ETSI TS 103 589 V1.1.1 (2018-06)

TECHNICAL SPECIFICATION

Higher Order Ambisonics (HOA) Transport Format

EBU
OPERATING EUROVISION

Reference

DTS/JTC-047

Keywords

audio, broadcasting, TV, UHDTV

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Technical Specification (TS) has been produced by Joint Technical Committee (JTC) Broadcast of the European Broadcasting Union (EBU), Comité Européen de Normalisation ELECtrotechnique (CENELEC) and the European Telecommunications Standards Institute (ETSI).

NOTE: The EBU/ETSI JTC Broadcast was established in 1990 to co-ordinate the drafting of standards in the specific field of broadcasting and related fields. Since 1995 the JTC Broadcast became a tripartite body by including in the Memorandum of Understanding also CENELEC, which is responsible for the standardization of radio and television receivers. The EBU is a professional association of broadcasting organizations whose work includes the co-ordination of its members' activities in the technical, legal, programme-making and programme-exchange domains. The EBU has active members in about 60 countries in the European broadcasting area; its headquarters is in Geneva.

European Broadcasting Union
CH-1218 GRAND SACONNEX (Geneva)
Switzerland
Tel: +41 22 717 21 11
Fax: +41 22 717 24 81

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

Higher Order Ambisonics (HOA) signals are able to deliver a significantly enhanced immersive sound compared to conventional stereo or 5.1 channel audio signals. However, there are some use cases where HOA signals cannot be transported because of the large number of HOA input channels. The present document provides an HOA transport format which allows unrestricted HOA order signals to be transported.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] ISO/IEC 23008-3:2015/AMD 1:2016: "Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, 3D Audio Profile and Levels".

NOTE: Available at https://www.iso.org/standard/67953.html.

[2] ISO/IEC 23008-3:2015/DAM 5: "Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, Audio Metadata Enhancements".

NOTE: Available at https://www.iso.org/standard/74433.html.

[3] ISO/IEC 23008-3:2015: "Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio".

NOTE: Available at https://www.iso.org/standard/63878.html.

[4] ISO/IEC 23008-3:2015/AMD 3:2017: " Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, MPEG-H 3D Audio Phase 2".

NOTE: Available at https://www.iso.org/standard/69561.html.

[5] ISO/IEC 13818-1:2015: "Information technology - Generic coding of moving pictures and associated audio information - Part 1: Systems".

NOTE: Available at https://www.iso.org/standard/67331.html.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] SMPTE Motion Imaging Journal: "Building The World's Most Complex TV Network: A Test Bed for Broadcasting Immersive and Interactive Audio" R. L. Bleidt et al.: pp. 26-34, 2017.

NOTE: Available at http://ieeexplore.ieee.org/document/7963945/.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

MPEG-H Audio Stream (MHAS): self-contained stream format to transport ISO/IEC 23008-3 data

MPEG-H 3DA: MPEG-H 3D Audio standard defined in ISO/IEC 23008-3 [1] to [4].

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACN	Ambisonic Channel Number
AGC	Adaptive Gain Control
AU	Access Unit
BG	Background (audio channel)
CRC	Cyclic Redundancy Check
DAW	Digital Audio Workstation
FG	Foreground (audio channel)
HDMI	High-Definition Multimedia Interface
HD-SDI	High-Definition Serial Digital Interface
HOA	Higher Order Ambisonics
HTF	HOA Transport Format
HTFAS	HOA Transport Format Audio Stream
ISO	International Organization for Standardization
MHAS	MPEG-H Audio Stream
MMT	MPEG media transport
MPEG	Moving Pictures Experts Group
MPEG-H LC	MPEG-H Audio Low Complexity profile

NOTE: As defined in ISO/IEC 23008-3 [1].

NOC	Network Operation Centre
OTA	Over The Air (media)
OTT	Over The Top (media)
PCM	Pulse Code Modulation
SDI	Serial Digital Interface
SID	Single Index Designation
SMPTE	Society of Motion Picture & Television Engineers
VHTF	Vector based HOA Transport Format

4 Higher Order Ambisonics (HOA) Transport Format

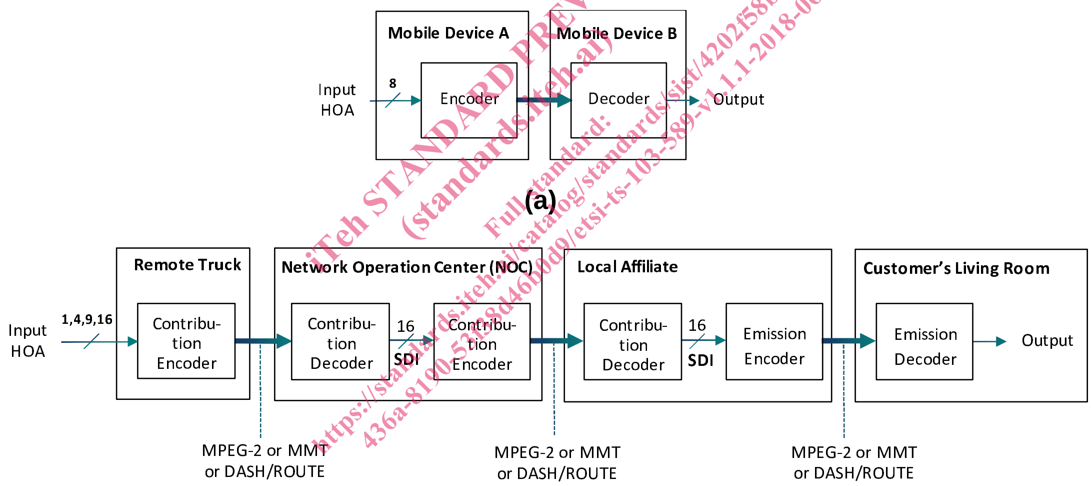
4.1 Introduction

Higher Order Ambisonics (HOA) signals are able to deliver a significantly enhanced immersive sound compared to conventional stereo or 5.1 channel audio signals. However, there are some use cases where HOA signals cannot be transported because of the large number of HOA input channels.

One use case is mobile devices where the number of input channels is limited by *N* Pulse-Code Modulation (PCM) channels. As shown in Figure 1 (a), if *N* is 8, a maximum of First Order Ambisonics (FOA which requires 4 PCM channels) can be transported.

Another use case is a typical broadcast workflow as shown in Figure 1 (b). Here, a contribution encoder can transmit 16 PCM channels from the remote truck to the Network Operation Centre (NOC) or local affiliate(s). However, the use of single High-Definition Serial Digital Interface (HD-SDI) link has a limitation of being able to transport only 16 PCM channels. This restricts the transport to a maximum of 3rd order HOA signals (requiring 16 PCM channels) with the additional restriction that there are no additional discrete audio elements to be transported. If additional audio elements are to be transported, only a maximum of 2nd order HOA (requiring 9 PCM channels) can be transported.

The present document aims to specify an HOA transport format which allows unrestricted HOA order signals to be transported. This not only includes the above two cases, but also any other cases with limitations in bandwidth and the number of transport channels. Other examples include High-Definition Multimedia Interface (HDMI) or other wired or wireless connectivity interfaces.



(b)

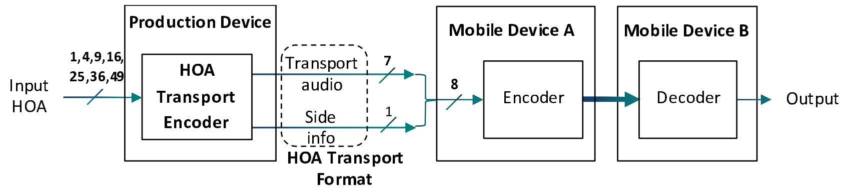
Figure 1: (a) conventional mobile devices and (b) conventional broadcast chain for order-restricted Ambisonics transport

4.2 Generic HOA Transport Format

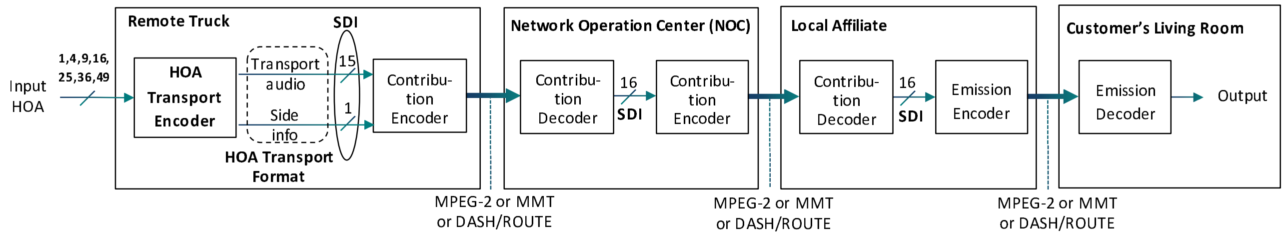
To transport higher than 1st order HOA over the mobile device as shown in Figure 1 (a), an HOA Transport Encoder is used in the production devices, such as a microphone array or a digital audio workstation (DAW). As shown in Figure 2 (a), the HOA transport encoder encodes the input HOA of any order into the **HOA transport format** (**HTF**) which contains *I* transport audio signals along with the HOA Side-info data. The number *I* of transport audio signals is usually much lower than the number of HOA input coefficients.

To transport higher than 3rd order HOA over the SDI framework, an HOA Transport Encoder is placed in front of the contribution encoder as shown in Figure 2 (b). For example, the HOA transport encoder converts input 49 HOA coefficients (6th order HOA signal) to the HOA transport format which contains 13 transport audio channels along with a single HOA Side-info channel. The 16 channel HD-SDI can carry these (13+1) channels with 2 empty channels.

For error protection in SDI transmission, the HOA Side-info can be modulated with communications modem technologies into a PCM control track signal that fits into the audio signal bandwidth [i.1].



(a)



(b)

Figure 2: HOA Transport Format for (a) mobile devices and (b) broadcast chain

Annex A presents an example guideline about HOA transport over Serial Digital Interface (SDI) utilizing communications modem technologies [i.1].

Annex B shows the HOA content production workflow where the HOA transport encoder is placed outside the broadcast chain.

In Table 1, the syntax of the configuration of Generic HOA transport format is defined as a binary representation format. In Table 2, the corresponding semantics of the configuration of Generic HOA transport format is defined.

Table 1: Syntax of HOATransportFormatConfig()

Syntax	No. of bits	Mnemonic
HOATransportFormatConfig(HoaTransportType)		
{		
if (HoaTransportType == 0) {		
InputSamplingFrequency;	**4**	**uimsbf**
InputAudioBitDepth = (**InputAudioBitDepthIdx**+1)*8;	**2**	**uimsbf**
HoaFrameLengthIdx;	**3**	**uimsbf**
NumOfHoaCoeffs = (**HoaOrder** + 1)^2;	**5**	**uimsbf**
NumOfTransportChannels = NumOfHoaCoeffs;		
HoaNormalization;	**2**	**uimsbf**
HoaCoeffOrdering;	**2**	**uimsbf**
IsScreenRelative;	**1**	**uimsbf**
} else if (HoaTransportType == 1) {		
HoaNormalization = 1;		
HoaCoeffOrdering = 0;		
NumOfTransportChannels = **CodedNumOfTransportChannels**+1;	**5**	**uimsbf**
HOAConfig();		
} else if (HoaTransportType == 2) {		
HoaNormalization = 0;		
HoaCoeffOrdering = 0;		
NumOfTransportChannels = **CodedNumOfTransportChannels**+1;	**5**	**uimsbf**
HOAConfig_SN3D();		
isScreenRelative = isScreenRelative_E;		
} else if (HoaTransportType == 3) {		
InputSamplingFrequency;	**4**	**uimsbf**
InputAudioBitDepth = (**InputAudioBitDepthIdx**+1)*8;	**2**	**uimsbf**
HoaFrameLengthIdx;	**3**	**uimsbf**
NumOfHoaCoeffs = (**HoaOrder** + 1)^2;	**5**	**uimsbf**
HoaNormalization = 0;		
HoaCoeffOrdering = 0;		
IsScreenRelative;	**1**	**uimsbf**
NumOfTransportChannels = **CodedNumOfTransportChannels**+1;	**5**	**uimsbf**
}		
if (IsScreenRelative) {		
if (**hasNonStandardScreenSize**) {	**1**	**bslbf**
if (**isCenteredInAzimuth**) {	**1**	**bslbf**
bsScreenSizeAz;	**9**	**uimsbf**
} else {		
bsScreenSizeLeftAz;	**10**	**uimsbf**
bsScreenSizeRightAz;	**10**	**uimsbf**
}		
bsScreenSizeTopEl;	**9**	**uimsbf**
bsScreenSizeBottomEl;	**9**	**uimsbf**
}		
}		
}		

Table 2: Semantics of HOATransportConfig()

HoaTransportType	This element contains information about HOA transport mode. 0: HOA coefficients (as defined in this clause) 1: ISO/IEC 23008-3-based HOA Transport Format as defined in clause 4.3 2: Modified ISO/IEC 23008-3-based HOA Transport Format for SN3D normalization as defined in clause 4.4 3: V-vector based HOA Transport Format as defined in clause 4.5
InputSamplingFrequency	This element contains information about input sampling frequency. 0: 24 kHz 1: 32 kHz 2: 44,1 kHz 3: 48 kHz 4: 96 kHz 5: 192 kHz 6 - 15: reserved

InputAudioBitDepthIdx	This element determines the input audio bit depth by InputAudioBitDepth = (**InputAudioBitDepthIdx**+1)*8.
HoaOrder	This element determines the HOA order of the coded signal.
HoaNormalization	This element contains information about HOA coefficient normalization. 0: SN3D normalization 1: N3D normalization 2: FuMa normalization 3: reserved
HoaCoeffOrdering	This element contains information about HOA coefficient ordering. 0: ACN 1: SID 2-3: reserved
IsScreenRelative	This element contains information about whether the content is: 0: not screen related 1: screen related
hasNonStandardScreenSize	This flag specifies whether the defined production screen size is different from the default screen size. The definition is done via viewing angles (in degrees) corresponding to the screen edges. The default screen size is defined with the following values (a 4K display at an optimal viewing distance): $\varphi_{\text{left}} = 29.0^\circ$, $\varphi_{\text{right}} = -29.0^\circ$ $\theta_{\text{top}} = 17.5^\circ$, $\theta_{\text{bottom}} = -17.5^\circ$
isCenteredInAzimuth	This flag defines whether the production screen is frontal and centered in azimuth (absolute values of the azimuth angles of the left and right screen edge are identical) or not.
bsScreenSizeAz	This field defines the azimuth angles (in degree) corresponding to the left and right screen edge: $\varphi_{\text{left}} = 0{,}5\ \text{bsScreenSizeAz}$ $\varphi_{\text{left}} = \min(\max(\varphi_{\text{left}}, 0), 180)$ $\varphi_{\text{right}} = -0{,}5\ \text{bsScreenSizeAz}$ $\varphi_{\text{right}} = \min(\max(\varphi_{\text{right}}, -180), 0)$
bsScreenSizeLeftAz	This field defines the azimuth angle (in degree) corresponding to the left screen edge: $\varphi_{\text{left}} = 0{,}5\ (\text{bsScreenSizeLeftAz} - 511)$ $\varphi_{\text{left}} = \min(\max(\varphi_{\text{left}}, -180), 180)$
bsScreenSizeRightAz	This field defines the azimuth angle (in degree) corresponding to the right screen edge: $\varphi_{\text{right}} = 0{,}5\ (\text{bsScreenSizeRightAz} - 511)$ $\varphi_{\text{right}} = \min(\max(\varphi_{\text{right}}, -180), 180)$
bsScreenSizeTopEl	This field defines the elevation angle (in degree) corresponding to the top screen edge: $\theta_{\text{top}} = 0{,}5\ (\text{bsScreenSizeTopEl} - 255)$ $\theta_{\text{top}} = \min(\max(\theta_{\text{top}}, -90), 90)$
bsScreenSizeBottomEl	This field defines the elevation angle (in degree) corresponding to the bottom screen edge: $\theta_{\text{bottom}} = 0{,}5\ (\text{bsScreenSizeBottomEl} - 255)$ $\theta_{\text{bottom}} = \min(\max(\theta_{\text{bottom}}, -90), 90)$
HoaFrameLengthIdx	This element contains information about the HOA frame length *L*. See also Table 5.
CodedNumOfTransportChannels	This element contains information about the coded number of transport channels.
NumOfTransportChannels	This element contains information about the number of transport channels.
HOAConfig()	This element contains information about the configuration for HOA spatial encoding as defined in ISO/IEC 23008-3 [1] to [4], clause 12.3.
HOAConfig_SN3D()	This element contains information about the configuration for HOA spatial encoding as defined in clause 4.4.

In Table 3, the syntax of the frame data of Generic HOA transport format is defined as a binary representation format. In Table 4, the corresponding semantics of the frame data of Generic HOA transport format is defined.

Table 3: Syntax of HOATransportFormatFrame()

Syntax	No. of bits	Mnemonic

`HOATransportFormatFrame(HoaTransportType)`
`{`
`if (HoaTransportType == 1) {`
`HOAFrame();`
`} else if (HoaTransportType == 2) {`
`HOAFrame_SN3D();`
`} else if (HoaTransportType == 3) {`
`HOAFrame_VvecTransportFormat();`
`}`
 	
`for (j=0;j< HoaFrameLength;j++) {`
`for (i=0;i< NumOfTransportChannels;i++) {`
`htfCoreAudioChannels[i][j];`
`}`
`}`
`}`
	
	InputAudioBitDepth
		
		bslbf

Table 4: Semantics of HOATransportFormatFrame ()

HOAFrame()	The HOAFrame() holds the information that is required to decode the *L* samples of an HOA frame of N3D normalization as described in clause 4.3.
HOAFrame_SN3D()	The HOAFrame() holds the information that is required to decode the *L* samples of an HOA frame of SN3D normalization as described in clause 4.4.
HOAFrame_VvecTransportFormat()	The HOAFrame() holds the information that is required to decode the *L* samples of an HOA frame based on the V-vectors as described in clause 4.5.
NumOfTransportChannels	This element contains information about the number of transport channels defined in Table 1.
HoaFrameLength	This element contains information about the HOA frame length *L* defined in Table 5.
htfCoreAudioChannels[i][j]	This element contains information about the audio data of a *j*-th sample in an *i*-th transport channel.

Table 5: Value of HOA frame length in samples, *HoaFrameLength* (*L*), depending on • `InputSamplingFrequency` and `HoaFrameLengthIdx`

InputSamplingFrequency (kHz)	HoaFrameLengthIdx							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
24	192	256	384	480	512	768	960	1 024
32	256	384	512	640	832	1 024	1 280	1 366
44,1	384	512	768	960	1 024	1 536	1 920	2 048
48	384	512	768	960	1 024	1 536	1 920	2 048
96	768	1 024	1 536	1 920	2 048	3 072	3 840	4 096
192	1 536	2 048	3 072	3 840	4 096	6 144	7 680	8 192

4.3 ISO/IEC 23008-3-based HOA Transport Format (HoaTransportType = 1)

4.3.1 Introduction

This clause defines Type 1 of the HOA Transport Format (`HoaTransportType` = 1) which is based on ISO/IEC 23008-3 (MPEG-H 3D Audio) [1] to [4].