



(12) 发明专利

(10) 授权公告号 CN 102413414 B

(45) 授权公告日 2015.03.25

(21) 申请号 201110321276.6

US 2009214045 A1, 2009.08.27,

(22) 申请日 2011.10.12

Mohan, A et al. Using computer vision to generate customized spatial audio. 《Multimedia and Expo, 2003. ICME'03. Proceedings. 2003》. 2003,

(30) 优先权数据

12/903,610 2010.10.13 US

(73) 专利权人 微软公司

审查员 刘珊

地址 美国华盛顿州

(72) 发明人 J·弗莱克斯 A·巴-泽埃夫

J·塔迪夫

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 胡利鸣

(51) Int. Cl.

H04S 3/00(2006.01)

(56) 对比文件

US 6500008 B1, 2002.12.31,

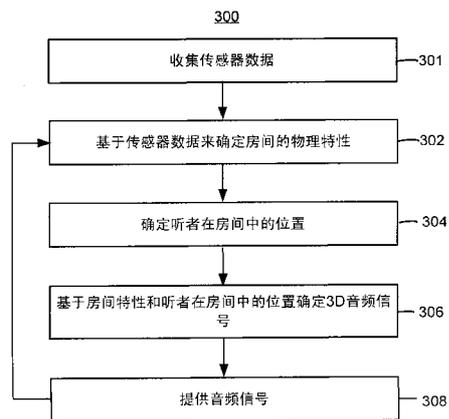
权利要求书2页 说明书15页 附图11页

(54) 发明名称

用于扩展现实的高精度 3 维音频的系统和方法

(57) 摘要

本发明涉及用于扩展现实的高精度 3 维音频的系统和方法。提供了用于提供 3D 音频的技术，该 3D 音频可以用于扩展的现实。3D 音频信号可以基于从听者所在的实际房间中所采集的传感器数据以及听者在该房间中的实际姿态来生成。该 3D 音频信号可以包括基于所采集的传感器数据和听者的位置所确定的多个分量。例如，可以确定虚拟声音源与听者之间的多个（虚拟）声音路径。该传感器数据可以用于估计房间中的材料，使得可以确定这些材料在声音在沿着这些路径传播时可能对声音造成的影响。在一些实施例中，传感器数据可以用于收集听者的物理特性，使得可以从 HRTF 库中确定合适的 HRTF。



1. 一种用于提供 3D 音频信号的方法,包括:
 - 收集关于房间以及听者在所述房间中的姿态的传感器数据,所述传感器数据包括关于所述听者的耳朵的深度信息;
 - 基于传感器数据确定房间的实际物理特性 (302);
 - 确定虚拟环境的虚拟特性;
 - 将所述虚拟特性应用于所述实际物理特性;
 - 构建包括应用于所述房间的实际物理特性的虚拟特性的所述房间模型,包括用来自所述虚拟环境的虚拟物体的虚拟物理特性替代所述房间中实际物体的实际物理特性;
 - 基于所述传感器数据确定听者在所述房间中的位置 (304);
 - 从关于所述听者的耳朵的深度信息确定所述听者的耳朵的物理特性;
 - 基于所述听者的耳朵的物理特性从与头部相关的传输函数 HRTF 库中确定 HRTF;
 - 基于包括应用于所述房间的实际物理特性的虚拟物理特性的所述房间模型和所述听者在所述房间中的位置以及所述听者的 HRTF 来确定 3D 音频信号,基于包括应用于所述房间的实际物理特性的虚拟物理特性的所述房间模型来确定 3D 音频信号包括确定虚拟物体的虚拟物理特性将会如何影响声音反射 (306)。
2. 如权利要求 1 所述的方法,其特征在于,确定 3D 音频信号包括:
 - 基于所述房间的物理参数以及所述听者在所述房间中的位置来确定虚拟声音源与所述听者的每一耳朵之间的多个声音路径。
3. 如权利要求 2 所述的方法,其特征在于,确定 3D 音频信号包括:
 - 确定 3D 音频信号的多个分量,其中为所述多个声音路径中的每一个声音路径确定分量;以及
 - 将针对所述听者的所述 HRTF 应用于所述多个分量中的每个 (308)。
4. 如权利要求 1 至 3 中的任一项所述的方法,其特征在于,还包括:基于所述传感器数据估计所述房间中的物体的材料,确定 3D 音频信号还基于所述材料将对声音造成的影响。
5. 如权利要求 1 所述的方法,其特征在于,收集传感器数据包括:
 - 收集图像传感器数据;以及
 - 确定所述听者在所述房间中的位置包括基于所述图像传感器数据确定所述听者在所述房间中的姿态和旋转。
6. 如权利要求 1 所述的方法,其特征在于,确定所述听者在所述房间中的位置包括:跟踪所述听者在所述房间中的姿态。
7. 如权利要求 1 所述的方法,其特征在于,还包括:
 - 在第一时间存储所述听者的耳朵的物理特性;
 - 在晚于第一时间的第二时间基于生物测定信息标识出所述听者,所述生物测定信息不同于所述听者的耳朵的物理特性;以及
 - 基于针对所标识出的听者的耳朵所存储的物理特性从所述库中选择 HRTF。
8. 如权利要求 1 所述的方法,其特征在于,所述传感器数据包括来自所述听者上的一个或多个话筒的传感器数据。
9. 一种用于提供 3D 音频信号的系统,包括:
 - 用于使用收集关于听者的耳廓的深度信息的传感器收集关于环境和听者的数据的装

置；

用于基于所收集的传感器数据确定所述环境中的实际物体的物理尺寸和位置的装

置；

用于从软件应用访问虚拟环境的虚拟物理特性的装置；

用于将所述虚拟物理特性应用于所述实际物体的所述物理尺寸和位置来构建模型的装置；

用于基于所收集的传感器数据确定所述听者在所述环境中的位置的装置；

用于根据所述深度信息确定所述听者的耳廓的物理特性的装置；

用于基于包括应用于所述环境中的实际物体的物理尺寸和位置的虚拟物理特性的模型以及所述听者在所述环境中的位置来确定 3D 音频信号的多个分量的装置,包括用于确定应用于所述实际物体的物理尺寸和位置的虚拟物理特性将会如何影响声音反射的装置；

用于访问与头部相关的传输函数 HRTF 的库并基于所述听者的耳廓的物理特性为所述听者选择 HRTF 的装置；

用于将针对所述听者的所选 HRTF 应用于所述多个分量中的每个的装置；以及

用于提供 3D 音频信号的装置。

用于扩展现实的高精度 3 维音频的系统和方法

技术领域

[0001] 本发明涉及用于扩展现实的高精度 3 维音频的系统和方法。

背景技术

[0002] 公知的是,人类具有使用他们的耳朵在即使没有任何视觉线索的情况下识别声音源的能力。人类通过采用从一只耳朵导出的线索并且通过比较在两只耳朵处所接收的线索(差异线索或双耳线索)来估计源的位置。在这些差异线索之中有到达的时间差和强度差。单耳线索来自声音源与人类解剖结构之间的交互,在人类解剖结构中,原始源声音在其进入耳道以前被修改,以供由听觉系统进行处理。

[0003] 在现实世界情况下,声音实际上是从特定位置处发出的。所期望的可能是,使听者能够感受到:由音频扬声器产生的声音似乎来自 3 维空间中的特定位置。一种可能的技术包括让用户佩戴亦称“头戴式耳机”的“头戴耳机”。也就是说,一个音频扬声器被放置在每个耳朵之上或附近。该技术可以采用使用“与头部相关的传输函数”(HRTF)来创建音频信号以制造声音源自 3D 空间中的某位置这一错觉。在此,制造声音来自 3D 空间中的某位置这一错觉的音频信号被称为 3D 音频信号。

[0004] HRTF 可以基于自由空间中的声音与到达耳膜时的声音之间的差异来定义。HRTF 描述:给定的声波输入(其被参数化为频率和声音位置)在该声音达到耳膜和内耳以前如何被头部和耳廓的衍射和反射性质过滤。HRTF 可以与人的头部形状和他们耳朵的物理特性密切相关。因此,HRTF 可能因人而异而显著不同。因此,尽管 HRTF 可以用于帮助创建 3D 音频信号,但是在使 HRTF 适应于每个用户方面仍然存在挑战。

[0005] 3D 音频的一种可能的应用是在扩展的现实场景中。扩展的现实可以被定义为使用某种计算机生成的技术来扩展现实世界情况。扩展的现实、以及其他 3D 应用需要精确的 3-D 音频。例如,用户应当能够将声音精确定位为来自虚拟声音源。

[0006] 尽管存在 3D 音频技术,但是期望改进。如已经提到的那样,一种改进是为用户提供精确的 HRTF。但是,还期望其它改进。3D 音频信号应当是精确的、消费者友好的、节省成本的,并且与现有音频系统兼容。

发明内容

[0007] 提供了用于提供 3D 音频的技术。3D 音频可以用在扩展的现实中,但是这不是必需的。在此所公开的技术是精确的、节省成本的、用户友好的,并且与现有音频系统兼容。技术可以使用一个或多个传感器来收集描述听者所处的环境(例如房间)以及听者在房间中的位置的现实世界数据。逼真的 3D 音频信号可以基于从传感器收集的数据生成。一种选择是使用传感器来收集描述听者的物理特性(例如头部和耳廓形状和大小)的数据以便为该听者确定合适的 HRTF。

[0008] 一个实施例包括一种方法,该方法包括:基于传感器数据确定房间的物理特性;确定听者在该房间中的位置;以及基于该房间的物理特性和听者在该房间中的位置确定

3D 音频信号。

[0009] 一个实施例包括一种装置,该装置包括一个或多个传感器、处理器以及计算机可读存储介质。该计算机可读存储介质其上存储有指令,所述指令在处理器上执行时致使该处理器使用传感器来收集关于环境和听者的数据。处理器基于传感器数据确定环境的物理特性以及听者在该环境中的位置。处理器基于环境的物理特性和听者在该环境中的位置来确定 3D 音频信号的不同分量。处理器将针对听者的与头部相关的传输函数 (HRTF) 应用于 3D 音频信号的每个分量,并且提供 3D 音频信号。

[0010] 一个实施例包括一种用于提供 3D 音频信号的方法。该方法可以包括:收集传感器数据,该传感器数据可以包括深度信息。基于深度信息确定听者的物理参数。可以基于 HRTF 库确定与头部相关的传输函数 (HRTF)——该确定可以基于听者的物理参数。可以收集包括关于房间的深度信息的传感器数据。可以基于该深度信息确定房间的物理参数。可以在房间中确定听者的位置。可以基于房间的物理参数以及听者在该房间中的位置来确定虚拟声音源与该听者之间的声音路径。基于房间的物理参数,可以为每个声音路径确定 3D 音频信号的分量。针对听者的 HRTF 可以应用于 3D 音频信号的每个分量,并且可以提供该 3D 音频信号。

[0011] 提供本发明内容以便以简化形式介绍将在以下的具体实施方式中进一步描述的一些概念。本发明内容并不旨在标识出所要求保护的主题的关键特征或必要特征,也不旨在用于限定所要求保护的主题的范围。

附图说明

[0012] 图 1 描绘了运动捕捉系统的示例性实施例。

[0013] 图 2 描绘了图 1 的运动捕捉系统的示例性框图。

[0014] 图 3 是用于提供 3D 音频信号的过程的一个实施例的流程图。

[0015] 图 4A 描绘了用于确定房间的模型的过程的流程图。

[0016] 图 4B 描绘了用于基于虚拟特性和真实特性来构建房间模型的过程的一个实施例的流程图。

[0017] 图 5A 是用于确定 3D 音频信号的音频分量的过程的一个实施例的流程图。

[0018] 图 5B 描绘了房间的顶视图以以 2 维示出可能的声音路径。

[0019] 图 6 示出了用于确定听者在房间中的位置和旋转的过程的流程图。

[0020] 图 7 描述了用于为特定听者确定 HRTF 的过程的一个实施例。

[0021] 图 8 描绘了用于基于之前收集的详细特性来为听者选择 HRTF 的过程的一个实施例的流程图。

[0022] 图 9 是描绘用于基于这样的数据修改房间模型的过程的一个实施例的流程图。

[0023] 图 10 描绘了用于生成 3D 音频信号的一个实施例的框图。

[0024] 图 11 描绘了可以在图 1 的运动捕捉系统中使用的计算环境的示例框图。

[0025] 图 12 描绘了可以在图 1 的运动捕捉系统中使用的计算环境的另一示例框图。

具体实施方式

[0026] 提供了用于提供 3D 音频的技术。3D 音频可以用于扩展现实,但是其他应用也是

可能的。在此所公开的技术是精确的、节省成本的、用户友好的，并且与现有音频系统兼容。3D 音频信号可以基于从听者所在的实际房间中所收集的传感器数据以及听者在该房间中的实际姿态来生成。该音频信号可以表示“虚拟声音”，该虚拟声音表示来自 3D 空间中的某个特定位置的声音。该位置可以表示在视频屏幕上所显示的某个物体、或者甚至听者房间中的真实实物物体。在一些实施例中，3D 音频信号通过一组头戴耳机被提供给听者。该 3D 音频信号可以包括多个分量，这些分量基于所收集的传感器数据和听者的位置来确定。例如，墙壁和家具的位置可以从传感器数据中确定。还可以确定虚拟声音源与听者之间的多个（虚拟）声音路径。该传感器数据可以用于估计房间中的材料，使得可以确定这些材料在声音在沿着这些路径传播时可能对声音造成的影响。在一些实施例中，可以分析传感器数据以确定听者的物理特性，使得可以从 HRTF 库中确定合适的 HRTF。针对听者的 HRTF 可以应用于 3D 音频信号的不同分量。下面讨论其他细节。

[0027] 在一些实施例中，生成 3D 音频信号被用于运动捕捉系统中。因此，将描述示例性的运动捕捉系统。然而，可以理解，本文所描述的技术不限于运动捕捉系统。图 1 描绘了运动捕捉和 3D 音频系统 10 的示例，其中房间（或其他环境）中的人与应用交互。运动捕捉和 3D 音频系统 10 包括显示器 196、深度相机系统 20、以及计算环境或装置 12。深度相机系统 20 可以包括图像相机组件 22，该图像相机组件 22 具有光发射器 24、光传感器 25、以及红绿蓝 (RGB) 相机 28。在一个实施例中，光发射器 24 发射准直光束。准直光束的示例可包括但不限于，红外 (IR) 和激光。在一个实施例中，光发射器 24 是 LED。从视野 6 内的听者 8、物体 33、墙壁 35 等等反射离开的光由光传感器 25 来检测。在一些实施例中，系统 10 使用该信息来确定如何生成 3D 音频信号。描述该房间的诸如 RGB 信息（其在下面予以讨论）之类其他信息可以用于确定如何生成 3D 音频信号。

[0028] 亦称为听者的用户站在深度相机系统 20 的视野 6 中。听者 8 佩戴了用来提供 3D 音频声音的头戴耳机 27。在该示例中，头戴耳机 27 包括两个音频扬声器 37，所述音频扬声器 37 之一佩戴在每只耳朵之上或附近。系统 10 可以提供驱动音频扬声器 37 的 3D 音频信号。该 3D 音频信号可以使用无线或有线连接来提供。在某个实施例中，系统 10 将 3D 音频信号提供给另一组件，比如高保真度立体声系统、HDTV 等等。

[0029] 对于听者 8 而言，3D 音频信号中的声音可能似乎源自某个虚拟声音源 29。作为一个示例，虚拟声音源 29a 可以是显示器 196 上所显示的物体。然而，虚拟声音源 29a 可以对应于房间中的某个真实物体 29b。例如，可以指示用户将玩偶侏儒 (gnome) 放置在他们面前的桌上，其中系统 10 可以使得对用户而言就好像该玩偶侏儒正在与他们交谈（由于通过头戴 27 所播放的 3D 音频的缘故）。虚拟声音源 29 甚至可能好像源自房间之外。

[0030] 在一些实施例中，用户“佩戴”一个或多个话筒 31，所述话筒可以被系统 10 用于确定房间的声学性质以提供更逼真的 3D 音频信号。在该示例中，话筒 31 位于头戴耳机 27 上，但是用户可以将话筒 31 “佩戴”在另一位置处。在一些实施例中，用户“佩戴”一个或多个惯性传感器 38，所述惯性传感器可以被系统 10 用于确定听者 8 的位置和旋转。在该示例中，惯性传感器 38 位于用户的头上，但是用户可以将惯性传感器 38 “佩戴”在另一位置处。例如，惯性传感器 38 可以集成到头戴耳机 27 中。在一些实施例中，用户 8 可以携带相机，该相机可以用于向系统 10 提供与深度相机系统 20 所生成的深度和 / 或 RGB 信息类似的深度和 / 或 RGB 信息。

[0031] 线 2 和 4 表示视野 6 的边界。可以定义笛卡儿世界坐标系,其包括:沿着深度相机系统 20 的焦距例如水平地延伸的 z 轴;垂直地延伸的 y 轴;以及侧向地且水平地延伸的 x 轴。注意,附图的透视被修改成简化表示,显示器 196 在 y 轴方向上垂直延伸,z 轴垂直于 y 轴和 x 轴且与用户所站立的地面平行地从深度相机系统 20 延伸出来。

[0032] 一般而言,运动捕捉系统 10 用于识别、分析和 / 或跟踪对象。计算环境 12 可包括计算机、游戏系统或控制台等等,以及执行应用程序的硬件组件和 / 或软件组件。

[0033] 深度相机系统 20 可以包括相机,该相机用于可视地监视诸如用户之类的一个或多个物体 8,使得可以捕捉、分析并跟踪用户所作出的姿势和 / 或运动,以执行应用中的一个或多个控制或动作,比如选中用户界面 (UI) 中的菜单项。

[0034] 运动捕捉系统 10 可以连接到诸如显示器 196 等可向用户提供视觉和音频输出的视听设备,如电视机、监视器、高清电视机 (HDTV) 等,或甚至是墙或其他表面上的投影。还可以经由单独的设备来提供音频输出。注意,3D 音频信号通常通过头戴耳机 27 来提供。为驱动显示器,计算环境 12 可包括提供与应用相关联的视听信号的诸如图形卡之类的视频适配器和 / 或诸如声卡之类的音频适配器。显示器 196 可以通过例如 S- 视频电缆、同轴电缆、HDMI 电缆、DVI 电缆、VGA 电缆等等连接到计算环境 12。

[0035] 图 2 描绘了图 1 的运动捕捉和 3D 音频系统 10 的示例性框图。系统 10 包括深度相机系统 20 和计算环境 12。在该实施例中,计算环境 12 具有 3D 音频生成 195。计算环境 12 输入来自深度相机系统 20 的深度信息和 RGB 信息,并且将 3D 音频信号输出给音频放大器 197。音频放大器 197 可以是诸如 HDTV、立体声系统等等之类的单独设备的一部分。3D 音频生成 195 可以通过在处理器 192 上执行指令来实现。注意,硬件执行的实施方式、以及混合型软件 / 硬件实施方式也是可能的。

[0036] 深度相机系统 20 可被配置成生成深度图像,该深度图像可以包括深度值。深度相机系统 20 可以将深度图像组织成“Z 层”,或者可以与从深度相机系统 20 沿其视线延伸的 Z 轴垂直的层。深度图像可包括所捕捉的场景的二维 (2-D) 像素区域,其中该 2-D 像素区域中的每个像素具有代表距离图像相机组件 22 的线性距离 (径向距离) 的或由像素观察的 3D 位置的 Z 分量 (垂直距离) 的相关联的深度值。

[0037] 图像相机组件 22 可以包括光发射器 24 和一个或多个光传感器 25 来捕捉从视野内的物体反射离开的光的强度。例如,深度相机系统 20 可使用光发射器 24 将光发射到物理空间上并使用光传感器 25 来检测从物理空间内一个或多个对象的表面所反射的光。在一些实施例中,深度值基于光强来确定。例如,随着时间的过去,越来越多的光子达到给定的像素。在收集时间段以后,每个像素处的光强被采样。深度图像中的深度值可以基于每个像素处的光强来确定。在一些实施例中,光发射器 24 发射脉冲红外光。在一些实施例中,光被调制到所期望的频率。

[0038] 红绿蓝 (RGB) 相机 28 可用于捕捉可见光图像。深度相机系统 20 还可以包括话筒 30,其包括例如接收声波并将其转换成电信号的换能器或传感器。另外,话筒 30 可用于接收也可由人提供的诸如声音之类的音频信号,以控制可由计算环境 12 运行的应用程序。音频信号可包括人的口声,如说的话、口哨声、喊声及其他发声,以及非口声,如掌声或跺脚。在一些实施例中,话筒 30 是话筒阵列,它可具有一起运行的任何数量的话筒。如图 1 中所述,还可以由用户 8 佩戴一个或多个话筒 31。这些话筒 31 的输出可以被提供给计算环境

12 以供由 3D 音频生成 195 来使用。如果期望的话,则话筒 30 的输出也可以由 3D 音频生成 195 来使用。

[0039] 深度相机系统 20 可包括与图像相机组件 22 进行通信的处理器 32。处理器 32 可以包括可执行例如包括用于生成 3D 音频信号的指令在内的指令的标准化处理器、专用处理器、微处理器等等。

[0040] 深度相机系统 20 还可包括存储器组件 34,存储器组件 34 可存储可由处理器 32 执行的指令、以及存储 RGB 相机所捕捉的图像或图像帧、或任何其他合适的信息、图像等等。根据一个示例实施方式,存储器组件 34 可包括随机存取存储器 (RAM)、只读存储器 (ROM)、高速缓存、闪存、硬盘或任何其他合适的有形计算机可读存储组件。存储器组件 34 可以是经由总线 21 与图像捕捉组件 22 和处理器 32 通信的单独的组件。根据另一实施例,存储器组件 34 可被集成到处理器 32 和 / 或图像捕捉组件 22 中。

[0041] 深度相机系统 20 可以通过通信链路 36 与计算环境 12 进行通信。通信链路 36 可以是有线和 / 或无线连接。根据一个实施方式,计算环境 12 可以经由通信链路 36 向深度相机系统 20 提供时钟信号,该时钟信号指出何时从位于深度相机系统 20 的视野中的物理空间捕捉图像数据。

[0042] 此外,深度相机系统 20 可通过通信链路 36 向计算环境 12 提供深度信息和由 RGB 相机 28 捕捉的图像。计算环境 12 然后可使用深度信息和所捕捉的图像来控制应用。例如,如图 2 所示,计算环境 12 可包括诸如姿势过滤器集合等姿势库 190,每一姿势过滤器具有关于可 (在用户移动时) 执行的姿势的信息。例如,可以为各种手势 (如手的猛击或投掷) 提供姿势过滤器。通过将检测到的运动与每一个过滤器进行比较,可以标识由人执行的指定的姿势或运动。还可以确定执行运动的程度。

[0043] 计算环境还可包括处理器 192,其用于执行存储在存储器 194 中的指令以向显示设备 196 提供音频 - 视频输出信号并实现其他功能。

[0044] 图 3 是用于提供 3D 音频信号的过程 300 的一个实施例的流程图。过程 300 可以在系统 10 内实现,但是可以使用不同的系统。在步骤 301,收集传感器数据。该传感器数据可以包括、但不限于深度信息、GRB 数据以及音频数据。例如,可以使用深度相机系统 20 来用光传感器 25 收集其 (使用光发射器 24) 所发射的光。还可以使用 RGB 相机 28。在一个实施例中,使用由用户 8 佩戴的一个或多个话筒 31 来收集传感器数据。也可以使用深度相机系统 20 中的话筒 30。在一个实施例中,用户 8 握住相机并且将其在四周移动以收集房间周围的传感器数据。该数据可以包括深度信息和 RGB 数据。

[0045] 在步骤 302,基于传感器数据来确定听者所在的房间或其他环境的物理特性。该传感器数据可以用于确定诸如墙壁和各个物体位于何处之类的信息。而且,该传感器数据可以用于估计房间中的材料。例如,传感器数据可以用于确定地面是硬木还是地毯。

[0046] 在步骤 304,确定听者在房间中的位置。在一个实施例中,使用传感器数据来确定听者的位置。例如,可以使用在步骤 302 所收集的传感器数据来确定听者的位置。

[0047] 在步骤 306,基于听者在房间中的位置以及该房间的一个或多个物理特性来确定 3D 音频信号。作为一个示例,可以确定虚拟声音源与听者之间的多个声音路径。此外,可以将房间的物理特性作为因素计入。作为一个示例,从硬木地面反射离开的声音将不同于从地毯反射的声音。因此,对于具有这样的路径的声音路径而言,这可以作为因素计入。在一

些实施例中,应用针对听者的 HRTF 以形成 3D 音频信号。在一些实施例中,基于传感器确定的特性来确定针对听者的 HRTF。例如,图像相机组件 20 中的传感器可以捕捉深度和 / 或 RGB 数据。可以存在 HRTF 库,其中基于匹配过程从该库中选择(或以其他方式确定)合适的 HRTF。

[0048] 在步骤 308 中,提供 3D 音频信号。例如,3D 音频信号被提供给用于驱动头戴耳机 27 的音频放大器 197。注意,过程 300 可以通过如下方式重复:收集更多传感器数据(301);重新确定房间的物理特性(步骤 302);重新确定听者的位置(步骤 304)等等。然而,不需要连续地重复所有步骤。例如,过程 300 可以以任何所期望的间隔来重新确定房间特性。可以预期某些信息保持不变(例如墙壁的位置)。然而,诸如物体位置之类的其他房间信息可能随时间改变。由于听者的位置可能非常频繁地改变,因此可以仔细地跟踪听者的位置。

[0049] 图 4A 描绘了用于确定房间模型的过程 400 的一个实施例的流程图。过程 400 可以用在过程 300 的步骤 301、302 和 306 中。例如,该模型可以从在步骤 301 所收集的传感器数据中构建,并且在步骤 306 被用于确定音频分量。在步骤 402,生成房间中的一个或多个物体的深度图像。在一个实施例中,通过由深度相机系统 20 将 IR 射束发射到视野内并且在一个或多个图像传感器处收集所反射的数据来形成深度图像。然后,处理该传感器数据以确定深度值(例如与各个物体相距的距离)。注意,由于视野可能是有限的,因此深度相机系统 20 可以调节视野并且重复收集附加的深度信息。在一些实施例中,图像相机组件 22 由电机来控制,该电机允许移动视野以捕捉房间的更完整的照片。如上所述,用户 8 可以握住相机并且将其用于扫描房间以收集深度数据。

[0050] 在步骤 404,生成房间中的一个或多个物体的 RGB 图像。在一个实施例中,由深度相机系统 20 使用红绿蓝(RGB)相机 28 来形成 RGB 图像。如上所述,用户 8 可以握住相机并且将其用于扫描房间以收集 RGB 数据。RGB 图像可以与深度图像一起从一个以上的数据收集步骤中形成。步骤 402 和 404 是步骤 301 的一个实施例。

[0051] 在步骤 406,确定房间和该房间中的物体的物理尺寸。还可以确定物体的物理位置。该信息可以基于在步骤 402 和 404 所收集的数据。在一些实施例中,基于所收集的数据来外插出该物理尺寸。如所述那样,深度相机系统 20 可能不能不收集整个房间的数据。例如,参考图 1,视野可能不能捕捉整个墙壁 35。在这样的情况下,一种选择是对所收集的数据进行外插以针对没有数据的区域估计墙壁 35 的位置。步骤 406 是步骤 302 的一个实施例。

[0052] 在步骤 408,对房间中的物体的材料进行估计。作为一个示例,对各件家具、墙壁、天花板、地面等等的材料进行估计。在一些实施例中,使用深度信息来辅助该确定。例如,该深度信息可以用于确定地板是光滑的(以及可能确定:是硬木的或是铺了地板砖的)还是粗糙的(可能确定:是铺了地毯的)。还可以使用 RGB 信息。注意,不需要估计实际材料,但是这是一种选择。估计材料的原因是能够确定该材料将如何影响声音。因此,可以确定和排序可用于确定材料将如何影响声音从物体离开的反射的任何参数。

[0053] 在步骤 410,基于在步骤 406 和 408 确定的物理尺寸和材料来构造房间模型。之后,可以基于该房间模型生成 3D 音频信号。例如,该模型可以用在过程 300 的步骤 306。因此,可以用 3D 音频信号来扩展用户房间的实际现实。步骤 406、408 和 410 是步骤 302 的一个实施例。

[0054] 图 4B 描绘了用于基于虚拟特性和真实特性来构建房间模型的过程 450 的一个实施例。作为一个示例,可以使用处理器 450 来使得对于听者而言好像他们的房间被以某种方式变换。例如,如果用户玩视频游戏,其中用户想象他们处于监狱牢房中,则处理器 450 可以用于构建具有监狱牢房的特性的房间模型。该模型可以使用用户房间的一些实际特性,比如物体的大小和位置。然而,替代于使用真实物体的实际材料,可以使用虚拟特性。例如,替代于实际的厚地毯,可以对水泥地面进行建模。因此,听者的房间的现实可以基于该模型由 3D 音频信号来扩展。

[0055] 在步骤 452,访问物体在实际房间中的物理尺寸和位置。这些特性可以已经使用过程 400 被确定。然而,实际房间特性在期望时可以重新确定。

[0056] 在步骤 454,确定虚拟环境的特性。例如,实现虚拟游戏的软件应用可以提供定义虚拟环境的参数。在本示例中,该应用可以提供描述墙壁、地面、天花板等等的虚拟材料的参数。注意,这些参数可以以另一方式来确定。

[0057] 在步骤 456,将虚拟特性应用于实际房间特性。因此,替代于确定用户的实际地面是铺了地毯的以及确定声音将如何受到地毯影响,将用户的地面建模为水泥。然后,作出关于水泥将如何影响声音反射的确定。如果期望的话,则可以让虚拟特性应用于房间中的各个物体。例如,如果期望让沙发模拟大石 (boulder),则可以让石头的特性应用于沙发。

[0058] 在步骤 458,基于来自步骤 456 的信息来构建用户房间的模型。该模型可以在生成 3D 音频信号时使用。例如,该模型可以用在图 3 的过程 300 的步骤 306。注意,用户房间中的实际物体(家具、墙壁、天花板等等)可以用于确定该模型。因此,可以通过 3D 音频信号来扩展用户的实际房间的现实。

[0059] 图 5A 是用于确定 3D 音频信号的分量的过程 500 的一个实施例的流程图。过程 500 是过程 300 的步骤 306 的一个实施例。在步骤 502,确定虚拟声音源 29 的位置。例如,如果用户与显示器 196 上所示的虚拟世界交互,则虚拟声音源 29 可以是该虚拟世界中所显示的某个物体。然而,虚拟声音源 29 可以是用户房间中的实际物体或虚拟物体。例如,用户可以将物体放置在该房间中的所期望的位置处。然后,系统可以标识出该物体的位置。作为特定的示例,系统可以指示用户将物体放置在用户想要放置的地方。作为响应,用户可以在桌上放置玩偶侏儒。然后,系统例如通过使用深度相机系统来确定物体的位置。如前面所讨论的那样,系统可以跟踪用户的物理位置。因此,系统能够通过跟踪用户的运动来确定用户已经将玩偶侏儒放置在了桌上。可以将其他技术用于该系统以确定虚拟声音源 29 的实际位置。虚拟声音源 29 甚至可以处于房间之外。例如,系统可以使得某人好像在敲门或者从门的另一侧谈话。

[0060] 在步骤 504,确定虚拟声音源 29 与听者 8 之间的声音路径。这可以包括确定直接路径和一个或多个间接路径。步骤 504 可以基于已经在过程 300 的步骤 302 确定的房间信息。图 5B 描绘了房间的顶视图以 2 维示出可能的声音路径。注意,系统 10 可以以 3 维确定声音路径;然而使用 2 维来简化说明。在步骤 504 以前,系统可以确定听者 8 和房间中的其他物体 33 的位置。作为一个示例,其他物体 33 可以是沙发。在该示例中,声音路径包括直接声音路径和两个间接声音路径。一个间接声音路径是包括从一个物体的声音反射的一阶路径。还描绘了包括从两个物体的反射的二阶路径。在该示例中,物体 33 阻挡潜在的一阶路径(其由指向物体 33 的虚线箭头来指示)。还可以确定三阶和更高阶的路径。注

意,可以考虑从墙壁以外的物体离开的反射。图 5B 的具体视图未描绘离开地面和天花板的
声音反射,但是也可以考虑这些声音路径。

[0061] 在步骤 506,为每个声音路径确定 3D 音频信号的分量。这些不同分量可以联合以
形成 3D 音频信号。关于房间中的材料的信息可以用在步骤 506。例如,如果已经确定:沿
着一阶路径存在关闭的窗,则可以将声音从玻璃反射离开的影响作为因素计入。另一方面,
可能确定:该窗当前为打开的,在这种情况下,一阶路径可以不加以考虑。作为另一示例,
窗帘可能被关闭,在这种情况下,考虑窗帘对一阶路径上传播的声音的影响。如之前所述,关
于房间的信息可以以任何所期望的间隔更新。因此,当用户进行交互时,所生成的 3D 音频
信号可能由于诸如用户打开窗、关闭窗帘等等之类的情况而改变。

[0062] 在步骤 508,将针对听者的 HRTF 应用于每个音频分量。下面讨论为听者确定合适
HRTF 的进一步细节。在将 HRTF 应用于每个音频分量以后,这些分量可以合并以生成 3D 音
频信号。注意,在输出 3D 音频信号以前可以执行其他处理。

[0063] 图 6 示出了用于确定听者在房间中的位置和旋转的过程 600 的流程图。例如,过程
600 可以用于确定用户的头旋转到哪个方位。过程 600 是过程 300 的步骤 304 的一个实施
例。注意,过程 600 不一定包括收集用于为听者 8 确定合适 HRTF 的信息。该信息可以如下
面所述那样以更加受限的基础来收集。该示例性方法例如可以使用深度相机系统 20 来实
现。可以扫描用户 8 来生成模型,比如骨架模型、网格人类模型、或人的任何其他合适的表
示。然后,该模型可以与房间信息一起用于确定用户在该房间中的位置。用户的旋转(例
如用户的头所取向的方位)也可以从该模型中确定。

[0064] 根据一个实施方式,在步骤 602 中,例如,从深度相机系统接收深度信息。可以将
深度图像下采样到较低的处理分辨率,使得其可以更容易地用较少的计算开销来使用和处
理。另外,可从深度图像中移除和/或平滑掉一个或多个高变度和/或含噪声的深度值;可
填入和/或重构缺少的和/或移除的深度信息的部分;和/或可对所接收的深度信息执行
任何其他合适的处理,使得该深度信息可用于生成诸如骨架模型等模型。

[0065] 在判定步骤 604,判定深度图像是否包括人类目标。这可以包括对深度图像中的每
一个目标或物体进行泛色填充,将该目标或物体与图案进行比较以判断深度图像是否包括
人类目标。例如,可以如上文所描述的那样将深度图像的选定区域或点中的像素的各种深
度值进行比较,以确定可以定义目标或对象的边缘。可基于所确定的边缘来对 Z 层的可能 Z
值进行泛色填充。例如,与确定的边缘相关联的像素和该边缘内的区域的像素可以彼此相
关联,以定义可以与图案相比较的捕捉区域中的目标或对象,这在下面将更详细地描述。

[0066] 如果判定步骤 604 为真,则执行步骤 606。如果判断步骤 604 为假,则在步骤 602
中接收附加的深度信息。

[0067] 每一个目标或对象与其比较的图案可包括一个或多个数据结构,这些数据结构具
有共同地定义人的典型身体的变量集合。可以将与例如视野内的人类目标和非人类目标的
像素相关联的信息与变量进行比较,以标识人类目标。在一个实施例中,可以基于身体部
位,对该集合中的每一个变量赋予权重。例如,图案中的诸如头和/或肩之类的各种身体部
位可以具有与其相关联的权重值,这些权重值可以大于诸如腿之类的其他身体部位的权
重值。根据一个实施方式,当将目标与变量进行比较以判断目标是否可能是人类以及哪些目
标可能是人类时,可以使用权重值。例如,变量和目标之间具有较大的权重值的匹配与具有

较小权重值的匹配相比可产生目标是人类的更大似然性。

[0068] 步骤 606 包括扫描人类目标以寻找身体部位。可以扫描人类目标,以提供与人的一个或多个身体部位相关联的诸如长度、宽度等等之类的测量值,以提供该人的准确模型。在一示例实施例中,人类目标可以被隔离,并且可以创建人类目标的位掩模来扫描一个或多个身体部位。可以通过例如对人目标进行泛色填充来创建位掩模,以便人类目标可以与捕捉区域元素中的其他目标或对象分离。然后,可以对于一个或多个身体部位分析位掩码,以生成人类目标的模型,如骨架模型、网格人类模型等等。

[0069] 步骤 608 包括生成人类目标的模型。在一个实施方式中,可以使用由扫描的位掩模确定的测量值来定义骨架模型中的一个或多个关节。一个或多个关节被用来定义对应于人类的身体部位的一个或多个骨头。一般而言,每个身体部位可被表征为定义骨架模型的关节和骨骼的数学向量。身体部位在关节处可以相对于彼此移动。模型可以包括描述用户的头的旋转的信息,使得得知用户的耳朵的取向。

[0070] 在步骤 610,在用户上的惯性传感器收集数据。在一个实施例中,至少一个惯性传感器位于用户的头上以允许跟踪用户的头。

[0071] 在步骤 611 中,通过每秒钟多次更新人的位置来跟踪模型。随着用户在物理空间中移动,使用来自深度相机系统的信息来调整骨架模型,以使该骨架模型表示人。来自惯性传感器的数据也可以用于跟踪用户。具体而言,可以向骨架模型的一个或多个受力面施加一个或多个力,以将骨架模型调整为更加紧密地对应于物理空间中的人类目标的姿态的姿态。一般而言,可使用用于跟踪一个或多个人的运动的任何已知技术。

[0072] 在步骤 612,基于对模型的跟踪来确定用户在房间中的位置。在步骤 614,基于对模型的跟踪来确定用户的头的旋转。过程 600 可以继续跟踪用户,使得位置和旋转可以更新。

[0073] 在一些实施例中,基于听者 8 的物理特性从 HRTF 库中确定针对用户 8 的 HRTF。这些物理特性可以基于诸如深度信息和 RGB 信息之类的来自传感器的输入来确定。图 7 描述了用于为特定听者 8 确定 HRTF 的过程 700 的一个实施例。该 HRTF 可以用在过程 300 的步骤 306 或者过程 500 的步骤 508。注意,HRTF 可以在任何时间确定。作为一个示例,为用户确定一次 HRTF,并且将其存储以供反复使用。当然,可能修订 HRTF(例如选择新的 HRTF)。

[0074] 在步骤 702,系统 10 指示用户 8 采取某个姿势或姿态。例如,系统指示用户向左看。在步骤 704,系统 10 收集用户处于该姿态的情况下的数据。例如,使用深度相机系统 20 来收集深度信息(利用传感器 25)以及 RGB 信息(利用传感器 28)。在步骤 706,系统数据是否有效。例如,如果系统曾预期针对右耳的数据,则系统确定该数据是否与针对右耳所预期的数据相匹配。如果否,步骤 702 可以重复,使得再次指示用户采取正确的姿态。如果数据有效(步骤 706 为“是”),则系统确定是否存在该用户要采取的更多姿势/姿态。在接下来的迭代中,可以要求用户向正前方看,向右看等等。可以收集针对多种多样的姿态的数据。

[0075] 当收集到合适的的数据时,过程 700 继续到步骤 710 以在步骤 710 为听者 8 确定 HRTF。在一些实施例中,存在可从中进行选择的 HRTF 库。这些 HRTF 可以与用户的各种物理特性相关联。示例包括、但不限于头大小和宽度、耳廓特性、身体尺寸。例如,特定的 HRTF 可以与同头大小和耳廓相关的特定测量结果相关联。这些测量结果可以是某个范围或单个

值。例如，一个测量结果可以是头宽度，这可以表达根据单个值或某个范围来表达。然后，系统可以通过将用户的物理特性与同库中 HRTF 相关联的物理特性进行匹配来为用户选择 HRTF。可以使用任何技术来确定最佳匹配。在一个实施例中，系统进行内插以为用户确定 HRTF。例如，用户的测量结果可以处于两个 HRTF 的测量结果之间，在这种情况下，可以通过对两个 HRTF 的参数进行内插来确定针对该用户的 HRTF。

[0076] 接着，系统可以执行附加的步骤以验证该 HRTF 确定是良好的，并且可能为该听者选择更好的 HRTF。在步骤 712，系统为该用户播放 3D 音频信号。这可以通过用户佩戴的头戴耳机来播放。在步骤 714，可以要求用户指向 3D 音频信号的明显的源。在一个实施例中，该过程被制成游戏，其中要求用户对着该声音射击。例如，系统在没有任何视觉表示的情况下播放鸭子声音。在步骤 716，系统确定用户所指向的位置。步骤 716 可以包括：使用深度相机系统来收集深度信息。系统还可以向用户要求语音输入，该语音输入可以由该系统使用语音识别来识别。可以针对其他声音重复步骤 712 - 716，直到在步骤 717 确定收集到足够的信息。

[0077] 在步骤 718，系统确定 HRTF 的有效程度。例如，系统确定用户能够多精确地对虚拟声音进行定位。在一个实施例中，如果用户命中声音源（例如用户射中鸭子），则系统在显示器 196 上显示该鸭子。然后，系统确定是否应当为该用户确定不同的 HRTF。如果是，则通过返回到步骤 710 来确定新的 HRTF。过程 700 可以重复步骤 712 - 718，直到找到令人满意的 HRTF。

[0078] 在步骤 722，为用户存储 HRTF。注意，这不一定是在过程 700 中被测试的最后一个 HRTF。也就是说，系统可以确定：之前在过程 700 曾被测试的 HRTF 之一可能是最好的。还应注意，可以为给定用户存储一个以上的 HRTF。例如，可以针对佩戴眼镜和未佩戴眼镜的用户重复过程 700，为每种情况存储一个 HRTF。

[0079] 如所述那样，用于确定听者 8 的详细特性使得可以为该用户存储 HRTF 的这个过程可以不频繁地进行——可能仅仅进行一次。图 8 描绘了用于基于之前收集的详细特性来为听者 8 选择 HRTF 的过程 800 的一个实施例的流程图。例如，过程 700 可以在过程 800 以前执行一次。然后，过程 800 可以执行许多次。在步骤 802，使用生物测定信息来标识出听者 8。注意，该信息不同于在过程 700 收集的信息。然而，可能的是，可能存在信息的一些重叠。收集生物测定信息可以包括收集深度信息和 RGB 信息。在一个实施例中，系统能够例如基于面部识别来识别听者。

[0080] 在步骤 804，为在步骤 802 所标识出的用户选择合适的 HRTF。在一个实施例中，选择在过程 700 为该用户存储的 HRTF。在另一实施例中，详细的用户特性已经在过程 700 被存储。然后，在过程 800，可以基于所存储的详细用户特性来选择 HRTF。如果期望的话，这些所存储的详细用户特性可以被当前所收集的信息扩充例如，用户可能在此时戴着帽子。因此，系统可以在过程 700 期间选择与用户未戴帽子的情况不同的 HRTF。

[0081] 如上所述，用户可能佩戴可收集关于房间的声学数据的一个或多个话筒。图 9 是描绘用于基于这样的数据修改房间模型的过程 900 的一个实施例的流程图。作为一个示例，过程 400 可能已经执行了至少一次以基于深度图像和 / 或 RGB 图像来确定房间模型。然后，过程 900 可以用于修改该房间模型。

[0082] 在步骤 902，通过位于听者所在的房间中的扬声器来播放声音。该步骤可以在任何

时间执行以帮助细化房间模型。在任选步骤 904, 指示用户在播放声音时在房间四周走动。不一定具体地告知用户应走到那里。另一方面, 可以指示用户走到房间四周的不同位置; 然而, 这不是必需的。注意, 步骤 904 是任选的。在一个实施例中, 不是指示用户他们应当在房间四周自动, 而是简单地假定: 用户作为正常玩游戏或其他交互的一部分而将在四周移动。在步骤 906, 在播放声音时收集来自用户所佩戴的一个或多个话筒 31 的数据。用户可以将这些话筒 31 佩戴在其耳朵附近, 但是这不是必需的。在步骤 908, 确定用户的位置并且将该位置与从话筒 31 收集的数据相关。一种选择是使用深度信息和 RGB 信息来对用户进行定位。在步骤 910, 基于在步骤 906 被收集、在步骤 908 与用户位置相关的数据来确定房间声学特性。在步骤 912, 基于在步骤 910 所确定的声学性质来更新房间模型。

[0083] 如所述那样, 过程 400 可以以所期望的频繁程度执行。因此, 一种选择是使用过程 400 构建房间模型。然后, 可以使用过程 900 来更新房间模型。接着, 可以再次使用过程 400 一次或多次来更新 (或重新创建) 房间模型。另一选择是将过程 900 与过程 400 相组合。例如, 所生成的初始房间模型可以基于过程 400 和 900 二者的使用。

[0084] 图 10 描绘了用于生成 3D 音频信号的一个实施例的框图。该框图提供了过程 300 的一个实施例的附加细节。声音源 1002 表示 3D 音频信号所基于的虚拟声音。例如, 该声音源可以是狗叫的 (录制的或计算机生成的) 数字数据。总的来说, 声音源 1002 是沿着若干路径——直接路径和若干反射路径——被处理的。图 5B 中提供了这些路径的一个示例。对于直接路径而言, 通过施加增益和滤波器 1006 来处理声音源 1002; 然后为听者应用 HRTF 1008。对于间接路径而言, 首先为每个反射路径计算 1004 方位角和仰角。然后, 该处理类似于针对直接路径所描述的处理。该结果可以在施加所估计的混响尾音 (reverb tail) 1010 以产生最终的 3D 音频信号以前相加, 该 3D 音频信号可以通过头戴耳机来播放 1012。

[0085] 图 10 的图描绘了可以出于不同原因使用的传感器输入。传感器输入可以用于计算用户的姿态和旋转, 这如框 1020 中所示。传感器输入可以用于构建房间模型, 以及用于估计房间材料, 这如框 1030 所示。最后, 传感器输入可以用于确定用户特性, 比如耳廓和头部特性, 这如框 1040 所示。这些用户特性可以用于为该用户确定 HRTF。注意, 针对用户的 HRTF 不要求是来自库的 HRTF。例如, 可以使用内插来从库中的两个或更多个 HRTF 中形成 HRTF。

[0086] 框 1020 的传感器输入 (其用于计算用户姿态) 可以包括、但不限于深度信息、RGB 信息、以及惯性数据 (其来自在用户上的惯性传感器)。框 1030 的传感器输入 (其用于计算房间模型) 可以包括、但不限于深度信息、RGB 信息、以及声学数据 (其例如来自用户佩戴的话筒)。框 1040 的传感器输入 (其用于确定 HRTF) 可以包括、但不限于深度信息和 RGB 信息。

[0087] 为了计算反射的方位角和仰角, 可以使用来自框 1020 和 1030 的数据。类似地, 增益和滤波器可以使用来自框 1020 和 1030 的数据。注意, 该传感器数据可以在任何时间更新。例如, 用户可能移动, 使得捕捉用户位置的传感器数据非常频繁地变化。这些改变例如可以被馈送给方位角和仰角计算 1004, 使得不断地因改变的用户姿态而更新 3D 音频信号。类似地, 用户姿态的改变可以实时地馈送给增益和滤波器 1006。在一些实施例中, 针对用户的 HRTF 不是实时地更新。然而, 实时地更新针对用户的 HRTF 是一种选择。

[0088] 在接近生成 3D 音频信号结束时添加的混响尾音可以基于房间模型和对材料的估计。因此,框 1030 可以是用于估计混响尾音 1010 的输入。在一个实施例中,系统存储与诸如房间大小和材料之类的因素相关的混响尾音的库。系统能够基于房间模型选择混响尾音之一。系统还可以在两个所存储的混响尾音之间内插。因此,通过选择所存储的混响尾音,节省了计算时间。

[0089] 图 11 描绘了可用于生成 3D 音频信号的计算环境的示例性框图。该计算环境可以用在图 1 的运动捕捉系统中。上文所描述的诸如计算环境 12 等的计算环境可包括诸如游戏控制台等的多媒体控制台 100。

[0090] 控制台 100 可以从图 2 的深度相机系统 20 接收输入。控制台还可以从话筒 31 和惯性传感器 38 接收输入,所述话筒 31 和惯性传感器 38 二者都可以由用户佩戴。控制台 100 可以将 3D 音频信号输出给音频放大器 197。

[0091] 多媒体控制台 100 包括具有 1 级高速缓存 102、2 级高速缓存 104 和闪存 ROM(只读存储器)106 的中央处理单元(CPU)101。一级高速缓存 102 和二级高速缓存 104 临时存储数据并因此减少存储器访问周期数,由此改进处理速度和吞吐量。CPU101 可以设置成具有一个以上的内核,以及由此的附加的一级和二级高速缓存 102 和 104。诸如闪存 ROM 之类的存储器 106 可存储当多媒体控制台 100 通电时在引导过程的初始阶段期间加载的可执行代码。

[0092] 图形处理单元(GPU)108 和视频编码器/视频编解码器(编码器/解码器)114 形成用于高速和高分辨率图形处理的视频处理流水线。经由总线从图形处理单元 108 向视频编码器/视频编解码器 114 运送数据。视频处理流水线向 A/V(音频/视频)端口 140 输出数据,用于传输至电视或其他显示器。存储器控制器 110 连接到 GPU108,以便于处理器对各种类型的存储器 112,比如 RAM(随机存取存储器)的处理器访问。A/V 端口 140 可以连接到显示器 196。

[0093] 多媒体控制台 100 包括可在模块 118 上实现的 I/O 控制器 120、系统管理控制器 122、音频处理单元 123、网络接口 124、第一 USB 主控制器 126、第二 USB 控制器 128 和前面板 I/O 子部件 130。USB 控制器 126 和 128 用作外围控制器 142(1)-142(2)、无线适配器 148、和外置存储器设备 146(例如闪存、外置 CD/DVD ROM 驱动器、可移动介质等)的主机。网络接口(NW IF)124 和/或无线适配器 148 提供对网络(例如,因特网、家庭网络等)的访问并且可以是包括以太网卡、调制解调器、蓝牙模块、电缆调制解调器等各种不同的有线或无线适配器组件中任何一种。

[0094] 提供系统存储器 143 来存储在引导过程期间加载的应用数据。提供了媒体驱动器 144,其可以包括 DVD/CD 驱动器、硬盘驱动器、或其他可移动媒体驱动器。介质驱动器 144 可以在多媒体控制台 100 的内部或外部。应用数据可经由介质驱动器 144 访问,以由多媒体控制台 100 执行、回放等。媒体驱动器 144 经由诸如串行 ATA 总线或其他高速连接等总线连接到 I/O 控制器 120。

[0095] 系统管理控制器 122 提供涉及确保多媒体控制台 100 的可用性的各种服务功能。音频处理单元 123 和音频编解码器 132 形成具有高保真度和立体声处理的对应的音频处理流水线。音频数据经由通信链路在音频处理单元 123 与音频编解码器 132 之间传输。音频处理流水线将数据输出到 A/V 端口 140 以供外置音频播放器或具有音频能力的设备再现。

在一些实施例中,3D 音频信号通过 A/V 端口 140 来提供,但是 3D 音频信号可以通过不同的连接来提供。

[0096] 前面板 I/O 子部件 130 支持暴露多媒体控制台 100 的外表面上的电源按钮 150 和弹出按钮 152 以及任何 LED(发光二极管)或其他指示器的功能。系统供电模块 136 向多媒体控制台 100 的组件供电。风扇 138 冷却多媒体控制台 100 内的电路。

[0097] CPU101、GPU108、存储器控制器 110、和多媒体控制台 100 内的各个其他组件经由一条或多条总线互连,包括串行和并行总线、存储器总线、外围总线、和使用各种总线架构中任一种的处理器或局部总线。

[0098] 当多媒体控制台 100 通电时,应用数据可从系统存储器 143 加载到存储器 112 和/或高速缓存 102、104 中并在 CPU101 上执行。应用可呈现在导航到多媒体控制台 100 上可用的不同媒体类型时提供一致的用户体验的图形用户界面。在操作中,介质驱动器 144 中包含的应用和/或其他媒体可从介质驱动器 144 启动或播放,以向多媒体控制台 100 提供附加功能。

[0099] 多媒体控制台 100 可通过将该系统连接到电视机或其他显示器而作为独立系统来操作。在该独立模式中,多媒体控制台 100 允许一个或多个用户与该系统交互、看电影、或听音乐。然而,随着通过网络接口 124 或无线适配器 148 可用的宽带连接的集成,多媒体控制台 100 还可作为较大网络社区中的参与者来操作。

[0100] 当多媒体控制台 100 通电时,可以保留指定量的硬件资源以供多媒体控制台操作系统作系统使用。这些资源可以包括存储器保留(例如,16MB)、CPU 和 GPU 周期(例如,5%)、网络带宽(例如,8kbs)等。因为这些资源是在系统引导时保留的,所以所保留的资源对应用程序而言是不存在的。

[0101] 具体地,存储器保留可以是足够大以包含启动内核、并发系统应用和驱动程序。CPU 保留可以是恒定的,使得若所保留的 CPU 使用不被系统应用使用,则空闲线程将消耗任何未使用的周期。

[0102] 对于 GPU 保留,通过使用 GPU 中断来显示由系统应用生成的轻量消息(例如,弹出窗口),以调度代码来将弹出窗口呈现为覆盖图。覆盖图所需的存储器量取决于覆盖区域大小,并且覆盖图可与屏幕分辨率成比例缩放。在并发系统应用使用完整用户界面的情况下,优选使用独立于应用分辨率的分辨率。定标器可用于设置该分辨率,从而无需改变频率并引起 TV 重新同步。

[0103] 在多媒体控制台 100 引导且系统资源被保留之后,就执行并发系统应用来提供系统功能。系统功能被封装在上述所保留的系统资源中执行的一组系统应用中。操作系统内核标识了是系统应用线程而非游戏应用线程的线程。系统应用可被调度为在预定时间并以预定时间间隔在 CPU101 上运行,来为应用提供一致的系统资源视图。进行调度是为了把对于控制台上运行的游戏应用的高速缓存分裂最小化。

[0104] 当并发系统应用需要音频时,则由于时间敏感性而异步调度音频处理给游戏应用。多媒体控制台应用管理器(如下所述)在系统应用活动时控制游戏应用的音频水平(例如,静音、衰减)。

[0105] 输入设备(例如,控制器 142(1) 和 142(2)) 由游戏应用和系统应用共享。输入设备不是所保留的资源,但却在系统应用和游戏应用之间切换以使其各自具有设备的焦点。

应用管理器可控制输入流的切换,而无需知晓游戏应用的知识,并且驱动程序维持有关焦点切换的状态信息。

[0106] 图 12 描绘了可用于提供 3D 音频信号的计算环境的另一示例性框图。计算环境可以从图 2 的深度相机系统 20 接收输入。计算环境还可以从话筒 31 和惯性传感器 38 接收输入,所述话筒 31 和惯性传感器 38 二者都可以由用户佩戴。计算环境可以将 3D 音频信号输出给头戴耳机 27。

[0107] 计算环境 220 包括通常包括各种有形计算机可读存储介质的计算机 241。这可以是能由计算机 241 访问的任何可用介质,而且包含易失性和非易失性介质、可移动和不可移动介质。系统存储器 222 包括易失性和 / 或非易失性存储器形式的计算机存储介质,如只读存储器 (ROM) 223 和随机存取存储器 (RAM) 260。基本输入 / 输出系统 224 (BIOS) 包括如在启动时帮助在计算机 241 内的元件之间传输信息的基本例程,它通常储存在 ROM 223 中。RAM 260 通常包含处理单元 259 可以立即访问和 / 或目前正在操作的数据和 / 或程序模块。图形接口 231 与 GPU 229 进行通信。作为示例而非局限,图 12 描绘了操作系统 225、应用程序 226、其他程序模块 227 和程序数据 228。

[0108] 计算机 241 也可以包括其他可移动 / 不可移动、易失性 / 非易失性计算机存储介质,例如,读写不可移动、非易失性磁性介质的硬盘驱动器 238,读写可移动、非易失性磁盘 254 的磁盘驱动器 239,以及读写诸如 CD ROM 或其他光学介质之类的可移动、非易失性光盘 253 的光盘驱动器 240。可以在该示例性操作环境中使用的其他可移动 / 不可移动、易失性 / 非易失性有形计算机可读存储介质包括但不限于,磁带盒、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 238 通常由例如接口 234 等不可移动存储器接口连接至系统总线 221,而磁盘驱动器 239 和光盘驱动器 240 通常由例如接口 235 等可移动存储器接口连接至系统总线 221。

[0109] 以上讨论并在图 12 中描绘的驱动器及其相关联的计算机存储介质为计算机 241 提供了对计算机可读指令、数据结构、程序模块和其他数据的存储。例如,硬盘驱动器 238 被描绘为存储了操作系统 258、应用程序 257、其他程序模块 256、以及程序数据 255。注意,这些组件可以与操作系统 225、应用程序 226、其他程序模块 227 和程序数据 228 相同,也可以与它们不同。在此给操作系统 258、应用程序 257、其他程序模块 256、以及程序数据 255 提供了不同的编号,以说明至少它们是不同的副本。用户可以通过诸如键盘 251 和定点设备 252 (通常被称为鼠标、跟踪球或触摸垫) 之类的输入设备向计算机 241 输入命令和信息。其他输入设备 (未示出) 可包括话筒、游戏杆、游戏手柄、圆盘式卫星天线、扫描仪等。这些和其他输入设备通常由耦合至系统总线的用户输入接口 236 连接至处理单元 259,但也可以由诸如并行端口、游戏端口或通用串行总线 (USB) 等其他接口和总线结构来进行连接。图 2 的包括相机 28 的深度相机系统 20 可以定义附加的输入设备。显示器 196 也经由诸如视频接口 232 之类的接口连接至系统总线 221。除监视器之外,计算机还可以包括可以通过输出外围接口 233 连接的诸如头戴耳机 27 和打印机 243 之类的其他外围输出设备。

[0110] 计算机 241 可使用至诸如远程计算机 246 之类的一个或多个远程计算机的逻辑连接在互联网环境中操作。远程计算机 246 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其他常见的网络节点,且通常包括许多或所有以上相对于计算机 241 描述的元件,但是在图 12 中仅示出了存储器存储设备 247。逻辑连接包括局域网 (LAN) 245 和广域网

(WAN) 249,但也可以包括其他网络。这样的联网环境在办公室、企业范围计算机网络、内联网和因特网中是常见的。

[0111] 当在 LAN 联网环境中使用时,计算机 241 通过网络接口或适配器 245 连接至 LAN237。当在 WAN 联网环境中使用时,计算机 241 通常包括调制解调器 250 或用于通过诸如因特网等 WAN249 建立通信的其他手段。调制解调器 250 可以是内置或外置的,它可以经由用户输入接口 236 或其他适当的机制连接至系统总线 221。在联网环境中,相对于计算机 241 所描述的模块或其部分可被存储在远程存储器存储设备中。作为示例而非限制,图 12 描绘了远程应用程序 248 驻留在存储器设备 247 上。应当理解,所示的网络连接是示例性的,并且可使用在计算机之间建立通信链路的其他手段。

[0112] 前面的对本技术的详细描述只是为了说明和描述。它不是为了详尽的解释或将本技术限制在所公开的准确的形式。鉴于上述教导,许多修改和变型都是可能的。所描述的实施例只是为了最好地说明本技术的原理以及其实际应用,从而使精通本技术的其他人在各种实施例中最佳地利用本技术,适合于特定用途的各种修改也是可以的。本技术的范围由所附的权利要求进行定义。

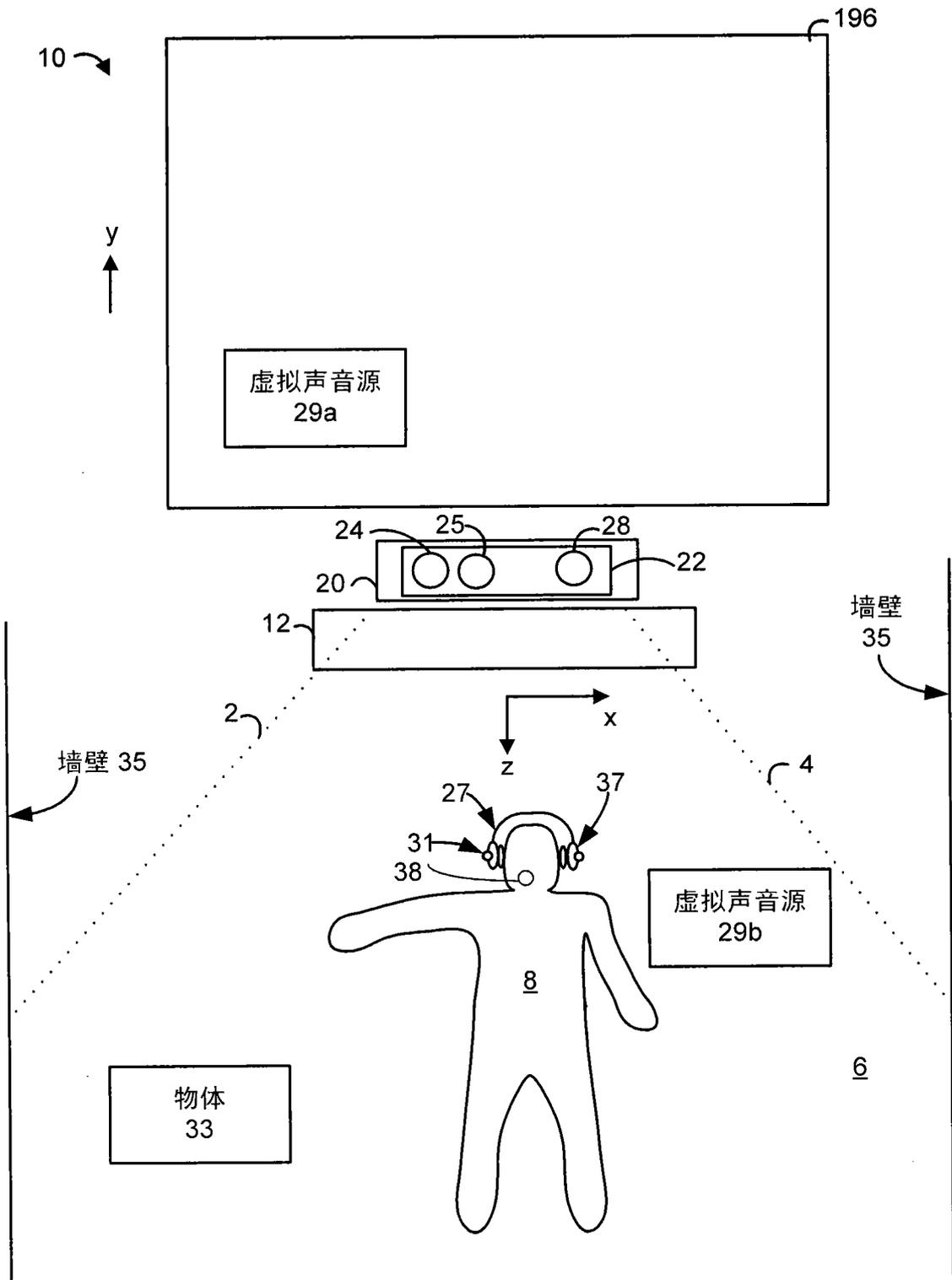


图 1

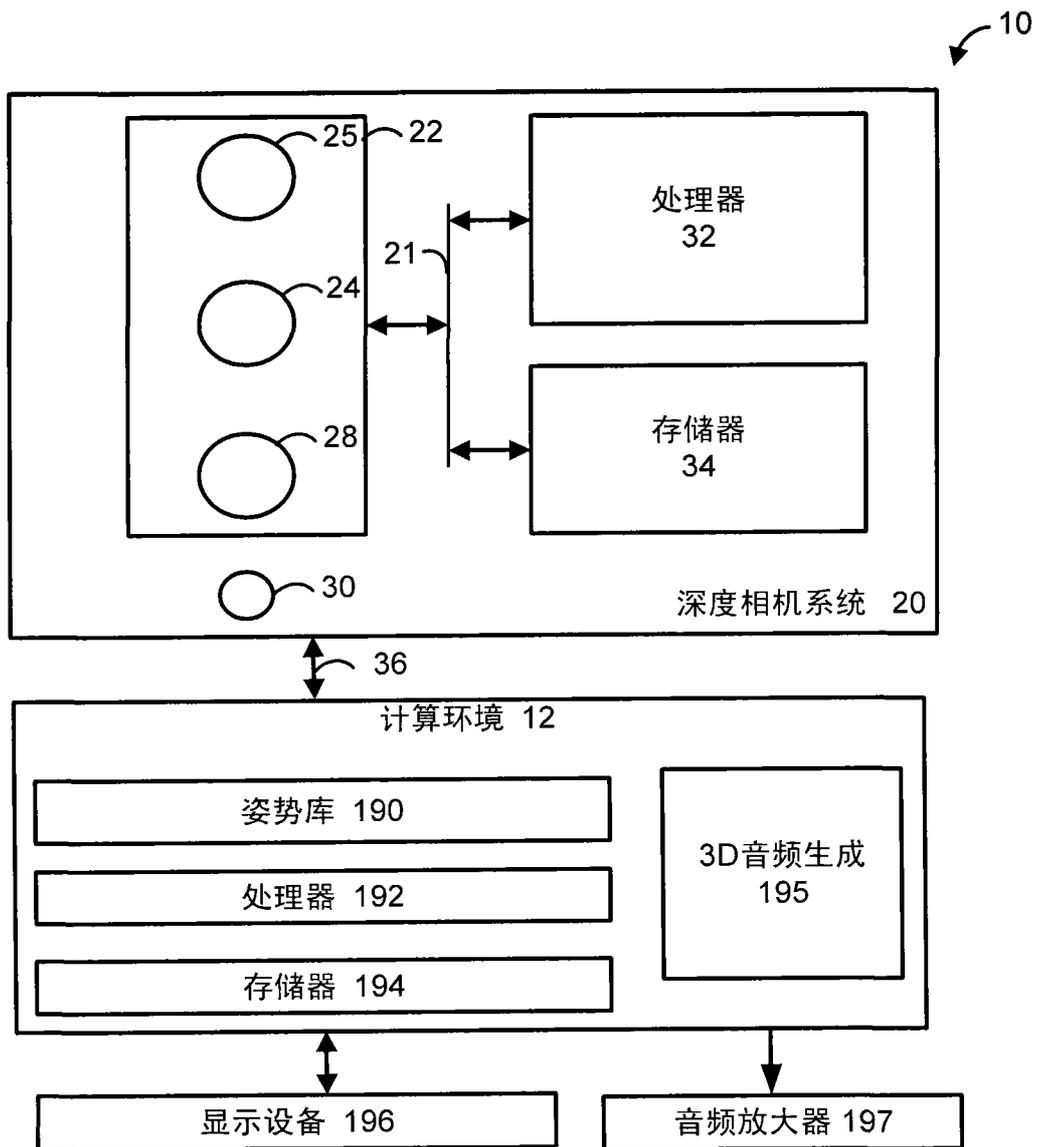


图 2

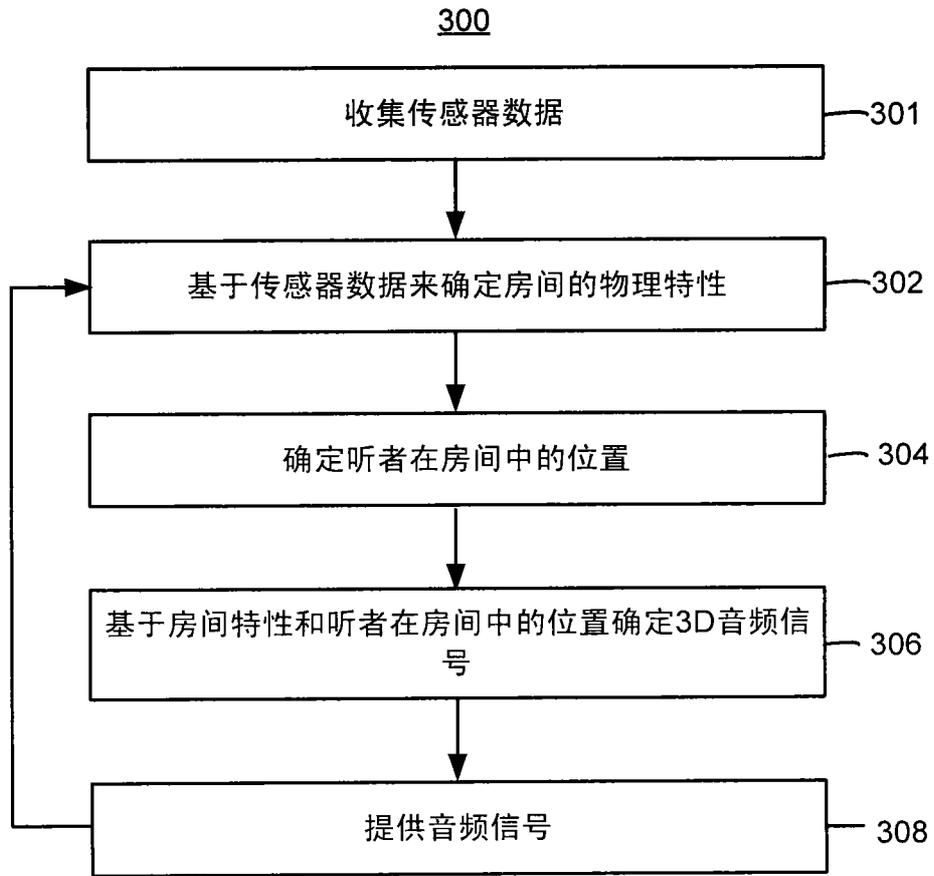


图 3

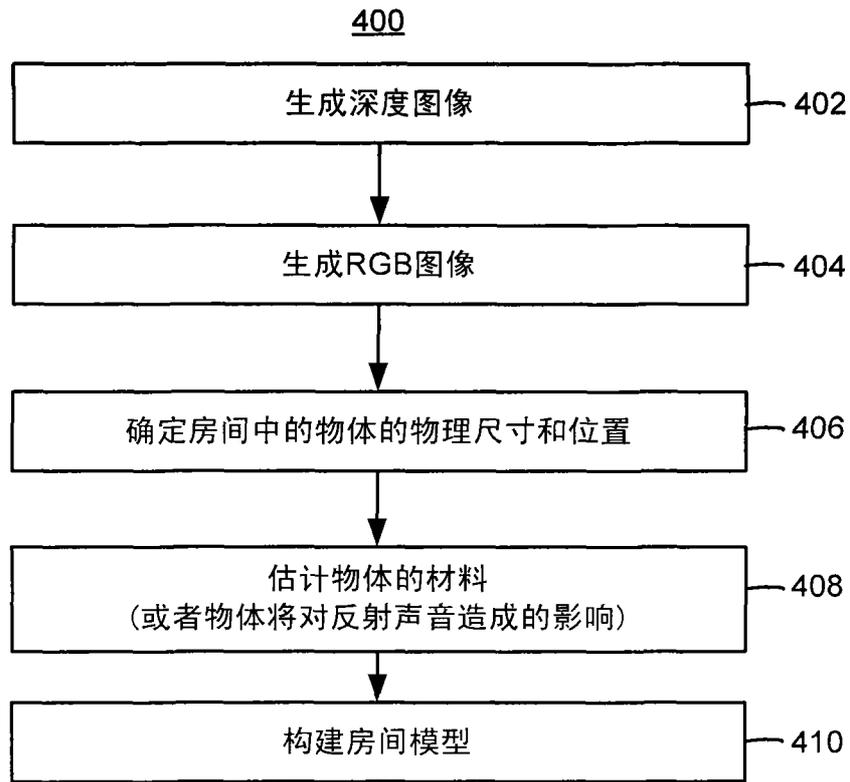


图 4A

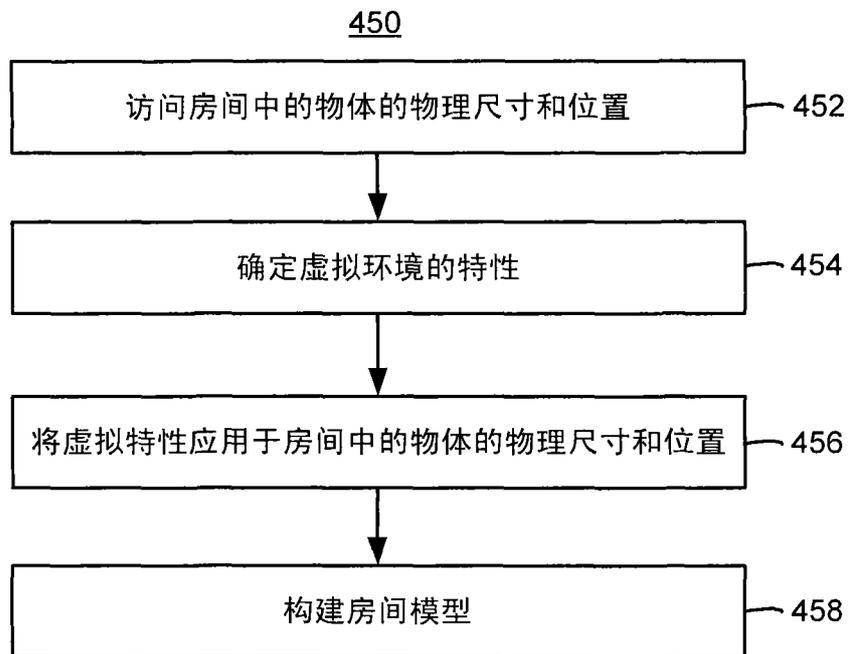


图 4B

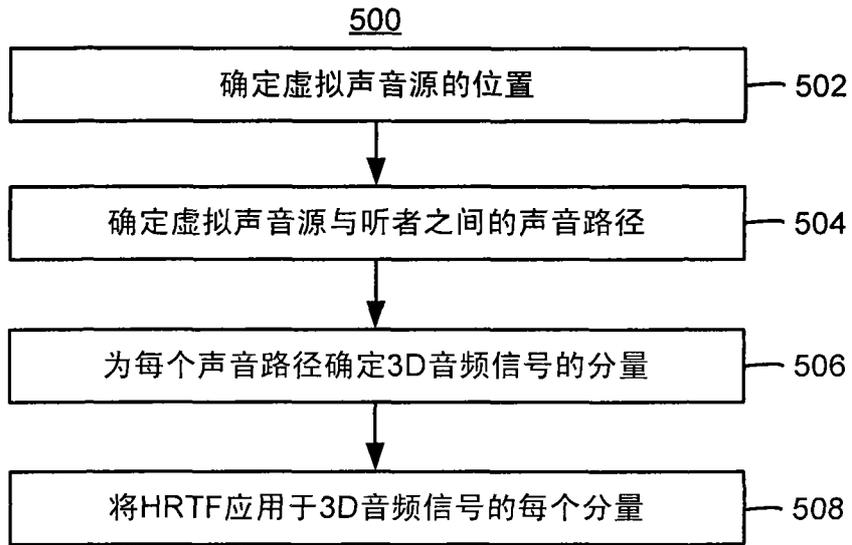


图 5A

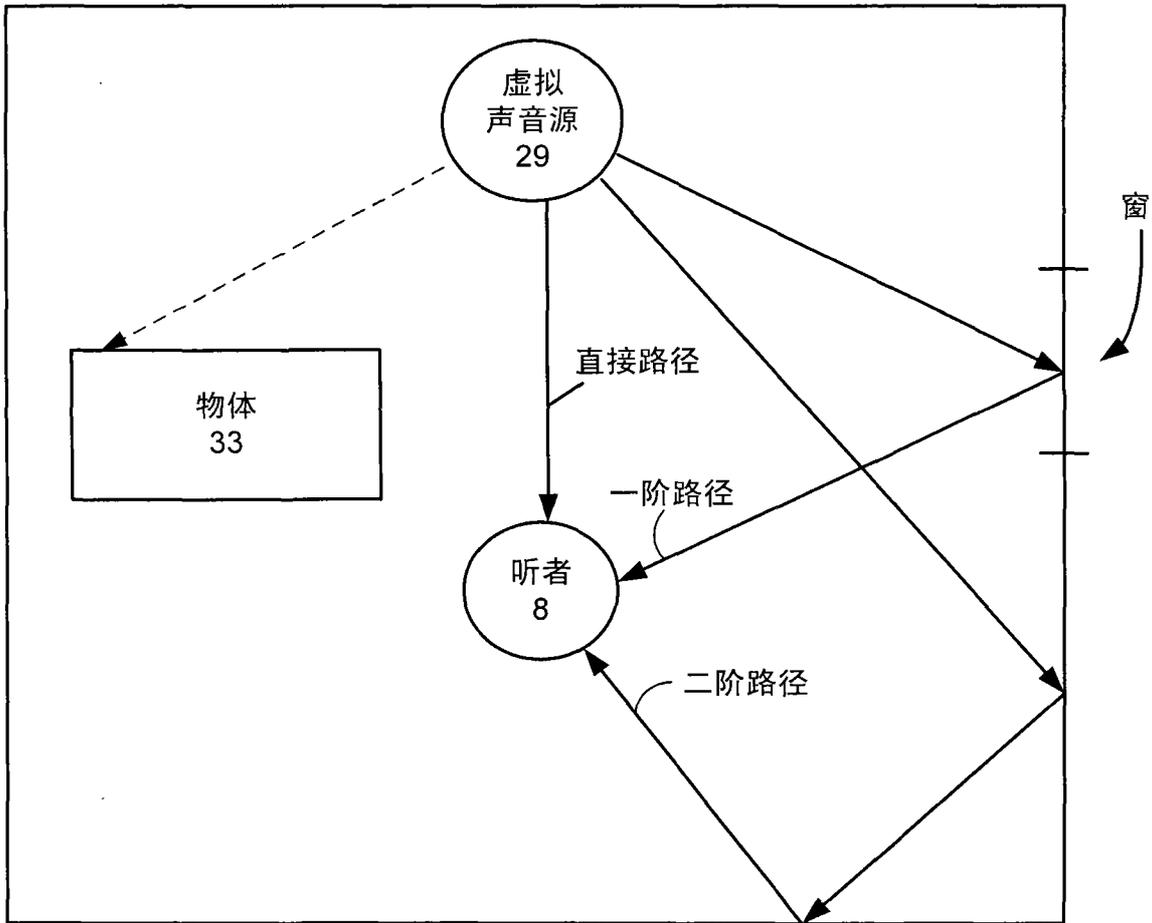


图 5B

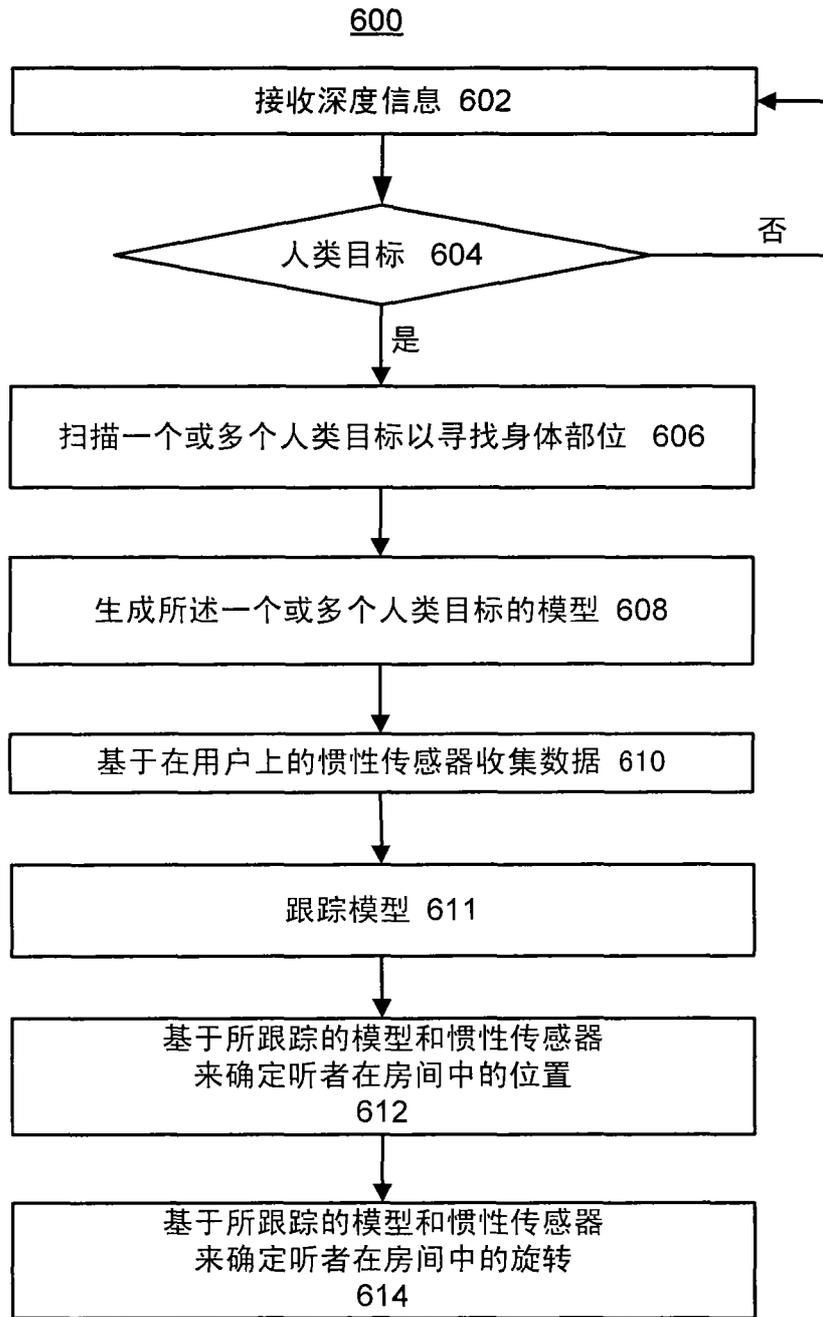


图 6

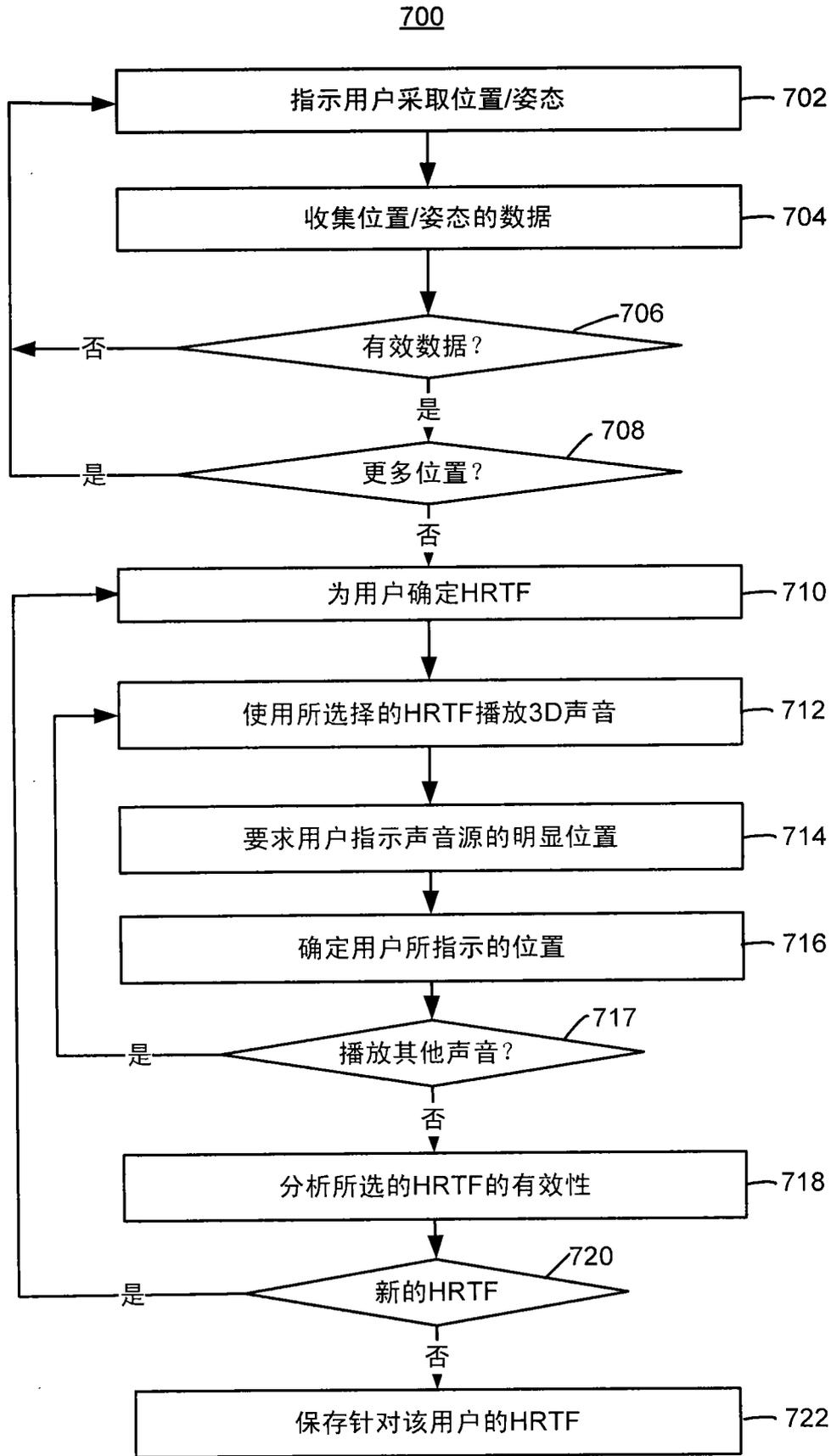


图 7

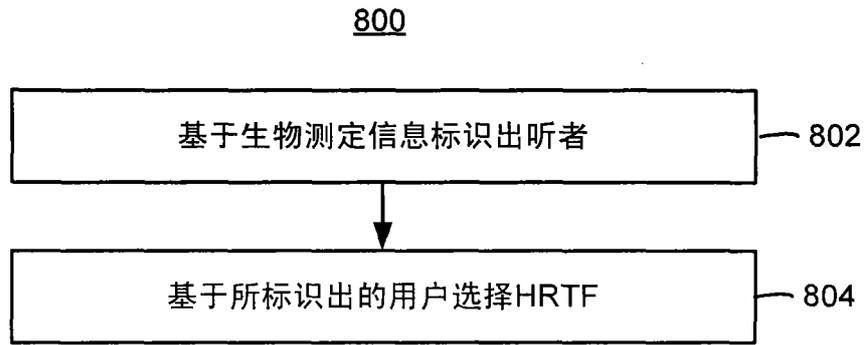


图 8

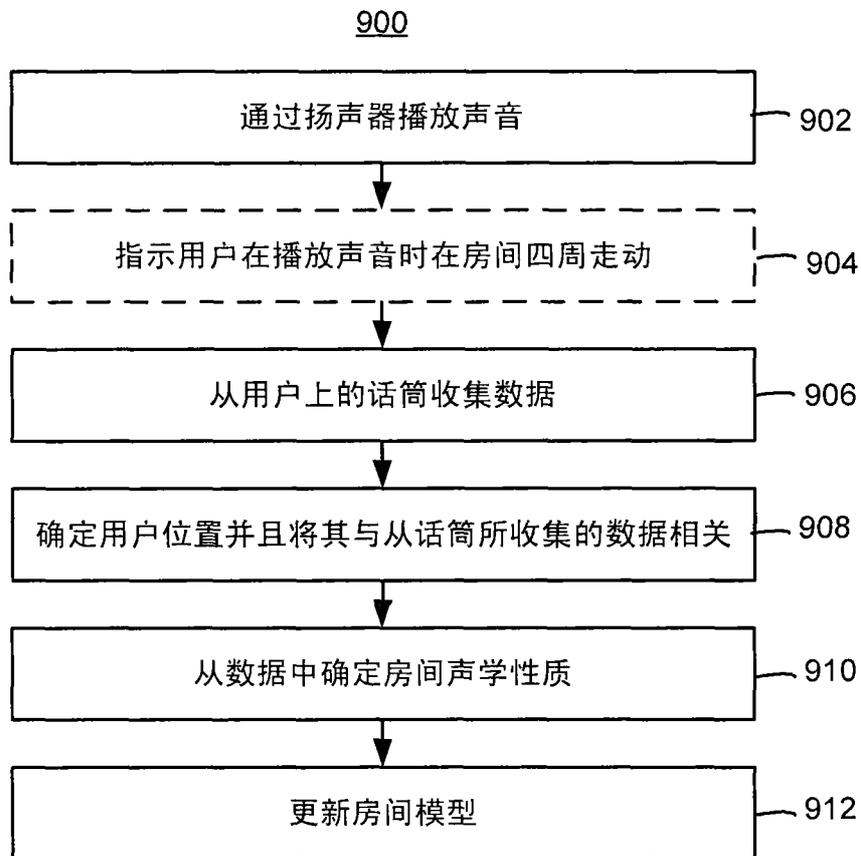


图 9

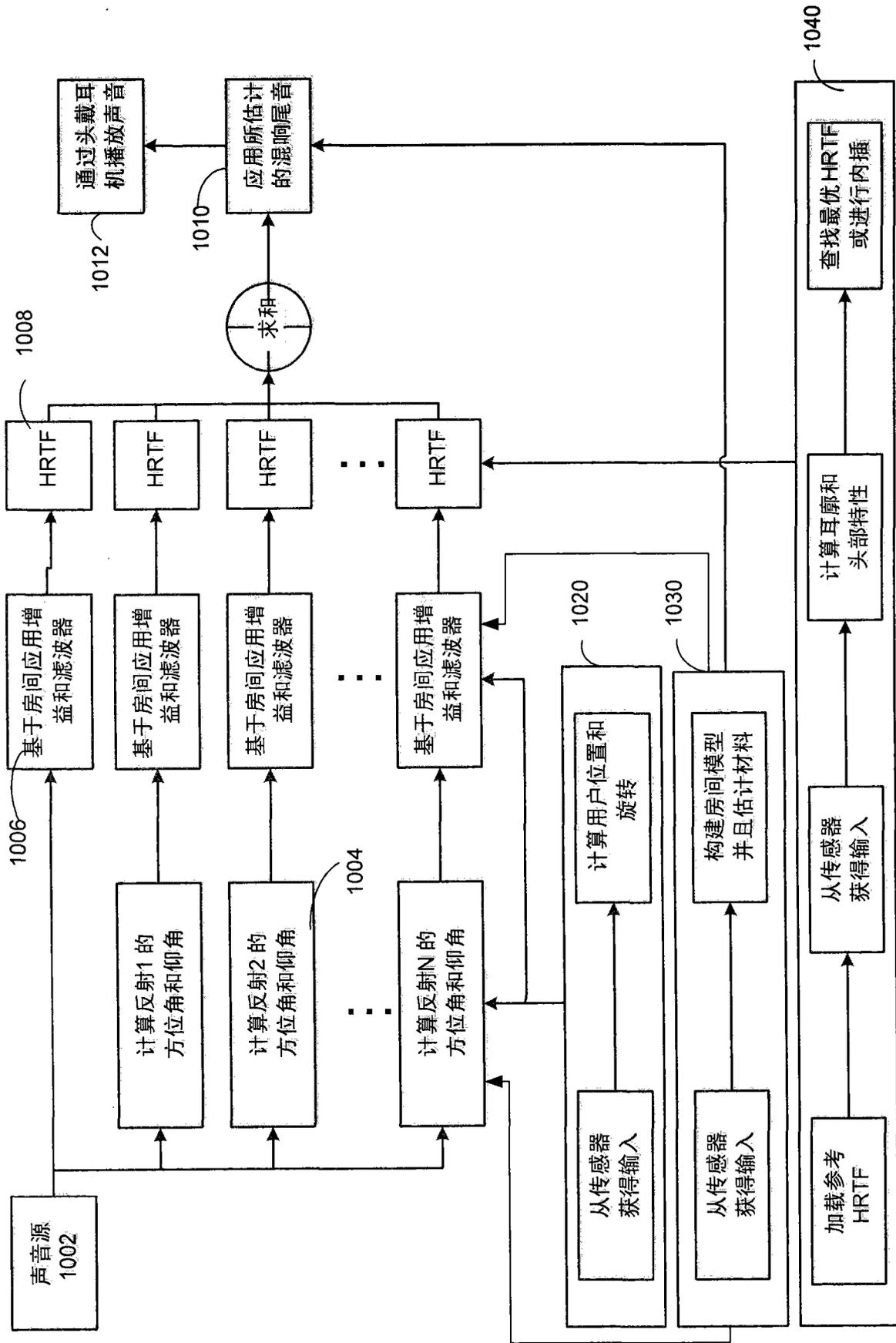


图 10

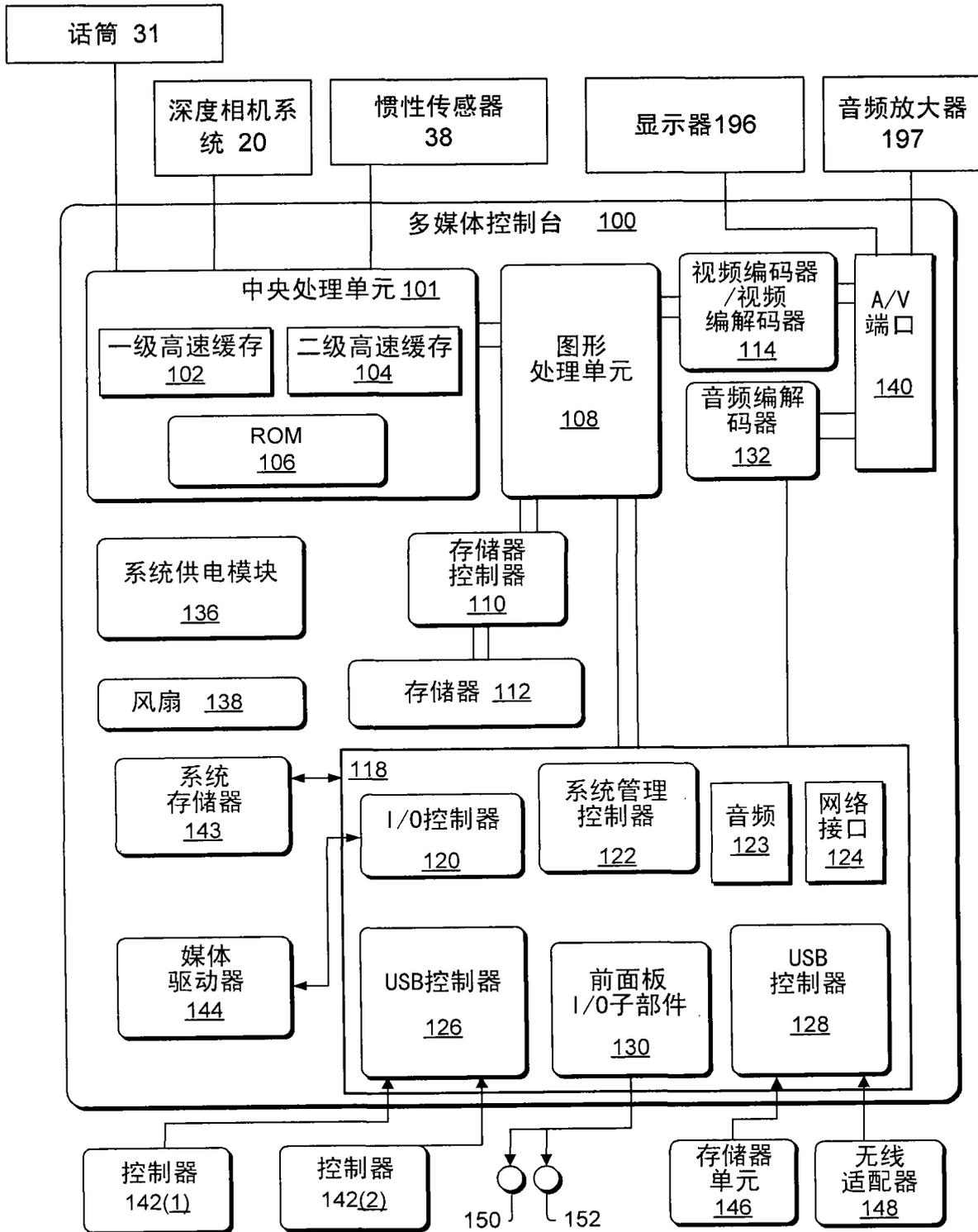


图 11

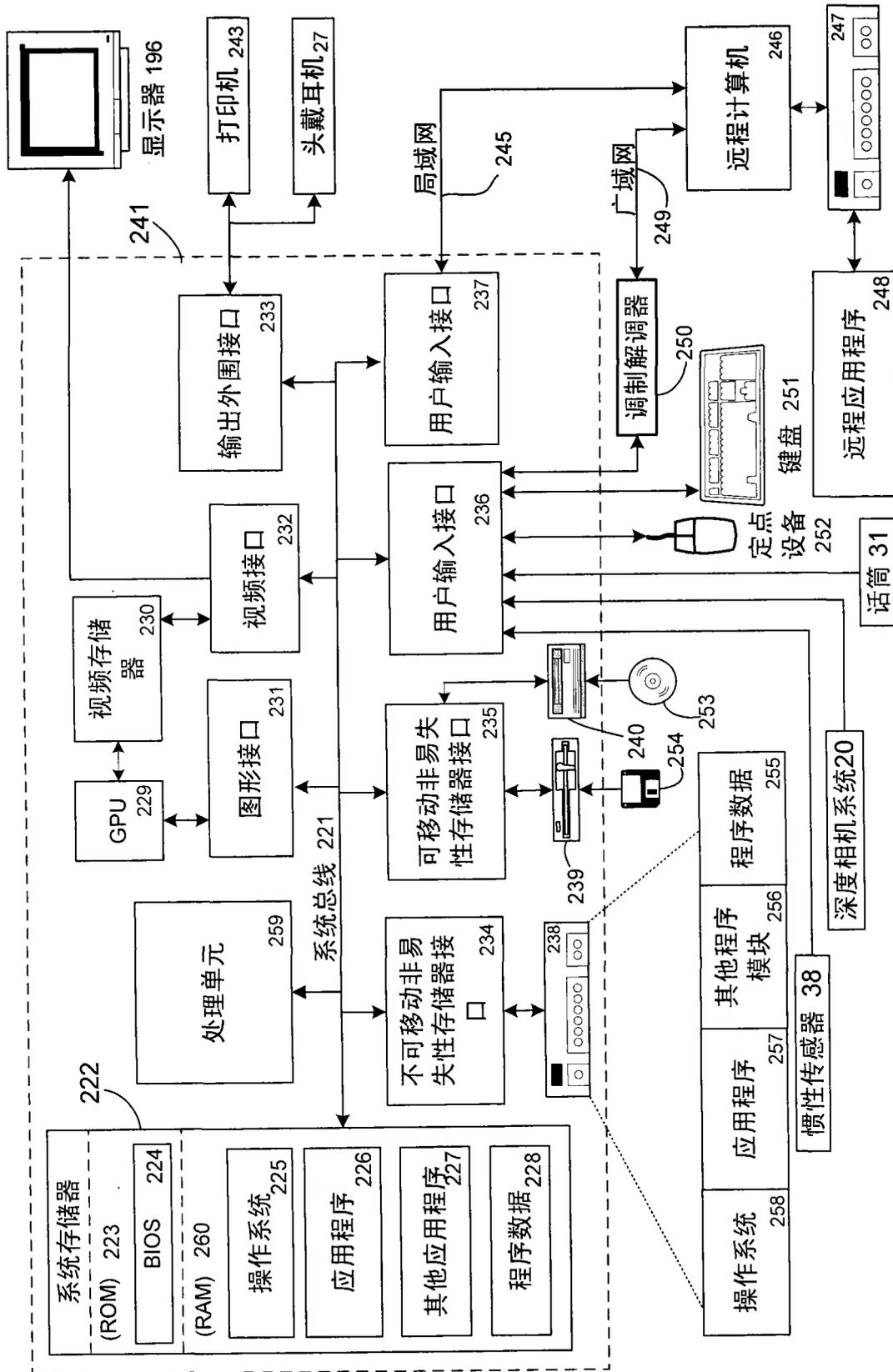


图 12