



(12) 发明专利申请

(10) 申请公布号 CN 104335207 A

(43) 申请公布日 2015. 02. 04

(21) 申请号 201380028728. 9

(51) Int. Cl.

(22) 申请日 2013. 06. 25

G06F 17/30 (2006. 01)

(30) 优先权数据

G06F 3/16 (2006. 01)

61/666, 659 2012. 06. 29 US

13/797, 937 2013. 03. 12 US

(85) PCT国际申请进入国家阶段日

2014. 11. 28

(86) PCT国际申请的申请数据

PCT/US2013/047668 2013. 06. 25

(87) PCT国际申请的公布数据

W02014/004544 EN 2014. 01. 03

(71) 申请人 苹果公司

地址 美国加利福尼亚州

(72) 发明人 C · B · 弗莱扎克 T · R · 格鲁伯

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华

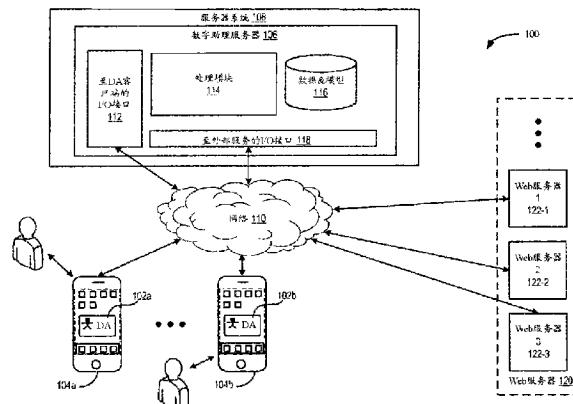
权利要求书3页 说明书21页 附图23页

(54) 发明名称

用于文档的语音激活导航和浏览的设备、方法和用户界面

(57) 摘要

本发明公开了具有存储器和一个或多个处理器的电子设备，该电子设备接收包括多个链接的第一文档。该电子设备输出对第一文档的至少一部分的语音阅读并且输出识别所述多个链接中的一个链接的可听信息。响应于输出识别所述链接的可听信息，该电子设备从用户接收关于链接的语音命令，并且响应于从用户接收到语音命令，输出对第二文档的与链接相关联的至少一部分的语音阅读。



1. 一种用于导航通过文档的方法,包括:

在具有存储器和一个或多个处理器的电子设备处:

接收包括多个链接的第一文档;

输出对所述第一文档的至少一部分的语音阅读;

输出识别所述多个链接中的一个链接的可听信息;

响应于输出识别所述链接的所述可听信息,从所述用户接收关于所述链接的语音命令;以及,

响应于从所述用户接收到所述语音命令,输出对第二文档的与所述链接相关联的至少一部分的语音阅读。

2. 根据权利要求1所述的方法,其中所述链接与所述第一文档的所述部分中的文本相关联。

3. 根据权利要求1-2中任一项所述的方法,包括:

响应于从所述用户接收到关于所述链接的所述语音命令,存储有关所述链接的信息;以及

输出对所述第二文档的一个或多个部分的语音阅读。

4. 根据权利要求3所述的方法,包括:

在输出对所述第二文档的至少一部分的所述语音阅读之后:

从所述用户接收请求导航回到所述第一文档的语音命令;以及

响应于从所述用户接收到请求导航回到所述第一文档的所述语音命令,输出对所述第一文档的继所述第一文档中的与所述链接相关联的文本之后的一个或多个部分的语音阅读。

5. 根据权利要求3所述的方法,包括:

在输出对所述第二文档的至少一部分的所述语音阅读之后,输出有关导航回到所述第一文档的语音询问;

响应于输出有关导航回到所述第一文档的所述语音询问,从所述用户接收请求导航回到所述第一文档的语音命令;以及

响应于从所述用户接收到请求导航回到所述第一文档的所述语音命令,输出对所述第一文档的包括所述第一文档中的与所述链接相关联的文本的一个或多个部分的语音阅读。

6. 根据权利要求3所述的方法,包括:

在输出对所述第二文档的至少一部分的所述语音阅读之后,响应于从所述用户接收到请求导航回到所述第一文档的所述语音命令,自动输出对所述第一文档的继所述第一文档中的与所述链接相关联的文本之后的一个或多个部分的语音阅读。

7. 根据权利要求1-6中任一项所述的方法,其中所述语音命令为第一类型的语音命令,所述方法包括:

从所述用户接收关于所述链接的不同于所述第一类型的第二类型的语音命令;以及

响应于从所述用户接收到关于所述链接的所述第二类型的所述语音命令:

从所述第二文档检索信息;以及

基于来自所述第二文档的所述信息,输出对应于所述第二类型的所述语音命令的语音信息。

8. 根据权利要求 7 所述的方法, 其中所述第二类型的所述语音命令包括对有关所述第二文档的作者的信息的请求。

9. 根据权利要求 7-8 中任一项所述的方法, 其中所述第二类型的所述语音命令包括对所述第二文档的概要的请求。

10. 根据权利要求 1-9 中任一项所述的方法, 还包括 :

将两个或更多个链接识别为对应于来自所述用户的语音命令的候选链接 ;

输出有关所述候选链接的可听信息 ; 以及

从所述用户接收对所述候选链接中的单个链接的选择。

11. 一种用于浏览文档的方法, 包括 :

在具有存储器和一个或多个处理器的电子设备处 :

接收具有多个部分的所述文档, 其中所述多个部分中的至少一些部分与相应的元数据相关联 ;

输出对所述文档的相应部分的语音阅读, 包括基于所述相应的元数据可听地区分所述相应部分 ;

从用户接收请求导航至与特定元数据相关联的特定部分的语音命令 ; 以及,

响应于接收到所述语音命令, 输出对与所述特定元数据相关联的所述特定部分的语音阅读。

12. 根据权利要求 11 所述的方法, 还包括 :

通过如下方式根据所述相应部分的相应样式来向所述用户输出对所述文本的所述相应部分的语音阅读 :

使用第一组语音特征向所述用户输出对所述文档中的所述文本的第一部分的语音阅读 ; 以及

使用不同于所述第一组语音特征的第二组语音特征向所述用户输出对所述文档中的所述文本的第二部分的语音阅读。

13. 根据权利要求 11-12 中任一项所述的方法, 还包括 :

在输出对所述文档的一个或多个部分的所述语音阅读的同时, 从所述用户接收所述语音命令的至少一部分。

14. 根据权利要求 11-13 中任一项所述的方法, 其中所述文档的所述相应部分中的至少一个部分为所述文档的相应段落。

15. 根据权利要求 11-14 中任一项所述的方法, 其中所述文档的所述相应部分中的至少一个部分为所述文档的相应标题。

16. 根据权利要求 11-15 中任一项所述的方法, 其中所述文档的所述相应部分中的至少一个部分为所述文档的相应句子。

17. 根据权利要求 11-16 中任一项所述的方法, 其中所述文档的第一部分具有第一样式, 并且所述文档的第二部分具有不同于所述第一样式的第二样式。

18. 根据权利要求 11-17 中任一项所述的方法, 其中所述文档的每个部分与单个链接相关联。

19. 一种用于识别一个或多个文档的方法, 包括 :

在具有存储器和一个或多个处理器的电子设备处 :

输出对多个文档中的一个文档的至少一部分的语音阅读；

在输出所述语音阅读的同时，从用户接收请求对应于特定标准的文档的语音命令；以及，

响应于从所述用户接收到所述语音命令：

识别所述多个文档中的对应于所述特定标准的一个或多个文档；以及

输出对所述一个或多个识别的文档的相应文档的至少一部分的语音阅读。

20. 根据权利要求 19 所述的方法，其中所述特定标准要求所述一个或多个识别的文档由被所述用户所识别的一位或多为作者创作。

21. 根据权利要求 19-20 中任一项所述的方法，其中所述特定标准要求所述一个或多个识别的文档与特定文档相关联。

22. 根据权利要求 19-21 中任一项所述的方法，其中所述特定标准要求所述一个或多个识别的文档是对特定文档的回复。

23. 根据权利要求 19-22 中任一项所述的方法，其中所述特定标准要求所述一个或多个识别的文档包括来自相应作者的最新消息。

24. 根据权利要求 19-23 中任一项所述的方法，其中所述特定标准要求所述一个或多个识别的文档对应于由所述用户识别的特定日期范围。

25. 根据权利要求 19-24 中任一项所述的方法，还包括：

输出指示所述一个或多个识别的文档的数量的可听信息。

26. 根据权利要求 19-25 中任一项所述的方法，还包括：

输出指示所述相应文档为所述一个或多个识别的文档的最后一个文档的可听信息。

27. 一种便携式电子设备或计算机系统，包括：

一个或多个处理器；和

存储器，所述存储器存储由所述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权利要求 1-26 所述的方法中的任一种方法的指令。

28. 一种存储一个或多个程序的计算机可读存储介质，所述一个或多个程序包括当由具有一个或多个处理器的便携式电子设备或计算机系统执行时使得所述设备执行权利要求 1-26 的所述方法中的任一种方法的指令。

29. 一种便携式电子设备或计算机系统上的图形用户界面，所述便携式电子设备或计算机系统具有显示器、存储器、以及执行存储在所述存储器中的一个或多个程序的一个或多个处理器，所述图形用户界面包括根据权利要求 1-26 所述的方法中的任一种方法显示的用户界面。

30. 一种便携式电子设备或计算机系统，包括：

用于执行权利要求 1-26 所述的方法中的任一种方法的装置。

31. 一种用于便携式电子设备或计算机系统中使用的信息处理装置，包括：

用于执行权利要求 1-26 所述的方法中的任一种方法的装置。

32. 一种便携式电子设备或计算机系统，包括：

被配置为执行权利要求 1-26 所述的方法中的任一种方法的处理单元。

用于文档的语音激活导航和浏览的设备、方法和用户界面

技术领域

[0001] 所公开的实施例整体涉及数字助理系统，且更具体地涉及执行文档的语音激活导航和浏览的数字助理系统。

背景技术

[0002] 正如人类个人助理一样，数字助理系统可执行所请求的任务并提供所请求的建议、信息或服务。数字助理系统满足用户请求的能力取决于数字助理系统对请求或指令的正确理解。在自然语言处理方面的最新进展已使得用户能够使用口头或文本形式的自然语言来与数字助理系统进行交互。此类数字助理系统可解释用户的输入以推断用户的意图、将所推断出的意图转换成可执行的任务和参数、执行操作或部署服务以执行任务，以及产生可被用户理解的输出。

[0003] 此类数字助理系统可被配置为协助在与电子设备进行交互方面具有有限访问性的用户。例如，视力下降的人群诸如低视力用户和盲人用户、存在阅读障碍的用户或具有学习障碍的其他用户，或者甚至是仅仅希望或需要在操作期间不必注视设备便能够使用设备的有视力的用户均可受益于向用户阅读信息的数字助理系统。在另一个实例中，对于具有触摸屏的电子设备而言，动作技能有限的人群，诸如某根手指或某只手受损的那些人，如果在触摸屏上执行触摸手势不是不可能的话，也可能发现执行该触摸手势很困难。然而，数字助理系统可接收语音命令，从而消除对触摸手势的需求。

[0004] 然而，对文档的导航和浏览仍然是麻烦且效率低下的，从而对视力受损和 / 或动作技能有限的用户带来明显的认知负担。

发明内容

[0005] 如上所述，需要用于向视力受损和 / 或动作技能有限的用户提供用于文档的导航和浏览的改进的用户界面的数字助理系统。这使用户能够有效地导航通过并浏览文档。

[0006] 本文所公开的实施例提供了提供文档的语音激活导航和浏览的方法、系统和计算机可读存储介质。

[0007] 一些实施例提供了一种用于导航通过在具有存储器和一个或多个处理器的电子设备处执行的文档的方法。该方法包括接收含有多个链接的第一文档、输出对第一文档的至少一部分的语音阅读、输出识别所述多个链接中的一个链接的可听信息，以及响应于输出识别该链接的可听信息，从用户接收关于该链接的第一类型的语音命令。该方法还包括，响应于从用户接收到语音命令，输出对第二文档的与该链接相关联的至少一部分的语音阅读。

[0008] 根据一些实施例，一种用于浏览在具有存储器和一个或多个处理器的电子设备处执行的文档的方法包括接收具有多个部分的文档，其中所述部分中的至少一些部分与相应的元数据相关联。该方法还包括输出对文档的相应部分的语音阅读，包括基于相应的元数据可听地区分所述相应部分。该方法还包括从用户接收请求导航至与特定元数据相关联的

特定部分的语音命令；以及，响应于接收到该语音命令，输出对与特定元数据相关联的特定部分的语音阅读。

[0009] 根据一些实施例，一种用于识别在具有存储器和一个或多个处理器的电子设备处执行的一组文档的方法包括输出对多个文档中的一个文档的至少一部分的语音阅读。该方法还包括，在输出语音阅读的同时，从用户接收请求对应于特定标准的文档的语音命令。该方法还包括，响应于从用户接收到语音命令，识别所述多个文档中的对应于特定标准的一个或多个文档，并且输出对所述一个或多个识别的文档的相应文档的至少一部分的语音阅读。

[0010] 根据一些实施例，电子设备包括存储用于被所述一个或多个处理器执行的一个或多个程序的存储器和一个或多个处理器。所述一个或多个程序包括用于执行上述方法中的任一中方法的操作的指令。根据一些实施例，在具有显示器、存储器、和执行存储在存储器中的一个或多个程序的一个或多个处理器的电子设备上的图形用户界面包括在上述方法的任一种方法中显示的元件中的一个或多个元件，这些元件响应于输入而被更新，如上述方法中的任一种方法中所述。根据一些实施例，计算机可读存储介质已在其中存储了指令，当该指令由具有存储器和一个或多个处理器的电子设备执行时，使得该设备执行上述方法中的任一种方法的操作。根据一些实施例，电子设备包括用于执行上述方法中的任一种方法的操作的装置。根据一些实施例，用于电子设备中的信息处理装置包括用于执行上述方法中的任一种方法的操作的装置。根据一些实施例，电子设备包括被配置为执行上述方法中的任一种方法的操作的处理单元。

[0011] 根据一些实施例，电子设备包括被配置为接收音频输入的音频输入单元。电子设备还包括被配置为输出可听信息的音频输出单元。电子设备包括耦接至音频输入单元和音频输出单元的处理单元。处理单元被配置为接收包括多个链接的第一文档。处理单元被配置为输出对第一文档的至少一部分的语音阅读。处理单元被配置为输出识别所述多个链接中的一个链接的可听信息。处理单元被配置为，响应于输出识别该链接的可听信息，从用户接收关于该链接的语音命令。处理单元被配置为，响应于从用户接收到语音命令，输出对第二文档的与该链接相关联的至少一部分的语音阅读。

[0012] 根据一些实施例，电子设备包括被配置为接收音频输入的音频输入单元。电子设备还包括被配置为输出可听信息的音频输出单元。电子设备包括耦接至音频输入单元和音频输出单元的处理单元。处理单元被配置为接收具有多个部分的文档，其中所述部分中的至少一些部分与相应的元数据相关联。处理单元被配置为输出对文档的相应部分的语音阅读，包括基于相应的元数据可听地区分所述相应部分。处理单元被配置为从用户接收请求导航至与特定元数据相关联的特定部分的语音命令。处理单元被配置为，响应于接收到语音命令，输出对与特定元数据相关联的特定部分的语音阅读。

[0013] 根据一些实施例，电子设备包括被配置为接收音频输入的音频输入单元。电子设备还包括被配置为输出可听信息的音频输出单元。电子设备包括耦接至音频输入单元和音频输出单元的处理单元。处理单元被配置为输出对多个文档中的一个文档的至少一部分的语音阅读。处理单元被配置为，在输出语音阅读的同时，从用户接收请求对应于特定标准的文档的语音命令。处理单元被配置为，响应于从用户接收到语音命令，识别所述多个文档中的对应于特定标准的一个或多个文档；并且输出对所述一个或多个识别的文档的相应文档

的至少一部分的语音阅读。

[0014] 因此,用启用文档的导航和浏览的新型且改进的方法来提供数字助理系统,从而改进针对具有有限访问性的用户的用户界面。此类方法和系统可补充或替换现有的方法和系统。

附图说明

- [0015] 图 1 为示出根据一些实施例的数字助理在其中操作的环境的框图。
- [0016] 图 2 为示出根据一些实施例的数字助理客户端系统的框图。
- [0017] 图 3 为示出根据一些实施例的独立式数字助理系统或数字助理服务器系统的框图。
- [0018] 图 4A-4N 示出根据一些实施例的在电子设备上显示的示例性用户界面。
- [0019] 图 5 为示出根据一些实施例的用于导航通过由电子设备执行的文档的操作的流程图。
- [0020] 图 6 为示出根据一些实施例的用于浏览由电子设备执行的文档的操作的流程图。
- [0021] 图 7 为示出根据一些实施例的用于识别由电子设备执行的一组文档的操作的流程图。
- [0022] 图 8 为根据一些实施例的电子设备的功能框图。
- [0023] 图 9 为根据一些实施例的电子设备的功能框图。
- [0024] 图 10 为根据一些实施例的电子设备的功能框图。
- [0025] 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

[0026] 图 1 为根据一些实施例的数字助理的操作环境 100 的框图。术语“数字助理”、“虚拟助理”、“智能自动化助理”或“自动数字助理”是指解释口头和 / 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者或者者：识别具有被设计来实现推断出的用户意图的步骤和参数的任务流（例如，识别任务类型），将来自推断出的用户意图的特定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，向服务提供方发送请求）；以及以可听（例如，语音）和 / 或可视形式来生成对用户的输出响应。

[0027] 具体地，数字助理系统能够接受至少部分地以自然语言命令、请求、声明、讲述和 / 或询问的形式的用户请求。通常，用户请求寻求数字助理系统的任务的信息性回答和执行。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一个问题，诸如“我现在在哪儿？”基于用户的当前位置，数字助理可回答：“您在西大门附近的中央公园”。用户还可请求执行一项任务，例如通过声明“请邀请我的好友们下周参加我女朋友的生日派对”。作为响应，数字助理可通过生成语音输出“好的，马上”来确认该请求，并随后将适当的日程邀请从用户的电子邮件地址发送至列于用户的电子通讯录中的用户好友中的每一位。存在与数字助理进行

交互以请求信息或执行各种任务的许多其他方法。除提供言语响应并采取经编程的动作之外,数字助理还可提供其他视频或音频形式(例如,作为文本、警报、音乐、视频、动画等)的响应。

[0028] 如图1中所示,在一些实施例中,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,并且与DA服务器106进行通信。DA服务器106为任意数量的DA客户端102提供服务器侧功能,所述任意数量的DA客户端各自驻留在相应的用户设备104(也称作客户端设备)上。

[0029] 在一些实施例中,DA服务器106包括面向客户端的I/O接口112、一个或多个处理模块114、数据与模型116,以及至外部服务的I/O接口118。面向客户端的I/O接口有助于数字助理服务器106的面向客户端的输入和输出处理。所述一个或多个处理模块114利用数据与模型116基于自然语言输入来确定用户的意图,并基于推断出的用户意图来执行任务执行。

[0030] 在一些实施例中,DA服务器106通过一个或多个网络110与外部服务(例如,一种或多种导航服务、一种或多种消息传送服务、一种或多种信息服务、日历服务、电话服务等)通信以用于完成任务或获取信息。至外部服务的I/O接口118有助于此类通信。

[0031] 用户设备104的实例包括但不限于手持式计算机、个人数字助理(PDA)、平板电脑、膝上型计算机、台式计算机、蜂窝电话、智能电话、增强型通用分组无线电服务(EGPRS)移动电话、媒体播放器、导航设备、游戏机、电视机、遥控器、或者这些数据处理设备中的任意两种或更多种的组合或任何其他合适的数据处理设备。有关用户设备104的更多细节参照图2中示出的示例性用户设备104提供。

[0032] 一个或多个通信网络110的实例包括局域网(“LAN”)和广域网(“WAN”),例如互联网。一个或多个通信网络110可使用任何已知的网络协议,包括各种有线或无线协议诸如以太网、通用串行总线(USB)、火线(FIREWIRE)、全球移动通信系统(GSM)、增强型数据GSM环境(EDGE)、码分多址(CDMA)、时分多址(TDMA)、蓝牙、Wi-Fi、互联网协议语音技术(VoIP)、Wi-MAX、或任何其他合适的通信协议来实施。

[0033] 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。

[0034] 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)和服务器侧部分(例如,DA服务器106)两者,但在一些实施例中,数字助理系统仅指服务器侧部分(例如,DA服务器106)。作为另外一种选择,在一些实施例中,可将数字助理的功能实施为安装在用户设备上的独立式应用程序。此外,数字助理的客户端部分与服务器部分之间的功能划分在不同的实施例中可以是变化的。例如,在一些实施例中,DA客户端102为仅提供面向用户的输入和输出处理功能且将数字助理的所有其他功能委托给DA服务器106的瘦客户端。在一些其他实施例中,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。

[0035] 图1还示出了web服务器120(例如,web服务器1(122-1)、web服务器2(122-2)、

web 服务器 3(122-3) 等)。尽管所述一个或多个 web 服务器 120 不是数字助理系统的一部分,但服务器系统 108 和 / 或一个或多个用户设备 104 可与 web 服务器 120 中的一者或多者进行通信以检索一个或多个文档和 / 或与所述一个或多个文档相关联的信息。

[0036] 图 2 为根据一些实施例的用户设备 104 的框图。用户设备 104 包括存储器接口 202、一个或多个处理器 204 以及外围设备接口 206。用户设备 104 中的各种部件通过一条或多条通信总线或信号线来耦接。用户设备 104 包括各种传感器、子系统、以及耦接至外围设备接口 206 的外围设备。传感器、子系统、以及外围设备采集信息和 / 或有助于用户设备 104 的各种功能。

[0037] 例如,在一些实施例中,运动传感器 210、光传感器 212、以及接近传感器 214 耦接至外围设备接口 206 以有助于取向、照明和接近感测功能。在一些实施例中,其他传感器 216 诸如定位系统 (例如, GPS 接收器)、温度传感器、生物计量传感器等,连接至外围设备接口 206,以有助于相关功能。

[0038] 在一些实施例中,用户设备 104 包括耦接至外围设备接口 206 的照相机子系统 220。在一些实施例中,照相机子系统 220 的光学传感器 222 有助于照相机功能,诸如拍摄照片和记录视频剪辑。在一些实施例中,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。在一些实施例中,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 以有助于支持语音的功能,诸如语音识别、语音复制、数字记录和电话功能。

[0039] 在一些实施例中,I/O 子系统 240 还耦接至外围设备接口 206。在一些实施例中,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技术来检测接触和移动或它们的间断,所述多种触摸灵敏度技术诸如电容性技术、电阻性技术、红外技术、表面声波技术、接近传感器阵列等。在一些实施例中,用户设备 104 包括不具有触敏表面的显示器。在一些实施例中,用户设备 104 包括独立的触敏表面。在一些实施例中,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,所述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、拇指滚轮、红外线端口、USB 端口、和 / 或指针设备诸如触笔。

[0040] 存储器接口 202 耦接至存储器 250。在一些实施例中,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)。

[0041] 在一些实施例中,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并用于执行硬件相关任务的指令。通信模块 254 有助于与一个或多个附加设备、一个或多个计算机和 / 或一个或多个服务器进行通信。图形用户界面模块 256 有助于图形用户界面处理。传感器处理模块 258 有助于与传感器相关的处理和功能 (例如,处理用一个或多个麦克风 228 所接收的语音输入)。电话模块 260 有助于与

电话相关的过程和功能。应用程序模块 262 有助于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。在一些实施例中,应用程序模块 262 包括 web 浏览器应用程序 270 或与 web 浏览器应用程序 270 交互。在一些实施例中,应用程序模块 262 包括电子消息传送应用程序或与电子消息传送应用程序交互。

[0042] 如上所述,在一些实施例中,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,用户专用的词汇数据、偏好数据,和 / 或其他数据诸如用户的电子通讯录、待办事项列表、购物清单等)以提供数字助理的客户端侧功能。

[0043] 在各种实施例中,数字助理客户端模块 264 能够通过用户设备 104 的各种用户界面(例如,I/O 子系统 244)接受语音输入、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。

[0044] 在一些实施例中,数字助理客户端模块 264 利用各种传感器、子系统和外围设备来从用户设备 104 的周围环境采集附加信息以建立与用户输入相关联的上下文。在一些实施例中,数字助理客户端模块 264 将上下文信息或其子集与用户输入一起提供至数字助理服务器(例如,数字助理服务器 106,图 1)以帮助推断用户的意图。

[0045] 在一些实施例中,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。在一些实施例中,上下文信息还包括设备的物理状态,例如,设备取向、设备位置、设备温度、功率电平、速度、加速度、运动模式、蜂窝信号强度等。在一些实施例中,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而提供至数字助理服务器(例如,数字助理服务器 106,图 1)。

[0046] 在一些实施例中,DA 客户端模块 264 响应于来自数字助理服务器的请求而选择性地提供存储在用户设备 104 上的信息(例如,用户数据 266 的至少一部分)。在一些实施例中,数字助理客户端模块 264 还当数字助理服务器 106(图 1)进行请求时经由自然语言对话或其他用户界面来引出来自用户的附加输入。数字助理客户端模块 264 将附加输入传递至数字助理服务器 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所表达的用户的意图。

[0047] 在一些实施例中,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 / 或在软件中实施,该硬件和 / 或软件包括在一个或多个信号处理集成电路和 / 或专用集成电路中,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。例如,在一些实施例中,用户设备 104 不包括触摸屏 246。

[0048] 图 3 为根据一些实施例的示例性数字助理系统 300 的框图。在一些实施例中,数字助理系统 300 在独立式计算机系统上实施。在一些实施例中,数字助理系统 300 跨多个计算机而分布。在一些实施例中,数字助理的模块和功能中的一些被划分成服务器部分和客户端部分,其中客户端部分驻留在用户设备(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）进行通信，例如如图 1 中所示。在一些实施例中，数字助理系统 300 为图 1 中所示的服务器系统 108（和 / 或数字助理服务器 106）的实施例。在一些实施例中，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。应当指出的是，数字助理系统 300 仅为数字助理系统的一个实例，且该数字助理系统 300 可具有比示出更多或更少的部件、可组合两个或更多个部件、或可具有部件的不同配置或布置。图 3 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 / 或专用集成电路），或它们的组合中实施。

[0049] 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。

[0050] 在一些实施例中，存储器 302 包括非暂态计算机可读介质，诸如高速随机存取存储器和 / 或非易失性计算机可读存储介质（例如，一个或多个磁盘存储设备、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）。

[0051] I/O 接口 306 将数字助理系统 300 的输入 / 输出设备 316 诸如显示器、键盘、触摸屏和麦克风耦接至用户界面模块 322。I/O 接口 306 与用户界面模块 322 结合，接收用户输入（例如，语音输入、键盘输入、触摸输入等）并相应地对这些输入进行处理。在一些实施例中，当数字助理在独立式用户设备上实施时，数字助理系统 300 包括相对于图 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 228）。在一些实施例中，数字助理系统 300 代表数字助理具体实施的服务器部分，并且通过驻留在用户设备（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。

[0052] 在一些实施例中，网络通信接口 308 包括一个或多个有线通信端口 312 和 / 或无线传输和接收电路 314。一个或多个有线通信端口经由一个或多个有线接口，例如以太网、通用串行总线 (USB)、火线等来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，这些通信标准、协议和技术诸如 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 实现数字助理系统 300 与网络以及其他设备之间的通信，该网络诸如互联网、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)。

[0053] 在一些实施例中，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。

[0054] 操作系统 318（例如，Darwin、RTXC、LINUX、UNIX、OS X、WINDOWS、或嵌入式操作系统诸如 VxWorks）包括用于控制和管理一般系统任务（例如，存储器管理、存储设备控制、电力管理等）的各种软件部件和 / 或驱动器，并有助于各种硬件、固件与软件部件之间的通信。

[0055] 通信模块 320 有助于通过通信网络接口 308 来进行数字助理系统 300 与其他设备之间的通信。例如，通信模块 320 可与图 2 中所示的设备 104 的通信模块 254 进行通信。通信模块 320 还包括用于处理被无线电路 314 和 / 或有线通信端口 312 所接收的数据的各种软件部件。

[0056] 在一些实施例中,用户界面模块 322 经由 I/O 接口 306 来从用户(例如,来自键盘、触摸屏、和 / 或麦克风)接收命令和 / 或输入,并将用户界面对象提供在显示器上。

[0057] 应用程序 324 包括被配置为由所述一个或多个处理器 304 执行的程序和 / 或模块。例如,如果数字助理系统在独立式用户设备上实施,则应用程序 324 可包括用户应用程序,诸如游戏、日历应用程序、导航应用程序、web 浏览器应用程序、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。

[0058] 存储器 302 还存储数字助理模块(或数字助理的服务器部分)326。在一些实施例中,数字助理模块 326 包括以下子模块、或者它们的子集或超集:输入 / 输出处理模块 328、语音文本转换(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、以及服务处理模块 338。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、任务流模型 354、以及服务模型 356。

[0059] 在一些实施例中,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:识别在从用户接收的自然语言输入中表达的用户意图;主动引出并获得推断用户意图所需的信息(例如,通过消除词、姓名、意图的歧义等);确定用于满足推断出的意图的任务流;以及执行该任务流以满足推断出的意图。在一些实施例中,当出于各种原因而未向或不能向用户提供令人满意的响应时,数字助理还采取适当的行动。

[0060] 在一些实施例中,I/O 处理模块 328 通过 I/O 设备 316 与用户进行交互,或通过网络通信接口 308 与用户设备(例如,图 1 中的用户设备 104)进行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。上下文信息包括用户专用的数据、词汇、和 / 或与用户输入有关的偏好。在一些实施例中,上下文信息还包括当接收到用户请求时所述设备(例如,图 1 中的用户设备 104)的软件和硬件状态,和 / 或与当接收到用户请求时用户的周围环境相关的信息。在一些实施例中,I/O 处理模块 328 还向用户发送有关用户请求的跟进问题,并从用户接收回答。在一些实施例中,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文本转换(STT)处理模块 330 以用于语音文本转换。

[0061] 在一些实施例中,语音文本转换处理模块 330 通过 I/O 处理模块 328 来接收语音输入(例如,在语音记录中捕捉的用户话语)。在一些实施例中,语音文本转换处理模块 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词或符号的序列。语音文本转换处理模块 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(Hidden Markov)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。在一些实施例中,语音文本转换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。一旦语音文本转换处理模块 330 获得语音文本转换处理的结果(例如,词或符号的序列),其便将结果传送至自然语言处理模块 332 以用于意图推断。

[0062] 数字助理 326 的自然语言处理模块 332 (“自然语言处理器”) 取得由语音文本转换处理模块 330 生成的词或符号的序列 (“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 执行并且具有在任务流模型 354 中实施的相关联的任务流的任务。相关联的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“一种或多种可执行意图”的能力。

[0063] 在一些实施例中, 除从语音文本转换处理模块 330 获得的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 / 或进一步定义包含在从语音文本转换处理模块 330 接收的符号序列中的信息。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。

[0064] 在一些实施例中, 自然语言处理基于知识本体 360。知识本体 360 是一种包含多个节点的层级结构, 每个节点要么代表“可执行意图”、要么代表与“可执行意图”中的一者或多者有关的一种“属性”或其他“属性”。如上所指出的, “可执行意图”代表数字助理系统 300 能够执行的任务 (例如, “可执行”或者可对其采取行动的任务)。“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。知识本体 360 中可执行意图节点与属性节点之间的连接定义由属性节点所代表的参数如何从属于由可执行意图节点所代表的任务。

[0065] 在一些实施例中, 知识本体 360 由可执行意图节点和属性节点组成。在知识本体 360 内, 每个可执行意图节点直接连接至或通过一个或多个中间属性节点连接至一个或多个属性节点。类似地, 每个属性节点直接连接至或通过一个或多个中间属性节点连接至一个或多个可执行意图节点。

[0066] 可执行的意图节点连同其所链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的意图相关联, 并涉及与特定可执行的意图相关联的一组节点 (以及其间的关系)。在一些实施例中, 知识本体 360 由多个域组成。每个域可与一个或多个其他域共享一个或多个属性节点。

[0067] 在一些实施例中, 知识本体 360 包括数字助理能够理解并对其起作用的所有域 (因此可执行的意图)。在一些实施例中, 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进行修改。

[0068] 在一些实施例中, 可将与多个相关的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相关的属性节点和可执行的意图节点的群集。与旅行相关的可执行的意图节点可包括“机票预订”、“酒店预订”、“汽车租赁”、“获取路线”、“寻找兴趣点”, 等等。同一超级域 (例如, “旅行”超级域) 下的可执行的意图节点可具有多个共用的属性节点。例如, 针对“机票预订”、“酒店预订”、“汽车租赁”、“获取路线”、“寻找兴趣点”的可执行的意图节点可共享属性节点“起始位置”、“目的地”、“出发日期 / 时间”、“到达日期 / 时间”及“同行人数”中的一个或多个。

[0069] 在一些实施例中，知识本体 360 中的每个节点和与由节点所代表的属性或可执行意图有关的一组词和 / 或短语相关联。与每个节点相关联的相应组的词和 / 或短语是与节点相关联的所谓的“词汇”。可将与每个节点相关联的相应组的词和 / 或短语存储在与由节点所代表的属性或可执行意图相关联的词汇索引 344 中。例如，与“餐厅”的属性的节点相关联的词汇可包括词诸如“食物”、“饮品”、“菜系”、“饥饿”、“吃”、“比萨”、“快餐”、“一餐”等。又如，与“发起电话呼叫”的可执行意图的节点相关联的词汇可包括词和短语诸如“呼叫”、“打电话”、“拨打”、“与 通电话”、“呼叫该号码”、“打电话给”等。词汇索引 344 任选地包括不同语言的词和短语。

[0070] 在一些实施例中，自然语言处理器 332 从语音文本转换处理模块 330 接收符号序列（例如，文本串），并确定符号序列中的词牵涉哪些节点。在一些实施例中，如果发现符号序列中的词或短语与知识本体 360 中的一个或多个节点相关联（经由词汇索引 344），则所述词或短语将“触发”或“激活”这些节点。当多个节点被“触发”时，基于已激活节点的数量和 / 或相对重要性，自然语言处理器 332 将选择可执行意图中的一个作为用户意图让数字助理执行的任务（或任务类型）。在一些实施例中，选择具有最多“已触发”节点的域。在一些实施例中，选择具有最高置信度（例如，基于其各个已触发节点的相对重要性）的域。在一些实施例中，基于已触发节点的数量和重要性的组合来选择域。在一些实施例中，在选择节点的过程中还考虑附加因素，诸如数字助理系统 300 先前是否已正确解释来自用户的类似请求。

[0071] 在一些实施例中，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名称时，自然语言处理器 332 将能够识别该名称涉及知识主体中的属性或子属性的特定实例。在一些实施例中，特定实体的名称是企业、餐厅、人、电影等的名称。在一些实施例中，数字助理系统 300 可从其他数据源中搜索并识别特定实体名称，所述其他数据源诸如用户的通讯录、电影数据库、音乐家数据库和 / 或餐厅数据库。在一些实施例中，当自然语言处理器 332 识别出符号序列中的词是特定实体的名称（诸如用户通讯录中的名称）时，在于用户请求的知识主体内选择可执行意图的过程中，为该词赋予附加的重要性。

[0072] 例如，当从用户请求中识别出词“Santo 先生”且当词汇索引 344 中发现姓“Santo”是用户联系人列表中的联系人之一时，则用户请求可能对应于“发送消息”或“发起电话呼叫”域。又如，当在用户请求中发现的词“ABC 咖啡馆”且当在词汇索引 344 中发现的词语“ABC 咖啡馆”是用户所在城市中的特定餐厅的名称时，则用户请求可能对应于“餐厅预订”域。

[0073] 用户数据 348 包括用户专用的信息，诸如用户专用的词汇、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 可使用用户专用的信息来补充包含在用户输入中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。

[0074] 一旦自然语言处理器 332 基于用户请求识别出可执行意图（或域），自然语言处理器 332 便生成结构化查询以表示所识别的可执行意图。在一些实施例中，结构化查询包括

针对可执行意图的域内的一个或多个节点的参数，并且所述参数中的至少一些参数填充有在用户请求中指定的特定信息和要求。例如，用户可以说：“通知寿司店预定晚上 7 点的座位。”在该情况下，自然语言处理器 332 能够基于用户输入将可执行意图正确地识别为“餐厅预订”。根据知识主体，“餐厅预订”域的结构化查询可包括参数诸如 {菜系}、{时间}、{日期}、{同行人数} 等。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=“寿司类”} 和 {时间=“下午 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，其他必要参数诸如 {同行人数} 和 {日期} 未基于当前可用的信息在结构化查询中指定。在一些实施例中，自然语言处理器 332 利用所接收的上下文信息来填充结构化查询中的一些参数。例如，如果用户请求“我附近”的寿司餐厅，则自然语言处理器 332 可利用来自用户设备 104 的 GPS 坐标来填充结构化查询中的 {位置参数}。

[0075] 在一些实施例中，自然语言处理器 332 将结构化查询（包括任何已完成的参数）传送至任务流处理模块 336（“任务流处理器”）。任务流处理器 336 被配置为执行以下中的一者或者者：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。在一些实施例中，完成这些任务所必需的各种过程在任务流模型 354 中提供。在一些实施例中，任务流模型 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。

[0076] 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 / 或弄清可能有模糊的话语。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 以进行与用户的对话。在一些实施例中，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。在一些实施例中，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，并接收经由口头或物理（例如，触摸手势）响应的来自用户的输入。继续上文的实例，当任务流处理器 336 调用对话处理模块 334 以针对与域“餐厅预订”相关联的结构化查询来确定“同行人数”和“日期”信息时，对话流处理器 334 生成问题诸如“共有多少人用餐？”和“具体哪天用餐？”以传送至用户。一旦从用户接收到回答，对话处理模块 334 便用缺失是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。

[0077] 在一些情况下，任务流处理器 336 可能接收到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”域的结构化查询可能指示预期接收人为“Bob”，并且用户可具有多个名为“Bob”的联系人。任务流处理器 336 将请求对话处理器 334 弄清结构化查询的这个属性。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。

[0078] 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 根据包含在结构化查询中的特定参数来执行任务流模型中的步骤和指令。例如，针对可执行意图“餐厅预订”的任务流模型可包括用于联系餐厅并实际上请求在特定时间针对特定同行人数的预订的步骤和指令。例如，通过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 = 下午 7 点，同行人数 = 5 人}，任务流处理器 336 可执行以下步骤：(1) 登

录 ABC 咖啡馆的服务器,或者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、时间和同行人数信息,(3) 提交表格,以及(4) 在用户日历中针对该预订制作日历条目。

[0079] 在一些实施例中,任务流处理器 336 在服务处理模块 338 (“服务处理器”) 的辅助下完成用户输入中所请求的任务或者提供用户输入中所请求的信息性回答。例如,服务处理器 338 可代替任务流处理器 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所述其他应用程序进行交互,以及调用第三方服务(例如,餐厅预订门户网站、社交网站、银行门户网站等)或与第三方服务进行交互。在一些实施例中,每项服务所需的协议和应用编程接口 (API) 可通过服务模型 356 间的相应服务模型来指定。服务处理器 338 针对服务访问适当的服务模型并依据该服务模型根据该服务所需的协议和 API 来生成针对该服务的请求。

[0080] 例如,如果餐厅已启用在线预订服务,则餐厅可提交一服务模型,该服务模型指定进行预订的必要参数以及用于将必要参数的值传送至在线预订服务的 API。当被任务流处理器 336 请求时,服务处理器 338 可使用存储在服务模型 356 中的 web 地址来建立与在线预订服务的网络连接,并将预订的必要参数(例如,时间、日期、同行人数)以根据在线预订服务的 API 的格式发送至在线预订接口。

[0081] 在一些实施例中,自然语言处理器 332、对话处理器 334 以及任务流处理器 336 共同且反复地使用以推断并限定用户的意图、获得信息以进一步明确并提炼用户意图、并最终生成响应(例如,将输出提供至用户,或完成任务)以满足用户的意图。

[0082] 在一些实施例中,在已执行满足用户请求所需的所有任务之后,数字助理 326 制定确认响应,并通过 I/O 处理模块 328 将该响应发送回用户。如果用户请求寻求信息性回答,则确认响应向用户呈现所请求的信息。在一些实施例中,数字助理还请求用户来指示用户是否对由数字助理 326 所产生的响应满意。

[0083] 图 4A-4N 示出根据一些实施例的在电子设备(例如,相对于图 2 所描述的用户设备 104,或相对于图 3 所描述的独立式设备 300) 上所显示的示例性用户界面。

[0084] 图 4A-4J 示出根据一些实施例的与导航通过和浏览一个或多个文档相关联的示例性用户界面。

[0085] 图 4A 示出包含第一文档的一部分的示例性用户界面。在一些实施例中,第一文档包括例如用下划线指示的多个链接(例如,“新国家”、“内战”、“一个大战场”和“战场的一角”)。

[0086] 图 4A 还示出,在一些实施例中,在输出对第一文档的语音阅读之前,电子设备输出指示第一文档上的当前位置的可听信息和 / 或包括有关第一文档的信息的可听信息。例如,电子设备可输出音频信号(例如,“哔哔声”)或语音指示,该语音指示指示第一文档的将要阅读的部分为第一文档的开头或者电子设备即将输出语音阅读。此外,或代替之,电子设备可输出有关第一文档的信息,诸如第一文档的标题、作者、日期和来源。

[0087] 图 4B 示出了电子设备输出对第一文档的一部分的语音阅读,该语音阅读对应于文本“八十七年前,我们的先辈在这个大陆上建立了”。

[0088] 图 4C 示出了电子设备输出识别与文本“新国家”相关联的链接的可听信息。例如,电子设备输出词“链接”的语音信号和 / 或类似于“哔哔声”的音频信号。在一些实施例中,

术语“链接”是指包含所链接的文档的位置的元数据。在一些实施例中，术语“链接”包括超文本锚点。语音信号和音频信号可在输出对与链接相关联的文本的语音阅读之前或之后输出。在一些实施例中，电子设备还输出有关链接的信息，诸如所链接的文档的标题、作者、日期和来源。

[0089] 在一些实施例中，电子设备在输出语音信号、音频信号、有关链接的信息和 / 或对与链接相关联的文本的语音阅读之后暂停预定义的时段（例如，一秒、两秒、三秒、四秒或五秒等）。

[0090] 图 4D 示出了电子设备输出对第一文档的继与所述链接相关联的文本之后的一部分的语音阅读。

[0091] 在一些实施例中，电子设备从用户接收语音命令。在一些实施例中，在电子设备输出语音信号、音频信号、有关链接的信息和 / 或对与链接相关联的文本的语音阅读的同时；在暂停期间；或者在电子设备输出对所述文档的继与所述链接相关联的文本之后的一部分的语音阅读的同时，该电子设备从用户接收语音命令。在一些实施例中，电子设备在输出有关不同于所述链接的第二链接的语音信号和 / 或音频信号之前从用户接收语音命令。

[0092] 在一些实施例中，语音命令是对导航至所链接的文档的请求（例如，“跟随链接”）。在一些实施例中，电子设备存储文档的最后一个语音输出部分。在一些实施例中，电子设备存储有关链接的信息（例如，链接的位置和 / 或标识）作为文档的最后一个语音输出部分的替代。

[0093] 图 4E 示出了电子设备显示所链接的文档的一部分并输出对所链接的文档的一部分的语音阅读。在一些实施例中，在输出对链接的文档的语音阅读之前，电子设备输出指示文档上的当前位置的可听信息和 / 或包括有关文档的信息的可听信息。例如，电子设备可输出音频信号（例如，“哔哔声”）或语音指示，该音频信号或语音指示指示文档的将要阅读的部分为文档的开头或者电子设备即将输出语音阅读。此外，或代替之，电子设备可输出有关文档的信息，诸如文档的标题、作者、日期和来源。此外，电子设备可输出音频信号或语音指示，该音频信号或语音指示指示文档的将要阅读的部分为标题、新行、新句子和 / 或新段落。

[0094] 图 4F 示出电子设备进一步输出对所链接的文档的后续部分的语音阅读。类似地，电子设备可继续输出对所链接的文档的剩余部分的语音阅读。

[0095] 图 4G 示出电子设备输出对所链接的文档的最后一个句子的语音阅读。

[0096] 在一些实施例中，在电子设备输出对所链接的文档的一个或多个部分的语音阅读的同时，电子设备从用户接收请求导航回到第一文档的语音命令。图 4H 示出了，作为响应，电子设备显示第一文档的一个或多个部分并输出对第一文档的一部分的语音阅读。

[0097] 在一些实施例中，电子设备输出有关导航回到第一文档的语音询问。在一些实施例中，在输出对所链接的文档的最后一个句子的语音阅读之后，电子设备输出有关导航回到第一文档的语音询问。当用户提供请求导航回到第一文档的语音命令时，电子设备输出对第一文档的一部分的语音阅读，如上所述（即，导航回到第一文档）。作为另外一种选择，在一些实施例中，例如当电子设备完成输出对所链接的文档的语音阅读时，电子设备在不从用户接收语音命令的情况下自动导航回到第一文档。

[0098] 在一些实施例中，当电子设备恢复输出对第一文档的语音阅读时，电子设备开始

输出对第一文档的对应于该第一文档的最后一个语音输出部分的一部分的语音阅读。在一些实施例中，电子设备通过输出对与链接相关联的文本的语音阅读来恢复。在一些实施例中，电子设备通过输出对第一文档的继与所述链接相关联的文本之后的一部分的语音阅读来恢复。在一些实施例中，电子设备通过输出对包括与所述链接相关联的文本的句子的语音阅读来恢复。在一些实施例中，电子设备通过输出对包括与所述链接相关联的文本的段落的语音阅读来恢复。

[0099] 图 4I 示出电子设备输出对第一文档中的下一个段落的一部分的语音阅读。在一些实施例中，在输出对下一个段落的一部分的语音阅读之前，电子设备输出指示文档上的当前位置的可听信息。例如，电子设备可输出音频信号（例如，“哔哔声”）或语音指示，该音频信号或语音指示指示文档的将要阅读的部分为新行、新句子和 / 或新段落。

[0100] 图 4J 示出了，在一些实施例中，文档包括具有相应样式的多个部分。换句话讲，文档的第一部分可包括第一样式并且文档的第二部分包括第二样式。例如，图 4J 中所示的文档的第一段落具有第一样式（例如，该段落为非斜体字体）并且文档的后续段落具有不同于第一样式的第二样式（例如，该段落为斜体字体）。电子设备使用第一组语音特征（例如，成年女性嗓音）来输出对文档的第一部分的语音阅读并且使用第二组语音特征（例如，成年男性嗓音）来输出对文档的第二部分的语音阅读。语音特征可包括音高、语速和音量中的一者或者，和 / 或对基于例如性别（例如，男性或女性）和年龄（例如，成年人或儿童）分类的特定组的说话人而言为典型的特征。

[0101] 图 4K-4N 示出了根据一些实施例的与识别一个或多个文档相关联的示例性用户界面。尽管图 4K-4N 中所示的示例性用户界面包括电子邮件消息，但类似方法可与其他类型的文档诸如书刊章节、百科全书条目等一起使用。

[0102] 图 4K 示出了电子消息的示例性用户界面。电子设备输出对消息的至少一部分的语音阅读（例如，“有一封从 David 发送给您、John、Karen 和 Paul 的电子邮件。主题为新功能。该消息说到‘这看起来真好。大伙儿觉得呢？’”）。

[0103] 在一些实施例中，电子设备从用户接收请求对应于特定标准的一个或多个文档的语音命令。例如，在一些实施例中，特定标准是一个或多个文档由一位或多作者创作（例如，“找到来自 David 的电子邮件消息”、“找到来自 David、John、Karen 和 Paul 的电子邮件消息”、“找到来自该电子邮件的地址的电子邮件消息”，等等）。在其他实施例中，特定标准是一个或多个文档与特定文档相关联（例如，“在该话题中找到电子邮件消息”、“找到对该电子邮件的回复”，等等）。在另外的实施例中，特定标准是一个或多个文档是对特定文档的回复。在另外的实施例中，特定标准是一个或多个文档包括来自相应作者的最新消息（例如，“找到来自 David 的最新电子邮件”）。在一些实施例中，特定标准是一个或多个文档包括来自相应作者的第一条消息。

[0104] 在一些实施例中，特定标准是一个或多个文档对应于特定日期范围（例如，“找到上周接所收的电子邮件”、“找到在 1 月 1 日至 3 月 31 日期间所接收的电子邮件消息”，等等）。在一些实施例中，特定日期范围对应于单个日期。

[0105] 在一些实施例中，特定标准包括上述提到的标准中的一个或多个（例如，“找到上周从 David 和 Karen 接收的电子邮件”）。

[0106] 图 4L 示出了响应于用户请求由特定作者例如 David 创作的电子邮件消息而显示

的示例性用户界面。电子设备输出对电子邮件消息列表的语音阅读。在一些实施例中，电子设备输出指示列表中电子邮件消息的数量的可听信息。

[0107] 图 4M 示出了响应于用户请求对先前所显示的电子邮件消息的回复而显示的另选的用户界面。电子设备输出对回复的列表的语音阅读。在一些实施例中，电子设备输出指示列表中的电子邮件消息的数量的可听信息。

[0108] 在一些实施例中，电子设备输出对列表（例如，图 4L 或图 4M 中所示的列表）中的电子邮件消息的语音阅读。在一些实施例中，当电子设备输出对列表中的最后一封电子邮件的语音阅读时，电子设备输出指示被读出的电子邮件是列表中的最后一封电子邮件的可听信息。

[0109] 尽管图 4A-4N 示出了在电子设备上显示的示例性用户界面，但在一些实施例中，电子设备可在不于电子设备上显示示例性用户界面的情况下输出可听信息（例如，语音阅读和 / 或可听信号）。在一些实施例中，电子设备完全不包括显示器。

[0110] 图 5 为示出了根据一些实施例的用于导航通过由电子设备（例如，相对于图 2 所述的用户设备 104，或相对于图 3 所述的独立式设备 300）所执行的文档的操作的流程图。在一些实施例中，电子设备包括便携式电子设备。在一些实施例中，电子设备包括计算机系统。

[0111] 这些操作仅仅是示例性的并且在各种实施例中可由电子设备执行较少或更少的操作。

[0112] 在一些实施例中，电子设备接收 (502) 第一文档。

[0113] 电子设备输出 (504) 对文档的一部分的语音阅读（例如，图 4B）。

[0114] 在一些实施例中，电子设备确定 (506) 文档的该部分是否包括链接。如果文档的该部分包括链接，则电子设备输出 (508) 识别所述多个链接中的一个链接的可听信息（例如，词“链接”的语音输出或音频信号“哔哔声”，和 / 或有关链接的信息诸如所链接的文档的标题、作者、日期和来源，如相对于图 4C 所述的）。

[0115] 在一些实施例中，电子设备从用户接收语音命令。在一些实施例中，电子设备确定 (510) 电子设备是否已接收到语音命令。

[0116] 在一些实施例中，当电子设备接收到语音命令时，电子设备确定所接收的语音命令是否为导航命令 (512)（例如，所接收的语音命令是否包括对导航至与链接相关联的第二文档（即，所链接的文档）的请求）。如果所接收的语音命令包括对导航至所链接的文档的请求，则电子设备检索 (514) 所链接的文档。电子设备然后输出 (504) 对所链接的文档的一个或多个部分的语音阅读（例如，图 4E）。

[0117] 在一些实施例中，当电子设备接收到语音命令时，电子设备确定所接收的语音命令是否为信息命令 (516)（例如，所接收的语音命令是否包括对有关所链接的文档的信息的请求）。如果所接收的语音命令包括对有关所链接的文档的信息（例如，标题、作者、日期、来源、摘要、概要、第一个句子、第一个段落，等等）诸如“作者是谁”的请求，则电子设备检索 (518) 所请求的信息。在一些实施例中，从所链接的文档获得所请求的信息的至少一部分（例如，标题、作者、日期和摘要）。在一些实施例中，通过处理所链接的文档（例如，对所链接的文档进行概述）来获得所请求的信息的至少一部分。在一些实施例中，由第三方服务器来提供所请求的信息的至少一部分（例如，查看所链接的文档）。电子设备输出

(520) 对所请求的信息的语音阅读。在其他实施例中，电子设备从所链接的文档的元数据接收所请求的信息的至少一部分，或者从与所链接的文档相关联的其他来源检索信息以获得细节。

[0118] 在一些实施例中，当用户在两个相邻链接之间或紧接在两个相邻链接之后提供语音命令时，电子设备将这两个链接识别为对应于语音命令的候选链接。由于用户的语音命令可能指向两个相邻链接的第一链接或第二链接，所以可能需要相对于用户的语音命令应当执行哪一个链接来进行阐明。在一些实施例中，电子设备输出有关候选链接的可听信息和 / 或有关用户想要针对哪个链接来执行语音命令的可听询问。例如，在电子设备输出对句子“学校 A 与学校 B 之间的下一场比赛将于 3 月 3 日举行”的语音阅读的同时，用户提供语音命令“跟随链接”。在一些实施例中，电子设备输出可听询问“哪一个链接？”并从用户接收后续语音命令。当后续语音命令为“学校 B”时，电子设备检索与文本“学校 B”链接的文档并输出对所链接的文档的一部分的语音阅读。

[0119] 在一些实施例中，在输出对文档的相应部分的语音阅读之后，电子设备确定 (522) 是否已到达文档的结尾（例如，电子设备是否已输出对整个文档的语音阅读）。如果尚未到达文档的结尾，则电子设备输出 (504) 对文档的后续部分的语音阅读。在一些实施例中，如果已到达文档的结尾，则电子设备等待来自用户的导航命令。当电子设备从用户接收到 (524) 请求导航至前一个文档或新文档的导航命令时，电子设备检索 (526) 该前一个文档或新文档。如先前所述，这些操作（例如，操作 522、524 和 526 等等）可由服务器或移动设备来执行。在一些实施例中，在从用户接收到导航命令之前，电子设备输出有关用户是否想要导航至另一个文档的语音询问（例如，如相对于图 4G 所述的）。在一些实施例中，当用户不想要导航至另一个文档时，电子设备停止输出语音阅读。

[0120] 图 6 为示出了根据一些实施例的用于浏览由电子设备（例如，相对于图 2 所述的用户设备 104 或相对于图 3 所述的独立式设备 300）所执行的文档的操作的流程图。在一些实施例中，电子设备包括便携式电子设备。在一些实施例中，电子设备包括计算机系统。

[0121] 这些操作仅仅是示例性的并且在各种实施例中可由电子设备执行较少或更少的操作。

[0122] 在一些实施例中，电子设备接收 (602) 具有多个部分的文档。所述部分中的至少一些部分与相应的元数据相关联。在一些实施例中，相应的元数据指示文档的结构（例如，段落、句子、标题、样式等等）。

[0123] 电子设备输出 (604) 对文档的相应部分的语音阅读。在输出对文档的相应部分的语音阅读的同时，电子设备基于该部分的相应元数据可听地区分每个相应部分。例如，电子设备确定 (606) 相应部分是否对应于位标。如本文所用，位标是指文档中的预定义的位置或预定义类型的位置。在一些实施例中，当相应部分对应于位标时，电子设备输出 (608) 语音信号或音频信号以指示该部分是否对应于位标。例如，在一些实施例中，在新段落的开始处，电子设备输出语音信号或音频信号以指示新段落的开始。类似地，在一些实施例中，使用语音信号或音频信号用于指示章节的开头或结尾和 / 或指示书签。

[0124] 继输出语音信号或音频信号以指示该部分是否对应于位标之后（例如，在于输出语音信号或音频信号以指示该部分是否对应于位标之后输出对数个段落的语音阅读之后），电子设备从用户接收语音命令并确定 (610) 语音命令是否为导航命令。当语音命令是

请求导航至位标的导航命令时,电子设备导航(614)至对应于位标的部分,并输出(604)对对应于位标的部分的语音阅读。

[0125] 在一些实施例中,在输出对文档的相应部分的语音阅读之后,电子设备确定(616)是否已到达文档的结尾(例如,电子设备是否已输出对整个文档的语音阅读)。如果尚未到达文档的结尾,则电子设备输出(604)对文档的后续部分的语音阅读。

[0126] 图7为示出了根据一些实施例的用于识别由电子设备(例如,相对于图2所述的用户设备104,或相对于图3所述的独立式设备300)所执行的一组文档的操作的流程图。在一些实施例中,电子设备包括便携式电子设备。在一些实施例中,电子设备包括计算机系统。

[0127] 这些操作仅仅是示例性的并且在各种实施例中可由电子设备执行较少或更少的操作。

[0128] 在一些实施例中,电子设备输出(702)对文档的至少一部分的语音阅读。

[0129] 电子设备从用户接收(704)请求对应于特定标准的文档的语音命令。在一些实施例中,电子设备在输出语音阅读的同时接收语音命令的至少一部分。在一些实施例中,特定标准要求所述一个或多个识别的文档由被用户所识别的一位或多作者创作(706)。在一些实施例中,特定标准要求所述一个或多个识别的文档与特定文档相关联(708)。在一些实施例中,特定标准要求所述一个或多个识别的文档是对特定文档的回复(710)。在一些实施例中,特定标准要求所述一个或多个识别的文档包括来自相应作者的最新消息(712)。在一些实施例中,特定标准要求所述一个或多个识别的文档对应于由用户所识别的特定日期范围(714)。

[0130] 在一些实施例中,在接收到请求对应于特定标准的文档的语音命令之前,电子设备从用户接收关于文档的信息(例如,作者、字数、最后更新的日期、文档中章节的数量等等)的请求。在一些实施例中,响应于对关于文档的信息的请求,电子设备输出包括关于文档的所请求的信息的可听信息。在一些实施例中,电子设备从一个或多个远程计算机系统(例如,一个或多个搜索引擎和/或数据库服务器)检索所请求的信息。在一些实施例中,电子设备从文档确定所请求的信息。

[0131] 在一些实施例中,电子设备从用户接收对存储所请求的信息的请求,并且作为响应,存储所请求的信息(例如,存储在电子设备或远程存储设备上)。在一些实施例中,电子设备在没有来自用户的对存储所请求信息的请求的情况下,存储所请求的信息。

[0132] 在一个实例中,该请求针对文档的长度(例如,“这篇文章有多长?”),并且响应于对文档长度的请求,电子设备确定文档的长度(例如,字数或页数)并输出包括文档的长度的可听信息。在另一个实例中,该请求针对文档是否更新过(例如,“这篇文章是否更新过?”),并且响应于对文档是否更新过的请求,电子设备确定文档是否已被更新并输出指示文档是否已被更新的可听信息(例如,“存在对该篇文章的更正”)。在又一个实例中,该请求针对文档的作者(例如,“谁创作了这篇文章?”)。在一些实施例中,响应于对文档作者的请求,电子设备输出包括文档的作者的可听信息。在一些情况下,作者信息从文档提取(例如,在如“由……创作”的文本部分中或在标记语言的作者字段中)。在一些情况下,作者信息从链接至该文档的网页提取。

[0133] 在一些实施例中,电子设备基于所请求的信息来接收对附加信息的请求,获得附

加信息，并输出附加信息。在上文相对于作者信息所描述的一些实例中，在输出包括文档作者的可听信息之后，电子设备接收对由同一位作者创作的其他文档的请求（例如，“这位作者还写过哪些别的文章？”），并且作为响应，获得包括由该文档的作者创作的其他文档的信息，并输出所述信息。在一些实施例中，电子设备通过向一个或多个远程计算机系统（例如，一个或多个搜索引擎和 / 或数据库服务器）发送请求并从所述一个或多个远程计算机系统的至少一个子组接收包括由该文档的作者创作的其他文档的信息的至少一部分来获得所述信息。

[0134] 在一些实施例中，在输出对文档的至少一部分的语音阅读的同时，电子设备从用户接收对存储关于文档的信息的请求。在一些实施例中，响应于接收到对存储关于文档的信息的请求，电子设备存储关于文档的信息。例如，在一些实施例中，该请求是为文档加书签（例如，“为这篇文章加书签”），并且该电子设备将文档的访问信息（例如，文档的统一资源定位符）存储为书签。在一些实施例中，电子设备将文档的访问信息存储在电子设备上。在一些实施例中，电子设备将访问信息存储在远程服务器处。

[0135] 对于本领域的普通技术人员而言将显而易见的是，在一些实施例中，电子设备在不执行后续操作（例如，下文所述的操作 716、718 和 720）的情况下提供所请求的信息。

[0136] 电子设备识别（716）对应于特定标准的一个或多个文档（例如，图 4L 和图 4M）。

[0137] 电子设备输出（718）对所述一个或多个识别的文档的相应文档的至少一部分的语音阅读。

[0138] 在一些实施例中，在输出对文档的相应部分的语音阅读之后，电子设备确定（720）是否已到达文档的结尾（例如，电子设备是否已输出对整个文档的语音阅读）。如果尚未到达文档的结尾，则电子设备输出（718）对文档的后续部分的语音阅读。

[0139] 在一些实施例中，如果已到达文档的结尾，则电子设备确定是否已阅读所述一个或多个识别的文档中的所有文档。如果尚未阅读所述一个或多个识别的文档中的所有文档，则电子设备输出对所述一个或多个识别的文档中的下一个文档的一部分的语音阅读。

[0140] 根据一些实施例，图 8 示出了根据如上所述的本发明的原理配置的电子设备 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实现以实行本发明的原理。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为或者被分离为子块以实现如上所述的本发明的原理。因此，本文的描述可支持本文所述的功能块的任何可能的组合或分离或进一步的定义。

[0141] 如图 8 中所示，电子设备 800 包括音频输入单元 804 和音频输出单元 806。在一些实施例中，电子设备 800 包括被配置为显示电子文档的一个或多个部分的显示单元 802。在一些实施例中，电子设备 800 包括触敏表面单元 808，该触敏表面单元被配置为检测触敏表面单元 808 上的一个或多个手势。电子设备还包括耦接至音频输入单元 804 和音频输出单元 806 的处理单元 810。在一些实施例中，处理单元 810 还耦接至显示单元 802 和触敏表面单元 808。在一些实施例中，处理单元 810 包括文档接收单元 812、输出单元 814、语音命令接收单元 816、存储单元 818、检索单元 820、识别单元 822 以及用户选择接收单元 824。

[0142] 处理单元 810 被配置为接收包括多个链接的第一文档（例如，利用文档接收单元 812）。处理单元 810 被配置为输出对第一文档的至少一部分的语音阅读（例如，利用输出单元 814 通过音频输出单元 806）。处理单元 810 被配置为输出识别所述多个链接中的一个

链接的可听信息（例如，利用输出单元 814 通过音频输出单元 806）。处理单元 810 被配置为，响应于输出识别该链接的可听信息，从用户接收关于该链接的语音命令（例如，利用语音命令接收单元 816 通过音频输入单元 804）。处理单元 810 被配置为，响应于从用户接收到语音命令，输出对第二文档的与该链接相关联的至少一部分的语音阅读（例如，利用输出单元 814 通过音频输出单元 806）。

[0143] 在一些实施例中，链接与第一文档的该部分中的文本相关联。

[0144] 在一些实施例中，处理单元 810 被配置为，响应于从用户接收到关于链接的语音命令，存储有关该链接的信息（例如，利用存储单元 818）并输出对第二文档的一个或多个部分的语音阅读（例如，利用输出单元 814 通过音频输出单元 806）。

[0145] 在一些实施例中，处理单元 810 被配置为，在输出对第二文档的至少一部分的语音阅读之后，从用户接收请求导航回到第一文档的语音命令（例如，利用语音命令接收单元 816 通过音频输入单元 804）；以及，响应于从用户接收到请求导航回到第一文档的语音命令，输出对第一文档的继第一文档中的与所述链接相关联的文本之后的一个或多个部分的语音阅读（例如，利用输出单元 814 通过音频输出单元 806）。

[0146] 在一些实施例中，处理单元 810 被配置为，在输出对第二文档的至少一部分的语音阅读之后，输出有关导航回到第一文档的语音询问（例如，利用输出单元 814 通过音频输出单元 806）；响应于输出有关导航回到第一文档的语音询问，从用户接收请求导航回到第一文档的语音命令（例如，利用语音命令接收单元 816 通过音频输入单元 804）；以及，响应于从用户接收到请求导航回到第一文档的语音命令，输出对第一文档的包括在第一文档中的与所述链接相关联的文本的一个或多个部分的语音阅读（例如，利用输出单元 814 通过音频输出单元 806）。

[0147] 在一些实施例中，处理单元 810 被配置为，在输出对第二文档的至少一部分的语音阅读之后，响应于从用户接收到请求导航回到第一文档的语音命令，自动输出对第一文档的继第一文档中的与所述链接相关联的文本之后的一个或多个部分的语音阅读（例如，利用输出单元 814 通过音频输出单元 806）。

[0148] 在一些实施例中，语音命令为第一类型的语音命令。在一些实施例中，处理单元 810 被配置为从用户接收关于链接的不同于第一类型的第二类型的语音命令（例如，利用语音命令接收单元 816 通过音频输入单元 804）；以及，响应于从用户接收到关于所述链接的第二类型的语音命令，从第二文档中检索信息（例如，利用检索单元 820）；以及基于来自第二文档的信息，输出对应于第二类型的语音命令的语音信息（例如，利用输出单元 814 通过音频输出单元 806）。

[0149] 在一些实施例中，第二类型的语音命令包括对有关第二文档的作者的信息的请求。

[0150] 在一些实施例中，第二类型的语音命令包括对第二文档的概要的请求。

[0151] 在一些实施例中，处理单元 810 被配置为将两个或更多个链接识别为对应于来自用户的语音命令的候选链接（例如，利用识别单元 822）；输出有关候选链接的可听信息（例如，利用输出单元 814 通过音频输出单元 806）；以及从用户接收对候选链接中的单个链接的选择（例如，利用用户选择接收单元 824 通过音频输入单元 804 或触敏表面单元 808）。

[0152] 根据一些实施例，图 9 示出了根据如上所述的本发明的原理配置的电子设备 900

的功能框图。设备的功能块可由硬件、软件、或者硬件和软件的组合实现以实行本发明的原理。本领域的技术人员能够理解，图 9 中所述的功能块可被组合为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的定义。

[0153] 如图 9 中所示，电子设备 900 包括音频输入单元 904 和音频输出单元 906。在一些实施例中，电子设备 900 包括被配置为显示电子文档的一个或多个部分的显示单元 902。在一些实施例中，电子设备 900 包括触敏表面单元 908，该触敏表面单元被配置为检测触敏表面单元 908 上的一个或多个手势。电子设备还包括耦接至音频输入单元 904 和音频输出单元 906 的处理单元 910。在一些实施例中，处理单元 910 还耦接至显示单元 902 和触敏表面单元 908。在一些实施例中，处理单元 910 包括文档接收单元 912、输出单元 914 以及语音命令接收单元 916。

[0154] 处理单元 910 被配置为接收具有多个部分的文档（例如，利用文档接收单元 912），其中所述部分中的至少一些部分与相应的元数据相关联。处理单元 910 被配置为输出对文档的相应部分的语音阅读，包括基于相应的元数据可听地区分相应部分（例如，利用输出单元 914 通过音频输出单元 906）。处理单元 910 被配置为从用户接收请求导航至与特定元数据相关联的特定部分的语音命令（例如，利用语音命令接收单元 916 通过音频输入单元 904）。处理单元 910 被配置为，响应于接收到语音命令，输出对与特定元数据相关联的特定部分的语音阅读（例如，利用输出单元 914 通过音频输出单元 906）。

[0155] 在一些实施例中，处理单元 910 被配置为根据相应部分的相应样式向用户输出对文本的相应部分的语音阅读，具体方式为：使用第一组语音特征向用户输出对文档中文本的第一部分的语音阅读（例如，利用输出单元 914 通过音频输出单元 906）；以及使用不同于第一组语音特征的第二组语音特征来向用户输出对文档中的文本的第二部分的语音阅读（例如，利用输出单元 914 通过音频输出单元 906）。

[0156] 在一些实施例中，处理单元 910 被配置为在输出对文档的一个或多个部分的语音阅读的同时，从用户接收语音命令的至少一部分（例如，利用语音命令接收单元 916 通过音频输入单元 904）。

[0157] 在一些实施例中，文档的相应部分中的至少一个部分为文档的相应段落。

[0158] 在一些实施例中，文档的相应部分中的至少一个部分为文档的相应标题。

[0159] 在一些实施例中，文档的相应部分中的至少一个部分为文档的相应句子。

[0160] 在一些实施例中，文档的第一部分具有第一样式并且文档的第二部分具有不同于第一样式的第二样式。

[0161] 在一些实施例中，文档的每个部分与单个链接相关联。

[0162] 根据一些实施例，图 10 示出了根据如上所述的本发明的原理配置的电子设备 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实行本发明的原理。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为或者被分离为子块以实现如上所述的本发明的原理。因此，本文的描述可支持本文所述的功能块的任何可能的组合或分离或进一步的定义。

[0163] 如图 10 中所示，电子设备 1000 包括音频输入单元 1004 和音频输出单元 1006。在一些实施例中，电子设备 1000 包括被配置为显示电子文档的一个或多个部分的显示单元

1002。在一些实施例中，电子设备 1000 包括触敏表面单元 1008，该触敏表面单元被配置为检测触敏表面单元 1008 上的一个或多个手势。电子设备还包括耦接至音频输入单元 1004 和音频输出单元 1006 的处理单元 1010。在一些实施例中，处理单元 1010 还耦接至显示单元 1002 和触敏表面单元 1008。在一些实施例中，处理单元 1010 包括输出单元 1012、语音命令接收单元 1014 以及识别单元 1016。

[0164] 处理单元 1010 被配置为输出对多个文档中的一个文档的至少一部分的语音阅读（例如，利用输出单元 1012 通过音频输出单元 1006）。处理单元 1010 被配置为，在输出语音阅读的同时，从用户接收请求对应于特定标准的文档的语音命令（例如，利用语音命令接收单元 1014 通过音频输入单元 1004）。处理单元 1010 被配置为，响应于从用户接收到语音命令，识别所述多个文档中的对应于特定标准的一个或多个文档（例如，利用识别单元 1016）；并且输出对所述一个或多个识别的文档的相应文档的至少一部分的语音阅读（例如，利用输出单元 1012 通过音频输出单元 1006）。

[0165] 在一些实施例中，特定标准要求所述一个或多个识别的文档由被用户识别的一位或多作者创作。

[0166] 在一些实施例中，特定标准要求所述一个或多个识别的文档与特定文档相关联。

[0167] 在一些实施例中，特定标准要求所述一个或多个识别的文档是对特定文档的回复。

[0168] 在一些实施例中，特定标准要求所述一个或多个识别的文档包括来自相应作者的最新消息。

[0169] 在一些实施例中，特定标准要求所述一个或多个识别的文档对应于被用户所识别的特定日期范围。

[0170] 在一些实施例中，处理单元 1010 被配置为输出指示所述一个或多个识别的文档的数量的可听信息（例如，利用输出单元 1012 通过音频输出单元 1006）。

[0171] 在一些实施例中，处理单元 1010 被配置为输出指示相应文档为所述一个或多个识别的文档的最后一个文档的可听信息（例如，利用输出单元 1012 通过音频输出单元 1006）。

[0172] 出于解释的目的，前面的描述是通过参考具体实施例来进行描述的。然而，上面的示例性讨论并旨在是穷尽的或要将本发明限制到所公开的精确形式。根据以上教导内容，很多修改和变型都是可能的。选择和描述实施例是为了充分阐明本发明的原理及其实际应用，以由此使得本领域的其他技术人员能够充分利用具有适合于所构想的特定用途的各种修改的本发明以及各种实施例。

[0173] 例如，相对于图 6 所述的操作的一个或多个方面可与相对于图 7 和图 8 所述的操作（例如，在到达文档的结尾之后输出可听询问）一起使用。类似地，相对于图 7 所述的操作的一个或多个方面可与相对于图 6 和图 8 所述的操作一起使用，并且相对于图 8 所述的操作的一个或多个方面可与相对于图 6 和图 7 所述的操作一起使用。为简明起见，这些细节不再重复。

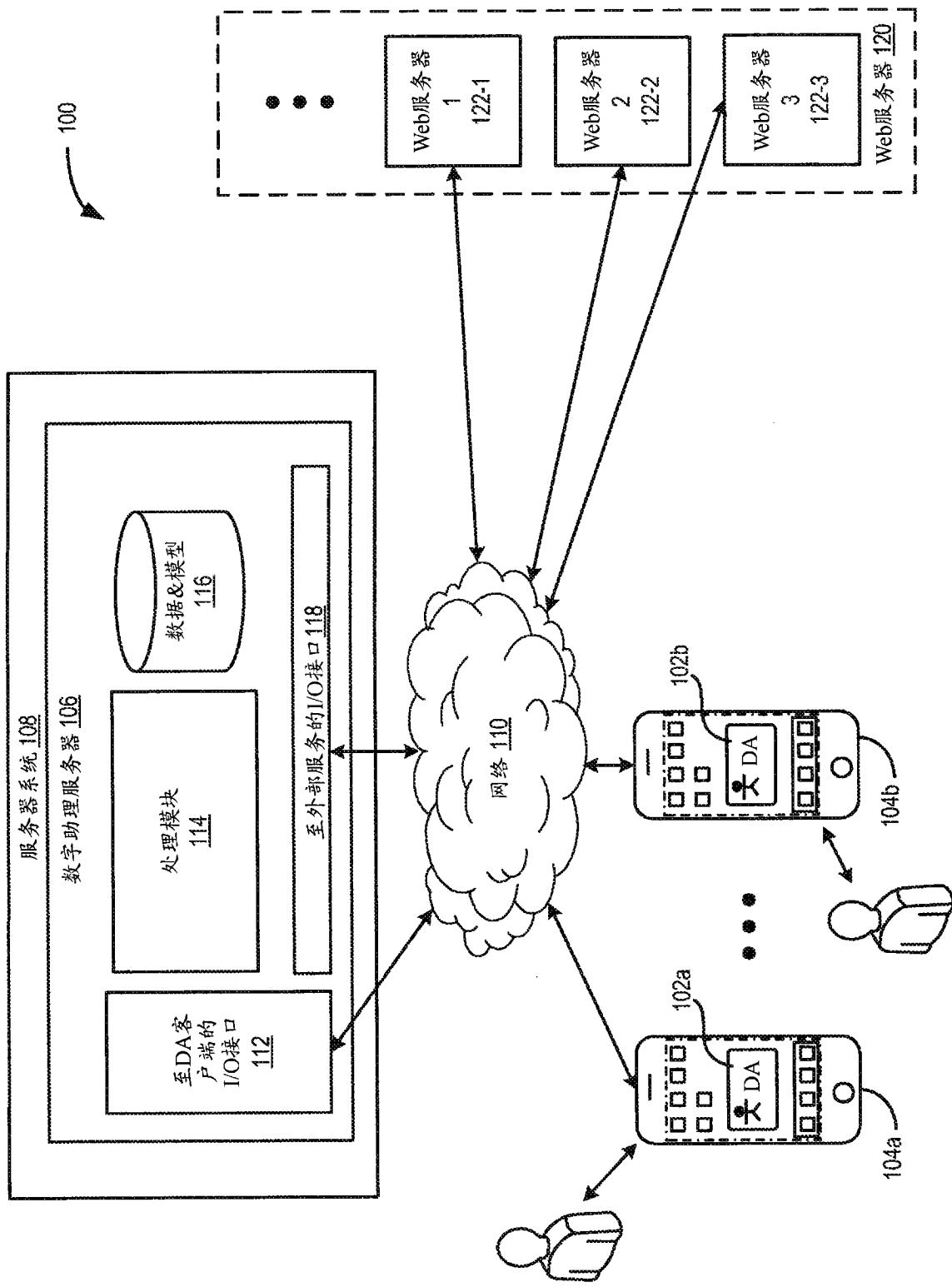


图 1

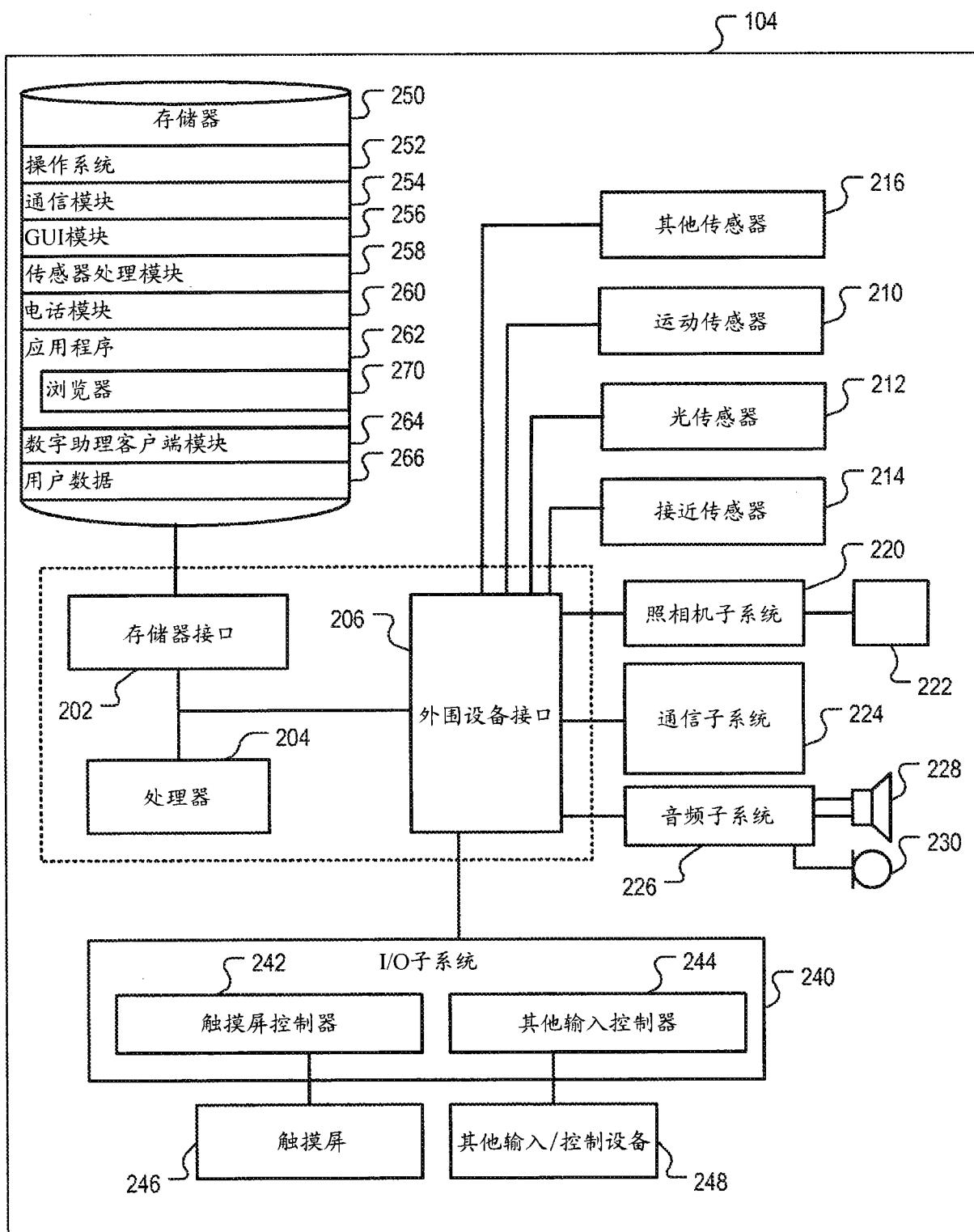


图 2

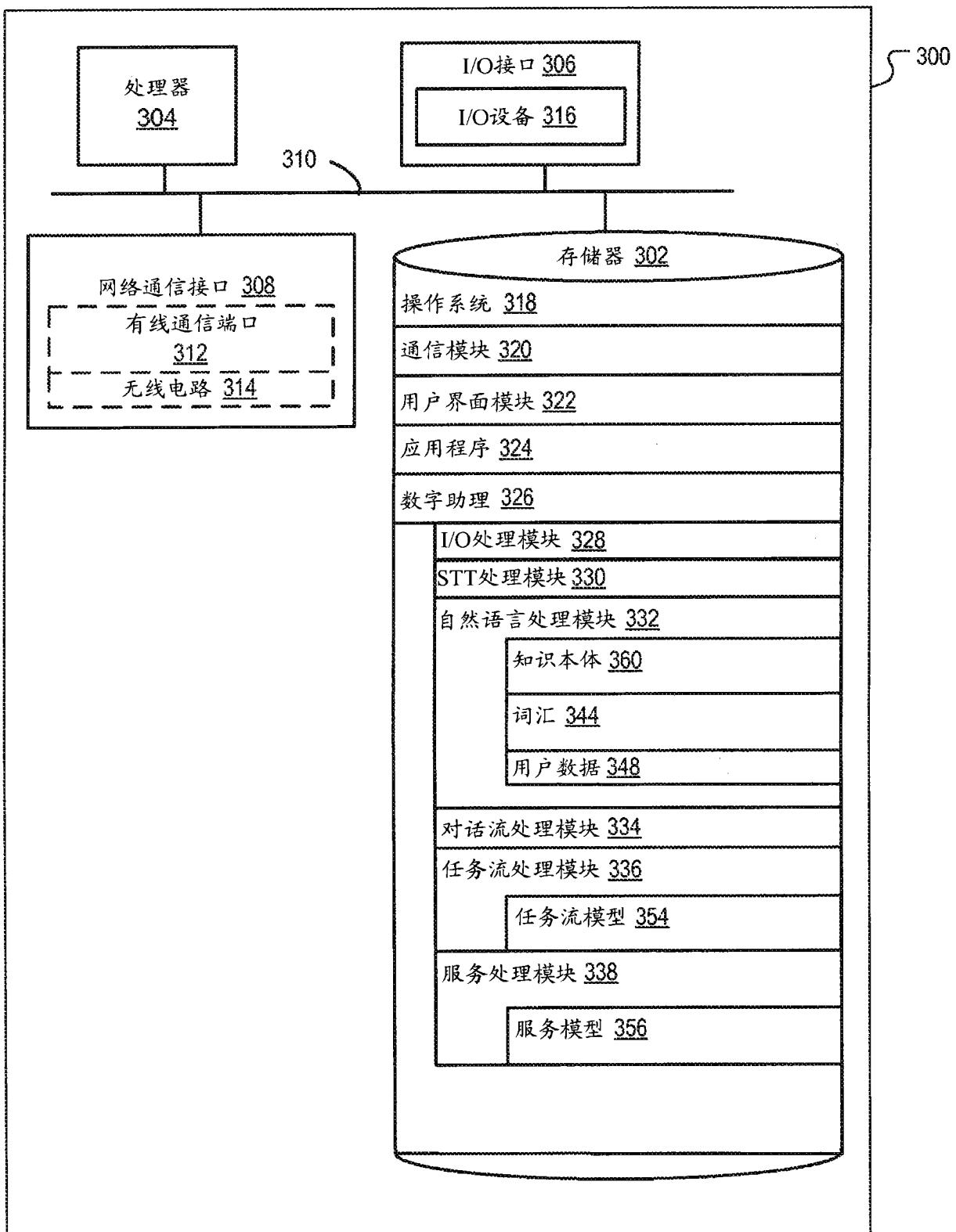


图 3

多功能设备100



八十七年前，我们的先辈在这块大陆上建立了一个以自由为构想并且奉献于人人平等的主张的新国家。

现在我们参与了一场考验这个国家或如此构想并且如此奉献的任何一个国家能否长久存在的大内战。我们在这场战争的一个大战场上相遇。我们把战场的一角奉献给为国家生存而献出他们的生命的那些人们，作为他们最终的安息之地。这是我们义不容辞、理所当然做的事。

但是，从更深刻的意义来说，我们不能使这一角战场成为圣地，我们不能使它流芳百世，我们不能使它永垂青史。因为在这里战斗过的勇士们，活着的和死去的，已经使战场神圣化了，我们微薄的力量远远不能为它增光或者使它减色。世人不太会注意，也不会

触摸屏 112

图 4A

多功能设备100 ↗



八十七年前，我们的先辈在这块大陆上建立了一个以自由为构想并且奉献于人人平等的主张的新国家。

现在我们参与了一场考验这个国家或如此构想并且如此奉献的任何一个国家能否长久存在的大内战。我们在这场战争的一个大战场上相遇。我们把战场的一角奉献给为国家生存而献出他们的生命的那些人们，作为他们最终的安息之地。这是我们义不容辞、理所当然做的事。

但是，从更深刻的意义来说，我们不能使这一角战场成为圣地，我们不能使它流芳百世，我们不能使它永垂青史。因为在这里战斗过的勇士们，活着的和死去的，已经使战场神圣化了，我们微薄的力量远远不能为它增光或者使它减色。世人不太会注意，也不会

触摸屏112 ↘

图 4B

多功能设备100 ↗



八十七年前，我们的先辈在这块大陆上建立了一个以自由为构想并且奉献于人人平等的主张的新国家。

现在我们参与了一场考验这个国家或如此构想并且如此奉献的任何一个国家能否长久存在的大内战。我们在这场战争的一个大战场上相遇。我们把战场的一角奉献给为国家生存而献出他们的生命的那些人们，作为他们最终的安息之地。这是我们义不容辞、理所当然做的事。

但是，从更深刻的意义来说，我们不能使这一角战场成为圣地，我们不能使它流芳百世，我们不能使它永垂青史。因为在这里战斗过的勇士们，活着的和死去的，已经使战场神圣化了，我们微薄的力量远远不能为它增光或者使它减色。世人不太会注意，也不会

触摸屏112 ↗

图 4C

多功能设备100 ↗



八十七年前，我们的先辈在这块大陆上建立了一个以自由为构想并且奉献于人人平等的主张的新国家。

现在我们参与了一场考验这个国家或如此构想并且如此奉献的任何一个国家能否长久存在的大内战。我们在这场战争的一个大战场上相遇。我们把战场的一角奉献给为国家生存而献出他们的生命的那些人们，作为他们最终的安息之地。这是我们义不容辞、理所当然做的事。

但是，从更深刻的意义来说，我们不能使这一角战场成为圣地，我们不能使它流芳百世，我们不能使它永垂青史。因为在这里战斗过的勇士们，活着的和死去的，已经使战场神圣化了，我们微薄的力量远远不能为它增光或者使它减色。世人不太会注意，也不会

触摸屏112 ↘

图 4D

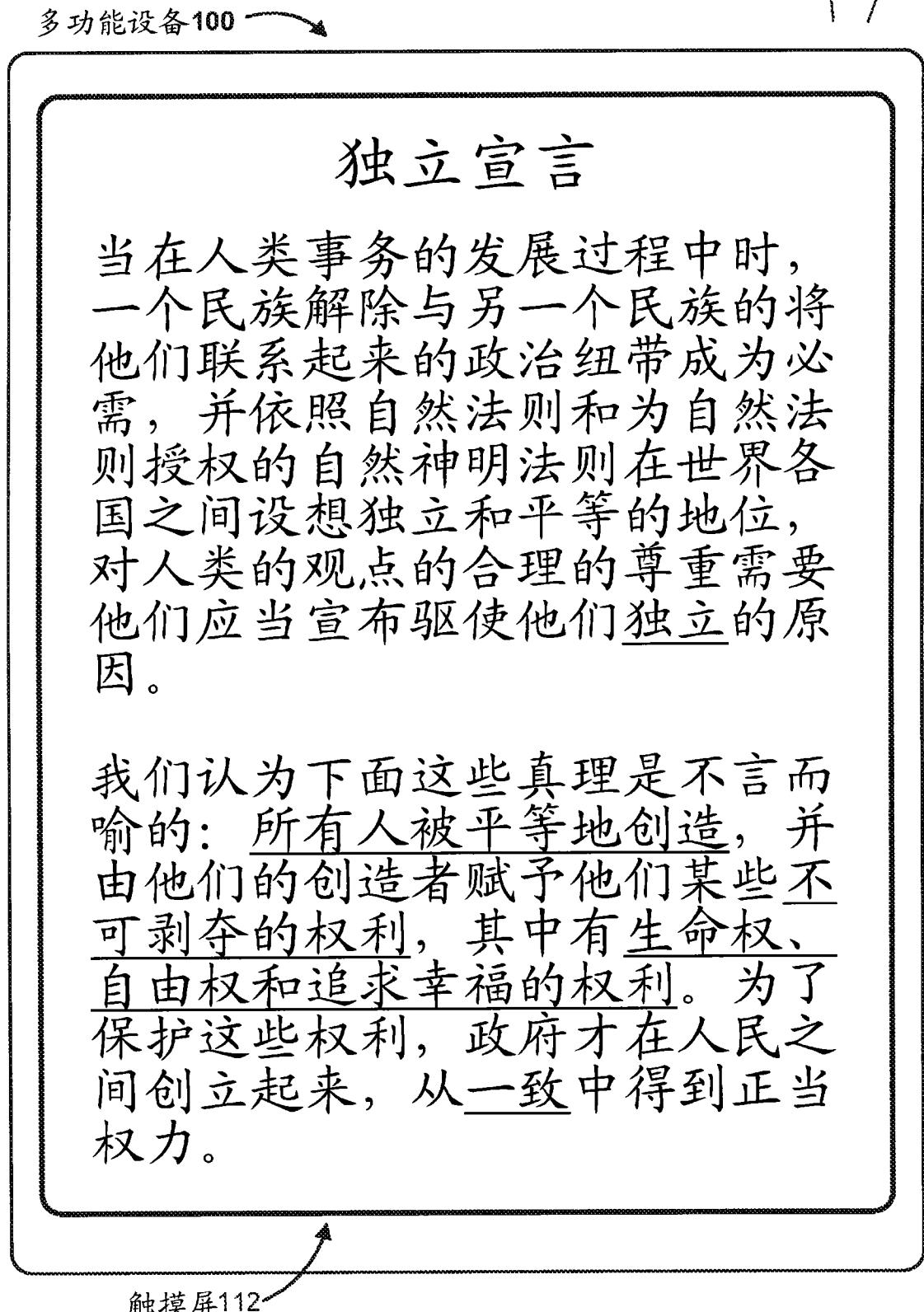


图 4E

多功能设备 100



独立宣言

当在人类事务的发展过程中时，一个民族解除与另一个民族的将他们联系起来的政治纽带成为必需，并依照自然法则和为自然法则授权的自然神明法则在世界各国之间设想独立和平等的地位，对人类的观点的合理的尊重需要他们应当宣布驱使他们独立的原因。

我们认为下面这些真理是不言而喻的：所有人被平等地创造，并由他们的创造者赋予他们某些不可剥夺的权利，其中有生命权、自由权和追求幸福的权利为了保护这些权利，政府才在人民之间创立起来，从一致中得到正当权力。

触摸屏 112

图 4F

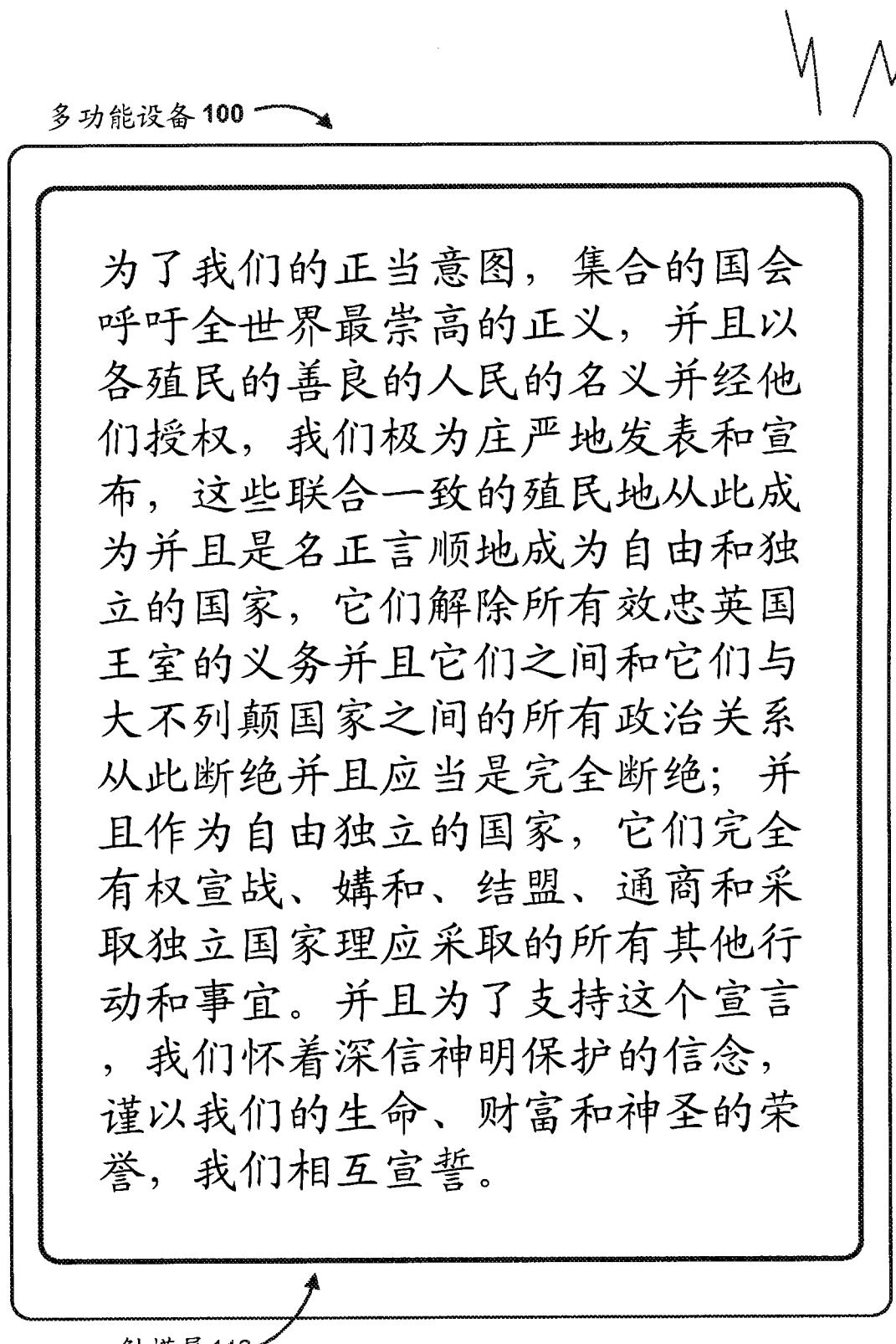


图 4G

多功能设备100 ↗



八十七年前，我们的先辈在这块大陆上建立了一个以自由为构想并且奉献于人人平等的主张的新国家。

现在我们参与了一场考验这个国家或如此构想并且如此奉献的任何一个国家能否长久存在的大内战。我们在这场战争的一个大战场上相遇。我们把战场的一角奉献给为国家生存而献出他们的生命的那些人们，作为他们最终的安息之地。这是我们义不容辞、理所当然做的事。

但是，从更深刻的意义来说，我们不能使这一角战场成为圣地，我们不能使它流芳百世，我们不能使它永垂青史。因为在这里战斗过的勇士们，活着的和死去的，已经使战场神圣化了，我们微薄的力量远远不能为它增光或者使它减色。世人不太会注意，也不会

触摸屏112 ↘

图 4H

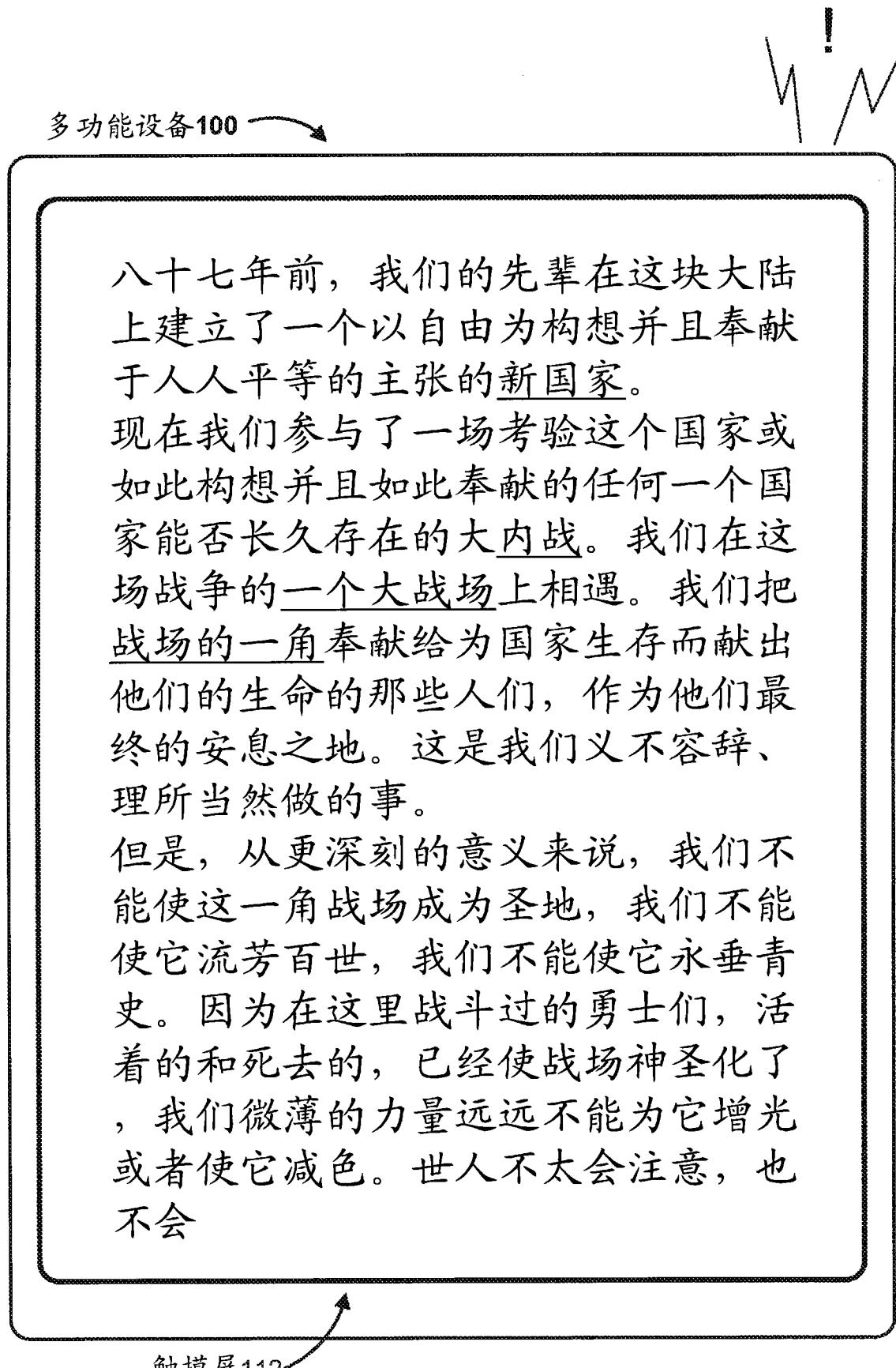


图 4I

多功能设备100 ↗

《葛底斯堡演讲》是由美国总统亚伯拉罕林肯发表的演讲。以下是《葛底斯堡演讲》的文本：

八十七年前，我们的先辈在这块大陆上建立了一个以自由为构想并且奉献于人人平等的主张的新国家。

现在我们参与了一场考验这个国家或如此构想并且如此奉献的任何一个国家能否长久存在的大内战。我们在这场战争的一个大战场上相遇。我们把战场的一角奉献给为国家生存而献出他们的生命的那些人们，作为他们最终的安息之地。这是我们义不容辞、理所当然做的事。

但是，从更深刻的意义来说，我们不能使这一角战场成为圣地，我们不能使它流芳百世，我们不能使它永垂青史

触摸屏112 ↗

图 4J

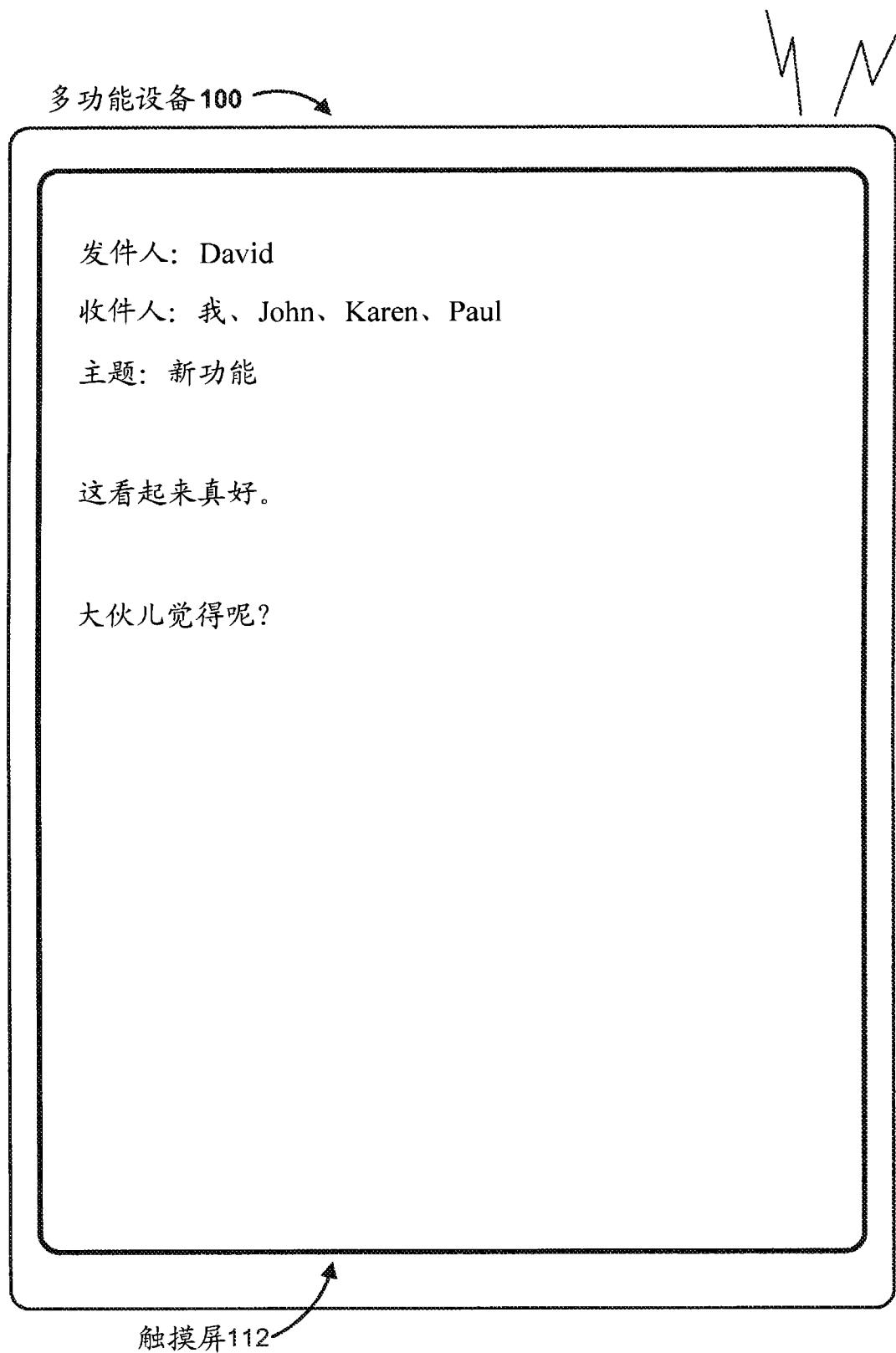


图 4K

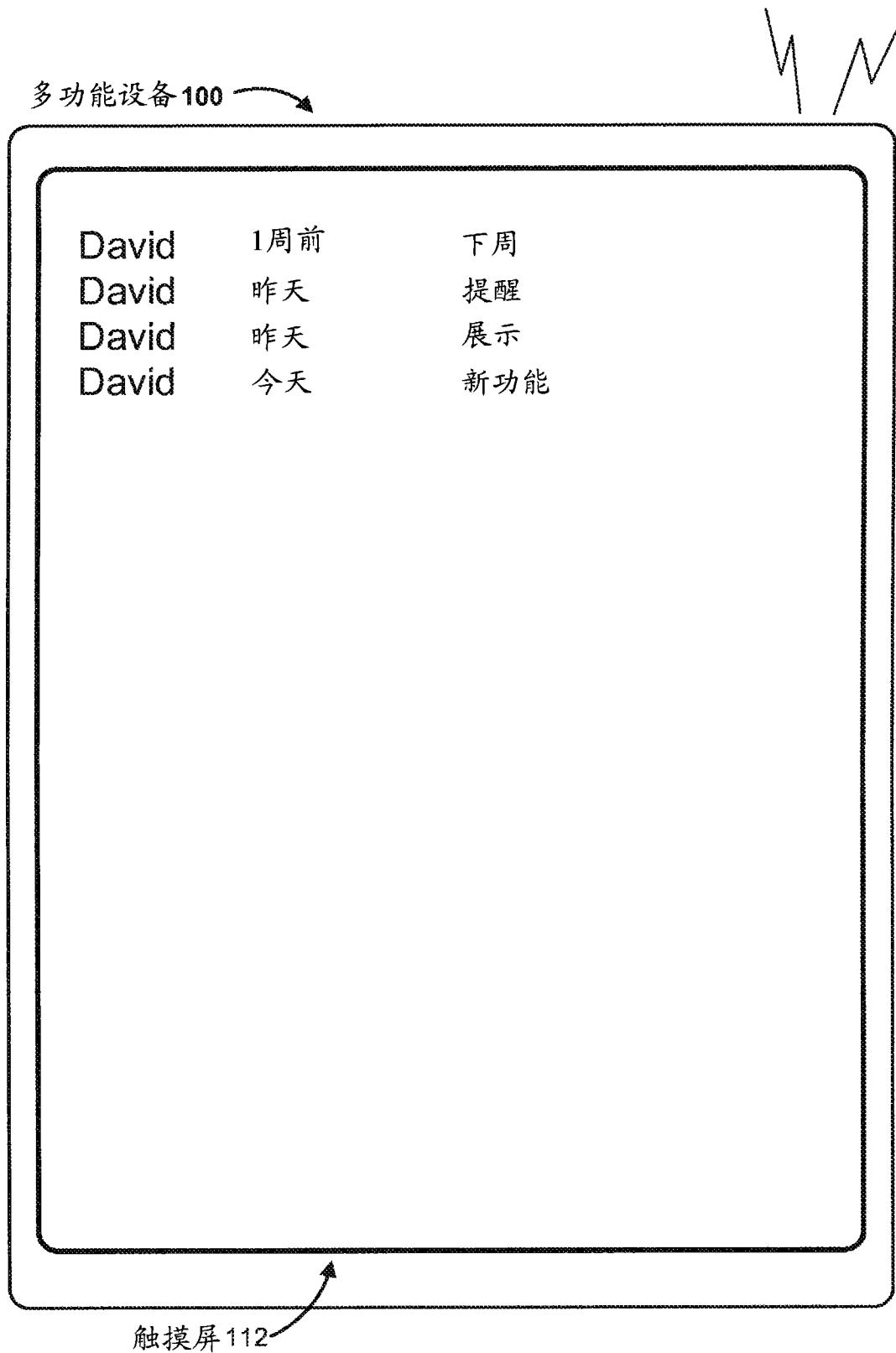


图 4L

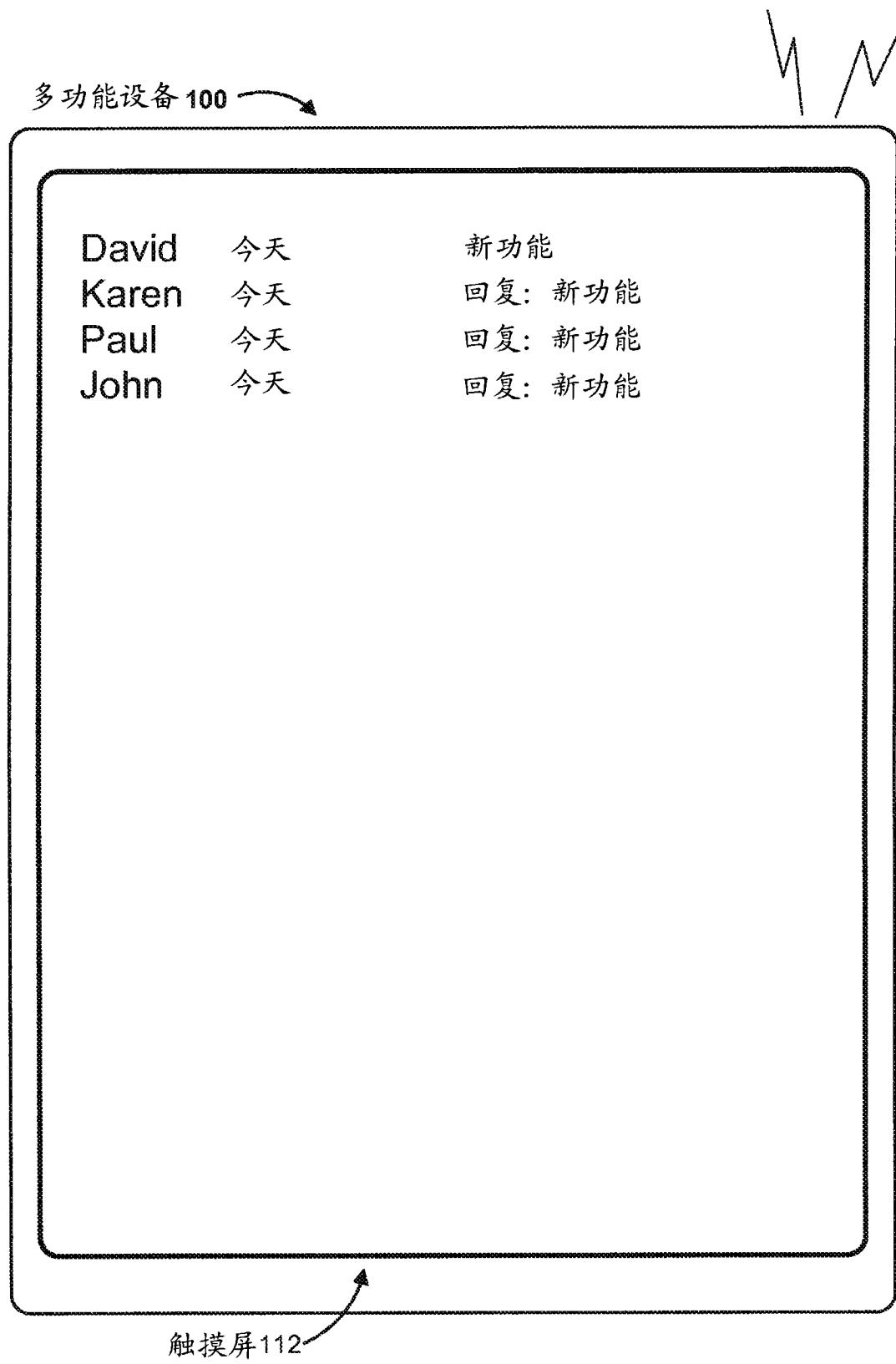


图 4M

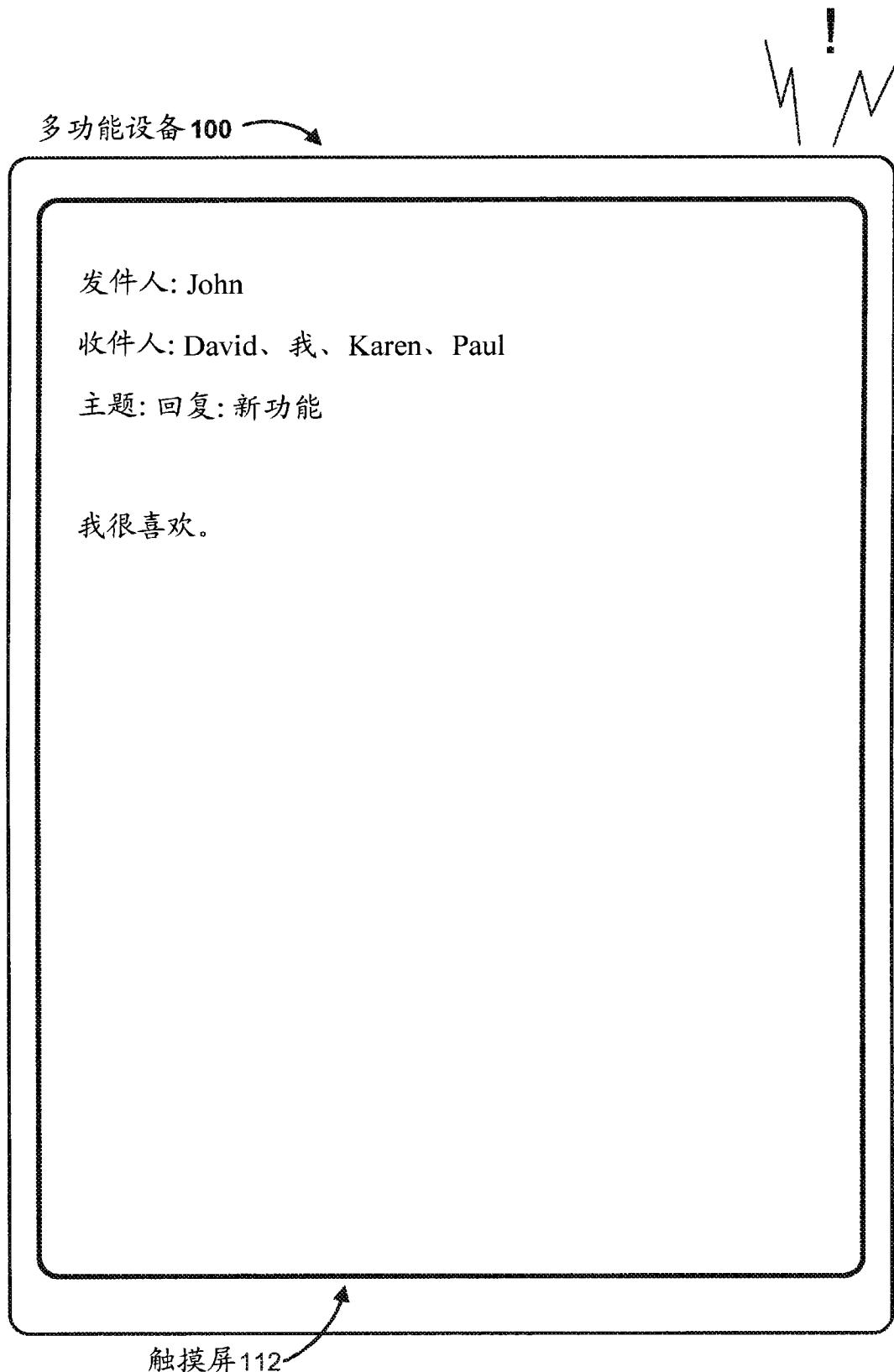


图 4N

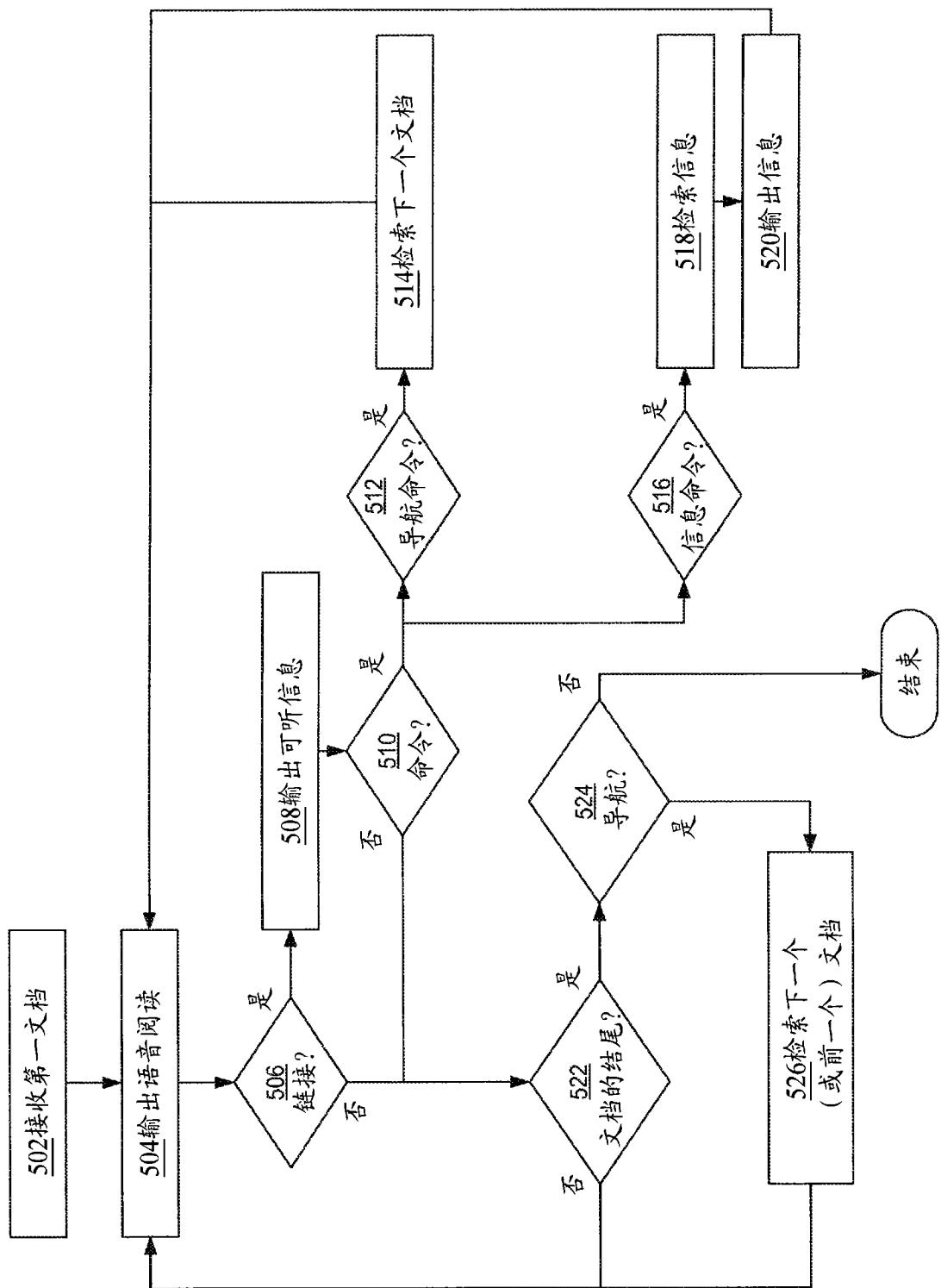


图 5

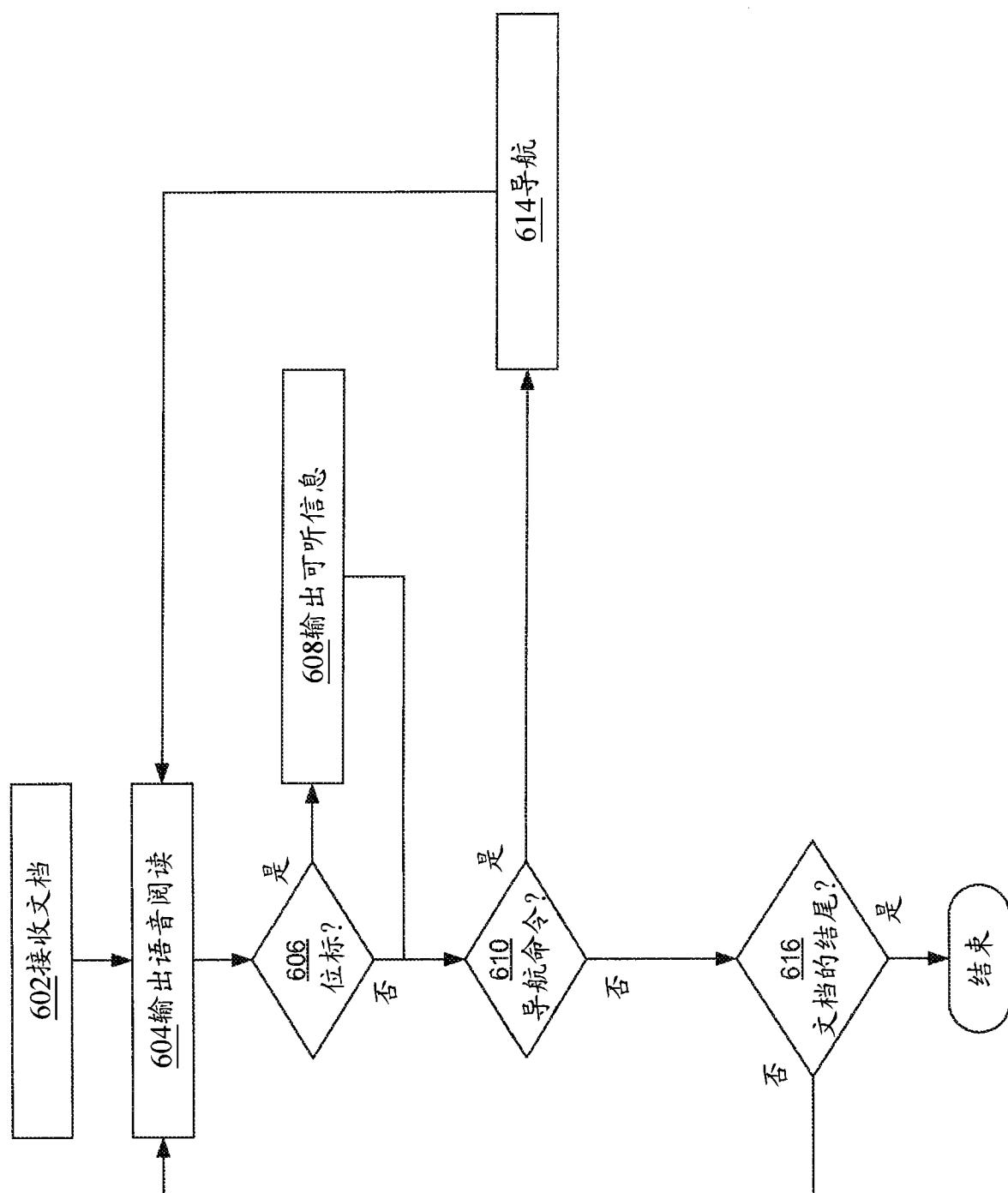


图 6

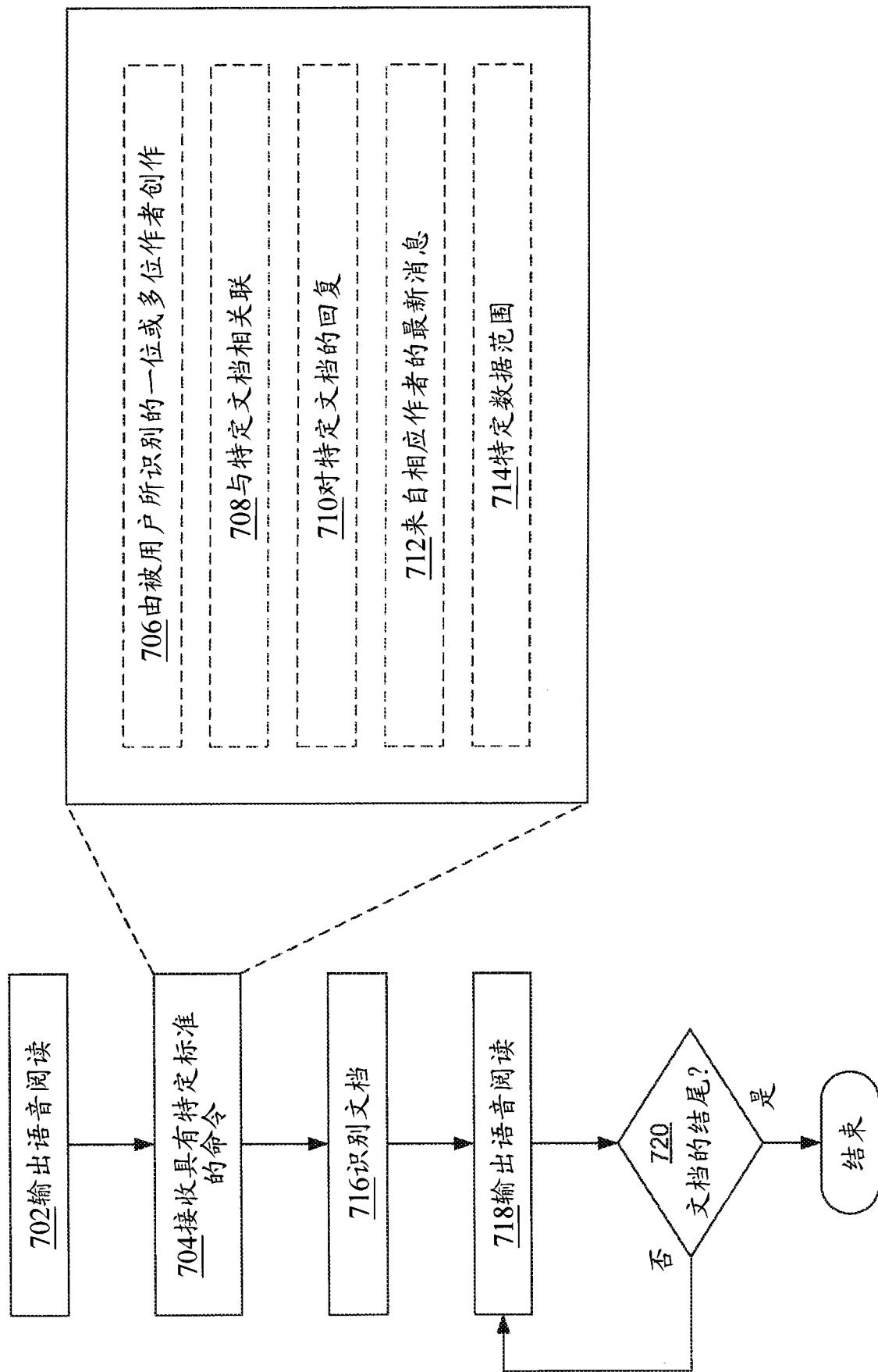


图 7

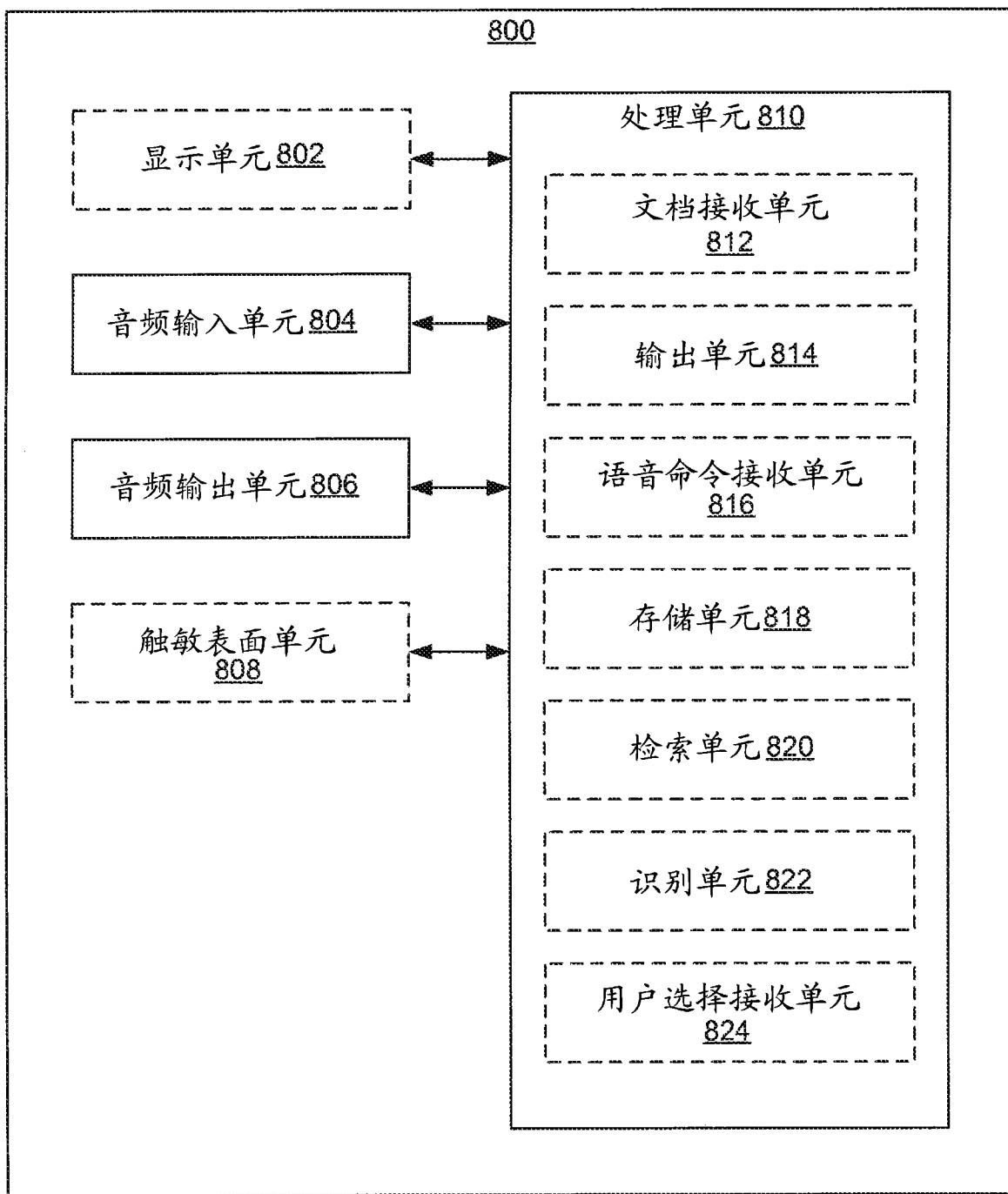


图 8

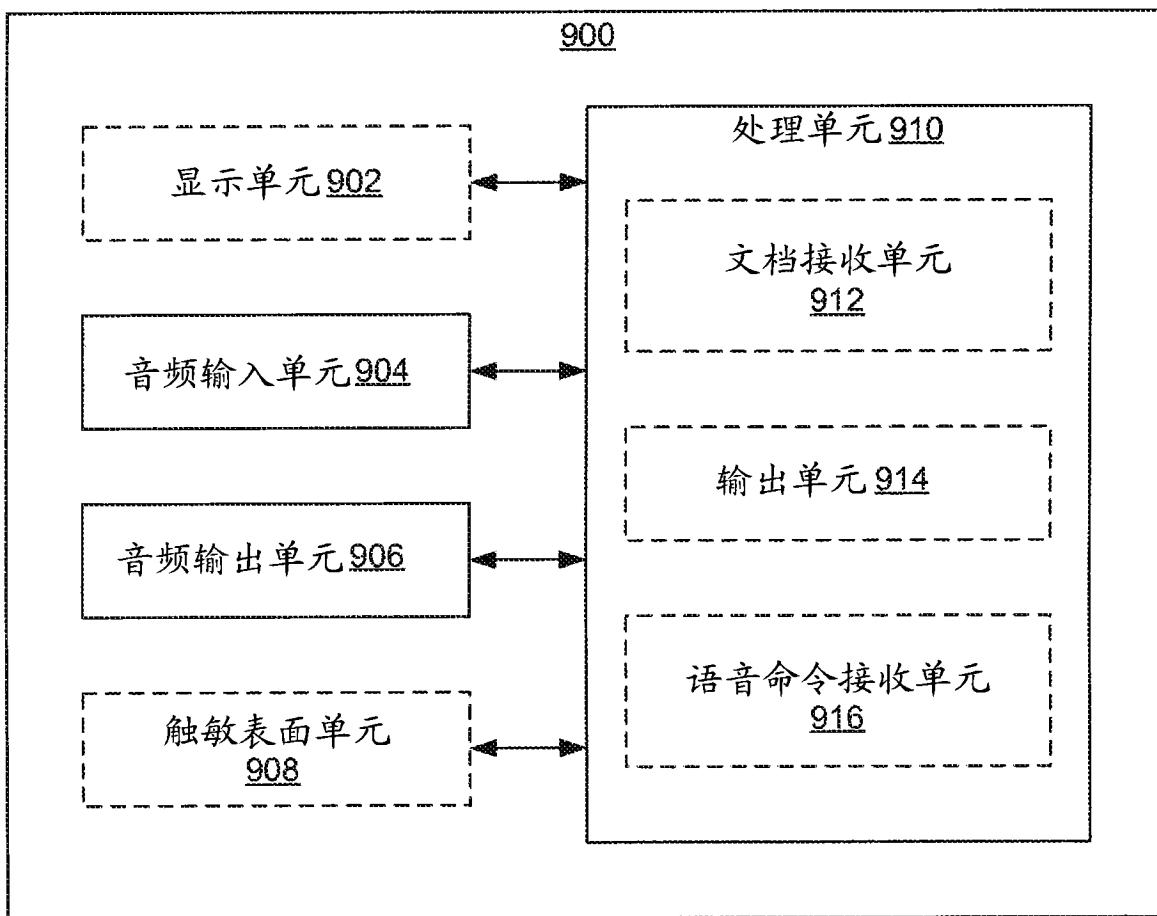


图 9

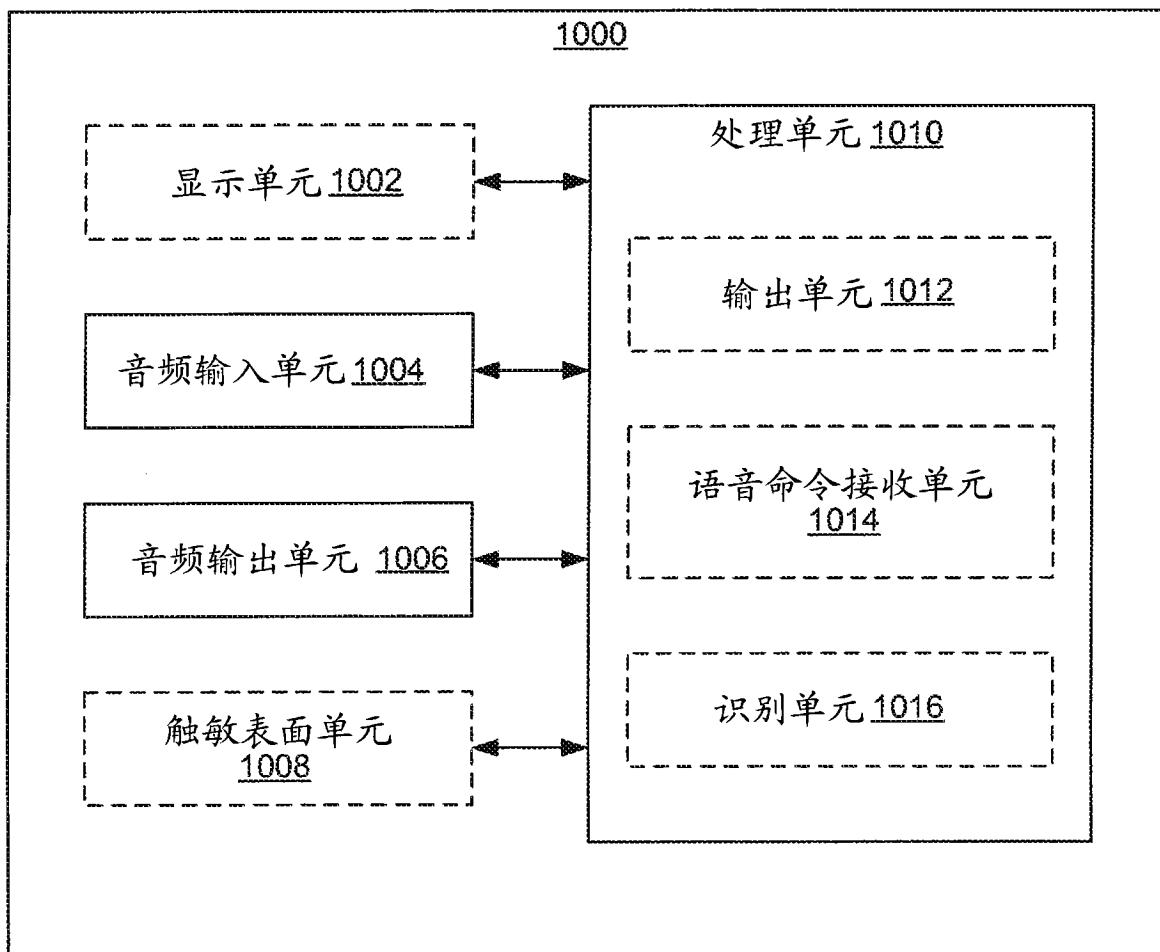


图 10