SCTE·ISBE STANDARDS

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 243-3 2017

Next Generation Audio Carriage Constraints for Cable Systems: Part 3 –MPEG-H Audio Carriage Constraints

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) Standards and Operational Practices (hereafter called "documents") are intended to serve the public interest by providing specifications, test methods and procedures that promote uniformity of product, interchangeability, best practices and ultimately the long term reliability of broadband communications facilities. These documents shall not in any way preclude any member or non-member of SCTE from manufacturing or selling products not conforming to such documents, nor shall the existence of such standards preclude their voluntary use by those other than SCTE members.

SCTE assumes no obligations or liability whatsoever to any party who may adopt the documents. Such adopting party assumes all risks associated with adoption of these documents, and accepts full responsibility for any damage and/or claims arising from the adoption of such documents.

Attention is called to the possibility that implementation of this document may require the use of subject matter covered by patent rights. By publication of this document, no position is taken with respect to the existence or validity of any patent rights in connection therewith. SCTE shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of this document have been requested to provide information about those patents and any related licensing terms and conditions. Any such declarations made before or after publication of this document are available on the SCTE web site at http://www.scte.org.

Table of Contents

1. Scope

This standard is part of a suite documenting carriage constraints of Next Generation Audio (NGA) codecs in MPEG-2 transport systems and in MPEG DASH. This part of the standard, in conjunction with Part 1 [2], defines carriage of MPEG-H Audio in MPEG-2 transport systems and in MPEG DASH.

The descriptors necessary to signal MPEG-H Audio in MPEG-2 transport systems are defined in ISO/IEC 13818-1 [4] and ETSI EN 300 468 [5], and their usage is described in this document.

Multiplexing and transport for cable using MPEG-2 systems are defined in SCTE 54 [8]. Coding constraints for MPEG-H Audio elementary streams are defined in SCTE 242-3 2017 [1].

2. Normative References

The following documents contain provisions, which, through reference in this text, constitute provisions of this document. At the time of Subcommittee approval, the editions indicated were valid. All documents are subject to revision; and while parties to any agreement based on this document are encouraged to investigate the possibility of applying the most recent editions of the documents listed below, they are reminded that newer editions of those documents might not be compatible with the referenced version.

2.1. SCTE References

[1]   SCTE 242-3 2017, Next Generation Audio Coding Constraints for Cable Systems: Part 3 – MPEG-H Audio Coding Constraints

[2]   SCTE 243-1 2017, Next Generation Audio Carriage Constraints for Cable Systems: Part 1 – Common Transport Signaling

2.2. Standards from Other Organizations

[3]   ATSC CS A/342 Audio, Part 1: Audio Common Elements

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

[5]   ETSI DVB BlueBook A038 (2016-10), Digital Video Broadcasting (DVB); Specification for Service Information (SI) in DVB systems (EN 300 468)

[6]   ISO/IEC 23008-3:2015: Information technology -- High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio, ISO/IEC 23008-3:2015/Amendment 2: Information technology -- High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio – Amendment 2, ISO/IEC 23008-3:2015/Amendment 3: Information technology -- High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio – Amendment 3

[7]   DASH IF: "Guidelines for Implementation: DASH-IF Interoperability Points for ATSC 3.0, Version 0.90," DASH Interoperability Forum, August 3, 2016. (work in progress) http://dashif.org/wp-content/uploads/2016/08/DASH-IF-IOP-for-ATSC3.0-v0.90.pdf

3. Informative References

The following documents might provide valuable information to the reader but are not required when complying with this document.

[8] ANSI/SCTE 54, Digital Video Service Multiplex and Transport System Standard for Cable Television

4. Compliance Notation

shall	This word or the adjective "***required***" means that the item is an absolute requirement of this document.
shall not	This phrase means that the item is an absolute prohibition of this document.
forbidden	This word means the value specified shall never be used.
should	This word or the adjective "*recommended*" means that there may exist valid reasons in particular circumstances to ignore this item, but the full implications should be understood and the case carefully weighted before choosing a different course.
should not	This phrase means that there may exist valid reasons in particular circumstances when the listed behavior is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.
may	This word or the adjective "*optional*" means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.
deprecated	Use is permissible for legacy purposes only. Deprecated features may be removed from future versions of this document. Implementations should avoid use of deprecated features.

5. Abbreviations and Definitions

5.1. Abbreviations

AU	access unit
DRC	Dynamic Range Control
MHAS	MPEG-H Audio Stream
PES	Packetized Elementary Stream
PMT	Program Map Table
PSI	Program Specific Information
RAP	Random Access Point
STD	System Target Decoder
TS	Transport Stream

5.2. Definitions

This document uses the terminology defined in Part 1 of this standard [2], and the mapping of the ATSC 3.0 Audio Glossary Terms to the MPEG-H Audio alternative terms as defined in ATSC A/342-1 Clause 4 [5].

6. MPEG-H Audio Stream (MHAS)

6.1. Elementary stream data encapsulation into MHAS packets

The MPEG-H Audio elementary stream data as defined in SCTE 242-3 2017 [1] shall be encapsulated into MPEG-H Audio Stream (MHAS) packets according to ISO/IEC 23008-3 Clause 14 [6].

MHAS packets of all types defined in ISO/IEC 23008-3 Clause 14 [6] may be present in an MHAS elementary stream, except for the following packet types, which shall not be present in the stream:

- PACTYP_CRC16
- PACTYP_CRC32
- PACTYP_GLOBAL_CRC16
- PACTYP_GLOBAL_CRC32

Other MHAS packets may be present in an MHAS elementary stream (e.g., PACTYP_SYNC, PACTYP_SYNCGAP).

If Audio Scene Information, according to ISO/IEC 23008-3 Clause 15 [6] is present, it shall always be encapsulated in an MHAS PACTYP_AUDIOSCENEINFO packet. Audio Scene Information shall not be included in the **mpegh3daConfig()** structure in the MHAS PACTYP_MPEGH3DACFG packet.

6.2. Configuration Change and Audio/Video Alignment

At each configuration change, the **MHASPacketLabel** shall be changed to a different value from the **MHASPacketLabel** in use before the configuration change occurred. The Access Unit (AU) that contains a configuration change shall be encoded as RAP as defined in SCTE 242-3 2017 Clause 8 [1].

The values of the **MHASPacketLabel** shall be set according to ISO/IEC 23008-3 Clause 14 [6].

The Access Unit that contains the configuration change and the last Access Unit before the configuration change may contain a truncation message embedded in an MHAS PACTYP_AUDIOTRUNCATION packet, as defined in ISO/IEC 23008-3 Clause 14 [6]. The usage of truncation messages enables synchronization between video and audio elementary streams at program boundaries. When used, sample-accurate splicing and reconfiguration of the audio stream are possible.

If MHAS packets of type PACTYP_AUDIOTRUNCATION are present, they shall be used as described in ISO/IEC 23008-3 Clause 14 [6].

6.3. Multi-Stream delivery

In case of multi-stream delivery (as described in SCTE 242-3 2017 Clause 8 [1]) the Audio Program Components of one Audio Program are not carried within one single MHAS elementary stream, but in two or more MHAS elementary streams.

The following constraints apply:

- The Audio Program Components of one Audio Program are carried in one main MHAS elementary stream, and one or more auxiliary MHAS elementary streams.
- The main MHAS stream shall contain at least the Audio Program Components corresponding to the default Audio Presentation, i.e., the Audio Scene Information is present and exactly one preset

shall have the **mae_groupPresetID** field set to "0", as specified in ISO/IEC 23008-3 Clause 15.3 [6].

- The **mae_isMainStream** field in the Audio Scene Information shall be set to "1" in the main MHAS stream, as specified in ISO/IEC 23008-3 Clause 15.3 [6]. This field shall be set to "0" in the auxiliary MHAS streams.
- In each additional MHAS stream (i.e., streams with **mae_isMainStream** field set to 0) the **mae_bsMetaDataElementIDoffset** field in the Audio Scene Information shall be set to the index of the first metadata element in the additional MHAS stream minus one as specified in ISO/IEC 23008-3 Clause 14.6 and Clause 15.3 [6].
- For the main and the auxiliary MHAS stream(s), the **MHASPacketLabel** shall be set according to ISO/IEC 23008-3 subclause 14.6 [6].
- The main and the auxiliary MHAS stream(s) that carry Audio Program Components of one Audio Program shall be time aligned.
- In each auxiliary MHAS elementary stream, RAPs shall be aligned to the RAPs present in the main stream.

7. MPEG-H Audio in MPEG-2 Transport Streams

7.1. Introduction

This section specifies the encapsulation and corresponding signaling of MPEG-H Audio encoded streams in an MPEG-2 Transport Stream environment. It shall be used in conjunction with the common signaling for NGA codescs described in Part 1 of this standard [2].

MPEG-H Audio elementary streams are carried in MPEG-2 Transport Streams (TS) as Packetized Elementary Streams (PES). Each PES packet may contain one or more Access Units (AU).

7.2. MPEG-H Audio Stream (MHAS)

The MPEG-H Audio elementary stream data as defined in SCTE 242-3 2017 [1] shall be encapsulated into MPEG-H Audio Stream (MHAS) packets as specified in subclause 6.1, with further specification in ISO/IEC 13818-1 subclause 2.19.2 [4].

7.2.1. PES Constraints

The PES packet header shall have a Presentation Time Stamp (PTS) associated with the first AU commencing in the PES packet. This PES packet header should have **data_alignment_indicator** set to '1', which requires that the first byte following the PES header is a sync word (per ISO/IEC 13818-1 subclause 2.4.3.7 [4]).

7.2.2. Configuration change and Audio/Video Alignment

MPEG-H Audio elementary streams shall comply with the constraints defined in Clause 6.2.

7.2.3. Multi-stream delivery

MPEG-H Audio elementary streams shall comply with the constraints defined in Clause 6.3.

7.3. Random Access Points

This section describes constraints and signaling for random access points in MPEG-H Audio streams. Elementary stream constraints for random access points are described in SCTE 242-3 2017 Clause 8 [1].

7.3.1. MHAS Random Access Constraints

All rules defined in ISO/IEC 13818-1 subclause 2.19.2 [4] regarding Random Access Points (RAP) shall apply. Particularly, this subclause specifies that a RAP into an MPEG-H Audio Stream consists of the following MHAS packets, in the following order:

- PACTYP_SYNC
- PACTYP_MPEGH3DACFG
- PACTYP_MPEGH3DAFRAME

Additionally, the following rules apply:

- An MHAS PACTYP_BUFFERINFO packet shall be present before the MHAS PACTYP_MPEGH3DAFRAME packet.
- If Audio Scene Information is present, an MHAS PACTYP_AUDIOSCENEINFO packet shall directly follow the MHAS PACTYP_MPEGH3DACFG packet as defined in ISO/IEC 23008-3 Clause 14 [6].

Furthermore, the audio data encapsulated in the MHAS packet PACTYP_MPEGH3DAFRAME shall follow the rules for a RAP as defined in ISO/IEC 23008-3 subclause 5.7 [6].

NOTE: Additional MHAS packets may be present in-between the above listed MHAS packets or after the MHAS PACTYP_MPEGH3DAFRAME packet, with one exception:

- the MHAS PACTYP_AUDIOSCENEINFO packet, when present, shall directly follow the MHAS PACTYP_MPEGH3DACFG packet.

7.3.2. MPEG-2 Transport Stream Random Access Constraints and Signaling

A TS packet containing the PES packet header of an MPEG-H Audio RAP shall have an adaptation field. The **payload_unit_start_indicator** bit shall be set to '1' in the TS packet header and the **adaptation_field_control** bits shall be set to '11' (as per ISO/IEC 13818-1 [4]).

In addition, the **random_access_indicator** bit in the adaptation field of the TS packet that contains the PES packet header of the MPEG-H Audio RAP shall be set to '1' and follow the constraints specified in ISO/IEC 13818-1 subclause 2.4.3.5 [4].

If the PES packet contains a RAP AU, then the RAP AU shall be the first AU in the PES packet and the **data_alignment_indicator** in the PES packet header shall be set to '1'.

7.3.3. Time Intervals between Random Access Points

MPEG-H Audio RAPs shall be inserted in the audio elementary stream at least once in every 2 seconds. The minimum distance between two RAPs shall be 500 ms.

It is recommended that those audio frames whose PTS values are closest to the PTS values of the RAPs of the associated video elementary stream are also coded as RAPs

7.4. PES Packet stream_id and stream_type

The value of the **stream_id** field for MHAS formatted MPEG-H Audio packetized elementary streams shall be '110x xxxx', where each x can be either 0, or 1. The value of **stream_type** for MPEG-H Audio packetized elementary streams shall be 0x2D or 0x2E (indicating ISO/IEC 23008-3 [6] Audio with MHAS transport syntax).

The **stream_type** value 0x2D shall be used for MPEG-H Audio single-stream delivery or for the main stream in case of MPEG-H Audio multi-stream delivery.

The **stream_type** value 0x2E shall be used for auxiliary streams in case of MPEG-H Audio multi-stream delivery (see SCTE 242-3 2017 Clause 10 [1] regarding MPEG-H Audio multi-stream delivery).

7.5. STD Audio Buffer Size

It is recommended that for MPEG-H Audio, the main audio buffer size (*BSn*) has a value as defined in ISO/IEC 13818-1 subclause 2.19.3 [4].

7.6. Signaling of MPEG-H Audio in MPEG-2 Transport Streams

The descriptors specified in this section are used for signaling of MPEG-H Audio codec specific information and NGA features in MPEG-2 transport systems. For signaling NGA features in MPEG-2 transport systems the concept of Audio Preselection is used, as described in Part 1 of this standard, subclause 7.1 [2].

7.6.1. MPEG-H_3dAudio_descriptor

The **MPEG-H_3dAudio_descriptor** provides information about individual MPEG-H Audio elementary streams within a transport stream that are to be identified in the PSI PMT sections. The intended purpose is to provide configuration information for the receiver. It should be noted that multiple MPEG-H Audio elementary streams may be present in a multiplex for a given program.

The **MPEG-H_3dAudio_descriptor** is defined in ISO/IEC 13818-1 subclause 2.6.106 [4] and is located in the PMT of the PSI Tables defined in ETSI EN 300 468 [5].

The **MPEG-H_3dAudio_descriptor** shall be included in a program map section at most once in each relevant ES_info descriptor loop that describes an elementary stream carrying an MPEG-H Audio stream, coded in accordance with ISO/IEC 23008-3 [6] that is included in a transport stream.

7.6.1.1. Profiles and levels

The profile and level value is signaled in the **mpegh3daProfileLevelIndication** field in the **MPEG-H_3dAudio_descriptor** as specified in ISO/IEC 13818-1 subclause 2.6.106 [4]**.**

The values for MPEG-H Audio Low Complexity (LC) Profile Level 1, Level 2 and Level 3 are "0x0B," "0x0C," and "0x0D", respectively, as specified in ISO/IEC 23008-3 subclause 5.3.2 [6].

7.6.2. audio_preselection_descriptor

The **audio_preselection_descriptor** provides information about the available Audio Preselections for one Audio Program contained in one or more MPEG-H Audio associated elementary streams within a transport stream that are to be identified in the PSI PMT sections.

All constraints specified in Part 1 of this standard, subclause 7.1.2 [2] shall apply.

The contents of the **audio_preselection_descriptor** and the Audio Scene Information carried in the MPEG-H Audio elementary should match. The Audio Scene Information is defined in ISO/IEC 23008-3 Clause 15 [6].

The fallowing mapping of the audio preselection descriptor fields to the Audio Scene Information fields should apply:

- The **num_preslections** field in the **audio_preselection_descriptor** should correspond to the **mae_numGroupPresets** field specified in ISO/IEC 23008-3 subclause 15.3 [6].
- The **preselection_id** field in the **audio_preselection_descriptor** should correspond to the **mae_groupPresetID** field specified in ISO/IEC 23008-3 subclause 15.3 [6].
- The first language indicated by the **ISO_639_language_code** field in the **audio_preselection_descriptor** should correspond to the information conveyed in **mae_contentLanguage** field of the default dialog element (i.e., the **maeGroup** which is marked as default in **mae_switchGroupDefaultGroupID** and is tagged in **mae_contentKind** as dialog) specified in ISO/IEC 23008-3 subclause 15.3 [6].
- The **spoken_subtitles** field in the **audio_preselection_descriptor** should be set to '1', if the **mae_contentKind** field specified in ISO/IEC 23008-3 subclause 15.3 [6] is set to '8'.

7.6.3. emergency_information_descriptor

The **emergency_information_descriptor** provides information about the available audio/aural representation of the emergency information contained in one or more MPEG-H Audio associated elementary streams within a transport stream that are to be identified in the PSI PMT sections.

All constraints specified in Part 1 of this standard, subclause 7.2.2 [2] shall apply.

7.6.4. Accesibility Information

For MPEG-H Audio the accessibility fields defined in Part 1 of this standard, subclause 7.3 [2] should correspond to the **mae_groupPresetKind** value in the **mae_GroupPresetDefinition()** structure and the **mae_contentKind** values in the **mae_ContentData()** structures in the **AudioSceneInformation()** of the MPEG-H Audio stream as specified in ISO/IEC 23008-3 [6]. The mapping from the MPEG-H Audio meta-data fields should be done as follows:

- **audio_description** field should be set to '1', if the **mae_contentKind** value of at least one Audio Element is set to '9' ("audio description/visually impaired").

- **dialogue_enhancement** field should be set to '1', if at least the dialog Audio Elements with a **mae_contentKind** value of '2' ("dialogue") have **mae_allowGainInteractivity** set to '1' and **mae_interactivityMaxGain** set to a non-zero value in the corresponding **mae_GroupDefinition()** structure.

- **audio_representation_emergency** field should be set to '1', if the **mae_contentKind** value of at least one Audio Element is set to '12' ("emergency").

8. MPEG-H Audio in MPEG DASH

8.1. Introduction

This section specifies the encapsulation and corresponding signaling of MPEG-H Audio encoded streams in MPEG DASH, additionally to common signaling of NGA codecs specified in Part 1 of this standard, Clause 8 [2].

8.2. MPEG-H Audio Stream (MHAS)

The MPEG-H Audio elementary stream data as defined in SCTE 242-3 2017 [1] shall be encapsulated into MPEG-H Audio Stream (MHAS) packets as specified in subclause 6.1.

The following packet types may be present in an MHAS elementary stream. If they are present, however, they shall be ignored by decoders:
- PACTYP_SYNC
- PACTYP_SYNCGAP

If text labels for Group of Elements, Switch Groups or Presets should be carried within an MPEG-H Audio Stream, they may be encapsulated either as part of the MHAS PACTYP_AUDIOSCENEINFO packet within an **mae_Description()** structure, or alternatively they may be encapsulated within an MHAS PACTYP_DESCRIPTOR packet carrying an **MPEG-H_3dAudio_text_label_descriptor** as defined in ISO/IEC 13818-1 subclause 2.6.112 [4].

If content identifiers should be carried within an MPEG-H Audio Stream, they may be encapsulated in an MHAS PACTYP_MARKER packet with the **marker_byte** set to "E0".

8.3. ISOBMFF Encapsulation

8.3.1. MPEG-H Audio Sample Entry

The sample entry "mhm1" shall be used for encapsulation of MHAS packets into ISOBMFF files, according to ISO/IEC 23008-3 subclause 20.6 [6].

The sample entry "mhm2" shall be used in cases of multi-stream or hybrid delivery, i.e., when the MPEG-H Audio Program is split into two or more streams for delivery as described in ISO/IEC 23008-3 subclause 14.6 [6].

If the **MHAConfigurationBox()** is present, the MPEG-H Profile-Level Indicator `mpegh3daProfileLevelIndication` in the **MHADecoderConfigurationRecord()** shall be set to "0x0B," "0x0C," or "0x0D" for MPEG-H Audio LC Profile Level 1, Level 2, or Level 3, respectively. The Profile-Level Indicator in the MHAS PACTYP_MPEGH3DACFG packet shall be set accordingly.

8.3.2. Random Access Point and Stream Access Point

A File Format sample containing a Random Access Point (RAP), i.e., a RAP into an MPEG-H Audio Stream, is a "sync sample" in the ISOBMFF and shall consist of the following MHAS packets, in the following order:

- PACTYP_MPEGH3DACFG
- PACTYP_AUDIOSCENEINFO (if Audio Scene Information is present)

- PACTYP_BUFFERINFO
- PACTYP_MPEGH3DAFRAME

Note that additional MHAS packets may be present between the MHAS packets listed above or after the MHAS packet PACTYP_MPEGH3DAFRAME, with one exception: when present, the PACTYP_AUDIOSCENEINFO packet shall directly follow the PACTYP_MPEGH3DACFG packet, as defined in ISO/IEC 23008-3 subclause 14.4 [6].

Additionally, the following constraints shall apply for sync samples:

- The audio data encapsulated in the MHAS packet PACTYP_MPEGH3DAFRAME shall follow the rules for random access points as defined in ISO/IEC 23008-3, subclause 5.7 [6].
- All rules defined in ISO/IEC 23008-3 subclause 20.6.1 [6] regarding sync samples shall apply.
- The first sample of an ISOBMFF file shall be a RAP. In cases of fragmented ISOBMFF files, the first samples of each Fragments shall be RAPs.
- In case of non-fragmented ISOBMFF files, a RAP shall be signaled by means of the File Format sync sample box "stss," as defined in ISO/IEC 23008-3 subclause 20.2 [6].
- In case of fragmented ISOBMFF files, the sample flags in the Track Run Box ('trun') are used to describe the sync samples. The "sample_is_non_sync_sample" flag SHALL be set to "0" for a RAP; it shall be set to "1" for all other samples.

8.3.3. Configuration Change

Additional to the MHAS constraints for a Configuration Change described in Clause 6.2 the following constraints apply:
- A configuration change may happen at the beginning of a new ISOBMFF file or Fragment or at any position within the file. In the latter case, the File Format sample that contains a configuration change shall be encoded as a sync sample (RAP) as defined above.
- A sync sample that contains a configuration change and the last sample before such a sync sample may contain a truncation message (PACTYP_AUDIOTRUNCATION) as defined in ISO/IEC 23008-3 Clause 14 [6].

8.3.4. Multi-stream delivery

MPEG-H Audio elementary streams shall comply with the constraints defined in Clause 6.3.

8.4. Signaling of MPEG-H Audio in MPEG DASH

All constraints specified in Part 1 of this standard, Clause 8 [2] shall apply for signaling of MPEG-H Audio in MPEG DASH.

Additional details on Attributes and Elements used with MPEG-H Audio and specified in DASH-IF IOP subclause 5.4.4.3 [7] shall apply.