(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112771479 B
(45)授权公告日 2024.06.25

(21)申请号 201980064097.3
(22)申请日 2019.09.12
(65)同一申请的已公布的文献号
申请公布号 CN 112771479 A
(43)申请公布日 2021.05.07
(30)优先权数据
62/742,324 2018.10.06 US
16/567,700 2019.09.11 US
(85)PCT国际申请进入国家阶段日
2021.03.29
(86)PCT国际申请的申请数据
PCT/US2019/050824 2019.09.12
(87)PCT国际申请的公布数据
WO2020/072185 EN 2020.04.09
(73)专利权人 高通股份有限公司
地址 美国加利福尼亚州
(72)发明人 M.Y.金 N.G.彼得斯
S.M.A.萨莱辛 S.G.斯瓦米纳坦
D.森
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 邓亚楠
(51)Int.Cl.
G06F 3/01(2006.01)
H04S 7/00(2006.01)
(56)对比文件
US 2018077451 A1,2018.03.15
US 2018098173 A1,2018.04.05
US 2018288558 A1,2018.10.04
审查员 马铭泽

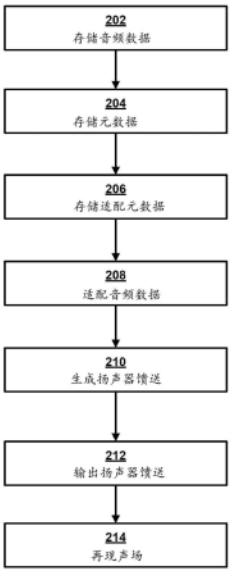
权利要求书2页 说明书19页 附图17页

(54)发明名称

六自由度和三自由度向后兼容性

(57)摘要

一种用于虚拟现实(VR)、混合现实(MR)、增强现实(AR)、计算机视觉和图形系统的向后兼容性的设备和方法。该设备和方法使得能够在支持较少自由度的设备上以更多自由度渲染音频数据。该设备包含存储器,其被配置为存储表示在多个捕获位置捕获的声场的音频数据、使音频数据能够被渲染以支持N个自由度的元数据、以及使音频数据能够被渲染以支持M个自由度的适配元数据。该设备还包含被耦接到该存储器上的一个或多个处理器,该处理器被配置为:基于该适配元数据,适配音频数据以提供M个自由度,以及基于适配的音频数据生成扬声器馈送。



CN 112771479 B

1.一种设备,包括:

存储器,被配置为存储表示在多个捕获位置捕获的声场的音频数据、使所述音频数据能够被渲染以支持N个自由度的元数据、以及使所述音频数据能够被渲染以支持M个自由度的适配元数据,其中,N是第一整数,M是不同于所述第一整数的第二整数;以及

一个或多个处理器,耦接到所述存储器,所述处理器被配置为:

通过使显示设备显示多个位置或轨迹,并从用户接收指示所述多个位置之一或指示所述轨迹的地点的输入,来确定用户位置;

将所述用户位置作为所述适配元数据存储到所述存储器;

基于所述适配元数据,适配所述音频数据以提供M个自由度;以及

基于适配的音频数据生成扬声器馈送。

2.根据权利要求1所述的设备,其中N等于6并且M小于N。

3.根据权利要求1所述的设备,还包括所述显示设备。

4.根据权利要求1所述的设备,其中,为了确定所述用户位置,所述一个或多个处理器被配置为:

基于所述轨迹的所述地点,选择多个位置之一作为所述用户位置。

5.根据权利要求2所述的设备,其中,所述6个自由度包括在二维空间坐标空间或三维空间坐标空间中限定的偏航、俯仰、翻滚和平移距离。

6.根据权利要求5所述的设备,其中,所述M个自由度包括三个自由度,所述三个自由度包括偏航、俯仰和翻滚。

7.根据权利要求5所述的设备,其中M等于0。

8.根据权利要求1所述的设备,其中,所述适配元数据包括所述用户位置和用户方位,并且其中,所述一个或多个处理器被配置为基于所述用户位置和所述用户方位来适配所述音频数据以提供零个自由度。

9.根据权利要求1所述的设备,其中所述适配元数据包括所述用户位置,并且其中所述一个或多个处理器被配置为:

基于所述用户位置,确定提供M个自由度的效果矩阵;以及

将所述效果矩阵应用于所述音频数据以适配所述声场。

10.根据权利要求9所述的设备,其中,所述一个或多个处理器还被配置为将所述效果矩阵乘以渲染矩阵以获得更新的渲染矩阵,并且其中,所述一个或多个处理器还被配置为将所述更新的渲染矩阵应用于所述音频数据以:

提供M个自由度;以及

生成所述扬声器馈送。

11.根据权利要求1所述的设备,其中,所述一个或多个处理器还被配置为获得指定所述适配元数据的比特流,所述适配元数据包括与所述音频数据相关联的所述用户位置。

12.根据权利要求1所述的设备,其中,所述一个或多个处理器还被配置为获得指示与所述设备接口连接的用户的旋转头部运动的旋转指示,并且其中,所述一个或多个处理器还被配置为基于所述旋转指示和所述适配元数据来适配所述音频数据以提供三个自由度。

13.根据权利要求1所述的设备,其中,所述一个或多个处理器耦接到可穿戴设备的扬声器,其中,所述一个或多个处理器被配置为将双耳渲染器应用于适配的更高阶立体混响

声音频数据以生成所述扬声器馈送,并且其中,所述一个或多个处理器还被配置为将所述扬声器馈送输出到所述扬声器。

14.根据权利要求13所述的设备,其中,所述可穿戴设备包括手表、眼镜、耳机、增强现实(AR)头戴式耳机、虚拟现实(VR)头戴式耳机或扩展现实(XR)头戴式耳机。

15.根据权利要求1所述的设备,其中,所述音频数据包括与具有一阶或更低阶的球基函数相关联的更高阶立体混响声系数、与具有混合阶和子阶的球基函数相关联的更高阶立体混响声系数、或者与具有大于一的阶数的球基函数相关联的更高阶立体混响声系数。

16.根据权利要求1所述的设备,其中,所述音频数据包括一个或多个音频对象。

17.根据权利要求1所述的设备,还包括:

被配置为基于所述扬声器馈送再现所述声场的一个或多个扬声器。

18.根据权利要求17所述的设备,其中,所述设备是车辆、机器人或手机中的一个。

19.根据权利要求1所述的设备,其中,所述一个或多个处理器包括处理电路。

20.根据权利要求19所述的设备,其中,所述处理电路包括一个或多个专用集成电路。

21.一种方法包括:

存储表示在多个捕获位置捕获的声场的音频数据;

存储使所述音频数据能够被渲染以支持N个自由度的元数据;

存储使所述音频数据能够被渲染以支持M个自由度的适配元数据,其中N是第一整数,M是不同于第一整数的第二整数;

通过使显示设备显示多个位置或轨迹,并从用户接收指示所述多个位置之一或指示所述轨迹的地点的输入,来确定用户位置;

将所述用户位置作为所述适配元数据进行存储;

基于所述适配元数据,适配所述音频数据以提供M个自由度;以及

基于所述适配的音频数据,生成扬声器馈送。

22.一种设备,包括:

用于存储表示在多个捕获位置捕获的声场的音频数据的部件;

用于存储使所述音频数据能够被渲染以支持N个自由度的元数据的部件;

用于存储使所述音频数据能够被渲染以支持M个自由度的适配元数据的部件,其中N是第一整数,并且M是不同于第一整数的第二整数;

用于通过使显示设备显示多个位置或轨迹,并从用户接收指示所述多个位置之一或指示所述轨迹的地点的输入,来确定用户位置的部件;

用于将所述用户位置作为所述适配元数据进行存储的部件;

用于基于所述适配元数据,适配所述音频数据以提供M个自由度的部件;以及

用于基于适配的音频数据,生成扬声器馈送的部件。

六自由度和三自由度向后兼容性

[0001] 交叉引用

[0002] 本申请要求于2019年9月11日提交的美国申请号16/567,700的优先权,其要求于2018年10月6日提交的美国临时申请号62/742,324的权益,通过引用将其全部内容并于本文。

技术领域

[0003] 本公开涉及媒体数据的处理,诸如音频数据的处理。

背景技术

[0004] 近年来,人们对增强现实(AR)、虚拟现实(VR)和混合现实(MR)技术越来越感兴趣。无线空间中图像处理和计算机视觉技术的进步已导致更好的渲染(render)和计算资源分配,以改善这些技术的视觉质量和身临其境的视觉体验。

[0005] 在VR技术中,可以使用头戴式显示器向用户呈现虚拟信息,使得用户可以在他们眼前的屏幕上从视觉上体验人造世界。在AR技术中,通过附加或叠加在现实世界中的物理对象上的视觉对象来增强现实世界。增强可以将新的视觉对象插入到现实世界环境中,或用视觉对象掩饰现实世界环境。在MR技术中,很难分辨出现实或合成/虚拟和用户的视觉体验之间的界限。

发明内容

[0006] 本公开总体上涉及以计算机为媒介的现实系统的用户体验的听觉方面,包含虚拟现实(VR)、混合现实(MR)、增强现实(AR)、计算机视觉和图形系统。更具体地,该技术可以使得能够在支持少于五个自由度的设备或系统上渲染用于VR、MR、AR等的音频数据,该音频数据占五个或更多个自由度。作为一个示例,在头部运动方面或在支持零个自由度的设备或系统上,该技术可以使得能够在仅支持三自个由度(偏航、俯仰和翻滚)的设备或系统上渲染占六个自由度(空间中的偏航、俯仰和翻滚加上用户的x、y和z平移)的音频数据。

[0007] 在一个示例中,一种设备包括:存储器,其被配置为存储表示在多个捕获位置捕获的声场的音频数据、使音频数据能够被渲染以支持N个自由度的元数据、以及使音频数据能够被渲染以支持M个自由度的适配元数据,其中,N是第一整数,M是不同于第一整数的第二整数,以及被耦接到存储器的一个或多个处理器,该一个或多个处理器被配置为:基于该适配元数据,适配音频数据以提供M个自由度;以及基于适配的音频数据生成扬声器馈送。

[0008] 在另一个示例中,一种方法包括:存储表示在多个捕获位置捕获的声场的音频数据;存储使音频数据能够被渲染以支持N个自由度的元数据;存储使音频数据能够被渲染以支持M个自由度的适配元数据,其中N是第一整数,M是不同于第一整数的第二整数;基于适配元数据,适配音频数据以提供M个自由度;以及基于适配的音频数据生成扬声器馈送。

[0009] 在又一个示例中,一种设备包括:用于存储表示在多个捕获位置捕获的声场的音频数据的部件;用于存储使音频数据能够被渲染以支持N个自由度的元数据的部件;用于存

储使音频数据能够被渲染以支持M个自由度的适配元数据的部件,其中N是第一整数,M是不同于第一整数的第二整数;用于基于该适配元数据,适配音频数据以提供M个自由度的部件;以及用于基于适配的音频数据生成扬声器馈送的部件。

[0010] 在附图和以下说明书中将详细阐述本公开的一个或多个示例。根据说明书和附图以及权利要求书,该技术的各个方面的其他特征、目的和优点将是显而易见的。

附图说明

[0011] 图1是示出各种阶数和子阶数的球谐基函数的图。

[0012] 图2A和图2B是示出可以执行本公开中描述的技术的各个方面的系统的图。

[0013] 图3是示出用户佩戴的VR设备的示例的图。

[0014] 图4是示出六自由度(6-DOF)头部运动方案的图。

[0015] 图5是更详细地示出在执行本公开中描述的效果技术的各个方面时图2A和2B中所显示的音频回放系统的框图。

[0016] 图6是更详细地示出图5的示例中所显示的效果单元如何根据本公开中描述的技术的各个方面获得效果矩阵的图。

[0017] 图7是示出图5所显示的深度图的图,根据本公开中描述的技术的各个方面,已经对图5进行了更新以反映锚点到深度图的映射。

[0018] 图8是描绘根据本公开的技术的流程图。

[0019] 图9是显示图2A和图2B的系统可如何处理音频数据的图。

[0020] 图10是示出图2A和图2B的系统可根据本公开的技术如何处理音频数据的图。

[0021] 图11是示出图2A和图2B的系统可根据本公开的技术如何处理音频数据的图。

[0022] 图12是示出图2A和图2B的系统可根据本公开的技术如何处理音频数据的图。

[0023] 图13是示出图2A和图2B的系统可根据本公开的技术如何处理音频数据的图。

[0024] 图14是示出可根据本公开中描述的技术的各个方面进行操作的可穿戴设备的示例的图。

[0025] 图15A和15B是示出可以执行本公开中描述的技术的各个方面的其他示例系统的图。

具体实施方式

[0026] 下面参考附图描述了本公开的特定实施方式。在说明书中,在整个附图中,共同的特征采用共同的附图标记指定。如本文中所使用的,各种术语仅出于描述特定实施方式的目的而并非旨在进行限制。例如,单数形式的“一”,“一个”和“该”也旨在包含复数形式,除非上下文另外明确指出。可以进一步理解,术语“包括”可以与“包含”互换使用。附加地,将理解的是,术语“其中”可以与“在此”互换使用。如本文中所使用的,“示例性”可以指示示例、实施方式和/或方面,并且不应被解释为限制或指示偏好或占优选实施方式。如本文中所使用的,用于修饰诸如结构、组件、操作等元素的序数术语(例如,“第一”、“第二”、“第三”等)本身并不指示该元素针对另一个元素的任何优先级或顺序,而仅仅是将该元素与具有相同名称(但使用序数术语)的另一个元素区分开。如本文中所使用的,术语“集合”是指一个或多个元素的分组,并且术语“多个”是指若干个元素。

[0027] 如本文中所使用的,“耦接”可以包含“通信耦接”、“电耦接”或“物理耦接”,并且还可以(或者可替代地)包含其任何组合。两个设备(或组件)可以直接地或经由一个或多个其他设备、组件、电线、总线、网络(例如,有线网络、无线网络、或其组合)等间接地耦接(例如通信耦接,电耦接或物理耦接)。如例示性的和非限制性的示例所示,电耦接的两个设备(或组件)可以包含在同一个设备中,也可以包含在不同的设备中,并且可以经由电子产品、一个或多个连接器或感应耦接进行连接。在一些实施方式中,通信耦接(诸如以电通信方式耦接)的两个设备(或组件),可以直接或间接地(诸如,经由一条或多条电线、总线、网络等)发送和接收电信号(数字信号或模拟信号)。如本文中所使用的,“直接耦接”可以包含两个设备耦接(例如,通信耦接、电耦接或物理耦接)但没有中间组件。

[0028] 如本文中所使用的,“集成的”可以包含“与之制造或一起出售”。如果用户购买了捆绑设备的封装或将该设备包含在内作为封装的一部分的封装,则该设备可以是集成的。在一些描述中,两个设备可以被耦接,但是不必被集成(例如,不同的外围设备可以不被集成到命令设备,但是仍然可以被“耦接”)。另一个示例可以是本文描述的任何收发器或天线可以“耦接”到处理器,但不一定是包含AR、VR或MR设备的封装的一部分。当使用术语“集成的”时,可以从本文公开的上下文(包含本段)推断出其他示例。

[0029] 如本文中所使用的,设备之间的“无线”连接可以基于各种无线技术,诸如蓝牙、无线保真(Wi-Fi)或Wi-Fi的变体(例如,Wi-Fi Direct)。设备可以基于不同的蜂窝通信系统,诸如长期演进(LTE)系统、码分多址(CDMA)系统、全球移动通信系统(GSM)、无线局域网(WLAN)系统或其他一些无线系统进行“无线连接”。CDMA系统可以实施宽带CDMA(WCDMA)、CDMA 1X、演进-数据优化(EVDO)、时分同步CDMA(TD-SCDMA)或CDMA的其他版本。另外,当两个设备在视线内时,“无线连接”也可以基于其他无线技术,诸如信号处理(例如,音频信号处理或射频处理)中使用的超声、红外、脉冲射频电磁能、结构光或到达方向技术。

[0030] 如本文中所使用的,A“和/或”B可以意味着“A和B”或“A或B”,或“A和B”与“A或B”两者均适用或可接受。

[0031] 术语“计算设备”在本文中一般指的是服务器、个人计算机、膝上型计算机、平板计算机、移动设备、蜂窝电话、智能本、超极本、掌上型计算机、个人数据助理(PDA)、无线电子邮件接收器、支持多媒体互联网的蜂窝电话、全球定位系统(GPS)接收器、无线游戏控制器以及类似的电子设备中的任何一个或全部,其中类似的电子设备包含可编程处理器和用于无线发送和/或接收信息的电路。

[0032] 市场上有各种基于“环绕声”信道的格式,范围例如从5.1家庭影院系统(在立体声基础上进军到客厅方面是最成功的)到由NHK(日本广播协会或日本广播公司)开发的22.2系统。内容创建者(例如好莱坞工作室)愿意为电影制作一次配音,而不愿意花费精力为每种潜在的扬声器配置针对配音进行再混合。运动图像专家组(MPEG)已经发布了标准,该标准允许使用元素的分层集合(例如,高阶立体混响声(Higher-Order Ambisonic)-HOA-系数)来表示声场,对于大多数扬声器配置(包含5.1和22.2配置),无论是在各种标准定义的位置还是在不均匀的位置,该元素都可以被渲染给扬声器馈送。

[0033] MPEG将该标准发布为MPEG-H 3D音频标准,正式标题为“Information technology-High efficiency coding and media delivery in heterogeneous environments-Part 3:3D audio”,由ISO/IEC JTC 1/SC 29提出,带有文档标识符ISO/IEC

DIS23008-3,并且日期为2014年7月25日。MPEG还发布3D音频标准的第二版,标题为“Information technology-High efficiency coding and media delivery in heterogeneous environments-Part 3:3D audio”,由ISO/IEC JTC 1/SC 29提出,带有文档标识符ISO/IEC 23008-3:201x(E),并且日期为2016年10月12日。在本公开中对“3D音频标准”的引用可以指的是以上标准中的一个或两个。

[0034] 如上所述,元素的分层集合的一个示例是球谐系数(spherical harmonic coefficient,SHC)集合。以下表达式展示了使用SHC对声场的描述或表示:

[0035] $$p_i(t,r_r,\theta_r,\varphi_r)=\sum_{\omega=0}^{\infty}\left[4\pi\sum_{n=0}^{\infty}j_n(kr_r)\sum_{m=-n}^{n}A_n^m(k)\,Y_n^m(\theta_r,\varphi_r)\right]e^{j\omega t},$$

[0036] 该表达式显示,在时间t处,声场任何一点$\{r_r,\theta_r,\varphi_r\}$的压力$p_i$都可以由SHC,即$A_n^m(k)$唯一地表示。在此,$k=\frac{\omega}{c}$,c是声速($\backsim 343\ \mathrm{m/s}$),$\{r_r,\theta_r,\varphi_r\}$是参考点(或观察点),$j_n(\bullet)$是n阶的球贝塞尔函数,并且$Y_n^m(\theta_r,\varphi_r)$是n阶和m子阶的球谐基函数(也可以称为球基函数)。可以认识到,方括号中的项是信号(即$S(\omega,r_r,\theta_r,\varphi_r)$)的频域表示,可以通过各种时频变换(诸如离散傅里叶变换(DFT)、离散余弦变换(DCT)或小波变换)对信号进行近似计算。分层集合的其他示例包含小波变换系数的集合和多分辨率基函数的系数的其他集合。

[0037] 图1是示出从零阶(n=0)到四阶(n=4)的球谐基函数的图。可以看出,对于每个阶数,存在子阶m的扩展,为便于说明,这些子阶在图1的示例中被显示但是未被明确地标记。

[0038] 可以通过各种麦克风阵列配置物理地获取(例如,记录)SHC $A_n^m(k)$,或者,可替换地,它们可以从声场的基于信道或基于对象的描述中导出。SHC(也可以称为高阶立体混响声-HOA-系数)表示基于场景的音频,其中SHC可以被输入到音频编码器以获得可以促进更高效的传输或存储的编码的SHC。例如,可以使用涉及$(1+4)^2$(25,因此是四阶)系数的四阶表示。

[0039] 如上所述,可以从使用麦克风阵列的麦克风记录中导出SHC。Poletti,M.在音频工程协会杂志2005年11月,53卷第11期,第1004-1025页刊登的“Three-Dimensional Surround Sound Systems Based on Spherical Harmonics”,对可以如何从麦克风阵列中导出SHC的各种示例进行了描述。

[0040] 为了说明可以如何从基于对象的描述中导出SHC,考虑以下等式。对应于单个音频对象的声场的系数$A_n^m(k)$可以表示为:

[0041] $$A_n^m(k)=g(\omega)(-4\pi ik)h_n^{(2)}(kr_s)Y_n^{m^*}(\theta_s,\varphi_s),$$

[0042] 其中i是$\sqrt{-1}$,$h_n^{(2)}(\bullet)$是n阶的球汉克尔函数(第二类),并且$\{r_s,\theta_s,\varphi_s\}$是对象的位置。知道对象源能量g(ω)作为频率函数(例如,使用时频分析技术,诸如对PCM流执行快速傅里叶变换),使得可以将每个PCM对象和相应的位置转换为SHC $A_n^m(k)$。此外,由于上面的是线性和正交分解,因此可以显示每个对象的$A_n^m(k)$系数是可加的。以

这种方式,许多PCM对象可以由$A_n^m(k)$系数表示(例如,作为单个对象的系数向量之和)。本质上,系数包含有关声场的信息(压力作为3D坐标的函数),并且上述表示在观察点$\{r_r,\theta_r,\varphi_r\}$附近从单个对象到整个声场表示的转换。其余图在下面基于SHC的音频译码的背景进行描述。

[0043] 图2A和图2B是示出可以执行本公开中所述技术的各个方面的系统的图。如在图2A的示例中显示,系统10包含源设备12和内容消费者设备14。尽管在源设备12和内容消费者设备14的背景下进行了描述,但是可以在任何背景下实施这些技术,其中声场的任何一种表示(包含基于场景的音频数据——诸如HOA系数、基于对象的音频数据以及基于信道的音频数据)都可以被编码以形成表示音频数据的比特流。

[0044] 而且,源设备12可以表示能够生成声场表示的任何形式的计算设备。尽管源设备12可以采用其他形式,但是在本文中,源设备12通常在作为VR内容创建者设备的背景下进行描述。同样,内容消费者设备14可以表示能够实施本公开中描述的技术以及音频回放的任何形式的计算设备。在本文中,内容消费者设备14通常在作为VR客户端设备的背景下进行描述,但是可以采用其他形式。

[0045] 源设备12可以由娱乐公司或生成多信道音频内容以供诸如内容消费者设备14之类的内容消费者设备的操作者消费的其他实体来操作。在许多VR场景中,源设备12结合视频内容生成音频内容。源设备12包含内容捕获设备300和内容捕获辅助设备302。内容捕获设备300可以被配置为与麦克风5接口连接或以其他方式与麦克风5通信。麦克风5可以表示能够捕获声场并将其表示为音频数据11的Eigenmike®或其他类型的3D音频麦克风。

[0046] 在一些示例中,内容捕获设备300可以包含集成到内容捕获设备300的外壳中的集成麦克风5。内容捕获设备300可以与麦克风5无线地接口连接或经由有线连接来接口连接。不是经由麦克风5捕获音频数据或与其结合捕获音频数据,而是在经由某种类型的可移动存储,无线地和/或经由有线输入过程输入音频数据11之后,内容捕获设备300可以处理音频数据11。这样,根据本公开,内容捕获设备300和麦克风5的各种组合都是可能的。

[0047] 内容捕获设备300也可以被配置为与声场表示生成器302接口连接或以其他方式与麦克风5通信。声场表示生成器302可以包含能够与内容捕获设备300接口连接的任何类型的硬件设备。声场表示生成器302可以使用由内容捕获设备300提供的音频数据11来生成由音频数据11表示的相同声场的各种表示。举例来说,为了使用音频数据11生成声场的不同表示,声场表示生成器302可以使用用于声场的立体混响声表示的译码方案,称为混合阶立体混响声(Mixed Order Ambisonics,MOA),如2017年8月8日提交并于2019年9月3日授权的美国专利10,405,126题为"MIXED-ORDER AMBISONICS(MOA)AUDIO DATA FOR COMPUTER-MEDIATED REALITY SYSTEMS"中更详细地讨论的。

[0048] 为了生成声场的特定MOA表示,声场表示生成器302可以生成完整的HOA系数集合的部分子集。举例来说,由声场表示生成器302生成的每个MOA表示可以提供针对声场的一些区域的精度,但是在其他区域中的精度较低。在一个示例中,声场的MOA表示可以包含八(8)个未压缩的HOA系数,而同一声场的三阶HOA表示可以包含十六(16)个未压缩的HOA系数。这样,作为HOA系数的部分子集生成的声场的每个MOA表示,可能比相应的从HOA系数生成的相同声场的三阶HOA表示的存储强度和带宽强度低(如果并且当作为通过所示出的传

输信道的比特流21的部分进行传输时)。

[0049] 尽管针对MOA表示进行了描述,但是本公开的技术也可以针对全阶立体混响声(full-order ambisonic,FOA)表示来执行,其中给定阶数N的所有HOA系数都用于表示声场。换句话说,不是使用音频数据11的部分非零子集来表示声场,而是声场表示生成器302可以针对给定阶数N使用所有音频数据11来表示声场,从而得到总共等于$(N+1)^2$的HOA系数。

[0050] 在这方面,高阶立体混响声音频数据11可以包含与具有一阶或更少阶的球基函数相关联的高阶立体混响声系数11(其可以被称为“第一阶立体混响声音频数据11”)、与具有混合阶和子阶的球基函数相关联的高阶立体混响声系数(可以被称为上述“MOA表示”)、或与具有大于一的阶数的球基函数相关联的高阶立体混响声系数(其是上面提及的“FOA表示”)。

[0051] 在一些示例中,内容捕获设备300可以被配置为与声场表示生成器302进行无线通信。在一些示例中,内容捕获设备300可以经由无线连接或有线连接中的一者或两者与声场表示生成器302进行通信。经由内容捕获设备300与声场表示生成器302之间的连接,内容捕获设备300可以以各种内容形式来提供内容,为了进行讨论,这里将其描述为音频数据11的一部分。

[0052] 在一些示例中,内容捕获设备300可以利用声场表示生成器302的各个方面(就声场表示生成器302的硬件或软件能力而言)。例如,声场表示生成器302可以包含专用硬件,配置为(或专门软件,在被执行时使一个或多个处理器)进行心理声学的音频编码(诸如由运动图像专家组(MPEG)或MPEG-H 3D音频编码标准提出的表示为“USAC”的统一语音和音频编码器)。内容捕获设备300可以不包含心理声学的音频编码器专用硬件或专门软件,而是以非心理声学的音频编码的形式提供内容301的音频方面。声场表示生成器302可以通过至少部分地针对内容301的音频方面执行心理声学的音频编码来辅助内容301的捕获。

[0053] 声场表示生成器302还可通过至少部分地基于从音频数据11生成的音频内容(例如,MOA表示和/或三阶HOA表示)来生成一个或多个比特流21,来辅助内容捕获和传输。比特流21可以表示音频数据11的压缩版本和任何其他不同类型的内容301(诸如球形视频数据、图像数据或文本数据的压缩版本)。

[0054] 作为一个示例,声场表示生成器302可以生成比特流21以用于跨传输信道传输,该传输信道可以是有线或无线信道、数据存储设备等。比特流21可以表示音频数据11的编码版本,并且可以包含主要比特流和另一个侧比特流,这可以称为侧信道信息。在某些情况下,表示音频数据的压缩版本的比特流21可以符合根据MPEG-H 3D音频编码标准产生的比特流。

[0055] 内容消费者设备14可以由个人操作并且可以表示VR客户端设备。尽管针对VR客户端设备进行了描述,但内容消费者设备14可以表示其他类型的设备,诸如增强现实(AR)客户端设备、混合现实(MR)客户端设备、标准计算机、头戴式耳机、耳机、或能够跟踪操作客户端消费者设备14的个人的头部运动和/或一般平移运动的任何其他设备。如在图2A的示例中显示,内容消费者设备14包含音频回放系统16,其可以指的是能够渲染音频数据的任何形式的音频回放系统,包含SHC(无论是以三阶HOA表示和/或MOA表示的形式)、音频对象和音频信道中的一个或多个,以作为多信道音频内容进行回放。

[0056] 虽然在图2A中显示比特流21被直接传输到内容消费者设备14,但源设备12可以将比特流21输出到位于源设备12与内容消费者设备14之间的中间设备。中间设备可以存储比特流21,以稍后传递到内容消费者设备14,内容消费者设备14可以请求比特流。中间设备可以包括文件服务器、web服务器、台式计算机、膝上型计算机、平板计算机、移动电话、智能电话或能够存储比特流21以供稍后由音频解码器取得的任何其他设备。中间设备可以驻留在内容传递网络中,该内容传递网络能够向诸如内容消费者设备14之类的请求比特流21的订户流传输比特流21(并且可能与传输相应的视频数据比特流相结合)。

[0057] 可替换地,源设备12可以将比特流21存储到诸如光盘、数字视频光盘、高清晰度视频光盘或其他存储介质的存储介质,其中大多数存储介质能够被计算机读取,因此,可以将其称为计算机可读存储介质或非暂时性计算机可读存储介质。在此上下文中,传输信道可以指的是用来传输存储在介质上的内容的信道(并且可以包含零售商店和其他基于商店的传递机制)。因此,无论如何,本公开的技术不应在这方面限于图2A的示例。

[0058] 如上所述,内容消费者设备14包含音频回放系统16。音频回放系统16可以表示能够回放基于信道的音频数据的任何系统。音频回放系统16可以包含许多不同的渲染器22。每个渲染器22可以提供不同形式的渲染,其中不同形式的渲染可以包含执行向量基幅度相移(vector-base amplitude panning,VBAP)的各种方式中的一种或多种,和/或执行声场合成的各种方式中的一种或多种。如本文所用,“A和/或B”意味着“A或B”或“A和B”两者。

[0059] 音频回放系统16可以进一步包含音频解码设备24。音频解码设备24可以表示被配置为对比特流21进行解码以输出音频数据15(再次,作为一个示例,其可以包含形成完整的三阶HOA表示的HOA或其形成相同声场或其分解的MOA表示的子集,诸如占优音频信号、环境HOA系数和MPEG-H3D音频编码标准中所述的基于向量的信号)的设备。这样,音频数据15可以类似于HOA系数的完整集合或部分子集,但是由于有损操作(例如,量化)和/或经由传输信道的传输而可能不同。音频回放系统16可以在对比特流21进行解码以获得音频数据15之后,对音频数据15进行渲染以输出扬声器馈送25。扬声器馈送25可以驱动一个或多个扬声器(为了便于说明,未在图2A的示例中显示)。声场的立体混响声表示可以以多种方式,包含N3D、SN3D、FuMa、N2D或SN2D被归一化。

[0060] 为了选择适当的渲染器,或者在一些情况中为了生成适当的渲染器,音频回放系统16可以获得指示扩音器的数量和/或扩音器的空间几何形状的扩音器信息13。在一些情况中,音频回放系统16可以使用参考麦克风并以动态确定扩音器信息13的方式驱动扩音器,来获得扩音器信息13。在其他情况下,或者结合扩音器信息13的动态确定,音频回放系统16可以提示用户与音频回放系统16接口连接并输入扩音器信息13。

[0061] 音频回放系统16可以基于扩音器信息13选择音频渲染器22之一。在一些情况下,当音频渲染器22中的任何一个都不在扩音器信息13中指定的扩音器几何形状的某个阈值相似性度量(就扩音器几何形状而言)内时,音频回放系统16可以基于扩音器信息13生成音频渲染器22之一。在一些情况下,音频回放系统16可以基于扩音器信息13生成音频渲染器22之一,而无需首先尝试选择音频渲染器22中的一个现有的音频渲染器。

[0062] 当将扬声器馈送25输出到耳机时,音频回放系统16可以利用渲染器22之一,该渲染器22使用头部相关传递函数(head-related transfer function,HRTF)或能够向左右扬声器馈送25渲染以进行耳机扬声器回放的其他函数来提供双耳渲染。术语“扬声器”或“换

能器”通常可以指的是任何扬声器,包含扩音器、耳机扬声器等。然后,一个或多个扬声器可以回放渲染的扬声器馈送25。

[0063] 尽管描述为根据音频数据11’渲染扬声器馈送25,但是对扬声器馈送25的渲染的引用可以指的是其他类型的渲染,诸如直接并入到来自比特流21的音频数据15的解码中的渲染。可替代的渲染的示例可以在MPEG-H 3D音频编码标准的附件G中找到,其中渲染在声场合成前、在占优信号形成和背景信号形成期间发生。这样,对音频数据15的渲染的引用应理解为指的是实际音频数据15的渲染或音频数据15的分解或其表示(诸如上述占优音频信号、环境HOA系数、和/或基于向量的信号-也可以称为V向量)的渲染。

[0064] 如上所述,内容消费者设备14可以表示其中人类可穿戴显示器安装在操作VR设备的用户的眼睛前面的VR设备。图3是示出用户402穿戴的VR设备400的示例的图。VR设备400被耦接到耳机404或以其他方式包含耳机404,耳机404可以通过扬声器馈送25的回放来再现由音频数据11’表示的声场。扬声器馈送25可以表示能够使耳机404的换能器内的膜以各种频率振动的模拟或数字信号,其中这种过程通常被称为驱动耳机404。

[0065] 视频、音频和其他感官数据可能会在VR体验中发挥重要作用。为了参与VR体验,用户402可以穿戴VR设备400(也可以称为VR头戴式耳机400)或其他可穿戴电子设备。VR客户端设备(诸如VR头戴式耳机400)可以跟踪用户402的头部运动,并且使经由VR头戴式耳机400显示的视频数据适应以考虑头部运动,从而提供身临其境的体验,在其中用户402可以体验以可视的三个维度显示在视频数据中的虚拟世界。

[0066] 尽管VR(以及AR和/或MR的其他形式)可以使用户402在视觉上驻留在虚拟世界中,但VR头戴式耳机400可能常常缺乏听觉上将用户置于虚拟世界中的能力。换句话说,VR系统(可能包含负责渲染视频数据和音频数据的计算机(为便于说明,在图3的示例中未显示)以及VR头戴式耳机400)可能无法支持完整的三维听觉沉浸。

[0067] 音频通常向用户提供零个自由度(zero degrees of freedom,0DOF),这意味着用户运动不会改变音频渲染。然而,VR可以为用户提供一些自由度,这意味着音频渲染可以根据用户运动而改变。VR的音频方面已经被分类为三个独立的沉浸类别。第一类别提供最低级的沉浸,称为三个自由度(3DOF)。3DOF指的是音频渲染,其考虑了头部在三个自由度(偏航、俯仰和翻滚)中的运动,从而使用户可以在任何方向上自由地环顾四周。然而,3DOF无法考虑其中头不在声场的光学和声学中心上的平移头部运动。

[0068] 第二类别称为3DOF加(3DOF+),除了由于头部运动远离声场的声学中心和光学中心而产生的有限的空间平移运动之外,3DOF+还提供了三个自由度(偏航、俯仰和翻滚)。3DOF+可以为诸如运动视差之类的感知效果提供支持,这可以增强沉浸感。

[0069] 第三类别称为六个自由度(6DOF),以一种在头部运动(偏航、俯仰和翻滚)方面考虑了三个自由度,但同时也考虑了用户在空间中的平移(x、y和z平移)的方式渲染音频数据。可以通过跟踪用户在物理世界中的位置的传感器或通过输入控制器来引入空间平移。

[0070] 图4是示出用于AVR和/或AR应用的六个自由度(6-DOF)头部运动方案的图。如图4显示,6-DOF方案包含3-DOF方案之外的三个附加运动线。更具体地,除了上面讨论的旋转轴之外,图4的6-DOF方案还包含三条线,用户的头部位置可以沿着这三条线平移运动或致动。三个平移方向是左右(L/R)、上下(U/D)和前后(F/B)。源设备12的音频编码设备和/或音频解码设备24可以实施视差处理以解决三个平移方向。举例来说,音频解码设备24可以应用

一个或多个传输因子调整各种前景音频对象的能量和/或方向信息,以基于VR/AR用户的6-DOF运动范围来实施视差调整。

[0071] 根据本公开的一个示例,源设备12可以生成表示在多个捕获位置捕获的声场的音频数据、使音频数据能够被渲染以支持至少五个自由度的元数据、以及使音频数据能够被渲染以支持少于五个自由度的适配元数据。内容消费者设备14可以接收和存储表示在多个捕获位置捕获的声场的音频数据、使音频数据能够被渲染以支持至少五个自由度的元数据、以及使音频数据能够被渲染以支持少于五个自由度的适配元数据。基于该适配元数据,内容消费者设备14可以适配音频数据以提供少于五个自由度,并且音频渲染器22可基于该适配的音频数据生成扬声器馈送。

[0072] 根据本公开的另一个示例,源设备12可以生成表示在多个捕获位置捕获的声场的音频数据、使音频数据能够被渲染以支持六个自由度的元数据、以及使音频数据能够被渲染以支持少于六个自由度的适配元数据。内容消费者设备14可以接收和存储表示在多个捕获位置捕获的声场的音频数据、使音频数据能够被渲染以支持六个自由度的元数据、以及使音频数据能够被渲染以支持少于六个自由度的适配元数据。基于该适配元数据,内容消费者设备14可以适配音频数据以提供少于六个自由度,并且音频渲染器22可以基于该适配的音频数据生成扬声器馈送。

[0073] 根据本公开的另一示例,内容消费者设备14可以存储表示在多个捕获位置捕获的声场的音频数据;确定用户位置;根据用户位置适配音频数据以提供M个自由度,其中M包括整数值;以及基于适配的音频数据生成扬声器馈送。为了确定用户位置,内容消费者设备14可以显示多个用户位置,并从用户接收指示多个位置之一的输入。为了确定用户位置,内容消费者设备14可以显示轨迹;从用户接收指示轨迹的地点的输入;并基于轨迹的地点选择多个位置之一作为用户位置。为了确定用户位置,内容消费者设备14可以检测用户的运动并基于该运动选择位置。内容消费者设备14可以基于来自多个位置的运动来选择位置。

[0074] 根据本公开的另一示例,内容消费者设备14可以存储表示在多个捕获位置捕获的声场的音频数据;适配音频数据以提供M个自由度;基于具有M个自由度的适配的音频数据生成扬声器馈送;适配音频数据以提供N个自由度;以及基于具有N个自由度的适配的音频数据生成扬声器馈送。内容消费者设备14可以进一步被配置为在基于具有M个自由度的适配的音频数据和基于具有N个自由度的适配的音频数据来生成扬声器馈送之间进行切换。例如,内容消费者设备14可以响应于用户输入或用户运动来执行这种切换。

[0075] 尽管针对如图3的示例中显示的VR设备进行了描述,但是,该技术可以由其他类型的可穿戴设备执行,包含手表(诸如所谓的“智能手表”)、眼镜(诸如所谓的“智能眼镜”)、耳机(包含经由无线连接耦接的无线耳机或经由有线或无线连接耦接的智能耳机)、以及任何其他类型的可穿戴设备。这样,可以通过任何类型的可穿戴设备来执行该技术,当用户穿戴该可穿戴设备时,用户可以通过该技术与可穿戴设备进行交互。

[0076] 图2B是示出可以执行本公开中描述的技术的各个方面的另一个示例系统100的框图。系统100类似于图2A显示的系统10,区别是图2A中显示的音频渲染器22被替换为能够使用一个或多个HRTF或能够向左右扬声器馈送103进行渲染的其他函数执行双耳渲染的双耳渲染器102。

[0077] 音频回放系统16可以将左右扬声器馈送103输出到耳机104,耳机104可以表示可

穿戴设备的另一个示例,并且可以耦接到附加可穿戴设备以促进声场的再现,诸如手表、上面提到的VR头戴式耳机、智能眼镜、智能服装、智能戒指、智能手镯或任何其他类型的智能珠宝(包含智能项链)等。耳机104可以无线地或经由有线连接耦接到附加可穿戴设备。

[0078] 附加地,耳机104可以经由有线连接(诸如标准的3.5mm音频插孔、通用系统总线(USB)连接、光学音频插孔或其他形式的有线连接)耦接至音频回放系统16,或无线地(诸如通过Bluetooth™连接、无线网络连接等)耦接至音频回放系统16。耳机104可以基于左右扬声器馈送103来重新创建由音频数据11表示的声场。耳机104可以包含左耳机扬声器和右耳机扬声器,它们由相应的左右扬声器馈送103供电(或者换句话说,被驱动)。

[0079] 图5是更详细地示出图2A和2B中所示的音频回放系统的框图。如在图5的示例中所显示,除了上述音频解码设备24之外,音频回放系统16还包含效果单元510和渲染单元512。效果单元510表示被配置为获得上述效果矩阵26(在图5的示例中显示为“EM 26”)的单元。渲染单元512表示被配置为确定和/或应用上述音频渲染22(在图5的示例中显示为“AR 22”)中的一个或多个的单元。

[0080] 如上所述,音频解码设备24可以表示被配置为根据MPEG-H 3D音频编码标准对比特流21进行解码的单元。音频解码设备24可以包含比特流提取单元500、逆增益控制和重新分配单元502、占优声音合成单元504、环境合成单元506和组合单元508。关于前述单元500-508中的每个的更多信息可以在MPEG-H 3D音频编码标准中找到。

[0081] 尽管在MPEG-H 3D音频编码标准中进行了详细描述,但是下面提供500-508中的每个单元的简要描述。比特流提取单元500可以表示被配置为提取音频数据的分解以及组成由音频数据11定义的声场的表示所需的其他语法元素或数据的单元。比特流提取单元500可以标识比特流11中的一个或多个输送信道501,每个输送信道可以指定环境音频信号或占优音频信号。比特流提取单元500可以提取输送信道501,并且将输送信道501输出到逆增益控制和重新分配单元502。

[0082] 尽管为了便于说明而在图5的示例中未显示,但是,音频解码设备24可以包含心理声学的音频解码器,该心理声学的音频解码器针对输送信道501执行心理声学的音频解码(例如,高级音频译码-AAC)。而且,音频解码设备24可以包含执行诸如在输送信道501之间进行衰减等的图5的示例中未显示的各种其他操作的更多单元。

[0083] 比特流提取单元500可以进一步提取侧信息(side information)521,该侧信息521定义语法元素和用于执行增益控制和分配的其他数据。比特流提取单元500可以将侧信息521输出到逆增益控制和重新分配单元502。

[0084] 比特流提取单元500还可以提取定义语法元素和用于执行占优声音合成的其他数据(包含定义在输送信道501中定义的相应的占优音频信号的空间特性(诸如宽度、方向和/或形状)的向量)的侧信息523。附加地,该比特流提取单元500可以提取侧信息525,该侧信息525定义语法元素和用于执行环境合成的其他数据。比特流提取单元500将侧信息523输出到占优声音合成单元504,并且将侧信息525输出到环境合成单元506。

[0085] 逆增益控制和重新分配单元502可以表示被配置为基于侧信息521执行针对输送信道501的逆增益控制和重新分配的单元。逆增益控制和重新分配单元502可以基于侧信息521确定增益控制信息,并将增益控制信息应用于每个输送信道501,以反转由声场表示生成302实施的应用在音频编码设备上的增益控制以减小输送信道501的动态范围。接下来,

逆增益控制和重新分配单元502可以基于侧信息523,确定每个输送信道501是指定占优音频信号503还是环境音频信号505。逆增益控制和重新分配单元502可以将占优音频信号503输出到占优声音合成单元504并且将环境音频信号505输出到环境合成单元506。

[0086] 占优声音合成单元504可以表示被配置为基于侧信息523来合成由音频数据11表示的声场的占优音频分量的单元。占优声音合成单元504可以将每个占优音频信号503乘以在侧信息523中指定的对应空间向量(也可以称为“基于向量的信号”)。占优声音合成单元504将乘法结果作为占优声音表示507输出到组合单元508。

[0087] 环境合成单元506可以表示被配置为基于侧信息525来合成由音频数据11表示的声场的环境分量的单元。环境合成单元506将合成结果作为环境声音表示509输出到组合单元508。

[0088] 组合单元508可以表示被配置为基于占优声音表示507和环境声音表示509来组成音频数据15的单元。在一些示例中,组合单元508可以将占优声音表示507添加到环境声音表示509以获得音频数据15。组合单元508可以将音频数据15输出到效果单元510。

[0089] 效果单元510可以表示被配置为执行本公开中描述的效果技术的各个方面,以基于平移距离17或者如下面更详细地描述的平移距离17和深度图513来生成EM 26的单元。

[0090] 从照片生成图5中显示的深度图513。深度图513中的区域越白,则其在拍摄照片时距离相机越近。深度图513中的区域越黑,则其在拍摄照片时距离相机越远。效果单元510可以将EM 26应用于音频数据15以获得适配的音频数据511。适配的音频数据511可以被适配以提供三个自由度加效果,其在声场中考虑由平移距离17指示的平移头部运动。效果单元510可以将适配的音频数据511输出到渲染单元512。

[0091] 渲染单元512可以表示被配置为将AR22中的一个或多个AR应用于适配的音频数据511,从而获得扬声器馈送25的单元。渲染单元512可以将扬声器馈送25输出到图3的示例中显示的耳机404。

[0092] 尽管被描述为分离单元510和512,但是效果单元510可以被合并在渲染单元512内,其中,EM 26以下面更详细描述的方式乘以AR 22中的所选择的一个。EM 26与AR 22中的所选择的一个相乘可以产生更新的AR(其可以被表示为“更新的AR 22”)。然后,渲染单元512可以将更新的AR 22应用于音频数据15以使音频数据15适配以提供考虑平移距离17的3DOF+效果并且渲染扬声器馈送25。

[0093] 图6是更详细地示出图5所示的示例中显示的效果单元如何根据本公开中描述的技术的各个方面获得效果矩阵的图。如在图6的示例中显示,用户402最初驻留在重建的声场600的中间,如图6中表示的“初始用户位置”的左面所示。重建的声场600,虽然显示为圆形,但被建模作为在参考距离602处围绕用户402的球体。在一些示例中,用户402可以在配置VR设备14以回放音频数据时输入参考距离602。在其他示例中,参考距离602是静态的,或者被定义为比特流21的语法元素。当使用语法元素定义时,参考距离602可以是静态的(诸如,单次发送,因此在体验期间是静态的)或动态的(诸如,在体验期间多次发送,例如,每个音频帧或每一些周期性或非周期性数量的音频帧)。

[0094] 效果单元510可以接收参考距离602并且在平移头部运动606之前确定位于距用户402的头部参考距离602处的锚点604。在图6的示例中显示锚点604作为“X”标记。效果单元510可以将锚点604确定为球形声场600表面上的多个个均匀分布的锚点,该球形声场具有

等于参考距离602的半径。在其他示例中,锚点可以由相机(未显示)确定,并且可以被提供给效果单元510。

[0095] 锚点604可以表示参考点,通过该参考点来确定平移头部运动606。换句话说,锚点604可以表示围绕球形声场600分布的参考点,通过该参考点可以确定平移头部运动606以适配该声场。锚点604不应与视觉图像搜索算法中理解的锚点或关键点混淆。此外,锚点604可以代表距用户402的头部参考距离处的参考点,用于确定相对于每个锚点604的平移头部运动606。相对于每个锚点604的平移头部运动606的程度可影响针对其中驻留锚点604的相应一个的声场部分的渲染。这样,锚点604也可以表示声场采样点,通过声场采样点来确定平移头部运动606并基于相对平移头部运动606来适配声场的渲染。

[0096] 无论如何,然后,用户402可以执行平移头部运动606,并如"平移运动后的用户位置"标题下的图6的示例所显示,使头部向右运动平移距离17。效果单元510可以在平移头部运动606之后确定相对于多个锚点604中的每一个的更新距离608。尽管在图6的示例中仅显示了单一更新距离608,效果单元510可以确定相对于每个锚点604的更新距离608。接下来,效果单元510可以基于每个更新距离608确定EM 26。

[0097] 效果单元510可以为每个平移的锚点计算距离相关的响度调整(以EM 26的形式)。针对每个参考点的计算可以表示为$g_{|}$,其中原始参考距离602表示为$dist_{ref}$,而更新距离608可以表示为$dist_{new,|}$。对于每个锚点604,效果单元510可以使用等式 $g_{|} = \left(\frac{dist_{ref}}{dist_{new,|}}\right)^{distPow}$ 计算$g_{|}$。distPow参数可以控制效果强度,其可以由用户402输入以控制效果强度的大小。虽然被描述为受用户402控制的变量,但是distPow参数也可以由内容创建者动态地或静态地指定。

[0098] 在数学上,围绕用户402的声场600可以表示为球体上的M个等距的锚点604(也可以称为"空间点604"),其中球体中心位于用户402的头部。通常选择变量"M"以使M大于或等于$(N+1)^2$,其中N代表与音频数据15相关联的最大阶数。

[0099] M个等距的空间点604导致从用户402的头部延伸到M个等距的空间点604中的每一个的M个空间方向。M个空间方向可以由$\boxtimes m$表示。效果单元510可以基于M个空间方向$\boxtimes m$获得应用于渲染矩阵的EM 26。在一个示例中,效果单元510获得从与M个空间方向中的每一个相关联的HOA系数计算出的EM 26。然后,效果单元510可以为空间方向1=1…M中的每一个执行响度补偿,其被应用于EM 26以生成补偿后的EM 26。虽然被描述为M个等距的空间点604,但是点604也可以是非等距的,或者换句话说,以非均匀的方式分布在球体周围。

[0100] 根据"DIS"版本的附件F.1.5的MPEG-H 3D音频编码标准所使用的变量,当响度补偿作为一个示例被讨论时,效果单元510可以如下从与M个空间方向相关联的HOA系数计算EM 26:

[0101] $$\tilde{F} = \left(\Psi^{(O,M)^T}\right) \dagger \left(\Psi_m^{(O,M)^T}\right)$$

[0102] 使用$\Psi^{(0,M)^T} := [S_1^O \; S_2^O \; \ldots \; S_M^O] \in \mathbb{R}^{O \times M}$

[0103] "†"符号可以代表伪逆矩阵运算。

[0104] 然后,效果单元510可以为空间方向1=1…M中的每一个执行响度补偿,其被根据

下面的等式应用于矩阵F：

[0105] $$A(l)=\sqrt{\frac{(RS_{m_l}^{O})^{T}(RS_{m_l}^{O})}{(\boldsymbol{R}\tilde{\boldsymbol{F}}S_l^{U})^{T}(\boldsymbol{R}\tilde{\boldsymbol{F}}S_l^{U})}}*g_l$$

[0106] 其中$\boldsymbol{F}=\left(\boldsymbol{\Psi}^{(O,M)^T}\right)\dagger diag(A)\left(\Psi_m^{(O,M)^T}\right).$

[0107] 然后，效果单元510可以将AR 22的所选的一个(在下文中由变量“R”表示)乘以EM 26(在上下文中由变量“F”表示)，以生成上文所讨论的更新的AR22(采用变量“D”表示)。

[0108] D＝RF

[0109] 当禁用了与距离有关的响度调整时，上述方法可以在数学上表示通过消除乘以g_l的乘法进行的与距离无关的响度调整，从而得到以下结果：

[0110] $$A(l)=\sqrt{\frac{(RS_{m_l}^{O})^{T}(RS_{m_l}^{O})}{(\boldsymbol{R}\tilde{\boldsymbol{F}}S_l^{U})^{T}(\boldsymbol{R}\tilde{\boldsymbol{F}}S_l^{U})}}$$

[0111] 在所有其他方面，当启用与距离无关的响度调整时(或换句话说，当与距离有关的响度调整被禁用时)，数学表示不变。

[0112] 用这种方法，效果单元510可以将EM 26提供给渲染单元512，其将音频渲染器22与补偿的EM 26相乘，以创建能够同时考虑三个自由度和平移头部运动606的适配空间渲染矩阵(在此称为“更新的AR 22”)，其中音频渲染器22将音频数据15从球谐域转换成空间域扬声器信号25(在这种情况下可以是将音频数据渲染为双耳音频耳机扬声器信号的双耳渲染器)。

[0113] 在一些情况中，效果单元510可以确定多个EM 26。例如，效果单元510可以确定用于第一频率范围的第一EM 26、用于第二频率范围的第二EM 26等等。第一EM 26和第二EM 26的频率范围可以重叠，或者可以不重叠(或换句话说，可能彼此不同)。这样，本公开中描述的技术不应限于单一EM 26，而应包含多个EM 26的应用，包含但不限于多个示例频率相关的EM 26。

[0114] 如上所述，效果单元510还可以基于平移距离17和深度图513确定EM 26。比特流21可以包含与音频数据16相对应的视频数据，其中这样的视频数据与音频数据16同步(使用例如帧同步信息)。尽管在图2-4的示例中未显示，客户端消费者设备14可以包含视频回放系统，该视频回放系统对提供视频数据的相应比特流进行解码，该视频数据可以包含深度图，诸如深度图513。如上所述，深度图513提供了360度虚拟现实场景的灰度表示，其中黑色表示非常远的距离，而白色表示近距离，灰色的各种阴影表示黑色和白色之间的中间距离。

[0115] 视频回放系统的视频解码设备可以利用深度图513从在视频比特流中指定的各个右眼视图或左眼视图来制定用于左眼或右眼的视图。视频解码设备可以基于深度图来更改右眼视图和左眼视图之间的横向距离的量，基于灰色阴影越深将横向距离缩放得更小。这样，深度图513中用白色或浅灰色阴影代表的近对象在左眼视图和右眼视图之间可具有较大的横向距离，而深度图513中用黑色或深灰色阴影代表的远对象在左眼视图和右眼视图之间可具有较小的横向距离(因此更接近类似于遥远的点)。

[0116] 效果单元510可以利用深度图513提供的深度信息来适配锚点604相对于用户402的头部位置。也就是说,效果单元510可以将锚点604映射到深度图513,并且利用深度图513内的映射位置处的深度图513的深度信息来标识针对每个锚点604的更准确的参考距离602。图7是示出图5所显示的深度图的图,根据本公开中描述的技术的各个方面,已经对图5进行了更新以反映锚点到深度图的映射。

[0117] 在这方面,不是假定单一参考距离602,而是效果单元510可以利用深度图513来估计用于每个锚点604的单个参考距离602。这样,效果单元510可以确定相对于锚点604的每个单个确定的参考距离602中的每个的更新距离608。

[0118] 虽然被描述为针对灰度深度图513执行,但是可以针对提供深度信息的其他类型的信息(诸如彩色图像、彩色或灰度立体图像、红外相机图像等)来执行该技术。换句话说,可以针对提供与对应的音频数据15相关联的场景的深度信息的任何类型的信息来执行该技术。

[0119] 图8是描绘根据本公开的技术的流程图。音频回放系统16可以接收比特流21,并且可以将其含有的音频数据存储在存储器中(202)。音频回放系统16还可将包含在比特流21中的元数据存储在存储器中(204)。另外,音频回放系统16可以将适配元数据存储在存储器中(206)。例如,适配元数据可以包含用户位置和用户方位中的一个或多个。在一些示例中,某些适配元数据,诸如用户位置和用户方位,包含在比特流21中。在其他示例中,用户位置和用户方位没有包含在比特流21中,而是通过用户输入来接收。

[0120] 然后,音频回放系统16可以基于适配元数据适配音频数据(208)。例如,音频回放系统16可以适配音频数据以提供比源设备12所创建的更少的自由度。在一些示例中,当适配音频数据时,音频回放系统16可以将诸如效果矩阵26之类的效果矩阵应用于音频数据。在一些示例中,音频回放系统16可以基于用户位置来确定效果矩阵。在一些示例中,音频回放系统16可以将效果矩阵26乘以渲染矩阵以获得更新的渲染矩阵。在一些示例中,音频回放系统16可以获得指示用户402的旋转头部运动的旋转指示,并且可以基于旋转指示和适配元数据来适配音频数据。

[0121] 音频回放系统16可以从适配的音频数据生成扬声器馈送(210)。扬声器馈送可以被配置为与耳机、扩音器或任何其他类型的扬声器一起使用。在一些示例中,音频回放系统16可以将更新的渲染矩阵应用于音频数据以生成扬声器馈送。在一些示例中,音频回放系统16可以将双耳渲染器应用于适配的更高阶立体混响声音频数据以生成扬声器馈送。

[0122] 在一些示例中,音频回放系统16可以将扬声器馈送输出到扬声器(212)。在一些示例中,音频回放系统可以在一个或多个扬声器(诸如耳机或一个或多个扩音器)上再现声场(214)。

[0123] 图9是显示图2A和图2B的系统可如何处理音频数据的图。在图9的示例中,源设备12对6DOF内容进行编码,并且将包含6DOF内容的比特流21传输到音频回放系统16。在图9的示例中,音频回放系统16支持6DOF内容,因此,渲染6DOF内容并基于6DOF内容生成扬声器馈送。用户可以利用所有6DOF运动来消费6DOF内容,这意味着用户可以在3D空间中自由运动。比特流21包含音频和元数据以解码所有可能的用户位置。

[0124] 图5、6、7和9已经描述了支持3DOF+和6DOF渲染的音频回放系统16的示例性实施方式。然而,图10-13描述了不支持3DOF+和6DOF渲染或者禁用了这种功能的音频回放系统16

的示例性实施方式。

[0125] 图10是示出图2A和图2B的系统可以根据本公开的技术如何处理音频数据的图。在图10的示例中,源设备12对6DOF内容进行编码,并且将包含6DOF内容的比特流21传输到音频回放系统16。在图10的示例中,音频回放系统16仅支持3DOF内容。因此,音频回放系统16将6DOF内容适配为3DOF内容,并基于3DOF内容生成扬声器馈送。用户可以利用3DOF运动(例如,改变俯仰、偏航和翻滚)来消费3DOF内容,这意味着用户不能平移运动但可以旋转运动。

[0126] 在图10的示例中,比特流21包含音频和元数据以解码所有可能的用户位置。比特流21还包含适配元数据,该适配元数据包含用户位置信息。在图10的示例中被假设为不支持6DOF内容的音频回放系统16使用用户位置信息来适配6DOF内容以提供3DOF并基于适配的音频数据生成扬声器馈送。当由支持6DOF的设备处理时,该设备可以忽略包含用户位置信息的适配元数据,这意味着支持6DOF的设备可以不需要此类信息。

[0127] 图11是示出图2A和图2B的系统可以根据本公开的技术如何处理音频数据的图。在图11的示例中,源设备12对6DOF内容进行编码,并且将包含6DOF内容的比特流21传输到音频回放系统16。在图11的示例中,音频回放系统16仅支持3DOF内容。因此,音频回放系统16将6DOF内容适配为3DOF内容,并基于适配的3DOF内容生成扬声器馈送。用户可以利用3DOF运动(例如,改变俯仰、偏航和翻滚)来消费3DOF内容,这意味着用户不能平移运动但可以旋转运动。

[0128] 在图11的示例中,比特流21包含音频和元数据以解码所有可能的用户位置。比特流21还包含适配元数据,该适配元数据包含用于多个位置的用户位置信息。在图11的示例中被假设为不支持6DOF内容的音频回放系统16使用来自多个位置之一的用户位置信息来适配6DOF内容以提供3DOF内容。音频回放系统16的用户可以选择多个位置之一,并且音频回放系统16可以基于所选位置的用户位置信息将6DOF内容适配为3DOF内容。当由支持6DOF的设备处理时,该设备可以忽略包含用于多个位置的用户位置信息的适配元数据,这意味着支持6DOF的设备可以不需要此类信息。

[0129] 图12是示出图2A和图2B的系统可以根据本公开的技术如何处理音频数据的图。在图12的示例中,源设备12对6DOF内容进行编码,并且将包含6DOF内容的比特流21传输到音频回放系统16。在图12的示例中,音频回放系统16仅支持3DOF内容。因此,音频回放系统16将6DOF内容适配为3DOF内容,并基于3DOF内容生成扬声器馈送。用户可以利用3DOF运动(例如,改变俯仰、偏航和翻滚)来消费3DOF内容,这意味着用户不能平移运动但可以旋转运动。

[0130] 在图12的示例中,音频回放系统16的用户可以从多个位置中选择位置,并且音频回放系统16可以将用户选择传输到源设备12。基于用户选择,源设备12可以生成比特流21,使得比特流21不包含用于解码所有可能的用户位置的音频和元数据。在图12的示例中,比特流21可以例如包含音频和元数据以解码由用户选择的用户位置。比特流21还包含适配元数据,该适配元数据包含用于选择的用户位置的用户位置信息。在图12的示例中,被假设为不支持6DOF内容的音频回放系统16使用用户位置信息来适配6DOF内容以提供3DOF并基于适配的音频数据生成扬声器馈送。

[0131] 图13是示出图2A和图2B的系统可以根据本公开的技术如何处理音频数据的图。在图13的示例中,源设备12对6DOF内容进行编码,并且将包含6DOF内容的比特流21传输到音频回放系统16。在图13的示例中,音频回放系统16仅支持0DOF内容。因此,音频回放系统16

将6DOF内容适配为0DOF内容,并基于0DOF内容生成扬声器馈送。用户可以在不使用任何6DOF运动的情况下消费0DOF内容,这意味着用户无法进行旋转运动或平移运动。

[0132] 在图13的示例中,比特流21包含音频和元数据以解码所有可能的用户位置。比特流21还包含适配元数据,该适配元数据包含用户位置信息和用户方位信息。在图13的示例中被假设为不支持6DOF内容的音频回放系统16使用用户位置信息来适配6DOF内容以提供0DOF并基于适配的音频数据生成扬声器馈送。当由支持6DOF的设备处理时,该设备可以忽略包含用户位置信息和用户方位信息的适配元数据,这意味着支持6DOF的设备可以不需要此类信息。当由支持3DOF的设备处理时,该设备可以基于用户位置将6DOF内容适配为3DOF,但忽略用户方位信息。

[0133] 图14是示出可根据本公开中描述的技术的各个方面进行操作的可穿戴设备800的示例图。在各种示例中,该可穿戴设备800可以表示VR头戴式耳机(诸如上述的VR头戴式耳机400)、AR头戴式耳机、MR头戴式耳机或扩展现实(XR)头戴式耳机。增强现实“AR”可以指的是叠加在用户实际所在的现实世界上的计算机渲染的图像或数据。混合现实“MR”可以指的是被世界锁定(world locked)到现实世界中特定位置的计算机渲染的图像或数据,也可以指的是VR上的一种变体,其中部分计算机渲染的3D元素和部分拍摄的现实元素被组合为模拟用户在环境中的实体存在的沉浸体验。扩展现实“XR”可以指的是VR、AR和MR的统称。有关XR术语的更多信息,请参阅Jason Peterson于2017年7月7日发布的文档,标题为“Virtual Reality,Augmented Reality,and Mixed Reality Definitions”。

[0134] 可穿戴设备800可以表示其他类型的设备,诸如手表(包含所谓的“智能手表”)、眼镜(包含所谓的“智能眼镜”)、耳机(包含所谓的“无线耳机”和“智能耳机”、智能服装、智能珠宝等。无论代表VR设备、手表、眼镜、和/或耳机,可穿戴设备800均可经由有线连接或无线连接与支持该可穿戴设备800的计算设备进行通信。

[0135] 在一些情况中,支持该可穿戴设备800的计算设备可以被集成在该可穿戴设备800内,这样,该可穿戴设备800可以被视为与支持该可穿戴设备800的计算设备相同的设备。在其他情况中,该可穿戴设备800可以与可以支持该可穿戴设备800的单独计算设备进行通信。在这方面,不应将术语“支持”理解为需要单独的专用设备,而是可以将被配置为执行本公开中描述的技术的各个方面的一个或多个处理器集成在该可穿戴设备800中或集成在与该可穿戴设备800分开的计算设备中。

[0136] 例如,当该可穿戴设备800表示VR设备400时,单独的专用计算设备(诸如包含一个或多个处理器的个人计算机)可以渲染音频和视频内容,而该可穿戴设备800可以确定平移头部运动,一旦确定了平移头部运动,专用计算设备可以基于平移头部运动,根据本公开中描述的技术的各个方面渲染音频内容(作为扬声器馈送)。作为另一个示例,当该可穿戴设备800表示智能眼镜时,该可穿戴设备800可以包含一个或多个处理器,该处理器既确定平移头部运动(通过与该可穿戴设备800的一个或多个传感器内的接口连接)又基于确定的平移头部运动渲染扬声器馈送。

[0137] 如图所示,该可穿戴设备800包含后置相机、一个或多个定向扬声器、一个或多个跟踪和/或记录相机以及一个或多个发光二极管(LED)灯。在一些示例中,LED灯可以被称为“超亮”LED灯。另外,该可穿戴设备800包含一个或多个眼动跟踪相机、高灵敏度音频麦克风和光学/投影硬件。该可穿戴设备800的光学/投影硬件可以包含耐用的半透明显示技术和

硬件。

[0138] 该可穿戴设备800还包含连接性硬件,该连接性硬件可以表示支持多模式连接性(诸如4G通信、5G通信等)的一个或多个网络接口。该可穿戴设备800还包含环境光传感器和骨传导换能器。在某些情况下,该可穿戴设备800还可以包含具有鱼眼镜头和/或远摄镜头的一个或多个无源和/或有源相机。根据本公开的各种技术,本公开的各种设备,诸如图2A的内容消费者设备14,可以使用可穿戴设备800的转向角来选择声场的音频表示(例如,MOA表示之一)以经由可穿戴设备800的定向扬声器-耳机404进行输出。将理解的是,该可穿戴设备800可以表现出多种不同的形状因子。

[0139] 此外,跟踪和记录相机以及其他传感器可以促进平移距离606的确定。尽管在图14的示例中未显示,但是,可穿戴设备800可以包含上面讨论的MEMS或用于检测平移距离606的其他类型的传感器。

[0140] 尽管针对诸如上文针对图3的示例所讨论的VR设备400和在图2A和2B的示例中阐述的其他设备的可穿戴设备的特定示例进行了描述,本领域普通技术人员将理解与图2A-3有关的描述可适用于可穿戴设备的其他示例。例如,诸如智能眼镜的其他可穿戴设备可以包含传感器,通过传感器可以获得平移头部运动。作为另一个示例,诸如智能手表的其他可穿戴设备可以包含传感器,通过传感器可以获得平移头部运动。这样,本公开中描述的技术不应限于特定类型的可穿戴设备,而是可以将任何可穿戴设备配置为执行本公开中描述的技术。

[0141] 图15A和15B是示出可以执行本公开中描述的技术的各个方面的示例系统的图。图15A示出了其中源设备12还包含相机200的示例。相机200可以被配置为捕获视频数据并将捕获的原始视频数据提供给内容捕获设备300。内容捕获设备300可以将视频数据提供给源设备12的另一组件,以进一步处理为视口划分的部分。

[0142] 在图15A的示例中,内容消费者设备14还包含可穿戴设备800。将理解的是,在各种实施方式中,可穿戴设备800可以被包含在内容消费者设备14中或从外部耦接到内容消费者设备14。如以上针对图14所讨论的,可穿戴设备800包含显示硬件和扬声器硬件,用于输出视频数据(例如,与各种视口相关联)并用于渲染音频数据。

[0143] 图15B示出了与图15A所示出的示例相似的示例,区别是图15A中显示的音频渲染器22被替换为能够使用一个或多个HRTF或能够向左右扬声器馈送103进行渲染的其他函数执行双耳渲染的双耳渲染器102。音频回放系统16可以将左右扬声器馈送103输出到耳机104。

[0144] 耳机104可以经由有线连接(诸如标准的3.5mm音频插孔、通用系统总线(USB)连接、光学音频插孔或其他形式的有线连接)耦接至音频回放系统16,或无线地(诸如通过Bluetooth™连接、无线网络连接等)耦接至音频回放系统16。耳机104可以基于左右扬声器馈送103来重新创建由音频数据表示的声场。耳机104可以包含左耳机扬声器和右耳机扬声器,它们由相应的左右扬声器馈送103供电(或者换句话说,被驱动)。

[0145] 应当认识到,根据示例,本文描述的任何技术的某些动作或事件可以以不同的顺序执行,可以被添加,合并或完全省略(例如,并非所有描述的动作或事件都是技术实践所必需的)。此外,在某些示例中,动作或事件可以例如通过多线程处理、中断处理或多个处理器并发地而不是顺序地执行。

[0146] 在一些示例中,VR设备(或流传输设备)可以使用耦接至VR/流传输设备的存储器的网络接口将交换消息传达到外部设备,其中,交换消息与声场的多个可用表示相关联。在一些示例中,VR设备可以使用耦接到网络接口的天线来接收无线信号,该无线信号包含与声场的多个可用表示相关联的数据分组、音频分组、视频协定或传输协议数据。在一些示例中,一个或多个麦克风阵列可以捕获该声场。

[0147] 在一些示例中,存储到存储器设备的声场的多个可用表示可以包含声场的多个基于对象的表示、声场的更高阶立体混响声表示、声场的混合阶立体混响声表示、声场的基于对象的表示和声场的更高阶立体混响声表示的组合、声场的基于对象的的表示和声场的混合阶立体混响声表示的组合、或者声场的混合阶表示和声场的更高阶立体混响声表示的组合。

[0148] 在一些示例中,声场的多个可用表示中的一个或多个声场表示可以包含至少一个高分辨率区域和至少一个较低分辨率区域,并且其中基于转向角的所选择展示针对至少一个高分辨率区域提供更大的空间精度以及针对较低分辨率区域提供更小的空间精度。

[0149] 在一个或多个示例中,可以以硬件、软件、固件或其任何组合来实施所描述的功能。如果以软件实施,则功能可以作为一个或多个指令或代码存储在计算机可读介质上或通过计算机可读介质传输,并由基于硬件的处理单元执行。计算机可读介质可以包括计算机可读存储介质,其对应于诸如数据存储介质的有形介质,或者通信介质,包括例如根据通信协议来促进将计算机程序从一个地方转移到另一个地方的任何介质。以这种方式,计算机可读介质通常可以对应于(1)非暂时性的有形计算机可读存储介质,或者(2)诸如信号或载波的通信介质。数据存储介质可以是可由一台或多台计算机或一个或多个处理器存取以取得指令、代码和/或数据结构以实现本公开中描述的技术的任何可用介质。计算机程序产品可以包括计算机可读介质。

[0150] 作为示例而非限制,这种计算机可读存储介质可以包括RAM、ROM、EEPROM、CD-ROM或其他光盘存储、磁盘存储或其他磁性存储装置,闪存或可以用来以指令或数据结构的形式存储所需的程序代码,并且可以由计算机存取的任何其他介质。而且,任何连接都适当地称为计算机可读介质。例如,如果使用同轴电缆、光纤电缆、双绞线、数字订户线(DSL)或无线技术(例如红外、无线电和微波)从网站、服务器或其他远程源传输指令,则介质的定义包括同轴电缆、光纤电缆、双绞线、DSL或诸如红外、无线电和微波之类的无线技术。然而,应当理解的是,计算机可读存储介质和数据存储介质不包括连接、载波、信号或其他瞬时介质,而是针对非瞬时的有形存储介质。本文使用的磁盘和光盘包括光碟(CD)、激光光盘、光学盘、数字多功能光盘(DVD)、软盘和蓝光光盘,其中磁盘通常以磁性方式再现数据,而光盘则通过激光光学方式再现数据。上述的组合也应包括在计算机可读介质的范围内。

[0151] 指令可以由一个或多个处理器执行,例如一个或多个数字信号处理器(DSP)、通用微处理器、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其他等效的集成或离散逻辑电路。因此,如本文所使用的术语“处理器”可以指任何前述结构或适合于实现本文描述的技术的任何其他结构。另外,在一些方面,本文描述的功能可以在被配置用于编码和解码的专用硬件和/或软件模块内提供,或结合在组合编解码器中。同样,该技术可以在一个或多个电路或逻辑元件中完全实现。

[0152] 本公开的技术可以在包括无线手机、集成电路(IC)或一组IC、(例如,芯片组)的多

种设备或装置中实现。在本公开中描述各种组件、模块或单元以强调配置为执行所公开技术的设备的功能方面,但不一定需要由不同硬件单元来实现。而是,如上所述,各种单元可以组合在编解码器硬件单元中,或者由包括如上所述的一个或多个处理器的互操作硬件单元的集合结合合适的软件和/或固件来提供。

[0153] 已经对各种示例进行了描述。这些示例以及其他示例都在下述权利要求的范围内。

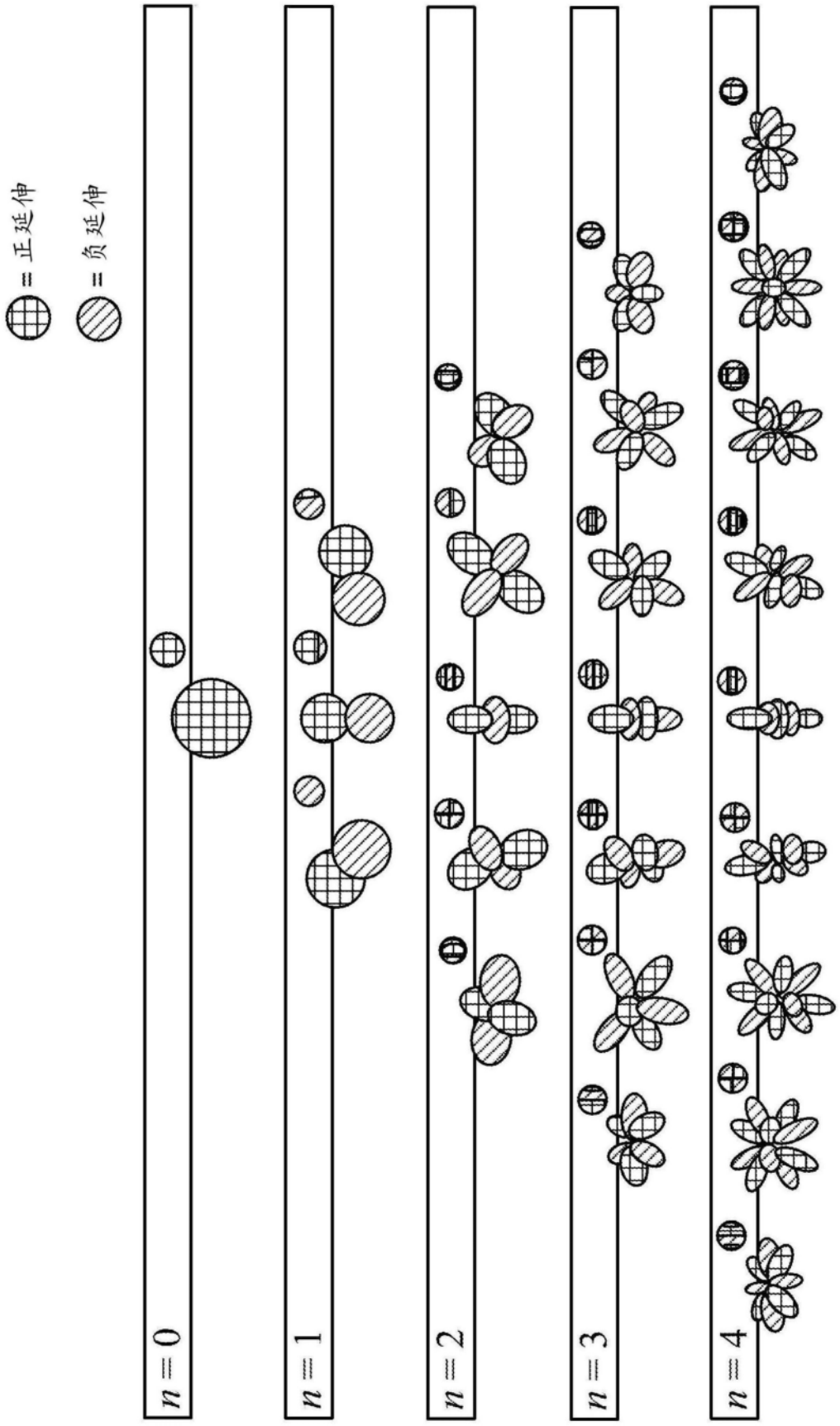


图1

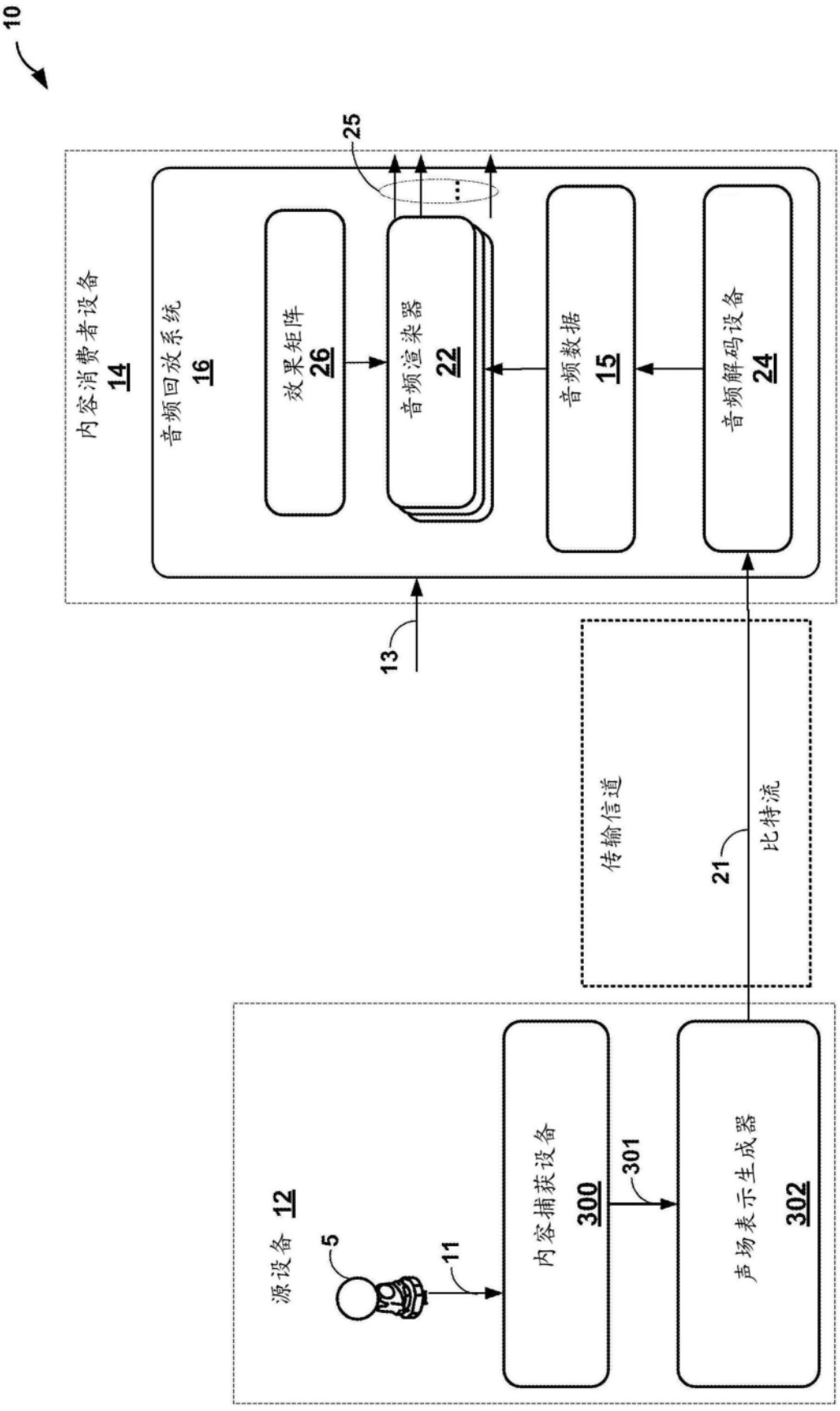

10
内容消费者设备
14
音频回放系统
16
效果矩阵
26
音频渲染器
22
音频数据
15
音频解码设备
24
25
13
传输信道
21
比特流
源设备
12
5
11
内容捕获设备
300
301
声场表示生成器
302

图2A

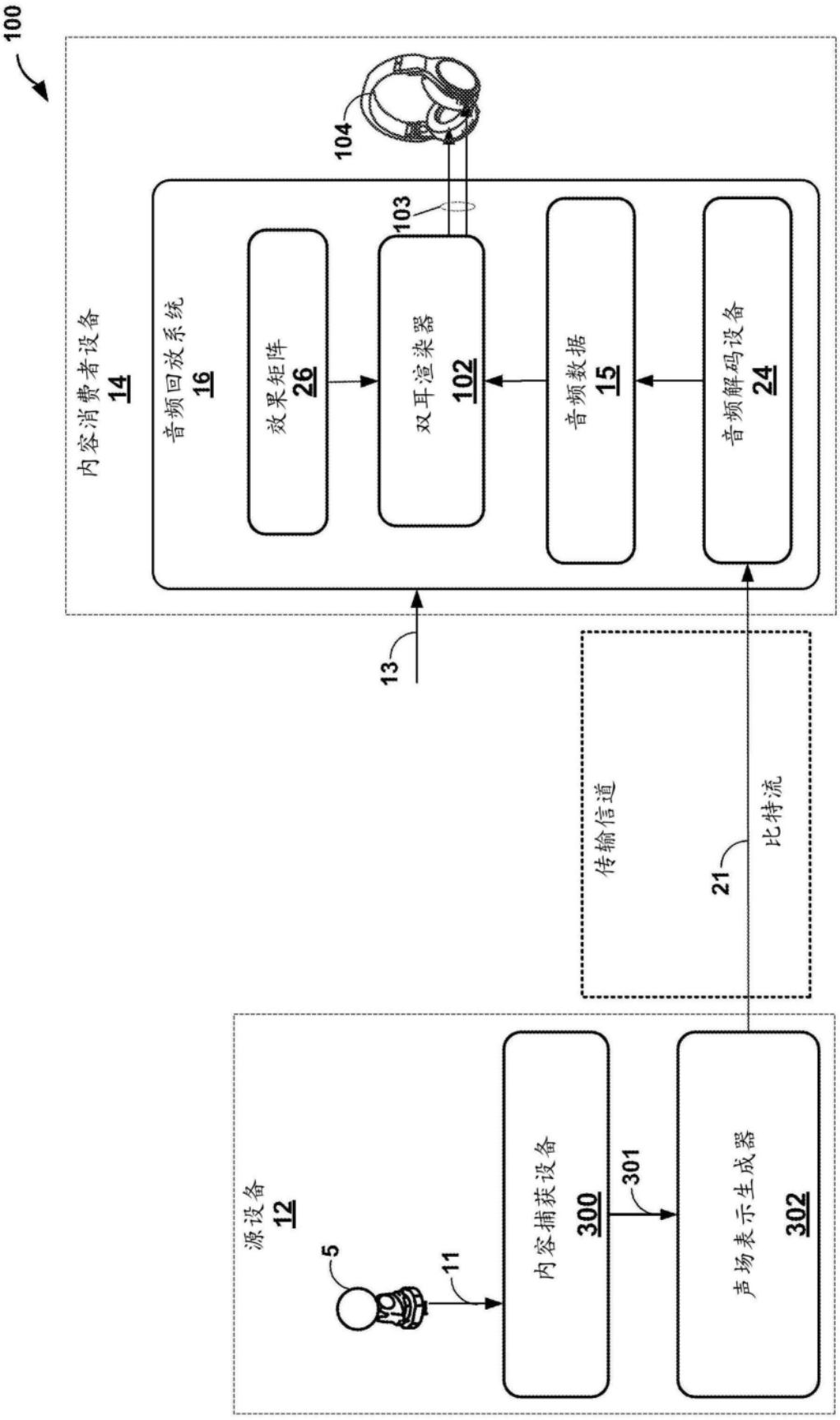

100
内容消费者设备
14
音频回放系统
16
效果矩阵
26
双耳渲染器
102
音频数据
15
音频解码设备
24
104
103
13
传输信道
21
比特流
源设备
12
5
11
内容捕获设备
300
301
声场表示生成器
302

图2B

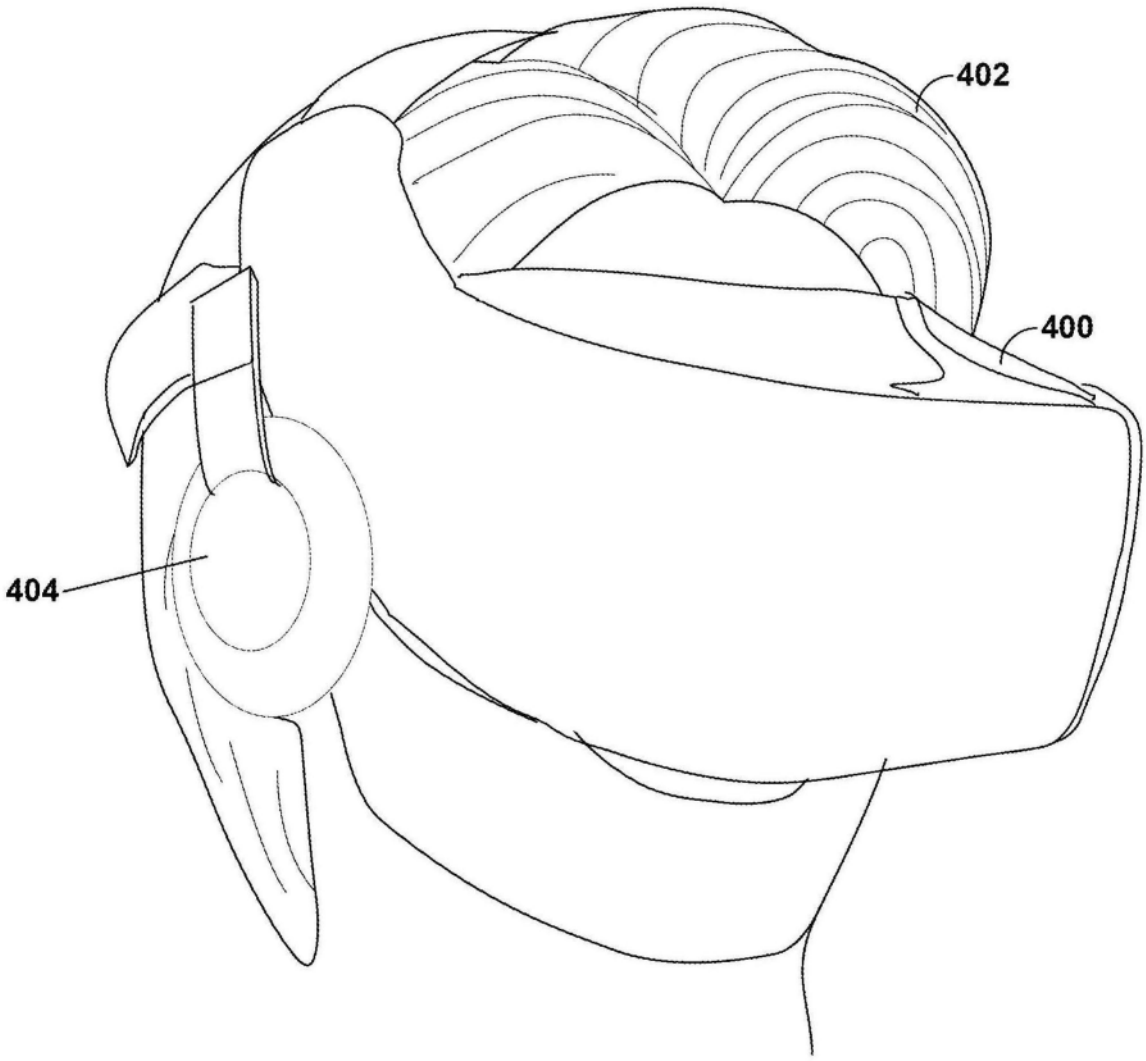


图3

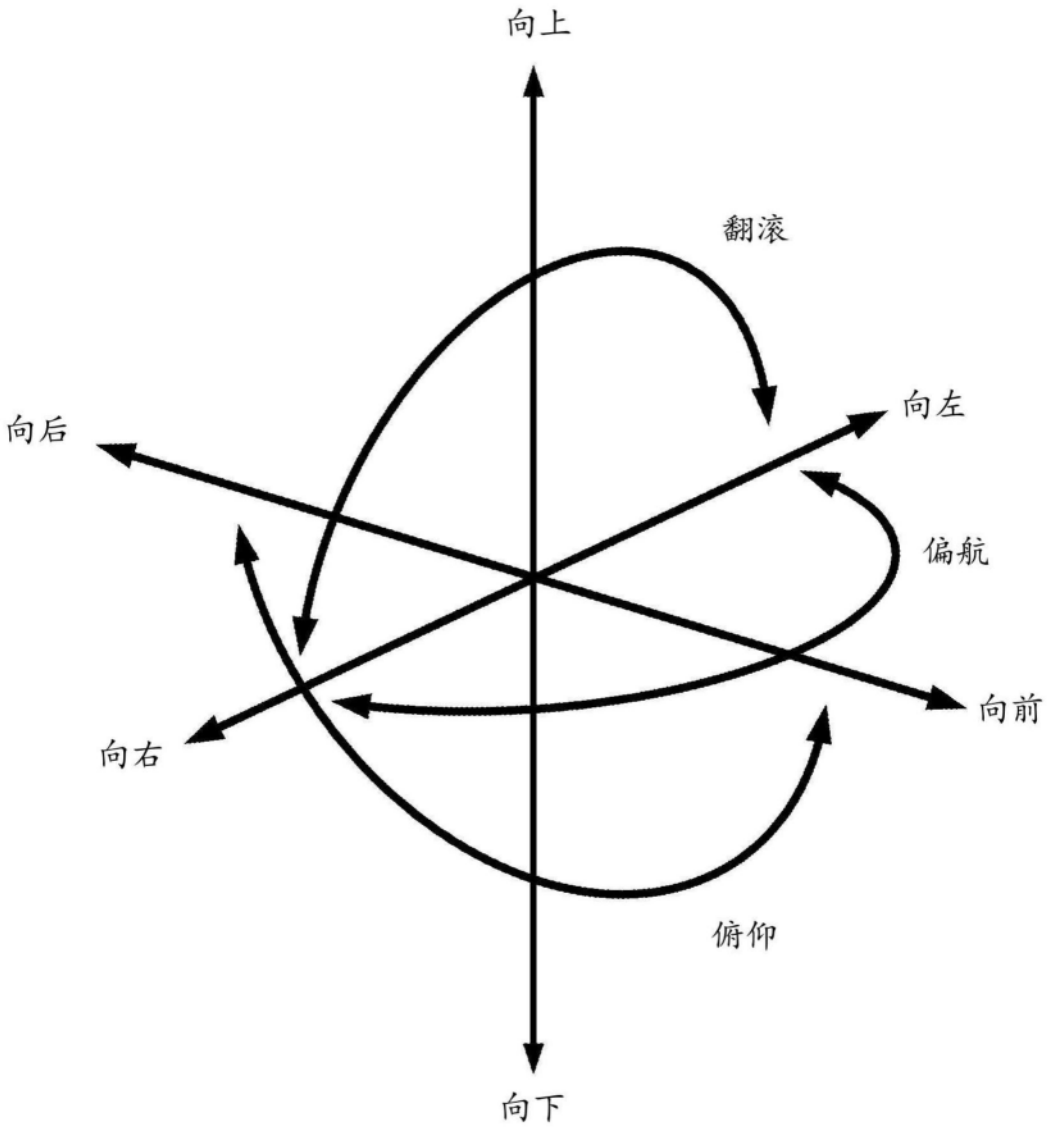


图4

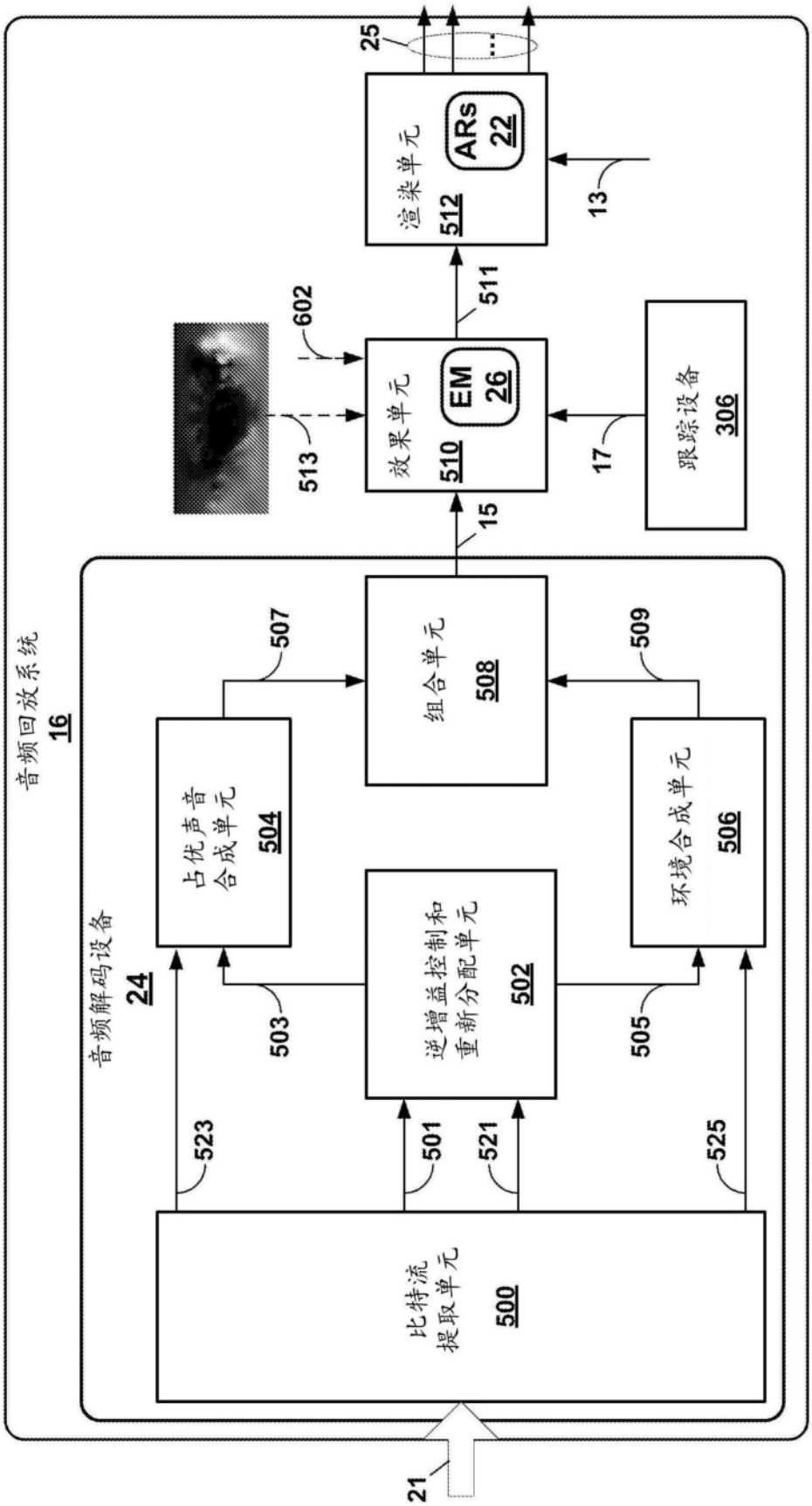

音频回放系统
16
音频解码设备
24
比特流提取单元
500
逆增益控制和重新分配单元
502
占优声音合成单元
504
环境合成单元
506
组合单元
508
效果单元
510
EM
26
渲染单元
512
ARs
22
跟踪设备
306
21
523
501
521
525
503
505
507
509
15
17
513
602
511
13
25

图5

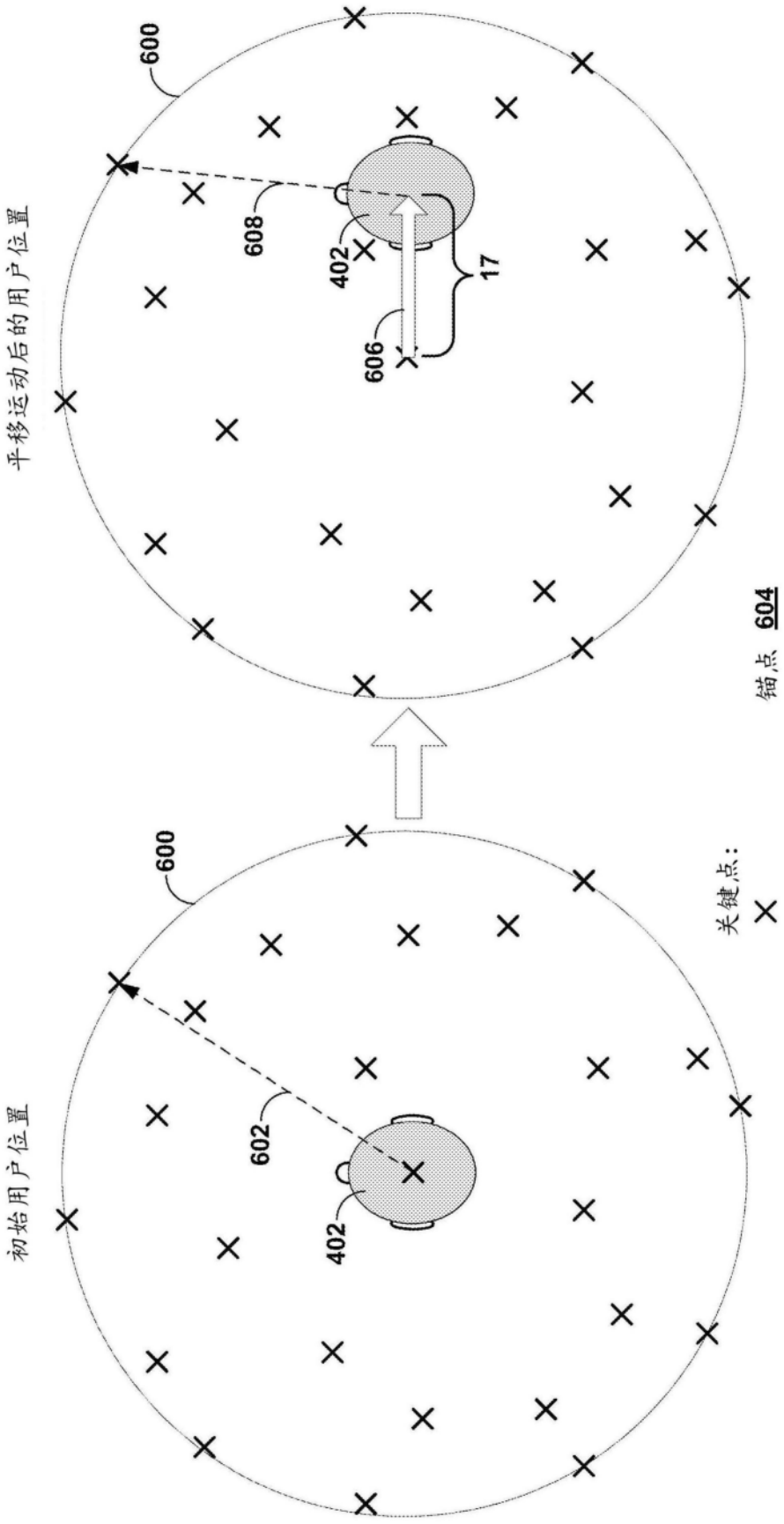


图6

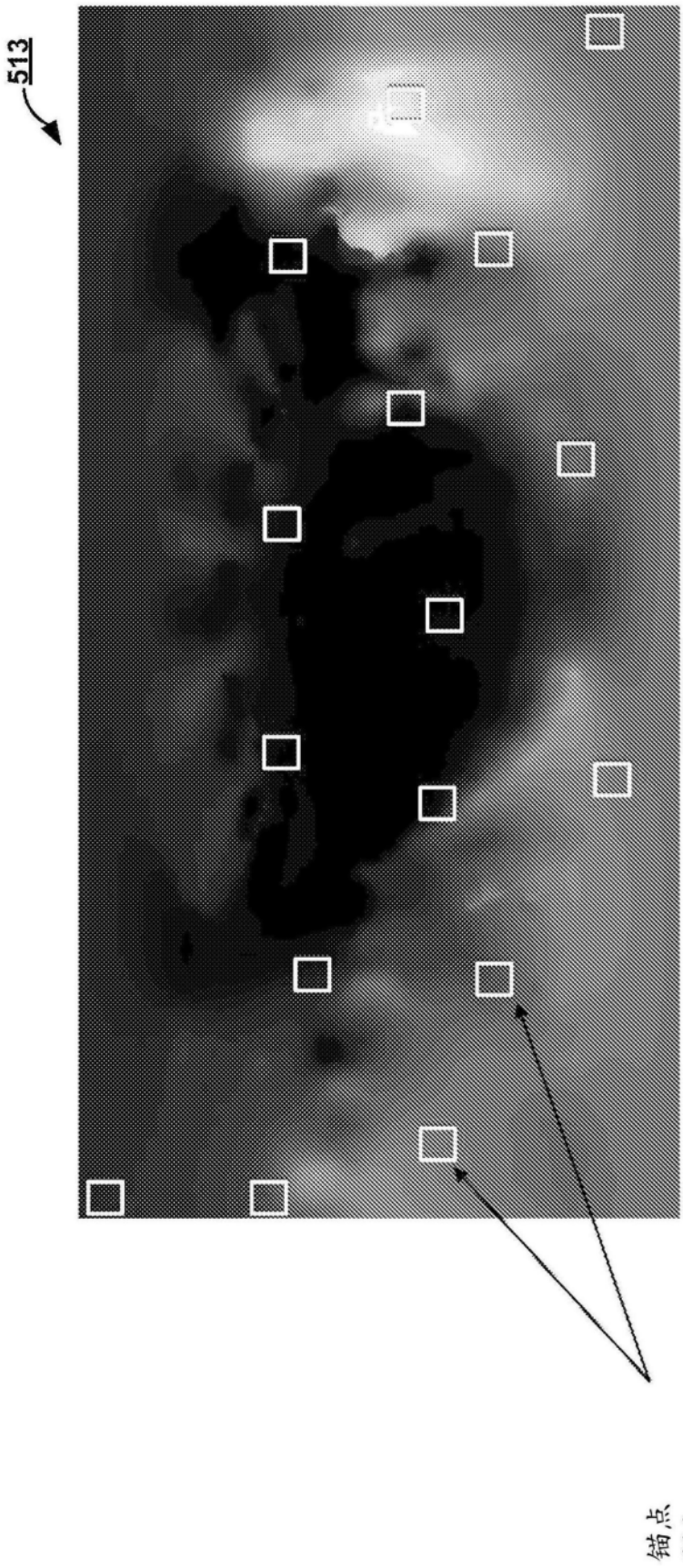


图7

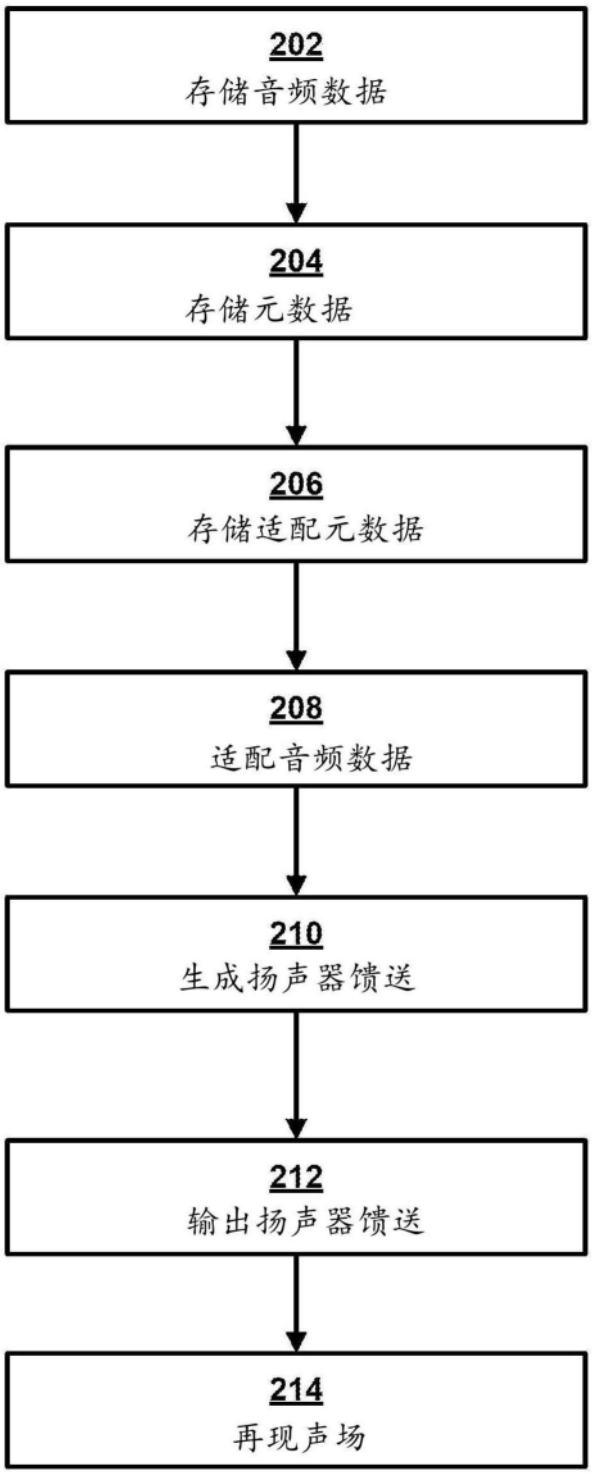

202
存储音频数据
204
存储元数据
206
存储适配元数据
208
适配音频数据
210
生成扬声器馈送
212
输出扬声器馈送
214
再现声场

图8

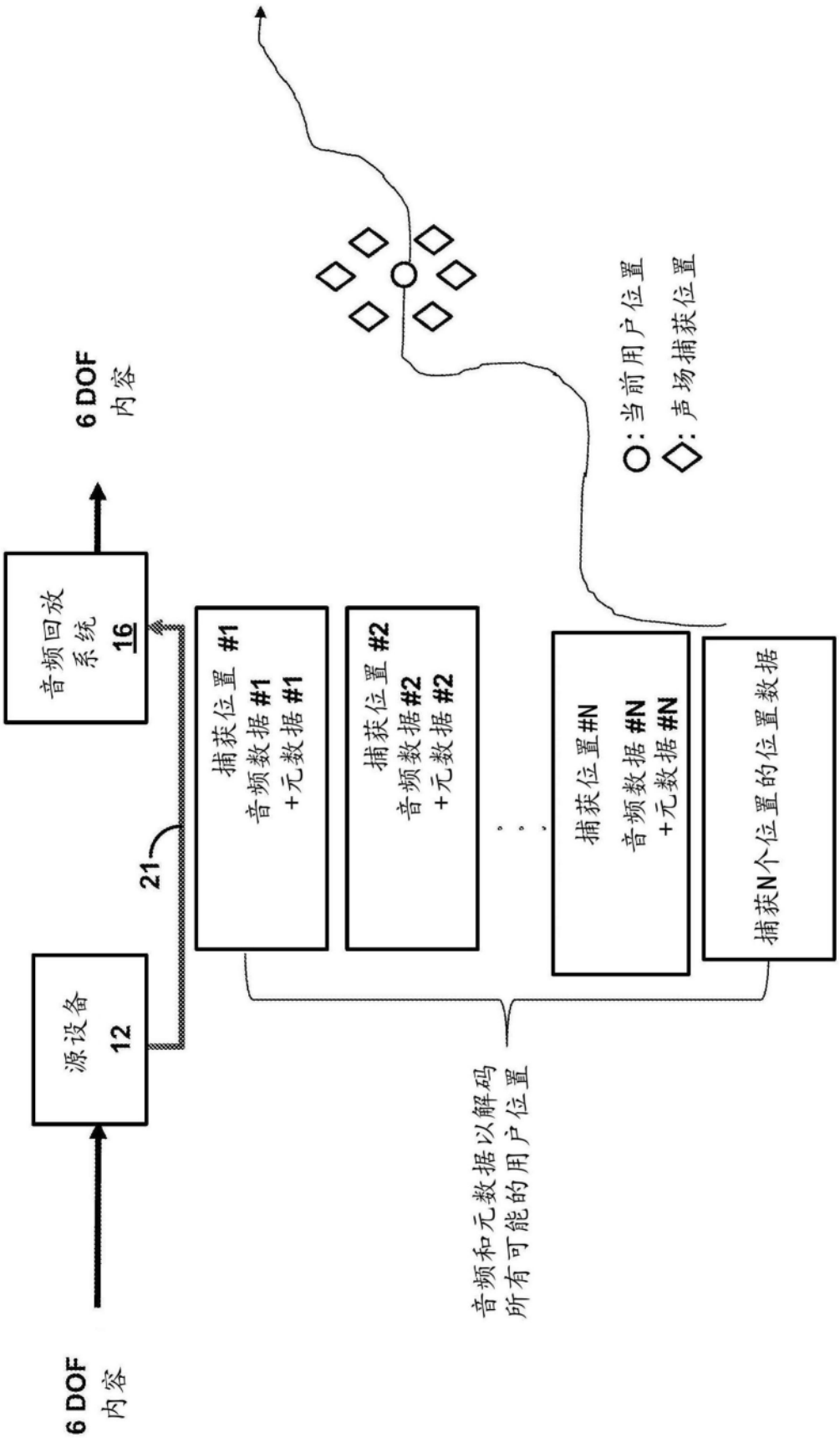

6 DOF
内容
源设备
12
21
音频回放
系统
16
6 DOF
内容
捕获位置 #1
音频数据 #1
+元数据 #1
捕获位置 #2
音频数据 #2
+元数据 #2
捕获位置 #N
音频数据 #N
+元数据 #N
捕获N个位置的位置数据
音频和元数据以解码
所有可能的用户位置
O: 当前用户位置
◇: 声场捕获位置

图9

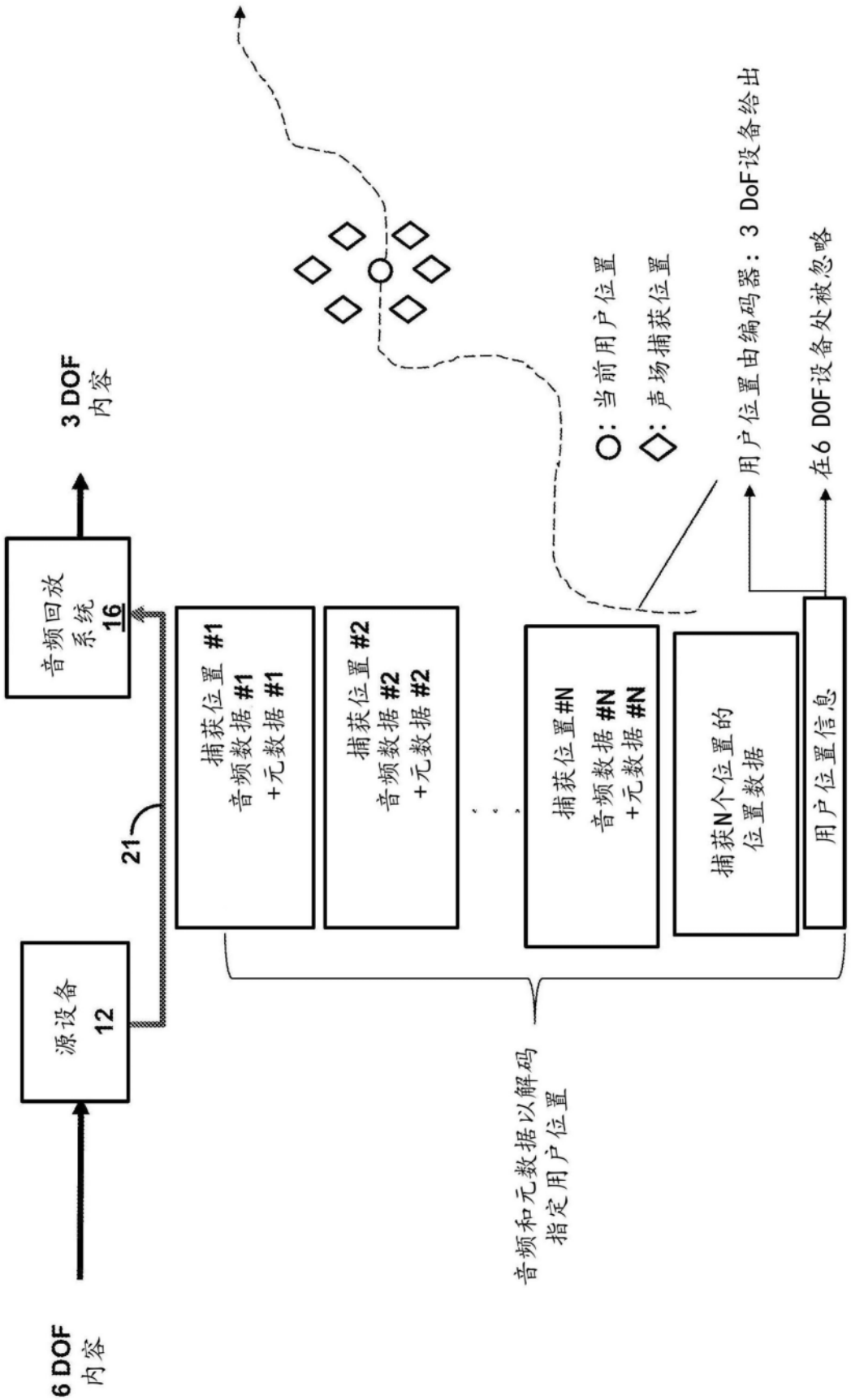

6 DOF
内容
源设备
12
音频和元数据以解码
指定用户位置
21
音频回放
系统
16
3 DOF
内容
捕获位置 #1
音频数据 #1
+元数据 #1
捕获位置 #2
音频数据 #2
+元数据 #2
捕获位置#N
音频数据 #N
+元数据 #N
捕获N个位置的
位置数据
用户位置信息
O: 当前用户位置
◇: 声场捕获位置
用户位置由编码器：3 DoF设备给出
在6 DOF设备处被忽略

图10

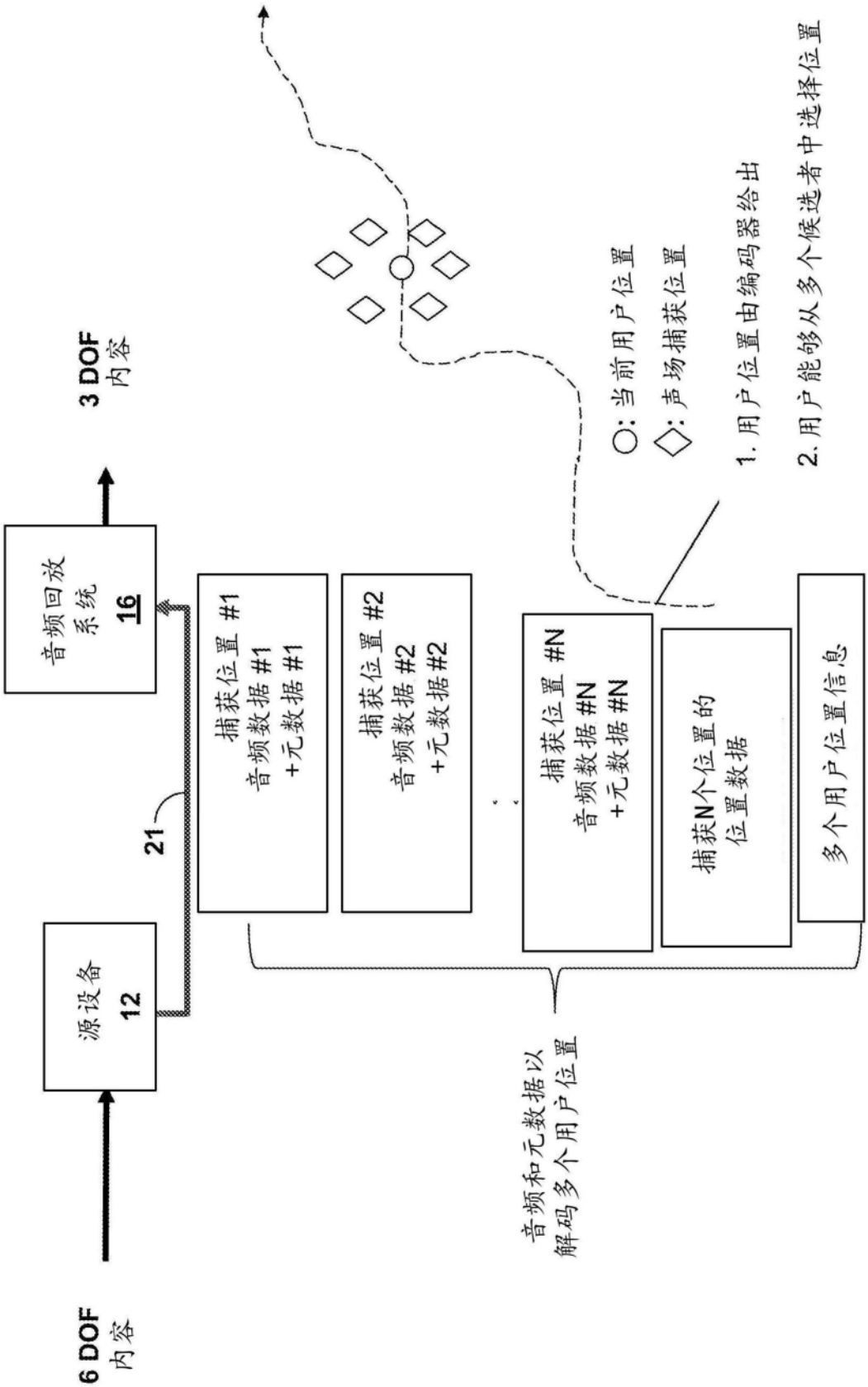

6 DOF
内容
源设备
12
音频和元数据以
解码多个用户位置
21
音频回放
系统
16
3 DOF
内容
捕获位置 #1
音频数据 #1
+元数据 #1
捕获位置 #2
音频数据 #2
+元数据 #2
捕获位置 #N
音频数据 #N
+元数据 #N
捕获N个位置的
位置数据
多个用户位置信息
○: 当前用户位置
◇: 声场捕获位置
1. 用户位置由编码器给出
2. 用户能够从多个候选者中选择位置

图11

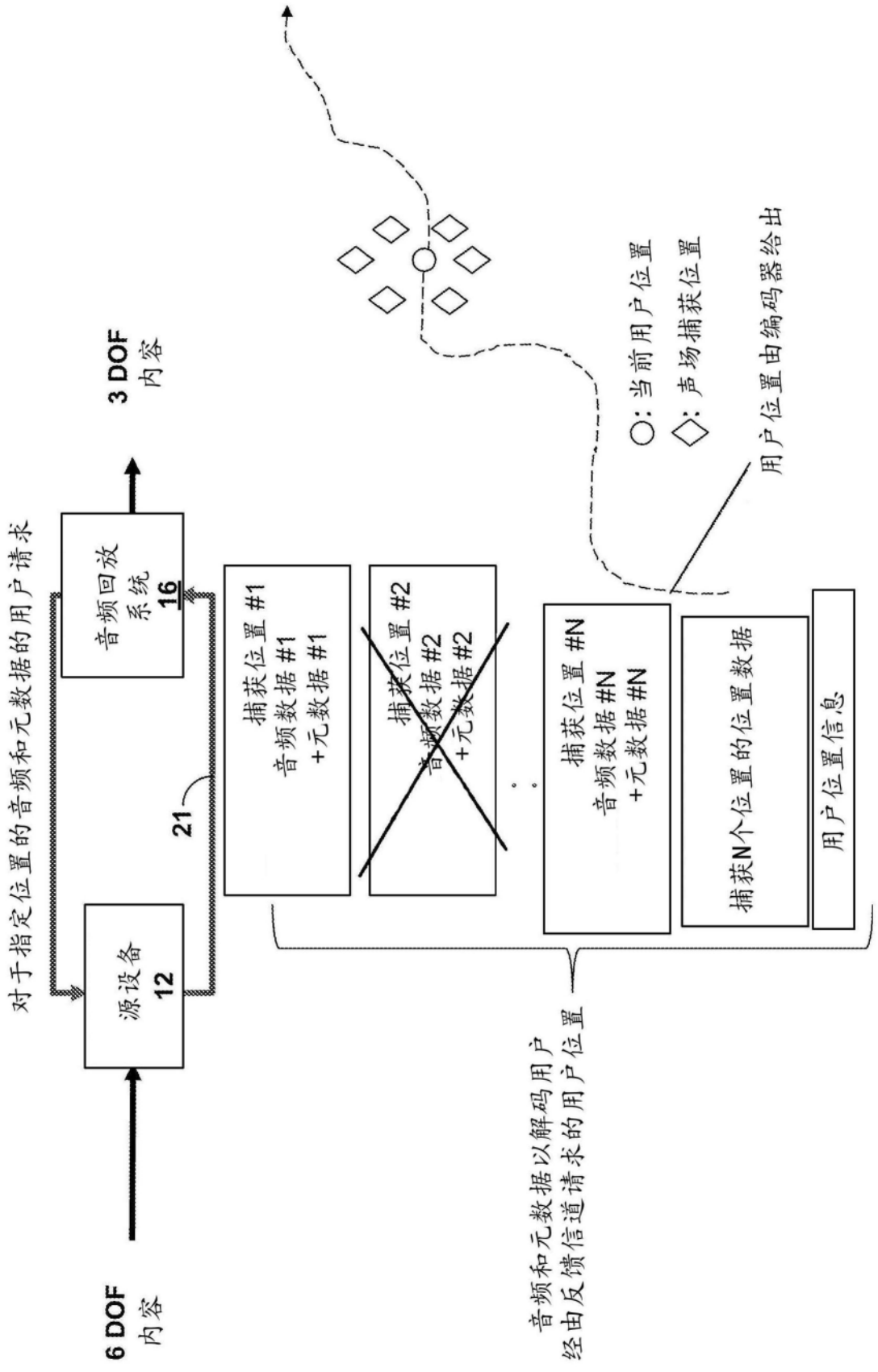

6 DOF
内容
源设备
12
音频回放
系统
16
3 DOF
内容
对于指定位置的音频和元数据的用户请求
21
音频和元数据以解码用户
经由反馈信道请求的用户位置
捕获位置 #1
音频数据 #1
+元数据 #1
捕获位置 #2
音频数据 #2
+元数据 #2
捕获位置 #N
音频数据 #N
+元数据 #N
捕获N个位置的位置数据
用户位置信息
○: 当前用户位置
◇: 声场捕获位置
用户位置由编码器给出

图12

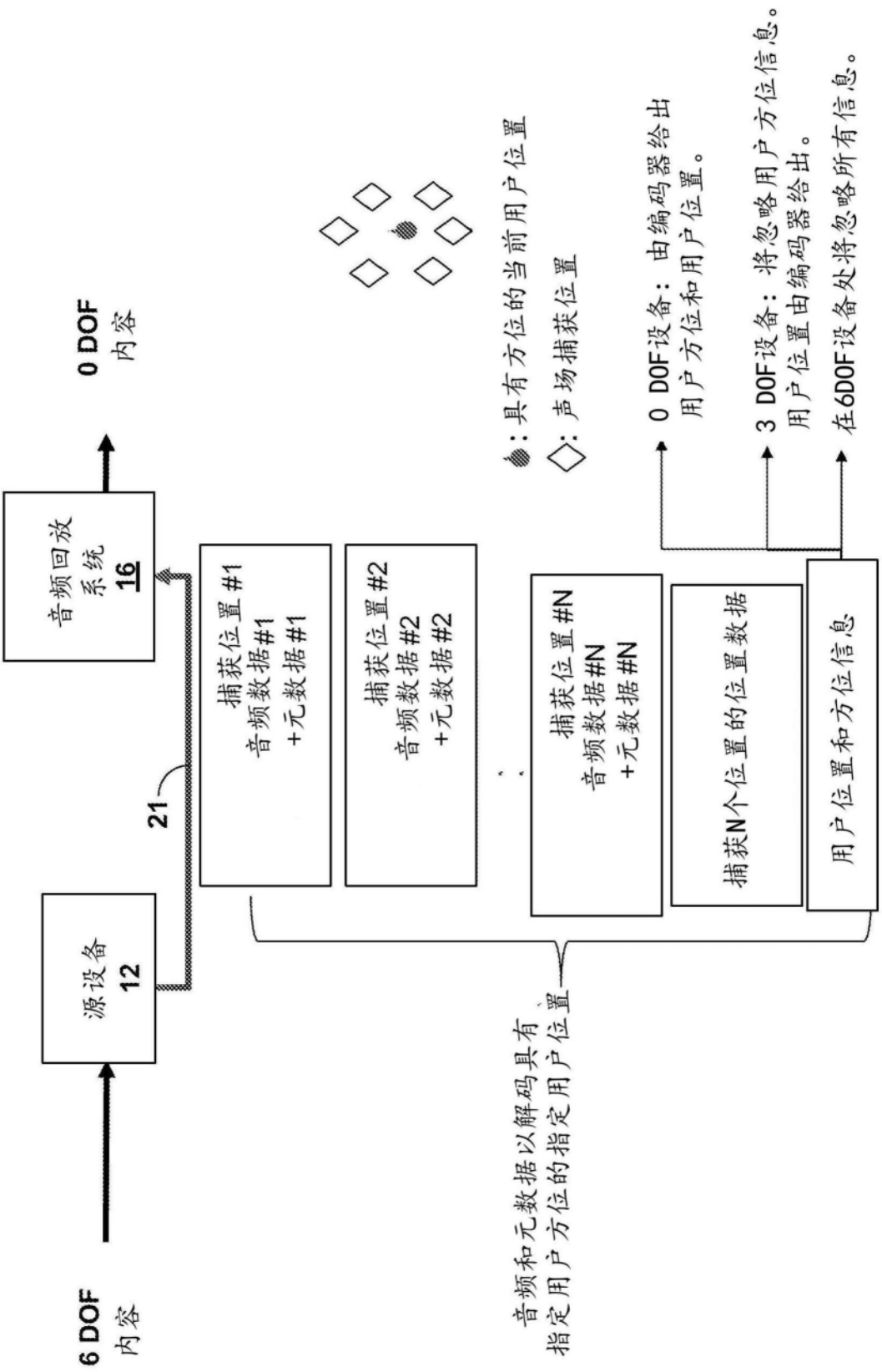

6 DOF
内容
源设备
12
音频和元数据以解码具有
指定用户方位的指定用户位置
21
音频回放
系统
16
0 DOF
内容
捕获位置#1
音频数据#1
+元数据#1
捕获位置#2
音频数据#2
+元数据#2
捕获位置#N
音频数据#N
+元数据#N
捕获N个位置的位置数据
用户位置和方位信息
●：具有方位的当前用户位置
◇：声场捕获位置
0 DOF设备：由编码器给出
用户方位和用户位置。
3 DOF设备：将忽略用户方位信息。
用户位置由编码器给出。
在6DOF设备处将忽略所有信息。

图13

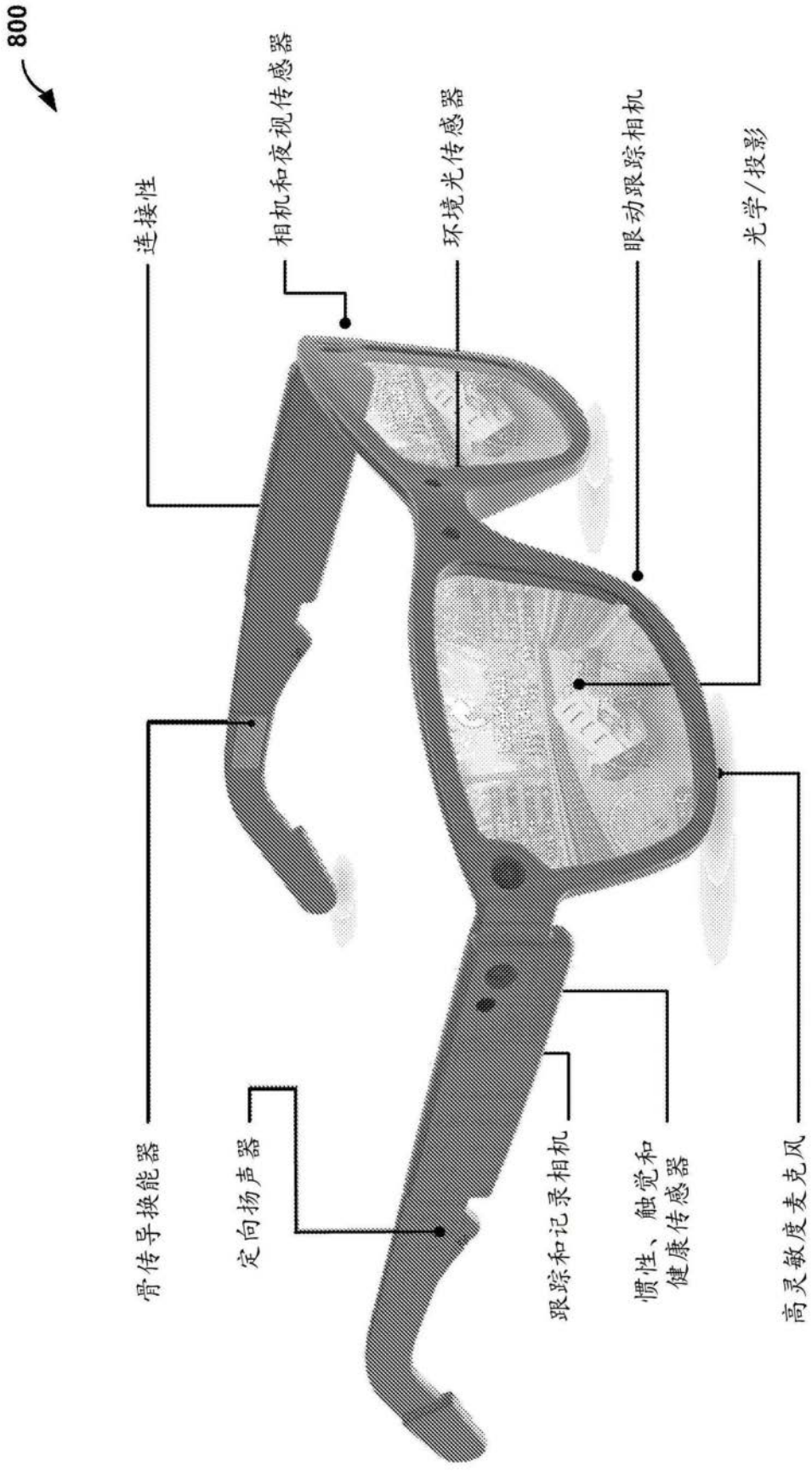

800
连接性
相机和夜视传感器
环境光传感器
眼动跟踪相机
光学/投影
骨传导换能器
定向扬声器
跟踪和记录相机
惯性、触觉和
健康传感器
高灵敏度麦克风

图14

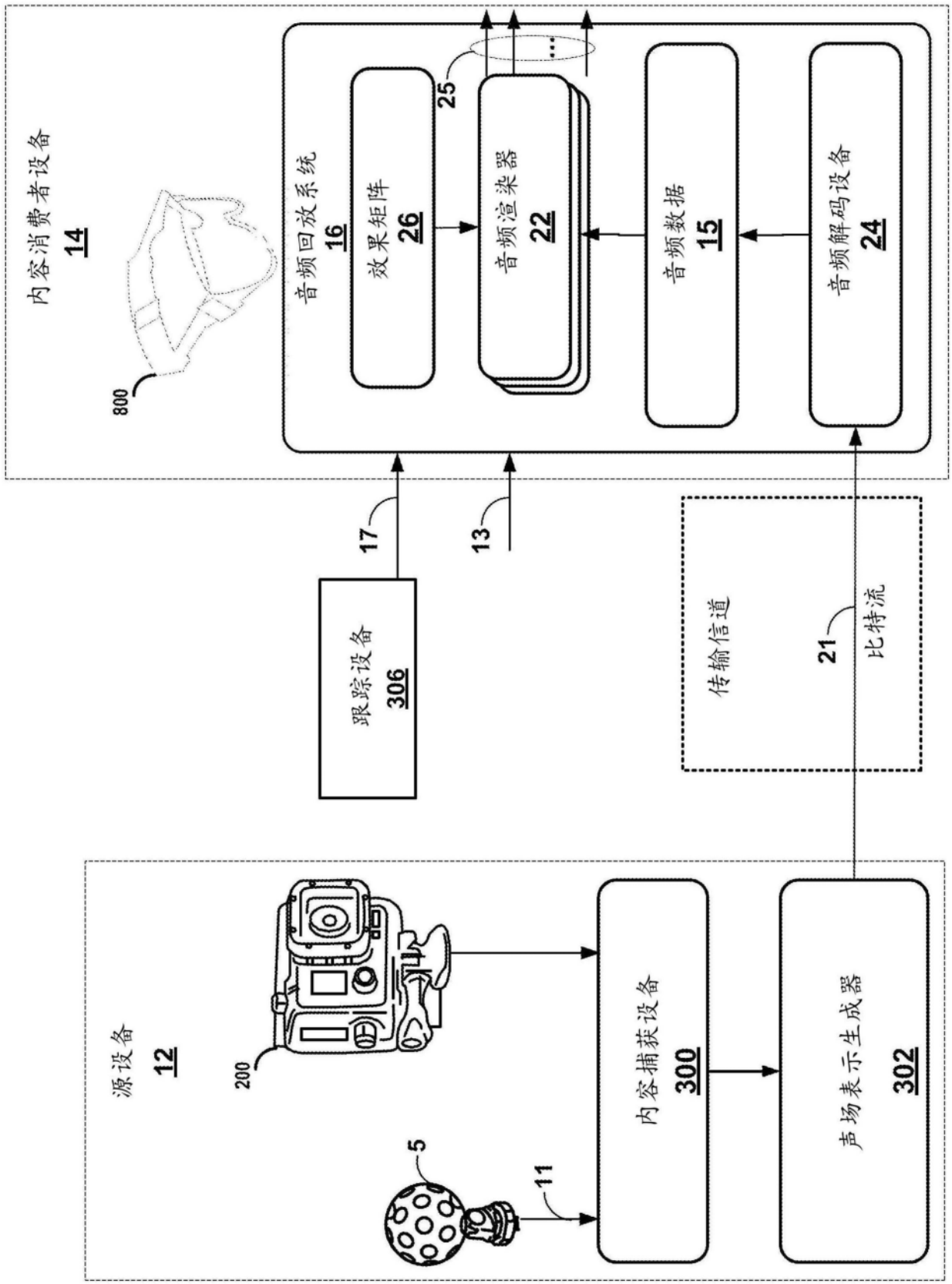

内容消费者设备
14
800
音频回放系统
16
效果矩阵
26
音频渲染器
22
25
音频数据
15
音频解码设备
24
17
13
跟踪设备
306
传输信道
21
比特流
源设备
12
200
5
11
内容捕获设备
300
声场表示生成器
302

图15A

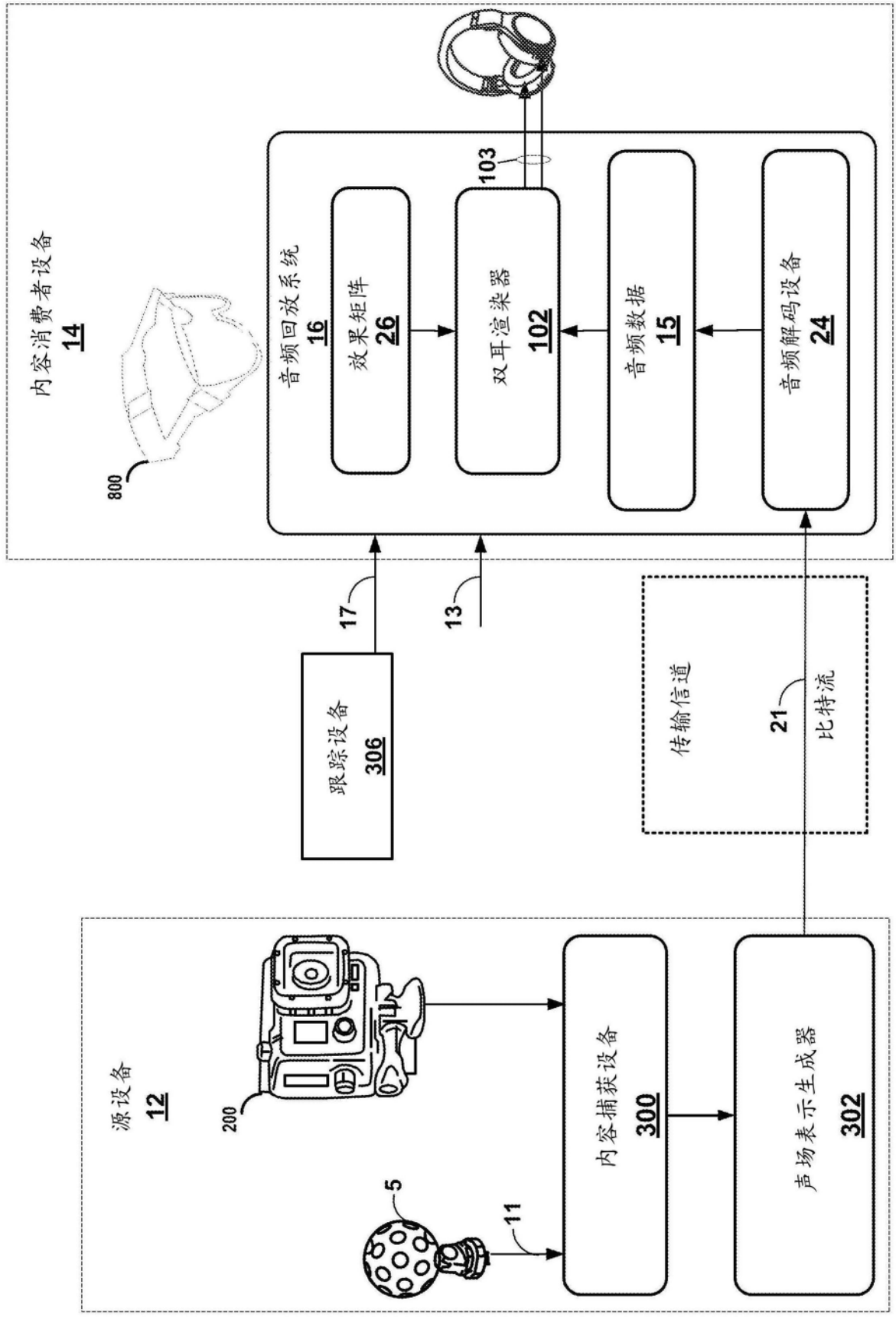

内容消费者设备
14
800
音频回放系统
16
效果矩阵
26
双耳渲染器
102
103
音频数据
15
音频解码设备
24
跟踪设备
306
17
13
传输信道
21
比特流
源设备
12
200
内容捕获设备
300
声场表示生成器
302
5
11

图15B