



(12) 发明专利

(10) 授权公告号 CN 110138654 B

(45) 授权公告日 2022.02.11

(21) 申请号 201910490608.X

G10L 15/26 (2006.01)

(22) 申请日 2019.06.06

G10L 13/04 (2013.01)

(65) 同一申请的已公布的文献号

G10L 13/033 (2013.01)

申请公布号 CN 110138654 A

G10L 21/043 (2013.01)

(43) 申请公布日 2019.08.16

(56) 对比文件

(73) 专利权人 北京百度网讯科技有限公司

CN 109801648 A, 2019.05.24

地址 100085 北京市海淀区上地十街10号

CN 109801648 A, 2019.05.24

百度大厦2层

CN 103811012 A, 2014.05.21

(72) 发明人 唐婉骐 康佳美 曾丽霞 周轶景

CN 103543979 A, 2014.01.29

谢寒梅 朱丽娜

CN 104157301 A, 2014.11.19

CN 106357929 A, 2017.01.25

(74) 专利代理机构 北京英赛嘉华知识产权代理

CN 106373558 A, 2017.02.01

有限责任公司 11204

CN 109524027 A, 2019.03.26

代理人 马晓亚 田丽

lianglingxu314. “自动清除录音文件中的空白段”. 《百度文库, <http://wenku.baidu.com/view/dc93f6160b4e767f5acfcda.html>》. 2010, 第1-4页.

(51) Int. Cl.

H04L 51/18 (2022.01)

H04L 51/04 (2022.01)

H04L 51/10 (2022.01)

审查员 李浩

权利要求书2页 说明书10页 附图7页

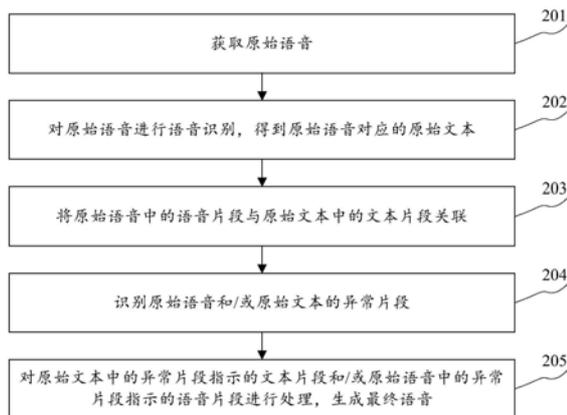
(54) 发明名称

用于处理语音的方法和装置

(57) 摘要

本申请实施例公开了用于处理语音的方法和装置。该方法的一具体实施方式包括：获取原始语音；对原始语音进行语音识别，得到原始语音对应的原始文本；将原始语音中的语音片段与原始文本中的文本片段关联；识别原始语音和/或原始文本的异常片段；对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理，生成最终语音。该实施方式将原始语音中的语音片段与原始文本中的文本片段关联，实现了语音的可视化处理。同时，避免了由于语音录制过程中出现异常状况而需要重新录制的问题，从而提高了语音录制效率。

200



CN 110138654 B

1. 一种用于处理语音的方法,包括:

获取原始语音;

对所述原始语音进行语音识别,得到所述原始语音对应的原始文本;

将所述原始语音中的语音片段与所述原始文本中的文本片段关联;

识别所述原始文本的异常片段;

对所述原始文本中的所述异常片段指示的文本片段和/或所述原始语音中的所述异常片段指示的语音片段进行处理,生成最终语音;

对所述最终语音进行平滑处理,包括:基于所述最终语音的音量和重音,确定所述最终语音对应的情绪级别;基于所述最终语音对应的情绪级别,调整所述最终语音的语速、音调 and 重音,其中,所述语速、所述音调的强度以及所述重音的数量均与所述情绪级别成正比。

2. 根据权利要求1所述的方法,其中,所述对所述原始语音进行语音识别,得到所述原始语音对应的原始文本,包括:

当所述原始语音中存在空白语音片段和/或拖长音语音片段时,将所述空白语音片段识别成第一预设符号,和/或将所述拖长音语音片段识别成第二预设符号。

3. 根据权利要求2所述的方法,其中,所述空白语音片段是所述原始语音中声音强度小于预设强度,不包含人声特征声波,并且持续时长大于第一预设时长的语音片段,所述拖长音语音片段是所述原始语音中声音强度不小于所述预设强度,不包含语义内容,并且持续时长大于第二预设时长的语音片段。

4. 根据权利要求3所述的方法,其中,所述将所述空白语音片段识别成第一预设符号,和/或将所述拖长音语音片段识别成第二预设符号,包括:

基于所述空白语音片段的时长与所述第一预设时长的比值,确定将所述空白语音片段识别成的所述第一预设符号的数量;和/或基于所述拖长音语音片段的时长与所述第二预设时长的比值,确定将所述拖长音语音片段识别成的所述第二预设符号的数量。

5. 根据权利要求4所述的方法,其中,所述对所述原始文本中的所述异常片段指示的文本片段和/或所述原始语音中的所述异常片段指示的语音片段进行处理,生成最终语音,包括:

删除所述原始文本中的所述异常片段指示的文本片段的至少部分,以及同步删除所述原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分,生成最终语音。

6. 根据权利要求5所述的方法,其中,所述删除所述原始文本中的所述异常片段指示的文本片段的至少部分,包括:

当所述原始文本中存在所述第一预设符号和/或所述第二预设符号时,将所述第一预设符号和/或所述第二预设符号从所述原始文本中删除;

当所述原始文本中存在连续重复出现次数超过预设次数的词语时,将除预设位置之外的所述词语从所述原始文本中删除。

7. 根据权利要求1所述的方法,其中,所述对所述最终语音进行平滑处理,包括:

基于所述最终语音的语音特征,确定所述最终语音对应的方言类别;

基于所述最终语音对应的方言类别,矫正所述最终语音中的音节,以及调整所述最终语音的重音。

8. 根据权利要求5所述的方法,其中,所述方法还包括:

响应于检测到对所述原始文本中的文本片段的至少部分的修订操作,同步修订所述原始语音中的、与所修订的文本片段的至少部分关联的语音片段的至少部分。

9. 根据权利要求1-8之一所述的方法,其中,所述原始语音是即时通讯应用中的第一用户发送的;以及所述方法还包括:

将所述最终语音发送至所述即时通讯应用的服务器,以使所述即时通讯应用的服务器将所述最终语音发送给所述即时通讯应用的第二用户。

10. 一种用于处理语音的装置,包括:

获取单元,被配置成获取原始语音;

第一识别单元,被配置成对所述原始语音进行语音识别,得到所述原始语音对应的原始文本;

关联单元,被配置成将所述原始语音中的语音片段与所述原始文本中的文本片段关联;

第二识别单元,被配置成识别所述原始文本的异常片段;

处理单元,被配置成对所述原始文本中的所述异常片段指示的文本片段和/或所述原始语音中的所述异常片段指示的语音片段进行处理,生成最终语音;

平滑单元,被配置成对所述最终语音进行平滑处理;

其中,所述平滑单元进一步被配置成:基于所述最终语音的音量和重音,确定所述最终语音对应的情绪级别;基于所述最终语音对应的情绪级别,调整所述最终语音的语速、音调和重音,其中,所述语速、所述音调的强度以及所述重音的数量均与所述情绪级别成正比。

11. 一种终端设备,包括:

一个或多个处理器;

存储装置,其上存储有一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-9中任一所述的方法。

12. 一种计算机可读介质,其上存储有计算机程序,其中,所述计算机程序被处理器执行时实现如权利要求1-9中任一所述的方法。

用于处理语音的方法和装置

技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及用于处理语音的方法和装置。

背景技术

[0002] 即时通讯(Instant Message,IM)是目前互联网上最为流行的通讯方式,能够实现即时发送和接收信息。目前大部分的移动终端都安装有即时通讯应用。即时通讯应用支持文字、图片、语音、视频等格式的信息的交互。对于语音交互,通常是发送用户录制语音,并直接发送给接收用户。

[0003] 然而,在语音录制过程中,不可避免地会出现异常状况,例如,发送用户存在语言表达缺陷(如说话磕巴),发送用户在语音录制过程中停下来思考说话的内容等。当出现异常状况时,通常需要发送用户重新录制语音,再发送给接收用户。

发明内容

[0004] 本申请实施例提出了用于处理语音的方法和装置。

[0005] 第一方面,本申请实施例提供了一种用于处理语音的方法,包括:获取原始语音;对原始语音进行语音识别,得到原始语音对应的原始文本;将原始语音中的语音片段与原始文本中的文本片段关联;识别原始语音和/或原始文本的异常片段;对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音。

[0006] 在一些实施例中,对原始语音进行语音识别,得到原始语音对应的原始文本,包括:当原始语音中存在空白语音片段和/或拖长音语音片段时,将空白语音片段识别成第一预设符号,和/或将拖长音语音片段识别成第二预设符号。

[0007] 在一些实施例中,空白语音片段是原始语音中声音强度小于预设强度,不包含人声特征声波,并且持续时长大于第一预设时长的语音片段,拖长音语音片段是原始语音中声音强度不小于预设强度,不包含语义内容,并且持续时长大于第二预设时长的语音片段。

[0008] 在一些实施例中,将空白语音片段识别成第一预设符号,和/或将拖长音语音片段识别成第二预设符号,包括:基于空白语音片段的时长与第一预设时长的比值,确定将空白语音片段识别成的第一预设符号的数量;和/或基于拖长音语音片段的时长与第二预设时长的比值,确定将拖长音语音片段识别成的第二预设符号的数量。

[0009] 在一些实施例中,对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音,包括:删除原始文本中的异常片段指示的文本片段的至少部分,以及同步删除原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分,生成最终语音。

[0010] 在一些实施例中,删除原始文本中的异常片段指示的文本片段的至少部分,包括:当原始文本中存在第一预设符号和/或第二预设符号时,将第一预设符号和/或第二预设符号从原始文本中删除;当原始文本中存在连续重复出现次数超过预设次数的词语时,将除

预设位置之外的词语从原始文本中删除。

[0011] 在一些实施例中,在生成最终语音之后,还包括:对最终语音进行平滑处理。

[0012] 在一些实施例中,对最终语音进行平滑处理,包括:基于最终语音的语音特征,确定最终语音对应的方言类别;基于最终语音对应的方言类别,矫正最终语音中的音节,以及调整最终语音的重音。

[0013] 在一些实施例中,对最终语音进行平滑处理,包括:基于最终语音的音量和重音,确定最终语音对应的情绪级别;基于最终语音对应的情绪级别,调整最终语音的语速、音调和重音。

[0014] 在一些实施例中,该方法还包括:响应于检测到对原始文本中的文本片段的至少部分的修订操作,同步修订原始语音中的、与所修订的文本片段的至少部分关联的语音片段的至少部分。

[0015] 在一些实施例中,原始语音是即时通讯应用中的第一用户发送的;以及该方法还包括:将最终语音发送至即时通讯应用的服务器,以使即时通讯应用的服务器将最终语音发送给即时通讯应用的第二用户。

[0016] 第二方面,本申请实施例提供了一种用于处理语音的装置,包括:获取单元,被配置成获取原始语音;第一识别单元,被配置成对原始语音进行语音识别,得到原始语音对应的原始文本;关联单元,被配置成将原始语音中的语音片段与原始文本中的文本片段关联;第二识别单元,被配置成识别原始语音和/或原始文本的异常片段;处理单元,被配置成对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音。

[0017] 在一些实施例中,第一识别单元进一步被配置成:当原始语音中存在空白语音片段和/或拖长音语音片段时,将空白语音片段识别成第一预设符号,和/或将拖长音语音片段识别成第二预设符号。

[0018] 在一些实施例中,空白语音片段是原始语音中声音强度小于预设强度,不包含人声特征声波,并且持续时长大于第一预设时长的语音片段,拖长音语音片段是原始语音中声音强度不小于预设强度,不包含语义内容,并且持续时长大于第二预设时长的语音片段。

[0019] 在一些实施例中,第一识别单元进一步被配置成:基于空白语音片段的时长与第一预设时长的比值,确定将空白语音片段识别成的第一预设符号的数量;和/或基于拖长音语音片段的时长与第二预设时长的比值,确定将拖长音语音片段识别成的第二预设符号的数量。

[0020] 在一些实施例中,处理单元进一步被配置成:删除原始文本中的异常片段指示的文本片段的至少部分,以及同步删除原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分,生成最终语音。

[0021] 在一些实施例中,处理单元进一步被配置成:当原始文本中存在第一预设符号和/或第二预设符号时,将第一预设符号和/或第二预设符号从原始文本中删除;当原始文本中存在连续重复出现次数超过预设次数的词语时,将除预设位置之外的词语从原始文本中删除。

[0022] 在一些实施例中,该装置还包括:平滑单元,被配置成对最终语音进行平滑处理。

[0023] 在一些实施例中,平滑单元进一步被配置成:基于最终语音的语音特征,确定最终

语音对应的方言类别;基于最终语音对应的方言类别,矫正最终语音中的音节,以及调整最终语音的重音。

[0024] 在一些实施例中,平滑单元进一步被配置成:基于最终语音的音量和重音,确定最终语音对应的情绪级别;基于最终语音对应的情绪级别,调整最终语音的语速、音调和重音。

[0025] 在一些实施例中,该装置还包括:修订单元,被配置成响应于检测到对原始文本中的文本片段的至少部分的修订操作,同步修订原始语音中的、与所修订的文本片段的至少部分关联的语音片段的至少部分。

[0026] 在一些实施例中,原始语音是即时通讯应用中的第一用户发送的;以及该装置还包括:发送单元,被配置成将最终语音发送至即时通讯应用的服务器,以使即时通讯应用的服务器将最终语音发送给即时通讯应用的第二用户。

[0027] 第三方面,本申请实施例提供了一种终端设备,该终端设备包括:一个或多个处理器;存储装置,其上存储有一个或多个程序;当一个或多个程序被一个或多个处理器执行,使得一个或多个处理器实现如第一方面中任一实现方式描述的方法。

[0028] 第四方面,本申请实施例提供了一种计算机可读介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如第一方面中任一实现方式描述的方法。

[0029] 本申请实施例提供的用于处理语音的方法和装置,首先对获取到的原始语音进行语音识别,以得到原始语音对应的原始文本;之后将原始语音中的语音片段与原始文本中的文本片段关联;然后识别原始语音和/或原始文本的异常片段;最后对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,以生成最终语音。将原始语音中的语音片段与原始文本中的文本片段关联,实现了语音的可视化处理。同时,避免了由于语音录制过程中出现异常状况而需要重新录制的问题,从而提高了语音录制效率。

附图说明

[0030] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本申请的其它特征、目的和优点将会变得更明显:

[0031] 图1是本申请可以应用于其中的示例性系统架构;

[0032] 图2是根据本申请的用于处理语音的方法的一个实施例的流程图;

[0033] 图3是根据本申请的用于处理语音的方法的又一个实施例的流程图;

[0034] 图4A是输入界面中的原始语音和原始文本的示意图;

[0035] 图4B是原始语音和原始文本的处理过程的示意图;

[0036] 图4C是输入界面中的最终语音和最终文本的示意图;

[0037] 图4D是聊天界面中的最终语音的示意图;

[0038] 图5是根据本申请的用于处理语音的装置的一个实施例的结构示意图;

[0039] 图6是适于用来实现本申请实施例的终端设备的计算机系统的结构示意图。

具体实施方式

[0040] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是,此处所描

述的具体实施例仅仅用于解释相关发明,而非对该发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与有关发明相关的部分。

[0041] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0042] 图1示出了可以应用本申请的用于处理语音的方法或用于处理语音的装置的实施例的示例性系统架构100。

[0043] 如图1所示,系统架构100可以包括设备101、102、103和网络104。网络104用以在设备101、102和设备103之间提供通信链路的介质。网络104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。

[0044] 设备101、102、103可以是支持网络连接从而提供各种网络服务的硬件设备或软件。当设备为硬件时,其可以是各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机、台式计算机和服务器等等。这时,作为硬件设备,其可以实现成多个设备组成的分布式设备群,也可以实现成单个设备。当设备为软件时,可以安装在上述所列举的电子设备中。这时,作为软件,其可以实现成例如用来提供分布式服务的多个软件或软件模块,也可以实现成单个软件或软件模块。在此不做具体限定。

[0045] 在实践中,设备可以通过安装相应的客户端应用或服务端应用来提供相应的网络服务。设备在安装客户端应用之后,其可以在网络通信中体现为客户端。相应地,在安装服务端应用之后,其可以在网络通信中体现为服务端。

[0046] 作为示例,在图1中,设备101、102体现为客户端,而设备103体现为服务端。具体地,设备101、102可以是安装有即时通讯应用的客户端,设备103可以是即时通讯应用的服务器。设备101可以获取原始语音;对原始语音进行语音识别,得到原始语音对应的原始文本;将原始语音中的语音片段与原始文本中的文本片段关联;识别原始语音和/或原始文本的异常片段;对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音;将最终语音发送至设备103。设备103可以将最终语音发送给设备102。

[0047] 需要说明的是,本申请实施例所提供的用于处理语音的方法可以由设备101执行。

[0048] 应该理解,图1中的网络 and 设备的数目仅仅是示意性的。根据实现需要,可以具有任意数目的网络和设备。

[0049] 继续参考图2,其示出了根据本申请的用于处理语音的方法的一个实施例的流程200。该用于处理语音的方法,包括以下步骤:

[0050] 步骤201,获取原始语音。

[0051] 在本实施例中,用于处理语音的方法的执行主体(例如图1所示的设备101)可以获取原始语音。通常,上述执行主体具有语音录制功能,以录制原始语音。

[0052] 步骤202,对原始语音进行语音识别,得到原始语音对应的原始文本。

[0053] 在本实施例中,上述执行主体可以对原始语音进行语音识别,以得到原始语音对应的原始文本。通常,上述执行主体可以提取原始语音的特征向量,并利用识别引擎对原始语音的特征向量进行识别,以得到原始语音对应的原始文本。

[0054] 在本实施例的一些可选的实现方式中,当原始语音中存在空白语音片段和/或拖长音语音片段时,上述执行主体可以将空白语音片段识别成第一预设符号(例如省略号),

和/或将拖长音语音片段识别成第二预设符号(例如长横线)。通常,上述执行主体可以利用声学模型识别原始语音中的空白语音片段和/或拖长音语音片段。其中,空白语音片段可以是原始语音中声音强度小于预设强度(例如15分贝),不包含人声特征声波,并且持续时长大于第一预设时长(例如1秒)的语音片段。拖长音语音片段可以是原始语音中声音强度不小于预设强度,不包含语义内容,并且持续时长大于第二预设时长(例如0.5秒)的语音片段。

[0055] 在本实施例的一些可选的实现方式中,上述执行主体可以基于空白语音片段的时长与第一预设时长的比值,确定将空白语音片段识别成的第一预设符号的数量;和/或基于拖长音语音片段的时长与第二预设时长的比值,确定将拖长音语音片段识别成的第二预设符号的数量。例如,空白语音片段中的每1秒识别成一个省略号。拖长音语音片段中的每0.5秒识别成一个长横线。

[0056] 步骤203,将原始语音中的语音片段与原始文本中的文本片段关联。

[0057] 在本实施例中,上述执行主体可以将原始语音中的语音片段与原始文本中的文本片段关联。通常,上述执行主体可以将原始语音切割成帧,以得到多个语音片段,并将每个语音片段与原始文本中对应的文本片段关联。这样,只要处理原始语音和原始文本之一,原始语音和原始文本均会被同步处理。

[0058] 步骤204,识别原始语音和/或原始文本的异常片段。

[0059] 在本实施例中,上述执行主体可以识别原始语音和/或原始文本中的异常片段。其中,异常片段可以包括但不限于以下至少一项:空白片段、重复片段、拖长音片段等等。

[0060] 步骤205,对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音。

[0061] 在本实施例中,上述执行主体可以对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,以生成最终语音。例如,上述执行主体可以首先删除原始文本中的异常片段指示的文本片段的至少部分,生成最终文本;然后对最终文本进行语音合成,生成最终语音。

[0062] 本申请实施例提供的用于处理语音的方法,首先对获取到的原始语音进行语音识别,以得到原始语音对应的原始文本;之后将原始语音中的语音片段与原始文本中的文本片段关联;然后识别原始语音和/或原始文本的异常片段;最后对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,以生成最终语音。将原始语音中的语音片段与原始文本中的文本片段关联,实现了语音的可视化处理。同时,避免了由于语音录制过程中出现异常状况而需要重新录制的问题,从而提高了语音录制效率。

[0063] 进一步参考图3,其示出了根据本申请的用于处理语音的方法的又一个实施例的流程300。该用于处理语音的方法,包括以下步骤:

[0064] 步骤301,获取即时通讯应用中的第一用户发送的原始语音。

[0065] 在本实施例中,用于处理语音的方法的执行主体(例如图1所示的设备101)可以获取即时通讯应用中的第一用户发送的原始语音。其中,上述执行主体可以安装有即时通讯应用。

[0066] 步骤302,对原始语音进行语音识别,得到原始语音对应的原始文本。

[0067] 步骤303,将原始语音中的语音片段与原始文本中的文本片段关联。

[0068] 步骤304,识别原始语音和/或原始文本的异常片段。

[0069] 在本实施例中,步骤302-304的具体操作已在图2所示的实施例中步骤202-204中进行了详细的介绍,在此不再赘述。

[0070] 步骤305,删除原始文本中的异常片段指示的文本片段的至少部分,以及同步删除原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分,生成最终语音。

[0071] 在本实施例中,上述执行主体可以删除原始文本中的异常片段指示的文本片段的至少部分。由于原始语音中的语音片段与原始文本中的文本片段关联。当删除原始文本中的异常片段指示的文本片段的至少部分时,上述执行主体会同步删除原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分。随后,上述执行主体可以基于删除处理后的原始语音,生成最终语音。通常,当异常片段是空白片段或拖长音片段时,上述执行主体删除原始文本中的异常片段指示的文本片段,并同步删除原始语音中的、与所删除的文本片段关联的语音片段。当异常片段是重复片段时,上述执行主体删除原始文本中的异常片段指示的文本片段的部分,并同步删除原始语音中的、与所删除的文本片段的部分关联的语音片段的部分。

[0072] 在本实施例的一些可选的实现方式中,当原始文本中存在第一预设符号和/或第二预设符号时,说明原始文本中存在空白文本片段和/或拖长音文本片段。此时,上述执行主体可以将第一预设符号和/或第二预设符号从原始文本中删除。

[0073] 在本实施例的一些可选的实现方式中,当原始文本中存在连续重复出现次数超过预设次数(例如4次)的词语时,说明原始文本中存在重复文本片段。此时,上述执行主体可以将除预设位置(例如重复文本片段中的最后一个词语)之外的词语从原始文本中删除。

[0074] 在本实施例的一些可选的实现方式中,响应于检测到对原始文本中的文本片段的至少部分的修订操作,上述执行主体会同步修订原始语音中的、与所修订的文本片段的至少部分关联的语音片段的至少部分。

[0075] 步骤306,对最终语音进行平滑处理。

[0076] 在本实施例中,上述执行主体可以对最终语音进行平滑处理,以使最终语音更加自然。

[0077] 在本实施例的一些可选的实现方式中,上述执行主体可以首先基于最终语音的语音特征,确定最终语音对应的方言类别;然后基于最终语音对应的方言类别,矫正最终语音中的音节,以及调整最终语音的重音。通常,上述执行主体可以将最终语音的语音特征与方言库中的各种类别的方言的语音特征进行匹配,并基于匹配结果确定最终语音对应的方言类别。

[0078] 在本实施例的一些可选的实现方式中,上述执行主体可以首先基于最终语音的音量和重音,确定最终语音对应的情绪级别;然后基于最终语音对应的情绪级别,调整最终语音的语速、音调和重音。通常,情绪级别越高,语速越快、重音越多、声音强度越强。例如,情绪级别可以被划分为5个等级。对于第1等级的情绪级别,其语速为正常语速的0.5倍,重音最少,声音强度比正常声音强度低10%。对于第3等级的情绪级别,其语速为正常语速,重音数量适中,声音强度为正常声音强度。对于第5等级的情绪级别,其语速为正常语速的1.5

倍,重音数量最多,声音强度比正常声音强度高10%。

[0079] 步骤307,将最终语音发送至即时通讯应用的服务器,以使即时通讯应用的服务器将最终语音发送给即时通讯应用的第二用户。

[0080] 在本实施例中,上述执行主体可以将最终语音发送至即时通讯应用的服务器(例如图1所示的设备103)。这样,即时通讯应用的服务器可以将最终语音发送给即时通讯应用的第二用户。

[0081] 本申请实施例还提供了用于处理语音的方法的一个应用场景。在该应用场景中,用户A和用户B的手机上安装有即时通讯应用。用户A的手机运行即时通讯应用,打开与用户B的聊天会话页。其中,聊天会话页包括输入界面和聊天界面。用户A点击语音录制控件,录制第一用户发送的原始语音。同时,对原始语音进行语音识别,得到原始语音对应的原始文本。这里,原始语音中的空白语音片段被识别成了省略号。原始语音中的语音片段与原始文本中的文本片段关联。具体如图4A所示,其示出了输入界面中的原始语音和原始文本的示意图。随后,用户A点击智能优化控件,识别原始文本中的异常片段。并删除原始文本中的异常片段指示的文本片段的至少部分,以及同步删除原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分。具体如图4B所示,其示出了原始语音和原始文本的处理过程的示意图。然后,用户点击优化完成控件,生成最终语音和最终文本。具体如图4C所示,其示出了输入界面中的最终语音和最终文本的示意图。最后,用户点击发送控件,将最终语音发送至即时通讯应用的服务器。即时通讯应用的服务器将最终语音发送给用户B。此时,用户A和用户B的聊天界面中均会显示最终语音。具体如图4D所示,其输出了聊天界面中的最终语音的示意图。

[0082] 从图3中可以看出,与图2对应的实施例相比,本实施例中的用于处理语音的方法的流程300突出了处理步骤。由此,本实施例描述的方案在删除原始文本中的内容时,同步删除原始语音中的相应内容,实现了语音的可视化处理。并且,对通过同步删除相应内容的原始语音而得到的最终语音进行平滑处理,使最终语音更加自然。此外,将用于处理语音的方法应用在即时通讯应用的语音交互过程中,提高了沟通效率。

[0083] 进一步参考图5,作为对上述各图所示方法的实现,本申请提供了一种用于处理语音的装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0084] 如图5所示,本实施例的用于处理语音的装置500可以包括:获取单元501、第一识别单元502、关联单元503、第二识别单元504和处理单元505。其中,获取单元501,被配置成获取原始语音;第一识别单元502,被配置成对原始语音进行语音识别,得到原始语音对应的原始文本;关联单元503,被配置成将原始语音中的语音片段与原始文本中的文本片段关联;第二识别单元504,被配置成识别原始语音和/或原始文本的异常片段;处理单元505,被配置成对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音。

[0085] 在本实施例中,用于处理语音的装置500中:获取单元501、第一识别单元502、关联单元503、第二识别单元504和处理单元505的具体处理及其所带来的技术效果可分别参考图2对应实施例中的步骤201、步骤202、步骤203、步骤204和步骤205的相关说明,在此不再赘述。

[0086] 在本实施例的一些可选的实现方式中,第一识别单元502进一步被配置成:当原始语音中存在空白语音片段和/或拖长音语音片段时,将空白语音片段识别成第一预设符号,和/或将拖长音语音片段识别成第二预设符号。

[0087] 在本实施例的一些可选的实现方式中,空白语音片段是原始语音中声音强度小于预设强度,不包含人声特征声波,并且持续时长大于第一预设时长的语音片段,拖长音语音片段是原始语音中声音强度不小于预设强度,不包含语义内容,并且持续时长大于第二预设时长的语音片段。

[0088] 在本实施例的一些可选的实现方式中,第一识别单元502进一步被配置成:基于空白语音片段的时长与第一预设时长的比值,确定将空白语音片段识别成的第一预设符号的数量;和/或基于拖长音语音片段的时长与第二预设时长的比值,确定将拖长音语音片段识别成的第二预设符号的数量。

[0089] 在本实施例的一些可选的实现方式中,处理单元505进一步被配置成:删除原始文本中的异常片段指示的文本片段的至少部分,生成最终文本;对最终文本进行语音合成,生成最终语音。

[0090] 在本实施例的一些可选的实现方式中,处理单元505进一步被配置成:删除原始文本中的异常片段指示的文本片段的至少部分,以及同步删除原始语音中的、与所删除的文本片段的至少部分关联的语音片段的至少部分,生成最终语音。

[0091] 在本实施例的一些可选的实现方式中,处理单元505进一步被配置成:当原始文本中存在第一预设符号和/或第二预设符号时,将第一预设符号和/或第二预设符号从原始文本中删除;当原始文本中存在连续重复出现次数超过预设次数的词语时,将除预设位置之外的词语从原始文本中删除。

[0092] 在本实施例的一些可选的实现方式中,用于处理语音的装置500还包括:平滑单元(图中未示出),被配置成对最终语音进行平滑处理。

[0093] 在本实施例的一些可选的实现方式中,平滑单元进一步被配置成:基于最终语音的语音特征,确定最终语音对应的方言类别;基于最终语音对应的方言类别,矫正最终语音中的音节,以及调整最终语音的重音。

[0094] 在本实施例的一些可选的实现方式中,平滑单元进一步被配置成:基于最终语音的音量和重音,确定最终语音对应的情绪级别;基于最终语音对应的情绪级别,调整最终语音的语速、音调和重音。

[0095] 在本实施例的一些可选的实现方式中,用于处理语音的装置500还包括:修订单元(图中未示出),被配置成响应于检测到对原始文本中的文本片段的至少部分的修订操作,同步修订原始语音中的、与所修订的文本片段的至少部分关联的语音片段的至少部分。

[0096] 在本实施例的一些可选的实现方式中,原始语音是即时通讯应用中的第一用户发送的;以及用于处理语音的装置500还包括:发送单元(图中未示出),被配置成将最终语音发送至即时通讯应用的服务器,以使即时通讯应用的服务器将最终语音发送给即时通讯应用的第二用户。

[0097] 下面参考图6,其示出了适于用来实现本申请实施例的终端设备(例如图1所示的设备101)的计算机系统600的结构示意图。图6示出的终端设备仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0098] 如图6所示,计算机系统600包括中央处理单元(CPU)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储部分608加载到随机访问存储器(RAM)603中的程序而执行各种适当的动作和处理。在RAM 603中,还存储有系统600操作所需的各种程序和数据。CPU 601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出(I/O)接口605也连接至总线604。

[0099] 以下部件连接至I/O接口605:包括键盘、鼠标等的输入部分606;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607;包括硬盘等的存储部分608;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609。通信部分609经由诸如因特网的网络执行通信处理。驱动器610也根据需要连接至I/O接口605。可拆卸介质611,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610上,以便于从其上读出的计算机程序根据需要被安装入存储部分608。

[0100] 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609从网络上被下载和安装,和/或从可拆卸介质611被安装。在该计算机程序被中央处理单元(CPU)601执行时,执行本申请的方法中限定的上述功能。需要说明的是,本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读介质或者是上述两者的任意组合。计算机可读介质例如可以是一—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(EEPROM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介质以外的任何计算机可读介质,该计算机可读介质可以发送、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。

[0101] 可以以一种或多种程序设计语言或其组合来编写用于执行本申请的操作的计算机程序代码,所述程序设计语言包括面向目标的设计语言—诸如Java、Smalltalk、C++,还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或终端设备上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)——连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0102] 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,该模块、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,框图和/或流程图中的每个方框、以及框图和/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。

[0103] 描述于本申请实施例中所涉及到的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。所描述的单元也可以设置在处理器中,例如,可以描述为:一种处理器包括获取单元、第一识别单元、关联单元、第二识别单元和处理单元。其中,这些单元的名称在某种情况下并不构成对该单元本身的限定,例如,获取单元还可以被描述为“获取原始语音的单元”。

[0104] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的终端设备中所包含的;也可以是单独存在,而未装配入该终端设备中。上述计算机可读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该终端设备执行时,使得该终端设备:获取原始语音;对原始语音进行语音识别,得到原始语音对应的原始文本;将原始语音中的语音片段与原始文本中的文本片段关联;识别原始语音和/或原始文本的异常片段;对原始文本中的异常片段指示的文本片段和/或原始语音中的异常片段指示的语音片段进行处理,生成最终语音。

[0105] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本申请中所涉及的发明范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

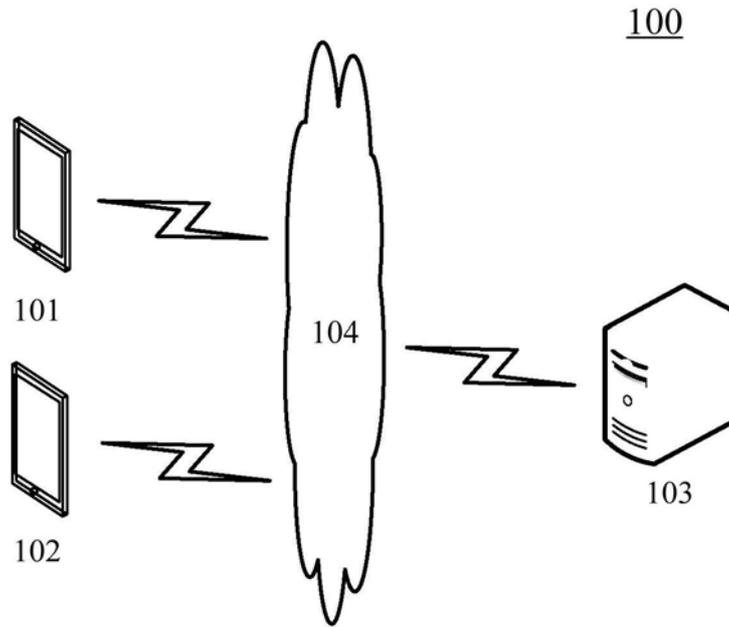


图1

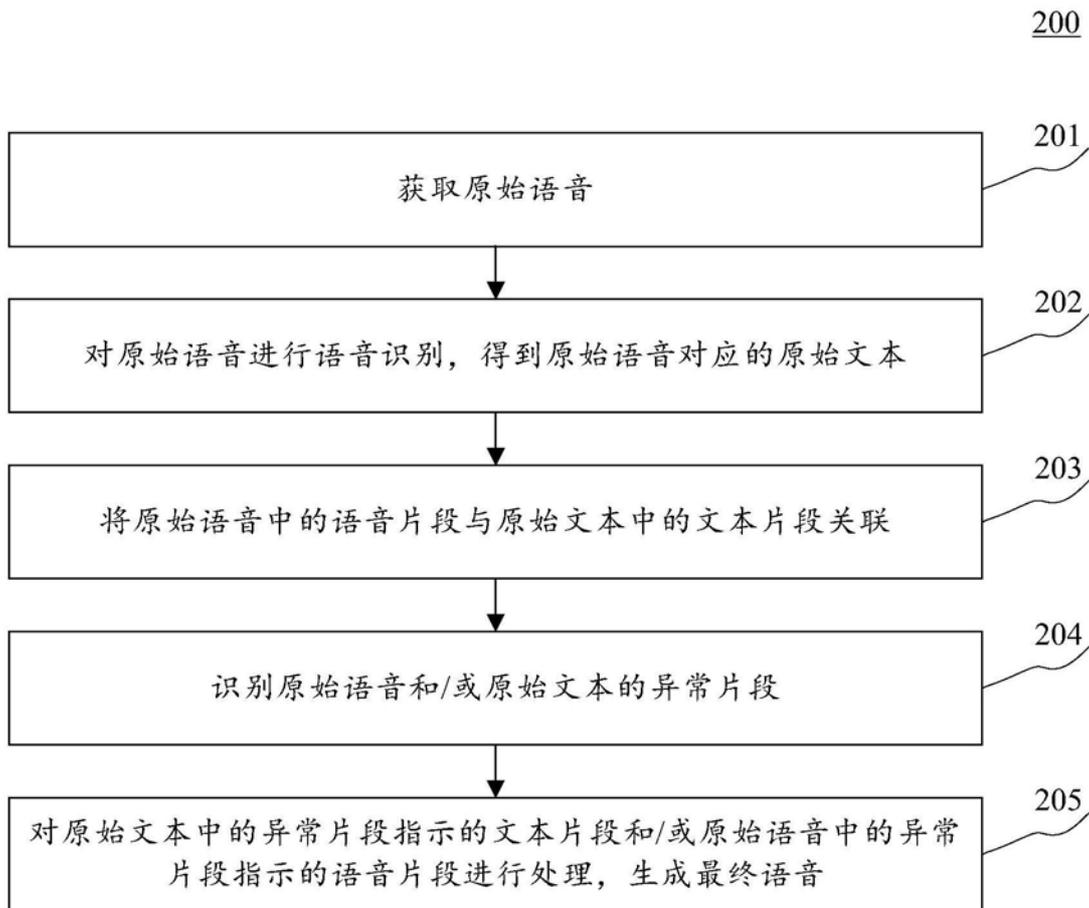


图2

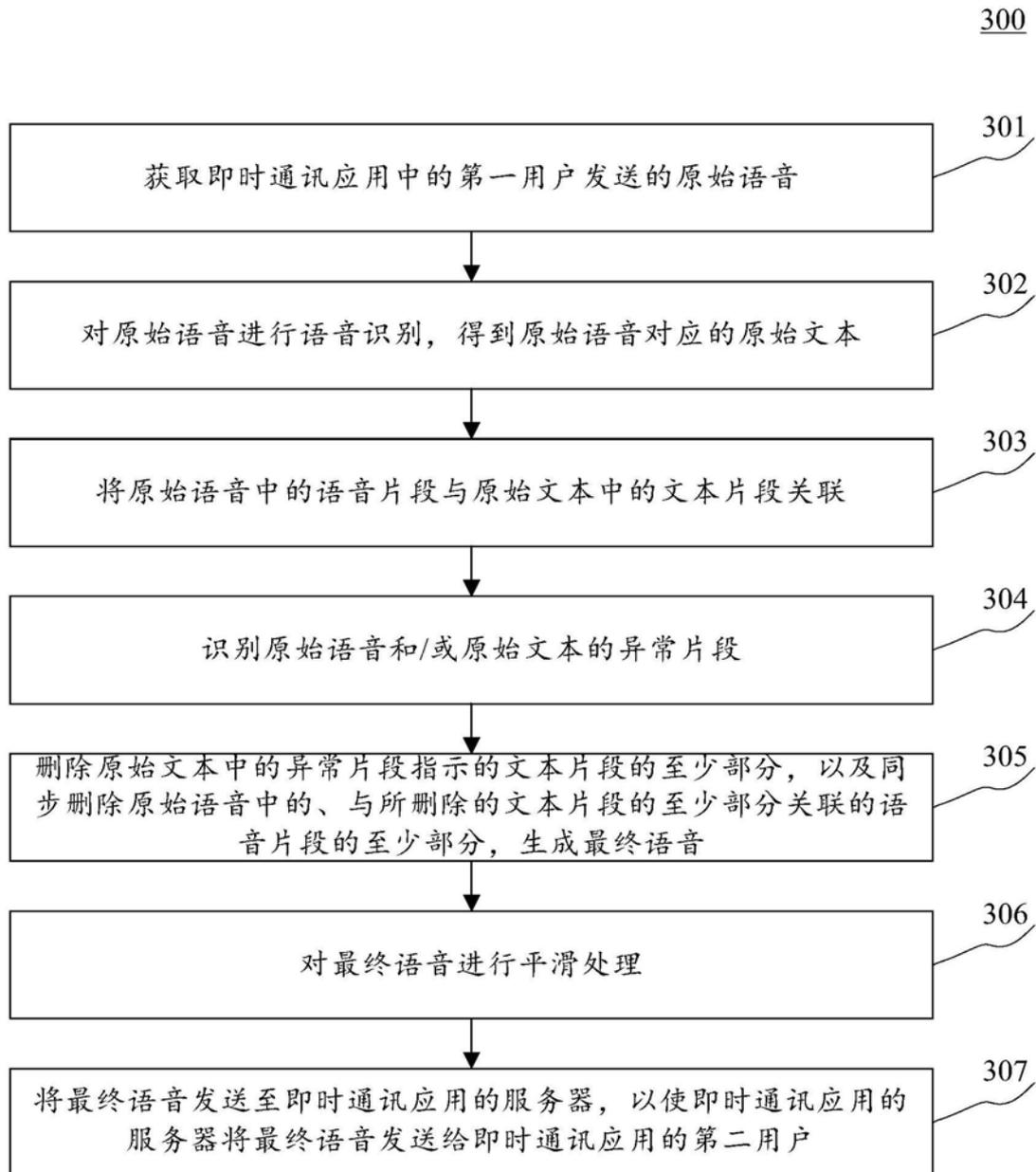


图3



图4A



图4B



图4C



图4D

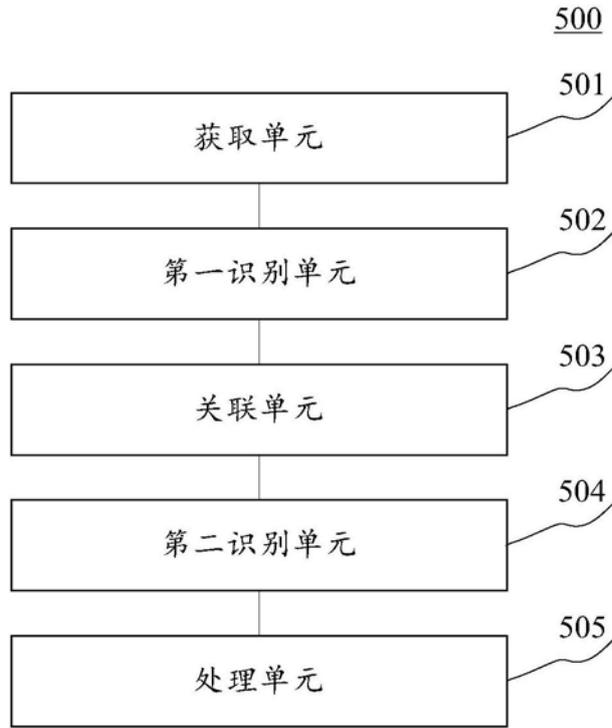


图5

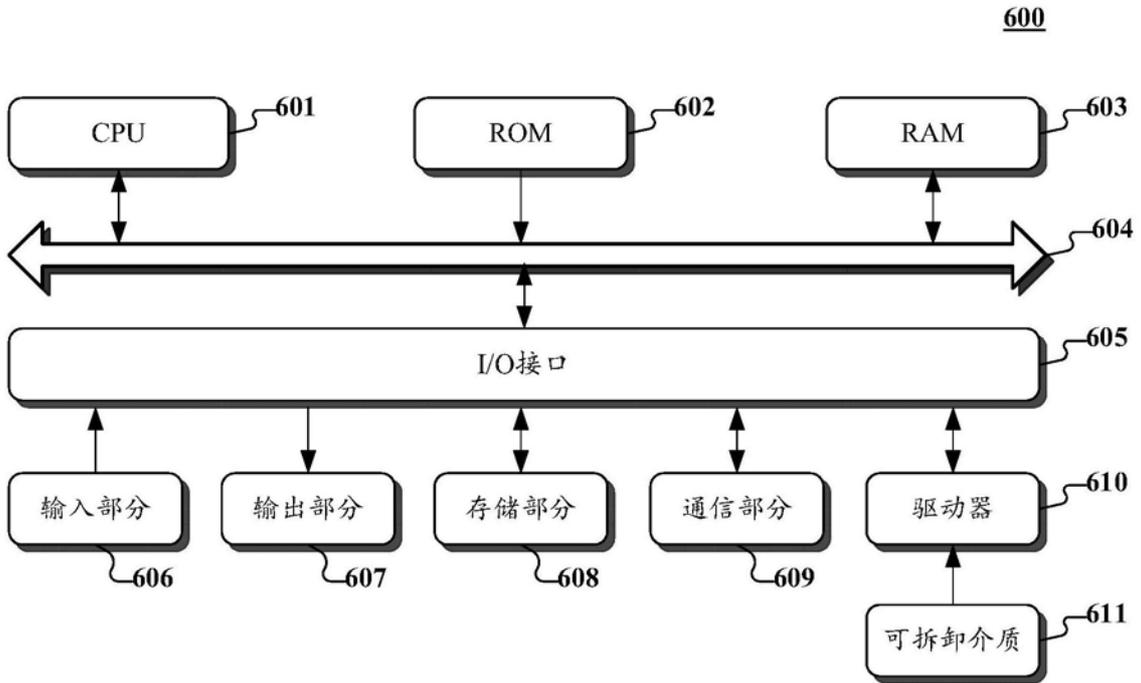


图6