SCTE · ISBE STANDARDS

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 243-1 2017

Next Generation Audio Carriage Constraints for Cable Systems: Part 1 – Common Transport Signaling

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) / International Society of Broadband Experts (ISBE) 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•ISBE 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•ISBE members.

SCTE•ISBE 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•ISBE 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•ISBE web site at http://www.scte.org.

Table of Contents

List of Figures

List of Tables

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. It is intended to be used in conjunction with the specific audio technologies described in subsequent Parts of this standard.

The common descriptors necessary to signal NGA in MPEG-2 transport systems are defined in ISO/IEC 13818-1 [2] and ETSI EN 300 468 [3], and their usage is described in this Part of the standard. Additional codec specific descriptors are defined in subsequent Parts (which defines codec specific carriage constraints).

Multiplexing and transport for cable using MPEG-2 transport systems are defined in SCTE 54 [5]. Coding constraints for NGA elementary streams are defined in SCTE 242-1 2017 [6] (which provides a general description of NGA systems) and subsequent Parts of the standard (which define codec specific coding constraints).

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. Standards from Other Organizations

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

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

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

[4]   DASH IF: "Guidelines for Implementation: DASH-IF Interoperability Points for ATSC 3.0, Version 1.0," DASH Interoperability Forum, January 31, 2016. http://dashif.org/wp-content/uploads/2017/02/DASH-IF-IOP-for-ATSC3-0-v1.0.pdf

3. Informative References

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

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

[6]   SCTE 242-1 2017, Next Generation Audio Coding Constraints for Cable Systems: Part 1 – Introduction and Common Constraints

[7]   ATSC CS A/331, Signaling, Delivery, Synchronization and Error Protection

ANSI/SCTE 243-1 2017

[8]   ISO/IEC 23009-1:2014/Amd 4:2017 Segment Independent SAP Signalling (SISSI), MPD chaining, MPD reset and other extensions

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

DE	Dialog Enhancement
EINFO	Emergency Information
NGA	Next Generation Audio
PES	Packetized Elementary Stream
PMT	Program Map Table
PSI	Program Specific Information
TS	Transport Stream
VDS	Video Description Service

5.2. Definitions

This document uses the terminology specific to the ATSC 3.0 audio system as defined in ATSC A/342-1 Clause 4 [1]. Additionally, the following definitions are used:

Audio Preselection: set of Audio Program Components representing a version of the Audio Program that may be selected by a user for simultaneous decoding. An Audio Preselection is a sub-selection from all available Audio Program Components of one Audio Program. An Audio Preselection can be considered the NGA equivalent of audio services in predecessor systems, whereby each audio service comprises a complete audio mix.

auxiliary NGA stream: NGA stream delivered using NGA multi-stream delivery, and containing additional Audio Program Components not contained in the main NGA stream.

main NGA stream: NGA stream delivered using NGA multi-stream delivery, and containing at least all the Audio Program Components corresponding to at least one Audio Preselection.

multi-stream delivery: method for carrying Audio Program Components in several NGA streams (e.g., when Audio Program Components offering additional languages are carried in separate elementary streams to facilitate re-multiplexing or service aggregation).

NGA stream: audio elementary stream containing one or more Audio Program Components of one Audio Program.

single-stream delivery: method for carrying all Audio Program Components in a single NGA stream.

6. Introduction

Next Generation Audio (NGA) systems provide Broadcasters, Operators and Content Providers with more flexibility to create and deliver their content. These NGA systems introduce a number of new concepts and techniques, including:

- The capability to provide Immersive Audio with the addition of height elements,
- The capability to provide personalization options to broadcasters and consumers,
- The introduction of Audio Objects to facilitate immersive and personalized audio.

6.1. Audio Preselection Information

Audio Preselections are used for offering alternative audio mixes to the user for "User Personalization". These mixes could include existing service types such as alternative language versions or Video Description Services (VDS), but could also include new services such as "home and away" commentary mix for a sports event or a team channel.

NGA systems enable user control of certain aspects of the Audio Scene (e.g., adjusting the relative level of dialogue with respect to the ambient music and effects) by combining the Audio Program Components, present in one or more NGA streams, at the receiver side. In this way several Audio Program Components can be shared between different Audio Preselections, allowing for very efficient delivery of additional services compared to legacy systems (e.g., the multi-language functionality can be achieved using supplementary streams in legacy systems). This is a major advantage of the NGA systems, where one stream contains much more than one complete audio main.

6.2. Carriage of NGA

Audio Program Components corresponding to one or more Audio Preselections can be delivered in a single elementary stream (i.e., NGA single-stream delivery) or in multiple elementary streams (i.e., NGA multi-stream delivery).

In case of single-stream delivery all Audio Program Components of one Audio Program are carried in a single NGA stream, together with the signaling information of the available Audio Preselections.

In case of multi-stream delivery the Audio Program Components of one Audio Program are not carried within one single NGA stream, but in two or more NGA streams. In this case the main NGA stream contains

at least all the Audio Program Components corresponding to one Audio Preselection. The auxiliary streams may contain additional Audio Program Components (e.g., additional language tracks).

The metadata formats embedded in the NGA streams shall be used only for carrying audio related data.

6.3. Audio Specific Capabilities

Typical differentiation of receiver capabilities for the audio decoding and rendering pipeline may use one or multiple of the following properties:

- Codec capabilities:
- Codec, Profile and Level
- Rendering capabilities/environment
- User preferences and settings (accessibility, language, role)
- User interaction and Personalization

7. Signaling of NGA codecs in MPEG-2 TS

7.1. Audio Preselection Information

For signaling the features of NGA systems in MPEG-2 Transport Streams the concept of Audio Preselection is used as defined in ETSI EN 300 468 subclause 6.4 [3]

7.1.1. audio_preselection_descriptor

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

The descriptor may be used by the receiver for selection of the appropriate Audio Preselection to present the Audio Program to the user, and to locate the constituent elementary streams. The **audio_preselection_descriptor** is defined in ETSI EN 300 468 subclause 6.4 [3] and is located in the PMT of the PSI Tables defined in ETSI EN 300 468 [3].

The **audio_preselection_descriptor,** when present, shall be used as specified in ETSI EN 300 468 subclause 6.4 [3] and shall be included in the program map section at most once in the ES_info descriptor loop that describes the elementary stream carrying the main NGA stream.

The **audio_preselection_descriptor** shall be present for all NGA elementary streams that contain Audio Preselections for the Audio Program.

In case of single-stream delivery, the use of the **audio_preselection_descriptor** is optional.

In case of multi-stream delivery, the following rules apply:

- the **audio_preselection_descriptor** shall be included in the program map section at most once in the ES_info descriptor loop that describes the elementary stream carrying the main NGA stream.
- the **stream_identifier_descriptor** (defined in ETSI EN 300 468 subclause 6.2 [3]) shall be included in the program map section in each relevant ES_info descriptor loop that describes an elementary stream carrying an additional NGA stream.

When the **audio_preselection_descriptor** is included in the program map section, the **ISO_639_Language_descriptor** (ISO/IEC 13818-1 subclause 2.6.18 [2]) shall not be included in the program map section in any relevant ES_info descriptor loop that describes an elementary stream carrying the main or an auxiliary NGA stream.

7.1.2. Usage of the audio_preselection_descriptor

The audio_preselection_descriptor is used for signaling the NGA features that the receiver may use for selection of the appropriate Audio Preselection and NGA streams to present the Audio Program to the user. For each available Audio Preselection, the descriptor provides information about the available accessibility features and languages, indication about the preferred reproduction layout and the interactivity options.

Figure 1 illustrates an example of one Audio Program containing several available Audio Preselections, assuming all Audio Program Components and the Audio Preselection information are contained in the same stream (single-stream delivery case).

Audio Preselection Descriptor	NGA Elementary Stream					
	Audio Program Components					
	M&E	Dialogue EN	Dialogue ES	VDS EN	VDS ES	Team Radio
Default Preselection (English)	M&E	Dialogue EN	Dialogue ES	VDS EN	VDS ES	Team Radio
English Video Description Preselection	M&E	Dialogue EN	Dialogue ES	VDS EN	VDS ES	Team Radio
Spanish Preselection	M&E	Dialogue EN	Dialogue ES	VDS EN	VDS ES	Team Radio
Spanish Video Description Preselection	M&E	Dialogue EN	Dialogue ES	VDS ES	VDS ES	Team Radio
English Team Radio Preselection	M&E	Dialogue EN	Dialogue ES	VDS EN	VDS ES	Team Radio
M&E Only Preselection	M&E	Dialogue EN	Dialogue ES	VDS EN	VDS ES	Team Radio

Figure 1 - Example Broadcast Audio Preselections in an Audio Program

7.2. Emergency Information

7.2.1. Introduction

The ATSC 3.0 Emergency Alert System structure, defined in ATSC A/331 Annex G [7], specifies several types of emergency information that can be delivered to TV receivers including audio/aural representation of the emergency information. This section provides appropriate signaling of the audio/aural representation of the emergency information for MPEG-2 transport systems.

7.2.2. emergency_information_descriptor

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

The descriptor shall be present if at least one of the Audio Preselections, available in the **audio_preselection_descriptor**, contains an audio/aural representation of the emergency information.

The **emergency_information_descriptor** shall be included in the program map section at most once in the ES_info descriptor loop that describes the elementary stream carrying the main NGA stream that is included in a transport stream.

Table 1 - Bit Stream Syntax for emergency information descriptor

Syntax	No. of Bits	Format
emergency_information_descriptor() {		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
num_preselections	5	uimsbf
audio_representation_emergency	1	uimsbf
reserved	2	'11'
for (i=0; i<num_preselections; i++) {		
preselection_id	5	uimsbf
reserved	3	'1 1111'
}		
emergency_information_start_time_present	1	bslbf
emergency_information_end_time_present	1	bslbf
reserved	6	'11 1111'
If (emergency_information_start_time_present) {		
emergency_information_start_time	32	uimsbf
reserved	6	'11 1111'
emergency_information_start_time_ms	10	uimsbf
}		
If (emergency_information_end_time_present) {		
emergency_information_end_time	32	uimsbf
reserved	6	'11 1111'
emergency_information_end_time_ms	10	uimsbf
}		
}		

descriptor_tag: The descriptor tag shall be set to 0xED to identify the descriptor as an **emergency_information_descriptor**.

descriptor_length: This 8-bit field shall specify the number of bytes of the descriptor immediately following the **descriptor_length** field.

num_preselections: This 5-bit unsigned integer field shall specify the number of Audio Preselections for which the emergency information described in this descriptor is valid. The minimum number of **num_preselections** shall be "1".

audio_representation_emergency: This 1-bit Boolean field shall indicate, when set to '1', that the Audio Preselection contains aural/audio representation of emergency information.

preselection_id: This 5-bit unsigned integer field identifies each Audio Preselection for which the emergency information described in this descriptor is valid.

emergency_information_start_time_present: This 1-bit Boolean flag shall indicate, when set to '1', that the fields **emergency_information_start_time** and **emergency_information_start_time_ms** are present. When set to '0', the fields **emergency_information_start_time** and **emergency_information_start_time_ms** shall not be present and the start time of the aural/audio representation of the emergency information shall be inferred to be equal to the presentation time of first media sample in presentation order in the audio asset.

emergency_information_end_time_present: This 1-bit Boolean flag shall indicate, when set to '1', that the fields **emergency_information_end_time** and **emergency_information_end_time_ms** are present. When set to '0', the fields **emergency_information_end_time** and **emergency_information_end_time_ms** shall not be present and the end time of the aural/audio representation of the emergency information shall be inferred to be equal to the presentation time of the last media sample in presentation order in the audio asset.

emergency_information_start_time: This 32-bit unsigned integer shall indicate the start time of the aural/audio representation of the emergency information, as the least-significant 32 bits of the count of the number of seconds since January 1, 1970 00:00:00, International Atomic Time (TAI).

emergency_information_start_time_ms: This 10-bit unsigned integer in the range 0 to 999 shall indicate the milliseconds offset from the time indicated in emergency_information_start_time, such that the formula:

$$emergency\_information\_start\_time + (emergency\_information\_start\_time\_ms/1000)$$

yields the start time of the audio/aural emergency information to the nearest 1 millisecond.

emergency_information_end_time: This 32-bit unsigned integer shall indicate the end time of the aural/audio representation of the emergency information, as the least-significant 32 bits of the count of the number of seconds since January 1, 1970 00:00:00, International Atomic Time (TAI).

emergency_information_end_time_ms: This 10-bit unsigned integer in the range 0 to 999 shall indicate the milliseconds offset from the time indicated in emergency_information_end_time, such that the formula:

$$emergency\_information\_end\_time + (emergency\_information\_end\_time\_ms/1000)$$

yields the end time of the audio/aural emergency information to the nearest 1 millisecond.

7.3. Accessibility Information

The accessibility support is signaled for each Audio Preselection. The accessibility fields are defined in the **audio_preselection_descriptor** and in the **emergency_information_descriptor**. Table 2 specifies the mapping of the accessibility fields to the ATSC 3.0 audio accessibility services.

Table 2 - Mapping of accesibility fields to Audio Accesibility Services

Accessibility fields	Location	Audio Accessibility Service
audio_description	**audio_preselection_descriptor**	Video Description Service (VDS) for Visually Impaired

dialogue_enhancement	audio_preselection_descriptor	Dialog Enhancement (DE) enabled
audio_representation_emergency	emergency_information_descriptor	Audio/Aural representation of Emergency Information (EINFO)

Video Description Service for Visually Impaired – indicates, when **audio_description** field is set to "1", the presence of the audio accessibility service for visually impaired for the Audio Preselection.

Dialog Enhancement enabled – indicates, when **dialogue_enhancement** field is set to "1", that the Audio Preselection enables the ability for a receiver to change the relative level of dialog to enhance dialog intelligibility.

Audio/Aural representation of Emergency Information – indicates, when **audio_representation_emergency** field is set to "1", that the Audio Preselection contains aural/audio representation of emergency information.

7.4. Example of descriptors usage in PMT

Table 3 specifies an example of how the descriptors declared or defined within the present document, can be used in the PSI PMT depending on the NGA features available in a single NGA stream. This does not imply that their use in other ways is restricted.

Table 3 - Possible usage in PMT SI tables of descriptors for single-stream delivery

Descriptor	Stereo	EINFO	Multi-lang.	DE EI	VDS
codec specific descriptors	M	M	M	M	M
audio_preselection_descriptor	O	O	M	M	M
emergency_information_descriptor	O	M	O	M	O
stream_identifier_descriptor	O	O	O	O	O
M = mandatory in PMT; O = optional in PMT; X = Not Allowed					

Table 4 and Table 5 specify a similar example of how the descriptors declared or defined within the present document, can be used in the PSI PMT depending on the NGA features available in the case of NGA multi-stream delivery. This does not imply that their use in other ways is restricted.

Table 4 - Possible usage in PMT SI tables of descriptors for main stream

Descriptor	Stereo	EINFO	Multi-lang.	DE EI	VDS
codec specific descriptors	M	M	M	M	M
audio_preselection_descriptor	O	M	M	M	M
emergency_information_descriptor	O	M	O	M	O
stream_identifier_descriptor	O	O	O	O	O
M = mandatory in PMT; O = optional in PMT; X = Not Allowed					

Table 5 - Possible usage in PMT SI tables of descriptors for auxiliary streams

Descriptor	Stereo	EINFO	Multi-lang.	DE EI	VDS
codec specific descriptors	N/A	O	O	O	O
audio_preselection_descriptor	N/A	X	X	X	X
emergency_information_descriptor	N/A	X	X	X	X
stream_identifier_descriptor	N/A	M	M	M	M
M = mandatory in PMT; O = optional in PMT; X = Not Allowed					

8. Signaling of NGA codecs in MPEG DASH

For signaling audio, the Preselection feature defined in ISO/IEC23009-1:2014/Amd.4:2017 [8] is used. It is specifically adapted to address the Next Generation Audio concepts (see A/342-1 [1]).

8.1. Background and Basic Use Cases

The use cases specified in DASH-IF IOP subclause 5.4.1 [4] are expected to be supported by the client reference model.

8.2. Assumptions and Definitions

The assumptions and definitions specified in DASH-IF IOP subclause 5.4.2 [4] shall apply.

8.3. Codec-Independent Mapping to DASH

The attributes available in Adaptation Sets and Media Content Components specified in DASH-IF IOP subclause 5.4.3 [4] shall be used.

8.4. Codec-specific Properties

The codec-specific properties on how codecs can be mapped on the generic data structure defined in Clause 8.3 are specified in subsequent Parts of this standard. This typically includes for each codec:

- Codecs parameter settings
- Usage of the Preselection elements
- Random Access Point and Switching Point requirements
- The definition of bitstream switching or media level switching
- File format encapsulation requirements