



**International
Standard**

ISO/IEC 23008-3

**Information technology — High
efficiency coding and media
delivery in heterogeneous
environments —**

**Part 3:
3D audio**

*Technologies de l'information — Codage à haute efficacité et
livraison des médias dans des environnements hétérogènes —*

Partie 3: Audio 3D

<https://standards.iteh.ai/catalog/standards/iso/c55536a5-4625-411a-9db3-9865f8d277d8/iso-iec-23008-3-2026>

**Fourth edition
2026-02**

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

ISO/IEC 23008-3:2026

<https://standards.itih.ai/catalog/standards/iso/c55536a5-4625-411a-9db3-9865f8d277d8/iso-iec-23008-3-2026>



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2026

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

Page

Foreword.....	xii
Introduction.....	xii
1 Scope	1
2 Normative references	1
3 Terms, definitions, symbols, abbreviated terms and conventions.....	1
3.1 Terms and definitions.....	1
3.2 Symbols, abbreviated terms and conventions	2
3.2.1 Symbols and abbreviated terms.....	2
3.2.2 Conventions.....	2
4 Technical overview.....	2
4.1 Decoder block diagram	2
4.2 Overview over the codec building blocks	3
4.3 Efficient combination of decoder processing blocks in the time domain and QMF domain.....	6
4.4 Rule set for determining processing domains	9
4.4.1 Audio core codec processing domain.....	9
4.4.2 Mixing	10
4.4.3 DRC-1 Operation domains (DRC in rendering context)	10
4.4.4 Audio core codec interface domain to rendering	10
4.4.5 Rendering context	10
4.4.6 Post-processing context	11
4.4.7 End-of-chain context.....	11
4.5 Sample rate converter	11
4.6 Decoder delay	11
4.7 Contribution mode of MPEG-H 3D audio.....	12
4.8 MPEG-H 3D audio profiles and levels	12
4.8.1 General	12
4.8.2 Profiles	13
5 MPEG-H 3D audio core decoder.....	27
5.1 Definitions.....	27
5.1.1 Joint stereo.....	27
5.1.2 MPEG surround based stereo (MPS 212)	28
5.2 Syntax.....	28
5.2.1 General	28
5.2.2 Decoder configuration	28
5.2.3 MPEG-H 3D audio core bitstream payloads	51
5.3 Data structure	72
5.3.1 General	72
5.3.2 General configuration data elements.....	72
5.3.3 Loudspeaker configuration data elements	75
5.3.4 Core decoder configuration data elements	77
5.3.5 Downmix matrix data elements	81
5.3.6 HOA rendering matrix data elements	84
5.3.7 Signal group information elements	87
5.3.8 Low frequency enhancement (LFE) channel element, mpegh3daLfeElement().....	87
5.3.9 Compatible profile and levels sets.....	88
5.4 Configuration element descriptions	88
5.4.1 General	88
5.4.2 Downmix configuration.....	88

5.4.3	HOA rendering matrix configuration	94
5.5	Tool descriptions	98
5.5.1	General.....	98
5.5.2	Quad channel element.....	98
5.5.3	Transform splitting	100
5.5.4	MPEG surround for mono to stereo upmixing.....	107
5.5.5	Enhanced noise filling	110
5.5.6	Audio pre-roll	134
5.5.7	Fullband LPD	137
5.5.8	Time-domain bandwidth extension.....	148
5.5.9	LPD stereo coding	161
5.5.10	Multichannel coding tool.....	169
5.5.11	Filterbank and block switching	179
5.5.12	Frequency domain prediction.....	180
5.5.13	Long-term postfilter.....	183
5.5.14	Tonal component coding.....	188
5.5.15	Internal channel on MPS212 for low complexity format conversion	198
5.5.16	High resolution envelope processing (HREP) tool.....	210
5.6	Buffer requirements	216
5.6.1	Minimum decoder input buffer	216
5.6.2	Bit reservoir	216
5.6.3	Maximum bit rate.....	217
5.7	Stream access point requirements and inter-frame dependency.....	217
6	Dynamic range control and loudness processing	218
6.1	General.....	218
6.2	Description	218
6.3	Syntax	219
6.3.1	Loudness metadata	219
6.3.2	Dynamic range control metadata.....	219
6.3.3	Data elements.....	220
6.4	Decoding process	222
6.4.1	General.....	222
6.4.2	Dynamic range control.....	224
6.4.3	Usage of downmixId in MPEG-H	224
6.4.4	DRC set selection process	225
6.4.5	DRC-1 for SAOC 3D Content	227
6.4.6	DRC-1 for HOA content.....	228
6.4.7	Loudness normalization.....	229
6.4.8	Peak limiter.....	229
6.4.9	Time-synchronization of DRC gains	230
6.4.10	Default parameters	230
7	Object metadata decoding	230
7.1	General.....	230
7.2	Description.....	230
7.3	Syntax	231
7.3.1	Object metadata configuration	231
7.3.2	Top level object metadata syntax	232
7.3.3	Subsidiary payloads for efficient object metadata decoding.....	233
7.3.4	Subsidiary payloads for object metadata decoding with low delay	238
7.3.5	Enhanced object metadata configuration	244
7.4	Data structure	247
7.4.1	Definition of ObjectMetadataConfig() payloads	247
7.4.2	Efficient object metadata decoding.....	247
7.4.3	Object metadata decoding with low delay	255

7.4.4	Enhanced object metadata	260
8	Object rendering.....	263
8.1	Description.....	263
8.2	Terms and definitions.....	263
8.3	Input data	264
8.4	Processing	265
8.4.1	General remark	265
8.4.2	Imaginary loudspeakers	265
8.4.3	Dividing the loudspeaker setup into a triangle mesh	266
8.4.4	Rendering algorithm	268
9	SAOC 3D	272
9.1	Description.....	272
9.2	Definitions.....	272
9.3	Delay and synchronization.....	274
9.4	Syntax.....	274
9.4.1	Payloads for SAOC 3D	274
9.4.2	Definition of SAOC 3D payloads.....	278
9.5	SAOC 3D processing	280
9.5.1	Compressed data stream decoding and dequantization of SAOC 3D data	280
9.5.2	Time/frequency transforms.....	280
9.5.3	Signals and parameters	281
9.5.4	SAOC 3D decoding.....	283
9.5.5	Dual mode	288
10	Generic loudspeaker rendering/format conversion	288
10.1	Description.....	288
10.2	Definitions.....	290
10.2.1	General remarks	290
10.2.2	Variable definitions	290
10.3	Processing	290
10.3.1	Application of transmitted downmix matrices.....	290
10.3.2	Application of transmitted equalizer settings	295
10.3.3	Downmix processing involving multiple channel groups.....	295
10.3.4	Initialization of the format converter	296
10.3.5	Audio signal processing.....	312
11	Immersive loudspeaker rendering/format conversion.....	318
11.1	Description.....	318
11.2	Syntax.....	320
11.3	Definitions.....	320
11.3.1	General remarks	320
11.3.2	Variable definitions	321
11.4	Processing	322
11.4.1	Initialization of the format converter	322
11.4.2	Audio signal processing.....	364
12	Higher order ambisonics (HOA).....	372
12.1	Technical overview	372
12.1.1	Block diagram	372
12.1.2	Overview of the decoder tools	373
12.2	Syntax.....	374
12.2.1	Configuration of HOA elements	374
12.2.2	Payloads of HOA elements.....	378
12.3	Data structure	391
12.3.1	Definitions of HOA Config	391

12.3.2	Syntax of getSubbandBandwidths()	395
12.3.3	Definitions of HOA payload	396
12.4	HOA tool description	403
12.4.1	HOA frame converter	403
12.4.2	Spatial HOA decoding	419
12.4.3	HOA renderer	448
12.4.4	Layered coding for HOA	457
13	Binaural renderer	460
13.1	General	460
13.2	Frequency-domain binaural renderer	460
13.2.1	General	460
13.2.2	Definitions	462
13.2.3	Parameterization of binaural room impulse responses	466
13.2.4	Frequency-domain binaural processing	478
13.3	Time-domain binaural renderer	485
13.3.1	General	485
13.3.2	Definitions	486
13.3.3	Parameterization of binaural room impulse responses	488
13.3.4	Time-domain binaural processing	492
14	MPEG-H 3D audio stream (MHAS)	493
14.1	Overview	493
14.2	Syntax	494
14.2.1	Main MHAS syntax elements	494
14.2.2	Subsidiary MHAS syntax elements	496
14.3	Semantics	496
14.3.1	mpeghAudioStreamPacket()	496
14.3.2	MHASPacketPayload()	497
14.3.3	Subsidiary MHAS packets	499
14.4	Description of MHASPacketTypes	499
14.4.1	PACTYP_FILLDATA	499
14.4.2	PACTYP_MPEGH3DACFG	499
14.4.3	PACTYP_MPEGH3DAFRAME	499
14.4.4	PACTYP_SYNC	500
14.4.5	PACTYP_SYNGAP	500
14.4.6	PACTYP_MARKER	500
14.4.7	PACTYP_CRC16 and PACTYP_CRC32	501
14.4.8	PACTYP_DESCRIPTOR	501
14.4.9	PACTYP_USERINTERACTION	501
14.4.10	PACTYP_LOUDNESS_DRC	501
14.4.11	PACTYP_BUFFERINFO	502
14.4.12	PACTYP_GLOBAL_CRC16 and PACTYP_GLOBAL_CRC32	502
14.4.13	PACTYP_AUDIOTRUNCATION	502
14.4.14	PACTYP_AUDIOSCENEINFO	503
14.4.15	PACTYP_EARCON	503
14.4.16	PACTYP_PCMCONFIG	504
14.4.17	PACTYP_PCMDATA	504
14.4.18	PACTYP_LOUDNESS	504
14.4.19	MHASPacketType specific requirements for MHASPacketLabel	504
14.5	Application examples	505
14.5.1	Light-weighted broadcast	505
14.5.2	MPEG-2 transport stream	506
14.5.3	CRC error detection	506
14.5.4	Audio sample truncation	507
14.6	Multi-stream delivery and interface	507

14.7	Carriage of generic data	510
14.7.1	Syntax.....	510
14.7.2	Semantics.....	511
14.7.3	Processing at the MPEG-H 3D audio decoder	512
15	Metadata audio elements (MAE)	512
15.1	General	512
15.2	Syntax.....	513
15.3	Semantics.....	522
15.4	Definition of mae_metaDataElementIDs	534
15.5	Loudness compensation after gain interactivity.....	535
16	Loudspeaker distance compensation.....	537
17	Interfaces to the MPEG-H 3D audio decoder.....	538
17.1	General	538
17.2	Interface for local setup information	538
17.2.1	General	538
17.2.2	WIRE output	538
17.2.3	Syntax for local setup information	539
17.2.4	Semantics for local setup information	539
17.3	Interface for local loudspeaker setup and rendering.....	540
17.3.1	General	540
17.3.2	Syntax for local loudspeaker signalling.....	540
17.3.3	Semantics for local loudspeaker signalling.....	541
17.4	Interface for binaural room impulse responses (BRIRs).....	542
17.4.1	General	542
17.4.2	Syntax of binaural renderer interface	542
17.4.3	Semantics.....	547
17.5	Interface for local screen size information	551
17.5.1	General	551
17.5.2	Syntax.....	551
17.5.3	Semantics.....	551
17.6	Interface for signaling of local zoom area.....	552
17.6.1	General	552
17.6.2	Syntax.....	552
17.6.3	Semantics.....	552
17.7	Interface for user interaction.....	553
17.7.1	General	553
17.7.2	Definition of user interaction categories.....	553
17.7.3	Definition of an interface for user interaction.....	554
17.7.4	Syntax of interaction interface	555
17.7.5	Semantics of interaction interface	556
17.8	Interface for loudness normalization and dynamic range control (DRC).....	558
17.9	Interface for scene displacement data.....	558
17.9.1	General	558
17.9.2	Definition of an interface for scene-displacement data	559
17.9.3	Syntax of the scene displacement interface	560
17.9.4	Semantics of the scene displacement interface	560
17.10	Interfaces for channel-based, object-based, and HOA metadata and audio data.....	561
17.10.1	General	561
17.10.2	Expectations on external renderers	561
17.10.3	Object-based metadata and audio data (object output interface)	561
17.10.4	Channel-based metadata and audio data	569
17.10.5	HOA metadata and audio data	573
17.10.6	Audio PCM data	577