

团 体 标 准

T/AI 109.7—2024

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

2024 – 11 – 29 发布

2024 – 11 – 29 实施

中关村视听产业技术创新联盟 发布

T/AI 109.7-2024

T/ALI 109.7-2024



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

T/AI 109.7-2024

目 次

前言	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	3
5 音频基本文件封装格式	4
6 CMAF 轨道和媒体配置	13
7 DASH 传输技术要求	15
8 传输流和节目流技术要求	18
9 SMT 传输技术要求	28
10 RTP 传输技术要求	36
附录 A（规范性） MIME 类型的'codecs'参数	41

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；
- 第7部分：音频封装与传输。

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、北京博雅睿视科技有限公司。

本部分主要起草人：窦维蓓、牟伦田、郑建铎、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、魏建超、黄铁军、高文、崔晓冉。

引 言

T/AI 109旨在规定智能媒体的编码、封装与传输规范，旨在为异构网络中的智能媒体数据提供编码、封装和传输服务。

T/AI 109拟由10个部分组成：

- 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的封装和传输。
- 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。

本文件是T/AI 109的第7部分。本文件第5章和第6章规定了符合AVS3智能媒体编码音频数据和AVS2的高效多媒体编码音频数据的基本文件封装格式和CMAF轨道和媒体配置，并进一步规定了基于第5章和第6章定义的AVS3/AVS2音频封装格式进行DASH传输的技术要求、传输流和节目流的技术要求、SMT传输的技术要求以及RTP传输的技术要求。

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2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	专利申请人/专利权人	标准条款涉及专利的(章、条编号)
1	CN202311676524.8	音频数据传输方法、装置、电子设备及可读存储介质	咪咕文化科技有限公司, 中国移动通信集团有限公司	10.1
2	CN202310150695.0	一种音频编解码方法、装置、设备及介质	腾讯科技(深圳有限公司)	5.1.3,9.1.2,9.1.3

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。

T/AI 109.7-2024

信息技术 智能媒体编码 第7部分：音频封装与传输

1 范围

本文件规定了符合 AVS3 智能媒体编码音频数据的基本文件封装格式、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；符合 AVS2 的高效多媒体编码音频数据的基本文件封装格式、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。

2 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 33475.3-2018 信息技术 高效多媒体编码 第3部分：音频

注：在本文件中 GB/T 33475.3-2018 简称为 AVS2 音频

GB/T 18793-2002 信息技术 可扩展置标语言(Extensible Markup Language (XML)) 1.0

T/AI 109.3-2023 信息技术 智能媒体编码 第3部分：沉浸式音频

注：在本文件中 T/AI 109.3 简称为 AVS3 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 13818-1:2023 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 14496-12:2022 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 23000-19:2024 信息技术 多媒体应用格式 第19部分：片段媒体的通用媒体应用格式 (Information technology -- Multimedia application format (MPEG-A) -- Part 19: Common media application format (CMAF) for segmented media)

ISO/IEC 23009-1:2022 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 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)

3 术语和定义

下列术语和定义适用于本文件。

3.1

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。

3.2

AVS3音频编码位流 AVS3 audio bitstream

符合T/AI 109.3的编码音频信号所形成的二进制数据流。

3.3

AVS2音频编码位流 AVS2 audio bitstream

符合GB/T 33475.3-2018的编码音频信号所形成的二进制数据流。

3.4

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）。

3.5

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。

3.6

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。

3.7

初始化片段 initialization segment

包含有媒体流解码所必需元数据的片段。

3.8

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。

3.9

轨道 track

文件中一系列相关样本的集合。

3.10

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。

3.11

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。

3.12

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。

3.13

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。

3.14

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。

3.15

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。

3.16

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。

4 缩略语

下列缩略语适用于本文件。

AVS	音视频编解码标准	(Audio and Video coding Standard)
AVS2	第二代音视频编解码标准	(the second generation AVS standard)
AVS3	第三代音视频编解码标准	(the third generation AVS standard)
AASF	AVS 音频存储格式	(AVS Audio Storage Format)
AATF	AVS 音频传输格式	(AVS Audio Transport Format)
CMAF	通用媒体应用格式	(Common Media Application Format)
DASH	基于 HTTP 的动态自适应流媒体	(Dynamic Adaptive Streaming over HTTP)
DTS	解码时间戳	(Decoding Time-Stamp)
HOA	高阶立体声场信号	(Higher Order Ambisonics)

HTTP	超文本传输协议	(HyperText Transfer Protocol)
ISO BMFF	ISO 基本媒体文件格式	(ISO Base Media File Format)
MIME	多用途互联网邮件扩展类型	(Multipurpose Internet Mail Extensions)
MP	媒体呈现	(Media Presentation)
MPD	媒体呈现描述	(Media Presentation Description)
MTU	最大传输单元	(Maximum Transmission Unit)
PES	分组化基本流	(Packetized Elementary Stream)
PMT	节目映射表	(Program Map Table)
PS	节目流	(Program Stream)
PSI	节目特定信息	(Program-Specific Information)
RTP	实时传输协议	(Real-time Transport Protocol)
SAP	流访问点	(Stream Access Point)
SDP	会话描述协议	(Session Description Protocol)
SMT	智能媒体传输	(Smart Media Transport)
STD	系统目标解码器	(System Target Decoder)
TS	传输流	(Transport Stream)
T-STD	传输系统目标解码器	(Transport System Target Decoder)
URI	统一资源标识符	(Uniform Resource Identifier)
URN	统一资源名称	(Uniform Resource Name)
UTC	协调世界时	(Coordinated Universal Time)
XML	可扩展置标语言	(eXtensible Markup Language)

5 音频基本文件封装格式

5.1 AVS3 音频的基本文件封装格式

5.1.1 AVS3 音频的基本流定义

5.1.1.1 通则

本文件对符合 T/AI 109.3 规范的 AVS3 音频的基本流结构和基本流格式进行定义。

5.1.1.2 AVS3 音频编码系统

本文件规定的符合 T/AI 109.3 的 AVS3 音频编码系统包括：声道信号编码系统、对象信号编码系统、HOA 信号编码系统、元数据编码系统。

5.1.1.3 基本流结构

本文件规定的符合 T/AI 109.3 的 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，对应的 `audio_codec_id` 分别为 2、0、1。

5.1.1.4 基本流格式

本文件规定的符合 T/AI 109.3 的 AVS3 音频基本流格式分为两种：存储格式 AASF 和传输格式 AATF。

5.1.2 AVS3 音频配置信息

5.1.2.1 AVS3 音频通用全码率音频编码特有配置

5.1.2.1.1 定义

本条定义符合 T/AI 109.3 的通用全码率音频编码内容的特有配置 Avs3AudioGASpecificConfig 的语法和语义。

5.1.2.1.2 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) number_objects;
        unsigned int(1) reserved;
    } else if(content_type==2) {
        unsigned int(7) channel_number_index;
        unsigned int(1) reserved;
        unsigned int(7) number_objects;
        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;
    }
}
```

5.1.2.1.3 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。

channel_number_index: 应符合 T/AI 109.3的附录 A。

number_objects: 表示音频对象数量, 等于 T/AI 109.3的附录 A 中 object_channel_number +1。

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。

resolution: 应符合T/AI 109.3的附录A

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	

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_per_channel 对应的比特率×number_objects
2	声道信号的 bitrate_index 对应的比特率+对象信号的 bitrate_index_per_channel 对应的比特率×number_objects
3	HOA 信号的 bitrate_index 对应的比特率
4-15	保留

表1中的coding_profile和soundBedType, 以及表2中的bitrate_index和bitrate_index_per_channel应符合 T/AI 109.3的附录A。

5.1.2.2 AVS3 音频通用高码率音频编码特有配置

5.1.2.2.1 定义

本条定义符合T/AI 109.3的通用高码率音频编码内容的特有配置Avs3AudioGHSpecificConfig的语法和语义。

5.1.2.2.2 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;
}

```

5.1.2.2.3 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。

5.1.2.3 AVS3 音频无损音频编码特有配置

5.1.2.3.1 定义

本条定义符合 T/AI 109.3的音频无损音频编码内容的特有配置 Avs3AudioLLSpecificConfig 的语法和语义。

5.1.2.3.2 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;
}

```

5.1.2.3.3 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。

5.1.3 ISO 基本媒体文件格式扩展

5.1.3.1 AVS3 音频解码器配置数据盒

5.1.3.1.1 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

本条根据5.1.2中定义的音频编码特有配置,定义了 AVS3音频解码器配置数据盒 CA3SpecificBox 的语法和语义。

5.1.3.1.2 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;
    }
}
```

5.1.3.1.3 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。

5.1.3.2 AVS3 音频样本入口

5.1.3.2.1 定义

样本入口类型: 'av3a'或'a3as'

容器: Sample Description Box ('stds')

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

本条定义了AVS3音频样本入口的语法和语义。其中，

对于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。

对于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。

5.1.3.2.2 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;
}
```

5.1.3.2.3 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。

5.1.3.3 AVS3 音频样本格式

本条定义了 AVS3音频样本格式，其中，

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。

5.2 AVS2 音频的基本文件封装格式

5.2.1 AVS2 音频基本流定义

5.2.1.1 通则

本文件对符合 GB/T 33475.3-2018规范的 AVS2音频基本流结构和基本流格式进行定义。

5.2.1.2 AVS2 音频编码系统

本文件规定的符合 GB/T 33475.3-2018的 AVS2音频编码系统包括：声道信号编码系统、对象信号编码系统和元数据编码系统。

5.2.1.3 基本流结构

本文件规定的符合 GB/T 33475.3-2018的 AVS2音频的基本流结构包括：通用音频编码位流和无损音频编码位流，对应的 audio_codec_id 分别为0、1。

5.2.1.4 基本流格式

本文件规定的符合 GB/T 33475.3-2018的 AVS2音频基本流格式分为两种：存储格式 AASF 和传输格式 AATF。

5.2.2 AVS2 音频配置信息

5.2.2.1 AVS2 音频通用音频编码特有配置

5.2.2.1.1 定义

本条定义符合 GB/T 33475.3-2018的通用音频编码内容的特有配置 AVSAGASpecificConfig 的语法和语义。

5.2.2.1.2 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();  
    }  
}
```

5.2.2.1.3 语义

sampling_frequency_index：应符合 GB/T 33475.3-2018的附录 A。

anc_data_index：应符合 GB/T 33475.3-2018的附录 A。

coding_profile：应符合 GB/T 33475.3-2018的附录 A。

bitstream_type：应符合 GB/T 33475.3-2018的附录 A。

channel_number_index：应符合 GB/T 33475.3-2018的附录 A。

bitrate_index：应符合 GB/T 33475.3-2018的附录 A。

raw_frame_length: 应符合 GB/T 33475.3-2018的附录 A。

resolution: 应符合 GB/T 33475.3-2018的附录 A。

anc_data_block(): 应符合 GB/T 33475.3-2018的附录 A。

5.2.2.2 AVS2 音频无损音频编码特有配置

5.2.2.2.1 定义

本条定义符合 GB/T 33475.3-2018的无损音频编码内容的特有配置 AVSALLSpecificConfig 的语法和语义。

5.2.2.2.2 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();
    }
}
```

5.2.2.2.3 语义

sampling_frequency_index: 应符合 GB/T 33475.3-2018的附录 A。

sampling_frequency: 应符合 GB/T 33475.3-2018的附录 A。

anc_data_index: 应符合 GB/T 33475.3-2018的附录 A。

coding_profile: 应符合 GB/T 33475.3-2018的附录 A。

channel_number: 应符合 GB/T 33475.3-2018的附录 A。

resolution: 应符合 GB/T 33475.3-2018的附录 A。

anc_data_block(): 应符合 GB/T 33475.3-2018的附录 A。

5.2.3 ISO 基本媒体文件格式扩展

5.2.3.1 AVS2 音频解码器配置数据盒

5.2.3.1.1 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

本条根据5.2.2中定义的音频编码特有配置，定义了 AVS2音频解码器配置数据盒 AVSASpecificBox 的语法和语义。

5.2.3.1.2 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();
}
```

5.2.3.1.3 语义

AVSAGASpecificConfig 在5.2.2.1中定义， AVSALLSpecificConfig 在5.2.2.25.1.2.2中定义。

audio_codec_id: 应符合 GB/T 33475.3-2018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 33475.3-2018。

5.2.3.2 AVS2 音频样本入口

5.2.3.2.1 定义

样本入口类型: 'cavs'或'a2as'

容器: Sample Description Box ('stds')

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

本条定义了AVS2音频样本入口的语法和语义。其中，

对于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。

对于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。

5.2.3.2.2 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;
}

```

5.2.3.2.3 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 33475.3-2018的附录 A。

5.2.3.3 AVS2 音频样本格式

本条定义了 AVS2音频样本格式，其中，

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 33475.3-2018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 33475.3-2018的附录 A。

6 CMAF 轨道和媒体配置

6.1 AVS3 音频 CMAF 轨道和媒体配置

6.1.1 通则

AVS3音频CMAF轨道格式应符合ISO/IEC 23000-19:2024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'。

6.1.2 AVS3 音频 CMAF 轨道约束

6.1.2.1 通则

任何符合AVS3音频媒体配置文件的CMAF轨道都应符合：

- a) 本文件5.1.3中定义的AVS3音频轨道；
- b) ISO/IEC 23000-19:2024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()。

6.1.2.2 样本描述数据盒 ('stsd')

AVS3音频轨道中的SampleDescriptionBox应包含符合ISO/IEC 14496-12:2022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。

6.1.3 AVS3 音频 CMAF 切换集约束

6.1.3.1 通则

对于符合AVS3音频媒体配置文件的CMAF切换集，适用以下约束：

- a) 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；
- b) CMAF切换集中的每个CMAF轨道应符合ISO/IEC 23000-19:2024中7.3.4规定的通用CMAF切换集约束要求；
- c) 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。

6.1.3.2 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：

- a) 应符合ISO/IEC 23000-19:2024中7.3.4规定的通用CMAF切换集约束要求；
- b) 应符合ISO/IEC 23000-19:2024中7.3.4.2规定的通用单一初始化约束；
- c) CMAF头部中的音频样本的audio_codec_id应保持不变。

6.1.4 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。

6.2 AVS2 音频 CMAF 轨道和媒体配置

6.2.1 通则

AVS2音频CMAF轨道格式应该符合ISO/IEC 23000-19:2024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’。

6.2.2 AVS2 音频 CMAF 轨道约束

6.2.2.1 通则

任何符合 AVS2音频媒体配置文件的CMAF轨道都应符合：

- a) 本文件5.2.3中定义的AVS2音频轨道；
- b) ISO/IEC 23000-19:2024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()。

6.2.2.2 样本描述数据盒 ('stsd')

AVS2音频轨道中的SampleDescriptionBox应包含符合ISO/IEC 14496-12:2022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。

6.2.3 AVS2 音频 CMAF 切换集约束

6.2.3.1 通则

对于符合AVS2音频媒体配置文件的CMAF切换集，适用以下约束：

- a) 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；
- b) CMAF切换集中的每个CMAF轨道应符合 ISO/IEC 23000-19:2024中7.3.4规定的通用 CMAF切换集约束要求；
- c) 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。

6.2.3.2 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：

- a) 应符合ISO/IEC 23000-19:2024的7.3.4中规定的通用CMAF切换集约束要求；
- b) 应符合ISO/IEC 23000-19:2024的7.3.4.2中规定的通用单一初始化约束；
- c) CMAF头部中的音频样本的audio_codec_id应保持不变。

6.2.4 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。

7 DASH 传输技术要求

7.1 AVS3 音频 DASH 传输技术要求

7.1.1 通则

本条规定符合T/AI 109.3的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 23009-1:2022）进行传输时的媒体呈现描述（MPD）与片段格式。

7.1.2 DASH 片段格式

7.1.2.1 通则

DASH 片段格式应符合第 5 章规定的 AVS3 音频的基本文件封装格式，其样本入口类型为'av3a'。

7.1.2.2 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。

7.1.2.3 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：

- a) 每个 AVS3 音频样本只能包含一帧且只有一帧 `aatf_frame()`；
- b) 每个 AVS3 音频样本的 `audio_codec_id` 应保持不变；
- c) 每个 AVS3 音频样本的 SAP 的类型，在 ISO/IEC 14496-12:2022 附录 I 定义，值都为 1。

7.1.2.4 索引片段

AVS3 音频索引片段应满足以下约束：

- a) 每个子片段由一个 ISO/IEC 14496-12:2022 中 8.16.3 中定义的 'sidx' 类型的片段索引数据盒索引；
- b) 'sidx' 数据盒指示引用的子片段的 `starts_with_SAP` 为 1，`SAP_type` 为 1。

7.1.3 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。

7.1.4 DASH MPD 描述子

7.1.4.1 通则

本条定义了媒体呈现描述（MPD）中用于描述 AVS3 音频的一些新的 XML (GB/T 18793-2002) 元素和属性，及命名空间和规则。

7.1.4.2 @mimeType 属性

`@mimeType` 属性应设置为 "audio/mp4"。

7.1.4.3 @audioSampleRate 属性

音频采样率的属性源于 CA3SpecificBox 的 `sampling_frequency_index`。

7.1.4.4 @startWithSAP 属性

`@startWithSAP` 属性应该设置成 1。

7.1.4.5 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：

- a) 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；
- b) 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;
- c) 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;

7.2 AVS2 音频 DASH 传输技术要求

7.2.1 通则

本条规定符合GB/T 33475.3-2018的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 23009-1:2022) 进行传输时的媒体呈现描述 (MPD) 与片段格式。

7.2.2 DASH 片段格式

7.2.2.1 通则

DASH 片段格式应符合本文件第 5 章规定的 AVS2 音频的基本文件封装格式, 其样本入口类型为 'cavs'。

7.2.2.2 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。

7.2.2.3 媒体片段

每个 DASH 媒体片段应包含一个或多个 GB/T 33475.3-2018 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:

- a) 每个 AVS2 音频样本只能包含一帧且只有一帧aatf_frame();
- b) 每个 AVS2 音频样本的audio_codec_id应保持不变;
- c) 每个 AVS2 音频样本的SAP的类型, 在ISO/IEC 14496-12:2022附录I定义, 值都为1。

7.2.2.4 索引片段

AVS2 音频索引片段应满足以下约束:

- a) 每个子片段由一个ISO/IEC 14496-12:2022中8.16.3定义的'sidx'类型的片段索引数据盒索引;
- b) 'sidx'数据盒指示引用的子片段的starts_with_SAP为1, SAP_type为1。

7.2.3 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。

7.2.4 DASH MPD 描述子

7.2.4.1 通则

本条定义了媒体呈现描述（MPD）中用于描述 AVS2 音频的一些新的 XML (GB/T 18793-2002)元素和属性，及命名空间和规则。

7.2.4.2 @mimeType 属性

@mimeType 属性应设置为"audio/mp4"。

7.2.4.3 @audioSampleRate 属性

音频采样率的属性源于 AVSASpecificBox 的 sampling_frequency_index。

7.2.4.4 @startWithSAP 属性

@startWithSAP 属性应该设置成 1。

7.2.4.5 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：

- a) 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；
- b) 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；

8 传输流和节目流技术要求

8.1 AVS3 音频传输流和节目流技术要求

8.1.1 通则

本条规定符合 T/AI 109.3 的 AVS3 音频编码位流的在传输流中封装格式与传输要求。

AVS3 音频编码位流的传输流应满足以下约束：

- a) AVS3 音频流应是 ISO/IEC 13818-1:2023 中节目的一个节目元素，基本流的 stream_type 字段值应等于 '0xD5'；
- b) AVS3 音频使用 AATF 的封装格式，即封装成 aatf_frame()；
- c) 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。

8.1.2 PES 分组的流标识

AVS3音频应作为PES_packet_data_bytes携带在PES分组数据包中，并通过PMT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：

- a) PES分组的stream_id取值'1111 1101'（extended_stream_id）；
- b) 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；
- c) 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；
- d) 一个PES包可以包含一帧或多帧aatf_frame()。

8.1.3 节目和节目元素描述符

8.1.3.1 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 13818-1中已规定为0x05。

本条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。

8.1.3.2 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。

8.1.3.3 注册描述符中各字段语义定义

registration_descriptor()的定义应参考ISO/IEC 13818-1:2023，AVS3音频的format_identifier应为0x41-56-53-41('AVSA')。

8.1.3.4 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) {		
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) {		

表 4 (续)

语 法	位 数	助 记 符
<code>hoa_order</code>	4	<code>uimsbf</code>
<code>reserved</code>	4	<code>bslbf</code>
<code>}</code>		
<code>total_bitrate</code>	16	<code>uimsbf</code>
<code>}</code>		
<code>resolution</code>	2	<code>uimsbf</code>
<code>reserved</code>	6	<code>bslbf</code>
<code>for (i=0; i<N; i++) {</code>		
<code> addition_info[i]</code>	8	<code>bslbf</code>
<code>}</code>		
<code>}</code>		

8.1.3.5 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。

`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。

`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。

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	

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_per_channel 对应的比特率 \times (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_per_channel 对应的比特率 \times (object_channel_number + 1)
3	HOA 信号的 bitrate_index 对应的比特率
4-15	保留

表5中的coding_profile和soundBedType,以及表6中的bitrate_index和bitrate_index_per_channel应符合T/AI 109.3的附录A

8.1.4 AVS3 音频 T-STD 模型拓展

本条定义了 AVS3 音频传输流的 T-STD 模型扩展。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。

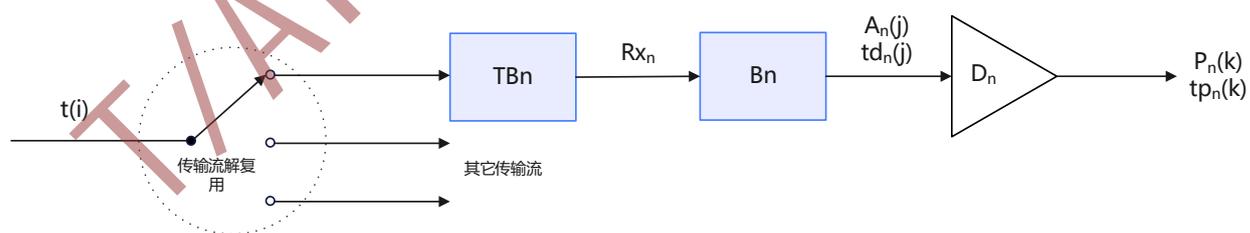


图1 面向 AVS3 音频的 T-STD 模型拓展

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。

8.1.4.1 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 R_{x_n} 计算公式如下:

$$R_{x_n} = 1.2 \times R_{\max} \times N$$

式中:

R_{\max} ——AVS3 音频单通道最大速率;

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 R_{x_n} 取值, 见表 7。

表 7 R_{x_n} 取值范围

通道数	R_{x_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:

$$BS_n = BS_{\text{mux}} + BS_{\text{dec}} + BS_{\text{oh}}$$

式中:

BS_{mux} ——额外复用缓存, 计算公式如下:

$$BS_{\text{mux}} = 0.004 \times R_{\max} \times N$$

BS_{dec} ——基本流存储单元缓存, 取值为 6144 bits;

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	BS_n [bytes]
1-8	7440
9-16	14 352
17-48	42 000
49-128	111 120

8.1.4.2 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 $td_n(j) - t(i) \leq 10$ (s)。

8.1.4.3 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。

8.2 AVS2 音频传输流和节目流技术要求

8.2.1 通则

本条规定符合GB/T 33475.3-2018的AVS2音频编码位流在传输流中封装格式与传输要求。

AVS2音频编码位流的传输流应满足以下约束：

- a) AVS2音频流应是ISO/IEC 13818-1:2023中节目的一个节目元素，基本流的stream_type字段值应等于'0xD3'；
- b) AVS2音频使用AATF的封装格式，即封装成aatf_frame()；
- c) 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。

8.2.2 PES 分组的流标识

AVS2音频应作为PES_packet_data_bytes携带在PES分组数据包中，并通过PMT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：

- a) PES分组的stream_id取值'1011 1101' (private_stream_1)。
- b) 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。
- c) 一个PES包可以包含一帧或多帧aatf_frame()。

8.2.3 节目和节目元素描述符

8.2.3.1 节目和节目元素描述符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 13818-1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。

8.2.3.2 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。

8.2.3.3 注册描述符中各字段语义定义

registration_descriptor()的定义应参考ISO/IEC 13818-1:2023, AVS2音频的format_identifier应为0x41-56-53-41('AVSA')。

8.2.3.4 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
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) {		

表 10 (续)

语 法	位 数	助 记 符
<code>language</code>	24	<code>bslbf</code>
<code>}</code>		
<code>if (avs_version_flag == 1) {</code>		
<code> audio_codec_id</code>	4	<code>uimsbf</code>
<code> anc_data_index</code>	1	<code>bslbf</code>
<code> coding_profile</code>	3	<code>uimsbf</code>
<code> if (audio_codec_id==0) {</code>		
<code> bitrate_index</code>	4	<code>uimsbf</code>
<code> bitstream_type</code>	1	<code>bslbf</code>
<code> reserved</code>	3	<code>bslbf</code>
<code> raw_frame_length</code>	16	<code>uimsbf</code>
<code> }</code>		
<code> resolution</code>	2	<code>uimsbf</code>
<code> reserved</code>	6	<code>bslbf</code>
<code> if (anc_data_index) {</code>		
<code> anc_data_block()</code>		
<code> }</code>		
<code>}</code>		
<code>for (i=0; i<N; i++) {</code>		
<code> addition_info[i]</code>	8	<code>bslbf</code>
<code>}</code>		
<code>}</code>		

8.2.3.5 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`的值为1, 表示音频流中包含文本; 若`text_present_flag`的值为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** 的值为1, 表示后面包含语言字段; 若 **language_present_flag** 的值为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。

8.2.4 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:

$$R_{Xn} = 1.2 \times R_{\max} \times N \text{ bps}$$

式中:

R_{\max} 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 ——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 R_{Xn} 的取值, 见表11。

表 11 R_{Xn} 取值范围

通道数	R_{Xn} [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 BS_n 大小定义为:

$$BS_n = BS_{\text{mux}} + BS_{\text{dec}} + BS_{\text{oh}}$$

式中:

BS_{mux} ——额外复用缓存, 计算公式如下:

$$BS_{\text{mux}} = 0.004 \times R_{\max} \times N;$$

BS_{dec} ——基本流存储单元缓存, **audio_codec_id** 为0时, $6144\text{bits} \times N$ 。

BS_{oh} ——PES头大小，528字节。

其中 R_{max} 和N的定义同上。

AVS2音频T-STD模型扩展中对应 BS_n 的取值，见表12。

表 12 BS_n 取值范围

通道数	BS_n [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

8.2.5 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。

9 SMT 传输技术要求

9.1 AVS3 音频 SMT 传输技术要求

9.1.1 通则

本条规定符合T/AI 109.3的AVS3音频编码位流基于智能媒体传输协议SMT（T/AI 114-2021）进行传输的技术要求，应符合如下约束：

- 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；
- 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；
- 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章的扩展定义。

9.1.2 AVS 音频媒体资源描述符

9.1.2.1 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。

9.1.2.2 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_tag		16	uimsbf
descriptor_length		16	uimsbf
audio_format_type		4	uimsbf
audio_codec_id		4	uimsbf
coding_profile		3	uimsbf
average_bitrate_flag		1	bslbf
hoa_order_flag		1	bslbf
channel_number_flag		1	bslbf
object_info_flag		1	bslbf
reserved	'1'	1	uimsbf
if(average_bitrate_flag==1)			
average_bitrate		16	uimsbf
else {			
max_bitrate		16	uimsbf
min_bitrate		16	uimsbf
}			
if(hoa_order_flag){			
max_hoa_order		8	uimsbf
}			
if(channel_number_flag){			
max_channel_number		8	uimsbf
}			
if(object_info_flag){			
max_object_channel_number		8	uimsbf
}			
bit_depth_resolution		8	uimsbf
sample_rate		24	uimsbf
}			

9.1.2.3 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。

9.1.3 交互反馈信令表

9.1.3.1 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:

- a) **table_type**应取值为3;
- b) **asset_group_flag**应取值为0。

9.1.3.2 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_id		8	uimsbf
version		8	uimsbf
length		16	uimsbf
table_payload {			
table_type		8	uimsbf
timestamp		32	uimsbf
message_source		1	bslbf
asset_group_flag		1	uimsbf
reserved		6	uimsbf
if(asset_group_flag){			
asset_group_id		8	
}			
else{			
asset_id()			
}			
if(table_type == 3){			
coordinate_type		8	uimsbf
if(coordinate_type == 0){			
ClientPosition()			
}			
if(coordinate_type == 1){			
azimuth		8	uimsbf
elevation		8	uimsbf
distance		8	uimsbf
}			
}			
}			
}			

9.1.3.3 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 $(-2^{15}, 2^{15} - 1)$ ，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 $(-2^{15}, 2^{15} - 1)$ ，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 $(-2^{15}, 2^{15} - 1)$ ，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。

9.2 AVS2 音频 SMT 传输技术要求

9.2.1 通则

本条规定符合GB/T 33475.3-2018的AVS2音频编码位流基于智能媒体传输协议SMT(T/AI 114-2021)进行传输的技术要求，应符合如下约束：

- 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；
- 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；
- 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章的扩展定义。

9.2.2 AVS 音频媒体资源描述符

9.2.2.1 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。

9.2.2.2 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_tag		16	uimsbf
descriptor_length		16	uimsbf
audio_format_type		4	uimsbf
audio_codec_id		4	uimsbf
coding_profile		3	uimsbf
average_bitrate_flag		1	bslbf
hoa_order_flag		1	bslbf
channel_number_flag		1	bslbf
object_info_flag		1	bslbf
reserved	'1'	5	uimsbf
if(average_bitrate_flag==1)			
average_bitrate		16	uimsbf
else {			
max_bitrate		16	uimsbf
min_bitrate		16	uimsbf
}			
if(hoa_order_flag){			
max_hoa_order		8	uimsbf
}			
if(channel_number_flag){			
max_channel_number		8	uimsbf
}			
if(object_info_flag){			
max_object_channel_number		8	uimsbf
}			
bit_depth_resolution		8	uimsbf
sample_rate		24	uimsbf
}			

9.2.2.3 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。

9.2.3 交互反馈信令表

9.2.3.1 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:

- a) **table_type**应取值为3;
- b) **asset_group_flag**应取值为0。

9.2.3.2 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_id		8	uimsbf
version		8	uimsbf
length		16	uimsbf
table_payload {			
table_type		8	uimsbf
timestamp		32	uimsbf
message_source		1	bslbf
asset_group_flag		1	uimsbf
reserved		6	uimsbf
if(asset_group_flag){			
asset_group_id		8	
}			
else{			
asset_id()			
}			
if(table_type == 3){			
coordinate_type		8	uimsbf
if(coordinate_type == 0){			
ClientPosition()			
}			
if(coordinate_type == 1){			
azimuth		8	uimsbf
elevation		8	uimsbf
distance		8	uimsbf
}			
}			
}			
}			

9.2.3.3 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 $(-2^{15}, 2^{15} - 1)$ ，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 $(-2^{15}, 2^{15} - 1)$ ，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 $(-2^{15}, 2^{15} - 1)$ ，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。

10 RTP 传输技术要求

10.1 AVS3 音频 RTP 传输技术要求

10.1.1 RTP 负载

10.1.1.1 通则

本条规定符合T/AI 109.3的AVS3音频编码位流基于RTP协议进行传输的技术要求。RTP协议将符合T/AI 109.3的AVS3音频编码位流封装为若干个RTP包，每个RTP包由RTP头和RTP负载组成。

10.1.1.2 封包规则

RTP负载封包规则应当符合以下任何一个约束：

- a) RTP负载携带的AVS3音频编码格式为aatf_frame();
- b) 如果一个aatf_frame()的长度超过了MTU，需要参考RFC 3550的6.1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。

10.1.2 RTP 头

10.1.2.1 通则

RFC3550 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 10.2.2 进行了扩展定义。

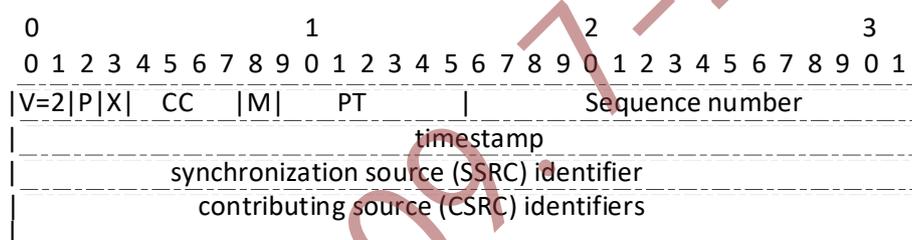


图 2 RTP 头结构

10.1.2.2 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。

10.1.3 SDP 定义

10.1.3.1 通则

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 10.1.3.2 中的字段进行扩展定义。

10.1.3.2 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。

10.1.4 MIME 类型

10.1.4.1 通则

当使用RFC 6381定义的MIME类型的'codecs'参数时，如果 MIME类型对应本文件中定义的文件格式，'codecs'参数值的样本入口类型应为'av3a'。

10.1.4.2 MIME 参数定义

MIME 媒体名字: audio

MIME 子名字: AV3A-AATF

必需参数:

a) **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:

b) **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的值为1)，则codec-nn-id=0x0201。

c) **config**: 16进制的字符串，对应5.2.2中定义的CA3SpecificBox；

d) **bitrate**: 音频流的码率，对应5.2.2中定义的total_bitrate。

10.2 AVS2 音频 RTP 传输技术要求

10.2.1 RTP 负载

10.2.1.1 通则

本条规定符合GB/T 33475.3-2018的AVS2音频编码位流基于RTP协议进行传输的技术要求。RTP协议将符合GB/T 33475.3-2018的AVS2音频编码位流封装为若干个RTP包，每个RTP包由RTP头和RTP负载组成。

10.2.1.2 封包规则

RTP负载封包规则应当符合以下任何一个约束：

- a) RTP负载携带的AVS2音频编码格式为aatf_frame()；
- b) 如果一个aatf_frame()的长度超过了MTU，需要参考RFC 3550的6.1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。

10.2.2 RTP 头

10.2.2.1 通则

RFC3550 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 10.2.2.2 进行了扩展定义。

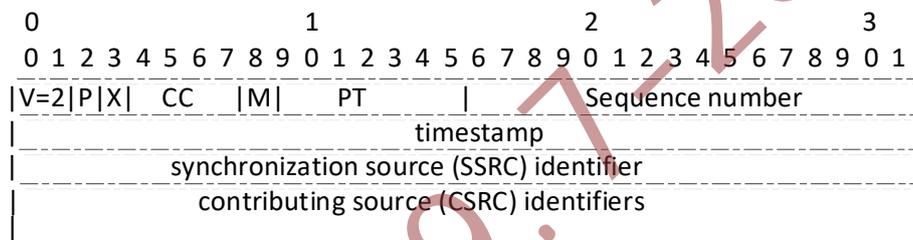


图 3 RTP 头结构

10.2.2.2 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。

10.2.3 SDP 定义

10.2.3.1 通则

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 10.2.3.2 中的字段进行扩展定义。

10.2.3.2 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。

10.2.4 MIME 类型

10.2.4.1 通则

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'参数值的样本入口类型为'cavs'。

10.2.4.2 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。

附 录 A
(规范性)
MIME 类型的'codecs'参数

A.1 通则

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

- a) 且'codecs'参数值起始于5.1.1定义的样本入口类型,则'codecs'参数的子参数应符合附录A.2;
- b) 且'codecs'参数值起始于5.1.2定义的样本入口类型,则'codecs'参数的子参数应符合附录A.3。

A.2 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。

A.3 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。