(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116348950 A
(43)申请公布日 2023.06.27

(21)申请号 202180071279.0
(22)申请日 2021.11.10
(30)优先权数据
17/102,687 2020.11.24 US
(85)PCT国际申请进入国家阶段日
2023.04.18
(86)PCT国际申请的申请数据
PCT/CN2021/129740 2021.11.10
(87)PCT国际申请的公布数据
WO2022/111282 EN 2022.06.02
(71)申请人 国际商业机器公司
地址 美国纽约阿芒克
(72)发明人 C·德克罗普 T·阿格拉瓦尔 J·R·福克斯 S·K·拉克希特
(74)专利代理机构 北京市柳沈律师事务所 11105
专利代理师 陈金林
(51)Int.Cl.
G10L 15/22(2006.01)
G10L 15/26(2006.01)
G10L 21/10(2013.01)
G10L 21/18(2013.01)

权利要求书2页 说明书10页 附图15页

(54)发明名称

在执行任何语音命令时从周围环境进行基于AR(增强现实)的选择性声音包括

(57)摘要

一种生成增强语音命令的方法、系统和装置,包括:识别从相应的多个换能器到智能扬声器设备的多个声音,使用增强现实设备生成声音的可视化,其中,能够使用可视化来选择声音中的一个或多个,以及生成用于智能扬声器设备的增强语音命令,其中,增强语音命令包括使用增强现实设备的可视化来选择的一个或多个声音。



CN 116348950 A

1.一种生成增强语音命令的方法,包括:

识别从相应的多个换能器到智能设备的多个声音;

使用增强现实设备生成声音的可视化,其中,能够使用所述可视化来选择声音中的一个或多个;以及

生成用于智能扬声器设备的所述增强语音命令,其中,所述增强语音命令包括使用所述增强现实设备的所述可视化来选择的所述一个或多个声音。

2.根据权利要求1所述的方法,还包括基于来自所述换能器中的一个或多个的历史声音的选择和取消选择历史来自动地选择所述一个或多个声音。

3.根据权利要求1所述的方法,其中,所述多个换能器中的每一个包括扬声器,并且所述智能设备包括智能扬声器设备,并且

其中,所述声音包括在增强现实空间中选择的口语内容或非口语内容。

4.根据权利要求1所述的方法,还包括基于存储在存储器中的历史信息的选择和取消选择历史来选择所述一个或多个声音。

5.根据权利要求1所述的方法,还包括使用增强现实设备将所述语音命令转译成所述声音的可视化的生成。

6.根据权利要求1所述的方法,还包括通过由增强现实系统对一个或多个声音的选择来执行所述增强语音命令。

7.根据权利要求1所述的方法,还包括:

在提交语音命令时或在提交语音命令之后,选择在所述增强现实设备中可视化的、并且相应地被考虑在所述增强语音命令中以用于执行的一个或多个口语内容;以及

基于保存在历史语料库中的来自所述换能器中的一个或多个的历史声音的选择和取消选择历史来选择所述一个或多个声音。

8.根据权利要求1所述的方法,在云基础设施中实现。

9.一种生成增强语音命令的系统,包括:

存储器,存储计算机指令;以及

处理器,被配置为执行所述计算机指令以:

识别从相应的多个换能器到智能扬声器设备的多个声音;

使用增强现实设备生成声音的可视化,其中,能够使用所述可视化来选择声音中的一个或多个;以及

生成用于所述智能扬声器设备的所述增强语音命令,其中,所述增强语音命令包括使用所述增强现实设备的所述可视化来选择的所述一个或多个声音。

10.根据权利要求8所述的系统,还包括基于来自所述换能器中的一个或多个的历史声音的选择和取消选择历史来自动地选择所述一个或多个声音。

11.根据权利要求8所述的系统,其中,所述多个换能器中的每一个包括扬声器,并且所述智能设备包括智能扬声器设备,并且

其中,所述声音包括在增强现实空间中选择的口语内容或非口语内容。

12.根据权利要求8所述的系统,还包括基于存储在所述存储器中的历史信息的选择和取消选择历史来选择所述一个或多个声音。

13.根据权利要求8所述的系统,还包括使用增强现实设备将所述语音命令转译成所述

声音的可视化的生成。

14.根据权利要求8所述的系统,还包括:

通过由增强现实系统对一个或多个声音的选择来执行所述增强语音命令;

在提交语音命令时或在提交语音命令之后,选择在所述增强现实设备中可视化的、并且相应地被考虑在所述增强语音命令中以用于执行的一个或多个口语内容;以及

基于保存在历史语料库中的来自所述换能器中的一个或多个的历史声音的选择和取消选择历史来选择所述一个或多个声音。

15.一种计算机程序产品,包括具有随其体现的程序指令的计算机可读存储介质,所述程序指令能够由计算机读取和执行以使所述计算机执行一种方法,包括:

识别从相应的多个换能器到智能扬声器设备的多个声音;

使用增强现实设备生成声音的可视化,其中,能够使用所述可视化来选择声音中的一个或多个;以及

生成用于所述智能扬声器设备的所述增强语音命令,其中,所述增强语音命令包括使用所述增强现实设备的所述可视化来选择的所述一个或多个声音。

16.根据权利要求15所述的计算机程序产品,还包括基于来自所述换能器中的一个或多个的历史声音的选择和取消选择历史来自动地选择所述一个或多个声音。

17.根据权利要求15所述的计算机程序产品,其中,所述多个换能器中的每一个包括扬声器,并且所述智能设备包括智能扬声器设备,并且

其中,所述声音包括在增强现实空间中选择的口语内容或非口语内容。

18.根据权利要求15所述的计算机程序产品,还包括基于存储在存储器中的历史信息的选择和取消选择历史来选择所述一个或多个声音。

19.根据权利要求15所述的计算机程序产品,还包括使用增强现实设备将所述语音命令转译成所述声音的可视化的生成。

20.根据权利要求15所述的计算机程序产品,还包括:

通过由增强现实系统对一个或多个声音的选择来执行所述增强语音命令;

在提交语音命令时或在提交语音命令之后,选择在所述增强现实设备中可视化的、并且相应地被考虑在所述增强语音命令中以用于执行的一个或多个口语内容;以及

基于保存在历史语料库中的来自所述换能器中的一个或多个的历史声音的选择和取消选择历史来选择所述一个或多个声音。

在执行任何语音命令时从周围环境进行基于AR(增强现实)的选择性声音包括

背景技术

[0001] 本发明涉及用于选择性声音包括的方法、装置和系统的实施例,并且更具体地但非限制性地,涉及用于在执行任何语音命令时从周围环境进行基于AR(增强现实)的选择性声音包括的方法、装置和系统。

[0002] 像AMAZON ALEXA、GOOGLE HOME等AI(人工智能)语音辅助系统正变得非常流行。例如,全球语音辅助市场被设置为达到空前的水平。其中一个驱动力包括对改进的客户体验的增加的需求,另一个驱动力是医疗保健市场中增加的使用情况(患者参与)。

[0003] 在这样的系统中,用户可以提交语音命令,并且相应地语音命令被执行。

[0004] 多人可以向单个AI语音辅助系统提交语音命令,并且AI语音辅助系统可以识别语音并且可以执行用户特定的语音命令。

[0005] 然而,存在用户提交任何语音命令并且周围环境的其他用户向第一用户的语音命令提供进一步建议或附加反馈的情况。当这种情况发生时,原始用户可能不想在他的语音命令中接受其他用户的一些建议。目前,AI语音辅助系统不能够理解它应当执行哪个语音命令,以及它可以忽略哪个命令。

[0006] 因此,需要一种能够处理更复杂输入的AI语音辅助系统。例如,需要一种方法和系统,通过该方法和系统,用户可以在其提交语音命令时选择性地识别来自周围环境的哪些口语内容或非口语内容应当被考虑。

发明内容

[0007] 鉴于上述背景技术的上述问题和其它问题、缺点和不足,所公开的发明的示例性方面提供了一种用于在执行任何语音命令时从周围环境进行基于AR的选择性声音包括的方法、装置和系统。

[0008] 根据本发明的实施例,一种生成增强语音命令的方法包括:识别从相应的多个换能器到智能扬声器设备的多个声音;使用增强现实设备生成声音的可视化,其中,能够使用可视化来选择声音中的一个或多个;以及生成用于智能扬声器设备的增强语音命令,其中,增强语音命令包括使用增强现实设备的可视化来选择的一个或多个声音。

[0009] 根据本发明的另一实施例,一种生成增强语音命令的系统包括存储计算机指令的存储器,以及处理器,该处理器被配置为执行计算机指令以识别从相应的多个换能器到智能扬声器设备的多个声音,使用增强现实设备生成声音的可视化,其中,能够使用可视化来选择声音中的一个或多个,以及生成用于智能扬声器设备的增强语音命令,其中,增强语音命令包括使用增强现实设备的可视化来选择的一个或多个声音。

[0010] 根据本发明的又一实施例,一种包括具有随其体现的程序指令的计算机可读存储介质的计算机程序产品,包括一种包括随其体现的程序指令的计算机可读存储介质的计算机程序产品,所述程序指令能够由计算机读取及执行以使所述计算机执行一种方法,包括识别从相应的多个换能器到智能扬声器设备的多个声音,使用增强现实设备生成声音的可

视化。能够使用可视化来选择声音中的一个或多个，以及生成用于智能扬声器设备的增强语音命令，其中，增强语音命令包括使用增强现实设备的可视化来选择的一个或多个声音。

[0011] 因此，已经相当广泛地概述了本发明的某些实施例，以便可以更好地理解本发明的具体实施方式，并且以便可以更好地理解本发明对现有技术的贡献。当然，本发明的另外的实施例将在下面描述，并且其将形成所附权利要求的主题。

[0012] 应当理解，本发明在其应用中不限于以下描述中阐述的或附图中示出的构造细节和部件布置。本发明能够具有除所描述的那些实施例之外的实施例，并且能够以各种方式实践和执行。而且，应当理解，本文所使用的措辞和术语以及摘要是为了描述的目的，而不应当被认为是限制。

[0013] 因此，本领域技术人员将理解，本公开所基于的概念可以容易地用作设计用于执行本发明的若干目的其他结构、方法和系统的基础。因此，重要的是，权利要求被认为包括这些等同结构，只要它们不偏离本发明的范围。

附图说明

[0014] 根据以下参照附图对本发明的示例性实施例的详细描述，将更好地理解本发明的示例性方面。

[0015] 图1示出了本发明的实施例的上下文图。

[0016] 图2示出了本发明的实施例的系统。

[0017] 图3示出了本发明的实施例的处理流程/工作流程图。

[0018] 图4示出了图3的系统对命令的确证。

[0019] 图5示出了由图3的系统对声音的位置的检测。

[0020] 图6示出了在图3的AR眼镜上显示命令。

[0021] 图7示出了在图3的本发明的实施例中在AR眼镜上选择命令。

[0022] 图8示出了AI语音助理执行来自图3的AR选择的命令。

[0023] 图9示出了图3的历史语料库的反馈循环。

[0024] 图10示出了本发明的实施例的高级图概述。

[0025] 图11示出了用于在其中结合本发明的示例性实施例的示例性硬件/信息处理系统。

[0026] 图12示出了用于存储实现根据本发明的示例性实施例的方法的程序的机器可读指令的信号承载存储介质。

[0027] 图13描绘了根据本发明的示例实施例的云计算节点。

[0028] 图14描绘了根据本发明的示例实施例的云计算环境。

[0029] 图15描绘了根据本发明的示例实施例的抽象模型层。

具体实施方式

[0030] 现在将参照附图描述本发明，其中相同的附图标记始终表示相同的部件。要强调的是，根据惯例，附图的各种特征不一定按比例绘制。相反，为了清楚起见，各种特征的尺寸可以任意地扩大或缩小。下面提供的示例性实施例用于说明的目的，而不是限制权利要求。此外，请注意，任何步骤可以以不同顺序执行，或组合执行，或同时执行。另外，所示的任何

结构和实施例都可以被修改或组合。

[0031] 如上所述，存在用户提交任何语音命令并且周围环境的其他用户向第一用户的语音命令提供进一步建议或附加反馈的情况。当这种情况发生时，原始用户可能不想在他的语音命令中接受其他用户的一些建议。目前，AI语音辅助系统不能够理解它应当执行哪个语音命令，以及它可以忽略哪个命令。本发明提供如下解决方案。

[0032] 图1示出了本发明的实施例的上下文图。

[0033] 本发明提供了一种方法和系统，通过该方法和系统，用户10可以在用户提交其语音命令时选择性地识别来自周围环境的哪些口语内容或非口语内容应当被考虑。

[0034] 当用户10提交语音命令时，同时，周围环境的人(用户12、14、16)也对语音命令提供附加建议。因此，用户10可以选择性地包括来自周围环境的一个或多个声音。

[0035] 例如，用户B12提供“牛奶”的语音提示，用户C14提供“苹果”的语音提示，以及用户D16提供“玩具”的语音提示。另一方面，主用户A 10向系统100提交“请定购米饭A、饼干B、冷饮”的语音命令。

[0036] AI(人工智能)语音辅助系统30将语音命令接收为“请定购米饭、饼干、冷饮、牛奶和苹果”。用户通过使用AR眼镜并观看AR环境中的语音命令和其他声音，来在AR选择18中选择“牛奶”22和“苹果”24，而不选择“玩具”26，并且还选择命令“请定购米饭、饼干和冷饮”20，以提供“请定购米饭、饼干、牛奶的冷饮和苹果”的最终输出。玩具26选择被移除。

[0037] 图2示出了本发明的实施例的示例系统。用户A 10、用户B12、用户C14和用户D16可以经由AR系统40、AI助理30和IoT(物联网)设备50的配对设备与系统100通信，这将在下面进一步描述。系统100可以包括系统计算机102，其包括处理器64和存储器66。历史语料库60可被包括在存储器66中，或者历史语料库可被分开地位于存储器设备62中。AR系统40还包括带有存储器44的处理器42，而AI助理30包括带有存储器34的处理器32。多个IoT设备50设备可与系统100配对。AR系统40可以包括增强现实眼镜46(或者其可以是单独的IoT设备50)。IoT设备50可包括带有存储器54的处理器52、以及诸如麦克风和扬声器之类的换能器56。IoT设备50可以是多个设备，每个设备也包括显示器58和多个传感器57。传感器57例如可以检测诸如智能手表或其他设备中的生物测定数据或其他信息。

[0038] 替代实施例可以包括以下AR系统40和AI助理30的系统中的任何一个或全部可以被并入系统计算机102。另外，AR眼镜46可以是与AR系统40配对的单独的IoT设备50。

[0039] 图3示出了本发明的实施例的处理流程/工作流程图。参考图3，以及参考图1和图2，在用于设备和用户标识的配对的第一部分110中，主用户(例如，来自图1和图2的用户A 10)选择进入模块并被标识112。用户10选择本发明，并且用户10的语音的历史语料库60、62被启动。在130处，语料库60、62经由路径134从系统100收集用户命令的节奏、用户命令的频率、以及对用户命令的上下文分析。

[0040] 然后，在114处，系统100被配对到AR(增强现实)系统40、AI(人工智能)助理30和IoT(物联网)设备50。AI语音辅助系统30与增强现实眼镜46配对。AI语音辅助系统40将基于语音签名来识别用户10，并且增强现实系统40将基于视网膜扫描IoT认证来识别用户。

[0041] 如果用户10、12、14、16佩戴诸如智能手表之类的另一IoT设备50，则从用户收集的生物测定将能够基于与AR系统40、AI语音助理30、以及其他生物测定数据的接近度来标识用户10、12、14、16。

[0042] 配对可以通过任何类型的通信协议来完成，诸如蓝牙、Wi-Fi、专有协议等。

[0043] 然后，主用户10陈述命令，周围环境中的成员跟随该命令116。当任何唤醒命令被提交时，AI语音辅助系统30将唤醒并将接收语音命令。

[0044] 当用户10正在提交语音命令时，AI语音辅助系统30将保持处于唤醒并将接收语音命令。

[0045] AR眼镜46的激活如下。一旦AI语音助理30唤醒，其将触发增强现实眼镜46“醒来”并收听其它声音(如果AR眼镜46包含麦克风48)。

[0046] 然后，系统100确证语音命令，并捕获准确的命令以及空间118中的位置。当用户正在提交语音命令时，AI语音辅助系统30将捕获周围环境中产生的其它声音。

[0047] 图4示出了由图3的系统对命令的确证。

[0048] 参考图4、图3和图2，命令118a的确证如下。AR系统40(如果它具有麦克风48)也将接收来自周围环境的声音，并将确证来自周围环境的辅助命令。这可以通过分析声音的差异并确保它们在相同的阈值内来完成

[0049] 利用傅立叶变换154，可以缩放音频文件以标识类似的文件：

[0050] $$f(t) = a_0 + \sum_{m=1}^{\infty} a_m \cos\left(\frac{2\pi m t}{T}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi m t}{T}\right)$$

[0051] 如果缩放的方程在定义的ε>0内，则两个音频波将被确证156。

[0052] 可以在文本分析158上进行进一步的确证。AI助理系统30注册的命令与来自增强现实系统40的命令进行比较。这可以通过使用例如词袋(Bag of Words，BOW)算法来比较每个系统30、40检测的词以检测命令中的相似性来完成。

[0053] IoT 50确证如下。连接到系统100的其他IoT设备50将能够确认该信息。

[0054] 图5示出了在图3的点118处由系统100检测声音的位置。参照图5、图2和图3，检测声音的位置118b如下。AR眼镜46还将识别声音以及谁(例如，用户10、12、14、16等)正在发出声音。

[0055] 可以通过社交媒体集成204进行标识。如果用户(例如，用户10、12、14和16)在社交媒体平台上或通过用户的联系人列表被标识为“朋友”，则他们可以以该方式被标识。

[0056] 空间中的人的对象检测206可以由AR系统40执行以标识将声音标记到谁(例如，用户10、12、14、16等)。

[0057] 声音的方向208可以近似为声波包含最高幅度的方向。

[0058] IoT确证210如下。另外，麦克风增强的配对的IoT设备50可帮助对命令的位置进行三角测量。IoT设备50可包括社交媒体集成204中存在的那些(即，同一网络上的“朋友”的设备可被利用来标识位置)。

[0059] 组合这些参数204、206、208和210将查明谁说了什么命令。

[0060] 然而，返回参考图3，在第二部分122中非命令被忽略120。忽略任意声音包括未被检测为命令的声音将被忽略。这可以包括但不限于微弱噪声或其他类型的噪声等。

[0061] 然后，在AR系统40中的AR屏幕上显示命令，并且主用户10选择最终命令124。

[0062] 图6示出了在图3的AR眼镜124a上显示命令。检测到的命令将显示在AR眼镜46上。

[0063] 哪个对象发出特定命令或声音将被显示220在AR眼镜46上。

[0064] AR眼镜46将显示周围环境声音的实际位置和方向以及AR环境222中的语音命令。

[0065] 当语音命令正在被提交时,AR眼镜将显示从周围环境产生的实时声音,并且将显示在AR眼镜224中。用户还可以可视化被转译成AR眼镜中的文本的其语音命令以及从周围环境生成的声音226。

[0066] 图7示出了在图3的本发明的实施例中在AR眼镜124b上选择命令。

[0067] 用户10可以使用手指姿势、眼睛接触、语音命令或面部表情以及任何周围环境声音,以允许这些声音被包括在语音命令中。基于用户的选择性动作,来自周围环境的一个或多个声音将被考虑,并且将被允许将这些声音与原始语音命令一起添加230。

[0068] 用户10可以以选择性的方式从声音中选择声音,并且相应地将这些声音与用户10已经提交的语音命令一起添加232。

[0069] 选择性声音可以与语音命令一起被包括,并且所述语音命令可以在用户提交语音命令时或者在提交语音命令之后被包括234。

[0070] 一部分声音的选择236如下。尽管以选择性的方式从周围环境中选择一个或多个声音,但是用户可以从周围环境中的任何源中选择一部分语音,并且相应地添加这些声音原始语音命令236。

[0071] 基于位置而不是文本的选择238如下。当从周围环境中选择任何声音时,用户10可以选择原始语音命令的位置,在该位置处,声音将与语音命令238一起添加。

[0072] 图8示出了AI语音助理执行来自图3的AR选择的命令。

[0073] 参考图2、图3和图8,然后,AI语音助理执行来自AR选择126的命令。聚合最终语音命令和执行242如下。所提出的系统100将聚合从周围环境中选择的语音,并且将与用户10提交的原始语音命令244聚合。来自AR系统40的AR显示器将向用户10显示最终语音命令,或者AI助理30将朗读最终语音命令用于执行246。则系统100在确认248之后执行250语音命令。

[0074] 图9示出了图3的历史语料库的反馈循环。

[0075] 参考图2、图3和图9,然后场景被捕获并被注入回到历史语料库132中。对历史语料库的反馈循环如下。历史上,AI语音辅助系统30和AR眼镜46将学习周围环境声音如何与提交的语音命令上下文相关,并且将识别哪个声音是合适的并且可以被添加。

[0076] 在提交语音命令时,如果从周围环境生成任何一个或多个声音,则所提出的系统100将是应当与语音命令252一起包括哪些声音。

[0077] 基于历史学习,所提出的系统将示出视图以示出所生成的声音将被包括在语音命令中的何处,因此用户将理解声音将被添加到何处254。

[0078] 社交媒体交互256如下。如果辅用户12、14或16说出许多有用的命令附加物,则它们将更可能被添加,因为主用户作出语音命令,反之亦然(如果辅用户添加许多无用的命令,则它们将不被添加到向前移动的语音命令中)256。

[0079] 返回参考图3,基于用户输入和来自路径134的与系统100的用户交互以及所捕获的场景132来构建历史语料库130。

[0080] 另一实施例可包括家庭成员模式。家庭成员声音将被识别,并且将具有用于传递语音命令的较低阈值。或者,语音助手30也可以被构建到AR系统40中。

[0081] 返回参考图2,制造中的使用情况可以如下。如果某人正在制造工厂中工作并且配备有AR系统40,则各种命令被表述为需要完成什么工作。这些命令出现在屏幕上,并且用户

可以选择哪些命令是需要遵循的命令，哪些命令仅仅是背景噪声。AR系统40的模块将这些命令聚合成一个命令，并将其作为最终命令提供给用户10。

[0082] 图10示出了本发明的实施例的高级图概述。

[0083] 本发明的一些特征的概述如下。首先，在步骤302中，系统100将语音命令转译给AR系统40。当提交任何语音命令时，AI语音辅助系统40将基于语音签名识别个体声音。因此，它将标识哪些命令来自用户10、12、14、16以及哪些命令是从周围环境生成的。此外，所提交的语音命令将被显示在增强现实眼镜46上，用于原始语音命令中一个或多个声音的选择性包括。

[0084] 然后，系统100选择AR空间304中的外部命令。在提交语音命令时或在提交语音命令之后，用户10可以选择性地包括在增强现实系统40中可视化的、并且因此将被考虑在语音命令中并且将被执行的一个或多个口语内容。

[0085] 然后，系统100提供AR空间306中的语音命令的执行。当由任何用户10、12、14、16提交任何语音命令时，增强现实眼镜46将显示从周围环境的实时声音生成，并且因此基于手指、面部和眼神接触交互，用户10可以选择性地将一个或多个声音包括在语音命令中，并且因此AI语音辅助系统40将执行完整的语音命令。

[0086] 用户10可在提交任何语音命令时从任何声音源选择完整或部分的口语内容，并且因此声音的选定部分将与语音命令包括在一起且完整的语音命令将被执行。在选择性地包括来自周围环境的一个或多个声音以及语音命令时，用户可以选择语音命令上的可以选择性地包括外部声音以完成语音命令的位置。

[0087] 然后，系统100提供用于选择命令的历史语料库308。使用关于将一个或多个声音与所提交的语音命令的选择性包括的历史学习，所提出的系统100将推荐哪些声音应当与语音命令一起包括，并且还将示出视图以推荐应当将声音添加到原始语音命令的何处。

[0088] 下面的图11至图15提供了可以实现的附加硬件配置。在图1至图15的不同附图中示出的不同特征可以在不同示例之间组合、改变或切换。

[0089] 图11示出了系统100的另一硬件配置，其中存在根据本发明的信息处理/计算机系统1100，并且其优选地具有可以实现本发明的技术的至少一个处理器或中央处理单元(CPU)1110。

[0090] CPU 1110经由系统总线1112互连到随机存取存储器(RAM)1114、只读存储器(ROM)1116、输入/输出(I/O)适配器1118(用于将诸如盘单元1121和磁带驱动器1140的外围设备连接到总线1112)、用户接口适配器1122(用于将键盘1124、鼠标1126、扬声器1128、麦克风1132和/或其它用户接口设备连接到总线1112)、用于将信息处理系统连接到数据处理网络、因特网、内联网、个人区域网(PAN)等的通信适配器1134、以及用于将总线1112连接到显示器设备1138和/或打印机1139(例如，数字打印机等)的显示器适配器1136。

[0091] 除了上述硬件/软件环境之外，本发明的不同方面包括用于执行上述方法的计算机实现的方法。作为示例，该方法可以在以上讨论的特定环境中实现。

[0092] 例如，可以通过操作由数字数据处理装置实现的计算机来执行机器可读指令序列，从而实现这种方法。这些指令可以驻留在各种类型的信号承载介质中。

[0093] 因此，本发明的这个方面涉及一种编程产品，包括信号承载存储介质，其有形地体现了可由结合CPU 1110和上述硬件的数字数据处理器执行的机器可读指令的程序，以执行

本发明的方法。

[0094] 这种信号承载存储介质可以包括例如包含在CPU 1110内的RAM,例如由快速访问存储器表示。

[0095] 或者,指令可包含在另一信号承载存储介质1200中,诸如可由CPU 1210直接或间接访问的磁性数据存储磁盘1210或光学存储磁盘1220(图12)。

[0096] 无论是包含在磁盘1210、光盘1220、计算机/CPU 1210中,还是其它地方,指令都可以存储在各种机器可读数据存储介质上。

[0097] 因此,本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括其上具有计算机可读程序指令的计算机可读存储介质(或多个介质),所述计算机可读程序指令用于使处理器执行本发明的各方面。

[0098] 计算机可读存储介质可以是能够保留和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质可以是例如但不限于电子存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或前述的任何合适的组合。计算机可读存储介质的更具体示例的非穷举列表包括以下:便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式光盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、诸如上面记录有指令的打孔卡或凹槽中的凸起结构的机械编码装置,以及上述的任何适当组合。如本文所使用的计算机可读存储介质不应被解释为暂时性信号本身,诸如无线电波或其他自由传播的电磁波、通过波导或其他传输介质传播的电磁波(例如,通过光纤线缆的光脉冲)或通过导线传输的电信号。

[0099] 本文描述的计算机可读程序指令可以从计算机可读存储介质下载到相应的计算/处理设备,或者经由网络,例如因特网、局域网、广域网和/或无线网络,下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光传输光纤、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或网络接口从网络接收计算机可读程序指令,并转发计算机可读程序指令以存储在相应计算/处理设备内的计算机可读存储介质中。

[0100] 用于执行本发明的操作的计算机可读程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据,或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言(诸如Smalltalk、C++等)以及常规的过程式编程语言(诸如“C”编程语言或类似的编程语言)。计算机可读程序指令可以完全在用户的计算机上执行,部分在用户的计算机上执行,作为独立的软件包执行,部分在用户的计算机上并且部分在远程计算机上执行,或者完全在远程计算机或服务器上执行。在后一种情况下,远程计算机可以通过任何类型的网络连接到用户的计算机,包括局域网(LAN)或广域网(WAN),或者可以连接到外部计算机(例如,使用因特网服务提供商通过因特网)。在一些实施例中,为了执行本发明的各方面,包括例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA)的电子电路可以通过利用计算机可读程序指令的状态信息来执行计算机可读程序指令以使电子电路个性化。

[0101] 在此参考根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述本发明的各方面。将理解,流程图和/或框图的每个框以及流程图和/或框图中

的框的组合可以由计算机可读程序指令来实现。

[0102] 这些计算机可读程序指令可以被提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器以产生机器，使得经由计算机或其他可编程数据处理装置的处理器执行的指令创建用于实现流程图和/或框图的一个或多个框中指定的功能/动作的装置。

[0103] 这些计算机可读程序指令还可以存储在计算机可读存储介质中，其可以引导计算机、可编程数据处理装置和/或其他设备以特定方式工作，使得其中存储有指令的计算机可读存储介质包括制品，该制品包括实现流程图和/或框图的一个或多个框中指定的功能/动作的各方面的指令。

[0104] 计算机可读程序指令还可以被加载到计算机、其他可编程数据处理装置或其他设备上，以使得在计算机、其他可编程装置或其他设备上执行一系列操作步骤，以产生计算机实现的过程，使得在计算机、其他可编程装置或其他设备上执行的指令实现流程图和/或框图的一个或多个框中指定的功能/动作。

[0105] 附图中的流程图和框图示出了根据本发明的各种实施例的系统、方法和计算机程序产品的可能实现的架构、功能和操作。在这点上，流程图或框图中的每个框可以表示指令的模块、段或部分，其包括用于实现指定的逻辑功能的一个或多个可执行指令。在一些替代实施方案中，框中所提及的功能可不按图中所提及的次序发生。例如，连续示出的两个框实际上可以基本上同时执行，或者这些框有时可以以相反的顺序执行，这取决于所涉及的功能。还将注意，框图和/或流程图图示的每个框以及框图和/或流程图图示中的框的组合可以由执行指定功能或动作或执行专用硬件和计算机指令的组合的专用的基于硬件的系统来实现。

[0106] 现在参考图13，示出了云计算节点的示例的示意图1400。云计算节点1400仅是合适的云计算节点的一个示例，并且不旨在对本文描述的本发明的实施例的使用范围或功能提出任何限制。无论如何，云计算节点1400能够被实现和/或执行上文阐述的任何功能。

[0107] 在云计算节点1400中，存在计算机系统/服务器1412，其可与许多其他通用或专用计算系统环境或配置一起操作。可适用于计算机系统/服务器1412的公知的计算系统、环境和/或配置的示例包括但不限于，个人计算机系统、服务器计算机系统、瘦客户机、胖客户机、手持或膝上型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费电子产品、网络PC、小型机系统、大型计算机系统、以及包括以上系统或设备中的任一个的分布式云计算环境等。

[0108] 计算机系统/服务器1412可以在计算机系统可执行指令的一般上下文中描述，诸如由计算机系统执行的程序模块。通常，程序模块可以包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、逻辑、数据结构等。计算机系统/服务器1412可以在分布式云计算环境中实践，其中任务由通过通信网络链接的远程处理设备执行。在分布式云计算环境中，程序模块可以位于包括存储器存储设备的本地和远程计算机系统存储介质中。

[0109] 如图13所示，云计算节点1400中的计算机系统/服务器1412以通用计算设备的形式示出。计算机系统/服务器1412的组件可包括但不限于一个或多个处理器或处理单元1416、系统存储器1428、以及将包括系统存储器1428在内的各种系统组件耦合至处理器1416的总线1418。

[0110] 总线1418表示若干类型的总线结构中的任何一种的一个或多个，包括存储器总线

或存储器控制器、外围总线、加速图形端口、以及使用各种总线体系结构中的任何一种的处理器或局部总线。作为示例而非限制，这些体系结构包括工业标准架构(ISA)总线、微通道架构(MCA)总线、增强型ISA(EISA)总线、视频电子技术标准协会(VESA)局部总线和外围部件互连(PCI)总线。

[0111] 计算机系统/服务器1412通常包括各种计算机系统可读介质。这样的介质可以是计算机系统/服务器1412可访问的任何可用介质，并且它包括易失性和非易失性介质、可移动和不可移动介质。

[0112] 系统存储器1428可以包括易失性存储器形式的计算机系统可读介质，诸如随机存取存储器(RAM)1430和/或高速缓冲存储器1432。计算机系统/服务器1412还可包括其它可移动/不可移动、易失性/非易失性计算机系统存储介质。仅作为示例，存储系统1434可被提供用于从不可移动、非易失性磁介质(未示出，并且通常被称为“硬盘驱动器”)读取和向其写入。尽管未示出，但是可以提供用于从可移动、非易失性磁盘(例如，“软盘”)读取和向其写入的磁盘驱动器，以及用于从诸如CD-ROM、DVD-ROM或其它光学介质等可移动、非易失性光盘读取或向其写入的光盘驱动器。在这样的实例中，每个都可以通过一个或多个数据介质接口连接到总线1418。如下面将进一步描绘和描述的，存储器1428可以包括至少一个程序产品，该程序产品具有一组(例如，至少一个)程序模块，该程序模块被配置成执行本发明的实施例的功能。

[0113] 作为示例而非限制，具有一组(至少一个)程序模块1442的程序/实用程序1440以及操作系统、一个或多个应用程序、其它程序模块和程序数据可被存储在存储器1428中。操作系统、一个或多个应用程序、其它程序模块和程序数据或其某种组合中的每一个可包括联网环境的实现。程序模块1442通常执行这里描述的本发明实施例的功能和/或方法。

[0114] 计算机系统/服务器1412还可以与一个或多个外部设备1414通信，诸如键盘、定点设备、显示器1424等；使用户能够与计算机系统/服务器1412交互的一个或多个设备；和/或使计算机系统/服务器1412能够与一个或多个其它计算设备通信的任何设备(例如，网卡、调制解调器等)。这种通信可以经由输入/输出(I/O)接口1422发生。此外，计算机系统/服务器1412可以经由网络适配器1420与一个或多个网络通信，所述网络诸如局域网(LAN)、通用广域网(WAN)和/或公共网络(例如，因特网)。如图所示，网络适配器1420经由总线1418与计算机系统/服务器1412的其它组件通信。应当理解，尽管未示出，但是其他硬件和/或软件组件可以与计算机系统/服务器1412结合使用。示例包括但不限于微代码、设备驱动器、冗余处理单元、外部磁盘驱动器阵列、RAID系统、磁带驱动器以及数据存档存储系统等。

[0115] 现在参考图14，描绘了说明性云计算环境1550。如图所示，云计算环境1550包括云消费者使用的本地计算设备可以与其通信的一个或多个云计算节点1400，所述本地计算设备诸如例如个人数字助理(PDA)或蜂窝电话1554A、台式计算机1554B、膝上型计算机1554C和/或汽车计算机系统1554N。节点1400可彼此通信。它们可以被物理地或虚拟地分组(未示出)在一个或多个网络中，诸如如上文描述的私有云、社区云、公有云或混合云或其组合。这允许云计算环境1550提供基础设施、平台和/或软件作为服务，云消费者不需要为其维护本地计算设备上的资源。应当理解，图14中所示的计算设备1554A-N的类型仅旨在说明，并且计算节点1400和云计算环境1550可以在任何类型的网络和/或网络可寻址连接上(例如，使用web浏览器)与任何类型的计算机化设备通信。

[0116] 现在参考图15,示出了由云计算环境1550(图14)提供的一组功能抽象层。应当预先理解,图15中所示的组件、层和功能仅旨在说明,并且本发明的实施例不限于此。如所描绘的,提供了以下层和相应的功能:

[0117] 硬件和软件层1660包括硬件和软件组件。硬件组件的示例包括大型机,在一个示例中是IBM® zSeries®系统;基于RISC(精简指令集计算机)架构的服务器,在一个示例中是IBM pSeries®系统;IBM xSeries®系统;IBM BladeCenter®系统;存储设备;网络和联网组件。软件组件的示例包括网络应用服务器软件,在一个示例中是IBM WebSphere®应用服务器软件;以及数据库软件,在一个示例中是IBM DB2®数据库软件。(IBM、zSeries、pSeries、xSeries、BladeCenter、WebSphere和DB2是在全世界许多管辖区注册的国际商业机器公司的商标)。

[0118] 虚拟化层1662提供抽象层,从该抽象层可以提供虚拟实体的以下示例:虚拟服务器;虚拟存储;虚拟网络,包括虚拟专用网络;虚拟应用和操作系统;以及虚拟客户端。

[0119] 在一个示例中,管理层1664可以提供以下描述的功能。资源供应提供了对被用来在云计算环境内执行任务的计算资源和其他资源的动态采购。计量和定价提供了在云计算环境中利用资源时的成本跟踪,以及用于消耗这些资源的开账单或开发票。在一个示例中,这些资源可以包括应用软件许可证。安全性为云消费者和任务提供身份验证,以及为数据和其他资源提供保护。用户门户为消费者和系统管理员提供对云计算环境的访问。服务级别管理提供云计算资源分配和管理,使得满足所需的服务级别。服务级别协议(SLA)规划和履行提供了云计算资源的预安排和采购,其中根据SLA来预计未来需求。

[0120] 工作负载层1666提供了可以利用云计算环境的功能的示例。可以从该层提供的工作负载和功能的示例包括这样的功能,诸如绘图和导航;软件开发和生命周期管理;虚拟教室教育递送;数据分析处理;事务处理;并且,更具体地,相对于本发明,基于上下文输入生成搜索自动完成建议的API和运行时系统组件。

[0121] 从详细的说明书中,本发明的许多特征和优点是显而易见的,因此,所附权利要求旨在覆盖落入本发明的真实范围内的本发明的所有这些特征和优点。此外,由于本领域技术人员将容易想到许多修改和变化,因此不期望将本发明限制于所图示和描述的确切构造和操作,并且因此,可以诉诸于落入本发明的范围内的所有合适的修改和等同物。

[0122] 应当理解,本发明在其应用中不限于以下描述中阐述的或附图中示出的构造细节和部件布置。本发明能够具有除所描述的那些实施例之外的实施例,并且能够以各种方式实践和执行。而且,应当理解,这里所使用的措辞和术语以及摘要是为了描述的目的,而不应当被认为是限制。

[0123] 因此,本领域技术人员将理解,本公开所基于的概念可以容易地用作设计用于执行本发明的若干目的其他结构、方法和系统的基础。因此,重要的是,权利要求被认为包括这些等同结构,只要它们不偏离本发明的范围。

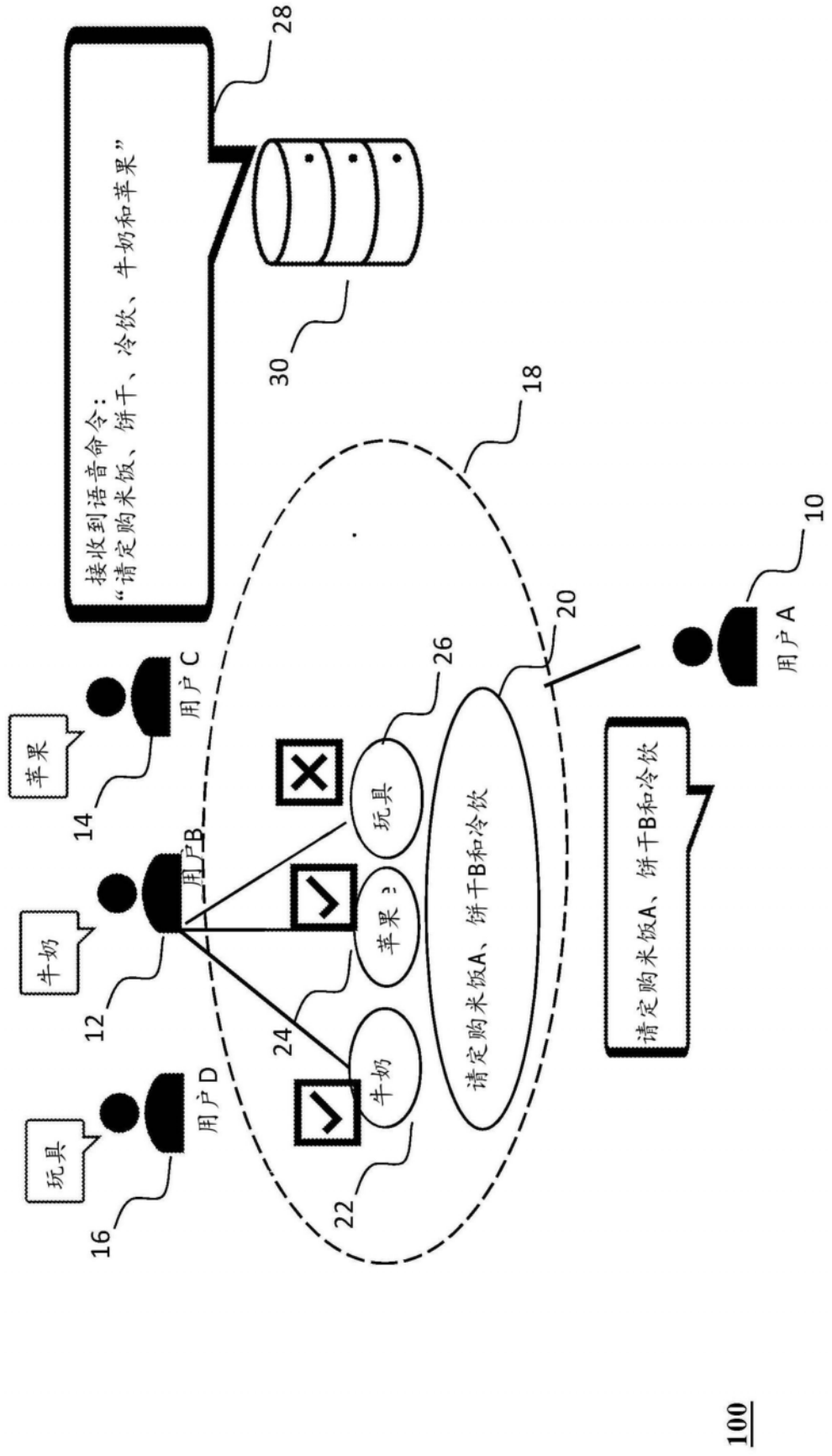

接收到语音命令：
“请定购米饭、饼干、冷饮、牛奶和苹果”
28
30
18
10
用户A
20
请定购米饭A、饼干B和冷饮
26
玩具
苹果
用户C
14
苹果
牛奶
12
用户B
24
玩具
用户D
16
牛奶
22
请定购米饭A、饼干B和冷饮
100

图1

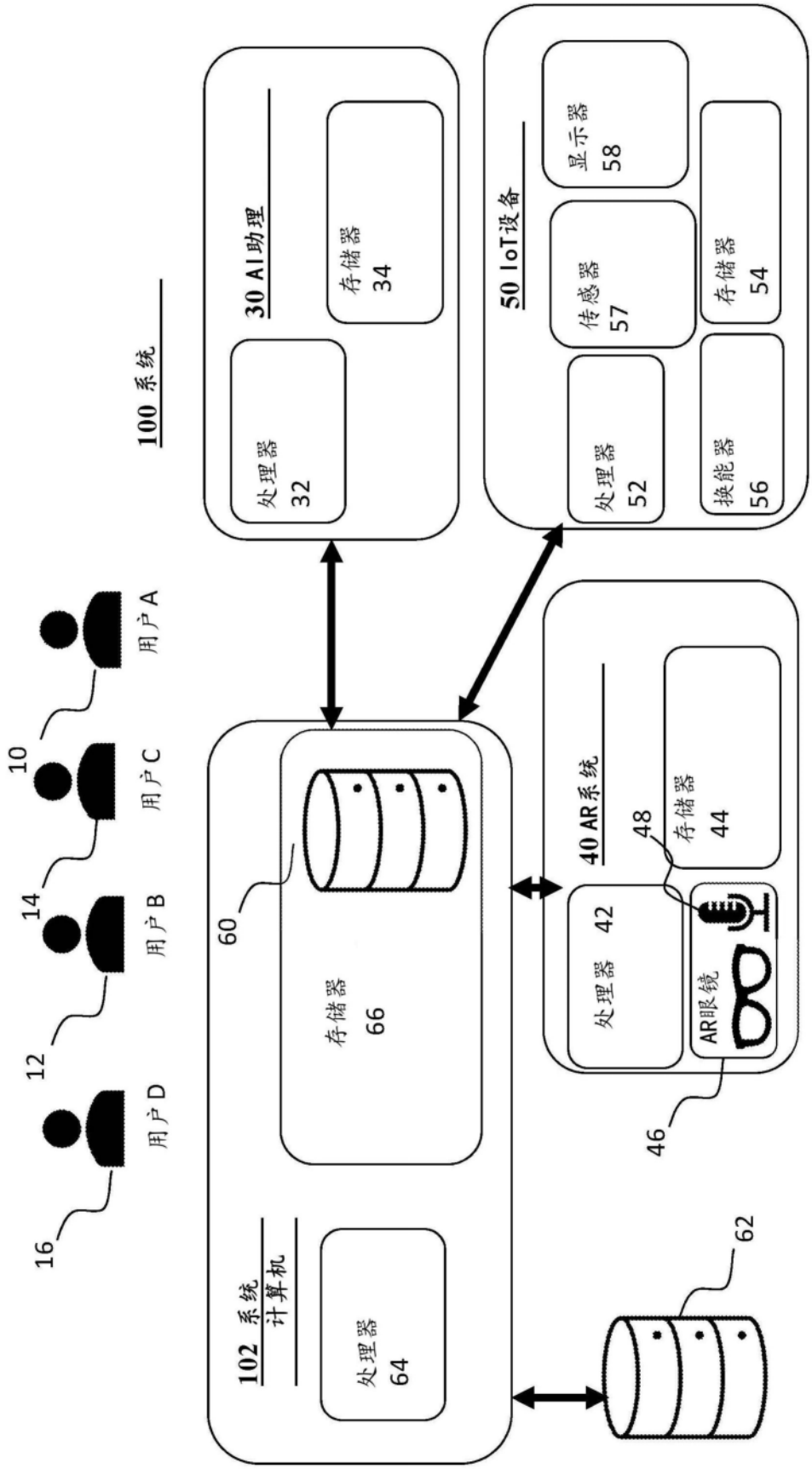

100 系统
30 AI助理
处理器
32
存储器
34
50 IoT设备
处理器
52
传感器
57
显示器
58
存储器
54
换能器
56
40 AR系统
处理器 42
48
存储器
44
AR眼镜
46
60
存储器
66
102 系统
计算机
处理器
64
62
用户A
用户B
用户C
用户D
10
12
14
16

图2

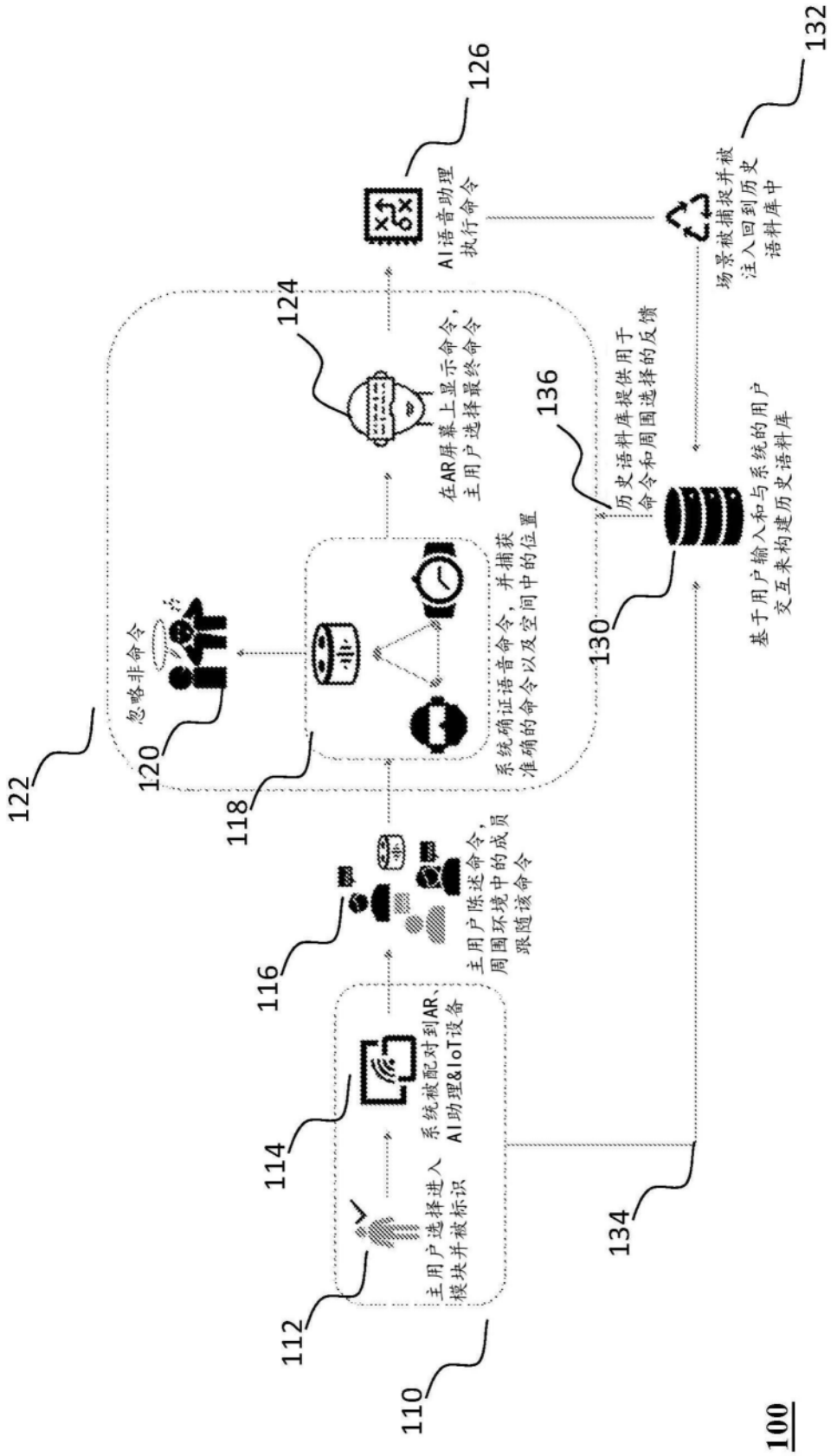

100
110
112
主用户选择进入模块并被标识
114
系统被配对到AR、AI助理&IoT设备
116
主用户陈述命令，周围环境中的成员跟随该命令
118
120
忽略非命令
122
系统确证语音命令，并捕获准确的命令以及空间中的位置
124
在AR屏幕上显示命令，主用户选择最终命令
126
AI语音助理执行命令
130
基于用户输入和与系统的用户交互来构建历史语料库
132
场景被捕捉并被注入回到历史语料库中
134
136
历史语料库提供用于命令和周围选择的反馈

图3

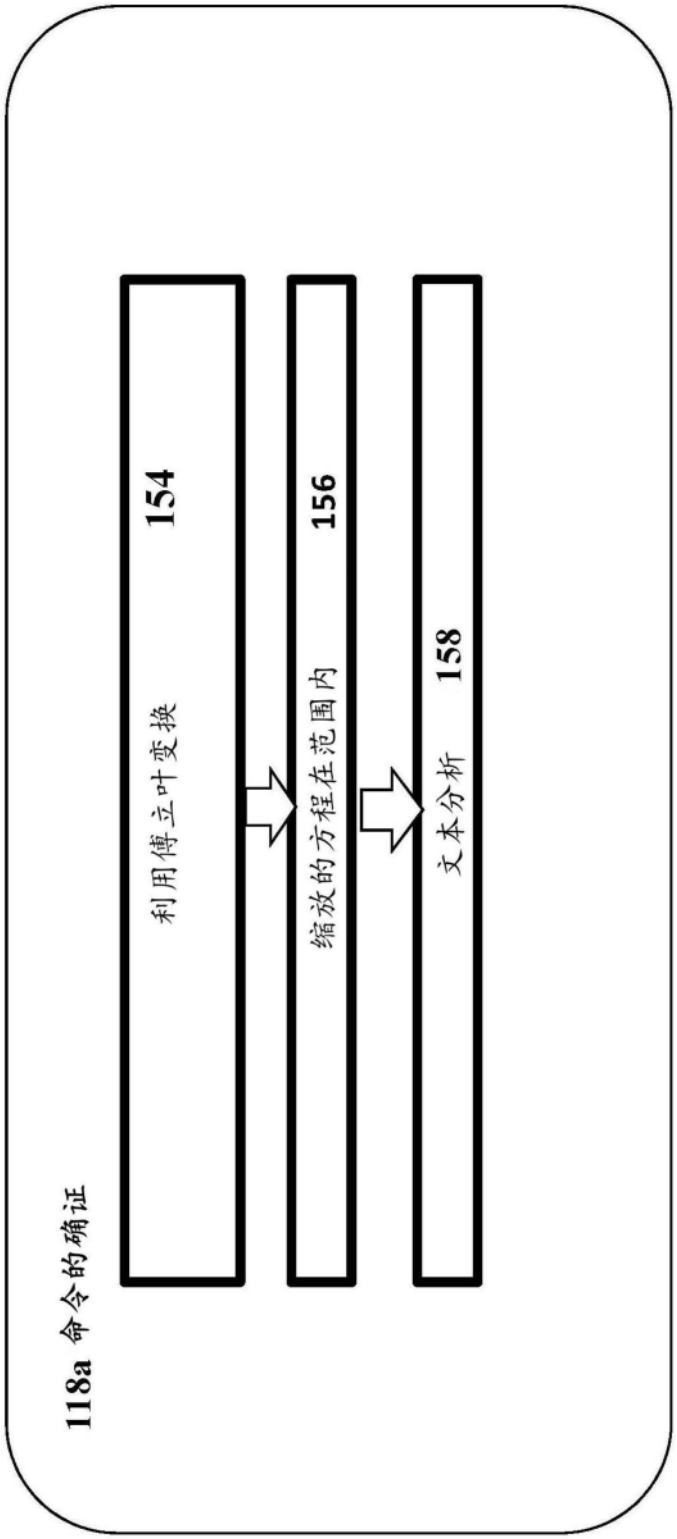


图4

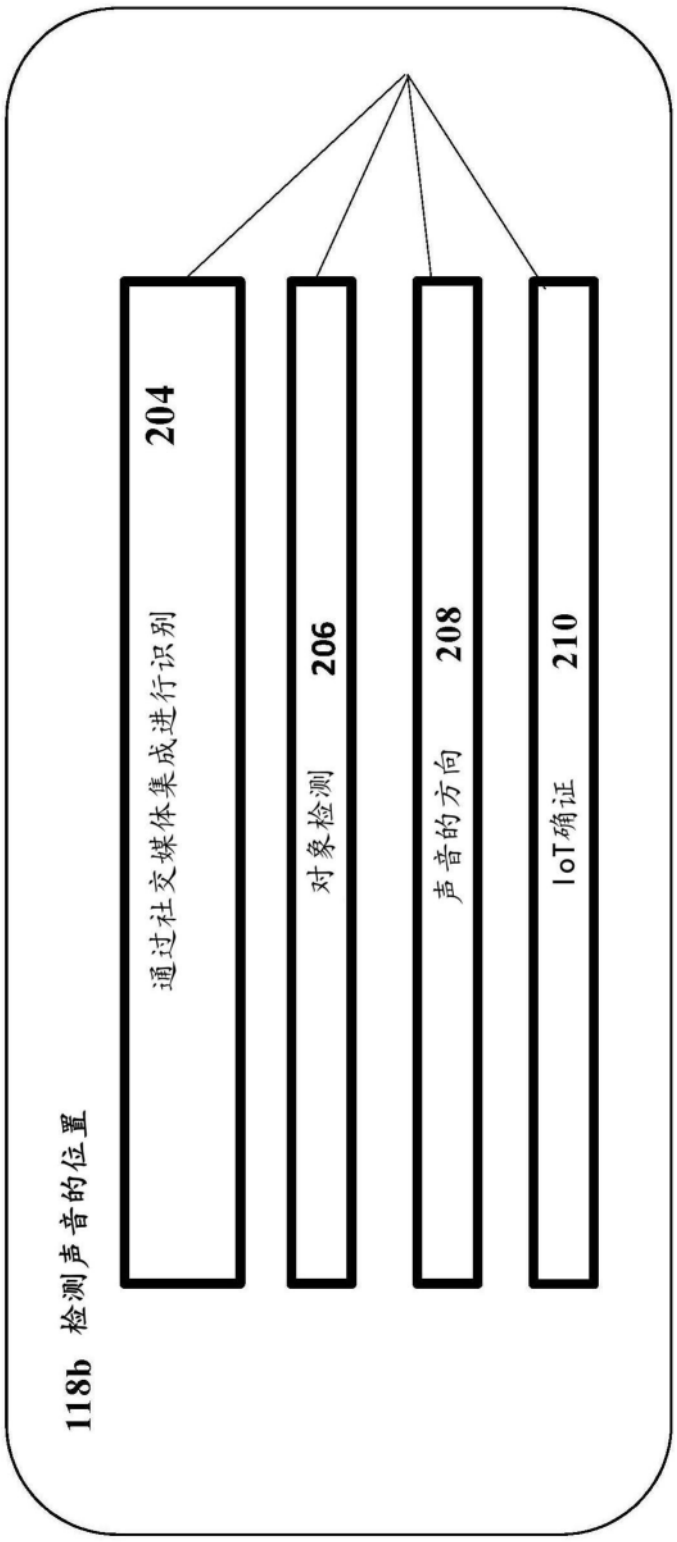


图5

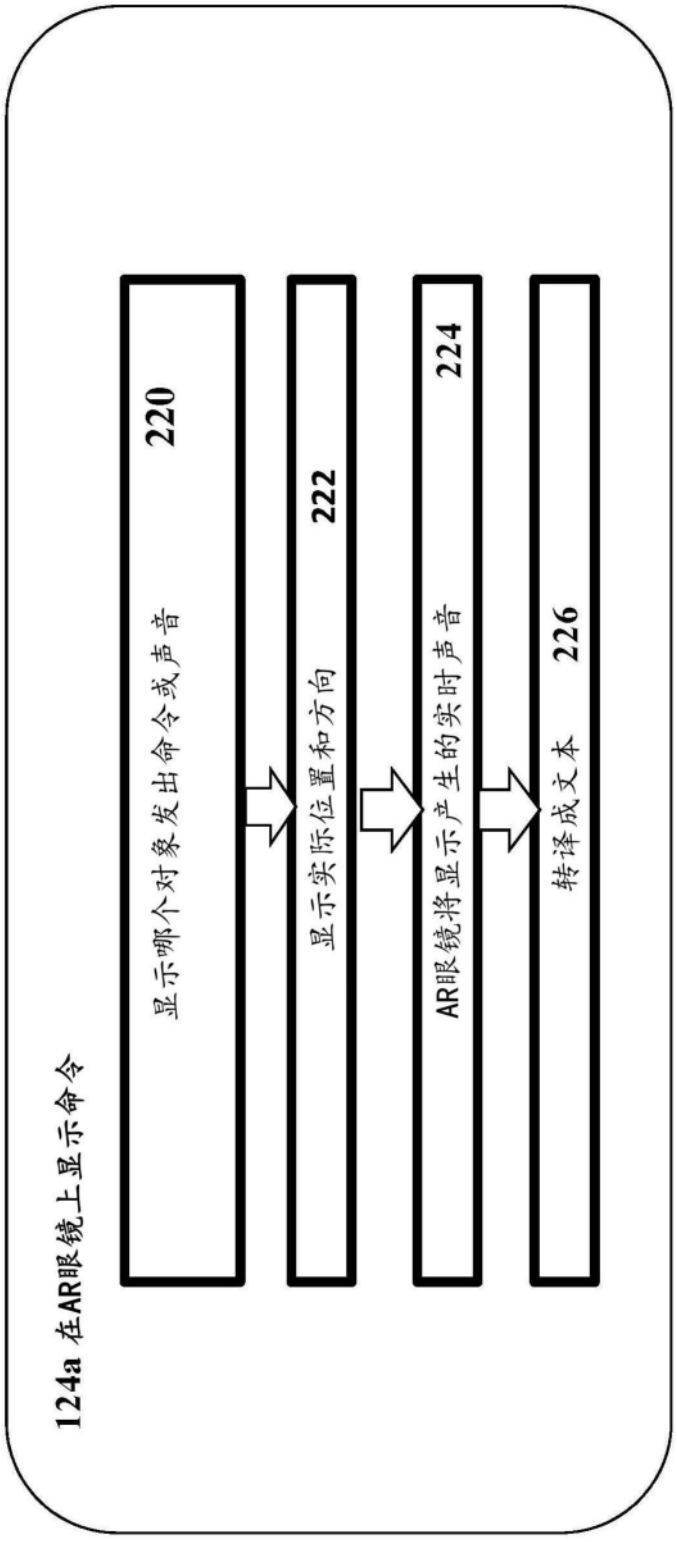

124a 在AR眼镜上显示命令
显示哪个对象发出命令或声音 220
显示实际位置和方向 222
AR眼镜将显示产生的实时声音 224
转译成文本 226

图6

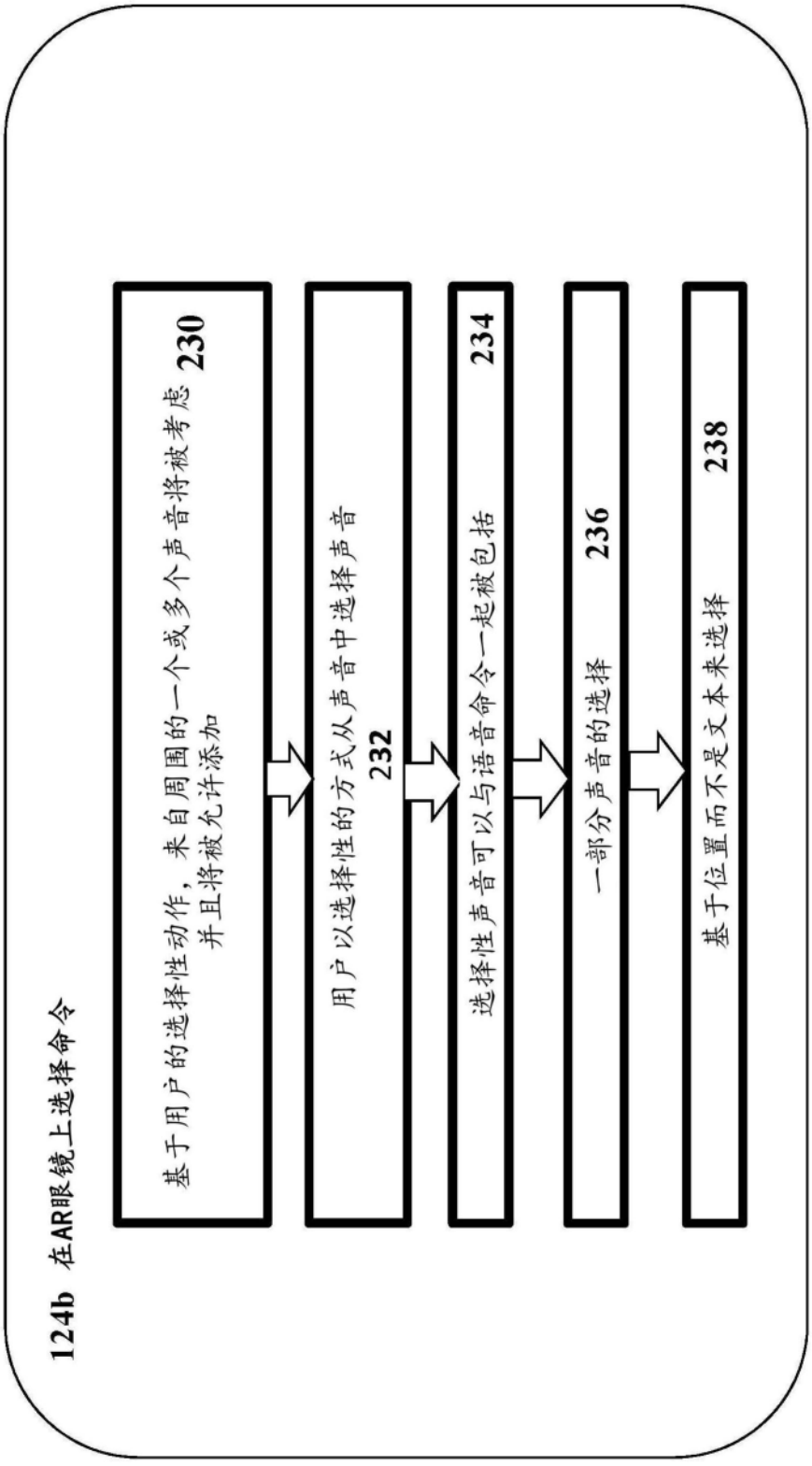

124b 在AR眼镜上选择命令
基于用户的选择性动作，来自周围的一个或多个声音将被考虑并且将被允许添加 230
用户以选择性的方式从声音中选择声音 232
选择性声音可以与语音命令一起被包括 234
一部分声音的选择 236
基于位置而不是文本来选择 238

图7

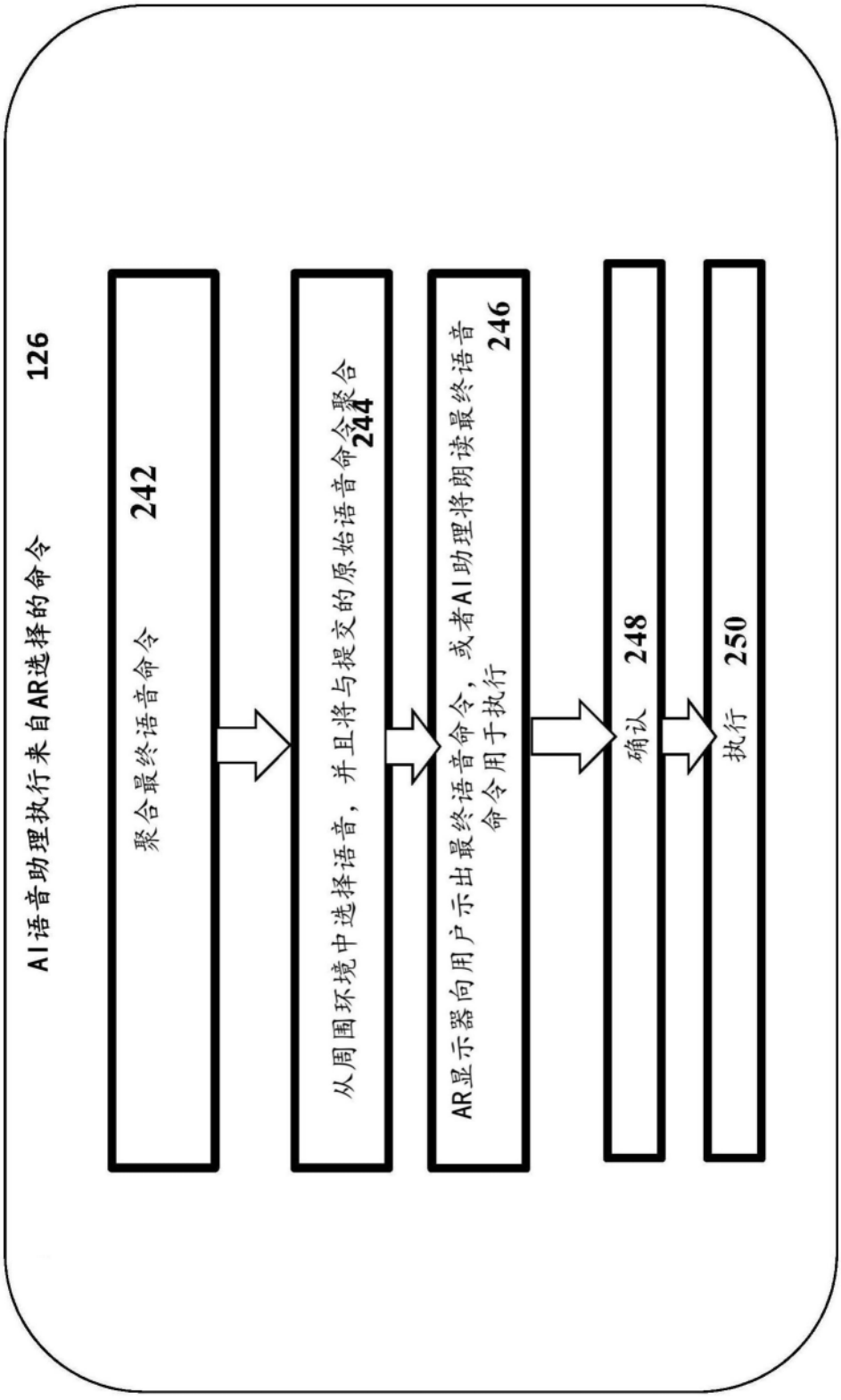

AI语音助理执行来自AR选择的命令 126
聚合最终语音命令 242
从周围环境中选择语音，并且将与提交的原始语音命令聚合 244
AR显示器向用户示出最终语音命令，或者AI助理将朗读最终语音命令用于执行 246
确认 248
执行 250

图8

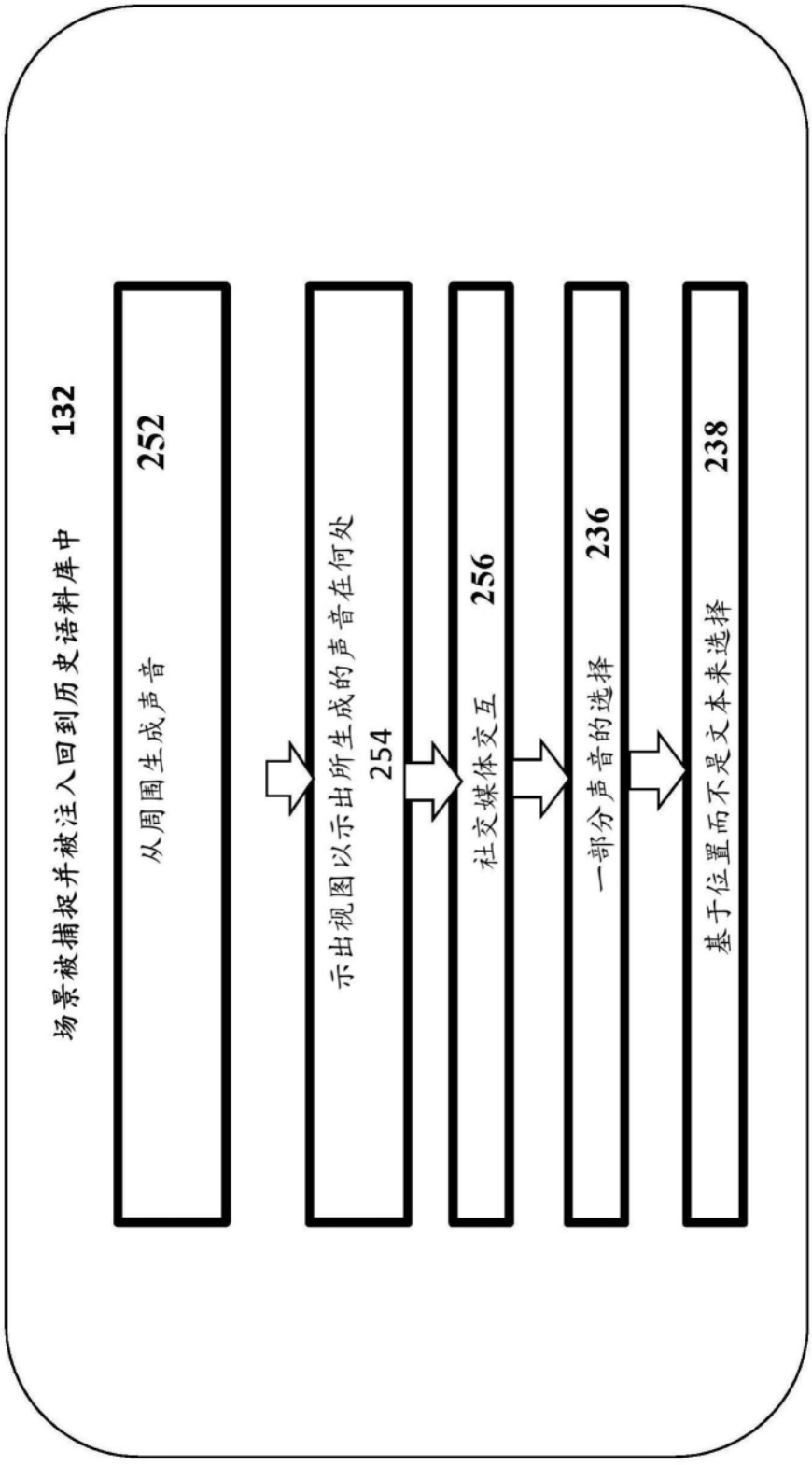

场景被捕捉并被注入回到历史话料库中 132
从周围生成声音 252
示出视图以示出所生成的声音在何处 254
社交媒体交互 256
一部分声音的选择 236
基于位置而不是文本来选择 238

图9


将语音命令转译给AR 302
选择AR空间中的外部命令 304
在AR空间中执行语音命令 306
用于选择命令的历史语料库 308
10

图10

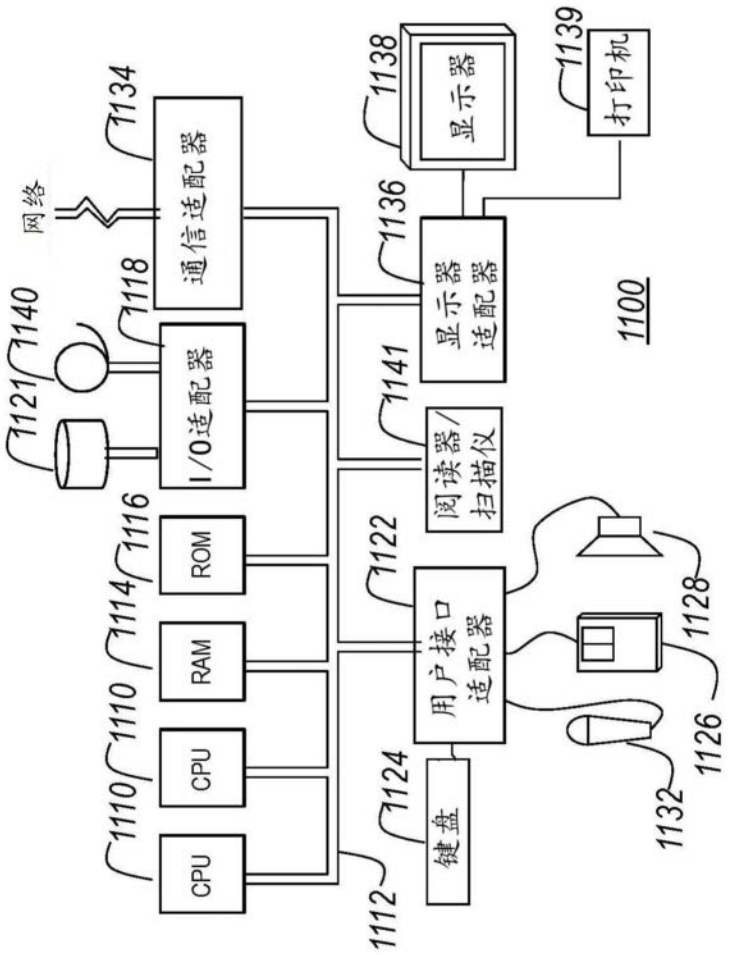

网络
通信适配器
I/O适配器
ROM
RAM
CPU
CPU
显示器
打印机
显示器适配器
阅读器/扫描仪
用户接口适配器
键盘
1100
1110
1112
1114
1116
1118
1121
1122
1124
1126
1128
1132
1134
1136
1138
1139
1140
1141

图11

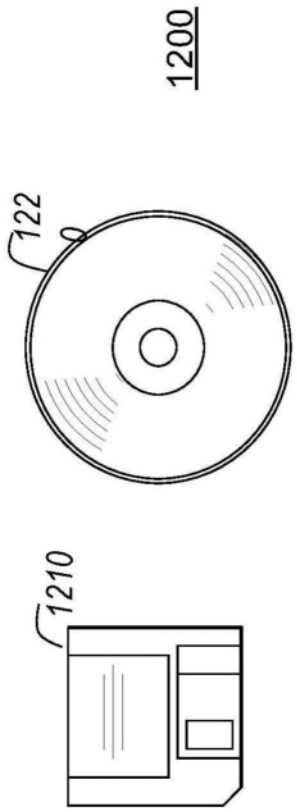


图12

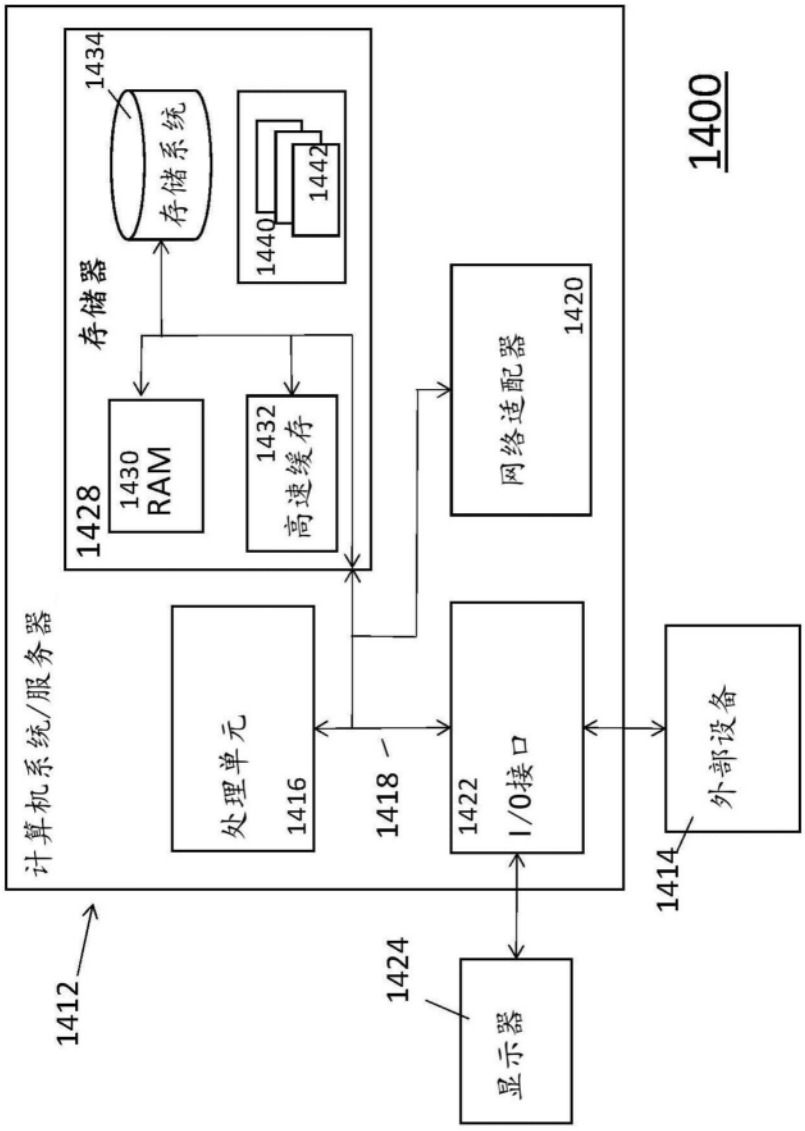

1400
计算机系统/服务器
1412
处理单元
1416
1418
1422
I/O接口
1414
外部设备
网络适配器
1420
1424
显示器
1428
存储器
1430
RAM
1432
高速缓存
1434
存储系统
1440
1442

图13

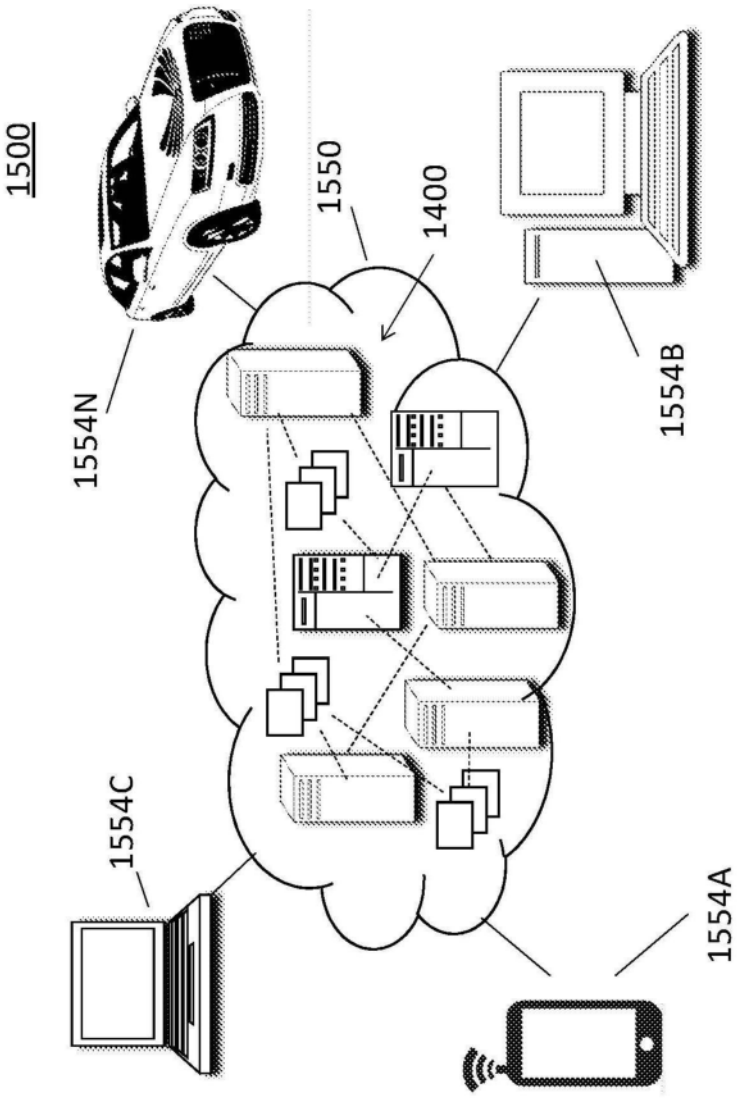


图14

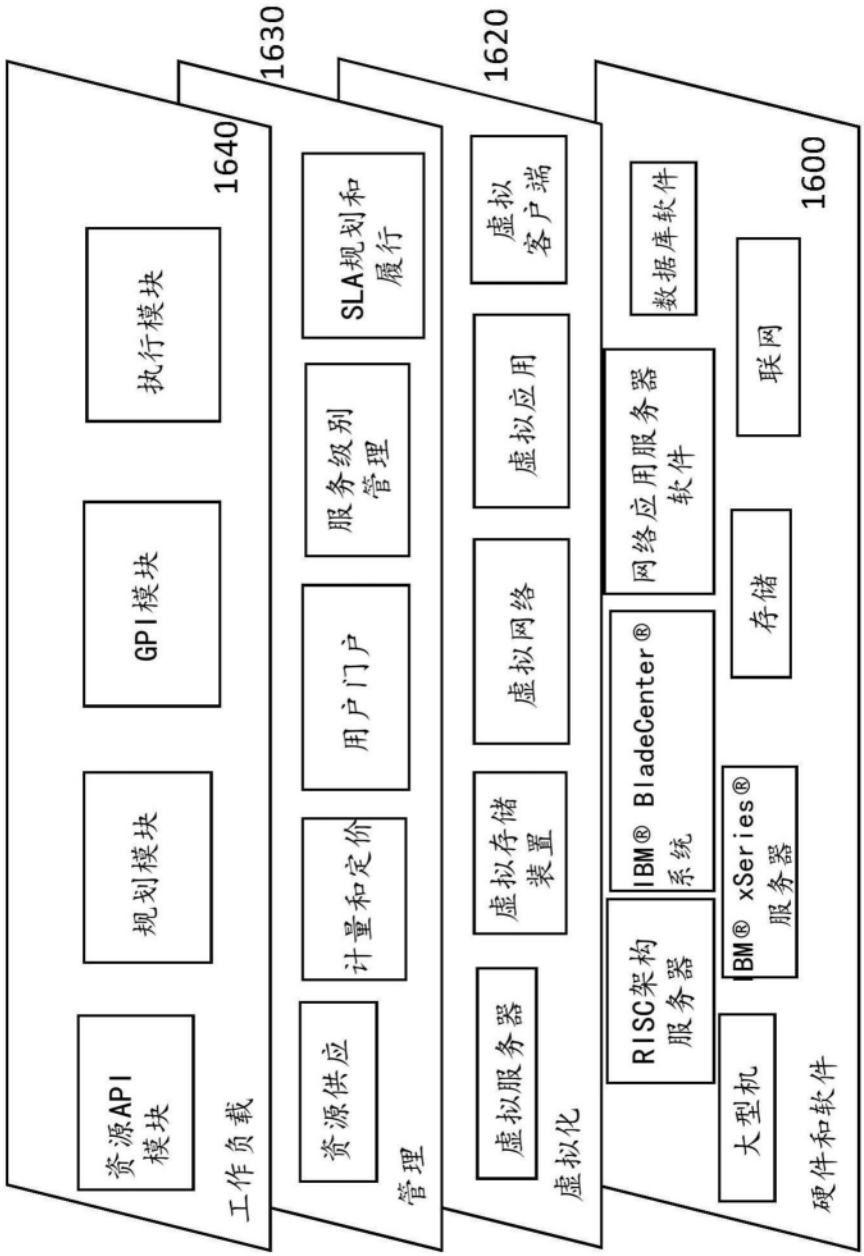

1640
1630
1620
1600
资源API模块
规划模块
GPI模块
执行模块
工作负载
资源供应
计量和定价
用户门户
服务级别管理
SLA规划和履行
管理
虚拟服务器
虚拟存储装置
虚拟网络
虚拟应用
虚拟客户端
虚拟化
RISC架构服务器
IBM® BladeCenter®系统
网络应用服务器软件
数据库软件
大型机
IBM® xSeries®服务器
存储
联网
硬件和软件

图15