**Coding of moving pictures and audio  
Convenorship: UNI (Italy)**

**Document type: Approved WG 11 document**

**Title: Definition of MPEG-I Phase 2 and subphases**

**Status: Approved**

**Date of document: 2020-01-17**

**Source: Requirements**

**Expected action: Review**

**No. of pages: 6**

**Email of convenor: leonardo@chiariglione.org**

**Committee URL: http://mpeg.chiariglione.org**

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N18970**

**January 2020, Brussels, BE**

|  |  |
| --- | --- |
| **Source** | **Requirements** |
| **Title** | **Definition of MPEG-I Phase 2 and subphases** |
| **Status** | **Approved; Publicly available** |

# Introduction

This document contains the current thinking on MPEG-I phase 2 sub phases. It is a living document that is open for comments and input.

The reason for defining phase 2a is as follows. MPEG-I phase 2 has been defined as tools for 6 degrees of freedom (6 DoF), but these elements are diverse and they are not all on the same time path. The goal of defining sub-phases in MPEG-I phase 2 is to deliver a meaningful and coherent subset of 6DoF functionality to be delivered in a reasonable timeframe. The reasonable timeframe has been defined as mid-2021 because a number of important standards will be available by then and it is a long enough period to provide missing functionality. Mid‑2021 also aligns with the completion schedule of 3GPP Rel-17. This is the third release for 5G, which is expected to be frozen in September 2021.

The document also starts to define a subphase 2b, and more sub-phases may be defined at a later moment, taking market developments into account, and encompassing more tools.

# What standards does industry need from MPEG for Immersive Media?

A brief and high-level assessment of relevant developments that guide what the industry requires from MPEG includes the following considerations:

* Augmented Reality (AR) is getting more and more attention.
* User-friendly 6DoF-enabled end-user devices are starting to appear on the market in the form of untethered headsets with inside-out tracking. A first step towards headset-based AR experiences are provided by VR headsets with cameras for inside-out tracking that can also be used to blend the real-world video with a virtual scene (e.g. the Oculus Quest). There are also glasses that project a rendered image into the user’s direct view of the real world (e.g. Hololens). These headsets are a good platform for AR experiences, and hybrid VR/AR applications and services.
* AR and VR experiences require means to compose graphics content and timed media together. Integrating timed media in workflows is becoming common. Not many tools exist for this today.
* Content and experiences for 6DoF devices (HMDs) require extending graphics with the additional support of other media.
* For the market adoption of these new media, it is key to provide technologies that would allow such media to be integrated in existing rendering engines such as Unity or Unreal in a developer-friendly manner.
* Immersive experiences will include “some kind of interactivity”; i.e., users are able to interact with the scene (switch on light, rotate object, etc.).
* Cloud access and streaming provide access to immersive media experiences feeling present (https://xinreality.com/wiki/Presence).

# MPEG-I Phase 2a definition

The mission of MPEG-I Phase 1a was to address storage and delivery of VR360 experiences in a 3DoF environment thanks to the definition of a new MPEG-I media element: OMAF.

The mission of MPEG-I Phase 1b was to build enhancements on Phase 1a VR360 experience, in a 3DoF+ environment. In the meantime, MPEG has been developing additional MPEG-I media elements such as VVC, Point Cloud Compression and MPEG Immersive Video.

It is agreed that it would be too ambitious to fulfil all of Phase 2’s requirements in a reasonable time frame, because there are clear signs that the industry is waiting for some sort of 6DoF experience. To meet these industry needs, MPEG should deliver a coherent set of tools in the next 18 months.

MPEG believes that defining a solution that allows bringing MPEG media objects into existing 6DoF environments is now more urgent than trying to provide a native 6DoF content format from scratch. A lot of 6DoF content is today computer-generated and limited to graphics; integrating MPEG media into these environments is an opportunity today.

Therefore, we define the mission of Phase 2a as:

**MPEG media for 6DoF environments**.

The MPEG-I Scene Description will be used for composition, and the following media types will be supported in immersive scenes:

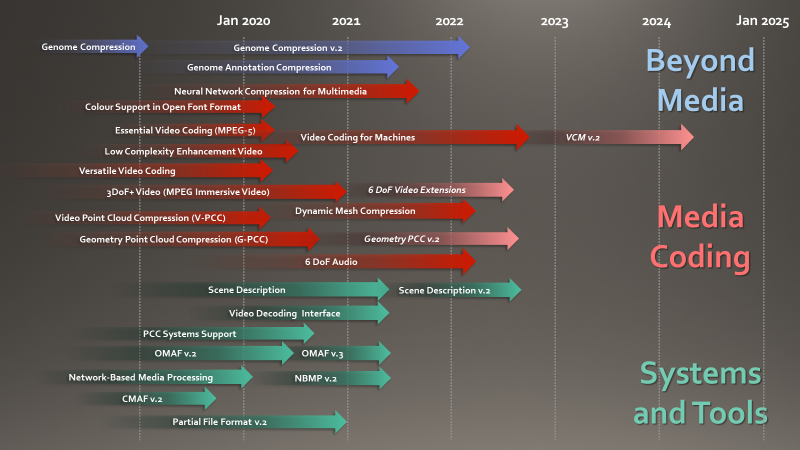
* 2D Visual media (enabling VVC, HEVC and other MPEG encoded video and images)
* Omnidirectional Video and images
* 3D (6DoF) visual media such as Dynamic Video-based Point Clouds
* Audio (enabling mono, stereo and MPEG-H immersive)
* Multiple media objects of the same or different types concurrently
* Network-based Media Processing, including interfaces for cloud-based rendering and hybrid cloud/local rendering (“split rendering”)
* Accessibility features such as subtitles/closed caption/timed text

MPEG-I Phase 2a will also enable efficient and adaptive streaming of these media types, be extensible, and include some form of interactivity with the media.

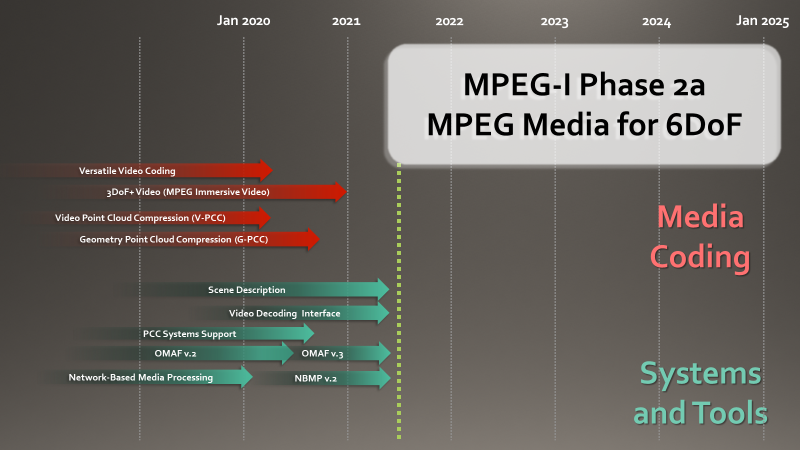
The MPEG-I Scene Description format will allow using MPEG media in both VR as well as AR environments.

MPEG expects to keep developing new media representation formats, more advanced Scene Description support as well as enable new delivery environments after July 2021 to expand the scope and functionality of MPEG-I.

MPEG’s Roadmap of major tools and standards under development is as follows (see N18975, a public document, for background):



 Of those tools, the following are relevant to MPEG-I Phase 2a:



## Example of a service enabled by MPEG-I Phase 2a

An example of the services that Phase 2a enables is a follows. A user with a VR headset sits in a 3DoF or 3DoF+ audiovisual scene. In the same scene, 3D volumetric objects are present. The user is able to inspect these objects and possibly interact with them.



If the video can be considered as far away with respect to the user, the user might even be able to move around the volumetric object(s). This might require a virtual floor to be projected as well, as in the picture below. Alternatively, the real environment can be partially blended in.



## First thinking on MPEG-I phase 2b

This section comprises the current thinking on MPEG-I Phase 2b. This Section is tentative and input is welcome, both from MPEG participants and industry representatives that do not take part in MPEG.

* The target is to have MPEG-I phase 2b ready at the end of 2022. Given the MPEG meeting schedule, this means the October 2022 meeting.
* Phase 2b will include:
  + MPEG-I audio
  + Updates to the Scene Description spec, including updates to support MPEG-I Audio
  + 6 DoF video extensions, if indeed ready according to the current planning
  + Dynamic Mesh compression
  + Geometry-based dynamic point clouds
  + Immersive Media Access and Delivery
  + More advanced cloud-based and split scene rendering, including integration with the Scene Description

The following elements of this section are still to be provided:

* a description of the new use cases enabled by Phase 2b
* if possible, a description on the types of products and services that are expected to be in the market by the release of MPEG-I phase 2b. One element is the expectation that there will be a significant 5G deployment by that time.

The picture below explains which elements will be ready by October 2022

