 EIF의 공간 구조, 음향 재료 등 메타데이터 및 SOFA 파일의 지향성 정보는 양자화 및 다중화되어 비트스트림을 형성하게 된다.

나. 렌더러의 구조

렌더러는 그림 6과 같이 비트스트림과 외부 인터페이스로부터 음원 신호, 클록 신호, 로컬 제어 정보, 청취자 위치 정보 등을 입력받아 렌더링을 수행하게 된다. 다음에서는 렌더러의 각 주요 모듈에 대하여 간단히 기술한다.

1) 장면 모듈

장면 모듈은 내부 혹은 외부의 모든 장면 정보의



출처 Reproduced from [7], WG6 MPEG Audio Coding contribution (M58913, Description of the MPEG-I Immersive Audio CfP submission of Ericsson, Fraunhofer IIS/AudioLabs and Nokia) © ISO/IEC [2022] – All rights reserved.

그림 6 RM0 base 렌더러 구조

변화를 처리하는데, 장면 정보는 음향 요소 및 물리 객체 등 장면의 6DoF 렌더링에 관련되는 모든 메타데이터의 현재 상태를 반영하는 모듈이다. 현재의 장면 정보는 비트스트림에 의해 전송된 장면 갱신 정보, 렌더러의 외부 인터페이스로부터 입력되는 청취자 위치 및 동적 갱신 정보, 그리고 AR 응용을 위한 LSDF에서 정의된 음향 요소들에 의한 갱신 정보를 기반으로 갱신된다.

2) 스트림 관리 모듈

스트림 관리 모듈은 장면 정보의 음향 요소에 관련된 음향 신호를 입력하는 인터페이스를 제공하며, 이 음향 신호는 미리 인코딩, 디코딩 처리된 음원 신호 혹은 로컬/원격 음원이다.

3) 렌더러 모듈

렌더러 모듈은 스트림 관리 모듈로부터 제공된 음향 신호를 현재의 장면 정보를 이용하여 렌더링하는 모듈이다. 렌더링을 수행하기 위해서 렌더링 대상이 되는 음향 신호, 즉 렌더 아이템의 렌더링 파라미터 처리 및 신호처리를 위한 렌더러 단계의 각 단계를 탄력적으로 처리한다.

4) 클록 모듈

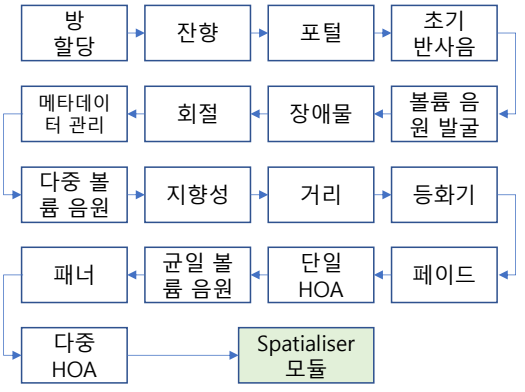
클록 모듈은 장면 모듈에 있어 장면의 현재 시간 정보를 제공한다. 이 클록 입력은 다른 외부 모듈과의 동기신호가 될 수 있고, 렌더러 내부의 기준 시간이 될 수 있다.

5) Spatializer 모듈

Spatializer 모듈은 렌더러의 마지막 단계로서, 렌더링되어야 할 렌더 아이템을 선택된 재생 방법에 따른 적절한 출력 신호를 생성한다. 기본적으로 제공된 SOFA 포맷의 HRIR(Head Related Impulse Response) 데이터에 기반한 바이노럴 스테레오 출력을 생성하며, 향후 멀티채널 스피커 재생을 위한 신호 출력을 제공하도록 확장될 수 있다.

다. 렌더러 단계 기능

렌더러 단계는 미리 설정된 순서로 실행되는데, 그 순서는 그림 7과 같다. 각 렌더러 단계는 활성화



출처 Reproduced from [7], WG6 MPEG Audio Coding contribution (M58913, Description of the MPEG-I Immersive Audio CfP submission of Ericsson, Fraunhofer IIS/AudioLabs and Nokia) © ISO/IEC [2022] – All rights reserved.

그림 7 RM0 base 렌더러 단계의 처리 순서

되어 있는 렌더 아이템을 렌더링 처리하게 되는데, 각 렌더러 단계에서 렌더 아이템을 선택적으로 비활성화 혹은 활성화할 수 있다.

1) 방 할당

방 할당 단계는 음향환경정보가 포함된 방에 청취자가 들어가는 경우, 그 방에 대한 음향환경정보의 메타데이터를 각 렌더 아이템에 적용하는 단계로서, 이후 잔향, 포털 단계에서 이 정보를 사용하여 관련 처리를 수행한다.

2) 잔향

잔향 단계는 현재 공간의 음향환경정보에 따라 잔향을 생성하는 단계로서, 비트스트림으로부터 잔향 파라미터를 읽어와서 FDN 잔향기의 감쇠 및 지연 파라미터를 초기화한다. AR의 경우, 렌더러에 직접 입력되는 LSDF의 음향환경정보에 의해 인코더보다 간단한 FDN 잔향기의 파라미터를 산출하여 사용한다. 잔향기의 출력은 몰입감을 높이기 위해 멀티채널 패너에 의해 청취자 둘레에 균등한 분포로 렌더링한다.

3) 포털

포털 단계는 후기 잔향에 대해 음향환경정보가 다른 공간 사이에서 부분적으로 개방된 음향전달 경로를 모델링하는 단계로서, 음원이 있는 공간 전체를 균일 볼륨 음원으로 모델링하며, 비트스트림에 포함된 포털의 형상 정보에 따라 벽을 장애물로 간주하여 균일 볼륨 음원 렌더링 방법으로 렌더링한다.

4) 초기 반사음

초기 반사음 단계에서는 고품질 및 저복잡도의 두 가지 초기 반사음 렌더링 방법이 제공된다. 품질과 연산량을 고려하여 선택하게 되며, 이 단계를 없애는 것도 가능하다.

- **고품질 초기 반사음**

비트스트림에 포함된 초기반사를 일으키는 초기반사 벽면에 대한 이미지 소스의 가시성을 판단하여 초기 반사음을 산출할 수 있다. 또한, 대안으로 인코더에서 생성된 음원 및 청취자 복셀 쌍에 대한 전파 경로 정보인 복셀 데이터를 사용함으로써 고속 연산이 가능하다. 복셀 데이터가 제공되는 경우 2차 반사음까지 실시간 처리가 가능하며, 복셀 데이터 없이 직접 산출하는 경우 1차 반사음까지 처리할 수 있다. 또한 이 단계에서는 장애물에 의한 반사 및 투과 손실을 함께 처리한다.

- **저복잡도 초기 반사음**

미리 정의한 간단한 초기 반사음 패턴들을 사용하여 초기 반사음 구간을 대체하는데, 후기 잔향의 시작 시간과 음원-청취자 사이의 거리 및 청취자의 위치에 기반하여 결정된다. 인코더에서 잠재적인 청취자 위치에 대한 기하학적 분석을 통해 요약된 파라미터가 전송되며, 이에 의하여 수평면의 초기 반사음 패턴이 적용된다.

5) 볼륨 음원 발굴

볼륨 음원 발굴 단계는 포털을 포함한 공간적 크기를 가지는 음원을 렌더링하기 위해 사방으로 방사된 음선이 각 포털/볼륨 음원에 교차하는 점을 찾고 이 정보를 장애물 및 균일 볼륨 음원 단계에서 사용한다.

6) 장애물

장애물 단계는 음원과 청취자 사이의 직선 경로에 대한 장애물 정보를 제공하는데, 장애물 경계에

서의 페이드인-아웃 처리를 위한 상태 플래그, 투과율에 의한 EQ(Equalizer) 파라미터가 해당 데이터 구조에서 갱신된다. 이후 단계인 회절과 균일 볼륨 음원 단계에서도 이 정보를 그대로 사용한다. 균일 볼륨 음원에 대해서는 청취자로부터 볼륨 음원으로 방사되는 음선다발이 장애물로 가려진 부분과 그렇지 않은 부분에 대해 가려진 부분은 투과율을 적용함으로써 최종 바이노럴 신호를 생성하도록 한다.

7) 회절

회절 단계는 장애물에 의해 가려진 음원으로부터 청취자에게 전달되는 회절 음원을 생성하는 데 필요한 정보를 제공한다. 비트스트림에 포함된 회절 경로 혹은 회절 에지 정보를 사용하는데, 고정된 음원에 대해서는 미리 산출된 회절경로를 사용하고, 이동 음원에 대해서는 잠재적인 에지로부터 현재 청취자에 대한 회절경로를 산출하여 사용한다.

8) 메타데이터 관리

메타데이터 관리 단계에서는 렌더 아이템 중 거리 감쇠 혹은 장애물에 의해 가청범위 아래로 감쇠될 경우, 이후 단계에서 연산량을 절약할 수 있도록 비활성화한다.

9) 다중 볼륨 음원

다중 볼륨 음원 단계는 공간적 크기를 가지며 다수의 음원 채널을 포함하는 음원을 렌더링하는 단계로서, 멀티채널 혹은 HOA 음원에 의해 내부 및 외부 볼륨 음원 표현으로 렌더링한다. HOA의 경우 일반적인 내부 볼륨 음원 표현으로부터 외부 볼륨 음원 표현을 생성하며, 객체음원의 경우 EIF에 규정된 최대 9개의 음원을 배열하여 사용자 혹은 객체 중심 표현을 제공하게 된다.

10) 지향성

지향성 단계에서는 지향성 정보가 정의된 렌더 아이템에 대하여 음원의 현재 방향에 대한 지향성 파라미터, 즉 대역별 이득을 기존의 EQ 값에 추가 적용하는 단계로서, 정보 압축을 위해 감소된 비트스트림의 지향성 정보를 인터폴레이션하여 EQ 대역과 일치시킬 필요가 있다.

11) 거리

거리 단계에서는 음원과 청취자 사이의 거리에 의한 지연, 거리 감쇠, 공기 흡음 감쇠를 적용하는 부분이다. 지연은 가변 지연 메모리 버퍼 및 인터폴레이션/재표본화를 이용하여 렌더 아이템의 물리적 지연 및 도플러 효과를 생성한다. 음원이 등속 이동하는 경우에는 블록 단위의 거리를 산출하여 갱신한다. 거리 감쇠의 경우 점음원인 경우 $1/r$ 감쇠율을 적용하며, 볼륨 음원의 경우 별도의 감쇠 커브를 적용한다. 공기 흡음 감쇠의 경우 ISO 9613-1 표준에 따르며, 온도 20℃, 습도 40%, 대기압 101.325kPa 상태를 기본으로 적용한다.

12) 등화기

등화기(EQ: Equalizer) 단계는 장애물 투과, 회절, 초기 반사, 지향성, 거리 감쇠 등에 의해 누적된 주파수 대역별 이득 값에 대하여 FIR(Finite Impulse Response) 필터를 적용하는 단계이다.

13) 페이드

페이드 단계는 렌더 아이템이 비활성화되거나 활성화되었을 때 혹은 청취자가 공간적으로 점프하였을 때, 페이드인-아웃 처리를 수행함으로써 발생할 수 있는 불연속 왜곡을 감소시키는 단계이다.

14) 단일 HOA

단일 HOA 단계는 하나의 HOA 음원에 의한 배경음을 렌더링하는 단계로서, MPEG-H 3D Audio 디코더로부터 입력되는 ESD(Equivalent Spatial Domain) 포맷의 신호를 HOA로 변환한 후 MagLS(Magnitude Least Squares) 디코더에 의해 바이노럴 신호로 변환한다.

15) 균일 볼륨 음원

균일 볼륨 음원 단계는 피아노와 같이 공명을 사용하는 대형 악기, 폭포, 빗소리, 방의 잔향 등 공간적인 크기를 가지며 단일 특성을 가지는 음원을 렌더링하는 단계로서, 비상관(Decorrelation)된 스테레오 음원으로 볼륨 음원 공간의 무수한 음원들의 효과를 모사한다. 장애물에 가려진 경우 장애물 단계의 정보를 기반으로 부분적으로 가려진 효과를 생성한다.

16) 패너

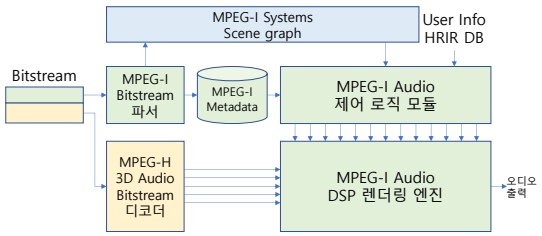
패너 단계는 멀티채널 잔향을 렌더링할 때 헤드트래킹 기반 글로벌 좌표에 각 채널 신호를 패닝 방법, 즉 VBAP(Vector Based Amplitude Panning) 기반으로 렌더링하는 단계이다.

17) 다중 HOA

다중 HOA 단계는 두 개 이상의 HOA 음원이 동시에 사용되는 콘텐츠의 6DoF의 음향을 생성하는 단계이다. ESD 포맷의 신호를 HOA로 변환하여 처리하게 되며, 인코더에서 미리 산출한 공간 메타데이터 프레임의 정보를 이용하여 청취자의 위치에 대한 6DoF 렌더링을 제공한다.

2. RM0 비트율 범주 기술(P27)

P27 렌더러의 상위 레벨 구조는 그림 8과 같으며,



출처 Reproduced from [8], WG6 MPEG Audio Coding contribution (M58851, Joint CfP proposals by Philips, Dolby and Qualcomm) © ISO/IEC [2022] – All rights reserved.

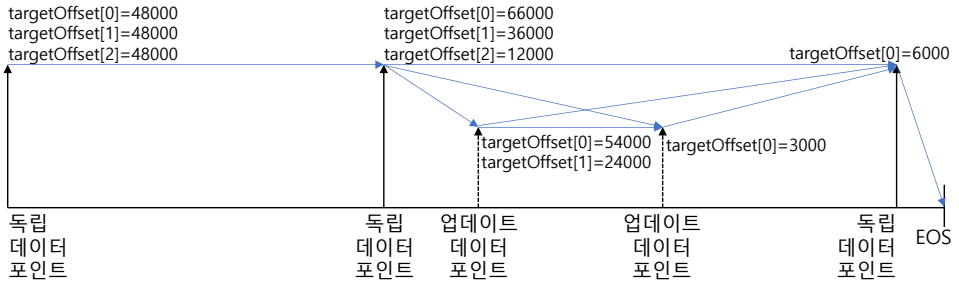
그림 8 P27 렌더러의 상위 레벨 구조

녹색으로 구분된 블록들이 MPEG-I 오디오 렌더러를 표현하고 있다. 다른 블록들은 다른 MPEG 표준들을 나타내며, 렌더러의 외부 인터페이스를 통해 연결된다. MPEG-I 비트스트림 파서는 비트스트림으로부터 오디오 메타데이터를 추출하여 데이터 구조에 저장하며, MPEG-I 시스템의 씬그래프에 관련 요소들을 추가한다.

제어 로직 모듈은 렌더링에 필요한 파라미터들을 씬그래프에 요청하여 받으며, 청취자의 현재 위치 및 방향에 기반하여 DSP(Digital Signal Processor) 렌더링 엔진을 구동하기 위한 다양한 제어 데이터를 생성한다[8].

전체적으로 P13과 P27 렌더러의 구조는 유사하다고 할 수 있으며, 명백히 다른 부분은 비트스트림의 임의접근을 위한 필수 렌더링 파라미터의 1초 주기의 반복 시 불필요한 중복을 피하기 위한 비트스트림 구조로 되어 있는 것과 MPEG-I 시스템의 씬그래프 정보를 사용한다는 것이다.

임의접근을 위한 비트스트림의 개념적인 블록 구조는 그림 9와 같으며, 데이터 포인트의 개념을 사용하고 있다. 각 데이터 포인트의 비트스트림은 다음 데이터 포인트까지의 장면 렌더링을 위한 데이터를 포함하고 있으며, 임의 길이를 가질 수 있다. 데이터 포인트에는 독립 데이터 포인트와 업데이트

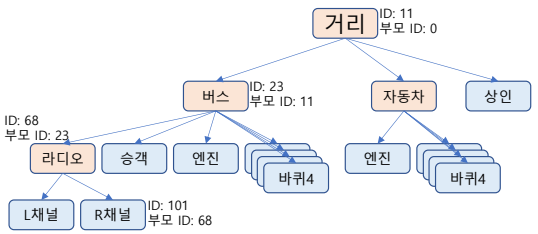


출처 Reproduced from [8], WG6 MPEG Audio Coding contribution (M58851, Joint CfP proposals by Philips, Dolby and Qualcomm) ⓒ ISO/IEC [2022] – All rights reserved.

그림 9 임의접근을 위한 비트스트림 블록 구조

데이터 포인트가 있으며, 독립 데이터 포인트는 임의접근을 위해 정의된 주기 사이의 장면을 디코딩하여 렌더링하는 데 필요한 모든 정보를 포함한다. 업데이트 데이터 포인트는 독립 데이터 포인트 주기 안에서 업데이트되는 정보를 보내기 위한 데이터 포인트로서, 음원 및 장면 객체의 이동 혹은 상태 변화를 처리하기 위한 메타데이터를 렌더러에 제공한다.

씬그래프는 MPEG-I 시스템 파트의 표준이며, 게임 콘텐츠에서 잘 알려진 개념인데, MPEG-I 표준에서는 시스템, 비디오, 오디오 파트 사이에 장면 정보를 공유하기 위해 사용한다. 그림 10과 같이 장면의 요소들, 즉 음원, 음향환경, 장애물 등을 계층적으로



출처 Reproduced from [8], WG6 MPEG Audio Coding contribution (M58851, Joint CfP proposals by Philips, Dolby and Qualcomm) ⓒ ISO/IEC [2022] – All rights reserved.

그림 10 음향 씬그래프의 한 예

표현하며, 각 요소는 자신의 유일한 ID(Identifier) 및 부모 ID를 가지며, 로컬 음원 및 원격 음원이 실시간으로 추가될 수 있다.

Ⅴ. 결론

지금까지 MPEG-I Immersive Audio의 표준화 및 제안된 기술의 개략적인 동향을 알아보았다. 현재의 RM0 base 기술로 선정된 P13 렌더러의 성능에서도 개선될 여지가 있으며, 향후 P27과의 통합 및 CE 절차를 통하여 국제표준 문서가 발간될 예정인 2024년 상반기까지 개선될 것이다.

MPEG-I 표준은 VR 및 AR을 대표적 응용 분야로 고려하고 있으며, 향후 메타버스, 원격 협력 작업, 디지털 트윈 등 소셜 실감 미디어 서비스에 활용될 수 있을 것이다. 소셜 미디어 서비스에 활용되기 위해서는 현장의 미디어를 획득하고, 실시간 인코딩하는 기능이 함께 필요하게 되는데, 현장의 미디어를 획득하기 위한 기술 및 미디어의 성능을 개선하기 위한 다양한 능동 잡음 제거 및 객체 분리 기술이 필요한 실정이다. 또한 VR 미디어의 실시간 인코딩을 위해서는 공간정보의 데이터베이스화, 분산된 미디어의 분산처리 기법 등 넘어야 할 허들이 매

우 많지만, 현재의 소프트웨어 기술 및 AI 기술의 발전 추세를 보면 머지않은 미래에 현실화될 수 있을 것으로 예상된다.

용어해설

Immersive Audio 몰입형 오디오 또는 공간음향과 동일한 의미로 사용되며, 기존의 3차원 입체음향과는 달리 자이로 및 가속도 센서에 의해 청취자의 움직임을 추적하면서 음향의 이미지가 글로벌 좌표에 고정되도록 함으로써 몰입감을 높일 수 있도록 공간적으로 안정된 음향을 제공하는 기술

Thurstone V Louis L. Thurstone이 개발한 심리 측정 점수 산출 방법으로서, 평가 대상 개체 쌍을 연속적으로 비교하도록 요청하고, 이를 다수의 피험자에 의해 반복적으로 시행하는 경우, 피험자들의 판단이 각 개체에 대한 잠재 척도 점수의 정규 분포에서 샘플링된 것이라는 가정하에 각 개체의 점수를 스케일링하는 방법

약어 정리

AI	Artificial Intelligence
AR	Augmented Reality
CD	Committee Draft
CE	Core Experiment
CfP	Call for Proposal
DDR	Diffuse to Direct Ratio
DOA	Direction Of Arrival
DoF	Degree of Freedom
DSP	Digital Signal Processor
DTR	Direct to Total energy Ratio
EIF	Encoder Input Format
EQ	Equalizer
ESD	Equivalent Spatial Domain
FDN	Feedback Delay Network
FIR	Finite Impulse Response
HOA	Higher Order Ambisonics
HRIR	Head Related Impulse Response
ID	Identifier
JOD	Just Objectionable Differences
LSDF	Listener Space Description Format
MagLS	Magnitude Least Squares
MPEG	Moving Picture Experts Group
RM	Reference Model
RT60	Reverberation Time 60dB
SOFA	Spatial Oriented Format for Audio
VBAP	Vector Based Amplitude Panning
VR	Virtual Reality
WD	Working Draft

참고문헌

[1] S.R. Quachenbush and J. Herre, “MPEG standards for compressed representation of immersive audio,” Proc. IEEE, vol. 109, no. 9, 2021, pp. 1578-1589.

[2] S.R. Quachenbush, Report on the 6th Meeting of MPEG Audio Coding, N0111, ISO/IEC JTC1/SC29/WG6, Jan. 2022.

[3] WG6 MPEG Audio Coding, MPEG-I Audio Architecture and Requirements, N0028, ISO/IEC JTC1/SC29/WG6, Jan. 2021.

[4] Audio Subgroup, MPEG-I Audio Architecture and Evaluation for 6DoF, N17177, ISO/IEC JTC1/SC29/WG11, Oct. 2017.

[5] WG6 MPEG Audio Coding, MPEG-I Immersive Audio Encoder Input Format, N0054, ISO/IEC JTC1/SC29/WG6, Apr. 2021.

[6] WG6 MPEG Audio Coding, Report on MPEG-I Immersive Audio Call for Proposals, N0119, ISO/IEC JTC1/SC29/WG6, Jan. 2022.

[7] S. Disch, E. Karlsson, and S. Mate, Description of the MPEG-I Immersive Audio CfP submission of Ericsson, Fraunhofer IIS/AudioLabs and Nokia, M58913, ISO/IEC JTC1/SC29/WG6, Jan. 2022.

[8] J. Koppens et al., Joint CfP proposals by Philips, Dolby and Qualcomm, M58851, ISO/IEC JTC1/SC29/WG6, Jan. 2022.