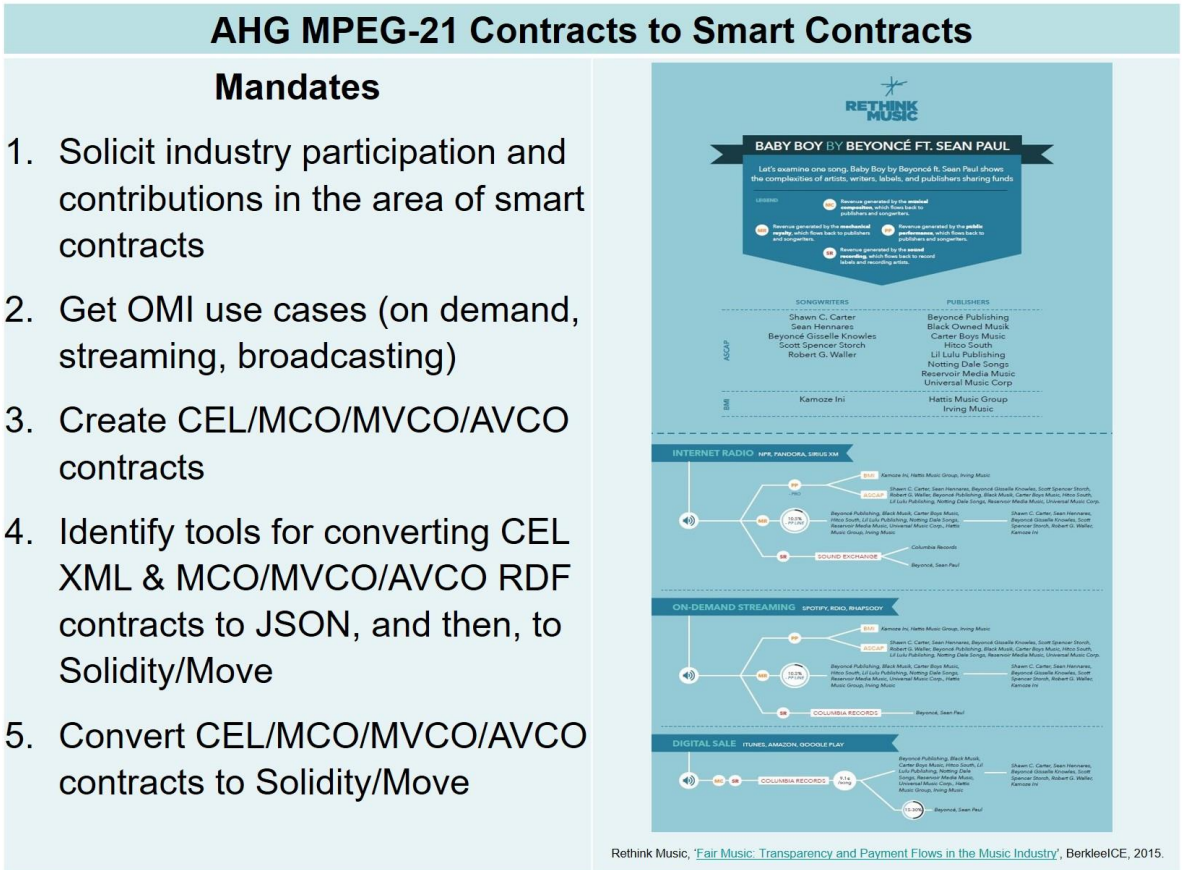
127th MPEG Gothenburg, Sweden, 8 - 12 July 2019, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 Smart Contracts

Under the auspices of the Requirements subgroup an exploration activity and an Ad Hoc Group (AHG) have been established on MPEG-21 Based Smart Contracts with the following mandates:



Rethink Music, 'Fair Music: Transparency and Payment Flows in the Music Industry', BerkleeICE, 2015.

Chairmen	Panos Kudumakis (QMUL), Xin Wang (MediaTek)
Duration	Until next meeting
Reflector(s)	smart-contracts@lists.aau.at
Subscribe	https://lists.aau.at/mailman/listinfo/smart-contracts

References

- Panos Kudumakis, Thomas Wilmering, Mark Sandler, Víctor Rodríguez-Doncel, Laurent Boch and Jaime Delgado, "MPEG IPR Ontologies", ISO/IEC JTC1/SC29/WG11/N18500, Geneva, CH, Mar. 2019.
- Panos Kudumakis, Thomas Wilmering, Mark Sandler and Jeremy Foss, "MPEG IPR Ontologies for Media Trading and Personalization", 1st International Workshop on Data-driven Personalization of Television (DataTV'19) at ACM International Conference on Interactive Experiences for Television and Online Video (TVX'19), Manchester, UK, 5-7 Jun. 2019.

Standards

Acronym	Standard	MPEG Document	Reference Software
MVCO	ISO/IEC 21000-19, 'Information technology -- Multimedia framework (MPEG-21) -- Part 19: Media value chain ontology', June 2010.	N11146 91st Kyoto	N/A
	ISO/IEC 21000-8/AMD2, 'Information Technology -- Multimedia Framework (MPEG-21) -- Part 8: Reference software / AMD2 Reference software for media value chain ontology', Nov. 2011.	N12135 97th Torino	https://tinyurl.com/y6tsr9as
AVCO	ISO/IEC 21000-19:2010/AMD1, 'Information Technology -- Multimedia Framework (MPEG-21) -- Part 19: Media Value Chain Ontology / AMD 1 Extensions on Time-Segments and Multi-Track Audio', June 2018.	N17170 120th Macau	N/A
	ISO/IEC 21000-8:2008/AMD4, 'Information Technology -- Multimedia Framework (MPEG-21) -- Part 8: Reference Software / AMD 4 Media Value Chain Ontology Extensions on Time-Segments and Multi-Track Audio', Oct. 2018.	N17404 121th Gwangju	https://standards.iso.org/iso-iec/21000/-8/ed-2/en/amd/4
MCO	ISO/IEC 21000-21 **(2nd Ed.)**, 'Information technology -- Multimedia framework (MPEG-21) -- Part 21: Media Contract Ontology', May 2017.	N15940 114th San Diego	https://standards.iso.org/iso-iec/21000/-21/ed-2
CEL	ISO/IEC 21000-20 **(2nd Ed.)**, 'Information technology -- Multimedia framework (MPEG-21) -- Part 20: Contract Expression Language', Dec. 2016.	N15994 114th San Diego	Included in N15994

2 Network-based Media Processing

Scope

The Network-based Media Processing framework defines the interfaces including both data formats and APIs among the entities connected through the digital networks for media processing. Users can access and configure their operations remotely for efficient intelligent processing. It describes and manages workflows to be applied to the media data. This process includes uploading of media data to the network, instantiation of the media processing tasks, and configuration of the tasks. The framework enables dynamic creation of media processing pipelines, access of processed media data and metadata in real-time or in a deferred way. The media and metadata formats used between the Media Source, Workflow Manager and Media Processing Entities in a media processing pipeline are also within the scope.

Introduction

Network-based Media Processing defines a framework that enables initializing and controlling media processing in the network. An NBMP Source describe the requested media processing and provide information about the nature and format of their media data. Based on that, an NBMP Workflow Manager establishes the media processing workflow and informs the NBMP Source that the workflow is ready and that media processing can start. The Media Source(s) can then start transmitting their media to the network for processing.

An NBMP Workflow can be understood as a connected graph of media processing Tasks, each of which performs a well-defined media processing operation. The Workflow Manager ensures the correct operation of the Workflow by configuring and monitoring each Task as well as the Workflow output. The Workflow manager is responsible for the selection of the media processing Functions and instantiating them as Tasks based on the Workflow Description that is received from the NBMP Source.

NBMP abstracts the underlying computing platform interactions to establish, load, instantiate and monitor the media processing entities that will run the media processing tasks. NBMP defines APIs between NBMP Source and Workflow Manager, Workflow Manager and Task(s), and an API to discover appropriate Function(s). NBMP is media format and protocol agnostic. However, it identifies and signals the media, metadata, and auxiliary information formats for data exchanged between Media Source, Workflow Manager and Tasks.

General

The Network-Based Media Processing (NBMP) framework enables the creators, service providers and consumers of digital media to describe media processing operations that are to be performed by the media processing entities in the network as shown in Fig. 1. It provides a method to describe a workflow by composing a set of the media processing functions that are accessible through NBMP Application Programming Interfaces (APIs). A Media Processing Entity contains processing tasks applied on the media data and the related metadata received from Media Sources or other Media Processing Entities. A Media Processing Entity provides control capabilities that configure and manage the processing tasks. A Media Processing Task produces media data and related metadata to be consumed by a Media Sink or other Media Processing Tasks.

The NBMP framework supports any format of media content, including the existing MPEG codecs and MPEG formats such as ISO/IEC 13818-1 (MPEG-2 TS), ISO/IEC 14496-12 (ISOBMFF), ISO/IEC 23008-1 (MMT) and ISO/IEC 23009-1 (DASH).

The NBMP framework supports the delivery over IP-based networks using common transport protocols such as TCP, UDP, RTP and HTTP.

The NBMP framework also support the existing delivery methods such as streaming, file delivery, push-based progressive download, hybrid delivery, multipath and heterogeneous network environments.

Architecture

NBMP specifies interfaces to create and control media processing workflows in the network. NBMP can be split into a control plane and a media plane. The control plane covers the following APIs:

- Workflow API is used by NBMP Source to create and control a media processing workflow
- Function Discovery API provides the means for Workflow Manager and/or NBMP Source to discover media processing Functions that can be loaded as part of a media processing Workflow.
- Task API is used by the Workflow Manager to configure and monitor a Task at runtime.

On the media plane, NBMP defines the media formats, the metadata, and the supplementary information formats between the NBMP Source and the Task, as well as between the Tasks themselves. The discovery of NBMP Workflow Manager and Function Repository is out of scope. The NBMP architecture is depicted by the following diagram as shown in the Figure 1.

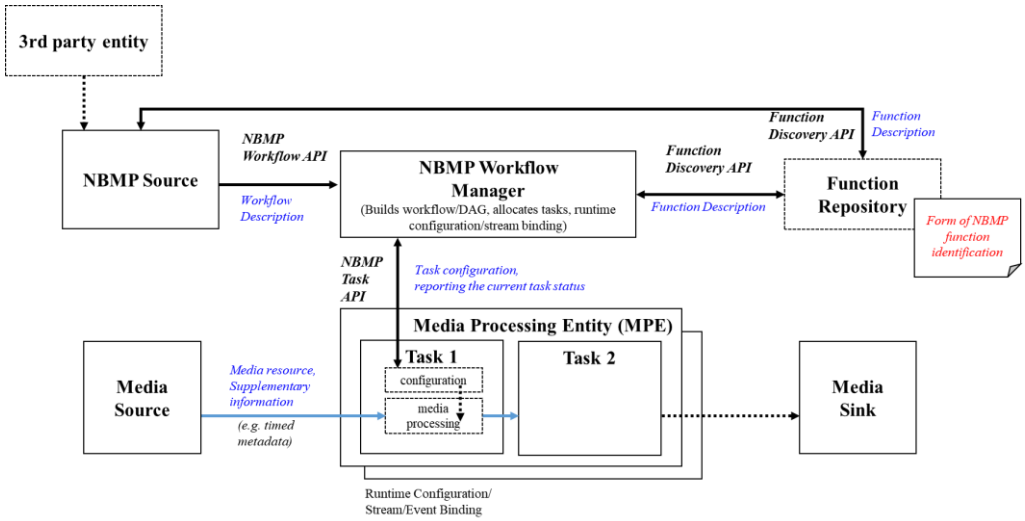


Figure 1 - NBMP Reference Architecture

Use Cases

Use Case #1: 360 Stitching
Use Case #2: 6DoF Pre-rendering
Use Case #3: Guided Transcoding
Use Case #4: E-Sports Streaming
Use Case #5 : OMAF Packager
Use Case #6 : Measurement (client preferences or network conditions)

NBMP and Big Data

Towards a possible collaboration between SC 29/WG 11 (MPEG) and SC 42/WG 2 (Big Data), an earlier contribution by JTC 1/WG 9 N16 on NIST Big Data Reference Architecture (NBD-RA) discussed. The outcome of the discussion is documented in form of a draft mapping from NBMP architecture to the NBD-RA, which is depicted in Fig. 2. That is, Big Data processing may be implemented as NBMP functions and deployed as NBMP Tasks on top of the Big Data Framework provider. The communication between the system orchestrator and the Big Data processing may be provided by the Task API. The interface between the Data Provider and the Big Data Application Provider is mapped to the NBMP Workflow API.

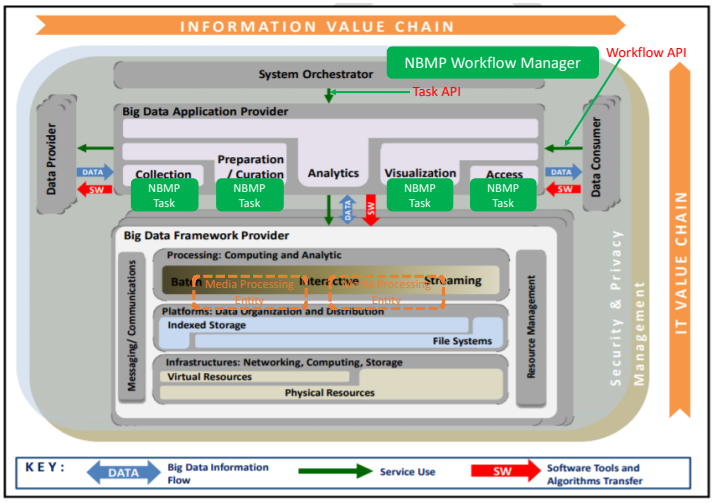


Figure 2 - Draft NBMP mapping into JTC1/WG9 N16 - NIST Big Data Reference Architecture (NBD-RA)

Output Documents
N18605 - DoC on ISO/IEC CD 23090-8 Network-based Media Processing
N18657 - Text of ISO/IEC DIS 23090-8 Network-based Media Processing
N18658 - Technologies under Considerations for NBMP
N18659 - Description of Core Experiments on Function Template validation
N18660 - Workplan for reference software and conformance on Network-based Media Processing
N18661 - Comparison of NBMP architecture and Big Data Reference Architecture (ISO/IEC 20547-3)

3 Immersive Audio

Fig. 3 shows the MPEG-I Audio Architecture. MPEG-I Audio Requirements followed by the draft Audio Evaluation Platforms (AEPs) for VR & AR are briefly described next.

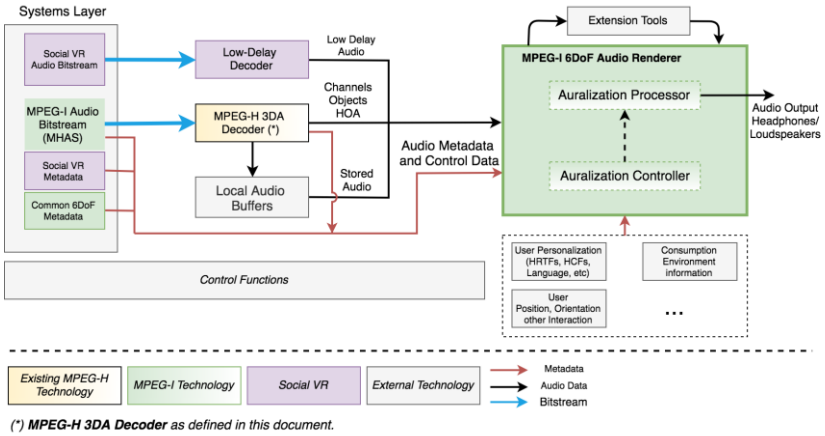


Figure 3 - MPEG-I Audio Reference Architecture

- MPEG-I Audio technology appearing in green is the technology to be standardized.
- MPEG-H 3DA Decoder is defined in the definitions section.
- MPEG-I Audio specification will not address the audio coding for Social VR, however the interface for Social VR audio and associated metadata must be standardized.
- The Common 6DoF Metadata represents metadata describing audio-only or audio-visual elements which may be delivered on systems layer (e.g., Scene Graph). The same metadata could be delivered in the MPEG-I Audio bitstream.
- Control Functions may contain functionality which will become clear as the standard is developed.
- Auralization Controller is a possible block within the audio renderer that handles audio scene changes and updates the underlying Auralization Processor.
- Auralization Processor is a possible functional block within the audio renderer taking care of the actual signal processing.
- Extension Tools can provide functionality realized by future MPEG technology or code points to external technology.

General Requirements

1. The specification shall support user 6DoF (Degrees of Freedom) so that a single user perceives an experience consistent with user's movement in the environment (e.g., low and non-perceivable motion-to-sound latency).

2. The specification shall support efficient representation and compression of media and metadata.

   2.1. Media coding shall be done according to the MPEG-H 3D Audio Low Complexity Profile (including 3D Audio metadata). The specification shall support any combination of channel-based, object-based and HOA content types.

   Note: Multiple HOA streams (e.g., sampled at different locations) may be supported simultaneously in one Audio Scene.

   2.2. The specification shall support Loudness and Dynamic Range Control using MPEG-H 3D Audio tools.

   2.3. The specification shall support additional metadata as needed to support user 6DoF.

   2.4. The specification shall support delivery of the audio scene in multiple audio streams.

   Note: This may already be fulfilled by MPEG-H 3D Audio.

3. The specification shall support rendering of the audio scene, consistent with the acoustic elements and acoustic environments, resulting in a user experience consistent with the scene.

   3.1. [SYS] Audio elements shall be rendered consistently with their corresponding visual elements, if such visual elements exist.

   3.2. The specification shall support signalling of audio elements that have a fixed position relative to the user orientation and position (e.g., non-diegetic content).

   Note: This may already be fulfilled using MPEG-H 3D Audio.

   3.3. The specification shall support earcons.

   Note: This may already be fulfilled using MPEG-H 3D Audio.

4. The specification shall support dynamic inclusion of audio elements in a sub-scene based on their relevance, e.g., audibility relative to the user location, orientation, direction and speed of movement or any other audio scene change.

   4.1. The specification shall support metadata to allow fetching of relevant sub-scenes, e.g., depending on the user location, orientation or direction and speed of movement.

   Note: A complete audio scene may be divided into a number of audio sub-scenes, defined as a set of audio elements, acoustic elements and acoustic environments. Each audio sub-scene could be created statically or dynamically.

Requirements on Audio Renderer

5. The specification shall support metadata describing the audio scene.

6. The specification shall support metadata for controlling and restricting the audio scene.

   Note: This may already be fulfilled by MPEG-H 3D Audio.

7. The specification shall support control (e.g., via metadata or interface) of the audio rendering parameters (e.g., consumption space, player capabilities, etc.).

8. The specification shall support random-access in time (e.g. every 0.5 seconds) and space (e.g. jump within a sub-scene or to a new sub-scene).

9. The specification shall support metadata for enabling transition effects on audio rendering during user jumps between two different listener positions in the audio scene (e.g., fade-out fade-in).

10. The specification shall support metadata for enabling audio zooming (e.g., adjustment of prominence, dialog enhancement, simulation of depth-of-field effect, etc.).

    Note: This may already be fulfilled by MPEG-H 3D Audio.

11. The specification shall support 3D spatial extent for audio objects, supported by metadata.

12. The specification shall support rendering of the radiation pattern of audio objects and channels, supported by metadata.

13. The specification shall support occlusion of audio elements, supported by metadata.
14. The specification shall support locally captured audio (e.g., user's own voice, side tone) in the audio scene.
15. The specification shall support accurate 3D spatial localization of audio objects (sound sources). Differences in location are with respect to what is perceivable.

Interfaces and extensions

16. The specification shall support input interfaces for changing the audio scene.

    Note: MPEG-H 3D Audio already provides interfaces enabling basic functionality that could be enhanced for MPEG-I Audio.

17. The specification shall enable extension of the rendering functionality (e.g., interfaces to external rendering tools, extension payloads, reserved bit fields etc.).
18. The specification shall support personal HRTFs in the audio renderer, including an interface for providing these filters.
19. The specification shall support personal headphone equalization filters including an interface for providing these filters.

Presentation Modes

20. The specification shall support 6DoF head-tracked binaural rendering for headphone reproduction.
21. The specification shall support 6DoF head-tracked rendering to loudspeakers (e.g. to immersive configurations such as 7.1 + 4H).

    Note: The user explores the scene by moving in the listening area and based on his position the audio is rendered.

22. The specification shall support 6DoF rendering to loudspeakers for the use case that the user's consumption position is fixed, while the virtual position changes.

    Note: The user is located in the sweet-spot and navigates the scene for example using a joystick.

23. The specification shall support rendering to a combination of 6DoF head-tracked binaural headphones reproduction and loudspeaker reproduction.

Social VR

24. The specification shall support rendering of speech and audio from other users in the virtual environment. The speech and audio may be immersive.
    - 24.1. The specification shall support low-latency conversation between users within a given virtual environment.
    - 24.2. The specification shall support low-latency conversation between a user within the given virtual environment and a user outside the given virtual environment.
    - 24.3. [SYS] The specification shall enable synchronization of audio and video of users and the scene.
    - 24.4. The specification shall support metadata specifying restrictions and recommendations for rendering of speech/audio from the other users (e.g. on placement and sound level).

Interoperability between 3DoF and 6DoF platforms

25. The specification shall support decoding and presentation of MPEG-H 3D Audio Low Complexity Profile content on an MPEG-I 6DoF platform with an experience as with an MPEG-H 3D Audio Low Complexity Profile decoder.
26. The specification should support consumption of MPEG-I Audio content on MPEG-H 3D Audio Low Complexity Profile decoder (3DoF).

    Note: By using the MHAS format this requirement can be fulfilled.

27. The specification shall enable consumption of MPEG-I Audio content on MPEG-I Audio platforms with reduced degrees of freedom e.g., 3DoF+, 3DoF, 0DoF platforms.

Draft MPEG-I Audio Evaluation Platform

The Moving Picture Experts Group (MPEG) is a working group developing standards to create coded representations of digital audio, video and 3D graphics. The MPEG-I standard specifically addresses design and coding requirements in order to render audio content for six degrees-of-freedom (/6-DoF) virtual and/or augmented reality content. Unlike traditional cinematic content, 6-DoF VR/AR allows all pitch, yaw, roll and translation along the X, Y and Z axis to be represented inside the VR environment. Typical use cases for this technology could be a tour of a virtual museum, or presentation of a holiday destination before purchase. In addition, the technology includes the potential for users to interact and manipulate elements of their surroundings.

Evaluation Platform

In order to assess the performance of submitted proponent technologies, an audio evaluation platform (AEP) has been developed. Using the AEP, multiple proponent technologies can be evaluated in parallel, in real-time, whilst the user is free to fully explore the virtual reality environment. The aim of the AEP is to allow assessors to capture quantitative subjective listening test data such that all aspects and experiences from a user can be considered when assessing quality for the next generation of audio coding for virtual and augmented reality. Fig. 4 shows a top-level overview how pre-processed offline data (left) coordinates with the real-time evaluation (right).

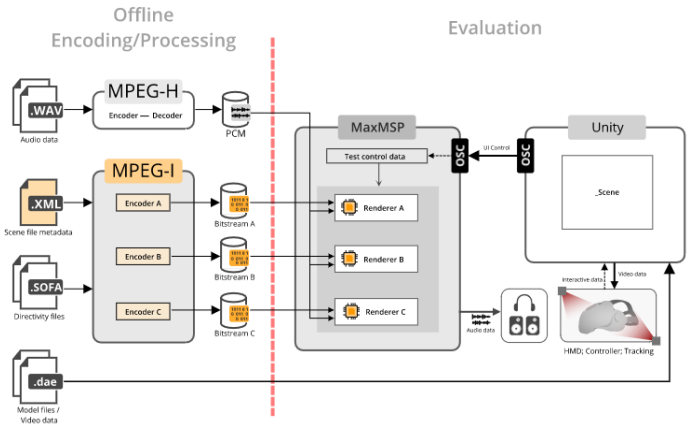


Figure 4 - Overview of Evaluation Platform

The AEP has been developed for intended use with the HTC Vive or HTC Vive PRO. In order to meet the requirements of VR applications, and to have a real-time system that can conduct traditional audio evaluation methodologies with multiple real-time renderers (VST3 plug-ins) running in parallel, the AEP is split into two major components.

- Unity3d: This is used as the graphical rendering engine, housing all VR components, mechanics, visual elements and user interface controls.
- MaxMSP 8: This is used as the audio rendering engine, housing all CfP renderers, all testing configuration settings, audio material, data handling and monitoring.

AR Extension to the Evaluation Platform

To facilitate AR testing using an AR HMD, an extension to the Evaluation Platform was made to allow the use of the Microsoft HoloLens instead of the HTC HMD.

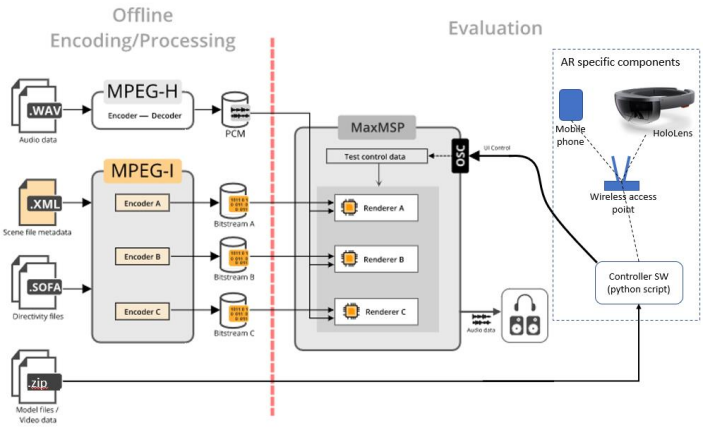


Figure 5 - Overview of Audio Evaluation Platform with the AR extension

The AR extension uses some of the existing components of the Evaluation Platform without modifications (MaxMSP and VSTs) and replaces some of the components (Unity, HTC Vive) with new ones (Fig. 5, above). The Unity and tracker components on the right side of Fig. 4 are replaced with the following components:

- Microsoft HoloLens with an app used for showing AR scene visual and performing tracking duties.
- Controller software running on the same computer as MaxMSP and a wireless access point to handle communication between the HoloLens app and MaxMSP.
- A mobile phone for viewing a web UI served by the controller script for user input (MUSHRA scores, moving between scenes).

In addition to the above, an external sound card is required for audio playback.

Output Documents
N18626 - Workplan on MPEG-I Audio
N18627 - Draft Documentation for the MPEG-I Audio Evaluation Platform
N18628 - Draft MPEG-I 6DoF Audio Encoder Input Format
N18629 - Draft MPEG-I Audio Test and Evaluation Procedures

4 Internet of Media Things

Mapping ISO/IEC SC29 IoMT to ISO/IEC SC41 IoT Architecture

Inside ISO, parallel efforts are carried out for standards related to Internet of Media Thing. For instance, SC 41/WG 3 is establishing standards in the areas of reference architectures, frameworks, and vocabulary/definitions for the Internet of Things.

The mapping of components between IoMT and SC41 IoT Reference Architecture is straightforward as the interfaces encompassed by the SC29 IoMT Architecture (Fig. 6) ensure data exchanges between the various domains considered by SC41 IoT Reference Architecture (Fig. 7). Specifically, the following mapping is established:

- Interface 1 & 1' of IoMT relate to the User Domain interaction with the things in SC41 IoT Reference Architectures
- Interface 2 & 2' of IoMT relate to the functionalities of the Sensing & Controlling Domain and to Application & Service Domain
- Interfaces 3 of IoMT relates to Resource Access & Interchange Domain.

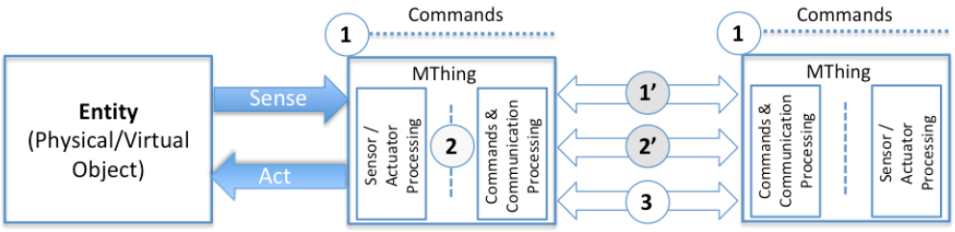


Figure 6 - IoMT Architecture

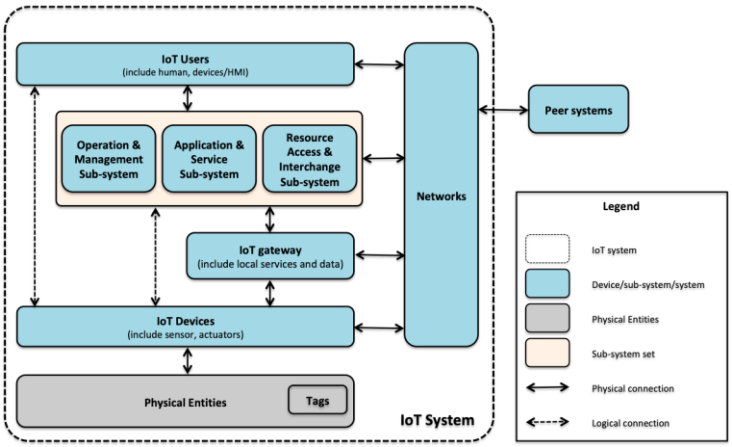


Figure 7 - Entity-based RM in ISO/IEC 30141:2018

References

- ISO/IEC 30141:2018 Internet of Things (IoT) - Reference Architecture