SMPTE ST 2020-1:2014
Revision of SMPTE 2020-1-2008

SMPTE STANDARD

Format of Audio Metadata and Description of the Asynchronous Serial Bitstream Transport

Table of Contents

Page

List of Figures

List of Tables

Approved
August 16, 2014

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 its Standards Operations Manual.

SMPTE ST 2020-1 was prepared by Technology Committee 24TB.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. 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 document.

Audio data-rate reduction technologies can use metadata, which is multiplexed into the encoded audio bitstream, to describe the encoded audio and convey information that precisely controls downstream encoders and decoders.

Audio metadata is first created during program creation or mastering. It might need to be carried in the vertical ancillary data space of a digital television signal.

There is no specific provision in this standard for ensuring that the relative timing between the video and its embedded VANC data is correct. The only timing relationship that exists is created when the data is embedded in the VANC. Once that relationship is established, the deterministic nature of the serial digital interface ensures that the relationship is preserved

There are currently two methods of mapping the metadata into the vertical ancillary (VANC) data space. The following suite of SMPTE Standards, define the basic characteristics of the audio metadata, its serial transport used to convey the metadata between devices, and both of the mappings into the VANC space.

SMPTE ST 2020-1, Format of Audio Metadata and Description of the Asychronous Serial Bitstream Transport

SMPTE ST 2020-2, Vertical Ancillary Data Mapping of Audio Metadata — Method A

SMPTE ST 2020-3, Vertical Ancillary Data Mapping of Audio Metadata — Method B

1 Scope

This Standard defines the basic structure and timing requirements of an asynchronous serial audio metadata stream with respect to an associated video signal. These requirements apply to the serial metadata streams either present on asynchronous serial data communications links between devices or present in the vertical ancillary (VANC) data packets transported by the video signal. The Standard also defines the Data ID (DID) and Secondary Data ID (SDID) values used to identify the VANC packets themselves and to associate the metadata in VANC packets with specific audio programs.

This Standard defines levels of operation for two methods of mapping the metadata into ancillary data packets.

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.

3 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.

ANSI TIA/EIA-485-A 1998 (R 2003), Electrical Characteristics of Generators and Receivers for Use in Balanced Digital Multipoint Systems

SMPTE ST 291-1:2011, Ancillary Data Packet and Space Formatting

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

4 Definitions and Acronyms

4.1
Audio Channel
An audio signal representing the part of a program to be reroduced by one of the speakers, earpieces etc. of a sound reproduction system. A stereo program, for example, is composed of a Left channel audio signal and a Right channel audio signal.

4.2
Audio Channel Pair
Two audio signals representing part of or all of a program. The term may also be used to refer to the two signals carried by an AES3 digital audio interface.

4.3
Audio Program
A collection of one or more audio channels.

4.4
FPS
Frames per second.

4.5
Metadata Frame
A burst or frame of data that carries the metadata information related to a group of audio samples, which are temporally related to a video frame.

4.6
Serial Metadata Stream
The format and transport of a metadata frame as an asynchronous serial data signal on an ANSI/TIA/EIA-485-A physical interface.

4.7
VANC
Vertical Ancillary data space (see SMPTE ST 291-1).

5 Serial Metadata Stream Structure and Timing on a Serial Data Link

Audio metadata on a serial data link shall be transmitted as an intermittent, asynchronous serial data stream running at 115,200 baud, with 1 Start bit, 8 data bits, 1 Stop bit and no Parity bit, at TIA/EIA-485-A (formerly RS-485) electrical levels. The repetition rate of the Metadata frames shall match the repetition rate of the associated Interlaced or Progressive Video frames for video frame rates less than or equal to 30 frames per second and the repetition rate of pairs of Progressive video frames for rates greater than 30 frames per second as shown in Figure 2. The transmission of the first and second metadata subframes should be timed to avoid the areas around the vertical interval switch points of the associated video signal to avoid corrupting the metadata stream when it is switched synchronously with the video signal. The duration and timing of these intervals around the switch points allows some tolerance in the relative timing of the video and metadata streams. This is detailed in Section 5.2.

5.1 Structure on an Asynchronous Serial Link

Each metadata burst or frame shall carry the audio metadata information related to a specific group of audio samples, which are temporally related to a specific video frame or frames. Metadata frames shall be subdivided into a first and second subframe, each of which shall be prefixed by a specific sync sequence. The number of bytes in each subframe depends on the number and configuration of the associated rate reduced audio programs.

Figure 1 shows the general structure of the serial metadata stream. The duration of the subframes depends on the number and format of the associated audio programs, but the repetition period of the complete metadata frame shall be equal to the frame period of the associated video signal operating at or below 30 frames per second, or to a pair of frame periods of a progressively scanned video signal operating at greater than 30 frames per second. The top part of Figure 1 shows the structure of the serial metadata stream when the duration of Metadata Subframe 1 is less than one video field (or one Progressive frame). The bottom part of Figure 1 shows the structure when the duration of Metadata Subframe 1 is greater than one video field (or one Progressive frame).

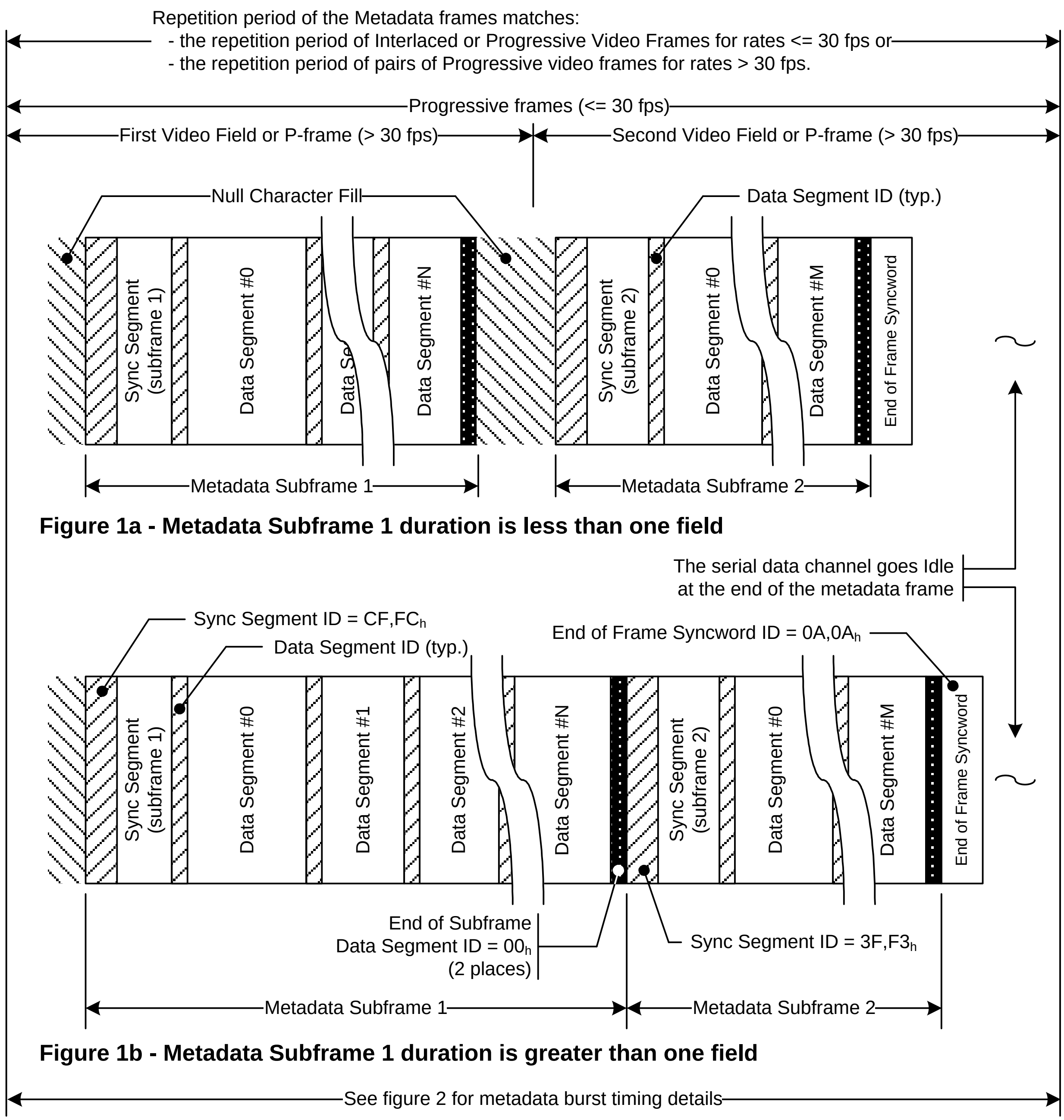

Figure 1 – Structure of a Metadata Frame in the Asynchronous Serial Stream

5.2 Metadata Frame Timing Requirements

The serial metadata stream is comprised of two subbframes, each of which is timed with respect to the associated video signal to avoid having a vertical interval switch point occur within either subframe.

Metadata subframe 1 shall begin no earlier than the beginning of the second line following the vertical interval switch point and no later than the end of the line following the vertical blanking interval of the associated video signal (see Figure 2).

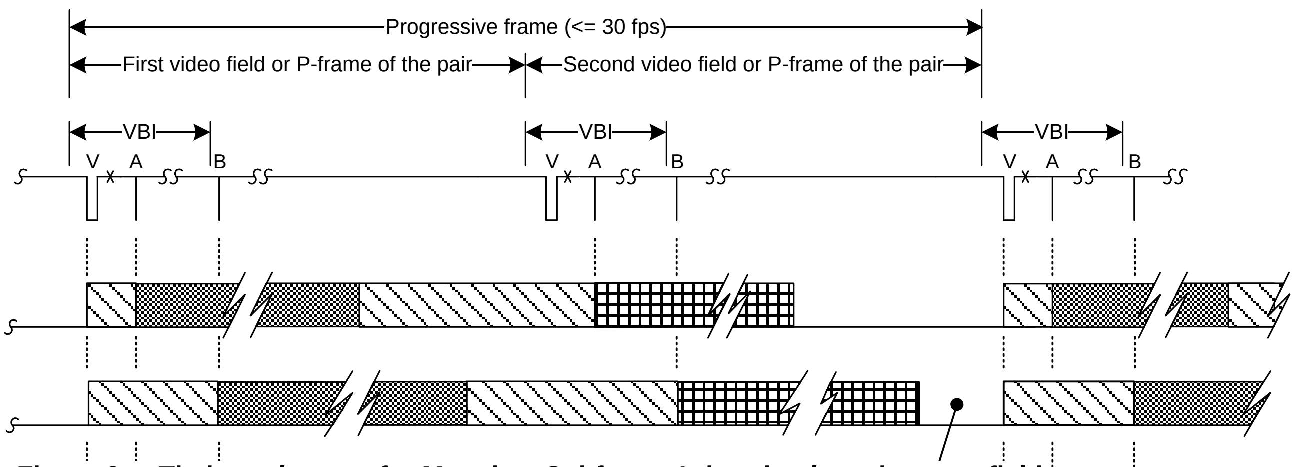

Figure 2a - Timing tolerance for Metadata Subframe 1 duration less than one field

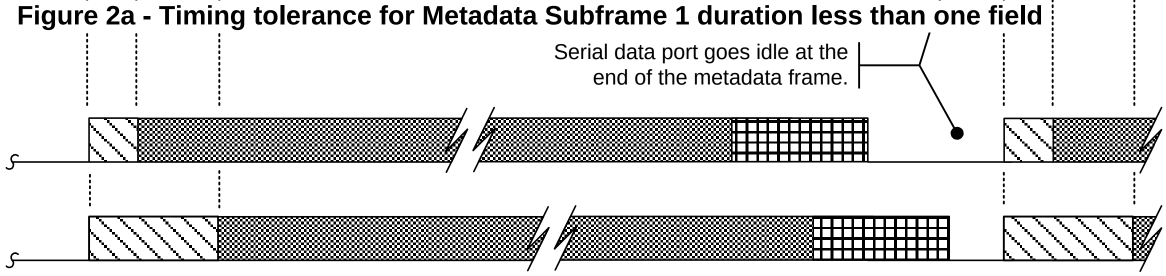

Figure 2b - Timing tolerance for Metadata Subframe 1 duration greater than one field

Legend

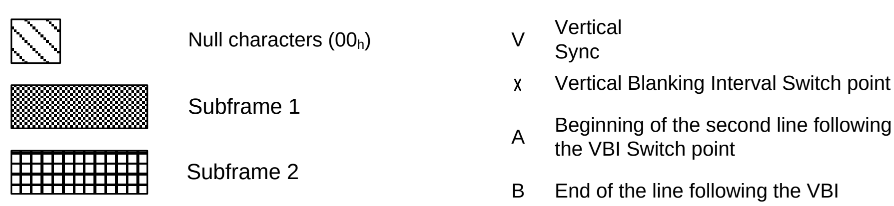

Figure 2 – Asynchronous Serial Metadata Stream Timing

If metadata subframe 1 ends before the start point of metadata subframe 2, then Null characters (00_h) shall be transmitted until metadata subframe 2 begins.

Metadata subframe 2 shall begin no earlier than the beginning of the second line following the vertical interval switch point and no later than the end of the line following the vertical blanking interval of the associated video signal, as shown in Figures 2a, unless the duration of subframe 1 is greater than that of video Field 1 or the first P frame, in which case subframe 2 shall begin when subframe 1 finishes, as shown in Figure 2b.

The asynchronous serial data channel shall go idle at the end of the second metadata subframe.

The signal alignment and Vertical Interval switching point relationships between interlaced and progressively scanned 1125, 750, 625 and 525-line systems are defined by SMPTE RP 168 (Definition of Vertical Interval Switching Point for Synchronous Video Switching).

For progressive video formats operating at rates greater than 30 frames per second, the first video frame of each pair of progressively scanned frames shall be the frame that is most closely aligned with the first character after the idle period in the serial metadata stream. All the timing details shall be referenced to the first frame of the progressive pair.

All text and diagrams in this section apply equally to video systems operating at integer and {integer/1.001} frame rates. The basic rate and timing of the recovered serial metadata stream shall match that of the associated video signal.

The serial metadata port should output a series of Null (00h) characters coincidentally with the Vertical sync period of the first field (or first progressive frame) and continuing until the beginning of metadata subframe 1. All receivers shall accept serial metadata streams with or without Null characters prepended to metadata subframe 1.

6 System Latency

To limit the VANC insertion latency to the duration of one metadata frame in interlaced and progressively scanned television systems operating at or below 30 frames per second, VANC transmitters shall insert the packet containing the received metadata into the next video frame after the end of the received Metadata Frame.

To limit the VANC insertion latency to the duration of one metadata frame in a progressively scanned system operating at frame rates greater than 30 frames per second, VANC transmitters shall insert the packet containing the received metadata into the next P frame after the end of the received Metadata frame. The VANC transmitter may also insert a duplicate packet into the following P frame.

VANC receivers shall output the received metadata as early as possible in the video frame in which it is received in order to minimize decoding latency, and to frame the serial data output with the timing specifications as described in Section 5.2.

In some cases, leading Null (00h) characters may have to be stripped from the recovered metadata to maintain the timing specified in Section 5.2.

7 Format of VANC Data Packets

Each data packet shall comply with the format defined in SMPTE ST 291-1 for a type 2 ANC packet. It consists of the ancillary data flag (ADF), the data ID (DID), the secondary data ID (SDID), the data count (DC), the user data words (UDW), and the checksum (CS).

The DID shall be set to the value 45h.

Note: The DC is variable and is defined in the associated mapping standards SMPTE ST 2020-2 and ST 2020-3

7.1 SDID and Audio Channel Pair Number

When multiple audio programs are carried or associated with a single video signal, the SDID value shall be used to identify the VANC data packet(s) carrying the audio metadata for each of the audio programs.

Table 1 specifies the relationship between an SDID value and the first audio channel pair of the audio channel pairs used to carry all of the audio channels required to make up an audio program. Audio channels of an audio program shall be carried on consecutive channel pairs because only the first channel pair of a program can to be signaled.

This identification system shall also be applied when several audio programs are carried on individual, rather than embedded AES3 signal paths. It shall also be applied to consecutively numbered pairs of analog signal paths associated with a single video program.

An SDID value of 01h shall be used when there is only one audio program associated with a video signal, and there is no intended association between VANC packets with an SDID value of 01h and a specific audio channel pair.

See Annex E for further information on how to assign SDID values

Table 1 – Association Between SDID Values and the First Audio Channel Pair of an Audio Program

Audio Channel Pair	SDID
No association	01h
Channel pair 1/2	02h
Channel pair 3/4	03h
Channel pair 5/6	04h
Channel pair 7/8	05h
Channel pair 9/10	06h
Channel pair 11/12	07h
Channel pair 13/14	08h
Channel pair 15/16	09h

8 Location of the Vertical Ancillary Data

The ANC data packet containing audio metadata shall be located in the active line portion of one line in the vertical ancillary space. Data may be located in any lines in the area from the second line after the line specified for switching, as defined in SMPTE RP 168, to the last line before active video, inclusive (see Annex B for further guidance).

When the ANC packets are carried in a high definition signal they shall be carried in the Y stream.

Notes:

1 Users are cautioned that although this standard specifies VANC as the location for the transport of information in an SD-SDI interface, there are legacy devices that will not pass this information if it is located on video lines where the V-bit (see SMPTE ST 125 and ITU-R BT.656) is set to logical one.

2 Receiving equipment ought to identify ANC packets for specific audio metadata streams on the basis of their ANC DID and SDID fields.

3 See Annex C for information about how to distinguish the two mapping methods.

9 Levels of Operation

To define the level of support for the two methods of mapping audio metadata defined in SMPTE ST 2020-2 and SMPTE ST 2020-3 by VANC transmitter and receiver equipment, one or more suffix letters shall be added to the basic Standard number ST 2020.

9.1 Support Levels

Level A – SMPTE ST 2020-2 Method A

Level B – SMPTE ST 2020-3 Method B

9.2 Examples of Nomenclature

A VANC transmitter or receiver device that inserts or extracts audio metadata VANC packets mapped according to Method A described in SMPTE ST 2020-2 would be said to conform to SMPTE ST 2020-A.

A VANC transmitter or receiver device that inserts or extracts audio metadata VANC packets mapped according to Method B described in SMPTE ST 2020-3 would be said to conform to SMPTE ST 2020-B.

A VANC transmitter or receiver device that can insert or extract audio metadata VANC packets mapped according to both Method A and Method B would be said to conform to SMPTE ST 2020-AB .

Annex A Bibliography (Informative)

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 274:2008) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 274M-2008). Documents with the same root number (e.g. 274) and publication year (e.g. 2008) are functionally identical.

SMPTE ST 125:2013, SDTV Component Video Signal Coding 4:4:4 and 4:2:2, for 13.5 MHz and 18 MHz Systems

SMPTE ST 274:2008, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE ST 293:2003 (Stabilized 2010), Television — 720 x 483 Active Line at 59.94-Hz Progressive Scan Production — Digital Representation

SMPTE ST 296:2012,1280 x 720 Progressive Image Sample Structure — Analog and Digital Representation and Analog Interface

SMPTE ST 2020-2:2014, Vertical Ancillary Data Mapping of Audio Metadata — Method A

SMPTE ST 2020-3:2014, Vertical Ancillary Data Mapping of Audio Metadata — Method B

SMPTE RP 291-2:2013, Ancillary Data Space Use — 4:2:2 SDTV and HDTV Component Systems and 4:2:2 2048 ×1080 Production Image Formats

SMPTE RDD 6:2008, Television — Description and Guide to the Use of the Dolby® E Audio Metadata Serial Bitstream

Recommendation ITU-R BT.656-5 (12/07), Interfaces for Digital Component Video Signals in 525-Line and 625-Line Television Systems Operating at the 4:2:2 Level of Recommendation ITU-R BT.601

Recommendation ITU-R BT.1358-1 (09/07), Studio Parameters of 625 and 525 Line Progressive Television Systems

Annex B Comments on the Preferred Location of VANC Packets (Informative)

Notwithstanding the possibility that the VANC packets may be placed on any line in the vertical blanking interval, it may be desirable to further constrain the location to a preferred line to improve the probability of successful passage through the production process, particularly when older devices are involved. System designers should be aware that some devices, particularly older recorders, will pass data on only a small subset of lines in the vertical ancillary space. This must be taken into account in the system design and choice of data location.

It has been observed that some devices do not test for existing ANC packets and consequently overwrite existing ANC packets. Typically, these devices insert data on the second line after the switching line.

Annex C Distinguishing Between Mapping Method A and Method B (Normative)

There are two methods for mapping Audio Metadata into VANC packets as described in SMPTE ST 2020-2 and in SMPTE ST 2020-3. Both methods use the same set of DID and SDID values.

The metadata mapping scheme described as Method A in SMPTE ST 2020-2 uses one descriptor byte at the beginning of each VANC packet and the value of this byte will not have the values 00_h or CF_h. A metadata receiver may recognize this format by inspecting the first audio metadata VANC packet (with the same DID/SDID values) of two consecutive video frames. If it is formatted according to Method A, it begins with a non-Null byte with a value other than CF_h.

The metadata mapping scheme described as Method B in SMPTE ST 2020-3 will always have a character with the value 00_h or CF_h as the first byte of the first VANC packet. If the first packet of the frame on two consecutive frames does not match these values it may be assumed to have been formatted according to Method A; otherwise, it may be assumed to have been formatted according to Method B.

Annex D Asynchronous Serial Data Communication (Informative)

Asynchronous start-stop describes an asynchronous transmission protocol in which a start signal is sent prior to each code symbol and a stop signal is sent after each code symbol. The start signal serves to prepare the receiving mechanism for the reception and registration of a symbol and the stop signal serves to bring the receiving mechanism to rest in preparation for the reception of the next symbol.

In the diagram, a start bit is sent, followed by eight data bits, no parity bit and one stop bit, for a 10-bit character frame. After the stop bit, the line may remain idle indefinitely, or another character may immediately be started.

Annex E Assigning SDID Values (Normative)

When multiple audio programs are carried by or associated with a single video signal, the SDID value shall be used to identify the VANC data packet(s) carrying the audio metadata for each of the audio programs.

The audio channels making up an audio program shall be assigned to consecutively numbered or arranged audio channel pairs. The VANC packets carrying the audio metadata for a specific audio program shall be identified by setting the SDID values of these packets to the value associated with the lowest numbered audio channel pair of the group of audio channel pairs carrying the complete audio program, as shown in Table 1.

For example, if there is a six channel program carried in the first three audio channel pairs, and a stereo program carried by the fourth audio channel pair, then the VANC packets carrying the metadata for the six channel program would have their SDID value set to 02h because the program starts in audio channel pair 1/2. The VANC packets carrying the metadata for the stereo program would have their SDID value set to 05h because the stereo program starts in audio channel pair 7/8.