

SMPTE STANDARD

for Television —

Formatting AES Audio and Auxiliary Data into Digital Video Ancillary Data Space



Page 1 of 17 pages

1 Scope

1.1 This standard defines the mapping of AES digital audio data, AES auxiliary data, and associated control information into the ancillary data space of serial digital video conforming to ANSI/SMPTE 259M or SMPTE 344M. The audio data and auxiliary data are derived from AES3, hereafter referred to as AES audio. The AES audio data may contain linear PCM audio or non-PCM data formatted according to SMPTE 337M.

1.2 Audio sampled at 48 kHz and clock locked (synchronous) to video is the preferred implementation for intrastudio applications. As an option, this standard supports AES audio at synchronous or asynchronous sampling rates from 32 kHz to 48 kHz.

1.3 The minimum, or default, operation of this standard supports 20 bits of audio data as defined in clause 3.5. As an option, this standard supports 24-bit audio or four bits of AES auxiliary data as defined in clause 3.10.

1.4 This standard provides a minimum of two audio channels and a maximum of 16 audio channels based on available ancillary data space in a given format (four channels maximum for composite digital). Audio channels are transmitted in pairs combined, where appropriate, into groups of four. Each group is identified by a unique ancillary data ID.

1.5 Several modes of operation are defined and letter suffixes are applied to the nomenclature for this standard to facilitate convenient identification of interoperation between equipment with various capabilities. The default form of operation is 48-kHz synchronous audio sampling carrying 20 bits of AES audio data and defined in a manner to ensure reception by all equipment conforming to this standard.

2 Normative references

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

AES3-2003, AES standard for Digital Audio — Digital Input-Output Interfacing — Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data (AES3)

ANSI/SMPTE 259M-1997, Television — 10-Bit 4:2:2 Component and 4fsc NTSC Composite Digital Signals — Serial Digital Interface

SMPTE 291M-1998, Television — Ancillary Data Packet and Space Formatting

SMPTE 337M-2000, Television — Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface

SMPTE 344M-2000, Television — 540 Mb/s Serial Digital Interface

SMPTE RP 165-1994, Error Detection Checkwords and Status Flags for Use in Bit-Serial Digital Interfaces for Television

SMPTE RP 168-2002, Definition of Vertical Interval Switching Point for Synchronous Video Switching

3 Definition of terms

3.1 AES audio: All the VUCP data, audio data and auxiliary data, associated with one AES digital stream as defined in AES3.

3.2 AES frame: Two AES subframes, one with audio data for channel 1 followed by one with audio data for channel 2.

3.3 AES subframe: All data associated with one AES audio sample for one channel in a channel pair.

3.4 audio control packet: An ancillary data packet occurring once a field in an interlaced system (once a frame in a progressive system) and containing data used in the operation of optional features of this standard.

3.5 audio data: 23 bits: 20 bits of AES audio associated with one audio sample, not including AES auxiliary data, plus the following 3 bits: sample validity (V bit), channel status (C bit), and user data (U bit).

3.6 audio data packet: An ancillary data packet containing audio data for 1 or 2 channel pairs (2 or 4 channels). An audio data packet may contain audio data for one or more samples associated with each channel.

3.7 audio frame number: A number, starting at 1, for each frame within the audio frame sequence. For the example in 3.8, the frame numbers would be 1, 2, 3, 4, 5.

3.8 audio frame sequence: The number of video frames required for an integer number of audio samples in synchronous operation. As an example: the audio frame sequence for synchronous 48-kHz sampling in an 30/1.001 frame/s system is 5 frames.

3.9 audio group: Consists of one or two channel pairs which are contained in one ancillary data packet. Each audio group has a unique ID as defined in clause 12.2. Audio groups are numbered 1 through 4.

3.10 auxiliary data: Four bits of AES audio associated with one sample defined as auxiliary data by AES3. The four bits may be used to extend the resolution of audio sample.

3.11 channel pair: Two digital audio channels, generally derived from the same AES audio source.

3.12 data ID: A word in the ancillary data packet which identifies the use of the data therein.

3.13 extended data packet: An ancillary data packet containing auxiliary data corresponding to, and immediately following, the associated audio data packet.

3.14 sample pair: Two samples of AES audio as defined in clause 3.1.

3.15 synchronous audio: Audio is defined as being clock synchronous with video if the sampling rate of audio is such that the number of audio samples occurring within an integer number of video frames is itself a constant integer number, as in the following examples:

<u>Audio sampling rate</u>	<u>Samples/frame, 30/1.001 fr/s video</u>	<u>Samples/frame, 25 fr/s video</u>
48.0 kHz	8008/5	1920/1
44.1 kHz	147147/100	1764/1
32.0 kHz	16016/15	1280/1

AES11 provides specific recommendations for audio and video synchronization.

NOTE – The video and audio clocks must be derived from the same source since simple frequency synchronization could eventually result in a missing or extra sample within the audio frame sequence.

4 Overview and levels of operation

4.1 Audio data derived from one or more AES frames and one or two channel pairs are configured in an audio data packet as shown in figure 1. Generally, both channels of a channel pair will be derived from the same AES audio source; however, this is not required. The number of samples per channel contained in one audio data packet will depend on the distribution of the data in a video field. As an example, the ancillary data space in some television lines may carry three samples, some may carry four samples. Other values are possible. Ancillary data space carrying no samples will not have an audio data packet.

NOTE – Receiver designers should recognize that some existing transmission equipment may transmit other sample counts, including zero. Receivers should handle correctly sample counts from zero to the limits of ancillary data space and receive buffer space.

4.2 Three types of ancillary data packets to carry AES audio information are defined.

The audio data packet carries all the information in the AES bit stream excluding the auxiliary data defined by AES3. The audio data packet is located in the ancillary data space of the digital video on most of the television lines in a field. An audio control packet is transmitted once per field in an interlaced system and once per frame in a progressive system. The audio control packet is optional for the default case of 48-kHz synchronous audio (20 or 24 bits), and is required for all other modes of operation. Auxiliary data are carried in an extended data packet corresponding to and immediately following the associated audio data packet.

4.3 Data IDs (see clauses 12.2, 13.1, and 14.1) are defined for four separate packets of each packet type. This allows for up to eight channel pairs in component video; however, there is ancillary data space for only two channel pairs (of 20 or 24 bit, 48-kHz audio) in composite video. In this standard, the audio groups are numbered 1 through 4 and the channels are numbered 1 through 16. Channels 1 through 4 are in group 1, channels 5 through 8 are in group 2, and so on.

4.4 If extended data packets are used, they are included on the same video line as the audio data packet which contains data from the same sample pair. The extended data packet follows the audio data packet and contains two 4-bit groups of auxiliary data per ancillary data word as shown in figure 1.

4.5 To define the level of support in this standard by a particular equipment, a suffix letter is added to the standard number. The default compliance is defined as level A and implements synchronous audio sampled at 48 kHz and carrying only the (20-bit) audio data packets. Distribution of samples on the television lines for level A specifically follows the uniform sample distribution as required by clause 9.1 in order to ensure interoperation with receivers limited to level A operation (see annex A for distribution analysis).

4.6 Levels of operation indicate support as listed:

A) Synchronous audio at 48 kHz, 20-bit audio data packets (allows receiver operation with a buffer size less than the 64 samples required by clause 9.2);

- B) Synchronous audio at 48 kHz, for use with composite digital video signals, sample distribution to allow extended data packets, but not utilizing those packets (requires receiver operation with a buffer size of 64 samples per clause 9.2);
- C) Synchronous audio at 48 kHz, audio and extended data packets;
- D) Asynchronous audio (48 kHz implied, other frequencies if so indicated);
- E) 44.1-kHz audio;
- F) 32-kHz audio;
- G) 32-kHz to 48-kHz continuous sampling rate range;
- H) Audio frame sequence (see clause 14.2);
- I) Time delay tracking;
- J) Non-coincident Z bits in a channel pair.

4.7 Examples of compliance nomenclature:

A transmitter that supports only 20-bit 48-kHz synchronous audio would be said to conform to SMPTE 272M-A. (Transmitted sample distribution is expected to conform to clause 9.)

A transmitter that supports 20-bit and 24-bit 48-kHz synchronous audio would be said to conform to SMPTE 272M-ABC. (In the case of level A operation, the transmitted sample distribution is expected to conform to clause 9, although a different sample distribution may be used when it is in operation conforming to levels B or C.)

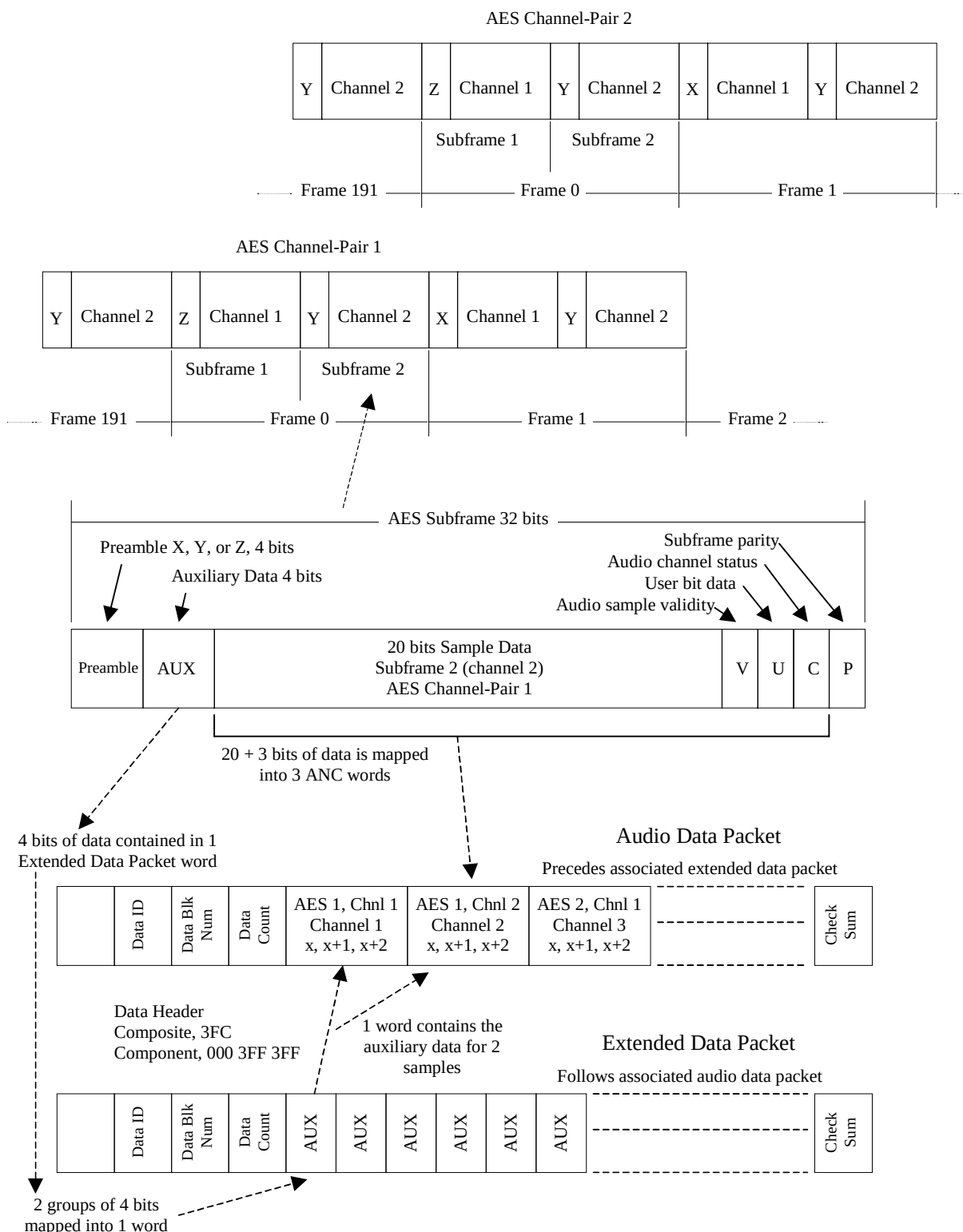
A receiver which can only accept 20-bit 48-kHz synchronous audio and requiring level A sample distribution would be said to conform to SMPTE 272M-A.

A receiver which only utilizes the 20-bit data but can accept the level B sample distribution would be said to conform to SMPTE 272M-AB since it will handle either sample distribution.

A receiver which accepts and utilizes the 24-bit data would be said to conform to SMPTE 272M-C.

Equipment that supports only asynchronous audio and only at 32 kHz, 44.1 kHz, and 48 kHz would be said to conform to SMPTE 272M-DEF.

NOTE – Implementations of this standard may achieve synchronous or asynchronous operation through the use of sample rate converters. Documented compliance levels for products should reference how AES audio is mapped into the ancillary data space only. It is recommended that product manufacturers clearly state when sample rate conversion is used to support multiple sample rates and/or asynchronous operation. It is also recommended that the use of sample rate conversion be user selectable. For example, when the AES audio data contains SMPTE 337M formatted data the use of sample rate conversion will corrupt the 337M data (see Annex B). This recommendation applies to both multiplexing (embedding) and demultiplexing (receiving) devices.



NOTE - See clause 15 and SMPTE 291M for ancillary data packet formatting

Figure 1 – Relation between AES data and audio extended data packets

5 Use of ancillary data space

5.1 For component video, audio and extended data shall be located in the data space between EAV and SAV (HANC) and may be on any line allowed by this standard.

5.2 For composite video, audio and extended data packets may be located in any ancillary data space, except that audio data shall not be present during equalizing pulses.

5.3 Audio and extended data shall not be transmitted during the horizontal ancillary data space following the normal video switching point; that is, the first horizontal interval subsequent to the switched line (see SMPTE RP 168).

5.4 Audio and extended data are not transmitted during the portion of the horizontal ancillary data space designated for error detection check-words defined in SMPTE RP 165.

NOTE — Receiver designers should recognize that some existing transmission equipment may not conform to the restrictions of clauses 5.2 through 5.4. Receivers should receive audio data transmitted in any ancillary data space.

5.5 In accordance with SMPTE 291M audio and extended data should be inserted immediately after the digital synchronization data (EAV or TRS-ID) in the available ancillary data space. For composite video, in the special case of the second vertical sync pulse in a television line, audio data shall be inserted at the earliest sample designated as ancillary data space (word 340 for 30/1.001 frame/s video rates, word 404 for 25 frame/s video rates).

6 Audio data packet formatting

6.1 The four audio channels from audio group 1 are ordered such that channels 1 and 2 make one channel pair and channels 3 and 4 make another. Audio group 2 contains channels 5 and 6 as one channel pair, and so on.

6.2 Where the audio data are derived from a single AES data stream, the data shall be ordered such that data from a subframe 1 is always transmitted before the data from a subframe 2 in the same channel pair. This means that data from subframe 1 would be placed in channel 1 (or 3, 5, ...) and data from subframe 2 would be placed in channel 2 (or 4, 6, ...).

6.3 The order that the channel pairs are transmitted within a group is not defined, however it is recommended that the channel pair containing channels 1 and 2 precede the channel pair containing channels 3 and 4.

NOTE — Receiver designers should recognize that some existing transmission equipment may not conform to the recommended channel pair ordering.

6.4 When only one channel of a channel pair is active, both channels shall still be transmitted. It is also recommended that both channel pairs within a group (all four channels) always be transmitted. If the audio signal is not derived from a single AES audio signal, then the accompanying inactive channel's audio sample bits must be set to all zeros with the V bit, C bit, and U bit set to appropriate values.

6.5 Audio channels within the same channel pair shall have the same sampling rate and synchronous or asynchronous status.

6.6 Channel pairs may be mixed with respect to their sampling rate and synchronous or asynchronous status. Each video frame will contain the appropriate number of AES audio samples for the rate used.

6.7 The audio packet length is variable. To meet the requirements of clause 8.1, the length must be short enough to allow room in the remaining ancillary data space for the extended data packet if auxiliary data are present.