



(12) 发明专利申请

(10) 申请公布号 CN 112204654 A

(43) 申请公布日 2021.01.08

(21) 申请号 201980026154.9

(22) 申请日 2019.02.15

(30) 优先权数据

62/630,979 2018.02.15 US

(85) PCT国际申请进入国家阶段日

2020.10.15

(86) PCT国际申请的申请数据

PCT/US2019/018235 2019.02.15

(87) PCT国际申请的公布数据

WO2019/161216 EN 2019.08.22

(71) 申请人 得麦股份有限公司

地址 美国加利福尼亚州

(72) 发明人 A·达恩

(74) 专利代理机构 上海专利商标事务所有限公
司 31100

代理人 何焜 张鑫

(51) Int.Cl.

G10L 15/22 (2006.01)

G10L 15/30 (2013.01)

G10L 25/63 (2013.01)

G06K 9/00 (2006.01)

G06F 16/332 (2019.01)

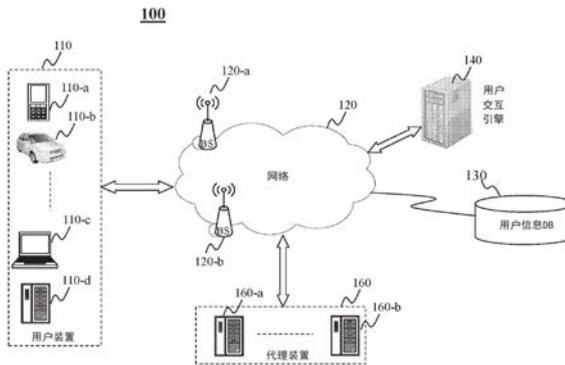
权利要求书3页 说明书25页 附图27页

(54) 发明名称

用于基于预测的先发式对话内容生成的系统和方法

(57) 摘要

本示教涉及一种用于管理用户机器对话的方法、系统、介质和实施方式。在装置上接收与对话有关的信息，其中，用户用该装置参与对话。基于与对话有关的信息，驻留在装置上的本地对话管理器关于与存储在装置上的预测对话路径相关联的预测响应，搜索将要给与用户的响应。预测对话路径、预测响应和本地对话管理器是基于驻留在服务器上的对话树先发式地生成的。如果响应被本地对话管理器识别出，响应被发送到装置。如果响应没有被本地对话管理器识别出，装置向服务器发送对于响应的请求。



1. 一种用于管理用户机器对话的、在至少一个机器上实现的方法,该机器包含至少一个处理器、存储器和能够连接到网络的通信平台,该方法包含:

在装置上接收与对话有关的信息,其中,用户用装置参与对话;

由驻留在装置上的本地对话管理器,基于与对话有关的信息,关于与存储在装置上的预测对话路径相关联的预测响应,搜索将要给与用户的响应,其中,预测对话路径、预测响应和本地对话管理器是基于驻留在服务器上的对话树先发式地生成的;

如果响应被本地对话管理器识别出,响应于语音,将响应发送给用户;以及

如果响应没有被本地对话管理器识别出,向服务器发送对于响应的请求。

2. 权利要求1的方法,其中,与对话有关的信息包括用户的言语、对对话周边情况的观察以及对观察的描绘中的至少一者。

3. 权利要求2的方法,其中,

对对话周边情况的观察包括对用户以及对话场景中至少一者的观察;

对用户的观察包括面部表情、姿态、用户的移动以及言语的音调中的一种以上;且

对对话场景的观察包括在场景中存在的对象以及对话场景中的声音中的一种以上。

4. 权利要求1的方法,其中,请求和与对话有关的信息一起被发送到服务器。

5. 权利要求1的方法,还包含:

在请求被发送到服务器后,从服务器接收响应、更新的预测对话路径、更新的预测响应以及更新的本地对话管理器;以及

将从服务器接收的响应发送给用户。

6. 权利要求5的方法,其中:

更新的预测对话路径由服务器基于响应、对话树和/或与对话有关的信息先发式地生成,

更新的预测响应由服务器基于更新的预测对话路径和对话树先发式地生成,且

更新的本地对话管理器由服务器基于更新的预测对话路径和更新的预测响应生成。

7. 权利要求5的方法,还包含:

基于从服务器接收的更新的预测对话路径,对预测对话路径进行更新;

基于更新的预测响应,对预测响应进行更新;以及

基于从服务器接收的更新的本地对话管理器,对装置上的本地对话管理器进行更新。

8. 一种机器可读的非暂时性介质,其上记录有用于管理用户机器对话的信息,其中,该信息在被机器读取时,使得机器执行:

在装置上接收与对话有关的信息,其中,用户用装置参与对话;

由驻留在装置上的本地对话管理器,基于与对话有关的信息,关于与存储在装置上的预测对话路径相关联的预测响应,搜索将要给与用户的响应,其中,预测对话路径、预测响应和本地对话管理器是基于驻留在服务器上的对话树先发式地生成的;

如果响应被本地对话管理器识别出,响应于语音,将响应发送给用户;以及

如果响应没有被本地对话管理器识别出,向服务器发送对于响应的请求。

9. 权利要求8的介质,其中,与对话有关的信息包括用户的言语、对对话周边情况的观察以及对观察的描绘中的至少一者。

10. 权利要求9的介质,其中,

对话周边情况的观察包括对用户以及对话场景中至少一者的观察；
对用户的观察包括面部表情、姿态、用户的移动以及言语的音调中的一种以上；且
对话场景的观察包括在场景中存在的对象以及对话场景中的声音中的一种以上。

11. 权利要求8的介质,其中,请求和与对话有关的信息一起被发送到服务器。

12. 权利要求8的介质,还包含:

在请求被发送到服务器后,从服务器接收响应、更新的预测对话路径、更新的预测响应以及更新的本地对话管理器;以及

将从服务器接收的响应发送给用户。

13. 权利要求12的介质,其中:

更新的预测对话路径由服务器基于响应、对话树和/或与对话有关的信息先发式地生成;

更新的预测响应由服务器基于更新的预测对话路径和对话树先发式地生成;且
更新的本地对话管理器由服务器基于更新的预测对话路径和更新的预测响应生成。

14. 权利要求12的介质,其中,该信息在被机器读取时,还使机器执行:

基于从服务器接收的更新的预测对话路径,对预测对话路径进行更新;

基于更新的预测响应,对预测响应进行更新;以及

基于从服务器接收的更新的本地对话管理器,对装置上的本地对话管理器进行更新。

15. 一种用于管理用户机器对话的系统,包含:

对话状态分析器,其被配置为,在装置上接收与对话有关的信息,其中,用户用装置参与对话;

驻留在装置上的本地对话管理器,其被配置为,基于与对话有关的信息,关于与存储在装置上的预测对话路径相关联的预测响应,搜索将要给与用户的响应,其中,预测对话路径、预测响应和本地对话管理器是基于驻留在服务器上的对话树先发式地生成的;

响应发送器,其被配置为,如果响应被本地对话管理器识别出,响应于语音,将响应发送给用户;以及

装置/服务器协调器,其被配置为,如果响应没有被本地对话管理器识别出,向服务器发送对于响应的请求。

16. 权利要求15的系统,其中,与对话有关的信息包括用户的言语、对话周边情况的观察以及对观察的描绘中的至少一者。

17. 权利要求16的系统,其中,

对话周边情况的观察包括对用户以及对话场景中至少一者的观察;

对用户的观察包括面部表情、姿态、用户的移动以及言语的音调中的一种以上;且
对话场景的观察包括在场景中存在的对象以及对话场景中的声音中的一种以上。

18. 权利要求15的系统,其中,请求和与对话有关的信息一起被发送到服务器。

19. 权利要求15的系统,其中:

装置/服务器协调器被进一步配置为,在请求被发送到服务器后,从服务器接收响应、更新的预测对话路径、更新的预测响应以及更新的本地对话管理器。

20. 权利要求19的系统,其中:

更新的预测对话路径由服务器基于响应、对话树和/或与对话有关的信息先发式地生

成，

更新的预测响应由服务器基于更新的预测对话路径和对话树先发式地生成，且更新的本地对话管理器由服务器基于更新的预测对话路径和更新的预测响应生成。

21. 权利要求19的系统，还包含本地对话信息更新器，其被配置为：

基于从服务器接收的更新的预测对话路径，对预测对话路径进行更新；

基于更新的预测响应，对预测响应进行更新；以及

基于从服务器接收的更新的本地对话管理器，对装置上的本地对话管理器进行更新。

用于基于预测的先发式对话内容生成的系统和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2018年2月15日提交的美国临时申请62/630,979的优先权,其内容全文并入此处作为参考。

[0003] 本申请与2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461788)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503025)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461789)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503026)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461790)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503027)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461808)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503028)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461809)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503029)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461810)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503030)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461819)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503031)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461811)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0502960)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0502961)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503035)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0502963)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0502966)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0502964)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0502965)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503032)、2019年2月15日提交的国际申请_____ (代理人案卷号047437-0461820)、2019年2月15日提交的美国专利申请_____ (代理人案卷号047437-0503017)有关,其全部内容并入此处作为参考。

技术领域

[0004] 本示教一般涉及计算机。具体而言,本示教涉及计算机化的对话代理。

背景技术

[0005] 由于互联网连接无处不在,带来人工智能技术的进步和基于互联网的通信的蓬勃发展,因此,计算机辅助的对话系统日渐普及。例如,越来越多的呼叫中心配置自动对话机器人来处理用户呼叫。酒店开始安装能够回答旅客或客人问题的多种售货亭。在线预订(无论是旅游住宿还是剧场票务等)也越来越频繁地用聊天机器人完成。近些年来,其他领域内

的自动人机通信也变得越来越普遍。

[0006] 基于不同领域中公知的会话模式,这样的传统计算机辅助对话系统通常用特定的问题和回答预先编程。不巧的是,人类会话者可能无法预测,有时不会遵从预先计划的对话模式。另外,在某些情况下,人类会话者可能在该过程中离题,继续固定的会话模式可能令人恼火或失去兴趣。当这种情况发生时,这种机器传统对话系统常常不能继续吸引人类会话者参与,故使人机对话或者中辍、将任务交给人类操作员,或者,人类会话者直接离开对话,而这是不希望看到的。

[0007] 另外,传统的基于机器的对话系统常常并非设计为处理人的情绪因素,更不用说在与人类进行会话时将如何处理这种情绪因素考虑在内。例如,传统的机器对话系统常常不会发起会话,除非人启动系统或问某些问题。即使传统对话系统发起会话,它具有开始会话的固定方式,不会因人而异或是基于观察进行调整。因此,尽管它们被编程为忠实遵循预先设计的对话模式,它们通常不能就会话的动态发展做出行动以及进行自适应,以便使会话以能吸引人参与的方式进行。在许多情况下,当涉入对话的人明显心烦或感到不满时,传统的机器对话系统浑然不觉,还以使那人心烦的同种方式继续会话。这不仅使得会话不愉快地结束(机器对此还是不知情),而且使那人在将来不愿意与任何基于机器的对话系统进行会话。

[0008] 在某些应用中,为了判断如何有效进行下去,基于从人观察到的内容来执行人机对话线程是至为重要的。一个例子是教育相关的对话。当聊天机器人用于教孩子阅读时,必须监视孩子是否以正被施教的方式具有感知力,并持续处理以便有效进行。传统对话系统的另一限制是它们对背景的毫无意识。例如,传统的对话系统不具有这样的能力:观察会话的背景并即兴产生对话策略,从而吸引用户参与,并改善用户体验。

[0009] 因此,存在对解决这些限制的方法和系统的需求。

发明内容

[0010] 这里公开的示教涉及用于计算机化的对话代理的方法、系统和程序设计。

[0011] 在一实例中,公开了一种在机器上实现的方法,该机器具有至少一个处理器、存储器和能够连接到网络的通信平台,该方法用于管理用户机器对话。与对话有关的信息在装置上被接收,其中,用户用该装置参与对话。基于与对话有关的信息,驻留在装置上的本地对话管理器关于与存储在装置上的预测对话路径相关联的预测响应,搜索将要给与用户的响应。预测对话路径、预测响应和本地对话管理器基于驻留在服务器上的对话树先发式地(preemptively)生成。如果本地对话管理器识别出响应,响应被发送给装置。如果本地对话管理器没有识别出响应,装置向服务器发送对响应的请求。

[0012] 在一不同的实例中,用于管理用户机器对话的系统。系统包含装置,该装置包含对话状态分析器、本地对话管理器、响应发送器以及装置/服务器协调器。对话状态分析器被配置为,在装置上,接收与对话有关的信息,其中,用户用该装置参与对话。本地对话管理器驻留在装置上,并被配置为,基于与对话有关的信息,关于与存储在装置上的预测对话路径相关联的预测响应,搜索将要给与用户的响应,其中,预测对话路径、预测响应以及本地对话管理器是基于驻留在服务器上的对话树先发式地生成的。响应发送器被配置为,响应于语音,将响应发送给用户,如果响应被本地对话管理器识别的话。装置/服务器协调器被配

置为,如果响应没有被本地对话管理器识别,向服务器发送对于响应的请求。

[0013] 其他的概念涉及实现本示教的软件。根据此概念的软件产品包含至少一个机器可读的非暂时性介质以及由该介质承载的信息。由该介质承载的信息可以是可执行程序代码数据、与可执行程序代码相关联的参数和/或与用户、请求、内容或其它附加信息有关的信息。

[0014] 在一实例中,机器可读的非暂时性有形介质上记录有用于管理用户机器对话的数据,其中,该介质在由机器读取时使得机器执行一系列的步骤。在装置上接收与对话有关的信息,其中,用户用装置参与对话。基于与对话有关的信息,驻留在装置上的本地对话管理器关于与存储在装置上的预测对话路径相关联的预测响应检索将要给予用户的响应。预测对话路径、预测响应和本地对话管理器基于驻留在服务器上的对话树先发式地生成。如果响应被本地对话管理器识别,响应被发送给装置。如果响应没有被本地对话管理器识别,装置向服务器发送对响应的请求。

[0015] 其他的优点和新特征将部分在下面的说明书中给出,部分将由本领域技术人员在检视下面的说明书和附图时明了或通过制造或运行实例来习得。本示教的优点可通过实践和应用下面讨论的详细实例中给出的方法、设备以及组合的多种实施形态来实现和获得。

附图说明

[0016] 这里介绍的方法、系统和/或程序设计进一步以示例性实施例的方式描述。这些示例性实施例参照附图详细介绍。这些实施例是非限制性的示例性实施例,其中,贯穿几幅附图,类似的参考标号代表类似的结构,其中:

[0017] 图1示出了根据本示教一实施例,用于促成操作用户装置的用户与连同用户交互引擎的代理装置之间的对话的网络环境;

[0018] 图2A-2B示出了根据本示教一实施例,在对话过程中,在用户装置、代理装置和用户交互引擎之间的连接;

[0019] 图3A示出了根据本示教一实施例,具有示例性类型的代理身体的代理装置的示例性结构;

[0020] 图3B示出了根据本示教一实施例的示例性代理装置;

[0021] 图4A示出了根据本示教多种实施例,用于自动伴侣的整体系统的示例性高层次系统图;

[0022] 图4B示出了根据本示教一实施例,具有基于自动伴侣与用户之间的交互采取的路径的正在进行的对话的对话树的一部分;

[0023] 图4C示出了根据本示教一实施例的示例性人类-代理装置交互和由自动伴侣执行的示例性处理;

[0024] 图5示出了根据本示教一实施例,在自动对话伴侣的不同处理层之间的示例性多层处理和通信;

[0025] 图6示出了根据本示教一实施例,用于基于人工智能的教育伴侣的示例性高层次系统框架;

[0026] 图7示出了人机对话系统的装置-服务器构造;

[0027] 图8示出了根据本示教一实施例,针对人机对话管理的示例性框架;

- [0028] 图9示出了根据本示教一实施例,用于人机对话管理的装置的的示例性高层次系统图;
- [0029] 图10为根据本示教一实施例,用于人机对话管理的装置的示例性过程的流程图;
- [0030] 图11为根据本示教一实施例,用于人机对话管理的服务器的示例性系统图;
- [0031] 图12为根据本示教一实施例,用于人机对话管理的服务器的示例性过程的流程图;
- [0032] 图13示出了根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器装置配置的示例性系统图;
- [0033] 图14示出了根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器的示例性系统图;
- [0034] 图15为根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器的示例性过程的流程图;
- [0035] 图16示出了根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器装置配置的不同的示例性系统图;
- [0036] 图17示出了根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的装置的示例性系统图;
- [0037] 图18为根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的装置的示例性过程的流程图;
- [0038] 图19示出了根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器的示例性系统图;
- [0039] 图20为根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器的示例性过程的流程图;
- [0040] 图21示出了根据本示教一实施例,经由先发式生成的对话内容,用于人机对话管理的服务器装置配置的又一不同的示例性系统图;
- [0041] 图22示出了根据本示教不同的实施例,经由先发式生成的对话内容,用于人机对话管理的服务器的示例性系统图;
- [0042] 图23为根据本示教不同的实施例,经由先发式生成的对话内容,用于人机对话管理的服务器的示例性过程的流程图;
- [0043] 图24为示例性移动装置架构的示例性图,其可用于实现根据多种实施例实施本示教的特定系统;
- [0044] 图25为示例性计算装置架构的示例性图,其可用于实现根据多种实施例实施本示教的特定系统。

具体实施方式

[0045] 在下面的具体介绍中,通过举例的方式,给出了多种具体细节,以便提供对相关示教的详尽理解。然而,本领域技术人员应当明了,本示教可以在没有这些细节的情况下实现。在其他的实例中,公知的方法、过程、部件和/或电路以相对较高的层次介绍而没有细节,从而避免不必要地模糊本示教的实施形态。

[0046] 本示教目标在于解决传统的人机对话系统的不足之处,并提供使得更为有效且真

实的人机对话成为可能的系统和方法。本示教将人工智能并入具有代理装置的自动伴侣，其与来自用户交互引擎的支柱支持 (backbone support) 配合，故使自动伴侣能够基于连续监视的指示对话周边情况的多模态数据来执行对话，自适应地推定对话参与者的心态/情绪/意图，并基于动态改变的信息/推定/背景信息来适应性地调整会话策略。

[0047] 通过多个面向的自适应，包括但不限于会话的主题、用于进行会话的硬件/部件以及用于向人类会话者发送响应的表情/行为/姿态，根据本示教的自动伴侣能够对对话进行个性化。通过基于人类会话者对于对话的接受度有多高的观察来灵活地改变会话策略，自适应控制策略将使得会话更加真实且富有成效。根据本示教的对话系统可被配置为实现目标驱动的策略，包括动态配置被认为是最适合实现预期目的的硬件/软件部件。这样的最优化基于学习来进行，包括从以前的会话进行学习，以及通过在会话期间关于某些预期目标连续评估人类会话者的行为/反应，从正在进行的会话进行学习。为实现目标驱动策略而开发的路径可被确定为使人类会话者保持参与会话，即使在某些实例中，某些时刻的路径可能看起来偏离预期目标。

[0048] 具体而言，本示教公开了一种用户交互引擎，其向代理装置提供支柱支持，以促成与人类会话者的更为真实且参与度更高的对话。图1示出了根据本示教一实施例的网络环境100，网络环境用于促成在操作用户装置的用户和与用户交互引擎协同的代理装置之间的对话。在图1中，示例性的网络环境100包含：一个以上的用户装置110，例如用户装置110-a、110-b、110-c和110-d；一个以上的代理装置160，例如代理装置160-a、……、160-b；用户交互引擎140；以及，用户信息数据库130，其中的每一个可以经由网络120彼此通信。在某些实施例中，网络120可对应于单个网络或者不同网络的组合。例如，网络120可以是局域网 (“LAN”)、广域网 (“WAN”)、公用网络、专用网络、公共交换电话网 (“PSTN”)、互联网、内联网、蓝牙网络、无线网络、虚拟网络和/或其任何组合。在一实施例中，网络120也可包含多个网络接入点。例如，环境100可包含有线或无线接入点，例如但不限于基站或互联网交换点120-a、……、120-b。基站120-a和120-b可促成例如到/来自用户装置110和/或代理装置160的与联网框架100中的一个以上的其他部件在不同类型的网络上的通信。

[0049] 用户装置 (例如110-a) 可以出于不同的类型，以促成操作用户装置的用户连接到网络120并发送/接收信号。这样的用户装置110-a可对应于任何合适类型的电子/计算装置，包括但不限于桌面计算机 (110-d)、移动装置 (110-a)、包含在运输工具 (110-b) 中的装置、……、移动计算机 (110-c)，或为固定装置/计算机 (110-d)。移动装置可包括但不限于移动电话、智能电话、个人显示装置、个人数字助理 (“PDA”)、游戏机/装置、例如手表、Fibit、别针/胸针、耳机等的可穿戴装置。带有一装置的运输工具可包括汽车、卡车、摩托车、客船、轮船、火车或飞机。移动计算机可包括笔记本电脑、超级本、手持式装置等。固定装置/计算机可包括电视机、机顶盒、智能家居装置 (例如冰箱、微波炉、洗衣机或干衣机、电子助理等) 和/或智能配件 (例如灯泡、电灯开关、电子相框等)。

[0050] 代理装置 (例如160-a、……、160-b中的任何一个) 可对应于不同类型装置中的一个，该装置可与用户装置和/或用户交互引擎140通信。如下面更为详细地介绍的，每个代理装置可被看作自动伴侣装置，其在例如来自用户交互引擎140的支柱支持下与用户接口。这里介绍的代理装置可对应于机器人，其可以是游戏装置、玩具装置、指定的代理装置，例如旅游代理或天气代理，等等。这里公开的代理装置能够促成和/或帮助与操作用户装置的用

户的交互。通过这样,代理装置可被配置为机器人,经由来自应用服务器130的后端支持,其能够控制其某些部件,例如,做出某些身体移动(例如头部),表现出特定的面部表情(例如笑弯眼睛),或以特定的嗓音或音调(例如兴奋的音调)说话以表现特定的情绪。

[0051] 当用户装置(例如用户装置110-a)被连接到例如160-a的代理装置时(例如经由接触或无接触连接),在例如110-a的用户装置上运行的客户端可以与自动伴侣(代理装置或用户交互引擎或二者)通信,以便使得操作用户装置的用户与代理装置之间的交互式对话成为可能。客户端可以在某些任务中独立动作,或者,可以由代理装置或用户交互引擎140远程控制。例如,为了响应来自用户的问题,代理装置或用户交互引擎140可以控制在用户装置上运行的客户端,以便向用户呈现响应的语音。在会话过程中,代理装置可包含一个以上的输入机制(例如照相机、麦克风、触摸屏、按钮等),其允许代理装置捕获与用户或与会话相关联的本地环境有关的输入。这样的输入可帮助自动伴侣建立对会话周围的氛围(例如用户的移动,环境的声音)以及人类会话者心态(例如用户捡起球,这可能表示用户厌倦了)的理解,从而使得自动伴侣能够相应地反应,并以将保持用户感兴趣和参与的方式进行会话。

[0052] 在所示的实施例中,用户交互引擎140可以是后端服务器,其可以是集中式的或分布式的。它被连接到代理装置和/或用户装置。它可被配置为向代理装置160提供支柱支持,并引导代理装置以个性化和定制化的方式执行会话。在某些实施例中,用户交互引擎140可接收来自所连接装置(代理装置或用户装置)的信息,分析这些信息,并通过向代理装置和/或用户装置发送指示来控制会话的流动。在某些实施例中,用户交互引擎140也可直接与用户装置通信,例如提供动态数据(例如,用于在用户装置上运行的客户端的控制信号,以呈现特定的响应)。

[0053] 一般而言,用户交互引擎140可控制用户与代理装置之间的会话的流动和状态。各个会话的流动可以基于与会话相关联的不同类型的信息受到控制,例如,关于参与会话的用户的信息(例如来自用户信息数据库130)、会话历史、会话周边情况信息、和/或实时用户反馈。在某些实施例中,用户交互引擎140可以被配置为获得多种传感器输入(例如但不限于音频输入、图像输入、触觉输入和/或背景输入),处理这些输入,阐述对人类会话者的理解,相应地基于这种理解生成响应,并控制代理装置和/或用户装置,以便基于响应进行会话。作为一说明性实例,用户交互引擎140可接收表征来自操作用户装置的用户的话语的音频数据,并生成响应(例如文本),于是,该响应可以以计算机生成言语的形式作为对用户的响应被传送给用户。作为另一实例,用户交互引擎140也可响应于该言语生成一个以上的指示,其控制代理装置执行特定的动作或动作组。

[0054] 如所示的,在人机对话过程中,用户作为对话中的人类会话者可通过网络120与代理装置或用户交互引擎140通信。这样的通信可涉及多种模态的数据,例如音频、视频、文本等。经由用户装置,用户可发送数据(例如请求、表征用户话语的音频信号、或用户周边的场景的视频)和/或接收数据(例如来自代理装置的文本或音频响应)。在某些实施例中,多种模态的用户数据在被代理装置或用户交互引擎140接收到时可被分析,以理解人类用户的语音或姿态,故使用户的情绪或意图可被推定并用于确定对用户的响应。

[0055] 图2A示出了根据本示教一实施例,在对话过程中,在用户装置110-a、代理装置160-a和用户交互引擎140之间的特定连接。可以看到,任何两方之间的连接全部可以是双

向的,如这里所讨论的。代理装置160-a可以与用户经由用户装置110-a接口,以便以双向通信执行对话。一方面,代理装置160-a可以由用户交互引擎140控制,以便向操作用户装置110-a的用户说出响应。另一方面,来自用户现场的输入,包括例如用户的言语或动作以及关于用户周边情况的信息,经由连接被提供给代理装置。代理装置160-a可被配置为处理这种输入,并动态调节其对用户的响应。例如,代理装置可被用户交互引擎140指示为在用户装置上呈现树。知道用户周围环境(基于来自用户装置的视觉信息)显示出绿色的树和草地,代理装置可以把将被呈现的树定制为繁茂的绿树。如果来自用户现场的场景显示正值冬季,代理装置可进行控制,以使用对于没有叶子的树的参数在用户装置上呈现树。作为另一实例,如果代理装置被指示为在用户装置上呈现鸭子,代理装置可从用户信息数据库130取得关于颜色偏好的信息,并在发送用于向用户装置呈现的指示之前生成以用户偏好颜色对鸭子进行定制的参数。

[0056] 在某些实施例中,来自用户现场的这些输入以及其处理结果也可被发送到用户交互引擎140,用于促成用户交互引擎140更好地理解与对话相关联的具体情况,使得用户交互引擎140可确定对话的状态、用户的情绪/心态,并生成基于对话的具体情况以及对话的预期目的(例如用于教小孩英语词汇)的响应。例如,如果从用户装置接收的信息指示用户看上去厌倦并变得不耐烦,用户交互引擎140可判断为将对话的状态改变为用户感兴趣的话题(例如基于来自用户信息数据库130的信息),以便继续使用户参与会话。

[0057] 在某些实施例中,在用户装置上运行的客户端可以被配置为,能够处理从用户现场获取的不同模态的原始输入,并将处理的信息(例如原始输入的相关特征)发送到代理装置或用户交互引擎,以便进一步处理。这将减少通过网络传输的数据的量,并增强通信效率。类似地,在某些实施例中,代理装置也可被配置为能够处理来自用户装置的信息,并提取用于例如定制化目的的有用信息。尽管用户交互引擎140可控制对话的状态和流动控制,使得用户交互引擎140轻量级更好地改善了用户交互引擎140的规模(scale)。

[0058] 图2B示出了与图2A所示相同的设置,且具有用户装置110-a的附加细节。如图所示,在用户和代理310之间的对话过程中,用户装置110-a可连续收集与用户及其周边情况有关的多模态传感器数据,其可被分析,以检测与对话有关的任何信息,并被用于以自适应的方式智能地控制对话。这可进一步增强用户体验或参与度。图2B示出了示例性的传感器,例如视频传感器230、音频传感器240、……、或触觉传感器250。用户装置也可发送文本数据,作为多模态传感器数据的一部分。这些传感器共同提供对话周边的背景信息,并可被用于由用户交互引擎140理解情况,以便管理对话。在某些实施例中,多模态传感器数据可首先在用户装置上被处理,不同模态的重要特征可被提取并发送到用户交互引擎140,故对话能以对背景的理解受到控制。在某些实施例中,原始多模态传感器数据可被直接发送到用户交互引擎140,用于处理。

[0059] 由图2A-2B可见,代理装置可对应于具有不同部分(包括其头部210及其身体220)的机器人。尽管图2A-2B所示的代理装置显示为人形机器人,其可以以其他的形式被构建,例如鸭子、熊、兔子等。图3A示出了根据本示教一实施例,具有示例性类型的代理身体的代理装置的示例性结构。如所示的,代理装置可包括头和身体,头附着到身体。在某些实施例中,代理装置的头可具有附加的部分,例如面部、鼻子和嘴,它们中的一些可被控制为例如做出移动或表情。在某些实施例中,代理装置上的面部可对应于显示屏,显示屏上可呈现面

部且该面部可以是人类的或动物的。这样显示的面部也可受到控制以表达情绪。

[0060] 代理装置的身体部分也可对应于不同形态,例如鸭子、熊、兔子等。代理装置的身体可以是固定的、可动的或半可动的。具有固定身体的代理装置可对应于能放置在例如桌子的表面上的装置,从而与坐在桌边的人类用户进行面对面的会话。具有可动身体的代理装置可以对应于能够在例如桌面或地板的表面上到处移动的装置。这样的可动的身体可包括能在运动学上被控制以进行实体移动的部件。例如,代理身体可包括脚,其可被控制为在需要时在空间中移动。在某些实施例中,代理装置的身体可以是半可动的,即,有些部分可动,有些不可动。例如,具有鸭子外表的代理装置的身体上的尾巴可以是可动的,但鸭子不能在空间中移动。熊形身体代理装置也可具有可动的手臂,但熊只能坐在表面上。

[0061] 图3B示出了根据本示教一实施例的示例性代理装置或自动伴侣160-a。自动伴侣160-a是使用语音和/或面部表情或身体姿态与人交互的装置。例如,自动伴侣160-a对应于具有不同部分的、电子操控(animatronic)的外围装置,包括头部310、眼部(照相机)320、具有激光器325和麦克风330的嘴部、扬声器340、具有伺服机构350的颈部、可用于无接触存在性检测的一个以上的磁体或其它部件360、以及对应于充电底座370的身体部分。在运行中,自动伴侣160-a可以连接到用户装置,其可包括经由网络连接的移动多功能装置(110-a)。一旦连接,自动伴侣160-a和用户装置经由例如语音、动作、姿态和/或经由用激光点指器的点指(pointing)而彼此交互。

[0062] 自动伴侣160-a的其他示例性功能可以包括响应于用户响应的反应式表达,例如,经由显示在例如作为自动伴侣面部的一部分的屏幕上的交互式视频卡通角色(例如化身/avatar)。自动伴侣可使用照相机(320)来观察用户的存在性、面部表情、注视方向、周边情况等。电子操控实施例可以通过点指其包含照相机(320)的头(310)来“看”,使用其麦克风(340)来“听”,通过引导能够经由伺服机构(350)移动的头(310)的方向来“点指”。在某些实施例中,代理装置的头也可由例如用户交互系统140或由用户装置(110-a)的客户端经由激光器(325)来远程控制。如图3B所示示例性自动伴侣160-a也可被控制为经由扬声器(330)来“说话”。

[0063] 图4A示出了根据本示教多种实施例,用于自动伴侣的整体系统的示例性高层次系统图。在此所示实施例中,整体系统可包含驻留在用户装置、代理装置和用户交互引擎140中的部件/功能模块。这里所述的整体系统包含多个处理层和分级结构,它们一起执行智能方式的人机交互。在所示的实施例中有5个层,包括用于前端应用以及前端多模态数据处理的层1、用于对话设定的描绘的层2、对话管理模块所驻留的层3、用于不同参与者(人、代理、装置等)的推定心态的层4、用于所谓效用(utility)的层5。不同的层可对应于不同等级的处理,从层1上的原始数据采集和处理到层5上的改变对话参加者的效用的处理。

[0064] 术语“效用”由此定义为基于与对话历史相关联地检测到的状态识别的参与者的偏好。效用可以与对话中的参与者相关联,无论该参与者是人、自动伴侣还是其他的智能装置。用于特定参与者的效用可表征世界的不同状态,无论是实物的、虚拟的或者甚至是精神的。例如,状态可以表征为对话所沿在世界的复杂地图中穿行的特定路径。在不同的实例中,当前状态基于多个参与者之间的交互而演化为下一状态。状态也可以是与参与者有关的,即,当不同参与者参加交互时,由这种交互带来的状态可能发生变化。与参与者相关联的效用可被组织为偏好的分级结构,且这样的偏好分级结构可基于在会话过程中做出的参

与者选择以及表露出的喜好而随时间演化。这样的偏好(其可被表征为从不同选项中做出的有顺序的选择序列)被称为效用。本示教公开了这样的方法和系统:通过该方法和系统,智能自动伴侣能够通过与人会话者的对话来学习用户的效用。

[0065] 在支持自动伴侣的整体系统中,层1中的前端应用以及前端多模态数据处理可驻留在用户装置和/或代理装置中。例如,照相机、麦克风、键盘、显示器、呈现器、扬声器、聊天泡泡、用户接口元件可以是用户装置的部件或功能模块。例如,可能有在用户装置上运行的应用或客户端,其可包括图4A所示外部应用接口(API)之前的功能。在某些实施例中,超出外部API的功能可以被考虑为后端系统,或驻留在用户交互引擎140中。在用户装置上运行的应用可取得来自用户装置的电路或传感器的多模态数据(音频、图像、视频、文字),对多模态数据进行处理,以生成表征原始多模态数据的特征的文字或其他类型的信号(例如检测到的用户面孔等对象、语音理解结果),并发送到系统的层2。

[0066] 在层1中,多模态数据可以经由例如为照相机的传感器、麦克风、键盘、显示器、扬声器、聊天泡泡、呈现器或其他用户接口元件来获取。可对这样的多模态数据进行分析,以推定或推断能用于推断更高层次的特性(例如表情、角色(character)、姿态、情绪、动作、注意力、意图等)的多种特征。这样的更高层次的特性可由层2上的处理单元获得,接着,经由图4A所示的内部API,例如,由更高层次的部件用于在更高的概念层次上智能地推断或推定与对话有关的附加信息。例如,在层2上获得的所推定的对话参加者的情绪、注意力或其它特性可用于推定参加者的心态。在某些实施例中,这种心态也可在层4上基于附加的信息来推定,例如,记录的周边环境或这种周边环境中的其他附加信息,例如声音。

[0067] 推定的参与者的心态,无论是与人还是自动伴侣(机器)有关,可被层3的对话管理所依赖,以便确定,例如,如何进行与人类会话者的会话。每个对话如何逐步发展常常表征人类用户的偏好。这样的偏好可以动态地在对话过程中在效用(层5)上被捕获。如图4A所示,层5上的效用表征演化的状态,其指示参与者的演化的偏好,它们也可由层3上的对话管理用于决定进行交互的合适或智能的方式。

[0068] 不同层之间的信息共享可以经由API实现。在图4A中所示的某些实施例中,层1和其他层之间的信息分享经由外部API,而层2-5之间的信息分享经由内部API。能够明了,这仅仅是一种设计上的选择,其他的实现方式也可以实现这里给出的示教。在某些实施例中,通过内部API,多种层(2-5)可访问由其他层产生或存储的信息,以支持处理。这样的信息可包括将被应用到对话的通用配置(例如代理装置的角色是化身、优选的嗓音或将为对话产生的虚拟环境,等等)、对话的当前状态、当前对话历史、已知的用户偏好、推定的用户意图/情绪/心态等等。在某些实施例中,可从外部数据库访问能经由内部API分享的某些信息。例如,可从例如开源数据库访问与代理装置(例如鸭子)的希望的角色有关的特定配置,其提供参数(例如,视觉上呈现鸭子的参数,和/或呈现来自鸭子的语音需要的参数)。

[0069] 图4B示出了根据本示教的实施例的正在进行的对话的对话树的一部分,其具有基于自动伴侣和用户之间的交互取得的路径。在此所示实例中,(自动伴侣的)层3中的对话管理可预测多种路径,与用户的对话(或一般地,交互)可以以该路径进行。在此实例中,各个节点可代表对话的当前状态的点,且节点的各个分支可代表来自用户的可能的响应。如此实例所示,在节点1上,自动伴侣可面临三种分立的路径,其可取决于从用户检测到的响应来取得。如果用户用肯定性的响应来回应,对话树400可从节点1进行到节点2。在节点2上,

响应于来自用户的肯定性响应,可以为自动伴侣生成响应,于是,响应可被呈现给用户,其可包括音频、视觉、文本、触觉或其任何组合。

[0070] 在节点1上,如果用户负面地响应,用于此阶段的路径是从节点1到节点10。如果用户在节点1上用“一般”响应来回应(例如,不是负面的,但也不是正面的),对话树400可进行到节点3,在节点3上,来自自动伴侣的响应可被呈现,可存在来自用户的三种分立的可能响应,“无响应”、“正面响应”、“负面响应”,分别对应于节点5、6、7。取决于关于在节点3上呈现的自动伴侣响应的、用户的实际响应,层3上的对话管理于是可相应地延续对话。例如,如果用户在节点3上用正面响应来回应,自动伴侣移动到在节点6上回应用户。类似地,取决于用户对自动伴侣在节点6上的响应的反应,用户可进一步用正确的回答来响应。在这种情况下,对话状态从节点6移动到节点8,等等。在此所示的实例中,这一阶段期间的对话状态从节点1移动到节点3、到节点6、并到节点8。节点1、3、6、8的遍历构成与自动伴侣和用户之间的底层会话一致的路径。如图4B所示,代表该对话的路径由连接节点1、3、6、8的实线表示,而在该对话过程中跳过的路径用虚线表示。

[0071] 图4C示出了根据本示教一实施例,由自动伴侣执行的示例性人-代理装置交互和示例性处理。如图4C所示,可以进行不同层上的操作,且它们一起以协调的方式促成智能对话。在所示的实例中,代理装置可首先在402处询问用户“你今天好吗?”以发起对话。响应于402处的言语,用户可以在404处用言语“好”来回应。为了管理对话,自动伴侣可以在对话过程中致动不同的传感器,以便做出对用户以及周边环境的观察。例如,代理装置可获取关于用户所处周边环境的多模态数据。这样的多模态数据可包括音频、视觉或文本数据。例如,视觉数据可捕获用户的面部表情。视觉数据也可揭示会话场景周围的背景信息。例如,场景的图像可揭示存在篮球、桌子和椅子,这提供了关于环境的信息,并可在对话管理中被利用,以便增强用户的参与度。音频数据可不仅捕获用户的语音响应,还捕获其他的周边信息,例如响应的音调、用户说出回应的方式或者用户的口音。

[0072] 基于所获取的多模态数据,分析可以由自动伴侣(例如由前端用户装置或由后端用户交互引擎140)进行,以评估用户的态度、情绪、心态和效用。例如,基于视觉数据分析,自动伴侣可检测到用户表现出悲哀、无笑容、用户语音缓慢且嗓音低沉。对对话中的用户状态的描绘可在层2上基于在层1上获取的多模态数据来进行。基于这样检测的观察,自动伴侣可以推断(在406上)用户对当前话题不是那么感兴趣且参与度不高。例如,对用户的情绪或精神状态的这样的推断可以在层4上基于对与用户相关联的多模态数据的描绘来进行。

[0073] 为了响应用户的当前状态(参与度不高),自动伴侣可判断为使用户振奋,以便使用户更好地参与。在此所示的实例中,自动伴侣可以通过在408处向用户说出问题“你想玩游戏吗?”来利用会话环境中可用的东西。这样的问题可以通过将文本转换为语音(例如,使用为用户个性化的定制嗓音)以音频形式作为语音给出。在这种情况下,用户可通过在410处说“好”来回应。基于连续获取的与用户有关的多模态数据,例如,经由层2的处理,可能观察到响应于玩游戏的邀请,用户的眼睛看上去左顾右盼,特别是,用户的眼睛可能注视篮球所在的地方。同时,自动伴侣也可观察到,一旦听到玩游戏的建议,用户的面部表情从“悲伤”变为“微笑”。基于这样观察到的用户的特性,自动伴侣可在412处推断为用户对篮球感兴趣。

[0074] 根据所获取的新信息以及基于其的推断,自动伴侣可以决定利用环境中可用的篮

球来使用户在对话中的参与度更高,同时仍又实现对用户的教育目的。在这种情况下,层3中的对话管理可对会话进行适应以谈论游戏,并利用用户注视房间里的篮球这一观察,使得对话对用户来说更加有趣,同时仍实现例如对用户进行教育的目标。在一示例性实施例中,自动伴侣生成响应,建议用户玩拼写游戏(在414处),并让用户拼写单词“篮球”。

[0075] 在给定自动伴侣的根据对用户和环境的观察的自适应对话策略的情况下,用户可做出响应,提供单词“篮球”的拼写(在416处)。可连续就用户在回答拼写问题时有多热情进行观察。基于例如在用户回答拼写问题时获取的多模态数据所判断,如果用户看起来以更为欢快的态度迅速响应,自动伴侣可以在418处推断为用户现在参与度更高。为了进一步鼓励用户积极参加对话,自动伴侣于是可生成正面响应“做得好!”,并指示将此响应用欢快、鼓励、积极的嗓音传送给用户。

[0076] 图5示出了根据本示教的多种实施例,在以对话管理器510为中心的自动对话伴侣的不同处理层之间的示例性通信。图中的对话管理器510对应于层3中的对话管理的功能部件。对话管理器是自动伴侣的重要部分,且其管理对话。按照传统,对话管理器将用户的言语取作输入,并判断如何对用户做出响应。这在不考虑用户偏好、用户的心态/情绪/意图或对话的周边环境的情况下做出,也就是说,不为相关世界的不同的可用状态授予任何权重。缺少对周边世界的了解常常限制了人类用户和智能代理之间会话的参与度或感知的真实性。

[0077] 在本示教的某些实施例中,充分运用与正在进行的对话有关的会话参与者的效用,以允许进行更为个性化、灵活且参与度更高的对话。这促进了智能代理扮演不同的角色,以便在不同的任务中更为有效,例如安排约会、预订旅行、订购设备和补给品、在线研究多种话题。当智能代理认识到用户的动态心态、情绪、意图和/或效用时,这使得代理能以更有目标且有效的方式使人类会话者参与对话。例如,当教育代理教孩子时,孩子的偏好(例如他喜欢的颜色)、观察到的情绪(例如,有时候孩子不想继续课程)、意图(例如,孩子将手伸向地板上的球,而不是专注于课程)都可允许教育代理灵活地将关注的主题调整到玩具,并可能调整继续与孩子会话的方式,以便给孩子休息时间,从而实现对孩子进行教育的整体目标。

[0078] 作为另一实例,本示教可用于,通过问在给定从用户实时观察到的东西的情况下更为适合的问题,增强用户服务代理的服务,并因此实现改善的用户体验。这根植于如这里所公开的本示教的本质方面,通过开发学习和适应参加对话的参与者的偏好或心态的方法和手段,使得对话能够以参与度更高的方式进行。

[0079] 对话管理器(DM) 510是自动伴侣的核心部件。如图5所示,DM 510(层3)取得来自不同层的输入,包括来自层2的输入以及来自更高的抽象层的输入,例如,用于推定涉入对话的参与者的心态的层4,以及基于对话以及评估的其性能来学习效用/偏好的层5。如所示的,在层1上,从不同模态的传感器获取多模态信息,其被处理,以便获得例如对数据进行描绘的特征。这可包括视觉、音学和文本模态的信号处理。

[0080] 这样的多模态信息可以在对话过程中由布置在用户装置(例如110-a)上的传感器获取。所获取的多模态信息可以与操作用户装置110-a的用户和/或对话场景周边情况有关。在某些实施例中,多模态信息也可在对话过程中由代理装置(例如160-a)获取。在某些实施例中,用户装置和代理装置二者上的传感器可获取相关信息。在某些实施例中,所获取

的多模态信息在层1上受到处理,如图5所示,其可包括用户装置和代理装置二者。取决于情况和配置,各个装置上的层1处理可以不同。例如,如果用户装置110-a用于获取对话的周边信息,包括关于用户以及用户周边情况的信息,原始输入数据(例如文本、视觉或音频)可在用户装置上被处理,接着,处理得到的特征可被发送到层2,用于进一步的分析(在更高的抽象层上)。如果关于用户和对话环境的某些多模态信息由代理装置获取,这样获取的原始数据的处理也可由代理装置(图5未示出)处理,于是,从这样的原始数据中提取的特征可从代理装置被发送到层2(其可以位于用户交互引擎140中)。

[0081] 层1也处理从自动对话伴侣到用户的响应的信息呈现。在某些实施例中,呈现由代理装置(例如160-a)执行,这种呈现的实例包括语音、表情(其可以是面部的)或执行的身体动作。例如,代理装置可将来自用户交互引擎140接收的文本串(作为对用户的响应)呈现为语音,使得代理装置可以向用户说出响应。在某些实施例中,文本串可被发送到代理装置,并具有附加的呈现指示,例如音量、音调、音高等,其可用于以特定方式将文本串转换为与内容的言语对应的声波。在某些实施例中,将要传送给用户的响应也可包括动画(animation),例如,用将要经由例如面部表情或身体动作(例如举起一只手臂等)传送的态度说出响应。在某些实施例中,代理可以被实现为用户装置上的应用。在这种情况下,来自自动对话伴侣的相应的呈现经由用户装置(例如110-a(图5未示出))实现。

[0082] 多模态数据的处理得到的特征可在层2上进一步处理,以实现语言理解和/或多模态数据理解,包括视觉、文字及其任何组合。某些这样的理解可能针对单一模态,例如语音理解,有些可以针对基于集成的信息对参与对话的用户的周边情况的理解。这样的理解可以是实物的(例如,识别场景中的特定对象)、认知上的(例如识别出用户说了什么,或某个明显的声音,等等)或精神上的(例如特定的情绪,例如基于语音的音调、面部表情或用户姿态推定出的用户的压力)。

[0083] 层2上生成的多模态数据理解可由DM 510用于判断如何响应。为了增强参与度和用户体验,DM 510也可基于来自层4的推定的用户心态和代理心态以及来自层5的参与对话的用户的效用来确定响应。涉入对话的参与者的心态可基于来自层2的信息(例如推定的用户情绪)以及对话的进展推定。在某些实施例中,用户和代理的心态可以在对话过程中动态推定,这样推定的心态于是可用于学习(与其他数据一起)用户的效用。所学习的效用代表用户在不同对话情境中的偏好,并基于历史对话及其结果而被推定。

[0084] 在特定话题的各个对话中,对话管理器510将其对对话的控制基于相关的对话树,对话树可能与、或者可能不与话题相关联(例如,可引入闲聊,以增强参与度)。为了生成对话中对用户的响应,对话管理器510也可考虑附加的信息,例如用户的状态、对话场景的周边情况、用户的情绪、用户和代理的推定的心态、以及已知的用户偏好(效用)。

[0085] DM 510的输出对应于相应地确定的对用户的响应。为了将响应传送给用户,DM 510也可以阐述传送响应的方式。响应被传送的形式可以基于来自多个源的信息来确定,例如,用户的情绪(例如,如果用户是不快乐的孩子,响应可以以温柔的嗓音呈现),用户的效用(例如,用户可能偏好与其父母类似的某种口音),或用户所处的周边环境(例如,嘈杂的地方,故响应需要以高音量传送)。DM 510可将所确定的响应与这些传送参数一起输出。

[0086] 在某些实施例中,这样确定的响应的传送通过根据与响应相关联的多种参数来生成各个响应的可传送形式来实现。在一般情况下,响应以某些自然语言的语音的形式传送。

响应也可以与特定非语言表达耦合的语音传送,非语言表达作为所传送响应的一部分,例如为点头、摇头、眨眼或耸肩。可能有听觉上的但是非语言的其他形式的可传送响应模式,例如口哨。

[0087] 为了传送响应,可传送的响应形式可以经由例如语言响应生成和/或行为响应生成来产生,如图5所示。出于其所确定的可传送形式的这种响应于是可由呈现器用于实际以其预期形式呈现响应。对于自然语言的可传送形式,响应的文本可用于,根据传送参数(例如音量、口音、风格等),经由例如文本到语音技术来合成语音信号。对于将要非语言形式(例如特定的表情)传送的任何响应或其部分,预期的非语言表达可以被翻译成为(例如经由动画)能用于控制代理装置(自动伴侣的有形体现)的特定部分的控制信号,从而执行特定的机械运动,以便传送响应的非语言表达,例如点头、耸肩或吹口哨。在某些实施例中,为了传送响应,特定的软件部件可被调用,以便呈现代理装置的不同面部表情。响应的这种演绎也可由代理同时进行(例如,用开玩笑的嗓音说出响应,并在代理的脸上浮现大大的笑容)。

[0088] 图6示出了根据本示教的多种实施例,用于基于人工智能的教育型伴侣的示例性高层次系统图。在此所示实施例中,存在五个处理层次,即装置层、处理层、论证层、教学或施教层以及教师层。装置层包含传感器(例如麦克风和照相机),或者媒体传送装置(例如伺服机构),其用于移动例如扬声器或机器人的身体部分,从而传送对话内容。处理层包含多种处理部件,其目的在于处理不同类型的信号,包括输入和输出信号。

[0089] 在输入侧,处理层可包括语音处理模块,用于基于从音频传感器(麦克风)获得的音频信号来进行例如语音识别,以便理解在说什么,从而确定如何响应。音频信号也可被识别,以便生成用于进一步分析的文本信息。来自音频传感器的音频信号也可被情绪识别处理模块使用。情绪识别模块可以被设计为,基于来自照相机的视觉信息和同步的音频信息,识别参与者的多种情绪。例如,快乐的情绪常常可伴有笑脸和特定的听觉线索。作为情绪指示的一部分,经由语音识别获得的文本信息也可由情绪识别模块用于推定所涉及的情绪。

[0090] 在处理层的输出侧,当特定的响应策略被确定时,这样的策略可被翻译成为将由自动伴侣做的具体动作,以便对另一参与者做出响应。这样的动作可以通过传送某种音频响应或经由特定姿态表达特定情绪或态度来进行。当响应以音频被传送时,具有需要被说出的词语的文本由文本到语音模块进行处理,以便产生音频信号,于是,这样的音频信号被发送到扬声器,以便呈现作为响应的语音。在某些实施例中,基于文本生成的语音可以根据其他的参数进行,例如,可用于以特定音调或嗓音对语音生成进行控制的参数。如果响应将作为实体动作被传送,例如,在自动伴侣上实现的身体移动,则将要采用的动作也可以是将用于生成这样的身体移动的指示。例如,处理层可包含根据某种指示(符号)来移动自动伴侣的头部(例如点头、摇头或头部的其他运动)的模块。为了遵从移动头部的指示,基于该指示,用于移动头部的模块可生成电信号,并发送到伺服机构,以便实体控制头部运动。

[0091] 第三层是论证层,其用于基于分析的传感器数据来执行高层次的论证。来自语音识别的文本或推定的情绪(或其他描绘)可被发送到推断程序,该程序可用于,基于从第二层接收的信息,推断多种高层次概念,例如意图、心态、偏好。推断的高层次概念于是可由基于效用的计划模块使用,在给定在教学层上定义的施教计划和当前用户状态的情况下,该模块设计出在对话中做出响应的计划。计划的响应于是可被翻译成将被执行以便传送计划

的响应的动作。该动作于是被动作生成器进一步处理,以便具体指向不同的媒体平台,从而实现智能响应。

[0092] 教学层和教师层都涉及所公开的教育型应用。教师层包含关于设计用于不同主题的课程表的活动。基于设计的课程表,教学层包含课程表调度器,其基于所设计的课程表来调度课程,基于课程表调度,问题设置模块可安排将基于特定的课程表调度而被提供的特定的问题设置。这样的问题设置可由论证层的模块用于辅助推断用户的反应,于是,基于效用和推断的心理状态来相应地规划响应。

[0093] 在用户机器对话系统中,图5中的对话管理器(例如510)扮演中心角色。它从具有(对用户言语、面部表情、周边情况等)观察的用户装置或代理装置接收输入,并确定在给定的对话当前状态和对话目的的情况下合适的响应。例如,如果特定对话的目的是向用户教授三角测量概念,由对话管理器510设计的响应不仅仅基于来自用户的前一通信,还基于确保用户学习该概念的目的来确定。传统而言,对话系统通过探索与对话的预期意图以及当前会话状态相关联的对话树来驱动与人类用户的通信。这在图7中示出,其中,用户700与装置710接口,以实现会话。在会话过程中,用户说出发送到装置710的某些语音,基于用户的言语,装置向服务器720发送请求,服务器720于是向装置提供响应(基于对话树750获得的),装置于是将响应呈现给用户。由于有限的计算力和内存,生成对用户的响应需要的大多数计算在服务器720上进行。

[0094] 在运行中,从装置710的角度来看,其获取来自与对话有关的用户700的言语,向服务器720发送具有所获取的用户信息的请求,然后,接收由服务器720确定的响应,并在装置710上向用户700呈现响应。在服务器侧,其包含控制器730和对话管理器740,控制器730可被配置为与装置110-a接口,对话管理器740基于合适的对话树750来驱动与用户的对话。对话树750可以基于当前对话从多个对话树中选择。例如,如果当前对话是为了预订航班,为对话管理器740选择以驱动会话的对话树可以被特别地构建,用于该预期的意图。

[0095] 当接收到用户的信息时,控制器730可分析所接收的用户信息(例如用户说了什么),以得出对话的当前状态。于是,它可以调用对话管理器740,以便基于对话的当前状态在对话树750中搜索,从而识别对用户的合适的响应。这样识别的响应于是从对话管理器740被发送到控制器730,控制器730于是可转发给装置710。这样的对话过程需要在装置710和服务器720之间往复的通信流量,花费时间和带宽。另外,在大多数情况下,服务器720可以是对多个用户装置和/或代理装置(如果它们与用户装置分立的话)的支柱支持。另外,每个用户装置可以在需要使用不同的对话树驱动的不同的对话中。给定这一点,当存在大量依赖于服务器720驱动其相应的对话的装置时,按照传统,服务器720需要为全部用户装置/代理装置做出决策,不断处理来自不同对话的信息以及搜索不同对话树以得到对不同对话的响应可能变得耗时,影响服务器扩大规模的能力。

[0096] 本示教公开了一种替代性配置,通过在装置(用户装置,或代理装置)上智能地缓存完整对话树750的相关片段,使得进行人机对话的分布式方法成为可能。这里,“相关性”可以基于不同时间帧上与各个对话有关的相应的时间和空间本地性而动态地定义。为了促成对缓存在装置上的本地对话树的利用,缓存的对话树可以与对话管理器的本地版本结合提供,对话管理器具有合适的一组功能,使得本地对话管理器能够在缓存的对话树上运行。关于将要缓存在装置上的各个本地对话树,与母对话树(由之开拓出本地对话树的整体对

话树)相关联的功能的子集可被动态地确定和提供。例如,使得本地对话管理器能够解析缓存的本地对话树并遍历本地对话树的功能。在某些实施例中,将要配置在装置上的本地对话管理器可以基于不同的判据最优化,例如本地装置类型、具体的本地对话树、对话的性质、从对话场景做出的观察、和/或特定的用户偏好。

[0097] 图8示出了根据本示教一实施例,针对分布式对话管理的示例性框架。如图所示,框架包含与用户800以及服务器840接口的装置810,装置和服务器以分布式的方式一起驱动与用户800的对话。取决于实际对话配置,装置710可以是被用户700操作的用户装置(例如110-a),或作为自动对话伴侣的一部分的代理装置(例如160-a),或它们的组合。装置用于与用户700或用户装置110-a接口,以便执行与用户的对话。装置和服务器一起构成自动对话伴侣,并以有效且高效的方式管理对话。在某些实施例中,服务器被连接到多个装置,以使用作这些装置的后端,驱动就不同话题与不同用户的不同对话。

[0098] 除了其他部件以外,装置810包含:关于当前对话状态为装置设计的本地对话管理器820;本地对话树,它是整体对话树750的一部分,并且基于对话的当前状态和发展为装置开拓出。在某些实施例中,缓存在装置810上的这样的本地对话树830基于这样的评估来确定和配置:给定对话的当前状态和/或用户的已知偏好,装置810可能在不远的将来需要这一部分的对话树,以驱动与用户800的对话。

[0099] 在对话树和对话管理器的本地版本被配置在装置810上的情况下,每当可行时,对话被本地对话管理器基于缓存的本地对话树830管理。正是通过这种方式,由装置810和服务器840之间的频繁通信导致的流量和带宽消耗得到降低。在运行中,如本地对话管理器820所判断,如果用户800的言语的内容在缓存的对话树830内,装置810于是从缓存的对话树830向用户提供响应,而不必与服务器通信。因此,响应于用户800的速度也可改善。

[0100] 如果存在缓存未命中(cache miss),即,给定用户的输入,本地对话管理器820没有在缓存的对话树830中找到响应,装置810,它向服务器840发送具有与当前对话状态有关的信息的请求,然后,接收基于完整对话树750由服务器840中的对话管理器860识别的响应。由于存在未命中,采用来自服务器840的响应,装置810也从服务器接收更新的本地对话树(DT)和本地对话管理器(DM),使得DT和DM的先前的本地版本能用基于对话的发展自适应地生成的更新版本得到更新。

[0101] 在此所示的实施例中,服务器840包含控制器850、对话管理器860、本地DM/DT生成器870(本地DM指的是本地对话管理器820,本地DT指的是本地对话树830)。对话管理器860的功能角色与传统系统中的相同,根据被选择为驱动对话的对话树750,基于来自用户的输入确定响应。在运行中,在接收到来自装置810的对于响应的请求(带有用户的信息)时,控制器850不仅调用对话管理器860生成所请求的响应,也调用本地DM/DT生成器870,以便基于所接收的用户信息,关于对话树850和由对话管理器860推定的当前对话状态,为请求的装置810生成更新的本地对话树830(DT)和本地对话管理器820(DM)。这样生成的本地DT/DM于是被发送到装置810,以更新先前在其中缓存的版本。

[0102] 图9示出了根据本示教一实施例,装置810的示例性高层次系统图。如这里所讨论的,装置810可以是用户装置、代理装置或其组合。图9示出了用于实现本示教的相关功能部件,且每个这样的部件可驻留在用户装置或是代理装置上,它们以协调的方式一起工作,以实现与本示教的装置810有关的功能的方面。在所示的实施例中,装置810包含传感器数据

分析器910、周边信息理解单元920、本地对话管理器820、装置/服务器协调器930、响应呈现单元940、本地对话管理更新器950以及本地对话树更新器960。图10为根据本示教一实施例的装置810的示例性过程的流程图。在运行中,传感器数据分析器910在图10的1005处接收来自用户800的传感器数据。这样接收的传感器数据可以是多模态的,包括例如表征用户语音的听觉数据,和/或对应于用户视觉表征(例如面部表情)和/或对话场景周边情况的视觉数据。

[0103] 在接收到传感器数据时,传感器数据分析器910在1010处分析所接收的数据,并从传感器数据提取相关特征,并发送到周边信息理解单元920。例如,基于从音频数据提取的听觉特征,周边信息理解单元920可确定与来自用户800的言语对应的文本。在某些实施例中,从视觉数据提取的特征也可用于理解对话中正在发生什么。例如,用户800的嘴唇移动可被跟踪,嘴唇形状的特征可被提取并用于在除音频数据之外理解用户800说出的语音的文本。周边信息理解单元920也可分析传感器数据的特征,以实现对话的其它方面的理解。例如,来自用户的语音的音调、用户的面部表情、对话场景中的物体等等也可被识别,并由本地对话管理器820用于确定响应。

[0104] 在得出对话的当前状态的理解时(例如,用户说了什么,或以何种方式),周边信息理解单元920可依赖于多种模型或传感器数据理解模型925,其可包括,例如,用于识别对话场景中的声音的听觉模型、用于识别说了什么的自然语言理解(NLU)模型、用于检测例如用户面部和场景中的其他对象(树、桌子、椅子等)的对象检测模型、用于检测面部表情或用于检测与人的不同情绪状态相关联的语音音调的情绪检测模型,等等。对话当前状态的这种理解于是可从周边信息理解单元920发送到本地对话管理器820,以便使其能够基于本地对话树830确定对用户的响应。

[0105] 在接收到当前对话状态的情况下,调用本地对话管理器(DM)820,以便在1015处在本地对话树(DT)830中搜索响应。如这里所讨论的,当前对话状态可包含一个以上类型的信息,例如用户当前言语、推定的用户情绪/用意、和/或对话场景的周边信息。对用户当前言语的响应一般基于言语的内容以及用于驱动对话的对话树(例如对话750)生成。根据本示教,本地DM 820一旦被调用,搜索本地DT 830,以查看本地DT 830是否能用于识别合适的响应。搜索是基于当前言语的内容的。配置本地DM 820和本地DT830的预期意图是,在大多数情况下,可以在本地找到响应,节省与服务器840通信以识别响应的时间和流量。如果是这种情况,如1020处所确定的,来自用户的当前言语的内容落在本地DT 830的非叶子节点上,响应是非叶子节点的分支中的一个。也就是说,本地DM 820在1025处基于本地DT 830的搜索生成响应,这样生成的响应于是被响应呈现单元940在1030处呈现给用户。

[0106] 在某些情况下,不能在本地DT 830中找到响应。当这种情况发生时,响应需要由服务器840根据整体对话树850生成。可能存在不同的情境,其中,本地DM 820不能基于本地DT 830找到响应。例如,来自用户的当前言语的内容可能在本地DT 830中找不到。在这种情况下,对来自用户的未识别言语的响应将由服务器840确定。在不同的情况下,在本地DT 830中找到当前言语,但其响应并非本地存储的(例如,当前对话状态对应于本地DT830的叶子节点)。在这种情况下,响应也不是本地可获得的。在这两种情境之下,缓存在830中的本地对话树不能用于进一步驱动对话,于是,本地DM 820调用装置/服务器协调器930,以便在1035处向服务器840发送具有与对话状态相关的信息的、对响应的请求,从而促使服务器识

别合适的响应。然后,装置/服务器协调器930在1040和1045处分别接收所寻求的响应以及更新的本地DM和本地DT。在接收到更新的本地DM和本地DT时,装置/服务器协调器930于是调用本地对话管理器更新其950和本地对话树更新器960,以便在1050处更新本地DM 820和本地DT 830。装置/服务器协调器930还在1055处将接收的响应发送到响应呈现单元940,使得响应可在1030处向用户呈现。

[0107] 图11示出了根据本示教一实施例,服务器840的示例性系统图。在此所示的实施例中,服务器840包含装置接口单元1110、当前本地DM/DT信息检索器1120、当前用户状态分析器1140、对话管理器860、更新本地DT确定器1160、更新本地DM确定器1150以及本地DM/DT生成器870。图12为根据本示教一实施例,服务器840的示例性过程的流程图。在运行中,当装置接口单元1110在图12的1210处接收到来自装置的、具有与对话的当前状态有关的信息的、寻求响应的请求时,其调用当前用户状态分析器1140,以便在1220处分析所接收的相关信息,从而理解用户的输入。为了识别对用户输入的响应,对话管理器860被调用,以便在1230处检索完整的对话树750,从而获得响应。

[0108] 如这里所讨论的,当服务器840被请求提供在装置上对对话的响应时,其指示先前在该装置上配置的本地DM 820和本地DT 830不再为该本地对话工作(它们已经导致未命中)。因此,除了为装置提供响应之外,服务器840还生成将要缓存在装置上的更新的本地DM和本地DT。在某些实施例中,为了实现这一点,装置接口单元1110还调用当前本地DM/DT信息检索器1120,以便在1240处检索与先前配置在装置上的本地DM/DT有关的信息。

[0109] 这样检索的关于先前配置的本地DM和本地DT的信息,连同当前服务器生成的响应和当前对话状态,被发送到更新本地DT确定器1160和更新本地DM确定器1150,以便在1250处,关于当前响应和当前对话状态确定更新的本地DT和更新的本地DM。这样确定的更新本地DM/DT于是被发送到本地DM/DT生成器870,于是,其在1260处生成将被发送到装置的、更新的本地DM/DT。生成的更新本地DM/DT于是被存档在本地DT/DM分发档案1130中,接着,被装置接口单元1110发送到装置。通过这种方式,每当出现未命中时,服务器840更新装置上的本地DM/DT,使得服务器支持装置所需要的通信流量和带宽可被减低,因此,在人机对话中响应于用户的速度可得到增强。

[0110] 按照传统,基于对话树的搜索,对话管理系统(例如对话管理器840)取得文本(例如基于语音理解生成的)并输出文本。在某种意义上,对话树对应于决策树。在基于这样的决策树驱动的对话的每一步骤中,可能存在表征当前言语的节点和从该节点分叉而出的表征连接到该节点的所有可能回答的多个选择。因此,从每个节点,可能的响应可以沿着多个路径中的任何一个。从这个意义上来说,对话的过程遍历对话树,并形成对话路径,如图4B所示。对话管理器的工作是,通过在底层对话方面最优化某种增益,确定每个节点(表征用户的言语)上的选择。基于来自不同来源的信息和理解对话周边情境的不同方面判断所选择的路径可能需要花费时间。

[0111] 另外,由于有限的计算力和内存,生成对用户的响应的大量计算工作在服务器(例如图7中的720)上进行。例如,当用户信息被接收时,服务器720可分析用户信息,以理解在说什么。驻留服务器的对话管理器740于是搜索对话树750,以识别合适的响应。如这里所讨论的,这一对话过程严重依赖于后端服务器,并需要装置710与服务器720之间往复的通信流量。这花费时间和带宽,影响服务器扩大规模进行与多个用户的同时实时对话的能力。

[0112] 本示教还公开了一种方法,通过预测对话树750中用户可能在不远的将来采取哪个或哪些路径,并先发式地沿着预测路径生成预测响应,使得人机对话响应时间的进一步减少成为可能。每个用户的路径预测可以基于描绘例如用户偏好的模型,并基于例如过去的对话历史和/或公知常识经由机器学习创建。这样的训练可以是在不同粒度等级上个性化的。例如,为预测对话路径学习的模型可以是基于关于个体所收集的过去的数据进行个性化的。也可通过基于相关训练数据来训练个性化模型,使得对于这种对话路径预测的模型适应个别需要,例如,为了为共有类似特性的一组用户对模型进行训练,可使用对于类似用户的训练数据。这种训练和预测可以离线进行,训练结果于是可应用于在线运行,以减少对话管理器的计算负荷以及响应时间,使得服务器能更好地扩大规模,处理高容量的请求。

[0113] 通过预测对话路径并先发式地生成可能的响应,当响应在先发式生成的响应之中时,预先生成的响应于是可被直接提供给用户,而不必调用对话管理器来搜索对话树,例如750。如果响应不在先发式生成的那些当中,于是,可通过请求对话管理器搜索对话树来做出请求,以找到响应。图13示出了根据本示教一实施例,使用先发式预测对话路径(整体对话树750的子部分)以及对话内容(响应)的框架的示例性实施例。在此所示的实施例中,用户1300经由装置1310通信,装置1310可以与图7所示类似地构建,并与服务器1320通信,服务器1320使用先发式预测对话路径和响应,增强响应人类用户时的延迟。在此所示的实施例中,服务器1320包含控制器1330、对话管理器1340、预测路径/响应生成器1350,预测路径/响应生成器1350生成预测对话路径1360并相应地生成先发式生成的响应1370。在运行中,当装置1310接收用户信息(言语、视频等)时,为了确定响应,装置1310向服务器1320发送请求,用与对话状态有关的信息(例如言语和/或对话周边情况的观察,例如用户的态度、情绪、用意、对话场景中的对象及其描绘)寻求响应。如果所请求的响应在预测路径1360中,于是,直接从预测响应1370之中检索1370中对应的先发式生成的响应,并发送到装置1310。通过这种方式,由于对话管理器1340不被调用以处理请求并搜索对话树750以获得请求,故而,提供响应的延迟得到改善。

[0114] 图14示出了根据本示教一实施例,服务器1320的示例性高层次系统图。在所示的实施例中,服务器1320包含对话状态分析器1410、响应来源确定器1420、对话管理器1340、预测响应检测器1430、响应发送器1440、预测路径生成器1460、预测响应生成器1450。图15为根据本示教一实施例,服务器1320的示例性过程的流程图。在运行中,对话状态分析器1410在图15的1505处接收请求,该请求具有与底层对话的状态有关的信息,该信息包括,例如,表征用户的语音的听觉数据,或分析的用户语音,以及视情况可选地,与对话状态有关的其他信息。这样接收的信息在1510处被分析。为了确定适合响应于用户言语的响应是否已经在以前被先发式地生成,响应来源确定器1420被调用,以便在1515处,基于预测路径1360中所存储的,确定与用户当前言语相关的预测路径是否存在。如果与用户当前言语相关的预测路径1360存在,其在1520处进一步检查,关于预测路径,对于当前言语的希望响应是否在预测路径中存在,即,对当前言语的希望响应是否已经被先发式地生成。如果想要的响应已经在先前生成,响应来源确定器1420调用预测响应检索器1430,以便在1525处,从预测响应1370中检索先发式地生成的响应,于是,调用响应发送器1440,以便在1530处将先发式地生成的响应发送到装置1310。

[0115] 如果如1515处所判断的,与言语相关的预测路径不存在,或者,如果如1520处所判

断的,希望的响应没有被先发式地生成的(在预测路径中),过程进行到在1535处调用对话管理器1340,以便生成关于当前用户言语的响应。这涉及搜索对话树750,以识别响应。在未命中的情况下(即,预测路径不存在,或者,已经存在的预测路径不包含响应),对话管理器1340也可致动预测路径生成器1460,以便在给定当前言语/所识别响应的情况下预测路径。在被致动时,为了生成预测路径,预测路径生成器1460可在1540处分析当前生成的响应,以及,视情况可选地,还分析当前涉入对话的用户的配置文件,该配置文件从用户配置文件存储器1470中取得。基于这样的信息,预测路径生成器1460在1545处基于当前言语/响应、对话树750以及视情况可选的用户配置文件来预测路径。基于预测的路径,预测响应生成器1450在1550处生成与新预测路径相关联的预测响应,即,先发式地生成响应。这样预测的新路径及其先发式生成的预测响应于是在1555处被预测路径生成器1460和预测响应生成器1450分别存储在预测路径存储器1360和预测响应存储器1370中。于是,如此识别的响应在1530处返回到装置,以便向用户响应。

[0116] 图16示出了根据本示教的实施例,在管理与用户的对话中,装置1610和服务器1650之间的不同的示例性配置。相比于图13所示的实施例,为了进一步增强性能并减小延迟和流量,图16的配置还在装置1610上配置本地对话管理器1620,其具有对应的本地预测路径1640以及对应的先发式生成的本地预测响应1630。本地对话管理器1620基于本地预测路径1640和本地预测响应1630运行,以便尽自己可能地驱动对话。当存在未命中时,装置向服务器1650发送具有与对话状态有关的信息的请求,以寻求响应。如图所示,服务器1650也存储预测路径1630的服务器版本以及预测响应1370的服务器版本。在某些实施例中,存储在服务器1650上的服务器预测路径1360和服务器预测响应1370可能不与本地版本1640和1630相同。例如,服务器预测路径1360可能比本地预测路径1640更为广泛。这样的区别可以基于不同的运行考虑,例如本地存储器限制,或传输大小的局限。

[0117] 在运行中,当装置上存在未命中时,装置1610向服务器1650发送请求,请求具有与对话有关的信息,且请求对于当前对话的响应。当这种情况发生时,服务器1650可识别合适的响应,并将响应发送到装置。服务器识别的这样的响应可以是服务器预测路径1360中的服务器预测响应1370中的一个。如果响应不能在服务器预测路径/响应中找到,服务器于是可以在整体对话树中搜索,以识别响应。采用两个层级的(装置和服务器)缓存的预测路径和响应,生成响应需要的时间进一步减少。

[0118] 如图16中的配置所示,装置1610包含本地对话管理器1620,其被配置为本地地发挥作用,以便通过搜索预测路径1360的本地版本1640以及先发式地生成的响应1630(其为存储在服务器1650上的预测响应的本地版本),生成对用户1600的响应。如果本地对话管理器1620基于预测路径1640和预测响应1630在本地找到响应,装置1610将向用户提供响应,而不从服务器1650请求响应。在这种配置中,当存在未命中时,装置1610请求服务器1650提供响应。在接收到请求时,服务器1650可进行到基于服务器预测路径1360和服务器预测响应1370生成响应。如果服务器预测路径1360及服务器预测响应1370相比于本地预测路径1640及对应的本地预测响应1630更为广泛,本地预测路径/响应中找不到的响应可能包含在服务器预测路径/响应中。仅仅在服务器1650不能在其预测路径1360和预测响应1370中找到响应的条件下,服务器1650进行到在整体对话树750中搜索响应。

[0119] 除了识别用于装置的响应以外,服务器1650也可生成更新的本地预测路径1640、

对应的本地预测响应1630以及更新的本地对话管理器1620,更新的本地对话管理器1620可关于更新的本地预测路径/响应运行。更新的本地预测路径/响应和更新的本地对话管理器于是可被发送到装置,用于将来的运行。预测路径和预测响应的更新的本地版本可以基于整体对话树750或者是服务器预测路径1360以及服务器预测响应1370生成。在某些情况下,当服务器不能从服务器预测路径1360以及服务器预测响应1370中识别合适响应时,在这种情况下,预测路径/响应的服务器和本地版本以及本地对话管理器都需要更新。如果合适的响应尽管没有在装置1610上找到,但从服务器预测路径/响应中识别出来,服务器预测路径/响应可能不需要更新。

[0120] 如这里所讨论的,当接收到对于响应的请求时,更新的本地预测路径/响应可由服务器生成。在某些情况下,更新的本地预测路径/响应可以由现有的服务器预测路径/响应生成。在某些实施例中,更新的服务器预测路径/响应也可能需要更新,使得更新的本地预测路径/响应于是可以基于更新的服务器预测路径/响应来生成,其中,更新的服务器预测路径/响应是基于对话树750生成的。在这种情况下,服务器可以生成预测路径和预测响应的更新的服务器版本和本地版本,即,对预测路径和预测响应的更新在服务器1650和装置1610二者之上发生。一旦更新的本地预测路径/响应被生成,更新的本地对话管理器于是可被相应地生成。一旦生成,更新的本地对话信息(包括更新的本地预测路径/响应和更新的本地对话管理器)于是从服务器被发送到装置,使得其可用于在装置上更新本地对话管理器1620、预测路径1640和预测响应1630。

[0121] 图17示出了根据本公开一实施例,装置1610的示例性高层次系统图。为了实现图16所示的示例性配置,装置1610的示例性构造包含对话状态分析器1710、响应来源确定器1720、本地响应管理器1620、预测响应检索器1730、响应发送器1740、装置/服务器协调器1750以及预测路径/响应更新器1760。装置1610也包含本地预测路径1640和本地预测响应1630,它们都被本地对话管理器1620用于驱动装置与用户之间的对话。如这里所讨论的,经由装置/服务器协调器1750,本地预测路径1640和本地预测响应1630可被预测路径/响应更新器1760基于从服务器1650接收的本地预测路径/响应的更新版本更新。

[0122] 图18为根据本示教一实施例,装置1610的示例性过程的流程图。在运行中,当对话状态分析器1710在图18的1810处接收到与正在进行的对话有关的信息(其包括用户的言语以及对话周边的其他信息)时,其在1820处判断对话的对话状态。与对话有关的周边信息可包括多模态信息,例如用户言语的音频、关于用户的视觉信息(例如用户的面部表情或姿态)、或其他类型的传感器数据,例如与用户的移动有关的触觉信息。由对话状态分析器1710基于所接收的周边信息确定的对话状态可包括用户言语的内容,基于例如用户面部表情和/或用户语音音调确定的用户的情绪状态,推定的用户的用意,对话环境中的相关对象,等等。

[0123] 基于当前对话状态中的用户言语,响应来源确定器1720确定对用户言语的响应是否能够基于本地存储的预测路径1640和本地存储的预测响应1630被识别出。例如,在1830处,判断本地预测路径是否与当前言语相关。例如,当其包含与当前言语对应的节点时,本地预测路径可能是相关的。如果本地预测路径是相关的,其可以在1840处进一步检查本地预测路径是否包含可用于对用户的言语进行响应的、先发式地生成的(预测的)响应。如果本地预测路径中先发式地生成的响应作为对用户的响应是合适的,本地对话管理器1620被

调用,以便基于本地存储的预测路径1640和本地存储的预测响应1630生成响应。在这种情况下,本地对话管理器1620调用预测响应检索器1730,以便在1850处检索先发式地生成的响应(例如根据本地对话管理器1620的指示),并将检索到的先发式生成的响应转发给响应发送器1740,以便在1855处将本地识别的响应发送到用户。在这种情境下,装置1610不需要请求服务器提供响应(节省时间),也不需要与服务器1650通信(减少流量),使其有效地在所需计算、带宽和延迟方面增强性能。

[0124] 如果本地预测路径与当前言语不相关,或者,对用户言语的合适的响应不能在本地预测响应中找到,装置/服务器协调器1750被调用,以便与服务器1650通信,以得到响应。为了做到这一点,装置/服务器协调器1750在1860处向服务器1650发送具有与对话状态有关的信息的对响应的请求,并等待接收反馈。当装置/服务器协调器1750接收到来自服务器的反馈时,反馈可包括在1870处接收的所寻求的响应,以及在1880处接收的具有更新的预测响应的更新的本地预测路径和相应地生成的更新的本地对话管理器。采用这样接收的本地对话信息,本地对话信息更新器1760进行到在1890处更新本地对话信息,其包括本地预测路径1640、本地预测响应1630、本地对话管理器1620。接收到的响应于是在1855处经由响应发送器1440被发送给用户。

[0125] 图19示出了根据本示教一实施例,服务器1650的示例性高层次系统图。在此所示的实施例中,服务器1650包含对话状态分析器1910、响应来源确定器1920、对话管理器1340、预测响应检索器1930、预测路径/响应生成器1960、本地对话管理器生成器1950、响应/本地对话信息发送器1940。图20为根据本示教一实施例,服务器1650的示例性过程的流程图。在运行中,当对话状态分析器1910在图20的2005处接收到来自装置的具有相关联的对话状态信息的对响应的请求时,其在2010处分析所接收的对话状态,并将该信息传送到响应来源确定器1920,以便确定将会在哪里识别出所寻求的响应。在某些情况下,响应可以从与服务器预测路径1360相关联的服务器预测响应中找到。在某些情况下,响应可能需要从整体对话树750识别。

[0126] 如果在2015处判断为服务器预测路径1360存在,进一步在2020处判断对当前对话状态的响应是否能在服务器预测路径1360中找到。如果响应能在服务器预测路径1360中被找到,预测响应检索器1930被调用,以便在2025处从1370中检索先发式生成的预测响应,检索到的响应被发送到响应/路径发送器1940,用于将响应与其他更新的对话信息一起发送,其他更新的对话信息包括更新的本地预测路径、更新的预测响应、更新的本地对话管理器。如果没有合适的服务器预测路径1360可用于生成响应(例如没有服务器预测路径,或现有的服务器预测路径1360与当前对话状态不相关)或对于当前对话状态合适的响应不能在服务器预测路径1360中找到,响应来源确定器1920调用用对话管理器1340,以便基于整体对话树750在2030处生成关于当前对话状态的响应。

[0127] 如这里所讨论的,每当服务器被要求生成响应时(即,装置上存在未命中),其表示本地预测路径和本地预测响应不再能使本地对话管理器驱动对话成为可能。因此,在响应于向装置提供响应的请求时,服务器1650也可为装置生成更新的本地预测路径和更新的预测响应。另外,更新的本地对话管理器也可能需要被相应地生成,以便与更新的本地预测路径及响应一致。这样的与更新本地对话有关的信息可以由服务器生成,并与所生成的响应一起被发送到装置。

[0128] 另外,由于服务器上也可能存在关于服务器预测路径1360以及服务器预测响应1370的未命中,服务器预测路径和服务器预测响应也可能在出现服务器层次的未命中时需要被更新。在这种情境下,服务器和本地版本的预测路径及响应可以被重新生成,并用于更新前面的版本。因此,在2035处,判断服务器预测路径和服务器预测响应是否需要更新。如果需要,预测路径/响应生成器1960被调用,以便分别在2040和2045处生成更新的服务器预测路径和服务器预测响应。在这种情境中,更新的服务器预测路径/响应用于在2050处生成更新的本地预测路径以及对应的更新的预测响应。

[0129] 如果如在2035处判断的,服务器预测路径/响应不需要被更新,更新的本地预测路径和响应于是在2050处基于当前版本的服务器预测路径和服务器预测响应被生成。更新的本地预测路径和更新的本地预测响应于是由本地对话管理器生成器1950用于在2055处,根据对话树750和对话管理器1340,基于更新的本地预测路径和更新的本地预测响应,生成更新的本地对话管理器1620。服务器生成的响应于是在2060处与更新的本地对话信息(其包括更新的本地预测路径、更新的本地预测响应以及更新的本地对话管理器)一起被发送给装置,使得它们能被本地对话信息更新器1760(图17)使用,以更新本地预测路径1640、本地预测响应1630、本地对话管理器1620。

[0130] 图21示出了根据本示教的实施例,管理与用户的对话中,在服务器和装置之间的又一示例性运行配置。在此所示的实施例中,代替在服务器上保持服务器预测路径和服务器预测(先发式地生成的)响应的副本的是,服务器保留将要与预测路径/响应/本地对话管理器有关地被分发到装置的内容的记录。在这种配置中,由于不存在预测路径和响应的服务器版本,每当服务器被请求提供响应时,服务器中的对话管理器将直接从整体对话树中识别这样的响应。基于这样识别的响应,服务器于是进行到生成更新的本地预测路径/响应和更新的本地对话服务器,于是,其可以与响应一起被发送到装置。所接收的预测路径/响应/对话管理器的更新本地版本于是被用于替换先前的本地对话管理器1620、先前的本地预测路径1640、先前的本地预测响应1630,以便促进装置上的进一步的本地对话管理。这在图21中示出,其中,这一配置中的服务器2110包含本地对话信息分发日志2120。

[0131] 采用这种配置,装置1610基于本地预测路径1640和对应的本地预测(先发式地生成的)响应1630(二者均由服务器2110预测并在装置1610上动态配置)来执行本地化的对话管理。服务器1670可以在接收到来自装置1610的请求和与当前对话状态有关的信息时,识别装置不能在先前配置的预测路径中找到的响应,于是,基于所接收的信息,先发式地生成预测对话路径和预测响应。在此实施例中,服务器1670可以不保持用于不同装置的预测对话路径并基于它们运行。相反,这样的预测对话路径和响应被发送到个体装置,以便使它们能够相应地管理其自己的本地对话。在这种配置中,服务器可在分发日志2120中保持信息,该信息记录被发送到不同装置的本地预测对话路径以及与之关联的先发式生成的响应。在某些实施例中,当先前的版本不能再用于驱动对话时,这样记入日志的信息可用于生成对应的更新本地预测路径以及先发式地生成的响应。

[0132] 图22示出了根据本示教一实施例,服务器2110的示例性高层次系统图。如图所示,服务器2110包含对话状态分析器2110、对话管理器1340、本地预测路径/响应生成器2220、本地对话管理器生成器2230、本地对话信息发送器2240、分发记录更新器2250。图23为根据本示教一实施例,服务器2110的示例性过程的流程图。在运行中,在2310处,当请求被对话

状态分析器2210接收到时,请求受到分析,且其被对话管理器1340用于在2320处基于对话状态和整体对话树750生成响应。如这里所讨论的,对话状态可包括操作装置的用户的话语以及对话周边的其他信息,例如面部表情、推定的用户情绪状态、用户的用意、对话场景中的相关对象及其描绘。在某些实施例中,当对话管理器1340生成响应时,其还可考虑对话周边的信息,例如用户的情绪状态和/或用户的配置文件信息,例如用户喜欢什么。例如,如果用户的话语并非回应性的,带有负面的情绪状态,对话管理器1340可识别更为基于用户配置文件驱动的反应,而不是遵循对话树750中的设定的路径。例如,如果用户的话语与对话不太相关且用户看起来不满,对话管理器可选择更为基于用户偏好驱动的反应,而不是由对话树750驱动的。如果用户喜欢篮球且对话场景中有篮球,对话管理器1340可决定与用户谈论篮球,从而在继续对话的初始话题之前使用户重新专注。

[0133] 这样生成的响应于是由本地预测路径/响应生成器2220用于在2230处生成更新的本地预测路径和更新的本地响应。这样的更新本地对话信息的生成可以不仅基于响应,还基于来自对话状态和/或用户配置文件的附加信息。通过这种方式,本地更新预测路径和响应与对话管理器1340所生成的响应、当前对话状态和/或用户的偏好一致。基于更新的本地预测路径和响应,在2340处,由本地对话管理器生成器生成更新的本地对话管理器。更新的本地对话信息(本地预测路径、本地预测响应和本地对话管理器)于是被发送到本地对话信息发送器,其于是在2350处将这样的信息发送到装置1610,使得本地预测路径、本地预测响应和本地对话管理器可用更新版本替换,从而在装置1610上本地地驱动未来的对话。于是,分发记录更新器2250在2360上更新对话信息分发日志2120。

[0134] 图24为示例性移动装置架构的示意图,该架构可用于根据多种实施例实现实施本示教的至少某些部分的特定系统。在此实例中,实现本示教的用户装置对应于移动装置2400,其包括但不限于智能电话、平板电脑、音乐播放器、手持游戏机、全球定位系统(GPS)接收器、可穿戴计算装置(例如眼镜、腕表等)或其它任何外形因素。移动装置2400可包括一个以上的中央处理单元(CPU)2440、一个以上的图形处理单元(GPU)2430、显示器2420、内存2460、诸如无线通信模块的通信平台2410、存储器2490以及一个以上的输入/输出(I/O)装置2440。任何其他合适的部件,包括但不限于系统总线或控制器(未示出),也可包含在移动装置2400中。如图24所示,移动操作系统2470(例如iOS、Android、Windows Phone等)以及一个以上的应用2480可从存储器2490被装载到内存2460中,以便由CPU 2440执行。应用2480可包括浏览器或任何其他合适的移动app,用于管理移动装置2400上的会话系统。用户交互可以经由I/O装置2440实现,并经由网络120被提供给应用客户端。

[0135] 为了实现本公开中介绍的多种模块、单元及其功能,计算机硬件平台可作用于这里介绍的一个或多个元件的硬件平台。硬件元件、操作系统和这种计算机的编程语言在性质上是传统的,且假设本领域技术人员足够熟悉它们,以便使这些技术适应于这里介绍的本示教。具有用户接口元件的计算机可用于实现个人计算机(PC)或其他类型的工作站或终端装置,但是,如果合适地编程的话,计算机也可作为服务器运行。相信本领域技术人员熟悉这种计算机设备的结构、编程和一般运行,因此,附图可能是不言自明的。

[0136] 图25为示例性计算装置架构的示意图,该架构可用于根据多种实施例实现实施本示教的至少某些部分的特定系统。实现本示教的这种特定系统具有硬件平台的功能框图,该硬件平台包括用户接口元件。计算机可以是通用计算机或专用计算机。二者都能用于实

施用于本示教的特定系统。这种计算机2500可用于实现如这里所介绍的会话或对话管理系统的任何部件。例如，会话管理系统可以在例如计算机2500的计算机上实现，经由其硬件、软件程序、固件或其组合。尽管为方便起见示出了仅仅一个这样的计算机，与这里介绍的会话管理系统有关的计算机功能可以以分布式方式在若干个类似的平台上实现，从而分散处理负荷。

[0137] 例如，计算机2500包括与连接于其上的网络相连接的COM端口2550，以促进数据通信。计算机2500还包括中央处理单元(CPU) 2520，其采用一个或多个处理器的形式，用于执行程序指令。示例性计算机平台包括：内部通信总线2510；不同形式的程序存储器和数据存储器(例如盘2570、只读存储器(ROM) 2530或随机访问存储器(RAM) 2540)，用于将要由计算机2500处理和/或进行通信的多种数据文件以及将由CPU 2520执行的可能的程序指令。计算机2500还包括I/O部件2560，其支持在计算机和这里的其他部件(例如用户接口元件2580)之间的输入/输出流。计算机2500也可经由网络通信接收编程和数据。

[0138] 因此，如上面所概述的对话管理方法和/或其他过程的实施形态可以在程序中实现。本技术的程序方面可被看作典型地出于可执行代码和/或相关数据的形式“产品”或“制品”，该可执行代码和/或相关数据被承载在一种机器可读介质上或在其中实现。有形非暂时性“存储器”类型介质包括任何或全部存储器或其他的用于计算机、处理器等的存储器或其相关模块，例如多种半导体存储器、带驱动器、盘驱动器等，其可在任何时候提供用于软件编程的存储。

[0139] 所有或部分软件有时可通过网络(例如互联网或多种其他电信网络)传送。例如，这种传送可使软件从一台计算机或处理器向另一台(例如与会话管理有关)的载入成为可能。因此，可承载软件元件的另一类型的介质包括光、电和电磁波，例如通过本地装置之间的物理接口、通过有线和光固定网络、通过多种空中链路使用。承载这种波的物理元件(例如有线或无线链路，光链路等)也被看作承载软件的介质。如这里所使用的，除了限制为有形的“存储”介质，例如计算机或机器“可读介质”的术语指参与向处理器提供指令以便执行的任何介质。

[0140] 因此，机器可读介质可采用多种形式，包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质包括例如光或磁盘，例如任何计算机等等之中的任何存储装置，其可用于实现附图所示的系统或其任何部件。易失性存储介质包括动态存储器，例如这种计算机平台的主存储器。有形传输介质包括：同轴电缆、铜线和光纤，其包括构成计算机系统内的总线的导线。载波传输介质可采用电或电磁信号或者是声或光波(例如在射频(RF)和红外(IR)数据通信期间生成的那些)的形式。计算机可读介质的一般形式因此包括例如软盘、可折叠盘、硬盘、磁带、任何其他磁介质、CD-ROM、DVD或DVD-ROM、任何其他光介质、穿孔卡片纸带、具有孔的图案的任何其他物理存储介质、RAM、PROM和EPROM、闪速EPROM、任何其他的存储器芯片或插装盒、传输数据或指令的载波、传送这样的载波的链路或电缆、或计算机可从之读取编程代码和/或数据的任何其他介质。许多这些形式的计算机可读介质可以涉入将一个或多个的指令的一个或多个的序列承载到物理处理器，以便执行。

[0141] 本领域技术人员将会明了，本示教适用于多种修改和/或增强。例如，尽管上面介绍的多种部件的实现可以在硬件装置中实现，其还可实现为仅仅使用软件的解决方案，例

如安装在已有的服务器上。另外,这里所公开的欺骗性网络检测技术也实现为固件、固件/软件组合、固件/硬件组合或是硬件/固件/软件组合。

[0142] 尽管上面已经介绍了本示教和/或其他实例,将会明了,可对之做出多种修改,且这里公开的主题可以以多种形式和实例实现,且本示教可以在多种应用中应用,这里仅仅介绍了其中的一些。所附权利要求旨在要求落入本示教真实范围内的任何以及全部应用、修改和变型。

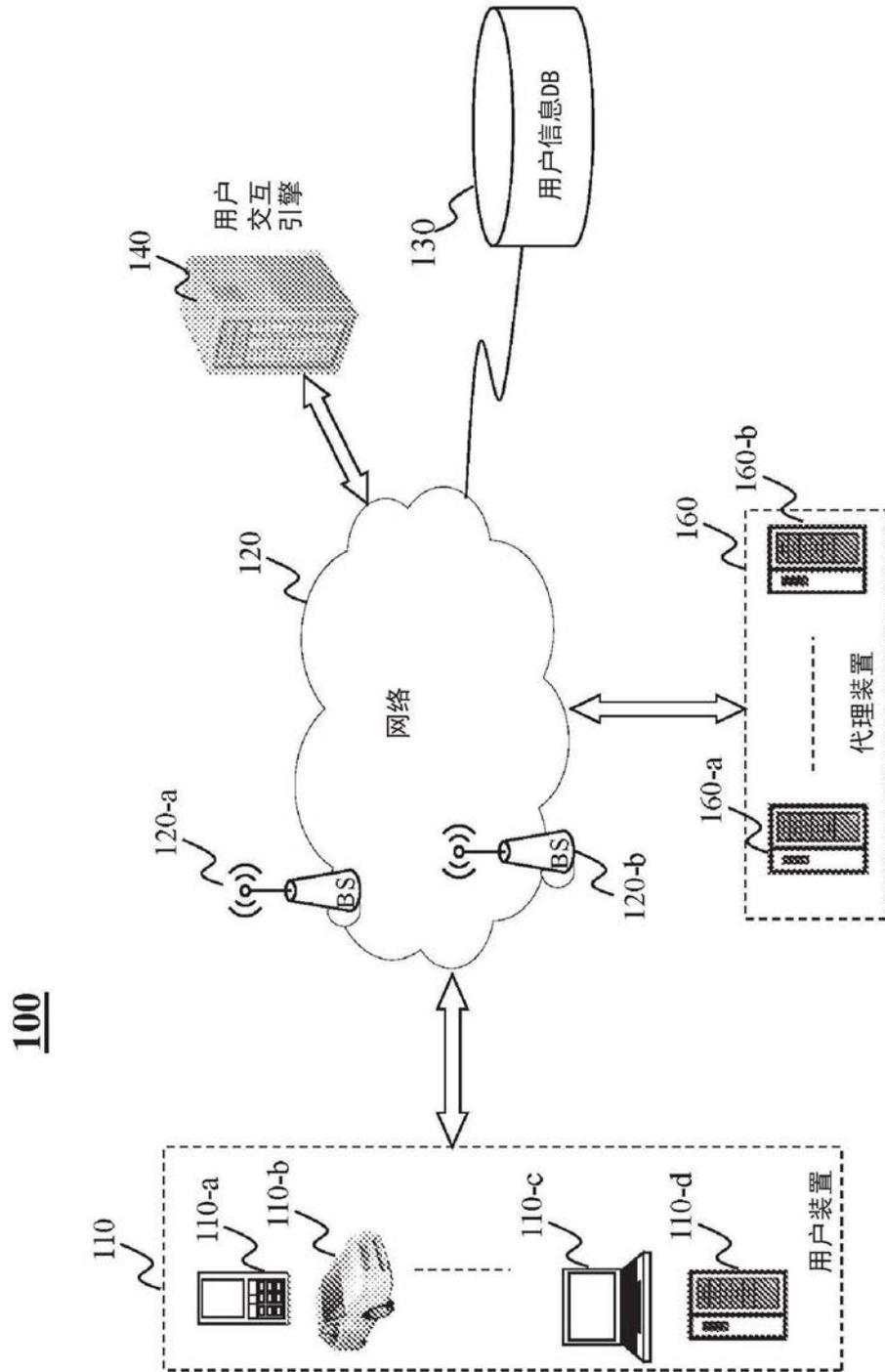


图1

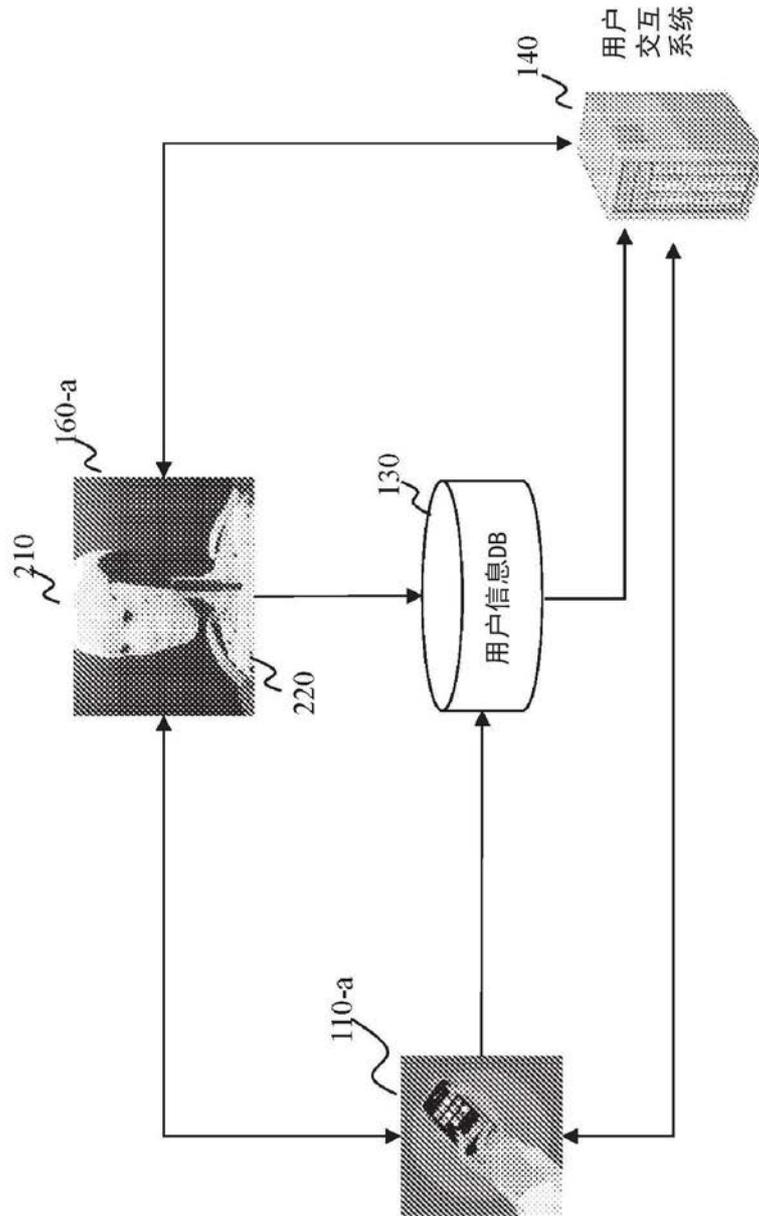


图2A

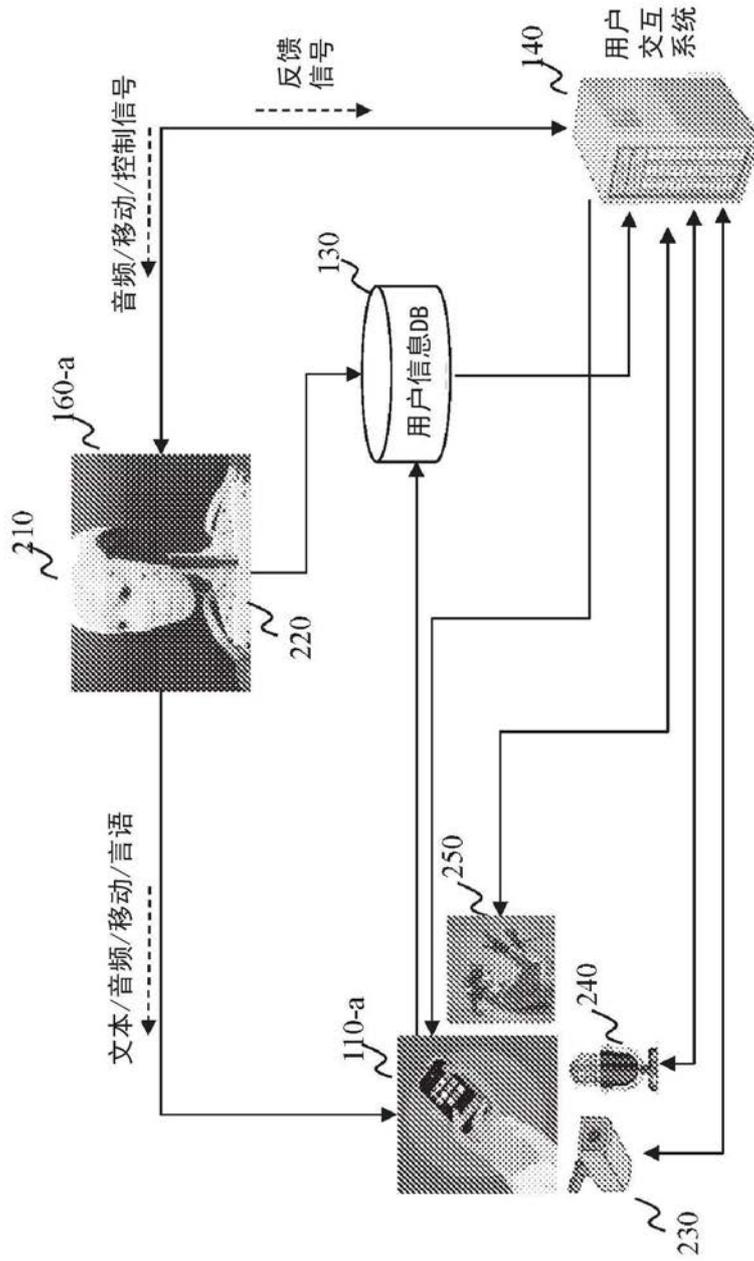


图2B

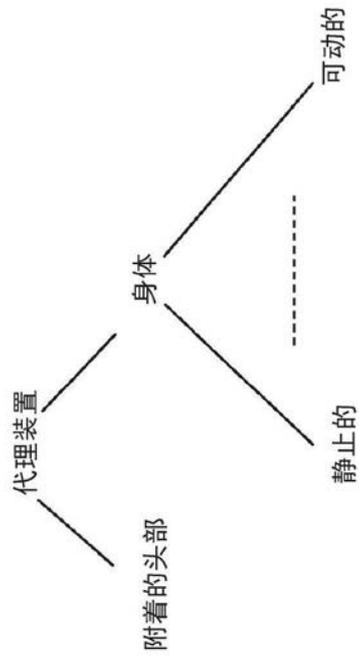


图3A

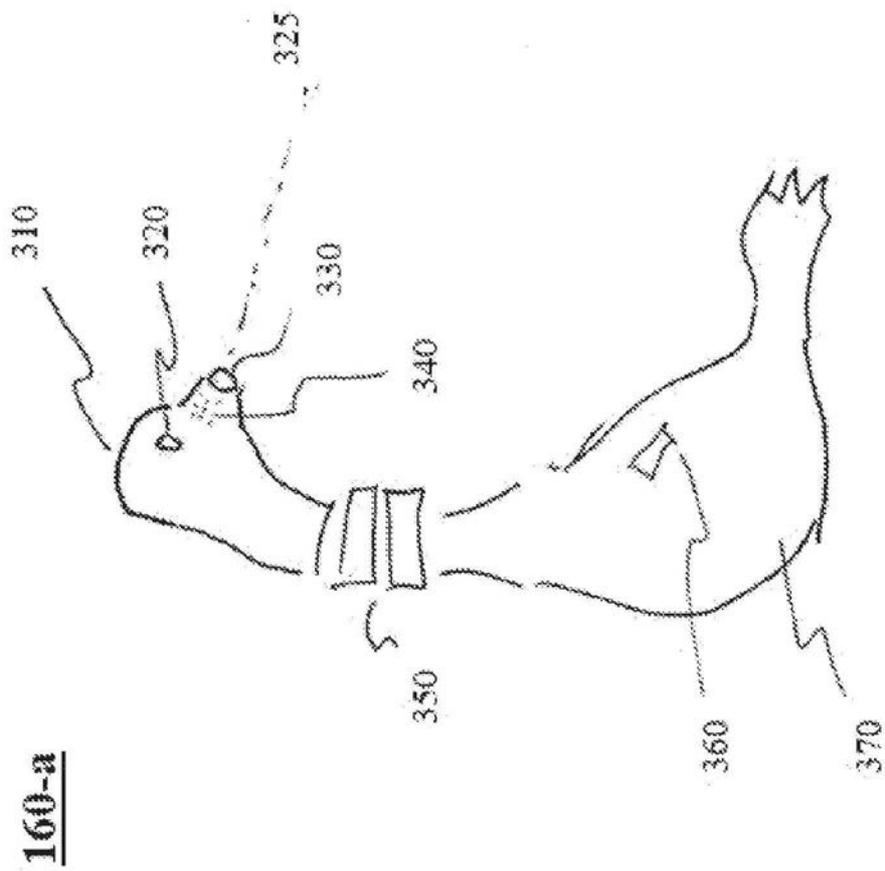


图3B

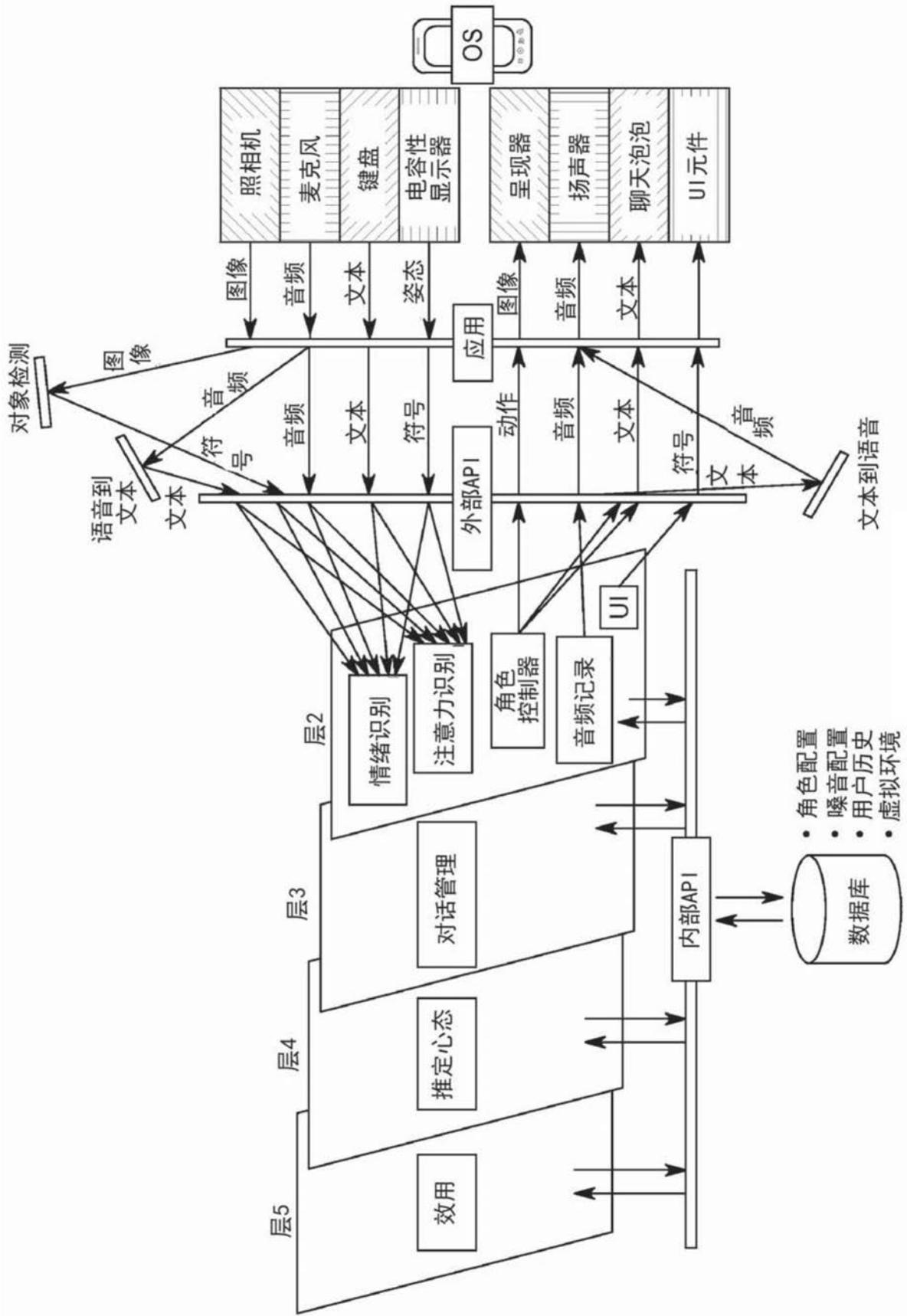


图4A

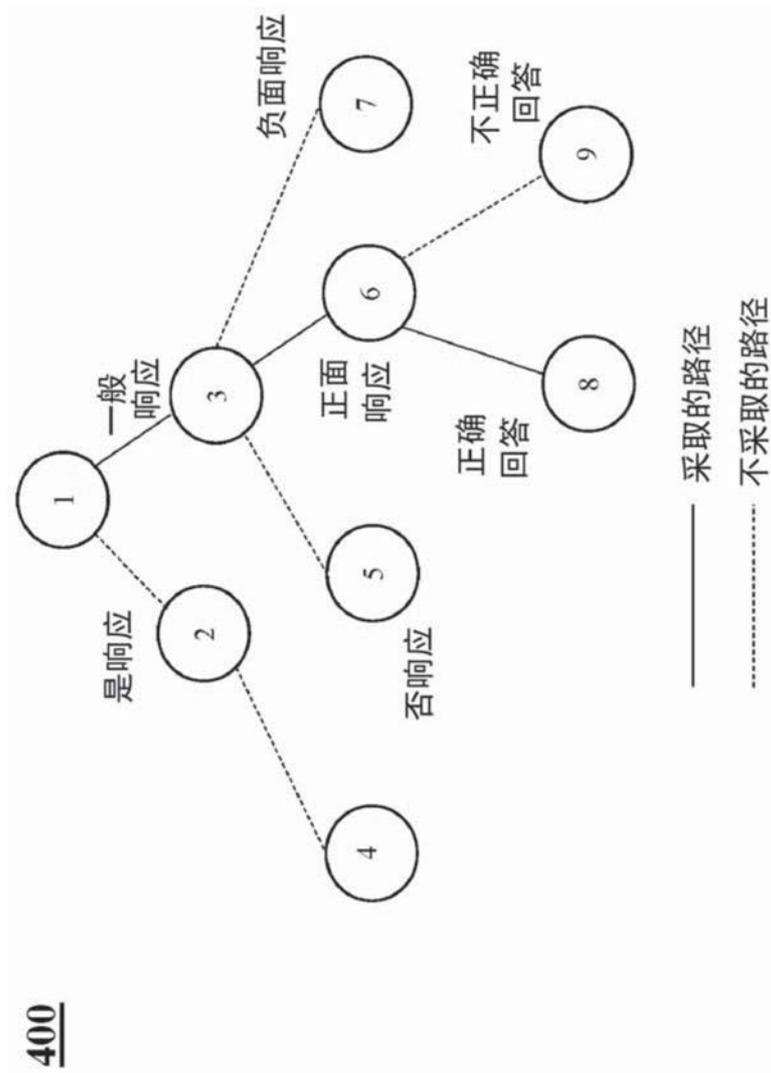


图4B

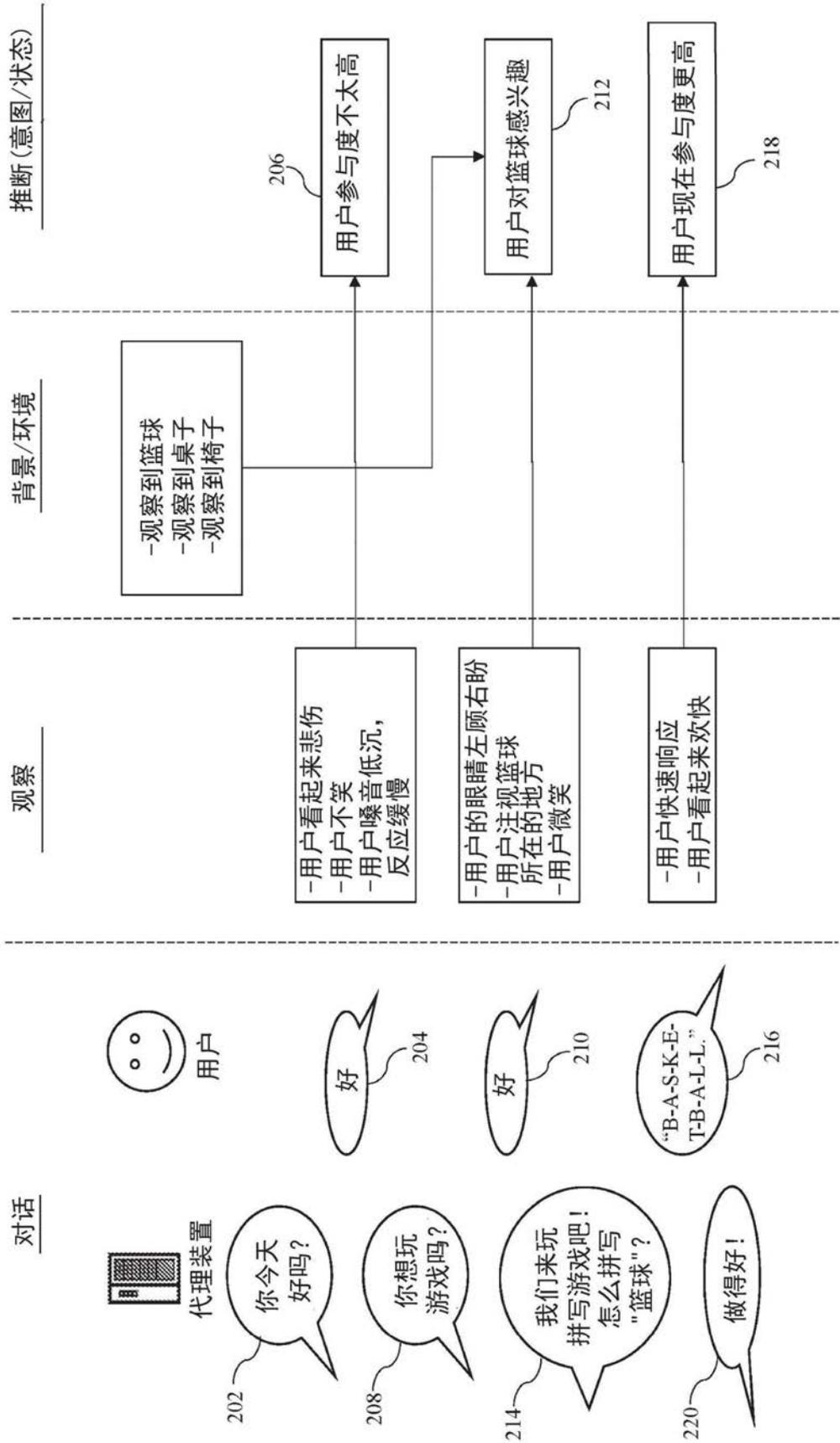


图4C

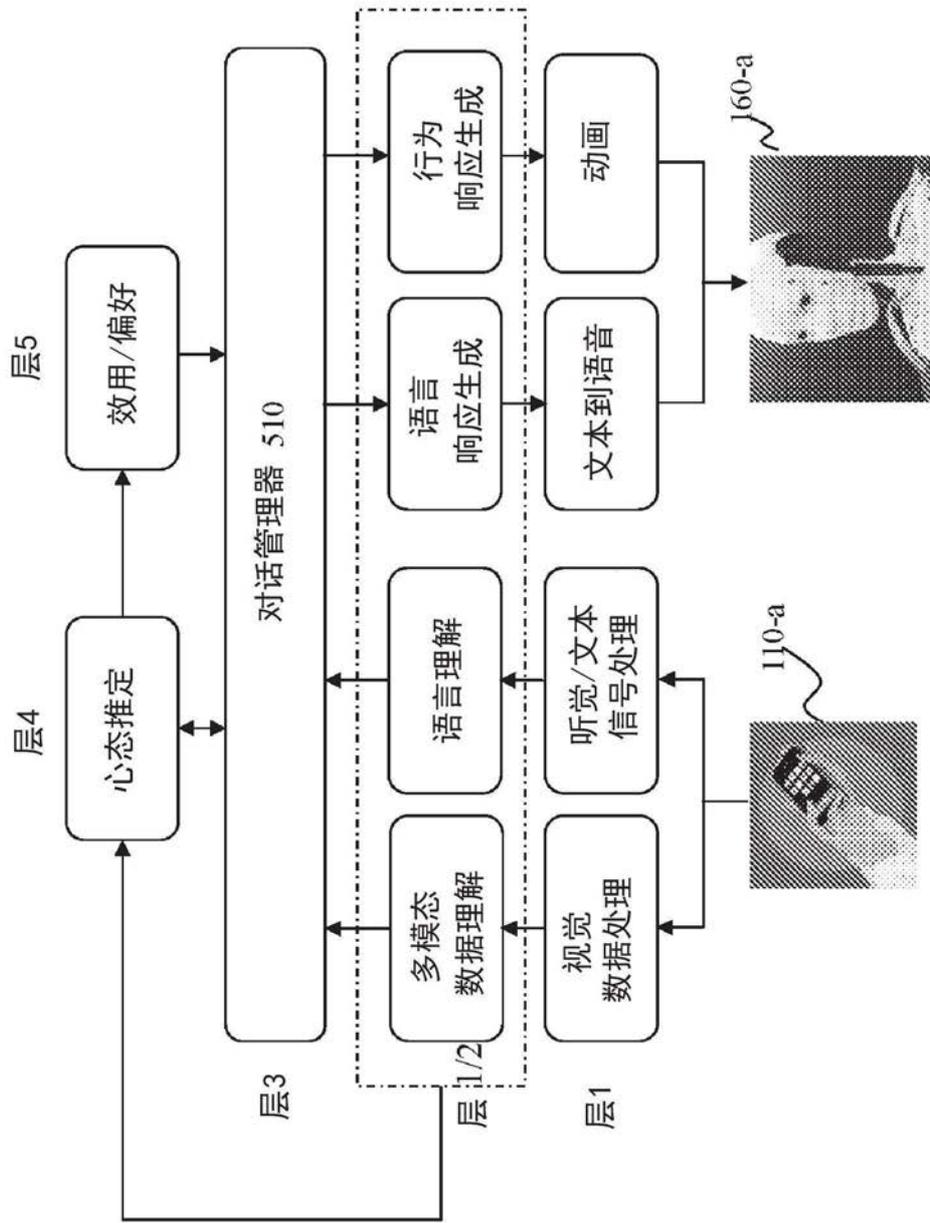


图5

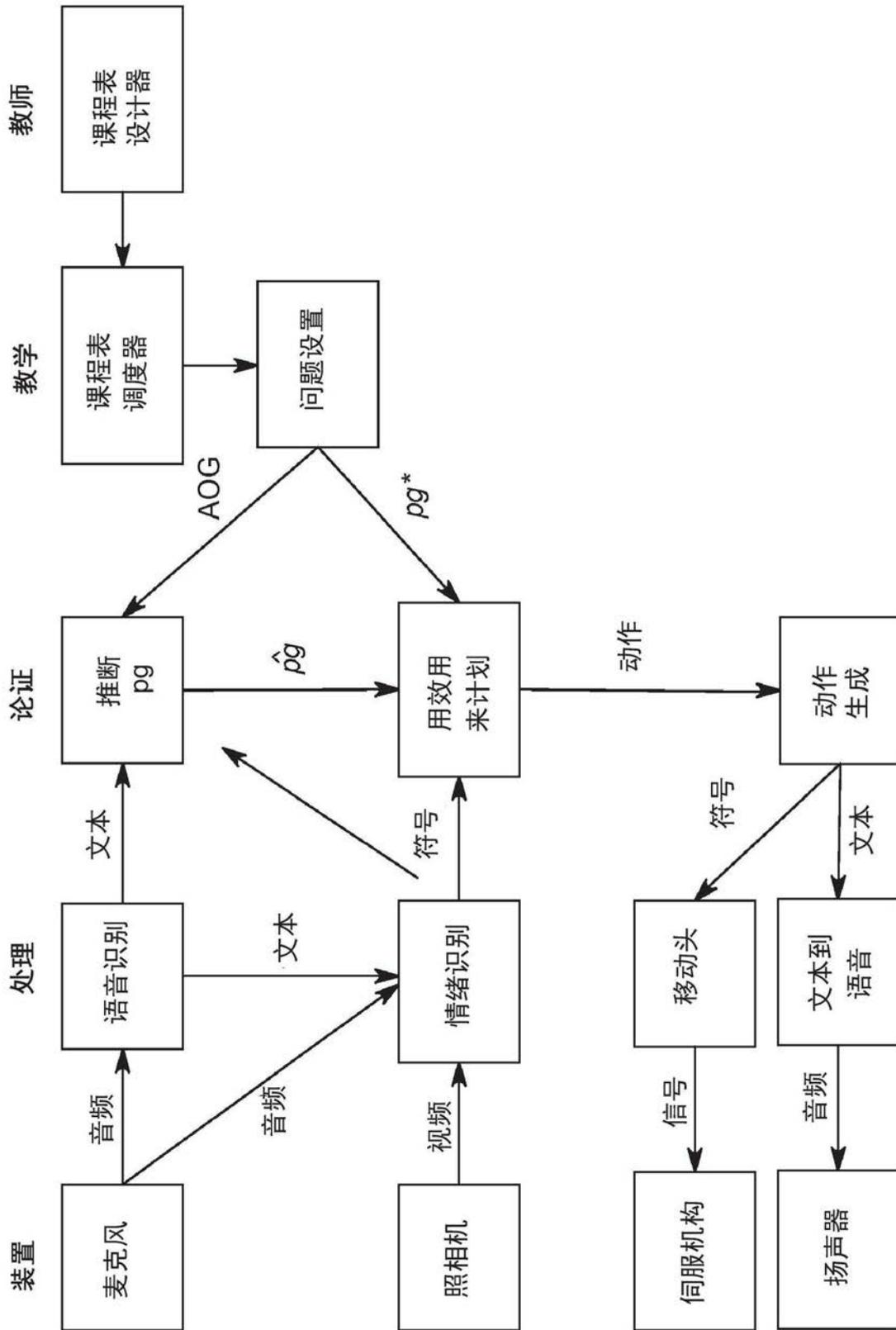


图6

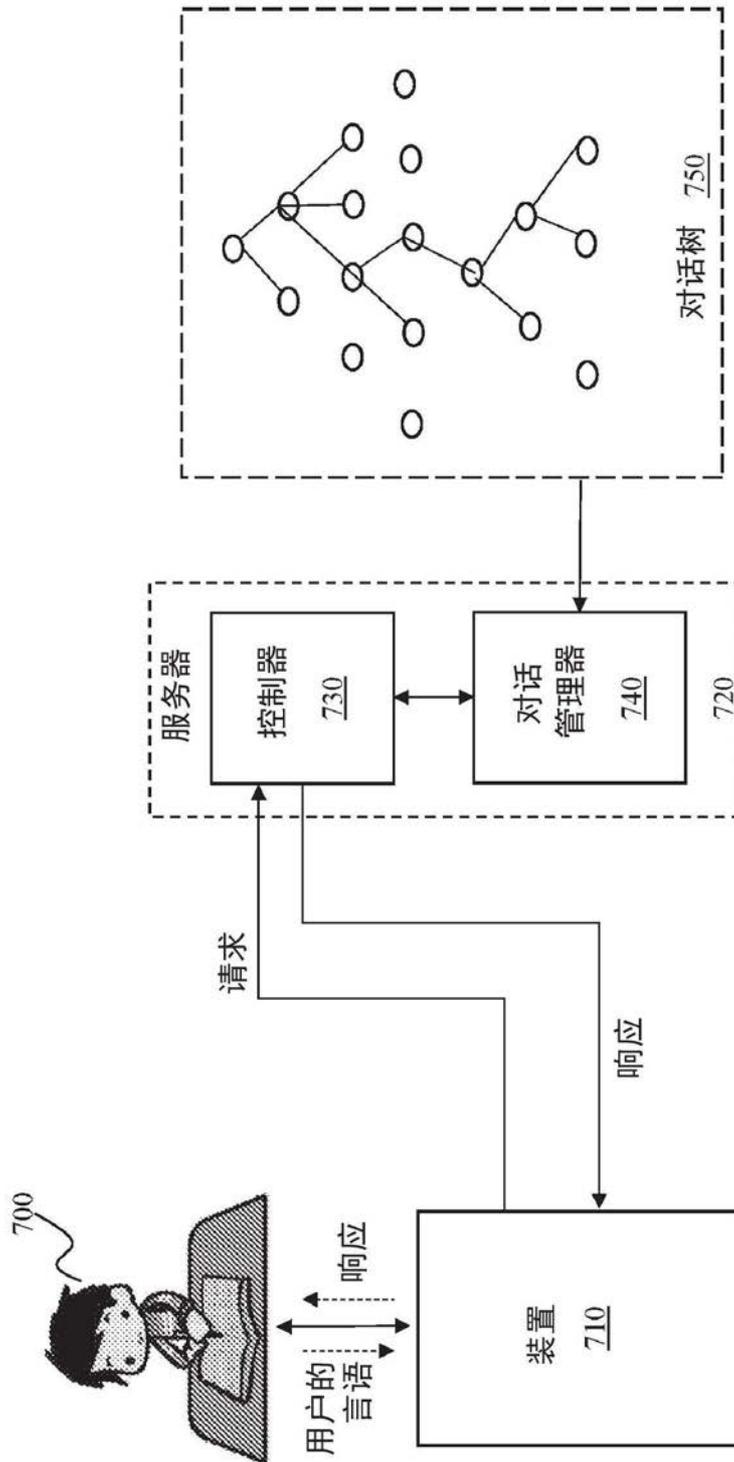


图7

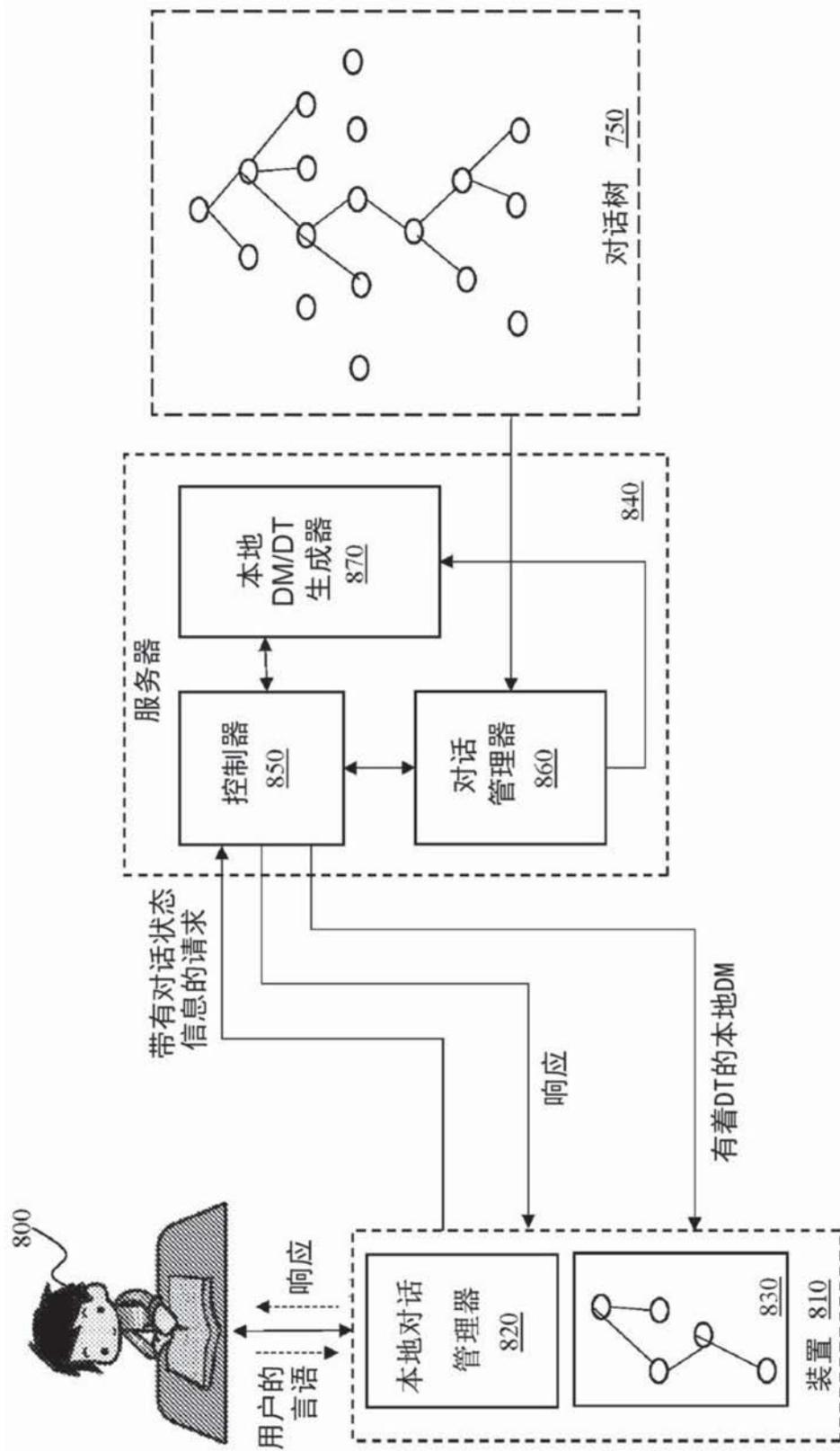


图8

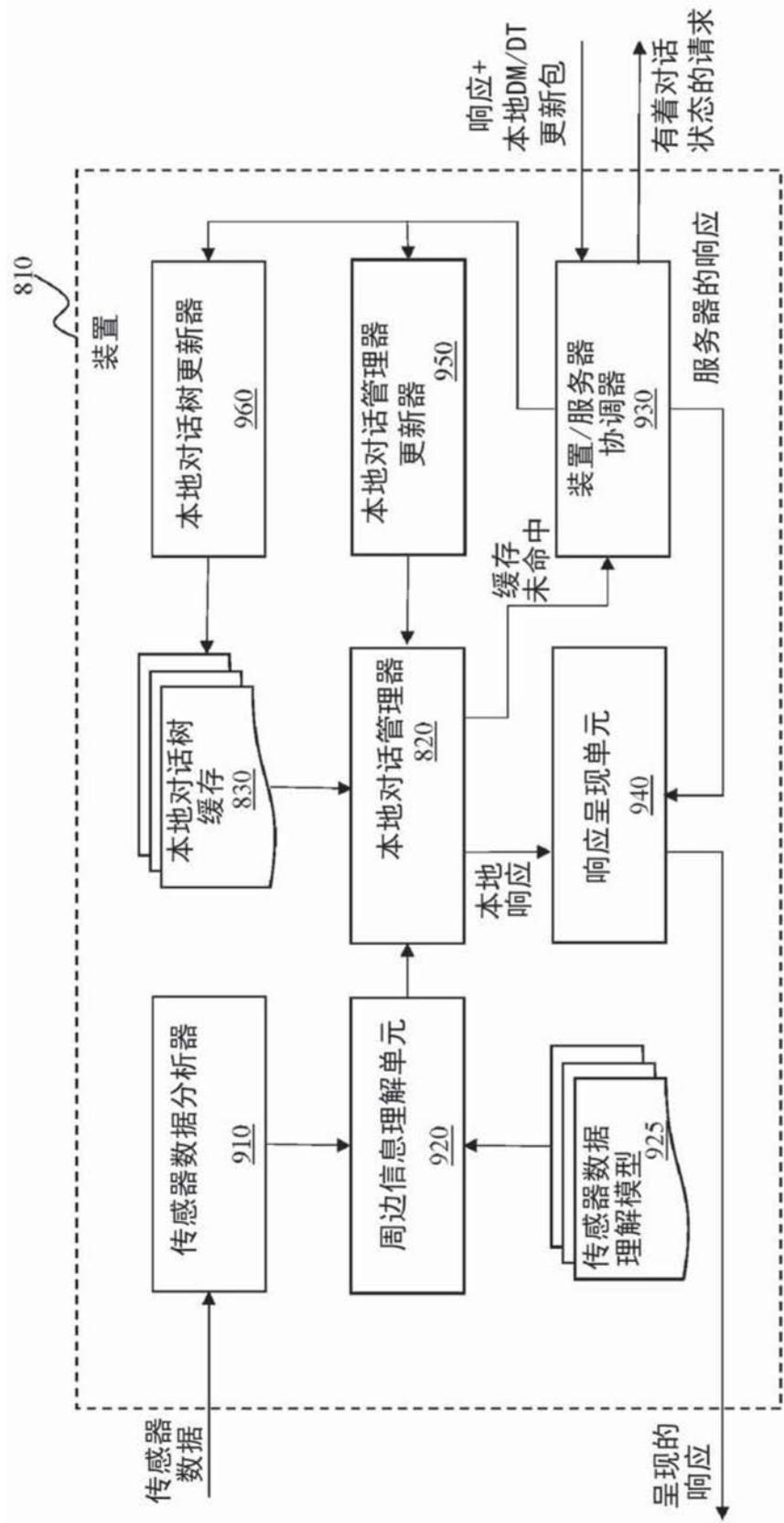


图9

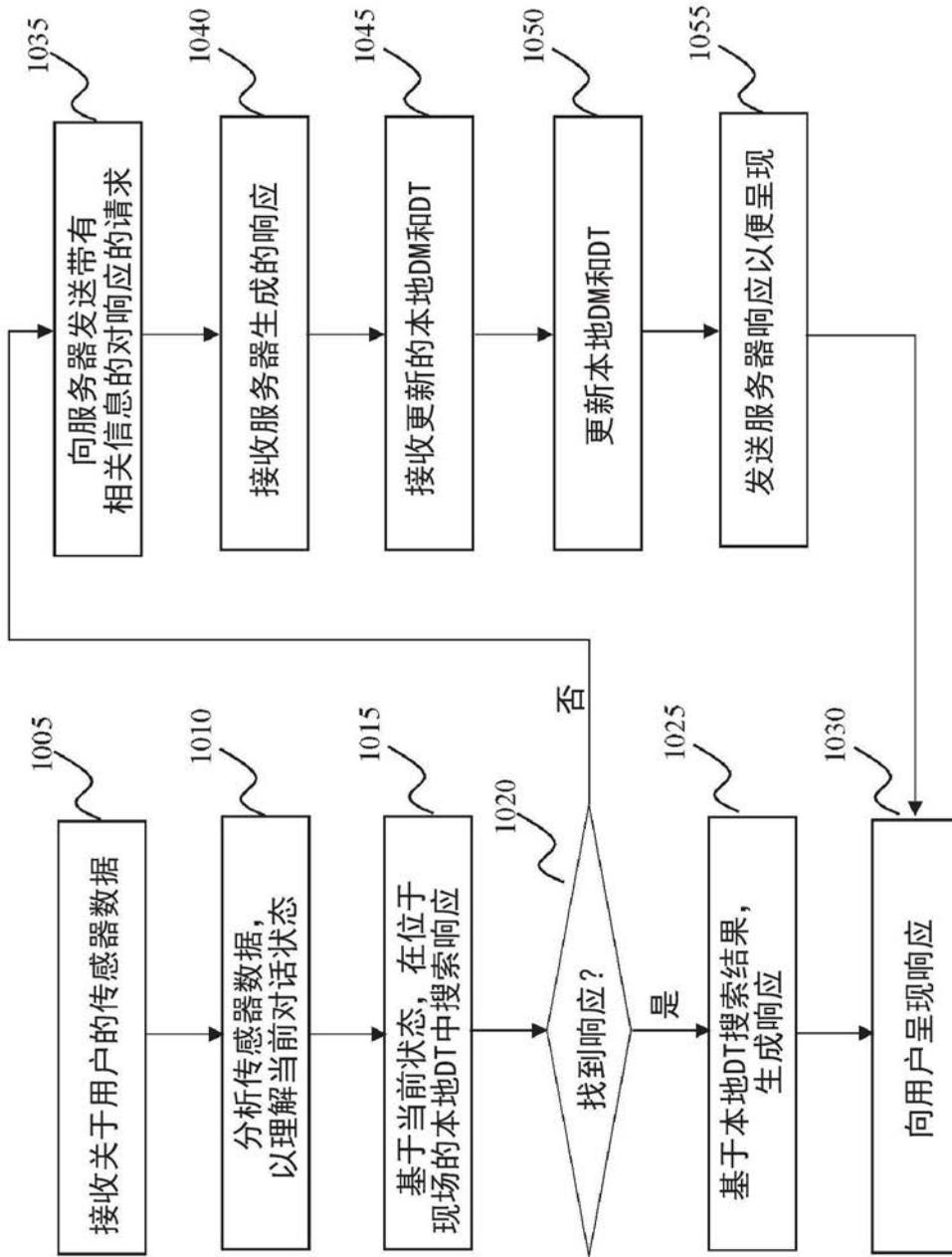


图10

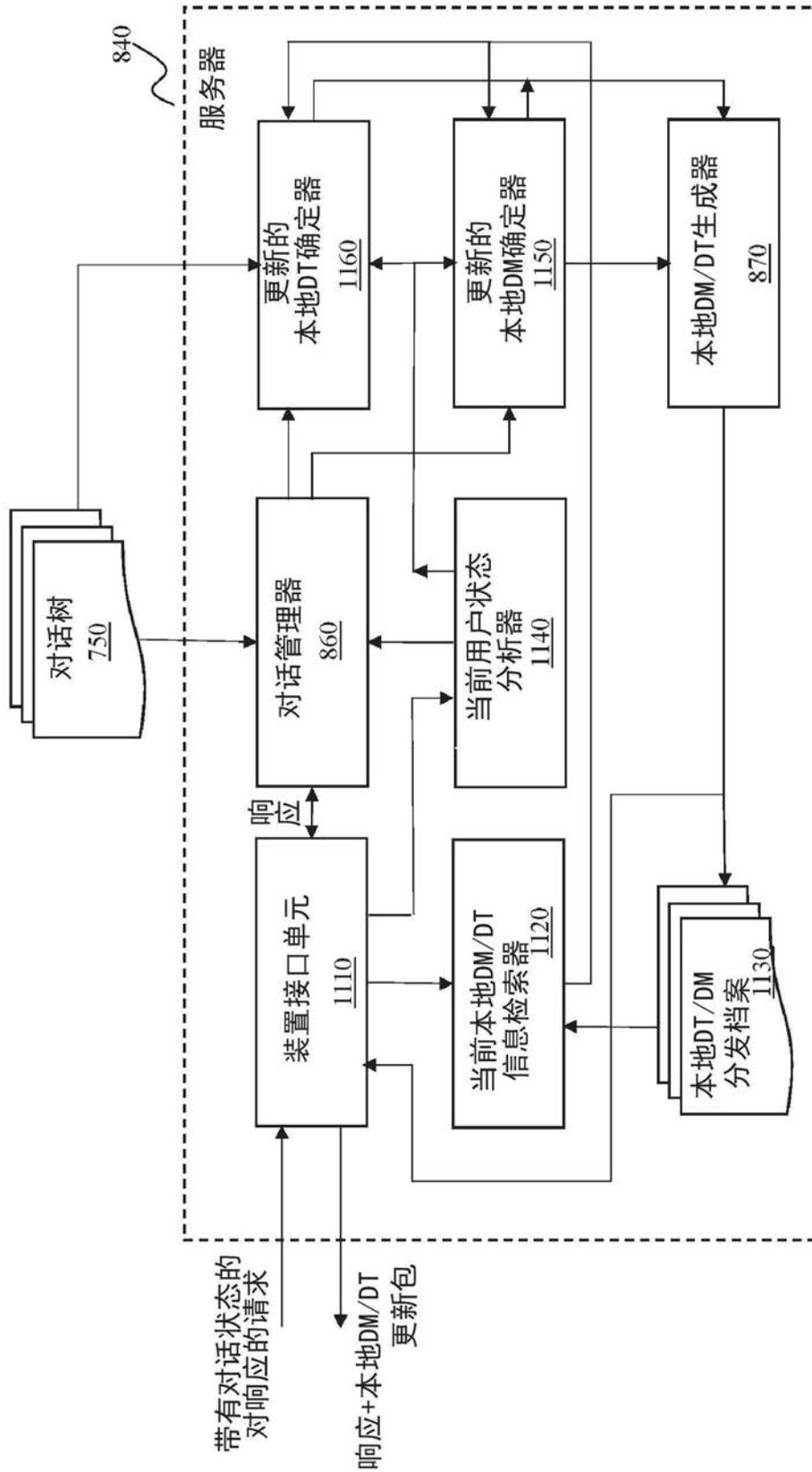


图11

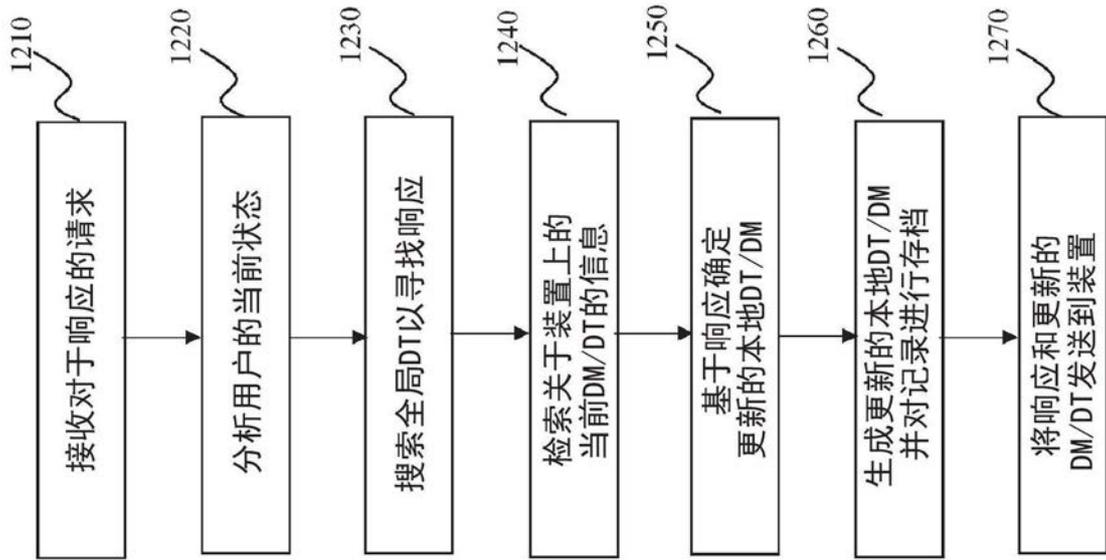


图12

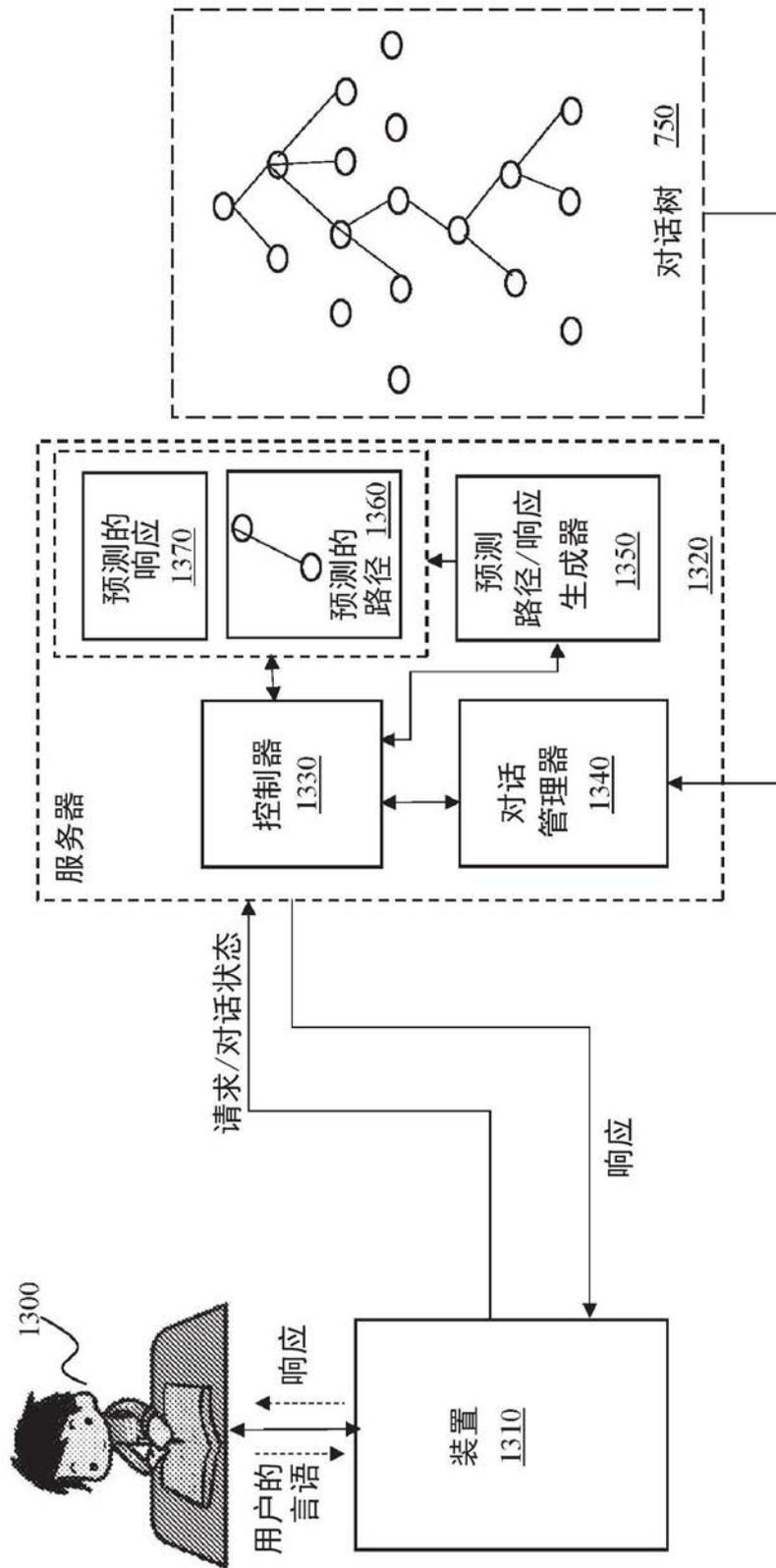


图13

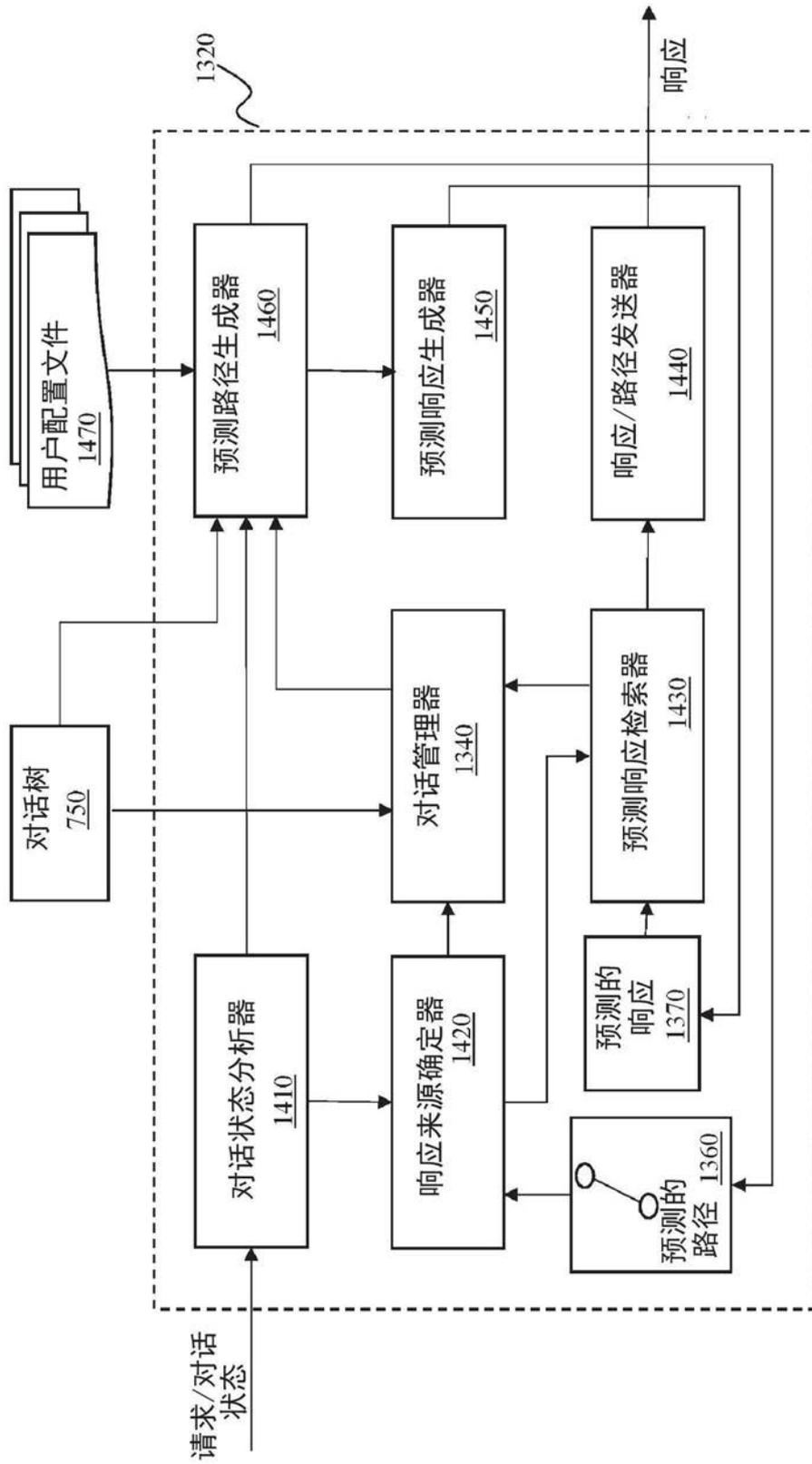


图14

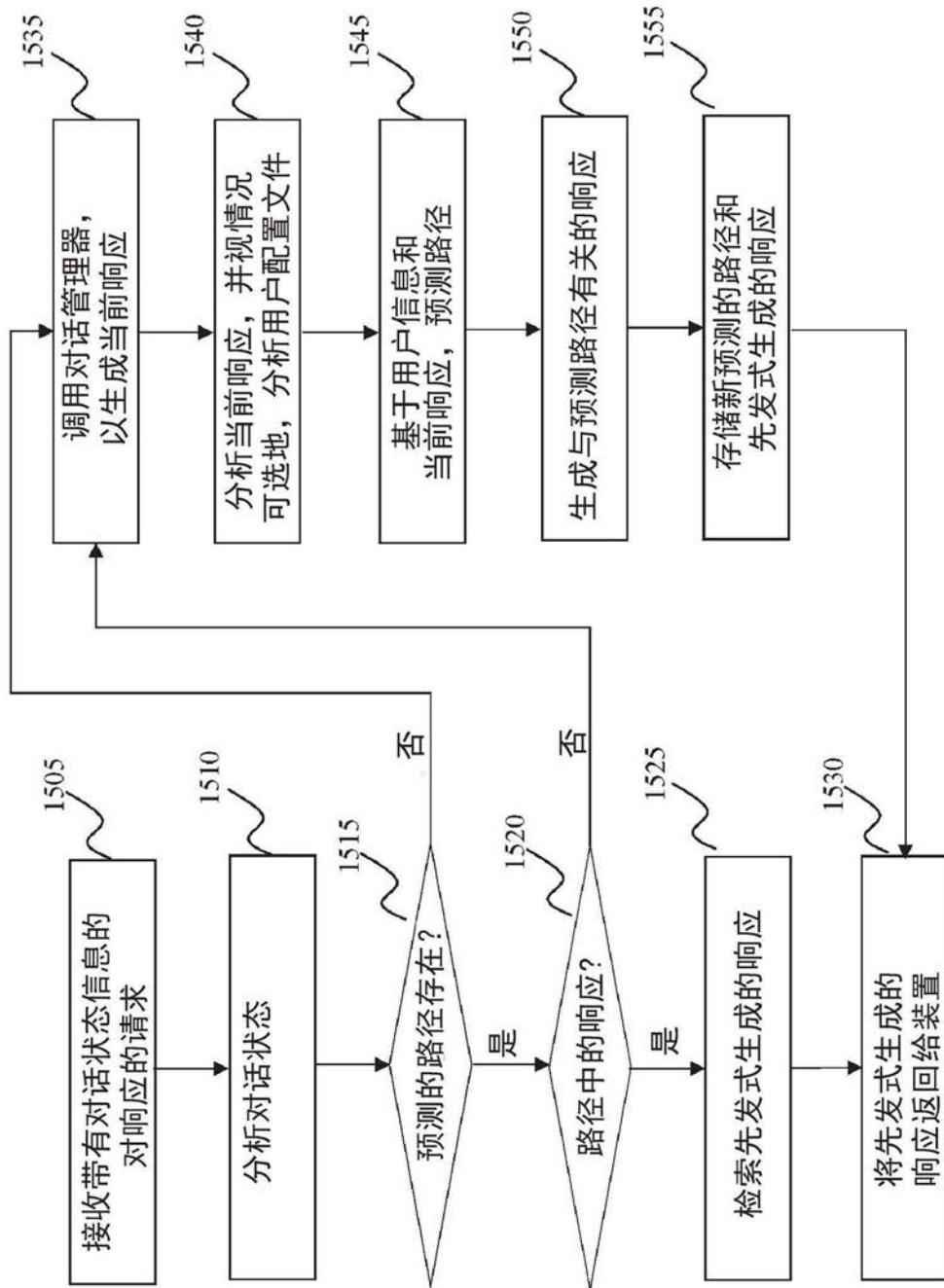


图15

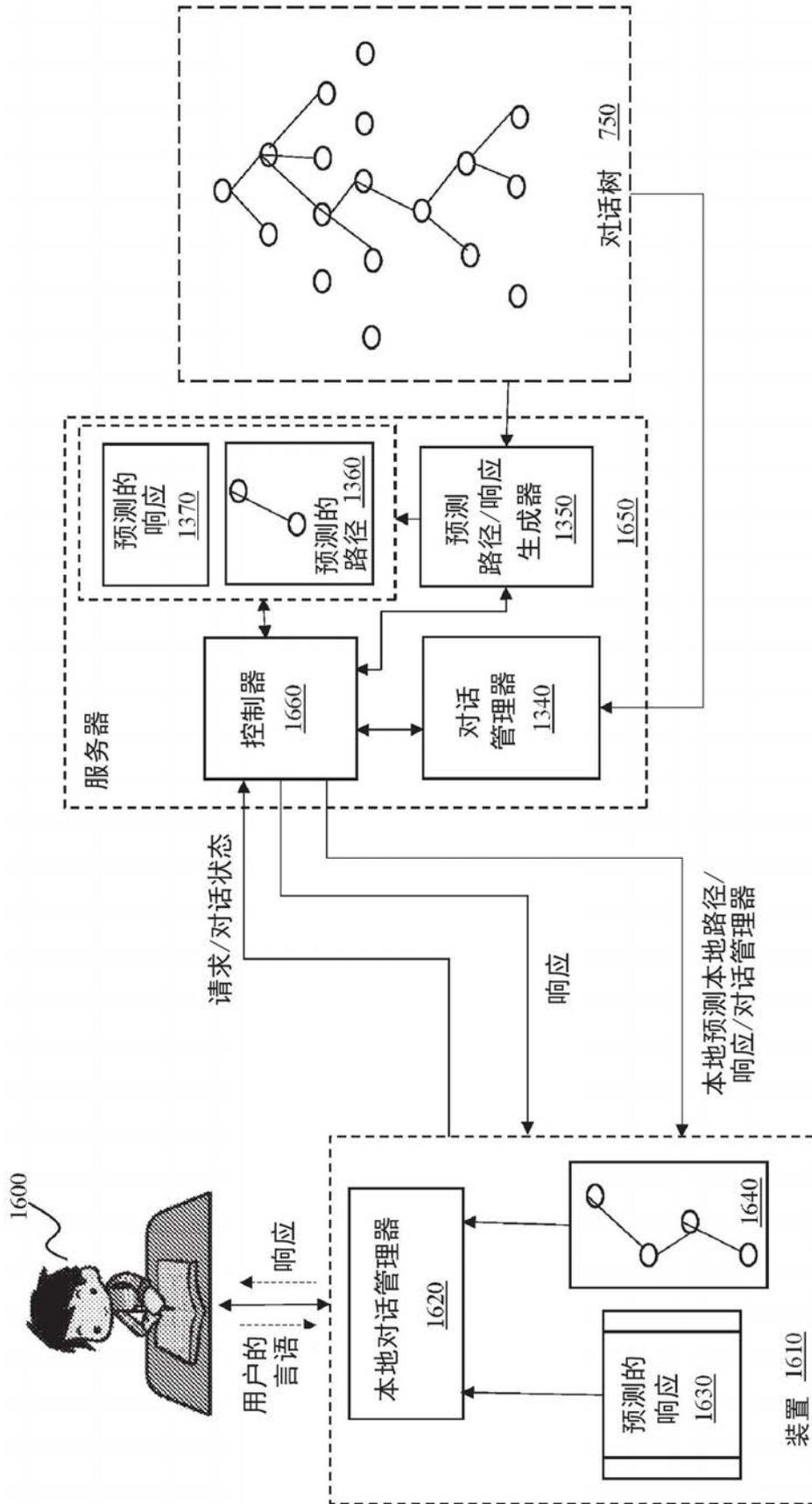


图16

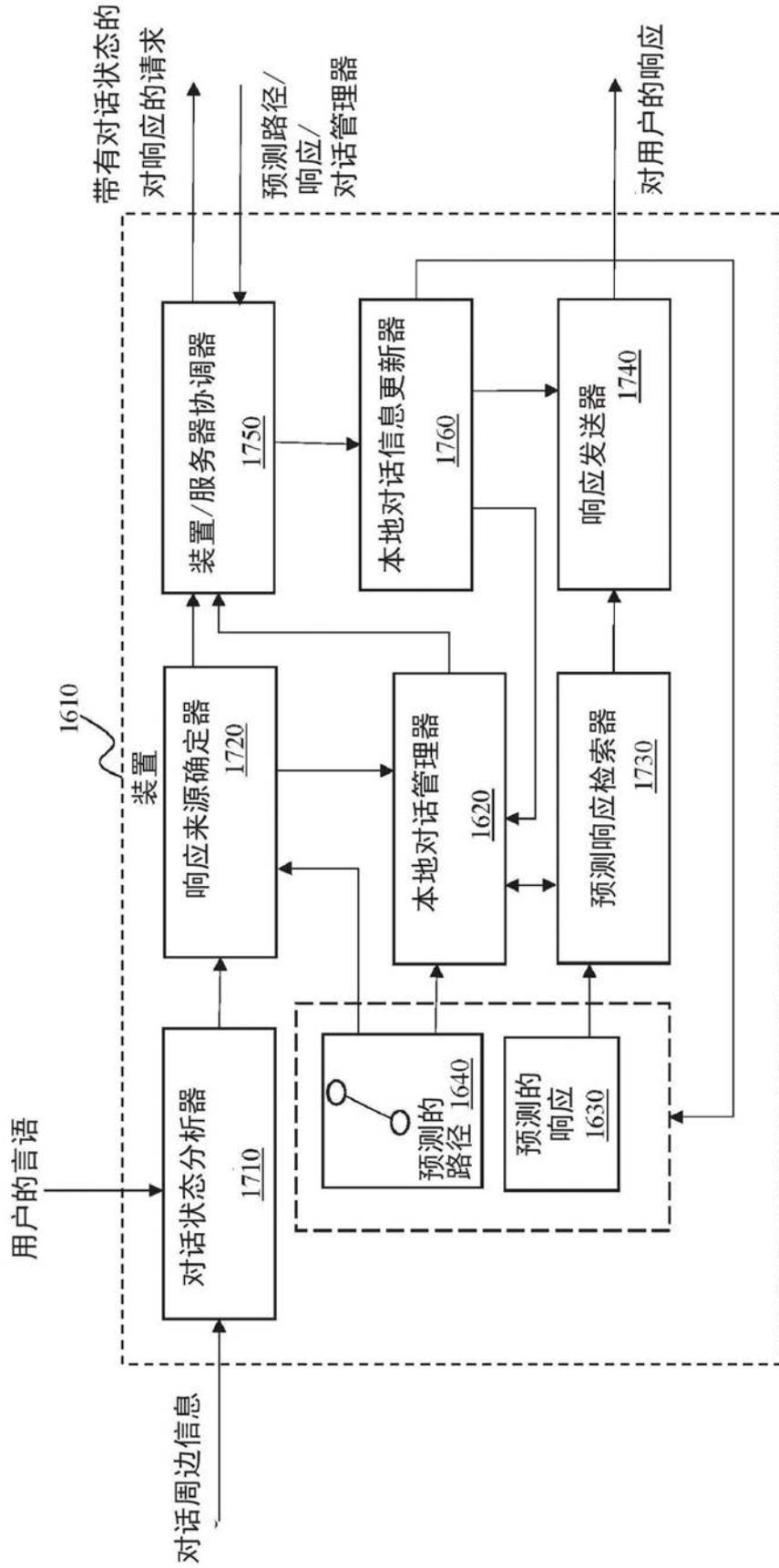


图17

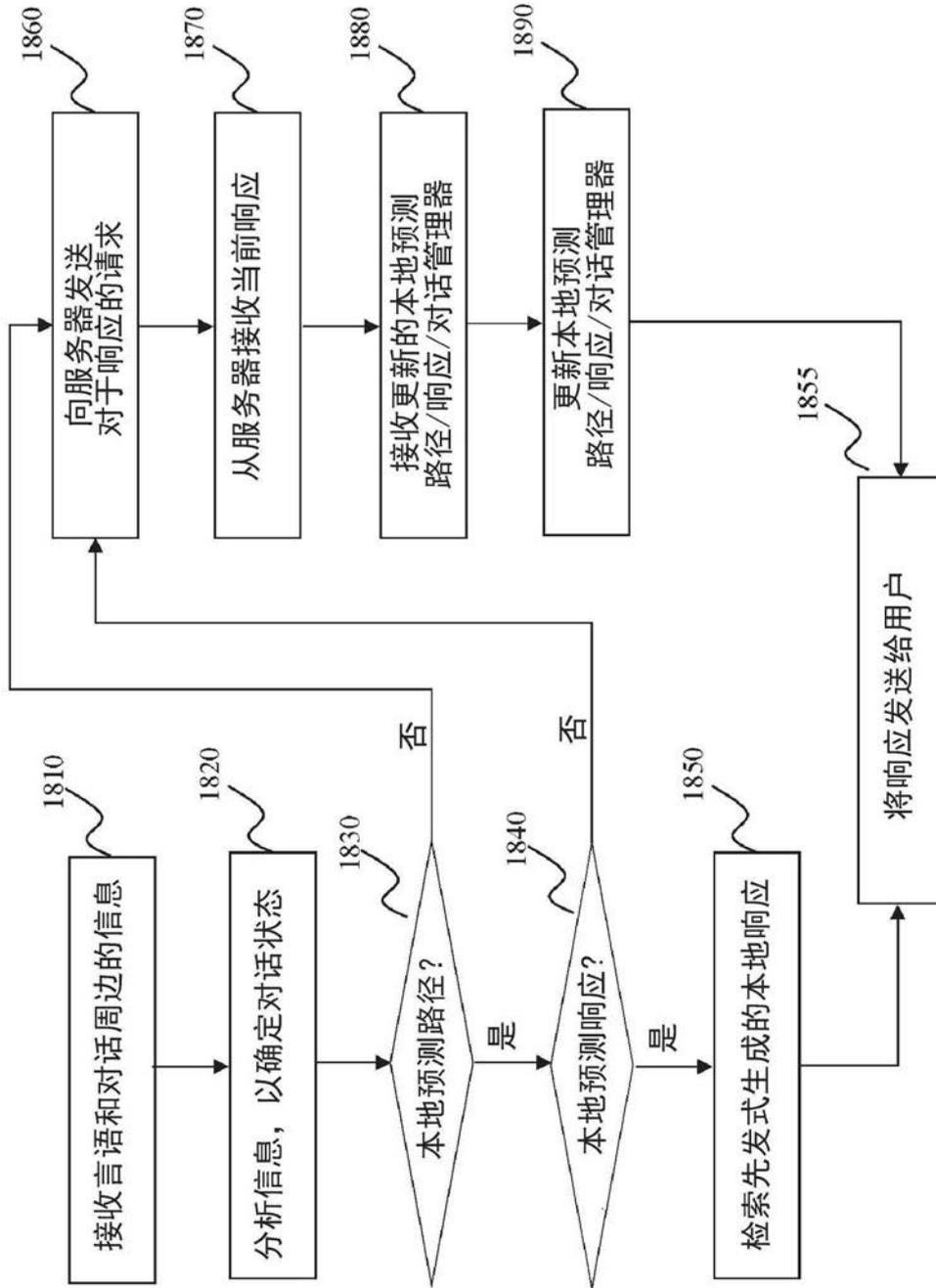


图18

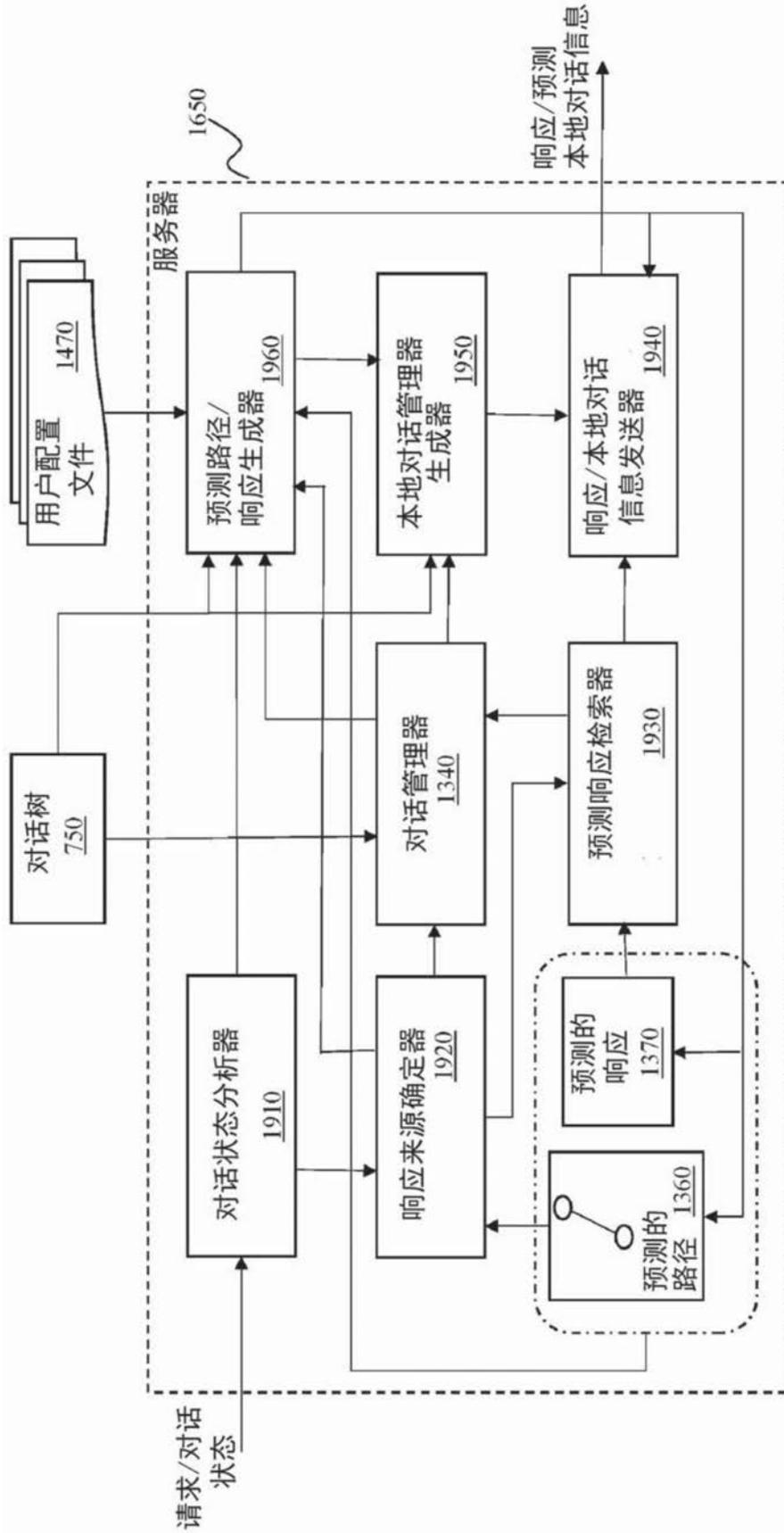


图19

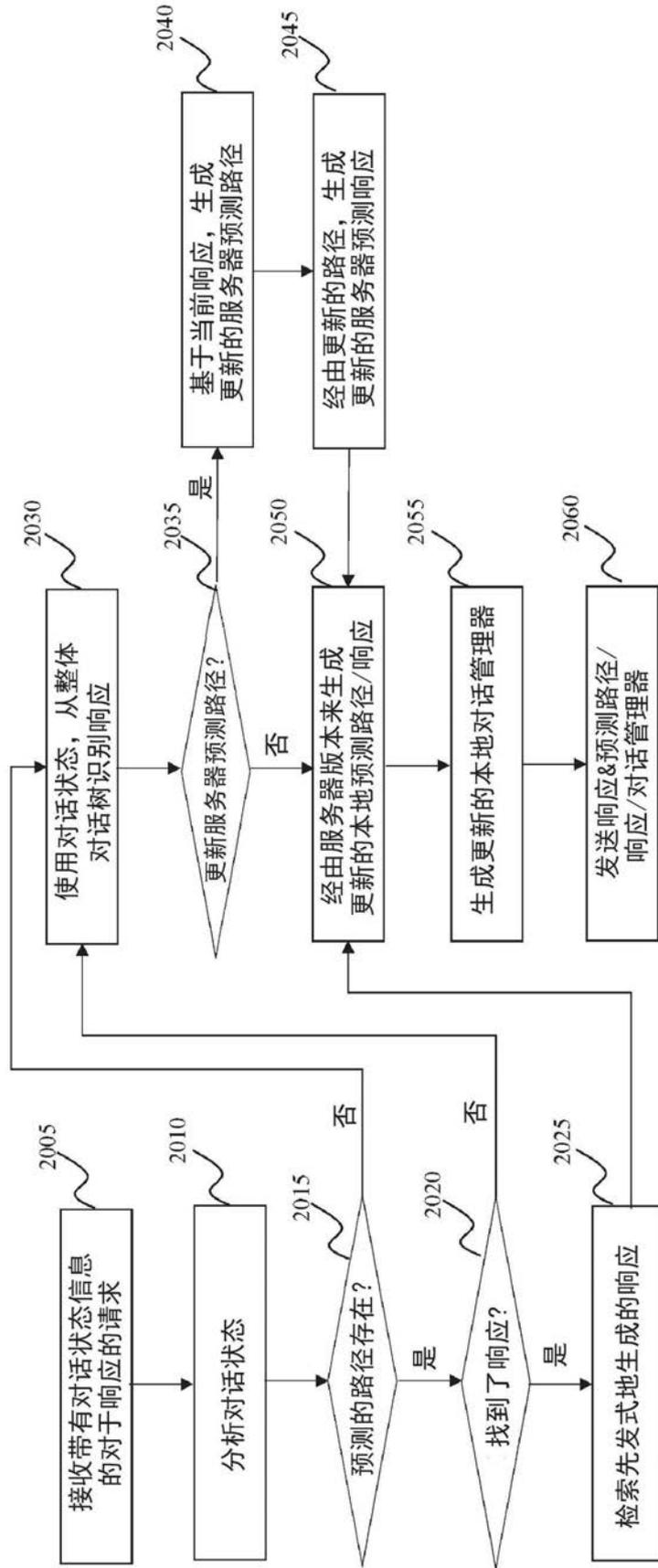


图20

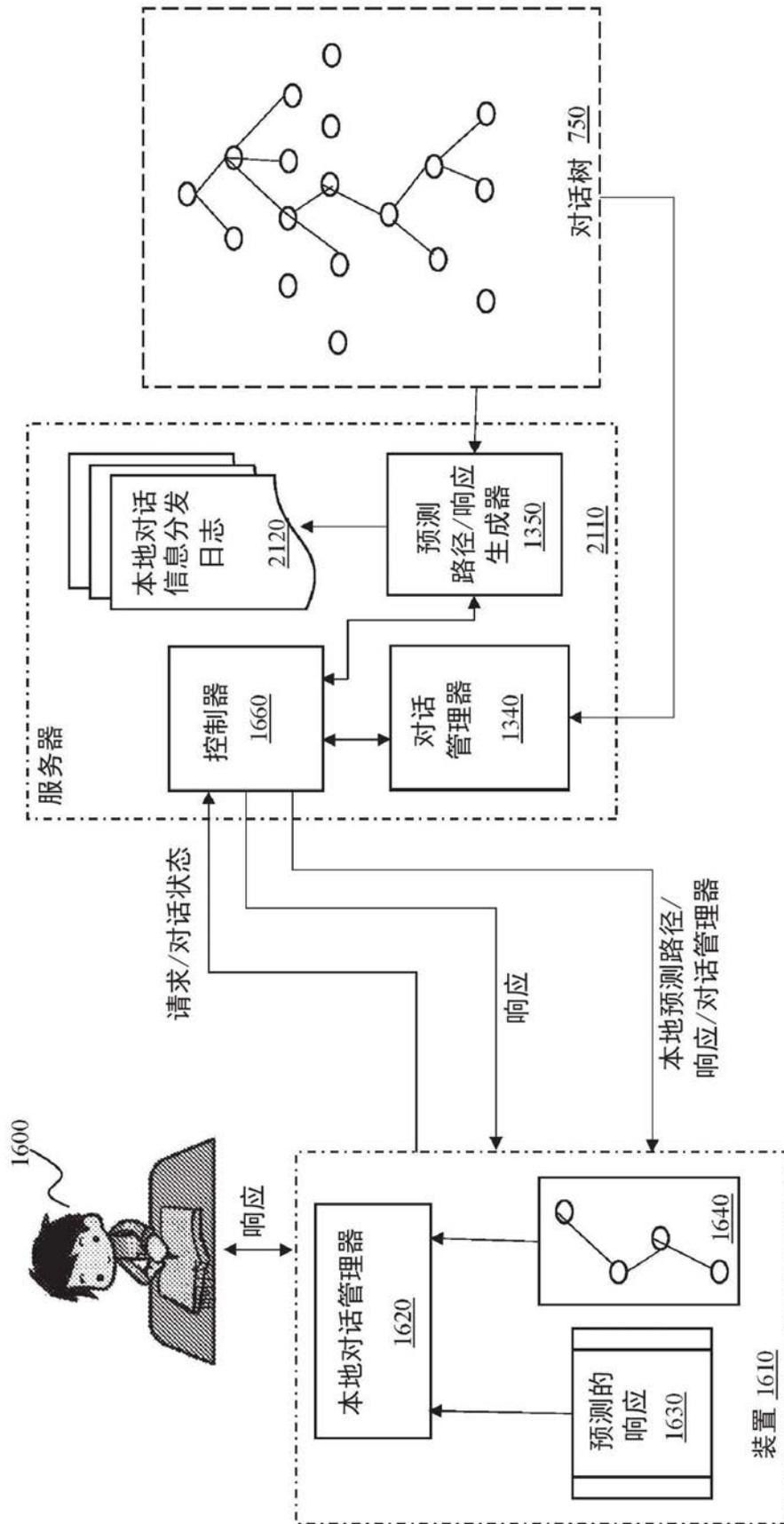


图21

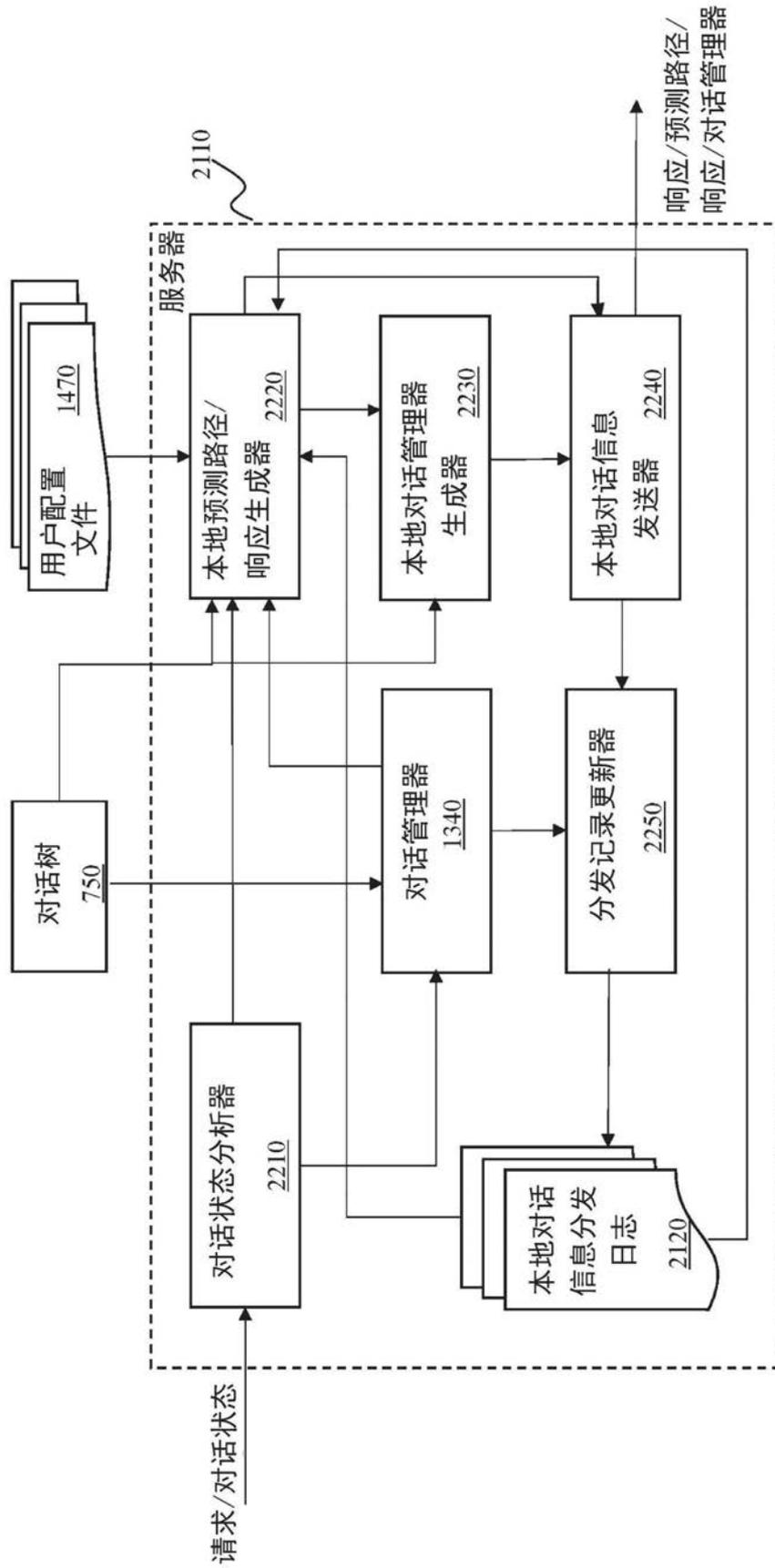


图22

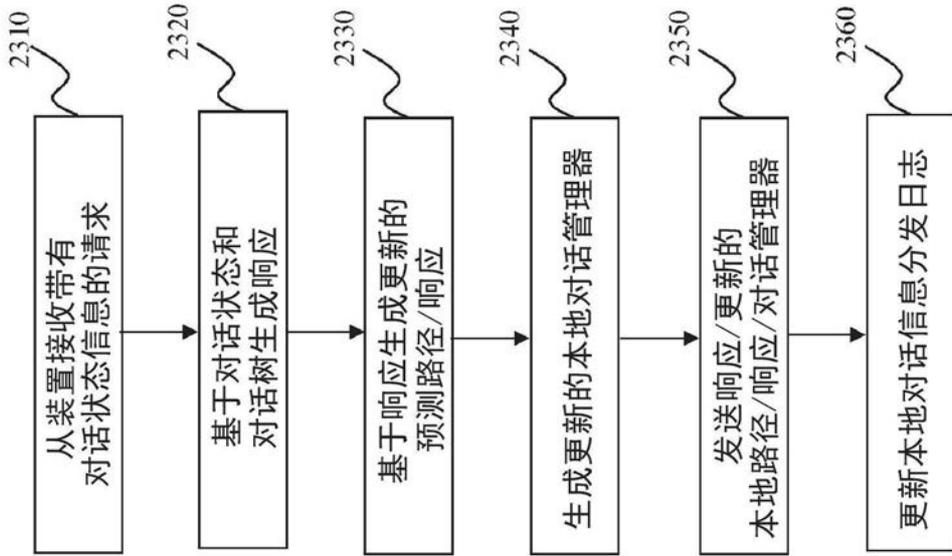
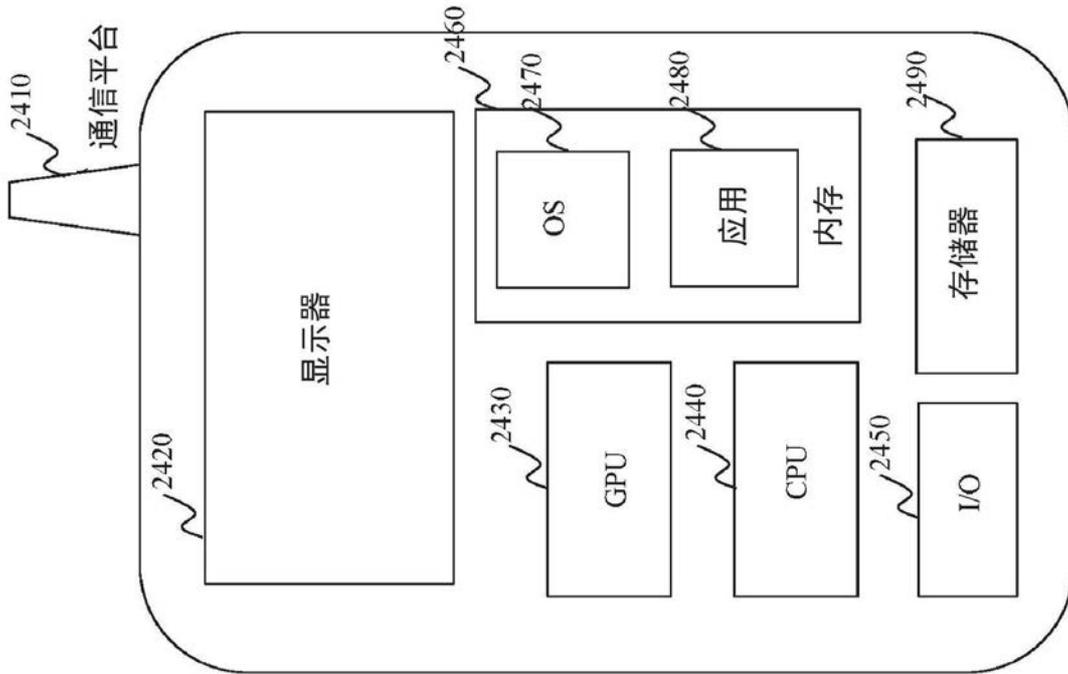
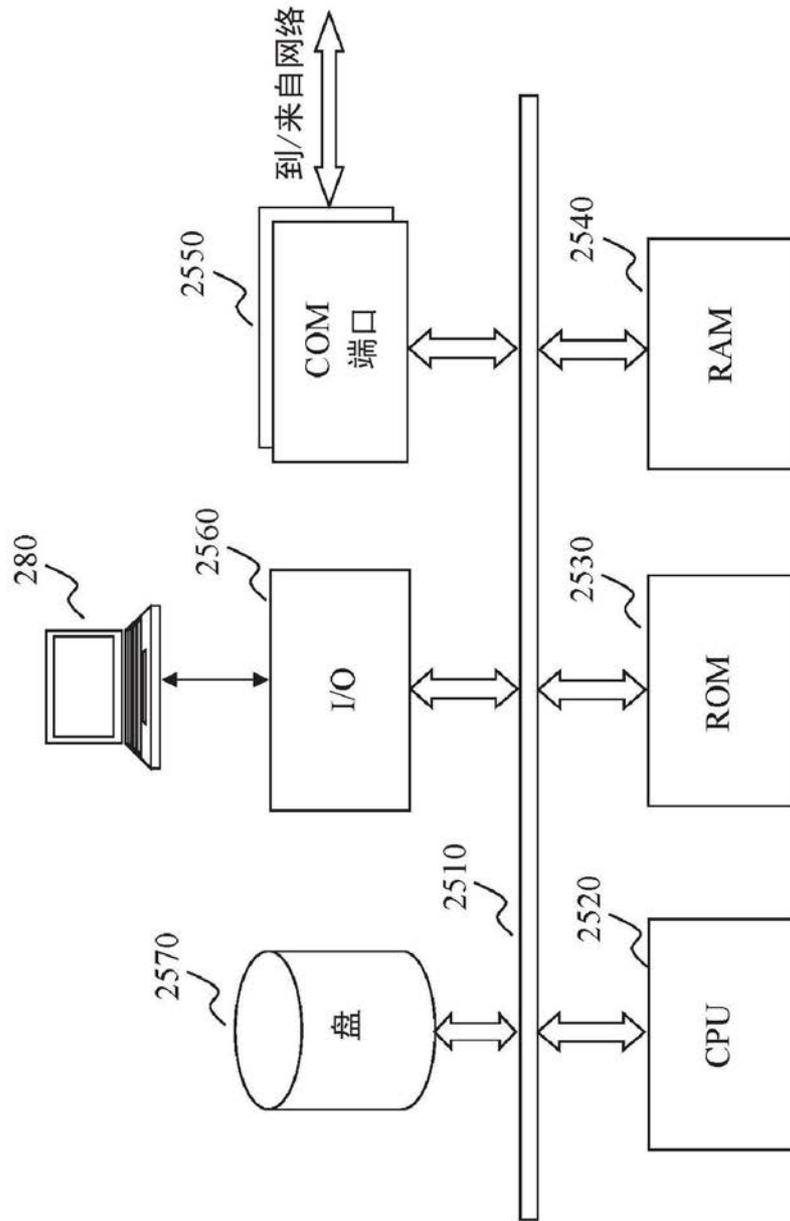


图23



2400

图24



2500

图25