



(12) 发明专利

(10) 授权公告号 CN 108351700 B

(45) 授权公告日 2022.08.12

(21) 申请号 201680066714.X

(22) 申请日 2016.09.13

(65) 同一申请的已公布的文献号  
申请公布号 CN 108351700 A

(43) 申请公布日 2018.07.31

(30) 优先权数据  
62/219,604 2015.09.16 US  
62/289,085 2016.01.29 US

(85) PCT国际申请进入国家阶段日  
2018.05.15

(86) PCT国际申请的申请数据  
PCT/US2016/051521 2016.09.13

(87) PCT国际申请的公布数据  
W02017/048713 EN 2017.03.23

(73) 专利权人 奇跃公司  
地址 美国佛罗里达州

(72) 发明人 S·V·曼加特 M·B·塔克  
A·A·塔吉克

(74) 专利代理机构 北京市中咨律师事务所  
11247  
专利代理师 杨晓光 于静

(51) Int.Cl.  
G06F 3/01 (2006.01)  
G06F 3/16 (2006.01)  
H04S 7/00 (2006.01)

(56) 对比文件  
US 2013236040 A1, 2013.09.12  
US 2014115468 A1, 2014.04.24  
US 2006028400 A1, 2006.02.09  
CN 102566756 B, 2015.07.22  
CN 102681657 A, 2012.09.19  
US 2014328505 A1, 2014.11.06

审查员 宰涛

权利要求书2页 说明书31页 附图19页

(54) 发明名称

音频文件的头部姿势混合

(57) 摘要

描述了能够向显示设备的用户呈现音频文件的听觉或视觉表示的可穿戴设备的示例,其中音频文件包括表示音频文件的不同音频内容的多个干音轨。描述如下的系统和方法:确定用户的姿势;基于用户的姿势来生成音频文件的多个干音轨中的至少一个干音轨的音频混合;基于用户的姿势和音频混合来生成音频混合的可视化;将表示音频混合的音频信号传送给扬声器;并将表示音频混合的可视化的视觉信号传送给显示器。

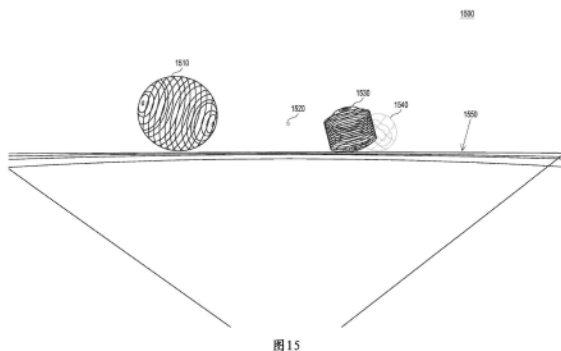


图15

1. 一种可穿戴设备,包括:

非瞬态存储器,其被配置为存储包括多个干音轨的音频文件,每个干音轨表示所述音频文件的不同音频内容;

传感器,其被配置为测量与所述可穿戴设备的用户的姿势相关联的信息;

显示器,其被配置为将图像呈现给所述可穿戴设备的用户的眼睛;

扬声器,其被配置为将声音呈现给所述可穿戴设备的用户;以及

处理器,其与所述非瞬态存储器、所述传感器、所述扬声器以及所述显示器通信,所述处理器编程有可执行指令从而:

在所述显示器上渲染包括与音频文件的不同音轨相关联的多个表面的对象;

确定相对于所述用户的所述姿势的所述对象的方位的改变,其中所述用户的所述姿势指示所述用户的注视方向或所述用户的位置和方位;

至少部分地基于所述对象的所述方位和所述用户的所述姿势,生成所述音频文件的所述多个干音轨中的至少一个干音轨的音频混合;

至少部分地基于所述用户的所述姿势和所述音频混合,生成所述音频混合的可视化;以及

向所述扬声器发送表示所述音频混合的音频信号。

2. 根据权利要求1所述的可穿戴设备,其中,所述多个干音轨中的干音轨与不同的乐器或人声音轨相关联。

3. 根据权利要求1所述的可穿戴设备,其中,所述多个干音轨中的至少一个干音轨包括与所述干音轨的空间位置和空间方向性相关联的信息。

4. 根据权利要求1所述的可穿戴设备,其中,所述传感器包括以下中的至少一个:陀螺仪、加速度计、惯性测量单元、眼睛跟踪照相机或被配置为对所述用户的环境成像的照相机。

5. 根据权利要求1所述的可穿戴设备,其中,所述姿势包括以下中的至少一个:头部姿势、眼睛姿势、身体姿势或手势。

6. 根据权利要求1所述的可穿戴设备,其中,所述可视化包括与所述多个干音轨的相应干音轨相关联的视觉图形。

7. 根据权利要求6所述的可穿戴设备,其中,所述视觉图形包括与所述干音轨的表现相关联的乐器或人的图形表示。

8. 根据权利要求6所述的可穿戴设备,其中,所述视觉图形中的至少一些被呈现在相对于所述用户的不同深度平面处。

9. 根据权利要求1所述的可穿戴设备,其中,所述可穿戴设备被配置为识别所述用户的环境中的物理对象,并且将所述多个干音轨中的干音轨的图形表示呈现为与所述物理对象相关联。

10. 根据权利要求1所述的可穿戴设备,其中,所述可穿戴设备被配置为接收来自用户输入设备的输入,并且所述处理器被配置为至少部分地基于所述输入来生成所述音频混合或所述可视化。

11. 一种用于与音频文件交互的方法,所述方法包括:

在可穿戴设备的控制下,所述可穿戴设备包括硬件计算机处理器,被配置为测量与所

述可穿戴设备的用户的姿势相关联的信息的传感器;被配置为将图像呈现给所述可穿戴设备的用户的眼睛的显示器;被配置为将声音呈现给所述可穿戴设备的用户的扬声器;

在所述显示器上渲染包括与音频文件的不同干音轨相关联的多个表面的对象;

检测相对于所述用户的所述姿势的所述对象的方位的改变,其中所述用户的所述姿势指示所述用户的注视方向或所述用户的位置和方位;

至少部分地基于所述对象的所述方位和所述用户的所述姿势,生成所述音频文件的所述多个干音轨中的至少一个干音轨的音频混合;

至少部分地基于所述用户的所述姿势或者所述音频混合中的一个,更新所述多个干音轨的可视化;以及

向所述扬声器发送表示所述音频混合的音频信号。

12. 根据权利要求11所述的方法,其中,所述多个干音轨的干音轨与不同的乐器或人声音轨相关联。

13. 根据权利要求11所述的方法,其中,所述多个干音轨中的至少一个干音轨包括与所述干音轨的空间位置和空间方向性相关联的信息。

14. 根据权利要求11所述的方法,其中,所述可视化包括与所述多个干音轨的相应干音轨对应的视觉图形。

15. 根据权利要求14所述的方法,其中,所述视觉图形包括与所述相应干音轨相关联的乐器或与所述相应干音轨相关联的演奏所述乐器的人。

16. 根据权利要求14所述的方法,其中,所述视觉图形中的至少一些被呈现在相对于所述用户的不同深度平面处。

17. 根据权利要求11所述的方法,其中,所述姿势包括以下中的至少一个:头部姿势、眼睛姿势、身体姿势或手势。

18. 根据权利要求11所述的方法,其中,更新所述可视化包括强调或去强调与多个干音轨中的干音轨相关联的视觉图形。

19. 根据权利要求11所述的方法,还包括:

接收所述多个干音轨中的干音轨的选择;

接收修改所述干音轨的指示,其中,所述指示包括所述用户的所述姿势的改变或者来自用户输入设备的输入;以及

至少部分地基于所述指示来修改所述干音轨。

20. 根据权利要求19所述的方法,其中,修改所述干音轨包括以下中的至少一个:调整音量,用另一干音轨代替所述干音轨,静音所述干音轨,强调所述干音轨超过所述多个干音轨中的其它干音轨。

## 音频文件的头部姿势混合

[0001] 相关申请的交叉引用

[0002] 本申请根据35USC§119(e)要求2015年9月16日提交的美国临时申请No.62/219,604题为“音频文件的头部姿势混合(HEAD POSE MIXING OF AUDIO FILES)”和2016年1月29日提交的美国临时申请No.62/289,085题为“音频文件的头部姿势混合(HEAD POSE MIXING OF AUDIO FILES)”的优先权权益。两个临时申请的全部内容通过引用全部并入在此。

### 技术领域

[0003] 本公开涉及虚拟现实和增强现实成像和可视化系统,尤其涉及基于用户姿势来混合音频文件的系统。

### 背景技术

[0004] 现代计算和显示技术促进了用于所谓的“虚拟现实”、“增强现实”或“混合现实”体验的系统的开发,其中数字再现的图像或其部分以其看起来似乎是或可能被认为是真实的方式呈现给用户。虚拟现实或“VR”场景通常涉及数字或虚拟图像信息的呈现,而不透明于其它实际的真实世界的视觉输入;增强现实或“AR”情景通常涉及将数字或虚拟图像信息呈现为对用户周围的现实世界的可视化的增强;混合现实或“MR”,其与真实世界和虚拟世界合并在一起,以产生物理对象和虚拟对象共存并实时交互的新的环境。事实证明,人类的视觉感知系统非常复杂,开发促进虚拟图像元素在其它虚拟或现实世界图像元素中的舒适的、感觉自然的、丰富的呈现的VR、AR或MR技术是有挑战性的。在此公开的系统和方法解决了与VR、AR和MR技术有关的各种挑战。

### 发明内容

[0005] 描述了能够向显示设备的用户呈现音频文件的可听或可视表示的可穿戴设备的示例。音频文件包括表示音频文件的不同音频内容的多个干音轨。

[0006] 可穿戴设备的一个实施例包括:非瞬态存储器,其被配置为存储包括多个干音轨的音频文件,其中每个干音轨表示音频文件的不同音频内容;传感器,其被配置为测量与可穿戴设备的用户的姿势相关联的信息;显示器,其被配置为将图像呈现给可穿戴设备的用户的眼睛;扬声器,其被配置为向可穿戴设备的用户呈现声音;以及与非瞬态存储器、传感器、扬声器以及显示器通信的处理器。处理器编程有可执行指令从而:确定用户的姿势;至少部分地基于用户的姿势生成音频文件的多个干音轨中的至少一个干音轨的音频混合;至少部分地基于用户的姿势和音频混合来生成音频混合的可视化;将表示音频混合的音频信号传送给扬声器;并将表示音频混合的可视化的视觉信号传送给显示器。

[0007] 另一方面,描述了用于与增强现实对象进行交互的方法。该方法在硬件计算机处理器的控制下执行。该方法包括生成供可穿戴系统的用户进行交互的增强现实对象;在用户与界面交互时检测用户的手势;将检测到的手势与对增强现实对象的特征的修改相关联;以及根据对增强现实对象的特征的修改来修改增强现实对象。可穿戴系统可以包括执

行用于与增强现实对象交互的方法的处理器。

[0008] 在附图和以下描述中阐述本说明书中描述的主题的一个或多个实施方式的细节。根据说明书、附图以及权利要求将清楚其它特征、方面和优势。该概述和以下详细的描述都不旨在限定或限制本发明主题的范围。

## 附图说明

[0009] 图1描绘了具有某些虚拟现实对象以及由人看到的某些物理对象的混合现实场景的示意图。

[0010] 图2示意性示出了可穿戴系统的示例。

[0011] 图3示意性示出了使用多个深度平面来模拟三维图像的方法的方面。

[0012] 图4示意性示出了用于向用户输出图像信息的波导堆叠的示例。

[0013] 图5示出了可以由波导输出的示例出射光束。

[0014] 图6是示出包括波导装置、将光耦合到波导装置或从波导装置光耦合光的光耦合器子系统、以及控制子系统的光学系统的示意图,该光学系统用于生成多焦点立体显示器、图像或光领域。

[0015] 图7是可穿戴系统的示例的框图。

[0016] 图8是渲染与识别的对象相关的虚拟内容的方法的示例的过程流程图。

[0017] 图9是可穿戴系统的另一示例的框图。

[0018] 图10是用于确定到可穿戴系统的用户输入的方法的示例的过程流程图。

[0019] 图11是用于与虚拟用户界面进行交互的方法的示例的过程流程图。

[0020] 图12-14示意性示出了向可穿戴系统的用户呈现音频文件的多个流轨道的可视化的用户界面的示例,其中音频文件至少部分地基于用户的姿势被动态地混合。

[0021] 图15示出了在用户环境中的不同深度处显示不同视觉图形的3D用户界面的示例。

[0022] 图16A和图16B示出了声源的方向性的示例。

[0023] 图17示出了创建声音拼贴效果的示例。

[0024] 图18示出了视觉和听觉上呈现音频文件的示例过程。

[0025] 在整个附图中,可以重新使用附图标记来指示参考元件之间的对应关系。提供附图以示出在此描述的示例实施例并且不旨在限制本公开的范围。

## 具体实施方式

### [0026] 概述

[0027] 音频文件可以包括表示音频信号的多个干音轨(stem track),用于例如声音、鼓、吉他、贝斯或其它声音。干音轨可以与多个乐器(诸如一组鼓或四重奏乐器)相关联,或与单个声源(诸如语音或一个乐器)相关联。单个干音轨可以表示单声道、立体声道或环绕声道。音频文件可以包括1、2、3、4、5、6、8、10、12个或更多的干音轨。除了干音轨之外,音频文件还可以包括用于标准回放的主音轨。

[0028] 用户可能想要与音频文件中的干音轨交互并通过混合干音轨来生成新的音频文件。然而,现有的用户界面对于这项任务通常是麻烦的,因为他们通常不向干音轨提供可视化,并且通常需要专业技能来组合多个干音轨。

[0029] 在此描述的可穿戴系统旨在通过提供与干音轨相关联的视觉图形来解决该问题。例如,与干音轨相关联的视觉图形可以是用于该干音轨的乐器的图形表示。如果干音轨与声音相关联,则视觉图形也可以是虚拟的人。

[0030] 可穿戴系统可以允许用户使用姿势(例如头部姿势、身体姿势、眼睛姿势或手势)容易地与干音轨进行交互。例如,用户可以通过移动他的手或改变他的头部位置在音频文件中混合多个干音轨,或者跨多个音频文件来混合干音轨。用户还可以修改音频文件,例如通过调整干音轨(例如调整干音轨的音量)或通过用其它干音轨替换干音轨。在一些实施例中,干音轨的某些混合可以与用户环境中的位置相关联。当用户移动到环境中的位置时,可穿戴系统可以播放与该位置关联的声音(或声音的混合)。与干音轨交互的另外的示例参考图12-18进一步描述。

[0031] 尽管在此参考音频文件描述了示例,但是可穿戴系统也可以被配置为允许与视频文件或者音频和视频文件的组合(诸如其中视频文件包括音频声道)的类似用户交互。

### [0032] 3D显示器

[0033] 可穿戴系统可以被配置成呈现三维(3D)用户界面,用于用户与虚拟内容(诸如音频文件中的干音轨的可视化)进行交互。例如,可穿戴系统可以是可穿戴设备的一部分,其可以单独或组合地呈现VR、AR或MR环境用于用户交互。

[0034] 图1描绘了具有某些虚拟现实对象以及由人看到的某些物理对象的混合现实场景的示意图。在图1中,描绘了MR场景100,其中MR技术的用户看到以背景中的人、树、建筑为特征的真实世界的公园状的设置110以及实体平台120。除了这些项目,MR技术的用户还感觉他“看到”站在真实世界的平台120上的机器人雕像130,以及看起来像飞行的蜜蜂的化身的卡通式的头像角色140,尽管这些元素不存在于真实世界中。

[0035] 为了使3D显示器产生真实的深度感觉,并且更具体地,模拟的表面深度感觉,期望显示器的视场中的每个点产生与其虚拟深度对应的适应响应。如果对显示点的适应响应不对应于该点的虚拟深度(由汇聚和立体视觉的双眼深度线索确定),则人眼可能经历适应冲突,导致成像不稳定、有害的眼部紧张、头痛,并且在没有适应信息的情况下,几乎完全缺乏表面深度。

[0036] VR、AR和MR体验可以通过具有显示器的显示系统来提供,其中与多个深度平面对应的图像被提供给观看者。对于每个深度平面,图像可以是不同的(例如,提供场景或对象的略微不同的呈现),并且可以由观看者的眼睛单独聚焦,从而有助于基于眼睛所需的适应向用户提供深度线索,为位于不同深度平面上的场景和/或基于观察不同深度平面上的不同图像特征失焦而聚焦不同图像特征。如在此其它地方所讨论的,这样的深度线索提供了可靠的深度感知。

[0037] 图2示出了可穿戴系统200的示例。可穿戴系统200包括显示器220,以及支持显示器220的功能的各种机械和电子模块和系统。显示器220可以耦合到可由用户、穿戴者或观看者210穿戴的框架230。显示器220可以定位在用户210的眼睛的前方。显示器220可以包括穿戴在用户头部上的头戴式显示器(HMD)。在一些实施例中,扬声器240耦合到框架230并且定位成邻近用户的耳道(在一些实施例中,未示出的另一个扬声器定位成邻近用户的另一个耳道以提供立体/可塑造声音控制)。如参照图12-16进一步描述的,可穿戴系统200可以通过扬声器240向用户播放音频文件,并使用显示器220呈现声音文件中的各种干音轨的3D

可视化。

[0038] 可穿戴系统200还可以包括观察用户周围环境中世界的面向外的成像系统464(在图4中示出)。可穿戴系统100还可以包括可以跟踪用户的眼睛运动的面向内的成像系统462(在图4中示出)。面向内的成像系统可以跟踪一只眼睛的运动或两只眼睛的运动。面向内的成像系统可以附接到框架230并且可以与处理模块260和/或270电通信,处理模块260和/或270可以处理由面向内的成像系统获取的图像信息以确定例如瞳孔直径和/或眼睛的方位或用户210的眼睛姿势。

[0039] 作为示例,可穿戴系统200可以使用面向外的成像系统464和/或面向内的成像系统462来获取用户的姿势的图像。图像可以是静止图像、视频的帧或视频、以组合的形式等。该姿势可用于混合音频文件的干音轨或用于确定应将哪些音频内容呈现给用户。

[0040] 显示器220诸如通过有线导线或无线连接可操作地耦合250到本地数据处理模块260,该本地数据处理模块260可以以各种配置安装,诸如固定地附接到框架230,固定地附接到由用户穿戴的头盔或帽子,嵌入到耳机中,或以其它方式可移除地附接到用户210(例如,以背包方式的配置,以带式耦合方式的配置)。

[0041] 本地处理和数据模块260可以包括硬件处理器以及诸如非易失性存储器(例如,闪存存储器)的数字存储器,二者都可用于辅助数据的处理、缓冲以及存储。数据可以包括如下数据:a)从传感器(其可以例如可操作地耦合到框架230或以其它方式附接到用户210)捕获的数据,例如图像捕获设备(例如,面向内的成像系统和/或面向外的成像系统中的照相机)、麦克风、惯性测量单元(IMU)、加速度计、罗盘、全球定位系统(GPS)单元、无线电设备和/或陀螺仪;和/或b)使用远程处理模块270和/或远程数据储存库280获取和/或处理的数据,可能在这样的处理或检索之后传递给显示器220。本地处理和数据模块260可以通过通信链路262和/或264(诸如经由有线或无线通信链路)可操作地耦合到远程处理模块270和远程数据储存库280,使得这些远程模块作为资源可用于本地处理和数据模块260。另外,远程处理模块280和远程数据储存库280可以相互可操作地耦合。

[0042] 在一些实施例中,远程处理模块270可以包括被配置为分析和处理数据和/或图像信息的一个或多个处理器。在一些实施例中,远程数据储存库280可以包括数字数据存储设施,其可以通过互联网或其它网络配置以“云”资源配置而可用。在一些实施例中,在本地处理和数据模块中存储全部数据,并且执行全部计算,允许从远程模块完全自主使用。

[0043] 例如,远程数据储存库280可以被配置为存储音频文件的内容,诸如与干音轨相关联的信息。本地处理和数据模块260和/或远程处理模块270可以检测用户的姿势,诸如用户的注视方向。处理模块260和270可以与远程数据储存库280进行通信以获得干音轨并且在用户的注视方向上生成干音轨的可视化。处理模块260和270可以进一步与显示器220通信并向用户呈现可视化。

[0044] 人类视觉系统复杂,并且提供深度的现实感知是具挑战性的。不受理论的限制,相信对象的观看者可能由于聚散度和适应性的组合而将该对象感知为“三维”。两只眼睛彼此相对的聚散运动(即,光瞳孔彼此相向或远离的滚动运动,以会聚眼睛的视线来注视对象)与眼睛晶状体的聚焦(或“适应性”)密切相关。在正常情况下,改变眼睛晶状体的焦点或适应眼睛,以将焦点从在不同距离处的一个对象改变到另一个对象,将会在称为“适应性聚散度反射(accommodation-vergence reflex)”的关系下自动地导致在聚散度上的匹配改变

达到相同的距离。同样,在正常情况下,聚散度的改变将引发适应性的匹配改变。提供适应性和聚散度之间的更好匹配的显示系统可以形成更逼真且舒适的三维图像模拟。

[0045] 图3示出了使用多个深度平面来模拟三维图像的方法的方面。参考图3,在z轴上距眼睛302和眼睛304的不同距离处的对象由眼睛302和眼睛304适应,以使得那些对象在焦点中。眼睛302和眼睛304呈现特定的适应状态,以使沿着z轴的不同距离处的对象进入焦点。因此,可以说特定的适应状态与深度平面306中的特定一个深度平面相关联,该特定深度平面具有相关联的焦距,以使得当眼睛处于该深度平面的适应状态时,特定深度平面中的对象或对象的部分被聚焦。在一些实施例中,可以通过为眼睛302和304中的每一只眼睛提供图像的不同呈现来模拟三维图像,并且还通过提供与深度平面中每一个深度平面对应的图像的不同呈现来模拟三维图像。尽管为了清楚说明而示出为分离的,但应理解的是,例如,随着沿着z轴的距离增加,眼睛302和眼睛304的视场可能重叠。另外,虽然为了便于说明而示出为平坦的,但应理解的是,深度平面的轮廓可以在物理空间中是弯曲的,使得深度平面中的所有特征在特定的适应状态下与眼睛对焦。不受理论的限制,可以相信的是,人类眼睛通常可以解释有限数量的深度平面以提供深度感知。因此,通过向眼睛提供与这些有限数量的深度平面中的每一个深度平面对应的图像的不同呈现,可以实现感知深度的高度可信的模拟。

#### [0046] 波导堆叠组件

[0047] 图4示出了用于向用户输出图像信息的波导堆叠的示例。可穿戴系统400包括可以用于采用多个波导432b、434b、436b、438b、440b向眼睛/大脑提供三维感知的波导堆叠或堆叠波导组件480。在一些实施例中,可穿戴系统400对应于图2的可穿戴系统200,图4更详细地示意性地示出了该可穿戴系统200的一些部分。例如,在一些实施例中,波导组件480可以被集成到图2的显示器220中。

[0048] 继续参考图4,波导组件480还可以包括在波导之间的多个特征458、456、454、452。在一些实施例中,特征458、456、454、452可以是透镜。在其它实施例中,特征458、456、454、452可以不是透镜。而是它们可以简单地是间隔物(例如,用于形成空气间隙的包层和/或结构)。

[0049] 波导432b、434b、436b、438b、440b和/或多个透镜458、456、454、452可以被配置为以各种级别的波前曲率或光线发散向眼睛发送图像信息。每个波导级别可以与特定的深度平面相关联,并且可以被配置为输出与该深度平面对应的图像信息。图像注入装置420、422、424、426、428可用于将图像信息注入到波导440b、438b、436b、434b、432b中,其中的每一个波导可以被配置为分配入射光穿过每一个相应的波导,用于向眼睛410输出。光从图像注入装置420、422、424、426、428的输出表面出射并被注入到波导440b、438b、436b、434b、432b的相应输入边缘。在一些实施例中,可以将单个光束(例如,准直光束)注入到每一个波导中,以便与特定波导相关联的深度平面对应的特定角度(和发散量)输出朝向眼睛410定向的克隆准直光束的整个视场。

[0050] 在一些实施例中,图像注入装置420、422、424、426、428是分立显示器,每个显示器产生用于分别注入到相应波导440b、438b、436b、434b、432b中的图像信息。在一些其它实施例中,图像注入装置420、422、424、426、428是单个复用显示器的输出端,其可以例如经由一个或多个光导管(诸如,光纤线缆)向图像注入装置420、422、424、426、428中的每一个图像

注入装置输送图像信息。

[0051] 控制器460控制堆叠波导组件480和图像注入装置420、422、424、426、428的操作。控制器460包括调节图像信息到波导440b、438b、436b、434b、432b的定时和提供的编程(例如,在非暂时性计算机可读介质中的指令)。在一些实施例中,控制器460可以是单个整体装置,或通过有线或无线通信通道连接的分布式系统。在一些实施例中,控制器460可以是处理模块260和/或270(图2所示)的部分。

[0052] 波导440b、438b、436b、434b、432b可以被配置为通过全内反射(TIR)在每一个相应的波导内传播光。波导440b、438b、436b、434b、432b可以各自是平面的或具有其它形状(例如,弯曲),具有主要的顶表面和底表面以及在这些主要的顶表面和底表面之间延伸的边缘。在所示的配置中,波导440b、438b、436b、434b、432b可以各自包括光提取光学元件440a、438a、436a、434a、432a,这些光提取光学元件被配置为通过将每一个相应波导内传播的光重定向而将光提取到波导外,以向眼睛410输出图像信息。提取的光也可以被称为外耦合的光,并且光提取光学元件也可以被称为外耦合光学元件。提取的光束在波导中传播的光照射光重定向元件的位置处被波导输出。光提取光学元件(440a,438a,436a,434a,432a)可以例如是反射和/或衍射光学特征。虽然为了便于描述和清晰绘图起见而将其图示设置在波导440b、438b、436b、434b、432b的底部主表面处,但是在一些实施例中,光提取光学元件440a、438a、436a、434a、432a可以设置在顶部和/或底部主表面处,和/或可以直接设置在波导440b、438b、436b、434b、432b的体积中。在一些实施例中,光提取光学元件440a、438a、436a、434a、432a可以形成在附接到透明基板材料层中以形成波导440b、438b、436b、434b、432b。在一些其它实施例中,波导440b、438b、436b、434b、432b可以是单片材料,并且光提取光学元件440a、438a、436a、434a、432a可以形成在那片材料的表面上和/或那片材料的内部中。

[0053] 继续参考图4,如在此所讨论的,每一个波导440b、438b、436b、434b、432b被配置为输出光以形成与特定深度平面对应的图像。例如,最接近眼睛的波导432b可以被配置为将如注入到这种波导432b中的准直光传送到眼睛410。准直光可以代表光学无限远焦平面。下一个上行波导434b可以被配置为将穿过第一透镜452(例如,负透镜)的准直光在其可以到达眼睛410之前发出。第一透镜452可以被配置为产生轻微凸面的波前曲率,使得眼睛/大脑将来自下一个上行波导434b的光解释为来自第一焦平面,该第一焦平面从光学无限远处更靠近向内朝向眼睛410。类似地,第三上波导436b将输出光在到达眼睛410之前穿过第一透镜452和第二透镜454。第一透镜452和第二透镜454的组合光功率可被配置为产生另一增量的波前曲率,以使得眼睛/大脑将来自第三波导436b的光解释为来自第二焦平面,该第二焦平面从光学无穷远比来自下一个上行波导434b的光更靠近向内朝向人。

[0054] 其它波导层(例如,波导438b、440b)和透镜(例如,透镜456、458)被类似地配置,其中堆叠中的最高波导440b通过它与眼睛之间的全部透镜发送其输出,用于代表最靠近人的焦平面的聚合(aggregate)焦度。当在堆叠波导组件480的另一侧上观看/解释来自世界470的光时,为了补偿透镜458、456、454、452的堆叠,补偿透镜层430可以设置在堆叠的顶部处以补偿下面的透镜堆叠458、456、454、452的聚合焦度。这种配置提供了与可用波导/透镜配对一样多的感知焦平面。波导的光提取光学元件和透镜的聚焦方面可以是静态的(例如,不是动态的或电激活的)。在一些替代实施例中,两者之一或者两者都可以是使用电激活特征

而动态的。

[0055] 继续参考图4,光提取光学元件440a、438a、436a、434a、432a可以被配置为将光重定向到它们相应的波导之外并且针对与波导相关联的特定深度平面输出具有适当的发散量或准直量的该光。结果,具有不同相关联深度平面的波导可具有不同配置的光提取光学元件,其取决于相关联的深度平面输出具有不同发散量的光。在一些实施例中,如在此所讨论的,光提取光学元件440a、438a、436a、434a、432a可以是体积或表面特征,其可以被配置为以特定角度输出光。例如,光提取光学元件440a、438a、436a、434a、432a可以是体积全息图、表面全息图和/或衍射光栅。在2015年6月25日公开的美国专利公开No. 2015/0178939中描述了诸如衍射光栅的光提取光学元件,其通过引用全部并入在此。

[0056] 在一些实施例中,光提取光学元件440a、438a、436a、434a、432a是形成衍射图案或“衍射光学元件”(在此也称为“DOE”)的衍射特征。优选地,DOE具有相对较低的衍射效率,以使得仅光束的一部分通过DOE的每一个交点偏转向眼睛410,而其余部分经由全内反射继续移动通过波导。携带图像信息的光因此被分成多个相关的出射光束,该出射光束在多个位置处离开波导,并且该结果对于在波导内反弹的该特定准直光束是朝向眼睛304的相当均匀图案的出射发射。

[0057] 在一些实施例中,一个或多个DOE可以在它们主动地衍射的“开”状态和它们不显著衍射的“关”状态之间可切换。例如,可切换的DOE可以包括聚合物分散液晶层,其中微滴在主体介质中包含衍射图案,并且微滴的折射率可以切换为基本上匹配主体材料的折射率(在这种情况下,图案不明显地衍射入射光),或者微滴可以切换为与主体介质的指数不匹配的指数(在这种情况下,该图案主动地衍射入射光)。

[0058] 在一些实施例中,深度平面的数量和分布和/或景深可以基于观看者的眼睛的瞳孔大小和/或方位而动态地改变。景深可与观看者的瞳孔大小成反比。因此,随着观看者眼睛瞳孔的大小减小,景深增加,使得由于平面的位置超出了眼睛的聚焦深度而不可辨别的该平面可能变得可辨别,并且随着瞳孔大小的减小和景深的相应增加表现为更聚焦。类似地,用于向观看者呈现不同图像的间隔开的深度平面的数量可随着瞳孔大小减小而减小。例如,观看者在不调整眼睛远离一个深度平面和到另一个深度平面的适应性的情况下,可能不能清楚地感知处于一个瞳孔大小的第一深度平面和第二深度平面的细节。然而,这两个深度平面可以在不改变适应性的情况下,对于处于另一瞳孔大小的用户同时充分地聚焦。

[0059] 在一些实施例中,显示系统可以改变接收图像信息的波导的数量,基于瞳孔大小和/或方位的确定或者基于接收特定瞳孔大小和/或方位的电信号指示。例如,如果用户的眼睛不能区分与两个波导相关联的两个深度平面,则控制器460可以被配置或编程为停止向这些波导中的一个提供图像信息。有利地,这可以减轻系统的处理负担,从而增加系统的响应性。在其中波导的DOE可在开启和关闭状态之间切换的实施例中,当波导确实接收图像信息时,DOE可切换到关闭状态。

[0060] 在一些实施例中,可能期望的是出射光束符合直径小于观看者眼睛的直径的条件。然而,考虑到观看者的瞳孔大小的可变性,满足这种条件可能是具有挑战性的。在一些实施例中,通过响应于观看者的瞳孔大小的确定而改变出射光束的大小,该条件在宽范围的瞳孔大小上满足。例如,随着瞳孔大小减小,出射光束的大小也可以减小。在一些实施例

中,可以使用可变光圈来改变出射光束大小。

[0061] 可穿戴系统400可以包括对世界470的一部分成像的面向外的成像系统464(例如,数字照相机)。世界470的这部分可以被称为视场(FOV),并且成像系统464有时被称为FOV照相机。可供观看者观看或成像的整个区域可被称为能视场(field of regard(FOR))。FOR可以包括围绕可穿戴系统400的立体角的 $4\pi$ 球面度。在可穿戴系统400的一些实施方式中,因为用户可以移动他们的头部和眼睛以观看用户周围的对象(位于用户的前面、后面、上面、下面或者旁边),所以FOR可以包括显示系统400的用户周围的基本上全部立体角。从面向外的成像系统464获得的图像可以用于跟踪用户做出的姿势(例如,手或手指姿势),检测用户前方的世界470中的对象等等。

[0062] 可穿戴系统400还可以包括面向内的成像系统466(例如,数字照相机),其观察用户的运动,诸如眼睛运动和面部运动。面向内的成像系统466可以用于捕获眼睛410的图像以确定眼睛304的瞳孔的大小和/或方位。面向内的成像系统466可以用于获得图像,这些图像用于确定用户正在观看的方向(例如,眼睛姿势)或用于用户的生物识别(例如,经由虹膜识别)。在一些实施例中,每只眼睛可以利用至少一个相机,以独立地分别确定每只眼睛的瞳孔大小和/或眼睛姿势,从而允许向每只眼睛呈现图像信息动态地适合该眼睛。在一些其它实施例中,只确定并假定单个眼睛410的瞳孔直径和/或方位(例如,每对眼睛仅使用单个照相机)与用户的两只眼睛类似。可以分析由面向内的成像系统466获得的图像以确定用户的眼睛姿势和/或情绪,其可以由可穿戴系统400用来决定应该向用户呈现哪些音频或视觉内容。可穿戴系统400还可以使用诸如IMU、加速度计、陀螺仪等的传感器来确定头部姿势(例如,头部位置或头部方位)。头部姿势可以单独使用或与眼睛姿势组合使用以与干音轨进行交互和/或呈现音频内容。

[0063] 可穿戴系统400可以包括用户输入设备466,用户可以通过该用户输入设备466向控制器460输入命令以与可穿戴系统400交互。例如,用户输入设备466可以包括触控板、触摸屏、操纵杆、多自由度(DOF)控制器、电容感测设备、游戏控制器、键盘、鼠标、方向板(D-pad)、棒、触觉设备、图腾(例如,用作虚拟用户输入设备)等等。在一些情况下,用户可以使用手指(例如,拇指)在触敏输入设备上按压或滑动以向可穿戴系统400提供输入(例如,向由可穿戴系统400提供的用户界面提供用户输入)。用户输入设备466可在使用可穿戴系统400期间由用户的手握持。用户输入设备466可以与可穿戴系统400进行有线或无线通信。

[0064] 图5示出了由波导输出的出射光束的示例。示出了一个波导,但是应该理解的是,波导组件480中的其它波导可以类似地起作用,其中波导组件480包括多个波导。光520在波导432b的输入边缘432c处被注入到波导432b中,并且通过TIR在波导432b内传播。在光520撞击在DOE 432a上的点处,一部分光如出射光束510离开波导。出射光束510被示出为基本上平行,但是取决于与波导432b相关联的深度平面,该出射光束510也可以以一定角度(例如,形成发散的出射光束)被重定向以传播到眼睛410。应该理解的是,基本上平行的出射光束可以指示具有光提取光学元件的波导,其中光提取光学元件将光外耦合以形成看起来被设置在距眼睛410较大距离(例如,光学无穷远)处的深度平面上的图像。其它波导或者其它光提取光学元件组可以输出更加发散的出射光束图案,这将需要眼睛410适应更近距离以将其聚焦在视网膜上并且将被大脑解释为光来自比光学无穷远更接近眼睛410的距离。

[0065] 图6是示出包括波导装置、将光耦合到波导装置或从波导装置光耦合光的光耦

合器子系统、以及控制子系统的光学系统的示意图,该光学系统用于生成多焦点立体显示器、图像或光领域。光学系统可以包括波导装置、将光耦合到波导装置或从波导装置光耦合光的光耦合器子系统、以及控制子系统。光学系统可用于生成多焦点立体、图像或光场。该光学系统可以包括一个或多个主平面波导632a(在图6中仅示出一个)以及与至少一些主波导632a中的每一个主波导相关联的一个或多个DOE 632b。平面波导632b可以类似于参考图4讨论的波导432b、434b、436b、438b、440b。光学系统可以使用分布波导装置沿着第一轴(图6所示的垂直轴或Y轴)中继光,并且沿着第一轴(例如,Y轴)扩展光的有效出射光瞳。分布波导装置可以例如包括分布平面波导622b和与分布平面波导622b相关联的至少一个DOE 622a(由双点划线示出)。分布平面波导622b在至少一些方面可以与主平面波导632b相似或相同,但具有与其不同的方位。类似地,至少一个DOE 622a在至少一些方面可以与DOE 632a相似或相同。例如,分布平面波导622b和/或DOE 622a可以分别由与主平面波导632b和/或DOE 632a相同的材料构成。图6所示的光学显示系统600的实施例可以集成到图2所示的可穿戴系统200中。

[0066] 中继的和出射光瞳扩展的光从分布波导装置被光耦合到一个或多个主平面波导632b中。主平面波导632b沿着优选地与第一轴正交的第二轴(例如,图6的视图中的水平轴或X轴)中继光。值得注意的是,第二轴可以是与第一轴非正交的轴。主平面波导632b沿着该第二轴(例如,X轴)扩展光的有效出射光瞳。例如,分布平面波导622b可以沿着垂直轴或Y轴中继和扩展光,并且将该光传递到沿着水平轴或X轴中继和扩展光的主平面波导632b。

[0067] 光学系统可以包括一个或多个彩色光源(例如,红色、绿色和蓝色激光)610,这些彩色光源可以光耦合到单模光纤640的近端中。可以穿过压电材料的中空管8来通过或接收光纤640的远端。远端作为非固定柔性悬臂644从管642突出。压电管642可以与四个象限电极(未示出)相关联。例如,电极可以镀在管642的外侧、外表面或外周或直径上。芯电极(未示出)也位于管642的芯、中心、内周或内径中。

[0068] 例如经由导线660电耦合的驱动电子器件650驱动相对的电极对独立地在两个轴上弯曲压电管642。光纤644的突出远端顶端具有机械谐振模式。谐振的频率可以取决于光纤644的直径、长度和材料特性。通过在光纤悬臂644的第一机械谐振模式附近振动压电管8,使得光纤悬臂644振动,并且可以扫过大的偏转。

[0069] 通过激发两个轴上的谐振,光纤悬臂644的顶端在遍及二维(2-D)扫描的区域中双轴扫描。通过与光纤悬臂644的扫描同步地调制一个或多个光源610的强度,从光纤悬臂644出射的光形成图像。美国专利公开No.2014/0003762中提供了这样的设置的描述,其通过引用全部并入在此。

[0070] 光学耦合器子系统的部件准直从扫描光纤悬臂644出射的光。准直光由镜面648反射到包含至少一个衍射光学元件(DOE)622a的窄分布平面波导622b中。准直光通过全内反射(TIR)沿分布平面波导622b垂直地(相对于图6的视图)传播,并且与DOE 622a重复相交。DOE 622a优选具有低衍射效率。这导致一部分光(例如,10%)在与DOE 622a的每个交点处被衍射朝向较大的主平面波导632b的边缘,并且一部分光通过TIR在其原始轨迹上向下分布平面波导622b的长度而继续。

[0071] 在与DOE 622a的每个交点处,附加光被衍射向主波导632b的入口。通过将入射光分成多个外耦合组,光的出射光瞳在分布平面波导622b中由DOE 4垂直地扩展。从分布平面

波导622b外耦合的该垂直扩展的光进入主平面波导632b的边缘。

[0072] 进入主波导632b的光经由TIR沿着主波导632b水平传播(相对于图6的视图)。由于光通过TIR沿着主波导632b的至少一部分长度水平传播,因此光在多个点处与DOE 632a相交。DOE 632a可以有利地被设计或构造成具有相位轮廓,该相位轮廓是线性衍射图案和径向对称衍射图案的总和,以产生光的偏转和聚焦。DOE 632a可以有利地具有低衍射效率(例如,10%),使得DOE 632a的每个交点只有一部分光束的光朝着视图的眼睛偏转,而其余的光经由TIR通过主波导632b继续传播。

[0073] 在传播光和DOE 632a之间的每个交点处,一部分光朝着主波导632b的相邻面衍射,从而允许光脱离TIR,并且从主波导632b的面出射。在一些实施例中,DOE 632a的径向对称衍射图案另外向衍射光赋予聚焦水平,既整形单个光束的光波前(例如,赋予曲率)以及以与设计的聚焦水平相匹配的角度将光束转向。

[0074] 因此,这些不同的路径可以通过多个DOE 632a以不同的角度、聚焦水平和/或在出射光瞳处产生不同的填充图案来使光耦合到主平面波导632b外。出射光瞳处的不同填充图案可以有利地用于创建具有多个深度平面的光场显示。波导组件中的每一层或堆叠中的一组层(例如3层)可用于产生相应的颜色(例如,红色、蓝色、绿色)。因此,例如,可以采用第一组的三个相邻层在第一焦深处分别产生红光、蓝光和绿光。可以采用第二组的三个相邻层在第二焦深处分别产生红光、蓝光和绿光。可以采用多组来产生具有各种焦深的全3D或4D彩色图像光场。

#### [0075] 可穿戴系统的其它部件

[0076] 在许多实施方式中,AR系统可以包括除了上述可穿戴系统的部件之外或者替代的其它部件。例如,可穿戴系统可以包括一个或多个触觉设备或部件。触觉设备或部件可以可操作以向用户提供触觉感觉。例如,触觉设备或部件可以在接触虚拟内容(例如,虚拟对象、虚拟工具、其它虚拟构造)时提供压力和/或纹理的触觉感觉。触觉感觉可以复制虚拟对象表示的物理对象的感觉,或者可以复制虚拟内容表示的想象对象或角色(例如,龙)的感觉。在一些实施方式中,触觉设备或部件可以由用户穿戴(例如,用户可穿戴手套)。在一些实施方式中,触觉设备或部件可以由用户持有。

[0077] 例如,可穿戴系统可以包括能由用户操纵以允许输入或与AR系统交互的一个或多个物理对象。这些物理对象在这里可以被称为图腾(totem)。一些图腾可采取无生命对象的形式,例如一块金属或塑料、墙、桌子的表面。在某些实施方式中,图腾可能实际上不具有任何物理输入结构(例如,键、触发器、操纵杆、轨迹球、摇杆开关)。相反,图腾可以简单地提供物理表面,并且AR系统可以渲染用户界面,以使用户看起来在图腾的一个或多个表面上。例如,AR系统可以渲染计算机键盘和触控板的图像从而看起来驻留在图腾的一个或多个表面上。例如,AR系统可以渲染虚拟计算机键盘和虚拟触控板从而看起来在用作图腾的薄铝矩形板的表面上。矩形板本身没有任何物理键或触控板或传感器。然而,AR系统可以检测用户操纵或交互或者触摸矩形板,作为经由虚拟键盘和/或虚拟触控板进行的选择或输入。用户输入设备466(图4中所示)可以是图腾的实施例,其可以包括触控板、触摸板、触发器、操纵杆、轨迹球、摇杆开关、鼠标、键盘、多自由度控制器,或者另一个物理输入设备。用户可以单独或与姿势一起使用图腾来与可穿戴系统和/或其它用户进行交互。

[0078] 在美国专利公开No.2015/0016777中描述了本公开的与可穿戴设备、HMD和显示系

统一起使用的触觉设备和图腾的示例,其全部内容通过引用并入本文。

[0079] 示例可穿戴系统、环境和界面

[0080] 可穿戴系统可以采用各种映射(mapping)相关技术以便在渲染的光场中实现高景深。在映射出虚拟世界时,了解现实世界中的所有特征和点以准确描绘与现实世界相关的虚拟对象是有利的。为此,从可穿戴系统的用户捕获的FOV图像可以通过包括新图片被添加到世界模型,这些新图片传达现实世界的各个点和特征的信息。例如,可穿戴系统可以收集一组地图点(例如2D点或3D点)并找到新的地图点以渲染更精确版本的世界模型。可以将第一用户的世界模型(例如,通过诸如云网络的网络)传送给第二用户,使得第二用户可以体验第一用户周围的世界。

[0081] 图7是MR环境700的示例的框图。MR环境700可以被配置为从一个或多个用户可穿戴系统(例如,可穿戴系统200和/或显示系统220)和/或静止房间系统(例如,房间照相机等)接收输入(例如,来自用户的可穿戴系统的视觉输入702,诸如房间照相机的静止输入704,来自各种传感器的感测输入706,来自用户输入设备504的用户输入、手势、图腾、眼睛跟踪等)。可穿戴系统可以使用各种传感器(例如,加速度计、陀螺仪、温度传感器、运动传感器、深度传感器、GPS传感器、面向内的成像系统、面向外的成像系统等)来确定用户的环境的位置以及各种其它属性。该信息可以进一步利用来自房间中的静止照相机的信息进行补充,这些信息可以从不同的角度提供图像和/或各种线索。由照相机(诸如房间照相机和/或面向外的成像系统的照相机)获取的图像数据可以缩减为一组映射点。

[0082] 一个或多个对象识别器708可以浏览接收到的数据(例如,点的收集)并且借助于地图数据库710来识别和/或映射点、标记图像,将语义信息附加到对象。地图数据库710可以包括随时间收集的各个点及其对应的对象。各种设备和地图数据库可以通过网络(例如LAN,WAN等)相互连接以访问云。

[0083] 基于该信息和地图数据库中的点的收集,对象识别器708a-708n可识别对象并用语义信息对其进行补充以赋予对象生命。例如,如果对象识别器识别一组点作为门,则系统可以附加一些语义信息(例如,门具有铰链并且在铰链周围具有90度的运动)。如果对象识别器识别出一组点作为镜子,则系统可以附加语义信息,即镜子具有可反射房间中对象的图像的反射表面。随着时间的推移,地图数据库会随着系统(可能驻留在本地或可能通过无线网络访问)积累更多来自世界的的数据而增大。一旦对象被识别,信息就可以被传送到一个或多个可穿戴系统。例如,MR环境700可以包括关于在加利福尼亚发生的场景的信息。该环境700可以被发送到纽约的一个或多个用户。基于从FOV照相机和其它输入接收的数据,对象识别器和其它软件部件可以映射从各种图像收集的点、识别对象等,使得场景可以准确地“传递”给可能在世界的不同地区的第二用户。环境700也可以使用拓扑图用于本地化目的。

[0084] 作为另一个示例,用户的环境可以与特定音频内容相关联。例如,音频文件的干音轨可以与用户房间的不同位置相关联。可穿戴系统可以将用户环境的地图与干音轨的位置信息一起传递给另一个用户,从而允许其它用户可以与干音轨交互。因此,这两个用户可以通过在用户房间中混合干音轨来协作创建新的音频文件。

[0085] 作为又一个示例,用户可以在他的环境中与干音轨交互以创建新的音频文件。用户的可穿戴系统可以将新的音频文件传递给其它可穿戴系统用于回放或播放新的音频文

件给其它人(例如通过扬声器)。

[0086] 图8是渲染与识别的对象相关的虚拟内容的方法800的示例的过程流程图。方法800描述如何将虚拟场景呈现给MR系统(例如,可穿戴系统)的用户。用户可能在地理上远离场景。例如,用户可能是纽约,但可能想要查看目前在加利福尼亚州正在进行的场景,或者可能想要与居住在加利福尼亚州的朋友散步。

[0087] 在框810处,AR系统可以从用户和其它用户接收关于用户的环境的输入。这可以通过各种输入设备和地图数据库中已有的知识来实现。在框810处,用户的FOV照相机、传感器、GPS、眼睛跟踪等向系统传达信息。在框820处,系统可以基于该信息确定稀疏点。稀疏点可用于确定姿势数据(例如,头部姿势、眼睛姿势、身体姿势和/或手势),这些姿势数据可用于显示和理解用户周围环境中各种对象的方位和位置。在框830处,对象识别器708a、708n可以使用地图数据库浏览这些收集的点并识别一个或多个对象。然后在框840处,该信息可以被传达给用户的个人可穿戴系统,并且在框850处,可以将期望的虚拟场景相应地显示给用户。例如,期望的虚拟场景(例如,加州的用户)可以与纽约的用户的各种对象和其它环境的相关地显示在适当的方位、位置等。

[0088] 图9是可穿戴系统的另一示例的框图。在该示例中,可穿戴系统900包括地图,该地图可以包括世界的地图数据。地图可以部分地驻留在可穿戴系统上,并且可以部分驻留在可以通过有线或无线网络(例如,在云系统中)访问的联网存储位置处。姿势处理910可以在可穿戴计算架构(例如,处理模块260或控制器460)上执行,并且利用来自地图的数据来确定可穿戴计算硬件或用户的位置和方位。姿势数据可以根据用户正在体验系统并在世界上操作时随时收集的数据来计算。数据可以包括图像,来自传感器(例如通常包括加速度计和陀螺仪组件的惯性测量设备)的数据以及与真实或虚拟环境中的对象有关的表面信息。

[0089] 稀疏点表示可以是同时定位和地图构建(SLAM或V-SLAM,参考其中输入只是图像/视觉的配置)过程的输出。该系统可以配置为不仅可以查找各个部件在世界中的位置,而且还了解世界由什么组成。姿势可以是实现许多目标的组成部件,包括填充地图和使用来自地图的数据。

[0090] 在一个实施例中,稀疏点位置本身可能并不完全足够,并且可能需要进一步的信息来产生多焦点AR、VR或MR体验。通常涉及深度地图信息的密集表示可以被用来至少部分地填补该缺口。这样的信息可以根据被称为立体(Stereo)940的过程来计算,其中使用诸如三角测量或飞行时间感测的技术来确定深度信息。图像信息和有源(active)图案(诸如使用有源(active)投影仪创建的红外图案)可以用作立体过程940的输入。大量的深度地图信息可以融合在一起,并且其中一些可以用表面表示来概括。例如,数学上可定义的表面是有效的(例如相对于大的点云)并且对诸如游戏引擎的其它处理设备是可消化输入。因此,可以在融合过程930中组合立体过程(例如,深度图)940的输出。姿势也可以是该融合过程930的输入,并且融合过程930的输出变成填充地图过程920的输入。例如在地形绘制中,子表面可以彼此连接以形成更大的表面,并且地图变成点和表面的大混合。

[0091] 为了解决混合现实过程960中的各个方面,可以使用各种输入。例如,在图9所示的实施例中,游戏参数可以是输入以确定系统的用户正在与各种位置处的一个或多个怪物进行怪物作战游戏,在各种条件下死亡或逃跑的怪物(诸如如果用户射击怪物),各种位置处的墙壁或其它对象等。世界地图可以包括关于这些对象在哪里彼此相关的信息,作为对混

合现实的另一个有价值的输入。与世界相关的姿势也成为一种输入，并且对几乎任何交互系统都起着关键作用。

[0092] 来自用户的控制或输入是可穿戴系统900的另一输入。如这里所述，用户输入可以包括视觉输入、手势、图腾、音频输入、感觉输入等。为了在周围移动或玩游戏，例如，用户可能需要指示可穿戴系统900关于他或她想做什么。除了仅仅在空间中移动自己之外，还存在可以使用的各种形式的用户控制。在一个实施例中，图腾、用户输入设备或诸如玩具枪的对象可由用户握持并由系统跟踪。系统优选地将被配置为知道用户正握持物品并且理解用户与物品具有什么样的交互（例如，如果图腾或对象是枪，则系统可以被配置为理解位置和方位，以及用户是否在点击触发器或可以配备有传感器（诸如IMU）的其它感测按钮或元件，这可以有助于确定正在发生什么，即使当这种活动不在任何照相机的视场内。）

[0093] 手势跟踪或识别还可以提供输入信息。可穿戴系统900可以被配置成跟踪和解释按钮按压的手势，用于动作表示左或右、停止、抓住、握持等。例如，在一种配置中，用户可能想要翻阅非游戏环境中的电子邮件或日历，或与另一个人或玩家进行“拳头碰撞”。可穿戴系统900可以被配置为利用最小量的手势，其可以是动态的也可以不是动态的。例如，手势可以是简单的静态手势，如张开的手用于停止，拇指向上用于ok，拇指向下用于非ok；或者手向右或左、或上/下轻击用于方向命令。

[0094] 眼睛跟踪是另一输入（例如，跟踪用户正在看的地方从而控制显示技术以在特定深度或范围渲染）。在一个实施例中，可以使用三角测量来确定眼睛的聚散度，然后使用针对特定人开发的聚散度/适应性模型，可以确定适应性。

[0095] 关于照相机系统，图9中示出的示例可穿戴系统900可以包括三对照相机：布置在用户脸部两侧的相对宽的FOV或无源SLAM照相机对，定向在用户前面的不同照相机对，以处理立体成像过程940并且还捕获手势和用户面前的图腾/对象跟踪。三对照相机中的照相机可以是面向外的成像系统464（在图4中示出）的一部分。可穿戴系统900可以包括眼睛跟踪照相机（其可以是图4中所示的面向内的成像系统462的一部分），其朝向用户的眼睛定位从而对眼睛向量和其它信息进行三角测量。可穿戴系统900还可以包括一个或多个纹理光投影仪（诸如红外（IR）投影仪）以将纹理注入到场景中。

[0096] 图10是用于确定向可穿戴系统的用户输入的方法1000的示例的过程流程图。在这个示例中，用户可以与图腾进行交互。用户可具有多个图腾。例如，用户可具有社交媒体应用的指定的一个图腾，玩游戏的另一个图腾等。在框1010处，可穿戴系统可以检测图腾的运动。图腾的运动可以通过用户的FOV照相机进行识别或者可以通过传感器（例如，触觉手套、图像传感器、手跟踪设备、眼部跟踪照相机、头部姿势传感器等）进行检测。

[0097] 在框1020处，至少部分地基于检测到的手势、眼睛姿势、头部姿势或通过图腾的输入，可穿戴系统检测图腾（或用户的眼睛或头部或手势）相对于参考系的位置、方位和/或移动。参考系可以是一组地图点，可穿戴系统根据该地图点将图腾（或用户）的运动转换为动作或命令。在框1030处，与图腾的用户交互被映射。基于用户交互相对于参考系1020的映射，在框1040处系统确定用户输入。

[0098] 例如，用户可以来回移动图腾或物理对象以表示翻动虚拟页面并移动到下一页面或者从一个用户界面（UI）显示屏幕移动到另一个UI屏幕。作为另一个示例，用户可以移动他们的头部或眼睛来查看用户FOR中的不同的现实或虚拟对象。如果用户注视特定现实或

虚拟对象的时间长于阈值时间,则可以选择现实或虚拟对象作为用户输入。在一些实施方式中,可以跟踪用户眼睛的聚散度并且可以使用适应性/聚散度模型来确定用户眼睛的适应状态,其提供用户正在聚焦的深度平面的信息。在一些实施方式中,可穿戴系统可以使用光线投射技术来确定哪些现实或虚拟对象是沿着用户的头部姿势或眼睛姿势的方向。在各种实施方式中,光线投射技术可包括投射具有基本上很小横向宽度的薄的铅笔射线或具有基本横向宽度(例如锥体或平截头体)的投射光线。

[0099] 用户界面可由在此所述的显示系统(例如图2中的显示器220)投影。它也可以使用各种其它技术显示,例如一个或多个投影仪。投影仪可以将图像投影到例如画布或球体的物理对象上。可以使用系统外部或系统的一部分的一个或多个照相机(例如,使用面向内的成像系统462或面向外的成像系统464)来跟踪与用户界面的交互。

[0100] 图11是用于与虚拟用户界面进行交互的方法1100的示例的过程流程图。方法1100可以由在此描述的可穿戴系统来执行。

[0101] 在框1110处,可穿戴系统可以识别特定UI。UI的类型可以由用户预先确定。可穿戴系统可以基于用户输入(例如,手势、视觉数据、音频数据、感测数据、直接命令等)识别特定UI需要填充。在框1120处,可穿戴系统可以生成针对虚拟UI的数据。例如,可以生成与UI的界限、一般结构、形状等相关的数据。另外,可穿戴系统可以确定用户的物理位置的地图坐标,使得可穿戴系统可以显示与用户的物理位置相关的UI。例如,如果UI是以身体为中心的,则可穿戴系统可以确定用户的身体姿态、头部姿势或眼睛姿势的坐标,使得可以在用户周围显示环形UI或者可以在墙上或在用户面前显示平面UI。如果UI是以手为中心的,则可以确定用户的手的地图坐标。这些地图点可以通过以下数据得出,通过FOV照相机、感测输入所接收的数据或任何其它类型的收集数据。作为示例,可穿戴系统可以在用户周围生成音频文件中的各种干音轨的图形表示。

[0102] 在框1130处,可穿戴系统可以将数据从云端发送到显示器,或者可以将数据从本地数据库发送到显示器部件。在框1140处,基于发送的数据向用户显示UI。例如,光场显示器可以将虚拟UI投影到用户的一只或两只眼睛中。一旦创建了虚拟UI,在框1150处,可穿戴系统可以简单地等待来自用户的命令以在虚拟UI上生成多个虚拟内容。例如,UI可以是围绕用户身体的身体中心环。然后,可穿戴系统可以等待命令(手势、头部或眼部运动、来自用户输入设备的输入等),并且如果识别出(框1160),则可以向用户显示与该命令相关联的虚拟内容(框1170)。作为示例,可穿戴系统可以在混合多个流轨道之前等待用户的手势。

[0103] 在美国专利公开No.2015/0016777中描述了AR系统、UI和用户体验(UX)的附加示例,其全部内容通过引用合并于此。

#### [0104] 干音轨的示例

[0105] 如参考图2和4所述,可穿戴系统可以向用户播放音频文件并呈现音频内容的可视化。可穿戴系统可以是配置为呈现AR、VR和/或MR环境的可穿戴设备(例如,头戴式设备/显示器)的一部分。音频文件可以使用本地处理和数据模块260、远程处理模块270或远程数据存储库280、进行组合等来存储和处理。

[0106] 音频文件可以包括多个音轨。例如,音频文件可以包括立体声主音轨、用于立体声系统的“右”或“左”声道的音轨、用于环绕声道的音轨等等。某些类型的音频文件可以包括表示不同乐器(例如吉他、贝斯、鼓、钢琴、合成器、喇叭)或人声(例如,主唱、伴唱)的多个音

轨。表示不同乐器或人声的音轨有时被称为干音轨或干。在各种实施方式中,干音轨的数量可以是1、2、3、4、5、6、7、8、10、12或更多。

[0107] 干音轨可以是音频文件的任何划分,其中每个干音轨表示音频文件的不同音频内容。作为示例,干音轨可以表示记录中的特定乐器或人声(例如,吉他或主唱)。例如,室内音乐(例如,弦乐四重奏)的记录可以包括代表第一小提琴、第二小提琴、中提琴和大提琴的四个单独的干音轨。作为另一示例,干音轨可以表示一组乐器或人声(例如,贝斯和鼓或合唱的节奏部分)。例如,对于管弦乐队来说,可能有不同的干音轨,表示小提琴、木管乐器、打击乐器、铜管乐器、贝斯等。干音轨不限于音乐记录,可用于语音的记录(例如,不同发言者的不同干音轨)。

[0108] 干音轨可以与用户的FOR中的空间位置相关联。因此,干音轨的声音可能听起来似乎来自其相关的空间位置。例如,在干音轨与用户的右侧相关联的情况下,干音轨的声音可能似乎来自用户的右侧。随着用户改变他的姿势,用户可以感知不同的声场或视场。在同一示例中,如果用户转动180度,声音可能似乎来自用户的左侧(而不是右侧)。这有利地提供独特的、用户可控的用户体验。

[0109] 干音轨的位置可以相对于用户(例如,用户的头部)来限定,诸如以相对于用户的特定距离和/或角度。另外或可选地,干音轨的位置可以全局限定。举例来说,可以将干音轨定位在前面的墙壁上,而将另一干音轨定位在用户房间的中间。

[0110] 干音轨的位置可随时间变化。例如,干音轨可与歌手的声音相关联。干音轨的位置可以改变,就好像歌手(虚拟地)在用户的环境中四处移动一样。干音轨的位置也可以基于用户交互而改变。如下面进一步描述的,用户可以移动干音轨,例如通过抓取与干音轨相关联的视觉图形。

[0111] 可穿戴系统可以提供单独干音轨和/或组合的一个或多个干音轨的可视化。因此,可穿戴系统可以有利地允许用户可视化音频空间化并且与干音轨交互。

[0112] 在一些实施方式中,音频文件可以包括干音轨加上附加信息(例如,元数据),附加信息可以由可穿戴设备用来渲染音乐的音频或视觉表示。例如,附加信息可以包括与舞台上的音乐演奏者的移动有关的信息(例如,在舞台周围移动或跳舞的主唱),关于演奏场所的音质或大小的信息(例如,舞台的大小、音乐厅的大小等)等等。可穿戴系统可以使用这些附加信息来呈现音乐艺术家对表演的视觉表示。例如,主唱的运动信息可用于确定当主唱在演出期间四处移动时,应当将主唱的表现(例如,图像或化身)显示在用户视野中的哪个位置。这样的实施例可以用于创建记录音频轨道所在的现场表演的视觉渲染。

[0113] 可以从音频文件的元数据中解析关于音频干的信息、干的时间序列位置数据、干音轨的可视化以及一条或多条附加的信息。例如,音频文件可以支持这种元数据的格式编码。元数据可以指定干音轨和音频文件的相关干音轨的可视化(例如,形状、动画等)。元数据还可以指定活动的(active)干音轨以及在音频文件的给定时间的干音轨的位置。音频文件可以是专有格式。为了允许音频文件与可能不具有提供在此所公开的可视化和混合干的功能(例如,提供向后兼容性)的音频设备一起使用,音频格式可以存储这样的元数据,以使得较旧的音频系统可以例如忽略元数据(例如,通过不从文件中读取)以及只读取声音文件的立体混合。

[0114] 混合干音轨的示例

[0115] 两个或多个干音轨可以进行组合以产生音频混合。音频混合可以模拟反映干音轨的空间位置的声场。音频混合也可以反映干音轨的混合,而不管它们的空间位置。

[0116] 用于不同乐器或人声的干音轨的组合(或混合)产生可呈现给用户以供可穿戴系统(例如,经由扬声器240)回放的音频混合。音频混合可以用于产生一个或多个音频通道(例如,立体声或环绕声(其可具有5、7或更多通道))。可穿戴系统还可以修改音频混合,诸如干音轨水平(例如,响度或柔和度)、频率内容、动态性(例如,风格(断奏或连奏)、速度)、全景位置(有时被称为平移,例如,在立体声或多声道声场之中的音乐声音的分布),并且可以添加效果(例如,混响),其组合或等。

[0117] 用户可以单独或组合地使用姿势或用户输入设备与干音轨交互。作为一个示例,混合的空间化或可视化可以(至少部分地)基于用户的手部运动或通过从用户输入设备466(例如,用户或第三方可控制的鼠标或游戏棒)接收的控制信号。可穿戴系统可以识别用户手的位置。随着手的移动,可穿戴系统可以生成不同的干音轨混合以向用户(听觉地或视觉地)呈现。在一些实施方式中,用户可以使用他或她的手(或指示笔或指挥棒)来“触摸”或“移动”通过可穿戴系统的显示器220看到的真实或虚拟对象,并且可穿戴系统可以响应于触摸或移动来修改或调整混合。作为另一个示例,用户可以识别干音轨(例如,使用用户输入设备)并且使用他的手臂沿着轨迹移动干音轨。干音轨可能会与另一干音轨相撞。可穿戴系统可以向用户呈现这两个干音轨的混合。

[0118] 在一些实施方式中,当用户正在收听音频文件时,可穿戴系统可以基于用户的姿势(诸如头部、眼睛、脚部、手势或其它身体姿势)动态地混合音频文件的干音轨。例如,随着用户的眼睛或头部移动,可穿戴系统(例如,经由处理模块260和/或270)可以动态地混合音频文件中的干音轨以创建“新”的一组声音,用于呈现给用户(例如,通过扬声器240)。因此,可穿戴系统可以有利地空间化音频文件的干音轨的混合。

[0119] 另外,可穿戴系统(例如,经由处理模块260和/或270)可以基于用户的姿势,向用户(例如,经由显示器200)动态地呈现单独干音轨的可视化或混合干音轨的可视化。这允许用户在用户的头部或眼睛(双眼)姿势改变时看到干音轨的音频混合的不同视觉表示。

[0120] 在一些实施方式中,音频文件可以包括干音轨以及与干音轨的一个或多个预混合组合相关联的信息。在这样的实施方式中,不是随着用户的姿势改变而动态地混合干音轨,而是可穿戴系统可以访问预混合的组合中的一个以呈现给用户。可穿戴系统可以单独或组合地基于用户位置、用户姿势、来自用户输入设备的输入等来选择预混合组合。例如,预混合组合可以如同音乐在用户的头部周围环绕一样来再现音乐。

[0121] 可穿戴系统的用户还可以“锁定”声音源在环境(诸如,用户的房间)中或相对于对象(诸如,用户的头或身体)的位置。例如,用户可以在他的环境中移动干音轨。一旦用户找到他喜欢的音频混合,他可以锁定干音轨在音频混合中的空间位置,例如通过启动用户输入设备或通过他的姿势。无论用户在其环境中的位置如何,音频混合都可以因此保持不变。用户可以将“锁定的”混合存储在音频文件中,以使用户将来可以通过回放音频文件来重新体验视场和声场。

[0122] 呈现干音轨的示例用户界面

[0123] 图12、13和14示意性地示出了向可穿戴系统的用户呈现音频文件的多个干音轨的可视化的用户界面的示例。可以至少部分地基于用户的姿势来动态地混合干音轨。

[0124] 图12示意性地示出了两个示例用户界面1210和1220。在示例界面1210和1220中，示出了五个干音轨的混合。在示例用户界面1210中，五个不同的视觉图形（例如，图标或者球体）1210a-1210e示出为表示相应的五个干音轨。视觉图形不需要是静态的，并且可以在播放音频文件时被移动或修改或改变。在该示例中，当用户的头部或眼睛姿势改变时，音频混合改变并且视觉表示1210相应地改变。例如，用户的眼睛（或头部）姿势可以指示用户正朝着图形1210a的方向看。在一些这样的情况下，图形1210a可以被显示为比不在用户注视方向上的图形1210b-1210e被强调（例如，更大、更亮等）。例如，通过增大音量也可以强调与图形1210a相关联的声音。当用户的姿势改变时，可以强调另一视觉图形，或者可以向用户呈现不同的声场（对应于干音轨的不同音频混合）和不同的视场。

[0125] 在一些实施方式中，可穿戴系统可以基于音频文件的内容来确定是否强调视觉图形。例如，可穿戴系统可以强调视觉图形1210a，因为视觉图形1210a对应的干音轨在音量上被强调或者是音频文件中的主导音轨。

[0126] 在示例性用户界面1220中，视觉图形1220a和1220c被呈现为利萨如(Lissajous)图。在该示例中，尽管在用户环境中可能有额外的干音轨，在用户的FOV中的音频文件中可能有5个干音轨（1220a到1220e）。然而，与视觉图形1220b、1220d和1220e对应的干音轨被静音。因此，可穿戴系统不在这些方向上显示（或大小或亮度显著减小）视觉图形（点1220b、1220d和1220e被显示为点以指示没有（或减少的）图形显示给用户）。用户将会听到（例如，通过扬声器240）仅仅与方向1220a、1220c相关联的干音轨的音频混合。

[0127] 在各种实施方式中，视觉图形1210a-1210e、1220a-1220e可以彼此不同或相同。可以呈现许多类型的视觉图形（例如参见图13和图14中的示例），并且这里描述的示例旨在是说明性的而非限制性的。例如，图12A中示出的利萨如图1220a、1220c可以是动态的，并且它们的图案可以随时间（例如，随着音乐的节拍）改变。

[0128] 在一些实施方式中，可穿戴系统可以（至少部分地）使用干音轨的音频频谱（频率）内容来确定用户的视场中对应的视觉图形的可视化。音频频谱内容可以用于确定视觉图形的放置。例如，可以在第一方向上显示高频，并且可以在不同的第二方向上显示低频。作为另一示例，具有不同频谱内容的声音可以由视觉图形在用户的FOR的不同部分中（包括用户将不得不转身看到或听到的方向）进行表示。音频频谱内容还可以用于确定视觉图形的外观。例如，可穿戴系统可以基于干音轨的音频频谱内容生成与该干音轨相对应的利萨如图。

[0129] 用户界面1210和1220可以以VR显示模式呈现，其中用户看到视觉图形1210a-1210e或1220a-1220e，但不看到外部世界470（在图4中示出）。用户界面1210和1220也可以以AR或MR显示模式呈现，其中用户看到叠加在外部世界470（图4中所示）上的视觉图形1210a-1210e或1220a-1220e。

#### [0130] 与干音轨的示例交互

[0131] 用户可以改变干音轨的个别声音特征。作为与处于AR显示模式中的用户界面1210、1220的用户交互的示例，音乐节目主持人(DJ)可以在通过可穿戴系统显示干音轨时使用手势来控制干音轨的个别声音特征（例如，音量）。利用可穿戴系统的DJ可以“抓取”视觉图形中的一个，诸如视觉图形1210a，以控制该干音轨的音量。如果视觉图形1210a表示人声音轨，则DJ可以压缩他的手以减小该干音轨的音量。干音轨的其它声音特征可能会改变或修改。例如，如果DJ要选择不同的声音特征模式（例如，通过做出一些其它手势、脚部姿

势),则DJ可以改变或修改由视觉图形1210a表示的人声音轨的另一个声音特征。如果声音特征模式改变为音调,则DJ在这种情况下压缩她的手将会降低音调。另一方面,如果DJ想要升高音调,则DJ可以使用另一手势来升高与视觉图形1210a相关联的音调。例如,DJ可通过在与图形1210a相关联的区域(例如,在图形上或附近)以顺时针或逆时针激情快速移动她的手来“旋转”图形1210a。

[0132] 改变或修改各种干音轨的其它声音特征是可能的。例如,其它声音特征可以包括音量、音调、低音(bass)、高音、音调质量、强度、持续时间等。虽然前面的示例已经在手势的上下文中进行了描述,但是也可以使用其它身体姿势来控制出现在可穿戴系统显示屏上的干音轨的声音特征。

[0133] 另外或替代地,一些干音轨可以以VR模式显示。姿势还可用于改变或修改VR模式中显示的干音轨。作为另一个示例,DJ可以轻敲他们脚中的一只脚两次(例如,在第一次轻敲的特定时间段内轻敲两次脚)以指示干音轨的声音特征要被改变或修改。通常,手势的各种实施例可以用于与视觉显示交互。尽管已经根据姿势描述了前述示例,但是来自用户输入设备的控制信号也可以改变干音轨。例如,用户可以利用操纵杆、图腾或触敏输入设备来改变被表示为视觉图形的特定干音轨的音调。

[0134] 干音轨或其部分也可以由可穿戴系统的用户生成。用户生成的干音轨可以替换音频文件的相应的一个(或多个)干音轨。例如,用户可以将用户生成的干音轨替换为音频的原始干音轨之一。用户可以录制(用户或其它人的)人声音轨,并用用户人声音轨替换录制的原始人声音轨。因此,正在播放的干音轨可能听起来好像用户在音频录制中唱歌一样。在另一个实施例中,如果两个干音轨对应于二重奏的相应人声音轨,则用户生成的人声音轨可以替代二重奏中的一个声音。作为又一示例,用户生成的吉他独奏可以替换音频录制的相应的吉他干音轨。任何数量的音频干音轨都可以利用用户生成的干音轨全部或部分替换。

[0135] 图13示意性地示出了包括干音轨的音频文件的MR呈现1300的示例。在该示例中,用户穿过显示器220(如图2所示)向外看向包含真实(物理)对象(诸如沙发1312)的房间1310。在图13中仅示出一个物理对象用于说明目的,而这不是限制。

[0136] 在该示例中,音频文件包括与人声、吉他、贝斯和鼓对应的干音轨。可穿戴系统呈现出:人弹奏吉他的形式的视觉图形1322用于吉他音轨、人弹奏贝斯吉他的形式的视觉图形1328用于贝斯音轨,人歌唱形式的视觉图形1326用于人声音轨,以及人弹奏架子鼓形式的视觉图形1324用于鼓音轨。与图12中所示的用户界面1210和1220相似,不同的视觉图形1322、1324、1326和1328在房间1310中的不同方向进行显示。在该MR呈现中,吉他演奏者的视觉图形1322被呈现为好像吉他演奏者正坐在实际存在于房间1310中的(物理)沙发1312上。视觉图形1324、1326和1328呈现在没有物理对象的房间的区域中。

[0137] 由于这里描述的可穿戴系统的实施例能够在用户的视场中的多个深度平面处呈现对象(参见例如图3的讨论,图4的显示系统400的波导组件480中的波导432b、434b、436b、438b,或者图6的显示系统600的波导632b),用户界面1300示出了房间中不同深度处的不同视觉图形1322、1326、1324和1328。例如,贝斯吉他手(视觉图形1328)朝向房间1310的前方,主唱(视觉图形1326)在房间1310中更向后但在鼓手(视觉图形1324)前方,鼓手(视觉图形1324)位于房间的后面,离用户最远。此外,在此示例中,不同视觉图形中的细节数量可以取

决于深度。例如,因为鼓手(视觉图形1324)看起来距离用户最远,所以视觉图形1324具有比看起来更接近用户的视觉图形1326和1328更少的细节。在音频文件包括与(例如,主唱的)运动有关的附加信息的实施例中,例如可穿戴系统可以将表示歌手的视觉图形1326呈现为在房间1310中四处移动。

[0138] 图13所示的示例性用户界面1300相应地示出了其中在用户的外部世界环境470(图4中示出)中可视地表示不同干音轨的示例。例如,可穿戴系统可以通过分析由面向外的成像系统获取的图像来识别环境470中的对象(例如,房间1310中的沙发1312),或者可以与设置在环境470中的可以用来辅助显示视觉图形的图腾或电子跟踪器进行通信。例如,可以在房间1310中的图腾的位置处可视化用于特定干音轨的视觉图形。

[0139] 随着用户在房间1310中四处移动,干音轨的呈现可以相应地改变。例如,当用户走近房间的后部时,可穿戴系统可以更详细地呈现视觉图形1324。可穿戴系统还可以增加与视觉图形1324相关联的鼓音轨的音量,以反映用户更接近鼓手。在一些实施方式中,可穿戴系统可以基于用户的位置来调整干音轨的空间化。例如,当视觉图形1326呈现在用户的前方时,可穿戴系统可以播放音频文件,就好像人声音轨来自用户的前方。然而,随着用户走过视觉图形1326并朝向视觉图形1324,可穿戴系统可以播放音频文件,就好像人声音轨来自用户的右侧。

[0140] 在某些实施方式中,当用户四处走动时,可穿戴系统可以更新视觉图形的位置。例如,参考图13,当用户四处移动时,视觉图形1322、1324、1326和/或1328可“跟随”房间1310中的用户。

[0141] 可穿戴系统的用户可以选择性地选择音频音轨的期望混合。例如,如果干音轨对应于人声、吉他、贝斯和鼓,则用户能够只收听其中一个干音轨(看到其中一个干音轨的视觉表示)(例如,只是人声或只是吉他)。可穿戴系统可以向用户(视觉和听觉)呈现仅用户FOV中的干音轨。当用户移动他或她的头部或眼睛时,干音轨可能会移入或移出用户的FOV。因此,用户可以听(或看)不同的干音轨或不同的干音轨混合。在一些实施方式中,用户可以通过使用适当的手势(例如,手部姿势)使音频音轨静音。例如,用户可能想要听吉他独奏,并且可能会将除吉他音轨之外的所有音轨静音。

[0142] 另外或可选地,如上面关于与视觉显示器交互以控制声音特征的示例所描述的,还可以使用手势在VR、AR和/或MR显示模式中修改或改变视觉图形。作为一个示例,在MR显示模式下,用户可以用手按压低音吉他手(例如视觉图形1328),以压缩与由视觉图形1322、1326、1324和1328所播放的音轨的声音相关联的低音。视觉图形1328的最终显示可以看起来小于其它图形1322、1326或1324。作为另一个示例,用户可以使用手势来控制由视觉图形1322、1326、1324和1328表示的干音轨的音量。用户可以朝向房间1310的前方“抓取”并“拖”鼓手的视觉图形1324以增加鼓的音量。相反,为了减小人声的音量,用户可以朝房间1310的后部“推回”鼓手的视觉图形1324。

[0143] 视觉图形也可以使用其它姿势来修改或改变。例如,耦合到可穿戴系统200的其它传感器,诸如惯性测量单元(IMU),可以用于检测用户的头部或身体姿势。例如,IMU可以检测到用户向后和向前点头,这可对应于头部姿势的改变。可穿戴系统可以使用头部姿势的这种改变来选择视觉图形1322、1324、1326和1328中的一个。例如,在用户正在观看贝斯的视觉图形1328并点头时,可穿戴系统可以选择与贝斯对应的干音轨。在一些情况下,足够

“急剧的”头部运动(例如,IMU测量高于阈值)可以指示视觉图形的选择,而更平缓的头部运动(IMU测量低于阈值)可以指示与图形相关联的性质的调整。例如,当用户朝房间1310的后部“推回”视觉图形1326时,用户可以朝视觉图形1328点头,以指示视觉图形1328的选择,例如贝斯干音轨。一旦被选择,用户可以以某种方式(例如,抑制贝斯干音轨)改变该干音轨,或者如上所述关于干音轨的声音特征的改变。

[0144] 另外地或可替换地,在用姿势改变选择的干音轨时,用户还可以执行特定的干音轨并且替换由可穿戴系统正在播放的音频音轨的对应干音轨(例如,如上述关于替换干音轨的示例)。例如,用户可以替换图形1326的人声,同时改变由其它视觉图形表示的其它干音轨,例如表示贝斯干音轨的视觉图形1328。此外,可以使用除头部姿势或手势之外的身体姿势。例如,脚部姿势的改变(例如,轻敲脚部)可以指示踏上吉他或鼓踏板、击打鼓、启动音频效果(例如,混响)等等。

[0145] 图14示意性地示出了可由可穿戴系统呈现的用户界面1400的另一示例。在这个示例中,音频文件包括演奏一首室内乐的弦乐四重奏(两把小提琴、一把中提琴和一把大提琴)。当可穿戴系统的用户收听音频文件时,可穿戴系统向用户呈现用户界面1400并且示出表示小提琴的视觉图形1410和1420、表示大提琴的视觉图形1430以及表示该中提琴的视觉图形1440。当用户的头部或眼睛姿势发生变化时,这四种乐器的不同组合将听觉地呈现给用户。例如,如果用户看向小提琴1410和1420的方向,则用户可以听到其中小提琴声音在大提琴和中提琴声音上被强调(其可以被静音或以比小提琴更低的音量播放)的音频。小提琴的视觉图形1410和1420可以相应地针对大提琴和中提琴的视觉图形1430和1440加以强调。例如,小提琴的弓可以显示为移动,而大提琴和中提琴的弓可以显示为不移动。尽管在该示例中,视觉图形1410、1420、1430和1440显示为在共同深度平面处显示,但这不是限制,并且在其它示例中,表示不同乐器的视觉图形可以呈现在距离用户不同的深度平面处。图15中示出这种3D表示的其它示例。

[0146] 可穿戴系统还可以被配置为允许与所显示的视觉图形进行用户交互。例如,用户界面1400可以呈现各种干音轨,用户可以以干音轨正被表示或显示的特定方式来与其交互。特定的视觉图形可以是可以被改变或修改的表示。作为示例,在MR显示模式下,如果用户“抓取”中提琴1440(例如,如上面关于与MR模式中显示的图形交互的示例所述的“抓取”中提琴),则用户可以“演奏”中提琴来修改中提琴声音的特征。继续该示例,在用户“抓取”中提琴之后,用户可以使用可穿戴系统生成的“弓”来“演奏”中提琴,由此修改中提琴声音的特征(例如音量)。作为MR显示模式的另一个示例,用户可以“演奏”中提琴以替换正在播放的相应的中提琴干音轨。例如,用户可以使用可穿戴系统生成的“弓”来演奏中提琴并生成中提琴干音轨。在一个实施例中,该用户生成的中提琴干音轨可以替换弦乐四重奏中的中提琴干音轨,使得对用户产生的可听声音是用户演奏弦乐四重奏的中提琴部分。

[0147] 虽然上述示例已经在室内乐器的视觉图形的上下文中进行了描述,但是其它视觉图形可以充当UI以改变或修改所生成的UI所表示的对象。例如,对于具有电吉他的音轨,UI可以包括具有琴弦、开关和旋钮的电吉他的表示。用户可以拨或弹奏吉他弦,调整拨动开关或旋转旋钮来控制音轨的各个方面。视觉图形可以是待修改的对象的仿实物(skeuomorphic)表示。例如,包括数字键盘的电话图形可以表示与可穿戴系统相关联的拨号功能,使得用户可以拨打电话号码。作为另一个示例,日历的AR表示可以模仿纸张桌面日

历史上的装订外观。一般而言,由可穿戴生成的任何视觉图形对象可以用作用户界面。

[0148] 在三维空间中呈现干音轨

[0149] 如参考图13所示,可穿戴系统可呈现3D用户界面,其在用户环境中的不同深度处显示视觉图形。图15示出3D用户界面的另一示例。

[0150] 在图15中,在用户的FOV中有4个视觉图形1510、1520、1530和1540。这些视觉图形可以是利萨如图或其它图形表示。视觉图形的外观可以对应于正在播放给用户的音频内容。例如,因为视觉图形1510相关的干音轨在音频文件中被强调(例如,在混合中播放得更响),用户可以将视觉图形1510感知为离他最近。视觉图形1520以减小的尺寸显示,因为其相关联的干音轨可以相对静音。视觉图形可以随着音频内容的变化而改变。例如,当与视觉图形1520相关联的干音轨不再被静音时,视觉图形1520可能不再以缩小尺寸出现。

[0151] 可穿戴系统可以基于视觉图形相关联的干音轨的空间位置来放置视觉图形。例如,视觉图形1520可以看起来很小,因为其相应的声音来自遥远的位置(例如,管弦乐队后面的鼓手),而视觉图形1510可看起来更大,因为其相应的声音来自更近的位置(例如,在管弦乐队前面的歌手)。作为另一示例,因为与视觉图形1540相关联的干音轨可能比与视觉图形1530相关联的干音轨更远离用户,所以视觉图形1540看起来在视觉图形1530后面。图形1510-1540的相对距离或位置可以被显示从而为用户提供3D体验,并且可以随着音乐改变或随用户姿势改变而动态地改变(在距离、位置、尺寸、颜色、形状等方面)。图形1510-1540可以是可交互的,使得用户可以选择它们中的一个或多个(例如,通过用户的手伸出)并且四处移动所选择的图形,这不仅可以改变可视化的属性而且可以改变声音的属性(例如,使得与所选图形相对应的干音轨更响、更柔和,具有或多或少的低音或高音等)。

[0152] 可穿戴系统还可以将视觉图形移出用户的FOV。图15示出了3D视图1500中的地平线1550。可穿戴系统可以显示虚拟图形淹没在地平线1550下或从地平线1550上升。作为示例,当干音轨的音量减小时,可穿戴系统可以显示其相关的视觉图形在地平线1550下方移动。因此,用户可能无法看到地平线1550下方的视觉图形部分。作为另一个示例,当干音轨的音量升高时,可穿戴系统可以显示相关的视觉图形从地平线1550上升。在一些实施方式中,可穿戴系统可以将地平线显示为反射表面(诸如模拟水面或玻璃),其中可穿戴系统可以在地平线1550上显示视觉图形1510、1520、1530和1540的反射图像。

[0153] 在一些实施例中,只有当用户足够靠近视觉图形时,可穿戴系统才可以显示视觉图形。例如,在图15中,系统可以被配置为不显示视觉图形1520,因为视觉图形1520离用户太远。如参照图16B进一步描述,可穿戴系统可以被配置为仅显示视觉图形,当用户还可以听到与视觉图形相关联的声源时,例如当用户在定向声源的可听区域内时。

[0154] 作为另一个示例,图16A示出了用户环境的鸟瞰图1600a。在可穿戴系统正在播放一段音乐的同时,可穿戴系统可以呈现五个(在该示例中)视觉图形1610a、1620a、1630a、1640a和1650a,每个视觉图形与音乐的干音轨相关联。在这个示例中,五个可视化图形被放置在用户环境的不同位置处。用户1670a可以站在房间的中间并被五个视觉图形包围。在给定时间,用户1670a也可以在他的FOV中看到视觉图形的子集,即使在他的环境中存在其它视觉图形。例如,用户1670a可以在他的FOV中同时看到视觉图形1630a和1640a,但因为视觉图形1610a在用户1670a后面所以没有看到视觉图形1610a。随着用户1670a改变他的姿势,其它视觉图形可变得可见。例如,当用户1670a转身时,用户可以感知视觉图形1610a,但是

因为视觉图形1630a和1640a在用户的FOV之外它们可能不再可感知。

[0155] 如参考图13所述,随着用户1670a在他的环境中移动,视觉图形的外观可能相应地改变。例如,随着用户1670a接近视觉图形1610a,可穿戴系统可以呈现视觉图形1610a的更多细节。当用户正在看视觉图形1610a时,可穿戴系统还可以强调视觉图形1610a,例如使其看起来更大且更亮。然而,当用户转向不同的方向时,视觉图形1610a可以减小尺寸或亮度,而用户的注视方向上的另一个视觉图形(例如1650b)可以被强调(如图16B所示)。

[0156] 作为另一个示例,用户1670a可以在他的FOV中初始感知视觉图形1630a和1640a。但是随着用户1670a移动靠近视觉图形1630a,视觉图形1640a可能在用户的FOV之外,并且因此用户1670a可不再能够感知视觉图形1640。

[0157] 在某些情况下,尽管视觉图形可以基于用户的位置而改变,但音频内容可以保持不变(例如,因为用户锁定了干音轨的空间位置)。例如,当视觉图形离用户1670a更近时,可穿戴系统可以使得视觉图形更亮,而当其离开用户1670a时更弱。同时,可穿戴系统可以被配置成在用户1670a四处移动时不更新与视觉图形相关联的干音轨(诸如不改变干音轨的音量)。

[0158] 在其它实现中,随着用户1670a在他的环境中四处移动,音频文件中的音频内容可以被更新。例如,随着用户1670a移动靠近视觉图形1610a的位置,与视觉图形1610a相关联的声音可能变得更加突出。类似地,用户1670可以初始地面对视觉图形1630a和1640a。因此,与视觉图形1630a和1640a相关联的干音轨可以显示为好像它们在用户1670a的前面。然而,在用户1670a转向时,与视觉图形1630a和1640a相关联的干音轨可以被更新好像它们在用户1670a的后面演奏。

#### [0159] 声源的方向性

[0160] 图16A和图16B还示出了声源的方向性的示例。如在鸟瞰视图1600a和1600b中所示,干音轨1610a、1620a和1650b可能仅在几个方向上可听见。例如,干音轨1610a仅在区域1612a中可听见;干音轨1620a仅在区域1622a和1624a中可听见;以及干音轨1650b(如图16B中示出)仅在区域1652b、1654b、1656b和1658b中可听见。

[0161] 当用户在可听区域内时,用户可以感知与可听区域相关联的声音。例如,在图16B中,当用户在区域1652b内时,用户可以感知与视觉图形1650b相关联的声音。然而,如果用户在区域1652b、1654b、1656b和1658b之外,则用户可能不能听到与视觉图形1650b相关联的声音。

[0162] 在一些实施方式中,干音轨的可听区域可以重叠。如图16A中所示,区域1624a和1612a在区域1672a处重叠。因此,当用户1670a在区域1672a内时,可穿戴系统可以呈现与视觉图形1620a和1610a相关联的声音的混合。例如,音频混合可以反映声音的空间位置。在视图1600a中,因为用户面对视觉图形1610a,所以在音频混合中,与视觉图形1610a相关联的声音可能看起来在用户的前面,而与视觉图形1622a相关联的声音可能看起来在用户的左侧。

[0163] 可穿戴系统可单独或与FOV中的视觉图形1620a组合显示视觉图形1610a。例如,尽管当用户在区域1672a中时,用户可以听到与视觉图形1610a和1620a相关联的声音的音频混合,但用户可能仅仅感知视觉图形1610a。在其它实施方式中,由于音频混合包括与两个视觉图形相关联的声音,所以可穿戴系统可呈现视觉图形二者(1610a和1620a)。

[0164] 在一些实施例中,当用户处于干音轨的可听区域内时,可穿戴系统可以被配置为仅显示与干音轨有关的视觉图形。作为示例,当用户位于区域1652b、1654b、1656b和1658b内时,可穿戴系统可以显示视觉图形1650b。随着用户接近视觉图形1650b,可穿戴系统可以将视觉图形1650b显示为更大且更亮。同时,可穿戴系统可以可选地增大音频混合中的视觉图形1650b的音量。然而,当用户在可听区域之外时,即使在一些情况下,用户正在看视觉图形1650b的方向,但可穿戴系统可以被配置为不示出视觉图形1650b。

[0165] 在某些实施方式中,只有用户环境中的干音轨的子集可具有定向声音效果。例如,在图16A中,与视觉图形1630a和1640a相关联的声音可能不具有定向声音效果。因此,尽管用户在区域1672a中,但是可穿戴系统仍然可以包括与视觉图形1630a和1640a相关联的声音,以及与音频混合中的视觉图形1620a和1610a相关联的声音。

[0166] 图16A和图16B中声音区域的形状和大小仅用于说明的目的,并非旨在限制。受定向声音效果影响的区域可能与除波瓣 (lobe) 以外的形状相关联。该区域也可以比图示更大或更小。例如,在一些实施例中,在多个声音之间没有重叠。因此,当用户在该声源的区域时只能感知一个声源。

#### [0167] 声音拼贴效果的示例

[0168] 图17示出了创建声音拼贴效果的示例。图17A示出了具有多个表面1710、1720和1730的虚拟3D对象1700。每个表面可以单独或组合地与音频干、音频文件、声音类型或其它声源相关联。如上所述,每个表面上的音频干可以是定向的。例如,如果用户位于表面1730的前方,则用户可以听到与表面1730相关联的干,但是用户可能无法听到与表面1710相关联的干,因为与表面1710相关联的声音是而不在用户的方向上。

[0169] 可穿戴系统的用户可以例如在方向1750上旋转3D虚拟对象1700 (或围绕其移动),以创建包括与表面1720、表面1730等相关联的声音 (或其一部分) 的音频混合。在一些实施方式中,音频混合可以与3D虚拟对象1700的移动对应。例如,在用户已经停止旋转3D虚拟对象1700的情况下,3D虚拟对象1700的移动可以逐渐减慢。因此,声音拼贴可以包括与某些表面相关的较大部分声音。

[0170] 虽然3D虚拟对象1700使用立方体形状示出,但是其它几何形状也可以用于3D虚拟对象。例如,3D虚拟对象可以是单独或组合的球体、锥体、立方体、多面体、金字塔、平截头体、不规则形状等。干音轨和/或声音文件也不需要与虚拟对象的表面相关联。例如,它们可以与3D虚拟对象的不同体积部分相关联。在一些实施方式中,可能存在与给定表面和/或体积相关联的多个干音轨和/或声音文件。

#### [0171] 呈现干音轨的示例过程

[0172] 图18示出了视觉和听觉上呈现音频文件的示例过程。过程1800可以由在此描述的可穿戴系统执行。

[0173] 在框1810处,可穿戴系统可以访问与音频文件相关联的干音轨。可穿戴系统可以识别音频文件中的干音轨。可穿戴系统还可以访问存储在例如远程数据储存库280 (图2所示) 处的干音轨,其可以不是音频文件的一部分,但可以用来与音频文件的其它干音轨混合。

[0174] 在框1820处,可穿戴系统生成干音轨的可视化。例如,可穿戴系统可以提供与干音轨相关的视觉图形。如参考图12-14所述,视觉图形可以表示乐器或演奏干音轨的人或者是

任何类型的图形形状、设计、颜色等。图形可以呈现2D或3D。视觉图形可以基于音乐的音频内容而改变。例如，视觉图形可以根据音量、音乐节拍等来扩大或缩小。

[0175] 可穿戴系统还可以允许使用姿势和/或用户输入设备与干音轨进行用户交互。在框1830处，可穿戴系统可以检测用户的姿势。该姿势可以包括手势、头部姿势、脚部姿势或其它身体姿势。可穿戴系统可以使用参照图2描述的传感器(诸如IMU和面向外的成像系统)来检测用户的姿势。

[0176] 在框1840处，可穿戴系统可以基于用户的姿势生成干音轨的音频混合。例如，用户可以抓取干音轨的视觉图形并引起两个视觉图形之间的碰撞。可穿戴系统可以相应地混合它们相关的干音轨以产生音频混合。作为另一个示例，用户可以转动他的身体并且感知在他的FOV中的一个或多个视觉图形。可穿戴系统可以混合与他的FOV中的视觉图形相关联的干音轨以生成音频混合。作为又一示例，用户可以调整干音轨，诸如通过使用姿势来调整干音轨的音量或空间位置。可穿戴系统可以基于对干音轨的调整来相应地产生音频混合。

[0177] 在框1850处，可穿戴系统可以基于用户的姿势来更新可视化。例如，可穿戴系统可以基于用户的位置和方位来显示在用户的FOV中的视觉图形的子集。可穿戴系统还可以基于用户的姿势突出或强调(或去强调)一个或多个视觉图形。当两个或多个视觉图形相撞时，可穿戴系统还可以生成并显示新的视觉图形。新的视觉图形可以反映先前的视觉图形之间的碰撞，例如通过将碰撞的视觉图形交错在一起。

[0178] 可穿戴系统还可以允许多个用户共享与视觉图形和干音轨的交互。例如，房间中可能有多个用户。每个用户可以混合干音轨，并生成的音频和视频音频混合可以呈现给房间中的所有用户。在一些实施方式中，每个用户的可穿戴系统可以基于用户的位置和/或方位呈现个性化的视觉图形。例如，面向东方的用户可能会感知与面向西方的用户不同的视觉图形子集。作为另一个示例，用户(例如，DJ)可以与他的环境中的干音轨交互以创建音频混合。用户的可穿戴系统可以将音频混合传送到其它可穿戴系统用于流式传输或回放。可穿戴系统还可以通过其他计算系统广播音频混合。例如，可穿戴系统可以通过房间内的扬声器将音频混合广播给其它人。

[0179] 虽然本文描述的示例是参照干音轨描述的，但是本文描述的技术不限于干音轨。可穿戴系统还可以呈现并允许与其它音频内容的用户交互。例如，可穿戴系统可以允许用户混合两个音频文件，其中一个音频文件具有背景音乐，而另一个音频文件包括人声。

#### [0180] 另外的实施例

[0181] 在第1方面，一种可穿戴显示设备，包括：非瞬态存储器，被配置为测量与可穿戴显示设备的穿戴者的姿势相关联的信息；显示器，被配置为将图像呈现给可穿戴显示设备的穿戴者的眼睛；扬声器，被配置为向可穿戴显示设备的穿戴者呈现声音；以及与非瞬态存储器、传感器、扬声器以及显示器通信的处理器，该处理器编程有可执行指令从而：确定穿戴者的姿势；至少部分基于穿戴者的姿势来生成音频文件的多个干音轨中的至少一个干音轨的音频混合；至少部分地基于穿戴者的姿势和音频混合来生成音频混合的可视化；将表示音频混合的音频信号传送给扬声器；并将表示音频混合的可视化的视觉信号传送给显示器。

[0182] 在第2方面，如方面1所述的可穿戴显示设备，其中多个干音轨包括至少四个干音轨。

[0183] 在第3方面中,如方面1或方面2所述的可穿戴显示设备,其中多个干音轨表示用于不同乐器的多个音轨。

[0184] 在第4方面,如方面3所述的可穿戴显示设备,其中多个干音轨还表示人声音轨。

[0185] 在第5方面,如方面4所述的可穿戴显示设备,其中多个干音轨表示吉他、低音吉他、鼓以及声音。

[0186] 在第6方面,如方面1-5中任一所述的可穿戴显示设备,其中传感器包括眼睛跟踪照相机、陀螺仪、加速度计或惯性测量设备。

[0187] 在第7方面,如方面1-6中任一所述的可穿戴显示设备,其中姿势包括头部姿势、眼睛姿势或头部姿势和眼睛姿势两者。

[0188] 在第8方面,如方面1-7中任一所述的可穿戴显示设备,其中姿势包括穿戴者的姿势。

[0189] 在第9方面,如方面1-8中任一所述的可穿戴显示设备,其中显示器被配置为向穿戴者呈现多个深度平面。

[0190] 在第10方面中,如方面1-9中任一所述的可穿戴显示设备,其中显示器被配置为向穿戴者呈现光场图像。

[0191] 在第11方面中,如方面1-10中任一所述的可穿戴显示设备,其中显示器包括多个堆叠的波导。

[0192] 在第12方面,如方面1-11中任一所述的可穿戴显示设备,其中可视化包括与多个干音轨中的每个干音轨相关联的视觉图形。

[0193] 在第13方面,如方面12所述的可穿戴显示设备,其中视觉图形包括与干音轨的表现相关联的乐器或人的图形表示。

[0194] 在第14方面,如方面12或方面13所述的可穿戴显示设备,其中视觉图形中的至少一些被呈现在相对于穿戴者的不同深度平面处。

[0195] 在第15方面中,如方面12-14中任一所述的可穿戴显示设备,其中可穿戴显示设备被配置为识别穿戴者附近的物理对象,并且将图形表示呈现为与物理对象相关联。

[0196] 在第16方面中,如方面1-15中任一所述的可穿戴显示设备,其中可穿戴显示设备被配置为接收来自电子部件的输入,并且处理器被配置为至少部分地基于输入来生成音频混合或可视化。

[0197] 在第17方面,如方面1-16中任一所述的可穿戴显示设备,其中多个干音轨中的至少一个干音轨包括与该干音轨的空间方向性相关联的信息。

[0198] 在第18方面,如方面1-17中任一所述的可穿戴显示设备,其中多个干音轨中的至少一个干音轨与相对于穿戴者的空间位置相关联。

[0199] 在第19方面,如方面18所述的可穿戴显示设备,其中处理器被配置为至少部分地基于穿戴者的姿势或者来自电子部件的输入,来锁定与多个干音轨中的至少一个干音轨相关联的空间位置。

[0200] 在第20方面,如方面1-19中任一所述的可穿戴显示设备,其中,音频文件以音频格式编码,该音频格式包括指示以下中的至少一个的元数据:多个干音轨,在给定时间与多个干音轨相关联的位置数据,或关于音频混合的可视化的信息。在另一方面,音频文件以音频格式编码,该音频格式包括指示以下中的至少一个的元数据:多个干音轨,在给定时间与多

个干音轨相关联的位置数据,或关于音频混合的可视化的信息。音频文件可以包括与多个干音轨相关联的信息和/或标识符。音频文件可以包括多个干音轨的音频混合。

[0201] 在第21方面,如方面1-20中任一所述的可穿戴显示设备,其中音频混合的可视化包括三维虚拟对象,该三维虚拟对象包括多个表面,其中每个表面与多个干音轨中的干音轨相关联。音频混合可以包括多个干音轨的声音拼贴效果。

[0202] 在第22方面,一种用于与增强现实对象交互的方法,该方法包括:在硬件计算机处理器的控制下:生成用于由可穿戴显示系统的穿戴者进行交互的增强现实对象,其中增强现实对象包括用于修改与增强现实对象相关联的特征并显示在可穿戴显示系统的显示器上的界面;在穿戴者与界面交互时检测穿戴者的手势;将检测到的手势与对增强现实对象的特征的修改相关联;以及根据对增强现实对象的特征的修改来修改增强现实对象。

[0203] 在第23方面,如方面22所述的方法,其中在穿戴者与界面交互的同时检测穿戴者的手势包括:利用可穿戴显示系统的面向外的图像捕获设备来捕获穿戴者的手的第一图像;利用可穿戴显示系统的面向外的图像捕获设备来捕获穿戴者的手的第二图像;以及比较第一图像和第二图像以检测穿戴者的手势。

[0204] 在第24方面,如方面22-23中任一所述的方法,其中增强现实对象表示电话,并且其中界面对应于数字键盘。

[0205] 在第25方面中,如方面22-24中任一所述的方法,其中将检测到的手势与对增强现实对象的特征的修改相关联包括:将电话上的手指按压关联以修改数字键盘的外观。

[0206] 在第26方面,如方面22-25中任一所述的方法,其中根据对增强现实对象的特征的修改来修改增强现实对象包括:在与相关联的手指按压对应的键盘上显示表示顺序按压的数字键盘。

[0207] 在第27方面,如方面22-26中任一所述的方法,其中增强现实对象的界面对应于增强现实对象的仿实物表示。

[0208] 在第28方面,如方面22-27中任一所述的方法,其中增强现实对象对应于乐器或演奏乐器的人。

[0209] 在第29方面中,如方面22-28中任一所述的方法,其中将检测到的手势与对乐器或演奏乐器的人的修改相关联包括:检测与乐器交互的穿戴者的手或手指移动以增加与乐器相关联的干音轨的属性。

[0210] 在第30方面,如方面22-29中任一所述的方法,其中增强现实对象与包括多个干音轨中的至少一个干音轨的音频文件相关联。

[0211] 在第31方面,如方面30所述的方法,其中增强现实对象是包括多个表面的三维虚拟对象,其中每个表面与多个干音轨中的干音轨相关联。

[0212] 在第32方面,如方面31所述的方法,还包括响应于穿戴者的姿势,使用多个干音轨创建声音拼贴效果。

[0213] 在第33方面,如方面30-32中任一所述的方法,其中与乐器相关联的干音轨是由耦合到可穿戴显示系统的音频系统播放的多个干音轨中的一个干音轨。

[0214] 在第34方面,如方面30-33中任一所述的方法,其中多个干音轨中的至少一个是由用户生成的干音轨替换的原始干音轨。

[0215] 在第35方面,如方面30至34中任一所述的方法,其中使用穿戴者的头部姿势来选

择另外的增强现实对象,用于改变由该另外的增强现实对象表示的对应干音轨。

[0216] 在第36方面,如方面30-35中任一所述的方法,其中多个干音轨中的至少一个干音轨包括与该干音轨的方向性相关联的信息。

[0217] 在第37方面,如方面30-36中任一所述的方法,其中多个干音轨中的至少一个干音轨与相对于穿戴者的空间位置相关联。

[0218] 在第38方面中,如方面37所述的方法,还包括:至少部分地基于穿戴者的姿势来锁定与多个干音轨中的至少一个干音轨相关联的空间位置。

[0219] 在第39方面,如方面30-38中任一所述的方法,其中音频文件以音频格式编码,其中音频格式包括指示以下中的至少一个的元数据:多个干音轨、在特定时间与多个干音轨相关联的位置数据、或关于多个干音轨的可视化的信息。

[0220] 在第40方面,一种用于与视觉图形交互的可穿戴显示系统,该系统包括:显示器,其被配置为显示视觉图形;至少一个面向外的照相机,其被配置为对可穿戴显示系统前方的区域成像;以及计算机处理器,其被配置为与网络通信并且被编程为:生成视觉图形以供可穿戴显示系统的穿戴者进行交互,其中视觉图形包括用于修改与视觉图形相关联的特征的界面;检测与可穿戴显示系统交互的穿戴者的手势;将检测到的手势与对视觉图形的特征的修改相关联;并根据对视觉图形的特征的修改来改变视觉图形。

[0221] 在第41方面,如方面40所述的可穿戴显示系统,其中视觉图形与至少一个干音轨相关联。

[0222] 在第42方面,如方面41的可穿戴显示系统,其中干音轨是定向的。

[0223] 在第43方面,如方面41-42中任一所述的可穿戴显示系统,其中干音轨与相对于穿戴者的空间位置相关联。

[0224] 在第44方面,如方面41-43中任一所述的可穿戴显示系统,其中计算机处理器还被配置为至少部分地基于穿戴者的姿势或来自电子部件的输入来锁定干音轨的空间位置。

[0225] 在第45方面,如方面41-44中任一所述的可穿戴显示系统,其中视觉图形包括多个表面,其中每个表面与多个干音轨的干音轨相关联,并且其中计算机处理器还被配置为响应于检测到的穿戴者的手势来创建多个干音轨的声音拼贴效果。

[0226] 在第46方面中,如方面1-21中任一所述的可穿戴显示设备,方面22-39中任一项所述的方法,方面40-45中任一所述的可穿戴显示系统,其中可穿戴显示设备和/或可穿戴显示系统包括头戴式显示器。在这些方面的任何一个中,可穿戴显示设备和/或可穿戴显示系统可以包括光场显示器。

[0227] 在第47方面,可穿戴设备包括:非瞬态存储器,被配置为存储包括多个干音轨的音频文件,每个干音轨表示音频文件的不同音频内容;传感器,其被配置为测量与可穿戴设备的用户的姿势相关联的信息;显示器,其被配置为将图像呈现给可穿戴设备的用户的眼睛;扬声器,其被配置为向可穿戴设备的用户呈现声音;以及与非瞬态存储器、传感器、扬声器和显示器通信的处理器,该处理器编程有可执行指令从而:确定用户的姿势;至少部分地基于用户的姿势生成音频文件的多个干音轨中的至少一个干音轨的音频混合;至少部分地基于用户的姿势和音频混合来生成音频混合的可视化;将表示音频混合的音频信号传送给扬声器;并将表示音频混合的可视化的视觉信号传送给显示器。

[0228] 在第48方面,如方面47所述的可穿戴设备,其中多个干音轨的干音轨与不同的乐

器或人声音轨相关联。

[0229] 在第49方面,如方面47-48中任一所述的可穿戴设备,其中多个干音轨中的至少一个干音轨包括与干音轨的空间位置和空间方向性相关联的信息。

[0230] 在第50方面,如方面47-49中任一所述的可穿戴设备,其中传感器包括以下中的至少一个:陀螺仪、加速度计、惯性测量单元、眼睛跟踪照相机或被配置为将用户的环境进行成像的照相机。

[0231] 在第51方面,如方面47-50中任一所述的可穿戴设备,其中姿势包括以下中的至少一个:头部姿势、眼睛姿势、身体姿势或手势。

[0232] 在第52方面,如方面47-51中任一所述的可穿戴设备,其中可视化包括与多个干音轨的相应干音轨相关联的视觉图形。

[0233] 在第53方面,如方面52所述的可穿戴设备,其中视觉图形包括与干音轨的表现相关联的乐器或人的图形表示。

[0234] 在第54方面,如方面52所述的可穿戴设备,其中视觉图形中的至少一些被呈现在相对于用户的不同深度平面处。

[0235] 在第55方面,如方面47-54中任一所述的可穿戴设备,其中可穿戴设备被配置为识别用户的环境中的物理对象,并且将多个干音轨中的干音轨的图形表示呈现为与物理对象相关联。

[0236] 在第56方面中,如方面47-55中任一所述的可穿戴设备,其中可穿戴设备被配置为接收来自用户输入设备的输入,并且处理器被配置为至少部分地基于输入来生成音频混合或可视化。

[0237] 在第57方面,一种用于与音频文件交互的方法,该方法包括:在可穿戴设备的控制下,该可穿戴设备包括硬件计算机处理器,配置为测量与可穿戴设备的用户的姿势相关联的信息的传感器,配置为将图像呈现给可穿戴设备的用户的眼睛的显示器,配置为向可穿戴设备的用户呈现声音的扬声器;获取与音频文件相关联的多个干音轨;生成多个干音轨的可视化;检测用户的姿势;至少部分地基于用户的姿势来生成音频文件的多个干音轨中的至少一个干音轨的音频混合;至少基于用户的姿势或音频混合中的至少一个来更新多个干音轨的可视化;将表示音频混合的音频信号传送给扬声器;以及将表示音频混合的可视化的视觉信号传送给显示器。

[0238] 在第58方面,如方面57所述的方法,其中多个干音轨的干音轨与不同的乐器或人声轨道相关联。

[0239] 在第59方面,如方面57-58中任一所述的方法,其中多个干音轨中的至少一个干音轨包括与干音轨的空间位置和空间方向性相关联的信息。

[0240] 在第60方面,如方面57-59中任一所述的方法,其中可视化包括对应于多个干音轨中的相应干音轨的视觉图形。

[0241] 在第61方面,如方面60所述的方法,其中视觉图形包括与相应干音轨相关联的乐器或与相应干音轨相关联的演奏乐器的人。

[0242] 在第62方面,如方面60所述的方法,其中至少一些视觉图形被呈现在相对于用户的不同深度平面处。

[0243] 在第63方面,如方面57-62中任一所述的方法,其中姿势包括以下中的至少一个:

头部姿势、眼睛姿势、身体姿势或手势。

[0244] 在第64方面,如方面57-63中任一所述的方法,其中更新可视化包括强调或去强调与多个干音轨的干音轨相关联的视觉图形。

[0245] 在第65方面,如方面57-64中任一所述的方法,还包括:接收对多个干音轨的干音轨的选择;接收修改干音轨的指示,其中指示包括用户的姿势的改变或来自用户输入设备的输入;以及至少部分基于指示来修改干音轨。

[0246] 在第66方面,如方面65所述的方法,其中修改干音轨包括以下中的至少一个:调整音量,用另一干音轨替代干音轨,静音干音轨,强调干音轨超过多个干音轨中的其它干音轨。

[0247] 结论

[0248] 在此描述的和/或在附图中描绘的过程、方法和算法中的每一个可以体现在由一个或多个物理计算系统、硬件计算机处理器、专用电路和/或被配置为执行具体和特定计算机指令的电子硬件所执行的代码模块中,并且由以上完全或部分自动化。例如,计算系统可以包括编程有特定计算机指令的通用计算机(例如,服务器)或专用计算机、专用电路等等。代码模块可以被编译并链接到可执行程序中,安装在动态链接库中,或者可以用解释的编程语言写入。在一些实施方式中,特定操作和方法可以由给定功能特定的电路来执行。

[0249] 此外,本公开的功能的某些实施方式在数学上、计算上或技术上是足够复杂的,以致于可能需要专用硬件或一个或多个物理计算设备(利用适当的专用可执行指令)来执行功能,例如由于所涉及的计算的数量或复杂性或为了基本实时地提供结果。例如,视频可以包括许多帧,每帧具有数百万个像素,并且特别编程计算机硬件需要处理视频数据,从而在商业上合理的时间量内提供期望的图像处理任务或应用。

[0250] 代码模块或任何类型的数据可以存储在任何类型的非瞬态计算机可读介质上,诸如物理计算机存储器,其包括硬盘驱动器、固态存储器、随机存取存储器(RAM)、只读存储器(ROM)、光盘、易失性或非易失性存储器、相同的组合和/或类似物。方法和模块(或数据)也可以在各种计算机可读传输介质上作为生成的数据信号(例如,作为载波或其它模拟或数字传播信号的一部分)传输,所述传输介质包括基于无线的和基于有线/线缆的介质,并且可以采取多种形式(例如,作为单个或多路复用模拟信号的一部分,或者作为多个离散数字分组或帧)。所公开的过程或过程步骤的结果可以持久地或以其它方式存储在任何类型的非瞬态、有形计算机存储器中,或者可以经由计算机可读传输介质来传送。

[0251] 在此所描述和/或附图中描绘的流程图中的任何过程、框、状态、步骤或功能应当被理解为潜在地表示代码模块、代码段或代码部分,这些代码模块、代码段或代码部分包括用于实现特定功能(例如,逻辑或算术)或步骤的一个或多个可执行指令。过程、框、状态、步骤或功能可以与在此提供的说明性示例相组合、重新排列、添加、删除、修改或以其它方式改变。在一些实施例中,附加的或不同的计算系统或代码模块可以执行在此描述的功能中的一些或全部。在此描述的方法和过程也不限于任何特定的顺序,并且与其相关的块、步骤或状态可以以适当的其它顺序来执行,例如串行、并行或以某种其它方式。任务或事件可以添加到所公开的示例实施例或者从中去除。此外,在此描述的实施方式中的各种系统部件的分离是出于说明的目的,并且不应该被理解为在所有实施方式中都需要这种分离。应该理解,所描述的程序组件、方法和系统通常可以一起集成在单个计算机产品中或者封装到

多个计算机产品中。许多实施方式变化是可能的。

[0252] 过程、方法和系统可以实现在网络(或分布式)计算环境中。网络环境包括企业范围的计算机网络、内联网、局域网(LAN)、广域网(WAN)、个人局域网(PAN)、云计算网络、众包(crowd-sourced)计算网络、互联网和万维网。网络可以是有线或无线网络或任何其它类型的通信网络。

[0253] 本公开的系统和方法各自具有若干创新性方面,其中没有单独一个对在此公开的期望属性完全负责或需要。上述各种特征和过程可以彼此独立地使用,或者可以以各种方式组合。所有可能的组合和子组合均旨在落入本公开的范围。对于本公开中所描述的实施方式的各种修改对于本领域技术人员来说可以是容易清楚的,并且在不脱离本公开的精神或范围的情况下,可将在此定义的一般原理应用于其它实施方式。因此,权利要求不旨在限于在此示出的实施方式,而是应被给予与本公开、在此公开的原理和新颖特征一致的最宽范围。

[0254] 本说明书中在分开的实现方式中描述的某些特征也可以在单个实现方式中组合地实现。相反地,在单个实施方式的上下文中描述的各种特征也可以分别在多个实施方式中分开地或以任何合适的子组合方式来实现。此外,尽管上文可以将特征描述为以某些组合起作用并且甚至最初如此要求,但是来自所要求的组合的一个或多个特征可以在一些情况下从组合中删除,并且所要求的组合可以针对子组合或子组合的变体。没有单个特征或特征组是每个实施例必需或不可缺少。

[0255] 除非另有特别说明,或者在所使用的上下文中进行理解,在此使用的条件语言,诸如“能(can)”、“能够(could)”、“可能(might)”、“可以(may)”、“例如(e.g.)”等等,一般意在表达某些实施例包括而其它实施例不包括某些特征、要素和/或步骤。因此,这样的条件语言通常不旨在暗示特征、要素和/或步骤以任何方式对于一个或多个实施例是必需的,或者一个或多个实施例必然包括用于在有或者没有作者输入或提示的情况下决定这些特征、要素和/或步骤是否包括在或执行在任何特定实施例中。术语“包括(comprising)”、“包括(including)”、“具有(having)”等是同义词,并且以开放式的方式包含性地使用,并且不排除附加的要素、特征、动作、操作等等。此外,术语“或”以其包含性含义(而不是其排他含义)使用,因此当用于例如连接要素列表时,术语“或”表示列表中的一个、一些或全部要素。另外,除非另有说明,否则本申请和所附权利要求书中使用的冠词“一”、“一个”和“所述”应被解释为表示“一个或多个”或“至少一个”。

[0256] 如在此所使用的,提及项目列表中的“至少一个”的短语是指这些项目的任何组合,包括单个成员。作为例子,“A,B或C中的至少一个”旨在涵盖:A、B、C、A和B、A和C、B和C,以及A、B和C。除非另有特别说明,否则诸如短语“X、Y和Z中的至少一个”的连接语言如所使用的在利用上下文进行理解,通常用于表达项目、术语等可以是X、Y或Z中的至少一个。因此,这样的连接语言通常不旨在暗示某些实施例需要X中的至少一个、Y中的至少一个和Z中的至少一个都存在。

[0257] 类似地,虽然可以在附图中以特定顺序示出操作,但是应该认为,不需要以所示出的特定顺序或按顺次顺序执行这样的操作,或者不需要执行所有示出的操作以实现期望的结果。此外,附图可以以流程图的形式示意性地描绘一个或多个示例过程。然而,未示出的其它操作可以并入示意性说明的示例性方法和过程中。例如,一个或多个附加操作可以在

任何所示操作之前、之后、同时或之间执行。另外,在其它实施方式中,操作可以重新安排或重新排序。在特定情况下,多任务和并行处理可能是有利的。而且,上述实施方式中的各种系统组件的分离不应当被理解为在所有实施方式中都需要这种分离,并且应该理解的是,所描述的程序组件和系统通常可以一起集成在单个软件产品中或者封装到多个软件产品。另外,其它的实施方式在以下权利要求的范围中。在一些情况下,权利要求中列举的动作能够以不同的顺序执行并且仍然实现期望的结果。

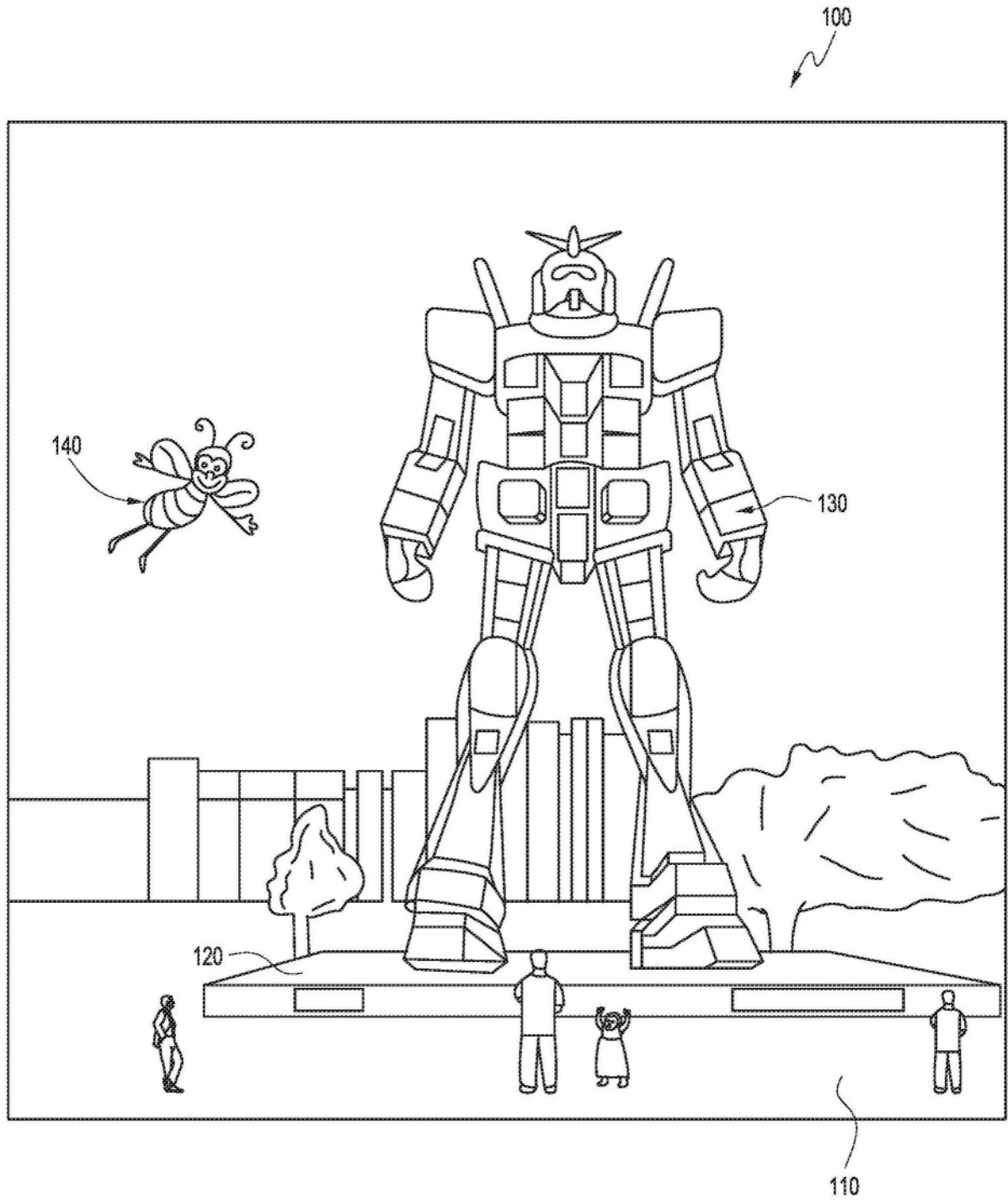


图1

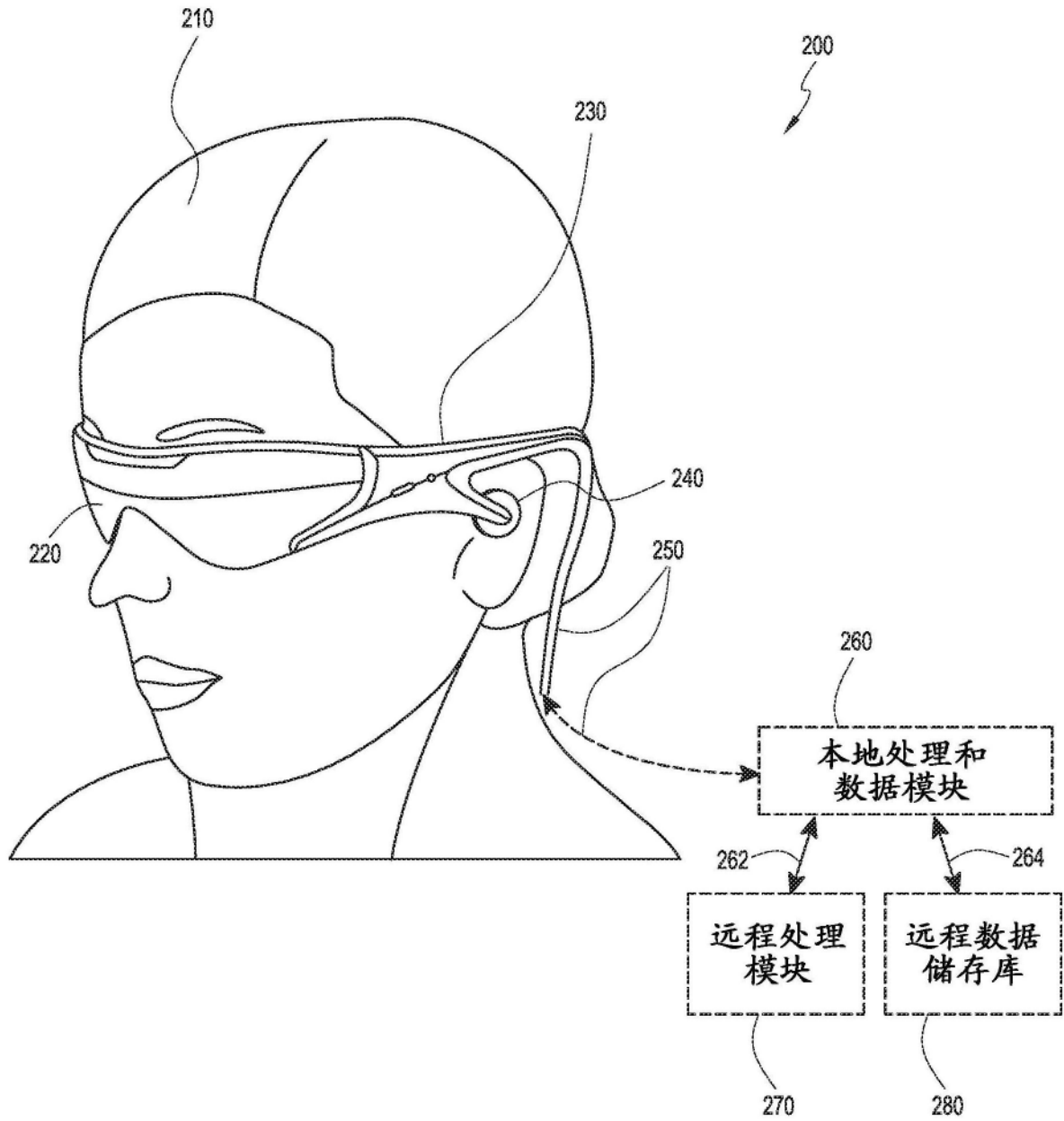


图2

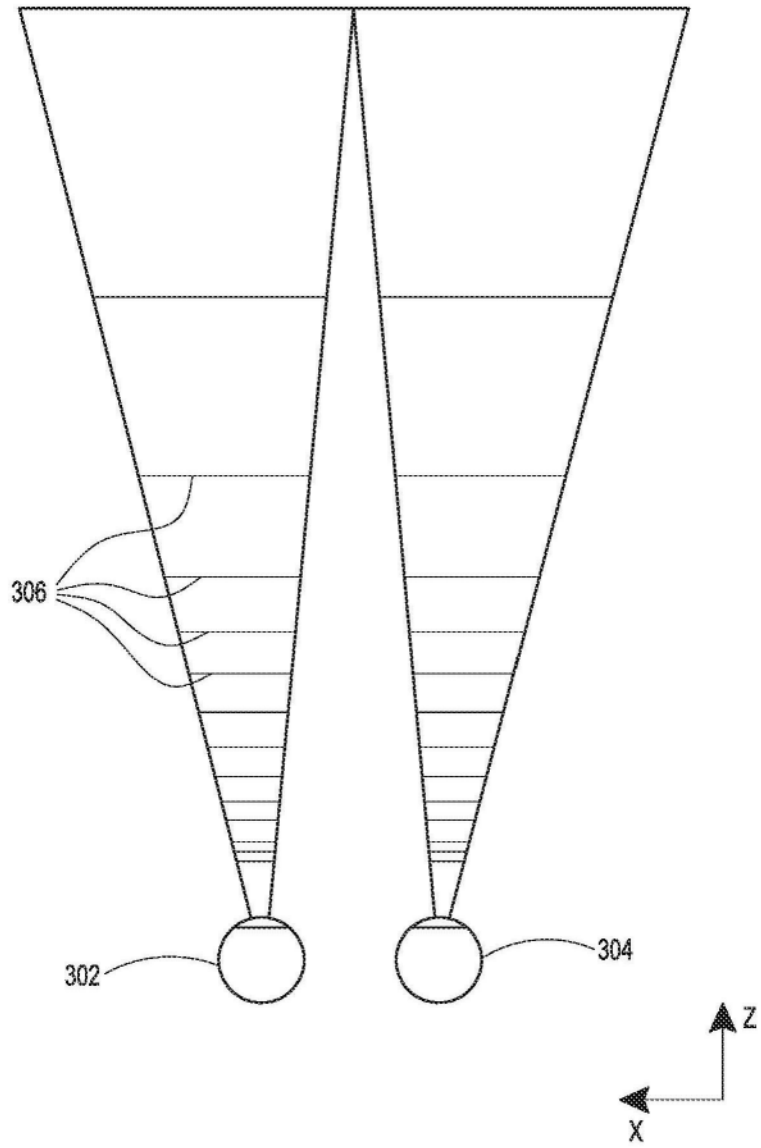


图3

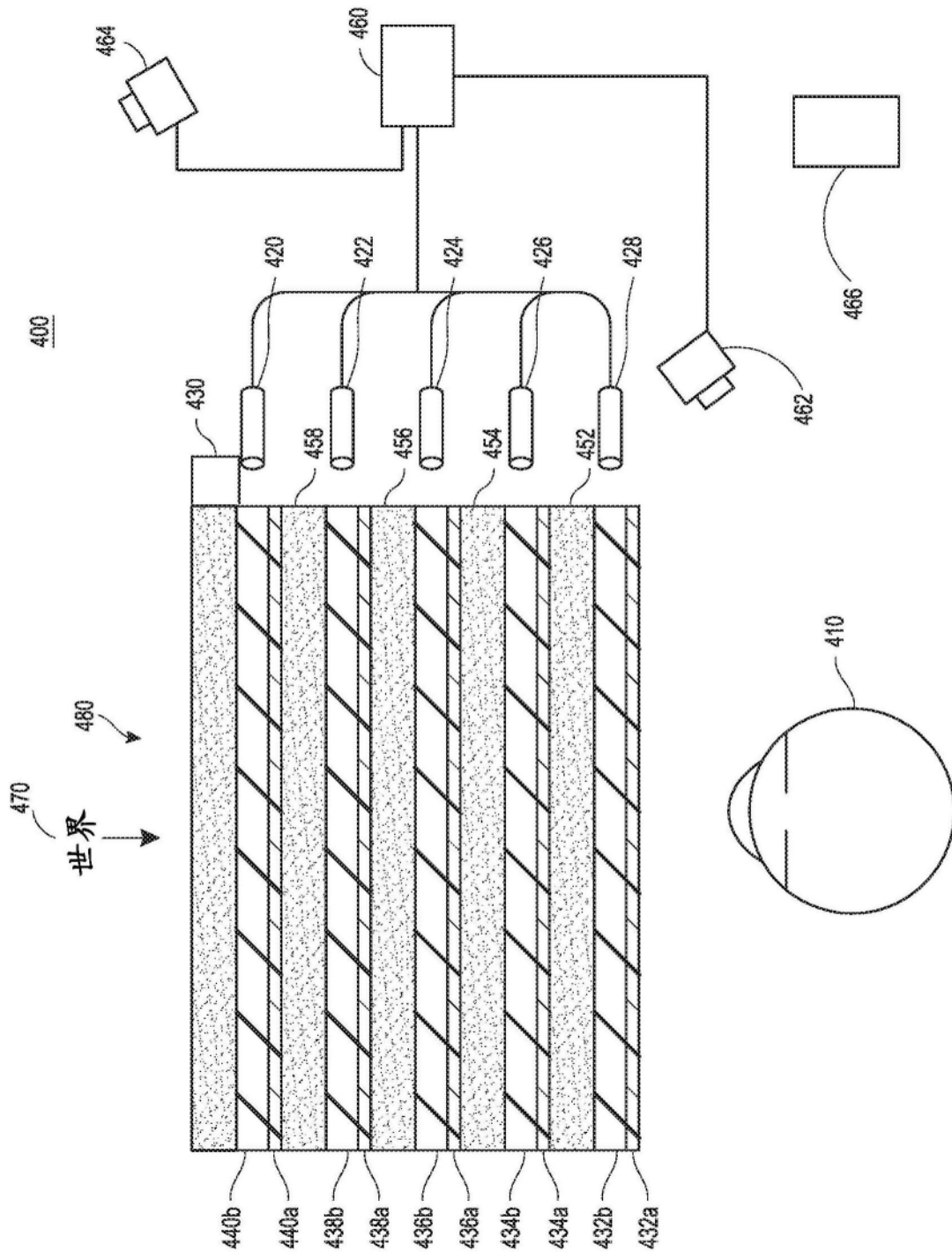


图4

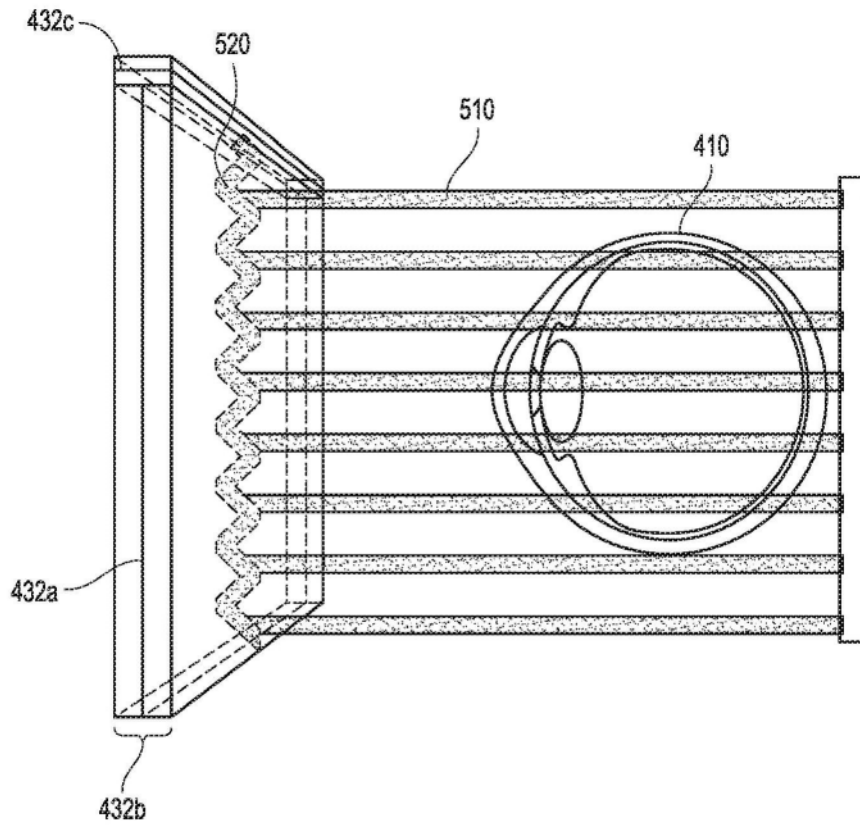


图5

600

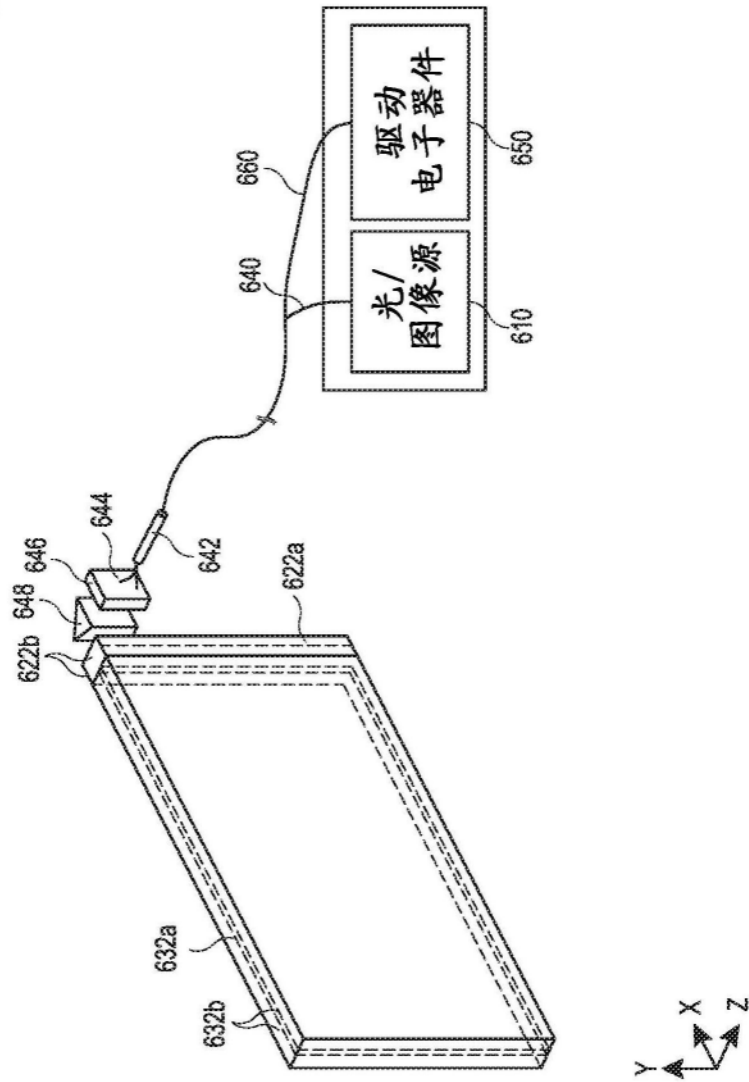


图6

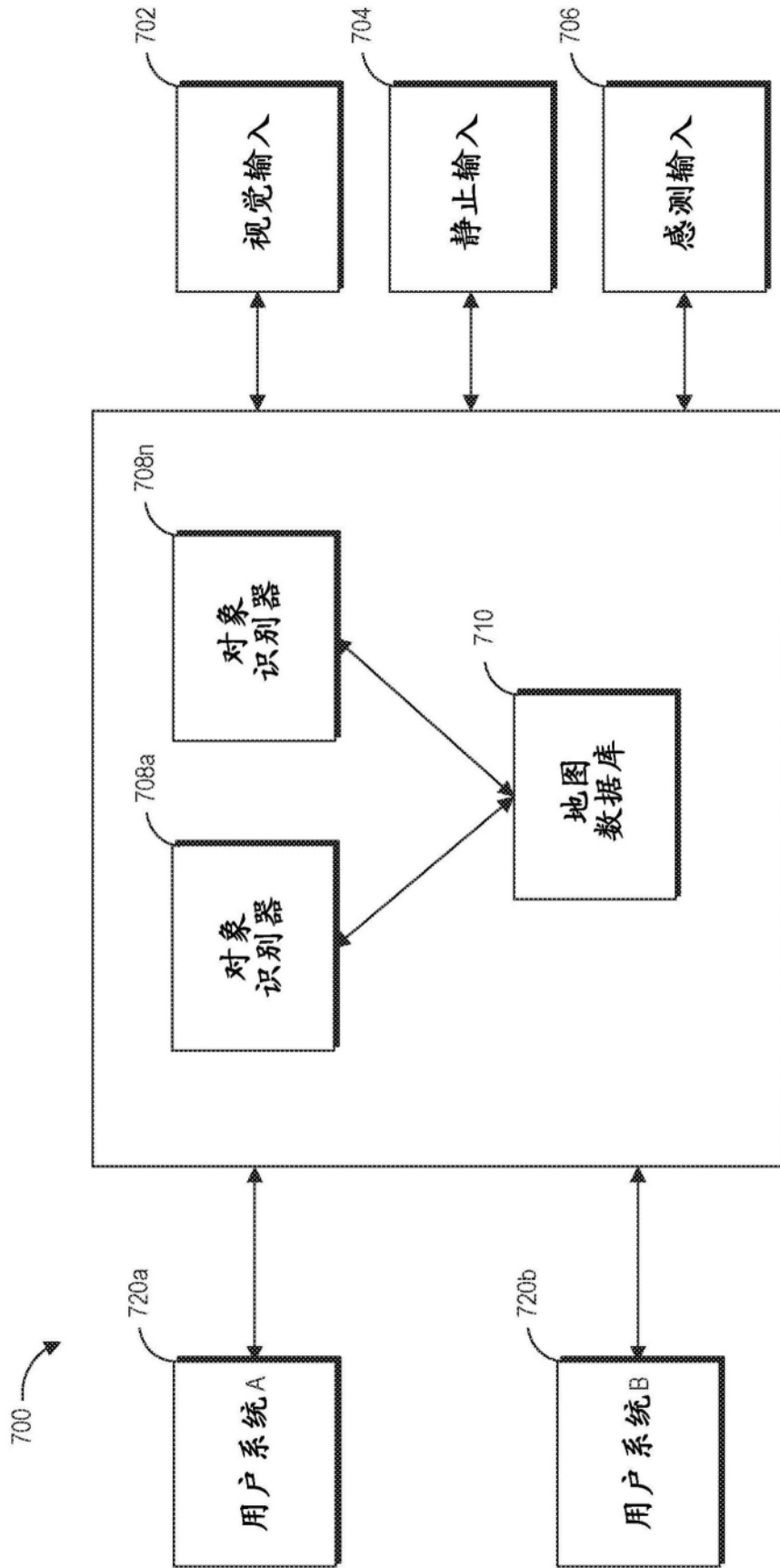


图7

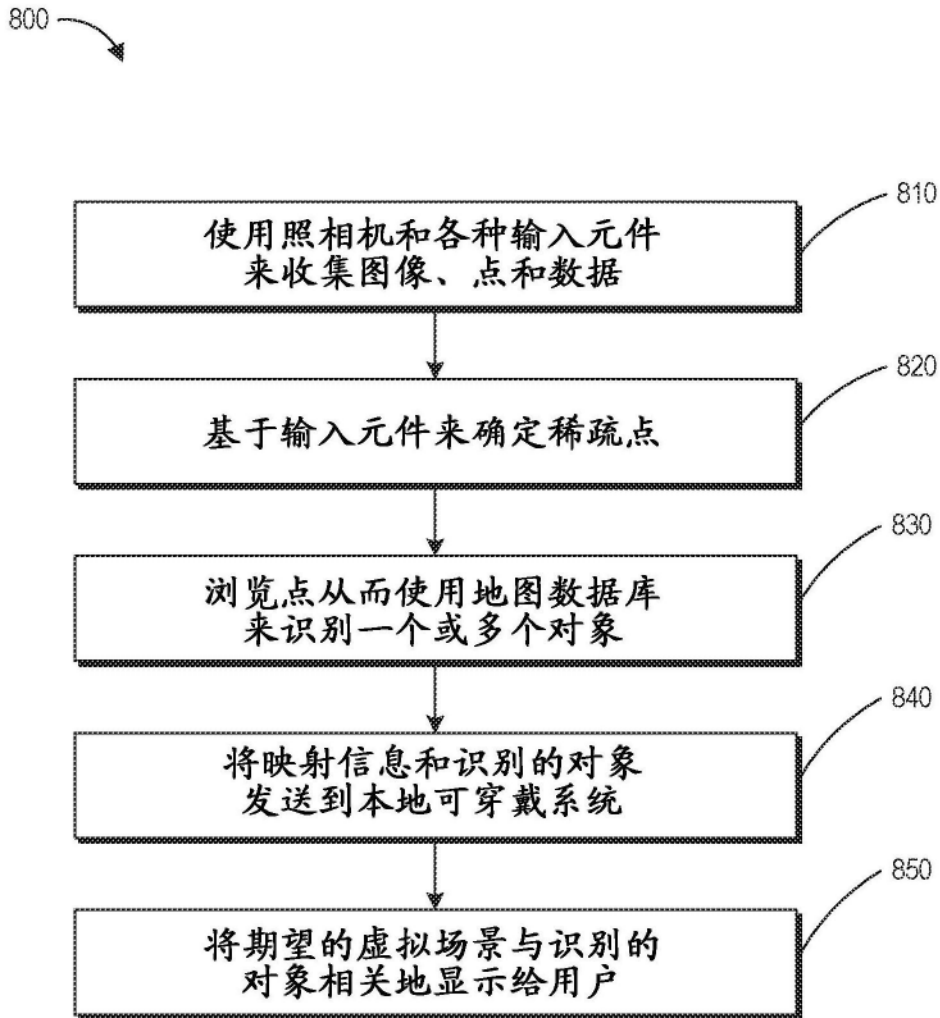


图8

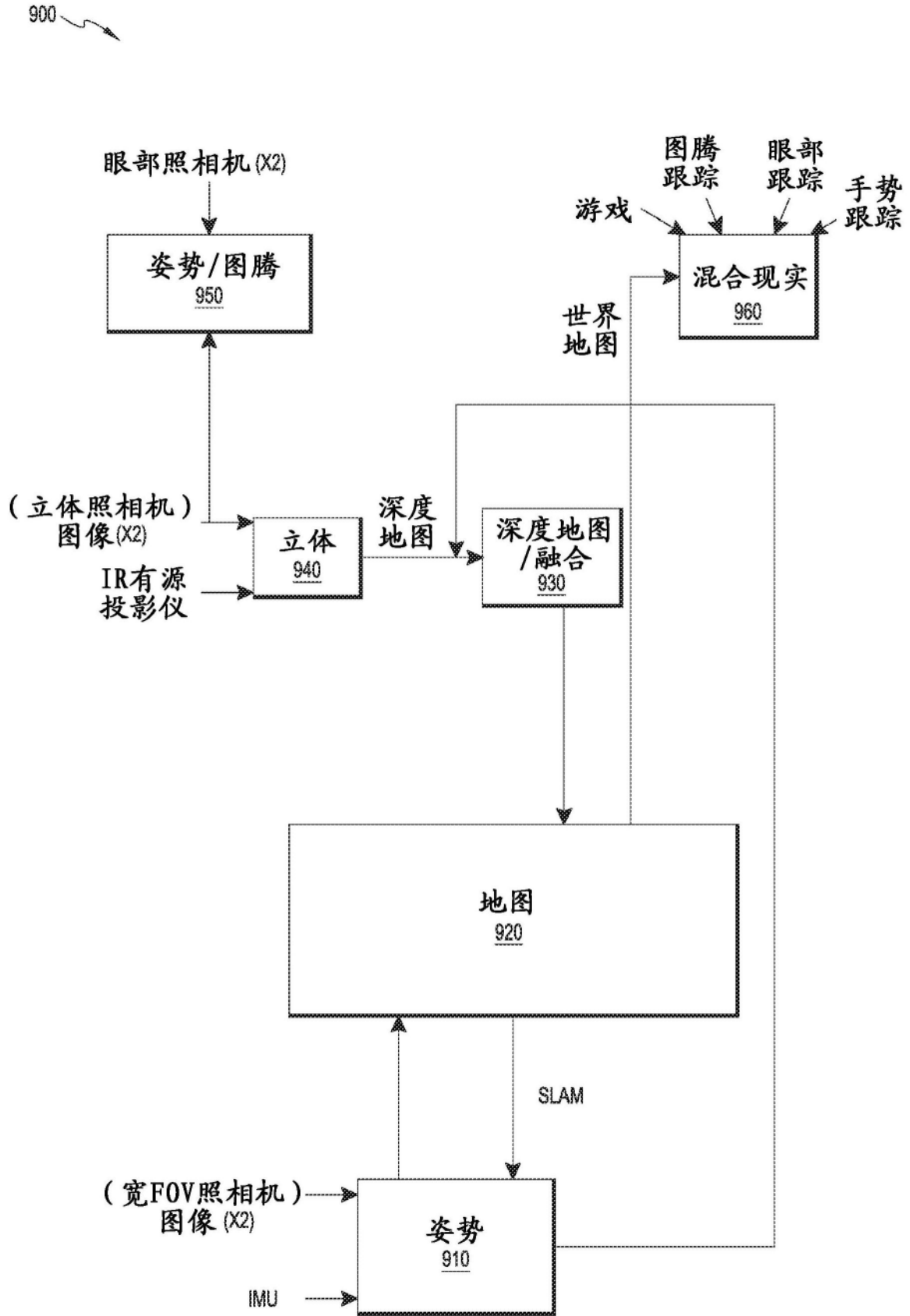


图9

1000

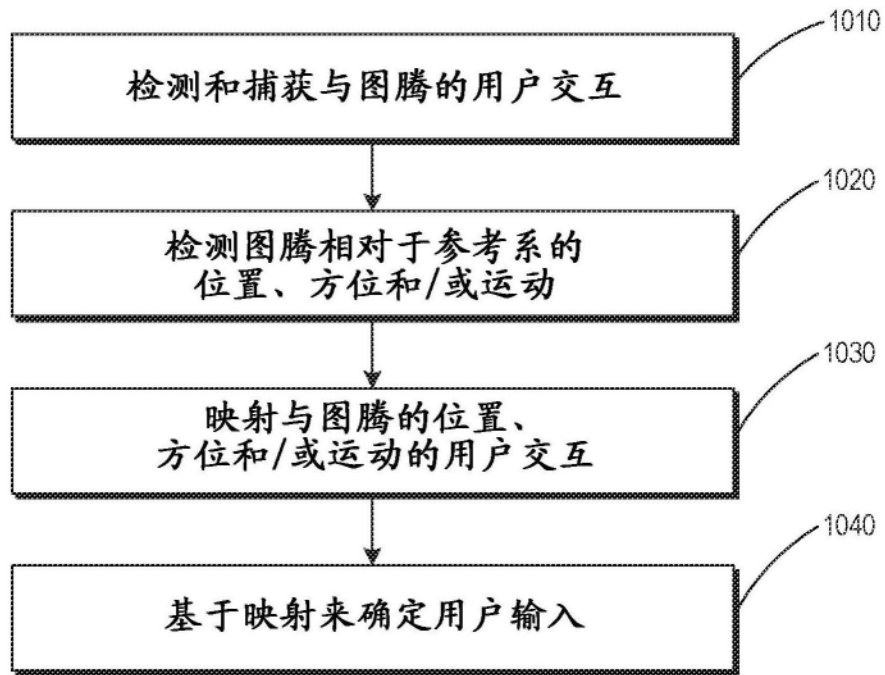


图10

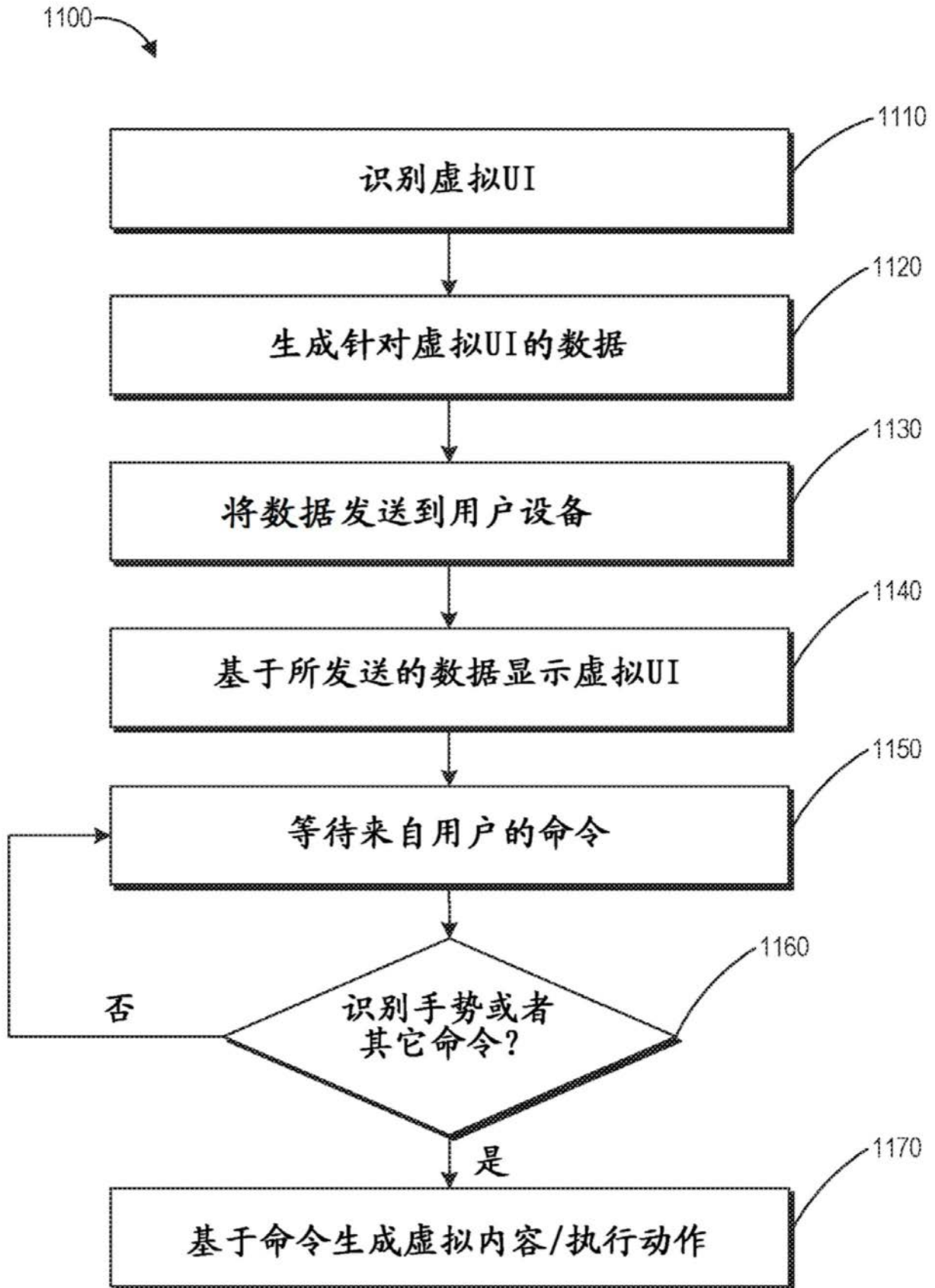


图11

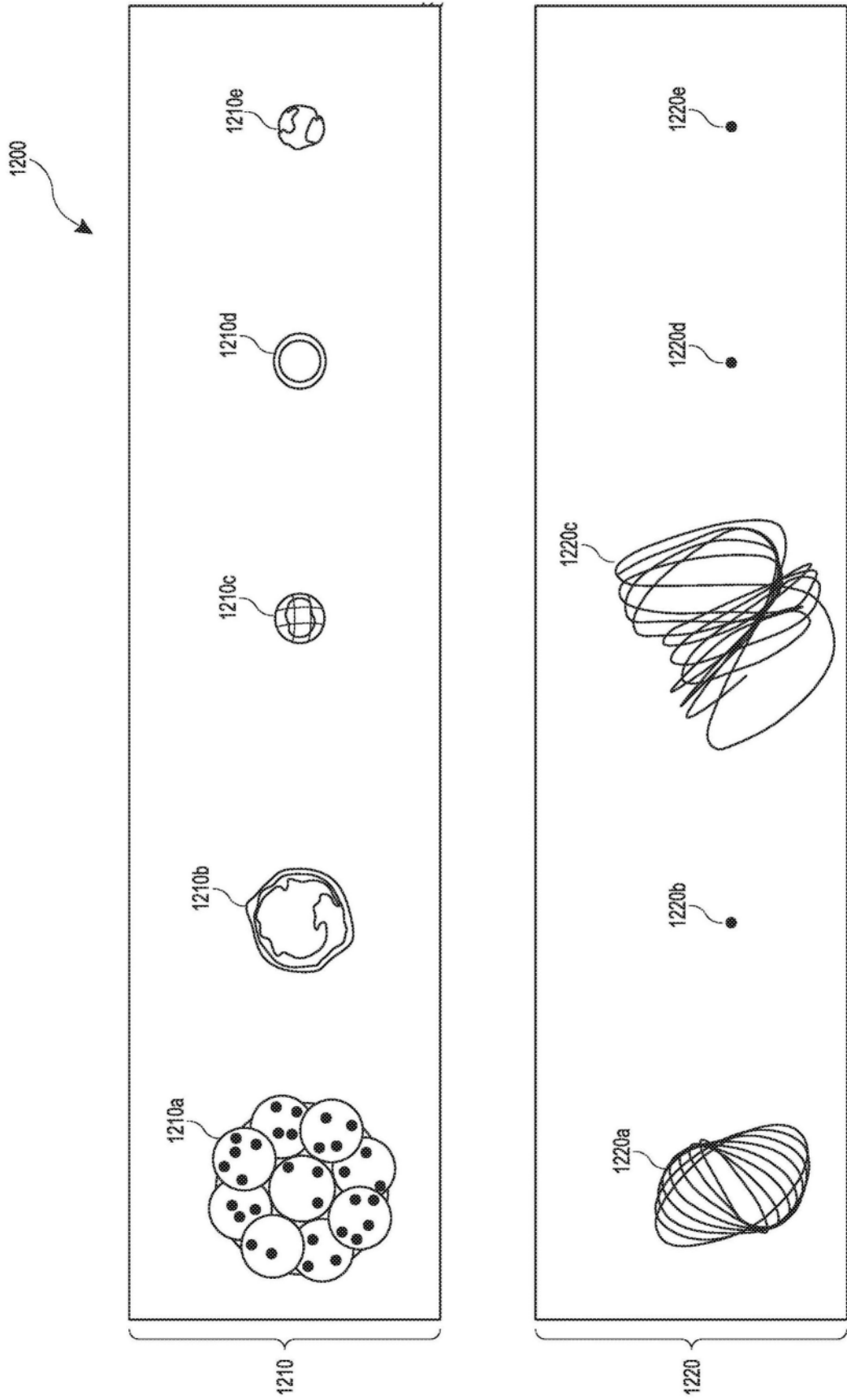


图12

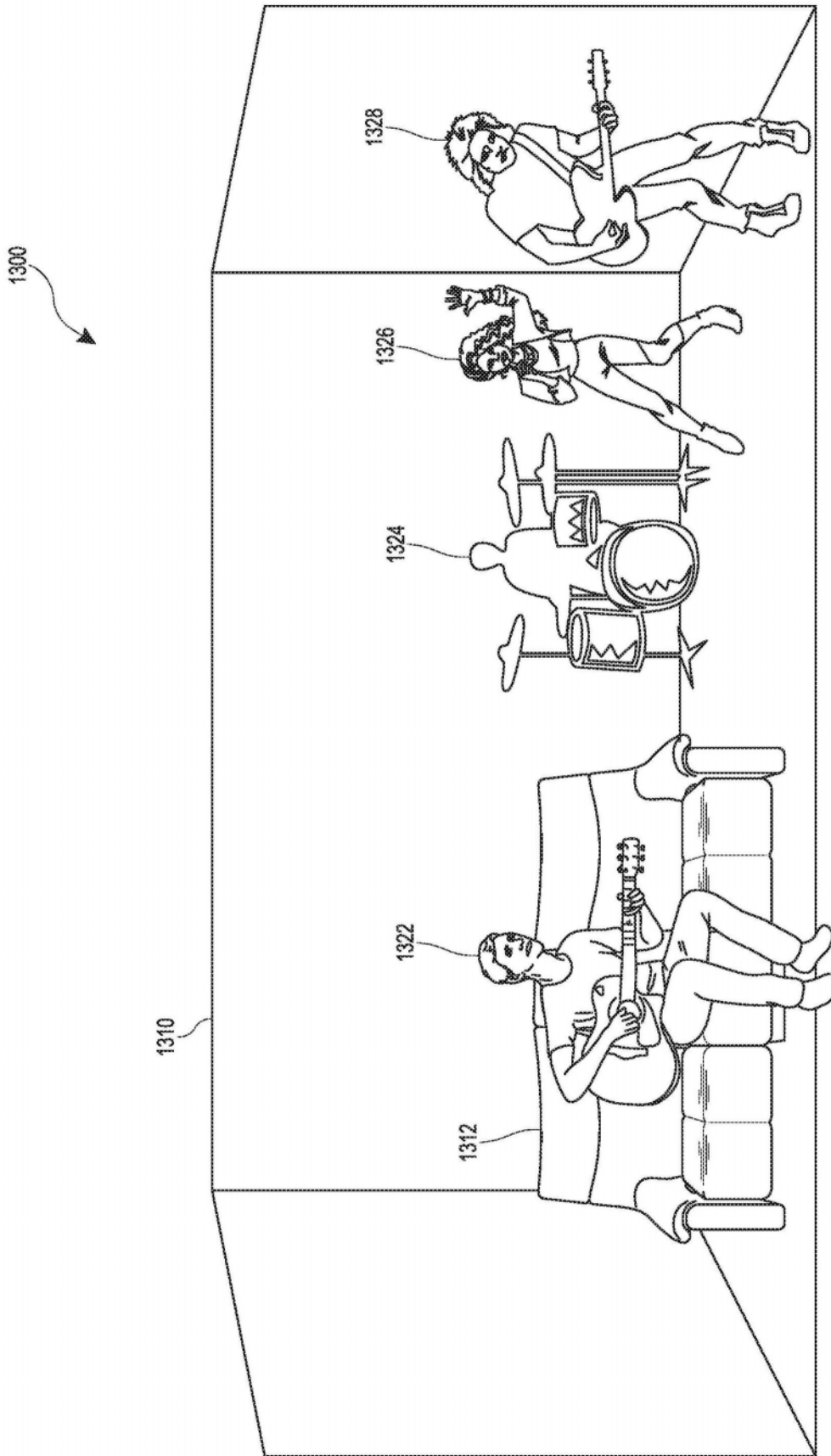


图13

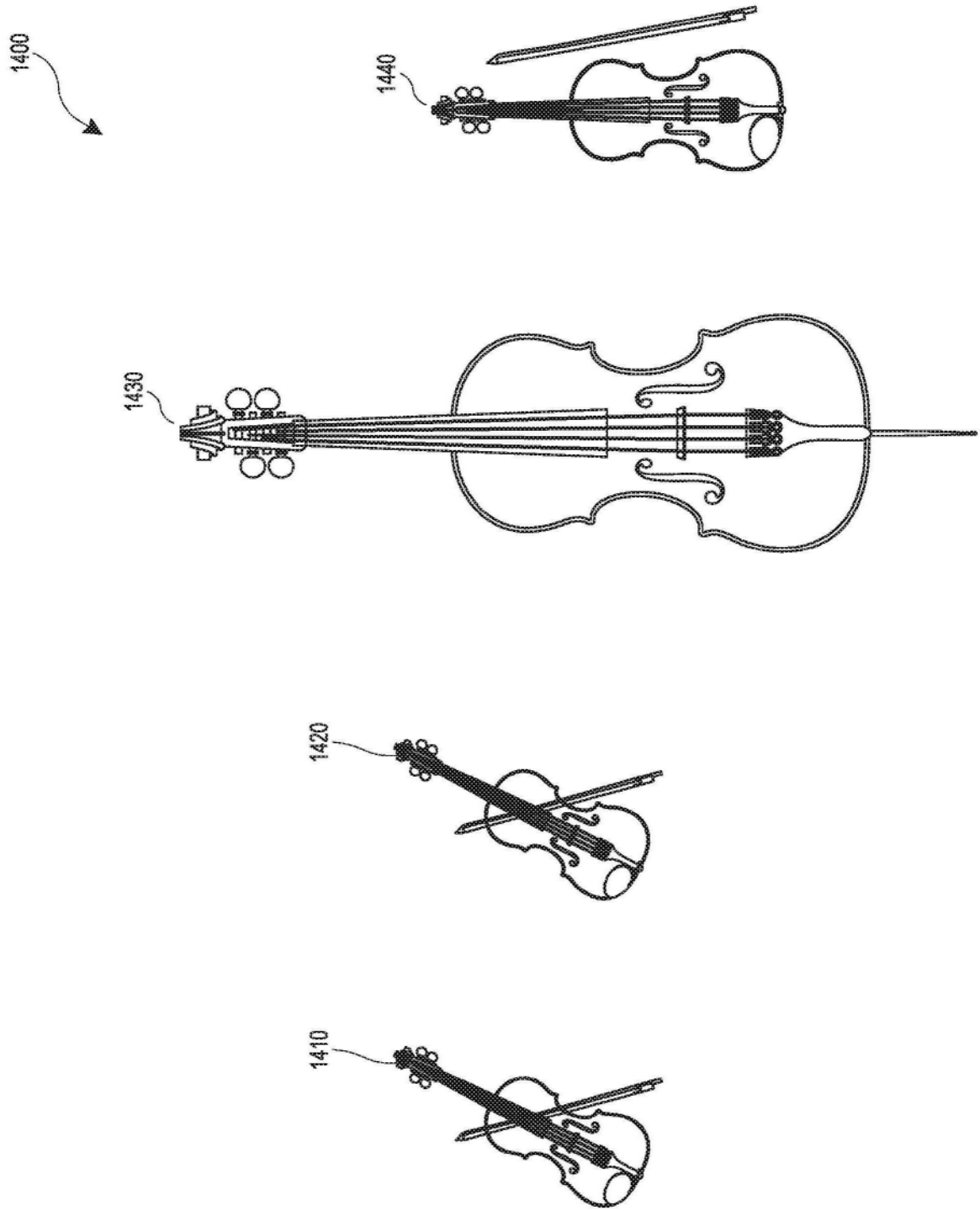


图14

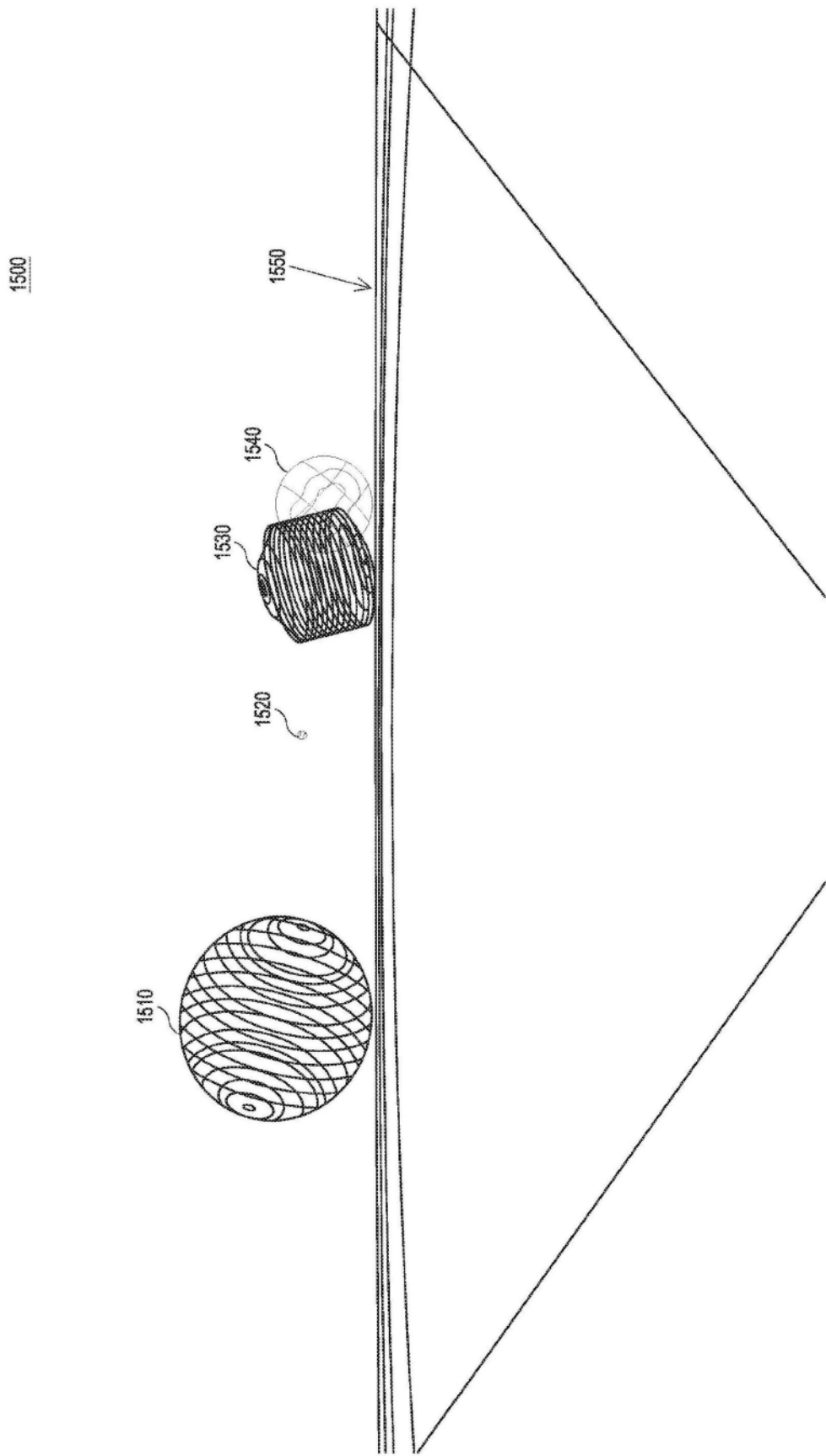


图15

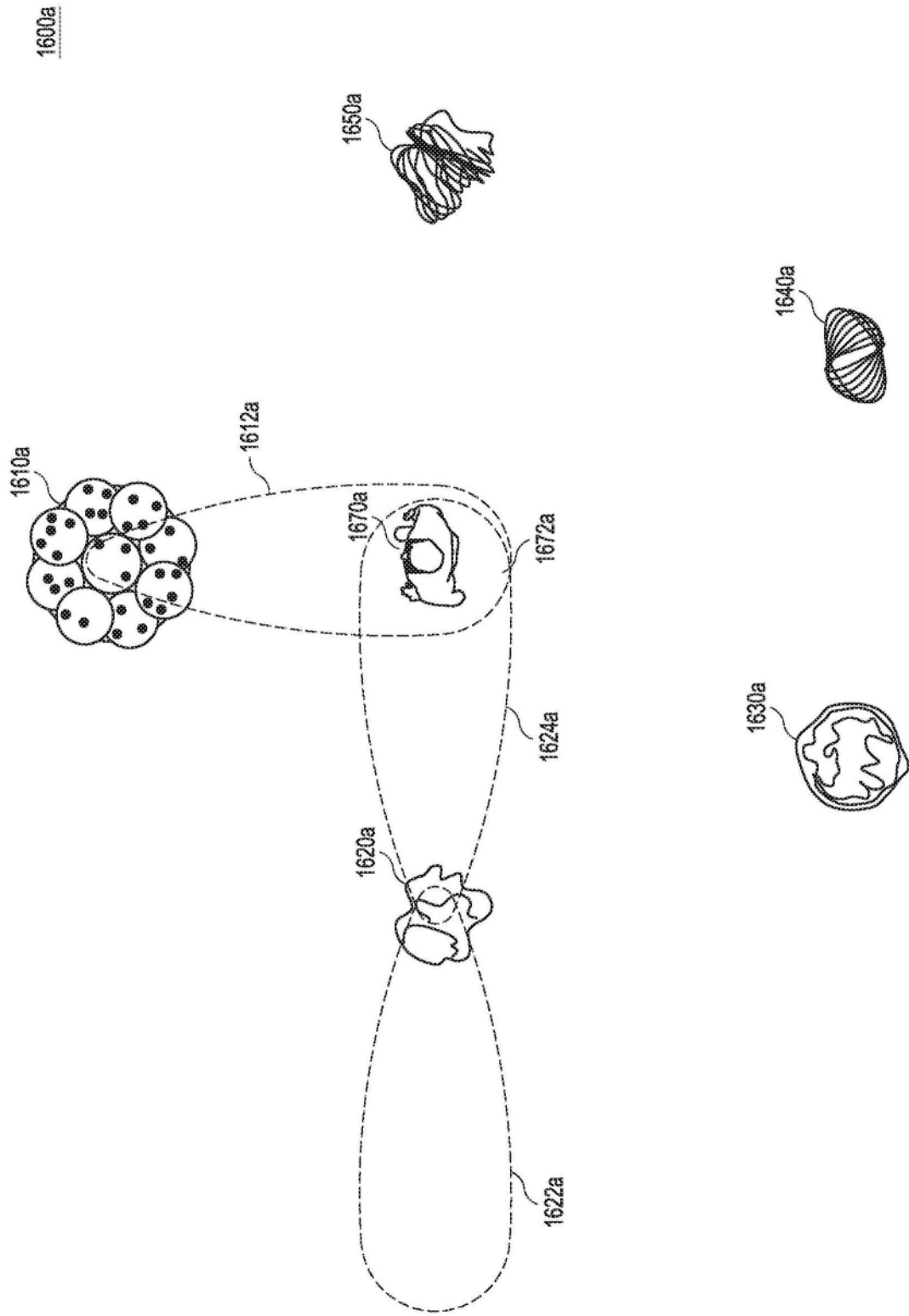


图16A

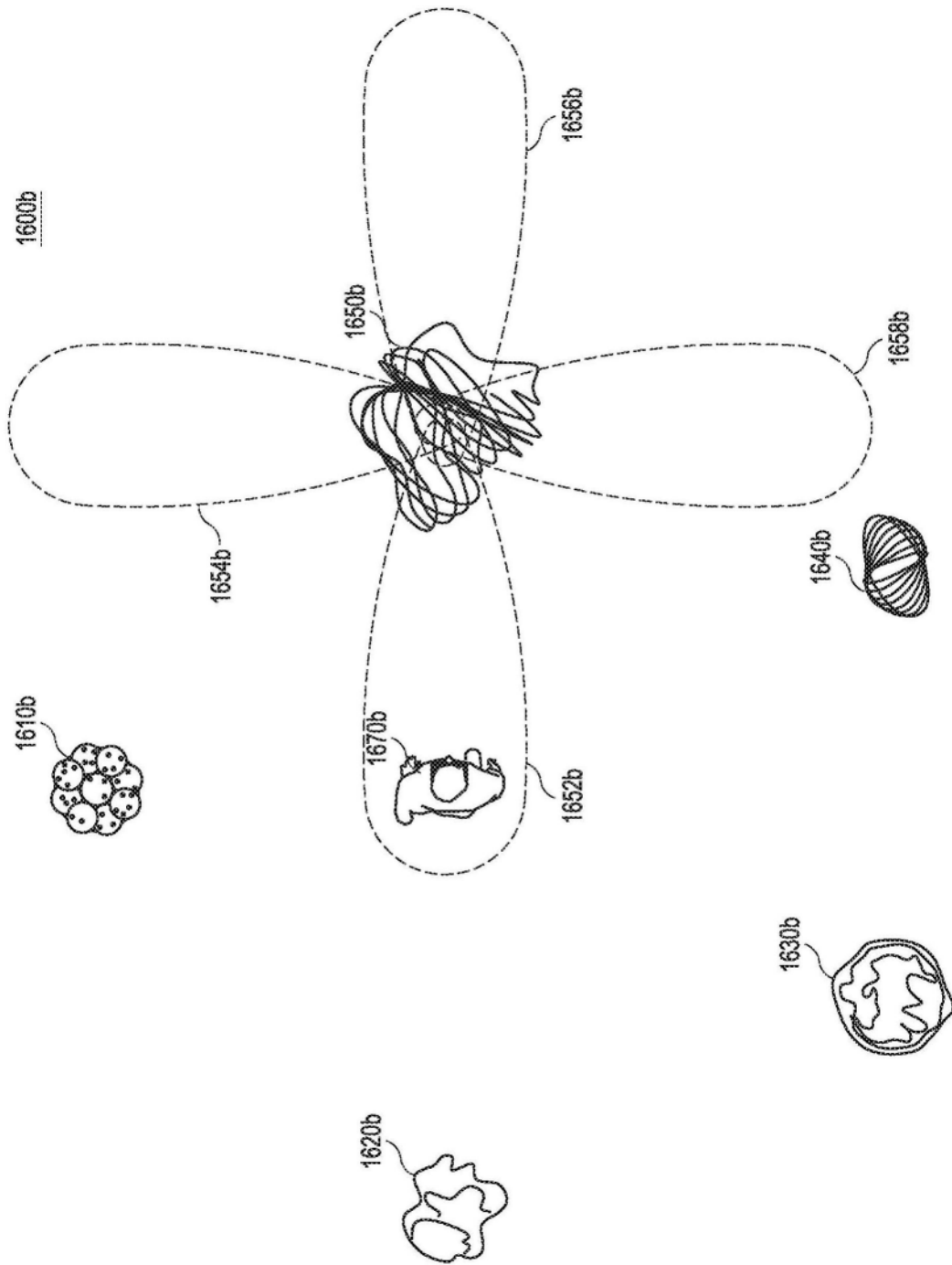


图16B

1700

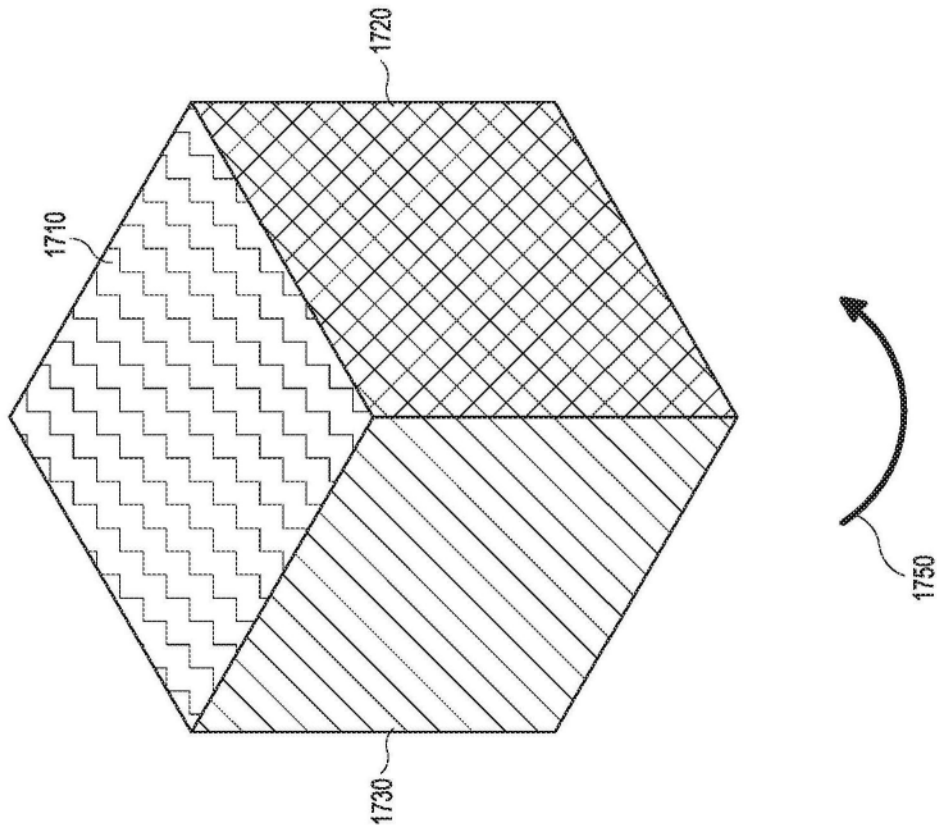


图17

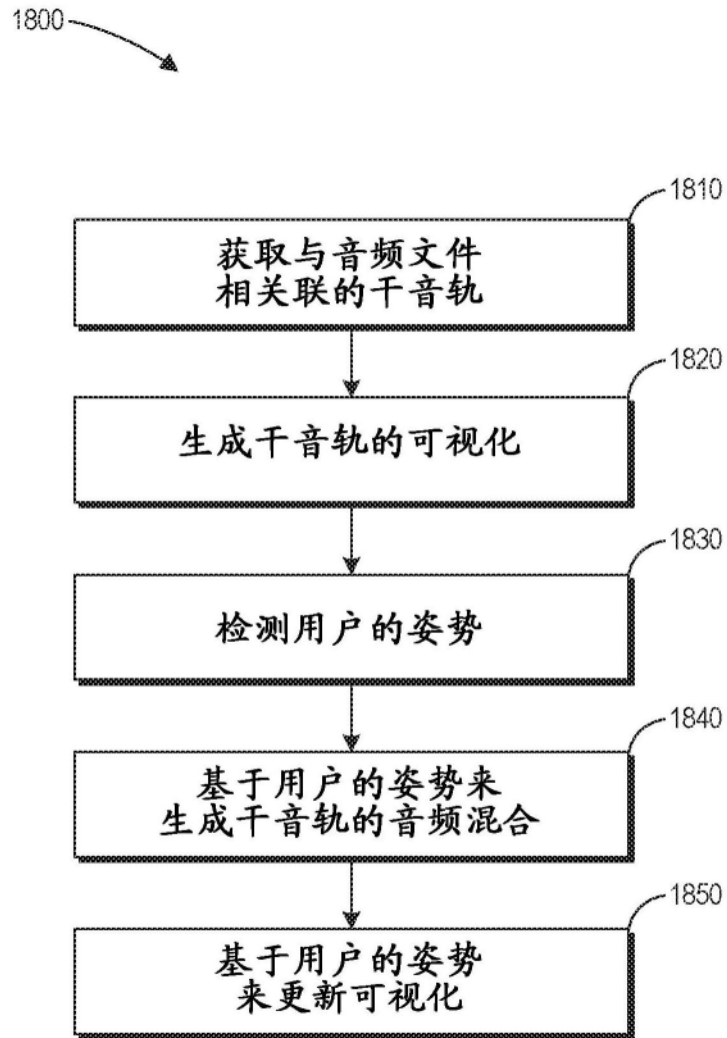


图18