SMPTE ST 2041-2:2010

SMPTE STANDARD

Format for Non-PCM Audio in AES3 — MPEG-2 AAC and HE AAC Audio in ADTS

Table of Contents

Page

Approved
August 2, 2010

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in Part XIII of its Administrative Practices.

This SMPTE ST 2041-2 was prepared by Technology Committee 32NF.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Standard. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

The MPEG Committee of ISO/IEC has produced a number of different audio compression technologies. Each of these is nominally an "emission" codec, which means that its design does not include consideration for maintaining audio quality in multiple decode – re-encode cycles. As a result many operators will desire to keep the original compressed bitstream intact while routing, switching, and other baseband manipulations are done with associated video signals.

The first of these codecs, MPEG-1 Layer I and II audio (documented in ISO/IEC 11172-3) and MPEG-2 Layers I, II, and III audio (documented in ISO/IEC 13818-3) were intended to be bit-wise compatible within Layers I and II, while MPEG-2 Layer III used "Non-Backwards Compatible" extensions. MPEG-2 Layer III audio has not been widely used in broadcast, however some recent extensions to it have been added via ISO/IEC 14496-3 Subpart 9 which may result in wider use.

The second of these codecs, MPEG-2 AAC (documented in ISO/IEC 13818-7) is intentionally non-compatible with Layer I, II, or III, and is known as "Advanced Audio Coding." A second set of documents, which extended MPEG-2 AAC were documented as a part of MPEG-4 in ISO/IEC 14496-3. Although MPEG-4 AAC was intended to be backwards compatible with MPEG-2 AAC, no assumption should be made that an MPEG-2 decoder can decode an MPEG-4 AAC bitstream (largely due to the different transport wrappers). When the SBR tool was added to AAC, creating "HE AAC," it was documented in Amendments to both ISO/IEC 13818-7 and ISO/IEC 14496-3.

It should be noted that there are two alternative "wrappers" standardized for AAC/HE AAC, one called ADTS (widely used in Japanese Digital Broadcasting) and the other known as LATM/LOAS, since the MPEG-4 (ISO/IEC 14496-3) AAC codecs introduced new features and capabilities that require a transport format which can signal their contents. In order to be able to pass the audio bitstream without the necessity to do partial decoding to locate flags the wrapper was devised. LATM/LOAS is specified as the wrapper for MPEG-4

AAC/HE AAC audio streams within DVB transport stream broadcasting applications. As a result, this suite of SMPTE standards will document carriage of both.

The following suite of SMPTE standards defines the carriage of MPEG compressed audio bitstreams within an AES3 carrier bitstream:

SMPTE ST 2041-1, Format for Non-PCM Audio in AES3 – MPEG-1/MPEG-2 Layers I, II, and III Audio

SMPTE ST 2041-2, Format for Non-PCM Audio in AES3 – MPEG-2 AAC/HE AAC Audio in ADTS

SMPTE ST 2041-3, Format for Non-PCM Audio and Data in AES3 – MPEG-4 AAC and HE AAC Compressed Digital Audio in ADTS and LATM/LOAS Wrappers

The bitstreams defined in this standard can be carried independently of video as AES3 bitstreams or embedded into SDI or HD-SDI bitstreams in the normal manner specified by other SMPTE standards.

1 Scope

This standard specifies an interface format for the transport of MPEG-2 AAC/HE AAC data rate reduced audio streams contained in the ADTS wrapper and placed into an AES3 serial digital audio stream. This Standard supports the MPEG-2 AAC LC Profile, with or without the use of the SBR tools. This Standard is limited to carriage of a single audio elementary stream.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

The following standards contain provisions which, through reference in this text, constitute provisions of this recommended practice. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this recommended practice are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

AES3-2009, AES Standard for Digital Audio Engineering — Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data

ISO/IEC 13818-7:2006, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 7: Advanced Audio Coding (AAC)

SMPTE 337-2008, Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface

SMPTE ST 338:2010, Format for Non-PCM Audio and Data in AES3 — Data Types

SMPTE RP 168-2009, Definition of Vertical Interval Switching Point for Synchronous Video Switching

4 Definitions and Acronyms

4.1 Definitions

4.1.1 Access Unit

Smallest entity to which timing information can be attributed. An access unit is the smallest individually decodable unit. A decoder consumes access units.

4.1.2 Video Sync Point

Signal Alignment Point as defined by Annex A of SMPTE RP 168.

4.2 Acronyms

ADTS: Audio Data Transport Stream, a wrapper structure defined in ISO/IEC 13818-7, and consisting of a fixed header, a variable header, an optional error check (CRC_check) and a specified number of raw_data_blocks( ). The fixed header of the ADTS contains the syncword plus all parts of the header which are necessary for decoding and which do not change from frame to frame. The variable header of the ADTS contains header data which changes from frame to frame. The ADTS only supports a raw_data_stream( ) with only one program. The program can have up to 7 channels plus an independently switched coupling channel.

LC: Low Complexity. In the context of this specification, this term refers to the MPEG-2 AAC Low Complexity Profile as specified in ISO/IEC 13818-7. This profile is used when there are restrictions on the usage of memory and processing complexity in MPEG-2 AAC decoders.

MDCT: Modified Discrete Cosine Transform

PCM: Pulse Code Modulation

SBR: Spectral Band Replication

5 Mapping of the MPEG-2 AAC ADTS Audio Bitstream onto AES3

5.1 Overview

MPEG-2 AAC coded audio shall be transported in an AES3 data stream as a series of Data Bursts. Each Data Burst shall start with a Burst Preamble as defined by SMPTE 337, containing information about the Burst Payload, which shall follow the Burst Preamble. The Burst Payload shall consist of a single ADTS frame. The Burst Payload shall be followed by enough padding words (which shall be PCM zeros, or digital silence) to make the resulting Data Burst duration exactly match that of 1024 samples of baseband (PCM) audio that the coded audio represents.

The resulting Data Bursts shall be placed in the audio sample word/aux data fields of AES3 subframes at regular intervals in either the frame or subframe mode (see SMPTE 337, section 5). Data Bursts shall be placed in the AES3 transport, using either 16, 20, or 24 bits of the available data space. While the 24-bit mode allows more efficient use of the AES3 capacity, the 16 and 20-bit modes allow use with existing equipment limited to 16- or 20-bit operation.

A single ADTS Frame shall form the Burst Payload, as shown by Figure 1. Each ADTS Frame begins with an ADTS fixed header, followed by a variable Header, and an optional ADTS error check (CRC_check) word, followed by the Raw Data Block of AAC coded audio that represents 1024 samples of baseband (PCM) audio. If the SBR tool (HE AAC) is used, the size of the Raw Data Block is doubled to 1920 or 2048 samples (respectively). The SBR present case is shown in Figure 2.

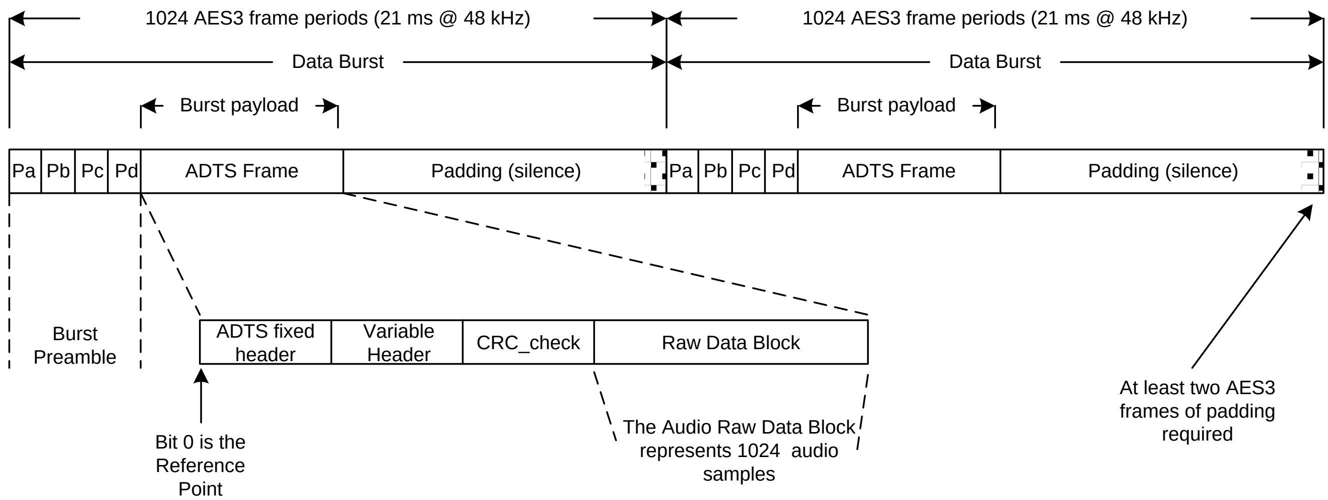

Figure 1 – MPEG-2 AAC audio data in an ADTS wrapper, transported in an AES3 stream

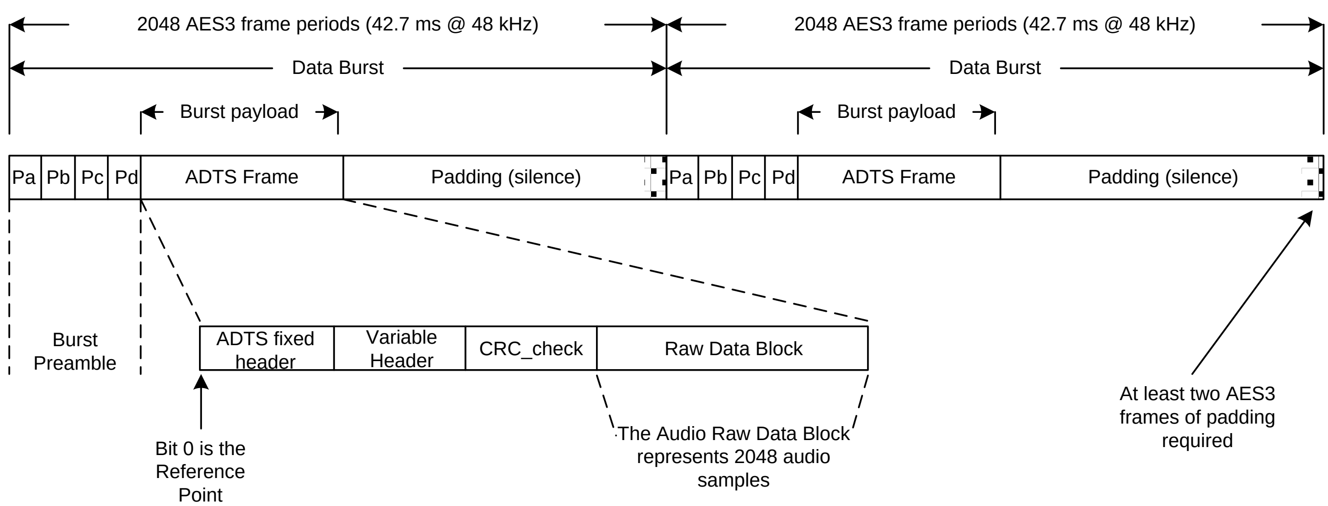

Figure 2 – MPEG-2 AAC audio data (with SBR) in an ADTS wrapper, transported in an AES3 stream

5.2 Burst_preamble

The Pc word (burst_info value) of the burst_preamble carries the data_type identifier, the data_type_dependent and the data_stream_number information (see SMPTE 337, Table 7)

5.2.1 data_type identifier

The data_type identifier shall be set to 7 per Table 1 of SMPTE ST 338, indicating that the audio is coded as an MPEG-2 AAC stream.

5.2.2 data_type_dependent

The values of the data_type_dependent bits shall be as shown in Table 1, and as described below.

Table 1 – Values of data_type_dependent field for MPEG-2 AAC in an ADTS wrapper

data_type_dependent bit number	Meaning
0-1	Reserved, must be set to '10'
2	0 => SBR data is not present 1 => SBR data is present
3	0 => Data Burst length of 1024 (2048) AES3 frames 1 => Reserved
4	0 => MPEG Surround is not used 1 => MPEG Surround is used

SBR – Bit 2 shall either be set to 0 to indicate that the SBR tool is not used or set to 1 to indicate that the SBR tool is used in the audio coding. As the ADTS header does not provide any mechanism to explicitly signal the presence of SBR data, such signaling may be either done outside the scope of this standard or by partially decoding the MPEG-2 AAC raw data to see if SBR data is present.

data_burst_length – Bit 3 shall be set to 0 to indicate that the Data Burst is 1024/2048 AES3 frames long (2048 if Bit 2 is 1).

mpeg_surround - Bit 4 shall either be set to 0 to indicate that MPEG Surround is not present or set to 1 to indicate that MPEG Surround is present.

5.2.3 data_stream_number

The data_stream_number may be set to any number from 0h to 6h. The value should be 0 (zero) indicating only one program is carried in the AES3 stream. 7h is a reserved value. See SMPTE 337, clause 7.1.3.5.

5.3 Burst Payload

The MPEG-2 AAC encoder produces a stream of Raw Data Blocks, as defined by ISO/IEC 13818-7. Each Raw Data Block contains audio data that represents 1024 audio samples of all the audio channels in a single program. The length of the Raw Data Block depends on the encoded bit rate. ADTS fixed and variable headers and an optional ADTS error check (CRC_check) word as defined by ISO/IEC 13818-7 are prepended to the Raw Data Block to form an ADTS frame. Each Burst Payload shall consist of a single ADTS frame.

Note: Some regional standards, such as ARIB B32 Part 2, require the presence of the CRC and specify other constraints upon the AAC coding used.

5.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport MPEG-2 AAC coded audio streams shall be the same as the rate at which the encoded audio was sampled. When SBR data is present the frame rate of the AES3 stream shall be the same as the output sample rate of the SBR tool.

5.5 Reference Point

The Reference Point of an MPEG-2 AAC Burst Payload shall be bit 0 of the Burst Payload, as shown in Figure 1 or Figure 2, as appropriate.

5.6 Payload Repetition Rate

MPEG-2 AAC Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur 1024 (the indicated data_block_length) AES3 frames apart when SBR data is not present in the stream, and 2048 AES3 frames apart when SBR data is present in the stream.

5.7 Decode Latency (Professional)

A reference decoder shall output the first PCM sample of the decoded audio exactly two Data Burst periods after the first bit of the first Data Burst is received by the decoder.

Note: The decoding latency of two Data Burst periods does not include the encoding latency. The encoding latency needs to be added to the decoding latency when calculating the total delay of the audio system.

5.8 Reference Position

The Reference Position of a Burst Payload is defined by the relationship of the decoded audio to an associated video signal. A Burst Payload is in the Reference Position when the decoded audio from that Burst Payload is in sync with the associated video.

The Reference Point of the Burst Payload carried in an AES3 stream whose sampling frequency is locked to the associated video signal shall therefore precede the Video Sync Point by two Data Burst periods.

Annex A Bibliography (Informative)

ARIB B32, Video Coding, Audio Coding, and Multiplexing Specifications for Digital Broadcasting, Version 2.1 (2007)

IEC 60958, Digital Audio Interface

IEC 61937-1 (2007-01), Digital Audio — Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 1: General

IEC 61937-2 (2007-05), Digital Audio – Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 2: Burst-Info

IEC 61937-4 (2003-05), Digital Audio – Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 4: Non-linear PCM Bitstreams According to the MPEG Audio Format

IEC 61937-6 (2006-01), Digital Audio – Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 6: Non-linear PCM bitstreams according to the MPEG-2 AAC and MPEG-4 AAC audio formats

IEC 61937-11 (2010-05), Digital Audio – Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 11: MPEG-4 AAC and its Extensions in LATM/LOAS

ISO/IEC 11172-3:1993, Information Technology — Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s — Part 3: Audio

ISO/IEC 13818-1 | ITU-T H.222.0, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems

ISO/IEC 13818-3:1998, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 3: Audio

ISO/IEC 14496-3:2009, Information Technology — Coding of Audio-Visual Objects — Part 3: Audio