



(12) 发明专利

(10) 授权公告号 CN 103250207 B

(45) 授权公告日 2016. 01. 20

(21) 申请号 201180053153. 7

H04S 3/00(2006. 01)

(22) 申请日 2011. 10. 26

(56) 对比文件

(30) 优先权数据

10306211. 3 2010. 11. 05 EP

CN 1495705 A, 2004. 05. 12,

CN 1677490 A, 2005. 10. 05,

EP 2205007 A1, 2010. 07. 07,

CN 101872618 A, 2010. 10. 27,

(85) PCT国际申请进入国家阶段日

2013. 05. 03

审查员 张飞弦

(86) PCT国际申请的申请数据

PCT/EP2011/068782 2011. 10. 26

(87) PCT国际申请的公布数据

W02012/059385 EN 2012. 05. 10

(73) 专利权人 汤姆逊许可公司

地址 法国伊西莱穆利诺

(72) 发明人 F. 基勒 S. 科登 J. 贝姆

H. 克罗普 J-M. 巴特克

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 吕晓章

(51) Int. Cl.

G10L 19/008(2013. 01)

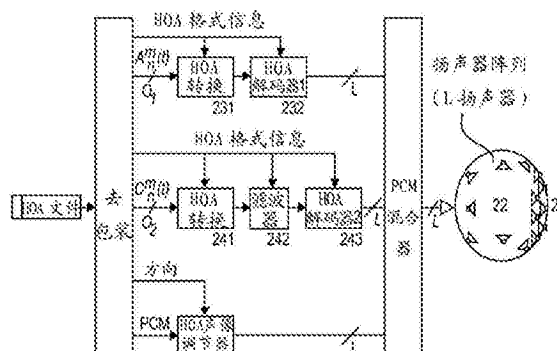
权利要求书1页 说明书41页 附图9页

(54) 发明名称

高阶高保真度立体声响复制音频数据的数据结构

(57) 摘要

本发明涉及一种高阶高保真度立体声响复制HOA音频数据的数据结构,该数据结构包括用于一个或多个不同HOA音频数据流描述的2D或3D空间音频内容。HOA音频数据具有大于'3'的阶,该数据结构另外还包括单音频信号源数据和/或来自固定或时变空间位置的麦克风阵列音频数据。



1. 一种音频呈现方法,其中,接收包含至少两个不同 HOA 音频数据信号的 HOA 音频数据流,并且使用 (231, 232) 它们中的至少第一个 HOA 音频数据信号以供位于呈现场所 (10) 的不同区域处的密集的扬声器布置 (11, 21) 进行呈现,使用 (241, 242, 243) 它们中的至少第二个并且是不同的 HOA 音频数据信号以供环绕所述呈现场所 (10) 的不密集的扬声器布置 (12, 22) 进行呈现,其中,用于所述密集的扬声器布置 (11, 21) 的所述音频数据表示球面波和第一高保真度立体声响复制阶,用于所述不密集的扬声器布置 (12, 22) 的所述音频数据表示平面波和 / 或小于所述第一高保真度立体声响复制阶的第二高保真度立体声响复制阶。

2. 如权利要求 1 所述的音频呈现方法,其中,所述呈现场所是电影院中的收听区域或座位区域。

3. 如权利要求 1 所述的音频呈现方法,其中,高保真度立体声响复制系数的至少部分被带宽降低,使得对于不同的 HOA 阶,相关的高保真度立体声响复制系数的带宽是不同的 (1221-122N)。

4. 如权利要求 3 所述的音频呈现方法,其中,所述带宽降低是基于 MDCT 处理 (1431-143M)。

5. 一种音频呈现装置,包括:

- 用于接收包含至少两个不同 HOA 音频数据信号的 HOA 音频数据流的部件;
- 用于处理它们中的至少第一个 HOA 音频数据信号以供位于呈现场所 (10) 的不同区域处的密集的扬声器布置 (11, 21) 进行呈现的部件 (231, 232);
- 用于处理它们中的至少第二个并且是不同的 HOA 音频数据信号以供环绕所述呈现场所 (10) 的不密集的扬声器布置 (12, 22) 进行呈现的部件 (241, 242, 243), 其中,用于所述密集的扬声器布置 (11, 21) 的所述音频数据表示球面波和第一高保真度立体声响复制阶,用于所述不密集的扬声器布置 (12, 22) 的所述音频数据表示平面波和 / 或小于所述第一高保真度立体声响复制阶的第二高保真度立体声响复制阶。

6. 如权利要求 5 所述的音频呈现装置,其中,高保真度立体声响复制系数的至少部分被带宽降低,使得对于不同的 HOA 阶,相关的高保真度立体声响复制系数的带宽是不同的 (1221-122N)。

7. 如权利要求 6 所述的音频呈现装置,其中,所述带宽降低是基于 MDCT 处理 (1431-143M)。

高阶高保真度立体声响复制音频数据的数据结构

技术领域

[0001] 本发明涉及一种高阶高保真度立体声响复制(Ambisonics)音频数据的数据结构, 其包括二维和 / 或三维空间音频内容数据, 并且也适用于具有大于 '3' 阶的高阶高保真度立体声响复制(Higher Order Ambisonics (HOA)) 音频数据。

背景技术

[0002] 3D 音频可以通过被称为如下所述的高阶高保真度立体声响复制(HOA) 的技术使用声场描述来实现。存储 HOA 数据需要一些协定和约定, 其约定特定解码器必须如何使用该数据才能够创建用于在给定再现扬声器设置上进行重播的扬声器信号。没有现有的存储格式定义 HOA 的所有这些约定。例如在 2009 年 3 月 30 日在 Martin Leese "File Format for B-Format", <http://www.ambisonia.com/Members/etienne/Members/mleese/file-format-for-b-format> 中描述的(基于可扩展 'Riff/wav' 结构)的 B- 格式及其 *.amb 文件格式实现是当今可用的最复杂的格式。.amb 文件格式在 2000 年在 R.W.Dobson 的 "Developments in Audio File Formates" ICMC 柏林 2000 年, 中呈现。

[0003] 在 2010 年 7 月 16 日, 在高保真度立体声响复制交换站点(Ambisonics Xchange Site) 上在 "Existing formats", <http://ambisonics.iem.at/xchange/format/existing-formats> 中公开了现有文件格式的概述, 并且在该站点上在 "A first proposal to specify, define and determine the parameters for an Ambisonics exchange format", <http://ambisonics.iem.at/xchange/format/a-first-proposal-for-the-format> 中还公开了对于高保真度立体声响复制交换格式的建议。

发明内容

[0004] 关于 HOA 信号, 对于 3D, 可以全部记录(编码)以相同频率从不同声源采集 $M=(N+1)^2$ 个(对于 2D 是 $(2N+1)$ 个)不同的音频对象, 并且将其再现为不同的声音对象, 假设他们是均匀空间分布的。这意味着, 1 阶高保真度立体声响复制信号可以携带四个 3D 或三个 2D 音频对象, 并且这些对象需要围绕着 3D 的球面或者围绕 2D 的圆均匀分隔开。在记录中空间重叠和多于 M 个信号将导致模糊, 仅仅可以将最大声信号再现为相干对象, 而其它弥漫性信号取决于空间上的重叠、频率和响度(loudness)相似度将在一定程度上使相干信号退化。

[0005] 关于电影院中的声学情况, 对于前屏幕区域, 要求高空间声音定位精度, 以便配合视觉场景。环绕声对象的感知较不重要(混响(reverb)、与视觉场景无关的声音对象)。这里, 扬声器的密度与前面区域相比可以更小。

[0006] (与前面区域相关的) HOA 数据的 HOA 阶需要大, 以便使得能够选择声音全息(holophonic)重播。典型的阶是 $N=10$ 。这要求 $(N+1)^2=121$ 个 HOA 系数。理论上, 如果这些音频对象将均匀地空间分布, 我们还可以编码 $M=121$ 个音频对象。但是, 在我们的情形

下,它们被限制到前面区域(因为我们仅仅在这里需要这样高阶)。实际上,在没有模糊的情况下,我们可能仅仅编码大约 $M=60$ 个音频对象(前面区域至多是球体方向的一半,因此是 $M/2$)。

[0007] 关于上述 B- 格式,其使能了仅仅高达 3 高保真度立体声响复制阶的描述,而文件大小被限制为 4GB。其它空间信息丢失,诸如对现代解码器重要的波类型或参考解码半径。不可能对不同的高保真度立体声响复制分量(声道)使用不同的采样格式(字宽度)和带宽。也没有对于存储用于高保真度立体声响复制的辅助信息和元数据的标准化。

[0008] 在现有技术中,使用麦克风阵列记录高保真度立体声响复制信号被限制为一阶。如果将开发 HOA 麦克风的试验原型,这将在将来改变。对于 3D 内容的创建,可以在一阶高保真度立体声响复制下使用麦克风阵列来记录环境声场的描述,由此方向性源可以被使用特写单声道(closed-up mono)麦克风或者高方向性麦克风与方向性信息一起(即该源的位置)来捕捉。然后,方向性信号可以被编码为 HOA 描述,或者这可以由复杂解码器执行。无论如后,需要新的高保真度立体声响复制文件格式能够一次存储多于一个声场描述,但是没有现有格式能够封装多于一个高保真度立体声响复制描述。

[0009] 本发明要解决的问题是提供一种高保真度立体声响复制文件格式,其能够一次存储两个或多个声场描述,其中高保真度立体声响复制阶可以大于 3。

[0010] 为了创建逼真的 3D 音频,下一代高保真度立体声响复制解码器将要求处理大量协定和约定以及存储数据,或者可以相干地存储所有相关参数和数据元素的单文件格式。

[0011] 空间声音内容的发明性文件格式可以存储一个或多个 HOA 信号、以及 / 或者方向性的单声道信号和方向性信息,其中高保真度立体声响复制阶大于 3 以及文件大于 4GB 是可行的。此外,发明性文件格式提供了现有格式没有提供的附加元素:

[0012] 1) 在该文件格式中存储了下一代 HOA 解码器所需的重要信息:

[0013] 一高保真度立体声响复制波信息(平面、球面、混合类型)、感兴趣区域(收听区域外部的源或收听区域内的源)、以及(用于球面波解码的)参考半径

[0014] 一可以存储相关的方向性的单声道信号。可以使用角度和距离信息、或者使用高保真度立体声响复制系数的编码矢量来描述这些方向性信号的位置信息。

[0015] 2) 在辅助信息中包含定义高保真度立体声响复制数据的所有参数,以便确保记录清楚:

[0016] 一高保真度立体声响复制缩放和规范化(SN3D、N3D、福尔斯马勒姆(Furse Malham)、B 格式、...、用户定义)、混合阶信息。

[0017] 3) 扩展高保真度立体声响复制数据的存储格式以允许灵活且经济地存储数据:

[0018] 一发明性的格式允许利用不同的 PCM- 字大小的分辨率并且使用限制的带宽来存储与高保真度立体声响复制阶(高保真度立体声响复制声道)相关的数据。

[0019] 4) 元字段允许存储关于文件的附属信息(如用于麦克风信号的记录信息):

[0020] 一记录参考坐标系、麦克风、源和虚拟收听者位置、麦克风方向性特性、房间和源信息。

[0021] 用于 2D 和 3D 音频内容的该文件格式覆盖高阶高保真度立体声响复制描述(HOA)以及具有固定或时变位置的单源的存储,并且包含使得下一代音频解码器能够提供逼真的 3D 音频的所有信息。

[0022] 使用适当设置,发明性的文件格式还适用于音频内容的流传输。因此,可以在文件创建者选择的时间实例处发送依赖于内容的辅助信息(头部数据)。发明性文件格式也用作场景描述,其中音频场景的音轨可以在任何时间处开始和结束。

[0023] 原理上,发明性的数据结构适用于高阶高保真度立体声响复制 HOA 音频数据,其数据结构包括用于一个或多个不同 HOA 音频数据流描述的 2D 和 / 或 3D 空间音频内容数据,并且其数据结构也适用于具有大于 '3' 阶的 HOA 音频数据,另外其数据结构可以包括单音频信号源数据和 / 或来自固定或时变空间位置的麦克风阵列音频数据。

[0024] 原理上,发明性的方法适用于音频呈现,其中接收包含至少两个不同 HOA 音频数据信号的 HOA 音频数据流,并且在具有位于呈现点的不同区域处的密集扬声器布置的情况下至少将其中的第一个用于呈现,并且在具有环绕所述呈现点的低密集扬声器布置的情况下至少将其中的至少第二个和不同的一个用于呈现。

附图说明

[0025] 参考附图描述本发明的实例实施例,在附图中:

[0026] 图 1 是在前面区域具有密集扬声器布置而围绕收听区域具有稀疏扬声器密度的电影院中的声音全息再现;

[0027] 图 2 是复杂的解码系统;

[0028] 图 3 是从麦克风阵列记录、单源记录、简单和复杂声场生成的 HOA 内容创建;

[0029] 图 4 是下一代浸入式内容创建;

[0030] 图 5 是简单环绕扬声器设置的 HOA 信号的 2D 解码、以及前面阶段的声音全息扬声器设置以及更稀疏 3D 环绕扬声器设置的 HOA 信号的 3D 解码;

[0031] 图 6 是内部域问题,其中源在感兴趣区域 / 有效区域外部;

[0032] 图 7 是球面坐标的定义;

[0033] 图 8 是外部域问题,其中源在感兴趣区域 / 有效区域内部;

[0034] 图 9 是简单示例的 HOA 文件格式;

[0035] 图 10 是包含多帧和多音轨的 HOA 文件的示例;

[0036] 图 11 是具有多个元数据块(MetaDataChunk)的 HOA 文件;

[0037] 图 12 是音轨区域编码处理;

[0038] 图 13 是音轨区域解码处理;

[0039] 图 14 是使用 MDCT 处理降低带宽的实现;

[0040] 图 15 是使用 MDCT 处理重构带宽的实现。

具体实施方式

[0041] 随着 3D 视频的日益蔓延,浸入式的音频技术正在成为用以区分的有趣特征。高阶高保真度立体声响复制(HOA)是这些能够提供以增量方式将 3D 音频引入到电影院中的方法的技术之一。使用 HOA 声音音轨和 HOA 解码器,电影院可以以现有的音频环绕扬声器设置开始并且逐步投入更多扬声器,逐步提高浸入式体验。

[0042] 图 1a 示出了在前面区域具有密集扬声器布置 11 而围绕收听或座位区域 10 具有稀疏扬声器密度 12 的电影院中的声音全息再现,提供了精确再现与视觉动作相关的声音

以及足够精确再现所再现的环境声音的方法。

[0043] 图 1b 示出了所再现的前面声波到达的感知方向,其中平面波到达的方向匹配不同的屏幕位置,即平面波适用于再现深度。

[0044] 图 1c 示出了所再现的球面波到达的感知方向,其导致所感知的声音方向和围绕屏幕的 3D 视觉动作的更高一致度。

[0045] 由于在电影院中主视觉动作在收听者的前面区域中发生的事实,引起了对两种不同的 HOA 流的需要。而且,对于前面声源的检测声音方向的感知精度比对于环绕源的检测声音方向的感知精度要高。因此,前面空间声音再现的精度需要比用于所再现的环境声音的空间精度高。对于前面屏幕区域要求用于声音再现的声音全息部件、大量扬声器、专用解码器和相关扬声器驱动器,而对于环境声音再现需要更低成本的技术(环绕收听区域的较低密度的扬声器、以及更不完美的解码技术)。

[0046] 由于内容创建和声音再现技术,有利的是为环境声音提供一个 HOA 再现并且为前景动作声音提供一个 HOA 再现,参见图 4。使用简单设置以及简单的稀疏再现声音装备的电影院可以在解码之前混合两个流(参见图 5 上部 分)。

[0047] 装备有完全浸入式再现部件的更复杂的电影院可以使用两个解码器,一个用于解码环境声音,而另一个专用解码器用于前景主动作的虚拟声源的高精度定位,如在图 2 的复杂解码系统以及图 5 下部分中示出的。

[0048] 具体 HOA 文件包含至少两个音轨,其呈现用于环境声音 $A_n^m(t)$ 以及用于与虚拟主动作相关的前面声音 $C_n^m(t)$ 的 HOA 声场。可以提供用于方向性效果的可选流。两个对应的解码器系统与声像调节器(panner)一起提供用于密集的前面 3D 声音全息扬声器系统 21 和较低密集的(即稀疏的) 3D 环绕系统 22 的信号。

[0049] 音轨 1 流的 HOA 数据信号呈现环境声音,并且在 HOA 转换器 231 中被转换以输入到专用于环境再现的解码器 1 (232)。对于音轨 2 数据流,HOA 信号数据(与虚拟场景相关的前面声音)在 HOA 转换器 241 中被转换以输入到距离校正(方程(26))滤波器 242,以便以专用解码器 2 (243)围绕屏幕区域更好地放置球面声音源。方向性的数据流被直接声像调节(pan)到 L 扬声器。三个扬声器信号被 PCM 混合以便利用 3D 扬声器系统进行联合再现。

[0050] 看上去没有已知的文件格式专用于这样的情形。已知的 3D 声场记录使用具有相关声音音轨的完整场景描述,或者在存储以用于稍后再现时使用单声音场描述。第一类的示例是 WFS (波场合成) 格式和各种容器格式。第二类的示例是像 B 或 AMB 格式的高保真度立体声响复制格式,参见上述文章“File Format for B-Format”。后者限于高保真度立体声响复制阶为三、固定传输格式、固定解码器型号和单声音场。

[0051] HOA 内容创建和再现

[0052] 在图 3 中描绘了生成 HOA 声音唱描述的处理。

[0053] 在图 3a 中,通过使用麦克风阵列来创建声音场的自然记录。采集信号(capsule signal)被矩阵化和均衡化,以便形成 HOA 信号。高阶信号(高保真度立体声响复制阶 >1)通常被带通滤波以降低由于采集距离效应引起的伪像,被低通滤波以降低高频处的空间混声(spatial alias),被高通滤波以在增加高保真度立体声响复制阶 $nh_n(kr_{d\text{mic}})$ 的情况下降低额外的低频电平,参见方程(34)。可选地,可以应用距离编码滤波,参见方程(25)和

(27)。在存储之前, HOA 格式信息被添加到音轨头部。

[0054] 通常使用多个方向性的单源流来创建人工声音场再现。如图 3b 中所示, 可以将单源信号捕获为 PCM 记录。这可以通过特写(close-up) 麦克风进行或者通过使用具有高方向度的麦克风进行。另外, 记录声音源相对于虚拟最佳收听位置的方向性参数(r_s, Θ_s, Φ_s) (HOA 坐标系, 或者用于稍后映射的任何参考点)。当呈现电影场景时, 还可以通过人工地放置声音来创建距离信息。如图 3c 中所示, 然后使用方向性信息(Θ_s, Φ_s) 来创建编码矢量 ψ , 将方向性的源信号编码为高保真度立体声响复制信号, 参见方程(18)。这等效于平面波再现。拖尾滤波处理可以使用距离信息来将球面源特征印记到高保真度立体声响复制信号(方程(19))中, 或者应用距离编码滤波(方程(25) (27))。在存储之前, HOA 格式信息被添加到音轨头部。

[0055] 如图 3d 中所描绘的, 通过 HOA 混合高保真度立体声响复制信号, 生成更复杂的波场描述。在存储之前, HOA 格式信息被添加到音轨头部。

[0056] 在图 4 中描绘了用于 3D 电影院的内容生成的处理。利用高空间精度编码与视频动作相关的前面声音, 并且将其与 HOA 信号(波场) $C_n^m(t)$ 混合并存储为音轨 2。所包含的编码器以高空间精度和最佳匹配视频场景所需的特定波类型来进行编码。音轨 1 包含声场 $A_n^m(t)$ 其与所编码的环境声音相关而不限源方向。通常, 环境声音的空间精度不需要与前面声音的空间精度那样高(因此, 高保真度立体声响复制阶可以更小), 并且波类型的建模将更不重要。环境声音场还可以包括前面声音信号的回响(reverberant)部分。两个音轨被多路复用以便存储和 / 或交换。

[0057] 可选地, 可以将方向性的声音(例如, 音轨 3)多路复用到文件。这些声音可以是特殊效果的声音、对话或运动信息(像用于视觉障碍的叙事演讲)。

[0058] 图 5 示出了解码的原理。如上部分所描绘的, 在简化的 HOA 解码之间, 具有稀疏扬声器设置的电影院可以将来自音轨 1 和音轨 2 的两个 HOA 信号混合, 并且可以截断音轨 2 的阶并且将两个音轨的维度降低为 2D。在出现方向性的流的情况下, 将其编码为 2D HOA。然后, 所有三个流被混合以形成单 HOA 表示, 其然后被解码并再现。

[0059] 下部分对应于图 2。装备有用于前面阶段的声音全息系统和稀疏的 3D 环绕系统的电影院将使用专用的复杂解码器, 并且混合扬声器供给。对于音轨 1 数据流, 表示环境声音的 HOA 数据被转换到专用于再现环境的解码器 1。对于音轨 2 数据流, HOA (与视觉场景相关的前面声音) 被转换并且被距离校正(方程(26))以便用于更好地围绕屏幕区域放置球面声音源以及用于专用解码器 2。方向性的数据流被直接声像调节到 L 扬声器。三个扬声器信号被 PCM 混合以便利用 3D 扬声器系统进行联合再现。

[0060] 使用高阶高保真度立体声响复制的声音场描述

[0061] 使用球面谐波(SH)的声音场描述

[0062] 当使用球面谐波 / 贝塞尔描述时, 在方程(1)中提供了声波方程的解, 参见 M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics", Journal of Audio Engineering Society, 53(11), pp. 1004 - 1025, 2005 年 11 月、以及 Earl G. Williams, "Fourier Acoustics", Academic Press, 1999 年。

[0063] 声压是球面坐标 r, Θ, Φ (对于其定义请参见图 7) 和空间频率的函数 $k = \frac{\omega}{c} = \frac{2\pi f}{c}$ 。

[0064] 该描述对于感兴趣区域或有效区域外部的音频声音源有效(内部域问题,如图6所示),并且假设正交规范化的球面谐波:

$$[0065] \quad p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(k) j_n(kr) Y_n^m(\theta, \phi) \quad (1)$$

[0066] $A_n^m(k)$ 被称为高保真度立体声响复制系数, $j_n(kr)$ 是第一类的球面贝塞尔函数, $Y_n^m(\theta, \phi)$ 被称为球面谐波(SH), n 是高保真度立体声响复制阶索引, m 指示程度。

[0067] 由于贝塞尔函数仅仅对于小 kr 值(距原点的小距离或者低频率)具有显著值的特性,因此序列可以在某阶 n 处停止并且以足够精度限制于值 N 。当存储 HOA 值时,通常存储直至阶 N 的高保真度立体声响复制系数 A_n^m, B_n^m 或一些导数(下面描述细节)。 N 被称为高保真度立体声响复制阶。

[0068] N 被称为高保真度立体声响复制阶,项“阶”通常也与贝塞尔 $j_n(kr)$ 和汉克尔 $h_n(kr)$ 函数中的 n 组合使用。

[0069] 对于 $r > r_{\text{Source}}$,在方程(2)中表达了用于外部情况的波方程的解,其中源位于感兴趣区域或有效区域内,如图8所描绘的。

$$[0070] \quad p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n B_n^m(k) h_n^{(1)}(kr) Y_n^m(\theta, \phi) \quad (2)$$

[0071] $B_n^m(k)$ 再次被称为高保真度立体声响复制系数, $h_n^{(1)}(kr)$ 表示第一类和第 n 阶的球面汉克尔函数。方程假设正交规范化的 SH。

[0072] 注释:通常,第一类的球面汉克尔函数 $h_n^{(1)}$ 被用来描述正频率的输出波(与 e^{ikr} 相关),而第二类的球面汉克尔函数 $h_n^{(2)}$ 被用来描述输入波(与 e^{-ikr} 相关),参见上述的“Fourier Acoustics”书。

[0073] 球面谐波

[0074] 球面谐波 Y_n^m 可以是复数值或实数值。HOA 的通常情况使用实数值的球面谐波。使用实数和复数球面谐波的高保真度立体声响复制的统一描述可以在 Mark Poletti, “Unified description of Ambisonics using real and complex spherical harmonics”, Proceedings of the Ambisonics Symposium 2009, Gras, 奥地利, 2009年6月中看到。

[0075] 存在将球面谐波规范化的不同方法(其与球面谐波是实数还是复数无关),参见以下关于(实数)球面谐波和规范化方案的网页: <http://www.ipgp.fr/~wiecsor/SHTTOOLS/www/conventions.html>, http://en.citiseindium.org/wiki/Spherical_harmonics。

[0076] 规范化对应于 Y_n^m 和 $Y_{n'}^{m'}^*$ 之间的正交关系。

[0077] 注释:

$$[0078] \quad \int_{S^2} Y_n^m(\Omega) Y_{n'}^{m'}(\Omega)^* d\Omega = \frac{N_{n,m}}{\sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}}} \frac{N_{n',m'}}{\sqrt{\frac{(2n'+1)(n'-|m'|)!}{4\pi(n'+|m'|)!}}} \delta_{nn'} \delta_{mm'}$$

[0079] 其中, S^2 是单位球面, 对于 $a=a'$, Kroneker delta 等于 1, 否则等于 0。

[0080] 通过下式来描述复数球面谐波:

$$[0081] \quad Y_n^m(\Theta, \phi) = s_m \Theta_n^m(\theta) e^{im\phi} = s_m N_{n,m} P_{n,|m|}(\cos(\theta)) e^{im\phi} \quad (3)$$

[0082] 其中, 像在上述“Fourier Acoustics”书中, 对于正 m , 对于交替符号,

$$[0083] \quad i = \sqrt{-1} \text{ 且 } s_m = \begin{cases} (-1)^m & m > 0 \\ 1 & \text{else} \end{cases}.$$

[0084] (注释: S_m 是传统项, 并且可以仅对于正 SH 省略)。 $N_{n,m}$ 是规范化项, 其采取正交规范化表达的形式 (! 表示阶乘)。

$$[0085] \quad N_{n,m} = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi (n+|m|)!}} \quad (4)$$

[0086] 下表 1 示出了用于复数值球面谐波的一些常用规范化方案。 $P_{n,|m|}(x)$ 是相关联的拉格朗日函数, 其中, 在符号后接着 $|m|$, 依据上面的文章“Unified description of Ambisonics using real and complex spherical harmonics”, 其避免了被称为康登肖特利相位的相位项 $(-1)^m$, 并且其有时被包括在其它符号内的 P_n^m 的表示中。可以使用方程将相关联的拉格朗日函数

[0087] $P_{n,|m|}: [-1,1] \rightarrow \mathbb{R}, n \geq |m| \geq 0$ 表达为:

$$[0088] \quad P_{n,|m|}(x) = \frac{1}{2^n n!} (1-x^2)^{\frac{|m|}{2}} \frac{d^{n+|m|}}{dx^{n+|m|}} (x^2-1)^n \quad (5)$$

[0089]

$N_{n,m}$ 复数SH的常用规范化方案			
未规范化	施密特半规范化 SN3D	4π 规范化 N3D, 测量 4π	正交规范化
1	$\sqrt{\frac{(n- m)!}{(n+ m)!}}$	$\sqrt{\frac{(2n+1)(n- m)!}{(n+ m)!}}$	$\sqrt{\frac{(2n+1)(n- m)!}{4\pi (n+ m)!}}$

[0090] 表 1 复数值的球面谐波的规范化因子

[0091] 在数值上, 有利的是从递归关系以渐进方式推导 $P_{n,|m|}(X)$, 参见 William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, “Numerical Recipes in C”, Cambridge University Press, 1992。在表 2 中给出了相关联的上至 $n=4$ 的拉格朗日函数:

[0092]

n m	0	1	2	3	4
0	$P_0^0(\cos\theta) = 1$	$P_1^0(\cos\theta) = \cos\theta$	$P_2^0(\cos\theta) = \frac{1}{2}(3\cos^2\theta - 1)$	$P_3^0(\cos\theta) = \frac{1}{2}(5\cos^3\theta - 3\cos\theta)$	$P_4^0(\cos\theta) = \frac{1}{8}(35\cos^4\theta - 30\cos^2\theta + 3)$
1		$P_1^1(\cos\theta) = \sin\theta$	$P_2^1(\cos\theta) = 3\cos\theta\sin\theta$	$P_3^1(\cos\theta) = \frac{3}{2}(5\cos^2\theta - 1)\sin\theta$	$P_4^1(\cos\theta) = \frac{5}{2}(7\cos^2\theta - 3\cos\theta)\sin\theta$
2			$P_2^2(\cos\theta) = 3\sin^2\theta$	$P_3^2(\cos\theta) = 15\cos\theta\sin^2\theta$	$P_4^2(\cos\theta) = \frac{15}{2}(7\cos^2\theta - 1)\sin^2\theta$
3				$P_3^3(\cos\theta) = 15\sin^3\theta$	$P_4^3(\cos\theta) = 105\cos\theta\sin^3\theta$
4					$P_4^4(\cos\theta) = 105\sin^4\theta$

[0093] 表 2 — 第一一些拉格朗日多项式

[0094] $P_{n,|m|}(\cos\theta)$, $n=0 \dots 4$

[0095] 通过组合与 m 的相反值对应的复数共轭 (conjugate) Y_n^m 来推导实数值 SH (引入定义 (6) 中的项 $(-1)^m$ 来获得实数 SH 的无符号表达, 其是高保真度立体声响复制中的常见情况):

[0096]

$$S_n^m(\theta, \phi) = \begin{cases} \frac{(-1)^m}{\sqrt{2}} (Y_n^m + Y_n^{m*}) & = \Theta_n^m(\theta) \sqrt{2} \cos(m\phi), & m > 0 \\ Y_n^0 & = \Theta_n^0(\theta), & m = 0 \\ \frac{(-1)^m}{i\sqrt{2}} (Y_n^{|m|} - Y_n^{|m|*}) & = \Theta_n^{|m|}(\theta) \sqrt{2} \sin(|m|\phi), & m < 0 \end{cases} \quad (6)$$

[0097] 令 $\Phi_m(\phi) = \Phi_{n=|m|}^m(\phi)$, 并且仅仅保持方位项, 方程 (6) 可以被重写为方程 (7), 以便突出与圆谐波的联系:

[0098]

$$S_n^m(\theta, \phi) = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) \Phi_m(\phi) \quad (7)$$

$$[0099] \quad \Phi_{n=|m|}^m(\phi) = \begin{cases} \cos(m\phi), & m > 0 \\ 1 & m = 0 \\ \sin(|m|\phi) & m < 0 \end{cases} \quad (8)$$

[0100] 用于给定高保真度立体声响复制阶 N 的球面分量 S_n^m 的总数量等于 $(N+1)^2$ 。在表 3 中给出了实数值球面谐波的常见规范化方案。

[0101]

$\tilde{N}_{n,m}$, 实数 SH 的常用规范化方案			
未规范化	施密特半规范化 , $\delta_{0,m}$	4π 规范化 NSD, 测量 4π	正交规范化
$\sqrt{2 - \delta_{0,m}}$	$\sqrt{(2 - \delta_{0,m}) \frac{(n - m)!}{(n + m)!}}$	$\sqrt{(2 - \delta_{0,m}) \frac{(2n + 1)(n - m)!}{(n + m)!}}$	$\sqrt{(2 - \delta_{0,m}) \frac{(2n + 1)(n - m)!}{4\pi (n + m)!}}$

[0102] 表 3 — 3D 实数 SH 规范化方案, 对于 $m=0$, $\delta_{0,m}$ 具有值 1, 否则具有值 0

[0103] 圆谐波

[0104] 对于二维表示,仅仅需要谐波子集。SH 程度可以仅仅取值 $m \in \{-n, n\}$ 。由于表示倾角 θ 的分量变为过时的并且球面谐波可以由方程(8)中给出的圆谐波替代,用于给定 N 的分量的总数目降到 $2N+1$ 。

[0105] 存在用于圆谐波的不同的规范化 N_m 方案,当将 3D 高保真度立体声响复制系数转换为 2D 系数时需要考虑这些方案。用于圆谐波的更一半的方程变为:

[0106]

$$\check{\Phi}_{n=|m|}^m(\phi) = N_m \Phi_m(\phi) = \begin{cases} N_m \cos(m\phi), & m > 0 \\ N_m & m = 0 \\ N_m \sin(|m|\phi) & m < 0 \end{cases} \quad (9)$$

[0107] 在表 4 中提供了一些用于圆谐波的常见规范化因子,其中通过水平项 $\Phi_m(\phi)$ 之前的因子引入规范化项:

[0108]

N_m , 圆谐波的常用规范化方案			
未规范化	3D 规范化	2D 规范化, N2D	正交规范化
$\sqrt{\frac{2 - \delta_{0,m}}{2}}$	1	$\sqrt{(2 - \delta_{0,m})}$	$\sqrt{(2 - \delta_{0,m})} \frac{1}{2\pi}$

[0109] 表 4 — 2D 圆谐波规范化方案,对于 $m=0$, $\delta_{0,m}$ 具有值 1, 否则具有值 0

[0110] 不同规范化之间的转换是直接了当的。通常,规范化对描述压力(参见方程(1)(2)) and 所有导出的考虑具有影响。规范化的类型也影响高保真度立体声响复制系数。还存在可以应用于缩放这些系数的权重,例如当使用 AMB- 格式存储文件时应用于高保真度立体声响复制系数的福尔斯马勒姆(FuMa)权重。

[0111] 关于 2D — 3D 变换,例如当利用用于 2D 扬声器设置的 2D 解码器解码高保真度立体声响复制表示(记录)时,CH (圆谐波)到 SH (球面谐波)变换也可以应用于高保真度立体声响复制系数,反之亦然。在下面的上至高保真度立体声响复制阶 4 的方案中描绘了用于 3D — 2D 变换的 S_n^m 和 $\check{\Phi}_{n=|m|}^m$ 之间的关系:

[0112]

				S_0^0				
			S_1^{-1}	S_1^0	S_1^1			
	S_2^{-2}	S_2^{-1}	S_2^0	S_2^1	S_2^2			
S_3^{-3}	S_3^{-2}	S_3^{-1}	S_3^0	S_3^1	S_3^2	S_3^3		
S_4^{-4}	S_4^{-3}	S_4^{-2}	S_4^{-1}	S_4^0	S_4^1	S_4^2	S_4^3	S_4^4

[0113] 可以如下地针对 $\theta = \frac{\pi}{2}$ 的水平面导出 2D 到 3D 变换因子:

[0114]

$$\alpha_{2D}^{3D} = \frac{S_{n=|m|}^m(\theta=\pi/2, \phi)}{\check{\Phi}_{n=|m|}^m(\phi)} = \frac{\tilde{N}_{m,m} (2m)!}{N_m m! 2^m} \quad (10)$$

[0115] 从 3D 到 2D 的变换使用 $1/\alpha_{\frac{2D}{3D}}$ 。结合下面的方程(28)(29)(30)给出细节。

[0116] 2D 规范化到正交规范化的变换变为：

[0117]

$$\alpha_{\frac{N2D}{ortho3D}} = \sqrt{\frac{(2m+1)!}{4\pi m!^2 2^{2m}}} \quad (11)$$

[0118] 高保真度立体声响复制系数

[0119] 高保真度立体声响复制系数具有声压的单位缩放：

$$[0120] \quad 1Pa = 1 \frac{N}{m^2} = 1 \frac{kg \cdot m}{s^2 \cdot m^2}。$$

[0121] 高保真度立体声响复制系数形成高保真度立体声响复制信号，并且通常是离散时间的函数。表 5 示出了维度表示、高保真度立体声响复制阶 N 以及高保真度立体声响复制系数(声道)的数量之间的关系：

[0122]

维数	高保真度立体声响复制系数的数量 高保真度立体声响复制声道的数量			
	N=1	N=2	N=3	N
2D	3	5	7	2 N+1
3D	4	9	16	(N+1) ²

[0123] 表 5 — 高保真度立体声响复制系数的数量

[0124] 当处理离散时间表示时，通常以交织方式存储高保真度立体声响复制系数，像用于多声道记录的 PCM 声道表示(声道=样本 v 的高保真度立体声响复制系数 A_n^m)，系数序列作为变换的问题。对于 3D、N=2 的示例为：

[0125]

$$A_0^0(v) \ A_1^{-1}(v) \ A_1^0(v) \ A_1^1(v) \ A_2^{-2}(v) \ A_2^{-1}(v) \ A_2^0(v) \ A_2^1(v) \ A_2^2(v) \ A_0^0(v+1) \ \dots \quad (12)$$

[0126] 对于 2D、N=2 的示例为：

[0127]

$$A_0^0(v) \ A_1^{-1}(v) \ A_1^1(v) \ A_2^{-2}(v) \ A_2^2(v) \ A_0^0(v+1) \ A_1^{-1}(v+1) \ \dots \quad (13)$$

[0128] $A_0^0(n)$ 信号可以被看作高保真度立体声响复制记录的单声道表示，没有方向性信息但是作为记录的总音色印象(timbre impression)的表示。

[0129] 高保真度立体声响复制系数的规范化通常是根据 SH 的规范化(如下面将变得明显的，参见方程(15))来执行的，这在解码外部记录时必须考虑(A_n^m 基于具有规范化因子 $N_{n,m}$ 的 SH， $\tilde{A}_n^m$ 基于具有规范化因子 $\tilde{N}_{n,m}$ 的 SH)：

[0130]

$$A_n^m = \frac{N_{n,m}}{\tilde{N}_{n,m}} \tilde{A}_n^m \quad (14)$$

[0131] 对于 SN3D 到 N3D 情况,其变为 $A_{N3D_n}^m = \sqrt{(2n+1)} \tilde{A}_{SN3D_n}^m$ 。

[0132] B 格式和 AMB 格式使用应用于系数的另外权重 (Gerson、福尔斯马勒姆 (FuMa)、MaxN 权重)。参考规范化然后通常是 SN3D, 参见 Jérôme Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", 博士论文 (PhD thesis), Université Paris 6, 2001, and Dave Malham, "3-D acoustic space and its simulation using ambisonics", http://www.dxarts.washington.edu/courses/567/current/malham_3d.pdf。

[0133] 下面的用于理想平面波或球面波的波方程的两种特定规范化示出了关于高保真度立体声响复制系数的更多细节:

[0134] 平面波

[0135] 求解平面波 A_n^m 的波方程变得独立于 k 和 r_s ; θ_s, ϕ_s 描述了源角度, r_s 表示共轭复数:

[0136]

$$A_{n \text{ plane}}^m(\theta_s, \phi_s) = 4\pi i^n P_{S_0} Y_n^m(\theta_s, \phi_s)^* = 4\pi i^n d_n^m(\theta_s, \phi_s) \quad (15)$$

[0137] 这里, P_{S_0} 用于描述在描述坐标系的原点处测量的源的缩放信号压力,其可以是时间的函数,并且对于正交规范化的球面谐波变为 $A_{0 \text{ plane}}^0 / \sqrt{4\pi}$ 。通常,高保真度立体声响复制假设平面波,并且传送或存储高保真度立体声响复制系数

$$d_n^m(\theta_s, \phi_s) = \frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n} = P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (16)$$

该假设提供了叠加不同方向性信号的能力以及简单的解码器设计。这对于以一阶 B 格式 (N=1) 记录的声音场™ (Soundfield™) 麦克风的信号也成立,当与均衡滤波器的相位演进相比时这变得明显(对于理论上的演进,参见上述文章 "Unified description of Ambisonics using real and complex spherical harmonics", chapter 2.1, 而对于专利保护的演进,参见 US4042779)。方程(1)变为:

[0138]

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) Y_n^m(\theta, \phi) 4\pi i^n P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (17)$$

[0139] 可以通过后处理麦克风阵列信号导出系数 d_n^m 或者可以使用单声道信号 $P_{S_0}(t)$ 合成地创建系数 d_n^m , 在该情况下方向性的球面谐波 $Y_n^m(\theta_s, \phi_s, t)^*$ 可以同样是依赖于时间的(移动源)。方程(17)对于每个时间采样实例 v 都有效。对于所选择的高保真度立体声响复制阶 N , 可以以矢量/矩阵形式(对于每个采样实例 v) 重写合成编码的处理。

[0140]

$$\mathbf{d} = \Psi P_{S_0} \quad (18)$$

[0141] 其中, $\mathbf{d}$ 是高保真度立体声响复制信号, 保持 $d_n^m(\theta_s, \phi_s)$ (对于 $N=2$ 的示例, $\mathbf{d}(t) = [d_0^0, d_1^{-1}, d_1^0, d_1^1, d_2^{-2}, d_2^{-1}, d_2^0, d_2^1, d_2^2]^T$), $\text{size}(\mathbf{d}) = (N+1)^2 \times 1 = 9 \times 1$, P_{S_0} 是参考原点处的源信号压力, Ψ 是编码矢量, 保持 $Y_n^m(\theta_s, \phi_s)^*$, $\text{size}(\Psi) = 9 \times 1$ 。可以从用于特定源方向(等于平面波的方向)的球面谐波导出编码矢量 $0_s, \phi_s$ 。

[0142] 球面波

[0143] 对于 $r < r_s$, 描述由点源(近场源)生成的进入球面波的高保真度立体声响复制系数高保真度立体声响复制系数为:

[0144]

$$A_{n \text{ spherical}}^m(k, \theta_s, \phi_s, r_s) = 4\pi \frac{h_n^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (19)$$

[0145] 该方程是与下面的方程(31)到(36)结合地导出的。 $P_{S_0} = p(0|r_s)$ 描述了原点处的声压, 并且再次变得与 $A_0^0/\sqrt{4\pi}$ 一致, $h_n^{(2)}$ 是第二类的阶 n 的球面汉克尔函数, $h_0^{(2)}$ 是第二类的零阶球面汉克尔函数。

[0146] 方程(19)与 Jérôme Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format", AES23rd International Conference, Denmark, May2003(AES第23次国际会议, 丹麦, 2003年5月)中的教导相似。这里,

$\frac{h_n(kr_s)}{h_0(kr_s)} = i^n \sum_{a=0}^n \frac{(n+a)!}{(n-a)!a!} \left(-\frac{ic}{2r_s\omega}\right)^a$, 顺便提及, 考虑

方程(11), $\frac{h_1(kr_s)}{h_0(kr_s)} = i(1 - \frac{ic}{r_s\omega})$ 这可以在 M. A. Gerson, "General metatheory of auditory localisation", 92th AES Convention, 1992, Preprint3306 中看到, 其中 Gerson 描述了第一度信号(first-degree signal)的近似效果。

[0147] 对于更高的高保真度立体声响复制阶 N 而言, 球面高保真度立体声响复制信号的合成创建更不常见, 这是因为 $\frac{h_n(kr_s)}{h_0(kr_s)}$ 的频率响应对于低频而言难以在数值上处理。通过考虑如下所述的用于解码/再现的球面模型, 可以克服这些数值问题。

[0148] 声音场再现

[0149] 平面波解码

[0150] 通常, 高保真度立体声响复制假设通过在圆上或球面上均匀分布的 L 扬声器再现声音场。当假设扬声器被放置在距收听者位置足够远时, 在中心($r_s > \lambda$)处平面波解码模型有效。 L 扬声器生成的声压由以下描述:

[0151]

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) Y_n^m(\theta, \phi) 4\pi i^n \sum_{l=1}^L w_l Y_n^m(\theta_l, \phi_l)^* \quad (20)$$

[0152] 其中, w_l 是用于扬声器 l 的信号并且具有单位规模的声压 1Pa。 w_l 通常被称为扬

声器 1 的驱动函数。

[0153] 期望方程 (20) 的声压与方程 (17) 描述的压力一致。这导致：

[0154]

$$\sum_{l=1}^L w_l Y_n^m(\theta_l, \phi_l)^* = d_n^m(\theta_s, \phi_s) = \frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n} \quad (21)$$

[0155] 这可以以矩阵形式重写, 已知为“重编码方程”(与方程 (18) 比较)：

[0156]

$$\mathbf{d} = \Psi \mathbf{y} \quad (22)$$

[0157] 其中 d 是高保真度立体声响复制信号, 保持 $d_n^m(\theta_s, \phi_s)$ 或者 $\frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n}$,

[0158] (对于 N=2 的示例, $\mathbf{d}(t) = [d_0^0, d_1^{-1}, d_1^0, d_1^1, d_2^{-2}, d_2^{-1}, d_2^0, d_2^1, d_2^2]^T$),
size(d)=(N+1)²x1=0x1, Ψ 是重编码矩阵, 保持 $Y_n^m(\theta_s, \phi_s)^*$ size(Ψ)=0x1, y 是扬声器信号 w_l , size(y(n), 1)=L。

[0159] 可以使用若干已知方法(例如, 模式匹配)来导出 y, 或者通过使特定扬声器声像调节函数最优化的方法来导出 y。

[0160] 球面波模型的解码

[0161] 更一般的解码模型再次假设围绕原点均匀地分布扬声器, 像球面波那样以距离 r_l 辐射点。依据方程 (1) 通过一般描述来给出高保真度立体声响复制系数 A_n^m , 并且根据方程 (19) 给出 L 扬声器生成的声压：

[0162]

$$A_n^m = \sum_{l=1}^L 4\pi \frac{h_n(kr_l)}{h_0(kr_l)} w_l Y_n^m(\theta_l, \phi_l)^* \quad (23)$$

[0163] 更复杂的解码器可以对高保真度立体声响复制系数 A_n^m 进行滤波, 以便提取

$C_n^m = A_n^m \frac{h_0(kr_l)}{4\pi h_n(kr_l)}$ 并且此后将 $\mathbf{d} = [C_0^0, C_1^{-1}, C_1^0, C_1^1, C_2^{-2}, C_2^{-1}, C_2^0, C_2^1, C_2^2, \dots]^T$ 应用于方程 (17) 以便导出扬声器权重。利用该模型, 通过原点处的压力确定扬声器信号 w_l 。

[0164] 存在替代方法, 其使用首先在上述文章 “Three-dimensional surround sound systems based on spherical harmonics” 中描述的简单源方法。假设扬声器均匀分布在球面上并且具有二次源特性。在 Jens Ahrens, Sascha Spors, “Analytical driving functions for higher order ambisonics”, Proceedings of the ICASSP, pages 373 - 376, 2008 (ICASSP 会议论文集, 第 373 - 376 页, 2008 年), 方程 (13) 中导出了解, 其可以针对用于高保真度立体声响复制阶 N 处的截断和扬声器增益 g_l 被重写以作为概括：

[0165]

$$w_l = \sum_{n=0}^N \sum_{m=-n}^n g_l \frac{A_n^m}{kr_l h_n^{(2)}(kr_l)} Y_n^m(\theta_l, \phi_l) \quad (24)$$

[0166] 距离编码的高保真度立体声响复制信号

[0167] 当(使用方程(18))建模或记录球面波时,在高保真度立体声响复制编码器处使用参考扬声器距离 $r_{1\text{ref}}$ 创建 C_n^m 可以求解的数值解:

[0168]

$$C_n^m = A_n^m \frac{h_0(kr_{l\text{ref}})}{4\pi h_n(kr_{l\text{ref}})} = \frac{h_0(kr_{l\text{ref}})}{h_n(kr_{l\text{ref}})} \frac{h_n(kr_s)}{h_0(kr_s)} P_{s_0} Y_n^m(\theta_s, \phi_s)^* \quad (25)$$

[0169] 传送或存储 C_n^m 参考距离 $r_{1\text{ref}}$ 以及用于使用球面距离编码系数的指示符。在解码器侧,只要实际扬声器距离 $r_1 \approx r_{1\text{ref}}$, 方程(22)给出的简单解码处理就是可行的。如果差异太大,则要求在高保真度立体声响复制解码之前通过滤波进行校正

$$D_n^m = C_n^m \frac{h_n(kr_{l\text{ref}})}{h_n(kr_l)} \quad (26)^\circ$$

[0170] 像方程(24)那样的其它解码模式导致用于距离编码高保真度立体声响复制的不同公式:

[0171]

$$\tilde{C}_n^m = \frac{A_n^m}{kr_{l\text{ref}} h_n(kr_{l\text{ref}})} = \frac{1}{kr_{l\text{ref}} h_n(kr_{l\text{ref}})} \frac{h_n(kr_s)}{h_0(kr_s)} P_{s_0} Y_n^m(\theta_s, \phi_s)^* \quad (27)$$

[0172] 而且,球面谐波的规范化可能对距离编码高保真度立体声响复制的公式有影响,即距离编码高保真度立体声响复制系数需要定义的上下文。

[0173] 上述 2D - 3D 转换的细节如下:

[0174] 用于通过乘法将 2D 圆分量转换为 3D 球形分量的转换系数 $\alpha_{2D/3D}$ 可以如下地导出:

[0175]

$$\alpha_{2D/3D} = \frac{S_{n=m}^m(\theta=\pi/2, \Phi)}{\Phi_{n=|m|}^m(\phi)} = \frac{N_{m,m} P_{|m|,|m|}(\cos(\theta=\pi/2)) \Phi_m(\phi)}{N_m \Phi_m(\phi)} \quad (28)$$

[0176] 使用常用等式(参见 2010 年 10 月 12 日维基百科的 "Associated Legendre polynomials",

[0177] http://en.wikipedia.org/w/index.php?title=Associated_Legendre_polynomials&oldid=363001511), $P_{l,l}(x) = (2l-1)!!(1-x^2)^{l/2}$, 其中 $(2l-1)!! = \prod_{i=1}^l (2i-1)$ 是二阶因子, $P_{|m|,|m|}$ 可以被表达为:

[0178]

$$P_{|m|,|m|}(\cos(\theta = \pi/2)) = (2m-1)!! = \frac{(2m)!}{m! 2^m} \quad (29)$$

[0179] 将方程(29)插入到方程(28)中,导致方程(30)。

[0180] 使用关系 $!! = \frac{(i+1)!}{i+1}$ 并且代入 $l=2m$, 从 2D 到正交-3D 的转换被导出为:

[0181]

$$\alpha_{N2D/ortho3D} = \sqrt{\frac{(2m+1)}{4\pi (2m)!}} \frac{(2m)!}{m! 2^m} = \sqrt{\frac{(2m+1)(2m)!}{4\pi m!^2 2^{2m}}} = \sqrt{\frac{(2m+1)!}{4\pi m!^2 2^{2m}}}, \quad (30)$$

[0182] 上述球面波扩展的细节如下:

[0183] 求解用于球面波的方程(1) (其是由 $r < r_s$ 的点源和进入波生成的) 更复杂, 这是因为需要使用体积流量 Q_s 来描述具有消失的无穷小的大小的点源, 其中 r 处场点和 r_s 处场点的辐射压力由以下给出 (参见上述书目 "Fourier Acoustics"):

[0184]

$$p(r|r_s) = -i \rho_0 c k Q_s G(r|r_s) \quad (31)$$

[0185] 其中, ρ_0 是具体密度, $G(r|r_s)$ 是格林函数

$$G(r|r_s) = \frac{e^{-ik|r-r_s|}}{4\pi|r-r_s|} \quad (32)$$

[0186] 对于 $r < r_s$, 在球面谐波中还可以通过下式来表达 $G(r|r_s)$:

[0187]

$$G(r|r_s) = i k \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) h_n^{(2)}(kr_s) Y_n^m(\theta, \phi) Y_n^m(\Theta_s, \Phi_s)^* \quad (33)$$

[0188] 其中, $h_n^{(2)}$ 是第二类的汉克尔函数。注意, 格林函数具有米⁻¹的单位大小 (由于 k , $\frac{1}{m}$)。可以将方程(31) (33) 与方程(1) 相比, 以导出球面波的高保真度立体声响复制系数:

[0189]

$$A_n^{m \text{ spherical}}(k, \Theta_s, \Phi_s, r_s) = \rho_0 c k^2 Q_s h_n^{(2)}(kr_s) Y_n^m(\Theta_s, \Phi_s)^* \quad (34)$$

[0190] 其中, Q_s 是单位为 $m^3 s^{-1}$ 的体积流量, ρ_0 是单位为 $kg m^{-3}$ 的具体密度。

[0191] 为了能够合成地创建高保真度立体声响复制信号并且能够与上面的平面波考虑有关, 使用在坐标系的原点处生成的声压来表达方程(34) 是合理的:

[0192]

$$P_{S_0} = p(0|r_s) = \frac{-i \rho_0 c k Q_s}{4\pi} \frac{e^{-ikr_s}}{r_s} = \frac{\rho_0 c k^2 Q_s}{4\pi} h_0^{(2)}(kr_s) \quad (35)$$

[0193] 这导致:

[0194]

$$A_n^{m \text{ spherical}}(k, \Theta_s, \Phi_s, r_s) = 4\pi \frac{h_n^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} P_{S_0} Y_n^m(\Theta_s, \Phi_s)^* \quad (36)$$

[0195] 交换存储格式

[0196] 根据本发明的存储格式允许将多于一个 HOA 表示和附加的方向性流一起存储在一个数据容器中。这使能 HOA 描述的不同格式, 其使得解码器能够将再现最优化, 并且对于大小 >4GB 提供了高效的数据存储。其它优点为:

[0197] A) 通过将使用不同格式的若干 HOA 描述与相关的存储格式信息一起存储, 高保真度立体声响复制解码器能够混合并解码两个表示。

[0198] B) 将下一代 HOA 解码器所需的信息项存储为格式信息:

[0199] 一维度, 感兴趣区域 (收听区域外部的源或收听区域内的源), 球面基本函数的规范化;

[0200] 一高保真度立体声响复制系数打包和缩放信息;

- [0201] 一高保真度立体声响复制波类型(平面,球面),参考半径(用于解码球面波);
- [0202] 一可以存储相关的方向性单声道信号。可以使用角度和距离信息、或者高保真度立体声响复制系数的编码矢量来描述这些方向性信号的位置信息。
- [0203] C) 扩展高保真度立体声响复制数据的存储格式以便允许灵活并经济地存储数据:
- [0204] 一利用不同 PCM- 字大小分辨率来存储与高保真度立体声响复制分量(高保真度立体声响复制声道)相关的高保真度立体声响复制数据;
- [0205] 一使用重采样或 MDCT 处理来利用降低的带宽存储高保真度立体声响复制数据。
- [0206] D) 元数据字段可用于将用于特定解码(前面、环境)的音轨相关联,并且允许存储关于文件的附属信息(诸如用于麦克风信号的记录信息):
- [0207] 一记录参考坐标系、麦克风、源和虚拟听众位置、麦克风方向性特性、房间和源信息。
- [0208] E) 该格式适用于存储包含不同音轨的多帧,允许在没有场景描述的情况下进行音频场景改变。(注释:一个音轨包含 HOA 声音场描述、或者带有位置信息的单源。一帧是一个或多个并列音轨的组合)。音轨可以在帧的开头处开始或者在帧的结束处结束,因此不需要时间码。
- [0209] F) 该格式便于快速访问音频音轨数据(快进或调到线索点)并且确定相对于文件数据的开头处的时间的的时间码。
- [0210] HOA 数据交换的 HOA 参数
- [0211] 表 6 总结了需要为 HOA 信号数据的非歧义交换(non-ambiguous exchange)定义的参数。球面谐波的定义固定于复数值和实数值情况,参见方程(3)(6)。
- [0212]

上下文	维度	2D/3D, 也影响高保真度立体声响复制系数(AC)的包装
	感兴趣区域	图 6, 图 8, 方程(1)(2)
	SH 类型	复数,实数值, 用于 2D 的圆
	SH 规范化	SN3D, N3D, 正交规范化
高保真度立体声响复制系数	AC 加权	B-格式, FuMa, maxN, 无加权,用户定义
	AC 序列和采样 分辨率	方程 (12)(13) 中的示例, 分辨率: 16/24 比特或浮点类型
	AC 类型	未规定 A_n^m , 平面波类型 d_n^m , 方程 (16), 距离编码类型 D_n^m 或 $\check{C}_n^m$, 方程 (26)(27)

[0213] 表 6 一用于 HOA 记录的非歧义交换的参数

[0214] 文件格式细节

[0215] 在下文中,具体描述用于存储由高阶高保真度立体声响复制(HOA)或者具有位置信息的单源构成的音频场景的文件格式。音频场景可以包含多个可能使用不同的规范化方案的 HOA 序列。因此,解码器可以将用于期望扬声器设置的对应扬声器信号计算为来自当前文件的所有音频音轨的叠加。该文件包含解码音频内容所需的所有数据。根据本发明的文件格式提供了在单个文件中存储多于一个 HOA 或单源信号的特征。该文件格式使用帧的组成,每个帧可以包含若干音轨,其中在一个或多个被称为音轨分组(TrackPackets)的分组中存储一个音轨的数据

[0216] 按照 little-endian 字节顺序存储所有整数类型,使得最低有效字节最先到达。比特顺序总是最高有效比特在前。整数数据类型的表示为'int'。前面的'u'指示无符号整数。比特的分辨率被写在定义的尾部。例如,无符号 16 比特整数字段被定义为'uint16'。整数格式的 PCM 样本和 HOA 系数被表示为定点数,小数点在最高有效比特处。

[0217] 所有浮点数据类型符合 IEEE 规范 IEEE-754, "Standard for binary floating-point arithmetic", <http://grouper.ieee.org/groups/754/>。浮点数据类型的表示为'float'。比特的分辨率被写在定义的尾部。例如,32 比特浮点字段被定义为'float32'。

[0218] 恒定标识符 ID (其表示帧、音轨或块、以及字符串的开头)被定义为数据类型字节。字节阵列的字节顺序是最高有效字节和比特在前。因此,在 32 比特字节字段中定义 ID' TRCK',其中,按照物理顺序'T','R','C'和'K'(<0x54;0x52;0x42;0x4b>)编写字节。

[0219] 十六进制值以'0x'开始(例如 0xAB64C5)。单比特被置入引号(例如'1'),多个二进制值'0b'以开始(例如 0b0011=0x3)。

[0220] 头部字段名称总是以头部名称开始,后面接着字段名称,其中,每个词的第一字母是大写字母(例如,TrackHeaderSize)。通过仅使用字段或头部名称的大写字母来创建字段或头部名称的缩写(例如 TrackHeaderSize=THS)。

[0221] HOA 文件格式可以包括多于一帧、分组或音轨。对于多个头部字段的区分,可以在字段或头部名称之后接有数字。例如,第三音轨的第二 TrackPacket (音轨分组)被命名为'Track3Packet2'。

[0222] HOA 文件格式可以包括复数值字段。这些复数值被存储为实部和虚部,其中实部被写在前面。'int8'格式的复数值 $1+i2$ 将被存储为'0x01'后面接着'0x02'。因此,与对应的实数值格式类型相比,复数值格式类型的字段或系数要求两倍存储大小。

[0223] 高阶高保真度立体声响复制文件格式结构

[0224] 单音轨格式

[0225] 高阶高保真度立体声响复制文件格式包括至少一个文件头部(FileHeader)、一个帧头部(FrameHeader)、一个音轨头部(TrackHeader)和一个音轨分组(TrackPacket),如图 9 中所描绘的,其示出了在一个或多个分组(Packets)中携带一个音轨(Track)的简单示例的 HOA 文件格式文件。

[0226] 因此,HOA 文件的基本结构是一个文件头部后接着帧,该帧包括至少一个音轨。音轨总是包括音轨头部以及一个或多个音轨分组。

[0227] 多帧和音轨格式

[0228] 与文件头部相反,HOA 文件可以包含多于一帧,其中帧可以包含多于一个音轨。如果超出帧的最大大小则使用新的帧头部,或者从一帧向另一帧添加音轨或移动音轨。在图 10 中示出多音轨和帧 HOA 文件的结构。

[0229] 对音轨帧的结构以帧头部(FrameHeader)开始,后面接着该帧(Frame)的所有音轨头部(TrackHeader)。因此,每个音轨的音轨分组被连续地发送给帧头部,其中音轨分组以与音轨头部相同的方式被交织。

[0230] 在多音轨帧中,在帧头部中定义样本中分组的长度,并且样本中分组的长度对于所有音轨都是恒定的。此外,每个音轨的样本是同步的,例如,音轨 1 分组 1 (Track1Packet1) 的样本与音轨 2 分组 1 (Track2Packet1) 的样本同步。特定的音轨编码类型(TrackCodingType)可能在解码器侧引起延迟,并且这样的特定延迟需要在解码器侧是已知的,或者要被包括在音轨头部(TrackHeader)的依赖音轨编码类型(TrackCodingType)的部分中,这是因为解码器将所有的音轨分组(TrackPacket)同步到帧的所有音轨的最大延迟。

[0231] 依赖文件的元数据

[0232] 指代完整 HOA 文件的元数据可以优选地被添加在文件头部(FileHeader)后的元数据块(MetaDataChunk)中。MetaDataChunk 以特定的一般用户 ID (GUID)开始,后面接着元数据块大小(MetaDataChunkSize)。MetaDataChunk 的实质内容(例如元数据信息)被打包到 XML 格式或任何用户定义格式。图 11 示出了使用若干 MetaDataChunk 的 HOA 文件格式的结构。

[0233] 音轨类型

[0234] HOA 格式的音轨(Track)在一般 HOA 音轨(HOATrack)和单源音轨(SingleSourceTrack)之间进行区分。HOATrack 包括被编码为 HOA 系数(HOACoefficient)的完整声音场。因此,在解码器不需要场景描述(例如编码源的位置)来解码系数。换句话说,在 HOACoefficient 内存储音频场景。

[0235] 与 HOATrack 相反,SingleSourceTrack 仅包括被编码为 PCM 样本的一个源以及该源在音频场景内的位置。SingleSourceTrack 的位置随着时间的推移可以固定或可变。源位置被发送为音轨 HOA 编码矢量(TrackHOAEncodingVector)或音轨位置矢量(TrackPositionVector)。TrackHOAEncodingVector 包含用于获得用于每个样本的 HOACoefficient 的 HOA 编码值。TrackPositionVector 将源的位置包含为相对于中央收听位置的角度和距离。

[0236] 文件头部(File Header)

[0237]

字段名称	大小 / 比特	数据类型	描述
FileID	32	字节	HOA文件格式的恒定文件标识符: <“H”; “O”; “A”; “F”> or <0x48; 0x4F; 0x41; 0x46>
FileVersionNumber	8	uint8	HOA格式的版本号0-255
FileSampleRate	32	uint32	对于所有帧和音轨恒定的Hs的采样率
FileNumberOfFrames	32	uint32	总帧数, 要求至少为 ‘1’
保留	8	字节	
总比特数	112		

[0238] FileHeader 包括完整 HOA 文件的所有恒定信息。FileID 用于标识 HOA 文件格式。采样率对所有音轨恒定, 即使它是在 FrameHeader 中发送的。其采样率从一帧到另一帧改变的 HOA 文件无效。在 FileHeader 中指示帧数, 以便向解码器指示帧结构。

[0239] 元数据块

[0240]

字段名称	大小 / 比特	数据类型	描述
ChunkID	32	字节	一般用户 ID (尚未定义)
ChunkSize	32	uint32	块中除了 ChunkID 和 ChunkSize 字段之外的以字节为单位的大小
ChunkData	8 * ChunkSize	字节	用户定义字段或取决于 ChunkID 的 XML-结构
总比特数	64 + 8 * ChunkSize		

[0241] 帧头部 (Frame Header)

[0242]

字段名称	大小 / 比特	数据类型	描述
FrameID	32	字节	所有帧头部的恒定标识符: <“F”; “R”; “A”; “M”> or <0x46; 0x52; 0x41; 0x4D>
FrameSize	32	uint32	帧中除了FrameID和FrameSize字段之外的以字节为单位的大小
FrameNumber	32	uint32	唯一帧编号, 其对于第一帧以0开始并且对于随后的帧增加。最后帧具有的该编号为FileNumberOffFrame-1。
FrameNumberOfSamples	32	uint32	在帧的每个音轨中存储的样本的数量
FrameNumberOfTracks	8	uint8	在帧中存储的音轨的数量
FramePacketSize	32	uint32	样本中分组的大小。分组大小对于所有音轨都是恒定的。
FrameSampleRate	32	uint32	对于所有帧和音轨恒定的Hs的采样率。必须与FileSampleRate一致 (对于不知道文件头部的流应用而言重新定义)
总比特数	200		

[0243] FrameHeader 保持帧的所有轨道的恒定信息, 并且指示 HOA 文件内的改变。FrameID 和 FrameSize 指示帧的开头和帧的长度。这两个字段允许容易访问每个帧和帧结构的交叉检查。如果帧长度要求多于 32 比特, 则一帧可以被划分为多帧。每帧具有唯一帧编号(FrameNumber)。FrameNumber 应以 0 开始, 并且应对于每个新帧增加 1。

[0244] 帧的样本的数量对于帧的所有轨道恒定。帧内轨道的数量对于帧恒定。发送新帧头部以在期望样本位置结束或开始音轨。

[0245] 在分组中存储每个音轨的样本。在样本中指示这些音轨分组(TrackPacket)的大小, 并且其对于所有音轨恒定。分组的数量等于存储帧的样本的数量所需的整数。因此, 音轨的最后分组可以包含比所指示的分组大小更少的样本。

[0246] 帧的采样率等于 FileSampleRate, 并且在 FrameHeader 中被指示以便允许在不知道 FileHeader 的情况下解码帧。这可以在从多帧文件的中间开始解码时在不知道的情况下(例如对于流应用)使用。

[0247] 音轨头部(Track Header)

[0248]

字段名称	大小 / 比特	数据类型	描述
TrackID	32	字节	所有音轨头部的恒定标识符: <“T”; “R”; “A”; “C”> or <0x54; 0x52; 0x41; 0x43>
TrackNumber	16	uint16	用于标识若干帧中相干音轨的唯一音轨编号
TrackHeaderSize	32	uint32	TrackHeader中除了TrackID和TrackNumber字段的大小 (到下一TrackHeader或第一TrackPacket的开头的偏移)
TrackMetaDataOffset	32	uint32	从该字段的结尾到TrackMetaData字段的开头的偏移。零 等于不包括TrackMetaData。
TrackSourceType	1	二进制	‘0’ = HOATrack 以及 ‘1’ = SingleSourceTrack
保留	7	二进制	0b0000000
条件: TrackSourceType == ‘0’			HOA Track的TrackHeader
<HOATrackHeader>	dyn	字节	参见HOA TrackHeader部分
条件: TrackSourceType == ‘1’			SingleSourceTrack的TrackHeader
<SingleSourceTrackHeader>	dyn	字节	参见单源固定位置TrackHeader和单源移动位置 TrackHeader
条件: TrackMetaDataOffset > 0			
TrackMetaData	dyn	字节	依赖音轨的元数据 (MetaData) 的XML字段, 参见 TrackMetaData表
总比特数	120 + dyn		

[0249] 项 ‘dyn’ 指代由于条件字段引起的动态字段大小。TrackHeader 保持用于特定音轨的分组的恒定信息。TrackHeader 被分为恒定部分和用于两个音轨源类型 (TrackSourceType) 的可变部分。TrackHeader 以恒定值 TrackID 开始, 以验证和标识 TrackHeader 的开头。向每个音轨分配唯一的音轨编号 (TrackNumber), 以在帧边界上指示相干轨道。因此, 具有相同 TrackNumber 的音轨可以在以下帧中出现。提供音轨头部大小 (TrackHeaderSize) 以跳到下一 TrackHeader, 并且其被指示为从 TrackHeaderSize 字段的结尾开始的偏移。音轨元数据偏移 (TrackMetaDataOffset) 提供样本的数量以便直接跳到 TrackMetaData 字段的开头, 其可以被用来跳过 TrackHeader 的可变长度部分。TrackMetaDataOffset 为零指示不存在 TrackMetaData 字段。依赖于音轨源类型 (TrackSourceType), 提供 HOA 音轨头部 (HOATrackHeader) 或单源音轨头部 (SingleSourceTrackHeader)。HOATrackHeader 提供用于描述完整声音场的标准 HOA 系数的辅助信息。SingleSourceTrackHeader 保持用于单声道 PCM 音轨的样本和源的位置的信息。对于 SingleSourceTrack 而言, 解码器无需将音轨包括在场景中。

[0250] 在 TrackHeader 结尾, 定义可选的 TrackMetaData 字段, 其使用 XML 格式来提供依

赖音轨的元数据,例如用于 A- 格式传输的附加信息(麦克风阵列信号)

[0251] HOA 音轨头部(HOA Track Header)

[0252]

字段名称	大小 / 比特	数据类型	描述
TrackComplexValueFlag	2	二进制	0b00: 仅实部 0b01: 实部和虚部 0b10: 仅虚部 0b11: 保留
TrackSampleFormat	4	二进制	0b0000 无符号整数 8比特 0b0001 有符号整数 8比特 0b0010 有符号整数 16比特 0b0011 有符号整数 24比特 0b0100 有符号整数 32比特 0b0101 有符号整数 64比特 0b0110 浮点32比特(二进制单精度) 0b0111 浮点64比特(二进制双精度) 0b1000 浮点128 比特(二进制四精度) 0b1001-0b1111 保留
保留	2	二进制	填充比特
TrackHOAParams	dyn	字节	参见 TrackHOAParams
TrackCodingType	8	uint8	,0* 利用恒定比特分辨率和恒定频率分辨率将HOA系数编码为PCM样本。 ,1* 利用依赖顺序的比特分辨率和频率分辨率将HOA系数编码 其它 为其它编码类型保留

[0253]

条件: TrackCodingType == '1'		用于编码类型1的辅助信息	
TrackBandwidthReductionType	8	uint8	0 用于所有顺序的全带宽 1 经由MDCT降低带宽 2 经由时域滤波器降低带宽

[0254]

TrackNumberOfOrderRegions	8	uint8	可以将带宽和比特分辨率适配于一定数量的区域, 其中每个数量具有开始和结束顺序。 TrackNumberOfOrderRegions指示所定义的区域的数量。
---------------------------	---	-------	--

[0255]

为每个区域写入以下字段			
TrackRegionFirstOrder	8	uint8	该区域的第一顺序
TrackRegionLastOrder	8	uint8	该区域的最后顺序
TrackRegionSampleFormat	4	二进制	0b0000 无符号整数 8比特 0b0001 有符号整数 8比特 0b0010 有符号整数 16比特 0b0011 有符号整数 24比特 0b0100 有符号整数 32比特 0b0101 有符号整数 64比特 0b0110 浮点32比特(二进制单精度) 0b0111 浮点64比特(二进制双精度) 0b1000 浮点128 比特(二进制四精度) 0b1001-0b1111保留
TrackRegionUseBandwidthReduction	1	二进制	'0' 全带宽用于该区域 '1' 利用TrackBandwidthReductionType降低用于该区域的带宽
保留	3	二进制	填充比特

[0256]

条件: TrackRegionUseBandwidthReduction == '1'			在该区域中降低带宽
条件: TrackBandwidthReductionType == 1			经由MDCT辅助信息降低带宽
TrackRegionWindowType	8	uint8	0: $\text{sine窗口: } W(t) = \sin\left(\frac{\pi(t+0.5)}{N}\right)$ 其它: 保留
TrackRegionFirstBin	16	uint16	第一次编码的MDCT 库(低截止频率)
TrackRegionLastBin	16	uint16	最后编码的 MDCT 库(高截止频率)
条件: TrackBandwidthReductionType == 2			经由时域滤波器辅助信息降低带宽
TrackRegionFilterLength	16	uint16	低通滤波器系数的数量

[0257]

<TrackRegionFilterCoefficients>	dyn	float32	<i>TrackRegionFilterLength</i> 低通滤波器系数
TrackRegionModulationFreq	32	float32	偏移信号谱所需的规范化的调制频率 Ω_{mod}/π
TrackRegionDownsampleFactor	16	uint16	下采样因子 M , 必须是 <i>FramePacketSize</i> 的除数
TrackRegionUpsampleFactor	16	uint16	上采样因子 $K < M$
TrackRegionFilterDelay	16	uint16	根据 (<i>FileSampleRate</i>) 的编码/解码带宽降低处理的样本中延迟

[0258] HOATrackHeader 是保持用于解码 HOATrack 的信息的 TrackHeader 的一部分。HOATrack 的 TrackPackets 传输编码轨道的整个声音场的 HOA 系数。基本上, HOATrackHeader 保持了在解码器侧解码用于给定扬声器设置的 HOA 系数所需的所有 HOA 参数。

[0259] 音轨复数值标志(TrackComplexValueFlag)和音轨样本格式(TrackSampleFormat)定义了每个 TrackPacket 的 HOA 系数的格式类型。对于编码或压缩系数, TrackSampleFormat 定义了解码或未压缩系数的格式。所有格式类型可以是实数或复数。

[0260] 在上面 File Format 细节部分中提供了关于复数值的更多信息。

[0261] 在音轨 HOA 参数(TrackHOAParams)中定义了所有依赖 HOA 的信息。在其它 TrackSourceTypes 中重新使用 TrackHOAParams。因此,在 TrackHOAParams 部分中定义并描述了 TrackHOAParams 的字段。

[0262] 音轨编码类型(TrackCodingType)字段指示 HOA 系数的编码(压缩)格式。HOA 文件格式的基本版本包括例如两种编码类型(CodingType)。

[0263] 一种 CodingType 是 PCM 编码类型(TrackCodingType== '0'),其中,以所选择的分组样本格式(TrackSampleFormat)将未压缩的实数或复数系数写入到分组中。在 TrackHOAParams 字段中定义了 HOA 系数的顺序和规范化。

[0264] 第二种 CodingType 允许改变样本格式,并且限制每种 HOA 顺序的系数的带宽。在音轨区域编码(TrackRegion Coding)部分中提供了该 CodingType 的具体描述,简短说明如下:

[0265] 音轨带宽降低类型(TrackBandwidthReductionType)确定已经被用来限制每种 HOA 顺序的带宽的处理的类型。如果不改变所有系数的带宽,则可以通过将 TrackBandwidthReductionType 字段设置为零来关掉带宽降低。定义了两种其它的带宽降低处理类型。格式包括频域 MDCT 处理以及可选地时域滤波器处理。关于更多关于 MDCT 处理的信息,参见经由 MDCT 降低带宽(Bandwidth reduction via MDCT)部分。

[0266] 可以将 HOA 顺序组合到具有相同样本格式和带宽的区域中。通过顺序区域音轨数量(TrackNumberOfOrderRegions)字段来指示区域的数量。对于每个区域,必须定义第一和最后顺序索引、样本格式和可选带宽降低信息。区域将获得至少一个顺序。使用在音轨样本格式(TrackSampleFormat)字段中指示的标准格式利用全带宽来编码未被任何区域覆

盖的顺序。特殊情况是没有使用区域(TrackNumberOfOrderRegions==0)。该情况可以用于PCM格式的去交织的HOA系数,其中HOA分量没有每个样本地被交织。区域的顺序的HOA系数被按照音轨区域样本格式(TrackRegionSampleFormat)编码。音轨区域使用带宽降低(TrackRegionUseBandwidthReduction)指示对于该区域的顺序的系数使用带宽降低处理。如果TrackRegionUseBandwidthReduction标志被设置,则带宽降低辅助信息将接在后面。对于MDCT处理,定义了窗口类型、以及第一和最后编码的MDCT库。这里,第一库等效于低截止频率,而最后库定义了高截止频率。也按照TrackRegionSampleFormat编码MDCT库,参见经由MDCT降低带宽(Bandwidth reduction via MDCT)部分。

[0267] 单源类型

[0268] 单源被划分为固定位置和移动位置源。在音轨移动源标志(TrackMovingSourceFlag)中指示源类型。移动位置源类型和固定位置源类型之间的区别在于:固定源的位置仅在TrackHeader中指示一次;而对于移动源,在每个TrackPackage中指示。可以在球面坐标下利用位置向量明确地指示源的位置,或者可以将源的位置隐含地指示为HOA编码矢量。源本身是PCM单声道音轨,在使用高保真度立体声响复制解码器进行重放的情况下,其必须在解码器侧被编码为HOA系数。

[0269] 单源固定位置音轨头部(Single Source fixed Position Track Header)

[0270]

字段名称	大小 / 比特	数据类型	描述
TrackMovingSourceFlag	1	二进制	常数 '0' 用于固定源
TrackPositionType	1	二进制	'0' 位置被作为角度TrackPositionVector [R, theta, phi]发送 '1' 位置被作为TrackHOAParamNumberOfCoeffs长度的HOA编码矢量发送
TrackSampleFormat	4	二进制	0b0000 无符号整数 8比特 0b0001 有符号整数 8比特 0b0010 有符号整数 16比特 0b0011 有符号整数 24比特 0b0100 有符号整数 32比特 0b0101 有符号整数 64比特 0b0110 浮点32比特(二进制单精度) 0b0111 浮点64比特(二进制双精度) 0b1000 浮点128 比特(二进制四精度) 0b1001-0b1111 保留
保留	2	二进制	填充比特

[0271]

条件: TrackPositionType == '0'			作为角度TrackPositionVector的位置如下
TrackPositionTheta	32	float32	rad [0..pi]内的倾斜角
TrackPositionPhi	32	float32	rad [0..2pi]内的方位角 (逆时针)
TrackPositionRadius	32	float32	以米为单位的距参考点的距离
条件: TrackPositionType == '1'			作为HOA编码矢量的位置
TrackHOAParams	dyn	字节	参见TrackHOAParams
TrackEncodeVectorComplex Flag	2	二进制	0b00: 仅实部 0b01: 实部和虚部 0b10: 仅虚部 0b11: 保留 用于编码矢量的数量类型
TrackEncodeVectorFormat	1	二进制	'0' float32 '1' float 64
保留	5	二进制	填充比特
条件: TrackEncodeVectorFormat == '0'			作为float 32的编码矢量

[0272]

<TrackHOAEncodingVector>	dyn	float32	按照TrackHOAParamCoeffSequence顺序的HOA编码矢量的TrackHOAParamNumberOfCoeffs条目
条件: TrackEncodeVectorFormat == '1'			作为float 64的编码矢量
<TrackHOAEncodingVector>	dyn	float64	按照TrackHOAParamCoeffSequence顺序的HOA编码矢量的TrackHOAParamNumberOfCoeffs条目

[0273] 通过音轨移动源标志(TrackMovingSourceFlag)为零定义固定位置源类型。第二字段指示音轨位置类型(TrackPositionType), 其将源位置的编码给出为球面坐标中的矢量或者 HOA 编码矢量。通过音轨样本格式(TrackSampleFormat) 字段指示单声道 PCM 样本的编码格式。如果源位置被发送为音轨位置矢量(TrackPositionVector), 则在音轨位置角度(TrackPositionTheta) (从 s 轴到 x-, y- 平面的倾斜度)、音轨位置方位角(TrackPositionPhi) (在 x 轴开始的逆时针方位角) 和音轨位置半径(TrackPositionRadius) 字段中定义源位置的球面坐标。

[0274] 如果源位置被定义为 HOA 编码矢量, 则首先定义音轨 HOA 参数(TrackHOAParam)。在 TrackHOAParam 部分中定义这些参数, 并且这些参数指示 HOA 编码矢量所使用的规范化和定义。音轨编码矢量复数标志(TrackEncodeVectorComplexFlag) 和音轨编码矢量格式(TrackEncodeVectorFormat) 字段定义了接下来的音轨 HOA 编码矢量(TrackHOAEncodingVector) 的格式类型。TrackHOAEncodingVector 由以 'float32' 或 'float64' 格式编码的系数音轨 HOA 参数数量(TrackHOAParamNumberOfCoeffs)的值组

成。

[0275] 单源移动位置音轨头部(Single Source moving Position Track Header)

[0276]

字段名称	大小 / 比特	数据类型	描述
TrackMovingSourceFlag	1	二进制	常数 '1' 用于移动源
TrackPositionType	1	二进制	'0' 位置被作为角度TrackPositionVector [R, theta, phi] 位置发送 '1' 位置被作为TrackHOAParamNumberOfCoeffs长度的HOA编码矢量发送

[0277]

字段名称	大小 / 比特	数据类型	描述
TrackSampleFormat	4	二进制	0b0000 无符号整数 8比特 0b0001 有符号整数 8比特 0b0010 有符号整数 16比特 0b0011 有符号整数 24比特 0b0100 有符号整数 32比特 0b0101 有符号整数 64比特 0b0110 浮点32比特(二进制单精度) 0b0111 浮点64比特(二进制双精度) 0b1000 浮点128 比特(二进制四精度) 0b1001-0b1111 保留
保留	2	二进制	填充比特

[0278]

条件: TrackPositionType == '1'			作为HOA编码矢量的位置
TrackHOAParams	dyn	字节	参见TrackHOAParams
TrackEncodeVectorComplex Flag	2	二进制	0b00: 仅实部 0b01: 实部和虚部 0b10: 仅虚部 0b11: 保留 用于编码矢量的数量类型
TrackEncodeVectorFormat	1	二进制	'0' float32 '1' float64
保留	5	二进制	填充比特

[0279] 通过 TrackMovingSourceFlag 为 '1' 来定义移动位置源类型。除了源位置数据字段 TrackPositionTheta, TrackPositionPhi, TrackPositionRadius 和 TrackHOAEncodingVector 没出现之外, 该头部与固定源头部相同。对于移动源, 这些字段位于 TrackPacket 中以便在每个分组中指示新(移动)源位置。

[0280] 具体音轨表(Special Track Table)

[0281] 音轨 HOA 参数(TrackHOAParam)

[0282]

字段名称	大小 / 比特	数据类型	描述
TrackHOAParamDimension	1	二进制	'0' = 2D, 以及 '1' = 3D

[0283]

字段名称	大小 / 比特	数据类型	描述
TrackHOAParamRegionOfInterest	1	二进制	‘0’ 对于感兴趣区域外部的源计算HOA系数（内部） ‘1’ 对于感兴趣区域内的源计算HOA系数（外部） （感兴趣区域不包含任何源）
TrackHOAParamSphericalHarmonicType	1	二进制	‘0’ 实数 ‘1’ 复数
TrackHOAParamSphericalHarmonicNorm	3	二进制	0b000 未规范化 0b001 施密特半规范化 0b010 4π 规范化或者2D 规范化 0b011 正交规范化 0b100 专用缩放 其它保留
TrackHOAParamFurseMalhamFlag	1	二进制	指示通过福尔斯马勒姆缩放来规范化HOA系数
TrackHOAParamDecoderType	2	二进制	0b00 平面波解码器缩放: $\frac{1}{(4\pi i^n)}$ 0b01 球面波解码器缩放（距离编码）: $\frac{1}{(ikh_n(kr_{ls}))}$ 0b10 球面波解码器缩放（用于所测量的声压的距离编码）: $\frac{h_0(kr_{ls})}{(ikh_n(kr_{ls}))}$ 0b11 清楚的HOA系数
TrackHOAParamCoeffSequence	2	二进制	0b00 B-格式顺序 0b01 数值向上 0b10 数值向下 0b11 保留
保留	5	二进制	填充比特
TrackHOAParamNumberOfCoeffs	16	uint16	每个样本的HOA系数的数量减一
TrackHOAParamHorizontalOrder	8	uint8	X/Y平面中高保真度立体声响复制顺序
TrackHOAParamVerticalOrder	8	uint8	用于3D维度的高保真度立体声响复制顺序（‘0’用于2D HOA系数）

[0284]

条件: TrackHOAParamSphericalHarmonicNorm == "dedicated" <0b101>			用于每个HOA系数的专用缩放值的字段
TrackComplexValueScalingFlag	2	二进制	0b00: 仅实部 0b01: 实部和虚部 0b10: 仅虚部 0b11: 保留 专用TrackScalingValues的数量类型
TrackScalingFormat	1	二进制	'0': float32 '1': float64
保留	5	二进制	填充比特

[0285]

条件: TrackScalingFormat = '0'		作为float32的TrackScalingFactors	
<TrackScalingFactors>	dyn	float32	TrackHOAParamNumberOfCoeffs 缩放因子 如果TrackComplexValueScalingFlag == 0b01, 则复数部分的顺序为 <[real1, imaginary1], [real2, imaginary2], ..., [realN, imaginary] >
条件: TrackScalingFormat = '1'		作为float64的TrackScalingFactors	
<TrackScalingFactors>	dyn	float64	TrackHOAParamNumberOfCoeffs 缩放因子 如果TrackComplexValueScalingFlag == 0b01, 则复数部分的顺序为<[real1, imaginary1], [real2, imaginary2], ..., [realN, imaginary] >
条件: TrackHOAParamDecoderType == 0b01 TrackHOAParamDecoderType == 0b10		定义用于距离编码的参考扬声器半径	
TrackHOAParamReferenceRadius	16	uint16	这是以mm为单位的参考扬声器半径 r_{ls} , 其已经应用于根据Poletti或Daniel的球面波解码器的HOA系数。

[0286] 过去已经讨论了用于HOA编码和解码的几种方法。然而, 尚没有用于编码HOA系数的任何结论或一致观点。有利地, 根据本发明的格式允许存储大多数已知的HOA表示。定义TrackHOAParam以便使得清楚在编码器侧已经使用了何种系数规范化和顺序序列。在解码器侧必须考虑这些定义, 以便混合HOA音轨以及应用解码器矩阵。

[0287] HOA系数可以应用于完整的三维声音场, 或者仅仅应用于二维x/y-平面。通过音轨HOA参数维度(TrackHOAParamDimension)字段定义HOA音轨(HOATrack)的维度。

[0288] 感兴趣区域的音轨HOA参数(TrackHOAParamRegionOfInterest)顺序地反映了两个声压扩张, 由此源驻留于感兴趣区域内部或外部, 并且感兴趣区域不包含任何源。分别在以上方程(1)和(2)中定义了用于内部情况和外部情况的声压计算, 由此通过共轭复数球面谐波函数 $Y_n^m(\theta, \phi)^*$ 确定HOA信号的方向性信息 $A_n^m(k)$ 。该函数是在复数和实数形式

下定义的。编码器和解码器必须应用等效数量类型的球面谐波函数。因此,音轨 HOA 参数球面谐波类型(TrackHOAParamSphericalHarmonicType)指示在编码器侧已经应用何种球面谐波函数。

[0289] 如上所述,基本上通过相关联的拉格朗日函数以及复数或实数三角函数来定义球面谐波函数。通过方程(5)定义相关联的拉格朗日函数。复数值球面谐波表示为:

$$[0290] \quad Y_n^m(\theta, \phi) = N_{n,m} P_{n,|m|}(\cos(\theta)) e^{im\phi} \begin{cases} (-1)^m & ; m \geq 0 \\ 1 & ; m < 0 \end{cases}$$

[0291] 其中, $N_{n,m}$ 是缩放因子(参见方程(3))。使用以下方程,可以将该复数值表示转换为实数值表示:

$$[0292] \quad S_n^m(\theta, \phi) = \begin{cases} \frac{(-1)^m}{\sqrt{2}} (Y_n^m + Y_n^{m*}) & = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) \cos(m\phi), & m > 0 \\ Y_n^0 & = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) & m = 0 \\ \frac{-1}{i\sqrt{2}} (Y_n^m - Y_n^{m*}) & = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) \sin(|m|\phi), & m < 0 \end{cases}$$

[0293] 其中,用于实数球面谐波的修改后的缩放因子为:

$$[0294] \quad \tilde{N}_{n,m} = \sqrt{2 - \delta_{0,m}} N_{n,m}, \quad \delta_{0,m} = \begin{cases} 1 & ; m = 0 \\ 0 & ; m \neq 0 \end{cases}$$

[0295] 对于 2D 表示,必须将圆谐波函数用于 HOA 系数的编码和解码。通过 $\tilde{Y}_m(\phi) = \tilde{N}_m e^{im\phi}$ 来定义圆谐波的复数值表示。

[0296] 通过 $\tilde{S}_m(\phi) = \tilde{N}_m \begin{cases} \cos(m\phi) & ; m \geq 0 \\ \sin(|m|\phi) & ; m < 0 \end{cases}$ 来定义圆谐波的实数值表示。

[0297] 几个规范化因子 $N_{n,m}$, $\tilde{N}_{n,m}$, $\tilde{N}_m$ 和 $\tilde{N}_m$ 被用于将球面或圆谐波函数适配于特定应用或要求。为了确保正确解码 HOA 系数,在解码器侧必须知道在编码器侧使用的球面谐波函数的规范化。下表 7 定义了可以利用音轨 HOA 参数球面谐波规范化(TrackHOAParamSphericalHarmonicNorm)字段选择的规范化。

[0298]

3D 复数值球面谐波规范化 $N_{n,m}$			
未规范化 0b000	施密特半规范化 SN3D 0b001	4π 规范化 测量 4 π 0b010	正交规范化 0b011
1	$\frac{(n- m)!}{(n+ m)!}$	$\frac{(2n+1)(n- m)!}{(n+ m)!}$	$\frac{(2n+1)(n- m)!}{4\pi(n+ m)!}$
3D 实数值球面谐波规范化 $N_{n,m}$			
未规范化 0b000	施密特半规范化 SN3D 0b001	4π 规范化 测量 4 π 0b010	正交规范化 0b011
$\sqrt{2-\delta_{0,m}}$	$\sqrt{(2-\delta_{0,m})} \frac{(n- m)!}{(n+ m)!}$	$\sqrt{(2-\delta_{0,m})} \frac{(2n+1)(n- m)!}{(n+ m)!}$	$\sqrt{(2-\delta_{0,m})} \frac{(2n+1)(n- m)!}{4\pi(n+ m)!}$
2D 复数值圆谐波规范化 N_n			
未规范化 0b000	施密特半规范化 SN2D 0b001	2D 规范化 H2D 0b010	正交规范化 0b011
$\sqrt{\frac{1}{2}}$	$\sqrt{\frac{1+\delta_{0,m}}{2}}$	1	$\sqrt{\frac{1}{2\pi}}$
2D 实数值圆谐波规范化 N_n			
未规范化 0b000	施密特半规范化 SN2D 0b001	2D 规范化 H2D 0b010	正交规范化 0b011
$\sqrt{\frac{2-\delta_{0,m}}{2}}$	1	$\sqrt{(2-\delta_{0,m})}$	$\sqrt{(2-\delta_{0,m})} \frac{1}{2\pi}$

[0299] 表 7 - 球面和圆谐波函数的规范化

[0300] 对于将来规范化, TrackHOAParamSphericalHarmonicNorm 字段的专用值是可用的。对于专用规范化, 在音轨 HOA 参数(TrackHOAParam) 的结尾定义了用于每个 HOA 系数的缩放因子。可以将专用缩放因子 TrackScalingFactor 作为实数或复数 'float32' 或 'float64' 值传送。在专用缩放的情况下, 在音轨复数值缩放标志(TrackComplexValueScalingFlag) 和音轨缩放格式(TrackScalingFormat) 字段中定义了缩放因子格式。

[0301] 福尔斯马勒姆规范化可以另外应用于编码的 HOA 系数, 以便将不同 HOA 顺序的系数的幅度均衡化为小于 '1' 的绝对值, 以便以整数格式类型传送。福尔斯马勒姆规范化被指定用于高达三阶系数的 SN3D 实数值球面谐波函数。因此, 仅仅推荐将福尔斯马勒姆规范化与 SN3D 实数值球面谐波函数组合使用。此外, 对于具有 HOA 阶高于三的音轨, 忽略音轨 HOA 参数福尔斯马勒姆标志。在解码器侧必须转换福尔斯马勒姆规范化以便解码 HOA 系数。表 8 定义了福尔斯马勒姆系数。

[0302]

n	m	福尔斯马勒姆权重
0	0	$1/\sqrt{2}$
1	-1	1
1	0	1
1	1	1
2	-2	$2/\sqrt{3}$
2	-1	$2/\sqrt{3}$
2	0	1
2	1	$2/\sqrt{3}$
2	2	$2/\sqrt{3}$
3	-3	$\sqrt{8/5}$
3	-2	$3/\sqrt{5}$
3	-1	$\sqrt{45/32}$
3	0	1
3	1	$\sqrt{45/32}$
3	2	$3/\sqrt{5}$
3	3	$\sqrt{8/5}$

[0303] 表 8 — 在编码器侧要应用的福尔斯马勒姆规范化因子

[0304] 音轨 HOA 参数解码器类型 (TrackHOAParamDecoderType) 定义了 在编码器侧假设在解码器侧出现了何种解码器。解码器类型确定要在解码器侧用来呈现声音场的扬声器模型 (球面或平面波)。由此, 通过将解码器方程的部分平移到解码器方程, 可以降低解码器的计算复杂度。另外, 可以降低编码器侧的数值问题。此外, 可以将解码器降低到对于所有 HOA 系数进行相同处理, 这是因为解码器侧的所有不一致可以被移动到编码器。然而, 对于球面谐波, 可以假设扬声器到收听位置具有恒定距离。因此, 在 TrackHeader 中指示了所假设的解码器类型, 并且在可选的音轨 HOA 参数参考半径字段 (TrackHOAParamReferenceRadius) 中以毫米为单位传送用于球面波解码器类型的扬声器半径 r_{ls} 。解码器侧的另外的滤波器可以均衡化所假设的扬声器半径和实际扬声器半径之间的差异。

[0305] HOA 系数 C_n^m 的音轨 HOA 参数解码器类型 (TrackHOAParamDecoder Type) 规范化取决于在 TrackHOAParamRegionOfInterest 中选择的顺序的内部或外部声音场扩张的使用。
 注释: 方程 (18) 和以下方程中的系数 d_n^m 对应于下面的系数 C_n^m 。在编码器侧, 依据在表 9 中定义的系数 A_n^m 或 B_n^m 来确定系数 C_n^m , 并且存储系数 C_n^m 。在 TrackHOAParam 头部的 TrackHOAParamDecoderTyp 字段中指示了所使用的规范化:

[0306]

TrackHOAParamDecoderType	HOA系数内部	HOA系数外部
0b00 平面波	$C_n^m = A_n^m / (4\pi r^n)$	—
0b01 球面波	$C_n^m = A_n^m / (ik h_n(kr_{10}))$	$C_n^m = A_n^m / (ik j_n(kr_{10}))$
0b10 球面波测量的声压	$C_n^m = A_n^m h_n(kr_{10}) / (h_n(kr_{10}))$	$C_n^m = A_n^m h_n(kr_{10}) / (j_n(kr_{10}))$
0b11 未规范化	$C_n^m = A_n^m$	$C_n^m = B_n^m$

[0307] 表 9 —用于几种解码器类型规范化的所传送的 HOA 系数

[0308] 用于一个时间样本的 HOA 系数包括系数 C_n^m 的数量 TrackHOAParamNumberOfCoeffs(0)。N 取决于 HOA 系数的维度。对于 2D 声音场，'0' 等于 2N+1，其中 N 等于来自 TrackHOAParam 头部的音轨 HOA 参数水平阶 (TrackHOAParamHorizontalOrder) 字段。2D HOA 系数被定义为 $C_{|m|}^m = C_m$ ，其中， $-N \leq m \leq N$ ，并且可以被表示为 3D 系数的子集，如表 10 所示。

[0309] 对于 2D 声音场，'o' 等于 $(N+1)^2$ ，其中 N 等于来自 TrackHOAParam 头部的音轨 HOA 参数垂直阶 (TrackHOAParamVerticalOrder) 字段。3D HOA 系数 C_n^m 被定义用于 $0 \leq n \leq N$ 和 $-n \leq m \leq n$ 在表 10 中给出了 HOA 系数的共同表示。

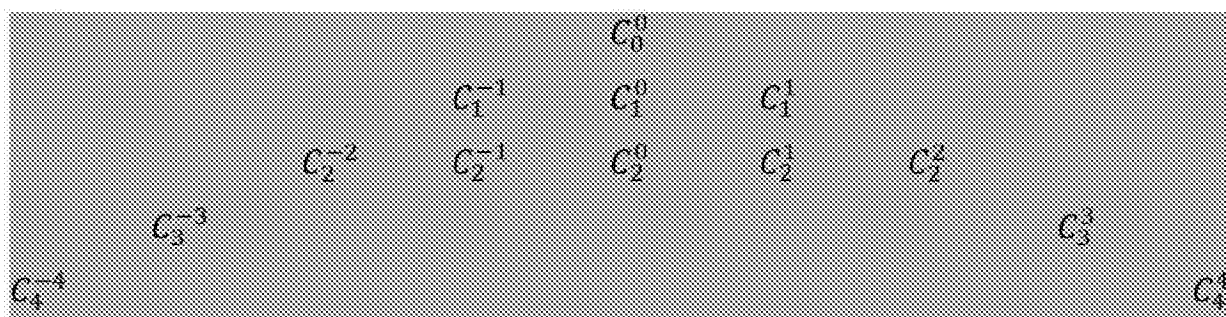
[0310]

				C_0^0				
			C_1^{-1}	C_1^0	C_1^1			
		C_2^{-2}	C_2^{-1}	C_2^0	C_2^1	C_2^2		
	C_3^{-3}	C_3^{-2}	C_3^{-1}	C_3^0	C_3^1	C_3^2	C_3^3	
C_4^{-4}	C_4^{-3}	C_4^{-2}	C_4^{-1}	C_4^0	C_4^1	C_4^2	C_4^3	C_4^4

[0311] 表 10 —高达四阶的 HOA 系数表示，将粗体的 2D 系数示出为 3D 系数的子集

[0312] 在 3D 声音场以及 TrackHOAParamHorizontalOrder 大于 TrackHOAParamVerticalOrder 的情况下，将执行混合阶解码。在混合阶信号中，仅在 2D 中传送一些更高阶系数。TrackHOAParamVerticalOrder 字段确定传送其所有系数的垂直阶。从该垂直阶到 TrackHOAParamHorizontalOrder，仅使用 2D 系数。因此，TrackHOAParamHorizontalOrder 等于或大于 TrackHOAParamVerticalOrder。在表 11 中绘制了四水平阶和二垂直阶的混合阶表示的示例：

[0313]



[0314] 表 11 — 用于二垂直阶和四水平阶的混合阶表示的 HOA 系数表示

[0315] 在音轨的分组中存储 HOA 系数 C_n^m 过去已经不同地定义了系数序列(例如哪个系数在先哪个系数在后)。因此,音轨 HOA 参数系数序列(TrackHOAParamCoeffSequence) 字段指示三种类型的系数序列。这三种序列是从表 10 的 HOA 系数排列中导出的。

[0316] B- 格式序列将特殊措辞(special wording)用于高达三阶的 HOA 系数,如表 12 所示:

[0317]

			W			
		Y	S	X		
	V	T	R	S	U	
Q	O	M	K	L	N	P

[0318] 表 12 — B- 格式 HOA 系数命名约定

[0319] 对于 B- 格式,从最低阶到最高阶传送 HOA 系数,其中按照字母表顺序传送每一阶的 HOA 系数。例如,按照序列 W, X, Y, S, R, S, T, U, V, K, L, M, N, O, P 和 Q 存储 HOA 阶三的 3D 设置的系数。B- 格式仅仅被定义高达第三 HOA 阶。对于水平(2D)系数的传送,忽略补充性的 3D 系数,例如 W, X, Y, U, V, P, Q。

[0320] 从最低到最高 HOA 阶 ($n = 0 \dots N$) 以数值向上或向下的方式在音轨 HOA 参数系数序列(TrackHOAParamCoeffSequence)中传送用于 3D HOA 的系数 C_n^m 。数值向上序列以 $m = -n$ 开始,并且增加到 $m = n$

[0321] $(C_0^0, C_1^{-1}, C_1^0, C_1^1, C_2^{-2}, C_2^{-1}, C_2^0, C_2^1, C_2^2, \dots)$, 其是在 Chris Travis, "Four candidate component sequences", <http://ambisonics.googlegroups.com/web/Four+candidate+component+sequences+V09.pdf>, 2008 中定义的 'CG' 序列。数值向下序列 m 按照从 $m = n$ 到 $m = -n$ ($C_0^0, C_1^1, C_1^0, C_1^{-1}, C_2^2, C_2^1, C_2^0, C_2^{-1}, C_2^{-2}, \dots$) 的另一种方式运行,其是在该发表物中定义的 'QM' 序列。

[0322] 对于 2D HOA 系数,TrackHOAParamCoeffSequence 数值向上和向下序列与 3D 情况相似,但其中省略具有 $|m| \neq n$ 的无用系数(即,仅仅表 10 中的部分 HOA 系数 $C_{|m|}^m = C_m$)。

因此,数值向上序列导致 $(C_0^0, C_1^{-1}, C_1^0, C_1^1, C_2^{-2}, C_2^{-1}, C_2^0, C_2^1, C_2^2, \dots)$, 数值向下序列导致

$(C_0^0, C_1^1, C_1^0, C_1^{-1}, C_2^2, C_2^1, C_2^0, C_2^{-1}, C_2^{-2}, \dots)$ 。

[0323] 音轨分组 (Track Packet)

[0324] HOA 音轨分组

[0325] PCM 编码类型分组

[0326]

字段名称	大小 / 比特	数据类型	描述
<PacketHOACoeffs>	dyn	dyn	以TrackSampleFormat和TrackHOAParamCoeffSequence存储的声道交织的HOA系数，例如 <[W(0), X(0), Y(0), S(0)], [W(1), X(1), Y(1), S(1)], ..., S(FrameNumberOfSamples - 1)]>

[0327] 该分组包含按照在 TrackHOAParamCoeffSequence 中定义的顺序的 HOA 系数 C_n^m ，其中连续地传送一个时间样本的所有系数。在 TrackSourceType 为零并且 TrackCodingType 为零的情况下，该分组用于标准 HOA 音轨。

[0328] 动态分辨率编码类型分组

[0329]

字段名称	大小 / 比特	数据类型	描述
<PacketHOACoeffsCode d>	dyn	dyn	根据TrackCodingType存储的声道去交织的HOA系数，例如<[W(0), W(1), W(2), ...], [X(0), X(1), X(2), ...], [Y(0), Y(1), Y(2), ...], [S(0), S(1), S(2), ...]>

[0330] 动态分辨率分组用于‘零’的音轨源类型 (TrackSourceType) 和‘一’的音轨编码类型 (TrackCodingType)。音轨阶区域 (TrackOrderRegion) 的不同分辨率导致用于每个 TrackOrderRegion 的不同存储大小。因此，以去交织方式存储 HOA 系数，例如连续地存储一个 HOA 阶的所有系数。

[0331] 单源音轨分组

[0332] 单源固定位置分组

[0333]

字段名称	大小 / 比特	数据类型	描述
<PacketMonoPCMTrack>	dyn	dyn	以TrackSampleFormat存储的单音频源的PCM样本

[0334] 单源固定位置分组用于‘一’的音轨源类型 (TrackSourceType) 和‘零’的音轨移动源标志 (TrackMovingSourceFlag)。该分组保持单声道源的 PCM 样本。

[0335] 单源移动位置分组

[0336]

字段名称	大小 / 比特	数据类型	描述
PacketDirectionFlag	1	二进制	如果方向已经改变则设置为‘1’。对于帧的第一分组，强制为‘1’。
保留	7	二进制	填充比特

条件: PacketDirectionFlag == ‘1’		随后为新位置数据	
条件: TrackPositionType == ‘0’		位置TrackPositionVector为角度TrackPositionVector	
theta	32	float32	rad [0..pi]内的倾斜角
phi	32	float32	rad [0..2pi]内的方位角（逆时针）
radius	32	float32	以米为单位的距参考点的距离

[0337]

条件: TrackPositionType == ‘1’	位置为 HOA编码矢量
------------------------------	-------------

[0338]

条件: TrackEncodeVectorFormat == ‘0’		编码矢量为float32	
<TrackHOAEncodingVector>	dyn	float32	按照TrackHOAParamCoeffSequence顺序的HOA编码矢量的TrackHOAParamNumberOfCoeffs项
条件: TrackEncodeVectorFormat == ‘1’		编码矢量为float64	
<TrackHOAEncodingVector>	dyn	float64	按照TrackHOAParamCoeffSequence顺序的HOA编码矢量的TrackHOAParamNumberOfCoeffs项

[0339]

<PacketMonoPCMTrack>	dyn	dyn	以TrackSampleFormat存储的单音频源的PCM 样本
----------------------	-----	-----	----------------------------------

[0340] 单源移动位置分组用于‘一’的音轨源类型(TrackSourceType)和‘一’的音轨移动源标志(TrackMovingSourceFlag)。其保持单声道PCM样本、以及用于TrackPacket的样本的位置信息。

[0341] 分组方向标志(PacketDirectionFlag)指示分组的方向是否已经改变、或者使用应使用前一分组的方向。为了确保从每帧的开头开始解码,对于帧的第一个移动源TrackPacket, PacketDirectionFlag等于‘一’。

[0342] 对于‘一’的PacketDirectionFlag,传送以下PCM样本源的方向信息。取决于TrackPositionType,方向信息被作为球面坐标下的音轨位置矢量(TrackPositionVector)发送,或者作为具有所定义的音轨编码矢量格式(TrackEncodingVectorFormat)的音轨HOA编码矢量(TrackHOAEncodingVector)发送。音轨编码矢量(TrackEncodingVector)生

成符合 HOA 参数头部(HOAParamHeader)字段定义的 HOA 系数。与方向性信息连续地,传送 TrackPacket 的 PCM 单声道样本的方向性信息。

[0343] 编码处理

[0344] 音轨区域编码

[0345] 可以利用麦克风阵列从声音场记录中导出 HOA 信号。例如,可以使用在 W003/061336A1 中公开的 Eigenmike 来获得阶三的 HOA 记录。然而,麦克风阵列的有限大小导致对所记录的 HOA 系数的约束。在 W003/061336A1、以及上述文章 "Three-dimensional surround sound systems based on spherical harmonics" 中,讨论了由有限麦克风阵列引起的问题。

[0346] 麦克风单元(capsule)的距离导致空间采样理论给出的上频率边界。在该上频率之上,麦克风阵列不能产生正确的 HOA 系数。此外,麦克风距 HOA 收听位置的有限距离要求均衡滤波器。这些滤波器对于低频获得高增益,甚至随着每个 HOA 阶增高。在 W003/061336A1 中,引入了用于高阶系数的低截止频率,以便操纵均衡滤波器的动态范围。这示出了不同 HOA 阶的 HOA 系数的带宽可能不同。因此,HOA 文件格式提供了音轨区域带宽降低(TrackRegionBandwidthReduction),其使得对于每个 HOA 阶能够仅传送所需要的频率带宽。

[0347] 由于均衡滤波器的高动态范围以及由于零阶系数基本上是所有麦克风信号之和的事实,不同 HOA 阶的系数可能具有不同的动态范围。因此,HOA 文件格式也提供了将格式类型适配于每个 HOA 阶的动态范围的特征。

[0348] 音轨区域编码处理(TrackRegion Encoding Processing)

[0349] 如图 12 所示,交织的 HOA 系数被送入第一去交织步骤或阶段 1211,其被指派给第一音轨区域(TrackRegion)并且将该 TrackRegion 的所有 HOA 系数分离到去交织缓冲器中以成为帧分组大小(FramePacketSize)的样本。从 HOA 音轨头部的音轨区域最后阶(TrackRegionLastOrder)和音轨区域第一阶(TrackRegionFirstOrder)字段中导出的 TrackRegion 系数。去交织意味着用于 n 和 m 的一个组合的系数 C_n^m 被编组到一个缓冲器中。从去交织步骤或阶段 1211,去交织的 HOA 系数被传送到音轨区域(TrackRegion)编码部分。其余的去交织的 HOA 系数被传送到下面的音轨区域去交织部分或阶段等等,直至去交织步骤或阶段 121N。去交织步骤或阶段的数量 N 等于 TrackNumberOfOrderRegions 加 '一'。另外的去交织步骤或阶段 125 将不是 TrackRegion 的部分的其余系数去交织到包括格式转换步骤或阶段 126 的标准处理路径中。

[0350] 音轨区域编码路径包括可选的带宽降低步骤或阶段 1221、以及格式转换步骤或阶段 1231,并且对于每个 HOA 系数缓冲器执行并行处理。如果 TrackRegionUseBandwidthReduction 字段被设置为 '一',则执行带宽降低。取决于所选择的 TrackBandwidthReductionType,选择处理以用于限制 HOA 系数的频率范围或者用于关键地对 HOA 系数进行下采样。这被执行来将 HOA 系数的数量降低到最小所需的样本数量。格式转换将当前的 HOA 系数格式转换为在 HOATrack 头部中定义的音轨区域样本格式(TrackRegionSampleFormat)。这是标准处理路径中仅有的将 HOA 系数转换为 HOA 音轨头部的所指示的 TrackSampleFormat 的步骤/阶段。

[0351] 多路复用器 TrackPacket 步骤或阶段 124 将 HOA 系数缓冲器多路复用到

TrackPacket 数据文件流中,如在所选择的 TrackHOAParamCoeffSequence 字段中定义的,其中用于 n 和 m 的一个组合的系数 C_n^m 索引(在一个缓冲器内)保持去交织。

[0352] 音轨区域解码处理(TrackRegion Decoding Processing)

[0353] 如图 13 所示,解码处理与编码处理相反。去多路复用器步骤或阶段 134 将来自所指示的 TrackHOAParamCoeffSequence 的 TrackPacket 数据文件或流去多路复用到去交织的 HOA 系数缓冲器(未示出)中。每个缓冲器包含用于 n 和 m 的一个组合的 FramePacketLength 系数 C_n^m 。

[0354] 步骤/阶段 134 初始化 TrackNumberOfOrderRegion 加‘一’处理路径,并且将去交织的 HOA 系数缓冲器的内容传送到适当的处理路径。通过 HOA 音轨头部的 TrackRegionLastOrder 和 TrackRegionFirstOrder 字段定义每个 TrackRegion 的系数。在包括格式转换步骤或阶段 136 和其余系数交织步骤或阶段 135 的标准处理路径中处理未被所选择的 TrackRegions 覆盖的 HOA 阶。标准处理路径对应于没有带宽降低步骤或阶段的音轨处理路径。

[0355] 在音轨处理 TrackProcessing 路径中,格式转换步骤/阶段 1331 到 133N 将以 TrackRegionSampleFormat 编码的 HOA 系数转换为用于解码器处理的数据格式。取决于 TrackRegionUseBandwidthReduction 数据字段,接着可选的带宽重构步骤或阶段 1321 到 132N,其中将带宽首先的并且关键采样的 HOA 系数重构至音轨的全带宽。在 HOA 音轨头部的 TrackBandwidthReductionType 字段中定义了重构处理的种类。

[0356] 在以下交织步骤或阶段 1311 到 131N 中,通过将一个时间样本的 HOA 系数编组,来将 HOA 系数的去交织的缓冲器的内容交织,将当前 TrackRegion 的 HOA 系数与之前 TrackRegion 的 HOA 系数组合。可以将 HOA 系数的结果序列适配至音轨的处理。此外,交织步骤/阶段处理使用带宽降低的音轨区域(TrackRegions)和不使用带宽降低的音轨区域之间的延迟,该延迟取决于所选择的音轨带宽降低类型(TrackBandwidthReductionType)处理。例如,MDCT 处理添加帧分组大小(FramePacketSize)样本的延迟,并且因此没有带宽降低的处理路径的交织步骤/阶段将其输出延迟一个分组。

[0357] 经由 MDCT 的带宽降低(Bandwidth reduction via MDCT)

[0358] 编码

[0359] 图 14 示出了使用 MDCT (修改的离散余弦变换)处理的带宽降低。经由缓冲器 1411 到 141M, FramePacketSize 样本的音轨区域(TrackRegion)的每个 HOA 系数经过对应的 MDCT 窗添加步骤或阶段 1421 到 142M。每个输入缓冲器包含 n 和 m 的一个组合的时间连续的 HOA 系数 C_n^m ,即,一个缓冲器被定义为

$$(\text{buffer}) C_n^m = [C_n^m(0), C_n^m(1), \dots, C_n^m(\text{FramePacketSize} - 1)]$$

[0360] 缓冲器的数量 M 与高保真度立体声响复制分量的数量相同(对于阶 N 的全 3D 声场,为 $((N+1)^2)$)。通过将之前缓冲器内容与当前缓冲器内容组合为用于对应步骤或阶段 1431 到 143M 中以下 MDCT 处理的新内容,缓冲器处理执行用于该 MDCT 处理的 50% 重叠,并且其存储当前缓冲器内容以用于下面缓冲器内容的处理。MDCT 处理在每帧的开头处重启,这意味着可以在不知道之前帧的情况下解码当前帧的音轨的所有系数,在当前帧的最后缓

冲器内容之后,处理零另外缓冲器内容。因此,MDCT 处理后的 TrackRegions 产生一个额外 TrackPacket。

[0361] 在窗口添加步骤 / 阶段中,将对应缓冲器内容与所选择的窗口函数 $w(t)$ 相乘,在用于每个 TrackRegion 的 HOATrack 头部字段 TrackRegionWindowType 中定义了该函数。

[0362] 在 J. P. Princen, A. B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-34, no. 5, pages 1153 - 1161, 1986 年 10 月中首次提到了修改的离散余弦变换。MDCT 可以被考虑为表示 FramePacketSize 子带的关键采样的滤波器组,并且其要求 50% 输入缓冲器重叠。输入缓冲器具有两倍于子带大小的长度。通过以下方程定义 MDCT,其中 T 等于 FramePacketSize,对于 $0 \leq k < T$,

$$[0363] \quad C_n^m(k) = \sum_{t=0}^{2T-1} w(t) C_n^m(t) \cos \left[\frac{\pi}{T} \left(t + \frac{T+1}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

[0364] 系数 $C_n^m(k)$ 被称为 MDCT 库。使用快速傅立叶变换,可以实现 MDCT 计算。

[0365] 在以下频率区域消减步骤或阶段 1441 到 144M 中,通过移除具有 $k < \text{TrackRegionFirstBin}$ 和 $k > \text{TrackRegionLastBin}$ 的所有 MDCT 库 $C_n^m(k)$ 来执行带宽降低,以便将缓冲器长度降低到 $\text{TrackRegionLastBin} - \text{TrackRegionFirstBin} + 1$,其中, TrackRegionFirstBin 是音轨区域的低截止频率, TrackRegionLastBin 是高截止频率。可以将 MDCT 库的忽略视作表示具有与 TrackRegionLastBin 和 TrackRegionFirstBin 频率对应的截止频率的带通滤波器。因此,仅仅传送所需的 MDCT 库。

[0366] 解码

[0367] 图 15 示出了使用 MDCT 处理的带宽解码或重构,其中,带宽受限的音轨区域的 HOA 系数被重构到音轨的全带宽。该带宽重构并行地处理在时间上去交织的 HOA 系数的缓冲器内容,其中,每个缓冲器包含系数 $C_n^m(k)$ 的 $\text{TrackRegionLastBin} - \text{TrackRegionFirstBin} + 1$ 个 MDCT 库。

[0368] 丢失频率区域添加步骤或阶段 1541 到 154M 通过利用使用零的丢失 MDCT 库的 TrackRegionFirstBin 和 TrackRegionLastBin 补充所接收的 MDCT 库,来重构帧分组长度 (FramePacketLength) 大小的完整 MDCT 缓冲器内容。此后,在对应的逆 MDCT 步骤或阶段 1531 到 153M 中执行逆 MDCT,以便重构时间域 HOA 系数 $C_n^m(t)$ 。逆 MDCT 可以被解释为合成滤波器组,其中,帧分组长度 (FramePacketLength) 的 MDCT 库被转换为两倍帧分组长度的时域系数。然而,时域样本的完整重构要求与在编码器中使用的窗口函数 $w(t)$ 与当前缓冲器内容的前一半和前一缓冲器内容的后半的重叠相加的乘法。通过以下方程定义逆 MDCT,对于 $0 < t < T$

$$[0369] \quad C_n^m(t) = \frac{w(t)}{2T} \sum_{k=0}^{T-1} C_n^m(k) \cos \left[\frac{\pi}{T} \left(t + \frac{T+1}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

[0370] 像 MDCT,使用快速傅立叶逆变换,可以实现逆 MDCT。

[0371] MDCT 窗口添加步骤或阶段 1521 到 152M 将重构的时域系数与由音轨区域窗口类型

(TrackRegionWindowType)定义的窗口函数相乘。下面的缓冲器 1511 到 151M 将当前音轨分组缓冲器内容的前一半到最后音轨分组缓冲器内容的后半相加,以便重构帧分组大小(FramePacketSize)的时域系数。存储当前音轨分组缓冲器内容的后半以用于下面音轨分组的处理,重叠相加处理移除了两个缓冲器内容的相反混叠分量。

[0372] 对于多帧 HOA 文件,禁止编码器使用前一帧的最后缓冲器内容,以便在新帧的开头处进行重叠相加例程。因此,在帧边界或者在新帧的开头处,重叠相加的缓冲器内容丢失,并且在第二音轨分组处可以执行帧的第一音轨分组的重构,由此与没有带宽降低的处理路径相比,引入了一个帧分组和解码一个额外音轨分组的延迟。通过结合图 13 描述的组织步骤 / 阶段来处理该延迟。

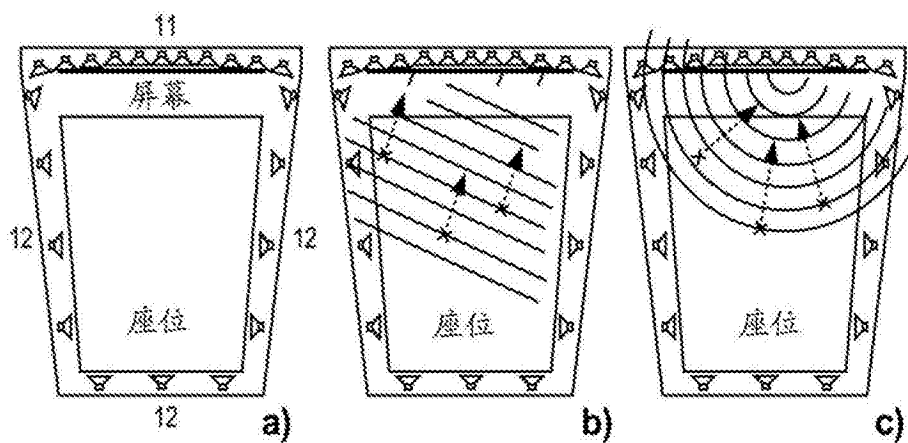


图 1

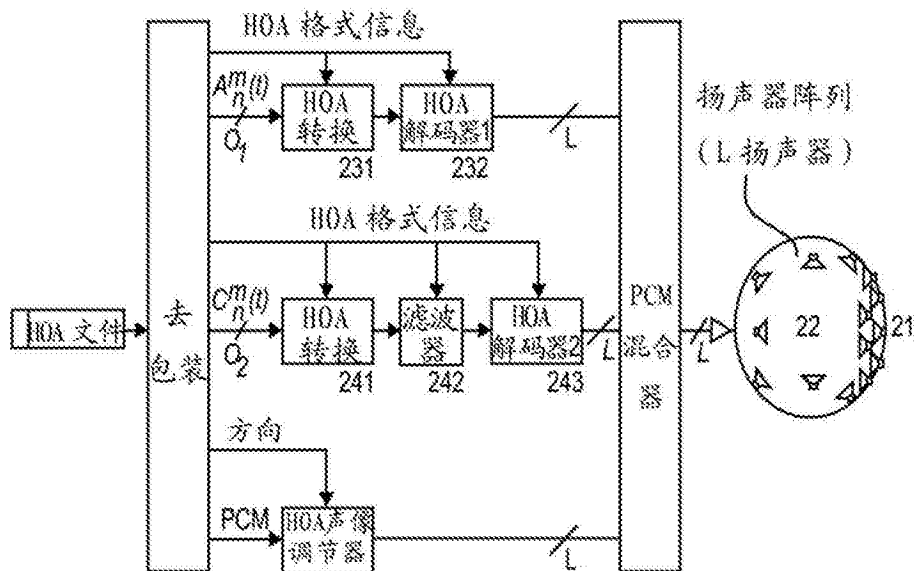


图 2

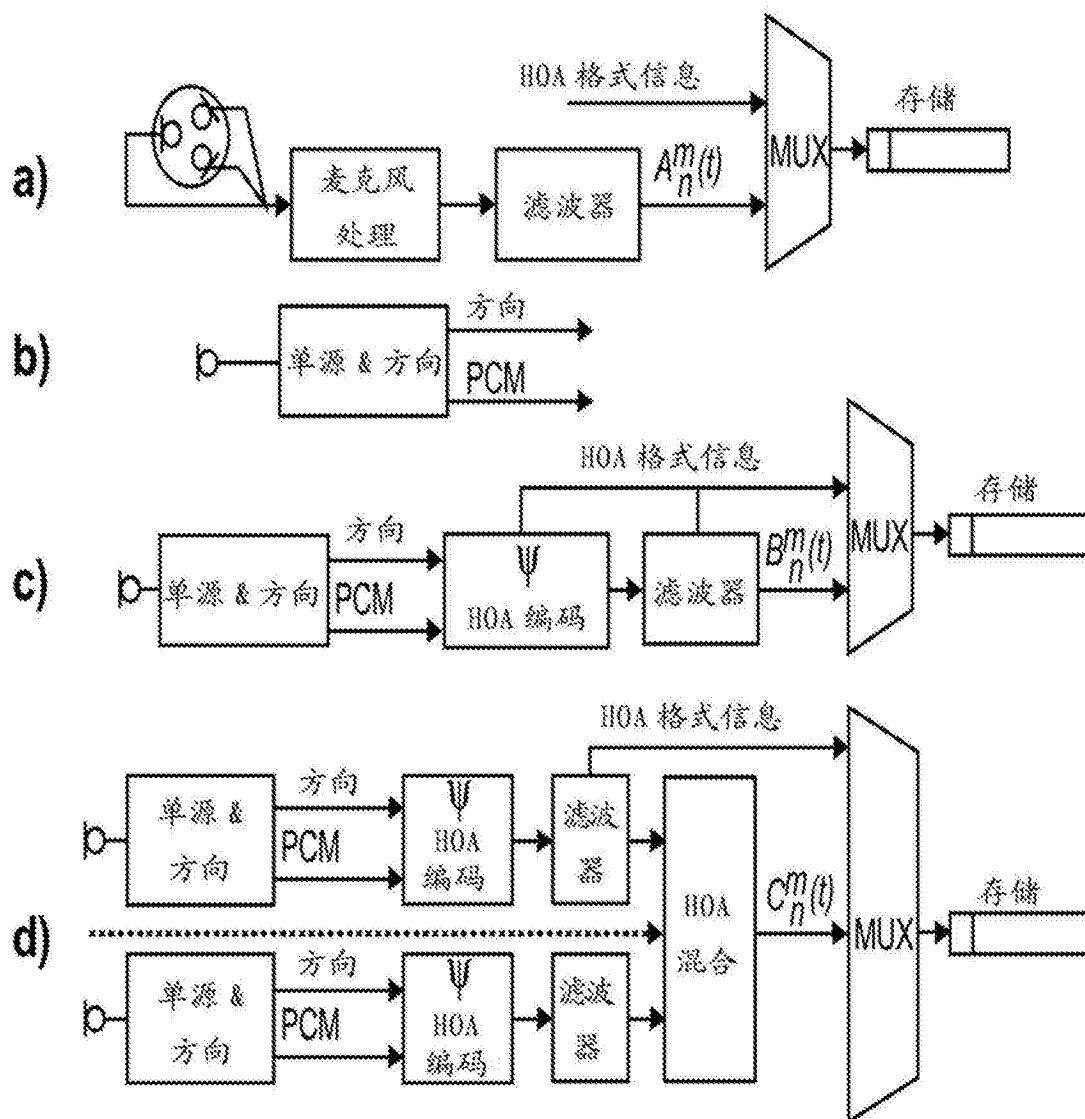


图 3

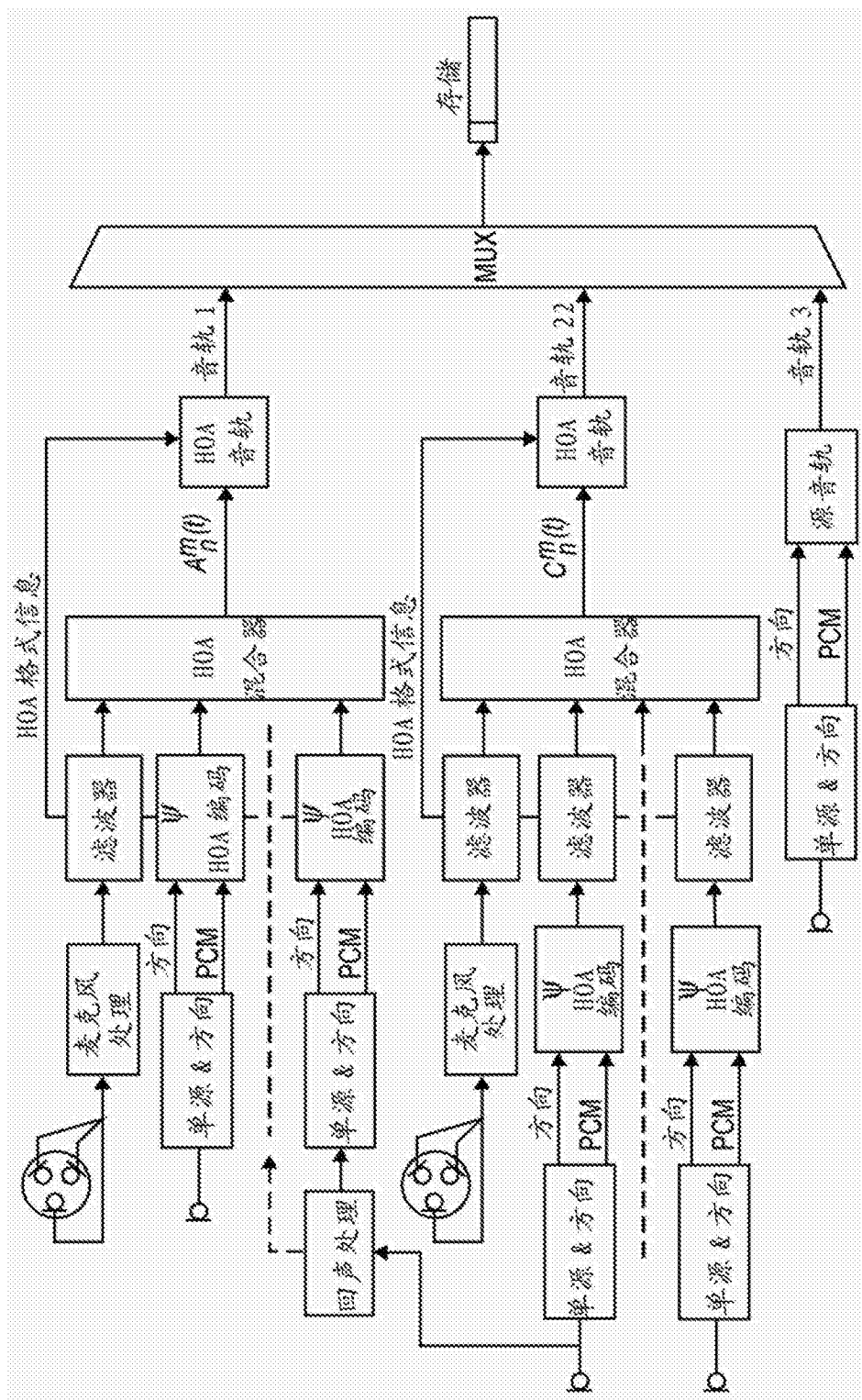


图 4

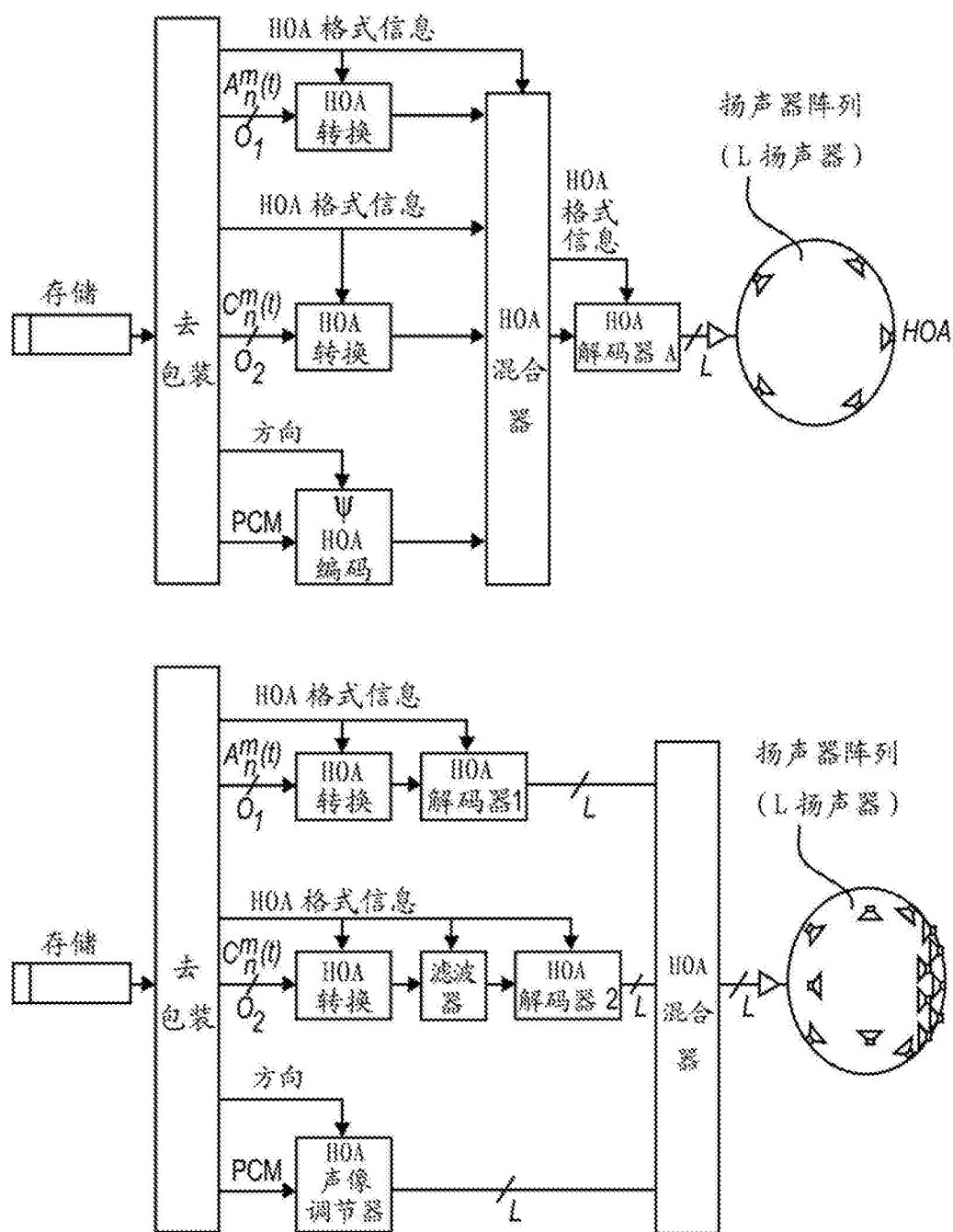


图 5

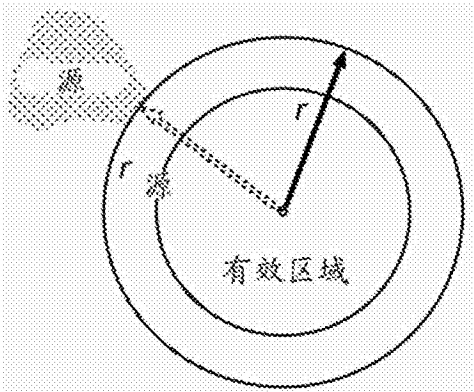


图 6

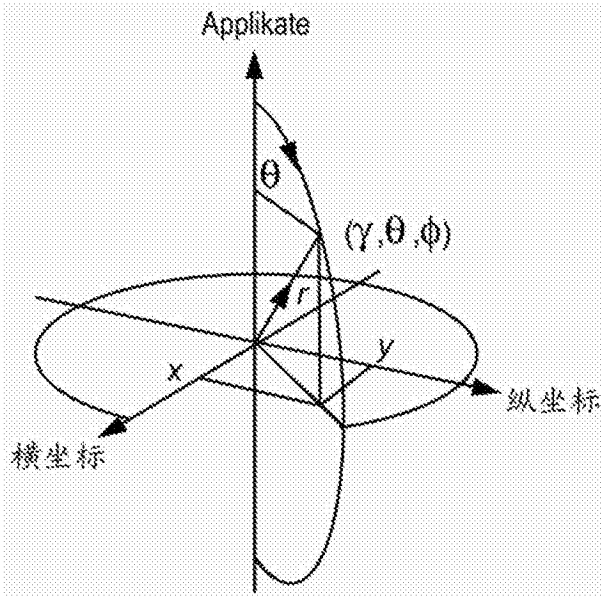


图 7

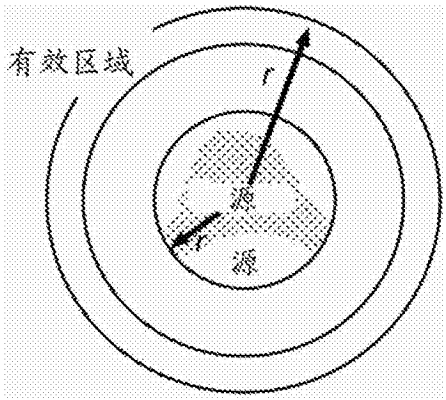


图 8

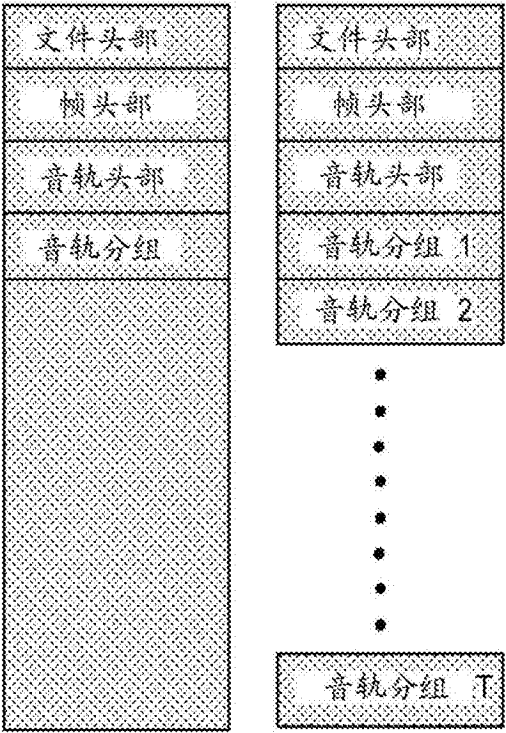


图 9

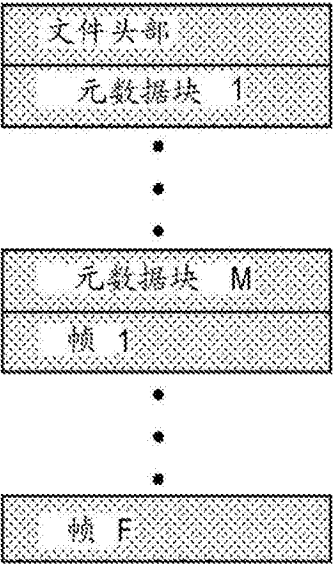


图 11

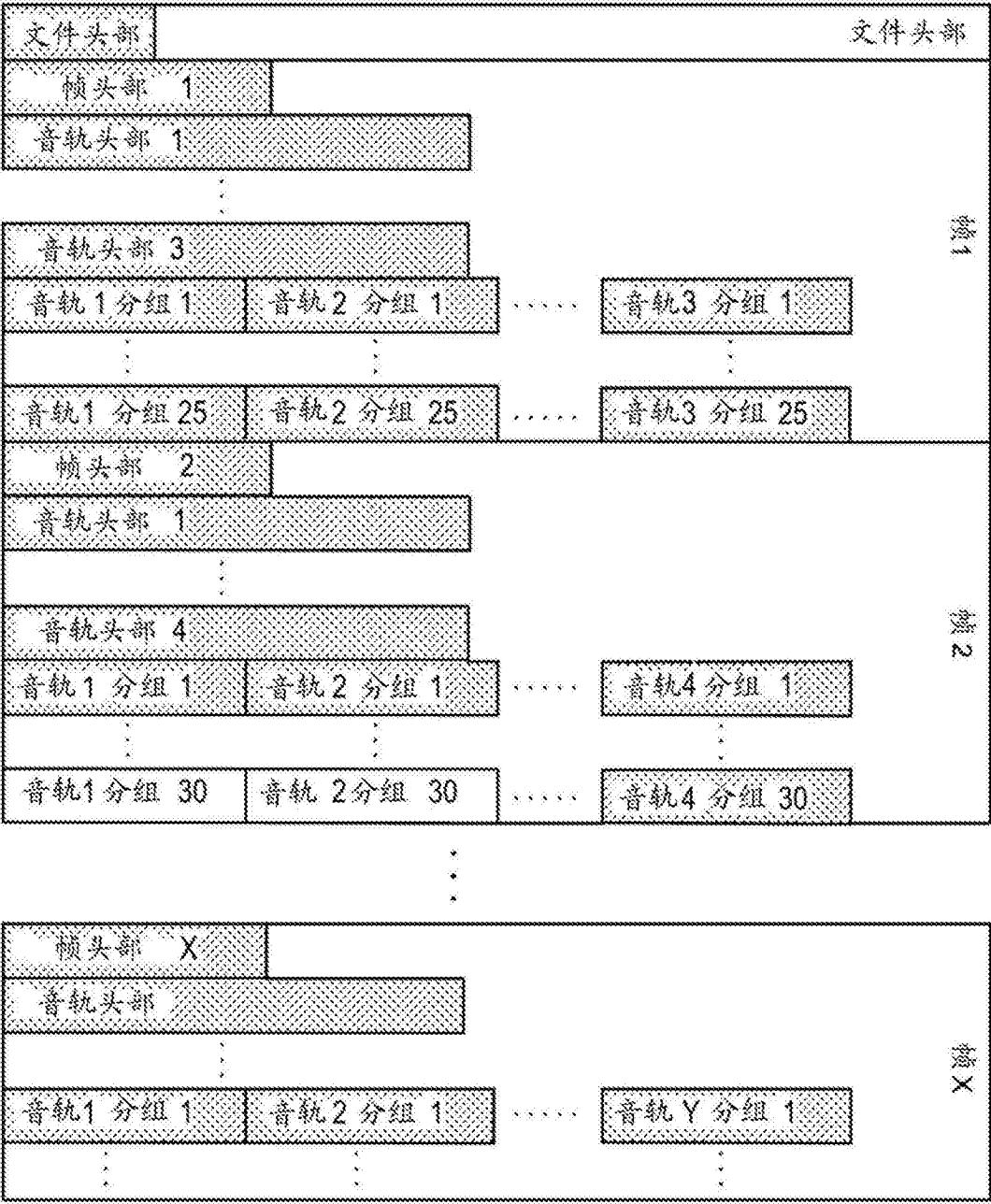


图 10

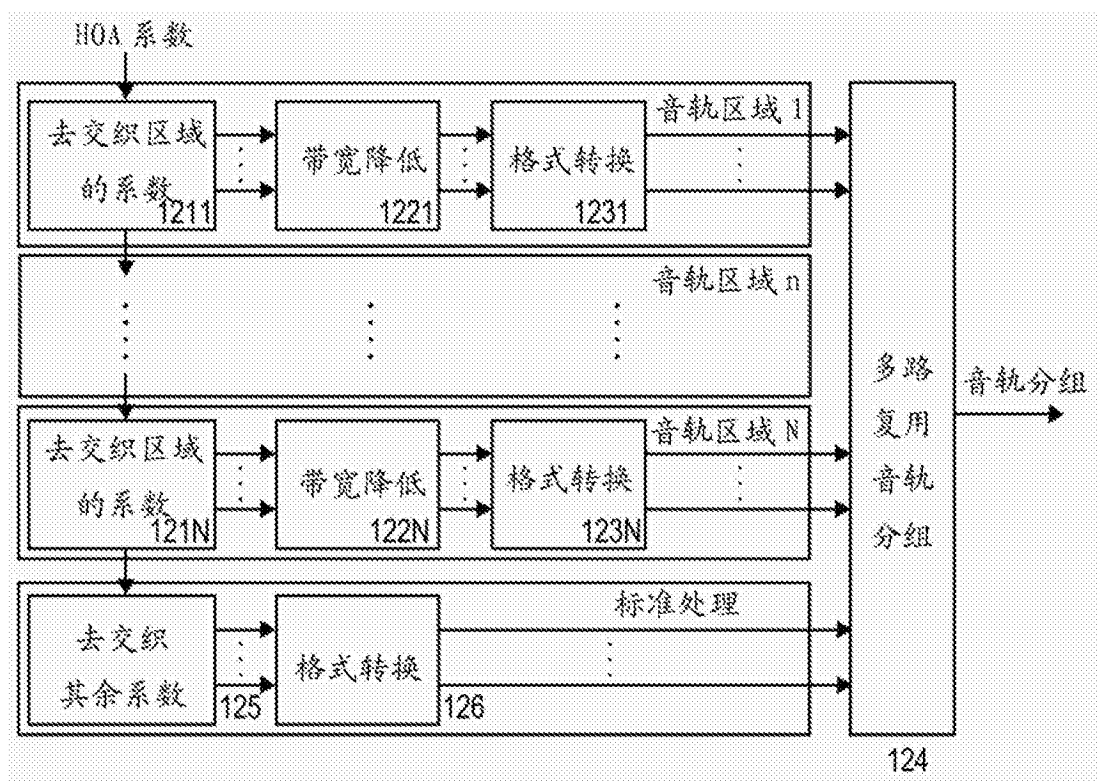


图 12

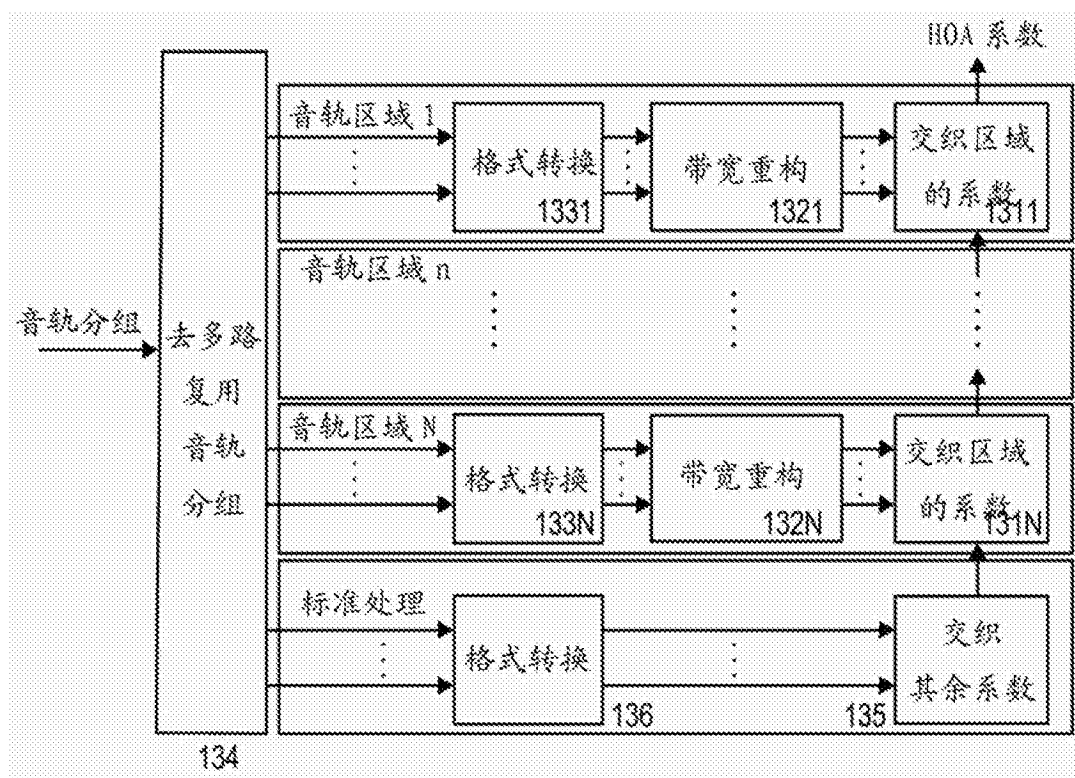


图 13

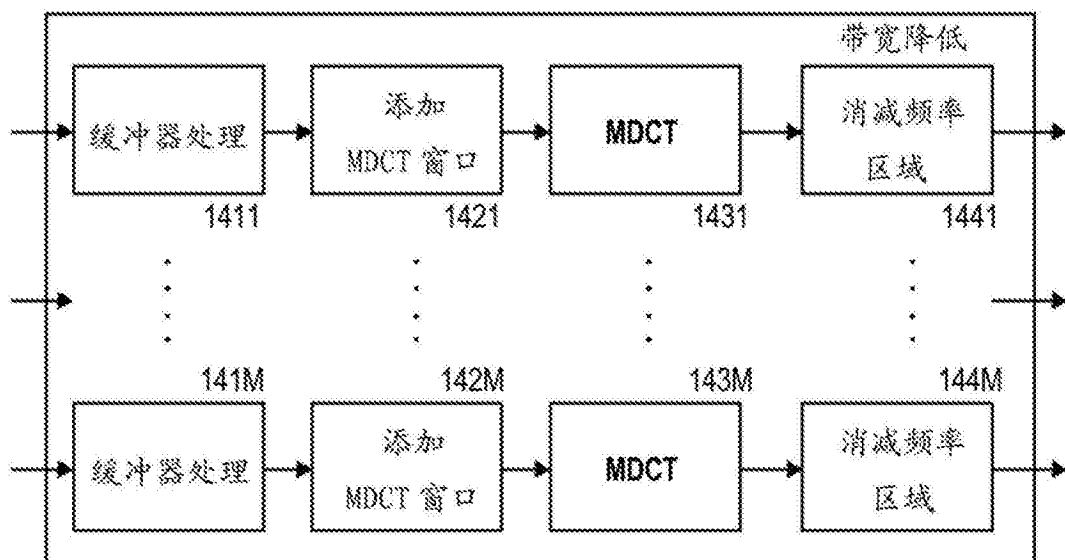


图 14

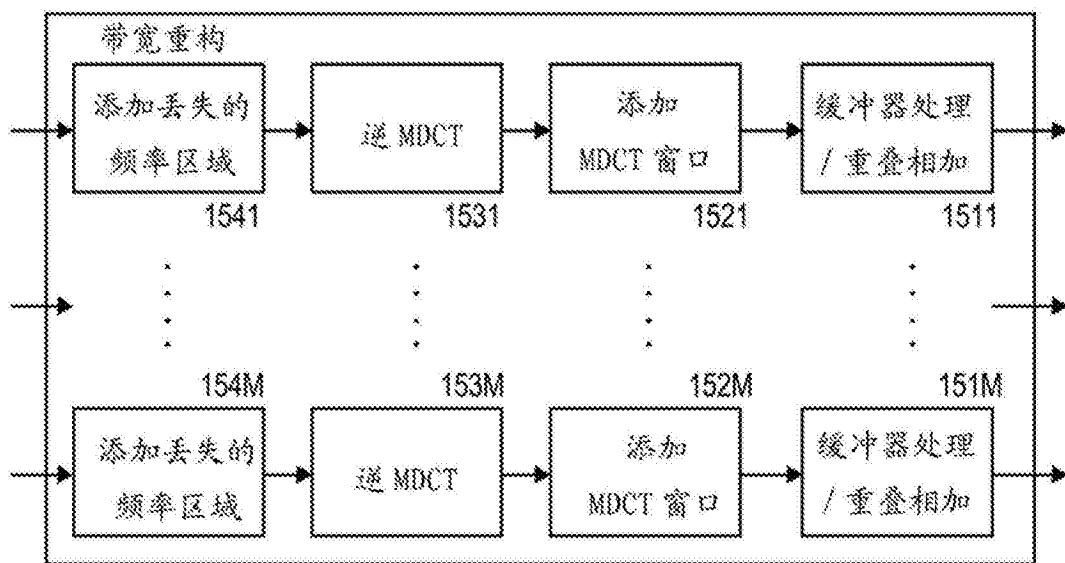


图 15