This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

INTERNATIONAL ISO/IEC

Second edition
2019-02

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 medias dans des environnements hétérogènes —

Partie 3: Audio 3D

Reference number
ISO/IEC 23008-3:2019(E)

© ISO/IEC 2019

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

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
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

 © ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

Contents

Page

Foreword x

Introduction xii

1 Scope 1

2 Normative references 1

3 Terms, definitions, symbols, abbreviations and mnemonics 2
3.1 Terms, definitions, symbols and abbreviated terms 2
3.2 Mnemonics 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 9
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 10
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 12

5 MPEG-H 3D audio core decoder 22
5.1 Definitions 22
5.1.1 Joint stereo 22
5.1.2 MPEG surround based stereo (MPS 212) 22
5.2 Syntax 22
5.2.1 General 22
5.2.2 Decoder configuration 23
5.2.3 MPEG-H 3D audio core bitstream payloads 41
5.3 Data structure 60
5.3.1 General 60
5.3.2 General configuration data elements 61
5.3.3 Loudspeaker configuration data elements 63
5.3.4 Core decoder configuration data elements 65
5.3.5 Downmix matrix data elements 69
5.3.6 HOA rendering matrix data elements 72
5.3.7 Signal group information elements 74
5.3.8 Low frequency enhancement (LFE) channel element, mpegh3daLfeElement() 75
5.4 Configuration element descriptions 75
5.4.1 General 75
5.4.2 Downmix configuration 76
5.4.3 HOA rendering matrix configuration 81
5.5 Tool descriptions 86
5.5.1 General 86
5.5.2 Quad channel element 86
5.5.3 Transform splitting 88

© ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

5.5.4 MPEG surround for mono to stereo upmixing....95
5.5.5 Enhanced noise filling....97
5.5.6 Audio pre-roll....121
5.5.7 Fullband LPD....124
5.5.8 Time-domain bandwidth extension....135
5.5.9 LPD stereo coding....148
5.5.10 Multichannel coding tool....155
5.5.11 Filterbank and block switching....166
5.5.12 Frequency domain prediction....166
5.5.13 Long-term postfilter....169
5.5.14 Tonal component coding....175
5.5.15 Internal channel on MPS212 for low complexity format conversion....184
5.5.16 High resolution envelope processing (HREP) tool....196
5.6 Buffer requirements....202
5.6.1 Minimum decoder input buffer....202
5.6.2 Bit reservoir....203
5.6.3 Maximum bit rate....203
5.7 Stream access point requirements and inter-frame dependency....203

6 Dynamic range control and loudness processing....205
6.1 General....205
6.2 Description....205
6.3 Syntax....205
6.3.1 Loudness metadata....205
6.3.2 Dynamic range control metadata....205
6.3.3 Data elements....206
6.4 Decoding process....207
6.4.1 General....207
6.4.2 Dynamic range control....209
6.4.3 Usage of downmixId in MPEG-H....209
6.4.4 DRC set selection process....210
6.4.5 DRC-1 for SAOC 3D Content....212
6.4.6 DRC-1 for HOA content....212
6.4.7 Loudness normalization....214
6.4.8 Peak limiter....214
6.4.9 Time-synchronization of DRC gains....214
6.4.10 Default parameters....214

7 Object metadata decoding....215
7.1 General....215
7.2 Description....215
7.3 Syntax....216
7.3.1 Object metadata configuration....216
7.3.2 Top level object metadata syntax....217
7.3.3 Subsidiary payloads for efficient object metadata decoding....218
7.3.4 Subsidiary payloads for object metadata decoding with low delay....222
7.3.5 Enhanced object metadata configuration....227
7.4 Data structure....230
7.4.1 Definition of ObjectMetadataConfig() payloads....230
7.4.2 Efficient object metadata decoding....230
7.4.3 Object metadata decoding with low delay....239
7.4.4 Enhanced object metadata....244

8 Object rendering....247
8.1 Description....247
8.2 Terms and definitions....247
8.3 Input data....248
8.4 Processing....249
8.4.1 General remark....249
8.4.2 Imaginary loudspeakers....249
8.4.3 Dividing the loudspeaker setup into a triangle mesh....250

 © ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

8.4.4 Rendering algorithm 252

9 SAOC 3D 256
9.1 Description 256
9.2 Definitions 256
9.3 Delay and synchronization 258
9.4 Syntax 258
9.4.1 Payloads for SAOC 3D 258
9.4.2 Definition of SAOC 3D payloads 262
9.5 SAOC 3D processing 264
9.5.1 Compressed data stream decoding and dequantization of SAOC 3D data 264
9.5.2 Time/frequency transforms 264
9.5.3 Signals and parameters 264
9.5.4 SAOC 3D decoding 266
9.5.5 Dual mode 271

10 Generic loudspeaker rendering/format conversion 272
10.1 Description 272
10.2 Definitions 273
10.2.1 General remarks 273
10.2.2 Variable definitions 273
10.3 Processing 274
10.3.1 Application of transmitted downmix matrices 274
10.3.2 Application of transmitted equalizer settings 278
10.3.3 Downmix processing involving multiple channel groups 278
10.3.4 Initialization of the format converter 279
10.3.5 Audio signal processing 294

11 Immersive loudspeaker rendering/format conversion 299
11.1 Description 299
11.2 Syntax 301
11.3 Definitions 301
11.3.1 General remarks 301
11.3.2 Variable definitions 302
11.4 Processing 303
11.4.1 Initialization of the format converter 303
11.4.2 Audio signal processing 343

12 Higher order ambisonics (HOA) 350
12.1 Technical overview 350
12.1.1 Block diagram 350
12.1.2 Overview of the decoder tools 351
12.2 Syntax 353
12.2.1 Configuration of HOA elements 353
12.2.2 Payloads of HOA elements 356
12.3 Data structure 368
12.3.1 Definitions of HOA Config 368
12.3.2 Syntax of getSubbandBandwidths() 373
12.3.3 Definitions of HOA payload 373
12.4 HOA tool description 381
12.4.1 HOA frame converter 381
12.4.2 Spatial HOA decoding 398
12.4.3 HOA renderer 428
12.4.4 Layered coding for HOA 436

13 Binaural renderer 439
13.1 General 439
13.2 Frequency-domain binaural renderer 439
13.2.1 General 439
13.2.2 Definitions 441
13.2.3 Parameterization of binaural room impulse responses 445
13.2.4 Frequency-domain binaural processing 457

© ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

13.3 Time-domain binaural renderer 464
13.3.1 General 464
13.3.2 Definitions 465
13.3.3 Parameterization of binaural room impulse responses 467
13.3.4 Time-domain binaural processing 471

14 MPEG-H 3D audio stream (MHAS) 472
14.1 Overview 472
14.2 Syntax 472
14.2.1 Main MHAS syntax elements 472
14.2.2 Subsidiary MHAS syntax elements 474
14.3 Semantics 475
14.3.1 mpeghAudioStreamPacket() 475
14.3.2 MHASPacketPayload() 475
14.3.3 Subsidiary MHAS packets 477
14.4 Description of MHASPacketTypes 477
14.4.1 PACTYP_FILLDATA 477
14.4.2 PACTYP_MPEGH3DACFG 477
14.4.3 PACTYP_MPEGH3DAFRAME 477
14.4.4 PACTYP_SYNC 478
14.4.5 PACTYP_SYNCGAP 478
14.4.6 PACTYP_MARKER 478
14.4.7 PACTYP_CRC16 and PACTYP_CRC32 479
14.4.8 PACTYP_DESCRIPTOR 479
14.4.9 PACTYP_USERINTERACTION 479
14.4.10 PACTYP_LOUDNESS_DRC 479
14.4.11 PACTYP_BUFFERINFO 479
14.4.12 PACTYP_GLOBAL_CRC16 and PACTYP_ GLOBAL_CRC32 480
14.4.13 PACTYP_AUDIOTRUNCATION 480
14.4.14 PACTYP_AUDIOSCENEINFO 481
14.5 Application examples 481
14.5.1 Light-weighted broadcast 481
14.5.2 MPEG-2 transport stream 482
14.5.3 CRC error detection 482
14.5.4 Audio sample truncation 483
14.6 Multi-stream delivery and interface 483
14.7 Carriage of generic data 486
14.7.1 Syntax 486
14.7.2 Semantics 487
14.7.3 Processing at the MPEG-H 3D audio decoder 487

15 Metadata audio elements (MAE) 488
15.1 General 488
15.2 Syntax 489
15.3 Semantics 496
15.4 Definition of mae_metaDataElementIDs 509
15.5 Loudness compensation after gain interactivity 510

16 Loudspeaker distance compensation 512

17 Interfaces to the MPEG-H 3D audio decoder 513
17.1 General 513
17.2 Interface for local setup information 513
17.2.1 General 513
17.2.2 WIRE output 513
17.2.3 Syntax for local setup information 514
17.2.4 Semantics for local setup information 514
17.3 Interface for local loudspeaker setup and rendering 514
17.3.1 General 514
17.3.2 Syntax for local loudspeaker signalling 515
17.3.3 Semantics for local loudspeaker signalling 516

 © ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

17.4 Interface for binaural room impulse responses (BRIRs) 517
17.4.1 General 517
17.4.2 Syntax of binaural renderer interface 517
17.4.3 Semantics 521
17.5 Interface for local screen size information 525
17.5.1 General 525
17.5.2 Syntax 525
17.5.3 Semantics 525
17.6 Interface for signaling of local zoom area 526
17.6.1 General 526
17.6.2 Syntax 526
17.6.3 Semantics 526
17.7 Interface for user interaction 527
17.7.1 General 527
17.7.2 Definition of user interaction categories 527
17.7.3 Definition of an interface for user interaction 527
17.7.4 Syntax of interaction interface 528
17.7.5 Semantics of interaction interface 529
17.8 Interface for loudness normalization and dynamic range control (DRC) 531
17.9 Interface for scene displacement data 532
17.9.1 General 532
17.9.2 Definition of an interface for scene-displacement data 532
17.9.3 Syntax of the scene displacement interface 533
17.9.4 Semantics of the scene displacement interface 533
17.10 Interfaces for channel-based, object-based, and HOA metadata and audio data 534
17.10.1 General 534
17.10.2 Expectations on external renderers 534
17.10.3 Object-based metadata and audio data (object output interface) 534
17.10.4 Channel-based metadata and audio data 540
17.10.5 HOA metadata and audio data 543
17.10.6 Audio PCM data 545

18 Application and processing of local setup information and interaction data and scene displacement data 546
18.1 Element metadata preprocessing 546
18.2 Interactivity limitations and restrictions 551
18.2.1 General information 551
18.2.2 WIRE interactivity 551
18.2.3 Position interactivity 552
18.2.4 Screen-related element remapping and object remapping for zooming 552
18.2.5 Closest loudspeaker playout 553
18.3 Screen-related element remapping 553
18.4 Screen-related adaptation and zooming for higher order ambisonics (HOA) 556
18.5 Object remapping for zooming 557
18.6 Determination of the closest loudspeaker 558
18.7 Determination of a list of loudspeakers for conditioned closest loudspeaker playback 559
18.8 Processing of scene displacement angles for channels and objects (CO) 561
18.9 Processing of scene displacement angles for scene-based content (HOA) 562
18.10 Determination of a reduced reproduction layout based on excluded sectors 564
18.11 Diffuseness rendering 565

19 MPEG-H 3D audio profile definition 566

20 Carriage of MPEG-H 3D audio in ISO base media file format 567
20.1 General 567
20.2 Random access and stream access 567
20.3 Overview of new box structures 567
20.4 MHA decoder configuration record 567
20.4.1 Definition 567
20.4.2 Syntax 568
20.4.3 Semantics 568

© ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

20.5 MPEG-H audio sample entry 568
20.5.1 Definition 568
20.5.2 Syntax 569
20.5.3 Semantics 569
20.6 MPEG-H audio MHAS sample entry 570
20.6.1 Definition 570
20.6.2 Syntax 571
20.7 MHA dynamic range control and loudness 571
20.7.1 Definition 571
20.7.2 Syntax 571
20.7.3 Semantics 573
20.8 MHA multi-stream signalling 573
20.8.1 Definition 573
20.8.2 Syntax 574
20.8.3 Semantics 574
20.9 Audio scene information 575
20.9.1 MHA group definition 575
20.9.2 MHA switch group definition 577
20.9.3 MHA group preset definition 578
20.9.4 MHA group description text label 579
20.9.5 MHA scene information 581
20.10 Track references 582

21 Sub-parameters for the MIME type 'Codecs' parameter 582
21.1 General 582
21.2 'Codecs' parameter for MPEG-H 3D audio 582

22 Timing considerations and decoder behaviour 582

23 Multi-stream handling 582
23.1 Restrictions on extension payloads 583

24 Low complexity generic loudspeaker rendering/format conversion 584
24.1 Description 584
24.2 Definitions 585
24.2.1 General remarks 585
24.2.2 Variable definitions 586
24.3 Processing 586
24.3.1 Application of transmitted downmix matrices 586
24.3.2 Application of transmitted equalizer settings 591
24.3.3 Downmix processing involving multiple channel groups 591
24.3.4 Initialization of the format converter 592
24.3.5 Audio signal processing 607

25 Low complexity immersive loudspeaker rendering/format conversion 610
25.1 Description 610
25.2 Syntax 611
25.3 Definitions 611
25.3.1 General remarks 611
25.3.2 Variable definitions 612
25.4 Processing 613
25.4.1 Initialization of the format converter 613
25.4.2 Audio signal processing 654

26 MPEG surround 657
26.1 Technical overview 657
26.2 Syntax and data structure 658
26.3 Tool description 658

Annex A (normative) **Tables for arithmetic decoding of IGF information** 659

Annex B (normative) **SAOC 3D Decorrelator pre-mixing matrices** 663

 © ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

Annex C (informative) **Encoder tools** **669**

Annex D (normative) **Peak limiter for unguided clipping prevention** **716**

Annex E (normative) **Compact template downmix matrices** **717**

Annex F (normative) **HOA tables** **718**

Annex G (informative) **Low complexity HOA rendering** **759**

Annex H (informative) **Information on delay and complexity of time-domain binauralization** **773**

Annex I (informative) **Determination of a rotation matrix for processing of scene displacement data** **778**

Annex J (informative) **Decorrelation filtering for 'diffuseness' processing** **779**

Annex K (informative) **Distance and depth spread rendering** **780**

Annex L (informative) **HREP encoder description** **782**

Annex M (informative) **Screen-related adaptation of HOA content in complexity constrained implementations** **786**

Annex N (normative) **Retaining original file length with MPEG-H 3D audio** **787**

Annex O (normative) **Codebook tables used to de-quantize high band time domain bandwidth extension parameters** **789**

Bibliography **798**

© ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

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.

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

This second edition cancels and replaces the first edition (ISO/IEC 23008-3:2015), which has been technically revised. It also incorporates ISO/IEC 23008-3:2015/Amd.1:2016, ISO/IEC 23008-3:2015/Amd.2:2016, ISO/IEC 23008-3:2015/Amd.3:2017 and ISO/IEC 23008-3:2015/Amd.4:2016.

The main changes compared to the previous edition are as follows:

— unreadable equations have been corrected;

— profiles have been defined;

— transport of MPEG-H 3D audio in MPEG-4 ISO Base Media File Format has been defined;

— coding efficiency, especially for low bitrate coding modes, has been improved (for scene-based as well as for object-based and for multichannel-based content);

— descriptive metadata has been added;

— MHAS description has been updated;

 © ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

— usage of MPEG-H 3D audio in broadcasting applications has been greatly improved;

— a tool for Advanced Loudness Control has been added;

— a layered coding mode for coding of scene-based content has been added;

— carriage of systems metadata has been defined.

A list of all parts in the ISO/IEC 23008 series can be found on the ISO website.

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.

© ISO/IEC 2019 – All rights reserved

This is a preview of "ISO/IEC 23008-3:2019". Click here to purchase the full version from the ANSI store.

Introduction

3D sound systems are able to realize a significantly enhanced sound experience relative to current widespread 5.1 channel audio programs and playback systems. These systems demand high quality audio coding and error-free transmission in order to keep the timbre, sound localization and sound envelopment of the original audio program. Presentation over headphones with suitable spatialization are also considered.

This document provides means for all scenarios where there is a need to compress a multi-channel audio program (e.g. 22.2 channel program) and to render it to the native target number of loudspeakers. In order to reach a wide market, a 3D audio program is able to be downmixed to a lower hierarchy of loudspeakers, for example 10.1 or 8.1 channels. In addition, all scenarios support a level of random access to facilitate broadcast break-in, and "trick modes" such as fast forward when playing from stored media.

This document focuses on applications such as audio for home theatres where the audio presentation is immersive, involving many loudspeakers (e.g. from 10 to more than 20) surrounding the listener and placed below, at and above ear-level. Moreover, applications as personal TV, TV for smartphones and multi-channel audio-only programs are envisioned. These require that 3D audio encoding/decoding systems are able to operate at low bitrates appropriate for efficient transmission over a cellular channel. At the same time, the sense of envelopment and accurate sonic localization even for systems having a tablet-sized visual displays with loudspeakers built into the device and headphone listening are maintained.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of patents. ISO and IEC take no position concerning the evidence, validity and scope of these patent rights.

The holders of these patent rights have assured ISO and IEC that they are willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statements of the holders of these patent rights are registered with ISO and IEC. Information may be obtained from:

Electronics and Telecommunications Research Institute (ETRI)	218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, KOREA
Koninklijke Philips N.V.	High Tech Campus 5, 5656AE Eindhoven, THE NETHERLANDS
Thomson Licensing	Suite 303, 4 Research Way, Princeton, NJ 08540, USA
Wilus Inc.	48 Mabang-ro, Seocho-gu, Seoul, 137-894, KOREA
Fraunhofer Gesellschaft zur Foerderung der angewandten Forschung e.V.	Am Wolfsmantel 33, 90158 Erlangen, GERMANY
Qualcomm Incorporated	5775 Morehouse Drive, San Diego, CA 92021, USA
Dolby Laboratories Licensing Corporation	100 Potrero Avenue, San Francisco, CA 94103-4938, USA
Dolby International AB	999 Brannan Street, San Francisco, CA 94103-4938, USA

© ISO/IEC 2019 – All rights reserved