**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2015/N15343**

**MPEG 112, June 2015, Warsaw, Poland**

|  |  |
| --- | --- |
| **Source** | **Requirements Subgroup** |
| **Status** | **Approved by Reqs**  |
| **Title** | **Requirements for Media Orchestration, v.1** |
| **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  |
| 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 | Stream of Media Data  |
| Metadata | Data about other data (such “other data” may be Media Data).  |
| Scene | Collection of orchestrated Media Data and MetadataNote that this can apply to sources as well as to sinks |
| Scene Processor | Non-human user of orchestrated Media *(Note: this definition was added to acknowledge the possibility of non-human consumption of Media Data; no current requirement or use case refers to it)* |
| Sink | Device that renders Media Data |
| Source | Device that captures and/or transmits data (Media Data and/or Metadata) |
| Timed Content | Content that has a temporal dimension  |
| 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

## General Requirements

1. The specification shall support 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.
2. The specification shall support multi-domain content distribution: accurately and dynamically controlling play-out across different delivery methods on a single device and on multiple devices

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

1. The specification shall support accurately and dynamically controlling play out across multiple Sinks, in time and in space.
2. The specification shall support protocols for exchanging device characteristics, including for device calibration and for instructing devices to change device settings
	1. The specification shall support describing, exchanging and controlling 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.
		1. These protocols shall support doing so dynamically, when the characteristic 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 describing, 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. The specification shall support describing, and 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. Dynamically aspects of such network parameters
3. The specification may support capturing and communicating User position and orientation

Note: Support for describing source and sink position and orientation is already listed by other requirements; capturing user position and orientation may be helpful as well; this is to be discussed. This is probably useful on the Sink side; its use on the Source side is yet to be established

1. The specification shall support 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

1. The specification shall support 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, coded bitstreams.
2. The specification shall support network-based media processing for orchestration purposes

## Requirements on Temporal Orchestration

1. The specification shall support “self-synchronisation”, in the sense that:
	1. The specification shall support synchronisation of sources without a common (master) 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 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 the orchestration of timed and non-timed media (e.g., video and stills into a single visual experience)

## Requirements on Spatial Orchestration

1. The specification shall support the spatial, dynamic orchestration of media coming from, and played across, multiple devices.
2. The specification shall support discovery of Sources, and accurately capturing and communicating their relative locations and direction (gaze) 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 discovery of Sinks (play-back devices) and their relative locations in a 3D environment, including their orientation
	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

## Requirements on Logical Orchestration

1. The specification shall support 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 synchronizing logical relationships among Media Components, whenever there are updates in temporal, spatial and logical dimensions.
	1. It shall be possible to express relationships in a fuzzy way, to enable “loose” relationships, in space and in time

# References

1. MPEG, ISO/IEC JTC1/SC29/WG11/N15342, Context and Objectives for Media Orchestration, MPEG 112, June 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