



International
Standard

ISO/IEC 23090-4

**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

**First edition
2025-11**

iteh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO/IEC 23090-4:2025

<https://standards.iteh.ai/catalog/standards/iso/fde38ce1-762e-4af1-bdb0-baa8a0508d0d/iso-iec-23090-4-2025>

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO/IEC 23090-4:2025

<https://standards.iteh.ai/catalog/standards/iso/fde38ce1-762e-4af1-bdb0-baa8a0508d0d/iso-iec-23090-4-2025>



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

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

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

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.

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.