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

Copyright Notification

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

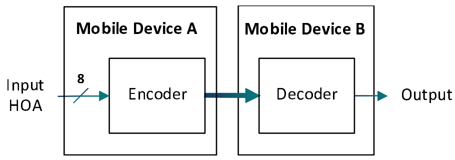
4.1 Introduction

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.

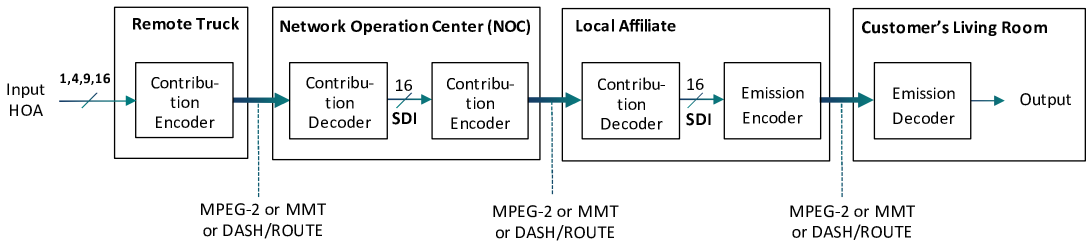
One use case is mobile devices where the number of input channels is limited by N Pulse-Code Modulation (PCM) channels. As shown in Figure 1 (a), if N is 8, a maximum of First Order Ambisonics (FOA which requires 4 PCM channels) can be transported.

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 3^{rd} 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 2^{nd} 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.



(a)



(b)

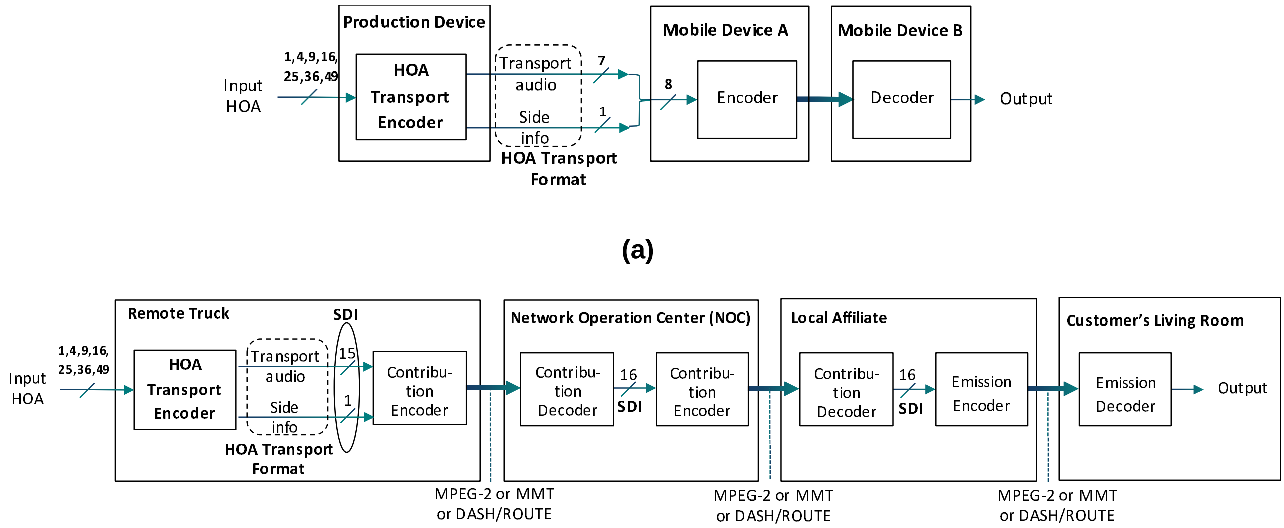
Figure 1: (a) conventional mobile devices and (b) conventional broadcast chain for order-restricted Ambisonics transport

4.2 Generic HOA Transport Format

To transport higher than 1^{st} 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 3^{rd} 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 (6^{th} 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].

4.3.2 HOA Transport Format defined in ISO/IEC 23008-3

In [2], the HOA input signal is analysed and encoded into the spatial coding parameters and the directional and ambient signals. The number of signals is usually lower than the number of HOA input coefficients. The HOA Frame Creater converts the resulting HOA spatial coding parameters to the HOA payloads HOAConfig() and HOAFrame().

In some environments (see e.g. Figure 2), the HOA spatial encoder is separated from the MPEG-H 3D Audio Core encoder. In this case, the HOA Transport Format consists of spatial coding parameters and the predominant and ambient signals. This HOA Transport Format can be transmitted from the HOA spatial encoder to the MPEG-H 3D Audio Core encoder. Compared with the input HOA, the HOA Transport Format usually requires a significantly reduced number of transport channels.

4.3.3 Implementation of HOA Transport Encoder (TE) and HOA Emission Encoder (EE)

Based on [2], the following terms are defined for simplicity:

- The combination of the **Spatial HOA Encoder** and the **HOA Frame Creater** is defined as the **HOA Encoder**.
- The predominant and ambient signals are defined as **HOA Transport Audio Signals**.
- The combination of the **HOAConfig** and **HOAFrame** is defined as the **HOA Side-info**.
- The combination of the **HOA Transport Audio Signals** and the **HOA Side-info** is defined as **HOA Transport Format**.

As shown in annexes A and B, there are several ways to design the broadcast chain. To make these systems working, it is beneficial to design the HOA Transport Encoder (TE) and the Emission Encoder (EE) such that:

- The bit-rate, the number of transport channels, and `hoaIndependencyFlag` are determined at EE. An `hoaIndependencyFlag` indicates whether a frame is independently decodable.
- Delay and complexity increase at TE and EE should be minimized.

Thus, three design criteria are defined:

1) Predominant audio channels (or called Foreground Audio (FG) channels): A V-vector represents the spatial distribution of the sound field for a particular vector-based predominant. FG audio channels and full V-vector elements are transmitted from TE to EE. At EE, a subset of FG audio channels and a subset of V-vector elements are selected and transmitted without any modification (no delay is required). If the EE modifies any FG audio channel, one frame delay is required for the adaptive gain correction (AGC) lookahead.
2) Ambient audio channels (or called background audio (BG) channels): As BG channels, H_BG, original HOA coefficients, H, are transmitted from TE to EE without applying any decorrelation and energy compensation: H_BG=H. At EE, a subset of BG channels is selected and transmitted without any modification (no delay is required). If the EE modifies any BG audio channel, one frame delay is required for the AGC lookahead.
3) To create random access points at EE, all the HOA Side-info parameter shall be encoded independently at TE. The predictive coding is not allowed at TE.

In the TE implementation, the total number of FGs and BGs are selected as 4 and 9, respectively. At EE, a subset of FGs and BGs is selected based on the EE bit rate. ChannelType is set to be 1 where Vector-based Signals are used to describe FGs or 2 where Additional Ambient HOA Coefficients are used to describe BGs. Table 6 describes different behaviours of FGs and BGs according to the parameter MinNumOfCoeffsForAmbHOA and codedVVecLength. The MinNumOfCoeffsForAmbHOA indicates the minimum number of ambient HOA coefficients for BG. The codedVVecLength value indicates the elements of the transmitted V-vector used to synthesize the vector-based signals:

0) Complete vector length (NumOfHoaCoeffs elements). Indicates that all of the coefficients for the predominant vectors (NumOfHoaCoeffs) are specified.
1) Vector elements 1 to MinNumOfCoeffsForAmbHOA and all elements defined in ContAddHoaCoeff are not transmitted.

2) Vector elements 1 to MinNumOfCoeffsForAmbHOA are not transmitted. Indicates that those coefficients of the predominant vectors corresponding to the number greater than a MinNumOfCoeffsForAmbHOA are specified.

Table 6: Different behaviour of FGs and BGs according to MinNumOfCoeffsForAmbHOA and codedVVecLength (the number of FGs and BGs are 4 and 9, respectively)

MinNumOfCoeffsForAmbHOA codedVVecLength	0	4	9
0	• H_BG = H - H_FG • Full V vectors • decorrMethod: not activated	• H_BG = H - H_FG • Full V vectors • decorrMethod = 1 (1~4)	• H_BG = H - H_FG • Full V vectors • decorrMethod = 1 (1~9)
1	• H_BG = H • No V for 1~9 • decorrMethod: not activated	• H_BG = H • No V for 1~9 • decorrMethod = 1 (1~4)	• H_BG = H • No V for 1~9 • decorrMethod = 1 (1~9)
2	• H_BG = H - H_FG • Full V vectors • decorrMethod: not activated	• H_BG = H for 1~4 • H_BG = H - H_FG for 5~9 • No V for 1~4 • decorrMethod = 1 (1~4)	• H_BG = H • No V for 1~9 • decorrMethod = 1 (1~9)

To meet the first two design criteria as outlined, the following conditions shall hold:

- For BG, H_BG = H instead of H_BG = H - H_FG. H_FG is the FG contribution to BG.
- For BG, decorrelation shall not be activated.
- For FG, full V-vector elements shall be transmitted.

Therefore, TE and EE are designed as shown in Figure 3.

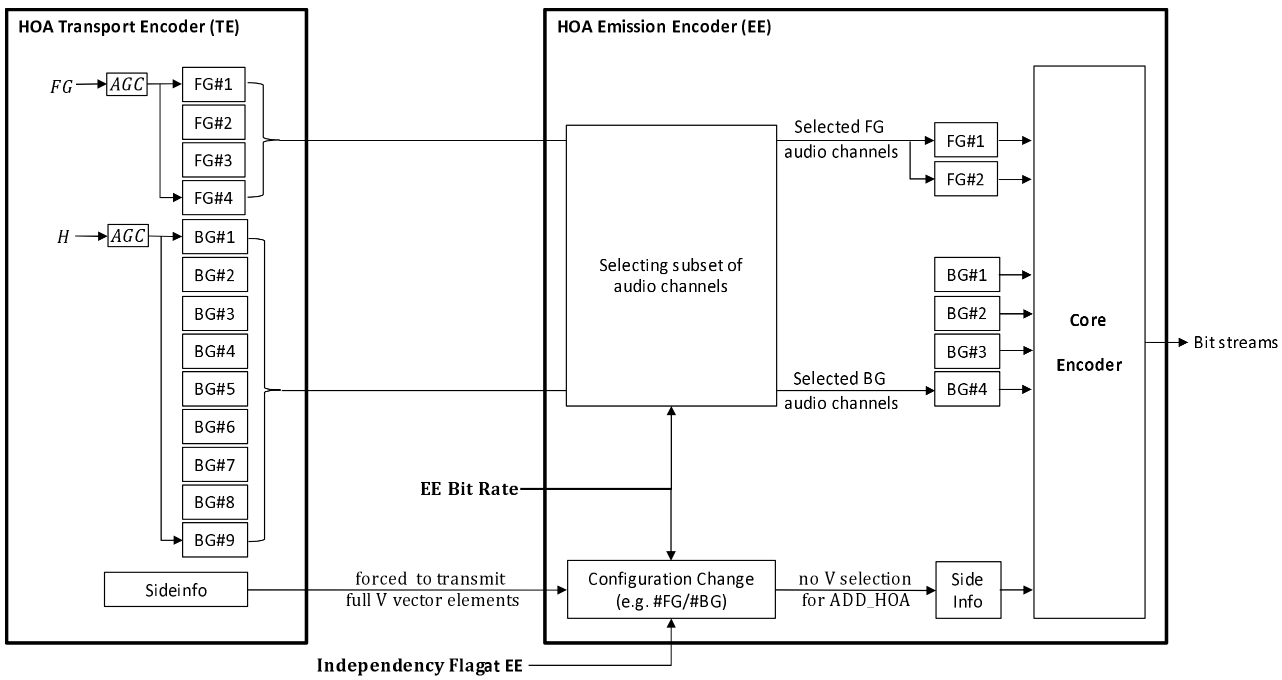


Figure 3: HOA Transport Encoder (TE) and Emission Encoder (EE) settings

At TE, codedVVecLength=1 and MinNumOfCoeffsForAmbHOA=0 are selected, 4 channels of FG and 9 channels of BG are transmitted to EE. Although codedVVecLength=1, full V-vectors are transmitted (instead of the elements between #BG+1 and #HOA coefficients).

At EE, codedVVecLength=1 and MinNumOfCoeffsForAmbHOA=0 are selected. Since codedVVecLength=1, only the V-vector elements between #BG+1 and #HOA coefficients are selected and transmitted. According to the bit rate at EE, the number of transport channels is determined for core encoding. If the subset of audio channels and the subset of HOA parameters are selected at EE without additional processing, there is no increase in complexity and delay for HOA spatial processing.

Table 7: Example of the numbers of FG and BG channels according to the HOA orders and bit rates

Order / Rate	1	2	3	4	5	6
Low	#FG=0 #BG=4	#FG=2 #BG=4	#FG=2 #BG=4	#FG=2 #BG=4	#FG=2 #BG=4	#FG=2 #BG=4
Mid	#FG=0 #BG=4	#FG=4 #BG=4	#FG=4 #BG=4	#FG=4 #BG=4	#FG=4 #BG=4	#FG=4 #BG=4
High	#FG=0 #BG=4	#FG=0 #BG=9	#FG=3 #BG=9	#FG=3 #BG=9	#FG=3 #BG=9	#FG=3 #BG=9

As an example, Table 7 shows the numbers of FG and BG channels according to the HOA orders and the bit rates. The total number of audio transport channels is the sum of the number of FG and BG channels.

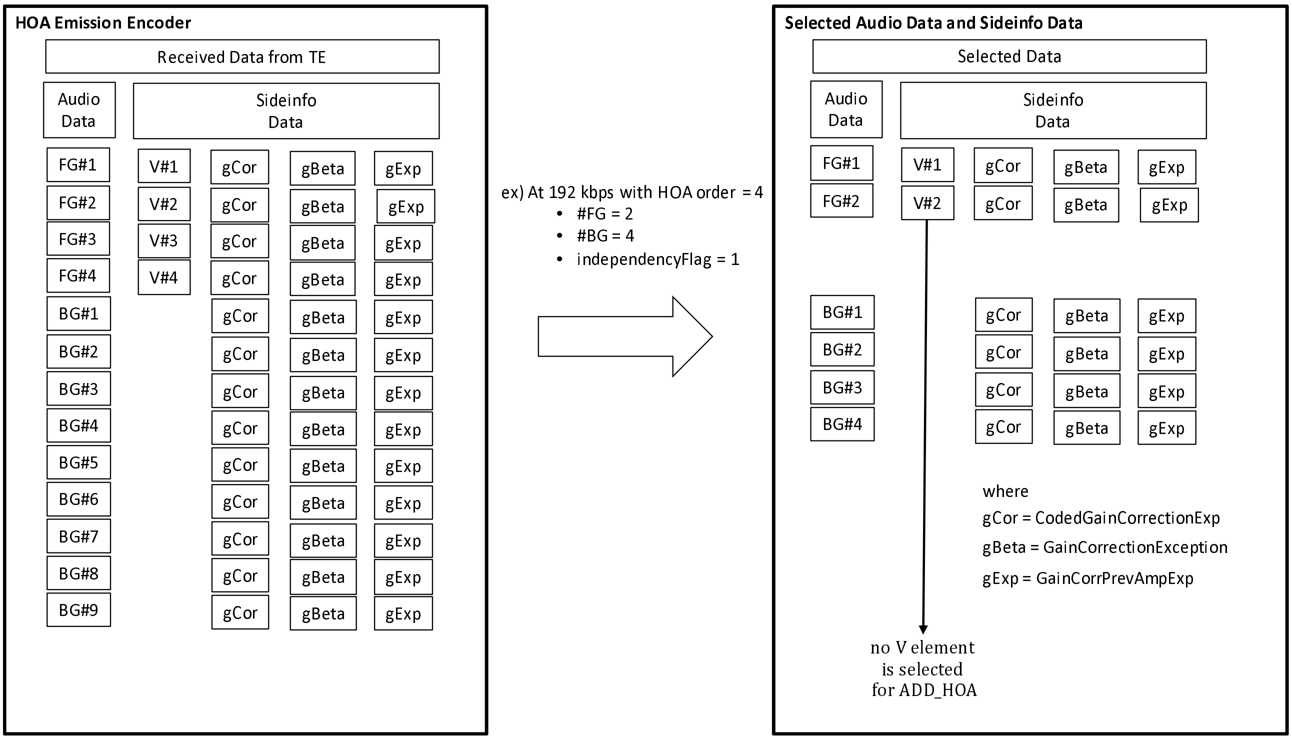


Figure 4: Example of HOA Transport Signals and HOA Side-info selection at the emission encoder

Example of HOA Transport Audio Signals and HOA Side-info selection at the emission encoder is shown in Figure 4:

- The numbers of FGs and BGs are determined based on Table 7. For rate = 192 kbps (low rate) and HOA order = 4, #FG=2 and #BG=4 are selected. Thus, from 4 FG and 9 BG audio channels transmitted from TE, first 2 FGs (FG#1 and FG#2) and 4 BGs (BG#1, BG#2, BG#3 and BG#4) are selected at EE. The selected audio channels are used as input to the core encoder without applying any decorrelation.

- Although codedVVecLength=1 and MinNumOfCoeffsForAmbHOA=0 at TE, TE transmits every element of V-vector. At EE, only the V-vector elements between 5 (=#BG+1) and 25 (=#HOA coefficients) are selected for the emission bit stream.

In summary, to implement the HOA Transport Encoder (TE) and the Emission Encoder (EE), the following configurations are used. More detailed description of each variable can be found at ISO/IEC 23008-3 [4].

- Configuration of TE:
    - `HoaOrder`: decided by the input HOA
    - `IsScreenRelative`: decided at TE
    - `UsesNfc`: content dependent:
        - This element determines whether or not the HOA Near Field Compensation (NFC) has been applied to the coded signal.
    - MinAmbHoaOrder = -
        - This element determines the minimum HOA order used for the coding of the ambient HOA representation.
    - numHOATransportChannels = 13
    - CodedSpatialInterpolationTime = 3
        - This element determines the time of the spatio-temporal interpolation of the Vector-based directional signals.
    - SpatialInterpolationMethod = 0
        - This element determines the spatial interpolation window.
    - CodedVVecLength = 1
    - ChannelType: decided at TE
        - This element stores the type of the i-th channel.
    - HOAFrameLengthIndicator = 1
        - Indicates the Frame Length *L* (number of samples) of the HOA Spatial Decoding relative to the Core Coder Frame Length.
    - hoaIndependencyFlag = 1
        - This flag signals that the current frame is an independent frame that can be decoded without having knowledge about the previous frame.
    - `NbitsQ` = 4~15
        - The type of dequantization of the V-vector is signalled by the word `NbitsQ`.
        - The `NbitsQ` value of 4 indicates vector-quantization.
        - When `NbitsQ` equals 5, a uniform 8 bit scalar dequantization is performed.
        - In contrast, an `NbitsQ` value of greater or equal to 6 indicates the application of Huffman decoding of a scalar-quantized V-vector.
- configurations of EE:
        - `HoaOrder`: sent from TE
        - `IsScreenRelative`: decided at EE (set to 0 to disable the Screen Relative functionality)
        - `UsesNfc` = sent from TE
        - `MinAmbHoaOrder` = -1
        - `numHOATransportChannels`: decided at EE
        - `CodedSpatialInterpolationTime` = 3

- SpatialInterpolationMethod = 0
- CodedVVecLength = 1
- ChannelType: sent from TE
- HOAFrameLengthIndicator = 1
- hoaIndependencyFlag = decided at EE
- NbitsQ = decided at EE

4.4 ISO/IEC 23008-3-based HOA Transport Format modified for SN3D Normalization (HoaTransportType = 2)

This clause defines Type 2 of the HOA Transport Format (HoaTransportType = 2). Implementation guidelines of HOA Transport Encoder (TE) and HOA Emission Encoder (EE) defined in clause 4.3.3 also apply for the HOA Transport Format Type 2.

The ISO/IEC 23008-3 syntax tables [4] are modified for implementing SN3D Normalization. Removing '_E' provides the same variable name as in the ISO/IEC 23008-3 syntax tables [4].

Table 8: Syntax of HOAConfig_SN3D

Syntax	No. of bits	Mnemonic
HOAConfig_SN3D()		
{		
NumOfHoaCoeffs_E = (**HoaOrder_E** + 1)^2;	**5**	**uimsbf**
IsScreenRelative_E;	**1**	**uimsbf**
HOADecoderConfig_SN3D(NumOfTransportChannels);		
}		

Table 9: Syntax of HOADecoderConfig_SN3D

Syntax	No. of bits	Mnemonic
HOADecoderConfig_SN3D(numHOATransportChannels)		
{		
MinAmbHoaOrder_E = **codedMinAmbHoaOrder_E** -1;	**3**	**uimsbf**
MinNumOfCoeffsForAmbHOA_E = (MinAmbHoaOrder_E + 1)^2;		
NumOfAdditionalCoders_E = numHOATransportChannels - MinNumOfCoeffsForAmbHOA_E;		
CodedSpatialInterpolationTime_E;	**3**	**uimsbf**
CodedVVecLength_E;	**2**	**uimsbf**
HOAFrameLengthIndicator_E;	**2**	**uimsbf**
if(MinAmbHoaOrder_E < HoaOrder_E) {		
DiffOrderBits_E = ceil(log2(HoaOrder_E - MinAmbHoaOrder_E +1));		
MaxHoaOrderToBeTransmitted_E = **DiffOrder_E** + MinAmbHoaOrder_E;	**DiffOrderBits_E**	**uimsbf**
}		
else {		
MaxHoaOrderToBeTransmitted_E = HoaOrder_E;		
}		
MaxNumOfCoeffsToBeTransmitted_E = (MaxHoaOrderToBeTransmitted_E + 1)^2;		
MaxNumAddActiveAmbCoeffs_E = MaxNumOfCoeffsToBeTransmitted _E - MinNumOfCoeffsForAmbHOA_E;		
VqConfBits_E =ceil(log2(ceil(log2(NumOfHoaCoeffs_E+1))));		
NumVVecVqElementsBits_E;	**VqConfBits_E**	**uimsbf**
AmbAsignmBits_E = ceil(log2(MaxNumAddActiveAmbCoeffs_E));		
}		
NOTE: HOAFrameLengthIndicator_E = 3 is reserved. CodedVVecLength_E = 3 is reserved.		

Table 10: Syntax of HOAFrame_SN3D

Syntax	No. of bits	Mnemonic
HOAFrame_SN3D()		
{		
NumOfVecSigs_E = 0;		
NumOfAddHoaChans_E = 0;		
hoaIndependencyFlag_E;	**1**	**uimsbf**
for(i=0; i< NumOfAdditionalCoders_E; ++i){		
ChannelSideInfoData_SN3D(i);		
switch ChannelType_E[i] {		
case 0:		
VecSigChannelIds_E[NumOfVecSigs_E] = i + 1;		
NumOfVecSigs_E++;		
break;		
case 1:		
AddHoaCoeff_E[NumOfAddHoaChans_E] = AmbCoeffIdx_E[i];		
NumOfAddHoaChans_E++;		
break;		
}		
default:		
}		
for(i=0; i< NumOfVecSigs_E; ++i){		
VVectorData_SN3D (VecSigChannelIds_E(i));		
}		
}		

Table 11: Syntax of ChannelSideInfoData_SN3D

Syntax	No. of bits	Mnemonic
ChannelSideInfoData_SN3D(i)		
{		
ChannelType_E[i]	**2**	**uimsbf**
switch ChannelType_E[i]		
{		
case 0:		
if(hoaIndependencyFlag_E){		
if(CodedVVecLength_E == 1){		
NewChannelTypeOne_E(k)[i];	**1**	**bslbf**
}		
NbitsQ_E(k)[i]	**4**	**uimsbf**
if (NbitsQ_E (k)[i] == 4) {		
CodebkIdx_E(k)[i];	**3**	**uimsbf**
NumVvecIndices_E(k)[i]++;	**NumVVecVqElementsBits_E**	**uimsbf**
}		
elseif (NbitsQ_E(k)[i] >= 6) {		
PFlag_E(k)[i] = 0;		
CbFlag_E(k)[i];	**1**	**bslbf**
}		
}		
else{		
if(CodedVVecLength_E == 1){		
NewChannelTypeOne_E(k)[i] =		
(0!=ChannelType_E(k-1)[i]);		
}		
bA_E;	**1**	**bslbf**
bB_E;	**1**	**bslbf**
if ((bA_E + bB_E) == 0) {		
NbitsQ_E(k)[i] = NbitsQ_E(k-1)[i];		
PFlag_E(k)[i] = PFlag_E(k-1)[i];		
CbFlag_E(k)[i] = CbFlag_E(k-1)[i];		
CodebkIdx_E(k)[i] = CodebkIdx_E(k-1)[i];		
NumVvecIndices_E(k)[i] = NumVvecIndices_E(k-1)[i];		
}		
else{		
NbitsQ_E(k)[i] = (8*bA_E)+(4*bB_E)+**uintC**_E;	**2**	**uimsbf**
if (NbitsQ_E(k)[i] == 4) {		
CodebkIdx_E(k)[i];	**3**	**uimsbf**
NumVvecIndices_E(k)[i]++;	**NumVVecVqElementsBits_E**	**uimsbf**
}		
elseif (NbitsQ_E(k)[i] >= 6) {		
PFlag_E(k)[i];	**1**	**bslbf**
CbFlag_E(k)[i];	**1**	**bslbf**
}		
}		
}		
break;		
case 1:		
AddAmbHoaInfoChannel_N3D(i);		
break;		
default:		
}		
}		
NOTE: CodebkIdx = 4 … 6 are reserved.		

Table 12: ChannelType definition

ChannelType:

- 0: Vector-based Signal
- 1: Additional Ambient HOA Coefficient
- 2: Empty
- 3: reserved

Table 13: AddAmbHoaInfoChannel_N3D definition

Syntax	No. of bits	Mnemonic
AddAmbHoaInfoChannel_N3D(i)		
{		
if(hoaIndependencyFlag_E){		
AmbCoeffTransitionState_E[i];	**2**	**uimsbf**
AmbCoeffIdx_E[i] = CodedAmbCoeffIdx_E + 1 + MinNumOfCoeffsForAmbHOA_E;	**AmbAsignmBits_E**	**uimsbf**
}		
else {		
if(**AmbCoeffIdxTransition_E** == 1) {	**1**	**bslbf**
if (AmbCoeffTransitionState_E[i] > 1) {		
AmbCoeffTransitionState_E[i] = 1;		
AmbCoeffIdx_E[i] = **CodedAmbCoeffIdx_E** + 1 + MinNumOfCoeffsForAmbHOA_E;	**AmbAsignmBits_E**	**uimsbf**
}		
else {		
AmbCoeffTransitionState_E[i] = 2;		
}		
}		
else {		
AmbCoeffTransitionState_E[i] = 0;		
}		
}		
}		

NOTE: The AmbCoeffIdx_E of the preceding frame is used under the following conditions:
if (AmbCoeffIdxTransitionState_E == 0 || AmbCoeffIdxTransitionState_E == 2)

AmbCoeffTransitionState_E:

0: No transition (continuous Additional Ambient HOA Coefficient)
1: Fade-in of Additional Ambient HOA Coefficient
2: Fade-out of Additional Ambient HOA Coefficient
3: reserved

Table 14: Syntax of VVectorData_SN3D

Syntax	No. of bits	Mnemonic
VVectorData_SN3D(i)		
{		
if (CodedVVecLength_E == 1) {		
VVecLengthUsed_E = VVecLength_E[i];		
VVecCoeffIdUsed_E = VVecCoeffId_E[i];		
} else {		
VVecLengthUsed_E = VVecLength_E;		
VVecCoeffIdUsed_E = VVecCoeffId_E;		
}		
if (NbitsQ_E(k)[i] == 4) {		
if (NumVvecIndices_E(k)[i] == 1) {		
VvecIdx_E[0] = **VvecIdx_E** + 1;	**nbitsIdx_E**	**uimsbf**
WeightVal_E[0] = ((**SgnVal_E** * 2)-1);	**1**	**uimsbf**
} else {		
WeightIdx_E;	**8**	**uimsbf**
for (j=0; j< NumVvecIndices_E(k)[i]; ++j) {		
VvecIdx_E[j] = **VvecIdx_E** + 1;	**nbitsIdx_E**	**uimsbf**
if (j<8) {		
WeightVal_E[j] = ((**SgnVal_E** *2)-1) *	**1**	**uimsbf**
WeightValCdbk_E[WeightIdx_E][j];		
} else {		
WeightVal_E[j] = ((**SgnVal_E** *2)-1) *	**1**	**uimsbf**
WeightValCdbk_E[WeightIdx_E][6+j%2];		
}		
}		
}		
}		
else if (NbitsQ_E(k)[i] == 5) {		
for (m=0; m< VVecLengthUsed_E; ++m){		
aVal_E[i][m] = (**VecVal_E** / 128.0) - 1.0;	**8**	**uimsbf**
}		
else if(NbitsQ_E(k)[i] >= 6) {		
for (m=0; m< VVecLengthUsed_E; ++m){		
huffIdx_E = huffSelect(VVecCoeffIdUsed_E[m],		
PFlag_E[i], CbFlag_E[i]);		
cid_E = huffDecode(NbitsQ_E[i], huffIdx_E,		**vlclbf**
huffVal_E);	**1..11**	
aVal_E[i][m] = 0.0;		
if (cid_E > 0) {		**bslbf**
aVal_E[i][m] = sgn_E = (**sgnVal_E** * 2) - 1;	**1**	
if (cid_E > 1) {		**uimsbf**
aVal_E[i][m] = sgn_E * (2.0^(cid_E -1)		
+ **intAddVal_E**);	**cid_E -1**	
}		
}		
}		
}		
}		

The variables in Table 15 are defined in ISO/IEC 23008-3 [4]. Removing '_E' provides the same variable name as in the ISO/IEC 23008-3 [4].

Table 15: Semantics of HOAConfig_SN3D()

HoaOrder_E	This element determines the HOA order of the coded signal.
NumOfHoaCoeffs_E	This element determines the number of HOA coefficients of the coded HOA representation, which is equal to the number of HOA coefficients to be reconstructed.
IsScreenRelative_E	This element indicates if the HOA representation shall be rendered with respect to the reproduction screen size as described in ISO/IEC 23008-3 [4], clause 18.4. 0: not screen related 1: screen related

MinAmbHoaOrder_E	This element determines the minimum HOA order used for the coding of the ambient HOA representation by MinAmbHoaOrder_E = **codedMinAmbHoaOrder_E** -1. The value -1 indicates that the number of decorrelated ambiance coefficients is equal to zero. Thus the HOA representation is transmitted without a residual ambiance HOA representation of lower order and all transport channels have a flexible channel type. The value 6 is used to extend the HOA order signaling.
MinNumOfCoeffsForAmbHOA_E	This element determines the minimum number of ambient HOA coefficients.
NumOfAdditionalCoders_E	This element determines the number of additional transport channels used for coding the directional and/or additional HOA coefficients of the ambient component. These transport channels have a flexible ChannelType_E.
CodedVVecLength_E	This element indicates the length of the transmitted data vector used to synthesize the vector-based signals.
MaxHoaOrderToBeTransmitted_E	This element indicates the maximum HOA order of the additional ambient HOA coefficients to be transmitted.
MaxNumOfCoeffsToBeTransmitted_E	This element indicates the maximum number of HOA coefficients to be transmitted, computed depending on MaxHoaOrderToBeTransmitted_E.
MaxNumAddActiveAmbCoeffs_E	This element signals the maximum index for the signaling of additional ambient HOA coefficients.
VqConfBits_E	This element indicates the number of bits necessary to signal the element **NumVVecVqElementsBits_E**.
NumVVecVqElementsBits_E	This element indicates the number of bits used to signal the element **NumVvecIndices_E** in ChannelSideInfoData_SN3D().
NumOfVecSigs_E	This element determines the number of active vector-based signals in the current HOAFrame_SN3D().
NumOfAddHoaChans_E	This element determines the total number of additional ambient HOA channels in the current HOAFrame_SN3D().
hoaIndependencyFlag_E	This flag signals that the current frame is an independent frame that can be decoded without having knowledge about the previous frame.
ChannelType_E[i]	This element stores the type of the i-th channel which is defined in Table ChannelType_SN3D().
VecSigChannelIds_E[NumOfVecSigs]	This element stores the channel index of each active vector-based signal of the current frame.
AmbCoeffIdx_E[i]	This element determines the index of the HOA signal where channel i contributes to the reconstructed HOA representation.
NewChannelTypeOne_E[i]	This flag indicates if in the previous frame (k-1) the transport channel was not initialized as a Vector-based Signal.
NbitsQ_E[i]	The NbitsQ_E[i] value determines the decoding method of the V-Vector associated with the Vector-based signal of the i-th channel. An NbitsQ_E[i] value of 4 determines the decoding of a vector-quantized V vector. The value 5 determines the decoding of a uniform 8bit scalar quantized V-vector. If the value is greater than 5, Huffman decoding of the V-Vector is determined.
CodebkIdx_E[i]	Signals the specific codebook used to dequantize the vector-quantized V vector associated with the Vector-based signal of the i-th channel.
NumVvecIndices_E(k)[i]	The number of vectors used to dequantize a vector-quantized V-vector.
PFlag_E[i]	The prediction flag used for the Huffman decoding of the scalar quantized V vector associated with the Vector-based signal of the i-th channel.
CbFlag_E[i]	The codebook flag used for the Huffman decoding of the scalar quantized V vector associated with the Vector-based signal of the i-th channel.
bA_E, bB_E	The msb (bA_E) and second msb (bB_E) of the NbitsQ_E [i] field.
uintC_E	The code word of the remaining two bits of the NbitsQ_E [i] field.

AmbCoeffTransitionState_E[i]	This decoder-internal variable tracks the state of the life-cycle of an additional ambient HOA coefficient. Those states are Fade-in, Continuous State, and Fade-out. The AmbCoeffIdxTransition_E signals a change of the state in the bitstream. When an additional ambient HOA coefficient is faded in, the CodedAmbCoeffIdx_E word is sent to signal the new AmbCoeffIdx_E. In all other states, the AmbCoeffIdx_E of the previous frame is used.
AmbCoeffIdxTransition_E	This element indicates that in this frame an additional ambient HOA coefficient is either being faded in or faded out. This flag will update the decoder-internal AmbCoeffTransitionState_E variable for this transport channel accordingly.
CodedAmbCoeffIdx_E	This element reads the coded index of the additional ambient HOA coefficient.
VVec_E(k)[i]	This is the V vector for the k-th frame for the i-th channel.
VVecLength_E	This variable indicates the number of vector elements to read out.
VVecCoeffId_E	This vector contains the indices of the transmitted V-vector coefficients.
VecVal_E	An integer value between 0 and 255.
aVal_E	A temporary variable used during decoding of the VVectorData_E.
huffVal_E	A Huffman code word, to be Huffman-decoded.
sgnVal_E	This is the coded sign value used during decoding.
intAddVal_E	This is additional integer value used during decoding.
WeightIdx_E	The index in WeightValCdbk_E used to dequantize a vector-quantized V vector.
nBitsQ_E	Field size for reading **WeightIdx_E** to decode a vector-quantized V vector.
WeightValCdbk_E	Codebook which contains a vector of positive real-valued weighting coefficients. Only necessary if NumVvecIndices_E is > 1.
VvecIdx_E	An index used to dequantize a vector-quantized V vector.
nbitsIdx_E	Field size for reading **VvecIdx_E** to decode a vector-quantized V vector.
WeightVal_E	A real-valued weighting coefficient to decode a vector-quantized V vector.

The variables in Table 16 are newly defined for the Type 2 HOA Transport Format.

Table 16: New semantics of HOAConfig_SN3D()

HOAFrameLengthIndicator_E	Indicates the Frame Length *L* (number of samples) of the HOA Spatial Decoding relative to the Core Coder Frame Length as defined in Table 17.
outputFrameLength_E	The number of samples per frame, using the original sampling frequency.
CodedSpatialInterpolationTime_E	This element determines the time of the spatio-temporal interpolation of the Vector-based directional signals as defined in Table 18.

Table 17: Value of HOA frame length *L* depending on `HOAFrameLengthIndicator_E` and `outputFrameLength_E`

outputFrameLength_E	**HOAFrameLengthIndicator_E**			
	0	**1**	**2**	**3**
768	768 samples	768 samples	768 samples	reserved
1 024	1 024 samples	1 024 samples	1 024 samples	20 ms*
2 048	2 048 samples	1 024 samples	1 024 samples	reserved
4 096	4 096 samples	2 048 samples	1 024 samples	reserved
N ms* = ceil ((sampling frequency * *N*)) in samples				

Table 18: Decoding of `CodedSpatialInterpolationTime_E`

L	**`CodedSpatialInterpolationTime_E`**							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
768	0	32	64	128	256	384	512	768
1 024	0	64	128	256	384	512	768	1 024
2 048	0	128	256	512	768	1 024	1 536	2 048
20 ms*	0	1,25 ms	2,5 ms	5 ms	7,5 ms	10 ms	15 ms	20 ms
N ms* = ceil((sampling frequency * *N*)) in samples								

Due to the signal normalization of SN3D the following decoder configuration is required:

- To decorrelate the first MinNumOfCoeffsForAmbHOA coefficients of the ambient HOA component the Ambience Synthesis processing includes an inverse spatial transform based on the mode matrix $\boldsymbol{\Psi}$**.** This mode matrix shall be SN3D normalized.
- When decoding vector-quantized V-vectors, a V-vector is represented by a weighted summation of the V-vector code vectors. These code vectors shall be SN3D normalized.

4.5 V-vector based HOA Transport Format (HoaTransportType = 3)

This clause defines the V-vector based HOA Transport Format (VHTF) (`HoaTransportType` = 3).

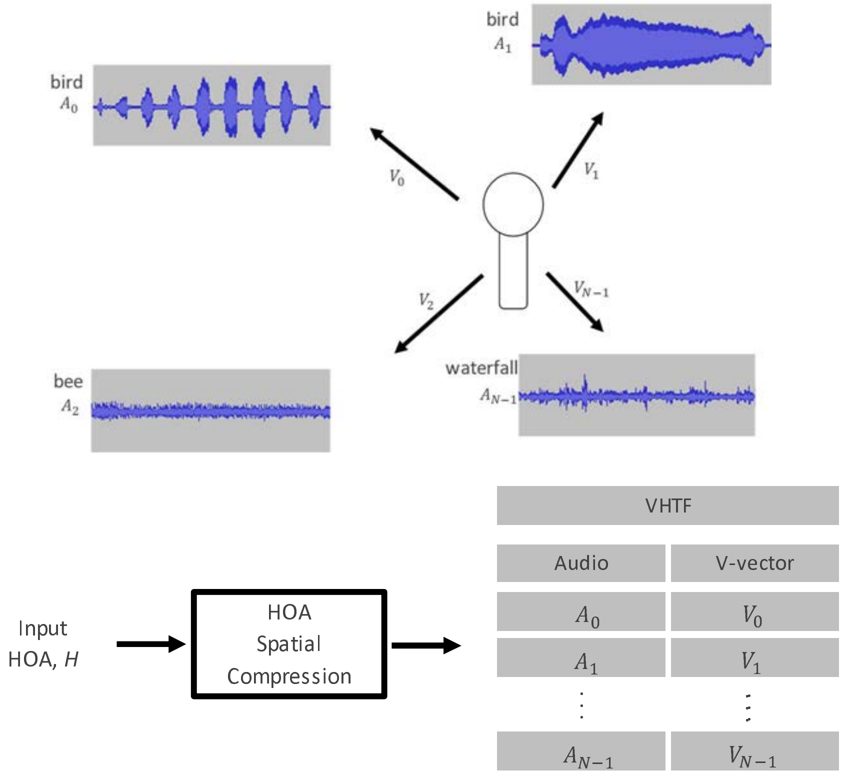


Figure 5: VHTF representation of HOA

As shown in Figure 5, VHTF is composed of audio signals, $\{A_i\}$, and the associated V-vectors, $\{V_i\}$. An input HOA signal, H, can be approximated by:

$$\widetilde{H} = \sum_{i=0}^{N-1} A_i V_i^T$$

where an i-th V-vector, V_i, is the spatial representation of the i-th audio signal, A_i. N is the number of transport channels. The dynamic range of each V_i is bound by [-1, 1]. Examples of V-vector based spatial representation are given in Figure 6.

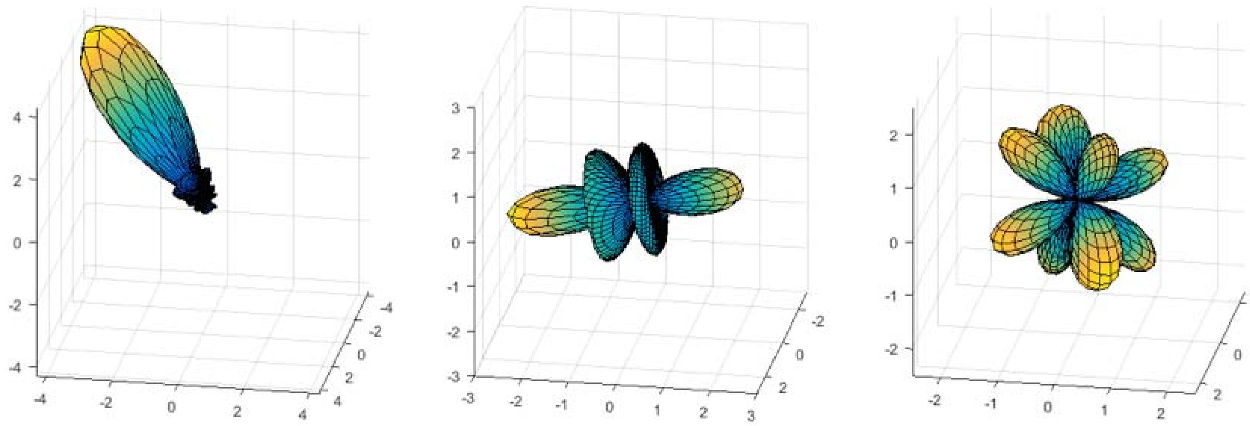


Figure 6: Examples for different spatial beam representations that can be expressed by V-vectors

VHTF can also represent an original input HOA, which means $\widetilde{H} = H$, in the following conditions:

- if V_i has all zero elements but a value of one at an i-th element index, i.e. $[0\ 0\ \cdots 1\ \cdots 0]^T$;
- and if A_i is the i-th HOA coefficients.

Thus, VHTF can represent both pre-dominant and ambient sound field elements.

As shown in Table 19, the HOAFrame_VvecTransportFormat() holds the information that is required to decode the *L* samples (HoaFrameLength in Table 1) of an HOA frame.

Table 19: Syntax of HOAFrame_VvecTransportFormat()

Syntax	No. of bits	Mnemonic
HOAFrame_VvecTransportFormat()		
{		
VvectorBits = **codedVvectorBitDepth***2+2;	**3**	**uimsbf**
PriorityBits = ceil(log2(NumOfTransportChannels));		
for (i=0;i<NumOfTransportChannels;i++) {		
priorityOfTC[i];	PriorityBits	**uimsbf**
interpolationOfTC[i];	**1**	**bslbf**
for (j=0;j<NumOfHoaCoeffs; j++){		
Vvector[i][j]=((**VvecVal**+1)*2^(1-VvectorBits))-1.0;	VvectorBits	**uimsbf**
}		
}		
}		

Table 20: Semantics of HOAFrame_VvecTransportFormat()

NumOfTransportChannels	This element contains information about the number of transport channels defined in Table 1.
codedVvectorBitDepth	This element contains information about the coded bit depth of a V-vector.
NumOfHoaCoeffs	This element contains information about the number of HOA coefficients defined in Table 1.
VvectorBits	This element contains information about the bit depth of a V-vector.
PriorityBits	This element contains information about the bit depth of HOA transport channel priority.
priorityOfTC[i]	This element contains information about the priority of the i-th transport channel (the channel with a lower priority value is more important, thus the channel with priorityOfTC[i] = 0 is the channel with the highest priority).
interpolationOfTc[i]	This flag indicates for i-th transport channel if the spatio-temporal interpolation with the VVector[i] of the previous frame shall be performed (interpolationOfTc[i]=1). If interpolationOfTc[i]=0, no spatio-temporal interpolation shall be performed. This spatio-temporal interpolation for an *n*-th frame and an *i*-th transport channel is performed as: $$H_i(n) = \left(w_{in} \cdot A_i(n)\right)V_i^T(n) + \left(w_{out} \cdot A_i(n)\right)V_i^T(n-1)$$ where w_{in} and w_{out} are the fade-in and fade-out windows as defined as: $$w_{in} = 0{,}5\left[1 - \cos\left(\frac{l}{L}\pi\right)\right] \quad l = [0, \ldots, L-1]$$ $$w_{out} = 0{,}5\left[1 - \cos\left(\frac{l+L}{L}\pi\right)\right] \quad l = [0, \ldots, L-1].$$
VvecVal	An integer value between 0 and $2^{VvectorBits}$-1.
Vvector[i][j]	This element contains information about a vector element representing spatial information. Its value is bounded by [-1,1].

5 HOA Transport Format Audio Stream

5.1 Introduction

This clause defines an HOA Transport Format Audio Stream (HTFAS) which is a self-contained audio stream format to transport HTF data. Similar to MPEG-H 3D audio stream (MHAS), the transport mechanism uses a packetized approach [4]. Both, configuration data as well as coded payload data is embedded into separate packets. Synchronization and length information is added to enable a self-synchronizing syntax.

5.2 Syntax of HOA Transport Format Audio Stream

In Tables 21, 22 and 23, the syntax elements of HTFAS are defined. In Table 24, the corresponding semantics of HTFAS is defined.

Table 21: Syntax of HOATransportFormatAudioStream()

Syntax	No. of bits	Mnemonic

HOATransportFormatAudioStream()
{
   while (bitsAvailable() != 0) {
      HOATransportFormatAudioStreamPacket();
   }
}		

Table 22: Syntax of HOATransportFormatAudioStreamPacket ()

Syntax	No. of bits	Mnemonic
HOATransportFormatAudioStreamPacket() {		
HTFASPacketType = escapedValue(3,8,8);	**3,11,19**	**uimsbf**
HTFASPacketLabel = escapedValue(2,8,32);	**2,10,42**	**uimsbf**
HTFASPacketLength = escapedValue(11,24,24);	**11,35,59**	**uimsbf**
HTFASPacketPayload(HTFASPacketType); }		

Table 23: Syntax of HTFASPacketPayload ()

Syntax	No. of bits	Mnemonic
HTFASPacketPayload(HTFASPacketType)		
{		
switch (HTFASPacketType) {		
case HTF_PACTYP_SYNC:		
0xA5; /* syncword */	**8**	**uimsbf**
break;		
case HTF_PACTYP_HTFCFG:		
HoaTransportType;	**5**	**uimsbf**
HOATransportFormatConfig(HoaTransportType);		
break;		
case HTF_PACTYP_HTFFRAME:		
HOATransportFormatFrame(HoaTransportType);		
break;		
case HTF_PACTYP_FILLDATA:		
for (i=0; i< HTFASPacketLength; i++) {		
htfas_fill_data_byte(i);	**8**	**bslbf**
}		
break;		
case HTF_PACTYP_SYNCGAP:		
syncSpacingLength = escapedValue(16,24,24);	**16,40,64**	**uimsbf**
break;		
case HTF_PACTYP_CRC16:		
htfasParity16Data;	**16**	**bslbf**
break;		
case HTF_PACTYP_CRC32:		
htfasParity32Data;	**32**	**bslbf**
break;		
case HTF_PACTYP_DESCRIPTOR:		
for (i=0; i< HTFASPacketLength; i++) {		
htfas_descriptor_data_byte(i);	**8**	**bslbf**
}		
break;		
case HTF_PACTYP_AUDIOTRUNCATION:		
isActive_AUDIOTRUNCATION;	**1**	**bool**
if (isActive_AUDIOTRUNCATION){		
truncFromBegin;	**1**	**bool**
nTruncSamples;	**13**	**uimsbf**
}		
break;		
}		
ByteAlign();		
}		
NOTE: A packet with HTFASPacketType 'HTF_PACTYP_HTFCFG' should be transmitted before packets with HTFASPacketType 'HTF_PACTYP_HTFFRAME'.		

Table 24: Semantics of HOATransportFormatAudioStream()

HTFASPacketType	This element contains information about the payload type in the actual packet. A decoder which does not support a certain HTFASPacketType shall skip this packet and continue with the next packet. The values of HTFASPacketType are given in Table 25.
HTFASPacketLabel	This element contains information about an indication of which packets belong together.
HTFASPacketLength	This element contains information about the length of the HTFASPacketPayload () in Bytes.
HoaTransportType	This element contains information about the HOA transport mode: 0: HOA coefficients as defined in Table 1. 1: ISO/IEC 23008-3 [1] to [4] based HOA Transport Format as defined in clause 4.3. 2: Modified ISO/IEC 23008-3 [1] to [4] 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.

htfTimeStamp	This element contains information about the time stamp.
HOATransportFormatConfig()	An HTF configuration structure as defined in Table 1.
HOATransportFormatFrame()	An HTF frame payload as defined in clauses 4.2, 4.3, 4.4 and 4.5.
htfas_fill_data_byte	8-bit data elements, no restrictions apply.
syncSpacingLength	This element contains information about the length in Bytes between the last two packets with HTF_PACTYP_SYNC.
htfasParity16Data	This element contains information about a 16-bit field that contains the CRC value that yields a zero output of the 16 registers in the decoder with the polynomial: $x^{16} + x^{15} + x^{5} + 1$ and the initial state of the shift register of 0xFFFF.
htfasParity32Data	This element contains information about a 32-bit field that contains the CRC value that yields a zero output of the 32 registers in the decoder with the polynomial: $x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^{8} + x^{7} + x^{5} + x^{4} + x^{2} + x + 1$ and the initial state of the shift register of 0xFFFFFFFF.
htfas_descriptor_data_byte	This element contains information used to embed MPEG-2 TS/PS descriptors in the streams. Data conveyed as htfas_descriptor_data_byte has the same syntax and semantics as defined for descriptor() in ISO/IEC 13818-1 [5].
isActive_AUDIOTRUNCATION	If 1 the audio truncation message is active, if 0 the decoder should ignore the message.
truncFromBegin	If 0 truncate samples from the end, if 1 truncate samples from the beginning.
nTruncSamples	number of samples to truncate.
ByteAlign()	Up to 7 fill bits to achieve byte alignment with respect to the beginning of the syntactic element in which ByteAlign() occurs.

Table 25: Value of HTFASPacketType

HTFASPacketType	Value
HTF_PACTYP_FILLDATA	0
HTF_PACTYP_HTFCFG	1
HTF_PACTYP_HTFFRAME	2
/* reserved */	3 - 5
HTF_PACTYP_SYNC	6
HTF_PACTYP_SYNCGAP	7
/* reserved */	8
HTF_PACTYP_CRC16	9
HTF_PACTYP_CRC32	10
HTF_PACTYP_DESCRIPTOR	11
/* reserved */	12 - 16
HTF_PACTYP_AUDIOTRUNCATION	17
/* reserved */	18 - max value

5.3 Application Examples of HOA Transport Format Audio Stream

A simple broadcast can be implemented as shown in Figure 7.

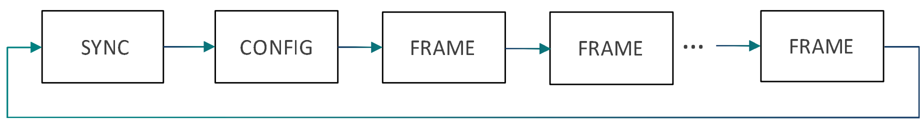


Figure 7: HTFAS - Example 1

As in [1] to [4], then embedding HTFAS into MPEG-2 Transport streams, fast synchronization to a stream at random access points is most important, while bitrate overhead is usually less critical. To improve synchronization to the stream, packets of type HTF_PACTYP_SYNC may be embedded more frequently and in addition the HTF_PACTYP_SYNCGAP type may also be embedded. Figures 8 and 9 indicate some possible solutions.

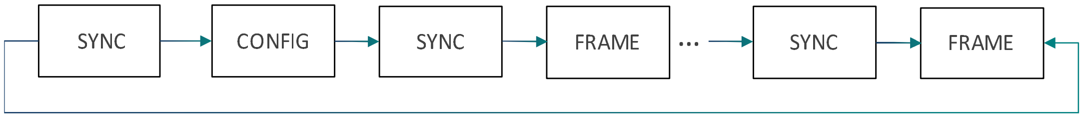


Figure 8: HTFAS - Example 2

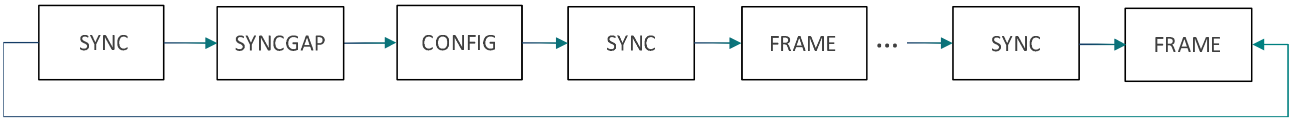


Figure 9: HTFAS - Example 3

Annex A (informative): Example guidelines for implementing HOA transport over SDI utilizing communications modem technologies

This annex presents guidelines about HOA transport over Serial Digital Interface (SDI) utilizing communications modem technologies [i.1]. This annex gives an example with Type 1 of the HOA Transport Format (`HoaTransportType` = 1), but it does not have to be limited by this HOA Transport Format. In a typical broadcast workflow, the use of single High-Definition SDI (HD-SDI) link has a limitation of being able to transport only 16 Pulse-Code Modulation (PCM) audio embedded channels. This restricts the transport to a maximum of 3rd order HOA signals (requiring 16 PCM channels) - and that, only if there are no additional data 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.

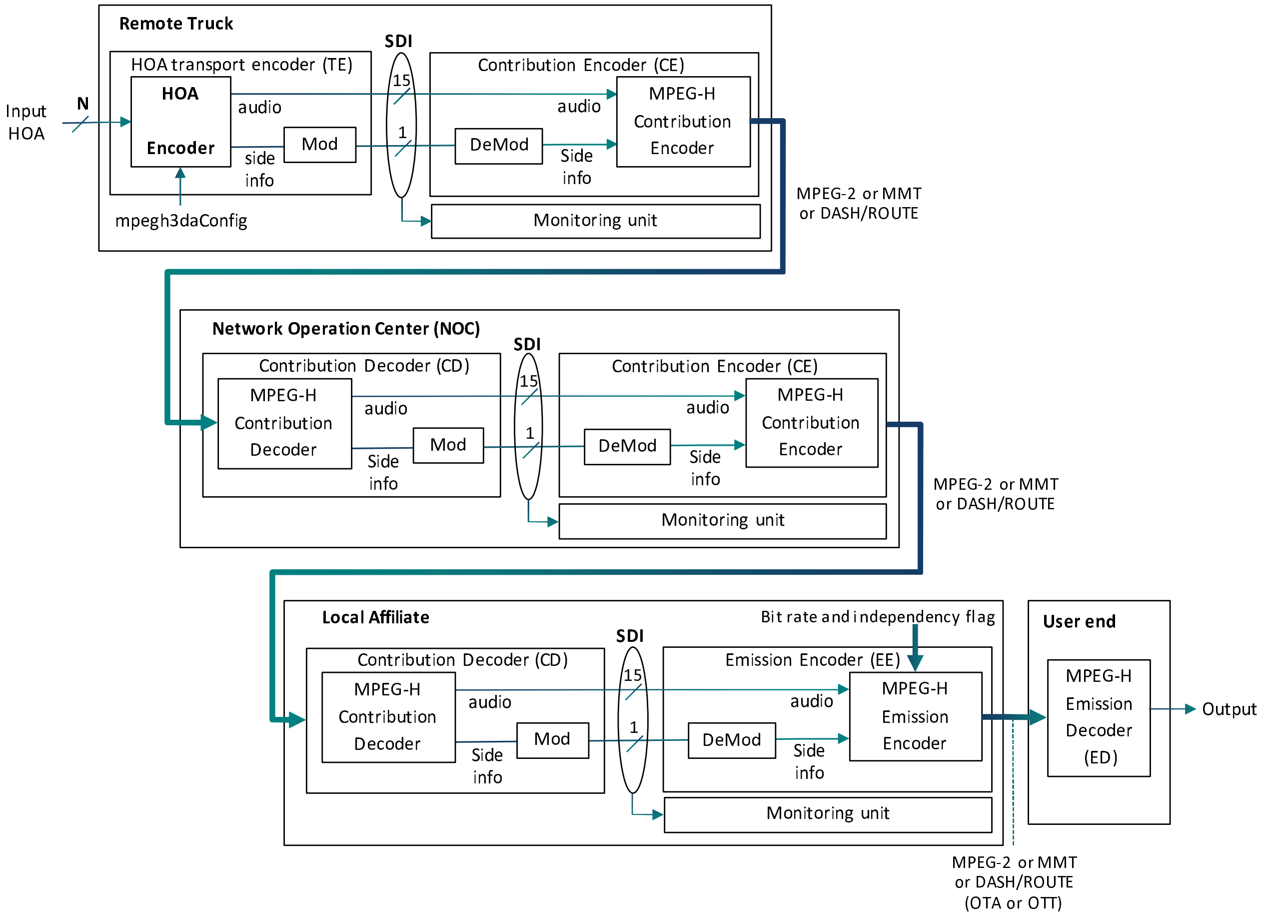


Figure A.1: Block diagram of broadcast chain type-I that transports HOA signals (for simplicity, conversion modules from/to MPEG-2 or MMT or DASH/ROUTE are not described)

The block diagram of broadcast chain that transports N-channel HOA signals is shown in Figure A.1.

At a remote truck, the HOA transport encoder converts the *N*-channel HOA input coefficients into 16 PCM audio channels. The HOA transport encoder consists of the HOA encoder and Side-info modulation block:

- The HOA encoder encodes the HOA input signal into the HOA transport audio signals and the HOA Side-info. More detailed description can be found in clause C.5 of [2].
- For error protection in SDI transmission, the HOA Side-info is modulated with communications modem technologies into a PCM control track signal that fits into the audio signal bandwidth.
- A combination of the HOA transport audio signals and the modulated HOA Side-info signal is called as HOA transport format for SDI transmission. In Figure A.1, 15 audio streams and a single Side-info stream are transported through HD-SDI.

At the contribution encoder, a demodulation block converts the modulated Side-info PCM data into the HOA Side-info binary data. The HOA Side-info and the 15 audio streams are provided to the MPEG-H 3DA contribution encoder as input signals. As described in the section 4.7 of [1], the Contribution Mode of MPEG-H 3DA specifies a generic transport mechanism for audio signals with accompanying metadata. It is designed to be unaware of the signal type and of the content and structure of the associated metadata. A value of speakerLayoutType == 3 in the signaling of the referenceLayout in the mpegh3daConfig() indicates that MPEG-H 3D audio shall operate in Contribution Mode. In Contribution Mode, the rendering context shall operate in a pass-through mode. Then, MPEG-2 TS, MMT, or DASH/ROUTE can be used as a transport format between a remote truck and a network operating centre (NOC). This conversion is not described in Figure A.1 for simplicity.

At NOC, the contribution decoder reconstructs the HOA transport format for SDI transmission. Between the contribution decoder and encoder at NOC, several operations can be performed including program switching and monitoring.

At local affiliate, the contribution decoder also reconstructs the HOA transport format for SDI transmission. The HOA emission encoder is depicted in Figure A.2. Emission bit-rate and hoaIndependencyFlag are determined at the emission encoder. Based on a given emission bit-rate:

1) a subset of audio channels is selected for MPEG-H 3DA core coding; and

2) a subset of HOA Side-info is selected for transmission.

Then, the bit-stream generated from the MPEG-H 3DA core encoder will be transmitted to the users with over-the-top (OTT) streaming services or over-the-air (OTA) TV services.

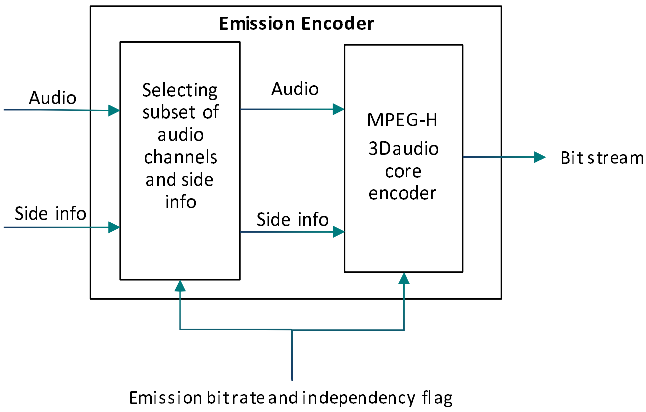


Figure A.2: Block diagram of HOA emission encoder

For example, a 6th order HOA input can be transported from a remote truck to a local affiliate as follows:

1) At a remote truck, the HOA encoder converts the input 49 PCM channels to the HOA transport format for SDI, which contains 13 HOA transport audio signals along with a single HOA Side-info. The HOA Side-info binary data is modulated into a PCM channel. The 16 channel HD-SDI can carry these (13+1) PCM channels and 2 empty channels to the contribution encoder.

2) The contribution encoder demodulates Side-info data, which are encoded with MPEG-H 3DA contribution mode.

3) At the NOC and local affiliates, the contribution decoder reconstructs the 16 SDI channel PCM data. The contribution encoder demodulates Side-info data, which are encoded with MPEG-H 3DA contribution mode.

4) The emission encoder controls the emission bit rate, hoaIndependencyFlag, and the number of transport channels for the emission to the users. If hoaIndependencyFlag == 1 then the present frame shall be independently decodable. From 16 channels, 13 transport audio channels and 1 demodulated control data are extracted. Based on the configuration (e.g. emission bit rate, hoaIndependencyFlag) at the emission encoder, a subset of audio channels is selected for core encoding and a subset of the HOA side info is selected as an HOA control data. Bit-stream packetization is followed for final emission.

Annex B (informative): Example guidelines for HOA production

This annex provides example guidelines for HOA production. This annex gives an example with Type 1 of the HOA Transport Format (`HoaTransportType` = 1), but it does not have to be limited by this HOA Transport Format.

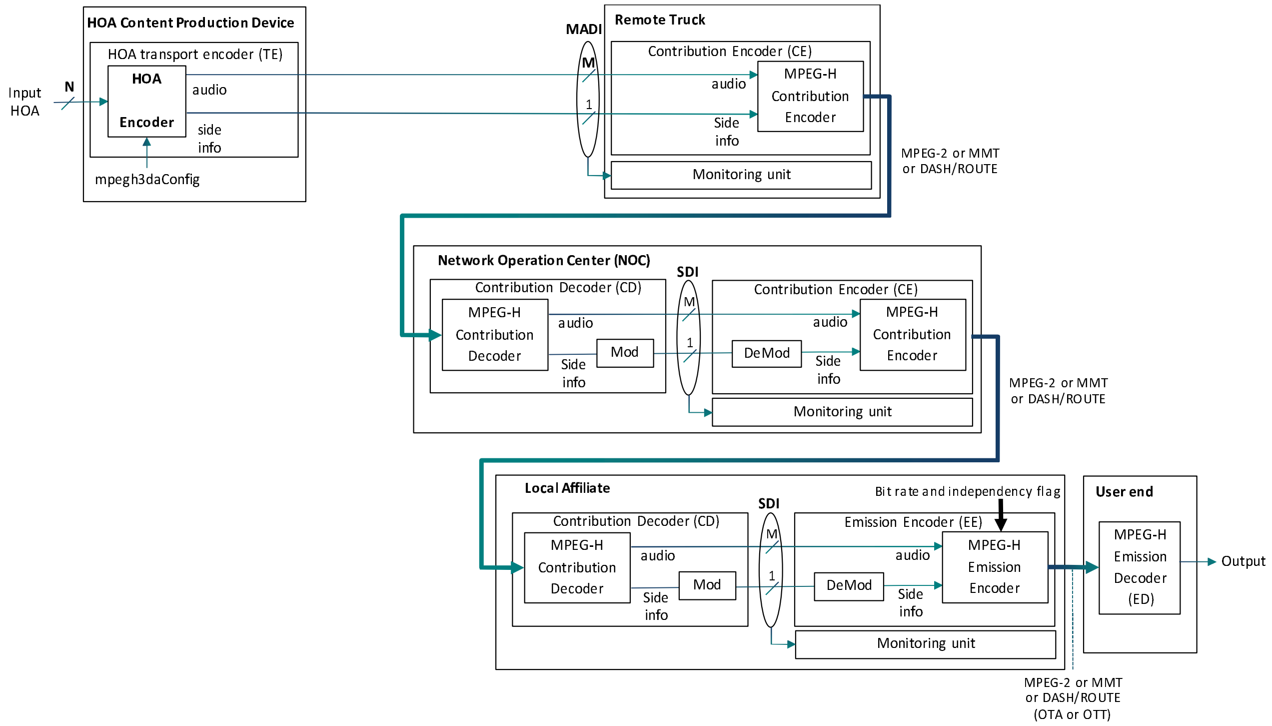


Figure B.1: Block diagram of HOA Content Production

Figure B.1 shows the block diagram of the HOA content production workflow. The HOA encoder can be placed outside the broadcast chain. In this case, the HOA transport format can be generated at an HOA content production device, such as a microphone or a digital audio workstation (DAW), and can be fed into the contribution encoder at the remote truck. The Side-info data modulation is an optional step depending on the input channel to the contribution encoder.

History

Document history		
V1.1.1	June 2018	Publication