



(12) 发明专利申请

(10) 申请公布号 CN 113993061 A

(43) 申请公布日 2022. 01. 28

(21) 申请号 202111294219.3

(51) Int.Cl.

(22) 申请日 2019.04.09

H04S 7/00 (2006.01)

(30) 优先权数据

62/654,915 2018.04.09 US

62/695,446 2018.07.09 US

62/823,159 2019.03.25 US

(62) 分案原申请数据

201980018139.X 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 克里斯托弗·费尔施

利昂·特连蒂夫 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公司

责任公司 11287

代理人 刘锋

权利要求书1页 说明书21页 附图5页

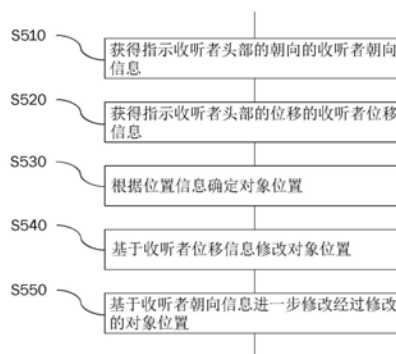
(54) 发明名称

用于MPEG-H 3D音频的三自由度(3DOF+)扩展的方法、设备和系统

(57) 摘要

本申请涉及用于MPEG-H 3D音频的三自由度(3DOF+)扩展的方法、设备和系统。描述了一种处理指示音频对象的对象位置的位置信息的方法,其中所述对象位置可用于渲染所述音频对象,所述方法包括:获得指示收听者头部的朝向的收听者朝向信息;获得指示所述收听者头部的位移的收听者位移信息;根据所述位置信息确定所述对象位置;通过对所述对象位置应用平移基于所述收听者位移信息修改所述对象位置;以及基于所述收听者朝向信息进一步修改经过修改的对象位置。进一步描述了一种用于处理指示音频对象的对象位置的位置信息的对应设备,其中所述对象位置可用于渲染所述音频对象。

500



1. 一种处理指示音频对象的对象位置的位置信息的方法,其中使用MPEG-H 3D音频解码器执行所述处理,其中所述对象位置能够用于渲染所述音频对象,所述方法包括:

获得指示收听者头部的朝向的收听者朝向信息;

经由MPEG-H 3D音频解码器输入接口获得指示所述收听者头部相对于标称收听位置的位移的收听者位移信息;

根据所述位置信息确定所述对象位置;

通过对所述对象位置应用平移基于所述收听者位移信息修改所述对象位置;以及

基于所述收听者朝向信息进一步修改经过修改的对象位置,其中,当所述收听者位移信息指示所述收听者头部相对于所述标称收听位置位移一定小的位置位移时,所述一定小的位置位移的绝对值为0.5米或小于0.5米,在所述收听者头部位移后,所述经过修改的音频对象位置与收听位置之间的距离保持为等于所述音频对象位置与所述标称收听位置之间的原始距离。

用于MPEG-H 3D音频的三自由度(3DOF+)扩展的方法、设备和系统

[0001] 分案申请的相关信息

[0002] 本案是分案申请。该分案的母案是申请日为2019年4月9日、申请号为201980018139.X、发明名称为“用于MPEG-H 3D音频的三自由度(3DOF+)扩展的方法、设备和系统”的发明专利申请案。

[0003] 相关申请交叉引用

[0004] 本申请要求以下优先权申请的优先权：于2018年4月9日提交的美国临时申请62/654,915(参考:D18045USP1)；于2018年7月9日提交的美国临时申请62/695,446(参考:D18045USP2)和于2019年3月25日提交的美国临时申请62/823,159(参考:D18045USP3)，所述美国临时申请通过引用并入本文。

技术领域

[0005] 本公开涉及用于处理指示音频对象位置的位置信息和指示收听者头部位置位移的信息的方法和设备。

背景技术

[0006] ISO/IEC 23008-3MPEG-H 3D音频标准的第一版本(2015年10月15日)和修正案1-4没有规定允许用户头部在三自由度(3DoF)环境中的一定小的平移移动。

发明内容

[0007] ISO/IEC 23008-3MPEG-H 3D音频标准的第一版本(2015年10月15日)和修正案1-4提供了用于3DoF环境的可能性的功能，其中用户(收听者)执行头部旋转动作。然而，此类功能最多仅支持旋转场景位移信号传递和对应的渲染。这意指在收听者头部朝向改变的情况下，音频场景可以保持空间上固定，这对应于3DoF性质。然而，在当前MPEG-H3D音频生态系统内，不可能考虑用户头部的一定小的平移移动。

[0008] 因此，需要用于处理音频对象的位置信息的方法和设备，所述方法和设备可以潜在地结合用户头部的旋转移动来考虑用户头部的一定小的平移移动。

[0009] 本公开提供了用于处理位置信息的设备和系统，所述设备和系统具有相应的独立权利要求和从属权利要求的特征。

[0010] 根据本公开的一方面，描述了一种处理指示音频对象位置的位置信息的方法，其中所述处理可以符合MPEG-H 3D音频标准。所述对象位置可以用于渲染所述音频对象。音频对象可以与其位置信息一起包含在基于对象的音频内容中。位置信息可以是音频对象的元数据(的一部分)。音频内容(例如，音频对象及其位置信息)可以在经编码音频位流中传送。所述方法可以包含接收音频内容(例如，经编码音频位流)。所述方法可以包含获得指示收听者头部的朝向的收听者朝向信息。收听者可以被称为(例如执行所述方法的音频解码器的)用户。收听者头部的朝向(收听者朝向)可以是收听者头部相对于标称朝向的朝向。所

述方法可以进一步包含获得指示收听者头部的位移的收听者位移信息。收听者头部的位移可以是相对于标称收听位置的位移。标称收听位置(或标称收听者位置)可以是默认位置(例如,预定位置、收听者头部的预期位置或扬声器布置的最佳点)。收听者朝向信息和收听者位移信息可以通过MPEG-H 3D音频解码器输入接口获得。收听者朝向信息和收听者位移信息可以基于传感器信息导出。朝向信息和位置信息的组合可以被称为姿势信息。所述方法可以进一步包含根据所述位置信息确定所述对象位置。例如,对象位置可以从位置信息中提取。对象位置的确定(例如,提取)可以进一步基于关于收听环境中一或多个扬声器的扬声器布置的几何形状的信息。对象位置也可以被称为音频对象的信道位置。所述方法可以进一步包含通过对所述对象位置应用平移基于所述收听者位移信息修改所述对象位置。修改对象位置可以涉及针对收听者头部从标称收听位置的位移纠正对象位置。换句话说,修改对象位置可以涉及对对象位置应用位置位移补偿。所述方法可以又进一步包含例如,通过对经过修改的对象位置应用旋转变换(例如,相对于所述收听者头部或所述标称收听位置的旋转)基于所述收听者朝向信息进一步修改经过修改的对象位置。进一步修改用于渲染音频对象的经过修改的对象位置可以涉及旋转音频场景位移。

[0011] 如以上所描述的进行配置,所提出的方法提供了更真实的收听体验,尤其是对于定位于靠近收听者头部的音频对象。除了常规地在3DoF环境中提供给收听者的三个(旋转)自由度之外,所提出的方法还可以考虑收听者头部的平移移动。这使得收听者能够从不同的角度并且甚至侧面接近靠近的音频对象。例如,收听者可能地除了旋转其头部之外还可以通过稍微移动其头部,来从不同的角度收听靠近收听者头部的“蚊子”音频对象。因此,所提出的方法可以为收听者实现经改善的、更真实的沉浸式收听体验。

[0012] 在一些实施例中,修改所述对象位置并进一步修改经过修改的对象位置可以被执行为使得在根据所述经过进一步修改的对象位置渲染到一或多个真实扬声器或虚拟扬声器之后,所述音频对象由所述收听者在心理声学上感知为源自相对于标称收听位置固定的位置,而不论所述收听者头部从所述标称收听位置的位移和所述收听者头部相对于标称朝向的所述朝向如何。因此,当收听者的头部经历从标称收听位置的位移时,音频对象可以感知为相对于收听者头部移动。同样地,当收听者的头部经历从标称朝向的朝向变化时,音频对象可以感知为相对于收听者的头部旋转。例如,所述一或多个扬声器可以是头戴式耳机的一部分,或者可以是扬声器布置(例如,2.1扬声器布置、5.1扬声器布置、7.1扬声器布置等)的一部分。

[0013] 在一些实施例中,基于收听者位移信息修改对象位置可以通过使对象位置平移一定向量来执行,所述向量与量值呈正相关并且与收听者头部从标称收听位置位移的向量的方向呈负相关。

[0014] 由此,确保由收听者感知到靠近的音频对象根据其头部移动而移动。这有助于为这些音频对象提供更真实的收听体验。

[0015] 在一些实施例中,所述收听者位移信息可以指示所述收听者头部从标称收听位置位移一定小的位置位移。例如,位移的绝对值可以不超过0.5m。位移可以用笛卡尔坐标(例如, x, y, z)或球面坐标(例如,方位角、仰角、半径)来表达。

[0016] 在一些实施例中,所述收听者位移信息可以指示所述收听者头部从标称收听位置的位移,所述位移可通过所述收听者移动其上身和/或头部来实现。因此,收听者可以在无

需移动其下部身体的情况下实现位移。例如,当收听者坐在椅子上时,可以实现收听者头部的位移。

[0017] 在一些实施例中,所述位置信息可以包含对所述音频对象距标称收听位置的的距离的指示。距离(半径)可以小于0.5m。例如,距离可以小于1cm。可替代地,可以由解码器将音频对象距标称收听位置的的距离设置为默认值。

[0018] 在一些实施例中,收听者朝向信息可以包含关于收听者头部的侧转、俯仰和横滚的信息。侧转、俯仰、横滚可以相对于收听者头部的标称朝向(例如,参考朝向)给出。

[0019] 在一些实施例中,所述收听者位移信息可以包含关于从标称收听位置的以笛卡尔坐标或以球面坐标表达的收听者头部位移的信息。因此,对于笛卡尔坐标,位移可以用x坐标、y坐标、z坐标表达,并且对于球面坐标,位移可以用方位角坐标、仰角坐标、半径坐标表达。

[0020] 在一些实施例中,所述方法可以进一步包含由可穿戴设备和/或固定式设备检测所述收听者头部的所述朝向。同样地,所述方法可以进一步包含由可穿戴设备和/或固定式设备检测所述收听者头部从标称收听位置的所述位移。可穿戴设备可以是、对应于和/或包含例如头戴式耳机或增强现实(AR)/虚拟现实(VR)头戴式耳机。例如,固定式设备可以是、对应于和/或包含相机传感器。这允许获得关于收听者头部的位移和/或朝向的准确信息,并且由此实现根据朝向和/或位移对靠近的音频对象的现实处理。

[0021] 在一些实施例中,所述方法可以进一步包含根据所述经过进一步修改的对象位置将所述音频对象渲染到一或多个真实扬声器或虚拟扬声器。例如,可以将音频对象渲染到头戴式耳机的左扬声器和右扬声器。

[0022] 在一些实施例中,所述渲染可以被执行为基于所述收听者头部的头部相关传递函数(HRTF)考虑所述音频对象距所述收听者头部小距离的声音闭塞。由此,渲染靠近的音频对象将会由收听者以甚至更真实的形式感知。

[0023] 在一些实施例中,所述经过进一步修改的对象位置可以被调整为由MPEG-H 3D音频渲染器使用的输入格式。在一些实施例中,所述渲染可以使用MPEG-H 3D音频渲染器执行。在一些实施例中,所述处理可以使用MPEG-H 3D音频解码器执行。在一些实施例中,所述处理可以由MPEG-H 3D音频解码器的场景位移单元执行。因此,所提出的方法允许在MPEG-H 3D音频标准的框架中实施有限的六自由度(6DoF)体验(即,3DoF+)。

[0024] 根据本公开的另一方面,描述了一种处理指示音频对象的对象位置的位置信息的另外的方法。所述对象位置可以用于渲染所述音频对象。所述方法可以包含获得指示所述收听者头部的位移的收听者位移信息。所述方法可以进一步包含根据所述位置信息确定所述对象位置。所述方法可以又进一步包含通过对所述对象位置应用平移基于所述收听者位移信息修改所述对象位置。

[0025] 如以上所描述的进行配置,所提出的方法提供了更真实的收听体验,尤其是对于定位于靠近收听者头部的音频对象。通过能够考虑收听者头部的一定小的平移运动,所提出的方法使得收听者能够从不同的角度并且甚至侧面接近靠近的音频对象。因此,所提出的方法可以为收听者实现经改善的、更真实的沉浸式收听体验。

[0026] 在一些实施例中,基于所述收听者位移信息修改所述对象位置被执行为使得在根据所述经过修改的对象位置渲染到一或多个真实扬声器或虚拟扬声器之后,所述音频对象

由所述收听者在心理声学上感知为源自相对于标称收听位置固定的位置,而不论所述收听者头部从所述标称收听位置的位移如何。

[0027] 在一些实施例中,基于收听者位移信息修改对象位置可以通过使对象位置平移一定向量来执行,所述向量与量值呈正相关并且与收听者头部从标称收听位置位移的向量的方向呈负相关。

[0028] 根据本公开的另一方面,描述了一种处理指示音频对象的对象位置的位置信息的另外的方法。所述对象位置可以用于渲染所述音频对象。所述方法可以包含获得指示收听者头部的朝向的收听者朝向信息。所述方法可以进一步包含根据所述位置信息确定所述对象位置。所述方法可以又进一步包含例如通过对所述对象位置应用旋转变换(例如,相对于所述收听者头部或所述标称收听位置的旋转)基于所述收听者朝向信息修改所述对象位置。

[0029] 如以上所描述的进行配置,所提出的方法可以考虑收听者头部的朝向以为收听者提供更加真实的收听体验。

[0030] 在一些实施例中,基于所述收听者朝向信息修改所述对象位置可以被执行为使得在根据所述经过修改的对象位置渲染到一或多个真实扬声器或虚拟扬声器之后,所述音频对象由所述收听者在心理声学上感知为源自相对于标称收听位置固定的位置,而不论所述收听者头部相对于标称朝向的所述朝向如何。

[0031] 根据本公开的另一方面,描述了一种用于处理指示音频对象的对象位置的位置信息的设备。所述对象位置可以用于渲染所述音频对象。所述设备可以包含处理器和存储器,所述存储器耦接到所述处理器。所述处理器可以适用于获得指示收听者头部的朝向的收听者朝向信息。所述处理器可以进一步适用于获得指示所述收听者头部的位移的收听者位移信息。所述处理器可以进一步适用于根据所述位置信息确定所述对象位置。所述处理器可以进一步适用于通过对所述对象位置应用平移基于所述收听者位移信息修改所述对象位置。所述处理器可以又进一步适用于例如,通过对所述经过修改的对象位置应用旋转变换(例如,相对于所述收听者头部或所述标称收听位置的旋转)基于所述收听者朝向信息进一步修改所述经过修改的对象位置。

[0032] 在一些实施例中,所述处理器可以适用于修改所述对象位置并进一步修改所述经过修改的对象位置,使得在根据所述经过进一步修改的对象位置渲染到一或多个真实扬声器或虚拟扬声器之后,所述音频对象由所述收听者在心理声学上感知为源自相对于标称收听位置固定的位置,而不论所述收听者头部从所述标称收听位置的位移和所述收听者头部相对于标称朝向的朝向如何。

[0033] 在一些实施例中,所述处理器可以适用于通过使所述对象位置平移一定向量而基于所述收听者位移信息修改所述对象位置,所述向量与量值呈正相关并且与所述收听者头部从标称收听位置位移的向量的方向呈负相关。

[0034] 在一些实施例中,所述收听者位移信息可以指示所述收听者头部从标称收听位置位移一定小的位置位移。

[0035] 在一些实施例中,所述收听者位移信息可以指示所述收听者头部从标称收听位置的位移,所述位移可通过所述收听者移动其上身和/或头部来实现。

[0036] 在一些实施例中,所述位置信息可以包含对所述音频对象距标称收听位置的距离

的指示。

[0037] 在一些实施例中,收听者朝向信息可以包含关于收听者头部的侧转、俯仰和横滚的信息。

[0038] 在一些实施例中,所述收听者位移信息可以包含关于从标称收听位置的以笛卡尔坐标或以球面坐标表达的收听者头部位移的信息。

[0039] 在一些实施例中,所述设备可以进一步包含可穿戴设备和/或固定式设备,所述可穿戴设备和/或固定式设备用于检测所述收听者头部的所述朝向。在一些实施例中,所述设备可以进一步包含可穿戴设备和/或固定式设备,所述可穿戴设备和/或固定式设备用于检测所述收听者头部从标称收听位置的所述位移。

[0040] 在一些实施例中,所述处理器可以进一步适用于根据所述经过进一步修改的对象位置将所述音频对象渲染到一或多个真实扬声器或虚拟扬声器。

[0041] 在一些实施例中,所述处理器可以适用于基于所述收听者头部的HRTF执行考虑所述音频对象距所述收听者头部小距离的声音闭塞的渲染。

[0042] 在一些实施例中,所述处理器可以适用于将所述经过进一步修改的对象位置调整为由MPEG-H 3D音频渲染器使用的输入格式。在一些实施例中,所述渲染可以使用MPEG-H 3D音频渲染器执行。也就是说,所述处理器可以实施MPEG-H 3D音频渲染器。在一些实施例中,所述处理器可以适用于实施MPEG-H 3D音频解码器。在一些实施例中,所述处理器可以适用于实施MPEG-H 3D音频解码器的场景位移单元。

[0043] 根据本公开的另一方面,描述了一种用于处理指示音频对象的对象位置的位置信息的另外的设备。所述对象位置可以用于渲染所述音频对象。所述设备可以包含处理器和存储器,所述存储器耦接到所述处理器。所述处理器可以适用于获得指示所述收听者头部的位移的收听者位移信息。所述处理器可以进一步适用于根据所述位置信息确定所述对象位置。所述处理器可以又进一步适用于通过对所述对象位置应用平移基于所述收听者位移信息修改所述对象位置。

[0044] 在一些实施例中,所述处理器可以适用于基于所述收听者位移信息修改所述对象位置,使得在根据所述经过修改的对象位置渲染到一或多个真实扬声器或虚拟扬声器之后,所述音频对象由所述收听者在心理声学上感知为源自相对于标称收听位置固定的位置,而不论所述收听者头部从所述标称收听位置的位移如何。

[0045] 在一些实施例中,所述处理器可以适用于通过使所述对象位置平移一定向量而基于所述收听者位移信息修改所述对象位置,所述向量与量值呈正相关并且与所述收听者头部从标称收听位置位移的向量的方向呈负相关。

[0046] 根据本公开的另一方面,描述了一种用于处理指示音频对象的对象位置的位置信息的另外的设备。所述对象位置可以用于渲染所述音频对象。所述设备可以包含处理器和存储器,所述存储器耦接到所述处理器。所述处理器可以适用于获得指示收听者头部的朝向的收听者朝向信息。所述处理器可以进一步适用于根据所述位置信息确定所述对象位置。所述处理器可以又进一步适用于例如,通过对所述经过修改的对象位置应用旋转变换(例如,相对于所述收听者头部或所述标称收听位置的旋转)基于所述收听者朝向信息修改所述对象位置。

[0047] 在一些实施例中,所述处理器可以适用于基于所述收听者朝向信息修改所述对象

位置,使得在根据所述经过修改的对象位置渲染到一或多个真实扬声器或虚拟扬声器之后,所述音频对象由所述收听者在心理声学上感知为源自相对于标称收听位置固定的位置,而不论所述收听者头部相对于标称朝向的朝向如何。

[0048] 根据又另一方面,描述了一种系统。所述系统可以包含一种根据上述方面中任何方面所述的设备和可穿戴设备和/或固定式设备,所述可穿戴设备和/或固定式设备能够检测收听者头部的朝向并且检测所述收听者头部的位移。

[0049] 应当理解,方法步骤和设备特征可以以多种方式互换。具体地,如本领域的技术人员所理解的,所公开的方法的细节可以被实施为适用于执行所述方法的步骤中的一些或全部步骤的设备,并且反之亦然。具体地,应当理解,根据本公开的设备可以涉及用于实现或执行根据以上实施例和其变体的方法的设备,并且关于所述方法作出的相应陈述类似地适用于对应设备。同样地,应当理解,根据本公开的方法可以涉及根据以上实施例和其变体操作设备的方法,并且关于所述设备作出的相应陈述类似地适用于对应方法。

附图说明

[0050] 以下参考附图以示范性方式解释本发明,其中

[0051] 图1示意性地展示了MPEG-H 3D音频系统的实例;

[0052] 图2示意性地展示了根据本发明的MPEG-H 3D音频系统的实例;

[0053] 图3示意性地展示了根据本发明的音频渲染系统的实例;

[0054] 图4示意性地展示了笛卡尔坐标轴(Cartesian coordinate axes)和其与球面坐标的关系的示例集合;以及

[0055] 图5是示意性地展示了根据本发明的处理音频对象的位置信息的方法的实例的流程图。

具体实施方式

[0056] 如本文所使用的,3DoF通常是可以正确处理用三个参数(例如,侧转、俯仰、横滚)指定的用户的头部移动(特别是头部旋转)的系统。此类系统通常可用于各种游戏系统中,如虚拟现实(VR)/增强现实(AR)/混合现实(MR)系统,或这种类型的其它声学环境中。

[0057] 如本文所使用的,(例如,音频解码器或包括音频解码器的再现系统的)用户也可以被称为“收听者”。

[0058] 如本文所使用的,3DoF+应意指,除了可以在3DoF系统中正确处理的用户的头部移动之外,还可以处理一定小的平移移动。

[0059] 如本文所使用的,“一定小”应指示移动限制在通常为0.5米的阈值以下。这意味着距用户的原始头部位置移动不大于0.5米。例如,用户移动受他/她坐在椅子上的约束。

[0060] 如本文所使用的,“MPEG-H 3D音频”应指在ISO/IEC 23008-3中标准化的说明书和/或ISO/IEC 23008-3标准的其任何未来修正案、版本或其它版本。

[0061] 在由MPEG组织提供的音频标准的上下文中,3DoF与3DoF+之间的区别可以如下定义:

[0062] • 3DoF:允许用户体验(例如,用户头部)的侧转移动、俯仰移动、横滚移动;

[0063] • 3DoF+:允许用户例如在坐在椅子上时体验(例如,用户头部)的侧转移动、俯仰

移动、横滚移动和有限的平移移动。

[0064] 有限(一定小)的头部平移移动可以是受某个移动半径约束的移动。例如,由于用户处于就座位置,例如,在不使用下部身体的情况下,移动可能受到约束。一定小的头部平移移动可以涉及或对应于用户头部相对于标称收听位置的位移。标称收听位置(或标称收听者位置)可以是默认位置(例如,预定位置、收听者头部的预期位置或扬声器布置的最佳点)。

[0065] 3DoF+体验可以与限制性的6DoF体验相当,其中平移移动可以被描述为有限的或一定小的头部移动。在一个实例中,还基于用户的头部位置和朝向,包含可能的声音闭塞来渲染音频。渲染可以执行为例如,基于收听者头部的头部相关传递函数(HRTF)考虑音频对象距收听者头部小距离的声音闭塞。

[0066] 关于与由MPEG-H 3D音频标准阐述的功能相容的方法、系统、设备和其它装置,其可以意指3DoF+能够用于MPEG标准的一或多个任何未来版本,如全向媒体格式的未来版本(例如,在MPEG-I的未来版本中标准化);和/或MPEG-H音频的任何更新(例如,基于MPEG-H 3D音频标准的修正案或更新的标准);或者可能需要更新的任何其它相关标准或配套标准(例如,指定某些类型的元数据消息和SEI消息的标准)。

[0067] 例如,可以将对于MPEG-H 3D音频说明书中阐述的音频标准来说是规范的音频渲染器扩展为包含对音频场景进行渲染,以准确地说明例如,在用户稍微侧向移动其头部时与音频场景的用户交互。

[0068] 本发明提供了各种技术优势,包含提供能够处理3DoF+用例的MPEG-H 3D音频的优势。本发明扩展MPEG-H 3D音频标准以支持3DoF+功能。

[0069] 为了支持3DoF+功能,音频渲染系统应考虑用户/收听者头部的有限/一定小的位置位移。应基于与初始位置(即,默认位置/标称收听位置)的相对偏移确定位置位移。在一个实例中,此偏移的量值(例如,可以基于 $r_{\text{offset}} = ||P_0 - P_1||$ 确定的半径偏移,其中 P_0 为标称收听位置,并且 P_1 为收听者头部的位移位置)听者头部的移位位置)最大为约0.5m。在另一个实例中,偏移的量值被限制为仅当用户坐在椅子上并且不执行下部身体移动(但是其头部相对于其身体移动)的时候才可实现的偏移。此(一定小的)偏移距离导致远音频对象的非常小的(感知的)水平差异和平移差异。然而,对于靠近对象,即使此类一定小的偏移距离也可能在感知上变得相关。事实上,收听者的头部移动可以对感知正确的音频对象定位的定位具有感知效果。只要(i)用户的头部位移(例如, $r_{\text{offset}} = ||P_0 - P_1||$)与到音频对象的距离(例如, r)之比用三角学方法产生在用户检测声音方向的心理声学能力的范围内的角度,此感知效果就可以保持显著(即,由用户/收听者在感知上可察觉)。对于不同的音频渲染器设置、音频材料和回放配置,此类范围可以有所不同。例如,假设定位准确度范围是例如 $\pm 3^\circ$,其中收听者头部的左右移动自由为 $\pm 0.25\text{m}$,这将对应用于 $\sim 5\text{m}$ 的对象距离。

[0070] 对于靠近收听者的对象(例如,距用户 $< 1\text{m}$ 的距离处的对象),正确处理收听者头部的位移对于3DoF+场景是至关重要的,因为在平移变化和水平变化两者期间有显著的感知效果。

[0071] 对靠近收听者的对象的处理的一个实例是,例如,当音频对象(例如,蚊子)非常靠近收听者的面部定位时。如提供VR/AR/MR能力的音频系统等音频系统,应允许用户从所有侧面和角度感知此音频对象,即使用户正在进行一定小的平移头部移动。例如,用户应能够

准确地感知对象(例如,蚊子),即使当用户在不移动其下部身体的情况下移动其头部。

[0072] 然而,与目前的MPEG-H 3D音频说明书相容的系统当前无法正确地处理此问题。相反,使用与MPEG-H 3D音频系统相容的系统会导致从相对于用户的错误位置中感知“蚊子”。在涉及3DoF+性能的场景中,一定小的平移移动应产生在对音频对象的感知方面的显著差异(例如,当向左移动一个用户的头部时,应从相对于用户头部的右侧感知“蚊子”音频对象等)。

[0073] MPEG-H 3D音频标准包含位流语法,所述位流语法允许通过位流语法,例如通过object_metadata()-语法元素(从0.5m开始)来用信号传递对象距离信息。

[0074] 可以将语法元素prodMetadataConfig()引入到由MPEG-H 3D音频标准提供的位流中,所述位流可以用于用信号传递对象距离非常靠近收听者。例如,语法prodMetadataConfig()可以用信号传递用户与对象之间的距离小于某个阈值距离(例如,<1cm)。

[0075] 图1和图2展示了基于耳机渲染的本发明(即,其中扬声器与收听者的头部共移动)。

[0076] 图1示出了符合MPEG-H 3D音频系统的系统行为100的实例。此实例假设收听者的头部在时间 t_0 时定位于位置 P_0 103处,并且在时间 $t_1 > t_0$ 时移动到位置 P_1 104。位置 P_0 和 P_1 周围的虚线圆圈指示可允许的3DoF+移动区域(例如,半径为0.5m)。位置A 101指示用信号传递的对象位置(在时间 t_0 和时间 t_1 时,即,假设用信号传递的对象位置随时间推移是恒定的)。位置A还指示由MPEG-H 3D音频渲染器在时间 t_0 时渲染的对象位置。位置B 102指示由MPEG-H 3D音频在时间 t_1 时渲染的对象位置。从位置 P_0 和 P_1 向上延伸的竖直线指示收听者的头部在时间 t_0 和 t_1 时的相应朝向(例如,观看方向)。用户头部在位置 P_0 与位置 P_1 之间的位移可以由 $r_{\text{offset}} = ||P_0 - P_1||$ 106表示。在收听者在时间 t_0 时定位于默认位置(标称收听位置) P_0 103处的情况下,他/她将在正确的位置A 101处感知到音频对象(例如,蚊子)。如果用户将在时间 t_1 时移动到位置 P_1 104,如果以当前标准化的形式应用MPEG-H 3D音频处理,则他/她将在位置B 102处感知到音频对象,这引入了所示误差 δ_{AB} 105。也就是说,尽管收听者的头部移动,但是音频对象(例如,蚊子)仍将被感知为定位于收听者头部的正前方(即,与收听者的头部基本上共移动)。值得注意的是,不论收听者头部的朝向如何,都会出现所引入误差 δ_{AB} 105。

[0077] 图2示出了根据本发明的相对于MPEG-H 3D音频的系统200的系统行为的实例。在图2中,收听者的头部在时间 t_0 时定位于位置 P_0 203处,并且在时间 $t_1 > t_0$ 时移动到位置 P_1 204。位置 P_0 和 P_1 周围的虚线圆圈再次指示可允许的3DoF+移动区域(例如,半径为0.5m)。在201处,指示位置A=B意指用信号传递的对象位置(在时间 t_0 和时间 t_1 时,即,假设用信号传递的对象位置随时间推移是恒定的)。位置A=B 201还指示由MPEG-H 3D音频在时间 t_0 和时间 t_1 时渲染的对象位置。从位置 P_0 203和 P_1 204向上延伸的竖直箭头指示收听者的头部在时间 t_0 和 t_1 时的相应朝向(例如,观看方向)。在收听者在时间 t_0 时定位于初始/默认位置(标称收听位置) P_0 203处的情况下,他/她将在正确的位置A 201处感知到音频对象(例如,蚊子)。如果用户将在时间 t_1 时移动到位置 P_1 203,他/她将仍然在位置B 201处感知到音频对象,所述位置类似于(例如,基本上等于)根据本发明的位置A 201。因此,本发明允许用户的位置随时间推移变化(例如,从位置 P_0 203到位置 P_1 204),同时仍然从相同的(空间上固定

的)定位(例如,位置A=B 201等)感知到声音。换句话说,音频对象(例如,蚊子)根据收听者的头部移动(例如,与所述头部移动呈负相关)相对于收听者的头部而移动。这使用户能够在音频对象(例如,蚊子)周围移动,并且从不同的角度或甚至侧面感知到音频对象。用户头部在位置 P_0 与位置 P_1 之间的位移可以由 $r_{\text{offset}} = ||P_0 - P_1||$ 206表示。

[0078] 图3展示了根据本发明的音频渲染系统300的实例。音频渲染系统300可以对应于或包含解码器,例如,MPEG-H 3D音频解码器。音频渲染系统300可以包含具有对应音频场景位移处理接口(例如,根据MPEG-H 3D音频标准的场景位移数据的接口)的音频场景位移单元310。音频场景位移单元310可以输出用于渲染相应的音频对象的对象位置321。例如,场景位移单元可以输出用于渲染相应的音频对象的对象位置元数据。

[0079] 音频渲染系统300可以进一步包含音频对象渲染器320。例如,渲染器可以由硬件、软件和/或通过云计算执行的任何部分或全部处理构成,所述部分或全部处理包含互联网上的通常被称为“云”的与由MPEG-H 3D音频标准阐述的说明书相容的各种服务,如软件开发平台、服务器、存储和软件。音频对象渲染器320可以根据相应的对象位置(这些对象位置可以是以下描述的经过修改的对象位置或经过进一步修改的对象位置)将音频对象渲染到一或多个(真实或虚拟)扬声器。音频对象渲染器320可以将音频对象渲染到耳机和/或扬声器。也就是说,音频对象渲染器320可以根据给定再现格式生成对象波形。为此,音频对象渲染器320可以利用经压缩对象元数据。每个对象可以根据其对象位置(例如,经过修改的对象位置,或经过进一步修改的对象位置)渲染到某些输出信道。因此,对象位置也可以被称为其音频对象的信道位置。音频对象位置321可以包含在由场景位移单元310输出的对象位置元数据或场景位移元数据中。

[0080] 本发明的处理可以符合MPEG-H 3D音频标准。如此,所述处理可以由MPEG-H 3D音频解码器执行,或者更具体地,由MPEG-H场景位移单元和/或MPEG-H 3D音频渲染器执行。因此,图3的音频渲染系统300可以对应于或包含MPEG-H 3D音频解码器(即,符合由MPEG-H 3D音频标准阐述的说明书的解码器)。在一个实例中,音频渲染系统300可以是包括处理器和耦接到处理器的存储器的设备,其中处理器适用于实施MPEG-H 3D音频解码器。具体地,处理器可以适用于实施MPEG-H场景位移单元和/或MPEG-H 3D音频渲染器。因此,处理器可以适用于执行本公开所描述的处理步骤(例如,以下参考图5描述的方法500的步骤S510到S560)。在另一个实例中,可以在云中执行处理或音频渲染系统300。

[0081] 音频渲染系统300可以获得(例如,接收)收听定位数据301。音频渲染系统300可以通过MPEG-H 3D音频解码器输入接口获得收听定位数据301。

[0082] 收听定位数据301可以指示收听者头部的朝向和/或位置(例如,位移)。因此,收听定位数据301(其也可以被称为姿势信息)可以包含收听者朝向信息和/或收听者位移信息。

[0083] 收听者位移信息可以指示收听者头部(例如,从标称收听位置)的位移。收听者位移信息可以对应于或包含收听者头部从标称收听位置的位移的量值的指示, $r_{\text{offset}} = ||P_0 - P_1||$ 206,如图2所展示。在本发明的上下文中,收听者位移信息指示收听者头部从标称收听位置的一定小的位置位移。例如,位移的绝对值可以不超过0.5m。通常,这是收听者的头部从标称收听位置的位移,所述位移可通过收听者移动其上身和/或头部来实现。也就是说,收听者可以在无需移动其下部身体的情况下实现位移。例如,如上文所指示的,当收听者坐在椅子上时,可以实现收听者头部的位移。位移可以以各种坐标系表示,例如,以笛卡尔坐

标(用 x 、 y 、 z 表示)或球面坐标(例如,用方位角、仰角、半径表示)表示。用于表示听众头部位移的替代性坐标系也是可行的并且应被理解为由本公开涵盖。

[0084] 收听者朝向信息可以指示收听者头部的朝向(例如,收听者头部相对于收听者头部的标称朝向/参考朝向的朝向)。例如,收听者朝向信息可以包括关于收听者头部的侧转、俯仰和横滚的信息。此处,侧转、俯仰和横滚可以相对于标称朝向给出。

[0085] 可以从可以提供关于用户的平移移动的信息的接收器连续地收集收听定位数据301。例如,最近可能已经从接收器收集在某一时间实例中使用的收听定位数据301。可以基于传感器信息导出/收集/生成收听定位数据。例如,收听定位数据301可以由具有适当传感器的可穿戴设备和/或固定式设备导出/收集/生成。也就是说,收听者头部的朝向可以由可穿戴设备和/或固定式设备检测。同样地,收听者头部(例如,从标称收听位置)的位移可以由可穿戴设备和/或固定式设备检测。例如,可穿戴设备可以是、对应于和/或包含头戴式耳机(例如,AR/VR头戴式耳机)。例如,固定式设备可以是、对应于和/或包含相机传感器。例如,固定式设备可以包含在电视机或机顶盒中。在一些实施例中,可以从可能已经获得(例如,接收到)传感器信息的音频编码器(例如,符合MPEG-H 3D音频的编码器)接收收听定位数据301。

[0086] 在一个实例中,用于检测收听定位数据301的可穿戴设备和/或固定式设备可以被称为支持头部位置估计/检测和/或头部朝向估计/检测的跟踪装置。存在各种允许使用计算机或智能电话相机(例如,基于面部识别和跟踪“FaceTrackNoIR”、“opentrack”)准确地跟踪用户的头部移动的解决方案。而且,若干个头戴式显示器(HMD)虚拟现实系统(例如,HTC VIVE、Oculus Rift)具有集成式头部跟踪技术。这些解决方案中的任何解决方案都可以在本公开的上下文中使用。

[0087] 同样重要的是要注意,物理世界中的头部位移距离不必与由收听定位数据301所指示的位移一一对应。为了实现超现实效果(例如,经过过度放大的用户运动视差效果),某些应用可以使用不同的传感器校准设置或指定真实空间中的运动与虚拟空间中的运动之间的不同映射。因此,可以预期,在一些用例中,一定小的物理移动在虚拟现实产生较大位移。在任何情况下,可以说,物理世界和虚拟现实中的位移(即,由收听定位数据301指示的位移)的量值呈正相关。同样地,物理世界和虚拟现实中的位移的方向呈正相关。

[0088] 音频渲染系统300可以进一步接收(对象)位置信息(例如,对象位置数据)302和音频数据322。音频数据322可以包含一或多个音频对象。位置信息302可以是音频数据322的元数据的一部分。位置信息302可以指示所述一或多个音频对象的相应的对象位置。例如,位置信息302可以包括对相应的音频对象相对于用户/收听者的标称收听位置的距离的指示。距离(半径)可以小于0.5m。例如,距离可以小于1cm。如果位置信息302不包含对给定音频对象距标称收听位置的指示,则音频渲染系统可以将此音频对象距标称收听位置的距离设置为默认值(例如,1m)。位置信息302可以进一步包括对相应的音频对象的仰角和/或方位角的指示。

[0089] 每个对象位置可以用于渲染其对应的音频对象。因此,位置信息302和音频数据322可以包含在基于对象的音频内容中或形成基于对象的音频内容。音频内容(例如,音频对象/音频数据322及其位置信息302)可以在经编码音频位流中传送。例如,音频内容可以采用从通过网络传输接收到的位流的格式。在这种情况下,可以说音频渲染系统(例如,

从经编码音频位流)接收音频内容。

[0090] 在本发明的一个实例中,元数据参数可以用于利用针对3DoF和3DoF+的向后兼容增强纠正对用例的处理。除了收听者朝向信息之外,元数据还可以包含收听者位移信息。此类元数据参数可以由图2和3以及本发明的任何其它实施例所示的系统利用。

[0091] 向后兼容增强可以允许基于规范性MPEG-H 3D音频场景位移接口纠正对用例(例如,本发明的实施方案)的处理。这意指传统的MPEG-H 3D音频解码器/渲染器将仍然产生输出,即使是不正确的输出。然而,根据本发明的增强MPEG-H 3D音频解码器/渲染器将正确地应用扩展数据(例如,扩展元数据)和处理,并且因此可能以正确的方式处理靠近收听者定位的对象的场景。

[0092] 在一个实例中,本发明涉及以不同于以下概述的格式的格式提供用于用户头部的一定小平移移动的数据,并且可能相应地对公式进行改编。例如,数据可以以如x坐标、y坐标、z坐标(笛卡尔坐标系中)等格式,而不是以方位角、仰角和半径(球面坐标系中)的格式提供。这些坐标系相对于彼此的实例如图4所示。

[0093] 在一个实例中,本发明涉及提供用于输入收听者头部平移移动的元数据(例如,包含在图3所示的收听定位数据301中的收听者位移信息)。元数据可以用于例如场景位移数据的接口。元数据(例如,收听者位移信息)可以通过部署支持3DoF+或6DoF跟踪的跟踪装置来获得。

[0094] 在一个实例中,元数据(例如,收听者位移信息,具体地收听者头部的位移,或等效地,场景位移)可以由以下三个参数表示:sd_azimuth、sd_elevation和sd_radius,所述三个参数涉及收听者头部的位移(或场景位移)的方位角、仰角和半径(球面坐标)。

[0095] 这些参数的语法由下表给出。

[0096] 表264b—mpegh3daPositionalSceneDisplacementData()的语法

语法	位数	助记符
mpegh3daPositionalSceneDisplacementData() {		
sd_azimuth;	8	Uimsbf
sd_elevation;	6	Uimsbf
sd_radius;	4	Uimsbf
}		

[0098] sd_azimuth此字段限定场景位移方位角位置。此字段可以取从-180到180的值。

[0099] $az_{\text{offset}} = (\text{sd_azimuth} - 128) \cdot 1.5$

[0100] $az_{\text{offset}} = \min(\max(az_{\text{offset}}, -180), 180)$

[0101] sd_elevation此字段限定场景位移仰角位置。此字段可以取从-90到90的值。

[0102] $el_{\text{offset}} = (\text{sd_elevation} - 32) \cdot 3.0$

[0103] $el_{\text{offset}} = \min(\max(el_{\text{offset}}, -90), 90)$

[0104] sd_radius此字段限定场景位移半径。此字段可以取从0.015626到0.25的值。

[0105] $r_{\text{offset}} = (\text{sd_radius} + 1) / 16$

[0106] 在另一个实例中,元数据(例如,收听者位移信息)可以由笛卡尔坐标中的以下三个参数表示:sd_x、sd_y、和sd_z,这会将数据的处理从球面坐标减少到笛卡尔坐标。元数据可以基于以下语法:

	语法	位数	助记符
	mpeg3daPositionalSceneDisplacementDataTrans() {		
[0107]	sd_x;	6	uimsbf
	sd_y;	6	uimsbf
	sd_z;	6	uimsbf
	}		

[0108] 如以上所描述的,上述语法或其等效物语法可以用信号传递与围绕x轴、y轴、z轴旋转相关的信息。

[0109] 在本发明的一个实例中,可以通过扩展说明用户头部的位置变化的等式来增强对信道和对象的场景位移角度的处理。也就是说,对对象位置的处理可以考虑(例如,可以至少部分地基于)收听者位移信息。

[0110] 图5的流程图中展示了处理指示音频对象的对象位置的位置信息的方法500的实例。此方法可以由解码器,如MPEG-H 3D音频解码器执行。图3的音频渲染系统300可以作为此类解码器的实例。

[0111] 作为第一步骤(图5中未示出),例如从经编码音频的位流中接收到包含音频对象和对应的位置信息的音频内容。然后,所述方法可以进一步包含对经编码音频内容进行解码以获得音频对象和位置信息。

[0112] 在步骤S510处,获得(例如,接收)收听者朝向信息。收听者朝向信息可以指示收听者头部的朝向。

[0113] 在步骤S520处,获得(例如,接收)收听者位移信息。收听者位移信息可以指示收听者头部的位移。

[0114] 在步骤S530处,根据位置信息确定对象位置。例如,可以从位置信息中提取对象位置(例如,用方位角、仰角、半径或x、y、z或其等效物表示)。对象位置的确定还可以至少部分地基于关于收听环境中一或多个(真实或虚拟)扬声器的扬声器布置的几何形状的信息。如果所述音频对象的位置信息中未包含半径,则解码器可以将半径设置为默认值(例如,1m)。在一些实施例中,默认值可以取决于扬声器布置的几何形状。

[0115] 值得注意的是,步骤S510、S520和S530可以以任何顺序执行。

[0116] 在步骤S540处,基于收听者位移信息修改在步骤S530处确定的对象位置。这可以通过根据位移信息(例如,根据收听者头部的位移)对对象位置应用平移来完成。因此,可以说修改对象位置涉及针对收听者头部的位移(例如,从标称收听位置的位移)纠正对象位置。具体地,基于收听者位移信息修改对象位置可以通过使对象位置平移一定向量来执行,所述向量与量值呈正相关并且与收听者头部从标称收听位置位移的向量的方向呈负相关。图2中示意性地展示了此类平移的实例。

[0117] 在步骤S550处,基于收听者朝向信息进一步修改在步骤S540处获得的经过修改的对象位置。例如,这可以通过根据收听者朝向信息将旋转变换应用于经过修改的对象位置来完成。此旋转可以是例如相对于收听者头部或标称收听位置的旋转。旋转变换可以通过场景位移算法来执行。

[0118] 如以上所指出的,当应用旋转变换时,考虑用户偏移补偿(即,基于收听者位移信息对对象位置的修改)。例如,应用旋转变换可以包含:

[0119] • 计算旋转变换矩阵(基于用户朝向,例如,收听者朝向信息),

- [0120] • 将对象位置从球面坐标转换成笛卡尔坐标;
- [0121] • 将旋转变换应用于经过用户-位置-偏移补偿的音频对象(即,应用于经过修改的对象位置),以及
- [0122] • 在旋转变换之后,将对象位置从笛卡尔坐标转换回球面坐标。
- [0123] 作为另外的步骤S560(图5中未示出),方法500可以包括根据经过进一步修改的对象位置将音频对象渲染到一或多个真实扬声器或虚拟扬声器。为此,经过进一步修改的对象位置可以被调整为由MPEG-H 3D音频渲染器(例如,以上所描述的音频对象渲染器320)使用的输入格式。上述一或多个(真实或虚拟)扬声器可以是例如头戴式耳机的一部分,或者是扬声器布置(例如,2.1扬声器布置、5.1扬声器布置、7.1扬声器布置等)的一部分。在一些实施例中,例如,可以将音频对象渲染到头戴式耳机的左扬声器和右扬声器。
- [0124] 以上所描述的步骤S540和S550的目的如下。即,修改对象位置并进一步修改经过修改的对象位置被执行为使得在根据经过进一步修改的对象位置渲染到一或多个(真实或虚拟)扬声器之后,音频对象被收听者在心理声学上感知为源自相对于标称收听位置固定的位置。音频对象的此固定位置应在心理声学上被感知,而不论收听者头部从标称收听位置的位移如何并且不论收听者头部相对于标称朝向的朝向如何。换句话说,当收听者的头部经历从标称收听位置的位移时,音频对象可以感知为相对于收听者头部移动(平移)。同样地,当收听者的头部经历从标称朝向的朝向变化时,音频对象可以感知为相对于收听者的头部移动(旋转)。由此,收听者可以通过移动其头部从不同角度和距离感知靠近的音频对象。
- [0125] 可以例如由以上所描述的音频场景位移单元310在(旋转/平移)音频场景位移的上下文中执行分别在步骤S540和S550处修改对象位置并进一步修改经过修改的对象位置。
- [0126] 应当注意的是,可以根据手边的特定用例省略某些步骤。例如,如果收听定位数据301仅包含收听者位移信息(但是不包含收听者朝向信息,或仅包含指示收听者头部的朝向从标称朝向没有偏差的收听者朝向信息),则可以省略步骤S550。然后,步骤S560处的渲染将根据在步骤S540处确定的经过修改的对象位置来执行。同样地,如果收听定位数据301仅包含收听者朝向信息(但是不包含收听者位移信息,或仅包含指示收听者头部的朝向从标称收听位置没有偏差的收听者位移信息),则可以省略步骤S540。然后,步骤S550将涉及基于收听者朝向信息修改在步骤S530处确定的对象位置。步骤S560处的渲染将根据在步骤S550处确定的经过修改的对象位置来执行。
- [0127] 广义地说,本发明基于收听者的收听定位数据301提出了对作为基于对象的音频内容(例如,位置信息302以及音频数据322)的一部分接收到的对象位置进行位置更新。
- [0128] 首先,确定对象位置(或信道位置) $p = (az, el, r)$ 。这可以在方法500的步骤530的上下文中(例如,作为所述步骤的一部分)执行。
- [0129] 对于基于信道的信号,半径 r 可以如下确定:
- [0130] 一如果再现扬声器设置中存在(基于信道的输入信号的信道的)预期扬声器,并且再现设置的距离是已知的,则将半径 r 设置为扬声器距离(例如,按cm计)。
- [0131] 一如果再现扬声器设置中不存在预期扬声器,但是再现扬声器(例如,与标称收听位置)的距离是已知的,则将半径 r 设置为最大再现扬声器距离。
- [0132] 一如果再现扬声器设置中不存在预期扬声器,并且再现扬声器距离不是已知的,

则将半径 r 设置为默认值(例如,1023cm)。

[0133] 对于基于对象的信号,半径 r 如下确定:

[0134] 一如果对象距离是已知的(例如,从生产工具和生产格式已知并在`prodMetadataConfig()`中传送),则将半径 r 设置为已知的对象距离(例如,根据MPEG-H 3D音频标准的表AMD5.7由`goa_bsObjectDistance[]`(按cm计)用信号传递)。

[0135] 表AMD5.7—`goa_Production_Metadata()`的语法

[0136]	语法	位数	助记符
[0137]	<pre> goa_Production_Metadata() { /* PRODUCTION METADATA CONFIGURATION */ goa_hasObjectDistance; if (goa_hasObjectDistance) { for (o = 0; o < goa_numberOfOutputObjects; o++) { goa_bsObjectDistance[o] } } } </pre>	1	Bslbf
		8	Uimsbf

[0138] 一如果对象距离从位置信息已知(例如,从对象元数据已知并在`object_metadata()`中传送),则将半径 r 设置为在位置信息中用信号传递的对象距离(例如,设置为与对象元数据一起传送的`radius[]`(按cm计))。半径 r 可以根据下文所示的章节用信号传递:“对象元数据的缩放”和“限制对象元数据”。

[0139] 对象元数据的缩放

[0140] 作为确定对象位置的上下文中的任选步骤,可以缩放根据位置信息确定的对象位置 $p=(az,el,r)$ 。这可以涉及应用缩放因子来为每个分量反转对输入数据的编码器缩放。这可以对每个对象执行。对象位置的实际缩放可以根据以下伪代码来实施:

```
descale_multidata()
{
    for (o = 0; o < num_objects; o++)
        azimuth[o] = azimuth[o] * 1.5;

    for (o = 0; o < num_objects; o++)
        elevation[o] = elevation[o] * 3.0;

    for (o = 0; o < num_objects; o++)
        radius[o] = pow(2.0, (radius[o] / 3.0)) / 2.0;

    for (o = 0; o < num_objects; o++)
        gain[o] = pow(10.0, (gain[o] - 32.0) / 40.0);
[0141]
    if (uniform_spread == 1)
    {
        for (o = 0; o < num_objects; o++)
            spread[o] = spread[o] * 1.5;
    }
    else
    {
        for (o = 0; o < num_objects; o++)
            spread_width[o] = spread_width[o] * 1.5;

        for (o = 0; o < num_objects; o++)
            spread_height[o] = spread_height[o] * 3.0;

        for (o = 0; o < num_objects; o++)
            spread_depth[o] = (pow(2.0, (spread_depth[o] / 3.0)) / 2.0) - 0.5;
[0142]
    }
    for (o = 0; o < num_objects; o++)
        dynamic_object_priority[o] = dynamic_object_priority[o];
}
```

[0143] 限制对象元数据

[0144] 作为确定对象位置的上下文中的另外的任选步骤,可以限制根据位置信息确定的(可能缩放的)对象位置 $p = (az, el, r)$ 。这可以涉及针对每个分量对经解码值施加限制,以使值保持在有效范围内。这可以对每个对象执行。对象位置的实际限制可以根据以下伪代码的功能来实施:

```
limit_range()
{
    minval = -180;
    maxval = 180;
    for (o = 0; o < num_objects; o++)
        azimuth[o] = MIN(MAX(azimuth[o], minval), maxval);

    minval = -90;
    maxval = 90;
    for (o = 0; o < num_objects; o++)
        elevation[o] = MIN(MAX(elevation[o], minval), maxval);

    minval = 0.5;
    maxval = 16;
    for (o = 0; o < num_objects; o++)
        radius[o] = MIN(MAX(radius[o], minval), maxval);

[0145] minval = 0.004;
        maxval = 5.957;
        for (o = 0; o < num_objects; o++)
            gain[o] = MIN(MAX(gain[o], minval), maxval);

    if (uniform_spread == 1)
    {
        minval = 0;
        maxval = 180;
        for (o = 0; o < num_objects; o++)
            spread[o] = MIN(MAX(spread[o], minval), maxval);
    }
    else
    {
        minval = 0;
        maxval = 180;
        for (o = 0; o < num_objects; o++)
            spread_width[o] = MIN(MAX(spread_width[o], minval), maxval);

        minval = 0;
```

```

maxval = 90;
for (o = 0; o < num_objects; o++)
    spread_height[o] = MIN(MAX(spread_height[o], minval), maxval);

minval = 0;
maxval = 15.5;
for (o = 0; o < num_objects; o++)
[0146]     spread_depth[o] = MIN(MAX(spread_depth[o], minval), maxval);
}
minval = 0;
maxval = 7;
for (o = 0; o < num_objects; o++)
    dynamic_object_priority[o] = MIN(MAX(dynamic_object_priority[o], minval),
maxval);
}

```

[0147] 之后,可以将所确定的(并且任选地,经缩放和/或限制的)对象位置 $p=(az, el, r)$ 转换成预定坐标系,例如,根据“公共约定”的坐标系,其中 0° 方位角处于右耳(正值逆时针方向),并且 0° 仰角处于头部顶部(正值向下)。因此,可以根据“公共”约定将对象位置 p 转换成位置 p' 。这利用以下产生对象位置 p' ：

[0148] $p' = (az', el', r)$

[0149] $az' = az + 90^\circ$

[0150] $el' = 90^\circ - el$

[0151] 其中半径 r 不变。

[0152] 同时,可以将收听者头部的由收听者位移信息 $(az_{offset}, el_{offset}, r_{offset})$ 指示的位移转换成预定坐标系。使用“公共约定”,这相当于

[0153] $az'_{offset} = az_{offset} + 90^\circ$

[0154] $el'_{offset} = 90^\circ - el_{offset}$

[0155] 其中半径 r_{offset} 不变。

[0156] 值得注意的是,可以在步骤S530或步骤S540的上下文中执行针对对象位置和收听者头部位移两者的到预定坐标系的转换。

[0157] 可以在方法500的步骤S540的上下文中(例如,作为所述步骤的一部分)执行实际位置更新。位置更新可以包括以下步骤:

[0158] 作为第一步骤,将位置 p 或者在已经执行到预定坐标系的传递的情况下的位置 p' 传递到笛卡尔坐标 (x, y, z) 。在下文中,在没有预期限制的情况下,将针对预定坐标系中的位置 p' 描述所述过程。而且,在没有预期限制的情况下,可以假设坐标轴的以下朝向/方向: x 轴指向右侧(当处于标称朝向时,从收听者头部观看), y 轴直指向前,并且 z 轴直指向上。同时,可以将收听者头部的由收听者位移信息 $(az'_{offset}, el'_{offset}, r_{offset})$ 指示的位移转换成笛卡尔坐标。

[0159] 作为第二步骤,以上述方式根据收听者头部的位移(场景位移)使笛卡尔坐标中的对象位置偏移(平移)。这可以通过以下进行:

[0160] $x = r \cdot \sin(el') \cdot \cos(az') + r_{offset} \cdot \sin(el'_{offset}) \cdot \cos(az'_{offset})$

[0161] $y = r \cdot \sin(el') \cdot \sin(az') + r_{offset} \cdot \sin(el'_{offset}) \cdot \sin(az'_{offset})$

[0162] $z = r \cdot \cos(el') + r_{offset} \cdot \cos(el'_{offset})$

[0163] 以上平移是基于方法500的步骤S540中的收听者位移信息对对象位置进行修改的实例。

[0164] 笛卡尔坐标中的经偏移对象位置转换成球面坐标并且可以被称为 p'' 。经偏移对象位置可以根据公共约定在预定坐标系中表达为 $p'' = (az'', e1'', r')$ 。

[0165] 当存在产生一定小的半径参数变化(即 $r' \approx r$)的收听者头部位移时,经过修改的对象位置 p'' 可以被重新定义为 $p'' = (az'', e1'', r)$ 。

[0166] 在另一个实例中,当存在可以产生相当大的半径参数变化(即 $r' \gg r$)的很大的收听者头部位移时,经过修改的对象位置 p'' 也可以被定义为 $p'' = (az'', e1'', r')$ 而不是具有经过修改的半径参数 r' 的 $p'' = (az'', e1'', r)$ 。

[0167] 可以从收听者的头部位移距离(即, $r_{\text{offset}} = ||P_0 - P_1||$)和初始半径参数(即, $r = ||P_0 - A||$)获得经过修改的半径参数 r' 的对应值(参见例如,图1和2)。例如,经过修改的半径参数 r' 可以基于以下三角关系确定:

$$[0168] \quad r' = \left(r^2 + r_{\text{offset}}^2 \right)^{1/2}$$

[0169] 将此经过修改的半径参数 r' 映射到对象/信道增益及其在随后的音频渲染中的应用可以显著地改善由于用户移动导致的水平变化的感知效果。允许对半径参数 r' 的此类修改实现“自适应最佳点”。这将意指MPEG渲染系统根据收听者的当前定位动态地调整最佳点位置。通常,根据经过修改的(或经过进一步修改的)对象位置对音频对象的渲染可以基于经过修改的半径参数 r' 。具体地,用于渲染音频对象的对象/信道增益可以基于经过修改的半径参数 r' (例如,基于所述经过修改的半径参数修改)。

[0170] 在另一个实例中,在扬声器再现设置和渲染期间(例如,在以上的步骤S560处),可以禁用场景位移。然而,场景位移的任选启用可以是可用的。这使得3DoF+渲染器能够根据收听者的当前定位和朝向创建动态可调整的最佳点。

[0171] 值得注意的是,将对象位置和收听者头部的位移转换为笛卡尔坐标的步骤是任选的,并且可以在任何合适的坐标系中执行根据收听者头部的位移(场景位移)的平移/偏移(修改)。换句话说,上文中对笛卡尔坐标的选择应理解为非限制性实例。

[0172] 在一些实施例中,可以通过位流中的标志(字段、元素、设置位)(例如,useTrackingMode元素)来启用或禁用场景位移处理(包含修改对象位置和/或进一步修改经过修改的对象位置)。ISO/IEC 23008-3中的子条款“17.3用于本地扬声器设置和渲染的接口”和“17.4用于双耳房间脉冲响应(BRIR)的接口”含有对激活场景位移处理的元素useTrackingMode的描述。在本公开的上下文中,useTrackingMode元素应限定(子条款17.3)对通过mpegh3daSceneDisplacementData()接口和mpegh3daPositionalSceneDisplacementData()接口发送的场景位移值的处理是否发生。可替代地或另外,(子条款17.4)useTrackingMode字段应限定是否连接了跟踪器装置并且是否应以特殊的头部跟踪模式处理双耳渲染,这意指对通过mpegh3daSceneDisplacementData()接口和mpegh3daPositionalSceneDisplacementData()接口发送的场景位移值的处理应发生。

[0173] 本文所描述的方法和系统可以实施为软件、固件和/或硬件。某些组件可以例如被实施为在数字信号处理器或微处理器上运行的软件。其它组件可以例如被实施为硬件和/

或专用集成电路。可以将所描述的方法和系统中遇到的信号存储于如随机存取存储器或光存储媒体等媒体上。所述信号可以通过网络,如无线网络、卫星网络、无线网络或有线网络,例如,互联网传递。利用本文所描述的方法和系统的典型装置是用于存储和/或渲染音频信号的便携式电子装置或其它消费者设备。

[0174] 虽然本文件参考了MPEG,并且具体地是MPEG-H 3D音频,但是本公开不应被解释为限于这些标准。相反,如本领域的技术人员将理解的,本公开还可以在其它音频编码标准中找到有利的应用。

[0175] 此外,虽然本文件频繁参考收听者头部(例如,从标称收听位置)的一定小的位置位移,但是本公开不限于一定小的位置位移,并且通常可以应用于收听者头部的任意位置位移。

[0176] 应当注意的是,说明书和附图仅展示了所提出的方法、系统和设备的原理。本领域的技术人员将能够实施各种布置,尽管本文没有明确描述或示出,但是所述布置体现了本发明的原理并且包含在本发明的精神和范围内。此外,本文件中概述的所有实例和实施例原则上明确旨在仅出于解释目的,以帮助读者理解所提出的方法的原理。此外,本文中提供本发明的原理、方面和实施例以及其具体实例的所有陈述旨在涵盖其等效物。

[0177] 除了以上之外,本发明的各种示例实施方案和示例实施例将从以下所列的枚举型示例实施例(EEE)中变得显而易见,所述枚举型示例实施例不是权利要求。

[0178] 第一EEE涉及一种用于解码经编码音频信号位流的方法,所述方法包括:由音频解码设备300接收所述经编码音频信号位流302、322,其中所述经编码音频信号位流包括经编码音频数据322和对应于至少一个对象-音频信号302的元数据;由所述音频解码设备300解码所述经编码音频信号位流302、322以获得多个声源的表示;由所述音频解码设备300接收收听定位数据301;由所述音频解码设备300生成音频对象位置数据321,其中所述音频对象位置数据321基于所述收听定位数据301描述相对于收听定位的多个声源。

[0179] 第二EEE涉及所述第一EEE的所述方法,其中所述收听定位数据301基于第一平位移置数据的第一集合和第二平位移置和朝向数据的第二集合。

[0180] 第三EEE涉及所述第二EEE的所述方法,其中所述第一平位移置数据或所述第二平位移置数据基于球面坐标集或笛卡尔坐标集中的至少一个。

[0181] 第四EEE涉及所述第一EEE的所述方法,其中通过MPEG-H 3D音频解码器输入接口获得收听定位数据301。

[0182] 第五EEE涉及所述第一EEE的所述方法,其中所述经编码音频信号位流包含MPEG-H 3D音频位流语法元素,并且其中所述MPEG-H 3D音频位流语法元素包含所述经编码音频数据322和对应于至少一个对象-音频信号302的所述元数据。

[0183] 第六EEE涉及所述第一EEE的所述方法,所述方法进一步包括由所述音频解码设备300将所述多个声源渲染到多个扬声器,其中渲染过程至少符合MPEG-H 3D音频标准。

[0184] 第七EEE涉及所述第一EEE的所述方法,所述方法进一步包括由所述音频解码设备300基于所述收听定位数据301的平移将对应于所述至少一个对象-音频信号302的位置 p 转换成对应于所述音频对象位置321的第二位置 p'' 。

[0185] 第八EEE涉及所述第七EEE的所述方法,其中所述音频对象位置在预定坐标系中的位置 p' 是基于以下各项确定(例如,根据公共约定)的:

$$[0186] \quad P' = (az', el', r)$$

$$[0187] \quad az' = az + 90^\circ$$

$$[0188] \quad el' = 90^\circ - el$$

$$[0189] \quad az'_{offset} = az_{offset} + 90^\circ$$

$$[0190] \quad el'_{offset} = 90^\circ - el_{offset}$$

[0191] 其中 az 对应于第一方位角参数, el 对应于第一仰角参数,并且 r 对应于第一半径参数,本文中 az' 对应于第二方位角参数, el' 对应于第二仰角参数并且 r' 对应于第二半径参数,其中 az_{offset} 对应于第三方位角参数, el_{offset} 对应于第三仰角参数,并且其中 az'_{offset} 对应于第四方位角参数, el'_{offset} 对应于第四仰角参数。

[0192] 第九EEE涉及所述第八EEE的所述方法,其中所述音频对象位置302的经偏移音频对象位置 p'' 321基于以下各项在笛卡尔坐标 (x, y, z) 中确定:

$$[0193] \quad x = r \cdot \sin(el') \cdot \cos(az') + x_{offset}$$

$$[0194] \quad y = r \cdot \sin(el') \cdot \sin(az') + y_{offset}$$

$$[0195] \quad z = r \cdot \cos(el') + z_{offset}$$

[0196] 其中所述笛卡尔位置 (x, y, z) 由 x 参数、 y 参数和 z 参数组成,并且其中 x_{offset} 涉及第一 x 轴偏移参数, y_{offset} 涉及第一 y 轴偏移参数,并且 z_{offset} 涉及第一 z 轴偏移参数。

[0197] 第十EEE涉及所述第九EEE的所述方法,其中所述参数 x_{offset} 、 y_{offset} 和 z_{offset} 基于以下各项:

$$[0198] \quad x_{offset} = r_{offset} \cdot \sin(el'_{offset}) \cdot \cos(az'_{offset})$$

$$[0199] \quad y_{offset} = r_{offset} \cdot \sin(el'_{offset}) \cdot \sin(az'_{offset})$$

$$[0200] \quad z_{offset} = r_{offset} \cdot \cos(el'_{offset})$$

[0201] 第十一EEE涉及所述第七EEE的所述方法,其中所述方位角参数 az_{offset} 涉及场景位移方位角位置并且基于以下各项:

$$[0202] \quad az_{offset} = (sd_azimuth - 128) \cdot 1.5$$

$$[0203] \quad az_{offset} = \min(\max(az_{offset}, -180), 180)$$

[0204] 其中 $sd_azimuth$ 是指示MPEG-H 3DA方位角场景位移的方位角元数据参数,其中所述仰角参数 el_{offset} 涉及场景位移仰角位置,并且基于以下各项:

$$[0205] \quad el_{offset} = (sd_elevation - 32) \cdot 3$$

$$[0206] \quad el_{offset} = \min(\max(el_{offset}, -90), 90)$$

[0207] 其中 $sd_elevation$ 是指示MPEG-H 3DA仰角场景位移的仰角元数据参数,其中所述半径参数 r_{offset} 涉及场景位移半径并且基于以下各项:

$$[0208] \quad r_{offset} = (sd_radius + 1) / 16$$

[0209] 其中 sd_radius 是指示MPEG-H 3DA半径场景位移的半径元数据参数,并且其中参数 X 和 Y 是标量变量。

[0210] 第十二EEE涉及所述第十EEE的所述方法,其中所述 x_{offset} 参数涉及到 x 轴方向中的场景位移偏位移置 sd_x ;所述 y_{offset} 参数涉及到 y 轴方向中的场景位移偏位移置 sd_y ;并且所述 z_{offset} 参数涉及到 z 轴方向中的场景位移偏位移置 sd_z 。

[0211] 第十三EEE涉及所述第一EEE的所述方法,所述方法进一步包括由所述音频解码设备以更新速率内插与所述收听定位数据301和所述对象音频信号302相关的所述第一位置

数据。

[0212] 第十四EEE涉及所述第一EEE的所述方法,所述方法进一步包括由所述音频解码设备300确定收听定位数据301的有效熵编码。

[0213] 第十五EEE涉及所述第一EEE的所述方法,其中基于传感器信息导出与收听定位数据301相关的所述位置数据。

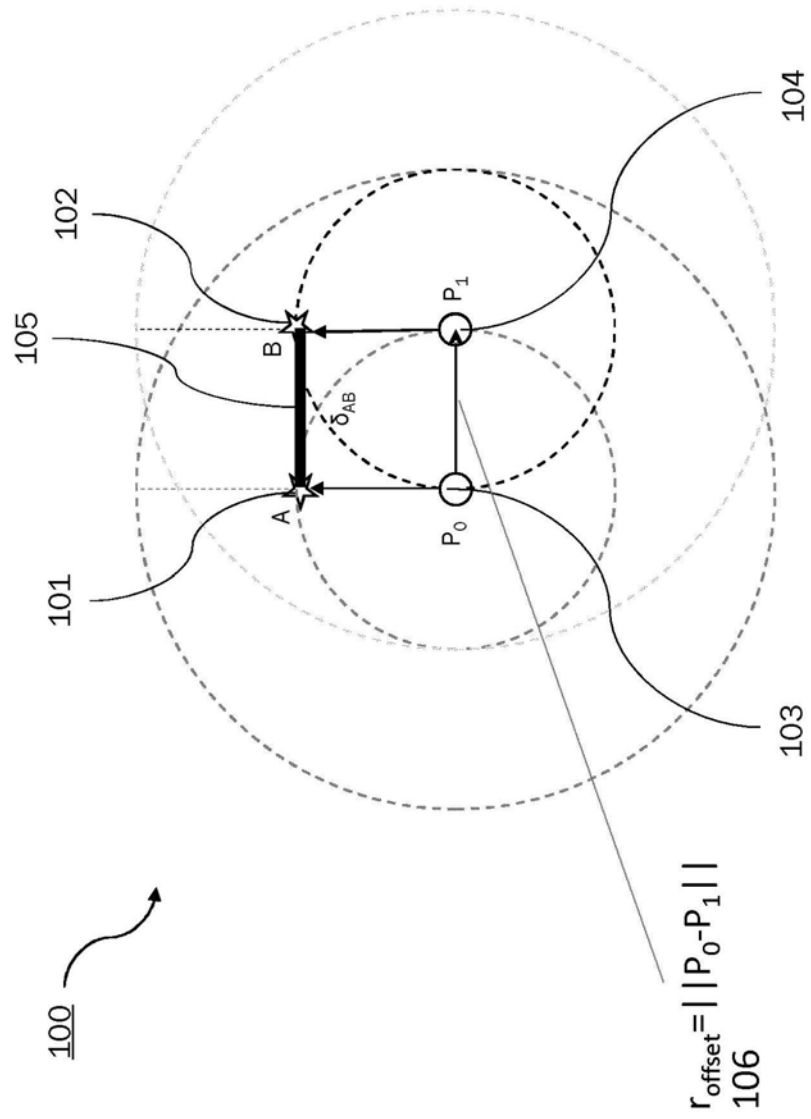


图1

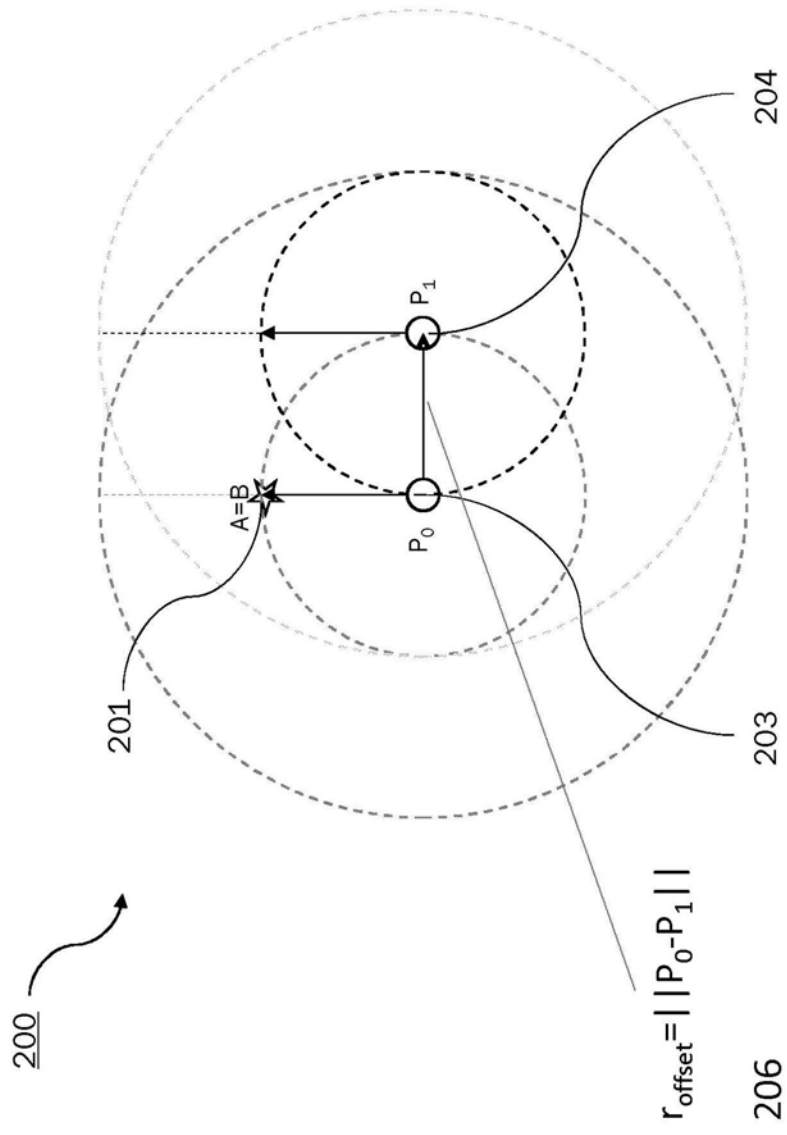


图2

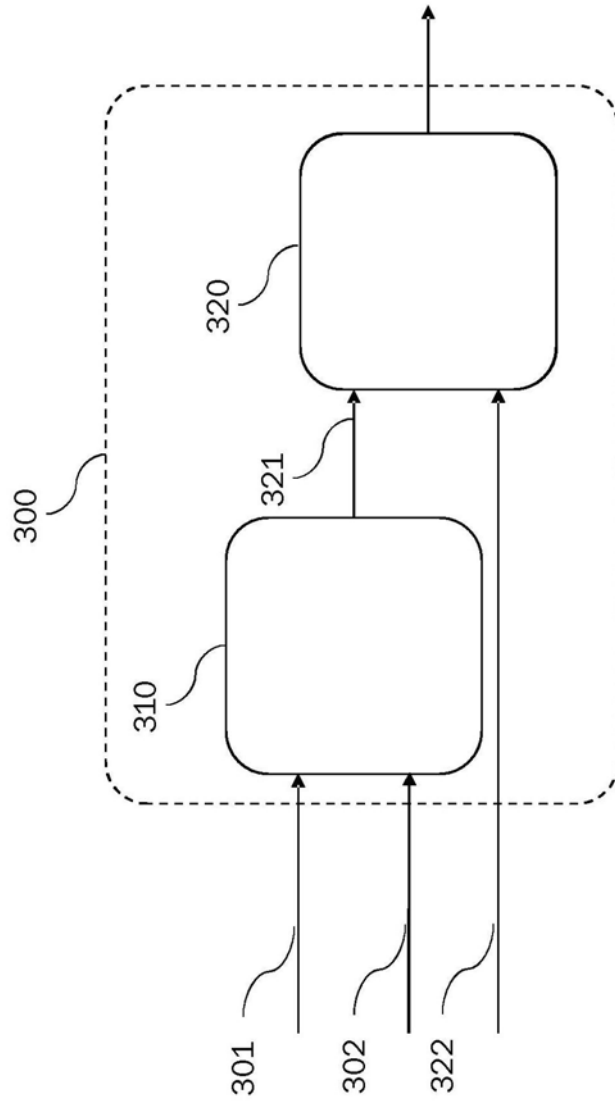


图3

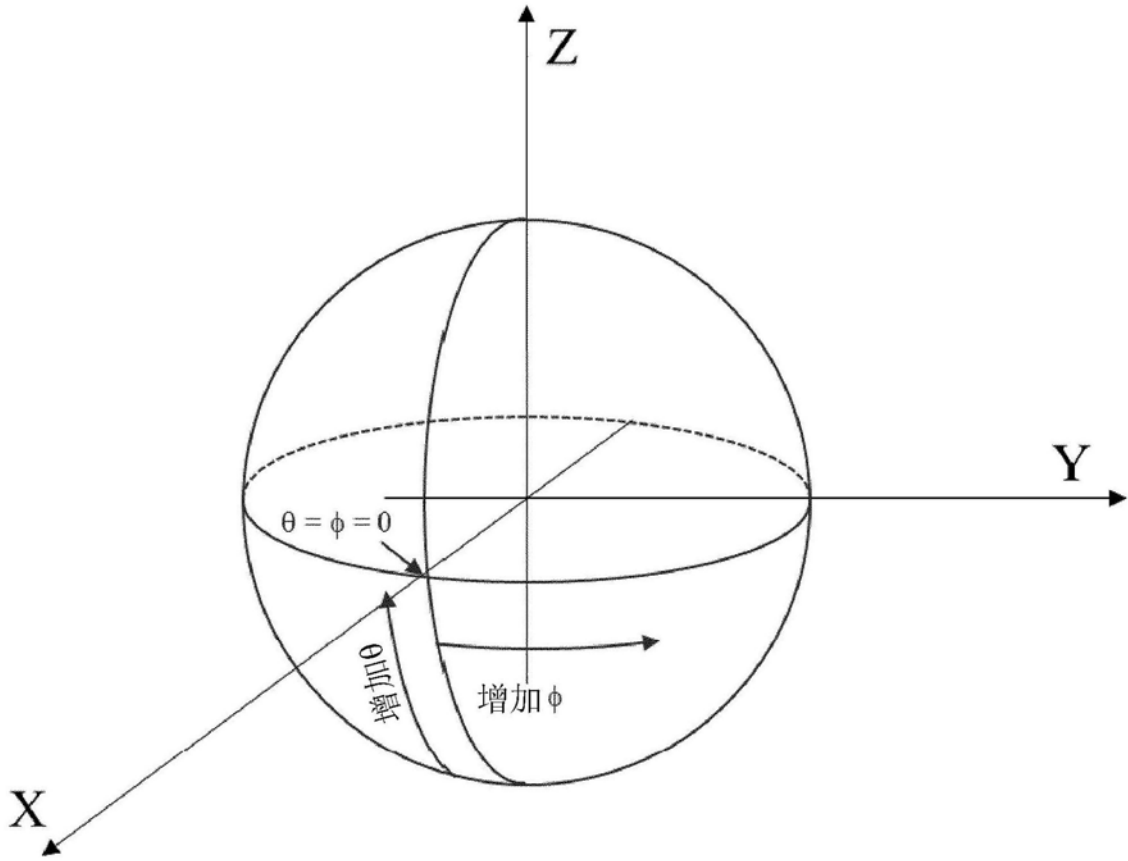


图4

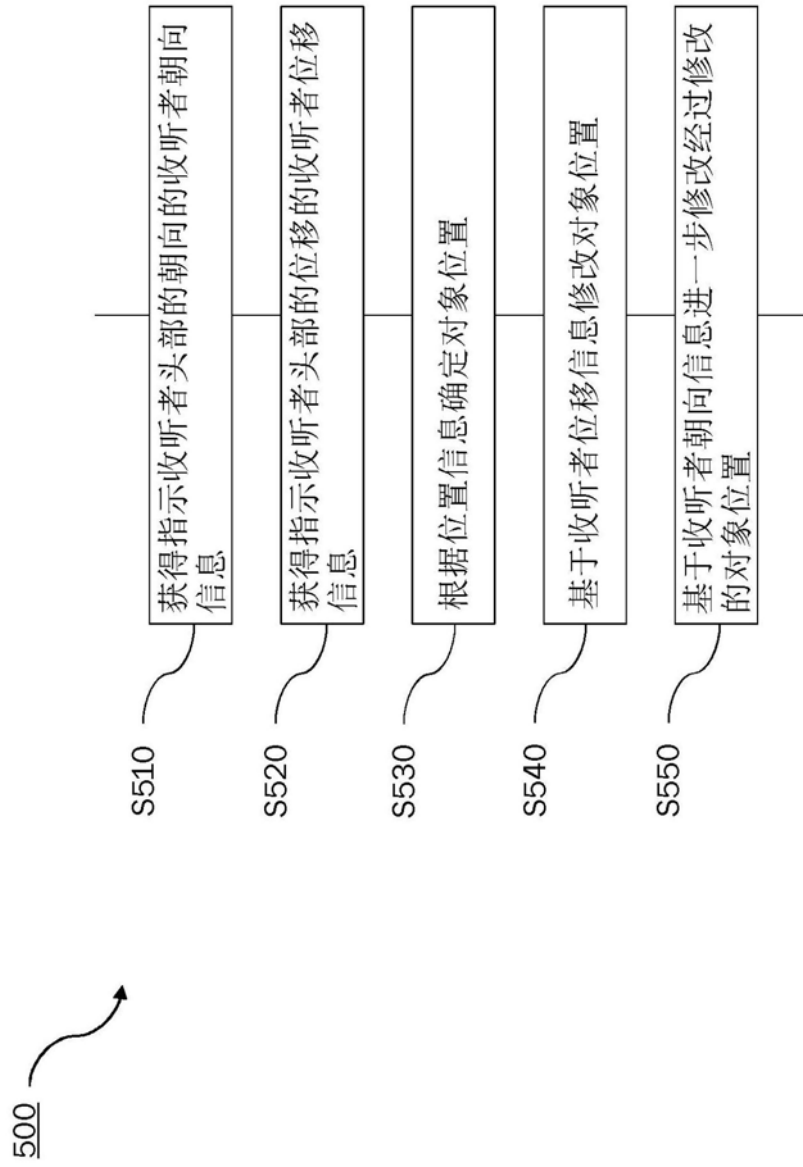


图5