SMPTE STANDARD

Audio to Video Synchronization Measurement — Fingerprint Transport

Table of Contents

Page

Approved
October 9, 2015

Foreword

The Society of Motion Picture and Television Engineers (SMPTE) 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 the Standards Operations Manual.

SMPTE ST 2064-2 was prepared by Technology Committee 24TB.

Intellectual Property

SMPTE draws attention to the fact that it is claimed that compliance with this Standard may involve the use of one or more patents or other intellectual property rights (collectively, "IPR"). The Society takes no position concerning the evidence, validity, or scope of this IPR.

Each holder of claimed IPR has assured the Society that it is willing to License all IPR it owns, and any third party IPR it has the right to sublicense, that is essential to the implementation of this Standard to those (Members and non-Members alike) desiring to implement this Standard under reasonable terms and conditions, demonstrably free of discrimination. Each holder of claimed IPR has filed a statement to such effect with SMPTE. Information may be obtained from the Director, Standards & Engineering at SMPTE Headquarters.

Attention is also drawn to the possibility that elements of this Standard may be subject to IPR other than those identified above. The Society shall not be responsible for identifying any or all such IPR.

Introduction

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

Errors in audio to video timing relationships have become commonplace in the industry. Such errors are known as “lip-sync” errors because of a typical viewer's sensitivity to inaccuracy in the synchronization of lip movement with the sound of speech. This situation arises because the audio and video portions of a program follow different processing paths, each with its own inherent timing factors. For various reasons, the audio and video delays through the program distribution chain may change dynamically. In order to correct such timing errors as they occur, a method is required to measure the change in audio to video synchronization without relying on out-of-service test signals.

The SMPTE 2064 suite of documents for Audio to Video Synchronization Measurement defines a process for extracting, packetizing, and transporting compact representations of audio and video essence, known as video and audio fingerprints, which change constantly as a function of changing picture and sound content. These fingerprints are extracted by non-intrusive analysis of the audio and video essence, and therefore can be used in a live on-air situation as well as in non-real-time systems. The fingerprints generated at a reference point where synchronization is known to be correct are intended to be transported through the program distribution chain, either bound to or separate from the associated essence. Where measurement and correction of synchronization errors is needed, a new set of fingerprints is extracted from the essence at a downstream location and compared with the fingerprints from the reference point. This comparison provides a dynamic measurement of the audio to video timing changes that have occurred, which may be used by other processes to display and/or correct any synchronization errors.

The SMPTE ST 2064-1 standard specifies the method for generating the audio and video fingerprints, and the SMPTE ST 2064-2 standard specifies the carriage of fingerprints using various transport methods.

1 Scope

This document defines the real-time transport of audio and video fingerprints used for audio to video timing measurement. These fingerprints and their containerization is defined in SMPTE ST 2064-1. Transport methods are defined for Serial Digital Interfaces (SDI), User Datagram Protocol (UDP) over Internet Protocol (IP), and MPEG-2 Transport 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 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.

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

SMPTE ST 292-1:2012, 1.5 Gb/s Signal/Data Serial Interface

SMPTE ST 2051:2014, Two-Frame Marker for 48(/1.001)-Hz, 50-Hz and 60(/1.001)-Hz Progressive Digital Video Signals on 1.5 Gb/s and 3 Gb/s Interfaces.

SMPTE ST 2064-1:2015, Audio to Video Synchronization Measurement — Fingerprint Generation

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

IETF STD 5, RFC 791-1981, DARPA Internet Program Protocol Specification

IETF STD 6, RFC 768-1980, User Datagram Protocol

IETF RFC 2460-1998, Internet Protocol, Version 6 (IPv6) Specification

ISO/IEC 13818-1:2007 (E), International Standard, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems.

4 Overview of Fingerprint Transport

Three methods of real-time transport are defined. The appropriate method should be selected for use as required in different parts of a system, depending upon system configuration and signal types. The audio and video fingerprint container is transport agnostic and may be carried unaltered through different concatenated transport methods.

5 Mapping into Vertical Ancillary Data Packets

5.1 Format of Vertical Ancillary (VANC) Data Packets

Each ancillary data (ANC) 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 UDW consists of the data payload, in this case, the Fingerprint Container.

The DID word shall be set to the value [41h]. The SDID word shall be set to the value of [0Bh].

DC is a count of the number of words in the UDW and shall be set to this word count.

5.1.1 UDW format

The fingerprint container shall be carried in the User Data Words of a SMPTE ST 291-1 ancillary data packet. The format of the fingerprint container is defined in the ”Container Structure” section of SMPTE ST 2064-1.

The ANC packet UDW shall be a sequence of 10-bit words up to the length of the fingerprint container per ”Container Structure” section of SMPTE ST 2064-1. The fingerprint container information is transmitted in bits b7 through b0 of the 10-bit data word. Bit b8 is even parity for bits b7 through b0 of the 10-bit data word, and bit b9 equals the complement of bit b8.

5.1.2 Location of the Vertical Ancillary Data

The fingerprint container shall be located in the VANC space of every frame. For interlaced video systems, it shall be located in the first field. For progressive segmented frame systems, it shall be located in the first frame segment as defined in SMPTE ST 2051. In all cases, the fingerprint container shall be located on one line in the range from the second line after the line specified for switching, as defined in SMPTE RP 168, to the last line before active video, inclusive.

Only one fingerprint container per interlaced, progressive or segmented frame shall be permitted at any point in the video chain.

In order to maximize the time available for processing the active video with which the fingerprint is associated, it is recommended that the fingerprint container should be located early in the VANC period (see Annex A for further guidance).

The fingerprint container shall be carried in the VANC space of the frame following the frame from which the fingerprint container was generated.

When the ANC packets defined in this standard are carried in a SMPTE ST 292-1 Interface, they shall be carried in the Luma (Y’) data channel.

Receiving equipment shall identify the fingerprint container on the basis of its DID and SDID fields. This is fundamental to the SMPTE ST 291-1 data concept, and all equipment intended to locate, identify, and/or process a fingerprint container shall comply with this requirement. The requirement is not intended to apply to equipment simply passing, switching, or recording the video signal, where the expectation is that the fingerprint containers are left unchanged wherever they are placed in the VANC space.

Notes:

1. Nothing in this standard prohibits the carriage of multiple VANC packets containing different categories of data on the same line in the VANC space.

2. For high definition signals, the packet location is constrained to the Y stream in order to ensure passage of this data through deployed devices that do not process VANC data in the C stream.

3. Designers should be aware that although this standard specifies the VANC data space 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 Recommendation ITU-R BT.656) is set to logical one.

6 Mapping into UDP / IP Packets

The fingerprint container may be transmitted using the User Datagram Protocol (UDP) over IP networks using either IPv4 or IPv6 transport.

The fingerprint container is transported according to RFC 768 in the user data space, beginning at octet 9 in the UDP message. Length and checksum in the UDP message are calculated accordingly and populated prior to transmission. Each fingerprint UDP datagram shall contain only one fingerprint container. Additional data shall not be included in the datagram payload.

6.1 UDP Transport Considerations

Each UDP/IP datagram sent by the transport layer shall completely adhere to the structure specified in the UDP Specification (RFC 768) and either the IPv4 (RFC 791) or IPv6 (RFC 2460), depending on which protocol is used.

The UDP is a low-overhead, unreliable protocol. This section discusses reliability issues inherent in the protocol that implementers and users should be aware of.

UDP does not provide any mechanism for the detection or correction of lost datagrams. Datagrams can be lost in transit due to congestion, corruption or any other intermittent network problem.

The UDP/IP datagrams can become corrupted in transit due to software, hardware or network errors. The checksums in UDP and IP do not provide a guarantee of corruption detection, and this transport mapping does not provide for a message acknowledgement or retransmission mechanism. The checksum in the fingerprint container may be used to further identify payloads with errors.

6.2 UDP Transport Requirements

The IP Transport used by the UDP does not guarantee that the sequence of datagram delivery will match the order in which the datagrams are sent. The sequence counter in the fingerprint container should be used for sequence ordering management.

7 Mapping Into MPEG-2 Transport Stream

7.1 Constraints on Program-Specific Information (PSI)

Fingerprint data is associated with a particular service and therefore is required to be an element of that particular program. Therefore, for each service with fingerprint data, the Program Map Table (PMT) instance shall include a program element of stream type 06h, Packetized Elementary Stream (PES) packets containing private data. An associated RID of “LIPS”, (4Ch 49h 50h 53h) shall be used.

7.2 PES Encapsulation of Fingerprint Container

For MPEG-2 Transport, the Fingerprint Container (FC) shall be encapsulated in a PES packet, using private_stream_2 as specified in Table 1.

7.2.1 Fingerprint Container Transport Syntax

Table 1 – Bit Stream Syntax for the Fingerprint Container Transport PES packet

Syntax	No. Bits	Format
FC_transport() {		
packet_start_code_prefix	24	bslbf
stream_id	8	BFh
PES_packet_length	16	bslbf
fingerprint_container()	var	
CRC_32	32	rpchof
}		

7.2.2 Fingerprint Data Transport Semantics

packet_start_code_prefix – The packet_start_code_prefix is a 24-bit code. Together with the stream_id that follows it constitutes a packet start code that identifies the beginning of a packet. The packet_start_code_prefix is the bit string '0000 0000 0000 0000 0000 0001' (0x000001).

stream_id – This 8 bit field shall be set to the value BFh, indicating private_stream_2.

PES_packet_length – This 16-bit field specifies the number of bytes immediately following in the PES packet following the last byte of this field (including the CRC_32 field).

fingerprint_container – The fingerprint container defined in SMPTE ST 2064-1.

CRC_32 – This is a 32-bit field that contains the CRC value that ensures a zero output from the registers in the decoder defined in Annex A of ISO/IEC-13818-1 after processing the entire FC_transport packet.

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

1. Notwithstanding the possibility that the VANC packets may be placed on any line in the vertical ancillary space, it is desirable to further constrain the location to a preferred line to improve the probability of successful passage through the production process. System designers should be aware that many devices in the studio processing chain will pass data on only a small subset of lines in the vertical ancillary space and some devices (e.g., production switchers, digital video effects units, and video servers) may not pass VANC packets at all. For successful implementation of fingerprint, system designers need to understand the characteristics of all relevant devices and it is essential that the VANC location chosen for inserting fingerprint containers takes into account downstream equipment in the video chain.

2. Some recording devices record only one line of data from the vertical ancillary space while others may record up to 11 lines. These recorders may select the lines that they record, or have other constraints. However, it is a common practice to set them to record three consecutive lines starting at the second line after the switching line. This typically constrains the available lines to the second line, the third line, and the fourth line after the switching line.

3. In Section 5.1.2 of this standard, it is recommended that the fingerprint containers be located early in the VANC period. Regardless of the line on which the fingerprint container is placed, SMPTE ST 291-1 compliant ANC data receivers will detect the packet by its DID and SDID.

4. Fingerprint containers are preferably embedded in the video signal early in the production process, ideally in the camera or image-generating device. It has been observed that some devices handling video signals do not test for existing ANC packets and consequently overwrite them. If such devices are present downstream of a fingerprint originating or processing device, it is possible that ANC data, including fingerprint data, could be deleted.

5. These considerations are similar to those for other data services carried in VANC.

Annex B Bibliography (Informative)

SMPTE ST 125:2013, SDTV Component Video Signal Coding 4:4:4 and 4:2:2 for 13 MHz and 18 MHz Systems

SMPTE ST 2016-3:2009, Vertical Ancillary Data Mapping of Active Format Description and Bar Data

Recommendation ITU-R BT.656-5:2007, 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