

世界超高清视频产业联盟标准

T/UWA 012.2—2022

“百城千屏”超高清视音频传播系统

视音频编码：系统

Coding of UHD video and audio broadcasting system for “Bai Cheng Qian Ping”:
system

XXXX – XX – XX 发布

XXXX – XX – XX 实施

目 次

前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	1
5 编码音视频复用传输要求.....	2
5.1 总体要求.....	2
5.2 stream_id 要求.....	2
5.3 stream_type 要求.....	2
5.4 AVS3 视频描述符要求.....	2
5.5 AVS2 视频描述符要求.....	5
5.6 扩展 PES 分组.....	6
6 传输流的 IP 传输要求.....	7
6.1 传输流作为 UDP 的负载传输.....	7
6.2 传输流作为 RTP 的负载传输.....	7
附录 A 独立音频传输（资料性附录）.....	错误！未定义书签。
A.1 独立音频互联网传输.....	错误！未定义书签。
A.2 独立音频同步要求.....	错误！未定义书签。

前 言

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

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这些专利的责任。

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：国家广播电视总局广播电视规划院、广东博华超高清创新中心有限公司、中央广播电视总台超高清视音频制播呈现国家重点实验室、华为技术有限公司、日本夏普株式会社、中国移动通信集团咪咕文化科技有限公司、海信视像科技股份有限公司、TCL实业控股股份有限公司、京东方科技集团股份有限公司、中国电子信息产业发展研究院、中兴通讯股份有限公司、北京市博汇科技股份有限公司、北京百度网讯科技有限公司、杭州当虹科技股份有限公司、利亚德光电股份有限公司、北京广播电视台、上海文化广播影视集团有限公司、上海交通大学、中国电子技术标准化研究院、中国信息通信研究院、上海数字电视国家工程研究中心有限公司、工业和信息化部电子第五研究所、北京爱奇艺科技有限公司、成都索贝数码科技股份有限公司、深圳创维-RGB电子有限公司、山东浪潮超高清视频产业有限公司、深圳酷开网络科技股份有限公司、北京数码视讯科技股份有限公司、索尼（中国）有限公司、北京三星通信技术研究有限公司、上海国茂数字技术有限公司、上海网达软件股份有限公司、北京锐马视讯科技有限公司、赛因芯微（北京）电子科技有限公司、北京淳中科技股份有限公司、成都德芯数字科技股份有限公司、重庆赛宝工业技术研究院有限公司、上海通维通讯网络科技有限公司、北京云晁科技有限公司

本文件主要起草人：姜文波、张文林、高山俊明、徐进、冯景峰、刘昕、于芝涛、严方红、原烽、张文刚、梅剑平、宁金辉、张伟民、崔俊生、潘晓菲、许春蕾、龙仕强、张鸿宇、温晓君、黄成、陈勇、刘莉、王荣芳、白雅贤、李法、陈家兴、罗峻兮、董云翔、李康敬、郭佩佩、潘兴浩、王之奎、李斌、顿胜堡、冯林、王立冬、范金慧、王振中、孙剑、宜玉栋、徐异凌、许哲、王亚军、薛勇、徐华伟、沈思宽、徐遥令、崔艳春、尹旭辉、李晓榕、赵显亮、邢怀飞、王立众、王国中、王威、袁谦、李庆喻、彭海、荣继、孟祥朋、吴健、吴强、孔令术、张振兴、李俊、曾贵修、潘月宏、霍杰、武智、沈培晶

“百城千屏”超高清视音频传播系统视音频编码：系统

1 范围

本文件规定了“百城千屏”超高清视音频传播系统中编码的视频基本流、音频基本流在GB/T 17975.1-2010定义的传输流中复用传输的要求，并规定了传输流在IP网络中的传输格式。

本文件适用于“百城千屏”超高清视音频传播系统中的视音频编码码流的复用与传输。

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。

ISO/IEC 13818-1:2019 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 13818-7: 2006 信息技术 运动图像及其伴音信息的通用编码 第1部分：先进音频编码
(Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC))

ISO/IEC 14496-3:2019 信息技术 音视频对象编码 第3部分：音频 (Information technology — Coding of audio-visual objects — Part 3: Audio)

ETSI TS 102 034 在IP网络中传输GB/T 17975.1的传输流 (Transport of MPEG-2 TS Based DVB Services over IP Based Networks)

RFC 7230 超文本传输协议消息语法与路由1.1 (Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing)

RFC 8216 HTTP实时流传输 (HTTP Live Streaming)

GB/T 17975.1-2010 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

GB/T 17975.3-2002 信息技术 运动图像及其伴音信号的通用编码 第3部分 音频

GB/T 33475.2-2016 信息技术 高效多媒体编码 第2部分：视频

T/AI 109.2 信息技术 智能媒体编码 第2部分：视频

3 术语和定义

下列术语和定义适用于本文件。

3.1

知识图像 library picture

T/AI 109.2中规定的知识位流中的图像，可以被其他位流中的图像参考

3.2

知识位流 library stream

T/AI 109.2中规定的包含知识图像的位流

3.3

主位流 sequence stream

T/AI 109.2中规定的可参考由该位流之外的信息提供的知识图像进行解码的位流

4 缩略语

下列缩略语适用于本文件。

AVS2 信息技术 高效音视频编码 第2部分: 视频 (Information Technology—High Efficiency Media Coding Part2: Video)

AVS3 信息技术 智能媒体编码 第2部分: 视频 (Information Technology Intelligent Media Coding Part2: Video)

HLS HTTP实时流传输 (HTTP Live Streaming)

RTP 实时应用传输协议 (Real-time Transport Protocol)

bslbf 比特串, 即二进位串, 左位在先 (Bit string, left bit first, where "left" is the order in which bit strings are written in the specification.)

uimsbf 无符号整数, 最高有效位在先 (Unsigned integer, most significant bit first)

UDP 用户数据报协议 (User Datagram Protocol)

UTC 世界标准时间 (Coordinated Universal Time)

5 编码音视频复用传输要求

5.1 总体要求

视频编码采用T/AI 109.2基准10位档、10.0.60级或GB/T 33475.2-2016基准10位档、8.0.60级。超高清视音频编码的视频基本流、音频基本流宜在GB/T 17975.1-2010定义的传输流中传输。

5.2 stream_id 要求

GB/T 17975.1-2010的“表25 stream_id赋值”基础上, stream_id为1110 xxxx的流编码规定为“GB/T 17975.2或GB/T20090.2或GB/T17191.2或T/AI 109.2或GB/T 33475.2, 视频流编号为xxxx”。

音频stream_id见ISO/IEC 13818-1:2019。

5.3 stream_type 要求

GB/T 17975.1-2010的“表37 流类型指定”基础上, 补充 stream_type值0xD2为GB/T 33475.2视频, T/AI109.2视频流的stream_type值为0xD4。

音频stream_type见ISO/IEC 13818-1:2019。

5.4 AVS3 视频描述符要求

5.4.1 AVS3 视频描述符语法

在GB/T 17975.1-2010的2.6节基础上增加AVS3视频描述。AVS3视频描述符语法应符合表1的要求。在GB/T 17975.1-2010的“表50节目和节目元素描述符”基础上, 补充标签值62, 标签值62应为AVS3视频描述的标签。

表1 AVS3 视频描述符语法

语 法	位 数	助 记 符
AVS3_video_descriptor () {		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
profile_id	8	uimsbf
level_id	8	uimsbf
multiple_frame_rate_flag	1	bslbf
frame_rate_code	4	uimsbf
sample_precision	3	uimsbf

表1 AVS3视频描述符语法 (续)

语 法	位 数	助 记 符
chroma_format	2	uimsbf
temporal_id_flag	1	bslbf
td_mode_flag	1	bslbf
library_stream_flag	1	uimsbf
reserved	3	bslbf
colour primaries	8	uimsbf
transfer_characteristics	8	uimsbf
matrix_coefficients	8	uimsbf
if (!library_stream_flag) {		
num_ref_library_stream	7	uimsbf
id_type_flag	1	bslbf
for (i=0; i<num_ref_library_stream; i++) {		
if (id_type_flag)		
ref_library_stream_PEID[i]	13	uimsbf
else {		
ref_library_stream_id[i]	8	uimsbf
reserved	5	bslbf
}		
reserved	3	bslbf
}		
}		

5.4.2 AVS3 视频描述各字段语义

档次字段 `profile_id`

该字段为8位。表示视频位流的档次，该字段与T/AI109.2视频位流中`profile_id`字段相同。

级别标识符字段 `level_id`

该字段为8位。表示视频位流的等级。该字段与T/AI109.2视频位流中`level_id`字段相同。

复合帧速率标志 `multiple_frame_rate_flag`

该字段为1位，置‘1’时表示视频流中可能有多个帧速率，置‘0’时表示只有单一帧速率。

帧速率码字段 `frame_rate_code`

该字段为4位，该字段与T/AI109.2视频位流中`frame_rate_code`字段定义相同。
`multiple_frame_rate_flag`字段置‘1’时，一个特定的帧速率意味着视频流中允许有某些其它的帧速率。
 应符合表2的要求。

表2 帧速率码

编码速率	同时允许的速率
23.976	23.976
24.0	
25.0	
29.97	23.976

表2 帧速率码（续）

编码速率	同时允许的速率
30.0	23.976 24.0 29.97
50.0	25.0
59.94	23.976 29.97
60.0	23.976 24.0 29.97 30.0 59.94
100.0	50.0
119.88	59.94
120.0	59.94 60.0 119.88

样本精度字段 sample_precision

该字段为3位。规定亮度和色度样本的精度。该字段与T/AI109.2视频位流中sample_precision字段相同。

色度格式字段 chroma_format

该字段为2位。规定色度分量的格式。该字段与T/AI109.2视频位流中chroma_format字段相同。

时间层标识允许标志 temporal_id_flag

该字段为1位。表示视频流是否允许使用时间层标识。该字段与T/AI109.2视频流中temporal_id_enable_flag字段相同。

立体视频模式标志 td_mode_flag

该字段为1位。表示视频流是单目视频流，或是多视点视频流。该字段与T/AI109.2视频流中td_mode_flag字段相同。

彩色三基色 colour_primaries

该字段为8位。说明视频流中源图像三基色的色度坐标。该字段与T/AI109.2视频流中colour_primaries字段相同。

光电转换特性 transfer_characteristics

该字段为8位。说明视频流中源图像的光电转换特性。该字段与T/AI109.2视频流中transfer_characteristics字段相同。

彩色信号转换矩阵 matrix_coefficients

该字段为8位。说明从红绿蓝三基色转换为亮度和色度信号时采用的转换矩阵。该字段与T/AI109.2视频流中matrix_coefficients字段相同。

知识位流标识 library_stream_flag

该字段为1位。指示节目中与该描述子对应的基本流是否为知识位流。值为‘1’表示描述子对应的基本流是知识位流；值为‘0’表示该描述子对应的基本流是主位流。

被参考的知识基本流数目字段 num_ref_library_stream

该字段为7位。定义了与该描述子对应的基本流是主位流时，该主位流依赖的知识位流的数量。

索引类型字段 id_type_flag

该字段为1位。指示与该描述子对应的基本流是主位流时，该主位流依赖的知识位流的索引类型。该字段值为‘1’表示被依赖的知识位流的索引使用该知识位流所在的传输流分组包的PEID；该字段值为‘0’表示被依赖的知识位流的索引使用该知识位流所在的PES分组包的stream_id。当描述子包含在节目流中时，该字段值应为‘0’。

被参考的知识基本流 PEID 字段 ref_library_stream_PEID[i]

该字段为13位。定义了与该描述子对应的基本流是主位流时，该主位流依赖的第i个知识位流所在传输流分组包中PEIT_indicator字段的值。

被参考的知识基本流索引字段 ref_library_stream_id[i]

该字段为8位。定义了与该描述子对应的基本流是主位流时，该主位流依赖的第i个知识位流所在PES分组包头中stream_id字段的值。

5.5 AVS2 视频描述符要求

5.5.1 AVS2 视频描述符语法

在GB/T 17975.1-2010的2.6节基础上补充AVS2视频描述符。AVS2视频描述符语法应符合表3的要求。在GB/T 17975.1-2010的“表50节目和节目元素描述符”基础上，补充标签值64，签值64应为AVS2视频描述的标签。

表 3 AVS2 视频描述符语法

语 法	位数	助记符
AVS2_video_descriptor () {		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
profile_id	8	uimsbf
level_id	8	uimsbf
extension_layer_number	8	uimsbf
for (i=1; i<=extension_layer_number; i++) {		
layer_profile_id[i]	8	uimsbf
layer_level_id[i]	8	uimsbf
layer_type[i]	8	uimsbf
dependent_layer_number[i]	8	uimsbf
for (j=0; j<dependent_layer_number[i]; j++) {		
dependent_layer_id[i][j]	8	uimsbf
}		
}		
multiple_frame_rate_flag	1	bslbf
frame_rate_code	4	uimsbf
AVS_still_present	1	bslbf
chroma_format	2	uimsbf
sample_precision	3	uimsbf
reserved	5	bslbf
colour_primaries	8	uimsbf
transfer_characteristics	8	uimsbf
matrix_coefficients	8	uimsbf
}		

5.5.2 AVS2 视频描述各字段语义

档次标识 **profile_id**

8位字段。表示比特流的档次。该字段与GB/T 33475.2视频流中profile_id字段相同。

等级标识 **level_id**

8位字段。表示比特流的等级。该字段与GB/T 33475.2视频流中level_id字段相同。

扩展层数 **extension_layer_number**

8位字段。表示比特流的扩展层数。该字段与GB/T 33475.2视频流中extension_layer_number字段相同。

层档次 **layer_profile_id**

8位字段。表示该比特流层的档次。该字段与GB/T 33475.2视频流中layer_profile_id字段相同。

层等级 **layer_level_id**

8位字段。表示该比特流层的等级。该字段与GB/T 33475.2视频流中layer_level_id字段相同。

层类型 **layer_type**

8位字段。该字段与GB/T 33475.2视频流中level_id字段相同。

依赖层数 **dependent_layer_number**

8位字段。表示当前层所依赖的其他层的数量。该字段与GB/T 33475.2视频流中dependent_layer_number字段相同。

依赖层索引 dependent_layer_id

8位字段。表示当前层所依赖的层的索引。该字段与GB/T 33475.2视频流中dependent_level_id字段相同。

复合帧速率标志 multiple_frame_rate_flag

1位字段，置'1'时表示视频流中可能有多个帧速率，置'0'时表示只有单一帧速率。

帧速率码字段 frame_rate_code

4位字段，该字段与GB/T 33475.2视频流中frame_rate_code字段定义相同。不同点在于multiple_frame_rate_flag字段置'1'时，一个特定的帧速率意味着视频流中允许有某些其它的帧速率。应符合表4的要求。

表4 帧速率码

编码速率	同时允许的速率
23.976	
24.0	23.976
25.0	
29.97	23.976
30.0	23.976 24.0 29.97
50.0	25.0
59.94	23.976 29.97
60.0	23.976 24.0 29.97 30.0 59.94

AVS 静态图象 AVS_still_present

1位字段。置'1'时表示该视频流只含静态图像数据；置'0'时，则可包含运动的或静态的图像数据。

色度格式 chroma_format

2位字段。规定色度分量的格式。该字段与GB/T 33475.2视频流中chroma_format字段编码方式相同。

采样精度 sample_precision

3位字段。规定亮度和色度样本的精度。该字段与GB/T 33475.2视频流中sample_precision字段编码方式相同。

彩色三基色 colour_primaries

8位字段。表示视频流中源图像三基色的色度坐标。该字段与GB/T 33475.2视频流中colour_primaries字段相同。

光电转移特性 transfer_characteristics

8位字段。表示视频流中源图像的光电转移特性。该字段与GB/T 33475.2视频流中transfer_characteristics字段相同。

彩色信号转换矩阵 matrix_coefficients

8位字段。表示从红绿蓝三基色转换为亮度和色度信号时采用的转换矩阵。该字段与GB/T 33475.2视频流中matrix_coefficients字段相同。

5.6 扩展 PES 分组

扩展PES分组，使用PES分组的私有数据段传输编码时间信息TimeStamp。

5.6.1 TimeStamp 语法

在GB/T 17975.1-2010基础上，PES_extension_flag置为1、PES_private_data_flag置为1，PES_private_data传输语法结构应符合表5的要求。该语法结构的应用见附录A。

表5 TimeStamp 语法结构

语 法	位数	助记符
TimeStamp () {		
syncword	12	uimsbf
version	2	uimsbf
utc_time_valid	1	uimsbf
reserved	1	uimsbf
reserved	64	uimsbf
utc_time	48	uimsbf
}		

5.6.2 TimeStamp 各字段语义

同步字 syncword

语法结构同步，应为0xFEE。

版本 version

表示该语法结构的版本，值为1。

reserved 保留

本语法结构中保留为全1。

UTC 时间有效标志位 utc_time_valid

是否设置utc时间的标志位，占1bit，1代表utc_time具有实际值，0代表utc_time无效。

世界标准时间 utc_time

生成该帧的时间，1970年1月1日0时0分0秒开始计数的毫秒数。

6 传输流的 IP 传输要求

6.1 传输流作为 UDP 的负载传输

应符合ETSI TS 102 034中7.1.2节的规定。

6.2 传输流作为 RTP 的负载传输

应符合ETSI TS 102 034中7.1.1节的规定。

附录 A (资料性) 独立音频传输

存在有的大屏不宜播放声音的情况,但是观众有观看大屏节目时听到同步音频的需求。针对此需求,需要单独传输一路与节目对应音频,以供观众收看大屏节目时用随身携带的移动设备同步收听节目的伴音,具体见图1。由于大屏位置不同、不同传输网络的延时不同、不同类型终端解码缓冲策略不同,造成解码视频播放延时不同;需要在节目码流中添加时间戳,具体语法结构见5.5节,解码终端需上报解码时间与对应时间戳;用户随身携带移动设备应用需获取设备所在位置以供服务平台找到对应大屏,并提供适合的音频流。

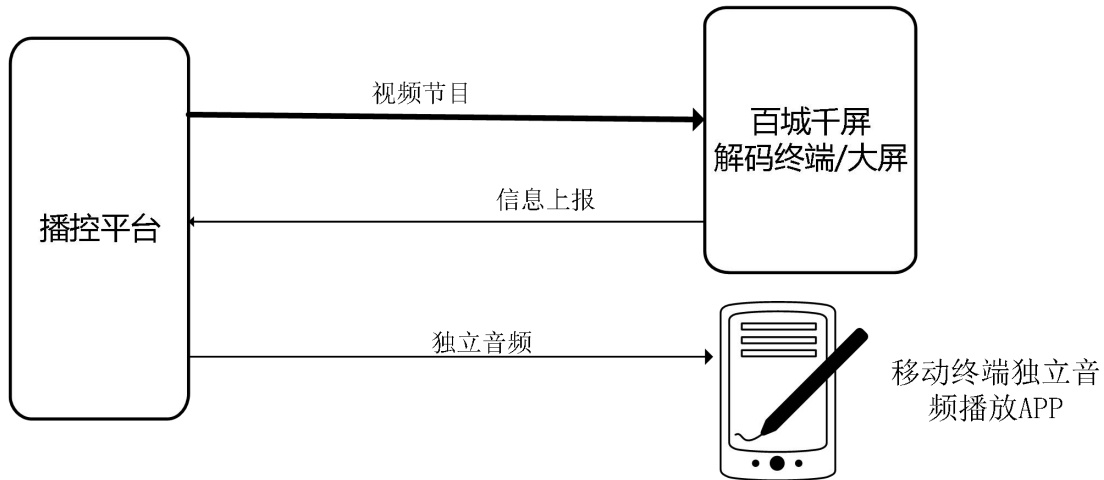


图 A.1 百城千屏独立音频传输示意图

A.1 独立音频互联网传输

为了降低延时和传输开销,建议采用RTMP或WebRTC传输,音频编码采用AAC。

为了控制独立音频播放终端可与对应大屏的视频同步播出,建议在音频中嵌入与生成该帧的时间对应的UTC时间戳,具体的时间戳嵌入方式根据独立音频流的分发方式和对应播放终端确定。

A.2 独立音频同步要求

在独立音频传输时,音频与对应的大屏视频时延应控制在 $(-120, 120)$ 毫秒内。

播放独立音频的移动终端APP应具备同步微调功能。