Society of Cable Telecommunications Engineers

ENGINEERING COMMITTEE

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 193-1 2014

MPEG-4 AAC Family Audio System – Part 1 Coding Constraints for Cable Television

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) Standards and Recommended 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, whether used domestically or internationally.

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 Standards.

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

MPEG-4 HE AAC – PART 1
CODING CONSTRAINTS FOR CABLE TELEVISION

1. SCOPE

This document defines the coding constraints on MPEG-4 AAC, HE AAC, and HE AAC v2 (referred to collectively in this document as the "AAC family") profile audio for cable television. It also discusses MPEG-2 AAC LC profile audio, which is closely related to MPEG-4 AAC profile audio. The carriage of the streams described in this specification is defined in SCTE 193-2 2014 [5]

2. NORMATIVE REFERENCES

The following documents contain provisions, which, through reference in this text, constitute provisions of this standard. At the time of Subcommittee approval, the editions indicated were valid. All standards are subject to revision; and while parties to any agreement based on this standard 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 may not be compatible with the referenced version.

[1] ISO/IEC 14496-3:2009: Information technology – Coding of audio-visual objects – Part 3: Audio

[2] ISO/IEC 14496-3:2009/Amendment 4:2013: Information technology – Coding of audio-visual objects – Part 3: Audio, Amendment 4: New levels for AAC profiles

[3] ISO/IEC 14496-26:2010: Information technology – Coding of audio-visual objects – Part 26: Audio Conformance

[4] ITU-R Recommendation BS.1770-3 (08/2012): Algorithms to measure audio programme loudness and true-peak audio level

3. INFORMATIVE REFERENCES

The following documents contain information that may be helpful in applying this standard.

[5] SCTE 193-2 2014, MPEG-4 AAC Family Audio System – Part 2: Constraints for Carriage over MPEG-2 Transport

[6] ISO/IEC 13818-7:2006, Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)

[7] ATSC A/85, 2013: "ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television"

[8] ETSI TS 101 154 V1.11.1 (2012-11): Digital Video Broadcasting (DVB) Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream

[9] ARIB STD-B32 Part 2, Video Coding, Audio Coding and Multiplexing Specifications for Digital Broadcasting, Part 2, Audio Signal and Coding System

[10] MPEG N2006, “Report on the MPEG-2 AAC Stereo Verification Tests”, http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w2006.zip

[11] MPEG N6009, “Report of Verification Tests of MPEG-4 High Efficiency AAC”, http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w6009.zip

[12] MPEG N7137, “Listening test report on MPEG-4 High Efficiency AAC v2”, http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/w3075.zip

4. COMPLIANCE NOTATION

“SHALL”	This word or the adjective “REQUIRED” means that the item is an absolute requirement of this specification.
“SHALL NOT”	This phrase means that the item is an absolute prohibition of this specification.
“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.

5. TERMS AND ACRONYMS

The following defined terms and acronyms are used in this document:

5.1 Terms

Audio Access Unit — An individually accessible portion of audio data within an elementary stream

Audio Object Type — Coding tools or modules, as specified in ISO/IEC 14496-3 [1] subclause 1.5.1.1, Table 1.1

Reserved — The term “reserved”, when used in the clauses defining the coded bit stream and associated signaling, indicates that the value may be used in the future for extensions to the standard. Unless otherwise specified, all

	reserved bits shall be set to '1'. Reserved fields and reserved values are intended to be ignored by receiving equipment.
Sequence	A series of audio access units that starts with an RAP AU up to but not including any subsequent RAP AU

5.2 Acronyms and Abbreviations

AAC	Advanced Audio Coding
AAC LC	AAC Low Complexity
AU	Access unit
DRC	Dynamic Range Control
HE AAC	High Efficiency Advanced Audio Coding
HE AAC v2	High Efficiency Advanced Audio Coding version 2
ID	Object Type identifier
LKFS	Loudness, K weighted, relative to nominal full scale
PS	Parametric Stereo
PNS	Perceptual Noise Substitution
RAP	Random Access Point
SBR	Spectral Band Replication

6. CODING CONSTRAINTS

6.1 Introduction

This section specifies constraints for use of MPEG-4 AAC, HE AAC, and HE AAC v2 audio coding as defined in ISO/IEC 14496-3 [1] with ISO/IEC 14496-3/Amendment 4 [2]. An overview of the profiles for the MPEG-4 AAC family can be found in ISO/IEC 14496-3 [1] subclause 1.5.2. MPEG-4 profiles consist of one or more Audio Object Types, as indicated in ISO/IEC 14496-3 [1] subclause 1.5.2.1, Table 1.3, where a name and a unique Object Type ID (ID) designate each Audio Object Type. Audio Object Types consist of coding tools or modules, as specified in ISO/IEC 14496-3 [1] subclause 1.5.1.1, Table 1.1.

It is desirable to support bitstreams that conform to MPEG-2 Advanced Audio Coding Low Complexity Profile (MPEG-2 AAC LC Profile). Such bitstreams can always be classified as MPEG-4 AAC Profile or MPEG-4 HE AAC Profile bitstreams so that the constraints specified in this document apply. The difference between the two bitstreams is that the MPEG-4 Perceptual Noise Shaping (PNS) tool is not defined in an MPEG-2 AAC LC Profile bitstream [6], but can exist in an MPEG-4 AAC Profile bitstream (see ISO/IEC 14496-3 [1] subclause 1.5.1.2.3). An MPEG-4 AAC Profile decoder is expected to decode both MPEG-2 AAC LC Profile and MPEG-4 AAC Profile bitstreams; likewise an MPEG-4 HE AAC Profile decoder is expected to decode both MPEG-4 HE AAC Profile bitstreams and MPEG-2 AAC LC Profile bitstreams with SBR.

MPEG-2 AAC bitstreams will be encapsulated in the ADTS ("Audio Data Transport Stream") structure as opposed to the LATM/LOAS ("Low Overhead Audio Transport Multiplex/ Low Overhead Audio Stream") structure used by MPEG-4 AAC. Decoders are expected to be able to process either encapsulation. This facilitates program exchange with material from countries using MPEG-2 AAC (such as Japan) as well as from countries using MPEG-4 AAC (such as Europe).

6.1.1 MPEG-4 AAC Profile

The MPEG-4 AAC Profile is defined in ISO/IEC 14496-3 [1] subclause 1.5.2.1 and consists of one Audio Object Type, AAC LC (ID 2). The levels of this profile are defined in ISO/IEC 14496-3 [1] subclause 1.5.2.3 Table 1.10. The AAC LC technology is defined in ISO/IEC 14496-3 [1] subclause 4. This subclause defines a range of technologies, but only specifications for Audio Object Type AAC LC are relevant.

6.1.2 MPEG-4 High Efficiency AAC Profile

The MPEG-4 High Efficiency AAC Profile (also referred to as the HE AAC Profile) is defined in ISO/IEC 14496-3 [1] subclause 1.5.2.1 and consists of two Audio Object Types, AAC LC (ID 2) and Spectral Band Replication (ID 5). The levels of this profile are defined in ISO/IEC 14496-3 [1] subclause 1.5.2.3 Table 1.11. An overview of signaling and carriage of the SBR data is in ISO/IEC 14496-3 [1] subclause 1.6.5. The SBR bitstream syntax is defined in ISO/IEC 14496-3 [1] subclause 4.4.2.8 Table 4.62 and subsequent tables. The SBR decoding semantics are defined in ISO/IEC 14496-3 [1] subclause 4.6.18.

6.1.3 MPEG-4 High Efficiency AAC v2 Profile

The MPEG-4 High Efficiency AAC v2 Profile (also referred to as the HE AAC v2 Profile) is defined in ISO/IEC 14496-3 [1] section 1.5.2.1 and consists of three Audio Object Types, AAC LC (ID 2), Spectral Band Replication (ID 5) and Parametric Stereo (ID 29). The levels of this profile are defined in ISO/IEC 14496-3 [1] subclause 1.5.2.3 Table 1.12. An overview of signaling and carriage of the PS data is in ISO/IEC 14496-3 [1] subclause 1.6.6. The PS bitstream syntax is defined in ISO/IEC 14496-3 [1] subclause 8.4.2 Table 8.9 and subsequent tables. The PS decoding semantics are defined in ISO/IEC 14496-3 [1] subclause 8.6.4. Additional information on the combination of the SBR and PS tools is in ISO/IEC 14496-3 [1] Annex 8.A.

6.1.4 Profiles and Levels

Table 1 and the associated notes are based on ISO/IEC 14496-3/Amendment 4 [2] Tables 1.10, 1.11, and 1.12 and is provided here for information. It shows how the levels of each of the three profiles for the AAC family relate. In particular, note that level 1 in the HE AAC and HE AAC v2 Profiles does not exist, so that level 2 can have a similar functionality (i.e., stereo at 48 kHz output sampling rate) across all three profiles. Similarly, level 3 in the AAC Profile does not exist, so that level 4 can have a similar functionality (i.e., 5 channels at 48 kHz output sampling rate) across all three profiles.

Since input sampling rates for this standard are constrained to 48 kHz, audio bitstreams coded in accordance with this standard shall use only Levels 2, 4, or 6. The Down Sampled mode of the SBR tool shall not be used.

Table 1: Levels for AAC, HE AAC and HE AAC v2 Profiles (Informative)

	AAC Profile		**HE AAC Profile**			**HE AAC v2 Profile** (Note 2, 6)		
Level	**Max. channels**	**Max. sampling rate [kHz]**	**Max. channels**	**Max. AAC sampling rate, SBR present [kHz]**	**Max. SBR sampling rate [kHz] (in/out)**	**Max. channels**	**Max. AAC sampling rate, SBR present [kHz]**	**Max. SBR sampling rate [kHz] (in/out)**
1	2	24	N/A	N/A	N/A	N/A	N/A	N/A
2	2	48	2	24	24/48	2	24	24/48
3	N/A	N/A	2	48	48/48 (Note 3)	2	24/48 (Note 4)	48/48 (Note 3)
4	5	48	5	24/48 (Note 5)	48/48 (Note 3)	5 (Note 6)	24/48 (Note 5)	48/48 (Note 3)
5	5	96	5	48	48/96	5 (Note 6)	48	48/96
6	7	48	7	24/48 (Note 5)	48/48 (Note 3)	7 (Note 6)	24/48 (Note 5)	48/48 (Note 3)
7	7	96	7	48	48/96	7 (Note 6)	48	48/96

Note 1: N/A indicates not applicable – the level does not exist for that profile.

Note 2: Level 2, 3, 4, 6 and 7 HE AAC v2 Profile decoders implement the baseline version of the parametric stereo tool. A level 5 decoder is not limited to the baseline version of the parametric stereo tool.

Note 3: Level 3, level 4 and level 6 decoders operate the SBR tool in downsampled mode if the sampling rate of the AAC core is higher than 24 kHz. Hence, if in the decoder the SBR tool operates on a 48 kHz AAC signal, the internal sampling rate of the SBR tool will be 96 kHz, however, the output signal will be downsampled by the SBR tool to 48 kHz.

Note 4: If Parametric Stereo data is present the maximum AAC sampling rate is 24 kHz. If Parametric Stereo data is not present the maximum AAC sampling rate is 48 kHz.

Note 5: For one or two channels the maximum AAC sampling rate, with SBR present, is 48 kHz. For more than two channels the maximum AAC sampling rate, with SBR present, is 24 kHz.

Note 6: Only bitstreams consisting of exactly one AAC single channel element are permitted to contain Parametric Stereo data. Bitstreams containing more than one channel in the AAC part are not permitted to contain Parametric Stereo data.

6.1.5 Sampling Rate

The system conveys digital audio sampled at a frequency of 48 kHz that shall be locked to the 27 MHz MPEG-2 system clock. The 48 kHz audio sampling clock is defined as:

$$48 \text{ kHz audio sample rate} = (2 \div 1125) \times (27 \text{ MHz MPEG-2 system clock})$$

If analog signal inputs are employed, the A/D converters shall sample at 48 kHz locked to the 27 MHz clock. If digital inputs are employed, the input sampling rate shall be 48 kHz locked to the system clock, or the audio encoder shall contain sampling rate converters which convert the sampling rate to 48 kHz locked to the system clock.

6.1.6 AAC Family Audio Bitrates

Independent listening tests[1] have shown that the AAC Profile delivers “broadcast quality” at 128 kbps (stereo). The HE AAC Profile is a superset of the AAC Profile and differs only in the use of one additional tool called SBR (Spectral Band Replication). Independent listening tests[2] have shown that HE AAC Profile codecs deliver “good to excellent audio quality” at 48 kbps (stereo). Similarly, the HE AAC v2 Profile is a superset of the HE AAC Profile and differs only in the use of one additional tool called PS (Parametric Stereo). Independent listening tests[3] have shown that HE AAC v2 Profile codecs deliver “good audio quality” at 24 kbps (stereo). AAC Family encoders typically switch off the SBR tool for bitrates at or above 96 kbps (stereo) and switch off the PS tool for bitrates above 48 kbps (stereo).

Since the PS tool is only defined for stereo, HE AAC v2 is not used for multichannel coding. Encoders will switch off the PS tool if the user selects multi-channel operation after having chosen “HE AAC v2.”

6.1.6.1 Operating Points

For the purposes of interoperability, the AAC Profile stereo operating points are defined as 64, 96, 112, 128, 182, 192, 224, 256 and 320 kbps. The HE AAC Profile stereo operating points are defined as 32, 48, 56, 64, 96, 112, and 128 kbps. The SBR tool shall not be used for bitrates greater than 128 kbps (stereo). The HE AAC v2 Profile stereo operating points are defined as 16, 24, 32, 40, 48, 56, and 64 kbps. Other bitrates are possible and legal.

For multi-channel operation, bitrate values scale directly with the number of channels (excluding the LFE) and the stereo operating point bitrate values are simply multiplied by 2.5 (for 5.1) or 3.5 (for 7.1).

---

[1] MPEG N2006 “Report on the MPEG-2 AAC Stereo Verification Tests.” [10]

[2] MPEG N6009 “Report of Verification Tests of MPEG-4 High Efficiency AAC.” [11]

[3] MPEG N7137 “Listening test report on MPEG-4 High Efficiency AAC v2.” [12]

6.2 Encoding Constraints

Bitstreams produced by the encoder shall comply with the MPEG-4 AAC Profile, the MPEG-4 High Efficiency AAC Profile, or the MPEG-4 High Efficiency AAC v2 Profile as defined in ISO/IEC 14496-3 [1], with level restrictions as follows.

Bitstreams including support for monaural and stereo MPEG-4 AAC or MPEG-4 HE AAC shall comply with the level 2 restrictions of the AAC or High Efficiency AAC Profiles, respectively.

Bitstreams including support for monaural, stereo, and parametric stereo MPEG-4 HE AAC v2 shall comply with the level 2 restrictions of the High Efficiency AAC v2 Profile.

Bitstreams including support for multichannel (up to 5.1 channels) MPEG-4 AAC, MPEG-4 HE AAC, or MPEG-4 HE AAC v2 shall comply with the level 4 restrictions of the AAC or High Efficiency AAC or High Efficiency AAC v2 Profiles, respectively. Coupling Channel Elements (CCEs) according to ISO/IEC 14496-3 [1] shall not be used.

Bitstreams including support for multichannel (up to 7.1 channels) MPEG-4 AAC, MPEG-4 HE AAC, or MPEG-4 HE AAC v2 shall comply with the level 6 restrictions of the AAC or High Efficiency AAC or High Efficiency AAC v2 Profiles, respectively. Coupling Channel Elements (CCEs) according to ISO/IEC 14496-3 [1] shall not be used.

It is strongly recommended that all bitstreams carry embedded metadata according to Section 7 herein. Bitstreams shall comply with the random access rules in Section 8 herein.

Bitstreams encapsulated in ADTS, intended for consumption by MPEG-2 AAC decoders, shall not use the PNS or PS tools.

6.3 Expectations for Decoders

It is expected that decoders will support High Efficiency AAC v2 Profile level 6. This includes support for lower levels of that profile, and support for level 6 and lower levels of the MPEG-4 AAC and MPEG-4 High Efficiency AAC profiles.

It should be noted that Level 4 (5.1 channel) decoders may not support decoding of bitstreams encoded in accordance with level 6 (7.1 channel). Operators should therefore consider the capabilities of decoders that may be available to their customers in deciding whether to transmit bitstreams coded in accordance with Level 6 or Level 4.

It is expected that all decoders will support dynamic range control (DRC) metadata according to ISO/IEC 14496-3 [1], subclause 4.5.2.7, matrix-mixdown according to ISO/IEC 14496-3 [1], subclause 4.5.1.2, and the use of ancillary metadata according to subclause 4.5.2.14 of ISO/IEC 14496-3/Amendment 4 [2]. It is also expected that all decoders will support both ADTS and LATM/LOAS transport formats.

6.3.1 Use of Metadata by Decoders

The **prog_ref_level** metadata parameter defined in subclause 4.5.2.7 of ISO/IEC 14496-3 [1] is set in the encoder and is used by decoders to adjust each item of decoded content to a uniform average loudness, thus avoiding annoying variations in average loudness from one item of content to another, e.g., between programs and commercials.

The DRC metadata defined in subclause 4.5.2.7 of ISO/IEC 14496-3 [1] is used by decoders to modify the dynamic range of decoded content, by reducing the level of very loud portions of the content and by raising the level of very quiet portions of the content, so that they are better adapted to the listening environment. If, however, the consumer has a home theatre or other listening environment that can reproduce the full dynamic range of the content, the consumer can choose to turn off the dynamic range control, allowing that consumer to enjoy the content exactly as it was originally mixed. The DRC parameters are set in the encoder, with values selected to be appropriate for the type of content.

The ancillary metadata defined in subclause 4.5.2.14 of ISO/IEC 14496-3/Amendment 4 [2] includes additional parameters for dynamic range control to enable decoders to provide additional compression when needed for use with audio systems or listening environments where heavy compression provides a better listening experience.

See further information on expectations for use of metadata by decoders in Section 7.4.

7. AAC METADATA

AAC audio carries metadata in three locations:

- MPEG-4 AAC dynamic range control tool as described in Section 7.1
- MPEG-4 AAC program config element as described in Section 7.2
- MPEG-4 AAC ancillary data as described in Section 7.3

7.1 Dynamic Range Control Tool

The dynamic range control (DRC) tool in AAC defines several parameters to enable control of loudness and dynamic range in the decoder. It is strongly recommended that these parameters be present in the bitstream for mono, stereo, and multichannel audio.

The DRC tool, including program reference level, is defined in ISO/IEC 14496-3 [1], subclause 4.5.2.7.

7.2 Program Config Element

The program config element (PCE) in MPEG-4 provides the capability to carry weighting factors for the matrix-mixdown method to enable decoders to create a 1- or 2-channel downmix from transmitted multi-channel content. This information may be present for multichannel audio.

The MPEG-4 AAC **program_config_element()** and **matrix_mixdown_idx** are defined in ISO/IEC 14496-3 [1], subclause 4.5.1.2.

See also downmixing parameters in Section 7.4.3.

7.3 Ancillary Data

The ancillary data tool is defined in ISO/IEC 14496-3/Amendment 4 [2], subclause 4.5.2.14.

The ancillary data tool extends the DRC tool and the matrix-mixdown method provided by the program config element (PCE) with specific information about the audio content, allowing the broadcaster to control rendering of the content. This data should be present for multichannel audio and may be present for stereo or monophonic audio.

It is strongly recommended that multichannel audio bitstreams carry MPEG-4 ancillary data, which includes downmixing parameters with finer granularity than the matrix-mixdown method, and use of that data by decoders is preferable to using the matrix-mixdown method. Further, the MPEG-4 ancillary data conveys downmix parameter for mixing 6 or 7 channel content to 5 channels, which is not supported by the PCE. See Section 7.4.3.1 for constraints on corresponding values if downmixing parameters are present in both the PCE and MPEG-4 ancillary data.

7.4 Use of DRC Tool and Ancillary Data

7.4.1 Loudness Normalization

Loudness normalization in MPEG-4 AAC audio coding is accomplished by utilizing the **prog_ref_level** parameter in the MPEG DRC tool, to carry the program reference level. The value of the **prog_ref_level** parameter shall indicate the loudness[4] of the corresponding encoded audio content using LKFS units. LKFS and its loudness measurement algorithm are specified in ITU-R Recommendation BS.1770 [4]. Decoders are expected to read this value and apply an appropriate level shift to audio essence to match the loudness of output audio to a target value set in the decoder (determined by the decoder parameter **target_ level**).

Decoders with a **target_level** set to higher levels than the transmitted **prog_ref_level** will increase the output audio level such that overloads might occur. Therefore appropriate DRC parameters shall be present in the bit stream as discussed in Sections 7.4.2 and 7.4.4.

7.4.2 Dynamic Range Compression

Dynamic range compression results in a reduction of the dynamic range of transmitted audio. While wide dynamic range audio may provide an exciting listen experience under excellent listening conditions, in some circumstances such audio is not appropriate. The DRC tool provides gain factors to reduce the dynamic range to suit specific listening environments. To

[4] Methods to measure loudness are explained in the ATSC Recommended Practice A/85 "Techniques for Establishing and Maintaining Audio Loudness for Digital Television" [7].

accommodate a wide range of scenarios, two different compression schemes are available: Light Compression and Heavy Compression.

Light Compression is intended for moderate reduction of the dynamic range in accordance with the artistic intent of the content provider. This can be considered as the default operation mode of decoders.

Heavy Compression is used only when a strong reduction of the dynamics, often in combination with an increased loudness level, is required.

7.4.2.1 Light Compression

Light Compression is accomplished by utilizing the MPEG-4 DRC tool defined in ISO/IEC 14496-3 [1], subclause 4.5.2.7. Light Compression values are generated by the encoder and transmitted in the bit stream to facilitate signal manipulation in the decoder. Such signal manipulation may be channel-selective and frequency-selective as set in the encoder. Using Light Compression, both positive and negative gain values may be scaled in the decoding process to take listener preferences and listening conditions into account. Limitations to this are defined in Section 7.4.4.

7.4.2.2 Heavy Compression

Heavy Compression is not possible using the MPEG DRC tool, therefore, if employed, Heavy Compression is accomplished by the use of **compression_value** in **MPEG4_ancillary_data()**. Heavy Compression values are generated by the encoder and transmitted in the bit stream to facilitate signal manipulation in the decoder. Such single parameter-driven signal manipulation is neither channel-selective nor frequency-selective. Unlike Light Compression, Heavy Compression is not scalable in the decoding process and can only be selected as either on or off. If Heavy Compression is selected, the decoder is expected to replace any Light Compression gains by gains derived from **compression_value**.

If Heavy Compression data is not present in the bit-stream, the decoder is expected to revert to the use of Light Compression gain values.

7.4.3 Consistency of Downmix Information

If downmix information for 5.1 to stereo is conveyed in both the **MPEG4_ancillary_data()** structure and the program config element (PCE), the value of the **matrix_mixdown_idx** (in **program_config_element()**) shall correspond to the **surround_mix_level_value** (in **MPEG4_ancillary_data()**) according to Table 6. For the **center_mix_level_value** no restrictions apply.

Table 6: Corresponding Values in MPEG4_ancillary_data() and program_config_element()

surround_mix_level_value	PCE matrix_mixdown_idx
"000" (-0.0 dB)	"00" (-3.0 dB)
"001" (-1.5 dB)	"00" (-3.0 dB)
"010" (-3.0 dB)	"00" (-3.0 dB)
"011" (-4.5 dB)	"01" (-6.0 dB)
"100" (-6.0 dB)	"01" (-6.0 dB)
"101" (-7.5 dB)	"10" (-9.0 dB)
"110" (-9.0 dB)	"10" (-9.0 dB)
"111" (- ∞ dB)	"11" (- ∞ dB)

7.4.4 Downmix of content in 7.1 channel configurations

Downmixing of 7.1 channel content is defined in a cascaded way: First, downmix 7.1 channels to 5.1 using the **dmix_a_idx** and **dmix_b_idx** fields defined in the **MPEG4_ancillary_data()** structure. If a downmix for 2 channel stereophonic playout is desired, downmix the generated 5.1 channel audio to 2 channels using the **center_mix_level_value** and the **surround_mix_level_value**.

7.4.5 DRC Presentation Mode

Dynamic range control may either be used to limit the dynamic range of an audio signal to improve intelligibility under noisy or difficult listening environments or to prevent highly undesirable clipping. The latter may occur when audio is played back at a higher target level than its **prog_ref_level** or when a reduction of the number of output channels is performed (i.e., downmixing).

To avoid this clipping, the downmixing methods and associated loudness levels can be monitored in production or appropriate dynamic range control values should be transmitted along with the audio as metadata. For this purpose, in addition to the ISO/IEC 14496-3 [1] **dynamic_range_info(),** the **compression_value** of the **MPEG4_ancillary_data()** should also be present.

In **MPEG4_ancillary_data()**, **drc_presentation_mode** shall be present to indicate the decoder **target_level** that was assumed by the encoder for use with Light and Heavy Compression. The Presentation Mode to be used is an operational choice, however, operators should be aware of the DRC capabilities of decoders that may be available to their customers that do not recognize presentation mode signaling. Table 7 shows the **drc_presentation_mode** values:

Table 7: DRC Presentation Mode

drc_presentation_mode	Description
"00"	DRC presentation mode not indicated
"01"	DRC presentation mode 1
"10"	DRC presentation mode 2
"11"	Reserved

7.4.5.1 DRC Presentation Mode 1

According to Section 4.5.2.14.2.4.1 of ISO/IEC 14496-3/Amendment 4 [2], encoders signaling DRC presentation mode 1 must ensure that light compression gains prevent clipping if a decoder is set to a **target_level** of 124 (-31 LKFS) or below and negative DRC coefficients were applied at full scale. For decoders set to a **target_level** of 80 (-20 LKFS) or below (including 96, or -24LKFS), heavy compression gain factors are sufficient to avoid clipping.

In both cases, transmitted gain factors shall be suitable to prevent clipping even if decoders perform a downmix to a reduced number of output channels. Regarding light compression, monophonic 1-channel downmixes need not to be considered.

7.4.5.2 DRC Presentation Mode 2

According to Section 4.5.2.14.2.4.1 of ISO/IEC 14496-3/Amendment 4 [2], encoders signaling DRC presentation mode 2 must ensure that light compression gains prevent clipping if a decoder is set to a **target_level** of 96 (-24 LKFS) or below and negative DRC coefficients were applied at full scale. This is true for playback of all channels and also for downmixing for 2-channel playback.

If DRC presentation mode 2 is signaled, decoders apply heavy compression only if a monophonic downmix (e.g., to feed RF modulated outputs) is desired.

Note 1: DRC presentation mode should be considered as a broadcaster preference choice and to avoid disturbances in the audio output, and should not be changed within an elementary stream.

Note 2: To assure consistent audio performance when the viewer changes to a different channel, all content should be encoded with the same value of **drc_presentation_mode**.

8. RANDOM ACCESS

MPEG-4 audio streams might not be decodable with full fidelity (i.e., full decoding) at each audio access unit (AU). The reason for this is that some static information might not be transmitted in a given AU in order to save bit-rate, and normally previously transmitted values are used by the decoder, thus full fidelity will not be possible until those values are received following the random access start of playback. The following section describes necessary constraints that permit full decoding of MPEG-4 audio at a given RAP AU.

8.1 Random Access with the AAC Profile; Encoding Constraints on the Audio Object Type AAC LC

If **channelConfiguration** (see Table 1.15 of ISO/IEC 14496-3 [1]) has the value 0, the p**rogram_config_element()** (PCE) containing the actual channel configuration shall be present in each RAP AU.

Note: Further constraints on the **AudioSpecificConfig()** (ASC) apply. As this structure is embedded into the transport layer, these constraints are described in [SCTE 193-2 2014].

8.2 Random Access with the HE AAC Profile; Further Encoding Constraints on the Audio Object Type SBR

The HE AAC Profile is based on the AAC Profile; therefore all constraints from the Audio Object Type AAC LC (see Section 8.1) are also satisfied. While an HE AAC Profile bitstream will typically contain SBR information, the hierarchical nature of the profiles means that it is not required to contain SBR information; e.g., an AAC Profile bitstream is also an HE AAC Profile bitstream.

In contrast to the AAC configuration that is completely described by the **AudioSpecificConfig()**, the Audio Object Type SBR decoder needs additional configuration parameters. These parameters are transmitted inside the **sbr_header()**, which is not required to be contained in every audio access unit.

An HE AAC Profile bit stream containing SBR information shall have an **sbr_header()** present in the first access unit of a Sequence (i.e., in the RAP AU) . Moreover, it is required to have the **bs_header_flag** set to 1 (signaling that the **sbr_header()** is present).

Note 1: Reliance on values transmitted in preceding frames (i.e., time differential coding of parameters) for frames containing an **sbr_header()** is prohibited by MPEG-4 Audio Conformance (see ISO/IEC 14496-26 [3], clause 7.17.1.2.1.4). This restriction assures that an RAP AU can be fully decoded and processed.

Note 2: According to ISO/IEC 14496-3 [1], subclause 4.5.2.8.2.1, "as long as no SBR header part is present, the SBR decoder performs upsampling and delay adjustment only." Therefore, audio playback may start even if **sbr_header()** was not yet received, but only with reduced quality. Therefore, ISO/IEC 14496-3 recommends: "In continuous broadcast applications, SBR extension data elements with an SBR header part are typically sent twice per second. In addition, a SBR header part can any time be inserted, if an instantaneous, possibly program dependent, change of header parameters is required." (See ISO/IEC 14496-3 [1], clause 4.5.2.8.2.1). Similarly, it is recommended that a RAP AU be transmitted twice per second.

8.3 Random Access with the HE AAC v2 Profile; Further Encoding Constraints on the Audio Object Type PS

The HE AAC v2 Profile is based on the HE AAC Profile; therefore all constraints from the HE AAC Profile (see Section 8.2) are also satisfied. While an HE AAC v2 Profile bitstream will typically contain PS information, the hierarchical nature of the profiles means that it is not required to contain PS information; e.g., a HE AAC Profile bitstream is also a HE AAC v2

Profile bitstream. In a manner similar to the Audio Object Type SBR, the configuration, or header, parameters for the Audio Object Type PS payload are transmitted inside the ps_data() information, and these configuration parameters are not required to be contained in every access unit.

An HE AAC v2 Profile bit stream containing PS information shall have **enable_ps_header** set to 1 (i.e., the PS decoder configuration parameters are present) in the first AU of a Sequence (i.e., in the RAP AU).

Note 1: Reliance on values transmitted in preceding frames (i.e., time differential coding of parameters) for frames with **enable_ps_header** = 1 is prohibited by MPEG-4 Audio Conformance (see ISO/IEC 14496-26 [3], clause 7.18.1.3). This restriction assures that an RAP AU can be fully decoded and processed.

Note 2: As Audio Object Type PS conformance requires **enable_ps_header** = 1 whenever **bs_header_flag** = 1, (i.e. PS Header information is present whenever an SBR header is present, see ISO/IEC 14496-26 [3], clause 7.18.1.3), the random access requirements for Audio Object Type PS additionally imply the random access requirements for Audio Object Type SBR.

Note 3: According to ISO/IEC 14496-3 [1], subclause 8.6.5.1, "a conformant decoder that receives PS data shall output the mono signal in the two output channels until a first **ps_data()** element with **enable_ps_header**=1 is received and in which for all enabled parameters frequency differential coding is employed and num_env>0, ensuring that the PS data can be decoded correctly." Therefore audio playback may start even if **ps_data()** cannot be decoded.

8.4 Random Access Constraints on Metadata

For a RAP, both **dynamic_range_info()** and **MPEG4_ancillary_data()** shall be present in the access unit. Further, **prog_ref_level_present** shall be set to 1, i.e., **prog_ref_level** is present in **dynamic_range_info**(). In **MPEG4_ancillary_data()**, **downmixing_levels_MPEG4_status** shall be set to 0x1.