



中华人民共和国国家标准

GB/T 44115.3—2025

信息技术 虚拟现实内容表达 第3部分：音频

Information technology—Virtual reality content representation—
Part 3: Audio

2025-02-28 发布

2025-09-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

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 符号和缩略语	3
5 串行流语法规则	6
6 系统框架	7
7 元数据定义	13
8 AAMSF 封装格式	57
9 耳机渲染器	60
10 扬声器渲染器	73
附录 A (资料性) HRTF 库的模式选择	86
附录 B (资料性) 一种基于 HRTF 线性插值的 VR 音频方法	87
附录 C (资料性) 人工混响模式	92
附录 D (资料性) 一种基于 Ambisonics 全景录音球的麦克风阵列摆放实现方案	96
附录 E (资料性) 一种基于 Ambisonics 全景录音球的空间编码方法	103
附录 F (资料性) 场景信号解码及渲染的通用技术	108
附录 G (资料性) 多声道音频生产系统规范建议及测试方案	112
附录 H (资料性) 基于头相关传输函数的 3DoF 三维音频渲染器	119
附录 I (资料性) 智能电视音箱环绕立体声系统	122
附录 J (资料性) 球谐 HRTF 通用制作技术	129
附录 K (资料性) 采样卷积混响模式	131
附录 L (资料性) 一种空间音频距离衰减模型	157
附录 M (资料性) 一种三维声近场效应实现方法	159
附录 N (资料性) 实时制作时元数据传输系统的案例	363
附录 O (资料性) 非 GY/T 316 标准扬声器布局的渲染器	369
附录 P (资料性) 扩展元数据示例代码	373
附录 Q (规范性) Surface 子元素支持的材料类型	376
附录 R (规范性) 注册渲染器 ID0 的耳机渲染规范	377
附录 S (规范性) 注册渲染器 ID128 的扬声器渲染规范	382
参考文献	387

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。

本文件是 GB/T 44115《信息技术 虚拟现实内容表达》的第 3 部分。GB/T 44115 已经发布了以下部分：

——第 2 部分：视频；

——第 3 部分：音频。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：清华大学、北京理工大学、中国电子技术标准化研究院、北京字跳网络技术有限公司、赛因芯微(北京)电子科技有限公司、北京全景声信息科技有限公司、中关村视听产业技术创新联盟、清华大学天津电子信息研究院、北京大学、咪咕文化科技有限公司、腾讯科技(深圳)有限公司、北京爱奇艺科技有限公司、小米通讯技术有限公司、全景声(北京)智能科技有限公司、华为技术有限公司、中国传媒大学。

本文件主要起草人：窦维蓓、王晶、耿一丹、朱博成、李婧欣、黄传增、柳德荣、吴健、吴强、许舒敏、潘兴德、曲天书、李岳鹏、商世东、王志航、刘长滔、黄为庆、刘孟美、赵天博、韩泽瑞、王宾、韩建、李琳、徐嵩、胡苏、王喆、高原、朱梦尧、靳聪、张伟民、高文、黄铁军、赵海英、崔晓冉。

引 言

随着虚拟现实(VR)技术的不断涌现,虚拟现实已成为增强沉浸感的新兴媒体形式,VR音频内容表达与渲染重建技术在虚拟现实沉浸感、真实感体验中占有重要地位。目前虚拟现实全景视频技术已经受到广泛关注和重视,但是由于缺乏VR音频内容表达和渲染重建的技术规范,难以实现各种VR音频采集制作与终端回放设备或系统间的互通共享。

GB/T 44115旨在确立适用于虚拟现实的视频和音频规范,拟由三个部分构成。

- 第1部分:系统。目的在于确立高效编码的虚拟现实视频和音频数据的传输格式与信令格式。
- 第2部分:视频。目的在于确立虚拟现实全景视频和自由视角视频的编码表示与重建方法。
- 第3部分:音频。目的在于确立虚拟现实设备及相关系统中的沉浸式音频内容的表达方式,包括元数据结构 and 渲染器的系统架构及接口规范。

本文件在遵循ITU-R BS.2076:2019的前提下,面向适应我国全景音频播出、沉浸式音频通信、虚拟现实音频交互等领域,确立VR音频元数据模型和渲染系统之间的协同关系,以及VR音频元数据流与音频流之间的可能的复接模式,形成VR音频内容表达的互联互通应用规范,支持3DoF和6DoF等虚拟现实场景下,涉及基于声道、对象、场景三类信号的采集制作和渲染重建,服务于虚拟现实(VR/AR/MR/XR)设备和相关系统中的音频应用。

本文件设计了音频元数据模型、音频编解码系统和VR音频渲染系统之间的解耦合架构,三个系统之间既相互独立,又紧密配合,形成VR音频内容表达的应用规范。本文件将对VR音频坐标系、元数据模型、渲染器系统、AVS音频元数据串行流封装格式进行规范。

本文件的发布机构提请注意,声明符合本文件时,可能涉及6.3、7.2、7.5、9.1、9.8、10.1、附录D、附录H、附录I、附录R、附录S中如下52项与虚拟现实音频相关的专利的使用。专利名称如下:

PCT/CN2021/100076,音频渲染系统、方法和电子设备;PCT/CN2022/098882,音频渲染系统、方法和电子设备;PCT/CN2021/100062,用于音频渲染的音频信号编码方法、装置和电子设备;PCT/CN2022/098850,音频渲染系统、方法和电子设备;PCT/CN2021/114366,音频元数据的处理方法和装置;PCT/CN2022/114219,音频信号的处理方法和装置;PCT/CN2021/121135,声音路径能量的淡入淡出方法、电子设备和介质;PCT/CN2022/122204,音频渲染方法、音频渲染设备和电子设备;PCT/CN2021/121718,一种动态估计场景近似长方体房间的方法;PCT/CN2022/122635,一种音频渲染系统和方法;PCT/CN2021/104309,混响时长的估计方法、音频信号的渲染方法和电子设备;PCT/CN2022/103312,音频信号的渲染方法、装置和电子设备;PCT/CN2021/115130,用于音频渲染的信号处理方法、装置和电子设备;PCT/CN2022/115194,用于音频渲染的信号处理方法、装置和电子设备;PCT/CN2021/121729,用于空间音频渲染的系统、方法和电子设备;PCT/CN2022/122657,用于空间音频渲染的系统、方法和电子设备;202110984837.4,一种音频制作模型和生成方法、电子设备及存储介质;202111102045.6,音频节目元数据和产生方法、电子设备及存储介质;202111100818.7,音频内容元数据和产生方法、电子设备及存储介质;202111102038.6,音频对象元数据和产生方法、电子设备及存储介质;202111205630.9,音轨唯一标识元数据和生成方法、电子设备及存储介质;202111204386.4,一种音频轨道元数据和生成方法、电子设备及存储介质;202111202898.7,一种音频流元数据和生成方法、电子设备及存储介质;202111308422.1,基于音床音频包格式元数据和产生方法、设备及介质;202111308430.6,基对象音频包格式元数据和产生方法、设备及介质;202111306844.5,基于场景音频包格式元数据和产生方法、设备及存储介质;202111308421.7,基于双耳音频包格式元数据和产生方法、设备及介质;202111021068.4,基于音床音频通道元数据和生成方法、设备及存储介质;202111020417.0,基于对象音频通道元数据和生成方

法、设备及存储介质；202111021066.5, 基于场景音频通道元数据和生成方法、设备及存储介质；202111021039.8, 基于双耳音频通道元数据和生成方法、设备及存储介质；202111666346.1, 一种广播音频格式文件生成方法、装置、设备及存储介质；202111666362.0, 音频元数据区块的生成方法、装置、设备及存储介质；202210588174.9, 生成渲染器内部数据结构的方法、装置、设备及存储介质；202210634563.0, 利用元数据对基于音床的音频进行渲染的方法及装置；202210762912.7, 共享渲染器组件的配置方法、装置、设备及存储介质；202210760302.3, 一种音床渲染项数据映射方法、装置、设备及存储介质；202210603204.9, 一种渲染器的渲染项确定方法、装置、设备及存储介质；202210600880.0, 一种音床输出渲染项确定方法、装置、设备及存储介质；202210603208.7, 一种对象输出渲染项确定方法、装置、设备及存储介质；202210603212.3, 一种场景输出渲染项确定方法、装置、设备及存储介质；202210603184.5, 音频渲染器的渲染项处理方法、装置、设备及存储介质；202210608202.9, 一种场景渲染项数据映射方法、装置、设备及存储介质；202210782056.1, 一种音频渲染器增益计算方法、装置、设备及存储介质；202210910129.0, 用于对象渲染器的元数据解析方法、装置、设备及介质；202210907370.8, 利用元数据对基于对象的音频进行渲染的方法及装置；202210912275.7, 利用元数据对基于场景的音频进行渲染的方法及装置；202211057713.2, 利用元数据对基于对象的音频进行渲染的方法及装置；202211063746.8, 利用元数据对基于场景的音频进行渲染的方法及装置；201510795213.2, 一种 3D 录音系统球面麦克风阵列分布方法；202011306975.9, 音频播放方法及装置、电子设备和存储介质；202210451743.5, 一种音频处理方法、装置、电子设备和可读存储介质。

本文件的发布机构对于该专利的真实性、有效性和范围无任何立场。

该专利持有人已向本文件的发布机构承诺, 他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。该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

通讯地址: 北京市海淀区颐和园路 5 号北京大学理科 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专利的责任。

信息技术 虚拟现实内容表达

第3部分：音频

1 范围

本文件规定了虚拟现实(VR)设备及相关系统中的沉浸式音频内容的表达方式,包括元数据结构和渲染器的系统架构及接口规范。

本文件适用于全景音频录播、沉浸式音频通信、虚拟现实音频交互等领域的音频采集、传输、渲染回放系统。

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 5271.1 信息技术 词汇 第1部分:基本术语

GB/T 5271.4 信息技术 词汇 第4部分:数据的组织

GB/T 5271.9 信息技术 词汇 第9部分:数据通信

GB/T 17191(所有部分) 信息技术 具有 1.5 Mbit/s 数据传输率的数字存储媒体运动图像及其伴音的编码

GB/T 17975.1 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

GB/T 33475.3 信息技术 高效多媒体编码 第3部分:音频

GY/T 316 用于节目制作的先进声音系统

ISO 639:2023 个别语种和语种组的代码(Code for individual languages and language groups)

ISO 3166-1 国家及下属地区名称代码 第1部分:国家代码(Codes for the representation of names of countries and their subdivisions—Part 1:Country code)

ITU-R BS.2051:2022 用于节目制作的高级音响系统(Advanced sound system for programme production)

ITU-R BS.2076:2019 音频定义模型(Audio definition model)

ITU-R BS.2088 带有元数据的音频节目素材国际交换的长文件格式(Long-form file format for the international exchange of audio programme materials with metadata)

ITU-R BS.2094:2017 音频定义模型的通用定义(Common definitions for the audio definition model)

ITU-R BS.2388:2024 音频定义模型和多声道音频文件的使用指南(Usage guidelines for the audio definition model and multichannel audio files)

3 术语和定义

GB/T 5271.1、GB/T 5271.4、GB/T 5271.9 和 GB/T 33475.3 界定的以及下列术语和定义适用于本文件。