
PROPOSED SMPTE STANDARD

SMPTE 299M

Revision of
SMPTE 299M-1997

for Television — 24-Bit Digital Audio Format for SMPTE 292M Bit-Serial Interface

Page 1 of 21 pages

1 Scope

1.1 This standard defines the mapping of 24-bit AES digital audio data and associated control information into the ancillary data space of a serial digital video conforming to SMPTE 292M. The audio 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 a clock frequency of 48 kHz 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 Audio channels are transmitted in groups of four, up to a maximum of 16 audio channels. Each group is identified by a unique ancillary data ID.

1.4 Audio data packets are multiplexed (embedded) into the horizontal ancillary data space of the C_b/C_r data stream, and audio control packets are multiplexed into the horizontal ancillary data space of the Y data stream.

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 referenced 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)

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

SMPTE 292M-1998, Television — Bit-Serial Digital Interface for High-Definition Television Systems

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

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 clock phase data: Audio clock phase is indicated by the number of video clocks between the first word of EAV and the video sample appearing at the same time as the audio sample at the input to the formatter.

3.5 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 process of decoding the audio data stream.

3.6 audio data: 29 bits: 24 bits of AES audio associated with one audio sample, including AES auxiliary data, plus sample validity bit (V), channel status bit (C), user data bit (U), even parity bit (P), and Z flag which is derived from the preamble of AES-3 stream. The Z bit is common to two channels of AES channel pair.

3.7 audio data packet: An ancillary data packet containing audio clock phase data, audio data for 2 channel pairs (4 channels) and error correction code. An audio data packet shall contain audio data of one sample associated with each audio channel.

3.8 audio frame number: A number, starting at 1, for each frame within the audio frame sequence. For the example in clause 3.9, 48-kHz sampling at 30.00/1.001 frames/s system, the frame numbers would be 1, 2, 3, 4, and 5.

3.9 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 at 30.00/1.001 frames/s system is 5 frames.

3.10 audio group: Consists of two channel pairs which are contained in one ancillary data packet. Each audio group has a unique ID as defined in clauses 5.1 and 6.1. Audio groups are numbered 1 through 4.

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

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

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

3.14 error correction code: BCH (31, 25) code (an error-correction method) in each bit sequence of b0-b7. Errors between the first word of ancillary data flag (ADF) through the last word of audio data of channel 4 (CH4) in user data words (UDW) will be corrected or detected within the capability of this code.

3.15 horizontal ancillary data block: An ancillary data space in the digital line blanking interval of one television line.

3.16 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. Examples are shown in table 1.

Table 1 – Examples of audio samples per frame for synchronous audio

Audio sampling rate	Samples/frame				
	30.00 frame/s	30.00/1.001 frame/s	25.00 frames/s	24.00 frames/s	24.00/1.001 frames/s
48.0 kHz	1600/1	8008/5	1920	2000	2002
44.1 kHz	1470/1	147147/100	1764	3675/2	147147/80
32.0 kHz	3200/3	16016/15	1280	4000/3	4004/3

AES11 provides specific recommendations for audio and video synchronization.

NOTE – Implementations of this standard may achieve synchronous or asynchronous operation through the use of sample rate converters. In the context of this standard, synchronous audio applies to the AES audio stream that is directly mapped into the ancillary data space, which may or may not be the AES audio stream present on device interfaces. 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 A). This recommendation applies to both multiplexing (embedding) and demultiplexing (receiving) devices.

4 Overview

4.1 Audio data derived from two channel pairs are configured in an audio data packet as shown in figure 1. Both channels of a channel pair are derived from the same AES audio source. The number of samples per channel used for one audio data packet is constant and is equal to one. The number of audio data packets in a given group is less than or equal to N_a in a horizontal ancillary data block. The definition and examples of N_a are described in clause 5.3.3.

4.2 Two types of ancillary data packets carrying AES audio information are defined and formatted per SMPTE 292M. Each audio data packet carries all of the information in the AES bit stream as defined by AES3. The audio data packet shall be located in the horizontal ancillary data space of the C_b/C_r data stream. An audio control packet shall be transmitted once per field in an interlaced system and once per frame in a progressive system in the horizontal ancillary data space of the second line after the switching point of the Y data stream.

4.3 Data ID are defined for four separate packets of each packet type. This allows for up to eight channel pairs. 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.

5 Audio data packet

5.1 Structure of audio data packet

5.1.1 The structure of the audio data packet shall be as shown in figure 2. Audio data packets shall be formatted according to the requirements of SMPTE 291M and shall include ancillary data flag (ADF), data identification (DID), data block number (DBN), data count (DC), user data words (UDW) and checksum (CS) fields as specified in SMPTE 291M. DC is always 218_h .

5.1.2 DID is defined as $2E7_h$ for audio group 1 (channel 1~4), $1E6_h$ for audio group 2 (channel 5~8), $1E5_h$ for audio group 3 (channel 9~12) and $2E4_h$ for audio group 4 (channel 13~16), respectively.

5.1.3 UDW is defined in clause 5.2. In this standard, UDW $_x$ means the xth user data word. There are always 24 words in the UDW of an audio data packet; i.e., UDW $_0$, UDW $_1$... UDW $_{22}$, UDW $_{23}$.

5.1.4 All audio channels in a given audio group shall have identical sampling rate, identical sampling phase and identical synchronous/asynchronous status.

5.1.5 For a given audio data packet, one sample of the audio data of each channel (CH1~CH4) is always transmitted. Even when only one of the four channels (CH1~CH4) is active, all audio data of the 4 channels shall be transmitted. In such case, the value of audio data, V, U, C and P bits of all inactive channels shall be set to zero.

5.2 Structure of user data words (UDW)

UDW consists of three kinds of data defined in clauses 5.2.1, 5.2.2, and 5.2.3. The description in this clause covers only audio group 1. The description for audio groups 2, 3, and 4 is similar to that for audio group 1 where channels 5, 9, and 13 correspond to channel 1; channels 6, 10, and 14 correspond to channel 2; channels 7, 11, and 15 correspond to channel 3; and channels 8, 12, and 16 correspond to channel 4, respectively.

5.2.1 CLK (audio clock phase data)

5.2.1.1 Bit assignment of CLK shall be as shown in table 2. Valid CLK data is required.

5.2.1.2 Bits of ck0 to ck12 indicate the number of video clocks between the first word of EAV and the video sample at the same time that audio sample appears at the input of the formatter. For example, the value of ck0~12 is in the range of 0 to 8191 for systems that use 74.25 MHz or 74.25/1.001 MHz clocks covered by SMPTE 292M. Examples of the relationship among video, sampling instants of digital audio and audio clock phase data for a 1080/60i system are shown in figure 3 (30 Hz frame rate) and figure 4 (30/1.001 Hz frame rate).

NOTE – Designers should recognize that some existing equipment may not recognize or support bit ck12 (see annex B).

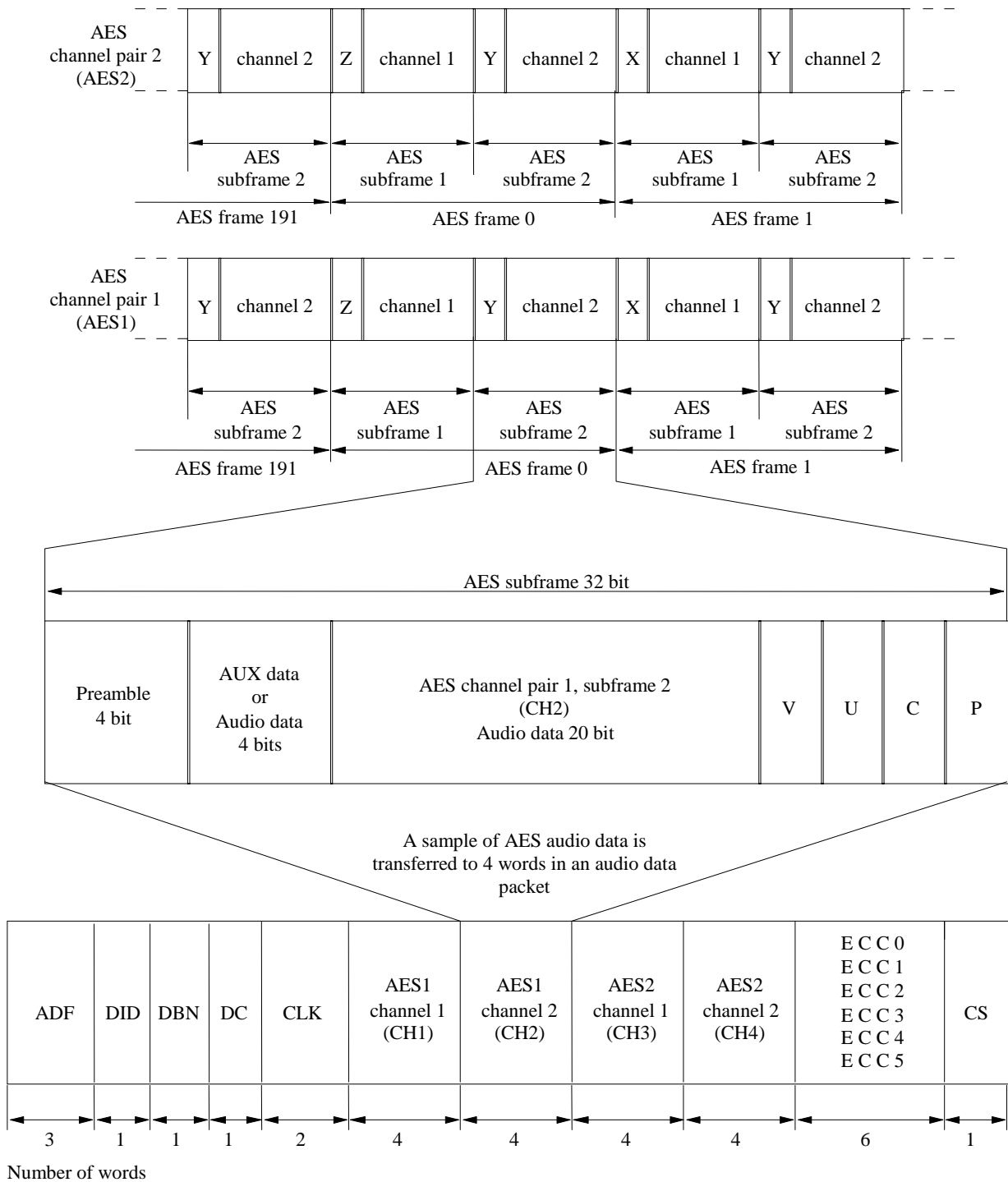


Figure 1 – Relationship between AES audio and audio data packets

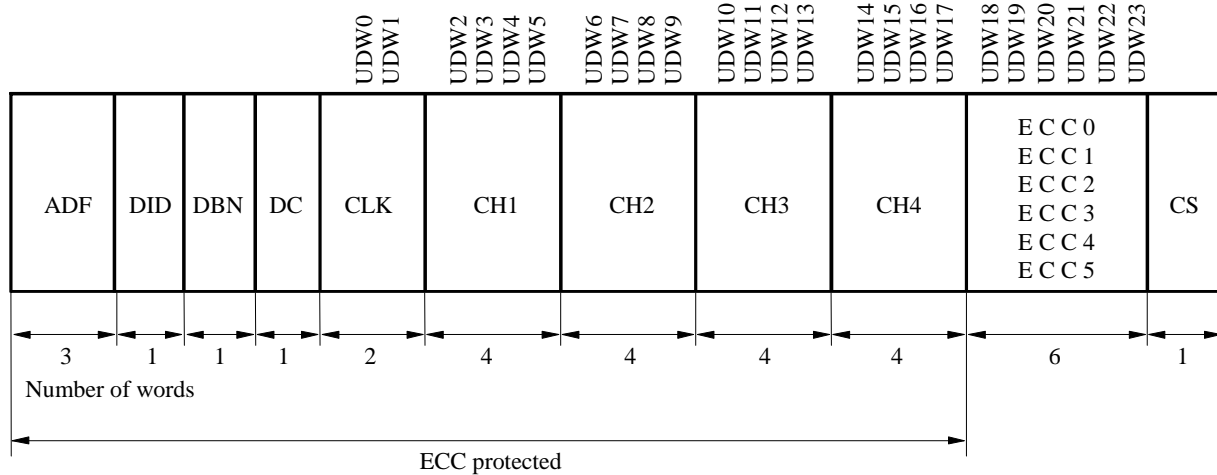


Figure 2 – Structure of audio data packets

5.2.1.3 The formatter places the audio data packet in the horizontal ancillary space following the video line during which the audio sample occurred. Following a switching point, the audio data packet is delayed one additional line to prevent data corruption.

Flag bit mpf defines the audio data packet position in the multiplexed output stream relative to the associated video data.

When bit mpf = 0, it indicates the audio data packet is located immediately after the video line during which the audio sample occurred.

When bit mpf = 1, it indicates the audio data packet is located in the second line following the video line during which the audio sample occurred.

The relationship between the multiplex position flag (mpf) and the multiplex position of the audio data packet is shown in figure 5.

5.2.2 CHn (audio data)

5.2.2.1 Bit assignment of CHn (n = 1~4) shall be as shown in table 3. All bits of an AES subframe are transparently transferred to four consecutive UDW words (UDW4n-2, UDW4n-1, UDW4n, UDW4n+1). UDW2 through UDW17 are always used for CHn in audio data packets.

5.2.2.2 Bit 3 of UDW2 and UDW10 indicates the status of the Z flag which corresponds to the AES block sync. The Z bit in UDW2 is for CH1 and CH2, and in UDW10 for CH3 and CH4, respectively.

5.2.2.3 Bits b0 through b2 in UDW2, UDW6, UDW10, and UDW14, and bit b3 in UDW6 and UDW14 are set to zero.

Table 2 – Bit assignment of CLK

Bit number	UDW0	UDW1
b9(MSB)	Not b8	Not b8
b8	Even parity ¹⁾	Even parity ¹⁾
b7	ck7 audio clock phase data	Reserved (set to 0)
b6	ck6 audio clock phase data	Reserved (set to 0)
b5	ck5 audio clock phase data	ck12 audio clock phase data (MSB)
b4	ck4 audio clock phase data	mpf multiplex position flag
b3	ck3 audio clock phase data	ck11 audio clock phase data
b2	ck2 audio clock phase data	ck10 audio clock phase data
b1	ck1 audio clock phase data	ck9 audio clock phase data
b0 (LSB)	ck0 audio clock phase data (LSB)	ck8 audio clock phase data

¹⁾ Even parity for b0 through b7.

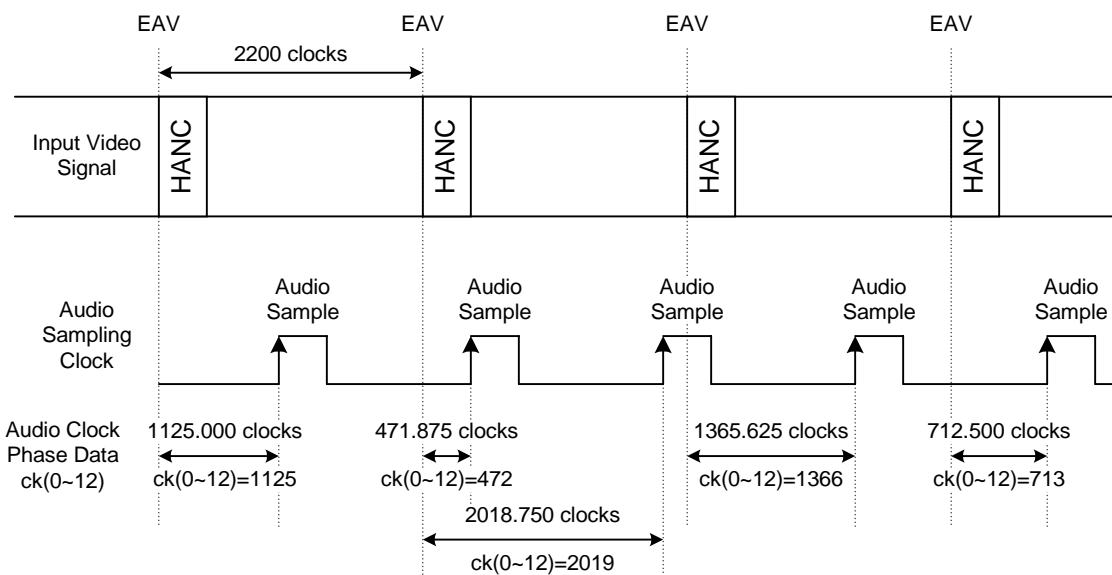


Figure 3 – Relationship between video lines, sampling instants of digital audio, and audio clock phase data (informative example – 1080/60i system with 48-kHz audio sampling rate and 30.00 Hz video frame rate)¹

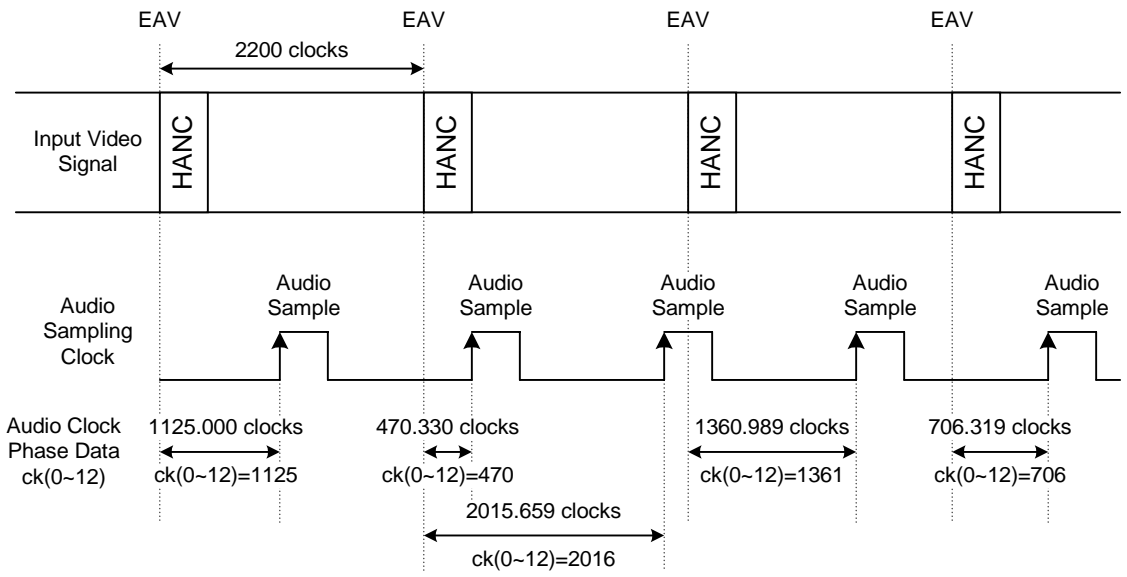
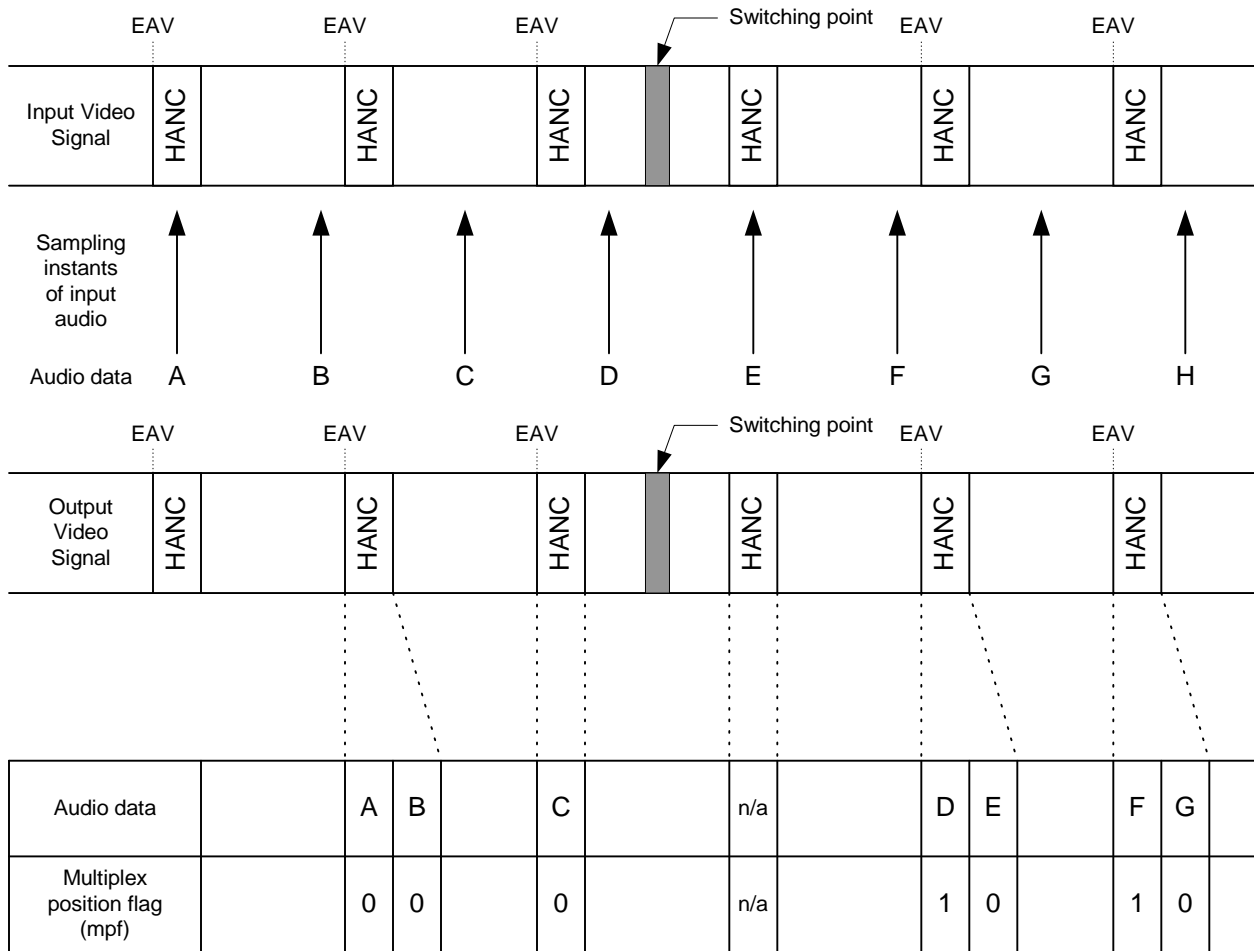


Figure 4 – Relationship between video lines, sampling instants of digital audio, and audio clock phase data (informative example - 1080/60i system with 48-kHz audio sampling rate and 30.00/1.001 Hz video frame rate)¹

NOTE – In figure 3 and figure 4, "clocks" refers to the video sampling clock as defined in clause 5.2.1.2.



NOTES:

- 1 For example, for samples A, B, C, E and G, mpf = 0 because the ancillary data packet is multiplexed in the horizontal ancillary data space of the next line relative to the input timing of the audio sample.
- 2 N/A shows that the line subsequent to the switching point precludes the insertion of ancillary data packets
3. For example, for samples D and F, mpf = 1 because the ancillary data packet is multiplexed in the horizontal ancillary data space of the second line relative to the input timing of the audio sample.

Figure 5 – Relationship between the multiplex position flag (mpf) and the multiplex position of audio data packets

Table 3 – Bit-assignment of audio data (CHn)

	Bit number	UDW2	UDW3	UDW4	UDW5
CH1	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity ¹	Even parity ¹	Even parity ¹	Even parity ¹
	b7	aud ₁ 3	aud ₁ 11	aud ₁ 19	P ₁
	b6	aud ₁ 2	aud ₁ 10	aud ₁ 18	C ₁
	b5	aud ₁ 1	aud ₁ 9	aud ₁ 17	U ₁
	b4	aud ₁ 0 (LSB)	aud ₁ 8	aud ₁ 16	V ₁
	b3	Z	aud ₁ 7	aud ₁ 15	aud ₁ 23(MSB)
	b2	0	aud ₁ 6	aud ₁ 14	aud ₁ 22
b1	0	aud ₁ 5	aud ₁ 13	aud ₁ 21	
b0 (LSB)	0	aud ₁ 4	aud ₁ 12	aud ₁ 20	
CH2	Bit number	UDW6	UDW7	UDW8	UDW9
	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity ¹	Even parity ¹	Even parity ¹	Even parity ¹
	b7	aud ₂ 3	aud ₂ 11	aud ₂ 19	P ₂
	b6	aud ₂ 2	aud ₂ 10	aud ₂ 18	C ₂
	b5	aud ₂ 1	aud ₂ 9	aud ₂ 17	U ₂
	b4	aud ₂ 0 (LSB)	aud ₂ 8	aud ₂ 16	V ₂
	b3	0	aud ₂ 7	aud ₂ 15	aud ₂ 23(MSB)
b2	0	aud ₂ 6	aud ₂ 14	aud ₂ 22	
b1	0	aud ₂ 5	aud ₂ 13	aud ₂ 21	
b0 (LSB)	0	aud ₂ 4	aud ₂ 12	aud ₂ 20	
CH3	Bit number	UDW10	UDW11	UDW12	UDW13
	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity ¹	Even parity ¹	Even parity ¹	Even parity ¹
	b7	aud ₃ 3	aud ₃ 11	aud ₃ 19	P ₃
	b6	aud ₃ 2	aud ₃ 10	aud ₃ 18	C ₃
	b5	aud ₃ 1	aud ₃ 9	aud ₃ 17	U ₃
	b4	aud ₃ 0 (LSB)	aud ₃ 8	aud ₃ 16	V ₃
	b3	Z	aud ₃ 7	aud ₃ 15	aud ₃ 23(MSB)
b2	0	aud ₃ 6	aud ₃ 14	aud ₃ 22	
b1	0	aud ₃ 5	aud ₃ 13	aud ₃ 21	
b0 (LSB)	0	aud ₃ 4	aud ₃ 12	aud ₃ 20	
CH4	Bit number	UDW14	UDW15	UDW16	UDW17
	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity ¹	Even parity ¹	Even parity ¹	Even parity ¹
	b7	aud ₄ 3	aud ₄ 11	aud ₄ 19	P ₄
	b6	aud ₄ 2	aud ₄ 10	aud ₄ 18	C ₄
	b5	aud ₄ 1	aud ₄ 9	aud ₄ 17	U ₄
	b4	aud ₄ 0 (LSB)	aud ₄ 8	aud ₄ 16	V ₄
	b3	0	aud ₄ 7	aud ₄ 15	aud ₄ 23(MSB)
b2	0	aud ₄ 6	aud ₄ 14	aud ₄ 22	
b1	0	aud ₄ 5	aud ₄ 13	aud ₄ 21	
b0 (LSB)	0	aud ₄ 4	aud ₄ 12	aud ₄ 20	

NOTES

- 1 Even parity for b0 through b7.
- 2 Z = AES block sync.
- 3 Un = AES user bit of CHn.
- 4 Pn = AES parity bit of CHn.
- 5 aud (0-23) =24-bit AES audio data of CHn.
- 6 Vn = AES sample validity bit of CHn.
- 7 Cn = AES channel status bit of CHn.
- 8 Value of Vn, Un, Cn and Pn is equal to that of AES subframe, respectively.

5.3 Multiplexing of audio data packet

5.3.1 Only the horizontal ancillary data space of the color-difference data stream (C_b/C_r) shall be used for transmission of the audio data packet.

5.3.2 The audio data packet shall not be multiplexed into the horizontal ancillary data space of the line subsequent to the switching point defined by the source format. As an example, the ancillary data space available for audio data packet in the 1080/60i system is shown in figure 7.

5.3.3 The number of samples per audio channel which can be multiplexed in one horizontal ancillary data space is less than or equal to N_a (Number of Audio samples), where N_a is defined in the following equation:

$$\begin{aligned}
 N_o &= \text{Int}(\text{audio sample rate}/\text{line frequency}) + 1 \\
 \text{if } &= N_o \times (\text{the number of total line per video frame} - \text{the number of switching line per video frame}) \\
 &< \text{the number of audio samples per video frame} \\
 \text{then } &N_a = N_o + 1 \\
 \text{else } &N_a = N_o
 \end{aligned}$$

When two or more samples of the audio data are transmitted in one horizontal ancillary data block, the packet of the audio sample which appears earlier at the input of the formatter shall be transmitted first.

NOTE – Some video formats may require up to 4 samples per data block (i.e. $N_a=4$).

5.3.4 An audio data packet shall be multiplexed in the horizontal ancillary data space of the first or second line following the line during which the audio sample occurred at the input of the formatter.

NOTE – Audio phase must be maintained across the audio groups carrying multiple-channel audio.

5.3.5 The audio data packet shall be multiplexed following the CRC, which is defined in SMPTE 292M.

5.3.6 When more than two audio data packets are transmitted in one horizontal ancillary data block, the audio data packets shall be contiguous with each other.

6 Audio control packet

6.1 Structure of audio control packet

6.1.1 The structure of the audio control packet shall be as shown in figure 8. Audio control packets shall be formatted according to the requirements of SMPTE 291M and shall include ancillary data flag (ADF), data identification (DID), data block number (DBN), data count (DC), user data words (UDW) and checksum (CS) fields as specified in SMPTE 291M. DC is always $10B_h$ and DBN is always 200_h .

6.1.2 DID is defined as $1E3_h$ for audio group 1 (channels 1~4), $2E2_h$ for audio group 2 (channels 5~8), $2E1_h$ for audio group 3 (channel 9~12), and $1E0_h$ for audio group 4 (channel 13~16), respectively.

6.1.3 UDW is defined in clause 6.2. In this standard, UDW $_x$ means the xth user data word. There are always 11 words in the UDW of an audio control packet; i.e., UDW $_0$, UDW $_1$... UDW $_9$, UDW $_{10}$.

6.2 Structure of user data words (UDW)

UDW consists of five types of data defined in clauses 6.2.1 through 6.2.5. The description in this clause covers only audio group 1. The description for audio groups 2, 3, and 4 is similar to audio group 1 where channels 5, 9, and 13 correspond to channel 1; channels 6, 10, and 14 correspond to channel 2; channels 7, 11, and 15 correspond to channel 3; channels 8, 12, and 16 correspond to channel 4, respectively.

	1920	1924	1926	1928	2195 2196	2199 0	Sample number	1919
Line number								
1								Vertical blanking
6								
7								Switching point
8								
9								Vertical blanking
20								
21	EAV	LN	CRCC	Available area	SAV			Active video
560								
561								Vertical blanking
568								
569								Switching point
570								
571								Vertical blanking
583								
584								Active video
1123								
1124								Vertical blanking
1125								

Figure 7 – Ancillary data space of C_b/C_r data stream available for transmission of audio data packets (1080/60i system)

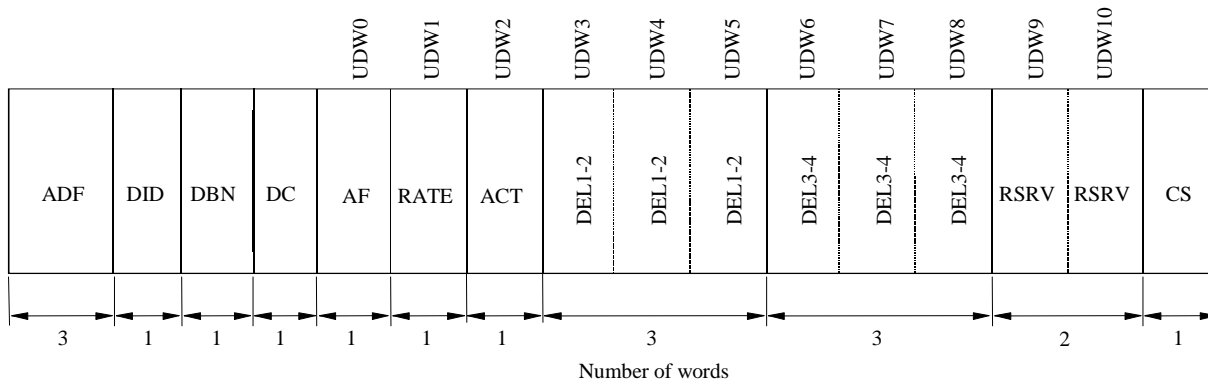


Figure 8 – Structure of audio control packets

6.2.1 AF (audio frame number data)

6.2.1.1 Audio frame number data (AF) provide a sequential numbering of video frames to indicate where they fall in the progression of non-integer number of samples per video frame (audio frame sequence). The first number of the sequence is always 1 and the final number is equal to the length of the audio frame sequence. A value of AF equal to all zeros indicates that frame numbering is not available.

6.2.1.2 The bit-assignment of the AF shall be as shown in table 5. The AF is common for all channels in a given audio group.

6.2.1.3 For correct use of the audio frame number, the audio frame sequence shall be defined. Three synchronous sampling rates are defined in this standard (see clause 3.16).

All audio frame sequences are based on two integer numbers of samples per frame (m and m+1) with audio frame numbers starting at 1 and proceeding to the end of the sequence. Odd-numbered audio frame (1, 3, 5, etc.) have the larger integer number of samples and even-numbered audio frames (2, 4, 6, etc.) have the smaller integer number of samples with the exception tabulated in table 6.

NOTE – Receiver designers should be aware that some existing equipment may not conform to the sequence restriction specifications of clause 6.2.1.3. Receivers should have the ability to receive audio data sequences correctly even when clause 6.2.1.3 is not implemented.

6.2.1.4 When channel pairs in a given audio group are operating in asynchronous mode, the AF word in the audio control packet is not used and b0 – b8 should be set to zero.

6.2.2 RATE (Sampling rate)

6.2.2.1 The sampling rate for all channel pairs is defined by the word RATE. The bit assignment of RATE shall be as shown in table 7.

6.2.2.2 The sync mode bit asx, when set to one, indicates that the channel pairs in a given audio group are operating asynchronously.

6.2.2.3 The rate code is currently defined as shown in table 8.

6.2.3 ACT

The word ACT indicates active channels. Bits a1 to a4 are set to one for each active channel in a given audio group; otherwise, they are set to zero. The bit assignment of ACT is shown in table 9.

6.2.4 DELm-n

6.2.4.1 The words DELm-n indicate the amount of accumulated audio processing delay relative to video, measured in audio sample intervals, for each channel pair of CHm and CHn.

6.2.4.2 The bit assignment of DELm-n shall be as shown in table 10. The e bit is set to one to indicate valid audio delay data. The delay words are referenced to the point where the AES/EBU data are input to the formatter. The delay words represent the average delay value, inherent in the formatting process, over a period no less than the length of the audio frame sequence plus any preexisting audio delay.

6.2.4.3 The audio delay data (del 0 – del 25) is represented in the format of 26-bit twos complement. Positive values indicate that the video leads the audio.

6.2.5 RSRV

6.2.5.1 The words marked RSRV are reserved for future use.

6.2.5.2 The bit assignment of RSRV word shall be as shown in table 11.

Table 5 – Bit-assignment of AF

Bit number	UDW0
	AF
b9 (MSB)	not b8
b8	f8 audio frame number (MSB)
b7	f7 audio frame number
b6	f6 audio frame number
b5	f5 audio frame number
b4	f4 audio frame number
b3	f3 audio frame number
b2	f2 audio frame number
b1	f1 audio frame number
b0 (LSB)	f0 audio frame number (LSB)

6.3 Multiplexing of the audio control packets

6.3.1 The audio control packets shall be transmitted once every field in an interlaced system and once per frame in a progressive system.

6.3.2 The audio control packets shall be transmitted in the horizontal ancillary data space of the second line after the switching point of Y data stream. For example, since the switching point for 1080/60i system exists in line 7 and 569, the audio control packets are transmitted in the horizontal ancillary data space of line 9 and line 571 of the Y data stream. Ancillary data space available for the transmission of audio control packets for this example is shown in figure 9.

Table 6 – Exceptions to audio frame sequences

Television system	Sampling rate (kHz)	Frame sequence	Basic numbering		Exceptions	
			Samples per odd audio frame (m)	Samples per even audio frame (m+1)	Frame number	Number of samples
30 frame/s	48.0	1	1 600		none	
	44.1	1	1 470		none	
	32.0	3	1 067	1 066	none	
30.00/1.001 frame/s	48.0	5	1 602	1 601	none	
	44.1	100	1 472	1 471	23, 47, 71	1 471
	32.0	15	1 068	1 067	4, 8, 12	1 068

Table 7 – Bit assignment of RATE

Bit number	UDW1
	RATE
b9 (MSB)	not b8
b8	Reserved (set to 0)
b7	Reserved (set to 0)
b6	Reserved (set to 0)
b5	Reserved (set to 0)
b4	Reserved (set to 0)
b3	X2 (MSB)
b2	X1 Rate code
b1	X0 (LSB)
b0 (LSB)	asx 0 = synchronous audio 1 = asynchronous audio

Table 8 – Assignment of rate code

X2	X1	X0	Sample rate
0	0	0	48.0 kHz
0	0	1	44.1 kHz
0	1	0	32.0 kHz
1	1	1	Free running
0	1	1	Reserved
	:		:
1	1	0	Reserved

Table 9 – Bit assignment of ACT

Bit number	UDW2
	ACT
b9 (MSB)	Not b8
b8	Even parity ¹⁾
b7	Reserved (set to 0)
b6	Reserved (set to 0)
b5	Reserved (set to 0)
b4	Reserved (set to 0)
b3	a4 active: 1, inactive: 0 (CH4)
b2	a3 active: 1, inactive: 0 (CH3)
b1	a2 active: 1, inactive: 0 (CH2)
b0 (LSB)	a1 active: 1, inactive: 0 (CH1)

¹⁾ Even parity for b0 through b7.

Table 10 – Bit assignment of DELm-n

Bit number	UDW3	UDW4	UDW5	UDW6	UDW7	UDW8
	DEL1-2			DEL3-4		
b9 (MSB)	Not b8	Not b8	Not b8	Not b8	Not b8	Not b8
b8	del 7	del 16	del 25 (±)	del 7	del 16	del 25 (±)
b7	del 6	del 15	del 24 (MSB)	del 6	del 15	del 24 (MSB)
b6	del 5	del 14	del 23	del 5	del 14	del 23
b5	del 4	del 13	del 22	del 4	del 13	del 22
b4	del 3	del 12	del 21	del 3	del 12	del 21
b3	del 2	del 11	del 20	del 2	del 11	del 20
b2	del 1	del 10	del 19	del 1	del 10	del 19
b1	del 0 (LSB)	del 9	del 18	del 0 (LSB)	del 9	del 18
b0 (LSB)	e	del 8	del 17	e	del 8	del 17

Table 11 – Bit assignment of RSRV

Bit number	UDW9	UDW10
	RSRV	RSRV
b9 (MSB)	Not b8	Not b8
b8	Reserved (set to 0)	Reserved (set to 0)
b7	Reserved (set to 0)	Reserved (set to 0)
b6	Reserved (set to 0)	Reserved (set to 0)
b5	Reserved (set to 0)	Reserved (set to 0)
b4	Reserved (set to 0)	Reserved (set to 0)
b3	Reserved (set to 0)	Reserved (set to 0)
b2	Reserved (set to 0)	Reserved (set to 0)
b1	Reserved (set to 0)	Reserved (set to 0)
b0 (LSB)	Reserved (set to 0)	Reserved (set to 0)

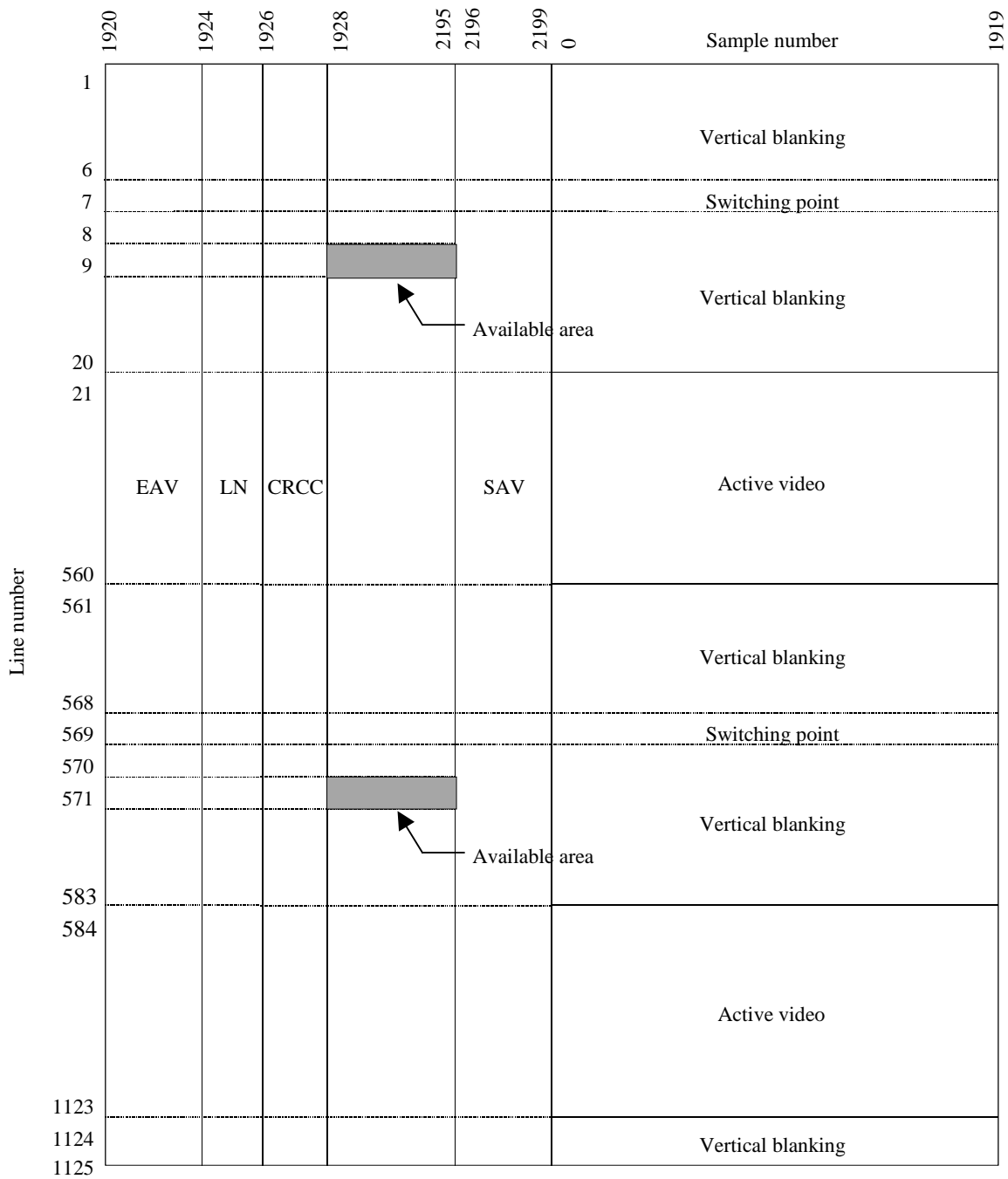


Figure 9 – Ancillary data space of Y data stream available for transmission of audio control packets (1080/60i system)

Annex A (informative)**Recommendations for handling of SMPTE 337M non-PCM data**

While this standard is written in terms of the AES data containing linear PCM audio, the AES data may contain SMPTE 337M formatted data which may include compressed (bit-rate reduced) audio or other types of non-audio data. Implementers should take this into consideration and when possible include support for SMPTE 337M data compatibility. Users should be aware that not all devices compliant with the standard may properly handle SMPTE 337M data. This section contains recommendations for compatibility with SMPTE 337M data that apply to both implementers and users.

A.1 Levels of operation

It is recommended that operation be restricted to 48 kHz synchronous modes when SMPTE 337M data is present in the AES data. Sample rate conversion should not be used to implement synchronous operation.

A.2 PCM processing

Any PCM type processing performed on the AES data stream within multiplexing or receiving devices should be defeated or bypassed when SMPTE 337M data is present as such processing will corrupt the SMPTE 337M data. Examples of PCM processing include gain changes, sample rate conversion, truncation, dithering, cross-fades, etc.

A.3 AES channel status data

AES channel status words contain useful information for detecting and identifying the presence of SMPTE 337M data with the AES data stream. It is recommended that all devices compliant with this standard maintain and transmit AES channel status information that is present on their input.

A.4 Additional receiving device recommendations

Receiving devices may include specific processing for handling dynamic bitstream changes, such as when the received SMPTE 292M signal has been switched. This processing may include handling of receive buffer overflow and underflow conditions resulting from the switch, especially in the case of 30/1.001 frame/s systems. Nominally this processing is meant to minimize the audibility of the disruption for linear PCM signals. Typical processing may include periodic AES data word (audio sample) drops or repeats to maintain receive buffer fullness and PCM type processing of additional AES data words to minimize the audibility of the drops or repeats. This processing is sub-optimal when dealing with SMPTE 337M data.

If possible it is recommended that receiving devices include the ability to detect the presence of SMPTE 337M data and restrict drop or repeat locations to AES data words not containing SMPTE 337M data. Any PCM processing should also be disabled to minimize modification of AES data words. If detection of SMPTE 337M data is not possible it is recommended that *whenever possible* drop or repeat locations be restricted to the AES data words immediately adjacent to the vertical interval switching area.

Annex B (informative)
Recommendations for handling legacy implementations

The 720/24p video format, as defined in SMPTE 296M, defines a total line length of 4125 pixels, which requires a resolution of 13 bits for the CLK audio clock phase data words. Previous versions of SMPTE 299M (299M-1997 and earlier) only offer a maximum of 12 bits resolution (ck(0~11)), providing a maximum CLK value of 4095. This revision of SMPTE 299M adds an additional bit in the audio data packet as the MSB (ck12) for the audio clock phase data, providing 13 bits resolution

NOTE – In SMPTE 299M (299M-1997 and earlier, ck12 was designated as the multiplex position flag. In this revision of SMPTE 299M, this bit is now renamed mpf and it's bit position in UDW1 remains unchanged from SMPTE 299M-1997.

Some legacy devices may not support this additional clock bit in the audio clock phase data. Some legacy audio formatting devices may hold the clock phase value at a maximum value when reached, until reset at the end of the line. This will produce a small amount of audio phase jitter for the period of one sample. Alternatively, the audio clock phase value may wrap around through zero in these legacy formatting devices.

To overcome these issues, it is recommended that audio receiver implementations should check for all of the above cases. On detection of the maximum value, a comparison can be made between previous clock phases and the correct position interpolated. If the clock phase data value starts to decrease within the same video line, the decoder should check to see if bit 5 (ck12) of UDW1 in the audio data packet is set. If ck12 is set, the correct 13-bit value of the audio clock phase data should be used. If ck12 is not set, the correct position should be interpolated.

Figure B.1 shows an example of the relationship between the video clock, audio sampling instants, and the audio clock phase data for a 720/24p video signal with 48-kHz audio sampling.

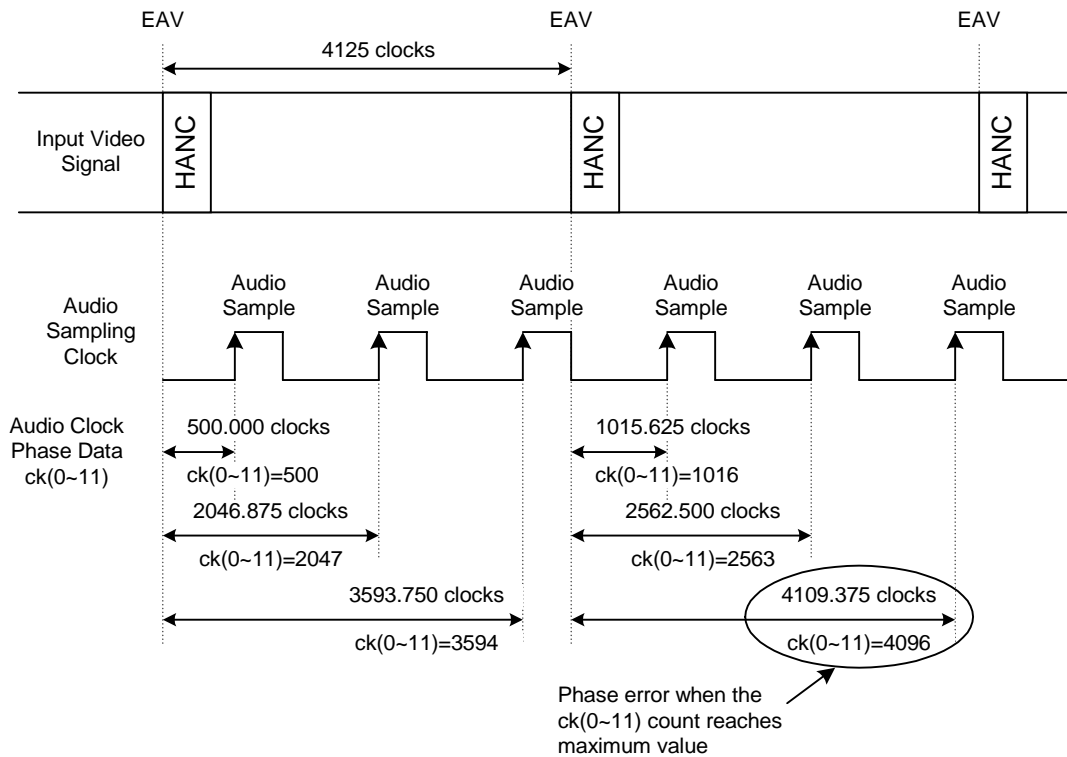


Figure B.1 – Relationship among video, audio sampling instants of digital audio, and audio clock phase data (example – 720/24p system with 48-kHz audio sampling rate and 24.00-Hz video frame rate)

Annex C (informative)

Bibliography

AES11-1997, AES Recommended Practice for Digital Audio Engineering — Synchronization of Digital Audio Equipment in Studio Operations

SMPTE 260M-1999, Television — High-Definition Production System — Digital Representation and Bit-Parallel Interface

SMPTE 272M-2004, Television — Formatting AES Audio and Auxiliary Data into Digital Video Ancillary Data Space

SMPTE 274M-2003, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE 296M-2001, Television — 1280 x 720 Progressive Image Sample Structure -- Analog and Digital Representation and Analog Interface

SMPTE 349M-2001, Television — Transport of Alternate Source Image Formats through SMPTE 292M