(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118824259 A
(43)申请公布日 2024.10.22

(21)申请号 202411189983.8

(22)申请日 2019.04.09

(30)优先权数据
62/655,990 2018.04.11 US

(62)分案原申请数据
201980013440.1 2019.04.09

(71)申请人 杜比国际公司
地址 荷兰阿姆斯特丹

(72)发明人 利昂·特连蒂夫
克里斯托弗·费尔施
丹尼尔·费希尔

(74)专利代理机构 北京律盟知识产权代理有限责任公司 11287
专利代理师 刘锋

(51)Int.Cl.
G10L 19/008(2013.01)
G10L 19/16(2013.01)
G10L 19/24(2013.01)
H04S 3/00(2006.01)
H04S 7/00(2006.01)

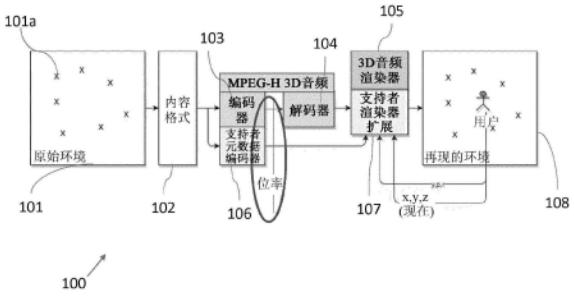
权利要求书1页 说明书16页 附图12页

(54)发明名称

用于6DOF音频渲染的方法、设备和系统及用于6DOF音频渲染的数据表示和位流结构

(57)摘要

本公开涉及用于6DOF音频渲染的方法、设备和系统及用于6DOF音频渲染的数据表示和位流结构。本公开涉及用于将音频信号编码到位流中(具体地说在编码器处)的方法、设备和系统,其包括:将与3DoF音频渲染关联的音频信号数据编码或包含到所述位流的一或多个第一位流部分中,以及将与6DoF音频渲染关联的元数据编码或包含到所述位流的一或多个第二位流部分中。本公开进一步涉及用于解码音频信号和基于所述位流的音频渲染的方法、设备和系统。



CN 118824259 A

1.一种用于在编码器处将音频信号编码到位流中的方法,所述方法包括:

将与3DoF音频渲染关联的音频信号数据编码或包含到所述位流的一或多个第一位流部分中;以及

将与6DoF音频渲染关联的元数据编码或包括到所述位流的一或多个第二位流部分中,其中所述方法进一步包含:

从一或多个音频源接收音频信号;

确定与距离衰减、遮挡和/或混响有关的环境特性和参数;

基于所述环境特性和所述参数来确定变换函数A的参数化并提供参数化的变换函数A,其中$A\ A^{-1}\approx 1$和$A^{-1}A\approx 1$;以及

通过使用所述变换函数A将来自所述一或多个音频源的所述音频信号变换为3DoF音频信号来生成与3DoF音频渲染关联的所述音频信号数据,其中

所述变换函数A将所述一或多个音频源的所述音频信号映射或投影到位于围绕默认3DoF听者位置的一或多个球体上的相应音频对象上。

用于6DOF音频渲染的方法、设备和系统及用于6DOF音频渲染的数据表示和位流结构

[0001] 分案申请的相关信息

[0002] 本案是申请日为2019年4月9日、申请号为201980013440.1、发明名称为“用于6DOF音频渲染的方法、设备和系统及用于6DOF音频渲染的数据表示和位流结构”的发明专利申请案的分案申请。

[0003] 相关申请案

[0004] 本申请要求2018年4月11日提交的美国临时申请序列号62/655,990的权益,该申请通过引用以其整体被并入本文中。

技术领域

[0005] 本公开涉及提供用于六自由度(6DoF)音频渲染的设备、系统和方法,具体地说,与用于6DoF音频渲染的数据表示和位流结构有关。

背景技术

[0006] 目前缺乏用于与用户的六自由度(6DoF)移动组合来渲染音频的适当解决方案。尽管存在用于与三自由度(3DoF)移动(偏航、俯仰、滚动)组合来渲染通道、对象和一阶/更高阶高保真度立体声响复制(HOA)信号的解决方案,但是缺乏对与用户的六自由度(6DoF)移动(偏航、俯仰、滚动和平移移动)组合来处置此类信号的支持。

[0007] 一般地,3DoF音频渲染提供声场,其中一或多个音频源在围绕预定听者位置(称为3DoF位置)的角位置处被渲染。3DoF音频渲染的一个示例被包含在MPEG-H 3D音频标准(缩写为MPEG-H 3DA)中。

[0008] 尽管MPEG-H 3DA被开发来支持用于3DoF的通道、对象和HOA信号,但是它还不能处置真正的6DoF音频。预想的MPEG-I 3D音频实现被期望以有效的方式(优选地包含有效的信号生成、编码、解码和/或渲染)将3DoF(和3DoF+)功能性向6DoF 3D音频设备扩展,同时优选地提供3DoF渲染向后兼容性。

[0009] 鉴于上述内容,本公开的目的是提供用于3D音频编码和/或3D音频渲染的方法、设备和数据表示和/或位流结构,其允许有效的6DoF音频编码和/或渲染,优选地带有用于例如根据MPEG-H 3DA标准的3DoF音频渲染的向后兼容性。

[0010] 本公开的另一个目的可以是提供用于3D音频编码和/或3D音频渲染的数据表示和/或位流结构,其允许有效的6DoF音频编码和/或渲染,优选地带有用于例如根据MPEG-H3DA标准的3DoF音频渲染的向后兼容性,以及用于有效的6DoF音频编码和/或渲染的编码和/或渲染设备,优选地带有用于例如根据MPEG-H 3DA标准的3DoF音频渲染的向后兼容性。

发明内容

[0011] 根据示例性方面,可以提供有一种用于将音频信号编码到位流中(具体地说在编码器处)的方法,该方法包括:将与3DoF音频渲染关联的音频信号数据编码和/或包含到位

流的一或多个第一位流部分中;和/或将与6DoF音频渲染关联的元数据编码和/或包含到位流的一或多个第二位流部分中。

[0012] 根据示例性方面,与3DoF音频渲染关联的音频信号数据包含一或多个音频对象的音频信号数据。

[0013] 根据示例性方面,一或多个音频对象位于围绕默认3DoF听者位置的一或多个球体上。

[0014] 根据示例性方面,与3DoF音频渲染关联的音频信号数据包含一或多个音频对象的方向数据和/或一或多个音频对象的距离数据。

[0015] 根据示例性方面,与6DoF音频渲染关联的元数据指示一或多个默认3DoF听者位置。

[0016] 根据示例性方面,与6DoF音频渲染关联的元数据包含或指示以下至少一个:6DoF空间的描述,任选地包含对象坐标;一或多个音频对象的音频对象方向;虚拟现实(VR)环境;和/或与距离衰减、遮挡和/或混响有关的参数。

[0017] 根据示例性方面,该方法可以进一步包含:从一或多个音频源接收音频信号;和/或基于来自一或多个音频源的音频信号和变换函数来生成与3DoF音频渲染关联的音频信号数据。

[0018] 根据示例性方面,通过使用变换函数将来自一或多个音频源的音频信号变换为3DoF音频信号来生成与3DoF音频渲染关联的音频信号数据。

[0019] 根据示例性方面,变换函数将一或多个音频源的音频信号映射或投影到位于围绕默认3DoF听者位置的一或多个球体上的相应音频对象上。

[0020] 根据示例性方面,该方法可以进一步包含:基于与距离衰减、遮挡和/或混响有关的环境特性和/或参数来确定变换函数的参数化。

[0021] 根据示例性方面,位流是MPEG-H 3D音频位流或使用MPEG-H 3D音频语法的位流。

[0022] 根据示例性方面,位流的一或多个第一位流部分表示位流的有效载荷,和/或一或多个第二位流部分表示位流的一或多个扩展容器。

[0023] 根据仍有的另一示例性方面,可以提供有一种用于解码和/或音频渲染(具体地说在解码器或音频渲染器处)的方法,该方法包括:接收位流,其在位流的一或多个第一位流部分中包含与3DoF音频渲染关联的音频信号数据,并且在位流的一或多个第二位流部分中进一步包含与6DoF音频渲染关联的元数据,和/或基于所接收的位流来执行3DoF音频渲染和6DoF音频渲染中的至少一个。

[0024] 根据示例性方面,在执行3DoF音频渲染时,基于在位流的一或多个第一位流部分中的与3DoF音频渲染关联的音频信号数据来执行3DoF音频渲染,同时丢弃在位流的一或多个第二位流部分中的与6DoF音频渲染关联的元数据。

[0025] 根据示例性方面,在执行6DoF音频渲染时,基于在位流的一或多个第一位流部分中的与3DoF音频渲染关联的音频信号数据和在位流的一或多个第二位流部分中的与6DoF音频渲染关联的元数据来执行6DoF音频渲染。

[0026] 根据示例性方面,与3DoF音频渲染关联的音频信号数据包含一或多个音频对象的音频信号数据。

[0027] 根据示例性方面,一或多个音频对象位于围绕默认3DoF听者位置的一或多个球体

上。

[0028] 根据示例性方面,与3DoF音频渲染关联的音频信号数据包含一或多个音频对象的方向数据和/或一或多个音频对象的距离数据。

[0029] 根据示例性方面,与6DoF音频渲染关联的元数据指示一或多个默认3DoF听者位置。

[0030] 根据示例性方面,与6DoF音频渲染关联的元数据包含或指示以下至少一个:6DoF空间的描述,任选地包含对象坐标;一或多个音频对象的音频对象方向;虚拟现实(VR)环境;和/或与距离衰减、遮挡和/或混响有关的参数。

[0031] 根据示例性方面,基于来自一或多个音频源的音频信号和变换函数来生成与3DoF音频渲染关联的音频信号数据。

[0032] 根据示例性方面,通过使用变换函数将来自一或多个音频源的音频信号变换为3DoF音频信号来生成与3DoF音频渲染关联的音频信号数据。

[0033] 根据示例性方面,变换函数将一或多个音频源的音频信号映射或投影到位于围绕默认3DoF听者位置的一或多个球体上的相应音频对象上。

[0034] 根据示例性方面,位流是MPEG-H 3D音频位流或使用MPEG-H 3D音频语法的位流。

[0035] 根据示例性方面,位流的一或多个第一位流部分表示位流的有效载荷,和/或一或多个第二位流部分表示位流的一或多个扩展容器。

[0036] 根据示例性方面,基于在位流的一或多个第一位流部分中与3DoF音频渲染关联的音频信号数据和在位流的一或多个第二位流部分中与6DoF音频渲染关联的元数据来执行6DoF音频渲染,其包含基于与3DoF音频渲染关联的音频信号数据和逆变换函数来生成与6DoF音频渲染关联的音频信号数据。

[0037] 根据示例性方面,通过使用逆变换函数和与6DoF音频渲染关联的元数据对与3DoF音频渲染关联的音频信号数据进行变换来生成与6DoF音频渲染关联的音频信号数据。

[0038] 根据示例性方面,逆变换函数是将一或多个音频源的音频信号映射或投影到位于围绕默认3DoF听者位置的一或多个球体上的相应音频对象上的变换函数的逆函数。

[0039] 根据示例性方面,基于与在位流的一或多个第一位流部分中的3DoF音频渲染关联的音频信号数据来执行3DoF音频渲染与在默认3DoF听者位置处,基于在位流的一或多个第一位流部分中的与3DoF音频渲染关联的音频信号数据和在位流的一或多个第二位流部分中的与6DoF音频渲染关联的元数据来执行6DoF音频渲染产生相同的生成的声场。

[0040] 根据仍有的另一示例性方面,可以提供有用于音频渲染的位流,该位流在位流的一或多个第一位流部分中包含与3DoF音频渲染关联的音频信号数据,并且在位流的一或多个第二位流部分中进一步包含与6DoF音频渲染关联的元数据。此方面可以与上述示例性方面中的任一项或多项组合。

[0041] 根据仍有的另一示例性方面,可以提供有一种包含处理器的设备,具体地说编码器,该处理器被配置成:将与3DoF音频渲染关联的音频信号数据编码和/或包含到位流的一或多个第一位流部分中;将与6DoF音频渲染关联的元数据编码和/或包含到位流的一或多个第二位流部分中。和/或输出编码的位流。此方面可以与上述示例性方面中的任一项或多项组合。

[0042] 根据仍有的另一示例性方面,可以提供有一种包含处理器的设备,具体地说解码

器或音频渲染器,该处理器被配置成:接收位流,其在位流的一或多个第一位流部分中包含与3DoF音频渲染关联的音频信号数据,并且在位流的一或多个第二位流部分中进一步包含与6DoF音频渲染关联的元数据,和/或基于所接收的位流来执行3DoF音频渲染和6DoF音频渲染中的至少一个。此方面可以与上述示例性方面中的任一项或多项组合。

[0043] 根据示例性方面,在执行3DoF音频渲染时,处理器被配置成基于在位流的一或多个第一位流部分中的与3DoF音频渲染关联的音频信号数据来执行3DoF音频渲染,同时丢弃在位流的一或多个第二位流部分中的与6DoF音频渲染关联的元数据。

[0044] 根据示例性方面,在执行6DoF音频渲染时,处理器被配置成基于在位流的一或多个第一位流部分中的与3DoF音频渲染关联的音频信号数据和在位流的一或多个第二位流部分中的与6DoF音频渲染关联的元数据来执行6DoF音频渲染。

[0045] 根据仍有的另一示例性方面,可以提供有一种包含指令的非暂时性计算机程序产品,所述指令在由处理器执行时促使处理器执行用于将音频信号编码到位流中(具体地说在编码器处)的方法,该方法包括:将与3DoF音频渲染关联的音频信号数据编码或包含到位流的一或多个第一位流部分中;和/或将与6DoF音频渲染关联的元数据编码或包含到位流的一或多个第二位流部分中。此方面可以与上述示例性方面中的任一项或多项组合。

[0046] 根据仍有的另一示例性方面,可以提供有一种包含指令的非暂时性计算机程序产品,所述指令在由处理器执行时促使处理器执行用于解码和/或音频渲染(具体地说在解码器或音频渲染器处)的方法,该方法包括:接收位流,其在位流的一或多个第一位流部分中包含与3DoF音频渲染关联的音频信号数据,并且在位流的一或多个第二位流部分中进一步包含与6DoF音频渲染关联的元数据,和/或基于所接收的位流来执行3DoF音频渲染和6DoF音频渲染中的至少一个。此方面可以与上述示例性方面中的任一项或多项组合。

[0047] 本公开的其它方面涉及对应计算机程序和计算机可读存储介质。

[0048] 将领会的是,方法步骤和设备特征可以以多种方式互换。具体地说,如本领域技术人员将领会的,所公开的方法的细节可以被实现为适于执行该方法的一些或全部步骤的设备,且反之亦然。具体地说,理解的是,关于所述方法所作的相应陈述同样地适用于对应设备,且反之亦然。

附图说明

[0049] 下面参考附图解释本公开的示例实施例,其中相同的附图标记可以指示相似或类似的元件,并且其中:

[0050] 图1示意性地图示了根据本公开的示例性方面的包含MPEG-H 3D音频解码器/编码器接口的示例性系统。

[0051] 图2示意性地图示了房间(6DoF空间)的6DoF场景的示例性顶视图。

[0052] 图3示意性地图示了根据本公开的示例性方面的图2的6DoF场景的示例性顶视图和3DoF音频数据和6DoF扩展元数据。

[0053] 图4A意性地图示了根据本公开的示例性方面的用于处理3DoF、6DoF和音频数据的示例性系统。

[0054] 图4B意性地图示了根据本公开的示例性方面的用于6DoF音频渲染和3DoF音频渲染的示例性解码和渲染方法。

[0055] 图5示意性地图示了根据图2至4B中的一或多个,在系统中的3DoF位置处的6DoF音频渲染和3DoF音频渲染的示例性匹配条件。

[0056] 图6A意性地图示了根据本公开的示例性方面的示例性数据表示和/或位流结构。

[0057] 图6B意性地图示了根据本公开的示例性方面的基于图6A的数据表示和/或位流结构的示例性3DoF音频渲染。

[0058] 图6C意性地图示了根据本公开的示例性方面的基于图6A的数据表示和/或位流结构的示例性6DoF音频渲染。

[0059] 图7A意性地图示了根据本公开的示例性方面的基于3DoF音频信号数据的6DoF音频编码变换A。

[0060] 图7B意性地图示了根据本公开的示例性方面的用于基于3DoF音频信号数据来近似/恢复6DoF音频信号数据的6DoF音频解码变换A^{-1}。

[0061] 图7C意性地图示了根据本公开的示例性方面的基于图7B的近似/恢复的6DoF音频信号数据的示例性6DoF音频渲染。

[0062] 图8示意性地图示了根据本公开的示例性方面的3DoF/6DoF位流编码的方法的示例性流程图。

[0063] 图9示意性地图示了根据本公开的示例性方面的3DoF和/或6DoF音频渲染的方法的示例性流程图。

具体实施方式

[0064] 在下文中,将参考附图更详细地描述优选的示例性方面。不同附图和实施例中相同或类似的特征可以用类似的附图标记表示。要理解的是,下面与各种优选示例性方面有关的详细描述不意味着限制本发明的范围。

[0065] 在本文中使用时,“MPEG-H 3D音频”将指如在ISO/IEC 23008-3和/或ISO/IEC 23008-3标准的任何过去和/或将来的修改、版次或其的其它版本中标准化的规范。

[0066] 在本文中使用时,MPEG-I 3D音频实现被期望将3DoF(和3DoF+)功能性向6DoF 3D音频扩展,同时优选地提供3DoF渲染向后兼容性。

[0067] 在本文中使用时,3DoF通常是能够正确处置通过三个参数(例如偏航、俯仰、滚动)指定的用户的头部移动(具体地说头部旋转)的系统。此类系统经常在各种游戏系统中可用,诸如虚拟现实(VR)/增强现实(AR)/混合现实(MR)系统,或其它此类类型的声环境。

[0068] 在本文中使用时,6DoF通常是能够正确处置3DoF和平移移动的系统。

[0069] 本公开的示例性方面涉及音频系统(例如,与MPEG-I音频标准兼容的音频系统),其中通过将有关的元数据转换为3DoF格式,诸如与MPEG标准(例如,MPEG-H 3DA标准)兼容的音频渲染器输入格式,音频渲染器将功能性向6DoF扩展。

[0070] 图1图示了示例性系统100,其被配置成除现有3DoF系统之外还使用元数据扩展和/或音频渲染器扩展,以便使得6DoF体验能实现。系统100包含原始环境101(其可以示例性地包含一或多个音频源101a)、内容格式102(例如,包含3D音频数据的位流)、编码器103和所提议的元数据编码器扩展106。系统100还可以包含3D音频渲染器105(例如3DoF渲染器)和支持者渲染器扩展107(例如,用于再现的环境108的6DoF渲染器扩展)。

[0071] 在通过3DoF的3D音频渲染的方法中,仅在预定3DoF位置处的用户的角取向的角度

(例如偏航角y、俯仰角p,滚动角r)可以被输入到3DoF音频渲染器105。通过扩展的6DoF功能性,用户的位置坐标(例如x、y和z)可以另外被输入到6DoF音频渲染器(扩展渲染器)。

[0072] 本公开的优点包含用于在编码器与解码器之间传送的位流的位速率改进。位流可以按照例如MPEG-I音频标准和/或MPEG-H 3D音频标准的标准来编码和/或解码,或者至少向后兼容诸如MPEG-H 3D音频标准的标准。

[0073] 在一些示例中,本公开的示例性方面涉及与多个系统兼容的单个位流(例如,MPEG-H 3D音频(3DA)位流(BS)或使用MPEG-H 3DA BS的语法的位流)的处理。

[0074] 例如,在一些示例性方面中,音频位流可以与两个或两个以上不同渲染器兼容,例如,可以与一个标准(例如,MPEG-H 3D音频标准)兼容的3DoF音频渲染器和可以与第二不同标准(例如,MPEG-I音频标准)兼容的新定义的6DoF音频渲染器或渲染器扩展。

[0075] 本公开的示例性方面涉及被配置成执行相同音频位流的解码和渲染的不同解码器,优选地以便产生相同音频输出。

[0076] 例如,本公开的示例性方面涉及3DoF解码器和/或3DoF渲染器和/或6DoF解码器和/或6DoF渲染器,其被配置成为相同的位流(例如,3DA BS或使用3DA BS的位流)产生相同的输出。示例性地,位流可以包含关于在VR/AR/MR(虚拟现实/增强现实/混合现实)空间中听者的定义位置的信息,例如作为6DoF元数据的一部分。

[0077] 本公开示例性地进一步涉及被配置成分别编码和/或解码6DoF信息(例如,与MPEG-I音频环境兼容)的编码器和/或解码器,其中本公开的此类编码器和/或解码器提供以下优点中的一或多个:

[0078] • VR/AR/MR有关的音频数据的质量和位率有效表示及其封装成音频位流语法(例如,MPEG-H 3D音频BS);

[0079] • 各种系统(例如,MPEG-H 3DA标准和预想的MPEG-I音频标准)之间的向后兼容性。

[0080] 为了优选地避免在3DoF与6DoF解决方案之间的竞争和提供在当前与未来技术之间的平滑过渡,向后兼容性是非常有益的。

[0081] 例如,在3DoF音频系统与6DoF音频系统之间的向后兼容性可以是非常有益的,诸如在诸如MPEG-I音频的6DoF音频系统中提供对诸如MPEG-H 3D音频的3DoF音频系统的向后兼容性。

[0082] 根据本公开的示例性方面,这能够通过例如在位流级别上为6DoF有关的系统提供向后兼容性来实现,所述6DoF有关的系统包括:

[0083] • 3DoF音频材料编码的数据和有关的元数据;以及

[0084] • 6DoF有关的元数据。

[0085] 本公开的示例性方面涉及标准3DoF位流语法,诸如第一类型的音频位流(例如,MPEG-H 3DA BS)语法,其例如在第一类型的音频位流(例如,MPEG-H 3DA BS)的一或多个扩展容器中封装6DoF位流元素,诸如MPEG-I音频位流元素。

[0086] 为了提供在性能级别上确保向后兼容性的系统,以下系统和/或结构可以是相关的并且可以出现:

[0087] 1a.3DoF系统(例如,与MPEG-H 3DA的标准兼容的系统)应当能够忽略所有6DoF有关的语法元素(例如,基于MPEG-H 3D音频位流语法的“mpegh3daExtElementConfig()”或

“mpegh3daExtElement()”的功能性而忽略MPEG-I音频位流语法元素),即,3DoF系统(解码器/渲染器)可以优选地被配置成忽略另外的6DoF有关的数据和/或元数据(例如,通过不读取6DoF有关的数据和/或元数据);以及

[0088] 2a.位流有效载荷的剩余部分(例如,含有与MPEG-H 3DA位流解析器兼容的数据和/或元数据的MPEG-I音频位流有效载荷)应是可由3DoF系统(例如,遗留MPEG-H 3DA系统)解码的,以便产生期望的音频输出,即,3DoF系统(解码器/渲染器)可以优选地被配置成解码BS的3DoF部分;以及

[0089] 3a.6DoF系统(例如,MPEG-I音频系统)应能够处理音频位流的3DoF有关的部分和6DoF有关的部分,并且在VR/AR/MR空间中在预定义的向后兼容的(一或多个)3DoF位置处产生与3DoF系统(例如,MPEG-H 3DA系统)的音频输出相匹配的音频输出,即,6DoF系统(解码器/渲染器)可以优选地被配置成在默认的(一或多个)3DoF位置处渲染与3DoF渲染的声场/音频输出相匹配的声场/音频输出;以及

[0090] 4a.6DoF系统(例如,MPEG-I音频系统)应围绕预定的向后兼容的(一或多个)3DoF位置提供音频输出的平滑变化(过渡)(即,在6DoF空间中提供连续的声场),即,6DoF系统(解码器/渲染器)可以优选地被配置成在默认的(一或多个)3DoF位置的周围渲染在默认的(一或多个)3DoF位置处平滑地过渡到3DoF渲染的声场/音频输出的声场/音频输出。

[0091] 在一些示例中,本公开涉及提供6DoF音频渲染器(例如,MPEG-I音频渲染器),其在一个、多个或一些3DoF位置中产生与3DoF音频渲染器(例如,MPEG-H 3D音频渲染器)相同的音频输出。

[0092] 目前,在直接将3DoF有关的音频信号和元数据直接传输到6DoF音频系统时存在缺陷,其包括:

[0093] 1.位率增大(即,除6DoF有关的音频信号和元数据之外还发送3DoF有关的音频信号和元数据);以及

[0094] 2.有限的有效性(即,(一或多个)3DoF有关的音频信号和元数据仅对(一或多个)3DoF位置有效)。

[0095] 本公开的示例性方面涉及克服上述缺点。

[0096] 在一些示例中,本公开涉及:

[0097] 1.使用3DoF兼容的(一或多个)音频信号和元数据(例如,与MPEG-H 3D音频兼容的信号和元数据)而不是原始音频源信号和元数据(或作为原始音频源信号和元数据的补充);和/或

[0098] 2.将适用性范围(对于6DoF渲染的使用)从(一或多个)3DoF位置增大到6DoF空间(由内容创建者定义),同时保持高级别的声场近似。

[0099] 本公开的示例性方面涉及有效地生成,编码,解码和渲染(一或多个)此类信号,以便实现这些目标和提供6DoF渲染功能性。

[0100] 图2图示了示例性房间202的示例性顶视图201。如图2中所示,示例性听者站在带有几个音频源和非平凡墙壁几何形状的房间的中间。在6DoF设备(例如,提供用于6DoF能力的系统)中,示例性听者能够四处移动,但是在一些示例中假设默认3DoF位置206可以对应于最佳VR/AR/MR音频体验的预期区域(例如,根据内容创建者的设置或意图)。

[0101] 具体地说,图2示例性地图示了墙壁203、6DoF空间204、示例性(可选的)方向性向

量205(例如,如果一或多个声源定向地发射声音)、3DoF听者位置206(默认3DoF位置206)和在图2中示例性地图示为星形的音频源207。

[0102] 图3图示了例如如在图2中的示例性6DoF VR/AR/MR场景,以及在3DoF音频位流302(例如,诸如MPEG-H 3D音频位流)和扩展容器303中含有的音频对象(音频数据+元数据)320。可以经由与MPEG标准(例如,MPEG-H或MPEG-I)兼容的设备或系统(例如,软件、硬件或经由云)对音频位流302和扩展容器303进行编码。

[0103] 本公开的示例性方面涉及在使用6DoF音频渲染器(例如,MPEG-I音频渲染器)时,以对应于3DoF音频渲染器(例如,MPEG-H音频渲染器)输出信号(可以与物理定律声音传播一致或不一致)的方式在“3DoF位置”中重建声场。此声场应优选地是基于原始的“音频源”,并且反映对应VR/AR/MR环境的复杂几何形状的影响(例如,“墙壁”、结构、声音反射、混响和/或遮挡等的效应)。

[0104] 本公开的示例性方面涉及由编码器以确保满足上述对应要求(1a)-(4a)的一个、多个或优选地全部的方式对描述该场景的所有相关信息进行参数化。

[0105] 如果并行地运行两个音频渲染模式(即,3DoF和6DoF)并且将内插算法应用于6DoF空间中的对应输出,则此类方案将是次优的,因为它将要求:

[0106] • 并行执行两个截然不同的渲染算法(即,一个用于特定的3DoF位置,一个用于6DoF空间);

[0107] • 大量音频数据(用于传输3DoF音频渲染器的另外音频数据)。

[0108] 本公开的示例性方面避免了上述缺点,因为优选地仅执行单个音频渲染模式(例如,而不是并行执行两个音频渲染模式)和/或3DoF音频数据优选地被用于带有用于恢复和/或近似(一或多个)原始声源信号的另外元数据的6DoF音频渲染(例如,而不是传送3DoF音频数据和(一或多个)原始声源数据)。

[0109] 本公开的示例性方面涉及(1)单个6DoF音频渲染算法(例如,与MPEG-I音频兼容),其优选地在(一或多个)特定位置处产生与3DoF音频渲染算法(例如,与MPEG-H 3DA兼容)完全相同的输出,和/或(2)表示音频(例如,3DoF音频数据)和6DoF有关的音频元数据,以最小化6DoF音频位流数据(例如,MPEG-I音频位流数据)的3DoF和VR/AR/MR有关的部分中的冗余。

[0110] 本公开的示例性方面涉及使用第一标准化格式位流(例如,MPEG-H 3DA BS)语法来封装第二标准化格式位流(例如,未来标准,例如,MPEG-I)或其部分和6DoF有关的元数据,以:

[0111] • 传输(例如,在3DoF音频位流语法的核心部分中)优选地由3DoF音频系统解码的音频源信号和元数据,所述3DoF音频系统优选地在(默认)(一或多个)3DoF位置中足够好地近似期望声场;以及

[0112] • 传输(例如,在3DoF音频位流语法的扩展部分中)被用来近似(恢复)用于6DoF音频渲染的原始音频源信号的6DoF有的关元数据和/或其它数据(例如,参数或/和信号数据)。

[0113] 本发明的一方面涉及确定期望的“(一或多个)3DoF位置”和在编码器侧3DoF音频系统(例如,MPEG-H 3DA系统)兼容信号。

[0114] 例如,如相对于图3所示,由于一些3DoF系统(例如MPEG-H 3DA系统)不能解决VR/

AR/MR环境效应(例如遮挡、混响等),因此,用于3DA的虚拟3DA对象信号可以在特定3DoF位置中产生相同的声场(基于信号x_{3DA}),其应优选地含有用于(一或多个)特定3DoF位置的VR环境的效应("湿"信号)。图3中所图示的方法和过程可以经由多种系统和/或产品来执行。

[0115] 在一些示例性方面,逆函数A^{-1}应优选地"不湿"(即,去除VR环境的效应),这些信号应是好的,如近似原始"干"信号x(其没有VR环境的效应)所必需的。

[0116] 可以优选地定义用于3DoF渲染的(一或多个)音频信号((x_{3DA})),以便例如基于以下内容为3DoF和6DoF音频渲染二者提供相同/类似的输出:

[0117] $F_{3DoF}(x_{3DA}) \rightarrow F_{6DoF}(x)\text{for 3DoF}$ 公式编号(1)

[0118] 音频对象可以被包含在标准化位流中。此位流可以按照诸如MPEG-H 3DA和/或MPEG-I的多种标准来编码。

[0119] BS可以包含关于对象信号、对象方向和对象距离的信息。

[0120] 图3进一步示例性地图示了扩展容器303,其可以例如在BS中含有扩展元数据。Bs的扩展容器303可以包含以下元数据中的至少一个:(i)3DoF(默认)位置参数;(ii)6DoF空间描述参数(对象坐标);(iii)(可选的)对象方向性参数;(iv)(可选的)VR/AR/MR环境参数;和/或(v)(可选的)距离衰减参数、遮挡参数和/或混响参数等。

[0121] 基于以下内容,可以包含有期望的音频渲染的近似:

[0122] $F_{6DoF}(x^{*}) \approx F_{6DoF}(x)\text{for 6DoF}$ 公式编号(2)

[0123] 该近似可以是基于VR环境,其中环境特性可以被包含在扩展容器元数据中。

[0124] 另外或任选地,优选地基于以下内容,可以提供用于6DoF音频渲染器(例如,MPEG-I音频渲染器)输出的平滑度:

[0125] $F_{6DoF} \subset G^{i\geq 0}\ for\ 3DoF+,\quad G^{i\geq 0}$ - 几何连续性类 公式编号(3)

[0126] 本公开的示例性方面涉及在编码器侧上定义3DoF音频对象(例如,MPEG-H 3DA对象),优选地基于:

[0127] $x_{3DA}:=A(x), ||F_{3DoF}(x_{3DA}) - F_{6DoF}(x)\text{for 3DoF}|| \rightarrow \min$ 公式编号(4)

[0128] 本公开的一方面涉及基于以下内容的在解码器上原始对象的恢复:

[0129] $x^{*}:=A^{-1}(x_{3DA})$ 公式编号(5)

[0130] 其中,x涉及声源/对象信号,x^{*}涉及声源/对象信号的近似,F(x)for 3DoF/for 6DoF涉及用于(一或多个)3DoF/6DoF听者位置的音频渲染函数,3DoF涉及(一或多个)给定参考兼容性位置∈6DoF空间;6DoF涉及(一或多个)任意允许位置∈VR场景;

[0131] • $F_{6FoF}(x)$涉及解码器指定的6DoF音频渲染(例如,MPEG-I音频渲染);

[0132] • $F_{3DoF}(x_{3DA})$涉及解码器指定的3DoF渲染(例如,MPEG-H 3DA渲染);以及

[0133] • A、A^{-1}涉及基于信号x及其逆(A^{-1})来近似信号x_{3DA}的函数(A)。

[0134] 优选地,以对应于3DoF音频渲染器输出信号的方式,"3DoF位置"中使用6DoF音频渲染器来重建近似的声源/对象信号。

[0135] 优选地基于声场来近似声源/对象信号,该声场是基于原始"音频源"并且反映对应VR/AR/MR环境(例如,"墙壁"、结构、混响、遮挡等)的复杂几何形状的影响。

[0136] 即,用于3DA的虚拟3DA对象信号优选地在特定3DoF位置中产生相同的声场(基于信号x_{3DA}),其含有用于(一或多个)特定3DoF位置的VR环境的效应。

[0137] 以下内容可以在渲染侧(例如,对于遵从标准(诸如MPEG-H或MPEG-I标准)的解码

器)可用:

[0138] • 用于3DoF音频渲染的(一或多个)音频信号:x_{3DA}

[0139] • 3DoF或6DoF音频渲染功能性:

[0140] $F_{3DoF}(x_{3DA})$或$F_{6DoF}(x)$ 公式编号(6)

[0141] 对于6DoF音频渲染,另外在渲染侧可以有6DoF元数据可用于6DoF音频渲染功能性(例如,基于3DoF音频信号x_{3DA}和6DoF元数据来近似/恢复一或多个音频源的音频信号x)。

[0142] 本公开的示例性方面涉及(i)3DoF音频对象(例如,MPEG-H 3DA对象)的定义和/或(ii)原始音频对象的恢复(近似)。

[0143] 音频对象可以示例性地被包含在3DoF音频位流(例如MPEG-H 3DA BS)中。

[0144] 位流可以包含关于对象音频信号、对象方向和/或对象距离的信息。

[0145] 扩展容器(例如,诸如MPEG-H 3DA Bs的位流的扩展容器)可以包含以下元数据中的至少一个:(i)3DoF(默认)位置参数;(ii)6DoF空间描述参数(对象坐标);(iii)(可选的)对象方向性参数;(iv)(可选的)VR/AR/MR环境参数;和/或(v)(可选的)距离衰减参数、遮挡参数、混响参数等。

[0146] 本公开可以提供以下优点:

[0147] • 向后兼容3DoF音频解码和渲染(例如,MPEG-H 3DA解码和渲染):6DoF音频渲染器(例如MPEG-I音频渲染器)输出对应于用于(一或多个)预定3DoF位置的3DoF渲染引擎(例如MPEG-H 3DA渲染引擎的3DoF渲染输出。

[0148] • 编码效率:对于此方案,能够有效地再使用遗留3DoF音频位流语法(例如MPEG-H 3DA位流语法)结构。

[0149] • 在(一或多个)预定(3DoF)位置处的音频质量控制:对于(一或多个)任何任意位置和对应6DoF空间,编码器能够明确地确保最佳感知音频质量。

[0150] 本公开的示例性方面可以涉及与MPEG标准(例如MPEG-I标准)位流兼容的格式的以下信令:

[0151] • 经由扩展容器机制(例如,MPEG-H 3DA BS)的隐式3DoF音频系统(例如MPEG-H3DA)兼容性信令,其使得6DoF音频(例如,MPEG-I音频兼容)处理算法能够恢复原始音频对象信号。

[0152] • 参数化,描述用于原始音频对象信号的近似的数据。

[0153] 6DoF音频渲染器可以指定如何在例如MPEG兼容系统(例如,MPEG-I音频系统)中恢复原始音频对象信号。

[0154] 此提议的概念:

[0155] • 关于近似函数(即A(x))的定义是通用的;

[0156] • 能够是任意复杂的,但在解码器侧应存在对应近似(即$\exists A^{-1}$);

[0157] • 近似地在数学上是"明确定义的"(例如在算法上稳定等);

[0158] • 在近似函数(即A(x))的类型方面是通用的;

[0159] • 近似函数可以是基于以下近似类型或这些方案的任意组合(以位率消耗增大的顺序列出):

[0160] -为信号x_{3DA}应用的(一或多个)参数化音频效应(例如参数控制的级别、混响、反射、遮挡等)

[0161] -(一或多个)参数编码的修改(例如用于传送的信号x_{3DA}的时间/频率变化修改增益)

[0162] -(一或多个)信号编码修改(例如近似残余波形($x-x_{3DA}$)的编码的信号);以及

[0163] •可扩展并适用于通用声场和声源表示(及其组合):对象、通道、FOA、HOA。

[0164] 图6A示意性地图示了根据本公开的示例性方面的示例性数据表示和/或位流结构。可以经由与MPEG标准(例如,MPEG-H或MPEG-I)兼容的设备或系统(例如,软件、硬件或经由云)已对数据表示和/或位流结构进行编码。

[0165] 位流BS示例性地包含第一位流部分302,其包含3DoF编码的音频数据(例如在位流的主要部分或核心部分中)。优选地,位流BS的位流语法与诸如MPEG-H 3DA位流语法的3DoF音频渲染的BS语法兼容或相符。3DoF编码的音频数据可以作为有效载荷被包含在位流BS的一或多个分组中。

[0166] 如前面例如结合上面图3所描述的,3DoF编码的音频数据可以包含一或多个音频对象的音频对象信号(例如,在默认3DoF位置周围的球体上)。对于定向音频对象,3DoF编码的音频数据可以进一步任选地包含对象方向,和/或任选地进一步指示对象距离(例如通过使用增益和/或一或多个衰减参数)。

[0167] 示例性地,BS示例性地包含第二位流部分303,其包含用于6DoF音频编码的6DoF元数据(例如在位流的元数据部分或扩展部分中)。优选地,位流BS的位流语法与诸如MPEG-H 3DA位流语法的3DoF音频渲染的BS语法兼容或相符。6DoF元数据可以作为扩展元数据被包含在位流BS的一或多个分组中(例如,在例如已经由MPEG-H 3DA位流结构提供的一或多个扩展容器中)。

[0168] 如前面例如结合上面图3所描述的,6DoF元数据可以包含一或多个3DoF(默认)位置的位置数据(例如坐标),进一步任选地包含6DoF空间描述(例如对象坐标),进一步任选地包含对象方向性,进一步任选地包含描述和/或参数化VR环境的元数据,和/或进一步任选地包含关于衰减、遮挡和/或混响等的参数化信息和/或参数。

[0169] 图6B示意性地图示了根据本公开的示例性方面的基于图6A的数据表示和/或位流结构的示例性3DoF音频渲染。如图6A中所示,可以经由与MPEG标准(例如,MPEG-H或MPEG-I)兼容的设备或系统(例如,软件、硬件或经由云)已对数据表示和/或位流结构进行编码。

[0170] 具体地,在图6B中示例性地图示了3DoF音频渲染可以由可以丢弃6DoF元数据的3DoF音频渲染器实现,以仅基于从第一位流部分302获得的3DoF编码的音频数据来执行3DoF音频渲染。即,例如,在MPEG-H 3DA向后兼容性的情况下,MPEG-H 3DA渲染器能够有效且可靠地忽略/丢弃位流的扩展部分(例如,(一或多个)扩展容器)中的6DoF元数据,以便仅基于从第一位流部分302获得的3DoF编码的音频数据来执行有效的常规MPEG-H 3DA 3DoF(或3DoF+)音频渲染。

[0171] 图6C示意性地图示了根据本公开的示例性方面的基于图6A的数据表示和/或位流结构的示例性6DoF音频渲染。如图6A中所示,可以经由与MPEG标准(例如,MPEG-H或MPEG-I)兼容的设备或系统(例如,软件、硬件或经由云)已对数据表示和/或位流结构进行编码。

[0172] 具体地,在图6C中示例性地图示了6DoF音频渲染可以由使用从第一位流部分302获得的3DoF编码的音频数据及从第二位流部分303获得的6DoF元数据的新颖的6DoF音频渲染器(例如,根据MPEG-I或以后的标准)实现,以基于从第一位流部分302获得的3DoF编码的

音频数据和从第二位流部分303获得的6DoF元数据来执行6DoF音频渲染。

[0173] 因此,在位流中没有或至少带有降低的冗余的情况下,相同的位流能够由遗留3DoF音频渲染器(其允许简单且有益的向后兼容性)用于3DoF音频渲染和由新颖的6DoF音频渲染器用于6DoF音频渲染。

[0174] 图7A示意性地图示了根据本公开的示例性方面的基于3DoF音频信号数据的6DoF音频编码变换A。可以根据与MPEG标准(例如,MPEG-H或MPEG-I)兼容的方法、过程、设备或系统(例如,软件、硬件或经由云)来执行变换(以及任何逆变换)。

[0175] 示例性地,类似于上面的图2和3,图7A示出了房间的示例性顶视图202,包含示例性地多个音频源207(其可以位于墙壁203之后,或者其声音信号可以被其它结构阻挡,这可以导致衰减、混响和/或遮挡效应)。

[0176] 为实现3DoF音频渲染目的,对多个音频源207的音频信号x进行变换,以便在围绕默认3DoF位置206(例如,3DoF声场中的听者位置)的球体S上获得3DoF音频信号(音频对象)。如上所述,3DoF音频信号被称为x_{3DA},并且可以通过使用变换函数A来获得,使得:

[0177] $x_{3DA}=A(x)$ 公式编号(6)

[0178] 在上述表达式中,x表示声源/对象信号,x_{3DA}表示用于在默认3DoF位置206中产生相同声场的3DA的虚拟3DA对象信号,以及A表示基于音频信号x近似音频信号x_{3DA}的变换函数。逆变换函数A^{-1}可以被用来恢复/近似用于6DoF音频渲染的声源信号,如已经在上面和下面进一步所讨论的。注意,$A\ A^{-1}=1$,并且$A^{-1}A=1$,或者至少$A\ A^{-1}\approx1$和$A^{-1}A\approx1$。

[0179] 一般地,在本公开的一些示例性方面中,变换函数A可以被认为是将音频信号x投影或至少映射到围绕默认3DoF位置206的球体S上的映射/投影函数。

[0180] 要进一步注意的是,3DoF音频渲染不知道VR环境(诸如现有的墙壁203等,或可以导致衰减、混响、遮挡效应等的其它结构)。因此,变换函数A可以优选地包含基于此类VR环境特性的效应。

[0181] 图7B示意性地图示了根据本公开的示例性方面的用于基于3DoF音频信号数据来近似/恢复6DoF音频信号数据的6DoF音频解码变换A^{-1}。

[0182] 通过使用逆变换函数A^{-1}和如上面图7A中获得的近似的3DoF音频信号x_{3DA},原始音频源207的原始音频信号x*能够被恢复/近似为:

[0183] $x^*=A^{-1}(x_{3DA})$。 公式编号(7)

[0184] 因此,图7B中的音频对象320的音频信号x*可以被恢复为与原始源207的音频信号x类似或相同,特别是在与原始源207相同的位置。

[0185] 图7C示意性地图示了根据本公开的示例性方面的基于图7B的近似/恢复的6DoF音频信号数据的示例性6DoF音频渲染。

[0186] 图7B中的音频对象320的音频信号x*随后能够被用于6DoF音频渲染,其中听者的位置也变得可变。

[0187] 在假设听者的听者位置是在位置206(与默认3DoF位置相同的位置)处时,6DoF音频渲染与基于音频信号x_{3DA}的3DoF音频渲染是渲染相同的声场。

[0188] 因此,在作为假设的听者位置的默认3DoF位置处的6DoF渲染$F_{6DoF}(x^*)$等于(或至少近似等于)3DoF渲染$F_{3DoF}(x_{3DA})$。

[0189] 此外,如果听者位置被移位,例如到图7C中的位置206',则在6DoF音频渲染中生成

的声场变得不同,但优选地可以平滑地出现。

[0190] 作为另一示例,可以假设第三听者位置206″,并且在6DoF音频渲染中生成的声场特别是对于左上音频信号变得不同,该左上音频信号对于第三听者位置206″未被墙壁203阻挡。优选地,这变得可能,因为逆函数A^{-1}恢复原始声源(没有诸如VR环境特性的环境效应)。

[0191] 图8示意性地图示了根据本公开的示例性方面的3DoF/6DoF位流编码的方法的示例性流程图。要注意的是,步骤的顺序是非限制性的,并且可以根据情况改变。而且,要注意的是,方法的一些步骤是可选的。方法例如可以由解码器、音频解码器、音频/视频解码器或解码器系统执行。

[0192] 在步骤S801中,方法(例如在解码器侧)接收一或多个音频源的(一或多个)原始音频信号x。

[0193] 在步骤S802中,方法(任选地)确定环境特性(诸如房间形状、墙壁、墙壁声音反射特性、对象、障碍物等)和/或确定参数(参数化效应,诸如衰减、增益、遮挡、混响等)。

[0194] 在步骤S803中,方法例如基于步骤S802的结果来(任选地)确定变换函数A的参数化。优选地,步骤S803提供参数化或预设的变换函数A。

[0195] 在步骤S804中,方法基于变换函数A将一或多个音频源的(一或多个)原始音频信号x变换为对应一或多个近似的3DoF音频信号x_{3DA}。

[0196] 在步骤S805中,方法确定6DoF元数据(其可以包含一或多个3DoF位置、VR环境信息和/或诸如衰减、增益、遮挡、混响等环境效应的参数和参数化)。

[0197] 在步骤S806中,方法将(一或多个)3DoF音频信号x_{3DA}到包含(嵌入)第一位流部分(或多个第一位流部分)中。

[0198] 在步骤S807中,方法将6DoF元数据包含(嵌入)到第二位流部分(或多个第二位流部分)中。

[0199] 随后,在步骤S808中,方法继续基于第一和第二位流部分来对位流进行编码,以提供包含第一位流部分(或多个第一位流部分)中的(一或多个)3DoF音频信号x_{3DA}和第二位流部分(或多个第二位流部分)中的6DoF元数据的编码的位流。

[0200] 随后,能够将编码的位流提供到3DoF解码器/渲染器以便仅基于第一位流部分(或多个第一位流部分)中的(一或多个)3DoF音频信号x_{3DA}进行3DoF音频渲染,或者到6DoF解码器/渲染器以便基于第一位流部分(或多个第一位流部分)中的(一或多个)3DoF音频信号x_{3DA}和第二位流部分(或多个第二位流部分)中的6DoF元数据进行6DoF音频渲染。

[0201] 图9示意性地图示了根据本公开的示例性方面的3DoF和/或6DoF音频渲染的方法的示例性流程图。要注意的是,步骤的顺序是非限制性的,并且可以根据情况改变。而且,要注意的是,方法的一些步骤是可选的。方法例如可以由编码器、渲染器、音频编码器、音频渲染器、音频/视频编码器或编码器系统或渲染器系统执行。

[0202] 在步骤S901中,接收包含第一位流部分(或多个第一位流部分)中的(一或多个)3DoF音频信号x_{3DA}和第二位流部分(或多个第二位流部分)中的6DoF元数据的编码的位流。

[0203] 在步骤S902中,从第一位流部分(或多个第一位流部分)获得(一或多个)3DoF音频信号x_{3DA}。这能够由3DoF解码器/渲染器以及也由6DoF解码器/渲染器完成。

[0204] 如果解码器/渲染器是用于3DoF音频渲染目的的遗留设备(或切换到3DoF音频渲

染模式的新3DoF/6DoF解码器/渲染器),则方法通过步骤S903继续,其中6DoF元数据被丢弃/忽略,并且随后继续到3DoF音频渲染操作,以基于从第一位流部分(或多个第一位流部分)获得的(一或多个)3DoF音频信号x_{3DA}来渲染3DoF音频。

[0205] 即,有利地保证了向后兼容性。

[0206] 另一方面,如果解码器/渲染器是用于6DoF音频渲染目的(诸如新的6DoF解码器/渲染器或切换到6DoF音频渲染模式的3DoF/6DoF解码器/渲染器),则方法通过步骤S905继续以从(一或多个)第二位流部分获得6Dof元数据。

[0207] 在步骤S906中,方法基于从第二位流部分(或多个第二位流部分)获得的6DoF元数据和逆变换函数A^{-1},从获得的来自第一位流部分(或多个第一位流部分)的(一或多个)3DoF音频信号x_{3DA}近似/恢复音频对象/源的音频信号x*。

[0208] 随后,在步骤S907中,方法继续以基于音频对象/源的近似/恢复的音频信号x*和基于听者位置(其在VR环境内可以是可变的)来执行6DoF音频渲染。

[0209] 在上述示例性方面中,能够提供有用于3D音频编码和/或3D音频渲染的有效且可靠的方法、设备和数据表示和/或位流结构,其允许有效的6DoF音频编码和/或渲染,有利地带有例如根据MPEG-H 3DA标准的用于3DoF音频渲染的向后兼容性。具体地,可能提供用于3D音频编码和/或3D音频渲染的数据表示和/或位流结构,其允许有效的6DoF音频编码和/或渲染,优选地带有例如根据MPEG-H 3DA标准的用于3DoF音频渲染的向后兼容性,以及提供用于有效的6DoF音频编码和/或渲染的对应编码和/或渲染设备,带有例如根据MPEG-H 3DA标准的用于3DoF音频渲染的向后兼容性。

[0210] 本文中所描述的方法和系统可以被实现为软件、固件和/或硬件。某些组件可以被实现为在数字信号处理器或微处理器上运行的软件。其它组件可以被实现为硬件和/或专用集成电路。在所描述的方法和系统中遇到的信号可以诸如随机存取存储器或光学存储介质的介质上。它们可以经由诸如无线电网络、卫星网络、无线网络或有线网络(例如因特网)的网络来传送。利用本文中所描述的方法和系统的典型装置是被用来存储和/或渲染音频信号的便携式电子装置或其它消费者设备。

[0211] 根据本公开的方法和设备的示例实现将通过以下列举的不是权利要求的示例实施例(EEE)而变得明显。

[0212] EEE1示例性地涉及一种用于对包括音频源信号、3DoF有关的数据和6DoF有关的数据的音频进行编码的方法,该方法包括:例如由诸如具体地说编码器的音频源设备对近似(一或多个)3DoF位置中的期望声场的音频源信号进行编码以确定3DoF数据;和/或例如由诸如具体地说编码器的音频源设备对6DoF有关的数据进行编码以确定6DoF元数据,其中元数据可以被用来近似用于6DoF渲染的原始音频源信号。

[0213] EEE2示例性地涉及EEE1的方法,其中3DoF数据涉及对象音频信号、对象方向和对象距离中的至少一个。

[0214] EEE3示例性地涉及EEE1或EEE2的方法,其中6DoF数据涉及以下的至少一个:3DoF(默认)位置参数、6DoF空间描述(对象坐标)参数、对象方向性参数、VR环境参数、距离衰减参数、遮挡参数和混响参数。

[0215] EEE4示例性地涉及一种用于传输数据,具体地说3DoF和6DoF可渲染音频数据的方法,该方法包括:例如在音频位流语法中传输音频源信号,该音频源信号例如在由3DoF音频

系统解码时优选地可以近似在(一或多个)3DoF位置中的期望声场;和/或例如在音频位流语法的扩展部分中传输6DoF有关的元数据以便近似和/或恢复用于6DoF渲染的原始音频源信号;其中6DoF有关的元数据可以是参数数据和/或信号数据。

[0216] EEE5示例性地涉及EEE4的方法,其中例如包含3DoF元数据和/或6DoF元数据的音频位流语法符合MPEG-H音频标准的至少一个版本。

[0217] EEE6示例性地涉及一种用于生成位流的方法,该方法包括:确定3DoF元数据,所述3DoF元数据是基于近似(一或多个)3DoF位置中的期望声场的音频源信号;确定6DoF有关的元数据,其中所述元数据可以被用来近似用于6DoF渲染的原始音频源信号;和/或将音频源信号和6DoF有关的元数据插入到位流中。

[0218] EEE7示例性地涉及一种用于音频渲染的方法,所述方法包括:

[0219] (一或多个)3DoF位置中的原始音频信号x的近似的音频信号x^*的6DoF元数据的预处理,其中6DoF渲染可以提供与用于3DoF渲染的所传输的音频源信号x_{3DA}的3DoF渲染相同的输出,该3DoF渲染近似在(一或多个)3DoF位置中的期望声场。

[0220] EEE8示例性地涉及EEE7的方法,其中基于以下内容来确定音频渲染:

[0221] $F_{6DoF}(x^*) \approx F_{3DoF}(x_{3DA}) \rightarrow F_{6DoF}(x)\text{for 3DoF}$

[0222] 其中$F_{6DoF}(x^*)$涉及用于(一或多个)6DoF听者位置的音频渲染功能,$F_{3DoF}(x_{3DA})$涉及用于(一或多个)3DoF听者位置的音频渲染功能,x_{3DA}是含有用于(一或多个)特定3DoF位置的VR环境的效应的音频信号,以及x^*涉及近似的音频信号。

[0223] EEE9示例性地涉及EEE8的方法,其中原始音频信号x的近似的音频信号x^*是基于:

[0224] $x^* := A^{-1}(x_{3DA})$

[0225] 其中A^{-1}涉及近似函数A的逆。

[0226] EEE10示例性地涉及EEE8或EEE9的方法,其中用来使用近似方法A获得原始音频源信号x的近似的音频信号x^*的元数据是基于以下内容来定义:

[0227] $x_{3DA} := A(x), ||F_{3DoF}(x_{3DA}) - F_{6DoF}(x)\text{for 3DoF}|| \rightarrow \min$

[0228] 其中,元数据的量小于传输原始音频源信号x所需的音频数据的量。

[0229] 其中,基于以下内容来确定音频渲染:

[0230] $F_{6DoF}(x^*) \approx F_{3DoF}(x_{3DA}) \rightarrow F_{6DoF}(x)\text{for 3DoF}$

[0231] 其中$F_{6DoF}(x^*)$涉及用于(一或多个)6DoF听者位置的音频渲染功能,$F_{3DoF}(x_{3DA})$涉及用于(一或多个)3DoF听者位置的音频渲染功能,x_{3DA}是含有用于(一或多个)特定3DoF位置的VR环境的效应的音频信号,以及x^*涉及近似的音频信号。

[0232] 本发明的示例性方面和实施例可以在硬件、固件或软件或二者的组合中被实现(例如,作为可编程逻辑阵列)。除非另外指明,否则作为本公开的一部分包含的算法或过程并不固有地与任何具体计算机或其它设备有关。具体地说,各种通用机器可以与根据本文中的教导编写的程序一起使用,或者可以更方便的是构造更专用的设备(例如,集成电路)来执行所要求的方法步骤。因此,本公开可以在一或多个可编程计算机系统上执行的一或多个计算机程序中被实现(例如,附图的元件的任一个的实现),每个可编程计算机系统包括至少一个处理器、至少一个数据存储系统(包含易失性和非易失性存储器和/或存储元件)、至少一个输入装置或端口及至少一个输出装置或端口。程序代码被应用于输入数据以执行本文中所描述的功能并生成输出信息。输出信息以已知的方式被应用于一或多个输出

装置。

[0233] 每个此类程序可以用任何期望的计算机语言(包含机器、汇编或高级过程、逻辑或面向对象的编程语言)来实现以与计算机系统进行通信。在任何情况下,语言可以是编译或解释语言。

[0234] 例如,在由计算机软件指令序列实现时,本公开的实施例的各种功能和步骤可以由在适当的数字信号处理硬件中运行的多线程软件指令序列实现,在此情况下,实施例的各种装置、步骤和功能可以对应于软件指令的部分。

[0235] 每个此类计算机程序优选地被存储在或下载到通用或专用可编程计算机可读的存储介质或装置(例如,固态存储器或介质,或磁性或光学介质),以便在存储介质或装置由计算机系统读取时配置和操作计算机以执行本文中所描述的规程。本发明的系统还可以被实现为计算机可读存储介质,其配置有(即存储)计算机程序,其中如此配置的存储介质促使计算机系统以特定和预定义的方式进行操作以执行本文中所描述的功能。

[0236] 上面描述了本公开的本发明的多个示例性方面和示例性实施例。然而,将理解的是,在不脱离本公开的本发明的精神和范围的情况下,可以进行各种修改。根据上述教导,本发明的许多修改和变化是可能的。要理解的是,在所附权利要求的范围内,本公开的本发明可以以不同于本文中具体描述的方式来实践。

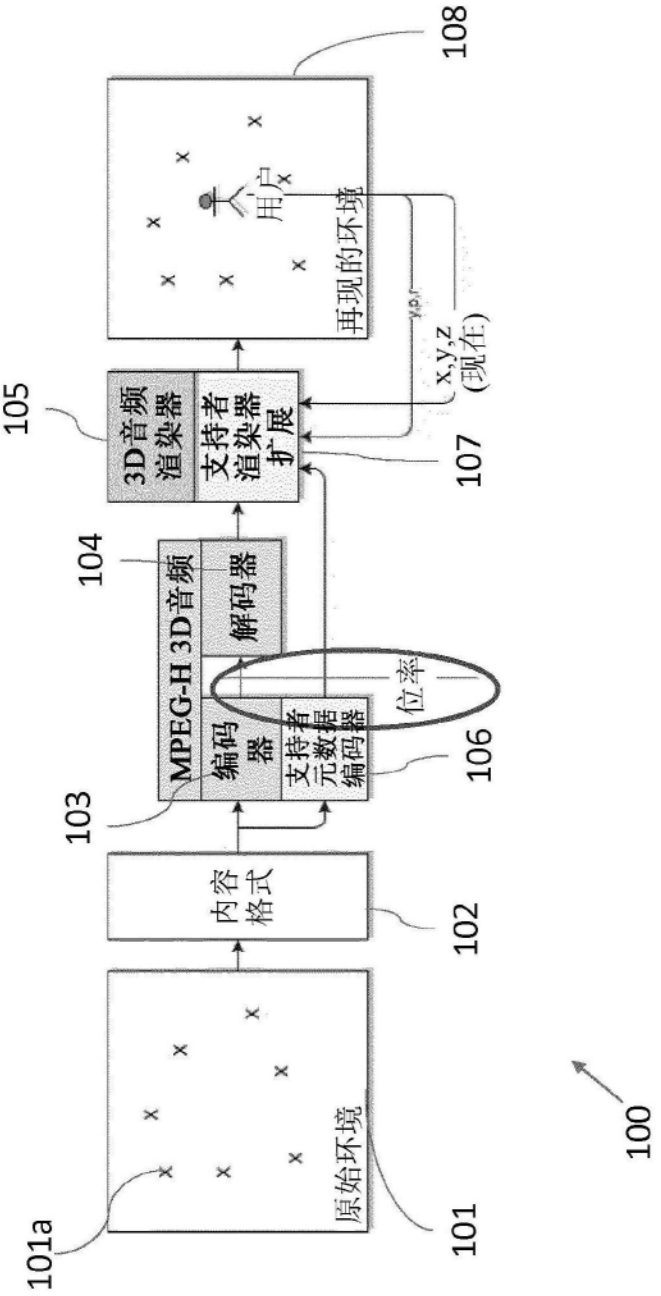

101a
原始环境
101
内容格式
102
MPEG-H 3D音频
编码器
解码器
支持者元数据编码器
103
104
位率
106
3D音频渲染器
支持者渲染器扩展
105
107
x,y,z
(现在)
再现的环境
用户
108
100

图1

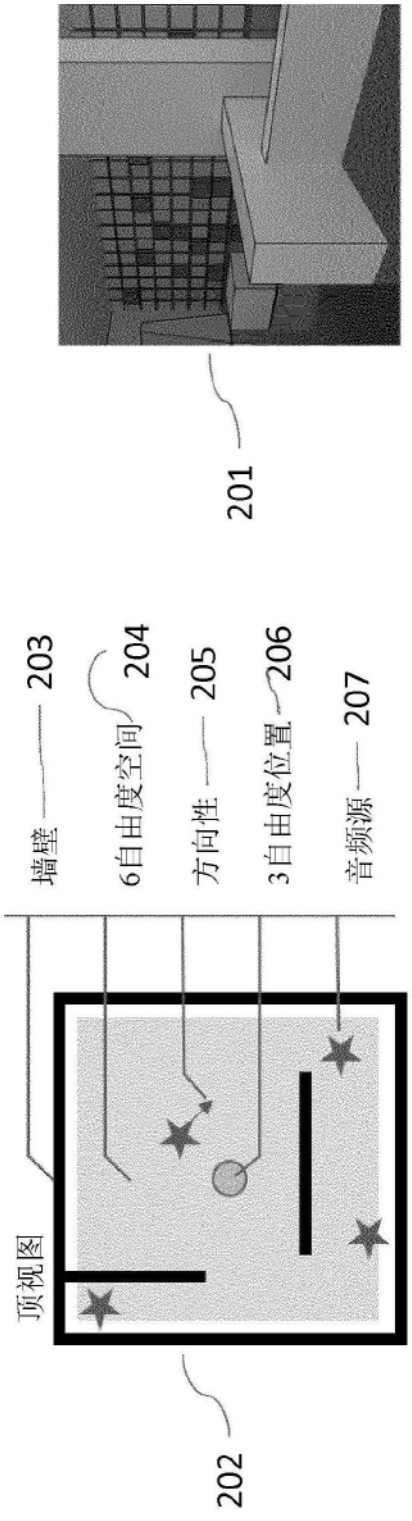


图2

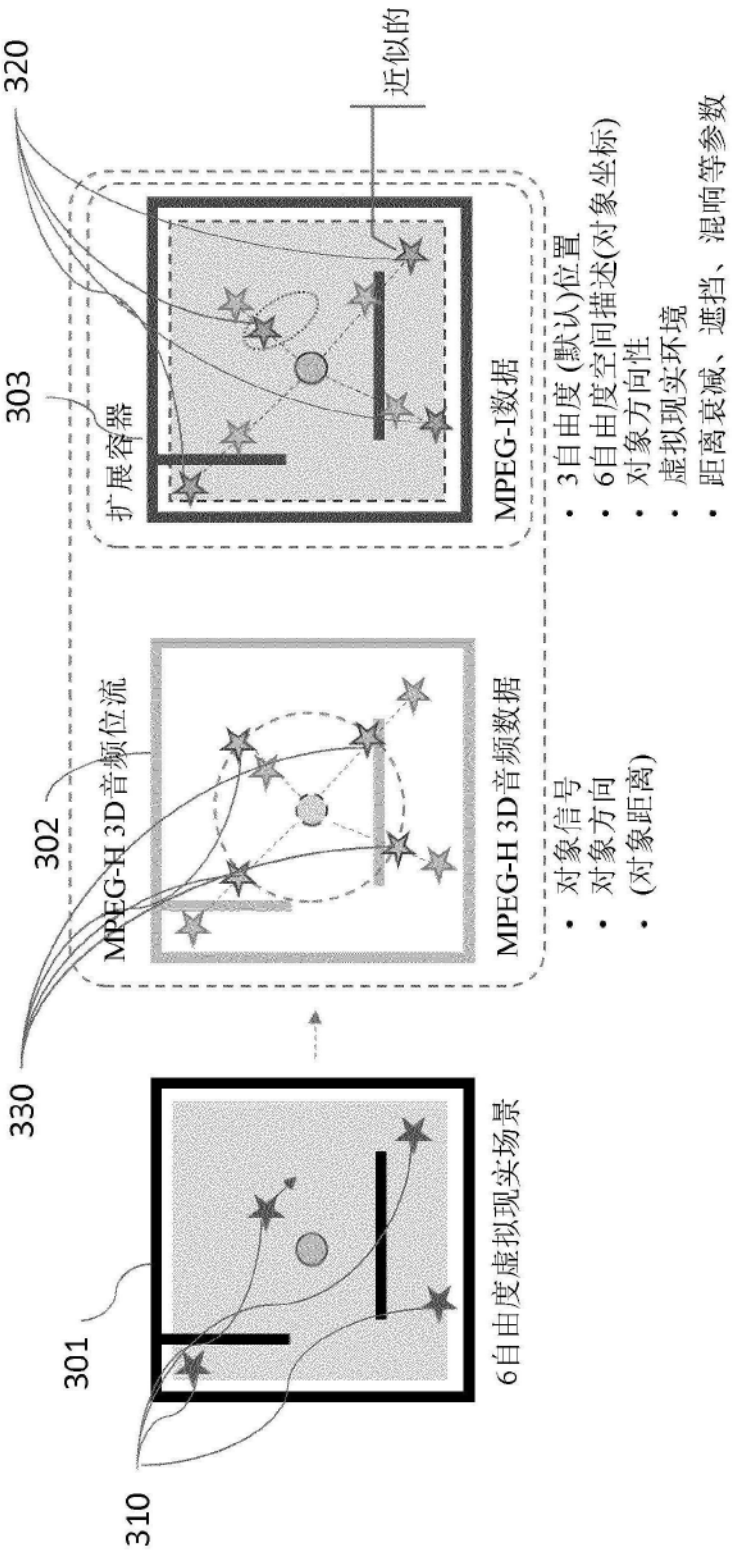

320
303
扩展容器
近似的
MPEG-I数据
• 3自由度 (默认)位置
• 6自由度空间描述(对象坐标)
• 对象方向性
• 虚拟现实环境
• 距离衰减、遮挡、混响等参数
302
MPEG-H 3D音频位流
MPEG-H 3D音频数据
• 对象信号
• 对象方向
• (对象距离)
330
301
6自由度虚拟现实场景
310

图3

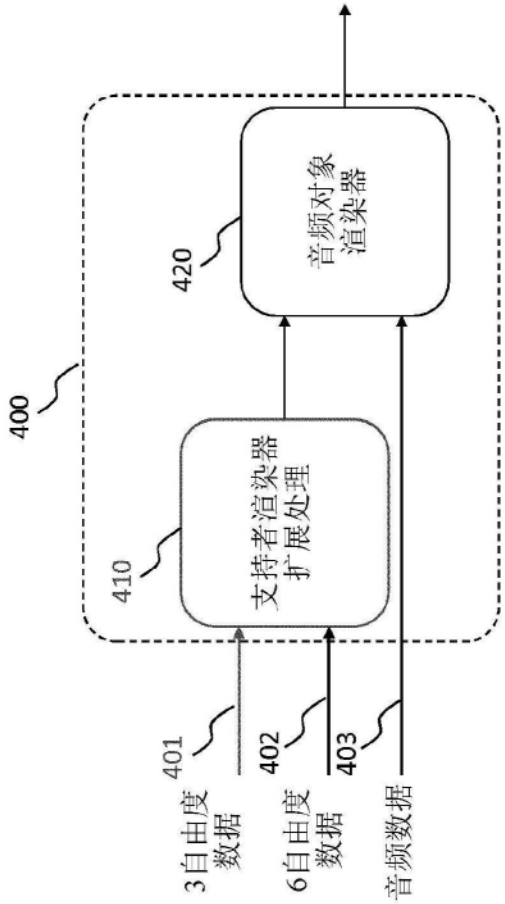

400
410
支持者渲染器
扩展处理
420
音频对象
渲染器
401
402
403
3自由度
数据
6自由度
数据
音频数据

图4A

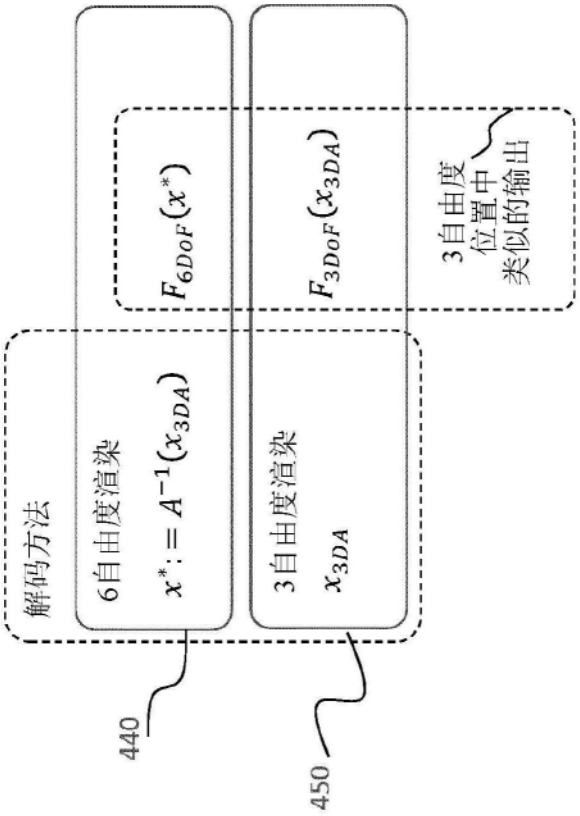

解码方法
6自由度渲染
$x^* := A^{-1}(x_{3DA})$
3自由度渲染
x_{3DA}
$F_{6DoF}(x^*)$
$F_{3DoF}(x_{3DA})$
3自由度
位置中
类似的输出
440
450

图4B

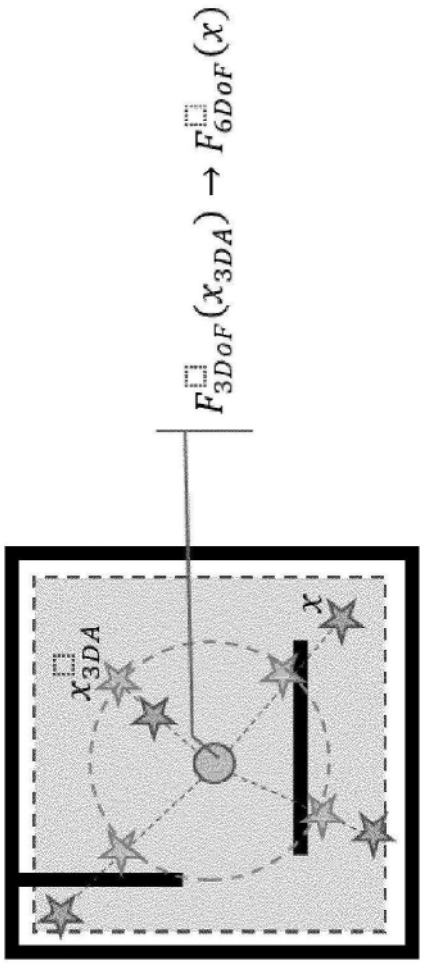

$F_{3DoF}^{\square}(x_{3DA}) \rightarrow F_{6DoF}^{\square}(x)$
$x_{3DA}^{\square}$
x

图5

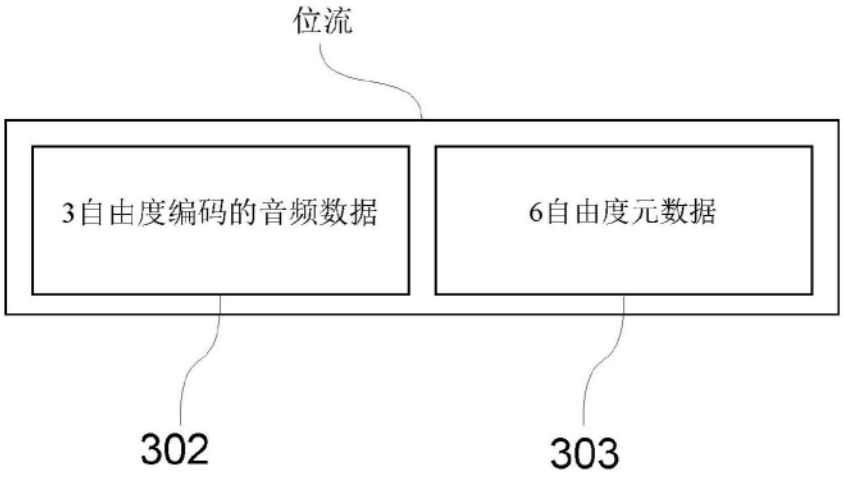

位流
3自由度编码的音频数据
6自由度元数据
302
303

图6A

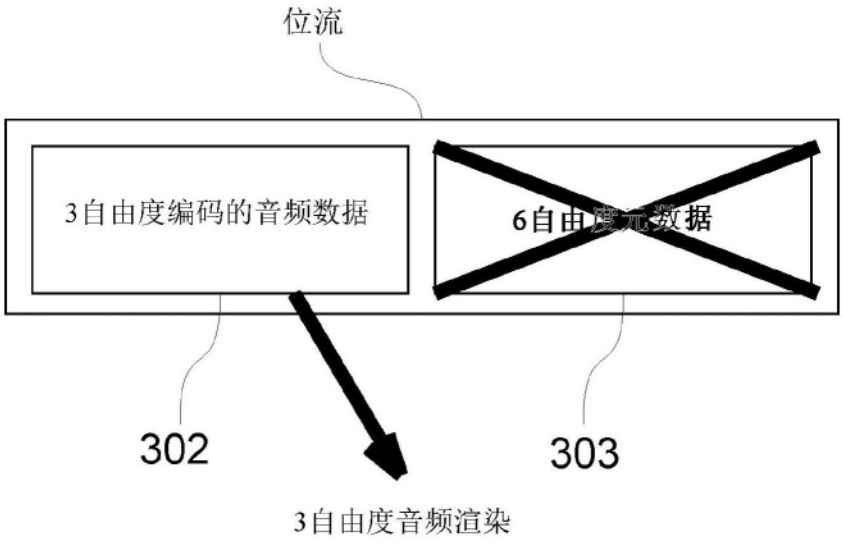

位流
3自由度编码的音频数据
6自由度元数据
302
303
3自由度音频渲染

图6B

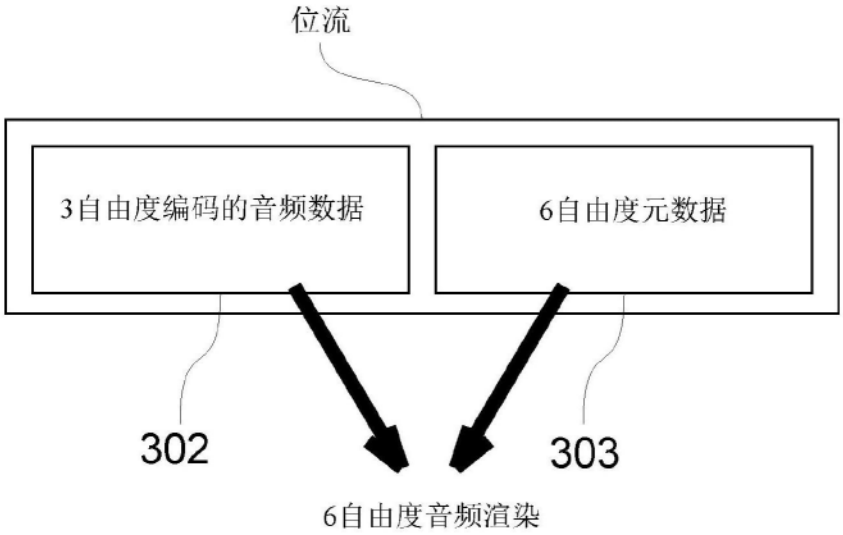

位流
3自由度编码的音频数据
6自由度元数据
302
303
6自由度音频渲染

图6C

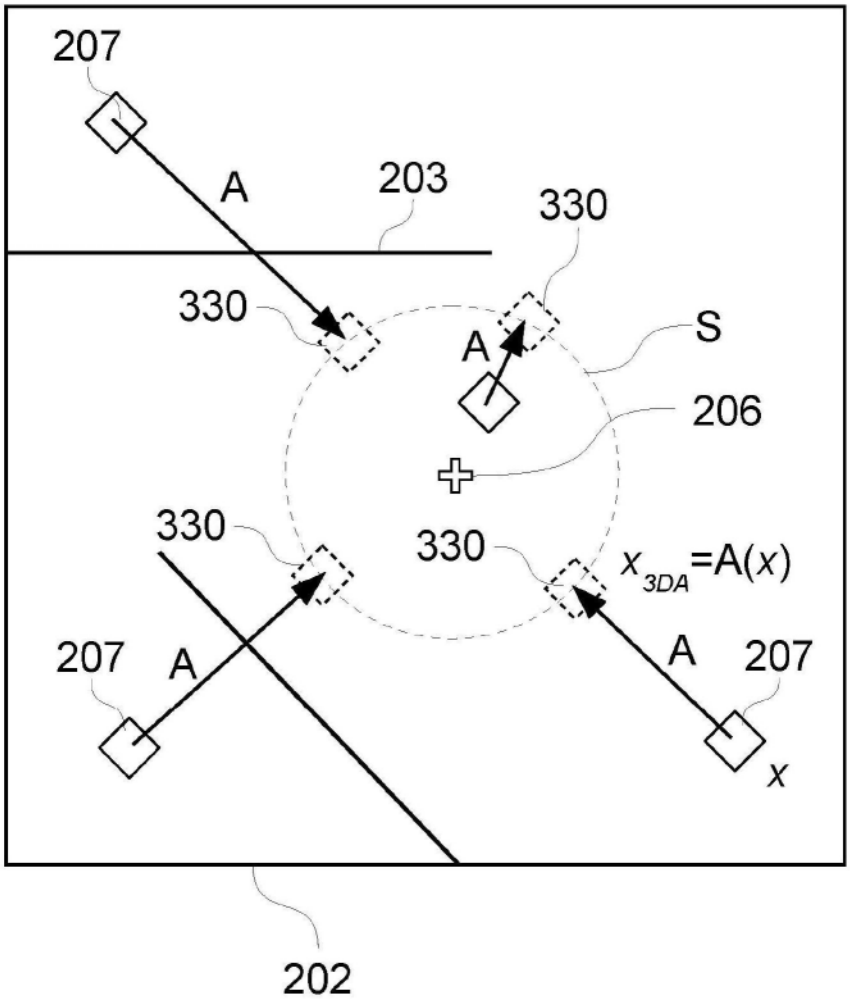


图7A

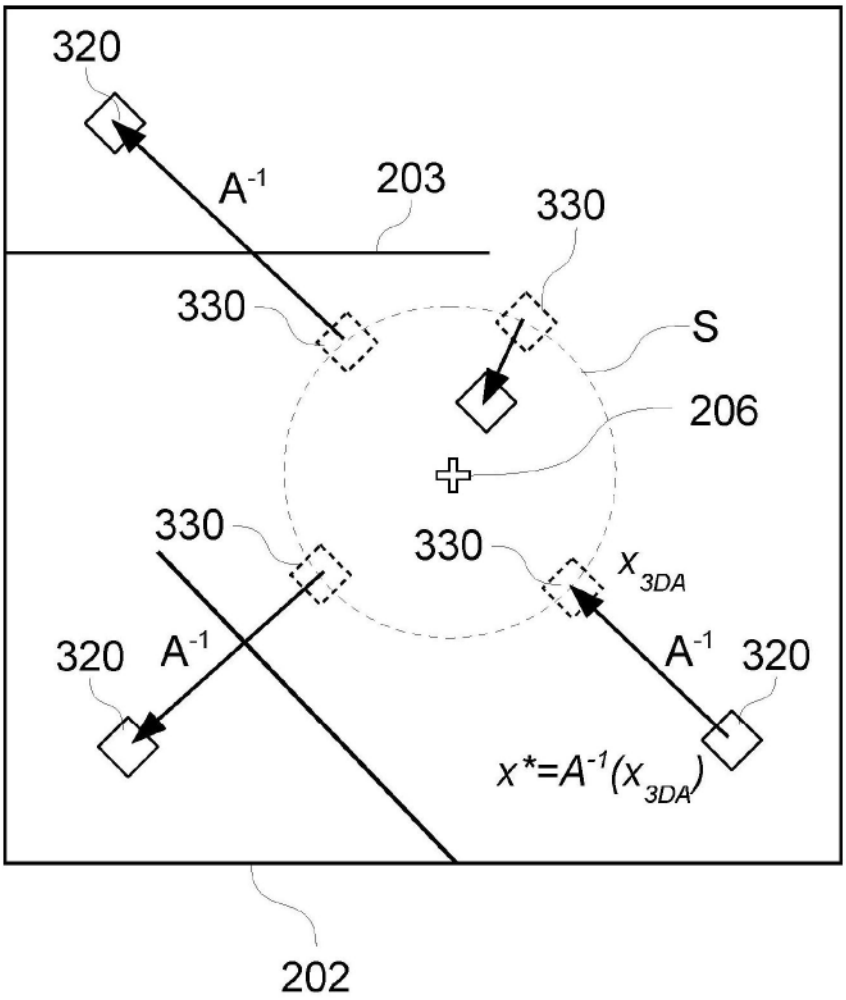


图7B

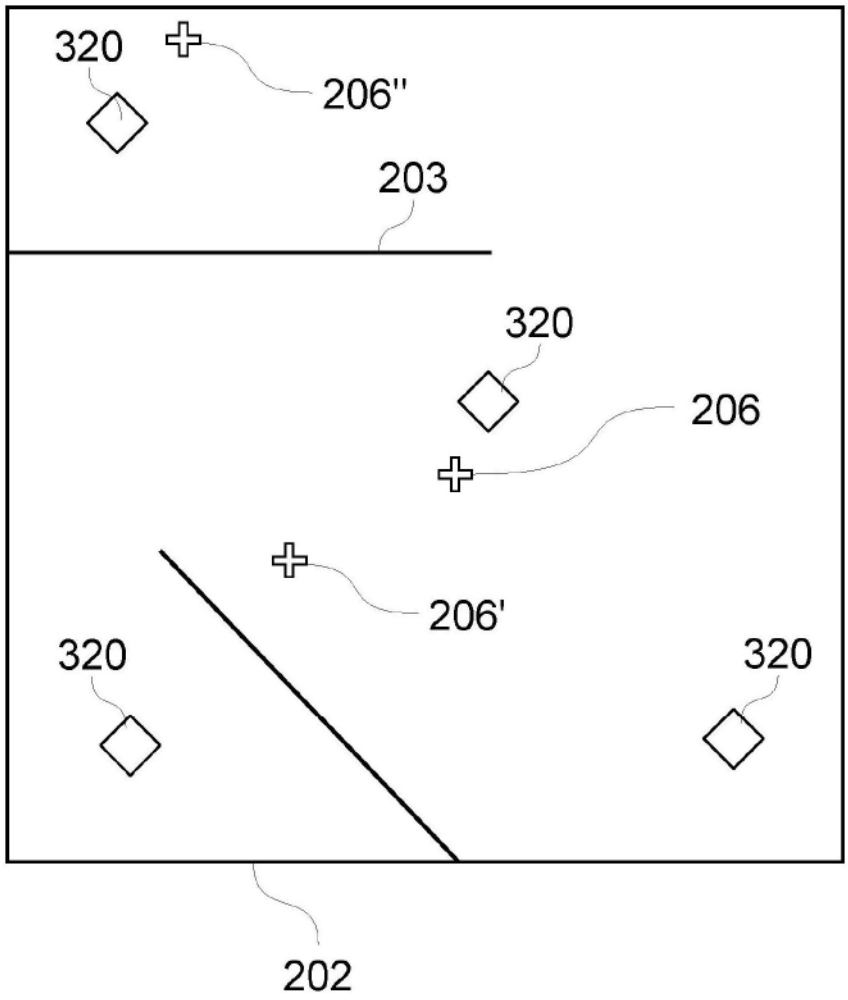


图7C

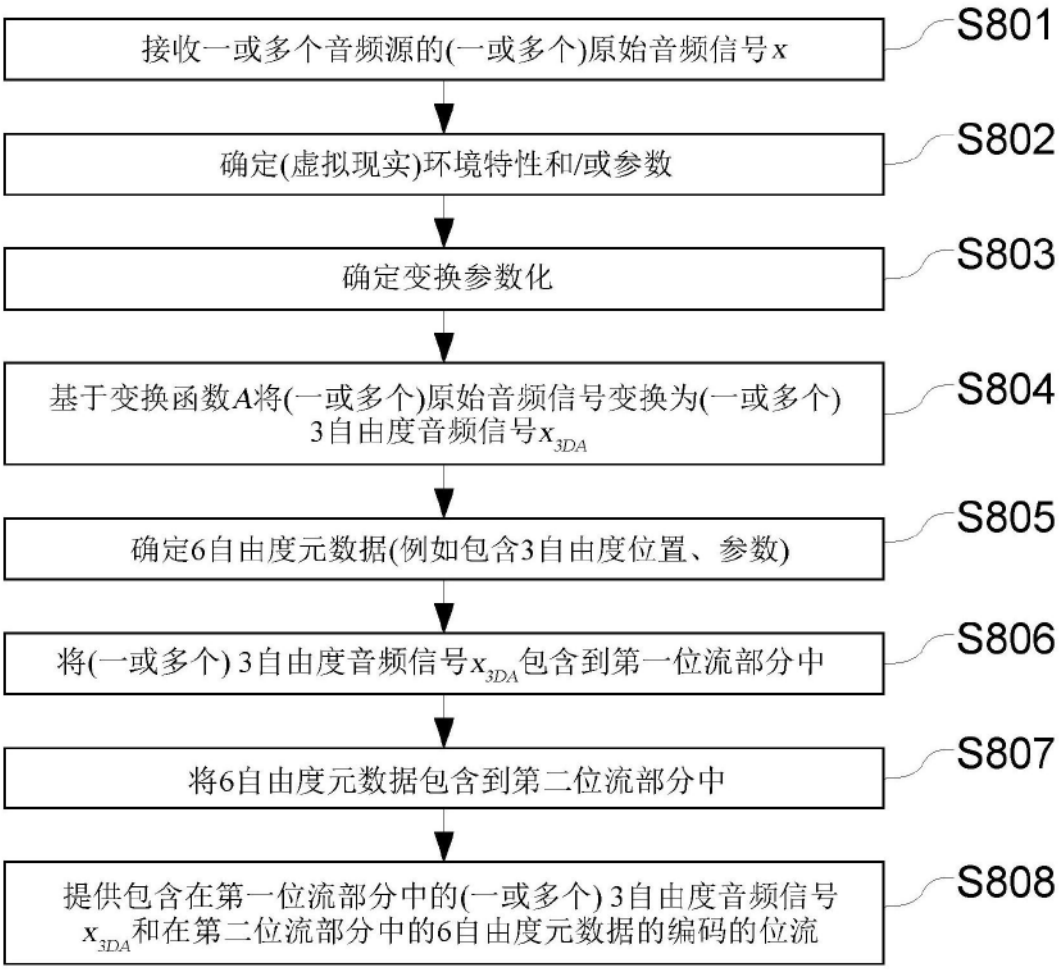

接收一或多个音频源的(一或多个)原始音频信号x
S801
确定(虚拟现实)环境特性和/或参数
S802
确定变换参数化
S803
基于变换函数A将(一或多个)原始音频信号变换为(一或多个)3自由度音频信号x_{3DA}
S804
确定6自由度元数据(例如包含3自由度位置、参数)
S805
将(一或多个)3自由度音频信号x_{3DA}包含到第一位流部分中
S806
将6自由度元数据包含到第二位流部分中
S807
提供包含在第一位流部分中的(一或多个)3自由度音频信号x_{3DA}和在第二位流部分中的6自由度元数据的编码的位流
S808

图8

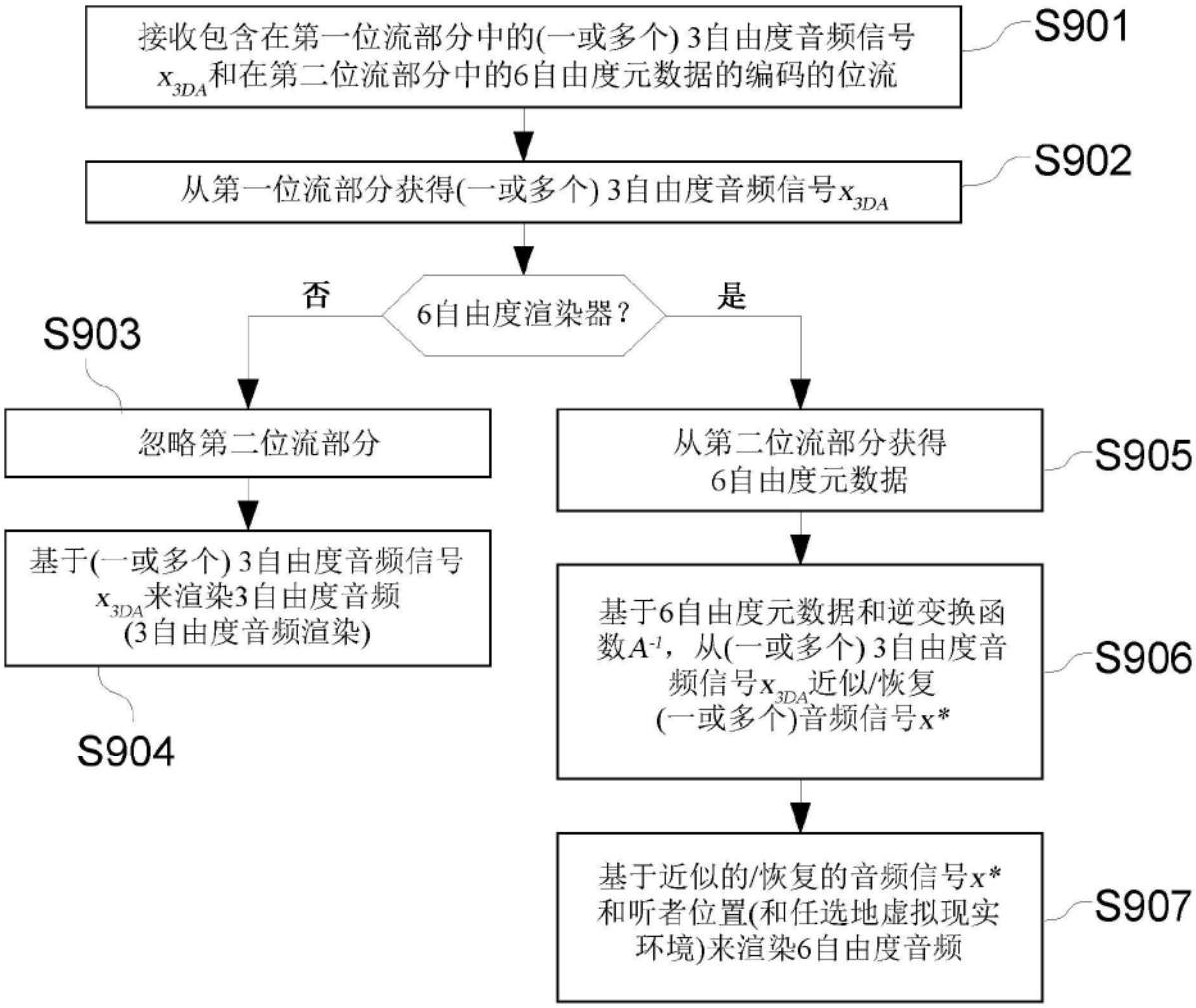

接收包含在第一位流部分中的(一或多个) 3自由度音频信号 x_{3DA}和在第二位流部分中的6自由度元数据的编码的位流
S901
从第一位流部分获得(一或多个) 3自由度音频信号x_{3DA}
S902
6自由度渲染器？
否
是
S903
忽略第二位流部分
基于(一或多个) 3自由度音频信号 x_{3DA}来渲染3自由度音频 (3自由度音频渲染)
S904
从第二位流部分获得 6自由度元数据
S905
基于6自由度元数据和逆变换函数A^{-1}，从(一或多个) 3自由度音频信号x_{3DA}近似/恢复 (一或多个)音频信号x^*
S906
基于近似的/恢复的音频信号x^*和听者位置(和任选地虚拟现实环境)来渲染6自由度音频
S907

图9