

ICS 33.160.25
CCS M 74

世界超高清视频产业联盟标准

T/UWA 009.2-2-2025

三维声技术规范 第 2-2 部分：应用指南 媒体格式

3D Audio technology specification Part 2-2:

application guide - media format

2025-06-11 发布

2025-06-11 实施

世界超高清视频产业联盟

目 次

目次	I
前言	III
引言	IV
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 Audio Vivid 文件格式	3
5.1 Audio Vivid 基本流定义	3
5.2 Audio Vivid 配置信息	4
5.3 ISO 基本媒体文件格式扩展	6
6 CMAF 轨道和媒体配置	7
6.1 通则	7
6.2 CMAF 轨道约束	7
6.3 CMAF 切换集约束	7
6.4 音频编解码参数	8
7 DASH 传输	8
7.1 概述	8
7.2 DASH 片段格式	8
7.3 DASH MPD 编码器参数	8
7.4 DASH MPD 描述符	9
8 传输流和节目流	9
8.1 通则	9
8.2 PES 分组流标识	10
8.3 节目和节目元素描述符	10
8.4 Audio Vivid T-STD 模型拓展	12
9 SMT 传输	14
9.1 通则	14
9.2 Audio Vivid 媒体资源描述符	14
9.3 交互反馈信令表	16
10 RTP 传输	18
10.1 RTP 负载	18
10.2 RTP 头	18

T/UWA 009.2-2—2025	
10.3 SDP 定义	19
10.4 MIME 类型	19
11 RTMP 传输	20
11.1 概述	20
11.2 AUDIODATA 定义	20
11.3 AudioTagBody 定义	20
11.4 AVS3AUDIODATA 定义	21
12 HLS 传输	21
12.1 概述	21
12.2 切片文件格式	21
12.3 HLS M3U8 扩展	21
参考文献	23

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准文件的结构和起草规则》给出的规则起草。

本文件是T/UWA 009《三维声技术规范》的第2-2部分，T/UWA 009已经发布了以下部分：

- 第1部分：编码、分发与呈现；
- 第3-1部分：技术要求和测试方法 家庭影音播放设备；
- 第3-2部分：技术要求和测试方法 便携式数字设备；
- 第3-3部分：技术要求和测试方法 超高清机顶盒；
- 第3-4部分：技术要求和测试方法 车载音频系统；
- 第3-5部分：技术要求和测试方法 菁彩声混音棚。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由世界超高清视频产业联盟提出并归口。

本文件起草单位：中国电子技术标准化研究院、华为技术有限公司、咪咕文化科技有限公司、上海交通大学、杭州当虹科技有限公司、深圳腾讯计算机系统有限公司、深圳创维显示技术有限公司、北京数字电视国家工程实验室有限公司、广东博华超高清创新中心有限公司、深圳市奥拓电子股份有限公司、康佳集团股份有限公司、深圳洲明科技股份有限公司。

本文件主要起草人：李婧欣、刘帅、韩建、徐异凌、王雪辉、李大龙、徐遥令、李思远、彭辉、梁锋、苏运全、李永杰、夏丙寅、朱梦尧、李康敬、陈珊、沈思康、张曼华、赵霞飞、白莹杰。

引 言

三维声技术规范旨在规范三维声音频的编码、分发与呈现方法，为三维声全流程中音频制作、元数据、渲染、控制、终端设备重放等关键环节提供指导，拟由八个部分构成。

——第1部分：编码、分发与呈现。目的在于规定三维声音频的编码、分发与呈现方法，兼容单声道、双声道立体声、多声道环绕声等声音格式。

——第2-1部分：应用指南 系统集成。目的在于帮助应用者构建符合第1部分标准要求的系统，实现互联互通。

——第2-2部分：应用指南 媒体格式。目的在于帮助应用者构建符合第1部分要求的媒体格式，以实现音视频录制、传输和播放等应用中三维声媒体文件的生成、接收与验证。

——第3-1部分：技术要求和测试方法 家庭影音播放设备。目的在于指导应用者设计、生产和校验符合第1部分要求的三维声家庭影音播放终端产品。

——第3-2部分：技术要求和测试方法 便携式数字设备。目的在于指导应用者设计、生产和校验符合第1部分要求的三维声便携数字终端产品。

——第3-3部分：技术要求和测试方法 超高清机顶盒。目的在于指导应用者设计、生产和校验符合第1部分要求的三维声超高清机顶盒。

——第3-4部分：技术要求和测试方法 车载音频系统。目的在于指导应用者设计、生产、验收、运行及维护符合第1部分要求的三维声车载音频系统及终端产品。

——第3-5部分：技术要求和测试方法 菁彩声混音棚。目的在于指导应用者制作符合第1部分要求的菁彩声音频内容，以及指导应用者设计、生产、验收、运行和维护符合第1部分要求的菁彩声制作系统。

三维声技术规范 第2-2部分：应用指南 媒体格式

1 范围

本文件规定了符合 T/UWA 009.1—2023 的三维声文件的封装格式、配置约束和传输要求。
本文件适用于音视频录制、播放等应用中三维声媒体文件的生成、接收与验证。

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 33475.6 信息技术 高效多媒体编码 第6部分：智能媒体传输

T/UWA 009.1—2023 三维声技术要求 第1部分：编码、分发与呈现

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 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)

ISO/IEC 23000-19: 2024 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)

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

用作数据编码表示有一定次序的一组二进制数据流。

3.2

Audio Vivid

T/UWA 009.1-2023规定的音频编码技术。

3.3

Audio Vivid编码位流 Audio Vivid bitstream

符合T/UWA 009.1—2023编码音频信号所形成的二进制数据流。

3.4

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。

3.5

初始化片段 initialization segment

包含有媒体流解码所必需元数据的片段。

3.6

表示 representation

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

3.7

轨道 track

文件中一系列相关样本的集合。

3.8

媒体呈现描述 media presentation description

用于提供流媒体服务规范化描述媒体呈现的文件。

3.9

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。

3.10

媒体片段 media segment

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

3.11

媒体资源 asset

任何与唯一标识符联系用作构建一个多媒体演示的多媒体数据实体。

3.12

样本 sample

在非提示轨道中，一个单独的音频帧、时间连续的一个音频帧序列或时间连续的一段压缩音频。

在提示轨道中，定义一个或多个流式分组构成的音频。

注：一个轨道中任何两个样本不能具有相同的时间戳。

3.13

切换集 switching set

同一媒体内容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。

CMAF: 通用媒体格式 (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)
 FLV: 闪存视频 (flash video)
 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 mark-up language)

5 Audio Vivid 文件格式

5.1 Audio Vivid 基本流定义

5.1.1 通则

Audio Vivid基本流即Audio Vivid编码位流, 应符合T/UWA 009.1—2023的规定。

5.1.2 Audio Vivid 音频编码特性

Audio Vivid编码系统支持声道信号编码、对象信号编码、HOA（包含FOA）信号编码、元数据编码。

Audio Vivid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、无损音频编码工具和元数据编码工具。

5.1.3 基本流结构

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

5.1.4 基本流格式

Audio Vivid 基本流格式 AATF，应符合 T/UWA 009.1—2023中附录 A 的规定。

5.2 Audio Vivid 配置信息

5.2.1 Audio Vivid 通用全码率音频编码特有配置

5.2.1.1 语法

Audio Vivid 通用全码率音频编码内容的特有配置应符合以下语法。

```
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_num_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_num_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;
    }
}
```

5.2.1.2 语义

sampling_frequency_index: 应符合 T/UWA 009.1—2023附录 A 的规定。

nn_type: 应符合 T/UWA 009.1—2023附录 A 的规定。

content_type: 表示音频内容类型, 见表1。

channel_number_index: 应符合 T/UWA 009.1—2023的附录 A。

number_objects: 表示音频对象数量, 为 T/UWA 009.1—2023的附录 A 中 object_channel_number +1。

hoa_order: 表示 HOA 阶数, 为 T/UWA 009.1—2023的附录 A 中 order+1。

total_bitrate: 表示总码率, 单位 kb/s, 根据 content_type 的值计算方式不同, 见表2。

resolution: 应符合 T/UWA 009.1—2023附录 A 的规定。

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	

其中, coding_profile和soundBedType应符合T/UWA 009.1—2023的附录A。

表 2 total_bitrate 配置表

content_type 值	total_bitrate 计算方式
0	T/UWA 009.1—2023 附录 A.2 中声道信号 bitrate_index 对应的比特率值
1	T/UWA 009.1 — 2023 附录 A.2 中对象信号 bitrate_index_per_channel 对应的比特率值 ×number_objects
2	T/UWA 009.1 — 2023 附录 A.2 中声道信号 bitrate_index 对应的比特率值 + 对象信号 bitrate_index_per_channel 对应的比特率值×number_objects
3	T/UWA 009.1—2023 附录 A.2 中 HOA 信号 bitrate_index 对应的比特率值
4-15	保留

5.2.2 Audio Vivid 无损音频编码特有配置

5.2.2.1 语法

Audio Vivid 无损音频编码内容的特有配置应符合以下语法。

```
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) {
```

T/UWA 009.2-2—2025

```
        bit(8*addition_info_length) addition_info;
    }
    unsigned int(2) reserved;
}
```

5.2.2.2 语义

sampling_frequency_index: 应符合 T/UWA 009.1—2023附录 A 的规定。

sampling_frequency: 应符合 T/UWA 009.1—2023附录 A 的规定。

anc_data_index: 应符合 T/UWA 009.1—2023附录 A 的规定, 取值应为0。

coding_profile: 应符合 T/UWA 009.1—2023附录 A 的规定。

channel_number: 应符合 T/UWA 009.1—2023附录 A 的规定。

resolution: 应符合 T/UWA 009.1—2023附录 A 的规定。

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。

5.3 ISO 基本媒体文件格式扩展

5.3.1 Audio Vivid 解码器配置数据盒

5.3.1.1 定义

数据盒类型: 'dca3'

容器: 'av3a'类型的样本入口

强制性: 强制包含于'av3a'类型的样本入口

数量: 一个

Audio Vivid 解码器配置数据盒 CA3SpecificBox 包含5.2中定义的音频编码特有配置。

5.3.1.2 语法

Audio Vivid 解码器配置数据盒应符合以下语法。

```
class CA3SpecificBox extends Box('dca3') {
    unsigned int(4) audio_codec_id;
    if (audio_codec_id == 2) {
        Avs3AudioGASpecificConfig() Avs3AudioGAConfig;
    }
    else if (audio_codec_id == 1) {
        Avs3AudioLLSpecificConfig() Avs3AudioLLConfig;
    }
}
```

5.3.1.3 语义

Avs3AudioSpecificGAConfig 在5.2.1中定义, Avs3AudioSpecificLLConfig 在5.2.2中定义。

audio_codec_id: 应符合 T/UWA 009.1—2023的附录 A。

5.3.2 Audio Vivid 样本入口

5.3.2.1 定义

样本入口类型: 'av3a'

容器: Sample Description Box ('stds')

强制性: 封装Audio Vivid编码位流的轨道必须包含一个'av3a'样本入口

数量: 一个

AATF类型的Audio Vivid编码位流在文件中应被存储为'av3a'类型的音频轨道，其轨道样本入口中应包含一个CA3SpecificBox数据盒。

5.3.2.2 语法

```
class AVS3ATSampleEntry() extends AudioSampleEntry ('av3a'){
    CA3SpecificBox    config
}
```

5.3.2.3 语义

CA3SpecificBox 提供 Audio Vivid 编码位流的解码配置信息。

CA3SpecificBox 提供并扩展了对 ChannelCount, SampleSize, SampleRate 的描述。

本文件规定解码器应忽略 AudioSampleEntry 中的 ChannelCount, SampleSize, SampleRate。

5.3.3 Audio Vivid 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/UWA 009.1—2023中附录 A 的规定。

6 CMAF 轨道和媒体配置

6.1 通则

Audio Vivid的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，同时应符合第5章中指定的Audio Vivid轨道格式以及第6章中的约束。

CMAF轨道的品牌标识定义为'ca3a'。

6.2 CMAF 轨道约束

6.2.1 通则

任何符合Audio Vivid媒体配置文件的CMAF轨道都应符合：

- a) 第5.3条中定义的Audio Vivid轨道。
- b) ISO/IEC 23000-19中定义的通用视频CMAF轨道格式，包括：
 - 每个 presentation 应对应一个唯一的 presentation_id；
 - 每个 Audio Vivid Sample 只能包含一帧且只有一帧 aatf_frame()。

6.2.2 样本描述数据盒 ('stds')

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

满足Audio Vivid特有数据盒 CMAF 轨道音频样本入口的语法和取值应符合第5.3条中定义的 AVS3ATSampleEntry ('av3a') 样本入口。

6.3 CMAF 切换集约束

6.3.1 通则

对于符合 Audio Vivid 媒体配置文件的 CMAF 切换集，适用以下约束：

- a) CMAF切换集中的每个CMAF轨道应符合6.2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中7.3.4规定的通用CMAF切换集约束要求。

c) 单一初始化Audio Vivid CMAF切换集应符合6.3.2中定义的约束。

6.3.2 单一初始化 CMAF 切换集约束

Audio VividCMAF切换集应符合如下单一初始化约束：

a) ISO/IEC 23000-19：2024中7.3.4规定的通用CMAF切换集约束要求。

b) 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保持不变。

6.4 音频编解码参数

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

Audio Vivid媒体配置的MIME类型的“编解码器”参数应符合本文件10.4.3的规定。

7 DASH 传输

7.1 概述

本章规定了Audio Vivid编码位流通过符合ISO/IEC 23009-1规定的“基于HTTP的动态自适应流媒体传输协议”进行传输时的MPD与片段格式。

7.2节定义了DASH片段格式，用于封装Audio Vivid数据的DASH片段格式应符合本文件第5章规定的Audio Vivid编码位流的文件格式，其样本入口类型应支持‘av3a’。

7.3节定义了Audio Vivid编码位流的MPD编码器参数。

7.4节定义并使用了一些新的XML元素和属性，并给出了其命名空间及规则。

7.2 DASH 片段格式

7.2.1 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。

7.2.2 媒体片段

每个DASH媒体片段应包含一个或多个T/UWA 009.1—2023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Audio Vivid样本只能包含一帧且只有一帧aatf_frame()；

b) 每个Audio Vivid样本的audio_codec_id保持不变；

c) 每个Audio Vivid样本的SAP的类型，在ISO/IEC 14496-12附录I定义取值为1；

7.2.3 索引片段

Audio Vivid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取值为1，SAP_type取值为1。

7.3 DASH MPD 编码器参数

Audio Vivid编码位流在MPD中的@codecs属性使用本文件10.4.3中定义的MIME类型'codecs'参

数。

7.4 DASH MPD 描述符

7.4.1 @mimeType 属性

@mimeType 属性应设置为"audio/mp4"。

7.4.2 @audioSampleRate 属性

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

7.4.3 @startWithSAP 属性

@startWithSAP 属性应设置成 1。

7.4.4 AudioChannelConfiguration 描述符

@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@codecs属性为'av3a.01', 则@value的属性值为:

- 第1个字节为0xF0;
- 第2个字节等于5.2.2.1中channel_number的值;
- 第3个字节为0;

b) 如果AVS3音频编码位流在MPD中@codecs属性为'av3a.02', 则@value的属性值为:

- 第1个字节的高4位为0xF, 低4位等于5.2.1.1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5.2.1.1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5.2.1.1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 5.2.1.1 中 object_channel_number + 1 的值; 否则, 第3个字节的值等于 0;

8 传输流和节目流

8.1 通则

本章规定了适用于Audio Vivid编码位流传输流的编码结构与参数。

Audio Vivid流应满足以下约束:

- a) Audio Vivid 流应是 ISO/IEC 13818-1 中的一个节目元素, 基本流的 stream_type 字段值应等于 '0xD5';
- b) Audio Vivid 使用 AATF 的封装格式, 即封装成 aatf_frame();
- c) Audio Vivid 流的常见编码参数, 如 audio_codec_id 应使用 Audio Vivid 流描述符标识。如果存在与 Audio Vivid 流相关联的 Audio Vivid 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 PES 分组流标识

Audio Vivid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。

Audio Vivid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 Audio Vivid。
- c) elementary stream 在 PES payload 是字节对齐，即 Audio Vivid 的首字节位于 PES payload 的首字节中。
- d) 一个 PES 包可包含一帧或多帧 aatf_frame()。

8.3 节目和节目元素描述符

8.3.1 注册描述符

PMT 中应包含 registration_descriptor()。registration_descriptor()的定义应符合 ISO/IEC 13818-1 的规定，其中，registration_descriptor()的 tag 值应为 0x05，Audio Vivid 的 format_identifier 应为 0x41-56-53-41('AVSA')

8.3.2 Audio Vivid 流描述符语法

AVS3_Audio_descriptor()位于 PMT 中的 ES_info_length 字段后面。其语法见表 3。

表 3 Audio Vivid 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==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

表3 (续)

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

8.3.3 Audio Vivid 流描述符中各字段的语义

descriptor_tag: Audio Vivid描述符的tag应为210(0xD2)。

descriptor_length: Audio Vivid描述符的长度。

audio_codec_id: 应符合T/UWA 009.1—2023的附录A。

anc_data_index: 应符合T/UWA 009.1—2023的附录A, 本部分取值应为0。

coding_profile: 应符合T/UWA 009.1—2023的附录A。

sampling_frequency_index: 应符合T/UWA 009.1—2023的附录A。

sampling_frequency: 应符合T/UWA 009.1—2023的附录A。

bitrate_index: 应符合T/UWA 009.1—2023的附录A。

channel_number: 应符合T/UWA 009.1—2023的附录A。

nn_type: 应符合T/UWA 009.1—2023的附录A。

content_type: 表示音频内容类型，见表 4。

表 4 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	

channel_num_index: 应符合T/UWA 009.1—2023的附录A。

object_channel_number: 应符合T/UWA 009.1—2023的附录A。

hoa_order: 表示HOA阶数，等于T/UWA 009.1—2023的附录A中order+1。

total_bitrate: 表示总码率，单位为 kb/s（千比特每秒），根据 content_type 的值计算方式不同，见表5。

表 5 total_bitrate 配置表

content_type 值	total_bitrate 计算方式
0	T/UWA 009.1—2023 附录 A.2 中声道信号 bitrate_index 对应的比特率值
1	T/UWA 009.1—2023 附录 A.2 中对象信号 bitrate_index_per_channel 对应的比特率值 ×(object_channel_number +1)
2	T/UWA 009.1—2023 附录 A.2 中声道信号 bitrate_index 对应的比特率值 + 对象信号的 bitrate_index_per_channel 对应的比特率值×(object_channel_number +1)
3	T/UWA 009.1—2023 附录 A.2 中 HOA 信号 bitrate_index 对应的比特率值
4-15	保留

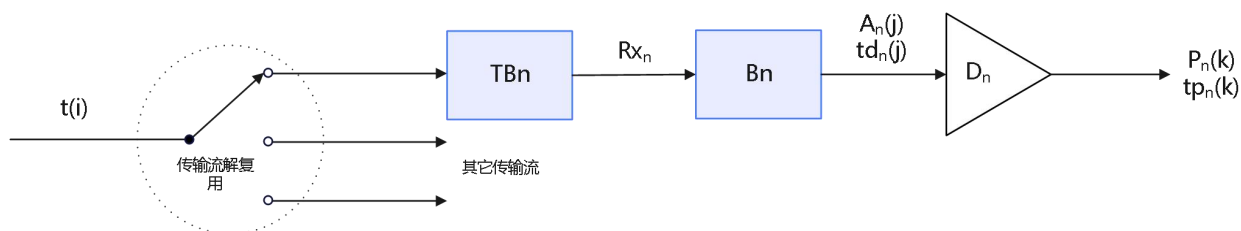
resolution: 应符合T/UWA 009.1—2023的附录A。

addition_info: 可选字段，附加信息。

8.4 Audio Vivid T-STD 模型拓展

8.4.1 T-STD 模型

Audio Vivid 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Audio Vivid的传输流应符合T-STD模型，见图1。



图中符号说明：

—— $t(i)$ ：传输流的第 i 个字节进入系统目标解码器的时间，单位 s 。

—— TB_n ：基本流 n 的传输缓存。

—— B_n ：基本流 n 的主缓存。

—— R_{x_n} ：从 TB_n 到 B_n 的传输速率。

—— $A_n(j)$ ：Audio Vivid 基本流的第 j 个存储单元。

—— $td_n(j)$ ： $A_n(j)$ 在系统目标解码器中解码的时间，单位 s 。

—— 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 s 。

图 1 面向 Audio Vivid 的 T-STD 模型拓展

8.4.2 缓存管理

Audio Vivid T-STD 模型扩展中从 TB_n 到 B_n 的传输速率 R_{x_n} 计算公式如公式（1）所示：

$$R_{x_n} = 1.2 \times R_{\max} \times N \dots \dots \dots (1)$$

式中：

$R_{\max}$ ——Audio Vivid 单通道最大速率；

N ——基本流中包含的音频通道总数；

表 6 给出了 Audio Vivid T-STD 模型扩展中对应 R_{x_n} 取值。

表 6 R_{x_n} 取值范围

通道数量	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

Audio Vivid T-STD 模型扩展中主缓存 BS_n 计算公式见公式（2）～（3）下：

$$BS_n = BS_{\text{mux}} + BS_{\text{dec}} + BS_{\text{oh}} \dots \dots \dots (2)$$

$$BS_{\text{mux}} = 0.004 \times R_{\max} \times N \dots \dots \dots (3)$$

式中：

BS_{mux} ——额外复用缓存：

T/UWA 009. 2-2—2025

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。

Audio Vivid T-STD 模型拓展中对应 BS_n取值见表 7。

表 7 BS_n取值范围

通道数量	BS _n [bytes]
1-8	7440
9-16	14 352
17-48	42 000
49-128	111 120

8.4.3 缓存延时

Audio Vivid传输流STD延时应满足：对于所有j对应存取单元A_n(j)中的所有字节i，对应 td_n(j) – t(i) ≤ 10 (s) 。

8.4.4 缓存管理条件

缓存管理条件满足：

- a) 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) B_n既不应上溢，也不应下溢；

9 SMT 传输

9.1 通则

Audio Vivid 编码位流基于 GB/T 33475.6 的 SMT 传输文件应遵循如下约束：

- a) Audio Vivid 编码位流符合第 5 章中基于 ISO BMFF 的文件封装格式；
- b) Audio Vivid 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；
- c) Audio Vivid 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。

9.2 Audio Vivid 媒体资源描述符

9.2.1 定义

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

9.2.2 语法

Audio Vivid媒体资源描述符语法见表8。

表8 Audio Vivid 媒体资源描述符语法

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
}			

9.2.3 语义

descriptor_tag: 用于标识描述符的类型。

descriptor_length: 指示描述符的长度, 单位为字节。

audio_format_type: 指示 Audio Vivid 编码位流的类别。该字段取值为 0 表示位流为 Audio Vivid AASF 存储格式的位流; 该字段取值为 1 表示位流为 Audio Vivid 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 Audio Vivid 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、**average_bitrate**、**min_bitrate** 分别指示音频媒体资源的最大码率、平均码率、最小码率, 以 kbps 为单位。

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。

9.3 交互反馈信令表

9.3.1 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Audio Vivid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取值为3；
- b) asset_group_flag取值为0。

9.3.2 语法

交互反馈信令表语法见表9。

表 9 交互反馈信令表

语法	值	比特数	类型
interaction_feedback_table() { table_id Version Length table_payload { table_type timestamp message_source asset_group_flag reserved if(asset_group_flag){ asset_group_id } else{ asset_id() } } }			uimsbf uimsbf 8 1 1 6 8

表 9 （续）

语法	值	比特数	类型
<pre>if(table_type == 3){ coordinate_type if(coordinate_type == 0){ ClientPosition() } if(coordinate_type == 1){ azimuth elevation distance } }</pre>		8	uimsbf
		8	uimsbf
		8	uimsbf
		8	uimsbf

9.3.3 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10。

表 10 交互反馈信令表类型取值含义

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;
```

T/UWA 009. 2-2—2025
}

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

10.1 RTP 负载

10.1.1 概述

Audio Vivid基于RTP协议进行传输时，通过RTP封装将音频帧封装为若干个RTP包，每个RTP包由RTP Header和RTP payload组成。

10.1.2 封包规则

RTP payload封包规则应符合以下任何一个约束：

- a) RTP 的 payload 携带的 Audio Vivid 编码格式为 aatf_frame()
- b) 如果一个 aatf_frame()的长度超过了 MTU，参考 RFC 3550 第 6.1 章的跨包规则来进行传输，即将该 aatf_frame()按照其长度之与 MTU 的倍数进行分割成多个 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 RTP 头

10.2.1 概述

RFC3550 第 5.1 条中定义的 RTP Header 结构定义见图 2,对于字段 Payload Type(PT), Marker(M) bit, Timestamp，其扩展定义见 10.2.2。

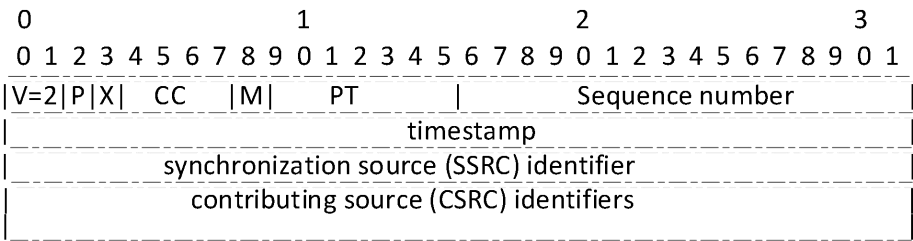


图 2 RTP Header 结构

10.2.2 RTP 头扩展

Payload Type (PT)：Audio Vivid 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。

10.3 SDP 定义

10.3.1 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 IETF RFC 8866 所述的规则, 需对第 10.3.2 章中的字段进行扩展定义。

10.3.2 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 Audio Vivid 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。

10.4 MIME 类型

10.4.1 概述

当使用 RFC 6381 定义的 MIME 类型的‘codecs’参数时, 如果 MIME 类型对应本文件定义的文件格式, ‘codecs’参数值应符合 Audio Vivid 编码位流格式, 其样本入口类型应支持“av3a”。

10.4.2 MIME 参数定义

MIME 媒体名字: audio

MIME 子名字: AV3A-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。

可选参数:

——**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/UWA 009.1—2023 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/UWA 009.1—2023 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。

示例:

codec-nn-id 使用举例如下:

——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。

——**config**: 16 进制的字符串, 对应第 5.3.1 中 CA3SpecificBox。

——**bitrate**: 音频流的码率, 对应第 5.2.1 中 total_bitrate。

10.4.3 MIME 类型的‘codecs’参数

10.4.3.1 概述

当使用 RFC 6381 定义的 MIME 类型的 ‘codecs’ 参数时, 如果 MIME 类型对应本文件中定义的文件格式, 且 ‘codecs’ 参数值起始于 5.3.2.1 定义的样本入口类型, 则 ‘codecs’ 参数的子参数应符合 10.4.3.2

T/UWA 009.2-2—2025
的规定。

10.4.3.2 Audio Vivid 编码参数

Audio Vividcodecs参数定义如下：

codecs = ‘av3a.audio_codec_id’

示例：

codecs= ‘av3a.02’

表示audio_codec_id为2。

11 RTMP 传输

11.1 概述

RTMP 使用的文件格式是 FLV，需要 FLV 格式进行扩展来支持 Audio Vivid 编码码流。

11.2 AUDIODATA 定义

AUDIODATA 的定义见表 11，Audio Vivid 的 SoundFormat 为 9，则将 FLV 中的 reserved 扩展为 Extended Format，在 AudioTagBody 前增加 SoundFormatEx，值为 17，用来指定 Audio Vivid。

表 11 AUDIODATA 定义

AudioTagHeader		
Field	Type	Comment
SoundFormat	UB[4]	9=Extended Format
SoundRate	UB[2]	Audio Vivid 中取值为 3
SoundSize	UB[1]	For Audio Vivid, Always 1
SoundType	UB[1]	For Audio Vivid, Always 1
SoundFormatEx	If SoundFormat=9 UB[16]	如果 SoundFormat=9，增加 SoundFormatEx 字段。 17 = Audio Vivid 其余为保留值
AudioPacketType	If SoundFormat==9 且 SoundFormatEx==17 UI8	The following values are defined(SoundFormat==9 且 SoundFormatEx==17): 0: Audio Vivid sequence header information 1: Audio Vivid Frame Data

11.3 AudioTagBody 定义

AudioTagBody 的定义见表 12，如果 AudioTagHeader 中 SoundFormat 为 9 且 SoundFormatEx 为 17，AudioTagBody 包含 AVS3AUDIODATA。

表 12 AudioTagBody 定义

AudioTagBody		
Field	Type	Comment
SoundData	IF SoundFormat==9 且 SoundFormatEx==17 AVS3AUDIODATA	新增 AVS3AUDIODATA 数据结构 0: Audio Vivid sequence header information 1: Audio Vivid Frame Data

11.4 AVS3AUDIODATA 定义

AVS3AUDIODATA的定义见表13。

表 13 AVS3AUDIODATA 定义

AVS3AUDIODATA		
Field	Type	Comment
Data	IF AudioPacketType==0 CA3SpecificBox () IF AudioPacketType==1 aatf_frame ()	CA3SpecificBox ()结构应符合本文件 5.3 中 CA3SpecificBox()的定义 aatf_frame ()结构应符合 T/UWA 009.1— 2023 中附录 A 的规定。

12 HLS 传输

12.1 概述

Audio Vivid编码码流在HLS协议中的扩展基于IETF RFC 8216规范，对切片文件格式以及HLS M3U8做扩展。

12.2 切片文件格式

Audio Vivid编码码流的切片格式使用两种格式：

- MPEG TS：用于封装Audio Vivid数据的TS切片格式应符合第8章的规定。
- fMP4：用于封装Audio Vivid数据的MP4切片格式应符合第5章的规定。

12.3 HLS M3U8 扩展

12.3.1 概述

HLS M3U8扩展规范遵循IETF RFC 8216所定义，扩展部分属性。

12.3.2 CODECS

Audio Vivid编码码流的CODECS参数应符合10.4.3节的规定。

示例：

```
#EXT-X-STREAM-INF:BANDWIDTH=3464568, CODECS="avc1.640028, av3a.02"
```

T/UWA 009.2-2—2025

example.m3u8

表示带有audio_codec_id为2的Audio Vivid的HLS流。

12.3.3 CHANNELS

Audio Vivid 编码码流的 CHANNELS 参数见 7.4.4 AudioChannelConfiguration 描述符。

参考文献

- [1] T/AI 109.7 信息技术智能媒体编码 第7部分:音频封装与传输
 - [2] IETF RFC 1738 统一资源定位符 (Uniform Resource Locators (URL))
 - [3] W3C XML 可扩展置标语言 (Extensible Markup Language (XML))
 - [4] W3C XML Schema Part 1 可扩展置标语言模式定义语言 第一部分: 结构 (XML Schema Definition Language (XSD) Part 1: Structures)
 - [5] W3C XML Schema Part 2 可扩展置标语言模式定义语言 第二部分: 数据类型 (XML Schema Definition Language (XSD) Part 2: Datatypes)
-