Adding Touch to Immersive Media: An Overview of the MPEG Haptics Coding Standard

Philippe Guillotel, Yeshwant Muthusamy, Quentin Galvane, Eric Vezzoli, Lars Nockenberg, Iraj Sodagar, Henry Da Costa, Alexandre Hulsken, Gurvan Lecuyer, Matthieu Perreira Da Silva, François Bouffard, Heather Culbertson, *Member, IEEE*, Sandeep Kollannur, and David Gueorguiev

***Abstract*—In May 2021, MPEG issued a call for proposals for the specification of a new coding format for haptic data. Following this call, a baseline reference design and associated software implementation were defined for the representation and coding of haptic data. It resulted in a standard that defines a complete generic framework for the delivery of haptic signals, allowing the development of current and future haptic applications in the mobile, gaming, and virtual reality domains. This paper introduces the results of the first phase of the MPEG haptics coding standard. It includes the description of the codec architecture, the current performances in terms of compression efficiency, and the plans for the coding representation and distribution of haptics. The publication of the final ISO international standard is expected in 2024.**

***Index Terms*—Haptics, MPEG, coding, distribution, performances, evaluation.**

I. Introduction

Haptics is the science and technology of transmitting and understanding information through touch, encompassing the generation, manipulation, and perception of tactile sensations, forces, and motions. In the context of immersive media, haptics involves encoding and decoding these diverse touch-related data to enrich user experiences, allowing for the integration of physical feedback into immersive virtual environments. Immersive media that hitherto engaged only the visual and audio senses can be expanded by adding the sense of touch through haptics.

High-Definition haptics actuators are becoming prominent in consumer electronics interfaces such as the SONY PS5 DualSense controller and the Meta Quest Pro Controller. These actuators can reproduce effects ranging from subtle to sharp, force feedback, and other textured effects that simulate different surfaces and sensations. HD refers to the capability to render detailed and delicate touch sensations with high fidelity, by opposition to standard haptics such as smartphone vibrations for notifications.

The increased adoption of different haptics technology and control stacks leads to market fragmentation where each commercial device implements a proprietary control and coding pipeline. Proprietary APIs, incompatible hardware-to-software interfaces, and different motor characteristics are creating barriers for application developers to easily incorporate haptics into their applications. Changes to lower levels of the haptic stack (e.g., replacing the actuator) typically necessitate changes to the upper levels (e.g., the codecs and APIs used). The high cost of such changes often discourages research activities and causes OEMs (Original Equipment Manufacturers) to scale back haptics innovation. This lack of interoperability ultimately hurts adoption by end-users, the most important link in the haptics value chain.

To alleviate this problem, MPEG (Motion Picture Experts Group), a technical subcommittee of the ISO/IEC Joint Technical Committee I (JTC 1), known for foundational media standards such as MP3 [1], [2], AAC [3], [4], and HEVC [5], embarked on the standardization of haptics in late 2020. The goal was to specify both a representation and coding format supporting existing parametric and sample-based input signals, for the efficient distribution and storage of Haptics. Haptics coding is being standardized as part of the MPEG-I suite of immersive media standards, using a three-step approach:

- Phase 1: Develop a core haptic codec standard
- Phase 2A: Add support for spatial haptics and scene integration

Received 15 March 2024; revised 2 August 2024, 1 December 2024, and 22 January 2025; accepted 26 January 2025. Date of publication 5 February 2025; date of current version 20 June 2025. This work was supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – under Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" of Technische Universität Dresden. This article was recommended for publication by Associate Editor J. Park and Editor-in-Chief D. Prattichizzo upon evaluation of the reviewers' comments. *(Corresponding author: Philippe Guillotel.)*

Philippe Guillotel, Quentin Galvane, and Gurvan Lecuyer are with InterDigital, 35510 Cesson-Sevigné, France (e-mail: philippe.guillotel@interdigital.com; quentin.galvane@interdigital.com; gurvan.lecuyer@interdigital.com).

Yeshwant Muthusamy is with Yeshvik Solutions, LLC, Allen, TX 75013 USA (e-mail: yeshwant@yeshvik.com).

Eric Vezzoli and Alexandre Hulsken are with Interhaptics, 59000 Lille, France (e-mail: eric.vezzoli@razer.com; alexandre.hulsken@razer.com).

Lars Nockenberg is with the Technical University of Munich, 80333 Munich, Germany, and also with the Centre for Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden, D-01062 Dresden, Germany (e-mail: lars.nockenberg@tum.de).

Iraj Sodagar is with Tencent America, Palo Alto, CA 94306 USA (e-mail: irajsodagar@global.tencent.com).

Henry Da Costa is with Immersion Corporation, Aventura, FL 33180 USA (e-mail: hdacosta@immersion.com).

Matthieu Perreira Da Silva and François Bouffard are with Nantes Université, 44000 Nantes, France (e-mail: matthieu.perreiradasilva@univ-nantes.fr; francois.bouffard@univ-nantes.fr).

Heather Culbertson and Sandeep Kollannur are with the University of Southern California, Los Angeles, CA 90007 USA (e-mail: heather.culbertson@usc.edu; sandeep.kollannur@usc.edu).

David Gueorguiev is with Sorbonne Université, CNRS, 75252 Paris, France (e-mail: david.gueorguiev@sorbonne-universite.fr).

Digital Object Identifier 10.1109/TOH.2025.3539026

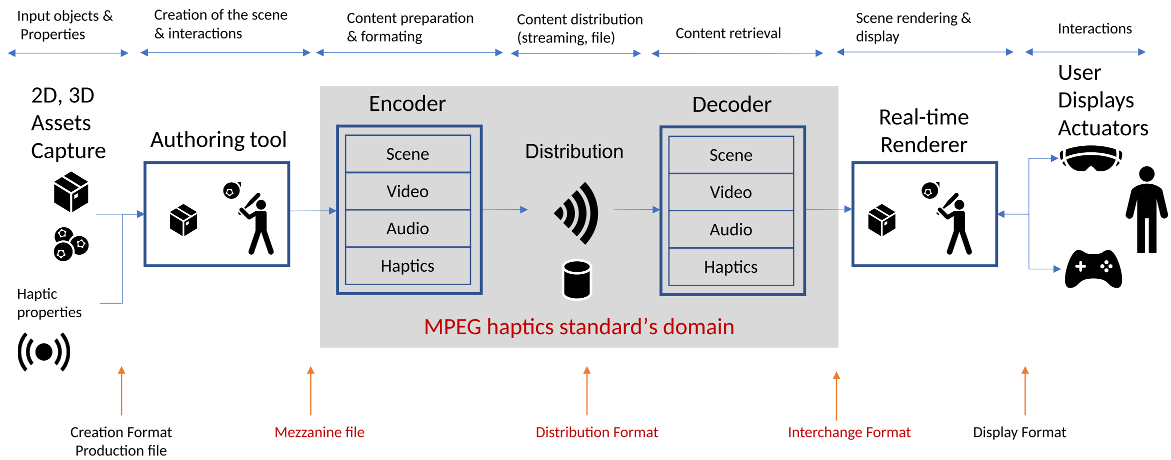

Fig. 1. Haptic media pipeline. It is similar to traditional A/V pipelines. MPEG standardizes the encoded representation and the decoder. The encoder and renderer are proprietary.

- Phase 2B: Add support for interactivity, avatars, and XR experiences (metaverse)

The codec in each phase will be built on technologies that were standardized in the previous phase.

Phase 1 of the MPEG haptics coding standard aims to standardize the representation and transmission of haptic data across the vibrotactile and kinesthetic modalities as a new media, enabling interoperable and high-quality haptic interactions across various devices and applications. This phase 1 focuses on the representation and encoding of haptics and less on interactive and closed-loop communication. The standardization process began with a Call for Proposals (CfP) in May 2021 [6]. Four organizations submitted their proposals in response to this CfP. These proposals were evaluated both subjectively and objectively and a baseline standardization candidate, termed Combined Reference Model 0 (CRM0), was selected in October 2021. The term 'Combined' refers to the fact that the baseline standardization candidate represented the merge of technologies from three of the four respondents to the CfP. Successive refinements of the standardization candidate were numbered accordingly. This paper describes the features of, and the results of the tests conducted on, CRM3, the refined standardization candidate that will be the Phase 1 MPEG Haptics codec standard.

This MPEG Haptics codec extends traditional MPEG media pipelines, starting from the capture or creation to the rendering, as shown in Fig. 1. The haptic experience authoring is usually performed through dedicated applications, possibly integrated into audio-visual creation tools. The creation process consists of manually synthesizing haptic effects or ingesting signals captured from dedicated sensors such as accelerometers, pressure sensors, or microphones. At the end of the pipeline, rendering software processes coded data to reconstruct the original signal, usually in real time. The device interface layer connects to appropriate user-facing hardware devices through specific haptic APIs such as game controllers, mobile phones, force feedback devices, and moving platforms. MPEG focuses on the middle of the chain: the encoded representation of the data, the interaction with other media like audio and video (including synchronization and transport), and the decoding of compressed data.

The different use cases and associated requirements for haptics are documented in [7] and [6], and can be summarized as follows:

- High quality preserving of frequency and amplitude independently.
- Being able to represent a drive signal between 5 Hz and 1000 Hz.
- Synchronization with audio and video.
- Playback transcoding to adapt to distinct target playback hardware with different capabilities, channel counts, physical placement, and arrangement relative to user.
- Supporting multiple simultaneous tracks and multiple encoded versions of a single track.
- Supporting lossy or lossless compression.
- Mixing multiple simultaneous tracks on a single output device with smooth transitions.
- User or application-controlled modulation of the overall haptic sensation.
- Haptic modality descriptor per track including vibrotactile and kinesthetic haptics.

This list was derived from a set of use cases addressing temporal haptic signals, such as mobile applications, movies/TV, gaming, and user interfaces (UI) where 3D spatialization is not required.

This paper describes MPEG Haptics Phase 1 core codec standard. Section II introduces haptic perception and describes previous work on haptic coding. Section III describes the different haptic effect types considered by the codec. Section IV details the codec architecture and the different tools provided, with the corresponding performances in Section V. Finally, Section VI concludes the paper with a short discussion of the MPEG roadmap for haptics.

II. Related Work

A. Haptic Perception

The human body is covered by four classes of mechanoreceptors: Merkel, Ruffini, Pacinian, and Meissner, each dedicated to perceive a range of tissue deformations [8]. Those receptors sense the influence that vibration, surface texture, deformation, and pressure create on the skin and inner tissues. Their pulsed responses are merged at the somatosensory cortex to create the tactile perception. Haptic sub-modalities are however quite different from audio and video. Unlike audio, human tactile sensors are distributed over $2\ \mathrm{m}^2$ of skin surface area all over the body. Skin, tendons and muscles receptors (esp. muscle spindles) from different body parts are activated to provide kinesthetic sensations, and the internal ear and body posture stimulates the perception of motion [9]. This leads to different properties to take into account when designing a coding format:

- The range of frequencies for dynamic haptics is $0.4\ \mathrm{Hz}$ to $800\ \mathrm{Hz}$ [8]. 5 Hz and below is particularly significant for kinesthetic haptics (force feedback). Compared to audio, haptics is thus a low frequency signal.
- The response depends on the receptive field and innervation density, leading to various response speeds. The type of signals to stimulate those receptors will thus be different [9].
- The human body is non-uniformly covered by different mechanoreceptors. The sensitivity to haptic stimuli depends on the innervation density of tactile afferents that varies considerably across different body regions [10]. Ref. [11] depicts this heterogeneous sensitivity with the "Sensory Homunculus", a representation of the body where body parts are scaled based on their sensitivity. This illustrates how important the localization of the stimuli on the body can be for the rendering and perception of haptic feedback.

B. Haptic Codecs

Ref. [12] provides an extensive overview of codecs for tactile and kinesthetic signals, and more generally of the various aspects of tactile and kinesthetic haptics. Ref. [12] lists use-cases in teleoperation, telepresence, and virtual reality and explains their requirements and challenges for haptic codecs. Since haptic and audio signals share similarities, audio codecs can be adapted to code haptic signals [13]. For instance, [14] used a speech codec but still achieved a relatively low compression ratio. Audio codecs are not optimized for haptics, especially considering the frequency dependence of the human sense of touch and perceptual masking, as well as the differing frequency ranges. This indicates the demand for specialized haptic codecs. Two waveform-based codecs for vibrotactile signals are [15] and [16], with the latter being incorporated into the codec presented in this paper.

Recent work on vibrotactile codecs also includes deep learning approaches like in [17], which uses an autoencoder architecture. Those promising architectures are still complex and not yet ready for market adoption.

The *IEEE Standards* Association developed a draft standard for the Tactile Internet (TI) as part of IEEE P1918.1 [18]. Haptics is considered as a sub-part of the specification (IEEE P1918.1.1) with codecs for tactile and kinesthetic signals [12]. There are two kinesthetic codecs for applications with negligible and significant delay, respectively. The codecs are based on [15], [19], and [20]. With its focus on teleoperation and remote touch applications, the main objectives for codec design were compression efficiency and stability. Therefore, those codecs do not support the parametric description of haptics, contrary to the codec presented here, which focuses on multimedia applications and their specific challenges like heterogeneity of replay devices.

In 2010, MPEG standardized an architecture and associated information representations to enable the interoperability between virtual worlds, e.g., digital content providers of a virtual world, (professional) gaming, simulation, and the real world, e.g., sensors, actuators, vision and rendering, and robotics [21], [22]. This standard, called MPEG-V, is very efficient for connecting different virtual worlds. Haptics is only a part of the coded information and is mainly handled through the description of the signals using the Sensory Effect Description Language (SEDL) (XML Schema-based). Therefore, it does not address all the requirements for haptic representation.

C. Evaluation and Metrics

Besides the PSNR (Peak signal-to-noise ratio, the log of the squared difference between two samples), traditionally used in audio and video coding for the objective performance of a haptic codec, there have been several proposals for subjective metrics. For kinesthetic signals, the HSSIM (Haptic Structural SIMilarity measure) [23] has been developed. The ST-SIM (Spectral Temporal SIMilarity) [24] and VibroMAF (Vibrotactile Multi-Method Assessment Fusion) [25] can be used for vibrotactile signals. The respective papers have shown that the metrics correlate well with perceptual tests. To get an evaluation result that closely represents real applications, it is preferred here to use a combination of PSNR and subjective experiments, as done for audio and video standards.

In addition, to evaluate the subjective performances, some test platforms for tactile and kinesthetic haptics have been proposed, based on custom [26] and available haptic devices [27].

Therefore, different coding schemes have been developed, but there is a need for a more generic codec, taking into account the different types of signals, modalities, and perceptions, and flexible enough to cope with variable networks and applications.

III. Types of Haptic Effects

There are mainly two ways to describe a haptic effect, which is a rendered haptic sensation for a certain duration. One is through a temporal signal captured with specific sensors, which we refer to as real haptics. The other way is through a description of the effect using dedicated primitives (Fig. 2). Those primitives can be either high-level primitives (e.g. an earthquake), or low-level primitives (e.g. a sinusoidal function of a given intensity and frequency). Usually, this is manually designed by a haptic artist and referred to as synthetic haptics.

Usually, real haptic signals are captured by sensors and represented with PCM (Pulse-code Modulation, i.e. time-based samples) [29] encoded data corresponding to a regularly sampled time series. PCM can be encoded with 8 or 16 bits per sample, with a sampling frequency ranging from 2 kHz to 16 kHz. Typical sensors are accelerometers, pressure sensors, temperature sensors, microphones, etc.

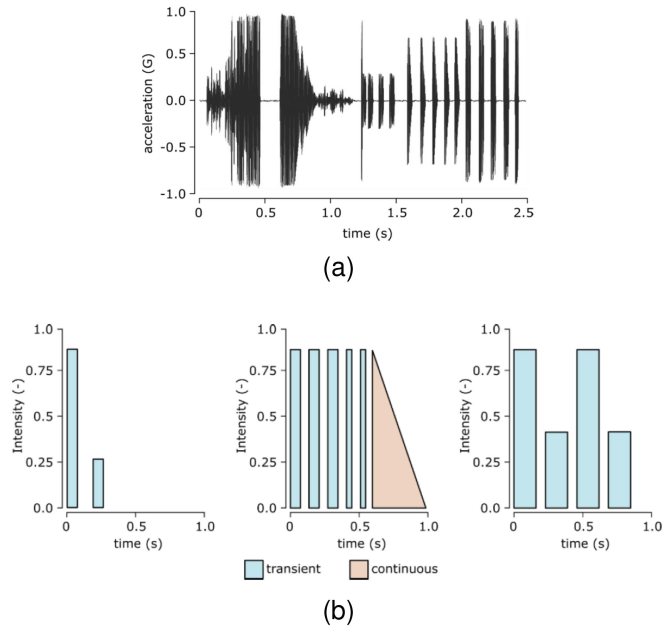

Fig. 2. Haptic effects. (a) Real haptics (captured temporal signal), (b) synthetic haptics (transient and continuous envelopes) [28].

Descriptive haptic is represented through parametric primitives, or descriptive signals (Fig. 2(b)). It can be described by an amplitude, a frequency, a duration, and a waveform shape such as sinusoid, or square. Two types of parametric effects are usually defined: transient and continuous. The descriptive coding represents transients as a short pulse with an amplitude, a position in time or space, and a reference frequency. A continuous signal is represented by a frequency and an amplitude information modulated on the timeline. Existing descriptive schemes use JSON or XML to describe haptic effects. Descriptive effects require a synthesizer to convert the primitives into a signal that a device can play back.

The codec described in this paper addresses these different types of haptic signals, as described in Section IV-B.

IV. MPEG Haptic Codec

The MPEG core Haptics coding standard has been designed to be generic, flexible, scalable, and multimodal to cover a wide range of multimedia applications. It addresses the coding of temporal and device-referenced spatial haptic data, supports localization on the user and is device agnostic. The next sections will detail those aspects of the codec, further detailed in the ISO/IEC 23090-31 "Haptics coding" standard specification [30].

A. Overview

The MPEG haptic standard was designed to support the encoding of both descriptive haptic data and PCM haptic signals. The descriptive input formats supported are the defined MPEG HJIF (Haptic JSON-based Interchange Format) format (the encoded MPEG format which can also be ingested) and two proprietary input formats (IVS and AHAP) (Section IV-B). This hybrid solution is based on the hierarchical data structure detailed in Section IV-C. This structure is shared by the two formats defined in the specification, namely the HJIF file format and the packetized MIHS (MPEG-I Haptic Streaming) format, detailed in Section IV-E. The global architecture of the codec is depicted in Fig. 3 and detailed in Section IV-D. It allows the use of different coding methods for different purposes, from low frequency effects to complex signals with high accuracy.

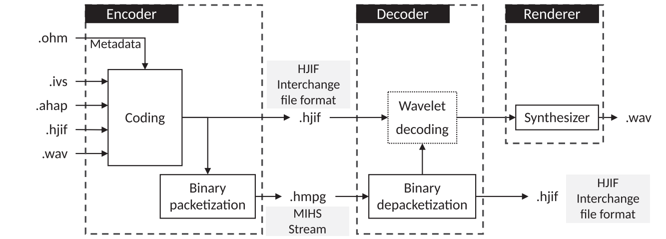

Fig. 3. MPEG haptics architecture (encoder and synthesizer are not normative, bitstreams and decoder are normative).

Both formats can be distributed with the appropriate network container: file, stream, and packet, and decoded at the receiver side. The network aspects are not part of the data coding standard, but a proper structuring of the data for the network layer is provided to prepare the data for network protocols, as described in Section IV-E. Appropriate network standards can then be extended to support this binary format.

Depending on the format, the receiver might decode the binary stream. The decoded data or original descriptive information is then interpreted by the rendering software to synthesize the original signal and convert it to device APIs, such as. wav files (series of PCM coded samples). The device adaptation is part of the manufacturer-specific device drivers.

Signal quality losses for descriptive input files might be generated only by the binary coding. The human-readable. hjif format is a transcoding method for descriptive haptics that does not introduce losses to the input descriptive file. For PCM data, this is different because instead of simple transcoding, lossy compression is applied to the input signal. Thus, both lossless coding and lossy coding are provided.

B. Input Formats

Both synthetic and real signals described in Section III are supported.

Synthetic haptic effects are typically generated with a dedicated authoring or editing tool [31], [32], [33], [34], by artists and designers trained in haptics, called hapticians [35] or haptographers. Those tools save the experience into an interchange file (a mezzanine file) for future editing or transfer to other production processes. The design process is usually associated with audio and video content. The interchange file is then ingested by the haptic encoder for storage or distribution.

Captured signals can be ingested into the aforementioned editing tools for potential editing or directly ingested by the encoder as PCM data [29].

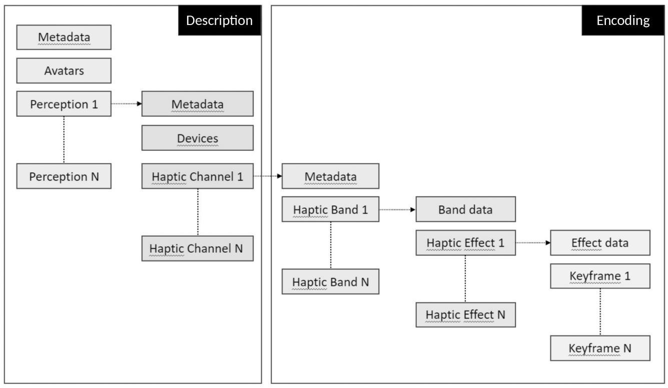

Fig. 4. Data structure.

TABLE I
LIST OF SUPPORTED MODALITIES [30]

Modality	Perception unit	Temporal or spatial
Pressure	Pa	Temporal
Acceleration	m/s^2	Temporal
Velocity	m/s	Temporal
Position	m	Temporal
Temperature	°C	Temporal
Vibrotactile	Normalized to [-1;1]	Temporal
Water	m^3	Temporal
Wind	m/s	Temporal
Force	N	Temporal
Electrotactile	Normalized to [-1;1]	Temporal
Vibrotactile Texture	Normalized to [-1;1]	Spatial
Stiffness	N/m	Spatial
Friction	Normalized to [-1;1]	Spatial
Humidity	Normalized to [-1;1]	Temporal
User-defined temporal	Normalized to [-1;1]	Temporal
User-defined spatial	Normalized to [-1;1]	Spatial

Effects can be of type temporal (time-dependent) or spatial (controlled by a spatial position).

The following input formats are supported by the specification for the reference implementation:

- Proprietary descriptive formats based on XML (.ivs) [36] or JSON (.ahap) [36], [37].
- MPEG descriptive format as described in this paper, based on JSON schema (.hjif).
- PCM sampled data (.wav). Ref. [38]

The content is described with the associated metadata file (.ohm) [36]. This optional metadata file includes the following information: name, description, version, number of haptic elements, and number of channels and their features such as gain and localization on the user body. This metadata file is intended to provide the necessary information for the configuration of an encoder.

C. Data Coding Structure

The coded representation of haptics defined in the standard is based on the hierarchical data structure depicted in Fig. 4 and shared by the two output formats HJIF and MIHS. This hierarchical construct organizes haptic data through five layers:

- **Experience**: Root of the data structure, it contains the complete haptic experience through one or more haptic perceptions.
- **Perception**: A perception defines the haptic signal for a given modality like vibrotactile, force, temperature, *etc.* Several perceptions may be encoded in the same stream/file and associated with a different type of end-user device.
- **Channels**: A perception may contain several channels. The different channels of a perception encode data of the same modality, but they may be associated with different actuators located in different places on the body like a haptic suit.
- **Bands**: The data contained in a channel is encoded through one or several frequency bands, each defining the data in a given frequency range. The standard defines four types of bands with different data representations.
- **Effect**: Each band contains one or several haptic events called effects. Effects are encoded through keyframes for descriptive content or binary wavelet streams for encoded PCM data.

Each level of the hierarchical structure contains metadata describing the corresponding data, providing information to configure the encoder or describing the encoded data for the rendering and synthesizer, as described bellow:

- **Experience metadata**
- Global information (date, version).
- A list of avatars representing the user. To allow mapping of an effect at a location on the user's body.
- A list of haptic perceptions.
- **Perception metadata**
- Information on the modality (from Table I) and associated avatar.
- A list of reference devices (including capabilities of the target actuators).
- A library of effects (for repeated effects).
- A list of haptic channels.
- **Channel metadata**
- Information on the channel (gain, mixing weight, body part, reference device).
- A list of haptic bands in a given frequency range.
  Three different body representations are available (Fig. 5), i) a simple mask, low memory, indicating semantic parts of the body (such as head front, upper back, right arm, left leg...), ii) a more complex but very accurate geometrical model (using a 3D mesh representation), and iii) a scalable approach based on the openXR representation [39].
- **Band data** Contains information about the band (frequency range) and type (*curve*, *transient*, *VectorialWave*, *WaveletWave*), and the list of haptic effects data.
- **Effect data** The encoded data for each band and channels. This is the encoded data information. The encoding process is described in Section IV-D.

This structure is encoded as a human-readable JSON-based file according to the JSON schemas specified. This Haptic JSON-based Interchange Format (HJIF) is ideal for authoring tools and can be manually edited. A .hjif file file can be ingested for further editing or binary encoding.

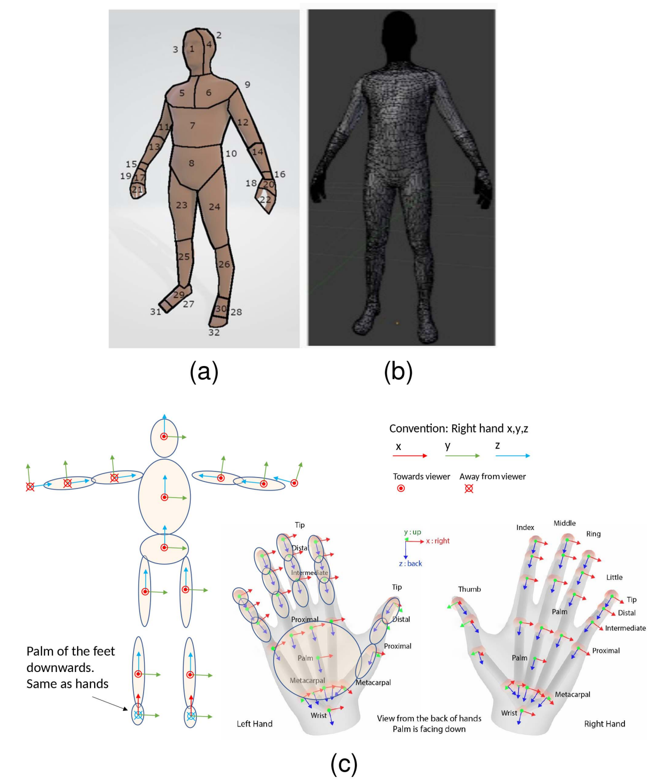

Fig. 5. User body representation. (a) Mask, (b) 3D mesh, (c) scalable (adapted from [39]).

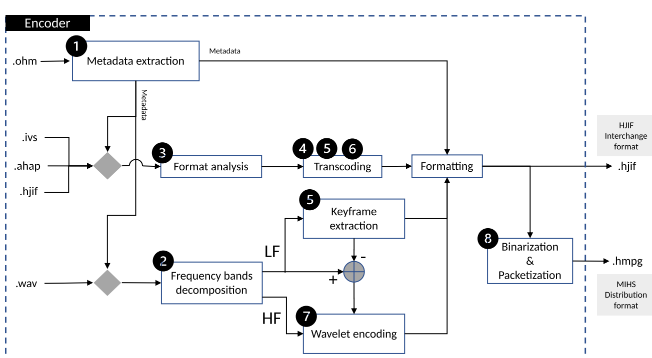

Fig. 6. Haptics coding architecture of the MPEG reference software (CRM). The numbers correspond to the sub-sections.

D. Encoding Haptic Effects

PCM and descriptive formats have distinct encoding processes described in this section. The elementary encoded information is an effect with additional information related to the rendering, included into the metadata (Section IV-C). The main rendering information includes the reference device (characteristics of the targeted device) and the body mapping (localisation on the user's body). The encoding format supports the description of the effect or a reference to a pre-designed effect (through a library of effects). Effects can also be combined. A generic encoder architecture is described in Fig. 6. It corresponds to the MPEG reference encoder (CRM), designed to demonstrate the codec capabilities. Different implementations can be designed corresponding to the targeted application.

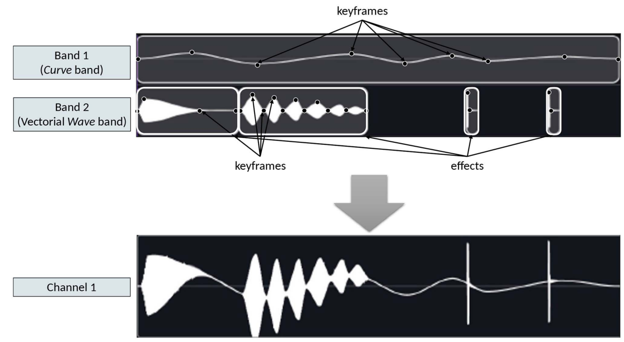

Fig. 7. Compositing of two bands (band 1 and band 2) to render final effect (channel 1).

Descriptive/parametric haptics is fundamentally based on the representation of a signal by a set of keyframes describing its amplitude, frequency and position. These keyframes encode a point at a given position in time. Intermediate temporal samples can be generated by interpolating with a given function at the rendering stage by a synthesizer. The combination of multiple primitive keyframes describes a full haptic effect. The type of keyframes and interpolating function are indicated in the haptic experience data (see Section IV-C/Band data). Three different keyframe coding methods are defined: *curve*, *transient*, *VectorialWave*; and five interpolation functions: *Linear*, *Cubic*, *Akima*, *Bézier*, and *BSpline*. Slowly varying effects (typical for force feedback) are generally coded using the *curve* band interpolation type. Vibrotactile effects are better encoded using the *VectorialWave* type through amplitude and frequency keyframes. And the *transient* type encodes shock effects through amplitude and frequency keyframes. To represent more complex signals, several bands (4 to 5) are generally designed and mixed together to synthesize the final signal. Fig. 7 describes this mixing with two bands, the first one encoded with *Curve* and the second one with *VectorialWave*. This descriptive format provides a framework to design haptic effects, but as depicted in Fig. 6, it is also designed to support other descriptive proprietary formats such as IVS and AHAP, using some transcoding.

PCM data (.wav) usually have a more complex frequency spectrum. Therefore, it can be encoded using one of the previous keyframe-based interpolation schemes (assuming adequate analysis of the input signal) or an additional encoding tool, called *WaveletWave*, using wavelet decomposition and coding able to cope with a large range of frequencies. In order to provide more flexibility, it is also possible to use any combination of those tools through scalable encoding with different bands. This is the choice made for the MPEG reference encoder (CRM) as described in Fig. 6, with two bands: a *Curve* band (to encode the low frequency part of the input signal) with an additional *WaveletWave* high frequency band. The cutoff frequency determines at which frequency the separation between base and high frequency bands is set.

The following subsections provide more details on those different encoding tools.

1) Metadata Extraction: The input. ohm file is a metadata file containing the information on the input data and the configuration of the encoder. Typically it includes the input format type and file name, the number of channels, the targeted end user device and a set of other information associated to the input file.

2) Frequency Bands Decomposition: This process splits the signal into frequency bands. The number and configuration of the low and high frequency bands is an encoder optimization choice. Here the proposed encoding structure as implemented in the reference encoder consists of one low frequency *Curve* band and one high frequency *WaveletWave* band as described in Fig. 6. The reference software also allows to skip this stage and only use a single band to encode the signal. Other encoders may use a different number of bands with different configurations.

The different frequency bands are then encoded using either one of the keyframe methods or the *WaveletWave* coding scheme, as described below. In general the *VectorialWave* and *WaveletWave* bands are more suited for temporal signals. The encoding of the low frequency band with a *Curve* band using the keyframe method is lossy, it only approximates the signal, therefore introducing a residual error (aka the error between the original signal and the reconstructed one from this band). To cope with this error, it is possible to compute this residual and add it to the high frequency band before encoding. This mechanism greatly improves the performances of the encoder especially at low bit-rates.

3) Format Analysis and Transcoding: Since several proprietary input formats are supported, an analysis is first performed to detect how haptic effects are described. Those proprietary descriptive files usually contain transient effects or curve effects. Transients are discrete short time effects, while curves are continuous time based effects. Then those effects are transcoded using the *transient* and *curve* coding tools, as described in the next sections.

4) Transient Coding: For a transient, the effect stores just a set of keyframes defining a position, an amplitude, and a frequency. As such, transient description are self contained effects.

5) Curve Coding: For a Curve band, each effect stores a set of keyframes defining a position and an amplitude plus an interpolation function to interpolate samples between the keyframes. The keyframes represent the control points of the curve. The type of curve representation or interpolation used to generate the full curve is specified in the metadata of the Band. Five traditional interpolation types are supported: *Linear*, *Cubic*, *Akima*, *Bézier*, and *BSpline*.

6) VectorialWave Coding: For VectorialWave bands, the effect stores a set of keyframes defining a position, an optional amplitude modulation, and an optional frequency. A signal waveform is also specified (five waveforms are defined: *Sine, Square, Triangle, sawToothUp, SawToothDown*) to synthesize the corresponding signal form.

7) WaveletWave Coding: The wavelet band processing is based on the codec presented in [16]. The encoder takes the high frequency band from the frequency band decomposition and the low frequency residual, and splits it into blocks of equal size, which is called block length. The block length has to be a power of 2 and at least 32 samples. The signal block is then analyzed in the Psychohaptic Model (PM). The lossy compression is applied by wavelet transforming the block using CDF9/7 filters and quantizing it using an embedded quantizer with individual bit depths for each wavelet band. The bit allocation is aided by the PM. In the end, each block is saved into a separate effect in a single band. The bitrate scaling is achieved by adjusting the sum of bit depths over all wavelet bands.

The PM is used to evaluate the perceptual importance of each wavelet band, and is tailored to vibrotactile signals. For this, masking in the frequency domain and the frequency dependence of the sensitivity of the human sense of touch are assessed. Masking in the frequency domain occurs, when content with higher amplitude overshadows other components of the signal with lower amplitude and similar frequency [40]. The PM transforms the input signal in the frequency domain using the Fourier transform and performs peak detection. The most prominent peaks are assumed to mask other frequency components and masker functions are generated for them. These masker functions are then combined with the absolute threshold of perception to form the global masking threshold. This threshold spectrum can be put into relation to the noise spectrum to assess the perceivability of noise. With this threshold, the PM can steer the quantization in such a way that perceptually less important signal components can be encoded more coarsely or be omitted. This PM can be modified to suit specific needs and haptic devices or be extended to cover more perceptual phenomena like temporal masking. Since the encoder is not standardized and the PM only influences the bit allocation, it will not interfere with standard conformance. The interested reader is referred to [16] for more details on the wavelet coding and the PM.

8) Binary Compression and Packetization: Binary coding is applied on all encoding methods, but in two different ways. The default way is to encode keyframes as integer numbers with fixed bit depth. For wavelet effects, lossless compression is applied using SPIHT and arithmetic coding. SPIHT stands for Set Partitioning In Hierarchical Trees and was introduced in [41]. It takes the quantized data from the wavelet processing module and leverages the similarities between wavelet bands to encode the data in an efficient bitstream. Arithmetic coding is then added to compress this bitstream further.

Packetization is then applied to provide a more useful bitstream for packet networks or other non-error-free communication media. The structure and process for this packetization is described in more details in Section IV-E.

E. Network-Ready Packetization Format

The MIHS (MPEG-I Haptic Stream) distribution format is a packetized binary stream stored as a binary file (.hmpg) or directly streamed. The format is designed to be easily encapsulated by any network protocol. The packets can be of variable duration and include two levels of packetization as shown in Fig. 8. The two levels of packetization are:

- MIHS unit that covers a duration of time and includes zero or more MIHS packets.
- MIHS packet that includes metadata or haptic effect data.

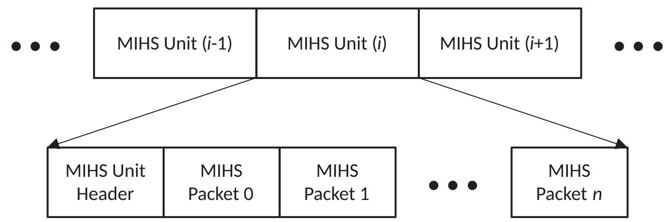

Fig. 8. MIHS packetization.

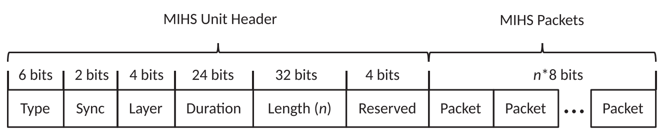

Fig. 9. Structure of an MIHS data unit.

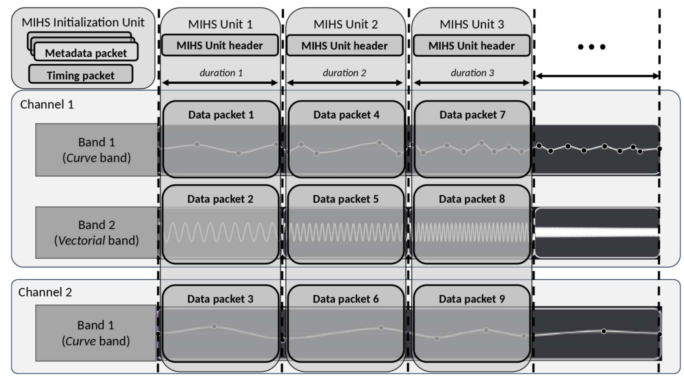

Fig. 10. Packetization of haptic effects into MIHS units. In this example, channel and band data is encapsulated into subsequent MIHS units. At any given point in time, all channel and band data is contained in a single MIHS unit. Specific MIHS initialization unit is used for synchronization and random access.

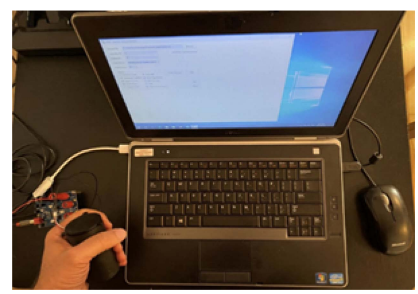

Fig. 11. Vibrotactile evaluation platform with the vibrotactile custom device held in the non-dominant hand.

Each MIHS unit covers a duration of haptic presentation time defined by a timestamp and a duration. Consecutive MIHS units are either temporally aligned, sharing the same timestamp and duration, or in sequence, i.e. an MIHS Unit starts at the end of the previous MIHS unit.

An MIHS unit comprises the unit header and MIHS packets, as shown in Fig. 9.

An MIHS unit may have one of the following types:

- **Initialization** which contains one MIHS initialization timing packet and may include one or more metadata MIHS packets. This unit starts a new time anchor in haptics presentation time by setting the timestamp and may also set or change the timescale. The duration of an initialization unit is zero.
- **Temporal** which contains one or more MIHS packets with temporal haptic effects. The duration of a temporal unit is a positive number.
- **Spatial** which contains one or more MIHS packets with spatial haptic effects. The duration of a spatial unit is zero.
- **Silent** which indicates that there is no effect that starts during this time interval. This unit does not include any MIHS packets other than a timing packet. The duration of a silent unit is a positive number.

The sync flag indicates whether the MIHS unit and the units that follow can be decoded independently from prior units, therefore whether the MIHS unit defines a random access point in the haptics presentation.

The timing of MIHS units and the timing parameters of MIHS packets are expressed in a time scale set in initialization MIHS units. Using this time scale, the haptics effects can be time-aligned with other media components, such as video and audio, with frame-level accuracy. All MIHS packets of an MIHS unit have the exact starting time and duration as the containing MIHS unit. The effect starting time is also signaled relative to the start time of the MIHS unit/packet that contains that effect. The starting time is negative for an effect started prior to the MIHS unit's start time, or positive for an effect starting during the MIHS unit's duration.

An example of MIHS packetization is illustrated in Fig. 10. An MIHS packet consists of a header and a payload. The header defines the type of packet, the layer that it belongs to in the case of layered coding, and its length. Various packet types are defined including timing packets, metadata packets, data packets, and packets for the effect library. An MIHS packet may carry one or more effects.

V. Evaluation

A. Methodology

Following the MPEG CfP methodology in [42], test streams, test platforms and protocols were defined to assess the quality of the proposed coding scheme with both objective metrics (PSNR) and subjective tests (MUSHRA [43]), as described below. The platforms designed for vibrotactile and kinesthetic signals are shown in Figs. 11 and 12. The objective of those platforms was to evaluate the potential perceived distortions induced by the signal encoding. Test subjects at three independent laboratories were asked to assess the subjective quality for three different bit-rates representative of the available range from moderate to high compression. The following compressed target bitrates were used: 64 kb/s, 16 kb/s, 8 kb/s, 2 kb/s.

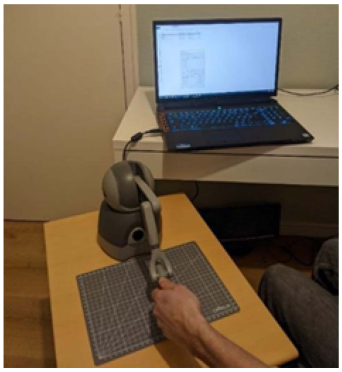

Fig. 12. Kinesthetic evaluation platform with the geomagic touch stylus held with the non-dominant hand.

TABLE II
FOSTER ELECTRIC CO. LTD ACTUATOR MODEL # 576865 SPECIFICATIONS [6]

Technology	Large bandwidth voice coil
Dimensions	$\phi 47 \times 13mm$
Resonance frequency	$90 \pm 18Hz$
Weight	57 g
Impedance	$5.5 \pm 0.8\Omega$ at 200 Hz

The codec presented here was evaluated in different configurations, but not to other codecs. The codec standardized in P1918.1.1 comes closest to it in terms of features and comparability. However, it is more focused on teleoperation and remote touch applications. This MPEG codec is targeting multimedia applications, leading to a much higher metadata overhead. Also, it includes descriptive formats that are not easily comparable to methods only taking waveform-based signals as input.

1) Codec Configuration: The version CRM3 of the MPEG reference software (https://github.com/MPEGGroup/HapticReferenceSoftware) was used with the default parameter set: cutoff frequency of 72.5 Hz and an automatic bit budget computed based on the target bitrate for wavelet bands.

2) Tactile Reference Device: The actuator used was the Foster Electric Co. Ltd model #576865 (Table II). The bandwidth of the actuator with a 100 g test mass is at least 1G-pp (peak-peak) of uniaxial acceleration from 60-500 Hz @ 1 Vrms [6].

The actuator has been integrated into a custom cylindrical mockup as displayed in Fig. 11. The cylinder shape was chosen since it provides maximum contact of the subject's palm and fingers with the mockup, ensuring that even minor differences in haptic effects can be adequately perceived. The build allowed us to fix the actuator within the custom-built cylinder for lateral vibrations relative to the palm. The actuator was placed at the end of the cylinder closest to the user's thumb. Some orientation markings were included in the device to allow different test labs to perform the same experiments. The cylinder was made from ABS plastic and measured Ø48x80 mm. The weight of the cylinder and actuator was 150 g [6].

The actuator was driven by a Texas Instruments TPA3112D1EVM audio amplifier. A UGREEN USB External Sound Card Audio Adapter was used between the computer and the amplifier. Test labs were instructed on how to calibrate the system to obtain a precise peak-to-peak output voltage from the amplifier for a given audio volume on the computer when playing a reference audio file [42].

TABLE III
NUMBER OF TESTS STREAMS PER TYPE

Test #	Type	# of Items
Test1-1	Vibrotactile Short Effects	8
Test1-2	Vibrotactile Long Effects	11
Test1-3	Kinesthetic Effects	10

While test labs could use any presentation system that complies with the MUSHRA methodology, it was highly recommended that the ARL STEP system be used [44]. Test labs were further instructed to equip subjects with headphones playing constant pink noise, and ask the subjects to hold the test device in their non-dominant hand [42].

3) Kinesthetic Reference Device: The 3DSystem Geomagic Touch[1] provides 3DoF force-feedback and inputs 6DoF position through a handheld stylus. It was used for evaluating kinesthetic content such as displacement and force feedback, as displayed in Fig. 12. The device was connected to a computer with an USB 2.0 interface. The content of a. wav file was played back by the MUSHRA test platform (see Section V-A6) that outputs the signal to an ASIO audio driver. This signal was then captured using a custom software based on the open-source API CHAI3D C++ [2] to set the force to be rendered by the kinesthetic device.

The users were trained to properly handle the stylus during the pre-screening test. In particular, for repositioning the stylus at the center position (using a marker on the table) and holding it flexibly and lightly to feel the stylus movements. The user's non-dominant hand was preferred for the test, the other hand was used to set the test value in the MUSHRA software (Section V-A6).

4) Test Streams + Descriptive Files: Three representative sets of test streams were provided by different companies and corresponded to market needs: Two sets for vibrotactile signals (short effects and long effects) and one set for kinesthetic signals (including force signals, acceleration, or movement). A vibrotactile short effect is an effect with a simple envelope (e.g., ADSR), where the envelope is the salient characteristic of the effect. A vibrotactile long effect is an effect where the envelopes and placement of its subcomponents are the salient characteristics of the effect. A total of 29 tests streams were used. The number of streams per test type is given in Table III. These test streams include both parametric and PCM input test files with duration ranging from 8 ms to 10 s. Further details on the test stream can be found in [6].

Each input file is associated with a description file (.ohm) providing relevant information of the input stream, such as file name, number of channels, body parts, or channel gain.

5) Objective Metrics: The Peak Signal-to-Noise Ratio (PSNR) is commonly used to measure the quality of reconstruction of lossy codecs (e.g., for audio or image compression). It is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of

[1] https://www.3dsystems.com/haptics-devices/touch
[2] https://www.chai3d.org/documentation/getting-started.

its representation. The PSNR is calculated to evaluate both vibrotactile and kinesthetic signals, following the formulation:

$$PSNR = 10 \times log_{10}\left(\frac{(v_{\max} - v_{\min})^2}{MSE}\right) \quad (1)$$

$$\text{with: } MSE = \frac{1}{N}.\sum_{i=0}^{N-1} |v(i) - \hat{v}(i)|^2 \quad (2)$$

where $v(i)$ is the modality (e.g., position, force, velocity, orientation...) at position i and $\hat{v}(i)$ its distorted counterpart. MSE is the Mean Squared Error.

In addition, the Bjontegaard metric [17] is used as a common way to compute an average PSNR or bit-rate gain for a whole range of bitrates, allowing results independent of either the bit-rate or the PSNR.

No additional metric was validated, therefore only the PSNR is used. However, since we are proposing incremental changes, the PSNR is considered acceptable to validate an improvement from a reference.

6) Subjective Evaluation: The MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) methodology [43] is a standardized approach for subjective audio quality assessment. It involves presenting listeners with multiple audio samples, including a reference and several test samples, some of which may be hidden or degraded. Participants rate each sample on a scale, allowing for detailed comparison and evaluation of perceived quality. MUSHRA has been successfully used within MPEG for several audio codec evaluations and a well-tested set of MUSHRA software tools was readily available [44].

MUSHRA was considered suitable for haptics testing because it allows for the direct comparison of multiple haptic stimuli against a known reference, facilitating nuanced quality assessments. The structured approach and inclusion of a reference help in capturing subtle differences in haptic feedback, ensuring reliable and repeatable subjective evaluations across both vibrotactile and kinesthetic haptic modalities.

The two test setups (vibrotactile and kinesthetic) described above were interfaced with the MUSHRA software to playback the tests streams on the respective platform. Test subjects were required to rank the signals from the different systems under test using the MUSHRA software interface. The different systems under test were anonymized and randomized.

The two test setups were provided to three independent university test labs (USC: University of South California, UNF: Nantes Université and SBN: Sorbonne university). The test labs were educated on the MUSHRA methodology and the test framework to ensure that tests at various labs followed the same rules and were comparable. Detailed instructions were provided regarding test subject seating and handling of the test devices. Subjects underwent a training session to familiarize themselves with the test setup and methodology, before starting their test sessions. Different training signals were used, corresponding to various types and levels of Haptic effects. Subjects wore headphones playing white noise to prevent external distractions but also to mask any potential sound from the devices.

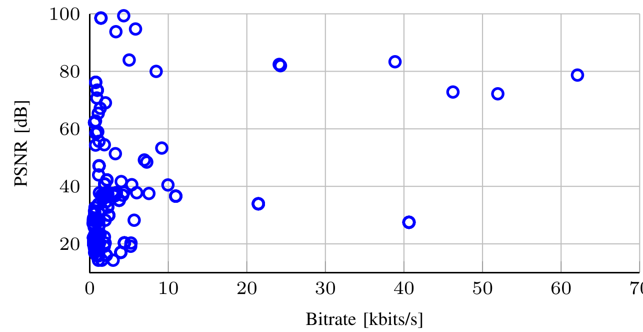

Fig. 13. Objective performances of the parametric input files. 8. AHAP files and 23. IVS files were used.

A total of 35 subjects participated in the tests, mainly students and researchers from the universities, with varying degrees of knowledge on Haptics. Each lab took appropriate action to have ethical approval and all participants gave informed consent. Data recorded from each subject was anonymized and no personally identifiable information was recorded.

Each individual user test session lasted about 30 minutes for vibrotactile and 1 h for kinesthetic. In each trial, the user experienced several versions of each test item, processed by a different system under test. Users were asked to judge the "Basic Haptic Quality" of the versions of the test item in each trial as compared to the open reference, with a linear ranking between 0 (bad) to 100 (excellent). Any perceived difference from the open reference was required to be rated down. One of the systems under test was always the hidden reference that was expected be given a rating of 100.

B. Performances

The performances given in this section were collected between July and September 2022, following the methodology described above and using the MPEG software reference model CRM3. For this round of user tests, the age distribution of the 35 test participants is: 25 participants aged [18-30], 7 aged [31-40] and 4 aged [41-50].

1) Objective Results: In order to assess the quality of the encoding and also to evaluate the required bit-rate for encoding high quality haptics, different evaluations were done. First, bitrate and PSNR were plotted for descriptive/parametric signals (.ivs and. ahap), as shown in Fig. 13. As depicted, the bit-rate ranges from a few kbps to roughly 60 kbps. Of course, for descriptive encoding, the bit-rate depends on the number of encoded effects. But for the set of tests streams used, corresponding to typical use cases, the distribution of bit-rates shown in Fig. 13 demonstrates that most content can be encoded with a bit-rate lower than 8kbps. PSNR values are indicative since the non normative synthesizer is included in the computation.

Second, for PCM tests streams, the PSNR of the decoded wave files was computed, using three configurations of the encoder:

- C2V: 1 band encoding using vectorial parameters
- C2W: 1 band encoding using wavelet decomposition

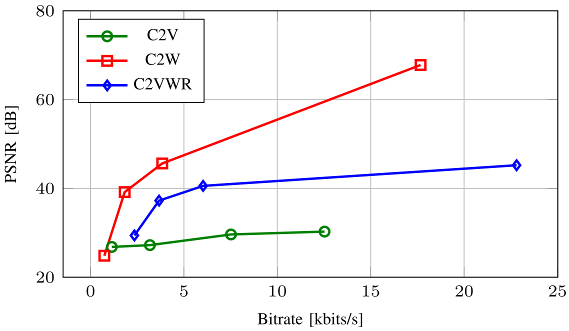

Fig. 14. Objective performance of the waveform based input files.

TABLE IV
TEST LABS AND SUBJECT COUNTS PER TEST (THE FIRST COLUMN UNDER EACH LAB IS THE TOTAL NUMBER OF SUBJECTS AND THE SECOND COLUMN IS THE NUMBER OF SUBJECTS THAT WERE POST-SCREENED)

Test Type	Test	Bitrate (kbps)	USC		SBN		UNF		Total Subjects	
VSE	Test1-1	2	7	3	9	8			16	11
VLE	Test1-2	2	4	1	9	5			13	6
	Test1-2	8	6	2	9	5			15	7
	Test1-2	16	6	2	9	6			15	8
KE	Test1-3	2	7	4			6	0	13	4
	Test1-3	8	6	3			5	0	11	3
	Test1-3	16	6	1			5	0	11	1

Test types are VSE (vibrotactile short effects), VLE (vibrotactile long effects), and KE (kinesthetic effects).

- C2VWR: 2 bands encoding, the low-frequency (LF) band using vectorial and the high-frequency (HF) band using wavelet decomposition and residual coding (corresponding to Fig. 6). Residual coding consists in computing the coding error from one band and adding it to higher band(s).

Fig. 14 depicts the resulting values for the full test streams. The wavelet encoding provides a more accurate reconstruction of the signal than parametric representation and interpolation, especially at high bit-rates. This is because arbitrary signals can be represented accurately using this method, when enough bits are spent for encoding. On the other hand, vectorial coding allows very low bit-rates with reasonable quality and reduced computational overhead. The residual coding provides a compromise between the advantages of both solutions, since errors in the low-frequency coding are compensated by the wavelet coding when the target bit-rate provides room for higher reconstruction fidelity. Additionally, one could also just decode the LF band to get a low frequency approximation of the signal.

2) Subjective Results: A rigorous statistical analyses of the MUSHRA results was performed, including a validity check using the post-screening rule from ITU-R BS.1534-3 standard [43]: *"If a subject scores the hidden reference below 90% for more than 15% of the test items in any test, then all the scores of that subject for that test are discarded."*. The final number of users per lab and test are given in Table IV. It is worth noting here that the MUSHRA methodology recommends a minimum of 8 test subjects per test. UNF expressed its inability to get at least 8 test subjects. Since this was the final round of verification testing (and prior rounds did indeed meet the 8 subjects/test criteria), it was decided to include UNF's test results as well.

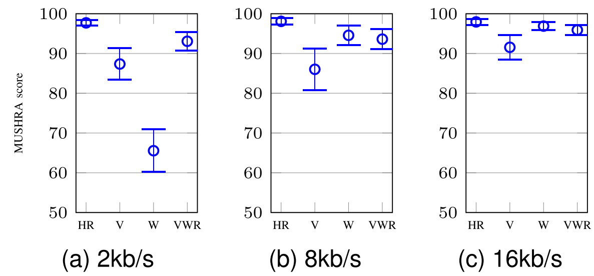

Fig. 15. Average subjective performances of the different encoder configurations for all input files & for three bit-rates, with the 2.5% confidence interval (HR: Hiden reference, V: Vectorial (C2V), W: Wavelet (C2W), VWR: Vectorial + Wavelet + Residual (C2VWR)).

The results are provided in Fig. 15 considering the 97.5% confidence interval. From the overall results (considering all bit-rates, streams and users), the wavelet encoding (C2W) exhibits lower results than both vectorial (C2V) and combined coding (C2WR) (i.e. 2 bands vectorial + wavelet residual coding). The combined C2VWR coding provides the highest score.

When looking more precisely on the results, per bit-rates, it can be further noticed that at low bit-rates C2W achieves a lower score, taking into account that the average bit-rate achieved was lower than with the other methods (due to a less accurate rate control, see Fig. 14). Looking at the results for the bit-rates separately, it can be seen that this only originates from the most aggressive compression setting, while C2W is performing better for the other bit-rates. C2VWR is overall a good compromise between the other two systems (taking the best of each) except for 2 kbps, where it achieved the best score, but also had the highest bitrate (see Fig. 14). Since the perceptual quality is high for all systems at medium and high bit-rates, the differences in score are lower. (Note however that the current version of the bit-rate control algorithm for wavelets, especially at low bit-rate, is not accurate and disadvantages this coding scheme).

C. Discussion

More statistical analysis were performed recently and detailed in [45]. This paper confirms that the task and methodology used in this work are suitable for haptics, considering what is done for other media such as audio or video signals. The final number of users in Table IV is however lower than those suggested by [45] (18 subjects for vibrotactile long effects or kinesthetic and 21 for vibrotactile short effects to obtain a satisfying degree of discriminability). Nevertheless, the results are consistent with the experts' assessment and sufficient to provide the general trend of the codec. This is especially the case for short effects with a high number of post-screened subjects. A one-way ANOVA test confirmed that for short effects it is difficult to discriminate between the different systems under test (probably due to the short duration). More complete analysis and tests will be done in the future.

In addition, the above performances were obtained with the reference software, which is not fully optimized. Thus higher performances can be expected for future implementations and these results represent a minimum expected efficiency. The configuration of the encoder is also left to the content provider; tuning the configuration is likely to result in higher performance, depending on the application. Finally, note that the average bit-rate is lower than the target bit-rate - which was used as maximum - set for the experiment since there is no real bit-rate control included into the reference model. Instead, encoder parameters were changed globally for all signals to scale the bit-rate.

VI. Conclusion

MPEG has finalized its development of a complete representation and coding format for Haptics. This standard is based on market needs and supports the main existing haptics formats. The codec architecture is flexible and generic to fulfill current and future applications and services. For instance, one may encode haptic effects using only a vectorial representation for very low bit-rate applications and simple synthetic effects, while a higher quality encoding can be achieved using an additional wavelet encoding band (or even using only wavelet encoding), for higher bit-rates and real signal recordings. From the evaluations performed during the development process, the following bit-rate ranges per channel can be recommended depending on the expected quality level:

- 2-4 kbps: perceptible distortion
- 8-16 kbps: some distortions, but not annoying
- 32-64 kbps: no perceptible distortion
- $\geq$ 64 kbps: perceptually lossless.

The standard is associated with a reference software (CRM), providing a reference implementation, and a conformance methodology for the validation of the standard developments (including conformance test streams). All the information is provided in the companion document ISO/IEC 23090-33 “Haptics coding conformance and reference software”, currently at the CD stage [46].

Additional efforts are ongoing for the integration of this MPEG Haptics format into global distribution systems. The MPEG Systems Working Group (ISO/IEC JTC1/SC 29/WG 3) is in the process of defining an ISOBMFF (ISO Base Media File Format) binding using the MIHS distribution format and support in DASH for the streaming of Haptics along other multimedia content. This will be a specific associated standard ISO/IEC 23090-32 “Carriage of haptics data”, currently in the final steps of the standardization (DIS) [47].

IETF is in the final stages of approving haptics as a new top-level media type [48], at the same level as audio, video, image, etc. Once published as an RFC, the two MPEG haptic formats ‘hjif’ and ’hmpg’ will be registered as sub-types under the ’haptics’ top-level media type. An internet-draft has also been submitted to specify transport of MPEG Haptics data into RTP packets [49]. These will facilitate adoption and deployment of Haptics for multimedia and communication applications.

While this first version of the MPEG standard (considering the coding of time-dependent haptic signals) is now ready for deployment, a new phase is started in MPEG to extend it to Haptic interactions, avatars and 3D objects for typically XR applications and the forthcoming metaverse. It will be integrated into version 2 of the MPEG Haptics standard.

Acknowledgment

This work was developed under the ISO/IEC JTC1/SC 29WG 02 and WG 07 MPEG groups.

References

[1] ISO/IEC JTC1/SC 29, *Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1.5 Mbit/s—Part 3: Audio, International Organization for Standardization*, Standard ISO/IEC 11172-3:1993, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 1993. [Online]. Available: https://www.iso.org/standard/22412.html

[2] ISO/IEC JTC1/SC 29, *Information Technology—Generic Coding of Moving Pictures and Associated Audio Information—Part 3: Audio, International Organization for Standardization*, Standard ISO/IEC 13818-3:1998, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 1998. [Online]. Available: https://www.iso.org/standard/26797.html

[3] ISO/IEC JTC1/SC 29, *Information Technology—Generic Coding of Moving Pictures and Associated Audio Information—Part 7: Advanced Audio Coding (AAC)*, International Organization for Standardization, Standard ISO/IEC 13818-7:2006, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2006. [Online]. Available: https://www.iso.org/standard/43345.html

[4] ISO/IEC JTC1/SC 29, *Information Technology—Coding of Audio-Visual Objects—Part 3: Audio, International Organization for Standardization*, Standard ISO/IEC 14496-3:2019, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2019. [Online]. Available: https://www.iso.org/standard/76383.html

[5] ISO/IEC JTC1/SC 29, *Information Technology—High Efficiency Coding and Media Delivery in Heterogeneous Environments—Part 2: High Efficiency Video, International Organization for Standardization*, Standard ISO/IEC 23008-2:2020, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2020. [Online]. Available: https://www.iso.org/standard/75484.html

[6] ISO/IEC JTC1/SC 29, *Call for Proposals for the Coded Representation of Haptics - Phase 1, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 02 N00070, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2021. [Online]. Available: https://www.mpeg.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20227.zip

[7] ISO/IEC JTC1/SC 29, *Updated MPEG-I Phase2 Haptics Use Cases, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 2 N00139, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2021. [Online]. Available: https://www.mpeg.org/wp-content/uploads/mpeg_meetings/136_OnLine/w20966.zip

[8] S. Bolanowski, G. Gescheider, R. Verrillo, and C. Checkosky, “Four channels mediate the mechanical aspects of touch,” *J. Acoust. Soc. Amer.*, vol. 84, pp. 1680–1694, 1988.

[9] F. P. McGlone and D. Reilly, “The cutaneous sensory system,” *Neurosci. Biobehavioral Rev.*, vol. 34, pp. 148–159, 2010.

[10] G. Corniani and H. P. Saal, “Tactile innervation densities across the whole body,” *J. Neuriophysiol.*, vol. 124, no. 4, pp. 1229–1240, 2020.

[11] W. Penfield and H. Jasper, *Epilepsy and the Functional Anatomy of the Human Brain*. Boston, MA, USA: Little, Brown, 1954. [Online]. Available: https://books.google.ca/books?id=mLOEAAAAIAAJ

[12] E. Steinbach et al., “Haptic codecs for the tactile internet,” *Proc. IEEE*, vol. 107, no. 2, pp. 447–470, Feb. 2019.

[13] X. Liu, M. Dohler, T. Mahmoodi, and H. Liu, “Challenges and opportunities for designing tactile codecs from audio codecs,” in *Proc. 2017 Eur. Conf. Netw. Commun.*, 2017, pp. 1–5.

[14] R. Chaudhari, B. Çizmeci, K. J. Kuchenbecker, S. Choi, and E. Steinbach, “Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation,” in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 409–418, doi: 10.1145/2393347.2393407.

[15] R. Hassen, B. Gülecyüz, and E. G. Steinbach, “PVC-SLP: Perceptual vibrotactile-signal compression based-on sparse linear prediction,” *IEEE Trans. Multimedia*, vol. 23, pp. 4455–4468, 2021.

[16] A. Noll, L. Nockenberg, B. Gülecyüz, and E. Steinbach, "VC-PWQ: Vibrotactile signal compression based on perceptual wavelet quantization," in *Proc. IEEE World Haptics Conf.*, 2021, pp. 427–432.
[17] Z. Li, R. Hassen, and Z. Wang, "Autoencoder for vibrotactile signal compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 4290–4294.
[18] O. Holland et al., "The IEEE 1918.1 'tactile internet'standards working group and its standards," *Proc. IEEE*, vol. 107, no. 2, pp. 256–279, Feb. 2019.
[19] X. Xu, M. Panzirsch, Q. Liu, and E. Steinbach, "Integrating haptic data reduction with energy reflection-based passivity control for time-delayed teleoperation," in *Proc. IEEE Haptics Symp.*, 2020, pp. 109–114.
[20] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, "Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 588–597, Feb. 2008.
[21] K. Yoon, S. K. Kim, J. J. Han, S. Han, and M. Preda, *MPEG-V : Bridging the Virtual and Real World*. London, U.K.: Academic Press, 2015. [Online]. Available: https://hal.science/hal-01271369
[22] ISO/IEC JTC1/SC 29, *Information Technology—Media Context and Control—Part 3: Sensory Information, International Organization for Standardization*, Standard ISO/IEC 23005-3:2019, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2019. [Online]. Available: https://www.iso.org/standard/71076.html
[23] R. Hassen and E. Steinbach, "HSSIM: An objective haptic quality assessment measure for force-feedback signals," in *Proc. Int. Conf. Qual. Multimedia Experience*, 2018, pp. 1–6.
[24] R. Hassen and E. Steinbach, "Subjective evaluation of the spectral temporal SIMilarity (ST-SIM) measure for vibrotactile quality assessment," *IEEE Trans. Haptics*, vol. 13, no. 1, pp. 25–31, Jan.–Mar. 2020.
[25] A. Noll, M. Hofbauer, E. Muschter, S.-C. Li, and E. Steinbach, "Automated quality assessment for compressed vibrotactile signals using multi-method assessment fusion," in *Proc. IEEE Haptics Symp.*, Santa Barabara, CA, USA, 2022, pp. 1–6.
[26] J. Kirsch, A. Noll, M. Strese, Q. Liu, and E. Steinbach, "A low-cost acquisition, display, and evaluation setup for tactile codec development," in *Proc. IEEE Int. Symp. Haptic, Audio Vis. Environments Games*, 2018, pp. 1–6.
[27] A. Bhardwaj et al., "A candidate hardware and software reference setup for kinesthetic codec standardization," in *Proc. IEEE Int. Symp. Haptic, Audio Vis. Environments Games*, 2017, pp. 1–6.
[28] E. Vezzoli, C. Ulrich, G. d. Butter, and R. Pijewski, "XR haptics, implementation and design guidelines. With enterprise VR application areas, use cases, and implementation examples," *Haptics Industry Forum*, Mar. 2022.
[29] Wikipedia, "Pulse-code modulation—Wikipedia, the free encyclopedia," 2023. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Pulse-code%20modulation&oldid=1143444419
[30] ISO/IEC JTC1/SC 29, *Text of ISO/IEC FDIS 23090-31 Haptics Coding, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 07 N00832, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2024.
[31] Interhaptics, "Haptic composer," 2022. [Online]. Available: https://www.interhaptics.com/tech/haptic-composer
[32] Apple, "Playing a custom haptic pattern from a file," 2023. [Online]. Available: https://developer.apple.com/documentation/corehaptics/playing_a_custom_haptic_pattern_from_a_file
[33] Haptrix, "Design, share & download cool haptic experiences," 2022. [Online]. Available: https://www.haptrix.com/
[34] F. Danieau, P. Guillotel, O. Dumas, T. Lopez, B. Leroy, and N. Mollet, "HFX studio: Haptic editor for full-body immersive experiences," in *Proc. 24th ACM Symp. Virtual Reality Softw. Technol.*, Tokyo, Japan, 2018, pp. 1–9.
[35] O. Schneider, K. MacLean, C. Swindells, and K. Booth, "Haptic experience design: What hapticians do and where they need help," *Int. J. Hum.-Comput. Stud.*, vol. 107, pp. 5–21, 2017.
[36] ISO/IEC JTC1/SC 29, *Encoder Input Format for MPEG Haptics, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 02 N00072, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2021. [Online]. Available: https://www.mpeg.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20229.zip
[37] Apple, "Representing haptic patterns in AHAP files," 2023. [Online]. Available: https://developer.apple.com/documentation/corehaptics/representing_haptic_patterns_in_ahap_files
[38] P. Kabal, "Audio file format specifications," 2022. [Online]. Available: https://www.mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/WAVE.html
[39] khronos, "The openxrspecification," 2017–2024. [Online]. Available: https://registry.khronos.org/OpenXR/specs/1.0/html/xrspec.html
[40] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach, "Perceptual and bitrate-scalable coding of haptic surface texture signals," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 3, pp. 462–473, Apr. 2015.
[41] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
[42] ISO/IEC JTC1/SC 29, *Submissions and Evaluation Procedures for Haptics CFP–Phase 1, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 02 N00071, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2021. [Online]. Available: https://www.mpeg.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20228.zip
[43] ITU SG 6, *Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems, International Telecommunications Union*, Standard ITU-R Recommendation BS.1534-3, International Telecommunication Union, Geneva, Switzerland, 2015. [Online]. Available: https://www.itu.int/rec/R-REC-BS.1534
[44] Audio Research Labs, "Subjective training and evaluation program (STEP) v1.09: A computer-controlled system for audio presentation and subjective evaluation," Jan. 2015. [Online]. Available: https://www.audioresearchlabs.com/arl-step/
[45] A. Pastor and P. L. Callet, "Towards guidelines for subjective haptic quality assessment: A case study on quality assessment of compressed haptic signals," in *Proc. IEEE Int. Conf. Multimedia Expo*. Jul. 2023, pp. 1667–1672.
[46] ISO/IEC JTC1/SC 29, *Text of ISO/IEC CD 23090-33 conformance and reference software, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 07 N00793, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2024.
[47] ISO/IEC JTC1/SC 29, *Text of ISO/IEC DIS 23090-32 Carriage of Haptics Data, International Organization for Standardization*, Standard ISO/IEC JTC 1/SC 29/WG 03 N01028, International Organization for Standardization, Geneva, Switzerland, International Electrotechnical Commission, Geneva, Switzerland, 2023.
[48] Y. K. Muthusamy and C. Ullrich, "RFC 9695: The 'haptics' top-level media type," Internet Engineering Task Force, IESG approved; publication pending, Dec. 27, 2024. [Online]. Available: https://datatracker.ietf.org/doc/draft-ietf-mediaman-haptics/
[49] H. Yang and X. d. Foy, "RTP payload for haptics," Internet Engineering Task Force, Active Internet-Draft, Oct. 16, 2023. [Online]. Available: https://datatracker.ietf.org/doc/draft-hsyang-avtcore-rtp-haptics/

Philippe Guillotel received the Engineering degree in electrical engineering, telecommunications, and computer science from the Ecole Nationale Superieure des Telecommunications, Paris, France, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the University of Rennes, Rennes, France, in 1986 and 2012, respectively. He is currently the Research Director of InterDigital. His research interests include UX/user experiences (immersive experiences, user sensing), media coding and distribution (2D, 3D, avatars, haptics), and AI for media representation and coding. He also leads a team of experts in the field of 2D and 3D video coding, haptics representation and coding, and avatar communication. He is chairing the MPEG Ad'hoc group responsible for the specification of the haptic coding standards.

Yeshwant Muthusamy received the B.Tech. degree in in computer science and engineering from Jawaharlal Nehru Technological University, Hyderabad, India, and the Ph.D. degree in computer science and engineering from Oregon Graduate Institute, Beaverton, OR, USA. He has held leadership positions with companies like Immersion Corporation, Toyota North America, Samsung Research, Nokia, and Texas Instruments. He is currently the Owner and Principal of Yeshvik Solutions LLC, an AI consulting firm based in Allen, TX, USA. At Yeshvik Solutions LLC, he helps clients with customized AI solutions, AI training, and intellectual property advisement in AI. He holds six patents and has authored or coauthored more than 30 peer-reviewed publications. As the First Chair of the MPEG Ad-hoc Group on Haptics, he led the specification of the standard from the Call for Proposals (CfP) through the Community Draft (CD) stages.

Quentin Galvane received the Engineering degree in computer science from the National Institutes of Science and Technology, France, in 2012, and the Ph.D. degree in computer graphics from INRIA, Paris, France, in 2015. After completing the Ph.D. and two Postdocs with Technicolor and INRIA, respectively, he joined Interdigital to work on avatars and the standardization of haptics. He is currently a Senior Researcher with InterDigital. He has significantly contributed to the standardization efforts led within MPEG, co-chairing the Haptic group and co-editing the ISO/IEC 23090-31 standard on haptic coding, as well as coordinating the associated reference and conformance software.

Eric Vezzoli received the Ph.D. degree in haptics from Lille University, Lille, France. He funded Interhaptics allowing the creation and deployment of HD haptics for any platform. Interhaptics was acquired by Razer in 2022. He is currently the Director of Technology of Razer. He is the Founder and co-host of the Haptics Club podcast that shares and documents best practices for haptics implementers, and the Director and XR Haptics working group chair with Haptics Industry Forum. He has authored more than 50 scientific papers and patents in haptics and human-machine interaction.

Lars Nockenberg studied electrical engineering from the Technical University of Munich, Munich, Germany, with focus on signal processing, haptics, and acoustics. He received the B.Sc. and the M.Sc. degrees in 2019 and 2022, respectively, from the Technical University of Munich, where he has been working toward the Ph.D. degree with the Chair of Media Technology since April 2022.

Iraj Sodagar received the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1994. He has participated, led, and managed various R&D projects, advanced architecture design and product development including image and video coding, media indexing and analysis, media storage, transport and delivery, and media transcoding on the cloud with various companies. He built and managed large technical teams for R&D, software development and productization, and standardization of developed technology. Dr. Sodagar had various leadership and engagement positions in ITU-T, JPEG, MPEG, AMQP, W3C, DVB, 3GPP, W3C, CTA WAVE, DLNA, and AOMedia. He started Wireless Media Forum, the first consortia on mobile video delivery, in the early 2000 s and DASH Industry Forum in 2014 and served as its President until 2024. He has been the chair of MPEG's DASH subgroup from its very start, and the co-chair of MPEG's CMAF, NBMP and haptics subgroups. His main research interests include media streaming and delivery, 5G and 6G architectures for media delivery and media content authentication/verification.

Henry Da Costa received the B.Eng. degree in computer engineering from Concordia University, Montreal, QC, Canada. He previously held leadership positions in software development with Softimage, providing computer graphics modeling, animation, and rendering software; and senior positions in computer graphics and finite element analysis software development at several other companies. He is currently the Director of Engineering with Immersion Corporation, where he has held leadership positions in haptic technology innovation for more than 22 years of experience and is an inventor on numerous patents in the field.

Alexandre Hulsken received the master's degree, specializing in computer vision and human-machine interaction, with a focus on natural interaction systems for VR/AR/XR. Since then, he has shifted his primary focus to haptic technologies, contributing significantly to the development of haptic file formats and adaptive rendering frameworks, enabling high-definition haptics across various platforms. He has been with Interhaptics since his masters studies in 2018 and is currently a Lead Engineer at Interhaptics, a Razer subsidiary since 2022.

Gurvan Lecuyer received the Ph.D. degree in AI on the analysis and automatic annotation of videos. Gurvan did Postdoc with the European Space Agency on the latest AI-based 3D rendering methods. Gurvan is currently a Researcher with InterDigital. Gurvan joined InterDigital to work specifically on Haptics. Gurvan has been a major contributor within the standardization group, particularly on the definition of the binary format and more specifically on the aspects of packetization of these binary data, essential for data management in streaming.

Matthieu Perreira Da Silva received the M.Sc. degree in image processing and the Ph.D. degree in computer science and applications from the University of La Rochelle, La Rochelle, France, in 2001 and 2010, respectively. From 2001 to 2006, he was a R&D Engineer in a private company dealing with biometric identification. From 2006 to 2011, he was successively an engineer, a Ph.D. student, and a teaching assistant with the University of La Rochelle. Since 2017, he has been the Co-Head of the Visual Computing international master track of Nantes University computer science master. Since 2020, he has been the Co-Head of XP-LAB, an interdisciplinary userlab located in Nantes Creative district. He has been an Associate Professor in the Image Perception Interaction (IPI) team of LS2N Lab since 2011. Apart from his research activities in the team, he is also teaching in the Computer Science Department, Polytech Nantes Engineering School. His research interests include AI, human perception, visual attention, human computer interaction, quality of experience, image processing, and computer vision.

François Bouffard is currently a Technical Manager based in Nantes with extensive experience in audiovisual technologies and digital innovation. Since 2019, he has been the Operational Manager of the Experience Lab with Halle 6 Ouest, an interdisciplinary research platform at the University of Nantes, Nantes, France, dedicated to user experience. His work involves overseeing technical operations, conducting technology watch, and supporting research projects, particularly those exploring interactions between technology, visual arts, and human perception. He also contributes to training and knowledge dissemination in these fields.

Heather Culbertson (Member, IEEE) received the B.S. degree in mechanical engineering from the University of Nevada, Reno, NV, USA, and the M.S. and Ph.D. degrees from the Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA, USA. She was a Research Scientist with Stanford University, Stanford, CA, USA. She is currently an Assistant Professor of computer science, aerospace and mechanical engineering, and biomedical engineering with the University of Southern California, Los Angeles, CA. Her research focuses on the design and control of haptic devices and rendering systems, human-robot interaction, and virtual reality.

Sandeep Kollannur received the B.Tech. degree in computer science from the Cochin University of Science and Technology, Kochi, India, in 2007, the M.S. degree in human-computer interaction with ergonomics from University College London, London, U.K., in 2015, and the second M.S. degree in computer science from the University of Calgary, Calgary, AB, Canada, in 2019. He is currently working toward the Ph.D. degree in computer science with the University of Southern California, Los Angeles, CA, USA.

David Gueorguiev studied physics and computational neuroscience. He received the Ph.D. degree in the neuroscience of touch from UCLouvain, Brussels, Belgium. After the Ph.D. he did postdocs with Inria, Lille, France, and the Max-Planck Institute. From 2020 to 2024, he was a Research Fellow with CNRS and Sorbonne University, Paris. He is currently a Research Associate with FNRS and UCLouvain. His scientific objectives relate to the role of tactile and multisensory sensations in decision-making and attentional processes.