



(12) 发明专利

(10) 授权公告号 CN 111443801 B

(45) 授权公告日 2023.10.13

(21) 申请号 202010220089.8

JP 2013099823 A, 2013.05.23

(22) 申请日 2020.03.25

US 2013335310 A1, 2013.12.19

(65) 同一申请的已公布的文献号

CN 110310631 A, 2019.10.08

申请公布号 CN 111443801 A

CN 108920640 A, 2018.11.30

(43) 申请公布日 2020.07.24

CN 110299152 A, 2019.10.01

(73) 专利权人 北京百度网讯科技有限公司

CN 109767774 A, 2019.05.17

地址 100085 北京市海淀区上地十街10号

CN 110765294 A, 2020.02.07

百度大厦2层

CN 107831903 A, 2018.03.23

(72) 发明人 李扬 李士岩

CN 109166575 A, 2019.01.08

(74) 专利代理机构 北京同立钧成知识产权代理

CN 103530501 A, 2014.01.22

有限公司 11205

CN 101362334 A, 2009.02.11

专利代理师 张娜 臧建明

CN 107016949 A, 2017.08.04

(51) Int. Cl.

CN 108063803 A, 2018.05.22

G06F 3/01 (2006.01)

CN 109166616 A, 2019.01.08

G06V 40/16 (2022.01)

CN 110689889 A, 2020.01.14

G10L 15/22 (2006.01)

US 10475106 B1, 2019.11.12

(56) 对比文件

张诗潮等. 体感技术现状和发展研究.《华东师范大学学报(自然科学版)》.2014, (第第2期), 全文.

CN 108235745 A, 2018.06.29

Sejima, Y.A Speech-Driven Embodied Communication System Based on an Eye Gaze Model in Interaction-Activated Communication.《LECTURE NOTES IN ARTIFICIAL INTELLIGENCE》.2017, 全文.

CN 110716634 A, 2020.01.21

CN 107103906 A, 2017.08.29

CN 110853619 A, 2020.02.28

JP 2004192653 A, 2004.07.08

CN 109992237 A, 2019.07.09

US 2016026342 A1, 2016.01.28

审查员 蒲韵霜

权利要求书2页 说明书10页 附图4页

(54) 发明名称

人机交互方法、装置、设备及存储介质

(57) 摘要

本申请公开了一种人机交互方法、装置、设备及存储介质,涉及智能语音识别技术领域。该方法包括:从目标用户的图像中获取目标用户的面部信息,该面部信息包括目标用户的面部朝向以及嘴唇状态;根据目标用户的面部信息,确定是否对目标用户进行响应。本申请实施例的技术方案,不需要在每轮对话的开始通过唤醒词进行唤醒,从而能够实现数字人与目标用户的高效自然的交流。

从目标用户的图像中获取目标用户的面部信息,该面部信息包括目标用户的面部朝向以及嘴唇状态 S210

根据目标用户的面部信息,确定是否对目标用户进行响应 S220

CN 111443801 B

1. 一种人机交互方法,其特征在于,包括:

从目标用户的图像中获取所述目标用户的面部信息,所述面部信息包括所述目标用户的面部朝向以及嘴唇状态;

若确定所述目标用户的视线方向为预定方向,并且所述目标用户的嘴唇状态为唇动状态,则确定通过虚拟对象对所述目标用户进行反馈,所述预定方向为虚拟对象的正面方向,所述虚拟对象为数字人;

所述方法还包括:

从所述目标用户的图像中获取所述目标用户的当前状态,所述当前状态包括:所述目标用户的动作、表情、环境中的一种或多种;

根据所述目标用户的当前状态,确定对所述目标用户的反馈优先级,所述反馈优先级用于表示对目标用户进行反馈的先后顺序;

基于所述反馈优先级,对所述目标用户进行响应;

所述对所述目标用户进行响应,包括:

获取所述目标用户的语音信息;

根据所述当前状态和所述语音信息,通过虚拟对象对所述目标用户进行反馈操作,所述反馈操作包括语音反馈,以及动作反馈和表情反馈中的至少一种;

或者,

根据所述当前状态以及所述语音信息,调用第三方应用程序对所述目标用户进行响应。

2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

若确定对所述目标用户进行响应,则获取所述目标用户发出的语音信息;

基于所述语音信息,对所述目标用户进行语音反馈。

3. 根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

接收所述目标用户通过图形用户界面输入的交互指令,根据所述交互指令对所述目标用户进行反馈。

4. 一种人机交互装置,其特征在于,包括:

信息获取模块,用于从目标用户的图像中获取所述目标用户的面部信息,所述面部信息包括所述目标用户的面部朝向以及嘴唇状态;

响应判断模块,用于若确定所述目标用户的视线方向为预定方向,并且所述目标用户的嘴唇状态为唇动状态,则确定通过虚拟对象对所述目标用户进行反馈,所述预定方向为虚拟对象的正面方向,所述虚拟对象为数字人;

状态获取模块,用于从所述目标用户的图像中获取所述目标用户的当前状态,所述当前状态包括:所述目标用户的动作、表情、环境中的一种或多种;

第一响应模块,用于根据所述目标用户的当前状态,确定对所述目标用户的反馈优先级,基于所述反馈优先级,对所述目标用户进行响应;所述反馈优先级用于表示对目标用户进行反馈的先后顺序;

所述第一响应模块,具体用于:

获取所述目标用户的语音信息;

根据所述当前状态和所述语音信息,通过虚拟对象对所述目标用户进行反馈操作,所

述反馈操作包括语音反馈,以及动作反馈和表情反馈中的至少一种;

或者,

根据所述当前状态以及所述语音信息,调用第三方应用程序对所述目标用户进行响应。

5. 根据权利要求4所述的装置,其特征在于,所述装置还包括:

语音信息获取模块,用于若确定对所述目标用户进行响应,则获取所述目标用户发出的语音信息;

第二响应模块,用于基于所述语音信息,对所述目标用户进行语音反馈。

6. 根据权利要求4或5所述的装置,其特征在于,所述装置还包括:

第三响应模块,用于接收所述目标用户通过图形用户界面输入的交互指令,根据所述交互指令对所述目标用户进行反馈。

7. 一种电子设备,其特征在于,包括:至少一个处理器,存储器,以及与其他电子设备通信的接口;所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1至3中任一项所述的人机交互方法。

8. 一种存储有计算机指令的非瞬时计算机可读存储介质,其特征在于,所述计算机指令用于使所述计算机执行权利要求1至3中任一项所述的人机交互方法。

人机交互方法、装置、设备及存储介质

技术领域

[0001] 本申请涉及智能语音识别技术领域,尤其涉及一种人机交互方法、装置、设备及存储介质。

背景技术

[0002] 随着人机交互技术的发展,智能语音设备例如智能音箱的应用越来越广泛,如何高效地进行人与机器的语音交互成为了关注的焦点。

[0003] 目前,智能语音设备均通过唤醒词唤醒进行交互,例如,通过预设唤醒词唤醒智能语音设备,再与智能语音设备进行语音交互。然而,在这种技术方案中,在多轮的语音交互过程中,每轮语音交互都需说唤醒词,因此,通过唤醒词唤醒的方式交互流程比较繁琐,不能达到像人与人之间的自然交互。

[0004] 因此,如何高效自然地与智能语音设备进行语音交互成为了亟待解决的技术难题。

发明内容

[0005] 本申请实施例提供一种人机交互方法、装置、设备及存储介质,用于解决如何高效自然地与智能语音设备进行语音交互的问题。

[0006] 第一方面,本申请提供一种人机交互方法,包括:

[0007] 从目标用户的图像中获取所述目标用户的面部信息,所述面部信息包括所述目标用户的面部朝向以及嘴唇状态;

[0008] 根据所述目标用户的面部信息,确定是否对所述目标用户进行响应。

[0009] 第二方面,本申请提供了一种人机交互装置,包括:

[0010] 信息获取模块,用于从目标用户的图像中获取所述目标用户的面部信息,所述面部信息包括所述目标用户的面部朝向以及嘴唇状态;

[0011] 响应判断模块,用于根据所述目标用户的面部信息,确定是否对所述目标用户进行响应。

[0012] 第三方面,本申请提供一种电子设备,包括:至少一个处理器,存储器,以及与其他电子设备通信的接口;所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行第一方面任一项所述的人机交互方法。

[0013] 第四方面,本申请提供一种存储有计算机指令的非瞬时计算机可读存储介质,所述计算机指令用于使所述计算机第一方面中任一项所述的人机交互方法。

[0014] 上述申请中的一个实施例具有如下优点或有益效果:一方面,根据目标用户的面部朝向以及嘴唇状态,确定是否对目标用户进行响应,能够高效准确地确定是否触发与目标用户的交互;另一方面,由于不需要在每轮对话的开始通过唤醒词进行唤醒,从而能够实现数字人与目标用户的高效自然的交流,提高用户的服务体验。

[0015] 上述可选方式所具有的其他效果将在下文中结合具体实施例加以说明。

附图说明

- [0016] 附图用于更好地理解本方案,不构成对本申请的限定。其中:
- [0017] 图1为根据本申请的一些实施例提供的人机交互方法的应用场景的示意图;
- [0018] 图2为根据本申请的一些实施例提供的人机交互方法的流程示意图;
- [0019] 图3为根据本申请的另一一些实施例提供的人机交互方法的流程示意图;
- [0020] 图4为根据本申请的又一些实施例提供的人机交互方法的流程示意图;
- [0021] 图5为根据本申请的一些实施例提供的人机交互装置的示意框图;
- [0022] 图6为根据本申请的一些实施例提供的响应判断模块的示意框图;
- [0023] 图7为根据本申请的一些实施例提供的人机交互装置的示意框图;
- [0024] 图8为用来实现本申请实施例的人机交互方法的电子设备的框图。

具体实施方式

[0025] 以下结合附图对本申请的示范性实施例做出说明,其中包括本申请实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本申请的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0026] 目前,智能语音设备均通过唤醒词唤醒进行交互,例如,通过预设唤醒词唤醒智能语音设备,再与智能语音设备进行语音交互。然而,在这种技术方案中,存在以下两个问题:(1)在多轮的语音交互过程中,每轮语音交互都需说唤醒词,因此,通过唤醒词唤醒的方式交互流程比较繁琐,不能达到像人与人之间的自然交互。(2)现有的智能语音模块只能对语音内容进行简易机械理解和认知,未能做到对外部例如服务对象的认知和理解,以及不能达到像人与人的复杂性交互。

[0027] 基于上述内容,本申请的基本思想在于:结合图像识别技术与语音识别技术对目标用户进行反馈。具体而言,一方面,通过图像识别获取目标用户的面部信息,面部信息包括目标用户的面部朝向以及嘴唇状态,根据面部信息确定是否对目标用户进行反馈;另一方面,若确定对目标用户进行反馈,则识别目标用户的当前状态例如动作、表情、环境等,根据目标用户的当前状态和/或语音信息对目标用户进行反馈。根据本申请实施例的技术方案,一方面,根据目标用户的面部朝向以及嘴唇状态,确定是否对目标用户进行反馈,能够实现数字人与目标用户的自然交流,不需要唤醒词进行唤醒;另一方面,根据目标用户的当前状态和/或语音信息进行反馈,能够根据对目标用户的当前状态的认知和理解进行反馈,从而能够实现像人与人之间的高效自然的交互。

[0028] 下面对本申请中涉及的名词等进行解释:

[0029] 当前状态:用于表示目标用户例如目标人物对象当前的动作、表情、周围环境的状态。

[0030] 反馈:用于表示虚拟对象例如数字人对目标人物对象的语言、肢体动作或表情动作中的一种或多种反馈。

[0031] 反馈优先级:用于表示对目标用户进行反馈的先后顺序,反馈优先级越高,则反馈

时间越早,例如,反馈优先级包括紧急处理等级、快速处理等级、普通处理等级。

[0032] 数字人:是数字角色技术与人工智能技术的结晶,人像建模、动作捕捉等数字角色技术为数字人带来了生动自然的形象表达,语音识别、自然语音理解、对话列举等人工智能技术为数字人带来了完善的认知、理解、表达能力。

[0033] 下面通过具体的实施方式对本申请提供的人机交互方法进行说明。

[0034] 图1为根据本申请的一些实施例提供的人机交互方法的应用场景的示意图。参照图1所示,该应用场景包括交互设备110,交互设备110包括图像处理模块112、语音处理模块114、反馈模块116。图像处理模块112用于拍摄装置例如摄像头获取周围环境中的目标用户120的图像,从获取的图像中提取目标用户的面部信息,根据目标用户的面部信息,确定是对目标用户进行反馈。语音处理模块114用于通过监听模块例如麦克风监听周围环境中的语音信息,若确定对目标用户进行反馈,则语音处理模块114获取目标用户发出的语音信息,根据该语音信息确定目标用户的意图,将所确定的目标用户的意图发送给反馈模块116例如数字人116。反馈模块116根据语音处理模块114发送的目标用户的意图对目标用户进行反馈。

[0035] 进一步地,图像处理模块112还可以从目标用户120的图像中获取目标用户的当前状态,例如,目标用户的动作、表情和周围环境,将目标用户的当前状态发送至反馈模块116。反馈模块116根据目标用户的当前状态和/或语音信息对目标用户进行反馈。

[0036] 下面结合图1的应用场景,参考图2来描述根据本申请的示例性实施例的人机交互方法。需要注意的是,上述应用场景仅是为了便于理解本申请的精神和原理而示出,本申请的实施例在此方面不受任何限制。相反,本申请的实施例可以应用于适用的任何场景。

[0037] 图2为根据本申请的一些实施例提供的人机交互方法的流程示意图。该人机交互方法包括步骤S210至步骤S220,该人机交互方法可以应用于人机交互设备例如图1的交互设备110,下面结合附图对示例实施例中的人机交互方法进行详细的说明。

[0038] 参照图2所示,在步骤S210中,从目标用户的图像中获取目标用户的面部信息,该面部信息包括目标用户的面部朝向以及嘴唇状态。

[0039] 在示例实施例中,通过图像获取装置例如摄像头获取目标用户的图像,从目标用户的图像中提取目标用户的面部图像,从目标用户的面部图像中获取目标用户的面部信息,目标用户的面部信息可以包括目标用户的面部朝向以及嘴唇状态。

[0040] 需要说明的是,目标用户的面部信息还可以包括其他适当的信息,例如,目标用户的视线信息或目标用户的表情信息等,这同样在本申请的保护范围内。

[0041] 在步骤S220中,根据目标用户的面部信息,确定是否对目标用户进行响应。

[0042] 在示例实施例中,根据目标用户的面部信息,确定目标用户的面部朝向以及嘴唇状态为唇动状态,根据目标用户的面部朝向以及嘴唇状态,确定是否向目标用户进行反馈。例如,若确定目标用户的面部朝向为预定方向,并且目标用户的嘴唇状态为唇动状态,则确定向目标用户进行反馈,预定方向可以为交互设备的显示屏的正面方向。

[0043] 根据图2的示例实施例中的技术方案,一方面,根据目标用户的面部朝向以及嘴唇状态,确定是否对目标用户进行响应,能够高效准确地确定是否触发与目标用户的交互;另一方面,由于不需要在每轮对话的开始通过唤醒词进行唤醒,从而能够实现数字人与目标用户的高效自然的交流,提高用户的服务体验。

[0044] 进一步地,对目标用户进行反馈可以包括:通过语音形式对目标用户进行反馈,例如可以通过虚拟对象例如数字人对目标用户进行语音反馈。例如,获取目标用户的语音信息,从目标用户的语音信息中提取关键字或槽值,确定目标用户的意图,根据关键字和目标用户的意图,对目标用户进行语音反馈。以车站场景为例,目标用户的语音信息包括“今天北京天气怎样”,从目标用户的语音信息中提取关键字“今天”、“北京”、“天气”,确定目标用户的意图为询问天气,根据关键字和意图获取今天北京的天气结果,通过语音的形式对目标用户进行反馈,例如,通过虚拟对象对目标用户进行语音反馈。通过数字人对目标用户进行反馈,能够以人与人之间的交互方式向用户提供服务,提高用户的服务体验。

[0045] 图3为根据本申请的另一一些实施例提供的人机交互方法的流程示意图。

[0046] 参照图3所示,在步骤S310中,从目标用户的图像中获取目标用户的面部信息,该面部信息包括目标用户的面部朝向以及嘴唇状态。

[0047] 在示例实施例中,通过图像获取装置例如摄像头获取目标用户的图像,从目标用户的图像中提取目标用户的面部图像,从目标用户的面部图像中获取目标用户的面部信息,目标用户的面部信息可以包括目标用户的面部朝向以及嘴唇状态。

[0048] 需要说明的是,目标用户的面部信息还可以包括其他适当的信息,例如,目标用户的视线信息或目标用户的表情信息等,这同样在本申请的保护范围内。

[0049] 在步骤S320中,根据目标用户的面部信息,确定是否对目标用户进行响应。

[0050] 在一些实施例中,根据目标用户的面部朝向以及嘴唇状态,确定是否向目标用户进行反馈。例如,若确定目标用户的面部朝向为预定方向,并且目标用户的嘴唇状态为唇动状态,则确定向目标用户进行反馈,预定方向可以为虚拟对象例如数字人的正面方向。

[0051] 在另一些实施例中,目标用户的面部朝向还包括视线方向信息,根据目标用户的视线方向以及嘴唇状态,确定是否对目标用户进行反馈,例如,若确定目标用户的视线方向为预定方向,并且目标用户的嘴唇状态为唇动状态,则确定性目标用户进行反馈,预定方向可以为虚拟对象的正面方向。

[0052] 在步骤S330中,若确定进行响应,则从目标用户的图像中获取目标用户的当前状态,当前状态包括:所述目标用户的动作、表情、环境中的一种或多种。

[0053] 在示例实施例中,从目标用户的图像中提取目标用户的动作信息、表情信息以及周围环境信息中的一种或多种,根据目标用户的动作信息、表情信息以及周围环境信息确定目标用户的当前动作、当前表情、当前环境。例如,若目标用户当前动作信息为携带行李信息,则确定目标用户的当前动作包括携带行李动作;若目标用户当前表情信息包含着急表情信息,则确定目标当前表情为着急表情;若目标用户的周围环境信息包含同伴信息,则确定目标用户的当前环境包括有同伴。

[0054] 需要说明的是,目标用户的当前状态还可以包括其他适当的状态信息,例如,本轮或上一轮会话的会话信息或目标用户的个人信息等,这同样在本申请的保护范围内。

[0055] 在步骤S340中,根据目标用户的当前状态,对目标用户进行响应。

[0056] 在示例实施例中,根据目标用户的当前状态结合目标用户的语音信息,对目标用户进行响应,例如,以车站的场景为例,若目标用户的当前状态为携带行李状态,目标用户的语音信息中包含“列车号”“候车室”“哪”的关键字,则确定目标用户的意图为去候车室,则向目标用户反馈该列车号对应的候车室的位置。

[0057] 进一步地,在示例实施例中,根据目标用户的当前状态,通过虚拟对象例如数字人对目标用户进行反馈操作,反馈操作包括语音反馈、动作反馈以及表情反馈中的一种或多种。通过数字人对目标用户进行反馈,能够以人与人之间的交互方式向用户提供服务,提高用户的服务体验。

[0058] 根据图3的示例实施例中的技术方案,根据目标用户的当前状态和/或语音信息进行反馈,能够根据对目标用户的当前状态的认知和理解进行反馈,从而能够实现像人与人之间的高效自然的交互。

[0059] 进一步地,在示例实施例中,对目标用户进行响应包括:获取目标用户的语音信息,根据目标用户的当前状态以及语音信息,调用第三方应用程序对目标用户进行响应。例如,以车站场景为例,若目标用户的当前状态为携带行李,语音信息为“D301列车的候车室在哪”,则确定需要对目标用户进行导航,调用第三方导航应用程序,向目标用户展示对应的候车室的位置和路线图。通过调用第三方应用程序,能够更快捷准确地对目标用户进行反馈,为用户提供更多的服务,提高用户的服务体验。

[0060] 图4为根据本申请的又一些实施例提供的人机交互方法的流程示意图。

[0061] 参照图4所示,在步骤S410中,从目标用户的图像中获取目标用户的面部信息,该面部信息包括目标用户的面部朝向以及嘴唇状态。

[0062] 在示例实施例中,通过图像获取装置例如摄像头获取目标用户的图像,从目标用户的图像中提取目标用户的面部图像,从目标用户的面部图像中获取目标用户的面部信息,目标用户的面部信息可以包括目标用户的面部朝向以及嘴唇状态。

[0063] 需要说明的是,目标用户的面部信息还可以包括其他适当的信息,例如,目标用户的视线信息或目标用户的表情信息等,这同样在本申请的保护范围内。

[0064] 在步骤S420中,根据目标用户的面部信息,确定是否对目标用户进行响应。

[0065] 在一些实施例中,根据目标用户的面部朝向以及嘴唇状态,确定是否向目标用户进行响应。例如,若确定目标用户的面部朝向为预定方向,并且目标用户的嘴唇状态为唇动状态,则确定向目标用户进行反馈,预定方向可以为虚拟对象例如数字人的正面方向。

[0066] 在另一些实施例中,目标用户的面部信息还包括视线信息,根据目标用户的视线方向以及嘴唇状态,确定是否对目标用户进行反馈,例如,若确定目标用户的视线方向为预定方向,并且目标用户的嘴唇状态为唇动状态,则确定性目标用户进行反馈,预定方向可以为虚拟对象的正面方向。

[0067] 在步骤S430中,若确定进行响应,则从目标用户的图像中获取目标用户的当前状态,当前状态包括:所述目标用户的动作、表情、环境中的一种或多种。

[0068] 在示例实施例中,从目标用户的图像中提取目标用户的动作信息、表情信息以及周围环境信息中的一种或多种,根据目标用户的动作信息、表情信息以及周围环境信息确定目标用户的当前动作、当前表情、当前环境。例如,若目标用户当前动作信息为携带行李信息,则确定目标用户的当前动作包括携带行李动作;若目标用户当前表情信息包含着急表情信息,则确定目标当前表情为着急表情;若目标用户的周围环境信息包含同伴信息,则确定目标用户的当前环境包括有同伴。

[0069] 需要说明的是,目标用户的当前状态还可以包括其他适当的的状态信息,例如,本轮或上一轮会话的会话信息或目标用户的个人信息等,这同样在本申请的保护范围内。

[0070] 在步骤S440中,根据目标用户的当前状态,确定对目标用户的反馈优先级。

[0071] 在示例实施例中,反馈优先级用于表示对目标用户进行反馈的先后顺序,反馈优先级越高,则反馈时间越早,例如,反馈优先级包括紧急处理等级、快速处理等级、普通处理等级。

[0072] 进一步地,根据目标用户的当前状态,确定目标用户的反馈优先级。例如,根据目标用户的当前状态例如抱小孩动作、着急表情,确定目标用户的反馈优先级为紧急处理等级,

[0073] 在步骤S450中,基于目标用户的反馈优先级,对目标用户进行响应。

[0074] 在示例实施例中,确定对目标用户的反馈优先级,基于目标用户的反馈优先级,对目标用户进行反馈。例如,根据目标用户的当前状态例如抱小孩动作、着急表情,确定目标用户的反馈优先级为紧急处理等级,若存在多个目标用户,则优先对该目标用户进行反馈。

[0075] 进一步地,根据目标用户的当前状态和/或反馈优先级,通过虚拟对象例如数字人对目标用户进行反馈操作,反馈操作包括语音反馈、动作反馈以及表情反馈中的一种或多种。

[0076] 根据图4的示例实施例中的技术方案,根据目标用户的当前状态和/或反馈优先级进行反馈,能够对紧急情况进行优先处理,从而能够实现更高效自然的人机交互。

[0077] 此外,在示例实施例中,在与目标用户的对话过程中,若接收到目标用户通过图形用户界面例如交互设备的屏幕输入的交互指令,则根据该交互指令对目标用户进行反馈。例如,在车站的场景下,与目标用户的对话过程中,若接收到目标用户在交互设备的屏幕上输入的打开车站地图的指令,根据该指令在交互设备的屏幕上显示车站地图导航。

[0078] 图5为根据本申请的一些实施例提供的人机交互装置的示意框图。参照图5所示,该人机交互装置500包括:

[0079] 信息获取模块510,用于从目标用户的图像中获取所述目标用户的面部信息,所述面部信息包括所述目标用户的面部朝向以及嘴唇状态;

[0080] 响应判断模块520,用于根据所述目标用户的面部信息,确定是否对所述目标用户进行响应。

[0081] 根据图5的示例实施例中的技术方案,一方面,根据目标用户的面部朝向以及嘴唇状态,确定是否对目标用户进行响应,能够高效准确地确定是否触发与目标用户的交互;另一方面,由于不需要在每轮对话的开始通过唤醒词进行唤醒,从而能够实现数字人与目标用户的高效自然的交流,提高用户的服务体验。

[0082] 图6为根据本申请的一些实施例提供的响应判断模块的示意框图。参照图6所示,所述响应判断模块520包括:

[0083] 信息确定单元610,用于根据所述目标用户的面部信息,确定所述目标用户的所述面部朝向以及所述嘴唇状态;

[0084] 响应确定单元620,用于若所述面部朝向的方向为预定方向,并且所述嘴唇状态为唇动状态,则确定对所述目标用户进行响应。

[0085] 在本申请的一些实施例中,所述预定方向为虚拟对象的正面方向,所述响应判断模块520具体还用于:

[0086] 通过所述虚拟对象对所述目标用户进行响应。

[0087] 图7为根据本申请的一些实施例提供的人机交互装置的示意框图。参照图7所示，所述装置500还包括：

[0088] 状态获取模块710，用于从所述目标用户的图像中获取所述目标用户的当前状态，所述当前状态包括：所述目标用户的动作、表情、环境中的一种或多种；

[0089] 第一响应模块720，用于根据所述当前状态和/或所述目标对象的语音信息，对所述目标用户进行响应。

[0090] 根据图7的示例实施例中的技术方案，根据目标用户的当前状态和/或语音信息进行反馈，能够根据对目标用户的当前状态的认知和理解进行反馈，从而能够实现像人与人之间的高效自然的交互。

[0091] 在本申请的一些实施例中，所述第一响应模块720具体还用于：

[0092] 根据所述目标用户的当前状态，确定对所述目标用户的反馈优先级；

[0093] 基于所述反馈优先级，对所述目标用户进行响应。

[0094] 在本申请的一些实施例中，所述第一响应模块720具体还用于：

[0095] 获取所述目标用户的语音信息；

[0096] 根据所述当前状态和所述语音信息，通过虚拟对象对所述目标用户进行反馈操作，所述反馈操作包括语音反馈、动作反馈以及表情反馈中的一种或多种。

[0097] 在本申请的一些实施例中，所述第一响应模块720具体还用于：

[0098] 获取所述目标用户的语音信息；

[0099] 根据所述当前状态以及所述语音信息，调用第三方应用程序对所述目标用户进行响应。

[0100] 在本申请的一些实施例中，所述装置500还包括：

[0101] 语音信息获取模块，用于若确定对所述目标用户进行响应，则获取所述目标用户发出的语音信息；

[0102] 第二响应模块，用于基于所述语音信息，对所述目标用户进行语音反馈。

[0103] 在本申请的一些实施例中，所述装置500还包括：

[0104] 第三响应模块，用于接收所述目标用户通过图形用户界面输入的交互指令，根据所述交互指令对所述目标用户进行反馈。

[0105] 上述几种实施方式提供的人机交互装置，用于实现前述任一方法实施例中的人机交互方法的技术方案，其实现原理和技术效果类似，在此不再赘述。

[0106] 需要说明的是，应理解以上几个实施例中提供的装置的各个模块的划分仅仅是一种逻辑功能的划分，实际实现时可以全部或部分集成到一个物理实体上，也可以物理上分开。且这些模块可以全部以软件通过处理元件调用的形式实现；也可以全部以硬件的形式实现；还可以部分模块通过处理元件调用软件的形式实现，部分模块通过硬件的形式实现。例如，收益指标确定模块可以为单独设立的处理元件，也可以集成在上述装置的某一个芯片中实现，此外，也可以以程序代码的形式存储于上述装置的存储器中，由上述装置的某一个处理元件调用并执行以上处理模块的功能。其它模块的实现与之类似。此外这些模块全部或部分可以集成在一起，也可以独立实现。这里所述的处理元件可以是一种集成电路，具有信号的处理能力。在实现过程中，上述方法的各步骤或以上各个模块可以通过处理器元件中的硬件的集成逻辑电路或者软件形式的指令完成。

[0107] 图8是用来实现本申请实施例的人机交互方法的电子设备的框图。如图8所示,是根据本申请实施例的电子设备的框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本申请的实现。

[0108] 如图8所示,该电子设备包括:一个或多个处理器810、存储器820,以及用于连接各部件的接口,包括高速接口和低速接口,以及与其他电子设备进行通信的接口。各个部件利用不同的总线互相连接,并且可以被安装在公共主板上或者根据需要以其它方式安装。处理器可以对在电子设备内执行的指令进行处理,包括存储在存储器中或者存储器上以在外部输入/输出装置(诸如,耦合至接口的显示设备)上显示图形用户接口(Graphical User Interface, GUI)的图形信息的指令。在其它实施方式中,若需要,可以将多个处理器和/或多条总线与多个存储器和多个存储器一起使用。同样,可以连接多个电子设备,各个设备提供部分必要的操作(例如,作为服务器阵列、一组刀片式服务器、或者多处理器系统)。图8中以一个处理器810为例。

[0109] 存储器820即为本申请所提供的非瞬时计算机可读存储介质。其中,所述存储器存储有可由至少一个处理器执行的指令,以使所述至少一个处理器执行本申请所提供的任一执行主体对应的人机交互方法。本申请的非瞬时计算机可读存储介质存储计算机指令,该计算机指令用于使计算机执行本申请所提供的方法。

[0110] 存储器820作为一种非瞬时计算机可读存储介质,可用于存储非瞬时软件程序、非瞬时计算机可执行程序以及模块,如本申请实施例中的人机交互方法中对应的程序指令/模块。处理器810通过运行存储在存储器820中的非瞬时软件程序、指令以及模块,从而执行服务器的各种功能应用以及数据处理,即实现上述方法实施例中的任一执行主体对应的人机交互方法。

[0111] 存储器820可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区则可以存储数据,例如,数据处理平台中存储的各方提供的数据,或者安全隔离区域中的叔叔等。此外,存储器820可以包括高速随机存取存储器,还可以包括非瞬时存储器,例如至少一个磁盘存储器件、闪存器件、或其他非瞬时固态存储器件。在一些实施例中,存储器820可选包括相对于处理器810远程设置的存储器,这些远程存储器可以通过网络连接至数据处理的电子设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0112] 此外,电子设备还可以包括:输入装置830和输出装置840。处理器810、存储器820、输入装置830和输出装置840可以通过总线850或者其他方式连接,图8中以通过总线连接为例。

[0113] 输入装置830可接收输入的数字或字符信息,以及产生与数据处理的电子设备的用户设置以及功能控制有关的键信号输入,例如触摸屏、小键盘、鼠标、轨迹板、触模板、指示杆、一个或者多个鼠标按钮、轨迹球、操纵杆等输入装置。输出装置840可以包括显示设备、辅助照明装置(例如,发光二极管)和触觉反馈装置(例如,振动电机)等。该显示设备可以包括但不限于,液晶显示器(Liquid Crystal Display, LCD)、发光二极管(Light

Emitting Diode,LED) 显示器和等离子体显示器。在一些实施方式中,显示设备可以是触摸屏。

[0114] 此处描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、专用集成电路 (Application Specific Integrated Circuit,ASIC)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0115] 这些计算程序(也称作程序、软件、软件应用、或者代码)包括可编程处理器的机器指令,并且可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。如本文使用的,术语“机器可读介质”和“计算机可读介质”指的是用于将机器指令和/或数据提供给可编程处理器的任何计算机程序产品、设备、和/或装置(例如,磁盘、光盘、存储器、可编程逻辑装置(Programmable Logic device,PLD)),包括,接收作为机器可读信号的机器指令的机器可读介质。术语“机器可读信号”指的是用于将机器指令和/或数据提供给可编程处理器的任何信号。

[0116] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,阴极射线管或者LCD监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0117] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(Local Area Network,LAN)、广域网(Wide Area Network,WAN)和互联网。

[0118] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。

[0119] 进一步地,本申请还提供一种存储有计算机指令的非瞬时计算机可读存储介质,用于存储计算机指令,该计算机指令被处理器执行后可实现前述任一方法实施例提供的技术方案。

[0120] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本申请公开的技术方案所期望的结果,本文在此不进行限制。

[0121] 上述具体实施方式,并不构成对本申请保护范围的限制。本领域技术人员应该明

白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本申请的精神和原则之内所作的修改、等同替换和改进等,均应包含在本申请保护范围之内。

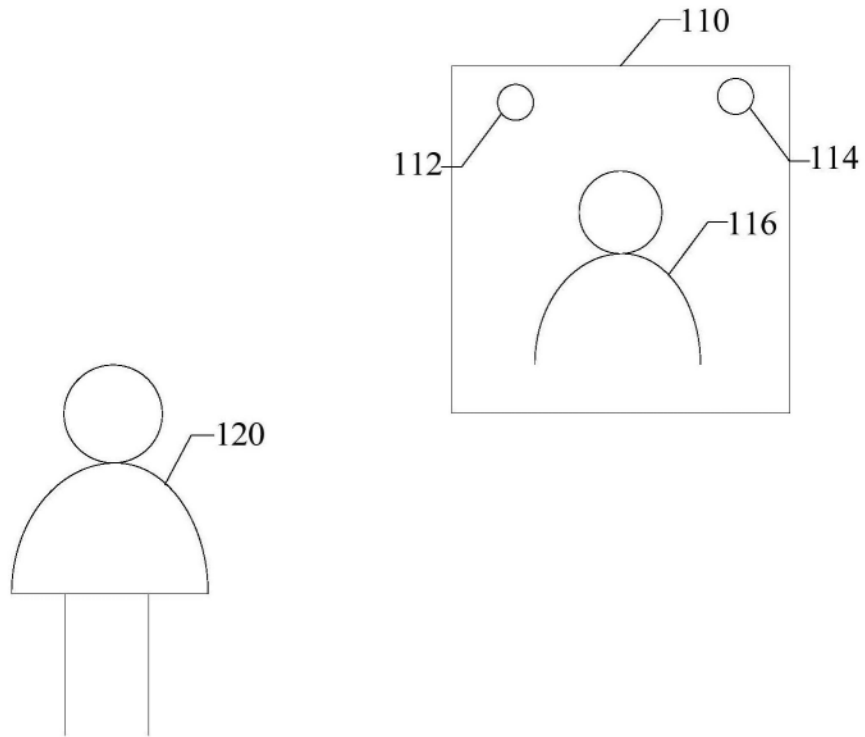


图1

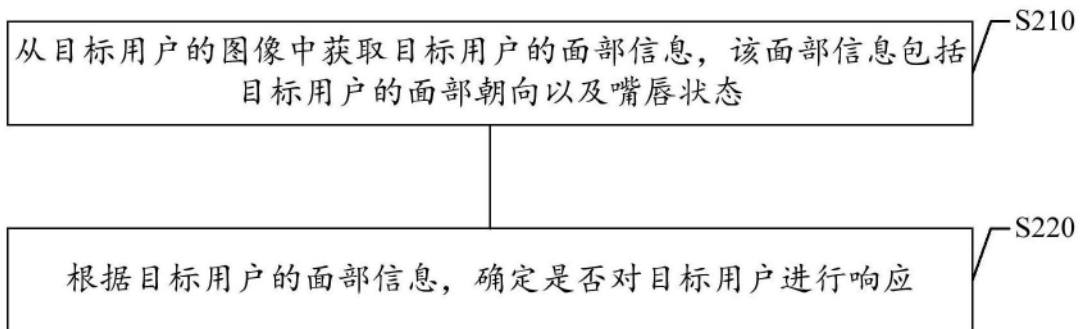


图2

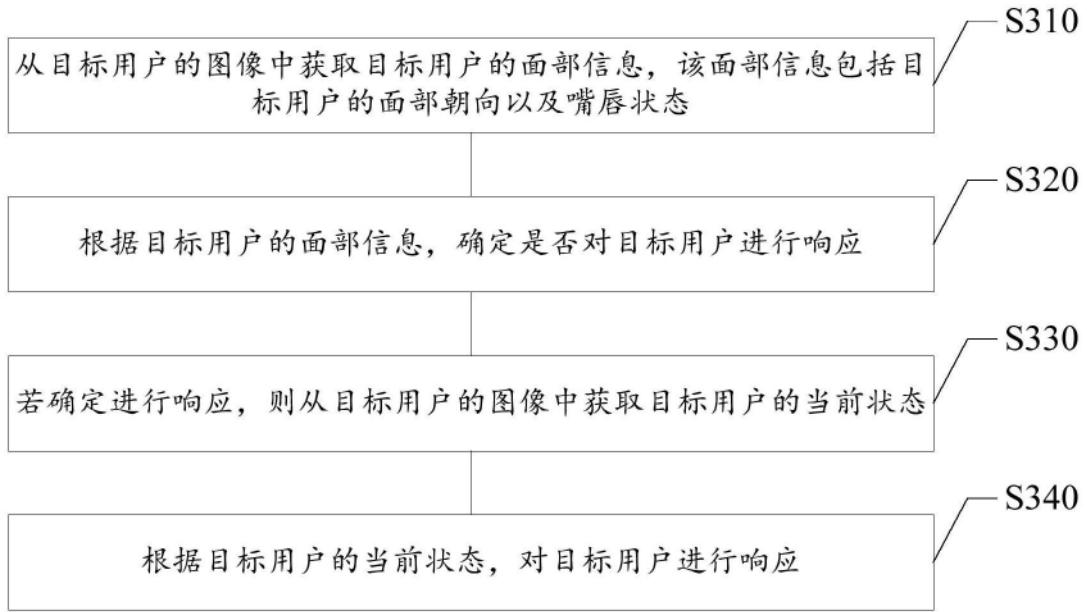


图3

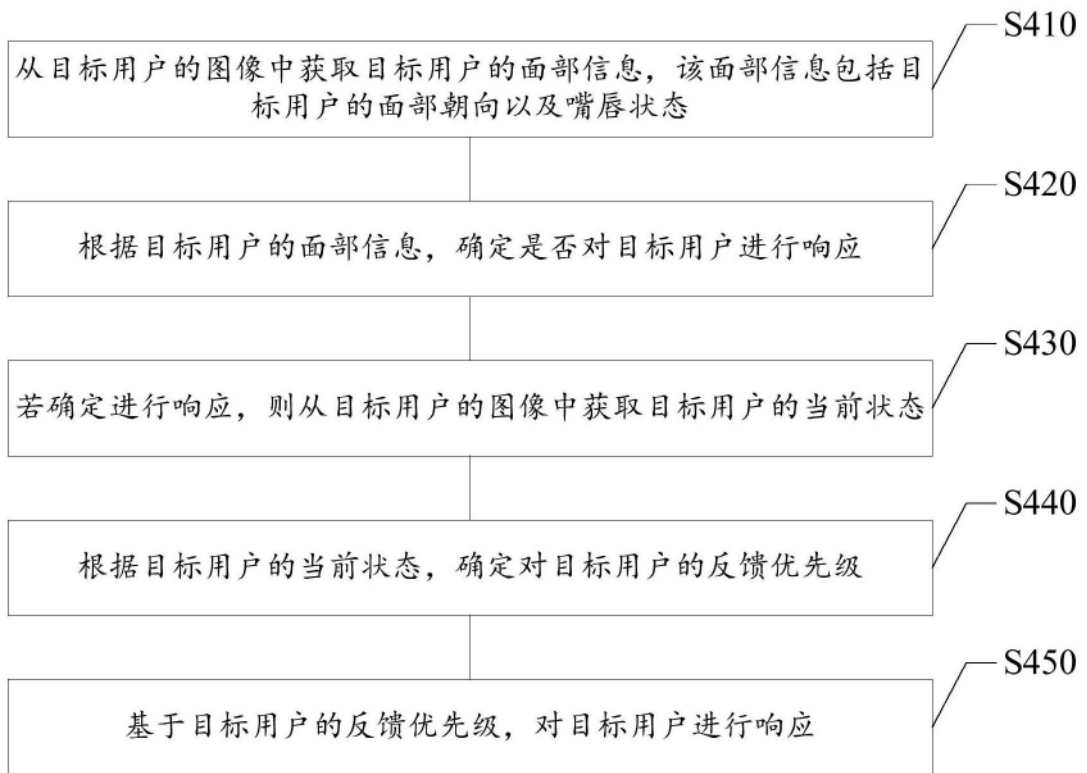


图4

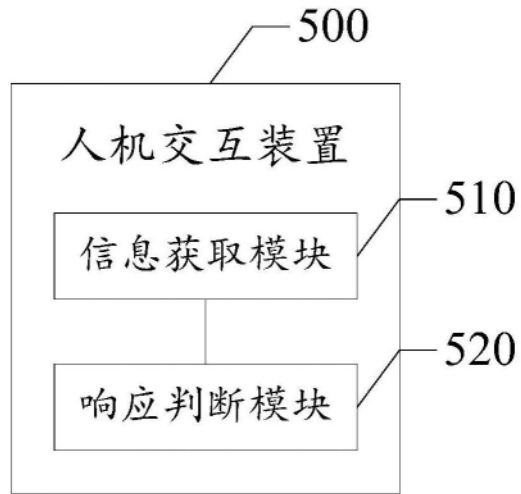


图5

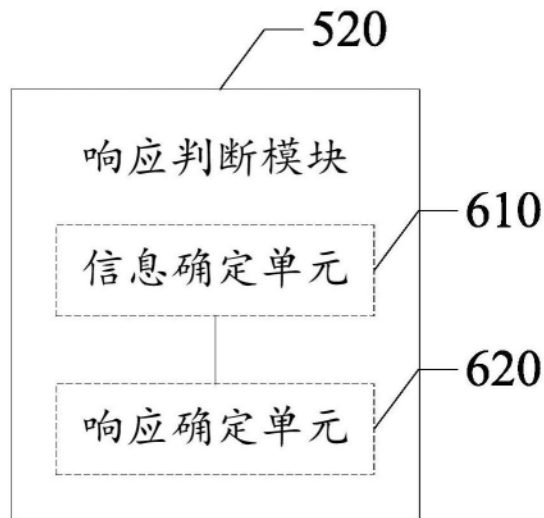


图6

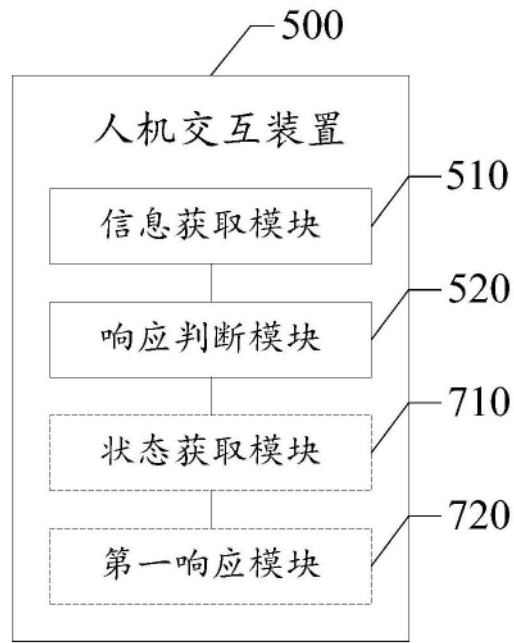


图7

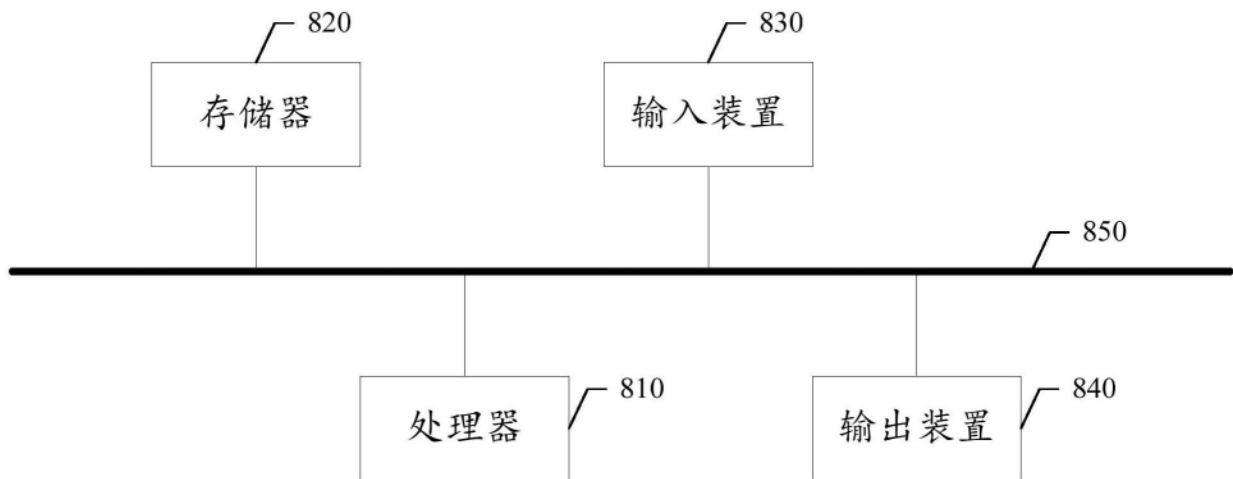


图8