ICS 33.160.25
M74

UHD World Association
世界超高清视频产业联盟

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

（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。

202x-xx-xx 发布 202x-xx-xx 实施

世界超高清视频产业联盟

目 次

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。

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：

本文件主要起草人：

三维声技术规范 第 2-2 部分：应用指南 媒体格式

1 范围

本文件规定了采用了 T/UWA 009.1-2023 规定的三维声文件的封装格式、文件配置约束等。

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 信息技术 高效多媒体编码 第 3 部分：音频

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 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)

ITU-R BS.2076-2 音频定义模型（Audio Definition Model）

ITU-R BS.2094-1 音频定义模型通用定义（Common definitions for the Audio Definition Model）

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

Audio Vivid编码位流 Audio Vivid bitstream

符合T/UWA 009.1的编码音频信号所形成的二进制数据流。

3.3

Audio Vivid编码表示 Audio Vivid representation

符合T/UWA 009.1的以其编码形式表示的数据单元。

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，可简称采样率。

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。

Audio Vivid	UWA 三维声编码技术规范	(Audio Vivid)
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)
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 kstream)
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 通则

本节定义适用于 Audio Vivid 通用全码率音频编码内容的特有配置。

5.2.1.2 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_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;
    }
}
```

5.2.1.3 语义

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

nn_type：应符合 T/UWA 009.1附录 A 的规定。

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/UWA 009.1的附录A。

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

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

hoa_order：表示 HOA 阶数，为 T/UWA 009.1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。

表 2 total_bitrate 配置表

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

resolution：应符合T/UWA 009.1附录A的规定

5.2.2 Audio Vivid 无损音频编码特有配置

5.2.2.1 定义

本节定义适用于 Audio Vivid 无损音频编码内容的特有配置。

5.2.2.2 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;
}
```

5.2.2.3 语义

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

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

anc_data_index：应符合 T/UWA 009.1附录 A 的规定，本部分取值为0。

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

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

resolution：应符合 T/UWA 009.1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

```
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 ('stsd')

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附录 A。

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) 本文件5.3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 必须对应一个唯一的 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（'stsd'）

Audio Vivid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Audio Vivid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5.3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) 应符合ISO/IEC 23000-19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。

6.4 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”参数应符合附录A的规定。

7 DASH 传输技术要求

7.1 概述

本章规定了Audio Vivid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 通则

用于封装 Audio Vivid 数据的 DASH 片段格式应符合第 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’。

7.2.2 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。

7.2.3 媒体片段

每个 DASH 媒体片段应包含一个或多个 T/UWA 009.1 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 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 应保持不变；

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

7.2.4 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 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

7.3 DASH MPD 编码器参数

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

7.4 DASH MPD 描述子

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 属性应该设置成 1。

7.4.4 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.01'，则@value的属性值为：

——第1个字节为0xF0；

——第2个字节等于5.2.2.2中channel_number的值；

——第3个字节为0；

b) 如果AVS3音频编码位流在MPD中的@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5.2.1.2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5.2.1.2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5.2.1.2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 5.2.1.2 中 object_channel_number + 1 的值；否则，第 3 个字节的值等于 0；

8 传输流和节目流技术要求

8.1 通则

本章规定了适用于Audio Vivid编码位流的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 中节目的一个节目元素，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

8.2.1 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 里应该是字节对齐的，即 Audio Vivid 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()。

8.3 节目和节目元素描述符

8.3.1 节目和节目元素描述子中各字段的语义定义

8.3.1.1 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 Audio Vivid 流描述符标签值，见表 4。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 4 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

8.3.1.2 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。

8.3.2 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。

8.3.3 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，Audio Vivid 的 format_identifier 应为 0x41-56-53-41('AVSA')。

8.3.4 Audio Vivid 流描述符

AVS3_Audio_descriptor()位于 PMT 中的 ES_info_length 字段后面。其语法见表 5。

表 5 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
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
}		
resolution	2	uimsbf
reserved	6	bslbf
for (i=0; i<N; i++) {		
addition_info[i]	8	bslbf
}		
}		

8.3.5 Audio Vivid 流描述符中各字段的语义定义

descriptor_tag：Audio Vivid描述符的tag应该为210(0xD2)。
descriptor_length：Audio Vivid描述符的长度。
audio_codec_id：应符合T/UWA 009.1的附录A。
anc_data_index：应符合T/UWA 009.1的附录A，本部分取值为0。
coding_profile：应符合T/UWA 009.1的附录A。
sampling_frequency_index：应符合T/UWA 009.1的附录A。
sampling_frequency：应符合T/UWA 009.1的附录A。
bitrate_index：应符合T/UWA 009.1的附录A。
channel_number: 应符合T/UWA 009.1的附录A。
nn_type：应符合T/UWA 009.1的附录A。
content_type：表示音频内容类型，见表 6。

表 6 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的附录A。
object_channel_number：应符合T/UWA 009.1的附录A。
hoa_order：表示HOA阶数，等于T/UWA 009.1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7。

表 7 total_bitrate 配置表

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

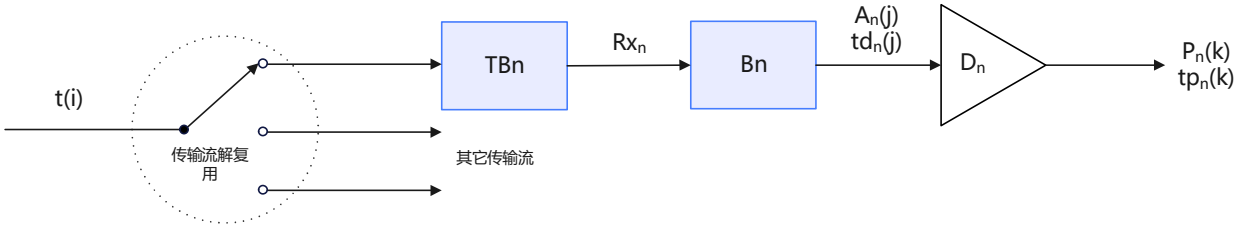
3	T/UWA 009.1 附录 A.2 中 HOA 信号的 bitrate_index 对应的比特率值
4-15	保留

resolution：应符合T/UWA 009.1的附录A。

addition_info：可选字段，附加信息。

8.4 Audio Vivid T-STD 模型拓展

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。

——$A_n(j)$：Audio Vivid 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。

图 1 面向 Audio Vivid 的 T-STD 模型拓展

8.4.1 缓存管理

Audio Vivid 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：

$$Rx_n = 1.2 \times R_{max} \times N \cdots\cdots (1)$$

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 8 给出了 Audio Vivid T-STD 模型拓展中对应Rx_n取值。

表 8 Rx_n取值范围

Number of Channels	Rx_n [bit/s]
1-8	2 000 000
9-16	3 686 400
17-48	11 059 200

49-128	29 491 200

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

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

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

式中：

BS_{mux}——额外复用缓存；

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 9。

表 9 BS_n取值范围

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

8.4.2 缓存延时

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

8.4.3 缓存管理条件

缓存管理条件应该满足：

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既不能上溢，也不能下溢；

9 SMT 传输技术要求

9.1 通则

Audio Vivid 编码位流基于 SMT 的传输文件需遵循如下约束：

a) Audio Vivid 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；

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10。

表 10 Audio Vivid 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
}			

9.2.3 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 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。

9.3.2 语法

交互反馈信令表语法见表11。

表 11 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			
message_source		1	
asset_group_flag		1	uimsbf
reserved		6	uimsbf
if(asset_group_flag){			
asset_group_id		8	
}			

语法	值	比特数	类型
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
}			
}			
}			
}			

9.3.3 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12。

表 12 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。

10 RTP 传输技术要求

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封包规则应当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，需要参考 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 章中定义的 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。

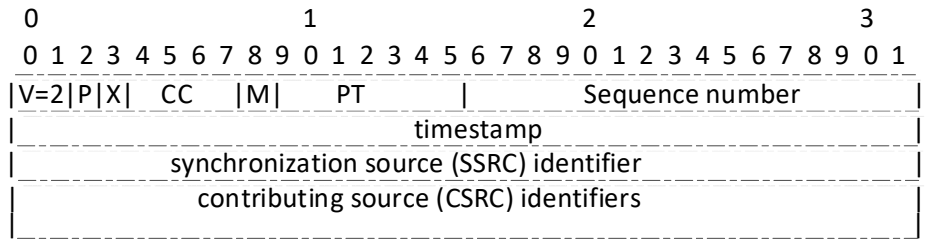


图 2 RTP Header 结构

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 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。

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 类型对应本标准中定义的文件格式，‘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 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/UWA 009.1 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。

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类型对应本标准中定义的文件格式，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 Audio Vivid 基本流在 RTMP 中的定义

11.1 概述

RTMP 使用的文件格式是 FLV。需要对该标准进行扩展来支持 Audio Vivid 编码码流。

11.2 AUDIODATA

AUDIODATA 的定义见表 13，Audio Vivid 的 SoundFormat 为 9，即原来的 reserved 扩展为 Extended Format，在 AudioTagBody 前增加 SourdFormatEx，值为 17，用来指定 Audio Vivid。

表 13 AUDIODATA 定义

AudioTagHeader		
Field	**Type**	**Comment**
SoundFormat	UB[4]	9=Extended Format
SoundRate	UB[2]	For AVS3 Audio, Always 3
SoundSize	UB[1]	For AVS3 Audio, Always 1
SoundType	UB[1]	For AVS3 Audio, Always 1
SourdFormatEx	If SoundFormat=9 UB[16]	如果 SoundFormat=9, 增加 SoundFormatEx 字段。 17 = AVS3 Audio 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: AVS3 Audio sequence header information 1: AVS3 Audio Frame Data

11.3 AudioTagBody

AudioTagBody 的定义见表 14，如果 AudioTagHeader 中 SoundFormat 为 9 且 SoundFormatEx 为 17，AudioTagBody 包含 AVS3AUDIODATA。

表 14 AudioTagBody 定义

AudioTagBody		
Field	**Type**	**Comment**
SoundData	IF SoundFormat==9 且 SoundFormatEx==17	新增 AVS3AUDIODATA 数据结构 0: AVS3 Audio sequence header information

AVS3AUDIODATA	1: AVS3 Audio Frame Data

11.4 AVS3AUDIODATA

AVS3AUDIODATA的定义见表15所。

表 15 AVS3AUDIODATA 定义

AVS3AUDIODADA		
Field	**Type**	**Comment**
Data	IF AudioPacketType==0 CA3SpecificBox () IF AudioPacketType==1 aatf_frame ()	CA3SpecificBox ()结构引用本文件的 5.3 的 CA3SpecificBox()定义 aatf_ frame ()结构引用本文件

12 Audio Vivid 基本流在 HLS 中的定义

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：

a) 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。

b) 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 CODEC

Audio Vivid编码码流的CODEC见A.2。

示例：

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

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 见 7.4.4 AudioChannelConfiguration 描述子。

参考文献

[1] T/AI 109.7 信息技术 智能媒体编码 第 7 部分：智能媒体格式-音频

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

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

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

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

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

[7]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)

[8]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)