
**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

This document is a preview generated by EMS



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2015, Published in Switzerland

All rights reserved. Unless otherwise specified, 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
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

	Page
Foreword	viii
Introduction.....	ix
1 Scope.....	1
2 Normative references.....	1
3 Terms, definitions and mnemonics	1
3.1 Terms and Definitions.....	1
3.2 Mnemonics.....	1
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 time domain and QMF domain.....	4
4.4 Rule set for determining processing domains	5
4.4.1 Audio Core Codec, Processing Domain	5
4.4.2 Mixing	6
4.4.3 Audio Core Codec, Interface Domain to Rendering	6
4.4.4 Rendering Context	6
4.4.5 Post-Processing Context.....	6
4.4.6 End-of-Chain Context.....	7
5 MPEG-H 3D Audio Core decoder	7
5.1 Terms and Definitions.....	7
5.1.1 Joint Stereo.....	7
5.1.2 MPEG Surround based stereo (MPS 212)	7
5.2 Syntax.....	7
5.2.1 General	7
5.2.2 Decoder configuration	7
5.2.3 MPEG-H 3D Audio Core bitstream payloads	22
5.3 Data Structure.....	30
5.3.1 General	30
5.3.2 General Configuration Data Elements.....	30
5.3.3 Loudspeaker Configuration Data Elements.....	32
5.3.4 Core Decoder Configuration Data Elements	34
5.3.5 Downmix Matrix Data Elements	37
5.3.6 HOA Rendering Matrix Data Elements	40
5.4 Configuration Element Descriptions	42
5.4.1 General	42
5.4.2 Downmix configuration.....	43
5.4.3 HOA rendering matrix configuration	47
5.5 Tool Descriptions	51
5.5.1 General	51
5.5.2 Quad Channel Element	52
5.5.3 Transform Splitting	53
5.5.4 MPEG Surround for Mono to Stereo upmixing	60
5.5.5 Enhanced Noise Filling.....	62
5.5.6 Audio Pre-Roll.....	82
5.6 Buffer requirements	86
5.6.1 Minimum decoder input buffer.....	86
5.6.2 Bit reservoir	86
5.6.3 Maximum bit rate.....	87
5.7 Stream Access Point requirements and inter-frame dependency	87
6 Dynamic Range Control and Loudness Processing.....	88

6.1	Introduction	88
6.2	Description	88
6.3	Syntax	88
6.3.1	Loudness Metadata	88
6.3.2	Dynamic Range Control Metadata	89
6.3.3	Data Elements	90
6.4	Decoding Process	91
6.4.1	General	91
6.4.2	Dynamic Range Control	93
6.4.3	Usage of downmixId in MPEG-H	93
6.4.4	DRC Set Selection Process	94
6.4.5	DRC-1 for SAOC 3D Content	95
6.4.6	DRC-1 for HOA Content	96
6.4.7	Loudness Normalization	98
6.4.8	Peak Limiter	98
6.4.9	Time-Synchronization of DRC gains	98
7	Object Metadata Decoding	98
7.1	Introduction	98
7.2	Description	98
7.3	Syntax	99
7.3.1	Object Metadata Configuration	99
7.3.2	Top level object metadata syntax	100
7.3.3	Subsidiary payloads for efficient object metadata decoding	100
7.3.4	Subsidiary payloads for object metadata decoding with low delay	104
7.4	Data Structure	108
7.4.1	Definition of ObjectMetadataConfig() payloads	108
7.4.2	Efficient Object Metadata Decoding	108
7.4.3	Object Metadata Decoding with Low Delay	113
8	Object Rendering	117
8.1	Description	117
8.2	Terms and Definitions	117
8.3	Input data	117
8.4	Processing	119
8.4.1	Imaginary Loudspeakers	119
8.4.2	Dividing the Loudspeaker Setup into a Triangle Mesh	120
8.4.3	Rendering Algorithm	121
9	SAOC 3D	125
9.1	Description	125
9.2	Definitions	125
9.3	Delay and synchronization	127
9.4	Syntax	127
9.4.1	Payloads for SAOC 3D	127
9.4.2	Definition of SAOC 3D payloads	131
9.5	SAOC 3D processing	133
9.5.1	Compressed data stream decoding and dequantization of SAOC 3D data	133
9.5.2	Time/frequency transforms	133
9.5.3	Signals and parameters	133
9.5.4	SAOC 3D decoding	135
9.5.5	Dual mode	140
10	Generic Loudspeaker Rendering/Format Conversion	141
10.1	Description	141
10.2	Definitions	142
10.2.1	General remarks	142
10.2.2	Variable definitions	142
10.3	Processing	143
10.3.1	Application of transmitted downmix matrices	143
10.3.2	Application of transmitted equalizer settings	148
10.3.3	Downmix processing involving multiple channel groups	148

10.3.4	Initialization of the format converter	149
10.3.5	Audio signal processing.....	165
11	Immersive Loudspeaker Rendering / Format Conversion	171
11.1	Description	171
11.2	Syntax.....	172
11.3	Definitions	173
11.3.1	General remarks	173
11.3.2	Variable definitions	173
12	Higher Order Ambisonics (HOA)	221
12.1	Technical Overview	221
12.1.1	Block Diagram	221
12.1.2	Overview of the decoder tools	222
12.2	Syntax.....	223
12.2.1	Configuration of HOA elements.....	223
12.2.2	Payloads of HOA elements.....	224
12.3	Data Structure.....	229
12.3.1	Definitions of HOA Config	229
12.3.2	Definitions of HOA payload	231
12.4	HOA Tool Description	234
12.4.1	HOA Frame Converter.....	234
12.4.2	Spatial HOA decoding.....	243
12.4.3	HOA Renderer.....	255
13	Binaural Renderer	263
13.1	Introduction.....	263
13.2	Frequency-Domain Binaural Renderer.....	264
13.2.1	Introduction.....	264
13.2.2	Definitions	266
13.2.3	Parameterization of Binaural Room Impulse Responses	270
13.2.4	Frequency-Domain Binaural Processing.....	282
13.3	Time-Domain Binaural Renderer	289
13.3.1	Introduction.....	289
13.3.2	Definitions	290
13.3.3	Parameterization of Binaural Room Impulse Responses	291
13.3.4	Time-Domain Binaural Processing.....	296
14	MPEG-H 3D audio stream (MHAS).....	297
14.1	Overview.....	297
14.2	Syntax.....	297
14.2.1	Main MHAS syntax elements.....	297
14.2.2	Subsidiary MHAS syntax elements	299
14.3	Semantics.....	299
14.3.1	mpeghAudioStreamPacket()	299
14.3.2	MHASPacketPayload()	300
14.4	Description of MHASPacketTypes.....	300
14.4.1	PACTYP_FILLDATA	300
14.4.2	PACTYP_MPEGH3DACFG	300
14.4.3	PACTYP_MPEGH3DAFRAME.....	301
14.4.4	PACTYP_SYNC	301
14.4.5	PACTYP_SYNGAP	301
14.4.6	PACTYP_MARKER	301
14.4.7	PACTYP_CRC16 and PACTYP_CRC32	302
14.4.8	PACTYP_DESCRIPTOR	302
14.4.9	PACTYP_USERINTERACTION	302
14.4.10	PACTYP_LOUDNESS_DRC	302
14.4.11	PACTYP_BUFFERINFO.....	303
14.5	Application Examples	303
14.5.1	Light-weighted broadcast.....	303
14.5.2	MPEG-2 Transport Stream.....	303
14.6	Multi-Stream Delivery and Interface	304

15	Metadata Audio Elements (MAE).....	306
15.1	Introduction	306
15.2	Syntax	307
15.3	Semantics	311
15.4	Definition of mae_metaDataElementIDs.....	319
16	Loudspeaker Distance Compensation	319
17	Interfaces to the MPEG-H 3D audio decoder	320
17.1	General.....	320
17.2	Interface for local setup information	321
17.2.1	General.....	321
17.2.2	WIRE output	321
17.2.3	Syntax for local setup information.....	321
17.2.4	Semantics for local setup information	321
17.3	Interface for local loudspeaker setup and rendering.....	322
17.3.1	General.....	322
17.3.2	Syntax for local loudspeaker signaling.....	322
17.3.3	Semantics for local loudspeaker signaling.....	323
17.4	Interface for binaural room impulse responses (BRIRs).....	324
17.4.1	Introduction	324
17.4.2	Syntax of Binaural Renderer Interface	324
17.4.3	Semantics	327
17.5	Interface for local screen size information	332
17.5.1	General.....	332
17.5.2	Syntax	332
17.5.3	Semantics	332
17.6	Interface for signaling of local zoom area.....	333
17.6.1	General.....	333
17.6.2	Syntax	333
17.6.3	Semantics	334
17.7	Interface for user interaction	334
17.7.1	Introduction	334
17.7.2	Definition of User Interaction Categories.....	334
17.7.3	Definition of an Interface for User Interaction	335
17.7.4	Syntax of interaction interface	335
17.7.5	Semantics of interaction interface	336
17.8	Interface for loudness normalization and dynamic range control (DRC).....	338
18	Application and processing of local setup information and interaction data.....	338
18.1	<i>Element Metadata</i> Preprocessing	338
18.2	Interactivity Limitations and Restrictions	341
18.2.1	General Information.....	341
18.2.2	WIRE Interactivity	341
18.2.3	Position Interactivity	342
18.2.4	Screen-Related Element Remapping and Object Remapping for Zooming	342
18.2.5	Closest Speaker Playback	342
18.3	Screen-Related Element Remapping	342
18.4	Object Remapping for Zooming	344
18.5	Determination of the Closest Speaker.....	345
Annex A	(normative) Tables for arithmetic decoding of IGF information	347
A.1	cf_se01[27]	347
A.2	cf_se10[27]	347
A.3	cf_se02[7][27]	347
A.4	short cf_se20[7][27]	347
A.5	short cf_se11[7][7][27]	348
A.6	cf_off_se01	349
A.7	cf_off_se10	349
A.8	cf_off_se02[7].....	349
A.9	short cf_off_se20[7].....	350
A.10	cf_off_se11[7][7]	350

Annex B (normative) SAOC 3D Decorrelator pre-mixing matrices	351
B.1 Premixing matrix for output configurations with small number of output channels.....	351
B.2 Premixing matrix for 22.2 output configuration.....	351
B.3 Algorithm for generating pre-mixing matrices.....	352
B.3.1 Input to the algorithm and representations.....	352
B.3.2 Algorithm steps.....	352
Annex C (informative) Encoder Tools	356
C.1 General Overview.....	356
C.1.1 Encoder block diagram.....	356
C.1.2 Overview of the encoder and decoder building blocks.....	356
C.2 Core Encoder Tools.....	357
C.2.1 Quad Channel Element.....	357
C.2.2 Transform Splitting.....	358
C.2.3 Calculation of Residual Signal for MPEG Surround with Hybrid Residual Coding.....	359
C.2.4 Enhanced Noise Filling.....	359
C.3 Object Metadata Encoding.....	360
C.3.1 Pre-Processing of the Object Metadata.....	360
C.3.2 Efficient Object Metadata Encoding.....	361
C.3.3 Object Metadata Encoding with Low Delay.....	361
C.3.4 Spatially skipping objects.....	361
C.4 SAOC 3D Encoder.....	361
C.4.1 Overview.....	361
C.4.2 Calculation of the SAOC 3D parameters.....	361
C.4.3 Time/frequency transform.....	362
C.4.4 Framing.....	362
C.4.5 Parameter quantization and coding.....	362
Annex D (informative) Peak limiter for unguided clipping prevention	384
Annex E (normative) Compact Template Downmix Matrices	385
Annex F (normative) HOA Tables	386
F.6 32 Uniformly Distributed Positions in Spherical Coordinates.....	390
F.12 Table of 256x8 weighting values, WeightValCdbk.....	404
Annex G (informative) Low Complexity HOA Rendering	421
G.1 Tool Description.....	421
G.2 Predominant Sound Rendering.....	421
G.3 Ambient Sound Rendering.....	422
G.4 Output Signal Composition.....	422
Annex H (informative) Information on delay and complexity of Time-Domain binauralization	423
H.1 Complexity and latency.....	423
H.1.1 Algorithm description.....	423
H.1.2 Complexity.....	423
H.1.3 Latency.....	424
H.2 Experimental results.....	425
H.3 Alternative low-delay implementations.....	426
Bibliography	428

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 on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: [Foreword - Supplementary information](#)

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 23008 consists of the following parts, under the general title *Information technology — High efficiency coding and media delivery in heterogeneous environments*:

- *Part 1: MPEG media transport (MMT)*
- *Part 2: High efficiency video coding*
- *Part 3: 3D audio*
- *Part 4: MMT Reference and Conformance Software*
- *Part 5: Reference software for high efficiency video coding*
- *Part 8: HEVC conformance testing*
- *Part 10: MPEG media transport forward error correction (FEC) codes*
- *Part 11: MPEG media transport composition information*
- *Part 12: Image file format*
- *Part 13: MMT Implementation Guidelines*

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 part of ISO/IEC 23008-3 “High Efficiency Coding and Media Delivery in Heterogeneous Environments — Part 3: 3D Audio” 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.

The main focus of this specification are 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 speakers built into the device and headphone listening are maintained.

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

1 Scope

This part of ISO/IEC 23008-3 specifies technology which supports the efficient transmission of 3D audio signals and flexible rendering for the playback of 3D audio in a wide variety of listening scenarios. These include 3D home theater setups, 22.2 loudspeaker systems, automotive entertainment systems and playback over headphones connected to a tablet or smartphone.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 13818-1:2013, *Information technology — Generic Coding of moving pictures and associated audio information: Systems*

ISO/IEC 14496-3:2009, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-11, *Information technology — Coding of audio-visual objects — Part 11: Scene description and application engine*

ISO/IEC 23001-8:2013, *Information technology — MPEG systems technologies — Part 8: Coding-independent code-points*

ISO/IEC 23001-8:2013/Amd.1, *Information technology — MPEG systems technologies — Part 8: Coding-independent code-points, AMENDMENT 1: New audio code points*

ISO/IEC 23003-1:2007, *Information technology — MPEG audio technologies — Part 1: MPEG Surround*

ISO/IEC 23003-2:2010, *Information technology — MPEG audio technologies — Part 2: Spatial Audio Object Coding (SAOC)*

ISO/IEC 23003-3:2012, *Information technology — MPEG audio technologies — Part 3: Unified speech and audio coding*

ISO/IEC 23003-4:2015, *Information technology — MPEG audio technologies — Part 4: Dynamic range control*

3 Terms, definitions and mnemonics

3.1 Terms and Definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-3:2009, 1.3 (Terms and definitions), in ISO/IEC 14496-3:2009, 1.4 (Symbols and abbreviations) and in ISO/IEC 23003-3:2012, 3.1 (Terms and definitions) apply.

3.2 Mnemonics

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