



## (12)发明专利

(10)授权公告号 CN 107146606 B

(45)授权公告日 2020.11.10

(21)申请号 201611235770.X

(22)申请日 2016.12.28

(65)同一申请的已公布的文献号

申请公布号 CN 107146606 A

(43)申请公布日 2017.09.08

(30)优先权数据

15/057,453 2016.03.01 US

(73)专利权人 谷歌有限责任公司

地址 美国加利福尼亚州

(72)发明人 王波 苏尼尔·韦缪里

巴纳比·约翰·詹姆斯

斯科特·B·赫夫曼

普拉维尔·库马尔·古普塔

(74)专利代理机构 中原信达知识产权代理有限

责任公司 11219

代理人 孙志湧 周亚荣

(51)Int.Cl.

G10L 15/22(2006.01)

G10L 15/26(2006.01)

G10L 21/06(2013.01)

G10L 21/10(2013.01)

G06F 3/16(2006.01)

(56)对比文件

US 2002/0133354 A1,2002.09.19

US 2006/0106614 A1,2006.05.18

US 2011/0131045 A1,2011.06.02

US 2008/0133244 A1,2008.06.05

US 8165886 B1,2012.04.24

US 2002/0133354 A1,2002.09.19

CN 105336326 A,2016.02.17

审查员 李海龙

权利要求书4页 说明书23页 附图4页

(54)发明名称

开发人员语音动作系统

(57)摘要

本发明涉及一种计算机实现的方法、一种系统以及一种非瞬时计算机可读存储设备。方法、系统和装置,用于由语音动作系统接收指定用于不同于该语音动作系统的新的语音动作的数据。至少基于该数据,生成用于该应用的语音动作意图,其中,语音动作意图包括当由该应用接收时,请求该应用执行为该新语音动作指定的一个或多个操作的数据。语音动作意图与对该新语音动作指定的触发项相关联。语音动作系统被配置为接收由已安装该应用的设备获得的用户话语的指示,并且确定用户话语的转录抄本对应于与该语音动作意图相关联的触发项。响应于该确定,语音动作系统将语音动作意图提供给该设备。



1. 一种用于生成语音动作的计算机实现的方法,包括:

由语音动作系统接收定义对安装在一个或多个设备上的软件应用当前不存在的新语音动作的数据,所述软件应用不同于所述语音动作系统,所述数据指示用于所述软件应用执行所述新语音动作的一个或多个操作和用于触发所述新语音动作的一个或多个触发项,其中,定义所述新语音动作的数据指定上下文,所述上下文指定用户设备或安装在所述用户设备上的所述软件应用的状态;

至少基于定义所述新语音动作的数据,由所述语音动作系统生成语音动作无源数据结构,其中,所述语音动作无源数据结构包括当被所述软件应用接收时,使得所述软件应用执行用于执行所述新语音动作的一个或多个操作的数据;

由所述语音动作系统将所述语音动作无源数据结构与所述上下文且与用于触发所述新语音动作的所述一个或多个触发项相关联,其中,多个语音动作无源数据结构被定义在所述语音动作系统中;

由所述语音动作系统接收 (i) 由所述用户设备获得的用户命令话语,所述用户设备安装有所述软件应用,以及 (ii) 关于所述用户设备的当前上下文信息;

使用所述当前上下文信息而不是所述用户命令话语的转录抄本从所述语音动作系统的所述多个语音动作无源数据结构中识别候选语音动作无源数据结构集合,所述候选语音动作无源数据结构集合包括由所述数据定义的所述语音动作无源数据结构并且基于与所述候选语音动作无源数据结构集合相关联的相应上下文来识别;

通过将所述用户命令话语的所述转录抄本与所述候选语音动作无源数据结构集合中的相应候选语音动作无源数据结构的触发项进行比较来缩小所识别的候选语音动作无源数据结构集合;

由所述语音动作系统确定所述用户命令话语的所述转录抄本对应于与所述语音动作无源数据结构相关联的所述一个或多个触发项;以及

响应于所述确定,由所述语音动作系统将所述语音动作无源数据结构提供给与所述语音动作系统远离的所述用户设备,由此使得安装在所述用户设备上的软件应用执行用于执行所述新语音动作的所述一个或多个操作。

2. 如权利要求1所述的计算机实现的方法,其中,所述新语音动作是所述软件应用未被编程以支持的启用语音的命令。

3. 如权利要求1所述的计算机实现的方法,其中,接收定义所述新语音动作的所述数据包括从发布了所述软件应用的开发人员接收所述数据。

4. 如权利要求1所述的计算机实现的方法,其中,所述语音动作系统从安装在所述用户设备上的所述软件应用未接收到定义所述新语音动作的所述数据。

5. 如权利要求1所述的计算机实现的方法,其中,所述上下文指定所述软件应用正执行的特定活动处于特定活动状态。

6. 如权利要求1所述的计算机实现的方法,包括:

由所述语音动作系统,确定所述上下文信息满足所述上下文;以及

其中,响应于确定所述用户命令话语的所述转录抄本对应于与所述语音动作无源数据结构相关联的所述一个或多个触发项并且所述上下文信息满足所述上下文,所述语音动作系统将所述语音动作无源数据结构提供给所述用户设备。

7. 如权利要求6所述的计算机实现的方法,其中,接收指示所述用户设备或安装在所述用户设备上的所述软件应用的状态的当前上下文信息包括:

由所述语音动作系统将用于特定上下文信息的请求提供给所述用户设备;以及  
响应于所述请求,由所述语音动作系统接收所述特定当前上下文信息。

8. 如权利要求6所述的计算机实现的方法,包括:

由所述语音动作系统确定所述当前上下文信息满足用于第二语音动作的上下文,并且所述用户命令话语的所述转录抄本对应于与用于所述第二语音动作的语音动作无源数据结构相关联的所述一个或多个触发项,其中,用于所述第二语音动作的所述语音动作无源数据结构包括当由与所述第二语音动作相关联的软件应用接收时使得与所述第二语音动作相关联的所述软件应用执行用于执行所述第二语音动作的一个或多个操作的数据;

响应于所述确定,由所述语音动作系统从所述新语音动作和所述第二语音动作中选择语音动作;以及

由所述语音动作系统将与所述选择的语音动作相关联的语音动作无源数据结构提供给所述用户设备,由此使得安装在所述用户设备上的软件应用执行用于执行所述选择的语音动作的一个或多个操作。

9. 如权利要求8所述的计算机实现的方法,其中,从所述新语音动作和所述第二语音动作中选择所述选择的语音动作包括响应于接收指示用户选择所述新语音动作或所述第二语音动作中的一个的数据,选择所述选择的语音动作。

10. 如权利要求8所述的计算机实现的方法,其中,从所述新语音动作和所述第二语音动作选择所述选择的语音动作包括:

向所述新语音动作和所述第二语音动作中的每一个分配得分;以及

至少基于分配给所述新语音动作和所述第二语音动作的每一个的得分,选择所述选择的语音动作。

11. 如权利要求8所述的计算机实现的方法,其中,从所述新语音动作和所述第二语音动作中选择所述选择的语音动作包括响应于确定与所述选择的语音动作相关联的软件应用正在前台中操作,选择所述选择的语音动作。

12. 如权利要求1所述的计算机实现的方法,其中,生成所述语音动作无源数据结构包括确定用于执行所述新语音动作的一个或多个操作能够由所述软件应用执行。

13. 如权利要求1所述的计算机实现的方法,包括:

由所述语音动作系统确定所述用户命令话语的所述转录抄本与所述语音动作无源数据结构相关联的一个或多个触发项类似;

响应于所述确定,由所述语音动作系统将指示确认所述用户命令话语是否对应于与所述语音动作无源数据结构相关联的所述一个或多个触发项,或旨在使所述软件应用执行所述新语音动作的用于用户输入的请求的数据提供给所述用户设备;

响应于所述请求,由所述语音动作系统并且从所述用户设备接收指示确认的数据;以及

响应于接收指示所述确认的数据,由所述语音动作系统将所述语音动作无源数据结构提供给所述用户设备,由此使得安装在所述用户设备上的软件应用执行用于执行所述新语音动作的一个或多个操作。

14. 如权利要求1所述的计算机实现的方法,包括:

由所述语音动作系统接收部署所述新语音动作的请求;以及

响应于所述请求,由所述语音动作系统部署所述新语音动作,其中,部署所述新语音动作启用触发所述新语音动作。

15. 如权利要求1所述的计算机实现的方法,包括:

由所述语音动作系统接收撤销部署所述新语音动作的请求;以及

响应于所述请求,由所述语音动作系统撤销部署所述新语音动作,其中,撤销部署所述新语音动作禁用触发所述新语音动作。

16. 如权利要求1所述的计算机实现的方法,包括:

由所述语音动作系统接收启用所述新语音动作的测试的请求,其中,所述请求指定应当为其启用所述新语音动作的一个或多个设备;以及

响应于所述请求,由所述语音动作系统启用触发用于一个或多个指定的设备的新语音动作,其中,对不包括在所述指定的一个或多个设备中的设备,禁用触发所述新语音动作。

17. 如权利要求1所述的计算机实现的方法,进一步包括:

基于所述数据确定所述新语音动作对所述软件应用是否有效;以及

基于所述新语音动作对所述软件应用是有效的所述确定,引入所述新语音动作以生成所述语音动作无源数据结构。

18. 如权利要求1所述的计算机实现的方法,其中,所述多个语音动作无源数据结构包括当建立所述软件应用时由至少第一应用开发人员提交的内置语音动作以及在建立所述软件应用之后由至少第二应用开发人员提交的其它语音动作。

19. 如权利要求18所述的计算机实现的方法,其中,所述多个语音动作无源数据结构进一步包括由所述设备中的一个或多个上的操作系统默认支持的专用语音动作。

20. 一种用于生成语音动作的系统,包括:

一个或多个计算机和一个或多个存储设备,所述一个或多个存储设备存储当由所述一个或多个计算机执行时,使得所述一个或多个计算机执行包括下述的操作的指令:

由语音动作系统接收定义对安装在一个或多个设备上的软件应用当前不存在的新语音动作的数据,所述软件应用不同于所述语音动作系统,所述数据指示用于所述软件应用执行所述新语音动作的一个或多个操作和用于触发所述新语音动作的一个或多个触发项,其中,定义所述新语音动作的数据指定上下文,所述上下文指定用户设备或安装在所述用户设备上的所述软件应用的状态;

至少基于定义所述新语音动作的数据,由所述语音动作系统生成语音动作无源数据结构,其中,所述语音动作无源数据结构包括当被所述软件应用接收时,使得所述软件应用执行用于执行所述新语音动作的一个或多个操作的数据;

由所述语音动作系统将所述语音动作无源数据结构与所述上下文且与用于触发所述新语音动作的所述一个或多个触发项相关联,其中,多个语音动作无源数据结构被定义在所述语音动作系统中;

由所述语音动作系统接收 (i) 由所述用户设备获得的用户命令话语,所述用户设备安装有所述软件应用,以及 (ii) 关于所述用户设备的当前上下文信息;

使用所述当前上下文信息而不是所述用户命令话语的转录抄本从所述语音动作系统

的所述多个语音动作无源数据结构中识别候选语音动作无源数据结构集合,所述候选语音动作无源数据结构集合包括由所述数据定义的所述语音动作无源数据结构并且基于与所述候选语音动作无源数据结构集合相关联的相应上下文来识别;

通过将所述用户命令话语的所述转录抄本与所述候选语音动作无源数据结构集合中的相应候选语音动作无源数据结构的触发项进行比较来缩小所识别的候选语音动作无源数据结构集合;

由所述语音动作系统确定所述用户命令话语的所述转录抄本对应于与所述语音动作无源数据结构相关联的一个或多个触发项;以及

响应于所述确定,由所述语音动作系统将所述语音动作无源数据结构提供给与所述语音动作系统远离的所述用户设备,由此使得安装在所述用户设备上的软件应用执行用于执行所述新语音动作的所述一个或多个操作。

21. 一种非瞬时计算机可读存储设备,所述设备存储指令,所述指令在由一个或多个计算机执行时,使得所述一个或多个计算机执行包括下述的操作:

由语音动作系统接收定义对安装在一个或多个设备上的软件应用当前不存在的新语音动作的数据,所述软件应用不同于所述语音动作系统,所述数据指示用于所述软件应用执行所述新语音动作的一个或多个操作和用于触发所述新语音动作的一个或多个触发项,其中,定义所述新语音动作的数据指定上下文,所述上下文指定用户设备或安装在所述用户设备上的所述软件应用的状态;

至少基于定义所述新语音动作的数据,由所述语音动作系统生成语音动作无源数据结构,其中,所述语音动作无源数据结构包括当被所述软件应用接收时,使得所述软件应用执行用于执行所述新语音动作的一个或多个操作的数据;

由所述语音动作系统将所述语音动作无源数据结构与所述上下文且与用于触发所述新语音动作的所述一个或多个触发项相关联,其中,多个语音动作无源数据结构被定义在所述语音动作系统中;

由所述语音动作系统接收 (i) 由所述用户设备获得的用户命令话语,所述用户设备安装有所述软件应用,以及 (ii) 关于所述用户设备的当前上下文信息;

使用所述当前上下文信息而不是所述用户命令话语的转录抄本从所述语音动作系统的所述多个语音动作无源数据结构中识别候选语音动作无源数据结构集合,所述候选语音动作无源数据结构集合包括由所述数据定义的所述语音动作无源数据结构并且基于与所述候选语音动作无源数据结构集合相关联的相应上下文来识别;

通过将所述用户命令话语的所述转录抄本与所述候选语音动作无源数据结构集合中的相应候选语音动作无源数据结构的触发项进行比较来缩小所识别的候选语音动作无源数据结构集合;

由所述语音动作系统确定所述用户命令话语的所述转录抄本对应于与所述语音动作无源数据结构相关联的所述一个或多个触发项;以及

响应于所述确定,由所述语音动作系统将所述语音动作无源数据结构提供给与所述语音动作系统远离的所述用户设备,由此使得安装在所述用户设备上的软件应用执行用于执行所述新语音动作的所述一个或多个操作。

## 开发人员语音动作系统

### 技术领域

[0001] 本说明书涉及语音动作,以及一个具体实现方式涉及配置语音动作。

### 背景技术

[0002] 应用中的任务能包括在用户交互以执行某一作业的软件中定义的一个或多个活动。活动是控制任务的执行的生命周期的类别,在任务内训练的多个活动能允许用户执行作业。在一些实现方式中,命令能与活动或有关活动的动作相关联,使得命令的用户提交能触发活动或动作。能激发特定意图来触发开始活动或执行动作。

### 发明内容

[0003] 本说明书描述允许应用开发人员部署用于先前安装的软件应用的新语音动作的平台。通常,使用该平台,应用开发人员能提交定义新语音动作的信息,其中,该信息指定该应用、将新语音动作指向触发的动作、用于触发动作的触发项以及触发项应当有效触发应用中的动作的上下文。

[0004] 当语音动作将要对触发应用中的动作有效时,由应用开发人员提交的上下文能指定用户设备状态。设备状态能例如包括哪一个或多个应用在前台(即,目前在设备的用户界面中有效)或后台(例如,目前在设备的用户界面中看不到)中操作,或能包括有关应用的特定信息,诸如什么活动目前正运行、运动活动的状态等。

[0005] 当应用开发人员提交定义新语音动作的信息时,服务或工具能验证定义新语音动作的信息来确定新语音动作是否与应用兼容,或者否则确定是否能实现新语音动作。如果新语音动作有效,通过引入定义新语音动作的信息,为该语音动作创建称为意图并且具有特定格式的无源数据结构的新实例。该意图能指定定义新语音动作的一些或全部信息,诸如应用、触发项、响应于检测触发项触发的活动或动作,和触发项触发活动或动作所需的上下文。然后,对该应用,能部署语音动作,使得语音动作变为用于该应用的启用的语音动作,无需对应用代码做额外变更来支持该语音动作。

[0006] 一旦部署了语音动作,操作用户设备的用户能提供语音输入。用户设备能提交用于该用户设备的上下文信息,并且能使用上下文信息和语音输入的转录抄本来识别意图和触发所识别的意图。当触发意图时,用于执行活动或动作的数据被确定,并且被传送到用户设备以响应于该语音输入,执行活动或动作。

[0007] 例如,媒体播放器应用的应用开发人员能定义使用触发项“播放下一首(play next)”,用于跳到下一首歌的新语音动作。当应当启用“播放下一首”语音动作时,应用开发人员能指定上下文。例如,应用开发人员可以指定,仅当媒体播放器应用在前台中运行并且处于使媒体播放器应用操作为音频播放器的模式中时,“播放下一首”语音动作应当导致媒体播放器应用跳到下一首歌。在通过应用开发人员部署语音动作后,将媒体播放器应用先前安装在他们的用户设备上的用户提交语音输入“播放下一首”。用户设备能提交指示用户设备或安装在用户设备上的应用的状态的上下文信息,诸如指示媒体播放器应用在前台中

运行并且处于音频播放器模式中的上下文信息。响应于接收语音输入“播放下一首”和指定上下文的信息,能将使得媒体播放器应用跳到下一首歌的数据传送到用户设备。相反,如果确定语音输入为说出“播放下一首”,但上下文信息指示社交网络应用在用户设备的前台运行,而不是媒体播放器应用,那么“播放下一首”语音输入不具有任何效果,或会导致在用户设备处执行不同操作。

[0008] 在本说明书中描述的主题的创新方面可以体现在方法中,该方法包括如下动作:由语音动作系统接收指定用于不同于所述语音动作系统的软件应用的新语音动作的数据,所述数据包括用于执行该新语音动作的一个或多个操作和用于触发该新语音动作的一个或多个触发项;至少基于接收的数据,由该语音动作系统生成用于该软件应用的上下文语音动作意图,其中,该上下文语音动作意图包括当被该软件应用接收时,请求该软件应用执行该新语音动作的一个或多个操作的数据;由该语音动作系统将该上下文语音动作意图与用于该新语音动作的一个或多个触发项相关联;其中,该语音动作系统被配置为:接收由安装有该软件应用的设备获得的用户话语的指示;确定该用户话语的转录抄本对应于与该上下文语音动作意图相关联的一个或多个触发项;以及响应于该确定,将该上下文语音动作意图提供给该设备,由此请求安装在该设备上的软件应用执行新语音动作的一个或多个操作。

[0009] 这些和其他实施例可以每一个可选地包括下述特征中的一个或多个。在各个示例中,接收的数据指定上下文,该上下文指定当启用该新语音动作时,该设备或该软件应用的状态;该上下文指定该软件应用正在安装有软件应用的设备的前台中操作;该上下文指定该软件应用正在安装有软件应用的设备的后台中操作;该上下文指定该软件应用正在执行特定活动;该上下文指定该软件应用正执行的特定活动处于特定活动状态。

[0010] 在其他示例中,这些和其他实施例可以每一个可选地包括下述特征:将该上下文语音动作意图与用于新语音动作的上下文相关联;由该语音动作系统接收指示安装有该软件应用的特定设备或安装在该特定设备上的软件应用的状态的上下文信息;确定该上下文信息满足用于新语音动作的上下文;以及响应于确定由特定设备获得的用户话语的转录抄本对应于与上下文语音动作意图相关联的一个或多个触发项以及该上下文信息满足与该上下文语音动作意图相关联的上下文,由该语音动作系统将上下文语音动作意图提供给该特定设备,由此请求安装在特定设备上的软件应用执行新语音动作的一个或多个操作;接收指示特定设备或安装在特定设备上的软件应用的状态的上下文信息包括:由语音动作系统将用于特定上下文信息的请求提供给特定设备;以及响应于该请求,接收特定上下文信息。

[0011] 在其他示例中,这些和其他实施例可以每一个包括一个或多个特征,包括:确定该上下文信息满足用于第二语音动作的上下文,并且用户话语的转录抄本对应于与用于该第二语音动作的意图相关联的一个或多个触发项,其中,用于第二语音动作的意图指定用于执行第二语音动作的一个或多个操作;响应于该确定,从新语音动作和第二语音动作中选择语音动作;以及由该语音动作系统将选择的语音动作相关联的意图提供给该特定设备,由此请求安装在该特定设备上的软件应用执行选择的语音动作的一个或多个操作;从该新语音动作和第二语音动作中选择选择的语音动作包括响应于接收指示用户选择新语音动作或第二语音动作中的一个的数据,选择选择的语音动作;从该新语音动作和第二语

音动作选择选择的语音动作包括：向该新语音动作和选择的语音动作中的每一个分配得分；以及至少基于分配给新语音动作和第二语音动作的每一个的得分，选择选择的语音动作；从该新语音动作和第二语音动作中选择选择的语音动作包括响应于确定与选择的语音动作相关联的软件应用正在前台中操作，选择选择的语音动作。

[0012] 在其他示例中，这些和其他实施例可以每一个包括一个或多个特征，包括：生成用于该软件应用的上下文语音动作意图包括确定新语音动作的一个或多个操作能够由该软件应用执行；确定由安装有该软件应用的特定设备获得的用户话语的转录抄本与该上下文语音动作意图相关联的一个或多个触发项类似；响应于该确定，由该语音动作系统将指示确认该用户话语是否指定一个或多个触发项，或旨在使该软件应用执行该新语音动作的用于用户输入的请求的数据提供给该特定设备；响应于该请求，由该语音动作系统并且从该特定设备接收指示确认的数据；以及响应于接收指示该确认的数据，由该语音动作系统将该上下文语音动作意图提供给特定设备，由此请求安装在特定设备上的软件应用执行该新语音动作的一个或多个操作。

[0013] 在其他示例中，这些和其他实施例可以每一个包括由该语音动作系统从与软件应用相关联的开发人员接收部署用于该软件应用的新语音动作的请求；以及响应于该请求，部署用于该软件应用的新语音动作，其中，部署用于该软件应用的新语音动作启用用于该软件应用的新语音动作；由该语音动作系统从与该软件应用相关联的开发人员接收撤销部署用于该软件应用的新语音动作的请求；以及响应于该请求，撤销部署用于该软件应用的新语音动作，其中，撤销部署用于软件应用的新语音动作禁用用于该软件应用的新语音动作。

[0014] 在其他示例中，这些和其他实施例可以每一个包括一个或多个特征，包括：由该语音动作系统从与该软件应用相关联的开发人员接收启用新语音动作的测试的请求，其中，该请求指定应当为其启用该新语音动作的一个或多个设备；以及响应于该请求，启用用于一个或多个指定的设备的新语音动作，其中，对不包括在指定的设备中的设备，禁用该新语音动作；由该语音动作系统接收由安装有软件应用的特定设备获得的用户话语的指示；由该语音动作系统确定该用户话语的转录抄本对应于与该上下文语音动作意图相关联的一个或多个触发项；以及响应于该确定，由该语音动作系统将该上下文语音动作意图提供给该特定设备，由此请求安装在该特定设备上的软件应用执行该新语音动作的一个或多个操作。

[0015] 在附图和以下描述中阐明了本说明书中所描述的主题的一个或多个实施例的细节。主题的其他特征、方面和优点将从说明书、附图和权利要求书变得显而易见。

## 附图说明

[0016] 图1图示用于语音动作开发系统和服务的示例系统。

[0017] 图2图示用于使用语音动作开发系统和服务，生成新语音动作的示例系统。

[0018] 图3图示用于使用语音动作的示例系统。

[0019] 图4是与语音动作开发系统和服务相关联的示例过程的流程图。

[0020] 各附图中相同的参考符号指示相同元件。



## 具体实施方式

[0021] 图1示出语音动作系统100的示例。语音动作系统100提供平台和服务,由此,应用开发人员能为先前已经安装在其他用户的设备上的应用,建立新语音动作。此后,已经在他们的设备上安装应用的该应用的用户能使用语音命令以便使应用执行如由新语音动作指定的特定动作。

[0022] 语音动作系统100从终端102a-102n接收定义由应用开发人员101a-101n提交的新动作的数据,其中,新语音动作用于一个或多个不同应用。语音动作系统100能确定每一新语音动作是否为有效语音动作。确定为有效的每一新语音动作能被引入来生成对应于新语音动作的意图,并且对应于新语音动作的意图能被存储在语音动作系统100的应用语音动作数据库110处。

[0023] 例如,语音动作系统100可以具有接收和验证定义新语音动作的数据的验证引擎。验证数据可以包括确定定义语音动作的数据的格式符合需求,确定新语音动作与特定应用兼容,或者否则确定基于从应用开发人员101a-101n接收的数据,形成语音动作。语音动作系统100的指示引擎能接收定义新语音动作的验证数据,并且能引入该数据来生成对应于新语音动作的意图。然后,将该意图存储在应用语音动作数据库110中。

[0024] 应用语音动作数据库110能存储属于能在许多不同软件应用内执行的活动或动作的语音动作。包括在应用语音动作数据库110中的语音动作可以包括当建立应用时,由应用开发人员101a-101n提交的内置语音动作,以及在建立应用后,由应用开发人员101a-101n提交的语音动作,用于由该应用支持的操作。

[0025] 语音动作还能包括通过缺省由操作系统支持的专用语音动作。这些预支持的语音动作可以是能够在应用中实现,不必由应用开发人员101a-101n提交给语音动作系统100的语音动作。例如,退出在前台运行的任务或活动的语音动作“退出应用”可以自动地用于任何应用,并且可以包括在用于每一应用的应用语音动作数据库110中,应用开发人员101a-101n不必将信息提交给语音动作系统100来定义“退出应用”语音动作。

[0026] 除应用语音动作数据库110外,语音动作系统100还能包括存储与任何特定应用或上下文不相关联的语音动作的操作系统(OS)语音动作数据库120。例如,使设备进入锁定状态的语音动作“锁定电话”可以是在OS语音动作数据库120中指定为与特定应用不相关联,或普遍可访问,即,当检测到“锁定设备”语音动作时,与设备的状态无关的语音动作。通常,在OS语音动作数据库120中存储的语音动作为不是基于应用开发人员101a-101n定义语音动作而生成的动作,因为在OS语音动作数据库120处存储的语音动作通用于用户设备操作环境,而不是由第三方应用开发人员101a-101n生成的任何特定应用。

[0027] 语音动作系统100使具有用户设备104的用户105能够将说出的语音输入提供给他们自己的用户设备104来使得由用户设备104或在他们的用户设备104上操作的应用执行动作。例如,当媒体播放器应用在用户设备104的前台中运行时,具有用户设备104的用户105具有在他们的用户设备104上运行的媒体播放器应用并且提供说出的输入“播放下一首”。

[0028] 对应于说出的输入和指定用户设备104的状态的上下文信息的音频数据,包括在用户设备104上运行的应用的状态,被传送到语音动作系统100。特别地,对图1所示的示例,对应于语音输入“播放下一首”的音频数据由语音动作系统100的语音识别引擎130接收,并且指示用户设备104的状态的上下文信息被语音动作系统100的上下文分析引擎140接收。

上下文信息可以指示媒体播放器应用正在用户设备104的前台中运行,媒体播放器应用目前处于音频播放器模式,并且可以指示关于用户设备104的状态和在用户设备104上安装的应用的其他信息。

[0029] 语音识别引擎130接收对应于语音输入的音频数据,生成语音输入的转录抄本,并且将语音输入的转录抄本提供给语音动作选择器150。上下文分析引擎140从用户设备104接收上下文信息并且处理上下文信息来确定相关上下文信息。例如,上下文分析引擎140可以分析上下文信息来识别安装在用户设备104上的应用,以及与那些应用的每一个相关联的元数据可以指定用于该应用的可用语音动作以及有关确定对给定设备状态,可以启用哪些语音动作的上下文信息。在一些实例中,基于该分析,上下文分析引擎140可以确定要求附加上下文信息来识别对用户设备104的特定状态,应当启用哪些语音动作,因此,上下文分析引擎140可以从用户设备104请求附加上下文信息。上下文分析引擎140可以将处理的上下文信息转发到语音动作选择器150。

[0030] 例如,如图1所示,语音识别引擎130可以接收对应于语音输入“播放下一首”的音频数据并且可以获得语音输入的转录抄本。上下文分析引擎140可以从用户设备104接收指定媒体播放器应用正在前台中运行、媒体播放器应用正在音频播放器模式中操作,并且指定关于用户设备104的其他信息的上下文信息。例如,上下文信息还可以指示社交网络应用正在用户设备104的后台中操作并且用户设备104的电池寿命目前为50%。上下文分析引擎140可以接收指示用户设备104具有媒体播放器和所安装的社交网络应用的信息,并且可以确定媒体播放器应用和社交网络应用均不会基于指示设备电池寿命的上下文信息启用语音动作。因此,上下文分析引擎140可以过滤上下文信息来仅指示媒体播放器应用正在用户设备104的前台中操作、媒体播放器应用处于音频播放器模式,以及社交网络应用正在用户设备104的后台中操作。然后,过滤的上下文信息和语音输入“播放下一首”的转录抄本被提供给语音动作选择器150。

[0031] 语音动作选择器150从语音识别引擎130接收说出的输入的转录抄本和从上下文分析引擎140接收包括处理的上下文信息的上下文。语音动作选择器150使用转录抄本和上下文来识别与在用户设备104处触发的特定语音动作相关联的特定意图。例如,语音动作选择器150能访问应用语音动作数据库110和OS语音动作数据库120来识别对用户设备104的当前上下文启用的候选语音动作的集合。识别候选语音动作的集合能包括识别在应用语音动作数据库110或OS语音动作数据库120处存储的、指定与由语音动作选择器150接收的上下文匹配的上下文的意图的集合。

[0032] 已经识别对提出的上下文启用的候选语音动作的集合,语音动作选择器150能将语音输入的转录抄本与启用的语音动作的每一个相关联的一个或多个触发短语比较。通常,触发短语能包括一个或多个触发项,并且操作为用于语音动作的标识符,使得检测特定触发短语中的一个或多个项导致识别和触发对应于特定触发短语的语音动作。例如,语音动作选择器150能将转录抄本与启用的语音动作的意图相关联的各个触发短语比较。语音动作选择器150基于确定转录抄本与由特定语音动作相关联的意图指定的触发项匹配,识别用于特定语音动作的特定意图。

[0033] 例如,语音动作选择器150能接收语音输入“播放下一首”的转录抄本以及用于用户设备104的上下文,并且能使用接收的上下文来基于用户设备104的当前状态,识别在用

户设备104处启动的候选语音动作,即,能在用户设备104处启动的语音动作。在确定用于用户设备104的上下文的候选语音动作的集合后,语音动作选择器150将语音输入“播放下一首”的转录抄本与由候选语音动作的意图指定的触发短语比较。语音动作选择器150能确定转录抄本“播放下一首”与由用于媒体播放器应用的“播放下一首”语音动作的意图指定的触发短语匹配。

[0034] 基于确定转录抄本的至少一部分与触发短语的一个或多个项匹配,转录抄本被识别为匹配于触发短语。例如,可以基于转录抄本的项的每一个与触发短语的项的每一个匹配,可以识别转录抄本和触发短语之间的匹配。在一些实现方式中,可以基于确定转录抄本的并非全部项与触发短语的项匹配,识别匹配。例如,即使触发短语的一些项不在转录抄本,即使转录抄本包括除那些触发短语外的项,或即使转录抄本的一个或多个项不同于触发短语的项,转录抄本也可以与触发短语匹配。由此,尽管转录抄本“新电子邮件”、“发送新电子邮件”或“打开新电子邮件”的每一个并非精确匹配于触发短语,这些转录抄本的每一个都可以识别为匹配于触发短语“打开新电子邮件”。在一些实例中,如果转录抄本包括以不同于由触发短语指定的顺序的触发短语的一些或全部项,转录抄本可以被识别为匹配于触发短语。例如,转录抄本“锁定电话”可以被识别为匹配于“电话锁定”触发短语。在一些实例中,基于确定转录抄本包括用于触发短语的一个或多个同义词或替选项,可以将转录抄本识别为匹配于触发短语。由此,基于“起草”是“写”的同义词,以及“打开”是用于“写”的识别的替选项,转录抄本“起草新电子邮件”或“打开新电子邮件”可以被识别为匹配于触发短语“写新电子邮件”。

[0035] 已经识别特定语音动作,语音动作选择器150将动作触发数据提供给用户设备104,使得启动与特定语音动作相关联的活动或动作。这样做,语音动作选择器150能获得用于控制用户设备104的信息以执行与选择的语音动作相关联的动作或活动。在一些实例中,控制用户设备104来执行与选择的语音动作相关联的活动或动作能包括激发选择的语音动作的意图。激发选择的语音动作的意图可以使得用于启动与选择的语音动作相关联的活动或动作的信息被提供给用户设备104。

[0036] 例如,用于选择的语音动作的意图能包括使得与语音动作相关联的活动或动作由与语音动作相关联的应用执行的数据。被选意图能被传送到用户设备104,使得由用户设备104接收意图能充当请求或能触发与语音动作相关联的活动或动作的执行。替选地,语音动作选择器150能确定操作为使选择的语音动作在用户设备104处被执行的动作触发数据的其他数据,并且能将该信息传送到用户设备104。例如,意图仅可以识别将由用户设备104执行的活动或动作来执行语音动作,以及语音动作选择器150能确定能控制用户设备104上的应用执行与语音动作相关联的活动或动作的动作触发数据。确定的动作触发数据能被传送到用户设备104,使得动作触发数据使得与语音动作相关联的活动或动作被执行。

[0037] 例如,在从候选语音动作的集合中选择“播放下一首”语音动作后,语音动作选择器150将用于“播放下一首”语音动作的意图传送到用户设备104,或获得用于控制在用户设备104上运行的媒体播放器应用跳到下一首歌的其他信息。语音动作选择器150将用于控制媒体播放器应用跳到下一首歌的数据传送到用户设备104,其依次启动在用户设备104上运行的媒体播放器应用跳到下一首歌,例如,向前跳到David Bowie的“Fame”的歌的过程。

[0038] 简而言之,如上所述,图1的系统包括对应于一个或多个应用开发人员101a-101n

的一个或多个终端102a-102n。终端102a-102n的每一个能例如在一个或多个有线或无线网络上,与语音动作系统100通信。终端102a-102n的每一个能是移动设备,诸如蜂窝电话、智能电话、平板计算机、手提计算机、个人数字助理(PDA)、上网本计算机或其他移动计算设备,或可以是任何固定的计算设备,诸如台式计算机或其他固定计算机。与用户105相关联的用户设备104还能在一个或多个有线或无线网络上,与语音动作系统100通信,其中,用户设备104也可以是移动或固定计算设备,诸如蜂窝电话、智能电话、平板计算机、上网本、个人数字助理、手提计算机、台式计算机或其他计算设备。

[0039] 图1所示的示例性语音动作系统100包括应用语音动作数据库110、操作系统(OS)语音动作数据库120、语音识别引擎130、上下文分析引擎140和语音动作选择器150。语音动作系统100的组件的每一个,包括应用语音动作数据库110、OS语音动作数据库120、语音识别引擎130、上下文分析引擎140和语音动作选择器150可以在一个或多个有线或无线数据通路上,与语音动作系统100的一个或多个其他组件通信,有线或无线数据通路支持电子通信交换。在一些实现方式中,可以合并语音动作系统100的组件中的一个或多个,使得它们的功能由单个组件执行,或可以由两个或以上组件表示,使得它们的功能分散在两个或以上组件上。语音动作系统100的组件可以实现在单个计算设备,诸如单个服务器系统上,或可以实现在组件之间支持电子通信交换的一个或多个有线或无线数据通路上通信的多个计算设备上。

[0040] 图2示出使应用开发人员能够生成用于在一个或多个用户设备上安装的应用的新语音动作的示例系统。例如,在应用已经被发行并且下载到一个或多个用户设备后,应用开发人员能使用语音动作开发服务,定义新语音动作。语音动作开发服务能部署新语音动作来支持用于应用的新语音动作,而不要求对应用本身的修改。

[0041] 简而言之,图2图示将语音动作开发服务提供给与终端202相关联的应用开发人员201的开发人员语音动作系统200。终端202可以在使得终端202能够与开发人员语音动作系统200交换电子通信的一个或多个有线或无线网络连接,诸如有线或无线互联网连接上,与开发人员语音动作系统200通信。开发人员语音动作系统200包括与验证标准数据库215通信的验证引擎210。开发人员语音动作系统200进一步包括与应用元数据后端225和应用语音动作数据库230通信的语法引入引擎220。应用语音动作数据库230可以与图1的应用语音动作数据库110类似,并且能存储由应用开发人员提交以定义用于应用的新语音动作并且基于由定义新语音动作的应用开发人员提交的信息的语法生成的意图。在一些实现方式中,为应用开发人员提供服务来建立新语音动作的开发人员语音动作系统200能是图1的语音动作系统100的一部分。在一些实现方式中,由开发人员语音动作系统200的组件执行的操作可以由开发人员语音动作系统200的不同组件来执行的操作。开发人员语音动作系统200可以包括更多、更少或不同的组件来执行如由开发人员语音动作系统200执行的所述操作。

[0042] 如图2所示,使用终端202,应用开发人员201能将定义新语音动作的信息提交给开发人员语音动作系统200。例如,终端202可以包括用于开发人员语音动作服务的应用,或终端202能够例如通过在互联网上,访问网站处的开发人员语音动作服务,在网络连接上,远程访问开发人员语音动作。在其他实现方式中,可以由开发人员语音动作系统200托管开发人员语音动作服务,使得终端202能在终端202和开发人员语音动作系统200之间的网络连

接上,访问开发人员语音动作系统200处的开发人员语音动作服务。

[0043] 为实现用于特定应用的新语音动作,应用开发人员201能将定义新语音动作的新语法提交给开发人员语音动作系统200,或能将定义新语音动作的其他信息提交给开发人员语音动作系统200。在图2所示的示例中,为实现新语音动作,应用开发人员201首先定义用于已经安装在一个或多个用户设备上的媒体播放器应用的新语音动作。这样做,应用开发人员201提交定义用于媒体播放器应用的新语音动作的语法。由应用开发人员201提交的语法指定应用,即媒体播放器应用、将与新语音动作相关联的触发短语“播放下一首”、响应于检测“播放下一首”触发短语,即跳到下一首歌,执行的动作、以及用于当新语音动作被启用,即媒体播放器应用必须在用户设备的前台中操作以及媒体播放器应用必须处于音频播放器模式,时的上下文。

[0044] 由应用开发人员201提交的语法可以是以使得应用开发人员201容易定义新语音动作的特定格式,无需熟悉用于语音动作意图的、由开发人员语音动作系统200使用的特定数据格式。在其他实现方式中,应用开发人员201能提交能由开发人员语音动作系统200的验证引擎210或另一组件接收并且被转换以生成定义新语音动作的语法的其他信息。例如,应用开发人员201可以利用包括应用开发人员201能完成以定义新语音动作的字段的表格来表示。表格可以使应用开发人员201能够指定有关新语音动作的信息,例如应用、用于新语音动作的触发短语和响应于触发短语,由应用执行或相对于应用的活动或动作。该表格能允许应用开发人员201提供对待启用的新语音动作,即,对将作为用户使用触发短语触发的候选语音动作的语音动作,要求用户设备的或在用户设备上操作的一个或多个应用的特定状态的上下文。可以在表格处输入有必要或可选地用来定义新语音动作的其他信息。由应用开发人员201提供的表格字段中的条目可以被转换成用于提交给开发人员语音动作系统200的语法,或条目可以由开发人员语音动作系统200接收并且在开发人员语音动作系统200处转换成语法。替选地,应用开发人员201可以通过使用提交给开发人员语音动作系统200的语法格式,或使用另一格式,诸如定界符分隔的值(DSV)格式,输入必要信息,来定义新语音动作。

[0045] 在一些实现方式中,可以开发新语音动作来执行在用户设备上操作的软件应用中的一个或多个操作、通常相对于用户设备,执行一个或多个操作(即,控制用户设备的硬件)、控制用户设备外部的应用(即,在不同设备上操作的应用)、或控制用户设备外部的另一设备。例如,应用开发人员201可以提供定义新语音动作的语法来控制用户设备上的特定应用(例如社交网络应用或媒体播放器应用)、控制用户设备(例如改变设备的屏幕的方位或使设备断电)、控制在不同设备上操作的应用(例如,控制在用户设备外的有线电视盒上操作的有线电视软件)、或控制与用户设备分离的另一设备(例如,使用提供给用户设备的语音输入,控制车库门)。在一些实现方式中,能定义语音动作,以便触发语音动作使得两个或以上不同活动或动作将在两个或以上不同应用或设备处被执行。

[0046] 由应用开发人员201指定的上下文识别对启用的语音动作,必须满足的一个或多个条件,即,不会由与语音动作相关联的触发短语来触发该语音动作,除非还满足上下文。应用开发人员201能将上下文定义为具有一个或多个不同需求或条件。

[0047] 在一些实例中,由应用开发人员201提交的上下文可以指定用户设备的状态或属性,诸如是否上电或锁定用户设备、用户设备是否具有相机、陀螺仪、气压计或其他组件或

特征。由此,例如如果从用户设备接收的上下文信息指示用户设备不具有相机,不会启用使用安装在用户设备上的应用拍摄照片的语音动作。

[0048] 在其他示例中,上下文要求特定应用在用户设备的前台或后台中操作。例如,如图2所示,仅当媒体播放器应用在用户设备204的前台中操作时,可以启用“播放下一首”语音动作,但如果媒体播放器应用关闭或在用户设备204的后台中操作时则不是如此。在其他示例中,利用当应用在后台中操作时启用语音动作的上下文,定义语音动作。例如,社交网络应用可以具有相关的语音动作来接受与另一用户连接的请求,以及用户能够触发语音动作来接受与另一用户连接的请求,即使社交网络应用仅在用户设备的后台中操作。

[0049] 上下文可以附加或可替代地要求启用在用于语音动作的特定模式中操作的特定应用。模式可以是应用正执行的特定活动或任务。例如,电子邮件应用可以被确定为处于电子邮件撰写模式,媒体播放器应用可以处于音频播放器模式,或相机应用可以处于相机模式或相册查看模式。用户设备能够确定特定应用操作在哪一模式,并且可以将那一信息包括在被用来确定是否启用特定语音动作的上下文信息中。

[0050] 附加地,在一些实例中,上下文可以要求应用模式具有某些状态。例如,上下文可以指示仅当媒体播放器应用处于电影播放器模式时以及当电影播放器模式中的应用的状态是应用目前正播放电影时,启用“暂停”电影的语音动作。在一些实例中,用户设备不能确定在特定模式中操作的应用的状态。在那些实例中,有必要配置该应用来在特定模式中提供指示该应用的状态的信息。由此,在用于特定应用的上下文中指定这样的条件可以要求修改应用本身,以便由用户设备提供的上下文信息包括所需的状态信息。

[0051] 在一些示例中,为语音动作指定的上下文要求在用户设备的屏幕上显示特定对象或对象类型来启用语音动作。例如,上下文可以指定仅当在用户设备的显示器输出图像,才启用“保存图像”语音动作,或可以指定如果显示器上呈现文本,才使得“高亮”语音动作可用。在另一示例中,仅当上下文信息指示在用户设备的显示器上呈现数字“1”或项目符号,或存在显示器上呈现的项目列表时,选择列表中的特定项目的语音动作,诸如“选择第一首”的语音动作才可用。在一些实例中,用户设备能够确定在给定时刻,例如当应用在显示器的控制中时,在显示器处呈现什么信息。由此,如果上下文指定对待启用的语音动作,必须在显示器处输出某些信息或信息类型,那么有必要修改该应用以提供那一信息。然后,用户设备能将指示在用户设备的显示器处输出什么的信息包括在被用来确定是否应当启用特定语音动作的上下文信息中。

[0052] 在其他实例中,上下文要求在用户设备处输出通知或特定类型的通知。例如,当收到新电子邮件时,向用户提供弹出通知的电子邮件应用可以具有对当通过应用新电子邮件通知被输出到用户设备的显示器时启用的“读取电子邮件”关联的语音动作。由此,上下文信息可以指示通知是否被输出到用户设备的显示器,以及信息可以被用于确定是否应当启用“读取电子邮件”。

[0053] 在一些示例中,上下文要求用户设备处于用于待启用的语音动作的特定地理位置。用户设备能够基于接入与特定地理位置相关联的互联网连接,使用蜂窝塔三角测量,使用全球定位系统(GPS)技术或使用其他手段,确定其地理位置。用户设备可以将指示其地理位置的信息包括在上下文信息中,以及可以基于满足上下文的地理位置需求的地理位置,启用语音动作。例如,与用于零售店的应用的语音动作相关联的上下文可以指定仅当具有

零售店应用的用户设备在零售店中的一个的范围内时,才应当处理某些语音动作。基于来自用户设备的指示用户设备的地理位置对应于在零售店中的一个的范围内的地理位置的上下文信息,启用用于零售店应用的语音动作。

[0054] 由应用开发人员定义的上下文还能指示在某些时间、日期、月、季期间,或当满足其他时间条件时,才启用与上下文相关联的语音动作。例如,上下文可以指定仅当从用户设备或从另一信息源接收的上下文信息满足时间条件时,才启用某些语音动作。例如,用于银行应用的语音动作可以与包括时间条件的上下文相关联,使得仅当上下文信息指示当提交语音动作时的时间在6:00AM和10:00PM之间时,才启用语音动作。

[0055] 上下文可以指定用于待启用的语音动作的其他需求。例如,开发人员语音动作系统200可以访问来自除用户设备外的源的信息,或能够从用户设备接收从一个或多个信息源获得的信息。从其他源接收的信息能是用于启用语音动作的要求的上下文信息。这样的信息能包括例如天气信息、用户的情绪信息、新闻信息、股票市场信息或其他信息。例如,开发人员语音动作系统200能够通过一个或多个有线或无线网络连接,例如互联网连接或与服务器系统的其他网络连接,访问其他信息源。用于语音动作的上下文可以指定仅当用于对应于相关用户设备的位置的地理区域的天气信息指示那一位置正下雨,才启用语音动作。为检查是否满足上下文条件,开发人员语音动作系统200能够在一个或多个网络连接上,访问用户设备的已知位置的天气信息。

[0056] 能将定义由应用开发人员201提交的新语音动作的语法或其他信息在网络上,从终端202提交给开发人员语音动作系统200。在开发人员语音动作系统200处,由验证引擎210接收提交的语法。如果应用开发人员201将定义语音动作,但不是以语法格式的信息提交给开发人员语音动作系统200,开发人员语音动作系统200可以将定义新语音动作的信息转换成语法。验证引擎210分析该语法来确定提交的语法是否可以被引入用于新语音动作的意图中。例如,验证引擎210能够确定由语法指定的触发短语是否是有效触发短语,确定由语法指定的活动是否能由特定的应用执行,能确定提交的语法的格式是否句法上有效,能确定由请求指定的上下文是否有效,或否则能确定提交的语法是否是能被引入以生成用于新语音动作的意图的有效语法。为执行该验证,验证引擎210访问包括被用来验证语法的规则的验证标准数据库215。

[0057] 例如,验证引擎210接收指定用于媒体播放器应用的新语音动作的语法。语法指定触发短语“播放下一首”,该触发短语“播放下一首”应当使媒体播放器应用跳到下一首歌,以及上下文指定当媒体播放器应用在前台操作并且是音频播放器模式时,才能启用新语音动作。验证引擎210能在验证标准数据库215处访问验证规则并且基于规则,验证接收的语法。例如,验证引擎210可以确定媒体播放器应用是否是现有应用,使得对特定应用,创建新语音动作。验证引擎210可以访问指定触发短语必须是长度大于1个语音音节的规则,并且可以基于确定触发短语“播放下一首”长于1个语音音节,验证触发短语“播放下一首”。验证规则可以指定媒体播放器应用能执行的可能动作列表,并且可以基于确定跳到下一首歌是验证引擎210能够执行的操作,验证跳到下一首歌的特定的动作。验证引擎210可以验证由语法指定的上下文以确保该上下文不包含矛盾。例如,验证引擎可以确保对待启用的新语音动作,特定的上下文不要求应用在前台操作并且也在后台操作,因为只能满足两个条件中的一个。其他验证标准可以应用于接收的语法来确定由应用开发人员201提交的语法是



否有效。

[0058] 可以由验证引擎210执行其他形式的验证。例如,验证引擎210可以确定是否授权应用开发人员201来生成用于特定应用的新语音动作,可以确定是否已经支付生成新语音动作所需的费用,或者否则可以确定由应用开发人员201提交的语法是否导致创建用于新语音动作的意图。

[0059] 基于由验证引擎210执行验证的结果,能将指示该语法是否有效的反馈提供给应用开发人员201。如果验证引擎210确定该语法无效,验证引擎210可以使终端202输出指示语法无效并且请求应用开发人员201校正语法的信息,或可以建议应用开发人员201能校正语法的方式。例如,如果语法被确定为无效,因为它指定不能由特定的应用执行动作(例如,如果语法指定新语音动作应当使媒体播放器应用打开新电子邮件),那么验证引擎210可以使指示该语法指定与特定的应用不兼容的动作的信息在终端202处输出,并且可以指示应用开发人员201可以通过改变特定的动作或特定的应用来校正语法。如果提交的语法有效,验证引擎210可以提供提交的语法有效的指示。

[0060] 附加地,验证引擎210可以从应用开发人员201请求附加输入。例如,如果语法被确定为有效,验证引擎210可以为应用开发人员201提供继续生成新语音命令的选项,使得不会由该语法生成用于新语音命令的意图,除非应用开发人员201确认他们希望继续生成新语音命令。这可以使得应用开发人员201确定他们是否将语法正确地提交给开发人员语音动作系统200,而不必继续新语音动作的语法引入或部署。

[0061] 基于由应用开发人员201提交的语法被验证,并且可选地,应用开发人员201提供指示应当继续语音动作生成过程的输入,能由验证引擎210将语法提供给语法引入引擎220。语法引入引擎220能引入接收的语法来生成用于新语音动作的意图。生成的意图能指定包括在定义新语音动作的语法中的信息。例如,语法引入引擎220能引入接收的语法来生成指定应用、触发短语、响应于检测到触发短语执行的动作或活动以及对该新语音动作指定的上下文中的一个或多个的意图数据结构的实例。语法引入引擎220能将用于新语音动作的意图存储在应用语音动作数据库230处。

[0062] 例如,如图2所示,语法引入引擎220能接收由应用开发人员201提交的、由验证引擎210验证的语法。语法引入引擎220能基于接收的语法,对新语音动作,生成新意图,“意图Y”。为生成该意图,语法引入引擎220可以生成意图数据结构的新实例并且可以用包括在接收的语法中的信息,包括指定媒体播放器应用的信息、指定触发短语“播放下一首”的信息、指定在媒体播放器应用中,“播放下一首”语音动作应当与跳到下一首歌相关联的信息,以及指定指示当媒体播放器应用在前台操作并且在音频播放器模式中操作时,应当启用“播放下一首”语音动作的上下文的信息填充意图数据结构。

[0063] 能在包括用于其他语音动作的其他意图,诸如“意图X”和“意图Z”的应用语音动作数据库230处,由语法引入引擎220存储新“意图Y”。“意图X”和“意图Z”的每一个可以与用于其他应用的其他语音动作相关联,或与用于媒体播放器应用的其他语音动作相关联。一旦新“意图Y”被存储在应用语音动作数据库230处,可以部署新语音动作,使得可以由将媒体播放器应用存储在他们的设备上的用户使用新语音动作。即,一旦部署,媒体播放器应用的用户可以使用新“播放下一首”语音动作,而不必更新他们的应用。

[0064] 在一些实现方式中,引入语法以生成意图还可以包含生成至少基于由应用开发人



员201提交的触发项的一个或多个备选触发项,并且将备选触发项包括在为新语音动作生成的意图中。例如,可以将一个或多个扩展规则、同义词规则、可选规则、替代规则、重写规则或其他规则应用于由应用开发人员201提交的触发项来确定备选触发项。语法引入引擎220可以将备选触发项包括在用于新语音动作的意图中,使得检测由应用开发人员201提交的触发项或备选触发项会导致触发语音动作。

[0065] 附加地,在一些实现方式中,语法引入引擎220会与意图数据结构分开地存储与新意图相关的信息。例如,可以与用于新语音动作的意图分开地存储指定用于待启用的新语音动作所需的上下文的信息,或与新语音动作相关联的触发项。在这些实例中,从用户设备接收的上下文信息或从用户设备接收的触发项可以与该意图相关联的上下文和触发项比较,而不必访问意图中的数据。可以基于确定接收的上下文信息和/或触发项与意图相关联的各自的上下文和/或触发项匹配,触发该意图。以这种方式,仅需要将触发与新语音动作相关联的活动或动作所需的信息包括在意图中,以及可以从意图外部地存储用来选择用于触发的意图,用来识别意图或其他信息的其他信息。例如,开发人员语音动作系统200可以存储查找表、链接列表或将上下文或触发项信息映射到特定意图的其他数据,以便上下文和触发项的特定配对的识别可以使与所述配对相关联的意图被触发。

[0066] 在一些实现方式中,开发人员语音动作系统200还可以为应用开发人员201提供能力来在被部署为公众使用前,测试新语音动作。例如,开发人员语音动作系统200可以为应用开发人员201提供能力来使用终端202测试新语音动作,而该语音动作对于其他用户不可用。开发人员语音动作系统200还可以为应用开发人员201提供能力来指定应当启用新语音动作的一个或多个用户设备以便测试。例如,应用开发人员201可以指定对已知用户设备的贝塔测试组,应当部署新语音动作,来在部署用于公众使用的新语音动作前,确保新语音动作的适当操作。

[0067] 在一些实现方式中,开发人员语音动作系统200可以通过改变与意图相关联的设置,支持新语音动作的测试,使得基于从指定用于测试的用户设备中的一个接收语音命令,识别和触发意图。例如,开发人员语音动作系统200可以存储用于在应用语音动作数据库230中存储的意图的每一个的设置,其中,设置指示用于特定语音动作的意图是否已经被部署为公众使用,或可以识别为测试目的部署意图的特定用户设备。在将意图存储在应用语音动作数据库230后,开发人员语音动作系统200可以从终端202接收指示为测试目的,对一个或多个用户设备的特定集合,应当启用该意图的信息。作为响应,可以更新用于该意图的设置来将该意图部署到特定用户设备集合,使得可以触发与该意图相关联的语音动作。一旦为测试启用,在该集合中,从用户设备接收的语音输入可以使意图被触发,由此允许测试语音动作。

[0068] 在其他实施例中,在用于新语音动作的意图被存储在应用语音动作数据库230中前,开发人员语音动作系统200可以启用测试新语音动作。例如,应用开发人员201可以指示他们希望在被部署前,测试新语音动作,并且可以指示应当暂时启用新语音动作以测试的多个用户设备或用户帐户。在生成用于新语音动作的意图后,开发人员语音动作系统200可以启用用于特定用户设备或用户帐户的语音动作,使得如果用户将与新语音动作相关联的触发短语提供为用户设备中的一个的输入,特定用户设备或与特定用户帐户相关联的用户设备会触发新语音动作。

[0069] 在又一实现方式中,在由语法引入引擎220引入用于新语音动作的语法前,开发人员语音动作系统200能提供用于新语音动作的测试能力。例如,在验证接收的语法后,验证引擎210或开发人员语音动作系统200的另一组件能够模拟由该语法定义的新语音动作的操作来使得应用开发人员201或其他设备的用户来测试新语音动作。在一些实例中,模拟新语音动作要求终端202或用于测试的其他用户设备来利用由开发人员语音动作系统200提供的模拟器或测试模式来模拟新语音动作的操作。在引入新语法前,提供测试能力能允许由应用开发人员201更快反复测试新语音动作,由此加速整个语音动作开发过程。

[0070] 还能由开发人员语音动作系统200提供部署语音动作和撤销语音动作的部署的能力。例如,在将定义新语音动作的语法由应用开发人员201提交给开发人员语音动作系统200、验证、引入以生成新意图,并且已经将新意图存储在应用语音动作数据库230后,开发人员语音动作系统200可以请求应用开发人员201是否希望部署新语音动作。应用开发人员201可以选择部署新语音动作,其中,部署新语音动作使得在安装与新语音动作相关联的应用的一般公众的用户设备处触发新语音动作。应用开发人员201能在终端202处提供输入来部署新语音动作。然后,能部署新语音动作,使得新语音动作能够在已经安装与新语音动作相关联的应用的其他用户设备上被触发。即,在部署新语音动作后,在已经安装应用的用户设备处,用于新语音动作的触发短语的输入将导致与新语音动作相关联的意图被触发。触发与新语音动作相关联的意图使得该意图或其他动作触发数据被传送到用户设备,以便该意图真正的数据或动作触发数据使得将由用户设备来执行语音动作的动作或活动。如上所述,部署该意图仅要求开发人员语音动作系统200更新与新语音动作的意图相关联的设置,以响应于语音输入,使得该意图、候选意图被触发。

[0071] 在已经部署语音动作后,应用开发人员201可以撤销新语音动作的部署,使得在已经安装与新语音动作相关联的应用的用户设备处,不再触发新语音动作。例如,应用开发人员201可以将识别语音动作和指示撤销语音动作的部署的意图的输入提供给开发人员语音动作系统200。作为响应,开发人员语音动作系统200例如可以通过更新与用于语音动作的意图相关联的设置,撤销语音动作的部署,使得不能触发用于语音动作的意图。由于部署或收回语音动作仅要求开发人员语音动作系统200更新在开发人员语音动作系统200处存储的信息,而不改变安装在用户的设备上的应用,应用开发人员201可以易于部署或撤销语音动作的部署。

[0072] 在一些实现方式中,开发人员语音动作系统200还能生成用于新语音动作的可发现性示例。可发现性示例能包括告知用户特定语音动作可用、能呈现给用户设备的用户的信息,可以包括在安装应用后,由开发人员提交的新语音动作。在一些示例中,响应于用户设备具有与语音动作相关联的上下文匹配的状态,或响应于用户询问哪些语音动作可用,输出可发现性示例。

[0073] 例如,基于接收定义新语音动作的语法,语法引入引擎220或开发人员语音动作系统200的另一组件能生成用于由该语法定义的语音动作的可发现性示例,并且能将可发现性示例存储在应用元数据后端225中。可以将可发现性示例提供给已经安装与新语音动作相关联的应用的用户设备。例如,可以定期地更新应用的元数据来包括在应用元数据后端225中存储的、用于该应用的可发现性示例的当前列表,备选地,基于用户设备提交指示其当前状态的上下文信息,可以将信息传送到用户设备,使得在用户设备处输出为用户设备

的状态启用的语音动作的可发现性示例。

[0074] 例如,除生成用于“播放下一首”语音动作的新“意图Y”,开发人员语音动作系统200可以生成用于“播放下一首”语音动作的一个或多个可发现性示例。可发现性示例可以包括在具有媒体播放器应用的用户设备处输出的信息。基于满足与“播放下一首”语音动作相关联的上下文,或基于其他条件,可以在用户设备处输出可发现性示例。例如,用于“播放下一首”语音动作的可发现性示例能是每当媒体播放器应用用在用户设备的前台中操作并且处于音频播放器模式时输出的消息“再试一次“播放下一首””。在一些实现方式中,可以响应于其他事件或条件,输出可发现性示例。例如,可以响应于用户设备的用户通过按压按钮或控制,或通过说出“我能说什么语音动作?”,向用户设备提供输入,输出可发现性示例消息“再试一次“播放下一首””。

[0075] 通常,可以将用于应用语音动作的可发现性示例提供给用户,无需应用本身的动作。例如,开发人员语音动作系统200可以将包括可发现性示例的数据提供给用户设备,以及用户设备可以使可发现性示例被输出,无需求与可发现性示例有关的应用执行操作。在一些实现方式中,在用户设备上操作的OS能够接收可发现性示例或访问指定可发现性示例的应用元数据,或使可发现性示例在用户设备处被输出。由此,可以提供用于与特定应用有关的语音动作的可发现性示例以便输出,无需修改特定应用。

[0076] 在一些实现方式中,能以用户将理解不是由应用本身提供可发现性示例的方式,输出用于应用语音动作的可发现性示例。例如,用户设备能以很显然可发现性示例由开发人员语音动作系统200或OS,而不是与语音动作有关的特定应用提供的方式,提供可发现性示例。例如,当提供可发现性示例时,用户设备可以使用识别为与用户设备的OS相关联的特定计算机化语音语调。替选地,可以以明确该应用不提供可发现性示例的方式,输出可发现性示例。例如,可以输出说出“再试一次“播放下一首”或“暂停”以及让媒体播放器应用处理你的请求”的可发现性示例来将OS的角色与由媒体播放器应用本身执行的动作分开。也可以实现指示由单独的系统或应用提供可发现性示例的其他方式。

[0077] 图3示出用于处理从用户设备的用户接收语音动作的示例性语音动作系统300。在一些实现方式中,在将应用下载到用户设备后,由语音动作系统300处理的语音动作可以是应用开发人员已经为应用创建的语音动作。可以使用图2的开发人员语音动作系统200,对该应用创建由应用开发人员创建的语音动作。

[0078] 简而言之,语音动作系统300在一个或多个有线或无线网络,诸如使用户设备304与语音动作系统300交换电子通信的一个或多个有线或无线互联网或蜂窝数据连接上,与属于用户305的用户设备304通信。语音动作系统300包括在一个或多个网络连接上,处理从用户设备304接收的信息的语音识别引擎330和上下文分析引擎340。语音动作系统包括与语音识别引擎330和上下文分析引擎340通信,并且还可以访问应用语音动作数据库310和OS语音动作数据库320的匹配器350。匹配器350与可以访问应用数据数据库370的动作触发生成器360通信。可以访问用户活动历史390的消除歧义引擎380还能在一个或多个网络连接上,从用户设备304发送或接收电子通信。

[0079] 为处理由用户305提交的语音输入,将对应于指示用户设备304和/或在用户设备304上操作的的应用的状态的语音输入和上下文信息的音频数据提供给语音动作系统300。在一些实例中,用户设备304可以响应于用户305选择使用户设备304开始录制用户语音的控

制,响应于用户305说出其检测使得用户设备304从用户305获得后续语音输入的短语,获得语音输入,或者否则可以确定录制用户305的语音输入。提供特定输入来触发用户语音的录制的需求能防止用户设备304不必要地处理用户语音,即通过减少必须处理的音频数据量,或防止录制过多用户语音以致侵犯用户305的隐私,即,通过仅录制用户305正说给用户设备304的特定语音。

[0080] 在图3所示的示例中,例如,用户305可以选择用户设备304上的控制来使得用户设备304开始录制用户305的语音输入,并且在用户305已经选择该控制后,用户305可以通过说出“播放下一首”,提供语音输入。在一个或多个网络上,用户设备304可以将对应于语音输入的音频数据传送到语音动作系统300。附加地,用户设备304可以确定用于用户设备304或在用户设备304上操作的应用的上下文信息,以及用于用户设备304的上下文信息可以在一个或多个网络上,被传送到语音动作系统300。

[0081] 在一些实例中,从用户设备304传送到语音动作系统300的上下文信息包括由在用户设备304处存储的信息识别的特定信息。例如,用户设备304能够将指示传送到语音动作系统300的特定状态信息的信息存储为上下文信息。由用户设备304提供给语音动作系统300的上下文信息可以包括上下文信息的标准化体,使得所有用户设备将相同的上下文信息提交给语音动作系统300,或由用户设备304提供的上下文信息可以是为用户设备304指定的特定上下文信息的集合。在一些实现方式中,语音动作系统300可以从用户设备304请求某些上下文信息,以及由用户设备传送的上下文信息可以是响应于该请求,提供给语音动作系统300的信息。在又一实例中,与安装在用户设备304上的应用相关联的信息,诸如应用元数据可以指定提供给语音动作系统300的上下文信息,以及用户设备304可以将上下文信息传送到包括由应用信息指定的信息的语音动作系统300。

[0082] 例如,在图3所示的示例中,由用户设备304传送到语音动作系统300的上下文信息可以包括指示媒体播放器应用目前正在用户设备304的前台中操作、媒体播放器应用在音频播放器模式中操作,以及用户设备304的当前电池寿命为50%的信息。

[0083] 可以由语音动作系统300的语音识别引擎330接收由用户设备304传送的音频数据,并且语音识别引擎330可以生成语音音频数据的转录抄本。例如,如关于图1的语音识别引擎130所述,语音识别引擎330能够接收音频数据并且生成对应于用户305的语音输入的音频数据的转录抄本。在生成音频数据的转录抄本后,语音识别引擎330能将转录抄本提供给匹配器350。

[0084] 在一些实现方式中,除接收音频数据外,语音识别引擎330还可以通过用户设备304,接收提供给语音动作系统300的上下文信息。语音识别引擎330可以基于接收的上下文信息,偏置语音识别。在一些实现方式中,基于上下文信息的偏置语音识别可以包括调整音频数据中检测到某些词的概率。例如,语音识别引擎330可以基于与由上下文信息指示的上下文的相关性,调整检测某些词的概率。

[0085] 例如,语音识别引擎330可以接收对应于话语“播放下一首”的音频数据以及指示媒体播放器应用目前在用户设备304的前台中运行的上下文信息。语音识别引擎330可以基于上下文信息,偏置语音识别,使得基于语音识别引擎330基于确定那些词与在用户设备304的前台中运行的媒体播放器应用相关或缺少关联,增加在音频数据中检测到词“播放(play)”或短语“播放下一首”的概率,或减小检测到词“星球(Planets)”的概率,生成“播放

下一首”的转录抄本,代替“星球”的转录抄本。

[0086] 语音动作系统的上下文分析引擎340接收由用户设备304传送的上下文信息并且处理该上下文信息以确定用户设备304或在用户设备304上运行的应用的上下文。与图1的上下文分析引擎140类似,语音动作系统300的上下文分析引擎340能处理接收的上下文信息来确定与识别待触发的语音动作有关的接收的上下文信息的部分,或能处理接收的上下文信息来确定是否有用户设备304识别触发的语音动作所需的其他上下文信息。如果需要附加上下文信息来确定执行的语音动作,上下文分析引擎340能从用户设备304请求附加上下文信息,使得用户设备304能利用附加上下文信息响应于该请求。在一些实现方式中,上下文分析引擎340能修改接收的上下文信息的格式来生成能由语音动作系统300用于识别触发的语音动作的上下文。上下文分析引擎340将处理的上下文信息提供给匹配器350。

[0087] 在图3所示的示例中,例如,上下文分析引擎340从用户设备304接收上下文信息。接收的上下文信息可以指示媒体播放器应用在用户设备304的前台中运行、媒体播放器应用在音频播放器模式中操作,以及用户设备304的电池寿命处于50%。上下文分析引擎340可以确定电池寿命信息与响应于用户305的语音输入,确定触发哪一语音动作无关,因此,上下文分析引擎340可以将上下文提供给仅指示媒体播放器应用在用户设备304的前台中操作并且处于音频播放器模式中的匹配器350。

[0088] 匹配器350从上下文分析引擎340接收上下文并且从语音识别引擎330接收转录抄本,并且使用接收的上下文和转录抄本来识别触发的一个或多个候选语音动作。例如,匹配器350可以将上下文分析引擎340接收的上下文与包括在应用语音动作数据库310和/或OS语音动作数据库320中的一个或多个意图相关联的上下文匹配。基于确定接收的上下文与一个或多个意图相关联的上下文匹配,匹配器350可以将那些意图识别为用于触发的候选意图。在基于将接收的上下文与候选意图相关联的上下文匹配,识别候选意图后,匹配器350可以通过将接收的转录抄本与候选意图相关联的触发短语比较,缩小候选意图的集合。匹配器350可以将识别缩小的候选意图的集合的信息提供给动作触发生成器360。

[0089] 在一些实例中,匹配器350可以接收语音输入的多个候选转录抄本,并且可以基于确定候选意图的每一个与从上下文分析引擎340接收的上下文和/或从语音识别引擎330接收的语音输入的一个或多个候选转录抄本匹配的上下文,识别缩小的候选意图的集合。例如,语音识别引擎330可以确定用于语音输入的多个候选转录抄本,并且可以将指示候选转录抄本的每一个的信息提供给匹配器350。然后,匹配器350可以识别每一个指定与由匹配器350接收的上下文兼容的、还指定匹配于语音输入的候选转录抄本的触发短语的上下文的缩小的候选意图的集合。

[0090] 在图3所示的示例中,匹配器350接收指示用户设备304具有在前台中操作的媒体播放器并且处于音频播放器模式的上下文。匹配器350还接收音频数据的两个转录抄本,即,转录抄本“播放下一首”和“星球”。匹配器350通过将接收的上下文和转录抄本与在应用语音动作数据库310和OS语音动作数据库320处存储的意图相关联的上下文和触发短语比较,识别候选意图来基于该信息激发。基于该比较,匹配器350识别两个候选意图,包括使媒体播放器应用跳到下一首歌的、用于“播放下一首”语音动作的意图,以及用于使称为“星球”的应用打开的、用于“星球”语音动作的意图。匹配器350能将信息提供给识别两个候选意图的动作触发生成器360,或匹配器350能将候选意图的拷贝提供给动作触发生成器360。

[0091] 动作触发生成器360接收指示响应于语音输入激发的一个或多个候选意图和从用户设备304接收的上下文信息的信息。动作触发生成器360对一个或多个特定意图生成动作触发,其中,每一动作触发包括能够控制用户设备304执行与相应的意图相关联的语音动作的数据。在一些实现方式中,动作触发生成器360生成用于由匹配器350识别的每一候选意图的动作触发。在其他实现方式中,动作触发生成器360可以从候选意图中,获得用于触发的特定意图的指示,并且可以生成仅用于该特定意图的动作触发。在任一实现方式中,动作触发生成器360可以基于从匹配器350接收的信息和在应用数据数据库370处访问的信息,生成动作触发。

[0092] 例如,动作触发生成器360可以接收用于候选语音动作的意图,以及动作触发生成器360能够访问应用数据数据库370处,使得在用户设备304处执行候选语音动作的信息。例如,在应用数据数据库370处访问的信息能包括用于执行或触发与语音动作相关联的活动或动作的信息,诸如有关用户设备304的信息,有关与语音动作相关联的应用的信息或有关在用户设备304上运行的操作系统的信息。动作触发生成器360能将访问的信息用作用于候选语音动作的动作触发,或能使用访问的信息来生成能够在用户设备304处,触发与候选语音动作相关联的动作或活动的动作触发数据。在一些实现方式中,语音动作系统300可以不包括动作触发生成器360或应用数据数据库370,使得可以将意图本身传送到用户设备304并且意图中的信息能使用户设备304执行与该意图相关联的活动或动作。

[0093] 在图3所示的示例中,动作触发生成器360可以接收用于媒体播放器应用的“播放下一首”语音动作的意图和用于打开“星球”应用的“星球”语音动作的意图。动作触发生成器360可以生成用于“播放下一首”和“星球”语音动作的每一个的动作触发。为生成用于“播放下一首”语音动作的动作触发,动作触发生成器360访问应用数据数据库370处,属于安装在用户设备304上的媒体播放器应用的信息,可以访问属于用户设备304或特定模型的用户设备304的信息,或可以访问属于在用户设备304上运行的操作系统的信息。动作触发生成器360使用接收的意图和在应用数据数据库370处访问的信息来生成用于“播放下一首”语音动作的动作触发。类似地,动作触发生成器360可以通过访问与“星球”应用、用户设备304或在用户设备304上运行的操作系统有关的信息,生成用于“星球”语音动作的动作触发,并且基于接收和访问的信息,生成用于“星球”语音动作的动作触发。

[0094] 如上所述,在一些示例中,匹配器350可以识别用于触发的一个以上的候选意图,并且在那些实例中,有必要从候选的集合中选择用于触发的特定意图。消除歧义引擎380能执行选择或接收从候选意图中选择的用户输入的操作。例如,消除歧义引擎380能将请求输出到要求用户305从与候选意图相关联的候选语音动作中选择特定语音动作的用户设备304。响应于该请求,用户305能将输入提供给用户设备304,以选择用于执行的特定候选语音动作。能由消除歧义引擎380接收指示用户选择的信息。

[0095] 替代地,消除歧义引擎380可以访问包括关于提交给用户设备304的先前语音动作或在用户设备304处执行的其他动作的信息的用户活动历史390。例如,用户活动历史390能指示用户305触发特定候选语音动作的次数、用户305在用户设备304处通过或不通过使用语音动作,执行特定动作的次数,或能够包括关于用户305使用用户设备304的其他信息。在一些实例中,用户活动历史390能包括指示除用户305外的用户的使用历史的信息。例如,用户活动历史390能包括指示具有安装在用户设备上的媒体播放器应用和/或“星球”应用的

所有用户的使用模式的信息。

[0096] 在一些实现方式中,用户活动历史390中的信息能由语音动作系统300使用来从候选意图中选择,而不请求从用户305的输入。例如,当提供类似语音输入时,消除歧义引擎380可以访问用户活动历史390处、指示用户305先前已经选择特定候选语音动作的信息,并且基于那一信息,消除歧义引擎380可以选择触发的特定候选意图,而不请求从用户305的输入。类似地,消除歧义引擎380可以确定已经提供与由用户305接收的语音输入类似的语音输入的一定比例或数量的用户已经选择触发特定语音动作,并且消除歧义引擎380可以基于那一信息,无需请求从用户305的输入,选择触发的特定候选语音意图。在一些实现方式中,候选意图的每一个具有指示特定触发短语和特定候选意图之间的关系强度的相关得分。可以基于指示候选意图和接收的触发短语之间的关系的强度的分配给候选意图的得分,选择特定候选意图。

[0097] 其他因素可以被考虑或包括在用于选择候选意图的用户活动历史390中。例如,在语音动作系统触发某一候选意图后,语音动作系统300可以跟踪用户设备304处的行为以确定触发的意图是否看起来与用户305的预期匹配。例如,如果在语音动作系统300触发某一语音动作后,用户305采取在用户设备304处执行另一动作的操作,指示语音动作系统300触发了错误的候选意图,因此,语音动作系统300可以将指示接收的语音输入不应当触发或应当不希望触发该触发的意图的信息添加到用户活动历史390。在其他示例中,用户活动历史390可以包括指示用户305或其他用户对他们的用户设备的给定的状态,例如,对在提供给语音动作系统300的上下文信息中指示的特定状态,有多频繁地选择特定候选意图的信息。包括在用户活动历史390中的其他信息能由消除歧义引擎380访问并且用来从候选意图的集合选择特定候选意图。

[0098] 对图3中所示的示例,基于确定用于由用户305提供的语音动作的候选意图包括用于媒体播放器应用的“播放下一首”语音动作的意图或用于打开“星球”应用的“星球”语音动作的意图,消除歧义引擎380能将请求特定候选语音动作的用户选择的信息提供给用户305。如所示,能在请求用户305选择“下一首歌”语音动作,或“打开‘星球’”的语音动作的用户设备304的显示器处输出请求。用户305在用户设备304处提供选择“下一首歌”选项的输入。指示用户选择的数据被传送到消除歧义引擎380,并且在接收指示用户选择的数据中,消除歧义引擎380确定触发用于“播放下一首”语音动作的意图。在一些示例中,消除歧义引擎380还可以将指示用户305已选取以在用于“星球”语音动作的意图上,触发用于“播放下一首”语音动作的意图的信息存储在用户活动历史390处。消除歧义引擎380可以依赖在用户活动历史390处存储的信息来自动地确定候选语音动作以触发是否稍后由语音动作系统300,从用户设备304接收类似的语音输入。

[0099] 在图3所示的示例的替选实现方式中,消除歧义引擎380可以基于访问用户活动历史390,选择触发的特定候选意图。例如,用户活动历史390可以指示用户设备304的用户305先前已经提供类似的语音输入并且已选取以触发用于“下一首歌”语音动作的意图。基于指示在接收类似的语音输入后,先前触发用于“播放下一首”语音动作的意图的用户活动历史390,消除歧义引擎380可以确定以触发用于“播放下一首”语音动作的意图,而不要求从用户305的附加输入。

[0100] 在已经选择特定候选意图后,如果需要,已经生成用于被选候选语音动作的动作



触发,语音动作系统300将使得执行选择的语音动作的信息传送到用户设备304。例如,语音动作系统300的消除歧义引擎380或其他组件能将用于选择的语音动作的意图或动作触发数据传送到用户设备304来使用户设备304执行选择的语音动作。

[0101] 在图3所示的示例中,例如,在用户305已经选择“下一首歌”选项来使媒体播放器应用跳到下一首歌的选择“播放下一首”语音动作后,并且在已经生成包括用于控制媒体播放器应用跳到下一首歌的数据的动作触发后,消除歧义引擎380将动作触发传送到用户设备304。用户设备304能接收动作触发,并且动作触发能通过使媒体播放器应用跳到下一首歌,使用户设备304执行“播放下一首”语音动作。

[0102] 在一些实现方式中,可以提供指示触发的语音动作的通知以在用户设备304处输出。例如,语音动作系统300或用户设备304的OS可以使识别已经触发的语音动作的通知在用户设备304处被输出。在一些实例中,可以以对用户305来说,很明确与触发的语音动作相关联的应用不负责检测语音动作或触发语音动作的方式,提供识别触发的语音动作的通知。例如,在用户设备304处提供的通知可以使用可识别为与语音动作系统300、用户设备304或用户设备304的OS相关联,并且与在用户设备304上操作的媒体播放器应用不相关联的语音语调。在另一示例中,以将处理来自用户305的语音输入的系统或应用与执行触发的语音动作的应用区分的方式,呈现通知。例如,通知可以声明“你已经触发“播放下一首”语音动作。我现在将使媒体播放器应用跳到下一首歌”,由此向用户305指示媒体播放器应用不负责处理语音输入。

[0103] 图4图示用于由语音动作系统执行的示例过程400的流程图。在一些示例中,图4的过程400能由图1的语音动作系统100,或由图1的语音动作系统100的组件执行。过程400描述一过程,通过该过程,应用开发人员能建立用于已经安装在一个或多个用户设备上的应用的新语音动作,安装该应用的用户设备的用户能触发新语音动作来使得与新语音动作相关联的操作由用户设备上的应用执行。

[0104] 初始地,由对不同语音动作系统的软件应用指定新语音动作的语音动作系统,接收数据,其中,该数据包括用于执行新语音动作的一个或多个操作和用于触发新语音动作的一个或多个触发项(402)。例如,使用终端102a-102n,与软件应用相关联的应用开发人员101a-101n能将定义新语音动作的数据提交给语音动作系统100。

[0105] 定义新语音动作的数据能包括指定用于操作新语音动作的一个或多个操作和用于触发新语音动作的一个或多个触发项的数据。例如,定义新语音动作的数据可以指定触发项“播放下一首”并且可以指定新语音动作应当执行使媒体播放器应用跳到下一首歌的操作。该数据还可以指定与新语音动作有关的附加信息,诸如新语音动作所属的应用、用户应当何时启用新语音动作以使用的上下文,或其他信息。例如,定义“播放下一首”语音动作的数据还可以指定语音动作属于媒体播放器应用,并且可以指定当媒体播放器应用在前台中操作并且处于音频播放器模式时,应当启用语音动作。

[0106] 至少基于接收的数据由语音动作系统生成用于软件应用的上下文语音动作意图,其中,上下文语音动作意图包括当由软件应用接收时,请求软件应用执行新语音动作中的一个或多个操作的数据(404)。例如,图1的语音动作系统100可以接收指定新语音动作的数据,并且可以基于接收的数据,生成用于新语音动作的上下文语音动作意图。语音动作系统100可以将上下文语音动作意图存储在应用语音动作数据库110中。在一些实现方式中,语



音动作系统100可以使用与类似参考图2所述的过程,通过验证接收的数据和引入接收的语法来生成上下文语音动作意图,生成该意图。在一些实现方式中,所生成的上下文语音动作意图能存储控制软件应用来执行用于执行新语音动作的一个或多个操作的信息。

[0107] 通过语音动作系统,上下文语音动作意图与用于新语音动作的一个或多个触发项相关联(406)。例如,在由语音动作系统100创建用于新语音动作的上下文语音动作意图后,语音动作系统100能将用于新语音动作的上下文语音动作意图与在接收的数据中指定的一个或多个触发项相关联。具有关联的一个或多个触发项的上下文语音动作意图可以被存储在应用语音动作数据库110中。在一些实现方式中,将一个或多个触发项与上下文语音动作意图相关联能包含指定上下文语音动作意图中的一个或多个触发项。在其他实施方式中,关联会受将一个或多个触发项与上下文语音动作意图链接影响,使得接收一个或多个触发项中的一个会导致上下文语音动作意图的识别。

[0108] 语音动作系统被配置为接收由安装有软件应用的设备获得的用户话语的指示(408)。例如,语音动作系统100可以被配置为接收已经从用户设备104接收话语的指示,其中,话语由用户设备104获得作为来自用户105的语音输入。用户设备104还具有与已经安装的新语音相关联的软件应用,例如参考图1所述的媒体播放器应用。

[0109] 语音动作系统被配置为确定用户话语的转录抄本对应于与上下文语音动作意图相关联的一个或多个触发项(410)。例如,语音动作系统100可以被配置为接收对应于用户话语的音频数据,并且能使用语音动作系统100的语音识别引擎130,生成用户话语的转录抄本。由语音识别引擎130生成的用户话语的转录抄本能与多个上下文语音动作意图,包括用于新语音动作的上下文语音动作意图和用于其他语音动作的上下文语音动作意图比较。

[0110] 在一些实例中,可以由图1的语音动作选择器150执行比较,其中,语音动作选择器150能将用户话语的转录抄本与在应用语音动作数据库110或OS语音动作数据库130处存储的上下文语音动作意图的触发项比较。基于该比较,语音动作选择器150可以确定用户话语的转录抄本与新语音动作的上下文语音动作意图相关联的一个或多个触发项中的一个或多个匹配。例如,语音识别引擎130可以生成用于从读取“播放下一首”的用户设备104接收的用户话语的转录抄本。语音动作选择器150将“播放下一首”语音动作与候选上下文语音动作意图相关联的触发项比较来确定该转录抄本与用于新的“播放下一首”语音动作的上下文语音动作意图相关联的触发短语“播放下一首”匹配。

[0111] 语音动作系统被配置成将上下文语音动作意图提供给特定设备,由此请求安装在特定设备上的软件应用执行新语音动作中的一个或多个操作(412)。例如,响应于确定由语音动作系统100的语音识别引擎130生成的用户话语的转录抄本与用于新语音动作的上下文语音动作意图相关联的一个或多个触发项匹配,语音动作选择器150能将用于新语音动作的上下文语音动作意图提供给用户设备104。用于新语音动作的上下文语音动作意图能有效地请求安装在用户设备104上的软件应用执行新语音动作的一个或多个操作。例如,上下文语音动作意图可以包括在用户设备104接收上下文语音动作意图时,触发与新语音动作相关联的一个或多个应用的操作。

[0112] 已经描述了多个实现方式。不必说,将理解在不背离本公开的精神和范围的情况下,可以做出各种修改。例如,可以使用如上所示的各种形式的流程,重新排序、添加或移除步骤。因此,其他实现方式在下述权利要求的范围内。

[0113] 例如,本文所述的系统和/或方法可以收集有关用户的个人信息,或可以使用个人信息,可以为用户提供控制程序或特征是否收集个人信息,例如,关于用户社交网络、社交动作或活动、职业、用户偏好或用户当前位置的信息,或控制系统和/或方法是否和/或如何执行与用户更相关的操作的机会。此外,某些数据可以在存储或使用其之前依一种或多种方式匿名化,使得移除个人可识别的信息。例如,可以匿名化用户身份使得无法确定用户的个人可识别信息,或获得地理位置信息的用户位置可以被一般化(诸如到城市、ZIP编码或州级),使得无法确定用户的特定位置。因此,用户可以具有对如何收集关于他或她的信息和如何使用所述信息的控制。

[0114] 尽管关于开发用于与安装在用户设备上的应用一起使用的语音动作,突出描述上述实施例,也可以关于机器、其他设备、机器人或其他系统,使用所述的特征。例如,可以利用图1的语音动作系统100来开发和实现用于与机器交互的语音动作,其中,机器具有相关的计算系统,可以被用来开发和实现用于与机器人或具有机器组件的系统交互的语音动作,可以被用来开发和实现用于与仪器、娱乐系统或其他设备交互的语音动作,或可以用来开发和实现用于与车辆或其他交通系统交互的语音动作。

[0115] 可以在数字电子电路中或者在包括本说明书中公开的结构及其结构等效物的计算机软件、固件或者硬件中或者在它们中的一项或者多项的组合中实现本说明书中描述的实施例和所有功能操作。实施例可以被实现为一个或者多个计算机程序产品、即在计算机可读介质上编码的用于由数据处理装置执行或者控制数据处理装置的操作的计算机程序指令的一个或者多个模块。计算机可读介质可以是机器可读存储设备、机器可读存储基板、存储器设备、影响机器可读传播信号的物质组成或者它们中的一项或者多项的组合。术语“数据处理装置”涵盖了用于处理数据的所有装置、设备和机器、例如包括可编程处理器、计算机或者多个处理器或者计算机。装置除了包括硬件之外还可以包括为讨论的计算机程序创建执行环境的代码、例如构成处理器固件、协议栈、数据库管理系统、操作系统或者它们中的一项或者多项的组合的代码。传播信号是人为生成的信号、例如由机器生成的电、光学或者电磁信号,该信号被生成以用于对信息进行编码一部向适当接收器装置发送。

[0116] 可以用包括编译或者解释语言的任何形式的编程语言编写计算机程序(也称为程序、软件、软件应用、脚本或者代码),并且可以用任何形式部署它、包括作为独立程序或者作为适合用于在计算环境中使用的模块、组件、子例程或者其它单元。计算机程序不一定对应于文件系统中的文件。程序可以存储于保持其它程序或者数据的文件的部分中(例如存储于标记语言文档中的一个或者多个脚本)、专用于讨论的程序的单个文件中或者多个协同文件(例如存储一个或者多个模块、子程序或者代码部分的文件)中。计算机程序可以被部署成在一个计算机上或者在位于一个地点或者分布于多个地点并且由通信网络互连的多个计算机上执行。

[0117] 在本说明书中描述的过程和逻辑流程可以由一个或者多个可编程处理器执行,该处理器执行一个或者多个计算机程序以通过对输入数据操作并且生成输出来执行功能。过程和逻辑流程也可以由专用逻辑电路,如FPGA(现场可编程门阵列)或者ASIC(专用集成电路)执行,并且装置也可以实现为该专用逻辑电路,如FPGA(现场可编程门阵列)或者ASIC(专用集成电路)。

[0118] 适合于执行计算机程序的处理器例如包括通用和专用微处理器两者和任何种类

的数字计算机的任何一个或者多个处理器。一般而言,处理器将从只读存储器或者随机存取存储器或者这二者接收指令和数据。

[0119] 计算机的基本元件是用于执行指令的处理器以及用于存储指令和数据的一个或者多个存储器设备。一般而言,计算机也将包括用于存储数据的一个或者多个海量存储设备如磁盘、光磁盘或者光盘或者操作地耦合从该海量存储设备接收数据或者向该海量存储设备传送数据或者这二者。然而计算机无需具有这样的设备。另外,计算机可以嵌入于另一设备中,仅举数例,该另一设备例如是平板计算机、移动电话、个人数字助理(PDA)、移动音频播放器、全球定位系统(GPS)接收器。适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储器设备、例如包括半导体存储器设备如EPROM、EEPROM和闪存存储器设备;磁盘如内部硬盘或者可拆卸盘;磁光盘;以及CD ROM和DVD-ROM盘。处理器和存储器可以由专用逻辑电路补充或者并入于专用逻辑电路中。

[0120] 为了提供与用户的交互,实施例可以实限于计算机上,该计算机具有用于向用户显示信息的显示器设备如CRT(阴极射线管)或者LCD(液晶显示器)监视器以及用户可以用来向计算机提供输入的键盘和指向设备如鼠标或者跟踪球。其它种类的设备也可以用来提供与用户的交互;例如向用户提供的反馈可以是任何形式的感官反馈如视觉反馈、听觉反馈或者触觉反馈;并且可以用包括声音、语音或者触觉输入的任何形式接收来自用户的输入。

[0121] 实施例可以实现于计算系统中,该计算系统包括后端组件如作为数据服务器或者包括中间件组件如应用服务器或者包括前端组件、例如具有用户通过其实现方式交互的图形用户界面或者网页浏览器的客户端计算机或者这样的后端、中间件或者前端组件中的一个或者多个组件的任何组合。系统的组件可以由任何数字数据通信形式或者介质如通信网络互连。通信网络的示例包括局域网(“LAN”)和广域网(“WAN”)如因特网。

[0122] 计算系统可以包括客户端和服务端。客户端和服务端一般相互远离并且通常通过通信网络交互。客户端和服务端的关系借助计算机程序来发生,这些计算机程序在相应计算机上运行并且相互具有客户端-服务器关系。

[0123] 尽管本说明书包含许多细节,但是这些不应解释为对公开内容的范围或者要求保护的内容的范围的限制、而应当作为对特定于实施例的特征的描述。也可以在单个实施例组合实现本说明书中在单独实施例的上下文中描述的某些特征。换言之,也可以在多个实施例中单独或者在任何适当子组合中实现在单个实施例的上下文中描述的各种特征。另外虽然上文可以将特征描述为在某些组合中作用并且甚至起初这样要求保护,但是在一些情况下可以从要求保护的组合中去除来自该组合的一个或者多个特征,并且要求保护的组合可以涉及子组合或者子组合的变形。

[0124] 类似地,尽管在附图中以特定顺序描绘操作,但是这不应理解为要求以所示特定顺序或者以序列顺序执行这样的操作或者执行所有所示操作以实现希望的结果。在某些境况中,多任务和并行处理可以是有利的。另外,在上文描述的实施例中分离各种系统组件不应理解为在所有实施例中要求这样的分离,并且应当理解描述的程序组件和系统一般可以一起集成于单个软件产品中或者封装到多个软件产品中。

[0125] 在其中提到HTML文件的每个实例中,可以替换为其它文件类型或者格式。例如HTML文件可以替换为XML、JSON、纯文本或者文件的其它类型。另外当提到表或者散列表时,

可以使用其它数据结构(比如电子数据表、关系数据库或者结构化文件)。

[0126] 由此,已经描述特定实施例。其它实施例在所附权利要求的范围内。例如在权利要求中记载的动作可以按不同顺序来执行并且仍然获得希望的结果。

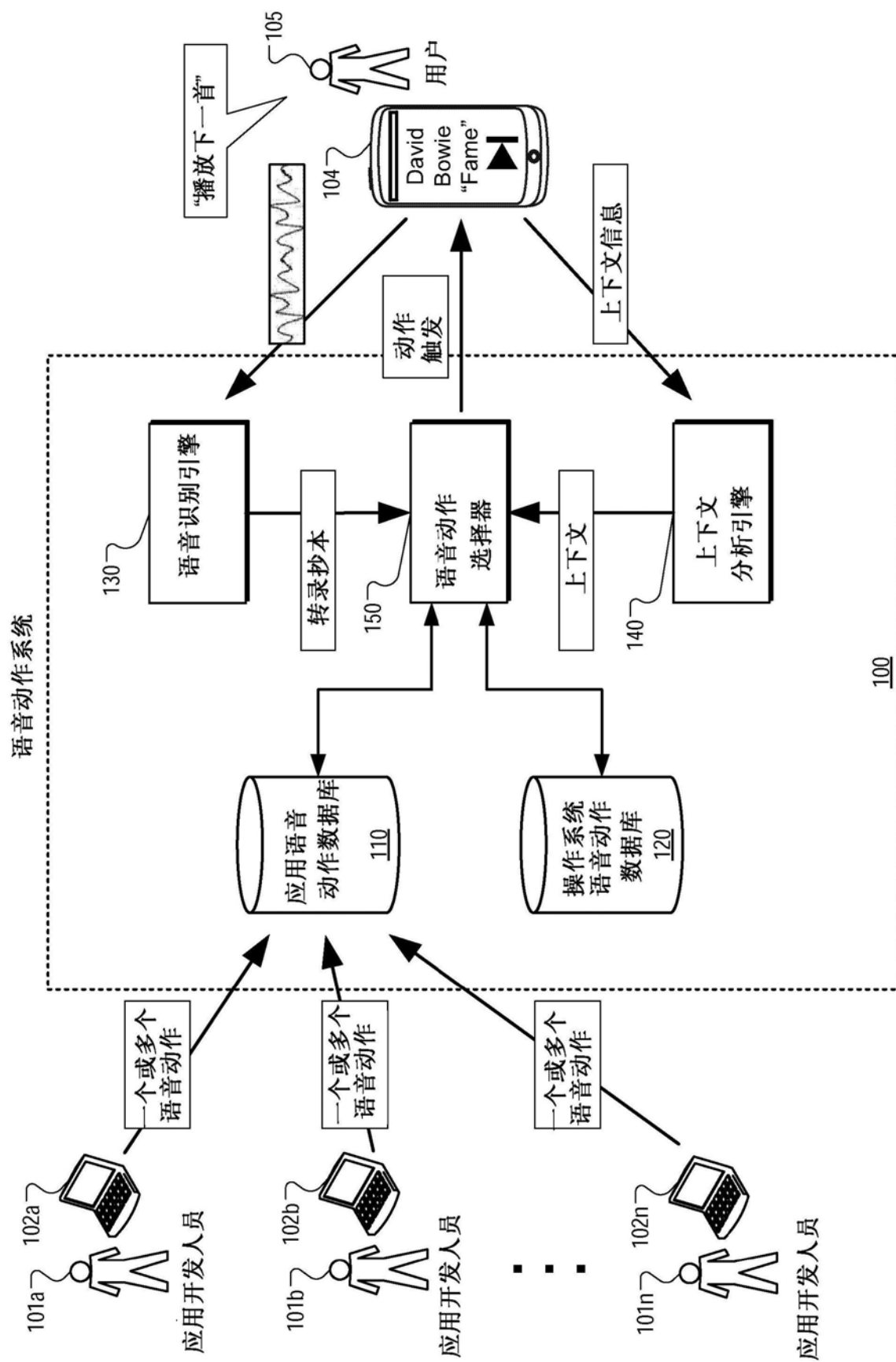


图1

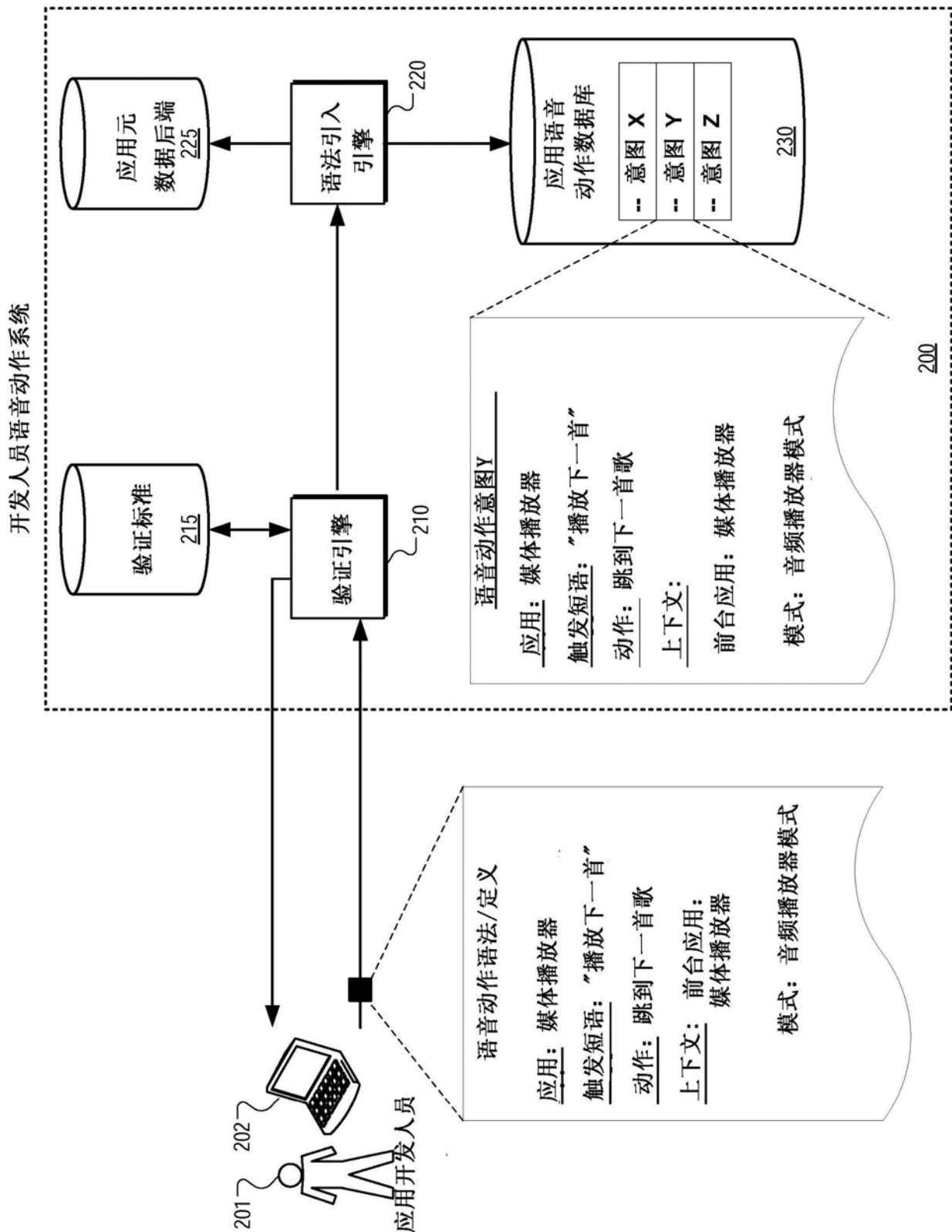


图2

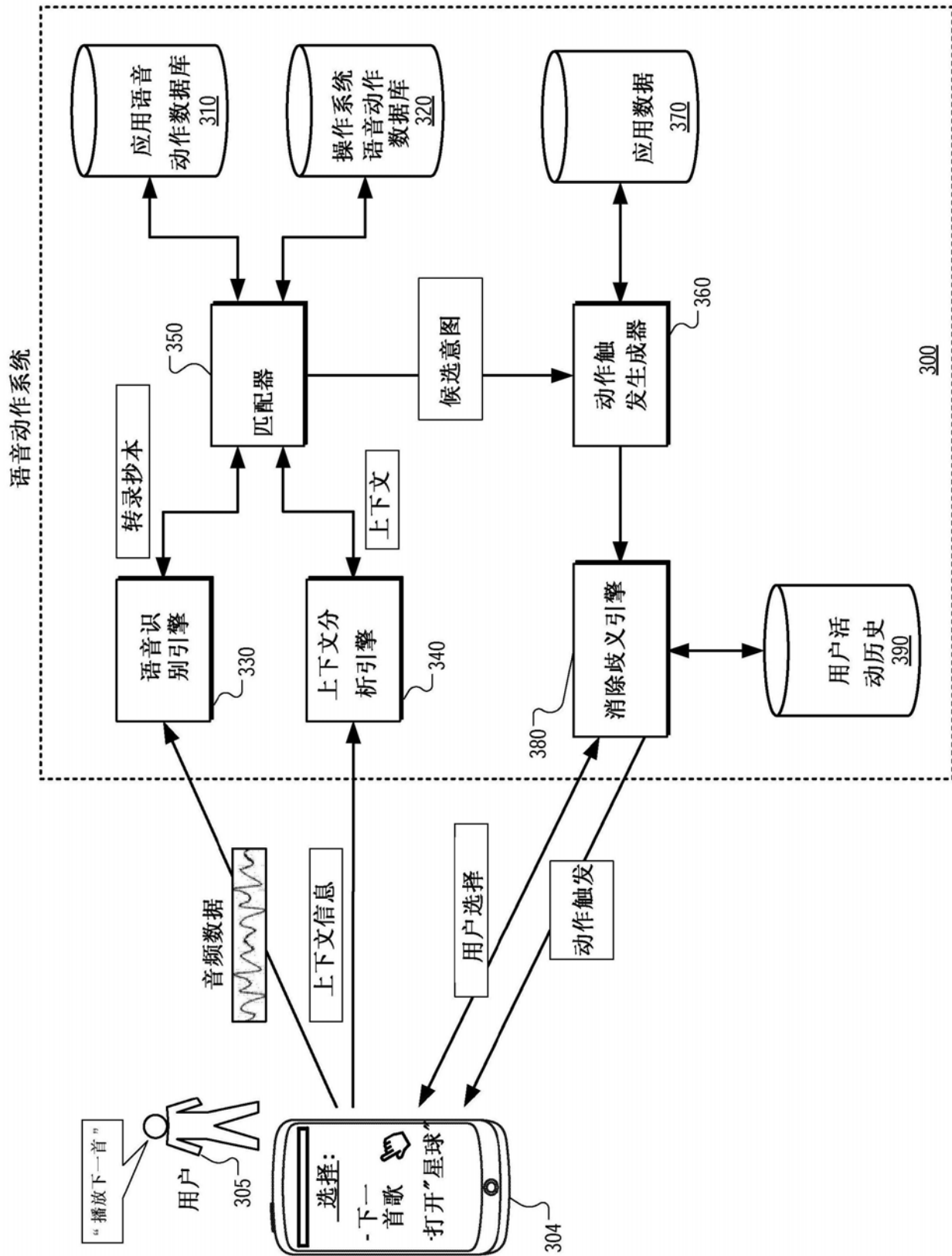


图3

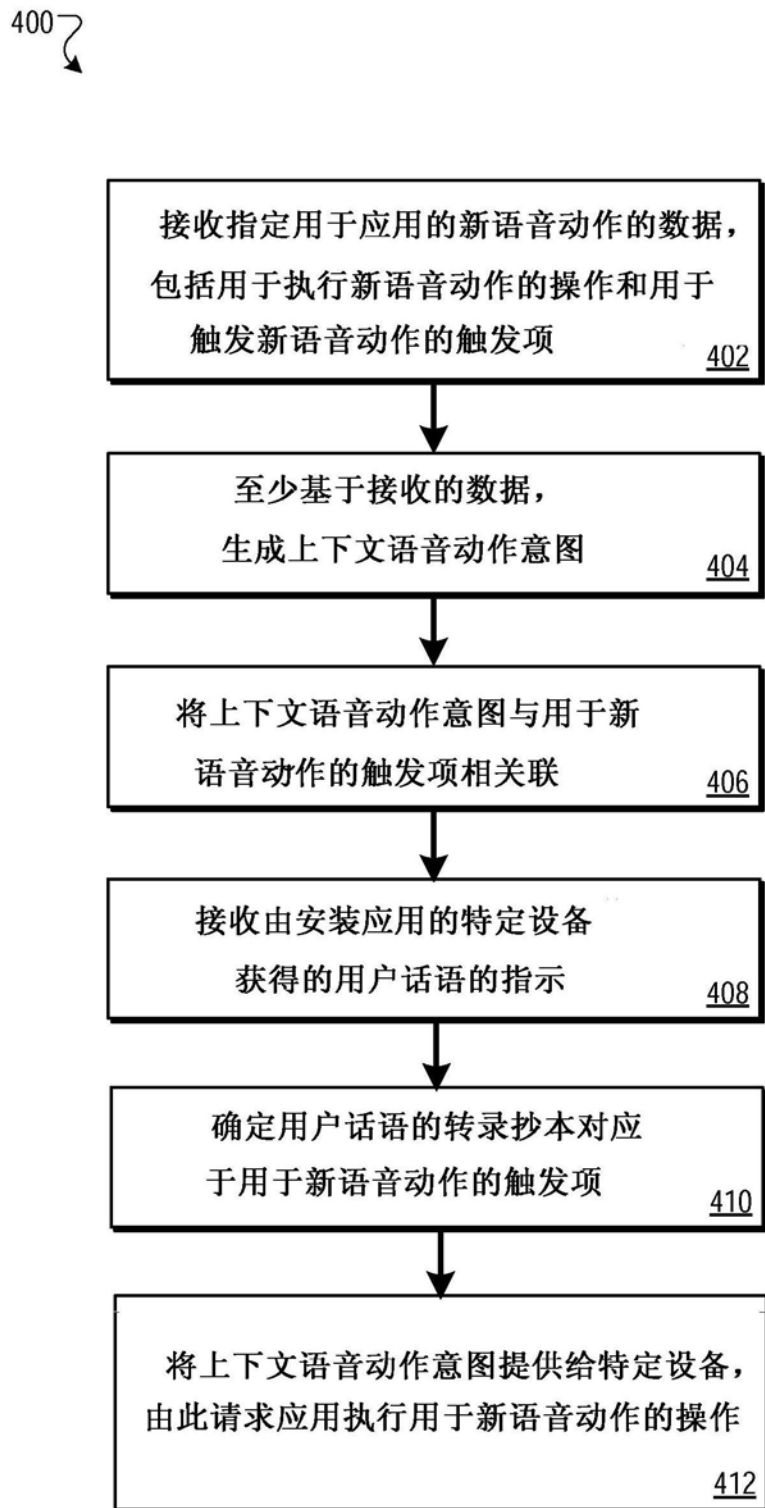


图4