**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** **17331**

**January 2018, Gwangju, KR**

|  |  |
| --- | --- |
| **Source:** | Requirements |
| **Title** | Requirements MPEG-I phase 1b |
| **Editor** | Rob Koenen / Mary-Luc Champel |
|  |  |

# Introduction

This document contains the requirements for MPEG-I phase 1b.

“The Specification” shall mean any future version(s) of the Omnidirectional Media Format (MPEG-I Part 2), planned for publication after the first version of this specification, and in MPEG-H Audio AMD5, as well as any supporting standards that may require updating, such as those specifying certain types of metadata and SEI messages.

Terms that use Initial Caps have the meaning given to them in the OMAF specification (W17235, ISO/IEC 23090-2) and in the Definitions Section below.

# Definitions

This Section contains definitions for terms that are not defined in the OMAF specification.

|  |  |
| --- | --- |
| **Audio Metadata** | Metadata that pertains to audio data |
| **Binocular Disparity** | difference in image location of an object seen by the left and right eyes or displayed on the left and right views of a display |
| **Earcon** | brief, distinctive sound used to represent a specific event or convey other information (Source: Wikipedia) |
| **Head-Motion Parallax** | displacement or difference in the apparent position of an object viewed from different viewing positions or viewing orientations. |
| **Motion-to-High-Quality Latency** | The time it takes between head motion and displaying content at high quality in a head-mounted device. |
| **Overlay** | Rendering of visual media over 360-degree video content. |
| **Transparency** | One or more overlays of visual media, where the overlay area is visually homogeneous, and the degree of visibility of the overlay(s) and the 360-video content may be arbitrary and not necessarily “binary”. |
| **Viewing Space** | 3D space of viewing positions within which rendering of image and video is enabled and VR experience is valid. |
| **Viewpoint** | the point from which the user views the scene; it usually corresponds to a camera position. Slight head motion does not imply a different Viewpoint |
| **Visual Media** | videos, images and text |

# Requirements for MPEG-I phase 1b

## Parallax

1. The Specification shall enable OMAF player implementations to support Head-Motion Parallax and Binocular Disparity in rendering of images and video, where the user perceives natural viewing conditions consistent with head motion.
   1. The Specification shall enable OMAF player implementations to support Head-Motion Parallax in rendering regardless of the viewing orientation.
      1. The Specification shall support constraining valid viewing orientations for head-motion parallax to Azimuth and Elevation ranges less than 360° and 180°, respectively.
   2. The Specification shall enable defining a Viewing Space that contains viewing positions constrained to head motion movements.

Note: For example, such Viewing Space may enable immersive experience for the following viewing situations:

* *A user sitting on a couch or a rotating chair in a natural way.*
* *A user standing in a natural way but without taking steps.*
  + 1. The specification shall support changing of the Viewing Space over time.
  1. The support for Head-Motion Parallax in rendering shall be enabled regardless of the distance between the viewer and the objects that are in the Viewport provided that no objects are within the Viewing Space.
  2. The Specification shall enable content authoring in a manner that Head-Motion Parallax and Binocular Disparity are supported in rendering without annoying artefacts while the user is in the Viewing Space.

Note: Testing methodology for visual quality of experience of rendering with Head-Motion Parallax and Binocular Disparity should be developed and/or chosen during the standardization.

* + 1. The Specification shall enable the support of Head-Motion Parallax and Binocular Disparity in rendering in a manner that any potential degradations in the quality of experience are not abrupt but graceful.
    2. The Specification shall enable an OMAF player implementation to detect viewing positions and/or viewing orientations where rendering is potentially unacceptable.

Note: Unacceptable rendering may occur when the viewing position is outside a Viewing Space and/or when the viewing orientation is outside the range of valid viewing orientations as discussed in 1.1.1.

Note: OMAF player implementations could indicate to the users when the viewing position/orientation is such that it is close to yield an unacceptable rendering quality. For example, the OMAF player may fade the image to grey if the head moves too far away from the sweet spot.

* 1. The Specification shall support creation of image or video content that i) enables rendering such content with head-motion parallax support, and ii) is structured in a manner that players not supporting head-motion parallax (e.g., OMAF v1 players) will decode and render monoscopic or stereoscopic omnidirectional projected images or video correctly.

Note: The Specification can allow such image or video content that enables rendering with Head-Motion Parallax and is not backward compatible with players not supporting head-motion parallax, but a non-backward compatible solution should have tangible benefits over a backward compatible solution.

## Composition and Spatial Alignment of Visual Media

1. The Specification shall support Transparency of Visual Media.

Note: transparency is meaningful only in conjunction with overlays, and not otherwise.

1. The Specification shall enable the Overlaying of visual media, such that their ordering and position are clear.

Examples:

* overlaying a logo (note: may not be rectangular and may use transparency);
* overlaying a sign language interpreter over the 360-degree video
* overlaying a small ERP of the entire 360-degree video as preview-window on top of the current Viewport to be used as guiding mechanism;
* overlaying a thumbnail of the recommended Viewport over the current Viewport.

Note: This requirement only pertains to simple overlays and is not intended to provide full-fledged composition

1. Any coordinate system used in the Specification shall be the same for audio and video media, and shall align with the coordinate system in OMAF version 1.
2. The specification shall define means to describe the relationships (spatial coordinates) between the different viewpoints.
3. The specification shall enable subtitles and Closed Caption rendering where such rendering takes Viewport information into account, for all phase 1b content, for example including video with depth.
4. The Specification shall support always rendering such Subtitles and Closed Captions in Viewport.

## Spatial and Temporal alignment of the virtual environment

1. The Specification shall support synchronisation between any combination of timed media data and timed metadata.
2. The Specification shall support minimizing the end-to-end delay for applications that need this.
   1. The specification shall enable an encoding to presentation delay that is similar to traditional linear TV.
   2. The specification shall support making a trade-off between efficiency and end-to-end delay.

Note: To a significant extent, this end-to-end delay is determined by factors that are beyond the present specification

1. The Specification shall support keeping motion-to-photon delay within the immersive limit

Note: motion-to-photon latency is probably fully determined by the rendering in the device.

1. The Specification shall not preclude temporally synchronizing (live) VR streams with other content, potentially distributed over an alternative distribution path, including broadcast streams.
2. The specification shall define means to describe the spatial relationships between the content corresponding to different Viewpoints.
3. The specification shall enable temporally synchronizing content corresponding to different Viewpoints.

Note: In addition to having the spatial relationship defined between the content captured in different Viewpoints, it is important to have temporal synchronization between the multi-location content. Switching between content from different Viewpoints would result in a bad user experience, in absence of temporal synchronization

1. The Specification shall support switching between content from different Viewpoints.
2. The specification shall enable signalling transition effects when switching to a content at a new location.

Example: a fade to black from the content at the old location and then fade in the new one; similar effects may apply to audio.

* 1. Transition effects shall be enabled to provide a comfortable and immersive experience.

1. The specification shall support alternative Viewport information that may comprise of single location or multiple location content. “Alternative” means that the Viewport would change independent of user input, e.g., by head motion.

Example: a director’s cut.

Note: This requirement would enable guidance to the player; it does not enforce any player behaviour. Also, note that this functionality may be easier on the user on 2D devices than in a Head-Mounted Display.

## Trick Play and Zooming

1. The Specification shall support zooming, where the processing is optimized for what is in the Viewport.

Note: when used in an HMD, zooming will be akin to using binoculars. When used on a “flat” device, the user can zoom by pinching in addition to the swiping that moves the Viewport

* 1. Zooming shall be supported during regular playback and trick play (at least during slow-motion and pause), while always seeing the Viewport in high quality
  2. The specification shall enable support of multiple “VR magnifiers” that allow for zooming an area less than the size of a Viewport.

Note: VR magnifiers may be for example algorithms, pre-magnified content, or parameterised, potentially with multiple types of VR magnifier controls, like user control, director control and a mix of these.

1. 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
   2. pausing while looking around and always seeing the Viewport in high quality
2. The Specification shall support zoom and trick play for content with and without Head Motion Parallax and in combination with Viewport-dependent processing

## Viewport-Dependent Processing

1. The specification shall support efficient Viewport-dependent processing (e.g., encoding, delivery, decoding, rendering) that can be adapted to different network conditions, as well as device capabilities and configurations.

Note: Device capabilities include decoder capability (codec, profile), display type (flat screen, HMD), display capability (field of view, refresh rate, colour volume, dynamic range and brightness, etc.)

* 1. The Specification shall support such Viewport-dependent processing with and without Head Motion Parallax.
  2. The Specification shall enable significant bitrate reduction over streaming of the full 360 sphere, where the goal is a reduction in the order of 75% in content without Head Motion Parallax. The Specification shall also provide a significant bitrate reduction over Viewport-dependent streaming using the OMAF v.1. These bitrate reductions shall be achieved at equivalent subjective quality levels assuming “normal” head motion.
     1. The Specification shall support minimizing any quality degradation that the user perceives because of motion-to-high-quality latency.
     2. The Specification shall support trading off motion-to-high-quality latency and bitrate efficiency.
  3. The Specification shall allow flexibility, in a storage-efficient manner, with regards to:
     1. implementing graceful degradation strategies, allowing the client to adapt the experience to the available bandwidth, where such strategies shall not be standardized

Note: this is similar to the implementation of bitrate-adaptation in DASH, which is not standardized by MPEG but rather left to implementers.

* + 1. preconfiguring content for low-complexity clients while allowing smarter clients to make their own decisions in retrieving certain areas at certain qualities
    2. supporting clients with different Viewport sizes in an efficient manner.
  1. The Specification shall enable keeping bitrate spikes due to head motion to a minimum
  2. The specification shall allow efficient distribution of VR360 content over Content Distribution Networks using http(s) without requiring changes to such CDNs.
     1. The Specification shall not require server-side per-Client processing
     2. The Specification should not preclude SAND-like optimizations
  3. The specification shall allow of Viewport-dependent streaming and consumption in media playback frameworks relying on Media Source Extensions and Encrypted Media Extensions.

## Audio

The audio requirements in this Section assume that the (normative) MPEG-H Renderer will be used and may be extended to allow rendering of the audio scene more accurately, or in ways that account for user interaction with the scene, e.g. when a user moves their head slightly sideways, or looks in a different direction.

Multi-location content support envisions delivering semantically related content, which has been captured via one or more cameras from different Viewpoints. For example: Viewpoints could refer to multiple cameras located within the same area (e.g., a basketball game) or different areas (e.g., multiple rooms).

Note that MPEG believes that the MPEG-H specification, including AMD5, already defines the Audio Metadata that is required to meet the requirements in this Section.

1. The specification shall support conveying Audio Metadata that describes the location of audio objects and the virtual environment.
   1. Audio Metadata describing the location of an audio object relative to the default position of the user.
   2. Positional Audio Metadata containing azimuth, elevation and distance describing all positions relative to the listener. The positional Audio Metadata shall provide sufficient spatial and temporal resolution that allows a renderer to form an output signal with high perceptual quality and without audible artefacts
   3. Audio Metadata about the acoustic environment of the sound objects shall be provided to the renderer
   4. Audio Metadata about the environment shall be provided with a spatial and temporal resolution that allows a renderer to form an output signal within a believable audible environment
2. The specification shall support conveying Audio Metadata that includes a description of the behaviour of a sound source dependent on the listener’s head orientation, head motion and distance to the sound source and must allow a renderer to form an output signal in a way that depends on the Viewport.

Example: when a user looks in a certain direction, the audio that is associated with that direction is amplified with respect to other audio sources in the scene, in a way that goes beyond just the normal directional rendering

1. The specification shall support conveying Audio Metadata describing the orientation and directivity of an audio object in such a way that a renderer can modify the level and/or the spectrum of a source dependent on the relative angle and the orientation of the source, if perceptually relevant.
2. The specification shall support conveying Audio Metadata describing the behaviour of a sound source when it is occluded, if perceptually relevant.
3. The specification shall enable signalling of audio objects that have a fixed position relative to the user orientation.
4. The specification shall enable signalling that permits zooming on audio rendering.
5. The specification shall support conveying Audio Metadata describing the size of an audio object.

## Earcons

1. The Specification shall support Earcons which can passively guide users of an immersive presentation via audio cues.
2. The Specification shall support easily adding and editing Earcons on top of existing audio content in post-processing.

## Optimized Viewing Experiences and Performance

1. The specification shall include video one or more Profile/Level combinations as well as other capabilities that enable rendering the Viewport at a resolution that meets or exceeds the capabilities of devices that are expected to ship in 2019. It is currently believed that this will include at least one Level with a resolution of 8k x 4k, possibly at 120 Hz frame rate.
2. The Specification shall support exploiting advanced device display capabilities for a natural and comfortable presentation including dynamic range, pixel density, colour space and frame rate capabilities.
3. The Specification shall support users adapting experiences to their individual viewing comfort.
4. The Specification shall support protecting users in terms of dynamically changing light conditions while still rendering specular highlights.

# Requirements for Metrics in Immersive Services

This section contains requirements on Metrics to be used with Immersive Services and Applications.

In this Section 3.1, “Specification” shall mean the part of the MPEG-I standard that will specify these metrics.

Please note that reporting metrics is not an obligation that any ISO standard will impose on any implementation of the Specification, and that some metrics may contain privacy-sensitive information (e.g. gaze direction in an omnidirectional video). Reporting such metrics may be subject to user consent.

## Viewport Metadata

1. The Specification shall support reporting the users’ Viewport, in head-mounted displays as well as “flat” devices.

## Other Metrics

1. The Specification shall support reporting different latency components of VR consumption systems.
2. The Specification shall support logging measurable events that can be logged by common VR360 consumption systems.
3. The Specification shall reuse already existing DASH Metrics whenever possible.