INTERNATIONAL STANDARD

ISO/IEC 23008-3

Second edition
2019-02

AMENDMENT 2
2020-11

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

Part 3: 3D audio

AMENDMENT 2: 3D Audio baseline profile, corrections and improvements

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

Partie 3: Audio 3D

AMENDEMENT 2: Profil de base audio 3D, corrections et améliorations

Reference number
ISO/IEC 23008-3:2019/Amd.2:2020(E)

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

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).

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) or the IEC list of patent declarations received (see https://patents.iec.ch).

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 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 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.

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

Part 3: 3D audio

AMENDMENT 2: 3D Audio baseline profile, corrections and improvements

Subclause 4.8.2 (Profiles)

After list item 3, add:

4) The baseline profile is a subset of the low-complexity profile which supports channel and object signals.

Replace Table 2 with:

Table 2 — Summary of the location of and normative reference to the definitions of MPEG-H 3D audio profiles

Tool/Module		Defined in ISO/IEC	Sub-clause	USAC 23003-3	MPEG-H 3D audio High profile	MPEG-H 3D audio Low-complexity profile	MPEG-H 3D audio Baseline profile
block switching		14496-3	4.6.11	**X**	X	X	X
window shapes	AAC based	14496-3	4.6.11	**X**	X	X	X
	Additional windows	23003-3	6.2.9.3	**X**	X	X	X
filter bank	AAC based	14496-3	4.6.11	**X**	X	X	X
	additional USAC	23003-3	7.9	**X**	X	X	X
TNS		14496-3	4.6.9	**X**	X	X	X
intensity		14496-3	4.6.8.2				
coupling		14496-3	4.6.8.3				
perceptual noise synthesis	PNS	14496-3	4.6.13				
	noise filling	23003-3	7.2	**X**	X	X	X

[a] Restrictions apply dependent on the levels.

[b] Implementation of binaural rendering is only mandated if headphone reproduction is supported.

[c] Multi-band DRC-1 shall be applied in the STFT domain of the TD format converter.

[d] The TD format converter downmix shall be applied for downmixing.

[e] In order to achieve target complexity for the LC profile at a given level, study Annex G.

[f] File format encapsulation is independent of the profile that is used for the bitstream. A profile level indicator is part of the file format specification (see subclause 20.4).

Table 2 *(continued)*

Tool/Module		Defined in ISO/IEC	Sub-clause	USAC 23003-3	MPEG-H 3D audio High profile	MPEG-H 3D audio Low-complexity profile	MPEG-H 3D audio Baseline profile
MS	basic mid/side coding	14496-3	4.6.8.1	**X**	X	X	X
	MDCT based complex prediction	23003-3	7.7.2	**X**	X	X	X
quantization	non-uniform	14496-3	4.6.1	**X**	X	X	X
	uniform	23003-3	7.1	**X**	X	X	
entropy coding	Huffman	14496-3	4.6.3				
	context adaptive arithmetic coding	23003-3	7.4	**X**	X	X	X
SBR	base	14496-3	4.6.18	**X**	X		
	enhanced	23003-3	7.5	**X**	X		
parametric stereo extension	parametric stereo	14496-3	8.6.4 / 8.A				
	MPEG surround 2-1-2 (incl. residual coding)	23003-3	6.2.13	**X**	X		
	quad channel element	23008-3	5.5		X		
ACELP		23003-3	7.14	**X**	X	X	
frequency domain noise shaping	scale factor based	14496-3	4.6.2	**X**	X	X	X
	LPC based	23003-3		**X**	X	X	
intelligent gap filling	IGF for FD	23008-3			X	X	X
improved LPD coding	IGF for TCX and TBE in ACELP	23008-3			X	X	
	LPD stereo	23008-3			X	X	
predictors for FD	frequency-domain prediction and time-domain post-filtering	23008-3			X	X	X
predictors for TCX	frequency-domain prediction and time-domain post-filtering	23008-3			X	X	

[a] Restrictions apply dependent on the levels.

[b] Implementation of binaural rendering is only mandated if headphone reproduction is supported.

[c] Multi-band DRC-1 shall be applied in the STFT domain of the TD format converter.

[d] The TD format converter downmix shall be applied for downmixing.

[e] In order to achieve target complexity for the LC profile at a given level, study Annex G.

[f] File format encapsulation is independent of the profile that is used for the bitstream. A profile level indicator is part of the file format specification (see subclause 20.4).

Table 2 *(continued)*

Tool/Module		Defined in ISO/IEC	Sub-clause	USAC 23003-3	MPEG-H 3D audio High profile	MPEG-H 3D audio Low-complexity profile	MPEG-H 3D audio Baseline profile
discrete multi-channel coding	MCT	23008-3			X	X	X
format converter	generic downmix	23008-3	10, 24		X	X[d]	X[d]
immersive rendering	immersive rendering within format converter	23008-3	11, 25		X	X[d]	X[d]
static metadata	metadata audio elements (MAE) and audio scene information (ASI)	23008-3	15		X	X	X
	decoder and renderer						
dynamic object metadata	object audio metadata (OAM)	23008-3	7, 8		X	X	X
	decoder and renderer						
MPS	MPEG surround extension	23003-1	10		X		
SAOC-3D	decoder and renderer	23008-3	9		X		
HOA	decoder and renderer	23008-3	12		X	X[e]	
	near field compensation	23008-3			X	X[a]	
	subband directional prediction	23008-3			X		
	parametric ambiance replication (PAR)	23008-3			X		
	phase-based decorrelation	23008-3			X		
Binaural	FD-binaural, TD-binaural	23008-3	13		X	X[b]	X[b]
	HOA2Binaural H2B	23008-3			X	X[b]	

[a] Restrictions apply dependent on the levels.

[b] Implementation of binaural rendering is only mandated if headphone reproduction is supported.

[c] Multi-band DRC-1 shall be applied in the STFT domain of the TD format converter.

[d] The TD format converter downmix shall be applied for downmixing.

[e] In order to achieve target complexity for the LC profile at a given level, study Annex G.

[f] File format encapsulation is independent of the profile that is used for the bitstream. A profile level indicator is part of the file format specification (see subclause 20.4).

Table 2 *(continued)*

Tool/Module		Defined in ISO/IEC	Sub-clause	USAC 23003-3	MPEG-H 3D audio High profile	MPEG-H 3D audio Low-complexity profile	MPEG-H 3D audio Baseline profile
DRC	DRC-1	23003-4			X	X[c]	X[c]
	DRC-2 (single band)	23003-4			X	X	X
	DRC-2 (multi band)	23003-4					
	DRC-3 (single band)	23003-4			X	X	X
sample rate converter		23008-3			X	X	X
peak limiter	unguided clipping prevention	23008-3 23003-4	D		X	X	X
loudness	loudness metada-ta and handling	23003-4	6		X	X	X
	loudness com-pensation	23008-3			X	X	X
MHAS	MPEG-H 3D Audio stream	23008-3	14		X	X	X
	truncation message and CRC packet type, ASI packet type	23008-3			X	X	X
file format	carriage of MPEG-H 3D audio in ISO base media file format	23008-3			f		
interfaces and process-ing	interfaces and processing for in-teraction data and local setup info	23008-3	17,18		X	X	X
carriage of generic data	carriage of generic data for the interaction with system engine	23008-3			X	X	
TCC	tonal component coding	23008-3			X		
IC	internal channel	23008-3			X		
HREP	high resolution envelope processing	23008-3			X		

[a] Restrictions apply dependent on the levels.

[b] Implementation of binaural rendering is only mandated if headphone reproduction is supported.

[c] Multi-band DRC-1 shall be applied in the STFT domain of the TD format converter.

[d] The TD format converter downmix shall be applied for downmixing.

[e] In order to achieve target complexity for the LC profile at a given level, study Annex G.

[f] File format encapsulation is independent of the profile that is used for the bitstream. A profile level indicator is part of the file format specification (see subclause 20.4).

The baseline profile is a subset of the low-complexity profile. If a decoder implementation supports decoding of low complexity profile level 3 bitstreams and supports the configuration extension CompatibleProfileLevelSet(), then the decoder shall support decoding of bitstreams encoded according to the baseline profile level 3. Bitstreams complying to the baseline profile may be signalled using:

— the mpegh3daProfileLevelIndication field set to indicate baseline profile as specified in Table 64, or alternatively,

— the mpegh3daProfileLevelIndication field set to indicate low complexity profile as specified in Table 64 and the CompatibleProfileLevelSet configuration extension for indicating compatibility to baseline profile, as described in Annex P.

Additionally, it is strongly recommended that low complexity profile bitstreams conforming to the baseline profile, are signalled using the profile and level values for mpegh3daProfileLevelIndication and CompatibleSetIndication given in Table P.1.

Subclause 4.8.2.4

Add new subclauses 4.8.2.5, 4.8.2.6 and 4.8.2.7 after subclause 4.8.2.4:

4.8.2.5 Levels of the baseline profile

4.8.2.5.1 General

Table AMD2.1 — Levels and their corresponding restrictions for the baseline profile

Level	Max. sampling rate	Max. number of core channels in compressed data stream	Max. number of decoder processed core channels	Max. number of channels in referenceLayout
1	48000	10	5	5
2	48000	18	9	9
3	48000	32	16[a] or 24[b]	16[a] or 24[b]
4	48000	56	28	24
5	96000	56	28	24

[a] No additional complexity restrictions are applied.

[b] Additional complexity restrictions given in 4.8.2.5.1 are applied.

— The use of switch groups determines the subset of core channels from the core channels in the bitstream that shall be decoded.

— If the mae_AudioSceneInfo() contains switch groups (mae_numSwitchGroups>0), then the elementLengthPresent flag shall be 1.

— The number of channels of the signalled referenceLayout shall not exceed the values defined in the levels in Table AMD2.1.

— Object renderer and binaural renderer that perform at least as well as the object and binaural renderer specified in Clauses 8 and 13 may be integrated using the output interfaces for un-rendered channels and objects described in subclause 17.10.

NOTE The performance recommendation covers the behaviour of the decoder over the complete decoding and rendering chain, especially for the case of configuration changes as described in subclause 5.5.6, mixing of channel and object content or DRC processing, loudness compensation and user interactivity.

— For Level 3 the maximum number of decoder processed core channels and maximum number of channels signalled in referenceLayout is:

  a) 16 if no additional complexity restrictions are applied,

  b) 24 if all the complexity restrictions in 4.8.2.5.1 are applied.

4.8.2.5.2 Complexity restrictions for Level 3 with more than 16 decoder processed core channels

— **signalGroupType** in Signals3d() shall indicate SignalGroupTypeObject (Objects only).

— **usacElementType[elemIdx]** in mpegh3daDecoderConfig() shall indicate ID_USAC_SCE or ID_USAC_EXT.

— **noiseFilling** and **enhancedNoiseFilling** in mpegh3daCoreConfig() shall be set to "0".

— **usacExtElementType** in mpegh3daExtElementConfig() shall not be set to ID_EXT_ELE_MCT.

— Long term prediction filter shall not be used, i.e., **ltpf_data_present** and **common_ltpf** shall be set to "0".

— Frequency domain predictor shall not be used, i.e., **fdp_data_present** shall be set to "0".

4.8.2.6 Restrictions for the baseline profile and levels

All restrictions defined for low complexity profile in subclause 4.8.2.2 shall apply.

The LPD path of the core coder and HOA path are not supported.

Restrictions defined in Table AMD2.2 shall apply.

Table AMD2.2 — Baseline profile restrictions

MPEG-H 3D audio bit field	Structure	Use description
phaseAlignStrength	*downmixConfig()*	Shall have the value “0”
SignalGroupType[grp]	*Signals3d()*	Shall have the value "*SignalGroupTypeChannels*" or "*SignalGroupTypeObject*"
qceIndex	*mpegh3daChannelPairElementConfig()*	Shall have the value “0”
lpdStereoIndex	*mpegh3daChannelPairElementConfig()*	Shall have the value “0”
tw_mdct	*mpegh3daCoreConfig()*	Shall have the value “0”
fullbandLpd	*mpegh3daCoreConfig()*	Shall have the value “0”
core_mode[ch]	*mpegh3daCoreCoderData()*	Shall have the value “0”
common_max_sfb	*StereoCoreToolInfo()*	Shall have the value “1”
tns_on_lr	*StereoCoreToolInfo()*	Shall have the value “1”
common_tw	*StereoCoreToolInfo()*	Shall have the value “0”
fac_data_present	*fd_channel_stream()*	Shall have the value “0”

4.8.2.7 Signalling of profile and level compatibility sets

MPEG-H 3d audio bitstreams may comply with multiple profiles and levels and the CompatibleProfileLevelSet() syntax element defined in Table AMD2.3 may be used to signal the compatibility to multiple profiles.

The CompatibleProfileLevelSet() syntax element contains a list of profile-level numbers the content is compatible with. Only the lowest level per profile needs to be present, as higher level decoders are inherently compatible with lower level content.

Table AMD2.3 — Syntax of CompatibleProfileLevelSet()

Syntax	No. of bits	Mnemonic
CompatibleProfileLevelSet()		
{		
bsNumCompatibleSets;	**4**	**uimsbf**
numCompatibleSets = bsNumCompatibleSets + 1;		
reserved;	**4**	**uimsbf**
for (idx = 0; idx < numCompatibleSets; idx++) {		
CompatibleSetIndication;	**8**	**uimsbf**
}		
}		

Subclause 5.2.2.3

Replace Table 24 with:

Table 24 — Syntax of mpegh3daConfigExtension()

Syntax	No. of bits	Mnemonic
mpegh3daConfigExtension()		
{		
numConfigExtensions = escapedValue(2,4,8) + 1;		
for (confExtIdx=0; confExtIdx<numConfigExtensions; confExtIdx++) {		
usacConfigExtType[confExtIdx] = escapedValue(4,8,16);		
usacConfigExtLength[confExtIdx] = escapedValue(4,8,16);		
switch (usacConfigExtType[confExtIdx]) {		
case ID_CONFIG_EXT_FILL:		
while (usacConfigExtLength[confExtIdx]--) {		
fill_byte[i]; /* should be '10100101' */	**8**	**uimsbf**
}		
break;		
case ID_CONFIG_EXT_DOWNMIX:		
downmixConfig();		
break;		
case ID_CONFIG_EXT_LOUDNESS_INFO:		
mpegh3daLoudnessInfoSet();		