**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N17875**

**July 2018, Ljubljana, SI**

|  |  |
| --- | --- |
| **Source** | **System SubGroup** |
| **Status** | **Approved** |
| **Title** | **Thoughts on Adaptive Delivery and Access to Immersive Media** |
| **Editors** | Emmanuel Thomas (TNO), Thomas Stockhammer (Qualcomm) |

# Preamble

The present document captures thoughts on the adaptive delivery and access for immersive media. The content of the document is the result of discussions in the System Subgroup during MPEG 123 triggered by the input contribution m43753, "Streaming-First design for the MPEG-I project" [1] .

Adaptive Delivery and Access   
to Immersive Media

New, immersive services call for new ways of organizing, accessing, delivering and consuming media data. As a result, MPEG should assess what the implications are for its coding and encapsulation formats.

This document seeks to communicate the challenge ahead, and refrains from phrasing solutions.

# Background

MPEG-encoded media historically retrieved from storage media (e.g. CD-I, DVD) or delivered using broadcast (e.g. ARIB, ATSC, DVB). Nowadays streaming has become the dominant force in the video industry; video streaming services drive the technical innovation in the media industry. MPEG Technologies (ISO BMFF, DASH) find massive adoptionfor unicast for live and on-demand services augmenting, complementing and replacing broadcast (ATSC, DVB, HbbTV, SCTE). Immersive experiences (VR/360, 3DoF+, etc.) favor unicast-based technologies such as DASH to tailor the data stream to the exact needs of the consuming user in real time.

# Challenging New Applications

The following are example of application bringing new challenges for streaming and accessing MPEG coded data:

* **Tiled 360 videos** in very high resolution
* Large **Point** **Clouds** that can be navigated in 6 DoF
* **Light fields** with lots and lots of small tiles
* A complicated **Scene** **Graph** with many objects to traverse
* Audio objects can be audible, or beyond the “**audio** **horizon**” in an immersive experience
* All likely retrieved from some sort of **cloud infrastructure**
* All of these can be available in **multiple quality/bitrate** variations
* All of those need to be **decoded and decrypted** with constrained devices at the receiver side

# Trends

Significant parts of the media data are **unique** **to the receiver**. Delivery shifts from sender-driven to **receiver-driven**. In addition, application requirements change dynamically in real-time, which makes **latency** a crucial aspect. Not all the components of an application have the same requirement in terms of end-to-end latency. Total media data is increasing, and *therefore* the media data requires more **fine grain access**. Since mobile and wearable devices are core of the consumption of immersive media, the usage of **general purpose receiving platforms** (decoders, hardware decryptions, protocol stacks, renderers) should be leveraged to offer energy-efficient consumption.

# Technical Challenges Ahead

Since the representation of the immersive experience constitutes a large amount of data (several TBs expected), it is desirable to allow client to flexibly retrieve **parts** of a large body of media data from a **cloud** resource to create a coherent user experience under **constrained resources**, where:

* **constraints** exist like bandwidth, access latency, decode resources (and where these can fluctuate dynamically)
* the client in charge of making **trade-offs** given such constraints
* fast **response** **times** and **efficiency** are crucial for the QoE
* inherently, data is accessed and retrieved in multiple **parallel** **streams**
* this data may need to be **protected** and/or encrypted
* this data may need to be **cached** close to the user for the best experience
* the data is stored in the **cloud** in a distributed manner

# Organization of Data

Traditionally, data has been organized to allow temporal access. This dimension will remain but will merely be one of the dimensions:

* **Temporal** random access – “as usual”
  + Different timelines
  + Addition of non-timed media
* **Spatial** random access – retrieving only the relevant parts of the media
  + Depending on user orientation
* **Quality** access – retrieving the suitable quality
  + Making quality/bitrate trade-offs in switching between quality levels
  + Depending on what is visible/audible
  + Depending on retrieval/device and resource constraints, including bandwidth, latency, decoder capability, things like video and audio reproduction capabilities (e.g. screen resolution and color space; speaker config)
* **Object** access – which objects to retrieve and possibly which parts of the objects
  + Descriptions, Nodes, etc.
  + Decoding capabilities, user preferences, etc.

# Design Goals and working assumptions

To face the challenges of delivering immersive media, the media data stored on a cloud resource should offer a flexible, fast, timely and efficient **random access** where the media dimensions include: Spatial, Temporal, Quality, Bitrate, Objects. In conjunction, the **coding** and **encapsulation** **formats** of immersive media should by design factor in these dimensions such that the interface between the delivery and the coded data itself is lightweight and efficient. From an architecture point of view, there is a need to define a receiver model combining user interactions, decryption, decoding, and rendering along with a retrieval model that leverages these multiple random-access dimensions. These two models would help to better understand how a client could **dynamically** and **efficiently** **tailor** the experience for the user. Lastly, the conformance points between the cloud resource and the decoders to consume such experiences should be specified to ensure interoperability in the ecosystem.

# References

1. ISO/IEC JTC1/SC29/WG11 MPEG2018/m43753, Streaming-First design for the MPEG-I project, Emmanuel Thomas, Rob Koenen (TNO), Ali C. Begen (OzU), Jill Boyce (Intel), July 2018, Ljubljana, SI