

# ISO/IEC 23008-3:2026-02 (E)

## Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio

---

<b>Contents</b>		<b>Page</b>
Foreword.....		xii
Introduction.....		xii
<b>1</b>	<b>Scope .....</b>	<b>1</b>
<b>2</b>	<b>Normative references .....</b>	<b>1</b>
<b>3</b>	<b>Terms, definitions, symbols, abbreviated terms and conventions.....</b>	<b>1</b>
<b>3.1</b>	<b>Terms and definitions.....</b>	<b>1</b>
<b>3.2</b>	<b>Symbols, abbreviated terms and conventions .....</b>	<b>2</b>
<b>3.2.1</b>	<b>Symbols and abbreviated terms.....</b>	<b>2</b>
<b>3.2.2</b>	<b>Conventions .....</b>	<b>2</b>
<b>4</b>	<b>Technical overview .....</b>	<b>2</b>
<b>4.1</b>	<b>Decoder block diagram .....</b>	<b>2</b>
<b>4.2</b>	<b>Overview over the codec building blocks .....</b>	<b>3</b>
<b>4.3</b>	<b>Efficient combination of decoder processing blocks in the time domain and QMF domain.....</b>	<b>6</b>
<b>4.4</b>	<b>Rule set for determining processing domains .....</b>	<b>9</b>
<b>4.4.1</b>	<b>Audio core codec processing domain .....</b>	<b>9</b>
<b>4.4.2</b>	<b>Mixing .....</b>	<b>10</b>
<b>4.4.3</b>	<b>DRC-1 Operation domains (DRC in rendering context).....</b>	<b>10</b>
<b>4.4.4</b>	<b>Audio core codec interface domain to rendering .....</b>	<b>10</b>
<b>4.4.5</b>	<b>Rendering context.....</b>	<b>10</b>
<b>4.4.6</b>	<b>Post-processing context.....</b>	<b>11</b>
<b>4.4.7</b>	<b>End-of-chain context.....</b>	<b>11</b>
<b>4.5</b>	<b>Sample rate converter .....</b>	<b>11</b>
<b>4.6</b>	<b>Decoder delay .....</b>	<b>11</b>
<b>4.7</b>	<b>Contribution mode of MPEG-H 3D audio.....</b>	<b>12</b>
<b>4.8</b>	<b>MPEG-H 3D audio profiles and levels .....</b>	<b>12</b>
<b>4.8.1</b>	<b>General .....</b>	<b>12</b>
<b>4.8.2</b>	<b>Profiles .....</b>	<b>13</b>
<b>5</b>	<b>MPEG-H 3D audio core decoder.....</b>	<b>27</b>
<b>5.1</b>	<b>Definitions.....</b>	<b>27</b>
<b>5.1.1</b>	<b>Joint stereo.....</b>	<b>27</b>
<b>5.1.2</b>	<b>MPEG surround based stereo (MPS 212) .....</b>	<b>28</b>
<b>5.2</b>	<b>Syntax.....</b>	<b>28</b>
<b>5.2.1</b>	<b>General .....</b>	<b>28</b>
<b>5.2.2</b>	<b>Decoder configuration .....</b>	<b>28</b>
<b>5.2.3</b>	<b>MPEG-H 3D audio core bitstream payloads .....</b>	<b>51</b>
<b>5.3</b>	<b>Data structure .....</b>	<b>72</b>
<b>5.3.1</b>	<b>General .....</b>	<b>72</b>
<b>5.3.2</b>	<b>General configuration data elements.....</b>	<b>72</b>
<b>5.3.3</b>	<b>Loudspeaker configuration data elements .....</b>	<b>75</b>
<b>5.3.4</b>	<b>Core decoder configuration data elements .....</b>	<b>77</b>
<b>5.3.5</b>	<b>Downmix matrix data elements .....</b>	<b>81</b>
<b>5.3.6</b>	<b>HOA rendering matrix data elements .....</b>	<b>84</b>
<b>5.3.7</b>	<b>Signal group information elements .....</b>	<b>87</b>
<b>5.3.8</b>	<b>Low frequency enhancement (LFE) channel element, mpegH3daLfeElement().....</b>	<b>87</b>

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_SYNCGAP	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

17.11	Interface for positional scene displacement data .....	577
17.11.1	General .....	577
17.11.2	Syntax of the positional scene displacement interface.....	578
17.11.3	Semantics of the positional scene displacement interface .....	578
17.11.4	Processing.....	578
18	Application and processing of local setup information and interaction data and scene displacement data .....	579
18.1	Element metadata preprocessing .....	579
18.1.1	General information .....	579
18.1.2	Initialization .....	580
18.1.3	Processing loop.....	581
18.1.4	Element routing.....	585
18.2	Interactivity limitations and restrictions .....	585
18.2.1	General information .....	585
18.2.2	WIRE interactivity .....	585
18.2.3	Position interactivity.....	586
18.2.4	Screen-related element remapping and object remapping for zooming .....	586
18.2.5	Closest loudspeaker payout.....	587
18.3	Screen-related element remapping .....	587
18.4	Screen-related adaptation and zooming for higher order ambisonics (HOA).....	590
18.5	Object remapping for zooming .....	592
18.6	Determination of the closest loudspeaker .....	594
18.7	Determination of a list of loudspeakers for conditioned closest loudspeaker playback .....	594
18.8	Processing of scene displacement angles for channels and objects (CO).....	596
18.9	Processing of scene displacement angles for scene-based content (HOA) .....	598
18.10	Determination of a reduced reproduction layout based on excluded sectors .....	600
18.11	Diffuseness rendering.....	601
19	MPEG-H 3D audio profile definition.....	603
20	Carriage of MPEG-H 3D audio in ISO base media file format .....	603
20.1	General.....	603
20.2	Random access and stream access .....	603
20.3	Overview of new box structures.....	603
20.4	MHA decoder configuration record.....	604
20.4.1	Definition .....	604
20.4.2	Syntax.....	604
20.4.3	Semantics.....	604
20.5	MPEG-H audio sample entry .....	604
20.5.1	Definition .....	604
20.5.2	Syntax.....	605
20.5.3	Semantics.....	605
20.6	MPEG-H audio MHAS sample entry .....	606
20.6.1	Definition .....	606
20.6.2	Syntax.....	607
20.7	MHA dynamic range control and loudness .....	607
20.7.1	Definition .....	607
20.7.2	Syntax.....	607
20.7.3	Semantics.....	609
20.8	MHA multi-stream signalling.....	610
20.8.1	Definition .....	610
20.8.2	Syntax.....	610
20.8.3	Semantics.....	610
20.9	Audio scene information .....	611

20.9.1	MHA group definition.....	611
20.9.2	MHA switch group definition.....	613
20.9.3	MHA group preset definition.....	614
20.9.4	MHA group description text label.....	615
20.9.5	MHA scene information.....	617
20.10	Track references.....	618
20.11	MPEG-H Audio profile and level compatibility sets.....	618
20.11.1	Definition.....	618
20.11.2	Syntax.....	618
20.11.3	Semantics.....	618
21	Sub-parameters for the MIME type 'Codecs' parameter.....	619
21.1	General.....	619
21.2	'Codecs' parameter for MPEG-H 3D audio.....	619
22	Timing considerations and decoder behaviour.....	619
23	Multi-stream handling.....	619
23.1	Restrictions on extension payloads.....	619
24	Low complexity generic loudspeaker rendering/format conversion.....	621
24.1	Description.....	621
24.2	Definitions.....	622
24.2.1	General remarks.....	622
24.2.2	Variable definitions.....	623
24.3	Processing.....	623
24.3.1	Application of transmitted downmix matrices.....	623
24.3.2	Application of transmitted equalizer settings.....	628
24.3.3	Downmix processing involving multiple channel groups.....	628
24.3.4	Initialization of the format converter.....	629
24.3.5	Audio signal processing.....	646
25	Low complexity immersive loudspeaker rendering/format conversion.....	648
25.1	Description.....	648
25.2	Syntax.....	649
25.3	Definitions.....	650
25.3.1	General remarks.....	650
25.3.2	Variable definitions.....	650
25.4	Processing.....	651
25.4.1	Initialization of the format converter.....	651
25.4.2	Audio signal processing.....	695
26	MPEG surround.....	698
26.1	Technical overview.....	698
26.2	Syntax and data structure.....	699
26.3	Tool description.....	699
27	Production metadata decoding.....	699
27.1	General.....	699
27.1.1	Object distance coding.....	699
27.1.2	Direct headphone signalling.....	699
27.1.3	Reference distance coding.....	700
27.2	Syntax.....	701
27.2.1	Production metadata configuration.....	701
27.2.2	Production metadata frame.....	702
27.3	Semantics.....	703
27.3.1	Production metadata configuration.....	703
27.4	Decoding process.....	705

<b>28</b>	<b>Earcon metadata .....</b>	<b>705</b>
<b>28.1</b>	<b>General.....</b>	<b>705</b>
<b>28.2</b>	<b>Syntax.....</b>	<b>705</b>
<b>28.3</b>	<b>Semantics.....</b>	<b>707</b>
<b>28.4</b>	<b>Decoding process .....</b>	<b>711</b>
<b>Annex A (normative)</b>	<b>Tables for arithmetic decoding of IGF information .....</b>	<b>713</b>
<b>Annex B (normative)</b>	<b>SAOC 3D Decorrelator pre-mixing matrices.....</b>	<b>717</b>
<b>Annex C (informative)</b>	<b>Encoder tools.....</b>	<b>724</b>
<b>Annex D (normative)</b>	<b>Peak limiter for unguided clipping prevention .....</b>	<b>772</b>
<b>Annex E (normative)</b>	<b>Compact template downmix matrices .....</b>	<b>773</b>
<b>Annex F (normative)</b>	<b>HOA tables.....</b>	<b>774</b>
<b>Annex G (informative)</b>	<b>Low complexity HOA rendering.....</b>	<b>816</b>
<b>Annex H (informative)</b>	<b>Information on delay and complexity of time-domain binauralization .....</b>	<b>830</b>
<b>Annex I (informative)</b>	<b>Determination of a rotation matrix for processing of scene displacement data .....</b>	<b>835</b>
<b>Annex J (informative)</b>	<b>Decorrelation filtering for ‘diffuseness’ processing.....</b>	<b>837</b>
<b>Annex K (informative)</b>	<b>Distance and depth spread rendering .....</b>	<b>838</b>
<b>Annex L (informative)</b>	<b>HREP encoder description.....</b>	<b>840</b>
<b>Annex M (informative)</b>	<b>Screen-related adaptation of HOA content in complexity constrained implementations .....</b>	<b>844</b>
<b>Annex N (normative)</b>	<b>Retaining original file length with MPEG-H 3D audio.....</b>	<b>845</b>
<b>Annex O (normative)</b>	<b>Codebook tables used to de-quantize high band time domain bandwidth extension parameters.....</b>	<b>847</b>
<b>Annex P (informative)</b>	<b>Implementation and usage guidelines for signalling of profile and level compatibility sets .....</b>	<b>856</b>
<b>Bibliography .....</b>		<b>856</b>