A picture containing drawing, food

Description automatically generated **ISO/IEC JTC 1/SC 29/ WG 11 N 19511**

**ISO/IEC JTC 1/SC 29/WG 11**

**Coding of moving pictures and audio**

**Convenorship: Japan (JISC)**

|  |  |
| --- | --- |
| **Document type:** | Approved WG 11 document |
| **Title:** | Requirements for MPEG-I Phase 2 |
| **Status:** | Approved |
| **Date of document:** | 2020-07-03 |
| **Source:** | Convenor, ISO/IEC JTC 1/SC 29/WG 11 |
| **No. of Pages:** | 17 |
| **Email of acting convenor** | ostermann@tnt.uni-hannover.de |
| **Committee URL:** | <http://isotc.iso.org/livelink/livelink/open/jtc1sc29> |
|  |  |

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N19511**

**MPEG 131, June 2020, Online**

|  |  |
| --- | --- |
| **Source** | **Requirements** |
| **Status** | **Approved** |
| **Title** | **Requirements for MPEG-I Phase 2** |
| **Editors** | **Mary-Luc Champel (Xiaomi), Igor D.D. Curcio (Nokia)** |

# Introduction

This document contains requirements for MPEG-I Phase 2 and replaces its earlier version in N18965. It includes requirements for different areas collected from [1-4]. This document replaces these earlier documents and will be the only source of requirements to be updated over time. There may be separate documents for requirements that are related to MPEG-I (e.g., requirements for Video Decoding Interface for Immersive Media, requirements for Media Access and Delivery). Such documents will evolve independently.

“The specification” shall mean any specification which may be targeted for MPEG-I Phase 2, as well as any supporting standards that may require updating, such as those specifying certain types of metadata.

For each requirements sub-section, and wherever possible, this document identifies the specifications that should fulfil them. These are listed as “target specifications”.

The MPEG-I Phase 2a requirements described in section 3 are final, and no further contributions are possible. The current estimate of completion of MPEG-I Phase 2a work is July 2021 [5]. The estimate of completion of MPEG-I Phase 2b is tentatively Q3/2022. For Phase 2b requirements new input contributions are still encouraged.

For some requirements Call for Proposals may be issued.

# Definitions

This Section contains definitions for terms that are not defined in specifications related to MPEG-I Phase 2.

|  |  |
| --- | --- |
| **Media Types** | Types of media including at least: 2D video, spherical video, meshes, point clouds and other visual objects (such as graphics), various audio media data and haptic data. |
| **6DoF scene** | A scene in which a user is able to move around freely with 6 degrees of freedom. |
| **Viewpoint** | The point from which the user observes (views and hears) the scene; it usually corresponds to a camera position. Slight head motion does not imply a different Viewpoint |
| **Audio Object** | Audio Signal and associated metadata as used in ISO/IEC 23008-3. |
| **Audio Channel** | Audio Signal and associated metadata as used in ISO/IEC 23008-3. |
| **HOA** | Audio Signal and associated metadata as used in ISO/IEC 23008-3. |
| **Earcon** | Audio Signal and associated metadata as used in ISO/IEC 23008-3:2019/AMD 1. |
| **Audio Element** | One or more audio signals and associated metadata. Audio Elements are audio objects, channels or HOA signals with associated MPEG-I 6DoF metadata and MPEG-H 3D Audio metadata if appropriate. |
| **Acoustic Environment** | Metadata describing the acoustic properties of the virtual scene to be rendered, e.g. room or locality. These might include reverberation times (RT60), zones of occlusion, etc. |
| **Acoustic Element** | Object in immersive space that reflects, diffracts and/or absorbs sound. |
| **Audio Scene** | All audio elements, acoustic elements and acoustic environment which are needed to render the sound in the scene.  Note: Audio scene is not the same as scene-based audio (e.g., HOA). |
| **User** | The listener whose position and orientation are used for rendering. |
| **Audio Scene change** | Changes in the audio scene which impact the audio rendering. These changes might be triggered by the user. |
| **Audio Renderer** | Normative functionality that renders sound for the user. The output of the Audio Renderer are the audio signals to be reproduced over headphones or loudspeakers. |
| **Audio Rendering Parameters** | Parameters that control the audio renderer, such as enabling/disabling specific features (e.g., room acoustics, Doppler shifts) or min/max values, gain values, default values, etc. |
| **MPEG-H 3DA Decoder** | MPEG-H 3D Audio Low Complexity (LC) Profile decoder that receives as input an MPEG-H 3D Audio LC Profile MHAS stream and provides as output decoded PCM audio together with all metadata available in the MHAS packets. Decoded PCM audio contains channels, objects and reconstructed HOA as described in ISO/IEC 23008-3:2018 Section 17.10.  *Note: Not all metadata provided has to be used for MPEG-I 6DoF rendering.* |
| **Haptics** | The science of touch. Haptics has two sub-modalities – vibrotactile (touch, texture, temperature) and kinesthetic (force feedback, motion). |
| Processing model / Scene Processing Model | A normative description of the association of information in a media bitstream with a scene description. |

# Requirements for MPEG-I Phase 2a

## MPEG-I Container for 6DoF Content Media

*Potential target specifications: OMAF, File Format.*

1. The specification shall support the storage of the defined Media Types.
2. The specification shall support the presentation and playback of different Media Types.
   1. The specification shall support the storage and presentation of multiple different Media Types such that they are synchronized both spatially and temporally.

*e.g. a 2D background video surrounding a point cloud object.*

1. The specification shall support the presentation of content media depending on the user-selected location and view within a 6DoF scene.

## Scene Description for 6DoF Content Media

*Potential target specifications: Scene Description for MPEG Media.*

### General

1. If possible, the solution shall define interfaces to integrate existing scene description formats rather than define a new scene description format
2. The solution shall reuse existing interfaces/API definitions (also from other SDOs) whenever possible and appropriate
3. It shall be possible to indicate object information in the scene description format such that one can derive a consistent relationship between object map information in a video bitstream and object information signaled in the scene description format.
4. The information signaled in the scene description format shall be consistent with presentation layer information signaled in the video/audio bitstream at the codec layer as well as information signaled in the media formats at the systems level via ISOBMFF, DASH, etc… In case of contradicting information, the scene description format shall take precedence.
5. It shall be possible to associate Information in the media bitstream targeted for the scene description with a processing model.
6. The scene description shall enable the option to describe the scene using geographical coordinate systems.
7. The scene description shall enable modular rendering, i.e. smaller portions of the scene can be independently accessed and rendered.

### Reference Scene Description Selection

*MPEG has decided not to develop a format for Scene Description itself, but rather to rely on an available industry standard. The following requirements were drafted to help selecting the system that MPEG seeks to reference.*

1. The scene description shall support audio and video formats as well as other media formats standardised by MPEG (including OMAF).
2. The scene description shall enable the support of other visual or audio media formats.
3. The scene description shall support definitions to indicate how sub-graphs and objects are related in terms of their temporal, spatial and logical relationships
4. The scene description shall support composition of digital representations of natural and synthetic objects.
5. The scene description shall support synchronisation between objects and attributes in the scene.
6. The scene description shall support spatial and temporal random access.
7. The scene description should support information to enable a renderer to perform path tracing.
8. The scene description shall support sub-graph representation that allows modular rendering e.g. leafs in the scene description tree can also be packaged and referenced individually from a parent scene description and container.
9. The scene description shall support references (e.g. URLs) to external media resources in place of embedded file references
10. The scene description shall support a mechanism to safely customize behavior for nodes like camera, texture, geometry, audio, and object placement nodes through sandboxed, validated domain specific shaders or scripts for these nodes without affecting the functionality or forcing changes to the root node graph or other node types; i.e. provide a mechanism to safely extend the scene description.

### Presentation engine interface

1. It shall be possible to update the whole scene-graph, a sub-graph, or a node in the scene description
2. It shall be possible to correctly render a 6DoF Presentation after a random access in time
3. It shall be possible to perform timed scene description updates
4. It shall be possible to associate a scene description update with the corresponding scene description
5. It shall be possible to use a scene description as the entry point to a 6DoF presentation.

### Media access Interface

1. It shall be possible to access timed and non-timed, 2D and 3D media (meshes, point clouds, audio elements, …), stored locally or over the network
2. It shall be possible to pre-fetch media that the presentation engine expects to be used in the presentation
3. It shall be possible to retrieve media depending on the desired level of detail
4. It shall be possible to retrieve and access referenced media partially in time and space
5. It shall be possible to describe position, orientation, and visual/acoustic characteristics when rendering referenced media
6. It shall be possible to synchronize media objects/resources and media components of a single object
7. Audio elements shall be rendered consistently with their corresponding visual elements, if such visual elements exist.
8. The specification shall enable synchronization of audio and video of users and the scene.

## Delivery of 6DoF Content Media

*Potential target specifications: OMAF, DASH, MMT, Carriage of Visual Volumetric Video-based Coding Data, Carriage of Geometry-based Point Cloud Compression Data, or any additional specification that MPEG may create for 6DoF media.*

1. The specification shall support low-delay delivery of different Media Types for 6DoF contents~~.~~
2. The specification shall support the partial delivery and presentation of content media depending on the user selected location and view within the scene.
3. The specification shall support full or partial delivery of media content depending on different network conditions, device capabilities and configurations.
4. The specification shall support differentiated service access by clients, through adaptation or partial delivery of the media content.

## Delivery Requirements for Content Interactivity

*Potential target specifications: OMAF, Scene Description for MPEG Media.*

1. The specification shall support low delay processing and presentation of object features for a scene, in order to minimize motion-to-photon latency.
2. The specification shall support defining conditional switching between viewports
3. The specification shall support hotspots that trigger actions like switching viewports.
4. The specification shall support signalling how content needs to loop back or continue playing, where this behaviour may be triggered by certain interactive conditions.

## Processing and Adaptation of 6DoF Content Media

*Potential target specifications: NBMP, OMAF, Carriage of Visual Volumetric Video-based Coding Data, Carriage of Geometry-based Point Cloud Compression Data.*

1. The specification shall support the network-based media processing framework.
2. The specification shall support the processing (for media distribution) of different Media Types.
3. The specification shall support the adaptation of multiple different Media Types to different types of devices in various network conditions.
   1. The specification shall support level of density adaptation of point cloud media depending on the user selected view location and end-to-end system capabilities.
4. The specification should support interoperable interfaces enabling the processing of different media data to create the scene.
5. The Specification shall provide efficient support for trick play:
   1. fast-forward and slow motion and rewind with support for changing the Viewport during the trick play.

## 

## Network (pre-)rendering of 6DoF Content Media

*Potential target specifications: NBMP.*

1. The specification shall support (pre-)rendering of immersive contents by a network, instead of the client which will consume the contents, where the output will be either 6DoF, 3DoF+, 3DoF, or 2D, and could be compatible with OMAF players.
   1. The specification shall support 6DoF content media metadata for (pre-)rendering by a network.
   2. The specification shall support metadata (such as different client network conditions, device capabilities and configurations) for (pre-)rendering by a network.
   3. The specification shall support remote encoding for converting immersive 6DoF content to a simpler representation, such as 2D image/video.

# Requirements for MPEG-I Phase 2b

## Audio Requirements for MPEG-I Phase 2b

*Potential target specifications: Immersive Audio Coding.*

Note: background information for the Audio requirements can be found in [1].

### General Requirements for Audio with 6 Degrees of Freedom

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.
   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.

* 1. The specification shall support Loudness and Dynamic Range Control using MPEG-H 3D Audio tools.
  2. The specification shall support additional metadata as needed to support user 6DoF.
  3. The specification shall support delivery of the audio scene in multiple audio streams.

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

1. 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.
   1. Audio elements shall be rendered consistently with their corresponding visual elements, if such visual elements exist.
   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.

* 1. The specification shall support earcons.

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

1. 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.
   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

1. The specification shall support metadata describing the audio scene.
2. The specification shall support metadata for controlling and restricting the audio scene.

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

1. The specification shall support control (e.g., via metadata or interface) of the audio rendering parameters (e.g., consumption space, player capabilities, etc.).
2. 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).
3. 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).
4. 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.*

1. The specification shall support 3D spatial extent for audio objects, supported by metadata.
2. The specification shall support rendering of the radiation pattern of audio objects and channels, supported by metadata.
3. The specification shall support occlusion of audio elements, supported by metadata.
4. The specification shall support locally captured audio (e.g., user's own voice, side tone) in the audio scene.
5. 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

1. 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.

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

### Presentation Modes

1. The specification shall support 6DoF head-tracked binaural rendering for headphone reproduction.
2. 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.

1. 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.

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

### Social VR

1. The specification shall support rendering of speech and audio from other users in the virtual environment. The speech and audio may be immersive.
   1. The specification shall support low-latency conversation between users within a given virtual environment.
   2. The specification shall support low-latency conversation between a user within the given virtual environment and a user outside the given virtual environment.
   3. The specification shall enable synchronization of audio and video of users and the scene.
   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 audio platforms

1. 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.
2. 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.

1. 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.

## Spatial and Temporal Alignment of 6DoF Content Media

*Potential target specifications: OMAF.*

*Note: This section is meant to target metadata which could be included into future OMAF amendments (not using scene graphs)*

1. The specification shall support metadata for enabling the client to synthesize a Viewpoint that was not captured by a camera, using omnidirectional videos from multiple camera-captured Viewpoints as inputs.
2. The specification shall support metadata which allows natural and smooth switching between real camera captured views and virtual (interpolated) camera views within the 6DoF scene.

## 

## Scene Description for 6DoF Content Media

*Potential target specifications: Scene Description for Immersive Media.*

### Reference Scene Description Selection

1. The scene description should support nodes and attributes in order to implement natural laws of light, energy propagation and physical kinematic operations.
2. The scene description should support nodes and attributes in order to implement natural laws of acoustic energy propagation and physical kinematic operations.
3. The scene description should support description of ray-traced camera parameters for rendering
4. The scene description shall support parametric models for use in rendering environmental acoustic behaviour (e.g. reverberation, occlusion and directivity).

### Local capture Interface

1. It shall be possible to discover and configure local capture modalities
2. It shall be possible to adjust the presentation based on local capture modality availability
3. It shall be possible to reference media objects that are captured locally using different capture modalities
4. It shall be possible to provide feedback through available actuators

### User inputs Interface

1. It shall be possible to discover user interactivity modules
2. it shall be possible to define custom interactivity procedures based on input from the user or from the user’s devices and sensors

### Export

1. The scene description shall support information to enable a renderer to output raster data (image, and video), volumetric data (point clouds, meshes, arrays of voxels, and reflectance fields) and audio.
2. The scene description shall support a scriptable export output node for asynchronously exporting (as a file stream or buffer) any or all parts of the scene description connected to a node into a simpler or flattened representation.

## 

## Descriptions for Content Interactivity

*Potential target specifications: OMAF, Scene Description for MPEG Media.*

1. The specification shall enable realistic composition of a 6DoF scene depending on the user-selected location and orientation.

Note: Such composition may, e.g., include delivering proper lighting information and some form of geometry information of the scene so the view is rendered with realistic lighting and shadows.

1. The specification shall support a user interacting with objects within a given virtual environment.
   1. The specification shall support giving an object a place in the virtual environment, moving it and changing some attributes

Example: a rectangular screen that represents a TV set on which a different channel could be chosen; a point cloud object that is inserted into a representation of the real world

1. The specification shall support metadata describing objects and their features within a 6DoF scene.
   1. This metadata will be defined in updated versions.
2. The specification shall support replacement of 6DoF content media data.
   1. The specification shall support metadata for the processing of object features, namely preserving and replacing features for objects in a scene, either by the client, or in the network.

*Note: A typical scene may have multiple objects, with indices which describe their features.*

## Descriptions for Multi-User Interactivity

*Potential target specifications: OMAF, NBMP, possibly Scene Description for MPEG Media.*

1. The specification shall enable realistic composition of user-embodiment within 6DoF content media.

Note: Such composition may, e.g., include delivering proper lighting information and some form of geometry information of the scene so user-embodiment is rendered with realistic lighting and shadows.

1. The specification shall support rendering of other users in 6DoF content media, including possible speech or audio from other users.
2. The specification shall enable multi-user immersive applications in which several users are experiencing the same immersive experience together.
   1. It shall be possible to detect & render interactions between users within the immersive environment.
3. The specification shall support conveying metadata about the spatial alignment of a camera and the person that this camera is capturing, in the physical environment. This metadata shall include at least:
   1. the distance between the camera and the captured person
   2. the orientation of the camera
   3. the focal length and possibly other lens parameters
   4. the location of the captured person’s head in the video

## 

## Delivery Requirements for Multi-User Interactivity

*Potential target specifications: OMAF, NBMP, possibly Scene Description for MPEG Media.*

1. The specification shall enable multi-user immersive applications in which several users are experiencing the same immersive experience together.
   1. The specification shall enable synchronous play-out between multiple users consuming the same scene, where the synchronization is accurate within 100 ms.

Example: multiple users viewing the same sports event and communicating about the event in real-time

Note: the end-to-end latency of the experience is much less important than the inter-user sync.

1. The specification shall support synchronization of audio and video of users and the scene.
   * 1. Synchronization among users of all interactions (with VR objects and between users) shall be supported.
2. The specification shall support the low-latency delivery of interactions between users within a given virtual environment.
3. The specification shall support interactions where it will be rendered to a certain group of users.
   1. The specification shall support privacy protection features related to the delivery of client specific metadata (e.g. position, orientation, etc.), especially for interactions related to bi-directional delivery applications (e.g. social VR).

*NOTE: This requirement needs to be further clarified.*

## Use Case Specific Requirements

Note: Requirements from this section may be updated to more generic requirements which match into the technology clustered sections above.

### Social VR

*Potential target specifications: OMAF, and possibly Scene Description for MPEG Media and/or a new specification that MPEG may want to develop to support the requirements in this section.*

1. The specification shall support interactions in case of capability mismatching of user devices.
2. The specification shall support that a user can recognize objects to interact with.
3. The specification shall support metadata that provides the position/orientations of remote users in the immersive environment of a local user
4. The specification shall support metadata that provides the direction of view of the local user in its immersive environment
5. The specification shall support metadata that indicates which remote user is being looked at by the local user
6. The specification shall support metadata that indicates whether a video stream corresponds to user A looking straight into the camera (main camera) or not (side camera)
7. The specification shall support containers (e.g. ISOBMFF) to carry the above metadata to remote users.
8. The specification shall support metadata that provides x/y/z and yaw/pitch/roll coordinates that provide the position/orientation of a camera or audio source with respect to the perceived immersive environment.
9. The specification shall support metadata that provides the type and other details of the camera or audio source.
10. The specification shall support metadata to identify and distinguish cameras and audio sources.
11. Metadata to signal details of the visual and/or auditive/haptic indication of the camera/audio source position/orientation in the immersive environment
12. Containers (e.g. ISOBMFF) to carry the above metadata to the other user(s).
13. Metadata that provides the angle of view of the camera
14. Metadata that provides the distance between the camera and the video-captured person
15. Metadata about the position of the video-captured person in the video frame
16. Containers (e.g. ISOBMFF) to carry the above metadata to remote users.
17. The specification shall support positioning multiple objects in the same omnidirectional environment where these objects are represented by an image or a video

Note: the primary use case of this requirement is to place multiple users in, e.g., a video conferencing setting in an omnidirectional environment, together with other elements, e.g., a video screen.

* 1. The specification shall enable making the omnidirectional environment consistent:
     1. for all users that are embedded in their omnidirectional environment
     2. *between* the users (they can look at one another, and can see when another person is looking at them)
     3. for multiple users looking at / pointing at a common element in the omnidirectional environment (e.g., a video screen)
  2. The specification shall support synchronization of user viewpoints and orientations (i.e. where each user is looking at), as well as content playback status, between users.

Note: these individual environments need not necessarily be the same, as long as they are internally consistent for all participants individually

Note: “consistent” means the right visual perspective for all objects as well as audio/visual alignment

* 1. The specification shall enable bringing multiple users together in the same omnidirectional environment even when they are captured using their own individual 3DoF coordinate system
  2. The specification shall enable positioning a rectangular object (e.g., a video screen) inside the omnidirectional environment with correct perspective and parallax. The surface of this object may be at an arbitrary distance from the user, and may have an arbitrary orientation in 3D space.

## Draft Haptics Requirements

*Potential target specifications: this likely requires a new specification. It may also need support in OMAF and File Format.*

Requirements for Haptics are currently under study.

MPEG currently envisages standardizing the coded representation of haptics but not necessarily the rendering part. MPEG is currently investigating which process will be followed.

1. The specification shall support rendering of haptic media based on available hardware at the client.
2. The specification shall support mixing and modulation of intensity of haptic tracks.
3. The specification shall support associating haptics with objects or object features in the scene.
4. The specification shall support transcoding as part of pre-rendering of haptic media
5. The specification shall support collision detection suitable for enabling haptic interaction
6. The specification shall support association of haptic texture with a 3D object for enabling haptic interaction.
7. The specification shall support server-side identification and encoding of material properties.
8. The specification shall support association of haptic media with media geometry types (e.g., by nested sphere, by solid angle, etc.)
9. The specification shall support orchestration of multimodal media presentation (e.g., audio, video, and haptics)
10. The specification shall support haptic effects to be associated with 2D video, and spatial subsections of the 2D video.
11. The specification shall support haptic effects to be associated with VR objects and VR environments.

# 

# Requirements for MPEG-I later Phases

This section will be filled with MPEG-I Phase 2 requirements when it is understood they cannot be met during Phase 2a or 2b. New requirements may also be added here.

# References

1. MPEG-I requirements on audio, w18158.
2. MPEG-I Phase 2 requirements, w18127.
3. MPEG-I requirements on scene description, w18339.
4. MPEG-I Phase 1b requirements, w17331.
5. MPEG Roadmap, w18975.