

**INTERNATIONAL ORGANISATION FOR STANDARDISATION  
ORGANISATION INTERNATIONALE DE NORMALISATION  
ISO/IEC JTC1/SC29/WG11  
CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2016/M39934  
January 2017, Geneva, CH**

**Source**    **Telecom ParisTech, Canon Research Centre France**  
**Status**    **For consideration during MPEG #117**  
**Title**      **Usage of 'hvt1' in DASH**  
**Author**    Cyril Concolato, Jean Le Feuvre, Frédéric Mazé, Franck Denoual, Nael Ouedraogo

## **1 Introduction**

This contribution discusses elements present in the DASH TuC N16465 [1]. Based on the TuC content, it proposes to add a normative annex to the DASH specification to specify how content using HEVC tile tracks, using the 'hvt1' code point, shall be described in an MPD.

## **2 Discussion**

### **2.1 Considered use cases**

Section 4 of the TuC document discusses different use cases for HEVC tile tracks as defined in the HEVC File Format [2], in MPEG-DASH.

This contribution concentrates only on the use cases that consider adaptive streaming of HEVC videos based on tile tracks. In these use cases, a player may decide to play all tiles, possibly with different qualities, or it may also decide to download and play only some tiles (possibly just one). We consider that these use cases are the most relevant. The use case of building a complete bitstream from a single tile track (using extractors) is not considered in this contribution, since it can already be achieved with standard DASH dependencies description.

In the considered use cases, the aggregation of the downloaded media segments may form a complete bitstream (i.e. that there are NALU for all the tiles); or the downloaded media segments may only contain data for some tiles (of interest) and not all, if the DASH client decides to fetch only a subset of the content. It should therefore be pointed out that:

- The reconstructed MP4 file (i.e. the one resulting from the aggregation of the tile media segments and the init segment, following the DASH dependency attributes) is conformant to the HEVCFF.
- The HEVC elementary stream that can be extracted from the MP4 file may not be complete, and therefore may not respect the HRD model, but is conformant. Additionally, it will be treated correctly by proper HEVC decoders (e.g. decoders that handle packet loss: a 'missing' tile is a special case of a 'loss', simpler to handle than general NALU loss).

From the TuC, it can be seen that there are several options to author the MPD (even for the considered use case). We propose to restrict the options, without loss of features.

## **2.2 About initialization segments**

The main choice concerns the content of the initialization segment. In the current TuC, there are the following options:

- a. either all representations use the same initialization segment, and therefore combining some or all representations is simple since the (single) initialization segment can be used;
- b. or some representations (possibly all) use a different initialization segment. For example, each tile representation may have an initialization segment declaring only the tile track and its base track. Combining two tile representations therefore requires either:
  - mixing initialization segments (which can be tricky)
  - or providing an additional representation with the corresponding initialization segment. This latter case would require as many representations as there are possible tile combinations, which may not be reasonable for large tiling grids and several bitrates.

Since there is no loss of functionality, and since this can improve interoperability, we propose to restrict the possible choices of DASH to the option a).

## **2.3 About SRD**

When using HEVC tile tracks in DASH, a representation pointing to the base tile track only could be used. Although that associated base tile track is fully conformant to ISOBMFF, it does not lead to a meaningful rendering on its own. We propose to mandate that when such representation is used, it is signaled as non-meaningful.

### 3 Proposed Text

#### **Annex XXXX** (normative) Use of HEVC Tile Tracks

#### **XXX.1 Introduction**

ISO/IEC 14496-15 defines the carriage of videos encoded according to ISO/IEC 23008-2, in particular using tiles. This annex specifies how some of the features related to tiles are declared in MPD.

We define the following terms:

- A **Tile Representation** is a Representation whose segments are conformant to ISO/IEC 14496-15 and whose media segments contain only media samples for one or more tile tracks, as defined in ISO/IEC 14496-15, i.e. whose type is "hvt1".
- A **Tile Base Representation** is a Representation whose segments are conformant to ISO/IEC 14496-15 and whose media segments contain only media samples for the base tile track, i.e. whose type is "hev2" or "hvc2".

NOTE 1: A Tile Representation is dependent on a Tile Base Representation because ISOBMFF tile tracks depend on ISOBMFF base tile track for processing.

NOTE 2: An MPD MAY offer all tile tracks in one single Tile Representation, or each tile track in its own single Tile Representation, or a combination of tile tracks in a single Tile Representation.

NOTE 3: The above definitions apply to HEVC tile tracks (i.e. of type "hvt1") and to the associated tile base track. ISO/IEC 14496-15 allows the carriage of tile content in other types of tracks. MPD authors may decide to provide Representations using these other track types. This Annex does not constraint these Representations.

#### **XXX.2 Constraints on initialization segments**

In an MPD, all Tile Representations depending on the same Tile Base Representation SHALL have the same initialization segment (i.e. declaring all tile tracks and the tile base track).

NOTE1: The media segments of those Representations will typically contain only media samples from a set of tile tracks.

For a given MPD time, the concatenation of a Tile Base Representation initialization segment, followed by the Tile Base Representation media segment for that time and the media segments for that time of one or more Tile Representations but identical initialization segment shall be a conforming HEVC file as defined in ISO/IEC 14496-15.

NOTE2: Given that all Representations use the same initialization segment and that this segment declares the tile base track and all tile tracks, the above rule makes it possible for a DASH client to select several Tile Representations and to combine them.

NOTE3: It is possible to offer different Tile Base Representations and different Tile Representations associated to these Tile Base Representations, for example to provide multiple resolutions. In this case, reconstructing a single complete HEVC bit stream is made by selecting all Tile Representations sharing the same initialization segment.

### **XXX.3 Constraints on Spatial Relationship Descriptions**

Spatial Relationship Descriptions shall be used in Tile Representations or Tile Base Representations to indicate to a DASH client how the different Representations (and containing Adaptation Sets) are combined in the source content.

SRD descriptors shall be used and the Tile Base Representation and its dependent Tile Representations shall have the same `source_id` parameter.

The Tile Base Representation shall have its width and height set to the values of the tile base track, i.e. the width and height of the full media.

SRD descriptors shall be used on Tile Base Representations as follows:

- an `EssentialProperty` SRD descriptor shall be used,
- this SRD's `object_x`, `object_y`, `object_width` and `object_height` shall be set to 0,

NOTE: The use of 0 for `object_width` and `object_height` is meant to signal to the DASH client that this Representation alone does not lead to a meaningful presentation.

- this SRD shall contain the optional `total_width` and `total_height` parameters, specified as the values of the full video width and height, possibly in arbitrary units.

SRD descriptors shall be used on Tile Representations as follows:

- a `SupplementalProperty` SRD descriptor should be used,
- this SRD's `object_x`, `object_y`, `object_width` and `object_height` shall be set to the horizontal position, vertical position, width and height values of the tile in the reconstructed picture, in the same units as the Tile Base Representation SRD descriptor.

## **4 Example of MPD**

An example MPD for a 2x1 tiling is as follows. It describes a Tile Base Representation corresponding to a 1280x640 full-resolution video with two Tile AdaptationSets, each proposing two Tile Representations with different bitrates.

```

<MPD>
<Period >
  <AdaptationSet maxWidth="1280" maxHeight="640" >
    <EssentialProperty schemeIdUri="urn:mpeg:dash:srd:2014" value="1,0,0,0,0,1280,640"/>
    <SegmentTemplate initialization="v_base.mp4" ... />
    <Representation id="1"
      mimeType="video/mp4"
      codecs="hev2.1.1.6.L186.0"
      width="1280" height="640"/>
  </AdaptationSet>
  <AdaptationSet maxWidth="640" maxHeight="640" ...>
    <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014" value="1,0,0,640,640"/>
    <SegmentTemplate initialization="v_base.mp4" ... />
    <Representation id="1_1"
      mimeType="video/mp4"
      codecs="hvt1.1.1.6.L186.0"
      dependencyId="1"
      bandwidth="128000"/>
    <Representation id="1_2"
      mimeType="video/mp4"
      codecs="hvt1.1.1.6.L186.0"
      dependencyId="1"
      bandwidth="768000"/>
  </AdaptationSet>
  <AdaptationSet maxWidth="640" maxHeight="640" ...>
    <SupplementalProperty schemeIdUri="urn:mpeg:dash:srd:2014" value="1,640,0,640,640"/>
    <SegmentTemplate initialization="v_base.mp4" ... />
    <Representation id="2_1"
      mimeType="video/mp4"
      codecs="hvt1.1.1.6.L186.0"
      dependencyId="1"
      bandwidth="128000"/>
    <Representation id="2_2"
      mimeType="video/mp4"
      codecs="hvt1.1.1.6.L186.0"
      dependencyId="1"
      bandwidth="768000"/>
  </AdaptationSet>
</Period>
</MPD>

```

## 5 Conclusion

We recommend MPEG to include the proposed text as a normative annex to MPEG DASH.

## 6 References

- [1] w16465 “Technologies under Consideration for Dynamic Adaptive Streaming over HTTP 23009, parts 1, 3, 4, 5 and 6” by T. Stockhammer, MPEG#116, Chengdu, October 2016
- [2] w16169 “Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format”, MPEG#115, Geneva, June 2016