UHD World Association
世界超高清视频产业联盟

UHD World Association
世界超高清视频产业联盟

UHD Word Association
www.theuwa.com

三维菁彩声（Audio Vivid）

技术白皮书

(V1.0)

世界超高清视频产业联盟

前 言

本文件由 UWA 联盟秘书处组织制订，并负责解释。

本文件发布日期：2022 年 8 月 29 日。

本文件由世界超高清视频产业联盟提出并归口。

本文件主要起草单位：

中央广播电视总台、中国电子技术标准化研究院、华为技术有限公司、北京抖音信息服务有限公司、央广新媒体文化传媒（北京）有限公司、赛因芯微(北京) 电子科技有限公司、中视广信科技有限公司、国家广播电视总局广播电视科学研究院、国家广播电视总局广播电视规划院、中国移动咪咕公司、中国移动智慧家庭运营中心、工业和信息化部电子第五研究所、深圳市腾讯计算机系统有限公司、北京爱奇艺科技有限公司、日本夏普株式会社、中国联合网络通信集团有限公司、TCL 实业控股股份有限公司、康佳集团股份有限公司、海信视像科技股份有限公司、深圳创维-RGB 电子有限公司、青岛海尔多媒体有限公司、四川长虹电器股份有限公司、京东方科技集团股份有限公司、广州视源电子科技股份有限公司、深圳市洲明科技股份有限公司、杭州当虹科技股份有限公司、北京数码视讯科技股份有限公司 、北京市博汇科技股份有限公司、优酷信息技术（北京）有限公司、北京百度网讯科技有限公司、北京小米电子产品有限公司、湖南广播电视台、广东广播电视台、北京流金岁月传媒科技股份有限公司、上海数字电视国家工程研究中心有限公司、浙江广播电视集团、浙江华策影视股份有限公司、中图云创智能科技（北京）有限公司、北京体育大学 5G 高新视频体育融合创新应用国家广播电视总局实验室、上海文化广播影视集团有限公司

本文件主要起草人：

姜文波、程多福、关朝洋、李岩、陈仁伟、李婧欣、王喆、黄传增、郭晓、吴健、徐建新、孙剑、宁金辉、周芸、焦健波、鹿楠楠、庞超、郑强、范胜利、张建东、柳德荣、吴强、王琦、潘兴浩、谢于贵、 李康敬、 程志鹏、 杨忠尧、单华琦、邢刚、韩建、郭佩佩、陈维、程剑、王鹏、林琳、张宏伟、杜正中、孙磊、朱子荣、李大龙、刘长滔、赵兴龙、苏运全、王海盈、徐遥令、童海、耿一丹、陈

迅、顿胜堡、泮利、陈纯丹、李广宙、熊昭民、赖凡、王子谦、吴晓东、傅斌星、李沄泓、韦胜钰、谭胜淋、陈勇、董杰、周骋、邹箭宇、谢杨、曾泽君、殷惠清、王雪辉、徐扬法、陈家兴、陈左乐、陈嘉欣、江建亮、王荣芳、李法、邢怀飞 、查丽、于磊、高伟标、秦宇 、唐迅

免责说明：

1， 本文件免费使用，仅供参考，不对使用本文件的产品负责。

2， 本文件刷新后上传联盟官网，不另行通知。

目录

1. 背景

音频技术随着人们对声音还原和品质的不断追求以及科技的创新而不断发展。从历史上第一台机械留声机的出现，发展到光学唱片、磁性录音和电子录音；从单声道发展到立体声、环绕声、三维声；音频处理技术从基于纯电子技术发展到和人工智能等最新处理技术的结合，依然处于不断发展中。从用户受众角度来看，近年来，随着超高清新技术（高分辨率、高帧率、高色深、宽色域、高动态范围、三维声）的不断演进，人们对“真实”、“沉浸”的音视频体验不断得以提升。三维声技术作为超高清体验的重要组成部分，带来空间感、方位感、高还原度、高沉浸度，带给观众更具感染力的临场感、个性化和交互体验。从产业发展角度，音频与视频产业的协同发展开始提速。如电视机发展从标清电视与立体声音频、高清电视与环绕声音频发展到超高清电视与三维声沉浸式音频配套发展；同时户外大屏也同步配套三维声技术，大力推动了产业的整体发展。因此现阶段抓住发展机遇，提升音频表现能力，快速追赶国际先进水平，同时也加快满足人们对高质量音频内容消费的需求，进而提升产业促进发展。因此，三维声技术的推广具有重要的经济和社会效应。

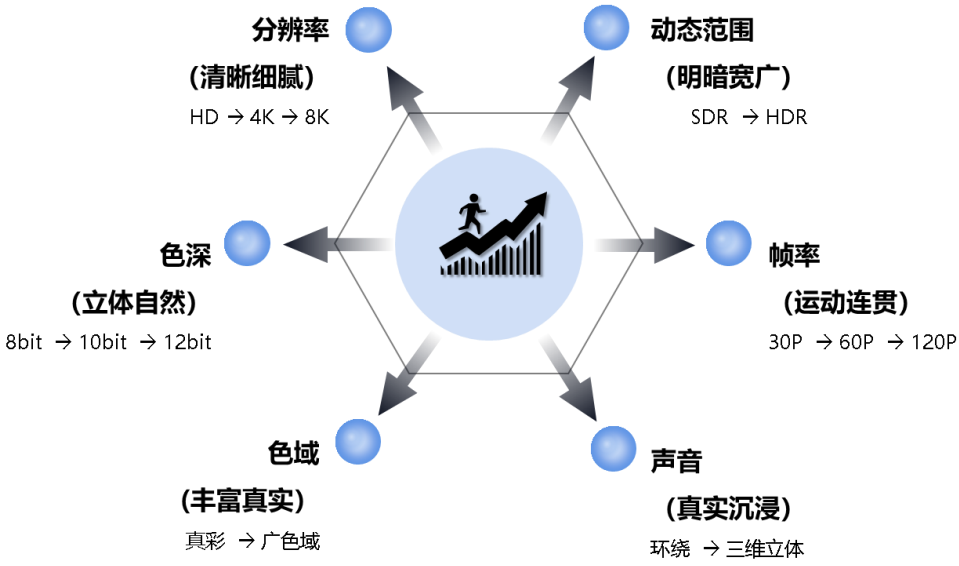

图 1 声音是超高清视频六维技术的重要组成之一

2. 技术简介

2.1 使用场景

三维声相对传统声音增加了空间感和方位感，使听众能再现在现实世界中所听到的声音，从而满足人们对声音高度还原，高度沉浸的体验需求，同时可具备个性化选择和交互体验。三维菁彩声（Audio Vivid）解决声音从构建到还原的整个环节，可在家庭环境、影院环境、个人、AR/VR 以及车载中得以应用。

2.2 声音重现技术

为实现重现声音的空间感和方位感，技术上可以依托三种形式。第一种基于声道的方式；第二种基于声床+对象的方式；第三种方式基于声场的方式。

1.基于声道的实现方式

该种方式需要对声道进行配置，将每个声道的音源映射到指定扬声器。该方法的局限性是一种声道配置对应一种扬声器布局，一种声道配置下的音频渲染只有在对应扬声器布局下才能获得最佳效果。

2.基于声床+对象的实现方式

声床承载基本环境声，对象是一系列单声道音频元素和对应元数据的集合。元数据表示对象的位置，强度，大小等信息。在回放时根据元数据信息，将对象映射到一个或多个扬声器或者双耳化渲染到耳机播放，以达到想要的空间音频效果。

3.基于声场的实现方式

声音作为一种波的属性，按照声波进行传递。对于给定时间的声场，每点均可以通过多个声波场的压力函数来表示。当获得该空间中该点的压力值时，便可对空间中的声音进行重构。为了确保和提升基于声场的音频质量，需要对该点的压力函数的各个系数准确获取以提高声场空间系数的编码质量。Higher Order Ambisonics（HOA） 是一种定义在球体表面上的 3D 声场建模格式，可以在任何设备（如耳机、扬声器、音箱）上对声场实现准确捕获、处理和重构。HOA 系统性能随着 HOA 阶数的增加而增加，但 HOA 信号数量也随之增加，因此需要信号处理技术和能力不断增加，如 AI 等新技术的引入。

4.扬声器重放

三维声重放环境在三维声音频制作中发挥着重要作用。扬声器使用数量的多少在很大程度上影响着最佳听音的区域范围，如扬声器数量少，最佳听音的区域范围也就越小，两者呈正相关。

5.双耳渲染

三维声的重放也可以依托耳机得以实现。为此，为了使这一技术的性能和优势得以充分发挥和体现，这需要对以上三种三维声实现方式进行双耳渲染，从而使普通立体声耳机可以聆听三维声音频。双耳渲染技术更好的推动三维声的普及, 也是 VR/AR 场景三维声呈现的主要渲染技术。

3. 三维菁彩声（Audio Vivid） 解决方案

3.1 三维菁彩声（Audio Vivid）引入

世界超高清视频产业联盟(UWA)牵头，与 AVS 编解码标准协同，联合产业端到端生态，推动发布三维菁彩声（Audio Vivid）技术团体标准草案，旨在快速推动超高清视频产业发展，提升超高清视频核心关键技术标准影响力。

对比业界现有的三维声技术，三维菁彩声（Audio Vivid）技术标准的目标是面向全球，技术先进，是一个更加开放的、具备产业安全要求的技术标准和方案，同时产业生态政策友好，更加适合超高清产业生态各方进行端到端的产业部署。在各方的联合支持下，现阶段三维菁彩声（Audio Vivid）技术标准已经完成了端到端的体系建设，将进入市场规模使用。

3.2 解决方案

三维菁彩声（Audio Vivid）涉及工具，芯片，电视机、手机等终端，平台等多方因素。 集成场景如下图所示：

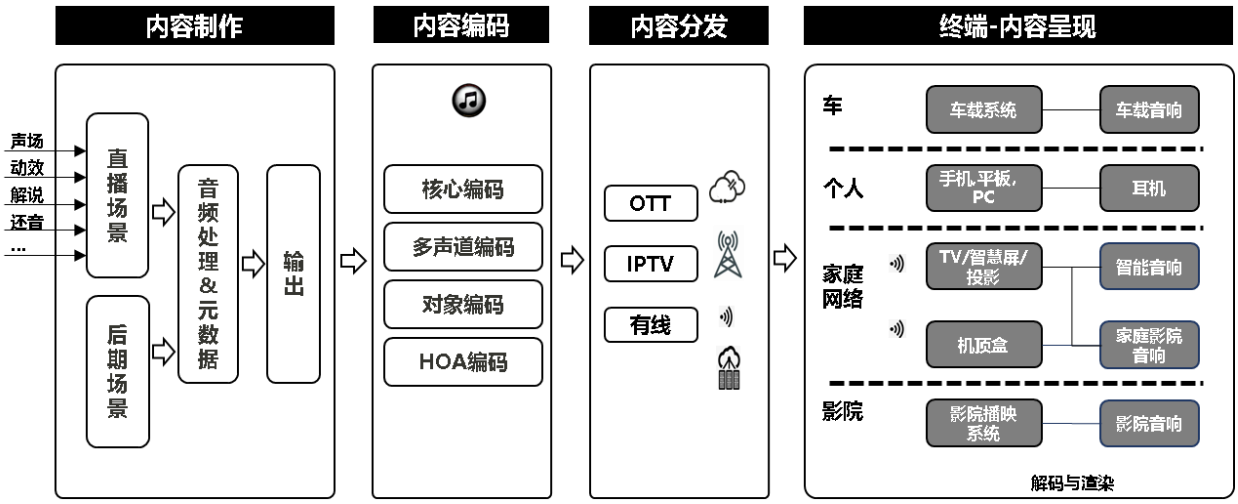

图 2 三维菁彩声（Audio Vivid）集成场景

其中三维菁彩声（Audio Vivid） 标准覆盖内容生成和内容重放相关的编解码和渲染部分。目前标准涉及的端到端生态包括：内容制作，内容编码，终端解码与渲染等。

3.3 技术原理

三维菁彩声（Audio Vivid） 针对不同的信号类型采用不同的技术工具对输入信号进行编解码。采用多声道编码技术去除多声道信号间的信息冗余。采用 HOA 空间编码技术去除 HOA 各声道信号间的空间几何信息冗余。采用基于心理声学模型的预处理和基于 AI 的量化，熵编码技术去除单声道、对象音频信号中的信息冗余。通过扬声器或耳机完成最终渲染输出。详细可参考 UWA 联盟发布的三维菁彩声（Audio Vivid）技术标准。

3.4 关键特性

三维菁彩声（Audio Vivid）技术作为全球领先的超高清音视频技术之一，具备以下关键特性

1. 支持单通道，立体声，环绕声，三维声（多声道声床，音频对象，Ambisonic 声场）
2. 支持有损和无损编解码
3. 支持扬声器和双耳渲染
4. 支持 Three degree of freedom(3DoF)呈现
5. 支持 HOA 空间编码工具，大幅提升 HOA 信号的编码效率
6. 解码器复杂度与业界标杆相当

7. 支持 16 个通道的编解码（HOA 最大支持到 3 阶，可扩展至更高阶）
8. 采样率支持 32kHz~192kHz
9. 量化精度支持 16 比特和 24 比特
10. 速率支持 32kbps~1.6Mbps
11. 编解码算法时延小于 80ms（典型场景）
12. 最大支持 7 阶 HOA 双耳渲染,支持 128 轨音频实时渲染
13. 双耳渲染时延小于 40ms（典型场景）

3.5 编解码方案

三维菁彩声（Audio Vivid）音频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。[1]

编码器由多种编码工具构成，如图 3 所示，包括：通用全码率音频编码工具和无损音频编码工具。

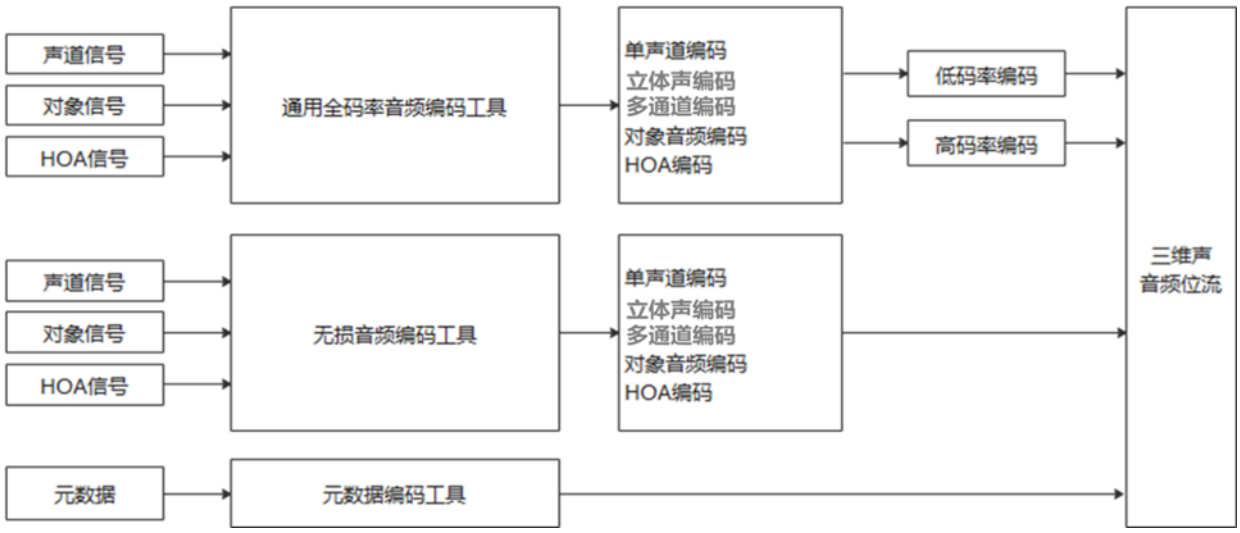

图 3　三维声音频编码器示意图

针对不同特征的音频信号或不同的应用场景，用户可以根据输入类型和码率范围，选择使用通用全码率音频编码、无损音频编码工具和元数据编码工具。

1. 通用全码率音频编码工具

采用神经网络变换、量化和熵编码技术，基于声道相关性的多声道下混和比特分配技术，基于虚拟扬声器的 HOA 空间编码技术等，适用于单声道、立体声、多声道编码、对象音频编码、混合音频编码、HOA 编码。

支持采样率 32kHz～192kHz，支持 16 比特和 24 比特采样精度。支持编码输出位流为：

1) 单声道/对象：32、44、56、64、72、80、96、128、144、164、192kb/s；

2) 立体声：32、48、64、80、96、128、144、192、256、320kb/s；

3) 5.1 多声道：96、128、144、160、192、256、320、384、448、512、640、720kb/s；

4) 7.1 多声道：128、160、192、256、384、480、576、640kb/s；

5) 5.1.4 三维声：176、256、384、448、576、704kb/s；

6) 7.1.4 三维声：240、384、512、608、832kb/s；

7) FOA：96、128、192、256kb/s；

8) 2 阶 HOA：192、256、320、384、480、512、640kb/s；

9) 3 阶 HOA：256、320、384、512、640、896kb/s；

10) 声床+对象混合信号：以上立体声/多声道/三维声信号和单声道/对象信号码率的各种组合；

特别的，为了确保三维菁彩声（Audio Vivid）商用上线系统具备必要的音频质量，本白皮书针对一些典型类型音频信号给出编码码率建议如下：

1) 5.1：不小于 192kbps，针对极高音质应用不小于 320kps

2) 5.1.2：不小于 320kbps，针对极高音质应用不小于 480kbps

3) 5.1.4：不小于 384kbps，针对极高音质应用不小于 576kbps

4) 7.1.4：不小于 384kbps，针对极高音质应用不小于 608kbps

5) 每个对象：不小于 96kbps，针对极高音质应用不小于 164kbps

2. 无损音频编码工具

无损音频编码工具支持最多 128 声道、任意采样频率。并支持 8 比特、16 比特和 24 比特采样精度。

3. 元数据编码工具

对音频元数据信息进行编码。

4. 解码和渲染系统

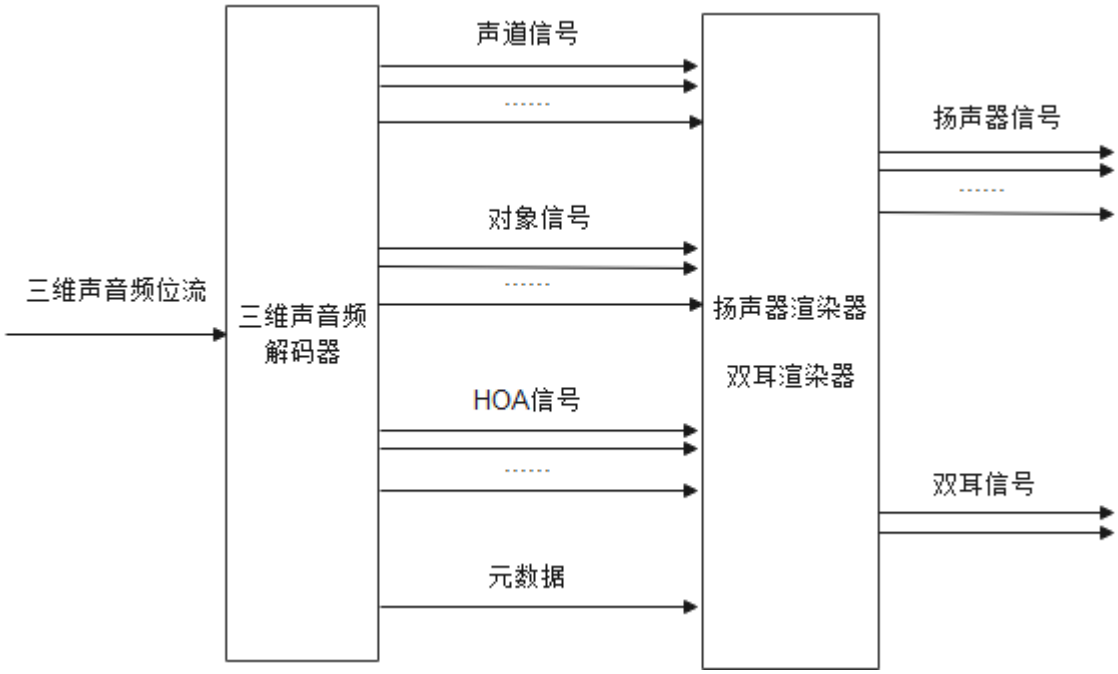

图 4　三维声音频解码和渲染系统框架示意图

5. 通用全码率音频编解码示例

通用全码率音频编解码包括单声道编解码、立体声编解码、多声道编解码、对象编解码、混合编解码、HOA 编解码和元数据编解码。

5.1 编码器架构

通用全码率音频编码器的基本构架如图 5 所示。通用全码率音频核心编码器由暂态检测、窗型判断、时频变换、频域噪声整形、时域噪声整形、频带扩展、下混、神经网络变换、量化和区间编码等构成，将声道信号和对象信号编码为位流。HOA 空间编码器和核心编码器将 HOA 信号编码为位流。元数据编码器将元数据编码为位流。

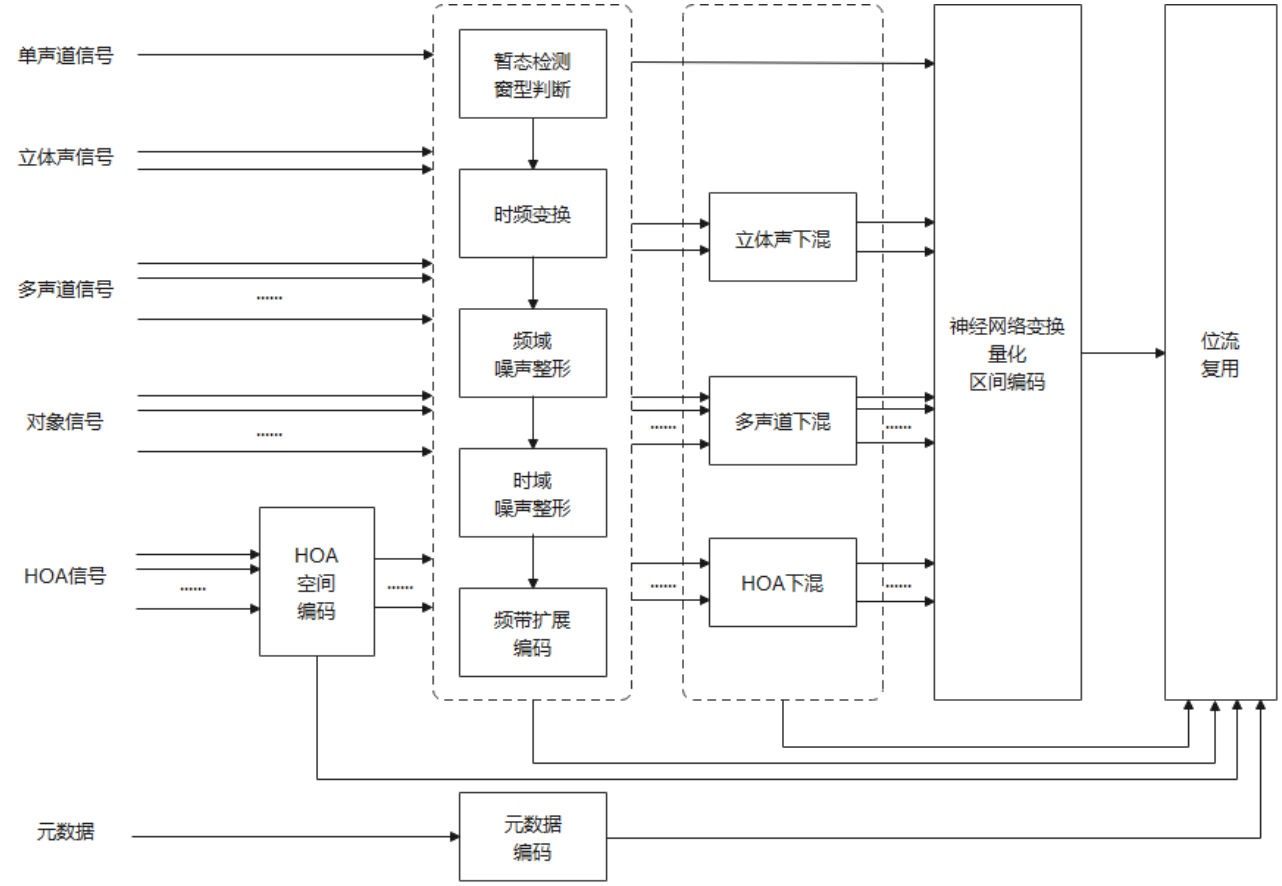

图 5 通用全码率音频编码器框架

5.2 解码器架构

通用全码率音频解码器的基本构架如图 6 所示。通用全码率音频解码器由区间解码、逆量化、神经网络逆变换、频带扩展解码、逆时域噪声整形、逆频域噪声整形、上混和逆时频变换等构成了核心解码器，将位流解码为声道信号和对象信号。HOA 空间解码器和核心解码器将位流解码为 HOA 信号。元数据解码器将位流解码为元数据。

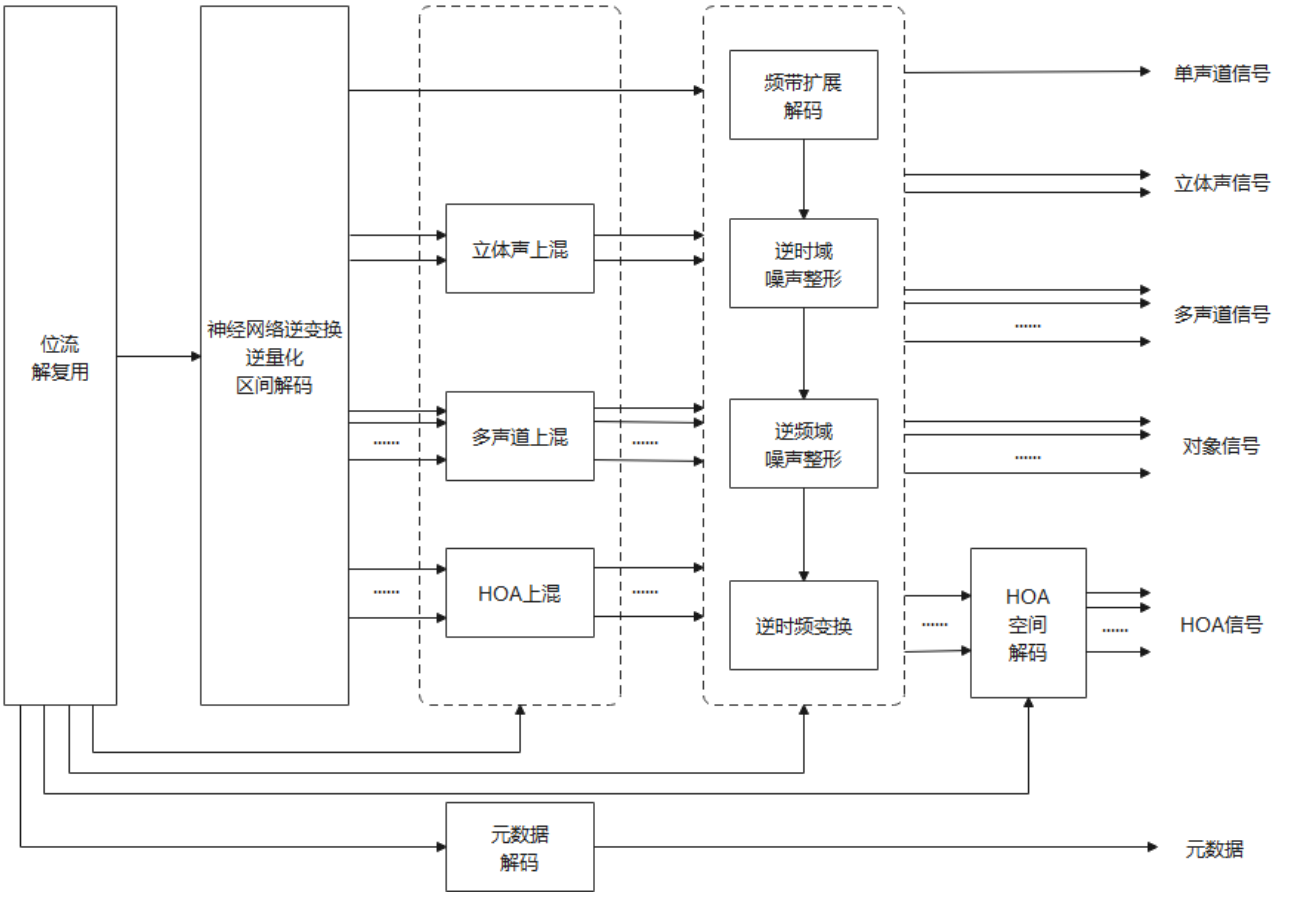

图 6 通用全码率音频解码器框架

5.3 流程

通用全码率音频编码器可以分为编码预处理、各模式信号下混、神经网络变换、量化和区间编码以及元数据编码。编码预处理将每个声道信号由时域变换到频域并进行预处理。信号下混根据不同编码模式对预处理后的频域信号进行下混，去除声道间的相关性。神经网络变换、量化和区间编码采用神经网络对每个下混后的声道进行变换和编码。元数据编码按照元数据结构、根据每个元数据的量化规范对元数据进行量化编码。通用全码率音频解码器可以分为编码后处理、各模式信号上混、神经网络逆变换、逆量化和区间解码以及元数据解码几个部分。解码是编码的逆过程。

3.6 元数据方案

三维声音频元数据兼容 ITU-R BS.2076 标准定义的音频模型，并增加扩展元数据。元数据系统由两部分组成——基础元数据部分<AudioformatExtended>和扩展元数据部分<VRext>。其中，基础元数据部分引用 ITU-R BS.2076-2 标准，扩展元数据部分为本文件的新增定义。基于此架构，本文件的元数据系统既能

前向兼容, 又能后向扩展，在满足元数据全球互联互通需求的同时, 又提供了足够的灵活性和可扩展性，能够为本文件的沉浸式音频系统提供强大的表征能力。该元数据系统架构和详细说明参考标准图 7。

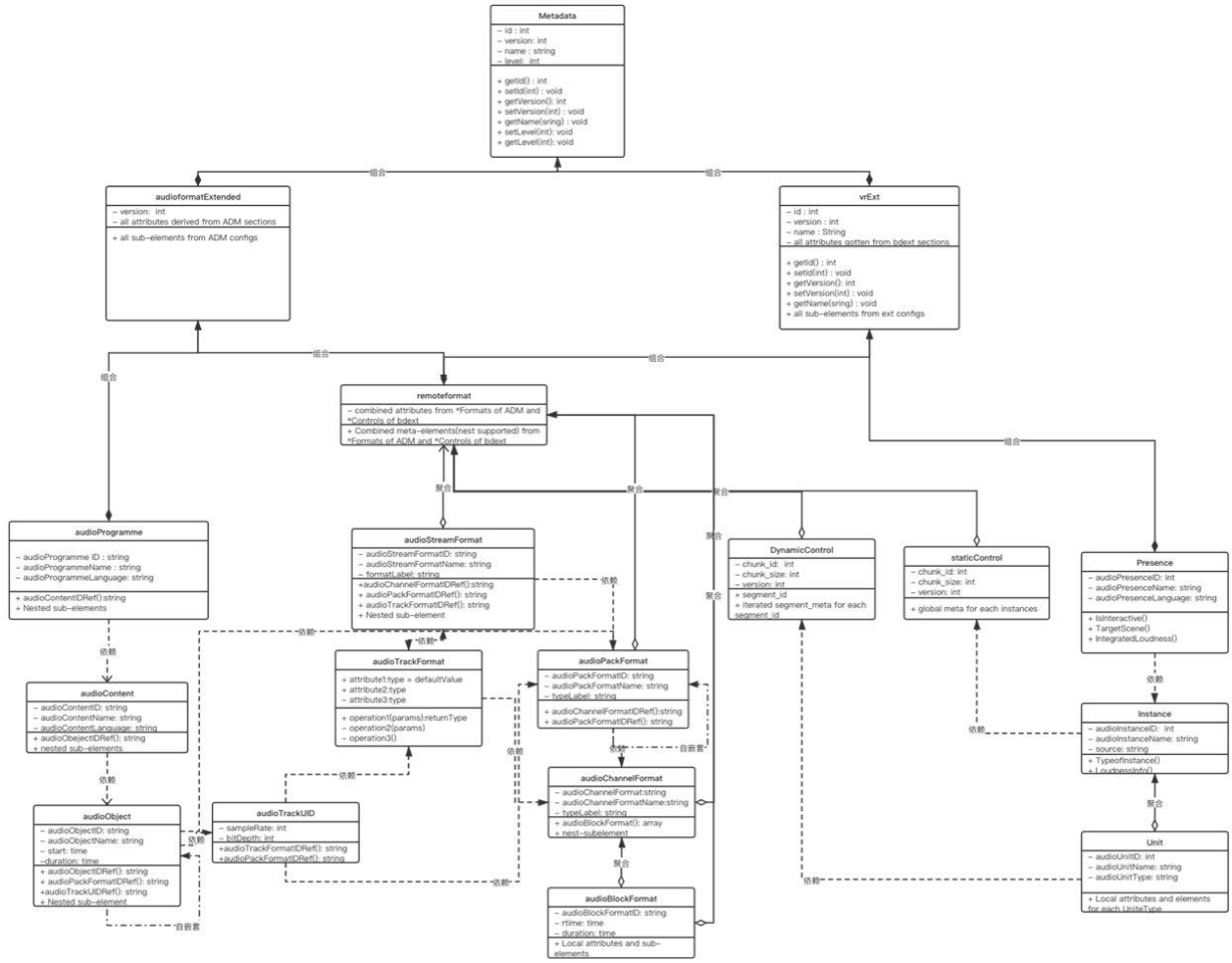

图 7 元数据系统架构图

三维菁彩声（Audio Vivid）详细编解码以及其他示例参考附录已发布的编码标准规范。

3.7 渲染方案

3.7.1 扬声器渲染

三维菁彩声（Audio Vivid）渲染器支持制作端音频内容制作、终端渲染回放。既可用于直接解析 ITU-R BS.2076 ADM 元数据，也可解析第三方的元数据；在音频评估、制作和监听期间，为 ADM 元数据解析提供一个开放的参考渲染器，将有利于音频内容生态系统的整体健康。

应用和处理元数据和相关音频数据，体现以下渲染原则：

1) 正确处理响度相关的元数据，保持目标响度

2) 保留艺术意图，如正确应用传输的上混、下混矩阵

3) 适当地呈现对象的空间属性（位置、空间范围等）

4) 正确处理漫反射、发散角方位距离和排除扇区元数据

5) 正确处理高优先级的音频对象

扬声器渲染器总体架构由上图几个核心组件和处理步骤组成，包括：

1) 将 ADM 元数据转换为可渲染的 TypeMetadata

2) 根据用于渲染的 TypeMetadata 进行增益计算等

根据 TypeMetadata 的 TypeDefinition，渲染过程分为：

✓ 基于对象的渲染

✓ 基于直接扬声器信号的渲染

✓ 基于 HOA 的渲染

矩阵类型在创建渲染项期间处理，并且作为直接扬声器类型渲染器的一部分；而双耳类型也作为直接扬声器类型中的耳机输出部分。

1. 基于对象的渲染

基于对象的渲染原理如下：

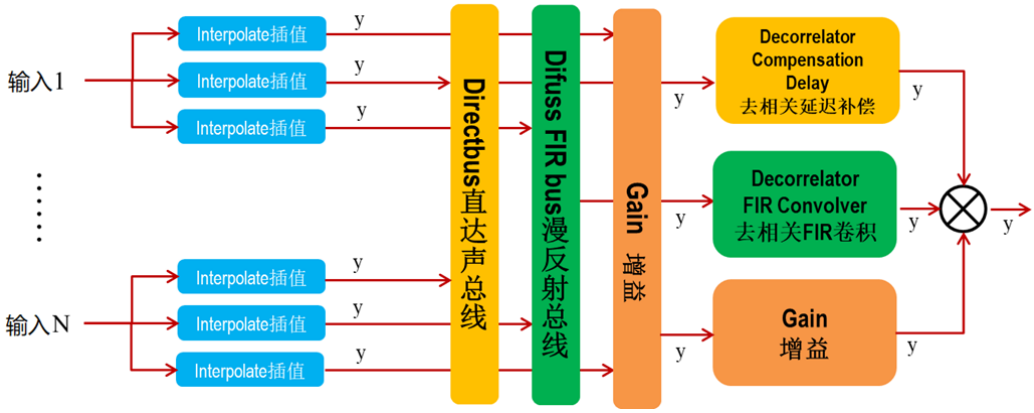

图 8 基于对象的渲染

在通过对象输出接口请求对象音频数据和元数据的输出的情况下，绕过漫反射处理提供元数据和音频数据的输入。元数据以 objectsTypemetadata 对象的输入形式进入渲染器；通过对元素元数据预处理模块的处理，经过 objects_gains 计算，获得基于对象内容的元数据和音频数据的输出， 当请求接口输出数据时，这些元素将启用以进行播放。

2.基于声道的渲染

基于声道 DirectSpeakers 的渲染原理如下：

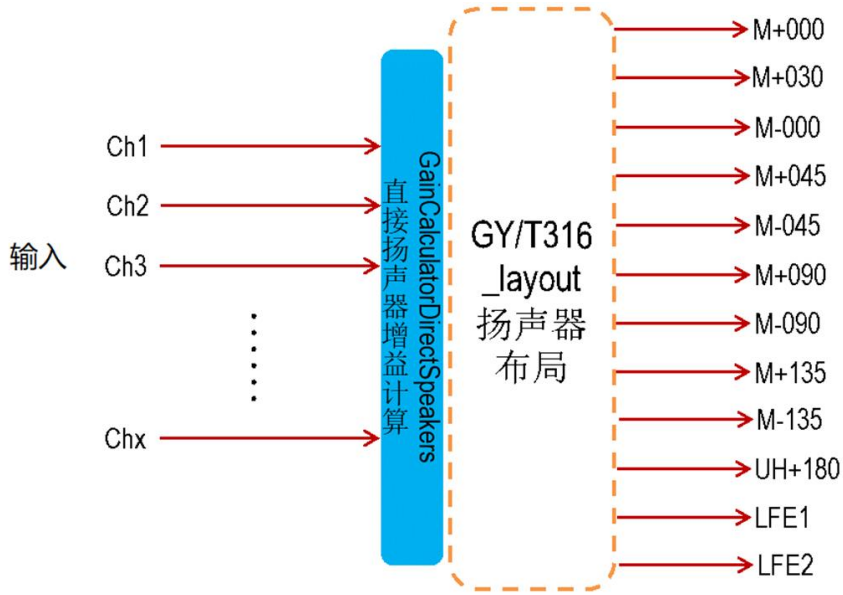

图 9 基于声道的渲染

使用 Gain_calculator_Direct_Speakers 为 DirectSpeakers 声道计算的增益时，增益将直接应用在音频输入声道，直接生成输出音频频道。DirectSpeakers 元数据不应是动态的，因此增益不应在 audioBlockFormat 内插值，但如果用户更改元数据，则应插值增益。

增益计算时，输入的声道数与输出回放的扬声器布局声道数无关，无论输入声道数与输出声道数是否相等，渲染器按 BS.2051 的输出声道数进行渲染。当输入的 DirectSpeaks 数量少于输出的时候，会发生上混的增益计算，当输入的 DirectSpeakers 数量多于输出声道的时候，会计算出下混的增益。对于基于矩阵的音频类型，因矩阵音频类型基本信号为中间和侧面周围两个信号声道，故一般对侧面周围信号进行上混。

3.基于 HOA 的渲染

基于 HOA 的渲染原理结构如下：

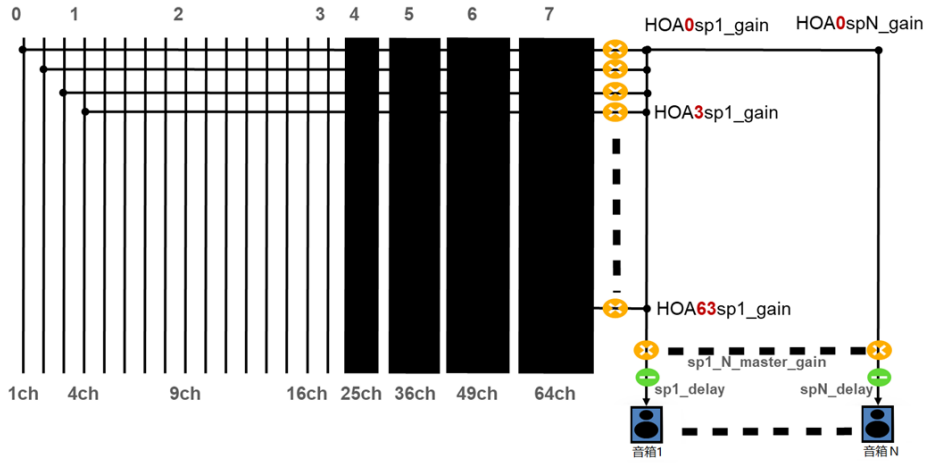

图 10 基于 HOA 的渲染

本渲染器选定 AllRAD 算法的 HOA 解码方法。通过 HOA 的解码器 D 计算出每一 HOA 轨道给每一个音箱的增益值 HOAnspN_gain；将多轨 HOA 信号输出给独立的音箱。使用 Gain_calculator_HOA 为 HOA 声道计算的解码矩阵 D 直接应用于输入音频信道，以产生输出音频信道。与 DirectSpeakers 一样，HOA 元数据不应是动态的（即每个 audioChannelFormat 中应有一个 audioBlockFormat），因此增益不应在块内插值，但如果用户更改元数据，则应插值增益。在接口上显示之前，调整接收信号的延迟 spN_delay，使其与常规输出信号具有相同的延迟。HOA 接口和 HOA 渲染器不同时运行：如果请求通过 HOA 接口输出，则 HOA 渲染器的输出应静音，而不再现音频，反之亦然。

3.7.2 双耳渲染

本音频渲染系统会使用 3.7 节元数据系统中描述音频内容和渲染技术的控制信息, 比如音频载荷的输入形式是单通道, 双声道, 多声道, 还是音频对象或者 HOA, 以及动态的声源和听者的位置信息，渲染的声学环境信息如房屋形状, 大小, 墙体材质等。核心渲染系统依据不同的音频信号表示形式和从元数据系统中解析出来的相应元数据, 作相应播放设备和环境的渲染。系统框架见下图：

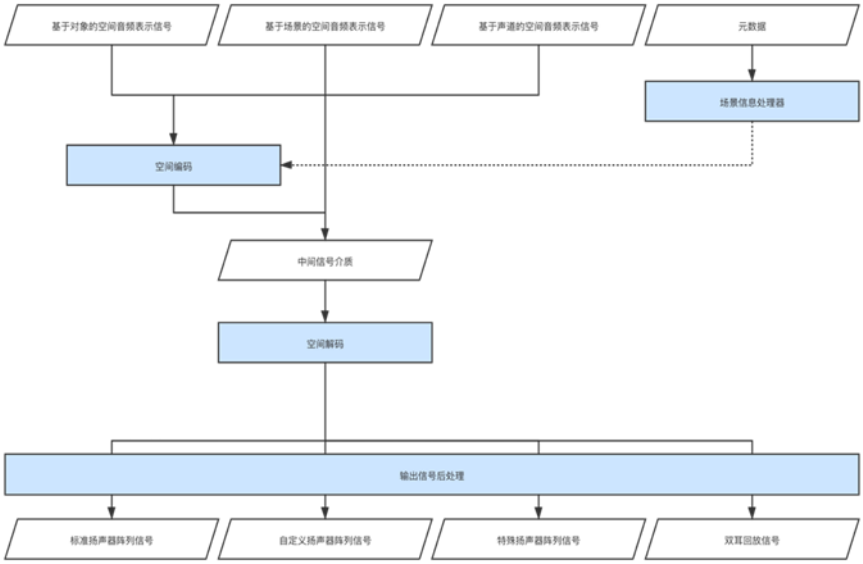

图 11 渲染器总体框架

处理链路引擎会以块（例如：1024 个采样点）为单位处理音频数据，块的大小应在引擎初始化时决定并不再更改。在实时处理链路上，应先对元数据进行解析并针对元数据调整场景信息。下面分别对多声道音频，对象音频，HOA 音频的渲染加以描述。

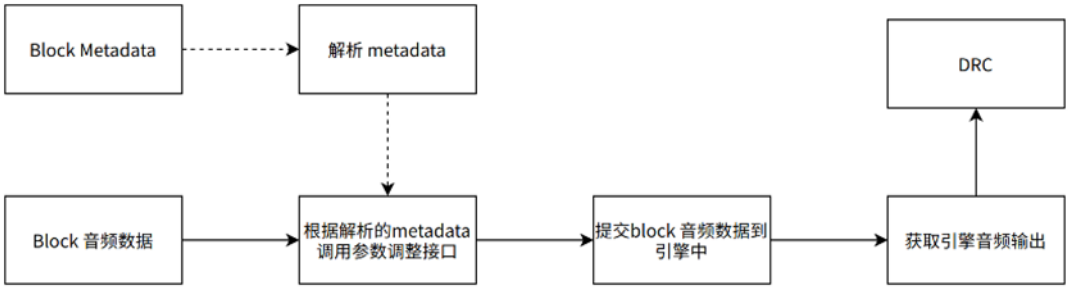

图 12 处理链路图

1.HOA 信号的渲染

渲染器目前支持至多 7 阶的 HOA 信号渲染。由于 ADM 中 HOA 信号的声道是分离的并标记了对应 order 和 degree。目前引擎支持 ACN 的声道排布方法，输入 FuMa 排列的声道暂不支持，需要在输入前进行 order 和 degree 的转换。对于正规化方法，引擎支持 N3D 和 SN3D 两种正则化方式，FuMa 的正则化需要预先进行转换。

渲染器支持同时渲染多个 HOA 信号，在 block 处理时如果有多个 channel 有同样的 order 和 degree 时，这两个信号会被混合。

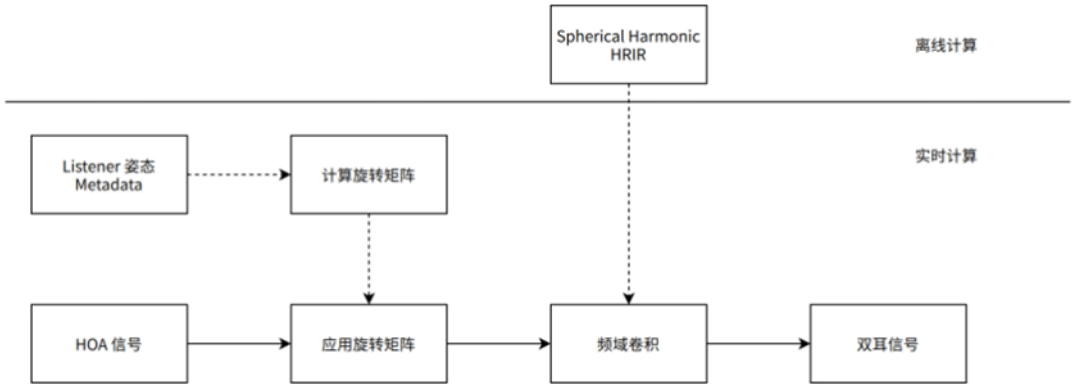

图 13 HOA 信号渲染图

上图总结了通过双耳回放 HOA 信号的流程。渲染器中自带了一组 1 ~ 7 阶的球谐 HRIR。在渲染过程中，首先通过 listener 姿态确定旋转矩阵，并转换 HOA 信号，在针对每个 HOA 的声道，和对应的 HRIR 进行卷积。

2.对象信号渲染

渲染器会将 Object 信号统一编码为 HOA 信号，然后复用 HOA 信号的渲染流程进行对象音频的双耳化渲染。

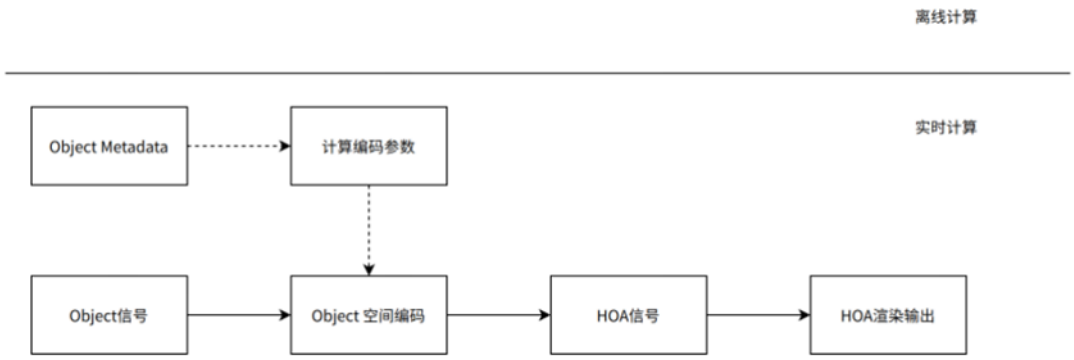

图 14 对象信号渲染图

3.多声道信号渲染

多声道音频渲染成双耳音频可以借鉴对象音频的渲染，将多声道音频当做多个对象音频，然后通过 HOA 信号渲染的方法进行双耳回放。

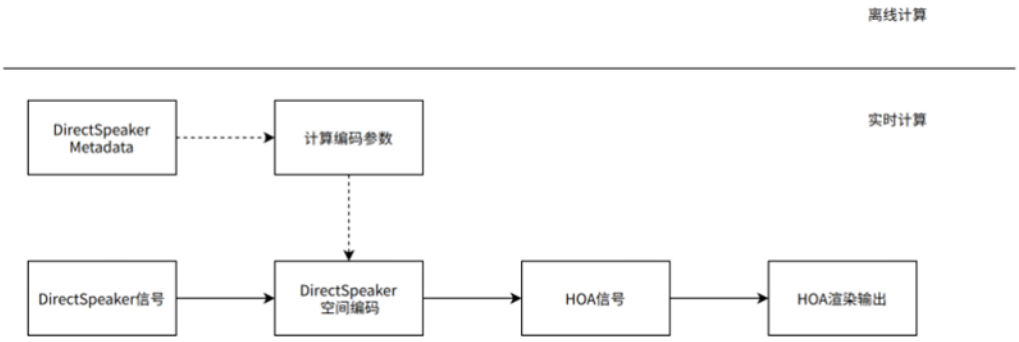

图 15 多声道信号渲染图

4. 三维菁彩声（Audio Vivid）用户体验

4.1 扬声器空间布放示例

以扬声器 5.1.4 声道为例（5.1.4 作为家庭影音的基础配置，在传统 5.1 环绕声系统(左右,中置, 左右后, 低音)的基础上, 增加 4 个头顶扬声器(左右上前, 左右上后)。如图 16 所示：

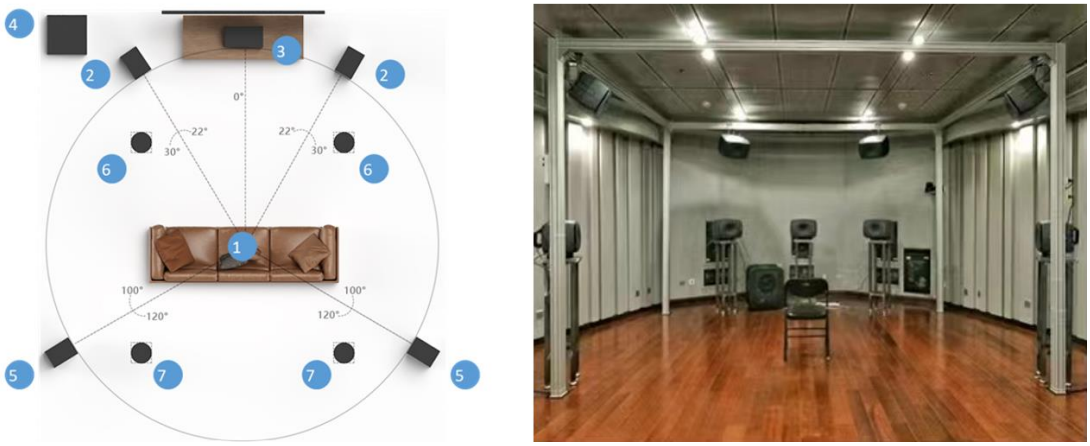

图 16 通用演示示意图

通过空间，设备，视音频文件的播放体验，使受众明确感知三维声的立体感，环绕感，空间感。

1. 最佳听音位
2. 左右扬声器

3. 中置扬声器
4. 低音扬声器
5. 左右后环扬声器
6. 左右前部天空扬声器
7. 左右后部天空扬声器

4.2 双耳空间声演示示例

声音进入耳朵时，在耳廓、头部附近时传播路线发生变化，到达左右耳的时间上也有微小的差别，所以真正听到的声音与原始音源并不一样，如下图所示：

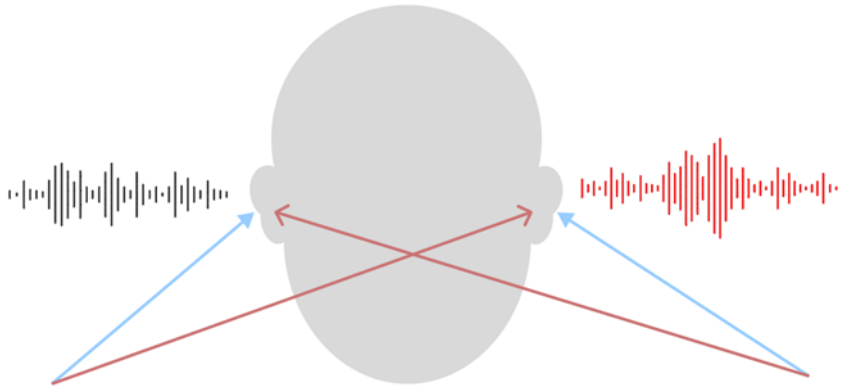

图 17 双耳渲染演示示意图

通过双耳渲染，可以使佩戴耳机时与不戴耳机获得一样的听觉感受，具有三维声空间感和沉浸感，也具备真实环境的房屋声学效果。

4.3 主观体验与验证

基于扬声器方案的通用体验示意如下图

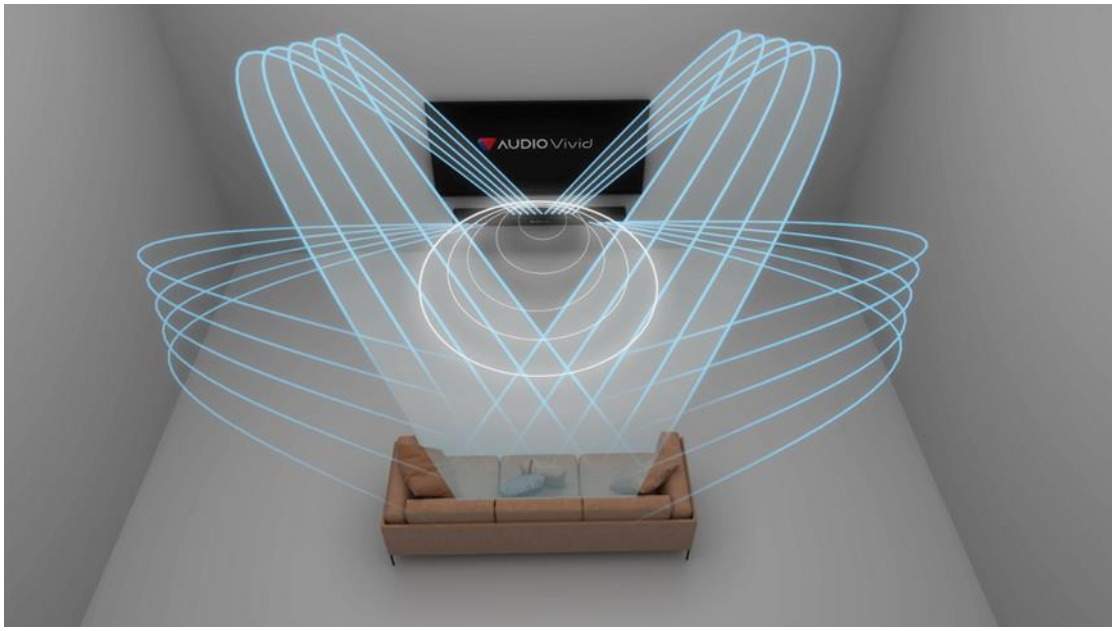

图 18 通用三维声体验示意图

通过受众体验，可明确感知方位，环绕，高度声音信息。并可通过以下进行主观评价。

功能点	评价点
音质	清晰度
	平衡度
	明亮度
	失真度
	亲切度（丰满度，圆润度，柔和度）
	动态感
空间感	混响感知
	空间感知
方向感	垂直度
	水平度
距离感	响度
	声压级别

声源体积	宽度，深度
外化感	（耳机）

表 1 主观评价表参考

5. 三维菁彩声（Audio Vivid）内容制作渲染器与使用示例（基于扬声器）

5.1 环境搭建

内容制作工具分为两种：作为 DAW 的插件渲染器和硬件渲染器设备.

作为 DAW 插件渲染器示例图片：

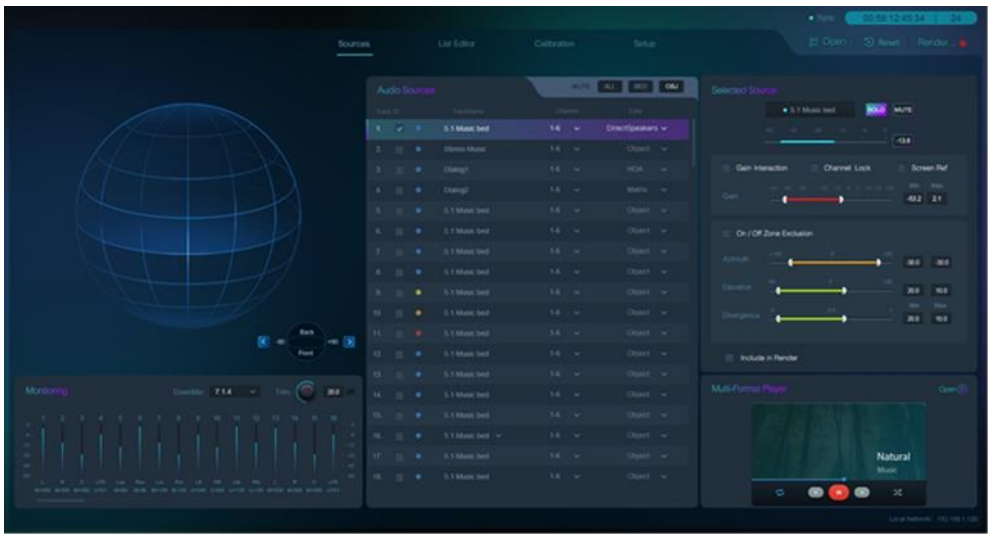

图 19 DAW 插件渲染器示例

硬件渲染器工作流程图：

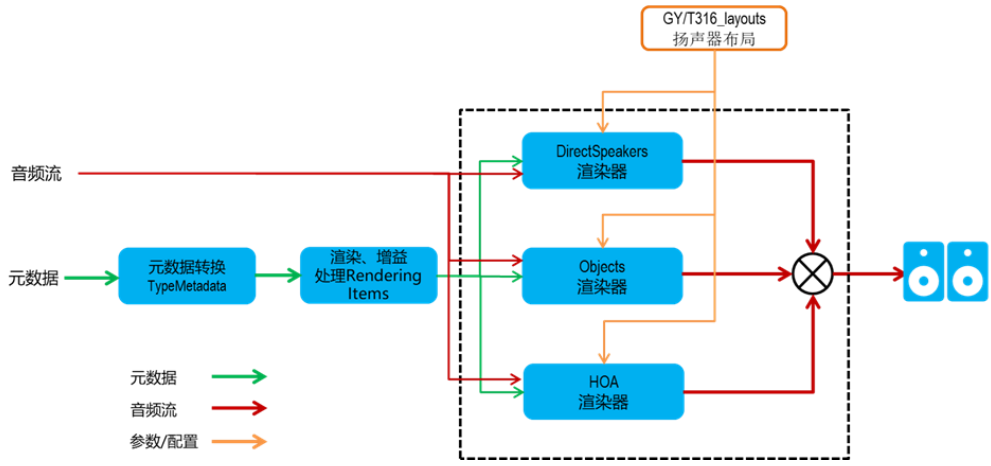

图 20 硬件渲染器工作流程图

5.2 工作流程

从制作端到播放终端工作流程如下图所示：

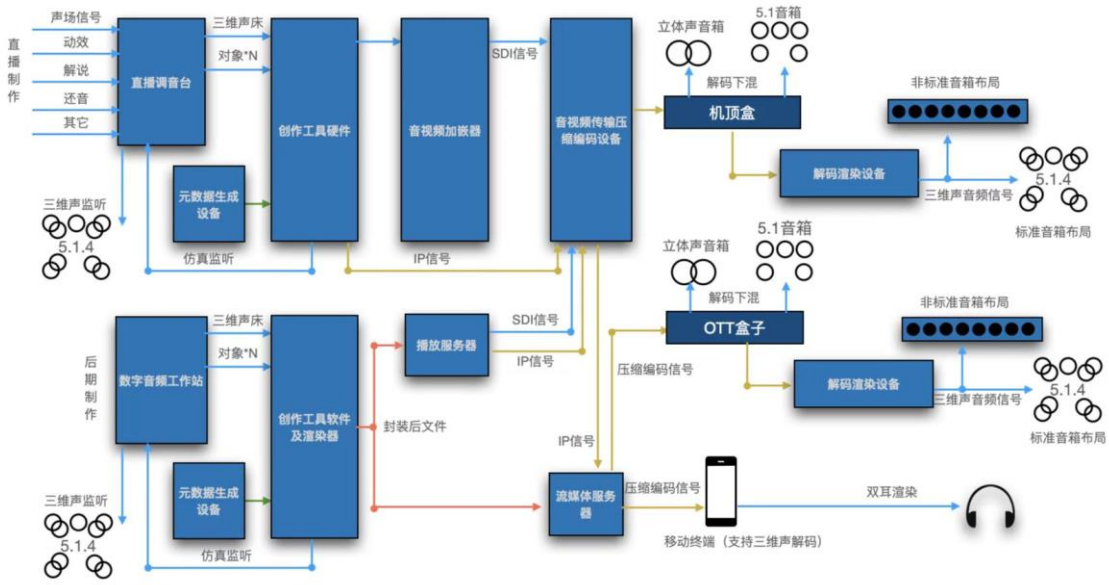

图 21 工作流程示例

5.3 操作步骤

1) DAW 中创建音频节目混音，执行内容制作时的编辑和信号处理。

2) 将音频声道路由到新的软件渲染器或渲染设备，而不是路由到 DAW 中的的输出总线。

3) 在 DAW 的每条音轨，使用内置的 3D Panner（如果可用）在 3D 空间中定位。

4) Panner 信息作为单独的元数据与音频信号一起发送到渲染器。

5) 将 Studio 扬声器（环绕高度扬声器）连接到渲染器的输出。渲染器软件实时处理音频信号及其单独的 pan 信息，将其作为扬声器上的合适的 3D 声场播放。

6) 将混音“存储”到一个特定的 ADM 文件中，仍然将音频信号和它们的 panning 信息分开。

7) 此后可以在渲染器中打开 ADM 文件以进行监听、编辑 ADM 元数据配置或将其导出到各种媒体文件。

8) 加载到渲染器中的 ADM 文件（代表 DAW 混音）可以导出为各种媒体文件格式，用作各种分发渠道的可交付文件。

9) 交付文件必须编码为特殊的比特流格式，以便通过流式服务将 DAW 混音发送给最终用户。

10) 最后，消费者接收 DAW 混音作为编码文件，该文件仍然保持音频和 pan 信息分开，回放设备对可用的扬声器布局或耳机执行实际渲染。

6. “百城千屏” 三维菁彩声（Audio Vivid）应用案例

6.1 “百城千屏”简介

"百城千屏”是工业和信息化部、中央宣传部、交通运输部、文化和旅游部、国家广播电视总局、中央广播电视总台等六部门联合部署开展的超高清视频落地推广活动。目前，在全国 35 个城市已有 100 多块超高清大屏落地，公众可通过公共大屏观看中央广播电视总台 8K 超高清电视频道。同时为了配合开展“百城千屏”超高清视频落地推广活动，在原有超高清视频观看的基础上，在不造成声音干扰的前提下，通过移动终端提供超高清视频的伴音服务—“百城千屏随身听”。2022 年 8 月，总台使用三维菁彩声（Audio Vivid）技

术，实现在“百城千屏随身听”移动端平台，可收听视音频精确同步的三维声信号。9 月 10 日，总台通过“百城千屏”播出平台，首次使用三维菁彩声（Audio Vivid）音频编解码同步播出中央广播电视总台中秋晚会，在喜庆的氛围里，给公众带来一场视听盛宴。

6.2 “百城千屏随身听”平台

“百城千屏随身听”平台主要解决面向移动端的电视伴音的编码、传输，实现了电视频道音视频内容分别传输，协同呈现，极大地提升了户外大屏用户的收视体验。随身听系统提供的音频服务已经与 2022 年 8 月 18 日由立体声升级为三维菁彩声（Audio Vivid）三维声，实现三维声直播的商用落地。系统结构如下图：

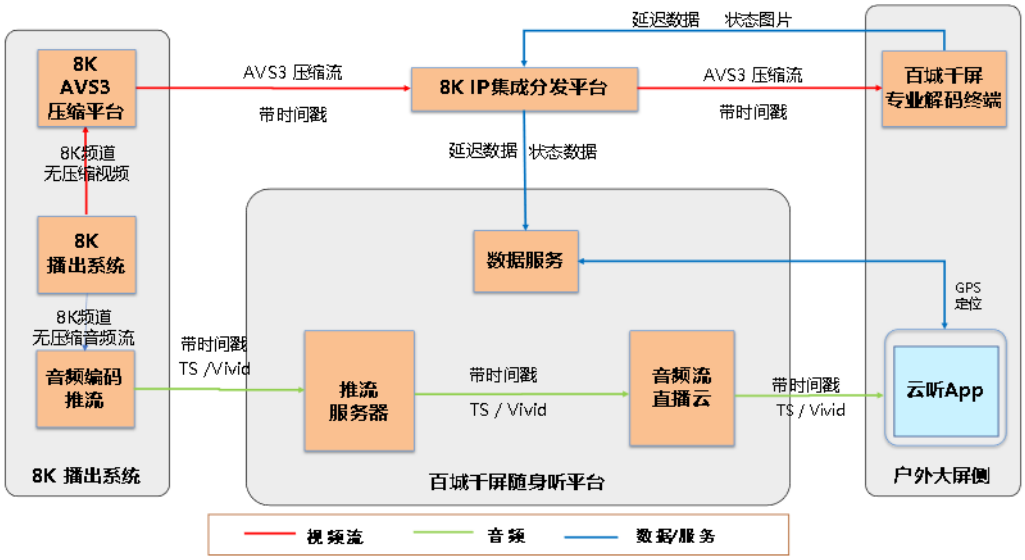

图 22 “百城千屏随身听”系统架构

“百城千屏随身听”平台已经实现了三维菁彩声（Audio Vivid）音频在编码、封装、传输、分发全链路的落地，并通过随身听客户端实现解码和双耳渲染的播放。其中重点解决问题如下：

1、研制了基于 X86 架构服务器的广播级三维菁彩声（Audio Vivid）编码器，实现支持多种协议封装和信源格式能力，实现在电视台 8K 播出系统中输出三维菁彩声（Audio Vivid）三维声音频流。

2、研制了基于 TS 封装适用于互联网及移动互联网的三维菁彩声（Audio Vivid）分发传输服务器，完成了三维菁彩声（Audio Vivid）音频流的封装、推流及流媒体服务等项功能，实现了通过互联网为用户提供三维菁彩声（Audio Vivid）音频服务。

3、由于“百城千屏”项目中音视频需要通过不同的链路传输，在不同的终端播出，所以项目在落地过程中有针对性的设计了音视频同步方式，采用异构网络视音频同步传输技术，在 AVS3 视频流中嵌入了播出绝对时间的时间戳，在解码终端解码时结合本地时间计算链路延迟并上报数据，同时通过移动终端的定位信息，平台为终端提供对应的解码终端的延迟时间。解码终端可以通过该延迟及三维菁彩声（Audio Vivid）音频流中的时间戳实现音视频同步功能，达到了国家标准规定的-90ms～120ms 的终端用户视音频同步要求。

6.3 “百城千屏随身听”终端

“百城千屏随身听”终端集成在“云听”客户端中。“云听”是中央广播电视总台推出的以 5G 技术作为支撑的移动音频客户端，以资讯、知识、文化为内容战略方向,集纳总台精品内容,自制音频 IP 节目,创作优质有声书,致力于为手机、车机、平板电脑、智能穿戴等多终端用户提供全场景的声音。

在“百城千屏”项目中，云听客户端首先通过扫码或用户选择关联到具体的大屏，然后通过总台的媒体网关接口获取针对该大屏的音频流 URL 和对应的视频解码播放延时 tvideo,然后再通过获取的 URL 读取经过三维菁彩声（Audio Vivid）编码和 TS 封装的音频流数据，最后经过接封装、解码、渲染等步骤完成大屏三维菁彩声（Audio Vivid）音频信号的播放工作。具体的播放流程如下：

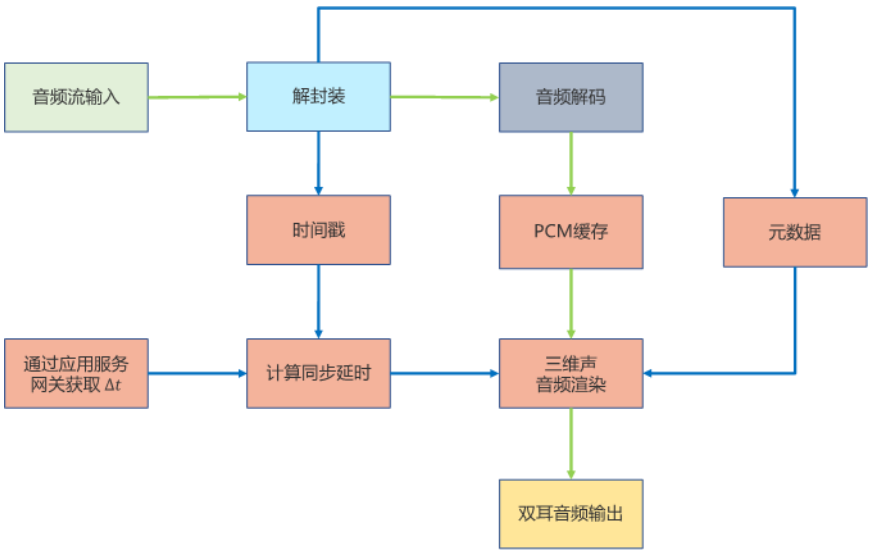

图 23 “百城千屏随身听”云听客户端播放流程图

1.音频流输入，音频播放器根据 URL 使用的网络协议获取到经过三维菁彩声（Audio Vivid）编码和 TS 封装的数字音频流；（目前使用的传输协议是 HTTP）

2.音频解封装，播放器从 TS 音频流中解封装出：三维菁彩声（Audio Vivid）音频编码数据、编码时间戳信息和音频元数据（如果包含有）；

3.音频解码，将解封装的三维菁彩声（Audio Vivid）音频编码数据解码成多声道 PCM 数据，然后存入渲染播放缓存队列；

4.音画同步，播放器周期性从总台媒体应用网关获取所关联大屏的视频解码播放延时 tvideo,同时播放器根据解封装获取的编码时间戳与本地时间戳计算差值，获得音频解码延时 taudio,并根据 tvideo 与 taudio 之间的差值计算出音频需要调整的同步延时时间来调整渲染播放缓存队列。

5.三维声渲染，播放器从渲染播放缓存队列中获取待播放的多声道 PCM 数据，然后根据包含了声床信息的元数据信息进行双耳音频渲染，并将渲染后的双声道 PCM 数据输出到声卡设备。

7. 三维菁彩声（Audio Vivid）发展建议

构建开放产业生态

吸引国内外尽可能多优质产业资源加入三维菁彩声（Audio Vivid）产业生态构建中，并促进国际标准的迭代。加快三维声技术标准在广播电视、新媒体、影视娱乐等方面的应用推广，扩大技术规范在垂直行业的影响力。推动产业链上下游深入合作，打通端到端产业链。

以应用示范推动三维菁彩声（Audio Vivid）普及

以重大赛事、活动推动标准应用与普及。推动在文娱、工业制造等领域的应用，打造标杆案例。以重大活动赛事为抓手，推动三维声标准先行先试，以具体落地案例促进产业链上下游的深度合作，提升观众认知。

鼓励企业加快应用与研发

面向未来 2-3 年，通过加快部署家庭影院，车载影音体验，个人影音体验，虚拟现实体验等业务，鼓励内容制作行业，分发平台加快三维菁彩声（Audio Vivid）的标准内容制作和分发，加速在产业链构筑，推进行业演进，构筑未来超高清音视频产业的技术和人才基础。

8. 附录

8.1 缩略语

下列术语和定义适用于本文件：

比特率 （bitrate ）

编码 （coding ）

编码器 （coder ）

编码音频位流 （coded Audio bitstream)

采样频率 （sampling frequency（fs）

解码 （decoding)

解码器 （decoder)

声道 （channel ）

数据单元 （data element ）

神经网络 （neural network ）

AV 音视频（Audio and Video）

dBFs 分贝满刻度（deciBel Full Scale）

C 前中声道（Front Center）

FFT 快速傅立叶变换（Fast Fourier Transform）

L 左声道（Left）

LFE 低频增强声道 (low frequency enhancement channel)

Lrs 左后环绕（Left Rear Surround）

Lss 左侧环绕（Left Side Surround）

Ltb 左上后（Left Top Back）

Ltf 左上前（Left Top Front）

MPEG 运动图像专家组（Moving Picture Experts Group）

PMT 节目映射表（Program Map Table）

R 右声道（Right）

Rrs 右后环绕（Right Rear Surround）

Rss 右侧环绕（Right Side Surround）

Rtb 右上后（Right Top Back）

Rtf 右上前（Right Top Front）

8.2 引用

[1] T/UWA 009.1-2022《三维声音技术规范 第 1 部分：编码分发与呈现》世界超高清视频产业联盟标准，2022.

[2] T/UWA 009.3-1-20《三维声音技术规范 第 3-1 部分：技术要求和测试方法 家庭影音播放设》世界超高清视频产业联盟标准，2022.

8.3 认证与授权

认证测试是产业生态发展的重要手段，三维菁彩声（Audio Vivid）具有权威、规范、科学、开放的测试认证体系，严格规范显示设备和便携式显示设备的呈现效果，对通过认证的企业提供最高质量的超高清技术支持，牵引终端产业向更高标准的显示技术迈进。

1. 为什么需要认证?

   UWA 联盟代表先进的音视频技术和生态标准，通过 UWA 认证的产品是高水平的音视频技术产品

   认证速度快，认证流程简单

2. 通过认证后能获得什么?

   UHW 认证证书和认证标志（三维菁彩声（Audio Vivid））的使用权

   列入 UWA 世界超高清视频产业联盟的公示认证名录

   优先参与联盟组织的获证企业联合营销活动

   长期有效的 三维菁彩声（Audio Vivid） 高质量技术支持

   参与联盟组织的人员、工程师的专业技术培训

3. 测试认证方法

对于支持三维菁彩声（Audio Vivid）的显示设备或系统，认证基本流程见下图，其中蓝色模块表示《认证实施规则》中提到的基本认证流程。具体认证方法登录世界超高清产业联盟官网，注册后即可下载文件《三维菁彩声（Audio Vivid） 认证实施规则》。

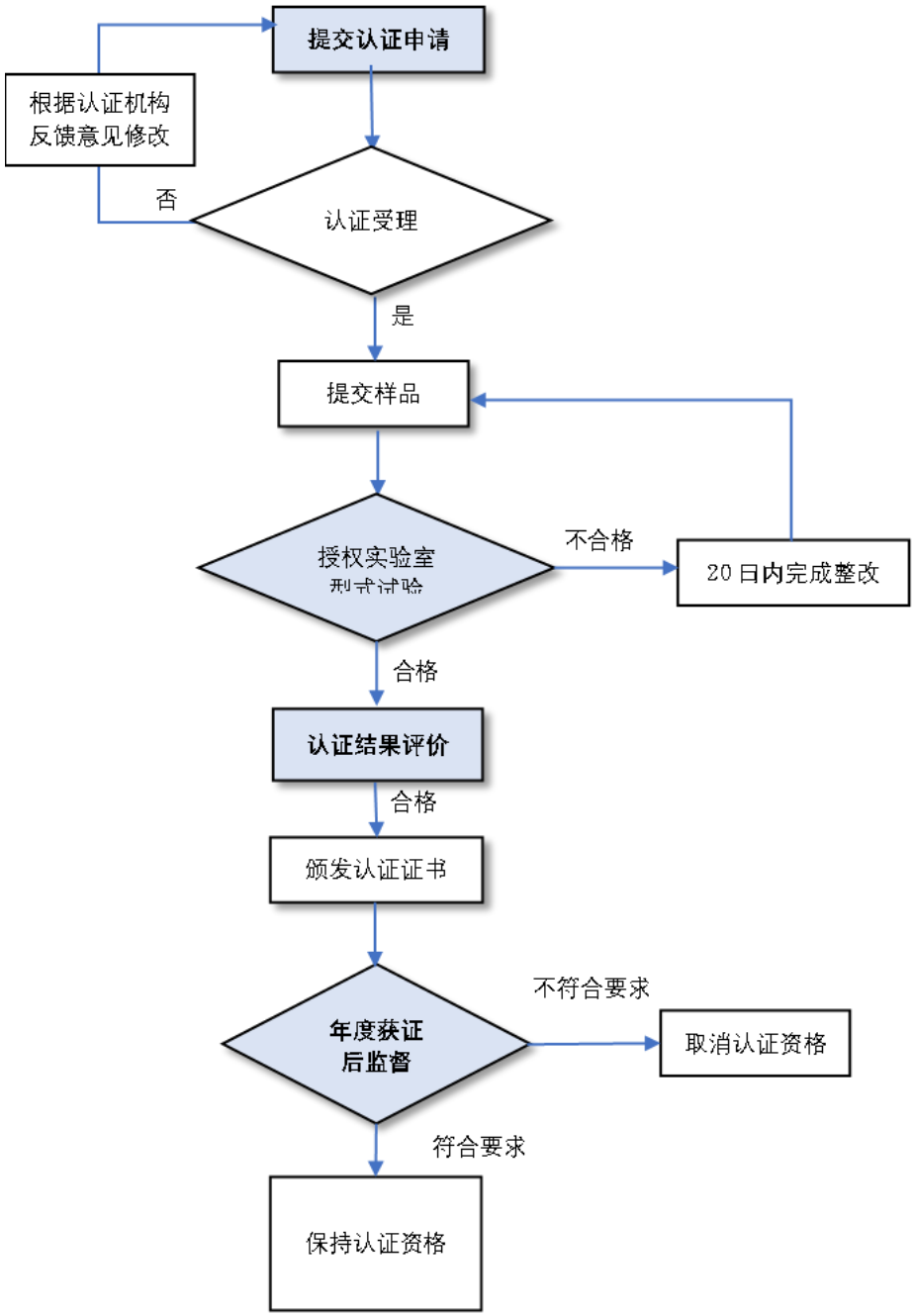

图 24 认证流程

8.4 测试与评估

1. 测试方式示例

三维声通用测试方案如下：通过标准测试流（编码以后的数据流）提供原始音源，进入被测影音设备，然后通过设备的自发声单元或者陪测发声设备进行测试，同时音频进入音频分析仪同步进行分析。本测试用例不包含双耳耳机，详细测试与评估参考三维菁彩声（Audio Vivid）标准测试文档。[2]

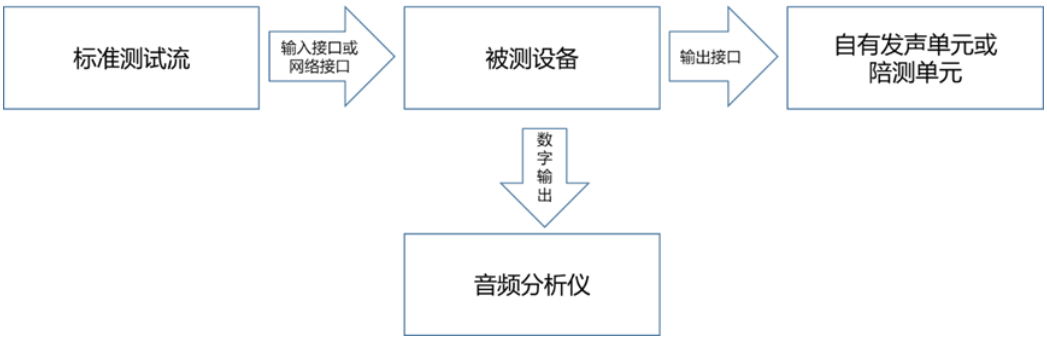

图 25 通用测试示意图

测试方法按以下几步进行：

a) 按图连接被测设备；

b) 被测设备根据标准测试流的输入

c）P1b 接口通过被测设备发声，同时通过 P3 接口介入音频分析仪分析。验证被测设备是否符合功能要求。

2.功能指标

功能将按照以下功能表进行验收测试

序号	项目		功能要求
1	三维菁彩声（Audio Vivid）音频识别		具备 UI 的设备应具备标识、解码 Audio Vivid 音频流的功能，能从一个节目中复用的多个音频流（Audio Vivid、MPEG-1 Layer II 音频）中正确标识并解码 Audio Vivid，同时不能将非 Audio Vivid 音频流标识为 Audio Vivid。
2	声床解码	声道映射	应能正确映射 Audio Vivid 音频的所有声道，包括立体声、5.1.4 多声道，所有正常声道信号均能正确复现。

3		输入采样频率	应能解码 32kHz、44.1kHz、48kHz 采样频率的 Audio Vivid 音频，宜能解码 96kHz 采样频率的 Audio Vivid 音频。
4		码率	应能解码 64kbps～832kbps 码率的 Audio Vivid 音频
5		采样精度	支持 16 比特，无损音频解码宜支持 24 比特采样精度
7	无损音频解码		宜支持无损音频解码
	HOA 解码		宜能正确还原三阶 HOA 信号，所有方位还原准确。
8	对象音频解码		应能支持对象音频还原，所有方位还原准确。
9	输出采样频率		具备数字音频输出的设备，应支持 48kHz 采样频率输出，宜支持 96kHz。
10	音效渲染		应能将 Audio Vivid 音频向下混合至设备自身最大音频播放能力集。

表 2 功能要求

3.技术指标

序号	项目	单位	性能要求
1	音频信噪比	dB	≥60
2	声道对增益差	dB	≤1
3	声道对的串扰	dB	≤-60
4	过载	——	无削波

表 3 电信号要求

序 号	项　目	单位	性能要求
1	声频率响应特性	dB	200Hz~8000Hz 范围内的不均匀度(波峰波谷)应小于 6dB。
2	额定输入时最大声压级	dBSPL	≥74dB
3	额定输入时声压总谐波失真	%	≤5（200Hz~8000Hz）

表 4 声信号要求

序号	接收终端类型	时间差范围 ms
1	音视频播放显示设备	-125~45
2	音视频播放设备	-30~20
3	音频接收播放设备	—
注：音视频信号时间差为-40ms 表示电视接收终端解码后音频信号落后视频信号 40ms； 音视频信号时间差为 20ms 表示电视接收终端解码后音频信号超前视频信号 20ms。		

表 5 时间差

序号	项目	功能要求
1	增益控制	应能正确解析响度元数据并正确控制响度，偏差不超过±2dB。
2	对白增益控制	应能正确解析对白响度元数据并正确控制对白响度，偏差不超过±2dB。

表 6 元数据处理性能


UWA
UHD World Association
世界超高清视频产业联盟