



## (12)发明专利申请

(10)申请公布号 CN 110460715 A

(43)申请公布日 2019.11.15

(21)申请号 201910373311.5

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038

(22)申请日 2019.05.07

代理人 刘玉洁

(30)优先权数据

PA201870384 2018.06.12 DK

62/668,154 2018.05.07 US

62/679,798 2018.06.02 US

16/032,487 2018.07.11 US

(51)Int.Cl.

H04M 1/725(2006.01)

G10L 15/22(2006.01)

G06F 3/041(2006.01)

G06F 3/01(2006.01)

(71)申请人 苹果公司

地址 美国加利福尼亚

(72)发明人 S·O·勒梅 M·R·巴斯蒂安

R·霍伦斯坦 M·郑

C·马阿卢夫 B·J·牛恩多普

H·涅托 T·派克 J·彼得森

S·斯库利 S·斯里达尔

B·威斯汀 S·赵

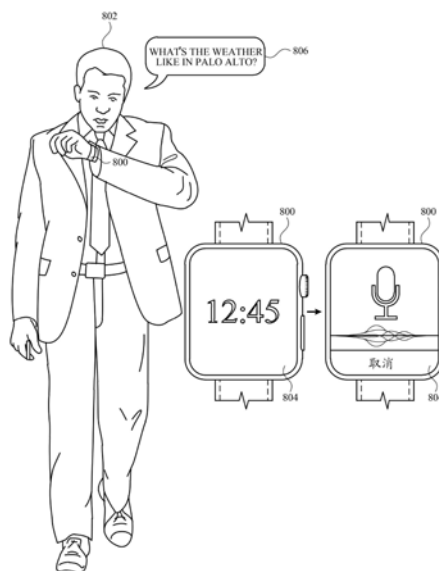
权利要求书5页 说明书59页 附图24页

(54)发明名称

用于操作数字助理的方法、设备和介质

(57)摘要

本发明题为“用于操作数字助理的方法、设备和介质”。本发明提供了用于操作智能自动化助理的系统 and 过程。示例性过程包括检测表示电子设备的运动的输入并利用该电子设备的麦克风对音频输入进行采样。该示例性过程还包括基于音频输入和表示电子设备的运动的输入来确定是否发起虚拟助理会话。根据确定发起虚拟助理会话，该示例性过程包括发起虚拟助理会话。根据确定不发起虚拟助理会话，该示例性过程包括放弃发起虚拟助理会话。



1. 一种方法,包括:  
在具有麦克风和显示器的电子设备处:  
检测表示所述电子设备的运动的输入;  
利用所述麦克风对音频输入进行采样;  
基于所述音频输入和表示所述电子设备的运动的输入,确定是否发起虚拟助理会话;  
根据确定发起所述虚拟助理会话,发起所述虚拟助理会话;以及  
根据确定不发起所述虚拟助理会话,放弃发起所述虚拟助理会话。
2. 根据权利要求1所述的方法,还包括:  
根据发起所述虚拟助理会话:  
基于所述音频输入确定用户意图;  
基于所述用户意图执行任务;以及  
提供与所述任务相关联的输出。
3. 根据权利要求1所述的方法,还包括:  
根据发起所述虚拟助理会话:  
根据不基于所述音频输入确定用户意图,放弃响应于所述音频输入提供输出。
4. 根据权利要求1至3中任一项所述的方法,其中所述音频输入不包括口头触发。
5. 根据权利要求1至4中任一项所述的方法,还包括:  
基于表示所述电子设备的运动的所述输入的第一部分来开启所述显示器;并且  
其中根据开启所述显示器来对所述音频输入进行采样。
6. 根据权利要求5所述的方法,其中表示所述电子设备的运动的所述输入的第一部分表示手腕举高手势或手腕旋转手势。
7. 根据权利要求1至6中任一项所述的方法,其中根据开启所述显示器来执行对所述音频输入的采样。
8. 根据权利要求1至7中任一项所述的方法,其中在检测表示所述电子设备的运动的所述输入的至少第二部分的同时对所述音频输入进行采样。
9. 根据权利要求1至8中任一项所述的方法,其中表示所述电子设备的运动的输入由所述电子设备的一个或多个加速度计检测。
10. 根据权利要求1至9中任一项所述的方法,其中表示所述电子设备的运动的输入由所述电子设备的一个或多个陀螺仪检测。
11. 根据权利要求1至10中任一项所述的方法,其中所述电子设备的所述运动包括所述电子设备在多个姿势中的两个或更多个姿势之间的转变。
12. 根据权利要求11所述的方法,其中所述多个姿势包括已垂下姿势、正在举高姿势、已举高姿势和正在垂下姿势。
13. 根据权利要求12所述的方法,其中确定是否发起所述虚拟助理会话包括:  
基于表示所述电子设备的运动的所述输入,确定所述电子设备的所述运动是否包括从所述正在举高姿势到所述已举高姿势的转变。
14. 根据权利要求13所述的方法,其中确定是否发起所述虚拟助理会话包括:  
确定从所述正在举高姿势到所述已举高姿势的所述转变的持续时间。
15. 根据权利要求13至14中任一项所述的方法,其中在第一时间对所述音频输入进行

采样,其中从所述正在举高姿势到所述已举高姿势的所述转变与第二时间相关联,并且其中确定是否发起所述虚拟助理会话包括基于所述第一时间和所述第二时间确定是否发起所述虚拟助理会话。

16.根据权利要求12至15中任一项所述的方法,其中确定是否发起所述虚拟助理会话包括:

基于表示所述电子设备的运动的所述输入,确定所述已举高姿势的持续时间是否超过第一预定持续时间。

17.根据权利要求12至16中任一项所述的方法,其中确定是否发起所述虚拟助理会话包括:

基于表示所述电子设备的运动的所述输入,确定所述电子设备的所述运动是否包括从所述已举高姿势到所述正在垂下姿势的第一转变。

18.根据权利要求12至17中任一项所述的方法,还包括:

根据发起所述虚拟助理会话:

基于表示所述电子设备的运动的所述输入,确定所述电子设备的所述运动是否包括从所述已举高姿势到所述正在垂下姿势的第二转变;以及

根据确定所述电子设备的所述运动包括从所述已举高姿势到所述正在垂下姿势的所述第二转变,将所述虚拟助理会话去激活。

19.根据权利要求18所述的方法,还包括:

确定与从所述已举高姿势到所述正在垂下姿势的所述第二转变相关联的时间;

确定与所述音频输入中的语音结束相关联的时间;并且

其中进一步根据确定与从所述已举高姿势到所述正在垂下姿势的所述第二转变相关联的时间在与所述音频输入中的所述语音结束相关联的时间之前,执行将所述虚拟助理会话去激活。

20.根据权利要求12至19中任一项所述的方法,还包括:

确定与从所述已举高姿势到所述正在垂下姿势的第三转变相关联的时间;以及

基于与从所述已举高姿势到所述正在垂下姿势的所述第三转变相关联的时间来确定所述音频输入的终点。

21.根据权利要求12至20中任一项所述的方法,其中确定是否发起所述虚拟助理会话包括基于运动模型确定是否发起所述虚拟助理会话。

22.根据权利要求21所述的方法,其中所述运动模型包括第一神经网络。

23.根据权利要求21至22中任一项所述的方法,其中所述运动模型特定于所述电子设备的第一用户。

24.根据权利要求21至23中任一项所述的方法,还包括:

使用所述运动模型确定所述电子设备在所述多个姿势中的两个或更多个姿势之间的转变的概率;以及

使用所述运动模型确定所述多个姿势中的姿势的概率。

25.根据权利要求1至24中任一项所述的方法,其中检测表示所述电子设备的运动的所述输入包括使用第一处理器对表示所述电子设备的运动的所述输入进行采样。

26.根据权利要求25所述的方法,其中对所述音频输入进行采样包括使用所述第一处

理器对所述音频输入进行采样。

27. 根据权利要求25所述的方法, 其中对所述音频输入进行采样包括使用第二处理器对所述音频输入进行采样。

28. 根据权利要求27所述的方法, 其中所述第二处理器在操作时比所述第一处理器消耗更多的功率。

29. 根据权利要求27至28中任一项所述的方法, 其中基于表示所述电子设备的运动的所述输入的至少第三部分来激活所述第二处理器。

30. 根据权利要求1至29中任一项所述的方法, 其中确定是否发起所述虚拟会话包括确定所述音频输入是否包括人类语音。

31. 根据权利要求30所述的方法, 其中确定所述音频输入是否包括人类语音包括确定所述音频输入是否包括由所述电子设备的授权用户说出的人类语音。

32. 根据权利要求1至31中任一项所述的方法, 其中确定是否发起所述虚拟助理会话包括确定与所述音频输入相关联的方向。

33. 根据权利要求1至32中任一项所述的方法, 其中确定是否发起所述虚拟助理会话包括确定所述音频输入与所述电子设备的接近度。

34. 根据权利要求1至33中任一项所述的方法, 其中确定是否发起所述虚拟助理会话包括基于音频模型确定是否发起所述虚拟助理会话。

35. 根据权利要求34所述的方法, 其中所述音频模型包括第二神经网络。

36. 根据权利要求34至35中任一项所述的方法, 其中所述音频模型特定于所述电子设备的第二用户。

37. 根据权利要求1至36中任一项所述的方法, 其中确定是否发起所述虚拟助理会话包括:

在第三时间, 基于所述音频输入确定是否发起所述虚拟助理会话;

在第四时间, 基于表示所述电子设备的运动的所述输入确定是否发起所述虚拟助理会话; 以及

确定所述第三时间和所述第四时间是否在第二预定持续时间内; 并且

其中根据确定所述第三时间和所述第四时间在所述第二预定持续时间内, 执行发起所述虚拟助理会话。

38. 根据权利要求1至37中任一项所述的方法, 还包括:

根据发起所述虚拟助理会话:

基于所述音频输入的语言内容确定是否将所述虚拟助理会话去激活;

根据确定将所述虚拟助理会话去激活, 将所述虚拟助理会话去激活; 以及

根据确定不将所述虚拟助理会话去激活, 放弃将所述虚拟助理会话去激活。

39. 根据权利要求1至38中任一项所述的方法, 还包括:

根据发起所述虚拟助理会话:

确定所述音频输入的持续时间是否小于阈值持续时间; 以及

根据确定所述音频输入的所述持续时间小于所述阈值持续时间, 将所述虚拟助理会话去激活。

40. 根据权利要求1至39中任一项所述的方法, 其中发起所述虚拟助理会话包括显示与

虚拟助理相关联的第一用户界面。

41. 根据权利要求1至40中任一项所述的方法, 其中发起所述虚拟助理会话包括提供音频输出。

42. 根据权利要求1至41中任一项所述的方法, 其中发起所述虚拟助理会话包括提供触觉输出。

43. 根据权利要求1至42中任一项所述的方法, 还包括:

根据发起所述虚拟助理会话, 将所述虚拟助理会话去激活。

44. 根据权利要求43所述的方法, 其中将所述虚拟助理会话去激活包括停止显示与所述数字助理相关联的第二用户界面。

45. 根据权利要求43至44中任一项所述的方法, 其中将所述虚拟助理会话去激活包括放弃对所述音频输入进行响应。

46. 根据权利要求1至45中任一项所述的方法, 其中所述电子设备是计算机、扬声器、智能手表、电话或它们的组合。

47. 一种电子设备, 包括:

麦克风;

显示器;

一个或多个处理器;

存储器; 和

一个或多个程序, 其中所述一个或多个程序存储在所述存储器中并且被配置为由所述一个或多个处理器执行, 所述一个或多个程序包括用于执行以下操作的指令:

检测表示所述电子设备的运动的输入;

利用所述麦克风对音频输入进行采样;

基于所述音频输入和表示所述电子设备的运动的输入, 确定是否发起虚拟助理会话;

根据确定发起所述虚拟助理会话, 发起所述虚拟助理会话; 以及

根据确定不发起所述虚拟助理会话, 放弃发起所述虚拟助理会话。

48. 一种存储一个或多个程序的非暂态计算机可读存储介质, 所述一个或多个程序包括指令, 所述指令当由具有麦克风和显示器的电子设备的一个或多个处理器执行时, 使得所述电子设备:

检测表示所述电子设备的运动的输入;

利用所述麦克风对音频输入进行采样;

基于所述音频输入和表示所述电子设备的运动的输入, 确定是否发起虚拟助理会话;

根据确定发起所述虚拟助理会话, 发起所述虚拟助理会话; 以及

根据确定不发起所述虚拟助理会话, 放弃发起所述虚拟助理会话。

49. 一种电子设备, 包括用于执行以下操作的装置:

检测表示所述电子设备的运动的输入;

对音频输入进行采样;

基于所述音频输入和表示所述电子设备的运动的输入, 确定是否发起虚拟助理会话;

根据确定发起所述虚拟助理会话, 发起所述虚拟助理会话; 以及

根据确定不发起所述虚拟助理会话, 放弃发起所述虚拟助理会话。

50. 一种电子设备,包括:

一个或多个处理器;

存储器;和

一个或多个程序,其中所述一个或多个程序存储在所述存储器中并且被配置为由所述一个或多个处理器执行,所述一个或多个程序包括用于执行根据权利要求1至46中任一项所述的方法的指令。

51. 一种存储一个或多个程序的非暂态计算机可读存储介质,所述一个或多个程序包括指令,所述指令当由电子设备的一个或多个处理器执行时,使得所述电子设备执行根据权利要求1至46中任一项所述的方法。

52. 一种电子设备,包括:

用于执行根据权利要求1至46中任一项所述的方法的装置。

## 用于操作数字助理的方法、设备和介质

[0001] 本申请要求2018年5月7日提交的美国专利申请No. 62/668,154 “RAISE TO SPEAK”以及2018年6月2日提交的美国专利申请No. 62/679,798“RAISE TO SPEAK”的优先权,这两个专利申请的内容据此全文以引用方式并入。

### 技术领域

[0002] 本文整体涉及智能自动化助理,并且更具体地讲,涉及服务诸如智能自动化助理来发起语音。

### 背景技术

[0003] 智能自动化助理(或数字助理)可在人类用户与电子设备之间提供有利界面。此类助理可允许用户使用自然语言以语音形式和/或文本形式与设备或系统进行交互。例如,用户可向正在电子设备上运行的数字助理提供包含用户请求的语音输入。数字助理可从该语音输入解译用户意图并且将用户意图操作化成任务。随后可通过执行电子设备的一项或多项服务来执行这些任务,并且可将响应于用户请求的相关输出返回给用户。

[0004] 操作数字助理需要电力,在依赖电池且其上通常运行数字助理的手持式或便携式设备上,电力是有限资源。因此,期望以节能的方式发起和操作数字助理。

### 发明内容

[0005] 本文公开了示例方法。示例方法包括在具有麦克风和显示器的电子设备处:检测表示电子设备的运动的输入,以及用麦克风来对音频输入进行采样。该方法还包括基于音频输入和表示电子设备的运动的输入来确定是否发起虚拟助理会话。该方法还包括根据发起虚拟助理会话的确定来发起虚拟助理会话,并且根据不发起虚拟助理会话的确定来放弃发起虚拟助理会话。

[0006] 本文公开了示例非暂态计算机可读介质。一种示例非暂态计算机可读存储介质存储一个或多个程序。该一个或多个程序包括指令,该指令在由具有麦克风和显示器的电子设备的一个或多个处理器执行时使得电子设备:检测表示电子设备的运动的输入;用麦克风来对音频输入进行采样;基于音频输入和表示电子设备的运动的输入来确定是否发起虚拟助理会话;根据发起虚拟助理会话的确定来发起虚拟助理会话;并且根据不发起虚拟助理会话的确定来放弃发起虚拟助理会话。

[0007] 本文公开了示例电子设备。示例电子设备包括麦克风、显示器、一个或多个处理器、存储器和一个或多个程序,其中一个或多个程序存储在存储器中并且被配置为由一个或多个处理器执行,该一个或多个程序包括用于以下操作的指令:检测表示电子设备的运动的输入;用麦克风来对音频输入进行采样;基于音频输入和表示电子设备的运动的输入来确定是否发起虚拟助理会话;根据发起虚拟助理会话的确定来发起虚拟助理会话;并且根据不发起虚拟助理会话的确定来放弃发起虚拟助理会话。

[0008] 示例电子设备包括用于以下操作的装置:检测表示电子设备的运动的输入;对音

频输入进行采样;基于音频输入和表示电子设备的运动的输入来确定是否发起虚拟助理会话;根据发起虚拟助理会话的确定来发起虚拟助理会话;并且根据不发起虚拟助理会话的确定来放弃发起虚拟助理会话。

[0009] 基于音频输入和表示电子设备的运动的输入(例如,“运动输入”)来确定是否发起虚拟助理会话提供了自然且有效的发起虚拟助理并与其进行交互的方式。例如,为了发起虚拟助理并基于口头请求让虚拟助理执行任务,用户仅仅需要将设备举高至他或她的嘴部附近并且在大致相同的时间处发出请求(例如,不需要包括口头触发(诸如“嘿,Siri”)的请求)。以这种方式促进虚拟助理交互能够实现有效的虚拟助理交互(例如,在每次交互之前不需要口头触发)以及对口头请求的快速虚拟助理响应。此外,使用音频输入和运动输入这二者来发起虚拟助理会话可防止基于音频输入和运动输入中任一者的虚拟助理意外激活。这降低了意外发起虚拟助理的几率并且节省了原本在非期望的情况下操作虚拟助理而消耗的电池电力。这样,使得用户-设备界面更有效(例如,通过减少用户提供的音频输入的量,通过减少虚拟助理响应时间,通过防止虚拟助理意外激活),此外,通过使得用户能够快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

## 附图说明

[0010] 图1是示出了根据各种示例的用于实现数字助理的系统和环境的框图。

[0011] 图2A是示出根据各种示例的实现数字助理的客户端侧部分的便携式多功能设备的框图。

[0012] 图2B是示出了根据各种示例的用于事件处理的示例性部件的框图。

[0013] 图3示出了根据各种示例的实现数字助理的客户端侧部分的便携式多功能设备。

[0014] 图4是根据各种示例的具有显示器和触敏表面的示例性多功能设备的框图。

[0015] 图5A示出了根据各种示例的便携式多功能设备上的应用程序的菜单的示例性用户界面。

[0016] 图5B示出了根据各种示例的具有与显示器分开的触敏表面的多功能设备的示例性用户界面。

[0017] 图6A示出了根据各种示例的个人电子设备。

[0018] 图6B是示出根据各种示例的个人电子设备的框图。

[0019] 图7A是示出了根据各种示例的数字助理系统或其服务器部分的框图。

[0020] 图7B示出了根据各种示例的图7A中所示的数字助理的功能。

[0021] 图7C示出了根据各种示例的知识本体的一部分。

[0022] 图8A至图8D示出了用于基于音频输入和表示电子设备的运动的输入来在电子设备处发起虚拟助理会话的示例性技术。

[0023] 图8E至图8F示出了用于去激活虚拟助理会话的示例性技术。

[0024] 图9示出了根据一些示例的被配置为处理音频输入和运动输入的系统的框图。

[0025] 图10A至图10D示出了根据各种示例的用于操作数字助理的过程。

[0026] 图11示出了根据各种示例的用于传输通信的过程。



## 具体实施方式

[0027] 在以下对示例的描述中将引用附图,在附图中以例示的方式示出了可被实施的特定示例。应当理解,在不脱离各个示例的范围的情况下,可使用其他示例并且可作出结构性变更。

[0028] 如上文所讨论,本公开涉及用于基于音频输入和运动输入发起虚拟助理会话的技术。例如,为了在手腕可佩戴设备(例如,智能手表)上发起虚拟助理会话,佩戴该设备的用户将该设备举高至用户的嘴部附近并且在该设备被举高至用户嘴部附近时(或者在之前或之后不久)提供音频输入(例如,“帕罗奥图天气怎么样?”)。在一些示例中,音频输入不包括用于发起虚拟助理的口头触发(例如,“嘿,Siri”)。基于检测到的运动输入和采样的音频输入在设备上发起虚拟助理会话,并且发起的虚拟助理会话响应音频输入(例如,通过提供响应“气温70度,晴天”)。因此,提供了直观且有效的发起虚拟助理会话并与其进行交互的方式。

[0029] 尽管以下描述使用术语“第一”、“第二”等来描述各种元件,但这些元件不应受所述术语的限制。这些术语只是用于将一个元件与另一元件区分开。例如,在不脱离各种所述示例的范围的情况下,第一输入可被称为第二输入,并且类似地,第二输入可被称为第一输入。第一输入和第二输入均为输入,并且在一些情况下为独立且不同的输入。

[0030] 在本文中各种所述示例的描述中所使用的术语仅仅是为了描述特定示例的目的,而并非旨在进行限制。如在对各种所述示例的描述和所附权利要求书中所使用的那样,单数形式“一个(“a”,“an”)”和“该”旨在也包括复数形式,除非上下文另外明确地指示。还将理解的是,本文中所使用的术语“和/或”是指并且涵盖相关联的所列出的项目中的一个或多个项目的任何和全部可能的组合。还将理解的是,术语“包括”(“includes”、“including”、“comprises”和/或“comprising”)在本说明书中使用是指定存在所陈述的特征、整数、步骤、操作、元件和/或部件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、部件和/或其分组。

[0031] 根据上下文,术语“如果”可被解释为意指“当...时”或“在...时”或“响应于确定...”或“响应于检测到...”。类似地,根据上下文,短语“如果确定...”或“如果检测到[所陈述的条件或事件]”可被解释为意指“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]”。

### [0032] 1. 系统和环境

[0033] 图1示出了根据各种示例的系统100的框图。在一些示例中,系统100 实现数字助理。术语“数字助理”、“虚拟助理”、“智能自动化助理”或“自动数字助理”是指解译口头形式和/或文本形式的自然语言输入来推断用户意图,并且基于推断出的用户意图来执行动作的任何信息处理系统。例如,为了作用于推断出的用户意图,系统执行下述步骤中的一个或多个:识别具有设计用于实现推断出的用户意图的步骤和参数的任务流,根据推断出的用户意图将特定要求输入到任务流中;通过调用程序、方法、服务、API等执行任务流;以及以可听(例如,语音)和/或可视形式来生成对用户的输出响应。

[0034] 具体地讲,数字助理能够接受至少部分地为自然语言命令、请求、声明、讲述和/或询问的形式的用户请求。通常,用户请求要么寻求数字助理作出信息性回答,要么寻求数字助理执行任务。对用户请求的令人满意的响应包括提供所请求的信息性回答、执行所请求

的任务或这两者的组合。例如,用户向数字助理提出问题,诸如“我现在在哪里?”。基于用户的当前位置,数字助理回答“你在中央公园西门附近。”用户还请求执行任务,例如“请邀请我的朋友们下周来参加我女朋友的生日聚会。”作为响应,数字助理可通过讲出“好的,马上”来确认请求,然后代表用户将合适的日历邀请发送到用户电子通讯录中列出的用户朋友中的每位朋友。在执行所请求的任务期间,数字助理有时在很长时间段内在涉及多次信息交换的持续对话中与用户进行交互。存在与数字助理进行交互以请求信息或执行各种任务的许多其他方法。除提供言语响应并采取经编程的动作之外,数字助理还提供其他视频或音频形式的响应,例如作为文本、警报、音乐、视频、动画等。

[0035] 如图1所示,在一些示例中,数字助理根据客户端-服务器模型来实现。数字助理包括在用户设备104上执行的客户端侧部分102(后文称作“DA客户端102”)以及在服务器系统108上执行的服务器侧部分106(后文称作“DA服务器106”)。DA客户端102通过一个或多个网络110与DA服务器106通信。DA客户端102提供客户端侧功能,诸如面向用户的输入和输出处理,以及与DA服务器106通信。DA服务器106为各自位于相应用户设备104上的任意数量的DA客户端102提供服务器侧功能。

[0036] 在一些示例中,DA服务器106包括面向客户端的I/O接口112、一个或多个处理模块114、数据与模型116以及到外部服务的I/O接口118。面向客户端的I/O接口112有利于DA服务器106的面向客户端的输入和输出处理。一个或多个处理模块114利用数据与模型116来处理语音输入,并基于自然语言输入来确定用户意图。此外,一个或多个处理模块114基于推断出的用户意图来执行任务执行。在一些示例中,DA服务器106通过一个或多个网络110与外部服务120通信以完成任务或采集信息。到外部服务的I/O接口118促成此类通信。

[0037] 用户设备104可以是任何合适的电子设备。在一些示例中,用户设备104是便携式多功能设备(例如,下文参考图2A所述的设备200)、多功能设备(例如,下文参考图4所述的设备400)或个人电子设备(例如,下文参考图6A至图6B所述的设备600)。便携式多功能设备是例如还包含其他功能(诸如,PDA和/或音乐播放器功能)的移动电话。便携式多功能设备的特定示例包括来自Apple Inc. (Cupertino, California)的Apple Watch<sup>®</sup>、iPhone<sup>®</sup>、iPod Touch<sup>®</sup>和iPad<sup>®</sup>设备。便携式多功能设备的其他示例包括但不限于耳塞式耳机/头戴式耳机、扬声器以及膝上型电脑或平板电脑。此外,在一些示例中,用户设备104是非便携式多功能设备。具体地讲,用户设备104是台式计算机、游戏机、扬声器、电视或电视机顶盒。在一些示例中,用户设备104包括触敏表面(例如,触摸屏显示器和/或触控板)。此外,用户设备104任选地包括一个或多个其他物理用户接口设备,诸如物理键盘、鼠标和/或操纵杆。下文更详细地描述了电子设备诸如多功能设备的各种示例。

[0038] 一个或多个通信网络110的示例包括局域网(LAN)和广域网(WAN),例如互联网。一个或多个通信网络110使用任何已知的网络协议来实现,包括各种有线或无线协议,诸如以太网、通用串行总线(USB)、火线(FIREWIRE)、全球移动通信系统(GSM)、增强型数据GSM环境(EDGE)、码分多址(CDMA)、时分多址(TDMA)、蓝牙、Wi-Fi、互联网协议语音(VoIP)、Wi-MAX或任何其他合适的通信协议。

[0039] 服务器系统108在一个或多个独立式数据处理装置或分布式计算机网络上实现。在一些示例中,服务器系统108还采用第三方服务提供方(例如,第三方云服务提供方)的各种虚拟设备和/或服务来提供服务器系统108的潜在计算资源和/或基础结构资源。

[0040] 在一些示例中,用户设备104经由第二用户设备122与DA服务器106 通信。第二用户设备122与用户设备104相似或相同。例如,第二用户设备122类似于下文参考图2A、图4以及图6A至图6B所述的设备200、400或600。用户设备104被配置为经由直接通信连接(诸如,蓝牙、NFC、BTLE等)或者经由有线或无线网络(诸如,局域Wi-Fi网络)而通信耦接到第二用户设备122。在一些示例中,第二用户设备122被配置为充当用户设备104与DA服务器106之间的代理。例如,用户设备104的DA 客户端102被配置为经由第二用户设备122向DA服务器106传输信息(例如,在用户设备104处接收的用户请求)。DA服务器106处理该信息,并经由第二用户设备122将相关数据(例如,响应于用户请求的数据内容) 返回到用户设备104。

[0041] 在一些示例中,用户设备104可被配置为将针对数据的缩略请求传送到第二用户设备122,以减少从用户设备104传输的信息量。第二用户设备 122被配置为确定添加到缩略请求的补充信息,以生成完整的请求来传输到 DA服务器106。该系统架构可有利地通过使用具有较强通信能力和/或电池电力的第二用户设备122(例如,移动电话、膝上型计算机、平板电脑等) 作为到DA服务器106的代理,允许具有有限通信能力和/或有限电池电力的用户设备104(例如,手表或类似的紧凑型电子设备) 访问DA服务器 106提供的服务。虽然图1中仅示出两个用户设备104和122,但应当理解,在一些示例中,系统100可包括在此代理配置中被配置为与DA服务器系统106通信的任意数量和类型的用户设备。

[0042] 虽然图1中所示的数字助理包括客户端侧部分(例如,DA客户端 102)和服务器侧部分(例如,DA服务器106)两者,但在一些示例中,数字助理的功能被实现为安装在用户设备上的独立式应用程序。此外,数字助理的客户端部分与服务器部分之间的功能划分在不同的具体实施中可变化。例如,在一些示例中,DA客户端为仅提供面向用户的输入和输出处理功能,并将数字助理的所有其他功能委派给后端服务器的瘦客户端。

## [0043] 2. 电子设备

[0044] 现在将注意力转至用于实现数字助理的客户端侧部分的电子设备的实施方案。图2A是示出根据一些实施方案的具有触敏显示器系统212的便携式多功能设备200的框图。有时为了方便将触敏显示器212称为“触摸屏”,并且有时将其称为或叫做“触敏显示器系统”。设备200包括存储器202(其任选地包括一个或多个计算机可读存储介质)、存储器控制器222、一个或多个处理单元(CPU) 220、外围设备接口218、RF电路208、音频电路210、扬声器211、麦克风213、输入/输出(I/O) 子系统206、其他输入控制设备216和外部端口224。设备200任选地包括一个或多个光学传感器264。设备200任选地包括用于检测设备200(例如,触敏表面诸如设备 200的触敏显示器系统212)上的接触的强度的一个或多个接触强度传感器 265。设备200任选地包括用于在设备200上生成触觉输出的一个或多个触觉输出发生器 267(例如,在触敏表面(诸如,设备200的触敏显示器系统 212或设备400的触摸板455)上生成触觉输出)。这些部件任选地通过一个或多个通信总线或信号线203进行通信。

[0045] 如在本说明书和权利要求书中所使用的那样,触敏表面上的接触的“强度”这一术语是指触敏表面上的接触(例如,手指接触)的力或压力(每单位面积的力),或者是指触敏表面上的接触的力或压力的替代物(代用物)。接触的强度具有值范围,该值范围包括至少四个不同的值并且更典型地包括上百个不同的值(例如,至少256个)。接触的强度任选地使用各种方法和各种传感器或传感器的组合来确定(或测量)。例如,在触敏表面下方或相邻于触敏表面的一个或多个力传感器任选地用于测量触敏表面上的不同点处的力。在一些具

体实施中,来自多个力传感器的力测量值被组合(例如,加权平均)以确定所估计的接触力。类似地,触笔的压敏顶端任选地用于确定触笔在触敏表面上的压力。另选地,在触敏表面上检测到的接触区域的大小和/或其变化、接触附近的触敏表面的电容和/或其变化以及/或者接触附近的触敏表面的电阻和/或其变化任选地被用作触敏表面上的接触的力或压力的替代物。在一些具体实施中,接触力或压力的替代物测量直接用于确定是否已经超过强度阈值(例如,强度阈值以对应于替代物测量的单位来描述)。在一些具体实施中,接触力或压力的替代物测量被转换成估计的力或压力,并且估计的力或压力用于确定是否已超过强度阈值(例如,强度阈值是以压力的单位进行测量的压力阈值)。使用接触的强度作为用户输入的属性,从而允许用户访问用户在实地面积有限的尺寸更小的设备上本来不可访问的附加设备功能,该尺寸更小的设备用于(例如,在触敏显示器上)显示能表示和/或接收用户输入(例如,经由触敏显示器、触敏表面或物理控件/机械控件,诸如旋钮或按钮)。

[0046] 如本说明书和权利要求书中所使用的那样,术语“触觉输出”是指将由用户通过用户的触摸感检测到的设备相对于设备的先前位置的物理位移、设备的部件(例如,触敏表面)相对于设备的另一个部件(例如,外壳)的物理位移或部件相对于设备的质心的位移。例如,在设备或设备的部件与用户对触摸敏感的表面(例如,手指、手掌或用户手部的其他部分)接触的情况下,通过物理位移生成的触觉输出将由用户解释为触感,该触感对应于设备或设备的部件的物理特征的所感知的变化。例如,触敏表面(例如,触敏显示器或触控板)的移动任选地由用户解释为对物理致动按钮的“按下点击”或“松开点击”。在一些情况下,用户将感觉到触感,诸如“按下点击”或“松开点击”,即使在通过用户的移动而物理地被按压(例如,被移位)的与触敏表面相关联的物理致动按钮没有移动时。又如,即使在触敏表面的光滑度无变化时,触敏表面的移动也会任选地由用户解释或感测为触敏表面的“粗糙度”。虽然用户对触摸的此类解释将受到用户的个体化感官知觉的限制,但是对触摸的许多感官知觉是大多数用户共有的。因此,当触觉输出被描述为对应于用户的特定感官知觉(例如,“按下点击”、“松开点击”、“粗糙度”)时,除非另外陈述,否则所生成的触觉输出对应于设备或其部件的物理位移,该物理位移将会生成典型(或普通)用户的所述感官知觉。

[0047] 应当理解,设备200仅为便携式多功能设备的一个示例,并且设备200任选地具有比所示出的更多或更少的部件,任选地组合两个或更多个部件,或者任选地具有这些部件的不同配置或布置。图2A中所示的各种部件以硬件、软件或硬件与软件两者的组合来实现,这包括一个或多个信号处理和/或专用集成电路。

[0048] 存储器202包括一个或多个计算机可读存储介质。这些计算机可读存储介质例如为有形的和非暂态的。存储器202包括高速随机存取存储器,并且还包括非易失性存储器,诸如一个或多个磁盘存储设备、闪存存储设备或其他非易失性固态存储设备。存储器控制器222控制设备200的其他部件访问存储器202。

[0049] 在一些示例中,存储器202的非暂态计算机可读存储介质用于存储指令(例如,用于执行下文所述的过程的各方面)以供指令执行系统、装置或设备诸如基于计算机的系统、包含处理器的系统或可从指令执行系统、装置或设备取出指令并执行指令的其他系统使用或与其结合使用。在其他示例中,指令(例如,用于执行下文描述的过程的各方面)存储在服务器系统108的非暂态计算机可读存储介质(未示出)上,或在存储器202的非暂态计算机可读存储介质与服务器系统108的非暂态计算机可读存储介质之间划分。

[0050] 外围设备接口218用于将设备的输入和输出外围设备耦接到CPU 220 和存储器202。该一个或多个处理器220运行或执行存储器202中所存储的各种软件程序和/或指令集以执行设备200的各种功能并处理数据。在一些实施方案中,外围设备接口218、CPU 220和存储器控制器222在单个芯片诸如芯片204上实现。在一些其他实施方案中,它们在独立的芯片上实现。

[0051] RF(射频)电路208接收和发送也被称为电磁信号的RF信号。Rf电路208将电信号转换为电磁信号/将电磁信号转换为电信号,并且经由电磁信号与通信网络及其他通信设备进行通信。RF电路208任选地包括用于执行这些功能的熟知的电路,包括但不限于天线系统、RF收发器、一个或多个放大器、调谐器、一个或多个振荡器、数字信号处理器、编解码芯片组、用户身份模块(SIM)卡、存储器等等。RF电路208任选地通过无线通信与网络以及其他设备进行通信,该网络为诸如互联网(也被称为万维网(WWW))、内联网和/或无线网络(诸如蜂窝电话网络、无线局域网(LAN)和/或城域网(MAN))。RF电路208任选地包括用于检测近场通信(NFC)场的熟知的电路,诸如通过近程通信无线电部件来进行检测。无线通信任选地使用多种通信标准、协议和技术中的任一种,包括但不限于全球移动通信系统(GSM)、增强型数据GSM环境(EDGE)、高速下行链路分组接入(HSDPA)、高速上行链路分组接入(HSUPA)、演进、纯数据(EV-DO)、HSPA、HSPA+、双小区HSPA(DC-HSPDA)、长期演进(LTE)、近场通信(NFC)、宽带码分多址(W-CDMA)、码分多址(CDMA)、时分多址(TDMA)、蓝牙、蓝牙低功耗(BTLE)、无线保真(Wi-Fi)(例如,IEEE 802.11a、IEEE 802.11b、IEEE 802.11g、IEEE 802.11n和/或IEEE 802.11ac)、互联网协议语音(VoIP)、Wi-MAX、电子邮件协议(例如,互联网消息访问协议(IMAP)和/或邮局协议(POP))、即时消息(例如,可扩展消息处理和存在协议(XMPP)、用于即时消息和存在利用扩展的会话发起协议(SIMPLE)、即时消息和存在服务(IMPS))和/或短消息服务(SMS),或者任何其他适当的通信协议,包括在本文档提交日期时尚未开发出的通信协议。

[0052] 音频电路210、扬声器211和麦克风213提供用户与设备200之间的音频接口。音频电路210从外围设备接口218接收音频数据,将音频数据转换为电信号,并且将电信号传输到扬声器211。扬声器211将电信号转换为人耳可听见的声波。音频电路210还接收由麦克风213根据声波转换来的电信号。音频电路210将电信号转换为音频数据,并将音频数据传输到外围设备接口218以用于处理。音频数据通过外围设备接口218检索自和/或传输至存储器202和/或RF电路208。在一些实施方案中,音频电路210还包括耳麦插孔(例如,图3中的312)。该耳麦插孔提供音频电路210与可移除的音频输入/输出外围设备之间的接口,该可移除的音频输入/输出外围设备诸如仅输出的耳机或者具有输出(例如,单耳耳机或双耳耳机)和输入(例如,麦克风)两者的耳麦。

[0053] I/O子系统206将设备200上的输入/输出外围设备诸如触摸屏212和其他输入控制设备216耦接到外围设备接口218。I/O子系统206任选地包括显示控制器256、光学传感器控制器258、强度传感器控制器259、触觉反馈控制器261,以及用于其他输入或控制设备的一个或多个输入控制器 260。该一个或多个输入控制器260从其他输入控制设备216接收电信号/将电信号发送到该其他输入控制设备。其他输入控制设备216任选地包括物理按钮(例如,下压按钮、摇臂按钮等)、拨号盘、滑动开关、操纵杆、点击轮等。在一些另选实施方案中,一个或多个输入控制器260任选地耦接到以下各项中的任一者(或不耦接到以下各项中的

任一者):键盘、红外线端口、USB端口以及指针设备诸如鼠标。该一个或多个按钮(例如,图3中的308)任选地包括用于扬声器211和/或麦克风213的音量控制的增大/减小按钮。该一个或多个按钮任选地包括下压按钮(例如,图3中的 306)。

[0054] 快速按下压按钮会解除对触摸屏212的锁定或者开始使用触摸屏上的手势来对设备进行解锁的过程,如在2005年12月23日提交的名称为“Unlocking a Device by Performing Gestures on an Unlock Image”的美国专利No.7,657,849的美国专利申请11/322,549中所述的那样,该美国专利申请据此全文以引用方式并入。较长地按下压按钮(例如,306)使设备 200开机或关机。用户能够自定义一个或多个按钮的功能。触摸屏212用于实现虚拟按钮或软按钮以及一个或多个软键盘。

[0055] 触敏显示器212提供设备和用户之间的输入接口和输出接口。显示控制器256从触摸屏212接收电信号并且/或者将电信号发送到该触摸屏。触摸屏212向用户显示视觉输出。视觉输出包括图形、文本、图标、视频及其任何组合(统称为“图形”)。在一些实施方案中,一些视觉输出或全部视觉输出对应于用户界面对象。

[0056] 触摸屏212具有基于触觉和/或触感接触来接受来自用户的输入的触敏表面、传感器或传感器组。触摸屏212和显示控制器256(与存储器202中的任何相关联的模块和/或指令集一起)检测触摸屏212上的接触(和该接触的任何移动或中断),并且将所检测到的接触转换为与被显示在触摸屏 212上的用户界面对象(例如,一个或多个软键、图标、网页或图像)的交互。在示例性实施方案中,触摸屏212和用户之间的接触点与用户的手指对应。

[0057] 触摸屏212使用LCD(液晶显示器)技术、LPD(发光聚合物显示器)技术或LED(发光二极管)技术,但在其他实施方案中可使用其他显示技术。触摸屏212和显示控制器256使用目前已知或以后将开发的多种触摸感测技术中的任何技术,以及其他接近传感器阵列或用于确定与触摸屏212接触的一个或多个点的其他元件来检测接触及其任何移动或中断,所述多种触摸感测技术包括但不限于电容式、电阻式、红外和表面声波技术。在一个示例性实施方案中,使用投射式互电容感测技术,诸如在来自 Apple Inc. (Cupertino, California)的iPhone®和iPod Touch®中使用的技术。

[0058] 在一些实施方案中,触摸屏212的触敏显示器类似于以下美国专利: 6,323,846 (Westerman等人)、6,570,557 (Westerman等人)和/或 6,677,932 (Westerman)和/或美国专利公开2002/0015024A1中所述的多触敏触摸板,这些专利申请均据此全文以引用方式并入。然而,触摸屏212 显示来自设备200的视觉输出,而触敏触摸板不提供视觉输出。

[0059] 在一些实施方案中,触摸屏212的触敏显示器如以下申请中所述:(1) 2006年5月2日提交的名称为“Multipoint Touch Surface Controller”的美国专利申请No.11/381,313;(2) 2004年5月6日提交的名称为“Multipoint Touchscreen”的美国专利申请No.10/840,862;(3) 2004年7月30日提交的名称为“Gestures For Touch Sensitive Input Devices”的美国专利申请No. 10/903,964;(4) 提交于2005年1月31日的名称为“Gestures For Touch Sensitive Input Devices”的美国专利申请No.11/048,264;(5) 2005年1月 18日提交的名称为“Mode-Based Graphical User Interfaces For Touch Sensitive Input Devices”的美国专利申请No.11/038,590;(6) 2005年9月 16日提交的名称为“Virtual Input Device Placement On A Touch Screen User Interface”的美国专利申请No.11/228,758;(7) 2005年9月16日提交的名称为“Operation Of A Computer With A Touch

Screen Interface”的美国专利申请No.11/228,700; (8) 2005年9月16日提交的名称为“Activating Virtual Keys Of A Touch-Screen Virtual Keyboard”的美国专利申请No. 11/228,737; 以及 (9) 2006年3月3日提交的名称为“Multi-Functional Hand- Held Device”的美国专利申请No.11/367,749。所有这些申请全文以引用方式并入本文。

[0060] 触摸屏212例如具有超过100dpi的视频分辨率。在一些实施方案中,触摸屏具有约160dpi的视频分辨率。用户使用任何合适的对象或附加物诸如触笔、手指等与触摸屏212进行接触。在一些实施方案中,将用户界面设计为主要通过基于手指的接触和手势来工作,由于手指在触摸屏上的接触区域较大,因此这可能不如基于触笔的输入精确。在一些实施方案中,设备将基于手指的粗略输入转化为精确的指针/光标位置或命令以用于执行用户所期望的动作。

[0061] 在一些实施方案中,除触摸屏之外,设备200还包括用于激活或去激活特定功能的触摸板(未示出)。在一些实施方案中,触控板是设备的触敏区域,与触摸屏不同,该触敏区域不显示视觉输出。触控板是与触摸屏 212分开的触敏表面,或者是由触摸屏形成的触敏表面的延伸。

[0062] 设备200还包括用于为各种部件供电的电力系统262。电力系统262包括电力管理系统、一个或多个电源(例如,电池、交流电(AC))、再充电系统、电力故障检测电路、功率转换器或逆变器、电力状态指示器(例如,发光二极管(LED))和与便携式设备中电力的生成、管理和分配相关联的任何其他部件。

[0063] 设备200还包括一个或多个光学传感器264。图2A示出了耦接到I/O 子系统206中的光学传感器控制器258的光学传感器。光学传感器264包括电荷耦合器件(CCD)或互补金属氧化物半导体(CMOS)光电晶体管。光学传感器264从环境接收通过一个或多个透镜而投射的光,并且将光转换为表示图像的数据。结合成像模块243(也叫做相机模块),光学传感器264 捕获静态图像或视频。在一些实施方案中,光学传感器位于设备200的后部,与设备前部的触摸屏显示器212相背对,使得触摸屏显示器被用作用于静态图像和/或视频图像采集的取景器。在一些实施方案中,光学传感器位于设备的前部,使得在用户在触摸屏显示器上查看其他视频会议参与者的同时获取该用户的图像以用于视频会议。在一些实施方案中,光学传感器264的位置可由用户改变(例如,通过旋转设备外壳中的透镜和传感器),使得单个光学传感器264与触摸屏显示器一起使用,以用于视频会议以及静态图像和/或视频图像采集两者。

[0064] 设备200任选地还包括一个或多个接触强度传感器265。图2A示出了耦接到I/O子系统206中的强度传感器控制器259的接触强度传感器。接触强度传感器265任选地包括一个或多个压阻应变仪、电容式力传感器、电气力传感器、压电力传感器、光学力传感器、电容式触敏表面或其他强度传感器(例如,用于测量触敏表面上的接触的力(或压力)的传感器)。接触强度传感器265从环境接收接触强度信息(例如,压力信息或压力信息的代用物)。在一些实施方案中,至少一个接触强度传感器与触敏表面(例如,触敏显示器系统212)并置排列或邻近。在一些实施方案中,至少一个接触强度传感器位于与位于设备200的前部上的触摸屏显示器212相背对的设备200的后部上。

[0065] 设备200还包括一个或多个接近传感器266。图2A示出了耦接到外围设备接口218的接近传感器266。另选地,接近传感器266耦接到I/O子系统206中的输入控制器260。接近



传感器266如以下美国专利申请中所述的那样执行:No.11/241,839,名称为“Proximity Detector In Handheld Device”;No.11/240,788,名称为“Proximity Detector In Handheld Device”;No.11/620,702,名称为“Using Ambient Light Sensor To Augment Proximity Sensor Output”;No.11/586,862,名称为“Automated Response To And Sensing Of User Activity In Portable Devices”;以及No. 11/638,251,名称为“Methods And Systems For Automatic Configuration Of Peripherals”,这些美国专利申请据此全文以引用方式并入本文。在一些实施方案中,当多功能设备被放置在用户的耳朵附近时(例如,当用户正在打电话时),接近传感器关闭并且禁用触摸屏212。

[0066] 设备200任选地还包括一个或多个触觉输出发生器267。图2A示出了耦接到I/O子系统206中的触觉反馈控制器261的触觉输出发生器。触觉输出发生器267任选地包括一个或多个电声设备(诸如,扬声器或其他音频部件)和/或用于将能量转换成线性运动的机电设备(诸如,电机、螺线管、电活性聚合器、压电致动器、静电致动器或其他触觉输出生成部件(例如,用于将电信号转换成设备上的触觉输出的部件))。接触强度传感器265从触觉反馈模块233接收触觉反馈生成指令,并且在设备200上生成能够由设备200的用户感觉到的触觉输出。在一些实施方案中,至少一个触觉输出发生器与触敏表面(例如,触敏显示器系统212)并置排列或邻近,并且任选地通过竖直地(例如,向设备200的表面内/外)或侧向地(例如,在与设备200的表面相同的平面中向后和向前)移动触敏表面来生成触觉输出。在一些实施方案中,至少一个触觉输出发生器传感器位于设备200的后部上,与位于设备200的前部上的触摸屏显示器212相对。

[0067] 设备200还包括一个或多个加速度计268。图2A示出了耦接至外围设备接口218的加速度计268。另选地,加速度计268耦接至I/O子系统206中的输入控制器260。加速度计268如以下美国专利公开中所述那样执行:美国专利公开20050190059,“Acceleration-based Theft Detection System for Portable Electronic Devices”和美国专利公开20060017692,“Methods And Apparatuses For Operating A Portable Device Based On An Accelerometer”,这两个美国专利公开全文以引用方式并入本文。在一些实施方案中,基于对从一个或多个加速度计接收的数据的分析来在触摸屏显示器上以纵向视图或横向视图显示信息。除加速度计268之外,设备200还任选地包括磁力计(未示出)和GPS(或GLONASS或其他全球导航系统)接收器(未示出),以用于获取关于设备200的位置和取向(例如,纵向或横向)的信息。

[0068] 在一些实施方案中,存储在存储器202中的软件部件包括操作系统 226、通信模块(或指令集)228、接触/运动模块(或指令集)230、图形模块(或指令集)232、文本输入模块(或指令集)234、全球定位系统(GPS)模块(或指令集)235、数字助理客户端模块229以及应用程序(或指令集)236。此外,存储器202存储数据与模型,诸如用户数据与模型231。此外,在一些实施方案中,存储器202(图2A)或470(图4)存储设备/全局内部状态257,如图2A、和图4中所示。设备/全局内部状态257包括以下中的一者或多者:活动应用程序状态,其指示哪些应用程序(如果有的话)当前是活动的;显示状态,用于指示什么应用程序、视图或其它信息占据触摸屏显示器212的各个区域;传感器状态,包括从设备的各个传感器和输入控制设备216获得的信息;以及关于设备的位置和/或姿态的位置信息。

[0069] 操作系统226(例如,Darwin、RTXC、LINUX、UNIX、OS X、iOS、WINDOWS或嵌入式操作



系统诸如VxWorks)包括用于控制和管理一般系统任务(例如,存储器管理、存储设备控制、功率管理等)的各种软件部件和/或驱动器,并且促进各种硬件部件和软件部件之间的通信。

[0070] 通信模块228有利于通过一个或多个外部端口224来与其他设备进行通信,并且还用于处理由RF电路208和/或外部端口224所接收的数据的各种软件部件。外部端口224(例如,通用串行总线(USB)、火线等)适于直接耦接到其他设备,或间接地通过网络(例如,互联网、无线LAN等)进行耦接。在一些实施方案中,外部端口是与iPod<sup>®</sup>(Apple Inc.的商标)设备上所使用的30针连接器相同的或类似的和/或与其兼容的多针(例如,30针)连接器。

[0071] 接触/运动模块230任选地检测与触摸屏212(结合显示控制器256)和其他触敏设备(例如,触摸板或物理点击式转盘)的接触。接触/运动模块230包括各种软件部件以用于执行与接触检测相关的各种操作,诸如确定是否已发生接触(例如,检测手指按下事件)、确定接触强度(例如,接触的力或压力,或者接触的力或压力的替代物)、确定是否存在接触的移动并跟踪在触敏表面上的移动(例如,检测一个或多个手指拖动事件),以及确定接触是否已停止(例如,检测手指抬起事件或接触断开)。接触/运动模块230从触敏表面接收接触数据。确定接触点的移动任选地包括确定接触点的速率(量值)、速度(量值和方向)和/或加速度(量值和/或方向的改变),所述接触点的移动由一系列接触数据表示。这些操作任选地被应用于单点接触(例如,单指接触)或者多点同时接触(例如,“多点触摸”/多个手指接触)。在一些实施方案中,接触/运动模块230和显示控制器256检测触摸板上的接触。

[0072] 在一些实施方案中,接触/运动模块230使用一组一个或多个强度阈值来确定操作是否已由用户执行(例如,来确定用户是否已“点击”图标)。在一些实施方案中,根据软件参数来确定强度阈值的至少一个子集(例如,强度阈值不是由具体物理致动器的激活阈值来确定的,并且可在不改变设备200的物理硬件的情况下来调整)。例如,在不改变触控板或触摸屏显示器硬件的情况下,触控板或触摸屏的鼠标“点击”阈值可被设定成预定义的阈值的大范围中的任一个阈值。另外,在一些具体实施中,向设备的用户提供用于调节一组强度阈值中的一个或多个强度阈值(例如,通过调节各个强度阈值和/或通过利用对“强度”参数的系统级点击来一次调节多个强度阈值)的软件设置。

[0073] 接触/运动模块230任选地检测用户的手势输入。触敏表面上的不同手势具有不同的接触模式(例如,所检测到的接触的不同运动、计时和/或强度)。因此,任选地通过检测特定接触模式来检测手势。例如,检测手指轻击手势包括检测手指按下事件,然后在与手指按下事件相同的位置(或基本上相同的位置)处(例如,在图标的位置处)检测手指抬起(抬离)事件。作为另一个示例,在触敏表面上检测手指轻扫手势包括检测手指按下事件,然后检测一个或多个手指拖动事件,并且随后检测手指抬起(抬离)事件。

[0074] 图形模块232包括用于在触摸屏212或其他显示器上呈现和显示图形的各种已知的软件部件,包括用于改变所显示的图形的视觉冲击(例如,亮度、透明度、饱和度、对比度或其他视觉属性)的部件。如本文所用,术语“图形”包括可被显示给用户的任何对象,非限制性地包括文本、网页、图标(诸如,包括软键的用户界面对象)、数字图像、视频、动画等。

[0075] 在一些实施方案中,图形模块232存储表示要使用的图形的数据。每个图形任选地被分配有对应的代码。图形模块232从应用程序等接收用于指定待显示的图形的一个或多

个代码,在必要的情况下还一起接收坐标数据和其他图形属性数据,并且然后生成屏幕图像数据,以输出到显示控制器256。

[0076] 触感反馈模块233包括用于生成指令的各种软件部件,这些指令由一个或多个触觉输出发生器267使用以便响应于用户和设备200的交互而在设备200上的一个或多个位置处产生触觉输出。

[0077] 在一些示例中作为图形模块232的部件的文本输入模块234提供用于在各种应用程序(例如,联系人237、电子邮件240、IM 241、浏览器247 和需要文本输入的任何其他应用程序)中输入文本的软键盘。

[0078] GPS模块235确定设备的位置,并提供该信息以供在各种应用程序中使用(例如提供给电话238以供在基于位置的拨号中使用;提供给相机243 作为图片/视频元数据;以及提供给提供基于位置的服务的应用,诸如天气桌面小程序、本地黄页桌面小程序和地图/导航桌面小程序)。

[0079] 数字助理客户端模块229包括各种客户端侧数字助理指令,以提供数字助理的客户端侧功能。例如,数字助理客户端模块229能够通过便携式多功能设备200的各种用户接口(例如,麦克风213、加速度计268、触敏显示器系统212、光学传感器229、其他输入控制设备216等)接受声音输入(例如,语音输入)、文本输入、触摸输入和/或手势输入。数字助理客户端模块229还能够通过便携式多功能设备200的各种输出接口(例如,扬声器211、触敏显示器系统212、触觉输出生成器267等)提供音频形式的输出(例如,语音输出)、视觉形式的输出和/或触觉形式的输出。例如,将输出提供为语音、声音、警报、文本消息、菜单、图形、视频、动画、振动和/或以上两者或更多者的组合。在操作期间,数字助理客户端模块229使用RF电路208与DA服务器106通信。

[0080] 用户数据与模型231包括与用户相关联的各种数据(例如,用户特定的词汇数据、用户偏好数据、用户指定的名称发音、来自用户电子通讯录的数据、待办事项、购物清单等)以提供数字助理的客户端侧功能。此外,用户数据与模型231包括用于处理用户输入并且确定用户意图的各种模型(例如,语音识别模型、统计语言模型、自然语言处理模型、知识本体、任务流模型、服务模型等)。

[0081] 在一些示例中,数字助理客户端模块229利用便携式多功能设备200 的各种传感器、子系统和外围设备来从便携式多功能设备200的周围环境采集另外信息,以建立与用户、当前用户交互和/或当前用户输入相关联的上下文。在一些示例中,数字助理客户端模块229将上下文信息或其子集与用户输入一起提供至DA服务器106以帮助推断用户意图。在一些示例中,数字助理还使用上下文信息来确定如何准备输出并将其传送给用户。上下文信息被称为上下文数据。

[0082] 在一些示例中,伴随用户输入的上下文信息包括传感器信息,例如照明、环境噪声、环境温度、周围环境的图像或视频等。在一些示例中,上下文信息还可包括设备的物理状态,例如设备取向、设备位置、设备温度、功率电平、速度、加速度、运动模式、蜂窝信号强度等。在一些示例中,与DA服务器106的软件状态相关的信息,例如便携式多功能设备200的运行过程、已安装程序、过去和当前的网络活动、后台服务、错误日志、资源使用等,作为与用户输入相关联的上下文信息提供给DA服务器 106。

[0083] 在一些示例中,数字助理客户端模块229响应于来自DA服务器106 的请求而选择

性地提供存储在便携式多功能设备200上的信息(例如,用户数据231)。在一些示例中,数字助理客户端模块229还在DA服务器 106请求时引来自用户经由自然语言对话或其他用户接口的附加输入。数字助理客户端模块229将该附加输入传送至DA服务器106,以帮助DA服务器106进行意图推断和/或满足在用户请求中表达的用户意图。

[0084] 下文参考图7A至图7C对数字助理进行更详细的描述。应当认识到,数字助理客户端模块229可包括下文所述的数字助理模块726的任意数量的子模块。

[0085] 应用程序236包括以下模块(或指令集)或者其子集或超集:

- [0086] • 联系人模块237(有时称为通讯录或联系人列表);
- [0087] • 电话模块238;
- [0088] • 视频会议模块239;
- [0089] • 电子邮件客户端模块240;
- [0090] • 即时消息(IM)模块241;
- [0091] • 健身支持模块242;
- [0092] • 用于静态图像和/或视频图像的相机模块243;
- [0093] • 图像管理模块244;
- [0094] • 视频播放器模块;
- [0095] • 音乐播放器模块;
- [0096] • 浏览器模块247;
- [0097] • 日历模块248;
- [0098] • 桌面小程序模块249,其在一些示例中包括以下各项中的一者或多者:天气桌面小程序249-1、股票桌面小程序249-2、计算器桌面小程序249-3、闹钟桌面小程序249-4、词典桌面小程序249-5和用户获取的其他桌面小程序以及用户创建的桌面小程序249-6;
- [0099] • 用于形成用户创建的桌面小程序249-6的桌面小程序创建器模块 250;
- [0100] • 搜索模块251;
- [0101] • 视频和音乐播放器模块252,其合并视频播放器模块和音乐播放器模块;
- [0102] • 记事本模块253;
- [0103] • 地图模块254;和/或
- [0104] • 在线视频模块255。

[0105] 存储在存储器202中的其他应用程序236的示例包括其他文字处理应用程序、其他图像编辑应用程序、绘图应用程序、呈现应用程序、支持 JAVA的应用程序、加密、数字版权管理、声音识别和声音复制。

[0106] 结合触摸屏212、显示控制器256、接触/运动模块230、图形模块232 和文本输入模块234,联系人模块237用于管理通讯录或联系人列表(例如,存储在存储器202或存储器470中的联系人模块237的应用程序内部状态292中),包括:将一个或多个姓名添加到通讯录;从地址簿删除姓名;将电话号码、电子邮件地址、物理地址或其他信息与姓名关联;将图像与姓名关联;对姓名进行归类和分类;提供电话号码或电子邮件地址来发起和/或促进通过电话238、视频会议模块239、电子邮件240或即时消息 241的通信;等等。

[0107] 结合RF电路208、音频电路210、扬声器211、麦克风213、触摸屏 212、显示控制器256、接触/运动模块230、图形模块232和文本输入模块 234,电话模块238用于输入对应于

电话号码的字符序列、访问联系人模块 237 中的一个或多个电话号码、修改已经输入的电话号码、拨打相应的电话号码、进行会话以及当会话完成时断开或挂断。如上所述,无线通信使用多种通信标准、协议和技术中的任一种。

[0108] 结合RF电路208、音频电路210、扬声器211、麦克风213、触摸屏 212、显示控制器256、光学传感器264、光学传感器控制器258、接触/运动模块230、图形模块232、文本输入模块234、联系人模块237和电话模块238,视频会议模块239包括根据用户指令来发起、进行和终止用户与一个或多个其他参与方之间的视频会议的可执行指令。

[0109] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232和文本输入模块234,电子邮件客户端模块240包括响应于用户指令来创建、发送、接收和管理电子邮件的可执行指令。结合图像管理模块244,电子邮件客户端模块240使得非常容易创建和发送具有由相机模块243拍摄的静态图像或视频图像的电子邮件。

[0110] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232和文本输入模块234,即时消息模块241包括用于以下操作的可执行指令:输入与即时消息对应的字符序列、修改先前输入的字符、传输相应即时消息(例如,使用短消息服务(SMS)或多媒体消息服务(MMS)协议以用于基于电话的即时消息或者使用XMPP、SIMPLE或IMPS 以用于基于互联网的即时消息)、接收即时消息以及查看所接收的即时消息。在一些实施方案中,所传输和/或接收的即时消息包括图形、照片、音频文件、视频文件和/或如MMS和/或增强型消息服务(EMS)中支持的其他附件。如本文所用,“即时消息”是指基于电话的消息(例如,使用SMS 或MMS发送的消息)和基于互联网的消息(例如,使用XMPP、SIMPLE 或IMPS发送的消息)两者。

[0111] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232、文本输入模块234、GPS模块235、地图模块254和音乐播放器模块,健身支持模块242包括用于以下操作的可执行指令:创建健身(例如,具有时间、距离和/或卡路里燃烧目标);与健身传感器(运动设备)进行通信;接收健身传感器数据;校准用于监视健身的传感器;为健身选择和播放音乐;以及显示、存储和传输健身数据。

[0112] 结合触摸屏212、显示控制器256、一个或多个光学传感器264、光学传感器控制器258、接触/运动模块230、图形模块232和图像管理模块 244,相机模块243包括用于进行以下操作的可执行指令:捕获静态图像或视频(包括视频流)并且将它们存储到存储器202中,修改静态图像或视频的特征,或从存储器202删除静态图像或视频。

[0113] 结合触摸屏212、显示控制器256、接触/运动模块230、图形模块 232、文本输入模块234和相机模块243,图像管理模块244包括用于进行以下操作的可执行指令:排列、修改(例如,编辑)、或以其他方式操控、加标签、删除、展示(例如,在数字幻灯片或相册中),以及存储静态图像和/或视频图像。

[0114] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232和文本输入模块234,浏览器模块247包括根据用户指令来浏览互联网(包括搜索、链接至、接收和显示网页或其部分,以及链接至网页的附件和其他文件)的可执行指令。

[0115] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232、文本输入模块234、电子邮件客户端模块240和浏览器模块247,日历模块248包括用于根据用户指令来创建、显示、修改和存储日历以及与日历相关联的数据(例如,日历条目、待办事

项等)的可执行指令。

[0116] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232、文本输入模块234和浏览器模块247,桌面小程序模块249是可由用户下载并使用的微型应用程序(例如,天气桌面小程序 249-1、股市桌面小程序249-2、计算器桌面小程序249-3、闹钟桌面小程序 249-4和字典桌面小程序249-5)或由用户创建的微型应用(例如,用户创建的桌面小程序249-6)。在一些实施方案中,桌面小程序包括HTML(超文本标记语言)文件、CSS(层叠样式表)文件和JavaScript文件。在一些实施方案中,桌面小程序包括XML(可扩展标记语言)文件和JavaScript 文件(例如,Yahoo!桌面小程序)。

[0117] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232、文本输入模块234和浏览器模块247,桌面小程序创建器模块250被用户用于创建桌面小程序(例如,使网页的用户指定部分变成桌面小程序)。

[0118] 结合触摸屏212、显示控制器256、接触/运动模块230、图形模块232 和文本输入模块234,搜索模块251包括用于根据用户指令来搜索存储器 202中匹配一个或多个搜索条件(例如,一个或多个用户指定的搜索词)的文本、音乐、声音、图像、视频和/或其他文件的可执行指令。

[0119] 结合触摸屏212、显示控制器256、接触/运动模块230、图形模块 232、音频电路210、扬声器211、RF电路208和浏览器模块247,视频和音乐播放器模块252包括允许用户下载和回放以一种或多种文件格式(诸如,MP3或AAC文件)存储的所录制的音乐和其他声音文件的可执行指令,以及用于显示、展示或以其他方式回放视频(例如,在触摸屏212上或在经由外部端口224连接的外部显示器上)的可执行指令。在一些实施方案中,设备200任选地包括MP3播放器诸如iPod(Apple Inc.的商标)的功能性。

[0120] 结合触摸屏212、显示控制器256、接触/运动模块230、图形模块232 和文本输入模块234,记事本模块253包括用于根据用户指令来创建和管理记事本、待办事项等的可执行指令。

[0121] 结合RF电路208、触摸屏212、显示控制器256、接触/运动模块 230、图形模块232、文本输入模块234、GPS模块235和浏览器模块247,地图模块254用于根据用户指令接收、显示、修改和存储地图以及与地图相关联的数据(例如,驾驶方向、与特定位置处或附近的商店及其他兴趣点有关的数据,以及其他基于位置的数据)。

[0122] 结合触摸屏212、显示控制器256、接触/运动模块230、图形模块 232、音频电路210、扬声器211、RF电路208、文本输入模块234、电子邮件客户端模块240和浏览器模块247,在线视频模块255包括允许用户访问、浏览、接收(例如,流式接收和/或下载)、回放(例如,在触摸屏上或在经由外部端口224连接的外部显示器上)、发送具有到特定在线视频的链接的电子邮件,以及以其他方式管理一种或多种文件格式(诸如,H.264)的在线视频的指令。在一些实施方案中,使用即时消息模块241而不是电子邮件客户端模块240来发送至特定在线视频的链接。在线视频应用程序的其他描述可在于2007年6月20日提交的名称为“Portable Multifunction Device,Method,and Graphical User Interface for Playing Online Videos”的美国临时专利申请60/936,562和于2007年12月31日提交的名称为“Portable Multifunction Device,Method,and Graphical User Interface for Playing Online Videos”的美国专利申请11/968,067中找到,这两个专利申请的内容据此

全文以引用方式并入本文。

[0123] 上述模块和应用程序中的每一者与用于执行上文所述的一种或多种功能以及在本申请中所述的方法(例如,本文中所述的计算机实现的方法和其他信息处理方法)的一组可执行指令对应。这些模块(例如,指令集)不必被实现为独立的软件程序、过程或模块,并因此在各种实施方案中可组合或以其他方式重新布置这些模块的各种子集。例如,视频播放器模块可与音乐播放器模块组合成单个模块(例如,图2A中的视频和音乐播放器模块252)。在一些实施方案中,存储器202存储上述模块和数据结构的子组。此外,存储器202存储上文未描述的另外的模块和数据结构。

[0124] 在一些实施方案中,设备200为该设备上的预定义的一组功能的操作唯一地通过触摸屏和/或触摸板来执行的设备。通过使用触摸屏和/或触控板作为用于设备200的操作的主要输入控制设备,减少设备200上的物理输入控制设备(诸如下压按钮、拨盘等)的数量。

[0125] 唯一地通过触摸屏和/或触摸板执行的该预定义的一组功能任选地包括用户界面之间的导航。在一些实施方案中,该触摸板在被用户触摸时将设备200从显示在设备200上的任何用户界面导航到主菜单、home菜单或根菜单。在此类实施方案中,使用触摸板来实现“菜单按钮”。在一些其他实施方案中,菜单按钮是物理下压按钮或者其他物理输入控制设备,而不是触摸板。

[0126] 图2B是示出了根据一些实施方案的用于事件处理的示例性部件的框图。在一些实施方案中,存储器202(图2A)或470(图4)包括事件分类器270(例如,在操作系统226中)以及相应的应用程序236-1(例如,前述应用程序237至251、255、480至490中的任一个应用程序)。

[0127] 事件分类器270接收事件信息并确定要将事件信息递送到的应用程序236-1和应用程序236-1的应用程序视图291。事件分类器270包括事件监视器271和事件分配器模块274。在一些实施方案中,应用程序236-1包括应用程序内部状态292,该应用程序内部状态指示当应用程序是活动的或正在执行时被显示在触敏显示器212上的当前应用程序视图。在一些实施方案中,设备/全局内部状态257被事件分类器270用于确定哪个(哪些)应用程序当前是活动的,并且应用程序内部状态292被事件分类器270用于确定要将事件信息递送到的应用程序视图291。

[0128] 在一些实施方案中,应用程序内部状态292包括另外的信息,诸如以下各项中的一者或多者:当应用程序236-1恢复执行时将被使用的恢复信息、指示正被应用程序236-1显示的信息或准备好用于被该应用程序显示的信息的用户界面状态信息、用于使得用户能够返回到应用程序236-1的前一状态或视图的状态队列以及用户采取的先前动作的重复/撤销队列。

[0129] 事件监视器271从外围设备接口218接收事件信息。事件信息包括关于子事件(例如,触敏显示器212上作为多点触摸手势一部分的用户触摸)的信息。外围设备接口218传输其从I/O子系统206或传感器诸如接近传感器266、加速度计268和/或麦克风213(通过音频电路210)接收的信息。外围设备接口218从I/O子系统206接收的信息包括来自触敏显示器212或触敏表面的信息。

[0130] 在一些实施方案中,事件监视器271以预先确定的间隔将请求发送到外围设备接

口218。作为响应,外围设备接口218传输事件信息。在其他实施方案中,外围设备接口218仅存在显著事件(例如,接收到高于预先确定的噪声阈值的输入和/或接收到超过预先确定的持续时间的输入)时才传输事件信息。

[0131] 在一些实施方案中,事件分类器270还包括命中视图确定模块272和/或活动事件识别器确定模块273。

[0132] 当触敏显示器212显示不止一个视图时,命中视图确定模块272提供用于确定子事件已在一个或多个视图内的什么地方发生的软件过程。视图由用户能够在显示器上看到的控件和其他元素构成。

[0133] 与应用程序相关联的用户界面的另一方面为一组视图,本文中有时也将其称为应用程序视图或用户界面窗口,在其中显示信息并且发生基于触摸的手势。在其中检测到触摸的(相应应用程序的)应用程序视图对应于应用程序的程序化分级结构或视图分级结构内的程序化水平。例如,在其中检测到触摸的最低水平视图被称为命中视图,并且被认为是正确输入的事件集至少部分地基于初始触摸的命中视图来确定,该初始触摸开始基于触摸的手势。

[0134] 命中视图确定模块272接收与基于触摸的手势的子事件相关的信息。当应用程序具有在分级结构中组织的多个视图时,命中视图确定模块272将命中视图识别为应对子事件进行处理的分级结构中的最低视图。在大多数情况下,命中视图是发起子事件(例如,形成事件或潜在事件的子事件序列中的第一子事件)在其中发生的最低水平视图。一旦命中视图被命中视图确定模块272识别,则命中视图通常接收与其被识别为命中视图所针对的同一触摸或输入源相关的所有子事件。

[0135] 活动事件识别器确定模块273确定视图分级结构内的哪个或哪些视图应当接收特定的子事件序列。在一些实施方案中,活动事件识别器确定模块273确定仅命中视图才应接收特定子事件序列。在其他实施方案中,活动事件识别器确定模块273确定包括子事件的物理位置的所有视图均为活跃参与的视图,并且由此确定所有活跃参与的视图均应接收特定子事件序列。在其他实施方案中,即使触摸子事件完全被局限到与一个特定视图相关联的区域,分级结构中的较高视图将仍然保持为活跃参与的视图。

[0136] 事件分配器模块274将事件信息分配到事件识别器(例如,事件识别器280)。在包括活动事件识别器确定模块273的实施方案中,事件分配器模块274将事件信息递送到由活动事件识别器确定模块273确定的事件识别器。在一些实施方案中,事件分配器模块274在事件队列中存储事件信息,该事件信息由相应事件接收器282进行检索。

[0137] 在一些实施方案中,操作系统226包括事件分类器270。另选地,应用程序236-1包括事件分类器270。在又一个实施方案中,事件分类器270为独立模块,或者为被存储在存储器202中的另一个模块(诸如,接触/运动模块230)的一部分。

[0138] 在一些实施方案中,应用程序236-1包括多个事件处理程序290以及一个或多个应用程序视图291,其中每个应用程序视图包括用于处理发生在应用程序的用户界面的相应视图内的触摸事件的指令。应用程序236-1的每个应用程序视图291包括一个或多个事件识别器280。通常,相应应用程序视图291包括多个事件识别器280。在其他实施方案中,事件识别器280中的一个或多个事件识别器是独立模块的一部分,该独立模块为诸如用户界面工具包(未示出)或应用程序236-1从中继承方法和其他属性的较高级别的对象。在一些实施

方案中,相应事件处理程序290包括以下各项中的一者或多者:数据更新器276、对象更新器277、GUI更新器278和/或从事件分类器270接收的事件数据279。事件处理程序290利用或调用数据更新器 276、对象更新器277或GUI更新器278来更新应用程序内部状态292。另选地,应用程序视图291中的一个或多个应用程序视图包括一个或多个相应的事件处理程序290。另外,在一些实施方案中,数据更新器276、对象更新器277和GUI更新器278中的一者或多者被包括在相应的应用程序视图291中。

[0139] 相应的事件识别器280从事件分类器270接收事件信息(例如,事件数据279),并且从该事件信息识别事件。事件识别器280包括事件接收器 282和事件比较器284。在一些实施方案中,事件识别器280还包括元数据 283和事件传递指令288(其包括子事件传递指令)的至少一个子集。

[0140] 事件接收器282从事件分类器270接收事件信息。事件信息包括关于子事件例如触摸或触摸移动的信息。根据子事件,事件信息还包括附加信息,诸如子事件的位置。当子事件涉及触摸的运动时,事件信息还包括子事件的速率和方向。在一些实施方案中,事件包括设备从一个取向旋转到另一取向(例如,从纵向取向旋转到横向取向,或反之亦然),并且事件信息包括关于设备的当前取向(也被称为设备姿态)的对应信息。

[0141] 事件比较器284将事件信息与预定义的事件或子事件定义进行比较,并且基于该比较来确定事件或子事件,或者确定或更新事件或子事件的状态。在一些实施方案中,事件比较器284包括事件定义286。事件定义286 包含事件的定义(例如,预定义的子事件序列),例如事件1(287-1)、事件2(287-2)以及其他事件。在一些实施方案中,事件(287)中的子事件包括例如触摸开始、触摸结束、触摸移动、触摸取消和多点触摸。在一个示例中,事件1(287-1)的定义是在显示的对象上的双击。例如,双击包括被显示对象上的预先确定时长的第一次触摸(触摸开始)、预先确定时长的第一次抬离(触摸结束)、被显示对象上的预先确定时长的第二次触摸(触摸开始)以及预先确定时长的第二次抬离(触摸结束)。在另一个示例中,事件2(287-2)的定义是在被显示对象上的拖动。例如,拖动包括在被显示对象上的预先确定时长的触摸(或接触)、触摸在触敏显示器212上的移动以及触摸的抬离(触摸结束)。在一些实施方案中,事件还包括用于一个或多个相关联的事件处理程序290的信息。

[0142] 在一些实施方案中,事件定义287包括用于相应用户界面对象的事件的定义。在一些实施方案中,事件比较器284执行命中测试,以确定哪个用户界面对象与子事件相关联。例如,在触敏显示器212上显示三个用户界面对象的应用程序视图中,当在触敏显示器212上检测到触摸时,事件比较器284执行命中测试,以确定这三个用户界面对象中哪一个用户界面对象与该触摸(子事件)相关联。如果每个所显示的对象与相应的事件处理程序290相关联,则事件比较器使用该命中测试的结果来确定哪个事件处理程序290应当被激活。例如,事件比较器284选择与子事件和触发该命中测试的对象相关联的事件处理程序。

[0143] 在一些实施方案中,相应事件(287)的定义还包括延迟动作,该延迟动作延迟事件信息的递送,直到已经确定子事件序列是对应于还是不对应于事件识别器的事件类型。

[0144] 当相应事件识别器280确定子事件串不与事件定义286中的任何事件匹配时,相应事件识别器280进入事件不可能、事件失败或事件结束状态,在此之后不理睬基于触摸的手势的后续子事件。在这种情况下,对于命中视图保持活动的其他事件识别器(如果有的话)继续跟踪并处理持续进行的基于触摸的手势的子事件。



[0145] 在一些实施方案中,相应事件识别器280包括元数据283,该元数据具有指示事件递送系统应该如何执行对活跃参与的事件识别器的子事件递送的可配置属性、标记和/或列表。在一些实施方案中,元数据283包括指示事件识别器彼此如何交互或如何能够交互的可配置属性、标志和/或列表。在一些实施方案中,元数据283包括指示子事件是否被递送到视图或程序化分级结构中的不同层级的可配置属性、标志和/或列表。

[0146] 在一些实施方案中,当事件的一个或多个特定的子事件被识别时,相应的事件识别器280激活与事件相关联的事件处理程序290。在一些实施方案中,相应事件识别器280将与事件相关联的事件信息递送到事件处理程序290。激活事件处理程序290不同于将子事件发送(和延期发送)到相应命中视图。在一些实施方案中,事件识别器280发出与所识别的事件相关联的标志,并且与该标志相关联的事件处理程序290获取该标志并执行预定义的过程。

[0147] 在一些实施方案中,事件递送指令288包括递送关于子事件的事件信息而不激活事件处理程序的子事件递送指令。相反,子事件递送指令将事件信息递送到与子事件序列相关联的事件处理程序或者递送到活跃参与的视图。与子事件序列或与活跃参与的视图相关联的事件处理程序接收事件信息并执行预先确定的过程。

[0148] 在一些实施方案中,数据更新器276创建和更新在应用程序236-1中使用的数据。例如,数据更新器276对联系人模块237中所使用的电话号码进行更新,或者对视频播放器模块中所使用的视频文件进行存储。在一些实施方案中,对象更新器277创建并更新在应用程序236-1中使用的对象。例如,对象更新器277创建新用户界面对象或更新用户界面对象的位置。GUI更新器278更新GUI。例如,GUI更新器278准备显示信息,并且将显示信息发送到图形模块232以用于显示在触敏显示器上。

[0149] 在一些实施方案中,事件处理程序290包括数据更新器276、对象更新器277和GUI更新器278,或具有对该数据更新器、对象更新器和GUI更新器的访问权限。在一些实施方案中,数据更新器276、对象更新器277和GUI更新器278被包括在相应的应用程序236-1或应用程序视图291的单个模块中。在其他实施方案中,它们被包括在两个或更多个软件模块中。

[0150] 应当理解,上述关于触敏显示器上的用户触摸的事件处理的讨论还适用于利用输入设备来操作多功能设备200的其他形式的用户输入,并不是所有用户输入均是在触摸屏上发起的。例如,任选地与单次或多次键盘按下或按住协作的鼠标移动和鼠标按钮按下;触控板上的接触移动,诸如轻击、拖动、滚动等;触笔输入;设备的移动;口头指令;检测到的眼睛移动;生物特征输入;和/或它们的任何组合任选地被用作对应于限定要识别的事件的子事件的输入。

[0151] 图3示出了根据一些实施方案的具有触摸屏212的便携式多功能设备200。触摸屏任选地在用户界面(UI)300内显示一个或多个图形。在本实施方案中以及下文所述的其他实施方案中,用户能够通过例如利用一个或多个手指302(在附图中未按比例绘制)或者利用一个或多个触笔303(在附图中未按比例绘制)在图形上做出手势来选择这些图形中的一个或多个图形。在一些实施方案中,当用户中断与一个或多个图形的接触时,将发生对一个或多个图形的选择。在一些实施方案中,手势任选地包括一次或多次轻击、一次或多次轻扫(从左向右、从右向左、向上和/或向下)和/或已与设备200发生接触的手指的滚动(从右向

左、从左向右、向上和/或向下)。在一些具体实施中或在一些情况下,不经意地与图形接触不会选择图形。例如,当与选择对应的手势是轻击时,在应用程序图标上方扫动的轻扫手势任选地不会选择对应的应用程序。

[0152] 设备200还包括一个或多个物理按钮,诸如“home”或菜单按钮 304。如前所述,菜单按钮304用于导航到在设备200上执行的一组应用程序中的任何应用程序236。另选地,在一些实施方案中,菜单按钮被实现为被显示在触摸屏212上的GUI中的软键。

[0153] 在一个实施方案中,设备200包括触摸屏212、菜单按钮304、用于使设备开启/关闭和用于锁定设备的下压按钮306、一个或多个音量调节按钮 308、用户身份模块(SIM)卡槽310、耳麦插孔312和对接/充电外部端口 224。下压按钮306任选地用于通过压下该按钮并且将该按钮保持在压下状态持续预定义的时间间隔来对设备进行开/关机;通过压下该按钮并在该预定义的时间间隔过去之前释放该按钮来锁定设备;和/或对设备进行解锁或发起解锁过程。在另选的实施方案中,设备200还通过麦克风213接受用于激活或去激活某些功能的语音输入。设备200还任选地包括用于检测触摸屏212上的接触的强度的一个或多个接触强度传感器265,和/或用于为设备200的用户生成触觉输出的一个或多个触觉输出发生器267。

[0154] 图4是根据一些实施方案的具有显示器和触敏表面的示例性多功能设备的框图。设备400不需要是便携式的。在一些实施方案中,设备400为膝上型计算机、台式计算机、平板电脑、多媒体播放器设备、导航设备、教育设备(诸如儿童学习玩具)、游戏系统或控制设备(例如,家用控制器或工业用控制器)。设备400通常包括一个或多个处理单元(CPU) 410、一个或多个网络或其他通信接口460、存储器470和用于使这些部件互连的一个或多个通信总线420。通信总线420任选地包括使系统部件互连并且控制系统部件之间的通信的电路(有时被称为芯片组)。设备400包括具有显示器440的输入/输出(I/O)接口430,该显示器通常为触摸屏显示器。I/O接口430还任选地包括键盘和/或鼠标(或其他指向设备) 450以及触控板 455、用于在设备400上生成触觉输出的触觉输出发生器457(例如,类似于上文参考图2A所述的一个或多个触觉输出发生器267)、传感器459(例如,光学传感器、加速度传感器、接近传感器、触敏传感器和/或类似于上文参考图2A所述的一个或多个接触强度传感器265的接触强度传感器)。存储器470包括高速随机存取存储器,诸如DRAM、SRAM、DDR RAM或其他随机存取固态存储器设备;并且任选地包括非易失性存储器,诸如一个或多个磁盘存储设备、光盘存储设备、闪存存储设备或其他非易失性固态存储设备。存储器470任选地包括远离一个或多个CPU 410定位的一个或多个存储设备。在一些实施方案中,存储器470存储与存储在便携式多功能设备200(图2A)的存储器202中的程序、模块和数据结构类似的程序、模块和数据结构或其子集。此外,存储器470任选地存储在便携式多功能设备200的存储器202中不存在的另外程序、模块和数据结构。例如,设备400的存储器470任选地存储绘图模块480、呈现模块 482、文字处理模块484、网站创建模块486、盘编辑模块488和/或电子表格模块490,而便携式多功能设备200(图2A)的存储器202任选地不存储这些模块。

[0155] 在一些示例中,图4中的上述元件中的每一者均存储在一个或多个先前提到的存储器设备中。上述模块中的每个模块对应于用于执行上述功能的指令集。上述模块或程序(例如,指令集)不必被实现为独立的软件程序、过程或模块,因此这些模块的各种子集在各种实施方案中组合或以其他方式重新布置。在一些实施方案中,存储器470存储上述模块和

数据结构的子组。此外,存储器470存储上文未描述的另外的模块和数据结构。

[0156] 现在将注意力转到可在例如便携式多功能设备200上实现的用户界面的实施方案。

[0157] 图5A示出了根据一些实施方案的便携式多功能设备200上的应用程序的菜单的示例性用户界面。类似的用户界面在设备400上实现。在一些实施方案中,用户界面500包括以下元件或者其子集或超集:

[0158] 用于一种或多种无线通信(诸如蜂窝信号和Wi-Fi信号)的一个或多个信号强度指示器502;

[0159] • 时间504;

[0160] • 蓝牙指示器505;

[0161] • 电池状态指示器506;

[0162] • 具有常用应用程序的图标托盘508,常用应用程序图标诸如:

[0163] ◦电话模块238的被标记为“电话”的图标516,该图标任选地包括未接来电或语音留言的数量的指示符514;

[0164] ◦电子邮件客户端模块240的被标记为“邮件”的图标518,该图标任选地包括未读电子邮件的数量的指示符510;

[0165] ◦浏览器模块247的被标记为“浏览器”的图标520;以及

[0166] ◦视频和音乐播放器模块252(也称为iPod (Apple Inc.的商标) 模块252)的被标记为“iPod”的图标522;以及

[0167] • 其他应用的图标,诸如:

[0168] ◦Im模块241的被标记为“消息”的图标524;;

[0169] ◦日历模块248的被标记为“日历”的图标526;;

[0170] ◦图像管理模块244的被标记为“照片”的图标528;;

[0171] ◦相机模块243的被标记为“相机”的图标530;;

[0172] ◦在线视频模块255的被标记为“在线视频”的图标532;;

[0173] ◦股市桌面小程序249-2的被标记为“股市”的图标534;;

[0174] ◦地图模块254的被标记为“地图”的图标536;;

[0175] ◦天气桌面小程序249-1的被标记为“天气”的图标538;;

[0176] ◦闹钟桌面小程序249-4的被标记为“时钟”的图标540;;

[0177] ◦健身支持模块242的被标记为“健身支持”的图标542;;

[0178] ◦记事本模块253的被标记为“记事本”的图标544;以及

[0179] ◦用于设置应用程序或模块的被标记为“设置”的图标546,该图标提供对设备200及其各种应用程序236的设置访问。

[0180] 应当指出的是,图5A中示出的图标标签仅为示例性的。例如,视频和音乐播放器模块252的图标522任选地被标记为“音乐”或“音乐播放器”。对于各种应用程序图标任选地使用其他标签。在一些实施方案中,相应应用程序图标的标签包括与该相应应用程序图标对应的应用程序的名称。在一些实施方案中,特定应用程序图标的标签不同于与该特定应用程序图标对应的应用程序的名称。

[0181] 图5B示出了具有与显示器550(例如,触摸屏显示器212)分开的触敏表面551(例

如,图4的平板或触摸板455)的设备(例如,图4的设备 400)上的示例性用户界面。设备400还任选地包括用于检测触敏表面551 上的接触的强度的一个或多个接触强度传感器(例如,传感器457中的一个或多个传感器)和/或用于为设备400的用户生成触觉输出的一个或多个触觉输出发生器459。

[0182] 尽管将参考触摸屏显示器212(其中组合了触敏表面和显示器)上的输入给出随后的示例中的一些示例,但在一些实施方案中,设备检测与显示器分开的触敏表面上的输入,如图5B中所示。在一些实施方案中,触敏表面(例如,图5B中的551)具有与显示器(例如,550)上的主轴(例如,图5B中的553)对应的主轴(例如,图5B中的552)。根据这些实施方案,设备检测在与显示器上的相应位置对应的位置(例如,在图5B中,560对应于568并且562对应于570)处的与触敏表面551的接触(例如,图5B中的560和562)。这样,在触敏表面(例如,图5B中的551)与多功能设备的显示器(例如,图5B中的550)分开时,由设备在触敏表面上检测到的用户输入(例如,接触560和562以及它们的移动)被该设备用于操纵显示器上的用户界面。应当理解,类似的方法任选地用于本文所述的其他用户界面。

[0183] 另外,虽然主要是参考手指输入(例如,手指接触、手指轻击手势、手指轻扫手势)来给出下面的示例,但应当理解,在一些实施方案中,这些手指输入中的一个或多个由来自另一个输入设备的输入(例如,基于鼠标的输入或触笔输入)替换。例如,轻扫手势任选地由鼠标点击(例如,而不是接触),之后是光标沿着轻扫的路径的移动(例如,而不是接触的移动)替代。又如,轻击手势任选地由在光标位于轻击手势的位置上方时的鼠标点击(例如,代替对接触的检测,之后是停止检测接触)替代。类似地,当同时检测到多个用户输入时,应当理解的是,多个计算机鼠标任选地被同时使用,或鼠标和手指接触任选地被同时使用。

[0184] 图6A示出了示例性个人电子设备600。设备600包括主体602。在一些实施方案中,设备600包括相对于设备200和400(例如,图2A-图4)所述的特征中的一些或全部特征。在一些实施方案中,设备600具有在下文中称为触摸屏604的触敏显示屏604。另选地或除触摸屏604之外,设备 600还具有显示器和触敏表面。与设备200和400的情况一样,在一些实施方案中,触摸屏604(或触敏表面)具有用于检测正在施加的接触(例如,触摸)的强度的一个或多个强度传感器。触摸屏604(或触敏表面)的一个或多个强度传感器提供表示触摸的强度的输出数据。设备600的用户界面基于触摸强度来对触摸作出响应,这意味着不同强度的触摸可调用设备600 上的不同的用户界面操作。

[0185] 用于检测和处理触摸强度的技术可例如在相关申请中找到:2013年5 月8日提交的名称为“Device,Method,and Graphical User Interface for Displaying User Interface Objects Corresponding to an Application”的国际专利申请序列号PCT/US2013/040061,以及2013年11月11日提交的名称为“Device,Method,and Graphical User Interface for Transitioning Between Touch Input to Display Output Relationships”的国际专利申请序列号 PCT/US2013/069483,这两个专利申请中的每个专利申请均据此全文以引用方式并入。

[0186] 在一些实施方案中,设备600具有一个或多个输入机构606和608。输入机构606和608(如果包括的话)是物理形式的。物理输入机构的示例包括下压按钮和可旋转机构。在一些实施方案中,设备600具有一个或多个附接机构。此类附接机构(如果包括的话)可允许将设备600与例如帽子、眼镜、耳环、项链、衬衣、夹克、手镯、表带、手链、裤子、皮带、鞋子、钱

包、背包等附接。这些附接机构允许用户佩戴设备600。

[0187] 图6B描绘了示例性个人电子设备600。在一些实施方案中,设备600 包括相对于图2A、图2B和图4所述的部件中的一些或全部部件。设备600 具有总线612,该总线将I/O部分614与一个或多个计算机处理器616和存储器618操作性地耦接。I/O部分614被连接到显示器604,该显示器可具有触敏部件622并且任选地还具有触摸强度敏感部件624。此外,I/O部分614与通信单元630连接,以用于使用Wi-Fi、蓝牙、近场通信(NFC)、蜂窝和/或其他无线通信技术来接收应用程序和操作系统数据。设备600包括输入机构606和/或608。例如,输入机构606是可旋转输入设备或者可按压输入设备以及可旋转输入设备。在一些示例中,输入机构608是按钮。

[0188] 在一些示例中,输入机构608任选地为麦克风。个人电子设备600包括例如各种传感器,诸如GPS传感器632、加速度计634、定向传感器640(例如,罗盘)、陀螺仪636、运动传感器638和/或它们的组合,所有这些设备均可操作性连接到I/O部分614。

[0189] 个人电子设备600的存储器618为用于存储计算机可执行指令的非暂态计算机可读存储介质,该指令在由一个或多个计算机处理器616执行时例如使得计算机处理器执行下文所述的技术和过程。该计算机可执行指令也例如在任何非暂态计算机可读存储介质内进行存储和/或传送,以供指令执行系统、装置或设备诸如基于计算机的系统、包含处理器的系统或可从指令执行系统、装置或设备获取指令并执行指令的其他系统使用或与其结合。个人电子设备600不限于图6B的部件和配置,而是可包括多种配置的其他部件或另外部件。

[0190] 如本文所用,术语“示能表示”是指例如在设备200、400、600和/或 800(图2A、图4、图6A至图6B和/或图8A至图8E)的显示屏上显示的用户交互式图形用户界面对象。例如,图像(例如,图标)、按钮和文本(例如,超链接)各自构成示能表示。

[0191] 如本文所用,术语“焦点选择器”是指指示用户界面中用户正在与其进行交互的当前部分的输入元件。在包括光标或其他位置标记的一些具体实施中,光标充当“焦点选择器”,使得当光标在特定用户界面元素(例如,按钮、窗口、滑块或其他用户界面元素)上方时在触敏表面(例如,图4中的触摸板455或图5B中的触敏表面551)上检测到输入(例如,按压输入)的情况下,该特定用户界面元素根据所检测到的输入而被调节。在包括能够实现与触摸屏显示器上的用户界面元素的直接交互的触摸屏显示器(例如,图2A中的触敏显示系统212或图5A中的触摸屏212)的一些具体实施中,触摸屏上的所检测到的接触充当“焦点选择器”,使得当在触摸屏显示器上在特定用户界面元素(例如,按钮、窗口、滑块或其他用户界面元素)的位置处检测到输入(例如,由接触进行的按压输入)时,根据所检测到的输入来调整特定用户界面元素。在一些具体实施中,焦点从用户界面的一个区域移动到用户界面的另一个区域,而无需光标的对应移动或触摸屏显示器上的接触的移动(例如,通过使用制表键或箭头键将焦点从一个按钮移动到另一个按钮);在这些具体实施中,焦点选择器根据焦点在用户界面的不同区域之间的移动而移动。不考虑焦点选择器所采取的具体形式,焦点选择器通常是由用户控制的以便递送与用户界面的用户预期的交互(例如,通过向设备指示用户界面的用户期望与其进行交互的元素)的用户界面元素(或触摸屏显示器上的接触)。例如,在触敏表面(例如,触摸板或触摸屏)上检测到按压输入时,焦点选择器(例如,光标、接触或选择框)在相应按钮上方的位置将指示用户期望激活相应按钮(而不是设备显示

器上示出的其他用户界面元素)。

[0192] 如说明书和权利要求中所使用的那样,接触的“特征强度”这一术语是指基于接触的一个或多个强度的接触的特征。在一些实施方案中,特征强度基于多个强度样本。特征强度任选地基于相对于预定义事件(例如,在检测到接触之后,在检测到接触抬离之前,在检测到接触开始移动之前或之后,在检测到接触结束之前,在检测到接触的强度增大之前或之后和/或在检测到接触的强度减小之前或之后)而言在预先确定的时间段(例如,0.05秒、0.1秒、0.2秒、0.5秒、1秒、2秒、5秒、10秒)期间采集的预定义数量的强度样本或一组强度样本。接触的特征强度任选地基于以下各项中的一者或多者:接触强度的最大值、接触强度的均值、接触强度的平均值、接触强度的前10%处的值、接触强度的半最大值、接触强度的90%最大值等。在一些实施方案中,在确定特征强度时使用接触的持续时间(例如,在特征强度是接触的强度在时间上的平均值时)。在一些实施方案中,将特征强度与一组一个或多个强度阈值进行比较,以确定用户是否已执行操作。例如,该组一个或多个强度阈值包括第一强度阈值和第二强度阈值。在该示例中,特征强度未超过第一阈值的接触导致第一操作,特征强度超过第一强度阈值但未超过第二强度阈值的接触导致第二操作,而特征强度超过第二阈值的接触导致第三操作。在一些实施方案中,使用特征强度与一个或多个阈值之间的比较来确定是否要执行一个或多个操作(例如,是执行相应操作还是放弃执行相应操作),而不是用于确定执行第一操作还是第二操作。

[0193] 在一些实施方案中,识别手势的一部分以用于确定特征强度。例如,触敏表面接收连续的轻扫接触,该连续的轻扫接触从起始位置过渡并到达结束位置,在该结束位置处,接触的强度增加。在该示例中,接触在结束位置处的特征强度仅基于连续轻扫接触的一部分,而不是整个轻扫接触(例如,仅轻扫接触在结束位置处的部分)。在一些实施方案中,在确定接触的特征强度之前向轻扫手势的强度应用平滑化算法。例如,平滑化算法任选地包括以下各项中的一种或多种:不加权滑动平均平滑化算法、三角平滑化算法、中值滤波器平滑化算法和/或指数平滑化算法。在一些情况下,这些平滑化算法消除了轻扫接触的强度中的窄的尖峰或凹陷,以实现确定特征强度的目的。

[0194] 相对于一个或多个强度阈值(诸如,接触检测强度阈值、轻按压强度阈值、深按压强度阈值和/或一个或多个其他强度阈值)来表征触敏表面上的接触的强度。在一些实施方案中,轻按压强度阈值对应于这样的强度:在该强度下设备将执行通常与点击物理鼠标或触摸板的按钮相关联的操作。在一些实施方案中,深按压强度阈值对应于这样的强度:在该强度下设备将执行与通常与点击物理鼠标或触控板的按钮相关联的操作不同的操作。在一些实施方案中,当检测到特征强度低于轻按压强度阈值(例如,并且高于标称接触检测强度阈值,比标称接触检测强度阈值低的接触不再被检测到)的接触时,设备将根据接触在触敏表面上的移动来移动焦点选择器,而不执行与轻按压强度阈值或深按压强度阈值相关联的操作。一般来讲,除非另有陈述,否则这些强度阈值在不同组的用户界面附图之间是一致的。

[0195] 接触的特征强度从低于轻按压强度阈值的强度增大到介于轻按压强度阈值与深按压强度阈值之间的强度有时被称为“轻按压”输入。接触特征强度从低于深按压强度阈值的强度增大到高于深按压强度阈值的强度有时被称为“深按压”输入。接触特征强度从低于接触检测强度阈值的强度增大到介于接触检测强度阈值与轻按压强度阈值之间的强度有

时被称为检测到触摸表面上的接触。接触特征强度从高于接触检测强度阈值的强度减小到低于接触检测强度阈值的强度有时被称为检测到接触从触摸表面抬离。在一些实施方案中,接触检测强度阈值为零。在一些实施方案中,接触检测强度阈值大于零。

[0196] 在本文所述的一些实施方案中,响应于检测到包括相应按压输入的手势或响应于检测到利用相应接触(或多个接触)执行的相应按压输入来执行一个或多个操作,其中至少部分地基于检测该接触(或多个接触)的强度增大到高于按压输入强度阈值来检测相应按压输入。在一些实施方案中,响应于检测到相应接触的强度增大到高于按压输入强度阈值(例如,相应按压输入的“向下冲程”)来执行相应操作。在一些实施方案中,按压输入包括相应接触的强度增大到高于按压输入强度阈值以及该接触的强度随后减小到低于按压输入强度阈值,并且响应于检测到相应接触的强度随后减小到低于按压输入强度阈值(例如,相应按压输入的“向上冲程”)来执行相应操作。

[0197] 在一些实施方案中,设备采用强度滞后以避免有时称为“抖动”的意外输入,其中设备限定或选择与按压输入强度阈值具有预定义关系的滞后强度阈值(例如,滞后强度阈值比按压输入强度阈值低X个强度单位,或者滞后强度阈值为按压输入强度阈值的75%、90%或某个合理比例)。因此,在一些实施方案中,按压输入包括相应接触的强度增大到高于按压输入强度阈值以及该接触的强度随后减小到低于对应于按压输入强度阈值的滞后强度阈值,并且响应于检测到相应接触的强度随后减小到低于滞后强度阈值(例如,相应按压输入的“向上冲程”)来执行相应操作。类似地,在一些实施方案中,仅在设备检测到接触强度从等于或低于滞后强度阈值的强度增大到等于或高于按压输入强度阈值的强度并且任选地接触强度随后减小到等于或低于滞后强度的强度时才检测到按压输入,并且响应于检测到按压输入(例如,根据环境,接触强度增大或接触强度减小)来执行相应操作。

[0198] 为了便于解释,任选地,响应于检测到以下各种情况中的任一种情况而触发对响应于与按压输入强度阈值相关联的按压输入或响应于包括按压输入的手势而执行的操作的描述:接触的强度增大到高于按压输入强度阈值、接触的强度从低于滞后强度阈值的强度增大到高于按压输入强度阈值的强度、接触的强度减小到低于按压输入强度阈值和/或接触的强度减小到低于与按压输入强度阈值对应的滞后强度阈值。另外,在将操作描述为响应于检测到接触的强度减小到低于按压输入强度阈值而执行的示例中,任选地响应于检测到接触的强度减小到低于对应于并且小于按压输入强度阈值的滞后强度阈值来执行操作。

### [0199] 3. 数字助理系统

[0200] 图7A示出了根据各种示例的数字助理系统700的框图。在一些示例中,数字助理系统700在独立式计算机系统上实现。在一些示例中,数字助理系统700跨多个计算机分布。在一些示例中,数字助理的模块和功能中的一些被划分成服务器部分和客户端部分,其中客户端部分位于一个或多个用户设备(例如,设备104、122、200、400、600或800)上并通过一个或多个网络与服务器部分(例如,服务器系统108)通信,例如,如图1 中所示。在一些示例中,数字助理系统700是图1中所示的服务器系统108 (和/或DA服务器106)的具体实施。应当指出,数字助理系统700仅为数字助理系统的一个示例,且该数字助理系统700具有比所示更多或更少的部件、组合两个或更多个部件,或者可具有部件的不同配置或布局。图 7A中所示的各种部件在硬件、用于由一个或多个处理器执行的软件指令、固件(包括一个或多



个信号处理集成电路和/或专用集成电路),或它们的组合中实现。

[0201] 数字助理系统700包括存储器702、一个或多个处理器704、输入/输出(I/O)接口706以及网络通信接口708。这些部件可通过一条或多条通信总线或信号线710彼此通信。

[0202] 在一些示例中,存储器702包括非暂态计算机可读介质,诸如高速随机存取存储器和/或非易失性计算机可读存储介质(例如,一个或多个磁盘存储设备、闪存存储设备或其他非易失性固态存储设备)。

[0203] 在一些示例中,I/O接口706将数字助理系统700的输入/输出设备716(诸如,显示器、键盘、触摸屏和麦克风)耦接到用户界面模块722。I/O接口706,与用户界面模块722结合,接收用户输入(例如,语音输入、键盘输入、触摸输入等)并相应地对这些输入进行处理。在一些示例中,例如,当数字助理在独立用户设备上实现时,数字助理系统700包括相对于图2A、图4、图6A至图6B和图8A至图8E中的设备200、400、600或800分别所述的部件和I/O通信接口中的任一者。在一些示例中,数字助理系统700表示数字助理具体实施的服务器部分,并且可通过位于用户设备(例如,设备104、200、400、600或800)上的客户端侧部分与用户进行交互。

[0204] 在一些示例中,网络通信接口708包括一个或多个有线通信端口712以及/或者无线传输和接收电路714。一个或多个有线通信端口经由一个或多个有线接口例如以太网、通用串行总线(USB)、火线等接收和发送通信信号。无线电路714从通信网络及其他通信设备接收RF信号和/或光学信号以及将RF信号和/或光学信号发送至通信网络及其他通信设备。无线通信使用多种通信标准、协议和技术中的任一种,诸如GSM、EDGE、CDMA、TDMA、蓝牙、Wi-Fi、VoIP、Wi-MAX、或任何其他合适的通信协议。网络通信接口708使数字助理系统700通过网络,诸如互联网、内联网和/或无线网络诸如蜂窝电话网络、无线局域网(LAN)和/或城域网(MAN),与其他设备之间的通信成为可能。

[0205] 在一些示例中,存储器702或存储器702的计算机可读存储介质存储程序、模块、指令和数据结构,其包括以下内容中的全部或其子集:操作系统718、通信模块720、用户界面模块722、一个或多个应用程序724和数字助理模块726。具体地讲,存储器702或存储器702的计算机可读存储介质存储用于执行上述过程的指令。一个或多个处理器704执行这些程序、模块和指令,并从数据结构读取数据或将数据写到数据结构。

[0206] 操作系统718(例如,Darwin、RTXC、LINUX、UNIX、iOS、OS X、WINDOWS或嵌入式操作系统诸如VxWorks)包括用于控制和管理一般系统任务(例如,存储器管理、存储设备控制、电源管理等)的各种软件部件和/或驱动器,并且有利于各种硬件、固件和软件部件之间的通信。

[0207] 通信模块720有利于数字助理系统700与其他设备之间通过网络通信接口708进行的通信。例如,通信模块720与电子设备(诸如分别在图2A、图4、图6A至图6B中所示的设备200、400或600)的RF电路208通信。通信模块720还包括各种部件,用于处理由无线电路714和/或有线通信端口712所接收的数据。

[0208] 用户界面模块722经由I/O接口706接收来自用户(例如,来自键盘、触摸屏、指向设备、控制器和/或麦克风)的命令和/或输入,并且在显示器上生成用户界面对象。用户界面模块722还准备输出(例如,语音、声音、动画、文本、图标、振动、触觉反馈、光照等)并将其经由I/O接口706(例如,通过显示器、音频通道、扬声器、触控板等)传送给用户。



[0209] 应用程序724包括被配置为由一个或多个处理器704执行的程序和/或模块。例如,如果数字助理系统在独立式用户设备上实施,则应用程序724 包括用户应用程序,诸如游戏、日历应用程序、导航应用程序或邮件应用程序。如果数字助理系统700在服务器上实现,则应用程序724包括例如资源管理应用程序、诊断应用程序、或调度应用程序。

[0210] 存储器702还存储数字助理模块726(或数字助理的服务器部分)。在一些示例中,数字助理模块726包括以下子模块或者其子集或超集:输入/输出处理模块728、语音转文本(STT)处理模块730、自然语言处理模块 732、对话流处理模块734、任务流处理模块736、服务处理模块738和语音合成处理模块740。这些模块中的每一者均具有对以下数字助理模块726 的系统或数据与模型中的一者或多者或者其子集或超集的访问权限:知识本体760、词汇索引744、用户数据748、任务流模型754、服务模型756 和ASR系统758。

[0211] 在一些示例中,使用在数字助理模块726中实现的处理模块、数据和模型,数字助理可执行以下中的至少一些:将语音输入转换成文本;识别在从用户接收的自然语言输入中表达的用户意图;主动引出并获得完全推断用户意图所需的信息(例如,通过消除字词、姓名、意图的歧义等);确定用于满足推断出的意图的任务流;以及执行该任务流以满足推断出的意图。

[0212] 在一些示例中,如图7B中所示,I/O处理模块728可通过图7A中的 I/O设备716与用户交互或者通过图7A中的网络通信接口708与用户设备(例如,设备104、200、400或600)交互,以获取用户输入(例如,语音输入)并提供对用户输入的响应(例如,作为语音输出)。I/O处理模块 728随同接收到用户输入一起或在接收到用户输入之后不久任选地获得与来自用户设备的用户输入相关联的上下文信息。上下文信息包括特定于用户的数据、词汇,和/或与用户输入相关的偏好。在一些示例中,该上下文信息还包括在接收到用户请求时的用户设备的软件状态和硬件状态,和/或在接收到用户请求时与用户的周围环境相关的信息。在一些示例中,I/O处理模块728还向用户发送与用户请求有关的跟进问题,并从用户接收回答。在用户请求被I/O处理模块728接收且用户请求包括语音输入时,I/O处理模块728将语音输入转发至STT处理模块730(或语音识别器)以进行语音文本转换。

[0213] STT处理模块730包括一个或多个ASR系统758。该一个或多个ASR 系统758可处理通过I/O处理模块728接收到的言语输入,以产生识别结果。每个ASR系统758可包括前端言语预处理器。前端语音预处理器从语音输入中提取代表性特征。例如,前端语音预处理器对语音输入执行傅里叶变换,以提取表征语音输入的光谱特征作为代表性多维向量的序列。另外,每个ASR系统758包括一个或多个言语识别模型(例如,声学模型和/ 或语言模型)并且实现一个或多个言语识别引擎。语音识别模型的示例包括隐马尔可夫模型、高斯混合模型、深层神经网络模型、n元语言模型以及其他统计模型。语音识别引擎的示例包括基于动态时间规整的引擎和基于加权有限状态变换器(WFST)的引擎。使用一个或多个语音识别模型和一个或多个语音识别引擎来处理前端语音预处理器的所提取的代表性特征以产生中间识别结果(例如,音素、音素串和子字词),并且最终产生文本识别结果(例如,字词、字词串、或符号序列)。在一些示例中,语音输入至少部分地由第三方服务处理或在用户的设备(例如,设备104、设备 200、设备400或设备600)上处理,以产生识别结果。一旦STT处理模块730产生包含文本串(例如,字词,或字词的序列,或符号序列)的识别结果,识别结果即被传送至自然语言处理模块732以供意图推断。在一些示例中,STT处理模块730产生言语输入的

多个候选文本表示。每个候选文本表示是与语音输入对应的字词或符号的序列。在一些示例中,每个候选文本表示与言语识别置信度得分相关联。基于语音识别置信度得分,STT 处理模块730对候选文本表示进行排序并将n个最佳(例如,n个排名最高)候选文本表示提供给自然语言处理模块732以供意图推断,其中n为大于零的预先确定的整数。例如,在一个示例中,仅将排名最高的(n=1)候选文本表示递送至自然语言处理模块732以供意图推断。又如,将5个排名最高的(n=5)候选文本表示传递给自然语言处理模块732以供意图推断。

[0214] 有关语音转文本处理的更多细节在提交于2011年9月20日的名称为“Consolidating Speech Recognition Results”的美国实用新型专利申请序列号13/236,942中有所描述,其全部公开内容以引用方式并入本文。

[0215] 在一些示例中,STT处理模块730包括可识别字词的词汇并且/或者经由语音字母转换模块731访问该词汇。每个词汇字词与语音识别语音字母表中表示的字词的一个或多个候选发音相关联。具体地讲,可识别字词的词汇包括与多个候选发音相关联的字词。例如,该词汇包括与/tə'meɪroʊ/和/tə'matoʊ/的候选发音相关联的字词“tomato”。另外,词汇字词与基于来自用户的先前语音输入的自定义候选发音相关联。此类自定义候选发音存储在STT处理模块730中,并且经由设备上的用户配置文件与特定用户相关联。在一些示例中,字词的候选发音基于字词的拼写以及一个或多个语言学和/或语音学规则确定。在一些示例中,候选发音手动生成,例如,基于已知的标准发音而手动生成。

[0216] 在一些示例中,基于候选发音的普遍性来对候选发音进行排名。例如,候选语音/tə'meɪroʊ/的排序高于/tə'matoʊ/,因为前者是更常用的发音(例如,在所有用户中,对于特定地理区域的用户而言,或者对于任何其他合适的用户子集而言)。在一些示例中,基于候选发音是否为与用户相关联的自定义候选发音来对候选发音进行排序。例如,自定义候选发音的排名高于标准候选发音。这可用于识别具有偏离规范发音的独特发音的专有名词。在一些示例中,候选发音与一个或多个语音特征(诸如地理起源、国家或种族)相关联。例如,候选发音/tə'meɪroʊ/与美国相关联,而候选发音/tə'matoʊ/与英国相关联。此外,候选发音的排序基于存储在设备上的用户配置文件中的用户的一个或多个特征(例如,地理起源、国家、种族等)。例如,可从用户配置文件确定该用户与美国相关联。基于用户与美国相关联,候选发音/tə'meɪroʊ/ (与美国相关联)可比候选发音/tə'matoʊ/ (与英国相关联)排名更高。在一些示例中,经排序的候选发音中的一个可被选作预测发音(例如,最可能的发音)。

[0217] 当接收到语音输入时,STT处理模块730被用来(例如,使用声音模型)确定对应于该语音输入的音素,然后尝试(例如,使用语言模型)确定匹配该音素的字词。例如,如果STT处理模块730首先识别对应于该语音输入的一部分的音素序列/tə'meɪroʊ/,那么它随后可基于词汇索引744确定该序列对应于字词“tomato”。

[0218] 在一些示例中,STT处理模块730使用模糊匹配技术来确定话语中的字词。因此,例如,STT处理模块730确定音素序列/tə'meɪroʊ/对应于字词“tomato”,即使该特定音素序列不是该字词的候选音素序列。

[0219] 数字助理的自然语言处理模块732(“自然语言处理器”)获取由STT处理模块730生成的n个最佳候选文本表示(“字词序列”或“符号序列”),并且尝试将每个候选文本表示与由数字助理所识别的一个或多个“可执行意图”相关联。“可执行意图”(或“用户意图”)表

示可由数字助理执行并且可具有在任务流模型754中实现的相关联的任务流的任务。相关联任务流是数字助理为了执行任务而采取的一系列经编程的动作和步骤。数字助理的能力范围取决于已在任务流模型754中实现并存储的任务流的数量和种类,或换言之,取决于数字助理所识别的“可执行意图”的数量和种类。然而,数字助理的有效性还取决于助理从以自然语言表达的用户请求中推断出正确的“一个或多个可执行意图”的能力。

[0220] 在一些示例中,除从STT处理模块730获取的字词或符号的序列之外,自然语言处理模块732还例如从I/O处理模块728接收与用户请求相关联的上下文信息。自然语言处理模块732任选地使用上下文信息来明确、补充和/或进一步限定在从STT处理模块730接收的候选文本表示中包含的信息。上下文信息包括例如用户偏好,用户设备的硬件和/或软件状态,在用户请求之前、期间或之后不久收集的传感器信息,数字助理与用户之间的先前交互(例如,对话),等等。如本文所述,在一些示例中,上下文信息是动态的,并且随对话的时间、位置、内容、以及其他因素而变化。

[0221] 在一些示例中,自然语言处理基于例如知识本体760。知识本体760 为包含许多节点的分级结构,每个节点表示“可执行意图”或与“可执行意图”或其他“属性”中的一者或多者相关的“属性”。如上所述,“可执行意图”表示数字助理能够执行的任务,即,该任务为“可执行的”或可被进行的。“属性”代表与可执行意图或另一属性的子方面相关联的参数。知识本体760中可执行意图节点与属性节点之间的连接定义由属性节点表示的参数如何从属于由可执行意图节点表示的任务。

[0222] 在一些示例中,知识本体760由可执行意图节点和属性节点组成。在知识本体760内,每个可执行意图节点直接连接至或通过一个或多个中间属性节点连接至一个或多个属性节点。类似地,每个属性节点直接连接至或通过一个或多个中间属性节点连接至一个或多个可执行意图节点。例如,如图7C所示,知识本体760包括“餐厅预订”节点(即,可执行意图节点)。属性节点“餐厅”、“日期/时间”(针对预订)和“同行人数”均直接连接至可执行意图节点(即,“餐厅预订”节点)。

[0223] 此外,属性节点“菜系”、“价格区间”、“电话号码”和“位置”是属性节点“餐厅”的子节点,并且各自通过中间属性节点“餐厅”链接至“餐厅预订”节点(即,可执行意图节点)。又如,如图7C所示,知识本体760还包括“设定提醒”节点(即,另一个可执行意图节点)。属性节点“日期/时间”(针对设定提醒)和“主题”(针对提醒)均连接至“设定提醒”节点。由于属性“日期/时间”与进行餐厅预订的任务和设定提醒的任务二者相关,因此属性节点“日期/时间”连接至知识本体760中的“餐厅预订”节点和“设定提醒”节点二者。

[0224] 可执行意图节点连同其链接的属性节点一起被描述为“域”。在本讨论中,每个域与相应的可执行意图相关联,并涉及与特定可执行意图相关联的一组节点(以及这些节点之间的关系)。例如,图7C中示出的知识本体760包括在知识本体760内的餐厅预订域762的示例以及提醒域764的示例。餐厅预订域包括可执行意图节点“餐厅预订”、属性节点“餐厅”、“日期/时间”和“同行人数”以及子属性节点“菜系”、“价格区间”、“电话号码”和“位置”。提醒域764包括可执行意图节点“设定提醒”和属性节点“主题”和“日期/时间”。在一些示例中,知识本体 760由多个域组成。每个域与一个或多个其他域共享一个或多个属性节点。例如,除了餐厅预订域762和提醒域764之外,“日期/时间”属性节点还与许多不同域(例如,行程安排域、旅行预订域、电影票域等)相关联。

[0225] 尽管图7C示出知识本体760内的两个示例域,但其他域包括例如“查找电影”、“发起电话呼叫”、“查找方向”、“安排会议”、“发送消息”以及“提供问题的回答”、“阅读列表”、“提供导航指令”、“提供针对任务的指令”等。“发送消息”域与“发送消息”可执行意图节点相关联,并且进一步包括属性节点诸如“一个或多个接收人”、“消息类型”和“消息正文”。属性节点“接收人”进一步例如由子属性节点诸如“接收人姓名”和“消息地址”来限定。

[0226] 在一些示例中,知识本体760包括数字助理能够理解并对其起作用的所有域(以及因而可执行意图)。在一些示例中,知识本体760诸如通过添加或移除整个域或节点,或者通过修改知识本体760内的节点之间的关系进行修改。

[0227] 在一些示例中,将与多个相关可执行意图相关联的节点群集在知识本体760中的“超级域”下。例如,“旅行”超级域包括与旅行相关的属性节点和可执行意图节点的群集。与旅行相关的可执行意图节点包括“机票预订”、“酒店预订”、“汽车租赁”、“获取路线”、“寻找兴趣点”,等等。同一超级域(例如,“旅行”超级域)下的可执行意图节点具有多个共用的属性节点。例如,针对“机票预订”、“酒店预订”、“汽车租赁”、“路线规划”和“寻找感兴趣的点”的可执行意图节点共享属性节点“起始位置”、“目的地”、“出发日期/时间”、“到达日期/时间”和“同行人数”中的一者或多者。

[0228] 在一些示例中,知识本体760中的每个节点与和由节点表示的属性或可执行意图相关的一组字词和/或短语相关联。与每个节点相关联的相应组的字词和/或短语是所谓的与节点相关联的“词汇”。将与每个节点相关联的相应组的字词和/或短语存储在与由节点所代表的属性或可执行意图相关联的词汇索引744中。例如,返回图7B,与“餐厅”属性的节点相关联的词汇包括字词诸如“美食”、“酒水”、“菜系”、“饥饿”、“吃”、“披萨”、“快餐”、“膳食”等。又如,与“发起电话呼叫”可执行意图的节点相关联的词汇包括字词和短语诸如“呼叫”、“打电话”、“拨打”、“与……通电话”、“呼叫该号码”、“打电话给”等。词汇索引 744任选地包括不同语言的字词和短语。

[0229] 自然语言处理模块732接收来自STT处理模块730的候选文本表示(例如,一个或多个文本字符串或一个或多个符号序列),并且针对每个候选表示,确定候选文本表示中的字词涉及哪些节点。在一些示例中,如果发现候选文本表示中的字词或短语(经由词汇索引744)与知识本体760中的一个或多个节点相关联,则所述字词或短语“触发”或“激活”这些节点。基于已激活节点的数量和/或相对重要性,自然语言处理模块732选择可执行意图中的一个可执行意图作为用户意图使数字助理执行的任务。在一些示例中,选择具有最多“已触发”节点的域。在一些示例中,选择具有最高置信度(例如,基于其各个已触发节点的相对重要性)的域。在一些示例中,基于已触发节点的数量和重要性的组合来选择域。在一些示例中,在选择节点的过程中还考虑附加因素,诸如数字助理先前是否已正确解译来自用户的类似请求。

[0230] 用户数据748包括特定于用户的信息,诸如特定于用户的词汇、用户偏好、用户地址、用户的默认语言和第二语言、用户的联系人列表以及每位用户的其他短期或长期信息。在一些示例中,自然语言处理模块732使用特定于用户的信息来补充用户输入中所包含的信息以进一步限定用户意图。例如,针对用户请求“邀请我的朋友参加我的生日派对”,自然语言处理模块732能够访问用户数据748以确定“朋友”是哪些人以及“生日派对”将于何时何地举行,而不需要用户在其请求中明确地提供此类信息。

[0231] 应当认识到,在一些示例中,利用一个或多个机器学习机构(例如,神经网络)来实现自然语言处理模块732。具体地,一个或多个机器学习机构被配置为接收候选文本表示和与候选文本表示相关联的上下文信息。基于候选文本表示和相关联的上下文信息,一个或多个机器学习机构被配置为基于一组候选可执行意图确定意图置信度得分。自然语言处理模块732 可基于所确定的意图置信度得分从一组候选可执行意图中选择一个或多个候选可执行意图。在一些示例中,还利用知识本体(例如,知识本体760)从一组候选可执行意图中选择一个或多个候选可执行意图。

[0232] 基于符号串搜索知识本体的其他细节在提交于2008年12月22日的名称为“Method and Apparatus for Searching Using An Active Ontology”的美国实用新型专利申请序列号12/341,743中有所描述,其全部公开内容以引用方式并入本文。

[0233] 在一些示例中,一旦自然语言处理模块732基于用户请求识别出可执行意图(或域),则自然语言处理模块732生成结构化查询以表示所识别的可执行意图。在一些示例中,结构化查询包括针对可执行意图的域内的一个或多个节点的参数,并且所述参数中的至少一些参数填充有用户请求中指定的特定信息和要求。例如,用户说“帮我在寿司店预订晚上7点的座位。”在这种情况下,自然语言处理模块732能够基于用户输入将可执行意图正确地识别为“餐厅预订”。根据知识本体,“餐厅预订”域的结构化查询包括参数诸如{菜系}、{时间}、{日期}、{同行人数}等。在一些示例中,基于语音输入和使用STT处理模块730从语音输入得出的文本,自然语言处理模块732针对餐厅预订域生成部分结构化查询,其中部分结构化查询包括参数{菜系=“寿司类”}以及{时间=“晚上7点”}。然而,在该示例中,用户话语包含不足以完成与域相关联的结构化查询的信息。因此,基于当前可用信息,在结构化查询中未指定其他必要参数诸如{同行人数}和{日期}。在一些示例中,自然语言处理模块732用所接收的上下文信息来填充结构化查询的一些参数。例如,在一些示例中,如果请求“附近的”寿司店,自然语言处理模块732用来自用户设备的GPS坐标来填充结构化查询中的{位置}参数。

[0234] 在一些示例中,自然语言处理模块732识别针对从STT处理模块730 接收的每个候选文本表示的多个候选可执行意图。另外,在一些示例中,针对每个所识别的候选可执行意图生成相应的结构化查询(部分地或全部地)。自然语言处理模块732确定针对每个候选可执行意图的意图置信度得分,并基于意图置信度得分对候选可执行意图进行排序。在一些示例中,自然语言处理模块732将所生成的一个或多个结构化查询(包括任何已完成的参数)传送到任务流处理模块736(“任务流处理器”)。在一些示例中,针对m个最佳(例如,m个排名最高的)候选可执行意图的一个或多个结构化查询被提供给任务流处理模块736,其中m为预先确定的大于零的整数。在一些示例中,将针对m个最佳候选可执行意图的一个或多个结构化查询连同对应的候选文本表示提供给任务流处理模块736。

[0235] 基于根据语音输入的多个候选文本表示所确定的多个候选可执行意图推断用户意图的其他细节在2014年6月6日提交的名称为“System and Method for Inferring User Intent From Speech Inputs”的美国实用新型专利申请序列号14/298,725中有所描述,其全部公开内容以引用方式并入本文。

[0236] 任务流处理模块736被配置为接收来自自然语言处理模块732的一个或多个结构化查询,在必要时完成结构化查询,以及执行“完成”用户最终请求所需的动作。在一些示例

中,完成这些任务所必需的各种过程在任务流模型754中提供。在一些示例中,任务流模型754包括用于获取来自用户的附加信息的过程,以及用于执行与可执行意图相关联的动作的任务流。

[0237] 如上所述,为了完成结构化查询,任务流处理模块736需要发起与用户的另外对话,以便获取另外信息和/或弄清可能有歧义的话语。当有必要进行此类交互时,任务流处理模块736调用对话流处理模块734来参与同用户的对话。在一些示例中,对话流处理器模块734确定如何(和/或何时)向用户请求附加信息,并且接收和处理用户响应。通过I/O处理模块728将问题提供给用户并从用户接收回答。在一些示例中,对话处理模块734经由音频和/或视频输出向用户呈现对话输出,并接收经由口头或物理(例如,点击)响应的来自用户的输入。继续上述示例,在任务流处理模块736调用对话流处理模块734来确定针对与域“餐厅预订”相关联的结构化查询的“同行人数”和“日期”信息时,对话流处理模块734生成诸如“一行几位?”和“预订哪天?”之类的问题传递给用户。一旦收到来自用户的回答,对话流处理模块734就用缺失信息填充结构化查询,或将信息传递给任务流处理模块736以根据结构化查询完成缺失信息。

[0238] 一旦任务流处理模块736已针对可执行意图完成结构化查询,则任务流处理模块736继续执行与可执行意图相关联的最终任务。因此,任务流处理模块736根据结构化查询中包含的特定参数来执行任务流模型中的步骤和指令。例如,针对可执行意图“餐厅预订”的任务流模型包括用于联系餐厅并实际上请求在特定时间针对特定同行人数的预订的步骤和指令。例如,使用结构化查询诸如:餐厅预订、{餐厅=ABC咖啡馆、日期=3/12/2012、时间=7pm、同行人数=5},任务流处理模块736可执行以下步骤:(1)登录ABC咖啡馆的服务器或诸如OPENTABLE®的餐厅预订系统,(2)以网站上的形式输入日期、时间和同行人数信息,(3)提交表单,以及(4)在用户的日历中形成针对预订的日历条目。

[0239] 在一些示例中,任务流处理模块736在服务处理模块738(“服务处理模块”)的辅助下完成用户输入中所请求的任务或者提供用户输入中所请求的信息性回答。例如,服务处理模块738代表任务流处理模块736发起电话呼叫、设定日历条目、调用地图搜索、调用用户设备上安装的其他用户应用程序或与所述其他应用程序进行交互,以及调用第三方服务(例如,餐厅预订门户网站、社交网站、银行门户网站等)或与第三方服务进行交互。在一些示例中,通过服务模型756中的相应服务模型指定每项服务所需的协议和应用程序编程接口(API)服务处理模块738针对服务访问适当的服务模型,并依据服务模型根据该服务所需的协议和API生成针对该服务的请求。

[0240] 例如,如果餐厅已启用在线预订服务,则餐厅提交服务模型,该服务模型指定进行预订的必要参数以及用于将必要参数的值传送至在线预订服务的API。在被任务流处理模块736请求时,服务处理模块738可使用存储在服务模型中的Web地址来建立与在线预订服务的网络连接,并将预订的必要参数(例如,时间、日期、同行人数)以根据在线预订服务的API的格式发送至在线预订界面。

[0241] 在一些示例中,自然语言处理模块732、对话处理模块734以及任务流处理模块736被共同且反复地使用,以推断并限定用户的意图、获取信息以进一步明确并细化用户意图、并最终生成响应(即,输出至用户,或完成任务)以满足用户的意图。所生成的响应是至少部分地满足用户意图的对语音输入的对话响应。另外,在一些示例中,所生成的响应被输出为

语音输出。在这些示例中,所生成的响应被发送到语音合成处理模块740 (例如,语音合成器),在语音合成处理模块中,可处理所生成的响应以将对话响应以语音形式合成。在其他示例中,所生成的响应是与满足语音输入中的用户请求相关的数据内容。

[0242] 在任务流处理模块736接收到来自自然语言处理模块732的多个结构化查询的示例中,任务流处理模块736首先处理所接收结构化查询的第一结构化查询以试图完成第一结构化查询和/或执行由第一结构化查询表示的一个或多个任务或动作。在一些示例中,第一结构化查询对应于排名最高的可执行意图。在其他示例中,第一结构化查询选自基于对应的语音识别置信度得分和对应的意图置信度得分的结合所接收的结构化查询。在一些示例中,如果任务流处理模块736在第一结构化查询的处理期间(例如,由于无法确定必要的参数)遇到错误,任务流处理模块736可继续选择和/或处理所接收结构化查询的与排名较低的可执行意图对应的第二结构化查询。例如基于对应候选文本表示的语音识别置信度得分、对应候选可执行意图的意图置信度得分、第一结构化查询中的缺失必要参数或它们的任何组合来选择第二结构化查询。

[0243] 语音合成模块740被配置为合成用于呈现给用户的语音输出。语音合成处理模块740基于数字助理提供的文本来合成人语音输出。例如,所生成的对话响应是文本串的形式。语音合成处理模块740将文本串转换成可听语音输出。语音合成处理模块740使用任何适当语音合成技术,以便从文本生成语音输出,包括但不限于:拼接合成、单元选择合成、双音素合成、特定于域的合成、共振峰合成、发音合成、基于隐马尔可夫模型(HMM)的合成以及正弦波合成。在一些示例中,语音合成处理模块740被配置为基于对应于这些字词的音素串来合成各个字词。例如,音素串与所生成的对话响应中的字词相关联。音素串存储在与人字词相关联的元数据中。语音合成处理模块740被配置为直接处理元数据中的音素串,以合成人语音形式的字词。

[0244] 在一些示例中,替代使用语音合成模块740(或除此之外),在远程设备(例如,服务器系统108)上执行语音合成,并且将合成的语音发送至用户设备以输出给用户。例如,这可发生在一些具体实施中,其中在服务器系统处生成数字助理的输出。而且由于服务器系统通常比用户设备具有更强的处理能力或更多的资源,其有可能获得比客户端侧合成将实现的质量更高的语音输出。

[0245] 有关数字助理的另外细节可见于提交于2011年1月10日的名称为“Intelligent Automated Assistant”的美国实用新型专利申请No.12/987,982 和提交于2011年9月30日的名称为“Generating and Processing Task Items That Represent Tasks to Perform”的美国实用新型专利申请No.13/251,088 中,其全部公开内容以引用方式并入本文。

#### [0246] 4. 发起和去激活虚拟助理的示例性技术

[0247] 图8A至图8D示出了用于基于音频输入和表示电子设备的运动的输入来在电子设备处发起虚拟助理会话的示例性技术。

[0248] 图8A示出了用户802的电子设备800。在一些示例中,设备800是设备104、106、200、400或600中的一者。在一些示例中,如图所示,设备800为可穿戴电子设备,诸如智能手表。在其他示例中,设备800为计算机、机顶盒、扬声器、电话或任何前述设备的组合。

[0249] 在图8A中,电子设备800处于已垂下的姿势。如下文所述,在一些示例中,用户802



将设备800从已垂下姿势抬起并提供音频输入,以在设备 800处发起虚拟助理会话并使得发起的虚拟助理会话响应音频输入。

[0250] 已垂下姿势是电子设备的多个姿势(例如,已垂下的姿势、正在举高的姿势、已举高的姿势和正在垂下的姿势)中的一个姿势。图8B示出了处于正在举高姿势中的设备800,图8C至图8E示出了处于已举高姿势中的设备800,并且图8F示出了处于正在垂下姿势中的设备800。然而,应当理解,图8A至图8F中示出的设备800的姿势仅仅是示例性的。也就是说,电子设备的已垂下姿势、正在举高姿势、已举高姿势和正在垂下姿势各自可分别不同于图8A至图8F中示出的已垂下姿势、正在举高姿势、已举高姿势和正在垂下姿势。类似地,电子设备在图8A至图8F中示出的姿势之间的转变也仅仅是示例性的。下文相对于图9讨论了用于确定电子设备的姿势和确定在姿势之间的转变的示例性技术。

[0251] 在本示例中,如图所示,由于电子设备800在用户802的手腕上并且用户802的手腕降低,电子设备800处于已垂下姿势。但如本文所讨论,在一些示例中,电子设备的已垂下姿势不同于所示的已垂下姿势。例如,如果用户802在其手腕上佩戴了设备800并且用户802正用他的手在键盘上打字,则设备800也处于已垂下姿势。又例如,如果用户802在其手腕上佩戴了设备800并且用户802正在躺下且期间其手腕保持在其一侧(例如,在臀部高度处),则设备800也处于已垂下姿势。因此,在一些示例中,如果未朝向或远离用户802的嘴部抬起设备800并且/或者远离用户 802的嘴部定位该设备(例如,定位成与嘴部相距0.5英尺、1英尺、1.5英尺或2英尺以上,定位成与嘴部相距用户802的一臂长等),则该设备处于已垂下的姿势中。

[0252] 在一些示例中,如图8A所示,设备800具有显示器804(例如,触摸屏)。当设备800处于已垂下姿势时,用户802不太可能正在看向显示器804或者正在提供口头请求。因此,在一些示例中,由于设备800处于已垂下姿势中,显示器804被关闭(例如,未显示)。在一些示例中,由于显示器804被关闭或者设备800的一个或多个麦克风继而被关闭(例如,不对音频输入进行采样)。这样,通过关闭在当前设备姿势下不太可能使用的设备部件来节约电池电力。然而,在其他示例中,当设备800处于已垂下姿势中时,显示器804和/或设备800的一个或多个麦克风被开启。

[0253] 在图8B中,用户802将设备800从已垂下姿势抬起到正在举高姿势(例如,用户802朝向其嘴部抬起他的手腕)。如下文所讨论,在一些示例中,此类抬起使得显示器804和设备800的一个或多个麦克风开启。开启一个或多个麦克风从而使得能够对音频输入(例如,下文相对于图8C所讨论的音频输入806“帕罗奥图天气怎么样?”)进行采样。如下文进一步所讨论的那样,在一些示例中,基于所提供的设备运动和音频输入806发起虚拟助理会话,并且发起的虚拟助理会话提供对音频输入806的响应。

[0254] 在一些示例中,当提供设备的运动时,设备检测表示其运动的输入,其在本文中称为运动输入。在一些示例中,通过设备的一个或多个运动传感器来检测运动输入。一个或多个运动传感器包括能够检测运动输入的任何一個或多个传感器,诸如一个或多个加速度计、一个或多个陀螺仪、一个或多个相机、一个或多个接近传感器、一个或多个超声传感器、一个或多个微波运动传感器等。

[0255] 在一些示例中,检测设备的运动输入包括使用第一处理器对运动输入进行采样。在一些示例中,第一处理器是与设备的主处理器分开的设备的低功率处理器。因此,在一些



示例中,第一处理器能够连续地对(例如,来自一个或多个陀螺仪和/或加速度计的)运动输入进行采样,而不显著损害电池寿命。例如,第一处理器通过将运动输入写入到存储缓冲器(例如,环缓冲器)中来连续地对运动输入进行采样。

[0256] 在一些示例中,电子设备使用电子设备的一个或多个麦克风来对音频输入进行采样。在一些示例中,一个或多个麦克风连续地对音频输入进行采样。在一些示例中,根据本文所述的技术供电的一个或多个麦克风对音频输入进行采样。

[0257] 在一些示例中,在检测运动输入的至少一部分时对音频输入进行采样。例如,如图8C所示,在检测运动输入(例如,使用一个或多个加速度计和/或一个或多个陀螺仪)时,设备800对音频输入806“帕罗奥图天气怎么样?”进行采样。尽管本示例示出了用户802在抬起设备800(例如,图8C)之后提供音频输入,但在一些示例中,用户802在抬起设备800(例如,图8B)时或者在抬起设备800之前(例如,图8A)提供音频输入806。

[0258] 在一些示例中,在设备处对音频输入进行采样包括使用设备的第一处理器(例如,与对设备的运动输入进行采样的处理器相同的处理器)对音频输入进行采样。例如,第一处理器能够通过将音频输入写入到存储缓冲器(例如,环缓冲器)中来连续地对音频输入(例如,从一个或多个麦克风)进行采样。

[0259] 在一些示例中,在设备处对音频输入进行采样包括使用第二处理器对音频输入进行采样。在一些示例中,第二处理器为设备的主处理器(例如,应用处理器)。

[0260] 在一些示例中,基于运动输入的至少一部分来激活第二处理器。在一些示例中,运动输入的一部分表示手腕举高和/或手腕旋转手势。例如,随着用户802将设备800从已垂下姿势抬起到正在举高姿势,用户802执行手腕举高手势。设备800的第一处理器检测手腕举高手势并激活第二处理器,第二处理器然后对音频输入806“帕罗奥图天气怎么样?”进行采样。用于确定手腕抬起和/或手腕旋转手势的示例性技术在2015年7月10日提交的名称为“Raise Gesture Detection in a Device”的美国专利申请No. 14/796,675中有所描述,其内容据此全文以引用方式并入本文。

[0261] 在一些示例中,激活第二处理器包括使得第二处理器退出低功率模式。在一些示例中,如果第二处理器在该模式下操作时可执行的操作的数量或类型受到限制,则第二处理器处于低功率模式下。在一些示例中,如果其正在与一些禁用的处理器电路一起操作,则第二处理器处于低功率模式下。在一些示例中,如果第二处理以相对于第二处理器在正常模式下操作时的值减小的时钟速度、电压或电流操作,则第二处理器处于低功率模式下。在一些示例中,第二处理器必须退出低功率模式以提供某种功能,诸如启动或执行应用程序、操作虚拟助理和/或对音频输入进行采样。

[0262] 在一些示例中,如图8B所示,显示器804基于设备800的运动输入的至少一部分来供电。在一些示例中,运动输入的一部分表示手腕举高和/或手腕旋转手势。例如,随着用户800将设备802从已垂下姿势抬起到正在举高姿势,用户802执行手腕举高手势和/或手腕旋转手势。设备800(例如,第一处理器)基于运动输入来确定手腕举高和/或手腕旋转手势,然后使得显示器804开启。例如,显示器804显示包括当前时间12:45的用户界面。

[0263] 在一些示例中,开启显示器804使得设备800的一个或多个麦克风开启。因此,在一些示例中,根据开启显示器804来对音频输入806进行采样。例如,在用户802发起手腕举高手势之后,设备800开启显示器804 以及一个或多个麦克风,并且使用该一个或多个麦克风

来对音频输入806“帕罗奥图天气怎么样?”进行采样。尽管在一些示例中将开启显示器804(并且从而开启一个或多个麦克风)描述为基于运动输入,但在其他示例中,开启显示器804基于除运动输入之外的输入或者除基于运动输入之外还基于其他输入。例如,在一些示例中,开启显示器804基于触觉输入(例如,在显示器804上的触摸,在设备800上的按钮按压)或音频输入(例如,接收足够振幅的音频输入)。

[0264] 作为上述技术的示例,设备800的第一处理器对运动输入进行连续地采样并且检测表示设备800的手腕举高和/或手腕旋转手势的运动输入。响应于检测到表示手腕举高和/或腕旋转手势的运动输入,开启设备800的显示器804和一个或多个麦克风,并且第一处理器激活第二处理器。第二处理器使用该一个或多个麦克风对音频输入(例如,“帕罗奥图天气怎么样?”)进行采样。这样,当用户802不太可能发出语音命令(例如,当设备800处于已垂下姿势中时),设备800的显示器804、一个或多个麦克风以及第二处理器被关闭(或者处于低功耗模式),从而节省电池功率。

[0265] 在图8C中,用户802将设备800抬起到已举高手势,如图所示。例如,用户802将设备800抬起(例如,从正在举高姿势抬起)到其嘴部附近。在一些示例中,设备800(例如,因为用户802不再抬起设备800)相对地稳定在已举高姿势中。如本文所讨论的那样,设备800还对用户802提供(例如,在将设备800抬起到已举高姿势之后不久由用户802提供)的音频输入806“帕罗奥图天气怎么样?”进行采样。设备800确定是否根据下文所讨论的示例性技术基于音频输入806并且基于运动输入来发起虚拟助理会话。如果设备800确定发起虚拟助理会话,则发起虚拟助理会话(例如,通过在显示器804上显示虚拟助理用户界面),如图所示。

[0266] 在一些示例中,采样的音频输入不包括口头触发。例如,音频输入806“帕罗奥图天气怎么样?”不包括用于发起虚拟助理的口头触发诸如“嘿,Siri”。因此,在一些示例中,用户简单地向虚拟助理发出请求(例如,在抬起设备时,或者在抬起设备之前或之后不久),使得虚拟助理发起并响应用户请求。

[0267] 在一些示例中,确定是否发起虚拟助理会话包括确定用于发起虚拟助理会话的一个或多个标准是否得到满足。在一些示例中,确定是否发起虚拟助理会话包括确定发起虚拟助理会话的概率。例如,如果用于发起虚拟助理会话的标准得到满足,则发起虚拟助理会话的概率增加。如果用于发起虚拟助理的标准未得到满足,则发起虚拟助理会话的概率降低。在一些示例中,如果发起虚拟助理会话的概率大于预定阈值并且/或者如果用于发起虚拟助理会话的一个或多个标准得到满足,则对发起虚拟助理会话进行确定。

[0268] 在一些示例中,上述确定中的一者或多者由运动模型、音频模型和/或虚拟助理发起模型(例如,举高至讲话模型)进行。例如,运动模型基于运动输入来确定发起虚拟助理会话的概率并且对发起虚拟助理会话进行确定。音频模型基于音频输入来确定发起虚拟助理会话的概率并且对发起虚拟助理会话进行确定。下文相对于图9讨论示例性运动模型、示例性音频模型和示例性虚拟助理发起模型。

[0269] 在一些示例中,分别由运动模型和音频模型确定的发起虚拟助理会话的相应概率和相应确定彼此分开。因此,在一些示例中,虚拟助理会话响应于运动模型和音频模型两者均确定发起虚拟助理会话而发起虚拟助理会话。在其他示例中,虚拟助理会话响应于运动模型和音频模型中一者确定发起虚拟助理会话而发起虚拟助理会话。这样,在一些示例中,

使用两个单独的确定(例如,一个基于运动输入和另一个基于音频输入)来发起虚拟助理会话。

[0270] 发起虚拟助理会话的示例性标准在下文中有所讨论。

[0271] 示例性标准包括电子设备是否在两个或更多个姿势之间转变。示例性转变是从正在举高姿势到已举高姿势,如图8B和图8C所示。在一些示例中,如果确定从正在举高姿势到已举高姿势的转变,则标准得到满足。在一些示例中,如果未确定从正在举高姿势到举高姿势的转变,则标准未得到满足。这样,检测从正在举高姿势到已举高姿势的转变(例如,使得设备靠近用户的嘴部的手臂/手腕移动)可为确定是否发起虚拟助理会话的重要标准。

[0272] 另一示例性转变是从已举高姿势到正在垂下姿势。图8E示出了电子设备的示例性正在垂下姿势。在一些示例中,如果确定从已举高姿势到正在垂下姿势的转变,则标准未得到满足。这样,如果设备800的运动包括用户手腕的正在垂下,例如在一些示例中,未发起虚拟助理。

[0273] 示例性标准包括姿势之间转变的相应时间是否在预定持续时间内。例如,确定从正在举高姿势转变到已举高姿势的第一时间(例如,当设备800 首先从正在举高姿势进入已举高姿势时),并且确定从已举高姿势转变到正在垂下姿势的第二时间(例如,当设备首先从已举高姿势进入正在垂下姿势时)。如果第一时间在第二时间的预定持续时间(例如,0.1秒、0.2 秒、0.3秒、0.4秒、0.5秒、0.6秒、0.7秒、0.8秒、0.9秒、1秒、1.5秒、2秒)内,则标准未得到满足。如果第一时间未在预定持续时间内,则标准得到满足。这样,如果正在举高姿势到已举高姿势的转变与已举高姿势到已垂下姿势的转变之间的持续时间相对较短,则不发起虚拟助理会话。在一些示例中,这防止因包括正在举高姿势到已举高姿势再到正在垂下姿势的转变的运动(例如,跑步,手臂锻炼等)而意外发起虚拟助理。应当理解,以这种方式将任何姿势之间的转变的相应时间进行比较与本文教导内容一致。

[0274] 示例性标准还包括姿势的相应持续时间。例如,确定已举高姿势的持续时间,并且将其与预定持续时间(例如,0.1秒、0.2秒、0.3秒、0.4 秒、0.5秒、0.6秒、0.7秒、0.8秒、0.9秒、1秒、1.5秒、2秒)进行比较。在一些示例中,如果已举高姿势持续时间超过预定持续时间,则标准得到满足。在一些示例中,如果举高姿势持续时间未超过预定持续时间,则标准未得到满足。这样,在确定是否发起虚拟助理会话时,考虑电子设备在已举高姿势中的稳定性(例如,当用户802将设备800举高至他或她的嘴部附近并且提供请求时,设备可能在已举高姿势中保持大于某预定持续时间的的时间)。尽管上述讨论集中于已举高姿势的持续时间,但在一些示例中,以类似的方式考虑任何姿势(例如,已垂下姿势、正在举高姿势、已举高姿势、正在垂下姿势)的持续时间。

[0275] 示例性标准还包括姿势之间的转变的相应持续时间。例如,确定正在举高姿势到已举高姿势的转变的持续时间并且将其与预定持续时间(例如,1秒、2秒、5秒)进行比较。如果正在举高姿势到已举高姿势的转变的持续时间超过预定持续时间,则标准未得到满足。如果正在举高姿势到已举高姿势的转变的持续时间未超过预定持续时间,则标准得到满足。以该方式将正在举高姿势到已举高姿势的转变的持续时间与预定持续时间进行比较可确保合适持续时间的正在举高姿势到已举高姿势的转变会发起虚拟助理会话。例如,如果用户802耗费太长时间(例如,3秒、4秒、5 秒、6秒、7秒、8秒、9秒、10秒)来将设备800从正在举高姿势举高到已举高姿势,则用户不太可能正在举高设备800以发出命令,因此未发起虚

拟助理会话。尽管上述讨论集中于正在举高姿势到已举高姿势的转变的持续时间,但在一些示例中,以类似的方式考虑任何两个姿势之间(例如,已垂下姿势到正在举高姿势、已垂下姿势到已举高姿势)的转变的持续时间。

[0276] 示例性标准还包括音频输入是否包括人类语音。例如,关于音频输入是否包括人类语音进行确定。如果音频输入包括人类语音,则标准得到满足。如果音频输入不包括人类语音,则标准未得到满足。这样,可避免响应于接收到不包括人类语音的音频输入(例如,背景噪声)意外发起虚拟助理会话。例如,如果用户802处于嘈杂环境中并且将设备800从正在举高姿势抬起到已举高姿势(例如,为了查看时间,响应消息等),在未向设备800发出口头请求的情况下,不会意外发起虚拟助理。

[0277] 在一些示例中,确定音频输入是否包括人类语音包括执行对音频输入的频域分析。例如,生成音频输入的声谱图(例如,使用傅里叶变换),并且分析音频输入的频谱分量以确定音频输入的至少一部分是否对应于人类语音。应当理解,在一些示例中,与本文教导内容一致地采用用于确定音频输入是否包括人类语音的任何合适的技术。

[0278] 示例性标准包括音频输入是否包括由设备的授权用户所讲的人类语音。在一些示例中,授权用户是与对音频输入进行采样的设备(例如,设备的所有者或主要用户)密切关联的用户。在一些示例中,授权用户是共享用户设备的受限的一组注册用户中的一个用户。确定音频输入是否包括由授权用户所讲的人类语音。如果音频输入包括由设备的授权用户所讲的人类语音,则标准得到满足。如果音频输入不包括由设备的授权用户所讲的人类语音,则标准未得到满足。这样,可防止响应于接收到不包括授权用户的语音的音频输入意外激活虚拟助理会话。例如,如果授权用户802处于包括背景语音(例如,多路重合噪声)的环境中并且将设备800从正在举高姿势抬起到已举高姿势(例如,为了查看时间,响应消息等),在未向设备800发出口头请求的情况下,不会意外发起虚拟助理。

[0279] 在一些示例中,确定音频输入是否包括由授权用户所讲的人类语音包括执行语音认证。例如,将音频输入的特征与一个或多个语音模型(例如,高斯混合模型、隐马尔可夫模型、神经网络等)的特征进行比较。在一些示例中,该一个或多个语音模型包括预定通用背景模型和授权用户的预定语音模型。在一些示例中,授权用户的预定语音模型基于授权用户的语音(例如,在对音频输入进行采样之前接收到的授权用户的语音)。应当理解,在一些示例中,与本文教导内容一致地采用用于确定音频输入是否包括授权用户所讲的人类语音的任何合适的技术。

[0280] 示例性标准包括与音频输入相关联的方向。例如,确定与音频输入相关联的方向。如果音频输入具有朝向设备(例如,朝向一个或多个麦克风或者朝向设备的屏幕)的方向,则标准得到满足。如果音频输入具有远离设备(例如,远离一个或多个麦克风或者远离设备的屏幕)的方向,则标准未得到满足。这样,可防止响应于接收到未指向设备(例如,来自站在用户800旁边的另一个人)的音频输入而意外激活虚拟助理。应当理解,在一些示例中,与本文教导内容一致地采用用于声音定位(例如,波束形成)的任何合适的技术。

[0281] 示例性标准包括音频输入与电子设备的接近度。例如,确定音频输入与电子设备的接近度并且将其与阈值接近度(例如,2cm、5cm、10cm、15cm或20cm)进行比较。如果所确定的接近度小于或等于阈值接近度,则标准得到满足。如果所确定的接近度大于阈值接近度,则标准未得到满足。这样,来自靠近设备的源的音频输入被认为是发起虚拟助理会话的

指示。来自不靠近设备的源(例如,来自不在设备800的阈值接近度内的另一个人)的音频输入可能会意外地发起虚拟助理会话。

[0282] 在一些示例中,确定音频输入与电子设备的接近度基于音频输入的振幅。在一些示例中,分析特定频谱范围内的音频输入的振幅(例如,人类语音的频率范围特性)以确定音频输入的接近度。在一些示例中,确定音频输入的接近度(另选地或除此之外还)基于音频输入的其他声学特征。例如,来自另外源的音频输入通常包括程度比来自更靠近的源的音频输入更大的回音。因此,在一些示例中,分析音频输入以确定回音的程度,从而确定音频输入的接近度。

[0283] 示例性标准包括音频输入的持续时间。例如,确定音频输入的持续时间,并且将其与预定持续时间(例如,0.1秒、0.2秒、0.3秒、0.4秒、0.5秒、0.6秒、0.7秒、0.8秒、0.9秒、1秒、1.5秒、2秒)进行比较。在一些示例中,音频输入的持续时间是期间音频输入在预定振幅和/或频率范围(例如,典型的人类语音和声音的振幅和/或频率范围)内的持续时间。如果音频输入的持续时间大于预定持续时间,则标准得到满足。如果音频输入的持续时间小于预定持续时间,则标准未得到满足。这样,在一些示例中,太短的音频输入(例如,不太可能为口头请求的音频输入)不会发起虚拟助理会话。例如,如果用户802在咳嗽或打喷嚏(例如,短的音频输入)时将设备800从正在举高姿势抬起到已举高姿势以覆盖他或她的嘴部,则咳嗽或打喷嚏不会意外发起虚拟助理。

[0284] 示例性标准基于与对音频输入进行采样和姿势之间的转变相关联的相应时间。例如,示例性标准包括与对音频输入进行采样和姿势之间的转变相关联的相应时间是否在预定持续时间内。例如,在第一时间(例如,设备800对输入806“帕罗奥图天气怎么样?”进行采样的时间)下对音频输入进行采样,并且设备从正在举高姿势转变到已举高姿势经历第二时间(例如,用户802将设备800抬起到图8C中所示的姿势中所花费的时间)。如果第一时间在第二时间的预定持续时间(例如,0.5秒、1秒、1.5秒、2秒、5秒)内,则标准得到满足。如果第一时间未在第二时间的预定持续时间内,则标准未得到满足。这样,在检测正在举高姿势到已举高姿势转变的适当持续时间内采样的音频输入被认为是发起虚拟助理的指示。在一些示例中,这防止基于未在大致相同的时间处分别提供的运动输入和音频输入而意外激活虚拟助理。例如,如果用户802将设备800从正在举高姿势抬起到已举高姿势并且然后在该抬起很久之后提供音频输入,则音频输入不太可能指向设备800。

[0285] 示例性标准包括对发起虚拟助理会话的单独确定的相应时间是否在预定持续时间内。例如,分别以第一时间和第二时间对发起虚拟助理会话进行单独的第一确定和第二确定(例如,用于发起虚拟助理会话的一个或多个标准得到满足的相应时间)。在一些示例中,运动模型进行第一确定并且音频模型进行第二确定(回想在一些示例中,运动和音频模型进行用于发起虚拟助理会话的相应确定)。如果第一时间在第二时间的预定持续时间(例如,0.5秒、1秒、1.5秒、2秒、5秒)内,则标准得到满足。如果第一时间未在第二时间的预定持续时间内,则标准未得到满足。

[0286] 因此,在一些示例中,发起虚拟助理会话基于检测用于发起虚拟助理会话的同步或大致同步的运动和音频输入。因此,即使运动输入和/或音频输入可各自适合用于发起虚拟助理(例如,分别满足用于发起虚拟助理的一个或多个标准),也可防止基于单独的运动输入和/或单独的音频输入而意外发起虚拟助理。因此,以这种方式提供虚拟助理激活提供

了有效且直观的发起虚拟助理会话的方式,同时还减少了错误发起的几率。例如,如图8A至图8C所示,为了发起设备800上的虚拟助理会话,用户802通过所示的姿势抬起设备800并且在抬起设备800之后不久(或者在抬起设备800时或在抬起设备800之前不久)提供音频输入806“帕罗奥图天气怎么样?”。

[0287] 在一些示例中,确定是否发起虚拟助理会话基于除上述运动和音频输入之外的输入或者除了基于上述运动和音频输入之外还基于其他输入。例如,确定这些其他输入是否满足用于发起虚拟助理会话的一个或多个标准。类似于上文所讨论的技术,在一些示例中,基于一个或多个标准是否得到满足来确定发起虚拟助理会话的概率。在一些示例中,如果概率超过阈值,则对发起虚拟助理会话进行确定。在一些示例中,用于发起虚拟助理会话的模型(例如,举高以讲话模型)进行这些确定中的一个或多个。下文相对于图9讨论用于发起虚拟助理会话的示例性模型。

[0288] 在一些示例中,电子设备包括在设备的前表面和/或后表面上一个或多个光传感器(例如,环境光线传感器、相机等)。将由光传感器检测到的光的量与阈值进行比较。如果光的量大于阈值,则用于发起虚拟助理的标准未得到满足。如果光的量小于阈值,则用于发起虚拟助理的标准得到满足。例如,当用户802将设备800抬起到已举高姿势以对其讲话时,设备800的前表面(例如,显示器804的表面)更靠近用户的嘴唇。因此,与设备800处于已垂下姿势时(例如,图8A)相比,设备802的前向光传感器检测到较少量的光(例如,小于预定阈值)。这样,较少量的光(例如,指示设备靠近用户的面部)被认为是发起虚拟助理会话的指示。

[0289] 在一些示例中,电子设备包括在设备的前表面和/或后表面上一个或多个颜色传感器(例如,相机、任何合适的RGB颜色传感器等)。将由一个或多个颜色传感器检测到的颜色与预定颜色(例如,人类皮肤、人类嘴唇等的颜色)进行比较。如果检测到的颜色与预定颜色匹配(例如,在预定波长内),则用于发起虚拟助理的标准得到满足。如果检测到的颜色不与预定颜色匹配,则用于发起虚拟助理会话的标准未得到满足。例如,当用户802将设备800抬起到已举高姿势以对其讲话时,设备800的前表面靠近用户的嘴唇。因此,设备800的前向颜色传感器检测匹配预定颜色的颜色。这样,对指示人类皮肤和/或嘴唇的颜色的检测被认为是发起虚拟助理会话的指示。

[0290] 在一些示例中,电子设备包括在设备的前表面和/或后表面上一个或多个相机。将由一个或多个相机检测到的面部特性(例如,下巴形状、嘴部形状、面部形状等)与预定面部特性(例如,人类面部特性、授权用户的面部特性)进行比较。如果检测到的面部特性与预定面部特性匹配,则用于发起虚拟助理的标准得到满足。如果检测到的面部特性不与预定面部特性匹配,则用于发起虚拟助理的标准未得到满足。例如,当用户802将设备800抬起到已举高姿势以对其讲话时,设备800的前表面朝向用户的面部。因此,设备800的前向相机检测与预定面部特性匹配的面部特性。

[0291] 在一些示例中,电子设备的一个或多个麦克风发射信号(例如,超声信号),并且该一个或多个麦克风检测基于发射信号反射的信号。在一些示例中,检测到的信号用于确定用于发起虚拟助理的一个或多个标准是否得到满足。例如,如果检测到的信号指示设备的阈值接近度(例如,2cm、5cm、10cm、20cm)内的对象(例如,人面部、人嘴唇),则标准得到满足。如果检测到的信号指示未在设备的阈值接近度内的对象,则标准未得到满足。这样,对

接近对象的检测被认为是发起虚拟助理会话的指示。

[0292] 在一些示例中,电子设备包括在设备的前表面和/或后表面上一个或多个接近传感器。例如,该一个或多个接近传感器包括任何合适的接近传感器(例如,电容传感器、光电传感器、感应传感器、飞行时间传感器等)。分析由一个或多个接近传感器检测到的输入,以确定其是否表示用于发起虚拟助理会话的指示。例如,如果来自一个或多个接近传感器的输入指示设备正在接近对象(例如,用户的面部/嘴部),则标准得到满足。如果来自一个或多个接近传感器的输入指示设备正在远离对象移动,则标准未得到满足。又如,如果来自一个或多个接近传感器的输入指示设备在对象的阈值接近度(例如,2cm、5cm、10cm、20cm)内,则标准得到满足。

[0293] 在一些示例中,在确定是否发起虚拟助理时,考虑虚拟助理交互上下文(例如,用户的上下文(例如,正在睡觉、在汽车中等)和关于先前虚拟助理状态的信息)。虚拟助理交互上下文提供了另外的设备和/或用户信息,这可有助于确定是否发起虚拟助理会话。例如,如果用户正在睡觉(例如,如由当天时刻、运动输入、用户指示等所确定的那样),与用户清醒时相比,更不易发起虚拟助理会话。

[0294] 下文讨论了示例性虚拟助理交互上下文。在一些示例中,基于虚拟助理交互上下文,确定发起虚拟助理会话的概率。在一些示例中,这些确定中的任一者由用于发起虚拟助理的模型(例如,举高以讲话模型)来制备。

[0295] 示例性虚拟助理交互上下文包括先前发起虚拟助理会话的时间。在一些示例中,如果先前发起虚拟助理会话的时间在当前时间的阈值持续时间(例如,30秒、1分钟、5分钟等)内,则发起虚拟助理会话的概率增大。在一些示例中,如果先前发起虚拟助理会话的时间在当前时间的阈值持续时间内,则发起虚拟助理会话的概率降低。

[0296] 示例性虚拟助理交互上下文包括先前的传感器输入(例如,上文所讨论的传感器输入中的任一者)是否表示迹近错失发起虚拟助理会话或者是否表示几乎发起虚拟助理会话。例如,基于一个或多个先前的传感器输入,确定发起虚拟助理会话的概率。如果该概率以小于第一量(例如,少量)的量超过发起虚拟助理会话的阈值概率,则该一个或多个先前的传感器输入表示迹近错失。如果该概率以最多为第二量(例如,少量)的量小于发起虚拟助理会话的阈值概率,则该一个或多个先前的传感器输入表示几乎发起。

[0297] 示例性虚拟助理交互上下文包括与发起虚拟助理会话的迹近错失相关联的时间。在一些示例中,将与发起虚拟助理会话的迹近错失相关联的时间与当前时间(例如,对当前输入进行采样的时间)进行比较。例如,与迹近错失相关联的时间是对一个或多个先前的传感器输入(表示迹近错失)进行采样的时间或者确定与一个或多个先前的传感器输入相关联的概率的时间。在一些示例中,如果与发起虚拟助理会话的迹近错失相关联的时间在当前时间的预定持续时间(例如,5秒、15秒、30秒、1分钟、5分钟等)内,则发起虚拟助理会话的概率增大。这样,在一些示例中,迹近错失发起虚拟助理会话的最新的先前输入指示基于当前输入应该更易于发起虚拟助理会话。在一些示例中,如果与发起虚拟助理会话的迹近错失相关联的时间在当前时间的预定持续时间内,则发起虚拟助理会话的概率降低。应当理解,以类似方式(例如,以增大或降低发起虚拟助理的概率)将与几乎发起虚拟助理会话相关联的时间与当前时间进行比较与本文的教导内容一致。

[0298] 示例性虚拟助理交互上下文包括与先前提供给虚拟助理的输入(例如,文本或口



头输入) 相关联的域。例如, 确定与先前输入相关联的域是否为预定类型的域。如果域为预定类型的域, 则发起虚拟助理会话的概率增大。如果域不为预定类型的域, 则发起虚拟助理会话的概率降低(或者未变化)。

[0299] 在一些示例中, 该预定类型的域为被配置为启用多轮虚拟助理交互的域。例如, 多轮虚拟助理交互包括其中将输入提供给虚拟助理并且虚拟助理通过引出进一步的输入来响应输入的虚拟助理交互。例如, “发送消息” 域被配置为响应于所提供的输入“发送消息” 启用响应“向谁发送消息?” (例如, 引出进一步的用户输入)。因此, 如果与先前的输入相关联的域被配置为启用多轮交互, 则可能用户旨在发起虚拟助理会话以提供进一步的输入。因此, 在一些示例中, 如果与先前输入相关联的域被配置为启用多轮交互, 则发起虚拟助理会话的概率增大。

[0300] 示例性虚拟助理交互上下文包括虚拟助理的先前状态。例如, 虚拟助理的状态表示虚拟助理是否正在等待进一步的输入。例如, 如果虚拟助理先前正在等待对输出“向谁发送消息?” 的响应, 则先前的状态为第一状态。例如, 如果虚拟助理先前未在等待响应(例如, 虚拟助理提供了输出, 诸如“今天是晴天”), 则先前的状态为第二状态。在一些示例中, 确定虚拟助理的先前状态。在一些示例中, 如果先前状态为第一状态, 则发起虚拟助理的概率增大。在一些示例中, 如果先前状态为第二状态, 则发起虚拟助理的概率降低。这样, 在一些示例中, 如果先前的虚拟助理状态指示虚拟助理正在等待响应, 则更易于发起虚拟助理。

[0301] 示例性虚拟助理交互上下文包括上下文信息(例如, 上文讨论的任何上下文信息)。在一些示例中, 从上下文信息来确定设备和/或用户的活动上下文(例如, 用户是否正在跑步、是否在汽车中、是否正在睡觉、是否在家中、是否在工作、是否正在旅行、是否在电影院、是否正在开会等)。某些活动上下文可指示用户更可能或者更不可能发起虚拟助理。因此, 在一些示例中, 基于确定的活动上下文, 发起虚拟助理的概率增大或降低。例如, 如果确定了用户正在开会(例如, 基于当前时间在用户的时间安排中的会议时间段内), 则发起虚拟助理的概率降低(例如, 因为用户可能不希望在会议期间被发起的虚拟助理打扰)。

[0302] 在一些示例中, 根据发起虚拟助理会话的确定, 发起虚拟助理会话。例如, 如图8C所示, 发起虚拟助理会话包括显示与虚拟助理会话相关联的用户界面(例如, 在显示器804上)。然而, 在一些示例中, 发起虚拟助理会话不包括显示虚拟助理用户界面。例如, 在一些示例中, 在显示器804 被关闭时, 发起虚拟助理会话。发起虚拟助理的另外示例性描述在2015年 8月31日提交的名称为“Virtual Assistant Activation”的美国专利申请No. 14/841,449中提供, 该专利申请的内容据此全文以引用方式并入本文。

[0303] 在一些示例中, 发起虚拟助理会话包括提供音频输出(例如, 设备 800提供指示虚拟助理正在侦听的音频输出)。在一些示例中, 发起虚拟助理会话包括提供触感输出(例如, 嗡嗡声或振动)。在一些示例中, 发起虚拟助理会话包括给被配置为操作虚拟助理的处理器电路(例如, 主处理器的电路)供电。在一些示例中, 发起虚拟助理包括发起一个或多个程序或模块(例如, 数字助理模块726)。

[0304] 在一些示例中, 根据不发起虚拟助理会话的确定, 不发起虚拟助理会话。例如, 未执行被包括在发起虚拟助理会话中的上述过程中的任一者并关闭显示屏, 并且/或者显示(或者保持显示)不为虚拟助理用户界面的用户界面(例如, 显示当前时间的用户界面)。

[0305] 图8D示出了设备800上发起的虚拟助理, 其提供对音频输入806的响应。具体地讲,



在用户802已经提供音频输入806“帕罗奥图天气怎么样?”(并且设备802已经对其进行采样)之后,发起的虚拟助理会话提供响应“气温70度,晴天”。

[0306] 因此,在一些示例中,根据发起虚拟助理会话的确定,基于音频输入来确定用户意图,基于该用户意图执行任务,并且提供与该任务相关联的输出。例如,如图8D所示,确定用于音频输入806“帕罗奥图天气怎么样?”的用户意图,执行任务(例如,检索天气信息),并且提供与该任务相关联的输出。例如,设备800在显示器804上显示响应“气温70度,晴天”。任选地,设备800提供音频输出808“气温70度,晴天”。上文相对图7A至图7C讨论了用于确定用户意图以及基于所确定的用户意图来执行任务的示例性技术。

[0307] 在一些示例中,根据在设备处发起虚拟助理会话,提供运动输入和音频输入以训练用于发起虚拟助理的模型(例如,举高以讲话模型)。在一些示例中,提供另外的输入(例如,除运动输入和音频输入之外的输入)和/或虚拟助理交互上下文来对模型进行训练(例如,如果确定基于另外的输入和/或上下文来发起虚拟助理会话)。例如,模型在设备上和/或在另一个电子设备(例如,服务器系统108)上实现,并且使用运动输入、音频输入、另外的输入和/或虚拟助理交互上下文作为训练数据来对模型进行训练。因此,成功发起虚拟助理会话的输入和/或虚拟助理交互上下文用于训练模型,从而增大成功发起的可能性并且降低意外发起的可能性。下文相对于图9讨论了训练用于发起虚拟助理的模型的示例性技术。

[0308] 在一些示例中,根据在设备处未发起虚拟助理会话,未提供输入和/或虚拟助理交互上下文来训练用于发起虚拟助理的模型。例如,丢弃输入和/或虚拟助理交互上下文。

[0309] 在一些示例中,根据提供与任务相关联的输出,执行提供输入和/或虚拟助理交互上下文来训练用于发起虚拟助理的模型。例如,在设备800提供音频输出808“气温70度,晴天”之后,提供运动输入和音频输入806来训练模型。以这种方式提供运动输入和音频输入可进一步确保适当的输入用于训练模型。例如,如果用户802意外发起虚拟助理会话,则他可在设备800提供输出808之前去激活虚拟助理会话。因此,如果提供了输出808,则可能用户旨在发起虚拟助理会话。因此,提供用于发起虚拟助理会话的正确输入来训练模型。

[0310] 在一些示例中,根据确定发起虚拟助理会话所持续的时间大于预定持续时间(例如,1秒、2秒、3秒、4秒、5秒、6秒、7秒、8秒、9秒、10秒)(例如,显示虚拟助理用户界面所持续的时间大于预定持续时间),执行提供输入和/或虚拟助理交互上下文来训练用于发起虚拟助理的模型。以这种方式提供输入和/或上下文可进一步确保适当的输入和/或上下文用于训练模型。例如,如果用户802意外发起虚拟助理会话,则他可在经过预定持续时间之前去激活虚拟助理会话。因此,如果经过预定持续时间,则可能用户802旨在发起虚拟助理会话。因此,提供用于发起虚拟助理会话的正确输入和/或上下文来训练模型。下文相对于图8D和图8E描绘了用于去激活虚拟助理会话的示例性技术。

[0311] 在一些示例中,提供意外发起虚拟助理会话的输入和/或虚拟助理交互上下文(例如,不正确的输入和/或上下文)来训练用于发起虚拟助理的模型。例如,不正确的输入和/或上下文训练模型,以使其不会基于未来不正确的输入和/或上下文而意外发起虚拟助理会话。因此,在一些示例中,根据确定输入和/或上下文对于发起虚拟助理会话来说是不正确的,执行提供输入和/或虚拟助理交互上下文来训练用于发起虚拟助理会话的模型。

[0312] 在一些示例中,确定一个或多个输入和/或上下文对于发起虚拟助理会话来说是否是不正确的包括确定是否基于该一个或多个输入和/或上下文来去激活虚拟助理会话。

例如,如果确定了根据下文所讨论的技术基于运动和/或音频输入来去激活虚拟助理会话,则确定运动输入和/或音频输入对于发起虚拟助理来说是不正确的。

[0313] 在一些示例中,确定一个或多个输入和/或上下文对于发起虚拟助理会话来说是否是不正确的包括确定(例如,通过数字助理模块726)用户反馈是否指示对虚拟助理的发起不满。在一些示例中,指示对虚拟助理会话不满的用户反馈包括提供给发起的虚拟助理的言语和/或文本用户输入(例如,用户输入诸如“现在不行”、“你怎么出现了?”“消失”等)。在一些示例中,指示对虚拟助理的发起不满的用户反馈包括在发起虚拟助理会话之后提供的(例如,在之后不久(例如,0.5秒、1秒、2秒、3秒、4 秒、5秒)提供的)去激活虚拟助理会话的用户输入。在一些示例中,去激活虚拟助理会话的用户输入包括按钮按压、口头输入和/或运动输入。

[0314] 图8E示出了用于去激活虚拟助理会话的示例性技术。具体来讲,用户 802已将设备800从正在举高姿势抬起到已举高姿势并且提供音频输入“TINA在这个星期天外出度假,她和她的丈夫一起去夏威夷看火山口和海滩上的海豚.....”。在该示例中,音频输入未指向设备800,而是指向用户 802正在与之交谈的另一个人(例如,站在用户802的左边并且用户802正看向其的人)。但是基于运动输入和音频输入,意外发起了虚拟助理会话。例如,在显示器804上显示虚拟助理用户界面。在本示例中,在根据下文所讨论的技术发起虚拟助理会话之后去激活虚拟助理会话(例如,显示器804停止显示虚拟助理用户界面)。

[0315] 在一些示例中,去激活虚拟助理会话包括停止显示与虚拟助理会话相关联的用户界面。在一些示例中,去激活虚拟助理会话包括放弃响应音频输入。在一些示例中,去激活虚拟助理会话包括关闭显示器。在一些示例中,去激活虚拟助理会话包括提供音频和/或触感输出。在一些示例中,去激活虚拟助理会话包括关闭被配置为操作虚拟助理的处理器电路(例如,主处理器的电路)。在一些示例中,去激活虚拟助理包括去激活一个或多个程序或模块(例如,数字助理模块726)。

[0316] 在一些示例中,在显示与虚拟助理会话相关联的用户界面之前执行去激活虚拟助理会话以及/或者确定是否去激活虚拟助理会话。在其他示例中,在显示与虚拟助理会话相关联的用户界面之后执行去激活虚拟助理会话以及/或者确定是否去激活虚拟助理会话。

[0317] 在一些示例中,确定是否将虚拟助理会话去激活包括确定用于将虚拟助理会话去激活的一个或多个标准是否得到满足。在一些示例中,确定是否将虚拟助理会话去激活包括确定将虚拟助理会话去激活的概率。例如,如果用于将虚拟助理会话去激活的标准得到满足,则将虚拟助理会话去激活的概率增加。如果用于将虚拟助理去激活的标准未得到满足,则将虚拟助理会话去激活的概率降低。在一些示例中,如果将虚拟助理会话去激活的概率大于预定阈值并且/或者如果用于将虚拟助理会话去激活的一个或多个标准得到满足,则确定将虚拟助理会话去激活。在一些示例中,上述确定中的任何一个是由用于将虚拟助理会话去激活的一个或多个模型(例如,音频模型、运动模型、举高以讲话模型)进行的。

[0318] 另外,在一些示例中,下面讨论的用于将虚拟助理会话去激活的任何技术以与确定是否发起虚拟助理会话类似的方式应用。例如,如果下面描述的用于将虚拟助理会话去激活的标准中的一个或多个(例如,如由运动模型、音频模型和/或举高以讲话模型确定的)得到满足,则不发起虚拟助理和/或发起虚拟助理会话的概率降低。类似地,在一些示例中,上面讨论的用于确定不发起虚拟助理会话的任何技术以与确定是否将虚拟助理会话去激

活类似的方式应用。例如,如果上面描述用于发起虚拟助理会话的标准中的一个或多个未得到满足,则将虚拟助理会话去激活和/或将虚拟助理会话去激活的概率增加。

[0319] 现在讨论用于将虚拟助理会话去激活的示例性标准。

[0320] 示例性标准包括音频输入的内容是否指示音频输入指向设备。在一些示例中,不太可能指向设备的音频输入应将虚拟助理会话去激活。被分析(例如,由设备800和/或由服务器系统108分析)以确定音频输入是否指向设备的示例性音频内容包括语言内容(例如,句子结构、句子长度,是否包括单词和/或关键词,句子类型(例如,疑问句或陈述句)等,以及设备是否可以执行音频输入中指定的任务。例如,基于其语言内容,确定音频输入“Tina在这个星期六外出度假,她和她的丈夫一起去夏威夷看火山口和海滩上的海豚……”不太可能直接指向设备。根据确定音频输入不太可能指向设备,用于将虚拟助理会话去激活的标准得到满足。因此,在一些示例中,虚拟助理会话被去激活(例如,服务器系统108向设备800发送控制信号,指示设备800将虚拟助理会话去激活)。

[0321] 在一些示例中,根据确定音频输入很可能指向设备,用于将虚拟助理会话去激活的标准未得到满足。因此,在一些示例中,虚拟助理会话不被去激活(例如,显示虚拟助理用户界面,设备对音频输入进行采样等)。

[0322] 示例性标准包括是否可以确定音频输入的意图(例如,由设备800和/或服务器系统108以足够的置信度确定)。例如,某些音频输入(例如,咳嗽、打喷嚏、长句子、诸如“嗯”、“哇”的短表达等)不能确定各自的意图,或者所确定的相应意图的相应置信度得分低(例如,小于预定阈值)。因此,在一些示例中,这些音频输入应将虚拟助理会话去激活。因此,在一些示例中,如果不能确定音频输入的意图,则满足标准。在一些示例中,如果音频输入的意图的置信度得分低,则满足标准。在一些示例中,如果确定了音频输入的意图,则不满足标准。在一些示例中,如果音频输入的意图的置信度得分高(例如,大于预定阈值),则不满足标准。

[0323] 示例性标准包括音频输入的持续时间。例如,短音频输入(例如,咳嗽、打喷嚏、诸如“嗯”、“哇”的短表达等)不太可能指向设备,因此这样的音频输入应该将虚拟助理会话去激活。因此,在一些示例中,将音频输入持续时间与预定持续时间(例如,0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1秒)进行比较。如果音频输入持续时间小于预定持续时间,则满足标准。如果音频输入持续时间大于预定持续时间,则不满足标准。

[0324] 图8F还示出了用于将虚拟助理会话去激活的示例性技术。具体地讲,在图8F中,用户802已将设备800从正在举高姿势提升到已举高姿势,并提供音频输入“星期六有个重要会议”。在该示例中,音频输入未指向设备800,而是指向用户802正在与之交谈的另一个人。但是基于运动输入和音频输入,意外发起了虚拟助理会话。用户802注意到这种意外的发起(例如,在显示器804上看到虚拟助理界面),并将设备800从已举高姿势下降到正在垂下姿势,如图所示。例如,用户800将设备从已举高姿势移动到远离他或她的嘴部/脸部。根据设备800的这种移动,虚拟助理会话被去激活,如图所示。

[0325] 因此,用于将虚拟助理会话去激活的示例性标准包括运动输入是否指示从已举高姿势到正在垂下姿势的转变。在一些示例中,如果运动输入指示从已举高姿势到正在垂下姿势的转变,则满足标准。如果运动输入包括未指示从已举高姿势到正在垂下姿势的转变,则不满足标准。这样,在一些示例中,将设备从已举高姿势下降到正在垂下姿势使所发起的

虚拟助理会话去激活。

[0326] 示例性标准包括与将设备从已举高姿势中下降相关联的时间是否在与音频输入中的语音结束相关联的时间之前。例如,确定与音频输入“星期六有个重要会议”的语音结束相关联的第一时间。确定与从已举高姿势到正在垂下姿势的转变相关联的第二时间。将第一时间与第二时间进行比较。在一些示例中,如果第二时间在第一时间之前,则满足标准。在一些示例中,如果第二时间在第一时间之后,则不满足标准。这样,如果用户 802在他说完之前放下他的手腕,则该语音不太可能指向设备800,因此虚拟助理会话被去激活。用于确定音频输入中的语音结束(例如,终止)的示例性技术在2015年9月4日提交的标题为“Context-Based Endpoint Detection”的美国专利申请14/846,667和2015年4月30日提交的标题为“Robust End-Pointing of Speech Signals Using Speaker Recognition”的美国专利申请14/701,147中有所描述,这两个申请的内容全文以引用方式并入本文。

[0327] 在一些示例中,使用从已垂下姿势到正在举高姿势的转变来确定音频输入中的语音终点。具体地讲,在一些示例中,与设备从已举高姿势到已垂下姿势的转变相关联的时间用于确定语音终点。在一些示例中,随后处理终止的语音输入以确定用户意图,并且根据上面参考图7A至C讨论的技术,基于所确定的用户意图来执行任务。因此,在一些示例中,用户802 通过放下设备800来指示他已完成讲话,如图8F所示。

#### [0328] 5. 用于传输通信的示例性技术

[0329] 在一些示例中,音频输入和运动输入用于调用除了发起虚拟助理之外的设备功能。这些功能包括例如在与外部设备的通信会话(例如,“无线对讲机”会话)中接听电话呼叫和打开通信信道。例如,如果设备参与与一个或多个其他(例如,外部)设备的通信会话,则用户可以如上所述在他或她的嘴部附近举起设备并提供音频输入。已经确定,运动输入和音频输入适合于发起与一个或多个其他设备的通信,并且音频输入被提供(例如,传输)到一个或多个其他设备。这样,为了与参与与用户的设备的通信会话的其他用户/设备通信,用户简单地将设备举高到他或她的嘴部附近,并提供传输到一个或多个其他设备的音频输入。

[0330] 由于在一些示例中,运动输入和音频输入的相同或类似组合可以调用多个设备功能(例如,发起虚拟助理,接听电话呼叫),所以确定调用哪个设备功能可能是重要的。否则,例如,用户意图发起虚拟助理会话的音频输入和运动输入可能意外地导致执行另一功能(例如,使另一用户的设备输出音频)。因此,在一些示例中,设备的上下文用于确定音频输入和运动输入应调用哪个设备功能。这样,如果确定的上下文适合于发起虚拟助理会话,则在确定是否发起虚拟助理会话时会考虑音频输入和运动输入。例如,如果确定的上下文适合于接听电话呼叫,则在确定是否接听电话呼叫时会考虑音频输入和运动输入。

[0331] 因此,在一些示例中,确定电子设备的上下文。确定上下文是否是表示与外部电子设备的通信的预定类型的上下文。在一些示例中,如果上下文是预定类型的上下文,则采样的音频输入和检测到的运动输入不用于确定是否发起虚拟助理会话,而是用于确定是否执行另一设备功能。在一些示例中,如果上下文不是预定类型的上下文,则根据上文所述的技术将采样的音频输入和检测到的运动输入用于确定是否发起虚拟助理会话。

[0332] 例如,当以下条件中的一个或多个得到满足时,设备的上下文是表示与外部电子设备的通信的预定类型的上下文。

[0333] 示例性条件包括设备当前正从外部电子设备接收通信(例如,电话呼叫、文本消息、电子邮件、语音邮件等)。如果设备当前正在接收通信,则满足条件。如果设备当前没有接收通信,则不满足条件。这样,如果设备当前正在接收电话呼叫并且用户提供运动输入和音频输入,则在确定是否接听电话呼叫时考虑音频输入和运动输入。例如,在确定是否发起虚拟助理会话时不考虑音频输入和运动输入。例如,如果确定接听电话呼叫,则将音频输入传输到外部设备(例如,呼叫者的设备)。这样,用户可以简单地举起他或她的设备并提供音频以接听电话呼叫并使音频被传输到呼叫者的设备。

[0334] 示例性条件包括设备已经从外部电子设备接收了通信(例如,电话呼叫、文本消息和电子邮件、语音邮件等)。例如,当设备接收到通信时,条件在接收到通信之后的短持续时间(例如,1、2、3、4、5、10、15、30 秒)内得到满足。在该短持续时间到期后,不满足条件。因此,在一些示例中,如果用户在接收到消息(例如,文本或电子邮件)之后不久提供运动输入和音频输入,则在确定是否响应该消息时考虑运动输入和音频输入。在确定是否发起虚拟助理会话时不考虑运动输入和音频输入。这样,如果用户接收到消息(例如,文本或电子邮件),则用户可以通过简单地举高他或她的设备并在接收到消息之后不久提供音频输入来响应该消息。在一些示例中,音频输入被转换为文本,并且该文本被传输到消息发送者的设备。

[0335] 示例性条件包括设备当前参与与外部电子设备的通信会话(例如,“无线对讲机”会话)。如果设备当前参与该通信会话,则满足条件。如果设备当前未参与该通信会话,则不满足条件。因此,如果设备当前参与通信会话(例如,被配置为提供该通信会话的一个或多个程序正在执行)并且用户提供运动输入和音频输入,则在确定是否传输通信时考虑运动输入和音频输入。这样,用户可以简单地通过举高他或她的设备并说出该音频来将音频传输到参与与用户的设备的通信会话的另一设备。

[0336] 在一些示例中,如果上下文是表示与外部电子设备的通信的预定类型的上下文,则基于采样的音频输入和检测到的运动输入确定是否将与该音频输入相关联的通信传输到外部电子设备。在一些示例中,确定是否传输通信的执行类似于上面讨论的用于确定是否发起虚拟助理会话(例如,基于音频输入、运动输入和/或其他输入)的技术。也就是说,如果运动输入、音频输入和/或其他输入适合于发起虚拟助理会话,则该运动输入、音频输入和/或其他输入同样适合于引起通信的传输。在一些示例中,音频输入、运动输入和/或其他输入被提供给模型(例如,举高以讲话模型 924),并且该模型确定是否传输通信。

[0337] 在一些示例中,如果上下文不是表示与外部电子设备的通信的预定类型的上下文,则基于音频输入和运动输入确定是否发起虚拟助理会话(例如,根据上面讨论技术)。

[0338] 在一些示例中,根据确定将与音频输入相关联的通信传输到外部电子设备,将与音频输入相关联的通信传输到外部电子设备。在一些示例中,与音频输入相关联的通信包括表示音频输入的音频内容的信号(例如,如果用户正在接听电话呼叫或参与通信会话)。在一些示例中,音频输入被转换为文本,并且与音频输入相关联的通信包括表示该文本的信号(例如,如果用户正在回复文本或电子邮件消息)。

[0339] 在一些示例中,根据确定不将与音频输入相关联的通信传输到外部电子设备,放弃将与音频输入相关联的通信传输到外部电子设备。例如,丢弃所采样的音频输入和检测到的运动输入和/或不生成要传输到外部电子设备的信号。

[0340] 在一些示例中,在确定将与音频输入相关联的通信传输到外部电子设备之后,用户放下设备(例如,从已举高姿势,如图8F所示)。在一些示例中,放下设备导致根据设备的状况执行一个或多个功能(例如,设备参与通信会话,设备参与电话呼叫,设备接收消息等)。例如,在一些示例中,当设备参与电话呼叫时放下设备导致电话呼叫终止。因此,为了终止电话呼叫(例如,在接听电话呼叫之后),用户只需放下该设备。

[0341] 又如,在用户已经举高设备以回复消息之后放下设备导致消息被发送(例如,已传输)。这样,为了回复消息,用户可以举高设备并发出对该消息的回复。用户的回复被转换为文本,并且放下设备导致消息被发送(例如,发送给消息的发送者)。

[0342] 又如,当设备参与与一个或多个外部设备的通信会话(例如,无线对讲机会话)时放下设备将关闭与一个或多个外部设备的通信信道(例如,这样,在设备处采样的音频输入不会被传输到一个或多个外部设备)。这样,为了将音频发送到参与与用户设备的通信会话的另一设备,用户举高他或她的设备并提供音频。然后用户可以放下他或她的设备以停止向另一设备传输音频。

[0343] 在一些示例中,提供音频输入、运动输入和/或确定的上下文以训练模型(例如,举高以讲话模型924)。例如,提供引起通信传输的音频输入、运动输入和/或确定的上下文来训练模型(例如,基于未来正确的音频输入、运动输入和/或上下文来成功地引起通信的传输)。又如,提供未引起通信传输的音频输入、运动输入和/或确定的上下文来训练模型(例如,以防止基于未来不正确的音频输入、运动输入和/或上下文的意外通信传输)。

[0344] 在一些示例中,确定是否将与音频输入相关联的通信传输到外部电子设备的模型(例如,模型924)是确定是否发起虚拟助理会话的相同模型。因此,训练这样的模型允许模型基于所提供的运动输入、所提供的音频输入以及上下文来正确地确定是否发起虚拟助理会话或者是否执行另一设备功能。如所讨论的,在运动输入和音频输入的相同或相似组合可以调用多个设备功能的示例中,这可能是期望的。

[0345] 在一些示例中,在确定将与音频输入相关联的通信传输到外部电子设备之后,通信的传输被阻止。这允许用户防止意外通信传输(例如,如果设备不正确地确定基于音频输入和运动输入而传输通信)。例如,在确定传输通信之后,设备在短持续时间内(例如,1秒、2秒、3秒等)输出指示(例如,视觉指示、音频指示和/或触觉指示),指示设备将传输通信(例如,在该短持续时间到期之后)。在该短持续时间期间,如果确定要阻止通信的传输,则通信的传输被阻止(例如,不传输通信和/或丢弃通信)。

[0346] 在一些示例中,确定是否阻止通信的传输的执行类似于用于确定是否将虚拟助理会话去激活的上述技术。例如,用户在该短持续时间内放下设备(例如,如图8F所示)以防止通信的传输。在一些示例中,如果通信的传输被阻止,则确定要从其传输通信的输入和上下文被认为是不正确的输入和上下文。在一些示例中,提供不正确的输入和/或上下文来训练模型(例如,以防止将来不正确的传输通信的确定)。

[0347] 图9示出了根据一些示例的被配置为处理音频输入和运动输入的系统900的框图。在一些示例中,系统900在独立计算机系统(例如,设备104、106、200、400、600或800中的任何一个)上实现。在一些示例中,系统900跨多个设备分布。例如,系统900的部件和功能中的一些被划分成服务器部分和客户端部分,其中客户端部分位于一个或多个用户设备(例如,设备104、106、200、400、600或800)上并通过一个或多个网络与服务器部分(例如,服

务器系统108)通信,例如,如图1中所示。

[0348] 下面讨论的图9中的每个框的相应功能可选地由硬件、软件或硬件和软件的组合来实现以执行本文所述的示例的原理。此外,应当注意,系统 900仅是用于处理音频输入和运动输入的系统的一个示例,并且系统900可以具有比所示更多或更少的部件,可以组合两个或更多个部件,或者可以具有不同的部件配置或布置。此外,尽管下面的讨论描述了在系统900的单个部件处执行的功能,但是应当理解,这些功能可以在系统900的其他部件处执行,并且这些功能可以在系统900的多于一个部件处执行。

[0349] 系统900包括耦接到处理单元916和模型单元922的输入单元902。输入单元包括各种输入设备,诸如一个或多个麦克风904、一个或多个加速度计906、以及可选地,一个或多个陀螺仪908、一个或多个相机910、一个或多个颜色传感器912、一个或多个光传感器914,和/或一个或多个接近传感器932。由输入单元902的输入设备分别收集的输入被提供给处理单元 916和模型单元922。

[0350] 处理单元916包括一个或多个处理器(例如,处理器220、616和704 中的任何一个)。例如,处理单元916包括主处理器918和低功率处理器 920。在一些示例中,主处理器918为执行例如操作系统和包括设备的数字助理等应用程序的应用处理器。在一些示例中,低功率处理器920在操作时比主处理器918消耗更少的功率,执行与主处理器918不同的计算机可执行指令,和/或在物理上比主处理器918小。

[0351] 在一些示例中,低功率处理器920对来自输入单元902的输入进行采样。在一些示例中,低功率处理器920在其从输入单元902采样的输入的类型上受到限制。例如,低功率处理器920仅对来自一个或多个加速度计 906和/或来自一个或多个陀螺仪908的运动输入进行采样。在一些示例中,低功率处理器920在其从输入单元902采样的输入的类型上不受限制。在一些示例中,根据上述示例,低功率处理器920基于来自输入单元 902的输入激活主处理器918。

[0352] 在一些示例中,主处理器918对来自输入单元902的输入进行采样。例如,主处理器918对来自一个或多个麦克风904的音频输入和/或来自一个或多个加速度计906和/或来自一个或多个陀螺仪908的运动输入进行采样。在一些示例中,根据上述示例,主处理器918根据被低功率处理器920 激活而对音频输入进行采样。

[0353] 模型单元922包括举高以讲话模型924、模型训练单元930和上下文单元934。举高以讲话模型924可选地包括运动模型926和音频模型928。本领域技术人员将理解,尽管模型单元922的功能与处理单元916分开描述,但在一些示例中,模型单元922的功能由处理单元916实现。例如,举高以讲话模型924和/或模型训练单元930的功能被实现为在主处理器918和/或低功率处理器920上执行的计算机可执行指令。

[0354] 上下文单元934确定电子设备的上下文。例如,上下文单元934根据上述技术确定电子设备的上下文是否是表示与外部电子设备的通信的预定类型的上下文。

[0355] 举高以讲话模型924接收来自输入单元902的输入。在一些示例中,基于由上下文单元934确定的上下文,举高以讲话模型924确定(1)是否发起虚拟助理会话或(2)是否将与音频输入相关联的通信传输到外部电子设备。

[0356] 如果电子设备的上下文不是预定类型的上下文,则举高以讲话924根据上面讨论的技术确定是发起虚拟助理会话还是将其去激活(例如,基于来自输入单元902的输入)。在



一些示例中,举高以讲话模型924确定和/或接收虚拟助理交互上下文(例如,从数字助理系统700),并且根据上面讨论的技术的任一种,基于虚拟助理交互上下文来确定是发起虚拟助理会话还是将其去激活。举高以讲话模型924可选地基于运动模型926和音频模型928进行的发起虚拟助理会话或将其去激活的相应确定来确定是发起虚拟助理会话还是将其去激活。在一些示例中,举高以讲话模型924发起虚拟助理会话和/或将其去激活。

[0357] 如果电子设备的上下文是预定类型的上下文,则举高以讲话模型924根据上面讨论的技术确定是否将与音频输入相关联的通信传输到外部电子设备(例如,基于来自输入单元902的输入)。在一些示例中,举高以讲话模型924使得与音频输入相关联的通信被传输到外部电子设备(例如,使用RF电路208)。在一些示例中,根据确定设备正在垂下(或已经垂下),举高以讲话模型924使得根据上面讨论的技术执行一个或多个设备功能。在一些示例中,举高以讲话模型924根据上述技术阻止通信的传输。

[0358] 在一些示例中,举高以讲话模型924包括神经网络(例如,递归神经网络(RNN)、卷积神经网络(CNN)等)。在一些示例中,举高以讲话模型924包括决策树(例如,梯度提升的决策树)和/或使用任何合适的机器学习技术来实现。

[0359] 在一些示例中,运动模型926接收运动输入并根据上述示例基于该运动输入确定是否执行设备功能(例如,发起虚拟助理会话,传输与音频输入相关联的通信)。例如,运动模型926使用运动输入确定电子设备的姿势(例如,已垂下、正在举高、已举高、正在垂下)和/或相应的姿势概率。在一些示例中,运动模型926还基于运动输入确定姿势之间的转变,这种转变的相应概率,姿势和姿势转变的相应持续时间,和/或与姿势之间的转变相关联的相应时间。

[0360] 在一些示例中,确定任何上述姿势特征包括分析加速度计和/或陀螺仪数据(例如,随时间推移)。例如,已垂下姿势(例如,图8A)与指示在一个或多个方向上很少或没有加速度的加速度计读数相关联。例如,正在举高姿势(例如,图8B)与指示在一个方向上的加速度和在另一个方向上的旋转加速度的加速度计读数相关联。例如,已举高姿势(例如,图8C)与在检测到与正在举高姿势一致的加速度计读数之后指示在所有方向上很少或没有加速度的加速度计读数相关联。例如,正在垂下姿势(例如,图8F)与在检测到与已举高姿势一致的加速度计读数之后指示一个方向上的加速度和另一个方向上的旋转加速度的加速度计读数相关联。

[0361] 在一些示例中,运动模型926包括神经网络(例如,递归神经网络(RNN)、卷积神经网络(CNN)等)。在一些示例中,运动模型包括决策树(例如,梯度提升的决策树)和/或使用任何合适的机器学习技术来实现。

[0362] 在一些示例中,运动模型926特定于电子设备的用户(例如,授权用户)。例如,可选地训练和/或优化运动模型926以基于来自特定用户的运动输入准确地确定是否执行设备功能(例如,发起虚拟助理会话,传输与音频输入相关联的通信)。例如,如果用户具有长/短手臂和/或具有举高设备以发起虚拟助理的特定方式,则运动模型926在确定姿势、姿势转变、姿势持续时间等时会考虑这些特定于用户的特性。因此,本公开设想用户注册以训练和/或优化运动模型926。例如,在注册期间,用户提供他或她打算引起功能的执行的设备动作,并且基于所提供的运动输入训练运动模型926。

[0363] 在一些示例中,音频模型928接收音频输入(例如,从一个或多个麦克风904),并根



据上述示例基于该音频输入确定是否执行设备功能(例如,发起虚拟助理会话,传输与音频输入相关联的通信)。例如,音频模型928被配置为尤其确定音频输入是否包括由授权用户说出的人类语音,并根据上面讨论的示例确定音频输入与设备的接近度。

[0364] 在一些示例中,音频模型928包括神经网络(例如,递归神经网络(RNN)、卷积神经网络(CNN)等)。在一些示例中,音频模型928包括决策树(例如,梯度提升的决策树)和/或使用任何合适的机器学习技术来实现。

[0365] 在一些示例中,音频模型928特定于电子设备的用户(例如,授权用户)。例如,可选地训练和/或优化音频模型928以准确地确定音频输入是否应基于用户特定的音频特性引起设备功能的执行。例如,如果用户具有在他或她打算发起虚拟助理会话时与设备说话的特定方式(例如,语调、音调、重音、说话速率等),则音频模型928在确定是否发起虚拟助理会话时会考虑这些特定于用户的特性。因此,本公开设想用户注册以训练和/或优化音频模型928。例如,在注册期间,用户提供音频输入,并且基于该音频输入训练和/或优化音频模型928。

[0366] 在一些示例中,基于音频输入和运动输入的注册是并行、同时或以换句话说重叠的方式执行的。例如,用户同时提供音频输入和运动输入,用户想要将这两种输入组合来引起功能的执行。在一些示例中,使用运动输入和音频输入训练举高以讲话模型924(例如,分别使用运动输入和音频输入训练运动模型926和音频模型928)。

[0367] 在一些示例中,训练举高以讲话模型924(例如,基于用于执行设备功能的用户特定偏好)包括调整上面讨论的一个或多个持续时间和/或一个或多个阈值。例如,基于虚拟助理交互上下文、设备上下文、运动、音频和/或其他输入,调整上面讨论的一个或多个持续时间和/或一个或多个阈值。这样,适应举高以讲话模型924适于(例如,随时间推移和/或在注册期间)基于用户特定偏好来引起设备功能的执行。

[0368] 模型训练单元930被配置为根据上述示例训练举高以讲话模型924。例如,模型训练单元接收来自输入单元902的输入,并且使用所接收的输入作为训练数据来训练举高以讲话模型924。例如,模型训练单元930被配置为例如基于相应的作为训练数据的运动输入和音频输入来训练运动模型926和/或音频模型928。

[0369] 例如,如果举高以讲话模型924包括神经网络,则训练举高以讲话模型924包括确定用于神经网络的节点之间的连接的一组最佳加权值。例如,如果举高以讲话模型924包括梯度提升树模型,则训练举高以讲话模型924包括最小化损失函数。以这种方式训练举高以讲话模型924增加了音频输入、运动输入、其他输入、设备上下文和/或虚拟助理交互上下文引起设备执行正确功能的几率并减少了执行不正确的功能的几率。例如,这样,旨在引起通信的传输的运动输入和音频输入不会意外地发起虚拟助理会话。又如,这样,旨在发起虚拟助理会话的运动输入和音频输入正确地发起虚拟助理会话,并且不适当的运动输入和音频输入不会意外地发起虚拟助理会话。

[0370] 图10A至图10D示出了根据各种示例的用于操作数字助理的过程1000。例如,过程1000使用实现数字助理的一个或多个电子设备(例如,设备104、106、200、400、600或800)执行。在一些示例中,使用客户端-服务器系统(例如,系统100)来执行过程1000,并且在服务器(例如,DA服务器106)和客户端设备之间以任何方式来划分过程1000的框。在其他示例中,在服务器和多个客户端设备(例如,移动电话和智能手表)之间划分过程1000的框。因

此,虽然过程1000的部分在本文中被描述为由客户端-服务器系统的特定设备执行,但应当理解,过程1000不限于此。在其他示例中,利用仅客户端设备(例如,用户设备104)或仅多个客户端设备来执行过程1000。在过程1000中,一些框被任选地组合,一些框的次序被任选地改变,并且一些框被任选地省略。在一些示例中,可结合过程 1000来执行附加步骤。

[0371] 如下所述,过程1000包括检测表示电子设备的运动的输入(例如,将电子设备朝向用户的嘴部举高)以及利用电子设备的麦克风对音频输入(例如,口头用户请求,比如“帕罗奥图的天气怎么样?”)进行采样。过程1000还包括基于音频输入和表示电子设备的运动的输入来确定是否发起虚拟助理会话。根据确定发起虚拟助理会话,过程1000包括发起虚拟助理会话(例如,显示与虚拟助理会话相关联的用户界面)。根据确定不发起虚拟助理会话,过程1000包括放弃发起虚拟助理会话。

[0372] 基于音频输入和表示电子设备的运动的输入(例如,“运动输入”)来确定是否发起虚拟助理会话提供了自然且有效的发起虚拟助理并与其进行交互的方式。例如,为了发起虚拟助理并基于口头请求让虚拟助理执行任务,用户仅仅需要将设备举高至他或她的嘴部附近并且在大致相同的时间处发出请求(例如,不需要包括口头触发(诸如“嘿,Siri”)的请求)。以这种方式促进虚拟助理交互能够实现有效的虚拟助理交互(例如,在每次交互之前不需要口头触发)以及对口头请求的快速虚拟助理响应。此外,使用音频输入和运动输入这二者来发起虚拟助理会话可防止基于音频输入和运动输入中任一者的虚拟助理意外激活。这降低了意外发起虚拟助理的几率并且节省了原本在非期望的情况下操作虚拟助理而消耗的电池电力。这样,使得用户-设备界面更有效(例如,通过减少用户提供的音频输入的量,通过减少虚拟助理响应时间,通过防止虚拟助理意外激活),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0373] 在框1002处,在具有麦克风和显示器的电子设备(例如,设备800)处,检测到表示该电子设备的运动的输入。在一些示例中,该电子设备是计算机、扬声器、智能手表、电话或它们的组合。

[0374] 在一些示例中,表示电子设备的运动的输入由电子设备的一个或多个加速度计(例如,906)检测。在一些示例中,表示电子设备的运动的输入由电子设备的一个或多个陀螺仪(例如,908)检测。在一些示例中,检测表示电子设备的运动的输入包括使用第一处理器(例如,920)对表示电子设备的运动的输入进行采样,如可选框1004中所示。

[0375] 在一些示例中,电子设备的运动包括电子设备在多个姿势中的两个或更多个姿势之间的转变。在一些示例中,这多个姿势包括已垂下姿势(例如,图8A中所示)、正在举高姿势(例如,图8B中所示)、已举高姿势(例如,图8C至图8E中所示)和已垂下姿势(例如,如图8F所示)。

[0376] 在可选框1006处,基于表示电子设备的运动的输入的第一部分来开启电子设备的显示器(例如,804)。在一些示例中,表示电子设备的运动的输入的第一部分表示手腕举高手势和/或手腕旋转手势(例如,如图8B所示)。

[0377] 在框1008处,利用电子设备的麦克风(例如,904)对音频输入(例如,806)进行采样。在一些示例中,根据开启显示器来执行对音频输入的采样。在一些示例中,在检测表示电子设备的运动的输入的至少第二部分时对音频输入进行采样(例如,在检测图8C中的运动输入时对音频输入 806进行采样)。在检测运动输入时对音频输入进行采样允许用户提

供(以及设备采样和/或检测)同时(或近似同时)的音频输入和运动输入来发起虚拟助理会话。如所讨论的,以这种方式提供虚拟助理发起提供了发起虚拟助理会话的自然方式和有效方式,并减少了意外发起虚拟助理的几率。因此,使得用户-设备接口更有效(例如,通过启用快速和简单的虚拟助理发起),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0378] 在一些示例中,对音频输入进行采样包括使用第一处理器对音频输入进行采样,如可选框1010中所示。在一些示例中,对音频输入进行采样包括使用第二处理器(例如,918)对音频输入进行采样,如可选框1012中所示。在一些示例中,第二处理器在操作时比第一处理器消耗更多的功率。在一些示例中,基于表示电子设备的运动的输入的至少第三部分来激活第二处理器。

[0379] 在一些示例中,音频输入(例如,806)不包括口头触发(例如,用于发起虚拟助理的语音触发诸如“嘿,Siri”)。对不包括口头触发的音频输入进行采样并使用这样的音频输入来发起虚拟助理会话使得能够进行自然且有效的虚拟助理发起/交互。例如,用户可以简单地向虚拟助理发出请求(不包括“嘿,Siri”)以发起虚拟助理并使其响应该请求。这样,使得用户-设备接口更有效(例如,通过减少用户提供的音频输入的量),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0380] 在框1014处,基于音频输入和表示电子设备的运动的输入来确定(例如,通过举高以讲话模型924)是否发起虚拟助理会话。如所讨论的,下面讨论的用于确定是否发起虚拟助理会话的示例性技术可以允许适当的运动输入和/或音频输入来发起虚拟助理会话并防止不适当的运动输入和/或音频输入发起虚拟助理会话。这样,改进了设备可操作性并且使得人-机接口更有效(例如,通过在请求时准确且有效地发起虚拟助理会话,通过防止用户被意外发起的虚拟助理会话打扰,通过减少或消除意外发起的虚拟助理消耗的电池电量,以及通过减少或消除提供来将意外发起的虚拟助理会话去激活的用户输入)。

[0381] 在一些示例中,确定是否发起虚拟助理会话包括基于运动模型(例如,926)确定是否发起虚拟助理会话,如可选框1016中所示。在一些示例中,运动模型包括第一神经网络。在一些示例中,运动模型特定于电子设备的第一用户。在一些示例中,使用运动模型确定电子设备在多个姿势中的两个或更多姿势之间的转变的概率,如可选框1018中所示。在一些示例中,使用运动模型确定多个姿势中的姿势的概率,如可选框1020中所示。

[0382] 在一些示例中,确定是否发起虚拟助理会话包括基于表示电子设备的运动的输入来(例如,通过模型926)确定电子设备的运动是否包括从正在举高姿势到已举高姿势的转变(例如,如图8B和图8C所示),如可选框 1022中所示。

[0383] 在一些示例中,确定是否发起虚拟助理会话包括确定(例如,通过模型926)从正在举高姿势到已举高姿势的转变的持续时间,如可选框1024 中所示。

[0384] 在一些示例中,在第一时间对音频输入进行采样,从正在举高姿势到已举高姿势的转变与第二时间相关联,并且确定(例如,通过模型924)是否发起虚拟助理会话基于第一时间和第二时间,如可选框1026中所示。这样,可以通过检测适合于发起虚拟助理的大致同步的运动输入和音频输入来发起虚拟助理会话。此外,可以防止单独基于运动输入或单独基于音频输入的意外虚拟助理发起(例如,如果在彼此的某个适当持续时间内没有接收到运动输入和音频输入,则这些输入不态可能用于发起虚拟助理)。

[0385] 在一些示例中,确定是否发起虚拟助理会话包括基于表示电子设备的运动的输入来确定(例如,通过模型926)已举高姿势的持续时间(例如,图8C中所示)是否超过第一预定持续时间,如可选框1028中所示。

[0386] 在一些示例中,确定是否发起虚拟助理会话包括基于表示电子设备的运动的输入来(例如,通过模型926)确定电子设备的运动是否包括从已举高姿势到正在垂下姿势的转变(例如,图8F中所示),如可选框1030中所示。

[0387] 在一些示例中,确定是否发起虚拟助理会话包括基于音频模型(例如,928)确定是否发起虚拟助理会话,如可选框1032中所示。在一些示例中,音频模型包括第二神经网络。在一些示例中,音频模型特定于电子设备的第二用户。

[0388] 在一些示例中,确定是否发起虚拟助理会话包括确定(例如,通过模型928)音频输入是否包括人类语音,如可选框1034中所示。在一些示例中,确定音频输入是否包括人类语音包括确定音频输入是否包括由电子设备的授权用户说出的人类语音,如可选框1036中所示。这样,可以防止由于采样噪声(和/或来自未授权用户的语音)而导致的意外虚拟助理发起。此外,可以提供基于接收来自人(和/或来自授权用户)的口头请求的准确虚拟助理发起。

[0389] 在一些示例中,确定是否发起虚拟助理会话包括确定(例如,通过模型928)与音频输入相关联的方向,如可选框1038中所示。这样,可以防止由于对未指向被配置为操作虚拟助理的设备的音频进行采样而导致的意外虚拟助理发起。此外,可以提供基于接收指向设备的音频输入的准确虚拟助理发起。

[0390] 在一些示例中,确定是否发起虚拟助理会话包括确定(例如,通过模型928)音频输入与电子设备的接近度,如可选框1040中所示。这样,可以防止由于对源自离被配置为操作虚拟助理的设备太远的源的音频进行采样而导致的意外虚拟助理发起(例如,因为在一些示例中,用户在设备靠近用户的嘴部时对设备说话)。此外,可以提供基于接收源自靠近设备的源的音频输入的准确虚拟助理发起。

[0391] 在一些示例中,确定是否发起虚拟助理会话包括在第三时间基于音频输入确定(例如,通过模型928)是否发起虚拟助理会话,如可选框1042 中所示。在一些示例中,确定是否发起虚拟助理会话包括在第四时间基于表示电子设备的运动的输入来确定(例如,通过模型926)是否发起虚拟助理会话,如可选框1044中所示。在一些示例中,确定是否发起虚拟助理会话包括确定(例如,通过模型924)第三时间和第四时间是否在第二预定持续时间内,如可选框1046中所示。在一些示例中,根据确定第三时间和第四时间在第二预定持续时间内来执行发起虚拟助理会话。因此,在一些示例中,发起虚拟助理会话是基于检测适合于发起虚拟助理会话的同步或大致同步的运动输入和音频输入。因此,即使运动输入和/或音频输入可各自适合用于发起虚拟助理,也可防止基于单独的运动输入和/或单独的音频输入而意外发起虚拟助理。因此,以这种方式提供虚拟助理激活提供了有效且直观的发起虚拟助理会话的方式,同时还减少了意外发起的几率。

[0392] 在框1048处,根据确定发起虚拟助理会话,发起虚拟助理会话(例如,通过模型924)。在一些示例中,发起虚拟助理会话包括显示与虚拟助理相关联的第一用户界面(例如,如图8C中最右侧的显示器804所示),如可选框1050中所示。在一些示例中,发起虚拟助理会话包括提供音频输出,如可选框1052中所示。在一些示例中,发起虚拟助理会话包括提

供触觉输出,如可选框1054中所示。

[0393] 在可选框1056处,确定(例如,通过模型924)与从已举高姿势到正在垂下姿势的第三转变相关联的时间。在可选框1058处,基于与从已举高姿势到正在垂下姿势的第三转变相关联的时间确定(例如,通过模型924)音频输入的终点。以这种方式确定音频输入的终点提供了有效且准确的终止音频输入的方式。例如,用户可以简单地放下他或她的手腕(例如,如图8F所示)来指示他或她已完成说话。

[0394] 在框1060处,根据确定不发起虚拟助理会话,放弃发起虚拟助理会话。

[0395] 在可选框1062处,根据发起虚拟助理会话,基于音频输入(例如,通过图7B中所示的数字助理)确定用户意图。在可选框1064处,根据发起虚拟助理会话,基于用户意图来执行任务。在可选框1066处,根据发起虚拟助理会话,提供与任务相关联的输出(例如,输出808)。

[0396] 在可选框1068处,根据发起虚拟助理会话并且根据不基于音频输入确定用户意图,放弃响应于音频输入提供输出。

[0397] 在可选框1070处,根据发起虚拟助理会话,虚拟助理会话被去激活(例如,通过模型924)。在一些示例中,将虚拟助理会话去激活包括停止显示与数字助理相关联的第二用户界面(例如,如图8E和图8F所示),如可选框1072中所示。在一些示例中,将虚拟助理会话去激活包括放弃响应音频输入,如可选框1074中所示。

[0398] 将虚拟助理会话去激活允许取消意外发起的虚拟助理会话。因此,减少了由于在不期望时操作虚拟助理而消耗的电池电量和处理能力。另外,在不期望时将虚拟助理会话去激活改善了设备可用性和可操作性(例如,用户可以继续将该设备用于其他目的而不受已发起的虚拟助理的打扰)。这样,使得用户-设备接口更有效(例如,通过在不期望时暂时将虚拟助理会话去激活),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0399] 另外,下面描述的用于将虚拟助理会话去激活的技术可以提供有关是否将虚拟助理会话去激活的准确和有效的确定,以及虚拟助理会话的准确和有效的去激活。以这种方式改进虚拟助理去激活改善了设备可用性和效率(例如,通过在不期望时将虚拟助理会话去激活并且通过防止期望的已发起的虚拟助理会话被去激活)。因此,使得用户-设备接口更有效(例如,通过准确地确定是否将虚拟助理会话去激活),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0400] 在可选框1076处,根据发起虚拟助理会话,基于表示电子设备的运动的输入,确定(例如,通过模型924)电子设备的运动是否包括从已举高姿势到正在垂下姿势的第二转变(例如,如图8F所示)。

[0401] 在可选框1078处,确定(例如,通过模型924)与从已举高姿势到正在垂下姿势的第二转变相关联的时间。

[0402] 在可选框1080处,确定(例如,通过模型924)与音频输入中的语音结束相关联的时间。

[0403] 在可选框1082处,根据确定电子设备的运动包括从已举高姿势到正在垂下姿势的第二转变,虚拟助理会话被去激活(例如,通过模型924)。在一些示例中,其中根据确定(例如,通过模型924)与从已举高姿势到正在垂下姿势的第二转变相关联的时间在与音频输入

中的语音结束相关联的时间之前,进一步执行将虚拟助理会话去激活,如可选框1084中所示。

[0404] 在可选框1086处,根据发起虚拟助理会话,基于音频输入(例如,图 8E中所示的音频输入)的语言内容确定(例如,通过模型924)是否将虚拟助理会话去激活。在可选框1088处,根据确定将虚拟助理会话去激活,虚拟助理会话被去激活(例如,通过模型924)。在可选框1090处,根据确定不将虚拟助理会话去激活(例如,通过模型924),放弃将虚拟助理会话去激活。

[0405] 在可选框1092处,根据发起虚拟助理会话,确定音频输入的持续时间是否小于阈值持续时间(例如,通过模型924)。在可选框1094处,根据确定音频输入的持续时间小于阈值持续时间,虚拟助理会话被去激活(例如,通过模型924)。

[0406] 以上参考图10A至图10D描述的操作任选地由图1至图4、图6A至图6B、图7A至图7C、图8A至图8F以及图9中所描绘的部件来实现。例如,过程1000的操作可由处理单元916和/或由模型单元922实现。本领域的普通技术人员会清楚地知道如何基于在图1至图4、图6A至图6B、图 7A至图7C、图8A至图8F以及图9中所描绘的部件来实现其他过程。

[0407] 图11示出了根据各种示例的用于传输通信的过程1100。例如,过程 1100使用一个或多个电子设备(例如,设备104、106、200、400、600或 800)执行。在一些示例中,使用客户端-服务器系统(例如,系统100)来执行过程1100,并且在服务器(例如,DA服务器106)和客户端设备之间以任何方式来划分过程1100的框。在其他示例中,在服务器和多个客户端设备(例如,移动电话和智能手表)之间划分过程1100的框。因此,虽然过程1100的部分在本文中被描述为由客户端-服务器系统的特定设备执行,但应当理解,过程1100不限于此。在其他示例中,利用仅客户端设备(例如,用户设备104)或仅多个客户端设备来执行过程1100。在过程1100中,一些框被任选地组合,一些框的次序被任选地改变,并且一些框被任选地省略。在一些示例中,可结合过程1100来执行附加步骤。

[0408] 如下所述,过程1100包括检测表示电子设备的运动的输入(例如,将电子设备朝向用户的嘴部举高)以及利用电子设备的麦克风对音频输入(例如,“嗨Shawn,你好吗?”)进行采样。过程1100还包括确定电子设备的上下文(例如,设备当前正在接听电话呼叫(例如,来自名为“Shawn”的联系人))。根据确定该上下文是表示与外部电子设备的通信的预定类型的上下文(例如,接听电话呼叫的上下文),过程1110包括基于音频输入和表示电子设备的运动的输入来确定是否将与该音频输入相关联的通信传输到外部电子设备。例如,是否向Shawn的设备传输音频“嗨 Shawn,你好吗?”。根据确定将与该音频输入相关联的通信传输到外部电子设备,过程1110包括将与该音频输入相关联的通信传输到外部电子设备。例如,设备接听来自Shawn的电话,并将“嗨Shawn,你好吗?”传输给Shawn的设备。根据确定不将与该音频输入相关联的通信传输到外部电子设备,过程1100包括放弃将与该音频输入相关联的通信传输到外部电子设备。

[0409] 确定电子设备的上下文允许电子设备基于音频输入和动作输入准确地确定执行哪个设备功能。具体地讲,在动作输入和音频输入的不同或相似组合可以引起多个设备功能的执行(例如,接听电话呼叫,发起虚拟助理会话)的示例中,重要的是区分(例如,基于确定的设备上下文)应基于音频输入和运动输入调用设备的哪个功能。否则,例如,旨在接听电话呼叫的音频输入和运动输入可能意外发起虚拟助理会话。因此,使用上下文信息来确

定要执行哪个设备功能使得能够执行正确的设备功能。这样,使得用户-设备接口更有效(例如,通过引起设备功能的正确执行,通过防止设备功能的不正确执行,通过减少或消除用以取消错误执行的设备功能的用户输入),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0410] 基于音频输入和表示电子设备的运动的输入(例如,“运动输入”)确定是否将与音频输入相关联的通信传输到外部电子设备提供了与外部电子设备通信的自然且有效的方式。例如,为了接听电话呼叫,用户只需将设备举高到他或她的嘴部附近并在大致相同的时间提供对电话呼叫的响应(例如,不提供触觉输入来接听该电话呼叫)。以这种方式促进与外部电子设备的通信使得能够与外部设备快速通信(例如,用户只需举高他或她的设备并说话来使通信被传输)。另外,使用音频输入和运动输入二者来传输通信可以防止仅基于音频输入或仅基于运动输入的意外通信传输。这降低了在不需要时传输通信的几率。这样,使得用户-设备接口更有效(例如,通过快速响应来自外部设备的通信,通过防止意外的通信传输),此外,通过使得用户能够更快速且有效地使用设备,减少了电力使用并且延长了设备的电池寿命。

[0411] 在框1102处,(例如,通过输入单元902)检测表示电子设备的运动的输入。

[0412] 在框1104处,利用电子设备的麦克风(例如,通过麦克风904)对音频输入进行采样。

[0413] 在框1106处,(例如,通过上下文单元934)确定电子设备的上下文。

[0414] 在框1108处,(例如,通过上下文单元934)确定电子设备的上下文是否是表示与外部电子设备的通信的预定类型的上下文。

[0415] 在框1110处,根据确定上下文是表示与外部电子设备的通信的预定类型的上下文(框1108,是),基于音频输入和表示电子设备的运动的输入确定(例如,通过模型924)是否将与音频输入相关联的通信传输到外部电子设备。

[0416] 在框1112处,根据确定将与音频输入相关联的通信传输到外部电子设备,与音频输入相关联的通信被传输到外部电子设备(例如,通过RF电路208)。

[0417] 在框1114处,根据确定不将与音频输入相关联的通信传输到外部电子设备,放弃将与音频输入相关联的通信传输到外部电子设备。

[0418] 在框1116处,根据确定上下文不是表示与外部电子设备的通信的预定类型的上下文(框1108,否),基于音频输入和表示电子设备的运动的输入确定是否发起虚拟助理会话(例如,基于以上关于图10A至图10D讨论的技术)。

[0419] 以上参考图11描述的操作任选地由图1至图4、图6A至图6B、图7A至图7C、图8A至图8F以及图9中所描绘的部件来实现。例如,过程1100的操作可由处理单元916和/或由模型单元922实现。本领域的普通技术人员会清楚地知道如何基于在图1至图4、图6A至图6B、图7A至图7C、图8A至图8F以及图9中所描绘的部件来实现其他过程。

[0420] 根据一些具体实施,提供一种计算机可读存储介质(例如,非暂态计算机可读存储介质),该计算机可读存储介质存储供电子设备的一个或多个处理器执行的一个或多个程序,该一个或多个程序包括用于执行本文所述方法或过程中的任一个的指令。

[0421] 根据一些具体实施,提供了一种电子设备(例如,便携式电子设备),该电子设备包括用于执行本文所述的方法和过程中的任一个的装置。



[0422] 根据一些具体实施,提供了一种电子设备(例如,便携式电子设备),该电子设备包括处理单元,该处理单元被配置为执行本文所述的方法和过程中的任一个。

[0423] 根据一些具体实施,提供了一种电子设备(例如,便携式电子设备),该电子设备包括一个或多个处理器和存储用以由一个或多个处理器执行的一个或多个程序的存储器,该一个或多个程序包括用于执行本文所描述的方法和过程中的任一个的指令。

[0424] 为了解释的目的,前面的描述是通过参考具体实施方案来进行描述的。然而,上面的例示性论述并非旨在是穷尽的或将本发明限制为所公开的精确形式。根据以上教导内容,很多修改形式和变型形式都是可能的。选择并描述这些实施方案是为了最好地解释这些技术的原理及其实际应用。本领域的其他技术人员由此能够最好地利用这些技术以及具有适合于所预期的特定用途的各种修改的各种实施方案。

[0425] 虽然参照附图对本公开以及示例进行了全面的描述,但应当注意,各种变化和修改对于本领域内的技术人员而言将变得显而易见。应当理解,此类变化和修改被认为被包括在由权利要求书所限定的本公开和示例的范围内。

[0426] 如上所述,本技术的一个方面是收集和使用可从各种来源获得的数据以改进用于发起虚拟助理/将虚拟助理去激活的模型。本公开预期,在一些实例中,这些所采集的数据可包括唯一地识别或可用于联系或定位特定人员的个人信息数据。此类个人信息数据可包括人口统计数据、基于位置的数据、电话号码、电子邮件地址、Twitter账号、家庭地址、与用户健康或健康水平相关的数据或记录(例如,生命体征测量、药物信息、锻炼信息)、出生日期或任何其他识别或个人信息。

[0427] 本公开认识到在本发明技术中使用此类个人信息数据可用于使用户受益。例如,个人信息数据可用于生成和/或训练用于发起虚拟助理的个性化模型。因此,使用这种个性化信息数据使得能够改进用于发起虚拟助理的模型,该模型根据特定用户发起虚拟助理的方式来定制。此外,本公开还预期个人信息数据有益于用户的其他用途。例如,健康和健身数据可用于向用户的总体健康状况提供见解,或者可用作使用技术来追求健康目标的个人的积极反馈。

[0428] 本公开设想了负责这种个人信息数据的收集、分析、公开、发送、存储或其他用途的实体将遵守已确立的隐私政策和/或隐私实践。具体地,此类实体应当实行并坚持使用被公认为满足或超出对维护个人信息数据的隐私性和安全性的行业或政府要求的隐私政策和实践。用户可以方便地访问此类策略,并应随着数据的采集和/或使用变化而更新。用户的个人信息应采集为实体的合法和合理使用,而不应在这些合法使用之外共享或出售。此外,在收到用户知情同意后,应进行此类采集/共享。此外,此类实体应考虑采取任何必要步骤,保卫和保障对此类个人信息数据的访问,并确保其他有权访问个人信息数据的人遵守其隐私政策和流程。另外,这种实体可使其本身经受第三方评估以证明其遵守广泛接受的隐私政策和实践。此外,应调整政策和实践,以便采集和/或访问的特定类型的个人信息数据,并适用于包括管辖范围的具体考虑的适用法律和标准。例如,在美国,某些健康数据的收集或访问可能受联邦和/或州法律诸如健康保险便利和责任法案(HIPAA)的约束;而其他国家的健康数据可能受其他法规 and 政策的约束,因此应予以相应处理。因此,在每个国家应保持不同的个人数据类型不同的隐私实践。

[0429] 不管前述情况如何,本公开还设想用户选择性地阻止使用或访问个人信息数据的



实施方案。即本公开预期可提供硬件元件和/或软件元件,以防止或阻止对此类个人信息数据的访问。例如,在收集用于发起虚拟助理的特定于用户的运动输入和音频输入(例如,在注册过程期间)的情况下,本技术可以被配置为允许用户在注册服务期间或之后的任何时间选择“选择加入”或“选择退出”参与个人信息数据的收集。又如,用户可以选择限制特定于用户的运动输入和/或音频输入被保持的时间长度,或者完全禁止收集特定于用户的运动输入和/或音频输入。除了提供“选择加入”和“选择退出”选项外,本公开设想提供与访问或使用个人信息相关的通知。例如,用户可在下载应用程序时通知用户,其个人信息数据将被访问,然后在个人信息数据被应用程序访问之前再次提醒用户。

[0430] 此外,本公开的目的是应管理和处理个人信息数据以最小化无意或未经授权访问或使用的风险。一旦不再需要数据,通过限制数据采集和删除数据可最小化风险。此外,并且当适用时,包括在某些健康相关应用中,数据去标识可用于保护用户的隐私。在适当的情况下,可以通过移除特定标识符(例如,出生日期等)、控制存储的数据的量或特征(例如,在城市级而非地址级收集位置数据)、控制数据的存储方式(例如,在用户之间聚合数据)和/或其他方法来促进去标识。

[0431] 因此,虽然本公开广泛地覆盖了使用个人信息数据来实现一个或多个各种所公开的实施方案,但本公开还设想各种实施方案也可在无需访问此类个人信息数据的情况下被实现。即,本发明技术的各种实施方案不会由于缺少此类个人信息数据的全部或一部分而无法进行。例如,可以使用用于基于非个人信息数据或最少量的个人信息发起虚拟助理的通用(例如,不特定于用户和/或特定用户组)模型来发起虚拟助理,诸如设备所请求的与用户相关联的内容,可供用于发起虚拟助理的模型使用的其他非个人信息,或公开提供的信息。

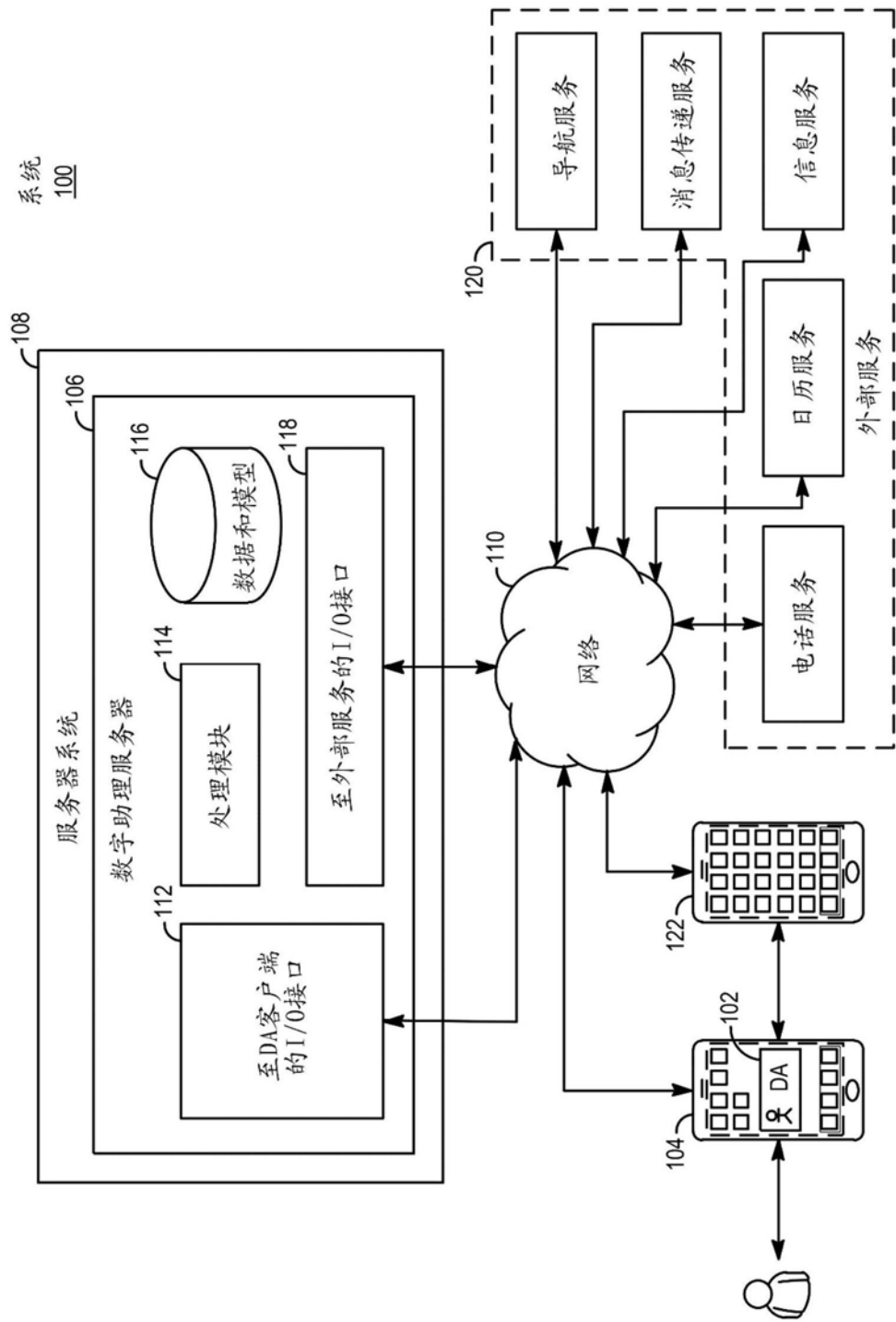


图1

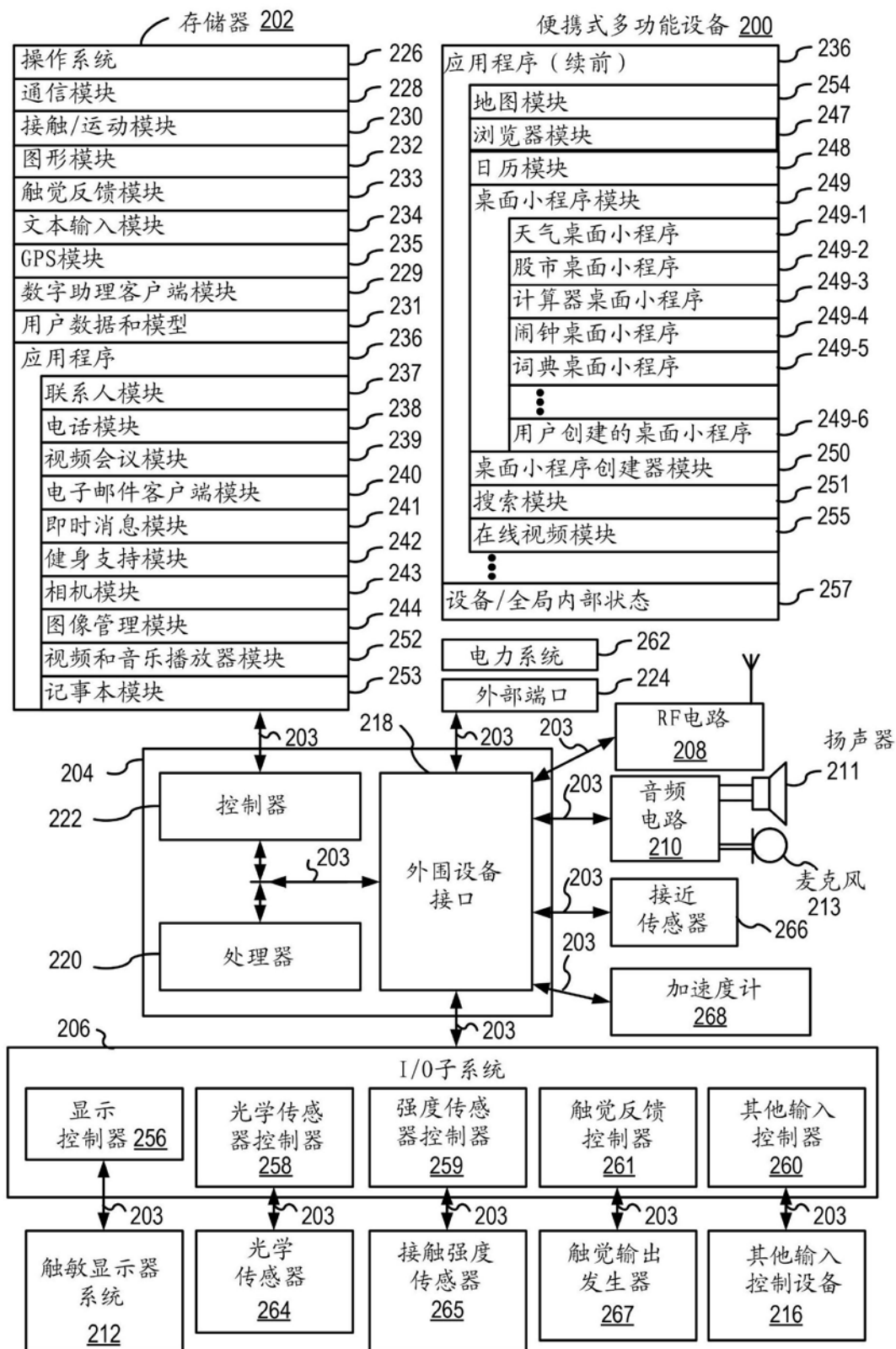


图2A

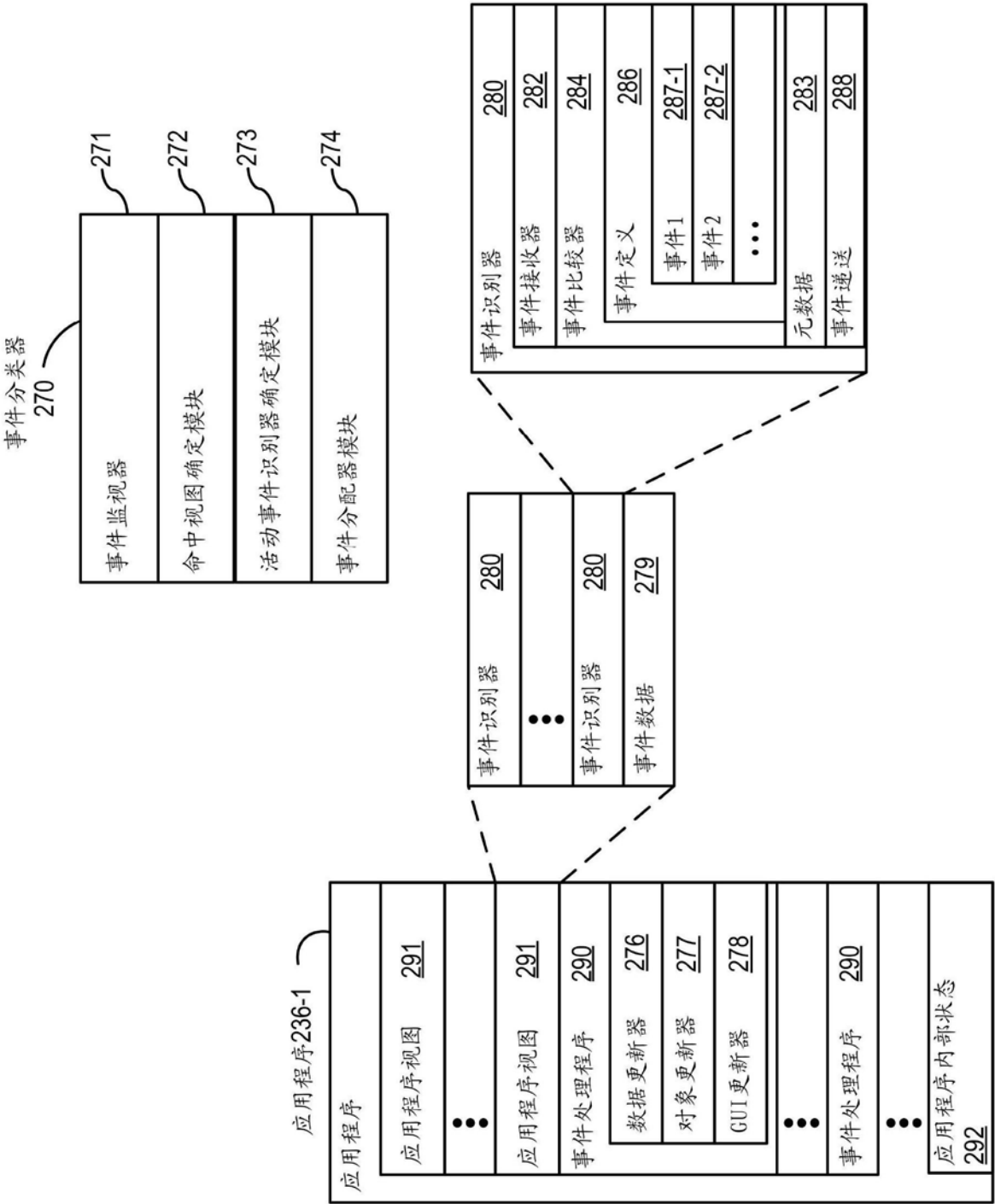


图2B

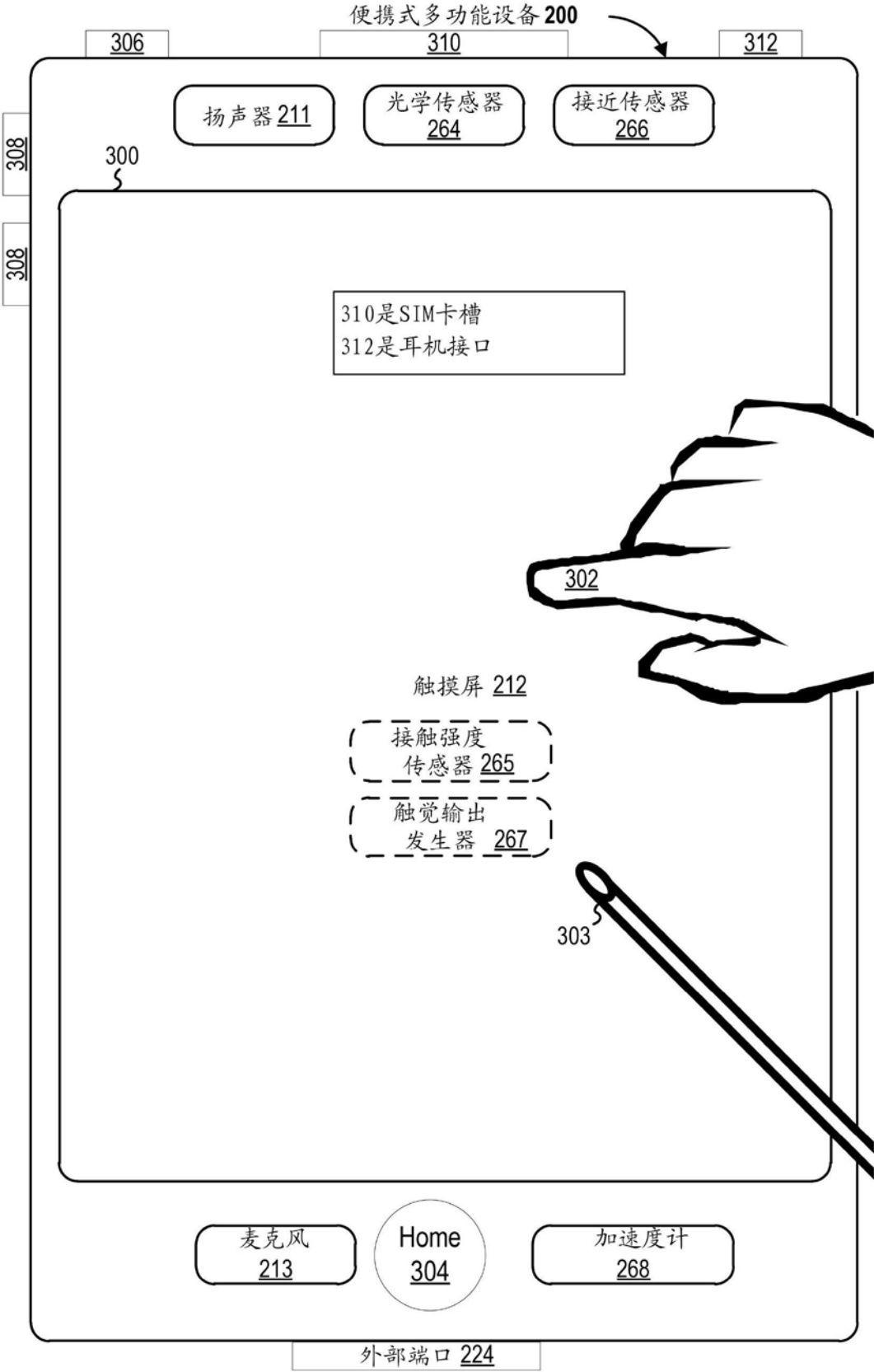


图3

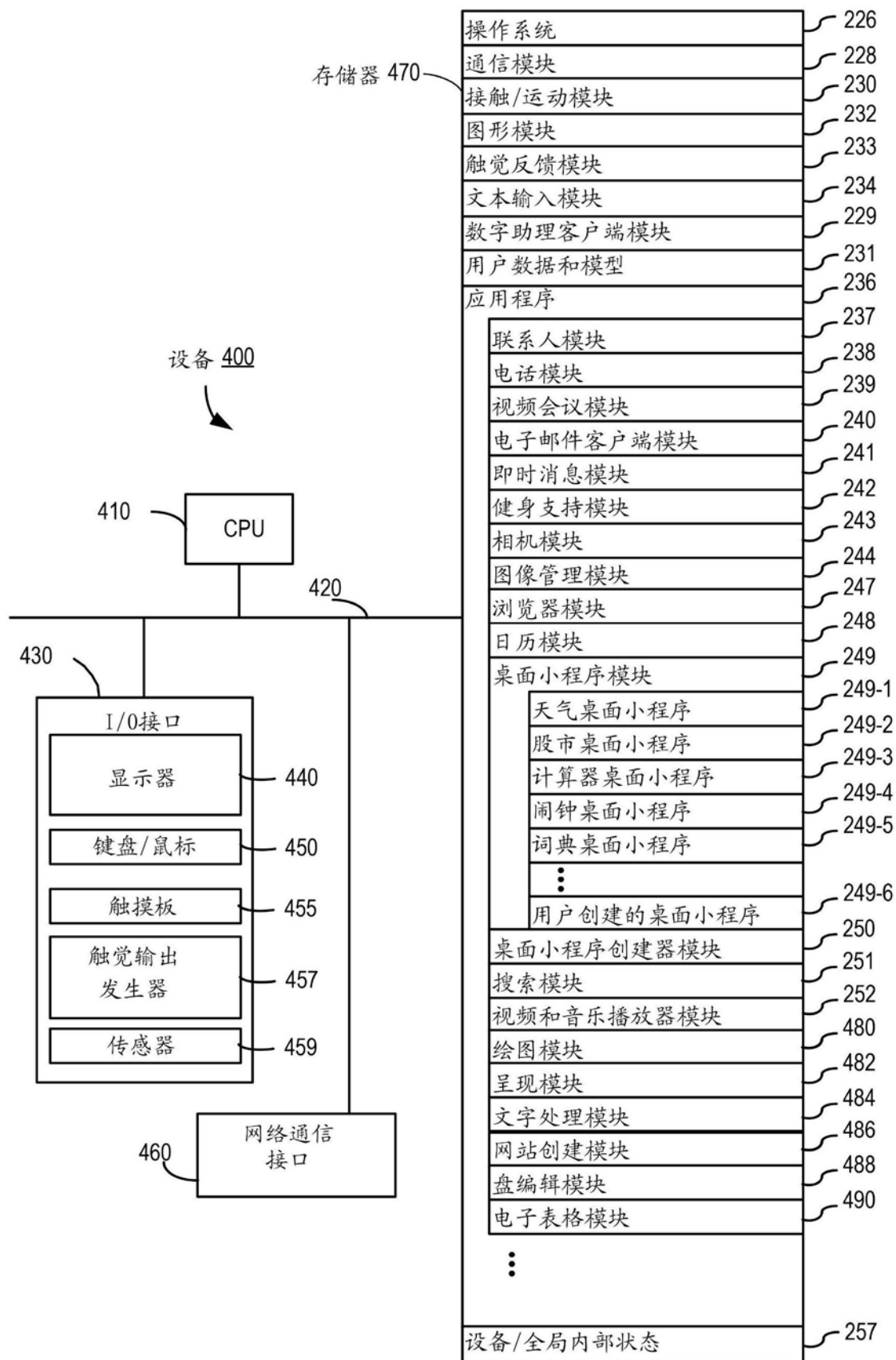


图4

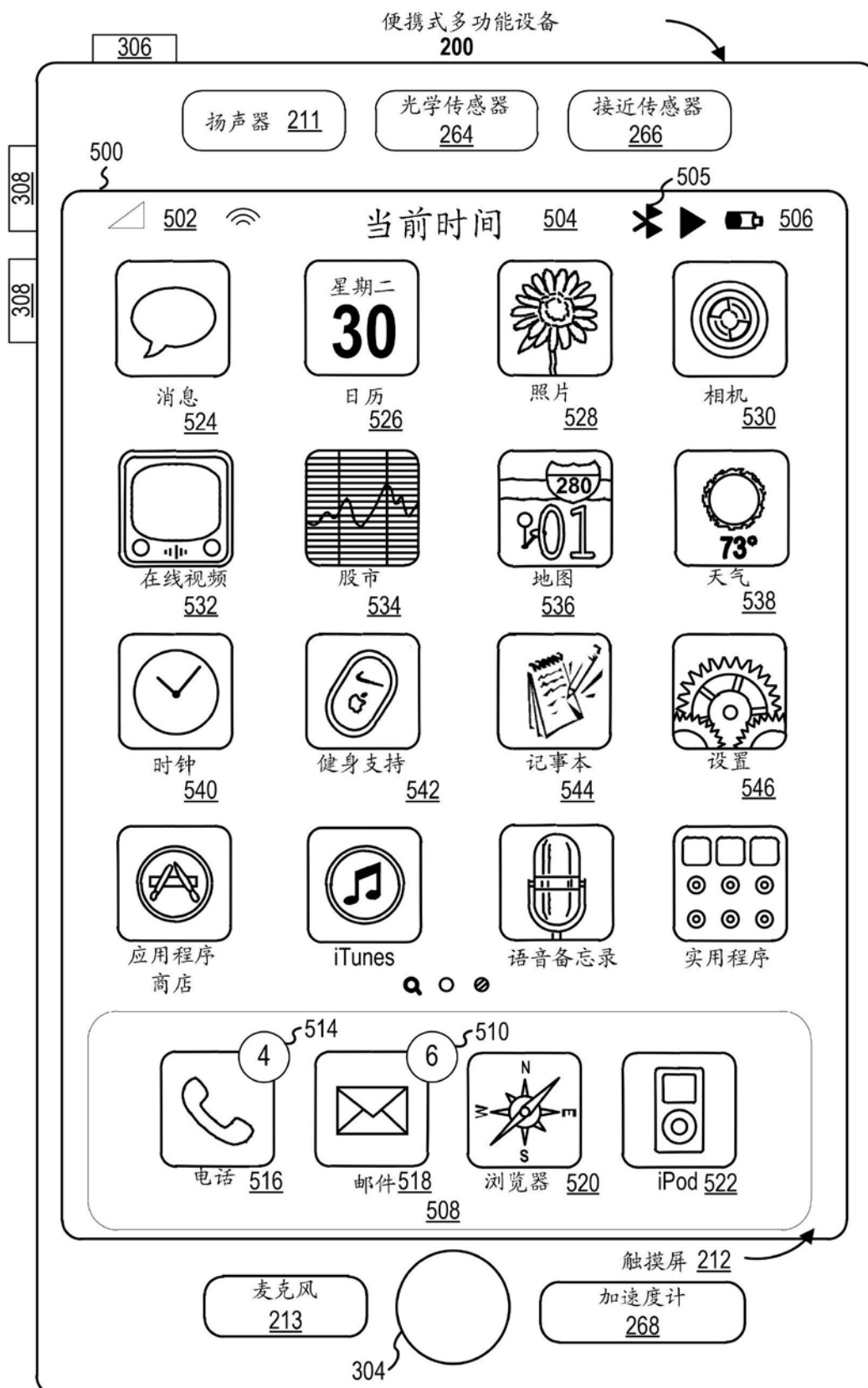


图5A

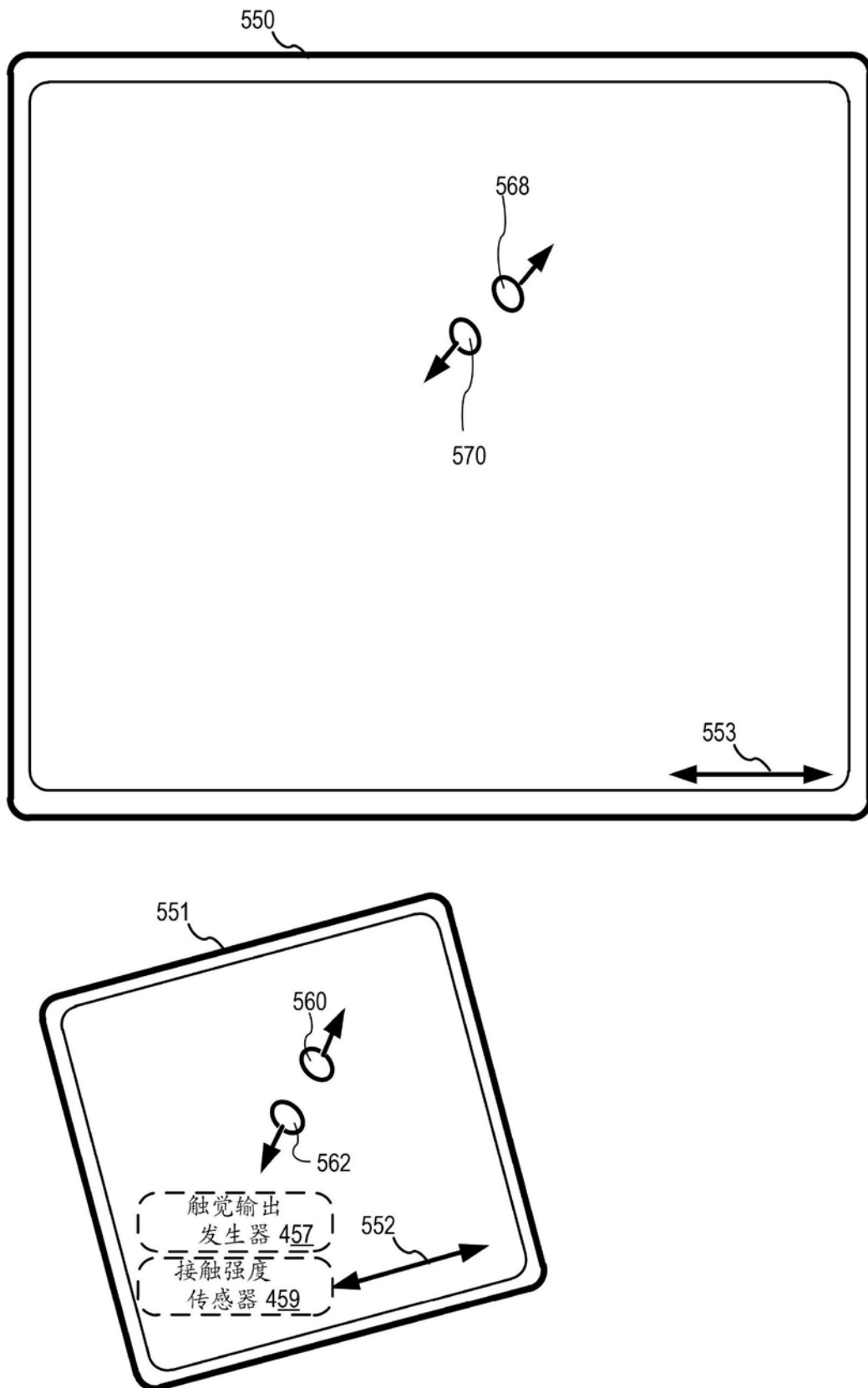


图5B



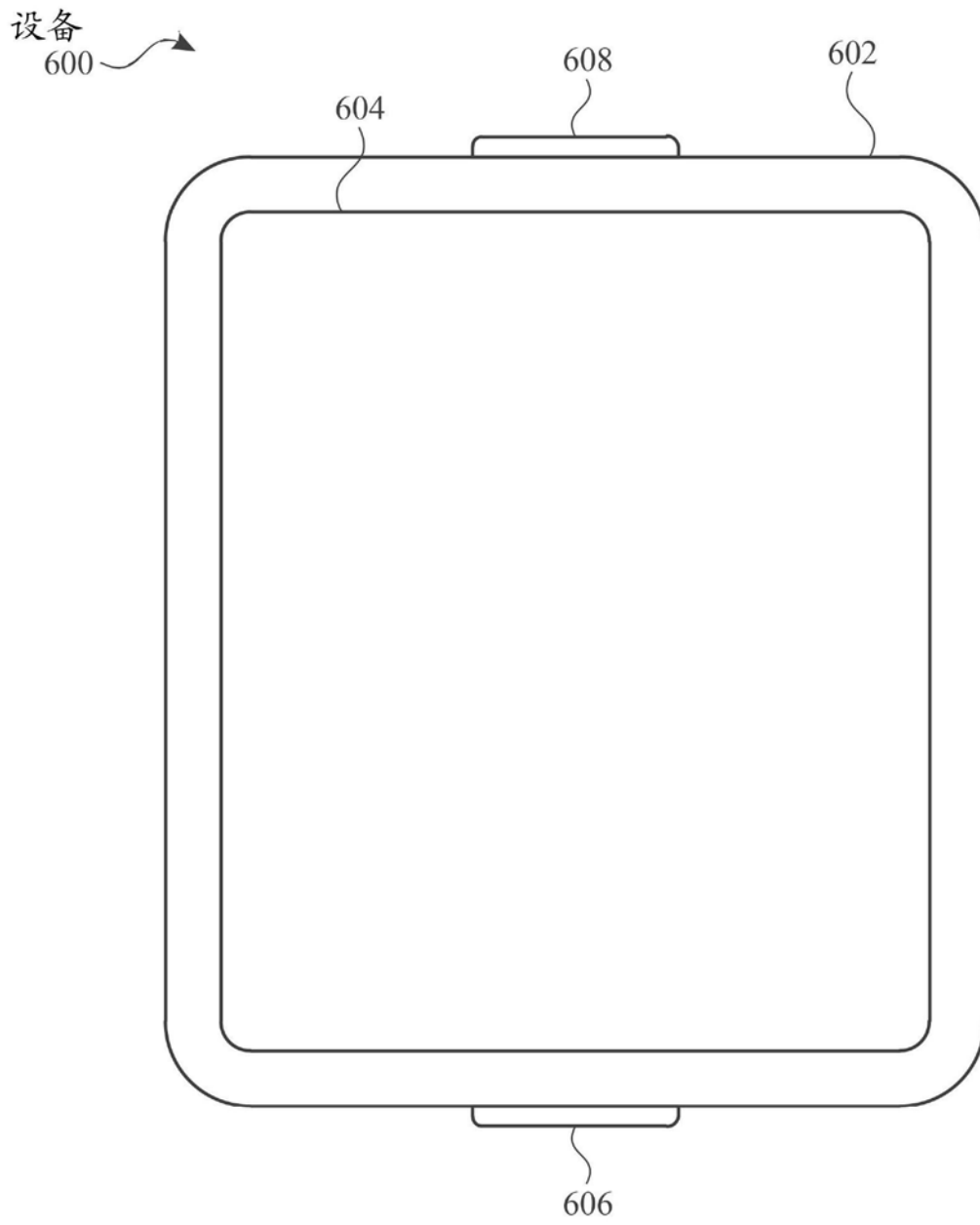


图6A

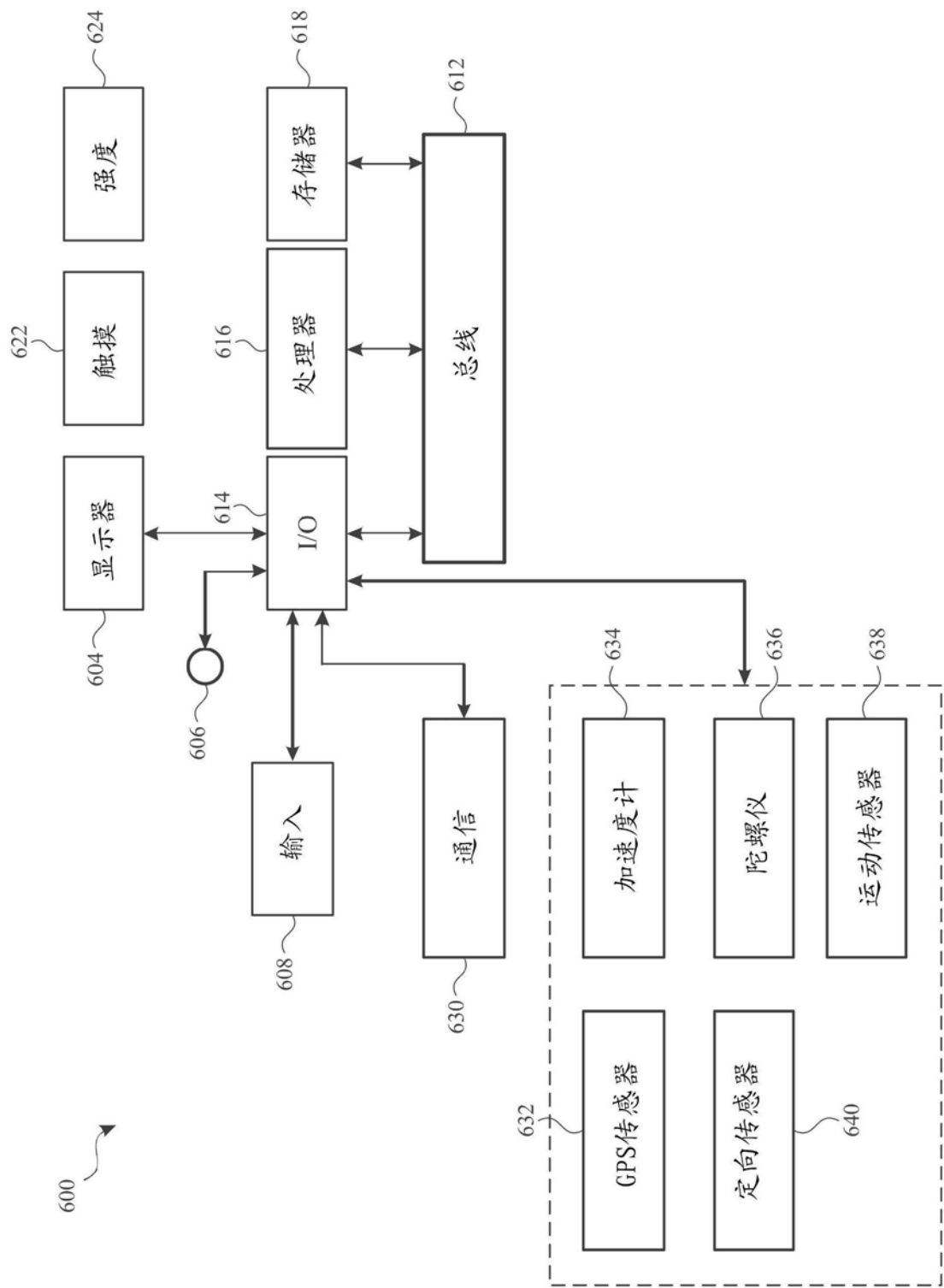


图6B

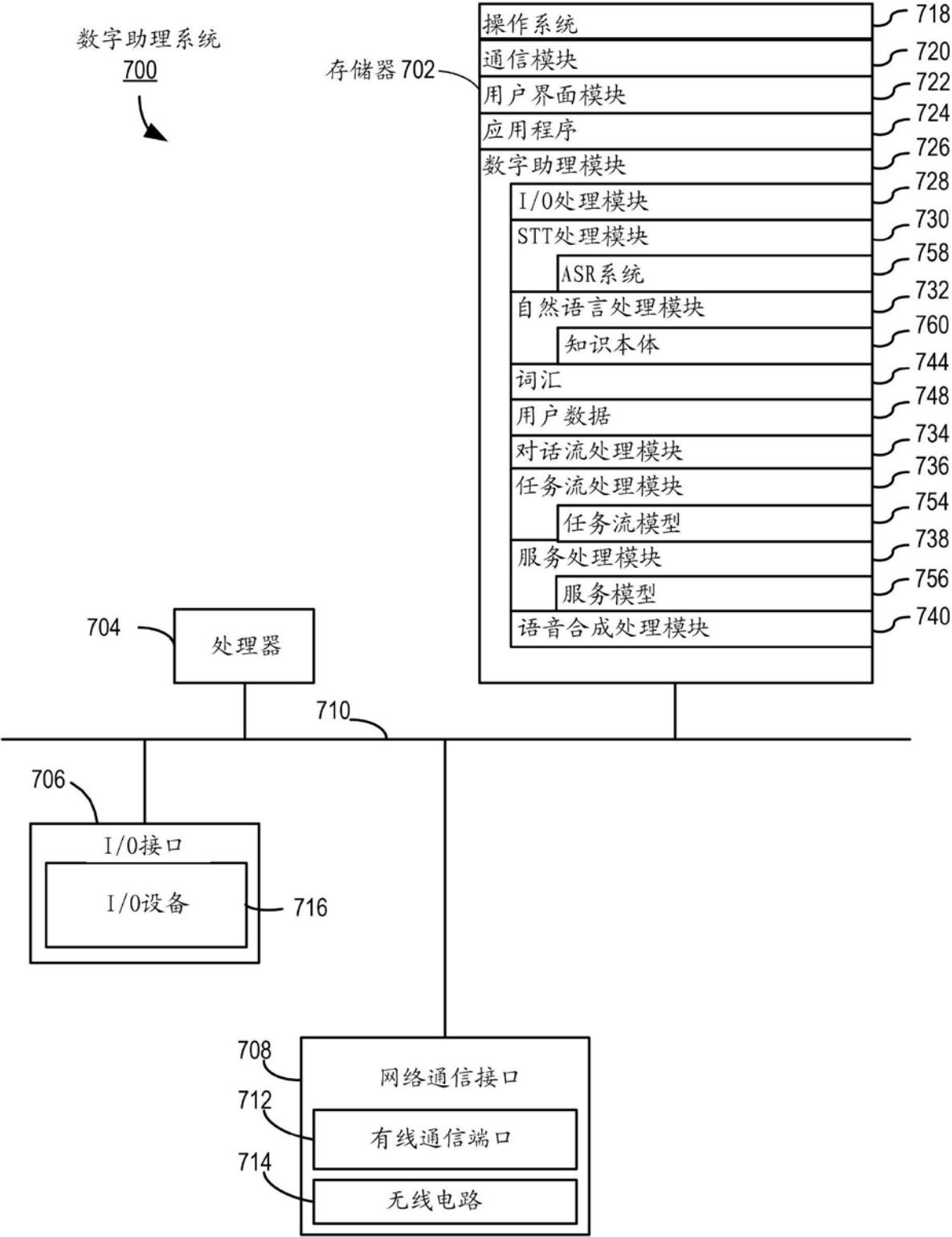


图7A

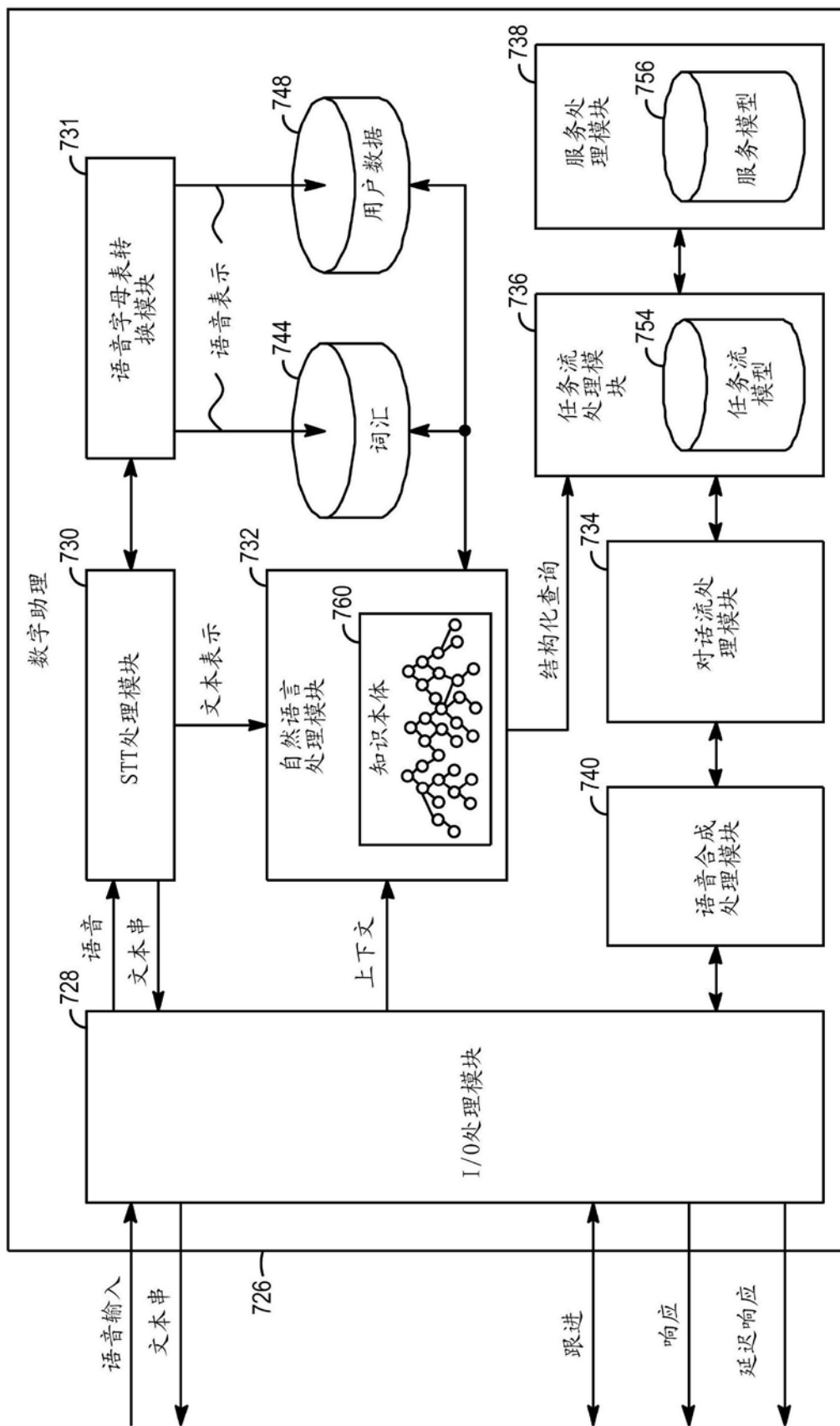


图7B

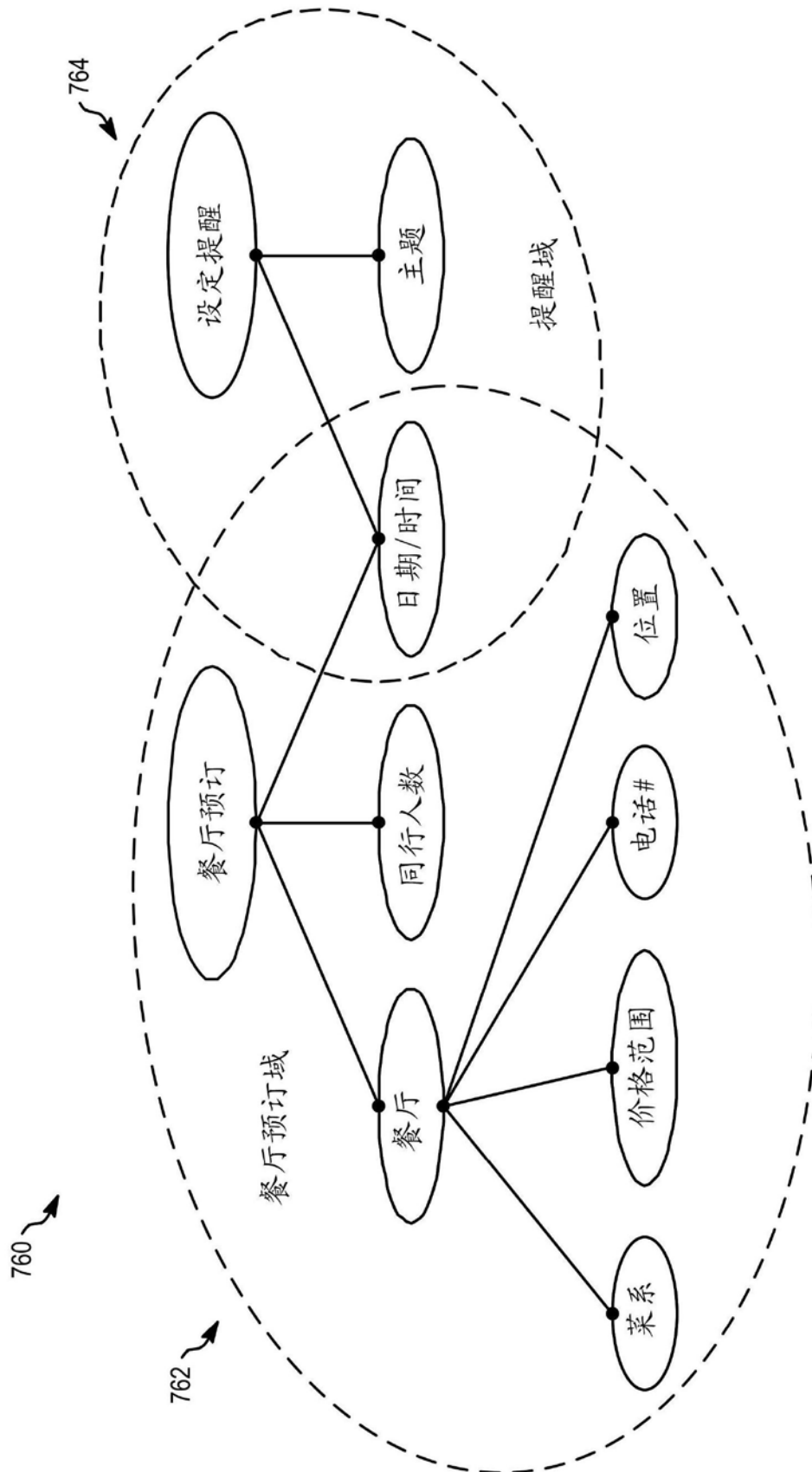


图7C

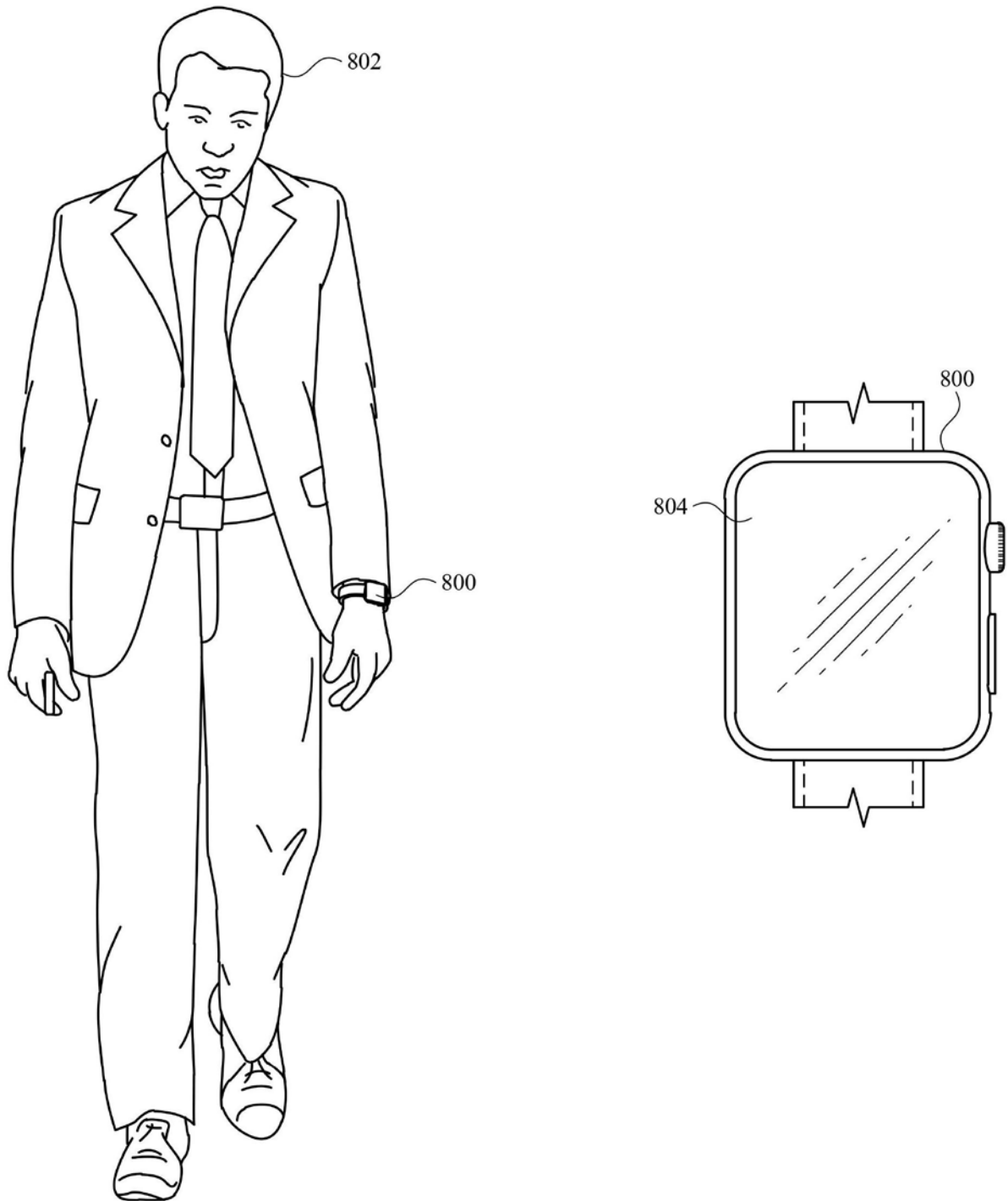


图8A

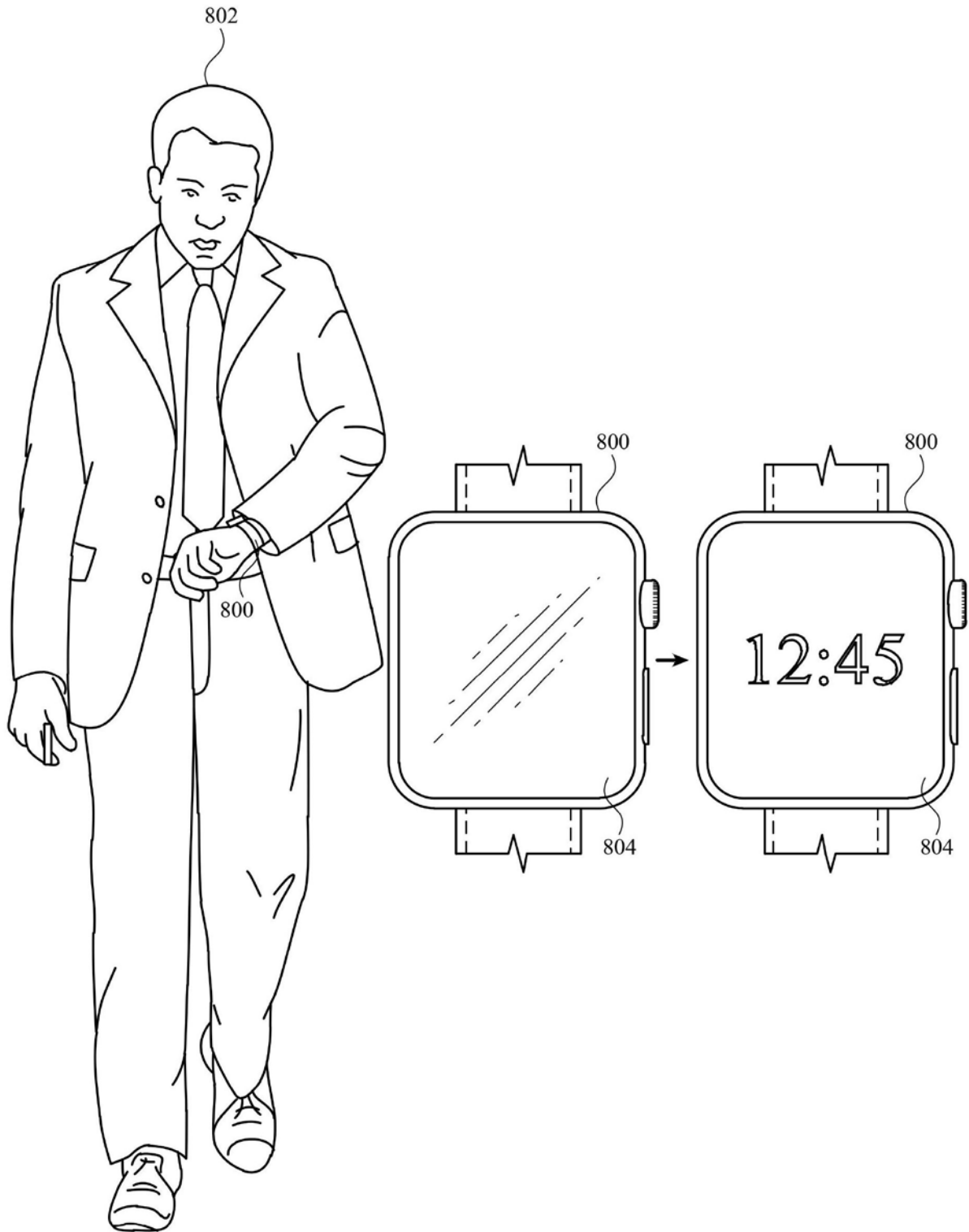


图8B

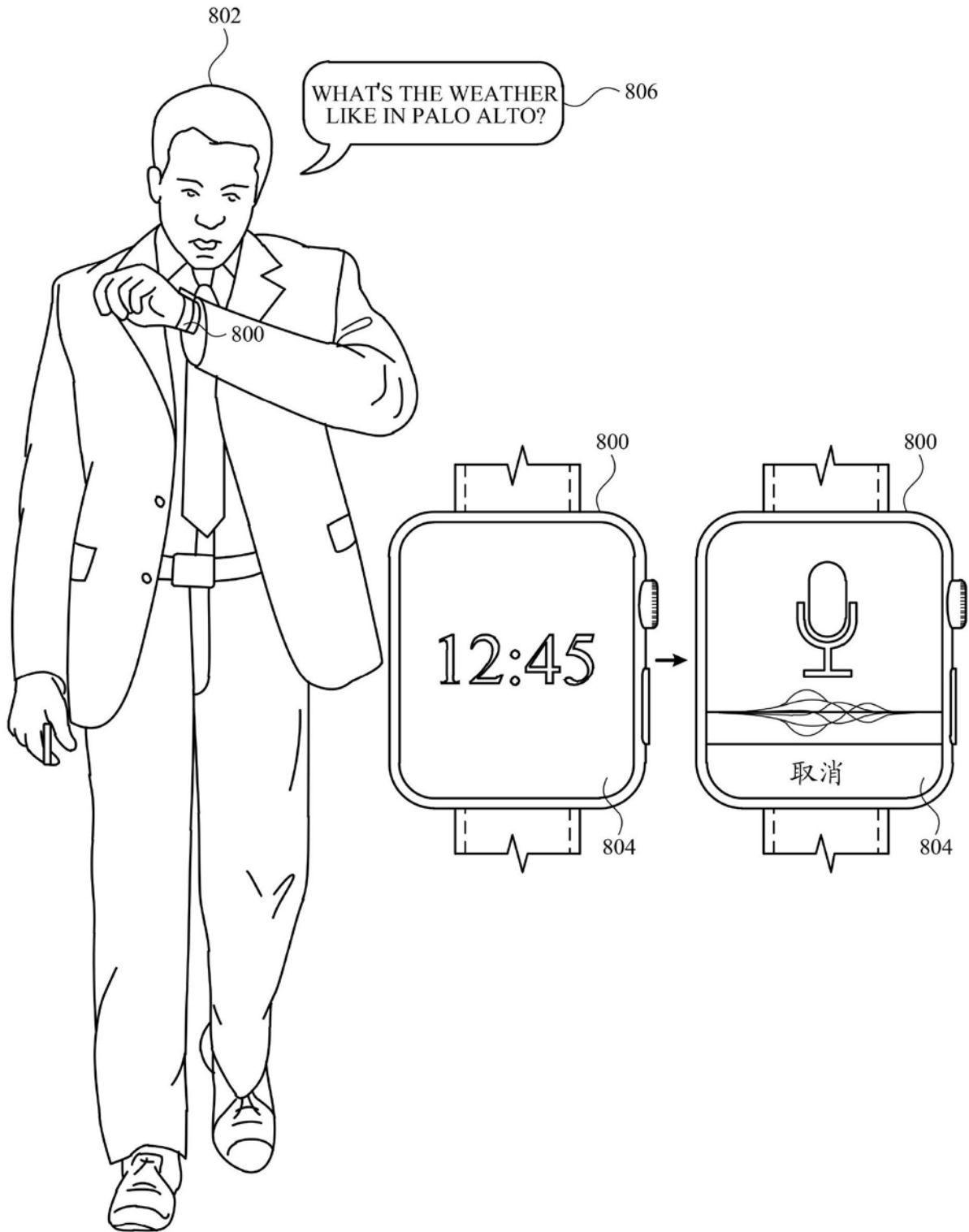


图8C



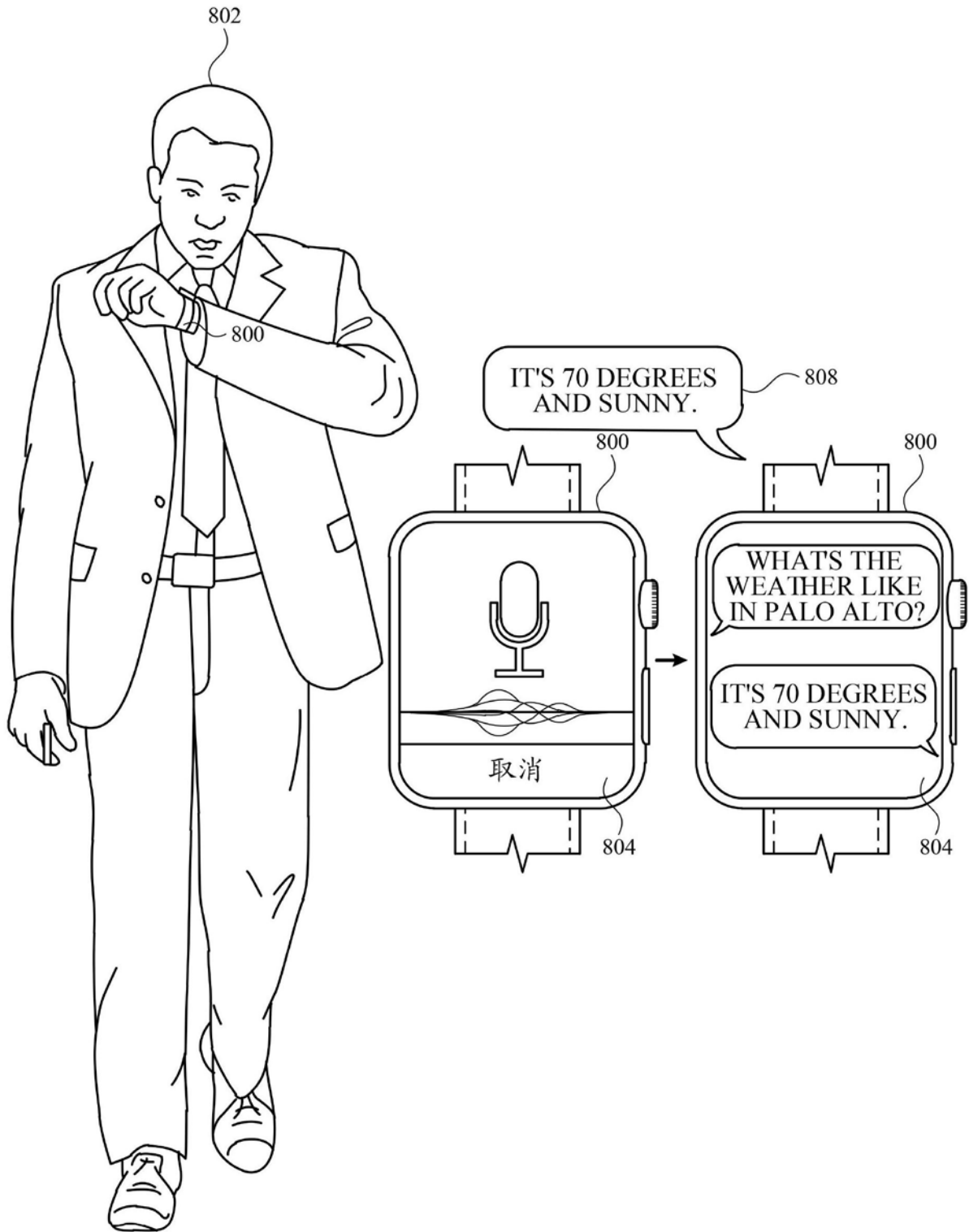


图8D

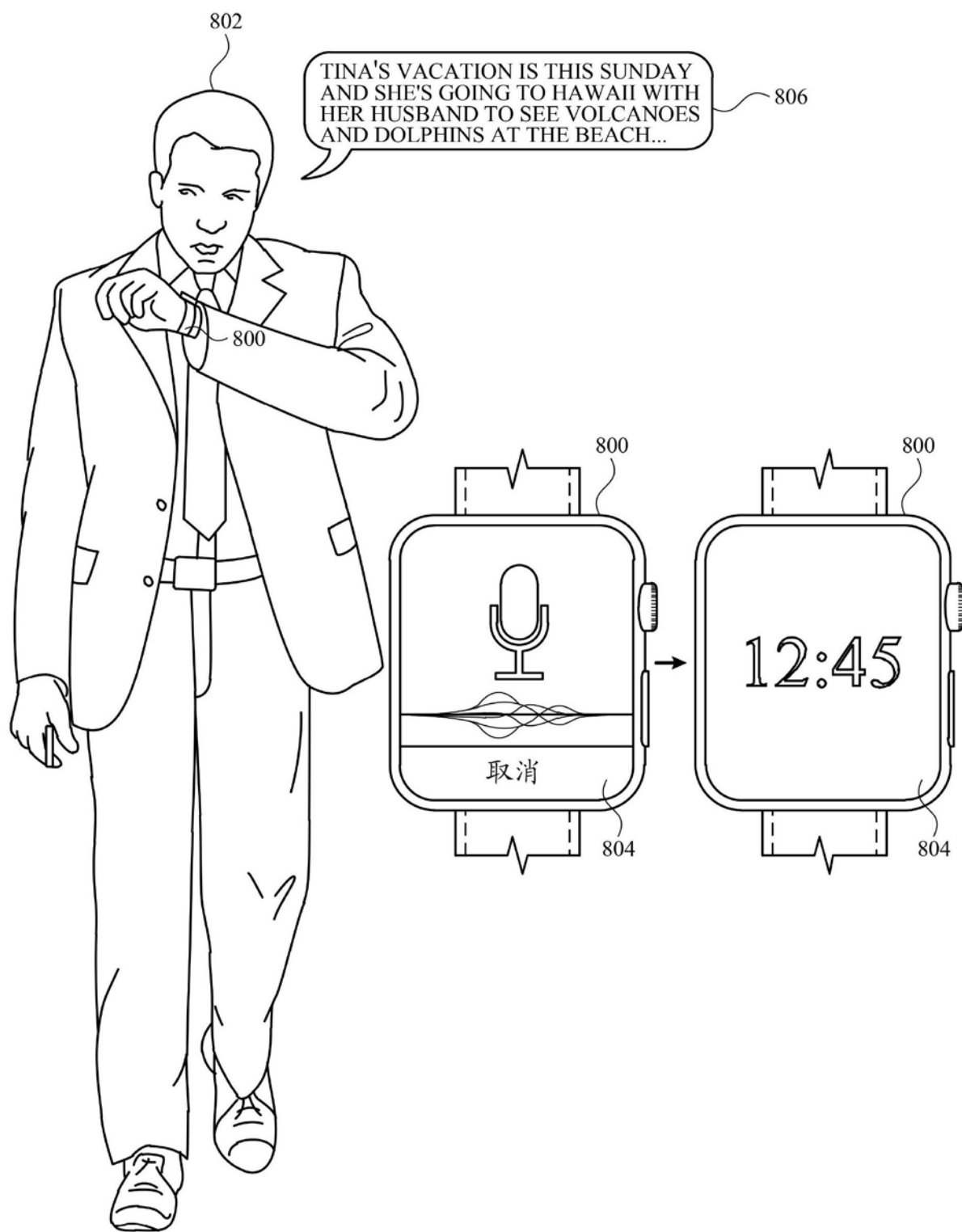


图8E

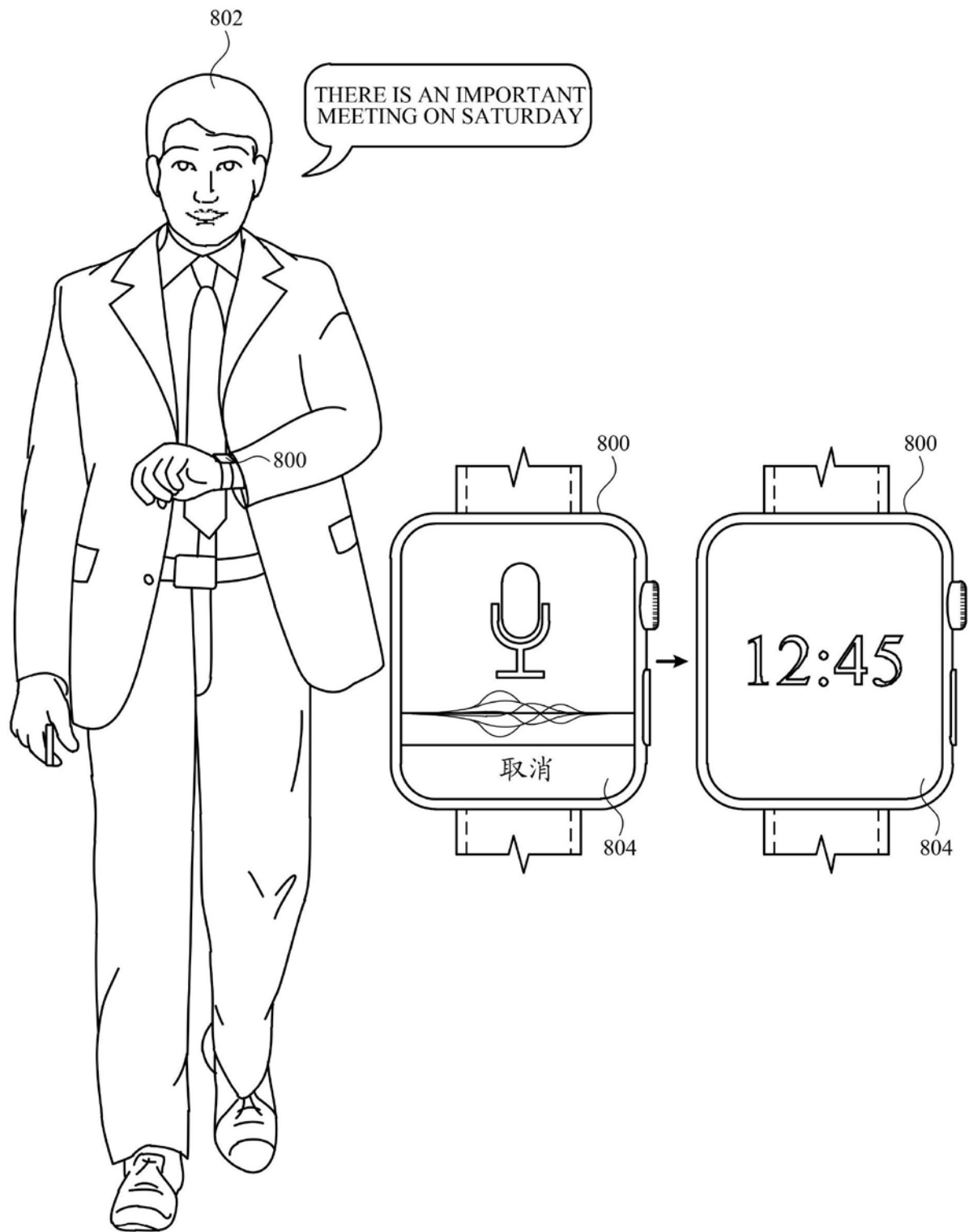


图8F

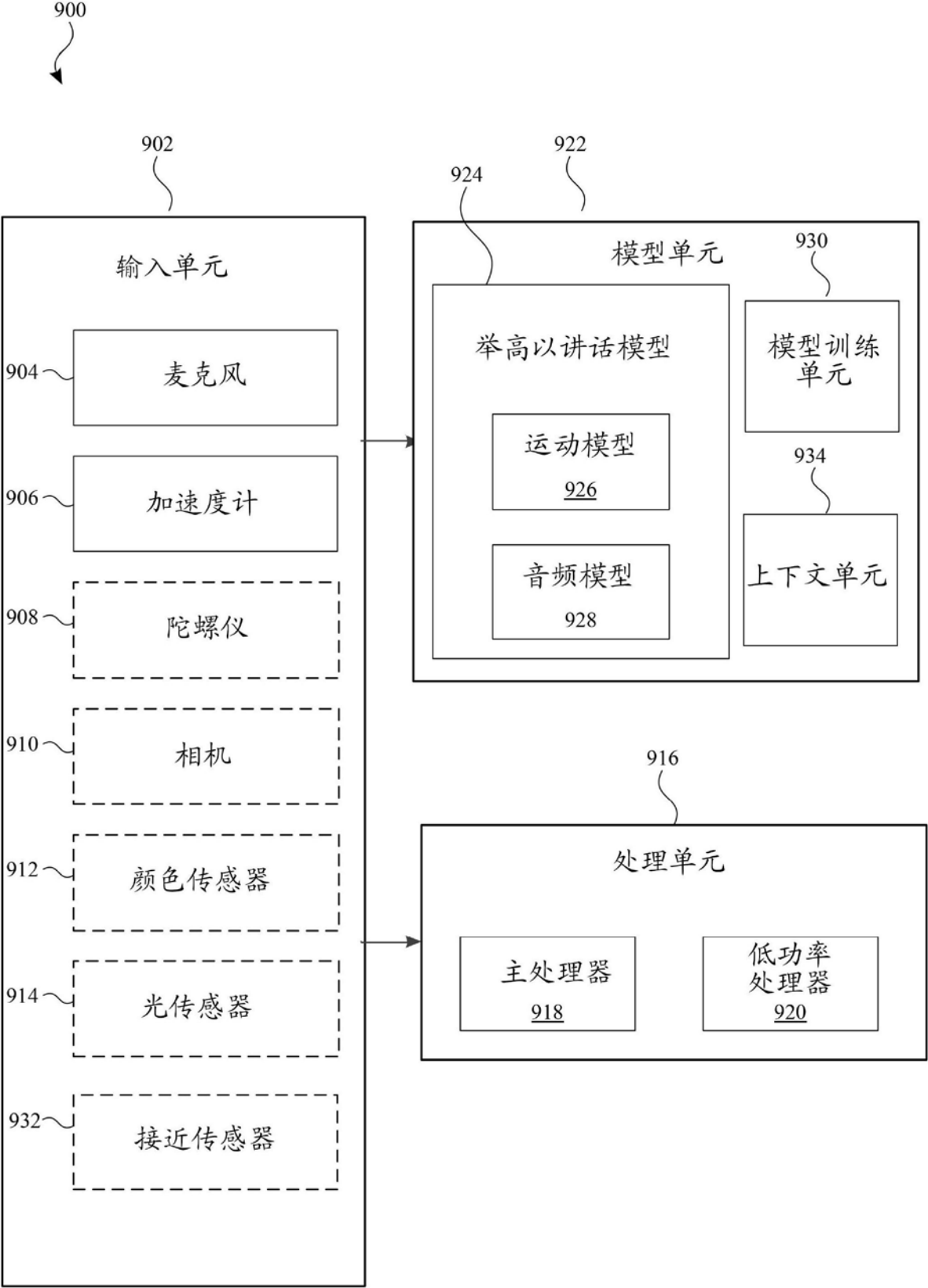


图9

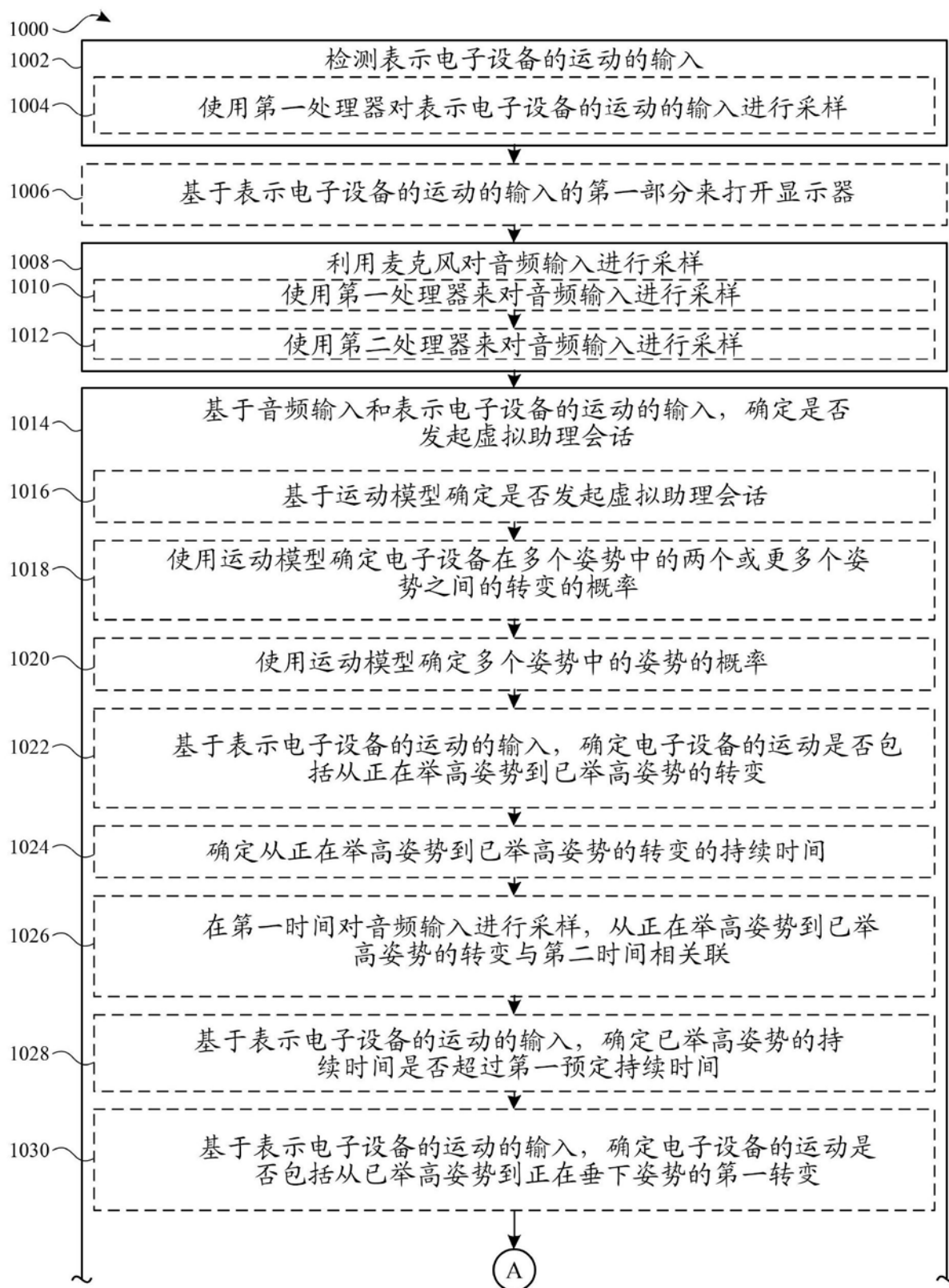


图10A

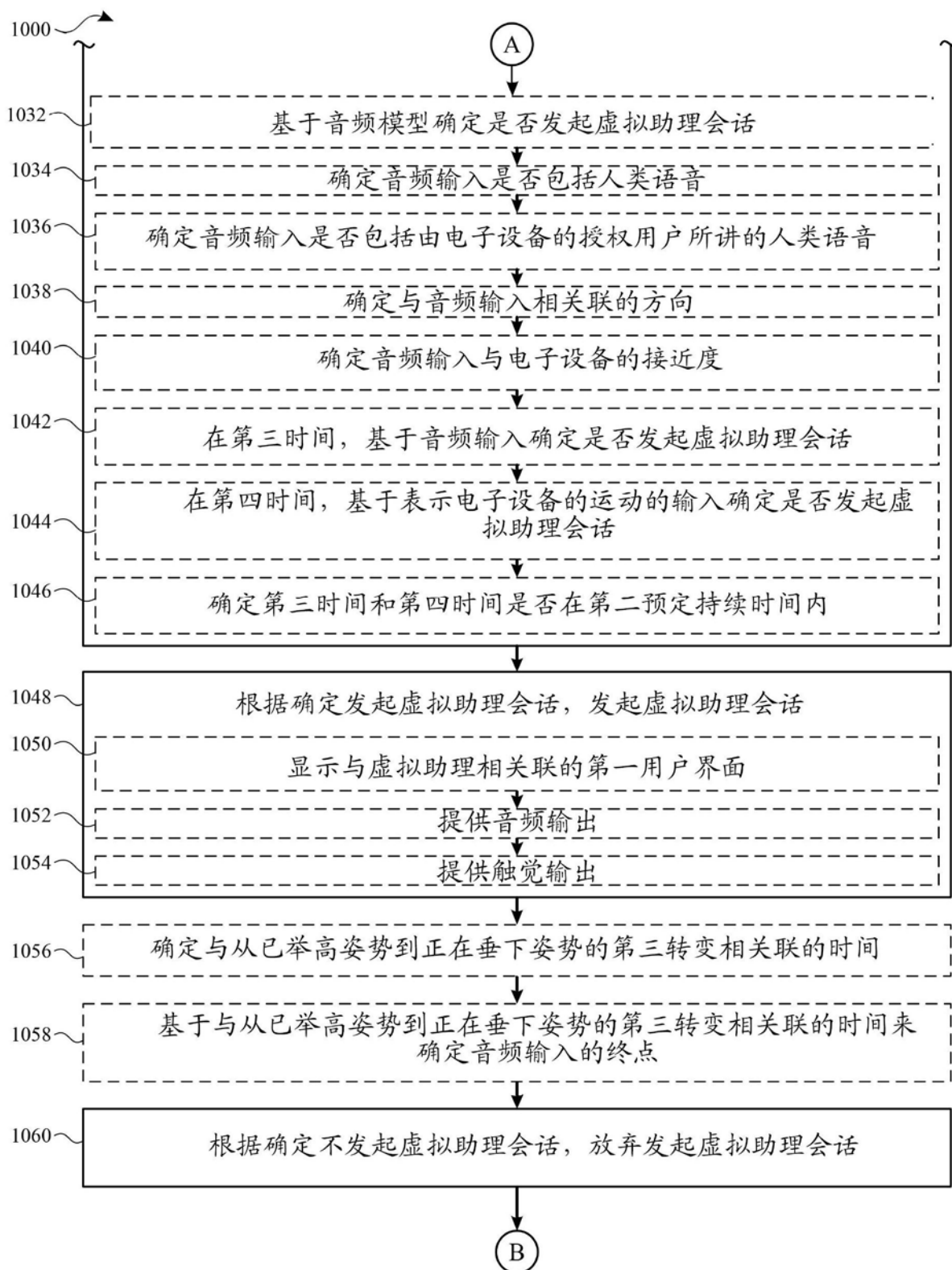


图10B

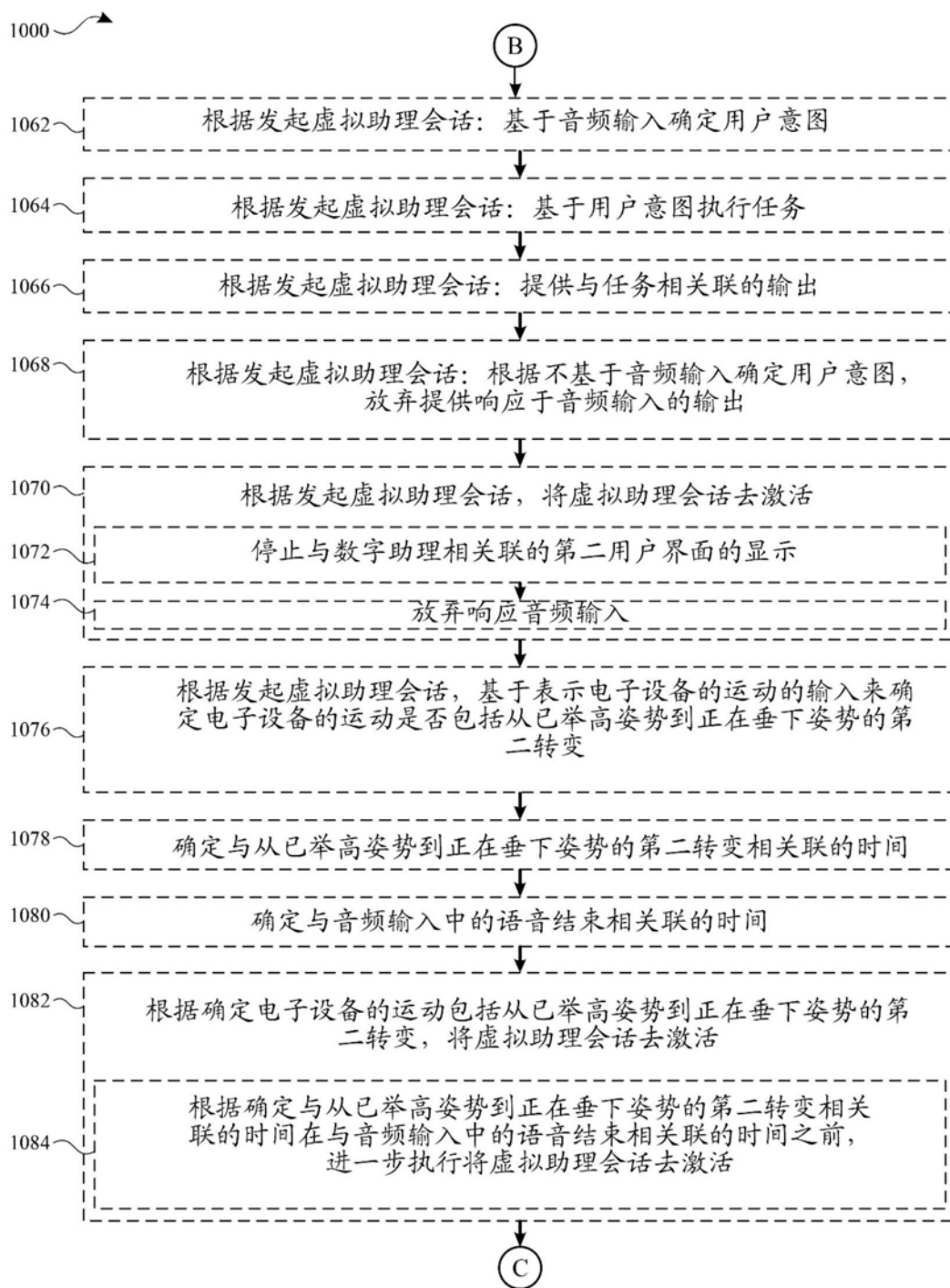


图10C

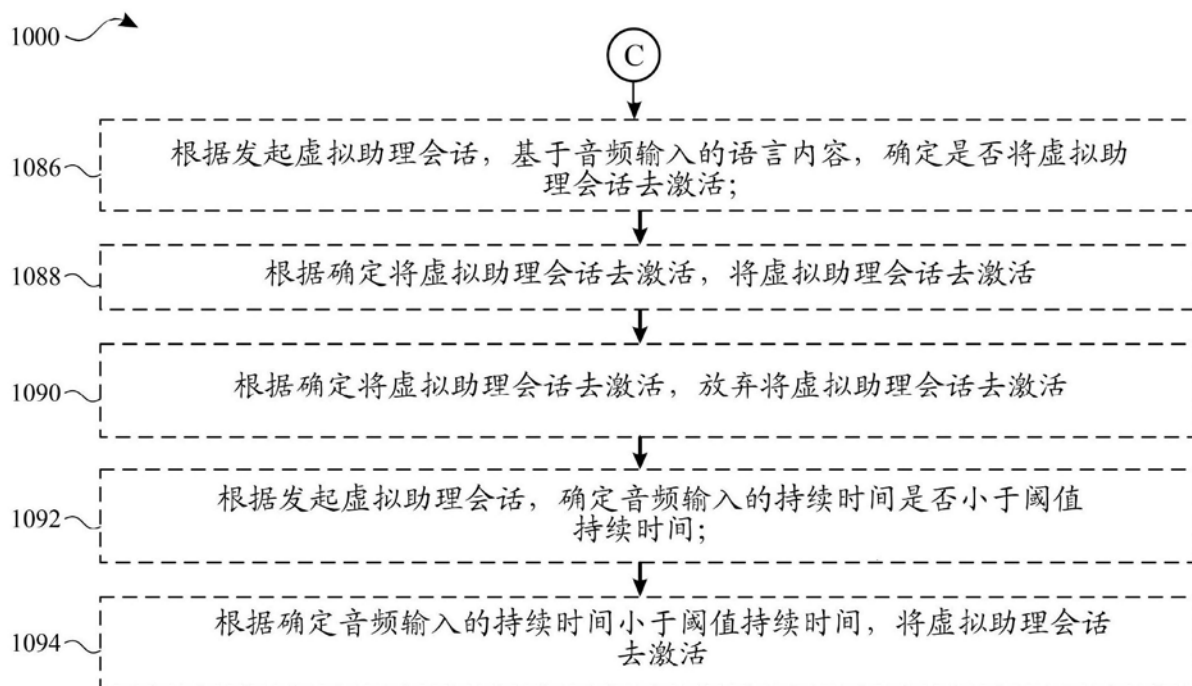


图10D



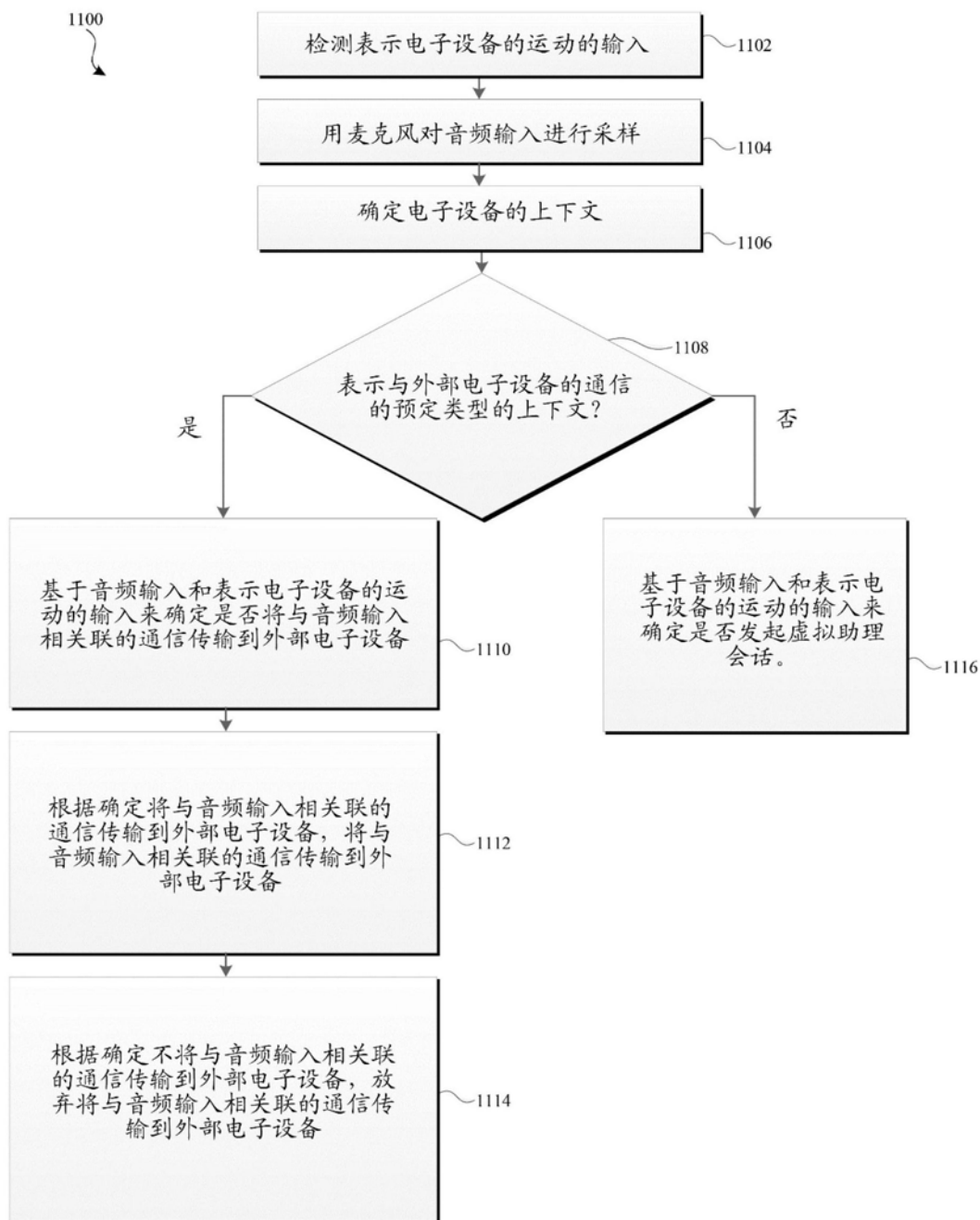


图11