**I****NTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N16952**

**July 2017, Torino, Italy**

|  |  |
| --- | --- |
| **Source** | **Systems** |
| **Title** | **Profiles under Considerations for ISO/IEC 23000-20 Omnidirectional Media Format** |
| **Author** | Yago Sanchez, Adrian Murtaza |

Contents Page

1 Introduction 3

2 Normative references 3

3 Media profiles 3

10 Media profiles 4

10.1 Video profiles 4

10.2 Audio profiles 9

4 Presentation Profiles under Consideration 20

4.2 Introduction 20

4.3 OMAF Baseline Viewport-Independent Presentation Profile 20

4.3.1 Introduction 20

4.3.2 Definition 20

4.3.3 ISO Base Media File format 20

Annex A. Additional media profiles under consideration 21

A.1. AVC viewport dependent profile 21

A.2. HEVC viewport independent fisheye video profile 22

A.3. Timed text profile 24

Annex B. Submission and requirements fulfilment information 26

B.1. HEVC viewport independent baseline profile 26

B.2. HEVC viewport dependent baseline profile 32

B.3. OMAF 3D Audio Baseline Profile 37

B.4. OMAF 2D Audio Legacy Profile 42

B.5. AVC viewport dependent media profile 48

B.6. Viewport-Independent Fisheye Video Profile 49

# Introduction

This document contains two video media profiles, and two audio media profile included in Section 3, which are included as well into the Study of DIS of OMAF [SoDIS-OMAF] as agreed profiles. Besides, a presentation profile is included in section 4 and 2 video media profiles and a timed text profile are specified in Annex A of this document, which correspond to proposed profiles that are under consideration, as agreed during the MPEG 119th meeting.

# Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

These normative references are intended to include corrigenda and amendments available at the time of use.

|  |  |
| --- | --- |
| [14496-24] | ISO/IEC TR 14496-24, Information technology — Coding of audio-visual objects — Part 24: Audio and systems interaction |
| [3DA] | ISO/IEC 23008-3, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio |
| [AAC] | ISO/IEC 14496-3, Information technology — Coding of audio-visual objects — Part 3: Advanced audio coding |
| [AVC] | ISO/IEC 14496-10, Information technology — Coding of audio-visual objects — Part 10: Advanced video coding |
| [CICPa] | ISO/IEC 23091-3, Information technology — Coding-independent code points — Part 3: Audio |
| [CICPv] | ISO/IEC 23091-2, Information technology — Coding-independent code points — Part 2: Video |
| [CMAF] | ISO/IEC 23000-19, Information technology — Multimedia application format (MPEG-A) — Part 19: Common media application format (CMAF) for segmented media |
| [DASH] | ISO/IEC 23009-1, Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats |
| [DRC] | ISO/IEC 23003-4, Information technology — MPEG audio technologies — Part 4: Dynamic range control |
| [HEVC] | ISO/IEC 23008-2, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding |
| [ISOM] | ISO/IEC 14496-12 Information technology — Coding of audio-visual objects — Part 12: ISO base media file format |
| [MMT] | ISO/IEC 23008-1, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 1: MPEG media transport (MMT) |
| [MP4FF] | ISO/IEC 14496-14, Information technology — Coding of audio-visual objects — Part 14, MP4 file format |
| [MP4SYS] | ISO/IEC 14496-1, Information technology — Coding of audio-visual objects — Part 1: Systems |
| [SoDIS-OMAF] | N16950, Study of ISO/IEC DIS 23000-20 Omnidirectional Media Format |

# Media Profiles included in the Study of DIS of OMAF

The media profiles and presentation profiles within this section have been agreed to be included into the Study of DIS of OMAF [SoDIS-OMAF]. Section 10.1 and 10.2 of the study of DIS are copied below:

# Media profiles

## Video profiles

[Ed. (YK): Align text formats, including paragraph spacing, with other clauses.]

### Overview

This clause defines media profiles for video. Table 10.1 provides an informative overview of the supported features. The detailed, normative specification for each video profile is subsequently provided in the referred clause.

**Table 10.10‑1 – Overview of OMAF media profiles for video (informative)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Media Profile** | **Codec** | **Profile** | **Level** | **Required scheme types** | **Viewport dependent delivery & decoding** | Brand | Clause |
| HEVC viewport independent baseline | HEVC | Main 10 | 5.1 | podv and erpv | no | ovib | 10.1.2 |
| HEVC viewport dependent baseline | HEVC | Main 10 | 5.1 | podv and at least one of erpv or ercm | yes | hevd | 10.1.3 |

### HEVC viewport independent baseline profile

#### General (informative)

Both monoscopic and stereoscopic spherical video up to 360 degrees are supported. The profile requires neither viewport dependent delivery nor viewpoint dependent decoding. Regular HEVC encoders, DASH packagers, DASH clients, file format parsers, and HEVC decoder engines can be used for encoding, distribution and decoding. The profile also minimizes the options for basic interoperability.

#### Elementary stream constraints

The NAL unit stream shall comply with HEVC Main 10 profile, Main tier, Level 5.1.

All pictures shall be encoded as coded frames, and shall not be encoded as coded fields.

The following fields shall be set as follows:

* general\_progressive\_source\_flag shall be set to 1.
* general\_frame\_only\_constraint\_flag shall be set to 1.
* general\_interlaced\_source\_flag shall be set to 0.

When VUI is present, aspect\_ratio\_info\_present\_flag shall be set to 1 and aspect\_ratio\_idc shall be set to 1 (square).

For each picture, there shall be an equirectangular projection SEI message present in the bitstream that applies to the picture.

When the video is stereoscopic, for each picture, there shall be a frame packing arrangement SEI message present in the bitstream that applies to the picture.

When the video does not provide full 360 coverage, for each picture, there shall be a region-wise packing SEI messages present in the bitstream that applies to the picture. [Ed. (YK): This might not be not needed if the coverage information is present in the equirectangular projection SEI message. Check this.]

When present, the frame packing arrangement SEI messages and the region-wise packing SEI messages shall indicate constraints that comply with the equirectangular projected video scheme type 'erpv' specified in 7.2.1.2.

#### ISO base media file format constraints

compatible\_brands in FileTypeBox shall include 'ovid'.

Video sample entry type shall be equal to 'resv'.

Constraints for 'resv' tracks as specified in clause 7 apply.

scheme\_type values equal to 'podv' and 'erpv' shall be present within the SchemeTypeBox and CompatibleSchemeTypeBox. [Ed. (YK): Some more details needed here. For example, can both be signalled in CompatibleSchemeTypeBox and nothing in SchemeTypeBox? (MH): I don't see a need for more details. For restricted video, exactly one SchemeTypeBox is required to be present as per ISOBMFF. (YK): Can the closed-ended 'erpv' can be included in SchemeTypeBox and consequently the SchemeTypeBox does not include 'podv'? Such details may affect the value of the 'codecs' parameter and IMO should be clarified. Since we did not have clear decision on these aspects at the previous MPEG meeting, we can leave these to be discussed and clarified as needed at the next MPEG meeting. Also a minor detail should be clarified: the wording may be understood as requiring both scheme type values to be present in both SchemeTypeBox and CompatibleSchemeTypeBox; this confusion should be avoided.]

The type of OriginalFormatBox within the RestrictedSchemeInfoBox shall be equal to 'hvc1'.

NOTE: Consequently, parameter sets are not present inband within samples.

LHEVCConfigurationBox shall not be present in OriginalFormatBox.

HEVCConfigurationBox in OriginalFormatBox shall indicate conformance to the elementary stream constraints specified in 10.1.2.2.

[Ed. (MH): This constraint needs improved phrasing. (YK): I think this constraint can be removed as requiring of the 'hvc1' sample entry type is sufficient and what is said here is redundant.] For the Decoder Configuration Record in the Sample Description Box, the following applies:

* It shall contain one or more decoding parameter sets. (Containing VPS, SPS, and PPS NALs for HEVC Video). Each video Sample in the track shall reference a parameter set in the Sample entry.

When the video elementary stream contains a frame packing arrangement SEI message, StereoVideoBox shall be present. When StereoVideoBox is present, it shall signal the frame packing format that is included in the frame packing arrangement SEI message(s) in the elementary stream.

When the video elementary stream contains a region-wise packing SEI message, RegionWisePackingBox shall be present. When present, RegionWisePackingBox shall signal the same information as in the region-wise packing SEI message(s). [Ed. (YK): This might not be needed if the coverage information is present in the equirectangular projection SEI message. Check this.]

When the playback is intended to be started using another orientation than the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes, the initial viewpoint region-on-sphere metadata, as specified in 7.4.4, shall be present.

#### Receiver requirements

Receivers conforming to this media profile shall be capable of processing either all referenced SEI messages in 10.1.2.2 or all allowed boxes within the SchemeInformationBox for the equirectangular projected video scheme type.

#### CMAF media profile

This clause defines the CMAF Media Profile for the HEVC viewport independent baseline profile. This media profile may be signalled with the compatibility brand 'cvid'.

The CMAF Media Profile Track for the HEVC viewport independent baseline profile shall conform to both of the following:

* The constraints specified in 10.1.2.3.
* HEVC CMAF Video Track as defined in [CMAF], Annex B.1.

Note that by the combination of the two, only a restricted set of the HEVC CMAF Video Track may be used for this profile. Only 'hvc1' may be used based on the ISO BMFF Track Constraints. The presence and absence of the VUI parameters is given by CMAF.

A CMAF Switching Set for the HEVC viewport independent baseline profile shall conform to the CMAF Switching Set constraints as defined in [CMAF], Annex B.2.1.

In addition, for a CMAF Switching Set for the HEVC viewport independent baseline profile, the following applies:

* The same projection format shall be used for all CMAF Tracks in one CMAF Switching Set. [Ed. (YK): Redundant.]
* The same frame packing format shall be used for all CMAF Tracks in one CMAF Switching Set.
* The same coverage information shall be used for all CMAF Tracks in one CMAF Switching Set.
* The same spatial resolution shall be used for all CMAF Tracks in one CMAF Switching Set.

The mapping to CMAF Addressable Objects follows the rules in [CMAF], clause 7.6.

#### DASH integration

An instantiation of the HEVC viewport independent baseline profile in DASH should be represented as one Adaptation Set, possibly with multiple Representations. If so, the Adaptation Set should provide the following signalling:

* @codecs='resv.podv.hvc1.1.6.L93.B0'
* @mimeType=’video/mp4 profiles="ovid"’ [Ed. (YK): The origformat and schemetypes optional MIME type parameters parameters, which may have values origformat=hvc1.1.6.L93.B0 schemetypes="podv,erpv", are under discussion and have been agreed at the Friday Systems plenary in Torino to be included into the output document on in-advance signalling.]
* A Supplemental Descriptor or Essential Descriptor providing the frame packing arrangement may be used.

NOTE: By the use of the restricted video scheme and the @profiles referring to this media profile, the DASH client has all information to identify if this media profile can be played back. For additional information, the Supplemental Descriptor is used to provide some details on the configuration of the contained Representations.

The concatenation of all DASH Segments on one Representation for HEVC viewport independent baseline media profile shall conform to all the constraints specified in 10.1.2.3.

Conformance to CMAF may be provided in addition by conforming to a HEVC CMAF Video Track as defined in [CMAF], Annex B.1.

In addition, for an Adaptation Set the following applies:

* The same projection format shall be used on all Representations in one Adaptation Set. [Ed. (YK): Redundant.]
* The same frame packing format shall be used on all Representations in one Adaptation Set.
* The same coverage information shall be used on all Representations in one Adaptation Set.
* The same spatial resolution shall be used on all Representations in one Adaptation Set.

When the playback is intended to be started using another orientation than the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes, a Representation containing initial viewpoint region-on-sphere metadata, as specified in clause 7.4.4, shall be present and associated with all related media Representations as specified in 8.2.6.

### HEVC viewport dependent baseline profile

#### General (informative)

This profile allows unconstrained use of rectangular region-wise packing. With the presence of region-wise packing, the resolution of the omnidirectional video can be emphasized in certain regions, e.g., according to the user's viewing orientation. In addition, the sample entry type 'hvc2' is allowed, making it possible to use extractors and get a conforming HEVC bitstream when tile-based streaming is used.

#### Elementary stream constraints

The NAL unit stream shall comply with the same constraints as the HEVC viewport independent baseline profile, with the following exceptions:

* For each picture, there shall be either an equirectangular projection SEI message or a cubemap projection SEI message present in the bitstream that applies to the picture. [Ed. (YK): How frequent can the projection format switches from one to the other? Within a CVS the projection format shall not change?]
* When present, the frame packing arrangement SEI messages and the region-wise packing SEI messages shall indicate constraints that comply with the equirectangular projected video scheme type 'erpv' specified in 7.2.1.2 or the packed equirectangular or cubemap projected video scheme type 'ercm' specified in 7.2.1.3.

#### ISO base media file format constraints

compatible\_brands in FileTypeBox shall include 'hevd'.

Video sample entry type shall be equal to 'resv'.

Constraints for 'resv' tracks as specified in clause 7 apply.

scheme\_type values equal to 'podv' and at least one of 'erpv' and 'ercm' shall be present within the SchemeTypeBox and CompatibleSchemeTypeBox. [Ed. (YK): Some more details needed here. For example, can both be signalled in CompatibleSchemeTypeBox and nothing in SchemeTypeBox? (MH): I don't see a need for more details. For restricted video, exactly one SchemeTypeBox is required to be present as per ISOBMFF. (YK): Can the closed-ended 'erpv' can be included in SchemeTypeBox and consequently the SchemeTypeBox does not include 'podv'? Such details may affect the value of the 'codecs' parameter and IMO should be clarified. Since we did not have clear decision on these aspects at the previous MPEG meeting, we can leave these to be discussed and clarified as needed at the next MPEG meeting. Also a minor detail should be clarified: the wording may be understood as requiring both scheme type values to be present in both SchemeTypeBox and CompatibleSchemeTypeBox; this confusion should be avoided.]

The type of OriginalFormatBox within the RestrictedSchemeInfoBox shall be equal to 'hvc1' or 'hvc2'. [Ed. (YS/MH): Allowing 'hvt1' is under consideration.]

LHEVCConfigurationBox shall not be present in OriginalFormatBox.

HEVCConfigurationBox in OriginalFormatBox shall indicate conformance to the elementary stream constraints specified in 10.1.3.2.

The track\_not\_intended\_for\_presentation\_alone flag of the TrackHeaderBox may be used to indicate that a track is not intended to be presented alone.

When the playback is intended to be started using another orientation than the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes, the initial viewpoint region-on-sphere metadata, as specified in 7.4.4, shall be present.

#### DASH integration

Requirements on the presence of Essential or Supplemental Property Descriptors are the same as for the HEVC viewport independent baseline profile.

When the MPD contains a Representation with a track for which the OriginalFormatBox is equal to 'hvc2', the following applies:

* Either the Representations carrying a track as specified in 10.1.3.3 with the original format 'hvc2' shall contain @dependencyId listing all dependent Representations that carry a track as specified in 10.1.3.3 with the original format 'hvc1' or a Preselection property descriptor shall be present and constrained as follows:
  + The Main Adaptation Set shall contain a Representation carrying a track as specified in 10.1.3.3 with the original format 'hvc2'.
  + The Partial Adaptation Sets shall contain Representations each carrying a track as specified in 10.1.3.3 with the original format 'hvc1'.

NOTE 1: When using the Preselection property descriptor, the number of Representations for carrying 'hvc2' tracks is typically smaller than when using @dependencyId. However, the use of @dependencyId might be needed for encrypted video tracks.

* The Initialization Segment of the Representation that contains @dependencyId or belongs to the Main Adaptation Set is constrained as follows:
  + Tracks are constrained as in 10.1.3.3.
  + The track corresponding to the 'hvc2' original format refers to the tracks indicated in the 'tref' box of the Initialization Segment.

NOTE 2: When Preselection is used, the sequence\_number integer values are not required to be processed and therefore the concatenation of the Subsegments (of the different Representations of the Adaptation Sets of a Preselection) in any order results in a conforming file.

The following applies for the use of @mimeType:

* @mimeType of the Main Adaptation Set shall include the profiles parameter and 'hevd' within the profiles parameter. [Ed. (YS): Do we need to do something with ‘hevd’? I mean changing to sth. not coding specific. (YK): Why? This is an HEVC media profile anyway.]
* When Preselection is used, the value of profiles of the main Adaptation Set shall be the same as the value of profiles of its partial Adaptation Sets.
* When @dependencyId is used, the values of profiles of the respective dependent and complementary Representations shall be the same.

When Preselection is used, the following applies:

* The value of @subsegmentAlignment in the Main Adaptation Set shall be an unsigned integer and equal to the value of @subsegmentAlignment of the each associated Partial Adaptation Set.
* The value of @segmentAlignment in the Main Adaptation Set shall be an unsigned integer and equal to the value of @segmentAlignment of the each associated Partial Adaptation Set.

NOTE 3: The HEVC viewport dependent baseline profile typically requires a low delay operation and fast switching. This requires frequent stream access points (e.g., lower than 1 second interval) to be available, which can be achieved by providing different representations with different **Switching**@interval values or with 'sidx' boxes having different starts\_with\_SAP values for each of the subsegments.

When low latency considerations are relevant for the HEVC viewport dependent baseline media profile, the following DASH profiles provide tools to support efficient low latency services:

* ISO Base Media File Format On-Demand profile: urn:mpeg:dash:profile:isoff-on-demand:2011
* ISO-Base Media File Format Broadcast TV profile: urn:mpeg:dash:profile:isoff-broadcast:2015

It is recommended that DASH clients consuming low latency services support either or both of the above profiles in order to support the latency requirements.

When the playback is intended to be started using another orientation than the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes, a Representation containing the initial viewpoint region-on-sphere metadata, as specified in clause 7.4.4, shall be present and associated with all related media Representations as specified in 8.2.6.

## Audio profiles

### Overview

This clause defines media profiles for audio in OMAF. Table 10.2 provides an informative overview of the supported features. The detailed, normative specification for each audio profile is subsequently provided in the referred clause.

**Table 10.2 - Overview of OMAF media profiles for audio (informative)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Media Profile | Codec | Profile | Level | Max Sampling Rate | 3D Metadata | Brand | Clause |
| 3D audio baseline | MPEG-H Audio | Low complexity | 1, 2 or 3 | 48 kHz | included in codec | oabl | 10.2.2 |
| 2D audio legacy | AAC | HE‑AACv2 | 4 | 48 kHz | no 3D metadata | - | - |

### OMAF 3D audio baseline profile

#### General (informative)

This media profile fulfills the requirements to support 3D audio. Channels, objects and Higher-Order Ambisonics (HOA) are supported, as well as combinations of those. The profile is based on MPEG-H 3D Audio [3DA].

MPEG-H 3D Audio [3DA] specifies coding of immersive audio material and the storage of the coded representation in an ISOBMFF track. The MPEG-H 3D Audio decoder has a constant latency, see Table 1 of ISO/IEC 23008-3 [3DA]. With this information, content authors can synchronize audio and video portions of a media presentation, e.g. ensuring lip-synch. When orientation sensor inputs (i.e. pitch, yaw, roll) of an MPEG-H 3D Audio decoder change, there will be some algorithmic and implementation latency (perhaps tens of ms) between user head movement and the desired sound field orientation. This latency will not impact audio/visual synchronization (i.e. lip synch), but only represents the lag of the rendered sound field with respect to the user head orientation.

MPEG-H 3D Audio specifies methods for binauralizing the presentation of immersive content for playback via headphones, as is needed for omnidirectinal media presentations. MPEG-H 3D Audio specifies a normative interface for the user’s orientation, as Pitch, Yaw, Roll, and permits low-complexity, low-latency rendering of the audio scene to any user orientation.

#### Elementary stream constraints

The audio stream shall comply with the MPEG-H 3D Audio Low Complexity (LC) Profile, Levels 1, 2 or 3 as defined in ISO/IEC 23008-3, clause 4.8 [3DA]. The values of the mpegh3daProfileLevelIndication for LC Profile Levels 1, 2 and 3 are "0x0B," "0x0C," and "0x0D", respectively, as specified in ISO/IEC 23008-3 [3DA], clause 5.3.2.

Audio data shall be encapsulated into MPEG-H Audio Stream (MHAS) packets according to ISO/IEC 23008-3, clause 14 [3DA].

All MHAS packet types defined in ISO/IEC 23008-3, clause 14 [3DA] may be present in the stream, except of the following packet types that shall not be present in the stream:

* PACTYP\_CRC16
* PACTYP\_CRC32
* PACTYP\_GLOBAL\_CRC16
* PACTYP\_GLOBAL\_CRC32

If Audio Scene Information per ISO/IEC 23008-3, clause 15 [3DA] is present, it always shall be encapsulated in an MHAS PACTYP\_AUDIOSCENEINFO packet [3DA]. Audio Scene Information shall not be included in the mpegh3daConfig() structure in the MHAS PACTYP\_MPEGH3DACFG packet.

#### ISO base media file format constraints

[Ed. (YK): A file brand should be defined here and reflected in the overviewing table. Currently, a brand is only defined for CMAF.]

##### General constraints

The sample entry 'mhm1' shall be used for encapsulation of MHAS packets into ISOBMFF files, per ISO/IEC 23008‑3, clause 20.6 [3DA].

The sample entry 'mhm2' shall be used in cases of multi-stream delivery, i.e., the MPEG‑H Audio Scene is split into two or more streams for delivery as described in ISO/IEC 23008-3, clause 14.6 [3DA].

If the MHAConfigurationBox() is present, the MPEG-H profile and level indicator mpegh3daProfileLevelIndication in the MHADecoderConfigurationRecord() shall be set to "0x0B", "0x0C", or "0x0D" for MPEG-H Audio LC Profile Level 1, Level 2, or Level 3, respectively, as specified in ISO/IEC 23008-3 [3DA], clause 5.3.2.

The first sample of the movie and the first sample of every fragment (when applicable) shall be a Stream Access Point (SAP) of type 1 (i.e., sync sample). For MPEG-H Audio a sync sample shall be properly signalled according to ISO/IEC 14496-12 [ISOM]. All rules defined in ISO/IEC 23008-3, clause 20.6.1 [3DA] regarding sync samples shall apply. In, addition, a sync sample shall consist of MHAS packets in the following order:

* PACTYP\_MPEGH3DACFG
* PACTYP\_AUDIOSCENEINFO (if Audio Scene Information is present)
* PACTYP\_BUFFERINFO
* PACTYP\_MPEGH3DAFRAME

Additional MHAS packets may be present between the MHAS packets listed above or after the MHAS packet PACTYP\_MPEGH3DAFRAME, with one exception: if present, the PACTYP\_AUDIOSCENEINFO packet shall directly follow the PACTYP\_MPEGH3DACFG packet, as defined in ISO/IEC 23008-3, clause 14.4 [3DA].

MPEG-H Audio sync samples contain Immediate Playout Frames (IPFs), as specified in ISO/IEC 23008-3, clause 20.2 [3DA], thus the audio data encapsulated in the MHAS packet PACTYP\_MPEGH3DAFRAME shall contain the AudioPreRoll() syntax element, as defined in sub-clause 5.5.6 of ISO/IEC 23008-3 [3DA], and shall follow the requirements for stream access points as defined in clause 5.7 of ISO/IEC 23008-3 [3DA]. The audio configuration is delivered as part of the MHAS packet PACTYP\_MPEGH3DACFG and, therefore, the AudioPreRoll() structure carried in the MHAS packet PACTYP\_MPEGH3DAFRAME shall not contain the Config()structure, i.e., the configLen field of the AudioPreRoll() shall be 0.

##### Configuration change constraints

A configuration change takes place in an audio stream when the content setup or the Audio Scene Information changes (e.g., when changes occur in the channel layout, the number of objects etc.), and therefore new PACTYP\_MPEGH3DACFG and PACTYP\_AUDIOSCENEINFO packets are required upon such occurrences. A configuration change usually happens at program boundaries, but it may also occur within a program.

The following constraints apply:

* At each configuration change, the MHASPacketLabel shall be changed to a different value from the MHASPacketLabel in use before the configuration change occurred. A configuration change may happen at the beginning of a new ISOBMFF file or at any position within the file. In the latter case, the File Format sample that contains a configuration change shall be encoded as a sync sample (RAP) as defined above.
* A sync sample that contains a configuration change and the last sample before such a sync sample may contain a truncation message (i.e., a PACTYP\_AUDIOTRUNCATION packet in the MHAS stream) as defined in ISO/IEC 23008-3, clause 14.4 [3DA]. If MHAS packets of type PACTYP\_AUDIOTRUNCATION are present, they shall be used as described in ISO/IEC 23008-3, clause 14.4 [3DA].

ISOBMFF tracks that belong to one Audio Programme use different configurations and a switch between two ISOBMFF tracks represents also a configuration change. Thus, the MHASPacketLabel needs to have different values for all ISOBMFF tracks that belong to one Audio Programme. Also, after a configuration change the MHASPacketLabel needs to have different values for all ISOBMFF tracks comprising an Audio Programme.

##### Multi-stream constraints

The multi-stream-enabled MPEG‑H Audio System is capable of handling Audio Programme Components delivered in several different elementary streams (e.g., the main MHAS stream containing one complete audio main, and one or more auxiliary MHAS streams, containing different languages and audio description). The MPEG-H Audio Metadata information (MAE) allows the MPEG‑H Audio Decoder to correctly decode several MHAS streams.

The following constraints apply for file formats using the sample entry 'mhm2':

* One MHAS stream shall be the main stream, i.e., in exactly one MHAS stream the Audio Scene Information shall have the mae\_isMainStream field set to 1. In all other MHAS streams the mae\_isMainStream shall be set to 0.
* In each auxiliary MHAS stream (i.e., streams with mae\_isMainStream field set to 0) the mae\_bsMetaDataElementIDoffset field in the Audio Scene Information shall be set to the index of the first metadata element in the auxiliary MHAS stream minus one.
* All MHAS elementary streams that carry Audio Programme Components of one Audio Programme shall be time aligned.
* In each auxiliary MHAS elementary stream (i.e., streams with mae\_isMainStream field set to 0), RAPs shall be aligned to the RAPs present in the main stream (i.e., the stream with mae\_isMainStream field set to 1).
* Presentation Description Manifests need to make sure that all streams that contribute to one Audio Programme can be identified as such.
* For the main and the auxiliary MHAS stream(s), the MHASPacketLabel shall be set according to ISO/IEC 23008-3, clause 14.6 [3DA]. ISOBMFF tracks that belong to one Switching Set need to use different MHASPacketLabel values within the same range of values associated to one stream, as specified in ISO/IEC 23008-3, clause 14.6 [3DA]. For example, all ISOBMFF tracks in the Switching Set for the main stream use different values between 1 and 16, all ISOBMFF tracks in the Switching Set for the first auxiliary stream use values between 17 and 32, and so on.

##### Loudness and dynamic range control

Loudness metadata shall be embedded within the mpegh3daLoudnessInfoSet() structure as defined in ISO/IEC 23008-3, Clause 6.3 [3DA]. Such loudness metadata shall include at least the loudness of the content rendered to the default rendering layout as indicated by the referenceLayout field (see ISO/IEC 23008-3, Clause 5.3.2 [3DA]). More precisely, the mpegh3daLoudnessInfoSet() structure shall include at least one loudnessInfo() structure with loudnessInfoType set to 0, whose drcSetId and downmixId fields are set to 0 and which includes at least one methodValue field with methodDefinition set to 1 or 2 (see ISO/IEC 23008-3, Clause 6.3.1 [3DA] and ISO/IEC 23003-4, Clause 7.3 [DRC]). The indicated loudness value shall be measured according to applicable regional loudness regulations.

DRC metadata shall be embedded in the mpegh3daUniDrcConfig() and uniDrcGain() structures as defined in ISO/IEC 23008-3, Clause 6.3 [3DA]. For each included DRC set the drcSetTargetLoudnessPresent field as defined in ISO/IEC 23003-4, Clause 7 [DRC] shall be set to 1.

The bsDrcSetTargetLoudnessValueUpper and bsDrcSetTargetLoudnessValueLower fields shall be configured to continuously cover the range of target loudness levels between -31 dB and 0 dB. The embedded DRC metadata should allow for a decoder output loudness of at least -16 LKFS.

Loudness compensation information (mae\_LoudnessCompensationData()), as defined in ISO/IEC 23008-3, Clause 15.5 [3DA] shall be present in the Audio Scene Information if the mae\_allowGainInteractivity field (according to ISO/IEC 23008-3, clause 15.3 [3DA]) is set to 1 for at least one group of audio elements.

#### CMAF media profile

This clause defines the CMAF Media Profile for OMAF 3D Audio Baseline Profile. This media profile may be signalled with the compatibility brand ‘oabl’.

The CMAF Media Profile Track for OMAF 3D Audio Baseline Profile shall conform

1. to the ISO BMFF Track Format Constraints as defined in 10.2.2.3, and
2. to a MPEG-H CMAF Audio Track as defined in [CMAF], Annex J.

Using the combination of the two, only a restricted set of the MPEG-H CMAF Audio Track may be used for this profile.

A CMAF Switching Set for OMAF 3D Audio Baseline Profile shall conform to the an CMAF Switching Set constraints as defined in [CMAF], Annex J.

The transformation to CMAF Resources follows the rules in [CMAF], clause 6.

#### DASH integration

An OMAF Audio Baseline Profile may be included in DASH Media Presentations [DASH] for Streaming delivery.

An instantiation of an OMAF Audio Baseline Profile in DASH should be represented as one Adaptation Set. If so the Adaptation Set should provide the following signalling according to [RFC6381] and ISO/IEC 23008‑3, clause 21 [3DA] as shown inTable 10‑3.

**Table 10‑3 - MPEG-H Audio MIME parameter according to RFC 6381 and ISO/IEC 23008‑3**

| **Codec** | **MIME type** | **codecs parameter** | **profiles** | **ISOBMFF Encapsulation** |
| --- | --- | --- | --- | --- |
| MPEG-H Audio LC Profile Level 1 | audio/mp4 | mhm1.0x0B | ”oabl” | ISO/IEC 23008-3 |
| MPEG-H Audio LC Profile Level 2 | audio/mp4 | mhm1.0x0C | ”oabl” | ISO/IEC 23008-3 |
| MPEG-H Audio LC Profile Level 3 | audio/mp4 | mhm1.0x0D | ”oabl” | ISO/IEC 23008-3 |
| MPEG-H Audio LC Profile Level 1, multi-stream | audio/mp4 | mhm2.0x0B | ”oabl” | ISO/IEC 23008-3 |
| MPEG-H Audio LC Profile Level 2, multi-stream | audio/mp4 | mhm2.0x0C | ”oabl” | ISO/IEC 23008-3 |
| MPEG-H Audio LC Profile Level 3, multi-stream | audio/mp4 | mhm2.0x0D | ”oabl” | ISO/IEC 23008-3 |

The mapping to DASH Segment formats should be done by using one of the CMAF Resources

* CMAF Single Fragment Segment Mode is mapped to either the ISO Base media file format live profile as defined in clause 8.4 of [DASH] or the ISO Base media file format extended live profile as defined in clause 8.9 of [DASH].
* CMAF Multiple Fragment Segment Mode is mapped to either the ISO Base media file format live profile as defined in clause 8.4 of [DASH] or the ISO Base media file format extended live profile as defined in clause 8.9 of [DASH].
* CMAF Chunk Mode is mapped to Broadcast TV profile as defined in Clause 8.11 of [DASH].

CMAF Track File Mode is mapped to either the ISO Base media file format On Demand profile as defined in clause 8.3 of [DASH] or the Extended ISO Base media file format On Demand profile as defined in clause 8.8 of [DASH].

##### Element and attribute settings

Table 10‑4 summarizes the mapping of relevant MPD elements and attributes to MPEG‑H Audio.

**Table 10‑4 - Summary of relevant MPD elements and attributes for MPEG-H Audio**

|  |  |
| --- | --- |
| Element or Attribute Name | Description |
| @codecs | The signalling of the codecs parameters is according to [RFC6381] and ISO/IEC 23008-3, clause 21 [3DA]. The value consists of the following two parts separated by a dot:   * the fourCC code (mhm1, mhm2), * ‘0x’ followed by the hex value of the profile-level-id, as defined in in ISO/IEC 23008-3 [3DA].   See Table 10‑3 for more details. |
| AdaptationSet@tag | This field lists the mae\_groupIDs as defined in ISO/IEC 23008-3 [3DA] that are contained in the Adaptation Set separated by white spaces. |
| Preselection@tag | This field indicates the mae\_groupPresetID as defined in ISO/IEC 23008-3 [3DA] that refers to a Preset in the scope of MPEG-H Audio. |
| ContentComponent@tag | This field indicates the mae\_groupID as defined in ISO/IEC 23008-3 [3DA] which is contained in the Media Content Component. |
| AudioChannelConfiguration | For MPEG-H Audio, the Audio Channel Configuration descriptor shall use the scheme URI “urn:mpeg:mpegB:cicp:ChannelConfiguration”. The value shall be taken from the ChannelConfiguration table as defined in ISO/IEC 23091-3 [CICPa]. Valid numbers for value are 1-7,9-12, 14-17 or 19. |
| @audioSamplingRate | Example: "48000" for 48 kHz  The indication shall correspond to the sampling frequency derived from the usacSamplingFrequencyIndex or usacSamplingFrequency as defined in ISO/IEC 23003-3 [3DA]. |
| RandomAccess | The type to be used with MPEG-H Audio shall be “closed”, i.e., the SAP type is 1. |
| @mimeType | The MIME type to be used with MPEG-H Audio shall be ”audio/mp4”. |
| Language | The language indicated should correspond to the information conveyed in mae\_contentLanguage of the default dialog element. The maeGroup which is marked as default in mae\_switchGroupDefaultGroupID and is tagged in mae\_contentKind as dialogue. This information is carried in the AudioScene-Information() of the MPEG-H Audio stream as defined in ISO/IEC 23008-3 [3DA]. |
| Role | The Role for a Preselection should be set by the content author. |
| Accessibility | If the mae\_contentKind value of at least one Audio Element is set to ‘9’ (“audio-description/visually impaired”), an Accessibility descriptor shall indicate “descriptions” according to the Role scheme defined in ISO/IEC 23009-1 [DASH].  If at least the Audio Elements with a mae\_contentKind value of ‘2’ (“dialogue”) have mae\_allowGainInteractivity set to ‘1’ and mae\_interactivityMaxGain set to a non-zero value in the corresponding mae\_GroupDefinition() structure, an Accessibility descriptor with the value “enhanced-audio-intelligibility” according to the Role scheme defined in ISO/IEC 23009-1 [DASH] may be used to indicate that the Preselection enables the ability for a receiver to change the relative level of dialog to enhance dialog intelligibility.  The mae\_contentKind value of at least one Audio Element is set to ‘12’ (“emergency”), an Accessibility descriptor shall indicate “emergency” according to the Role scheme defined in ISO/IEC 23009-1 [DASH].  The accessibility information indicated for a Preselection should also correspond to the mae\_groupPresetKind.  The mae\_contentKind field and all other fields mentioned above that start with a “mae\_” prefix are carried in the AudioSceneInformation() of the MPEG-H Audio stream as defined in ISO/IEC 23008-3 [3DA]. |
| Label | The Label for a Preselection should be set by the content author. |

The concept of “Preselections” as defined in ISO/IEC 23009-1 [DASH] allows to offer different combinations of those Audio Components, either for automatic selection based on user preferences or for manual selection by the user. The Audio Components may be delivered in a single stream or in multiple streams.

Two different methods are defined to signal Preselections in the MPD: The Preselection Descriptor and the Preselection Element. The Preselection descriptor is defined in 5.3.11.2 of ISO/IEC23009-1 [DASH]. It enables simple setups and backward compatibility, but may not be suitable for advanced use cases.

The Preselection Element is defined in 5.3.11.3 and 5.3.11.4 of ISO/IEC23009-1 [DASH]. The Role and Accessibility descriptors on the Preselection Element, as well as other parameters, such as a profile & level indication on the @codecs attribute are related only to that Preselection and not to the stream(s) referenced by the Preselection element.

### OMAF 2D audio legacy profile

#### General (informative)

This media profile fulfills requirements to support 2D channel-based audio. The delivery of up to 5.1 channels is supported. The profile is based on MPEG-4 AAC [AAC], which defines coding of general audio content. The delivery of up to 5.1 audio channels allows 2D rendering according to user’s head orientation.

HE-AAC is used worldwide in the most successful streaming services and supported by all major streaming and media platforms. Due to the wide reach, MPEG-4 AAC can be used for VR services and platforms, which use either mono, stereo, 4.0, or 5.1 surround channel configurations. The 2D Audio Legacy profile does not require any new signalling for the audio codec and its configuration. Therefore, it is compatible with all decoder implementations in the market.

#### Elementary stream constraints

##### General encoding constraints

The audio stream shall comply with MPEG-4 AAC-LC, HE-AAC or HE-AACv2 profiles, Level 4, as defined in ISO/IEC 14496-3 [AAC].

For HE-AAC encoded tracks, the first sample of the ISO BMFF movie and the first sample of every ISO BMFF movie fragment (when applicable) shall be a SAP of type 1, notably, the SBR configuration information shall be present in the audio access unit.

ISO BMFF tracks containing AAC audio as defined in ISO/IEC 14496-3 [AAC] shall conform to the following AAC audio encoding constraints:

* The elementary stream shall be a raw data stream, i.e., ADTS and ADIF headers shall not be present.
* Each AAC elementary stream shall be encoded using MPEG-4 AAC LC, HE-AAC, HE-AACv2, Level 4. Use of the MPEG-4 HE-AACv2 for stereo configuration is recommended for 32 kbps or lower.
* When using HE-AAC and HE-AACv2, explicit backwards compatible signalling shall be used to indicate the use of the SBR and PS coding tools.
* AAC elementary streams shall not exceed 48kHz sampling rate.
* The channel count of an AAC ISO BMFF track shall not exceed six audio channels, including the LFE channel. [Ed. (AM): Consider improving the text.]
* AAC ISO BMFF fragments containing HE-AAC shall start with a type 1 SAP, notably, the SBR configuration information shall be in the first packet.
* The transform length of the IMDCT for AAC shall be 1024 audio PCM samples for long blocks, and 128 audio PCM samples for short blocks.
* The following parameters shall not change within the elementary stream
  + Audio Object Type
  + Sampling Frequency
  + Channel Configuration

The channelConfiguration parameter carried in the AudioSpecificConfig shall be set according to one of the following specified values:

* channelConfiguration = 1 (for mono audio),
* channelConfiguration = 2 (for stereo audio),
* channelConfiguration = 4 (for four channel audio),
* channelConfiguration = 5 (for five channel audio),
* channelConfiguration = 6 (for six channel audio, i.e., 5.1 audio).

Producing audio content capable of seamless bitrate adaptation with OMAF 2D Audio Legacy media profile (AAC‑LC, HE-AAC, HE-AACv2) requires constrained encoding at fragment boundaries. For such scenarios, each AAC elementary stream shall be encoded following the constraints provided in ISO/IEC 23000-19 [CMAF], from clause 10.5.2 to clause 10.5.6.

Encoding recommendations for AAC audio tracks are provided in ISO/IEC 23000-19 [CMAF], Annex G.

##### Syntax and values of syntactic elements

The syntax and values for syntactic elements shall conform to ISO/IEC 14496-3 [AAC]. The following element shall not be present in an MPEG-4 HE-AAC or HE-AACv2 elementary stream:

* coupling\_channel\_element (CCE)

If the program\_config\_element (PCE) element is present then it shall only list a set of channels corresponding to one of the fixed channel configurations specific in ISO/IEC 14496-3 [AAC], Table 1.19, and the element shall not change for the duration of the track.

The arrangement of syntactic elements shall be according to Table 1.19 of ISO/IEC 14496-3 [AAC]. For convenience, the arrangement of elements for the allowed channel configurations is reported in Table 10‑5.

**Table 10‑5 – Arrangement of Audio syntactic elements**

|  |  |  |
| --- | --- | --- |
| Channel Configuration | Number of Channels | Audio syntactic elements |
| 1 | 1 | <SCE>, <optional additional elements>, <TERM>, for HE-AAC v2, and mono HE-AAC or AAC-LC |
| 2 | 2 | <CPE>, <optional additional elements>, <TERM>, for stereo HE-AAC or AAC‑LC |
| 4 | 4 | <SCE>, <CPE>, <SCE>, <optional additional elements>, <TERM> |
| 5 | 5.0 | <SCE>, <CPE>, <CPE>, <optional additional elements>, <TERM> |
| 6 | 5.1 | <SCE>, <CPE>, <CPE>, <LFE>, <optional additional elements>, <TERM> |

NOTE Angled brackets (<>) are used above to indicate separate syntactic elements, not stream syntax.

The syntax and values for individual\_channel\_stream shall conform to ISO/IEC 14496-3 [AAC]. The following fields shall be set as defined:

* gain\_control\_data\_present = 0

##### AAC presentation timing

The AAC codec uses audio frames of a fixed length, and a transform which applies over two frames. To obtain correct audio from a frame, both frames in the transform are needed, and hence the prior encoded frame and the current encoded frame need to be decoded to output the first frame. This is sometimes called “priming” and may be signalled using the ‘roll’ sample group.

A full reconstruction of the first encoded audio frame is sometimes not possible since there is no previous access unit. To still achieve a full reconstruction, a common practice is to add silence to the beginning of the audio signal. A more detailed explanation of this approach can be found in ISO/IEC 14496-24 [14496-24].

In practice, an encoder might prepend an arbitrary amount of (invalid) audio waveform samples to the signal. This portion of the audio signal is sometimes called “encoder delay” and varies depending on the implementation.

Presentation delay is compensated according to one of the following options:

* The most common approach to compensate for inserted extra audio is to add an offset edit list to the ISO BMFF header. In the case where padding has been added to the start of an audio stream, the media\_time in the edit list is the length (in audio samples, as measured by the timescale of the track) of the inserted audio samples; 2112 is a common example for AAC.
* If the content has been generated according to ISO/IEC 23000-19 [CMAF], Annex G.5, no EditListBox is present.
* If the SBR and PS coding tools are present, they shall not be considered for the purpose of delay compensation.

##### Loudness and dynamic range control

The audio stream should contain DRC and loudness metadata according to ISO/IEC 14496-3 [AAC]. The audio encoder should set the Program Reference Level to the loudness level of the audio stream.

The audio encoder should generate DRC metadata for light compression encoded in the dyn\_rng\_ctl and dyn\_rng\_sgn fields of dynamic\_range\_info() in the FIL element and DRC metadata for heavy compression in the compression\_value field of MPEG4\_ancillary\_data() in the data stream element (DSE).

NOTE It is expected that the audio decoder will use the Program Reference Level, if available, to achieve a desired target loudness, if applicable. It is expected that the audio decoder will apply the DRC metadata, if present, according to ISO/IEC 14496-3 [AAC] including the DRC Presentation Mode value of the drc\_presentation\_mode fields.

##### Maximum bitrate

The maximum bitrate of AAC elementary streams shall be calculated in accordance with the AAC buffer requirements as defined in ISO/IEC 14496-3 [AAC], clause 4.5.3. Only the raw data stream shall be considered in determining the maximum bitrate (system-layer descriptors are excluded).

#### ISO base media file format constraints

[Ed. (YK): A file brand should be defined here and reflected in the overviewing table.]

The syntax and values of the AudioSampleEntry shall conform to MP4AudioSampleEntry ('mp4a') as defined in ISO/IEC 14496-14 [MP4FF]. Table 10‑6 lists the allowed AAC profiles.

Table 10‑6 – AAC profiles

|  |  |  |
| --- | --- | --- |
| AAC profile | codingname | SampleEntry Type |
| MPEG-4 AAC (AAC-LC) | mp4a | MP4AudioSampleEntry |
| MPEG-4 High Efficiency AAC (HE-AAC) | mp4a | MP4AudioSampleEntry |
| MPEG-4 High Efficiency AAC v2 (HE-AACv2) | mp4a | MP4AudioSampleEntry |

The SampleEntry format in the SampleDescriptionBox is the same for each AAC audio profile.

##### Storage of AAC media samples

The following additional constraints apply:

* All audio media samples shall consist of one AAC audio access unit.
* All AAC access units in an ISOBMFF track shall be encoded with one of AAC LC, HE-AAC or HE-AACv2.
* The values given in AudioSampleEntry, DecoderConfigDescriptor, and DecoderSpecificInfo shall match the corresponding values in the AAC audio bitstream.

##### AAC audio sample entry

The syntax and values of the AudioSampleEntry shall conform to MP4AudioSampleEntry ('mp4a') as defined in [MP4FF].

The sample entry and fields specified in this section shall not change within an ISOBMFF track.

The value of the channelcount parameter in the AudioSampleEntry box defined in ISO/IEC 14496-3 [ISOM] shall be set to one of the following specified values:

* channelcount = 1 (for mono audio),
* channelcount = 2 (for stereo audio),
* channelcount = 4 (for four channel audio),
* channelcount = 5 (for five channel audio),
* channelcount = 6 (for six channel audio, i.e., 5.1 audio).

The value of the channelcountparameter in the AudioSampleEntry box shall correspond to the values of channelConfiguration field of AudioSpecificConfig according to Table 10‑7:

Table 10‑7 - Mapping of channelcount parameter in the AudioSampleEntry to channelConfiguration field of AudioSpecificConfig

|  |  |
| --- | --- |
| channelcount | channelConfiguration |
| 1 | 1 |
| 2 | 2 |
| 4 | 4 |
| 5 | 5 |
| 6 | 6 |
| The channel to loudspeaker mapping for each channelConfiguration index is given in Table 1.19, ISO/IEC 14496-3 [AAC]. The informative geometric speaker positions for channelConfiguration = 4 (Quadrophonic speaker layout) is 0, 90°, -90°, 180° deg (azimuth) | |

###### **ES\_Descriptor**

The syntax and values for ES\_Descriptor shall conform to ISO/IEC 14496-1 [MP4SYS], and the fields of the ES\_Descriptor shall be set to the following values.

* ES\_ID = 0
* streamDependenceFlag = 0
* URL\_Flag = 0;
* OCRstreamFlag = 0
* streamPriority = 0
* decConfigDescr = DecoderConfigDescriptor
* slConfigDescr = SLConfigDescriptor, predefined type 2

Descriptors other than those specified in 10.2.3.3.2.2 through 10.2.3.3.2.4 shall not be used.

###### **DecoderConfigDescriptor**

The syntax and values for DecoderConfigDescriptor shall conform to ISO/IEC 14496-1 [MP4SYS], and the fields of this descriptor shall be constrained to the following values.

* decoderSpecificInfo shall be used, and ProfileLevelIndicationIndexDescriptor shall not be used.
* objectTypeIndication = 0x40 (Audio)
* streamType = 0x05 (Audio Stream)
* upStream = 0
* decSpecificInfo = AudioSpecificConfig

###### **AudioSpecificConfig**

The syntax and values for AudioSpecificConfig shall conform to ISO/IEC 14496-3 [AAC].

The following fields of AudioSpecificConfig shall be set according to ISO/IEC 14496-3 [AAC] and 10.2.3.2:

* audioObjectType
* channelConfiguration
* extensionAudioObjectType
* GASpecificConfig

###### **GASpecificConfig**

The syntax and values for GASpecificConfig shall conform to ISO/IEC 14496-3 [AAC], and the fields of GASpecificConfig shall be set to the following values:

* frameLengthFlag = 0 (1024 lines IMDCT)
* dependsOnCoreCoder = 0
* extensionFlag = 0

#### DASH integration

An OMAF 2D Audio Legacy Profile may be included in DASH Media Presentations ISO/IEC 23009-1 [DASH] for Streaming delivery.

An instantiation of an OMAF 2D Audio Legacy Profile in DASH should be represented as one Adaptation Set. If so, the Adaptation Set shall provide the following signalling according to [RFC6381] as shown in Table 10‑8. The 3rd value of the codecs parameter is the audioObjectType, as if explicit hierarchical signalling were used.

Table 10‑8 – AAC MIME “Codecs” parameter according to RFC6381

|  |  |  |
| --- | --- | --- |
| AAC profiles | MIME type | codecs parameter |
| MPEG-4 AAC (AAC-LC) | audio/mp4 | mp4a.40.2 |
| MPEG-4 High Efficiency AAC (HE-AAC) | audio/mp4 | mp4a.40.5 |
| MPEG-4 High Efficiency AAC v2 (HE-AACv2) | audio/mp4 | mp4a.40.29 |

NOTE: HE-AAC is a superset of AAC-LC, and HE-AACv2 is a superset of HE-AAC. An HE-AACv2 decoder is capable of decoding HE-AAC or AAC-LC, and an AAC-LC decoder is capable of partially decoding HE-AAC and HE-AACv2 conforming to CMAF constraints (without reproduction of high frequencies coded with SBR).

# Presentation Profiles under Consideration

# Introduction

Presentation Profiles provide a full omnidirectional experience.

# OMAF Baseline Viewport-Independent Presentation Profile

# Introduction

The OMAF Baseline Viewport-Independent Presentation Profile is intended to provide highest interoperability and quality on the mobile-powered Head-Mounted Displays.

This profile fulfils basic requirements to support 3D Audio and omnidirectional and 3D video. Both monoscopic and stereoscopic video is supported. The profile does neither require viewport dependent decoding nor viewpoint dependent delivery.

The profile also minimizes the options for basic interoperability.

# Definition

Requirements of OMAF Main Presentation Profile.

* If containing video, it shall contain at least one component following the HEVC viewport independent baseline profile as defined in 10.1.2
* If containing audio, it shall contain at least one component following the OMAF 3D Audio Baseline Profile as defined in 10.2.

# ISO Base Media File format

An ISO BMFF file for which the content author considers that the VR experience is included in this one file using the technologies for the OMAF Main Presentation Profile may be offered using the ISO BMFF file brand ‘ompp’.

For a file with compatibility brand ‘ompp’ the following holds.

* The file shall conform to the ‘iso9’ brand.
* If containing video, the file shall contain at least one track following the HEVC viewport independent baseline profile track format as defined in 10.1.2.3.
* If containing audio, the file shall contain at least one track following the OMAF 3D Audio Baseline Profile track format as defined in 10.2.2.3.

1. Additional media profiles under consideration
   1. AVC viewport dependent profile
      1. General (informative)

This media profile allows unconstrained use of rectangular region-wise packing with AVC. With the presence of region-wise packing, the resolution of the omnidirectional video can be emphasized in certain regions, e.g. according to the user's viewing orientation.

* + 1. Elementary stream constraints

The video NAL unit stream shall comply with AVC High profile, Level 5.1.

[Ed. (YS/MH): More constraints on elementary streams are needed, e.g. progressive only.]

* + 1. ISO base media file format constraints

compatible\_brands in FileTypeBox shall include 'avde'.

Video sample entry type shall be equal to 'resv'.

Constraints for 'resv' tracks as specified in clause 7 apply.

scheme\_type values in SchemeTypeBox and CompatibleSchemeTypeBox(es) shall include 'podv'. [Ed. (YS/MH): A new scheme\_type should be specified similarly to those for done for HEVC profiles, also indicating constraints on projection.]

projection\_type shall be equal to 0 (the equirectangular projection) in the ProjectionFormatBox within the SchemeInformationBox.

The type of OriginalFormatBox within the RestrictedSchemeInfoBox shall be equal to 'avc1' or 'avc3'.

AVCConfigurationBox in OriginalFormatBox shall indicate conformance to the elementary stream constraints specified in C.1.1.

RegionWisePackingBox and StereoVideoBox may be present in SchemeInformationBox.

When the playback is intended to be started using another orientation than the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes, the initial viewpoint region-on-sphere metadata, as specified in 7.4.4, shall be present.

* + 1. DASH integration

The **AdaptationSet**@SegmentAlignment attribute shall be present and shall have a value of 'true' or '1'.

The **AdaptationSet**@startsWithSAP attribute shall be present and shall have a value of 'true' or '1'.

The @duration in the segment list or segment template and the **MPD**@minBufferTime should not exceed 2 seconds.

The ISOBMFF live profile or the ISOBMFF main profile may be used for viewport-dependent streaming in DASH.

When the playback is intended to be started using another orientation than the orientation (0, 0, 0) in (yaw, pitch, roll) relative to the global coordinate axes, a Representation containing the initial viewpoint region-on-sphere metadata, as specified in clause 7.4.4, shall be present and associated with all related media Representations as specified in 8.2.6.

* 1. HEVC viewport independent fisheye video profile
     1. General (Informative)

This media profile fulfils basic requirements to support omnidirectional video via multiple circular images captures by fisheye cameras. The profile requires neither viewport dependent delivery nor viewpoint dependent decoding. Regular HEVC encoders, DASH packagers, DASH clients, file format parsers, and HEVC decoder engines can be used for encoding, distribution and decoding. The profile also minimizes the options for basic interoperability.

* + 1. Elementary stream constraints

The NAL unit stream shall comply with HEVC Main 10 profile, Main tier, Level 5.1.

All pictures shall be encoded as coded frames, and shall not be encoded as coded fields.

The following fields shall be set as follows:

* general\_progressive\_source\_flag shall be set to 1.
* general\_frame\_only\_constraint\_flag shall be set to 1.
* general\_interlaced\_source\_flag shall be set to 0.

When VUI is present, aspect\_ratio\_info\_present\_flag shall be set to 1 and aspect\_ratio\_idc shall be set to 1 (square).

* + 1. ISO base media file format constraints

compatible\_brands in FileTypeBox shall include 'fodv'. [Ed. (MH): It might be better to use some other 4CC here to avoid confusion with scheme type 'fodv'.]

Video sample entry type shall be equal to 'resv'.

Constraints for 'resv' tracks as specified in clause 7 apply.

scheme\_type value equal to 'fodv' shall be present within the SchemeTypeBox and CompatibleSchemeTypeBox.

The type of OriginalFormatBox within the RestrictedSchemeInfoBox shall be equal to 'hvc1'.

NOTE: Consequently, parameter sets are not present inband within samples.

LHEVCConfigurationBox shall not be present in OriginalFormatBox.

HEVCConfigurationBox in OriginalFormatBox shall indicate conformance to the elementary stream constraints specified in C.2.2.

[Ed. (MH): This constraint needs improved phrasing. (YK): I think this constraint can be removed as requiring of the 'hvc1' sample entry type is sufficient and what is said here is redundant.] For the Decoder Configuration Record in the Sample Description Box, the following applies:

* It shall contain one or more decoding parameter sets. (Containing VPS, SPS, and PPS NALs for HEVC Video). Each video Sample in the track shall reference a parameter set in the Sample entry.
  + 1. Receiver requirements

Receivers conforming to this media profile shall be capable of processing all allowed boxes within the SchemeInformationBox for the fisheye omnidirectional video scheme type.

* + 1. CMAF media profile

This clause defines the CMAF Media Profile for the HEVC viewport independent fisheye video profile. This media profile may be signalled with the compatibility brand 'cvid'. [Ed. (MH): 'cvid' is used for HEVC viewport independent baseline profile. Should a different 4CC used here?]

The CMAF Media Profile Track for the HEVC viewport independent fisheye video profile shall conform to both of the following:

* The constraints specified in C.2.3.
* HEVC CMAF Video Track as defined in [CMAF], Annex B.1.

Note that by the combination of the two, only a restricted set of the HEVC CMAF Video Track may be used for this profile. Only 'hvc1' may be used based on the ISO BMFF Track Constraints. The presence and absence of the VUI parameters is given by CMAF.

A CMAF Switching Set for the HEVC viewport independent fisheye video profile shall conform to the CMAF Switching Set constraints as defined in [CMAF], Annex B.2.1.

In addition, for a CMAF Switching Set for the HEVC viewport independent fisheye video profile, the following applies:

* The same fisheye parameters shall be used for all CMAF Tracks in one CMAF Switching Set.

The mapping to CMAF Addressable Objects follows the rules in [CMAF], clause 7.6.

* + 1. DASH integration

An instantiation of the HEVC viewport independent fisheye video profile in DASH should be represented as one Adaptation Set, possibly with multiple Representations. If so, the Adaptation Set should provide the following signalling:

* @codecs='resv.fodv.hvc1.1.6.L93.B0'
* @mimeType=’video/mp4 profiles="fodv"’

NOTE: By the use of the restricted video scheme and the @profiles referring to this media profile, the DASH client has all information to identify if this media profile can be played back.

The concatenation of all DASH Segments on one Representation for HEVC viewport independent fisheye video media profile shall conform to all the constraints specified in C.2.3.

Conformance to CMAF may be provided in addition by conforming to a HEVC CMAF Video Track as defined in [CMAF], Annex B.1.

In addition, for an Adaptation Set the following applies:

* The same fisheye parameters shall be used for all Representations in one Adaptation Set.
  1. Timed text profile
     1. General (Informative)

This media profile can be used for providing subtitles and closed captions for videos.

* + 1. Elementary stream constraints

The elementary stream shall conform to either WebVTT or TTML/IMSC1.

[Ed. (MH): The following normative references should be added, if this profile is accepted:

WebVTT: The Web Video Text Tracks Format, W3C Working Draft 08 December 2015

W3C Working Draft, Timed Text Markup Language 2 (TTML2)

W3C Recommendation, TTML Profiles for Internet Media Subtitles and Captions 1.0 (IMSC1)]

[Ed. (MH): It should be considered whether one or two timed text profiles are specified.]

* + 1. ISO base media file format constraints

[Ed. (MH): It is probably reasonable to specified a brand 4CC here.]

Tracks shall conform to ISO/IEC 14496-30.

Rendering of the timed text may be done on the projected 2D video prior to VR rendering, in which case the text cue/region positions are relative to the full ERP video, or it can be done on the current viewport, in which case the text/cue region positions are relative to the current viewport.

When WebVTT is used, WVTTSampleEntry shall include ExtendedConfigBox .

When TTML or IMSC1 is used, ExtendedConfigBox shall be present in XMLSubtitleSampleEntry. The disparity information in ExtendedConfigBox is superceded by the tts:disparity attribute of TTML.

[Ed. (MH): If this profile is accepted, the definition of ExtendedConfigBox should probably be moved to clause 7.]

ExtendedConfigBox has the following syntax and semantics:

class ExtendedConfigBox extends Box(‘ttec’) {  
 unsigned int(1) viewport\_relative\_flag;  
 unsigned int(7) reserved;  
 if (viewport\_relative\_flag == 0) {  
 unsigned int(1) spherical\_region\_flag;  
 unsigned int(7) reserved;  
 if (spherical\_regions\_flag == 1) {  
 RegionOnSphereStruct();  
 unsigned int(16) region\_depth;  
 }  
 } else {  
 unsigned int(1) relative\_disparity\_flag;  
 unsigned int(7) reserved;  
 if (relative\_disparity\_flag)  
 signed int(16) disparity\_in\_percent;  
 else  
 signed int(16) disparity\_in\_pixels;  
 }  
}

[Ed. (MH): The semantics below should be phrased more precisely. Currently they leave some ambiguity.]

viewport\_relative\_flag: flag that indicates if the text cues and regions are positioned relative to the viewport or to the full video.

spherical\_region\_flag: if the text cue boxes and regions are relative to the full video, this flag indicates if the regions are given in spherical coordinates.

RegionOnSphereStruct provides a definition of a timed text region on the sphere. This region is used as anchor for all text cues in this track.

region\_depth: indicates the depth (z-value) of the region on which the timed text is to be rendered. This value is relative to a normalized sphere with radius 1.0 and is scaled by a factor of 65536.

relative\_disparity\_flag: this flag indicates if the disparity is provided as a percentage value of the width of the half view, when the flag is set to 1, or as pixel value, when the flag is set to 0.

disparity\_in\_percent: provides a percentage value of the width of the half view scaled by a factor of 32768 that indicates the disparity. The value may be negative.

disparity\_in\_pixels: provides the disparity in pixel value scaled by a factor of 32768. The value may be negative.

[Ed. (MH): Further study is ongoing on allowing the use of a timed metadata track for sphere region metadata to define the timed text region.]

1. Submission and requirements fulfilment information
   1. HEVC viewport independent baseline profile
      1. General submission information

For the submission of a profile, the following information is required

* Please attach a full technical description of the proposal to the submission, based on the text of the DIS of OMAF (as provided in N16824). The profile should be limited to referencing tools of the DIS of OMAF and any other specifications that are either complete or expected to be published by the end of 2017.
  + The complete proposal is in 10.1.2
* Please indicate if the submission addresses
  + Media Profile for Video
* Please provide a summary of the proposal
  + This media profile fulfils basic requirements to support omnidirectional video. Both monoscopic and stereoscopic 360 video is supported. The profile does neither require viewport dependent decoding nor viewpoint dependent delivery. Regular DASH clients, file format parsers and HEVC decoders engines can be used for distribution and decoding. The profile also minimizes the options for basic interoperability.
* Please provide information on
* Why a new profile is required?
  + This is an initial profile in order to serve basic interoperability using existing decoders and distribution infrastructures.
  + According to 3GPP TR26.917 (S4-170752), a 4K spatial resolution wit ERP already provides good quality and statistically significant better quality cannot be achieved by on existing setups by increasing spatial resolution.
* How is it significantly different from the existing profiles?
  + No profile exists yet
* If an existing profile can be adapted to accommodate the use case?
  + No profile exists yet
* Supporters for the Profile (Industry Backing and Interest for Deployment with preferably multiple companies, e.g. at least 3, interested in an interoperable end-to-end solution)
  + Qualcomm Incorporated
  + Technicolor
  + Samsung
  + Ericsson
  + Fraunhofer HHI
* Information by when test vectors and conformance software would be available (if the profiles is adopted) and if a reference receiver may be available as well.
  + The development of test content is currently considered for Sep 2017 in context of the VR-IF/DASH-IF interoperability efforts
  + The development of conformance software is currently considered for Sep 2017 in context of the VR-IF/DASH-IF interoperability efforts
  + The development of a reference receiver is currently not considered, but the extension of dash.js to support this media profile for VR is under consideration.
* The response to the requirements table as provided in clause 2.2.
  + See below

For the submission of a profile, the following information is recommended

* Any information on existing implementations, e.g. demos
  + Implementations exist based on proprietary services, details to be added
  + Demos are planned to be done in the context of the VR-IF efforts.
* How the profile addresses certain use cases as documented in clause 2.3.
  + See below

Any additional information may be provided such as how the performance can be improved with optional metadata included in OMAF.

* + 1. Addressing the requirements

Table B.1 provides the fulfilment matrix. Please submit along with your submission.

Table B.1 Requirements and Fulfilment Matrix (FF = y/es, n/o, p/atrial, d/oes not apply, o/ther – add comment)

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Requirement | FF | Comment |
| General | | | |
| 1 | The Specification shall provide for interoperable exchange of VR360 content | y | The profile is fully based on OMAF, ISO BMFF, CMAF and DASH and provides minimum options to fulfill interoperability. |
| 2 | The Specification shall avoid providing multiple tools for the same functionality to reduce implementation burden and improve interoperability. | y | The profile has minimum options and therefore fulfills the basic requirements for interoperability. |
| 3 | The Specification shall enable good quality and performance. | y | HEVC level 5.1 supports up to 4k by 2k at reasonable frame rates and therefore provides satisfactory quality at maximum simplicity. |
| 4 | The Specification shall enable full interoperability between services/content and clients. | y | The specification provides a full specification for download and streaming delivery for video. The interoperability is provided by the strict definition of the media profile requirements for the bitstream, the receiver and the signalling of the media profile. |
| 4.1 | The Specification shall contain a very low number of fully specified interoperability points that include what is traditionally known as Profile and Level information. | y | This is the first media profile |
| 4.2 |  |  |  |
| 4.1.1 | The existence of more than one interoperability point shall be justified if intended to target devices with different capabilities. | y | This is the first media profile |
| 4.2 | Interoperability points shall address a Media Profile including:   * file format tracks and elementary stream * 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 | y | All issues are specified in the specification above in the relevant clauses |
| 4.3 | 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 | P | This addresses a media profile which may be added to a presentation profile. |
| 4.4 | These interoperability points shall enable conformance to be tested, inside and outside of MPEG. | y | The description of the requirements are clearly documented in the profiles. |
| 4.5 | The Specification may contain partial interoperability points (e.g., a file format box, a visual media profile) at a lower level of granularity, to enable external bodies to specify their own full interoperability points. | P | The media profile may be used in a full presentation profile as defined in 4.3. |
| 4.6 | 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. | n/a |  |
| 4.7 | The specification shall define at least one media profile for audio. | n/a |  |
| 4.8 | The specification shall define at least one media profile for video | y | Fulfilled by the media profile |
| 4.9 | The specification shall define at least one presentation profile that includes one audio and one video media profile. | n/a |  |
| 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. | Y | The supported features and options are supported by the referenced devices. |
| 6 | The Specification shall support the representation, storage, delivery and rendering of   * Omnidirectional (up to 360° spherical) coded image/video (monoscopic and stereoscopic) with 3 DoF * Both 3D and 2D audio | P | The media profile addresses the first aspect |
| 7 | The specification shall work with existing MPEG storage and delivery formats | Y | The media profile describes on how to carry the data in ISO BMFF, CMAF and DASH. Existing parsers, CMAF Players and DASH clients can be used as long as the rendering pipeline supports SEI messages documented in clause 2.2.2. |
| 8 | The Specification shall support temporal synchronization and spatial alignment between different media types, in particular between audio and video. | P | Addressed by the usage of OMAF |
| 9 | The Specification shall support metadata for describing initial viewpoints and for the playback of omnidirectional video/image and audio according to that metadata. | P | The media profile does not prevent this information. |
| 10 | The Specification shall support the following interfaces:   * encoding and decoding for each media type * delivery for download and streaming | Y | Clause 2.2.2 and 2.2.3 and 2.2.4 address the aspects |
| 11 | The Specification shall enable applications to use hardware-supported or pre-installed independently manufactured decoders and renderers through defined MPEG conformance points. | Y | HEVC level 5.1 is broadly supported on existing HMDs as indicated by the LS from VR-IF |
| 12 | The Specification shall support viewport-dependent processing (this may include delivery, decoding and rendering).   * The Specification shall support dynamically changing viewports * The Specification should enable responsiveness to changing viewport in a way that doesn’t detract from the immersive experience | Y | In all cases the full 360 scene is decoded and is made available to the renderer. The renderer will be able to adjust the viewport without involving the DASH client or the network. |
| 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* | Y | This media profile address this issue for video. |
| Delivery | | | |
| 14 | The Specification shall support the following methods of distribution:   * File-based delivery * DASH-based streaming * MMT-based streaming | P | This mapping to ISO BMFF and DASH streaming is provided for this media profile. |
| 15 | 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 |  | Template Error |
| Visual | | | |
| 16 | The Specification shall enable content exchange with high visual perceptual quality. | Y | HEVC level 5.1. permits decent quality for 360 video. |
| 16.1 | 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). | P |  |
| 16.2 | The Specification shall support a framerate of at least 60fps | Y | HEVC level 5.1. permits 4K at 60 fps. |
| 17 | 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. | P | The content would need to be downsampled to be decodable by an HEVC level 5.1 decoder. The subsampling may be done in spatial or temporal domain or by reducing the coverage. |
| 18 | The Specification shall support metadata for the rendering of spherical video on a 2D screen | Y | Supported by OMAF |
| 19 | The Specification shall support fisheye-based video with a configuration of 2 cameras | n/a | Not addressed by the profile, but no requirement for each media profile. |
| 20 | The Specification shall support encoding of equirectangular projection (ERP) maps for monoscopic and stereoscopic video, in an efficient manner. | Y | Fully supported by the media profile. |
| 20.1 | Other projection maps than ERP for distribution should only be provided if consistent benefits over ERP is demonstrated | n/a | This is an issue for the main specification, but ERP is supported. |
| Audio | | | |
| 21 | Each audio media profile in the Specification shall   1. support immersive rendering with sufficiently low latency   Note: this is also expressed in requirement 12   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. | n/a |  |
| 22 | There may be one audio media profile that supports only 2D audio to cater to existing devices  All other audio media profiles defined in the Specification shall:   * support 3D Audio distribution, decoding & rendering. * support immersive content, e.g. 12ch or 3rd order Ambisonics, * support a combination of diegetic and non-diegetic content sources. * be capable to ingest and carry all content types:   + audio channels,   + audio objects,   + scene-based audio,   + and combinations of the above. * be able to carry dynamic meta-data for combining, presenting and rendering all content types. | n/a |  |
| Security | | | |
| 23 | The Specification shall not preclude:   * Decoding and rendering to support secure media pipelines * Efficient distribution for multiple DRM systems (e.g. using common encryption´)   The Specification should enable a secure media pipeline to be implemented. | Y | The use of a linear decoding flow w/o requiring to do complex pixel-based modification permits to use regular common encryption.  Also the usage of a single video stream provides the ability to apply regular common encryption.  Compatibility to HEVC CMAF ensure broad support for encrypted content as well. |

* + 1. Response to certain Scenario

Consider the following scenario. A content provider wants to provide a 360VR service with 3D audio to mobile HMD with head motion tracking. The content enables to change the field-of-view based on user interaction.

The HMD support HEVC Main-10 level 5.1 video decoder and an MPEG audio decoder. The original content is

* Basic VR content: as low as 4k x 2k (ERP), 8 or 10bit, BT.709, as low as 30fps, monoscopic and stereoscopic
* High-quality: up to 8k x 4k (ERP), 10 bit, possibly advanced transfer characteristics and colour transforms, sufficiently high frame rates, etc.
* Spatial audio content for immersive experiences, provided in the following formats:
  + Channel-based audio
  + Object-based audio
  + Scene-based audio
  + Or a combination of the above
  + Sufficient metadata for encoding, decoding and rendering the spatial audio scene permitting dynamic interaction with the content. The metadata may include additional metadata that is also used in regular TV applications, such as for loudness management.
  + Diegetic and non-diegetic audio content.

The content may need encryption for DRM purposes. The content may be downloaded and played locally, or the content may be made available by the use of DASH-based streaming or other adaptive streaming technologies.

How does the proposed profile address this scenario. Please provide a summary

* to what extent the profile addresses and improve such a scenario
  + The scenario is addressed for video for basic quality video as well as for high-quality video to at least some extent.
* on how the service provider using the profile for distribution can integrate it into the content generation, e.g. what are the necessary interfaces from the content production to the encoder
  + The content can be provided in typical formats such as ERP and can be prepared for distribution. The content processing is simple.
* on how a VR application can use a receiver implemented according to the profile and what information and interfaces are needed to enable the profile in a client
  + The VR application always requests the full video and provides it to the decoder. Only the rendering process does the viewport selection.
* on the quality and bitrate aspects for download cases for which a single ISO BMFF file is generated
  + Only a single video track is added to the file. The bitrate is expected to be in the range of 10-20 Mbit/s, depending on the content. For details see 3GPP TR26.918 ([S4-170752](http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_94/Docs/S4-170752.zip)).
* on the typical number of DASH Representations (and their bitrates) and Adaptation Sets in a DASH streaming environment.
  + A regular DASH client can be used and the offering is also following regular DASH offerings.
* How would the technology work if a secure media pipeline is required
  + A regular decryption module can be used.
  1. HEVC viewport dependent baseline profile
     1. General submission information

For the submission of a profile, the following information is required

* Please attach a full technical description of the proposal to the submission, based on the text of the DIS of OMAF. The profile should be limited to referencing tools of the DIS of OMAF and any other specifications that are either complete or expected to be published by the end of 2017.
  + The complete proposal is in clauses 10.1.3
* The submission addresses:
  + Media Profile for Video
* Summary of the proposal:
  + This media profile provides a viewpoint dependent solution based on HEVC tiles that allows for a higher resolution viewport than the viewport independent profile.
* Please provide information on
* Why a new profile is required?
  + This profile allows for a higher resolution than the viewport independent one; matching resolution of displays already available in the market.
* How is it significantly different from the existing profiles?
  + No profile exists yet
* If an existing profile can be adapted to accommodate the use case?
  + No profile exists yet
* Supporters for the Profile (Industry Backing and Interest for Deployment with preferably multiple companies, e.g. at least 3, interested in an interoperable end-to-end solution)
  + Fraunhofer HHI
  + Deutsche Telekom AG
  + Nokia
  + Samsung
  + Canon
  + Huawei
* Information by when test vectors and conformance software would be available (if the profiles is adopted) and if a reference receiver may be available as well.
  + The test vectors are provided with this proposal
  + Conformance software can be provided
  + The development of a reference receiver is currently not considered.
* The response to the requirements table is provided in clause 2.2.3.7
  + See below

For the submission of a profile, the following information is recommended

* Any information on existing implementations, e.g. demos
* Fraunhofer HHI (and others) has shown a demo on this profile

Any additional information may be provided such as how the performance can be improved with optional metadata included in OMAF.

* + 1. Addressing the requirements

Table B.2 provides the fulfilment matrix of the requirements for OMAF.

Table B.2 Requirements and Fulfilment Matrix (FF = y/es, n/o, p/atrial, d/oes not apply, o/ther – add comment)

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Requirement | FF | Comment |
| General | | | |
| 1 | The Specification shall provide for interoperable exchange of VR360 content | y | The profile is fully based on OMAF, ISO BMFF and DASH. |
| 2 | The Specification shall avoid providing multiple tools for the same functionality to reduce implementation burden and improve interoperability. | y | The profile is the first proposed viewport dependent profile based on tiles. |
| 3 | The Specification shall enable good quality and performance. | y | This profile allows for a higher viewport resolution than the HEVC viewport independent baseline profile. |
| 4 | The Specification shall enable full interoperability between services/content and clients. | y | The interoperability is provided by defining the media profile. This can be tested by checking the test vectors for tile-based streaming. |
| 4.1 | The Specification shall contain a very low number of fully specified interoperability points that include what is traditionally known as Profile and Level information. | y | There are only 2 proposed interoperability points including this one. |
| 4.1.1 | The existence of more than one interoperability point shall be justified if intended to target devices with different capabilities. | y | Devices exist that have display resolutions beyond the one that can be provided with a viewport independent profile based on HEVC Level 5.1 |
| 4.2 | Interoperability points shall address a Media Profile including:   * file format tracks and elementary stream * 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 | y | See clauses above. |
| 4.3 | 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 | P | This media profile can be part of a presentation profile. |
| 4.4 | These interoperability points shall enable conformance to be tested, inside and outside of MPEG. | y | There is a description and test vectors provided for this purpose. |
| 4.5 | The Specification may contain partial interoperability points (e.g., a file format box, a visual media profile) at a lower level of granularity, to enable external bodies to specify their own full interoperability points. | P | The media profile can be included in any presentation profile. |
| 4.6 | 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. | n/a |  |
| 4.7 | The specification shall define at least one media profile for audio. | n/a |  |
| 4.8 | The specification shall define at least one media profile for video | y | Fulfilled by the media profile |
| 4.9 | The specification shall define at least one presentation profile that includes one audio and one video media profile. | n/a |  |
| 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. | Y | Required decoding capabilities are HEVC Level 5.1 and the profile targets enabling higher resolutions available in the market and higher that might come soon. |
| 6 | The Specification shall support the representation, storage, delivery and rendering of   * Omnidirectional (up to 360° spherical) coded image/video (monoscopic and stereoscopic) with 3 DoF * Both 3D and 2D audio | P | The first aspect can be fulfilled as for the viewport independent profile. |
| 7 | The specification shall work with existing MPEG storage and delivery formats | Y | The media profile describes on how to carry the data in ISO BMFF and DASH. |
| 8 | The Specification shall support temporal synchronization and spatial alignment between different media types, in particular between audio and video. | P | Addressed by the usage of OMAF |
| 9 | The Specification shall support metadata for describing initial viewpoints and for the playback of omnidirectional video/image and audio according to that metadata. | P | The media profile does not prevent this information. |
| 10 | The Specification shall support the following interfaces:   * encoding and decoding for each media type * delivery for download and streaming | Y | Both interfaces are supported. |
| 11 | The Specification shall enable applications to use hardware-supported or pre-installed independently manufactured decoders and renderers through defined MPEG conformance points. | Y | HEVC Level 5.1 is broadly supported on existing HMDs. |
| 12 | The Specification shall support viewport-dependent processing (this may include delivery, decoding and rendering).   * The Specification shall support dynamically changing viewports * The Specification should enable responsiveness to changing viewport in a way that doesn’t detract from the immersive experience | Y | It is supported. |
| 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* | n/a | There is another profile proposed on this. |
| Delivery | | | |
| 14 | The Specification shall support the following methods of distribution:   * File-based delivery * DASH-based streaming * MMT-based streaming | P | This mapping to ISO BMFF and DASH is discussed in this contribution. |
| 15 | 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 |  | Template Errorr |
| Visual | | | |
| 16 | The Specification shall enable content exchange with high visual perceptual quality. | Y | With this profile and HEVC Level 5.1. we allow for a higher resolution than the viewport independent profile, providing a high visual quality. |
| 16.1 | 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). | y | A resolution matching the one of the display can be offered and still keep HEVC Level 5.1. |
| 16.2 | The Specification shall support a framerate of at least 60fps | Y | HEVC level 5.1. permits 4K at 60 fps. |
| 17 | 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. | y | Achieved by using tiles at different resolutions. |
| 18 | The Specification shall support metadata for the rendering of spherical video on a 2D screen | Y | Supported by OMAF |
| 19 | The Specification shall support fisheye-based video with a configuration of 2 cameras | n/a | Not addressed by the profile, but no requirement for each media profile. |
| 20 | The Specification shall support encoding of equirectangular projection (ERP) maps for monoscopic and stereoscopic video, in an efficient manner. | Y | Fully supported by the media profile. |
| 20.1 | Other projection maps than ERP for distribution should only be provided if consistent benefits over ERP is demonstrated | n/a | Not discussed now. However, If OMAF allowed further projections, the profile could consider usage of those. |
| Audio | | | |
| 21 | Each audio media profile in the Specification shall   1. support immersive rendering with sufficiently low latency   Note: this is also expressed in requirement 12   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. | n/a |  |
| 22 | There may be one audio media profile that supports only 2D audio to cater to existing devices  All other audio media profiles defined in the Specification shall:   * support 3D Audio distribution, decoding & rendering. * support immersive content, e.g. 12ch or 3rd order Ambisonics, * support a combination of diegetic and non-diegetic content sources. * be capable to ingest and carry all content types:   + audio channels,   + audio objects,   + scene-based audio,   + and combinations of the above. * be able to carry dynamic meta-data for combining, presenting and rendering all content types. | n/a |  |
| Security | | | |
| 23 | The Specification shall not preclude:   * Decoding and rendering to support secure media pipelines * Efficient distribution for multiple DRM systems (e.g. using common encryption´)   The Specification should enable a secure media pipeline to be implemented. | Y | This profile does not preclude any of those. |

* 1. OMAF 3D Audio Baseline Profile

MPEG-H 3D Audio is the latest audio codec developed by ISO/MPEG standardization group for efficient coding and rendering of high-quality Spatial Audio. MPEG-H 3D Audio carries all known popular audio representations (i.e., channel-based, object-based and scene-based) and offers high-quality reproduction to any output formats, thus providing an immersive experience. The wide range of embedded metadata types allows for personalization and dynamic interaction with the content. Based on the metadata, rendering technology of audio content is fully specified for both loudspeakers and headphones including interfaces to motion tracker data.

The Audio subgroup has concluded that the Low Complexity (LC) Profile of MPEG-H 3D Audio is the most suitable audio technology for offering a 360 audiovisual experience with 3DoF and it fulfils all requirements for Omnidirectional Media Format specified in section 3.4 of N16773 [1]. Consequently, this input contribution proposes the complete technical description of the corresponding Media Profile for MPEG-H Audio LC Profile described in N16826 [2], and include the Media Profile into the OMAF DIS text (as provided in N16824 [3]). Additionally, it is proposed to include the OMAF 3D Audio Baseline Media Profile into the OMAF Baseline Viewport-Independent Presentation Profile.

* + 1. Summary of the proposal:

The proposal provides complete technical description for:

* The OMAF 3D Audio Baseline Profile, including:
  + Elementary Stream Constraints
  + ISO BMFF Track Format Constraints
  + DASH Integration
  + CMAF Media Profile for OMAF 3D Audio Baseline Profile
* The OMAF Baseline Viewport-Independent Presentation Profile
* The response is based on the submission templates for OMAF profiles N16827 [4] and the requirements table as provided in clause 2.2 of N16773 [1]
  + 1. Additional Information
* Supporters for the Audio Media Profile:
  + Fraunhofer IIS,
  + Qualcomm Incorporated,
  + Samsung,
  + ETRI.
* The Audio subgroup has approved specification of the 3D Audio Baseline media profile based on MPEG-H Audio Low Complexity Profile, Level 3 with no additional tools.
* Fraunhofer IIS and Qualcomm Technologies, Inc. have demonstrated 3DOF VR consumption of MPEG-H 3D Audio LC profile encoded content on mobile phones at multiple trade shows (e.g., NAB 2016, IBC 2016, CES 2017, MWC 2017, NAB 2017). Demonstrations will also be available during the OMAF AhG.
* The proposed OMAF Media Profile is based on to the MPEG-H Audio CMAF Media Profile and it can be expected that the test vectors and conformance software will be made available in the same time with the CMAF test vectors and conformance software.
  + 1. Addressing the requirements

Table B.3 provides the fulfilment matrix. Please submit along with your submission.

Table B.3 Requirements and Fulfilment Matrix (FF = y/es, n/o, p/atrial, d/oes not apply, o/ther – add comment)

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Requirement | FF | Comment |
| General | | | |
| 1 | The Specification shall provide for interoperable exchange of VR360 content | y | The media profile is proposed for MPEG-H Audio LC Profile which specifies the ISO BMFF encapsulation. Based on CMAF and DASH, it provides maximum interoperability. |
| 2 | The Specification shall avoid providing multiple tools for the same functionality to reduce implementation burden and improve interoperability. | y | The media profile is based on MPEG-H Audio and specifies no additional tools for fulfilling all requirements and for offering high quality omnidirectional audio delivery. |
| 3 | The Specification shall enable good quality and performance. | y | MPEG-H Audio LC Profile provides excellent sound quality for 2D and 3D program material as shown in N16584 3D Audio Verification Test Report [5]. |
| 4 | The Specification shall enable full interoperability between services/content and clients. | y | Complete technical specification of download and streaming delivery for audio is provided, including requirements for the bitstream, and media profile signalling. |
| 4.1 | The Specification shall contain a very low number of fully specified interoperability points that include what is traditionally known as Profile and Level information. | y | The profile and level indication is provided for the proposed Media Profile:  - MPEG-H Audio LC Profile, Level 3. |
| 4.2 |  |  |  |
| 4.1.1 | The existence of more than one interoperability point shall be justified if intended to target devices with different capabilities. | d | This is the first media profile. |
| 4.2 | Interoperability points shall address a Media Profile including:   * file format tracks and elementary stream * 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 | p | The media profile specifies the requirements and constraints on file format and elementary stream for MPEG-H Audio LC Profile.  Specification of projections does not apply to audio. |
| 4.3 | 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 | p | The media profile may be added to a Baseline Presentation Profile. |
| 4.4 | These interoperability points shall enable conformance to be tested, inside and outside of MPEG. | y | The media profile specifies complete technical description which allows for conformance testing. |
| 4.5 | The Specification may contain partial interoperability points (e.g., a file format box, a visual media profile) at a lower level of granularity, to enable external bodies to specify their own full interoperability points. | d |  |
| 4.6 | 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. | d |  |
| 4.7 | The specification shall define at least one media profile for audio. | y | The complete technical description of the Media Profile for MPEG-H Audio LC Profile is proposed. |
| 4.8 | The specification shall define at least one media profile for video | d |  |
| 4.9 | The specification shall define at least one presentation profile that includes one audio and one video media profile. | y |  |
| 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. | y | See Section 1.2 above |
| 6 | The Specification shall support the representation, storage, delivery and rendering of   * Omnidirectional (up to 360° spherical) coded image/video (monoscopic and stereoscopic) with 3 DoF * Both 3D and 2D audio | y | MPEG-H Audio LC Profile supports both 2D and 3D Audio. |
| 7 | The specification shall work with existing MPEG storage and delivery formats | y | The MPEG-H Audio Stream is encapsulated into ISOBMFF. |
| 8 | The Specification shall support temporal synchronization and spatial alignment between different media types, in particular between audio and video. | y | The OMAF specification ensures A/V temporal alignment.  The MPEG-H 3D Audio decoder has a constant latency, see Table 1 of [3DA], thus audio and video portions of a media presentation can be synchronized. |
| 9 | The Specification shall support metadata for describing initial viewpoints and for the playback of omnidirectional video/image and audio according to that metadata. | p | The MPEG-H Audio LC Profile specifies sufficient metadata for describing initial viewpoints for audio based on the Audio Scene Information. The metadata enables omnidirectional playback of audio. |
| 10 | The Specification shall support the following interfaces:   * encoding and decoding for each media type * delivery for download and streaming | y | The encoding constraints and ISO BMFF Track Format Constraints are specified in sections 10.1.2.2 and 10.1.2.3.  The specification enables decoding according to MPEG-H Audio LC Profile Level 3. |
| 11 | The Specification shall enable applications to use hardware-supported or pre-installed independently manufactured decoders and renderers through defined MPEG conformance points. | y |  |
| 12 | The Specification shall support viewport-dependent processing (this may include delivery, decoding and rendering).   * The Specification shall support dynamically changing viewports * The Specification should enable responsiveness to changing viewport in a way that doesn’t detract from the immersive experience | y | The complete Audio Scene is decoded by the MPEG-H Audio LC decoder and provided to the renderer. The renderer adjusts the Audio Scene to any viewport. |
| 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* | d |  |
| Delivery | | | |
| 14 | The Specification shall support the following methods of distribution:   * File-based delivery * DASH-based streaming * MMT-based streaming | y | The media profiles specifies ISO BMFF encapsulation for MPEG-H Audio LC Profile. |
| 15 | 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 | d |  |
| Visual | | | |
| 16 | The Specification shall enable content exchange with high visual perceptual quality. | d |  |
| 16.1 | 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). | d |  |
| 16.2 | The Specification shall support a framerate of at least 60fps | d |  |
| 17 | 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. | d |  |
| 18 | The Specification shall support metadata for the rendering of spherical video on a 2D screen | d |  |
| 19 | The Specification shall support fisheye-based video with a configuration of 2 cameras | d |  |
| 20 | The Specification shall support encoding of equirectangular projection (ERP) maps for monoscopic and stereoscopic video, in an efficient manner. | d |  |
| 20.1 | Other projection maps than ERP for distribution should only be provided if consistent benefits over ERP is demonstrated | d |  |
| Audio | | | |
| 21 | Each audio media profile in the Specification shall   1. support immersive rendering with sufficiently low latency   Note: this is also expressed in requirement 12   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. | y | MPEG-H Audio LC Profile supports immersive and binaural rendering with sufficiently low latency.  MPEG-H Audio LC Profile provides excellent sound quality for 2D and 3D program material as shown in N16584 3D Audio Verification Test Report [5] |
| 22 | There may be one audio media profile that supports only 2D audio to cater to existing devices  All other audio media profiles defined in the Specification shall:   * support 3D Audio distribution, decoding & rendering. * support immersive content, e.g. 12ch or 3rd order Ambisonics, * support a combination of diegetic and non-diegetic content sources. * be capable to ingest and carry all content types:   + audio channels,   + audio objects,   + scene-based audio,   + and combinations of the above. * be able to carry dynamic meta-data for combining, presenting and rendering all content types. | y | MPEG-H Audio LC Profile supports:   * 2D and 3D Audio distribution, decoding & rendering. * immersive content * a combination of diegetic and non-diegetic content sources. * ingestion and carriage of all content types:   + audio channels,   + audio objects,   + scene-based audio (FOA and HOA),   + and combinations of the above.   MPEG-H Audio LC Profile specifies static and dynamic meta-data which allows for combining, presenting and rendering of all content types. |
| Security | | | |
| 23 | The Specification shall not preclude:   * Decoding and rendering to support secure media pipelines * Efficient distribution for multiple DRM systems (e.g. using common encryption´)   The Specification should enable a secure media pipeline to be implemented. | d |  |

* 1. OMAF 2D Audio Legacy Profile

In 1997, AAC was first introduced as MPEG-2 Part 7, i.e., ISO/IEC 13818-7:1997. The AAC Profile (AAC-LC) and HE‑AAC Profile (AAC-LC with SBR) were first standardized in ISO/IEC 14496-3:2001/Amd 1:2003 [7] while the HE‑AACv2 Profile (HE‑AAC with PS) was standardized in ISO/IEC 14496-3:2001/Amd 2:2004. Since then, MPEG-4 HE-AAC (High Efficiency Advanced Audio Coding) has become one of the most widely deployed and important enabling technologies for media delivery.

AAC and HE-AAC have been adopted by application standards like 3GPP, AES, ARIB, ATSC, SCTE, DLNA, DVB, DASH-IF, EBU, GSMA, HbbTV, HDMI, IEC, IMDA, WiFi Alliance, and WorldDMB. Furthermore, operating systems and browsers with support for HE-AAC include iOS, Android, Windows 7/8/10, Mac OS, IE9, IE10, Safari, Chrome.

Also, HE-AAC is used worldwide in the most successful streaming services and supported by all major streaming and media platforms. HE-AAC-powered streaming services include Netflix, YouTube, BBC iPlayer, Hulu, Amazon video, Pandora, Google Play, China Mobile, KDDI and many more.

Due to the wide reach, (HE-)AAC has been chosen for VR services and platforms. There, either stereo, quad or 5.1 surround channel configurations are used. For example, AAC is used currently in these configurations by the Samsung VR platform and Hulu VR.

Therefore, this input contribution proposes a complete technical description of the corresponding Media Profile, the OMAF 2D Audio Legacy Profile specified in N16826.

* + 1. Summary of the proposal:

The proposal provides complete technical description for:

* The OMAF 2D Audio Legacy Profile, including:
  + Elementary Stream Constraints
  + ISO BMFF Track Format Constraints
  + DASH Integration
* The response is based on the submission templates for OMAF profiles N16827 [4] and the requirements table as provided in clause 2.2 of N16773 [1].
  + 1. Additional Information
* Supporters for the Audio Media Profile:
  + Fraunhofer IIS,
  + Qualcomm Incorporated,
  + Samsung.
* The Audio subgroup has approved specification of the 2D Audio Legacy media profile based on MPEG-4 AAC, Level 4 with no additional tools.
* Fraunhofer IIS have demonstrated VR consumption of AAC encoded 2D content on mobile phones at multiple trade shows (e.g., NAB 2016, IBC 2016, CES 2017, MWC 2017, NAB 2017). Demonstrations will also be available during the MPEG 119th meeting.
* The proposed OMAF Media Profile is based on to the AAC Core CMAF Media Profile, N16819 [10], and it can be expected that the test vectors and conformance software will be made available in the same time with the CMAF test vectors and conformance software.
  + 1. Quality aspects

The quality of AAC and HE-AAC(v2) in stereo and 5.1 configuration is well proven and tested by MPEG and by several application standards.

Within MPEG, HE-AAC has been tested in mono and stereo configuration in the USAC Verification test. There, it was shown, that HE-AACv2 provides good quality for 24kbps mono. For stereo configuration, HE-AACv2 provides good at 64 kbps and excellent quality at 96 kbps, as shown in N12232 [8]. Note that the HE-AACv2 profile includes HE-AAC and AAC-LC.

The European Broadcasting Union conducted a test on multi-channel audio codecs in 2007 [9]: “It can be concluded that, at the moment, the MPEG HE-AAC seems to be the most favourable choice for a broadcaster requiring a good scalability of bitrate versus quality, down to relatively low bit rates. In addition, the AAC-based codec family offers excellent audio quality at higher bitrates, e.g. at 320 kbit/s (with the exception of "applause"). Our study shows that excellent quality (on average) can be achieved even at half the bitrate, i.e. 160 kbit/s, or even less, for all test items except for the most critical items.”

* + 1. Addressing the requirements

Table B.4 provides the fulfilment matrix. Please submit along with your submission.

Table B.4 Requirements and Fulfilment Matrix (FF = y/es, n/o, p/atrial, d/oes not apply, o/ther – add comment)

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Requirement | FF | Comment |
| General | | | |
| 1 | The Specification shall provide for interoperable exchange of VR360 content | y | The media profile is proposed for MPEG-4 AAC which specifies the ISO BMFF encapsulation, with no additional signaling on file format level.  Based on CMAF and DASH, it provides maximum interoperability. |
| 2 | The Specification shall avoid providing multiple tools for the same functionality to reduce implementation burden and improve interoperability. | y | The media profile is based on MPEG-4 AAC and specifies no additional tools for fulfilling all requirements for the 2D OMAF media profile. |
| 3 | The Specification shall enable good quality and performance. | y | MPEG-4 AAC provides excellent sound quality for 2D program material as shown in [8][9] |
| 4 | The Specification shall enable full interoperability between services/content and clients. | y | Complete technical specification of download and streaming delivery for audio is provided, including requirements for the bitstream, and media profile signaling. |
| 4.1 | The Specification shall contain a very low number of fully specified interoperability points that include what is traditionally known as Profile and Level information. | y | The profile and level indication is provided for the proposed Media Profiles: AAC-LC, HE-AAC, HE-AACv2, Level 4. |
| 4.2 | The existence of more than one interoperability point shall be justified if intended to target devices with different capabilities. | d | Only one 2D media profile was proposed and approved by the Audio subgroup. |
| 4.3 | Interoperability points shall address a Media Profile including:   * file format tracks and elementary stream * 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 | p | The media profile specifies the requirements and constraints on file format and elementary stream.  Specification of projections does not apply to audio. |
| 4.4 | 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 | d | Issues pertaining to synchronization of audio signal are fully described in N8837 [11], Technical Report on Audio and Systems Interaction |
| 4.5 | These interoperability points shall enable conformance to be tested, inside and outside of MPEG. | y | The media profile specifies complete technical description which allows for conformance testing. |
| 4.6 | The Specification may contain partial interoperability points (e.g., a file format box, a visual media profile) at a lower level of granularity, to enable external bodies to specify their own full interoperability points. | d |  |
| 4.7 | 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. | d |  |
| 4.8 | The specification shall define at least one media profile for audio. | y | The complete technical description of the Media Profile is proposed. |
| 4.9 | The specification shall define at least one media profile for video | d |  |
| 4.10 | The specification shall define at least one presentation profile that includes one audio and one video media profile. | y |  |
| 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. | y | The devices run operating systems that natively support MPEG-4 AAC (up to HE-AACv2 profile) in mono, stereo and 5.1 configurations as specified in this proposed media profile. |
| 6 | The Specification shall support the representation, storage, delivery and rendering of   * Omnidirectional (up to 360° spherical) coded image/video (monoscopic and stereoscopic) with 3 DoF * Both 3D and 2D audio | p | MPEG-4 AAC supports the delivery of 2D audio formats, only. |
| 7 | The specification shall work with existing MPEG storage and delivery formats | y | The media profile specifies encapsulation into ISOBMFF. |
| 8 | The Specification shall support temporal synchronization and spatial alignment between different media types, in particular between audio and video. | y | The OMAF specification ensures A/V temporal alignment. |
| 9 | The Specification shall support metadata for describing initial viewpoints and for the playback of omnidirectional video/image and audio according to that metadata. | d |  |
| 10 | The Specification shall support the following interfaces:   * encoding and decoding for each media type * delivery for download and streaming | y | The encoding constraints and ISO BMFF Track Format Constraints are specified. |
| 11 | The Specification shall enable applications to use hardware-supported or pre-installed independently manufactured decoders and renderers through defined MPEG conformance points. | y | The output of hardware-supported or pre-installed decoders are stereo or 5.1 audio signals. Conformance for these is specified in 14496-26:2010, MPEG-4 Audio Conformance |
| 12 | The Specification shall support viewport-dependent processing (this may include delivery, decoding and rendering).   * The Specification shall support dynamically changing viewports * The Specification should enable responsiveness to changing viewport in a way that doesn’t detract from the immersive experience | y | The output of the audio decoder can be rendered viewport-dependent by using existing device-dependent binaural rendering technology. The binaural rendering method is out of scope for this profile. |
| 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* | d |  |
| Delivery | | | |
| 14 | The Specification shall support the following methods of distribution:   * File-based delivery * DASH-based streaming * MMT-based streaming | y | The media profiles specifies ISO BMFF encapsulation. |
| 15 | 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 | d |  |
| Visual | | | |
| 16 | The Specification shall enable content exchange with high visual perceptual quality. | d |  |
| 16.1 | 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). | d |  |
| 16.2 | The Specification shall support a framerate of at least 60fps | d |  |
| 17 | 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. | d |  |
| 18 | The Specification shall support metadata for the rendering of spherical video on a 2D screen | d |  |
| 19 | The Specification shall support fisheye-based video with a configuration of 2 cameras | d |  |
| 20 | The Specification shall support encoding of equirectangular projection (ERP) maps for monoscopic and stereoscopic video, in an efficient manner. | d |  |
| 20.1 | Other projection maps than ERP for distribution should only be provided if consistent benefits over ERP is demonstrated | d |  |
| Audio | | | |
| 21 | Each audio media profile in the Specification shall   1. support immersive rendering with sufficiently low latency   Note: this is also expressed in requirement 12   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. | y | The 2D media profile  representation enables support for immersive binaural rendering with sufficiently low latency of 2D channel-based audio content up to 5.1.  MPEG-4 AAC provides excellent sound quality for 2D program material as shown in tests according to ITU-R BS.1534 in [8][9]  The output of the audio decoder can be rendered viewport-dependent by using existing device-dependent binaural rendering technologies, although binaural rendering technology is out of scope for this profile. |
| 22 | There may be one audio media profile that supports only 2D audio to cater to existing devices  All other audio media profiles defined in the Specification shall:   * support 3D Audio distribution, decoding & rendering. * support immersive content, e.g. 12ch or 3rd order Ambisonics, * support a combination of diegetic and non-diegetic content sources. * be capable to ingest and carry all content types:   + audio channels,   + audio objects,   + scene-based audio,   + and combinations of the above. * be able to carry dynamic meta-data for combining, presenting and rendering all content types. | y | The OMAF 2D Audio Legacy Media Profile is limited to 2D channel-based audio formats (mono, stereo, 5.1) and does not fulfill the requirements specified for all other media profiles |
| Security | | | |
| 23 | The Specification shall not preclude:   * Decoding and rendering to support secure media pipelines * Efficient distribution for multiple DRM systems (e.g. using common encryption´)   The Specification should enable a secure media pipeline to be implemented. | d |  |

* 1. AVC viewport dependent media profile
     1. General submission information
* Full technical description of the proposal to the submission:
* The complete proposal is in clauses A.1.
* The submission addresses:
* Media Profile for Video
* Summary of the proposal:
* This media profile provides a viewport dependent solution based AVC with region-wise packing that allows for a higher quality/resolution viewport than the viewport independent profile.

Please provide information on

* Why a new profile is required?
  + This profile allows for a higher resolution than the viewport independent one; matching resolution of displays already available in the market.
* How is it significantly different from the existing profiles?
  + No profile exists yet
* If an existing profile can be adapted to accommodate the use case?
  + No profile exists yet
* Supporters for the Profile (Industry Backing and Interest for Deployment with preferably multiple companies, e.g. at least 3, interested in an interoperable end-to-end solution)
  + Fraunhofer HHI
  + Deutsche Telekom AG
  + Nokia
  + Samsung
  + Canon
  + Huawei
* Information by when test vectors and conformance software would be available (if the profiles is adopted) and if a reference receiver may be available as well.
  + To be provided
* The response to the requirements table has been provided.
  + To be provided

For the submission of a profile, the following information is recommended

* Any information on existing implementations, e.g. demos
  1. Viewport-Independent Fisheye Video Profile
     1. General submission information

For the submission of a profile, the following information is required

* Please attach a full technical description of the proposal to the submission, based on the text of the DIS of OMAF (as provided in N16824[3]). The profile should be limited to referencing tools of the DIS of OMAF and any other specifications that are either complete or expected to be published by the end of 2017.
  + The complete proposal is in clause A.2
* Please indicate if the submission addresses
  + Media Profile for Video
* Please provide a summary of the proposal
  + This media profile fulfils basic requirements to support omnidirectional video via multiple circular images captures by fisheye cameras. The profile does neither require viewport dependent decoding nor viewpoint dependent delivery Regular DASH clients, file format parsers and HEVC decoders engines can be used for distribution and decoding. The profile also minimizes the options for basic interoperability.
* Please provide information on
* Why a new profile is required?
  + This is an initial profile in order to serve basic interoperability for fisheye vides using existing decoders and distribution infrastructures.
* How is it significantly different from the existing profiles?
  + No profile exists yet
* If an existing profile can be adapted to accommodate the use case?
  + No profile exists yet
* Supporters for the Profile (Industry Backing and Interest for Deployment with preferably multiple companies, e.g. at least 3, interested in an interoperable end-to-end solution)
  + Samsung
* Information by when test vectors and conformance software would be available (if the profiles is adopted) and if a reference receiver may be available as well.
  + The development of test content is currently considered for Sep 2017.
  + The development of a reference receiver is currently not considered.
* The response to the requirements table as provided in clause 0.
  + See below

For the submission of a profile, the following information is recommended

* Any information on existing implementations, e.g. demos
  + Implementations exist based on proprietary services, details to be added
* How the profile addresses certain use cases as documented in clause 0.
  + See below

Any additional information may be provided such as how the performance can be improved with optional metadata included in OMAF.

* + 1. Addressing the requirements

Table B.5 provides the fulfilment matrix. Please submit along with your submission.

TableB.5 Requirements and Fulfilment Matrix (FF = y/es, n/o, p/atrial, d/oes not apply, o/ther – add comment)

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Requirement | FF | Comment |
| General | | | |
| 1 | The Specification shall provide for interoperable exchange of VR360 content | y | The profile is fully based on OMAF, ISO BMFF, CMAF and DASH and provides minimum options to fulfill interoperability. |
| 2 | The Specification shall avoid providing multiple tools for the same functionality to reduce implementation burden and improve interoperability. | y | The profile has minimum options and therefore fulfills the basic requirements for interoperability. |
| 3 | The Specification shall enable good quality and performance. | y | HEVC level 5.1 supports up to 4k by 2k at reasonable frame rates and therefore provides satisfactory quality at maximum simplicity. |
| 4 | The Specification shall enable full interoperability between services/content and clients. | y | The specification provides a full specification for download and streaming delivery for video. The interoperability is provided by the strict definition of the media profile requirements for the bitstream, the receiver and the signalling of the media profile. |
| 4.1 | The Specification shall contain a very low number of fully specified interoperability points that include what is traditionally known as Profile and Level information. | y | This is the first media profile |
| 4.2 |  |  |  |
| 4.1.1 | The existence of more than one interoperability point shall be justified if intended to target devices with different capabilities. | y | This is the first media profile |
| 4.2 | Interoperability points shall address a Media Profile including:   * file format tracks and elementary stream * 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 | y | All issues are specified in the specification above in the relevant clauses |
| 4.3 | 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 | P | This addresses a media profile which may be added to a presentation profile. |
| 4.4 | These interoperability points shall enable conformance to be tested, inside and outside of MPEG. | y | The description of the requirements are clearly documented in the profiles. |
| 4.5 | The Specification may contain partial interoperability points (e.g., a file format box, a visual media profile) at a lower level of granularity, to enable external bodies to specify their own full interoperability points. | P | The media profile may be used in a full presentation profile as defined in 4.3. |
| 4.6 | 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. | n/a |  |
| 4.7 | The specification shall define at least one media profile for audio. | n/a |  |
| 4.8 | The specification shall define at least one media profile for video | y | Fulfilled by the media profile |
| 4.9 | The specification shall define at least one presentation profile that includes one audio and one video media profile. | n/a |  |
| 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. | Y | The supported features and options are supported by the referenced devices. |
| 6 | The Specification shall support the representation, storage, delivery and rendering of   * Omnidirectional (up to 360° spherical) coded image/video (monoscopic and stereoscopic) with 3 DoF * Both 3D and 2D audio | P | The media profile addresses the first aspect |
| 7 | The specification shall work with existing MPEG storage and delivery formats | Y | The media profile describes on how to carry the data in ISO BMFF, CMAF and DASH. Existing parsers, CMAF Players and DASH clients can be used as long as the rendering pipeline supports SEI messages documented in clause 2.2.2. |
| 8 | The Specification shall support temporal synchronization and spatial alignment between different media types, in particular between audio and video. | P | Addressed by the usage of OMAF |
| 9 | The Specification shall support metadata for describing initial viewpoints and for the playback of omnidirectional video/image and audio according to that metadata. | P | The media profile does not prevent this information. |
| 10 | The Specification shall support the following interfaces:   * encoding and decoding for each media type * delivery for download and streaming | Y | Clause 2.2.2 and 2.2.3 and 2.2.4 address the aspects |
| 11 | The Specification shall enable applications to use hardware-supported or pre-installed independently manufactured decoders and renderers through defined MPEG conformance points. | Y | HEVC level 5.1 is broadly supported on existing HMDs as indicated by the LS from VR-IF |
| 12 | The Specification shall support viewport-dependent processing (this may include delivery, decoding and rendering).   * The Specification shall support dynamically changing viewports * The Specification should enable responsiveness to changing viewport in a way that doesn’t detract from the immersive experience | Y | In all cases the full 360 scene is decoded and is made available to the renderer. The renderer will be able to adjust the viewport without involving the DASH client or the network. |
| 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* | Y | This media profile address this issue for video. |
| Delivery | | | |
| 14 | The Specification shall support the following methods of distribution:   * File-based delivery * DASH-based streaming * MMT-based streaming | P | This mapping to ISO BMFF and DASH streaming is provided for this media profile. |
| 15 | 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 |  | Template Errorr |
| Visual | | | |
| 16 | The Specification shall enable content exchange with high visual perceptual quality. | Y | HEVC level 5.1. permits decent quality for 360 video. |
| 16.1 | 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). | P |  |
| 16.2 | The Specification shall support a framerate of at least 60fps | Y | HEVC level 5.1. permits 4K at 60 fps. |
| 17 | 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. | P | The content would need to be downsampled to be decodable by an HEVC level 5.1 decoder. The subsampling may be done in spatial or temporal domain or by reducing the coverage. |
| 18 | The Specification shall support metadata for the rendering of spherical video on a 2D screen | Y | Supported by OMAF |
| 19 | The Specification shall support fisheye-based video with a configuration of 2 cameras | Y | Fully supported by the media profile. |
| 20 | The Specification shall support encoding of equirectangular projection (ERP) maps for monoscopic and stereoscopic video, in an efficient manner. | n/a | This media profile deals with fisheye video. |
| 20.1 | Other projection maps than ERP for distribution should only be provided if consistent benefits over ERP is demonstrated | n/a | This media profile deals with fisheye video. |
| Audio | | | |
| 21 | Each audio media profile in the Specification shall   1. support immersive rendering with sufficiently low latency   Note: this is also expressed in requirement 12   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. | n/a |  |
| 22 | There may be one audio media profile that supports only 2D audio to cater to existing devices  All other audio media profiles defined in the Specification shall:   * support 3D Audio distribution, decoding & rendering. * support immersive content, e.g. 12ch or 3rd order Ambisonics, * support a combination of diegetic and non-diegetic content sources. * be capable to ingest and carry all content types:   + audio channels,   + audio objects,   + scene-based audio,   + and combinations of the above. * be able to carry dynamic meta-data for combining, presenting and rendering all content types. | n/a |  |
| Security | | | |
| 23 | The Specification shall not preclude:   * Decoding and rendering to support secure media pipelines * Efficient distribution for multiple DRM systems (e.g. using common encryption´)   The Specification should enable a secure media pipeline to be implemented. | Y | The use of a linear decoding flow w/o requiring to do complex pixel-based modification permits to use regular common encryption.  Also the usage of a single video stream provides the ability to apply regular common encryption.  Compatibility to HEVC CMAF ensure broad support for encrypted content as well. |

* + 1. Response to certain Scenario

<This section needs to be updated.>

Consider the following scenario. A content provider wants to provide a 360VR service with 3D audio to mobile HMD with head motion tracking. The content enables to change the field-of-view based on user interaction.

The HMD support HEVC Main-10 level 5.1 video decoder and an MPEG audio decoder. The original content is

* Basic VR content: as low as 4k x 2k (ERP), 8 or 10bit, BT.709, as low as 30fps, monoscopic and stereoscopic
* High-quality: up to 8k x 4k (ERP), 10 bit, possibly advanced transfer characteristics and colour transforms, sufficiently high frame rates, etc.

# References

1. MPEG N16773, Requirements for Omnidirectional Media Format
2. MPEG N16826, Profiles under Considerations in OMAF (ISO/IEC 23090-2)
3. MPEG N16824, ISO/IEC DIS 23090-2 Omnidirectional MediA Format
4. MPEG N16827, Proposed Submission Template for Profiles in OMAF (ISO/IEC 23090-2)
5. MPEG N16584, 3D Audio Verification Test Report
6. MPEG N16821, Text on ISO/IEC 23000-19 DAM 1 SHVC media profile and additional audio media profiles
7. MPEG N5570, Text of ISO/IEC 14496-3:2001/FDAM1, Bandwidth Extension.
8. MPEG N12232, USAC Verification Test Report
9. EBU evaluation of multi-channel audio codecs; EBU tech 3324; EBU, Geneva, Sept 2007
10. MPEG N16819 Text of ISO/IEC FDIS 23000-19 Common Media Application Format
11. MPEG N8837, ISO/IEC 14496-3/FDAM 7, Technical Report on Audio and Systems Interaction, 79th meeting
12. MPEG N16784, Technologies under Consideration for ISOBMFF