International Standard

ISO/IEC 23090-4

First edition
2025-11

Information technology — Coded representation of immersive media —

Part 4: MPEG-I immersive audio

Technologies de l'information — Représentation codée de média immersifs —

Partie 4: Audio immersif MPEG-I

Reference number
ISO/IEC 23090-4:2025(en)

© ISO/IEC 2025



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2025

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Foreword vi
Introduction vii
1 Scope 1
2 Normative references 1
3 Terms, definitions and abbreviated terms 1
3.1 Terms and definitions 1
3.2 Mnemonics 5
3.3 Abbreviated terms 6
4 Overview 7
5 MPEG-I immersive audio transport 10
5.1 Overview 10
5.2 Definitions 11
5.3 MHAS syntax 11
5.3.1 Audio stream 11
5.3.2 Audio stream packet 12
5.4 Semantics 21
6 MPEG-I Immersive audio renderer 31
6.1 Definitions 31
6.2 Syntax 31
6.2.1 General 31
6.2.2 Generic codebook 31
6.2.3 Directivity payloads syntax 33
6.2.4 Diffraction payload syntax 36
6.2.5 Voxel payload syntax 42
6.2.6 Early reflection payload syntax 47
6.2.7 Portal payload syntax 50
6.2.8 Reverberation payload syntax 52
6.2.9 Audio plus payload syntax 54
6.2.10 Dispersion payload syntax 54
6.2.11 Scene plus payload syntax 54
6.2.12 Airflow payload syntax 71
6.2.13 Granular payload syntax 72
6.2.14 RasterMap payload syntax 75
6.2.15 Support elements 76
6.3 Data structure 82
6.3.1 General 82
6.3.2 Renderer payloads data structure 82
6.3.3 Generic codebook 134
6.4 Renderer framework 134
6.4.1 Control workflow 134
6.4.2 Rendering workflow 146
6.5 Geometry data decompression 159
6.5.1 General 159
6.5.2 Metadata extraction 159
6.5.3 Geometry 160
6.5.4 Materials 163
6.6 Renderer stages 164
6.6.1 Effect activator 164
6.6.2 Acoustic environment assignment 165

© ISO/IEC 2025 – All rights reserved

6.6.3 Granular synthesis 167
6.6.4 Reverberation 181
6.6.5 Portals 242
6.6.6 Early reflections 256
6.6.7 Airflow simulation 270
6.6.8 DiscoverSESS 277
6.6.9 Occlusion 279
6.6.10 Diffraction 284
6.6.11 Voxel-based occlusion and diffraction 301
6.6.12 Multi-Path voxel-based diffraction with RasterMaps 326
6.6.13 Voxel-based early reflections 332
6.6.14 Metadata culling 340
6.6.15 Heterogeneous extent 345
6.6.16 Directivity 374
6.6.17 Distance 380
6.6.18 Directional focus 390
6.6.19 Consolidation of render items 391
6.6.20 Equalizer (EQ) 398
6.6.21 Low-complexity early reflections (LC-ERs) 399
6.6.22 Fade 408
6.6.23 Single point higher order ambisonics (SP-HOA) 411
6.6.24 Homogeneous extent 416
6.6.25 Panner 421
6.6.26 Multi-point higher order ambisonics (MP-HOA) 428
6.6.27 Low-complexity MP-HOA 468
6.7 Spatializer 475
6.7.1 Binaural spatializer 475
6.7.2 Adaptive loudspeaker rendering 494
6.8 Limiter 528
6.8.1 General 528
6.8.2 Data elements and variables 528
6.8.3 Description 528
6.9 Interface for audio utilization information 530
6.9.1 General 530
6.9.2 Syntax and semantics of an interface for renderer audio utilization 530
Annex A (normative) Tables and additional algorithm details 531
A.1 Panner default output positions 531
A.2 Adaptive loudspeaker rendering calibration guide 531
A.3 RIR analysis: loudspeaker source directivity factor 535
A.4 Default acoustic environment presets 535
A.5 VR filter design initialization vector 541
A.6 Octave band neural network parameters 542
A.7 Third-octave band neural network parameters 545
A.8 Third-octave GEQ design filter bandwidths 574
A.9 Constants for feedback matrix calculation 574
A.10 Dispersion filter coefficient template 575
A.11 freqVec(b) - STFT band centre frequencies 580
A.12 Closest centre frequency bin for each one-third octave band frequency 581

© ISO/IEC 2025 – All rights reserved

A.13 EQbin ... 581
A.14 Fast convolution ... 582
A.14.1 Uniformly partitioned overlap-save convolution ... 582
A.14.2 Fast stereo convolution ... 583
A.15 Support element lookup tables ... 583
A.16 Airflow default frequency profiles ... 593
A.17 Reverberation extent mesh definitions ... 605
A.18 Headphone equalization preset responses ... 607
A.19 Listener voice default directivity pattern ... 607
A.20 Portal LoS data decoder tables ... 608
A.21 Reverberator output directions ... 609
A.22 Variable delay line anti-aliasing IIR lowpass filters ... 612
Annex B (informative) Encoder, interfaces and feature guidance ... 613
B.1 Encoder overview ... 613
B.2 Encoder modules ... 614
B.2.1 Scene configuration parameters ... 614
B.2.2 Audio plus metadata creation ... 614
B.2.3 Reverberation parametrization ... 614
B.2.4 Default acoustic environment (Default AE) ... 614
B.2.5 Low complexity early reflection parametrization ... 615
B.2.6 Portal creation in implicit portal mode ... 617
B.2.6.1 Creation of the geometry of the portal ... 617
B.2.6.2 Identification of the connection state between two portals ... 617
B.2.6.3 Creation of the portal struct containing all its metadata to be encoded ... 618
B.2.7 Line-of-sight data creation in explicit portal mode ... 619
B.2.8 Source/geometry staticity analysis ... 619
B.2.9 Diffraction edges and paths analysis ... 620
B.2.10 Early reflection surfaces and sequences analysis ... 620
B.2.11 Module data collection ... 620
B.2.12 Module data serialization ... 620
B.3 Listener space description format (LSI) ... 620
B.4 Encoder input format (EIF) ... 620
B.5 Accessibility user interface ... 620
B.6 Guidance on own voice usage – influence of system delay ... 621
Bibliography ... 624

© ISO/IEC 2025 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

A list of all parts in the ISO/IEC 23090 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

© ISO/IEC 2025 – All rights reserved

Introduction

0.1 General

MPEG-I Immersive audio reproduction with six degrees of freedom (6DoF) movement of the listener in an audio scene enables the experience of virtual acoustics in a Virtual Reality (VR) or Augmented Reality (AR) simulation. Audio effects and phenomena known from real-world acoustics like, for example, localization, distance attenuation, reflections, reverberation, occlusion, diffraction and the Doppler effect are modelled by a renderer that is controlled through metadata transmitted in a bitstream with additional input of interactive listener position data.

Along with other parts of MPEG-I (i.e. ISO/IEC 23090-12, “Immersive Video”, ISO/IEC 23090-5, “Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression” and ISO/IEC 23090-2, “Systems Support”), the ISO/IEC 23090 series of standards supports a complete audio-visual VR or AR presentation in which the user can navigate and interact with the simulated environment using 6DoF, that being spatial navigation (x, y, z) and user head orientation (yaw, pitch, roll).

While VR presentations impart the feeling that the user is actually present in the virtual world, AR enables the enrichment of the real world by virtual elements that are perceived seamlessly as being part of the real world. The user can interact with the virtual scene or virtual elements and, in response, cause sounds that are perceived as realistic and matching the users’ experience in the real world.

This document provides means for rendering a real-time interactive audio presentation while permitting the user to have 6DoF movement. It defines metadata to support this rendering and a bitstream syntax that enables efficient storage and streaming of the MPEG-I Immersive Audio content.

0.2 Typesetting of variables

For improved text readability, algorithmic variables are distinguished from descriptive text by their dedicated typesetting. Throughout the text, variables that are input directly from the bitstream are typeset in boldface letters. In the bitstream syntax the use of a boldface variable represents reading bits from the bitstream and converting to an appropriate type (based on the indicated mnemonics and number of bits), and assigning the resulting value to the boldfaced variable. Variables that are derived by computation are typeset in pseudo-code font or italics.

© ISO/IEC 2025 – All rights reserved

Information technology — Coded representation of immersive media —

Part 4:
MPEG-I immersive audio

1 Scope

This document specifies technology that supports the real-time interactive rendering of an immersive virtual or augmented reality audio presentation while permitting the user to have 6DoF movement in the audio scene. It defines metadata to support this rendering and a bitstream syntax that enables efficient storage and streaming of immersive audio content.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 266, *Acoustics — Preferred frequencies*

ISO/IEC 23008-3, *High efficiency coding and media delivery in heterogeneous environments — Part 3: 3D audio*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1.1
audio processing step
execution of DSP audio processing for one block of PCM audio at audio block rate

3.1.2
auralization
PCM waveform synthesis of immersive audio effects

3.1.3
audio element
audio objects, channels or HOA signals with associated MPEG-I 6DoF metadata and MPEG-H 3D Audio metadata if appropriate

© ISO/IEC 2025 – All rights reserved

3.1.4
audio scene
all audio elements, acoustic elements and Acoustic environments needed to render the sound in the scene

3.1.5
control workflow
all components of the renderer that are not part of the rendering workflow

Note 1 to entry: For example, the scene controller and all related components belong to the control workflow.

3.1.6
cover
frequency dependant quasi-uniform grid on the surface of a unit sphere whereupon each point has an associated scalar value in dB

3.1.7
conditional update
scene update initiated by the renderer when a certain condition is TRUE

3.1.8
Doppler effect
pitch change of sound perceived when distance between sound source and listener changes

3.19
dynamic update
update triggered by external entity that includes the values of the attributes to be updated

3.1.10
effective spatial extent
part of an extent that is acoustically relevant given a specific listening position relative to the extent

3.1.11
exterior representation
representation based on a source-centric format used for rendering when the listener is outside the extent of the audio element

3.1.12
frame length
length of a PCM audio block in samples (B) to be processed in one audio processing step

3.1.13
HOA group
recording or synthesis of an audio scene with one or more HOA sources

3.1.14
interior representation
representation based on a listener-centric format used for rendering when the listener is inside the extent of the audio element

3.1.15
listener centric format
audio format for representing an audio element that is rendered to reproduce a sound field around the listener

Note 1 to entry: HOA or channel-based formats like, e.g., 5.1 channels or 7.4.1 channels.

3.1.16
limiter
processing block that prevents clipping of the multi-channel output signal of the spatializer

© ISO/IEC 2025 – All rights reserved

3.1.17
location
3D location of an object in space in Cartesian coordinates

3.1.18
metadata
all input and state parameters that are used to calculate the acoustic events of a virtual environment

3.1.19
MPEG-H 3DA decoder
MPEG-H 3D Audio Low Complexity (LC) Profile decoder

Note 1 to entry: The decoder receives as input an MPEG-H 3D Audio LC Profile MHAS stream and provides as output decoded PCM audio together with all metadata available in the MHAS packets whereas decoded PCM audio contains channels, objects and reconstructed HOA as described in ISO/IEC 23008-3:2018, Subclause 17.10.

3.1.20
orientation
3DoF rotation of an object in Tait-Bryan angles (yaw, pitch, roll)

3.1.21
portal
model of spatial transfer regions where there is an acoustic link that can transfer acoustic energy from one AE into another AE

3.1.22
position
6DoF location and orientation

3.1.23
primary ray
rays used by the system to explore the scene geometry, the results of casting primary rays stored and addressed individually by unique ray IDs

3.1.24
primary RI
RI that is directly derived from an audio element in the scene

3.1.25
renderer
entire software specified in this document

3.1.26
renderer pipeline
collection of stages, which sequentially perform audio and metadata processing

3.1.27
rendering workflow
control of renderer pipeline and spatializer

3.2.28
render item
RI
any audio element in the renderer pipeline

3.2.29
RI type
type associated to RI, denoting its relevance for pipeline stages

© ISO/IEC 2025 – All rights reserved

3.2.30
sample frequency
fs
number of discrete audio PCM samples per second

3.2.31
scene
aggregate of all entities that represent all acoustically relevant data of a virtual environment modeled in the renderer through metadata

3.2.32
scene controller
renderer processing block that holds and updates the scene state

3.2.33
scene object
entity in the scene

Note 1 to entry: For example, geometry, audio element or listener.

3.2.34
scene state
reflection of the current state of all 6DoF metadata of the scene

3.2.35
secondary ray
auxiliary ray used to refine the results based on casting a primary ray

Note 1 to entry: The results of casting secondary rays are aggregated and stored under the respective primary ray IDs.

3.2.36
secondary RI
RI representing additional aspects of its primary RI

Note 1 to entry: Secondary RIs are e.g. mirror sources for modelling early reflections, sources at diffracting edges for modelling diffraction, extended sources corresponding to ray bundles having the same occlusion material lists.

3.2.37
source-centric format
audio format for representing an audio element that is rendered such that all direct sound from the audio element appears to radiate from a bounded region in space that does not include the listener

3.2.38
spatializer
processing block situated directly after the renderer pipeline to produce a multi-channel output signal for a specific playback method (e.g. headphones or loudspeakers)

3.2.39
spatially-heterogeneous audio element
audio element which has an extent and a source signal with more than one channel

3.2.40
stage
processing block in the renderer pipeline that addresses a dedicated rendering aspect

3.2.41
teleport
instant change of listener position triggered by user interaction within VR environment

© ISO/IEC 2025 – All rights reserved

3.2.42
timed update
scene update executed by the renderer once at a fixed predefined time

3.2.43
triggered update
scene update triggered from an external entity and executed by the renderer immediately after receiving the trigger

3.2.44
update step
calculation of updated scene control data at a control rate

3.2.45
user
listener whose position in the scene is input to the renderer

3.2.46
voxel
geometry element defining a volume on a regular three-dimensional grid

3.2.47
voxel sub-scene
scene containing scene geometry representation based on voxels and including one or more independent voxel sub-scenes

3.2.48
voxel sub-scene update
scene update executed by the renderer and applied to a particular available voxel sub-scene including the currently rendered one

3.2 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bitstream payload.

3.2.1
bslbf
bit string, left bit first, where “left” is the order in which bit strings are written in ISO/IEC 14496 (all parts)

Note 1 to entry: Bit strings are written as a string of 1s and 0s within single quote marks, for example '1000 0001'. Blanks within a bit string are for ease of reading and have no significance.

3.2.2
uimsbf
unsigned integer, most significant bit first

3.2.3
vlclbf
variable length code, left bit first, where “left” refers to the order in which the variable length codes are written

3.2.4
tcimsbf
two’s complement integer, most significant (sign) bit first

3.2.5
cstring
UTF-8 string

© ISO/IEC 2025 – All rights reserved

3.2.6
float
IEEE 754 floating single point precision number

3.3 Abbreviated terms

6DoF	6-degrees-of-freedom
AE	Acoustic environment
AR	augmented reality
CoRI	consolidation of render items
CV	configuration variable
DOA	direction-of-arrival
DSR	diffuse-to-source energy ratio
EIF	encoder input format
EP	extent processor
EQ	equalizer
ER	early reflections
ES	extended source
ESD	equivalent spatial domain
FDN	feedback delay network
HRIR	head-related impulse responses
HRTF	head related transfer function
IACC	interaural cross correlation
IALD	interaural level differences
IAPD	interaural phase differences
ID	identifier of an entity in the scene
IIR	infinite impulse response
LC-ER	low complexity early reflections
LCS	listener coordinate system
LoS	line-of-sight
LSI	listener space information
MP-HOA	multi point higher order ambisonics
PCM	Pulse Code Modulation
RDR	reverberant-to-direct ratio
RT60	reverberation time for reverberation energy to drop 60 dB
RI	render item
RIR	room impulse response
SESS	spatially extended sound sources
SN3D	data exchange format for ambisonics (schmidt semi-normalisation)
SO	scene object
SOFA	spatially oriented format for acoustics

© ISO/IEC 2025 – All rights reserved

SOS	second-order-section
SP-HOA	single point higher order ambisonics
SRIR	spatial room impulse response
VBAP	vector-base-amplitude-panning
VDL	variable delay line
VR	virtual reality

4 Overview

The renderer operates with a global sampling frequency of fs ∈ [32kHz, 44.1kHz, 48kHz]. Input PCM audio data with other sampling frequencies must be resampled to the global sample frequency before processing. The granular synthesis databases with audio grains have to be authored at the global sampling frequency, fs, before their usage at the renderer. A block diagram of the MPEG-I architecture overview is shown in Figure 2. The overview illustrates how the renderer is connected to external units like MPEG-H 3DA coded Audio Element bitstreams, the metadata MPEG-I bitstream and other interfaces. The MPEG-H 3DA coded Audio Elements are decoded by the MPEG-H 3DA Decoder. The MPEG-H 3DA Decoder shall be configured to decode the MPEG-H 3DA coded audio content but skip the rendering of the decoded audio elements in accordance with ISO/IEC 23008-3:2022, Subclause 17.10, where Subclause 17.10 provides interface for objects, Subclause 17.10.4 provides interface for channels and Subclause 17.10.5 provides interface for HOA. These decoded PCM samples are provided as input to the MPEG-I immersive audio renderer. These clauses describe the PCM output, included processing to be applied to the PCM outputs while skipping the peak limiter, and interface for accompanying MPEG-H 3DA pass through metadata (e.g., loudness). All Audio Elements (channels, objects and HOA) that are to be input into the renderer have a counterpart in the MPEG-I Immersive audio standard, namely so-called source types (see Figure 1): Objects are represented as object sources, equipped with many VR/AR specific properties. Channels are channel sources that are played back in the virtual world through a virtual loudspeaker setup. Finally, HOA sources can be rendered into the virtual world in two different ways: rendering one or more HOA sources individually, where the rendering has three degrees of freedom (user orientation only) for listening positions within a possible associated spatial extent of the HOA source and six degrees of freedom for listening positions outside of the spatial extent, or rendering one or more HOA sources as a group, with six degrees of freedom. For all three paradigms, encoded waveforms can preferably be carried over from MPEG-H 3D audio to MPEG-I Immersive audio directly without the need for any re-encoding and associated loss in quality.

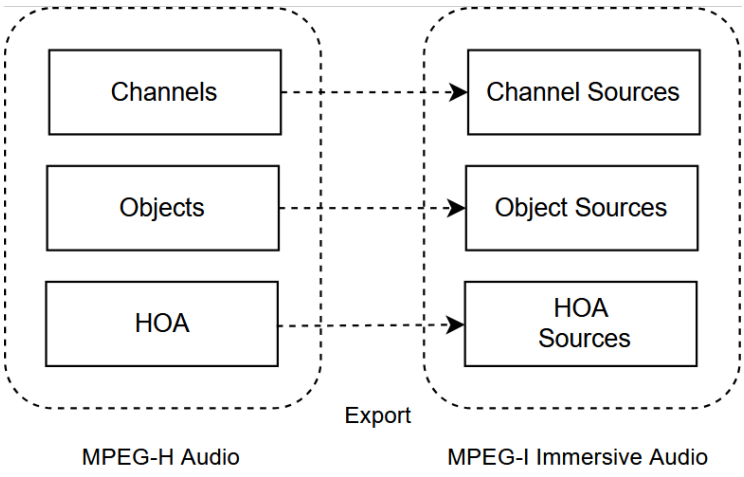


Figure 1 — Correspondences in MPEG-H audio and MPEG-I immersive audio

The decoded audio is subsequently rendered together with the MPEG-I bitstream, which is described in Clause 5. The MPEG-I bitstream carries an encoded representation of the Audio Scene description and other metadata used by the renderer. In addition, the renderer has access to listening space information,

© ISO/IEC 2025 – All rights reserved