

# SCTE • ISBE<sup>®</sup>

## S T A N D A R D S

---

Digital Video Subcommittee

---

AMERICAN NATIONAL STANDARD

ANSI/SCTE 242-1 2017

**Next Generation Audio Coding Constraints for Cable  
Systems: Part 1 – Introduction and Common  
Constraints**

## 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>.

All Rights Reserved

© Society of Cable Telecommunications Engineers, Inc. 2017  
140 Philips Road  
Exton, PA 19341

# Table of Contents

<b>Title</b>	<b>Page Number</b>
NOTICE	2
Table of Contents	3
1. Scope	4
2. Normative References	4
2.1. Standards from Other Organizations	4
3. Informative References	4
4. Compliance Notation	4
5. Abbreviations and Definitions	5
5.1. Abbreviations	5
5.2. Definitions	5
6. System overview	5
6.1. Features of NGA Systems	5
6.2. NGA Concepts	5
6.2.1. Audio Program Components and Preselections	5
6.2.2. Audio Element Formats	5
6.2.3. Audio Rendering	6
7. Encoding Constraints	7
7.1. Sampling Rate	7
7.2. Audio Program Structure	7
7.3. General Elementary Stream Structure	7

## 1. Scope

This document is part of a suite documenting coding constraints of Next Generation Audio (NGA) systems for cable television. It is intended to be used in conjunction with the specific audio technologies described in subsequent Parts of this standard.

## 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

## 3. Informative References

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

## 4. Compliance Notation

<i>shall</i>	This word or the adjective “ <i>required</i> ” means that the item is an absolute requirement of this document.
<i>shall not</i>	This phrase means that the item is an absolute prohibition of this document.
<i>forbidden</i>	This word means the value specified shall never be used.
<i>should</i>	This word or the adjective “ <i>recommended</i> ” 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.
<i>should not</i>	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.
<i>may</i>	This word or the adjective “ <i>optional</i> ” 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.
<i>deprecated</i>	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

HOA	Higher Order Ambisonics
NGA	Next Generation Audio
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.

## 6. System overview

### 6.1. Features of NGA Systems

A description of the main features of Next Generation Audio systems is provided in ATSC A/342-1 subclause 5.1 [1].

### 6.2. NGA Concepts

Several concepts are common to all NGA systems (e.g., systems supported by ATSC 3.0). This section describes these common concepts.

#### 6.2.1. Audio Program Components and Preselections

Audio Program Components are separate pieces of audio data that are combined to compose an Audio Preselection. A simple Audio Preselection may consist of a single Audio Program Component, such as a Complete Main Mix for a television program. Audio Preselections that are more complex may consist of several Audio Program Components, such as ambient music and effects, combined with dialog and video description.

Audio Preselections are combinations of Audio Program Components representing versions of the audio program that may be selected by a user. For example, a complete audio with English dialog, a complete audio with Spanish dialog, a complete audio (English or Spanish) with video description, or a complete audio with alternate dialog may all be selectable Preselections for a Program.

The Components of a Preselection can be delivered in a single audio Elementary Stream or in multiple audio Elementary Streams.

#### 6.2.2. Audio Element Formats

The NGA systems support three fundamental Audio Element Formats:

- 1) Channel Sets are sets of Audio Elements consisting of one or more Audio Signals presenting sound to speaker(s) located at canonical positions. These include configurations such as mono, stereo, or 5.1, and extend to include non-planar configurations, such as 7.1+4.
- 2) Audio Objects are Audio Elements consisting of audio information and associated metadata representing a sound's location in space (as described by the metadata). The metadata may be dynamic, representing the movement of the sound.
- 3) Scene-based audio (e.g., HOA) consists of one or more Audio Elements that make up a generalized representation of a sound field.

### **6.2.3. Audio Rendering**

Audio Rendering is the process of composing an Audio Preselection and converting all the Audio Program Components to a data structure appropriate for the audio outputs of a specific receiver. Rendering may include conversion of a Channel Set to a different channel configuration, conversion of Audio Objects to Channel Sets, conversion of scene-based sets to Channel Sets, and/or applying specialized audio processing such as room correction or spatial virtualization.

#### **6.2.3.1. Video Description Service (VDS)**

Video Description Service is an audio service carrying narration describing a television program's key visual elements. These descriptions are inserted into natural pauses in the program's dialog. Video description makes TV programming more accessible to individuals who are blind or visually impaired. The Video Description Service may be provided by sending a collection of "Music and Effects" components, a Dialog component, and an appropriately labeled Video Description component, which are mixed at the receiver. Alternatively, a Video Description Service may be provided as a single component that is a Complete Mix, with the appropriate label identification.

#### **6.2.3.2. Multi-Language**

Traditionally, multi-language support is achieved by sending Complete Mixes with different dialog languages. For NGA systems, multi-language support can be achieved through a collection of "Music and Effects" streams combined with multiple dialog language streams that are mixed at the receiver.

#### **6.2.3.3. Personalized Audio**

Personalized audio consists of one or more Audio Elements with metadata, which describes how to decode, render, and output "full" Mixes. Each personalized Audio Preselection may consist of an ambience "bed", one or more dialog elements, and optionally one or more effects elements. Multiple Audio Preselections can be defined to support a number of options such as alternate language, dialog or ambience, enabling height elements, etc.

There are two main concepts of personalized audio:

- 1) Personalization selection – The bit stream may contain more than one Audio Preselection where each Audio Preselection contains pre-defined audio experiences (e.g., "home team" audio experience, multiple languages, etc.). A listener can choose the audio experience by selecting one of the Audio Preselections.
- 2) Personalization control – Listeners can modify properties of the complete audio experience or parts of it (e.g., increasing the volume level of an Audio Element, changing the position of an Audio Element, etc.).

## **7. Encoding Constraints**

The following constraints are defined on all NGA systems for cable television.

### **7.1. Sampling Rate**

The sampling frequency of Audio Signals shall be 48 kHz.

### **7.2. Audio Program Structure**

An Audio Program shall consist of one or more Audio Preselections. One Audio Preselection shall be signaled as the default (main).

The default Audio Preselection shall have all of its Audio Program Components present in the broadcast stream.

Note: This constraint is intended to facilitate future applications in which additional Program Components are delivered by other means (e.g., hybrid use case in ATSC 3.0).

The main Audio Preselection is intended to be the default in cases where no other selection guidance (user-originated or otherwise) exists.

Audio Preselections shall consist of at least one Audio Program Component of any Audio Element Format.

Audio Program Components may be delivered in more than one Elementary Stream. Audio Preselections other than the default Preselection may include Audio Program Components from multiple Elementary Streams. Audio Preselections shall not utilize Audio Program Components from more than three Elementary Streams.

Further constraints specific to individual codecs are defined in subsequent Parts of this standard.

### **7.3. General Elementary Stream Structure**

The carriage of the streams described in this specification is defined in SCTE 243-1 [2], with further specifications defined in subsequent Parts of this standard.