(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 108022590 B
(45)授权公告日 2023.10.31

(21)申请号 201711071055.1

(22)申请日 2017.11.03

(65)同一申请的已公布的文献号
申请公布号 CN 108022590 A

(43)申请公布日 2018.05.11

(30)优先权数据
62/417,281 2016.11.03 US
15/801,307 2017.11.01 US

(73)专利权人 谷歌有限责任公司
地址 美国加利福尼亚州

(72)发明人 肯尼斯·米克斯特 托默·谢凯尔 图安·安赫·恩古耶

(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
专利代理师 周亚荣 安翔

(51)Int.Cl.
G10L 15/22(2006.01)
G10L 15/30(2013.01)

(56)对比文件
US 8768712 B1,2014.07.01
US 2015287411 A1,2015.10.08
CN 105393302 A,2016.03.09
US 2015325234 A1,2015.11.12
US 2016189717 A1,2016.06.30
US 2014258942 A1,2014.09.11

审查员 罗朋

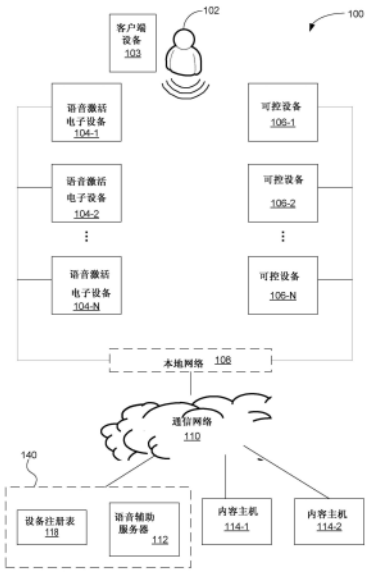
权利要求书3页 说明书24页 附图13页

(54)发明名称

语音接口设备处的聚焦会话

(57)摘要

本申请涉及语音接口设备处的聚焦会话。一种已连接电子设备的本地组中的第一电子设备处的方法包括:接收包括对第一操作的请求的第一语音命令;从所述本地组当中确定用于所述第一操作的第一目标设备;相对于所述第一目标设备建立聚焦会话;使所述第一操作由所述第一目标设备执行;接收包括对第二操作的请求的第二语音命令;确定所述第二语音命令不包括对第二目标设备的显式指定;确定所述第二操作可由所述第一目标设备执行;确定所述第二语音命令是否满足一个或多个聚焦会话维持准则;以及如果所述第二语音命令满足所述聚焦会话维持准则,则使所述第二操作由所述第一目标设备执行。



CN 108022590 B

1.一种用于将语音命令定向到目标设备的方法,包括:

在具有一个或多个麦克风、扬声器、一个或多个处理器以及存储由所述一个或多个处理器执行的一个或多个程序的存储器的第一电子设备处,所述第一电子设备是通信地耦合到服务器实现的公共网络服务的已连接电子设备的本地组的成员:

接收包括对第一操作的请求的第一语音命令;

基于所述第一语音命令的内容从已连接电子设备的所述本地组当中确定用于所述第一操作的第一目标设备;

相对于所述第一目标设备建立聚焦会话,其中,相对于所述第一目标设备建立聚焦会话包括将所述第一目标设备指派作为用于执行所述第一操作的对焦设备;

根据将所述第一目标设备指派作为用于执行所述第一操作的所述对焦设备,经由所述服务器实现的公共网络服务的操作使所述第一操作由所述第一目标设备执行;

在所述聚焦会话是活动的并且所述第一目标设备正在执行所述第一操作时:

接收包括对第二操作的请求的第二语音命令;

确定所述第二语音命令不包括对第二目标设备的显式指定;

确定所述第二操作可由所述第一目标设备执行;

确定所述第二语音命令是否满足一个或多个聚焦会话维持准则;以及

根据(i)所述第二语音命令满足所述一个或多个聚焦会话维持准则的确定,(ii)所述第二语音命令不包括对第二目标设备的显式指定的确定,以及(iii)所述第二操作可由所述第一目标设备执行的确定,维持相对于所述第一目标设备的所述聚焦会话,包括:

将所述第一目标设备指派作为用于执行所述第二操作的所述对焦设备;

根据将所述第一目标设备指派作为用于执行所述第二操作的所述对焦设备,经由所述服务器实现的公共网络服务的操作使所述第二操作由所述第一目标设备执行。

2.根据权利要求1所述的方法,其中从已连接电子设备的所述本地组当中确定用于所述第一操作的第一目标设备包括:

从所述第一语音命令获得对所述第一目标设备的显式指定。

3.根据权利要求1所述的方法,其中从已连接电子设备的所述本地组当中确定用于所述第一操作的第一目标设备包括:

确定所述第一语音命令不包括对所述第一目标设备的显示指定;

确定所述第一操作可由已连接电子设备的所述本地组当中的第二电子设备执行;以及

选择所述第二电子设备作为所述第一目标设备。

4.根据权利要求1所述的方法,进一步包括:

根据所述第二语音命令满足所述一个或多个聚焦会话维持准则的所述确定,相对于所述第一目标设备延长所述聚焦会话。

5.根据权利要求1所述的方法,其中相对于所述第一目标设备建立所述聚焦会话包括:

存储所述第一语音命令的时间戳;以及

存储所述第一目标设备的标识符。

6.根据权利要求1所述的方法,进一步包括:

接收包括对第三操作的请求和对已连接电子设备的所述本地组当中的第三目标设备的显式指定的第三语音命令;

相对于所述第一目标设备结束所述聚焦会话；

相对于所述第三目标设备建立另一聚焦会话；以及

经由所述服务器实现的公共网络服务的操作使所述第三操作由所述第三目标设备执行。

7.根据权利要求1所述的方法，其中第一目标设备是所述第一电子设备；以及

所述方法进一步包括：

接收包括对第四操作的请求和对已连接电子设备的所述本地组当中的第四目标设备的显式指定的第四语音命令，其中所述第四目标设备是作为已连接电子设备的所述本地组中的成员的第三电子设备，所述第三电子设备与所述第一电子设备不同；

相对于所述第一目标设备维持所述聚焦会话；

经由所述服务器实现的公共网络服务的操作使所述第四操作由所述第四目标设备执行。

8.根据权利要求7所述的方法，其中：

所述第二语音命令在使所述第四操作由所述第四目标设备执行之后被接收；

所述第一操作是媒体播放操作；以及

所述第二操作是媒体中止操作；以及

所述方法进一步包括：

接收包括对第五操作的请求和对已连接电子设备的所述本地组当中的第五目标设备的显式指定的第五语音命令，其中所述第五目标设备是所述第三电子设备；

相对于所述第一目标设备结束所述聚焦会话；

相对于所述第五目标设备建立另一聚焦会话；以及

经由所述服务器实现的公共网络服务的操作使所述第五操作由所述第五目标设备执行。

9.根据权利要求1所述的方法，进一步包括：

接收包括预定义操作终止请求的第五语音命令；以及

根据接收到所述第五语音命令：

使所述第一操作停止由所述第一目标设备执行；以及

相对于所述第一目标设备结束所述聚焦会话。

10.根据权利要求1所述的方法，其中：

所述第一操作是媒体播放操作；以及

所述第二操作是以下各项中的一个：媒体中止操作、媒体倒回操作、媒体快进操作、调高音量操作和调低音量操作。

11.根据权利要求1所述的方法，其中：

所述第一操作是到多个设备状态中的第一设备状态的设备状态改变操作；以及

所述第二操作是到所述多个设备状态中的第二设备状态的设备状态改变操作。

12.根据权利要求1所述的方法，其中：

所述第一操作是在幅度标尺的第一方向上的幅度改变操作；以及

所述第二操作是在所述幅度标尺的与所述第一方向相反的第二方向上的幅度改变操作。

13.根据权利要求1所述的方法,其中所述第一电子设备包括一个或多个LED的阵列;以及

所述方法进一步包括:

通过照亮所述LED阵列中的所述LED中的一个或多个LED来指示所述聚焦会话的状态。

14.一种电子设备,包括:

一个或多个麦克风;

扬声器;

一个或多个处理器;以及

存储指令的存储器,所述指令在由所述一个或多个处理器执行使得所述一个或多个处理器执行根据权利要求1至13中的任一项所述的方法。

15.一种存储指令的非暂时性计算机可读存储介质,所述指令在由具有一个或多个麦克风、扬声器和一个或多个处理器的电子设备执行时,使得所述电子设备执行根据权利要求1至13中的任一项所述的方法的操作。

语音接口设备处的聚焦会话

技术领域

[0001] 所公开的实施方式一般地涉及语音接口和相关设备，包括但不限于用于在目标设备从语音命令本身是未知的或模糊不清时将语音命令定向到目标设备的方法和系统。

背景技术

[0002] 具有语音接口的电子设备已广泛地用于从用户收集语音输入并根据这些语音输入执行不同的语音激活功能。这些语音激活功能可以包括指示或者命令目标设备执行操作。例如，用户可以向语音接口设备发出语音输入以指示目标设备打开或者关闭，或者以控制目标设备处的媒体播放。

[0003] 通常，如果用户希望作出指示目标设备执行操作的语音输入，则该用户将在该语音输入中指定目标设备。然而，必须为所有此类语音输入显式地指定目标设备对用户而言是乏味且麻烦的。即便当语音输入未指定目标或者指定了模糊不清的目标时，也期望语音接口设备具有用于语音输入的目标设备。

发明内容

[0004] 因此，需要具有语音辅助系统和/或语音辅助服务器系统的电子设备，所述语音辅助系统和/或语音辅助服务器系统包含用于当在语音输入中对目标设备的指定不存在或模糊不清时为语音输入确定或者指派目标设备的方法和系统。在本申请中所描述的各种实施方式中，操作环境包括向语音辅助服务提供接口的语音激活电子设备，以及可以经由所述语音辅助服务通过语音输入来控制的多个设备(例如，投射设备(cast device)、智能家居设备)。所述语音激活电子设备被配置成记录所述语音辅助服务(例如，语音辅助服务器系统)用来确定用户语音请求(例如，媒体播放请求、电力状态改变请求)的语音输入。所述语音辅助服务器系统然后将所述用户语音请求定向到如通过所述语音输入所指示的目标设备。所述语音激活电子设备被配置成记录对目标设备的指示不存在或模糊不清的后续语音输入。所述电子设备或所述语音辅助服务器系统为此语音输入指派目标设备，确定包括在此语音输入中的用户语音请求，并且将所述用户语音请求定向到所指派的目标设备。

[0005] 根据一些实施方式，在具有一个或多个麦克风、扬声器、一个或多个处理器以及存储由所述一个或多个处理器执行的一个或多个程序的存储器的第一电子设备处执行方法。所述第一电子设备是通信地耦合到公共网络服务的已连接电子设备的本地组的成员。所述方法包括：接收包括对第一操作的请求的第一语音命令；从已连接电子设备的所述本地组当中确定用于所述第一操作的第一目标设备；相对于所述第一目标设备建立聚焦会话；经由所述公共网络服务的操作使所述第一操作由所述第一目标设备执行；接收包括对第二操作的请求的第二语音命令；确定所述第二语音命令不包括对第二目标设备的显式指定；确定所述第二操作可由所述第一目标设备执行；确定所述第二语音命令是否满足一个或多个聚焦会话维持准则；以及根据所述第二语音命令满足所述聚焦会话维持准则的确定，经由所述公共网络服务的操作使所述第二操作由所述第一目标设备执行。

[0006] 根据一些实施方式，电子设备包括一个或多个麦克风、扬声器、一个或多个处理器以及存储待由所述一个或多个处理器执行的一个或多个程序的存储器。所述一个或多个程序包括用于执行上述的所述方法的指令。

[0007] 根据一些实施方式，非暂时性计算机可读存储介质存储一个或多个程序。所述一个或多个程序包括指令，所述指令当由具有一个或多个麦克风、扬声器和一个或多个处理器的电子设备执行时，使所述电子设备执行上述的所述方法的操作。

附图说明

[0008] 为了更好地理解各种描述的实施方式，应该结合以下附图参考下面的具体实施方式，在附图中相同的附图标记在所有图中指代对应的部分。

[0009] 图1图示根据一些实施方式的示例操作环境。

[0010] 图2图示根据一些实施方式的示例语音激活电子设备。

[0011] 图3A至图3B图示根据一些实施方式的示例语音辅助服务器系统。

[0012] 图4A至图4D图示根据一些实施方式的聚焦会话的示例。

[0013] 图5图示根据一些实施方式的建立聚焦会话并根据聚焦会话对语音输入作出响应的示例过程的流程图。

[0014] 图6A和图6B是根据一些实施方式的语音激活电子设备的前视图和后视图。

[0015] 图6C是根据一些实施方式的示出按照开放配置包含在电子设备190的基座中的扬声器的语音激活电子设备190的透视图。

[0016] 图6D是根据一些实施方式的示出包含在其中的电子组件的语音激活电子设备的侧视图。

[0017] 图6E(1)至图6E(4)示出根据一些实施方式的在语音激活电子设备的触摸感测阵列上检测到的四个触摸事件。

[0018] 图6E(5)示出根据一些实施方式的用户按压在语音激活电子设备的后侧的按钮。

[0019] 图6F是根据一些实施方式的语音激活电子设备的顶视图。

[0020] 图6G示出根据一些实施方式的通过用于指示语音处理状态的全色LED的阵列所显示的示例视觉图案。

[0021] 相同的附图标记在附图的数个视图中自始至终指代对应的部分。

具体实施方式

[0022] 虽然数字革命已提供了范围从公开共享信息到全球社区意义的许多好处，但是新兴的新技术常常在消费者当中引发混淆、怀疑和恐惧，从而防止消费者从本技术中受益。电子设备被方便地用作语音接口来从用户接收语音输入并发起语音激活功能，并且因此提供免视(eyes-free)和免手操(hands-free)方案以接近现有技术和新兴技术两者。具体地，即使用户的视线被遮挡并且他的手不得闲，在电子设备处接收到的语音输入也可以承载指令和信息。为了实现免提和免视体验，语音激活电子设备不变地或者仅在被触发时侦听环境(即，持续不变地对从环境收集到的音频信号进行处理)。另一方面，用户身份与用户的语音和由该用户使用的语言链接。为了保护用户身份，通常在作为受保护的、受控制的且亲密的空间(例如，家庭和汽车)的非公共场所中使用这些语音激活电子设备。

[0023] 根据一些实施方式，当语音命令中对目标设备的指示不存在或模糊不清时，语音激活电子设备确定或者将目标设备指派给在语音命令中作出的请求。语音激活电子设备相对于在语音命令中显式地指定或者指示的目标设备建立聚焦会话。当语音激活电子设备接收到对目标设备的指定或指示不存在或模糊不清的后续语音命令时，如果该语音命令满足一个或多个准则，则该电子设备将聚焦会话的目标设备指派给该语音命令。

[0024] 在一些实施方式中，当用户与语音接口设备对话以控制另一设备时，该语音接口设备存储哪一个设备正被用户作为目标(例如，在聚焦会话中)。在那之后的一段时期内，用于控制的默认目标设备是所存储的设备。例如，如果用户首先发出语音命令“打开厨房灯”，然后发出“关闭灯”，则在第一命令之后不久接收到第二语音命令的情况下用于第二语音命令的目标设备默认为“厨房灯”。作为另一示例，如果第一命令是“在客厅扬声器上播放音乐”，并且后续命令是“停止音乐”，则在第一命令之后不久接收到第二命令的情况下用于第二语音命令的目标设备默认为“客厅扬声器”。

[0025] 附加地，在一些实施方式中，如果在语音输入之间存在较长的时间间隙，则用户可能被要求确认或者验证最后使用的目标设备是预定目标设备。例如，如果第一语音命令是“在客厅扬声器上播放音乐”，并且在从第一语音命令起较长的时间间隙之后接收到的后续命令是“停止音乐”，则语音接口设备可以问用户“你想要停止客厅扬声器上的音乐吗?”以确认目标设备是“客厅扬声器”。

[0026] 以这种方式，用户可以被免去必须在每一个语音输入中指定他的请求的完整场境(context)的负担(例如，免去必须在请求待执行的操作的每一个语音输入中包括对目标设备的指定)。

[0027] 语音辅助操作环境

[0028] 图1是根据一些实施方式的示例操作环境。操作环境100包括一个或多个语音激活电子设备104(例如，语音激活电子设备104-1至104-N，在下文中被称为“语音激活设备”)。所述一个或多个语音激活设备104可以位于一个或多个位置中(例如，全部在一个结构的房间或空间中、遍布在一个结构内的多个空间中或者遍布在多个结构中(例如，一个在住所中并且一个在用户的汽车中))。

[0029] 环境100也包括一个或多个可控电子设备106(例如，电子设备106-1至106-N，在下文中被称为“可控设备”)。可控设备106的示例包括媒体设备(智能电视、扬声器系统、无线扬声器、机顶盒、媒体流设备、投射设备)和智能家居设备(例如、智能相机、智能恒温器、智能灯、智能危险检测器、智能门锁)。

[0030] 语音激活设备104和可控设备106通过通信网络110通信地耦合到语音辅助服务140(例如，到语音辅助服务140的语音辅助服务器系统112)。在一些实施方式中，语音激活设备104和可控设备106中的一个或多个通信地耦合到本地网络108，所述本地网络108通信地耦合到通信网络110；语音激活设备104和/或可控设备106经由本地网络108通信地耦合到通信网络110(并且，通过通信网络110，耦合到语音辅助服务器系统112)。在一些实施方式中，本地网络108是在网络接口(例如，路由器)处实现的局域网。通信地耦合到本地网络108的语音激活设备104和可控设备106也可以通过本地网络108彼此通信。

[0031] 可选地，语音激活设备104中的一个或多个通信地耦合到通信网络110并且不在本地网络108上。例如，这些语音激活设备不在与本地网络108相对应的Wi-Fi网络上，但是通

过蜂窝连接连接到通信网络110。在一些实施方式中，在本地网络108上的语音激活设备104与不在本地网络108上的语音激活设备104之间的通信通过语音辅助服务器系统112来完成。语音激活设备104(无论在本地网络108上还是在网络110上)被注册在语音辅助服务140的设备注册表118中并且因此为语音辅助服务器系统112所知。类似地，不在本地网络108上的语音激活设备104可以通过语音辅助服务器系统112与可控设备106进行通信。可控设备106(无论在本地网络108还是在网络110上)也被注册在设备注册表118中。在一些实施方式中，语音激活设备104与可控设备106之间的通信通过语音辅助服务器系统112。

[0032] 在一些实施方式中，环境100也包括一个或多个内容主机114。内容主机114可以是根据包括在用户语音输入或命令中的请求来流式传输或者以其它方式获得内容的远程内容源。内容主机114可以是语音辅助服务器系统112根据用户语音请求从其中检索信息的信息源。

[0033] 在一些实施方式中，可控设备106能够接收用于执行指定操作或者转变到指定状态的命令或请求(例如，来自语音激活设备104和/或语音辅助服务器系统112)并且将根据所接收到的命令或请求来执行操作或转变状态。

[0034] 在一些实施方式中，可控设备106中的一个或多个是被布置在操作环境100中以向一个或多个用户提供媒体内容、新闻和/或其它信息的媒体设备。在一些实施方式中，由媒体设备提供的内容被存储在本地内容源中，从远程内容源(例如，内容主机114)流式传输，或者在本地生成(例如，通过读取定制新闻简报、电子邮件、文本、本地天气报告等给操作环境100的一个或多个占用者的本地文本到语音处理器)。在一些实施方式中，媒体设备包括将媒体内容直接输出给受众(例如，一个或多个用户)的媒体输出设备，以及被联网以将媒体内容流式传输到媒体输出设备的投射设备。媒体输出设备的示例包括但不限于电视(TV)显示设备和音乐播放器。投射设备的示例包括但不限于机顶盒(STB)、DVD播放器、电视盒和媒体流设备，诸如谷歌的Chromecast™媒体流设备。

[0035] 在一些实施方式中，可控设备106也是语音激活设备104。在一些实施方式中，语音激活设备104也是可控设备106。例如，可控设备106可以包括到语音辅助服务140(例如，也可接收用户语音输入、对用户语音输入进行处理并且对用户语音输入作出响应的媒体设备)的语音接口。作为另一示例，语音激活设备104也可以根据语音输入中的请求或命令来执行特定操作并转变到特定状态(例如，也可播放流音乐的语音接口设备)。

[0036] 在一些实施方式中，语音激活设备104和可控设备106与具有相应的账户的用户相关联，或者与在用户域中具有相应的用户账户的多个用户(例如，相关用户组，诸如家庭中或组织中的用户;更一般地，主用户和一个或多个授权的附加用户)相关联。用户可以向语音激活设备104作出语音输入或语音命令。语音激活设备104从用户(例如，用户102)接收这些语音输入，并且语音激活设备104和/或语音辅助服务器系统112继续确定语音输入中的请求并且生成对该请求的响应。

[0037] 在一些实施方式中，包括在语音输入中的请求是对可控设备106执行操作(例如，播放媒体、暂停媒体、快进或倒回媒体、改变音量、改变屏幕亮度、改变灯亮度)或者转变到另一状态(例如，改变操作模式、打开或关闭、进入睡眠模式或者从睡眠模式唤醒)的命令或请求。

[0038] 在一些实施方式中，语音激活电子设备104通过以下步骤来对语音输入作出响应：

生成并提供对语音命令的口语响应(例如,响应于问题“现在是什么时间?”而说出当前时间);流式传输由用户请求的媒体内容(例如,“播放海滩男孩歌曲”);阅读为用户准备的新闻故事或每日新闻简报;播放存储在个人辅助设备上或者在本地网络上的媒体项;改变状态或者操作操作环境100内的一个或多个其它已连接设备(例如,将灯、电器或媒体设备打开/关闭、上锁/开锁、打开窗户等);或者经由网络110向服务器发出对应的请求。

[0039] 在一些实施方式中,所述一个或多个语音激活设备104被布置在操作环境100中以收集用于发起各种功能(例如,媒体设备的媒体播放功能)的音频输入。在一些实施方式中,这些语音激活设备104(例如,设备104-1至104-N)被布置为与可控设备104(例如,媒体设备)接近,例如,在与投射设备和媒体输出设备相同的房间中。可替选地,在一些实施方式中,语音激活设备104被布置在具有一个或多个智能家居设备而不是任何媒体设备的结构中。可替选地,在一些实施方式中,语音激活设备104被布置在具有一个或多个智能家居设备和一个或多个媒体设备的结构中。可替选地,在一些实施方式中,语音激活设备104被布置在没有联网的电子设备的位置中。另外,在一些实施方式中,结构中的房间或空间可以具有多个语音激活设备104。

[0040] 在一些实施方式中,语音激活设备104包括至少一个或多个麦克风、扬声器、处理器以及存储由该处理器执行的至少一个程序的存储器。扬声器被配置成允许语音激活设备104将语音消息和其它音频(例如,可听音调)递送到语音激活设备104位于操作环境100中所在的位置,从而广播音乐、报告音频输入处理的状态、与语音激活设备104的用户有对话或者将指令给予给语音激活设备104的用户。作为语音消息的替代方案,视觉信号也能用于向语音激活设备104的用户提供与音频输入处理的状态有关的反馈。当语音激活设备104是移动设备(例如,移动电话或平板计算机)时,其显示屏幕被配置成显示与音频输入处理的状态有关的通知。

[0041] 在一些实施方式中,语音激活设备104是连网以借助于语音辅助服务器系统112提供语音识别功能的语音接口设备。例如,语音激活设备104包括向用户提供音乐并且允许免视和免提访问语音辅助服务(例如,Google Assistant)的智能扬声器。可选地,语音激活设备104是台式或膝上型计算机、平板、包括麦克风的移动电话、包括麦克风并可选择地包括扬声器的投射设备、包括麦克风和扬声器的音频系统(例如,立体声系统、扬声器系统、便携式扬声器)、包括麦克风和扬声器的电视以及包括麦克风和扬声器并可选地包括显示器的汽车中的用户接口系统中的一个。可选地,语音激活设备104是简单且低成本的语音接口设备。一般地,语音激活设备104可以是能够连网并且包括麦克风、扬声器以及用于与语音辅助服务交互的程序、模块和数据的任何设备。考虑到语音激活设备104的简单性和低成本,语音激活设备104包括发光二极管(LED)的阵列而不是全显示屏幕,并且在LED上显示视觉图案以指示音频输入处理的状态。在一些实施方式中,LED是全色LED,并且可以采用LED的颜色作为待在LED上显示的视觉图案的一部分。例如,在下面参考图6描述使用LED来显示视觉图案以便传达信息或设备状态(例如,与指示聚焦会话是否已被发起、是活动的、已被扩展和/或多个用户中的哪些单独的用户与特定聚焦会话相关联有关的状态)的多个示例。在一些实施方式中,使用在与正在执行语音处理操作的语音激活设备相关联的常规显示器上示出的特征图像来显示指示语音处理操作的状态的视觉图案。

[0042] 在一些实施方式中,LED或其它视觉显示器用于传达多个参与电子设备的集体语

音处理状态。例如，在存在多个语音处理或语音接口设备(例如，如‘566应用的图4A中所示出的多个电子设备600;图1的多个语音激活设备104)的操作环境中，与相应的电子设备相关联的彩色LED组(例如，如图6中所示出的LED 404)可用于传达这些电子设备中的哪一个正在侦听用户，并且侦听设备中的哪一个是领导者(其中“领导者”设备一般地在对由用户发出的口语请求作出响应时起带头作用)。

[0043] 更一般地，‘566应用描述(例如，参见段落[0087]-[0100])用于使用LED的合集来在视觉上指示电子设备的各种语音处理状态(诸如“热词检测状态和侦听状态”、“思维模式或工作模式”以及“响应模式或说话模式”)的“LED设计语言”。在一些实施方式中，本文中所描述的语音处理操作的唯一状态是根据‘566应用的“LED设计语言”的一个或多个方面使用LED组来表示的。这些视觉指示器也可与由正在执行语音处理操作的电子设备所生成的一个或多个可听指示器组合。结果得到的音频和/或视觉指示器将使得语音交互环境中的用户能够理解该环境中的各种语音处理电子设备的状态并且以自然直观的方式有效地与这些设备交互。

[0044] 在一些实施方式中，当语音激活设备104的语音输入用于经由投射设备来控制媒体输出设备时，语音激活设备104有效地实现对支持投射的媒体设备的新的控制水平。在特定示例中，语音激活设备104包括具有远场语音接入的休闲享受扬声器并且充当语音辅助服务的语音接口设备。语音激活设备104能被布置在操作环境100中的任何区域中。当多个语音激活设备104分布在多个房间中时，它们变成被同步以从这些房间提供语音输入的投射音频接收器。

[0045] 具体地，在一些实施方式中，语音激活设备104包括具有连接到语音激活语音辅助服务(例如，Google Assistant)的麦克风的Wi-Fi扬声器。用户可经由语音激活设备104的麦克风发出媒体播放请求，并且要求语音辅助服务在语音激活设备104它本身上或者在另一已连接媒体输出设备上播放媒体内容。例如，用户可通过向Wi-Fi扬声器说“OK Google，play cat videos on my Living room TV(OK Google，在我的客厅电视上播放猫视频)”来发出媒体播放请求。语音辅助服务然后通过使用默认或指定的媒体应用在所请求的设备上播放所请求的媒体内容来履行媒体播放请求。

[0046] 在一些实施方式中，用户可经由语音激活设备104的麦克风发出与在显示设备上已经播放或者正在播放的媒体内容有关的语音请求(例如，用户可要求与媒体内容有关的信息，通过在线商店购买媒体内容，或者组成并发出与媒体内容有关的社交帖子)。

[0047] 在一些实施方式中，用户可能想随着他们移动通过住所而与他们进行当前媒体会话并且可从语音激活设备104中的一个或多个请求这样的服务。这需要语音辅助服务140将当前媒体会话从第一投射设备转移到未直接连接到第一投射设备或者不知道第一投射设备的存在的第二投射设备。继媒体内容转移之后，在在第一输出设备上放弃了媒体内容的播放的情况下耦合到第二投射设备的第二输出设备继续从音乐乐曲或视频剪辑内的确切点起播放耦合到第一投射设备的第一输出设备上先前播放的媒体内容。在一些实施方式中，接收到转移媒体会话的请求的语音激活设备104可满足该请求。在一些实施方式中，接收到转移媒体会话的请求的语音激活设备104将该请求中继到另一设备或系统(例如，语音辅助服务器系统112)以供处理。

[0048] 另外，在一些实施方式中，用户可以经由语音激活设备104的麦克风发出对信息的

或对动作或操作的执行的请求。所请求的信息可以是个人的(例如,用户的电子邮件、用户的日历事件、用户的航班信息等)、非个人的(例如,比赛分数、新闻故事等)或其之间的(例如,用户偏爱的团队或比赛的分数、来自用户的优选源的新闻故事等)。所请求的信息或动作/操作可以涉及对个人信息的访问(例如,利用由用户提供的支付信息购买数字媒体项、购买物理商品)。语音激活设备104以对用户的语音消息响应对请求作出响应,其中响应可以包括例如对履行请求的附加信息的请求、已履行了请求的确认、不可履行请求的通知等。

[0049] 在一些实施方式中,除语音激活设备104和可控设备106之中的媒体设备之外,操作环境100还可以包括可控设备106之中的一个或多个智能家居设备。集成智能家居设备包括在智能家居网络中与彼此并且/或者与中央服务器或云计算系统无缝地集成以提供各种有用的智能家居功能的智能多感测连网的设备。在一些实施方式中,智能家居设备被布置在操作环境100的与投射设备和/或输出设备相同的位置处,并且因此,位于与投射设备和输出设备接近或者相对于投射设备和输出设备在已知距离上。

[0050] 操作环境100中的智能家居设备可以包括但不限于一个或多个智能多感测连网的恒温器、一个或多个智能连网的多感测危险检测器、一个或多个智能多感测连网的入口接口设备(在下文中被称为“智能门铃”和“智能门锁”)以及一个或多个智能多感测连网的警报系统、一个或多个智能多感测连网的相机系统、一个或多个智能多感测连网的墙壁开关、一个或多个智能多感测连网的电源插座和一个或多个智能多感测连网的灯。在一些实施方式中,图1的操作环境100中的智能家居设备包括多个智能多感测连网的电器(在下文中被称为“智能电器”),诸如冰箱、电炉、烤箱、电视、洗衣机、烘干机、灯、立体声系统、对讲系统、车库门开启器、落地风扇、吊扇、壁式空调、水池加热器、灌溉系统、安全系统、空间加热器、窗户AC单元、电动通风口等。在一些实施方式中,这些智能家居设备类型中的任何一种可配备有如本文中所描述的麦克风和一个或多个语音处理能力,以便整个地或部分地对来自占用者或用户的语音请求作出响应。

[0051] 在一些实施方式中,可控设备104和语音激活设备104中的每一个能够与其它可控设备106、语音激活电子设备104、中央服务器或云计算系统和/或连网的其它设备(例如,客户端设备)进行数据通信和信息共享。可以使用各种定制或标准无线协议(例如,IEEE 802.15.4、Wi-Fi、ZigBee、6LoWPAN、Thread、Z-Wave、Bluetooth Smart、ISA100.11a、WirelessHART、MiWi等)和/或各种定制或标准有线协议中的任一种(例如,以太网、HomePlug等)或者任何其它适合的通信协议(包括在本文档的提交日期时尚未开发的通信协议)来执行数据通信。

[0052] 通过通信网络(例如,互联网)110,可控设备106和语音激活设备104可以与服务器系统(在本文中也被称作中央服务器系统和/或云计算系统)进行通信。可选地,服务器系统可以与和可控设备相关联的制造商、支持实体或服务提供者以及向用户显示的媒体内容相关联。因此,服务器系统包括对由语音激活设备104收集到的音频输入进行处理的语音辅助服务器112、提供所显示的媒体内容的一个或多个内容主机114(可选地基于分布式设备终端创建虚拟用户域的云投射服务服务器)以及保持虚拟用户环境中的分布式设备终端的记录的设备注册表118。分布式设备终端的示例包括但不限于可控设备106、语音激活设备104和媒体输出设备。在一些实施方式中,这些分布式设备终端链接到虚拟用户域中的用户账户(例如,谷歌用户账户)。应该了解,可在语音激活设备104处、在语音辅助服务器112处、在

另一智能家居设备(例如,中枢设备或可控设备106)处或者在上述的全部或子集的某个组合处在本地执行由语音激活设备104收集到的音频输入的处理(包括对那些输入的响应的生成)。

[0053] 应当了解,在一些实施方式中语音激活设备104也在没有智能家居设备的环境中起作用。例如,即使在智能家居设备不存在的情况下,语音激活设备104也可对对信息的或动作的执行和/或发起或者控制各种媒体播放功能的用户请求作出响应。语音激活设备104也可在各式各样的环境(包括但不限于车辆、船舶、商业或制造环境)中起作用。

[0054] 在一些实施方式中,语音激活设备104通过包括热词(也被称作"唤醒词")的语音输入被"唤醒"(例如,以激活语音激活设备104上用于语音辅助服务的接口、以将语音激活设备104置于语音激活设备104准备好接收对语音请求服务的语音请求的状态中)。在一些实施方式中,如果语音激活设备104相对于语音输入的接收在至少预定义量的时间(例如,5分钟)内一直空闲则语音激活设备104需要唤醒;预定义量的时间对应于在语音接口会话或对话超时之前允许的空闲时间的量。热词可以是词或短语,并且可以是预定义默认的和/或可以由用户定制(例如,用户可以将特定语音激活设备104的昵称设置为该设备的热词)。在一些实施方式中,可以存在可唤醒语音激活设备104的多个热词。用户可以说出热词,等待来自语音激活设备104的肯定应答(acknowledgement)响应(例如,语音激活设备104输出问候语),然后作出第一语音请求。可替选地,用户可以在一个语音输入中组合热词和第一语音请求(例如,语音输入包括后面是语音请求的热词)。

[0055] 在一些实施方式中,语音激活设备104与根据一些实施方式的操作环境100的可控设备106(例如,媒体设备、智能家居设备)、客户端设备或服务器系统交互。语音激活设备104被配置成从接近语音激活设备104的环境接收音频输入。可选地,语音激活设备104存储音频输入并且至少部分地在本地对这些音频输入进行处理。可选地,语音激活设备104经由通信网络110将所接收到的音频输入或经部分地处理的音频输入传送到语音辅助服务器系统112以用于进一步处理。语音激活设备104或语音辅助服务器系统112确定在音频输入中是否存在请求以及该请求是什么,确定并生成对该请求的响应,并且将该请求传送到一个或多个可控设备106。接收到响应的可控设备106被配置成根据响应执行操作或者改变状态。例如,媒体设备被配置成根据对音频输入中的请求的响应从一个或多个内容主机114获得媒体内容或互联网内容以供显示在耦合到该媒体设备的输出设备上。

[0056] 在一些实施方式中,可控设备106和语音激活设备104在用户域中彼此链接,并且更具体地,经由用户域中的用户账户彼此相关联。关于可控设备106(无论在本地网络108上还是在网络110上)和语音激活设备104(无论在本地网络108上还是在网络110上)的信息被与用户账户相关联地存储在设备注册表118中。在一些实施方式中,存在用于可控设备106的设备注册表和用于语音激活设备104的设备注册表。可控设备注册表可以引用在用户域中相关联的语音激活设备注册表中的设备,并且反之亦然。

[0057] 在一些实施方式中,语音激活设备104(和一个或多个投射设备)中的一个或多个以及可控设备106中的一个或多个经由客户端设备103被委用(commissioned)给语音辅助服务140。在一些实施方式中,语音激活设备104不包括任何显示屏幕,并且依靠客户端设备103来在委用过程期间提供用户接口,并且类似地对于可控设备106也一样。具体地,客户端设备103被安装有使得用户接口能够促进被设置在接近客户端设备的新语音激活设备104

和/或可控设备106的委用的应用。用户可以在客户端设备103的用户接口上发送用于对需要被委用的新电子设备104/106发起委用过程的请求。在接收到委用请求之后，客户端设备103与需要被委用的新电子设备104/103建立短距离通信链路。可选地，该短距离通信链路是基于近场通信(NFC)、蓝牙、低功耗蓝牙(BLE)等而建立的。客户端设备103然后将与无线局域网(WLAN)(例如，本地网络108)相关联的无线配置数据传达给新电子设备104/106。无线配置数据包括至少WLAN安全代码(即，服务集标识符(SSID)口令)，并且可选地包括SSID、网际协议(IP)地址、代理配置和网关配置。在经由短距离通信链路接收到无线配置数据之后，新电子设备104/106对无线配置数据进行解码和恢复，并且基于无线配置数据加入WLAN。

[0058] 在一些实施方式中，在客户端设备103上显示的用户界面上录入附加的用户域信息，并且用于将新电子设备104/106链接到用户域中的账户。可选地，附加用户域信息经由短距离通信链路结合无线通信数据被传送到新电子设备104/106。可选地，在新设备加入WLAN之后，附加的用户域信息经由WLAN被传送到新电子设备104/106。

[0059] 一旦已经将电子设备104/106委用到用户域中，就可以经由多个控制路径来控制其它设备及其相关联的活动。根据一个控制路径，安装在客户端设备103上的应用用于控制其它设备及其相关联的活动(例如，媒体播放活动)。可替选地，根据另一控制路径，电子设备104/106用于实现对其它设备及其相关联的活动的免视和免提控制。

[0060] 在一些实施方式中，语音激活设备104和可控设备106可以由用户(例如，由在用户域中与设备相关联的主用户)指派昵称。例如，可以给客厅中的扬声器设备指派昵称“客厅扬声器”。以这种方式，用户可以通过说出设备的昵称在语音输入中更容易地指代设备。在一些实施方式中，设备昵称和到对应设备的映射被存储在语音激活设备104(其将存储仅仅与和语音激活设备相同的用户相关联的设备的昵称)和/或语音辅助服务器系统112(其将存储与不同的用户相关联的设备的设备昵称)处。例如，语音辅助服务器系统112存储许多设备昵称以及跨越不同的设备和用户的映射，而与特定用户相关联的语音激活设备104下载与该特定用户相关联的设备的昵称和映射以用于本地存储。

[0061] 在一些实施方式中，用户可以将语音激活设备104和/或可控设备106中的一个或多个聚组成由用户创建的设备组。可以给予该组一名称，并且可以按组名称引用该组设备，类似于按昵称参考单独的设备。类似于设备昵称，可以将设备组和组名称存储在语音激活设备104和/或语音辅助服务器系统112处。

[0062] 来自用户的语音输入可以针对该语音输入中的请求显式地指定目标可控设备106或目标设备组。例如，用户可以发出语音输入“在客厅扬声器上播放古典音乐”。该语音输入中的目标设备是“客厅扬声器”；该语音输入中的请求是让“客厅扬声器”播放古典音乐的请求。作为另一示例，用户可以发出语音输入“在住所扬声器上播放古典音乐”，其中“住所扬声器”是设备组的名称。该语音输入中的目标设备组是“住所扬声器”；该语音输入中的请求是让组“住所扬声器”中的设备播放古典音乐的请求。

[0063] 来自用户的语音输入可能不具有目标设备或设备组的显式指定；在语音输入中不存在按名称对目标设备或设备组的引用。例如，紧跟以上示例语音输入“在客厅扬声器上播放古典音乐”之后，用户可以发出后续语音输入“暂停”。该语音输入不包括对暂停操作的请求的目标设备指定。在一些实施方式中，该语音输入中的目标设备指定可以是模糊不清的。

例如，用户可能已不完全地发出设备名称。在一些实施方式中，如下所述，可以将目标设备或设备组指派给显式目标设备指定不存在或者目标设备指定模糊不清的语音输入。

[0064] 在一些实施方式中，当语音激活设备104接收到具有目标设备或设备组的显式指定的语音输入时，语音激活设备104相对于所指定的目标设备或设备组建立聚焦会话。在一些实施方式中，语音激活设备104针对聚焦会话存储会话开始时间(例如，开始聚焦会话所基于的语音输入的时间戳)，并且作为聚焦会话的对焦设备存储所指定的目标设备或设备组。在一些实施方式中，语音激活设备104也在聚焦会话中记录后续语音输入。语音激活设备104记录至少聚焦会话中最近的语音输入并且也可选地记录并保持聚焦会话内前面的语音输入。在一些实施方式中，语音辅助服务器系统112建立聚焦会话。在一些实施方式中，可以通过显式地指定不同的目标设备或设备组的语音输入来结束聚焦会话。

[0065] 当相对于设备的聚焦会话是活动的并且语音激活设备接收到语音输入时，语音激活设备104相对于该语音输入作出一个或多个确定。在一些实施方式中，确定包括：语音输入是否包括显式目标设备指定、语音输入中的请求是否是可由对焦设备履行的请求以及与聚焦会话中的最后语音输入的时间和/或会话开始时间相比的语音输入的时间。如果语音输入不包括显式目标设备指定，包括可由对焦设备履行的请求，并且满足相对于聚焦会话中的最后语音输入的时间和/或会话开始时间的预定义时间准则，则对焦设备被指派为用于语音输入的目标设备。在下面对有关聚焦会话的进一步细节进行描述。

[0066] 操作环境中的设备

[0067] 图2是图示被作为语音接口应用来在根据一些实施方式的操作环境(例如，操作环境100)中收集用户语音命令的示例语音激活设备104的框图。语音激活设备104通常包括一个或多个处理单元(CPU)202、一个或多个网络接口204、存储器206以及用于互连这些组件(有时被称作芯片集)的一个或多个通信总线208。语音激活设备104包括促进用户输入的一个或多个输入设备210，诸如按钮212、触摸感测阵列214和一个或多个麦克风216。语音激活设备104也包括一个或多个输出设备218，包括一个或多个扬声器220，可选地包括LED阵列222，并且可选地包括显示器224。在一些实施方式中，LED阵列222是全色LED阵列。在一些实施方式中，语音激活设备104取决于该设备的类型而具有LED阵列222或显示器224或两者。在一些实施方式中，语音激活设备104也包括位置检测设备226(例如，GPS模块)和一个或多个传感器228(例如，加速度计、陀螺仪、光传感器等)。

[0068] 存储器206包括高速随机存取存储器，诸如DRAM、SRAM、DDR RAM或其它随机存取固态存储器设备；并且可选地，包括非易失性存储器，诸如一个或多个磁盘存储设备、一个或多个光盘存储设备、一个或多个闪速存储器设备或一个或多个其它非易失性固态存储设备。存储器206可选地包括远离一个或多个处理单元202的一个或多个存储设备。存储器206或可替选地存储器206内的非易失性存储器包括非暂时性计算机可读存储介质。在一些实施方式中，存储器206或存储器206的非暂时性计算机可读存储介质存储以下程序、模块和数据结构，或者其子集或超集：

[0069] ●包括用于处理各种基本系统服务并用于执行硬件相关任务的过程的操作系统232；

[0070] ●用于经由一个或多个网络接口204(有线的或无线的)和一个或多个网络110(诸如互联网、其它广域网、局域网(例如，本地网络108)、城域网等)将语音激活设备104连接到

其它设备(例如,语音辅助服务140、一个或多个可控设备106、一个或多个客户端设备103和其它语音激活设备104)的网络通信模块234;

[0071] ●用于经由一个或多个输入设备接收输入并且使得能够经由一个或多个输出设备218在语音激活设备104处呈现信息的输入/输出控制模块236,包括:

[0072] ○用于对在语音激活设备104周围的环境中收集的音频输入或语音消息进行处理或者准备所收集到的音频输入或语音消息以供在语音辅助服务器系统112处处理的语音处理模块238;

[0073] ○用于根据语音激活设备104的设备状态在LED 222上生成视觉图案的LED控制模块240;以及

[0074] ○用于感测语音激活设备104的顶面(例如,在触摸传感器阵列214上)的触摸事件的触摸感测模块242;

[0075] ●用于存储至少与语音激活设备104相关联的数据的语音激活设备数据244,包括:

[0076] ○用于存储与语音激活设备104它本身相关联的信息的语音设备设置246,包括公共设备设置(例如,服务层、设备模型、存储容量、处理能力、通信能力等)、用户域中的一个或多个用户账户的信息、设备昵称和设备组、有关在对待非注册用户时的限制的设置以及与由LED 222显示的一个或多个视觉图案相关联的显示规格;以及

[0077] ○用于存储音频信号、语音消息、响应消息以及与语音激活设备104的语音接口功能有关的其它数据的语音控制数据248;

[0078] ●用于执行包括在由语音辅助服务器系统112生成的语音请求响应中的指令并且在一些实施方式中生成对某些语音输入的响应的响应模块250;以及

[0079] ●用于相对于设备建立、管理和结束聚焦会话的聚焦会话模块252。

[0080] ●

[0081] 在一些实施方式中,语音处理模块238包括以下模块(未示出):

[0082] ●用于对向语音激活设备104提供语音输入的用户进行识别并消除歧义的用户识别模块;

[0083] ●用于确定语音输入是否包括用于唤醒语音激活设备104的热词并且在语音输入中识别此类热词的热词识别模块;以及

[0084] ●用于确定包含在语音输入中的用户请求的请求识别模块。

[0085] ●

[0086] 在一些实施方式中,存储器206也存储未完成(outstanding)聚焦会话的聚焦会话数据254,包括以下各项:

[0087] ●用于存储在未完成聚焦会话中在焦点上的设备或设备组的标识符

[0088] (例如,设备的设备昵称、设备组名称、MAC地址)的会话对焦设备256;

[0089] ●用于存储未完成聚焦会话开始的时间戳的会话开始时间258;以及

[0090] ●用于存储聚焦会话中的先前的请求或命令(包括至少最近的请求/命令)的日志的会话命令历史260。所述日志至少包括所记录的先前的请求/命令的时间戳。

[0091] 上面标识的元件中的每一个可以被存储在先前提及的存储器设备中的一个或多个中,并且对应于用于执行上述的功能的指令集。上面标识的模块或程序(即,指令集)未必

作为单独的软件程序、过程、模块或数据结构被实现，并且因此可以在各种实施方式中组合或者以其它方式重新安排这些模块的各个子集。在一些实施方式中，存储器206可选地存储上面所标识的模块和数据结构的子集。此外，存储器206可选地存储上面未描述的附加模块和数据结构。在一些实施方式中，存储在存储器206中的程序、模块和/或数据的子集可被存储在语音辅助服务器系统112上和/或由语音辅助服务器系统112执行。

[0092] 在一些实施方式中，上述的存储器206中的模块中的一个或多个是模块的语音处理库的一部分。语音处理库可以被实现并嵌入在各式各样的设备上。图3A至图3B是图示根据一些实施方式的操作环境(例如，操作环境100)的语音辅助服务140的示例语音辅助服务器系统112的框图。服务器系统112通常包括一个或多个处理单元(CPU)302、一个或多个网络接口304、存储器306以及用于互连这些组件(有时被称作芯片集)的一个或多个通信总线308。服务器系统112可以包括促进用户输入的一个或多个输入设备310，诸如键盘、鼠标、语音命令输入单元或麦克风、触摸屏显示器、触敏输入板、手势捕获相机或其它输入按钮或控件。此外，服务器系统112可以使用麦克风和语音识别或相机和手势识别来补充或者替换键盘。在一些实施方式中，服务器系统112包括用于捕获例如印刷在电子设备上的图形系列代码的图像的一个或多个相机、扫描器或照片传感器单元。服务器系统112也可以包括使得能够呈现用户界面和显示内容的一个或多个输出设备312，包括一个或多个扬声器和/或一个或多个视觉显示器。

[0093] 存储器306包括高速随机存取存储器，诸如DRAM、SRAM、DDR RAM或其它随机存取固态存储器设备；并且可选地，包括非易失性存储器，诸如一个或多个磁盘存储设备、一个或多个光盘存储设备、一个或多个闪速存储器设备或一个或多个其它非易失性固态存储设备。存储器306可选地包括远离一个或多个处理单元302的一个或多个存储设备。存储器306或可替选地存储器306内的非易失性存储器包括非暂时性计算机可读存储介质。在一些实施方式中，存储器306或存储器306的非暂时性计算机可读存储介质存储以下程序、模块和数据结构，或者其子集或超集：

[0094] ●包括用于处理各种基本系统服务并用于执行硬件相关任务的过程的操作系统316；

[0095] ●用于经由一个或多个网络接口304(有线的或无线的)和一个或多个网络110(诸如互联网、其它广域网、局域网、城域网等)将服务器系统112连接到其它设备(例如，客户端设备103、可控设备106、语音激活设备104)的网络通信模块318；

[0096] ●用于使得能够在客户端设备处呈现信息的用户界面模块320(例如，用于呈现应用322-328、微件(widget)、网站及其web页面和/或游戏、音频和/或视频内容、文本等的图形用户界面)；

[0097] ●在服务器侧执行的命令执行模块321(例如，游戏、社交网络应用、智能家居应用和/或用于控制客户端设备103、可控设备106、语音激活设备104和智能家居设备并且审查由此类设备所捕获的数据的其它基于web或非web的应用)，包括以下各项中的一个或多个：

[0098] ○被执行来提供与投射设备相关联的设备供应、设备控制和用户账户管理的服务器侧功能性的投射设备应用322；

[0099] ○被执行来提供与对应的媒体源相关联的媒体显示和用户账户管理的服务器侧功能性的一个或多个媒体播放器应用324；

[0100] ○被执行来提供对应的智能家居设备的设备提供、设备控制、数据处理和数据审查的服务器侧功能性的一个或多个智能家居设备应用326;以及

[0101] ○被执行来安排从语音激活设备104接收到的语音消息的语音处理或者直接对语音消息进行处理以提取用户语音命令和该用户语音命令的一个或多个参数(例如,投射设备或另一语音激活设备104的指定)的语音辅助应用328;以及

[0102] ●存储至少与(例如,在自动媒体输出模式和跟随模式下)媒体显示的自动控制相关联的数据以及其它数据的服务器系统数据330,所述其它数据包括以下各项中的一个或多个:

[0103] ○用于存储与一个或多个客户端设备相关联的信息的客户端设备设置332,包括公共设备设置(例如,服务层、设备模型、存储容量、处理能力、通信能力等)以及用于自动媒体显示控制的信息;

[0104] ○用于存储与投射设备应用322的用户账户相关联的信息的投射设备设置334,包括账户访问信息、设备设置(例如,服务层、设备模型、存储容量、处理能力、通信能力等)的信息以及用于自动媒体显示控制的信息中的一个或多个;

[0105] ○用于存储与一个或多个媒体播放器应用324的用户账户相关联的信息的媒体播放器应用设置336,包括账户访问信息、媒体内容类型的用户偏好、审查历史数据以及用于自动媒体显示控制的信息中的一个或多个;

[0106] ○用于存储与智能家居应用326的用户账户相关联的信息的智能家居设备设置338,包括账户访问信息、一个或多个智能家居设备的信息(例如,服务层、设备模型、存储容量、处理能力、通信能力等)中的一个或多个;

[0107] ○用于存储与语音辅助应用328的用户账户相关联的信息的语音辅助数据340,包括账户访问信息、一个或多个语音激活设备104的信息(例如,服务层、设备模型、存储容量、处理能力、通信能力等)中的一个或多个;

[0108] ○用于存储与用户域中的用户相关联的信息的用户数据342,包括用户的订阅(例如,音乐流服务订阅、视频流服务订阅、时事通讯订阅)、用户设备(例如,在与相应的用户、设备昵称、设备组相关联的设备注册表118中注册的设备)、用户账户(例如,用户的电子邮件账户、日历账户、金融账户)和其它用户数据;

[0109] ○用于存储用户域中的用户的语音简档的用户语音简档344,包括例如用户的语音模型或语音指纹以及用户的舒适音量水平阈值;以及

[0110] ○用于存储多个设备的聚焦会话数据的聚焦会话数据346。

[0111] ●用于管理设备注册表118的设备注册模块348;

[0112] ●用于对在电子设备104周围的环境中收集的音频输入或语音消息进行处理的语音处理模块350;以及

[0113] ●用于相对于设备建立、管理和结束聚焦会话的聚焦会议模块352。

[0114] 参考图3B,在一些实施方式中,存储器306也存储一个或多个未完成聚焦会话3462-1至3462-M的聚焦会话数据346,包括以下各项:

[0115] ●用于存储建立了聚焦会话的设备的标识符的会话源设备3464;

[0116] ●用于存储在未完成聚焦会话中在焦点上的存储设备或设备组的标识符(例如,设备的设备昵称、设备组名称、MAC地址)的会话对焦设备3466;

[0117] ●用于存储未完成聚焦会话开始的时间戳的会话开始时间3468;以及

[0118] ●用于存储聚焦会话中的先前的请求或命令(包括至少最近的请求/命令)的日志的会话命令历史3470。

[0119] 在一些实施方式中,语音辅助服务器系统112主要负责语音输入的处理,并且因此上面参考图2所描述的存储器206中的程序、模块和数据结构中的一个或多个被包括在存储器306中的相应的模块中(例如,与语音处理模块238包括在一起的程序、模块和数据结构被包括在语音处理模块350中)。语音激活设备104要么将捕获的语音输入传送到语音辅助服务器系统112以用于处理,要么首先对语音输入进行预处理并且将经预处理的语音输入传送到语音辅助服务器系统112以用于处理。在一些实施方式中,语音辅助服务器系统112和语音激活设备104具有有关语音输入的处理的一些共享的和一些划分的责任,并且图2所示的程序、模块和数据结构可以被包括在语音辅助服务器系统112和语音激活设备104两者中或者在语音辅助服务器系统112和语音激活设备104之间进行划分。图2所示的其它程序、模块和数据结构或其类似物也可以被包括在语音辅助服务器系统112中。

[0120] 上面标识的元件中的每一个可以被存储在先前提及的存储器设备中的一个或多个中,并且对应于用于执行上述的功能的指令集。上面标识的模块或程序(即,指令集)未必作为单独的软件程序、过程、模块或数据结构被实现,并且因此可以在各种实施方式中组合或者以其它方式重新安排这些模块的各个子集。在一些实施方式中,存储器306可选地存储上面标识的模块和数据结构的子集。此外,存储器306可选地存储上面未描述的附加模块和数据结构。

[0121] 示例聚焦会话

[0122] 图4A至图4D图示根据一些实施方式的聚焦会话的示例。在具有语音激活设备104和多个可控设备106的操作环境(例如,操作环境100)中,当环境中的用户作出将可控设备106中的一个指定为目标设备的语音输入时,可以与作为对焦设备的目标设备建立聚焦会话。

[0123] 图4A示出操作环境(例如,操作环境100)中的语音激活设备404(例如,语音激活设备104)以及三个可控设备406、408和410(例如,可控设备106)。这些设备可以在与用户402相同的空间中(例如,在相同房间中)或者遍布用户所位于的结构。设备406是昵称为“主卧室扬声器”的扬声器系统。设备408是昵称为“客厅电视”的媒体设备。设备410是昵称为“游戏室电视”的媒体设备。此刻不存在聚焦会话;聚焦会话418是空的。

[0124] 用户402发出语音输入403“play cat videos on game room TV(在游戏室电视上播放猫视频)”,并且语音激活设备404接收该语音输入。语音激活设备404确定语音输入403中的请求是播放猫视频的请求,并且目标设备是在语音输入403中显式地指定的“game room TV(游戏室电视)”设备410。如图4B所示,在语音激活设备404处建立其中对焦设备为“游戏室电视”设备410的会话418。播放猫视频的命令(由设备404或语音辅助服务器系统112)发送到“游戏室电视”设备410,并且设备410执行操作416。

[0125] 参考图4C,随后,当与在焦点上的“游戏室电视”410的会话418是活动的并且设备410正在执行操作416时,用户402发出另一语音输入“暂停”420。语音激活设备404确定语音输入420是否包括目标设备的指定以及语音输入420中的请求是否可由对焦设备410执行。在特定语音输入420“暂停”的情况下,语音激活设备404确定语音输入420不包括目标设备

的指定并且语音输入中的请求(无论正在播放都“暂停”)可由对焦设备执行。在一些实施方式中,确定语音输入420是否包括目标设备的指定包括在语音输入中查找与设备昵称的匹配(例如,对语音输入执行语音到文本识别并且对该文本进行解析以查找设备昵称)。在一些实施方式中,确定语音输入中的请求是否可由对焦设备执行包括确定语音输入中的请求是什么并且就与会话中的最后命令的一致性而将该请求与当前聚焦会话418的命令历史(例如,历史260)相比较(例如,“暂停音乐”请求与为“暂停音乐”的最近命令不一致),以及就与对焦设备的能力的一致性而对请求进行比较(例如,“暂停音乐”请求与智能灯的能力不一致)。

[0126] 在一些实施方式中,语音激活设备404也确定语音输入420是否满足一个或多个聚焦会话维持准则。在一些实施方式中,聚焦会话维持准则是语音输入420的时间戳在从活动会话中的最后语音输入403的时间戳起的一定时间内(例如,在前面的第一语音输入的一定时间内接收到第二语音输入)。在一些实施方式中,对于此准则来说存在多个时间阈值。例如,可以存在第一较短时间阈值(例如,20分钟)和第二较长时间阈值(例如,4小时)。如果在最后语音输入403的第一较短阈值内接收到语音输入420,并且满足上面的另一个准则,则对焦设备被设置为语音输入420的目标设备(并且,在一些实施方式中,在将语音输入420传送到语音辅助服务器系统112以用于处理时也传送此目标设备设置)。例如,语音输入420被确定为不包括目标设备指定并且请求“暂停”与最后命令“播放猫视频”一致。如果在语音输入403的较短时间阈值内接收到语音输入420,则对焦设备“游戏室电视”设备410被设置为语音输入420的目标设备,并且正在“游戏室电视”设备410处执行的操作416是根据语音输入420暂停猫视频,如图4D所示。

[0127] 如果在最后语音输入403的第一较短阈值之后并且在最后语音输入403的第二较长阈值内接收到语音输入420,并且满足上面的另一个准则,则语音激活设备404输出用于从用户请求对焦设备为语音输入420的期望目标设备的确认的语音提示。语音激活设备404在接收到对焦设备是期望目标设备的确认时,维持会话418并且将对焦设备设置为语音输入420的目标设备(并且,在一些实施方式中,在将语音输入420传送到语音辅助服务器系统112以用于处理时也传送此目标设备设置)。如果用户不确认目标设备,则语音激活设备404可以请求用户提供目标设备指定,请求用户再次说出语音输入但是包括目标设备指定,并且/或者结束会话418。在一些实施方式中,如果在自最后语音输入403起的第二较长阈值之后接收到语音输入420或者不满足上述的另一个准则,则会话418结束。在一些实施方式中,这些时间阈值的值被存储在存储器206和/或存储器306中。在语音输入之间经过的时间被与这些阈值相比较。

[0128] 在一些实施方式中,语音输入中的显式地指定的目标设备的缺乏以及语音输入中的请求与最后语音输入并与对焦设备的能力的一致性也被认为是聚焦会话维持准则。

[0129] 示例过程

[0130] 图5是图示根据一些实施方式的对用户的语音输入作出响应的方法500的流程图。在一些实施方式中,在具有一个或多个麦克风、扬声器、一个或多个处理器以及存储由所述一个或多个处理器执行的一个或多个程序的存储器的第一电子设备(例如,语音激活设备104)处实现方法500。此第一电子设备是通信地耦合(例如,通过网络110)到公共网络服务(例如,语音辅助服务140)的已连接电子设备(例如,与用户账户相关联的语音激活设备104

和可控设备106;与特定语音激活设备104相关联的可控设备106)的本地组的成员。

[0131] 第一电子设备接收(502)包括对第一操作的请求的第一语音命令。例如,语音激活设备404接收第一语音输入403。

[0132] 第一电子设备从已连接电子设备的本地组当中确定用于第一操作的第一目标设备(504)。语音激活设备404从设备406、408和410当中确定(例如,基于由语音处理模块238处理)语音输入403的目标设备(或设备组)。语音激活设备404将语音输入403中的目标设备指定“游戏室电视”识别为“游戏室电视”设备410。

[0133] 第一电子设备相对于第一目标设备(或设备组)建立聚焦会话(506)。语音激活设备404(例如,聚焦会话模块252)与作为对焦设备的“游戏室电视”设备410建立聚焦会话418。

[0134] 第一电子设备经由公共网络服务的操作使第一操作由第一目标设备(或设备组)执行(508)。语音激活设备404或语音辅助服务器系统112经由语音辅助服务140向设备410传送用于执行语音输入403中所请求的操作的命令。

[0135] 第一电子设备接收包括对第二操作的请求的第二语音命令(510)。语音激活设备404接收第二语音输入420。

[0136] 第一电子设备确定第二语音命令不包括第二目标设备(或设备组)的显式指定(512)。语音激活设备404确定(例如,基于由语音处理模块238处理)语音输入420的目标设备,并且识别语音输入420不包括目标设备指定。

[0137] 第一电子设备确定第二操作可由第一目标设备(或设备组)执行(514)。语音激活设备404确定语音输入420中所请求的操作能够由对焦设备410执行并且与语音输入403中所请求的且正由对焦设备410执行的最后操作一致。

[0138] 第一电子设备确定第二语音命令是否满足一个或多个聚焦会话维持准则(516)。语音激活设备404确定是否在语音输入403的一定时间内接收到语音输入420。

[0139] 根据第二语音命令满足聚焦会话维持准则的确定,第一电子设备经由公共网络的操作使第二操作由第一目标设备(或设备组)执行(518)。语音激活设备404确定在语音输入403的第一较短时间阈值内接收到语音输入420,并且根据该确定将语音输入420的目标设备设置为对焦设备410。语音激活设备404或语音辅助服务器系统112经由语音辅助服务140向设备410传送用于执行语音输入420中所请求的操作的命令。

[0140] 在一些实施方式中,从已连接电子设备的本地组当中确定用于第一操作的第一目标设备包括从所述第一语音命令获得所述第一目标设备的显式指定。语音激活设备404可以对语音输入403进行预处理以确定语音输入403是否包括目标设备的显式指定。可替选地,话音激活设备404可以从对语音输入403进行了处理的语音辅助服务器系统112接收目标设备的显式指定。

[0141] 在一些实施方式中,从已连接电子设备的本地组当中确定用于第一操作的第一目标设备包括确定第一语音命令不包括第一目标设备的显式指定,确定第一操作可由已连接电子设备的本地组当中的第二电子设备执行,并且选择所述第二电子设备作为所述第一目标设备。如果第一语音输入不包括目标的显式指定,但是包括在第一语音输入内的请求是可由组内的单个设备执行的请求(例如,视频相关命令并且在该组中仅有一个支持视频的设备),则该单个设备被设置为用于第一语音输入的目标设备。另外,在一些实施方式中,如

果除了语音激活设备之外还存在仅一个可控设备，则该可控设备是未显式地指定目标设备并且其请求的操作可由该可控设备执行的语音输入的默认目标设备。

[0142] 在一些实施方式中，可以(例如，通过语音辅助服务器系统112或语音激活设备104)对用户的语音输入历史(例如，由语音辅助服务器系统112收集并存储在存储器306中、由语音激活设备104收集并存储在存储器206中)进行分析以确定该历史是否示出特定语音激活设备104频繁地用于控制特定可控设备106。如果该历史确实示出这样的关系，则可以将该特定可控设备设置为给语音激活设备的语音输入的默认目标设备。

[0143] 在一些实施方式中，默认目标设备的指定(例如，标识符)被存储在语音激活设备104和/或语音辅助服务器系统112处。

[0144] 在一些实施方式中，根据第二语音命令满足聚焦会话维持准则的确定，相对于第一目标设备延长聚焦会话。在一些实施方式中，聚焦会话在一定量的时间之后超时(即，结束)。如果第二语音输入420满足聚焦会话维持准则，则聚焦会话418可以按时间延长(例如，重置超时定时器)。

[0145] 在一些实施方式中，相对于第一目标设备建立聚焦会话包括存储第一语音命令的时间戳，并且存储第一目标设备的标识符。当在接收到语音输入403之后建立聚焦会话时，语音激活设备404存储语音输入403的时间(例如，在会话命令历史260中)和对焦设备410的标识符(例如，在会话对焦设备256中)。

[0146] 在一些实施方式中，聚焦会话维持准则包括第二语音命令在相对于接收到第一语音命令的第一预定义时间间隔内或者在相对于接收到第一语音命令的第二预定义时间间隔内由第一电子设备接收到的准则，所述第二预定义时间间隔接续(succeeding)所述第一预定义时间间隔；并且确定第二语音命令是否满足一个或多个聚焦会话维持准则包括确定是否在第一预定义时间间隔或预定义第二时间间隔中的任一个内接收到第二语音命令。语音激活设备404确定语音输入420是否满足一个或多个聚焦会话维持准则，包括是否在语音输入403的第一时间阈值或第二时间阈值内接收到语音输入420。

[0147] 在一些实施方式中，根据在第一预定义时间间隔内接收到第二语音命令的确定，第一电子设备选择第一目标设备作为第二语音命令的目标设备。如果语音输入420被确定为从语音输入403起在第一较短时间阈值内被接收，则对焦设备410被设置为语音输入420的目标设备。

[0148] 在一些实施方式中，根据在第二预定义时间间隔内接收到第二语音命令的确定，第一电子设备输出用于确认第一目标设备作为第二语音命令的目标设备的请求；并且根据响应于对确认的请求对第一目标设备的肯定确认，选择第一目标设备作为第二语音命令的目标设备。如果语音输入420被确定为从语音输入403起在第一较短时间阈值之外但是在第二较长时间阈值内被接收，则语音激活设备提示用户确认目标设备(例如，问用户对焦设备410是否是预定目标设备)。如果用户确认对焦设备410是预定目标设备，则对焦设备410被设置为语音输入420的目标设备。

[0149] 在一些实施方式中，第一电子设备接收包括对第三操作的请求和已连接电子设备的本地组当中的第三目标设备的显式指定的第三语音命令，相对于第一目标设备结束聚焦会话，相对于第三目标设备建立聚焦会话，并且经由公共网络服务的操作使第三操作由第三目标设备执行。语音激活设备404可以在语音输入420之后接收包括除设备410以外的目

标设备(例如,设备406或408)的显式指定的新语音输入。根据该语音输入的接收,与在焦点上的设备410的聚焦会话418结束,并且与在焦点上的新目标设备的新会话被建立。语音激活设备404或语音辅助服务器系统112经由语音辅助服务140向新目标设备传送用于执行新语音输入中所请求的操作的命令。

[0150] 在一些实施方式中,第一目标设备是第一电子设备。第一电子设备接收包括对第四操作的请求和已连接电子设备的本地组当中的第四目标设备的显式指定的第四语音命令,其中第四目标设备是已连接电子设备的本地组的第三电子设备成员,第三电子设备与第一电子设备不同;相对于第一目标设备维持聚焦会话;并且经由公共网络服务的操作使第四操作由第四目标设备执行。如果在语音激活设备404处的活动聚焦会话418的对焦设备是语音激活设备404它本身,然后在语音输入420之后接收到将不同的设备指定为目标的新语音输入,则语音激活设备404或语音辅助服务器系统112经由语音辅助服务140向该不同的目标设备传送用于执行新语音输入中所请求的操作的命令,但是与在焦点上的语音激活设备404维持聚焦会话。

[0151] 在一些实施方式中,在使第四操作由第四目标设备执行之后接收第二语音命令,第一操作是媒体播放操作,并且第二操作是媒体中止操作。第一电子设备接收包括对第五操作的请求和已连接电子设备的本地组当中的第五目标设备的显式指定的第五语音命令,其中第五目标设备是第三电子设备;相对于第一目标设备结束聚焦会话;相对于第五目标设备建立聚焦会话,并且经由公共网络服务的操作使第五操作由第五目标设备执行。如果在语音激活设备404处的活动聚焦会话418的对焦设备是语音激活设备404它本身,并且语音输入403包括了发起媒体播放的请求,并且语音输入403包括了作为语音输入403的结果暂停媒体播放的请求,并且在语音输入420之后接收到将不同的设备指定为目标的新语音输入,则语音激活设备404或语音辅助服务器系统112经由语音辅助服务140向该不同的目标设备传送用于执行新语音输入中所请求的操作的命令,并且与在焦点上的语音激活设备的聚焦会话结束,并且与在焦点上的新目标设备的新聚焦会话被建立。

[0152] 在一些实施方式中,第一电子设备接收包括预定义操作终止请求的第五语音命令,并且根据接收到第五语音命令,使第一操作停止由第一目标设备执行,并且相对于第一目标设备结束聚焦会话。如果语音激活设备404接收到预定义终止命令(例如,“STOP”),则语音激活设备404或语音辅助服务器系统112经由语音辅助服务140向设备410传送用于停止执行操作416的命令,并且聚焦会话418结束。

[0153] 在一些实施方式中,第一操作是媒体播放操作,并且第二操作是以下各项中的一个:媒体中止操作、媒体倒回操作、媒体快进操作、调高音量操作和调低音量操作。语音输入403中的请求可以是发起媒体内容(例如,视频、音乐)的播放的请求,并且语音输入420中的请求可以是控制播放(例如,暂停、倒回,快进、调高/调低音量、下一项/乐曲、上一项/乐曲等)的请求。

[0154] 在一些实施方式中,第一操作是到多个设备状态中的第一状态的设备状态改变操作,并且第二操作是到多个设备状态中的第二状态的设备状态改变操作。语音输入403中的请求可以是转变到第一状态(例如,打开灯或设备、转向睡眠模式)的请求,并且语音输入420中的请求可以是转变到第二状态(例如,关闭灯或设备、从睡眠模式唤醒)的请求。

[0155] 在一些实施方式中,第一操作是在第一方向上针对幅度标尺的幅度改变操作,并

且第二操作是在与第一方向相反的第二方向上针对幅度标尺的幅度改变操作。语音输入403中的请求可以是在一个方向上改变幅度(例如,使灯发亮、调高音量)的请求,并且语音输入420中的请求可以是在相反方向上改变幅度(例如,使灯变暗、调低音量)的请求。

[0156] 在一些实施方式中,第一电子设备包括一个或多个LED的阵列。第一电子设备通过点亮LED阵列中的LED中的一个或多个来指示聚焦会话的状态。语音激活设备404可以通过在LED阵列上显示图案来指示存在活动聚焦会话或与该聚焦会话相关联的其它状态和其它信息(例如,聚焦会话活动了多久或者自最后语音输入以来已经过多少时间的指示)。

[0157] 在一些实施方式中,可以按标识的用户建立聚焦会话。例如,如果用户说出指定目标设备的语音输入,则该用户被标识并且相对于所标识的用户建立聚焦会话,其中在语音输入中指定的目标设备在焦点上。如果不同的用户说出语音输入并指定不同的目标设备,则该不同的用户被标识并且相对于所标识的不同的用户建立另一聚焦会话,其中该不同的目标设备在焦点上。可以基于与相应的标识的用户相对应的活动聚焦会话给由不同的用户说出并且未指定目标设备的语音输入指派不同的目标设备。

[0158] 聚焦会话的附加示例

[0159] 以下实施方式在与作为媒体设备的一个或多个可控设备106相同的房间中的语音激活设备104的场境中对实施方式进行描述。应该了解,在下面所描述的实施方式可以适于其它类型的可控设备106(例如,智能家居设备)并且适于其它设备布局设置。

[0160] 在一些实施方式中,如果不存在已经在语音激活设备上播放的媒体,则可在对焦设备为除该语音激活设备以外的可控设备情况下开始聚焦会话。在一些实施方式中,如果在语音激活设备上播放的媒体被暂停,则可以与除作为对焦设备的语音激活设备以外的可控设备开始聚焦会话。

[0161] 在一些实施方式中,如果用户发出具有指向与语音激活设备相关联的设备或设备组(并且可选地在与语音激活设备相同的WiFi网络上)的显式目标设备的任何有效请求则开始聚焦会话。此类有效请求的示例包括“在我的客厅扬声器上播放一些音乐”、“调高卧室电视上的音量”、“我的家庭组上的下一首歌曲”和“暂停客厅扬声器”。显式目标设备变成聚焦会话的对焦设备。

[0162] 在一些实施方式中,如果请求清楚地是与视频相关联的请求,并且在相关联的可控设备之中存在单个支持视频的设备,则可以与作为对焦设备的支持视频的设备建立聚焦会话。

[0163] 在一些实施方式中,如果在语音激活设备正在积极地播放媒体的同时接收到作为目标设备的另一设备的请求,则焦点将仍然在语音激活设备上,但是一旦语音激活设备停止或者暂停了其会话,在另一设备上播放或者控制媒体的任何新请求将焦点移动到该另一设备。

[0164] 例如,用户请求“播放嘎嘎小姐(Lady Gaga)”,并且语音激活设备开始播放嘎嘎小姐音乐并且与在焦点上的语音激活设备开始聚焦会话。用户然后请求“暂停”,并且语音激活设备暂停嘎嘎小姐音乐(并且维持聚焦会话达假定2小时)。在已经过1小时之后,用户然后请求“在我的电视上播放猫视频”。焦点移动到电视,并且电视开始播放猫视频。

[0165] 作为另一示例,用户请求“播放嘎嘎小姐”,并且语音激活设备开始播放嘎嘎小姐音乐并且开始与在焦点上的语音激活设备的聚焦会话。用户然后请求“在我的电视上示出

猫视频”，并且猫视频开始在电视上示出，但是焦点仍然保持在语音激活设备上。用户然后请求“下一首”，语音激活设备根据该请求前进到嘎嘎小姐音乐中的下一首乐曲。用户然后请求“暂停”，并且语音激活设备处的音乐被暂停。用户然后请求“我的电视上的下一个幻灯片”，并且下一个幻灯片在电视上开始并且焦点转移到电视。

[0166] 在一些实施方式中，有效请求包括发起音乐、发起视频、发起新闻阅读(例如，读出新闻文章)、发起播客、发起照片(例如，照片显示或幻灯片放映)以及任何媒体控制命令(除结束任何当前聚焦会话的预定义STOP命令以外。)

[0167] 在一些实施方式中，当发生下列中的任一个时聚焦会话结束：

[0168] ●聚焦会话被转移到不同的设备(经由语音输入，例如，显式地指定该不同的设备的语音输入)，并且在这种情况下与该不同的设备开始聚焦会话；

[0169] ●聚焦会话经由语音输入或从另一设备投射(例如，经由语音：“在<语音接口设备的昵称>上播放嘎嘎小姐”、“在本地播放嘎嘎小姐”等；经由投射：用户经由客户端设备上的应用将内容投射到语音激活设备)在语音激活设备上开始或者恢复(离开暂停状态)；

[0170] ○然而，如果语音激活设备是将播放媒体的组的成员(跟随者或领导者)，则它将不停止焦点(即使它正在播放)。所以焦点将仍然在所述组的领导者(其可以是另一语音激活设备)上；

[0171] ●当请求是给在焦点上的可控设备的预定义“STOP”命令(包括所有相关语法)时；

[0172] ●超时相关命令：

[0173] ○可以根据给予给可控设备(无论该可控设备是否是基于聚焦会话的对焦设备来显式地指定或者设置的)的最后请求或命令而不是预定义“停止”命令来测量超时；

[0174] ○超时可以是跨越各种可能的命令240分钟；以及

[0175] ●当用户按语音激活设备上用于暂停/播放的按钮时(此外这也将在本地在语音激活设备上恢复任何暂停的内容)。

[0176] 在一些实施方式中，语音激活设备请求目标设备的用户确认。用户被提示以便确认他是否想要在可控设备上播放媒体如下：

[0177] ●提示是为媒体发起而触发的(例如，开始尚未在播放的音乐)(与媒体控制相对，诸如快进或下一首乐曲)；

[0178] ●当聚焦会话活动时提示被触发；以及

[0179] ●提示在从来自当前语音激活设备的给予给可控设备(无论该可控设备是否是基于聚焦会话的对焦设备来显式地指定或者设置的)的最后语音命令而不是预定义“STOP”命令起已经过一些时间(例如，20分钟)之后被触发。

[0180] 用于确认的提示可以是，例如：

[0181] ●语音激活设备输出“你愿意让我在<可控设备名称>上播放吗?”

[0182] ○用户响应“是”。则在对焦可控设备上播放所请求的媒体并且在该设备上维持焦点。

[0183] ○

[0184] ○用户响应“否”。则在语音激活设备上播放所请求的媒体并且聚焦会话结束。

[0185] ○

[0186] ○其它：如果例如用户的响应是不清楚的，则语音激活设备可以会输出“抱歉，无

法理解你的响应”。

[0187] 在一些实施方式中，当聚焦会话被发起时，媒体发起和基于语音的控制命令被应用于对焦可控设备。非媒体请求(例如，搜索、问题)由语音激活设备回答，并且非媒体请求确实不结束聚焦会话。

[0188] 在一些实施方式中，即便当聚焦会话已开始时，物理交互也将仍然控制语音激活设备，所以与语音激活设备的用于改变音量和暂停/播放的物理交互(例如，按按钮、触摸触敏区域)影响语音激活设备，而不一定是可控设备。

[0189] 在一些实施方式中，向语音激活设备上的定时器/闹钟/文本到语音播放发出的请求或命令与给对焦可控设备的类似请求或命令相比具有更高的优先级。例如，如果语音激活设备由于定时器或闹钟正在响铃，并且用户发出“停止”，则语音激活设备停止定时器或闹钟响铃。如果用户然后发出“<调高/调低>音量”，则定时器或闹钟响铃仍然被停止，并且可控设备上的音量被调高或者调低。

[0190] 作为另一示例，如果语音激活设备正在播放文本到语音(例如，读出用户的电子邮件)，并且用户发出“停止”，则语音激活设备停止文本到语音阅读。如果用户然后发出“<调高/调低>音量”，则语音激活设备上的音量被调高或者调低。

[0191] 作为另一示例，如果语音激活设备是空闲的、被暂停或app加载，并且用户发出“停止”，则可控设备处的媒体播放被停止并且聚焦会话结束。如果用户然后发出“<调高/调低>音量”，则可控设备上的音量被调高或者调低。

[0192] 语音激活电子设备的物理特征

[0193] 图6A和图6B是根据一些实施方式的语音激活电子设备104(图1)的前视图600和后视图620。电子设备104包括一个或多个麦克风602和全色LED 604的阵列。全色LED 604能被隐藏在电子设备104的顶面下方并且在它们未点亮时对用户不可见。在一些实施方式中，全色LED 604的阵列在物理上按照环环形布置。另外，电子设备104的后侧可选地包括被配置成耦合至电源的电源连接器608。

[0194] 在一些实施方式中，电子设备104呈现没有可见按钮的干净样子，并且与电子设备104的交互基于语音和触摸手势。可替选地，在一些实施方式中，电子设备104包括有限数目的物理按钮(例如，在其后侧的按钮606)，并且与电子设备104的交互除了基于语音和触摸手势之外还基于对按钮的按压。

[0195] 一个或多个扬声器被布置在电子设备104中。图6C是根据一些实施方式的示出按照开放配置包含在电子设备104的基座610中的扬声器622的语音激活电子设备104的立体图660。电子设备104包括全色LED 604的阵列、一个或多个麦克风602、扬声器622、双频带WiFi802.11ac无线电、蓝牙LE无线电、环境光传感器、USB端口、处理器以及存储由该处理器执行的至少一个程序的存储器。

[0196] 参考图6D，电子设备104还包括被配置成检测电子设备104的顶面的触摸事件的触摸感测阵列624。触摸感测阵列624可以被布置并隐藏在电子设备104的顶面下方。在一些实施方式中，触摸感测阵列被布置在包括通孔的阵列的电路板的顶面上，并且全色LED 604被布置在电路板的通孔内。当电路板被设置在电子设备104的顶面正下方时，全色LED 604和触摸感测阵列624两者也被布置在电子设备104的顶面正下方。

[0197] 图6E(1)至图6E(4)示出根据一些实施方式的在语音激活电子设备104的触摸感测

阵列624上检测到的四个触摸事件。参考图6E(1)和图6E(2)，触摸感测阵列624检测语音激活电子设备104的顶面上的旋转扫掠。响应于检测到顺时针扫掠，语音激活电子设备104增加其音频输出的音量，并且响应于检测到逆时针扫掠，语音激活电子设备104降低其音频输出的音量。参考图6E(3)，触摸感测阵列624检测语音激活电子设备104的顶面上的单轻敲触摸。响应于检测到第一轻敲触摸，语音激活电子设备104执行第一媒体控制操作(例如，播放特定媒体内容)，而响应于检测到第二轻敲触摸，语音激活电子设备104实现第二媒体控制操作(例如，暂停当前正在播放的特定媒体内容)。参考图6E(4)，触摸感测阵列624检测语音激活电子设备104的顶面上的双轻敲触摸(例如，两个连续触摸)。两个连续触摸被分开时间小于预定长度的持续时间。然而，当它们被分开时间大于预定长度的持续时间时，两个连续触摸被认为是两个单轻敲触摸。在一些实施方式中，响应于检测到双轻敲触摸，语音激活电子设备104发起电子设备104侦听并识别一个或多个热词(例如，预定义关键词)的热词检测状态。在电子设备104识别热词之前，电子设备104不向语音辅助服务器112或云投射服务服务器118发送任何音频输入。在一些实施方式中，聚焦会话是响应于检测到一个或多个热词而发起的。

[0198] 在一些实施方式中，全色LED 604的阵列被配置成根据LED设计语言显示视觉图案集合，指示对语音激活电子设备104的顶面上的顺时针扫掠、逆时针扫掠、单轻敲或双轻敲的检测。例如，全色LED 604的阵列可以顺序地点亮以分别像图6E(1)和图6E(2)中所示出的那样跟踪顺时针扫掠或逆时针扫掠。在下面参考图6F和图6G(1)至图6G(8)来说明关于与电子设备104的语音处理状态相关联的视觉图案的更多细节。

[0199] 图6E(5)示出根据一些实施方式的用户对在语音激活电子设备104的后侧的按钮606的示例触摸或按压。响应于对按钮606的第一用户触摸或按压，电子设备104的麦克风被静音，而响应于对按钮606的第二用户触摸或按压，电子设备104的麦克风被激活。

[0200] 用于语音用户接口的视觉可视性的LED设计语言

[0201] 在一些实施方式中，电子设备104包括全色发光二极管(LED)的阵列而不是全显示屏幕。LED设计语言被采纳来配置全色LED的阵列的照明并且实现指示电子设备104的不同语音处理状态的不同视觉图案。LED设计语言由被应用于一组固定的全色LED的颜色、图案和特定运动的语法构成。语言中的元素被组合以在电子设备104的使用期间在视觉上指示特定设备状态。在一些实施方式中，全色LED的照明目的旨在除了其他重要的状态之外还清楚地刻画电子设备104的被动侦听和主动收听状态。可使用类似的LED设计语言元素来通过LED(例如，LED 604)在视觉上指示的状态包括一个或多个焦点会话的状态、与一个或多个特定焦点会话相关联的一个或多个用户的身份和/或一个或多个活动的焦点会话的持续时间。例如，在一些实施方式中，可以使用LED 604的不同的灯图案、颜色组合和/或特定运动来指示聚焦会话是活动的，由于检测到第二语音输入已被扩展，并且/或者由于缺乏与电子设备104的用户语音交互最近已失效。与特定焦点会话相关联的一个或多个用户的一个或多个身份也可用在视觉上标识特定用户的LED 604的不同的灯图案、颜色组合和/或特定运动来指示。全色LED的放置遵照电子设备104的物理约束，并且能基于特定技术(例如Google Assistant)在由第三方原始设备制造商(OEM)所制造的扬声器中使用全色LED的阵列。

[0202] 在语音激活电子设备104中，当电子设备104对从其周围环境收集的音频输入进行处理但不存储这些音频输入或者将这些音频输入传送到任何远程服务器时发生被动侦听。

相比之下，当电子设备104存储从其周围环境收集的音频输入并且/或者与远程服务器共享这些音频输入时发生主动侦听。根据本申请的一些实施方式，电子设备104仅在不违反电子设备104的用户的隐私的情况下被动地侦听其周围环境中的音频输入。

[0203] 图6F是根据一些实施方式的语音激活电子设备104的顶视图，并且图6G示出根据一些实施方式的通过用于指示语音处理状态的全色LED的阵列所显示的六个示例视觉图案。在一些实施方式中，电子设备104不包括任何显示屏幕，并且与全显示屏幕相比全色LED 604提供简单且低成本的视觉用户接口。全色LED可以被隐藏在电子设备的顶面下方并且在它们未点亮时对用户不可见。参考图6F和图6G，在一些实施方式中，全色LED 604的阵列在物理上按照环形布置。例如，如图6G(6)中所示，全色LED 604的阵列可以顺序地点亮以分别像图6E(1)和图6E(2)中所示出的那样跟踪顺时针扫掠或逆时针扫掠。

[0204] 用于在视觉上指示语音处理状态的方法被实现在电子设备104处。电子设备104经由一个或多个麦克风602收集来自接近该电子设备的环境的音频输入，并且对这些音频输入进行处理。处理包括标识并对来自环境中的用户的语音输入作出响应中的一个或多个。电子设备104从多个预定义语音处理状态当中确定处理的状态。对于全色LED604中的每一个，电子设备104标识与所确定的语音处理状态相关联的相应的预定LED照明规格。照明规格包括LED照明持续时间、脉冲速率、占空比、颜色序列和亮度中的一个或多个。在一些实施方式中，电子设备104确定语音处理状态(在一些实施方式中包括聚焦会话的状态)与多个用户中的一个相关联，并且通过根据所述多个用户中的一个的身份来定制全色LED 604的预定LED照明规格中的至少一种(例如，颜色序列)而标识全色LED 604的预定LED照明规格。

[0205] 另外，在一些实施方式中，根据所确定的语音处理状态，全色LED的颜色包括预定颜色集合。例如，参考图6G(2)、6G(4)和6G(7)-(10)，该预定颜色集合包括包括有蓝色、绿色、黄色和红色的Google品牌颜色，并且全色LED的阵列被划分成各自与Google品牌颜色中的一种相关联的四个象限。

[0206] 根据所标识的全色LED的LED照明规格，电子设备104使全色LED的阵列的照明同步以提供指示所确定的语音处理状态(在一些实施方式中包括聚焦会话的状态)的视觉图案。在一些实施方式中，指示语音处理状态的视觉图案包括多个离散的LED照明像素。在一些实施方式中，视觉图案包括起始段、环段和终止段。环段持续与全色LED的LED照明持续时间相关联的时间长度并且被配置成与语音处理状态的长度(例如，活动聚焦会话的持续时间)匹配。

[0207] 在一些实施方式中，电子设备104具有通过LED设计语言所表示的多于二十个不同的设备状态(包括所述多个预定义语音处理状态)。可选地，所述多个预定义语音处理状态包括热词检测状态、侦听状态、思考状态和响应状态中的一个或多个。在一些实施方式中，如上所述，所述多个预定义语音处理状态包括一个或多个聚焦会话状态。

[0208] 已经详细地参考了实施方式，其示例被图示在附图中。在上面的详细描述中，已经阐述了许多特定细节以便提供对各种描述的实施方式的透彻理解。然而，对于本领域的普通技术人员而言将显而易见的是，可以在没有这些特定细节的情况下实践各种描述的实施方式。在其它实例中，尚未详细地描述众所周知的方法、过程、组件、电路和网络以免不必要地使实施方式的各方面混淆。

[0209] 也应理解，尽管在一些实例中，在本文中使用术语第一、第二等来描述各种元件，

然而这些元件不该应受这些术语限制。这些术语仅用于区分一个元件和另一元件。例如，在不脱离各种所描述的实施方式的范围的情况下，第一设备能被称作第二设备，并且类似地，第二设备能被称作第一设备。第一设备和第二设备是两种类型的设备，但是它们不是同一设备。

[0210] 在本文的各种描述的实施方式的描述中使用的术语仅用于描述特定实施方式的目的，而不旨在为限制性的。如各种描述的实施方式和所附权利要求的描述中所使用的，除非上下文另外清楚地指示，否则单数形式“一(a)”、“一个(an)”和“所述(the)”也旨在包括复数形式。也应理解，如本文中所使用的术语“和/或”指代并包含相关联的列举项中的一个或多个的任何和所有可能的组合。还应理解，术语“包含”、“包含有”、“包括”和/或“包括有”当被用在本说明书中时，指定陈述的特征、整数、步骤、操作、元件和/或组件的存在，但是不排除一个或多个其它特征、整数、步骤、操作、元件、组件和/或其组的存在或添加。

[0211] 如本文中所使用的，取决于上下文，术语“如果”被可选地解释成意指“当…时”或“在…时”或“响应于确定”或“响应于检测到”或“根据…的确定”。类似地，取决于上下文，短语“如果确定了”或“如果检测到[陈述的条件或事件]”被可选地解释成意指“在确定…时”或“响应于确定”或“在检测到[所陈述的条件或事件]时”或“响应于检测到[所陈述的条件或事件]”或“根据检测到[陈述的条件或事件]的确定”。

[0212] 对于上面所讨论的系统收集关于用户的信息的情形，可以给用户提供用于选择参加/退出可以收集个人信息(例如，关于用户的偏好或智能设备的使用的信息)的程序或特征的机会。此外，在一些实施方式中，某些数据可以在它被存储或者使用之前被以一个或多个方式匿名，使得个人可识别的信息被移除。例如，可以使用户的身份匿名，使得对于该用户来说不可确定个人可识别的信息或者个人可识别的信息不可与用户相关联，并且使得用户偏好或用户交互被一般化(例如，基于用户人口统计资料泛化)，而不是与特定用户相关联。

[0213] 尽管各个附图中的一些以特定次序图示许多逻辑阶段，然而可以对不是次序相关的阶段重新排序并且可以组合或者取出其它阶段。虽然特别提及了一些重新排序或其它分组，但是其余的对于本领域的普通技术人员而言将是显而易见的，所以本文中所呈现的排序和分组不是替代方案的详尽列表。此外，应该认识到，这些阶段能用硬件、固件、软件或其任何组合加以实现。

[0214] 出于说明的目的，已经参考特定实施方式描述了上述描述。然而，上面的说明性讨论不旨在为详尽的或者将权利要求的范围限于所公开的精确形式。鉴于以上教导许多修改和变化是可能的。实施方式被选取以便最好地说明作为权利要求基础的原理及其实际应用，以因此使得本领域的技术人员能够按如适于所设想的特定用途的各种修改而最佳地使用这些实施方式。

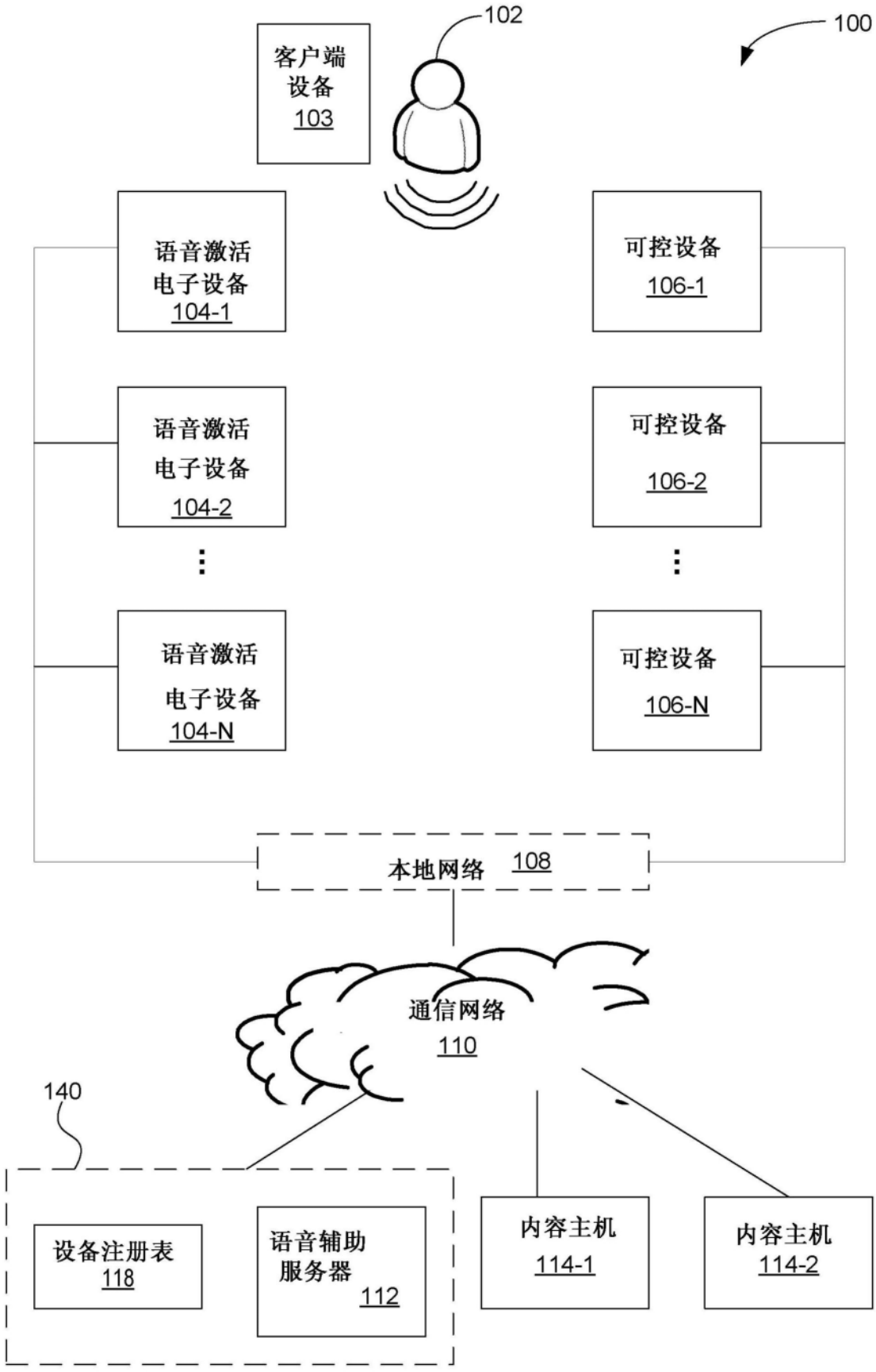

102
100
客户端
设备
103
语音激活
电子设备
104-1
语音激活
电子设备
104-2
语音激活
电子设备
104-N
可控设备
106-1
可控设备
106-2
可控设备
106-N
本地网络 108
通信网络
110
140
设备注册表
118
语音辅助
服务器
112
内容主机
114-1
内容主机
114-2

图1

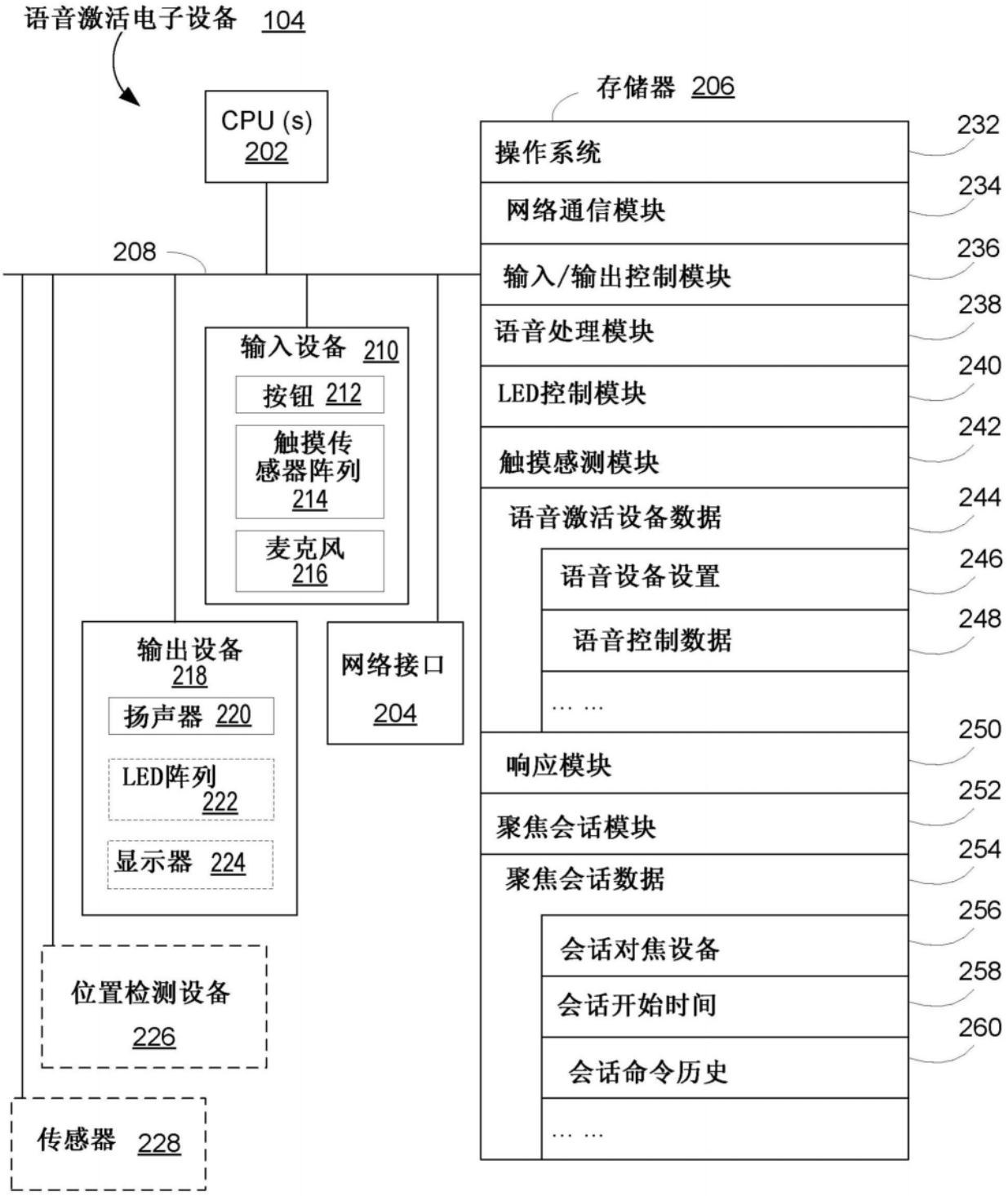

语音激活电子设备 104
CPU (s) 202
208
存储器 206
操作系统
232
网络通信模块
234
输入/输出控制模块
236
语音处理模块
238
LED控制模块
240
触摸感测模块
242
语音激活设备数据
244
语音设备设置
246
语音控制数据
248
… …
响应模块
250
聚焦会话模块
252
聚焦会话数据
254
会话对焦设备
256
会话开始时间
258
会话命令历史
260
… …
输入设备 210
按钮 212
触摸传感器阵列 214
麦克风 216
输出设备 218
扬声器 220
LED阵列 222
显示器 224
网络接口 204
位置检测设备 226
传感器 228

图2

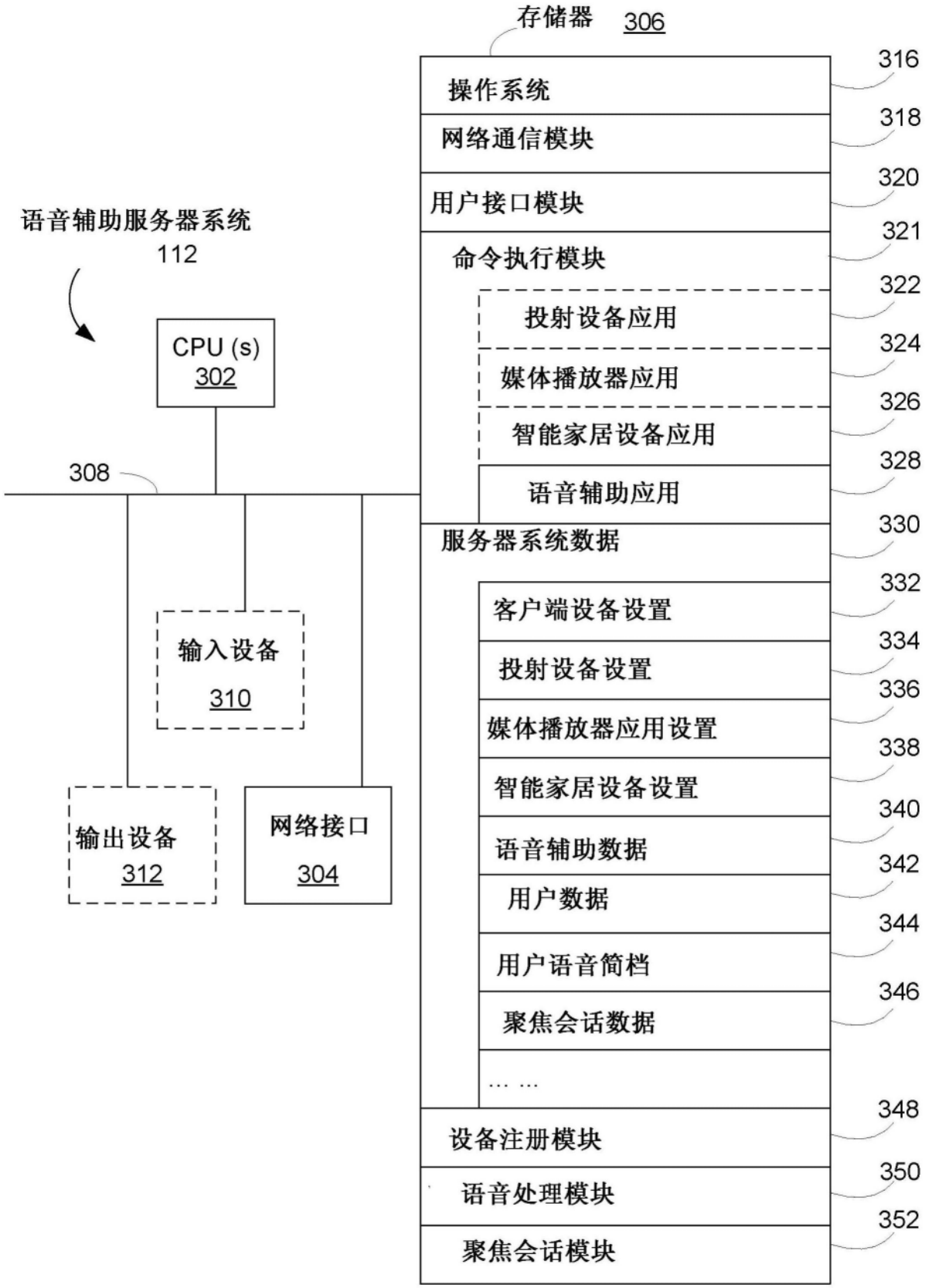

存储器 306
操作系统
316
网络通信模块
318
用户接口模块
320
命令执行模块
321
投射设备应用
322
媒体播放器应用
324
智能家居设备应用
326
语音辅助应用
328
服务器系统数据
330
客户端设备设置
332
投射设备设置
334
媒体播放器应用设置
336
智能家居设备设置
338
语音辅助数据
340
用户数据
342
用户语音简档
344
聚焦会话数据
346
… …
设备注册模块
348
语音处理模块
350
聚焦会话模块
352
语音辅助服务器系统
112
CPU (s)
302
308
输入设备
310
输出设备
312
网络接口
304

图3A


聚焦会话数据
346
会话1
3462-1
会话源设备
3464-1
会话对焦设备
3466-1
会话开始时间
3468-1
会话命令历史
3470-1
… …
会话2
3462-2
会话源设备
3464-2
会话对焦设备
3466-2
会话开始时间
3468-2
会话命令历史
3470-2
… …
… …
会话M
3462-M

图3B

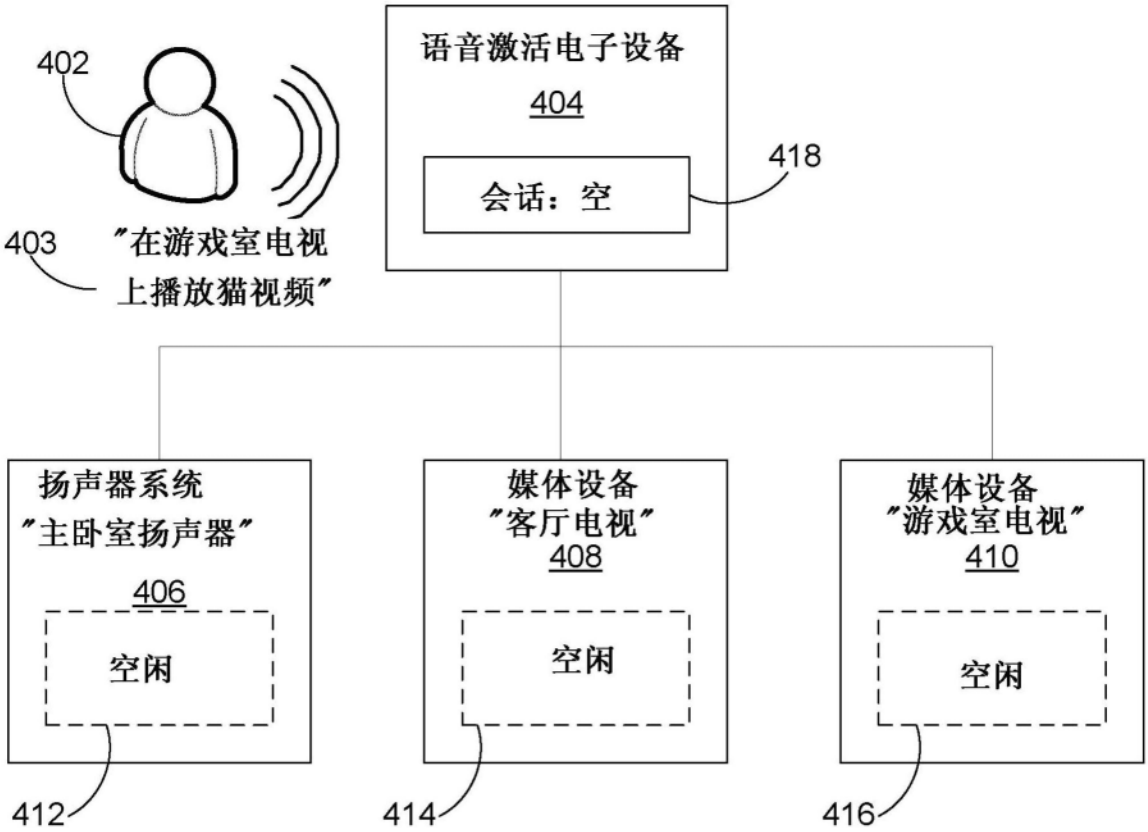

402
403
"在游戏室电视上播放猫视频"
语音激活电子设备
404
会话：空
418
扬声器系统
"主卧室扬声器"
406
空闲
412
媒体设备
"客厅电视"
408
空闲
414
媒体设备
"游戏室电视"
410
空闲
416

图4A

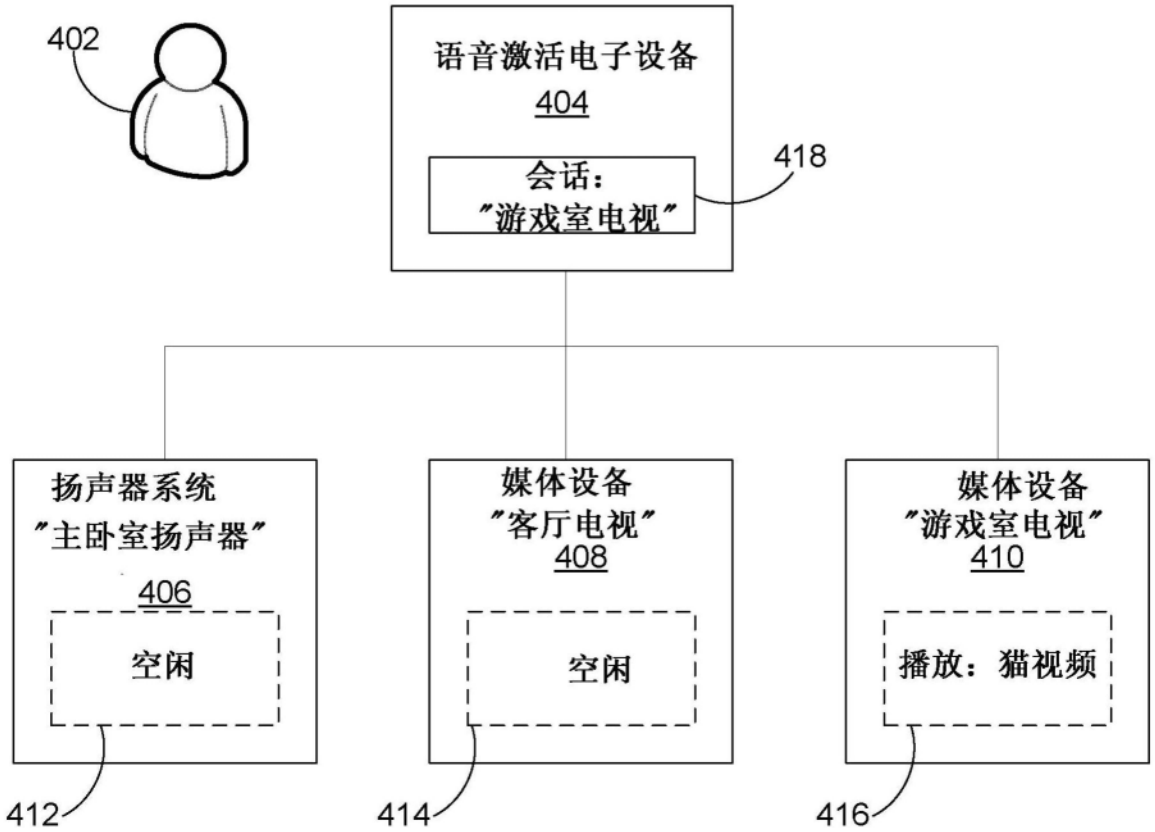

402
语音激活电子设备
404
会话：
"游戏室电视"
418
扬声器系统
"主卧室扬声器"
406
空闲
412
媒体设备
"客厅电视"
408
空闲
414
媒体设备
"游戏室电视"
410
播放：猫视频
416

图4B

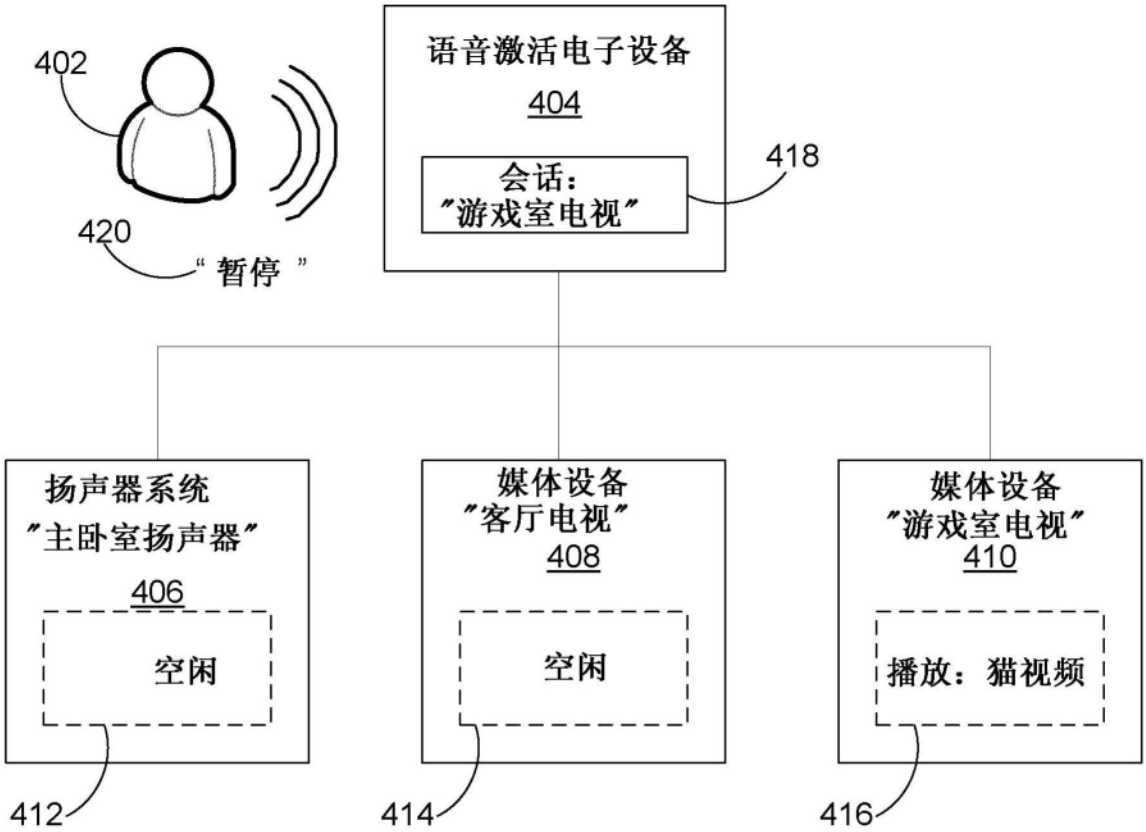

402
420
“ 暂停 ”
语音激活电子设备
404
会话：
″游戏室电视″
418
扬声器系统
″主卧室扬声器″
406
空闲
412
媒体设备
″客厅电视″
408
空闲
414
媒体设备
″游戏室电视″
410
播放：猫视频
416

图4C

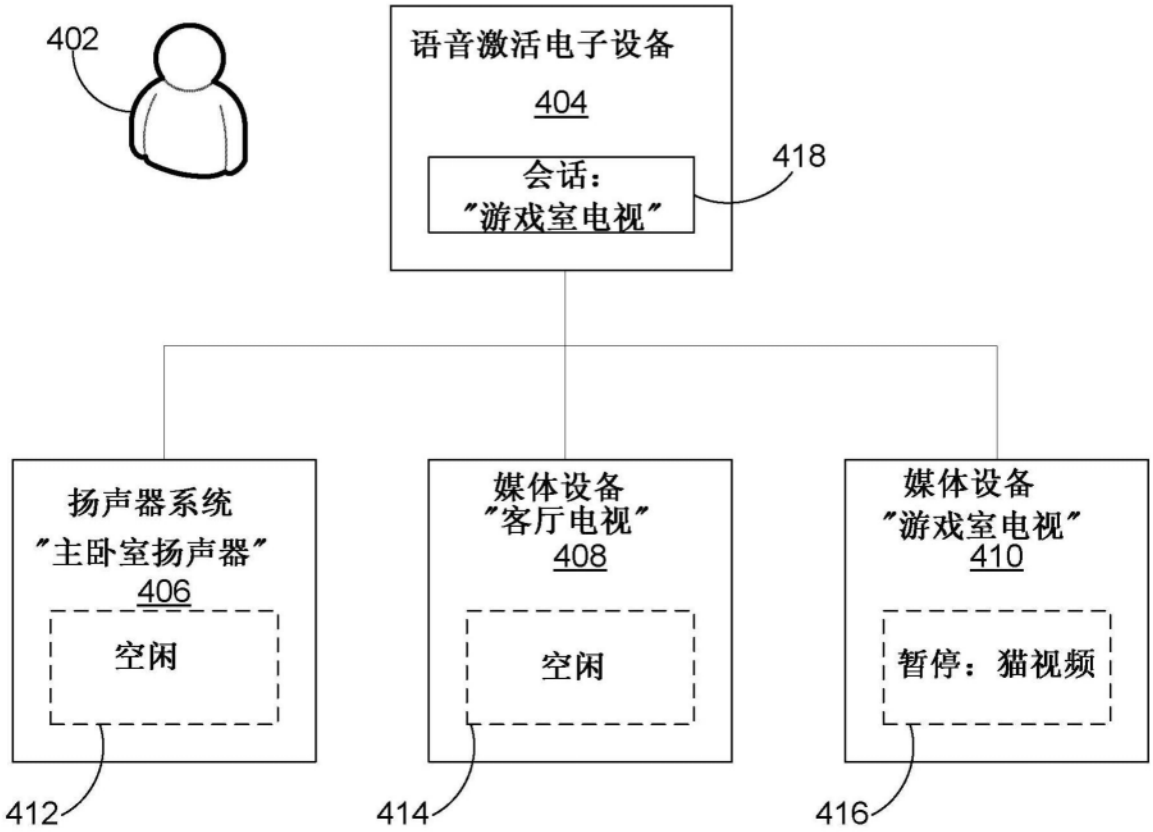

402
语音激活电子设备
404
会话：
"游戏室电视"
418
扬声器系统
"主卧室扬声器"
406
空闲
412
媒体设备
"客厅电视"
408
空闲
414
媒体设备
"游戏室电视"
410
暂停：猫视频
416

图4D

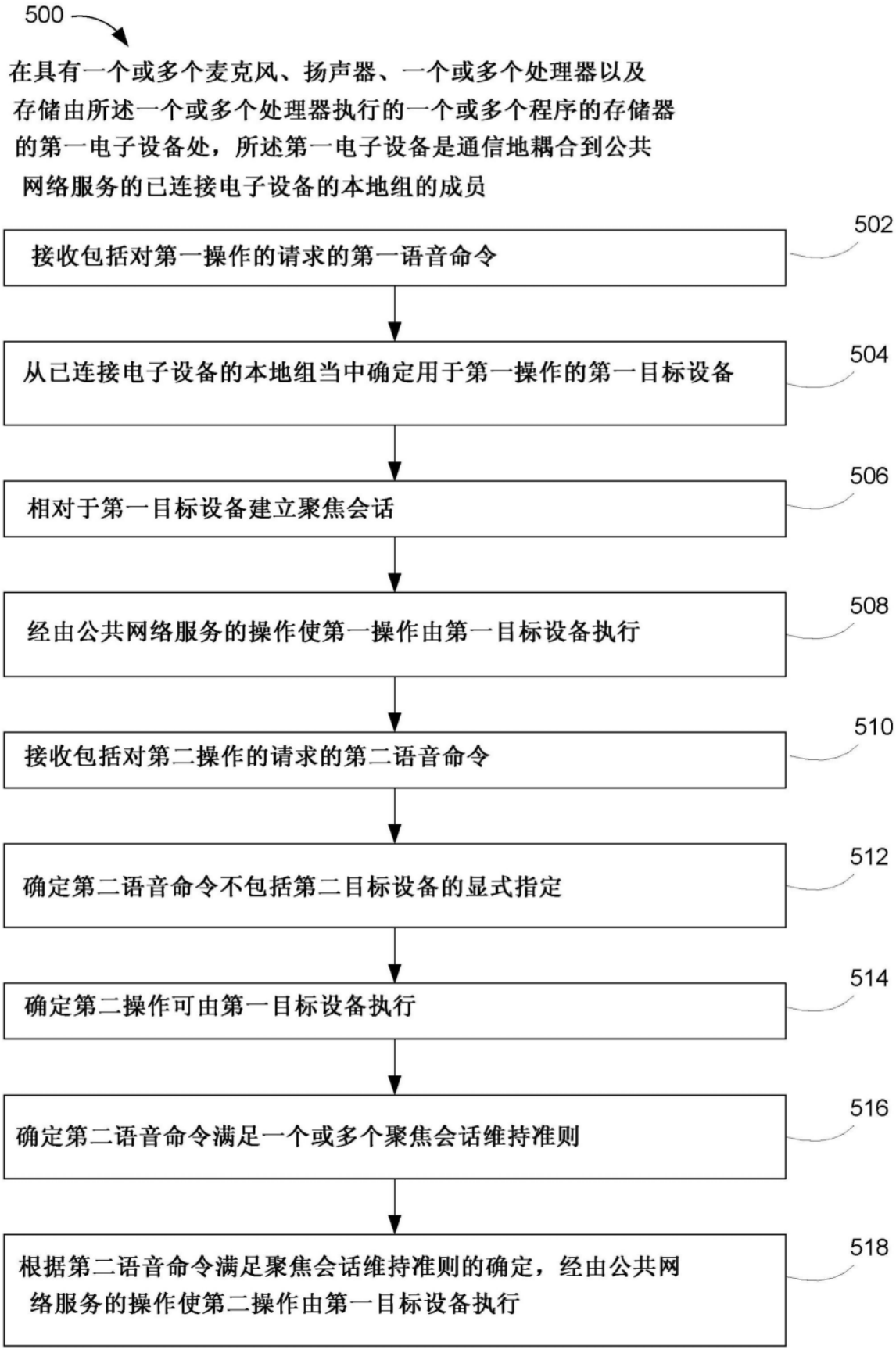

500
在具有一个或多个麦克风、扬声器、一个或多个处理器以及存储由所述一个或多个处理器执行的一个或多个程序的存储器的第一电子设备处，所述第一电子设备是通信地耦合到公共网络服务的已连接电子设备的本地组的成员
502
接收包括对第一操作的请求的第一语音命令
504
从已连接电子设备的本地组当中确定用于第一操作的第一目标设备
506
相对于第一目标设备建立聚焦会话
508
经由公共网络服务的操作使第一操作由第一目标设备执行
510
接收包括对第二操作的请求的第二语音命令
512
确定第二语音命令不包括第二目标设备的显式指定
514
确定第二操作可由第一目标设备执行
516
确定第二语音命令满足一个或多个聚焦会话维持准则
518
根据第二语音命令满足聚焦会话维持准则的确定，经由公共网络服务的操作使第二操作由第一目标设备执行

图5

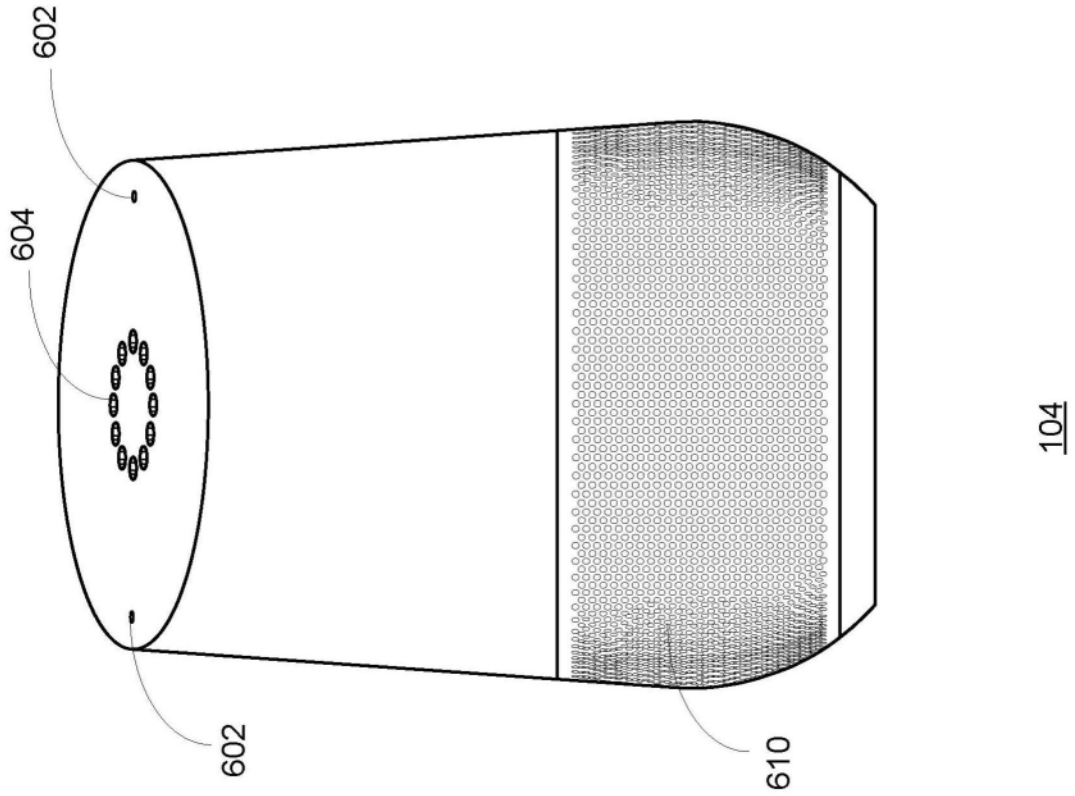


图6A

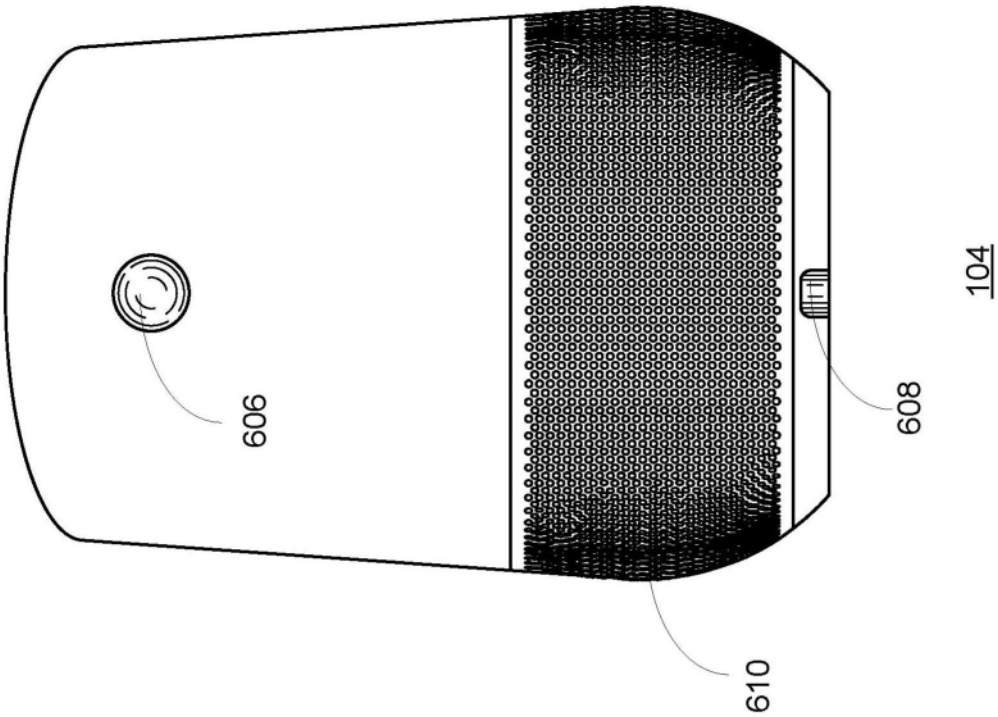


图6B

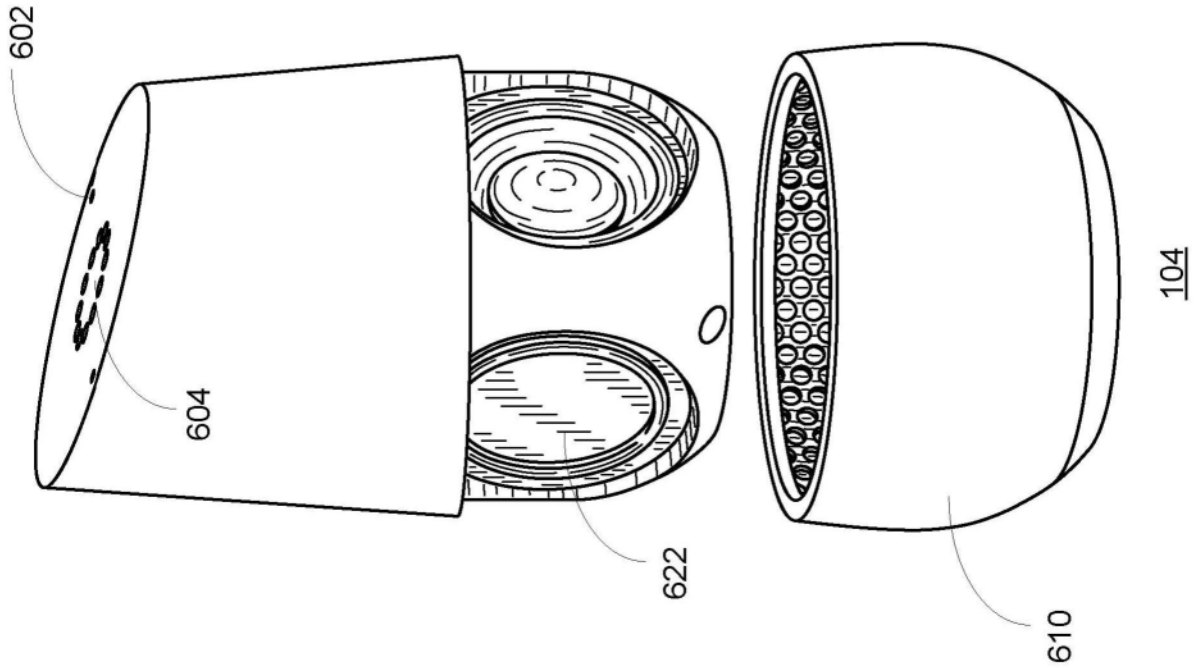


图6C

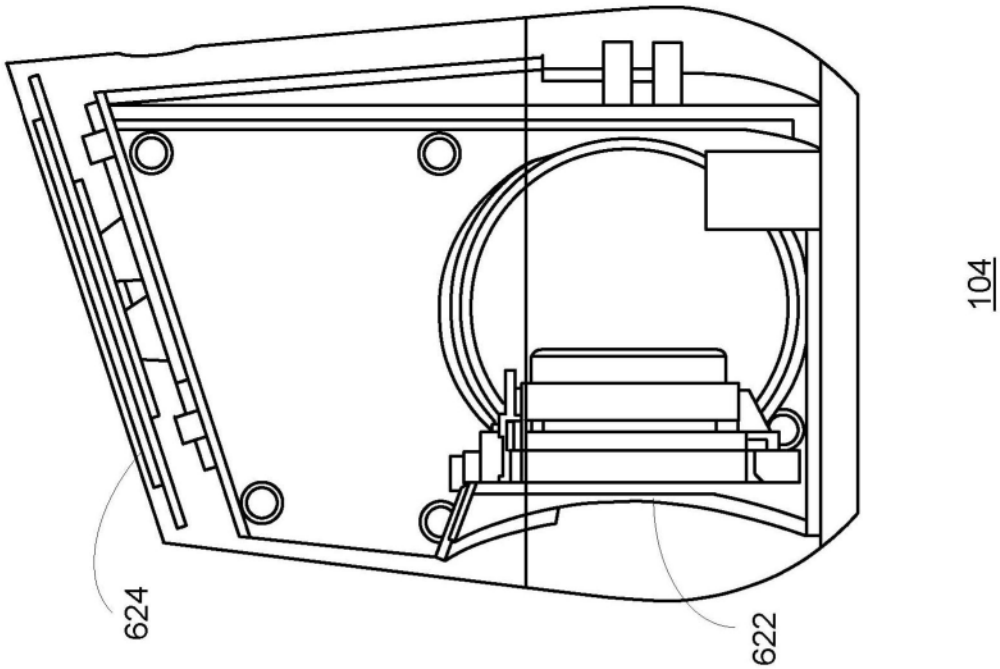


图6D

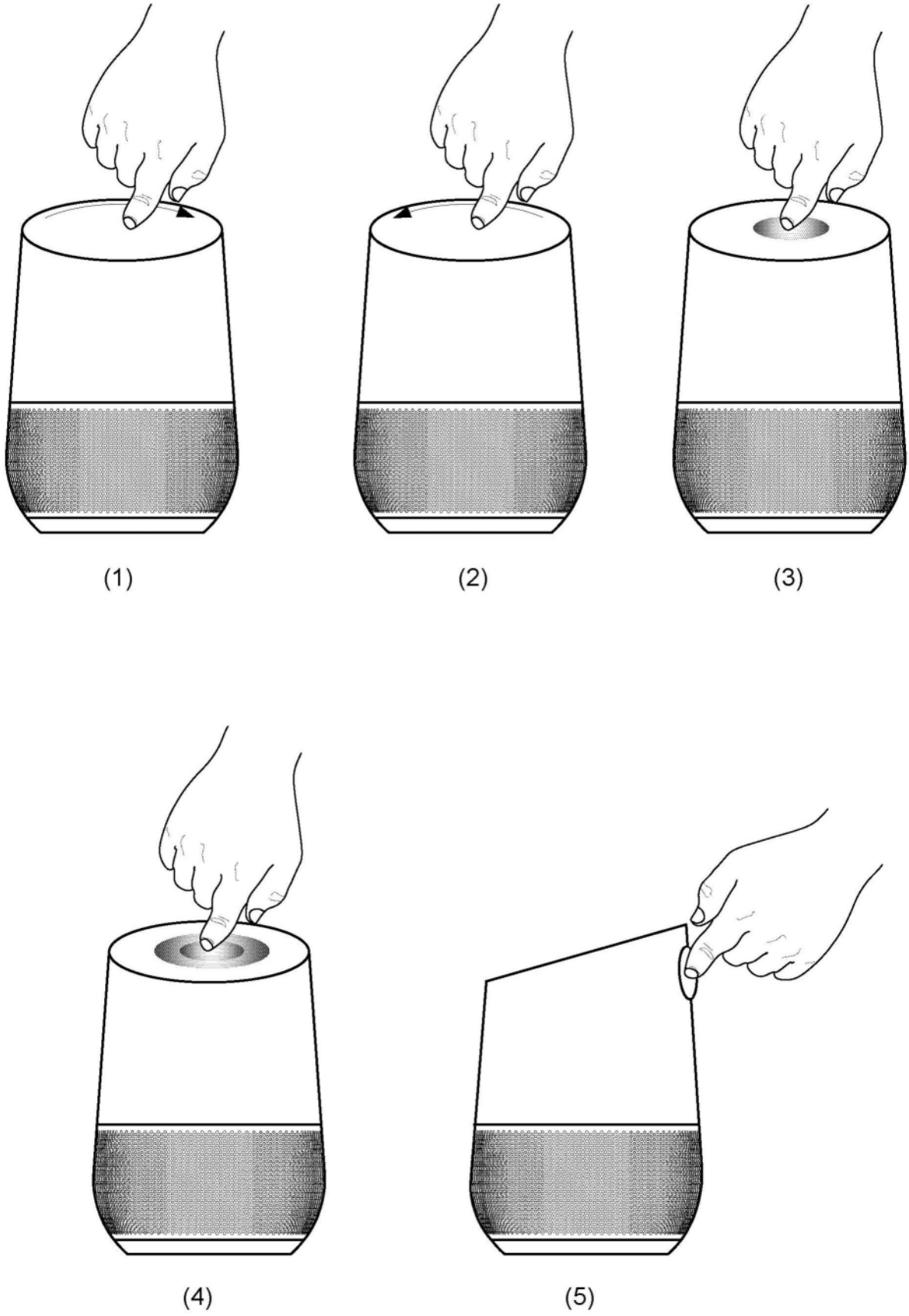


图6E

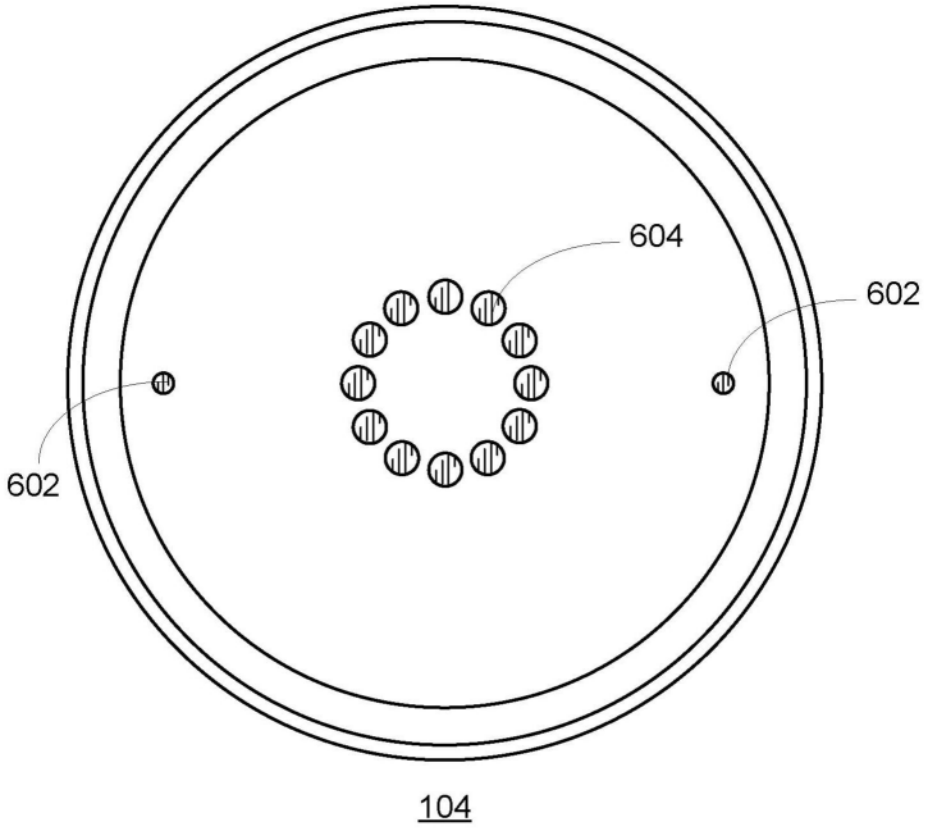

604
602
602
104

图6F

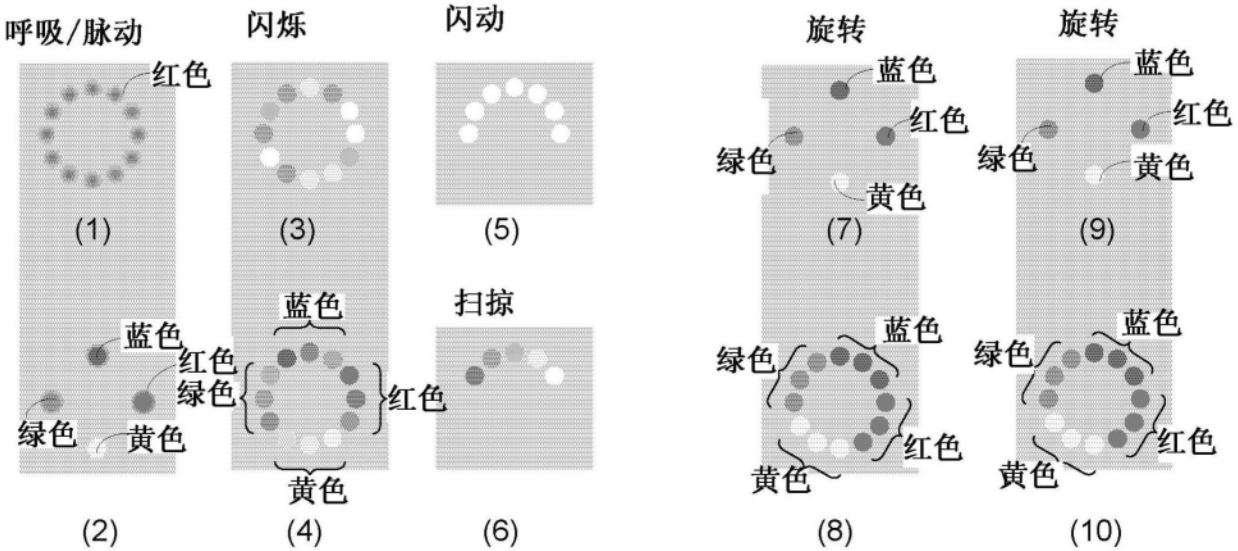

呼吸/脉动
闪烁
闪动
旋转
旋转
红色
蓝色
红色
绿色
黄色
(1)
(3)
(5)
(7)
(9)
蓝色
红色
绿色
绿色
黄色
红色
黄色
扫掠
蓝色
绿色
红色
黄色
蓝色
绿色
红色
黄色
(2)
(4)
(6)
(8)
(10)

图6G