 ISO/IEC JTC 1/SC 29/WG 2 N00115

**ISO/IEC JTC 1/SC 29/WG 2  
MPEG Technical requirements   
Convenorship: SFS (Finland)**

**Document type:** Output Document

**Title:** MPEG Haptics Phase 2 – Motivation, Issues, and Use Cases

**Status:** Approved

**Date of document:** 2021-07-16

**Source:** ISO/IEC JTC 1/SC 29/WG 2

**Expected action:** None

**Action due date:** None

**No. of pages:** 10 (with cover page)

**Email of Convenor:** igor.curcio@nokia.com

**Committee URL:** <https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-2>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2 N** **00115**

**Online – July 2021**

|  |  |
| --- | --- |
| **Title** | **MPEG Haptics Phase 2: Motivation, Issues, and Use Cases** |
| **Source** | **WG 2, MPEG Technical requirements** |
| **Status** | **Approved** |
| **Serial Number** | **20705** |

Table of Contents

[1 Introduction 1](#_Toc77324902)

[2 Motivation for Phase 2 2](#_Toc77324903)

[3 Relation to Existing MPEG Standards 2](#_Toc77324904)

[3.1 Immersive Audio and Video Standards 2](#_Toc77324905)

[3.2 Scene description 3](#_Toc77324906)

[4 State of the industry 3](#_Toc77324907)

[5 Unique challenges of haptic media 5](#_Toc77324908)

[5.1 Perceptual Variance 5](#_Toc77324909)

[5.2 Avatar vs Physical Body 6](#_Toc77324910)

[5.3 Physical Device Mapping 6](#_Toc77324911)

[5.4 Closed vs. Open Loop Feedback 6](#_Toc77324912)

[5.5 Synchronization with Other Media 6](#_Toc77324913)

[6 Interactivity Models 6](#_Toc77324914)

[7 Additional Use Cases 7](#_Toc77324915)

[7.1 Virtual Prototyping 7](#_Toc77324916)

[7.2 Training and Simulation 8](#_Toc77324917)

[7.3 Tele-existence/Telerobotics 8](#_Toc77324918)

[8 References 8](#_Toc77324919)

# Introduction

The Call for Proposals on the Coded Representation of Haptics – Phase 1 [3] is in progress and well on track to select an RM0 technology at MPEG136 in October 2021. The Phase 1 CfP outlined the requirements for Phase 2 and highlighted the differences between Phase 1 (Basic Haptics) and Phase 2 (Advanced Haptics) in Annex 1. In this document, we reiterate the motivation for Phase 2, delve into the details of how haptics functionality in Phase 2 relates to existing MPEG-I standards (like Scene Description [8], Immersive Audio [5], OMAF [10]), explore some of the key issues and challenges in advanced haptics, and provide new use cases that were not mentioned in the MPEG-I Haptics Use Cases document [9].

# Motivation for Phase 2

The Phase 1 requirements (in Annex 1 of the CfP [3]) address the coding of time-dependent haptic signals and are suitable for the coding of timed-haptic experiences that may be synchronized with audio and/or video media. In particular, Phase 1 targets current haptic applications where a haptic effect is represented as a stream targeting a haptic device that is handled, grasped, touched, or worn by a user. As such, it includes the possibility to address devices with one or multiple actuators by means of a multi-track haptic stream. Such multiple actuator devices are currently available in gamepads, haptic suits, or other controllers. In addition, since there is a need to playback the appropriate experience, the localization relative to the user also needs to be coded. It allows the proprietary rendering application to appropriately manage the rendering of the right effects at the right placement relative to the user.

However, Phase 1 does not include any information about the location of the haptic effect in space, i.e., a scene where the user is playing or moving, nor any notion of objects and interaction such as in XR applications. Phase 1 is limited to channels. That is where Phase 2 comes in. Phase 2 requirements address objects, scenes, and interactions. In particular, the objects’ physical properties need to be added, interaction between objects/events and the user are also addressed through the avatar representation of the user, or any other interaction model. Further, Phase 2 dovetails well with the ongoing work in the MPEG-I scene description and MPEG-I Immersive Audio standards. It adds haptics functionality to virtual scenes – essential to complete the overall immersive experience (audio, video, and haptics) for users. Given that the MPEG-I scene description and Immersive Audio standardization work is further along than the haptics effort, one of the goals of Phase 2 is to ensure that the haptics architecture is consistent and “plays nicely” with those of scene description and Immersive Audio.

From a commercial perspective, there has been a significant increase in the availability of low-cost haptic devices suitable for immersive experiences. The recent increase in investment in virtual reality has resulted in pervasive availability of haptic devices such as the Oculus Quest and the HTC Vive Pro both of which feature vibration haptics in a bi-manual controller configuration. The Sony DualSense controller features quad-channel haptic functionality including two kinesthetic triggers and two voice coil motors and is currently one of the best available haptic experiences at the consumer electronics (CE) level. At the enterprise VR level, whole-hand devices such as SenseGlove provide multiple degree-of-freedom vibration haptics or optionally multiple-finger kinesthetic feedback. Other devices are in-development and are expected to be available commercially within the next few years. MPEG-I Phase 2 has the opportunity to establish media standards that can be rendered across all these devices, and which will enable immersive, interactive tactile experiences.

# Relation to Existing MPEG Standards

## Immersive Audio and Video Standards

Current MPEG standards address the coding and distribution of audio and video content. Recently, new immersive standards (referred as ISO/IEC 23090 MPEG-I) for both audio and video are in development, namely MPEG-I Part 4: Immersive Audio [5] and a set of video standards, including MPEG-I Part 12: MPEG Immersive Video (MIV) [6] and MPEG-I Part 5: Video-based Point Cloud Compression (V-PCC) [7].

Those standards specify how to encode volumetric content, but still lack the possibility to interact and freely navigate with the environment using 6 degrees of freedom (6 DoF). MPEG-I Phase 2 is considering those limitations and planning for extensions of the current standards. The audio group already published a Call for Proposals for MPEG-I Immersive Audio [5] to address those needs, and the plan is to do the same for video later.

However, immersive media should also consider the sense of touch, aka haptics. This new media has been considered since 2020 by the MPEG requirements group and integrated into the current MPEG roadmap. A Call for Proposals for the Coded Representation of Haptics has just been issued at the MPEG134 meeting [3]. It is considered as an additional media signal that extensions of the MPEG system standard will need to address the proper synchronization and encapsulation with A/V content at a later stage.

## Scene description

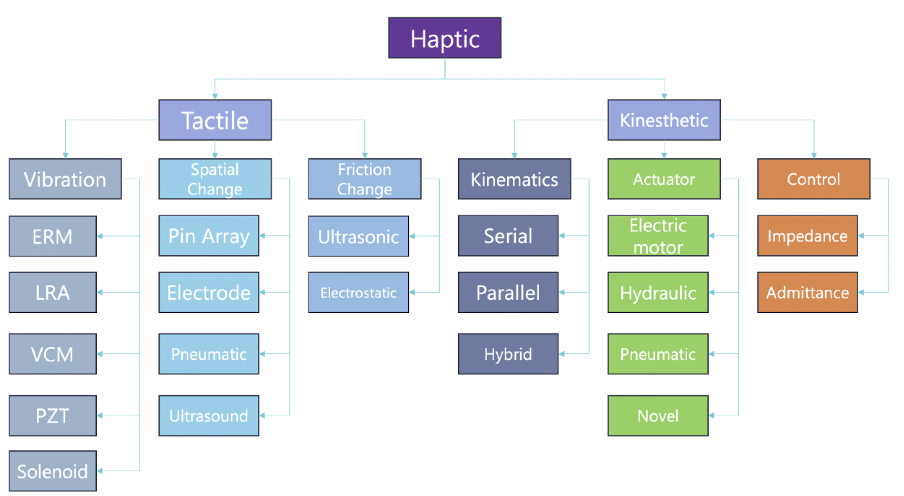
To interact with and navigate within a virtual or real scene, there is a need for a description of those scenes. MPEG is currently working on the finalization of such a standard called MPEG-I: Part 14 Scene Description (ISO/IEC 23090-14) [8]. It is based on the glTF format (GL Transmission Format) [2], a royalty-free specification for the efficient transmission and loading of 3D scenes and models by engines and applications. MPEG specifies extensions to this format to support its own formats for audio and video. In particular, the management of the A/V buffers and the synchronization between the different A/V elements and the scene.

Extension might also be necessary to integrate the coded haptic media (result of the ongoing CfP [3]) and be able to exchange information between the scene description and the haptic data.

# State of the industry

Haptics is increasingly used to bring physicality back in digital interfaces and products. The coupling of advanced sensing and tracking with physical feedback allows for a new class of experiences that look, sound, and feel more real than ever. Haptics can provide crucial sensory feedback to guide user’s perceptual understanding and assist basic motor control. Software buttons feel like mechanical parts, interfaces come alive, and game interactions evoke visceral reaction and sensation that approach real-world references. High quality haptic experiences are challenging to create thought, and the current offerings tend to be highly specialized and integrated, and single purpose. Proprietary solutions abound and experiences designed on one device are typically not portable, reusable, or extendable. Haptic feedback is typically a small part of the whole user experience, where it is carefully designed and orchestrated with regards to different input controls, sound, and visual elements.

In [1], a basic taxonomy of commercially available haptic devices is presented. It is adapted here for reference.



**Figure 1: Taxonomy of Haptic Modalities**

Haptic feedback covers a wide range of possible stimulation embodiments but is broadly divided into tactile and kinesthetic. Tactile feedback refers to sensations that stimulate subcutaneous mechanoreceptors such as vibration, friction, or micro-deformation. Kinesthetic feedback refers to sensations that provide force sensations which can stimulate both subcutaneous mechanoreceptors as well as proprioceptive mechanoreceptors.

The state of sophistication of the key categories of haptic devices and their suitability for standardization is summarized in the table below.

**Table 1: Suitability for Standardization of Haptic Device Categories**

|  |  |  |
| --- | --- | --- |
| **Haptic Modality** | **Technology Sophistication** | **Considerations for Standardization** |
| Vibration | Extremely sophisticated technologies with more than 20 years of commercial embodiments.  Current market leading products include iPhone and PlayStation 5.  Most emerging displays for XR utilize this technology.  Single vs multiple actuators. Both object that you hold (phone, game controller, steering wheel) vs interfaces that you wear (watch, vest) | Currently solutions are highly fragmented by product and market. Content signals are typically device specific or rely on proprietary coding and rendering frameworks.  Elementary stream coding is well covered by the MPEG-I Haptic Phase 1 CfP, but support of more complex arrangements and embodiments is needed for Phase 2. |
| Spatial Change | Non-contact ultrasound-based technologies are the most common commercial embodiment of spatial change haptics. These devices are still in early stages of development and commercialization. | Interactions with non-contact displays rely on a combination of hand-tracking and rendering of spatial point-cloud distributions. Likely a good candidate for support of an immersive MPEG coding standard. |
| Friction Change | Typically, 2D, these displays can modulate friction during contact interactions. There are no known use cases for friction modulation in immersive experiences and the technologies are still in the commercial adoption phase. | While not widely used, content for 2D friction displays can be derived from other 2D spatial haptic coded representations suitable for rendering on vibration or kinesthetic devices. |
| Kinesthetic | There are a small number of commercial kinesthetic displays, typically gloves with passive or active exoskeletons. These technologies are widely used in enterprise VR and XR use cases and are mature as a professional product category.  In addition, the PlayStation 5 DualSense controller incorporates 2 kinesthetic triggers and is a mature and widely available technology.  Simulators (driving, flying, medical training) and immersive furniture/room (entertainment, theme park, D-Box) | Interaction with 3D objects and the associated haptic sensations is the appropriate framework for kinesthetic displays. A suitable candidate for MPEG-I haptic coding. |

The current issues faced today are in the content creation and the device rendering steps. As the different assets and encoding are sparse, it increases the content creation complexity if you want the experience to cover a wide range of haptic capabilities. A standardized data format would unify the creation pipeline and drastically simplify the process.  
However, the standard data format must be flexible enough for content creation needs (spatial haptics, interaction-based haptics, etc.) but should also cover all relevant use cases on the devices market (XR controllers, exoskeletons, haptic gloves/vest, wearable haptic devices, automotive, game controllers, smartphones), even if the data rendering/device rendering is not part of the standard itself.

# Unique challenges of haptic media

Haptic experiences are very dynamic, relying on motor control and direct skin/bodily contact with the product/interface. They are inherently personal and do not exist or extend across space or distance, like audio and video do. Feeling haptics from a distance or remotely is a synthetic activity with no real-world analog or reference. Haptic media lacks intrinsic broadcast or shareable quality, as it is typically only personal (the individual touching/grasping the device).

Haptic media has several unique challenges that distinguish it from video and audio media. Some of these challenges are outlined below.

## Perceptual Variance

Touch is sensed principally by four simultaneous mechanoreceptor systems, each of which is attenuated to different types of stimulation. There are distinct types of receptors for transient vs. sustained contact, for sliding, for pressure, etc. Each of these sensation channels has distinct Weber fractions and frequency response characteristics. In addition, the distribution of mechanoreceptors varies across the 2m2 of skin area on the adult human body. Once these signals enter the brain, there is further psychophysical processing and integration. All these degrees of sensation variance represent opportunities to optimize coded representations but there are few existing academic or standards in this area.

## Avatar vs Physical Body

In existing 3DoF OMAF standards, the user is essentially represented as a head surrounded by a sphere on which video and audio are rendered. For tactile feedback in immersive environments, it is necessary to develop an explicit means for representing the user in a virtual environment. This representation is necessary since the display surface for tactile stimulation is the surface of the avatar. For example, if a virtual avatar’s leg intersects with a virtual chair, then a tactile sensation should be rendered on the real user’s leg. This representation and the mapping of this avatar to the user’s physical body is necessary but not currently standardized.

## Physical Device Mapping

In addition to the virtual-to-physical mapping, there is a further mapping to the available tactile display devices. Unlike audio and video, there are no existing universal haptic displays that can stimulate even a reasonable fraction of the skin surface across all mechanoreceptors. As such, there must exist a mechanism to remap or transcode a desired experience onto a typically small subset of body locations through available hardware.

## Closed vs. Open Loop Feedback

Kinesthetic haptic devices are often controlled using a closed-loop rendering algorithm that operates at 1kHz or higher and which has ultra-low latency requirements. This is distinct from audio and video in the sense that buffering of the high-rate signal is not permissible due to the interactive nature of the rendered experience. It is essential to ensure that this rendering loop rate is maintained and is responsive to user interactions. As a result, certain types of haptic experiences are specified as a function of user action, motion, or input (sensor value), instead of as a function of time.

## Synchronization with Other Media

The perception of a haptic sensation can be strongly influenced by simultaneous presentation of audio and video stimuli. Both the nature of these stimuli and their relative timing have a strong impact on the resulting haptic perception. Small variances in presentation, on the order of 10’s of ms, can cause haptic stimuli to be perceived as distinct events from audio and can lead to subliminal confusion or fatigue for users. Any standard that addresses multi-modal media needs to provide a means to accommodate this requirement.

# Interactivity Models

Haptic feedback is usually associated in a virtual environment with contact interactions between the user’s avatar and the environment, although this is not the only possible interaction paradigm. It is sometimes useful to categorize haptic feedback based on the intent of the experience designer:

**Immersive Haptics** - These are haptics that enable and provide a sense of presence for the user and are typically associated with environmental stimuli. For example, an explosion may be felt by a user in a virtual training environment. This type of haptic feedback is often well suited for *trigger and forget* presentation and pre-defined sensation. It is analogous to immersive 3D audio, in which predefined base media are transformed based on spatial position and other environmental effects.

**Agency Haptics** - These are haptics that enable a sense of agency for the user. These include contact and manipulation-related tactile sensations generated as a result of the user’s actions in the virtual environment. For example, picking up a virtual rubber ball and squeezing it suggests haptics that are generated during presentation based on the interaction between the user’s avatar and the virtual environment. This type of feedback is well suited for a functional dependence on the scene graph (e.g., force is a function of squeeze). For high quality haptic feedback, the actual rendering of this type of closed loop feedback is highly sensitive to loop rate and loop rate jitter.

The audio notions of diegetic and non-diegetic are related, but distinct concepts from this. In fact, this notion provides another level of interactivity as illustrated below.

|  |  |  |
| --- | --- | --- |
|  | Diegetic | Non-Diegetic |
| Immersive | User feels the rumble of a passing vehicle or nearby explosion. | Haptic effect is played during a scene transition or when the user selects alternate streams. |
| Agency | User feels the stiffness of a rubber ball they are holding in their virtual hand. | Haptic effect is rendered during interaction with a 2D UI overlaid on the virtual scene. |

There are likely distinct implications for the presentation engine and the media access framework in MPEG-I. Immersive haptics may be managed in a similar fashion as immersive audio, but Agency haptics is a new type of interaction for the framework.

# Additional Use Cases

In addition to the use cases described in the MPEG-I haptics use cases document [9], there are use cases that specifically utilize haptics, which are described in the HIF Enterprise VR Recommended Practices document [4]. These use cases, and others that are currently in development, are expected to drive updated requirements for the Phase 2 activities.

## Virtual Prototyping

A common use case for immersive VR technologies is for virtual prototyping. In the case of complex, time consuming, and expensive products such as airplanes, cars, and buildings, the use of pure virtual prototypes during early stages of product development is an essential and proven cost-saving and efficient replacement for physical prototypes. In many cases, it is further beneficial to incorporate human interaction into these virtual prototypes particularly if key aspects of the product design depend on human factors, aesthetics, or other uniquely human attributes. Haptics plays an important role in enabling tangible interaction and sensation for these virtual prototypes and has seen increased use as both haptic peripherals have become less expensive and as the prototypes have become more realistic.

## Training and Simulation

VR training (VRT) defines a solution using virtual or mixed reality to transfer useful skills towards the trainee with the use of an extended reality solution. Well-designed applications including haptic feedback in VR training can generate positive learning reinforcement, enhancing the effectiveness of training. This results in a lower rate of error during the training process.

VR training is commonly used in scenarios where there is a high cost of failure or for which generating the appropriate coverage of scenarios is complex or cost inefficient. Firefighter training, military training, and surgical training are all mature and common uses of virtual reality technologies that also typically make use of haptic feedback technologies and devices.

## Tele-existence/Telerobotics

Tele-existence is the ability to reproduce human capabilities in another physical avatar. Tele-existence is advantageous in dangerous situations/environments, sterile environments, and geographically distant locations. The advantages of tele-existence are the ability for real-time expertise, interactions, and oversight. This can vary from CBRNE (Chemical, Biological, Radiological, Nuclear, and high yield Explosives) to maintenance and medical use cases.

One of the most common examples of Telerobotics is found in modern surgical robotic systems, in which a commanded device is controlled by an expert surgeon from a remote console. These systems provide basic haptic feedback to the operator when the commanded device encounters bones or other hard objects within the surgical field.

# References

1. ‘Haptic display for virtual reality: progress and challenges,’ Dangxiao WANG, Yuan GUO, Shiyi LIU, Yuru ZHANG, Weiliang XU, Jing XIAO, Virtual Reality & Intelligent Hardware, Volume 1, Issue 2, 2019, Pages 136-162, ISSN 2096-5796,<https://doi.org/10.3724/SP.J.2096-5796.2019.0008>
2. [glTF Overview - The Khronos Group Inc](https://www.khronos.org/gltf/)
3. WG 02 N00070, “Call for Proposals for the Coded Representation of Haptics; Phase 1“, MPEG134, May 2021.
4. “Recommended Practices for Haptics in Enterprise VR”, Haptics Industry Forum, May 2021, <https://bit.ly/365gO0S>.
5. WG 06 N00056, “MPEG-I Immersive Audio Call for Proposals”, MPEG134, May 2021.
6. WG 04 N00049, “Text of ISO/IEC DIS 23090-12 MPEG Immersive Video”, MPEG133, January 2021.
7. WG 07 N00065, “Text of ISO/IEC DIS 23090-5 Visual Volumetric Video-based Coding and Video-based Point Cloud Compression 2nd Edition”, MPEG133, January 2021.
8. WG 03 N00264, “Text of ISO/IEC DIS 23090-14 Scene Description for MPEG Media”, MPEG134, May 2021.
9. N19513, “MPEG-I Phase 2 Haptics Use Cases”, MPEG131, July 2020.
10. ISO/IEC 23090-2:2019, “Information technology — Coded representation of immersive media — Part 2: Omnidirectional media format”.