



(12) 发明专利

(10) 授权公告号 CN 107111516 B

(45) 授权公告日 2021.07.30

(21) 申请号 201580072926.4

(22) 申请日 2015.12.29

(65) 同一申请的已公布的文献号
申请公布号 CN 107111516 A

(43) 申请公布日 2017.08.29

(30) 优先权数据
14/593,584 2015.01.09 US

(85) PCT国际申请进入国家阶段日
2017.07.10

(86) PCT国际申请的申请数据
PCT/US2015/067759 2015.12.29

(87) PCT国际申请的公布数据
W02016/111881 EN 2016.07.14

(73) 专利权人 微软技术许可有限责任公司
地址 美国华盛顿州

(72) 发明人 V·S·坎南 A·乌瑟拉克
D·J·黄 R·L·钱伯斯
T·索米欧 A·M·特鲁芬尼斯库
K·沙希德 A·艾玛米

(74) 专利代理机构 上海专利商标事务所有限公司 31100
代理人 陈小刚 陈斌

(51) Int.Cl.
G06F 9/48 (2006.01)
G10L 15/22 (2006.01)

(56) 对比文件
US 8595642 B1, 2013.11.26
US 2014278419 A1, 2014.09.18
US 2014365227 A1, 2014.12.11
US 2013176328 A1, 2013.07.11

审查员 李健壮

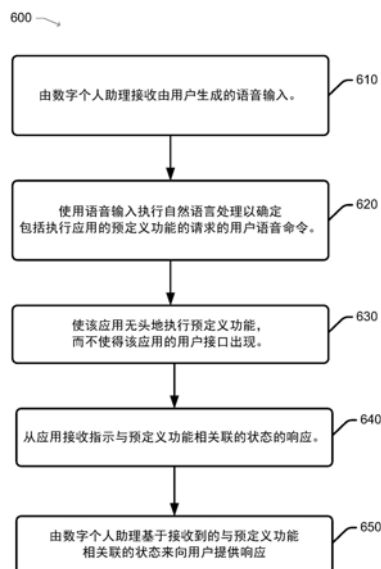
权利要求书3页 说明书18页 附图10页

(54) 发明名称

数字个人助理内的无头任务完成

(57) 摘要

描述了用于在数字个人助理的后台无头地完成应用的任务的技术。例如,一种方法可包括经由话筒接收语音输入。可以使用该语音输入执行自然语言处理以确定用户语音命令。用户语音命令可包括执行应用的任务的请求。可使得该应用将该任务作为后台进程来执行,而不出现该应用的用户接口。数字个人助理的用户接口可以基于接收到的与该任务相关联的状态向用户提供响应,以使得响应来自数字个人助理的用户接口的上下文内,而不出现该应用的用户接口。



1. 一种计算设备,包括:

处理单元;

存储器;以及

一个或多个话筒;

所述计算设备配置有语音控制的数字个人助理,各操作包括:

UI线程经由所述一个或多个话筒接收用户所生成的语音输入;

使用所述语音输入执行语音识别以确定所讲命令,其中所讲命令包括执行第三方应用的任务的请求,并且其中所述任务是使用定义能由所讲命令调用的第三方应用的任务的数据结构来标识的;

在执行语音识别的同时以及在确定所讲命令的完成之前发起所述第三方应用的热身序列,所述热身序列包括与控制线程通信的应用服务线程,所述控制线程与所述UI线程对接;

确定所述第三方应用的任务是否能够被无头地执行;

在确定所述第三方应用的任务能够被无头地执行时,使得所述控制线程开始新的应用服务线程作为后台进程以无头地执行所述任务;

接收来自所述第三方应用的指示与所述任务相关联的状态的响应;以及

通过所述语音控制的数字个人助理的用户接口基于接收到的与所述任务相关联的状态来向所述用户提供响应,以使得所述响应来自所述语音控制的数字个人助理的用户接口的上下文内,而不出现所述第三方应用的用户接口。

2. 如权利要求1所述的计算设备,其特征在于,确定所述第三方应用的任务能够被无头地执行包括确定所述数据结构将所述任务定义为后台任务。

3. 如权利要求1所述的计算设备,其特征在于,确定所述第三方应用的任务能够被无头地执行包括确定所述用户正以免手模式使用所述计算设备。

4. 如权利要求1所述的计算设备,其特征在于,所述热身序列包括分配所述存储器的一部分、预取指令、建立通信会话、从数据库检索信息、启动新执行线程、或唤起中断。

5. 如权利要求1所述的计算设备,其特征在于,定义能由所讲命令调用的第三方应用的任务的所述数据结构包括与所述任务相关联的状态到对所述用户的响应之间的映射。

6. 如权利要求1所述的计算设备,其特征在于,所述操作还包括:

在使得所述第三方应用作为后台进程来执行时,启动定时器;以及

如果所述定时器期满则终止所述后台进程。

7. 如权利要求1所述的计算设备,其特征在于,来自所述第三方应用的响应指示确认状态,并且基于所述确认状态的给所述用户的响应提示所述用户以是或否回答来作出响应。

8. 一种由包括话筒的计算设备实现的方法,所述方法包括:

由语音控制的数字个人助理的UI线程接收由用户生成的数字语音输入,其中所述数字语音输入是经由所述话筒来接收的;

使用所述数字语音输入执行自然语言处理以确定用户语音命令,其中所述用户语音命令包括执行第三方启用语音的应用的预定义功能的请求,并且其中所述预定义功能是使用定义可用的第三方启用语音的应用使用语音输入所支持的功能的数据结构来标识的;

在执行自然语言处理的同时以及在确定所述用户语音命令的完成之前发起所述第三

方启用语音的应用的热身序列,所述热身序列包括与控制线程通信的应用服务线程,所述控制线程与所述UI线程对接;

使得所述控制线程开始新的应用服务线程作为后台进程执行所述预定义功能,而不在所述计算设备的显示器上出现所述第三方启用语音的应用的用户接口;

接收来自所述第三方启用语音的应用的指示与所述预定义功能相关联的状态的响应;以及

通过所述语音控制的数字个人助理的用户接口基于接收到的与所述预定义功能相关联的状态来向所述用户提供响应,以使得所述响应来自所述语音控制的数字个人助理的用户接口的上下文内,而不出现所述第三方启用语音的应用的用户接口。

9.如权利要求8所述的方法,其特征在于,发起所述热身序列包括将投机性功能发送给所述第三方启用语音的应用。

10.如权利要求8所述的方法,其特征在于,与所述预定义功能相关联的状态是从热身、初始、进行中、确认、消歧或最终状态中选择的。

11.如权利要求8所述的方法,其特征在于,所述数据结构是能经由命令模式来扩展的,所述命令模式能用于将来自所述用户的请求关联到一个或多个第三方启用语音的应用的预定义功能。

12.如权利要求8所述的方法,其特征在于,所述数据结构包括:

语音命令到可用第三方启用语音的应用所支持的功能的第一映射;以及

从功能接收到的状态到从所述语音控制的数字个人助理提供给所述用户的响应的第二映射。

13.如权利要求8所述的方法,其特征在于,从所述语音控制的数字个人助理提供给所述用户的响应是从以下类型中选择的:显示文本、文本-到-语音、超文本标记语言HTML、列表模板、可被大声读出的文本、回到该应用的深链接、到网页或网站的链接、基于超文本标记语言的内容、图像、超链接、或者能够经由数字个人助理的用户接口显示给用户的其他合适信息。

14.如权利要求8所述的方法,其特征在于,所述第三方启用语音的应用是远程web服务。

15.如权利要求8所述的方法,其特征在于,使用所述数字语音输入来执行自然语言处理以确定用户语音命令包括解析上下文信息以使得所述用户语音命令是上下文无关的。

16.一种存储计算机可执行指令的计算机可读存储介质,所述计算机可执行指令使得计算设备执行用于在语音控制的数字个人助理的上下文内完成启用语音的应用的任务的操作,所述操作包括:

由所述语音控制的数字个人助理的UI线程接收由用户生成的数字语音输入,其中所述数字语音输入是经由话筒来接收的;

使用所述数字语音输入执行自然语言处理以确定用户语音命令,其中所述用户语音命令包括执行启用语音的应用的任务的请求,并且其中所述任务是使用将用户语音命令映射到启用语音的应用的任务的可扩展数据结构来标识的;

在执行自然语言处理的同时以及在确定所述用户语音命令的完成之前发起所述启用语音的应用的热身序列,所述热身序列包括与控制线程通信的应用服务线程,所述控制线

程与所述UI线程对接；

确定所述启用语音的应用的任务是前台任务还是后台任务；

在确定所述任务是后台任务时，使得所述控制线程开始新的应用服务线程作为后台任务并在所述语音控制的数字个人助理的上下文内执行，而不出现所述启用语音的应用的用户接口；

接收来自所述启用语音的应用的指示与所述任务相关联的状态的响应；以及

基于接收到的与所述任务相关联的状态向所述用户提供响应，在确定所述任务是后台任务时，所述响应是在所述语音控制的数字个人助理的上下文内提供的，而没有出现所述启用语音的应用的用户接口。

17. 如权利要求16所述的计算机可读存储介质，其特征在于，确定所述启用语音的应用的任务是前台任务还是后台任务包括参考所述可扩展数据结构。

数字个人助理内的无头任务完成

[0001] 背景

[0002] 随着计算技术的发展,越来越强大的计算设备已经变得可用。例如,计算设备正越来越多地增加各特征,例如语音识别。对于用户来说,语音可以是一种与计算设备进行通信的有效方式,并且正在开发语音控制应用,诸如语音控制的数字个人助理。

[0003] 数字个人助理可以被用于执行用于个体的任务或服务。例如,数字个人助理可以是在移动设备或台式计算机上运行的软件模块。可由数字个人助理执行的任务和服务的示例可包括检索天气状况和预报、赛事比分、交通指导和状况、本地和/或国内新闻事件、以及股票价格;通过创建新时间表条目来管理用户的时间表,以及向用户提醒即将到来的事件;以及存储和检索提醒事项。

[0004] 然而,数字个人助理可能不能执行用户可能想要执行的每一任务。因此,存在足够机会来改进涉及语音控制的数字个人助理的技术。

[0005] 概述

[0006] 提供本概述以便以简化的形式介绍以下在详细描述中进一步描述的一些概念。本概述并不旨在标识所要求保护主题的关键特征或必要特征,也不旨在用于限制所要求保护主题的范围。

[0007] 描述了用于在数字个人助理的后台无头地完成应用的的技术和工具。例如,一种方法可由包括话筒的计算设备实现。该方法可包括由语音控制的数字个人助理接收由用户生成的数字语音输入。数字语音输入可经由话筒来接收。可以使用该数字语音输入执行自然语言处理以确定用户语音命令。用户语音命令可包括执行第三方启用语音的应用的预定义功能的请求。预定义功能可以使用定义可用第三方启用语音的应用使用语音输入所支持的功能的数据结构来标识。可以使得所述第三方启用语音的应用作为后台进程执行所述预定义功能,而不在所述计算设备的显示器上出现所述第三方启用语音的应用的用户接口;可以接收来自所述第三方启用语音的应用的指示与所述预定义功能相关联的状态的响应。所述语音控制的数字个人助理的用户接口可基于接收到的与所述预定义功能相关联的状态向所述用户提供响应,以使得所述响应来自所述语音控制的数字个人助理的用户接口的上下文内,而不出现所述第三方启用语音的应用的用户接口。

[0008] 作为另一示例,可提供包括处理单元、存储器以及一个或多个话筒的用于执行本文所述的操作的计算设备。例如,一种由该计算设备执行的方法可包括经由该一个或多个话筒接收用户所生成的语音输入。可使用该语音输入执行语音识别以确定所讲命令。所讲命令可包括执行第三方应用的任务的请求。该任务可以使用定义能由所讲命令调用的第三方应用的任务的数据结构来标识。可以确定所述第三方应用的任务是否能够在无头地执行。在确定所述第三方应用的任务能够在无头地执行时,可使得所述第三方应用作为后台进程执行以无头地执行所述任务。可接收来自所述第三方应用的指示与所述任务相关联的状态的响应。所述语音控制的数字个人助理的用户接口可基于接收到的与所述任务相关联的状态向所述用户提供响应,以使得所述响应来自所述语音控制的数字个人助理的用户接口的上下文内,而不出现所述第三方应用的用户接口。

[0009] 作为另一示例,可提供包括处理单元和存储器的用于执行本文所述的操作的计算设备。例如,一种计算设备可执行用于在语音控制的数字个人助理的上下文内完成启用语音的应用的任务的操作。所述操作可包括在语音控制的数字个人助理处接收由用户生成的数字语音输入。数字语音输入可经由话筒来接收。可以使用该数字语音输入执行自然语言处理以确定用户语音命令。用户语音命令可包括执行启用语音的应用的任务的请求。任务可使用将用户语音命令映射到启用语音的应用的任务的可扩展数据结构来标识。可以确定所述启用语音的应用的任务是前台任务还是后台任务。在确定所述任务是后台任务时,可使得所述启用语音的应用将所述任务作为后台任务并在所述语音控制的数字个人助理的上下文内执行,而不出现所述启用语音的应用的用户接口。可接收来自启用语音的应用的响应。该响应可指示与该任务相关联的状态。可基于接收到的与该任务相关联的状态来向用户提供响应。在确定任务是后台任务时,响应可以在语音控制的数字个人助理的上下文内提供,而不出现启用语音的应用的用户接口。

[0010] 如本文所述的,各种其它特征和优点可按照需要被结合到所述技术中。

[0011] 附图简述

[0012] 图1是描绘用于在数字个人助理的后台无头地完成应用的任务的系统的示例的示图。

[0013] 图2是描绘用于在数字个人助理的后台无头地完成应用的任务的示例软件体系结构的示图。

[0014] 图3是用于与数字个人助理对接的应用的示例状态机的示图。

[0015] 图4是可被用来创建用于启用应用与数字个人助理之间的接口的数据结构的数据结构的命令定义的示例。

[0016] 图5是示出用于从数字个人助理内无头地执行应用的任务的多个线程的通信的示例序列图。

[0017] 图6是用于在数字个人助理的后台无头地完成应用的任务的示例方法的流程图。

[0018] 图7是用于确定在用户正向数字个人助理讲话时是否使应用热身的示例方法的流程图。

[0019] 图8是可用于实现所描述的一些实施例的示例计算系统的图示。

[0020] 图9是可以结合本文所述的技术来使用的示例移动设备。

[0021] 图10是可结合本文中描述的技术来使用的示例云支持环境。

[0022] 详细描述

[0023] 概览

[0024] 随着用户对使用数字个人助理变得更加舒适,该用户可能更喜欢在数字个人助理的上下文内执行更多动作。然而,数字个人助理的提供者不能预测或花费时间来开发用户可能想要使用的每一应用。因而,数字个人助理能够调用或启动已由该数字个人助理的提供者以外的实体创建的第三方应用是合乎需要的。

[0025] 在典型的解决方案中,在数字个人助理启动应用时该应用的用户接口出现,并且程序控制从该数字个人助理传递到该应用。一旦该应用的用户接口出现,用户就可验证该请求的状态并且用户可从该应用内执行附加任务。为返回该数字个人助理的用户接口,用户必须在控制可被返回到该数字个人助理之前退出该应用。

[0026] 作为使用移动电话的数字个人助理的一个具体示例,用户可请求使用安装在移动电话上的影片应用将影片添加到该用户的队列中。例如,用户可以向该数字个人助理的用户接口说出“影片应用,将影片X添加到我的队列 (Movie-Application,add Movie-X to my queue)”。在该命令被说出并被该助理识别之后,该助理可以启动影片应用,该影片应用将呈现该影片应用的用户接口。影片可被添加到用户的队列并且该队列可被呈现给用户作为影片已被添加的验证。用户可继续使用影片应用或者用户可以关闭影片应用以返回到数字个人助理的用户接口。

[0027] 在数字个人助理将控制转移到应用时,将该应用及其用户接口加载到存储器可花费能感知到的时间量。该延迟可潜在地影响用户的生产力,诸如通过延迟用户完成后续任务和/或通过中断用户的思路。例如,用户的注意力可被引导到在返回到数字个人助理的用户接口之前关闭该应用。此外,通过将控制转移到该应用,数字个人助理可用的上下文信息可能对该应用而言是不可用的。例如,数字个人助理可能理解用户的配偶的身份和联系信息、用户住宅或办公室的位置、或者用户的日托提供者的位置,但该应用可能不具有对该上下文信息的访问权。

[0028] 在本文描述的技术和解决方案中,数字个人助理可以确定第三方应用的任务是否可以在后台执行,以便用于执行该任务读操作在数字个人助理的上下文内执行并且不出现启用语音的应用的用户接口。因而,用户可体验到给定任务集合在数字个人助理的上下文内执行,与正在做用户任务的应用的上下文形成对比。此外,设备的功率可潜在地降低(并且电池寿命延长),因为在该应用的任务在后台执行时该应用的用户接口没有被加载到存储器。

[0029] 各应用可以向数字个人助理注册以扩展该助理提供的本机能力的列表。各应用可被安装在设备上或通过网络(诸如因特网)作为服务来被调用。模式定义可以使各应用能够注册语音命令,在用户请求该命令/任务时,一请求将被无头地启动。例如,应用可包括可由数字个人助理访问的语音命令定义(VCD)文件,其中VCD文件标识可被无头地启动的任务。该定义可以指定该应用的任务要总是无头地启动,或者该定义可以指定该应用的任务要在特定情形下无头地启动。例如,如果用户正在请求在不具有显示表面的设备(诸如无线健身手环)上执行任务,或者在用户正以免手模式操作时(诸如在用户连接到蓝牙耳机时),应用可选择无头地做一些事。

[0030] 各应用可提供与所请求的任务的进行中、失败、以及成功完成有关的响应,并且与状态相关的输出可由数字个人助理的用户接口来提供。各应用可以将许多不同类型的数据提供回数字个人助理,包括例如显示文本、可被大声读出的文本、回到该应用的深链接、到网页或网站的链接、以及基于超文本标记语言(HTML)的web内容。从应用到助理的数据可如同来自助理的本机功能一样来经由该助理的用户接口呈现。

[0031] 如果用户向应用提供可能具有多个含义或结果的请求,则该应用可向数字个人助理提供选择列表并且该助理的用户接口可被用来消除这些选择之间的歧义。如果用户向应用提供可能是破坏性或重要的请求(诸如在用户请求银行应用执行余额代偿),该助理的确认接口可被使用以在完成该破坏性或重要任务之前确认该请求。

[0032] 随着命令被讲出,应用可被投机地加载或热身。例如,在用户完成来自命令“影片应用,将影片X添加到我的队列”的短语“影片应用 (Movie-Application)”时,可分配存储器

并且可从存储中检索已安装影片应用的各个子例程并将其加载到所分配的存储器中以准备该命令变完整时使用这些子例程。当应用是web服务时,热身可包括例如建立通信会话并从远程服务器处的数据库检索用户专用信息。通过将应用热身,对用户作出响应的时间可潜在地被降低,使得交互更加自然并且使得用户可更快地移至下一任务,从而使得用户更具生产力。

[0033] 使用本文的技术,希望使用影片应用来将影片添加到用户的队列的用户可具有与在使用启动影片应用并将控制传递给该应用的典型解决方案时不同的体验。在这一示例中,影片应用的将影片添加到队列命令可在命令数据结构(诸如VCD文件)中被定义为是无头的。当用户说出来自命令“影片应用,将影片X添加到我的队列”的“影片应用”时,该影片应用可被热身以使得对用户的响应时间可被降低。在该命令变完整时,可使用影片应用将影片添加到用户的队列,但没有出现该影片应用的用户接口。影片可被添加到用户的队列并且数字个人助理可(使用该助理的用户接口)确认影片已被添加。用户可体验到更快的响应时间并可执行更少步骤来完成任务(例如,影片应用不必被关闭)。

[0034] 包括数字个人助理的示例系统

[0035] 图1是描绘用于在数字个人助理120的后台无头地完成启用语音的应用110的任务112的系统100的示例的系统图。启用语音的应用110和数字个人助理120可以是安装在计算设备130上的软件模块。计算设备130可以是例如台式计算机、膝上型设备、移动电话、智能电话、可穿戴设备(诸如手表或无线电子手环)、或平板计算机。计算设备130可包括用于标识可由数字个人助理120启动的应用和应用的任务的命令数据结构140。应用可由数字个人助理120在前台(诸如在该应用被启用时该应用的用户接口出现的情形中)和/或在后台(诸如在该应用被启动时该应用的用户接口不出现的情形中)启动。例如,应用的一些任务可以在前台启动且同一应用的不同任务可以在后台启动。命令数据结构140可以定义该应用和/或该应用的任务应当如何从数字个人助理120启动。

[0036] 计算设备130可包括用于将声音转换成电信号的话筒150。话筒150可以是分别使用电磁感应、电容或压电中的相应改变从气压变化中产生电子信号的动态的电容器或压电话筒。话筒150可以包括放大器、一个或多个模拟或数字滤波器和/或模-数转换器以产生数字声音输入。数字声音输入可以包括用户语音的再现,诸如在用户正命令数字个人助理120完成任务时。计算设备130可包括用于允许用户录入文本输入的触摸屏或键盘(未示出)。

[0037] 数字声音输入和/或文本输入可由数字个人助理120的自然语言处理模块122来处理。例如,自然语言处理模块122可以接收数字声音输入并将用户所说的词语转换成文本。所提取的文本可以被语义分析以确定用户语音命令。通过分析数字声音输入并响应于所说的命令来采取动作,数字个人助理120可以是语音控制的。例如,数字个人助理120可以将所提取的文本与可能用户命令的列表相比较以确定最可能匹配用户意图的命令。该匹配可以是基于统计或概率方法、决策树或其它规则、其它合适的匹配准则或其组合。可能用户命令可以是数字个人助理120的本机命令和/或在命令数据结构140中所定义命令。因而,通过在命令数据结构140中定义命令,可以代表用户由数字个人助理120所执行的任务的范围可被扩展。可能命令可包括执行启用语音的应用110的任务112,这可在命令数据结构140中被定义为无头或后台任务。

[0038] 自然语言处理模块122可以在处理语音时生成文本流,以使得中间文本串可在用

户话语完成之前被分析。因而,如果用户以应用的名称来开始命令,则该应用可在话语中提早标识,并且该应用可在用户说完命令之前被热身。使应用热身可包括从相对较慢的非易失性存储器(诸如硬盘驱动器或闪存)检索该应用的各指令并将这些指令存储在相对较快的易失性存储器(诸如主存储器或高速缓存存储器)中。

[0039] 在数字个人助理120确定命令与应用的任务相关联时,该应用的任务可被执行。如果数字个人助理120确定该应用的任务要作为后台进程来执行(诸如通过分析命令数据结构140中的定义),则该应用可以在后台执行。该应用(诸如启用语音的应用110)可以与数字个人助理120通信。例如,该应用可以顺序通过与任务的完成相关联的状态集,并且该应用的状态可被传递给数字个人助理120。例如,该应用能以“初始”状态开始,在正在执行任务时转移到“进行中”状态,并随后在完成任务时转移到“最终”状态。

[0040] 数字个人助理120可经由用户接口124报告任务的进度。用户接口124可以按各种方式向用户传递信息,诸如通过在计算设备130的显示器上呈现文本、图形或超链接,从计算设备130的扬声器生成音频输出,或者生成其他传感输出,诸如来自连接到计算设备130的离心重量的电机的振动。例如,用户接口124可以使得在任务处于进行中状态时在计算设备130的显示屏上呈现转轮。作为另一示例,用户接口124可以在任务处于最终状态并且任务被成功完成时生成指示任务的成功完成的仿真语音。通过使用数字个人助理120的用户接口124来报告任务的状态,响应可来自用户接口124的上下文内而不出现该应用的用户接口。

[0041] 应当注意,启用语音的应用110可由数字个人助理120的生产者或与该生产者不同的第三方来创建。数字个人助理120与启用语音的应用110的互操作可通过符合应用-到-应用软件合同并通过在命令数据结构140中定义功能性来实现。启用语音的应用110可能作为独立应用或只做为数字个人助理120的组件来操作。作为独立应用,启用语音的应用110可作为前台进程在数字个人助理120外部启动,诸如通过轻击或双击与启用语音的应用110相关联且显示在计算设备130的显示屏上的图标。启用语音的应用110可以在被启动时呈现用户接口并且用户可与该用户接口交互以执行任务。交互可以只使用语音输入,或者也可使用其他输入模式,诸如文本输入或做姿势。由数字个人助理120调用的应用可被安装在计算设备130上或者可以是web服务。

[0042] 数字个人助理120可以调用web服务,诸如在远程服务器计算机160上执行的web服务162。web服务是在网络(诸如网络170)上的网络地址处提供的软件功能。网络170可包括局域网(LAN)、广域网(WAN)、因特网、内联网、有线网络、无线网络、蜂窝网络、其组合、或适用于提供在计算设备130和远程服务器计算机160之间进行通信的信道的任何网络。应当明白,图1所示的网络拓扑结构已经被简化,并且可利用多个网络和联网设备来互连此处所公开的各种计算系统。web服务162可作为数字个人助理120的内核的一部分或主要部分来被调用。例如,web服务162可作为自然语言处理模块122的子例程来被调用。作为补充或替换,web服务162可以是命令数据结构140中定义的应用且可能从数字个人助理120无头地启动。

[0043] 包括数字个人助理的示例软件体系结构

[0044] 图2是描绘用于在数字个人助理120的后台无头地完成应用的任务的示例软件体系结构200的示图。在无头地执行应用的任务时,该任务可以在后台执行并且该应用的用户

接口不作为该任务被执行的结果而出现。相反,数字个人助理120的用户接口可被用来提供输出给用户和/或来自用户的输入,以使得用户在数字个人助理120的上下文内而非该应用的上下文内进行交互。因而,无头地执行的的应用的任务可以在后台执行达该任务的执行历时,并且该应用的用户接口从不出现。计算设备(诸如计算设备130)可以执行根据体系结构200来组织的数字个人助理120、操作系统(OS)内核210以及应用230的软件。

[0045] OS内核210一般提供计算设备130的软件组件和硬件组件之间的接口。OS内核210可包括用于渲染的组件(例如,渲染给显示器的视觉输出、生成用于扬声器的音频输出和其他声音、以及生成用于电机的振动输出)、用于联网的组件、用于进程管理的组件、用于存储器管理的组件、用于位置跟踪的组件、以及用于语音识别和其他输入处理的组件。OS内核210可以管理计算设备130的用户输入功能、输出功能、存储访问功能、网络通信功能、存储器管理功能、进程管理功能、以及其他功能。OS内核210可以向数字个人助理120和应用230提供对这样的功能的访问,诸如通过各种系统调用。

[0046] 用户可以生成用户输入(诸如语音、触觉以及运动)来与数字个人助理120交互。可经由OS内核210(它可包括响应于用户输入来创建消息的功能性)使得数字个人助理120知晓用户输入。该消息可由数字个人助理120或其他软件使用。用户输入可包括触觉输入(诸如触摸屏输入)、按钮按压或按键按压。OS内核210可包括用于识别来自触觉输入的对触摸屏的轻击、手指姿势等、按钮输入、或按键按压输入的功能性。OS内核210可以接收来自话筒150的输入并可包括用于识别来自语音输入的所讲命令和/或词语的功能性。OS内核210可以接收来自加速度计的输入并可包括用于识别取向或运动(诸如摇动)的功能性。

[0047] 数字个人助理120的用户接口(UI)输入处理引擎222可等待来自OS内核210的用户输入事件消息。UI事件消息可指示来自语音输入的所识别的词语、平移姿势、轻拂姿势、拖动姿势、或设备的触摸屏上的其他姿势、触摸屏上的轻击、键击输入、摇动姿势、或其他UI事件(例如,方向按钮或跟踪球输入)。UI输入处理引擎222可以将来自OS内核210的UI事件消息转换成发送给数字个人助理120的控制逻辑224的信息。例如,UI输入处理引擎222可包括自然语言处理能力并可指示特定应用名称已被讲出或键入或者用户已给出语音命令。或者,自然语言处理能力可被包括在控制逻辑224中。

[0048] 控制逻辑224可以接收来自数字个人助理120的各模块的信息,诸如UI输入处理引擎222、个性化信息存储226、以及命令数据结构140,并且控制逻辑224可以基于接收到的信息来作出决策并执行操作。例如,控制逻辑224可以确定数字个人助理120是否应当代表用户执行任务,诸如通过解析所讲文本流来确定是否已给出语音命令。

[0049] 控制逻辑224可在按命令来行动之前等待整个用户命令被讲出,或者控制逻辑224可以在命令仍然正被讲出且在它被讲完之前就开始按该命令来行动。例如,控制逻辑224可以分析所讲命令的中间串并尝试将这些串与命令数据结构140中定义的一个或多个应用相匹配。在应用将被调用的概率超出阈值时,该应用可被热身以使得该应用可更及时地对用户作出响应。可在多个应用和/或功能预期到被调用时投机性地被热身,并且如果确定应用将不被调用则可中止该应用。例如,在用户以特定应用的名称来开始所讲命令时,有很高概率该特定应用将被调用,并且所以该应用可被热身。作为另一示例,一些部分命令串可被限于命令数据结构140中定义的一小组应用,并且当存在与该部分命令串的匹配时,该组应用可被并行地热身。具体而言,命令数据结构140可只具有带词语“take”的命令的两个应用,

诸如具有命令“take a picture (拍摄图片)”的相机应用和具有命令“take a memo (记录备忘)”的备忘应用。控制逻辑224可在词语“take”被识别时开始热身相机应用和备忘应用两者,并且随后在完整命令“take a picture”被识别时可中止备忘应用。将应用热身可包括分配存储器、预取指令、建立通信会话、从数据库检索信息、启动新执行线程、唤起中断、或其他合适的因应用而异的操作。OS内核210的服务可在热身期间被调用,诸如例如进程管理服务、存储器管理服务、以及网络服务。

[0050] 所讲文本可包括上下文信息且控制逻辑224可以解析该上下文信息以使得用户语音命令是上下文无关的。上下文信息可包括计算设备130的当前位置、当前时间、取向以及个性化信息存储226中存储的个人信息。个人信息可包括:用户关系,诸如用户的、配偶的或孩子的名字;因用户而异的位置,诸如住宅、工作、学校、日托、或医生地址;来自用户联系人列表或日历的信息,;用户最喜欢的颜色、餐馆、或交通方法;重要生日、周年纪念、或其他日期;以及其他因用户而异的信息。用户可以给出具有上下文信息的命令且控制逻辑224可以将该命令转换成上下文无关的命令。例如,用户可以给出命令“Bus-app,tell me the busses home within the next hour (公交应用,告诉我下一小时内回家的公交)”。在这一示例中,该命令中的上下文信息是当前日期和时间、当前位置以及用户住宅的位置。

[0051] 控制逻辑224可以从维护实时时钟或具有对实时时钟的访问权的OS内核210获得当前时间。控制逻辑224可从OS内核210得到计算设备130的当前位置数据,而OS内核210可从计算设备130的本地组件获得当前位置数据。例如,位置数据可基于来自全球定位系统(GPS)的数据、通过蜂窝网络的蜂窝塔之间的三角测量、通过参考附近Wi-Fi路由器的物理位置、或者通过另一机制来确定。控制逻辑224可以从个性化信息存储226得到用户住宅的位置。个性化信息存储226可被存储在计算设备130的辅助或其他非易失性存储中。因而,控制逻辑224可以经由可访问存储资源(例如,个性化信息存储226)的OS内核210接收个性化信息。在上下文信息可被解析时,该命令可被转换成上下文无关的命令。例如,如果是星期五下午6点,用户在Main Street 444,并且用户的住宅是Pleasant Drive 123,则上下文无关命令可以是“Bus-app,tell me the busses arriving near 444Main Street and passing near 123Pleasant Drive between 6:00and 7:00p.m.on Fridays (公交应用,告诉我星期五下午6点和7点之间到达Main Street 444附近且经过Pleasant Drive 123附近的公交)”。

[0052] 该用户命令可由控制逻辑224(诸如在该命令是数字个人助理120的本机命令时)、安装在计算设备130上的应用230(诸如在该命令与应用230相关联时)、或者web服务162(诸如在该命令与web服务162相关联时)执行。命令数据结构140可以指定哪些命令与哪些应用相关联以及该命令可以在前台还是后台执行。例如,命令数据结构140可以将用户语音命令映射到由可用第三方启用语音的应用所支持的功能。

[0053] 在控制逻辑224确定用户命令与应用230的预定义的功能232相关联时,控制逻辑224可以使得应用230的预定义的功能232被执行。如果控制逻辑224确定应用230的预定义的功能232要作为后台进程来执行,则预定义的功能232可以在后台执行。例如,控制逻辑224可以通过唤起中断、写到共享存储器、写到消息队列、传递消息、或启动新执行线程(诸如经由OS内核210的进程管理组件)来向预定义的功能232发送请求240。应用230可以执行预定义的功能232并通过唤起中断、写到共享存储器、写到消息队列或传递消息来向控制逻辑

辑224返回响应242。该响应可包括应用230的状态和/或对用户命令作出响应的其他信息。

[0054] 在控制逻辑224确定该命令与web服务162相关联时,控制逻辑224可以使web服务162被调用。例如,请求260可通过OS内核210的联网组件被发送给web服务162。联网组件可以格式化该请求并通过网络170将该请求转发(诸如通过根据网络170的协议将该请求封装在网络分组中)到web服务162以执行该用户命令。请求260可包括多个步骤,诸如打开控制逻辑224与web服务162之间的通信信道(例如,套接字)以及发送与用户命令相关的信息。web服务162可以用能通过网络170传送并由联网组件作为回复262来转发到控制逻辑224的响应来对请求260作出响应。来自web服务162的响应可包括web服务162的状态以及对用户命令作出响应的其他信息。

[0055] 控制逻辑224可以在UI输出渲染引擎228和OS内核210的渲染组件的辅助下)基于来自各应用的响应生成要呈现给用户的输出。例如,命令数据结构140可以将各功能接收到的状态映射到来自语音控制的数字个人助理120的提供给用户的响应。一般而言,控制逻辑224可以将高级输出命令提供给UI输出渲染引擎228,UI输出渲染引擎228可产生送给OS内核210的渲染组件的较低级输出原语以用于显示器上的视觉输出、通过扬声器或耳机的音频和/或语音输出、以及来自电机的振动输出。例如,控制逻辑224可以将具有文本串的文本-到-语音命令发送给可生成模拟所讲语音的数字音频数据的UI输出渲染引擎228。

[0056] 控制逻辑224可以基于应用的状态来确定什么信息要提供给用户。各状态可对应于用户命令的开始、处理、确认、消歧、或完成。命令数据结构140可将应用的状态映射到要提供给用户的不同响应。可被提供的信息的类型包括例如显示文本、仿真语音、回到应用的深链接、到网页或网站的链接、以及基于超文本标记语言(HTML)的web内容。

[0057] 示例应用状态

[0058] 图3是用于以无头方式与数字个人助理120对接的应用的示例状态机300的示图。该应用可以在热身状态310或初始状态320中开始。在数字个人助理120使得该应用热身时可进入热身状态310,诸如在知晓应用名称但所讲命令尚未讲完时。该应用将保持在热身状态310,直至热身操作完成。在热身操作完成时,该应用可转移到初始状态320。

[0059] 在热身状态310完成之后或在数字个人助理120提供了对该应用的用户命令之后,可进入初始状态320。在初始状态320期间,用户命令被该应用处理。如果命令是没有歧义的但将花费比预定时间量更多的时间来完成(诸如5秒),则在命令正被执行时,状态可被转移到进行中状态330。如果命令是没有歧义的且可导致重要的或破坏性的操作被执行,则状态可转移到确认状态340。如果命令有点歧义,但歧义可通过在几个选项之间作出选择来澄清,则状态可转移到消歧状态350。如果命令是有歧义的且不能用几个选项来消歧,则状态可转移到最终状态360,诸如失败状态或重定向状态。如果命令不能被执行,则状态可转移到最终状态360,诸如失败状态。如果命令能在少于预定时间量的时间内完成且不需要请求来自用户的确认,则状态可转移到最终状态360,诸如成功状态。应当注意,最终状态360可以是具有多个条件的单个状态(诸如其中这些条件是成功、失败、重定向以及超时)或一群最终状态(诸如其中各状态是成功、失败、重定向以及超时)。

[0060] 进行中状态330可指示用户命令的操作正被执行或正被尝试。应用可在进行中状态330期间通过向数字个人助理120发送文本-到-语音(TTS)串或图形用户界面(GUI)串以使得信息可使用数字个人助理120的用户接口而被呈现给用户,来向用户提供信息。作为补

充或替换,默认信息(诸如转轮、沙漏、和/或取消按钮)可在进行中状态330期间使用数字个人助理120的用户接口被呈现给用户。

[0061] 在进行中状态330期间,应用可以监视各操作的进度并确定该应用是可停留在进行中状态330中还是转移到最终状态360。在一个实施例中,该应用可以启动定时器(诸如达5秒),并且如果该应用在定时器期满之前没有做出足够进展,则状态可转移到最终状态360,诸如超时状态。如果该应用作出了足够进展,则定时器可被重启并且可在下一定时器期满时再次检查进度。应用可具有停留在进行中状态330中的最大时限,并且如果超过该最大时限,状态可转移到最终状态360,诸如超时状态。与用户命令相关联的操作可以完成(成功或不成功地)并且状态可转移到适当的最终状态360。当应用处于进行中状态330中时,用户可以通过向数字个人助理120的用户接口给出命令来终止该应用。例如,用户可以按下或点击显示器上的“取消”或“后退”按钮或者说出“取消”。取消该命令可以使得数字个人助理120停止该应用,并且显示数字个人助理120的主屏幕或者退出。

[0062] 确认状态340可以指示该应用在完成任务之前正等待来自用户的确认。在数字个人助理120检测到应用处于确认状态340时,可使用数字个人助理120的用户接口向用户呈现用于是/否响应的提示。该应用可以向数字个人助理120提供作为具有是或否回答的问题的TTS串。数字个人助理120可以说出该应用所提供的TTS串并且可监听“是/否”回答。如果用户响应没有决定是或否回答,则数字个人助理120可继续询问用户该问题直至预定义次数(诸如三次)。如果所有尝试已被耗尽,则数字个人助理120可以说出默认短语,诸如“I'm sorry, I don't understand. Tap below to choose an answer(对不起,我不明白。轻击下方来选择回答)”并且数字个人助理120可以停止监听。如果用户轻击是或否,则数字个人助理120可以将该用户的选择发送给应用。如果用户轻击话筒图标,则数字个人助理120可再次尝试识别所讲回答(诸如通过复位对口头回答的尝试的次数进行计数的计数器)。数字个人助理120可以循环,直至存在匹配或者用户取消或点击显示屏上的后退按钮。如果应用接收到来自数字个人助理120的肯定响应,则该应用可以尝试完成任务。如果任务成功完成,则状态可转移到具有成功条件的最终状态360。如果任务未能成功完成或者该应用被取消,则状态可转移到具有失败条件的最终状态360。如果任务将花费高于预定时间量的时间才能完成,则状态可转移到进行中状态330,同时任务正被执行。

[0063] 消歧状态350可以指示应用在完成任务之前正等待用户在有限数目(诸如10个或更少)的选项之间进行澄清。应用可以向数字个人助理120提供TTS串、GUI串和/或用户从中选择的项列表。项列表可作为具有针对每一项提供给用户的一个或多个信息段(诸如标题、描述、和/或图标)的模板来提供。数字个人助理120可以使用该应用所提供的信息来向用户呈现项列表。数字个人助理120可以提示并监听来自用户的选择。用户可以使用灵活或非灵活选择来从列表中进行选择。非灵活选择意指用户只能以一种方式从列表选择,而灵活选择意指用户能以多种不同的方式从列表中进行选择。例如,用户可以基于各项被列出的数字次序来从列表中进行选择,诸如通过说出“第一”或“第二”以分别选择第一项或第二项。作为另一示例,用户可以基于各项之间的空间关系从列表进行选择,诸如“底部那个”、“顶部那个”、“右边那个”或者“从底部开始第二个”。作为另一示例,用户可以通过说出项的标题来从列表进行选择。

[0064] 作为消歧的具体示例,用户可以向数字个人助理120说出“影片应用,将影片X添加

到我的队列 (Movie-Application, add Movie-X to my queue)”。然而,可存在影片X (Movie-X) 的三个版本,诸如原始版本以及两个续集:影片-X I、影片-X II、以及影片-X III。响应于该所讲命令,数字个人助理120可以在后台使用将影片X添加到队列的命令来启动影片应用。影片应用可搜索影片X并确定存在三个版本。因而,影片应用可转移到消歧状态350并且将这三个备选选择发送给数字个人助理120。数字个人助理120可通过其用户接口向用户呈现这三个选择并且可以从列表选择一个。在用户作出正确选择时,数字个人助理120可以将该响应发送给影片应用并且正确的影片可被添加到队列。

[0065] 如果用户响应不能被决定到列表上的项,则数字个人助理120可继续询问用户该问题直至预定义次数。如果所有尝试已被耗尽,则数字个人助理120可以说出默认短语,诸如“I'm sorry, I don't understand. Tap below to choose an answer (对不起,我不明白。轻击下方来选择回答)”并且数字个人助理120可以停止监听。如果用户轻击所显示的列表上的各项之一,则数字个人助理120可以将该用户的选择发送给应用。如果用户轻击话筒图标,则数字个人助理120可再次尝试识别所讲回答(诸如通过复位对口头回答的尝试的次数进行计数的计数器)。数字个人助理120可以循环,直至存在匹配或者用户取消或点击显示屏上的后退按钮。如果应用接收到来自数字个人助理120的有效响应,则该应用可以尝试完成任务。如果任务在采取动作之前需要用户确认,则状态可转移到确认状态340。如果任务成功完成,则状态可转移到具有成功条件的最终状态360。如果任务未能成功完成或者该应用被取消,则状态可转移到具有失败条件的最终状态360。如果任务将花费高于预定时间量的时间才能完成,则状态可转移到进行中状态330,同时任务正被执行。

[0066] 应当理解,示例状态机300可以用附加或另选状态来扩展以启用用户与应用之间的各种多轮对话。消歧(经由消歧状态350)和确认(经由确认状态340)是多轮对话的具体示例。一般而言,在多轮对话中,无头应用可向用户请求附加信息而不使其用户接口出现。相反,可以通过代表该应用的数字个人助理120从用户获得信息。因而,数字个人助理120可充当用户与应用之间的管道。

[0067] 最终状态360可以指示该应用已经成功完成任务、未能完成任务、已超时、或正指出该应用应当在前台启动(重定向)。如上所述,最终状态360可以是具有多个条件的单个状态(例如,成功、失败、重定向以及超时)或一群最终状态(例如,成功、失败、重定向以及超时)应用可以向数字个人助理120提供TTS串、GUI串、项列表(经由模板提供)和/或启动参数。数字个人助理120可以使用数字个人助理120的用户接口向用户呈现由该应用提供的信息。作为补充或替换,数字个人助理120可以呈现与不同条件相关联地预定义或录制响应。例如,如果发生超时或任务失败,则数字个人助理120可以说出“Sorry! I couldn't get that done for you. Can you please try again later? (对不起!我不能为你完成。能请您稍候重试吗?)”。作为另一示例,如果应用正请求重定向,则数字个人助理120可以说出“Sorry. <appName> is not responding. Launching <appName> (对不起.<应用名>没有响应。正启动<应用名>)”并且数字个人助理120可尝试用初始语音命令和启动参数(如果该应用提供了启动参数的话)在前台启动该应用。作为另一示例,如果应用成功完成任务,则数字个人助理120可以说出“I've done that for you (我已为你完成)”。

[0068] 示例命令定义

[0069] 图4是符合一模式的命令定义400的示例,它可被用来创建用于允许第三方应用与

数字个人助理120之间的对接的数据结构(诸如命令数据结构140)。命令定义400可以用各种语言来编写,诸如由模式定义的可扩展标记语言(XML)或XML子集。例如,该模式可以定义命令定义的结构,诸如法定元素、元素的分层结构、每一元素的法定和可任选属性、以及其他合适的准则。命令定义400可由数字个人助理120用来辅助将用户话语解析成不同分量,诸如应用、命令或任务、以及数据项或时隙,其中数据项是可任选的。例如,命令“MovieAppService,add MovieX to my queue(影片应用服务,将影片X添加到我的队列)”可被解析成应用(“MovieAppService(影片应用服务)”、命令(“Add(添加)”以及数据项(“MovieX(影片X)”)。命令定义400可包括用于定义应用名、该应用的任务或命令、用于自然语言处理的备选短语、以及与不同应用状态相关联的响应的各元素。

[0070] 一个或多个应用可被定义在命令定义400中。应用可以是安装在该计算设备上或是web服务的第三方或其他应用。与该应用相关的信息可以用定义该应用的元素来界定。例如,应用名可以由<AppName>元素来定义,并且<AppName>元素之间的元素可以与开头<AppName>元素相关联。在命令定义400中,应用名是“MovieAppService(影片应用服务)”,且<AppName>元素之后的元素与“MovieAppService”应用相关联。

[0071] 应用名之后的命令是该应用的命令。命令可以用<Command>(命令)元素来标识。命令元素的属性可包括该命令的名称(例如,“Name”)和该命令的激活类型(例如,“ActivationType”)。例如,对于要在前台启动的命令,激活类型可以是“前台”,且对于要在后台启动的命令,激活类型可以是“后台”。“ActivationType”属性可以是可任选的,其中默认激活类型是前台。

[0072] <ListenFor>(监听)元素可被嵌套在<Command>元素内并且可被用来定义可讲出该命令的一种或多种方式。在执行自然语言处理时,可任选或载体词语可作为提示被提供给数字个人助理120。载体词语可被标识在方括号:[]内。数据项可被标识在花括号:{}内。在命令定义400中,一般存在调用“Add(添加)”命令的两种备选方式,如两个<ListenFor>元素所定义的。例如,说出“add MovieX to my queue(将影片X添加到我的队列)”或“add MovieX to my MovieAppService queue(将影片X添加到我的影片应用服务队列)”可被用来使数字个人助理120在后台启动MovieAppService(影片应用服务)的“Add”命令。应当注意,预定义短语可以用一组括号内的关键字“builtIn:(内置)”来标识:{builtIn:<phrase identifier>}(内置:<短语标识符>)。

[0073] <Feedback>(反馈)元素可被嵌套在<Command>元素内并且可被用来定义在数字个人助理120已成功识别来自用户的所讲命令时要向用户说出的短语。作为补充或替换,<Feedback>元素可以定义在所讲命令正被数字个人助理120解析时要显示给用户的文本串。

[0074] <Response>(响应)元素可被嵌套在<Command>元素内且可被用来定义由数字个人助理120提供给用户的一个或多个响应。每一响应与由“State(状态)”属性所定义的应用状态相关联。状态可以用于最终状态(诸如成功和失败)或用于中间状态(诸如进行中)。可定义多种类型的响应,诸如例如用于将文本显示在屏幕上的<DisplayString>(显示串)、用于将被说给用户的文本的<TTSSstring>、用于到网站的深链接的<AppDeepLink>、以及用于到网站的较不深的链接的<WebLink>。由<Response>元素定义的响应可以用由该应用提供的附加响应信息来扩充。

[0075] 示例序列图

[0076] 图5是解说用于从数字个人助理120内无头地执行第三方应用的功能的多个执行线程(510、520以及530)的通信的示例序列图500。UI线程510和控制线程520可以是数字个人助理120的多线程实施例的并行线程。UI线程510可主要负责捕捉来自数字个人助理120的用户接口的输入并将输出显示到该用户接口。例如,语音输入、触觉输入、和/或文本输入可由UI线程510捕捉。在一个实施例中,UI线程510可以对该输入执行自然语言处理并可以将用户所讲的命令与命令数据结构140中的命令相匹配。在所讲命令被确定与命令数据结构140中的命令相匹配时,该命令可被传递到控制线程520以供进一步处理。在一替换实施例中,UI线程510可以捕捉语音到文本输入,并且各单独词语可被传递给控制线程520,控制线程520可对该输入执行自然语言处理并可将用户所讲的命令与命令数据结构140中的命令相匹配。

[0077] 控制线程520可主要负责与应用进行通信并跟踪应用的进度并且与UI线程510对接。例如,UI线程510可以向控制线程520通知用户已向数字个人助理120的用户接口讲话。词语或命令可由控制线程520接收并且控制线程520可以向UI线程510通知用户命令何时已被控制线程520识别。UI线程510可以经由数字个人助理120的用户接口向用户指示对该命令作出的进度。UI线程510或控制线程520可以通过检索来自命令数据结构140的命令的属性来确定该命令要被无头地启动。在命令要被无头地启动时,控制线程520可以开始新线程或与现有线程(诸如AppService(应用服务)线程530)通信。为降低对用户的响应时间,AppService线程530是现有线程可以是合乎需要的,而非使控制线程520开始新线程。例如,AppService线程530可以在使应用热身时或者在计算设备130的引导期间被启动。

[0078] AppService线程530可以在计算设备130上执行或可以在远程服务器(诸如远程服务器计算机160)上执行。AppService线程530可主要负责完成由用户命令指定的功能。AppService线程530可维持状态机(诸如状态机300)来跟踪该功能的执行进度,并且可以向控制线程520提供与状态有关的更新。通过向控制线程520提供状态更新,AppService线程530可以是无头的,其中给用户的输出由数字个人助理120提供而非AppService线程530的用户接口。

[0079] 控制线程520可以通过接收来自应用的状态更新并检查该应用是否有进展来跟踪该应用(例如,AppService线程530)的进度。例如,控制线程520可以每次它与AppService线程530通信(发送信息给AppService线程530或从AppService线程530接收信息)时就启动预定义历时(诸如5秒)的定时器。如果定时器在AppService线程530作出响应之前期满,则控制线程520可以向UI线程510指示该应用未能作出响应并且UI线程510可以经由数字个人助理120的用户接口向用户呈现失败消息。在定时器期满之后,AppService线程530可被控制线程520终止或忽略。或者,如果AppService线程530在定时器期满之前作出响应,则在预期来自该应用的另一响应的情况下(诸如在应用以进行中状态来作出响应时),定时器可被复位,或者定时器可被取消(诸如在该应用已完功能(最终状态)时或在请求用户响应时(确认或消歧状态))。

[0080] 在控制线程520从AppService线程530接收到确认或消歧状态时,控制线程520可以向UI线程510指示向用户请求确认或消歧。UI线程510可以经由数字个人助理120的用户接口将该确认或消歧选择呈现给用户。在用户作出响应或未能作出响应时,UI线程510可以向控制线程520提供该用户响应或者确定没有响应。控制线程520可将用户响应传递给

AppService线程530以使得AppService线程530可以执行功能。如果用户未能作出响应,则控制线程520可以终止AppService线程530。

[0081] UI线程510可以经由数字个人助理120的用户接口显示各种类型的输出。例如,UI线程510可以生成音频输出,诸如来自文本的数字仿真语音输出。数字仿真语音可被发送给可将该数字仿真语音转换成模拟信号(诸如使用数模转换器)的音频处理芯片,该模拟信号可经由扬声器或耳机被输出为声音。作为另一示例,UI线程510可以提供视觉输出,诸如用于由用户在计算设备130的显示屏上查看的图像、动画、文本输出、以及超链接。如果超链接被轻击或点击,UI线程510可启动浏览器应用以查看与所选超链接相对应的网站。作为另一示例,UI线程510可以生成触觉输出,诸如通过向可使得计算设备130振动的电机发送振动信号。

[0082] 用于无头任务完成的示例方法

[0083] 图6是用于在数字个人助理120的后台无头地完成应用的任务的示例方法600的流程图。在610,可由数字个人助理120接收由用户生成的语音输入。语音输入可在计算设备130本地捕捉或从计算设备130远程地捕捉。作为一个示例,由用户生成的语音输入可由计算设备130的话筒150在本地捕捉并由模数转换器来数字化。作为另一示例,由用户生成的语音输入可由无线地连接到计算设备130的话筒(诸如由蓝牙伴随设备)远程地捕捉。数字个人助理120可通过在数字个人助理120的用户接口处录入的语音和/或文本来控制。

[0084] 在620,可以执行对语音输入的自然语言处理以确定用户语音命令。用户语音命令可包括执行应用(诸如第三方启用语音的应用)的预定义功能的请求。预定义功能可以使用数据结构来标识,该数据结构定义由数字个人助理120支持的应用和应用的功能。例如,兼容应用可被标识在命令定义文件中,诸如命令定义400。通过使用可扩展命令定义文件来定义可由数字个人助理120无头地执行的第三方应用的功能,数字个人助理120可以使用户能够使用数字个人助理120的用户接口执行更多任务。

[0085] 在630,数字个人助理120可以使该应用无头地执行预定义功能,而不使得该应用的用户接口出现在计算设备130的显示器上。数字个人助理120可以确定无头地执行该应用,因为该应用在命令数据结构140中被定义为无头的或者因为用户正以免手模式使用计算设备且在前台执行应用可能潜在地使用户分心。例如,数字个人助理120可以调用web服务来执行该应用的预定义功能。作为另一示例,数字个人助理120可以在确定了用户命令之后在计算设备130上启动新线程来执行该应用的预定义功能。作为又一示例,数字个人助理120可以与现有线程(诸如在该应用的热身期间启动的线程)通信,以执行该应用的预定义功能。预定义功能可作为后台进程来执行。应用可以监视预定义功能的进度,诸如通过跟踪预定义功能的状态。

[0086] 在640,可以从应用接收指示与预定义功能相关联的状态的响应。例如,状态可包括热身、初始、进行中、确认、消歧以及最终状态。响应可包括附加信息,诸如模板化列表、文本串、文本-到-语音串、图像、超链接、或可经由数字个人助理120的用户接口显示给用户的其他合适信息。

[0087] 在650,数字个人助理120的用户接口可以基于接收到的与预定义功能相关联的状态来向用户提供响应。以此方式,响应可来自数字个人助理120的用户接口的上下文内,而不出现该应用的用户接口。此外,数字个人助理120的确认和消歧能力可被用来确认和/或

澄清针对该应用的用户命令。

[0088] 用于确定是否使应用热身的示例方法

[0089] 图7是用于确定在用户正向数字个人助理120讲话时是否使应用热身的示例方法700的流程图。在710,用户可以向数字个人助理120键入、发声或讲话。可使用自然语言处理技术来分析用户的文本或语音并且可以从该语音识别出各单独词语。各单独词语可被分开地且在它们正被讲出的中间阶段中分析。例如,用户可以说出“hey Assistant, MyApp, do... (嗨助理,我的应用,做……)”。词语“hey”可以是载体词并且被丢弃。词语“Assistant”可被用来使数字个人助理120知晓用户正请求它执行动作。词语“MyApp”可被解释为应用。

[0090] 在720,所键入或讲出的词语可以与数字个人助理120的本机功能以及可扩展命令定义中提供的功能相比较。本机功能和命令定义文件中定义的功能可被统称为“已知AppService (应用服务)”。所讲词语可在各词语正被说出时被分析并与已知AppService相比较。换言之,对语音的分析可以发生在整个短语被用户讲出或键入之前。如果没有已知AppService是匹配的,则在730,数字个人助理120可打开web浏览器以使用与未被识别的所讲短语相对应的搜索串来检索搜索引擎网页。程序控制可被转移到web浏览器以使得用户可以细化web搜索和/或查看结果。然而,如果已知AppService是匹配的,则方法700可在740继续。

[0091] 在740,可以确定AppService应用是前台还是后台任务。例如,命令定义可包括将AppService应用定义为前台或后台应用的属性。如果AppService应用是前台任务,则在750,AppService应用可以在前台启动并且控制可被转移到AppService应用以完成该命令。如果AppService应用是后台任务,则方法700可以用并行步骤760和770继续。

[0092] 在760,数字个人助理120可以向用户提供与语音分析有关的信息。具体而言,数字个人助理120可以生成用于数字个人助理120的用户接口的进行中屏幕的输出。输出可被定义在例如命令定义的嵌套在<Command>元素内的<Feedback>元素中。输出可以是文本串且可以随着用户继续讲话而持续更新。

[0093] 在770,数字个人助理120可以使AppService应用热身而不等待用户话语结束。使AppService应用热身可包括分配存储器、预取指令、建立通信会话、从数据库检索信息、启动新执行线程、唤起中断、或其他合适的因应用而异的操作。该应用可基于投机性功能来热身。例如,与投机性功能相对应的指令可被获取,即使不确信该功能是所知道的。通过在用户完成所讲命令之前使应用热身,对用户作出响应的的时间可潜在地降低。

[0094] 在780,数字个人助理120可继续解析部分语音识别结果,直至话语完成。可基于被解析的命令和/或基于来自用户的暂停达预定时间量以上,来检测话语的结束。例如,可在识别出词语“queue (队列)”时检测到命令“MovieAppService, add MovieX to my queue”的结束。作为另一示例,命令“TextApp, text my wife that I will be home late for dinner (文本应用,用文本通知我妻子我将晚回家吃饭)”的结束可能更难以检测,因为该命令以未知长度的数据项结束。因而,暂停可被使用以向数字个人助理120指示该命令完成。

[0095] 在790,所讲命令的结束可被检测到且最终语音识别结果可被传递给该应用。该应用和数字个人助理120可彼此通信以完成所讲命令,如参考在前附图所描述的。

[0096] 计算系统

[0097] 图8描绘了其中可实现所描述的创新的合适的计算系统800的一般化示例。计算系统800并不旨在对使用范围或功能提出任何限制,因为这些创新可以在不同的通用或专用计算系统中实现。

[0098] 参考图8,计算系统800包括一个或多个处理单元810、815和存储器820、825。在图8中,该基本配置830被包括在虚线内。处理单元810、815执行计算机可执行的指令。处理单元可以是通用中央处理单元(CPU)、专用集成电路(ASIC)中的处理器或任意其它类型的处理器。在多处理系统中,多个处理单元执行计算机可执行指令以提高处理能力。例如,图8示出中央处理单元810以及图形处理单元或协处理单元815。有形存储器820、825可以是可由(诸)处理单元存取的易失性存储器(例如,寄存器、高速缓存、RAM)、非易失性存储器(例如,ROM、EEPROM、闪存等)或者两者的某一组合。存储器820、825以适合被(诸)处理单元执行的计算机可执行指令的形式,存储实现此处描述的一个或多个发明的软件880。

[0099] 计算系统可具有附加的特征。例如,计算系统800包括存储840、一个或多个输入设备850、一个或多个输出设备860以及一个或多个通信连接870。诸如总线、控制器或网络之类的互连机制(未示出)将计算系统800的各组件互连。通常,操作系统软件(未示出)为在计算系统800中执行的其它软件提供操作环境,并协调计算系统800的各组件的活动。

[0100] 有形存储840可以是可移动或不可移动的,并包括磁盘、磁带或磁带盒、CD-ROM、DVD或可用于储存信息并可在计算系统800内访问的任何其他介质。存储器840存储用于软件880的指令,所述软件880实现此处描述的一个或多个发明。

[0101] (诸)输入设备850可以是触摸输入设备(诸如键盘、鼠标、笔或跟踪球)、语音输入设备、扫描设备或向计算系统800提供输入的另一设备。对于视频编码,(诸)输入设备850可以是相机、视频卡、TV调谐卡或接受模拟或数字形式的视频输入的类似设备,或将视频样本读入计算系统800的CD-ROM或CD-RW。(诸)输出设备860可以是显示器、打印机、扬声器、CD刻录机或提供来自计算系统800的输出的另一设备。

[0102] (诸)通信连接870允许在通信介质上到另一计算实体的通信。通信介质传达诸如计算机可执行指令、音频或视频输入或输出、或已调制数据信号中的其他数据之类的信息。已调制数据信号是使其一个或多个特征以在信号中编码信息的方式设置或改变的信号。作为示例而非限制,通信介质可以使用电的、光学的、RF或其它载体。

[0103] 各创新可在计算机可执行指令(诸如包括在程序模块中的在目标现实或虚拟处理器上在计算系统中执行的那些计算机可执行指令)的一般上下文中描述。一般而言,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、库、对象、类、组件、数据结构等。如各实施例中描述的,这些程序模块的功能可以被组合,或者在这些程序模块之间拆分。针对各程序模块的计算机可执行指令可以在本地或分布式计算系统中执行。

[0104] 术语“系统”和“设备”在此被互换地使用。除非上下文明确指示,否则,术语并不暗示对计算系统或计算设备的类型的任何限制。一般说来,计算系统或计算设备可以是本地的或分布式的,并且可以包括具有实现本文中描述的功能的软件的专用硬件和/或通用硬件的任意组合。

[0105] 为了呈现起见,本详细描述使用了如“确定”和“使用”等术语来描述计算系统中的计算机操作。这些术语是对由计算机执行的操作的高级抽象,且不应与人类所执行的动作混淆。对应于这些术语的实际的计算机操作取决于实现而不同。

[0106] 移动设备

[0107] 图9是描述一示例移动设备900的系统示意图,该示例移动设备900包括在902概括示出的各种任选的硬件和软件组件。移动设备中的任何组件902可以与任何其他组件通信,然而为容易说明未示出所有连接。该移动设备可以是各种计算设备(例如,蜂窝电话、智能电话、手持式计算机、个人数字助理(PDA)等)中的任一个,并且可允许与诸如蜂窝、卫星或其他网络的一个或多个移动通信网络904进行无线双向通信。

[0108] 图示的移动设备900可以包括用于执行任务的控制器或处理器910(例如,信号处理器、微处理器、ASIC或者其他控制和处理逻辑电路),所述任务诸如信号编码、数据处理、输入/输出处理、功率控制、和/或其他功能。操作系统912可以控制各组件902的分配和使用,并且支持数字个人助理120和一个或多个应用程序914。应用程序可以包括常见的移动计算应用(例如,电子邮件应用、日历、联系人管理器、web浏览器、消息收发应用、影片应用、银行应用)、或者任何其他计算应用。应用程序914可包括具有可以由数字个人助理120无头地执行的的任务的应用。例如,任务可被定义在命令数据结构140中。用于访问应用存储的功能913还可以用于获取和更新应用程序914。

[0109] 图示的移动设备900可以包括存储器920。存储器920可以包括不可移动存储器922和/或可移动存储器924。不可移动存储器922可以包括RAM、ROM、闪存、硬盘或者其他公知的存储器存储技术。可移动存储器924可以包括闪存或订户身份模块(SIM)卡,这在GSM通信系统或者其他公知的存储器存储技术中是公知的,诸如“智能卡”。存储器920可用于存储运行操作系统912和应用914的数据和/或代码。示例数据可以包括要经由一个或多个有线或无线网络被发送至和/或接收自一个或多个网络服务器或其他设备的网页、文本、图像、声音文件、视频数据或者其他数据集。存储器920可用于存储订户标识符和设备标识符,所述订户标识符诸如国际移动订户身份(IMSI),所述设备标识符诸如国际移动设备标识符(IMEI)。这种标识符可以被发射至网络服务器以标识用户和设备。

[0110] 移动设备900可以支持一个或多个输入设备930以及一个或多个输出设备950,所述输入设备诸如触摸屏932、话筒934、相机936、物理键盘938和/或轨迹球940,所述输出设备诸如扬声器952和显示器954。其他可能的输出设备(未示出)可以包括压电或其他触觉输出设备。一些设备可以用于多于一个输入/输出功能。例如,触摸屏932和显示器954可以在单个输入/输出设备内被组合。

[0111] 输入设备930可以包括自然用户界面(NUI)。NUI是使用户能以“自然”方式与设备交互、免受诸如鼠标、键盘、远程控件等输入设备所施加的人工约束的任一界面技术。NUI方法的示例包括依赖于语音识别、触摸和触笔识别、屏上及邻近屏的姿势识别、空中姿势、头部和眼部跟踪、说话和语音、视觉、触摸、姿势以及机器智能的那些方法。NUI的其他示例包括使用加速度计/陀螺仪、面部识别、3D显示器、头部、眼部和凝视跟踪的运动手势检测、沉浸式增强的现实和虚拟现实系统,这些中的全部都提供更自然的界面,还包括使用电场传感电极(EEG及相关方法)来感测脑部活动的技术。因此,在一个具体示例中,操作系统912或应用914可以包括语音识别软件作为语音用户界面的一部分,该语音用户界面允许用户经由语音命令来操作设备900。而且,设备900可以包括允许经由用户的空间手势进行用户交互的输入设备和软件,诸如检测和解释手势以便向游戏应用提供输入。

[0112] 无线调制解调器960可被耦合到天线(未示出),并可支持处理器910和外部设备之

间的双向通信,如本领域中清楚理解的。调制解调器960被一般性地示出,并且可以包括用于与移动通信网络904的蜂窝调制解调器和/或其它基于无线电的调制解调器(例如蓝牙964或Wi-Fi 962)。无线调制解调器960一般被配置成与一个或多个蜂窝网络(诸如GSM网络)通信,用于单个蜂窝网络内、多个蜂窝网络之间、或者在移动设备和公共交换电话网(PSTN)之间的数据和语音通信。

[0113] 移动设备还可以包括至少一个输入/输出端口980、电源982、卫星导航系统接收机984(诸如全球定位系统(GPS)接收机)、加速度计986、和/或物理连接器990,物理连接器990可以是USB端口、IEEE 1394(火线)端口和/或RS-232端口。图示的组件902不是必须的或全包含的,因为任何组件可以被删除且其他组件可以被添加。

[0114] 云支持环境

[0115] 图10示出了其中可实现所描述的实施例、技巧和技术的合适的云支持环境1000的一般化示例。在示例环境1000中,由云1010提供各种类型的服务(例如,计算服务)。例如,云1010可以包括多个计算设备的集合,多个计算设备可以是中央式或分布式定位的,所述多个计算设备向经由诸如互联网这样的网络连接的各类用户和设备提供基于云的服务。实现方式环境1000可以不同方式被用来完成计算任务。例如,一些任务(例如,处理用户输入和呈现用户界面)可以在本地计算设备(例如,所连接的设备1030、1040、1050)上执行,而其他任务(例如,要在后续处理中使用的数据的存储)可以在云1010中执行。

[0116] 在示例环境1000中,云1010为具有各种屏幕能力的所连接的设备1030、1040、1050提供服务。所连接的设备1030代表具有计算机屏幕1035(例如,中等尺寸屏幕)的设备。例如,所连接的设备1030可以是诸如台式计算机、膝上型电脑、笔记本、上网本等个人计算机。所连接的设备1040代表具有移动设备屏幕1045(例如,小尺寸屏幕)的设备。例如,连接的设备1040可以是移动电话、智能电话、个人数字助理、平板计算机等。所连接的设备1050代表具有大屏幕1055的设备。例如,所连接的设备1050可以是电视机屏幕(例如,智能电视机)或连至电视机的另一设备(例如,机顶盒或游戏控制台)等等。所连接的设备1030、1040、1050中的一者或多者可以包括触摸屏能力。触摸屏可以以不同方式接受输入。例如,电容式触摸屏在对象(例如,指尖或触笔)跨表面上流动的电流扭曲或中断时检测到触摸输入。举另一个示例,触摸屏可以使用光学传感器在来自光学传感器的波束被中断时检测到触摸输入。对于一些触摸屏所检测的输入而言,与屏幕表面的物理接触不是必要的。在示例环境1000中也可以使用没有屏幕能力的设备。例如,云1010可以为没有显示器的一个或多个计算机(例如,服务器计算机)提供服务。

[0117] 服务可由云1010通过服务提供者1020提供、或通过其他在线服务提供者(未图示)而提供。例如,云服务可以被定制为特定的所连接设备(例如,所连接的设备1030、1040、1050)的屏幕大小、显示器能力和/或触摸屏能力。

[0118] 在示例环境1000中,云1010至少部分使用服务提供者1020,将此处所述的技术和解决方案提供给各种所连接的设备1030、1040、1050。例如,服务提供者1020可以为各种基于云的服务提供集中式解决方案。服务提供者1020可以管理对于各用户和/或设备(例如,对于所连接的设备1030、1040、1050和/或它们相应的用户)的服务预订。

[0119] 示例实现

[0120] 尽管以特定的顺序次序描述所公开方法的一些的操作以便于方便呈现,但应当理

解,这一描述方式包含重排,除非以下提出的具体语言要求特定的次序。例如,顺序描述的操作可以在一些情况下并行地重排或同时执行。此外,为简洁起见,附图可能不示出所公开的方法可以结合其他方法使用的各种方式。

[0121] 所公开的方法中的任何方法可被实现为被存储在一个或多个计算机可读存储介质上并在计算设备(例如任何可用计算设备,包括智能电话或其他包括计算硬件的移动设备)上执行的计算机可执行指令或计算机程序产品。计算机可读存储介质是在计算环境内可访问的任何可用的有形介质(例如,诸如DVD或CD之类的一个或多个光学介质盘、易失性存储器组件(诸如DRAM或SRAM)或非易失性存储器组件(诸如闪存或硬件驱动器))。作为示例并参考图8,计算机可读存储介质包括存储器820和825以及存储840。作为示例并参考图9,计算机可读存储介质包括存储器以及存储920、922和924。术语计算机可读存储介质不包括信号和载波。此外,术语计算机可读存储介质不包括通信连接(例如,870、960、962和964)。

[0122] 用于实现所公开技术的计算机可执行指令中的任一个以及在实现所公开的实施例期间创建和使用的任何数据可以被存储在一个或多个计算机可读存储介质上。计算机可执行指令可以是例如专用软件应用或者经由web浏览器或其他软件应用(诸如远程计算应用)访问和下载的软件应用的一部分。这种软件可以例如在单个本地计算机(例如,任何适当的商业可购买计算机)上或者使用一个或多个网络计算机在网络环境(例如,经由互联网、广域网、局域网、客户端-服务器网络(诸如云计算网络)、或者其他这样的网络)中执行。

[0123] 为清楚起见,仅描述了基于软件的实现方式的特定所选方面。省略了本领域公知的其他细节。例如,应当理解,所公开的技术不限于任何具体的计算机语言或程序。例如,所公开的技术可以通过以C++、Java、Perl、JavaScript、Adobe Flash或者其他适当的编程语言编写的软件来实现。同样,所公开的技术不限于任何特定的计算机或硬件类型。适当计算机和硬件的特定细节是公知的,并且不需要在本公开中详细提出。

[0124] 而且,基于软件的实施例(包括例如用于使计算机执行所公开方法的任一种的计算机可执行指令)中的任一者可以通过适当的通信手段被上载、下载或远程地访问。这种适当的通信手段包括例如互联网、万维网、内联网、软件应用、电缆(包括光纤电缆)、磁通信、电磁通信(包括RF、微波和红外通信)、电子通信或者其他这样的通信手段。

[0125] 所公开的方法、装置和系统不应以任何方式被限制。相反,本公开针对各种公开的实施例(单独和彼此的各种组合和子组合)的所有新颖和非显而易见的特征和方面。所公开的方法、装置和系统不限于任何具体方面或特征或它们的组合,所公开的实施例也不要求存在任一个或多个具体优点或者解决问题。

[0126] 来自任何示例的技术可以与在任何一个或多个其他示例中所描述的技术相组合。鉴于可应用所公开的本技术的原理的许多可能的实施例,应当认识到,所示实施例仅是所公开的技术的示例,并且不应被用作是对所公开的技术的范围的限制。

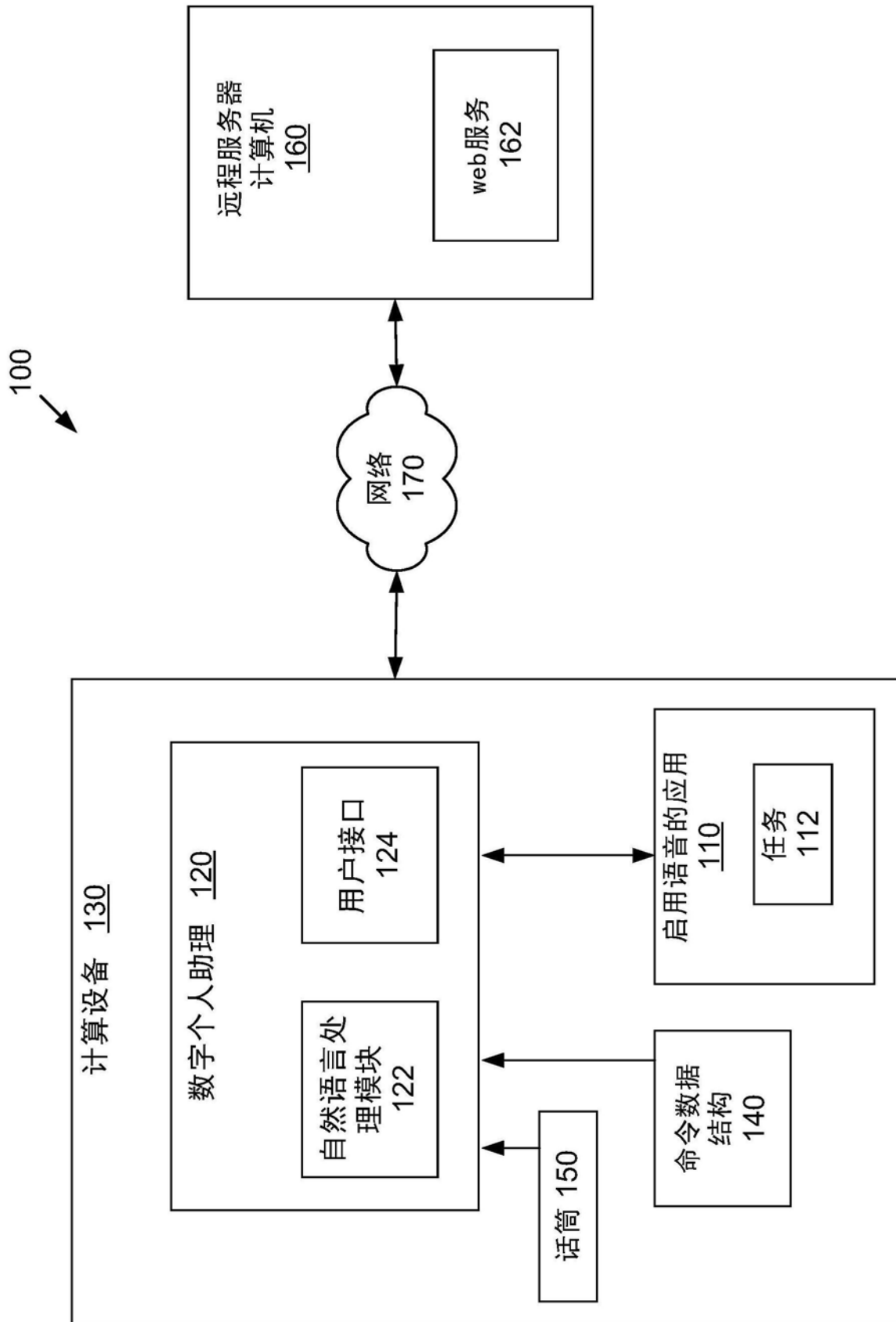


图1

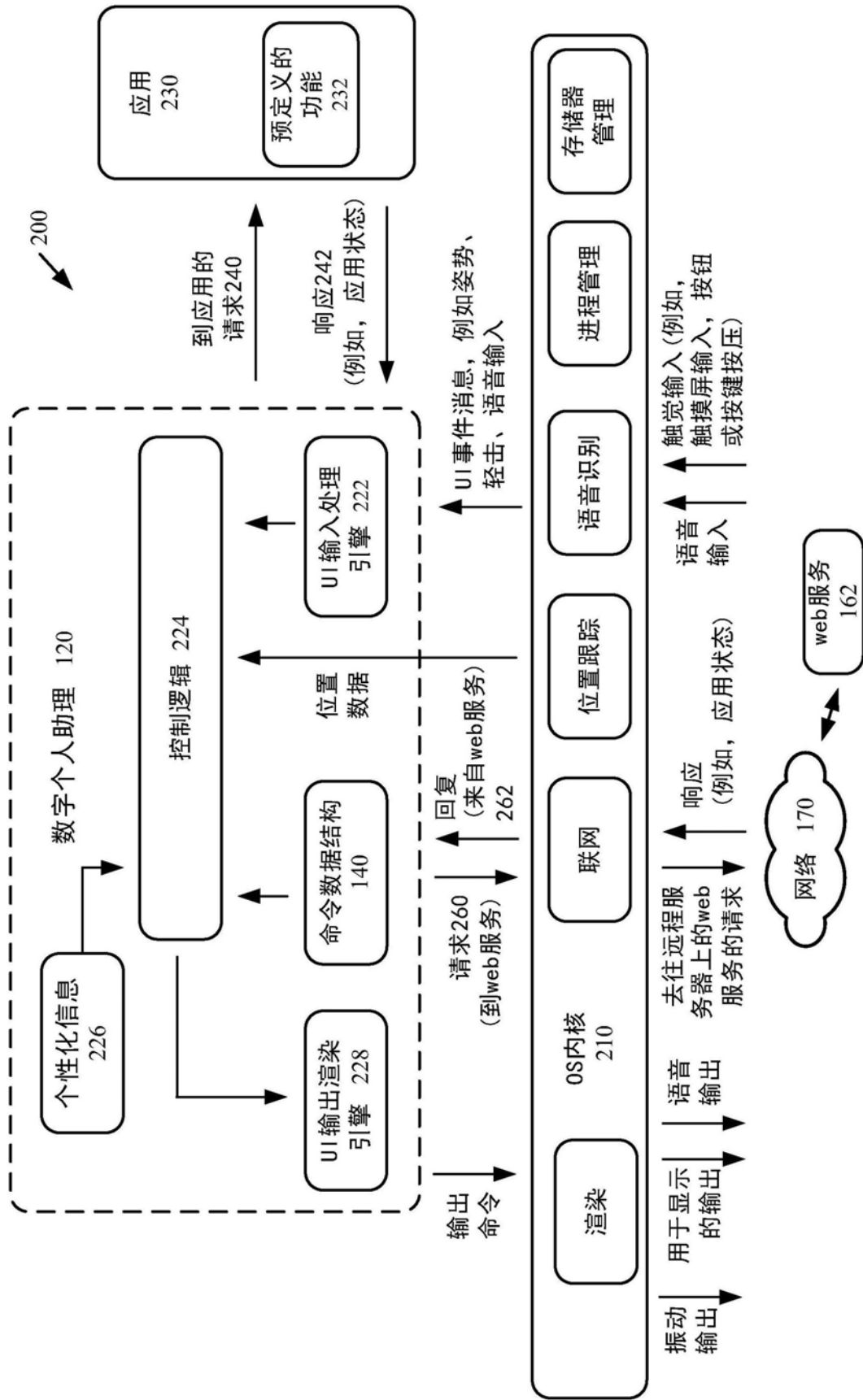


图2

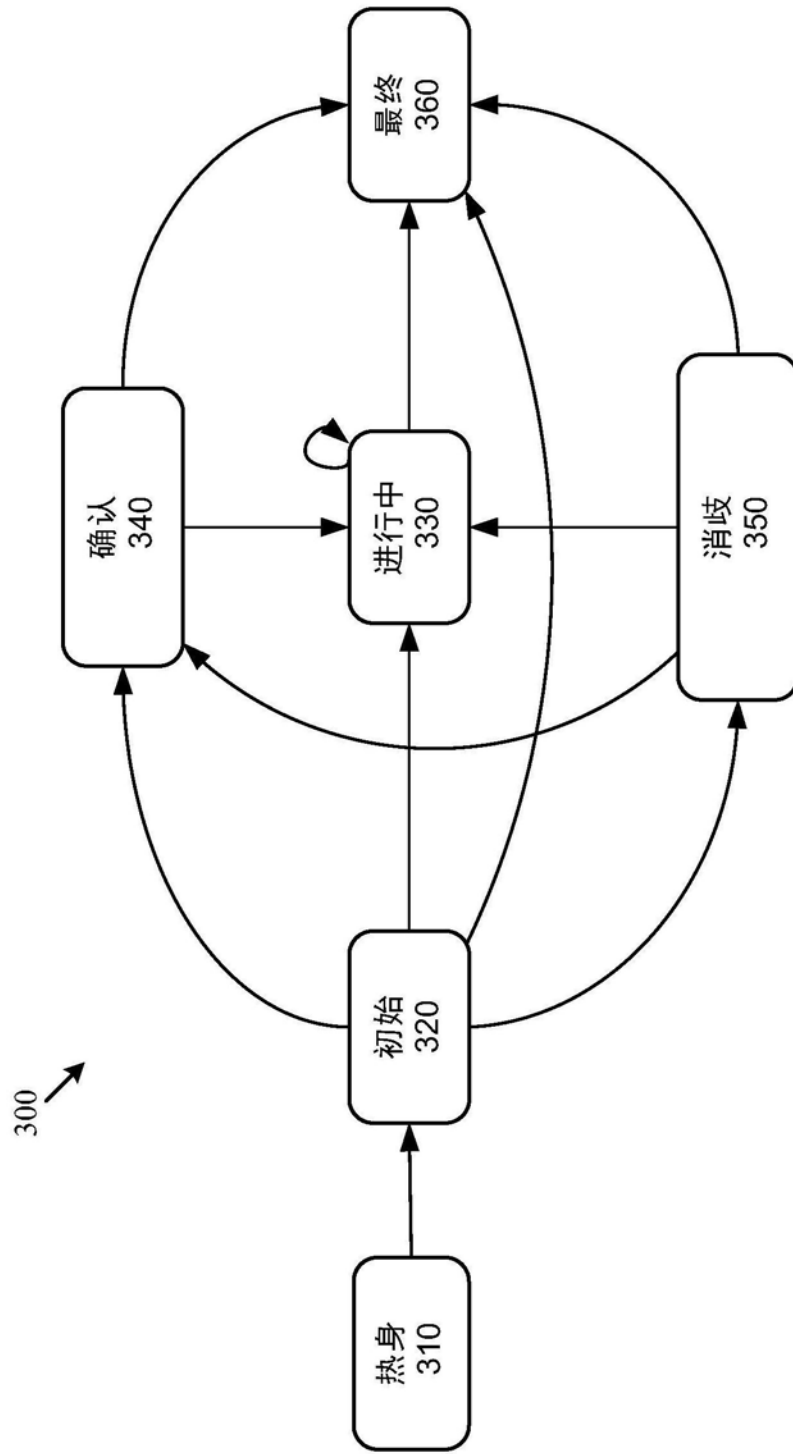


图3

400
↙

```

<CommandSet Name="MovieAppServiceTest">

  <AppName>MovieAppService</AppName>
  <Example>Add movieName</Example>

  <!--与命令有关的激活类型可以是后台或前台。作为示例，它可以默认地作为前台来处理。-->
  <Command Name="Add" ActivationType="background">
    <Example>Add the movieName to my MovieAppService queue</Example>
    <ListenFor>Add [the] {movieName} to [my] Queue</ListenFor>
    <ListenFor>Add [the] {movieName} to [my] {builtin:AppName} Queue</ListenFor>
    <Feedback>{builtin:ackPhrase}, adding {movieName} to your queue</Feedback>

    <!--响应可以使可任选的且所以可具有单独的响应子状态。如果应用没有提供对特定状态的响应，
    则可以显示录制的响应。 -->
    <Responses>

      <Response Name="SuccessMovieAdded" State="Success">
        <DisplayString>{builtin:success}, I've got that done for you</DisplayString>
        <TTSSString>{builtin:success}, I've got that done for you</TTSSString>
        <AppDeepLink URL="/showQueue.xml?movieAdded=true">View my MovieAppService
Queue
        <WebLink URL="http://www.MovieAppService.com/myQueue">Manage queue on
MovieAppService.com
</WebLink>
      </Response>

      <Response Name="FailureMovieNotAdded" State="Failure">
        <DisplayString>{builtin:failure}, that didnt work. Let's try to do this in the
{builtin:appName} app</DisplayString>
        <TTSSString>{builtin:failure}, that didnt work. Let's try to do this in the
{builtin:appName} app</TTSSString>
        <AppDeepLink></AppDeepLink>
        <WebLink></WebLink>
      </Response>

    </Responses>

  <!--在无头调用的情形中，导航元素可以指向应用服务以在这一命令被命中时启动。-->
  <Navigate Target="AppName-VoiceCommandService"/>
</Command>

</CommandSet>

```

图4

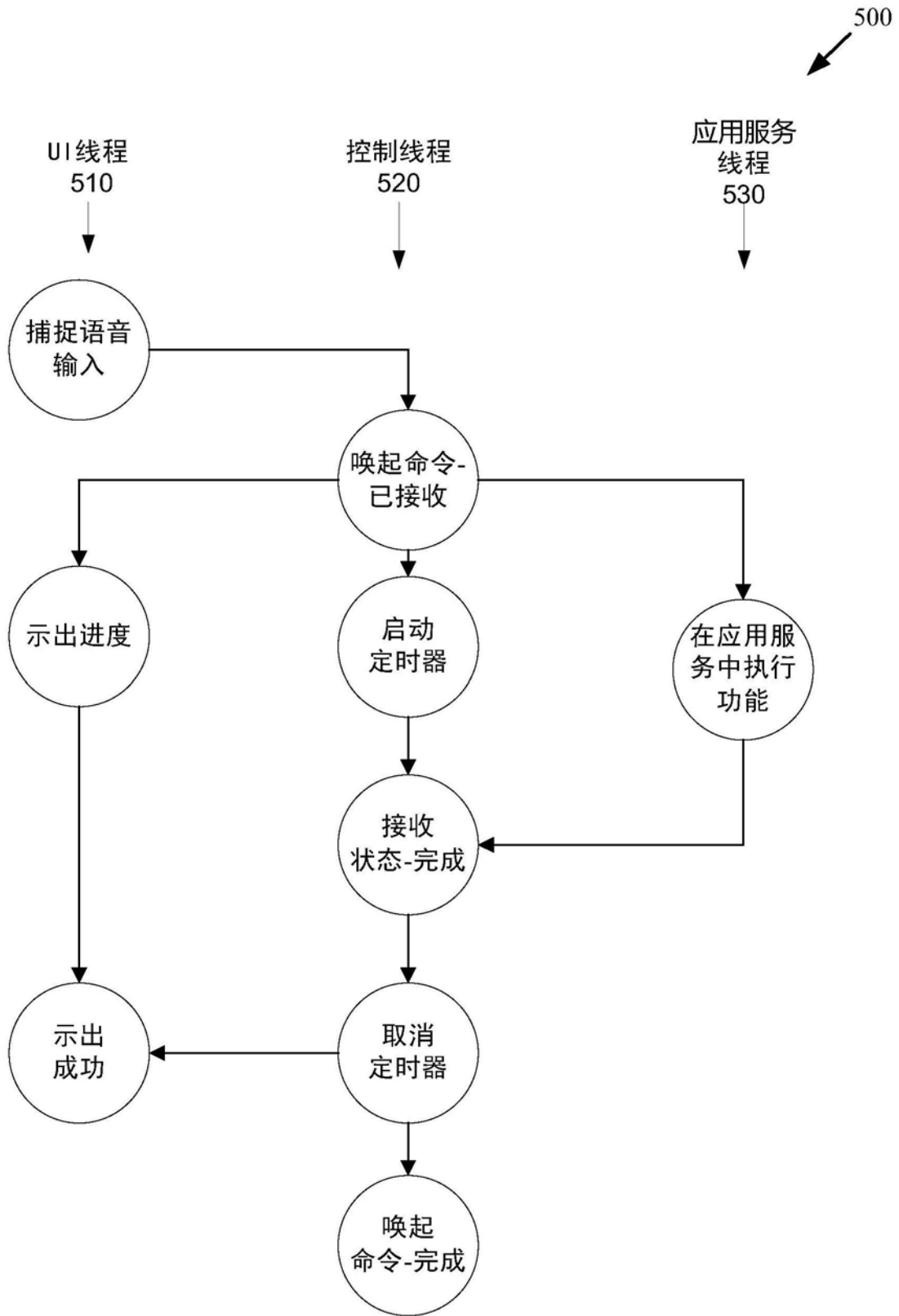


图5

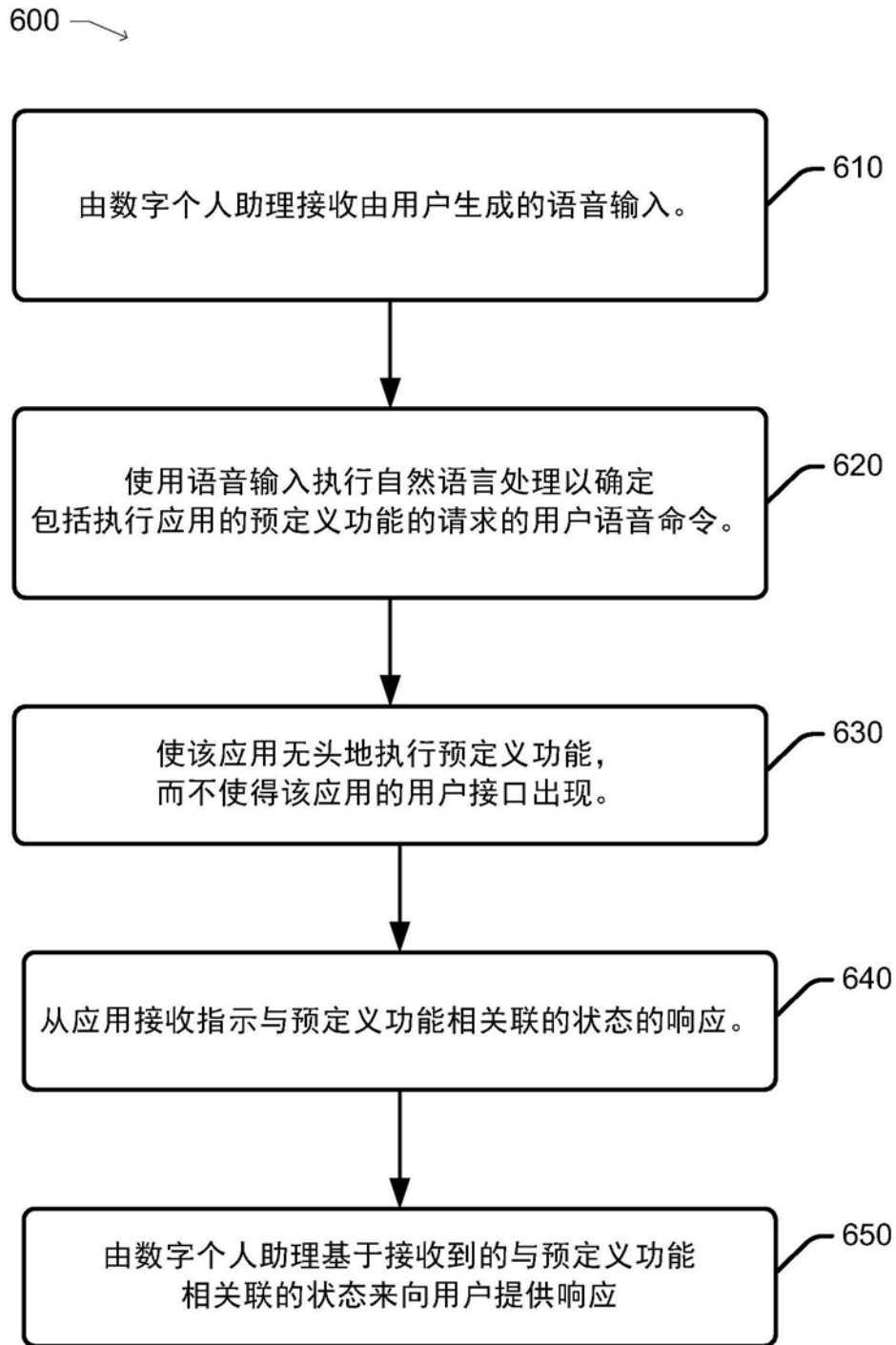


图6

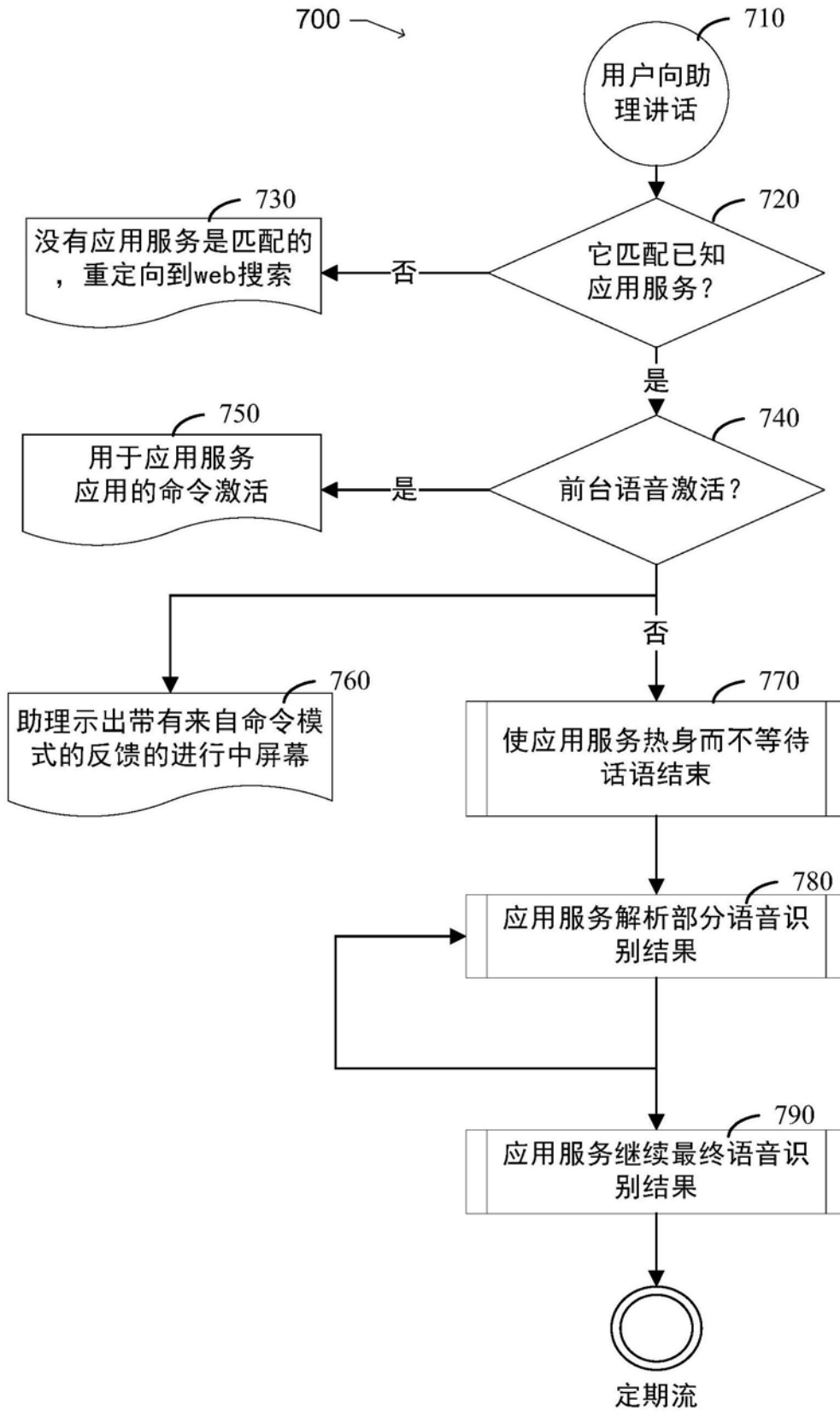


图7

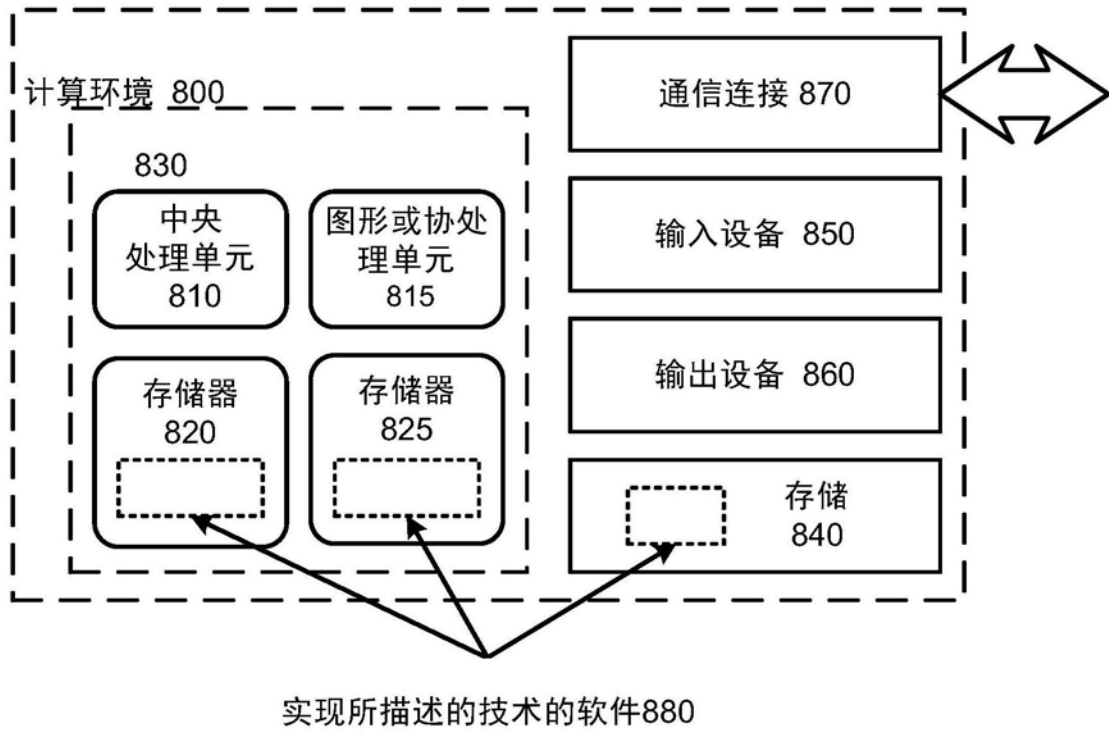


图8

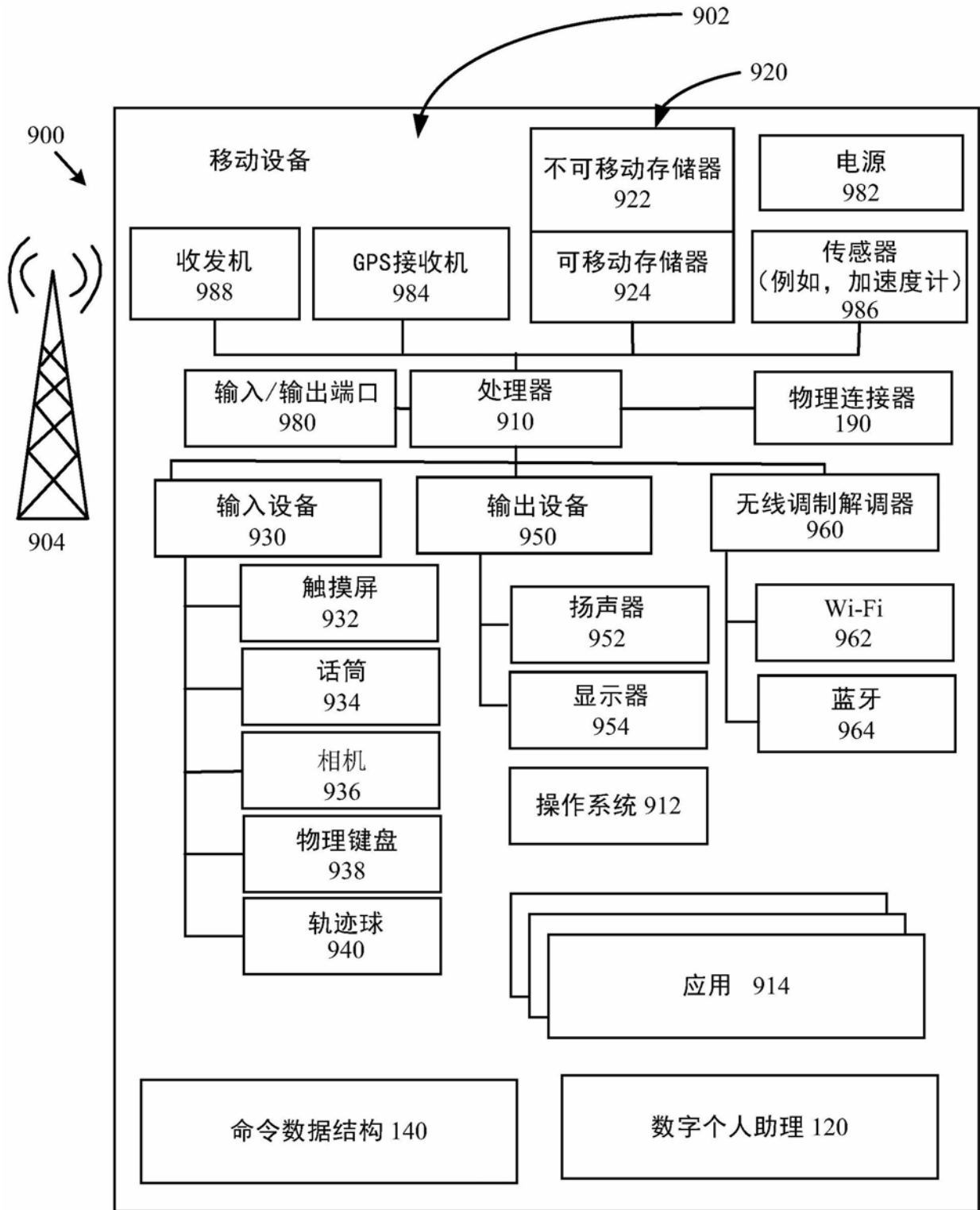


图9

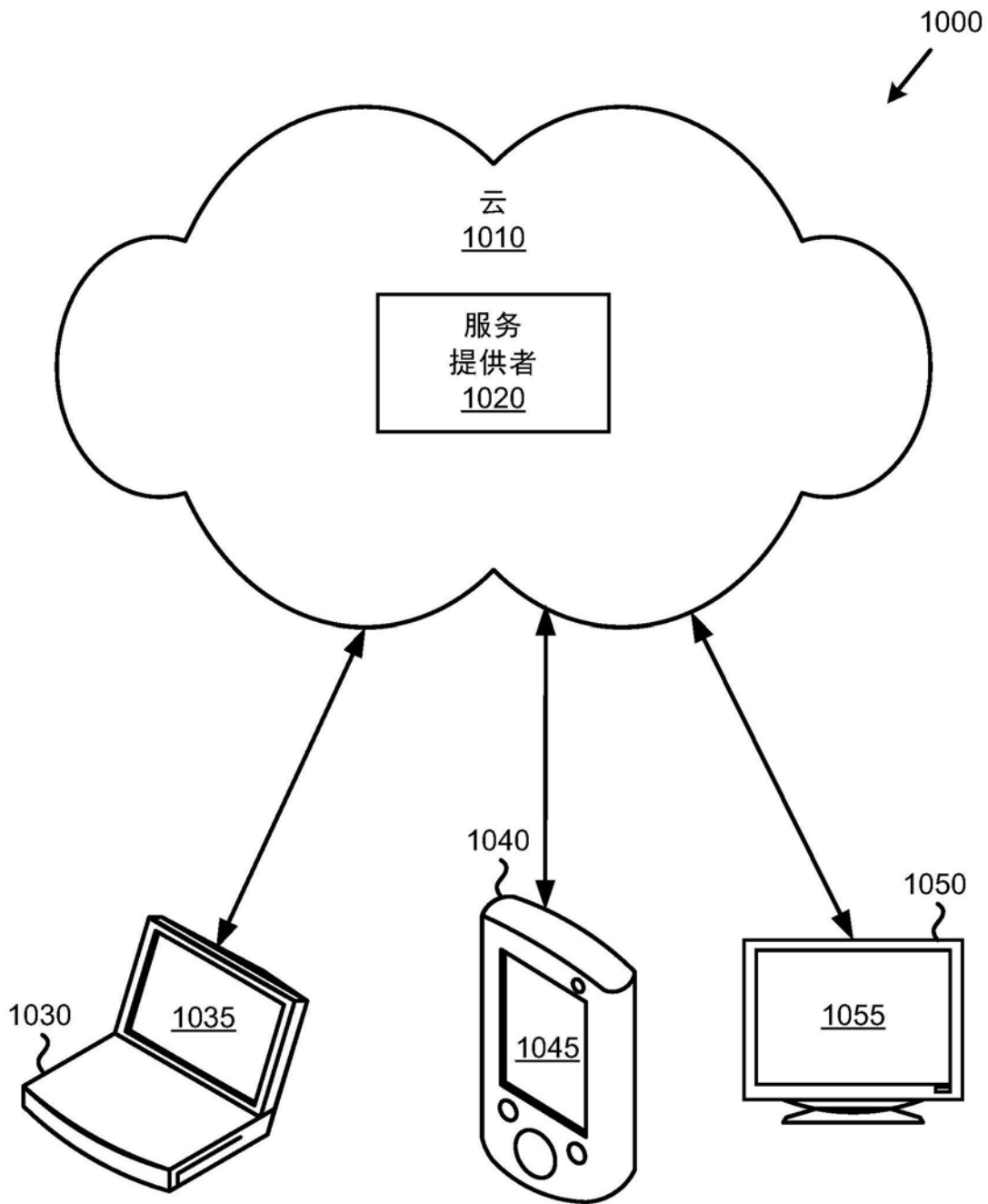


图10