



# (12) 发明专利申请

(10) 申请公布号 CN 115376487 A

(43) 申请公布日 2022. 11. 22

(21) 申请号 202211003279.X

(22) 申请日 2022.08.19

(71) 申请人 北京百度网讯科技有限公司  
地址 100085 北京市海淀区上地十街10号  
百度大厦2层

(72) 发明人 郭紫垣

(74) 专利代理机构 北京市汉坤律师事务所  
11602  
专利代理师 姜浩然 吴丽丽

(51) Int. Cl .  
G10L 13/10 (2013.01)  
G10L 13/047 (2013.01)  
G10L 25/63 (2013.01)

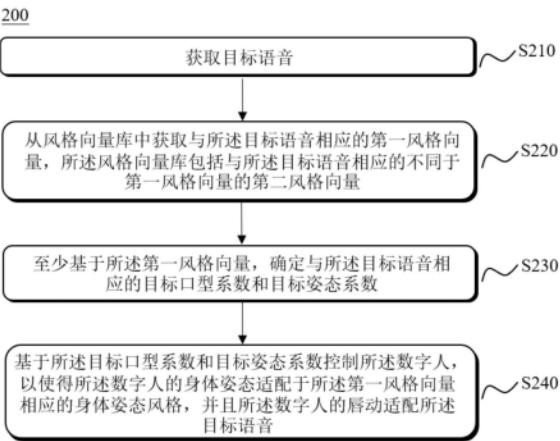
权利要求书3页 说明书13页 附图6页

## (54) 发明名称

数字人的控制方法、模型训练方法和装置

## (57) 摘要

本公开提供了一种数字人的控制方法、模型训练方法和装置,涉及人工智能技术领域,尤其涉及虚拟数字人、增强现实、虚拟现实、混合现实、扩展现实、元宇宙等技术领域。实现方案为:获取目标语音;从风格向量库中获取与所述目标语音相应的第一风格向量,所述风格向量库包括与所述目标语音相应的不同于第一风格向量的第二风格向量;至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数;以及基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动适配所述目标语音。



1. 一种数字人的控制方法,包括:

获取目标语音;

从风格向量库中获取与所述目标语音相应的第一风格向量,所述风格向量库包括与所述目标语音相应的不同于第一风格向量的第二风格向量;

至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数;以及

基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动适配所述目标语音。

2. 根据权利要求1所述的方法,其中,所述风格向量库包括多个风格向量,所述多个风格向量中不同风格向量相应的身体姿态风格不同。

3. 根据权利要求1或2所述的方法,还包括:

获取所述目标语音的语音特征,

所述至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:

基于所述目标语音的语音特征和所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数。

4. 根据权利要求1-3中任一项所述的方法,其中,所述第一风格向量为从风格向量库中所获取的一个风格向量。

5. 根据权利要求1-3中任一项所述的方法,还包括:

获取所述目标语音相应的目标文本,所述目标文本包括多个发音单元;以及

获取所述目标语音的每个发音单元相应的时间戳,

所述至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:

基于所述目标语音所包括的多个发音单元各自相应的时间戳以及所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数。

6. 根据权利要求5所述的方法,其中,所述第一风格向量包括从风格向量库中所获取的多个风格向量,基于所述目标语音所包括的多个发音单元各自相应的时间戳以及所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:

基于所述目标语音自身的上下文信息,确定所述目标语音中与所述多个风格向量中的每一者相应的起始发音单元和结束发音单元;

对于所述多个风格向量中的每一者,基于相应的起始发音单元和结束发音单元的时间戳,确定所述风格向量相应的起始帧和结束帧;以及

基于所述多个风格向量、所述多个风格向量中的每一者的起始帧和结束帧,确定与所述目标语音在时间上逐帧地对齐的目标口型系数和目标姿态系数。

7. 根据权利要求6所述的方法,其中,所述目标语音的上下文信息包括有关下列项中的一者或多者的信息:张口音、闭口音、爆破音、重复语段、语气词、转折词。

8. 根据权利要求1-7中任一项所述的方法,还包括:

获取目标情绪系数;以及

基于目标情绪系数调整所述目标姿态系数,以调整所述数字人的身体姿态幅度,  
所述基于所述目标口型系数和目标姿态系数控制所述数字人包括:  
基于所述目标口型系数和调整后的目标姿态系数控制所述数字人。

9. 根据权利要求1-7中任一项所述的方法,其中,从风格向量库中获取与所述目标语音相应的第一风格向量包括:

获取与所述目标语音相关的预设发音单元;

从风格向量库中获取所述预设发音单元相应的多个风格向量;

基于所述多个风格向量确定所述预设发音单元相应的多个口型系数和所述预设发音单元相应的多个姿态系数;

基于所述多个口型系数和所述多个姿态系数控制数字人;以及

基于相应的所述多个口型系数和所述多个姿态系数所得到的数字人的控制结果,从所述多个风格向量中确定所述第一风格向量。

10. 根据权利要求1-7中任一项所述的方法,其中,至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:

至少将所述第一风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的与所述目标语音相应的目标口型系数和目标姿态系数。

11. 一种用于数字人的模型训练方法,包括:

获取样本语音以及所述样本语音所包括的至少一个发音单元;

获取与所述样本语音相应的真实口型系数和真实姿态系数;

获取所述样本语音所包括的每个发音单元,以及与每个发音单元相应的起始帧和结束帧;

确定所述样本语音相应的风格向量,其中所述风格向量包括所述至少一个发音单元各自相应的子风格向量,所述样本语音中的每个发音单元的起始帧与结束帧之间的所有帧共用相同的子风格向量;

将所述样本语音和所述至少一个发音单元相应的至少一个子风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的预测口型系数和预测姿态系数;以及

基于所述真实口型系数和预测口型系数之间的损失以及所述真实姿态系数和预测姿态系数之间的损失,更新所述身体姿态与唇动模型的网络参数,并且更新所述风格向量。

12. 根据权利要求11所述的方法,还包括:

在所述数字人身体姿态与唇动模型的训练完成之后,基于经更新的风格向量确定风格向量库。

13. 一种数字人的控制装置,包括:

第一获取模块,所述第一获取模块被配置获取目标语音;

第二获取模块,所述第二获取模块被配置为从风格向量库中获取与所述目标语音相应的第一风格向量,所述风格向量库包括与所述目标语音相应的不同于第一风格向量的第二风格向量;

确定模块,所述确定模块被配置为至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数;以及

控制模块,所述控制模块被配置为基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动适配所述目标语音。

14.一种用于数字人的模型训练装置,包括:

第三获取模块,所述第三获取模块被配置为获取样本语音以及所述样本语音所包括的至少一个发音单元;

第四获取模块,所述第四获取模块被配置为获取与所述样本语音相应的真实口型系数和真实姿态系数;

第五获取模块,所述第五获取模块被配置为获取所述样本语音所包括的每个发音单元,以及与每个发音单元相应的起始帧和结束帧;以及

第一确定模块,所述第一确定模块被配置为确定所述样本语音相应的风格向量,其中所述风格向量包括所述至少一个发音单元各自相应的子风格向量,所述样本语音中的每个发音单元的起始帧与结束帧之间的所有帧共用相同的子风格向量;

第二确定模块,所述第二确定模块被配置为将所述样本语音和所述至少一个发音单元相应的至少一个子风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的预测口型系数和预测姿态系数;以及

更新模块,所述更新模块被配置为基于所述真实口型系数和预测口型系数之间的损失以及所述真实姿态系数和预测姿态系数之间的损失,更新所述身体姿态与唇动模型的网络参数,并且更新所述风格向量。

15.一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中

所述存储器存储有计算机程序,所述计算机程序在被所述至少一个处理器执行时实现根据权利要求1-12中任一项所述的方法。

16.一种存储有计算机程序的非瞬时计算机可读存储介质,其中,所述计算机程序在被处理器执行时实现根据权利要求1-12中任一项所述的方法。

17.一种计算机程序产品,包括计算机程序,其中,所述计算机程序在被处理器执行时实现根据权利要求1-12中任一项所述的方法。

## 数字人的控制方法、模型训练方法和装置

### 技术领域

[0001] 本公开涉及人工智能技术领域,尤其涉及虚拟数字人、增强现实、虚拟现实、混合现实、扩展现实、元宇宙等技术领域,具体涉及一种数字人的控制方法以及用于数字人的模型训练方法、装置、电子设备、计算机可读存储介质和计算机程序产品。

### 背景技术

[0002] 人工智能是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科,既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术。人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

[0003] 数字人高精形象生成一直是学术界的研究热点,并有着广泛的应用,最为常见的应用领域有虚拟主播、虚拟客服、虚拟助理、虚拟教师、虚拟偶像以及其他互动游戏和娱乐等。

[0004] 在此部分中描述的方法不一定是之前已经设想到或采用的方法。除非另有指明,否则不应假定此部分中描述的任何方法仅因其包括在此部分中就被认为是相关技术。类似地,除非另有指明,否则此部分中提及的问题不应认为在任何相关技术中已被公认。

### 发明内容

[0005] 本公开提供了一种数字人的控制方法以及用于数字人的模型训练方法、装置、电子设备、计算机可读存储介质和计算机程序产品。

[0006] 根据本公开的一方面,提供了数字人的控制方法,包括:获取目标语音;从风格向量库中获取与所述目标语音相应的第一风格向量,所述风格向量库包括与所述目标语音相应的不同于第一风格向量的第二风格向量;至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数;以及基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动适配所述目标语音。

[0007] 根据本公开的一方面,提供了用于数字人的模型训练方法,包括:获取样本语音以及所述样本语音所包括的至少一个发音单元;获取与所述样本语音相应的真实口型系数和真实姿态系数;获取所述样本语音所包括的每个发音单元,以及与每个发音单元相应的起始帧和结束帧;确定所述样本语音相应的风格向量,其中所述风格向量包括所述至少一个发音单元各自相应的子风格向量,所述样本语音中的每个发音单元的起始帧与结束帧之间的所有帧共用相同的子风格向量;将所述样本语音和所述至少一个发音单元相应的至少一个子风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的预测口型系数和预测姿态系数;以及基于所述真实口型系数和预测口型系数之间的损失以及所述真实姿态系数和预测姿态系数之间的损失,更新所述身体姿态与唇动模型的网络参数,并

且更新所述风格向量。

[0008] 根据本公开的另一方面,提供了数字人的控制装置,包括:第一获取模块,所述第一获取模块被配置获取目标语音;第二获取模块,所述第二获取模块被配置为从风格向量库中获取与所述目标语音相应第一风格向量,所述风格向量库包括与所述目标语音相应的不同于第一风格向量的第二风格向量;以及确定模块,所述确定模块被配置为至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数;以及控制模块,所述控制模块被配置为基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动适配所述目标语音。

[0009] 根据本公开的另一方面,提供了用于数字人的模型训练装置,包括:第三获取模块,所述第三获取模块被配置为获取样本语音以及所述样本语音所包括的至少一个发音单元;第四获取模块,所述第四获取模块被配置为获取与所述样本语音相应的真实口型系数和真实姿态系数;第五获取模块,所述第五获取模块被配置为获取所述样本语音所包括的每个发音单元,以及与每个发音单元相应的起始帧和结束帧;以及第一确定模块,所述第一确定模块被配置为确定所述样本语音相应的风格向量,其中所述风格向量包括所述至少一个发音单元各自相应的子风格向量,所述样本语音中的每个发音单元的起始帧与结束帧之间的所有帧共用相同的子风格向量;第二确定模块,所述第二确定模块被配置为将所述样本语音和所述至少一个发音单元相应的至少一个子风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的预测口型系数和预测姿态系数;以及更新模块,所述更新模块被配置为基于所述真实口型系数和预测口型系数之间的损失以及所述真实姿态系数和预测姿态系数之间的损失,更新所述身体姿态与唇动模型的网络参数,并且更新所述风格向量。

[0010] 根据本公开的另一方面,提供了一种计算设备,所述计算设备包括:处理器;以及存储程序的存储器,所述程序包括指令,所述指令在由所述处理器执行时使所述处理器执行上述的方法。

[0011] 根据本公开的另一方面,提供了一种存储程序的计算机可读存储介质,所述程序包括指令,所述程序包括指令,所述指令在由计算设备的处理器执行时,致使所述计算设备执行上述的方法。

[0012] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,其中,所述计算机程序在被处理器执行时实现根据上述的方法。

[0013] 根据本公开的一个或多个实施例,通过针对一种身体姿态风格使用相应风格向量进行控制,能够实现数字人的身体姿态风格的控制。

[0014] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

## 附图说明

[0015] 附图示例性地示出了实施例并且构成说明书的一部分,与说明书的文字描述一起用于讲解实施例的示例性实施方式。所示出的实施例仅出于例示的目的,并不限制权利要求的范围。在所有附图中,相同的附图标记指代类似但不一定相同的要素。

[0016] 图1示出了根据本公开的实施例可以将本文描述的各种方法和装置在其中实施的示例性系统的示意图；

[0017] 图2示出了根据本公开的实施例的数字人的控制方法的流程图；

[0018] 图3示出了根据本公开的用于确定目标口型系数和目标姿态系数的方法的流程图；

[0019] 图4示出了根据本公开的实施例的用于数字人的模型训练方法的流程图；

[0020] 图5示出了根据本公开的驱动数字人的过程图；

[0021] 图6示出了根据本公开的身体姿态与唇动模型的框图；

[0022] 图7示出了根据本公开的实施例的数字人的控制装置的组成框图；

[0023] 图8示出了根据本公开的实施例的用于数字人的模型训练装置的组成框图；

[0024] 图9示出了能够用于实现本公开的实施例的示例性电子设备的结构框图。

### 具体实施方式

[0025] 以下结合附图对本公开的示范性实施例做出说明，其中包括本公开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，而不会背离本公开的范围。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] 在本公开中，除非另有说明，否则使用术语“第一”、“第二”等来描述各种要素不意图限定这些要素的位置关系、时序关系或重要性关系，这种术语只是用于将一个元件与另一元件区分开。在一些示例中，第一要素和第二要素可以指向该要素的同一实例，而在某些情况下，基于上下文的描述，它们也可以指代不同实例。

[0027] 在本公开中对各种所述示例的描述中所使用的术语只是为了描述特定示例的目的，而并非旨在进行限制。除非上下文另外明确地表明，如果不特意限定要素的数量，则该要素可以是一个也可以是多个。此外，本公开中所使用的术语“和/或”涵盖所列出的项目中的任何一个以及全部可能的组合方式。

[0028] 下面将结合附图详细描述本公开的实施例。

[0029] 图1示出了根据本公开的实施例可以将本文描述的各种方法和装置在其中实施的示例性系统100的示意图。参考图1，该系统100包括一个或多个客户端设备101、102、103、104、105和106、服务器120以及将一个或多个客户端设备耦接到服务器120的一个或多个通信网络110。客户端设备101、102、103、104、105和106可以被配置为执行一个或多个应用程序。

[0030] 在本公开的实施例中，服务器120可以运行使得能够执行数字人的控制以及用于数字人的模型训练的方法的一个或多个服务或软件应用。

[0031] 在某些实施例中，服务器120还可以提供可以包括非虚拟环境和虚拟环境的其他服务或软件应用。在某些实施例中，这些服务可以作为基于web的服务或云服务提供，例如在软件即服务(SaaS)模型下提供给客户端设备101、102、103、104、105和/或106的用户。

[0032] 在图1所示的配置中，服务器120可以包括实现由服务器120执行的功能的一个或多个组件。这些组件可以包括可由一个或多个处理器执行的软件组件、硬件组件或其组合。操作客户端设备101、102、103、104、105和/或106的用户可以依次利用一个或多个客户端应

用程序来与服务器120进行交互以利用这些组件提供的服务。应当理解,各种不同的系统配置是可能的,其可以与系统100不同。因此,图1是用于实施本文所描述的各种方法的系统的一个示例,并且不旨在进行限制。

[0033] 用户可以使用客户端设备101、102、103、104、105和/或106来对数字人身体姿态风格进行控制。客户端设备可以提供使客户端设备的用户能够与客户端设备进行交互的接口。客户端设备还可以经由该接口向用户输出信息。尽管图1仅描绘了六种客户端设备,但是本领域技术人员将能够理解,本公开可以支持任何数量的客户端设备。

[0034] 客户端设备101、102、103、104、105和/或106可以包括各种类型的计算机设备,例如便携式手持设备、通用计算机(诸如个人计算机和膝上型计算机)、工作站计算机、可穿戴设备、智能屏设备、自助服务终端设备、服务机器人、游戏系统、瘦客户端、各种消息收发设备、传感器或其他感测设备等。这些计算机设备可以运行各种类型和版本的软件应用程序和操作系统,例如MICROSOFT Windows、APPLE iOS、类UNIX操作系统、Linux或类Linux操作系统(例如GOOGLE Chrome OS);或包括各种移动操作系统,例如MICROSOFT Windows Mobile OS、iOS、Windows Phone、Android。便携式手持设备可以包括蜂窝电话、智能电话、平板电脑、个人数字助理(PDA)等。可穿戴设备可以包括头戴式显示器(诸如智能眼镜)和其他设备。游戏系统可以包括各种手持式游戏设备、支持互联网的游戏设备等。客户端设备能够执行各种不同的应用程序,例如各种与Internet相关的应用程序、通信应用程序(例如电子邮件应用程序)、短消息服务(SMS)应用程序,并且可以使用各种通信协议。

[0035] 网络110可以是本领域技术人员熟知的任何类型的网络,其可以使用多种可用协议中的任何一种(包括但不限于TCP/IP、SNA、IPX等)来支持数据通信。仅作为示例,一个或多个网络110可以是局域网(LAN)、基于以太网的网络、令牌环、广域网(WAN)、因特网、虚拟网络、虚拟专用网络(VPN)、内部网、外部网、公共交换电话网(PSTN)、红外网络、无线网络(例如蓝牙、WIFI)和/或这些和/或其他网络的任意组合。

[0036] 服务器120可以包括一个或多个通用计算机、专用服务器计算机(例如PC(个人计算机)服务器、UNIX服务器、中端服务器)、刀片式服务器、大型计算机、服务器群集或任何其他适当的布置和/或组合。服务器120可以包括运行虚拟操作系统的一个或多个虚拟机,或者涉及虚拟化的其他计算架构(例如可以被虚拟化以维护服务器的虚拟存储设备的逻辑存储设备的一个或多个灵活池)。在各种实施例中,服务器120可以运行提供下文所描述的功能的一个或多个服务或软件应用。

[0037] 服务器120中的计算单元可以运行包括上述任何操作系统以及任何商业上可用的服务器操作系统的一个或多个操作系统。服务器120还可以运行各种附加服务器应用程序和/或中间层应用程序中的任何一个,包括HTTP服务器、FTP服务器、CGI服务器、JAVA服务器、数据库服务器等。

[0038] 在一些实施方式中,服务器120可以包括一个或多个应用程序,以分析和合并从客户端设备101、102、103、104、105和/或106的用户接收的数据馈送和/或事件更新。服务器120还可以包括一个或多个应用程序,以经由客户端设备101、102、103、104、105和/或106的一个或多个显示设备来显示数据馈送和/或实时事件。

[0039] 在一些实施方式中,服务器120可以为分布式系统的服务器,或者是结合了区块链的服务器。服务器120也可以是云服务器,或者是带人工智能技术的智能云计算服务器或智

能云主机。云服务器是云计算服务体系中的一项主机产品,以解决传统物理主机与虚拟专用服务器(VPS,Virtual Private Server)服务中存在的管理难度大、业务扩展性弱的缺陷。

[0040] 系统100还可以包括一个或多个数据库130。在某些实施例中,这些数据库可以用于存储数据和其他信息。例如,数据库130中的一个或多个可用于存储诸如音频文件和视频文件的信息。数据库130可以驻留在各种位置。例如,由服务器120使用的数据库可以在服务器120本地,或者可以远离服务器120且可以经由基于网络或专用的连接与服务器120通信。数据库130可以是不同的类型。在某些实施例中,由服务器120使用的数据库例如可以是关系数据库。这些数据库中的一个或多个可以响应于命令而存储、更新和检索到数据库以及来自数据库的数据。

[0041] 在某些实施例中,数据库130中的一个或多个还可以由应用程序使用来存储应用程序数据。由应用程序使用的数据库可以是不同类型的数据库,例如键值存储库,对象存储库或由文件系统支持的常规存储库。

[0042] 图1的系统100可以以各种方式配置和操作,以使得能够应用根据本公开所描述的各种方法和装置。

[0043] 相关技术中,通过使模型学习样本音频到人体姿态系数和口型系数的直接映射来对新音频进行推理。尽管这种直接端到端的映射学习方法是直截了当的,但一般只能将数字人(例如,3D数字人等)固定为具有一种身体姿态风格,如果需要在模型推理时更换姿态风格,则需要花费时间来重新训练模型。由于不同的演员有不同的演讲风格,不同演员讲话时的身体姿态也因而各有特点,因此端到端的映射学习方法使不同风格的身体姿态数据无法得到充分利用,进而导致模型的推理表现与稳定性不足。

[0044] 除此之外,发明人发现引起模型对姿态的推理性能差的另一个原因在于,同一个演员在不同状态下也会存在多种姿态风格。由于演员在阅读过程中,对于不同话术的感受度不同,即使同一个演员,在不同文字的阅读阶段的姿态风格也可能是不断变化的。于是,发明人洞察到对单个发音单元而言,存在与之对应的姿态风格。相关技术方案并没有就此进行探索,导致利用模型生成的身体姿态动画其实是多种结果的耦合,进而导致模型总是趋向于输出较为平均的结果,最终表现为整体身体姿态动画变化平缓,对爆破音、闭口音、张口音的变化不敏感。

[0045] 为了解决上述技术问题,本公开通过针对一种身体姿态风格使用相应风格向量进行控制,能够实现数字人的身体姿态风格的控制。

[0046] 图2示出了根据本公开的实施例的数字人的控制方法的流程图,如图2所示,方法200可以包括

[0047] 步骤S210:获取目标语音;

[0048] 步骤S220:从风格向量库中获取与所述目标语音相应的第一风格向量,所述风格向量库包括与所述目标语音相应的不同于第一风格向量的第二风格向量;

[0049] 步骤S230:至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数;以及

[0050] 步骤S240:基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动

适配所述目标语音。

[0051] 由此,针对一种身体姿态风格使用相应风格向量进行控制,能够实现数字人的身体姿态风格的控制。

[0052] 根据一些实施例,目标语音可以是自然语言输入,也可以是文本到语音(text-to-speech,TTS)输入。

[0053] 根据一些实施例,所述风格向量库可包括多个风格向量,每一个风格向量存储与所对应的对象(例如,单个发音单元,其中,发音单元可以是单个字、单个词语、单个句子、单个语段等)有关的风格信息,所述风格信息是在模型的训练阶段习得的。例如,在一风格向量存储与单个字“您”有关的风格信息的情况下,该风格向量所存储的风格信息可包括模型训练阶段期间当给定“您”的多个正常发音状态时的不同身体姿态真值(例如,由不同语音输入对象发出“您”字的正常读音时所做出的不同的身体姿态)之间的差异性。在该示例中,这种差异性可体现在以单个发音单元的正常读音发音的情况下个体语音输入对象在身体姿态的展示上的动作类型差异、动作幅度差异、动作方向差异、动作持续时间差异等等。

[0054] 根据一些实施例,第一风格向量可以指当被应用于驱动数字人进行展示时所得的身体姿态风格相对符合应用场景上下文和/或预设准则的一个或多个风格向量。例如,第一风格向量可以是在模型推理之前对风格向量库中的每一个风格向量进行遍历并且针对数字人进行仿真之后所确定的符合(例如,最契合)应用场景上下文和/或预设准则的风格向量。在一些示例中,应用场景上下文可包括目标文本(即,模型推理时输入到模型中的目标语音相应的文本)的文体、体裁、句式、应用场合等等。在另一些示例中,预设准则可包括风格向量的仿真结果是否符合目标语音中的爆破音的发音习惯(例如,发闭口音时嘴部闭合是否足够严密)、风格向量的仿真结果是否符合目标语音对应的身体姿态动作的习惯执行(例如,说“嗨”时通常伴随的挥手动作)、风格向量的仿真结果是否符合目标语音对应的身体姿态动作的执行幅度(例如,说“请”时的手臂伸展幅度是否明显)等等。

[0055] 根据一些实施例,可至少基于第一风格向量来确定目标语音相应的目标口型系数和目标姿态系数。目标口型系数和目标姿态系数用于驱动数字人进行与目标语音的播放同步的数字人唇动和身体姿态展示。在一些示例中,数字人唇动和身体姿态展示可以是语音帧同步的(即,逐帧的)。在一些示例中,例如在画面帧率要求不高的情况下,数字人唇动和身体姿态展示可以是隔帧的,即,相邻两帧的数字人唇动和身体姿态展示动画之间的一个或多个帧无需模型推理所输出的目标口型系数和目标姿态系数来驱动,而是被保持为与该相邻两帧中的一者或两者相一致。在另一些实施例中,数字人唇动和身体姿态展示可以是补间的,即,相邻两帧的数字人唇动和身体姿态展示动画之间的一个或多个帧无需模型推理所输出的目标口型系数和目标姿态系数来驱动,而是改由客户端或服务器基于该相邻两帧来自动生成。

[0056] 在一些实施例中,所述风格向量库包括多个风格向量,所述多个风格向量中不同风格向量相应的身体姿态风格不同。在模型训练阶段,针对训练文本中的每个发音单元配置相应发音单元向量,这些发音单元向量随着模型的训练而逐渐地具有如上所述的差异性。由此,风格向量库可随训练文本中所包括的发音单元数目的增加而得到扩展,这不仅节省了对训练文本及对应姿态真值进行分类和标记所需的人工和/或计算成本,而且置备了在模型推理时可供选取的就风格差异性而言细粒度较小的精细风格候选。

[0057] 在一些实施例中,上述方法200还包括获取所述目标语音的语音特征,并且所述至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:基于所述目标语音的语音特征和所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数。在一些示例中,语音特征可以为自相关特征,由于自相关特征与语音音素强相关,音素与嘴型强相关,因而能够获得较好的模型推理结果。可以理解的是,所述语音特征也可以为其他特征,诸如MFCC特征。例如,对于每一个大小为385ms的语音窗口,把语音分成64个语音片段,对每个语音片段提取32个分量系数,组成64x32维的特征作为模型的语音特征输入。选择MFCC特征的原因是其同样与语音音素强相关,并且对固定音色的表达能力更好。由此,提供了在模型推理时以语音特征输入为主导,并辅以风格向量的选择与控制机制来确定用于驱动数字人的唇动与身体姿态风格的目标口型系数和目标姿态系数的实施方式,从而使目标口型和姿态系数的选取更准确。

[0058] 在一些实施例中,获取所述目标语音的语音特征包括获取所述目标语音所包括的一个或多个帧;以及基于所述一个或多个帧,获取所述目标语音的语音特征。由此,相比采用卷积等特征聚合手段而言,基于目标语音所包括的一个或多个帧来获取目标语音的语音特征可提供逐帧的语音特征向量,以便于训练数据集(其包括由逐帧语音或逐帧的语音特征向量以及与之对应的逐帧的口型系数和姿态系数真值所构成的训练数据对)的置备,并促成对3D数字人唇动和半身姿态展示动画的逐帧驱动。

[0059] 在一些实施例中,所述第一风格向量为从风格向量库中所获取的一个风格向量。在相关技术中,通常预先录制一段视频以用作训练集,视频中每一帧的语音和画面用作预训练模型的输入和输出,以期模型可基于视频训练集来学习到语音与身体姿态和唇动之间的关联关系,即,端到端映射关系。然而,为了获得尚可的模型推理结果,相关技术中的此类方法不仅需要大量高时长训练视频,也不可避免地使训练样本中各种身体姿态风格混合在一起,从而使模型所习得的针对一个发音单元的关联关系中包含各种不同身体姿态风格的对其的影响。由此,与相关技术中因直接端到端的映射学习方法所致的模型推理阶段3D数字人所展示出的半身姿态风格糅杂的状况相比,通过获取语音的与某一风格相对应的风格向量,从而能够提供半身姿态风格统一的3D数字人唇动和半身姿态展示动画。

[0060] 在一些实施例中,上述方法200还包括:获取所述目标语音相应的目标文本,所述目标文本包括多个发音单元;以及获取所述目标语音的每个发音单元相应的时间戳,其中所述时间戳中的每一者确定所述时间戳相应的发音单元的起始帧和结束帧,所述至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:基于所述目标语音所包括的多个发音单元各自相应的时间戳以及所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数。由此,通过时间戳来标识目标语音中每个发音单元所占用的时间窗口并确定每个发音单元的发音所持续的帧数,从而当在单次模型推理期间使用不止一个风格向量(例如,第一风格向量)进行推理时促成所选不同风格向量所跨越(例如,映射到)的帧与相应风格向量的对齐。

[0061] 在一些实施例中,第一风格向量包括从风格向量库中所获取的多个风格向量。由此,与相关技术中因直接端到端的映射学习方法所致的模型推理阶段数字人所展示出的半身姿态风格糅杂的状况相比,能够提供具有多种半身姿态风格的明确且精准切换的数字人唇动和半身姿态展示动画。

[0062] 图3示出了根据本公开的用于确定目标口型系数和目标姿态系数的方法的流程图,如图3所示,方法300可以包括步骤S310:基于所述目标语音自身的上下文信息,确定所述目标语音中与所述多个风格向量中的每一者相应的起始发音单元和结束发音单元;步骤S320:对于所述多个风格向量中的每一者,基于相应的起始发音单元和结束发音单元的时间戳,确定所述风格向量相应的起始帧和结束帧;以及步骤S330:基于所述多个风格向量、所述多个风格向量中的每一者的起始帧和结束帧,确定与所述目标语音在时间上逐帧地对齐的目标口型系数和目标姿态系数。根据一些实施例,所述多个风格向量用于使得所述3D数字人的身体姿态适配与所述多个风格向量相应的多个身体姿态风格。由此,通过基于目标语音自身的上下文来将目标语音分割成不同语段并使这些不同语段在时间上与多种不同风格向量中相应一种风格向量对齐,来致使模型推理出风格有所差异的逐帧对齐的目标口型和姿态系数对,从而藉此驱动数字人来展示出与所分割的不同语段的音频播放相适应的身体姿态风格。

[0063] 在一些实施例中,所述目标语音的上下文信息包括有关下列项中的一者或多者的信息:张口音、闭口音、爆破音、重复语段、语气词、转折词。由此,可基于目标语音中较为显著的表现特征(例如,长闭口、重复语段导致的姿态幅度递进、语气词导致的姿态强弱、转折词导致的姿态停顿等等)来在模型推理之前预先确定候选风格向量,并且为最终选取馈送给模型进行推理的风格向量提供了预设准则。

[0064] 在一些实施例中,上述方法中的任一者还包括:获取目标情绪系数;以及基于目标情绪系数调整所述目标姿态系数,以调整所述数字人的身体姿态幅度,所述基于所述目标口型系数和目标姿态系数控制所述数字人包括:基于所述目标口型系数和调整后的目标姿态系数控制所述数字人。目标情绪系数可以是馈送给模型的附加输入,其在模型的训练阶段可以或者可以不用作模型的输入。由此,当出现风格向量库中的所有风格向量都未能满足应用场景上下文和/或预设准则的情况时,提供了用于调整数字人身体姿态风格(例如,身体动作类型、幅度、方向等)的冗余控制机制。

[0065] 在一些实施例中,从风格向量库中获取与所述目标语音相应的第一风格向量包括:获取与所述目标语音相关的预设发音单元;从风格向量库中获取所述预设发音单元相应的多个风格向量;基于所述多个风格向量确定所述预设发音单元相应的多个口型系数和所述预设发音单元相应的多个姿态系数;基于所述多个口型系数和所述多个姿态系数控制数字人;以及基于相应的所述多个口型系数和所述多个姿态系数所得到的数字人的控制结果,从所述多个风格向量中确定所述第一风格向量。由此,可通过在模型训练阶段与模型推理阶段之间的测试时间段期间基于目标语音中预设发音单元(例如,爆破音)来遍历风格向量库中的风格向量并针对数字人进行仿真,从而选取仿真测试结果最优和/或最合需的风格向量以便更好地针对目标语音进行模型推理。

[0066] 在一些实施例中,至少基于所述第一风格向量,确定与所述目标语音相应的目标口型系数和目标姿态系数包括:至少将所述第一风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的与所述目标语音相应的目标口型系数和目标姿态系数。需要注意的是,所述身体姿态与唇动模型可以是本领域已知的任何合适的机器学习模型、深度学习模型、神经网络等,它的一个非限制性示例可以是下文参考图6所描述的身体姿态与唇动模型600。

[0067] 图4示出了根据本公开的实施例的用于数字人的模型训练方法的流程图,如图4所示,方法400可以包括步骤S410:获取样本语音以及所述样本语音所包括的至少一个发音单元;步骤S420:获取与所述样本语音相应的真实口型系数和真实姿态系数;步骤S430:获取所述样本语音所包括的每个发音单元,以及与每个发音单元相应的起始帧和结束帧;步骤S440:确定所述样本语音相应的风格向量,其中所述风格向量包括所述至少一个发音单元各自相应的子风格向量,所述样本语音中的每个发音单元的起始帧与结束帧之间的所有帧共用相同的子风格向量;步骤S450:将所述样本语音和所述至少一个发音单元相应的至少一个子风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的预测口型系数和预测姿态系数;以及步骤S460:基于所述真实口型系数和预测口型系数之间的损失以及所述真实姿态系数和预测姿态系数之间的损失,更新所述身体姿态与唇动模型的网络参数,并且更新所述风格向量。由此,通过辅以训练语音中每个字相应的风格向量以使其用于存储风格差异性,将训练语音(不论是TTS输入还是真人音频输入)馈送给网络以获取网络推理出的两种系数(逐帧),并将系数对齐之后直接用于驱动数字人,从而能够使得该数字人针对所讲的话具有口型和身体姿态(例如,关节)的恰适风格变化。

[0068] 在一些实施例中,所述风格向量中所包括的每个子向量的元素是按照高斯分布来初始化的。由此,风格向量能够在模型训练阶段的误差传播(例如,loss梯度更新)/网络参数更新期间逐渐地存储当给定相应字的多个正常发音状态时的不同半身姿态真值之间的差异性,进而保存对一种风格的记录。

[0069] 在一些实施例中,在所述数字人身体姿态与唇动模型的训练完成之后,基于经更新的风格向量确定风格向量库。由此,可为模型推理阶段所用到的风格向量输入提供候选池且有助于在模型推理前的测试阶段对多个候选风格向量的检索之用途。

[0070] 图5示出了根据本公开的驱动数字人的过程500的示意图。

[0071] 过程500包括视频录制501、视频数据获取502、音频数据获取503、数据处理504、语音特征处理505、身体姿态/口型系数真值获取506、网络训练507、以及数字人驱动508。

[0072] 在视频录制501中,可使用普通摄像头录制模特身体姿态视频(例如,半身视频),模特在朗读内置的文本话术的同时其上半身可做自然运动来表达当前的内容。其中,内置的文本话术例如为120条包含各种发音音节的中文语句,当然本公开可采用含任何语言/音节的内置文本话术,而不偏离本公开的保护范围。身体(例如,半身)动作可以为与内容和情绪相关的手臂与手势以及头部(点头/摇头)自然运动。在录制画面的同时需要同步记录模特声音信息。

[0073] 在视频数据获取502,获取并保存从视频录制501中录制的逐帧的视频。

[0074] 在音频数据获取503,获取并保存从视频录制501中录制的逐帧的音频。通常,视频数据和音频数据是帧对齐的,当出现某帧的视频或音频缺失时,可选择丢弃该帧。

[0075] 在数据处理504中,示例方案采用3D人体参数化模型和3D人脸参数化模型来分别进行人体身体姿态重建和人脸面部重建。通过对数字人复用重建后的姿态系数(例如,pose系数)和口型系数(例如,blendshape系数),可以用低成本实现数字人的身体动作驱动和口型驱动。相关技术的方案包括基于相机阵列的真人3D数字高精形象面部扫描重建,通过音频与blendshape系数的数据组合来进行3D唇动模型的训练。当实际应用时,通过输入角色的音频,直接输出人脸参数化模型系数,进而形成3D人脸口型动画。这种基于相机阵列的真

人3D数字高精形象面部扫描重建需要依赖昂贵的采集设备、高配的电脑硬件、复杂繁重的解算流程,导致普用性较差。另外,由于整体流程无法同步完成半身动作的驱动,数字人整体自然度和真实度较低。

[0076] 在示例方案中,3D人体参数化模型通过建模的方式来获得,例如,可采集50个人(每个个体可具有高矮胖瘦的差异性)的形貌数据。当需要重建某个人时,从采集到的50个人中,通过“插值”的方式把该人表现出来(譬如某个人的腿型或者某几个人的腿型的混合后的效果跟该人的腿型是相似的,或者某一个人的身高或者某几个人的身高的混合后的结果跟该人是相似的,包括手臂的粗细等等)。接着,通过多幕的方式并利用三角测量,可以获取身体中主要关节在空间中的位置(通过人体关键点识别算法),在获得该位置以后便可采用3D人体参数化模型去表达,即,只需要让参数化模型本身(该参数化模型是提前做好绑定的,即,其内部初始地具有关节)与通过三角测量测得的关节进行匹配,匹配后即可大致对齐。在对齐(拟合)后,除了得到身体的形状系数,更重要的是得到姿态系数。在一些示例中,姿态系数表征关节的角度变化。该姿态系数用于输入语音相应的(逐帧)姿态系数真值。

[0077] 在示例方案中,人脸参数化模型由一系列形状组成,当期望重建某个人的唇部或面部的一些动作时,也通过“插值”来得到。此处需要注意的是,口型系数的拟合流程与姿态系数的拟合流程基本上一致。口型系数可以是广义的,包括眉眼的变化等在内。在一些示例中,下眼睑以下的部位落入口型系数的表征范围内。

[0078] 需要注意的是,采用3D人体参数化模型和3D人脸参数化模型分别进行人体半身姿态重建和人脸面部重建属于离线数据采集阶段,目的是获得训练数据对,以供后续训练阶段使用。

[0079] 在语音特征处理505中,对训练语音进行特征提取和计算以得到供训练模型的较高层使用的语音特征。

[0080] 在身体姿态/口型系数真值获取506中,获取并保存从数据处理504得到的身体姿态/口型系数真值。当获得身体姿态/口型系数真值之后,视频数据获取502中所保存的逐帧的视频可被丢弃。

[0081] 在网络训练507中,利用训练数据对(其包括输入和输出)来对网络进行训练,如上文所描述。

[0082] 在数字人驱动508中,将推理数据(其包括目标语音和第一风格向量)输入训练好的网络,以得到目标身体姿态/口型系数用以驱动数字人进行与目标语音的播放同步的动画展示。

[0083] 图6示出了根据本公开的身体姿态与唇动模型600的框图。

[0084] 模型600包括模型输入601、语音子模型602、唇动子模型603、身体姿态子模型604以及模型输出605。

[0085] 根据一些实施例,语音子模型可包括Formant Analysis Network(共振峰分析网络)、Articulation Network(发音网络)AwareAttention network(关注网络)等;唇动子模型可包括AwareAttention network(关注网络)、Face Fully Connected Layer(面部全连接层)、Face Output Layer(面部输出层)等;并且身体姿态子模型可包括AwareAttention network(关注网络)、Face Output Layer(面部输出层)、Body Fully Connected Layer(身体全连接层)、Body Output Layer(身体输出层)等。

[0086] 发明人发现唇动子模型的结果可以用来表征语音中的节奏信息,相较于语音特征的输出更能高效表现爆破音等特征。在爆破音情况下,身体动作也会有明显的节奏与幅度。因此,在一优选实施方式中,可将语音子模型的输出与唇动子模型的输出两者联合作为身体姿态子模型的输入,从而有效提升爆破音、急促音下的身体动作的匹配度。

[0087] 注意,模型输入601中模型输出605中所用到的涉及模型训练和模型推理两个阶段的各种输入和输出数据已在上文进行了详细描述,在此不再赘述。

[0088] 根据本公开的另一方面,参见图7,还提供了一种数字人的控制装置700,包括:第一获取模块701,所述第一获取模块701被配置获取目标语音;第二获取模块702,所述第二获取模块702被配置为从风格向量库中获取与目标语音相应的第一风格向量,所述风格向量库包括与目标语音相应的不同于第一风格向量的第二风格向量;确定模块703,所述确定模块703被配置为至少基于所述第一风格向量,确定与目标语音相应的目标口型系数和目标姿态系数;以及控制模块704,所述控制模块704被配置为基于所述目标口型系数和目标姿态系数控制所述数字人,以使得所述数字人的身体姿态适配于所述第一风格向量相应的身体姿态风格,并且所述数字人的唇动适配所述目标语音。

[0089] 根据本公开的另一方面,参见图8,还提供一种用于数字人的模型训练装置800,包括:第三获取模块801,所述第三获取模块801被配置为获取样本语音以及所述样本语音所包括的至少一个发音单元;第四获取模块802,所述第四获取模块802被配置为获取与样本语音相应的真实口型系数和真实姿态系数;第五获取模块803,所述第五获取模块803被配置为获取所述样本语音所包括的每个发音单元,以及与每个发音单元相应的起始帧和结束帧;以及第一确定模块804,所述第一确定模块804被配置为确定所述样本语音相应的风格向量,其中所述风格向量包括所述至少一个发音单元各自相应的子风格向量,所述样本语音中的每个发音单元的起始帧与结束帧之间的所有帧共用相同的子风格向量;第二确定模块805,所述第二确定模块805被配置为将所述样本语音和所述至少一个发音单元相应的至少一个子风格向量输入身体姿态与唇动模型,并获取所述身体姿态与唇动模型所输出的预测口型系数和预测姿态系数;以及更新模块806,所述更新模块806被配置为基于所述真实口型系数和预测口型系数之间的损失以及所述真实姿态系数和预测姿态系数之间的损失,更新所述身体姿态与唇动模型的网络参数,并且更新所述风格向量。

[0090] 根据本公开的实施例,还提供了一种能够实现上述方法中的任一者的电子设备、可读存储介质和计算机程序产品。

[0091] 本公开所提出的支持身体姿态风格切换的语音驱动数字人身体动作以及讲话的方法和系统,属于低成本且具有强泛化效果的数字人身体动作和口型驱动方案,通过捕捉语音信息中细粒度的特征变化,支持身体姿态风格切换的同时产生合理自然的身体动作以及口型效果。这种低成本且强泛化的驱动方案,能够直接用于直播场景以替换复杂的面部口型捕捉设备,极大降低了前期设备投入成本以及后期动画修改的人力成本。

[0092] 参考图9,现将描述可以作为本公开的服务器或客户端的电子设备的结构框图,其是可以应用于本公开的各方面的硬件设备的示例。电子设备旨在表示各种形式的数字电子的计算机设备,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作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0093] 如图9所示,设备900包括计算单元901,其可以根据存储在只读存储器(ROM) 902中的计算机程序或者从存储单元808加载到随机访问存储器(RAM) 903中的计算机程序,来执行各种适当的动作和处理。在RAM903中,还可存储设备9900操作所需的各种程序和数据。计算单元901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出(I/O) 接口905也连接至总线904。

[0094] 设备900中的多个部件连接至I/O接口905,包括:输入单元906、输出单元807、存储单元808以及通信单元909。输入单元906可以是能向设备9900输入信息的任何类型的设备,输入单元906可以接收输入的数字或字符信息,以及产生与电子设备的用户设置和/或功能控制有关的键信号输入,并且可以包括但不限于鼠标、键盘、触摸屏、轨迹板、轨迹球、操作杆、麦克风和/或遥控器。输出单元907可以是能呈现信息的任何类型的设备,并且可以包括但不限于显示器、扬声器、视频/音频输出终端、振动器和/或打印机。存储单元908可以包括但不限于磁盘、光盘。通信单元909允许设备900通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据,并且可以包括但不限于调制解调器、网卡、红外通信设备、无线通信收发机和/或芯片组,例如蓝牙TM设备、1302.11设备、WiFi设备、WiMax设备、蜂窝通信设备和/或类似物。

[0095] 计算单元901可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元901的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI) 计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元901执行上文所描述的各个方法和处理,例如三维人脸的唇动控制方法。例如,在一些实施例中,方法三维人脸的唇动控制方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元908。在一些实施例中,计算机程序的部分或者全部可以经由ROM 902和/或通信单元909而被载入和/或安装到设备900上。当计算机程序加载到RAM 903并由计算单元901执行时,可以执行上文描述的三维人脸的唇动控制方法的一个或多个步骤。备选地,在其他实施例中,计算单元901可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行三维人脸的唇动控制方法。

[0096] 本文中以上描述的系统和技术和各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、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。

[0097] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理单元或控制器,使得程序代码当由处理单元或控制器执行时使流程图和/或框图所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件

包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0098] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0099] 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,CRT(阴极射线管)或者LCD(液晶显示器)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。

[0100] 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(LAN)、广域网(WAN)和互联网。

[0101] 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器的关系。

[0102] 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,本公开中记载的各步骤可以并行地执行、也可以顺序地或以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0103] 虽然已经参见附图描述了本公开的实施例或示例,但应理解,上述的方法、系统和设备仅仅是示例性的实施例或示例,本发明的范围并不由这些实施例或示例限制,而是仅由授权后的权利要求书及其等同范围来限定。实施例或示例中的各种要素可以被省略或者可由其等同要素替代。此外,可以通过不同于本公开中描述的次序来执行各步骤。进一步地,可以以各种方式组合实施例或示例中的各种要素。重要的是随着技术的演进,在此描述的很多要素可以由本公开之后出现的等同要素进行替换。

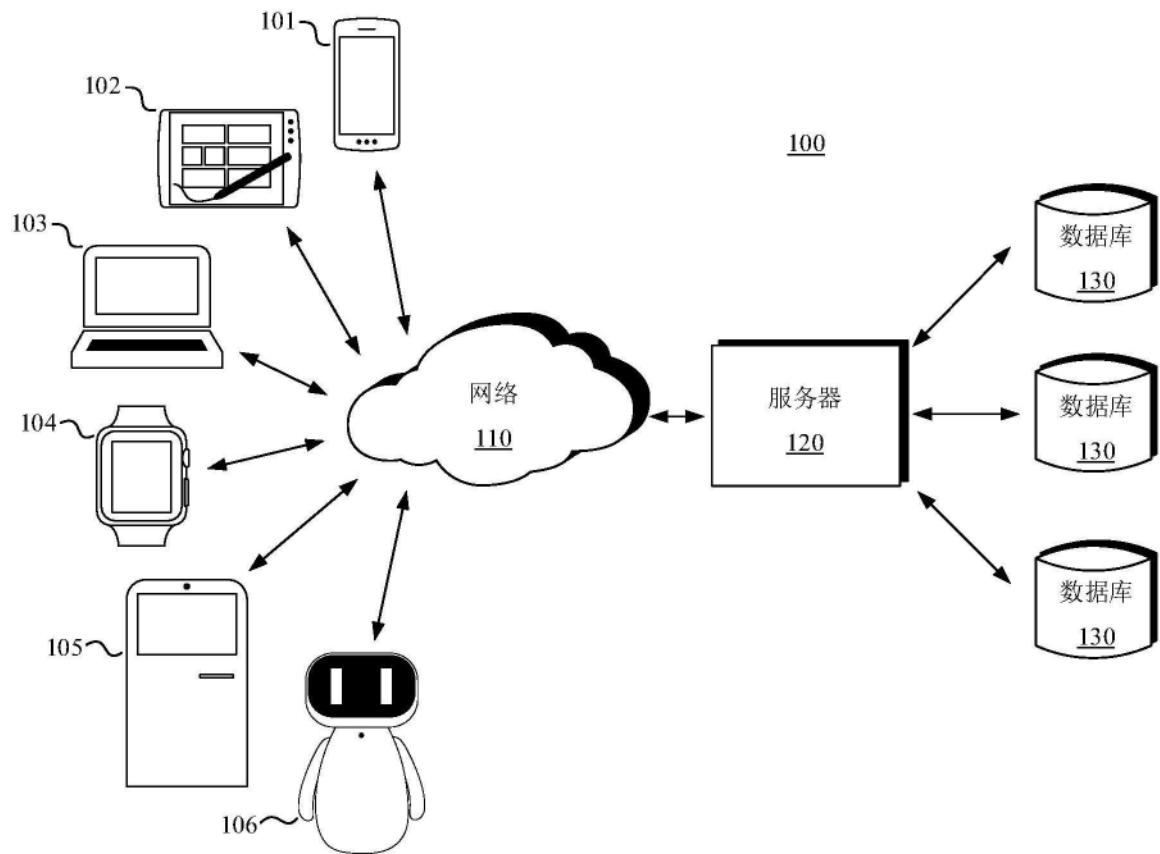


图1

200

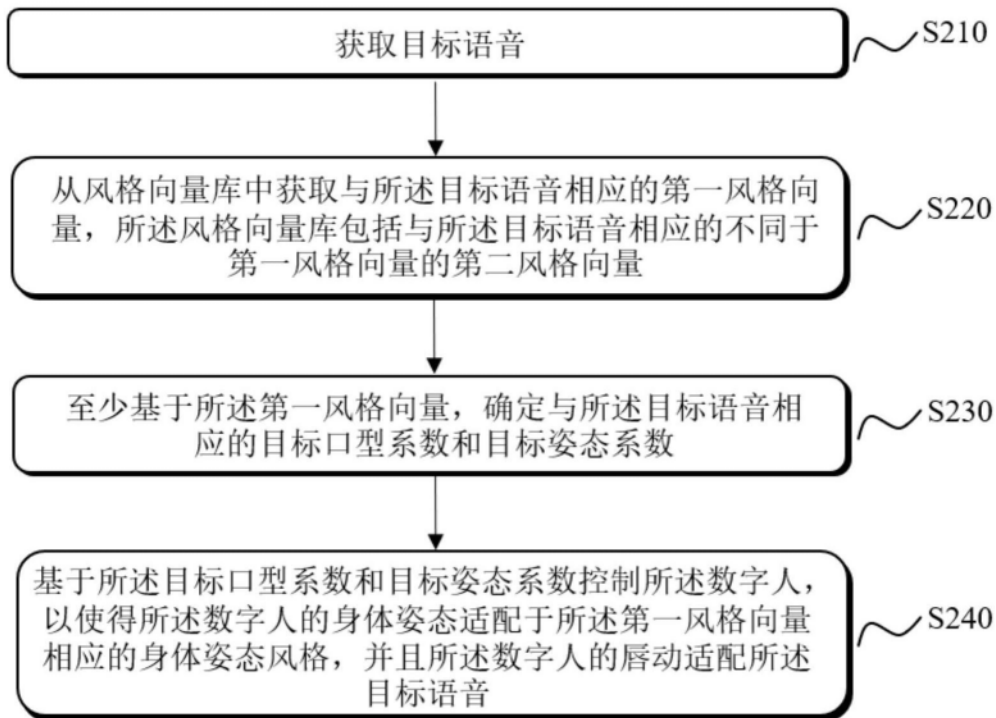


图2

300

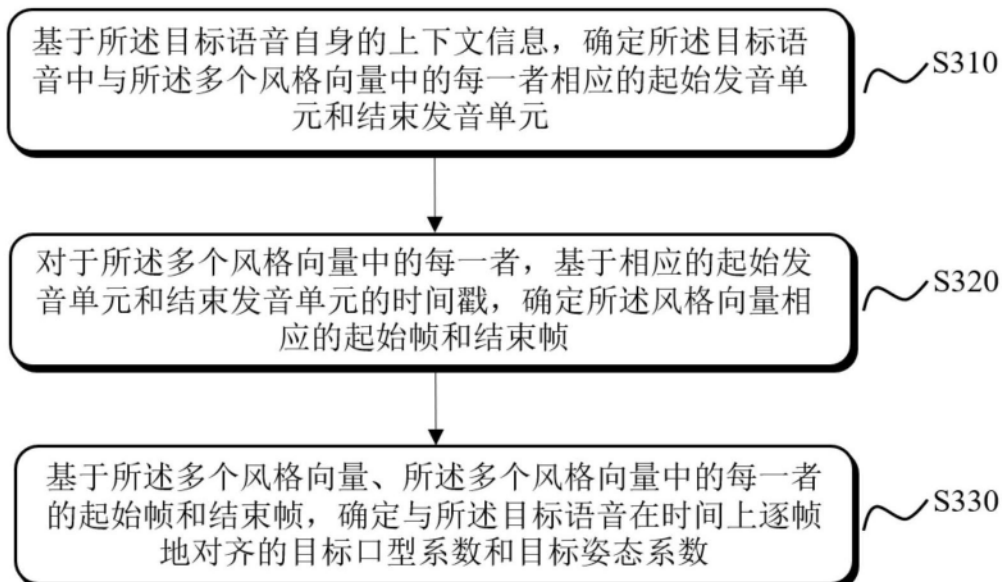


图3

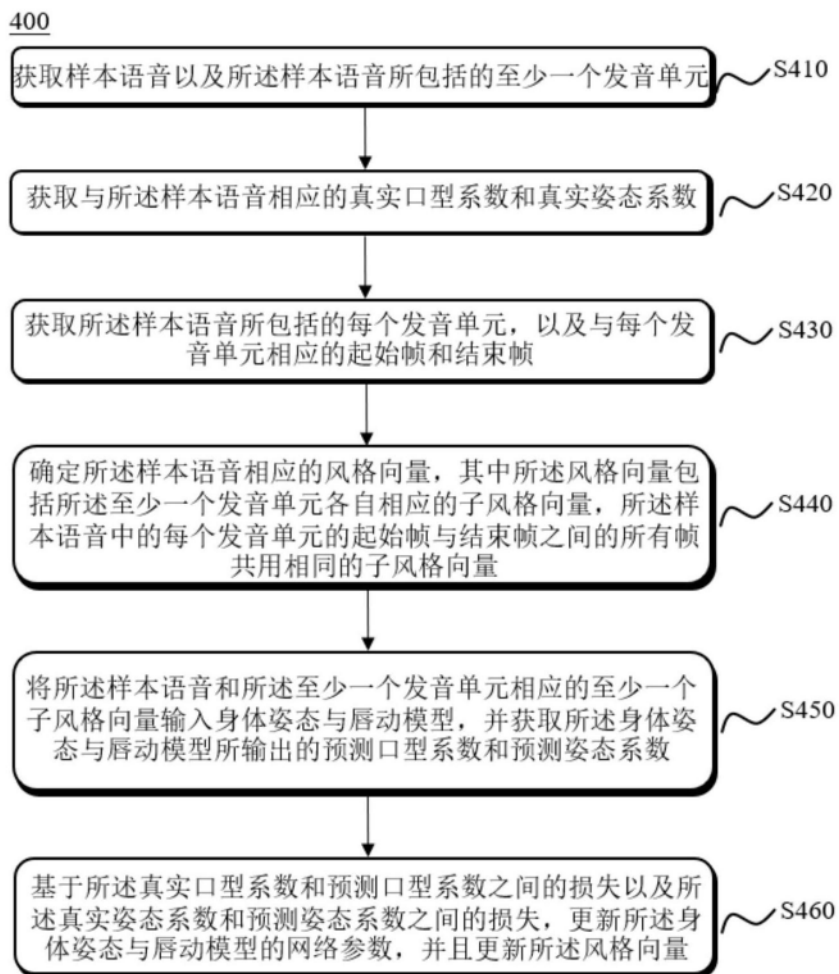


图4

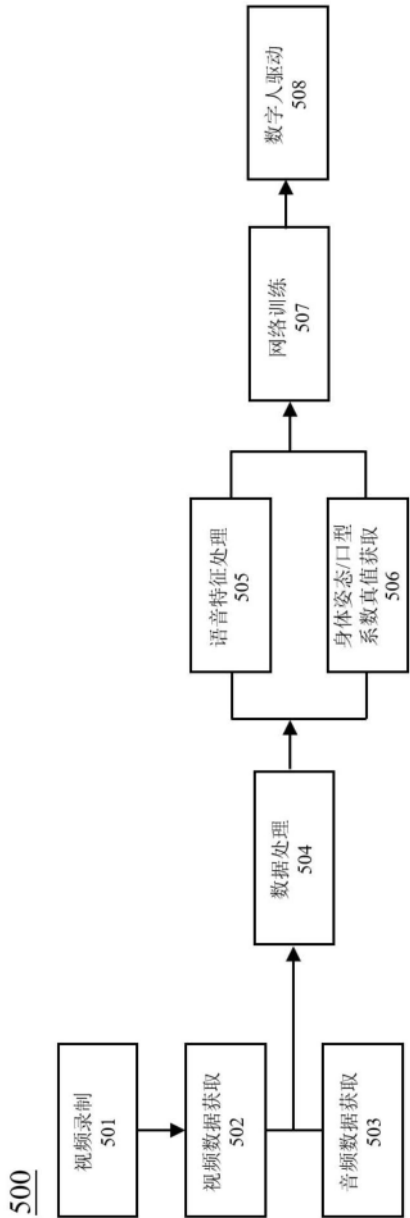


图5

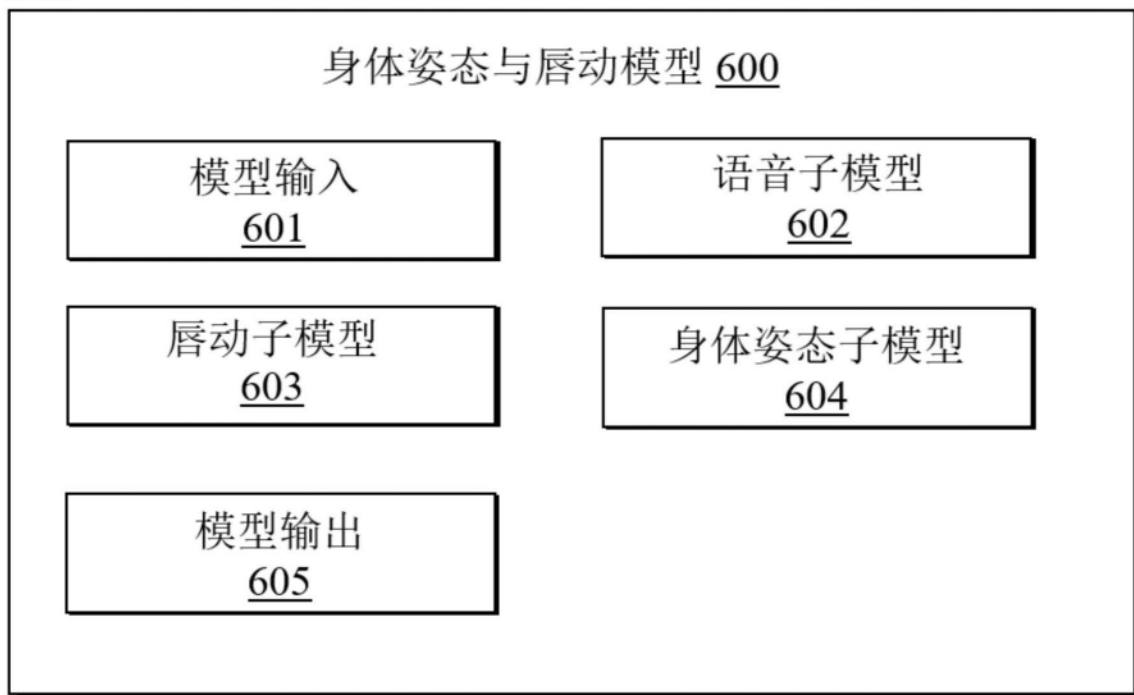
600

图6

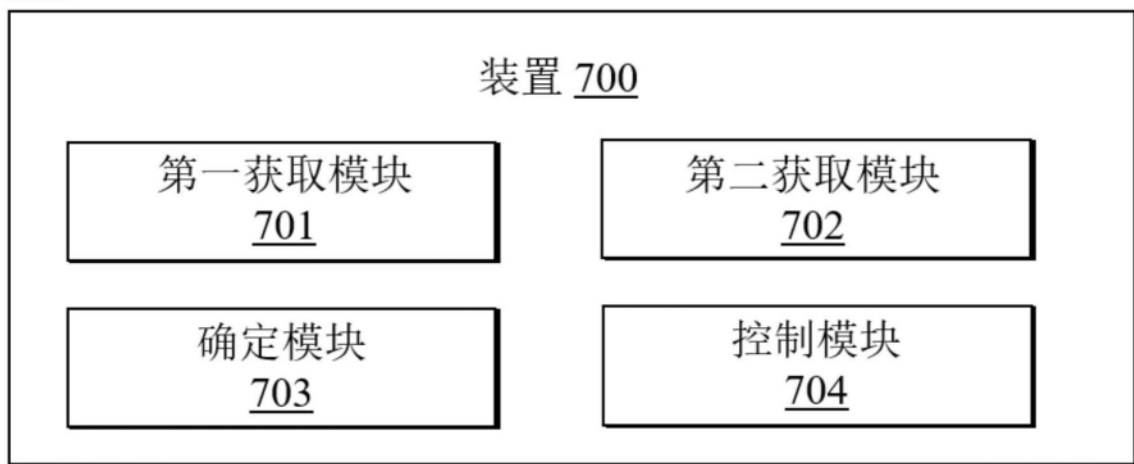
700

图7

800

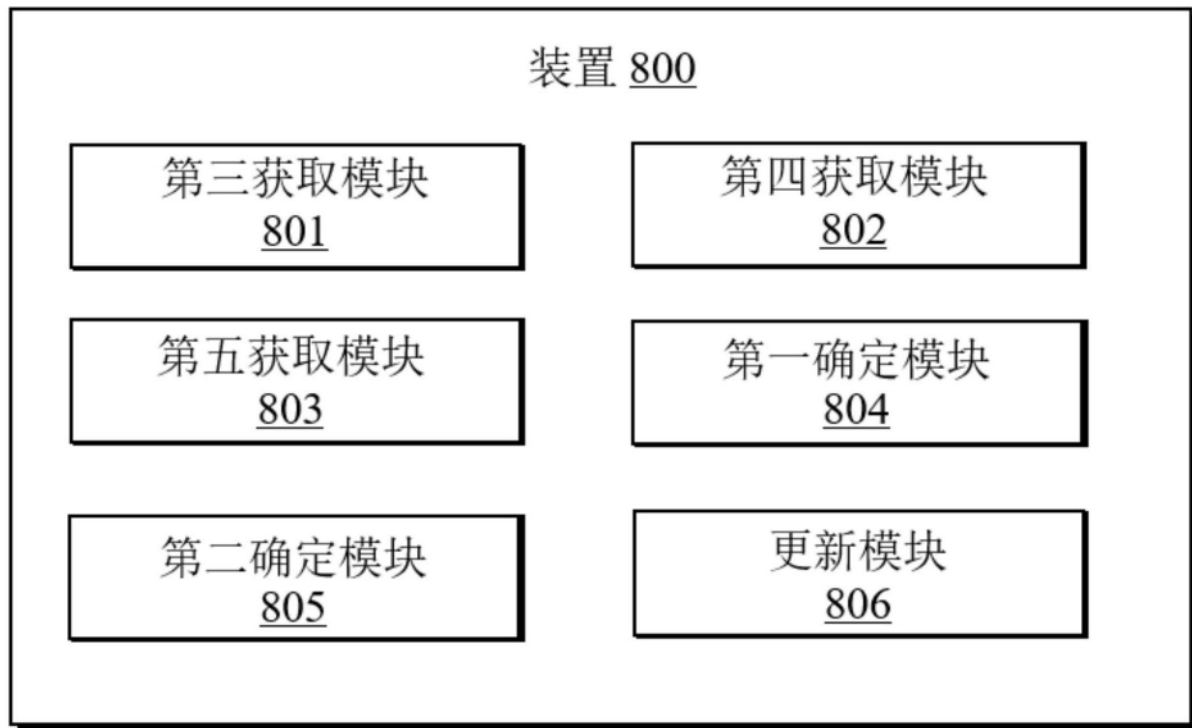


图8

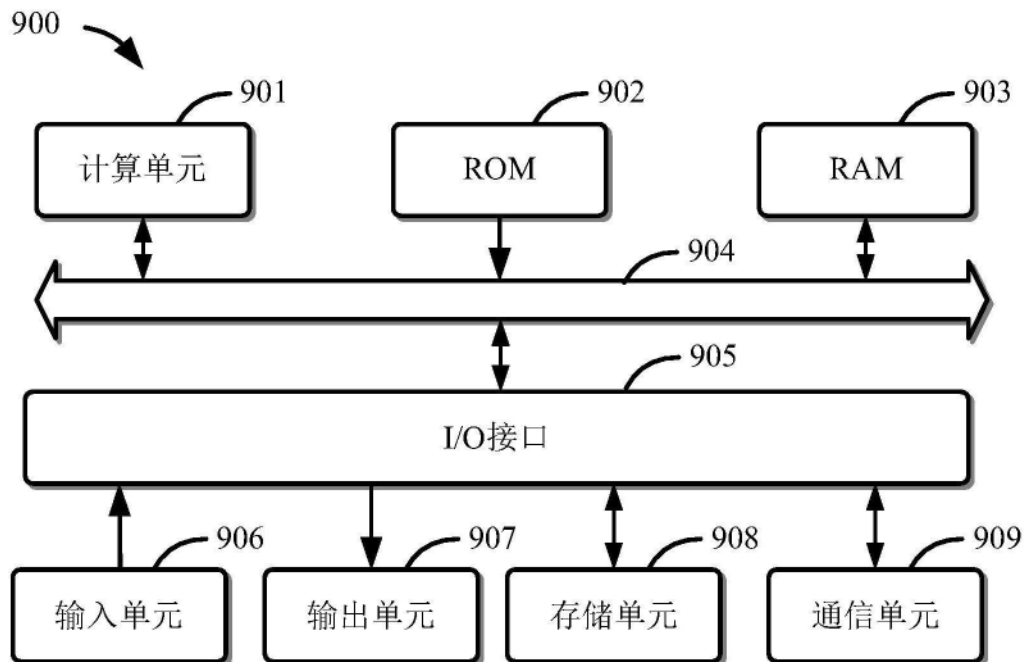


图9