(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107112014 A
(43)申请公布日 2017.08.29

(21)申请号 201580068237.6
(22)申请日 2015.12.11
(30)优先权数据
14/578,056 2014.12.19 US
(85)PCT国际申请进入国家阶段日
2017.06.14
(86)PCT国际申请的申请数据
PCT/US2015/065372 2015.12.11
(87)PCT国际申请的公布数据
WO2016/100139 EN 2016.06.23
(71)申请人 亚马逊技术股份有限公司
地址 美国华盛顿州
(72)发明人 P·S·万兰德 K·W·佩尔索尔 J·D·迈耶斯 J·M·辛普森 V·K·贡德蒂 D·R·托马斯 A·C·迈尔斯
(74)专利代理机构 北京市磐华律师事务所 11336
代理人 高伟 卜璐璐
(51)Int.Cl.
G10L 15/22(2006.01)

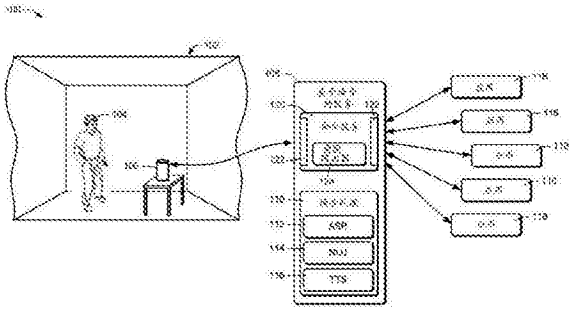
权利要求书2页 说明书14页 附图7页

(54)发明名称

在基于语音的系统中的应用焦点

(57)摘要

基于语音的系统包括在用户建筑物中的音频设备和通过多个应用来支持音频设备的使用的基于网络的服务。音频设备目的可以在于播放音频内容，例如音乐、音频书等。音频设备目的也可在于通过语音与用户交互作用。基于网络的服务监测从音频设备接收的事件消息以确定多个应用中的哪个当前具有语音焦点。当从用户接收到语音时，服务首先向当前具有主要语音焦点的应用提供对应的含义，如果有的话。如果没有当前具有主要语音焦点的应用，或如果具有主要语音焦点的应用不能够对含义做出响应，则服务向当前具有次要语音焦点的应用提供用户含义。



CN 107112014 A

1.一种方法,其包括:

向音频设备提供执行活动的命令,其中所述命令从多个应用当中标识有责任的应用;

从所述音频设备接收关于由所述音频设备显现的声音的事件消息,所述事件消息标识所述有责任的应用;

如果所述事件消息指示所述声音是用户交互的部分,则将所述有责任的应用指定为主要活动的;

接收由所述音频设备捕获的语音;

确定所述语音的含义;以及

如果在所述多个应用当中存在能对所述含义做出响应的主要活动应用,则请求所述主要活动应用对所述含义做出响应。

2.如权利要求1所述的方法,其还包括:

如果所述事件消息不指示所述音频是用户交互的部分,则将所述有责任的应用指定为次要活动的;以及

如果在所述多个应用当中没有能对所述含义做出响应的主要活动应用,则请求所述多个应用的次要活动应用对所述含义做出响应。

3.如权利要求2所述的方法,其还包括,如果在所述多个应用当中没有能对所述含义做出响应的主要活动应用,则:

确定所述次要活动应用能对所述含义做出响应;以及

将所述次要活动应用指定为主要活动的。

4.如权利要求2所述的方法,其还包括:

从所述主要活动应用接收所述主要活动应用将不对所述含义做出响应的指示;以及

响应于从所述主要活动应用接收到所述指示,请求所述次要活动应用对所述含义做出响应。

5.如权利要求1所述的方法,其还包括在请求所述主要活动应用对所述含义做出响应之前,确定所述主要活动应用能对所述含义做出响应。

6.如权利要求1所述的方法,其中所述分类指示所述音频是下列项中的至少一个:

是用户交互的部分的语音;

不是用户交互的部分的语音;

是用户交互的部分的音频内容;

不是用户交互的部分的音频内容;或

响应于由所述音频设备检测到条件而给出的音频通知。

7.如权利要求6所述的方法,其中所述音频通知包括:

不是用户交互的部分的背景音频通知;或

是用户交互的部分的前景音频通知。

8.如权利要求1所述的方法,其中:

所述命令指定标识所述有责任的应用的应用标识符;以及

所述事件消息指定所述应用标识符以标识所述有责任的应用。

9.如权利要求1所述的方法,其还包括:

确定在预定义时间段期间没有接收到标识所述有责任的应用的事件消息;以及

除去所述有责任的应用作为主要活动的的所述指定。

10.一种方法，其包括：

从设备接收关于由所述设备执行的第一行动的第一事件消息，所述第一事件消息从多个应用当中标识第一有责任的应用，其中所述多个应用中的每个能对由用户语音表达的一个或多个含义做出响应；

确定所述第一行动是用户交互的部分；

将所述第一有责任的应用指定为主要活动的；

标识第一用户语音的第一含义；以及

确定在所述多个应用当中有能对所述第一含义做出响应的主要活动应用；以及

选择所述主要活动应用以对所述第一含义做出响应。

11.如权利要求10所述的方法，其还包括：

从所述设备接收关于由所述设备执行的第二行动的第二事件消息，所述第二事件消息从所述多个应用当中标识第二有责任的应用；

确定所述第二行动不是用户交互的部分；

将所述第二有责任的应用指定为次要活动的；

确定第二用户语音的第二含义；

确定在所述多个应用当中没有能对所述第二含义做出响应的主要活动应用；以及

选择所述次要活动应用以对所述第二含义做出响应。

12.如权利要求11所述的方法，其还包括：

确定第三用户语音的第三含义；

确定所述主要活动应用将不对所述第三含义做出响应；以及

请求所述次要活动应用对所述第三含义做出响应。

13.如权利要求11所述的方法，其还包括：

确定第三用户语音的第三含义；

从所述主要活动应用接收所述主要活动应用将不对所述第三含义做出响应的指示；以及

请求所述次要活动应用对所述第三含义做出响应。

14.如权利要求10所述的方法，其中所述第一事件消息指示所述音频的分类，所述分类指示所述音频：

是用户交互的部分的语音；

不是用户交互的部分的语音；

是用户交互的部分的音频内容；

不是用户交互的部分的音频内容；或

是响应于由所述音频设备检测到条件而给出的音频通知。

15.如权利要求10所述的方法，其中所述第一事件消息指定标识所述第一有责任的应用的应用标识符。

在基于语音的系统中的应用焦点

[0001] 本申请要求于2014年12月19日提交的、标题为“Application Focus In Speech-Based Systems”的美国专利申请号14/578,056的优先权，该专利申请通过引用被全部并入本文。

[0002] 背景

[0003] 家、办公室、汽车和公共空间正变得更有线并与激增的计算设备例如上网本计算机、平板计算机、娱乐系统和便携式通信设备连接。当计算设备发展时，用户与这些设备交互的方式继续发展。例如，人可通过机械设备(例如键盘、鼠标等)、电气设备(例如触摸屏、触控板等)和光学设备(例如运动检测器、摄像机等)与计算设备交互。与计算设备交互的另一方式是通过音频设备，其理解人类语音并对人类语音做出响应。

[0004] 附图简述

[0005] 参考附图描述详细描述。在附图中，参考数字的最左边的数字标识参考数字首次出现的附图。相同的参考数字在不同附图中的使用指示相似或相同的部件或特征。

[0006] 图1是包括本地音频设备和远程基于语音的服务的话音交互计算体系结构的方框图。

[0007] 图2是示出在本地音频设备和远程基于语音的服务之间的信息流的例子的方框图。

[0008] 图3是示出与将含义路由到不同应用有关的信息流的例子的方框图。

[0009] 图4是示出选择和/或指定主要活动和次要活动应用的示例方法的流程图。

[0010] 图5是示出实现对主要活动应用的超时的示例方法的流程图。

[0011] 图6是示出处理来自本地音频设备的音频以确定由用户表示的含义并对含义做出响应的示例方法的流程图。

[0012] 图7是示出路由从用户话语得到的含义的示例方法的流程图。

[0013] 图8是示出本地音频设备的选定功能部件的方框图。

[0014] 图9是示出可部分地用于实现本文所述的基于语音的服务的服务器设备的部件的方框图。

[0015] 详细描述

[0016] 本公开描述用于与用户交互以提供服务的设备、系统和技术。如本文公开的系统可配置成接收用户语音并基于从在不同用户的家中的音频设备接收的音频来对用户语音做出响应。

[0017] 系统可包括基于语音的服务，其由基于网络的应用访问以结合家中音频设备来提供服务。应用可作为基于语音的服务的部分或由第三方提供者实现。基于语音的服务允许应用从家中音频设备接收信息并使用家中音频设备来执行操作。

[0018] 应用可将指令音频设备执行音频活动的命令发送到音频设备。例如，应用可指令音频设备播放音乐。作为音频活动的另一例子，应用可指令音频设备使用基于语音的服务或音频设备的文字到语音能力来播放语音。

[0019] 应用也可通过音频设备进行与用户的语音对话。语音对话包括与用户的特定行动

或意图有关的一序列语音问题、答案和/或陈述。更具体地，语音对话可包括一系列语音表达，其可包括用户的话语和由基于语音的服务产生的语音消息。语音对话例如可在初始的用户话语时开始。基于语音的服务可通过问问题例如“你想做什么”来做出响应。用户可通过在回答问题时做出陈述来做出响应。这个过程可迭代，直到基于语音的服务能够确定要采取的特定行动或要调用的功能为止。

[0020] 应用还可配置音频设备以响应于由音频设备本身检测或监测的条件来发出可听得见的通知。例如，音频设备可配置成在一天的指定时间或在指定的时间段之后发出警报。作为另一例子，音频设备可配置成响应于结合家自动化或家安全系统检测的事件而发出通知。通知可以是在背景中播放且不要求即时的用户注意或交互的被动通知。通知可以可选地包括比前景通知更大声或更显著并要求用户的更即时的行动或确认的主动或前景通知。

[0021] 音频设备的用户可通过讲话来向应用提供指令。音频设备捕获包括用户语音的声音并向基于语音的服务提供对应的音频信号。基于语音的服务在音频上执行自动语音识别(ASR)和自然语言理解(NLU)以确定用户语音的含义。含义作为例子可包括“播放音乐”、“暂停”、“停止”、“设置警报”、“呼叫Bob”、“播放天气简语”、“播放当前新闻摘要”、“订购披萨”、“创作电子邮件”、“音量调大”、“音量调小”、“消音”、“设置警报”、“取消”等。

[0022] 响应于确定用户语音的含义，基于语音的服务确定多个可用或活动应用中的哪个应被选择来对含义做出响应。单独的应用可向基于语音的服务注册以指示它们能够操纵的含义。可注册用于操纵单独的含义的多个应用。作为例子，可注册几个应用以通过执行关于由应用正执行的活动的“停止”行动来对“停止”含义做出响应。注意，从“停止”含义产生的行动取决于最终被请求操纵含义或对含义做出响应的应用。例如，一个应用可停止播放音乐而另一音乐可停止或取消警报。更一般地，从任何特定的含义产生的行动可以不同，取决于接收含义并对含义做出响应的应用。在一些情况下，应用可通过发起随后的对话打开例如通过产生对用户语音的含义的语音响应而对特定的含义做出响应。语音响应可请求澄清允许基于语音的服务完全确定用户的意图的信息。在其它情况下，应用可通过执行至少部分地通过含义指示的行动来做出响应，例如“由艺术家A演奏音乐”。

[0023] 当单独的应用指令音频设备发起活动时，应用提供与应用相关联的应用标识符。当音频设备执行活动时，音频设备将关于活动的事件消息发送到基于语音的服务。例如，事件消息可指示所请求的音乐已开始播放，音乐家列表的特定音轨已开始播放，语音已开始或结束，通知被给出，等等。每个事件消息指示应用的应用标识符，其负责事件所相关的活动。事件消息被传递到对应于应用标识符的应用。

[0024] 当用户讲话时，基于语音的系统执行ASR和NLU以识别用户的语音并确定语音的含义。然而，它可以是语音本身和语音的所确定的含义都不指示语音指向多个可用应用中的哪个。相应地，为了对用户语音的所确定的含义做出响应的目的，基于语音的系统具有监测音频设备的活动以留意哪些应用应被考虑为当前活动的路由部件。

[0025] 路由部件通过监测从音频设备接收的事件消息以确定哪个应用或哪些应用应当前被考虑为活动的来工作。更具体地，路由部件跟踪哪些应用负责由音频设备报告的最近音频事件。响应于标识有责任的应用的事件消息，路由部件将有责任的应用指定为主要活动的或次要活动的。主要活动应用被认为具有主要语音焦点。次要活动应用被认为具有次要语音焦点。所识别的用户语音的所确定的含义首先被提供到当前具有主要语音焦点的应

用。如果没有应用具有主要语音焦点或如果具有主要语音焦点的应用不能够操纵含义，则含义被提供到具有次要语音焦点的应用。

[0026] 为了留意哪些应用当前具有主要和次要语音焦点，路由部件监测来自音频设备的关于音频设备播放的音频的事件消息。单独的事件消息标识负责音频的应用，且也指示音频的类别。作为例子，分类可指示音频是否是用户交互的部分。如果分类指示音频是用户交互的部分，则路由部件将有责任的应用指定为具有主要语音焦点。如果分类指示音频不是用户交互的部分，则路由部件将有责任的应用指定为具有次要语音焦点。在所述实施方案中，只有一个应用(例如最近被指定为主要活动的应用)具有主要语音焦点，且只有一个应用(例如最近被指定为次要活动的应用)具有次要语音焦点。

[0027] 更一般地，基于涉及或打算发起双向用户交互的活动例如语音对话和活动用户通知的出现来准许主要语音焦点。基于不涉及双向用户交互的活动例如被动通知和音乐重放的出现来准许次要语音焦点。

[0028] 图1示出环境100，这些技术可在该环境中被实践。环境100可包括房间或其它用户建筑物102。用户建筑物可包括房屋、办公室、汽车和其它空间或区域。

[0029] 在用户建筑物102内的是用户104和一个或多个音频设备106。音频设备106在一些实施方案中可包括具有一个或多个麦克风、扬声器和网络接口或其它通信接口的基于网络的或网络可访问的设备。在某些实施方案中，音频设备106也可具有为了用户交互而设计的其它元件，包括按钮、旋钮、灯、指示器和各种类型的传感器、输入元件和输出元件。

[0030] 音频设备106从用户104接收口头命令并响应于该命令而提供服务。所提供的服务可包括执行行动或活动、再现媒体、得到和/或提供信息、监测本地条件并基于本地条件来提供通知、通过音频设备106经由所产生的或合成的语音来提供信息、代表用户104发起基于互联网的服务，等等。

[0031] 在图1所示的实施方案中，音频设备106与网络可访问的基于语音的服务108通信。基于语音的服务108可被实现为相对于音频设备106远程地定位的基于网络或基于云的服务。例如，基于语音的服务108可由企业组织和/或服务提供者实现以支持位于不同的用户建筑物102中的多个音频设备106，用户建筑物又可位于广泛变化的地理位置上。

[0032] 基于语音的服务108在一些实例中可以是经由广域网例如互联网来维护和可访问的网络可访问计算平台的部分。网络可访问计算平台例如这可以使用术语例如“立即响应式计算”、“软件即服务(SaaS)”、“平台计算”、“网络可访问平台”、“云服务”、“数据中心”等被提到。

[0033] 在音频设备106和基于语音的服务108之间的通信可通过各种类型的数据通信网络(包括局域网、广域网和/或公共互联网)来实现。蜂窝和/或其它无线数据通信技术也可用于与基于语音的服务108通信。用户建筑物102可包括本地网络支持设备以便于与基于语音的服务108通信，例如无线接入点、网络路由器、通信集线器等。

[0034] 基于语音的服务108可与各种服务和/或应用交互，支持多个音频设备106。作为例子，这样的服务可包括语音处理服务110。语音处理服务110可配置成从音频设备106接收实时音频或语音信息，以便识别用户语音，确定由语音处理的用户含义，并在用户含义的履行中执行行动或提供服务。例如，用户可以讲预定义的命令(例如“醒来”；“睡眠”)或可在与音频设备106交互时使用更随便的说话风格(例如，“我想去看电影。请告诉我在本地电影院正

播放什么”)。用户命令本质上可以是任何类型的操作,例如数据库查询、请求和消费娱乐(例如游戏、找到并播放音乐、电影或其它内容等)、个人管理(例如记入日历、做笔记等)、在线购物、财务交易等。

[0035] 语音和语音相关信息可以用很多不同的形式被提供到语音处理服务110。在一些实现中,语音相关信息可包括来自音频设备106的连续音频信号或流。可选地,语音相关信息可包括响应于在用户建筑物102内的检测到的声音而被提供到语音处理服务110的音频剪辑或段。在一些情况下,音频设备106可执行语音识别并向基于语音的服务108提供以文本的形式的用户语音。在一些实现中,基于语音的服务108可通过产生或指定语音来与用户104交互,语音又由音频设备106再现。语音合成可由语音处理服务110或由音频设备106执行。

[0036] 在所述实施方案中,语音处理服务110包括用于识别语音、理解所识别的语音的含义并用于产生语音的部件或功能。具体地,语音处理服务110包括自动语音识别(ASR)服务112、自然语言理解(NLU)服务114和文本到语音(TTS)服务116。也可提供各种其它类型的语音处理功能。

[0037] ASR服务112可使用各种技术来创建在音频信号中表示的语音字的完全的转录物。例如,ASR服务112可参考各种类型的模型,例如声模型和语言模型,以识别在音频信号中表示的语音的字。在很多情况下,通过培训例如通过对很多不同类型的语音采样并手动地分类来创建模型,例如这些模型。

[0038] 声模型可将语音表示为对应于音频波形随着时间的过去的特征的一系列矢量。特征可对应于频率、音高、振幅和时间模式。可基于培训数据的大集合来创建统计模型例如隐马尔科夫模型(HMM)和高斯混合模型。所接收的语音的模型接着与培训数据的模型比较以找到匹配。

[0039] 语言模型描述诸如语法规则、公共字使用和模式、字典含义等的东西,以建立字序列和组合的概率。使用语言模型的语音的分析可取决于上下文,例如出现在当前正被分析的语音的任何部分之前或之后的字。

[0040] ASR可提供识别候选项,其可包括字、短语、句子或语音的其它段。候选项可伴随有统计概率,每个统计概率指示在对应的候选项的准确度中的“置信度”。一般,具有最高置信度分数的候选项被选择为语音识别的输出。

[0041] NLU服务114分析由ASR服务112提供的字流,并产生字流的含义的表示。例如,NLU服务114可使用分析程序和语法规则来分析句子并用以容易由计算机处理的方式传达概念的正式定义的语言产生句子的含义的表示。例如,含义可实质上被表示为槽的分级集合或帧和槽值,其中每个槽对应于在语义上定义的概念。因此,句子的含义可在语义上由槽的帧和槽值表示。NLU也可使用从培训数据产生的统计模型和模式来利用在一般语音中的字之间的统计相关性。

[0042] 基于语音的服务108可配置成支持多个基于网络的应用118。应用118通过基于语音的服务108与音频设备106交互以至少部分地基于由音频设备106捕获或提供的用户语音结合音频设备106来提供功能。更特别地,应用118配置成通过基于语音的服务108的命令服务120进行通信,命令服务120充当设备代理以从音频设备106接收信息并向音频设备106提供指令、信息和内容。在一些情况下,命令服务120可使用第一组数据格式和/或协议来与音

频设备106通信，允许相对低级别或详细数据的传输。命令服务120可使用第二组数据格式和/或协议来与应用118通信，允许信息在相对较高的抽象级别处或使用不同类型的通信协议来传输。

[0043] 应用118可在一些情况下被实现为基于web的或基于网络的应用或服务。例如，特定的应用118可由基于语音的服务108的提供者或由第三方提供者实现为服务器或服务，并可通过网络例如互联网与命令服务120通信。在其它情况下，应用118可存在或安装在与用户104相关联的物理设备例如用户104的计算机或移动设备上，并可通过互联网或其它广域网与命令服务120通信。

[0044] 基于语音的服务108和命令服务120可配置成根据web服务模型来与音频设备106和/或应用118交互，且基于语音的服务108的功能可被实现为一个或多个web服务。通常，web服务可包括任何类型的计算服务，其经由包括一个或多个基于互联网的应用层数据传输协议例如一种版本的超文本传输协议(HTTP)或另一适当的协议的请求接口而对请求客户端变得可用。

[0045] 命令服务120可暴露一个或多个网络可访问API或应用接口122。API 122可被实现为具有统一资源定位器(URL)例如http://storageservice.domain.com的web服务端点。

[0046] 应用118可由各种卖方和/或提供者设计并提供以结合音频设备106来工作和/或使用音频设备106通过API 122和相关联服务来提供服务。应用118可提供范围从电子邮件到游戏的功能。应用118可包括启用语音的应用，其响应于用户语音和从用户语音得到的含义来执行行动。相应地，应用118可使它们的服务部分地基于语音和由音频设备106和语音处理服务110提供的语音相关信息，包括所识别的语音、从语音得到的含义和已从用户语音解释的意图或命令。此外，应用118可提供在音频设备106上被再现为语音的文本，并可经由命令服务120和API 122向或为音频设备106提供其它指令和命令。

[0047] 在一些实现中，所示应用118可以是其它应用的部件例如所谓的“小型应用”。每个应用或小型应用可由应用标识符标识。应用标识符可由基于语音的服务108分配或由应用本身提供。

[0048] 作为一个例子，应用可包括向音频设备106提供音乐或其它内容以由音频设备106显现的音频应用。

[0049] 每个应用118可与命令服务120通信以指示或记录它能够操纵的语音含义。多于一个应用118可能能够操纵任何给定含义或对任何给定含义做出响应。可选地，命令服务120可查询单独的应用以从应用接收关于它们是否可或将对某些含义做出响应的指示。

[0050] 命令服务120包括向适当的应用118提供所识别或所标识的语音含义的路由部件124。如将在下面更详细描述的，路由部件124根据当前正由音频设备106执行的活动来分配主要语音焦点和次要语音焦点。当含义被确定时，具有主要焦点(如果有的话)的应用首先被给予对含义做出响应的机会。

[0051] 图2示出可出现在应用118和音频设备106之间的通信的例子。为了清楚的目的，没有示出充当通信媒介物的命令服务120。

[0052] 应用118可使命令202被发送到音频设备106。命令202包括或指定对应于并标识应用118的应用标识符，其在图2中被称为AppID。命令202可指定将被音频设备106进行或执行的活动。例如，命令可指定由音频设备106播放的音频内容，例如音乐。作为另一例子，命令

202可指定将由音频设备106转换成语音并播放为音频的文本。作为另一例子，命令202可配置将由音频设备106实现的通知。

[0053] 在一些情况下，命令202可指定所命令的活动或由音频设备106响应于活动而产生的音频是否被考虑为交互式的。形成用户交互的部分的音频例如作为用户对话的部分的语音可被考虑为交互式的。不是用户交互的部分的音频例如音乐可被考虑为非交互式的。某些类型的再现的语音当不是用户交互的部分时可被考虑为非交互式的。例如，应用可产生语音以描述当前天气或交通条件，其不是语音交互的部分且将因此被考虑为非交互式的。

[0054] 在操作期间，音频设备106产生事件消息204并将事件消息204发送回到命令服务120。每个事件消息204描述音频事件或已出现在音频设备106处的其它事件。例如，事件消息204可指定某个类型的声音被播放，文本到语音重放已开始或结束，非交互式内容已开始或停止，内容或媒体的重放已进行到某个点，媒体项的重放已结束以及随后的媒体项的重放已开始，等等。事件消息204也可指定音频通知已由音频设备发起。

[0055] 每个事件消息指示负责活动的应用的应用标识符(AppID)，所述音频事件是该活动的一部分。事件消息204可由命令服务120传递到有责任的应用，如由AppID指定的，使得有责任的应用可监测它已请求的活动的进展。

[0056] 每个事件消息204也可指定所述音频是交互式的还是非交互式的。交互式音频包括是用户交互的部分的音频。非交互式音频是不是用户交互的部分的音频。一些事件消息可明确地指定对应的事件是否是交互式的。在其它情况下，事件的性质可内在地指示对应的事件是否是交互式的。例如，与音乐重放的状态有关的某些事件可被考虑为非交互式事件，即使关于这样的事件的事件消息可以不明确地将事件分类为交互式的或非交互式的。

[0057] 图3示出基于语音的服务108如何处理所接收的用户话语以向适当的应用118提供所确定的含义。音频设备106捕获作为音频信号被传输到基于语音的服务108的用户话语或语音302。语音处理服务110使用ASR和NLU来分析音频信号以确定用户语音302的含义304。路由部件124接收含义304的语义表示。路由部件130也接收并监测事件消息204。

[0058] 路由部件130监测事件消息204(当它们由音频设备106产生时)以确定哪个应用118被考虑为当前活动的。响应于事件消息204，路由部件124可标识主要活动应用和/或次要活动应用。被标识为主要活动应用的应用被考虑为具有主要语音焦点。被标识为次要活动应用的应用被考虑为具有次要焦点。在本文所述的实施方案中，只有单个应用被考虑为在任何给定时间是主要活动的，以及只有单个应用被考虑为在任何给定时间是次要活动的，虽然在某些其它实施方案中情况可能并不总是这样。

[0059] 当接收到含义304的语义表示时，路由部件124基于主要和次要活动的应用的以前标识向应用118之一提供含义304的表示。通常，主要活动的应用被给予操纵含义的第一机会，如果它能够。否则，如果没有当前是主要活动的应用或如果当前是主要活动的应用不能够操纵含义，则当前是次要活动的应用被给予操纵含义的机会。

[0060] 响应于从音频设备106接收到事件消息204，作为背景操作来执行将应用指定为主要或次要活动的。当从音频设备106接收到话语时，与将应用指定为主要或次要活动的过程独立和异步地执行含义的路由。

[0061] 图4示出可由路由部件124执行来基于从音频设备106接收的事件消息选择主要活动应用和次要活动应用的示例方法400。

[0062] 行动402包括从音频设备接收关于作为活动的部分的由音频设备播放的音频的事件消息204。事件消息204可包括事件描述404和对应于应用118的应用标识符406，应用118负责音频事件和/或活动，所述音频事件是该活动的部分。

[0063] 事件消息204可在一些情况下也包含指示音频事件的音频是否被考虑为交互式的或非交互式的事件分类408。交互式音频可包括作为与用户的语音对话或交互的部分的语音。其它类型的音频例如音乐或不是与用户的语音对话或交互的部分的语音可被考虑为背景或非交互式音频。在一些情况下，事件分类408可从事件消息204省略，且与事件消息204一起提供的事件或其它元数据的性质可指示对应的事件是否是交互式的。

[0064] 除了对话语音以外，响应于由音频设备106监测的条件而由音频设备106产生的某些类型的通知可被考虑为交互式的。虽然这样的通知不一定是语音对话的部分，但是它们可被考虑为用户交互的部分，因为它们请求即时的用户输入。例如，通知可包括用户被期望回复的可听得见的警报声音，例如通过说词“停止警报”。

[0065] 由音频设备106响应于由音频设备106监测的条件而产生的其它类型的通知可被考虑为非交互式的。例如，通知可包括打算向用户警告非关键条件例如消息或电子邮件的接收的背景声，这并不打算请求即时用户输入。

[0066] 通常，分类408或与事件消息相关联的其它信息可指示对应的音频包括：

[0067] 是用户交互的部分的语音；

[0068] 不是用户交互的部分的语音；

[0069] 是用户交互的部分的音频内容；

[0070] 不是用户交互的部分的音频内容；或

[0071] 响应于由音频设备检测到条件而给出的音频通知。

[0072] 音频通知可包括不是用户交互的部分的背景音频通知或是用户交互的部分的前景音频通知。

[0073] 行动410包括确定事件分类408或事件消息204的其它数据是否指示所接收的事件消息是针对交互式事件或非交互式事件。在事件消息204明确提供分类408的情况下，这可涉及检查分类408。否则，行动410可包括基于事件的类型或描述来确定对应的事件是否是交互式的，其中某些事件或某些类型的事件被定义为交互式的，而其它事件或其它类型的事件被定义为非交互式的。在一些情况下，例如与媒体例如音乐的重放有关的事件可按照定义被考虑为非交互式的。

[0074] 如果事件是交互式的，则执行行动412，其将任何当前指定的主要活动应用而不是有责任的应用指定为不再是主要活动。此外，执行行动414，其将有责任的应用(由应用标识符406指示)指定为现在是主要活动的并具有主要焦点。

[0075] 如果事件是非交互式的和/或有责任的应用未被行动414指定为主要活动的，则执行行动416，其将任何当前指定的主要活动应用而不是有责任的应用指定为不再是次要活动的。此外，执行行动418，其将有责任的应用(由应用标识符406指示)指定为现在是次要活动的并具有次要焦点。

[0076] 注意，某些类型的事件可内在地与对应的应用相关联，且应用标识符在这些情况下可被省略。例如，与从音频设备106的Bluetooth®外围设备接收的音频的重放有关的消息可内在地与应用118的特定应用相关联。

[0077] 图5示出示例方法500，其可关于已被指定为主要活动的有责任的应用执行，如在块502指示的，例如可根据图4的方法400发生。行动504包括确定是否预定义时间段已过去或超时已到期。如果该时间段已过去或超时已到期，则执行行动506，其除去将有责任的应用作为主要活动的指定。如果该时间段已过去或超时未到期，则循环地重复行动504。每当将当前指定的主要活动应用最新指定为主要活动的时，可重置时间段，诸如响应于最新接收的事件消息，事件消息导致通过图4的行动416重新分配主要焦点。

[0078] 方法500确保主要活动应用将不失去语音焦点，如果指定应用的交互式事件的事件消息未在指定的时间段期间被接收到。应用可稍后复得主要焦点，如果指定应用的应用标识符并指定交互式事件分类的新事件消息被接收到。

[0079] 图6示出处理用户语音的示例方法600。行动602包括接收包含用户语音的音频信号。行动604包括使用ASR来分析音频信号以识别用户语音并产生用户语音的转录物。行动606包括使用NLU来分析所识别的语音以确定用户语音的含义并产生用户语音及其含义的语义表示。行动608包括路由应用118的一个或多个的表示。

[0080] 图7示出将语音含义的语义表示路由到多个应用118之一的示例方法700。行动702包括接收含义的表示。行动704包括确定在多个应用118当中是否有被已指定为主要活动的且因此具有主要焦点的应用。如果有这样的主要活动应用，则执行确定主要活动应用是否可对含义做出响应的行动706。可通过参考指示哪些含义可由哪些应用操纵的应用的以前记录来执行行动706。可选地，可查询主要活动应用以确定它当前是否可对含义做出响应。如果主要活动应用可以或将对含义做出响应，则执行向应用提供含义的语义表示和/或请求主要活动应用对含义做出响应的行动708。在一些情况下，可组合行动706和708：含义的表示可连同使应用对含义做出响应的请求一起传递到主要活动应用，且应用可通过接受请求或指示应用将不对含义做出响应来做出响应。

[0081] 如果没有当前主要活动的应用，如果主要应用指示它将不或不能够对所确定的含义做出响应，或如果否则确定主要活动应用将不对含义做出响应，则执行行动710，其确定在多个应用118当中是否有已被指定为次要活动的并因此具有次要焦点的应用。如果有这样的次要活动应用，则执行行动712，其确定次要活动应用是否能够对所确定的含义做出响应。可通过参考指示哪些含义可由哪些应用操纵的应用的以前记录来执行行动712。可选地，可查询次要活动应用以确定它是否可当前对所确定的含义做出响应。如果次要活动应用可以或将对含义做出响应，则执行行动714，其向次要活动应用提供含义的语义表示和/或请求次要活动应用对含义做出响应。在一些情况下，可组合行动710和712：含义的语义表示可连同使次要活动应用对含义做出响应的请求一起传递到次要活动应用，且应用可通过接受请求或谢绝该请求来做出响应。

[0082] 当次要活动应用对含义做出响应时或当次要活动应用指示它可对含义做出响应时，也可执行行动716。行动716包括将次要活动应用指定为现在是主要活动的且因此具有主要语音焦点。当应用被指定为主要活动的时，以前被指定为主要活动的任何其它应用然后被指定为不再是主要活动的。注意，在某些实施方案中可只对某些类型的应用或事件执行行动716。作为例子，含义“提高音量”可被考虑为短暂命令或事件，且可以不导致对应的应用被给予主要焦点。

[0083] 如果没有当前是次要活动的应用，如果次要活动应用指示它将不或不能够对所确

定的含义做出响应，或如果否则确定次要活动应用将不对含义做出响应，则执行行动718，其确定在多个应用当中是否有可操纵所确定的含义的另一应用。可通过参考指示哪些含义可由哪些应用操纵的应用的以前记录来执行行动718。可选地或此外，可查询其它应用以确定它们是否可当前对含义做出响应。如果另一应用可操纵含义，则执行行动720，其向其它应用提供含义的表示和/或请求其它应用对含义做出响应。

[0084] 当其它应用之一对含义事件做出响应时或当否则非活动应用指示它可对含义做出响应时，也可执行行动722。行动722包括将响应应用指定为是主要活动的且因此具有主要语音焦点。当应用被指定为主要活动的时，以前被指定为主要活动的任何其它应用然后被指定为不再是主要活动的。注意，在某些实施方案中可只对不考虑为短暂的某些类型的应用或事件执行行动722。

[0085] 行动718可包括向不同的应用以它们向命令服务120注册的顺序提供含义的语义表示，较早的已注册应用被给予优于稍后注册的应用的优先级。可选地，每个应用可被请求提供指示含义被预期针对应用的可能性的置信度水平。例如，音乐重放应用在它当前不播放音乐时可以将本身考虑是“暂停”含义的相对不可能的接收方，即使它以前指示操纵“暂停”含义的能力。含义可接着被提供到提供最高置信度水平的应用。

[0086] 图8示出音频设备106的示例配置。在图8的例子中，音频设备106具有操作逻辑，其包括处理器802和存储器804。存储器804可包含以指令的形式的应用和程序，指令由处理器802执行以执行实现音频设备106的期望功能的动作或行动。存储器804可以是一种类型的计算机存储介质，并可包括易失性和非易失性存储器。因此，存储器804可包括但不限于RAM、ROM、EEPROM、闪存或其它存储器技术。

[0087] 图8示出可由音频设备106提供并由存储器804存储以实现音频设备106的功能的应用和/或程序的几个例子，但是可在各种实施方案中提供功能的很多其它应用和类型。

[0088] 音频设备106可具有配置成管理在音频设备106内并耦合到音频设备106的硬件和服务的操作系统806。此外，音频设备106可包括音频处理模块808，其从用户建筑物102接收音频并处理所接收的音频以执行行动并响应于用户语音而提供服务。在一些情况下，音频处理模块808可执行语音识别和关于所接收的音频的自然语言理解。在其它情况下，音频处理模块可将所接收的音频传送到基于语音的服务108，其可使用语音处理服务110来执行语音处理，例如语音识别和自然语言理解。音频处理模块808可执行各种类型的音频处理，包括过滤、压缩等，并可利用数字信号处理器或信号处理的其它方法。

[0089] 音频处理模块808也可负责制造或产生语音。例如，音频设备106可从基于语音的服务108接收文本，并可将文本转换成语音。可选地，音频设备106可接收由音频处理模块808处理的音频信号用于由音频设备106再现。

[0090] 音频设备106可具有配置成建立与基于语音的服务108的通信信道的通信部件810。各种类型的通信协议可由通信部件810支持。在一些情况下，通信部件810可配置成使用各种类型的网络通信技术之一通过API 122来建立与基于语音的服务108的安全和/或加密通信信道。

[0091] 音频设备106也可具有配置成响应于由音频设备106执行的音频活动来提供如上所述的事件消息的事件报告模块812。在一些实现中，音频设备106可向基于语音的服务108前摄地提供事件消息。在其它实现中，基于语音的服务可轮询或查询音频设备106以得到事

件消息。

[0092] 除了上面所述的软件功能以外，音频设备106还可实现各种类型的其它应用、功能和/或服务814。例如，其它服务814可包括在图8中被称为媒体播放器816的音频功能或应用，其用于响应于用户指令或在基于语音的服务108或应用118的指导下播放歌曲或其它类型的音频。媒体播放器816可从基于语音的服务108、从应用118的一个或多个或从第三方服务例如音乐服务、podcast服务等接收音频。例如，基于语音的服务108和/或应用118之一可指令音频设备106得到并播放来自第三方服务的特定歌曲。当接收到这个指令时，音频设备106的媒体播放器816可联系第三方服务，发起歌曲的流式传送或下载，并可接着播放歌曲而没有来自基于语音的服务108或应用118的指令音频设备106播放歌曲的另外的指令或信息。类似地，可将音乐家列表提供到媒体播放器816用于由音频设备106的媒体播放器816重放。

[0093] 音频设备106还可包括各种类型的基于硬件的部件或功能，包括设备接口818和通信接口820。设备接口818可提供到辅助设备例如Bluetooth™设备、远程显现设备、远程传感器等的连接。通信接口820可包括网络接口和允许音频设备106连接到基于语音的服务108并与基于语音的服务108通信的其它类型的接口。

[0094] 音频设备106可具有各种类型的指示器822，例如用于将操作信息传递给用户104的灯。指示器822可包括LED(发光二极管)、平板显示元件、文本显示器等。

[0095] 音频设备106还可具有可包括按钮、旋钮、滑块、触摸传感器等的各种类型的物理控件824。物理控件824可用于基本功能，例如启用/禁用音频设备106，设置音频设备106的音频输入音量，等等。

[0096] 音频设备106可包括麦克风单元826，其包括一个或多个麦克风以接收音频输入，例如用户话音输入。麦克风单元826在一些实现中可包括定向麦克风阵列，使得来自不同方向的声音可选择性地被接收和/或增强。音频设备106还可包括用于音频的输出的扬声器828。

[0097] 除了物理控件824和麦克风单元826以外，音频设备106还可具有各种其它类型的传感器830，其可包括静止和视频摄像机、深度传感器、3D(三维)摄像机、红外传感器、接近度传感器、用于测量周围声音和光的水平的传感器等。音频设备106还可具有分析能力，其利用来自传感器839的信息来确定用户建筑物102的特性和在用户建筑物102内的环境条件。例如，音频设备106可能能够分析光信息以确定房间的3D特性，包括在房间内的人或物体的存在和/或身份。作为另一例子，音频设备106可能能够检测并评估房间的音频特性，以便优化音频重放。

[0098] 音频设备106还可具有用于与用户104交互的其它用户接口(UI)元件832。其它UI元件可包括显示面板、投影仪、触控板、键盘等。

[0099] 在某些情况中，音频设备106可包括移动设备，例如智能电话、平板计算机、眼镜、手表等。移动设备可具有传感器，例如罗盘、加速度计、陀螺仪、全球定位接收器等以及具有基于应用来确定各种环境信息并访问基于网络的信息资源的能力。

[0100] 图9示出可用于实现基于语音的服务108的功能的服务器900的相关部件和/或可用于提供如本文所述的服务的其它部件。通常，功能元件可由一个或多个服务器实现，上面所述的各种功能以各种方式分布在不同的服务器当中。服务器可一起或单独地被定位，并

被组织为虚拟服务器、服务器组和/或服务器场。所述功能可由单个实体或企业的服务器提供，或可利用多个实体或企业的服务器和/或服务。

[0101] 在非常基本的配置中，示例服务器900可包括由一个或多个处理器组成的处理单元902和相关联存储器904。根据服务器900的配置，存储器904可以是一种类型的计算机存储介质并可包括易失性和非易失性存储器。因此，存储器904可包括但不限于RAM、ROM、EEPROM、闪存或其它存储器技术。

[0102] 存储器904可用于存储由处理单元902可执行的任何数量的功能部件。在很多实施方案中，这些功能部件包括由处理单元902可执行且当被执行时实现用于执行上面所述的行动的操作逻辑的指令或程序。

[0103] 在存储器904中存储的功能部件可包括操作系统906和与远程设备例如计算机、媒体消费设备等交互的web服务部件908。存储器904还可具有实现语音处理服务110、命令服务120、API 122和路由部件124的指令。在一些情况下，应用118的一个或多个也可被实现为存储在存储器904中的功能部件。

[0104] 服务器900当然可包括在图9中没有示出的很多其它逻辑、编程和物理部件。

[0105] 注意，虽然音频设备106在本文被描述为在家里使用的话音控制的或基于语音的音频设备，但是本文所述的技术可结合各种不同类型的设备例如电信设备和部件、免提设备、娱乐设备、媒体重放设备、平板计算机、个人计算机、专用设备等来实现。

[0106] 上面所述的实施方案可例如使用计算机、处理器、数字信号处理器、模拟处理器等编程地实现。然而，在其它实施方案中，可使用专门或专用电路——包括模拟电路和/或数字逻辑电路——来实现部件、功能或元件的一个或多个。

[0107] 而且，虽然已用某些特征所特有的语言描述了主题，但是应理解，在所附权利要求中定义的主题不一定限于所描述的特定特征。更确切地，特定特征被公开为实现权利要求的说明性形式。

[0108] 条款

[0109] 1.一种系统，其包括：

[0110] 命令服务，其配置成：与多个应用通信，与音频设备通信，并将命令发送到音频设备以为音频应用执行提供音频内容以由音频设备播放的活动，其中命令指定对应于音频应用的应用标识符；

[0111] 控制逻辑，其配置成执行包括以下项的动作：

[0112] 从音频设备接收关于由音频设备播放的声音的事件消息，其中事件消息指定对应于音频应用的应用标识符；

[0113] 如果事件消息指示由音频设备播放的声音是与用户的语音交互的部分，则将音频应用指定为主要活动的；

[0114] 如果事件消息指示由音频设备播放的声音不是与用户的语音交互的部分，则将音频应用指定为次要活动的；

[0115] 语音识别服务，其配置成从音频设备接收音频信号并识别在音频信号中的用户语音；

[0116] 语言理解服务，其配置成确定用户语音的含义；

[0117] 控制逻辑，其配置成执行包括以下项的另外的行动：

[0118] 如果在多个应用当中存在主要活动应用，则请求主要活动应用通过(a)执行至少部分地通过用户语音的含义指示的第一行动或(b)产生对用户语音的第一语音响应来对用户语音做出响应；以及

[0119] 如果在多个应用当中没有主要活动应用且如果在多个应用当中存在次要活动应用，则请求次要活动应用通过(a)执行至少部分地通过用户语音的含义指示的第二行动或(b)产生对用户语音的第二语音响应来对用户语音做出响应。

[0120] 2.如条款1所述的系统，其中事件消息指定指示声音是否是与用户的语音交互的部分的事件分类，分类指示声音包括下列项中的至少一个：

[0121] 是用户交互的部分的语音；

[0122] 不是用户交互的部分的语音；

[0123] 是用户交互的部分的音频内容；

[0124] 不是用户交互的部分的音频内容；或

[0125] 响应于由音频设备检测到条件而给出的音频通知。

[0126] 3.如条款1所述的系统，其中事件消息指示第二音频是响应于由音频设备检测到条件而给出的通知，动作还包括将音频应用指定为主要活动的。

[0127] 4.如条款1所述的系统，行动还包括：

[0128] 确定在预定义时间段期间没有接收到标识音频应用的事件消息；以及

[0129] 除去音频应用作为主要活动的指定。

[0130] 5.一种方法，其包括：

[0131] 向音频设备提供执行活动的命令，其中命令从多个应用当中标识有责任的应用；

[0132] 从音频设备接收关于由音频设备显现的声音的事件消息，事件消息标识有责任的应用；

[0133] 如果事件消息指示声音是用户交互的部分，则将有责任的应用指定为主要活动的；

[0134] 接收由音频设备捕获的语音；

[0135] 确定语音的含义；以及

[0136] 如果在多个应用当中存在可对含义做出响应的主要活动应用，则请求主要活动应用对含义做出响应。

[0137] 6.如条款1所述的方法，其还包括：

[0138] 如果事件消息不指示音频是用户交互的部分，则将有责任的应用指定为次要活动的；以及

[0139] 如果在多个应用当中没有可对含义做出响应的主要活动应用，则请求多个应用的次要活动应用对含义做出响应。

[0140] 7.如条款2所述的方法，其还包括，如果在多个应用当中没有可对含义做出响应的主要活动应用，则：

[0141] 确定次要活动应用可对含义做出响应；以及

[0142] 将次要活动应用指定为主要活动的。

[0143] 8.如条款2所述的方法，其还包括：

[0144] 从主要活动应用接收主要活动应用将不对含义做出响应的指示；以及

[0145] 响应于从主要活动应用接收到指示，请求次要活动应用对含义做出响应。

[0146] 9.如条款1所述的方法，其还包括在请求主要活动应用对含义做出响应之前确定主要活动应用可对含义做出响应。

[0147] 10.如条款1所述的方法，其中分类指示音频是下列项中的至少一个：

[0148] 是用户交互的部分的语音；

[0149] 不是用户交互的部分的语音；

[0150] 是用户交互的部分的音频内容；

[0151] 不是用户交互的部分的音频内容；或

[0152] 响应于由音频设备检测到条件而给出的音频通知。

[0153] 11.如条款6所述的方法，其中音频通知包括：

[0154] 不是用户交互的部分的背景音频通知；或

[0155] 是用户交互的部分的前景音频通知。

[0156] 12.如条款1所述的方法，其中：

[0157] 命令指定标识有责任的应用的应用标识符；以及

[0158] 事件消息指定应用标识符以标识有责任的应用。

[0159] 13.如条款1所述的方法，其还包括：

[0160] 确定在预定义时间段期间没有接收到标识有责任的应用的事件消息；以及

[0161] 除去有责任的应用作为主要活动的指定。

[0162] 14.一种方法，其包括：

[0163] 从设备接收关于由设备执行的第一行动的第一事件消息，第一事件消息从多个应用当中标识第一有责任的应用，其中多个应用中的每个可对由用户语音表达的一个或多个含义做出响应；

[0164] 确定第一行动是用户交互的部分；

[0165] 将第一有责任的应用指定为主要活动的；

[0166] 标识第一用户语音的第一含义；以及

[0167] 确定在多个应用当中有可对第一含义做出响应的主要活动应用；以及

[0168] 选择主要活动应用以对第一含义做出响应。

[0169] 15.如条款10所述的方法，其还包括：

[0170] 从设备接收关于由设备执行的第二行动的第二事件消息，第二事件消息从多个应用当中标识第二有责任的应用；

[0171] 确定第二行动不是用户交互的部分；

[0172] 将第二有责任的应用指定为次要活动的；

[0173] 确定第二用户语音的第二含义；

[0174] 确定在多个应用当中没有可对第二含义做出响应的主要活动应用；以及

[0175] 选择次要活动应用以对第二含义做出响应。

[0176] 16.如条款11所述的方法，其还包括：

[0177] 确定第三用户语音的第三含义；

[0178] 确定主要活动应用将不对第三含义做出响应；以及

[0179] 请求次要活动应用对第三含义做出响应。

[0180] 17.如条款11所述的方法，其还包括：

[0181] 确定第三用户语音的第三含义；

[0182] 从主要活动应用接收主要活动应用将不对第三含义做出响应的指示；以及

[0183] 请求次要活动应用对第三含义做出响应。

[0184] 18.如条款10所述的方法，其中事件消息指示音频的分类，分类指示音频是：

[0185] 是用户交互的部分的语音；

[0186] 不是用户交互的部分的语音；

[0187] 是用户交互的部分的音频内容；

[0188] 不是用户交互的部分的音频内容；或

[0189] 响应于由音频设备检测到条件而给出的音频通知。

[0190] 19.如条款14所述的方法，其中音频通知包括：

[0191] 不是用户交互的部分的背景音频通知；或

[0192] 是用户交互的部分的前景音频通知。

[0193] 20.如条款10所述的方法，其中第一事件消息指定标识第一有责任的应用的应用标识符。

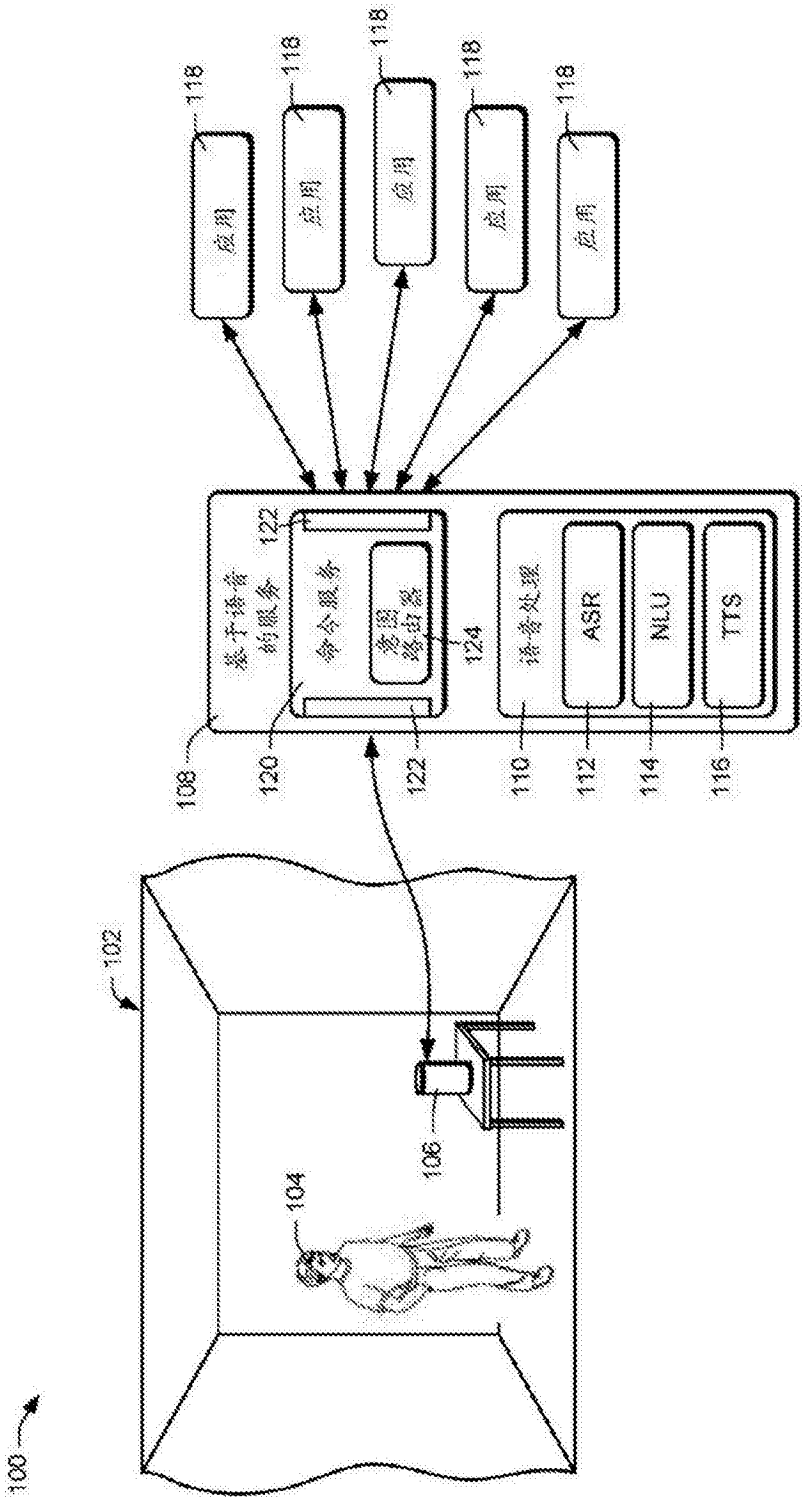

118
118
118
118
118
应用
应用
应用
应用
应用
基于语音的服务
122
命令服务
意图路由器
124
语音处理
ASR
NLU
TTS
108
120
122
110
112
114
116
102
106
104
100

图1

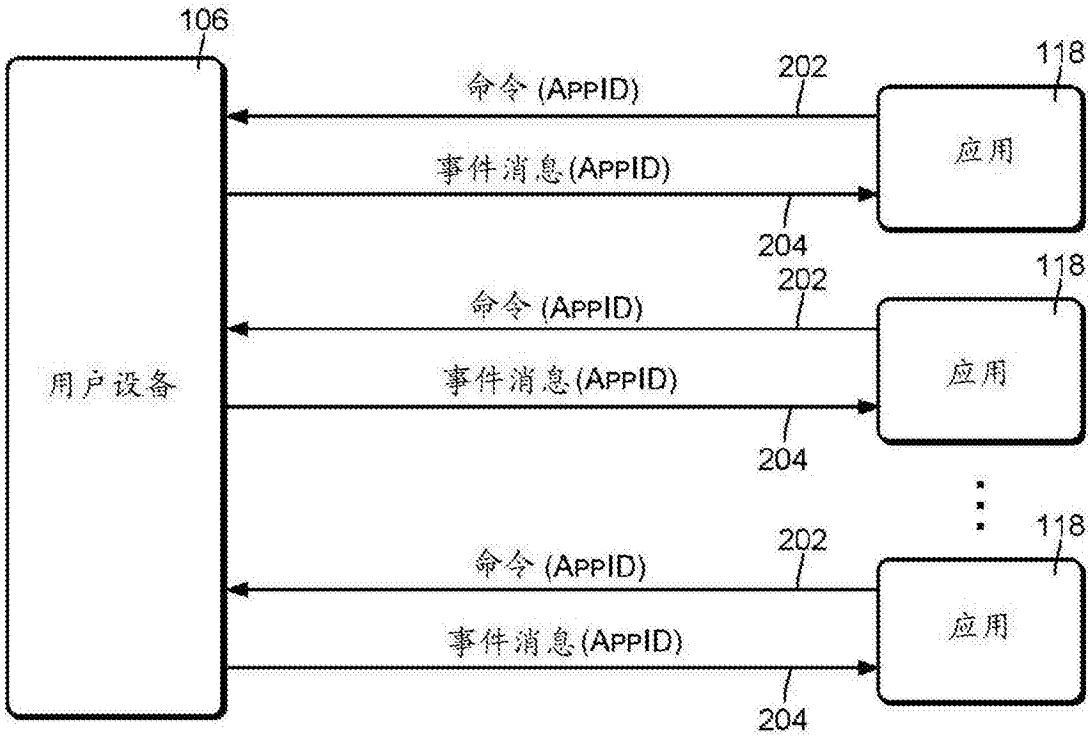

106
用户设备
命令 (APPID)
202
事件消息(APPID)
204
118
应用
命令 (APPID)
202
事件消息(APPID)
204
118
应用
命令 (APPID)
202
事件消息(APPID)
204
118
应用

图2

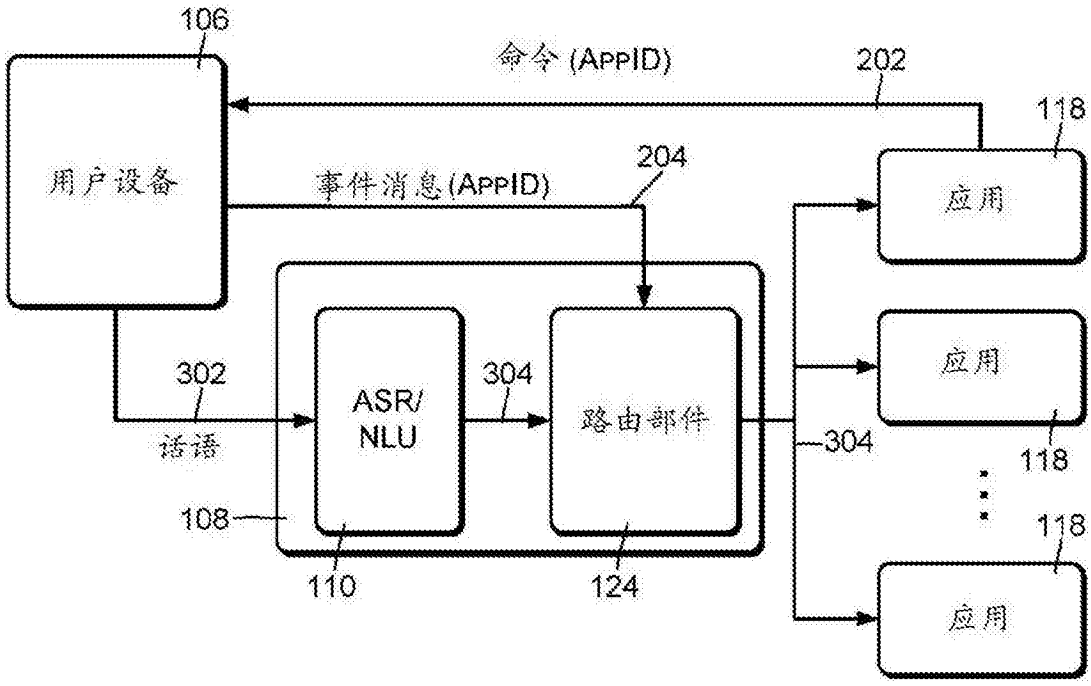

106
用户设备
命令 (APPID)
202
事件消息(APPID)
204
302
话语
ASR/
NLU
304
路由部件
108
110
124
304
118
应用
118
应用
118
应用

图3

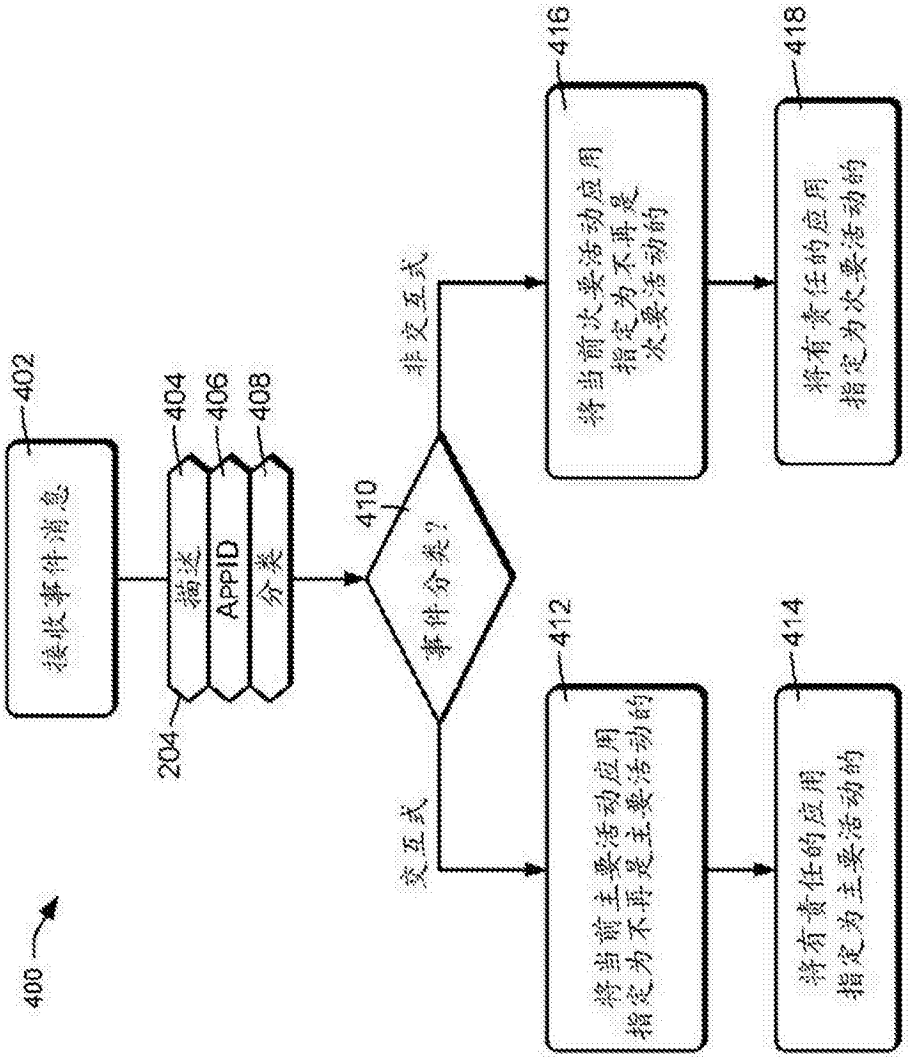

接收事件消息
402
描述
APPID
分类
404
406
408
204
事件分类?
410
非交互式
交互式
将当前次要活动应用指定为不再是次要活动的
416
将有责任的应用指定为次要活动的
418
将当前主要活动应用指定为不再是主要活动的
412
将有责任的应用指定为主要活动的
414
400

图4

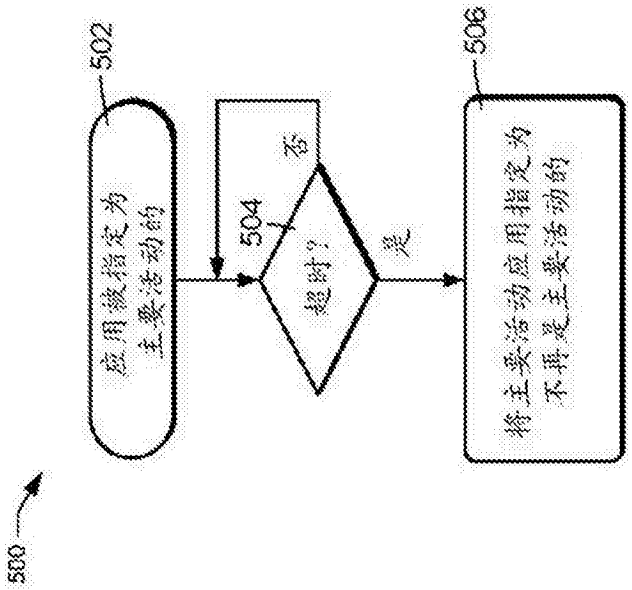

502
应用被指定为主要活动的
504
超时?
否
是
506
将主要活动应用指定为不再是主要活动的
500

图5

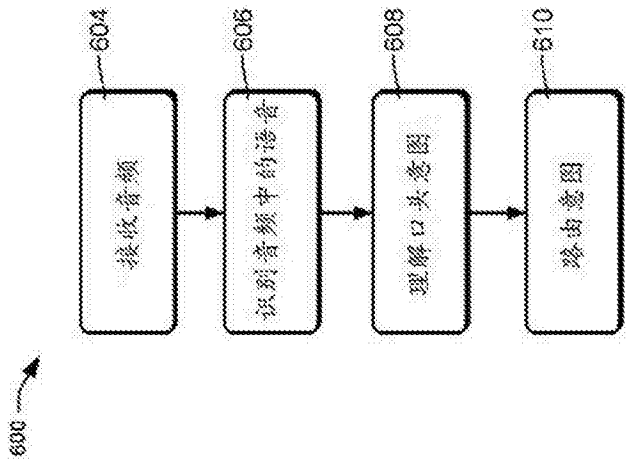

604
接收音频
606
识别音频中的语音
608
理解口头意图
610
响应意图
600

图6

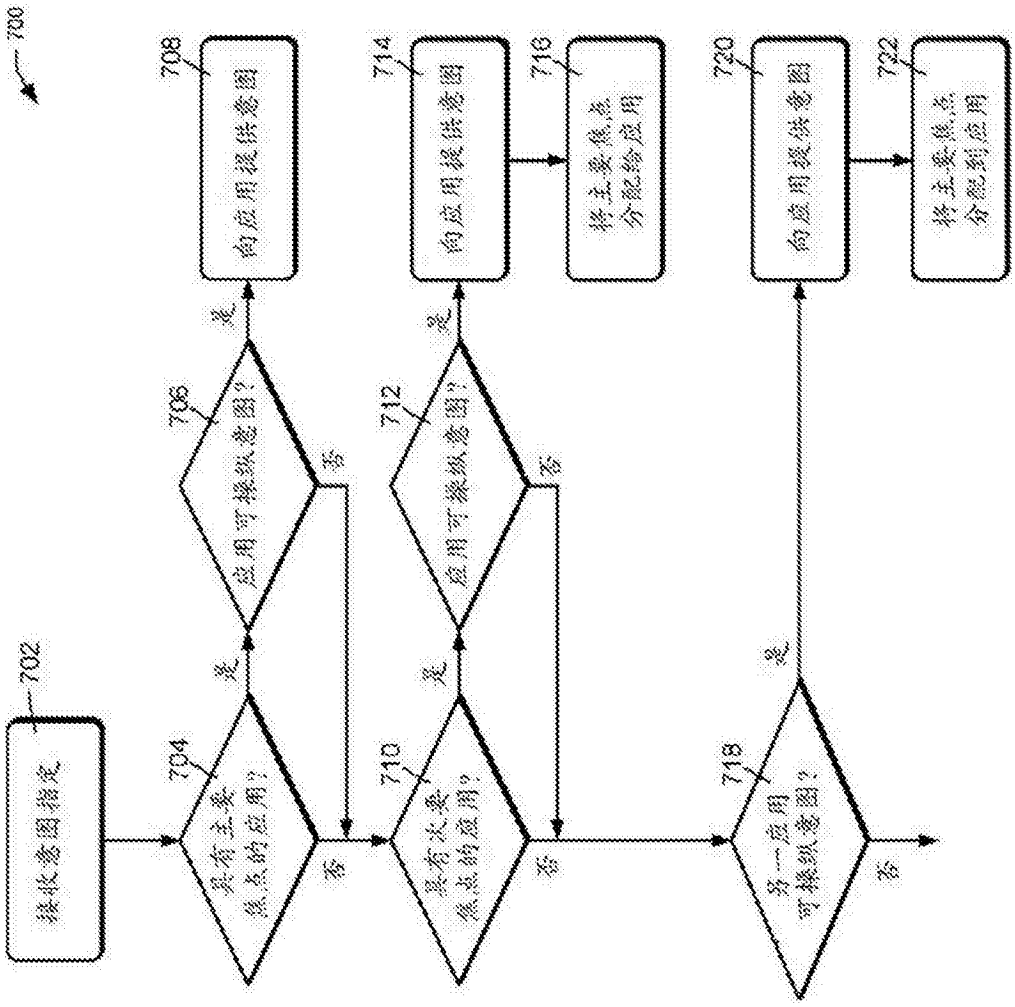

700
702
接收意图指定
704
具有主要焦点的应用?
706
应用可操纵意图?
708
向应用提供意图
710
具有次要焦点的应用?
712
应用可操纵意图?
714
向应用提供意图
716
将主要焦点分配给应用
718
另一应用可操纵意图?
720
向应用提供意图
722
将主要焦点分配到应用
是
否

图7

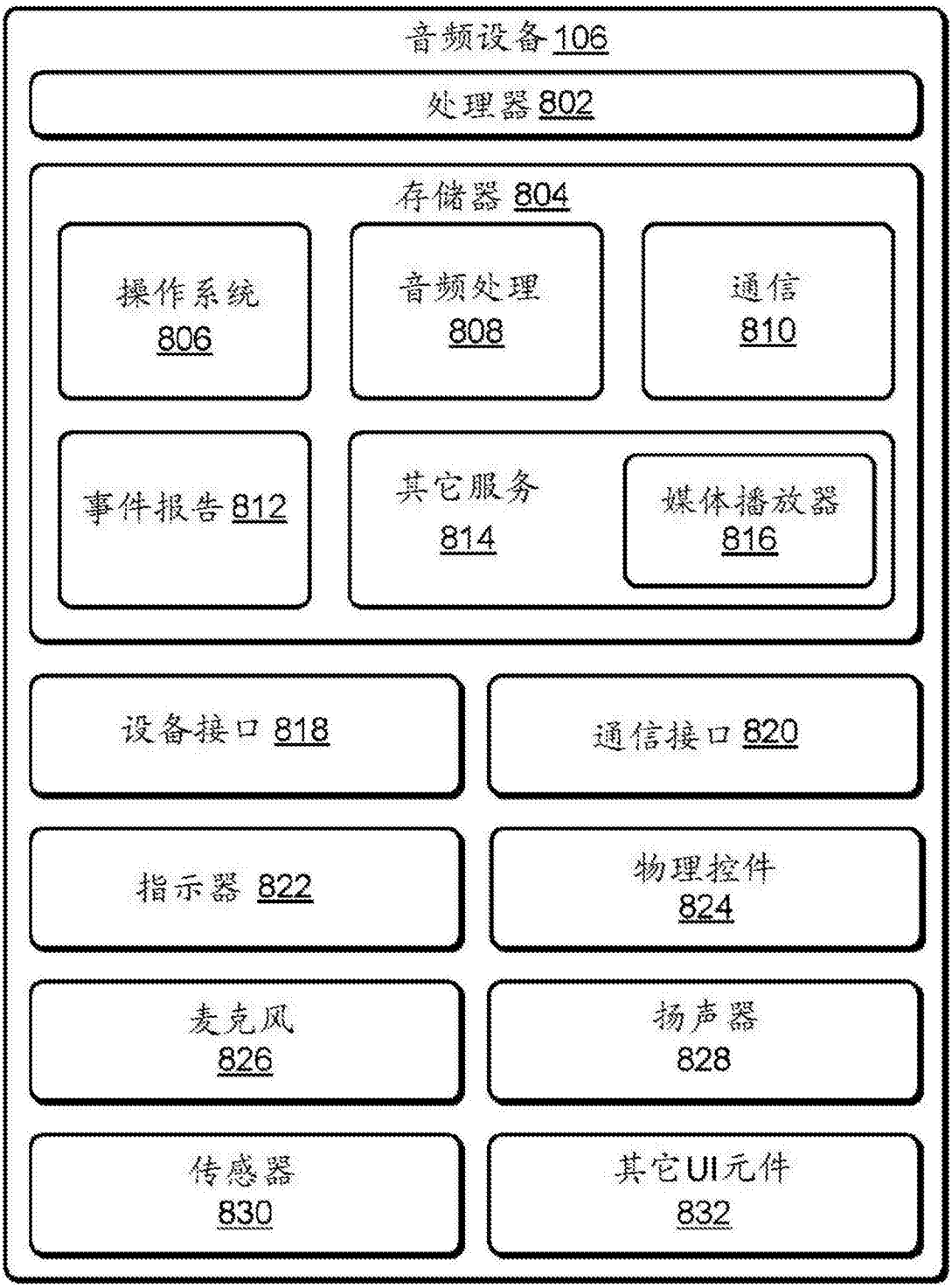

音频设备106
处理器802
存储器804
操作系统
806
音频处理
808
通信
810
事件报告812
其它服务
814
媒体播放器
816
设备接口 818
通信接口 820
指示器 822
物理控件
824
麦克风
826
扬声器
828
传感器
830
其它UI元件
832

图8

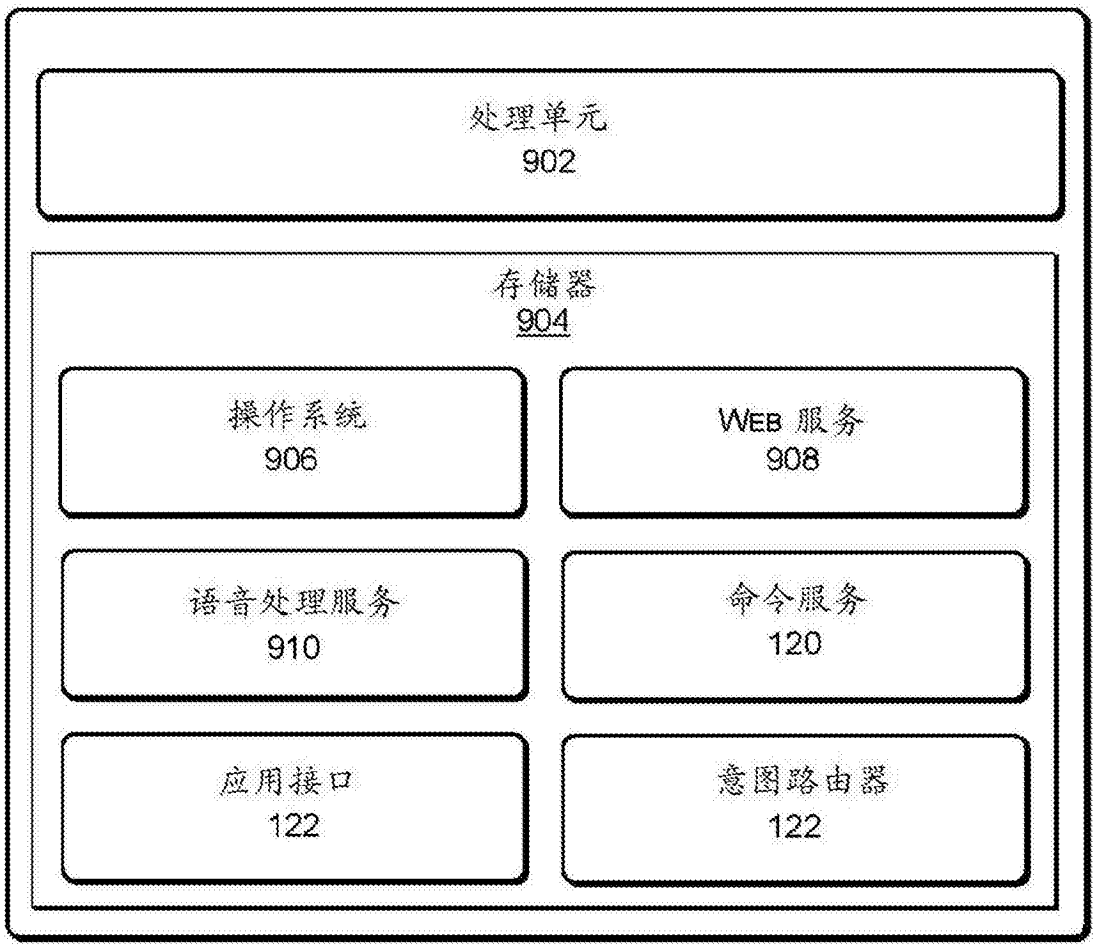

处理单元
902
存储器
904
操作系统
906
WEB 服务
908
语音处理服务
910
命令服务
120
应用接口
122
意图路由器
122

图9