

SMPTE STANDARD

Format for Non-PCM Audio in AES3 — MPEG Layer I, II, and III Audio



Page 1 of 11 pages

Table of Contents	Page
Foreword	2
Intellectual Property	2
Introduction.....	2
1 Scope	4
2 Conformance Notation	4
3 Normative References	4
4 Definitions and Acronyms	5
4.1 Definitions	5
4.2 Acronyms	5
5 MPEG-1 and MPEG-2 Layer I and Layer II Audio	5
5.1 Background	5
5.2 Burst_preamble.....	6
5.3 Burst Payload.....	7
5.4 AES3 Frame Rate (Sampling Frequency)	7
5.5 Reference Point	7
5.6 Payload Repetition Rate	7
5.7 Decode Latency (Professional)	7
5.8 Reference Position.....	8
6 MPEG-1 and MPEG-2 Layer III Audio	8
6.1 Background	8
6.2 Burst_preamble.....	9
6.3 Burst Payload.....	9
6.4 AES3 Frame Rate (Sampling Frequency)	9
6.5 Reference Point	9
6.6 Payload Repetition Rate	9
6.7 Decode Latency (Professional)	9
6.8 Reference Position.....	10
Annex A Bibliography (Informative)	11

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.

SMPTE ST 2041-1 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-1 Layer I and II or MPEG-2 Layers I, II, and III audio in professional applications using the AES3 serial digital audio interface. 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 11172-3:1993, Information Technology — Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mb/s — Part 3: Audio

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

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

AES3: Serial digital audio per AES3-2009.

PCM: Pulse Code Modulation

5 MPEG-1 and MPEG-2 Layer I and Layer II Audio

5.1 Background

MPEG-1 or MPEG-2 Layer I and Layer II 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 coded audio 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 either 384 or 1152 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, Overview section). 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 coded audio frame shall form the Burst Payload, as shown by Figure 1 and Figure 2. Each coded audio frame begins with a fixed header, followed by an optional CRC word, followed by the Audio Data Block of coded audio that represents either 384 or 1152 samples of baseband (PCM) audio.

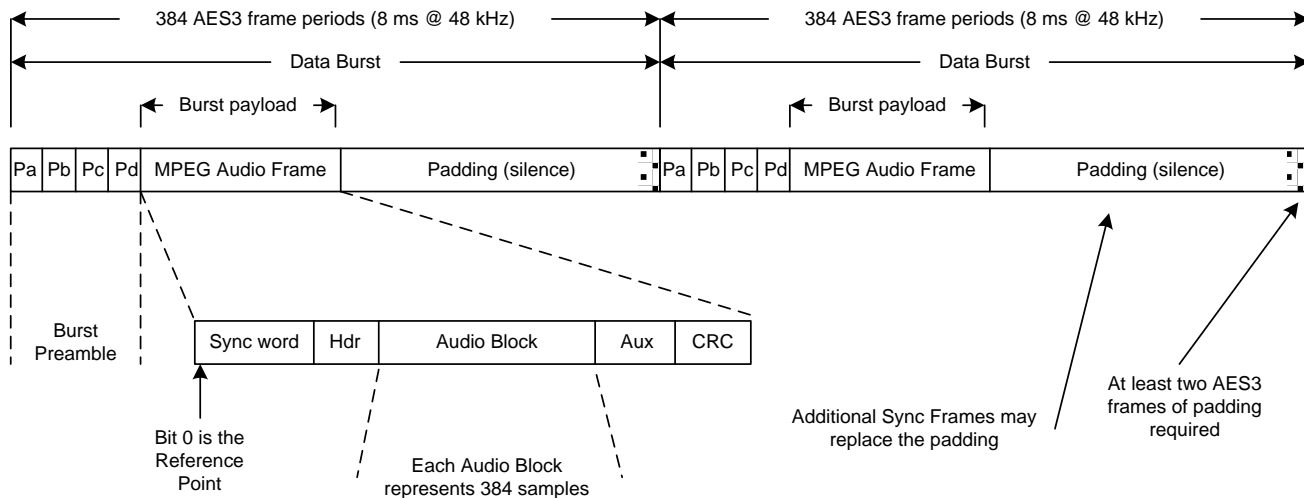


Figure 1 – MPEG-1 Layer I audio data, transported in an AES3 stream

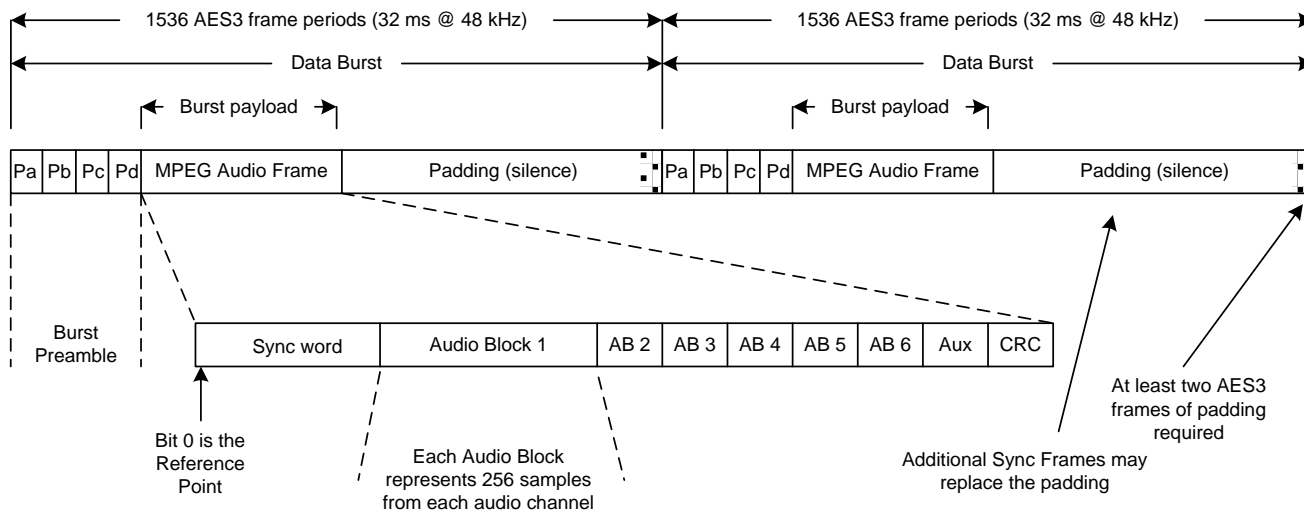


Figure 2 – MPEG-1 Layer II audio data, transported in an AES3 stream

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)

5.2.1 data_type identifier

The data_type identifier shall be set to 4, 5, or 6 per SMPTE ST 338, indicating that the compressed audio is Layer I or Layer II coded audio (either MPEG-1 or MPEG-2 as the case may be).

The data_type value of 6 indicates the coded audio is MPEG-2 Layer II with an extension field. Types 4 and 5 have no extension field.

5.2.2 data_type_dependent

When the value of data_type identifier is 4 (indicating MPEG-1 layer I audio), the value of data_type_dependent shall be zero.

When the value of data_type identifier is either 5 or 6 (indicating MPEG Layer II audio) the values of the data_type_dependent bits shall be as shown in Table 1.

Table 1 – Values of data_type_dependent field for MPEG Layer II or III

data_type_dependent bit number	Meaning
0-4	Reserved, must be set to '00000'

5.3 Burst Payload

The MPEG Layer I or II encoder produces a stream of Audio Frames, as defined by ISO/IEC 11172-7 or ISO/IEC 13818-3. Each Audio Frame contains 2 channels of audio data that represents either 384 or 1152 audio samples of each encoded audio channel in a single program. MPEG Layer I Audio Frames are 384 samples, while MPEG Layer II are 1152 samples. The Audio Frame is optionally followed by an MPEG-2 Layer II extension frame.

5.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport MPEG Layer I or II coded audio streams shall be the same as the rate at which the encoded audio was sampled.

5.5 Reference Point

The Reference Point of an MPEG Layer I or Layer II 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 Layer I or Layer II Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur 384 or 1152 (the indicated data_block_length) AES3 frames apart.

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 that is locked to the associated video signal must therefore precede the Video Sync Point by two Data Burst periods.

6 MPEG-1 and MPEG-2 Layer III Audio

6.1 Background

MPEG-1 or MPEG-2 Layer III 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 coded audio 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 1152 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, Overview section). 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 coded audio frame shall form the Burst Payload, as shown by Figure 3. Each coded audio frame begins with a fixed header, followed by a variable Header, and an optional CRC word, followed by the Raw Data Block of coded audio that represents 1152 samples of baseband (PCM) audio, followed by an optional Extension data block.

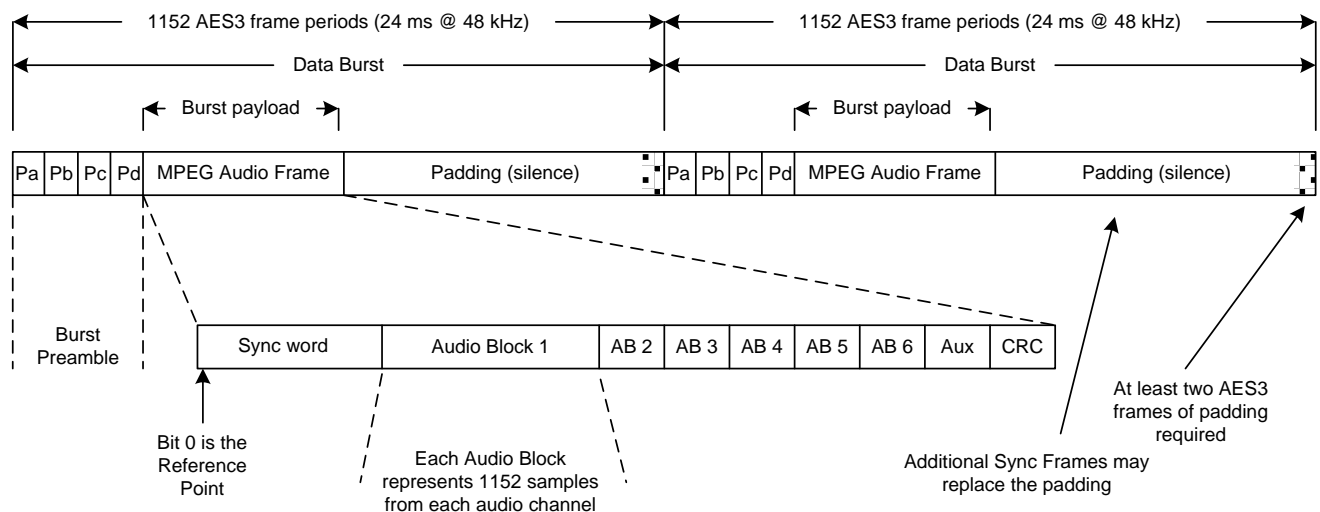


Figure 3 – MPEG-2 Layer III audio data, transported in an AES3 stream

6.2 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)

6.2.1 data_type identifier

The data_type identifier shall be set to 5 or 6 per SMPTE ST 338, indicating that the compressed audio is Layer III coded audio (either MPEG-1 or MPEG-2 as the case may be).

The data_type value of 6 indicates the coded audio has an extension field. Type 5 has no extension field.

6.2.2 data_type_dependent

The values of the data_type_dependent bits shall be as shown in Table 2.

Table 2 – Values of data_type_dependent field for MPEG Layer III

data_type_dependent bit number	Meaning
0-4	Reserved, must be set to '00000'

6.3 Burst Payload

The MPEG-2 Layer III encoder produces a stream of Audio Frames, as defined by ISO/IEC 13818-3. Each Audio Frame contains 2 channels of audio data that represents 1152 audio samples of each encoded audio channel in a single program. The Audio Frame is optionally followed by an MPEG-2 Layer III extension frame.

6.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport MPEG-2 Layer III coded audio streams shall be the same as the rate at which the encoded audio was sampled.

6.5 Reference Point

The Reference Point of an MPEG Layer III Burst Payload shall be bit 0 of the Burst Payload, as shown in Figure 3.

6.6 Payload Repetition Rate

MPEG Layer III Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur 1152 AES3 frames apart.

6.7 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 must be added to the decoding latency when calculating the total delay of the audio system.

6.8 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 that is locked to the associated video signal must therefore precede the Video Sync Point by two Data Burst periods.

Annex A Bibliography (Informative)

DVB/ETSI TS 101 154 v 1.9.1, Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG 2 Transport Stream

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 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 IEC 60958 — Part 11: MPEG-4 AAC and its Extensions in LATM/LOAS

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

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

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