

# SMPTE STANDARD



## Professional Media Over Managed IP Networks: PCM Digital Audio

---

Page 1 of 9 pages

### Table of Contents

<b>1</b>	<b>Scope</b>	<b>3</b>
<b>2</b>	<b>Conformance Notation</b>	<b>3</b>
<b>3</b>	<b>Normative References</b>	<b>3</b>
<b>4</b>	<b>Terms and Definitions</b>	<b>4</b>
<b>5</b>	<b>Textual Conventions</b>	<b>4</b>
5.1	SDP Parameters and Values	4
<b>6</b>	<b>Uncompressed PCM Digital Audio RTP Essence Format</b>	<b>4</b>
6.1	Media Clock, RTP Clock, and RTP Timestamps	4
6.2	AES67 Constraints	5
<b>7</b>	<b>Conformance Levels</b>	<b>7</b>
	<b>Bibliography (Informative)</b>	<b>9</b>

## 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. This SMPTE Engineering Document 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 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 Engineering Document.

The capability and capacity of IP networking equipment has improved steadily, enabling the use of IP switching and routing technology to transport and switch video, audio, and metadata essence within television facilities. Existing standards such as SMPTE ST 2022-6 have gained use in this application, but there is a desire in the industry to switch different essence elements separately.

This family of SMPTE engineering documents builds on the work of VSF TR03 and TR04, and of AES67, documenting a system for transporting various essence streams over IP networks, capturing the timing relationships between those streams. The system is intended to be extensible to a variety of essence types.

SMPTE ST 2110-10 covers the system as a whole, the timing model, and common requirements across all essence types.

SMPTE ST 2110-30 (this standard) documents and constrains of use of IP-encapsulated PCM audio in a manner compatible to AES67 in such systems.

## 1 Scope

This standard specifies the real-time, RTP-based transport of PCM digital audio streams over IP networks by reference to AES67. An SDP-based signaling method is defined for metadata necessary to receive and interpret the stream.

Non-PCM digital audio signals including compressed audio signals are outside the scope of this standard.

## 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; then 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 engineering document. At the time of publication, the editions indicated were valid. All standards are subject

to revision, and parties to agreements based on this engineering document are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

AES67-2015 AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability

Internet Engineering Task Force (IETF) RFC 3190 RTP Payload Format for 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio [online, viewed 2017-08-10] Available at <https://www.ietf.org/rfc/rfc3190.txt>

Internet Engineering Task Force (IETF) RFC 4566 SDP: Session Description Protocol [online, viewed 2017-08-10] Available at <https://www.ietf.org/rfc/rfc4566.txt>

SMPTE ST 2110-10:2017 Professional Media over Managed IP Networks – System Timing and Definitions

SMPTE ST 2036-2 Ultra High Definition Television – Audio Characteristics and Audio Channel Mapping for Program Production

## **4 Terms and Definitions**

The terms and definitions of SMPTE ST 2110-10 apply.

## **5 Textual Conventions**

### **5.1 SDP Parameters and Values**

The names and values of SDP Media Type parameters within the text of this document are formatted using a monospaced font (such as Courier) except when they appear in section headings.

## **6 Uncompressed PCM Digital Audio RTP Essence Format**

### **6.1 Media Clock, RTP Clock, and RTP Timestamps**

The Media Clock and RTP Clock shall comply with SMPTE ST 2110-10 sections 7.3 (titled “Media Clocks”) and 7.4 (titled “RTP Clocks”) respectively. The rate of the Media Clock and RTP Clock shall be the same as the digital audio sampling rate.

The RTP timestamp shall comply with SMPTE ST 2110-10 section 7.5 (titled “RTP Timestamps”).

All senders and receivers of PCM digital audio conforming to this standard shall support a digital audio sampling rate of 48 kHz, and should support digital audio sampling rates of 44.1 kHz and/or 96 kHz. Other sampling rates are out of scope.

All senders and receivers conforming to Levels above level A as defined in Table 2 shall support the sampling rates, packet times, and numbers of channels stated for their level in Table 2.

Note 1: The media clock and the network timebase share the SMPTE ST 2059-1 epoch, with an offset of zero between the Media Clock and the RTP Clock, as specified in SMPTE ST 2110-10.

Note 2: To maintain compatibility with other RTP implementations such as AES67, implementers ought to be mindful of the offset provisions in those RTP standards, and the possibility that the RTP Clock could be offset from the Media Clock.

Note 3: All of the IETF RFC documents which are listed in section 3 (Normative References) are Standards-Track Documents within the IETF; they are however at varying phases of standardization within the IETF process, not all of them have reached the final phase of "Internet Standard". The IETF standardization phases are described in IETF RFC 6410.

## 6.2 AES67 Constraints

### 6.2.1 General Provisions

Digital audio streams shall conform to AES67, including the Session Description Protocol (SDP) as described in IETF RFC 4566, subject to the constraints in this document.

Notwithstanding the provisions regarding SIP in AES67, digital audio receivers need not support Session Initiation Protocol (SIP) nor any other specific method mentioned in AES67 for connection management.

Audio senders and receivers shall comply with the provisions of AES67 section 7.1 titled "Payload format and sampling rate".

Audio senders and receivers shall observe the timing provisions of AES67 section 7.5 titled "Sender timing and receiver buffering".

The Standard UDP Datagram Size Limit as defined in SMPTE ST 2110-10 shall be used.

### 6.2.2 Channel Order Convention

If channel order is signaled in the SDP, the syntax of IETF RFC 3190 for the parameter `channel-order` shall be used. The <convention> of the `channel-order` should be `SMPTE2110`.

In the `SMPTE2110` channel-order convention, the <order> shall be a listing of Channel Grouping Symbols contained within parenthesis and separated by commas. The symbol of a Channel Grouping, and the quantity and order of the audio channels within a Channel Grouping shall be as defined in Table 1.

For example:

```
a=fmtp:101 channel-order=SMPTE2110.(51,ST)
```

This example defines the first six channels as a 5.1 surround group, and the next two channels as a stereo group.

Another example:

```
a=fmtp:101 channel-order=SMPTE2110.(M,M,M,M,ST,U02)
```

This example defines the first four channels as mono channels, and the next two channels as a stereo group, and the last two channels are an undefined group of two channels.

Any audio channels that do not match the quantity or order of the defined Soundfield Groups in Table 1 shall be identified and grouped as Undefined. Undefined is a special designation which is used to signal groups of channels whose mix is unknown or un-representable, or otherwise non-conformant to the other defined groupings and orderings in the convention. The Undefined channel grouping symbol shall be a “U” followed by a two-digit integer number with a leading zero if required, and the number shall define the number of channels carried in the undefined channel group. For example, “U05” refers to a five channel Undefined channel group.

If the channel-order parameter is not present, the audio channels shall be treated as Undefined. If there is a mismatch between the quantity of channels declared in the channel-order parameter and the quantity of channels actually in the stream, the audio channels that are not declared in the channel-order parameter shall be treated as Undefined.

The Channel Grouping Symbols shall be as defined in Table 1. The Audio Channel names and symbols used in Table 1 are as defined in SMPTE ST 2067-8, except for the 22.2 Channel Grouping and the Undefined Channel Grouping which are defined in this standard. The 22.2 Channel Grouping has channel names and order as per SMPTE ST 2036-2, Table 1. The Undefined channel grouping in Table 1 is not a Soundfield Group but is used to indicate the absence of a specific group for the referenced channels.

**Table 1 -- Channel Order Convention Grouping Symbols**

<b>Channel Grouping Symbol</b>	<b>Quantity of Audio Channels in group</b>	<b>Description of group</b>	<b>Order of Audio Channels in group</b>
M	1	Mono	Mono
DM	2	Dual Mono	M1, M2
ST	2	Standard Stereo	Left, Right
LtRt	2	Matrix Stereo	Left Total, Right Total
5.1	6	5.1 Surround	L, R, C, LFE, Ls, Rs
7.1	8	7.1 Surround	L, R, C, LFE, Lss, Rss, Lrs, Rrs
22.2	24	22.2 Surround	Order shall be per SMPTE ST 2036-2, Table 1
SGRP	4	One SDI audio group	1, 2, 3, 4
U01...U64	As indicated by Grouping Symbol (Unn where nn is the number of channels in the group)	Undefined	None specified, the order of channels in this group is Undefined.

Note 1: The Channel Order described above was developed such that phase-coherent multichannel audio groups (or simply mono channels) can have their channels clearly defined within SDP. It does not attempt to provide any “higher level” description of the purpose of the audio channels (such as “secondary language”). It only addresses fixed multichannel audio groups commonly used at the time of publication of this document, and does not attempt to deal with object-based audio.

Note 2: The 5.1 Surround and 7.1 Surround channel orders defined in Table 1 are based on typical ordering of those channels found in SMPTE ST 2035, EBU R 123, and SMPTE ST 429-2.

## 7 Conformance Levels

Receivers conforming to this standard shall implement at a minimum the Level A requirements in Table 2. Receivers conforming to this standard should specify the Table 2 Conformance Levels which are supported when indicating conformance to this standard.

**Table 2- Conformance Levels**

<b>Level</b>	<b>Supported by the Receiver</b>
A	Reception of 48 kHz streams with from 1 to 8 channels at packet times of 1 ms.
AX	Reception of 48 kHz streams with from 1 to 8 audio channels at packet times of 1 ms. Reception of 96 kHz streams with from 1 to 4 channels at packet times of 1ms.
B	Reception of 48 kHz streams with from 1 to 8 channels at packet times of 1 ms or 1 to 8 channels at packet times of 125 $\mu$ s.
BX	Reception of 48 kHz streams with from 1 to 8 channels at packet times of 1 ms or 1 to 8 channels at packet times of 125 $\mu$ s. Reception of 96 kHz streams with from 1 to 4 channels at packet times of 1ms or 1 to 8 channels at packet times of 125 $\mu$ s.
C	Reception of 48 kHz streams with from 1 to 8 channels at packet times of 1 ms or 1 to 64 channels at packet times of 125 $\mu$ s.
CX	Reception of 48 kHz streams with from 1 to 8 channels at packet times of 1 ms or 1 to 64 channels at packet times of 125 $\mu$ s. Reception of 96 kHz streams with from 1 to 4 channels at packet times of 1ms or 1 to 32 channels at packet times of 125 $\mu$ s.

Note: SDI allows the carriage of at least 16 embedded audio channels. Senders wishing to remain in compliance with level A, can do so by sending groups of channels organized into multiple AES67 streams, ideally grouped into logical channel groupings such as those listed in section 6.2.2.



## **Bibliography (Informative)**

EBU -- Recommendation R 123 EBU Audio Track Allocation for File Exchange

IETF RFC 6410 Reducing the Standards Track to Two Maturity Levels [online, viewed 2017-08-10] Available at <https://www.ietf.org/rfc/rfc6410.txt>

SMPTE ST 428-12:2013 D-Cinema Distribution Master Common Audio Channels and Soundfield Groups

SMPTE ST 429-2:2013 D-Cinema Packaging – DCP Operational Constraints

SMPTE ST 2035:2009 Audio Channel Assignments for Digital Television Recorders

SMPTE ST 2059-1:2015 Generation and Alignment of Interface Signals to the SMPTE Epoch

SMPTE ST 2067-8:2013 Interoperable Master Format — Common Audio Labels

Video Services Forum (VSF) TR03 Transport of Uncompressed Elementary Stream Media over IP [online, viewed 2017-08-10] Available at [http://www.videoservicesforum.org/download/technical\\_recommendations/VSF\\_TR-03\\_2015-11-12.pdf](http://www.videoservicesforum.org/download/technical_recommendations/VSF_TR-03_2015-11-12.pdf)