ICS 35.040

CCS L 71

团体标准

T/AI 109.7—202X

|  |
| --- |
|  |

信息技术 智能媒体编码 第7部分

音频封装与传输

Information technology – Intelligent media coding –

Part 7: Audio encapsulation and transport

点击此处添加与国际标准一致性程度的标识

|  |
| --- |
| 征求意见稿 |
| （在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上） |

XXXX - XX - XX实施

中关村视听产业技术创新联盟 发布

XXXX - XX - XX发布

目  次

[目次 I](#_Toc176373413)

[前言 II](#_Toc176373414)

[引言 III](#_Toc176373415)

[1 范围 1](#_Toc176373417)

[2 规范性引用文件 1](#_Toc176373418)

[3 术语和定义 1](#_Toc176373419)

[4 缩略语 3](#_Toc176373420)

[5 音频文件格式 4](#_Toc176373421)

[6 CMAF轨道和媒体配置 13](#_Toc176373422)

[7 DASH传输技术要求 15](#_Toc176373423)

[8 传输流和节目流技术要求 18](#_Toc176373424)

[9 SMT传输技术要求 29](#_Toc176373425)

[10 RTP传输技术要求 37](#_Toc176373426)

[附录A （规范性） MIME类型的'codecs'参数 42](#_Toc176373427)

 前  言

本文件按照GB/T 1.1—2020《标准工作导则 第1部分：标准文件的结构和起草规则》的规定起草。

本文件是T/AI 109《信息技术 智能媒体编码》的第7部分。T/AI 109已经发布了如下部分：

——第2部分: 视频；

——第3部分: 沉浸式音频；

——第4部分: 符合性测试；

——第6部分：智能媒体格式。

本文件由数字音视频编解码技术标准工作组提出。

本文件由中关村视听产业技术创新联盟归口。

本部分起草单位：清华大学、北京工业大学、鹏城实验室、上海交通大学、华为技术有限公司、北京全景声信息科技有限公司、咪咕文化科技有限公司、杭州当虹科技股份有限公司、中兴通讯股份有限公司、北京大学、中国电子技术标准化研究院、广东博华超高清创新中心有限公司、腾讯科技（深圳）有限公司、上海数字电视国家工程研究中心有限公司、上海海思技术有限公司、清华大学天津电子信息研究院、全景声(北京)智能科技有限公司。

本部分主要起草人：窦维蓓、牟伦田、郑建铧、徐异凌、李婧欣、刘帅、许舒敏、李琳、王雪辉、韩建、高原、王一帆、杨开发、黄成、陈智敏、张伟民、王尧、胡颖、邹志铭、马思伟、潘兴德、夏丙寅、王喆、袁乐、孙彦龙、徐嵩、单华琦、李康敬、吴迪、杨川、赵海英、朱博成。

引  言

T/AI 109.7规定了音频封装与传输规范，旨在为异构网络中的音频数据提供封装和传输服务。

T/AI 109拟由以下部分组成：

——第1部分：系统。目的在于确立沉浸媒体虚拟现实数据的系统信令。

——第2部分：视频。目的在于确立智能媒体高效视频压缩方法的解码过程。

——第3部分：沉浸式音频。目的在于确立适用沉浸式音频高效压缩方法的解码过程。

——第4部分：符合性测试。目的在于确立测试验证编码位流和解码器是否满足T/AI 109所规定的要求。

——第5部分：参考软件。目的在于确立对智能媒体高效视频压缩方法和沉浸式音频高效压缩方法的参考实现。

——第6部分： 智能媒体格式。目的在于确立异构网络中智能媒体编码数据的存储格式和传输信令。

——第7部分；音频封装与传输。目的在于确立异构网络中的沉浸式音频数据提供封装和传输。

——第10部分；实时语音。目的在于确立面向下一代实时语音通信的解码过程。

本文件的发布机构提请注意，声明符合本文件时，可能涉及如下XX项与数字音视频编解码技术相关的专利的使用。专利申请号及名称如下：

|  |  |  |
| --- | --- | --- |
| 序号 | 专利申请号 | 专利名称 |
| 1 |  |  |

本文件的发布机构对上述专利的真实性、有效性和范围无任何立场。

上述专利持有人已向本文件的发布机构保证，愿意同任何申请人在合理且无歧视的条款和条件下，就专利授权许可进行谈判。上述专利持有人的声明已在本文件的发布机构备案，相关信息可以通过以下联系方式获得：

联 系 人：黄铁军（数字音视频编解码技术标准工作组秘书长）

通讯地址：北京大学理科2号楼2641室

邮政编码：100871

电子邮件：tjhuang@pku.edu.cn

电话：+8610-62756172

传真：+8610-62751638

网址：http://www.avs.org.cn

请注意除上述专利外，本文件的某些内容仍可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

信息技术 智能媒体编码 第7部分：音频封装与传输

1. 范围

本文件规定了智能媒体编码数据的AVS3音频文件格式，高效多媒体编码数据的AVS2音频文件格式、AVS3/AVS2音频的CMAF轨道和媒体配置、DASH传输技术要求、传输流和节目流技术要求、SMT传输技术要求以及RTP传输技术要求。

本文件适用于智能媒体编码系统中的音视频直播、音视频点播、网络流媒体等应用。

1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 33475.3 信息技术 高效多媒体编码 第3部分：音频

T/AI 109.3-2023信息技术 智能媒体编码 第3部分：沉浸式音频

T/AI 114-2021 信息技术 高效多媒体编码 第6部分：智能媒体传输

ISO/IEC 13818-1信息技术 运动图像及其伴音信息的通用编码 第1部分:系统 (Information technology -- Generic coding of moving pictures and associated audio information -- Part 1: Systems)

ISO/IEC 14496-12 信息技术 音视频对象的编码 第12部分：ISO基本媒体文件格式 (Information technology – Coding of audio-visual objects – Part 12: ISO base media file format)

ISO/IEC 23000-19 信息技术 多媒体应用格式 第19部分：片段媒体通用媒体应用格式 (Information technology — Multimedia application format (MPEG-A) — Part 19: Common media application format (CMAF) for segmented media)

ISO/IEC 23009-1信息技术 基于HTTP的动态自适应流媒体 第1部分：媒体呈现描述和片段格式(Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats)

IETF RFC 1738 统一资源定位符 (Uniform Resource Locators (URL))

IETF RFC 3550 实时传输协议 (A Transport Protocol for Real-Time Applications)

IETF RFC 6381 "Bucket"媒体类型'Codecs'和'Profiles'参数 (The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types)

IETF RFC 8866 会话描述协议 (SDP: Session Description Protocol)

W3C XML 可扩展置标语言 (Extensible Markup Language (XML))

W3C XML Schema Part 1 可扩展置标语言模式定义语言 第1部分：结构 (XML Schema Definition Language (XSD) Part 1: Structures)

W3C XML Schema Part 2 可扩展置标语言模式定义语言 第2部分：数据类型 (XML Schema Definition Language (XSD) Part 2: Datatypes)

1. 术语和定义

下列术语和定义适用于本文件。

* 1.

位流 bitstream

用作数据编码表示的有一定次序的一组二进制数据流。

AVS3音频编码位流 AVS3 audio bitstream

符合T/AI 109.3规范的编码音频信号所形成的二进制数据流。

注：T/AI 109.3可简称为AVS3音频。

AVS2音频编码位流 AVS2 audio bitstream

符合GB/T 33475.3规范的编码音频信号所形成的二进制数据流。

注：GB/T 33475.3可简称为AVS2音频。

采样频率 sampling frequency

每秒从连续信号中提取离散信号的采样个数。

注：单位为赫兹（Hz）。

声道 channel

声音在录制或播放时在不同空间位置采集或回放的相互独立的音频信号。

保留 reserved

在文件格式或传输信令中的暂时未被使用的字段，可能在将来的标准扩展中被用到。

初始化片段 initialization segment

包含有媒体流解码所必需元数据的片段。

表示 representation

封装有一个或多个具有描述性元数据的媒体成分（编码的音频、视频等）的结构化数据集合。

轨道 track

文件中一系列相关样本的集合。

媒体呈现描述 media presentation description

用于提供流媒体服务的规范化描述媒体呈现的文件。

媒体片段 media segment

符合一定的媒体格式、可播放的片段。播放时可能需要与其前面的0个或多个片段以及初始化片段配合。

* 1.

媒体资源 asset

任何与唯一标识符联系的用作构建一个多媒体演示的多媒体数据实体。

片段 segment

媒体呈现描述中的HTTP统一资源定位符引用的媒体单元。

* 1.

样本 sample

在非提示轨道中，一个样本是一个单独的音频帧，时间连续的一个音频帧序列，或者时间连续的一段压缩音频。在提示轨道中，一个样本定义了一个或多个流式分组的构成。一个轨道中任何两个样本不能具有相同的时间戳。

智能媒体传输协议 smart media transport protocol

用于在IP网络上传输有效载荷的应用层传送协议。

切换集 switching set

同一媒体内容的多个可切换的编码版本的集合。

1. 缩略语

下列缩略语适用于本文件。

|  |  |  |
| --- | --- | --- |
| AVSAVS2AVS3CMAFDASHDTSHOAHTTPISO BMFFMIMEMPMPDMTUPESPMTPSPSIRTPSAPSDPSMTSTDTST-STDURIURNUTCXML | 音视频编解码标准第二代音视频编解码标准第三代音视频编解码标准通用媒体格式基于HTTP的动态自适应流媒体解码时间戳高阶立体声场信号超文本传输协议ISO基本媒体文件格式多用途互联网邮件扩展类型媒体呈现媒体呈现描述最大传输单元分组化基本流节目映射表节目流节目特定信息实时传输协议流访问点会话描述协议智能媒体传输系统目标解码器传输流传输系统目标解码器统一资源标识符统一资源名称协调世界时可扩展置标语言 | (audio and video coding standard)(the second generation AVS standard)(the third generation AVS standard)(common media application format)(dynamic adaptive streaming over HTTP)(decoding time-stamp)(higher order Ambisonics)(hypertext transfer protocol)(ISO base media file format)(multipurpose internet mail extensions)(media presentation)(media presentation description)(maximum transmission unit)(packetized elementary stream)(program map table)(program stream)(program-specific information)(real-time transport protocol)(stream access point)(session description protocol)(smart media transport)(system target decoder)(transport stream)(transport system target decoder)(uniform resource identifier)(uniform resource name)(coordinated universal time)(extensible mark-up language) |

1. 音频文件格式
	1. AVS3音频文件格式
		1. AVS3音频基本流定义
			1. 概述

AVS3音频基本流即AVS3音频编码位流，应符合T/AI 109.3规范。

* + - 1. AVS3音频编码特性

AVS3音频编码系统支持声道信号编码、对象信号编码、HOA信号编码、元数据编码。

AVS3音频编码器由多种编码工具构成，包括：通用全码率音频编码工具、通用高码率音频编码工具、无损音频编码工具和元数据编码工具。

* + - 1. 基本流结构

根据AVS3音频的编码特性，AVS3音频的基本流结构包括：通用全码率音频编码位流、通用高码率音频编码位流和无损音频编码位流。

* + - 1. 基本流格式

AVS3音频基本流可以分为两种：存储格式AASF和传输格式AATF，应符合T/AI 109.3的附录A。

* + 1. AVS3音频配置信息
			1. AVS3音频通用全码率音频编码特有配置
				1. 定义

本节定义适用于AVS3音频通用全码率音频编码内容的特有配置。

* + - * 1. 语法

class Avs3AudioGASpecificConfig {

unsigned int(4) sampling\_frequency\_index;

unsigned int(3) nn\_type;

unsigned int(1) reserved;

unsigned int(4) content\_type;

if (content\_type==0) {

unsigned int(7) channel\_number\_index;

unsigned int(1) reserved;

} else if(content\_type==1) {

unsigned int(7) object\_channel\_number;

unsigned int(1) reserved;

} else if(content\_type==2) {

unsigned int(7) channel\_number\_index;

unsigned int(1) reserved;

unsigned int(7) object\_channel\_number;

unsigned int(1) reserved;

 } else if(content\_type==3) {

unsigned int(4) hoa\_order;

}

unsigned int(16) total\_bitrate;

unsigned int(2) resolution;

if (content\_type==3) {

unsigned int(2) reserved;

} else {

unsigned int(6) reserved;

}

}

* + - * 1. 语义

sampling\_frequency\_index：应符合T/AI 109.3的附录A。

nn\_type：应符合T/AI 109.3的附录A。

content\_type：表示音频内容类型，见表1。

表1 content\_type配置表

|  |  |  |
| --- | --- | --- |
| content\_type值 | 音频内容类型 | 映射关系 |
| 0 | 声道信号 | coding\_profile值为0时 |
| 1 | 对象信号 | coding\_profile值为1且soundBedType值为0时 |
| 2 | 声道信号和对象信号混合 | coding\_profile值为1且soundBedType值为1时 |
| 3 | HOA信号 | coding\_profile值为2时 |
| 4-15 | 保留 |

注：coding\_profile和soundBedType应符合T/AI 109.3的附录A。

channel\_number\_index：应符合T/AI 109.3的附录A。

object\_channel\_number：应符合T/AI 109.3的附录A。

hoa\_order：表示HOA信号阶数，应符合T/AI 109.3的附录A中order的计算方法。

total\_bitrate：表示音频总比特率，单位kbps，根据content\_type的值计算方式不同，见表2。

表2 total\_bitrate配置表

|  |  |
| --- | --- |
| content\_type值 | total\_bitrate计算方式 |
| 0 | 声道信号的bitrate\_index对应的比特率 |
| 1 | 对象信号的bitrate\_index对应的比特率×（object\_channel\_number + 1） |
| 2 | 声道信号的bitrate\_index对应的比特率＋对象信号的bitrate\_index对应的比特率×（object\_channel\_number + 1） |
| 3 | HOA信号的bitrate\_index对应的比特率 |
| 4-15 | 保留 |

注：bitrate\_index应符合T/AI 109.3的附录A。

resolution：应符合T/AI 109.3的附录A

* + - 1. AVS3音频通用高码率音频编码特有配置
				1. 定义

本节定义适用于AVS3音频通用高码率音频编码内容的特有配置。

* + - * 1. 语法

class Avs3AudioGHSpecificConfig {

unsigned int(4) sampling\_frequency\_index;

unsigned int(1) anc\_data\_index;

unsigned int(3) coding\_profile;

unsigned int(1) bitstream\_type;

 unsigned int(7) channel\_number\_index;

 unsigned int(4) bitrate\_index;

 unsigned int(16) raw\_frame\_length;

unsigned int(2) resolution;

 unsigned int(16) addition\_info\_length;

if (addition\_info\_length > 0) {

bit(8\*addition\_info\_length) addition\_info;

}

unsigned int(6) reserved;

}

* + - * 1. 语义

sampling\_frequency\_index：应符合T/AI 109.3的附录A。

anc\_data\_index：应符合T/AI 109.3的附录A。

coding\_profile：应符合T/AI 109.3的附录A。

bitstream\_type：应符合T/AI 109.3的附录A。

channel\_number\_index：应符合T/AI 109.3的附录A。

bitrate\_index：应符合T/AI 109.3的附录A。

raw\_frame\_length：应符合T/AI 109.3的附录A。

resolution：应符合T/AI 109.3的附录A。

addition\_info\_length：指示addition\_info的长度，以字节为单位

addition\_info：指示Avs3AudioGHSpecificConfig配置中的额外信息。

* + - 1. AVS3音频无损音频编码特有配置
				1. 定义

本节定义适用于AVS3音频无损音频编码内容的特有配置。

* + - * 1. 语法

class Avs3AudioLLSpecificConfig {

unsigned int(4) sampling\_frequency\_index;

if (sampling\_frequency\_index==0xF) {

unsigned int(24) sampling\_frequency;

}
unsigned int(1) anc\_data\_index;

unsigned int(3) coding\_profile;

unsigned int(8) channel\_number;

unsigned int(2) resolution;

unsigned int(16) addition\_info\_length;

if (addition\_info\_length > 0) {

bit(8\*addition\_info\_length) addition\_info;

}

unsigned int(2) reserved;

}

* + - * 1. 语义

sampling\_frequency\_index：应符合T/AI 109.3的附录A。

sampling\_frequency：应符合T/AI 109.3的附录A。

anc\_data\_index：应符合T/AI 109.3的附录A。

coding\_profile：应符合T/AI 109.3的附录A。

channel\_number：应符合T/AI 109.3的附录A。

resolution：应符合T/AI 109.3的附录A。

addition\_info\_length：指示addition\_info的长度，以字节为单位

addition\_info：指示Avs3AudioLLSpecificConfig配置中的额外信息。

* + 1. ISO基本媒体文件格式扩展
			1. AVS3音频解码器配置数据盒
				1. 定义

**数据盒类型**：'dca3'
**容器：**'av3a'或'a3as'类型的样本入口
**强制性：**强制包含于'av3a'或'a3as'类型的样本入口
**数量：**一个

AVS3音频解码器配置数据盒CA3SpecificBox包含5.1.2中定义的音频编码特有配置。

* + - * 1. 语法

class CA3SpecificBox extends Box('dca3') {

unsigned int(4) audio\_codec\_id;
 if (audio\_codec\_id == 2){

 Avs3AudioGASpecificConfig() Avs3AudioGAConfig;

 }

else if (audio\_codec\_id == 0) {

 Avs3AudioGHSpecificConfig() Avs3AudioGHConfig;

 }

else if (audio\_codec\_id == 1) {

 Avs3AudioLLSpecificConfig() Avs3AudioLLConfig;

 }
}

* + - * 1. 语义

Avs3AudioSpecificGAConfig在5.1.2.1中定义，Avs3AudioSpecificGHConfig在5.1.2.2中定义, Avs3AudioSpecificLLConfig在5.1.2.3中定义。

audio\_codec\_id：应符合T/AI 109.3的附录A。

* + - 1. AVS3音频样本入口
				1. 定义

**样本入口类型：**'av3a'或'a3as'
**容器：**Sample Description Box ('stsd')
**强制性：**封装AVS3音频编码位流的轨道必须包含一个'av3a'或'a3as'样本入口
**数量：**一个

AATF类型的AVS3音频编码位流在'av3a'类型的音频轨道中进行传输，其轨道样本入口中应包含一个CA3SpecificBox数据盒。

AASF类型的AVS3音频编码位流在文件中应被存储为'a3as'类型的音频轨道，其轨道样本入口中应包含aasf\_header()以及一个CA3SpecificBox数据盒。

* + - * 1. 语法

class AVS3ATSampleEntry() extends AudioSampleEntry ('av3a'){

CA3SpecificBox config;

}

class AVS3ASSampleEntry() extends AudioSampleEntry ('a3as'){

CA3SpecificBox config;

unsigned int(16) avs3\_as\_header\_length;

bit(8\*avs3\_as\_header\_length) avs3\_as\_header;

}

* + - * 1. 语义

CA3SpecificBox 提供AVS3音频编码位流的解码配置信息。

由于CA3SpecificBox已经提供并扩展了对ChannelCount，SampleSize，SampleRate的描述，所以本文件规定解码器应忽略AudioSampleEntry中的ChannelCount，SampleSize，SampleRate。

avs3\_as\_header\_length：指示avs3\_as\_header的长度，以字节为单位。

avs3\_as\_header：指示aasf\_header()的信息，aasf\_header()应符合T/AI 109.3的附录A。

* + - 1. AVS3音频样本格式

若音频轨道的样本入口类型为'av3a'，则其轨道中的每个样本对应一个aatf\_frame()，其中aatf\_frame()的定义应符合T/AI 109.3附录A。

若音频轨道的样本入口类型为'a3as'，则其轨道中的每个样本对应aasf\_sequence()的一部分数据，其中aasf\_sequence()的定义应符合T/AI 109.3的附录A。

* 1. AVS2音频文件格式
		1. AVS2音频基本流定义
			1. 概述

AVS2音频基本流即AVS2音频编码位流，应符合GB/T 33475.3规范。

* + - 1. AVS2音频编码特性

AVS2音频编码系统支持声道信号编码、对象信号编码、HOA信号编码和元数据编码。

AVS2音频编码器由多种编码工具构成，包括：通用音频编码工具、无损音频编码工具和元数据编码工具。

* + - 1. 基本流结构

根据AVS2音频的编码特性，AVS2音频的基本流结构分为两种类型：通用音频编码位流、无损音频编码位流。

* + - 1. 基本流格式

AVS2音频基本流可以组织为两种格式：存储格式AASF和传输格式AATF，应符合GB/T 33475.3的附录A。

* + 1. AVS2音频配置信息
			1. AVS2音频通用音频编码特有配置
				1. 定义

本节定义适用于AVS2音频通用音频编码内容的特有配置。

* + - * 1. 语法

class AVSAGASpecificConfig {

unsigned int(4) sampling\_frequency\_index;

unsigned int(1) anc\_data\_index;

unsigned int(3) coding\_profile;

unsigned int(1) bitstream\_type;

 unsigned int(7) channel\_number\_index;

 unsigned int(4) bitrate\_index;

 unsigned int(16) raw\_frame\_length;

unsigned int(2) resolution;

 if (anc\_data\_index == 1) {

anc\_data\_block();

}

}

* + - * 1. 语义

sampling\_frequency\_index：应符合GB/T 33475.3的附录A。

anc\_data\_index：应符合GB/T 33475.3的附录A。

coding\_profile：应符合GB/T 33475.3的附录A。

bitstream\_type：应符合GB/T 33475.3的附录A。

channel\_number\_index：应符合GB/T 33475.3的附录A。

bitrate\_index：应符合GB/T 33475.3的附录A。

raw\_frame\_length：应符合GB/T 33475.3的附录A。

resolution：应符合GB/T 33475.3的附录A。

anc\_data\_block()：应符合GB/T 33475.3的附录A。

* + - 1. AVS2音频无损音频编码特有配置
				1. 定义

本节定义适用于AVS2音频无损音频编码内容的特有配置。

* + - * 1. 语法

class AVSALLSpecificConfig {

unsigned int(4) sampling\_frequency\_index;

if (sampling\_frequency\_index==0xF){

unsigned int(24) sampling\_frequency;

}
unsigned int(1) anc\_data\_index;

unsigned int(3) coding\_profile;

unsigned int(8) channel\_number;

unsigned int(2) resolution;

if (anc\_data\_index == 1) {

anc\_data\_block();

}

}

* + - * 1. 语义

sampling\_frequency\_index：应符合GB/T 33475.3的附录A。

sampling\_frequency：应符合GB/T 33475.3的附录A。

anc\_data\_index：应符合GB/T 33475.3的附录A。

coding\_profile：应符合GB/T 33475.3的附录A。

channel\_number：应符合GB/T 33475.3的附录A。

resolution：应符合GB/T 33475.3的附录A。

anc\_data\_block()：应符合GB/T 33475.3的附录A。

* + 1. ISO基本媒体文件格式扩展
			1. AVS2音频解码器配置数据盒
				1. 定义

**数据盒类型：**'avac'
**容器：**'cavs'或'a2as'类型的样本入口
**强制性：**强制包含于'cavs'或'a2as'类型的样本入口
**数量：**一个

AVS2音频解码器配置数据盒AVSASpecificBox包含5.2.2中定义的音频编码特有配置。

* + - * 1. 语法

class AVSASpecificBox extends Box('avac') {

unsigned int(4) audio\_codec\_id;
 if (audio\_codec\_id == 0){

 AVSAGASpecificConfig();

 }

else if (audio\_codec\_id == 1){

 AVSALLSpecificConfig();

 }

unsigned int(16) addition\_info\_length;

if (addition\_info\_length > 0) {

bit(8\*addition\_info\_length) addition\_info;

}

byte\_alignment();
}

* + - * 1. 语义

AVSAGASpecificConfig在5.2.2.1中定义，AVSALLSpecificConfig在5.2.2.2中定义。

audio\_codec\_id：应符合GB/T 33475.3的附录A。

addition\_info\_length：指示addition\_info的长度，以字节为单位。

addition\_info：指示AVS2音频位流配置中的额外信息。

byte\_alignment()：应符合GB/T 33475.3。

* + - 1. AVS2音频样本入口
				1. 定义

**样本入口类型：**'cavs'或'a2as'
**容器：**Sample Description Box ('stsd')
**强制性：**封装AVS3音频编码位流的轨道必须包含一个'cavs'或'a2as'样本入口
**数量：**一个

AATF类型的AVS2音频编码位流在'cavs'类型的音频轨道中进行传输，其轨道样本入口中应包含一个AVSASpecificBox数据盒。

AASF类型的AVS2音频编码位流在文件中应被存储为'a2as'类型的音频轨道，其轨道样本入口中应包含aasf\_header()以及一个AVSASpecificBox数据盒。

* + - * 1. 语法

class AVSAATSampleEntry() extends AudioSampleEntry ('cavs'){

AVSASpecificBox config

}

class AVSAASSampleEntry() extends AudioSampleEntry ('a2as'){

AVSASpecificBox config;

unsigned int(16) avsa\_as\_header\_length;

bit(8\*avsa\_as\_header\_length) avsa\_as\_header;

}

* + - * 1. 语义

AVSASpecificBox提供AVS2音频编码位流的解码配置信息。

由于AVSASpecificBox已经提供并扩展了对ChannelCount，SampleSize，SampleRate的描述，所以本文件规定解码器应忽略AudioSampleEntry中的ChannelCount，SampleSize，SampleRate。

avsa\_as\_header\_length：指示avsa\_as\_header的长度，以字节为单位。

avsa\_as\_header：指示aasf\_header()的信息，aasf\_header()应符合GB/T 33475.3的附录A。

* + - 1. AVS2音频样本格式

若音频轨道的样本入口类型为'cavs'，则其轨道中的每个样本对应一个aatf\_frame()其中aatf\_frame()应符合GB/T 33475.3附录A。

若音频轨道的样本入口类型为'a2as'，则其轨道中的每个样本对应aasf\_sequence()的一部分数据，其中aasf\_sequence()的应符合GB/T 33475.3的附录A。

1. CMAF轨道和媒体配置
	1. AVS3音频CMAF轨道和媒体配置
		1. 概述

AVS3音频CMAF轨道格式应符合ISO/IEC 23000-19中9.2指定的通用音频CMAF轨道格式，同时应符合本文件第5章中指定的AVS3音频轨道格式以及第6章中指定的约束。

如果CMAF轨道符合上述技术要求，则它被称为AVS3音频CMAF轨道，并且其品牌标识定义为'ca3a'。

* + 1. AVS3音频CMAF轨道约束
			1. 概述

任何符合AVS3音频媒体配置文件的CMAF轨道都应符合:

1. 本文件5.1.3中定义的AVS3音频轨道；
2. ISO/IEC 23000-19中定义的通用视频CMAF轨道格式：
	* 每个presentation必须对应一个唯一的presentation\_id；
	* 每个AVS3音频Sample只能包含一帧且只有一帧aatf\_frame()。
		+ 1. 样本描述数据盒 ('stsd')

AVS3音频轨道中的SampleDescriptionBox应包含符合ISO/IEC 14496-12中规定的一个音频样本入口。

符合 AVS3 音频特有数据盒的 CMAF 轨道的音频样本入口的语法和取值应符合5.1.3中定义的AVS3ATSampleEntry ('av3a') 样本入口。

* + 1. AVS3音频CMAF切换集约束
			1. 概述

对于符合AVS3音频媒体配置文件的CMAF切换集，适用以下约束：

1. CMAF交换集中的每个CMAF轨道应符合6.1.2中定义的 CMAF 轨道约束；
2. CMAF切换集中的每个CMAF轨道应符合ISO/IEC 23000-19中7.3.4规定的通用CMAF切换集约束要求；
3. 单一初始化AVS3音频CMAF切换集应符合6.1.3.2中定义的约束。
	* + 1. 单一初始化CMAF切换集约束

AVS3音频CMAF切换集应符合如下单一初始化约束：

1. 应符合ISO/IEC 23000-19中7.3.4规定的通用CMAF切换集约束要求；
2. 应符合ISO/IEC 23000-19中7.3.4.2规定的通用单一初始化约束；
3. CMAF头部中的音频样本的audio\_codec\_id应保持不变。
	* 1. 音频编解码参数

呈现应用程序应使用符合RFC 6381中规定的参数发送AVS3音频CMAF轨道和CMAF切换集的音频编解码器配置和级别。

AVS3音频媒体配置的MIME类型的“编解码器”参数应符合本文件附录A。

* 1. AVS2音频CMAF轨道和媒体配置
		1. 概述

AVS2音频CMAF轨道格式应该符合ISO/IEC 23000-19中9.2指定的通用音频CMAF轨道格式，同时应符合本文件第5章中指定的AVS2音频轨道格式以及第6章中指定的约束。

如果CMAF轨道符合上述技术要求，则它被称为AVS2音频CMAF轨道，并且其品牌标识定义为‘ca2a’。

* + 1. AVS2音频CMAF轨道约束
			1. 概述

任何符合 AVS2音频媒体配置文件的CMAF轨道都应符合:

1. 本文件5.2.3中定义的AVS2音频轨道；
2. ISO/IEC 23000-19中定义的通用视频CMAF轨道格式：
	* 每个presentation必须对应一个唯一的presentation\_id；
	* 每个AVS2音频Sample只能包含一帧且只有一帧aatf\_frame()。
		+ 1. 样本描述数据盒 ('stsd')

AVS2音频轨道中的SampleDescriptionBox应包含符合ISO/IEC 14496-12中规定的一个音频样本入口。

符合AVS2音频特有数据盒的 CMAF 轨道的音频样本入口的语法和取值应符合5.2.3中定义的AVSAATSampleEntry ('cavs') 样本入口。

* + 1. AVS2音频CMAF切换集约束
			1. 概述

对于符合AVS2音频媒体配置文件的CMAF切换集，适用以下约束：

1. CMAF交换集中的每个CMAF轨道应符合6.2.2中定义的CMAF轨道约束；
2. CMAF切换集中的每个CMAF轨道应符合 ISO/IEC 23000-19中7.3.4规定的通用 CMAF切换集约束要求；
3. 单一初始化AVS2音频CMAF切换集应符合6.2.3.2节中定义的约束。
	* + 1. 单一初始化CMAF切换集约束

AVS2音频CMAF切换集应符合如下单一初始化约束：

1. 应符合ISO/IEC 23000-19 7.3.4中规定的通用CMAF切换集约束要求；
2. 应符合ISO/IEC 23000-19 7.3.4.2中规定的通用单一初始化约束；
3. CMAF头部中的音频样本的audio\_codec\_id应保持不变。
	* 1. 音频编解码参数

呈现应用程序应使用符合RFC 6381中规定的参数发送AVS2音频CMAF轨道和CMAF切换集的音频编解码器配置和级别。

AVS2音频媒体配置的MIME类型的“编解码器”参数应符合本文件附录A。

1. DASH传输技术要求
	1. AVS3音频DASH传输技术要求
		1. 概述

本部分规定AVS3音频编码位流通过基于HTTP的动态自适应流媒体传输协议（ISO/IEC 23009-1）进行传输时的媒体呈现描述（MPD）与片段格式。

* + 1. DASH片段格式
			1. 概述

用于封装AVS3音频数据的DASH片段格式应符合本文件第5章规定的AVS3音频编码位流的文件格式，其样本入口类型应支持'av3a'。

* + - 1. 初始化片段

每个DASH初始化片段应包含一个CA3SpecificBox解码器配置记录。

* + - 1. 媒体片段

每个DASH媒体片段应包含一个或多个T/AI 109.3标准中规定的音频编码数据。

每个DASH媒体片段中的第一个媒体样本应符合以下任意一个约束：

1. 每个AVS3音频样本只能包含一帧且只有一帧aatf\_frame()；
2. 每个AVS3音频样本的audio\_codec\_id应保持不变；
3. 每个AVS3音频样本的SAP的类型，在ISO/IEC 14496-12附录I定义，值都为1。
	* + 1. 索引片段

AVS3音频索引片段应满足以下约束：

1. 每个子片段由一个ISO/IEC 14496-12中8.16.3中定义的'sidx'类型的片段索引数据盒索引；
2. sidx'数据盒指示引用的子片段的starts\_with\_SAP为1，SAP\_type为1。
	* 1. DASH MPD编码器参数

AVS3音频编码位流在MPD中的@codecs属性使用本文件附录A中定义的MIME类型的'codecs'参数。

* + 1. DASH MPD描述子
			1. 概述

本部分定义并使用了一些新的XML元素和属性，其命名空间及规则请参考本文件7.1.4。

* + - 1. @mimeType属性

@mimeType属性应设置为"audio/mp4"。

* + - 1. @audioSampleRate属性

音频采样率的属性源于CA3SpecificBox的sampling\_frequency\_index。

* + - 1. @startWithSAP属性

@startWithSAP属性应该设置成1。

* + - 1. AudioChannelConfiguration描述子

@schemeIdUri属性设置为"urn:avs:avs3:p7:2024:audio\_channel\_configuration"，用于描述编码位流包含的音频配置。

@value属性值为3个字节，计算方式如下：

1. 如果AVS3音频编码位流在MPD中的@codecs属性为'av3a.00'，则@value的属性值
	* 第1个字节为0xF0；
	* 第2个字节的最高1位等于0，低7位等于5.1.2.2.2中channel\_number\_index的值；
	* 第3个字节为0；
2. 如果AVS3音频编码位流在MPD中的@codecs属性为'av3a.01'，则@value的属性值
	* 第1个字节为0xF0；
	* 第2个字节等于5.1.2.3.2中channel\_number的值；
	* 第3个字节为0；
3. 如果AVS3音频编码位流在MPD中的@codecs属性为'av3a.02 '，则@value的属性值：
	* 第1个字节的高4位为0xF，低4位等于5.1.2.1.2中content\_type的值，其中，content\_type的取值范围为0~3;
	* 第2个字节：如果第1个字节等于0xF0、0xF2或者0xF3，则第2个字节最高1位等于0，低7位等于5.1.2.1.2中channel\_number\_index的值；否则，如果第1个字节等于0xF1，第2个字节等于5.1.2.1.2中object\_channel\_number + 1的值；
	* 第3个字节：如果第1个字节的值等于0xF2，则第3个字节的值等于5.1.2.1.2中object\_channel\_number + 1的值；否则，第3个字节的值等于0；
	1. AVS2音频DASH传输技术要求
		1. 概述

本章规定AVS2音频编码位流通过基于HTTP的动态自适应流媒体传输协议（ISO/IEC 23009-1）进行传输时的媒体呈现描述（MPD）与片段格式。

* + 1. DASH片段格式
			1. 概述

用于封装AVS2音频数据的DASH片段格式应符合本文件第5章规定的AVS2音频编码位流的文件格式，其样本入口类型应支持'cavs'。

* + - 1. 初始化片段

每个DASH初始化片段应包含一个AVSASpecificBox解码器配置记录。

* + - 1. 媒体片段

每个DASH媒体片段应包含一个或多个GB/T 33475.3标准中规定的音频编码数据。

每个DASH媒体片段中的第一个媒体样本应符合以下任意一个约束：

1. 每个AVS2音频样本只能包含一帧且只有一帧aatf\_frame()；
2. 每个AVS2音频样本的audio\_codec\_id应保持不变；
3. 每个AVS2音频样本的SAP的类型，在ISO/IEC 14496-12附录I定义，值都为1。
	* + 1. 索引片段

AVS2音频索引片段应满足以下约束：

1. 每个子片段由一个ISO/IEC 14496-12中8.16.3定义的'sidx'类型的片段索引数据盒索引；
2. sidx'数据盒指示引用的子片段的starts\_with\_SAP为1，SAP\_type为1。
	* 1. DASH MPD编码器参数

AVS2音频编码位流在MPD中的@codecs属性使用本文件附录A中定义的MIME类型的'codecs'参数。

* + 1. DASH MPD描述子
			1. 概述

本部分中定义并使用了一些新的XML元素和属性，其命名空间及规则请参考本文件7.2.4。

* + - 1. @mimeType属性

@mimeType属性应设置为"audio/mp4"。

* + - 1. @audioSampleRate属性

音频采样率的属性源于AVSASpecificBox的sampling\_frequency\_index。

* + - 1. @startWithSAP属性

@startWithSAP属性应该设置成1。

* + - 1. AudioChannelConfiguration描述子

@schemeIdUri属性设置为"urn:avs:avs3:p7:2024:audio\_channel\_configuration"，用于描述编码位流包含的音频配置。

@value属性值为3个字节，计算方式如下：

1. 如果AVS2音频编码位流在MPD中的@codecs属性为'cavs.00'，则@value的属性值
	* 第1个字节为0xF0；
	* 第2个字节的最高1位等于0，低7位等于5.2.2.1.2中channel\_number\_index的值；
	* 第3个字节为0；
2. 如果AVS2音频编码位流在MPD中的@codecs属性为'cavs.01'，则@value的属性值
	* 第1个字节为0xF0；
	* 第2个字节等于5.2.2.2.2中channel\_number的值；
	* 第3个字节为0；
3. 传输流和节目流技术要求
	1. AVS3音频传输流和节目流技术要求
		1. 概述

本章规定适用于AVS3音频编码位流的传输流的编码结构与参数。

AVS3音频流应满足以下约束：

1. AVS3音频流应是ISO/IEC 13818-1中节目的一个节目元素，基本流的stream\_type字段值应等于 '0xD5'；
2. AVS3音频使用AATF的封装格式，即封装成aatf\_frame()；
3. AVS3音频流的常见编码参数，如audio\_codec\_id应使用AVS3音频流描述符标识。如果存在与AVS3音频流相关联的AVS3音频流描述符，则该描述符应包含在节目映射表中相应基本流条目的描述符循环中。
	* 1. PES分组
			1. 流标识

AVS3音频应作为PES\_packet\_data\_bytes携带在PES分组数据包中，并通过节目映射表中分配的stream\_type字段值（0xD5）标识。

AVS3音频的PES数据包应满足以下约束：

1. PES分组的stream\_id取值'1111 1101'（extended\_stream\_id）；
2. PES分组包头中stream\_id\_extension\_flag取值'0'，stream\_id\_extension字段的取值'100 1111'用于表示AVS3音频；
3. elementary stream在PES payload里应该是字节对齐的，即AVS3音频的首字节必须位于PES payload的首字节中；
4. 一个PES包可以包含一帧或多帧aatf\_frame()。
	* 1. 节目和节目元素描述符
			1. 节目和节目元素描述子中各字段的语义定义

以下语义适用于从8.1.3.2到8.1.3.3中定义的描述符。

描述符标签字段 descriptor\_tag

该字段为8位，用于标识每一描述符，其中registration\_descriptor描述符标签值在ISO/IEC 138181-1中已规定为0x05。

本部分定义的注册描述符和AVS3音频流描述符标签值，见表3。TS或PS栏中'X'表示该描述符可分别用于传输流或节目流。

表3 节目和节目元素描述子

|  |  |  |  |
| --- | --- | --- | --- |
| 描述符标签值 | TS | PS | 标识 |
| 210 | X | X | AVS3\_audio\_descriptor |

描述子长度字段 descriptor\_length

该字段为8位。规定了紧跟在该字段之后的描述子的字节数。

* + - 1. 注册描述符

registration\_descriptor 提供了一种唯一且明确地识别私有数据格式的方法。

* + - 1. 注册描述符中各字段语义定义

registration\_descriptor()的定义请参考ISO/IEC 13818-1，AVS3音频的format\_identifier应为0x41-56-53-41('AVSA')。

* + - 1. AVS3 音频流描述符

AVS3\_audio\_descriptor()位于PMT中的ES\_info\_length字段后面，其语法见表4。

表4 AVS3音频流描述符语法

|  |  |  |
| --- | --- | --- |
| **语 法** | **位 数** | **助 记 符** |
| AVS3\_audio\_descriptor(){ |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length**  | **8** | **uimsbf** |
| **audio\_codec\_id** | **4** | **uimsbf** |
| **sampling\_frequency\_index**  | **4** | **uimsbf** |
| if(audio\_codec\_id == 0){ |  |  |
| **anc\_data\_index** | **1** | **bslbf** |
|   **coding\_profile** | **3** | **uimsbf** |
|  **bitrate\_index** | **4** | **uimsbf** |
|  **bitstream\_type** | **1** | **bslbf** |
|  **channel\_number\_index** | **7** | **uimsbf** |
|  **raw\_frame\_length** | **16** | **uimsbf** |
| } |  |  |
| if(audio\_codec\_id==1) { |  |  |
| if(sampling\_frequency\_index==0xf) { |  |  |
| **sampling\_frequency** | **24** | **uimsbf** |
| } |  |  |
| **anc\_data\_index** | **1** | **bslbf** |
|   **coding\_profile** | **3** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
|  **channel\_number** | **8** | **uimsbf** |
| } |  |  |
|  if(audio\_codec\_id==2) { |  |  |
| **nn\_type** | **3** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
|  **content\_type** | **4** | **uimsbf** |
|  if(content\_type==0) { |  |  |

表4（续）

|  |  |  |
| --- | --- | --- |
| **语 法** | **位 数** | **助 记 符** |
|   **channel\_number\_index** | **7** | **Uimsbf** |
| **reserved** | **1** | **bslbf** |
|  }else if(content\_type==1) { |  |  |
| **object\_channel\_number** | **7** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
| }else if(content\_type==2) { |  |  |
|   **channel\_number\_index** | **7** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
|  **object\_channel\_number** | **7** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
| }else if(content\_type==3) { |  |  |
| **hoa\_order** | **4** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
|  } |  |  |
|   **total\_bitrate** | **16** | **uimsbf** |
|  } |  |  |
|   **resolution** | **2** | **uimsbf** |
| **reserved** | **6** | **bslbf** |
| for (i=0; i<N; i++) { |  |  |
| **addition\_info[i]**  | **8** | **bslbf** |
| } |  |  |
| } |  |  |

* + - 1. AVS3音频流描述符中各字段的语义定义

descriptor\_tag：该字段为8位，取值为210(0xD2)。

descriptor\_length：该字段为8位，表示AVS3音频描述符的长度。

audio\_codec\_id：该字段为4位，应符合T/AI 109.3的附录A。

sampling\_frequency\_index：该字段为4位，应符合T/AI 109.3的附录A。

anc\_data\_index：该字段为1位，应符合T/AI 109.3的附录A。

coding\_profile：该字段为3位，应符合T/AI 109.3的附录A。

bitrate\_index：该字段为4位，应符合T/AI 109.3的附录A。

bitstream\_type：该字段为1位，应符合T/AI 109.3的附录A。

channel\_number\_index：该字段为7位，应符合T/AI 109.3的附录A。

raw\_frame\_length：该字段为16位，应符合T/AI 109.3的附录A。

sampling\_frequency：该字段为24位，应符合T/AI 109.3的附录A。

channel\_number: 该字段为8位，应符合T/AI 109.3的附录A。

nn\_type：该字段为3位，应符合T/AI 109.3的附录A。

content\_type：该字段为4位，表示音频内容类型，见表5。

表5 content\_type配置表

|  |  |  |
| --- | --- | --- |
| **content\_type值** | **音频内容类型** | **映射关系** |
| 0 | 声道信号 | coding\_profile值为0时 |
| 1 | 对象信号 | coding\_profile值为1且soundBedType值为0时 |
| 2 | 声道信号和对象信号混合 | coding\_profile值为1且soundBedType值为1时 |
| 3 | HOA信号 | coding\_profile值为2时 |
| 4-15 | 保留 |

注：coding\_profile和soundBedType应符合T/AI 109.3的附录A。

object\_channel\_number：该字段为7位，应符合T/AI 109.3附录A。

hoa\_order：该字段为4位，表示HOA信号阶数，应符合T/AI 109.3的附录A中order的计算方法。

total\_bitrate：该字段为16位，表示音频总比特率，单位kbps，计算方法见表6。

表6 total\_bitrate配置表

|  |  |
| --- | --- |
| **content\_type值** | **total\_bitrate计算方式** |
| 0 | 声道信号的bitrate\_index对应的比特率 |
| 1 | 对象信号的bitrate\_index对应的比特率×(object\_channel\_number + 1) |
| 2 | 声道信号的bitrate\_index对应的比特率＋对象信号的bitrate\_index对应的比特率×(object\_channel\_number + 1) |
| 3 | HOA信号的bitrate\_index对应的比特率 |
| 4-15 | 保留 |

注：bitrate\_index应符合T/AI 109.3的附录A。

resolution：该字段为2位，应符合T/AI 109.3的附录A。

addition\_info：可选字段，附加信息。

* + 1. AVS3音频T-STD模型拓展

AVS3音频T-STD模型拓展中访问单元AU（Access Unit）表示使用AATF封装格式的一个音频帧。对于包含AVS3音频的传输流应符合T-STD模型，见图1。



图1 面向AVS3音频的T-STD模型拓展

图1中符号说明：

* $t(i)$：传输流的第i个字节进入系统目标解码器的时间，单位秒。
* $TB\_{n}$：基本流n的传输缓存。
* $B\_{n}$：基本流n的主缓存。
* $Rx\_{n}$：从$TB\_{n}$到$B\_{n}$的传输速率。
* $A\_{n}(j)$：AVS3音频基本流的第j个存储单元。
* $td(j)$：$A\_{n}(j)$在系统目标解码器中解码的时间，单位秒。
* $D\_{n}$：基本流n对应的解码器。
* $P\_{n}(k)$：基本流n中第k个呈现单元。
* $tp\_{n}(k)$：基本流n中第k个呈现单元对应的时间，单位秒。
	+ - 1. 缓存管理

AVS3音频T-STD模型扩展中从$TB\_{n}$到$B\_{n}$的传输速率$Rx\_{n}$计算公式如下：

$$Rx\_{n}=1.2 ×R\_{max}×N$$

其中：

1. $R\_{max}$：AVS3音频单通道最大速率；
2. $N$：基本流中包含的音频通道总数。

AVS3音频T-STD模型拓展中对应$Rx\_{n}$取值，见表7。

表7 $Rx\_{n}$取值范围

|  |  |
| --- | --- |
| **通道数** | $Rx\_{n}$**[bit/s]** |
| 1-8 | 2 000 000 |
| 9-16 | 3 686 400 |
| 17-48 | 11 059 200 |
| 49-128 | 29 491 200 |

AVS3音频T-STD模型扩展中主缓存$BS\_{n}$计算公式如下：

$$BS\_{n}= BS\_{mux}+BS\_{dec} + BS\_{oh} $$

其中：

1. $BS\_{mux}$：额外复用缓存，计算公式如下：

$$BS\_{mux}=0.004×R\_{max}×N$$

1. $BS\_{dec}$：基本流存储单元缓存，取值为6144 bits；
2. $BS\_{oh}$：PES包头缓存，取值为528 bytes。

AVS3音频-T-STD模型拓展中对应$BS\_{n}$取值，见表8。

 表8 $ BS\_{n}$取值范围

|  |  |
| --- | --- |
| **通道数** | **BSn [bytes]** |
| 1-8 | 7440 |
| 9-16 | 14 352 |
| 17-48 | 42 000 |
| 49-128 | 111 120 |

* + - 1. 缓存延时

AVS3音频传输流STD延时应该满足：对于所有j对应的存取单元$A\_{n}(j)$中的所有字节i，对应$ td\_{n}\left(j\right)-t\left(i\right)\leq 10 (s)$ 。

* + - 1. 缓存管理条件

缓存管理条件应该满足：

* $TB\_{n}$不应上溢，且应在每一秒中至少有一次处于清空状态；
* $B\_{n}$既不能上溢，也不能下溢。
	1. AVS2音频传输流和节目流技术要求
		1. 概述

本章规定适用于AVS2音频编码位流的传输流的编码结构与参数。

AVS2音频流应满足以下约束：

1. AVS2音频流应是ISO/IEC 13818-1中节目的一个节目元素，基本流的stream\_type字段值应等于 '0xD3'；
2. AVS2音频使用AATF的封装格式，即封装成aatf\_frame()；
3. AVS2音频流的常见编码参数，如audio\_codec\_id应使用AVS2音频流描述符标识。如果存在与AVS2音频流相关联的AVS2音频流描述符，则该描述符应包含在节目映射表中相应基本流条目的描述符循环中。
	* 1. PES分组
			1. 流标识

AVS2音频应作为PES\_packet\_data\_bytes携带在PES分组数据包中，并通过节目映射表中分配的stream\_type字段值（0xD3）标识。

AVS2音频的PES数据包应满足以下约束：

1. PES分组的stream\_id取值'1011 1101'（private\_stream\_1）。
2. elementary stream在PES payload里应该是字节对齐的，即AVS2音频的首字节必须位于PES payload的首字节中。
3. 一个PES包可以包含一帧或多帧aatf\_frame()。
	* 1. 节目和节目元素描述符
			1. 节目和节目元素描述子中各字段的语义定义

以下语义适用于从8.2.3.2到8.2.3.3中定义的描述符。

描述符标签字段 descriptor\_tag

该字段为8位，用于标识每一描述符，其中registration\_descriptor描述符标签值在ISO/IEC 138181-1中已规定为0x05。

本部分定义的注册描述符和AVS2音频流描述符标签值，见表9。TS或PS栏中‘X’表示该描述符可分别用于传输流或节目流。

表9 节目和节目元素描述子

|  |  |  |  |
| --- | --- | --- | --- |
| 描述符标签值 | TS | PS | 标识 |
| 211 | X | X | AVS2\_audio\_descriptor |

描述子长度字段 descriptor\_length

该字段为8位。规定了紧跟在该字段之后的描述子的字节数。

* + - 1. 注册描述符

registration\_descriptor 提供了一种唯一且明确地识别私有数据格式的方法。

* + - 1. 注册描述符中各字段语义定义

registration\_descriptor()的定义请参考ISO/IEC 13818-1，AVS2音频的format\_identifier应为0x41-56-53-41('AVSA')。

* + - 1. AVS2 音频流描述符

AVS2\_audio\_descriptor()位于PMT中的ES\_info\_length字段后面，其语法见表10。

表10 AVS2音频流描述符语法

|  |  |  |
| --- | --- | --- |
| **语 法** | **位 数** | **助 记 符** |
| AVS2\_audio\_descriptor() { |  |  |
|  **descriptor\_tag** | **8** | **uimsbf** |
|  **descriptor\_length** | **8** | **uimsbf** |
| **num\_channels** | **8** | **uimsbf** |
|  **sample\_rate\_index** | **4** | **uimsbf** |
|  **avs\_version\_flag** | **1** | **bslbf** |
|  **text\_present\_flag** | **1** | **bslbf** |

表10（续）

|  |  |  |
| --- | --- | --- |
| **语 法** | **位 数** | **助 记 符** |
|   **language\_present\_flag** | **1** | **Bslbf** |
|   **reserved** | **1** | **‘0’** |
|  If (text\_present\_flag == 1) { |  |  |
|  **text\_length** | **8** | **uimsbf** |
|  for (i=0; i<text\_length; i++) { |  |  |
|   **text[i]** | **16** | **bslbf** |
|  } |  |  |
|  } |  |  |
|  if (language\_present\_flag == 1) { |  |  |
|  **language** | **24** | **bslbf** |
|  } |  |  |
| if (avs\_version\_flag == 1) { |  |  |
| **audio\_codec\_id** | **4** | **uimsbf**  |
| **anc\_data\_index** | **1** | **bslbf** |
| **coding\_profile** | **3** | **uimsbf** |
| if (audio\_codec\_id==0) { |  |  |
|  **bitrate\_index** | **4** | **uimsbf** |
| **bitstream\_type** | **1** | **bslbf** |
| **reserved** | **3** | **bslbf** |
| **raw\_frame\_length** | **16** | **uimsbf** |
| } |  |  |
| **resolution** | **2** | **uimsbf** |
| **reserved** | **6** | **bslbf** |
| if (anc\_data\_index) { |  |  |
|  anc\_data\_block() |  |  |
| } |  |  |
| } |  |  |
|  for (i=0; i<N; i++) { |  |  |
|  **addition\_info[i]** | **8** | **bslbf** |
|  } |  |  |
| } |  |  |

* + - 1. AVS2音频流描述符中各字段的语义定义

descriptor\_tag：该字段为8位，取值为0xD3。

descriptor\_length：该字段为8位，表示AVS2音频流描述子长度。对于AVS2音频流描述子，最小值是2，即该字段后至少需要2字节的描述子内容。

num\_channels：该字段为8位，表示AVS2音频流声道数。

sample\_rate\_index：该字段为4位，应符合GB/T33475.3附录A.2。

avs\_version\_flag：该字段为1位，表示AVS2版本号，通常设置为1。

text\_present\_flag：该字段为1位，表示AVS2音频流描述子中是否包含描述性的文本。若text\_present\_flag=1，表示音频流中包含文本；若text\_present\_flag=0，则不包含。

language\_present\_flag：该字段为1位，说明AVS2音频流描述子中是否包含3字节的语言字段。若language\_present\_flag=1，表示后面包含语言字段；若language\_present\_flag=0，则不包含语言字段。

text\_length：该字段为8位，表示描述性文本字段长度。

text[i]：文本字段，AVS2音频流的文本描述，使用2字节的unicode字符集编码。

language：该字段为24位，编码规定应符合GB/T4880.2—2000。

audio\_codec\_id: 该字段为4位，应符合GB/T33475.3的附录A.2。

anc\_data\_index: 该字段为1位，应符合GB/T33475.3的附录A.2。

coding\_profile: 该字段为3位，应符合GB/T33475.3的附录A.2。

bitrate\_index: 该字段为4位，应符合GB/T33475.3的附录A.2。

bitstream\_type: 该字段为1位，应符合GB/T33475.3的附录A.2。

raw\_frame\_length：该字段为16位，应符合GB/T33475.3的附录A.2。

resolution: 该字段为2位，应符合GB/T33475.3的附录A.2。

anc\_data\_block(): 应符合GB/T33475.3的附录A.2。

additional\_info：可选字段，附加信息。

* + 1. STD音频缓冲器大小

AVS2基本流有一个或多个声道组成，每个声道的最大码率为audio\_codec\_id为0时,192kbps，因此AVS2音频在传输缓存中的漏率计算为：

 RXn = 1.2×Rmax×N bps

式中：

Rmax恒为audio\_codec\_id为0时，192kbps，AVS2音频码流每个声道的码率上限的值；

N是AVS2音频基本流中包含的音频声道总数。

AVS2音频T-STD模型扩展中对应RXn的取值，见表11。

表11 $Rx\_{n}$取值范围

|  |  |
| --- | --- |
| **通道数** | $Rx\_{n}$**[bit/s] (audio\_codec\_id=0)** |
| 1-8 | 2 000 000 |
| 9-16 | 3 686 400 |
| 17-48 | 11 059 200 |
| 49-128 | 29 491 200 |

AVS2音频T-STD模型扩展中主缓存BSn大小定义为：

BSn = BSmux + BSdec + BSoh

式中：

BSmux——额外复用缓存，计算公式如下：

BSmux = 0.004×Rmax×N；

BSdec——基本流存储单元缓存，audio\_codec\_id为0时，6144bits×N。

BSoh——PES头大小，528字节。

其中Rmax和N的定义同上。

AVS2音频T-STD模型扩展中对应BSn的取值，见表12。

表12 $BS\_{n}$取值范围

|  |  |
| --- | --- |
| **通道数** | **BSn [bytes]****(audio\_codec\_id=0)** |
| 1-8 | 7 440 |
| 9-16 | 14 352 |
| 17-48 | 42 000 |
| 49-128 | 111 120 |

对于节目流，设置主缓存大小在P-STD\_buffer\_scale和P-STD\_buffer\_size中，见表13。

表13 P-STD\_buffer\_size取值范围

|  |  |  |
| --- | --- | --- |
| **Number of Channels** | **P-STD\_buffer\_scale** | **P-STD\_buffer\_size****(audio\_codec\_id=0)** |
| 1-8 | 0 | 59 |
| 9-16 | 0 | 113 |
| 17-48 | 0 | 329 |
| 49-128 | 0 | 869 |

* + 1. 字节对齐

音频ES流在GB/T 17975.1-2010或MPEG-2 TS传输流中应是字节对齐的，即编码帧的最初8比特应包含在GB/T 17975.1-2010或MPEG-2 TS传输流中的一个完整的独立字节中。

1. SMT传输技术要求
	1. AVS3音频SMT传输技术要求
		1. 概述

AVS3音频编码位流基于SMT的传输文件需遵循如下约束：

1. AVS3音频编码位流应符合本文件第5章中基于ISO BMFF的文件封装格式；
2. AVS3音频文件应符合SMT的文件封装要求，以通用封装单元的形式通过SMT进行传输；
3. AVS3音频文件传输过程中使用的信令消息，应符合SMT中信令消息的定义以及本部分的扩展定义。
	* 1. AVS音频媒体资源描述符
			1. 定义

AVS音频媒体资源描述符用于指示AVS3音频编码位流的编码类别、编码档次、存储模式等信息。AVS音频媒体资源描述符在SMT的MP表中进行扩展，用于解决AVS3音频在SMT协议下灵活传输与个性化消费的需求。

* + - 1. 语法

AVS音频媒体资源描述符语法，见表14。

表14 AVS音频媒体资源描述符

|  |  |  |  |
| --- | --- | --- | --- |
| **语法** | **值** | **比特数** | **备注** |
| Audio\_info\_descriptor () {descriptor\_tagdescriptor\_lengthaudio\_format\_typeaudio\_codec\_idcoding\_profileaverage\_bitrate\_flaghoa\_order\_flagchannel\_number\_flagobject\_info\_flagreservedif(average\_bitrate\_flag==1)average\_bitrateelse {max\_bitrate min\_bitrate}if(hoa\_order\_flag){ max\_hoa\_order}if(channel\_number\_flag){ max\_channel\_number}if(object\_info\_flag){ max\_object\_channel\_number}bit\_depth\_resolutionsample\_rate} | ‘1’   | 161644311111161616888824 | uimsbfuimsbfuimsbfuimsbfuimsbfbslbfbslbfbslbfbslbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbf |

* + - 1. 语义

descriptor\_tag：用于标识descriptor的类型。

descriptor\_length：指示标识符的长度，单位为字节。

audio\_format\_type：指示AVS音频编码位流的类别。该字段取值为0表示位流为AVS3音频AASF存储格式的位流；该字段取值为1表示位流为AVS3音频AATF传输格式的位流；该字段取值为2表示位流为AVS2音频AASF存储格式的位流；该字段取值为3表示位流为AVS2音频AATF传输格式的位流。

audio\_codec\_id：指示音频媒体资源的编码类别。对于AVS3音频位流，该字段取值为0时表示媒体资源为通用高码率音频编码数据；取值为1表示媒体资源为无损音频编码数据；该字段取值为2表示媒体资源为通用全码率音频编码数据；其余取值保留。

coding\_profile：指示音频媒体资源的编解码档次。该字段取值为0表示音频媒体资源的编解码遵循基本框架；该字段取值为1表示音频媒体资源的编解码遵循对象元数据编码框架；该字段取值为2表示音频媒体资源的编解码遵循HOA数据编码框架。

average\_bitrate\_flag：取值为0时表示音频媒体资源不具备平均码率；取值为1时表示音频媒体资源具备平均码率。

max\_bitrate：指示音频媒体资源的最大码率，单位kbps。

average\_bitrate：指示音频媒体资源的平均码率，单位kbps。

min\_bitrate：指示音频媒体资源的最小码率，单位kbps。

hoa\_order\_flag：取值为1时表示当前描述符中指示HOA阶数；取值为0时表示当前描述符中不指示HOA阶数。

channel\_number\_flag：取值为1时表示当前描述符中指示声道数；取值为0时表示当前描述符中不指示声道数。

object\_info\_flag：取值为1时表示当前描述符中指示声音对象信息；取值为0时表示当前描述符中不指示声音对象信息。

max\_hoa\_order：指示当前媒体资源支持的最大HOA阶数。

max\_channel\_number：指示当前媒体资源支持的最大声道数。

max\_object\_channel\_number：指示当前媒体资源包含的全部对象支持的最大声道数量。

bit\_depth\_resolution：指示音频输入信号的量化比特数。

sample\_rate：指示音频输入信号的采样频率。

* + 1. 交互反馈信令表
			1. 定义

交互反馈消息提供沉浸式媒体消费时，服务器与客户端之间的交互反馈。当沉浸式媒体消费中的服务器与客户端之间需要发送交互反馈信息时，使用此消息进行会话。一个交互反馈消息信令中可包含一个或多个交互反馈信令表。交互反馈信令表中包含了服务器和客户端之间交互反馈的信息，不同类型的交互反馈信令表用于指示不同类型的交互反馈信息。

对于AVS3音频编码位流的媒体资源，若其包含可交互的声音对象，则用户对于声音对象的交互操作可以通过交互反馈信令表进行反馈,其中声音对象的交互反馈信令表的字段取值应遵循如下约束：

1. table\_type应取值为3；
2. asset\_group\_flag应取值为0。
	* + 1. 语法

交互反馈信令表语法，见表15。

表15 交互反馈信令表

|  |  |  |  |
| --- | --- | --- | --- |
| **语法** | **值** | **比特数** | **类型** |
| interaction\_feedback\_table() {table\_idversionlengthtable\_payload {table\_typetimestampmessage\_sourceasset\_group\_flagreservedif(asset\_group\_flag){asset\_group\_id}else{asset\_id()}if(table\_type == 3){coordinate\_typeif(coordinate\_type == 0){ClientPosition()}if(coordinate\_type == 1){azimuthelevationdistance}}}} |  | 8816811688888 | uimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbf |

* + - 1. 语义

table\_type指示交互反馈信令表携带的信息类型。其取值含义见表16。

表16 交互反馈信令表类型

|  |  |
| --- | --- |
| 取值 | 描述 |
| 0 | 全景视频用户位置变动信息 |
| 1 | 容积视频用户位置变动信息 |
| 2 | 自由视角视频用户位置变动信息 |
| 3 | 音频声音对象交互信息 |
| 4-255 | 未定义 |

Timestamp：指示当前交互产生的时间，使用UTC时间。

message\_source：指示消息源，0表示交互反馈消息是客户端发往服务器，1表示交互反馈消息是服务器发往客户端。该值此处置0。

asset\_group\_flag：指示当前消费内容是否属于一个媒体资源组。取值为1表示客户端当前消费内容属于一个媒体资源组；取值为0表示客户端当前消费内容不属于媒体资源组。

asset\_group\_id：指示客户端当前消费内容的媒体资源组标识符

asset\_id：指示客户端当前消费内容的媒体资源标识符。

coordinate\_type：指示用户交互位置的坐标类型，该字段取值为0表示交互位置以笛卡尔坐标系指示；该字段取值为1表示交互位置以球面坐标系指示。

ClientPosition()指示全局坐标系下用户交互位置的x,y,z坐标，其具体定义如下。

aligned(8) class ClientPosition () {

 signed int(16) position\_x;

 signed int(16) position\_y;

 signed int(16) position\_z;

}

其中，position\_x指示用户实时位置相对起始位置沿着x轴位移，取值范围为（-215，215 − 1），以毫米为单位。

position\_y指示用户实时位置相对起始位置沿着y轴位移，取值范围为（-215，215 − 1），以毫米为单位。

position\_z指示用户实时位置相对起始位置沿着z轴位移，取值范围为（-215，215 − 1），以毫米为单位。

Azimuth、elevation、distance分别指示用户交互位置的方位角、高度和距离。

* 1. AVS2音频SMT传输技术要求
		1. 概述

AVS2音频编码位流基于SMT的传输文件需遵循如下约束：

1. AVS2音频编码位流应遵循第5章中的文件封装格式，封装为基于ISOBMFF的文件；
2. AVS2音频文件应遵循SMT的文件封装要求，以通用封装单元的形式通过SMT进行传输；
3. AVS2音频文件传输过程中使用的信令消息，需遵循SMT中信令消息的定义以及本部分的扩展定义。
	* 1. AVS音频媒体资源描述符
			1. 定义

AVS音频媒体资源描述符用于指示AVS2音频编码位流的编码类别、编码档次、存储模式等信息。AVS音频媒体资源描述符在SMT的MP表中进行扩展，用于解决AVS2音频在SMT协议下灵活传输与个性化消费的需求。

对于AVS2音频编码位流，相应AVS音频媒体资源描述符应满足如下约束：

——hoa\_order\_flag字段取值应为0。

* + - 1. 语法

AVS音频媒体资源描述符语法，见表17。

表17 AVS音频媒体资源描述符

|  |  |  |  |
| --- | --- | --- | --- |
| 语法 | 值 | 比特数 | 备注 |
| Audio\_info\_descriptor () {descriptor\_tagdescriptor\_lengthaudio\_format\_typeaudio\_codec\_idcoding\_profileaverage\_bitrate\_flaghoa\_order\_flagchannel\_number\_flagobject\_info\_flagreservedif(average\_bitrate\_flag==1)average\_bitrateelse {max\_bitrate min\_bitrate}if(hoa\_order\_flag){ max\_hoa\_order}if(channel\_number\_flag){ max\_channel\_number}if(object\_info\_flag){ max\_object\_channel\_number}bit\_depth\_resolutionsample\_rate} | ‘1’   | 161644311115161616888824 | uimsbfuimsbfuimsbfuimsbfuimsbfbslbfbslbfbslbfbslbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbf |

* + - 1. 语义

descriptor\_tag：用于标识descriptor的类型。

descriptor\_length：指示标识符的长度，单位为字节。

audio\_format\_type：指示AVS音频编码位流的类别。该字段取值为0表示位流为AVS3音频AASF存储格式的位流；该字段取值为1表示位流为AVS3音频AATF传输格式的位流；该字段取值为2表示位流为AVS2音频AASF存储格式的位流；该字段取值为3表示位流为AVS2音频AATF传输格式的位流。

audio\_codec\_id：指示音频媒体资源的编码类别。对于AVS2音频位流，该字段取值为0时表示媒体资源为通用音频编码数据；取值为1表示媒体资源为无损音频编码数据；其余取值保留。

coding\_profile：指示音频媒体资源的编解码档次。该字段取值为0表示音频媒体资源的编解码遵循基本框架；该字段取值为1表示音频媒体资源的编解码遵循对象元数据编码框架；该字段取值为2表示音频媒体资源的编解码遵循HOA数据编码框架。

average\_bitrate\_flag：取值为0时表示音频媒体资源不具备平均码率；取值为1时表示音频媒体资源具备平均码率。

max\_bitrate：指示音频媒体资源的最大码率，单位kbps。

average\_bitrate：指示音频媒体资源的平均码率，单位kbps。

min\_bitrate：指示音频媒体资源的最小码率，单位kbps。

hoa\_order\_flag：取值为1时表示当前描述符中指示HOA阶数；取值为0时表示当前描述符中不指示HOA阶数。

channel\_number\_flag：取值为1时表示当前描述符中指示声道数；取值为0时表示当前描述符中不指示声道数。

object\_info\_flag：取值为1时表示当前描述符中指示声音对象信息；取值为0时表示当前描述符中不指示声音对象信息。

max\_hoa\_order：指示当前媒体资源支持的最大HOA阶数。

max\_channel\_number：指示当前媒体资源支持的最大声道数。

max\_object\_channel\_number：指示当前媒体资源包含的全部对象支持的最大声道数量。

bit\_depth\_resolution：指示音频输入信号的量化比特数。

sample\_rate：指示音频输入信号的采样频率。

* + 1. 交互反馈信令表
			1. 定义

交互反馈消息提供沉浸式媒体消费时，服务器与客户端之间的交互反馈。当沉浸式媒体消费中的服务器与客户端之间需要发送交互反馈信息时，使用此消息进行会话。一个交互反馈消息信令中可包含一个或多个交互反馈信令表。交互反馈信令表中包含了服务器和客户端之间交互反馈的信息，不同类型的交互反馈信令表用于指示不同类型的交互反馈信息。

对于AVS2音频编码位流的媒体资源，若其包含可交互的声音对象，则用户对于声音对象的交互操作可以通过交互反馈信令表进行反馈,其中声音对象的交互反馈信令表的字段取值应遵循如下约束：

1. table\_type应取值为3；
2. asset\_group\_flag应取值为0。
	* + 1. 语法

交互反馈信令表语法，见表18。

表18 交互反馈信令表

|  |  |  |  |
| --- | --- | --- | --- |
| **语法** | **值** | **比特数** | **类型** |
| interaction\_feedback\_table() {table\_idversionlengthtable\_payload {table\_typetimestampmessage\_sourceasset\_group\_flagreservedif(asset\_group\_flag){asset\_group\_id}else{asset\_id()}if(table\_type == 3){coordinate\_typeif(coordinate\_type == 0){ClientPosition()}if(coordinate\_type == 1){azimuthelevationdistance}}}} |  | 8816811688888 | uimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbfuimsbf |

* + - 1. 语义

table\_type指示交互反馈信令表携带的信息类型。其取值含义，见表19。

表19 交互反馈信令表类型

|  |  |
| --- | --- |
| **值** | **描述** |
| 0 | 全景视频用户位置变动信息 |
| 1 | 容积视频用户位置变动信息 |
| 2 | 自由视角视频用户位置变动信息 |
| 3 | 音频声音对象交互信息 |
| 4-255 | 未定义 |

timestamp：指示当前交互产生的时间，使用UTC时间。

message\_source：指示消息源，0表示交互反馈消息是客户端发往服务器，1表示交互反馈消息是服务器发往客户端。该值此处置0。

asset\_group\_flag：指示当前消费内容是否属于一个媒体资源组。取值为1表示客户端当前消费内容属于一个媒体资源组；取值为0表示客户端当前消费内容不属于媒体资源组。

asset\_group\_id：指示客户端当前消费内容的媒体资源组标识符

asset\_id：指示客户端当前消费内容的媒体资源标识符。

coordinate\_type：指示用户交互位置的坐标类型，该字段取值为0表示交互位置以笛卡尔坐标系指示；该字段取值为1表示交互位置以球面坐标系指示。

ClientPosition()：指示全局坐标系下用户交互位置的x,y,z坐标，其具体定义如下。

aligned(8) class ClientPosition () {

 signed int(16) position\_x;

 signed int(16) position\_y;

 signed int(16) position\_z;

}

其中，position\_x指示用户实时位置相对起始位置沿着x轴位移，取值范围为（-215，215 − 1），以毫米为单位。

position\_y指示用户实时位置相对起始位置沿着y轴位移，取值范围为（-215，215 − 1），以毫米为单位。

position\_z指示用户实时位置相对起始位置沿着z轴位移，取值范围为（-215，215 − 1），以毫米为单位。

Azimuth、elevation、distance分别指示用户交互位置的方位角、高度和距离。

1. RTP传输技术要求
	1. AVS3音频RTP传输技术要求
		1. RTP负载
			1. 概述

AVS3音频基于RTP协议进行传输时，通过RTP封装将AVS3音频帧封装为若干个RTP包，每个RTP包由RTP头和RTP负载组成。

* + - 1. 封包规则

RTP负载封包规则应当符合以下任何一个约束：

1. RTP负载携带的AVS3音频编码格式为aatf\_frame()；
2. 如果一个aatf\_frame()的长度超过了MTU，需要参考RFC 3550第6.1章的跨包规则来进行传输：将aatf\_frame()按照MTU的倍数分割成多个MTU进行传输。

对于上述分割的MTU，使用RTP头字段里的Marker (M) bit来指明当前该MTU分帧是否是最后一帧。

* + 1. RTP头
			1. 概述

RFC3550第5.1章中定义的RTP头结构定义见图2，对于字段Payload Type(PT), Marker(M) bit, Timestamp，本文件第10.2.2章进行了扩展定义。



图2 RTP头结构

* + - 1. RTP头扩展

Payload Type（PT）：AVS3音频的PT值采用动态范围分配，即PT值属于[96, 127]范围。

Marker (M) bit：M指定aatf\_frame()包的边界。M等于1意味该RTP包包含一个完整的aatf\_frame()，或者aatf\_frame()的最后一个分包。

Timestamp：时间戳默认为90kHz。

* + 1. SDP定义
			1. 概述

SDP会话描述协议（Session Description Protocol）用于在媒体会话中传递媒体流信息，并允许会话描述的接收者参与会话, 在支持RTP的扩展时，除满足RFC 8866所述的规则，需要对第10.1.3.2章中的字段进行扩展定义。

* + - 1. SDP字段定义

"m="的media name对应MIME媒体名字audio。

"a=rtpmap"的encoding name对应MIME子名字AV3A-AATF，表示媒体流为AVS3音频位流

"a=rtpmap"的clock rate对应必需参数rate。

可选参数codec-nn-id、config、bitrate都是"a=fmtp"的属性。

* + 1. MIME类型
			1. 概述

当使用RFC 6381定义的MIME类型的'codecs'参数时，如果 MIME类型对应本标准中定义的文件格式，'codecs'参数值应符合AVS3音频编码位流格式，其样本入口类型应支持'av3a'。

* + - 1. MIME参数定义

MIME媒体名字：audio

MIME子名字：AV3A-AATF

必需参数：

1. **rate**：RTP时间戳的尺度，可以与音频的采样率相同，没有指定的话，默认是90000；

可选参数：

1. **codec-nn-id**：以十六进制两个字节表达，第一个字节（MSB）含义对应于T/AI 109.3的附录A定义的audio\_codec\_id，第二个字节则用来表示该编码器里的nn\_type类型（对应于T/AI 109.3的附录A定义的nn\_type），如果编码器不包含神经网络模型，则该字节的值默认为0；

codec-nn-id使用举例如下：

* + 使用audio\_codec\_id为0的通用高码率编码器，该编码器不包含神经网络模型，则codec-nn-id=0x0000；
	+ 使用audio\_codec\_id为1的无损音频编码器，该编码器不包含神经网络模型，则codec-nn-id=0x0100；
	+ 使用audio\_codec\_id为2的通用全码率编码器，并使用低复杂度神经网络模型(对应nn\_type为1)，则codec-nn-id=0x0201。
1. **config**：16进制的字符串，对应第5.2.2章中定义的CA3SpecificBox；
2. **bitrate**：音频流的码率，对应第5.2.2章中定义的total\_bitrate。
	1. AVS2音频RTP传输技术要求
		1. RTP负载
			1. 概述

AVS2音频基于RTP协议进行传输时，通过RTP封装将AVS2音频帧封装为若干个RTP包，每个RTP包由RTP头和RTP负载组成。

* + - 1. 封包规则

RTP负载封包规则应当符合以下任何一个约束：

1. RTP负载携带的AVS2音频编码格式为aatf\_frame()；
2. 如果一个aatf\_frame()的长度超过了MTU，需要参考RFC 3550第6.1章的跨包规则来进行传输：将aatf\_frame()按照MTU的倍数分割成多个MTU进行传输。

对于上述分割的MTU，使用RTP Header字段里的Marker (M) bit来指明当前该MTU分帧是否是最后一帧。

* + 1. RTP头
			1. 概述

RFC3550第5.1章中定义的RTP头结构定义见图3，对于字段Payload Type(PT), Marker(M) bit, Timestamp，本文件第10.2.2.2章进行了扩展定义。



图3 RTP头结构

* + - 1. RTP头扩展

Payload Type（PT）：AVS2音频的PT值采用动态范围分配，即PT值属于[96, 127]范围。

Marker (M) bit：M指定aatf\_frame()包的边界。M等于1意味该RTP包包含一个完整的aatf\_frame()，或者aatf\_frame()的最后一个分包。

Timestamp：时间戳默认为90kHz。

* + 1. SDP定义
			1. 概述

SDP会话描述协议（Session Description Protocol）用于在媒体会话中传递媒体流信息，并允许会话描述的接收者参与会话, 在支持RTP的扩展时，除满足RFC 8866所述的规则，需要对第10.3.2章中的字段进行扩展定义。

* + - 1. SDP字段定义

"m="的media name对应MIME媒体名字audio。

"a=rtpmap"的encoding name对应MIME子名字AVSA-AATF，表示媒体流为AVS2音频位流。

"a=rtpmap"的clock rate对应必需参数rate。

可选参数codec-nn-id、config、bitrate都是"a=fmtp"的属性。

* + 1. MIME类型
			1. 概述

当使用RFC 6381定义的MIME类型的'codecs'参数时，如果 MIME类型对应本标准中定义的文件格式，' codecs'参数值应符合AVS2音频编码位流格式，其样本入口类型应支持'cavs'。

* + - 1. MIME参数定义

MIME媒体名字：audio

MIME子名字：AVSA-AATF

必需参数：

**——rate**：RTP时间戳的尺度，可以与音频的采样率相同，没有指定的话，默认是90000。

1. （规范性）
MIME类型的'codecs'参数
	1. 概述

当使用RFC 6381定义的MIME类型的'codecs'参数时，如果MIME类型对应本标准中定义的文件格式，

1. 且'codecs'参数值起始于本标准5.1.1章定义的样本入口类型，则'codecs'参数的子参数应符合附录A.2；
2. 且'codecs'参数值起始于本标准5.1.2章定义的样本入口类型，则'codecs'参数的子参数应符合附录A.3。
	1. AVS3音频编码参数

AVS3音频codecs参数定义如下：

codecs = 'av3a.audio\_codec\_id'

示例：

codecs= 'av3a.02'

表示audio\_codec\_id为2。

* 1. AVS2音频编码参数

AVS2音频codecs参数定义如下：

codecs = 'cavs.audio\_codec\_id'

示例：

codecs= 'cavs.01'

表示audio\_codec\_id为1。