**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N** **16773**

**April 2017, Hobart, Australia**

|  |  |
| --- | --- |
| **Source:** | Requirements |
| **Title** | Requirements for Omnidirectional MediA Format |
| **Editors** | Mary-Luc Champel / Rob Koenen |
| **Public** | Yes |

# Introduction

This document contains the requirements for MPEG-I.

This first version focuses on MPEG-I phase 1a; requirements which should be fulfilled by MPEG-I part 2, the Omni-directional Media Format (also known as OMAF). These requirements take the advanced state of development of OMAF into account, as well as the timeline that has been established for MPEG-I Phase 1a.

# Definitions

|  |  |
| --- | --- |
| 2D Audio | Presentation of sound elements positioned in a plane |
| 3D Audio | presentation of sound elements from all directions including above and below |
| 3DoF | See TR |
| Immersive Audio | Concept and format for audio with the goal to capture, mix and reproduce 3D Audio, i.e. sound from all directions around a user. Immersive audio can be formatted as channels e.g. 7.1+4, scene-based audio (HOA), object-based audio, as well as combinations thereof. MPEG-H 3D Audio has been designed to these formats and their rendering for VR applications. |
| Diegetic audio | Sound whose source is visible on the screen or whose source is implied to be present by the action of the film  Note: Diegetic sound is any sound presented as originated from source within the film's world. Examples of diegetic sounds are: voices of characters, sounds made by objects in the story, music represented as coming from instruments in the story space. Diegetic sound can be either on screen or off screen depending on whether its source is within the frame or outside the frame. |
| Non-diegetic Audio | Sound whose source is neither visible on the screen nor has been implied to be present in the action of the film  Note: Non-diegetic sound is represented as coming from a source outside story space. Examples of non-diegetic sounds are: narrator's commentary, sound effects which are added for the dramatic effect, mood music. |
| Media Profile (from [1]) | Track for the file format, including elementary stream constraints |
| DASH Profile (from [1]) | Mapping of media to a DASH Media Presentation |
| Presentation Profile (from [1]) | Combination of different tools, including audio, video, encryption, subtitles |
| Viewport | The part of the Virtual Reality panorama that is visible to the user |

# Requirements for Part 2

In this Section 3, “Specification” shall mean the first version of MPEG-I Part 2 also known as OMAF that is planned for publication end of 2017.

## General Requirements for Part 2

1. The Specification shall provide for interoperable exchange of VR360 content.
2. The Specification shall avoid providing multiple tools for the same functionality to reduce implementation burden and improve interoperability.
3. The Specification shall enable good quality and performance.
4. The Specification shall enable full interoperability between services/content and clients.
   1. The Specification shall contain a very low number of fully specified interoperability points that include what is traditionally known as Profile and Level information.
      1. The existence of more than one interoperability point shall be justified if intended to target devices with different capabilities.
   2. Interoperability points shall address a Media Profile including:
      1. file format tracks and elementary stream
      2. rendering: The Specification shall provide interoperability points that include equirectangular projection. Other projection formats shall only be included if there are proven benefits and industry support;
   3. Interoperability points shall address a Presentation Profile for a full VR experience including different media (Video, Audio and Subtitles), enabling their temporal synchronization and spatial alignment
   4. These interoperability points shall enable conformance to be tested, inside and outside of MPEG.
   5. The Specification may contain partial interoperability points (e.g., a file format box, a visual media profile) at a lower level of granularity.
   6. The Specification may contain optional elements (like a description of the Director’s recommended viewport) when such options do not affect basic interoperability; Profiles can make such features mandatory but these features are not necessarily included in a Profile.
   7. The specification shall define at least one media profile for audio.
   8. The specification shall define at least one media profile for video.
   9. The specification shall define at least one presentation profile that includes one audio and one video media profile.
5. The Specification should take into account the capabilities of high quality devices such as HMDs that are on the market today (including Vive, Oculus, Gear VR, and Daydream) or that are on the market by the time the specification is stable, i.e., Q4 2017.
6. The Specification shall support the representation, storage, delivery and rendering of:
   1. Omnidirectional (up to 360° spherical) coded image/video (monoscopic and stereoscopic) with 3 DoF
   2. Both 3D and 2D audio
7. The specification shall work with existing MPEG storage and delivery formats
8. The Specification shall support temporal synchronization and spatial alignment between different media types, in particular between audio and video.
9. The Specification shall support metadata for describing initial viewpoints and for the playback of omnidirectional video/image and audio according to that metadata.
10. The Specification shall support the following interfaces:
    1. encoding and decoding for each media type
    2. delivery for download and streaming.
11. The Specification shall enable applications to use hardware-supported or pre-installed independently manufactured decoders and renderers through defined MPEG conformance points.
12. The Specification shall support viewport-dependent processing (this may include delivery, decoding and rendering).
    1. The Specification shall support dynamically changing viewports.
    2. The Specification should enable responsiveness to changing viewport in a way that doesn’t detract from the immersive experience.
13. The Specification shall support at least one Presentation Profile that requires support for neither viewport-dependent delivery nor viewport-dependent decoding.

Note: it is obvious that there will be viewport-dependent rendering, both for visual and audio components

## Delivery

1. The Specification shall support the following methods of distribution:
   1. File-based delivery
   2. DASH-based streaming
   3. MMT-based streaming

## Visual

1. The Specification shall enable content exchange with high visual perceptual quality.
   1. Taking the display resolution of existing headsets into consideration, the Specification shall support a visible viewport resolution beyond which the increase in resolution is no longer noticeable on these headsets.

Note: This may equate to a source resolution (for the full 360 video) of around 6k x 3k or 8k x 4k for equirectangular pictures (where the viewport is only the visible part of the panorama at a given point of time).

* 1. The Specification shall support a framerate of at least 60fps

1. The Specification shall support distribution of full panorama resolutions beyond 4K (e.g. 8K, 12K), to decoders capable of decoding only up to 4K@60fps, if sufficient interoperability can be achieved.
2. The Specification shall support metadata for the rendering of spherical video on a 2D screen.
3. The Specification shall support fisheye-based video with a configuration of 2 cameras.
4. The Specification shall support encoding of equirectangular projection (ERP) maps for monoscopic and stereoscopic video, in an efficient manner.
   1. Other projection maps than ERP for distribution should only be provided if consistent benefits over ERP is demonstrated.

## Audio

1. Each audio media profile in the Specification shall:
   1. support immersive rendering with sufficiently low latency

Note: this is related to requirement 12.2

* 1. support Excellent sound quality (as assessed per ITU-R BS.1534)
  2. support binauralization

Note: binauralization implies adaptivity to user head motion, such that the user experiences directional audio that is consistent with such head motion.

1. There may be one audio media profile that supports only 2D audio to cater to existing devices.
2. All other audio media profiles defined in the Specification shall:
   1. support 3D Audio distribution, decoding & rendering.
   2. support immersive content, e.g. 12ch or 3rd order Ambisonics,
   3. support a combination of diegetic and non-diegetic content sources.
   4. be capable to ingest and carry all content types:
      1. audio channels,
      2. audio objects,
      3. scene-based audio,
      4. and combinations of the above.
   5. be able to carry dynamic meta-data for combining, presenting and rendering all content types.

## Security

1. The Specification shall not preclude:
   1. Decoding and rendering to support secure media pipelines
   2. Efficient distribution for multiple DRM systems (e.g. using common encryption)
2. The Specification should enable a secure media pipeline to be implemented.

# Requirements for future versions of MPEG-I part 2

This Section will be added in the next revision of this document.

# References

1. MPEG, N16632, Study of ISO/IEC DIS 23000-19 Common Media Application Format, January 2017