**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N16337**

**MPEG 115, Geneva, CH, June 2016**

|  |  |
| --- | --- |
| **Source** | **Requirements Subgroup** |
| **Status** | **Draft** |
| **Title** | **Requirements for Media Orchestration, v.4** |
| **Editor** | **Rob Koenen** |

Requirements for Media Orchestration

# Introduction

This document contains a first draft of the requirements for Media Orchestration. Please refer to the Context and Objectives for Media Orchestration [1] for background. Many of the requirements originate from the study on Uniform Timeline Alignment [2], albeit sometimes in modified form.

For simplicity, the requirements are captured in the form of requirements on “the specification”. This may be a single MPEG standard, but also multiple standards. Some requirements may already be fulfilled, possibly by standards published by other bodies. MPEG’s sub groups can do the appropriate analysis.

Underlined keywords shall have the meaning as in [3].

# Definitions

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Content | Media Data |
| Controller | See [1] |
| Data | Media Data or Metadata or Orchestration Data |
| Media Component | A component of the Media Data, e.g. an audio stream belonging to an audiovisual Scene |
| Media Data | Data that can be rendered, including audio, video, text, graphics, images, haptic and tactile information (NB: this data can be timed or non-timed) |
| Media Stream | Timed Media Data |
| Metadata | Data about other Data, that cannot be rendered independently and may affect rendering, processing or orchestration of the associated Media Data. |
| Orchestrator | See [1] |
| Orchestration Data | Data that orchestrates any number of Timed Data Streams  *Note that Orchestration Data itself may also be Timed Data* |
| Sink | Device that presents Media Data |
| Source | Device that captures and/or transmits data (Media Data and/or Metadata) |
| Timed Data | Data that has an intrinsic timeline |
| Timed Metadata | Metadata that has an intrinsic timeline |
| User | The human that uses one or more device(s) and/or service(s). A User could be a producer and consumer of media, and both at the same time. |

# Requirements

This document specifies requirements for metadata, protocols and packaging/multiplexing and in some cases coded representation of certain times of data. The protocols that the document asks for are application-level protocols. MPEG does not currently envisage specifying network protocols for Media Orchestration

## General Requirements

1. Reuse of and compatibility with existing standards:
   1. The specification should reuse existing standards to the extent possible;
   2. The specification shall be compatible with relevant existing media delivery methods, including MPEG‑2 TS, MPEG DASH, and MMT;
   3. The specification shall be compatible with relevant existing storage formats, which includes at least the ISO Base Media File Format.
2. The specification shall support protocols and metadata for multi-Source content capture
   1. The Specification shall support orchestrating a single media experience from multiple independent Sources
   2. The specification shall support discovery and coordination of heterogeneous Sources.
   3. The specification shall support the discovery and coordination of Sources whose availability is dynamic, in that they may become available and unavailable from time to time during an event that is being captured.
3. The specification shall support orchestration of live Sources as well as of recorded Media Data.
   1. The specification shall support metadata for the description of spatial and/or temporal characteristics of recorded Media Data
4. The specification shall support protocols and metadata for multi-domain content distribution: accurately and dynamically controlling play-out across different delivery methods on a single device and on multiple devices
   1. The specification shall support dynamic control of media streams characteristics according to user interaction.

Note: one example is HbbTV-style delivery to a primary TV screen and secondary companion screen.

1. The specification shall support protocols and metadata for accurately and dynamically controlling play out across multiple Sinks, in time and in space.
   1. The specification shall enable sinks to dynamically play out media streams according to the dynamically changing location and orientation of sinks.

Note: note that this includes, e.g., Sink A adjusting playback in response to change in position of Sink B

1. The specification shall support protocols and metadata for exchanging device information and characteristics, including for device calibration and for instructing devices to change device settings
   1. The specification shall support Sources reporting about the perceptual quality of captured, and possibly encoded, encoded audio and video;
   2. The specification shall support metadata for describing, and protocols for exchanging and controlling, the characteristics of sources and sinks.
   3. The specification shall support metadata for describing characteristics of sources and sinks. This includes capabilities and settings like frame rate, bit rate, resolution, size (for displays), pixels per inch, colour gamut, audio capture and reproduction capabilities, brightness, contrast, volume, loudness, focus distance, focal length, aperture, etc.
   4. The specification shall support protocols for doing so dynamically, when the characteristics described above may change over time, or when there is a need to actively make them change

Note: calibration means that sources can be mutually “harmonized”, e.g. in dynamic range and colour space. This also applies to sinks

1. The specification shall support protocols and metadata for exchanging dynamically changing media stream characteristics
   1. The specification shall support metadata for describing, and protocols for exchanging and controlling, characteristics of media streams. This includes frame rate, bit rate (audio and video), resolution, focus distance, focal length, aperture, sampling frequency etc.
      1. These protocols shall support dynamic characteristics, when the characteristic described above may change over time, or when there is a need to actively make them change
      2. The specification shall support signalling variations in the audio-visual sampling rate (frame rate), e.g., as it happens in user-generated content capturing devices such as smart phones.
2. Moved to Section with optional requirements; see requirement 26)
3. The specification shall support metadata for the description of content
   1. The specification shall support descriptions that allow searching for relevant sources
   2. The specification shall support description of content to allow matching of sources, e.g., to understand which sources can be stitched together (video), or combined (audio) into a single coherent experience
   3. The specification shall support descriptions that allow Scene Processors to create a coherent experience from multiple, diverse sources of Media Data

Note: MPEG-7 obviously provides a rich set for content description that will be used to the extent possible. This includes Compact Descriptors for Visual Search (CDVS) and Compact Descriptors for Video Analysis (CDVA).

1. Moved to Section with optional requirements; see requirement 27)
2. Moved to Section with optional requirements; see requirement 28)
3. The specification shall support a standardised representation of the Timed Metadata and Orchestration Data.
4. The specification shall support multiplexing of (Timed) Metadata and Orchestration Data in transport containers, including MPEG-2 TS and ISOBMFF.
5. The specification shall define APIs that allow Media Orchestration to be performed in a web environment.
6. The specification shall support metadata to enable the dynamic generation of composition information for existing composition standards (e.g., HTML5).

NB: the generation of such composition information may happen in response to user actions

1. The specification shall support the orchestration of communication channels, enabling a shared media-consumption experience.
2. The specification shall support identifying, in a composite signal, a certain subsignal, in order to be able to cancel that specific signal out, or to replace it with a clean copy of that signal.

Note: such clean copy may be used to suppress audio crosstalk or replace a flickering-screen capture.

## Requirements on Temporal Orchestration

1. The specification shall support protocols and metadata for “self-synchronisation”, in the sense that:
   1. The specification shall support synchronisation of sources with different media timelines, without a common clock.
   2. The specification shall support a posteriori synchronisation, i.e., alignment of captured and recorded media data, even if there was no intention to synchronise such data at the time of capture.
2. The specification shall support protocols and metadata for accurate synchronisation in the presence of delay between decoding and presentation through processing delays in, for example, high-end screens.
3. The specification shall support protocols and metadata for the orchestration of timed and non-timed media (e.g., video and stills into a single visual experience).
4. The specification shall support protocols for synchronisation in distribution systems that alter time stamps.

## Requirements on Spatial Orchestration

1. The specification shall support protocols and metadata for the spatial, dynamic orchestration of media coming from, and played across, multiple devices.
2. The specification shall support protocols and metadata for discovery of Sources, and accurately capturing and communicating their absolute and/or relative locations in a 3D environment and direction (gaze), also in a 3D environment,
   1. The specification shall support dynamically and accurately tracking these coordinates, and communicating such coordinates
      1. It shall be possible to dynamically signal the confidence in such coordinates
   2. It shall be possible to sync such coordinates with the media streams from those Sources

Note: Metadata may be added by sources, or it may be required to infer parameters through processing. Such metadata is often already available from professional equipment

1. The specification shall support protocols and metadata for discovery of Sinks (play-back devices) and their absolute and/or relative locations in a 3D environment
   1. The specification shall support protocols and metadata for capturing and communicating user and Sink position and orientation.
   2. It shall be possible to enable each Sink to select and render media streams or a part thereof according to a user’s viewpoint (position and gaze orientation).

Example: rendering, from a 360 degree video, only the part that the user actually sees

* 1. The specification shall support dynamically and accurately tracking these coordinates, and communicating such coordinates
     1. It shall be possible to dynamically signal the confidence in such coordinates
     2. It shall be possible to sync such coordinates with the media streams going to these Sinks

Note: even when sources and sinks are incorporated into the same device, with the same position and orientation, their “gaze” may differ (e.g., the gaze of the camera and screen in a mobile device may be in opposite directions)

## Requirements on Logical Orchestration

1. The specification shall support metadata for expressing logical relationships among Media Components, such as:
   1. membership and subset relationships in grouping Media Components
   2. hierarchical relationships in structuring Media Components
2. The specification shall support protocols for synchronizing logical relationships among Media Components, whenever there are updates in temporal, spatial and logical dimensions.
   1. The specification shall support metadata for expressing relationships in a fuzzy way, to enable “loose” relationships, in space and in time.

# Optional Requirements

The requirements in this Section are considered “nice to have” and would add to the capabilities of Media Orchestration, but they are not considered of core importance to Media Orchestration, and some fall outside of MPEG’s core expertise.

## General Requirements

1. The specification should support metadata for describing, protocols for communicating about, device network interface characteristics, for individual devices and for the captured scene, including at least:
   1. Network type, in use and available, including ad-hoc networking options
   2. Network parameters including bandwidth, delay, network location;
      1. Dynamic aspects of such network parameters
2. The specification should support protocols and metadata for the dynamic sharing of computational resources for distributed media processing;
   1. The specification shall support synchronising media processors in different locations, such that the output of such processors is correlated. Example: distributed encoders that produce temporally correlated, encoded bitstreams.
3. The specification should support protocols for network-based media processing for orchestration purposes.

# References

1. MPEG, ISO/IEC JTC1/SC29/WG11/N16131, *Context and Objectives for Media Orchestration, MPEG 114, October 2015*
2. MPEG, ISO/IEC JTC1/SC29/WG11 N14644, Exploration of “*Uniform Signalling for Timeline Alignment*”, MPEG 109, July 2014
3. S. Bradner (ed.) IETF RFC 2119, *Key words for use in RFCs to Indicate Requirement Levels*, https://www.ietf.org/rfc/rfc2119.txt