



(12) 发明专利申请

(10) 申请公布号 CN 105408952 A

(43) 申请公布日 2016. 03. 16

(21) 申请号 201480013019. 8

(22) 申请日 2014. 01. 24

(30) 优先权数据

13/772, 373 2013. 02. 21 US

(85) PCT国际申请进入国家阶段日

2015. 09. 08

(86) PCT国际申请的申请数据

PCT/US2014/012922 2014. 01. 24

(87) PCT国际申请的公布数据

W02014/130205 EN 2014. 08. 28

(71) 申请人 谷歌技术控股有限责任公司

地址 美国加利福尼亚州

(72) 发明人 K·A·格雷

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 鄢迅

(51) Int. Cl.

G10L 15/187(2006. 01)

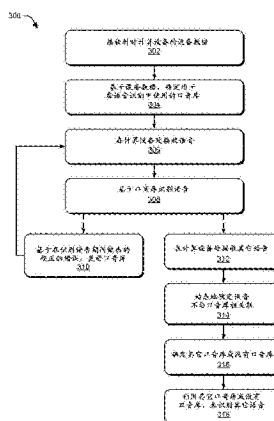
权利要求书2页 说明书8页 附图6页

(54) 发明名称

识别带口音的语音

(57) 摘要

描述了用于识别带口音的语音的技术 (300, 400, 500) 和装置 (100, 200, 700)。在一些实施例中, 口音模块使用基于设备数据的口音库来识别 (308) 带口音的语音、基于识别的词语被设置为要提供到其中的应用字段而使用不同的语音识别校正水平、或者基于对未正确识别的语音做出的校正来更新 (310) 口音库。



1. 一种方法,包括:  
接收针对计算设备的设备数据;  
基于所述设备数据,确定用于在语音识别中使用的口音库;  
在所述计算设备处接收语音;以及  
基于所述口音库,识别所述语音。
2. 根据权利要求 1 所述的方法,其中接收针对所述计算设备的设备数据包括检索或者使得检索与所述计算设备的用户相关联的个人数据。
3. 根据权利要求 2 所述的方法,其中所述个人数据包括联系人姓名,并且所述联系人姓名指示所述用户的原籍国。
4. 根据权利要求 2 所述的方法,其中所述个人数据包括消息地址,并且所述消息地址包括其中说的母语与所述计算设备被设置在的语言不同的国家中的目的地。
5. 根据权利要求 1 所述的方法,其中所述设备数据包括一个或多个安装的应用的身份、与所述计算设备被设置在的语言不同的语言相关联的键盘的类型、输入到搜索引擎中的搜索项、或者所述计算设备的类型。
6. 根据权利要求 1 所述的方法,其中所述口音库与所述计算设备被设置在的语言以及以所述语言说出的口音相关联。
7. 根据权利要求 6 所述的方法,其中以所述语言说出的所述口音与物理特性、文化或兴趣相关联。
8. 根据权利要求 1 所述的方法,其中基于所述口音库识别所述语音使用与所述计算设备被设置在的语言相关联的语言库,并且所述口音库包括对所述语言库的算法的更改。
9. 根据权利要求 1 所述的方法,进一步包括接收与所述计算设备相关联的用户的挖掘数据,所述挖掘数据包括与所述用户相关联的个人数据,并且其中确定所述口音库进一步基于所述挖掘数据。
10. 根据权利要求 1 所述的方法,进一步包括:  
在所述计算设备处接收其它语音;  
基于所述其它语音,动态地确定所述其它语音不与所述口音库相关联;  
确定针对所述其它语音的另一口音库或没有口音库;以及  
利用所述另一口音库或没有口音库,来识别所述其它语音。
11. 根据权利要求 10 所述的方法,其中动态地确定所述其它语音不与所述口音库相关联包括确定所述其它语音不是由与所述计算设备相关联的用户做出的。
12. 根据权利要求 1 所述的方法,进一步包括基于从所述语音未正确识别的、由用户校正的词语,更新所述口音库。
13. 一个或多个计算机可读存储介质,在其上具有处理器可执行指令,响应于由处理器的执行,所述指令执行操作,所述操作包括:  
接收语音;  
基于识别的文本被设置为要提供到的应用字段,确定语音识别校正水平;  
在所述语音识别校正水平下识别所述语音,以产生所述识别的文本;以及  
向所述应用字段提供所述识别的文本。
14. 根据权利要求 13 所述的计算机可读存储介质,其中所述应用字段是地址字段,并

且所述语音识别校正水平提供最大校正水平。

15. 根据权利要求 14 所述的计算机可读存储介质,其中所述最大校正水平包括使用口音库,并且进一步包括基于以下项来确定所述口音库:

接收的所述语音;

与通过其接收所述语音的计算设备相关联的设备数据;或者

与所述计算设备的用户相关联的个人数据。

16. 根据权利要求 13 所述的计算机可读存储介质,其中所述应用字段是仅数字字段,并且所述语音识别校正水平提供最大校正水平。

17. 根据权利要求 13 所述的计算机可读存储介质,其中所述应用字段是电子邮件、博客、社交网络入口、或文字处理文档的主体,并且所述语音识别校正水平提供低于最大的校正水平。

18. 根据权利要求 17 所述的计算机可读存储介质,其中所述语音识别引擎的所述低于最大的校正水平放弃使用口音库,并且使用比最大校正水平更少的计算资源。

19. 一种方法,包括:

从远程计算设备接收对语音元素的校正,所述语音元素在所述远程计算设备处使用口音库未正确识别;

更改所述口音库以提供更新的口音库,所述更新的口音库能够更准确地识别所述语音元素;以及

向所述远程计算设备或另一远程计算设备提供所述更新的口音库,以有效地使得所述远程计算设备或所述另一远程计算设备能够使用所述更新的口音库比使用所述口音库更准确地识别所述语音元素。

20. 根据权利要求 19 所述的方法,进一步包括:

从所述远程计算设备接收所述远程计算设备的设备数据,并且其中提供所述更新的口音库基于所述另一远程计算设备的至少一些设备数据与所述远程计算设备的所述设备数据相同或相似来向所述另一远程计算设备提供所述更新的口音库。

## 识别带口音的语音

### 背景技术

[0001] 在识别带口音说出时的语音方面,当前的语音识别技术是相当差的。为了解决这一问题,一个部分解决方案跟踪由用户响应于当前技术未能正确地识别词语所做出的校正。这一部分解决方案可能对于带口音的用户来说是令人沮丧的,因为在这些当前技术改善它们的识别之前,用户往往必须校正许多未正确识别的词语,往往如此多次以至于用户完全放弃声音识别。即使对于那些花时间和经得起挫折的用户,当该用户带有口音时,当前的许多技术仍然未充分地识别用户的语音。

[0002] 用于解决这一问题的另一部分解决方案要求用户去向专用用户接口并且说出特定词语的列表。要求带口音的用户找到这一专用用户接口并且说出词语列表无法提供优越的用户体验,并且因而往往将根本不会被用户执行。进一步地,要求来自用户的这一努力无法使得当前技术能够足够好地识别口音。再进一步地,即使拥有设备的用户向这方面努力,也不太可能由借用所有者的设备的另一用户来执行,诸如当设备的所有者在开车并且乘客使用所有者的设备时。

### 附图说明

[0003] 参照附图描述用于识别带口音的语音的技术和装置。贯穿附图,相同的数字用于引用同样的特征和部件:

[0004] 图 1 图示其中可以实现用于识别带口音的语音的技术的示例环境。

[0005] 图 2 图示图 1 的示例语言和口音库。

[0006] 图 3 图示用于使用基于设备数据确定的口音库来识别带口音的语音的示例方法。

[0007] 图 4 图示用于更改口音库以更准确地识别带口音的语音的示例方法。

[0008] 图 5 图示用于在基于应用字段的语音识别水平下识别语音的示例方法,其可以使用口音库。

[0009] 图 6 图示具有应用字段的示例应用。

[0010] 图 7 图示其可以实现用于识别带口音的语音的技术的示例装置的各种部件。

### 具体实施方式

[0011] 在识别带口音说出时的语音方面,用于识别带口音的语音的当前技术往往是相当差的。本公开描述用于使用口音库来识别带口音的语音的技术和装置,并且在一些实施例中,使用基于应用字段的不同语音识别校正水平,识别的词语被设置为提供到该应用字段中。

[0012] 以下讨论首先描述操作环境,其后是可以在这一环境中采用的技术、具有应用字段的示例应用、并且继续进行到示例装置。

[0013] 图 1 图示其中可以实现用于识别带口音的语音的技术的示例环境 100。示例环境 100 包括计算设备 102,计算设备 102 具有一个或多个处理器 104、计算机可读存储介质(存储介质)106、显示器 108 和输入机制 110。

[0014] 计算设备 102 被示出为具有集成麦克风 112 作为输入机制 110 的一个示例的智能电话。然而,可以使用各种类型的计算设备和输入机制,诸如具有分立独立麦克风的个人计算机、连接到具有麦克风的微微网(例如,Bluetooth™)耳机的蜂窝电话、或者具有集成立体声麦克风的平板和膝上型计算机(仅举几例)。

[0015] 计算机可读存储介质 106 包括口音模块 114、设备数据 116、挖掘数据 118 和应用 120。口音模块 114 包括语言库 122 和一个或多个口音库 124。口音模块 114 可以在有语音识别引擎(未示出)的情况下操作、在无语音识别引擎的情况下操作、包括语音识别引擎、与语音识别引擎集成、和/或补充语音识别引擎。口音模块 114 能够识别带口音的语音,诸如通过基于设备数据 116 确定与语言库 122 结合用于识别语音的口音库 124 中的口音库。

[0016] 语言库 122 与语言或其方言相关联,诸如澳大利亚英语、美国(US)英语、英国(皇家)英语等。语言库 122 和已知的语音识别引擎可以操作为执行已知的语音识别,虽然不要求使用任一者或两者。因而,在一些实施例中,口音模块 114 使用口音库 124 之一来补充使用已知类型的语言库 122 的已知语音识别引擎。

[0017] 通过示例的方式,考虑图 2,其图示图 1 的示例语言库 122 和口音库 124。这里示出两个示例语言库:澳大利亚英语 204 和 US 英语 206。与这些语言库 204 和 206 中的每个相关联的分别是众多口音库 208 和 210。

[0018] 口音库 208 包括八个示例(虽然本技术设想更多),包括澳大利亚(AU)英语-国语 208-1、AU 英语-新南(N.S.)威尔士 208-2、AU 英语-新西兰(NZ)奥克兰 208-3、AU 英语-NA 克赖斯特彻奇 208-4、AU 英语-水肺-潜水员 208-5、AU 英语-内陆 208-6、AU 英语-佩斯 208-7 和 AU 英语-印度尼西亚 208-8。如从名字清楚的,这些口音库中的每个口音库与大的语言组(澳大利亚英语)和该语言组内存在的口音相关联,无论其是最近说国语的移民还是参与水肺潜水的人。

[0019] 相似地,口音库 210 包括八个示例:US 英语-国语 210-1、US 英语-粤语 210-2、US 英语-波士顿 210-3、US 英语-冲浪者 210-4、US 英语-听力障碍 210-5、US 英语-农村 210-6、US 英语-南部 210-7 和 US 英语-阿拉斯加 210-8。注意,国语口音库 208-1 和 210-1 可以不同,因为每个与不同语言库相关联。然而,无论以澳大利亚方言或 US 方言说出英语,由于说国语者的共同品质,在口音库之间可能存在一些共同元素。注意,这些口音库在数目和处理的口音方面几乎是无限的。地区口音,小的移民群体或大的移民群体、兴趣和亚文化、以及甚至共同的物理特性所共有的口音,诸如听力障碍的人在口音方面具有一些共性。

[0020] 在图 2 的示例中,口音库 124 中的每个包含用于由语言库 122 使用的补充信息和算法。这里语言库 122 用于大语言组(例如,对于更大数目的人,其具有更多的、平均值或中值),其由口音库 124 中的一个或多个来补充。尽管图 2 的这一示例使口音库与语言库相关联,但是口音模块 114 可以放弃使用语言库或已知的语音识别引擎。代之,口音模块 114 可以提供其自己的算法和引擎,而不使用其它引擎或库,代之依靠口音库 124 而非语言库 122,但是包括对于识别大量人的语音有用的算法或信息。

[0021] 口音模块 114 可以基于设备数据 116 和/或挖掘数据 118(图 1 的两者)来确定口音库 124 中的哪个要用于识别带口音的语音。设备数据 116 可以包括设备个人数据 126 以及特定于计算设备 102 的数据。特定于计算设备 102 的数据可以包括计算设备 102 的制

造或购买日期（例如，最近发布的移动电话或平板）以及诸如制造商、硬件能力等之类的关于计算设备 102 的信息。

[0022] 设备个人数据 126 包括基于用户与计算设备 102 的交互所创建或确定的数据，诸如联系人的姓名、安装的应用、消息的接收国家或地区、用户的姓名、联系信息、非标准键盘（例如，针对除了计算设备 102 针对其设置的语言之外的特定语言）、以及上下文应用信息（例如，搜索项）。因而，联系人的姓名可以指示用户的原籍国，或者非标准类型的键盘可以指示除了针对计算设备设置的语言之外的语言是用户的母语。进一步地，消息的接收国家或地区可以包括其中针对计算设备设置的语言不是说得最多的语言的国家中的地址，例如，在设置澳大利亚英语情况下（诸如，图 2 所示的在澳大利亚英语 204 和 AU 英语 - 印度尼西亚 208-8 的情况下）为印度尼西亚的接收国家。

[0023] 更详细地，用户的联系方式中的电子邮件或地址可以指示用户的国籍或种族本源（例如，斯拉夫人的名或姓）。地址可以指示用户的出生位置或当前位置、以及关于用户的可以用于确定针对用户的口音库 124 的其它细节。电子邮件地址行中的名称或者那些电子邮件中的正文可以指示用户的朋友的国籍、出身、亚文化、或者用户的业务、或者用户的兴趣。如下面进一步指出的，这些兴趣可以指示口音，诸如用户在冲浪、水肺潜水或烹饪方面的兴趣。一些词语以及如何说出这些词语可以依赖于这些兴趣、并且因而亚文化。

[0024] 例如，参与水肺潜水的人可以使用术语“换气器”和“气压伤”，如果没有与水肺潜水相关联的口音库，这些术语可能不能正确识别。相似地，参与冲浪的人可以使用术语“高飞脚 (goofy foot)”、“回冲浪峰”或者“一道浪整个同时踏掉 (closed out)”，它们也可能未从用户的语音正确识别。最后，对于烹调爱好者，“La Creuset”、“烧烤”和“炖”可能在无当前技术的情况下未正确识别。

[0025] 设备个人数据 126 还可以包括在确定口音且因而口音库中有用的其它信息，诸如用户的电子书库中的斯拉夫语言的书、斯拉夫语言的新闻文章、关于波兰的文章和书、保存的波兰华沙的天气频道、关于在爱沙尼亚钓鱼的信息、针对手风琴音乐的 web 搜索条目、在用户的音乐库中的波尔卡音乐等。

[0026] 挖掘数据 118 也可以或代之由口音模块 114 用于确定口音库 124 中的哪个要用于识别语音。挖掘数据 118 包括挖掘个人数据 128，其可以包括可以通过因特网或其他方式找到的关于计算设备 102 的用户的任何个人数据。因而，挖掘个人数据 128 可以包括用户的搜索项、所购物、位置、人口统计特征、收入等。

[0027] 如指出的，计算机可读存储介质 106 还包括应用 120，诸如全部图 1 的电子邮件应用 130、社交网络应用 132 或电子表格应用 134。应用 120 中的每个应用包括一个或多个应用字段 136，其在一些实施例中用于确定语音识别校正水平。通过示例的方式，考虑电子表格应用 134。这里仅数字单元格 138 和一般文本单元格 140 均是应用字段 136 的示例。仅数字单元格 138 可以要求比一般文本单元格 140 更精确的文本、并且因而不同的语音识别校正水平。

[0028] 图 3 图示用于使用基于设备数据确定的口音库来识别带口音的语音的示例方法 300。描述这些和其它方法的块的顺序并不旨在被解释为限制，并且任何数目或组合的本文中这些和其它方法中描述的块可以以任何顺序组合，以实现方法或替代方法。

[0029] 在块 302 处，接收针对计算设备的设备数据。设备数据可以响应于在块 302 处执

行的主动检索而接收。因而,使用图 1 的环境 100 作为示例,口音模块 114 可以在块 302 处检索设备数据 116,诸如通过搜索计算设备 102 上的联系人数据以及关于计算设备 102 的技术细节。

[0030] 如上面部分指出的,设备数据 116 可以包括设备个人数据 126 以及与计算设备 102 相关联的其它非个人数据。通过一个进行中的示例的方式,假设设备数据 116 指示计算设备 102 是仅仅 30 天前发布的、具有显著计算能力的智能电话。这可以部分地用于基于人口统计特征指示至少当最近发布时这一智能电话的用户是早期采用者、精通技术、并且年龄在 18 和 32 之间,确定适当的口音库 124。

[0031] 假设设备个人数据 126 包括联系人姓名和地址,从而指示亚洲人的姓和亚洲人的名的统计相关数量。这一统计相关性可以以各种方式来确定,诸如通过与使用相同语言库 122 的典型的人的联系人列表进行比较。因而,尽管针对美国 (US) 英语语言库用户的联系人列表的亚洲人的名的平均数目可以是 1.3% 并且亚洲人的姓可以是 11%, 这里假设这一用户的联系人列表具有 14% 的亚洲人的名和 29% 的亚洲人的姓。基于它是从平均值的一个或多个标准偏差,统计分析考虑这一统计相关。这指示如下可能性,即用户可能不是以英语为母语者,或者用户的家庭成员很可能不是以英语为母语者,特别是亚洲人的名的统计相关数量,因为比起亚洲人的姓,亚洲人的名更可能指示第一代移民。

[0032] 除了来自用户的联系人列表的这一信息,假设设备个人数据 126 指示用户的姓名是“Molly Chin”、到海滩旅行的大量次数和持续时间、购买冲浪用具、以及用户住在南加州。

[0033] 在块 304 处,基于接收的设备数据,确定口音库。这一口音库被确定用于在语音识别中使用。继续进行中的实施例,假设口音模块 114 使设备数据 116 与这一类型的设备数据相关联的已知口音关联,由此确定两个不同的口音库 124 是可能的,两者为图 2 的 US 英语 - 国语 210-1 和 US 英语 - 冲浪者 210-4。假设基于针对用户推断的年轻年龄 (如早期采用者等)、到海滩旅行、英国人的名 (Molly)、基于冲浪者的购买等,冲浪者口音库被确定为更加可能的。在这一进行中的示例中,口音模块 114 基于设备数据 116 来确定口音库 124, 虽然口音模块 114 还可以或代之使这一确定基于挖掘数据 118 以及关于先前由计算设备 102 接收的语音的信息。

[0034] 在块 306 处,在计算设备处接收语音。语音可以以各种方式接收,诸如上面描述的输入机制 110。继续进行中的示例,假设用户讲出以下内容以进入向朋友的文本消息“Jean, is it closed out?”

[0035] 在块 308 处,基于口音库识别语音。结束进行中的示例,连同语音识别引擎,口音模块 114 使用基于设备数据选择的语言库 US 英语 206 和口音库 US 英语 - 冲浪者 210-4, 如上面指出的。这里假设如果没有口音库,语音识别引擎将会将 Molly 的语音“Jean, is it closed out?”识别为“Jean, is it close now?”然而,由于口音库 US 英语 - 冲浪者 210-4, 口音模块 114 动作以将 Molly 的语音正确识别为“Jean, is it close now?”然后口音模块 114 将这一文本传递到文本字段。

[0036] 在这一示例中,这一识别是由于口音模块 114 能够针对如何识别 Molly 的语音在多个选项之间进行选择,包括如下选项,如果没有口音库,相对于“close now”、“hosed out”和“closet”的其它可能选项,将被认为是用于当前语音识别引擎的低概率选项。这里

口音库 US 英语 - 冲浪者 210-4 添加词语、改变词语和短语的概率、并且更改算法,以改变如何解释某些声音(例如,冲浪者具有不同的语音模式,这是口音的一部分,不只是使用的词语)。

[0037] 替代地或附加地,方法 300 继续进行到块 310 和 / 或块 312-318。在块 310 处,基于在识别语音期间做出的校正的错误,更新口音库。块 310 可以与如下面描述的方法 400 结合或分立地工作。在上面的示例方法 300 中,校正识别了 Molly 的语音。如果不正确,通过用户 (Molly Chin) 的校正可以被记录并且用于更新口音库。

[0038] 在块 312 处,在计算设备接收其它语音,接收的该其它语音来自与在 302 处接收的语音不同的说话者。通过示例的方式,假设 Molly 将她的智能电话递给了她父亲,因为她在开车。假设 Molly 要她父亲请求好的泰国餐厅。还假设她父亲是以国语为母语者,并且英语是他的第二语言。进一步地,假设像许多以国语为母语者那样,Molly 的父亲使用音调来区分词语,而说英语者使用语调(句子中的音高模式)。进一步地,假设 Molly 的父亲像许多说国语者那样具有在音节的结尾发“l”音的问题。因而,Molly 的父亲将“why”发音为“wiw”、将“fly”发音为“flew”、并且将“pie”发音为“piw”。因而,当 Molly 的父亲通过讲出“Find Thai Restaurant”来要智能电话找到泰国餐厅时,但是由于他的口音,对于以 US 英语为母语者(或者仅使用 US 英语库的语音识别引擎)来说,其听起来像“Find Tew Restaurant”。

[0039] 在块 314 处,另一语音被动态确定为不与块 304 处确定的口音库相关联。口音模块 114 在接收语音“Find Tew Restaurant”时实时地确定说话者不是 Molly 并且因而口音库 US 英语 - 冲浪者 210-4 不适用。口音模块 114 可以基于“Tew”或其它指示符来确定这点,诸如词语“Restaurant”内的音调变化,其是说国语者和说粤语者两者所共有的,或者简单地从 Molly 接收的语音历史指示不是 Molly。这可以以众多方式来执行,诸如 Molly 具有通常高音高声音并且 Molly 的父亲不具有这一高音高、Molly 和 Molly 的父亲之间的说话速度差异等。

[0040] 在块 316 处,另一口音库或者没有口音库被确定用于该另一语音。继续这一示例,假设口音模块 114 基于词语“Restaurant”内的音调变化确定 Molly 的父亲是以国语为母语者或者以粤语为母语者。进一步地,假设口音模块 114 确定 Molly 的个人数据指示她有与其中国语是主导语言的中国地区(例如,北京)(而不是与粤语相关联的地区(例如,香港))更加密切相关联的朋友和地址。如上面指出的,这一信息可能已经在块 304 处确定。

[0041] 在块 318 处,在另一口音库或没有口音库(如上面确定的)的情况下识别另一语音。结束进行中的示例,通过使用图 2 的口音库 US 英语 - 国语 210-1,口音模块 114 将 Molly 的父亲的语音“Find Tew Restaurant”识别为“Find Thai Restaurant”,而不是将这一语音不正确地识别为“Find Two Restaurants”。

[0042] 图 4 图示用于更改口音库以更准确地识别带口音的语音的示例方法 400。

[0043] 在块 402 处,接收对语音元素的校正。这一校正使用口音库未正确识别的语音元素。校正可以从远程计算设备接收,虽然这不是要求的。如块 310 中指出的,使用口音库的语音识别可能是不正确的,并且然后由用户进行校正。与口音库相关联的一个或多个校正可以诸如从数以千计的远程计算设备(例如,智能电话、膝上型计算机、平板计算机、台式计算机等)接收。计算设备可以是图 1 的计算设备 102,但是在这一实施例中是与计算



设备 102 远程的服务器计算机,并且在该处校正被记录并且口音库 124 被更新以改善识别。

[0044] 在块 404 处,口音库被更改以提供更新的口音库,更新的口音库能够更准确地识别语音元素。为了说明使用上面的示例之一,假设口音库 US 英语 - 国语 210-1 将 Molly 的父亲的语音不正确地识别为“Find The Restaurant”而非“Find Thai Restaurant”。还假设 Molly 的父亲将不正确的识别校正为“Thai”。这一校正以及相同口音库的像它那样的许多其它校正可以被发送到更新实体并且由更新实体接收。更新实体可以是计算设备 102 上的口音模块 114,或者服务器计算机上的另一口音模块或其它实体。

[0045] 在块 406 处,更新的口音库被提供到一个或多个远程计算设备,以有效使得一个或多个远程计算设备能够更准确地识别语音元素。因而,使用更新的口音库,语音元素“Tew”将更有可能被正确地识别为“Thai”而非“The”。

[0046] 此外,还可以从一个或多个远程计算设备接收设备数据,该设备数据与远程计算设备的用户相关联,并且基于该设备数据来确定口音库以用于来自用户的语音的语音识别。因而,可以提供关于 Molly 的用于对口音库 US 英语 - 冲浪者 210-4 进行校正的信息或者关于 Molly 的父亲的用于对口音库 US 英语 - 国语 210-1 进行校正的信息。

[0047] 然后可以为某些设备数据或其它数据定制对适当的口音库的更新。实际上,随着时间的推移这可以动作以提供口音库的子类别。因而,诸如与 Molly Chin 具有相似性的人之类的说话者可以基于她在年龄 (18-30) 和地区 (南加州) 方面的相似性,接收 US 英语 - 冲浪者 210-4 的更新,而使用 US 英语 - 冲浪者 210-4 的另一说话者将不会接收,诸如生活在不同地区 (佛罗里达州迈阿密) 的男性 (年龄 45-60)。通过这样做,可以基于用户或他们的计算设备是否具有一个或多个与从其接收校正的远程计算设备的设备或挖掘数据相同的设备或挖掘数据的元素,向用户提供更新。

[0048] 图 5 图示用于在基于应用字段的语音识别水平下识别语音的示例方法 500,其可以使用口音库。

[0049] 在块 502 处,在计算设备接收语音。这可以如上面各种示例中阐述的那样。

[0050] 在块 504 处,基于识别的文本被设置为要提供到的应用字段,确定语音识别校正水平。这个的一个示例可以是图 1 的示例应用字段 136,即电子表格应用 134 的仅数字单元格 138 和一般文本单元格 140。如上面指出的,口音模块 114 可以基于应用字段来确定语音识别校正水平,诸如它可能需要高度准确的语音识别或者较不准确和 / 或更快的识别。

[0051] 通过示例的方式,考虑图 6,其图示具有应用字段 604 和 606 的示例电子邮件应用的用户界面 602。应用字段 604 是地址字段,并且应用字段 606 是主体字段。例如假设来自上面示例的 Molly Chin 讲出“Surf Girl Seven Seven Seven At Gee Mail Dot Com”。

[0052] 当打开要发送给朋友的新电子邮件时,假设电子邮件应用将首先将识别的文本接收到应用字段 604 处示出的电子邮件地址字段中。当说话时,并且在电子邮件地址完成之后,假设电子邮件应用将会将识别的文本接收到应用字段 606 处的电子邮件的主体中。在这一示例中,口音模块 114 确定最大校正水平应该用于地址字段。在这种情况下,口音模块 114 使用适当的口音库 124 或者做出改善准确性的其它细化。然而,改善准确性可能以识别文本的时间和计算资源 (处理器和电池) 方面为代价 (仅举几例)。因此,更高的语音校正水平可能并不总是适当的。

[0053] 例如,还要注意,通过确定使用零个、一个、或多个口音库 114 (诸如国语和冲浪者

口音库两者),口音模块 114 可以应用不同的校正水平。进一步地,口音模块 114 可以确定不使用或缺乏使用口音库 124 的校正水平。例如,口音模块 114 可以使用不同的语言库 122 以用于一些应用字段,或者使用指向说出的数字而不是正常语音中的口音的口音库 124。因而,语言库 122 之一可以指向识别其是数字或针对地址的语音,并且另一语言库 122 指向识别其是对话的语音。在本文中阐述的这些和其它方式中,技术可以动作以改善语音识别。

[0054] 在块 506 处,在语音识别校正水平下识别接收的语音,以产生识别的文本。因而,对于应用字段 604(电子邮件地址字段),使用指向预期语音的一个或多个口音库 124 和/或替代语言库 122,口音模块 114 在确定的语音识别水平下识别语音,这里在最大水平下。

[0055] 在块 508 处,识别的词语和其它文本被提供到应用字段。结束进行中的针对 Molly Chin 的示例,在块 508 处,口音模块 114 不是将语音“Surf Girl Seven Seven Seven At Gee Mail Dot Com”识别为词语,而是基于口音库 124 和/或语言库 122 识别为词语和文本的组合,并且还因为它是电子邮件的地址字段,“at”被识别为“@”符号。因而,语音被识别为“surfgirl777@GMail.com”。

[0056] 尽管不要求,在一些实施例,当应用字段是电子邮件、博客、社交网络入口、或文字处理文档的主体时,技术使用低于最大的语音校正水平。相反地,对于地址字段、电子表格中的仅数字字段、电话号码等,技术可以使用最大语音校正水平和/或替代语言库 122 或口音库 124。

[0057] 图 7 图示包括口音模块 114 的示例设备 700 的各种部件,口音模块 114 包括或具有对其它模块的访问,这些部件以硬件、固件和/或软件来实现,并且如参照先前的图 1 至图 6 中的任何一个所描述的。

[0058] 示例设备 700 可以以作为以下设备之一或组合的固定或移动设备来实现:媒体设备、计算设备(例如,图 1 的计算设备 102)、电视机顶盒、视频处理和/或渲染设备、器具设备(例如,封闭和密封的计算资源,诸如一些数字录像机或者全球定位卫星设备)、游戏设备、电子设备、车辆、和/或工作站。

[0059] 示例设备 700 可以与运行整个设备所需的电子电路、微处理器、存储器、输入输出(I/O)逻辑控制、通信接口和部件、其它硬件、固件和/或软件集成。示例设备 700 还可以包括耦合计算设备的各种部件以用于部件之间的数据通信的集成数据总线(未示出)。

[0060] 示例设备 700 包括各种部件,诸如输入输出(I/O)逻辑控制 702(例如,用于包括电子电路)和(多个)微处理器 704(例如微控制器或数字信号处理器)。示例设备 700 还包括存储器 706,其可以是任何类型的随机存取存储器(RAM)、低延迟非易失性存储器(例如,闪速存储器)、只读存储器(ROM)、和/或其他合适的电子数据存储。存储器 706 包括或具有对口音模块 114、语言库 122 和口音库 124 以及在一些实施例中对语音识别引擎(未示出)的访问。

[0061] 示例设备 700 还可以包括各种固件和/或软件,诸如操作系统 708,其连同其它部件可以由存储器 706 保持并且由微处理器 704 执行的计算机可执行指令。示例设备 700 还可以包括其他各种通信接口和部件、无线 LAN(WLAN)或无线 PAN(WPAN)部件、其他硬件、固件和/或软件。

[0062] 这些模块的其它示例性能和功能参照图 1 和图 2 所示的元素进行了描述。与其它模块或实体独立地或组合地,这些模块可以实现为由存储器 706 保持并且由微处理器 704

执行的计算机可执行指令,以实现本文中描述的各种实施例和 / 或特征。替代地或附加地,任何或所有这些部件可以实现为硬件、固件、固定逻辑电路、或它们的任何组合,其与 I/O 逻辑控制 702 和 / 或示例设备 700 的其它信号处理和控制电路结合实现。此外,这些部件中的一些部件可以与设备 700 分立地动作,诸如当远程(例如,基于云的)库执行口音模块 114 的服务时。

[0063] 虽然本发明已经以特定于结构特征和 / 或方法动作的语言进行了描述,但是要理解的是,在所附权利要求中限定的本发明不必要限于所描述的具体特征或动作。更确切地说,具体特征和动作被公开作为实现所要求保护的发明的示例形式。

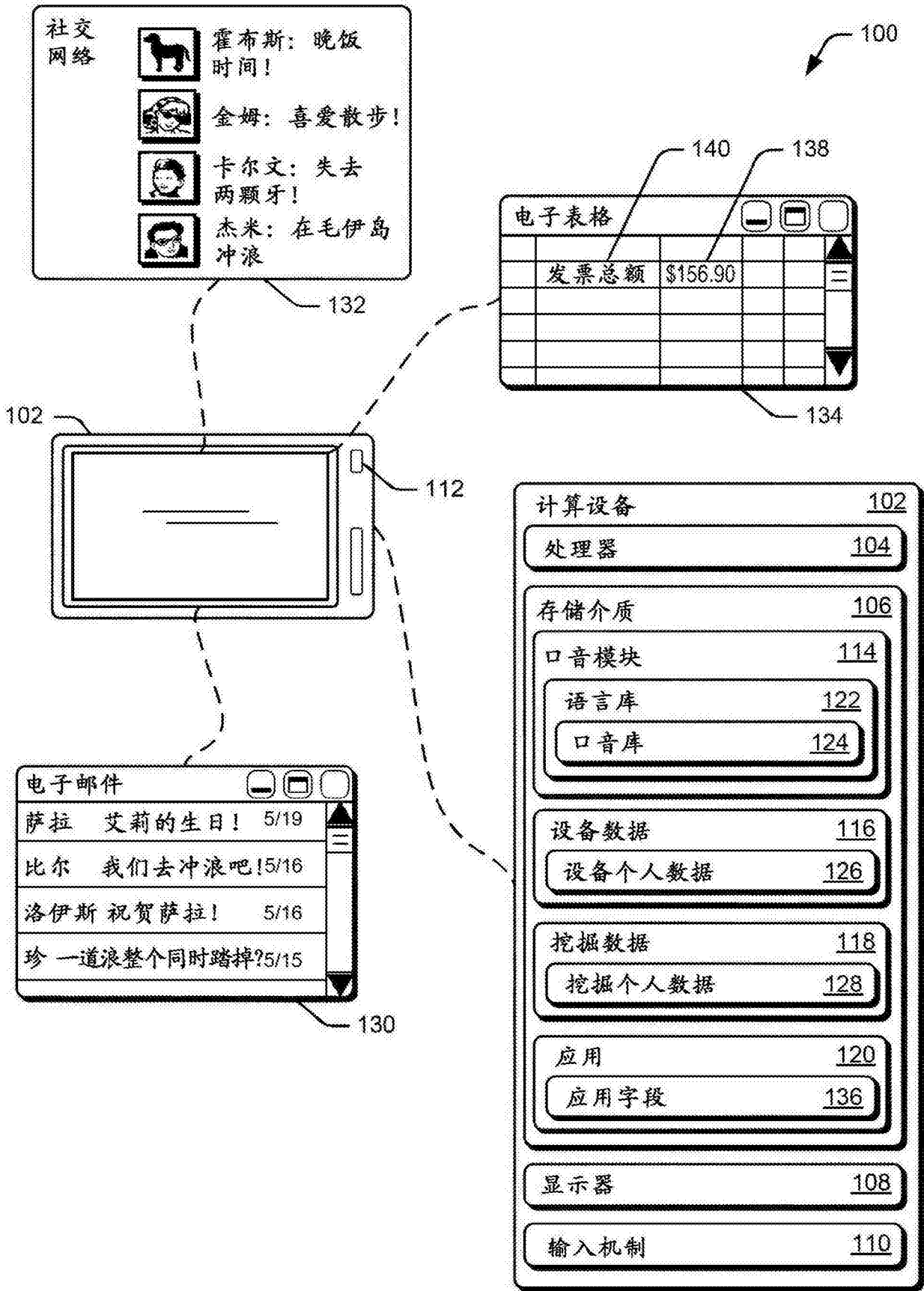


图 1

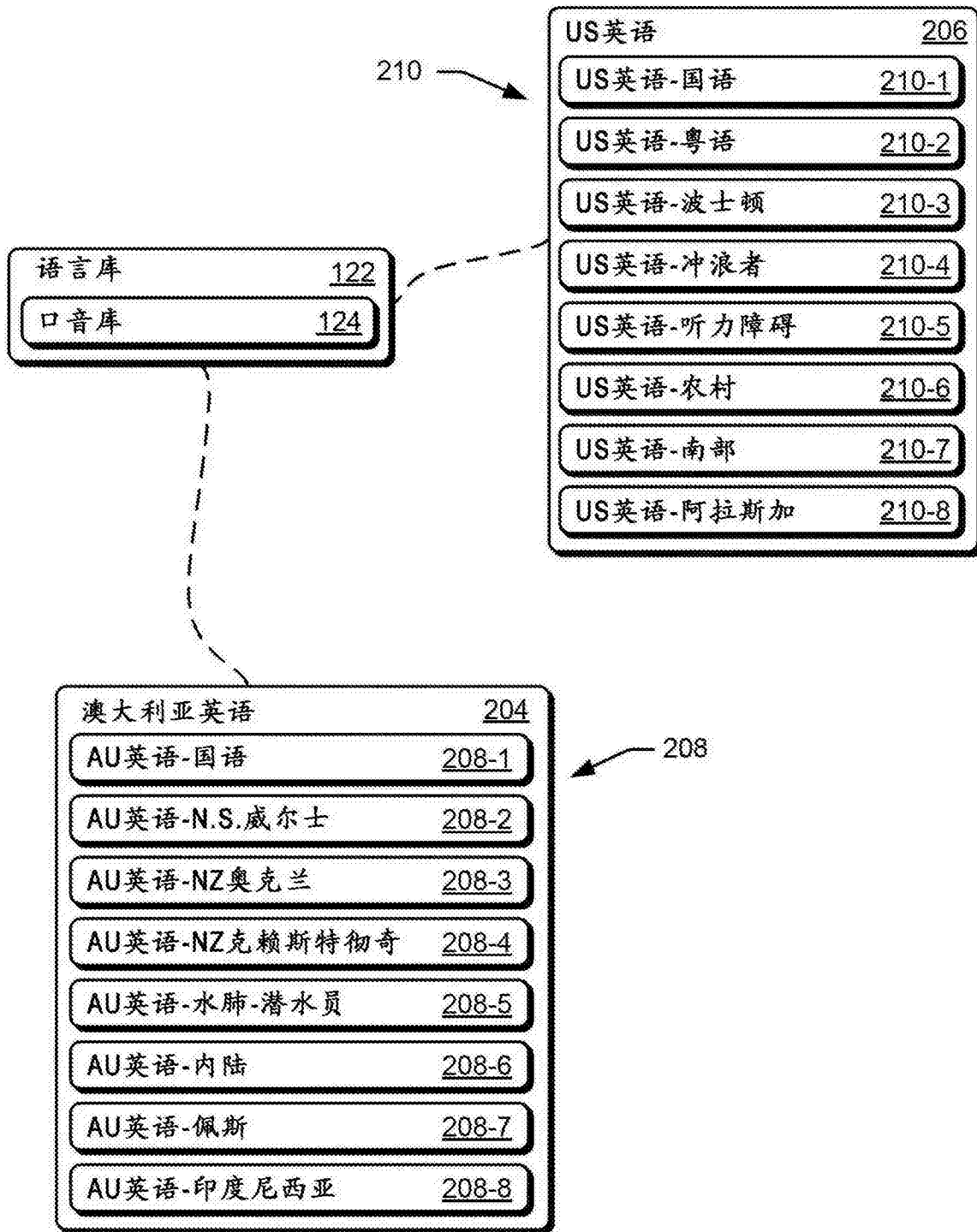


图 2

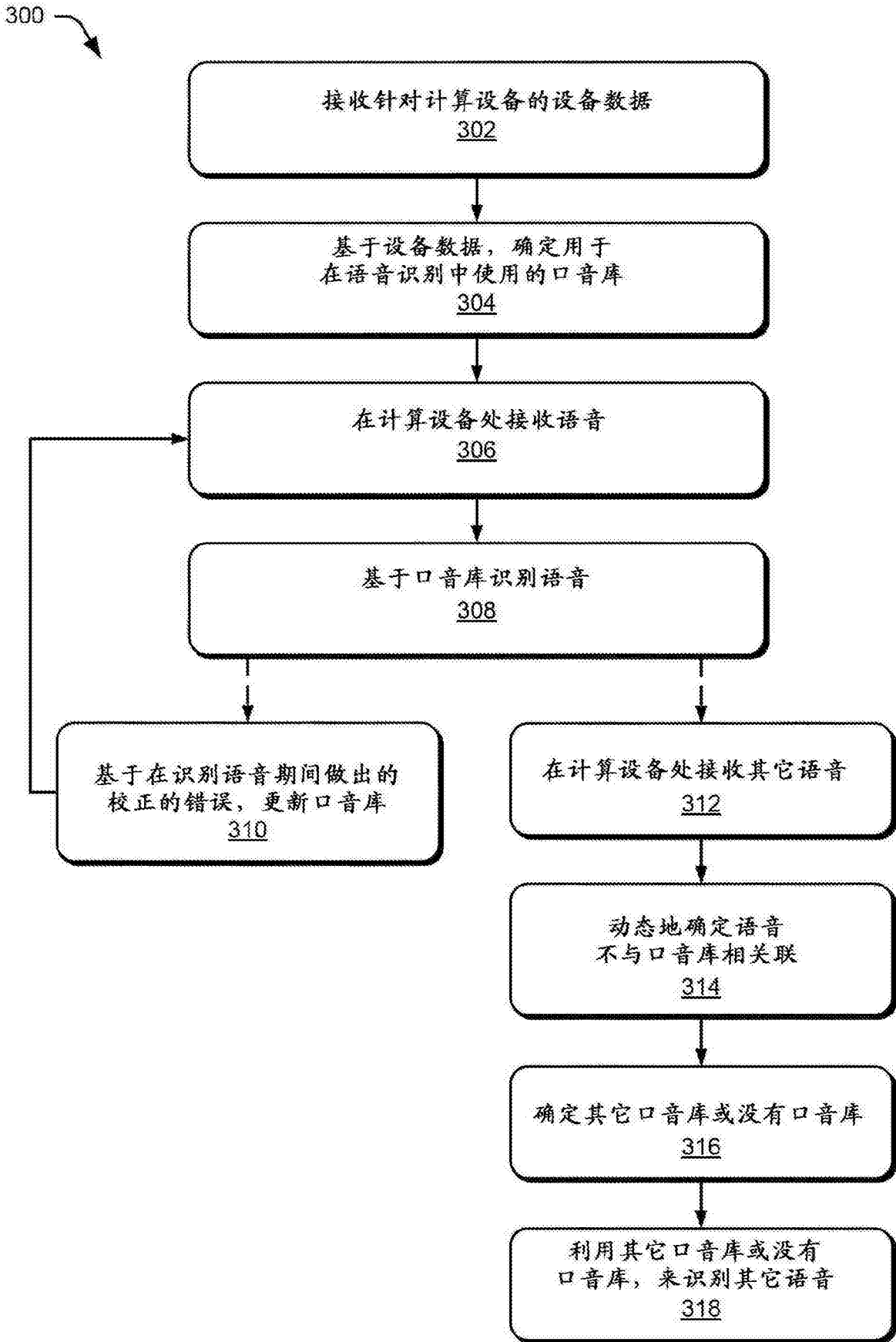


图 3

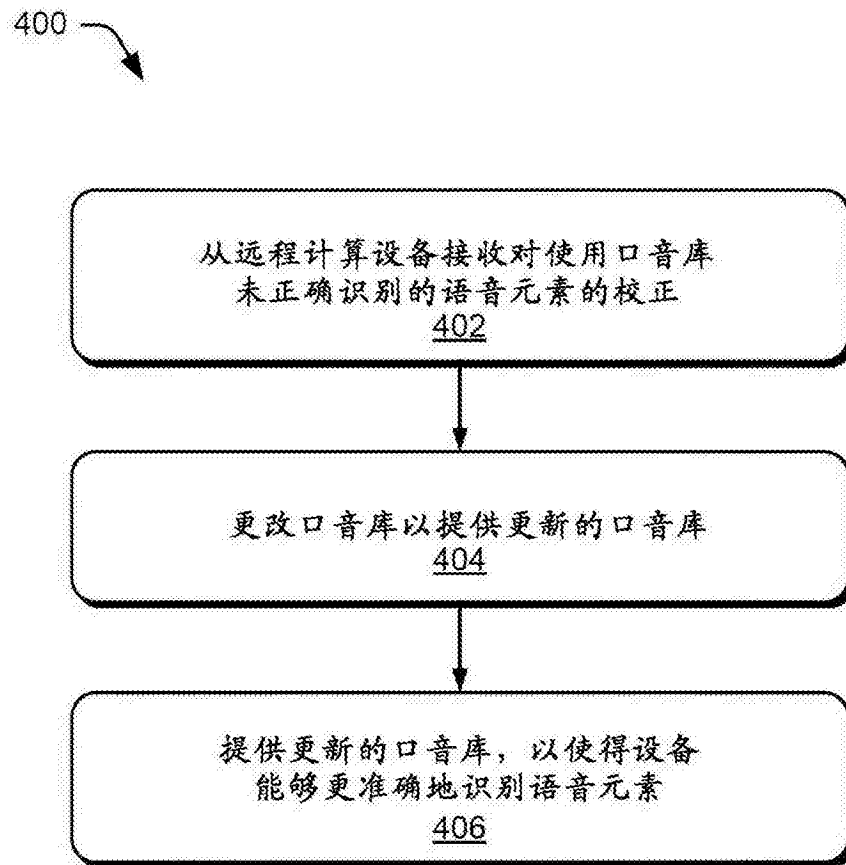


图 4

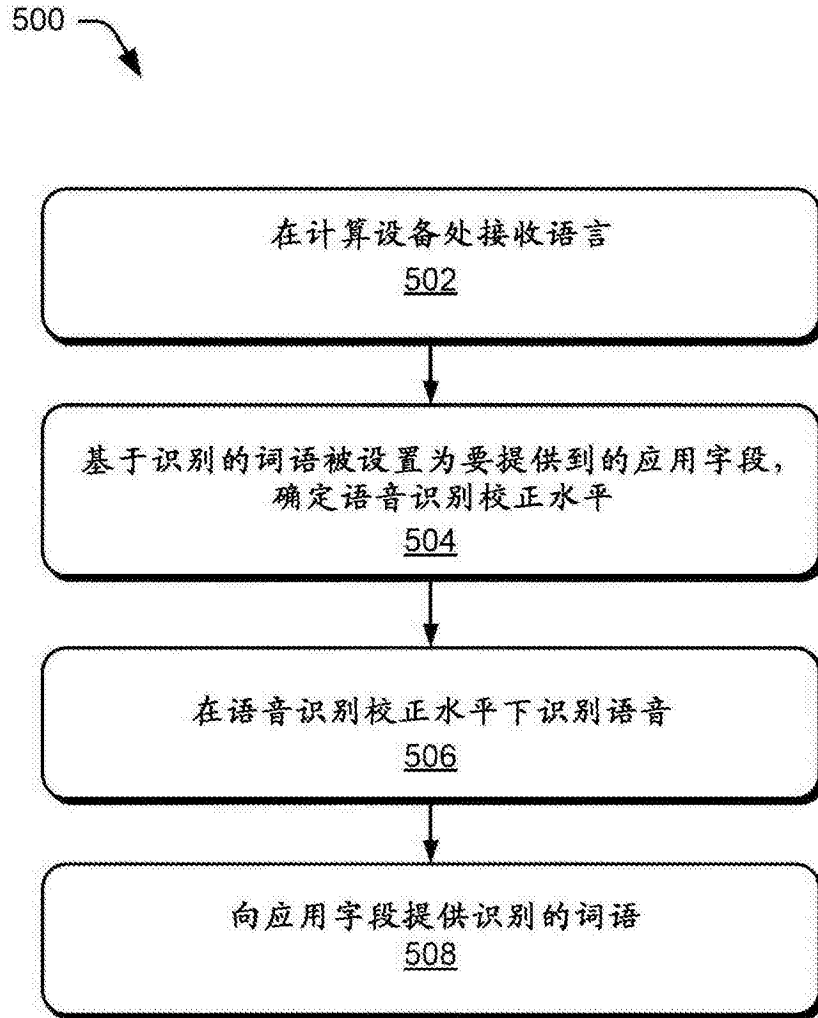


图 5

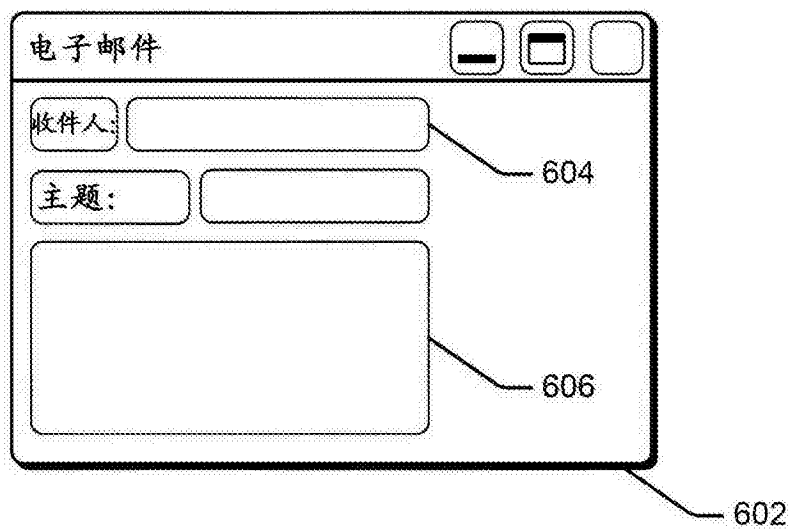


图 6



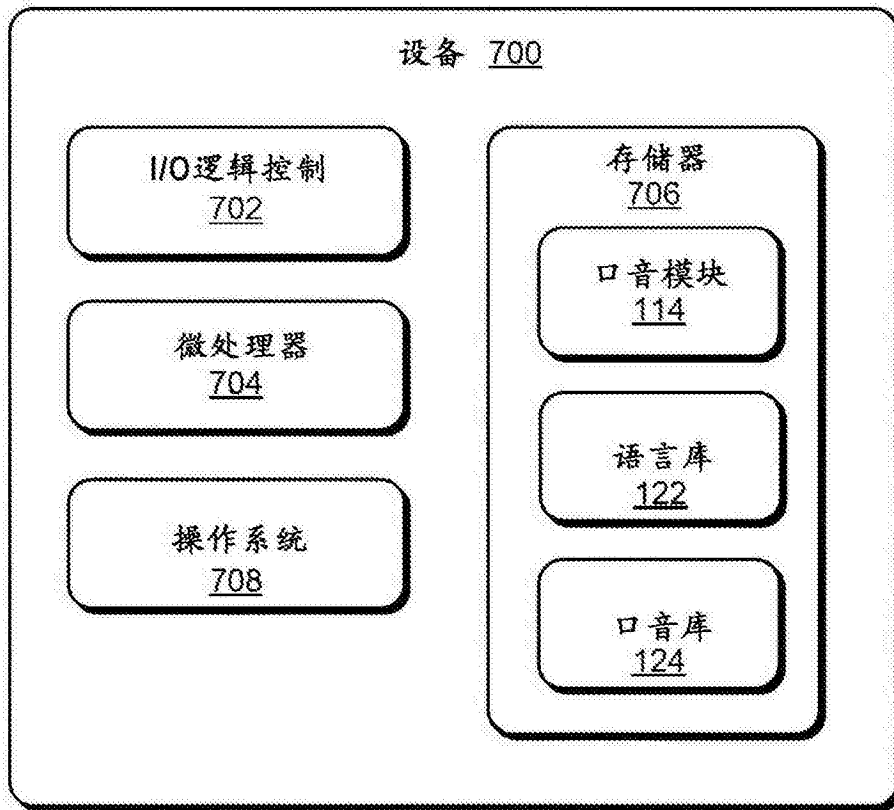


图 7