

# 沉浸式声音影院内容标准关系分析 与一种全景声影院内容制作的推测性方法

@izwb003

2023/12/20

## 摘要

沉浸式声音在十余年的发展历程中已经为行业带来了巨大的改变。在院线中，带有沉浸式声音的影厅之流行同样为电影爱好者带来了更具沉浸感的观影体验。无论是基于 Dolby Atmos®的杜比全景声影厅、基于 DTS:X 的临境音™影厅，还是 WANOS 等新兴国产技术，沉浸式声音在影院中的应用使得创作者的初衷得以更好地还原给观众。然而，符合沉浸式声音标准的影院内容的制作能力依然只被相关商业公司所掌握，普通内容制作者仍不能在不经商业公司昂贵的认证和授权的前提下制作在影院播放的沉浸式声音内容。这虽然有助于控制内容质量和获取必要的应得利益，但却对如观影活动等非专业场景下发挥沉浸式声音影院的最大作用造成了阻碍。

本文旨在通过对已公开的文献和资料进行相关的分析，来证明在 IAB 等新兴标准的驱动下凭已公开资料制作在沉浸式音频影厅播放沉浸式音频内容的可行性。同时，通过对沉浸式声音比特流配置文件 1 的理解给出对最常见的杜比全景声影厅的内容制作的推测性指导。

**关键词：**沉浸式声音，全景声，杜比全景声，数字特效影院

## Abstract

Immersive audio has brought tremendous changes to the industry over the past decade of development. The popularity of cinemas with immersive audio in theaters also brings movie enthusiasts a more immersive viewing experience. Whether based on Dolby Atmos ® Cinema and DTS: X ™ based advanced sound cinemas, as well as emerging domestic technologies such as WANOS, allow the application of immersive sound in cinemas to better convey the creator's original intention to the audience. However, the production capability of cinema content that meets the immersive sound standards is still only controlled by relevant commercial companies, and ordinary content producers still cannot produce immersive sound content played in cinemas without expensive certification and authorization from commercial companies. Although this helps to control content quality and obtain necessary benefits, it hinders the maximum effectiveness of immersive sound cinemas in non-professional settings such as movie watching activities.

This article aims to demonstrate the feasibility of producing immersive audio content in immersive audio theaters using publicly available materials, driven by emerging standards such as IAB, through relevant analysis of publicly available files and materials. At the same time, provide speculative guidance on the content production of the most common Dolby Atmos theater through an understanding of the immersive audio bitstream application profile 1.

**Keywords:** Immersive Audio; Atmos; Dolby Atmos; Digital special effects cinema

# 目录

引言 .....	5
声明与警示 .....	7
一、知识产权 .....	7
二、声明 .....	7
第一章 影院沉浸式声音的概念与实现 .....	8
第一节 沉浸式声音的相关概念 .....	8
第二节 沉浸式声音的影院实现 .....	9
第二章 影院沉浸式声音的内容制作 .....	11
第一节 沉浸式声音内容的打包 .....	11
第二节 沉浸式声音标准家族 .....	12
2.2.1 ST 2098-2 与 RDD57 .....	12
2.2.2 ST 2098-2 与 ST 2067-201 .....	14
2.2.3 RDD28 与 RDD29 .....	15
2.2.4 其它参考性信息 .....	16
2.2.5 母版格式 ADM BWF .....	16
第三章 标准内容的分析 .....	16
第一节 浅析 ST 2098-2 .....	16
3.1.1 比特流整体结构 .....	17
3.1.2 基本比特流元素 .....	18
3.1.2.1 帧定义 (IAFrame) .....	19
3.1.2.2 音床定义 (BedDefinition) .....	20
3.1.2.3 声音对象定义 (ObjectDefinition) .....	21
3.1.2.4 音床重映射 (BedRemap) .....	23
3.1.2.5 其它 IAB 新增元素 .....	24
3.1.2.6 音频数据的记录 .....	24
3.1.2.7 元素关系 .....	25
3.1.2.8 条件元素 .....	26
3.1.2.9 从已有信息中得到的兼容性思考 .....	27
第二节 浅析配置文件 1 .....	27
第三节 浅析 ST 429-18: IAB 的文件打包 .....	28
第四节 浅析 0 级插件 .....	29
第五节 浅析 0 级插件的杜比实现 .....	30
第六节 浅析 ADM BWF .....	31

第四章 最佳实践推测 .....	32
第一节 AS-DCP 库 .....	32
第二节 正规的杜比全景声影院内容制作流程 .....	33
第三节 正规的 IAB 影院内容制作流程 .....	35
第四节 所期待的最佳实践 .....	36
第五节 现行的可行解决方案 .....	39
结论 .....	40
参考文献 .....	41

## 引言

有声电影已经有了近百年的历史。在有声电影的百年发展历程中，工程人士一直没有停止对让电子复现的声音更为真实细腻的研究。长期以来，“堆积声道”一直是主流的解决方案。从单声道、立体声到 5.1、7.1 环绕，对声道数量的提升在一定意义上对增加沉浸感起到了积极作用。

2012 年 4 月，杜比实验室在 CinemaCon 上发布了新一代的电影声音格式：杜比全景声（Dolby Atmos®）。它采用了全新的“基于对象”的声音处理概念，让声音的展现正式摆脱了声道的限制。在博得了行业从业人员和社会各界人士的广泛关注后，杜比率先将“基于对象”的沉浸式声音概念推广到了影院行业。在全球各地兴建的“杜比全景声”影院通过增加天空声道和采用全景声声音处理器，让影院声音体验彻底摆脱了声道带来的局限性。观众得以体会到对象化的音频出现在影厅任何角落及天空声道带来的声音从头顶经过的震撼，让影院声音还原达到了传统基于声道的声音系统达不到的细致、沉浸和精确，得到了优良的社会赞誉。

不仅对于观众，杜比全景声对于音频内容创作者也有划时代的意义。传统的基于声道的音频编辑模式的各种局限在基于对象的音频编辑模式下皆被打破。音频工作者可充分发挥创意，让声音内容准确地出现在创作初衷所希望展现的空间位置上。这种编辑模式直观、清晰、简介等优点也很快得到了各类内容制作者的青睐。

对于对杜比全景声内容制作的掌控者，杜比实验室非常重视对工作流程的优化和对普及程度的关注。近十年时间以来，杜比除了多次维护更新全景声内容创作的相关工具，使得其更为简便易用之外，还针对影院、家庭影院、流媒体终端等不同的播放场景以及流媒体、蓝光、数字院线文件包等不同的记载方式开发了各种不同的全景声交付与回放、记录技术。此外，杜比实验室一直坚持大力推广全景声平民化。从与各大流媒体平台通力合作支持杜比全景声，到推出教程、指南与合作优惠鼓励 UGC 等内容创作，在杜比实验室的努力下，全景声已俨然成为了新一代高质量声音标准的代名词。

目前，针对作为非专业内容创作者的普通大众，杜比以较低门槛提供了一系列的工具来制作针对家庭近场娱乐内容的制作和交付。这些工具支持创建 Dolby Atmos ADM BWF, Dolby Atmos IMF IAB 等母版格式用于彼此交换内容，以及将这些母版编码为 Enhanced AC-3 with Joint Object Coding, TrueHD with Joint Object

Coding, AC-4 等用于回放和播放的编码格式以上传流媒体平台和用于播放欣赏。然而,杜比依然对制作用于嵌入数字院线包(DCP)的在杜比全景声影院播放的内容文件(Dolby Atmos Cinema MXF)给予严格的限制。杜比要求只有经过认证的授权工作室才能访问到此类内容文件的制作工具单元。文章作者认为,这是杜比为保证影院回放内容质量而采取的一种措施。但是,在此限制下,普通大众将无法在全景声影院回放自制的全景声内容(除非采取一些替代方案,下文将详细叙述)。这将对一些影院包场活动等应用场景下发挥全景声影院的完整作用造成阻碍。长期以来,文章作者经常收到“如何在全景声影院播放全景声内容”“如何制作全景声 DCP”等相关提问,而文章作者也往往只能给出“做不到,除非联系专业单位”的无奈答案。但是,文章作者认为一种可能可行的解决方案已经出现,但尚未有人对其加以研究以最终落实。这也正是本文撰写之目的。文章作者希望了解该解决方案是否切实可行,以及鼓励相关感兴趣的、了解专业知识的人士实现这一方案,从而使得在全景声影院播放自制的全景声内容变得可行。

该方案的背景如下:

伴随着杜比全景声的成功,全景声的概念俨然在音频行业掀起了一场革命。多家企业迅速跟进并推出了各自的研究成果。十年时间里,DTS:X、22.2ch、Barco Auro、WANOS 等相似概念的声音格式相继出现,并在业界建立起了一套全新的规格,统称“沉浸式声音”(Immersive Audio)。就此问题很快出现:各家不同标准的沉浸式声音规格均有各不相同的编解码方式和回放环境。这对内容创作者来说造成了极大的困难。为了使自己制作的内容能够在更多的系统上完成回放,创作者需疲于为自己的内容制作各种各样不同规格的交付档案,这显然不利于提高工作效率和创作内容的质量。

就此,电影电视工程师协会(Society of Motion Pictures and Television Engineers, SMPTE)作为行业技术规范的倡导者,于 2017 年联合相关行业人士、企业和协会成员开展起了一项工作,旨在统一一个能在不同的沉浸式声音系统上通用的沉浸式声音规范,并定名其为“沉浸式音频比特流”(Immersive Audio Bitstream),简称“IAB”。2018 年,SMPTE ST 2098 系列(SMPTE 25CSS Immersive Audio Suite)标准得以推出。电影电视工程师协会(下称“SMPTE”)希望日后各大厂商开发的沉浸式声音系统都能够兼容该系列标准,从而使得内容创作者可以瞄准该标准交付内容,节省了适配不同规格系统的工作。

目前，该系列标准的推广和适配工作仍在进行中。但是凭借文章作者已知的资料，国内相关行业似乎尚未跟进该标准。但是，该系列标准对于解决上述问题具有一个极大的参考意义：沉浸式音频比特流（下称“**IAB**”）标准正是基于未公开的杜比全景声影院内容标准而制订的。这也就意味着，相关感兴趣的经验丰富的开发人员极有可能可以通过 **SMPTE ST 2098** 系列标准以及一系列公开的建议性文件实现编写一个 **IAB**，或进一步地，杜比全景声影院内容的“编码器”，从而使得编码在全景声影院播放的内容变得可行，并制作全景声 **DCP** 用于影院播放。下文中将对其可行性给予详细的解释。

## 声明与警示

### 一、知识产权

提请读者注意，本文的绝大多数要素均来自公开文件。但其中的某些要素可能是一些专利权或知识产权、商业机密的主题。文章作者并未识别任何或所有此类权利。若本文内容有涉及到任何权益性问题，请及时同作者联系。作者将根据要求处理文章内容。

杜比、杜比全景声、杜比影院是杜比实验室国际有限公司的商标。

**SMPTE** 是电影与电视工程师协会的注册商标。

所有其它商标皆为各自所有者的财产。

### 二、声明

本文旨在于描述一种凭借已有资料的实现推测，旨在鼓励感兴趣相关内容的人士，并提供进一步的思考方向。本文**不是**学术论文，更**不是**标准或指导文档。本文的任何内容皆不应作为任何专业领域的工作参考。作者或其他当事人不对文章内容的有效性及实现效果做出任何保证。

文章作者深知自身能力及水平之不足。故须提醒读者注意，本文所述内容或包含大量谬误。文章作者欢迎并极其期待相关的批评及指正，以推进认知及进展。

在所适用的法律允许的范围内，本文所述内容不存在任何的保证。除非另有书面声明，作者和/或其他当事人“现以”不做出任何种类的、明示或默示的保证的



方式提供本文所述一切内容，包括但不限于适用于特定目的的默示保证。参照本文进行操作的一切风险（包括但不限于数据丢失、数据篡改、设备损坏及经济损失或赔偿等）均应由读者自行承担，即使作者或其他当事人已被告知此类损害的可能性。

## 第一章 影院沉浸式声音的概念与实现

### 第一节 沉浸式声音的相关概念

我们生活在三维的环境中，声音始终环绕于我们四周而被我们所感受到。但是，通过技术手段记录下来的声音只能通过存在于空间固定位置上的扬声器来单一、定点地被回放。这会导致声音失去其本应具有的方向信息。长期以来，解决这一局限性的方法是依赖于音响数量的增加。从单声道到立体声，再到 5.1 和 7.1，增加回放声道数量的方法对增强声音方位感起到了一定的积极作用。然而，声道概念下的回放系统仍然存在诸多明显的缺点。一是在某个声源发生声像的移动或变化时，由于声音的特性和人耳的构造，一系列反射和衍射效应会导致听者无法识别声音的连贯性；二是缺少位于头顶上方的声音回放信息，无法准确反映位于天空角度的声音细节，导致无法还原真实的听感，听觉效果无法与视觉感知相对应。

对此，沉浸式声音采用了一种截然不同的思考方式。其将播放的声音分为“音床”（Bed）和“对象”（Object）两个概念。其中，音床的理念同传统的基于声道的理念相同。系统可以定义多种不同规格的音床，包括但不限于 5.1, 7.1, 5.1.2, 5.1.4, 7.1.2, 7.1.4 等等，通过将不同声道的声音信号直接路由送至指定的扬声器来回放声音效果。这种设计可以将声音扩散到整个影厅，尤其适合用于渲染背景声音。在此基础上，影院声音处理器还会处理大量的声音对象。一个声音对象可以视作一个音频元素，但它可以自由平移到三维空间中的任何位置。处理器会通过一系列的算法和对应影厅的校准数据精确控制每个扬声器应当回放的声音，从而使得观众的听感产生某个声音对象的声音就来自于影厅三维空间某个位置的效果。这样的设计理念彻底让声音突破了声道的限制，打破了基于声道的回放系统存在的诸多局限，产生了更佳的声音重现效果。两者的结合构成了沉浸式声音的整体理念。

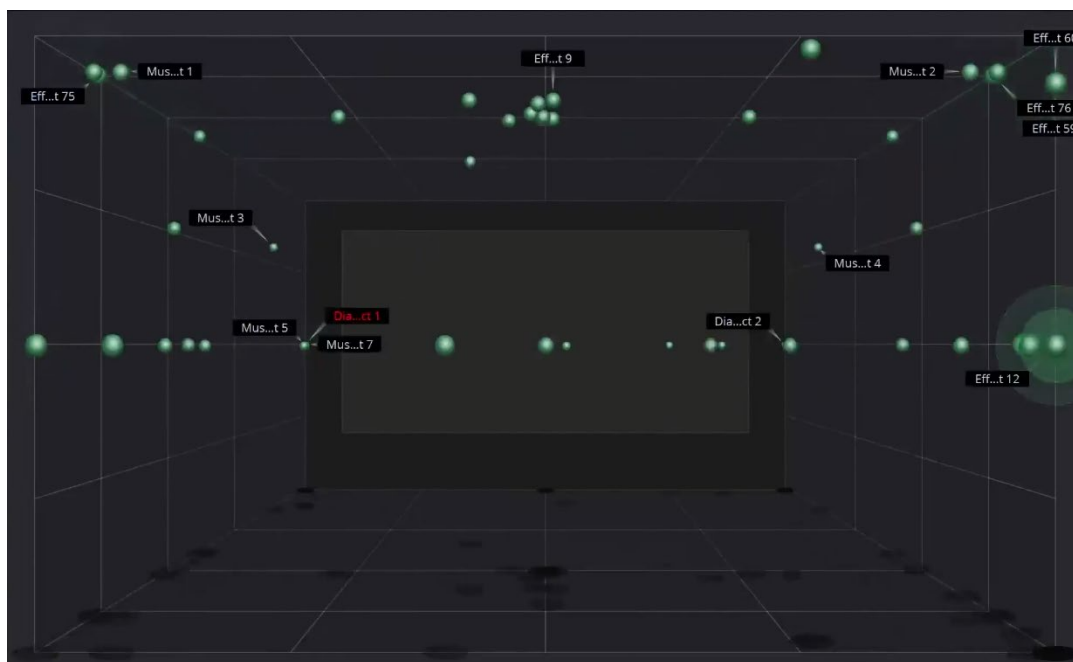


图 1-1 制作工具视角下的全景声声音对象分布



图 1-2 “Dolby Atmos Next-Generation Audio for Cinema (Issue 3)”（杜比全景声白皮书第三版）指出，音床和声音对象的结合构成了杜比全景声的独特回放效果。图片截取自原文

以杜比全景声为例，其不再以声道数量衡量回放效果的精细程度，而是最大允许支持 10 个声道的音床内容和 118 个声音对象。就此，制作者可以很自由地在可视化的工作环境中准确地将自己期望的声音重现位置还原。

## 第二节 沉浸式声音的影院实现

沉浸式声音影院的显著特征是更多的扬声器数量和天空声道的存在。而其区别于传统回放系统影厅的地方依然在于声音的处理方式。

传统的回放系统也可能具有很多数量的扬声器。但是，多个扬声器将使用同一个声道的信号回放同样的声音。也就是说，在影厅左后侧墙壁上的所有扬声器都将播放相同的来自 Ls 的声音，右后侧墙壁上的所有扬声器都将播放相同的来自 Rs 的声音，等等。最终，声音信息依然仅仅来自 6 或 8 个声道的信息，对回放效果和创作空间有较大的限制。

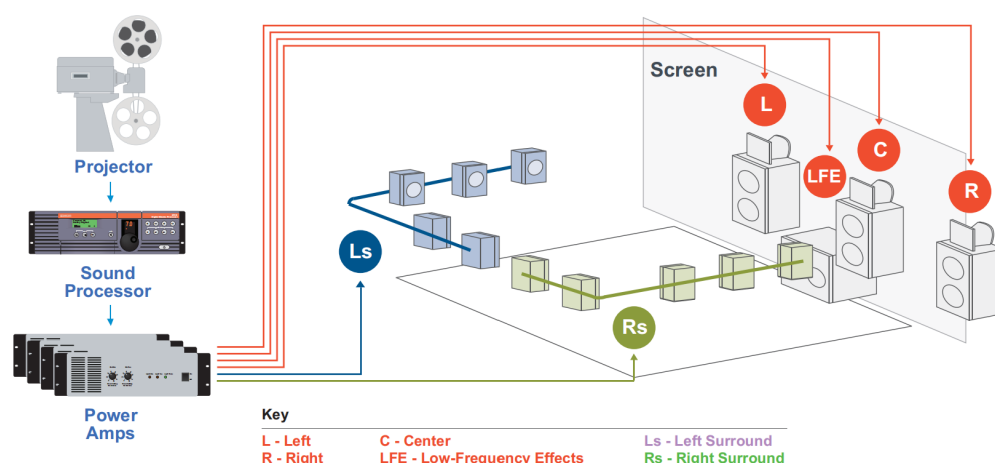


图 1-3 “Dolby Atmos Next-Generation Audio for Cinema (Issue 3)”（杜比全景声白皮书第三版）对传统 5.1 声道影院声音回放系统的示意。图片截取自原文

沉浸式音频系统则不再局限于只发送声道信息。相反，它更加依赖声音处理器的实时运算。处理器除了需要将音床内容输出到指定的扬声器外，还需要根据影厅的校准信息计算各个声音对象应当出现在影厅内的位置，并通过空间音频算法将声音输出到对应的扬声器，营造声音对象就在所处位置的声音效果。

仍以杜比全景声为例。如图 1-4 所示，全景声影院处理器在接收到声音信息输入后，会在渲染时分别渲染音床内容和对象内容，并经过电平管理和阵列校正。显著的不同是，处理器会为每一个扬声器单独运算一个 B 链处理，并输出完全不同的声音信号给最多 64 个扬声器。就此，每个扬声器都可以准确发出自身位置所应当发出的声音，极大提高了内容回放的精确性。

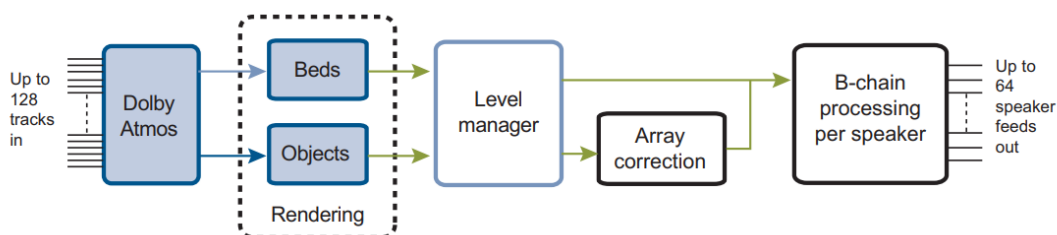


图 1-4 “Dolby Atmos Next-Generation Audio for Cinema (Issue 3)”（杜比全景声白皮书第三版）给出的杜比全景声输出链路图。图片截取自原文

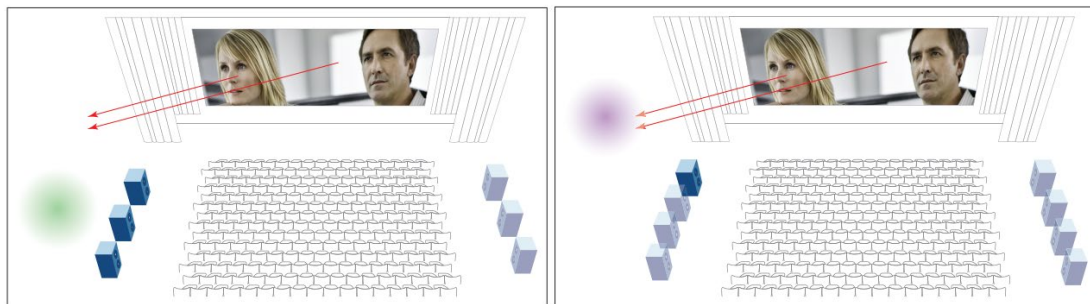


图 1-5 “Dolby Atmos Next-Generation Audio for Cinema (Issue 3)”（杜比全景声白皮书第三版）指出，为每个扬声器提供单独的 B 链处理可以使得声音出现在影厅中更精准的位置，从而更好的配合画面内容。图片截取自原文

仍然需要强调的是，并非所有的沉浸式声音系统都强调“音床+对象”的概念。如 Barco Auro, IMAX12 或 22.2ch 等沉浸式声音系统依然是基于声道处理声音的。这些系统更强调扬声器数量和天空声道带来的优势。

## 第二章 影院沉浸式声音的内容制作

### 第一节 沉浸式声音内容的打包

无论是何种规格的沉浸式声音，其内容必定需要在被打包后送至影院进行播放。在 IAB 的统一规范落地之前，不同的沉浸式声音系统可能遵照不同的协议和标准处理内容的制作和打包。下面仍然以杜比全景声为例进行介绍。

当代数字影院放映的内容依靠一种名为“数字院线包”（DCP）的数字文件包进行内容的交付。数字院线包内记录有一部放映内容的音频、视频和字幕等信息。这些信息本身被称为“精华”（Essence），连同一些元数据被封装在 MXF 容器内。这些被封装了的 MXF 容器文件就是“资源”（Asset）。

杜比全景声内容（Dolby Atmos Cinema MXF）是一种资源。如图 2-1 所示，它独立于普通的 5.1 音频而单独存在。正因如此，其记录方式同 5.1 或 7.1 音频的记录方式完全不同。

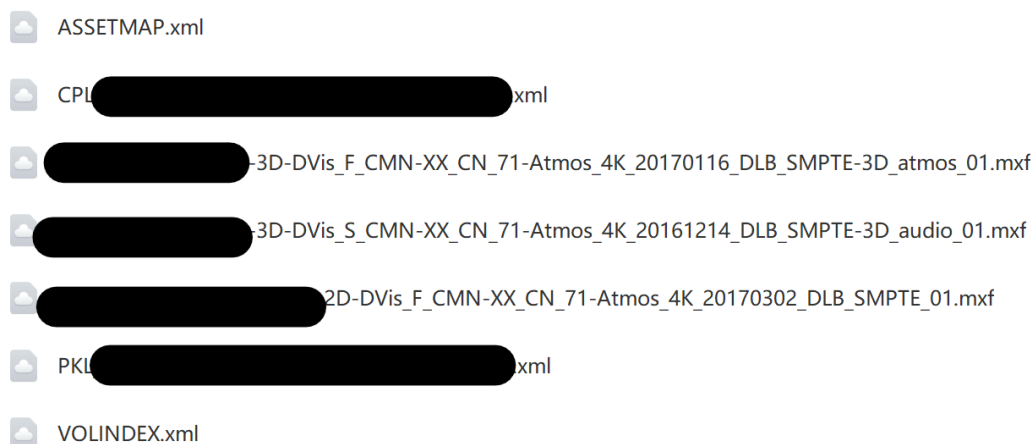


图 2-1 杜比全景声 DCP 包的文件组成。全景声内容、普通 7.1 音频内容、视频内容分别被封装保存在三个 mxf 文件内，分为三个 Asset 而彼此独立存在。

影院设备识别 DCP 包中包含的内容依靠由“合成播放列表”（CPL）所建立的合成与必要的资源映射。故此，CPL 文件中需指定让影院设备识别到全景声内容的信息。相关信息将于下文中详细探讨。下面，我们将重点放在这个杜比全景声内容上。如何借助已公开信息制作该内容为下文将重点探讨的核心。

在正规流程下，该内容应当由杜比实验室认证的内容创作工作室所拥有的母带渲染单元（RMU）编码，或通过 DCDM 在专业制版公司的处理下被封装制作为 DCP。这些流程的细节并不对外公开，并必然包含有商业机密或专利信息。但是，前文所叙述到的沉浸式音频全新标准 IAB 或许可以提供替代方案。

## 第二节 沉浸式声音标准家族

### 2.2.1 ST 2098-2 与 RDD57

任何可交换数字内容的正常交互均需要依赖彼此认可的标准或协议。作为 IAB 的领导者，SMPTE 下属的工程师团队为 IAB 制定有若干标准和文件。设备制造商可以参照这些标准完成相关软件/硬件的开发和测试。下文将通过分析这些标准彼此的关联引出进一步的结论。

IAB 全称“沉浸式音频比特流”。它同影院杜比全景声一样，是通过一种名为“比特流”的方式记录信息的。该比特流的完整细节被定义在了 SMPTE 标准 ST 2098-2:Immersive Audio Bitstream Specification（《沉浸式音频比特流规范》）中。ST 2098-2 定义了制造符合 IAB 标准的内容精华的全部细节。经验丰富的相关人士可以通过阅读 ST 2098-2 制造 IAB 内容的编码器，并用其编码 IAB 内容，打

包为 IAB 资源嵌入 DCP 交付，制造 IAB DCP 包。

由于 IAB 尚未完全落地，IAB DCP 包无法被目前的大多数影院设备识别播放。但是极具参考价值的一点是，IAB 正是由未公开的杜比全景声技术细节信息进行少量拓展而得来。也就是说，杜比全景声影院内容的精华比特流是 IAB 的子集。每一个杜比全景声比特流都符合 IAB，但并不是每一个 IAB 都符合杜比全景声比特流。

这也就意味着，通过对 ST 2098-2 加以必要的限制，取消 IAB 标准相对杜比全景声比特流标准的拓展部分，即可制造等价的杜比全景声比特流，作为精华嵌入 MXF 用于制作杜比全景声 DCP。

幸运的是，这一限制的具体内容同样得以披露公开。SMPTE 注册披露文件 RDD57:2021 SMPTE ST 2098-2 Immersive Audio Bitstream and Packaging Constraints: IAB Application Profile 1（《沉浸式音频比特流和包约束：IAB 应用配置文件 1》）定义了一些针对 ST 2098-2 的限制，并称被限制了的 IAB 为“IAB 配置文件 1”（IAB Profile 1）。该披露文件很有可能是为了在现有系统上及时推动 IAB 普及而公开的指导说明。文章作者认为，也许日后会出现 IAB 配置文件 2、3 等。值得注意的一点是，RDD57:2021（下称“配置文件 1”）明确指出，遗留内容（Dolby Atmos）遵循 IAB 配置文件 1 的指导方针。故此我们可以推断，不同类型的比特流之间的关系可以由图 2-2 反映出来。

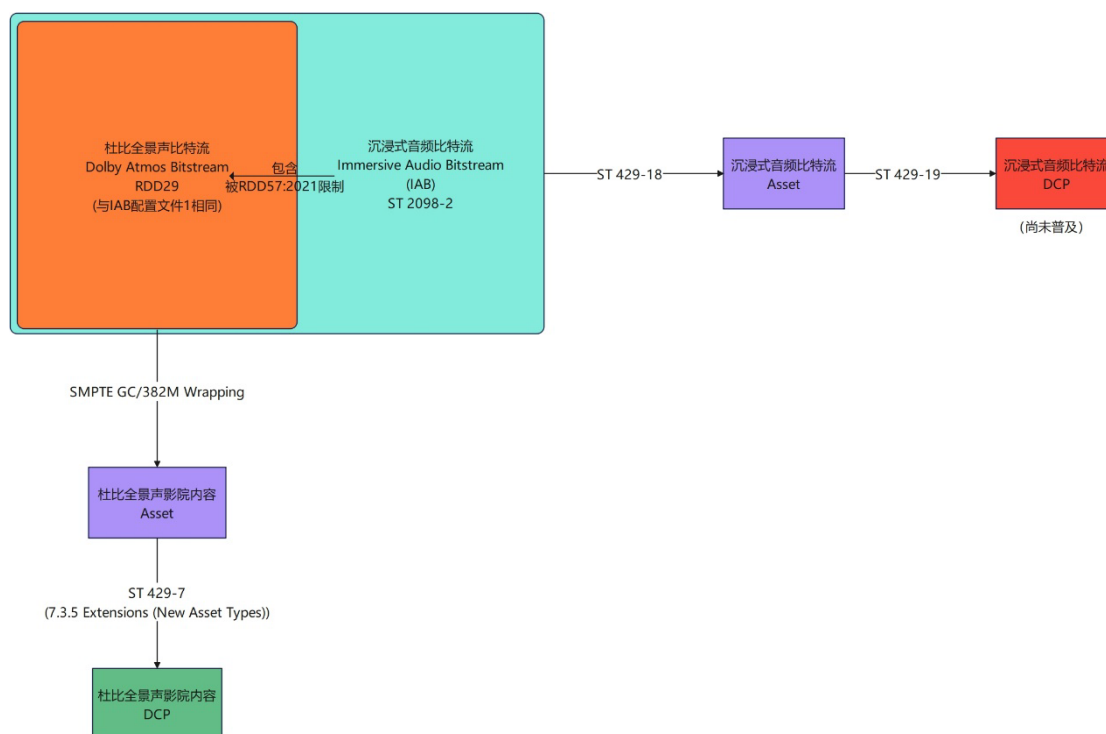


图 2-2 对杜比全景声比特流和沉浸式音频比特流之间相关关系的推断

就此，我们认为，感兴趣的相关人士如果在参考 ST 2098-2 编写 IAB 编码器的同时，限制编码器的功能在配置文件 1 的范围内，就可以认为该编码器可编码的内容与杜比全景声影院内容比特流一致，而相关资料都是可以公开访问或购买到的。

## 2.2.2 ST 2098-2 与 ST 2067-201

除了我们所希望探讨的杜比全景声影院内容外，读者或许还会熟悉另外一种格式叫做 Dolby Atmos IMF IAB。Dolby Atmos Conversion Tool、Davinci Resolve 等常见软件均支持导出或转换为该格式的文件。由于该格式同样封装于 MXF 文件中，许多人会误以为它同 DCP 中的杜比全景声影院内容相同，故尝试将其置于 DCP 中但以失败告终。事实上，虽同样使用 MXF 包装，这两类文件在内容和用途上都截然不同，自然无法混用。

该类型的文件是为一种名为“交付母版文件”（Interoperable Master Format, IMF）的文件包设计的。这种包从 DCP 标准修改而来，结构与 DCP 相似，但并非用于交付影院播放而是用于向流媒体平台等机构交付影片母版，用于制作不同的分发。故此，IMF IAB 实为一种和 ADM BWF 等同的母版格式。



阻碍将 Dolby Atmos IMF IAB 嵌入 DCP 的最大障碍在于，Dolby Atmos IMF IAB 虽同样是 IAB，但同配置文件 1 相反，它是从 ST 2098-2 经过进一步的拓展而得到的。定义该拓展的标准为 SMPTE ST 2067-201 Immersive Audio Bitstream Level 0 Plug-in (《沉浸式音频比特流 0 级插件》)。ST 2067-201 (下称“0 级插件”) 在 ST 2098-2 的基础上增加了大量拓展内容，使得 IMF IAB 更适合用于母版交付。但这也使得它和影院杜比全景声比特流的差异更大了。不仅如此，ST 2067-201 仅仅定义了 IMF IAB 的特点。使用杜比工具创作的 IMF IAB 称 Dolby Atmos IMF IAB，它的标准被杜比实验室在 0 级插件的基础上进行了进一步的调整和修改。具体的调整可在杜比专业支持网站的这份文档中获得详细信息：[https://professionalsupport.dolby.com/s/article/Dolby-Atmos-IMF-IAB-interoperability-guidelines?language=en\\_US](https://professionalsupport.dolby.com/s/article/Dolby-Atmos-IMF-IAB-interoperability-guidelines?language=en_US)。

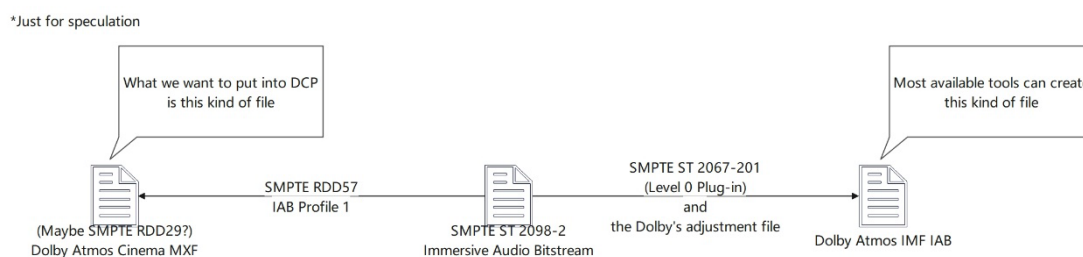


图 2-3 对已有格式之间关系的推断

### 2.2.3 RDD28 与 RDD29

事实上，杜比全景声比特流的规范亦早已有所披露。SMPTE 注册披露文件 RDD29:2014 (现已更新为 RDD29:2019) 已经对杜比全景声比特流做出了若干解释。同样地，杜比全景声 DCDM Print Master File 的比特流格式在 RDD28:2014 中得以披露。

SMPTE 在 RDD28 和 RDD29 中声明，作为注册披露文件，这些信息只能用作对杜比全景声比特流的解释。它并不打算支持创建或处理这些文件的硬件或软件应用程序的开发。故此，这些信息只能用作参考，且文章作者不认为其详细程度足以支撑制作真正的杜比全景声比特流编码器。同时，这些强调也提醒我们，即使按照下文所述的解决方案完成了开发工作，成品也只能被称作 IAB 配置文件 1 的编码器，而不能声称为杜比全景声比特流的编码器。



## 2.2.4 其它参考性信息

除上述标准及文档外，还有一系列的规范性内容可以提供参考：

- ST 2098-1 给出了沉浸式音频比特流的元数据记录方式。这可能在 ST 2098-2 中被引用。
- EG 2098-3 给出了沉浸式音频渲染器的技术期望与调试指导。
- ST 2098-5 定义了沉浸式音频的音频通道和声场组规格。这可能在 ST 2098-2 中被引用。
- ST 429-18 给出了将沉浸式音频比特流封装为轨道文件的方法。由于 IAB 尚未普及，该规范并不一定影响我们所期望的方法。
- ST 429-19 给出了沉浸式音频在 DCP 中的描述方式。由于 IAB 尚未普及，该规范并不一定影响我们所期望的方法。

## 2.2.5 母版格式 ADM BWF

Dolby Atmos ADM BWF 是最常见的母版格式。其为消费者内容制作而设计，并非用于影院内容。杜比专业支持网站给出了其详细技术信息：[https://professionalsupport.dolby.com/s/article/Dolby-Atmos-ADM-Profile-specification?language=en\\_US](https://professionalsupport.dolby.com/s/article/Dolby-Atmos-ADM-Profile-specification?language=en_US)。这可能对我们所期望的方法有参考价值。

# 第三章 标准内容的分析

## 第一节 浅析 ST 2098-2

ST 2098-2 作为我们所探讨的所有内容的核心，其定义了沉浸式音频比特流的技术规范，并可从中推断杜比全景声比特流的一些特点。

由于 ST 2098-2 是 SMPTE 拥有版权的出版物，在此我们不会，也没有能力对其进行完整的剖析，而是从 SMPTE Education 的一次公开讲座，由杜比影院行业高级工程师 Stan Cossette 主持的 The Ins and Outs of ST 2098-2 Immersive Audio Bitstream 中窥知其一二。在该讲座中所公开分享的内容理应是可以被讨论的。

### 3.1.1 比特流整体结构

不同于固有认知，IAB 中的所有信息是以帧为单位进行记录的。这些帧被称之为“IA Bitstream Frame”。IA Bitstream Frame 在影片节目播放的过程中同视频帧同步，并理应与视频内容具有相同的帧率。视频中的每一帧播放的瞬间，对应的 IA Bitstream Frame 也被解码并播放。这些帧在比特流中被顺序记录。

每一个 IA Bitstream Frame 由一小段前序（Preamble）和 IAFrame 组成。前序是杜比全景声比特流中的遗留内容，在 IAB 流中的作用是保持兼容性，故此可能不出现或保留为空。当前帧的关键信息于 IAFrame 中被呈现。

IAFrame 承载了当前帧的所有沉浸式音频元数据和精华。其顺序记录了一个（或多个）音床的定义、多个空间对象的定义，并紧随其音频内容。每个 IAFrame 可能以一些创作信息收尾。

一个值得注意和强调的事实是，对象是可能在空间中不断运动的。而如果每个 IAFrame 只记录某个对象在这一帧时刻的位置，对象的位置信息将只能以同步帧率的速度逐次刷新。由于当代电影依然以 24 帧的帧率为主，这样的记录方式会使得空间对象的位置信息只能以每秒 24 次的速度刷新。这样如果这一对象在运动，其位置变化就将不连贯。

对此，IAB 的解决方案是允许为空间对象的定义添加若干子块。每个子块允许为一帧内的对象记录不同的元数据。这样，通过按顺序在这一帧的时间内不断根据子块中的元数据调整对象的信息，就可以使得诸如运动等对象信息的变化变得更加连贯。

沉浸式音频比特流的整体结构可如图 3-1 被描述。

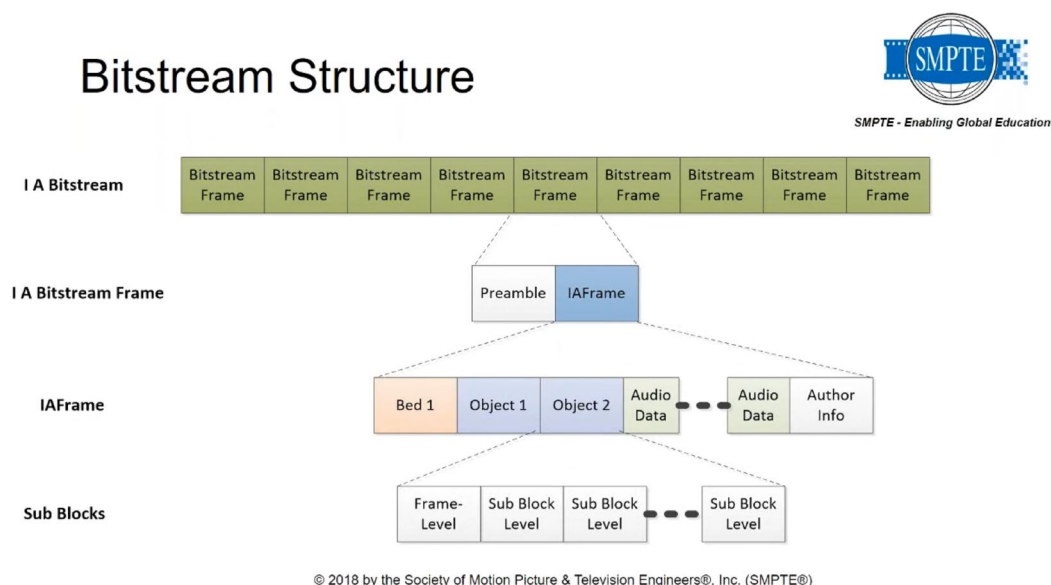


图 3-1 Stan 在 SMPTE Education 讲座中所描绘的沉浸式音频比特流结构。图片截取自视频

### 3.1.2 基本比特流元素

每一段必要的信息（IAFrame、音床的定义、声音对象的定义、音频内容等）在比特流中以一个元素（Element）的形式被记录，称 IAElement。IAB 实际上就是将无数元素以流的形式顺序记录，每一个元素都可以按指定规则携带若干子元素。每一个元素必然包含 ElementID 和 ElementSize 两个参数。ElementID 用于让解码器识别该元素所记录的信息类型。每一种不同的信息都具有一个独一无二的 ID 以填入 ElementID 用于识别。解码器被允许跳过 ElementID 不能被识别的元素。ElementSize 表明了该元素的大小。其可以被用于在解码器跳过不能被识别的元素时知道跳过比特流的多少长度。其具体包含内容根据内容的不同紧随其后。简而言之，每一个 IAElement 的结构可如如下伪代码表示：

```
IAElement()
{
    ElementID;
    ElementSize;

    switch(ElementID)
    {
        case(IAFrame):
            IAFRame;
            break;
        case(BedDefinition):
```

```

        BedDefinition;
    break;
    case(BedRemap):
        BedRemap;
    break;
    case(ObjectDefinition):
        ObjectDefinition;
    break;
    case(ObjectZoneDefinition19):
        ObjectZoneDefinition19;
    break;
    case(AuthoringToolInfo):
        AuthoringToolInfo;
    break;
    case(UserData):
        UserData;
    break;
    case(AudioDataDLC):
        AudioDataDLC;
    break;
    case(AudioDataPCM):
        AudioDataPCM;
    break;
    default:
        ElementSize * 8 未知信息;
}
}

```

下文中将对其所携带的每一种信息给出描述。请注意，每一个描述都是经过简化的，旨在理解其大体框架。再次强调本文不能作为任何专业指导使用：

### 3.1.2.1 帧定义 (IAFrame)

每一帧的相关定义在 IAFrame 元素中给定。其结构可如如下伪代码所述：

```

IAFrame()
{
    ElementID;
    ElementSize;

    SampleRate; //所记录的音频信息的采样率
    BitDepth;   //所记录的音频信息的位深
}

```

```

FrameRate; //内容的帧率。应与视频等内容的帧率相同以保持同步
MaxRendered; //回放期间可以渲染的通道和对象的最大数量。例如，杜比全景声允许最大 9.1 音床+118 个声音对象，则 MaxRendered 应定义为 128.
SubElementCount; //子元素数量
for(i=0; i < SubElementCount; i++)
{
    SubElement[i]; //子元素
}
}

```

每一个 IAFRAME 都可以独立解码。任何一个 IAFRAME 都与其它 IAFRAME 没有关联。这对内容变化提供了很大的自由。实际应用中，它可能被施以各种各样的约束，例如强制每一个 IAFRAME 都具有相同的 FrameRate 以禁止中途更改帧率等。RDD57 正是这样的约束。

下述的所有其它元素都是作为 IAFRAME 的子元素（SubElement）而存在的。

### 3.1.2.2 音床定义（BedDefinition）

作为 IAFRAME 的子元素，BedDefinition 用于定义音床。其结构可如如下伪代码所述：

```

BedDefinition()
{
    ElementID;
    ElementSize;

    MetaID; //当前音床的 ID。在前后多个 IAFRAME 帧中，同一个音床应始终具有同一个 MetaID。
    ConditionalBed; //是否为“条件音床”。
    if(ConditionalBed)
    {
        BedUseCase; //条件音床下的音床用例代码
    }
    ChannelCount; //声道数量
    for(i=0; i < ChannelCount; i++) //对于每个声道
    {
        ChannelID[i]; //当前声道的 ID，对应 SMPTE ST 428-12 或 SMPTE ST 2098-5 或 ITU-R BS.2051-2 定义的声道名称以确认当前声道将被路由到的目的地（L,R,C,LFE 等）。
        AudioDataID[i]; //当前声道的音频数据 ID。同某一个 AudioDataDLC 或 AudioDataPCM 元素的 AudioDataID 对应，表明该音床的音频对应于哪一个声音数据。
    }
}

```

```

        ChannelGain;    //当前声道的增益值。
        ChannelDecor;   //当前声道的解相关设置。
    }
    AudioDescription;    //对音频内容的描述
    SubElementCount;     //子元素数量
    for(i=0; i < SubElementCount; i++)
    {
        SubElement[i];  //子元素
    }
}

```

IAB 允许多个音床的存在。故此，给定 MetaID 以确保在播放过程中解码器始终能索引到同一个音床是有必要的。

AudioDataID 可以被设定为 0，表明没有对应音频。

### 3.1.2.3 声音对象定义 (ObjectDefinition)

作为 IAFRAME 的子元素，ObjectDefinition 用于定义一个声音对象。其结构可如下伪代码所述：

```

ObjectDefinition()
{
    ElementID;
    ElementSize;

    MetaID; //当前对象的 ID。在前后多个 IAFRAME 帧中，同一个对象应始终具有同一个 MetaID。
    AudioDataID;    //当前对象的音频数据 ID。同某一个 AudioDataDLC 或 AudioDataPCM 元素的 AudioDataID 对应，表明该对象的音频对应于哪一个声音数据。
    ConditionalObject; //是否为“条件对象”
    if(ConditionalObject)
    {
        ObjectUseCase; //条件对象下的对象用例代码
    }
    for(sb=0; sb < NumPanSubBlocks; sb++) //（对于每个）记录一帧内对象空间运动信息的子块
    {
        if(sb == 0)
            PanInfoExists = 1; //值为 1 表示存在新的平移信息，需要更新当前对象的空间位置。由于第一个子块始终是新的平移信息，故此第一个子块的该值必须为 1。
        else

```

```

        PanInfoExists;    //值为 0 表示不存在新的平移信息，仍然使用上一个
        值为 1 的子块记录的空间位置。

        if(PanInfoExists == 1) //如果信息需要更新
        {
            ObjectGain; //声音增益
            ObjectPos;   //对象位置
            ObjectSnap;  //对象捕捉控制相关定义，用于定义一个对象是更注重声音
            还原的生动还是位置的精准
            ObjectZoneControl; //区域（指定的一组扬声器）控制相关定义，可用
            于限制发出该对象声音的扬声器。
            ObjectSpread; //对象扩散定义，相当于对象的大小
            ObjectDecor;  //当前对象的解相关设置
        }
    }
    AudioDescription;    //对音频内容的描述
    SubElementCount;     //子元素数量
    for(i=0; i < SubElementCount; i++)
    {
        SubElement[i]; //子元素
    }
}

```

AudioDataID 可以被设定为 0，表明没有对应音频。

一般而言，每一个 ObjectDefinition 中会包含 8 个子块。这样，大约每 5ms，对象的空间位置等信息就将会更新一次。

在上述简化描述中，ObjectPos（对象位置）分三个数据进行描述：ObjectPosX, ObjectPosY, ObjectPosZ。它们分别代表一个从影厅空间中建立起的笛卡尔坐标系的 X, Y, Z 坐标值。该坐标系以屏幕所在墙壁、左侧（相对面向屏幕的观众）墙壁和除天空声道外的扬声器所在的平面（Base Layer）的交点为原点，向屏幕方向延伸为 X 轴、向放映室方向延伸为 Y 轴、向天花板方向延伸为 Z 轴。除地板外的每面墙壁都坐落于坐标值为 0 或 1 的地方。图 3-2 反映了该坐标系的简单模型。

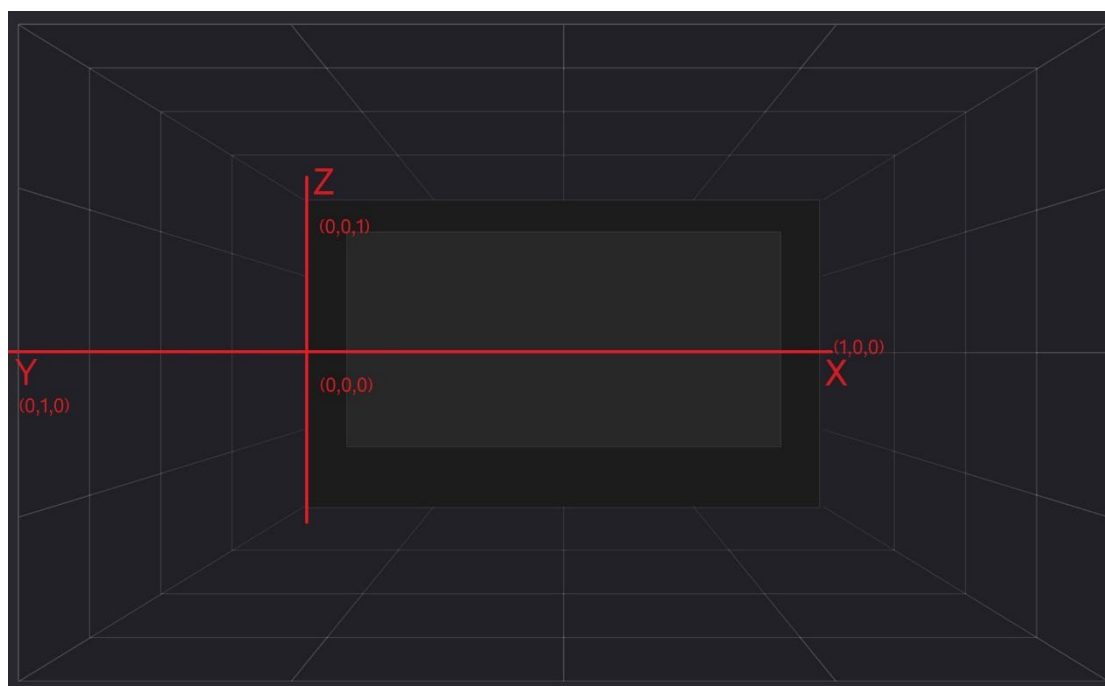
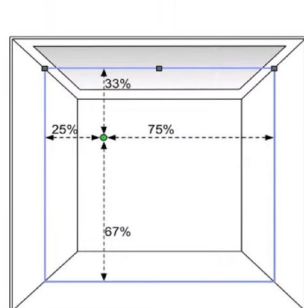


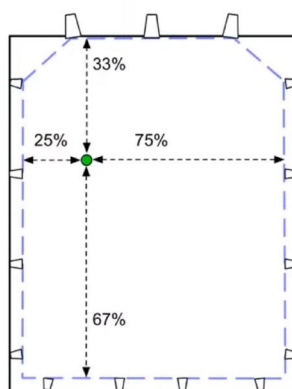
图 3-2 空间坐标系示意图

每个对象的坐标被限制在 0 和 1 的范围内，以避免移动到空间外。

## Mapping location to a cinema



Position  $[x,y,z] = [0.25, 0.33, 0]$



• Figure on left shows idealized location

© 2018 by the Society of Motion Picture & Television Engineers®, Inc. (SMPTE®)

图 3-3 Stan 在 SMPTE Education 讲座中给出的一个对象位置的坐标例子。针对这个例子，ObjectPosX=0.25, ObjectPosY=0.33, ObjectPosZ=0。值得注意的是，由于不允许负值的存在，空间对象不会移动到 Base Layer 下方的空间里。图片截取自视频

### 3.1.2.4 音床重映射 (BedRemap)

IAB 允许多种不同规格的音床的存在。音床可能是 5.1, 7.1, 5.1.2, 7.1.2,



5.1.4 等各种各样的规格。但是，回放时目标播放环境的扬声器配置可能只是其中的一种。故此，**BedDefinition** 允许一个叫做 **BedRemap**（音床重映射）的子元素存在。该元素将会解释针对不同的回放系统的配置，音床应该被如何重新缩混，进行上/下变换以适应目标回放环境。

由于 IAB 配置文件 1 和杜比全景声比特流都只允许一个 9.1（7.1.2）音床存在，并不支持进行音床重映射，故该元素暂不属于本文的讨论范围。

### 3.1.2.5 其它 IAB 新增元素

IAB 相比杜比全景声比特流增加了如下元素：

- **ObjectZoneDefinition19**：作为 **ObjectDefinition** 的子元素，该元素用于定义 Auro 扬声器布局的区域设定。如果该子元素存在，**ObjectZoneControl** 中的相关设置将被该元素中的设置替代。
- **AuthoringToolInfo**：保存该比特流的制作工具的信息。
- **UserData**：用于携带其它未在 ST 2098-2 中定义的数据。这些数据用其中的 **UserID** 字段识别其类型。不支持的解码器会将其跳过。影院系统中不会使用到该元素。

杜比全景声比特流不支持这些新增的元素。它们同样不应该出现在 IAB 配置文件 1 中。故此这些元素暂不属于本文的讨论范围。

### 3.1.2.6 音频数据的记录

前文所述的各个元素均记录的是元数据信息。真正的音频信息紧随其后在 **AudioDataDLC** 或 **AudioDataPCM** 元素中被记录。每一个 **AudioDataDLC** 或 **AudioDataPCM** 元素记录一个 **AudioDataID**。对应前面 **BedDefinition** 或 **ObjectDefinition** 中某个声道或对象的 **AudioDataID**，表明该声道或对象的声音数据应在具有哪个 **AudioDataID** 的 **AudioDataDLC** 或 **AudioDataPCM** 元素中获取。

**AudioDataDLC** 和 **AudioDataPCM** 的区别是，**AudioDataDLC** 使用一种无损声音编码方式记录音频（该编码方式在 ST 2098-2 的附录 B 中给出了定义、解释与编码器示例），**AudioDataPCM** 使用 PCM 编码方式记录音频。有关其具体记录方法应参阅 ST 2098-2 中给出的解释。

针对影院内容，只有 AudioDataDLC 元素被支持。故此，完整影院实现仍需考虑实现 DLC 无损声音编码。

### 3.1.2.7 元素关系

IAB 中，各元素之间的关系可如图 3-4 所示 UML 类图表示。

对 IAB 的实现不应忽略下列事实：

- 所有所述元素都是为了组成完整 IAFrame 而定义。
- 所有的元素都是可选的。
- 一些元素可能包含子元素，构成父子元素关系。
- 子元素如果被指定，其可能替换掉父元素中的部分或全部内容。

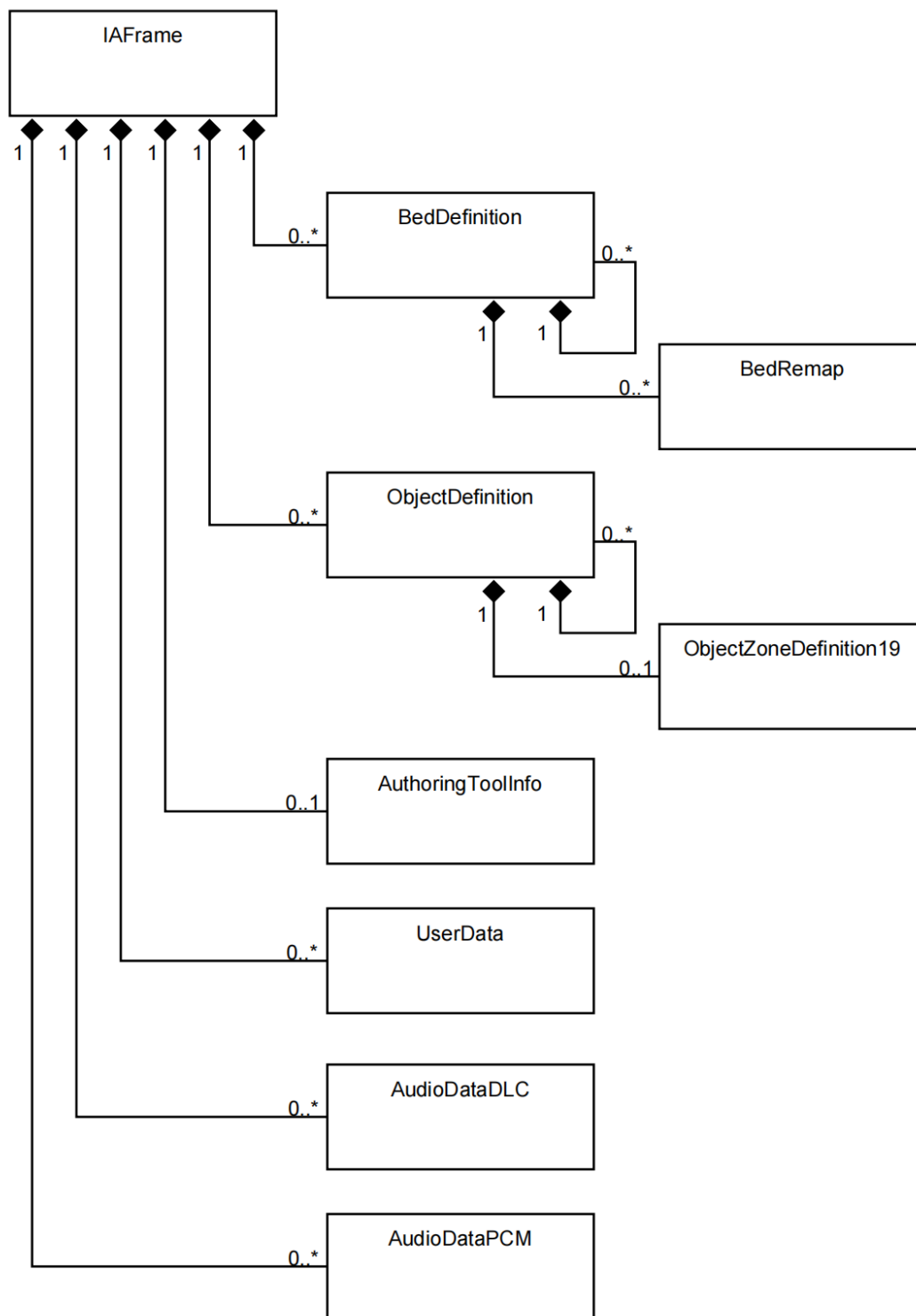


图 3-4 ST 2098-2 给出的 IAB 的各元素父子与顺序关系 UML 类图。图片截取自原文

### 3.1.2.8 条件元素

由于目标回放环境多种多样，创作者可能需要对比特流做一系列的补充以适应目标播放环境。条件元素为这种情况而设计。

条件元素包含条件音床、条件对象以及音床重映射。它们可能作为顶级元素（次于 IAFRAME）存在，也可能是某个元素的子元素。作为子元素存在时其如果生效，其将替代其父元素（或其一部分）而作用。每一个条件元素会指定一个用例代码表明自己用于何种回放环境。该回放环境下的解码器发现该条件元素的应用例代码所描述的环境符合自己所工作的环境后，就会使用该条件元素中的信息，而不是其父元素中的信息。

配置文件 1 和杜比全景声比特流不支持条件元素。

### 3.1.2.9 从已有信息中得到的兼容性思考

Stan 在 “The Ins and Outs of ST 2098-2 Immersive Audio Bitstream” 上指出，IAB 标准的向前兼容性可以被如下思考：

- 许多元素都来自已有的杜比全景声比特流。
- 没有新增加的元素是必须存在的。
- 已有元素的结构没有改变，除了一些新的内容被添加（这些内容可以被指定不存在）。
- 解码器可以跳过不能被识别的元素。

以上结论可见，新式的 IAB 对传统杜比全景声比特流的兼容性极佳。每一个杜比全景声比特流都是 IAB，但由于新增内容的添加，不一定每一个 IAB 都是杜比全景声比特流。参照 ST 2098-2 编码的新比特流可能不能在已有的解码器上良好工作。为使其兼容，需要对 ST 2098-2 加以约束。下文将要探讨的配置文件 1 解释了这种约束。

## 第二节 浅析配置文件 1

RDD57:2021 定义了一种对 ST 2098-2 的约束，以使得已有的解码器可以解码符合配置文件 1 的 IAB。ISDCF 和 EDCF 在 RDD57 中指出了其披露背景：当前已有的内容渲染器并不支持 ST 2098-2 中的完整规范。故此，需要一个工程文档来指导新的内容创作工具的部署。配置文件 1 旨在指导创作产品，以限制其使用一些现场设备无法呈现的特性。

可以推断，“一些现场设备”很有可能是指已有的杜比全景声解码器（如

CP850, CP950A 等)。故此,符合配置文件 1 的比特流理应可以被用于在这些解码器中播放。

具体的限制内容应当参阅 RDD57:2021 来了解。一些很明显的限制包括:

- 采样率必须为 48kHz。
- BedDefinition 中 ChannelCount 应为 10, 指代一个 9.1 (7.1.2) 音床。
- 最多有 1 个 BedDefinition。
- 最多有 118 个 ObjectDefinition。
- MaxRendered 最多为 128。
- 不支持条件音床及条件对象。
- 不支持音床重映射。
- 不支持各新增元素。
- 等等……

文章作者认为,配置文件 1 可以与 RDD29 一同阅读,以进一步发现配置文件 1 和杜比全景声比特流之间仍然可能存在的一些隐含差异。

除了对比特流的限制外,配置文件 1 同样对打包和 CPL 定义加以指导。这些指导将在下文中加以明确说明。由于杜比全景声比特流并不遵照这一指导,这些指示对我们所期待的流程可能没有参考意义。

### 第三节 浅析 ST 429-18: IAB 的文件打包

为作为资源嵌入 DCP, IAB 需要被打包为 MXF 文件。这种打包携带沉浸式音频数据精华的方法在 ST 429-18 中被给定,以便于嵌入 DCP 中使用并保证音画同步。由于 IAB 尚未普及,该方法并不是我们所期待的流程中将应用的方法。

IAB 的打包遵照帧打包 (Frame Wrapping) 要求。帧打包为每个帧中的精华添加 KL 标头使之成为 MXF 帧。若干 MXF 帧顺序相连。帧打包的原理如图 3-5 所示。

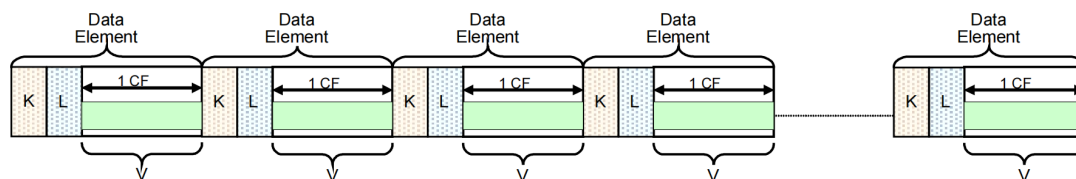


图 3-5 ST 429-18 中所解释的帧打包原理。图片截取自原文

图 3-5 的解释中, K 标头定义当前帧数据元素的键值。IAB 数据元素的键值在 ST 429-18 的第 6.1 节中被定义。L 标头给定当前帧数据内容的长度。其实现在 SMPTE ST 377 中被定义。V 即 ST 2098-2 中所定义的 I A Bitstream Frame。这种方式称为 KLV 编码。

帧打包后的内容在被打包为 MXF 时将被添加一个文件头和一个文件尾。文件尾中定义了一个对所有 MXF 帧的索引, 以便于对流中的帧进行随机读写, 以及及时索引到某一特定帧。完整过程如图 3-6 所示。

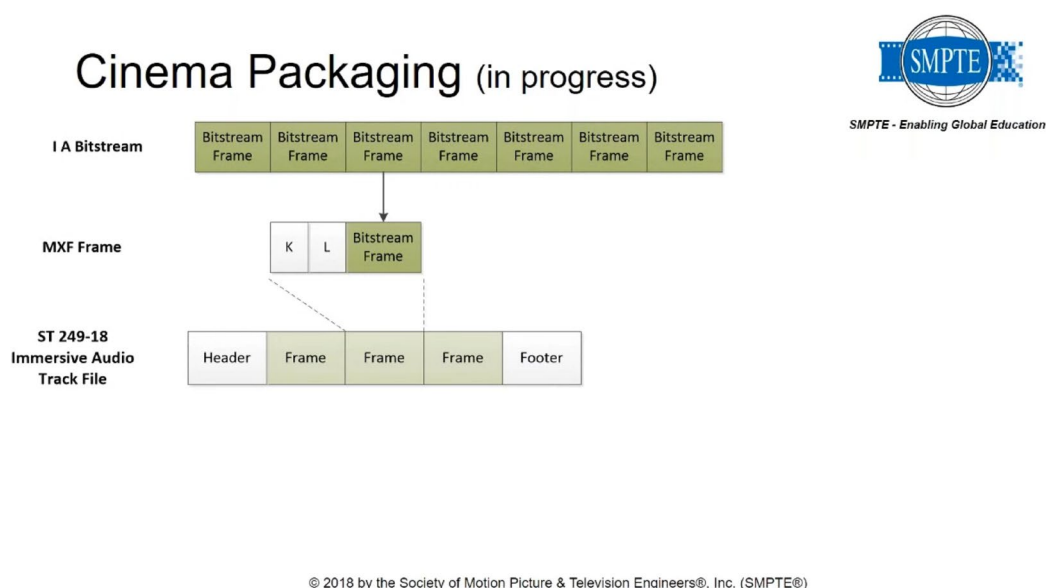


图 3-6 Stan 在 SMPTE Education 中对 IAB 内容影院打包的解释。图片截取自视频。

杜比全景声比特流的打包与此相似, 遵守 SMPTE GC Frame Wrapping。

## 第四节 浅析 0 级插件

0 级插件指定了 IMF IAB 的规范。IMF IAB 作为母版格式, 需要涵盖对后期流程更有帮助的信息。

0 级插件除了对 ST2098-2 所定义的比特流有若干修改(这些调整包括对一些参数的限制、对一些元素的调整或删除等)外, 差异主要存在于打包。0 级插件内容的打包遵照片段打包要求。与帧打包不同, 片段打包不是为单独的一个 I A Bitstream Frame 添加 KL 标头, 而是对一系列 I A Bitstream Frame 添加 KL 标头产生 KLV 编码。这样的设计更方便后期处理流程中的一些工作。

除此之外, 0 级插件规定的 MXF 封装在文件尾前有一段数据。这其中涵盖

一些 MCA 或 RFC 标准下定义的文件信息，诸如语言、内容类型、标题、版本等，以方便后期工具识别其内容。图 3-7 反映了 0 级插件的 MXF 打包方式。

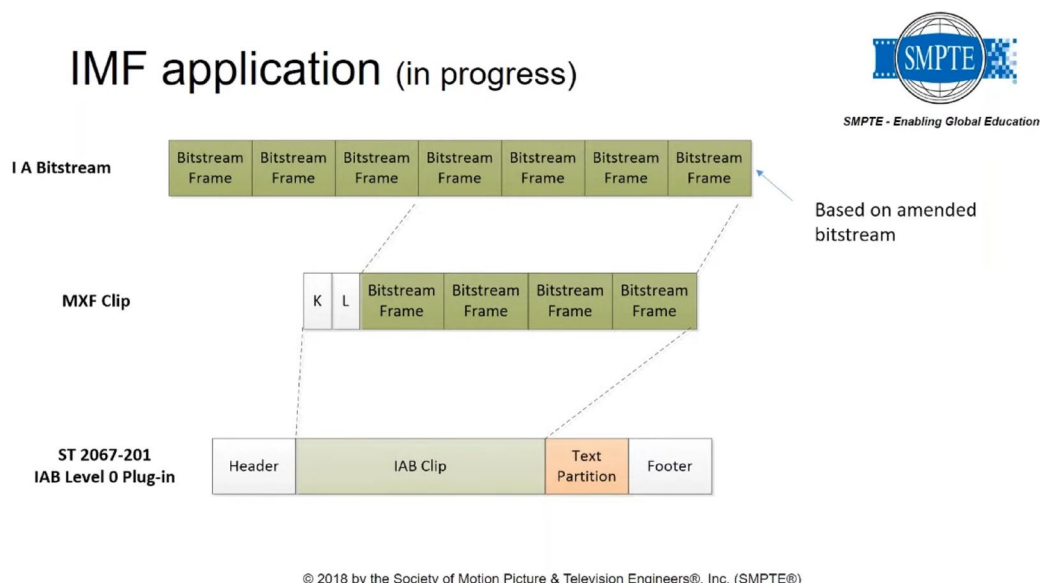


图 3-7 Stan 在 SMPTE Education 中对 IMF 应用下的 IAB 打包方式解释。图片截取自视频

## 第五节 浅析 0 级插件的杜比实现

为适应杜比开发的工具的编码与内容交换需求，使用杜比工具创建的 IMF IAB 在 0 级插件的基础上有进一步的修改。该修改在杜比官方文件《IMF IAB 互用性指导方案》（IMF IAB Interoperability Guidelines）中被定义。相关修改可以被简单描述如下：

杜比实现认为，由于 IAB 是逐帧记录数据的，当每条音轨被编码进 IAFRAME 时，音轨内容将被按帧切分。故此，必须有一个机制来跨帧重新组装每条轨道。IAB 中，在不同帧之间索引到同一条轨道的方法是依靠 MetaID。这样的方法并不方便。杜比的工具使用另一种方法来原因轨道的组装。此外，IAB 不限制每一帧中对象元素的数量，故此对象数量可能在不同帧之间发生变化。这种自由虽然提供了很大的灵活性，但可能导致编码、解码工具出现问题。故此，杜比工具重新做出了一些设计来简化其实现。

在杜比实现中，IAB 并不依赖 MetaID 来前后索引连续的音频。而是依赖于各个元素的编码顺序。例如，杜比首先编译 BedDefinition 来定义音床，而在杜比实现中音床各声道的 AudioDataPCM 会出现在 BedDefinition 前面，且顺序同 BedDefinition 中前后出现的各声道顺序相同。同样，编译 ObjectDefinition 来定

义声音对象时，声音对象的 `AudioDataPCM` 也会出现在 `ObjectDefinition` 前面来表示接下来的 `ObjectDefinition` 元素的声音数据。也就是说，杜比通过这样的方式将很多乱序的声音数据排列有序，以在帧与帧的行进过程中可以连续索引到同一轨道的声音。

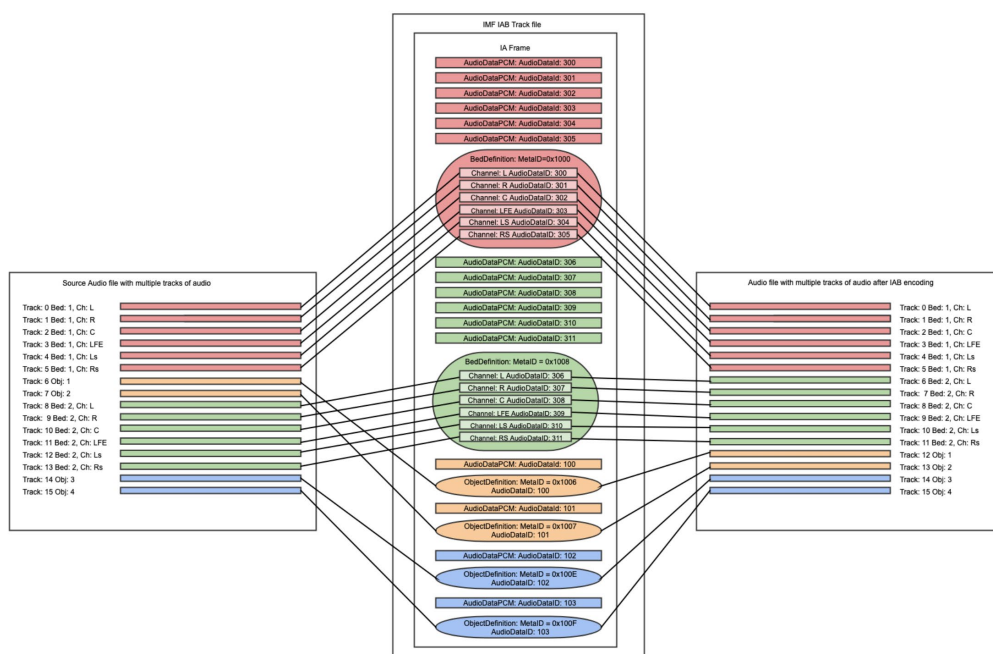


图 3-8 “IMF IAB Interoperability Guidelines”给出的杜比实现的示意图。图中可见，杜比将乱序的若干音轨通过按顺序排序，保证了前后对内容索引的连续性。图片截取自原文

此外，杜比实现限制了 IAB 的一些自由性。例如，杜比强制要求仅允许出现 1 个 7.1.2 音床和 118 个声音对象。同时，帧与帧之间声音对象的数量不能改变。

杜比实现同样针对不同的输入音床格式提供了一种名叫“扁平化”的方法来转换为 7.1.2 音床。它通过对不同音床输入的位置进行一些平方根运算来使转换后的 7.1.2 音床同原音床听感相同。

杜比实现的这些修改使得杜比工具制作处理为杜比全景声工作流程而交付的 IMF IAB 母版文件更为方便，其与其它母版格式的相似性也使得它更方便进行转换。0 级插件外加杜比实现构成了我们常见的 IMF IAB 格式母版。其常被用于嵌入 IMF 包中进行交付，制作各种各样的发行媒体。

## 第六节 浅析 ADM BWF

ADM BWF 是音频定义模型广播波形格式（Audio Definition Model Broadcast Wave Format）的简称。它是一种嵌有 ITU-R BS.2076 规定的音频定义模型数据的 EBU Tech 3285 广播波形格式。杜比全景声 ADM 配置文件的要求、建议和约



束在杜比官方文件《Dolby Atmos® Master ADM Profile》（杜比全景声®母版 ADM 配置文件）中被定义。

ADMBWF 中的声音信息同样是被离散地记录的。ADMBWF 以对象为基础记录声音数据和各项元数据。它使用一系列的 XML 元素和属性来定义音频内容的结构和元数据。这些元素包括但不限于 `audioTrackFormat`、`audioStreamFormat`、`audioChannelFormat`、`audioPackFormat` 和 `audioObject`。每个元素都有其特定的属性和限制，用于详细描述声音数据和元数据的组织方式。例如，`audioObject` 元素可以包含对应的 `audioPackFormat` 和 `audioTrackUID` 的引用，而 `audioPackFormat` 则定义了声音的空间配置。此外，`audioChannelFormat` 用于定义通道的类型和配置，如直接扬声器或对象。这种结构确保了声音和元数据的精确和灵活的表示，适应于杜比全景声等沉浸式音频体验的要求。

音频块作为离散的元数据事件，记录方式同 IAB 相似，但并非以帧为单位。每个音频块的长度以其“`interpolationLength`”属性定义。这个属性的值一般为 0.005208 秒，大约相当于 48kHz 采样率下的 250 个样本，或者 96kHz 采样率下的 500 个样本。这意味着，每个音频块通常包含大约 5.208 毫秒的音频，同时也意味着大约以每 5.208 毫秒为单位的对象空间位置等元数据的刷新。这种设计使得音频块可以非常灵活地处理音频数据和元数据，尤其适用于复杂的空间音频布局 and 动态的声音场景。

## 第四章 最佳实践推测

### 第一节 AS-DCP 库

AS-DCP（Academy / SMPTE - Digital Cinema Package）是一个由 SMPTE 工作组 DC28.20（现为 TC 21DC）开发的 C++编写的用于处理和交换数字影院包 (DCP) 和与之相关的文件的开源库。AS-DCP 提供了一系列的工具和 API，使得开发者可以创建、读取、写入和转换 DCP 和相关文件，例如影片、字幕和音频轨道。这对电影制作、发行和放映的数字化过程来说是非常重要的，因为 DCP 是电影行业广泛接受的标准格式。

在下文所述的最佳实践中，AS-DCP 库的内容打包封装、解封装功能和杜比全景声同步声道生成功能将起到帮助，这可以帮我们节省开发相关功能的麻烦。最新版本所引入的 AS-02 拓展还可以用于分析 0 级插件。

截至文章编写，AS-DCP 库被 CineCert 托管在 GitHub 上。任何人均可通过 <https://github.com/cinecert/asdcplib> 访问到其源代码。文章作者维护有一个它的 Windows x86\_64 编译版本，使用 Windows 的读者可通过 <https://github.com/izwb003/asdcplib/releases> 下载使用以省去编译它的麻烦。请读者留意，为避免与 fork 版本号冲突，我发布的 release 版本号会比最新的官方 release 高一个版本。

## 第二节 正规的杜比全景声影院内容制作流程

经过杜比认证的工作室所拥有的 RMU 单元可以用于编码杜比全景声比特流。杜比全景声比特流还可以从杜比全景声 Print Master File 经 DCDM 处理流程编码。编码后的比特流会被设备进行 SMPTE GC (或 SMPTE 382M) 打包封入 MXF 文件，生成 Dolby Atmos Cinema MXF。

Dolby Atmos Cinema MXF 会作为一个资源映射在 DCP 中嵌入合成，与普通 5.1/7.1 声音资源彼此独立存在。不兼容 Dolby Atmos 的影院声音处理器或服务器会直接忽略 Dolby Atmos 资源而使用普通 5.1/7.1 声音资源。

Dolby Atmos Cinema MXF 并不是在 SMPTE ST 429-7 中定义的资源类型，故对其进行 DCP 打包遵照 ST 429-7:2006 的 7.3.5 节“Extensions (New Asset Types)”的规定编写合成播放列表。一个示例的杜比全景声 DCP 的合成播放列表片段如下：

```
<Reellist>
  <Reel>
    <Id>urn:uuid:acaf9a9b-76ec-4cc4-8ac5-f927cd3cac08</Id>
    <AssetList>
      <MainPicture>
        <Id>urn:uuid:59fe9d81-ae8d-412b-929e-65ca64df1a35</Id>
        <EditRate>120 1</EditRate>
        <IntrinsicDuration>36000</IntrinsicDuration>
        <EntryPoint>0</EntryPoint>
        <Duration>36000</Duration>
        <KeyId>urn:uuid:f271b785-51f2-4712-9b09-f8d778c8c6a1</KeyId>
        <Hash>eMqUfSuPS88qj6lT+7GSr0UpV00=</Hash>
        <FrameRate>120 1</FrameRate>
        <ScreenAspectRatio>4096 1716</ScreenAspectRatio>
      </MainPicture>
      <MainSound>
```

```

        <Id>urn:uuid:c4ccb104-2205-4049-9fb5-92ccbf9fe175</Id>
        <EditRate>120 1</EditRate>
        <IntrinsicDuration>36000</IntrinsicDuration>
        <EntryPoint>0</EntryPoint>
        <Duration>36000</Duration>
        <KeyId>urn:uuid:8eb5550e-c3c7-4eb0-9154-
ab80065a0d78</KeyId>
        <Hash>8bY0u/caTZWieazqVY8i5tqvXZU=</Hash>
        <Language>en</Language>
    </MainSound>
    <axd:AuxData
xmlns:axd="http://www.dolby.com/schemas/2012/AD">
        <Id>urn:uuid:3d97f274-4cb3-452e-a69f-ca5a46f10851</Id>
        <EditRate>120 1</EditRate>
        <IntrinsicDuration>263105</IntrinsicDuration>
        <EntryPoint>0</EntryPoint>
        <Duration>36000</Duration>
        <KeyId>urn:uuid:246a30a8-2ea9-4e03-a9b8-
300140763398</KeyId>
        <Hash>SxdANLqsbvfsohvL0YiJqAK40xs=</Hash>
        <axd:DataType>urn:smppte:ul:060e2b34.04010105.0e090604.00
000000</axd:DataType>
    </axd:AuxData>
</AssetList>
</Reel>
</ReelList>

```

由此可见，杜比全景声内容单独使用标签 `axd:AuxData` 指定，并用单独的 `axd:DataType` 标签标记 SMPTE UL 060e2b34.04010105.0e090604.00000000 来让设备识别其为全景声内容。

除了 CPL 的正确映射外，杜比全景声 DCP 还依赖在普通 5.1/7.1 音频中第 14 或 16 轨道中记录的杜比全景声同步音轨。杜比全景声 DCP 的普通 5.1/7.1 资源共有 16 个声音轨道，其中第 14 或 16 轨道为杜比全景声同步信号所记录的位置。杜比全景声影院处理器依赖实时读取该轨道信息来确保读取全景声时音画同步。

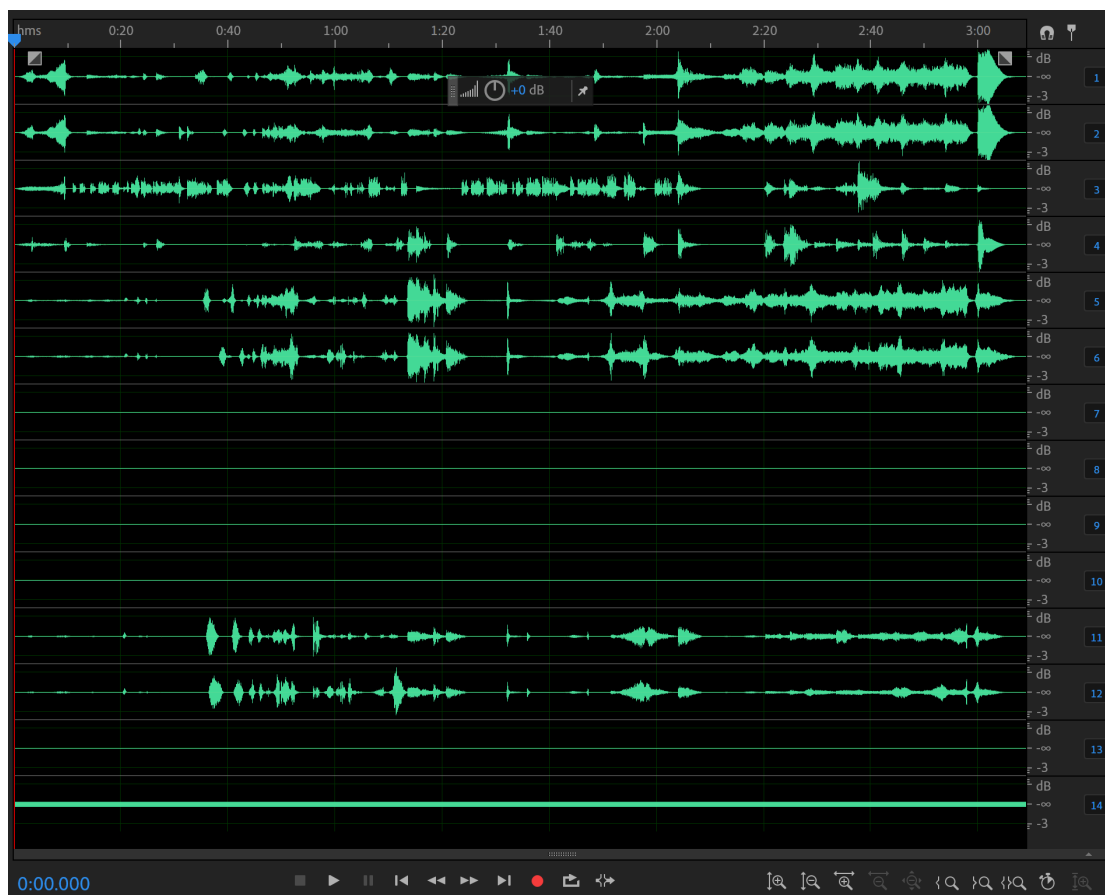


图 4-1 杜比全景声 DCP 的普通 7.1 音轨组成。可以看到，第 14 音轨中记录有同步信息。

最后需要留意的是为杜比全景声 DCP 给定正确的 ISDCF 命名。杜比全景声要求在 Audio Type 字段中添加子字段-Atmos。

杜比全景声 DCP 必须是 SMPTE DCP。它不能是 Interop DCP。

一个正确的杜比全景声 DCP 的要求如下：

- 正确的杜比全景声比特流和正确的 MXF 封装
- 正确的资源映射；
- 正确的合成播放列表信息所构建的正确的合成；
- 正确在 5.1/7.1 普通音频资源中写入杜比全景声同步音轨；
- 正确的 ISDCF DCNC 命名。
- 正确的 SMPTE DCP 结构。

### 第三节 正规的 IAB 影院内容制作流程

由于 IAB 尚未普及，该流程尚未能得到实际应用。在 IAB 得以铺开，IAB

影院内容的制作将按照如下流程进行：

1. 专门的 IAB 渲染器用于编码 IAB。
2. 专门的 IAB 打包工具用于根据 ST 429-18 封装 IAB 为 MXF 文件。
3. DCP 封装工具根据 ST 429-19，参考 ST 429-18 建立资源映射、合成播放列表。
4. 如果 IAB 是符合配置文件 1 的，则其还需要根据配置文件 1 的要求建立合成播放列表。
5. 正确的 ISDCF DCNC 命名。IAB 要求在 Audio Type 字段中添加子字段-IAB。

ST 429-18 中对合成播放列表的建立也是根据 ST 429-7:2006 的 7.3.5 节“Extensions (New Asset Types)”的规定编写的。

## 第四节 所期待的最佳实践

该最佳实践的最终目标是在杜比全景声影院中播放自制的能达到完整全景声效果的内容。

目前，已知的结论如下：

- IAB 标准已公开，IAB 配置文件 1 已披露。
- IAB 配置文件 1 与杜比全景声比特流等价。
- 杜比全景声比特流可以使用 AS-DCP 封装为 MXF。这个流程同样应该作用于 IAB 配置文件 1。
- DCP-o-matic 2 等常见的 DCP 打包软件支持导入杜比全景声影院内容。这个流程同样应该作用于 IAB 配置文件 1 自行封装的 MXF。

故此，我们可提供下述的一种方法：

- 根据 ST 2098-2 和配置文件 1 编写 IAB 配置文件 1 的编码器。
- 使用该编码器编码 IAB 配置文件 1 比特流。
- 使用 AS-DCP 封装 IAB 配置文件 1 比特流为 Dolby Atmos Cinema MXF。
- 将该 MXF 导入 DCP-o-matic 制作完整 DCP。

该实践的所有难度皆在于步骤 1，这也正是本文所希望鼓励的进一步探索方向。ST 2098-2 和配置文件 1 是公开的，而并非商业机密或严格专利。熟悉比特流操作的人士理应可以根据两份文件编写 IAB 配置文件 1 的编码器。但是，其难度还是相当大的，这一点值得留意。

步骤 1 的编码器应当可以通过读取 ADM BWF 或 IMF IAB 文件来作为输入。针对前者应参考《Dolby Atmos® Master ADM Profile v1.1》进行解码，针对后者应参考 0 级插件和《Dolby Atmos IMF IAB Interoperability Guidelines》进行解码。由于 0 级插件和 ST 2098-2 比特流高度相似，后者的办法应当可以复用大量内容。

核心难点应当在于 DLC 无损音频编码。ST 2098-2 在附录 B 中给出了该编码的解释，和一个最简单的编码器示例。但是，实现它需要熟悉音频编码。这需要经验丰富的人士来完成。不过，如果不使用 DLC 编码的一些进阶特性，该编码可以通过对 PCM 进行简单的移位工作完成，故此其实现难度不应过大。

截至文章编写，AS-DCP 库还不支持根据 ST 429-18 封装 IAB。故此，虽然 IAB DCP 理应可以被全景声影院解析，但完整的 IAB 制作流程可行性较低。故此，我们需利用配置文件 1 同杜比全景声比特流等价的特点，直接将 IAB 配置文件 1 封装为 Dolby Atmos Cinema MXF。其可行实践如下：

将每个 I A Bitstream Frame 编写为一个单独的文件。文件名需以 .atmos 作为后缀名，且文件名应为包含前导 0 帧编号。例如，DCPIAB000009.atmos，代表比特流的第九个 I A Bitstream Frame。将所有编码后的 I A Bitstream Frame 文件置于同一个文件夹中，确保其前后顺序为按帧顺序排列（包含前导 0 的帧编号可以实现这一点）。

使用 AS-DCP 编写程序完成封装。也可以使用 asdcplib 所提供的实用程序 asdcp-wrap：

```
asdcp-wrap <Folder Name> <Output File Name>
```

打包成品即为 Dolby Atmos Cinema MXF。如图 4-2 所示。



图 4-2 asdcplib 封装的 Dolby Atmos Cinema MXF。

对封装成品的测试不一定必须交付影院。可以在本机使用 Dolby Atmos Conversion Tool 或 Dolby Atmos Renderer 完成。两个工具均可于杜比官方网站获取（前者甚至完全免费）。这些工具都可以识别 Dolby Atmos Cinema MXF。如果使用这些工具打开生成的 Dolby Atmos Cinema MXF 没有发生报错，即可以推测封装是成功的。

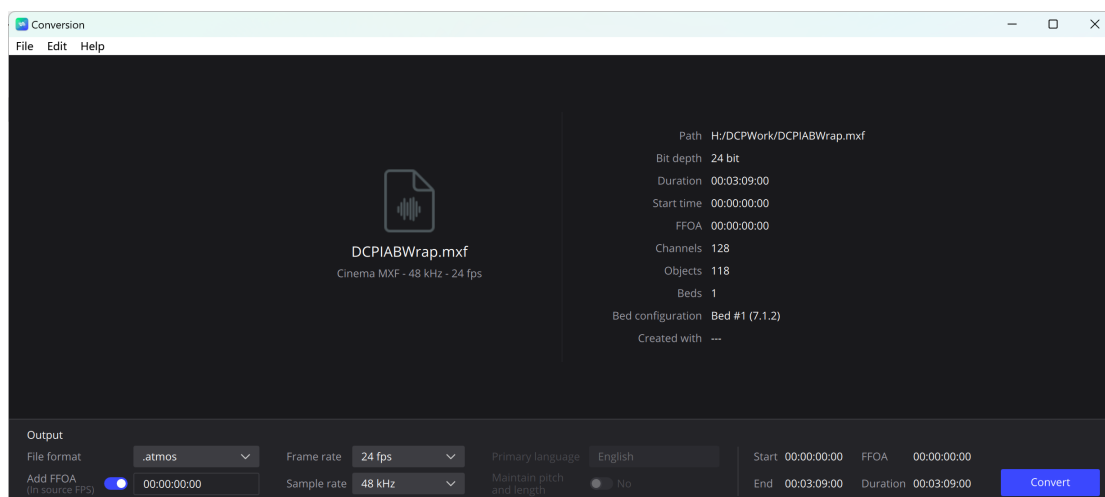


图 4-3 使用 Dolby Atmos Conversion Tool 打开封装的 MXF 没有报错。

封装后的内容可以直接使用 DCP-o-matic 2 进行 DCP 封装。最新版本的 DCP-o-matic 2 支持完整的杜比全景声内容封装流程。直接将生成的 Dolby Atmos

Cinema MXF 当作内容添加进 DCP-o-matic 2 即可。DCP-o-matic 2 将处理 CPL 映射与同步音轨添加等工作。

应特别注意制作的内容的响度问题。DCP-o-matic 2 无法像普通音频一样调整全景声的响度等参数。

## 第五节 现行的可行解决方案

在正式的解决方案出现之前, 现行可以通过市面上杜比全景声处理器 (CP850, CP950A) 等的“影院增值内容 (Alternative Content)”功能播放民用产品格式。通过使用 HDMI 连接这些处理器, 设备可以解码来自 HDMI 的源码全景声音频信号。使用支持的全景声回放设备 (如蓝光机、兼容杜比全景声功能的计算机) 等便可以通过 HDMI 连接来播放全景声内容。

但是, 民用产品格式的全景声质量不能同影院内容相媲美。以 Enhanced AC-3 with Joint Object Coding 为例, 其声音对象数量被缩减为仅有十几个。相比母版级别可以做到完整 128 个渲染内容的杜比全景声影院内容, 其听感会大幅下降。

有关具体细节, 应参阅这些影院处理器产品的用户手册。



## 结论

得益于统一标准的制定，本文证明了凭公开资料制作本为商业内容的全景声影院内容的可行性。但是，相关实现仍然需要有比特流处理和音频编码经验的人士来完成。

文章作者仍需提醒，虽然其操作理论可行，但进行相关推进时仍需仔细检查并思考其中可能存在的知识产权、商业机密或专利权风险。实现者应广泛查找资料和阅读相关披露信息来注意实现中可能存在的权益性问题。再次强调，本文仅为实现可行性分析，旨在对相关感兴趣人士提供参考帮助，不能作为任何专业工作的工程指导。

文章作者认为，在数字化深刻普及的今天，为充分发挥创造性，内容的制作应当具有足够便利的渠道以方便创作者。统一标准的制订和披露公开对此有着促进和推动作用，是应当被支持的。

## 参考文献

- [1]欧阳珣.听得见的传奇——杜比的品牌之路[M].北京:企业管理出版社,2015:44-50.
- [2]侯佳俊.论沉浸式声音技术在现代音频产业链中的应用[D].上海音乐学院,2023.DOI:10.27319/d.cnki.gsyty.2023.000147.
- [3]Dolby Laboratories. Dolby Atmos Next-Generation Audio for Cinema (Issue 3)[EB/OL]. <https://www.mixagefou.com/IMG/pdf/dolby-atmos-next-generation-audio-for-cinema-white-paper.pdf>,2014.
- [4]RDD57:2021, SMPTE ST 2098-2 Immersive Audio Bitstream and Packaging Constraints: IAB Application Profile 1[S].
- [5]Dolby Laboratories. Dolby Atmos IMF IAB Interoperability Guidelines[EB/OL]. [https://professionalsupport.dolby.com/s/article/Dolby-Atmos-IMF-IAB-interoperability-guidelines?language=en\\_US](https://professionalsupport.dolby.com/s/article/Dolby-Atmos-IMF-IAB-interoperability-guidelines?language=en_US),2021-09-24.
- [6]RDD29:2014, Dolby Atmos® Bitstream Specification[S].
- [7]ST 2098-2:2019, Immersive Audio Bitstream Specification[S].
- [8]ST 2067-201:2021, Immersive Audio Bitstream Level 0 Plug-in[S].
- [9]Dolby Laboratories. Dolby Atmos® Master ADM Profile v1.1[EB/OL]. [https://professionalsupport.dolby.com/s/article/Dolby-Atmos-ADM-Profile-specification?language=en\\_US](https://professionalsupport.dolby.com/s/article/Dolby-Atmos-ADM-Profile-specification?language=en_US),2022-01-05.