



(12) 发明专利申请

(10) 申请公布号 CN 104969289 A

(43) 申请公布日 2015. 10. 07

(21) 申请号 201480007349. 6

(74) 专利代理机构 北京市金杜律师事务所

(22) 申请日 2014. 02. 07

11256

代理人 王茂华 黄倩

(30) 优先权数据

61/762, 260 2013. 02. 07 US

(51) Int. Cl.

G10L 17/24(2006. 01)

(85) PCT国际申请进入国家阶段日

G06F 3/16(2006. 01)

2015. 08. 04

(86) PCT国际申请的申请数据

PCT/US2014/015418 2014. 02. 07

(87) PCT国际申请的公布数据

W02014/124332 EN 2014. 08. 14

(71) 申请人 苹果公司

地址 美国加利福尼亚州

(72) 发明人 J·G·宾德 O·塔金

S·D·波斯特 T·R·格鲁伯

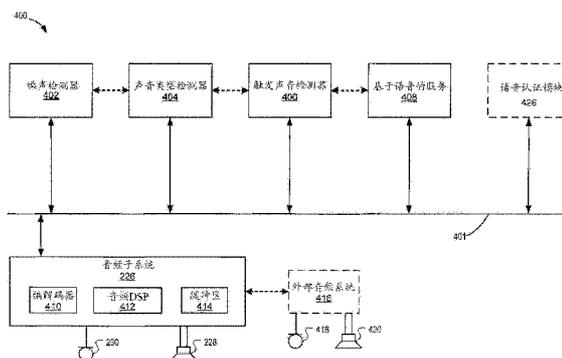
权利要求书3页 说明书32页 附图9页

(54) 发明名称

数字助理的语音触发器

(57) 摘要

本发明提供了一种用于操作语音触发器的方法。在一些具体实施中，该方法在包括存储器和一个或多个处理器的电子设备上执行，该存储器存储由一个或多个处理器执行的指令。该方法包括接收声音输入。声音输入可对应于口头字词或短语，或者它们的一部分。该方法包括确定声音输入的至少一部分是否对应于预先确定的声音类型，诸如人声。该方法包括在确定声音输入的至少一部分对应于预先确定的类型时，确定声音输入是否包括预先确定的内容，诸如预先确定的触发字词或短语。该方法还包括在确定声音输入包括预先确定的内容时，启动基于语音的服务，诸如基于语音的数字助理。



1. 一种用于操作语音触发器的方法,所述方法在包括存储器和一个或多个处理器的电子设备上执行,所述存储器存储由所述一个或多个处理器执行的指令,所述方法包括:

接收声音输入;

确定所述声音输入的至少一部分是否对应于预先确定的声音类型;

在确定所述声音输入的至少一部分对应于所述预先确定的类型时,确定所述声音输入是否包括预先确定的内容;以及

在确定所述声音输入包括所述预先确定的内容时,启动基于语音的服务。

2. 根据权利要求 1 所述的方法,其中所述确定所述声音输入是否对应于预先确定的声音类型由第一声音检测器来执行,其中所述确定所述声音输入是否包括预先确定的内容由第二声音检测器来执行,并且其中所述第一声音检测器在操作时比所述第二声音检测器消耗更少的功率。

3. 根据权利要求 2 所述的方法,其中所述第二声音检测器响应于所述第一声音检测器确定所述声音输入对应于所述预先确定的类型来启动。

4. 根据权利要求 2 所述的方法,其中在所述第一声音检测器确定所述声音输入对应于所述预先确定的类型之后,所述第二声音检测器操作至少预先确定的时间量。

5. 根据权利要求 1 所述的方法,其中所述预先确定的类型为人声并且所述预先确定的内容为一个或多个字词。

6. 根据权利要求 1 所述的方法,其中所述预先确定的内容为一个或多个预先确定的音素。

7. 根据权利要求 6 所述的方法,其中所述一个或多个预先确定的音素构成至少一个字词。

8. 根据权利要求 1 所述的方法,还包括在确定所述声音输入是否对应于预先确定的声音类型之前,确定所述声音输入是否满足预先确定的条件。

9. 根据权利要求 8 所述的方法,其中所述预先确定的条件为幅值阈值。

10. 根据权利要求 8 所述的方法,其中所述确定所述声音输入是否满足预先确定的条件由第三声音检测器来执行,其中所述第三声音检测器在操作时比所述第一声音检测器消耗更少的功率。

11. 根据权利要求 1 所述的方法,还包括:

将所述声音输入的至少一部分存储在存储器中;以及

一旦所述基于语音的服务启动,就将所述声音输入的所述部分提供至所述基于语音的服务。

12. 根据权利要求 1 所述的方法,还包括确定所述声音输入是否对应于特定用户的语音。

13. 根据权利要求 12 所述的方法,其中所述基于语音的服务在确定所述声音输入包括所述预先确定的内容以及所述声音输入对应于所述特定用户的所述语音时启动。

14. 根据权利要求 13 所述的方法,其中所述基于语音的服务在确定所述声音输入包括所述预先确定的内容以及所述声音输入不对应于所述特定用户的所述语音时在受限访问模式下启动。

15. 根据权利要求 13 所述的方法,还包括在确定所述声音输入对应于所述特定用户的

所述语音时,输出包括所述特定用户的姓名的语音提示。

16. 根据权利要求 1 所述的方法,还包括:

确定所述电子设备是否处于预先确定的取向;以及

在确定所述电子设备处于所述预先确定的取向时,启用所述语音触发器的预先确定的模式。

17. 一种用于操作语音触发器的方法,所述方法在包括存储器和一个或多个处理器的电子设备上执行,所述存储器存储由所述一个或多个处理器执行的指令,所述方法包括:

在第一模式下操作语音触发器;

通过检测到所述电子设备的麦克风和相机中的一者或多者被遮挡来确定所述电子设备是否处于大体上封闭的空间中;以及

在确定所述电子设备处于大体上封闭的空间中时,将所述语音触发器切换至第二模式。

18. 根据权利要求 17 所述的方法,其中所述第二模式为待机模式。

19. 根据权利要求 17 中任一项所述的方法,其中所述第一模式为监听模式。

20. 一种用于操作语音触发器的方法,所述方法在包括存储器和一个或多个处理器的电子设备上执行,所述存储器存储由所述一个或多个处理器执行的指令,所述方法包括:

确定所述电子设备是否处于预先确定的取向;以及

在确定所述电子设备处于所述预先确定的取向时,启用语音触发器的预先确定的模式。

21. 根据权利要求 20 所述的方法,其中所述预先确定的取向对应于所述设备的大体上水平且正面朝下的显示屏,并且所述预先确定的模式为待机模式。

22. 根据权利要求 20 所述的方法,其中所述预先确定的取向对应于所述设备的大体上水平且正面朝上的显示屏,并且所述预先确定的模式为监听模式。

23. 一种计算机可读存储介质,所述计算机可读存储介质存储由电子设备的一个或多个处理器执行的一个或多个程序,所述一个或多个程序包括用于执行以下操作的指令:

接收声音输入;

确定所述声音输入的至少一部分是否对应于预先确定的声音类型;

在确定所述声音输入的至少一部分对应于所述预先确定的类型时,确定所述声音输入是否包括预先确定的内容;以及

在确定所述声音输入包括所述预先确定的内容时,启动基于语音的服务。

24. 一种电子设备,包括:

声音接收单元,所述声音接收单元被配置为接收声音输入;和

处理单元,所述处理单元耦接至所述声音接收单元,所述处理单元被配置为:

确定所述声音输入的至少一部分是否对应于预先确定的声音类型;

在确定所述声音输入的至少一部分对应于所述预先确定的类型时,确定所述声音输入是否包括预先确定的内容;以及

在确定所述声音输入包括所述预先确定的内容时,启动基于语音的服务。

25. 根据权利要求 24 所述的电子设备,所述处理单元被进一步配置为:在确定所述声音输入是否对应于预先确定的声音类型之前,确定所述声音输入是否满足预先确定的条

件。

## 数字助理的语音触发器

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2013 年 2 月 7 日提交的名称为“VOICE TRIGGER FOR A DIGITAL ASSISTANT”的美国临时申请 61/762,260 的权益,该申请特此全文以引用方式并入以用于所有目的。

### 技术领域

[0003] 本发明所公开的具体实施整体涉及数字助理,并且更具体地涉及一种用于数字助理的语音触发器的方法和系统。

### 背景技术

[0004] 近来,基于语音的数字助理诸如苹果的 SIRI 已被引入市场来处理各种任务诸如网页搜索和导航。这种基于语音的数字助理的一个优点在于用户能够以免持方式与设备进行交互,而无需触摸乃至查看设备。免持操作在人无法或不能物理地操纵设备的情况下,诸如在他们正在开车的情况下,可以是尤其有利的。然而,为了启动基于语音的助理,用户通常必须在触摸屏上按下按钮或选择图标。这一触觉输入有损于免持体验。因此,提供一种使用语音输入或信号而非触觉输入来启用基于语音的数字助理(或其他基于语音的服务)的方法和系统将是有益的。

[0005] 使用语音输入来启用基于语音的助理需要监测音频通道以检测语音输入。该监测消耗电力,而电力是手持式设备或便携式设备上的有限资源,该手持式设备或便携式设备依靠于电池并且此类基于语音的数字助理常常运行于其上。因此,提供一种可用于启动设备上基于语音的服务的高能效语音触发器将是有益的。

### 发明内容

[0006] 因此,需要可在不过度消耗有限电力资源的情况下提供“随时监听”语音触发功能性的低功率语音触发器。下述的具体实施提供用于使用位于电子设备上的语音触发器来启动基于语音的助理的系统和方法。与基于语音的数字助理(或其他基于语音的服务,诸如语音转文字转录服务)的交互常常开始于用户按下设备上的示能表示(例如,按钮或图标)以启用数字助理时,之后设备向用户提供数字助理为活动的并在监听的某一指示,诸如光、声音(例如,哔哔声)或发声输出(例如,“我能帮你做些什么?”)。如本文所述,也可将语音触发器实现为使得其响应于特定预先确定的字词、短语或声音而启用,并且无需用户的物理交互。例如,用户能够通过叙读短语“嘿, SIRI”来启用 IPHONE 上的 SIRI 数字助理(二者由 Apple Inc., 本申请的受让人提供)。作为响应,设备输出哔哔声、声音或语音输出(例如,“我能帮你做些什么?”)来向用户表明监听模式是活动的。因此,用户可通过数字助理发起交互,而不必物理地触摸提供数字助理功能的设备。

[0007] 一种用于通过语音触发器启动基于语音的服务的技术为使基于语音的服务连续监听预先确定的触发字词、短语或声音(其任一种在本文中可称为“触发声音”)。然而,连

续操作基于语音的服务（例如，基于语音的数字助理）需要大量的音频处理和电池电力。为了降低由提供语音触发功能所消耗的功率，可采用几种技术。在一些具体实施中，电子设备的主处理器（即，“应用处理器”）保持在低功率或无功率状态，而使用较低功率（例如，因为它们不依赖于应用处理器）的一个或多个声音检测器保持为活动的。（当处于低功率或无功率状态时，应用处理器或任何其他处理器、程序或模块可被描述为非活动的或处于待机模式。）例如，即使在应用处理器为非活动的时，低功率声音检测器也用于监测触发声音的音频通道。该声音检测器在本文中有时称为触发声音检测器。在一些具体实施中，其被配置为检测特定声音、音素和 / 或字词。触发声音检测器（包括硬件和 / 或软件部件）被设计成识别特定字词、声音或短语，但就需要更大计算资源和功率资源的任务而言，通常不能提供完整的语音转文本功能或未针对该功能进行优化。因此，在一些具体实施中，触发声音检测器识别语音输入是否包括预定义的模式（例如，匹配字词“嘿，SIRI”的声波模式），但其无法（或未被配置为）将语音输入转换成文本或识别大量其他字词。一旦检测到触发声音，则使数字助理离开待机模式，使得用户能够提供语音命令。

[0008] 在一些具体实施中，触发声音检测器被配置为检测几种不同的触发声音，诸如一组字词、词组、声音和 / 或它们的组合。用户继而可使用那些声音中的任一种来启动基于语音的服务。例如，语音触发器被预先配置为对词组“嘿，SIRI”、“醒来，SIRI”、“调用我的数字助理”或“你好，HAL，听到了吗，HAL？”进行响应。在一些具体实施中，用户必须选择预先配置的触发声音中的一个作为唯一的触发声音。在一些具体实施中，用户选择预先配置的触发声音的子组，使得用户可利用不同的触发声音来启动基于语音的服务。在一些具体实施中，所有的预先配置的触发声音均保持为有效触发声音。

[0009] 在一些具体实施中，使用另一声音检测器使得甚至触发声音检测器也可在大部分时间内保持在低功率或无功率模式。例如，使用不同类型的声音检测器（例如，比触发声音检测器使用更低功率的声音检测器）来监测音频通道以确定声音输入是否对应于某一声音类型。基于声音的某些可辨识特征将声音分类成不同“类型”。例如，属于“人声”类型的声音具有一定的频谱内容、周期性、基频等。其他类型的声音（例如，口哨声、拍手声等）具有不同特征。不同类型的声音使用本文所述的音频和 / 或信号处理器技术来识别。该声音检测器在本文中有时称为“声音类型检测器”。例如，如果预先确定的触发短语为“嘿，SIRI”，则声音类型检测器确定输入是否有可能对应于人的语音。如果触发声音为非语音声音，诸如口哨声，则声音类型检测器确定声音输入是否有可能对应于口哨声。当检测到适当的聲音类型时，声音类型检测器启动触发声音检测器以进一步处理和 / 或分析该声音。并且由于声音类型检测器比触发声音检测器需要更少的功率（例如，由于其使用比触发声音检测器具有更低功率需求和 / 或更高效音频处理算法的电路），因此语音触发功能比只用触发声音检测器消耗更少的功率。

[0010] 在一些具体实施中，使用另一声音检测器使得上述的声音类型检测器和触发声音检测器两者均可在大部分时间内保持在低功率或无功率模式。例如，使用比声音类型检测器使用更低功率的声音检测器来监测音频通道以确定声音输入是否满足预先确定的条件，诸如幅值（例如，音量）阈值。该声音检测器在本文中可称为“噪声检测器”。当噪声检测器检测到满足预先确定阈值的聲音时，噪声检测器启动声音类型检测器以进一步处理和 / 或分析该声音。并且由于噪声检测器比声音类型检测器或触发声音检测器需要更少的功率

(例如,由于其使用具有更低功率需求和/或更高效音频处理算法的电路),因此语音触发功能比不使用噪声检测器而使用声音类型检测器和触发声音检测器的组合消耗更少的功率。

[0011] 在一些具体实施中,上述声音检测器中的任何一者或多者根据占空比进行操作,其中它们在“接通”和“关断”状态之间循环。这就进一步有助于减少语音触发器的功率消耗。例如,在一些具体实施中,噪声检测器“接通”(即,积极地监测音频通道)10毫秒,并且在接下来的90毫秒内“关断”。这样,噪声检测器90%的时间“关断”,而同时仍有效地提供连续噪声检测功能。在一些具体实施中,选择声音检测器的接通和关断持续时间使得所有的检测器在触发声音一直保持输入的情况下启用。例如,对于触发短语“嘿,SIRI”,可配置声音检测器使得不管触发短语开始于一个或多个占空比的哪个位置,触发声音检测器都被及时启用以分析足够的输入量。例如,触发声音检测器将被及时启用以接收、处理并分析声音“嘿,SIRI”,这足以确定声音与触发短语匹配。在一些具体实施中,声音输入在被接收并传送至上游检测器时存储在存储器中,使得较大部分的声音输入可被分析。因此,即使直到触发短语发出之后触发声音检测器才启动,其仍可分析整个已记录的触发短语。

[0012] 一些具体实施提供一种用于操作语音触发器的方法。该方法在包括存储器和一个或多个处理器的电子设备上执行,该存储器存储由一个或多个处理器执行的指令。该方法包括接收声音输入。该方法还包括确定声音输入的至少一部分是否对应于预先确定的声音类型。该方法还包括在确定声音输入的至少一部分对应于预先确定的类型时,确定声音输入是否包括预先确定的内容。该方法还包括在确定声音输入包括预先确定的内容时,启动基于语音的服务。在一些具体实施中,基于语音的服务为基于语音的数字助理。在一些具体实施中,基于语音的服务为听写服务。

[0013] 在一些具体实施中,确定声音输入是否对应于预先确定的声音类型由第一声音检测器执行,并且确定声音输入是否包括预先确定的内容由第二声音检测器执行。在一些具体实施中,第一声音检测器在操作时比第二声音检测器消耗更少的功率。在一些具体实施中,第一声音检测器执行声音输入的频域分析。在一些具体实施中,确定声音输入是否对应于预先确定的声音类型是在确定声音输入满足预先确定的条件(例如,由下述的第三声音检测器确定)时执行的。

[0014] 在一些具体实施中,第一声音检测器根据占空比周期性地监测音频通道。在一些具体实施中,占空比包括约20毫秒的接通时间以及约100毫秒的关断时间。

[0015] 在一些具体实施中,预先确定的类型为人声并且预先确定的内容为一个或多个字词。在一些具体实施中,确定声音输入的至少一部分是否对应于预先确定的声音类型包括确定声音输入的至少一部分是否包括人声的频率特征。

[0016] 在一些具体实施中,第二声音检测器是响应于第一声音检测器确定声音输入对应于预先确定的类型而启动的。在一些具体实施中,在第一声音检测器确定声音输入对应于预先确定的类型之后,第二声音检测器操作至少预先确定的时间量。在一些具体实施中,预先确定的时间量对应于预先确定内容的持续时间。

[0017] 在一些具体实施中,预先确定的内容为一个或多个预先确定的音素。在一些具体实施中,一个或多个预先确定的音素构成至少一个字词。

[0018] 在一些具体实施中,该方法包括在确定声音输入是否对应于预先确定的声音类型

之前,确定声音输入是否满足预先确定的条件。在一些具体实施中,预先确定的条件为幅值阈值。在一些具体实施中,确定声音输入是否满足预先确定的条件是由第三声音检测器来执行的,其中第三声音检测器在操作时比第一声音检测器消耗更少的功率。在一些具体实施中,第三声音检测器根据占空比周期性地监测音频通道。在一些具体实施中,占空比包括约 20 毫秒的接通时间以及约 500 毫秒的关断时间。在一些具体实施中,第三声音检测器执行声音输入的时域分析。

[0019] 在一些具体实施中,该方法包括将声音输入的至少一部分存储在存储器中,并且一旦基于语音的服务启动,就将声音输入的该部分提供至基于语音的服务。在一些具体实施中,使用直接存储器存取将声音输入的一部分存储在存储器中。

[0020] 在一些具体实施中,该方法包括确定声音输入是否对应于特定用户的语音。在一些具体实施中,基于语音的服务是在确定声音输入包括预先确定的内容以及声音输入对应于特定用户的语音时启动的。在一些具体实施中,基于语音的服务在确定声音输入包括预先确定的内容以及声音输入不对应于特定用户的语音时在受限访问模式下启动。在一些具体实施中,该方法包括在确定声音输入对应于特定用户的语音时,输出包括特定用户的姓名的语音提示。

[0021] 在一些具体实施中,确定声音输入是否包括预先确定的内容包括将声音输入的代表与参考表示进行比较,以及在声音输入的代表与参考表示相匹配的情况下确定声音输入是否包括预先确定的内容。在一些具体实施中,如果声音输入的代表以预先确定的置信度与参考表示匹配,则确定匹配。在一些具体实施中,该方法包括接收多个声音输入,该多个声音输入包括声音输入;以及响应于确定相应声音输入包括预先确定的内容,使用多个声音输入中的相应声音输入反复地调节参考表示。

[0022] 在一些具体实施中,该方法包括确定电子设备是否处于预先确定的取向,以及在确定电子设备处于预先确定的取向时,启用语音触发器的预先确定模式。在一些具体实施中,预先确定的取向对应于设备的大体上水平且正面朝下的显示屏,并且预先确定的模式为待机模式。在一些具体实施中,预先确定的取向对应于设备的大体上水平且正面朝上的显示屏,并且预先确定的模式为监听模式。

[0023] 一些具体实施提供一种用于操作语音触发器的方法。该方法在包括存储器和一个或多个处理器的电子设备上执行,该存储器存储由一个或多个处理器执行的指令。该方法包括在第一模式下操作语音触发器。该方法还包括通过检测到电子设备的麦克风和相机中的一者或多者被遮挡来确定电子设备是否处于大体上封闭的空间中。该方法还包括在确定电子设备处于大体上封闭的空间中时,将语音触发器切换至第二模式。在一些具体实施中,第二模式为待机模式。

[0024] 一些具体实施提供一种用于操作语音触发器的方法。该方法在包括存储器和一个或多个处理器的电子设备上执行,该存储器存储由一个或多个处理器执行的指令。该方法包括确定电子设备是否处于预先确定的取向,以及在确定电子设备处于预先确定的取向时,启用语音触发器的预先确定模式。在一些具体实施中,预先确定的取向对应于设备的大体上水平且正面朝下的显示屏,并且预先确定的模式为待机模式。在一些具体实施中,预先确定的取向对应于设备的大体上水平且正面朝上的显示屏,并且预先确定的模式为监听模式。

[0025] 根据一些具体实施,电子设备包括被配置为接收声音输入的声音接收单元和耦接至声音接收单元的处理单元。处理单元被配置为确定声音输入的至少一部分是否对应于预先确定的声音类型;在确定声音输入的至少一部分对应于预先确定的类型时,确定声音输入是否包括预先确定的内容;以及在确定声音输入包括预先确定的内容时,启动基于语音的服务。在一些具体实施中,处理单元被进一步配置为在确定声音输入是否对应于预先确定的声音类型时,确定声音输入是否满足预先确定的条件。在一些具体实施中,处理单元被进一步配置为确定声音输入是否对应于特定用户的语音。

[0026] 根据一些具体实施,电子设备包括被配置为在多个模式中的第一模式下操作语音触发器的语音触发单元;以及耦接至语音触发单元的处理单元。在一些具体实施中,处理单元被配置为:通过检测到电子设备的麦克风和相机中的一者或多者被遮挡来确定电子设备是否处于大体上封闭的空间中;并且在确定电子设备处于大体上封闭的空间中时,将语音触发器切换至第二模式。在一些具体实施中,处理单元被配置为确定电子设备是否处于预先确定的取向;以及在确定电子设备处于预先确定的取向时,启用语音触发器的预先确定模式。

[0027] 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中的任一种方法的指令。

[0028] 根据一些具体实施,提供一种电子设备(例如,便携式电子设备),其包括用于执行本文所述方法中的任一种方法的装置。

[0029] 根据一些具体实施,提供一种电子设备(例如,便携式电子设备),其包括被配置为执行本文所述方法中的任一种方法的处理单元。

[0030] 根据一些具体实施,提供一种电子设备(例如,便携式电子设备),其包括存储器和一个或多个处理器,该存储器存储供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中的任一种方法的指令。

[0031] 根据一些具体实施,提供一种用于在电子设备中使用的信息处理装置,该信息处理装置包括用于执行本文所述方法中的任一种方法的装置。

## 附图说明

[0032] 图 1 为示出根据一些具体实施的数字助理在其中操作的环境的框图。

[0033] 图 2 为示出根据一些具体实施的数字助理客户端系统的框图。

[0034] 图 3A 为示出根据一些具体实施的独立式数字助理系统或数字助理服务器系统的框图。

[0035] 图 3B 为示出根据一些具体实施的图 3A 中所示的数字助理的功能的框图。

[0036] 图 3C 为示出根据一些具体实施的知识本体的一部分的网络图。

[0037] 图 4 为示出根据一些具体实施的语音触发系统的部件的框图。

[0038] 图 5-7 为示出根据一些具体实施的用于操作语音触发系统的方法的流程图。

[0039] 图 8-9 为根据一些实施例的电子设备的功能框图。

[0040] 相似的附图标号是指整个附图中的对应部件。

## 具体实施方式

[0041] 图 1 为根据一些具体实施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”、“基于语音的数字助理”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（例如，识别对应于自然语言输入的任务类型）并基于推断出的用户意图来执行动作（例如，执行对应于所识别的任务类型的任务）的任何信息处理系统。例如，为遵照推断出的用户意图来执行动作，系统可执行以下操作中的一项或多项：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（例如，识别任务类型），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（例如，发送请求至服务提供方）；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。

[0042] 具体地讲，数字助理系统一旦启动，就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系统作出信息性回答，要么寻求数字助理系统执行任务。对用户请求的令人满意的响应通常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，用户可向数字助理系统提出诸如“我现在在哪里？”之类的问题。基于用户的当前位置，数字助理可能回答“你在中央公园西门附近。”用户还可请求执行任务，例如通过叙述“请邀请我的朋友下周来参加我女朋友的生日聚会。”作为响应，数字助理可通过生成语音输出“好的，马上”来确认请求，并且然后将合适的日历邀请从用户的电子邮件地址发送到用户的电子通讯录或联系人列表中列出的用户的每个朋友。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。除了提供口头应答并进行程序化动作之外，数字助理还可提供其他视觉或音频形式的应答（例如，像文本、警报、音乐、视频、动画等）。

[0043] 如图 1 中所示，在一些具体实施中，数字助理系统根据客户端 - 服务器模型来实施。数字助理系统包括在用户设备（例如，104a 和 104b）上执行的客户端侧部分（例如，102a 和 102b）（后文称作“数字助理 (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“数字助理 (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 进行通信。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 的通信。DA 服务器 106 为任意数量的 DA 客户端 102 提供服务器侧功能，该任意数量的 DA 客户端 102 各自位于相应的用户设备 104（也称作客户端设备或电子设备）上。

[0044] 在一些具体实施中，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、到外部服务的 I/O 接口 118、照片与标签数据库 130，以及照片 - 标签模块 132。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。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。照片与标签数据库 130 存储数字照片的指纹并任选地存储数字照片本身，以及与数字照片相关联的标签。照片 - 标签模块 132 创建标签，存储与照片相关联的标签和 / 或指纹，自动标记照片，并将标签连接到照片中的位置。

[0045] 在一些具体实施中，DA 服务器 106 通过一个或多个网络 110 与外部服务 120（例如，一种或多种导航服务 122-1、一种或多种消息型服务 122-2、一种或多种信息服务

122-3、日历服务 122-4、电话服务 122-5、一种或多种照片服务 122-6 等) 进行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。

[0046] 用户设备 104 的例子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业务 (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任何其他合适的处理设备。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。

[0047] 一个或多个通信网络 110 的例子包括局域网 (LAN) 和广域网 (WAN) 例如互联网。一个或多个通信网络 110 可使用任何已知的网络协议来实施, 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、互联网电话协议 (VoIP)、Wi-MAX、或任何其他合适的通信协议。

[0048] 服务器系统 108 可在计算机的至少一个数据处理装置和 / 或分布式网络上实施。在一些具体实施中, 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。

[0049] 尽管图 1 中示出的数字助理系统包括客户端侧部分 (例如, DA 客户端 102) 和服务侧部分 (例如, DA 服务器 106) 两者, 但在一些具体实施中, 数字助理系统仅指服务器侧部分 (例如, DA 服务器 106)。在一些具体实施中, 数字助理的功能可以实现为安装在用户设备上的独立式应用。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具体实施中可变化。例如, 在一些具体实施中, DA 客户端 102 为仅提供面向用户的输入和输出处理功能并且将数字助理的所有其他功能委派给 DA 服务器 106 的瘦客户端。在一些其他具体实施中, DA 客户端 102 被配置为执行或协助 DA 服务器 106 的一个或多个功能。

[0050] 图 2 为根据一些具体实施的用户设备 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并且 / 或者促成用户设备 104 的各种功能。

[0051] 例如, 在一些具体实施中, 运动传感器 210 (例如, 加速度计)、光传感器 212、GPS 接收器 213、温度传感器和接近传感器 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。在一些具体实施中, 其他传感器 216, 诸如生物传感器、气压计等, 连接至外围设备接口 206, 以促成相关功能。

[0052] 在一些具体实施中, 用户设备 104 包括耦接至外围设备接口 206 的相机子系统 220。在一些具体实施中, 相机子系统 220 的光学传感器 222 促成相机功能, 诸如拍摄照片和记录视频剪辑。在一些具体实施中, 用户设备 104 包括提供通信功能的一个或多个有线和 / 或无线通信子系统 224。通信子系统 224 通常包括各种通信端口、射频接收器和发射器、和 / 或光 (例如, 红外) 接收器和发射器。在一些具体实施中, 用户设备 104 包括音频子系统 226, 该音频子系统耦接至一个或多个扬声器 228 及一个或多个麦克风 230 以促成启用语音的功能, 诸如语音识别、语音复制、数字录制和电话功能。在一些具体实施中, 音频

子系统 226 耦接至语音触发系统 400。在一些具体实施中,语音触发系统 400 和 / 或音频子系统 226 包括用于接收和 / 或分析声音输入的低功率音频电路和 / 或程序 (即,包括硬件和 / 或软件),包括例如一个或多个模数转换器、数字信号处理器 (DSP)、检声器、存储缓冲器、编解码器等。在一些具体实施中,低功率音频电路 (单独地或连同用户设备 104 的其他部件) 提供用户设备 104 的一个或多个方面的语音 (或声音) 触发功能,诸如基于语音的数字助理或其他基于语音的服务。在一些具体实施中,低功率音频电路即使在用户设备 104 的其他部件诸如一个或多个处理器 204、I/O 子系统 240、存储器 250 等关闭和 / 或处于待机模式时提供语音触发功能。参考图 4 对语音触发系统 400 作进一步详细描述。

[0053] 在一些具体实施中,I/O 子系统 240 还耦接至外围设备接口 206。在一些具体实施中,用户设备 104 包括触摸屏 246,并且 I/O 子系统 240 包括耦接至触摸屏 246 的触摸屏控制器 242。当用户设备 104 包括触摸屏 246 和触摸屏控制器 242 时,触摸屏 246 和触摸屏控制器 242 通常被配置为,例如,使用多种触摸灵敏度技术中的任一种来检测它们的接触和移动或断开,该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。在一些具体实施中,用户设备 104 包括不具有触敏表面的显示器。在一些具体实施中,用户设备 104 包括独立的触敏表面。在一些具体实施中,用户设备 104 包括一个或多个其他输入控制器 244。当用户设备 104 包括一个或多个其他输入控制器 244 时,该一个或多个其他输入控制器 244 通常耦接至其他输入 / 控制设备 248,诸如一个或多个按钮、摇臂开关、拇指旋轮、红外线端口、USB 端口、和 / 或指针设备诸如触笔。

[0054] 存储器接口 202 耦接至存储器 250。在一些具体实施中,存储器 250 包括非暂态计算机可读介质,诸如高速随机存取存储器和 / 或非易失性存储器 (例如,一个或多个磁盘存储设备、一个或多个闪存存储设备、一个或多个光学存储设备,和 / 或其他非易失性固态存储设备)。在一些具体实施中,存储器 250 存储操作系统 252、通信模块 254、图形用户界面模块 256、传感器处理模块 258、电话模块 260 和应用程序 262 以及它们的子集或超集。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。图形用户界面模块 256 促成图形用户界面处理。传感器处理模块 258 促成与传感器有关的处理和功能 (例如,处理用一个或多个麦克风 228 接收的语音输入)。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。在一些具体实施中,用户设备 104 将每个与外部服务提供方中的至少一个相关联的一个或多个软件应用程序 270-1 和 270-2 存储在存储器 250 中。

[0055] 如上所述,在一些具体实施中,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或联系人列表、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。

[0056] 在各种具体实施中,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、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、视觉和 / 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(例如,数字助理服务器 106,图 1)进行通信。

[0057] 在一些具体实施中,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输入相关联的上下文。在一些具体实施中,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(例如,数字助理服务器 106,图 1)以帮助推断用户的意图。

[0058] 在一些具体实施中,可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。在一些具体实施中,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、功率电平、速度、加速度、运动模式、蜂窝信号强度等。在一些具体实施中,还将与用户设备 106 的软件状态相关的信息,例如,用户设备 104 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至数字助理服务器(例如,数字助理服务器 106,图 1)。

[0059] 在一些具体实施中,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 的至少一部分)。在一些具体实施中,数字助理客户端模块 264 还在数字助理服务器 106(图 1)请求时引出来自用户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。

[0060] 在一些具体实施中,存储器 250 可包括附加指令或更少的指令。此外,用户设备 104 的各种功能可在硬件和/或在固件(包括在一个或多个信号处理集成电路和/或特定于应用的集成电路)中实施,因此,用户设备 104 不需要包括图 2 中示出的所有模块和应用程序。

[0061] 图 3A 为根据一些具体实施的示例性数字助理系统 300(也称为数字助理)的框图。在一些具体实施中,数字助理系统 300 在独立式计算机系统上实施。在一些具体实施中,数字助理系统 300 跨多个计算机分布。在一些具体实施中,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。在一些具体实施中,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。在一些具体实施中,数字助理系统 300 在用户设备(例如,用户设备 104,图 1)中实施,从而消除了对客户端-服务器系统的需求。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 可具有比示出的更多或更少的部件、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在硬件、软件、固件(包括一个或多个信号处理集成电路和/或特定于应用的集成电路),或它们的组合中实现。

[0062] 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。

[0063] 在一些具体实施中,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、一个或多个闪

存存储器设备、一个或多个光学存储设备、和 / 或其他非易失性固态存储器设备)。

[0064] 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。在一些具体实施中, 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 (例如, 一个或多个麦克风 230)。在一些具体实施中, 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。

[0065] 在一些具体实施中, 网络通信接口 308 包括一个或多个有线通信端口 312 和 / 或无线发射和接收电路 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 通常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 能够通过网络, 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), 与其他设备进行通信。

[0066] 在一些具体实施中, 存储器 302 的非暂态计算机可读存储介质存储程序、模块、指令和数据结构, 这些程序、模块、指令和数据结构包括以下中的全部或子集: 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、以及数字助理模块 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。

[0067] 操作系统 318 (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、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。

[0068] 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 的通信模块 254 通信。通信模块 320 还包括各种软件部件, 该各种软件部件用于处理由无线电路 314 和 / 或有线通信端口 312 所接收的数据。

[0069] 在一些具体实施中, 用户界面模块 322 经由 I/O 接口 306 (例如, 从键盘、触摸屏, 和 / 或麦克风) 接收来自用户的命令和 / 或输入, 并将用户界面对象提供在显示器上。

[0070] 应用程序 324 包括被配置为由一个或多个处理器 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。

[0071] 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。在一些具体实施中, 数字助理模块 326 包括以下子模块、或者它们的子集或超集: 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 以及照片模块 132。这些处理模块中的每一个处理模块

均具有对数字助理 326 的以下数据与模型中的一者或多者、或者它们的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、分类模块 349、消歧模块 350、任务流模型 354、服务模型 356、照片标签模块 358、搜索模块 360，以及本地标签 / 照片存储装置 362。

[0072] 在一些具体实施中，使用处理模块（例如，输入 / 输出处理模块 328、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、和 / 或服务处理模块 338）、数据以及在数字助理模块 326 中实施的模型，数字助理系统 300 执行以下操作中的至少一些操作：识别从用户接收的自然语言输入中所表达的用户的意图；积极地引出并获得用于充分推断用户的意图所需的信息（例如，通过消除字词、名称、意向等的歧义）；确定用于实现推断出的意图的任务流；以及执行任务流以实现推断出的意图。在一些具体实施中，当由于各种原因而未向或不能向用户提供令人满意的响应时，数字助理也采取适当的行动。

[0073] 在一些具体实施中，如下所述，数字助理系统 300 从自然语言输入中识别出标记数字照片的用户意图，并且对自然语言输入进行处理以使用适当的信息来标记数字照片。在一些具体实施中，数字助理系统 300 也执行与照片相关的其他任务，诸如使用自然语言输入搜索数字照片、自动标记照片等。如图 3B 所示，在一些具体实施中，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）并提供对用户输入的响应。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数据、词汇，和 / 或与用户输入相关的偏好。在一些具体实施中，上下文信息还包括在接收到用户请求时设备（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。在一些具体实施中，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。在一些具体实施中，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 以进行语音文本转换。

[0074] 在一些具体实施中，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，在语音录制中捕捉的用户话语）。在一些具体实施中，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 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、基于动态时间规整（DTW）的语音识别以及其他统计和 / 或分析技术，来加以实施。在一些具体实施中，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。数字助理 326 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。如本文所用，“可执行意图”表示可由数字助理 326 和 / 或数字助理系统 300（图 3A）执行的任务并且具有在任务流模型 354 中实施的相关联任务流。相关联任务流是数字助理系统 300 为了执行任务而采取的一系列经编程的动作和步骤。数字助理系统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,取决于数字助理系统 300 所识别的“可执行意图”的数量和种类。然而,数字助理系统 300 的有效性还取决于数字助理系统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。

[0075] 在一些具体实施中,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。

[0076] 在一些具体实施中,自然语言处理基于知识本体 360。知识本体 360 为包含多个节点的分级结构,每个节点表示与“可执行意图”或其他“属性”中的一者或多者相关的“可执行意图”或“属性”。如上所述,“可执行意图”表示数字助理系统 300 能够执行的任务(例如,“可执行”或可被进行的任务)。“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。在知识本体 360 中的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。在一些具体实施中,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,图 3C 中所示的知识本体 360 包括“餐厅预订”节点,其为可执行意图节点。属性节点“餐厅”、“日期/时间”(针对预订)和“用餐人数”各自直接连接至“餐厅预订”节点(即,可执行意图节点)。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,并且各自通过中间属性节点“餐厅”连接至“餐厅预订”节点(即,可执行意图节点)。又如,图 3C 中所示的知识本体 360 还包括“设定提醒”节点,其为另一可执行意图节点。属性节点“日期/时间”(针对设定该提醒)和“主题”(针对该提醒)各自连接至“设定提醒”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“设定提醒”节点。

[0077] 可执行意图节点连同其连接的概念节点一起,可被描述为“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 的实例以及提醒域 364 的实例。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“用餐人数”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“设定提醒”和属性节点“主题”和“日期/时间”。在一些具体实施中,知识本体 360 由许多域组成。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其他域(例如,行程安排域、旅游预订域、电影票域等)相关联。尽管图 3C 示出知识本体 360 内的两个示例性域,但知识本体 360 可包括其他域(或可执行意图),诸如“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”、“标记照片”等。例如,“发送消息”域与“发送消息”可执行意图节点相关联,并且还可包括属性节点,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”可进一步例如由子属性

节点诸如“接收人姓名”和“消息地址”来定义。

[0078] 在一些具体实施中,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(并由此包括可执行意图)。在一些具体实施中,知识本体 360 可诸如通过添加或移除域或节点,或者通过修改知识本体 360 内的节点之间的关系来修改。

[0079] 在一些具体实施中,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“机票预定”、“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许多共用的属性节点。例如,针对“机票预定”、“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“用餐人数”中的一者或多者。

[0080] 在一些具体实施中,知识本体 360 中的每个节点与跟由节点所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”。与每个节点相关联的相应的一组字词和/或短语可存储在与由节点所表示的属性或可执行意图相关联的词汇索引 344(图 3B)中。例如,返回图 3B,与“餐厅”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”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”、“给……打电话”、“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。在一些具体实施中,图 3B 中所示的自然语言处理器 332 接收来自语音转文本处理模块 330 的符号序列(例如,文本串),并且确定符号序列中的字词牵涉哪些节点。在一些具体实施中,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),则该字词或短语将“触发”或“启用”那些节点。当多个节点被“触发”时,基于已启用节点的数量和/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作为用户预期让数字助理执行的任务(或任务类型)。在一些具体实施中,选择具有最多“已触发”节点的域。在一些具体实施中,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。在一些具体实施中,基于已触发节点的数量和重要性的组合来选择域。在一些具体实施中,在选择节点的过程中还考虑附加因素,诸如数字助理系统 300 先前是否已正确解译来自用户的类似请求。

[0081] 在一些具体实施中,数字助理系统 300 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时,自然语言处理器 332 将能够识别出该名称涉及知识主体中属性或子属性的特定实例。在一些具体实施中,特定实体的名称是企业、餐厅、人、电影等的名称。在一些具体实施中,数字助理系统 300 可从其他数据源中搜索并识别特定实体名称,该其他数据源诸如用户的通讯录或联系人列表、电影数据库、音乐家数据库和/或餐厅数据库。在一些具体实施中,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(诸如用户通讯录或联系人列表中的姓名)时,在于用户请求的知识主体内选择可执行意图的过程中,该字词被赋予附加重要性。例如,当从用户请求中识别出字词“Santo 先生”并且在词汇索引 344 中发现姓“Santo”是用户联系人列表中的联系人之一时,则用户请求很可能对应于“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”，并且在词汇索引 344 中发现词语“ABC 咖啡馆”是用户所在城市中的特定餐厅的名称时，则用户请求很可能对应于“餐厅预订”域。

[0082] 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。自然语言处理器 332 可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。

[0083] 在一些具体实施中，自然语言处理器 332 包括分类模块 349。在一些具体实施中，如下文更详细论述的，分类模块 349 确定文本串（例如，对应于与数字照片相关联的语音输入）中一个或多个词语中的每个词语是否均为实体、活动或位置中的一者。在一些具体实施中，分类模块 349 将一个或多个词语中的每个词语分类为实体、活动或位置中的一者。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。在一些具体实施中，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，并且这些参数中的至少一些填充有在用户请求中所指定的特定信息和要求。例如，用户可以说“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 }、{ 用餐人数 } 等。基于包含在用户话语中的信息，自然语言处理器 332 可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” } 以及 { 时间 = “晚上 7 点” }。然而，在该实例中，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，其他必要参数诸如 { 用餐人数 } 和 { 日期 } 并未在结构化查询中指定。在一些具体实施中，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，如果用户请求“我附近的”寿司店，则自然语言处理器 332 可用来自用户设备 104 的 GPS 坐标来填充结构化查询中的 { 位置 } 参数。

[0084] 在一些具体实施中，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 被配置为执行以下中的一者或多者：从自然语言处理器 332 接收结构化查询，完成结构化查询，以及执行“完成”用户的最终请求所需的动作。在一些具体实施中，完成这些任务所必需的各种过程在任务流模型 354 中提供。在一些具体实施中，任务流模型 354 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。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 / 或消除可能模糊的话语的歧义。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”）以进行与用户的对话。在一些具体实施中，对话处理模块 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。在一些具体实施中，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。例如，对话处理模块 334 经由音频和 / 或视频输出向用户呈现对话输出，并经由口头或物理（例如，触摸手势）响应接收来自用户的输入。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“用餐人数”和“日期”信息时，对话处理器 334 生成诸如“共有

多少人用餐？”和“具体哪天用餐？”之类的问题以传送至用户。一旦从用户接收到回答，对话处理模块 334 使用缺失信息填充结构化查询，或将信息传送至任务流处理器 336 以完成结构化查询中的缺失信息。

[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，针对“发送消息”域的结构化查询可指示预期接收人为“Bob”，并且用户可能具有多个名为“Bob”的联系人。任务流处理器 336 将请求会话处理器 334 消除结构化查询的该属性的歧义。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] 在一些具体实施中，对话处理器 334 包括消歧模块 350。在一些具体实施中，消歧模块 350 消除一个或多个模糊词语（例如，对应于与数字照片相关联的语音输出的文本串中的一个或多个模糊词语）的歧义。在一些具体实施中，消歧模块 350 识别出一个或多个词语中的第一词语具有多个候选含义，提示用户关于第一词语的附加信息，响应于提示接收来自用户的附加信息并且根据附加信息来识别与第一词语相关联的实体、活动或位置。

[0087] 在一些具体实施中，消歧模块 350 消除代词的歧义。在此类具体实施中，消歧模块 350 将一个或多个词语中的一个词语识别为代词并确定该代词所指代的名词。在一些具体实施中，消歧模块 350 通过使用与电子设备的用户相关联的联系人列表来确定代词所指代的名词。另选地或除此之外，消歧模块 350 还将代词所指代的名词确定为在与先前标记过的数字照片相关联的先前语音输入中所识别的实体、活动或位置的名称。另选地或除此之外，消歧模块 350 还将代词所指代的名词确定为基于与先前标记过的数字照片相关联的先前语音输入所识别的人名。在一些具体实施中，消歧模块 350 访问从手持式电子设备（例如，用户设备 104）的一个或多个传感器（例如，接近传感器 214、光传感器 212、GPS 接收器 213、温度传感器 215 和运动传感器 210）获得的信息以确定一个或多个词语的含义。在一些具体实施中，消歧模块 350 识别各自与实体、活动或位置中的一者相关联的两个词语。例如，两个词语中的第一个是指人，并且两个词语中的第二个是指位置。在一些具体实施中，消歧模块 350 识别各自与实体、活动或位置中的一者相关联的三个词语。

[0088] 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用餐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，用餐人数 = 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被配置为接受针对多个餐厅诸如 ABC 咖啡馆的预订的餐厅预订系统，(2) 在网站上的表格中输入日期、时间和用餐人数信息，(3) 提交表格，以及 (4) 在用户日历中制作针对该预订的日历条目。又如，如下文中详细描述，任务流处理器 336 响应于语音输入执行与标记或搜索数字照片相关联的步骤和指令，例如与照片模块 132 一起执行。在一些具体实施中，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 可代替任务流处理器 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这些其他应用程序交互，以及调用第三方服务（例如，餐厅预订门

户网站、社交网站或服务、银行门户网站等)或与这些第三方服务交互。在一些具体实施中,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。

[0089] 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 356 中的网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、用餐人数)以根据在线预订服务的 API 的格式发送至在线预订接口。

[0090] 在一些具体实施中,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(例如,将输出提供至用户,或完成任务)以满足用户的意图。

[0091] 在一些具体实施中,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。在一些具体实施中,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。

[0092] 现在来关注图 4,其为示出根据一些具体实施的语音触发系统 400 的部件的框图。(语音触发系统 400 并不限于语音,并且本文所述的具体实施同样适用于非语音的声音)。语音触发系统 400 由电子设备 104 内的各种部件、模块和/或软件程序构成。在一些具体实施中,语音触发系统 400 包括噪声检测器 402、声音类型检测器 404、触发声音检测器 406 和基于语音的服务 408 以及音频子系统 226,每个模块均耦接至音频总线 401。在一些具体实施中,使用这些模块中的更多或更少模块。声音检测器 402、404 和 406 可被称为模块,并且可包括硬件(例如,电路、存储器、处理器等)、软件(例如,程序、片上软件、固件等),和/或用于执行本文所述功能的它们的任何组合。在一些具体实施中,如图 4 中虚线所示出的,声音检测器可通信地、可编程地、物理地和/或可操作地彼此耦接(例如,经由通信总线)。(为了便于举例说明,图 4 示出仅与相邻的声音检测器耦接的每个声音检测器。应当理解,每个声音检测器也可耦接至其他声音检测器中的任一个声音检测器。)

[0093] 在一些具体实施中,音频子系统 226 包括编解码器 410、音频数字信号处理器(DSP)412 和存储缓冲器 414。在一些具体实施中,音频子系统 226 耦接至一个或多个麦克风 230(图 2) 和一个或多个扬声器 228(图 2)。音频子系统 226 将声音输入提供至声音检测器 402、404、406 和基于语音的服务 408(以及其他部件或模块,诸如电话和/或电话的基带子系统)以进行处理和/或分析。在一些具体实施中,音频子系统 226 耦接至包括至少一个麦克风 418 和至少一个扬声器 420 的外部音频系统 416。

[0094] 在一些具体实施中,基于语音的服务 408 为基于语音的数字助理,并且对应于上文参考图 1-3C 所描述的数字助理系统的一个或多个部件或功能。在一些具体实施中,基于语音的服务为语音转文本服务、听写服务等。在一些具体实施中,噪声检测器 402 监测音频通道以确定来自音频子系统 226 的声音输入是否满足预先确定的条件,诸如幅值阈值。音频通道对应于由一个或多个拾音装置诸如一个或多个麦克风 230(图 2)所接收的音频信息流。音频通道是指音频信息而不论其处理状态如何或者处理和/或传输该音频信息的特定

硬件如何。例如,音频通道可指来自麦克风 230 的模拟电脉冲(和/或电脉冲在其上传播的电路),以及由对模拟电脉冲进行处理(例如,通过音频子系统 226 和/或电子设备 104 的任何其他音频处理系统)所产生的数字编码的音频流。

[0095] 在一些具体实施中,预先确定的条件是在预先确定的时间量内声音输入是否超过一定音量。在一些具体实施中,噪声检测器使用声音输入的时域分析,该时域分析相对于其他类型的分析(例如,由声音类型检测器 404、触发字词检测器 406 和/或基于语音的服务 408 所执行的)需要相对较少的计算资源和电池资源。在一些具体实施中,使用其他类型的信号处理和/或音频分析,包括例如频域分析。如果噪声检测器 402 确定声音输入满足预先确定的条件,它便启动上游声音检测器,诸如声音类型检测器 404(例如,通过提供控制信号来启动一个或多个处理程序并且/或者通过提供电力至上游声音检测器)。在一些具体实施中,上游声音检测器是响应于满足其他条件而启动的。例如,在一些具体实施中,上游声音检测器是响应于确定设备未被存储在封闭空间中(例如,基于检测光的阈值水平的光检测器)而启动的。

[0096] 声音类型检测器 404 监测音频通道以确定声音输入是否对应于某一类型的声音,诸如具备人声、口哨声、拍手声等特性的声音。声音类型检测器 404 被配置为进行识别的声音类型将对应于语音触发器被配置为进行识别的一个或多个特定触发声音。在触发声音为口头字词或短语的具体实施中,声音类型检测器 404 包括“语音活动检测器”(VAD)。在一些具体实施中,声音类型检测器 404 使用声音输入的频域分析。例如,声音类型检测器 404 生成所接收的声音输入的频谱图(例如,使用傅里叶变换),并分析声音输入的频谱分量以确定声音输入是否可能对应于特定类型或类别的声音(例如,人类语音)。因此,在触发声音为口头字词或短语的具体实施中,如果音频通道正拾取环境声音(例如,交通噪声)而非人类语音,则 VAD 将不启动触发声音检测器 406。在一些具体实施中,只要满足任何下游声音检测器(例如,噪声检测器 402)的预先确定的条件,声音类型检测器 404 就保持为活动的。例如,在一些具体实施中,只要声音输入包括超过预先确定的幅值阈值的聲音(由噪声检测器 402 来确定),声音类型检测器 404 就保持为活动的,并且在声音降至预先确定的阈值以下时被停用。在一些具体实施中,一旦被启动,声音类型检测器 404 就保持为活动的直至满足条件,诸如定时器到期(例如,为时 1 秒、2 秒、5 秒或 10 秒,或任何其他适当持续时间)、声音类型检测器 404 的一定数量的开/关周期结束、或发生事件(例如,由噪声检测器 402 和/或声音类型检测器 404 所确定的,声音的幅值降至第二阈值以下)。

[0097] 如上所述,如果声音类型检测器 404 确定声音输入对应于预先确定的声音类型,它便启动上游声音检测器(例如,通过提供控制信号来启动一个或多个处理程序,并且/或者通过提供电力至上游声音检测器)诸如触发声音检测器 406。

[0098] 触发声音检测器 406 被配置为确定声音输入是否包括某些预先确定内容的至少一部分(例如,触发字词、短语或声音的至少一部分)。在一些具体实施中,触发声音检测器 406 将声音输入的代表(“输入表示”)与触发字词的一个或多个参考表示进行比较。如果输入表示以可接受置信度与一个或多个参考表示中的至少一者相匹配,触发声音检测器 406 便启动基于语音的服务 408(例如,通过提供控制信号来启动一个或多个处理程序,并且/或者通过提供电力至上游声音检测器)。在一些具体实施中,输入表示和一个或多个参考表示为频谱图(或它们的数学表示),这些频谱图表示出信号的频谱密度怎样随时间

变化。在一些具体实施中,表示是其他类型的音频特征或声纹。在一些具体实施中,启动基于语音的服务 408 包括使一个或多个电路、程序和 / 或处理器离开待机模式,并且调用基于声音的服务。基于声音的服务继而准备好提供更全面的语音识别、语音转文本处理,和 / 或自然语言处理。在一些具体实施中,语音触发系统 400 包括语音认证功能,使其能够确定声音输入是否对应于特定人诸如设备的所有者 / 使用者的语音。例如,在一些具体实施中,声音类型检测器 404 使用声纹技术来确定声音输入是由授权用户发出的。语音认证和声纹在转让给本申请受让人的美国专利申请 13/053,144 中更详细描述,该文献特此全文以引用方式并入。在一些具体实施中,语音认证包括在本文所述的任一种声音检测器中(例如,噪声检测器 402、声音类型检测器 404、触发声音检测器 406 和 / 或基于语音的服务 408)。在一些具体实施中,将语音认证实现为独立于上文列出的声音检测器的模块(例如,作为语音认证模块 428,图 4),并且可被可操作地定位在噪声检测器 402 之后、声音类型检测器 404 之后、触发声音检测器 406 之后、或定位在任何其他适当的位置处。

[0099] 在一些具体实施中,只要满足任何一个或多个下游声音检测器(例如,噪声检测器 402 和 / 或声音类型检测器 404)的条件,触发声音检测器 406 就保持为活动的。例如,在一些具体实施中,只要声音输入包括超过预先确定的阈值的聲音(如由噪声检测器 402 来检测),触发声音检测器 406 就保持为活动的。在一些具体实施中,只要声音输入包括某种类型的声音(如由声音类型检测器 404 来检测),触发声音检测器 406 就保持为活动的。在一些具体实施中,只要同时满足上述条件,触发声音检测器 406 就保持为活动的。

[0100] 在一些具体实施中,一旦被启动,触发声音检测器 406 就保持为活动的直至满足条件,诸如定时器到期(例如,为时 1 秒、2 秒、5 秒或 10 秒,或者任何其他适当持续时间)、触发声音检测器 406 的一定数量的开 / 关周期结束、或发生事件(例如,声音的幅值降至第二阈值以下)。在一些具体实施中,当一个声音检测器启动另一检测器时,这两个声音检测器均保持为活动的。然而,声音检测器可在不同时间为活动的或非活动的,并且不必为了使上游声音检测器为活动的而使所有上游(例如,较低功率和 / 或复杂性)声音检测器均为活动的(或满足它们的相应条件)。例如,在一些具体实施中,在噪声检测器 402 和声音类型检测器 404 确定满足它们的相应条件并且触发声音检测器 406 被启动之后,在触发声音检测器 406 操作时,噪声检测器 402 和声音类型检测器 404 中的一者或两者被停用并且 / 或者进入待机模式。在其他具体实施中,在触发声音检测器 406 操作时,噪声检测器 402 和声音类型检测器 404 两者(或者其中一者或另一者)保持为活动的。在各种具体实施中,声音检测器的不同组合在不同时间为活动的,并且一个声音检测器为活动的还是非活动的可取决于另一声音检测器的状态,或者可独立于另一声音检测器的状态。

[0101] 尽管图 4 描述了三个单独的声音检测器,其各自被配置为检测声音输入的不同方面,但在语音触发器的各种具体实施中可使用更多或更少的声音检测器。例如,在一些具体实施中,仅使用触发声音检测器 406。在一些具体实施中,触发声音检测器 406 与噪声检测器 402 或声音类型检测器 404 一起使用。在一些具体实施中,使用所有检测器 402-406。在一些具体实施中,还包括另外的声音检测器。

[0102] 此外,可在不同时间使用声音检测器的不同组合。例如,声音检测器的特定组合以及它们的交互方式可取决于一个或多个条件,诸如设备的上下文或操作状态。举个具体例子,如果设备接通电源(并从而不仅仅依赖于电池电力),触发声音检测器 406 便为活动的,

而噪声检测器 402 和声音类型检测器 404 保持为非活动的。又如,如果设备在口袋或背包里,则所有的声音检测器均为非活动的。通过级联如上所述的声音检测器,其中需要较低功率的检测器仅在必要时调用需要更多功率的检测器,这就可提供高效语音触发功能。如上所述,额外的功率效率通过根据占空比操作声音检测器中的一者或多者来实现。例如,在一些具体实施中,噪声检测器 402 根据占空比进行操作,使得噪声检测器即使在至少一部分时间内关闭,也可有效地执行连续噪声检测。在一些具体实施中,噪声检测器 402 接通 10 毫秒并关断 90 毫秒。在一些具体实施中,噪声检测器 402 接通 20 毫秒并关断 500 毫秒。其他开关持续时间也是可能的。

[0103] 在一些具体实施中,如果噪声检测器 402 在其“接通”时段期间检测到噪声,则噪声检测器 402 将保持接通以进一步处理和 / 或分析声音输入。例如,噪声检测器 402 可被配置为在其于预先确定的时间量内(例如,100 毫秒)检测到超过预先确定幅值的聲音的情况下启动上游声音检测器。因此,如果噪声检测器 402 在其 10 毫秒的“接通”时段期间检测到超过预先确定幅值的聲音,则它将不立即进入“关断”时段。相反,噪声检测器 402 保持为活动的并继续处理声音输入以确定其是否在整个预先确定的持续时间内(例如,100 毫秒)超过阈值。

[0104] 在一些具体实施中,声音类型检测器 404 根据占空比进行操作。在一些具体实施中,声音类型检测器 404 接通 20 毫秒并关断 100 毫秒。其他开关持续时间也是可能的。在一些具体实施中,声音类型检测器 404 能够在其占空比的“接通”时段内确定声音输入是否对应于预先确定类型的聲音。因此,如果声音类型检测器 404 在其“接通”时段期间确定声音为某种类型,声音类型检测器 404 将启动触发声音检测器 406(或任何其他上游声音检测器)。另选地,在一些具体实施中,如果声音类型检测器 404 在“接通”时段期间检测到可能对应于预先确定的类型的聲音,则检测器将不立即进入“关断”时段。相反,声音类型检测器 404 保持为活动的并继续处理声音输入以及确定其是否对应于预先确定的聲音类型。在一些具体实施中,如果声音检测器确定已检测到预先确定的聲音类型,它便启动触发声音检测器 406 以进一步处理声音输入并确定是否已检测到触发声音。类似于噪声检测器 402 和声音类型检测器 404,在一些具体实施中,触发声音检测器 406 根据占空比进行操作。在一些具体实施中,触发声音检测器 406 接通 50 毫秒并关断 50 毫秒。其他开关持续时间也是可能的。如果触发声音检测器 406 在其“接通”时段期间检测到存在可能对应于触发聲音的的聲音,则检测器将不立即进入“关断”时段。相反,触发声音检测器 406 保持为活动的并继续处理声音输入以及确定其是否包括触发聲音。在一些具体实施中,如果检测到这种聲音,触发声音检测器 406 便保持为活动的以在预先确定的持续时间内诸如 1 秒、2 秒、5 秒或 10 秒,或任何其他适当的持续时间内处理音频。在一些具体实施中,持续时间是基于特定触发字词的长度或被配置为进行检测的聲音来选择的。例如,如果触发短语是“嘿, SIRI”,则触发字词检测器在约 2 秒内进行操作以确定声音输入是否包括该短语。

[0105] 在一些具体实施中,声音检测器中的一些声音检测器根据占空比进行操作,而其他检测器在为活动的时连续操作。例如,在一些具体实施中,仅第一声音检测器(例如,图 4 中的噪声检测器 402)根据占空比进行操作,并且上游声音检测器一旦被启动就连续操作。在一些其他具体实施中,噪声检测器 402 和声音类型检测器 404 根据占空比进行操作,而触发声音检测器 406 连续操作。特定的声音检测器是连续操作还是根据占空比进行操作取决

于一个或多个条件,诸如设备的上下文或操作状态。在一些具体实施中,如果设备接通电源并不仅仅依赖于电池电力,则所有的声音检测器一旦被启动便连续操作。在其他具体实施中,如果设备在口袋或背包中(例如,由传感器和/或麦克风信号来确定),则噪声检测器 402(或声音检测器中的任一种)根据占空比进行操作,但在确定该设备很可能未被存放的情况下,噪声检测器连续地操作。在一些具体实施中,特定的声音检测器是连续操作还是根据占空比进行操作取决于设备的电池充电水平。例如,在电池充电量超过 50%时,噪声检测器 402 连续操作,并且在电池充电量低于 50%时,噪声检测器 402 根据占空比进行操作。在一些具体实施中,语音触发器包括噪声、回声,和/或声音消除功能(统称为噪声消除)。在一些具体实施中,噪声消除是由音频子系统 226(例如,由音频 DSP 412)来执行的。噪声消除在声音输入由声音检测器进行处理之前,从声音输入中减少或消除了不必要噪声或声音。在一些情况下,不必要噪声为来自用户环境的背景噪声,诸如来自风扇或键盘发出的点击声。在一些具体实施中,不必要噪声为上述、下述或预先确定幅值或频率下的任何声音。例如,在一些具体实施中,将超过典型人类音域(例如,3,000Hz)的声音从信号中滤除或去除。在一些具体实施中,使用多个麦克风(例如,麦克风 230)来帮助确定应减小和/或去除所接收声音的哪些成分。例如,在一些具体实施中,音频子系统 226 使用波束形成技术来识别似乎源于空间(例如,用户口中)中的单个点的声音或声音输入的部分。音频子系统 226 继而通过从声音输入中去除由所有麦克风等同接收的声音(例如,好像并不源于任何特定方向的环境声音)来集中注意力于该声音。

[0106] 在一些具体实施中,DSP 412 被配置为从声音输入中消除或去除由数字助理操作于其上的设备正输出的声音。例如,如果音频子系统 226 正在输出音乐、广播、播客、语音输出或任何其他音频内容(例如,通过扬声器 228),则 DSP 412 去除由麦克风拾取并包括在声音输入中的任何输出的声音。因此,声音输入不含输出的音频(或至少包含较少的输出的音频)。因此,提供至声音检测器的声音输入将比较清晰,并且触发更准确。噪声消除的方面在转让给本申请受让人的美国专利 7,272,224 中更详细描述,该文献特此全文以引用方式并入。

[0107] 在一些具体实施中,不同声音检测器需要声音输入以不同方式进行滤除和/或预处理。例如,在一些具体实施中,噪声检测器 402 被配置为在 60Hz 和 20,000Hz 之间分析时域音频信号,并且声音类型检测器被配置为在 60Hz 和 3,000Hz 之间执行音频的频域分析。因此,在一些具体实施中,音频 DSP412(和/或设备 104 的其他音频 DSP)根据声音检测器的相应需要对所接收的音频进行预处理。在一些具体实施中,另一方面,声音检测器被配置为根据它们的具体需要来滤除和/或预处理来自音频子系统 226 的音频。在这种情况下,音频 DSP 412 仍可在将声音输入提供至声音检测器之前执行噪声消除。在一些具体实施中,可使用电子设备的上下文来帮助确定是否操作语音触发器以及如何操作语音触发器。例如,当设备存放在用户的口袋、钱包或背包中时,他们将不太可能调用基于语音的服务,诸如基于语音的数字助理。另外,当用户在听喧闹的摇滚音乐会时,他们将不太可能调用基于语音的服务。对于一些用户来说,他们将不太可能在一天中的某些时候(例如,深夜)调用基于语音的服务。另一方面,还存在用户将很可能使用语音触发器来调用基于语音的服务的上下文。例如,一些用户将很可能在他们正在开车时、在他们独自一人时、在他们工作时等情况下使用语音触发器。使用各种技术来确定设备的上下文。在各种具体实施中,设备使用来自以

下部件或信息源中的任何一者或多者的信息来确定设备的上下文:GPS 接收器、光传感器、麦克风、接近传感器、方向传感器、惯性传感器、相机、通信电路和 / 或天线、充电和 / 或功率电路、开关位置、温度传感器、指南针、加速度计、日历、用户偏好等。设备的上下文继而能够用于调节语音触发器如何操作以及语音触发器是否操作。例如,在某些上下文中,只要保持该上下文,语音触发器就将被停用(或在不同模式下操作)。例如,在一些具体实施中,在电话处于预先确定的取向(例如,正面朝下放在一表面上)时、在预先确定的时间段内(例如,在晚上 10:00 和上午 8:00 之间)、在电话处于“静音”或“勿扰”模式时(例如,基于开关位置、模式设置或用户偏好)、在设备处于大体上封闭空间(例如,口袋、书包、钱包、抽屉或手套箱)中时、在设备位于具有语音触发器和 / 或基于语音的服务的其他设备附近时(例如,基于接近传感器、声 / 无线 / 红外通信)等情况下,语音触发器被停用。在一些具体实施中,语音触发系统 400 在低功率模式下被操作(例如,通过根据具有 10 毫秒“接通”时段和 5 秒的“关断”时段的占空比操作噪声检测器 402),而不是被停用。在一些具体实施中,当语音触发系统 400 在低功率模式下被操作时,以更低频率监测音频通道。在一些具体实施中,当语音触发器处于低功率模式下时,语音触发器使用与之处于正常模式下不同的声音检测器或声音检测器的组合。(语音触发器能够进行多种不同模式或操作状态,每种可使用不同的功率值,并且不同具体实施将根据它们的具体设计来使用这些模式或操作状态)。

[0108] 另一方面,当设备处于一些其他上下文中时,语音触发器将被启用(或在不同模式下被操作),只要保持该上下文即可。例如,在一些具体实施中,在语音触发器接通电源的情况下、在电话处于预先确定的取向时(例如,正面朝上放在一表面上)、在预先确定的时间段内(例如,在上午 8:00 和晚上 10:00 之间)、在设备正行进和 / 或在车里(例如,基于 GPS 信号、蓝牙连接或与车辆对接等)等情况下,语音触发器保持为活动的。设备在车辆中的检测方面在转让给本申请受让人的美国临时专利申请 61/657,744 中更详细地描述,该文献特此全文以引用方式并入。如何确定某些上下文的若干具体实例在下文中提供。在各种实施例中,使用不同技术和 / 或信息源来检测这些和其他上下文。

[0109] 如上所述,语音触发系统 400 是否为活动的(例如,监听)可取决于设备的物理取向。在一些具体实施中,当设备“正面朝上”位于一表面上时(例如,显示器和 / 或触摸屏表面可见),语音触发器为活动的,并且 / 或者当设备“正面朝下”时,语音触发器为非活动的。这就提供给用户一种在无需操纵设置菜单、开关或按钮的情况下启用和 / 或停用语音触发器的简单方法。在一些具体实施中,设备使用光传感器(例如,基于到设备 104 的正面和背面的入射光的差异)、接近传感器、磁性传感器、加速度计、陀螺仪、倾斜传感器、相机等检测其是正面朝上还是正面朝下地位于一表面上。在一些具体实施中,其他操作模式、设置、参数或偏好受到设备的取向和 / 或位置的影响。在一些具体实施中,语音触发器的特定触发声音、字词或短语取决于设备的取向 / 或位置进行监听。例如,在一些具体实施中,当设备处于第一取向时(例如,正面朝上放在一表面上),语音触发器监听第一触发字词、短语或声音,并且当设备处于另一取向时(例如,正面朝下放置),语音触发器监听不同的触发字词、短语或声音。在一些具体实施中,针对正面朝下取向的触发短语比针对正面朝上取向的触发短语长和 / 或复杂。因此,用户在他们周围有其他人或处于嘈杂环境中时可使设备正面朝下,使得语音触发器仍可在减少错误接受率的情况下进行操作,这可能更频繁地需要更短或更简单的触发字词。举个具体例子,正面朝上触发字词可以是“嘿, SIRI”,而正面朝

下触发短语可以是“嘿, SIRI, 我是 Andrew, 请醒过来”。较长触发短语还为声音检测器和 / 或语音认证器用于处理和 / 或分析提供了较大的语音样本, 从而提高了语音触发器的准确度并降低了错误接受率。

[0110] 在一些具体实施中, 设备 104 检测其是否在车辆 (例如, 汽车) 中。语音触发器对在用户处于车辆中的情况下调用基于语音的服务尤其有益, 因为它有助于减少操作设备和 / 或基于语音的服务所必需的物理交互。实际上, 基于语音的数字助理的有益效果之一在于其可用于在查看和触摸设备将不太可能或不够安全的情况下执行任务。因此, 语音触发器可在设备处于车辆中时使用, 使得用户不必触摸设备来调用数字助理。在一些具体实施中, 设备通过检测其已连接至和 / 或配对于车辆, 诸如通过蓝牙通信 (或其他无线通信) 或通过扩展坞接口或缆线, 来确定其处于车辆中。在一些具体实施中, 设备通过确定设备的位置和 / 或速度 (例如, 使用 GPS 接收器、加速度计和 / 或陀螺仪) 来确定其处于车辆中。如果例如由于设备正以每小时 20 英里以上的速度行进并且被确定正沿道路行进而确定设备可能处于车辆中, 则语音触发器保持为活动的和 / 或处于高功率或较敏感状态。

[0111] 在一些具体实施中, 设备通过确定其是否处于大体上封闭的空间中来检测其是否被存放 (例如, 在口袋、钱包、书包、抽屉等中)。在一些具体实施中, 设备使用光传感器 (例如, 专用环境光传感器和 / 或相机) 来确定其被存放。例如, 在一些具体实施中, 如果光传感器检测到少许光或未检测到光, 则设备很可能被存放。在一些具体实施中, 还对一天中的时间和 / 或设备的位置进行考虑。例如, 如果光传感器在预期较高光照水平的情况下 (例如, 在白天期间) 检测到较低光照水平, 则设备可能处于存放状态并且无需语音触发系统 400。因此, 将使语音触发系统 400 处于低功率或待机状态。在一些具体实施中, 位于设备的不同面上的传感器检测到的光的差异可用于确定设备的位置, 并由此确定其是否被存放。具体地, 用户很可能在设备放在桌子或表面上时而不是在设备存放在口袋或书包中时去尝试启用语音触发器。然而当设备正面朝下 (或正面朝上) 放在一表面诸如桌子或办公桌上时, 设备的一个表面将被遮挡使得少许光或没有光到达该表面, 而另一表面将暴露于环境光中。因此, 如果位于设备的正面和背面的光传感器检测到明显不同的光照水平, 则设备确定其并非被存放。另一方面, 如果位于相对面的光传感器检测到相同或类似的光照水平, 设备便确定其存放在大体上封闭的空间中。另外, 如果两个光传感器在白天期间 (或在设备预期电话处于明亮环境中时) 检查到较低光照水平, 设备便以更大置信度确定其被存放。

[0112] 在一些具体实施中, 还使用其他技术 (代替或除了光传感器) 来确定设备是否被存放。例如, 在一些具体实施中, 设备从扬声器或换能器 (例如, 扬声器 228) 发出一种或多种声音 (例如, 音调声、点击声、撞击声等), 并且监测一个或多个麦克风或换能器 (例如, 麦克风 230) 来检测发出的一种或多种声音的回声。(在一些具体实施中, 设备发出听不见的信号, 诸如人类听觉范围之外的声音)。通过回声, 设备确定周边环境的特征。例如, 相对较大环境 (例如, 房间或车辆) 将反射不同于相对较小的封闭环境 (例如, 口袋、钱包、书包、抽屉等) 的声音。

[0113] 在一些具体实施中, 在语音触发系统位于其他设备 (诸如具有语音触发器和 / 或基于语音的服务的其他设备) 附近的情况下与在其不位于其他设备附近的情况下, 语音触发系统 400 操作有所不同。例如在许多设备彼此靠近的情况下, 关闭或降低语音触发系统 400 的敏感性是有用的, 以使得在一个人发出触发字词时, 其他周边设备不同样被触发。在

一些具体实施中,设备使用 RFID、近场通信、红外 / 声信号等来确定接近其他设备。如上所述,当设备在免持模式下操作时,诸如当用户正在开车时,语音触发器尤其有用。在这种情况下,用户常常使用外部音频系统,诸如有线或无线耳机、带有扬声器和 / 或麦克风的手表、车辆的内置麦克风和扬声器等,来使自身不必将设备靠近其面部来进行呼叫或指示文本输入。例如,无线耳机和车辆音频系统可使用蓝牙通信或任何其他适当的无线通信连接至电子设备。然而,由于通过无线附件来保持打开音频通道所需的功率方面的原因,语音触发器通过无线音频附件来监测所接收的音频可能效率低下。具体地,无线耳机可在其电池中保持足够充电量以提供几个小时的连续通话时间,并因此优选的是储备电池以用于在实际通信需要耳机的情况下,而不是使用它来简单地监测环境音频并等待可能的触发声音。此外,有线外部耳机附件比单独使用车载麦克风可能需要明显更多电力,并且使耳机麦克风保持为活动的将耗尽设备的电池充电量。就由无线或有线耳机所接收的环境音频通常将主要由无声或不相干声音构成而言,尤其是如此。因此,在一些具体实施中,即使当设备耦接至外部麦克风(有线或无线)时,语音触发系统 400 也监测来自设备上的麦克风 230 的音频。然后,当语音触发器检测到触发字词时,设备初始化具有外部麦克风的活动的音频链路以通过外部麦克风而不是设备上的麦克风 230 来接收后续的声音输入(诸如对基于语音的数字助理的命令)。在满足某些条件时,但也可在外部音频系统 416(其可经由有线或无线方式通信地耦接至设备 104)与设备之间保持活动通信链路使得代替设备上的麦克风 230(或除其之外),语音触发系统 400 还可通过外部音频系统 416 来监听触发声音。例如,在一些具体实施中,电子设备和 / 或外部音频系统 416 的运动特征(例如,如由相应设备上的加速度计、陀螺仪等确定的)用于确定语音触发系统 400 应使用设备上的麦克风 230 还是使用外部麦克风 418 来监测环境声音。具体地,设备和外部音频系统 416 的运动之间的差异提供关于外部音频系统 416 是否实际在使用中的信息。例如,如果设备和无线耳机两者都在大体上相同地移动(或未移动),则可确定耳机未在使用中或未被佩戴。这可能是例如由于这两个设备彼此靠近并闲置(例如,放在桌子上或存放在口袋、书包、钱包、抽屉等中)而出现。因此,在这些情况下,语音触发系统 400 监测设备上的麦克风,因为耳机实际上正被使用是不太可能的。然而,如果无线耳机和设备之间的运动存在差异,则确定耳机正被用户佩戴。这些情况可能是例如由于设备已被放下(例如,在一表面上或在书包里),而耳机正佩戴在用户的头上(即使在佩戴者相对静止时,耳机也将可能至少少许地移动)而出现。在这些情况下,由于耳机很可能正被佩戴着,因此语音触发系统 400 保持活动通信链路并且代替设备上的麦克风 230(或除其之外),还监测耳机的麦克风 418。并且由于该技术重点关注设备和耳机的运动上的差异,因此两个设备所共有的运动可被抵消。这对以下情况是很有用的:例如当用户正在移动的车辆上使用耳机,其中设备(例如,蜂窝电话)正位于杯托中、空座位上或用户的口袋中,并且耳机正佩戴在用户的头上。一旦两个设备共有的运动相抵消(例如,车辆的运动),则可确定耳机相比于设备的相对运动(如果有的话)以确定耳机是否可能在使用中(或者,耳机是否未被佩戴着)。尽管以上论述涉及无线耳机,但类似的技术也可应用于有线耳机。

[0114] 由于人的声音存在很大区别,因此调整语音触发器来提高其在识别特定用户的语音中的准确性是必要或有利的。另外,人的声音可随时间变化,例如,由于生病、与年龄变化或激素变化相关的自然声音变化等。因此,在一些具体实施中,语音触发系统 400 能够调整

其针对特定用户或用户群的语音和 / 或声音识别配置文件。如上所述,声音检测器(例如,声音类型检测器 404 和 / 或触发声音检测器 406)可被配置为将声音输入(例如,用户提供的声音或话语)与一个或多个参考表示进行比较。例如,如果输入表示以预先确定的置信度与参考表示匹配,则声音检测器将确定声音输入对应于预先确定的声音类型(例如,声音类型检测器 404),或者确定声音输入包括预先确定的内容(例如,触发声音检测器 406)。为了调整语音触发系统 400,在一些具体实施中,设备对输入表示与之比较的参考表示进行调节。在一些具体实施中,参考表示被调节(或创建)成语音登记或“训练”程序的一部分,其中用户多次输出触发声音以便设备能够调节(或创建)参考表示。设备继而可使用此人的实际语音来创建参考表示。

[0115] 在一些具体实施中,设备使用在正常使用条件下所接收的触发声音来调节参考表示。例如,在成功的语音触发事件之后(例如,在发现满足所有触发标准的声音输入的情况下),设备将使用来自声音输入的信息来调节和 / 或调整参考表示。在一些具体实施中,只有以某一置信度被确定满足所有或一些触发标准的声音输入用于调节参考表示。因此,当语音触发器在声音输入对应于或包括触发声音方面置信度较低时,出于调节参考表示的目的,语音输入可被忽略。另一方面,在一些具体实施中,以较低置信度满足语音触发系统 400 的声音输入用于调节参考表示。

[0116] 在一些具体实施中,随着接收到越来越多的声音输入,设备 104 反复地调节参考表示(使用这些或其他技术),使得用户语音随时间的微小变化可得以适应。例如,在一些具体实施中,在每个成功触发事件之后,设备 104(和 / 或相关联的设备或服务)调节参考表示。在一些具体实施中,设备 104 分析与每个成功触发事件相关联的声音输入并确定是否应基于该输入调节参考表示(例如,在满足某些条件的情况下),并且在适当的情况下仅调节参考表示。在一些具体实施中,设备 104 随时间保持参考表示的移动平均值。在一些具体实施中,语音触发系统 400 检测不满足触发标准中的一者或多者(例如,由声音检测器中的一者或多者确定)但实际上可能是由授权用户所进行尝试的声音。例如,语音触发系统 400 可被配置为对触发短语诸如“嘿, SIRI”进行响应,但如果用户的语音发生变化(例如,由于生病、年龄、口音 / 音调变化等),语音触发系统 400 便可能不识别用户启用设备的尝试。(这也可能发生在语音触发系统 400 尚未针对用户的特定声音进行适当调整的情况下,诸如在将语音触发系统 400 设定为默认状况并且 / 或者用户尚未执行初始化或训练程序以针对其语音来定制语音触发系统 400 的情况下)。如果语音触发系统 400 未对用户启用语音触发器的首次尝试进行响应,则用户很可能重复触发短语。设备检测到这些重复的声音输入相互类似并且 / 或者它们类似于触发短语(尽管没有相似到足以致使语音触发系统 400 启用基于语音的服务)。如果满足这种条件,设备则确定声音输入对应于启用语音触发系统 400 的有效尝试。因此,在一些具体实施中,语音触发系统 400 使用那些接收到的声音输入来调节语音触发系统 400 的一个或多个方面,使得用户的类似话语未来将作为有效触发器被接受。在一些具体实施中,只有在满足某些条件或条件的组合的情况下,才使用这些声音输入来调整语音触发系统 400。例如,在一些具体实施中,在连续接收到预先确定数量的声音输入(例如,2 次、3 次、4 次、5 次或任何其他适当数量)时、在声音输入足够类似于参考表示时、在声音输入相互间足够类似时、在声音输入挨得很近(例如,它们是在预先确定的时间段内接收到的和 / 或以预先确定的时间间隔或接近预先确定的时间间隔接收

到的),并且/或者这些或其他条件的任何组合的情况下,使用声音输入来调整语音触发系统 400。在一些情况下,语音触发系统 400 可检测不满足触发标准中的一者或多者的一个或多个声音输入,之后手动启动基于语音的服务(例如,通过按下按钮或图标)。在一些具体实施中,由于基于语音的服务是在接收到声音输入不久之后启动的,因此语音触发系统 400 确定声音输入实际上对应于失败的语音触发尝试。因此,如上所述,语音触发系统 400 使用那些接收到的声音输入来调节语音触发系统 400 的一个或多个方面,使得用户的话语未来将作为有效触发器被接受。

[0117] 尽管上述调整技术涉及调节参考表示,但除了或代替调节参考表示,触发声音检测技术的其他方面也可以相同或类似的方式来调节。例如,在一些具体实施中,设备对如何滤除声音输入和/或将哪些滤波器应用于声音输入进行调节,以重点关注和/或消除声音输入的某些频率或频率范围。在一些具体实施中,设备对用于比较输入表示与参考表示的算法进行调节。例如,在一些具体实施中,对用于确定输入表示与参考表示之间差异的数学函数中的一个或多个项进行变化、增加或删除,或者替换不同的数学函数。在一些具体实施中,调整技术诸如上述那些需要的资源比语音触发系统 400 能够提供或被配置为进行提供的资源更多。具体地,声音检测器可不具有或具有对用于执行迭代调整参考表示和/或声音检测算法(或语音触发系统 400 的任何其他适当方面)所必需的数量或类型的处理器、数据或存储器的访问权限。因此,在一些具体实施中,上述调整技术中的一者或多者由更强大的处理器诸如应用处理器(例如,一个或多个处理器 204)或由不同设备(例如,服务器系统 108)来执行。然而,语音触发系统 400 被设计成即使在应用处理器处于待机模式下也进行操作。因此,用于调整语音触发系统 400 的声音输入在应用处理器为不活动的并且无法处理声音输入的情况下被接收。因此,在一些具体实施中,声音输入由设备进行存储以便在其被接收之后能够得到进一步处理和/或分析。在一些具体实施中,声音输入存储在音频子系统 226 的存储缓冲器 414 中。在一些具体实施中,使用直接存储器存取(DMA)技术(包括例如使用 DMA 引擎以便复制或移动数据,而无需启动应用处理器)将声音输入存储在系统存储器(例如,存储器 250,图 2)中。然后,一旦应用处理器(或服务器系统 108 或另一适当设备)启动,则存储的声音输入就被提供至应用处理器或由应用处理器访问存储的声音输入,使得应用处理器能够执行上述调整技术中的一者或多者。在一些具体实施中,

[0118] 图 5-7 为表示根据某些具体实施的用于操作语音触发器的方法的流程图。这些方法任选地由存储在计算机存储器或非暂态计算机可读存储介质(例如,客户端设备 104 的存储器 250、与数字助理系统 300 相关联的存储器 302)中并且由数字助理系统的一个或多个计算机系统的一个或多个处理器执行的指令进行控制,该数字助理系统包括但不限于服务器系统 108 和/或用户设备 104a。计算机可读存储介质可包括磁盘或光盘存储设备、固态存储设备诸如闪存存储器、或一个或多个其他非易失性存储设备。存储在计算机可读存储介质上的计算机可读指令可包括下述的一者或多者:源代码、汇编语言代码、目标代码、或由一个或多个处理器解译的其他指令格式。在各种具体实施中,可对每个方法中的一些操作进行结合并且/或者一些操作的次序可由图中所示的次序进行改变。另外,在一些具体实施中,可将各个附图中所示的和/或结合各个方法所论述的操作相结合以形成其他方法,并且可将同一附图中所示的和/或结合同一方法所论述的操作分成不同方法。此外,在一些具体实施中,方法中的一个或多个操作由数字助理系统 300 和/或电子设备(例如,

用户设备 104) 的模块来执行,这些模块包括例如自然语言处理模块 332、对话流处理模块 334、音频子系统 226、噪声检测器 402、声音类型检测器 404、触发声音检测器 406、基于语音的服务 408 和 / 或它们的任何子模块。图 5 示出根据一些具体实施的用于操作语音触发系统 (例如,语音触发系统 400,图 4) 的方法 500。在一些具体实施中,方法 500 在包括存储器和一个或多个处理器的电子设备 (例如,电子设备 104) 上执行,该存储器存储由一个或多个处理器执行的指令。电子设备接收声音输入 (502)。声音输入可对应于口头话语 (例如,字词、短语或句子)、人类发出的声音 (例如,口哨声、啞舌声、打响指声、拍手声等) 或任何其他声音 (例如,电子产生的啁啾声、机械噪声源等)。在一些具体实施中,电子设备通过音频子系统 226 (包括例如参考图 4 所描述的编解码器 410、音频 DSP 412 和缓冲器 414,以及麦克风 230 和 418) 接收声音输入。

[0119] 在一些具体实施中,电子设备确定声音输入是否满足预先确定的条件 (504)。在一些具体实施中,电子设备对声音输入应用时域分析以确定声音输入是否满足预先确定的条件。例如,电子设备在一时间段内分析声音输入以便确定声音幅值是否达到预先确定的水平。在一些具体实施中,如果声音输入的幅值 (例如,音量) 满足和 / 或超过预先确定的阈值,则满足阈值。在一些具体实施中,如果声音输入在预先确定的时间量内满足和 / 或超过预先确定的阈值,则满足阈值。如下更详细论述的,在一些具体实施中,确定声音输入是否满足预先确定的条件 (504) 由第三声音检测器 (例如,噪声检测器 402) 来执行。(在这种情况下,第三声音检测器用于将该声音检测器与其他声音检测器 (例如,下述的第一声音检测器和第二声音检测器) 区别开来,并且不一定指示声音检测器的任何操作位置或次序)。

[0120] 电子设备确定声音输入是否对应于预先确定的声音类型 (506)。如上所述,基于声音的某些可辨识特征将声音分成不同“类型”。确定声音输入是否对应于预先确定的类型包括确定声音输入是否包括或表现出特定类型的特征。在一些具体实施中,预先确定的声音类型为人声。在此类具体实施中,确定声音输入是否对应于人声包括确定声音输入是否包括人声的频率特征 (508)。如下更详细论述的,在一些具体实施中,确定声音输入是否对应于预先确定的声音类型 (506) 由第一声音检测器 (例如,声音类型检测器 404) 来执行。在确定声音输入对应于预先确定的声音类型时,电子设备确定声音输入是否包括预先确定的内容 (510)。在一些具体实施中,预先确定的内容对应于一个或多个预先确定的音素 (512)。在一些具体实施中,一个或多个预先确定的音素构成至少一个字词。在一些具体实施中,预先确定的内容为声音 (例如,口哨声、点击声或拍手声)。在一些具体实施中,如下所述,确定声音输入是否包括预先确定的内容 (510) 由第二声音检测器 (例如,触发声音检测器 406) 来执行。

[0121] 在确定声音输入包括预先确定的内容时,电子设备启动基于语音的服务 (514)。在一些具体实施中,如上详述的,基于语音的服务为基于语音的数字助理。在一些具体实施中,基于语音的服务为听写服务,其中语音输入被转换成文本并被包括在和 / 或显示在文本输入域 (例如,电子邮件、文本消息、文字处理或笔记应用程序等的文本输入域) 中。在基于语音的服务为基于语音的数字助理的具体实施中,一旦基于语音的数字助理被启动,就会向用户发出指示用户可向数字助理提供语音输入和 / 或命令的提示 (例如,声音或语音提示)。在一些具体实施中,启动基于语音的数字助理包括启用应用处理器 (例如,一个

或多个处理器 204, 图 2)、启动一个或多个程序或模块 (例如, 数字助理客户端模块 264, 图 2), 和 / 或建立到远程服务器或设备 (例如, 数字助理服务器 106, 图 1) 的连接。

[0122] 在一些具体实施中, 电子设备确定声音输入是否对应于特定用户的语音 (516)。例如, 对声音输入应用一种或多种语音认证技术以确定其是否对应于设备的授权用户的语音。语音认证技术在上文中更详细地描述。在一些具体实施中, 语音认证由声音检测器中的一个 (例如, 触发声音检测器 406) 来执行。在一些具体实施中, 语音认证由专用语音认证模块 (包括任何适当硬件和 / 或软件) 来执行。在一些具体实施中, 基于声音的服务响应于确定声音输入包括预先确定的内容并且声音输入对应于特定用户的语音而启动。因此, 例如, 基于声音的服务 (例如, 基于语音的数字助理) 将仅在授权用户讲出触发字词或短语时才启动。这就减少了未授权用户可调用服务的可能性, 并且其在多个电子设备很靠近的情况下尤其有用, 因为一个用户的触发声音的话语将不启用另一用户的语音触发器。

[0123] 在一些具体实施中, 在基于语音的服务为基于语音的数字助理的情况下, 响应于确定声音输入包括预先确定的内容但不对应于特定用户的语音, 基于语音的数字助理以受限访问模式启动。在一些具体实施中, 受限访问模式允许数字助理仅访问数字助理能够以其他方式提供的数据、服务和 / 或功能的子集。在一些具体实施中, 受限访问模式对应于只写模式 (例如, 使得数字助理的未授权用户无法访问来自日历、任务列表、联系人、照片、电子邮件、文本消息等的的数据)。在一些具体实施中, 受限访问模式对应于基于语音的服务的沙箱实例, 使得基于语音的服务将不从用户数据读取或对用户数据写入, 该用户数据诸如设备 104 上的用户数据 266 (图 2) 或任何其他设备上的用户数据 (例如, 图 3A 的用户数据 348, 其可存储在远程服务器诸如图 1 的服务器系统 108 上)。

[0124] 在一些具体实施中, 响应于确定声音输入包括预先确定的内容并且声音输入对应于特定用户的语音, 基于语音的数字助理输出包括特定用户姓名的提示。例如, 当通过语音认证识别出特定用户时, 基于语音的数字助理可输出提示诸如“我能帮你做些什么, Peter?”, 而不是更通用的提示诸如音调、哔哔声或非个性化语音提示。如上所述, 在一些具体实施中, 第一声音检测器确定声音输入是否对应于预先确定的声音类型 (在步骤 506), 并且第二声音检测器确定声音检测器是否包括预先确定的内容 (在步骤 510)。在一些具体实施中, 第一声音检测器在操作时比第二声音检测器消耗更少的功率, 例如因为第一声音检测器使用了处理器强度比第二检测器小的技术。在一些具体实施中, 第一声音检测器为声音类型检测器 404, 并且第二声音检测器为触发声音检测器 406, 两者均相对于图 4 在上文中有所论述。在一些具体实施中, 如上参考图 4 所描述的, 在操作时, 第一声音检测器和 / 或第二声音检测器根据占空比周期性地监测音频通道。

[0125] 在一些具体实施中, 第一声音检测器和 / 或第二声音检测器执行声音输入的频域分析。例如, 这些声音检测器执行拉普拉斯变换、Z 变换或傅里叶变换以生成声音输入或其一部分的频谱或确定声音输入或其一部分的频谱密度。在一些具体实施中, 第一声音检测器为语音活动检测器, 其被配置为确定声音输入是否包括具备人声的特性 (或声音输入的具有人声特性的其他特征、方面或属性) 的频率。

[0126] 在一些具体实施中, 第二声音检测器关闭或为非活动的直至第一声音检测器检测到预先确定类型的声音输入。因此, 在一些具体实施中, 方法 500 包括响应于确定声音输入对应于预先确定的类型来启动第二声音检测器。(在其他具体实施中, 第二声音检测器响

应于其他条件进行启动,或者不管第一声音检测器确定与否都连续操作。)在一些具体实施中,启动第二声音检测器包括启用硬件和/或软件(包括例如电路、处理器、程序、存储器等)。在一些具体实施中,第二声音检测器在其启动之后操作至少预先确定的时间量(例如,为活动的并且正监测音频通道)。例如,在第一声音检测器确定声音输入对应于预先确定的类型(例如,包括人声)时,第二声音检测器进行操作以便确定声音输入是否还包括预先确定的内容(例如,触发字词)。在一些具体实施中,预先确定的时间量对应于预先确定内容的持续时间。因此,如果预先确定的内容为短语“嘿,SIRI”,则预先确定的时间量将为足以确定是否发出该短语的长度(例如,1秒或2秒,或者任何其他适当的持续时间)。如果预先确定的内容较长,诸如为短语“嘿,SIRI,请醒过来帮帮我”,则预先确定的时间将更长(例如,5秒或另一个适当的持续时间)。在一些具体实施中,只要第一声音检测器检测到对应于预先确定类型的声音,第二声音检测器就进行操作。在此类具体实施中,例如,只要第一声音检测器在声音输入中检测到人的语音,第二声音检测器便将处理声音输入以确定其是否包括预先确定的内容。

[0127] 如上所述,在一些具体实施中,第三声音检测器(例如,噪声检测器402)确定声音输入是否满足预先确定的条件(在步骤504)。在一些具体实施中,第三声音检测器在操作时比第一声音检测器消耗更少的功率。在一些具体实施中,如上相对于图4所论述的,第三声音检测器根据占空比周期性地监测音频通道。另外,在一些具体实施中,第三声音检测器执行声音输入的时域分析。在一些具体实施中,第三声音检测器比第一声音检测器消耗更少的功率是因为时域分析在处理器强度上小于第二声音检测器所应用的频域分析。

[0128] 类似于有关响应于第一声音检测器(例如,声音类型检测器404)的确定来启动第二声音检测器(例如,触发声音检测器406)的以上论述,在一些具体实施中,第一声音检测器是响应于第三声音检测器(例如,噪声检测器402)的确定进行启动的。例如,在一些具体实施中,声音类型检测器404响应于噪声检测器402确定声音输入满足预先确定的条件(例如,在足够的持续时间内超过某一音量)而启动。在一些具体实施中,启动第一声音检测器包括启用硬件和/或软件(包括例如电路、处理器、程序、存储器等)。在其他具体实施中,第一声音检测器响应于其他条件而启动,或者连续操作。在一些具体实施中,设备将声音输入的至少一部分存储在存储器中(518)。在一些具体实施中,存储器为音频子系统226的缓冲器414(图4)。存储的声音输入允许设备进行声音输入的非实时处理。例如,在一些具体实施中,声音检测器中的一者或多者读取和/或接收存储的声音输入以处理存储的声音输入。这在上游声音检测器(例如,触发声音检测器406)直到音频子系统226部分接收到声音输入才启动的情况下尤其有用。在一些具体实施中,一旦基于语音的服务启动,声音输入的存储部分就被提供至基于语音的服务(520)。因此,基于语音的服务能够转录、处理或以其他方式对声音输入的存储部分进行操作,尽管基于语音的服务直到接收到该部分声音输入才是完全可操作的。在一些具体实施中,将声音输入的存储部分提供至电子设备的调整模块。

[0129] 在各种具体实施中,步骤(516)-(520)在方法500中的不同位置处执行。例如,在一些具体实施中,步骤(516)-(520)中的一者或多者在步骤(502)和(504)之间、在步骤(510)和(514)之间、或者在任何其他适当位置处执行。

[0130] 图6示出根据一些具体实施的用于操作语音触发系统(例如,语音触发系统400,

图 4) 的方法 600。在一些具体实施中,方法 600 在包括存储器和一个或多个处理器的电子设备(例如,电子设备 104)上执行,该存储器存储由一个或多个处理器执行的指令。电子设备确定其是否处于预先确定的取向(602)。在一些具体实施中,电子设备使用光传感器(包括相机)、麦克风、接近传感器、磁性传感器、加速度计、陀螺仪、倾斜传感器等来检测其取向。例如,电子设备通过将入射到前置摄像头的传感器的光的量或亮度与入射到后置摄像头的传感器的光的量或亮度进行比较来确定其是正面向下还是正面朝上放在一表面上。如果前置摄像头检测到的量和/或亮度远远大于后置摄像头检测到的量和/或亮度,则电子设备将确定其为正面朝上。另一方面,如果后置摄像头检测到的量和/或亮度远远大于前置摄像头检测到的量和/或亮度,则设备将确定其为正面朝下。在确定电子设备处于预先确定的取向时,电子设备启用语音触发器的预先确定模式(604)。在一些具体实施中,预先确定的取向对应于设备的大体上水平且正面朝下的显示屏,并且预先确定的模式为待机模式(606)。例如,在一些具体实施中,如果智能电话或平板电脑位于桌子或办公桌上使得屏幕正面朝下,语音触发器则置于待机模式(例如,关闭)以避免语音触发器无意中启用。

[0131] 另一方面,在一些具体实施中,预先确定的取向对应于设备的大体上水平且正面朝上的显示屏,并且预先确定的模式为监听模式(608)。因此,例如,如果智能电话或平板电脑位于桌子或办公桌上使得屏幕正面朝上,语音触发器则置于监听模式以便其在检测到触发时能够对用户进行响应。

[0132] 图 7 示出了根据一些具体实施的用于操作语音触发器(例如,语音触发系统 400,图 4)的方法 700。在一些具体实施中,方法 700 在包括存储器和一个或多个处理器的电子设备(例如,电子设备 104)上执行,该存储器存储由一个或多个处理器执行的指令。电子设备在第一模式下操作语音触发器(例如,语音触发系统 400)(702)。在一些具体实施中,第一模式为正常监听模式。

[0133] 电子设备通过检测到电子设备的麦克风和相机中的一者或多者被遮挡来确定其是否处于大体上封闭的空间中(704)。在一些具体实施中,大体上封闭的空间包括口袋、钱包、书包、抽屉、手套箱、公文包等。

[0134] 如上所述,在一些具体实施中,设备通过从扬声器或换能器发出一种或多种声音(例如,音调声、点击声、撞击声等)并且监测一个或多个麦克风或换能器来检测发出的一种或多种声音的回声,检测到麦克风被遮挡。例如,相对较大环境(例如,房间或车辆)将反射不同于相对较小的大体上封闭环境(例如,钱包或口袋)的声音。因此,如果设备基于回声(或没有回声)检测到麦克风(或发出声音的扬声器)被遮挡,设备便确定其处于大体上封闭的空间中。在一些具体实施中,设备通过检测麦克风正在拾取具备封闭空间特征的声音而检测到麦克风被遮挡。例如,当设备放在口袋中时,由于麦克风接触到或接近于口袋的织物,因此麦克风可检测特有的沙沙作响的噪声。在一些具体实施中,设备基于传感器所接收的光照水平,或者通过确定其是否能够获得聚焦图像来检测相机被遮挡。例如,如果相机传感器在预期较高光照水平的一时间内(例如,在白天期间)检测到较低光照水平,设备便确定相机被遮挡,并且设备处于大体上封闭的空间中。又如,相机可尝试通过其传感器获得聚焦图像。通常,在相机处于极其黑暗的地方(例如,口袋或背包)或在其过于靠近试图聚焦于其上的物体(例如,钱包或背包的内部)的情况下,这将是困难的。因此,如果相机无法获得聚焦图像,则确定设备处于大体上封闭的空间中。

[0135] 在确定电子设备处于大体上封闭的空间中时,电子设备将语音触发器切换至第二模式(706)。在一些具体实施中,第二模式为待机模式(708)。在一些具体实施中,在处于待机模式时,语音触发系统400将继续监测环境音频,但将不对所接收的声音进行响应,不管它们是否将以其他方式触发语音触发系统400。在一些具体实施中,在待机模式下,语音触发系统400被停用,并且不处理用于检测触发声音的音频。在一些具体实施中,第二模式包括根据不同于第一模式的占空比来操作语音触发系统400的一个或多个声音检测器。在一些具体实施中,第二模式包括操作声音检测器的不同于第一模式的组合。

[0136] 在一些具体实施中,第二模式对应于较敏感的监测模式,使得语音触发系统400即使处于大体上封闭的空间中也能够检测到触发声音并对其进行响应。在一些具体实施中,一旦语音触发器切换到第二模式,设备就通过检测电子设备的麦克风和相机中的一者或多者是否被遮挡(例如,使用上文相对于步骤(704)所描述的技术中的任一种)来周期性地确定该电子设备是否仍处于大体上封闭的空间中。如果设备仍处于大体上封闭的空间中,则语音触发系统400将保持在第二模式。在一些具体实施中,如果设备离开大体上封闭的空间,则电子设备将使语音触发器返回至第一模式。

[0137] 根据一些具体实施,图8示出根据如上所述的本发明的原理进行配置的电子设备800的功能框图。设备的功能块可以由硬件、软件、或者硬件和软件的组合来实现,以执行本发明的原理。本领域的技术人员能够理解,图8中所述的功能块可被组合或者被分离为子块,以实现如上所述的本发明的原理。因此,本文中的描述可支持本文所述功能块的任何可能的组合或分离或进一步限定。

[0138] 如图8所示,电子设备800包括被配置为接收声音输入的声音接收单元802。电子设备800还包括耦接至语音接收单元802的处理单元806。在一些实施例中,处理单元806包括噪声检测单元808、声音类型检测单元810、触发声音单元812、服务启动单元814,以及语音认证单元816。在一些具体实施中,噪声检测单元808对应于上述噪声检测器402,并且被配置为执行上文中参考噪声检测器402所描述的任何操作。在一些具体实施中,声音类型检测单元810对应于上述声音类型检测器404,并且被配置为执行上文中参考声音类型检测器404所描述的任何操作。在一些具体实施中,触发声音检测单元812对应于上述触发声音检测器406,并且被配置为执行上文中参考触发声音检测器406所描述的任何操作。在一些具体实施中,语音认证单元816对应于上述语音认证模块428,并且被配置为执行上文中参考语音认证模块428所描述的任何操作。处理单元806被配置为:确定声音输入的至少一部分是否对应于预先确定的声音类型(例如,利用声音类型检测单元810);在确定声音输入的至少一部分对应于预先确定的类型时,确定声音输入是否包括预先确定的内容(例如,利用触发声音检测单元812);并且在确定声音输入包括预先确定的内容时,启动基于语音的服务(例如,利用服务启动单元814)。

[0139] 在一些具体实施中,处理单元806还被配置为在确定声音输入是否对应于预先确定的声音类型之前,确定声音输入是否满足预先确定的条件(例如,利用噪声检测单元808)。在一些具体实施中,处理单元806还被配置为确定声音输入是否对应于特定用户的语音(例如,利用语音认证单元816)。

[0140] 根据一些具体实施,图9示出了根据如上所述的本发明的原理进行配置的电子设备900的功能框图。设备的功能块可以由硬件、软件、或者硬件和软件的组合实现,以执行

本发明的原理。本领域的技术人员应当理解,图 9 中描述的功能块可被组合或者被分离为子块,以实现如上所述的本发明的原理。因此,本文中的描述可支持本文所述功能块的任何可能的组合或分离或进一步限定。

[0141] 如图 9 所示,电子设备 900 包括语音触发单元 902。语音触发单元 902 可在不同模式下进行操作。在第一模式下,语音触发单元接收声音输入并确定它们是否满足某些标准(例如,监听模式)。在第二模式下,语音触发单元 902 不接收和 / 或不处理声音输入(例如,待机模式)。电子设备 900 还包括耦接至语音触发单元 902 的处理单元 906。在一些具体实施中,处理单元 906 包括环境检测单元 908 和模式切换单元 910,其中环境检测单元可包括一个或多个传感器(例如,包括麦克风、相机、加速度计、陀螺仪等)和 / 或与该一个或多个传感器进行交互。在一些具体实施中,处理单元 906 被配置为:通过检测到电子设备的麦克风和相机中的一者或多者被遮挡(例如,利用环境检测单元 908)来确定电子设备是否处于大体上封闭的空间中;并且在确定电子设备处于大体上封闭的空间中时,将语音触发器切换至第二模式(例如,利用模式切换单元 910)。

[0142] 在一些具体实施中,处理单元被配置为:确定电子设备是否处于预先确定的取向(例如,利用环境检测单元 908);并且在确定电子设备处于预先确定的取向时,启用语音触发器的预先确定模式(例如,利用模式切换单元 910)。

[0143] 根据一些具体实施,图 10 示出根据如上所述的本发明的原理进行配置的电子设备 1000 的功能框图。设备的功能块可以由硬件、软件、或者硬件和软件的组合实现,以执行本发明的原理。本领域的技术人员能够理解,图 10 中所述的功能块可被组合或者被分离为子块,以实现如上所述的本发明的原理。因此,本文中的描述可支持本文所述功能块的任何可能的组合或分离或进一步限定。

[0144] 如图 10 所示,电子设备 1000 包括语音触发单元 1002。语音触发单元 1002 可在不同模式下进行操作。在第一模式下,语音触发单元接收声音输入并确定它们是否满足某些标准(例如,监听模式)。在第二模式下,语音触发单元 1002 不接收和 / 或不处理声音输入(例如,待机模式)。电子设备 1000 还包括耦接至语音触发单元 1002 的处理单元 1006。在一些具体实施中,处理单元 1006 包括环境检测单元 1008 和模式切换单元 1010,其中环境检测单元可包括麦克风和 / 或相机并且 / 或者与麦克风和 / 或相机进行交互。

[0145] 处理单元 1006 被配置为:通过检测到电子设备的麦克风和相机中的一者或多者被遮挡(例如,利用环境检测单元 1008)来确定电子设备是否处于大体上封闭的空间中;并且在确定电子设备处于大体上封闭的空间中时,将语音触发器切换至第二模式(例如,利用模式切换单元 1010)。为了解释的目的,前面的描述是通过参考特定具体实施来进行描述的。然而,上面的示例性的讨论并非意图是详尽的,也并非意图要将所公开的具体实施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。选择并描述具体实施是为了最好地说明所公开思路的原理和实际应用,以由此使得本领域的其他技术人员能够最好地利用具有适合于所构想的特定用途的各种修改的所公开思路的原理和实际应用。

[0146] 应当理解,虽然术语“第一”、“第二”等可能在本文中用来描述各种元素,但是这些元素不应当被这些术语限定。这些术语只是用来将一个元素与另一元素区分开。例如,在不改变描述意图的情况下,第一声音检测器可被称为第二声音检测器,并且类似地,第二声

音检测器可被称为第一声音检测器,只要所有“第一声音检测器”出现之处统一重命名并且所有“第二声音检测器”出现之处统一重命名即可。第一声音检测器和第二声音检测器均为声音检测器,但它们并非相同的声音检测器。

[0147] 本文中所使用的术语仅仅是为了描述特定具体实施并非旨在对权利要求进行限制。如具体实施的描述和所附权利要求中所使用的那样,单数形式的“一个”、“一种”和“这个”旨在也涵盖复数形式,除非上下文清楚地有相反指示。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地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。还应当理解,术语“包括”和 / 或“包含”当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、元素和 / 或部件,但是并不排除存在或添加一个或多个其它特征、整数、步骤、操作、元素、部件和 / 或其分组。如本文所用,术语“如果”可根据上下文被理解为意思是“当”或“在 . . . . . 时”或“响应于确定”或“根据确定”或“响应于检测到”陈述的先决条件属实。类似地,短语“如果确定 [ 陈述的先决条件属实 ]”或“如果 [ 陈述的先决条件属实 ]”或“当 [ 陈述的先决条件属实 ]”可根据上下文被理解为意思是“在确定时”或“在确定了 . . . . . 时”或“响应于确定”或“根据确定”或“在检测到 . . . . . 时”或“响应于检测到”陈述的先决条件属实。

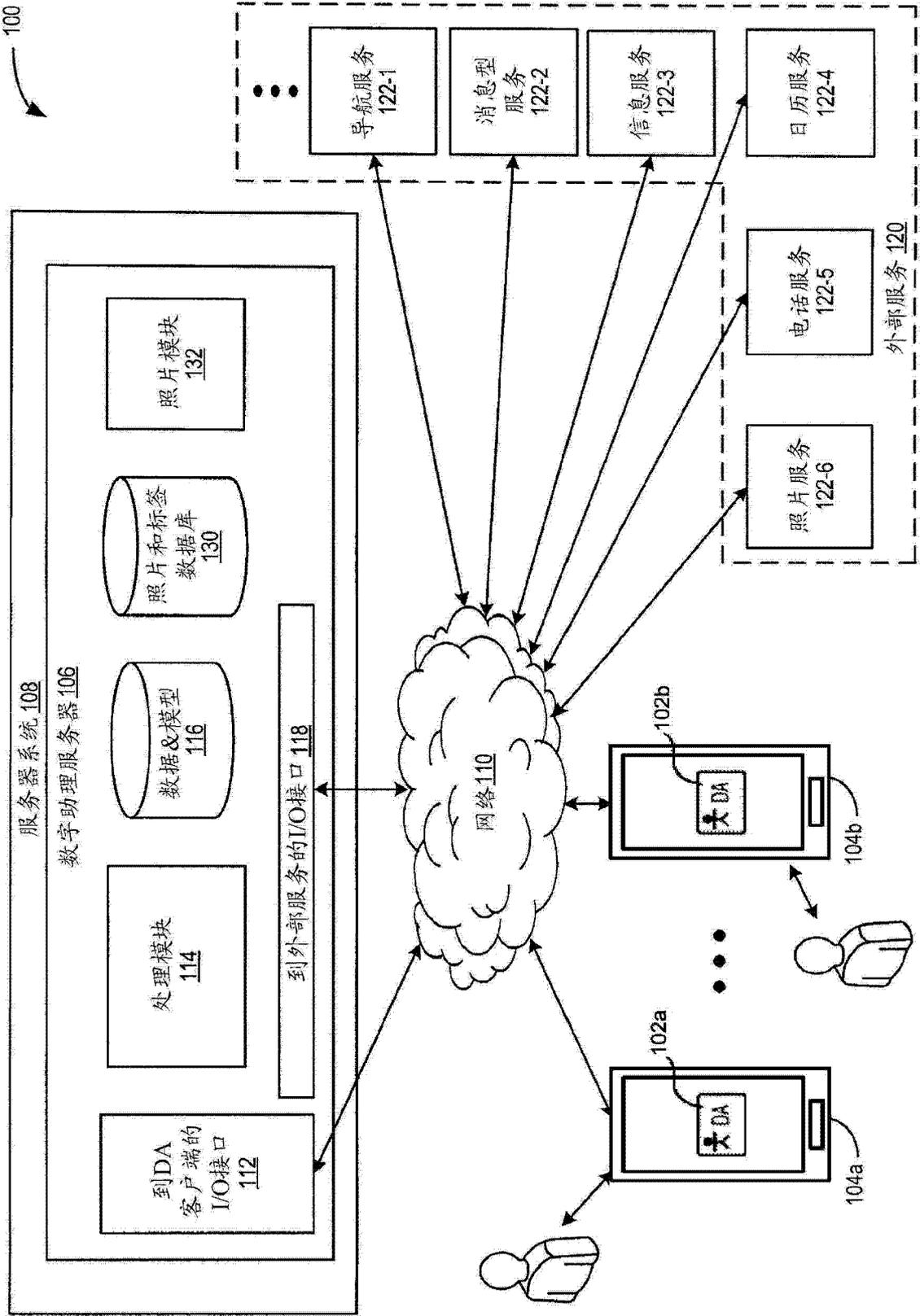


图 1

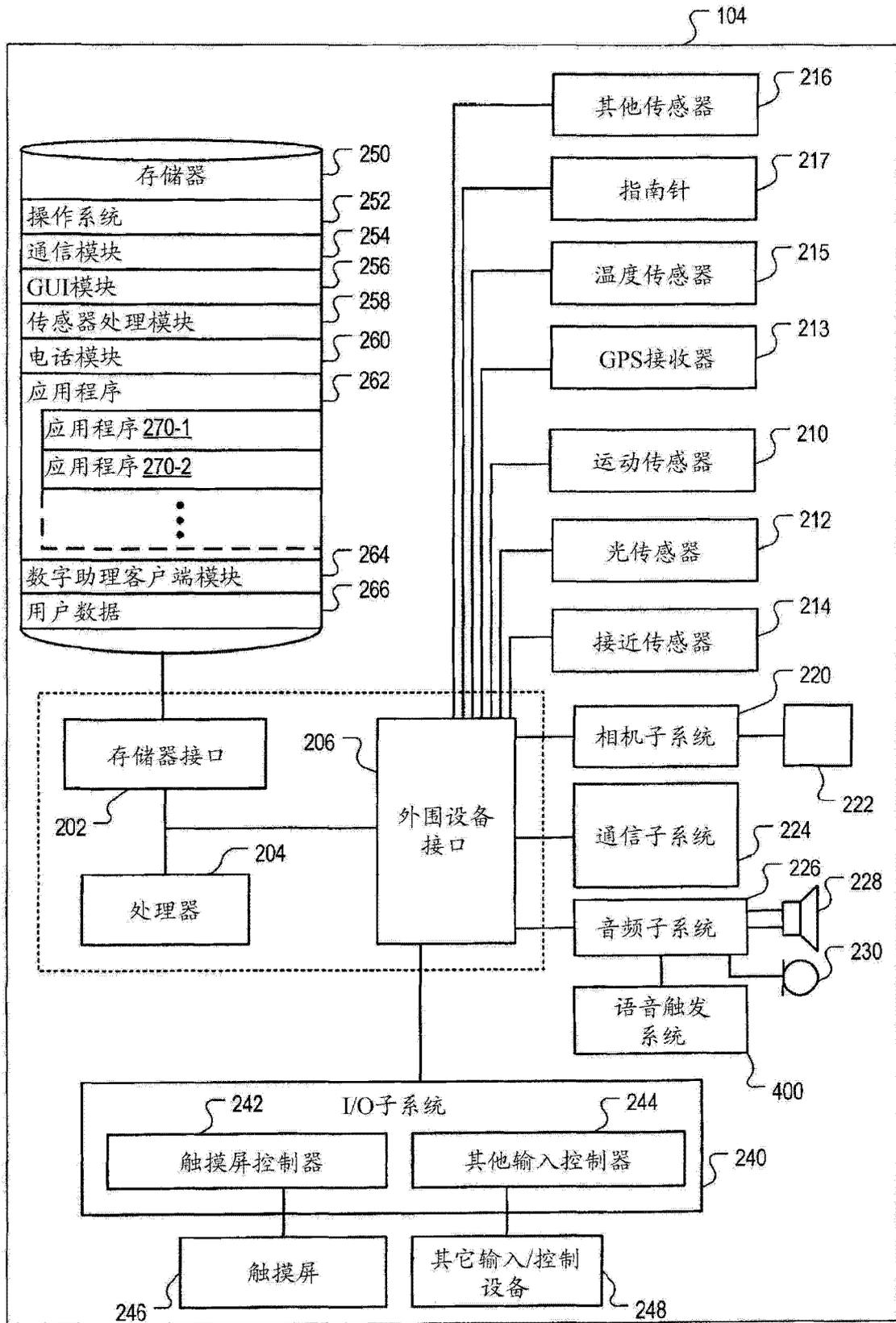


图 2

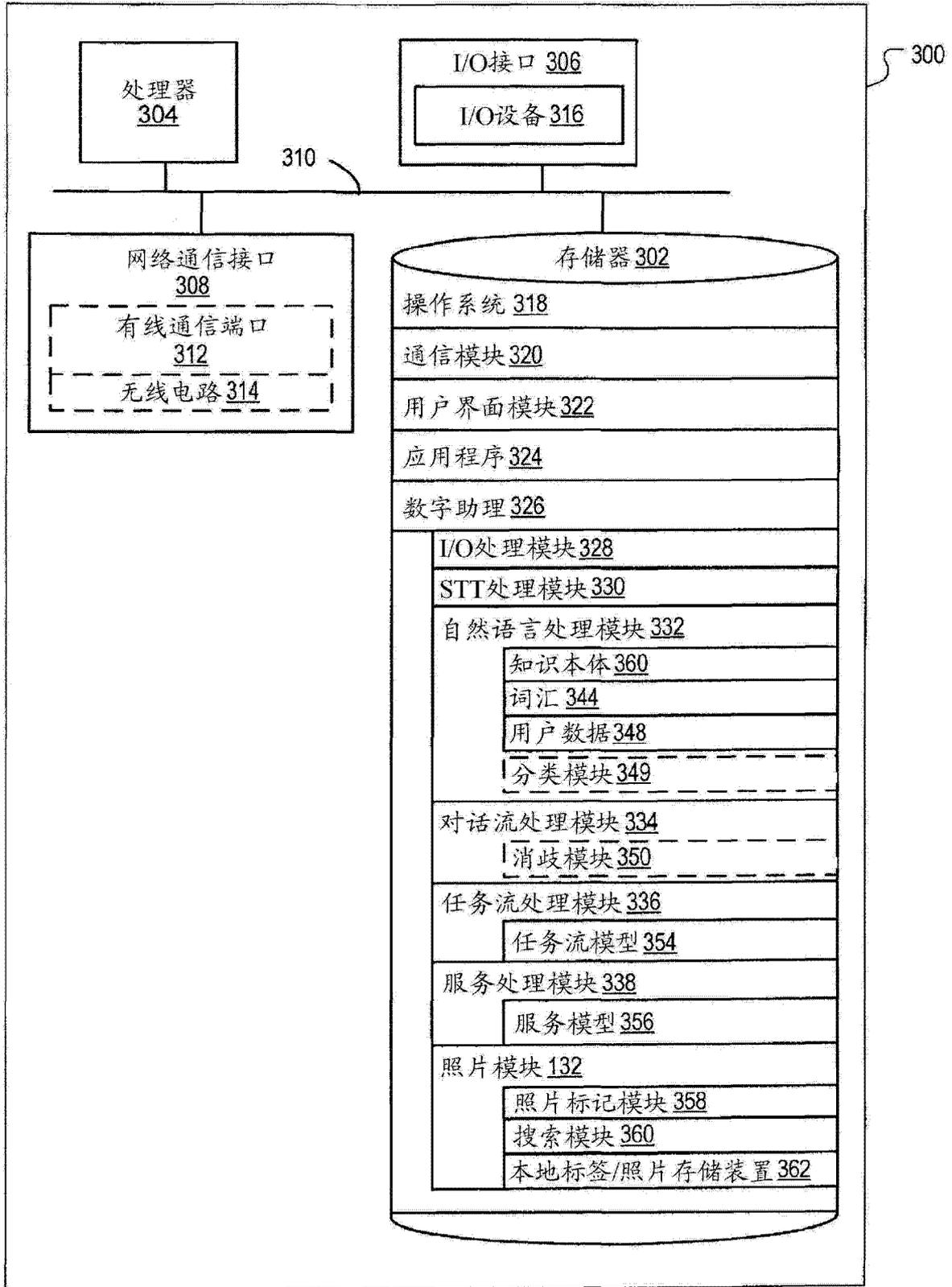


图 3A

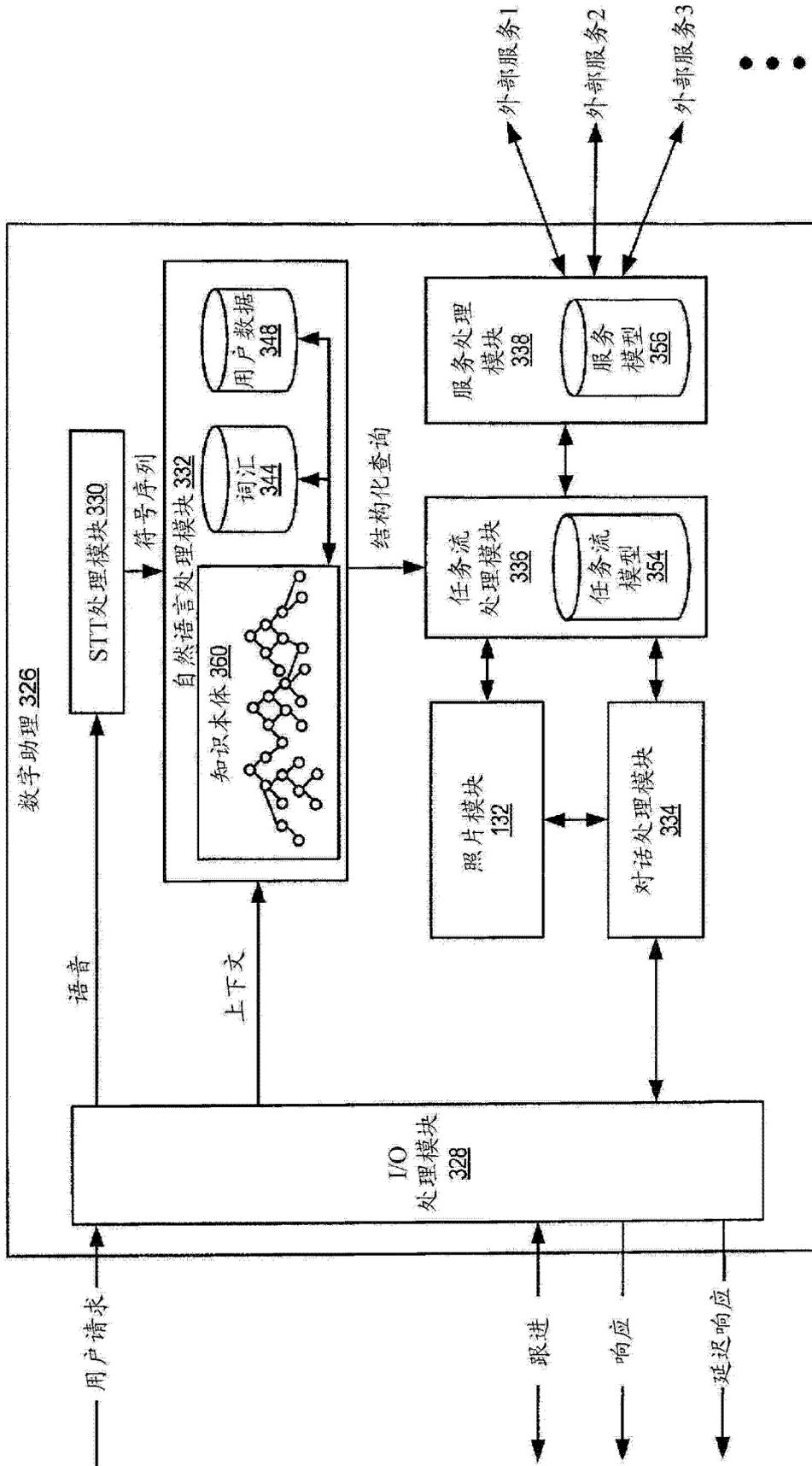


图 3B

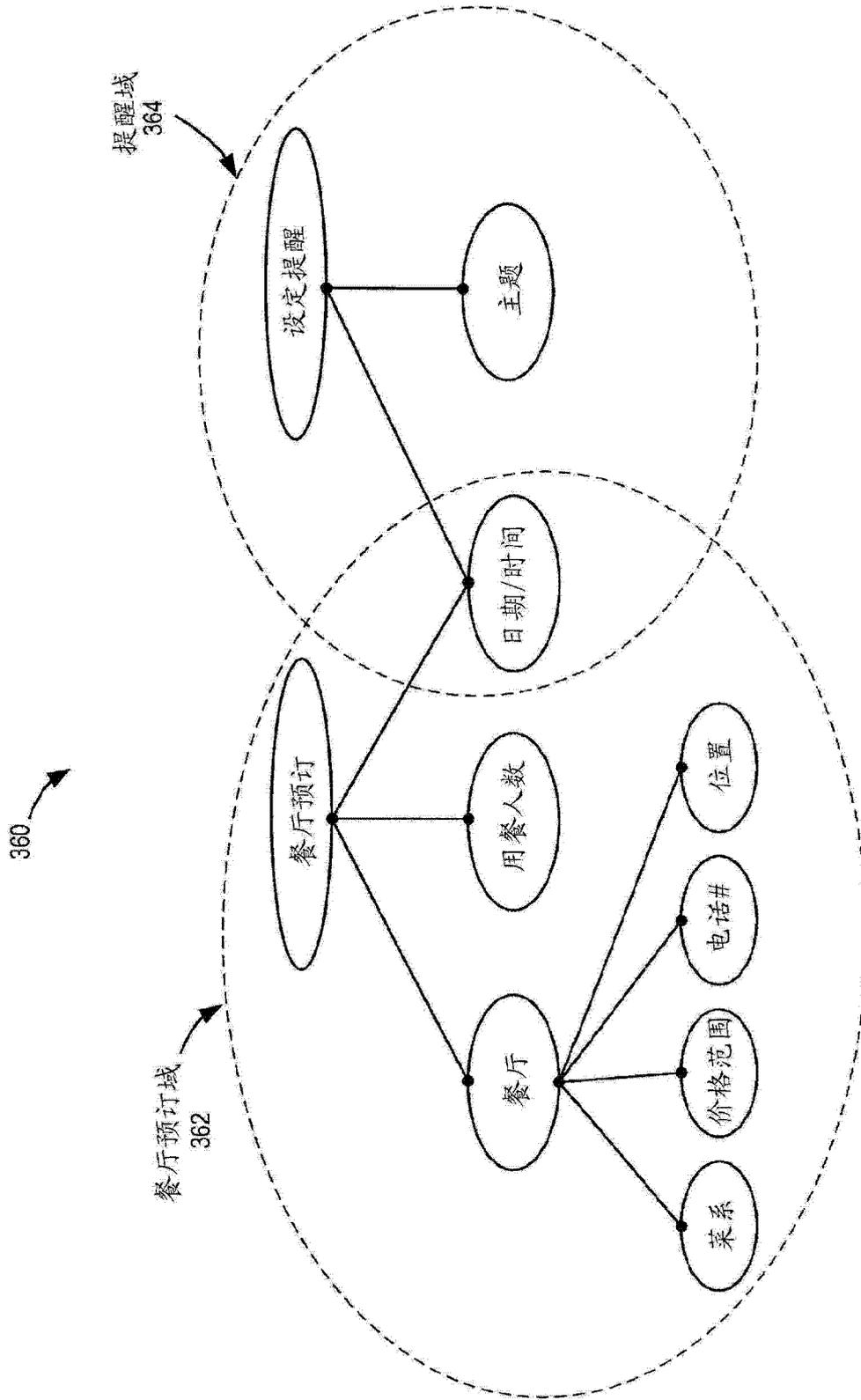


图 3C

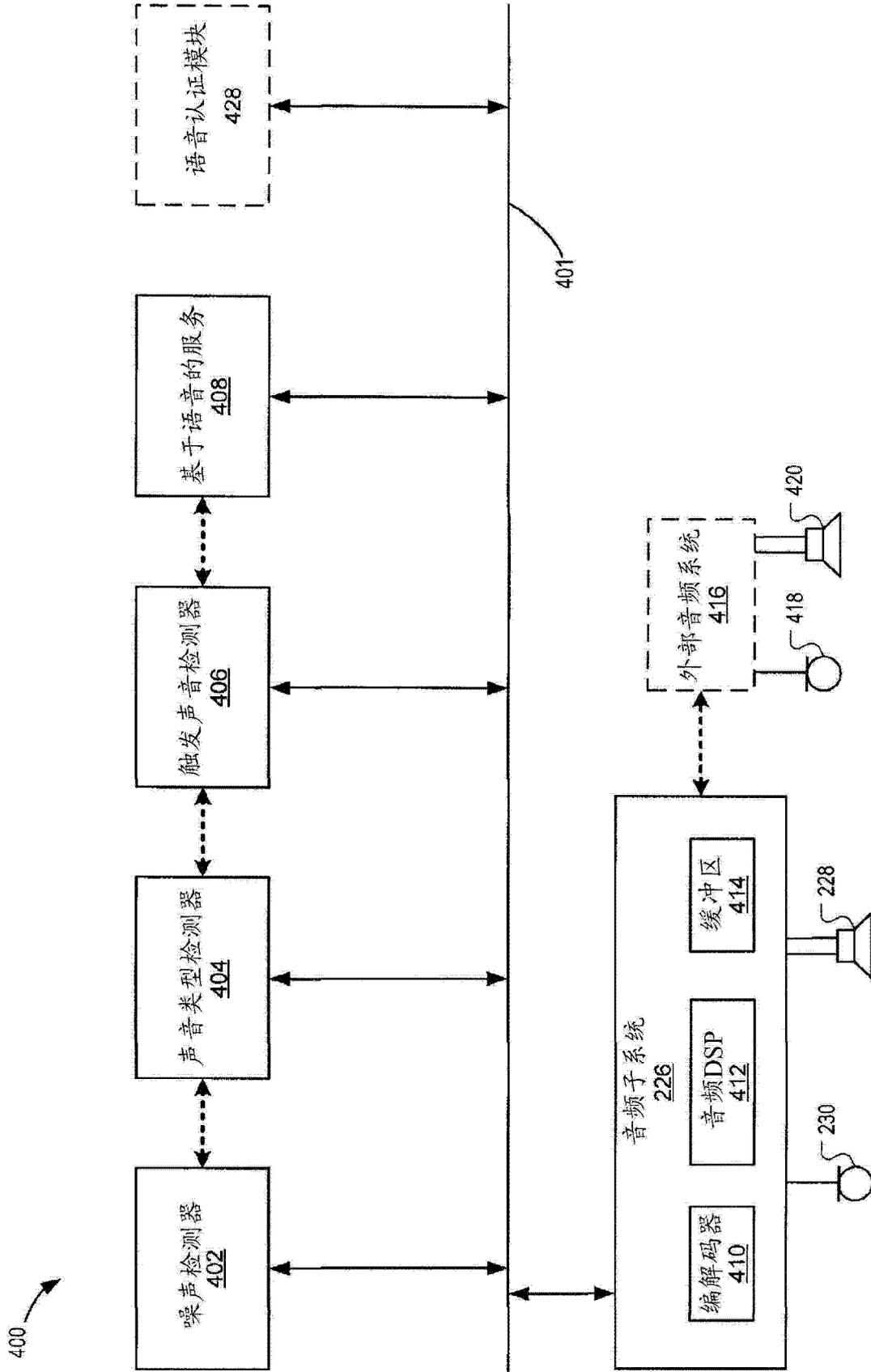


图 4

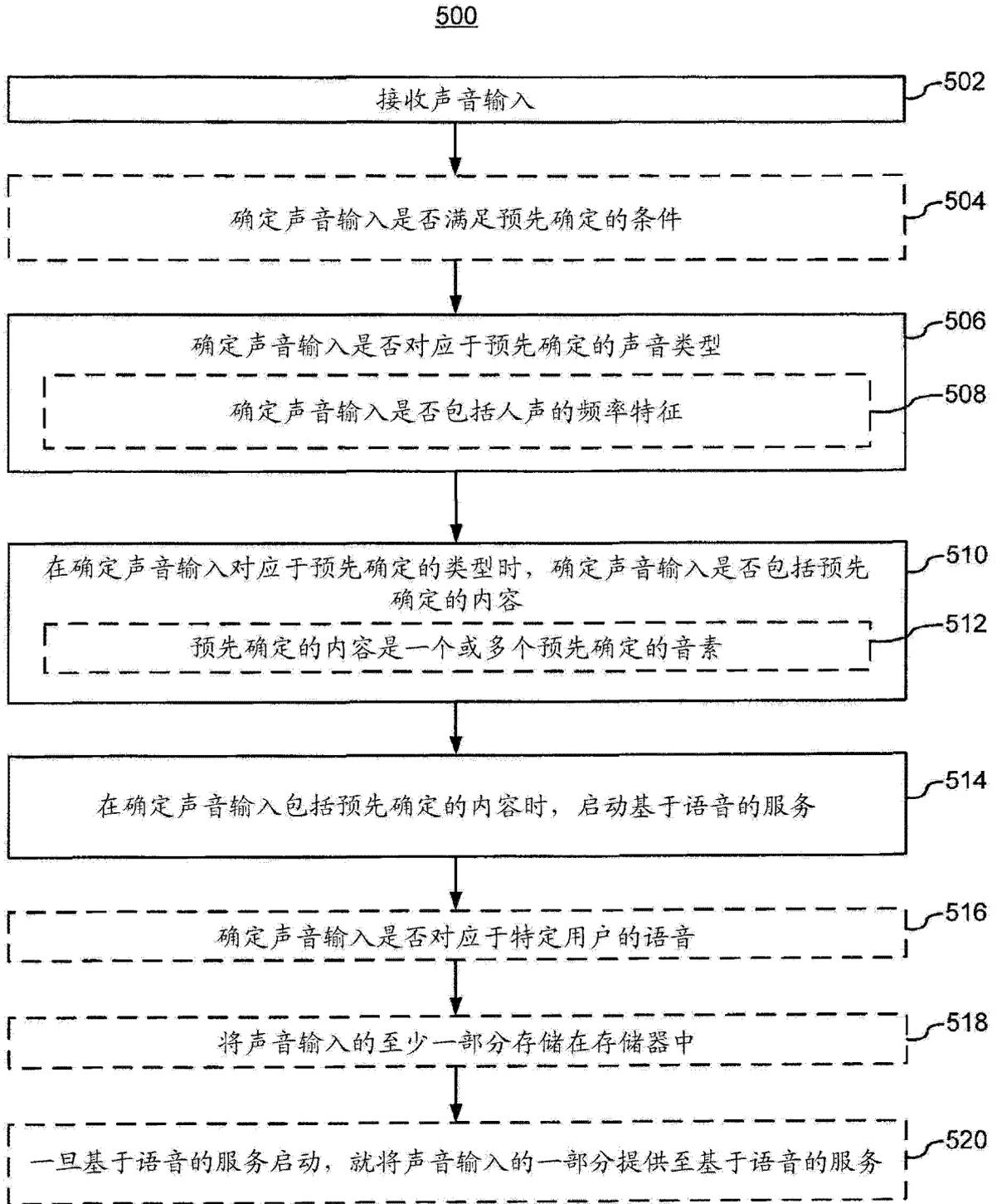


图 5

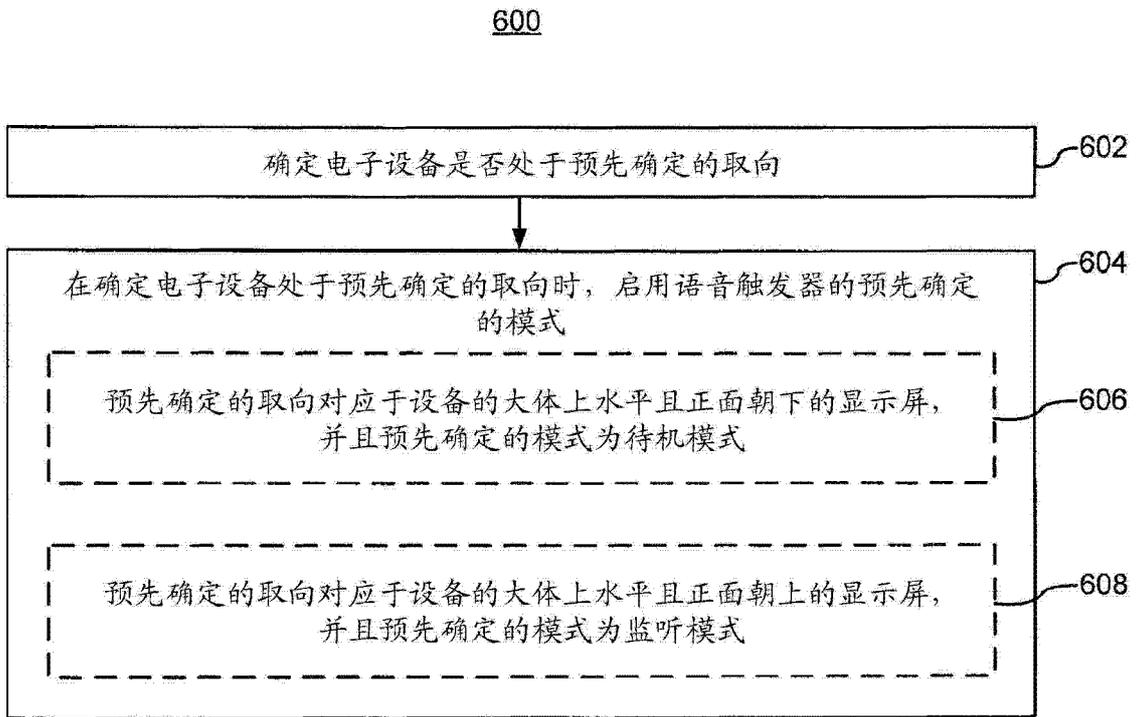


图 6

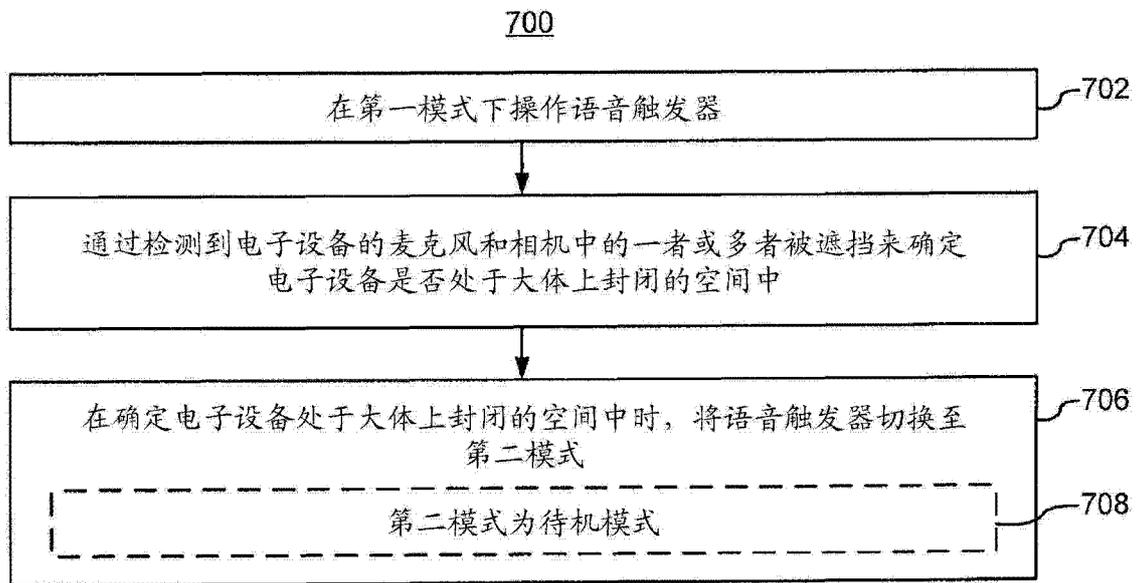


图 7

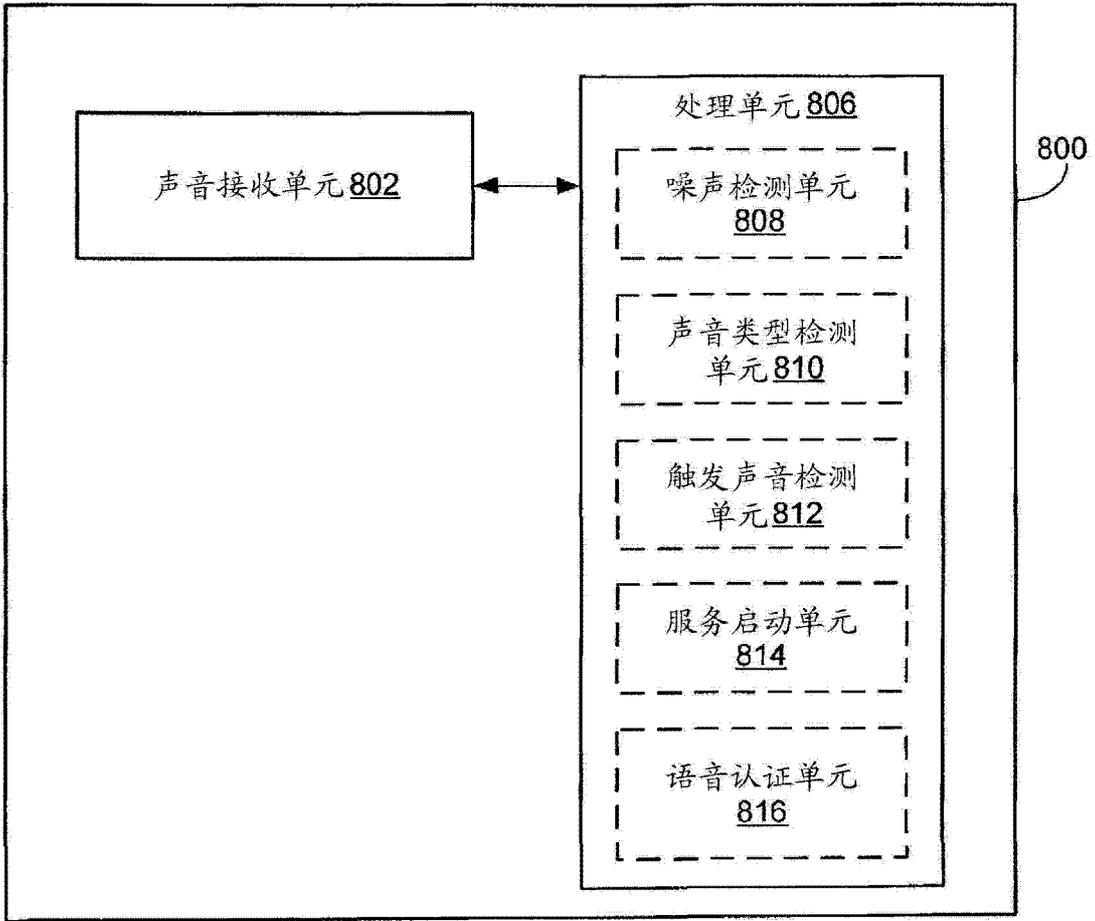


图 8

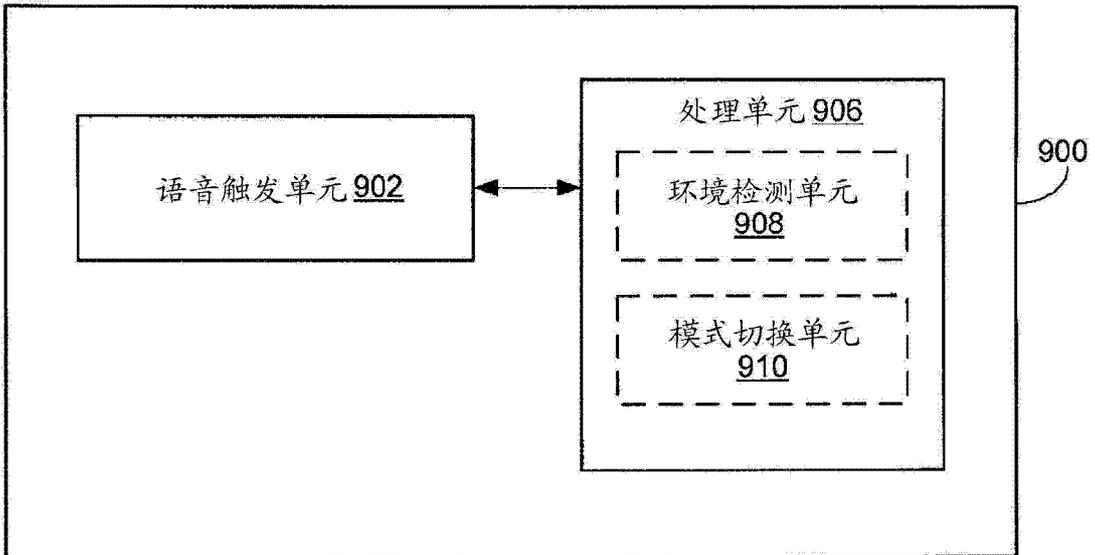


图 9