Society of Cable Telecommunications Engineers

ENGINEERING COMMITTEE
Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 193-2 2014

MPEG-4 AAC Family Audio System – Part 2 Constraints for Carriage over MPEG-2 Transport

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 2
CONSTRAINTS FOR CARRIAGE OVER MPEG-2 TRANSPORT

1. SCOPE

This document describes the carriage of MPEG-4 AAC, MPEG-4 HE AAC and MPEG-4 HE AAC v2 (referred to collectively in this document as the "AAC family") profile audio in MPEG-2 transport systems. It also discusses MPEG-2 AAC LC profile audio, which is closely related to MPEG-4 AAC profile audio.

The descriptor necessary to signal AAC family audio and information for signaling mixing of main and associated services in the receiver are defined in this document. Multiplexing and transport for cable using MPEG-2 systems are defined in SCTE 54 [8]. Coding constraints for AAC family audio elementary streams are defined in SCTE 193-1 2014 [1].

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] SCTE 193-1 2014, MPEG-4 AAC Family Audio System – Part 1, Coding Constraints for Cable Television

[2] ISO/IEC 13818-1:2013, Information technology – Generic coding of moving pictures and associated audio information: Systems

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

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

[5] ISO 639-2:1998, Codes for the representation of names of languages – Part 2: Alpha-3 code

[6] ISO/IEC 8859-1:1998, Information technology – 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1

[7] ISO/IEC 10646:2012, Information technology – Universal coded Character Set (UCS)

3. INFORMATIVE REFERENCES

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

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

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

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

[11] ETSI EN 300 468, "Digital Video Broadcasting (DVB); Specification for Service Information (SI) in DVB systems"

[12] ETSI TS 101 154, "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast Applications based on the MPEG-2 Transport Stream"

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

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.

5.2 Acronyms and Abbreviations

AAC	Advanced Audio Coding
ADTS	Audio Data Transport Stream
AU	Access unit
bslbf	Bit string, left bit first, where "left" is the order in which bit strings are written in this standard
HE AAC	High-Efficiency Advanced Audio Coding
HE AAC v2	High Efficiency Advanced Audio Coding version 2
LATM	Low overhead Audio Transport Multiplex
LOAS	Low Overhead Audio Stream
PES	Packetized Elementary Stream
PMT	Program Map Table
PSI	Program Specific Information
PNS	Perceptual Noise Substitution
PTS	Presentation Time Stamp
RAP	Random Access Point
STD	System Target Decoder
TS	Transport Stream
uimsbf	Unsigned integer, most significant bit first

6. AAC FAMILY IN MPEG-2 TRANSPORT STREAMS

6.1 Introduction

This section specifies the encapsulation and corresponding signaling of AAC family encoded audio in an MPEG-2 transport stream environment.

AAC family 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).

The AAC family elementary stream data as defined in SCTE 193-1 2014 [1] shall be encapsulated either in the LATM/LOAS AudioSyncStream() transport syntax or in the ADTS adts_frame() transport syntax per ISO/IEC 14496-3 [3]. The preferred encapsulation is LATM/LOAS.

6.2 AAC Elementary Stream Formatted as LATM/LOAS

When the AAC family elementary stream data is formatted according to the LATM multiplex format and subsequently encapsulated in the LOAS transmission format per ISO/IEC 14496-3 [3] the following constraints shall apply:

The **AudioSyncStream()** version shall be used.

Note: **AudioSyncStream()** adds a sync word to the audio stream to allow for synchronization.

The **AudioMuxElement()** multiplex element format shall be used.

The semantics of the **AudioMuxElement()** and **AudioSyncStream()** formatting are described in subclause 1.7 of ISO/IEC 14496-3 [3].

The following constraints to the LATM multiplex shall apply:

- **audioMuxVersion** shall be '0';
- **allStreamsSameTimeFraming** shall be '1', as all payloads belong to the same AU;
- **numSubFrames** shall be '0', as there is only one **PayloadMux()** (AU) per LATM **AudioMuxElement()**;
- **numProgram** shall be '0', as there is only one audio program per LATM multiplex;
- **numLayer** shall be '0', as no scalable profile is used;
- the field **latmBufferFullness** shall be set to 0xFF, indicating that the stream is variable rate and that buffer fullness measures are not used;
- the value for **frameLengthFlag** contained in the **GASpecificConfig** shall be set to '0', indicating that the transform length of the IMDCT for AAC is 1024 samples for long and 128 for short blocks.

6.2.1 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 [2], subclause 2.4.3.7).

6.3 AAC Elementary Stream Formatted as ADTS

When an AAC family elementary stream is formatted according to the ADTS multiplex format per ISO/IEC 14496-3 [3] the following constraints shall apply:

- the semantics of ISO/IEC 14496-3 [3], section 1.A.4.3 shall be followed,
- the **ID** field should be set to '1'

- the **adts_error_check()** element format should be used,
- the **adts_header_error_check()** element format should be used.

Furthermore, the use of PNS and PS tools are precluded. The SBR tool may be used. See ISO/IEC 13818-7 [9] and ARIB B.32 Part 2 [10] for additional details.

6.3.1 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 [2], subclause 2.4.3.7).

6.4 Random Access Points

This section describes constraints and signaling for random access points in AAC family streams. Elementary stream constraints for random access points are described in SCTE 193-1 2014 [1].

6.4.1 LATM/LOAS Random Access Constraints

The LATM/LOAS transport format carries StreamMuxConfig (SMC). This structure carries essential, but quasi-static, information such as sampling rate and channel configuration. This structure need not be transmitted with each AU, however, periodic transmission allows for random access to the stream.

Both **StreamMuxConfig()** and **AudioSpecificConfig()** shall be present in a Random Access Point (RAP) AU. In **AudioMuxElement()**, **useSameStreamMux** shall be set to '0', which requires that **StreamMuxConfig()** is present in **AudioMuxElement()**. In **StreamMuxConfig**(), since **numProgram** and **numLayer** are both set to '0', **useSameConfig** is set to '0'. Additionally, since **audioMuxVersion** is set to '0', **AudioSpecificConfig()** is present in **StreamMuxConfig()**.

6.4.2 ADTS Random Access Constraints

The ADTS transport format carries adts_fixed_header() and adts_variable_header(). These headers carry essential, but quasi-static, information such as sampling rate and channel configuration. These headers need not be transmitted with each AU, however, periodic transmission allows for random access to the stream.

The adts_fixed_header() and adts_variable_header() shall be present in a Random Access Point (RAP) AU.

6.4.3 MPEG-2 Transport Stream Random Access Constraints and Signaling

A TS packet containing the PES packet header of an AAC family 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 [2]). In addition, the **random_access_indicator** bit in the Adaptation field of the TS packet that contains the PES packet header of the AAC family RAP shall be set to '1' and follow the constraints specified in ISO/IEC 13818-1 [2] in subclause 2.4.3.5.

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'.

6.4.4 Time Intervals between Random Access Points

The encoder shall place AAC family RAPs in the audio elementary stream at least every 2 seconds. It is recommended that AAC family RAPs occur in the audio elementary stream at least every 500 ms. Further it is recommended those audio frames in the audio elementary stream whose PTS values are closest to the PTS values of the RAPs of the associated video elementary stream should also be coded as RAPs.

6.5 PES Packet stream_id and stream_type

In accordance with subclause 2.4.3.7 of ISO/IEC 13818-1 [2] the value of the **stream_id** field for LATM/LOAS and ADTS formatted AAC family packetized elementary streams shall be '110n nnnn', where each n can be either 0 or 1 and the stream number is given by the value taken by the n's.

In accordance with ISO/IEC 13818-1 [2] subclause 2.4.4.9, the value of **stream_type** for LATM/LOAS formatted AAC family packetized elementary streams shall be 0x11. In accordance with ISO/IEC 13818-1 [2] subclause 2.4.4.9, the value of **stream_type** for ADTS formatted AAC family packetized elementary streams, constrained according to section 6.3, shall be 0x0F.

6.6 STD Audio Buffer Size

It is recommended that for AAC family audio in a SCTE system, the main audio buffer size (BS_n) should have a value of 8 976 bytes in accordance with ISO/IEC 13818-1:2013 [2], subclause 2.11.2.2.

6.7 MPEG_AAC_descriptor

The **MPEG_AAC_descriptor** provides information about an AAC family elementary stream that has been coded in accordance with ISO/IEC 14496-3 [3] and ISO/IEC 14496-3/Amd 4 [4], as constrained by this document. The intended purpose is to provide configuration information for the receiver. It should be noted that an AAC family elementary stream contains the encoded representation of one audio service. Multiple AAC family elementary streams may be present in a multiplex for a given program.

One descriptor for each AAC family elementary stream shall be placed in the descriptor loop that immediately follows the **ES_info_length** field in the TS_program_map_section() describing that elementary stream.

Note: The syntax and semantics of the **MPEG_AAC_descriptor** is derived from the descriptor defined in Annex H of ETSI EN 300 468 [11].

6.7.1 **MPEG_AAC_descriptor** Syntax

Table 1 defines the syntax of the **MPEG_AAC_descriptor**.

Table 1: MPEG_AAC_descriptor Syntax

Syntax	Number of Bits	Identifier
MPEG_AAC_descriptor(){		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
AAC_profile	4	uimsbf
AAC_level	4	uimsbf
if (descriptor_length > 1){		
channel_service_flag	1	bslbf
mainid_flag	1	bslbf
ascv_flag	1	bslbf
language_flag	1	bslbf
component_name_flag	1	bslbf
AAC_extension_data_flag	1	bslbf
mixinfoexists	1	bslbf
reserved_zero	1	'0'
if (channel_service_flag == 1){		
channel_config	5	uimsbf
AAC_service_type	4	uimsbf
receiver_mix_rqd	1	uimsbf
reserved_zero	6	'000000'
}		
if (mainid_flag == 1){		
Reserved	5	'11111'
mainid	3	uimsbf
}		
if (asvc_flag == 1){		
asvc	8	uimsbf
}		
if (language_flag == 1){		
language	3*8	uimsbf
}		
if (component_name_flag == 1){		
component_name_length	8	uimsbf
for (i=0;i< component_name_length;i++) {		
component_name_text_char[i]	8	bslbf
}		
}		
if (AAC_extension_data_flag == 1){		
AAC_extension_data_length	8	uimsbf
for(i=0;i< AAC_extension_data_length;i++){		
AAC_extension_data_config[i]	8	uimsbf
}		
}		
}		
}		

6.7.2 Semantics for the **MPEG_AAC_descriptor**

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

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

AAC_profile This 4-bit unsigned integer field shall specify the profile used in the AAC family encoded stream, as defined in section 1.5.2.1 of ISO/IEC 14496-3 [3], and coded according to Table 2.

Table 2: AAC Profile

AAC_profile	Meaning	
0x0	MPEG-4 AAC Profile in LATM/LOAS	MPEG-2 AAC LC Profile in ADTS
0x1	MPEG-4 HE AAC Profile in LATM/LOAS	MPEG-2 AAC LC Profile with SBR in ADTS
0x2	MPEG-4 HE AAC v2 Profile in LATM/LOAS	N/A
0x3-0xF	Reserved	Reserved

AAC_level This 4-bit unsigned integer field in the range 1 to 7 shall specify the AAC coding level as defined in Table 1.10, Table 1.11, or Table 1.12 of ISO/IEC 14496-3 Amendment 4 [4], and used in the AAC family encoded stream.

channel_service_flag: This 1-bit field shall be set to '1' when the **channel_config**, **AAC_service_type** and **receiver_mix** fields are present in the descriptor.

mainid_flag: This 1-bit field shall be set to '1' when the **mainid** field is present in the descriptor.

ascv_flag: This 1-bit field shall be set to '1' when the **ascv** field is present in the descriptor.

language_flag: This 1-bit field shall be set to '1' when the **language** field is present in the descriptor.

component_name_flag: This 1-bit field shall be set to '1' when the component name is present in the descriptor.

AAC_extension_data_flag: This 1-bit field shall be set to '1' when **AAC_extension_data_config()** is present in the descriptor

mixinfoexists: This 1-bit field shall be set to '1' when information on the mixing parameters for receiver mix is available as AS_control_data, carried in PES private data per Table 6 in section 7.3.

channel_config: This 5-bit unsigned integer field shall specify the channel configuration according to configurations in Table 1.19 of ISO/IEC 14496-3 Amendment 4 [4]. Channel configuration values shall be constrained to 1, 2, 3, 4, 5, 6, 7, 11, 12, and 14. Table 3 below shows the allowed configurations extracted from Table 1.19, for information.

Table 3: Channel Configurations

Value	Number of channels	channel to speaker mapping
1	1	centre front speaker
2	2	left, right front speakers
3	3	centre front speaker, left, right front speakers
4	4	centre front speaker, left, right front speakers, rear centre speaker
5	5	centre front speaker, left, right front speakers, left surround, right surround speakers
6	5.1	centre front speaker, left, right front speakers, left surround, right surround speakers, low frequency enhancement speaker
7	7.1 Front	center front speaker left, right front center speakers, left, right front speakers, left surround, right surround speakers, low frequency enhancement speaker
11	6.1	center front speaker, left, right front speakers, left surround, right surround speakers, rear center speaker, low frequency enhancement speaker
12	7.1 Back	center front speaker left, right front speakers, left surround, right surround speakers, rear surround left, right speakers, low frequency enhancement speaker
14	7.1 Top	center front speaker left, right front speakers, left surround, right surround speakers, left, right front vertical height speakers, low frequency enhancement speaker

AAC_service_type: This 4-bit field indicates the type of audio service being conveyed in the AAC audio stream. The **AAC_service_type** field shall be coded as shown in Table 4. See section 6.6 for definitions of each service type.

Table 4: AAC_service_type Coding

AAC_service_type	Description
0	Complete Main (CM)
1	Music and Effects (ME)
2	Visually Impaired (VI)
3	Hearing Impaired (HI)
4	Dialogue (D)
5	Commentary (C)
6	Reserved
7	Voiceover (VO)
8-15	Reserved

receiver_mix_rqd: This bit shall be set to '1' when the audio service is a partial service that must be combined with another audio service before presentation. It shall be set to '0' when the audio service is a full service.

mainid: This optional 3-bit field contains a number in the range 0 to 7, which identifies a main audio service. The **mainid** field shall be included for a main service if an associated service is present that must be mixed in the receiver with the main service. If there is more than one main service to be mixed, each shall be tagged with a unique number. This value is used as an identifier to link associated services with particular main services. See Table 4 and section 7.4 for more information.

asvc: This 8-bit field is optional, but shall be present for an elementary stream carrying an associated service that is associated with one or more main audio services carried in the same program. Each bit (0 to 7) identifies with which main service(s) this associated service is associated. For example, to associate an associated audio service with a main audio service that has a **mainid** value of '0', the value of the **asvc** field is set to '00000001' (0x01). To associate an associated audio service with a main audio service that has a **mainid** value of '3', the value of the **asvc** field is set to '00001000' (0x08).

language: When present, this 3-byte field shall indicate the language of this audio service. The **language** field shall be coded as a threecharacter code in accordance with the bibliographic code "/B" of ISO 639-2/ [5]. Each character is coded into 8 bits according to ISO 8859-1 [6] and inserted in order into the 24-bit field. The coding is identical to that used in the MPEG-2 **ISO_639_language_code** value in the **ISO_639_language_descriptor**() specified in ISO/IEC 13818-1 [2].

component_name_length: When present, this 7-bit field shall indicate the length in bytes of the **component_name_text_character[]** array.

component_name_text_char[]: When present, this character array shall contain a brief textual description of the audio service. Text shall be encoded as UTF-8 per ISO/IEC 10646 [7]

AAC_extention_data_length: When present, this 8-bit field indicates the length in bytes of the AAC_extension_data_config[] structure.

AAC_extension_data_config[]: When present, this data structure provides additional information on extension data embedded in the MPEG AAC elementary stream.
The syntax of this structure is not yet defined and reserved for future use. Decoders are expected to parse and ignore this data.

6.8 Further Information and Constraints for AAC Audio Services

An AAC family elementary stream contains the encoded representation of one audio service, which may be a *main service* or various types of *associated service*[1].

It should be noted that a main service may consist of a complete program mix (i.e., it is a complete main service), but a main service may also consist of a partial program mix (e.g., music and effects), which is normally accompanied by an associated service (e.g., dialogue). It should also be noted that an associated service may be an enhancement service intended to be mixed with a main service in the receiver, but an associated service may itself also contain a complete program mix.

When multiple audio elementary streams of the same service type (CM, ME, VI, HI, D, etc.) are present in the same program, the **language** code(s) shall be included in the **MPEG_AAC_descriptor** associated with each such audio component.

When multiple audio elementary streams of the same service type and having the same language are present in the same program, the **component_name_flag** shall be set to '1' so that component name information associated with each such audio component is included in the **MPEG_AAC_descriptor**.

[1] It should be noted that the term "Supplementary Audio Service" is used elsewhere (e.g., in DVB) to describe these associated services, with basically the same functionality.

Table 5 indicates the values that some **MPEG_AAC_descriptor** parameters shall take for each type of services. An explanation of "broadcast mix" and "receiver mix" is provided in section 7.

Table 5: MPEG_AAC_descriptor Parameter Values for AAC Services

Service		Values			
		AAC_service_type	receiver_mix_rqd	mainid_flag	ascv_flag
Complete Main (CM)		0	0	'1' if an associated service is present that must be mixed with the CM service, otherwise '0'	0
Main Music and Effects (ME)	Receiver Mix	1	1	1	0
Associated Visually Impaired (VI)	Broadcast Mix	2	0	0	0
	Receiver Mix		1	0	1
Associated Hearing Impaired (HI)	Broadcast Mix	3	0	0	0
	Receiver Mix		1	0	1
Associated Dialogue (D)	Receiver Mix	4	1	0	1
Associated Commentary (C)	Broadcast Mix	5	0	0	0
	Receiver Mix		1	0	1
Associated Voiceover (VO)	Receiver Mix	7	1	0	1

Definition of the service types are as follows.

6.8.1 Complete Main (CM) Audio Service

The CM type of main audio service shall contain a complete audio program (including dialogue, music, and effects). The CM service may contain from 1 to 7.1 audio channels. A CM service can be further enhanced by mixing with an associated service in the receiver (see Table 4).

6.8.2 Main Audio Service, Music and Effects (ME)

The ME type of main audio service shall contain the music and effects portion of an audio program. The ME service may contain from 1 to 7.1 audio channels. The primary program dialogue shall not be included in an ME service. The ME service may be mixed with an associated D service to produce a complete mix in the receiver.

6.8.3 Visually Impaired (VI)[2]

[2] Note that the terms "Video Description" (or "Audio Description (AD)" in DVB) is sometimes used elsewhere to refer to this type of service, but VI is used in this standard for compatibility with terminology in ATSC and FCC documents.

The VI type of associated audio service is designed to improve the experience of the visually impaired, typically by insertion of narrated descriptions of a television program's key visual elements into natural pauses between the program's dialogue. The VI associated service may be a complete program mix containing music, effects, dialogue, and a special narration that describes the picture, using any number of channels up to 7.1. Alternatively, the VI associated service may be a special narration that describes the picture content, intended to be mixed with a Complete Main service in the receiver, in which case the number of channels shall be equal to or less than the main service.

6.8.4 Hearing Impaired (HI)

The HI type of associated audio service is designed to improve the experience of the hearing impaired by improving the intelligibility of the dialogue, and in addition can serve as improvement for listening in noisy environments. The HI associated service typically is a complete program mix with enhanced intelligibility, using any number of channels up to 7.1. Alternatively, the HI associated service may be an additional audio stream, intended to be mixed with a Complete Main service or ME main audio service in the receiver, in which case the number of channels shall be equal to or less than the main service.

6.8.5 Dialogue (D)

The D type of associated audio service is intended for use with an ME main audio service. A complete audio program is formed by simultaneously decoding a D and an ME service and mixing the audio from the D service with the audio from the ME service. Multiple D services (e.g., in different languages) may be associated with a single ME service. In which case, the language of each D service is indicated in the **MPEG_AAC_descriptor**. The number of channels of the D service shall be equal to or less than the number in the ME main service.

6.8.6 Commentary (C)

The C type of associated audio service is a complete program mix containing music, effects, dialogue, and some special commentary, using any number of channels up to 7.1. Alternatively, the C associated service may be an additional special commentary audio stream, intended to be mixed with a Complete Main service or ME main audio service in the receiver, in which case the number of channels shall be equal to or less than the main service.

6.8.7 Voice-Over (VO)

The VO type of associated audio service is a monophonic service to add alternative or additional dialogue to the program and intended to be mixed with a Complete Main service in the receiver.

7. MIXING OF MAIN AND ASSOCIATED SERVICES

7.1 Overview

Associated audio services as described in section 6.5.3 are audio soundtracks that provide an additional feature or function over and above that provided by the main audio stream. The associated service may be provided using one of two schemes:

- "Broadcast mix": pre-mixed by the broadcaster and offered as an alternative audio stream.

- "Receiver mix": mixed in the receiver under the control of signaling provided by the broadcaster plus some limited control from the user.

Broadcast mix associated services are signaled in the **MPEG_AAC_descriptor** using **receiver_mix_rqd** = '0'. No special treatment is required for presentation in the receiver.

It should be noted that support for the encoding of receiver mix associated services is optional but, if present, receiver mix associated services are signaled in the **MPEG_AAC_descriptor** using **receiver_mix_rqd** = '1'. Further signaling and other information relating to receiver mix associated services are discussed in the rest of this section.

Information to control the relative levels of the main and associated services is transmitted within the header of the relevant associated service elementary stream, as described in section 7.3. The program provider is expected to determine the optimum balance of the program sound and narration, etc., and author this control information.

While a particular application for receiver mix is for the VI service, the techniques described herein are applicable to all receiver mix associated service applications.

It should be noted that mixing of an associated audio service with main audio is performed in the PCM domain and does not employ specific codec properties.

7.2 Signaling of Services to be Mixed in the Receiver

Each available main service is tagged with the appropriate identifier number (0–7) using the **mainid** parameter in the **MPEG_AAC_descriptor** for the main service. The main service or services that a receiver mix associated service may be mixed with is signaled using the **asvc** parameter of the **MPEG_AAC_descriptor** for the associated service.

Up to eight main services may be signaled as being available for mixing. However, the system is intended to enable mixing of one associated service and one main service at a time because receivers are expected to have the capability to decode only two audio streams simultaneously.

7.3 Receiver Mix Control Information Syntax and Semantics

Receiver mix associated service (AS) control information, when present, shall be coded as **AS_control_data** according to Table 6. The presence of this data is indicated by the

mixinfoexists flag in the **MPEG_AAC_descriptor**. This information shall be carried as PES_private_data within the PES packet header of the associated service in accordance with ISO/IEC 13818-1 [2], table 2-21. Because a PES_packet() always carries exactly 16 bytes as PES_private_data, fill bytes are required.

Note: The syntax and semantics of **AS_control_data** are derived from the structure named "AD_descriptor" defined in Annex E of ETSI TS 101 154 [12].

Table 6: AS_control_data

Syntax	value	No. of Bits	Identifier
AS_control_data {			
Reserved	1111	4	
AS_control_data_length		4	
AS_text_tag	0x4454474144	40	bslbf
version_text_tag		8	bslbf
AS_fade_byte	0xXX	8	bslbf
AS_pan_byte	0xYY	8	bslbf
switch (version_text_tag) {			
case '0x31':			
/* nothing */			
break;			
case '0x32':			
AS_gain_byte center	0xUU	8	bslbf
AS_gain_byte front	0xVV	8	bslbf
AS_gain_byte surround	0xWW	8	bslbf
break;			
}			
for (i=0; i<N; i++) {			
Fill_byte	0xFF	8	bslbf
}			
}			

Semantics for AS_control_data shall be:

AS_control_data_length: The number of significant bytes (excluding any fill bytes) following the length field (i.e., 11 bytes for version_text_tag == 0x31 or 8 bytes for version_text_tag == 0x32).

AS_text_tag: A string of 5 bytes, which shall be 0x4454474144, that forms a simple and unambiguous means of distinguishing this from any other PES_private_data.

version_text_tag: The AS_text_tag is extended by a single ASCII character version designator (here "1" =0x31indicates revision 1). **AS_control_data** parameter sets with the same **AS_text_tag** but a higher version number shall be backwards compatible with the present document - the syntax and semantics of the fade and pan fields will be identical but some of the reserved bytes may be used for additional signaling.

AS_fade_byte: Takes values between 0x00 (representing no fade of the main program sound) and 0xFF (representing a full fade). Over the range 0x00 to 0xFE one lsb represents a step in attenuation of the program sound of 0.3 dB giving a range of 76.2 dB. The fade value of 0xFF represents no program sound at all (i.e., mute). The rate of signaling and the expected behavior of a decoder to changes in fade byte are described below.

AS_pan_byte: Takes values between 0x00, representing a central forward presentation of the associated service, and 0xFF, each increment representing a 360/256 degree step clockwise looking down on the listener (i.e., just over 1.4 degrees. For a stereo main service the pan value shall be restricted to ±30° of the center front (i.e., to the range 0x00..0x15 and 0xEB..0xFF). The rate of signaling and the expected behavior of a decoder are described below. This parameter is used only for a mono associated service.

AS_gain_byte_center: Applied to the main program center channel. Takes values between 0x00 and 0xFF, which represents a signed value in dB, with the gain value of 0x00 representing 0dB change in level. Over the range 0x00 to 0x7F one lsb represents a step in boost of the program center of 0.6 dB giving a maximum boost of +76.2 dB. The gain value of 0x80 represents no main center level at all (i.e., mute). The gain value of 0x81 represents a maximum attenuation of -76.2 dB. Over the range 0x81 through 0xFF and 0x00 one lsb represents a step in attenuation of the program center of 0.6 dB. The gain and attenuation for each value is further clarified in Table 7.

Table 7: Gain and Attenuation Values

0x80	0x81	…	0xFE	0xFF	0x00	0x01	0x02	…	0x7F
-∞ dB	-76.2 dB	…	-1.2 dB	-0.6 dB	0 dB	+0.6 dB	+1.2 dB	…	+76.2 dB

The rate of signaling and the expected behavior of a decoder to changes in gain byte are described below.

AS_gain_byte_front: As **AS_gain_byte_center**, applied to main program left and right front channels.

ASD_gain_byte_surround: As **AS_gain_byte_center**, applied to main program all surround channels.

The maximum rate of signaling of fade, pan and gain values is determined by the number of audio PES packets per second for that associated service stream. For efficiency, several access units (AUs) of audio are typically encapsulated within one PES packet and the fade and pan values in each set of **AS_control_data set** are deemed to apply to each AU encapsulated within, and which commences in, that PES packet. All PES packets containing associated service audio streams intended for receiver mix shall include **AS_control_data**. This allows fade, pan and gain to be relatively gradual or to be abrupt, as the program material requires.

7.4 Receiver Mix Processing for Main and Associated Audio Services

This section describes the receiver mix of an associated service having up to 5.1 channels with a main service having up to 5.1 channels. The number of channels of the associated service audio stream shall be less than or equal to the number of channel of the main audio service. The particular case of mixing a mono associated service is covered in section 7.5.

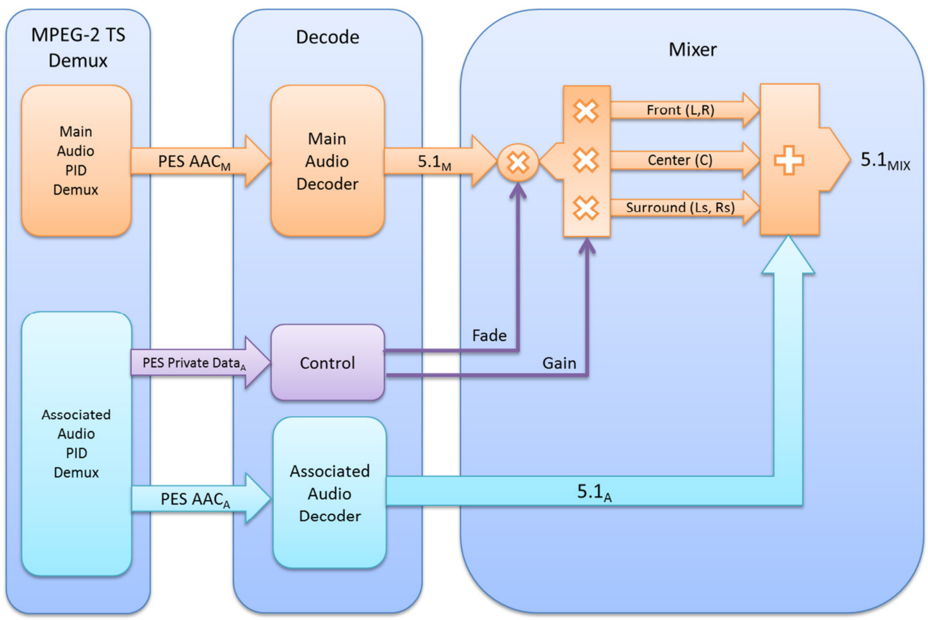

Figure 1: Mixing of 5.1 Channel Main with 5.1 Associated Audio

As shown in Figure 1, control of the relative levels of the main and associated audio services is provided by the gain metadata parameters **AS_fade_byte, AS_gain_byte_front, AS_gain_byte_center, and AS_gain_byte_surround** from the **AS_control_data** information of the associated service elementary stream. Note that **AS_pan_byte** is transmitted but in this case is not used.

7.5 Receiver Mix Processing for Mono Associated Service

VI video description associated service content is usually voice-only and is very often distributed as a mono signal. This VI signal (or any other mono associated service) may be mixed with a 5.1 or stereo main service in the receiver.

The principles of processing in a receiver mix decoder in the case of mono associated service, when main audio is 5.1 (or stereo), are similar to those described in section 7.4, with the addition of a pan control for the mono signal, and are shown diagrammatically in Figure 2.

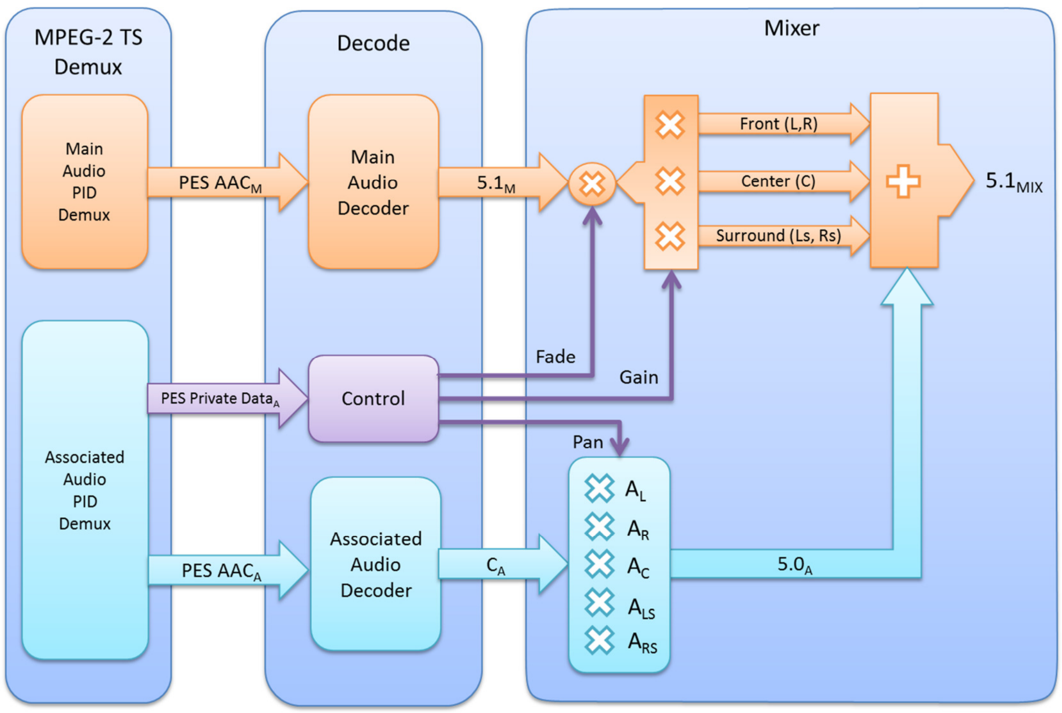

Figure 2: Mixing of 5.1 Channel Main with Mono Associated Audio

Again, control of the relative levels of the main and associated audio services is provided by the gain metadata parameters **AS_fade_byte, AS_gain_byte_front, AS_gain_byte_center, and AS_gain_byte_surround** from the **AS_control_data** information of the associated service elementary stream.

The **AS_pan_byte** from the **AS_control_data** enables the mono decoded associated service signal to be panned around the sound stage of the main program sound, thus allowing the program maker to place a narrator at any preferred position within the sound field. Table 8 shows the pan values for typical 5.1 loudspeaker locations.

Table 8: Interpretation of Audio Description Pan Values

pan value	degree	position	Main Audio 5.1					Main Audio Stereo	
			A_L	A_C	A_R	A_{LS}	A_{RS}	A_L	A_R
0x0	0°	Center	0	1	0	0	0	0.7	0.7
0x15	30°	Front Right	0	0	1	0	0	0	1
0x4E	110°	Rear Right	0	0	0	0	1	0	1
0xB2	-110°	Rear Left	0	0	0	1	0	1	0
0xEB	-30°	Front Left	1	0	0	0	0	1	0

7.6 Receiver Mix Decoder Expected Behavior

It is expected that all decoders will support decoding of receiver mix associated services. Suggestions for implementations are as follows.

A receiver mix audio decoder must maintain the relative timing between the decoded associated service signal and the decoded main program signal and between the decoded audio and the appropriate fade, pan and gain values.

If there is a valid **AS_control_data** in the encoded description signal for the selected service, the receiver mix associated decoder should present the appropriate mix of main program audio and the associated service audio to the user, attenuating the program sound by 0.3 dB per fade value increment and 0.6 dB per gain value step. If the audio decoder cannot support such small steps then the implemented attenuation should match the intended attenuation as closely as possible. For example if only -1 dB steps are possible, then fade values of 0x00 and 0x01 should map to 0 dB, 0x02, 0x03 and 0x04 should map to -1 dB, 0x05, 0x06, 0x07 and 0x08 to -2 dB, etc.

When fade and gain values are 0x00 (or in the absence of a receiver mix associated service stream) the main program audio level should be unattenuated.

It is important that a mono associated service audio is matrixed to the stereo output of a decoder so as to achieve a constant perceived volume as the audio is panned from stereo left through stereo center to stereo right. This may be achieved, for example, by using a model based on constant power as the associated service narrator is panned across the stereo sound stage.

In a stereo environment, the associated service decoder should interpret any pan values outside the ranges 0xEB..0xFF and 0x00..0x15 in the following manner. Pan values from 0x16 to 0x7F inclusive should be mapped to the value 0x15 (i.e., stereo hard right). Pan values from 0x80 to 0xEA should be mapped to the value 0xEB (i.e., stereo hard left).

If the associated service decoder detects an error in, or absence of, the **AS_control_data** in the associated service bitstream, the associated service decoder should mute the decoded associated service audio, restoring the program sound to its default unfaded amplitude, and setting the effective fade, pan and gain values to 0x00. This restoration should not be abrupt – it is recommended that under such conditions the value of fade and of pan are ramped to the default values (0x00) over a period of at least one second. Equally, if the associated service stream component is suddenly regained, the implemented value of fade, pan and gain should be ramped to the signaled values from the default values (0x00) over a similar period.

During programs for which there is no VI description or other associated service audio there is little reason to transmit an associated service stream of continual silence; in these cases the bitrate accorded to the associated service may be reassigned for other purposes. Decoders should therefore be able to respond promptly to the restoration of the associated service component at the start of a described program.

It is strongly recommended that receivers also provide the capability for the user to further adjust the level of the associated service audio relative to the main service, in order to suit particular listening conditions and personal preferences.

It is expected that receiver mix decoders will support use of audio metadata loudness normalization and dynamic range control both for the main audio service and for the associated service.