



(12)发明专利申请

(10)申请公布号 CN 112802474 A

(43)申请公布日 2021.05.14

(21)申请号 201911028867.7

(22)申请日 2019.10.28

(71)申请人 中国移动通信有限公司研究院

地址 100053 北京市西城区宣武门西大街
32号

申请人 中国移动通信集团有限公司

(72)发明人 刘娟

(74)专利代理机构 北京派特恩知识产权代理有
限公司 11270

代理人 王姗姗 张颖玲

(51)Int.Cl.

G10L 15/26(2006.01)

G10L 15/06(2013.01)

G10L 15/22(2006.01)

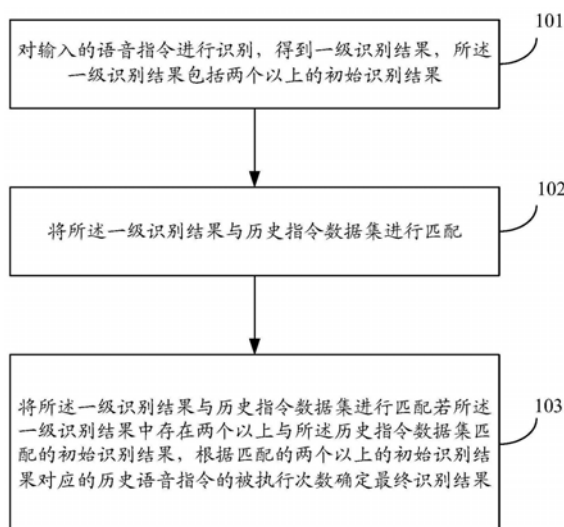
权利要求书2页 说明书11页 附图3页

(54)发明名称

语音识别方法、装置、设备及存储介质

(57)摘要

本发明公开了一种语音识别方法、装置、设备及存储介质。其中，该方法包括：对输入的语音指令进行识别，得到一级识别结果，所述一级识别结果包括两个以上的初始识别结果；将所述一级识别结果与历史指令数据集进行匹配，其中，所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数；若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果，根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。



1. 一种语音识别方法,其特征在于,包括:

对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果;

将所述一级识别结果与历史指令数据集进行匹配,其中,所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数;

若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。

2. 根据权利要求1所述的方法,其特征在于,每个所述初始识别结果具有对应的属性,所述方法还包括:

若所述一级识别结果中没有与所述历史指令数据集匹配的初始识别结果,将所述一级识别结果中各初始识别结果的属性与历史指令属性集进行匹配,根据匹配结果确定最终识别结果,其中,所述历史指令属性集包括所述历史语音指令对应的属性及所述历史语音指令对应的属性的被执行次数。

3. 根据权利要求2所述的方法,其特征在于,所述将所述一级识别结果与历史指令属性集进行匹配,根据匹配结果确定最终识别结果,包括:

若所述一级识别结果中仅存在一个与所述历史指令属性集匹配的初始识别结果,将所述初始识别结果确定为最终识别结果;

若所述一级识别结果中存在两个以上的与所述历史指令属性集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的属性的被执行次数确定最终识别结果;

若所述一级识别结果中没有与所述历史指令属性集匹配的初始识别结果,根据所述一级识别结果中各初始识别结果的语音识别匹配度确定最终识别结果。

4. 根据权利要求2所述的方法,其特征在于,所述方法还包括:

将所述最终识别结果输出给执行模块,根据所述执行模块反馈的信息更新所述历史指令数据集和所述历史指令属性集。

5. 根据权利要求4所述的方法,其特征在于,所述根据所述执行模块反馈的信息更新所述历史指令数据集和所述历史指令属性集,包括:

接收所述执行模块反馈的成功执行所述最终识别结果的反馈结果;

若所述历史指令数据集存在与所述反馈结果对应的历史语音指令,更新所述历史语音指令的被执行次数,若所述历史指令数据集不存在与所述反馈结果对应的历史语音指令,将所述最终识别结果加入所述历史指令数据集;

若所述历史指令属性集存在与所述反馈结果对应的属性,更新所述属性的被执行次数,若所述历史指令属性集不存在与所述反馈结果对应的属性,将所述最终识别结果的属性加入所述历史指令属性集。

6. 根据权利要求5所述的方法,其特征在于,所述历史指令数据集还包括历史语音指令的被执行时间,所述方法还包括:

基于所述历史语音指令的被执行时间确定不同区间内所述历史语音指令对应的被执行次数,根据不同区间内所述历史语音指令对应的被执行次数校正所述历史语音指令的被执行次数。

7. 根据权利要求5所述的方法,其特征在于,所述历史指令属性集还包括属性的被执行

时间,所述方法还包括:

基于所述属性的被执行时间确定不同区间内所述属性对应的被执行次数,根据不同区间内所述属性对应的被执行次数校正所述属性的被执行次数。

8.一种语音识别装置,其特征在于,包括:

语音识别模块,用于对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果;

匹配模块,用于将所述一级识别结果与历史指令数据集进行匹配,其中,所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数;若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。

9.一种语音识别设备,其特征在于,包括:处理器和用于存储能够在处理器上运行的计算机程序的存储器,其中,

所述处理器,用于运行计算机程序时,执行权利要求1至7任一项所述方法的步骤。

10.一种存储介质,所述存储介质上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时,实现权利要求1至7任一项所述方法的步骤。

语音识别方法、装置、设备及存储介质

技术领域

[0001] 本发明涉及语音识别技术,尤其涉及一种语音识别方法、装置、设备及存储介质。

背景技术

[0002] 目前,各类智能设备越来越成为人们日常生活必不可少的工具,如智能手表、智能手环、智能音箱等等,其中智能音箱除了能够音频播放的基本功能外,还具备了信息查询、定时闹钟、娱乐互动等更加丰富多彩的能力,为大家的生活带来了很大便利。

[0003] 相关技术中,智能音箱底层采用了语音识别的技术,对用户的语音指令进行识别和理解,从而转变为相应的文本或命令并进行执行。然而由于用户语调、发音等的变化,距离的远近,或者环境中其它声音的影响,非常容易导致智能音箱对用户输入的语音指令的识别不够精准,从而播放的内容并不是用户想要的,大大降低了用户体验,用户往往需要重复若干次语音指令才能得到满意的响应,给用户带来了不便。

发明内容

[0004] 有鉴于此,本发明实施例提供了一种语音识别方法、装置、设备及存储介质,旨在提高语音识别的准确率。

[0005] 本发明实施例的技术方案是这样实现的:

[0006] 本发明实施例提供了一种语音识别方法,包括:

[0007] 对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果;

[0008] 将所述一级识别结果与历史指令数据集进行匹配,其中,所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数;

[0009] 若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。

[0010] 本发明实施例还提供了一种语音识别装置,包括:

[0011] 语音识别模块,用于对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果;

[0012] 匹配模块,用于将所述一级识别结果与历史指令数据集进行匹配,其中,所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数;若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。

[0013] 本发明实施例又提供了一种语音识别设备,包括:处理器和用于存储能够在处理器上运行的计算机程序的存储器,其中,所述处理器,用于运行计算机程序时,执行本发明任一实施例所述方法的步骤。

[0014] 本发明实施例还提供了一种存储介质,所述存储介质上存储有计算机程序,所述

计算机程序被处理器执行时,实现本发明任一实施例所述方法的步骤。

[0015] 本发明实施例提供的技术方案,对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果,将所述一级识别结果与历史指令数据集进行匹配,若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果,可以对语音识别的一级识别结果进行纠偏,提高了语音识别的准确率,使得确定的最终识别结果符合用户的行为习惯和心理需求。

附图说明

[0016] 图1为本发明实施例语音识别方法的流程示意图;

[0017] 图2为本发明实施例语音识别装置的结构示意图;

[0018] 图3为本发明应用实施例语音识别装置的结构示意图;

[0019] 图4为本发明应用实施例中指令匹配的流程示意图;

[0020] 图5为本发明实施例语音识别设备的结构示意图。

具体实施方式

[0021] 下面结合附图及实施例对本发明再作进一步详细的描述。

[0022] 除非另有定义,本文所使用的所有的技术和科学术语与属于本发明的技术领域的技术人员通常理解的含义相同。本文中在本发明的说明书中所使用的术语只是为了描述具体的实施例的目的,不是旨在于限制本发明。

[0023] 相关技术中,智能音箱底层采用了语音识别的技术,对用户的语音指令进行识别和理解,从而转变为相应的文本或命令并进行执行。由于用户语调、发音等的变化,距离的远近,或者环境中其它声音的影响,非常容易导致智能音箱对用户输入的语音指令的识别不够精准。

[0024] 为了改善智能音箱的语音识别技术,存在以下相关技术:

[0025] 一、通过在音箱中嵌入噪声消除相关模块,从而在用户输入语音指令时,对用户的回声和空间中的背景噪音进行过滤消除,从而提高用户语音指令的辨识度;

[0026] 二、在房间中靠近用户处内置相关声音采集或者声音增强设备,并与智能音箱独立存在,该设备可对用户的声音进行采集,并传输到智能音箱,从而使得声音采集功能不局限于在智能音箱所在区域执行,扩大了智能音箱的拾音范围,进一步提高指令识别准确率;

[0027] 三、利用蓝牙耳机等额外的智能设备实现语音识别功能,以解决利用智能音箱进行语音识别所存在的语音识别距离有限及语音识别准确率低的问题。

[0028] 相关技术虽然能够部分解决由于环境嘈杂、用户距离太远等引起的智能音箱语音指令识别率低的问题,但对于用户偶发的发音不标准、咬字不清晰、音调变化等导致的识别率低的问题,没有有效的解决方案。比如,用户几乎每天晚上,都会在睡前通过语音指令“播放Kiss Baby”来控制智能音箱,给自己的宝宝播放摇篮曲“Kiss Baby”,轻柔舒缓又充满爱意的旋律和歌词可以很快让宝宝得到安抚,从而从疯狂的玩乐模式自然过渡到睡眠模式。然而有些时候,由于用户发音稍有不准,或者智能音箱语音识别模块的技术水平限制,就会将用户的语音指令识别错误,例如将“Kiss Baby”识别成“Kiss Me Baby”等相似的关键词,

从而播放了与“Kiss Baby”毫不相关的音乐,用户往往需要重复好几遍“播放 Kiss Baby”,直到智能音箱识别准确为止,不仅耗费了用户大量的时间和精力,且播放了错误的音乐有可能对用户的生活造成不便和干扰。

[0029] 基于此,在本发明的各种实施例中,对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果,将所述一级识别结果与历史指令数据集进行匹配,若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果,可以对语音识别的一级识别结果进行纠偏,提高了语音识别的准确率,使得确定的最终识别结果符合用户的行为习惯和心理需求。

[0030] 本发明实施例提供了一种语音识别方法,如图1所示,该方法包括:

[0031] 步骤101,对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果;

[0032] 这里,对输入的语音指令采用语音识别技术进行识别,得到识别概率满足设定要求的多个初始识别结果,该多个初始识别结果构成一级识别结果,该一级识别结果至少包括两个初始识别结果,从而需要执行后续的初始识别结果的纠偏。

[0033] 实际应用时,可以扩大识别结果的覆盖范围,对扩大覆盖范围后的多个识别结果选取识别概率达到设定概率以上的识别结果作为一级识别结果,或者选取设定数量个识别概率靠前的识别结果作为一级识别结果。

[0034] 步骤102,将所述一级识别结果与历史指令数据集进行匹配;

[0035] 这里,所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数。

[0036] 实际应用时,历史指令数据集可以本地获取或者通过网络获取。

[0037] 步骤103,若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。

[0038] 实际应用时,若所述一级识别结果与所述历史指令数据集中历史语音指令的匹配结果仅存在一个初始识别结果,则确定该初始识别结果对应的历史语音指令为最终识别结果。

[0039] 若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。具体地,选择被执行次数最多的历史语音指令为最终识别结果。这样,使得最终识别结果更加准确,更加符合用户的喜好及使用习惯。

[0040] 在一实施例中,每个所述初始识别结果具有对应的属性,所述方法还包括:

[0041] 若所述一级识别结果中没有与所述历史指令数据集匹配的初始识别结果,将所述一级识别结果中各初始识别结果的属性与历史指令属性集进行匹配,根据匹配结果确定最终识别结果,其中,所述历史指令属性集包括所述历史语音指令对应的属性及所述历史语音指令对应的属性的被执行次数。

[0042] 这里,属性可以为音乐的曲风属性。示例性地,音乐的曲风属性可以包括:儿歌、流行音乐、民谣、摇滚等。

[0043] 在一实施例中,所述将所述一级识别结果与历史指令属性集进行匹配,根据匹配结果确定最终识别结果,包括:

[0044] 若所述一级识别结果中仅存在一个与所述历史指令属性集匹配的初始识别结果,将所述初始识别结果确定为最终识别结果;

[0045] 若所述一级识别结果中存在两个以上的与所述历史指令属性集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的属性的被执行次数确定最终识别结果;

[0046] 若所述一级识别结果中没有与所述历史指令属性集匹配的初始识别结果,根据所述一级识别结果中各初始识别结果的语音识别匹配度确定最终识别结果。

[0047] 这样,可以在一级识别结果与历史指令数据集匹配失败的情况下,根据一级识别结果与历史指令属性集的匹配结果确定最终识别结果,从而有效识别出最终识别结果。若在一级识别结果与历史指令属性集匹配失败的情况下,还可以根据各初始识别结果的语音匹配度(即识别概率)确定最终识别结果,即选择识别概率最高的初始识别结果作为最终识别结果。

[0048] 为了满足历史指令数据集和历史指令属性集能够根据用户的使用习惯进行更新,基于此,在一实施例中,所述方法还包括:

[0049] 将所述最终识别结果输出给执行模块,根据所述执行模块反馈的信息更新所述历史指令数据集和所述历史指令属性集。

[0050] 在一实施例中,所述根据所述执行模块反馈的信息更新所述历史指令数据集和所述历史指令属性集,包括:

[0051] 接收所述执行模块反馈的成功执行所述最终识别结果的反馈结果;

[0052] 若所述历史指令数据集存在与所述反馈结果对应的历史语音指令,更新所述历史语音指令的被执行次数,若所述历史指令数据集不存在与所述反馈结果对应的历史语音指令,将所述最终识别结果加入所述历史指令数据集;

[0053] 若所述历史指令属性集存在与所述反馈结果对应的属性,更新所述属性的被执行次数,若所述历史指令数据集不存在与所述反馈结果对应的属性,将所述最终识别结果的属性加入所述历史指令属性集。

[0054] 实际应用时,若执行模块成功执行了最终识别结果,则根据执行模块的反馈的执行该最终识别结果的信息(即反馈结果),更新历史指令数据集和历史指令属性集。具体地,对于历史指令数据集中对应的历史语音指令的执行次数进行加1操作,对于历史指令属性集中对应的属性的执行次数进行加1操作。若所述历史指令数据集不存在与所述反馈结果对应的历史语音指令,将所述最终识别结果加入所述历史指令数据集,且执行次数赋值为1;若所述历史指令属性集不存在与所述反馈结果对应的属性,将所述最终识别结果的属性加入所述历史指令属性集,且执行次数赋值为1。

[0055] 在一实施例中,所述历史指令数据集还包括历史语音指令的被执行时间,所述方法还包括:

[0056] 基于所述历史语音指令的被执行时间确定不同区间内所述历史语音指令对应的被执行次数,根据不同区间内所述历史语音指令对应的被执行次数校正所述历史语音指令的被执行次数。

[0057] 这里,可以将历史有效时长 T 划分为相等的 n 个时长区间, n 为大于2的自然数,根据

各历史语音指令的被执行时间将各历史语音指令划分至对应的区间,且对各区间赋予相应的影响权重,基于不同区间内所述历史语音指令对应的被执行次数和相应的影响权重,校正所述历史语音指令的被执行次数。

[0058] 在一实施例中,所述历史指令属性集还包括属性的被执行时间,所述方法还包括:

[0059] 基于所述属性的被执行时间确定不同区间内所述属性对应的被执行次数,根据不同区间内所述属性对应的被执行次数校正所述属性的被执行次数。

[0060] 同理,可以根据各属性的被执行时间将各属性划分至对应的区间,且对各区间赋予相应的影响权重,基于不同区间内所述属性对应的被执行次数和相应的影响权重,校正所述属性的被执行次数。

[0061] 通过对历史指令数据集中历史语音指令的被执行次数进行校正,可以在根据历史语音指令的被执行次数确定最终识别结果时,更符合用户的历史行为习惯,提高识别结果的准确率。

[0062] 通过对历史指令属性集中属性的被执行次数进行校正,可以在根据历史指令属性集中属性的被执行次数确定最终识别结果时,更符合用户的历史行为习惯,提高识别结果的准确率。

[0063] 在一实施例中,影响权重可以基于时间因子确定,还可以基于不同的用户群体进行确定。以基于时间因子为例,可以将影响权重设置为随着远离当前时间而减少,接近当前时间而增大。

[0064] 为了实现本发明实施例的方法,本发明实施例还提供一种语音识别装置,该语音识别装置可以设置于智能音箱处,或者单独设置并与智能音箱通信连接。

[0065] 如图2所示,该装置包括:

[0066] 语音识别模块201,用于对输入的语音指令进行识别,得到一级识别结果,所述一级识别结果包括两个以上的初始识别结果;

[0067] 匹配模块202,用于将所述一级识别结果与历史指令数据集进行匹配,其中,所述历史指令数据集包括语音识别设备执行的历史语音指令及所述历史语音指令的被执行次数;若所述一级识别结果中存在两个以上与所述历史指令数据集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的历史语音指令的被执行次数确定最终识别结果。

[0068] 在一实施例中,每个所述初始识别结果具有对应的属性,所述匹配模块202 还用于:若所述一级识别结果中没有与所述历史指令数据集匹配的初始识别结果,将所述一级识别结果中各初始识别结果的属性与历史指令属性集进行匹配,根据匹配结果确定最终识别结果,其中,所述历史指令属性集包括所述历史语音指令对应的属性及所述历史语音指令对应的属性的被执行次数。

[0069] 在一实施例中,所述匹配模块202具体用于:

[0070] 若所述一级识别结果中仅存在一个与所述历史指令属性集匹配的初始识别结果,将所述初始识别结果确定为最终识别结果;

[0071] 若所述一级识别结果中存在两个以上的与所述历史指令属性集匹配的初始识别结果,根据匹配的两个以上的初始识别结果对应的属性的被执行次数确定最终识别结果;

[0072] 若所述一级识别结果中没有与所述历史指令属性集匹配的初始识别结果,根据所述一级识别结果中各初始识别结果的语音识别匹配度确定最终识别结果。

[0073] 在一实施例中,所述装置还包括更新模块203,用于将所述最终识别结果输出给执行模块,根据所述执行模块反馈的信息更新所述历史指令数据集和所述历史指令属性集。

[0074] 在一实施例中,更新模块203具体用于:

[0075] 接收所述执行模块反馈的成功执行所述最终识别结果的反馈结果;

[0076] 若所述历史指令数据集存在与所述反馈结果对应的历史语音指令,更新所述历史语音指令的被执行次数,若所述历史指令数据集不存在与所述反馈结果对应的历史语音指令,将所述最终识别结果加入所述历史指令数据集;

[0077] 若所述历史指令属性集存在与所述反馈结果对应的属性,更新所述属性的被执行次数,若所述历史指令属性集不存在与所述反馈结果对应的属性,将所述最终识别结果的属性加入所述历史指令属性集。

[0078] 在一实施例中,所述历史指令数据集还包括历史语音指令的被执行时间,所述更新模块203具体用于:基于所述历史语音指令的被执行时间确定不同区间内所述历史语音指令对应的被执行次数,根据不同区间内所述历史语音指令对应的被执行次数校正所述历史语音指令的被执行次数。

[0079] 在一实施例中,所述历史指令属性集还包括属性的被执行时间,所述更新模块203具体用于:基于所述属性的被执行时间确定不同区间内所述属性对应的被执行次数,根据不同区间内所述属性对应的被执行次数校正所述属性的被执行次数。

[0080] 实际应用时,语音识别模块201、匹配模块202及更新模块203,可以由语音识别装置中的处理器来实现。当然,处理器需要运行存储器中的计算机程序来实现它的功能。

[0081] 需要说明的是:上述实施例提供的语音识别装置在进行语音识别时,仅以上述各程序模块的划分进行举例说明,实际应用中,可以根据需要而将上述处理分配由不同的程序模块完成,即将装置的内部结构划分成不同的程序模块,以完成以上描述的全部或者部分处理。另外,上述实施例提供的语音识别装置与语音识别方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0082] 下面结合应用实施例对本发明再作进一步详细的描述。

[0083] 如图3所示,本应用实施例语音识别装置包括:用户输入语音识别模块(相当于语音识别模块)、指令匹配模块(相当于匹配模块)、结果输出及反馈模块(具有更新模块的功能)、用户历史指令数据存储模块及用户历史指令属性存储模块。下面对各模块的功能说明如下:

[0084] 1、用户输入语音识别模块

[0085] 用户输入语音识别模块对用户通过语音输入的指令I进行识别,并初步对应到相应的可执行的指令,例如用户说“我想听Kiss Baby”,则该模块解析到“听”对应为播放动作,“Kiss Baby”对应到播放的相关内容如音乐、故事等。注意此处与传统智能音箱的语音识别要求不同,为了扩大识别结果的覆盖范围,以便后续再进一步结合历史指令数据进行匹配,此模块需要针对每一条用户的语音指令产生多个概率较高的识别结果,并记录这些结果所对应的属性,即一级识别结果集合 $R = \{r_1, r_2, \dots, r_n\}$,及其属性 $\{c_1, c_2, \dots, c_n\}$,其中 r_i 为每一条可能的识别结果, c_i 为该指令对应的属性, n 为一级识别结果的数量, n 可以根据经验值设置为一个固定值例如3,也可以根据识别匹配度来限定,例如取识别匹配度大于 95%的所有识别结果。例如语音识别模块可能将用户说的“Kiss Baby”识别

为以下几种结果:“Kiss Me Baby”,其中匹配度99%,”Kiss Baby”,其中匹配度98%,”Cheers Baby”,其中匹配度97%。

[0086] 2、用户历史指令数据存储模块

[0087] 用户历史指令数据存储模块存储了用户在该智能音箱上执行过的所有历史指令数据D(即历史指令数据集),代表了用户使用智能音箱的偏好,包括但不限于用户通过语音输入执行的指令、用户通过智能终端上操控智能音箱的APP 来执行的指令等,从而统计得出每个指令被执行的次数,即 $D = \{(d_1, f_{d_1}), (d_2, f_{d_2}), \dots, (d_n, f_{d_n})\}$ 。

[0088] 3、用户历史指令属性存储模块

[0089] 用户历史指令属性存储模块则存储了用户在该智能音箱上执行过的所有历史指令所对应的属性数据 $C = \{c_1, c_2, \dots, c_n\}$ (即历史指令属性集),例如对于播放音乐类的指令,则存储播放的音乐属性,如儿歌、流行音乐、摇滚等,从而统计得出每类属性的指令被执行的次数,即 $\{f_{c_1}, f_{c_2}, \dots, f_{c_n}\}$ 。

[0090] 在实际应用中还可以考虑到这些数据代表的用户使用智能音箱的偏好的准确程度随着时间不断往前追溯而不断衰减的情况,例如用户在2019年2月在智能音箱上执行过的历史指令数据,对用户偏好的反应准确度,必然比2018年3 月的数据要高。这里,可建立历史指令数据影响力的时间衰减模型。此时用户历史指令数据存储模块需要记录每条指令被执行的时间,并据此计算每段时间内(例如过去每个月)每个指令被执行的次数(每成功的播放一次则次数加1)。用户历史指令属性存储模块则据此计算每段时间内(例如过去每个月)每类属性的指令被执行的次数(每成功的播放一次则次数加1)。

[0091] 若考虑时间周期T内(例如2018年3月到2019年2月的12个月)用户的历史指令数据,可将T划分为n个阶段(例如从2018年3月起以每个月划分为一个阶段,一直划分到2019年2月),对这n个阶段按照从近到远的时间顺序进行排序,其次序记为集合 $\{1, 2, \dots, n\}$,则对于某个指令 d_i ,分别存储每个阶段内该指令的执行次数 $f_{d_i} = \{f_{d_i,1}, f_{d_i,2}, \dots, f_{d_i,n}\}$ 。对于某类属性的指令 c_i ,分别存储每个阶段内该类属性指令的执行次数 $f_{c_i} = \{f_{c_i,1}, f_{c_i,2}, \dots, f_{c_i,n}\}$ 。

[0092] 同时考虑到每个阶段的指令对整个周期内用户偏好的影响不同,引入影响力权重,即每个时间阶段i对整个周期的影响权重为 λ_i ,

[0093] $0 < \lambda_i \leq 1$,

[0094] $\sum_{i=1}^n \lambda_i = 1$,

[0095] λ_i 的平均值 $\bar{\lambda}_i = 1/n$

[0096] 则考虑到时间衰减因素后,周期T内某个指令 d_i 的校正播放次数可定义为:

[0097] $\overline{f_{d_i}} = \sum_{j=0}^n \lambda_j * f_{d_i,j}$

[0098] 周期T内某类属性指令 c_i 的校正播放次数可定义为:

[0099] $\overline{f_{c_i}} = \sum_{j=0}^n \lambda_j * f_{c_i,j}$

[0100] 其中权重 λ_i 在实际应用过程中其取值可根据各指令影响力随着时间周期衰减的

实际程度来灵活调整。即引入权重调整因子 η_i ,

$$[0101] \quad \{\lambda_1, \lambda_2, \dots, \lambda_n\} = \{\bar{\lambda}_1 + \eta_1, \bar{\lambda}_2 + \eta_2, \dots, \bar{\lambda}_n + \eta_n\},$$

$$[0102] \quad \eta_1 > \eta_2 > \dots > \eta_n$$

$$[0103] \quad -1/n < \eta_i < (n-1)/n,$$

$$[0104] \quad \sum_{i=1}^n \eta_i = 0。$$

[0105] 在调整时,可以考虑不同特征属性的用户其适用的指令影响力衰减模型的不同,即对于偏好比较固定的群体如老年人等,其历史指令数据对用户偏好的反应程度受到时间衰减的影响比较小,因此每个阶段 λ_i 的取值可平均一些,即各调整因子范围窄一些,例如可赋予限定条件: η_i 的标准差 <0.01 ;而对于另外一部分偏好变化快的群体如年轻人等,其历史指令数据对用户偏好的反应程度受到时间衰减的影响较大,因此每个阶段 λ_i 取值的差距可大一些,即各调整因子范围可大一些,例如可赋予限定条件 η_i 的标准差 >0.1 。

[0106] 4、指令匹配模块

[0107] 指令匹配模块则基于用户输入语音识别模块输出的一级识别结果集合 $R = \{(r_1, c_1), (r_2, c_2), \dots, (r_n, c_n)\}$,与用户历史指令数据存储模块和用户历史指令属性存储模块中的数据进行匹配,从而最终输出唯一的识别结果(即最终识别结果)。

[0108] 指令匹配的具体流程如图4所示。例如,语音识别模块可能将用户说的“Kiss Baby”识别为以下几种结果:“Kiss Me Baby”,其中匹配度99%,”Kiss Baby”,其中匹配度98%,”Cheers Baby”,其中匹配度97%,其中“Kiss Baby”的指令属性为“儿歌”,“Kiss Me Baby”的指令属性为“摇滚音乐”,“Cheers Baby”的属性为“民俗音乐”。具体匹配方法包括:

[0109] 对于用户语音输入的指令I,先经过用户输入语音识别模块进行初步识别,得到一级识别结果 $R = \{r_1, r_2, \dots, r_n\}$,及每条结果对应的属性 $\{c_1, c_2, \dots, c_n\}$ 。然后先用一级识别结果同用户历史指令数据存储模块中的数据进行匹配,匹配方法为取一级识别结果R与历史指令数据D的交集 $R \cap D = \{d_1, d_2, \dots, d_m\}$,即当前可能的结果中有哪些是用户在周期T内播放过的:

[0110] 1、若一级识别结果R与历史指令数据D的交集只有一条指令,即当前可能的结果中只有一条是用户在周期T内播放过的,则最终匹配结果 R_{final} 即为这条指令,输出到结果输出及反馈模块。例如用户历史指令数据存储模块中只有“Kiss Baby”这一条指令播放记录,则最终匹配结果为“Kiss Baby”。

[0111] 2、若一级识别结果R与历史指令数据D的交集中有多条指令,即当前可能的结果中有多条是用户在周期T内播放过的,则可取最能代表用户偏好的那条指令最为匹配结果。即进一步根据这些指令在周期T内每条指令的校正播放次数 $\{\overline{f_{d_1}}, \overline{f_{d_2}}, \dots, \overline{f_{d_m}}\}$,取次数最

高的一条指令作为匹配结果: $R_{final} = d_i, d_i \in R \cap D$, 且 $\overline{f_{d_i}} = \max \{\overline{f_{d_1}}, \overline{f_{d_2}}, \dots, \overline{f_{d_m}}\}$ 。例如

用户历史指令数据存储模块中有“Kiss Baby”和“Kiss Me Baby”这两条指令播放记录,且“Kiss Baby”在过去一年内的校正播放次数为132次,“Kiss Me Baby”在过去一年内的校正播放次数为20次,则最终匹配结果为“Kiss Baby”。

[0112] 若次数最高的指令也有多个,即有好几条指令对用户偏好的代表程度都一样,即

$R_{middle} = \{d_i, d_j, \dots, d_a\}, \{d_i, d_j, \dots, d_a\} \in R \cap D,$

[0113] 且 $\overline{f_{d_i}} = \overline{f_{d_j}} = \dots = \overline{f_{d_a}} = \max \{\overline{f_{d_1}}, \overline{f_{d_2}}, \dots, \overline{f_{d_m}}\}$

[0114] 则可在这些指令中,取最能代表用户对指令的属性的偏好程度的那条指令作为最为匹配结果。即将这些指令再去与用户指令属性存储模块进行匹配,匹配方法为分别取 $R_{middle} = \{d_i, d_j, \dots, d_a\}$ 的属性 $\{c_i, c_j, \dots, c_a\}$, 根据用户历史指令属性存储模块所存储的数据,得出每类属性的指令在周期T内的校正播放次数 $\{\overline{f_{c_i}}, \overline{f_{c_j}}, \dots, \overline{f_{c_e}}\}$, 取次数最高的那类属性,并输出该属性所对应的那条指令作为最终匹配结果进行输出即 $R_{final} = d_n, d_n \in R_{middle}$, 且 $\overline{f_{c_n}} = \max \{\overline{f_{c_i}}, \overline{f_{c_j}}, \dots, \overline{f_{c_e}}\}$ 。例如用户历史指令数据存储模块中有“Kiss Baby”和“Kiss Me Baby”这两条指令播放记录,这两条指令在过去一年内的校正播放次数都为为86次,其中“Kiss Baby”的指令属性为儿歌,“Kiss Me Baby”的指令属性为摇滚音乐,且在过去一年内儿歌属性的指令的校正播放次数为1256次,摇滚音乐属性的指令的校正播放次数为326次,则最终匹配结果为儿歌属性对应的指令“Kiss Baby”。

[0115] 3、若一级识别结果R与历史指令数据D的交集为空,即当前可能的结果中没有一条是用户在周期T内播放过的,此处无法根据用户历史指令数据对一级识别结果进行进一步匹配,则可根据用户历史指令属性数据进行匹配,即在一级识别结果中取最能代表用户对指令的属性的偏好程度的那条指令作为最为匹配结果。匹配方法为分别取一级识别结果 $R = \{r_1, r_2, \dots, r_n\}$ 的属性 $\{c_1, c_2, \dots, c_n\}$, 与用户历史指令属性存储模块中所存储的数据 $C = \{c_1, c_2, \dots, c_n\}$ 取交集,得到属性匹配结果 $\{c_i, c_j, \dots, c_e\}$, 以及每类属性的指令在周期T内的校正播放次数 $\{\overline{f_{c_i}}, \overline{f_{c_j}}, \dots, \overline{f_{c_e}}\}$, 即当前可能的结果中有哪些指令的属性是用户在周期T内播放过的:

[0116] 1)、若一级识别结果R与历史指令属性数据D的交集只有一条指令,即当前可能的结果中只有一条指令的属性是用户在周期T内播放过的,则最终匹配结果 R_{final} 即为这条指令,输出到结果输出及反馈模块。例如“Kiss Baby”、“Kiss Me Baby”和“Cheers Baby”这三条指令均不在用户历史指令数据存储模块中,且用户历史指令属性存储模块中只有“儿歌”属性类的指令播放记录,没有“摇滚音乐”和“民俗音乐”属性类的指令播放记录,则最终匹配结果为“儿歌”属性类对应的指令“Kiss Baby”。

[0117] 2)、若一级识别结果R与历史指令属性数据D的交集中有多条指令,即当前可能的结果中有多条指令的属性是用户在周期T内播放过的,则可取最能代表用户偏好的那类属性所对应的指令最为匹配结果。即取属性匹配结果 $\{c_i, c_j, \dots, c_e\}$ 中在周期T内的校正播放次数 $\{\overline{f_{c_i}}, \overline{f_{c_j}}, \dots, \overline{f_{c_e}}\}$ 最高的那类属性,并输出该属性所对应的那条指令作为最终匹配结果

$R_{final} = d_n, d_n \in R$, 且 $\overline{f_{c_n}} = \max \{\overline{f_{c_i}}, \overline{f_{c_j}}, \dots, \overline{f_{c_e}}\}$ 。例如“Kiss Baby”、“Kiss Me Baby”和“Cheers Baby”这三条指令均不在用户历史指令数据存储模块中,且在过去一年内儿歌属性的指令的校正播放次数为1256次,摇滚音乐属性的指令的校正播放次数为326次,没有“民俗音乐”属性类的指令播放记录,则最终匹配结果为儿歌属性对应的指令“Kiss Baby”。

[0118] 3)、若一级识别结果R与历史指令属性数据C的交集也为空,即当前可能的结果中

没有一条指令的属性是用户在周期T内播放过的,则此处即无法根据用户历史指令数据对识别结果进行进一步匹配,也无法根据用户历史指令属性数据进行匹配。此时可认为当前一级识别结果R中没有任何指令可以体现用户的偏好,则可选择一级识别结果中匹配度最高的指令之间输出。例如“Kiss Me Baby”的匹配度99%为最高,则最终输出结果为“Kiss Me Baby”。

[0119] 5、结果输出及反馈模块

[0120] 该模块接收指令匹配模块输出的匹配结果,输出到智能音箱的相关指令执行模块从而进行指令的执行,并且在指令执行完成(例如播放一次音乐)后,将该条指令和其播放时间反馈到用户历史指令数据存储模块进行存储,将该条指令的属性和其播放时间反馈到用户历史指令属性存储模块进行存储。此时用户历史指令数据存储模块和用户历史指令属性存储模块也需要根据此反馈对所存储的指令和指令属性数据进行更新,即对播放次数进行对应增加。

[0121] 基于上述程序模块的硬件实现,且为了实现本发明实施例的方法,本发明实施例还提供一种语音识别设备。图5仅仅示出了该语音识别设备的示例性结构而非全部结构,根据需要可以实施图5示出的部分结构或全部结构。

[0122] 如图5所示,本发明实施例提供的语音识别设备500包括:至少一个处理器501、存储器502、和至少一个网络接口503。语音识别设备500中的各个组件通过总线系统504耦合在一起。可以理解,总线系统504用于实现这些组件之间的连接通信。总线系统504除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见,在图5中将各种总线都标为总线系统504。

[0123] 本发明实施例中的存储器502用于存储各种类型的数据以支持语音识别设备的操作。这些数据的示例包括:用于在语音识别设备上操作的任何计算机程序。

[0124] 本发明实施例揭示的语音识别方法可以应用于处理器501中,或者由处理器501实现。处理器501可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,语音识别方法的各步骤可以通过处理器501中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器501可以是通用处理器、数字信号处理器(DSP, Digital Signal Processor),或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。处理器501可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者任何常规的处理器等。结合本发明实施例所公开的方法的步骤,可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于存储介质中,该存储介质位于存储器502,处理器501 读取存储器502中的信息,结合其硬件完成本发明实施例提供的语音识别方法的步骤。

[0125] 在示例性实施例中,语音识别设备可以被一个或多个应用专用集成电路(ASIC, Application Specific Integrated Circuit)、DSP、可编程逻辑器件(PLD, Programmable Logic Device)、复杂可编程逻辑器件(CPLD, Complex Programmable Logic Device)、FPGA、通用处理器、控制器、微控制器(MCU, Micro Controller Unit)、微处理器(Microprocessor)、或者其他电子元件实现,用于执行前述方法。

[0126] 可以理解,存储器502可以是易失性存储器或非易失性存储器,也可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(ROM, Read Only

Memory)、可编程只读存储器(PROM,Programmable Read-Only Memory)、可擦除可编程只读存储器(EPROM,Erasable Programmable Read-Only Memory)、电可擦除可编程只读存储器(EEPROM,Electrically Erasable Programmable Read-Only Memory)、磁性随机存取存储器(FRAM,ferromagnetic random access memory)、快闪存储器(Flash Memory)、磁表面存储器、光盘、或只读光盘(CD-ROM,Compact Disc Read-Only Memory);磁表面存储器可以是磁盘存储器或磁带存储器。易失性存储器可以是随机存取存储器(RAM,Random Access Memory),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用,例如静态随机存取存储器(SRAM,Static Random Access Memory)、同步静态随机存取存储器(SSRAM,Synchronous Static Random Access Memory)、动态随机存取存储器(DRAM,Dynamic Random Access Memory)、同步动态随机存取存储器(SDRAM,Synchronous Dynamic Random Access Memory)、双倍数据速率同步动态随机存取存储器(DDRSDRAM,Double Data Rate Synchronous Dynamic Random Access Memory)、增强型同步动态随机存取存储器(ESDRAM,Enhanced Synchronous Dynamic Random Access Memory)、同步连接动态随机存取存储器(SLDRAM,SyncLink Dynamic Random Access Memory)、直接内存总线随机存取存储器(DRRAM,Direct Rambus Random Access Memory)。本发明实施例描述的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

[0127] 在示例性实施例中,本发明实施例还提供了一种存储介质,即计算机存储介质,具体可以是计算机可读存储介质,例如包括存储计算机程序的存储器 502,上述计算机程序可由语音识别设备的处理器501执行,以完成本发明实施例方法所述的步骤。计算机可读存储介质可以是ROM、PROM、EPROM、EEPROM、Flash Memory、磁表面存储器、光盘、或CD-ROM等存储器。

[0128] 需要说明的是:“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。

[0129] 另外,本发明实施例所记载的技术方案之间,在不冲突的情况下,可以任意组合。

[0130] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

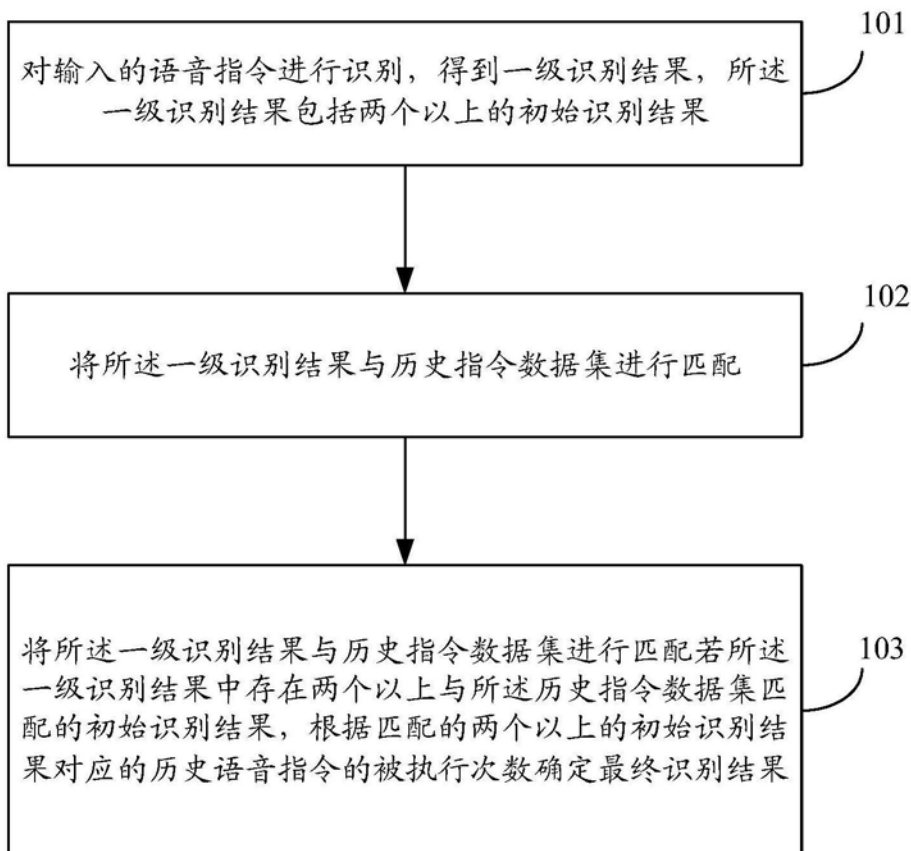


图1

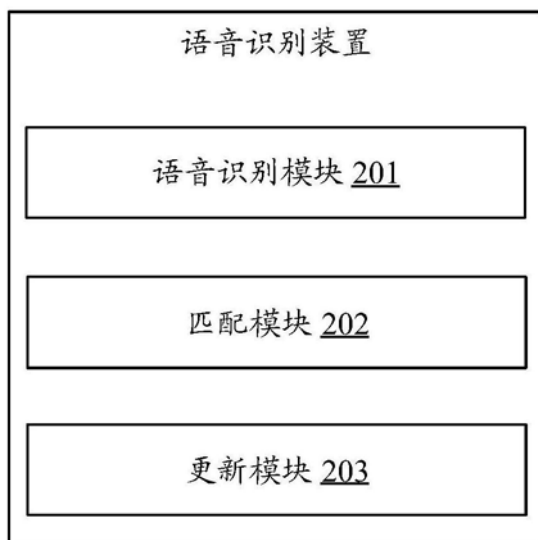


图2

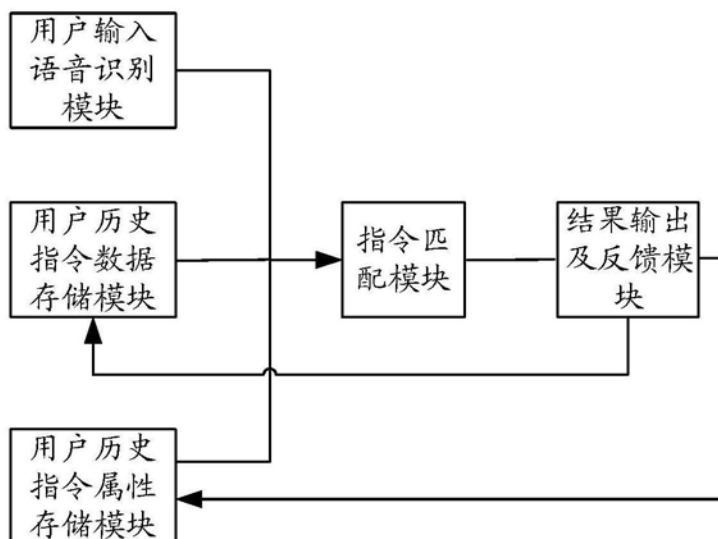


图3

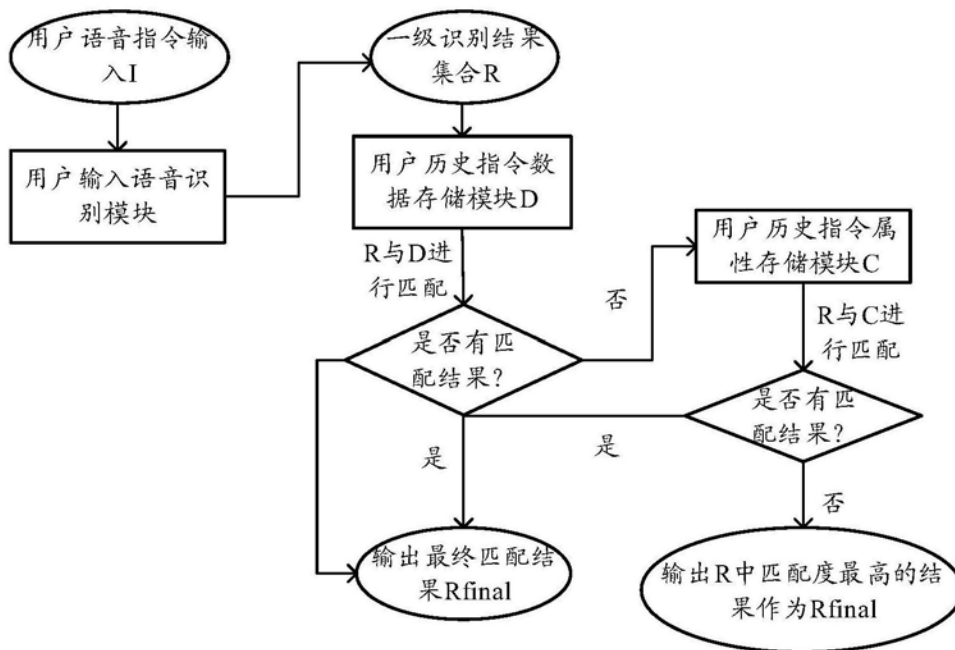


图4

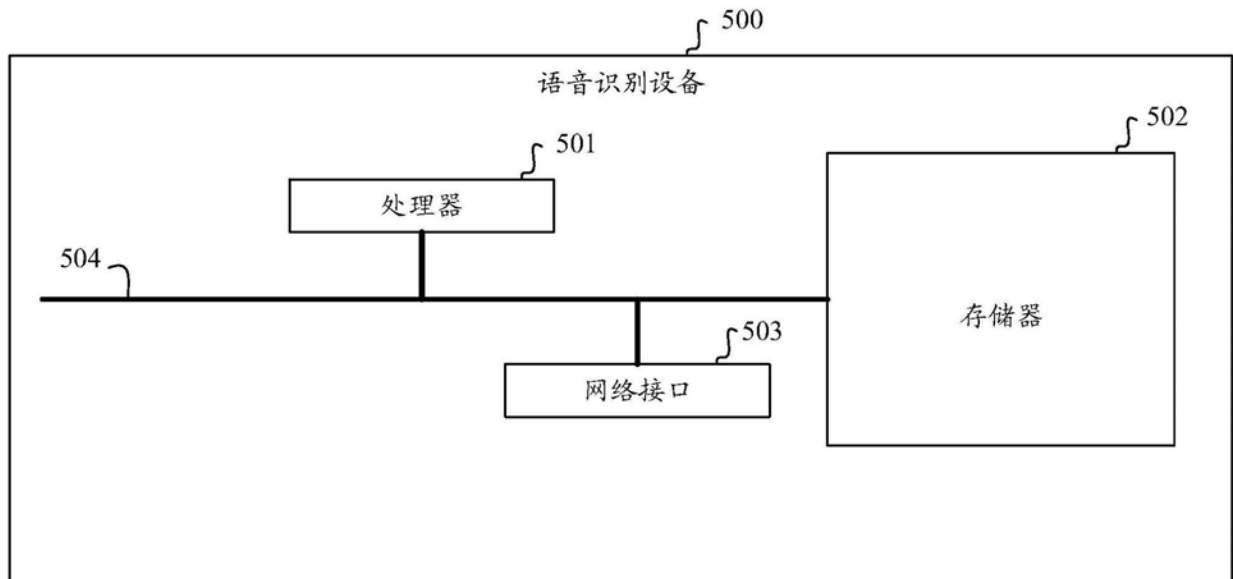


图5