



(12) 发明专利申请

(10) 申请公布号 CN 105027195 A

(43) 申请公布日 2015. 11. 04

(21) 申请号 201480011327. 7

(51) Int. Cl.

(22) 申请日 2014. 03. 11

G10L 15/00(2006. 01)

G10L 15/22(2006. 01)

(30) 优先权数据

61/785, 805 2013. 03. 14 US

(85) PCT国际申请进入国家阶段日

2015. 08. 28

(86) PCT国际申请的申请数据

PCT/US2014/023822 2014. 03. 11

(87) PCT国际申请的公布数据

W02014/159472 EN 2014. 10. 02

(71) 申请人 苹果公司

地址 美国加利福尼亚州

(72) 发明人 A · L · 拉森 S · R · 戴夫

D · 瓦罗格卢

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华 张凡

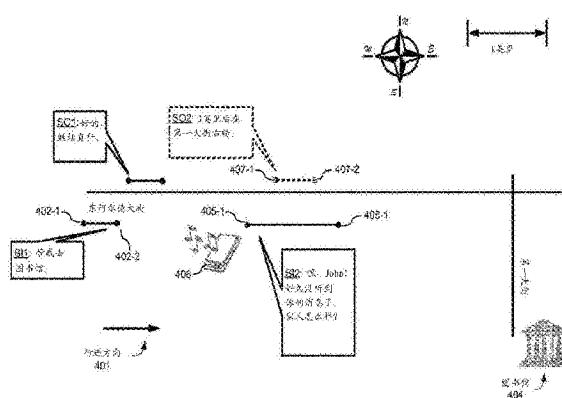
权利要求书2页 说明书18页 附图12页

(54) 发明名称

中断的上下文相关处理

(57) 摘要

本发明描述了接收要向设备的用户提供的语音输出。之后，确定所述设备当前是否正从用户接收语音输入。在确定所述设备当前未从用户接收语音输入时，向所述用户提供所述语音输出。另一方面，在确定所述设备正从用户接收语音输入时，确定提供所述语音输出是否紧急。在语音输出紧急时，向用户提供语音输出。在语音输出不紧急时，中止向用户提供语音输出。



1. 一种操作数字助理的方法,包括 :

在具有存储器和一个或多个处理器的设备处 :接收要向所述设备的用户提供的语音输出 ;

确定所述设备当前是否正从用户接收语音输入 ;

在确定所述设备当前未从所述用户接收语音输入时,向所述用户提供所述语音输出 ;

在确定所述设备正从所述用户接收语音输入时 :确定提供所述语音输出是否紧急 ;

在确定提供所述语音输出紧急时,向所述用户提供所述语音输出 ;以及

在确定提供所述语音输出不紧急时,中止向所述用户提供所述语音输出。

2. 根据权利要求 1 所述的方法,还包括在接收所述语音输出之前 :从所述用户接收执行数字助理任务的请求。

3. 根据权利要求 2 所述的方法,还包括在接收所述语音输出之前,向数字助理服务器发送所述请求,其中响应于所述请求而从所述服务器接收所述语音输出。

4. 根据权利要求 1 所述的方法,其中接收所述语音输出包括在所述设备处生成所述语音输出。

5. 根据权利要求 1 所述的方法,其中在确定提供所述语音输出紧急时,无延迟地向所述用户提供所述语音输出。

6. 根据权利要求 1 所述的方法,其中所述设备是电话,并且其中确定所述设备当前是否正从所述用户接收语音输入包括确定所述用户是否正在参与与远程用户的电话交谈。

7. 根据权利要求 6 所述的方法,其中在确定所述设备当前未从所述用户接收语音输入时,同时提供从所述远程用户接收的音频数据以及至所述用户的所述语音输出,

而不会由于所接收的音频数据而中止提供所述语音输出。

8. 根据权利要求 1 所述的方法,还包括在与所述远程用户进行电话交谈期间,向所述远程用户发送用于请求避免数字助理中断的请求。

9. 根据权利要求 1 所述的方法,还包括在确定所述设备不再从所述用户接收语音输入时,向所述用户提供所述语音输出。

10. 根据权利要求 9 所述的方法,其中确定所述设备不再从所述用户接收语音输入包括确定在上一次语音输入的时间和当前时间之间已经过去预定义的时间量。

11. 根据权利要求 10 所述的方法,其中所述预定义的时间量是所述语音输出的紧急度量的函数。

12. 根据权利要求 11 所述的方法,其中所述预先确定的时间量是所述语音输出的紧急度量的单调递减函数,由此在更少时间量中以更大紧急度量提供语音输出。

13. 根据权利要求 1 所述的方法,其中确定所述设备当前是否正从所述用户接收语音输入包括确定上一次语音输入是否是在预先确定的时间段内接收的。

14. 根据权利要求 13 所述的方法,其中所述预先确定的时间段是所述输出的紧急度量的函数。

15. 根据权利要求 1 所述的方法,还包括 :

确定所述输出是否满足消息跳过标准 ;以及

在确定所述语音输出满足消息跳过标准时,完全放弃提供所述输出。

16. 根据权利要求 15 所述的方法,其中在所述紧急度量低于预定义阈值时,满足所

述消息跳过标准。

17. 根据权利要求 15 所述的方法, 其中在所述语音输出是一组逐向指示中的导航命令并且所述设备计划在稍晚时间给出对应的导航命令时, 满足所述消息跳过标准。

18. 根据权利要求 1 所述的方法, 还包括在提供所述语音输出期间 : 从所述用户接收语音输入 ; 以及

中断语音输出。

19. 根据权利要求 16 所述的方法, 还包括 :

确定是否已满足对应于所述语音输出的完成标准 ; 在确定尚未满足所述完成标准时, 在稍晚时间中止所述语音输出的至少一部分 ;

在确定已满足所述完成标准时, 完全放弃输出所述语音输出的其余部分。

20. 根据权利要求 19 所述的方法, 其中在已经向所述用户提供了预定义百分比的所述语音输出时, 满足所述完成标准。

21. 根据权利要求 19 所述的方法, 其中所述预定义百分比的语音输出选自包括 50%、60%、70% 和 80% 的组中。

22. 根据权利要求 1 所述的方法, 其中所述设备包括显示器 ; 并且所述方法还包括在确定提供所述语音输出不紧急时, 提供对应于所述语音输出的显示输出。

23. 根据权利要求 1 所述的方法, 其中在所述设备处于特殊工作模式时, 无延迟地提供所述语音输出。

24. 根据权利要求 23 所述的方法, 其中所述特殊工作模式为包括保持工作模式和静音工作模式的组中的一者或者者。

25. 根据权利要求 1 所述的方法, 其中在所述语音输出满足用于即刻提供的用户可配置标准时, 无延迟地提供所述语音输出。

26. 根据权利要求 25 所述的方法, 其中在所述设备从先前已由所述用户识别为重要人物的人接收电子消息 (例如, SMS 或电子邮件消息) 时, 满足所述用户可配置标准。

27. 根据权利要求 25 所述的方法, 其中在所述设备接收股价更新并且其中所述用户先前已配置所述设备来即刻提供所述股价更新时, 满足所述用户可配置标准。

28. 根据权利要求 1 所述的方法, 其中在所述设备处于请勿打扰工作模式时, 放弃提供所述语音输出。

29. 一种方法, 包括方法 1-28 的任意组合的特征。

30. 一种其上存储有指令的非暂态计算机可读介质, 在由一个或多个处理器执行时, 所述指令使所述处理器执行根据权利要求 1-29 所述的方法中的任一种方法。

31. 一种系统, 包括 :

一个或多个处理器 ; 和

其上存储有指令的存储器, 在由所述一个或多个处理器执行时, 所述指令使所述处理器执行根据权利要求 1-29 所述的方法中的任一种方法。

中断的上下文相关处理

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2013 年 3 月 14 日提交的名称为“CONTEXT-SENSITIVE HANDLING OF INTERRUPTIONS”的美国临时专利申请号 61/785,805 的权益，所述专利申请以引用方式全文并入本文以用于所有目的。

技术领域

[0003] 所公开的实施例总体上涉及数字助理，并且更具体而言涉及基于当前上下文智能地处理用户发起和 / 或系统发起的中断的数字助理。

背景技术

[0004] 正如人类个人助理一样，数字助理或虚拟助理可执行所请求的任务并提供所请求的建议、信息或服务。助理满足用户请求的能力取决于助理对请求或指令的正确理解。自然语言处理的近期发展已使得用户能够利用口语或文本形式的自然语言与数字助理交互，而不是采用常规的用户界面（例如，菜单或编制命令）。此类数字助理可解译用户的输入以推断用户的意图、将推断出的意图转换成可执行的任务和参数、执行操作或部署服务以执行任务，以及产生可被用户理解的输出。理想的是，由数字助理所产生的输出应实现用户在用户和数字助理之间的自然语言交互期间所表达的意图。

[0005] 数字助理系统对用户请求产生令人满意的响应的能力取决于系统实施的自然语言处理、知识库和人工智能。设计良好的响应流程能够改善用户与系统交互时的体验并提升用户对系统服务和能力的信心。

[0006] 很多数字助理能够以语音输出的形式输送响应。例如，在某些情况下，语音输出包括由文本到语音引擎向用户大声读出的一个或多个逐向指示（例如，“在惠普尔大街左转”）。这些语音输出通常是在预先确定的时间（例如，在到达惠普尔大街之前的 1/4 英里处）或即刻（例如，在对诸如“现在什么时间？”的问题做出响应的情况下）提供的。这些数字助理响应流程的缺点是可能会在不合适的时间提供语音输出，诸如通过打断在电话交谈期间正向装置讲话的用户或向数字助理发出新请求的用户。提供不紧急信息的打断对于用户而言是令人沮丧且不方便的。此外，尽管能够听到两个不同的音频流，但人们难以在其自己讲话的同时聆听。因此，在数字助理尝试在用户讲话的同时输送语音输出时，抑制了用户理解那些语音输出的能力。

[0007] 因此，存在对智能且直观确定是否提供语音输出的操作数字助理的方法的需要。具体而言，存在对确定用户是否在讲话以及语音输出是否紧急到足以准许中断的操作数字助理的方法的需要。

发明内容

[0008] 下文描述的实施例为向用户提供语音输出提供了改进的方法。具体而言，在一些实施例中，在提供语音输出之前，数字助理会在用户结束讲话之后等待预先确定的时间量。

当然,一些语音输出比其他语音输出有更大紧迫性。例如,在数字助理计划通知用户需要在10秒钟之后在惠普尔大街左转时,用户必须要在错过转弯之前接收信息。在其他环境中,用户可能已经让数字助理在每场比赛结束时通知他或她Knicks的比分。提供指示Knicks最新比分的语音输出不那么紧急。因此,在一些实施例中,预先确定的时间量基于语音输出的紧急度。例如,在一些实施例中,默认的预先确定的时间为一秒(例如,足够长的时间,以合理确保用户已完成讲话),而相当紧急的输出(“1/4英里后左转”)将在用户已完成讲话之后半秒钟提供(例如,对应于讲话暂停的时间量)。在某些情况下,语音输出绝对紧急(“现在左转”),并且预先确定的时间量为零(例如,数字助理插话并打断用户)。

[0009] 所公开实施例通过仅以紧急消息打断正在讲话的用户而改善了语音输出响应流程和方法。这些实施例是由如下一般观察激励而出的:基于用户对两个音频输出进行上下文化而不必理解每个词语的能力,用户将具有某种能力理解两组音频输出。另一方面,用户将具有非常困难的时间在自己讲话的同时理解音频输出。因此,所公开实施例可以实现用户对语音输出助理的更大理解以及更令人满意的用户体验。

[0010] 本文公开的实施例提供了方法、系统、计算机可读存储介质和用户界面,供数字助理智能且动态确定是否提供语音输出。该方法包括接收要向设备的用户提供的语音输出以及确定设备当前是否正从用户接收语音输入。该方法还包括,在确定所述设备当前未从用户接收语音输入时,向所述用户提供所述语音输出。另一方面,在确定设备正从用户接收语音输入时,该方法还包括确定提供所述语音输出是否紧急。在语音输出紧急时,向用户提供语音输出。在语音输出不紧急时,(例如,至少暂时)中止向用户提供语音输出。

[0011] 在一些实施例中,该方法还包括在接收所述语音输出之前,从所述用户接收请求以执行数字助理任务。在一些实施例中,该方法还包括在接收所述语音输出之前,向数字助理服务器发送请求。在一些实施例中,响应于该请求从服务器接收语音输出。在一些实施例中,接收语音输出包括在设备处生成语音输出。

[0012] 在一些实施例中,该方法还包括在确定提供语音输出紧急时,无延迟地向所述用户提供语音输出。

[0013] 在一些实施例中,该设备是电话,并且确定设备当前是否正从用户接收语音输入包括确定用户是否正在参与与远程用户的电话交谈。

[0014] 在一些实施例中,在确定设备当前未从用户接收语音输入时,同时提供从远程用户接收的音频数据以及至用户的语音输出而不会由于所接收的音频数据而中止提供语音输出。

[0015] 在一些实施例中,该方法还包括在与远程用户进行电话交谈期间,向所述远程用户发送用于请求避免数字助理中断的请求。

[0016] 在一些实施例中,该方法还包括在确定所述设备不再从用户接收语音输入时,向所述用户提供所述语音输出。在一些实施例中,确定设备不再从用户接收语音输入包括确定在上一次语音输入的时间和当前时间之间已经过去预定义的时间量。在一些实施例中,预定义的时间量是语音输出的紧急度度量的函数。在一些实施例中,预先确定的时间量是语音输出的紧急度度量的单调递减函数,由此在更少时间量中以更大紧急度度量提供语音输出。

[0017] 在一些实施例中,确定设备当前是否正从用户接收语音输入包括确定上一次语音

输入是否是在预先确定的时间段内接收的。在一些实施例中，预先确定的时间段是输出的紧急度度量的函数。

[0018] 在一些实施例中，该方法还包括确定输出是否满足消息跳过标准，并且在确定语音输出满足消息跳过标准时，彻底中止提供输出。在一些实施例中，在紧急度度量低于预定义阈值时，满足消息跳过标准。在一些实施例中，在语音输出是一组逐向指示中的导航命令并且设备在稍晚时间给出对应的导航命令时，满足消息跳过标准。

[0019] 在一些实施例中，该方法还包括在提供所述语音输出期间：从用户接收语音输入；以及中止语音输出。在一些实施例中，该方法还包括确定是否已经满足对应于语音输出的完成标准。在这样的实施例中，该方法还包括在确定尚未满足完成标准时，在稍晚时间中止语音输出的至少一部分。在这样的实施例中，该方法还包括在确定已满足所述完成标准时，完全放弃输出所述语音输出的其余部分。在一些实施例中，在已经向用户提供预定义百分比的语音输出时，满足完成标准。在一些实施例中，预定义百分比的语音输出选自：50%、60%、70% 和 80%。

[0020] 在一些实施例中，该设备包括显示器。在这样的实施例中，该方法还包括在确定提供语音输出不紧急时，提供对应于语音输出的显示输出。

[0021] 在一些实施例中，该方法还包括在设备处于特殊工作模式时，无延迟地提供语音输出。在一些实施例中，特殊工作模式为选自保持工作模式和静音工作模式中的一者或者者。

[0022] 在一些实施例中，在语音输出满足用于即刻提供的用户可配置标准时，无延迟地提供语音输出。在这样的实施例中，在所述设备从先前已由用户识别为重要人物的人接收电子消息（例如 SMS 或电子邮件消息）时，满足所述用户可配置标准。在这样的实施例中，在设备接收股价更新并且其中用户先前已经配置设备即刻提供股价更新时，满足用户可配置标准。

[0023] 在一些实施例中，在设备处于请勿打扰工作模式时，放弃提供语音输出。

[0024] 在附图以及下文的描述中阐述了本说明书中所描述主题的一个或多个实施例的细节。根据说明书、附图及权利要求，本主题的其他特征、方面和优点将显而易见。

附图说明

[0025] 图 1 为示出根据一些实施例的数字助理在其中操作的环境的框图。

[0026] 图 2 为示出根据一些实施例的数字助理客户端系统的框图。

[0027] 图 3A 为示出根据一些实施例的数字助理系统或其服务器部分的框图。

[0028] 图 3B 为示出根据一些实施例的图 3A 中所示的数字助理的功能的框图。

[0029] 图 3C 为根据一些实施例的知识本体的一部分的图示。

[0030] 图 4A-4C 示出了根据一些实施例的示例性情形，其中数字助理确定是否提供语音输出。

[0031] 图 5A-5D 为根据一些实施例由数字助理实施的示例性方法的流程图，该方法基于确定装置当前是否正从用户接收语音输入以及语音输出的紧急性，确定是否向用户提供语音输出。

[0032] 相似的参考标号是指整个附图中的相应部件。

具体实施方式

[0033] 图 1 为根据一些实施例的数字助理的操作环境 100 的框图。术语“数字助理”、“虚拟助理”、“智能自动化助理”或“自动数字助理”是指解译口头和 / 或文本形式的自然语言输入以推断用户意图并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,为了进行推断出的用户意图,系统可执行以下内容中的一者或者 :通过设计用以实现所推断出的用户意图的步骤和参数来识别任务流,将来自推断出的用户意图的具体要求输入到任务流中;通过调用程序、方法、服务、API 等来执行任务流;以及生成对用户的听觉(例如,语音)和 / 或视觉形式的输出响应。

[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。针对用户请求的令人满意的响应是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,用户可能向数字助理问问题,诸如“我现在在哪里?”基于用户当前位置,数字助理可以回答“您在中央公园西门附近。”用户也可以请求执行任务,例如,“请邀请我的朋友参加下周我女朋友的生日聚会。”作为响应,数字助理可以通过说“好的,马上”来确认请求,并随后代表用户向用户电子通讯录中列出的用户的每个朋友发送适当的日历邀请。在执行请求的任务期间,数字助理有时会在连续对话中与用户交互,涉及在延长的时间段内多次交换信息。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] 在申请人 2011 年 1 月 10 日提交的针对“Intelligent Automated Assistant”的美国实用新型申请序列号 12/987,982 中描述了数字助理的实例,在此通过引用将其整个公开并入本文。

[0036] 如图 1 中所示,在一些实施例中,数字助理根据客户端 - 服务器模型来实施。数字助理包括在用户设备 104a,104b 上执行的客户端侧部分 102a,102b(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 为任意数量的 DA 客户端 102 提供服务器侧功能,所述任意数量的 DA 客户端 102 各自位于相应的用户设备 104 上。

[0037] 在一些实施例中,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 接口促成数字助理服务器 106 的面向客户端的输入和输出处理。所述一个或多个处理模块 114 利用数据与模型 116 基于自然语言输入来确定用户的意图,并基于推断出的用户意图来进行任务执行。在一些实施例中,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。

[0038] 用户设备 104 的实例包括但不限于手持式计算机、个人数字助理(PDA)、平板电脑、膝上型计算机、台式计算机、蜂窝电话、智能电话、增强型通用分组无线电服务(EGPRS)移动电话、媒体播放器、导航设备、游戏控制器、电视机、遥控器,或者这些数据处理设备中

任意两种或更多种的组合或其他数据处理设备。有关用户设备 104 的更多细节参照图 2 中示出的示例性用户设备 104 提供。

[0039] 一个或多个通信网络 110 的实例包括局域网（“LAN”）和广域网（“WAN”），例如互联网。该一个或多个通信网络 110 可使用任何已知的网络协议，包括各种有线或无线协议，诸如例如以太网、通用串行总线（USB）、火线、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、互联网电话协议（VoIP）、Wi-MAX、或任何其他合适的通信协议来实施。

[0040] 服务器系统 108 实现于一个或多个独立的数据处理设备或计算机的分布式网络上。在一些实施例中，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。

[0041] 尽管图 1 中示出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，但在一些实施例中，数字助理的功能被实现为安装于用户设备上的独立应用。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实施例中可变化。例如，在一些实施例中，DA 客户端为仅提供面向用户的输入和输出处理功能，且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。

[0042] 图 2 为根据一些实施例的用户设备 104 的框图。用户设备 104 包括存储器接口 202、一个或多个处理器 204，以及外围设备接口 206。用户设备 104 中的各种部件通过一条或多条通信总线或信号线耦接。用户设备 104 包括各种传感器、子系统、以及耦接至外围设备接口 206 的外围设备。传感器、子系统、以及外围设备采集信息和 / 或促成用户设备 104 的各种功能。

[0043] 例如，运动传感器 210、光传感器 212、以及接近传感器 214 耦接至外围设备接口 206 以方便取向、照明和接近感测功能。一个或多个其他传感器 216，诸如定位系统（例如，GPS 接收器）、温度传感器、生物识别传感器、陀螺仪、罗盘、加速度计等，也连接至外围设备接口 206，以促成相关功能。

[0044] 在一些实施例中，利用相机子系统 220 和光学传感器 222 促成相机功能，诸如拍摄照片和记录视频剪辑。通过一个或多个有线和 / 或无线通信子系统 224 来方便通信功能，所述通信子系统可包括各种通信端口、射频接收器与发射器和 / 或光学（如红外）接收器与发射器。将音频子系统 226 耦合到扬声器 228 和麦克风 230 以有利于启用语音的功能，诸如语音识别、语音复制、数字记录和电话功能。

[0045] 在一些实施例中，I/O 子系统 240 还耦接至外围设备接口 206。I/O 子系统 240 包括触摸屏控制器 242 和 / 或一个或多个其他输入控制器 244。触摸屏控制器 242 耦接至触摸屏 246。触摸屏 246 和触摸屏控制器 242 能够例如使用多种触摸灵敏度技术中的任一种来检测接触和移动或它们的间断，所述多种触摸灵敏度技术诸如电容技术、电阻技术、红外技术、表面声波技术、接近传感器阵列等。可将一个或多个其他输入控制器 244 耦合到其他输入 / 控制设备 248，例如一个或多个按钮、摇臂开关、拇指滚轮、红外端口、USB 端口和 / 或指针装置（诸如触笔）。

[0046] 在一些实施例中，存储器接口 202 耦合至存储器 250。存储器 250 可包括高速随机存取存储器和 / 或非易失性存储器，例如一个或多个磁盘存储设备、一个或多个光学存储

设备,和 / 或闪存存储器(例如NAND、NOR)。

[0047] 在一些实施例中,存储器250存储操作系统252、通信模块254、用户界面模块256、传感器处理模块258、电话模块260和应用程序262。操作系统252包括用于处理基础系统服务以及用于执行硬件相关任务的指令。通信模块254促成与一个或多个附加设置、一个或多个计算机和 / 或一个或多个服务器的通信。用户界面模块256利用其他输出通道(例如,扬声器)促成图形用户界面处理和输出处理。传感器处理模块258促成与传感器有关的处理和功能。电话模块260促成与电话有关的过程和功能。应用程序模块262促成用户应用程序的各种功能,诸如电子消息传送、网页浏览、媒体处理、导航、成像和 / 或其他过程和功能。

[0048] 如上所述,在本说明书中,存储器250还存储客户端侧数字助理指令(例如,在数字助理客户端模块264中)以及各种用户数据266(例如,特定于用户的词汇数据、偏好数据,和 / 或其他数据诸如用户的电子通讯录、待办事项列表、购物清单等)以提供数字助理的客户端侧功能。

[0049] 在各种实施例中,数字助理客户端模块264能够通过用户设备104的各种用户界面(例如,I/O子系统244)接受语音输入(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264还能够提供音频(例如,语音输入)、视频和 / 或触觉形式的输出。例如,可将输出提供为语音、声音、警报、文本消息、菜单、图形、视频、动画、振动、和 / 或以上中的两者或更多者的组合。在操作期间,数字助理客户端模块264使用通信子系统224来与数字助理服务器通信。

[0050] 在一些实施例中,数字助理客户端模块264利用各种传感器、子系统和外围设备来从用户设备104的周围环境采集附加信息以建立与用户、当前用户交互和 / 或当前用户输入相关联的上下文。在一些实施例中,数字助理客户端模块264将上下文信息或其子集与用户输入一起提供至数字助理服务器以帮助推断用户的意图。在一些实施例中,数字助理还使用上下文信息确定如何准备并向用户输送输出。

[0051] 在一些实施例中,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。在一些实施例中,上下文信息还包括设备的物理状态,例如,设备取向、设备位置、设备温度、功率水平、速度、加速度、运动模式、蜂窝信号强度等。在一些实施例中,向数字助理服务器提供与用户设备104的软件状态相关的信息,例如用户设备104的运行中的过程、安装的程序、过去和当前的网络活动、背景服务、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。

[0052] 在一些实施例中,DA客户端模块264响应于来自数字助理服务器的请求而选择性地提供存储在用户设备104上的信息(例如,用户数据266)。在一些实施例中,数字助理客户端模块264还在数字助理服务器106请求时引出来自用户的经由自然语言对话或其他用户界面的附加输入。数字助理客户端模块264将所述附加输入传送至数字助理服务器106,以帮助数字助理服务器106进行意图推断和 / 或满足在用户请求中表达的用户意图。

[0053] 在各实施例中,存储器250包括附加指令或更少的指令。此外,可在硬件和 / 或固件中,包括在一个或多个信号处理和 / 或专用集成电路中,执行用户设备104的各种功能。

[0054] 图3A为根据一些实施例的实例数字助理系统300的框图。在一些实施例中,数字助理系统300在独立式计算机系统上实施。在一些实施例中,数字助理系统300跨多个计

算机分布。在一些实施例中，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用户设备（例如，用户设备 104）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。在一些实施例中，数字助理系统 300 为图 1 中所示的服务器系统 108（和 / 或数字助理服务器 106）的实施例。应当指出，数字助理系统 300 仅为数字助理系统的一个实例，并且该数字助理系统 300 可具有比示出更多或更少的部件、可组合两个或更多个部件、或可具有部件的不同配置或布局。图 3A 中所示的各种部件可以在硬件、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。

[0055] 数字助理系统 300 包括存储器 302、一个或多个处理器 304、输入 / 输出 (I/O) 接口 306，以及网络通信接口 308。这些部件通过一条或多条通信总线或信号线 310 彼此通信。

[0056] 在一些实施例中，存储器 302 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、闪存存储器设备、或其他非易失性固态存储器设备）。

[0057] 在一些实施例中，I/O 接口 306 将数字助理系统 300 的输入 / 输出设备 316 诸如显示器、键盘、触摸屏、和麦克风耦接至用户界面模块 322。I/O 接口 306，与用户界面模块 322 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。在一些实施例中，例如，当数字助理在独立式用户设备上实施时，数字助理系统 300 包括关于图 2 中的用户设备 104 所描述的部件和 I/O 接口及通信接口中的任一者。在一些实施例中，数字助理系统 300 代表数字助理具体实施的服务器部分，并且通过位于用户设备（例如，图 2 中所示的用户设备 104）上的客户端侧部分与用户进行交互。

[0058] 在一些实施例中，网络通信接口 308 包括一个或多个有线通信端口 312 和 / 或无线传输和接收电路 314。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 314 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。无线通信可使用多种通信标准、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 308 使数字助理系统 300 与网络，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)，以及其他设备之间的通信成为可能。

[0059] 在一些实施例中，存储器 302 或存储器 302 的计算机可读存储介质存储程序、模块、指令和数据结构，包括以下内容中的全部或其子集：操作系统 318、通信模块 320、用户界面模块 322、一个或多个应用程序 324、和数字助理模块 326。一个或多个处理器 304 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。

[0060] 操作系统 318（例如，Darwin、RTXC、LINUX、UNIX、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、电力管理等）的各种软件部件和 / 或驱动器，并促成各种硬件、固件与软件部件之间的通信。

[0061] 通信模块 320 促成数字助理系统 300 与其他设备之间通过网络通信接口 308 进行的通信。例如，通信模块 320 可与图 2 中所示的设备 104 的通信接口 254 通信。通信模块 320 还包括各种部件，用于处理被无线电路 314 和 / 或有线通信端口 312 接收的数据。

[0062] 用户界面模块 322 经由 I/O 接口 306 接收来自用户（例如，来自键盘、触摸屏、定点

设备、控制器和 / 或麦克风) 的命令和 / 或输入，并在显示器上生成用户界面对象。用户界面模块 322 还准备并经由 I/O 接口 306(例如，通过显示器、音频通道、扬声器和触摸板等) 向用户输送输出 (例如，语音、声音、动画、文本、图标、振动、触觉反馈和光等) 。

[0063] 应用程序 324 包括被配置为由所述一个或多个处理器 304 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 324 可包括用户应用程序，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 300 在服务器场上实施，则应用程序 324 可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。

[0064] 存储器 302 还存储数字助理模块 (或数字助理的服务器部分) 326 。在一些实施例中，数字助理模块 326 包括以下子模块或其子集或超集：输入 / 输出处理模块 328 、语音转文本 (STT) 处理模块 330 、自然语言处理模块 332 、对话流处理模块 334 、任务流处理模块 336 、服务处理模块 338 和中断处理模块 340 。这些模块中的每一个均具有对以下数字助理 326 的数据和模型中的一者或者其子集或超集的访问权限：知识本体 360 、词汇索引 344 、用户数据 348 、任务流模型 354 、服务模型 356 和优先级参数数据库 358 。

[0065] 在一些实施例中，利用数字助理模块 326 中实现的处理模块、数据和模型，数字助理执行如下操作中的至少一些：识别从用户接收的以自然语言输入表达的用户意图；主动引出并获得完整推断用户意图所需的信息 (例如，通过不模棱两可的词汇、姓名、意图等) ；确定完成所推断意图的任务流；以及执行任务流以完成推断的意图。在本说明书中，稍后提供关于中断处理模块及其优先级参数使用的更多细节。

[0066] 如图 3B 中所示，在一些实施例中， I/O 处理模块 328 通过图 3A 中的 I/O 设备 316 与用户交互或通过图 3A 中的网络通信接口 308 与用户设备 (例如，图 1 中的用户设备 104) 交互以获得用户输入 (例如，语音输入) 以及提供对用户输入的响应 (例如，语音输出) 。 I/O 处理模块 328 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 data 、词汇、和 / 或与用户输入有关的偏好。在一些实施例中，上下文信息还包括在接收到用户请求时所述设备 (例如，图 1 中的用户设备 104) 的软件和硬件状态，和 / 或与在接收到用户请求时用户的周围环境相关的信息。在一些实施例中， I/O 处理模块 328 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。当用户请求被 I/O 处理模块 328 接收到且用户请求包含语音输入时， I/O 处理模块 328 将语音输入转发至语音转文本 (STT) 处理模块 330 以进行语音转文本转换。

[0067] 语音转文本处理模块 330 通过 I/O 处理模块 328 接收语音输入 (例如，在语音记录中捕捉的用户话语) 。在一些实施例中，语音转文本处理模块 330 使用各种声音和语言模型来将语音输入识别为音素的序列，并最终将其识别为以一种或多种语言书写的字词或符号的序列。语音转文本处理模块 330 可以使用任何合适的语音识别技术、声音模型以及语言模型，诸如隐马尔可夫模型、基于动态时间规正 (DTW) 的语音识别以及其他统计和 / 或分析技术，来加以实施。在一些实施例中，语音转文本处理可至少部分地由第三方服务执行或在用户的设备上执行。一旦语音转文本处理模块 330 获得语音转文本处理的结果，例如字词或符号的序列，它便将结果传送至自然语言处理模块 332 以进行意图推断。

[0068] 数字助理的自然语言处理模块 332 (“自然语言处理器”) 取得由语音转文本处理

模块 330 生成的字词或符号的序列（“符号序列”），并尝试将该符号序列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”表示可由数字助理执行并且具有在任务流模型 354 中实施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 354 中实施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。

[0069] 在一些实施例中，除从语音转文本处理模块 330 获得的字词或符号的序列之外，自然语言处理器 332 还接收与用户请求相关联的上下文信息，例如，来自 I/O 处理模块 328。自然语言处理器 332 任选地使用上下文信息来明确、补充和 / 或进一步定义在从语音转文本处理模块 330 接收的符号序列中包含的信息。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。如本说明书中所述，上下文信息是动态的，并可能随着时间、位置、对话内容和其他因素变化。

[0070] 在一些实施例中，自然语言处理是基于例如知识本体 360 进行的。知识本体 360 为包含很多节点的分级结构，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，即其“可执行”或可被进行。“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 360 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代表的参数如何从属于由可执行意图节点代表的任务。

[0071] 在一些实施例中，知识本体 360 由可执行意图节点和属性节点组成。在知识本体 360 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 3C 中所示，知识本体 360 可以包括“餐厅预定”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）。此外，属性节点“菜系”、“价格范围”、“电话号码”和“位置”是属性节点“餐厅”的子节点，并均通过中间属性节点“餐厅”连接到“餐厅预定”节点（即，可执行意图节点）。作为另一个实例，如图 3C 中所示，知识本体 360 还可以包括“设置提醒”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“设置提醒”节点。由于属性“日期 / 时间”同时与进行餐厅预订的任务和设定提醒的任务相关，因此属性节点“日期 / 时间”同时连接至知识本体 360 中的“餐厅预订”节点和“设置提醒”节点。

[0072] 可以将可执行意图节点，连同其连接的概念节点描述为“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（以及其间的关系）。例如，图 3C 中示出的知识本体 360 包括在知识本体 360 内的餐厅预订域 362 的实例和提醒域 364 的实例。餐厅预定域包括可执行意图节点“餐厅预定”、属性节点“餐厅”、“日期 / 时间”和“同行人数”以及子属性节点“菜系”、“价格范围”、“电话号码”和“位置”。提醒域 364 包括可执行意图节点“设置提醒”和属性节点“主体”和“日期 / 时间”。在一些实施例中，知识本体 360 由多个域组成。每个域可以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 362 和提醒域 364 之外，“日期 / 时间”属性节点还可与许多不同的域（例如，行程安排域、旅游预订域、电影票域等）相关联。

[0073] 尽管图 3C 示出知识本体 360 内的两个实例域，其他域（或可执行意图）包括例如“发起电话呼叫”、“查找方向”、“安排会议”、“发送消息”、和“提供对问题的回答”、“阅读列表”、“提供导航指令”、“提供任务指示”等。“发送消息”域与“发送消息”可执行意图节点相关联，并且还可以包括诸如“一个或多个接收人”、“消息类型”和“消息主体”的属性节点。例如，可以由诸如“接收人姓名”和“消息地址”的子属性节点进一步定义属性节点“接收人”。

[0074] 在一些实施例中，知识本体 360 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）。在一些实施例中，知识本体 360 可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 360 内的节点之间的关系来修改。

[0075] 在一些实施例中，可将与多个相关可执行意图相关联的节点群集在知识本体 360 中的“超级域”下。例如，“旅行”超级域可包括与旅行有关的属性节点和可执行意图节点的群集。与旅行有关的可执行意图节点可包括“机票预订”、“酒店预订”、“汽车租赁”、“路线规划”、“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可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”、“寻找兴趣点”的可执行意图节点可共享属性节点“起始位置”、“目的地”、“出发日期 / 时间”、“到达日期 / 时间”及“同行人数”中的一个或多个。

[0076] 在一些实施例中，知识本体 360 中的每个节点与跟由节点代表的属性或可执行意图有关的一组字词和 / 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”。可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 344 中。例如，返回图 3B，与“餐厅”的属性的节点相关联的词汇可包括字词诸如“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”的可执行意图的节点相关联的词汇可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 344 任选地包括不同语言的字词和短语。

[0077] 自然语言处理器 332 从语音转文本处理模块 330 接收符号序列（例如，文本串），并确定符号序列中的字词牵涉哪些节点。在一些实施例中，如果发现符号序列中的字词或短语与知识本体 360 中的一个或多个节点相关联（经由词汇索引 344），则所述字词或短语将“触发”或“激活”那些节点。基于已激活节点的数量和 / 或相对重要性，自然语言处理器 332 将选择可执行意图中的一个可执行意图作为用户意图让数字助理执行的任务。在一些实施例中，选择具有最多“已触发”节点的域。在一些实施例中，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。在一些实施例中，基于已触发节点的数量和重要性的组合来选择域。在一些实施例中，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。

[0078] 在一些实施例中，数字助理还将特定实体的名称存储在词汇索引 344 中，使得当在用户请求中检测到这些名称中的一个名称时，自然语言处理器 332 将能够识别该名称涉及知识主体中属性或子属性的特定实例。在一些实施例中，特定实体的名称是企业、餐厅、人、电影等的名称。在一些实施例中，数字助理从其他数据源中搜索并识别特定实体名称，

所述其他数据源诸如用户的通讯录、电影数据库、音乐家数据库和 / 或餐厅数据库。在一些实施例中,当自然语言处理器 332 识别出符号序列中的字词是特定实体的名称(诸如用户通讯录中的名称)时,在于用户请求的知识主体内选择可执行意图的过程中,为该字词赋予附加重要性。

[0079] 例如,在从用户请求识别出字词“Santo 先生”且在词汇索引 344 中发现姓“Santo”是用户联系人列表中的联系人之一时,则用户请求可能对应于“发送消息”或“发起电话呼叫”域。又如,当在用户请求中发现字词“ABC 咖啡馆”且在词汇索引 344 中发现术语“ABC 咖啡馆”是用户所在城市中的特定餐厅的名称时,则用户请求可能对应于“餐厅预订”域。

[0080] 用户数据 348 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。在一些实施例中,自然语言处理器 332 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”,自然语言处理器 332 能够访问用户数据 348 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。

[0081] 在 2008 年 12 月 22 日提交的针对“Method and Apparatus for Searching Using an Active Ontology”的美国实用新型申请序列号 12/341,743 中描述了基于令牌串搜索知识本体的其他细节,在此通过引用将其整个公开并入本文。

[0082] 在一些实施例中,一旦自然语言处理器 332 基于用户请求识别出可执行意图(或域),自然语言处理器 332 便生成结构化查询以表示所识别的可执行意图。在一些实施例中,结构化查询包括域内针对可执行意图的一个或多个节点的参数,并且所述参数中的至少一些填充有在用户请求中指定的特定信息和要求。例如,用户可以说:“通知寿司店预定晚上 7 点的座位”。在该情况下,自然语言处理器 332 能够基于用户输入将可执行意图正确地识别为“餐厅预订”。根据知识主体,“餐厅预订”域的结构化查询可包括参数诸如{菜系}、{时间}、{日期}、{同行人数}等。在一些实施例中,基于包含在用户话语中的信息,自然语言处理器 332 针对餐厅预订域生成部分结构化的查询,其中部分结构化的查询包括参数{菜系=“寿司类”} 和 {时间=“晚上 7 点”}。然而,在本实例中,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}。在一些实施例中,自然语言处理器 332 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,在一些实施例中,如果用户请求“附近的”寿司店,自然语言处理器 332 用来自用户设备 104 的 GPS 坐标来填充结构化查询中的{位置}参数。

[0083] 在一些实施例中,自然语言处理器 332 将结构化查询(包括任何已完成的参数)传送至任务流处理模块 336(“任务流处理器”)。任务流处理器 336 被配置为接收来自自然语言处理器 332 的结构化查询,完成结构化查询,以及如有必要执行“完成”用户最终请求所需的动作。在一些实施例中,完成这些任务所必需的各种过程在任务流模型 354 中提供。在一些实施例中,任务流模型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。

[0084] 如上所述,为了完成结构化查询,任务流处理器 336 可能需要发起与用户的附加对话,以便获得附加信息,和 / 或消除可能有歧义的话语的歧义。当此类交互有必要时,任

务流处理器 336 调用对话处理模块 334 (“对话处理器 334”) 以进行与用户的对话。在一些实施例中,对话处理器 334 确定如何 (和 / 或何时) 向用户询问附加信息,并接收和处理用户响应。通过 I/O 处理模块 328 将问题提供给用户并从用户接收回答。在一些实施例中,对话处理器 334 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。继续上文的实例,当任务流处理器 336 调用对话流处理器 334 以针对与域“餐厅预订”相关联的结构化查询来确定“同行人数”和“日期”信息时,对话流处理器 335 生成问题诸如“共有多少人用餐?”和“具体哪天用餐?”以传送至用户。一旦接收到来自用户的回答,对话流处理器 334 则可以用缺失信息填充结构化查询,或将信息传递给任务流处理器 336 以根据结构化查询完成缺失信息。

[0085] 在一些情况下,任务流处理器 336 可接收具有一个或多个模糊属性的结构化查询。例如,用于“发送消息”域的结构化查询可以指出,预期的接收人为“Bob”,并且用户可能有多位名为“Bob”的联系人。任务流处理器 336 将请求对话处理器 334 消除结构化查询的这一属性的歧义。继而,对话处理器 334 可询问用户“哪个 Bob ?”,并显示 (或读出) 名为“Bob”的联系人的列表,用户可从该列表中进行选择。

[0086] 一旦任务流处理器 336 已针对可执行意图完成结构化查询,任务流处理器 336 就继续执行与可执行意图相关联的最终任务。因此,任务流处理器 336 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”的任务流模型可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 = 2012/3/12, 时间 = 晚上 7 点, 同行人数 = 5 人 },任务流处理器 336 可执行以下步骤:(1) 登录 ABC 咖啡馆的服务器,或者餐厅预订系统诸如 OPENTABLE[®], (2) 在网站上的表格中输入日期、时间和同行人数信息,(3) 提交表格,以及(4) 在用户日历中针对该预订制作日历条目。

[0087] 在一些实施例中,任务流处理器 336 在服务处理模块 338 (“服务处理器”) 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,服务处理器 338 可代表任务流处理器 336 发起电话呼叫、设置日历条目、调用地图搜索、调用用户设备上安装的其他应用程序或与所述其他应用程序交互,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) 或与第三方服务交互。在一些实施例中,可通过服务模型 356 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)。服务处理器 338 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。

[0088] 例如,如果餐厅已启用在线预订服务,则餐厅可提交服务模型,该服务模型指定进行预订的必要参数以及将必要参数的值传送至在线预订服务的 API。在被任务流处理器 336 请求时,服务处理器 338 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接口。

[0089] 在一些实施例中,自然语言处理器 332、对话处理器 334 以及任务流处理器 336 共同且反复地使用以推断并限定用户的意图、获得信息以进一步明确并提炼用户意图、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。

[0090] 在一些实施例中,在已执行满足用户请求所需的所有任务之后,数字助理 326 制

定确认响应，并通过 I/O 处理模块 328 将该响应发送回用户。如果用户请求寻求信息性回答，则确认响应向用户呈现所请求的信息。在一些实施例中，数字助理还请求用户指示用户是否对数字助理 326 产生的响应满意。

[0091] 可以在 2010 年 1 月 18 日提交的名称为“Intelligent Automated Assistant”的美国实用新型申请号 12/987,982、2011 年 6 月 3 日提交的名称为“Generating and Processing Data Items That Represent Tasks to Perform”的美国实用新型申请号 61/493,201 中获得关于数字助理的更多详情，在此通过引用将其完整公开并入本文。

[0092] 在大部分情形中，在数字助理从用户接收用户输入时，数字助理尝试以尽可能小的延迟提供对用户输入的适当响应。例如，假设用户通过提供语音输入（例如，“现在交通状况如何？”）来请求特定信息（例如，当前交通信息）。数字助理刚刚接收并处理语音输入之后，数字助理任选地提供语音输出（例如，“查看交通信息……”），确认接收到用户请求。在数字助理响应于用户请求获得所请求信息之后，数字助理继续向用户提供所请求信息而没有更多延迟。例如，响应于用户的交通信息请求，数字助理可以在产生语音输出之后立刻提供一系列一个或多个分立的由简短暂停分开的语音输出，（例如，“路上有两起事故。<暂停>一起事故在惠普尔大街附近 101 北向路段。<暂停>并且第二起事故在 280 附近 85 北向路段。”）。

[0093] 出于本说明书的目的，初始确认用户请求以及响应于用户请求提供的一系列一个或多个分立语音输出都被视为对用户请求的完整响应的子响应。换言之，数字助理在接收到用户请求时为用户请求发起信息提供过程，并且在信息提供过程期间，数字助理准备并提供对用户请求的完整响应的每个子响应，而无需来自用户的进一步提示。

[0094] 有时，在能够获得请求的信息之前，需要额外的信息或说明（例如，路线信息）。在这样的情形中，数字助理向用户输出问题（例如，“您要去哪里？”），询问额外的信息或说明。在一些实施例中，数字助理提供的问题被视为对用户请求的完整响应，由于在从用户接收新输入之前，数字助理将不会采取进一步动作或提供对用户请求的任何额外响应。在一些实施例中，一旦用户提供了额外信息或说明，数字助理就为基于初始用户请求和额外的用户输入建立的“新”用户请求发起新的信息提供过程。

[0095] 在一些实施例中，数字助理在接收到每个新用户输入时发起新的信息提供过程，并且在（1）已经向用户提供了对用户请求的完整响应的所有子响应时，或（2）在数字助理向用户提供了关于开始现有信息提供过程的先前用户请求的额外信息或说明的请求时，每个现有的信息提供过程终止。

[0096] 通常，在数字助理接收到用户对信息或执行任务的请求之后，希望数字助理尽快提供响应（例如，包含所请求信息的输出、所请求任务的确认或请求说明的输出）。数字助理的实时响应性是评估数字助理性能的关键因素之一。在这种情况下，尽快准备响应，并且响应的默认提供时间是紧跟准备好响应之后的时间。

[0097] 不过，有时在接收到用户输入之后立刻提供初始子响应之后，数字助理在延长的时间段内的某一时间提供剩余的一个或多个子响应。在一些实施例中，在比逐个提供每个子响应所需时间之和更长的延长时间段内扩展针对用户请求的信息提供过程。例如，在一些实施例中，在通过音频输出通道向用户提供子响应时，在相邻对的子响应（例如，一对连贯的语音输出）之间插入短暂停（即，简短的静默时期）。

[0098] 在一些实施例中，在准备好子响应之后搁置它，仅在满足预定条件时才提供。在一些实施例中，在根据系统时钟已到达预先确定的触发时间和 / 或已发生预先确定的触发事件时，满足预先确定的条件。例如，如果用户向数字助理说“为我设置 5 分钟的计时器”，数字助理在接收到用户请求时发起信息提供过程。在信息提供过程期间，数字助理马上提供第一子响应（例如，“好的，计时器启动。”），并直到 5 分钟之后才提供第二和最后子响应（例如，“好，五分钟到了”）。在这种情况下，第一子响应的默认提供时间是准备好第一子响应之后紧跟的时间，并且第二最后子响应的默认提供时间是紧跟发生触发事件的时间（例如，从启动计时器之后过去 5 分钟）。在数字助理完成向用户提供最后子响应时，信息提供过程终止。在各实施例中，在第二子响应的默认提供时间之前的任何时间（例如，刚好在准备第一子响应之后，或者直到第二子响应的默认提供时间之前不久）准备第二子响应。

[0099] 如本说明书中稍后将要更详细描述的，在用于提供对用户请求的响应和 / 或用于提供提醒和通知的警报项的默认规则之上实现上下文相关中断处理程序（例如，图 3A 中的中断处理模块 340）。在一些实施例中，中断处理程序实时收集关于当前上下文的信息，并实时确定由于设备当前正在接收用户的语音输入，是否需要改变用于提供响应、提醒和 / 或通知的默认规则。例如，在一些上下文中，会更适于延迟（例如，中止，至少暂时）提供非紧急语音输出，由于用户正在向设备中讲话，而在其他时间，可能更适于立即提供紧急的语音输出（例如，“插话”或打断用户）。此外，在一些上下文中，可以接受并实际上更适于完全放弃提供语音输出。

[0100] 图 4A-4C 示出了根据一些实施例的示例性情形，其中，数字助理提供语音输出或不提供语音输出。在图 4A-4C 中，对应于语音输出的实线框（例如，图 4A 中的 S01）表示由设备实际提供的语音输出。另一方面，虚线框（例如，图 4A 中的 S02）表示在对应时间和位置未由设备实际提供，而是如果未检测到用户的语音输入而由设备提供的语音输出，如参考各个图更详细解释的那样。

[0101] 图 4A 示出了设备永久放弃语音输出的示例性情形。一开始，用户在东阿尔德大街上向东（401）行进。在由 402-1 表示的第一位置，用户通过指出“带我去图书馆”作为语音输入 SI1，请求设备提供去图书馆的逐向指示。用户结束讲话的位置由 402-2 指示，由于用户在运动的同时讲话，所以这与 402-1 不同。于是，对应于相应语音输入和输出的条（例如，对应于 SI1 的 402-1 和 402-2 之间的条）表示距离，或者等价地，相应输入 / 输出叙述（例如，讲话）所需的时间长度。

[0102] 设备接收语音输入并执行必要的操作，例如，以确定最近的图书馆的位置，如参考方法 500 和图 5A-5D 更详细所述。在本实例中，该设备确定图书馆 404 是最近的图书馆，并利用语音输出 S01（“好的，继续直行”）迅速响应。

[0103] 在 405-1 表示的位置处，与数字助理位于相同设备上的电话特征结构接收来电，如铃声图标 406 所示。用户通过提供语音输入 SI2 来接听电话，讲道“嘿，John！好久没听到你的消息了。家人怎么样？”在完成语音输入 SI2 时，用户处在由 405-2 指示的位置。不过，在 405-1 和 405-2 之间的过渡期，设备（例如从服务器或同一设备上的不同模块）接收语音输出 S02，指出“在 3 英里后的第一大街右转”。在本实例中，语音输出 S02 具有低紧急度量，由于设备计划在用户到达第一大街之前一次或多次警告用户即将转弯（例如，额外的警告诸如“1 英里后右转”和 / 或“现在右转到第一大街上”）。由于在要输出语音输出

S02 时 (407-1 直到 407-2) 设备正在接收语音输入 SI2, 所以设备中止输出语音输出 S02。此外, 由于与语音输出 S02 相关联的优先级非常低 (例如, 由于本实例中与逐向指示相关联的冗余度), 这种中止实际上完全放弃了输出语音输出 S02 (例如, 一直不输出三英里后左转的命令, 而是依赖于 1 英里的即刻警告)。

[0104] 图 4B 示出了根据一些实施例即刻向用户提供语音输出的示例性情形。图 4A 和 4B 中共用的相似附图标记是指相应情形的类似方面。于是, 为了简洁起见, 这里不会重复那些相似的方面。

[0105] 图 4B 与图 4A 不同之处在于, 电话在比图 4A 中距第一大街近得多处接听来电, 如铃声图标 406 所示。在用户通过语音输入 SI2 接听电话的同时, 设备接收对应于逐向指示命令的语音输出 S03, 指出用户很快就应当右转 (例如, 在本实例中, 100 英尺之后)。由于消息紧急, 设备会“插话” (例如, 在用户讲话的同时打断用户) 以输出“100 英尺后右转”。

[0106] 图 4C 示出了根据一些实施例暂时中止语音输出, 稍后向用户提供语音输出的示例性情形。图 4A 和 4C 中共用的相似附图标记是指相应情形的类似方面。于是, 为了简洁起见, 这里不会重复那些相似的方面。

[0107] 在语音输入 SI3 期间, 用户请求无论何时比赛应当结束时, 设备就通知用户 Knicks 的比分, 指出“比赛结束时告诉我 Knicks 的比分”。设备立即做出响应, 确认请求, 指出“好的, 在比赛结束时我将告诉您 Knicks 的比赛分数”。如前所述, 电话接收来电, 用户在语音输入 SI2 中回答, 说道, “嘿, John ! 好久没听到你的消息了。家人怎么样?”在 SI2 期间, Knicks 的比赛结束, 设备接收指明比分的语音输出 S05, 按照请求, 向用户提供。在本实例中, 认为语音输出 S05 不紧急, 由于在用户讲话的时间内 (例如, 在设备正在接收语音输入期间), Knicks 的比分不会变化。因此, 设备中止语音输出 S05, 如箭头 408 所示, 直到用户完成讲话, 然后输出语音输出 S05。不过, 在一些实施例中, 设备以不可闻的方式对用户请求做出响应, 诸如在设备显示器上显示 Knicks 的比分。在一些实施例中, 由于显示的响应不会打断用户讲话, 所以无延迟地提供这样的响应。在一些实施例中, 结合或者作为另外一种选择替代中止的语音输出来提供这样的显示的响应 (例如, 在显示的响应替代语音输出时, 完全放弃语音输出)。

[0108] 图 5A-5D 为由数字助理实施的示例性方法 500 的流程图, 该方法基于确定或装置当前是否正从用户接收语音输入以及语音输出的紧急性来确定是否向用户提供语音输出。在一些实施例中, 由数字助理的中断处理程序 (例如, 图 3A 中的中断处理程序 340) 基于当前上下文实时动态地执行是否提供语音输出的确定。

[0109] 在一些实施例中, 在接收语音输出 (参考 506) 之前, 设备从用户接收执行数字助理任务的请求 (502)。例如, 用户通过指明, 例如“帮我找到吃晚餐的地方, 不要太贵。”作为语音输入, 请求数字助理找到附近便宜的餐厅。或者, 用户通过例如指明“为我预定四点钟 Boulevard 的座位,”作为语音输入, 请求数字助理在特定餐厅进行预定, 或者, 用户询问到达本地地标的逐向指示 (“到金门大桥的方向”) 或询问棒球比分 (“Sox 打得怎么样?”) 或股价 (“Apple 的股票今天怎么样?”)。

[0110] 在一些实施例中, 在接收语音输出 (参考 506) 之前, 设备向数字助理服务器发送请求 (504)。响应于该请求从服务器接收语音输出。在一些实施例中, 在向服务器发送请求之前, 设备执行语音转文本操作 (例如, 利用 STT 处理模块 330)。在一些实施例中, 在服务

器处执行语音转文本。在一些实施例中，设备执行自然语言处理（例如，利用自然语言处理模块 322），包括利用用户数据进行知识本体、词汇分析和上下文匹配（例如，基于偏好诸如最爱、浏览历史和 / 或数字助理请求历史来弄清楚用户感兴趣的是那支“Sox”队）。服务器然后执行为请求服务所需的任何剩余操作（例如，从可执行属性识别一个或多个可执行项、一个或多个缺失属性，搜索一个或多个数据库和 / 或因特网以发现缺失信息等）。在一些实施例中，服务器准备响应（例如，文本串）并向用户返回响应。在一些实施例中，服务器准备语音响应（例如，音频数据）并向数字助理发送语音响应。

[0111] 在任何事件中，设备都接收要向设备的用户提供的语音输出（506）。在一些实施例中，响应于请求从服务器接收语音输出（例如，语音输出是对用户所做的请求的适当响应，假设其是请求晚餐预定或逐向指示）。在一些实施例中，接收语音输出包括在设备处生成语音输出（508）（例如，服务器响应于请求返回文本串，并且设备利用文本转语音引擎从文本生成语音输出）。应当理解，在一些实施例中，接收语音输出表示从服务器接收（任选地包括额外的处理操作，例如文本转语音操作）。替代地或此外，接收语音输出表示在第一设备部件（例如，模块诸如中断处理模块 340 或执行模块诸如中断处理模块 340 中保存的指令的处理器 304）处从第二设备部件（例如，模块诸如自然语言处理模块 332 或执行模块诸如自然语言处理模块 332 中保存的指令的处理器 304）接收。

[0112] 设备确定设备当前是否正从用户接收语音输入（510）。例如，在一些实施例中，设备是电话，并且确定设备当前是否正从用户接收语音输入包括确定用户是否正在参与与远程用户的电话交谈（512）。在这样的实施例中，如果用户当前正在交谈中讲话，设备确定其当前正从用户接收语音输入。在一些实施例中，在电话交谈另一端的一方正在讲话时，或者如果用户和另一方正在进行其他事项而有静默期，设备确定其当前未在接收语音输入（例如，在一些具体实施中，活动电话交谈足以确定设备正在接收语音输入，而在替代具体实施中，在用户实际上是交谈中讲话那个人时，设备确定正在接收语音输入）。

[0113] 在一些实施例中，确定设备当前是否正从用户接收语音输入包括确定上一次语音输入是否是在预先确定的时间段内接收的（514）。例如，由于在交谈消长中会有自然的暂停（例如，暂停以歇口气，暂停以考虑接下来讲什么），在一些实施例中，设备在得出用户未讲话的结论之前等候预先确定的时间量，而不是以即时或接近即时的方式检测语音输入。在一些实施例中，预先确定的时间段是输出的紧急度度量的函数（516）。例如，在设备在输出队列中具有紧急消息（“现在右转！”）时，设备将在确定用户未讲话之前等待更短时间量，从而在用户暂停歇口气或考虑接下来讲什么的时刻插话。在一些实施例中，确定设备当前是否正接收语音输入包括缄默确定（例如，基于设备麦克风处特定强度或方向性阈值），以例如消除背景噪声和 / 或用户讲出但并非意在作为语音输入的歧义（例如，在电话交谈期间，在用户暂停交谈而私下向另一方讲话时）。

[0114] 在确定设备当前未从用户接收语音输入时，设备向用户提供语音输出（518）。在一些实施例中，设备同时提供从远程用户接收的音频数据（参考 512，在用户正在通过电话参与时）以及至用户的语音输出，而不会由于所接收的音频数据而中止提供语音输出（520）。例如，在这样的实施例中，在远程用户（即，另一方）在电话交谈期间正在讲话时，设备将仍然提供来自数字助理的语音输出。在一些实施例中，同时提供从远程用户接收的音频数据（例如，语音）以及至用户的语音输出意味着在提供语音输出的同时暂时使来自远程用户

的音频数据静音。例如,在这样的实施例中,在远程用户讲道“八十七年以前,我们的祖先在这大陆上建立了一个国家,它孕育于自由,并且献身给一种理念,即所有人都是声来平等的。”时,输出队列中的语音输出是“左转”,实际提供给用户的音频将是“八十七年以前,我们的祖先在这大陆上建立了……“左转”……它孕育于自由,并且献身给一种理念,即所有人都是声来平等的。”用户从而将意识到,远程用户正在背诵林肯在 Gettysburg 的演讲,并且还将理解左转的指令。在一些实施例中,利用不同的腔调和 / 或音量提供从远程用户接收的音频数据以及语音输出,以将远程用户与数字助理区分开。

[0115] 在某些情况下,用户将已经配置设备超驰语音输出的提供。例如,在设备处于请勿打扰工作模式时,放弃提供语音输出 (522)。在一些实施例中,在用户已经配置设备处于请勿打扰工作模式时,设备处于请勿打扰工作模式。在一些实施例中,在用户已经配置设备工作于与请勿打扰不同的模式中时,设备处于请勿打扰工作模式,尽管如此,仍包括请勿打扰作为特征(例如,设备处于飞行模式,或静音模式,或者用户已经配置设备在一天中特定几个小时期间处于静音模式,等等。)

[0116] 在一些实施例中,在提供语音输出期间 (524),设备接收来自用户的语音输入 (526)。例如,在用户作为电话交谈的一部分讲话而中断或讲出针对数字助理操作的另一请求时,设备处于提供语音输出的中间。作为后一种情形的实例,在用户先前已请求过设备定位附近的中餐馆时,用户可以打断响应,指出他或她还需要向同事发送 SMS 消息。在这样的实施例中,设备将中断语音输出 (528)。在这样的实施例中,设备将确定是否已经满足对应于语音输出的完成标准 (530)。在一些实施例中,在已经向用户提供预定义百分比的语音输出时,满足完成标准 (532)。例如,在一些实施例中,预定义百分比的语音输出为选自 50%、60%、70% 和 80% 的百分比 (534)。在确定未满足完成标准时,设备在稍晚时间中止语音输出的至少一部分,并且在确定已满足完成标准时,设备完全放弃输出语音输出的其余部分 (536)。在一些实施例中,在设备确定消息的其余部分是讨论时满足完成标准(例如,在请求中餐之后,在设备叙述本地中餐馆的列表期间,用户声明,“没关系,我希望吃泰国菜。”)。

[0117] 在确定设备正从用户接收语音输入时,设备确定提供语音输出是否紧急 (538)。在一些实施例中,在语音输出满足用于即刻提供的用户可配置标准时,语音输出是紧急的 (540)。例如,在一些实施例中,在设备从先前已由用户识别为重要人物 (VIP) 的人接收电子消息时,满足用户可配置标准 (542)。替代地或此外,在一些实施例中,在设备接收股价更新且用户先前已经配置设备即刻提供股价更新时,满足用户可配置标准 (544)(例如,用户已经配置设备在特定股价超过特定值时警告他或她,使得用户能够尽快考虑卖出股票)。在一些实施例中,基于上下文确定提供语音输入是否紧急。例如,在语音输出包括不久就转弯的指示时(“现在左转!”),设备识别出消息是紧急的。在确定提供语音输出是紧急的时候,设备向用户提供语音输出 (546)(例如,设备“插话”,并尽管从用户接收语音输入仍然提供语音输出)。在一些实施例中,设备无延迟地向用户提供语音输出 (548)(例如,考虑到用户正在讲话,在产生输出所需的任何处理时间之上增加额外的延迟)。

[0118] 在确定提供语音输出不紧急时,设备中止向用户提供语音输出 (550)。如下文更详细所述,在某些情况下,中止提供语音输出意味着延迟提供语音输出直到稍晚时间,并然后提供语音输出,而在其他环境中,中止意味着完全放弃提供语音输出,并永不提供该特定语音输出。在某些情况下,中止意味着暂时延迟提供语音输出还是永久放弃提供语音输出取

决于特定实施例、具体实施和语音输出周围的上下文（参考 562）。在一些实施例中，在设备处于特殊工作模式时，设备无延迟地提供语音输出（552）（例如，即使设备当前正从用户接收语音输入）。例如，在一些实施例中，设备包括“打断我”工作模式，由此，不论设备是否在接收语音输入，用户都会被数字助理打断（例如，数字助理将插话）。在一些实施例中，特殊工作模式为选自保持工作模式和静音工作模式中的一者或者（553）。

[0119] 流程 553-1、553-2 和 553-3 代表额外的操作，根据方法 500 的一些实施例，在确定提供语音输出不紧急时，任选地执行额外操作。应当理解，结合流程 553 描述的各种操作未必是彼此排斥的，并且在某些情况下是组合的。

[0120] 例如，根据一些实施例，在确定设备不再从用户接收语音输入时，设备向用户提供语音输出（554）。在一些实施例中，确定设备不再从用户接收语音输入包括确定在上一次语音输入的时间和当前时间之间已经过去预定义的时间量（556）。在一些实施例中，预定义的时间量是语音输出的紧急度度量的函数。在一些实施例中，预先确定的时间量是语音输出的紧急度度量的单调递减函数，由此在更少时间量中以更大紧急度度量提供语音输出（560）。例如，在这些实施例中，在用户完成讲话之后，提供紧急语音输出之前，比语音输出较不紧急的情况相比，设备等待更短的时间量。

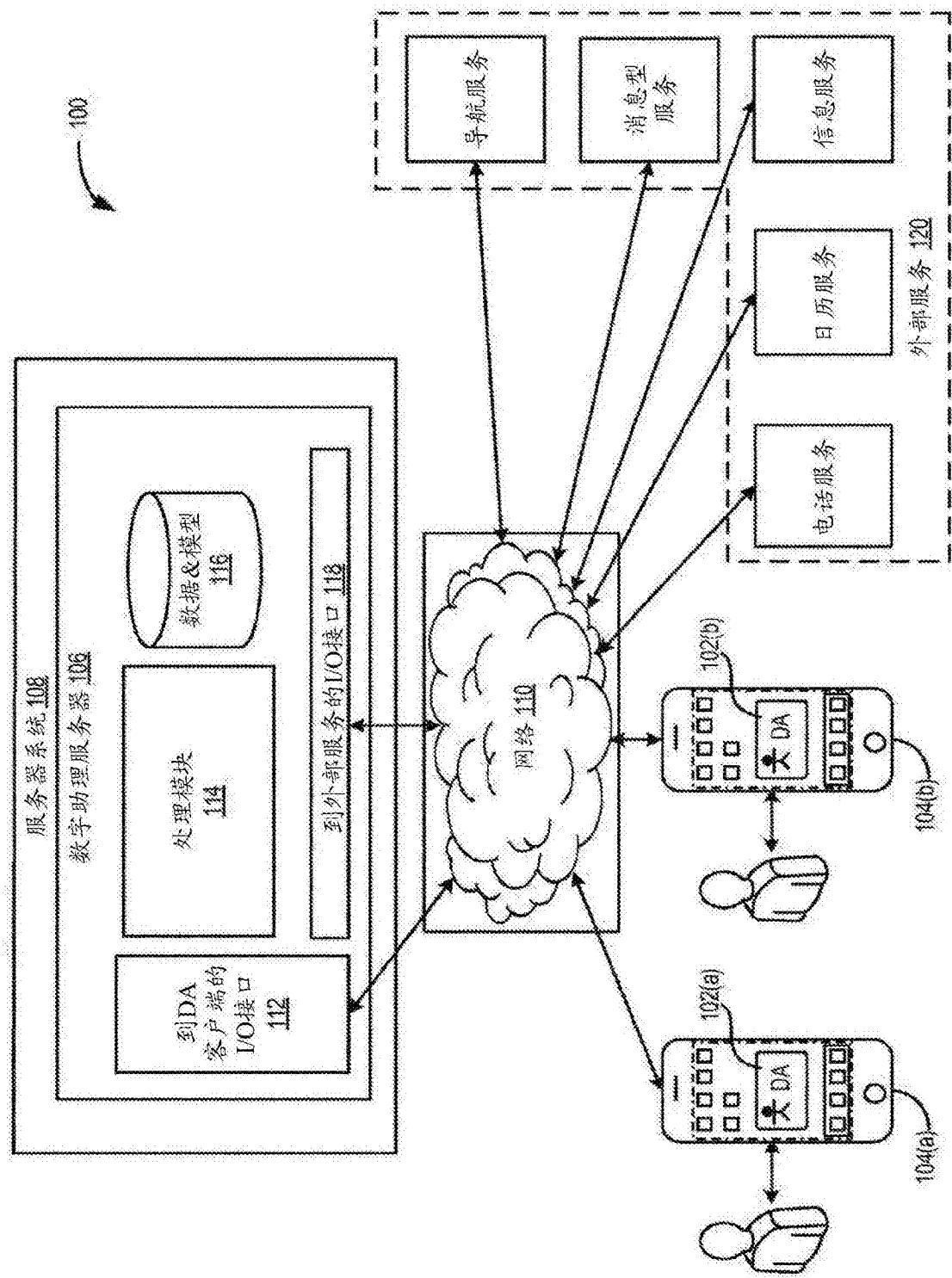
[0121] 在一些实施例中，设备确定输出是否满足消息跳过标准（562）。在一些实施例中，在紧急度度量低于预定义阈值时，满足消息跳过标准（564）。例如，在语音输出是警告序列中几个警告之一时，在某些情况下，不必为用户提供警告序列中的每个警告。在一些实施例中，在语音输出是一组逐向指示中的导航命令并且设备计划在稍晚时间给出对应的导航命令时，满足消息跳过标准（566）。例如，设备计划在转弯之前 2 英里、1 英里、1/2 英里和转弯时刻提供导航命令。在这样的情况下，在本来要叙述 1 英里命令时，用户正在提供语音输入时，设备完全放弃提供 1 英里命令。驾驶员将相应地仍然被 1/2 英里命令和转弯前一刻通知转弯。

[0122] 在一些实施例中，在设备包括显示器时，在确定提供语音输出不紧急时，设备提供对应于语音输出的显示输出（568）。在一些实施例中，

[0123] 以上参考图 5A-5D 所述的操作任选地由图 2 和 / 或图 3 中所描绘的部件来实现。例如，接收操作 504、提供操作 520、接收操作 526 任选地由上文详细所述的数字助理 326、I/O 处理模块 328、中断处理模块 340 和 / 或自然语言处理模块 332 实施。类似地，本领域技术人员会清楚地知道基于在图 2 和 / 或图 3 中所示的部件可如何实现其它过程。

[0124] 应当理解，其中上文已描述的操作的特定顺序仅仅是示例性的，并且并非旨在表明所述顺序是这些操作可被执行的唯一顺序。本领域的普通技术人员会想到多种方式来对本文所述的操作进行重新排序。

[0125] 为了解释的目的，前面的描述是通过参考具体实施例来进行描述的。然而，上面的示例性的讨论并非意图是详尽的，也并非意图要将本发明限制到所公开的精确形式。根据以上教导内容，很多修改形式和变型形式都是可能的。选择和描述实施例是为了充分阐明本发明的原理及其实际应用，以由此使得本领域的其他技术人员能够充分利用具有适合于所构想的特定用途的各种修改的本发明以及各种实施例。



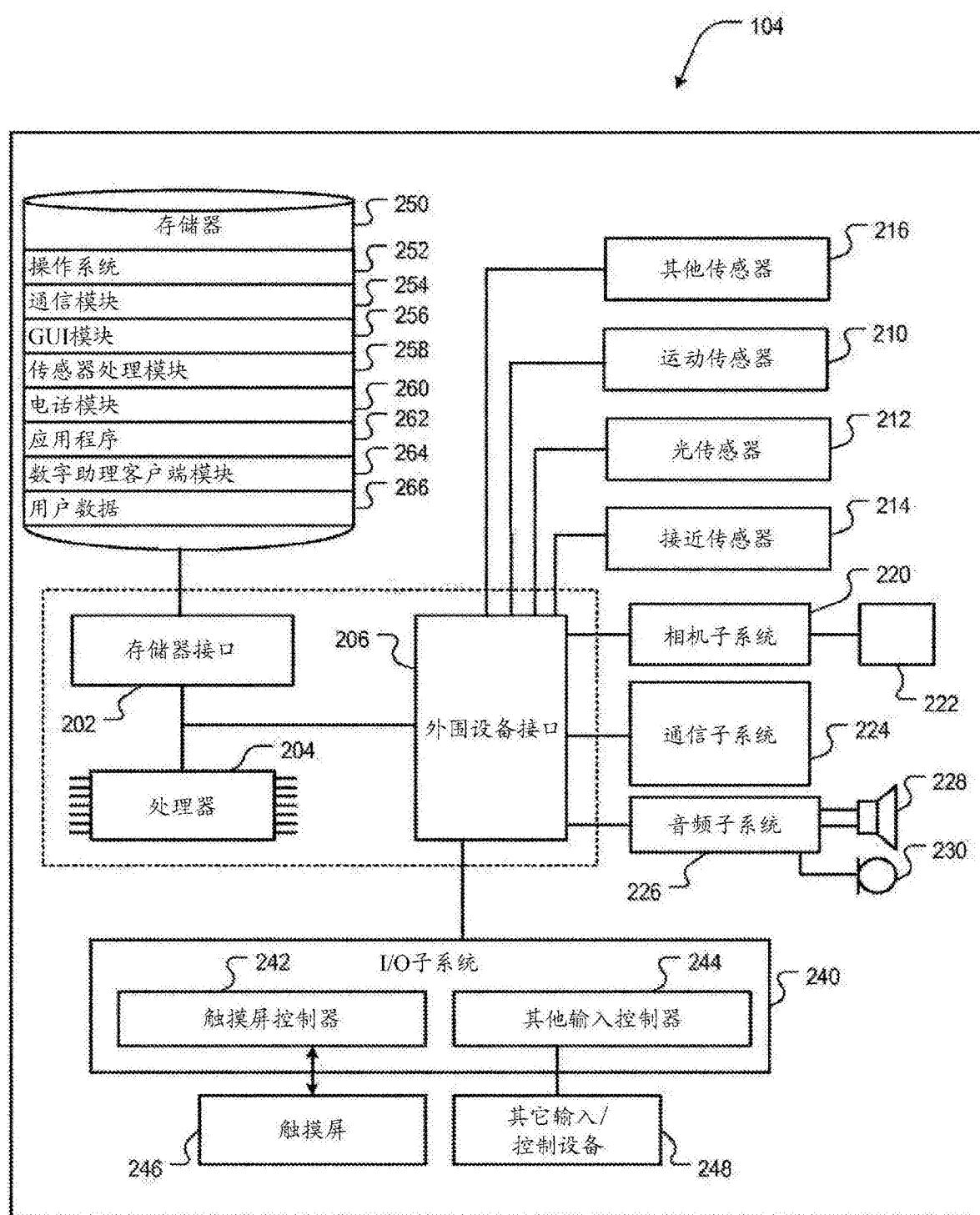


图 2

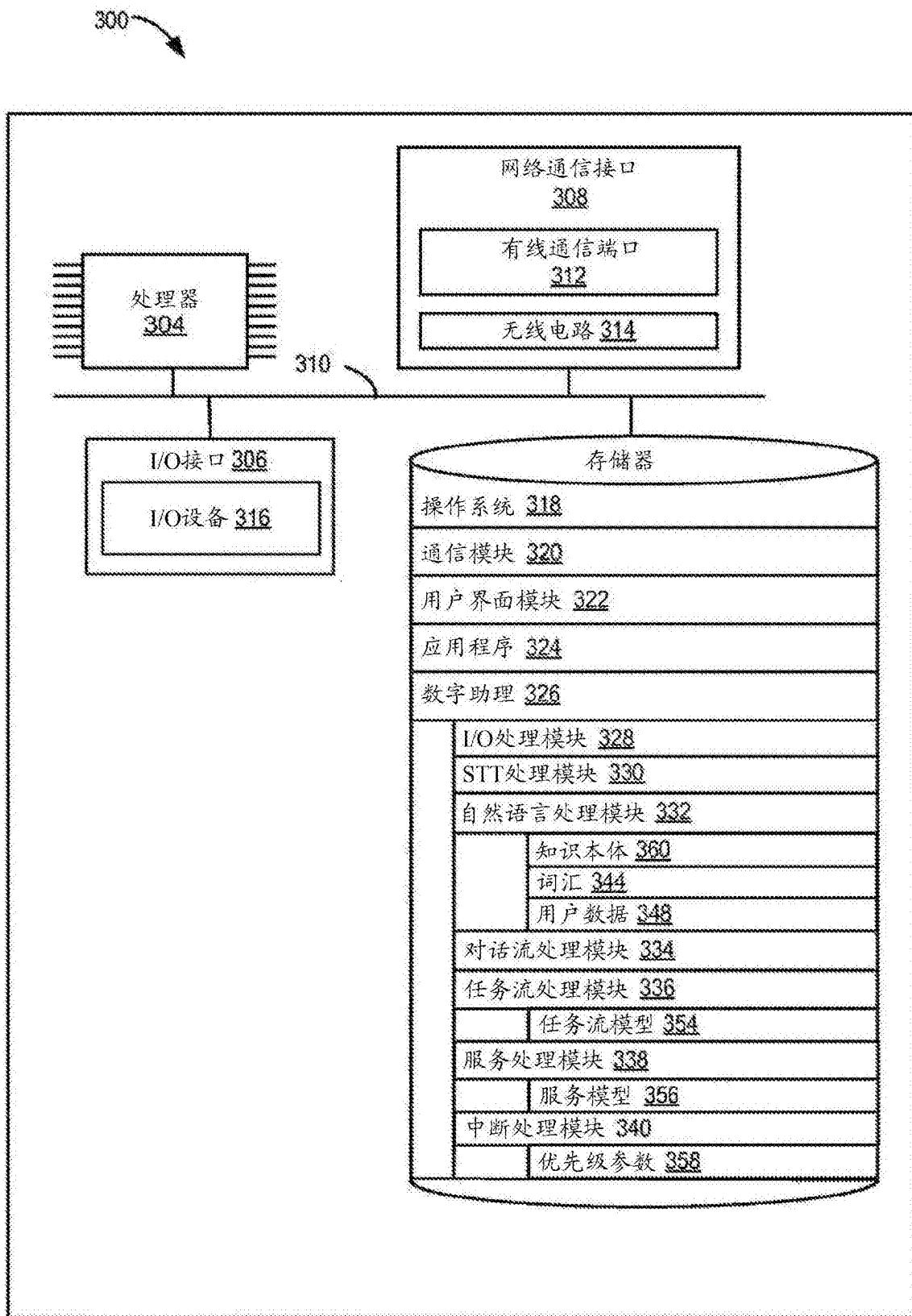


图 3A

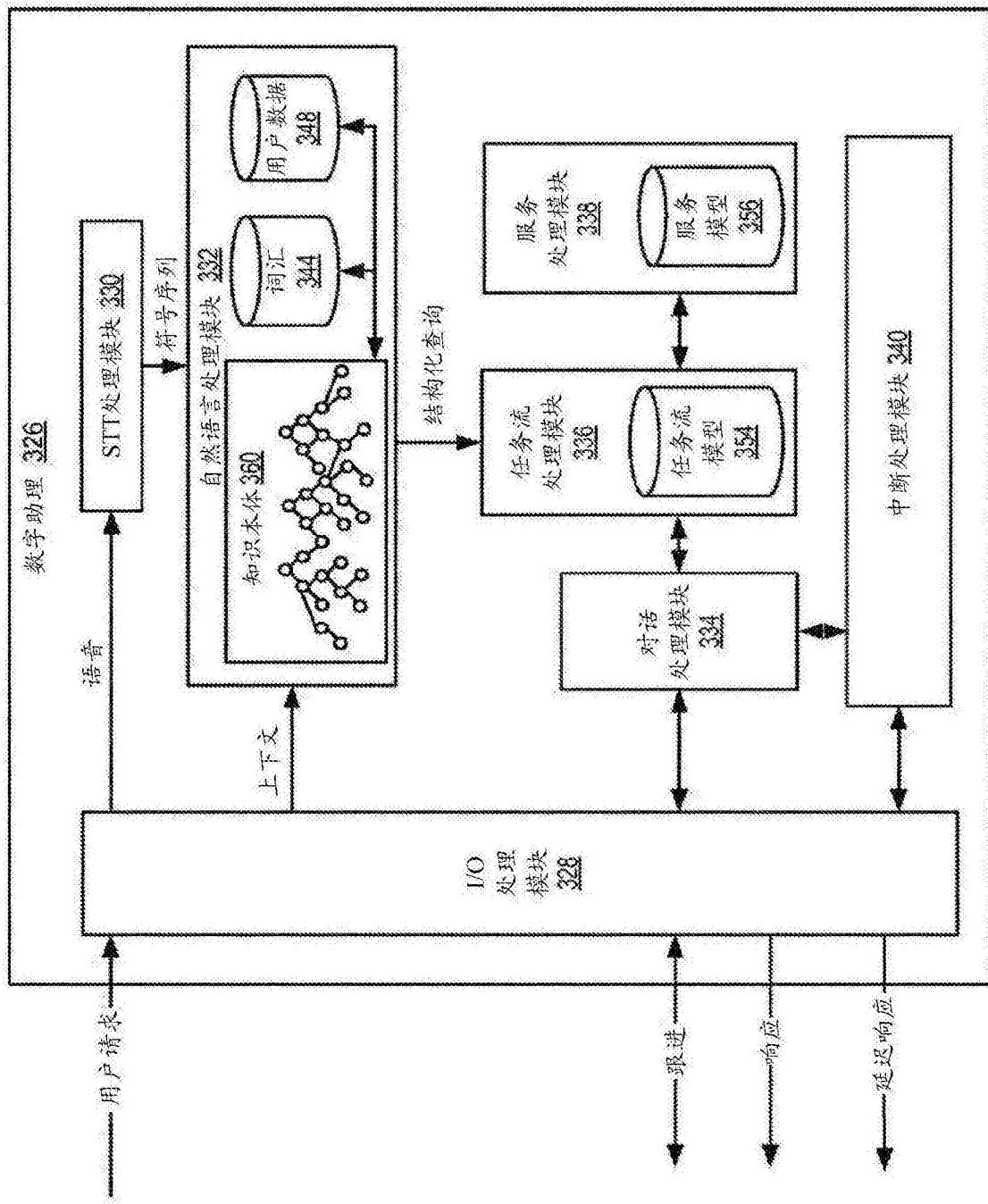


图 3B

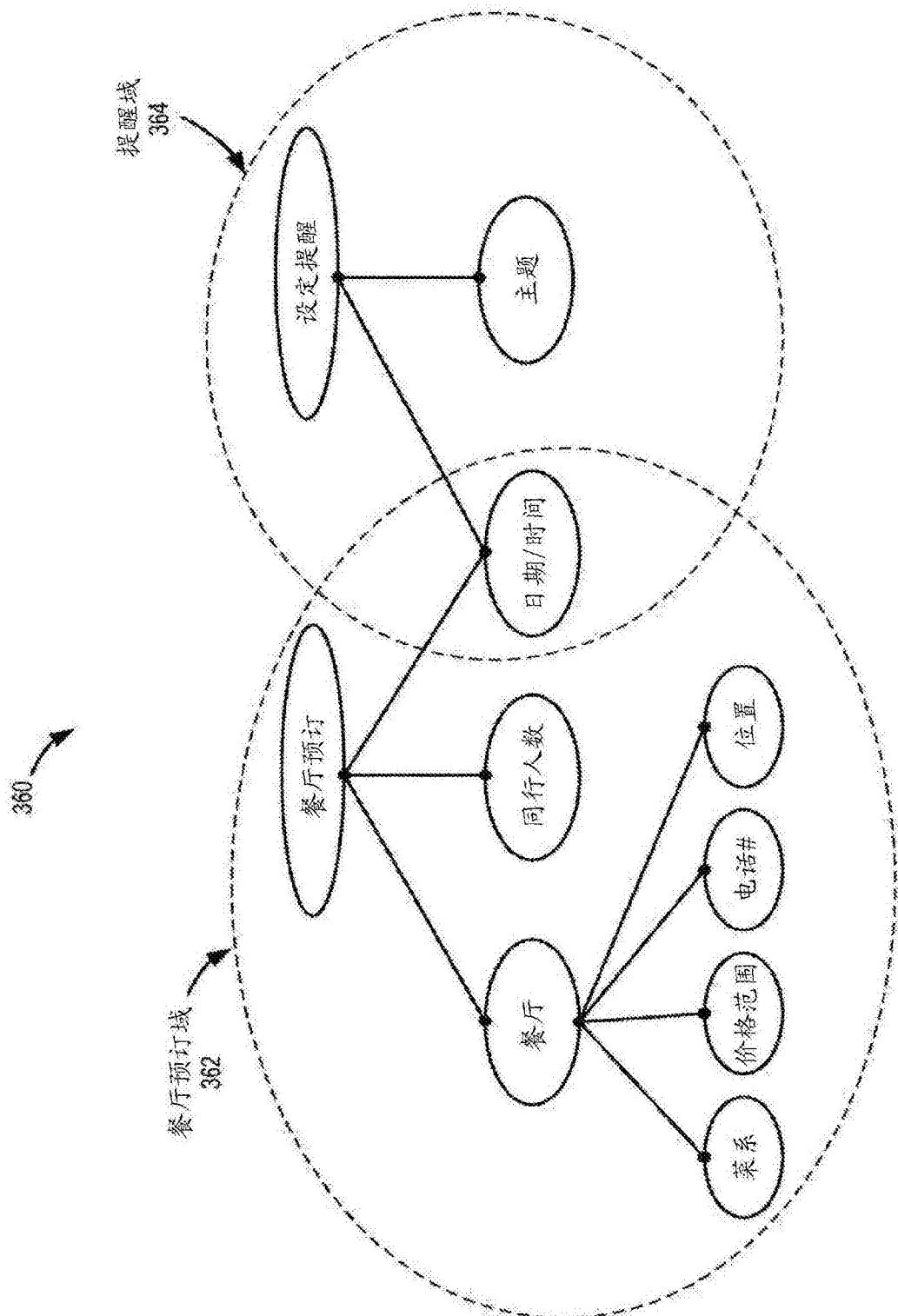


图 3C

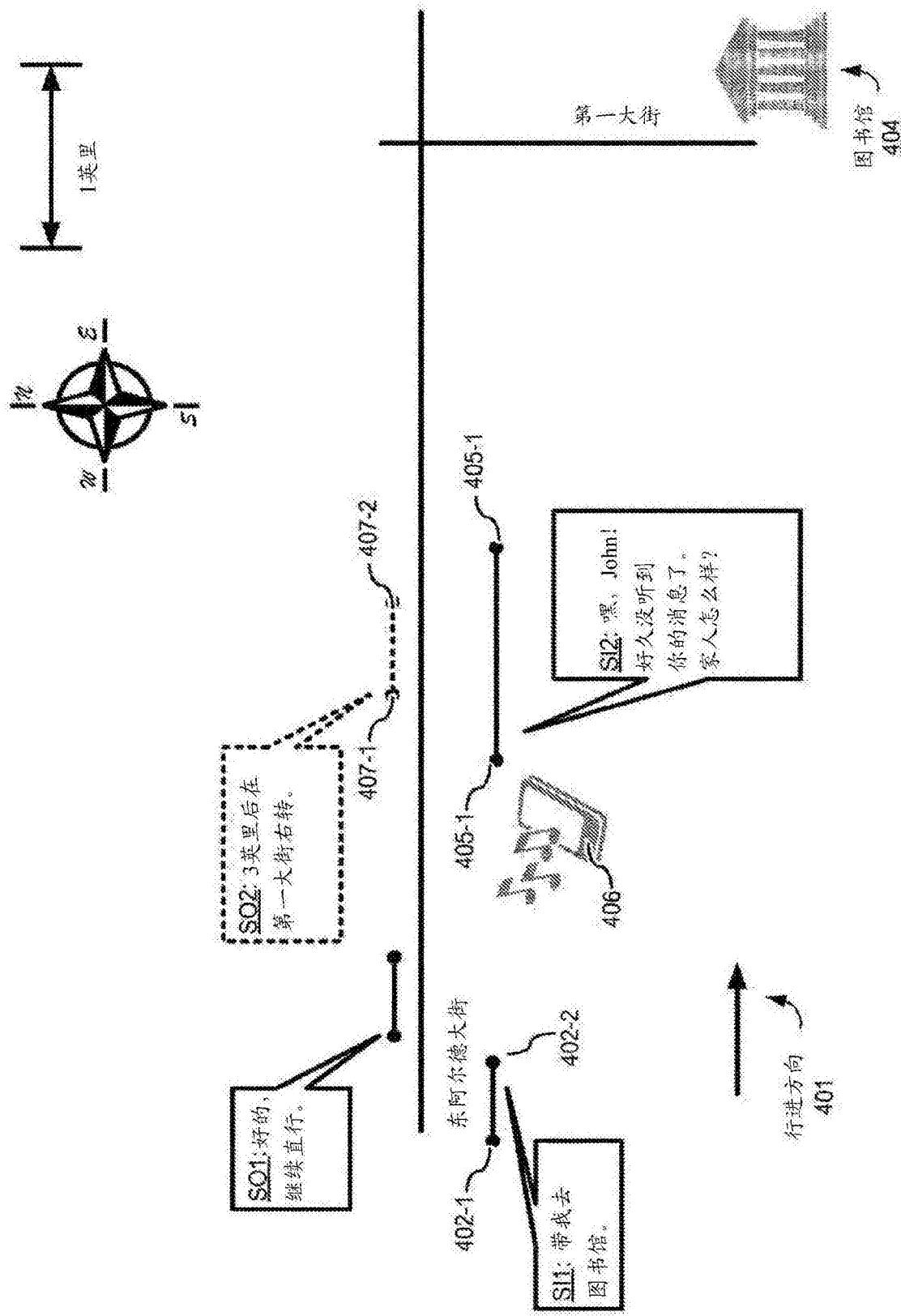


图 4A

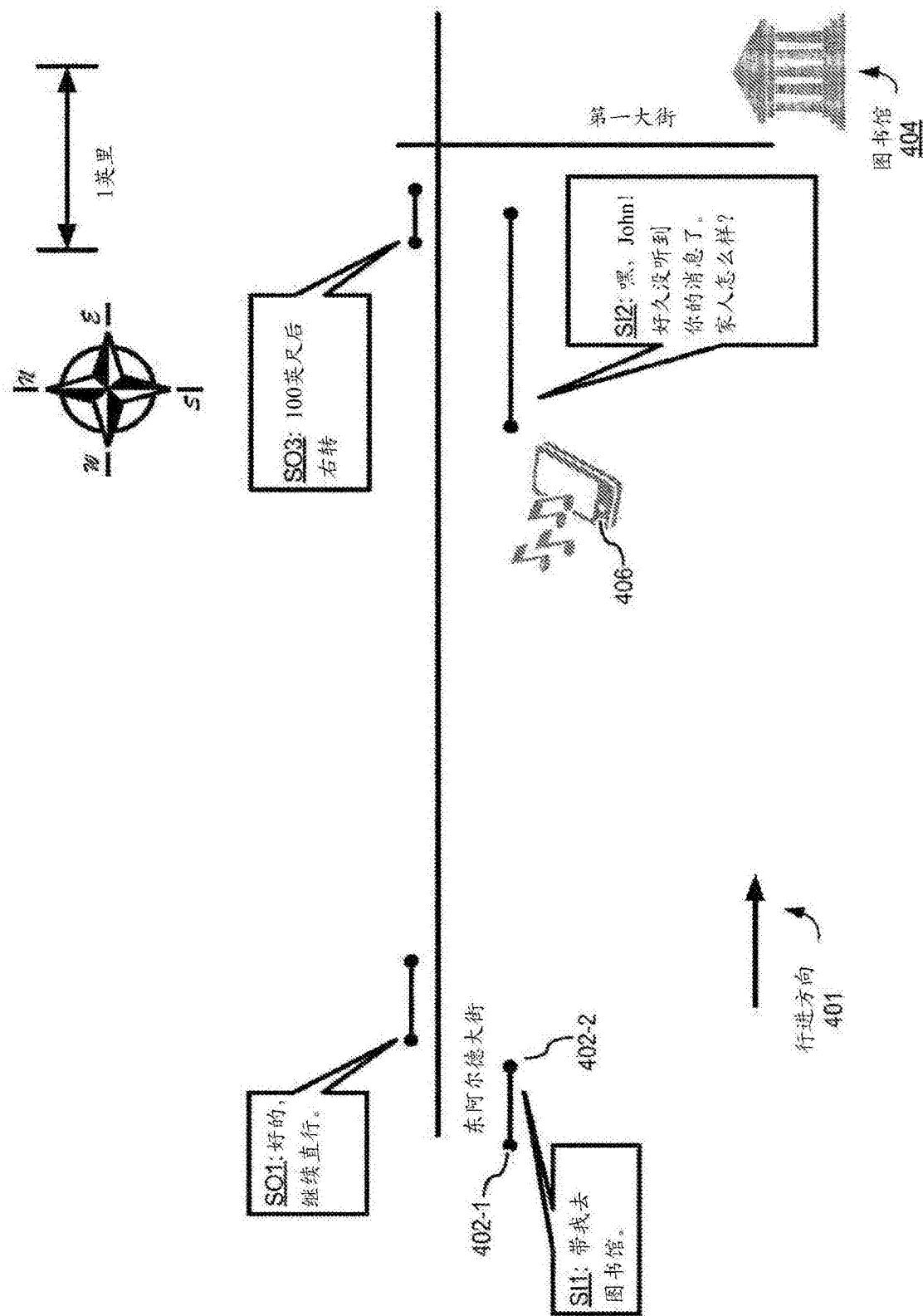


图 4B

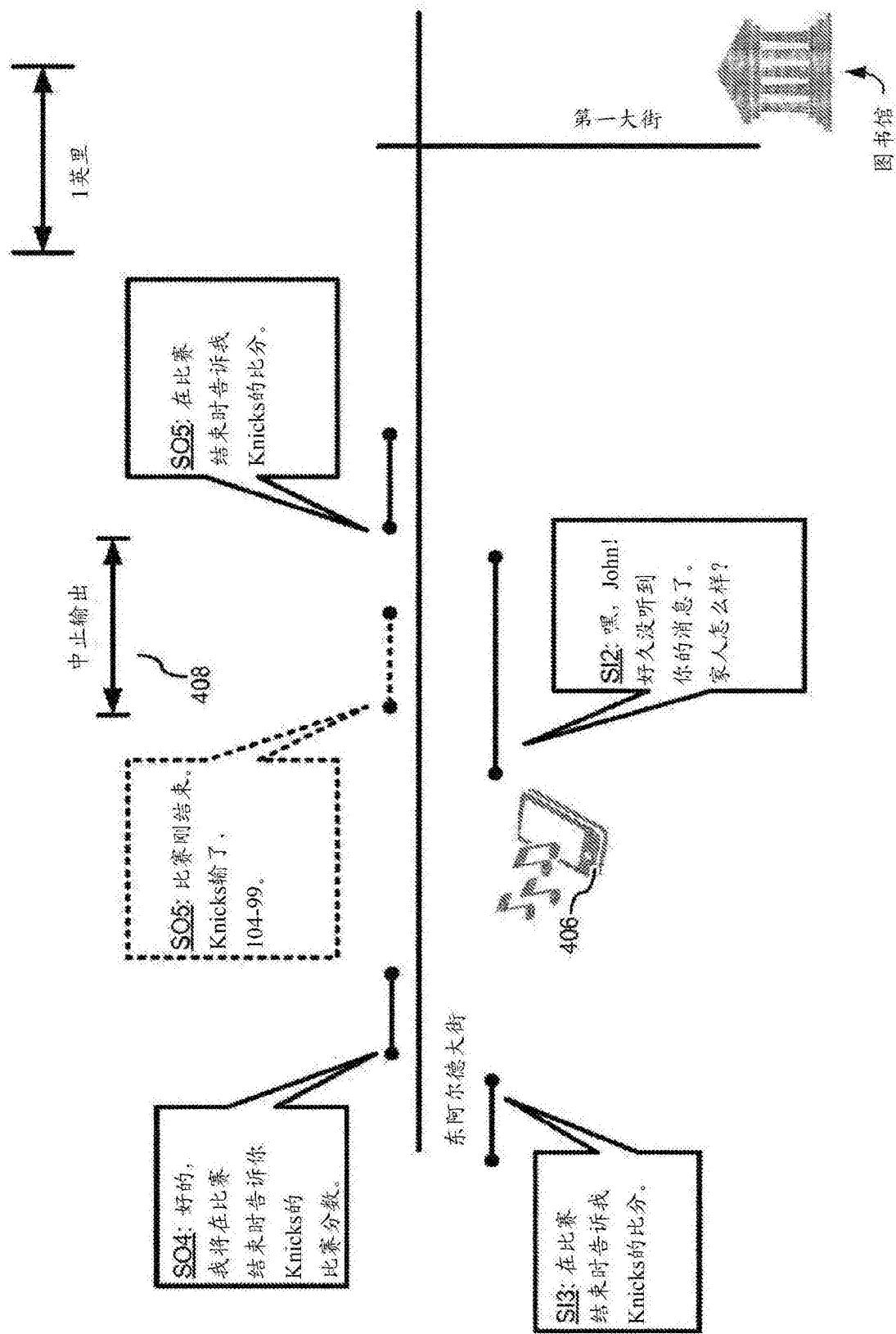


图 4C

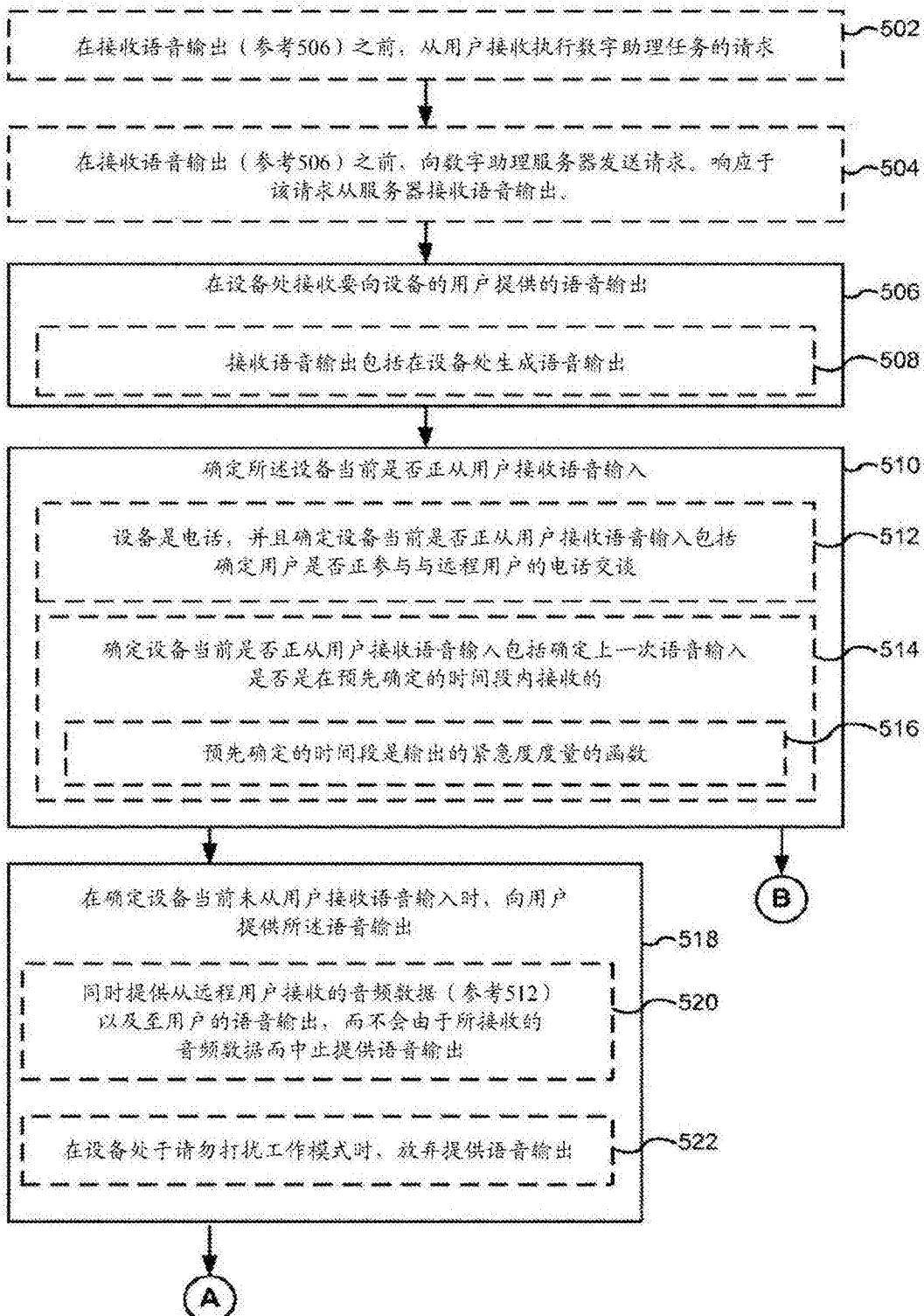
500

图 5A

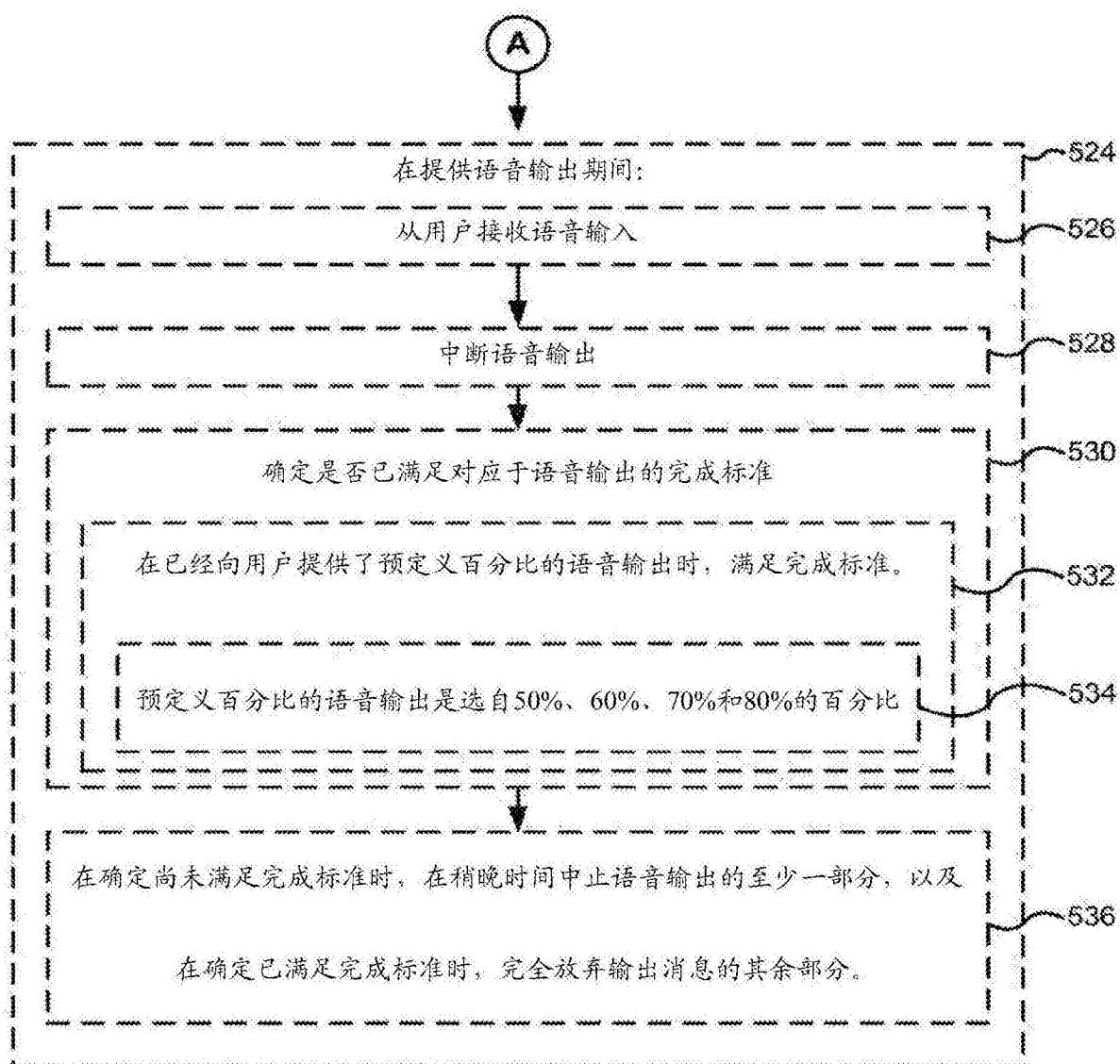


图 5B

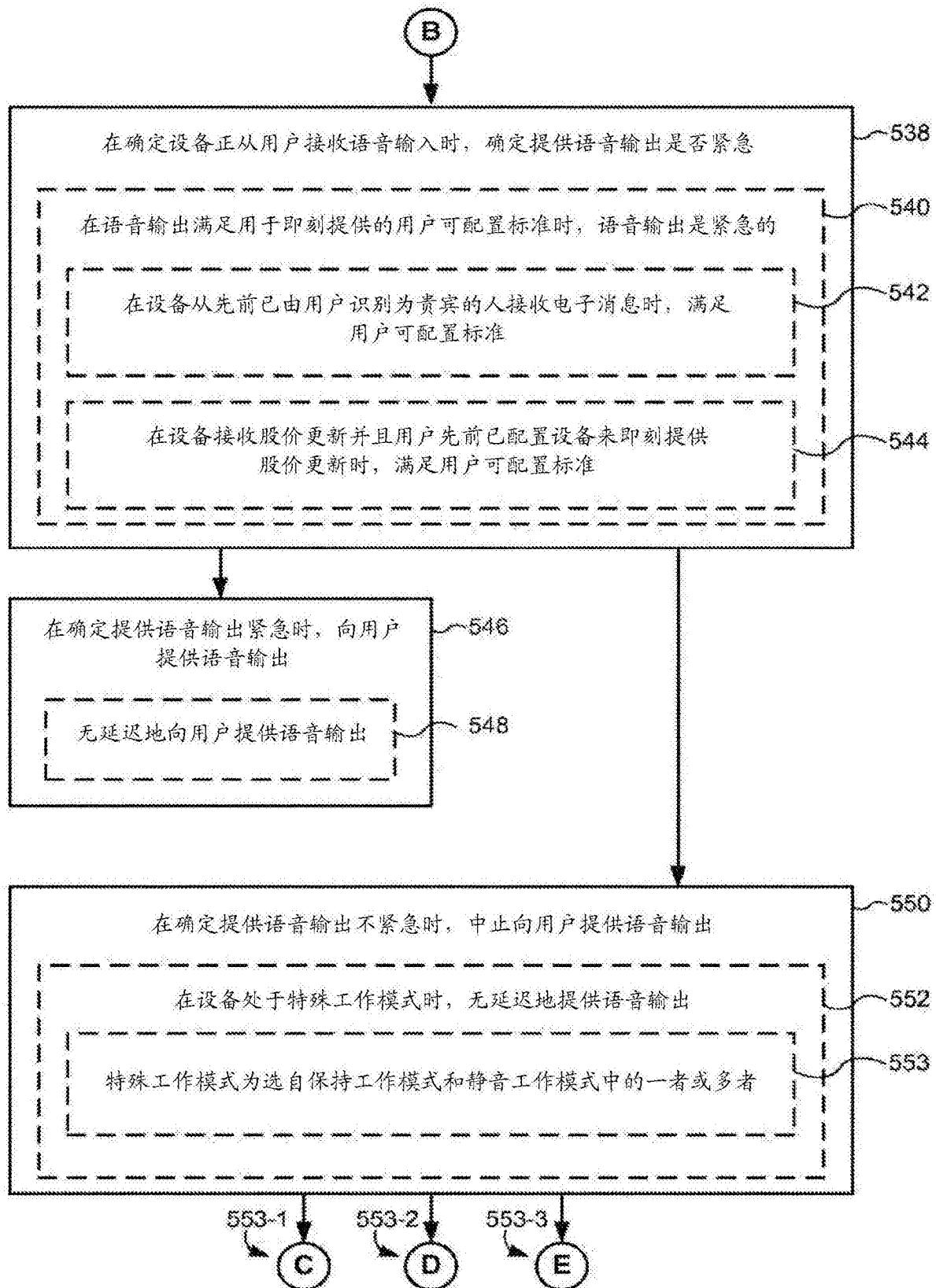


图 5C

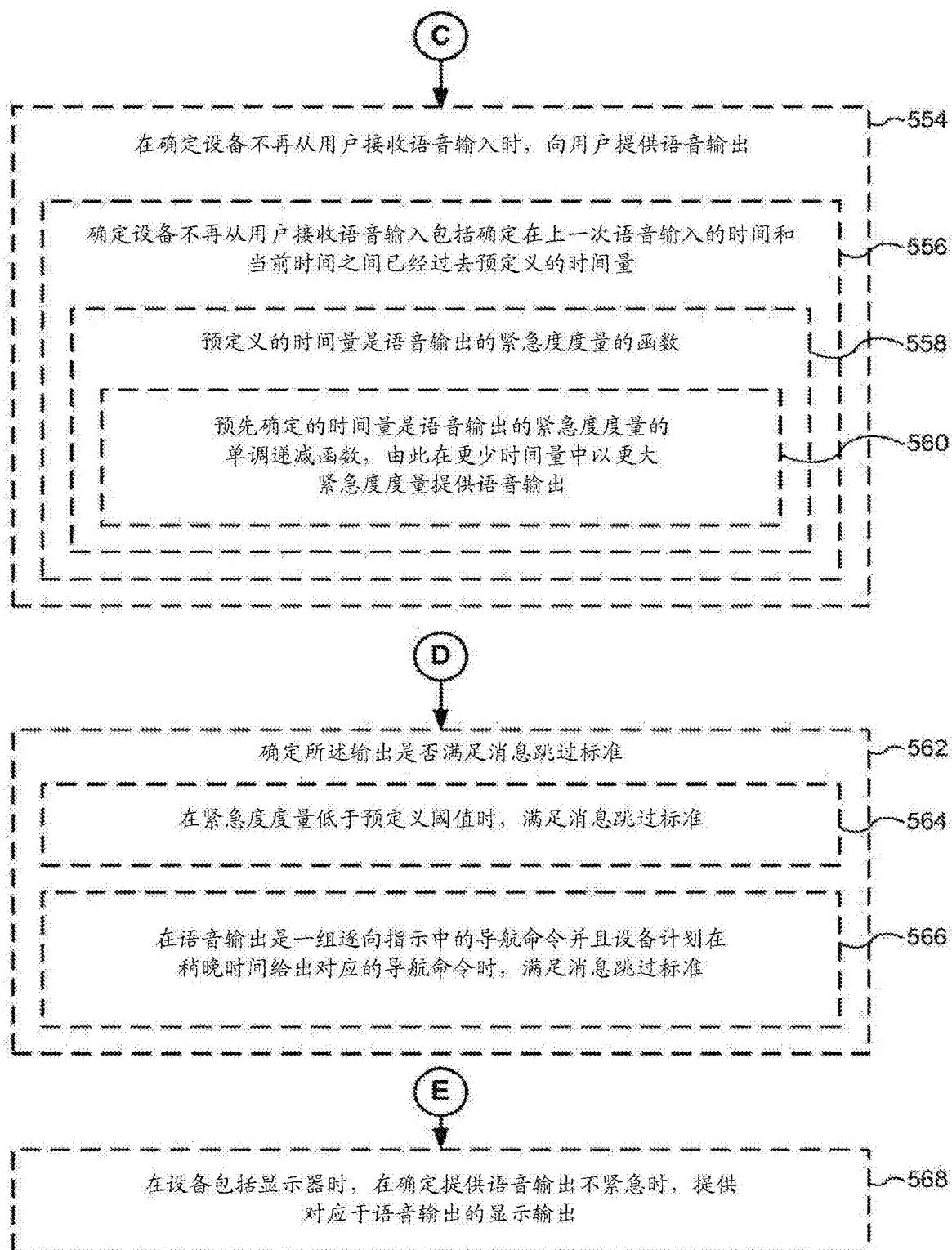


图 5D