

특집논문 (Special Paper)

방송공학회논문지 제25권 제5호, 2020년 9월 (JBE Vol. 25, No. 5, September 2020)

<https://doi.org/10.5909/JBE.2020.25.5.723>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

MPEG-I Immersive Audio 표준화 동향

강 경 옥^{a)†}, 이 미 숙^{a)}, 이 용 주^{a)}, 유 재 현^{a)}, 장 대 영^{a)}, 이 태 진^{a)}

MPEG-I Immersive Audio Standardization Trend

Kyeongok Kang^{a)†}, Misuk Lee^{a)}, Yong Ju Lee^{a)}, Jae-hyoun Yoo^{a)}, Daeyoung Jang^{a)},
and Taejin Lee^{a)}

요 약

본 고에서는 현재 탐색단계의 표준화가 진행 중인 MPEG-I Immersive Audio 표준화 동향을 소개한다. 이 표준은 5G/6G와 같은 초연결 환경에서 킬러 어플리케이션으로 기대되는 가상현실(Virtual Reality; VR) 및 증강현실(Augmented Reality; AR)에서, 이용자가 가상환경과 상호작용을 통해 6 자유도(Degrees of freedom; DoF)로 소리를 실감나게 느끼고 실제 환경에서 경험하는 것과 같은 공간음향 체험을 제공하는 것을 목표로 한다. 이를 위하여, MPEG Audio Working Group에서는 가상현실 및 증강현실에서 공간음향 체험을 위한 시스템 구조 및 요구사항을 정의하였다. 이를 기반으로 요구사항에 대한 제안 기술 선정을 위한 오디오 평가 플랫폼(Audio evaluation platform; AEP), 인코더 입력 포맷(Encoder input format; EIF) 및 평가 절차 등에 대한 논의를 진행하고 있으며, 본 고에서는 그 주요 내용을 요약 기술한다.

Abstract

In this paper, MPEG-I Immersive Audio Standardization and related trends are presented. MPEG-I Immersive Audio, which is under the development of standard documents at the exploration stage, can make a user interact with a virtual scene in 6 DoF manner and perceive sounds realistic and matching the user's spatial audio experience in the real world, in VR/AR environments that are expected as killer applications in hyper-connected environments such as 5G/6G. In order to do this, MPEG Audio Working Group has discussed the system architecture and related requirements for the spatial audio experience in VR/AR, audio evaluation platform (AEP) and encoder input format (EIF) for assessing the performance of submitted proponent technologies, and evaluation procedures.

Keyword : MPEG, Immersive audio, Virtual Reality, Augmented Reality

a) 한국전자통신연구원 통신미디어연구소 미디어연구본부 미디어부호화연구실(Media Coding Research Section, Media Research Division, Communication & Media Research Laboratory, ETRI)

† Corresponding Author : 강경옥(Kyeongok Kang)

E-mail: kokang@etri.re.kr

Tel: +82-42-860-5521

ORCID: <https://orcid.org/0000-0002-7673-652X>

※ 본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [20ZH1200, 초실감 입체공간 미디어·콘텐츠 원천기술 연구]

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [20ZH1200, The research of the basic media contents technologies]

· Manuscript received July 20, 2020; Revised August 28, 2020; Accepted August 28, 2020.

Copyright © 2020 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

통신 및 미디어 기술의 발전에 따라 미디어 서비스는 실재감을 제공하는 초실감 미디어 서비스로 진화하고 있다. 이러한 미디어 서비스는 일상생활은 물론 엔터테인먼트, 교육, 의료 등 다양한 산업에 적용되어 사회에 변화를 가져올 것으로 생각된다. 즉, 현실의 모든 사물이 디지털화되어 미디어를 통해 상호 기능적으로 연결되는 서비스가 일상화될 것으로 예상된다. 가상현실(Virtual reality; VR) 및 증강현실(Augmented reality; AR)은 5G/6G 등 초연결 환경에서 킬러 어플리케이션으로 기대되고 있으며, 6DoF (Degrees of freedom) 기반의 초실감 공간표현 및 체험을 가능하게 할 것이다. 또한, 미래사회는 가상의 공간에서 실시간으로 소통하고 체험하는 문화가 주류를 이룰 것으로 예상되는 가운데, 최근 COVID-19와 같은 사회재난 확산으로 디지털 기술을 활용한 비대면 방식의 정보 제공과 커뮤니케이션을 통한 새로운 소통방식에 대한 관심이 더욱 고조되고 있다.

한편, MPEG에서는 VR, AR, 라이트필드 등을 포함한 공간미디어 서비스를 위해 필요한 시스템, 비디오, 오디오 기술을 대상으로 하는 새로운 MPEG-I (Coded Representation of Immersive Media, ISO/IEC 23090) 프로젝트를 2016년 10월에 ISO에 제안하여 2017년 3월에 최종 승인받은 상태이다. 현재 “Immersive Media Architecture” (Part 1), “Versatile Video Coding” (Part 3) 및 “Immersive Audio Coding” (Part 4) 등으로 구성된 총 12개의 파트(Part)에 대

한 표준화가 진행되고 있다. 또한, 이용자에게 제공되는 자유도에 따라 3DoF부터 6DoF까지 단계적인 표준화 로드맵 수립하여 표준화를 진행하고 있다. 그림 1과 같이 제공되는 자유도에 따라 Phase 1(3DoF), Phase 2a(3DoF+) 및 Phase 2b(6DoF)로 단계를 구분하여 각 파트별로 표준화를 진행하고 있다^{[1][2]}.

Phase 1에서는, 의자에 앉아 있는 상황처럼 사용자의 위치가 고정된 상태에서, 각각 X, Y, Z 축을 중심으로 머리를 회전(Pitch, Yaw, Roll)할 수 있는 3DoF의 자유도를 제공한다(그림 1(a)). Phase 2a에서는 3DoF의 머리 회전에 추가로 제한된 범위 내에서 이동(병진 운동)이 허용된 3DoF+의 자유도를 제공한다(그림 1(b)). 한편 Phase 2b에서는 머리 회전과 함께 전후, 좌우, 상하 방향으로 이동이 가능한 것과 같이 사용자의 움직임이 자유로운 6DoF 자유도를 제공한다(그림 1(c)). 다만, 3DoF와 3DoF+까지는 기존의 MPEG-H 3D Audio 표준의 저복잡도(LC) 프로파일로 충분히 서비스 가능하다고 판단하여, MPEG-I Immersive Audio(이하 MPEG-I 오디오로 표기) 표준에서는 Phase 2b의 6DoF를 지원하는 기술에 대한 표준화를 진행하기로 하였다. MPEG-I 오디오 표준화는 현재 탐색 단계(Exploration)에 있으며, 2021년 1월 회의에서 기술제안요구서(Call for Proposals; CfP)를 제정할 예정이다.

MPEG Audio WG (Working Group)에서는 MPEG-I 오디오 표준의 채널, 객체 그리고 앰비소닉스(Ambisonics) 신호의 압축 및 복원을 위하여 MPEG-H 3D 오디오 기술을 사용하기로 하였다. 따라서, MPEG-I 오디오 표준화에서는

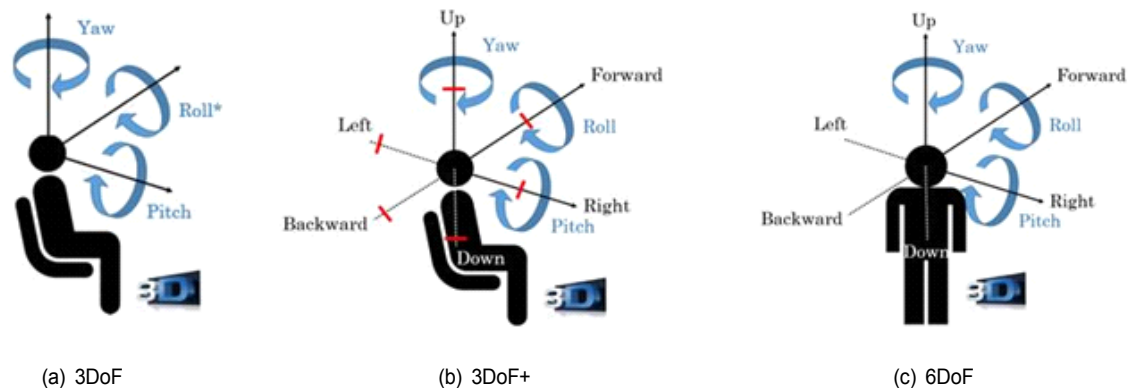


그림 1. MPEG-I 단계별 자유도

Fig. 1. Degree of freedoms defined in MPEG-I according to the standardization phases

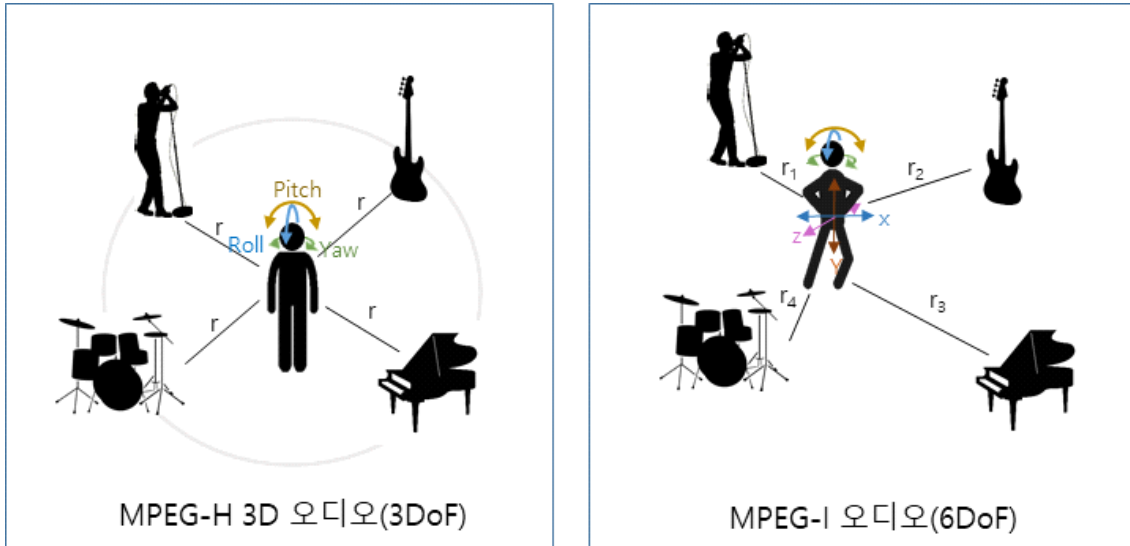


그림 2. MPEG-H 3D 오디오와 MPEG-I 오디오 기술을 이용한 서비스 환경 예(출처: [2])
 Fig. 2. 3DoF and 6DoF services by MPEG standards

6DoF 서비스를 위한 메타데이터(metadata) 및 렌더링(rendering) 기술에 대한 표준화를 진행하기로 하였다. 그림 2는 그림 1의 개념을 오디오 서비스에 적용한 예로서^[2], 전송한 바와 같이 MPEG-H 3D 오디오 기술로는, 사용자는 고정된 자리에서 머리만 각각 X, Y, Z축을 중심으로 움직일 수 있는 상태에서 오디오를 청취할 수 있다(3DoF). 반면에 MPEG-I 오디오 기술을 사용하면 사용자가 머리의 회전뿐만 아니라 공간 상에서 자유롭게 이동하면서 오디오를 청취할 수 있게 된다(6DoF).

본고에서는 MPEG-I 오디오 기술의 표준화 진행 상황에 대해 살펴본다. 이를 위하여 II장에서는 VR/AR 환경에서 6DoF 기반의 공간음향 체험을 위한 MPEG-I 오디오의 구조 및 요구사항을, III장, IV장 및 V장에서는 CfP에 따른 제안기술의 평가를 위한 오디오 평가 플랫폼(Audio evaluation platform; AEP), 인코더 입력 포맷(Encoder input format; EIF) 및 평가 절차에 대해 각각 살펴보고, 끝으로 VI장에서 표준화 전망 등을 포함한 결론을 맺는다.

II. 구조 및 요구사항

MPEG Audio WG에서는, 6DoF 오디오 서비스를 위한

구조로서 MPEG-I 오디오 참조 아키텍처를 정의하고 있으며, 이를 그림 3에 나타낸다^[3].

그림에서 연두색으로 표시된 “MPEG-I Audio Bitstream”, “Common 6DoF Metadata” 및 “MPEG-I 6DoF Audio Renderer”가 MPEG-I 오디오의 표준화 대상 기술이다. Social VR은 가상의 공간에서 두 사용자가 만나는 사용자 시나리오(use case)를 지원하기 위한 기술로, 이를 위한 부호화 기술 대신 social VR 오디오 인터페이스와 관련 메타데이터만 표준화 범위에 포함된다. 공통 6DoF 메타데이터는 신 그래프(scene graph)와 같이 시스템 계층으로 전달되는 오디오 또는 오디오비주얼 엘리먼트(elements)를 기술하는 메타데이터로서, MPEG-I 오디오 비트스트림으로 전달될 수도 있다. 가청화 처리기(auralization processor)는 MPEG-I 6DoF 오디오 렌더러 내에서 실제로 신호처리를 수행하고, 가청화 제어기(auralization controller)도 오디오 렌더러 내에서 오디오 장면 전환과 가청화 처리기의 업데이트 담당하는 블록이다. 미래의 MPEG 기술이 실현할 기능을 제공하거나 외부 기술에 대한 코드 포인터로서 확장툴(extension tools)에 대한 인터페이스도 제공한다.

또한 사용자가 가상 공간에서 움직일 때 현실 공간에서 움직이는 것과 같은 느낌이 들도록 하기 위한 요구사항을 정의하였다. 요구사항에는 가상공간의 사실적 표현 등에

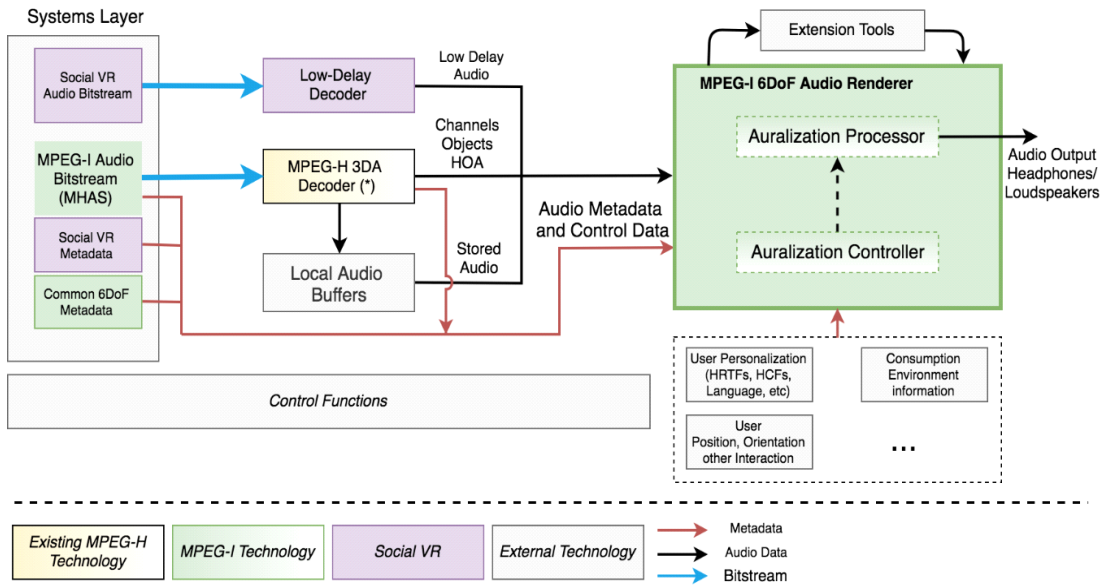


그림 3. MPEG-I 오디오 참조 아키텍처(출처: [3])
 Fig. 3. MPEG-I Audio reference architecture

대한 일반적인 내용, 오디오 렌더러, 인터페이스 및 확장, 재현 모드, Social VR, 재현 환경, 그리고 3DoF와 6DoF 플랫폼간의 호환성 등과 관련된 27개 항목이 정의되어 있다.

III. 오디오 평가 플랫폼 (AEP)

MPEG-I 오디오 표준은 사용자가 6DoF 자유도로 움직일 수 있는 VR/AR 환경에서 오디오 콘텐츠를 표현하기 위한 렌더링 구조 및 부호화에 대해 기술한다. 전통적인 영화와 같은 콘텐츠와는 달리, 6DoF VR/AR의 경우에는 사용자 방위에서의 움직임(Pitch, Yaw, Roll)과 VR 환경에서의 병진 운동(공간 상의 X, Y, Z 축을 따른)이 가능하다. 이러한 기술을 이용한 대표적 사용자 시나리오로, 가상 박물관 여행, 관광 목적지의 사전 가상 투어 등을 생각할 수 있다.

이러한 기술의 성능을 평가하기 위하여(최종적으로 표준 기술 제정을 위하여), MPEG Audio WG에서는 오디오 평가 플랫폼(AEP)^[4]을 개발하고 있으며, AEP의 안정화를 위하여 매 표준화 회의 전에 다수의 기관이 파일럿 테스트를

수행하고 있다. AEP를 이용하여, 사용자는 자유롭게 VR/AR 환경을 탐험하면서 다수의 제안 기술을 동시에 실시간으로 평가할 수 있다. 평가 플랫폼을 위해 VR 동작이 가능한 PC, HTC Vive[®]와 같은 VR 시스템 및 헤드폰 등을 포함한 하드웨어 스펙, Unity3d[®], Max/MSP[®] 등을 포함한 소프트웨어 스펙 등을 정의하고 있다. 그림 4는 오디오 평가 플랫폼의 개요도로서, 평가에 필요한 모든 요소를 보여주고 있다. 그림에서 확인할 수 있듯이 오프라인 인코딩/처리(좌)와 평가를 위한 실시간 AEP 자체(우)의 2개의 분야로 나누어 질 수 있다. AEP는 Unity3d[®]와 Max/MSP[®]를 활용하여 구현하고 있는데, 주요 구성요소로서는 테스트 형상 파일(test config file: TCF), 음원 파일(WAV), PCM 데이터, 음원의 지향성(directivity) 파일, 비트스트림 등을 들 수 있다. 향후 CFP가 제정되면 기술을 제안하고자 하는 기관은 MPEG-I 인코더(그림에서의 MPEG-I Encoder A, B, C, ...) 및 Max external로 구현된 렌더러(그림에서의 Renderer A, B, C, ...)를 제출해야 한다. 인코더는 6DoF를 제공하는 메타데이터를 압축하는 기능을 수행하여, 렌더러는 Unity3d[®] 및 Max/MSP[®]와 연동하여 오디오 신호를 실시간으로 렌더링하는 역할을 수행한다.

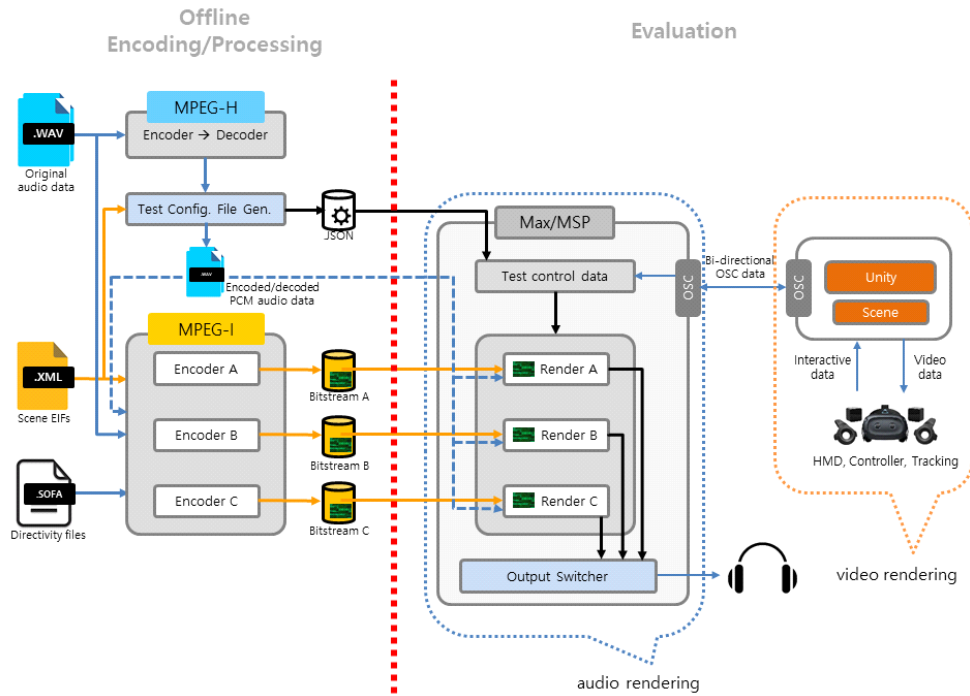


그림 4. 오디오 평가 플랫폼 및 프로세스 개요 (VR 경우) (출처: [4])
 Fig. 4. Overview of audio evaluation platform and process (only in VR case)

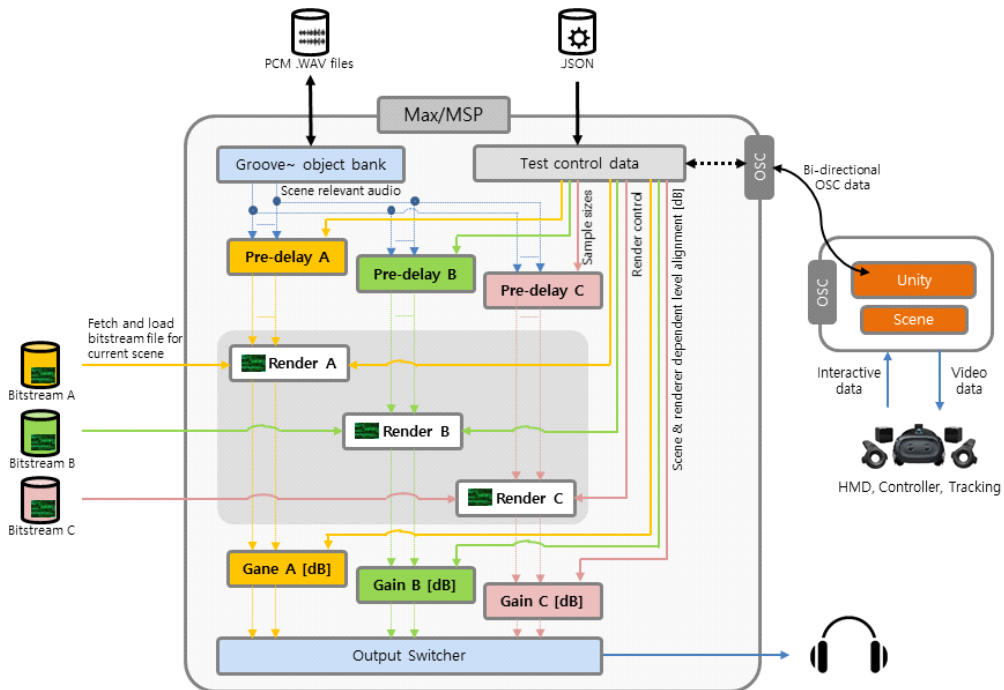


그림 5. MAX/MSP 내부 동작 (출처: [4])
 Fig. 5. Internal workings of Max/MSP

그림 5는 그림 4의 좌측의 비트스트림이 전달되었을 경우에 Max/MSP[®] 내에서의 구체적인 데이터 흐름 등을 포함한 내부 동작을 좀 더 자세하게 나타낸다. Unity3d[®]는 영상 및 사용자 인터페이스 제어 등을 수행하는 그래픽 렌더링 엔진으로 사용되고, Max/MSP[®]는 오디오 렌더링 엔진으로 사용된다. Max/MSP[®]는 테스트 구성 설정, 오디오, 데이터 처리 및 모니터링 등을 수행하며, CIP에 따라 제안된 각 기관별 렌더러를 Max External 형태로 포함하여 오디오 렌더링을 수행하고, 평가방법에 따른 주관평가를 수행하여 각 기관별 렌더러의 성능평가를 수행한다. 이때, 각 기관에서 제안한 렌더러 별로 신호에 대한 지연시간 및 신호레벨에 차이가 있을 수 있기 때문에 이를 보정하기 위한 기능으로 각각의 렌더러에 대한 지연시간 보정 및 이득 조정 등의 기능도 포함하고 있다.

IV. 인코더 입력 포맷 (EIF)

MPEG Audio WG에서는 MPEG-I 오디오 표준화를 위한 CIP 제안 기술 등의 평가 목적으로 인코더 입력 포맷 (EIF)을 규정하고 있다^[5]. EIF는 오디오 장면을 표현하는데 필요한 메타데이터로서, MPEG-I 오디오 인코더에의 입력 형식을 정의하고 있는 XML 파일이다. 인코더 입력 데이터는 MPEG-I 6DoF 오디오 장면(scene)을 기술하는 정보를 포함하고 있으며, 이는 모든 가상 청각(auditory) 장면 콘텐츠, 즉 장면에 포함된 모든 사운드 소스 및 리소스 데이터(오디오 waveforms, 소스의 방사 패턴, 음향 환경에 대한 정보 등)를 포함한다. 입력 데이터는 또한 장면 내에서의 변화(changes)를 기술할 수 있는데, 업데이트(updates)이라고 하는 이러한 변화는 서로 다른 시간에 발생하여 장면에 적용할 수 있다. 이외에도 수동 또는 어떤 조건(예로 청취자가 특정 객체 근처에 근접하는 것과 같은)에 의해 트리거되거나 외부 엔터티(entity)에 의해 동적으로 업데이트될 수 있다.

1. 장면 메타 모델

EIF 문서에서는 인코더 입력 포맷에서 사용할 메타 모델

을 그림 6과 같이 정의하고 있으며, 주요사항은 다음과 같다.

- 오디오 장면은 엔터티로 구성(sources, geometry, resources 등)
- 엔터티는 ID (string)로 식별(ID 포맷은 임의이나 유일해야 함)
- 엔터티는 장면의 모든 시간에 걸쳐 존재하며, 런 타임 중에 동적으로 생성되거나 제거되지 않음
- 엔터티는 이들 ID를 사용하는 다른 것들과 관련될 수 있음(예로, 지향성을 갖는 소스)
- 엔터티는 속성(property)을 가짐(예로, position, volume, material 등)
- 엔터티는 특정 엔터티 형식(entity type)을 가지며, 선언에 의해 결정됨
- 엔터티 형식에 따라 엔터티의 속성 및 속성값의 변경 여부가 결정됨

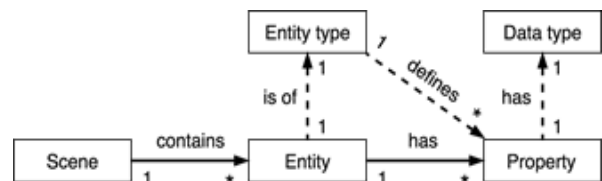


그림 6. 장면 메타 모델 (출처: [5])
Fig. 6. Scene meta model

장면내에서의 변화는 업데이트로 표현하며, 다음과 같이 정의한다:

- 업데이트는 특정 시간에 발생하고, 하나 이상의 수정(modification)을 포함
- 업데이트는 조건에 의해 시작될 수 있고, 외부 엔터티에 의해 수동으로 트리거되거나 실제 동작 시 다이나믹하게 업데이트 될 수 있음
- 수정은 엔터티의 한 개 또는 그 이상의 속성을 변경하지만, 모든 속성을 변경할 수는 없음(ID 변경 불가)
- 업데이트는 서로 다른 엔터티에 대한 여러 개의 수정을 묶어 처리할 수 있음

2. 장면 선언 및 업데이트

EIF 파일은 단일 `<AudioScene>` 노드로 구성되며, 여기에서 모든 엔터티가 선언되고, 업데이트가 정의된다. 현재 다음과 같은 형태의 업데이트가 정의되어 있다.

- **Timed update (L1):** 미리 정의된 특정 시간에 렌더러에 의해 한번 시행
- **Conditional update (L2):** 측정 조건이 충족될 때 렌더러에 의해 개시
- **Triggered update (L2):** 외부 엔터티로부터 OSC (Open

sound control) 메시지를 이용하여 수동으로 트리거됨

- **Dynamic update (L3):** 업데이트 할 속성값을 포함한 외부 엔터티로부터의 업데이트

업데이트에 의해 수정되는 속성(attributes)과 값(values)은 EIF 파일에 정의되고 인코더에 알려진다. 유일한 예외는 dynamic update로, 수정될 속성 값을 위한 매개 변수(parameter variables)를 삽입할 수 있다. 이 값은 Unity로부터 업데이트 메시지에 의해 렌더러에게 제공되며, 매개 변수는 '\$' 다음에 숫자로 표시된다. Cfp의 경우, dynamic update (L3)는 최대 3개의 오디오 엘리먼트의 위치, 방향 또는

표 1. `<Update>` and `<Modify>` nodes

Table 1. `<Update>` and `<Modify>` nodes

<code><Update></code>				
Declares one or more changes to the audio scene. The update is performed, when the specified time is reached, or the condition changed its state to the logical value expressed by fireOn, the update is triggered by its ID or index by an external entity				
The <code>fireOn</code> parameter determines whether the update fires when the condition changes from false-to-true (<code>fireOn="true"</code>) or from true-to-false (<code>fireOn="false"</code>). This is helpful for if-else type conditional updates.				
An <code><Update></code> node has one or more <code><Modify></code> child nodes.				
Child node	Count	Description		
<code><Modify></code>	<code>>=1</code>	Modifications (see below)		
Attribute	Type	Flags	Default	Description
<code>id</code>	ID	O*	none	Identifier
<code>index</code>	Integer	O*	none	Index identifying the update (globally unique)
<code>time</code>	Value	O*	none	Time when the update is initiated (seconds). Note: Must be less than or equal to the duration attribute of the AudioScene, and the update doesn't affect the scene before this time stamp.
<code>condition</code>	Condition ID	O*	none	Condition
<code>fireOn</code>	Boolean	O*	true	Update fires when this state is reached
<code>delay</code>	Float <code>>= 0</code>	O*	0	Postpone the update (seconds)
<code><Modify></code>				
Declares a modification of modifiable parameters of a single entity. The target entity is selected by the <code>id</code> attribute. Following attributes must be attributes of the corresponding entity. The attribute values are assigned the entities property values.				
For scene updates where attribute values are not available until rendering, typically user interactions, the attribute values are replaced by parameter indicators to specify in which order the attribute values are expected on the renderer interface. Parameter indicators are \$0, \$1, \$2, etc.				
When the target entity also has attributes 'transition' or 'duration' (see below), these can be modified by specifying them two times in the modification. The first occurrence controls the modification parameter, while the second marks the destination value of the entities' property.				
Example:				
<code><Modify id="src1" position="1 2 3" orientation="-20 5 0" /></code>				
sets the attributes <code>position</code> and <code>orientation</code> for the entity with ID <code>src1</code>				
Attribute	Type	Flags	Default	Description
<code>id</code>	ID	R		Target entity to be modified
<code>duration</code>	Float <code>>= 0</code>	O	0	Period for adapting from the current values to the new values (seconds)
*	*	*	*	Attribute of the target entity

O(Optional)

표 2. 업데이트 형태에 따른 허용가능한 속성 및 변수값 (출처: [5])

Table 2. Allowed attributes and variables for different types of updates

Update type	Attributes id, index	Attribute time	Attributes condition, fireOn, delay	Parameter variables \$0, \$1, \$2, ...
Timed update	optional	required	not allowed	not allowed
Conditional update	optional	not allowed	required	not allowed
Triggered update	required	not allowed	not allowed	not allowed
Dynamic update	required	not allowed	not allowed	present

이득만을 수정할 수 있고, transform 엔터티 내부에 존재하는 오디오 엘리먼트는 해당하지 않는다.

예로서 <Update> 노드에 대한 속성값은 표 1과 같다. <Update> 노드는 서로 다른 업데이트 형태를 명확하게 규정할 수 있도록 특별한 방법으로 사용하여야 한다.

위에서 언급한 업데이트 형태에 따른 허용 가능한 속성 및 변수값에 대해 표 2에 기술한다.

3. 엔터티 형식

EIF에서는 MPEG-I 요구사항서^[1] 기반으로 엔터티와 그 속성에 대해 정의한다.

대표적으로 아래와 같은 엔터티로 구분되며, 각각의 엔터티에 대한 노드 이름 및 의미(또는 활용 용도)를 간략하게 기술한다.

- Audio scene
 - 노드명: <AudioScene>
 - 의미(역할): 임의의 장면(scene)을 기술하는 XML 파일은 루트 노드에 해당하는 <AudioScene> 노드를 정확히 한 개 포함해야 함
- Audio elements
 - <ObjectSource>, <HOASource>, <ChannelSource>
 - 가상환경에서 재생해야 하는 사운드 소스를 의미하며, 객체신호(object), 앰비소닉스(Higher-order Ambisonics; HOA)^[6] 및 채널신호(channel)가 존재함
- Spatial transforms
 - <Transform>
 - 다수의 공간상의 엔터티를 함께 이동하거나 회전시킬

목적으로 <Transform> 노드를 정의하고 변환좌표계(Transform coordinate system)를 정의함

- Conditions
 - <ListenerProximityCondition>
 - “conditional update”를 위한 조건을 정의
- Audio resources
 - <AudioStream>
 - 로컬 오디오 파일(음원)에 의해 제공되는 오디오 스트림으로서, Audio elements 엔터티의 신호 소스로 사용되며, 원음 신호와 MPEG-H 3D Audio LC 프로 파일을 이용해 부복화된 신호가 있음
- Source radiation pattern
 - <SourceDirectivity>
 - 특정 방향과 주파수에서의 ObjectSources의 소리 방사 패턴(directionality)을 정의
- Geometric elements (Primitives, Meshes)
 - <Box>, <Sphere>, <Cylinder>, <Mesh>
 - 회절, 반사, 장애물 등 공간에서의 실내음향 효과를 표현하거나, 점 음원이 아닌 공간상의 폭(width)을 가진 ObjectSources를 표현하기 위한 기하적인 정보로 사용
- Acoustic materials
 - <AcousticMaterial>
 - 반사계수, 투과계수 등의 공간상의 매질특성을 기술
- Acoustic Environments
 - <AcousticEnvironment>
 - pre-delay, 잔향시간, Diffuse-to-Direct ratio (DDR)와 같은 직접음에 대한 반사비율 등의 음향적인 조건을 기술

V. 평가 절차

MPEG Audio WG에서는 기술 평가를 위하여 적합한 테스트 장면을 각 기관으로부터 제안받고 있으며, 현재 다수의 기관으로부터 객체신호, 채널신호 또는 HOA신호로 구성된 다양한 테스트 장면이 제안되어 있는 상황이다. 뿐만 아니라, CFP에 따른 제안된 기술에 대한 평가 절차 또한 정의하고 있다. 이에 따르면, MPEG-I 오디오 응용으로는 VR 및 AR을 고려하고 있기 때문에 아래와 같이 3가지 유형의 테스트를 검토하고 있고, 주관평가를 위한 테스트 방법으로도 3개의 방법을 검토하고 있는 단계이며, 추가적인 논의를 통하여 가장 최적의 방법이 선택될 예정이다^[7].

- 테스트 유형
 - Test 1: VR with Objects, Channels, 3DoF HOA
 - Test 2: AR with Objects, Channels
 - Test 3: VR with 6DoF HOA (interior/exterior, multi-point), Objects, Channels
- 테스트 방법(주관평가)
 - MUSHRA-VR: MPEG 오디오 부호화 표준 제정을 위하여 전통적으로 사용해 오던 방식인 MUSHRA (Multi Stimulus test with Hidden Reference and Anchor)를 VR 테스트에 맞게 변형한 것으로서, 기준 신호(Hidden reference) 없이 주관평가 대상(일반적으로 stimulus라고 하며, 본 표준의 경우에는 렌더러)이 제공하는 주관적인 품질을 평가하는 방법^[8]
 - A-B: CCR (Comparison Category Rating) 기반의 평가 방식으로, 두 개의 자극(double stimuli)을 쌍(pair)으로 제시하여 2개의 자극의 상대적 차이를 주관적으로 평가하는 방법^[9]
 - MAACR (Multiple Attributes/Absolute Category Rating): 주관적인 품질에 영향을 미치는 다양한 속성^{[10][11]}(attributes, 예로서 MPEG-I 오디오의 경우에, basic audio quality, externalization, consistency, plausibility 등)을 평가할 때, 각각의 속성에 대해 독립적으로 한번에 하나의 자극만을 제시하여 절대적으로 평가하는 방법^[9]

이와 같은 주관(subjective) 평가 외에 객관(objective) 평

가 방법 또한 정의하고 있으며, 주요 대상으로는 계산 복잡도, 지연시간(신호 지연뿐만 아니라 사용자 상호작용에 따른 사운드 지연인 motion to sound latency), 비트율 및 메모리 사용량 등이 고려되고 있다. 정의된 요구사항 항목에 따라서는 주관평가만으로 만족여부를 판단할 수 없기 때문에, 일부 요구사항에 대해서는 각 제안 기술이 이를 만족하는지 여부를 기술하는(descriptive input) 부분도 포함하고 있다.

통상적으로 오디오 표준화의 경우에는 CFP 제안기술에 대한 평가를 통하여 1위 기술을 선정하며, 이를 RM0 (reference model 0)로 선정한다. 그후 CE (core experiment) 과정을 통하여 최종 표준기술을 선정하고 있다. 이에 평가 절차에서는 RM0를 선정하기 위한 기준 및 구체적인 CE 방법에 대해서도 논의하고 있다.

VI. 결론

본고에서는 현재 MPEG에서 진행 중인 MPEG-I Immersive Audio 표준화 동향에 대해 살펴보았다. MPEG Audio WG에서는 6DoF 오디오 서비스를 위한 메타데이터 및 렌더링 기술을 MPEG-I 오디오의 표준화 대상으로 결정하였다. 이를 위하여 각 기관으로부터 기술제안을 받기 위해 필요한 문서 중 구조 및 요구사항 등의 문서는 완료되었으나, AEP, EIF, 평가절차 및 CFP 문서 등은 계속 논의 중에 있다. CFP의 경우에는 현재 초안(draft)^[12] 상태이며 2021년 1월 회의에서 제정할 예정이지만, 현재의 표준화 일정이 COVID-19로 인하여 영향을 받은 상황이라 확정이라고 단정할 수는 없다.

5G/6G 등 초연결 환경에서의 킬러 어플리케이션으로 기대되고 있는 VR/AR 등이 관심을 받고 있는 상황에서 최근 COVID-19와 같은 사회재난 확산으로 디지털 기술을 활용한 비대면 방식의 정보 제공 기술에 대한 관심이 더욱 고조됨에 따라, 몰입형 미디어 시장에 대한 관심과 이에따른 시장 성장은 더욱 가속화될 것으로 예상된다. 또한, 미디어 기술은 일상생활은 물론 엔터테인먼트, 교육, 의료 등 다양한 산업에 적용될 수 있어 사회 경제적 영향을 무시할 수 없다. 따라서 호환성을 중요하게 생각하는 표준의 중요성을 고려할 때, 현재 진행 중인 MPEG-I Immersive Audio의 표준에 대한 지속적인 관심이 요구된다.

참 고 문 헌 (References)

- [1] ISO/IEC JTC1/SC29/WG11/N17685, "Proposed Draft 1.0 of TR: Technical Report on Architectures for Immersive Media", April 2018.
- [2] Misuk Lee, Yongju Lee, Taejin Lee, "Audio standard for immersive media service - MPEG-H 3D Audio and MPEG-I Immersive Audio", TTA Journal, Vol. 185, pp. 26~32, October 2019. *(in Korean)*
- [3] ISO/IEC JTC1/SC29/WG11/N18518, "MPEG-I Audio Architecture and Requirements", Jan. 2019.
- [4] ISO/IEC JTC1/SC29/WG11/N19411, "Documentation for the MPEG-I Audio Evaluation Platform", July 2020.
- [5] ISO/IEC JTC1/SC29/WG11/N19409, "Draft MPEG-I 6DoF Audio Encoder Input Format", July, 2020.
- [6] Y.J. Lee, J. Yoo, D. Jang, M. Lee, T. Lee, "Spatial Audio Technologies for Immersive Media Service", ETRI Electronics and Telecommunications Trends, Vol. 34, pp. 14~22, June 2019. *(in Korean)*
- [7] ISO/IEC JTC1/SC29/WG11/N19502, "Draft MPEG-I Audio Test and Evaluation Procedures", July, 2020.
- [8] Recommendation ITU-R BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems", 2001~2003.
- [9] Recommendation ITU-T P.913, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment", March, 2016.
- [10] Ulrike Sloma *et.al.*, ISO/IEC JTC1/SC29/WG11/MPEG2020/m52368, "Thoughts on test methods for MPEG-I audio", Brussels, January 2020.
- [11] Recommendation ITU-R BS.1284-2, "General methods for the subjective assessment of sound quality", January 2019.
- [12] ISO/IEC JTC1/SC29/WG11/N19412, "Draft Call for Proposals for MPEG-I Immersive Audio", July, 2020.

저 자 소 개

강 경 옥



- 1985년 2월 : 부산대학교 물리학과 (이학사)
- 1988년 2월 : 부산대학교 물리학과 (이학석사)
- 2004년 2월 : 한국항공대학교 항공전자공학과 (공학박사)
- 2006년 4월 ~ 12월 : 영국 University of Southampton ISVR 방문연구원
- 2014년 8월 ~ 2015년 7월 : KAIST 문화기술대학원 방문연구원
- 1991년 2월 ~ 현재 : ETRI 미디어부호화연구실 책임연구원
- ORCID : <https://orcid.org/0000-0002-7673-652X>
- 주관심분야 : 실감음향, 음향신호처리, MPEG 오디오 표준화

이 미 숙



- 1991년 2월 : 호서대학교 전자공학과(공학사)
- 1993년 2월 : 호서대학교 전자공학과(공학석사)
- 2001년 2월 : KAIST(공학박사)
- 2002년 2월 ~ 현재 : ETRI 책임연구원
- ORCID : <https://orcid.org/0000-0003-4295-7575>
- 주관심분야 : 음성/오디오 신호처리

이 용 주



- 1999년 2월 : 경북대학교 전자공학과 (공학사)
- 2001년 2월 : 경북대학교 전자공학과 (공학석사)
- 2001년 2월 ~ 현재 : ETRI 오디오연구실 선임연구원
- ORCID : <https://orcid.org/0000-0003-0753-1616>
- 주관심분야 : 오디오 신호처리, 바이노럴 오디오 렌더링

저 자 소 개



유 재 현

- 2003년 2월 : 홍익대학교 전자전기공학부 학사
- 2005년 2월 : 서울대학교 대학원 전기컴퓨터공학부 석사
- 2015. 9월 ~ 현재 : KAIST 전기및전자공학부 박사과정
- 2005년 2월 ~ 현재 : ETRI 오디오연구실 선임연구원
- 2011년 11월 ~ 2012년 11월 : 일본 NHK 방송기술연구소 방문연구원
- ORCID : <https://orcid.org/0000-0002-9326-9353>
- 주관심분야 : 3D 오디오 및 음장 재생 기술, 오디오 신호 처리



장 대 영

- 1991년 2월 : 부경대학교 전자공학과 (공학사)
- 2000년 2월 : 배재대학교 컴퓨터공학과 (공학석사)
- 2008년 2월 : 배재대학교 컴퓨터공학과 (공학박사)
- 2004년 10월 ~ 2005년 9월 : 일본 동경전기대학/DiMagic Co. Ltd. 방문연구원
- 2019년 4월 ~ 2020년 3월 : 호주 울런공대학교 방문연구원
- 1991년 1월 ~ 현재 : ETRI 미디어부호화연구실 책임연구원
- ORCID : <https://orcid.org/0000-0003-2572-4374>
- 주관심분야 : 실감음향, 객체기반 오디오, 디지털 방송, 대화형 미디어



이 태 진

- 2014년 : 충남대학교 전자전파정보통신공학과 공학박사
- 2002년 ~ 2003년 : 일본 Tokyo Denki University, 방문연구원
- 2000년 ~ 현재 : ETRI 미디어부호화연구실 실장
- ORCID : <https://orcid.org/0000-0002-6067-511X>
- 주관심분야 : 오디오 부호화, 실감음향, 오디오 신호처리