



(12) 发明专利申请

(10) 申请公布号 CN 116524929 A

(43) 申请公布日 2023. 08. 01

(21) 申请号 202310552378.1

(22) 申请日 2023.05.16

(71) 申请人 京东科技信息技术有限公司

地址 100176 北京市大兴区北京经济技术
开发区科创十一街18号院2号楼6层
601

(72) 发明人 杨久东 陈蒙 冯明超 王培英

(74) 专利代理机构 北京律智知识产权代理有限
公司 11438

专利代理师 孙宝海

(51) Int. Cl.

G10L 15/22 (2006.01)

G10L 15/16 (2006.01)

G10L 15/02 (2006.01)

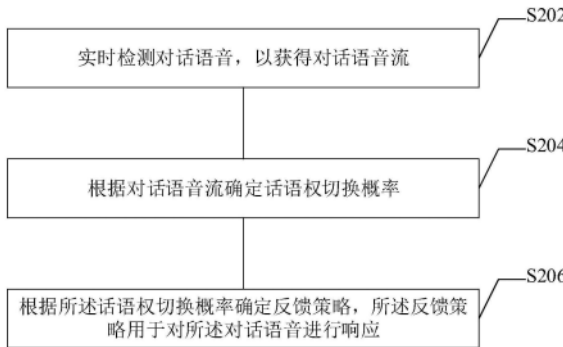
权利要求书2页 说明书12页 附图8页

(54) 发明名称

人机对话方法、装置、电子设备和计算机可读存储介质

(57) 摘要

本公开提供一种人机对话方法、装置、电子设备和计算机可读存储介质,涉及语音处理技术领域。人机对话方法包括:实时检测对话语音,以获得对话语音流;根据所述对话语音流确定话语权切换概率;根据所述话语权切换概率确定反馈策略,所述反馈策略用于对所述对话语音进行响应。本公开实施例可以提高人机对话过程中的流畅性,从而提高用户体验。



1. 一种人机对话方法,其特征在于,包括:
实时检测对话语音,以获得对话语音流;
根据所述对话语音流确定话语权切换概率;
根据所述话语权切换概率确定反馈策略,所述反馈策略用于对所述对话语音进行响应。

2. 根据权利要求1所述方法,其特征在于,根据所述对话语音流确定话语权切换概率,包括:

确定所述对话语音流的间断时间超过第一阈值;
从所述对话语音间断处向前截取目标时间段的对话语音;
对所述目标时间段的对话语音进行特征提取,以获得所述目标时间段的对话语音的音频特征;

根据所述目标时间段的对话语音的音频特征确定所述话语权切换概率。

3. 根据权利要求2所述方法,其特征在于,根据所述目标时间段的对话语音的音频特征确定所述话语权切换概率,包括:

通过目标神经网络的特征提取结构对所述目标时间段的对话语音的音频特征进行特征提取,以确定所述目标时间段的对话语音的语义特征;

通过所述目标神经网络的第一分支结构对所述目标时间段的对话语音的语义特征进行处理,确定所述话语权切换概率。

4. 根据权利要求3所述方法,其特征在于,所述根据所述话语权切换概率确定反馈策略,包括:

通过所述第一分支结构确定所述话语权切换概率大于第二阈值且小于或者等于第三阈值;

通过所述目标神经网络的第二分支结构对所述目标时间段的对话语音的语义特征进行处理,以确定是否给出响应词;

在根据所述目标时间段的对话语音的语义特征确定给出响应词时,在预设的多个所述响应词中随机确定一目标响应词;

基于所述目标响应词,对所述对话语音进行响应。

5. 根据权利要求3所述方法,其特征在于,所述根据所述话语权切换概率确定反馈策略,包括:

通过所述第一分支结构确定所述话语权切换概率所述话语权切换概率大于第三阈值;

通过所述目标神经网络的第二分支结构对所述目标时间段的对话语音的语义特征进行处理,确定预设的多个填充词对应的发出概率;

将发出概率最大的填充词作为目标填充词;

基于所述目标填充词,对所述对话语音进行响应。

6. 根据权利要求5所述方法,其特征在于,所述方法还包括:

根据所述话语权切换概率确定话语权切换损失;

根据是否给出响应词确定响应词损失;

根据所述各个填充词的发出概率确定填充词损失;

根据所述话语权切换损失、所述响应词损失和所述填充词损失确定目标损失;

通过所述目标损失对所述目标神经网络进行训练。

7. 根据权利要求2所述方法, 其特征在于, 对所述目标时间段的对话语音进行特征提取, 以获得所述目标时间段的对话语音的音频特征, 包括:

对所述目标时间段的对话语音进行特征提取;

确定所述目标时间段的对话语音的音调特征、强度特征、功率特征和频谱特征。

8. 一种人机对话装置, 其特征在于, 包括:

语音检测模块, 用于实时检测对话语音, 以获得对话语音流;

话语权切换概率确定模块, 用于根据所述对话语音流确定话语权切换概率;

响应词发出模块, 用于根据所述话语权切换概率确定反馈策略, 所述反馈策略用于对所述对话语音进行响应。

9. 一种电子设备, 其特征在于, 包括:

存储器和处理器;

所述存储器用于存储程序指令; 所述处理器调用所述存储器存储的所述程序指令, 用于实现如权利要求1-7任一项所述的人机对话方法。

10. 一种计算机可读存储介质, 其上存储有程序指令, 该程序指令被处理器执行时实现如权利要求1-7任一项所述的人机对话方法。

人机对话方法、装置、电子设备和计算机可读存储介质

技术领域

[0001] 本公开涉及语音处理技术领域,尤其涉及一种人机对话方法、装置、电子设备和计算机可读存储介质。

背景技术

[0002] 在人机对话过程中,对话流畅性是语音对话系统中至关重要的环节,怎样使得对话更流畅且自然地进行,是所有语音对话系统面临的一个难点。

[0003] 因此,如何在人机对话过程中保证对话流畅性是本发明要解决的技术问题。

[0004] 需要说明的是,在上述背景技术部分公开的信息仅用于加强对本公开的背景的理解。

发明内容

[0005] 本公开的目的在于提供一种人机对话方法、装置、电子设备以及计算机可读存储介质,能够提高人机对话过程的流畅度。

[0006] 本公开的其他特性和优点将通过下面的详细描述变得显然,或部分地通过本公开的实践而习得。

[0007] 本公开实施例提供了一种人机对话方法,包括:实时检测对话语音,以获得对话语音流;根据所述对话语音流确定话语权切换概率;根据所述话语权切换概率确定反馈策略,所述反馈策略用于对所述对话语音进行响应。

[0008] 在一些实施例中,根据所述对话语音流确定话语权切换概率,包括:确定所述对话语音流的间断时间超过第一阈值;从所述对话语音间断处向前截取目标时间段的对话语音;对所述目标时间段的对话语音进行特征提取,以获得所述目标时间段的对话语音的音频特征;根据所述目标时间段的对话语音的音频特征确定所述话语权切换概率。

[0009] 在一些实施例中,根据所述目标时间段的对话语音的音频特征确定所述话语权切换概率,包括:通过目标神经网络的特征提取结构对所述目标时间段的对话语音的音频特征进行特征提取,以确定所述目标时间段的对话语音的语义特征;通过所述目标神经网络的第一分支结构对所述目标时间段的对话语音的语义特征进行处理,确定所述话语权切换概率。

[0010] 在一些实施例中,所述反应词包括多个响应词;其中,根据所述话语权切换概率确定反馈策略,包括:通过所述第一分支结构确定所述话语权切换概率大于第二阈值且小于或者等于第三阈值;通过所述目标神经网络的第二分支结构对所述目标时间段的对话语音的语义特征进行处理,以确定是否给出响应词;在根据所述目标时间段的对话语音的语义特征确定给出响应词时,在预设的多个所述响应词中随机确定一目标响应词;基于所述目标响应词,对所述对话语音进行响应。

[0011] 在一些实施例中,所述反应词包括多个填充词;其中,根据所述话语权切换概率确定反馈策略,包括:通过所述第一分支结构确定所述话语权切换概率所述话语权切换概率

大于所述第三阈值;通过所述目标神经网络的第三分支结构对所述目标时间段的对话语音的语义特征进行处理,确定预设的多个填充词对应的发出概率;将发出概率最大的填充词作为目标填充词;基于所述目标填充词,对所述对话语音进行响应。

[0012] 在一些实施例中,所述方法还包括:根据所述话语权切换概率确定话语权切换损失;根据是否给出响应词确定响应词损失;根据所述各个填充词的发出概率确定填充词损失;根据所述话语权切换损失、所述响应词损失和所述填充词损失确定目标损失;通过所述目标损失对所述目标神经网络进行训练。

[0013] 在一些实施例中,对所述目标时间段的对话语音进行特征提取,以获得所述目标时间段的对话语音的音频特征,包括:对所述目标时间段的对话语音进行特征提取;确定所述目标时间段的对话语音的音调特征、强度特征、功率特征和频谱特征。

[0014] 本公开实施例提供了一种人机对话装置,包括:语音检测模块、话语权切换概率确定模块和响应词发出模块。

[0015] 其中,所述语音检测模块用于实时检测对话语音,以获得对话语音流;所述话语权切换概率确定模块可以用于根据所述对话语音流确定话语权切换概率;所述响应词发出模块可以用于根据所述话语权切换概率确定反馈策略,所述反馈策略用于对所述对话语音进行响应。

[0016] 本公开实施例提出一种电子设备,该电子设备包括:存储器和处理器;所述存储器用于存储程序指令;所述处理器调用所述存储器存储的所述程序指令,用于实现上述任一项所述的人机对话方法。

[0017] 本公开实施例提出一种计算机可读存储介质,其上存储有计算机程序,所述程序被处理器执行时实现如上述任一项所述的人机对话方法。

[0018] 本公开实施例提出一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述人机对话方法。

[0019] 本公开实施例提供的人机对话方法、装置、电子设备和计算机可读存储介质,通过对目标对象的对话语音进行实时检测,以预测该对话语音的话语权切换概率,从而使得电子设备根据该话语权切换概率确定反馈策略,以便在口语对话过程中向目标对象给出反应,从而使得口语对话更加流畅,提高人机对话的流畅性和客户体验。

[0020] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性的,并不能限制本公开。

附图说明

[0021] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理。显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0022] 图1示出了可以应用于本公开实施例的人机对话方法或人机对话装置的示例性设备的示意图。

- [0023] 图2是根据一示例性实施例示出的一种人机对话方法的流程图。
- [0024] 图3是根据一示例性实施例示出的一种人机对话方法的流程图。
- [0025] 图4是根据一示例性实施例示出的一种人机对话方法的流程图。
- [0026] 图5是根据一示例性实施例示出的一种目标神经网络的示意图。
- [0027] 图6是根据一示例性实施例示出的一种人机对话方法的流程图。
- [0028] 图7是根据一示例性实施例示出的一种通过话语权切换概率将话语权切换时机、响应词发出时机和填充词发出时间串联起来的示意图。
- [0029] 图8是根据一示例性实施例示出的一种神经网络方法的流程图。
- [0030] 图9是根据一示例性实施例示出的一种人机对话装置的框图。
- [0031] 图10示出了适于用来实现本公开实施例的电子设备的结构示意图。

具体实施方式

[0032] 现在将参考附图更全面地描述示例实施例。然而,示例实施例能够以多种形式实施,且不应被理解为限于在此阐述的实施例;相反,提供这些实施例使得本公开将全面和完整,并将示例实施例的构思全面地传达给本领域的技术人员。在图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。

[0033] 本公开所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中,提供许多具体细节从而给出对本公开的实施方式的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而省略特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知方法、装置、实现或者操作以避免模糊本公开的各方面。

[0034] 附图仅为本公开的示意性图解,图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。附图中所示的一些方框图不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0035] 附图中所示的流程图仅是示例性说明,不是必须包括所有的内容和步骤,也不是必须按所描述的顺序执行。例如,有的步骤还可以分解,而有的步骤可以合并或部分合并,因此实际执行的顺序有可能根据实际情况改变。

[0036] 在本申请的描述中,除非另有说明,“/”表示“或”的意思,例如,A/B可以表示A或B。本文中的“和/或”仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。此外,“至少一个”是指一个或多个,“多个”是指两个或两个以上。“第一”、“第二”等字样并不对数量和执行次序进行限定,并且“第一”、“第二”等字样也并无限定一定不同;用语“包含”、“包括”和“具有”用以表示开放式的包括在内的意思并且是指除了列出的要素/组成部分/等之外还可存在另外的要素/组成部分/等。

[0037] 为了能够更清楚地理解本发明的上述目的、特征和优点,下面结合附图和具体实施方式对本发明进行进一步的详细描述,需要说明的是,在不冲突的情况下,本申请的实施例及实施例中的特征可以相互组合。

[0038] 在介绍本申请实施例之前,可以对本申请实施例中涉及的动词或者名词进行解释。

[0039] 打断:也叫barge-in,指用户(如目标对象)说话时,系统也开始了讲话,系统语音和用户语音有重叠的情况。

[0040] 话语权决策:也叫endpointing,指用户(如目标对象)讲话时出现短暂暂停时,系统需要准确判断用户是否讲完话,如果用户讲完了则系统需要接过话语权,如果系统未讲完则系统不能抢话,且需要等待用户继续讲话。

[0041] 填充词:也叫fillerwords,主要填充在系统接过话语权之后、正式讲话之前,是一个简短回复,如“好的”,“是的”、“明白”等,由于系统语音处理中间耗时较高,填充词的主要作用在于快速响应,避免用户长时间等待。上述填充词可以用来缓解系统高延时的问题,同时提高口语对话的流畅度。

[0042] 响应词,也可称之为鼓励词,也叫backchannel,主要产生于用户讲话的过程中,给用户积极正向的响应,目的是为了鼓励用户继续讲话,但是并不会抢话,话语权仍然在用户侧。通过上述响应词可以向用户表示系统正在接收用户的语音信息,从而对用户进行激励,如“嗯”、“额”、“是的”等。

[0043] 本申请提出在人机对话过程中,设备会根据目标对象的对话语音及时给出响应词或反应词,以保证对话的流畅性并提高用户体验。

[0044] 本申请在话语权决策、填充词、响应词之间建立了关系,并将这种关系应用到真实的对话场景中。

[0045] 下面结合附图对本公开示例实施方式进行详细说明。

[0046] 图1示出了可以应用于本公开实施例的人机对话方法或人机对话装置的示例性设备的示意图。

[0047] 如图1所示,该示例性设备可以包括终端设备101,该终端设备可以用于与目标对象102进行人机对话。该终端设备可以通过网络与服务器(图1未示出)进行通信。网络可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0048] 其中,终端设备101可以是任意电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机、台式计算机、可穿戴设备、虚拟现实设备、智能家居等等。

[0049] 服务器可以是提供各种服务的服务器,例如对用户利用终端设备所进行操作的装置提供支持的后台管理服务器。后台管理服务器可以对接收到的请求等数据进行分析等处理,并将处理结果反馈给终端设备。

[0050] 服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN(ContentDeliveryNetwork,内容分发网络)、以及大数据和人工智能平台等基础云计算服务的云服务器等,本公开对此不做限制。

[0051] 在图1所示实施例中,终端设备101可以与目标对象102进行人机语音对话(实施场景可以参考人与机器对话过程)。在对话过程中,终端设备101或者为终端设备101提供后台服务的服务器可以实现以下过程:实时检测对话语音,以获得对话语音流;根据对话语音流确定话语权切换概率;根据话语权切换概率确定反馈策略,反馈策略用于对对话语音进行响应。

[0052] 在上述系统架构下,本公开实施例中提供了一种人机对话方法,该方法可以由任意具备计算处理能力的电子设备执行。

[0053] 图2是根据一示例性实施例示出的一种人机对话方法的流程图。本公开实施例所提供的方法可以由任意具备计算处理能力的电子设备来执行,例如该方法可以由上述图1实施例中的服务器或终端设备来执行,也可以由服务器和终端设备共同执行,在下面的实施例中,以终端设备为执行主体为例进行举例说明,但本公开并不限于此。

[0054] 参照图2,本公开实施例提供的人机对话方法可以包括以下步骤。

[0055] 步骤S202,实时检测对话语音,以获得对话语音流。

[0056] 在一些实施例中,电子设备可以对人机语音对话过程中目标对象的语音进行实时检测。

[0057] 步骤S204,根据对话语音流确定话语权切换概率。

[0058] 在一些实施例中,可以根据目标对象的语音确定话语权切换概率。

[0059] 其中,该话语权切换概率可以指的是目标对象讲话完成的概率。

[0060] 在一些实施例中,当话语权切换概率较高时,代表目标对象讲话完成,电子设备可以从目标对象处接过话语权,从而与目标对象进行对话。

[0061] 在一些实施例中,当话语权切换率较低时,代表目标对象继续讲话的欲望较高,电子设备不要打断目标对象讲话。

[0062] 可以理解的是,在目标对象继续说话的欲望较低(即话语权切换概率较高),很可能要结束说话的时候,电子设备接过话语权继续说话。

[0063] 接过话语权可以指的是电子设备在目标对象之后开始说话。例如可以是电子设备打断目标对象后开始说话(目标对象因为被打断了,所以不再说话);还可以是目标对象主动停止说话后电子设备接着进行说话。

[0064] 在一些实施例中,当话语权切换概率较低时,表明目标对象还要继续表达,电子设备不要打断目标对象的对话。

[0065] 步骤S206,根据话语权切换概率确定反馈策略,反馈策略用于对对话语音进行响应。

[0066] 在一些实施例中,上述反馈策略可以包括在电子设备讲实质性内容之前给出一些反应词,该反应词可以包括响应词、填充词等。

[0067] 上述反应词可以用于对目标对象的语音内容进行反应,比如进行激励、总结等,本申请对此不做限制。

[0068] 在一些实施例中,电子设备可以根据上述话语权切换概率发出响应词,适时的激励或者响应于目标对象话语,以模仿人与人之间的对话,从而在提高对话流畅性的同时提高用户体验。

[0069] 上述实施例,通过对目标对象的对话语音进行实时检测,以预测确定目标对象的话语权切换概率,从而使得电子根据该话语权切换概率发出反应词,以便在口语对话过程中向目标对象给出反应,从而使得口语对话更加流程,提高人机对话的流畅性和客户体验。

[0070] 图3是根据一示例性实施例示出的一种人机对话方法的流程图。

[0071] 参考图3,上述人机对话方法可以包括以下步骤。

[0072] 步骤S302,实时检测对话语音,以获得对话语音流。

[0073] 步骤S304,确定对话语音流的间断时间超过第一阈值。

[0074] 上述第一阈值可以小于或者等于1秒,可例如是0.1秒、0.2秒、0.01秒或者0.02秒等,本领域技术人员可以根据实际需要设置间断时间的第一阈值。

[0075] 步骤S306,从对话语音间断处向前截取目标时间段的对话语音。

[0076] 可以理解的是,当目标对象的对话语音超过第一阈值时,那么目标对象可能是要结束对话,也有可能是暂时的停下来想一想。

[0077] 例如,当目标对象说“这个问题我得想一想~~”后可能会存在一段时间语音停止,但是这段语音停止并不代表目标对象已经结束了对话,而是目标对象正在思考,因此电子设备不能打断目标对象。

[0078] 因此,本申请截取对话语音间断处之前的一段语音进行语音判断,以确定话语权切换概率。

[0079] 上述目标时间段的对话语音可以指的是2秒时长的对话语音,本领域技术人员可以根据实际需求设置目标时间段的时长,本申请对此不做限制。

[0080] 在一些实施例中,当判断目标对象的对话语音流间断时间超过第一阈值时,可以从语音间断处向前截取目标时间段的对话语音,以便判断目标对象是否还要继续讲话。

[0081] 步骤S308,对目标时间段的对话语音进行特征提取,以获得目标时间段的对话语音的音频特征。

[0082] 在一些实施例中,上述音频特征可以包括音调特征(如pitch特征)、强度特征(如intensity特征)、功率特征(如power特征)和频谱特征(如mfcc (Mel-scaleFrequency Cepstral Coefficients,梅尔倒谱系数))。

[0083] 步骤S310,根据目标时间段的对话语音的音频特征确定话语权切换概率。

[0084] 步骤S312,根据话语权切换概率确定反馈策略,反馈策略用于对对话语音进行响应。

[0085] 上述实施例提供的技术方案,可以在目标对象的语音对话发生间断时,根据间断前的语音片段确定话语权切换概率,从而根据话语权切换概率判断电子设备是否需要给出反应词,以及时地给目标对象反应。

[0086] 图4是根据一示例性实施例示出的一种人机对话方法的流程图。

[0087] 参考图4,上述人机对话方法可以包括以下步骤。

[0088] 步骤S402,实时检测对话语音,以获得对话语音流。

[0089] 步骤S404,确定对话语音流的间断时间超过第一阈值。

[0090] 步骤S406,从对话语音间断处向前截取目标时间段的对话语音。

[0091] 步骤S408,对目标时间段的对话语音进行特征提取,以获得目标时间段的对话语音的音频特征。

[0092] 步骤S410,通过目标神经网络的特征提取结构对目标时间段的对话语音的音频特征进行特征提取,以确定目标时间段的对话语音的语义特征。

[0093] 上述特征提取结构可以是任意一种可以进行特征提取的神经网络结构,如图5所示,上述特征提取结构501可以是由长短记忆循环神经网络LSTM和全连接神经网络构成的网络结构,本申请对此不做限制。

[0094] 步骤S412,通过目标神经网络的第一分支结构对目标时间段的对话语音的语义特

征进行处理,确定话语权切换概率。

[0095] 如图5所示,上述第一分支结构502可以包括卷积层(未示出)、全连接层或者归一化层(如sigmoid)等。

[0096] 在一些实施例中,可以通过第一分支结构对目标时间段的对话语音的语义特征进行处理,以预测话语权切换概率。

[0097] 步骤S414,根据话语权切换概率确定反馈策略,反馈策略用于对对话语音进行响应。

[0098] 上述实施例提供的技术方案可以通过目标神经网络的第一分支结构对目标时间段的对话语音的语义特征进行处理,以准确的确定话语权切换概率。

[0099] 图6是根据一示例性实施例示出的一种人机对话方法的流程图。

[0100] 参考图6,上述人机对话方法可以包括以下步骤。

[0101] 步骤S602,通过目标神经网络的特征提取结构对目标时间段的对话语音的音频特征进行特征提取,以确定目标时间段的对话语音的语义特征。

[0102] 步骤S604,通过目标神经网络的第一分支结构对目标时间段的对话语音的语义特征进行处理,确定话语权切换概率。

[0103] 在一些实施例中,反应词可以包括多个响应词。

[0104] 步骤S606,判断话语权切换概率是否大于第二阈值。

[0105] 图7是根据一示例性实施例示出的一种通过话语权切换概率将话语权切换时机、响应词发出时机和填充词发出时间串联起来的示意图。

[0106] 图7示出了所示的话语权决策、填充词、响应词三者的关系,首先这个图展示了这三个模块在口语对话系统中的先后顺序以及相互之间的依赖关系。首先他们的基础是进行话语权轮换的概率,我们把概率分为高中低三个区间。当概率值处于低区间时,不进行话语权轮换,继续让用户讲话。当概率值处于中间区间时,系统产生响应词(例如“好的”,“对的”等),但是并不接过话语权,继续让用户讲话。当概率值处于高区间时,由于整个系统的高延迟问题,这时系统首先产生填充词(例如“收到”,“明白了”等),给用户一个快速响应,然后再接过话语权开始讲话。

[0107] 如图7所示,在确定话语权切换概率不大于第二阈值的情况下,不做任何处理。

[0108] 在确定话语权切换概率大于第二阈值的情况下,继续通过步骤S608判断话语权切换概率是否大于第三阈值。

[0109] 如图7所示,在确定话语权切换概率大于第二阈值但是不大于第三阈值的情况下,可以考虑产生一个响应词,具体过程可以参考步骤S610~步骤S614。

[0110] 如图7所示,在确定话语权切换概率大于第三阈值的情况下,可以考虑产生一个填充词,具体过程可以参考步骤S616~步骤S620。

[0111] 步骤S610,通过目标神经网络的第二分支结构对目标时间段的对话语音的语义特征进行处理,以确定是否给出响应词。

[0112] 如图5所示,上述第二分支结构503可以包括卷积层(未示出)、全连接层或者归一化层(如sigmoid)等。

[0113] 在一些实施例中,可以通过第二分支结构对目标时间段的对话语音的语义特征进行处理,以确定是否给出响应词。

[0114] 步骤S612,在根据目标时间段的对话语音的语义特征确定给出响应词时,在多个响应中随机确定一目标响应词。

[0115] 步骤S614,基于目标响应词,对对话语音进行响应。

[0116] 在一些实施例中,反应词可以包括多个填充词。

[0117] 在一些实施例中,可以发出目标响应词,以便在不打断目标对象的语音对应的同时,向目标对象给出响应。

[0118] 步骤S616,通过目标神经网络的第三分支结构对目标时间段的对话语音的语义特征进行处理,确定预设的多个填充词对应的发出概率。

[0119] 如图5所示,上述第三分支结构504可以包括卷积层(未示出)、全连接层或者归一化层(如sigmoid)等。

[0120] 在一些实施例中,可以通过第三分支结构对目标时间段的对话语音的语义特征进行处理,确定预设的多个填充词对应的发出概率。

[0121] 步骤S618,将发出概率最大的填充词作为目标填充词。

[0122] 步骤S620,基于目标填充词,对对话语音进行响应。

[0123] 在一些实施例中,可以在接过目标对象的话语权开始讲话前,发出目标响应词。

[0124] 上述实施例提供的技术方案,一方面将响应词和填充词的发出实际通过话语权切换概率联系起来,使得电子设备与目标对象的对话更加流畅;另一方面,通过神经网络的多个分支对话话语权切换概率、响应词发出概率以及填充词发出概率进行判断,可以提高反应词发出时机的预测准确性;另外,将第一分支结构、第二分支结构以及第三分支结构放在同一个神经网络中进行训练,可以给更好的挖掘话语权切换概率、响应词发出概率以及填充词发出概率之间的关系,以便更好的完成预测。

[0125] 图8是根据一示例性实施例示出的一种神经网络方法的流程图。

[0126] 参考图8,上述神经网络方法可以包括以下步骤。

[0127] 步骤S802,根据话语权切换概率确定话语权切换损失。

[0128] 在一些实施例中,可以根据话语权切换概率和电子设备和目标对象的实际话语权切换情况确定话语权切换损失。

[0129] 步骤S804,根据是否给出响应词确定响应词损失。

[0130] 在一些实施例中,可以根据预测处理的是否给出响应词的预测结果,与实际情况下是否需要给出响应词的实际结果,计算响应词损失。

[0131] 步骤S806,根据各个填充词的发出概率确定填充词损失。

[0132] 根据各个填充词的发出概率以及实际操作中是否会发出该填充词的实际情况,计算填充词损失。

[0133] 步骤S808,根据话语权切换损失、响应词损失和填充词损失确定目标损失。

[0134] 步骤S810,通过目标损失对目标神经网络进行训练。

[0135] 通过上述实施例可以提高目标神经网络的预测准确性,从而提高语音对话的流畅度。

[0136] 通过图7,可以看到话语权决策、填充词、响应词三者之间存在强依赖的关系,于是本申请实施例对这三者进行统一建模,以获得图5所示的模型结构。本申请提出的图5所示的模型是一个多任务学习的模型,它们共享一个编码器。首先,该模型的输入是语音特征,

由于考虑到末尾音频包含重要的信息,于是本申请截取了用户音频间断前的最后2s的音频,进行切片分帧,每50ms为一帧,然后提取pitch(音调)、intensity(强度)、power(功率)、mfcc(梅尔倒谱系数)等音频特征作为模型的输入。如图5所示,本申请实施例可以采用长短记忆循环神经网络LSTM作为编码器模型,之后拼接了多层全连接层,然后接话语权决策、填充词、响应词三个子任务,多任务损失函数公式如公式(1)所示。其中话语权决策是一个回归任务,输出是进行话语权切换的概率值。填充词是一个多分类任务,决定输出什么类型的填充词。响应词是一个二分类任务,决定是否产生响应词,由于响应词不具有区分性,具体的响应词恶意从列表中随机产生。

[0137] 在本实施例中,可以先通过话语权决策任务确定话语权切换概率;当确定话语权切换概率大于第二阈值且小于或者等于第三阈值时,再通过响应词任务确定是否需要给出响应词,在根据目标时间段的对话语音的语义特征确定给出响应词时,在多个响应中随机确定一目标响应词,以便发出目标响应词;当确定话语权切换概率大于第三阈值时,通过填充词任务对目标时间段的对话语音的语义特征进行处理,确定预设的多个填充词对应的发出概率,将发出概率最大的填充词作为目标填充词,并基于目标填充词,对对话语音进行响应。

$$[0138] \quad l_{\text{loss}} = l_{\text{ep}} + l_{\text{filler}} + l_{\text{bc}} \quad (1)$$

[0139] 基于同一发明构思,本公开实施例中还提供了一种人机对话装置,如下面的实施例。由于该装置实施例解决问题的原理与上述方法实施例相似,因此该装置实施例的实施可以参见上述方法实施例的实施,重复之处不再赘述。

[0140] 图9是根据一示例性实施例示出的一种人机对话装置的框图。参照图9,本公开实施例提供的人机对话装置900可以包括:语音检测模块901、话语权切换概率确定模块902和响应词发出模块903。

[0141] 其中,语音检测模块901可以用于实时检测对话语音,以获得对话语音流;话语权切换概率确定模块902可以用于根据对话语音流确定话语权切换概率;响应词发出模块903可以用于根据话语权切换概率确定反馈策略,反馈策略用于对对话语音进行响应。

[0142] 此处需要说明的是,上述语音检测模块901、话语权切换概率确定模块902、响应词发出模块903对应于方法实施例中的S202~S204,上述模块与对应的步骤所实现的示例和应用场景相同,但不限于上述方法实施例所公开的内容。需要说明的是,上述模块作为装置的一部分可以在诸如一组计算机可执行指令的计算机系统中执行。

[0143] 在一些实施例中,话语权切换概率确定模块902可以包括:第一阈值判断子模块、语音截取子模块、特征提取子模块和话语权切换概率预测子模块。

[0144] 其中,第一阈值判断子模块可以用于确定对话语音流的间断时间超过第一阈值;语音截取子模块可以用于从对话语音间断处向前截取目标时间段的对话语音;特征提取子模块可以用于对目标时间段的对话语音进行特征提取,以获得目标时间段的对话语音的音频特征;话语权切换概率预测子模块可以用于根据目标时间段的对话语音的音频特征确定话语权切换概率。

[0145] 在一些实施例中,话语权切换概率预测子模块可以包括:语义特征提取单元和第一分支处理单元。

[0146] 其中,语义特征提取单元可以用于通过目标神经网络的特征提取结构对目标时间

段的对话语音的音频特征进行特征提取,以确定目标时间段的对话语音的语义特征;第一分支处理单元可以用于通过目标神经网络的第一分支结构对目标时间段的对话语音的语义特征进行处理,确定话语权切换概率。

[0147] 在一些实施例中,反应词可以包括多个响应词;其中,响应词发出模块903可以包括:第三阈值判断子模块、响应词判断子模块和目标响应词确定子模块和目标响应词发出子模块。

[0148] 其中,第三阈值判断子模块可以用于通过第一分支结构确定话语权切换概率大于第二阈值且小于或者等于第三阈值;响应词判断子模块可以用于通过目标神经网络的第二分支结构对目标时间段的对话语音的语义特征进行处理,以确定是否给出响应词;目标响应词确定子模块可以用于在根据所述目标时间段的对话语音的语义特征确定给出响应词时,在预设的多个所述响应词中随机确定一目标响应词;目标响应词发出子模块可以用于基于目标响应词,对对话语音进行响应。

[0149] 在一些实施例中,反应词包括多个填充词;其中,响应词发出模块903可以包括:大于第三阈值判断子模块、填充词发出概率确定子模块、目标填充词确定子模块和目标填充词发出子模块。

[0150] 其中,大于第三阈值判断子模块可以用于通过第一分支结构确定话语权切换概率大于第三阈值;填充词发出概率确定子模块可以用于通过目标神经网络的第三分支结构对目标时间段的对话语音的语义特征进行处理,确定预设的多个填充词对应的发出概率;目标填充词确定子模块可以用于将发出概率最大的填充词作为目标填充词;目标填充词发出子模块可以用于基于目标填充词,对对话语音进行响应。

[0151] 在一些实施例中,人机对话装置还可以包括:话语权切换损失确定模块、响应词损失确定模块、填充词损失确定模块、目标损失确定模块和训练模块。

[0152] 其中,话语权切换损失确定模块可以用于根据话语权切换概率确定话语权切换损失;响应词损失确定模块可以用于根据是否给出响应词确定响应词损失;填充词损失确定模块可以用于根据各个填充词的发出概率确定填充词损失;目标损失确定模块可以用于根据话语权切换损失、响应词损失和填充词损失确定目标损失;训练模块可以用于通过目标损失对目标神经网络进行训练。

[0153] 在一些实施例中,特征提取子模块可以包括:特征提取单元和特征确定单元。

[0154] 其中,特征提取单元可以用于对目标时间段的对话语音进行特征提取;特征确定单元可以用于确定目标时间段的对话语音的音调特征、强度特征、功率特征和频谱特征。

[0155] 由于装置900的各功能已在其对应的方法实施例中予以详细说明,本公开于此不再赘述。

[0156] 描述于本申请实施例中所涉及到的模块和/或子模块和/或单元可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块和/或子模块和/或单元也可以设置在处理器中。其中,这些模块和/或子模块和/或单元的名称在某种情况下并不构成对该模块和/或子模块和/或单元本身的限定。

[0157] 附图中的流程图和框图,图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个

用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0158] 此外,上述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,上述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

[0159] 图10示出了适于用来实现本公开实施例的电子设备的结构示意图。需要说明的是,图10示出的电子设备1000仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0160] 如图10所示,电子设备1000包括中央处理单元(CPU)1001,其可以根据存储在只读存储器(ROM)1002中的程序或者从储存部分1008加载到随机访问存储器(RAM)1003中的程序而执行各种适当的动作和处理。在RAM 1003中,还存储有电子设备1000操作所需的各种程序和数据。CPU 1001、ROM 1002以及RAM 1003通过总线1004彼此相连。输入/输出(I/O)接口1005也连接至总线1004。

[0161] 以下部件连接至I/O接口1005:包括键盘、鼠标等的输入部分1006;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分1007;包括硬盘等的储存部分1008;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分1009。通信部分1009经由诸如因特网的网络执行通信处理。驱动器1010也根据需要连接至I/O接口1005。可拆卸介质1011,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器1010上,以便于从其上读出的计算机程序根据需要被安装入储存部分1008。

[0162] 特别地,根据本公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本公开的实施例包括一种计算机程序产品,其包括承载在计算机可读存储介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分1009从网络上被下载和安装,和/或从可拆卸介质1011被安装。在该计算机程序被中央处理单元(CPU)1001执行时,执行本申请的系统中限定的上述功能。

[0163] 需要说明的是,本公开所示的计算机可读存储介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不

限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读存储介质,该计算机可读存储介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读存储介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0164] 作为另一方面,本申请还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读存储介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备可实现功能包括:实时检测对话语音,以获得对话语音流;根据对话语音流确定话语权切换概率;根据话语权切换概率确定反馈策略,反馈策略用于对对话语音进行响应。

[0165] 根据本申请的一个方面,提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述实施例的各种可选实现方式中提供的方法。

[0166] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,本公开实施例的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中,包括若干指令用以使得一台计算设备(可以是个人计算机、服务器、移动终端、或者智能设备等)执行根据本公开实施例的方法。

[0167] 本领域技术人员在考虑说明书及实践在这里公开的公开后,将容易想到本公开的其他实施例。本公开旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未申请的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由权利要求指出。

[0168] 应当理解的是,本公开并不限于这里已经示出的详细结构、附图方式或实现方法,相反,本公开意图涵盖包含在所附权利要求的精神和范围内的各种修改和等效设置。

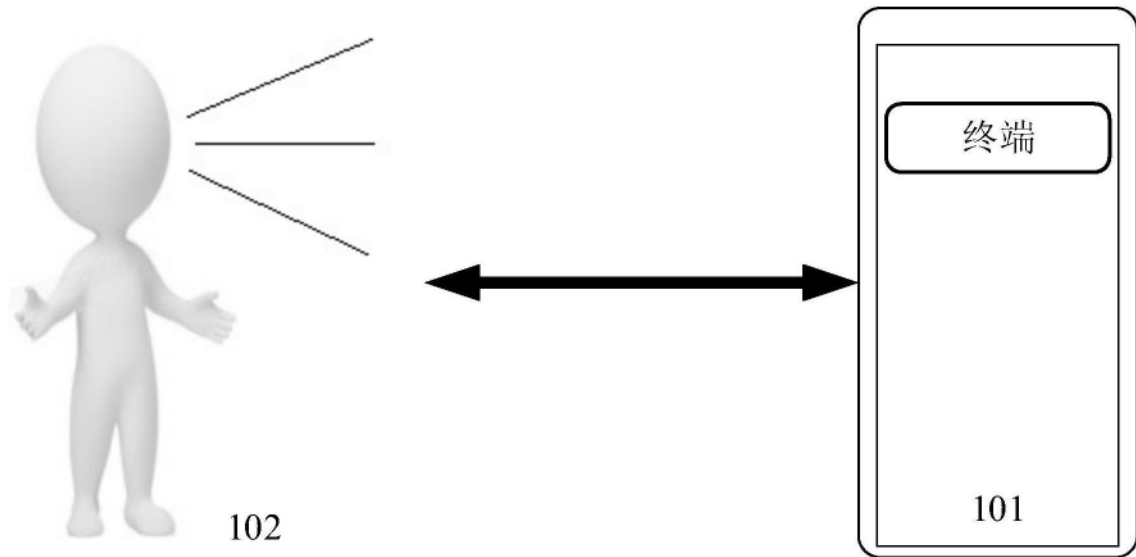


图1

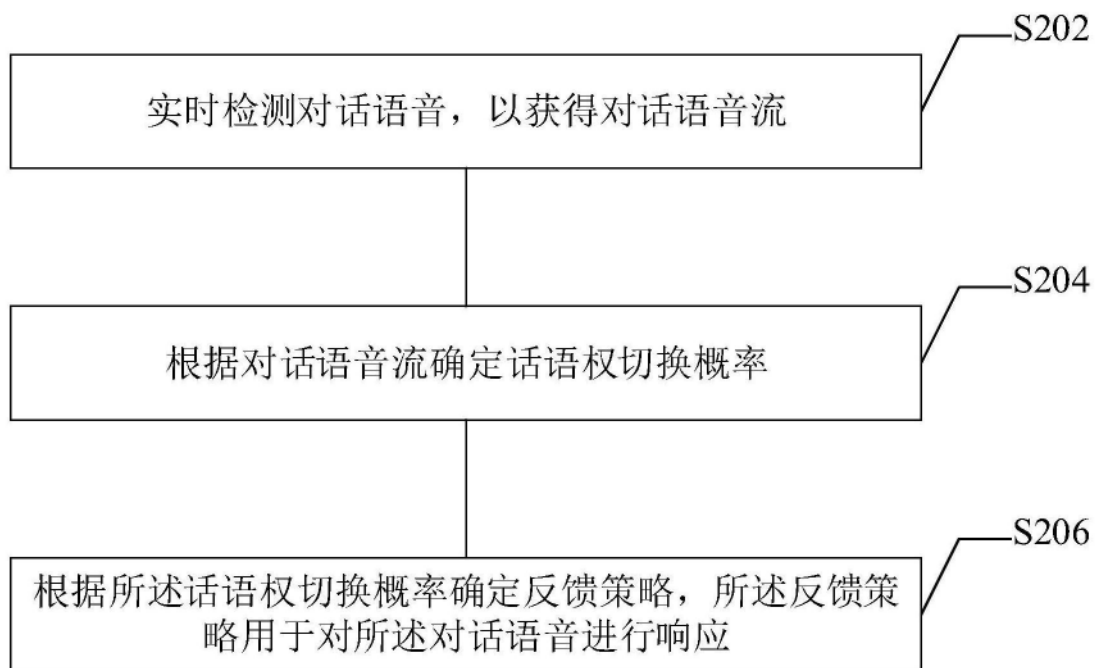


图2

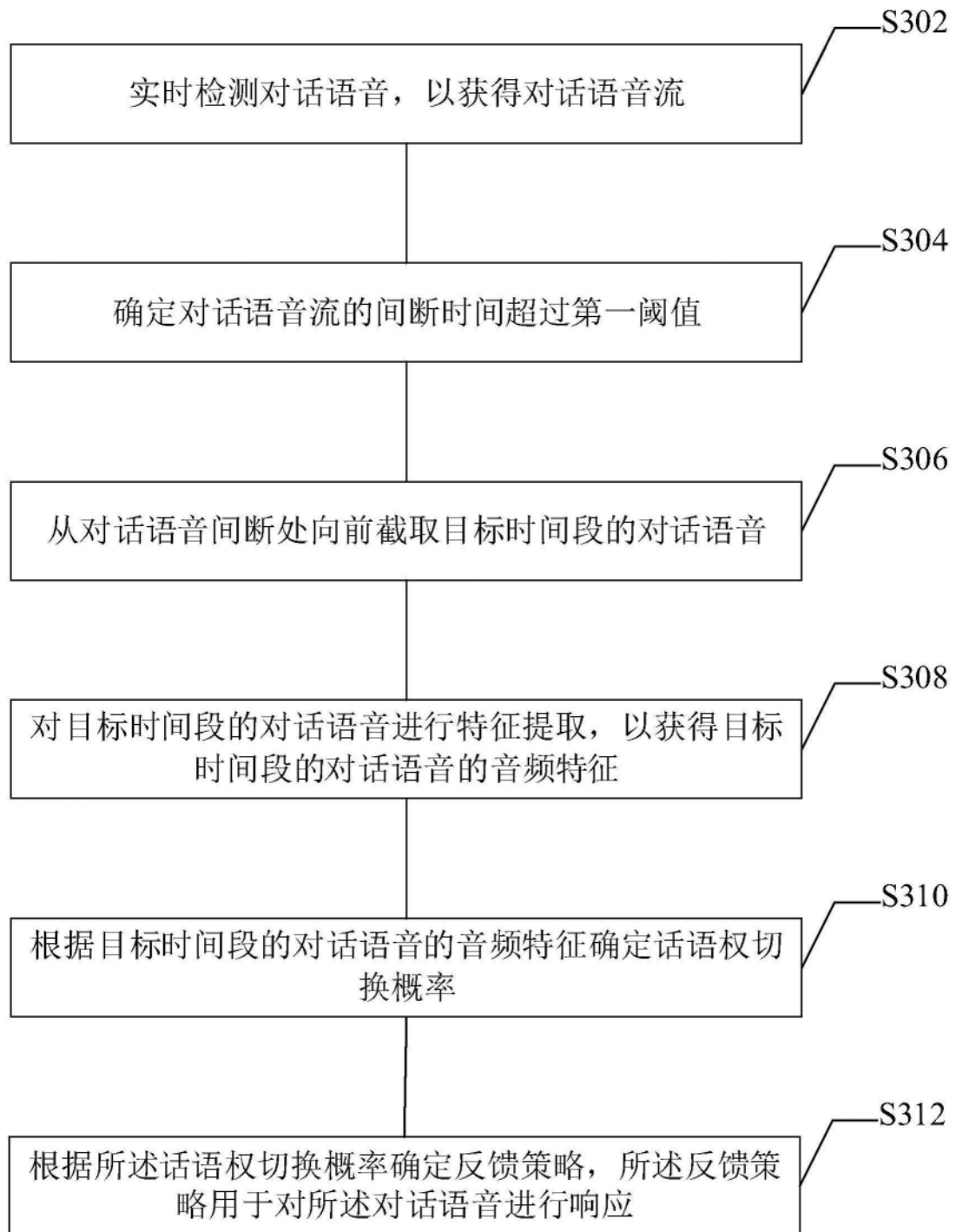


图3

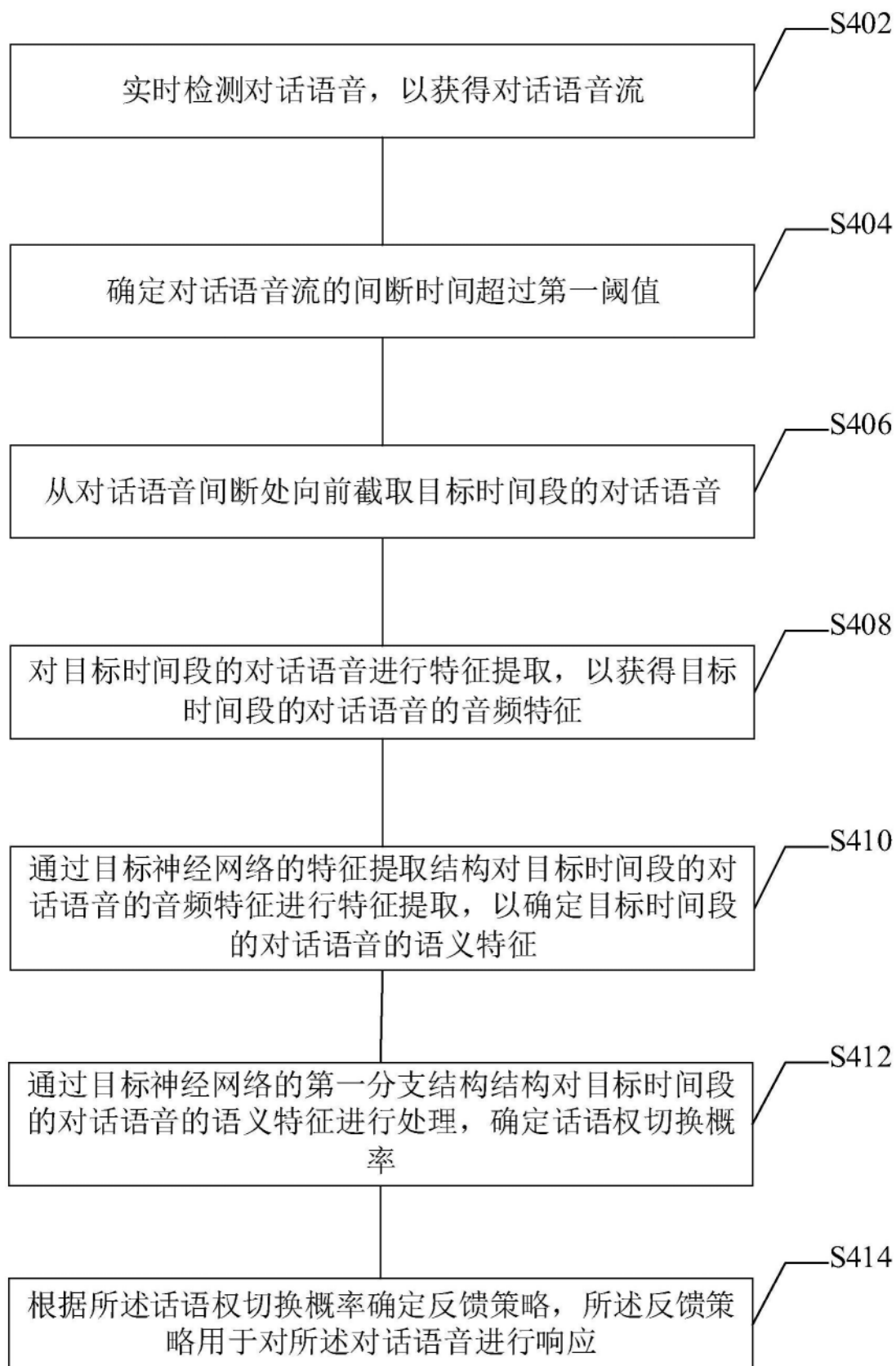


图4

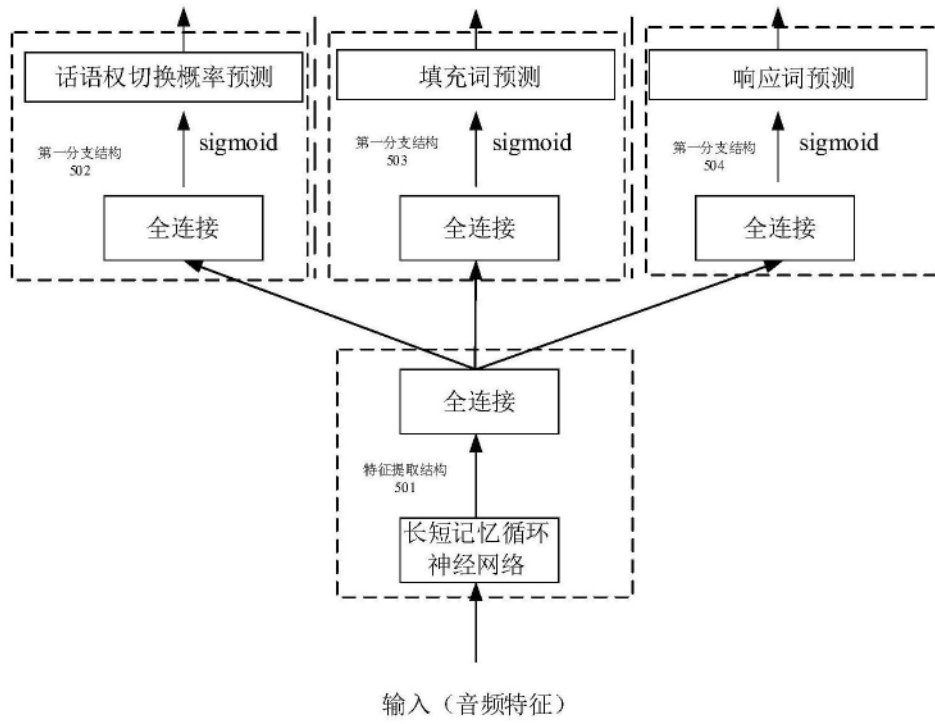


图5

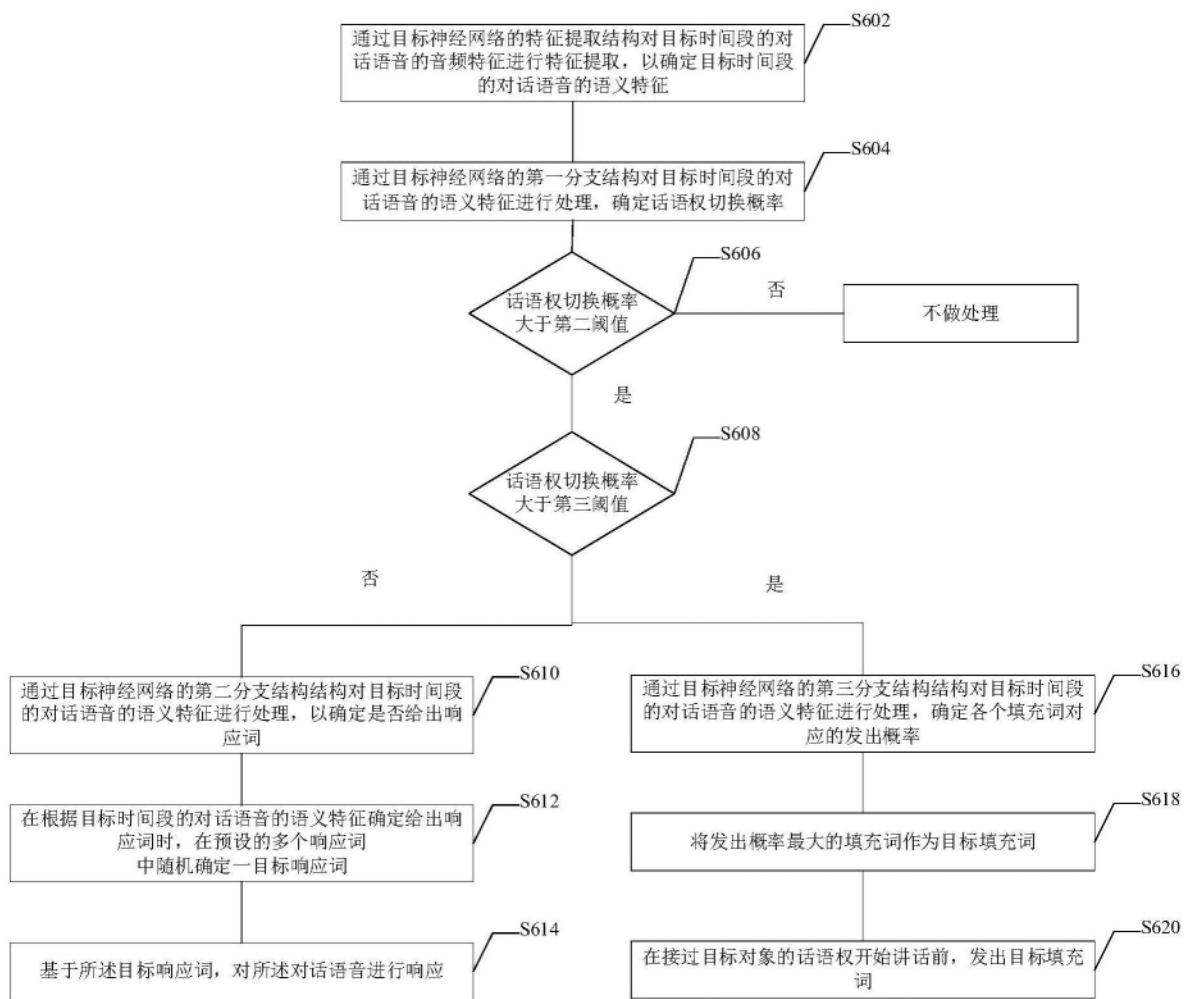


图6

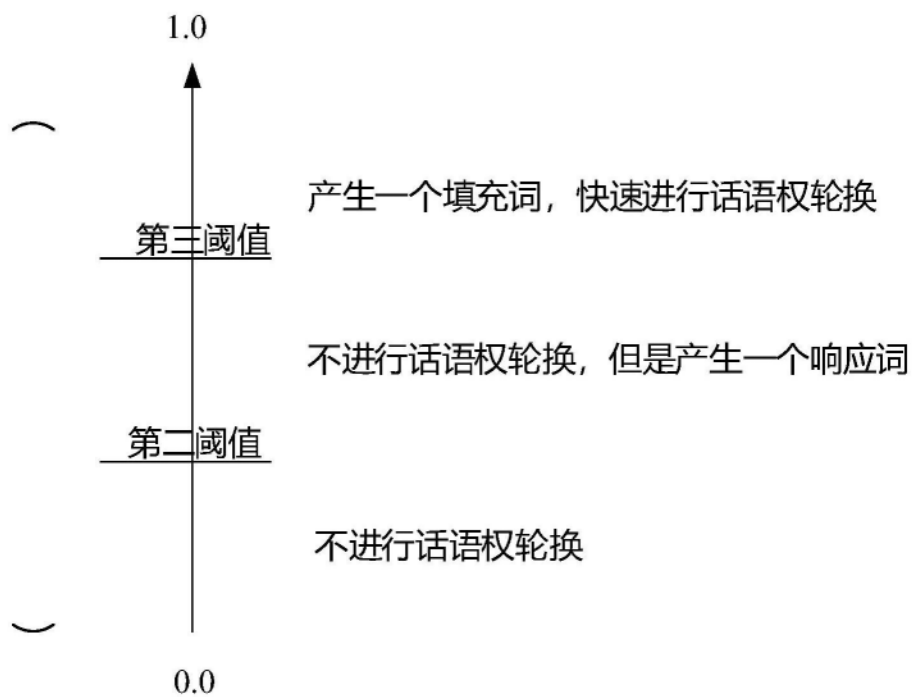


图7

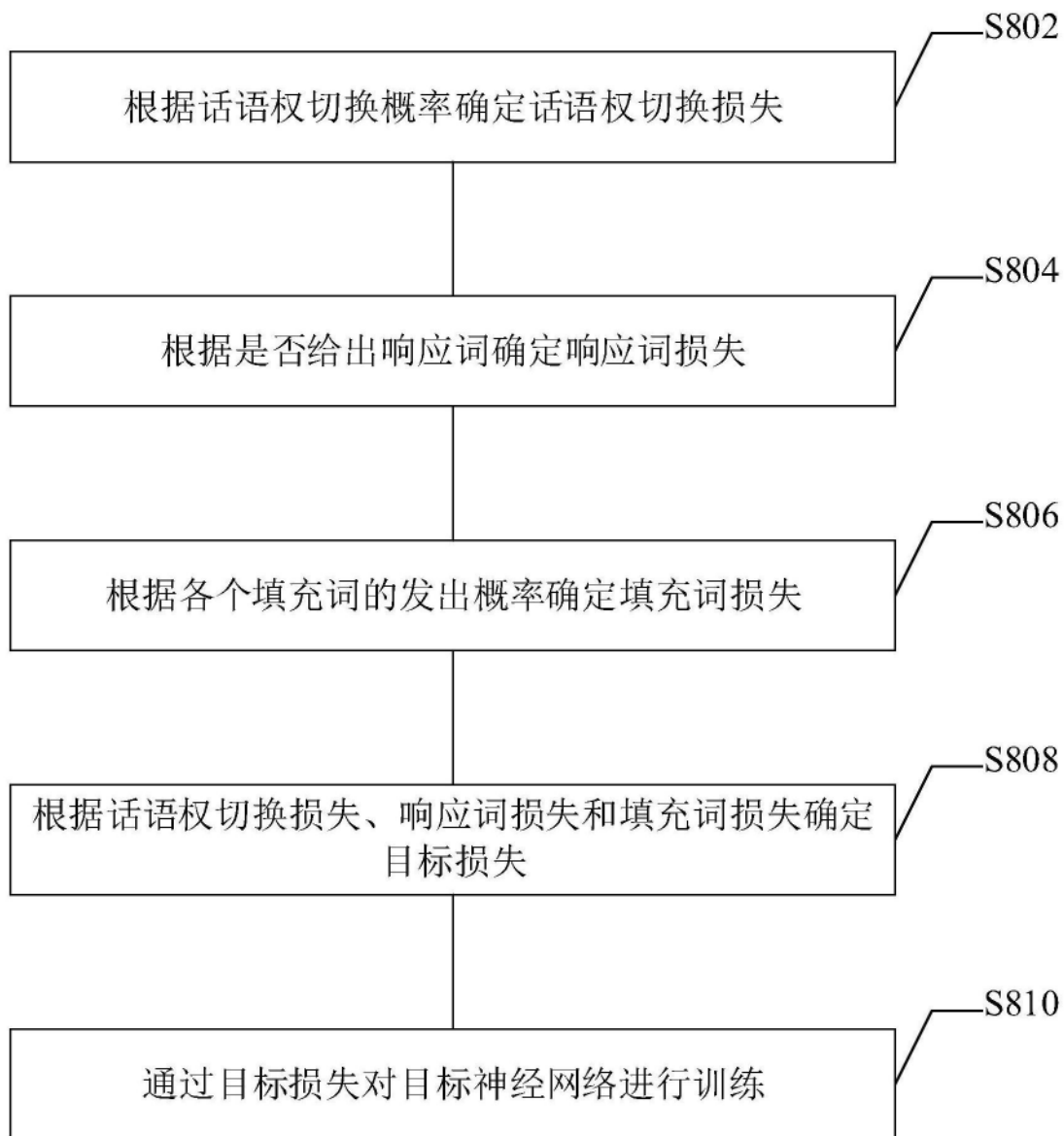


图8

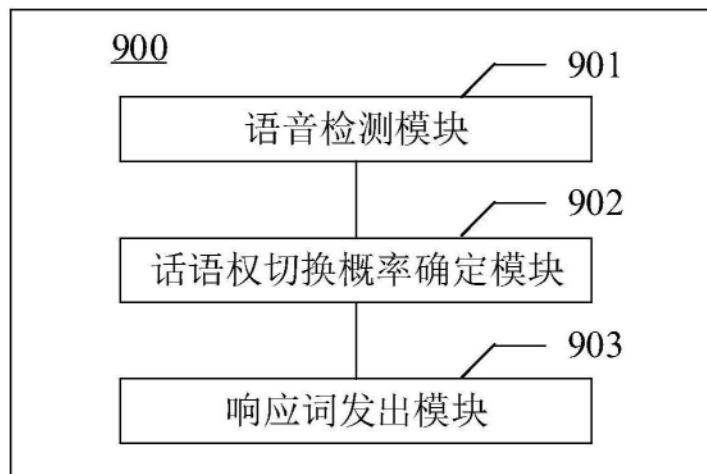


图9

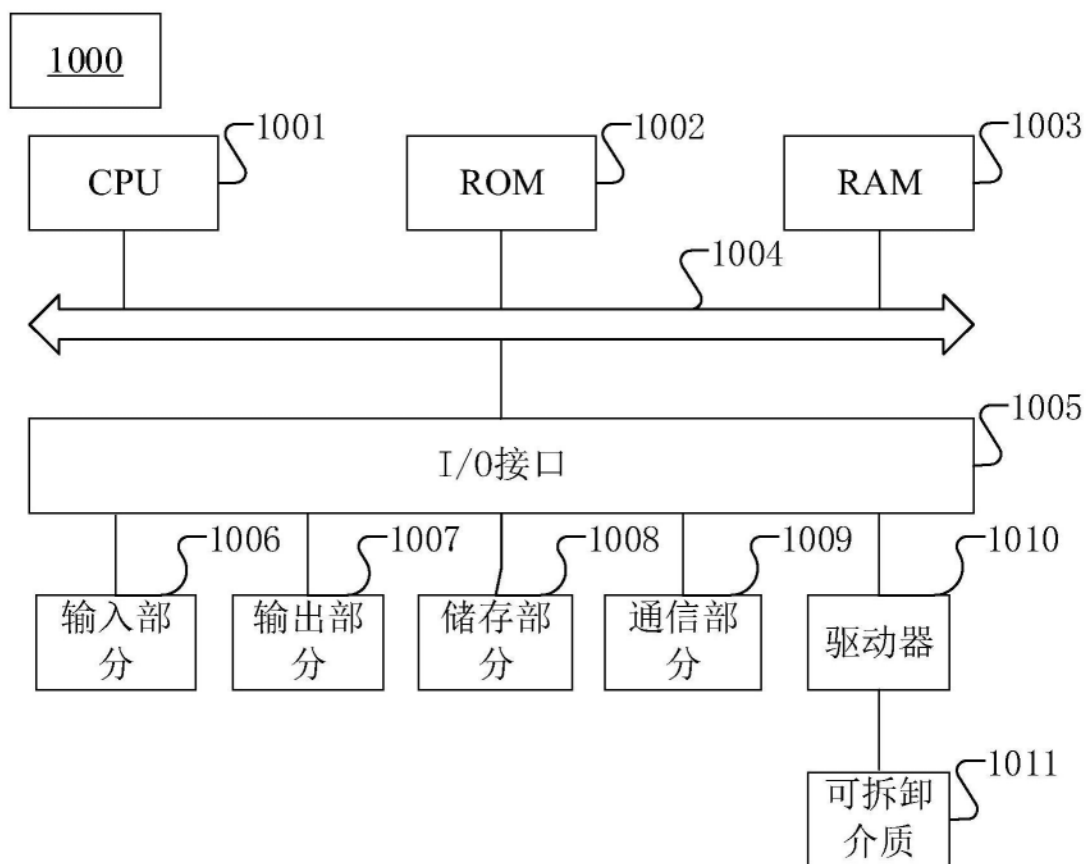


图10