



(12) 发明专利

(10) 授权公告号 CN 103035240 B

(45) 授权公告日 2015. 11. 25

(21) 申请号 201210369739. 0

US 5909666 A, 1999. 06. 01,

(22) 申请日 2012. 09. 28

US 6311157 B1, 2001. 10. 30,

US 7315818 B2, 2008. 01. 01,

(30) 优先权数据

13/247, 912 2011. 09. 28 US

审查员 祝晔

(73) 专利权人 苹果公司

地址 美国加利福尼亚州

(72) 发明人 陈力

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华

(51) Int. Cl.

G10L 15/08(2006. 01)

(56) 对比文件

CN 101183525 A, 2008. 05. 21,

CN 1864204 A, 2006. 11. 15,

US 2007038436 A1, 2007. 02. 15,

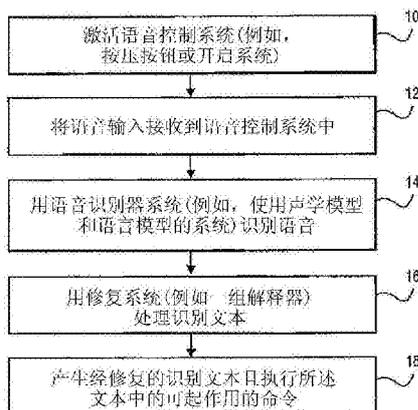
权利要求书3页 说明书19页 附图9页

(54) 发明名称

用于使用上下文信息的语音识别修复的方法和系统

(57) 摘要

本发明涉及用于使用上下文信息的语音识别修复的方法和系统。本发明提供一种语音控制系统,其可识别口述命令和相关联的字词(例如“呼叫在家里的妈妈”)且可致使选定应用程序(例如电话拨号器)执行所述命令以致使例如智能电话等数据处理系统执行基于所述命令的操作(例如查找妈妈在家里的电话号码并拨打所述号码以建立电话呼叫)。所述语音控制系统可使用一组解释器来修复来自语音识别系统的经识别文本,且来自所述组的结果可被合并为最终经修复转录,所述最终经修复转录被提供到所述选定应用程序。



1. 一种机器实施的用于语音识别修复的方法,其包括:
从数据处理系统的用户接收语音输入;
在所述数据处理系统中确定所述语音输入的上下文;
通过语音识别系统在所述语音输入中识别文本,所述文本识别产生文本输出;
将所述文本输出存储为具有多个标记的经剖析数据结构,所述多个标记各自表示所述文本输出中的字词;
用一组解释器处理所述标记中的每一者,其中每一解释器经设计以修复所述文本输出中的特定类型的错误,搜索一个或一个以上数据库以识别所述数据库中的一个或一个以上项目与所述标记中的每一者之间的匹配,且根据所述所识别的匹配和所述上下文确定所述解释器是否能修复所述文本输出中的标记;
合并由所述组解释器产生的选定结果以产生经修复的语音转录,所述经修复的语音转录表示所述文本输出的经修复版本;以及
基于所述经修复的语音转录中的命令而将所述经修复的语音转录提供到一组应用程序中的选定应用程序,其中所述选定应用程序经配置以执行所述命令。
2. 根据权利要求1所述的方法,其中所述上下文包括先前用户输入历史,且其中所述一个或一个以上数据库包括通讯录数据库,所述通讯录数据库存储姓名、地址和电话号码中的至少一者。
3. 根据权利要求1到2中任一权利要求所述的方法,其中所述上下文包括会话历史,其中所述一个或一个以上数据库包括媒体数据库,所述媒体数据库存储歌曲、题目和艺术家中的至少一者,且其中所述组解释器中的解释器在评估可能的匹配时使用至少两个字词的字符串。
4. 根据权利要求1到2中任一权利要求所述的方法,其中所述组解释器中的第一解释器使用第一算法来确定是否修复字词,且所述组解释器中的第二解释器使用第二算法来确定是否修复字词,所述第一算法不同于所述第二算法。
5. 根据权利要求4所述的方法,其中所述组解释器中的第三解释器使用第三算法来搜索所述一个或一个以上数据库,且所述组解释器中的第四解释器使用第四算法来搜索所述一个或一个以上数据库,所述第三算法不同于所述第四算法。
6. 根据权利要求1到2中任一权利要求所述的方法,其中所述组解释器中的所述解释器不试图修复所述命令。
7. 根据权利要求1到2中任一权利要求所述的方法,其中所述合并仅合并来自所述组解释器的不重叠结果,并且将来自所述组解释器的重叠结果排列在分级组中,且选择所述分级组中的一个结果并将其合并到所述经修复的语音转录中。
8. 根据权利要求1到2中任一权利要求所述的方法,其中每一解释器经设计以修复的所述特定错误类型是基于由所述解释器搜索的所述一个或一个以上数据库中的一个或一个以上字段来确定的。
9. 根据权利要求1到2中任一权利要求所述的方法,其中所述组解释器在确定是否修复所述文本输出中的一个或一个以上字词时搜索所述一个或一个以上数据库以将所述文本输出中的字词与所述一个或一个以上数据库中的一个或一个以上项目进行比较。
10. 根据权利要求1到2中任一权利要求所述的方法,其中语法剖析器根据所述文本输

出确定所述命令。

11. 根据权利要求 1 到 2 中任一权利要求所述的方法, 其中所述组应用程序包括以下各项中的至少两者: (a) 电话拨号器, 其使用所述经修复的语音转录来拨打电话号码; (b) 媒体播放器, 其用于播放歌曲或其它内容; (c) 文本消息接发应用程序; (d) 电子邮件应用程序; (e) 日历应用程序; (f) 本地搜索应用程序; (g) 视频会议应用程序; 或 (h) 人员或物体定位应用程序。

12. 一种数据处理系统, 其包括:

语音识别器, 其可操作以在语音输入中识别文本且产生文本输出;

上下文确定模块, 其可操作以确定所述语音输入的上下文;

麦克风, 其耦合到所述语音识别器以将所述语音输入提供到所述语音识别器;

存储装置, 其用于将所述文本输出存储为具有多个标记的经剖析数据结构, 所述多个标记各自表示所述文本输出中的字词;

一组解释器, 其耦合到所述语音识别器和所述上下文确定模块, 其中每一解释器经设计以修复所述文本输出中的特定类型的错误, 搜索一个或一个以上数据库以识别所述数据库中的一个或一个以上项目与所述标记中的每一者之间的匹配, 且根据所述所识别的匹配和所述上下文确定所述解释器是否能修复所述文本输出中的标记; 以及

控制器, 其用于合并由所述组解释器产生的选定结果以产生经修复的语音转录且用于基于所述经修复的语音转录中的命令来将所述经修复的语音转录提供到一组应用程序中的选定应用程序, 其中所述经修复的语音转录表示所述文本输出的经修复版本, 且所述选定应用程序经配置以执行所述命令。

13. 根据权利要求 12 所述的系统, 其中所述上下文包括先前用户输入历史, 且其中所述一个或一个以上数据库包括通讯录数据库, 所述通讯录数据库存储姓名、地址和电话号码中的至少一者。

14. 根据权利要求 12 到 13 中任一权利要求所述的系统, 其中所述上下文包括会话历史, 其中所述一个或一个以上数据库包括媒体数据库, 所述媒体数据库存储歌曲、题目和艺术家中的至少一者, 且其中所述组解释器中的解释器在评估可能的匹配时使用至少两个字词的字符串。

15. 根据权利要求 12 到 13 中任一权利要求所述的系统, 其中所述组解释器中的第一解释器使用第一算法来确定是否修复字词, 且所述组解释器中的第二解释器使用第二算法来确定是否修复字词, 所述第一算法不同于所述第二算法。

16. 根据权利要求 15 所述的系统, 其中所述组解释器中的第三解释器使用第三算法来搜索所述一个或一个以上数据库, 且所述组解释器中的第四解释器使用第四算法来搜索所述一个或一个以上数据库, 所述第三算法不同于所述第四算法。

17. 根据权利要求 12 到 13 中任一权利要求所述的系统, 其中所述组解释器中的所述解释器不试图修复所述命令。

18. 根据权利要求 12 到 13 中任一权利要求所述的系统, 其中所述合并仅合并来自所述组解释器的不重叠结果, 并且来自所述组解释器的重叠结果被排列在分级组中, 且所述分级组中的一个结果被选择且合并到所述经修复的语音转录中。

19. 根据权利要求 12 到 13 中任一权利要求所述的系统, 其中每一解释器经设计以修复

的所述特定错误类型是基于由所述解释器搜索的所述一个或一个以上数据库中的一个或一个以上字段来确定的。

20. 根据权利要求 12 到 13 中任一权利要求所述的系统,其进一步包含语法剖析器,所述语法剖析器用于根据所述文本输出确定所述命令。

用于使用上下文信息的语音识别修复的方法和系统

技术领域

[0001] 本发明涉及语音识别系统,且在一个实施例中,涉及用以控制数据处理系统的语音识别系统。

背景技术

[0002] 许多语音识别系统所具有的常见问题是准确性。用户可对着语音识别器说话,且系统可用识别文本做出响应,但所述识别文本通常可能含有许多错误,因为语音识别器未能恰当地识别人类用户的话语。

[0003] 语音识别可用以在电话上调用语音拨号,例如当用户在电话上口述命令“call mom(呼叫妈妈)”时。使用语音来控制数据处理系统可在来自语音识别器系统的转录错误决定用户口述“call Tom(呼叫汤姆)”而非“call mom(呼叫妈妈)”时导致异常系统行为。转录错误可由硬件缺点(例如不能够经由蓝牙头戴式耳机俘获高质量音频记录)或用户错误(例如不正确或不完整的发音或背景噪声)造成。一些语音识别系统可采用使用上下文来改善语音识别系统;美国专利 7,478,037 提供可采用上下文来辅助语音识别过程的语音识别系统的实例。

发明内容

[0004] 本发明的实施例提供一种语音控制系统,其可识别口述命令和相关联字词(例如“呼叫在家里的妈妈”)且可致使选定应用程序(例如电话拨号器)执行所述命令以致使系统(其可为智能电话)执行基于所述命令和相关联字词的操作(例如,向在家里的妈妈发出电话呼叫)。在一个实施例中,所述语音控制系统可使用包括常规声学模型和常规语言模型的语言识别器来根据从人类用户的语音获得的数字化输入产生文本输出。在一个实施例中,所述语音控制系统可由用户可调用的语音辅助应用程序来启动,且此语音辅助应用程序可解释并修复来自所述语音识别器的所述文本输出且将经修复的文本输出提供到一组应用程序中的选定应用程序;所述组应用程序可包括(例如)一个或一个以上应用程序,例如电话应用程序(用以拨号并建立语音电话呼叫连接)和媒体播放器应用程序(例如,iTunes)和 SMS(短消息服务)“文本消息”应用程序和视频会议(例如,“面对面时间(FaceTime)”)或聊天应用程序和用以找到或定位例如朋友等个人的应用程序和其它应用程序。

[0005] 在一个实施例中,语音辅助应用程序使用一组解释器来解释所述文本输出,所述解释器中的每一者经设计以解释所述组应用程序所使用的特定类型的文本。举例来说,名字解释器经设计以解释地址簿或通讯录数据库中的名字(在名字字段中),姓氏解释器经设计以解释地址簿或通讯录数据库中的姓氏(在姓氏字段中),全名解释器经设计以解释地址簿或通讯录数据库中的全名,且公司名称解释器经设计以解释地址簿或通讯录数据库中的公司名称。在一个实施例中,这些解释器可经配置以使用不同算法或过程来解释文本输出中的每一字词;举例来说,全名解释器可使用模拟匹配(使用编辑距离相似性测量)

算法来将文本输出中的字词与地址簿或通讯录数据库中的字词进行比较,但在一个实施例中,不使用 n 码组 (n-gram) 算法来将文本输出中的字词与地址簿或通讯录数据库中的字词进行比较,而名字解释器使用 n 码组算法来将文本输出中的字词与地址簿或通讯录数据库中的字词进行比较。此外,在一个实施例中,这些解释器可在搜索地址簿或通讯录数据库以查找匹配时使用不同搜索算法。在一个实施例中,每一解释器还可在解释字词时使用上下文(例如,所述上下文可包括媒体播放器正在播放歌曲的指示)。在一个实施例中,所述上下文可包括用户输入历史(例如会话历史(例如,先前识别语音))或所述组应用程序中的应用程序的状态等。在一个实施例中,所述组中的每一解释器可处理文本输出中的每一字词以试图确定其是否能修复所述字词,且在一个实施例中,每一解释器自行决定其是否能修复每一字词;所述解释器产生指示其是否能修复所述字词的得分或置信度。

[0006] 在一个实施例中,所述组解释器的控制器可通过排列所得的经修复解释(使用每一解释器的得分或置信度来执行所述排列)且接着合并所述经排列的解释来处理所述组解释器的结果。在一个实施例中,所述合并设法避免解释的重叠,使得仅使用来自一个解释器的输出来修复特定字词。

[0007] 在一个实施例中,所述语音辅助应用程序可基于字词在字词串中的位置来确定来自语音识别器系统的文本输出中的命令或通过语法剖析器来确定所述命令,且所述命令连同经修复的语音转录可由语音辅助应用程序传递到一组应用程序中的特定应用程序以供所述特定应用程序使用经修复的语音转录执行所述命令。在此实施例中,语音辅助应用程序可基于所述命令来选择所述特定应用程序;举例来说,识别文本中的“呼叫”命令致使语音辅助应用程序通过 API 将所述“呼叫”命令连同经修复的语音转录传递到电话拨号器或电话应用程序,而识别文本中的“停止”命令致使语音辅助应用程序通过 API 将“停止”命令传递到媒体播放器(例如 iTunes)以停止播放当前正在播放的歌曲。在此实例中,提供到所述组解释器中的媒体播放器解释器的上下文可包括媒体的状态(例如,上下文包括在语音识别器系统接收到含有识别字词“停止”的语音输入时披头士歌曲“Come Together(一起来)”当前正在播放的状态)。在此实例中,用户不需要在口述命令之前选择特定的所要应用程序;而是,用户在将语音辅助应用程序作为最前端应用程序(且具有语音输入焦点)的情况下进行口述且语音辅助应用程序接着自动地(不需要用户直接指定应用程序)基于所述命令来在所述组应用程序中选择恰当的应用程序,且接着通过 API 将所述命令传递到选定应用程序。

[0008] 在一个方面中,一种机器实施方法包括:从数据处理系统的用户接收语音输入;在所述数据处理系统中确定所述语音输入的上下文;通过语音识别系统在所述语音输入中识别文本,所述文本识别产生文本输出;将所述文本输出存储为具有多个标记的剖析数据结构,所述多个标记各自表示所述文本输出中的字词;用一组解释器处理所述标记中的每一者,其中每一解释器经设计以修复所述文本输出中的特定类型的错误,搜索一个或一个以上数据库以识别所述数据库中的一个或一个以上项目与所述标记中的每一者之间的匹配,且根据所述所识别的匹配和所述上下文确定所述解释器是否能修复所述文本输出中的标记;合并由所述组解释器产生的选定结果以产生经修复的语音转录,所述经修复的语音转录表示所述文本输出的经修复版本;以及基于所述经修复的语音转录中的命令而将所述经修复的语音转录提供到一组应用程序中的选定应用程序,其中所述选定应用程序经配置

以执行所述命令。

[0009] 在一些实施例中,所述上下文包括先前用户输入历史,且其中所述一个或一个以上数据库包括通讯录数据库,所述通讯录数据库存储姓名、地址和电话号码中的至少一者。

[0010] 在一些实施例中,所述上下文包括会话历史,其中所述一个或一个以上数据库包括媒体数据库,所述媒体数据库存储歌曲、题目和艺术家中的至少一者,且其中所述组解释器中的解释器在评估可能的匹配时使用至少两个字词的字符串。

[0011] 在一些实施例中,所述组解释器中的第一解释器使用第一算法来确定是否修复字词,且所述组解释器中的第二解释器使用第二算法来确定是否修复字词,所述第一算法不同于所述第二算法。

[0012] 在一些实施例中,所述组解释器中的第三解释器使用第三算法来搜索所述一个或一个以上数据库,且所述组解释器中的第四解释器使用第四算法来搜索所述一个或一个以上数据库,所述第三算法不同于所述第四算法。

[0013] 在一些实施例中,所述组解释器中的所述解释器不试图修复所述命令。

[0014] 在一些实施例中,所述合并仅合并来自所述组解释器的不重叠结果,并且将来自所述组解释器的重叠结果排列在分级组中,且选择所述分级组中的一个结果并将其合并到所述经修复的语音转录中。

[0015] 在一些实施例中,每一解释器经设计以修复的所述特定错误类型是基于由所述解释器搜索的所述一个或一个以上数据库中的一个或一个以上字段来确定的。

[0016] 在一些实施例中,所述组解释器在确定是否修复所述文本输出中的一个或一个以上字词时搜索所述一个或一个以上数据库以将所述文本输出中的字词与所述一个或一个以上数据库中的一个或一个以上项目进行比较。

[0017] 在一些实施例中,语法剖析器根据所述文本输出确定所述命令。

[0018] 在一些实施例中,所述组应用程序包括以下各项中的至少两者:(a) 电话拨号器,其使用所述经修复的语音转录来拨打电话号码;(b) 媒体播放器,其用于播放歌曲或其它内容;(c) 文本消息接发应用程序;(d) 电子邮件应用程序;(e) 日历应用程序;(f) 本地搜索应用程序;(g) 视频会议应用程序;或(h) 人员或物体定位应用程序。

[0019] 在一些实施例中,所述方法包括上文所陈述的特征的任何组合。

[0020] 在一个方面中,一种数据处理系统包括:语音识别器,其可操作以在语音输入中识别文本且产生文本输出;上下文确定模块,其可操作以确定所述语音输入的上下文;麦克风,其耦合到所述语音识别器以将所述语音输入提供到所述语音识别器;存储装置,其用于将所述文本输出存储为具有多个标记的剖析数据结构,所述多个标记各自表示所述文本输出中的字词;一组解释器,其耦合到所述语音识别器和所述上下文确定模块,其中每一解释器经设计以修复所述文本输出中的特定类型的错误,搜索一个或一个以上数据库以识别所述数据库中的一个或一个以上项目与所述标记中的每一者之间的匹配,且根据所述所识别的匹配和所述上下文确定所述解释器是否能修复所述文本输出中的标记;以及控制器,其用于合并由所述组解释器产生的选定结果以产生经修复的语音转录且用于基于所述经修复的语音转录中的命令来将所述经修复的语音转录提供到一组应用程序中的选定应用程序,其中所述经修复的语音转录表示所述文本输出的经修复版本,且所述选定应用程序经配置以执行所述命令。

[0021] 在一些实施例中,所述上下文包括先前用户输入历史,且所述一个或一个以上数据库包括通讯录数据库,所述通讯录数据库存储姓名、地址和电话号码中的至少一者。

[0022] 在一些实施例中,所述上下文包括会话历史,其中所述一个或一个以上数据库包括媒体数据库,所述媒体数据库存储歌曲、题目和艺术家中的至少一者,且其中所述组解释器中的解释器在评估可能的匹配时使用至少两个字词的字符串。

[0023] 在一些实施例中,所述组解释器中的第一解释器使用第一算法来确定是否修复字词,且所述组解释器中的第二解释器使用第二算法来确定是否修复字词,所述第一算法不同于所述第二算法。

[0024] 在一些实施例中,所述组解释器中的第三解释器使用第三算法来搜索所述一个或一个以上数据库,且所述组解释器中的第四解释器使用第四算法来搜索所述一个或一个以上数据库,所述第三算法不同于所述第四算法。

[0025] 在一些实施例中,所述组解释器中的所述解释器不试图修复所述命令。

[0026] 在一些实施例中,所述合并仅合并来自所述组解释器的不重叠结果,并且来自所述组解释器的重叠结果被排列在分级组中,且所述分级组中的一个结果被选择且合并到所述经修复的语音转录中。

[0027] 在一些实施例中,每一解释器经设计以修复的所述特定错误类型是基于由所述解释器搜索的所述一个或一个以上数据库中的一个或一个以上字段来确定的。

[0028] 在一些实施例中,所述系统进一步包含语法剖析器,所述语法剖析器用于根据所述文本输出确定所述命令。

[0029] 在一些实施例中,所述系统包括上文所陈述的特征的任何组合。

[0030] 本文中所描述的实施例可实施为机器可读非暂时存储媒体或方法或数据处理系统。

[0031] 以上概述并不包括本发明的所有方面的详尽列举。期望本发明包括可根据上文所概述的各种方面以及下文具体实施方案中所揭示的那些方面的所有合适组合实践的所有系统和方法。

附图说明

[0032] 在附图的图式中借助于实例而非限制来说明本发明,在附图中相同参考标号指示相似元件。

[0033] 图 1 展示说明根据本发明的一个实施例的方法的流程图。

[0034] 图 2 展示根据本发明的一个实施例的可包括软件模块和数据结构的架构的实例。

[0035] 图 3 为展示根据本发明的一个实施例的方法的流程图。

[0036] 图 4 展示根据本发明的一个实施例的架构的实例,其中控制器模块用以排列且合并来自根据本发明的一个实施例的一组解释器的经修复结果。

[0037] 图 5A 展示描绘根据本发明的一个实施例的方法的流程图。

[0038] 图 5B 展示可在本文中所描述的一个或一个以上实施例中采用的包括一个或一个以上 API 的软件架构。

[0039] 图 6 展示根据一个实施例的架构,其中在语音识别系统中确定并使用当前上下文。

[0040] 图 7 展示可在本文中所描述的一个或一个以上实施例中的修复过程中使用的数据结构实例。

[0041] 图 8 展示可在解释器确定是否修复已由语音识别系统识别的特定字词时由本文中所描述的解释器中的一者或一者以上使用的特定算法的实例。

[0042] 图 9 展示根据本发明的一个实施例的数据处理系统的实例。

[0043] 图 10 为可在本发明的一些实施例中使用的软件堆叠的实例。

[0044] 图 11 为说明可在本发明的一些实施例中使用的示范性 API 架构的框图。

具体实施方式

[0045] 将参考下文所论述的细节来描述本发明的各种实施例和方面，且附图将说明各种实施例。以下描述和附图说明本发明且不应解释为限制本发明。描述众多具体细节以提供对本发明的各种实施例的透彻理解。然而，在某些例子中，未描述众所周知或常规的细节以便提供对本发明的实施例的简明论述。

[0046] 说明书中对“一个实施例”或“一实施例”的参考意味着在本发明的至少一个实施例中可包括结合所述实施例描述的特定特征、结构或特性。在说明书的各种地方中出现短语“在一个实施例中”未必全部指代同一实施例。由包含硬件（例如，电路、专用逻辑等）、软件或两者的组合的处理逻辑来执行在以下图式中所描绘的过程。虽然下文中根据一些顺序操作来描述所述过程，但应了解，所描述的一些操作可按不同次序来执行。此外，一些操作可并行地而非顺序地来执行。

[0047] 本发明的一个实施例提供一组解释器，其每一者经设计或配置以修复由语音识别器系统提供的识别文本中的特定类型的错误。语音识别器系统可为常规的基于软件的语音识别系统，其包括声学模型和语言模型，且语音识别器系统中的这些模型的组合产生文本输出，所述文本输出接着由所述组解释器修复。解释器可经配置以与特定数据库和所述数据库中的内容以及可使用那些数据库的特定应用程序一起操作。在一个实施例中，所述组解释器与语音识别系统的分离（使得所述组解释器在语音识别系统提供输出之后进行操作）允许在设计语音控制系统方面具有较大灵活性。特定应用程序和 / 或那些数据库的任何改变可反映在恰当且对应的解释器的改变中，而不必改变基础的语音识别系统。举例来说，数据处理系统可使用现有的常规语音识别系统，且接着提供定制解释器，所述定制解释器针对特定应用程序和含有将针对数据处理系统上的每一应用程序或一组应用程序在口述命令中出现的内容的特定数据库进行定制。举例来说，例如“call John Smith on mobile（呼叫约翰·史密斯的移动电话）”等命令使用估计可能应当在用户的通讯录或地址簿数据库中出现的字词。名字“John（约翰）”应当在数据库中出现且姓氏“Smith（史密斯）”应当在数据库中出现；此外，数据库应包括指示一个电话号码是 John Smith（约翰·史密斯）的移动电话号码的字段识别符。可能需要命令“Call（呼叫）”位于口述命令的开始处，或数据处理系统可使用语法剖析器来确定来自口述命令的命令的位置。如果通讯录数据库改变或电话应用程序改变（例如，添加或删除或修改命令），那么可改变用于所述数据库和应用程序的解释器而不必修改语音识别系统（例如，不必修改语音识别系统的语言模型）。可通过（例如）改变解释器与之交互的（数据库中的）字段或改变用以将（来自语音识别系统的）文本输出中的字词与数据库中的字段进行匹配的算法或通过改变用以搜索数据库

的搜索算法来改变解释器。

[0048] 图 1 展示根据本发明的一个实施例的方法的实例,其可使用一组解释器来修复已由语音识别器系统(例如使用声学模型和语言模型的系统)提供的文本输出。所述方法可在操作 10 中开始,在操作 10 中激活语音控制系统。举例来说,在一个实施例中,用户可按压按钮或按压并固持按钮或选择或启动语音辅助应用程序或仅仅开启调用语音辅助应用程序作为始终运行的后台守护程序的数据处理系统。在已经激活语音控制系统(例如,语音辅助应用程序为最前端的且具有语音输入焦点)之后,语音控制系统接收语音输入(12)。在一个实施例中,用户可口述例如“call John Smith on mobile(呼叫约翰·史密斯的移动电话)”或“tell John Smith that I am in traffic and will be late for the meeting(告诉约翰·史密斯:我在途中且将见面迟到)”或“play all songs by the Beatles(播放披头士的所有歌曲)”或“tell my son to pick up milk if he goes to Safeway after school(让我儿子去取牛奶,如果他放学后去西夫韦的话)”等命令。接着,在操作 14 中,常规的语音识别系统或语音识别器可识别在口述输入中所接收的字词,所述口述输入已经使用可采用声学模型和语言模型两者的常规语音识别系统来数字化并处理以产生可呈单一代码(Unicode)或 ASCII 格式或编码或其它字符编码的文本输出。常规的语音控制或语音识别系统此时使用所得输出而不进行进一步处理。在本发明的至少一些实施例中,进一步处理所述输出以便确定是否修复由操作 14 所提供的来自语音识别器系统的识别文本输出中的一个或一个以上字词。举例来说,在本发明的一个实施例中,通过处理识别文本(其可呈单一代码编码形式)以确定是否可修复识别文本中的一个或一个以上字词来执行操作 16。在一个实施例中,由一组解释器执行所述修复,其中每一解释器经设计或配置以修复特定类型的错误,例如数据库的数据结构的特定字段中的错误。举例来说,一个解释器可经配置且设计以修复通讯录数据库的名字中的错误,而另一解释器可经设计以修复通讯录数据库中的公司名称中的错误。下文中通过使用不同算法(包括不同处理算法或搜索算法)来进一步描述每一解释器经配置以修复特定字段中的特定类型的错误的方式。作为操作 16 中的处理的结果,在操作 18 中提供经修复文本,且接着所述经修复文本可作为实际命令提供到特定应用程序,所述特定应用程序可为一组应用程序内的一个应用程序。

[0049] 在一个实施例中,数据处理系统可在所述组中包括两个应用程序,例如由语音输入控制的电话拨号器和由语音输入控制的媒体播放器(例如 iTunes)。在另一实施例中,所述组应用程序可包括那些应用程序以及文本消息接发(SMS-短消息接发服务)应用程序,和电子邮件应用程序,和日历应用程序,和备忘录应用程序,和本地搜索应用程序和视频会议应用程序和人员或物体定位应用程序。本地搜索应用程序是其中用户指令数据处理系统提供关于在地理上靠近用户当前位置的本地公司或本地实体的信息的应用程序。举例来说,本地搜索口述命令可为“find a Chinese restaurant(寻找中式餐厅)”,其可调用通过网络浏览器进行搜索以基于用户当前位置来查找本地中式餐厅。或者,在本地搜索应用程序的情况下,口述命令可为“call DNJ Auto Repair(呼叫 DNJ 汽车修理厂)”。如果用户系统中的通讯录数据库未包括 DNJ 汽车修理厂的条目,那么系统可作为响应来调用网络搜索以在用户当前位置(例如,由 GPS 接收器确定的位置)本地的区域中查找称为 DNJ 汽车修理厂的公司。

[0050] 图 2 展示数据处理系统的架构的实例,其可包括多个软件模块或硬件子系统来实

施图 2 中所展示的每一块,所述块还包括数据结构,例如数据库和来自所述模块的输出。在一个实施例中,元件 201、205、207、211、215 和 219 中的每一者可实施为软件模块或软件应用程序,其通过一个或一个以上 API 进行交互以便执行图 3 中所展示的方法或图 5A 中所展示的方法或图 3 和 5A 中所展示的方法的组合。图 2 中所展示的架构还可包括语音辅助应用程序,其将经数字化的语音输入提供到语音识别器系统 201;在一个实施例中,语音辅助应用程序可包括经展示为元件 207 的所述组解释器和经展示为元件 215 的控制器,且语音辅助应用程序可作为经展示为元件 205 的预处理器进行操作。另外,语音辅助应用程序还可包括经展示为元件 211 的上下文确定模块。

[0051] 图 2 中的元件 201 可包括常规的语音识别器系统,其采用声学模型和语言模型两者来识别来自人类用户的经数字化口述命令或输入中的字词。在一个实施例中,麦克风收集来自人类用户的口述声音,且那些声音经数字化并提供到经展示为元件 201 的语音识别器系统,所述语音识别器系统又产生经展示为元件 203 的呈例如单一代码等字符编码格式的识别文本输出。此文本输出 203 接着被提供到元件 205,所述元件 205 可为创建修复数据结构的预处理器,所述修复数据结构可在一个实施例中为使用标记的剖析数据结构,下文中结合图 7 进一步描述所述剖析数据结构,图 7 提供在(例如)图 3 中所展示的方法或图 5A 中所展示的方法的修复过程中所使用的此剖析数据结构的实例。在一个实施例中,标记可在数据结构中用以表示文本输出 203 中的每一字词,且元件 207 中的所述组解释器可对那些标记或字词进行操作以便确定是否修复文本输出 203 中的每一字词。在一个实施例中,可在元件 207 中包括任选的语法剖析器以便确定短语中的哪个字词是可用以从所述组应用程序中选出特定应用程序的命令,如下文中将结合图 5A 进一步描述。在图 4 中展示可在元件 207 中使用的一组解释器的实例,其包括可使用不同算法来搜索其对应数据库或处理字词以确定文本输出中的字词与对应数据库中的字词之间是否存在匹配的一组解释器。

[0052] 图 8 展示可由所述组解释器中的一个或一个以上解释器使用以便确定在文本输出 203 中的字词与一个或一个以上数据库(例如图 4 中所展示的通讯录数据库 415)中的字词之间是否存在匹配的算法的实例。下文中将结合图 4 和 8 进一步描述这些各种算法。元件 211 可为上下文确定模块,例如图 6 中所展示的上下文确定模块 601。来自元件 211 中的此上下文确定模块的输出被提供到元件 207 中所展示的所述组解释器中的一个或一个以上解释器,以便供这些解释器在确定文本输出 203 中的字词是否可由每一解释器修复时使用上下文。

[0053] 所述解释器中的每一者可经配置或设计以与一个或一个以上数据库(例如元件 209 中的数据库)交互。这些数据库可包括通讯录或地址簿数据库、电子邮件数据库、文本消息接发数据库、媒体数据库(例如 iTunes 数据库或者歌曲或电影或歌曲与电影的组合的数据库)等。在本发明的一个实施例中还可包括其它数据库和用以在那些数据库中交互的对应解释器。在典型操作中,经设计以与特定数据库交互(而不与其它数据库交互)的解释器将处理除了命令字词以外的每一字词以确定所述字词是否匹配其对应数据库中的现有字词以及其匹配程度。举例来说,名字解释器可使用如图 8 中所展示的 n 码组算法通过搜索通讯录数据库以查找可能是名字的字词且接着经由使用一个或一个以上算法确定是否应执行修复来在通讯录数据库中搜索此字词的匹配,所述一个或一个以上算法经设计以确定所述数据库中的字词与当前正由解释器处理的字词之间的匹配程度。在一个实施例

中,每一解释器处理文本输出 203 中的除了命令字词以外的每个字词以确定所述解释器是否能修复所述字词。此外,每一解释器可提供指示匹配程度或是否应当用在数据库中找到的替代字词来修复字词的得分或置信度。

[0054] 元件 207 中所展示的所述组解释器可在一个实施例中提供一组替代解释 213,且在一个实施例中,这些替代解释由元件 215 中所示的控制器来处理,所述元件 215 可排列且合并解释器的结果以便提供经合并的解释 217,所述经合并的解释 217 可接着被提供到作为语音输入中的命令的目标的应用程序。

[0055] 在一个实施例中,元件 215 中的控制器可为图 4 中所展示的控制器模块 411,其与图 4 中所展示的一组解释器交互,所述组解释器又通过在一个或一个以上数据库中执行搜索并使用一个或一个以上算法处理那些数据库中的匹配来与所述一个或一个以上数据库交互,如下文中进一步描述。例如图 5B 中所展示的语音辅助应用程序 511 等语音辅助应用程序可向 API 调用经展示为元件 219 的目标应用程序且可提供所述命令和经修复的转录(其在一个实施例中为经合并的解释 217)作为那些调用的参数。

[0056] 图 3 中展示用于修复识别文本输出(例如识别文本输出 203)中的字词的方法。图 3 的方法可用图 2 中所展示的架构和用图 4 中所展示的架构来执行,且使用图 5B 中所展示的一个或一个以上 API。此外,图 3 的方法可使用语音辅助应用程序,其可基于在识别文本输出 203 中所检测到的命令来从一组应用程序中选出应用程序。图 3 的方法可在操作 301 中开始,在操作 301 中语音识别系统产生初始转录。此可为图 2 中所展示的识别文本输出 203。操作 301 中所使用的语音识别系统可为用以在经数字化的语音输入中识别字词的语音识别器系统 201,其包括常规的声学模型和语言模型。在操作 303 中,正在执行图 3 的方法的数据处理系统可创建用于初始转录的剖析数据结构。此剖析数据结构可在图 3 的修复过程中使用,且此数据结构的实例在图 7 中展示,下文将进一步描述图 7。在操作 305 中,系统确定转录中的命令,且还确定用户和/或系统上下文。所述命令可通过要求用户首先口述命令或通过使用剖析文本输出(例如文本输出 203)的语法剖析器确定命令且因此命令字词本身在文本输出(例如文本输出 203)中的位置来确定。

[0057] 另外,在图 3 中所展示的实施例中,操作 305 还包括确定用户和/或系统上下文。上下文信息可包括哪些应用程序被启动且正在运行以及哪些应用程序未被启动且不在运行的列表、媒体播放器是否正在播放媒体(例如歌曲或电影),且还可包括基于传感器(例如接近传感器、定向传感器、加速计和其它传感器)的用户状态。另外,上下文信息还可包括先前会话历史,其可包括(对于所述组应用程序中的每一应用程序)先前所识别的文本,例如播放披头士专辑“阿比大街(Abbey Road)”等。在一个实施例中,上下文可包括先前会话中提及的应用程序领域,且还可包括当前应用程序状态是否期望来自用户的确认(例如是或否或取消等)。(用于确认的)选择值可由系统基于当前会话上下文来指定。举例来说,用户要求系统向朋友发送电子邮件。在撰写消息之后,系统请求用户进行确认。此时,确认选择值由“是”、“取消”和“对其进行改变”填充。在一个实施例中,上下文信息还可包括用户的当前位置,例如可在如本文中所描述用户请求本地搜索的情况下使用的当前 GPS 位置。上下文信息还可包括场所上下文和/或语言上下文;举例来说,输入语言上下文可由所述组解释器用来辅助语音修复。在一个实施例中,当语言上下文(在一个实施例中,其是根据用户的偏好设置来确定的)为英语时,那么解释器可将文本输出(来自语音识别系统

的初始转录)中的“yet(仍)”修复为“yes(是)”。

[0058] 在操作 307 中,在一个实施例中,系统执行所述组解释器中的每一解释器以便确定是否需要修复所述转录(例如,识别文本输出 203)且其是否能被修复。在一个实施例中,在操作 307 中执行所述组解释器中的所有解释器。在另一实施例中,仅执行用于当前正在执行的应用程序的那些解释器来确定是否需要针对仅那些当前正在执行的应用程序修复所述转录。在一个实施例中,每一解释器基于其算法来自行决定其是否能修复由语音识别器系统(例如图 2 的元件 201 中的语音识别器系统)所提供的识别文本输出中的一个或一个以上字词。此操作经展示为元件 309。如果没有解释器能修复或决定不需要修复,那么在操作 311 中,使用由语音识别器系统所提供的初始转录(例如识别文本输出 203)且将其提供到选定应用程序。另一方面,如果已经确定一个或一个以上字词为可修复的,那么提供一组替代解释 313,其包括初始转录(例如,识别文本输出 203)以及经修复的解释。举例来说,如果确定用户在其通讯录数据库中没有“John(约翰)”而是在其通讯录数据库中具有“Jon(乔恩)”,那么字词“Jon(乔恩)”将为字词“John(约翰)”的替代解释。每一解释器维持指示一个或一个以上替代解释的匹配程度的得分或置信度,所述解释器可将所述得分或置信度提供到(例如)控制器,例如图 4 中所展示的控制模块 411。所述得分或置信度可在排列各种解释时使用以便选择最高匹配解释。所述得分或置信度可基于每个字词或基于每个短语(例如,两个或三个字词)来确定。接着,在操作 315 中,控制模块或其它模块可执行合并操作,其在一个实施例中试图基于由每一解释器所提供的置信得分或匹配或排列得分来合并不重叠的解释。可接着在操作 317 中将作为已被修复的最终转录的经合并解释提供到选定应用程序。在一个实施例中,所述选定应用程序是基于在操作 305 中识别或确定的命令来选择的。

[0059] 图 4 展示使用一组解释器和控制器模块来修复初始转录(例如识别文本输出 203)中的字词的架构的实例。在一个实施例中,每一解释器经配置或设计以通过其恰当使用算法来处理一个或一个以上数据库的某些字段中的字词。举例来说,在图 4 中所展示的实施例中,解释器 401 经配置以使用算法 A 修复通讯录数据库 415 的名字字段中的字词,所述算法 A 可在一个实施例中为 n 码组算法,例如图 8 中所展示的算法。在一个实施例中,解释器可采用若干个算法或仅一个算法。除了 n 码组算法之外,算法还可包括模糊匹配算法,其可使用测量两个文本之间的相似性的编辑距离或可使用语音学匹配算法,例如双变音位算法或探测算法。另外,可使用前缀、后缀部分标记算法,且还可使用此项技术中已知的用于确定两个文本之间的匹配或相似性程度的其它算法。在一个实施例中,不同解释器使用不同算法,使得一个解释器可使用算法 A,而另一解释器使用算法 B 而非算法 A。在一个实施例中,所述算法经定制以在对应数据库中找到匹配且搜索数据库,且明确地说,针对每一解释器经设计以校正的特定字段而定制。解释器 403 可为使用算法 A 的姓氏解释器,且解释器 405 可为使用算法 B 的全名解释器。另外,图 4 中所展示的所述组解释器可包括公司名称解释器 407,其使用与算法 B 和算法 A 不同的算法 C。解释器 401、403、405 和 407 中的每一者能够存取通讯录数据库 415 而非数据库 414 以便搜索以查找其对应字段中的每一者中的匹配。除了针对不同字段使用不同算法之外,每一解释器可在搜索其对应数据库时采用不同搜索算法。图 4 中所展示的所述组解释器还包括媒体播放器解释器 409,其经设计以搜索媒体数据库 414(例如歌曲和 / 或电影以及其它媒体的 iTunes 数据库等)中的一个或一个以

上字段。

[0060] 图 4 中所展示的架构中的每一解释器可将一个或一个以上替代解释（例如所述组替代解释 213）提供到控制器模块 411。举例来说，名字解释器 401 可提供在口述命令中似乎是名字的内容的两个不同替代解释，且那两个不同解释将各自包括指示所述解释正确的置信度或概率的打分或等级。在一个实施例中，得分或等级是基于匹配或相似性的级别。图 8 展示具有不同得分的两个解释的实例。

[0061] 图 8 展示可如何使用 n 码组算法来提供用于排列匹配的得分的实例。在此实例中，来自语音识别器系统的文本（例如识别文本输出 203）包括字词“cream（奶酪）”801。接着将来自语音识别器系统的此字词与在用户的地址簿中找到的至少两个不同字词 803 和 805 进行比较，且明确地说，将字词 801 与字词 803 和 805 进行比较，如图 8 中所展示。算法通过将字符对与文本 801 进行比较来提供得分。如从图 8 中可以看到，名字“Kream（克林姆）”是比在地址簿中找到的另一名字（具有得分 0 的名字 805）接近的匹配（因为其具有得分 3）。

[0062] 所述组解释器中的每一解释器可使用由上下文确定模块（例如元件 211 中所展示的上下文确定模块或图 6 中的上下文确定模块 601）所提供的上下文信息。上下文可包括指示是否针对所述组应用程序中的一个或一个以上应用程序口述命令的先前会话历史以及任选地，命令本身和字词本身。先前会话历史 603 提供此信息，其还可包括例如触摸屏或键盘上的用户输入等先前用户输入等。上下文确定模块可根据先前会话历史而且还根据应用程序的状态 605 确定上下文，所述应用程序的状态 605 可包括指示哪些应用程序被启动且正在执行、哪些应用程序未被启动且因此不在执行以及媒体是否正在播放等的指示符。举例来说，媒体播放器解释器 409 可使用媒体正在播放时的上下文指示符来将字词“stock（股票）”的初始转录修复为“stop（停止）”，因为用户先前已经致使媒体开始播放且在媒体正在播放时的那个上下文中，由媒体播放器解释器 409 将字词“stock（股票）”解释为“stop（停止）”。上下文确定模块可确定语言或场所上下文，如本文中所描述。上下文确定模块 601 还可包括来自传感器（例如定向传感器或接近传感器或光传感器等）的输入作为上下文确定过程的一部分。另外，上下文确定模块 601 可包括先前用户输入历史。上下文确定模块 601 收集此各种关于上下文的信息，且将其提供到解释器，所述解释器使用所述上下文来帮助决定是否能在口述命令输入中修复字词。

[0063] 现将结合图 7 提供根据本发明的实施例的具体实施方案。数据结构 701 用标记 703 表示识别文本输出中的字词。

[0064] 语音识别过程获取语音音频记录且将其转录为一个或一个以上文本解释。初始转录被展示为文本字符串 705。这些转录文本存储在表状数据结构中，所述表状数据结构在一个实施例中被称为识别且展示在图 7 中。

[0065] 识别的基本构造是标记。标记是表示转录的原子单位的不可变字符串。如果转录由一序列标记 703 组成，那么每一标记被包封在称为短语 707 的第二级数据结构中。短语是支柱 - 主要数据结构。短语对象的有序列表形成识别。短语数据结构的存​​在将允许替代转录。

[0066] 举例来说，当用户说“Call John Smith on mobile（呼叫约翰·史密斯的移动电话）”时，语音辅助应用程序可产生图 7 中所展示的识别：

[0067] 语音修复过程获取识别对象（图 7 中所展示）作为输入且在原始识别的任何部分需要且能修复的情况下产生经修改的识别对象。

[0068] 可创建称为元修复的内部数据结构以辅助语音修复过程。此数据结构可由以下各项组成：原始识别对象（图 7 中所展示）、修复对象和相对于原始转录的标记位置。

[0069] 此处是用于图 7 中所展示的数据结构的标记位置查找表的实例：

[0070] Pair<Start,End> :特定标记字符串相对于原始识别文本的开始和结束位置

[0071] “Call John Smith on Mobile(呼叫约翰·史密斯的移动电话)”

[0072] 元修复的标记位置列表：

[0073] [0]:Pair<0, 3>

[0074] [1]:Pair<5, 8>

[0075] [2]:Pair<10, 14>

[0076] [3]:Pair<16, 17>

[0077] [4]:Pair<19, 24>

[0078] 元修复的修复对象由语音修复程序所产生的替代解释列表组成。用以表示替代解释的数据结构称为修复解释。

[0079] 修复解释由作为用于原始识别中的子字符串的看似可能的替换的文本以及所述子字符串的开始和结束位置组成。举例来说,如果“Jon(乔恩)”应为用于“John(约翰)”的替代,那么用于图 7 中所展示的数据结构的修复解释可被描述为如下：

[0080] 修复解释：

[0081] 文本：“Jon”

[0082] 开始：5

[0083] 结束：8

[0084] 元修复对象含有用以执行解释合并的信息。合并逻辑在将原始转录在一个实施例中传递通过所有解释器之后且在产生一个或一个以上修复解释的情况下发生。以下伪码提供可用以合并来自所述组解释器的不重叠的解释的合并函数的实例。在图 7 中展示“INPUT:original:Recognition(输入：原始：识别)”。

[0085]

FUNCTION merge

INPUT: original::Recognition

repair::Repair

OUTPUT: original::Recognition

BEGIN

IF repair.getRepairInterpretation() is EMPTY

RETURN original;

ELSE

String originalText = repair.getOriginalText();

List interpretations = repair.getRepairInterpretations();

List ranked = rank(originalText, interpretations);

List nonOverlap = findNonOverlapInterpretations(ranked);

FOR each RepairInterpretation i in nonOverlap

[0086]

```

List<Phrase> phrases = findRepairPhrases(original, i)
List<Token> tokens = makeRepairTokens(i);
addTokensToPhraseHead(phrase, tokens);
END FOR

```

```

RETURN original;

```

```

END

```

```

FUNCTION rank

```

```

    INPUT:  originalText::String

```

```

           interpretations::List<RepairInterpretation>

```

```

    OUTPUT: ranked::List<RepairInterpretation>

```

RETURN a sorted interpretation list using a custom Java Comparator that imposes a total ordering on a collection of objects.

```

/**

```

```

    Collections.sort(sorted, new Comparator<Repair.Interpretation>() {

```

```

        @Override

```

```

        public int compare(Repair.Interpretation o1,

```

```

Repair.Interpretation o2) {

```

```

            if (scoreInterpretation(o1) > scoreInterpretation(o2)) {

```

```

                return -1;

```

```

            } else if (scoreInterpretation(o1) <

```

```

scoreInterpretation(o2)) {

```

```

                return 1;

```

```

            } else {

```

```

                return 0;

```

```

            }

```

```

        }

```

```

        private float scoreInterpretation(Repair.Interpretation itp) {

```

```

            float score = 0;

```

```

            if (itp.isPerfectMatch(getOriginalText())) {

```

[0087]

```
        score = getOriginalText().length();
    }
    score += (float) itp.getLength() / getOriginalText().length();
    return score;
}
});
**/
```

END

[0088] 图 5A 说明本发明的一个实施例，其可使用允许系统基于在口述命令中检测到的命令来选择确定为口述命令的目标的恰当应用程序的语音辅助应用程序。在一个实施例中，语音辅助应用程序可使用语法剖析器来在口述命令中检测命令且进而选择恰当的应用程序或者可要求用户在每个口述命令中将命令口述为第一个字词，使得系统可确定哪个字词相对于口述输入中的其它字词来说是命令。在此实施例中，系统可基于口述命令来选择特定应用程序且进而不要求用户在口述命令之前选择应用程序以便使得所述应用程序成为系统的具有语音输入焦点的最前端应用程序。换句话说，在此实施例中，语音辅助应用程序可基于对于一个应用程序为恰当的命令来从一组应用程序中选出所述一个应用程序。在操作 501 中，系统可接收可指向一组应用程序中的一个应用程序的语音输入。所述应用程序可正在执行或不在执行。在一个实施例中，所述方法可经配置以使得仅正在执行的应用程序将在所述组应用程序中，而在另一实施例中，所有应用程序不管是否正在执行均可在所述组应用程序中使得每一应用程序可接收语音输入。在操作 503 中，语音辅助应用程序可接着确定语音输入中的命令，且可接着基于所确定的命令来选择恰当的应用程序。举例来说，如果命令是“call(呼叫)”，那么在一个实施例中恰当的应用程序是用以与语音输入(例如，call mom at home(呼叫在家里的妈妈))中所指定的人员建立电话呼叫的电话拨号器。所述命令可通过使用语法剖析器定位命令来确定，其中可通过指令用户使用动词来要求所述命令为动词，或系统可要求用户将命令放置在口述字词序列中的固定位置中。在一个实施例中，如果命令为字词“tell(告诉)”，那么选定应用程序是文本消息接发(SMS)应用程序；如果命令是字词“播放”或字词“停止”，那么选定应用程序是媒体播放器应用程序，等等。

[0089] 接着，在操作 505 中，执行解释器以修复识别语音输入中的一个或一个以上字词。如果在执行解释器之前在操作 503 中选择应用程序，那么可仅执行经设计以与所述特定应用程序一起工作的那些解释器而非执行所述系统中对于能够通过语音输入接收口述命令的所有应用程序可用的所有解释器。操作 505 类似于操作 307，且可结合图 6 中所展示的上下文确定模块来使用图 4 中所展示的架构。在一个实施例中，不修复在操作 503 中确定或检测到的命令。在此情况下，解释器在对解释器与之交互的一个或一个以上数据库执行搜索时将文本输入中的命令分析为结束字词。接着，在操作 507 中，语音辅助应用程序可将命令传递到在操作 503 中确定的选定应用程序且可将经修复的转录(其通过执行解释器和合并替代解释来产生)传递到选定应用程序。在一个实施例中，语音辅助应用程序可执行将

命令连同经修复的转录传递通过 API,例如如图 5B 中所展示的一个或一个 API。

[0090] 图 5B 中所展示的语音辅助应用程序 511 可与执行图 5A 的一个或一个以上方法的语音辅助应用程序相同。语音辅助应用程序 511 可通过经由 API 514 对操作系统 516 进行上下文调用来确定上下文,所述操作系统 516 又返回上下文信息,例如上文描述和 / 或图 6 中展示的上下文信息。上下文信息还可包括哪些应用程序正在执行以及哪些应用程序先前接收到用户输入或先前接收到口述命令的列表。语音辅助应用程序 511 还可对语音识别器系统进行调用,所述语音识别器系统可为在图 5B 中所展示的系统上执行的软件应用程序,其表示包括操作系统 516 以及语音辅助应用程序 511 和一组应用程序中的一个或一个以上应用程序(例如应用程序 518 和 520)的软件堆叠。应用程序 518 和 520 可接收从语音辅助应用程序传递通过 API 512 的命令。

[0091] 下文是在一个实施例中由语音辅助应用程序所进行的语音修复的三个使用实例。

[0092] (1) “对齐网格”语音拨号。语音辅助应用程序允许用户使用语音呼叫地址簿数据库中的联系人。用户在地址簿中具有名为“Marc Dickinson(玛克·迪金森)”的联系人,而没有名为“马克(Mark)”或“迪克(Dick)”的联系人。当用户说“Call Marc Dickinson(呼叫玛克·迪金森)”时,语音识别不正确地将输入转录为“Call Mark Dick son(呼叫马克·迪克儿子)”。代替告诉用户所述辅助程序因为其无法在数据库中找到“Mark Dick son(马克·迪克儿子)”而无法完成所述操作,语音修复可利用联系人姓名拼写且使用模糊匹配算法来产生较为看似可能的替代转录“Call Marc Dickinson(呼叫玛克·迪金森)”。(2) 消除用户意图的歧义。辅助语音应用程序允许用户发送 SMS 消息且做出语音拨号请求。当用户说“Tell my wife to pick up milk and fruits if she goes to Safeway after work(让我妻子去取牛奶和水果,如果她在下班后去西夫韦的话)”时,辅助程序自动地向用户的妻子编写文本消息。归因于识别错误,语音系统可不正确地将动作字词“tell(让)”转录为“call(呼叫)”或“tall(高)”。因为请求“Call my wife to pick up milk and fruits if she goes to Safeway after work(呼叫我妻子去取牛奶和水果,如果她在下班后去西夫韦的话)”或“tall my wife to pick up milk and fruits if she goes to Safeway after work(高我妻子去取牛奶和水果,如果她在下班后去西夫韦的话)”在一个实施例中不会映射到辅助程序中的任何可起作用的任务,所以默认响应通常是“Sorry! I don't know what you meant(对不起!我不知道你的意思)”。语音修复可通过使用上下文消除语音意图的歧义来解决此问题。举例来说,在已知字词“tell(让)”和“tall(高)”之间的编辑距离较短且语音拨号命令通常在目标人员标记之后没有长连续字符串的情况下,解释器可将原始转录重写为“Tell my wife to pick up milk and fruits if she plan to visit Safeway after work(让我妻子去取牛奶和水果,如果她打算在下班后去西夫韦的话)”。(3) 消除命令 / 系统关键字词的歧义。语音系统可不正确地转录短关键词发音。举例来说,用户说“Stop(停止)”而初始转录为“Stock(股票)”;用户说“Yes(是)”而初始转录为“Yet(仍)”。语音修复可通过基于一个或一个以上上下文线索而在原始转录文本为较不看似可能的解释时提供替代转录来解决这些问题。举例来说,当辅助程序正提示用户以获得是 / 否确认时,用户将不可能说“Yet(仍)”作为跟随响应。代替返回“Yet(仍)”作为最终转录,语音修复可用“Yes(是)”来将其盖写作为较看似可能的语音输入。类似的修复逻辑适用于媒体播放器领域。如果用户刚刚请求播放歌曲而紧接的语音转录为“Stock(股

票)”,那么语音修复可用“Stop(停止)”来将其盖写作为较看似可能的命令转录。

[0093] 图 9 展示可与本发明的一个实施例一起使用的数据处理系统 900 的实例。举例来说且在一个实施例中,系统 900 可实施为便携式数据处理装置,例如智能电话或输入板(例如,iPad)装置或膝上型计算机或娱乐系统。图 9 中所展示的数据处理系统 900 包括处理系统 911,其可为一个或一个以上微处理器或其可为单芯片系统(集成电路),且所述系统还包括用于存储数据和程序以供处理系统执行的存储器 901。存储器 901 可存储(例如)结合图 2 所描述的软件组件,且存储器 901 可为任何已知形式的机器可读非暂时存储媒体,例如半导体存储器(例如,快闪;DRAM;SRAM;等等)。系统 900 还包括音频输入/输出子系统 905,其可包括麦克风和扬声器,以用于例如重放音乐或通过扬声器和麦克风提供电话功能性。麦克风可接收本文中所描述的语音输入,且所述输入可经数字化并提供到如本文中所描述的语音识别器系统。

[0094] 显示器控制器和显示装置 909 可为用户提供视觉用户接口;此接口可包括图形用户接口,其类似于当在 iPhone 或 iPad 上运行 OS X 操作系统软件或 iOS 软件时在麦金托什计算机上所展示的图形用户接口。系统 900 还包括一个或一个以上无线收发器 903 来与另一数据处理系统通信。无线收发器可为 WLAN 收发器(例如,WiFi)、红外线收发器、蓝牙收发器和/或无线蜂窝式电话收发器。将了解,额外组件(未图示)在某些实施例中也可作为系统 900 的一部分,且在某些实施例中,比图 9 中所展示的组件少的组件也可在数据处理系统中使用。系统 900 进一步可包括一个或一个以上通信端口 917 以与另一数据处理系统通信。通信端口可为 USB 短路、火线端口、蓝牙接口、对接端口等。

[0095] 数据处理系统 900 还包括一个或一个以上输入装置 913,其经提供以允许用户向系统提供输入。这些输入装置可为小键盘或键盘或触摸板或多点触摸板,其与例如显示装置 909 等显示装置重叠并集成。数据处理系统 900 还可包括任选的输入/输出装置,其可为用于对接口的连接器。将了解,一个或一个以上总线(未图示)可用以互连各种组件,如此项技术中众所周知的。图 9 中所展示的数据处理系统可为手持式计算机或个人数字助理(PDA),或具有 PDA 类功能性的蜂窝式电话,或包括蜂窝式电话的手持式计算机,或媒体播放器(例如 iPod),或游戏或娱乐装置,或组合这些装置的多个方面或功能的装置(例如在一个装置中与 PDA 和蜂窝式电话组合的媒体播放器或嵌入式装置或其它消费型电子装置)。在其它实施例中,数据处理系统 900 可为网络计算机或位于另一装置内的嵌入式处理装置,或具有比图 9 中所展示的组件少的组件或可能多的组件的其它类型的数据处理系统。

[0096] 数据处理系统 900 可任选地包括一个或一个以上硬件装置,其经设计以数字化并存储由音频 I/O 905 中的麦克风所接收的人类语音。

[0097] 本发明的至少某些实施例可为数字媒体播放器(例如便携式音乐和/或视频媒体播放器)的一部分,所述数字媒体播放器可包括用以呈现媒体的媒体处理系统、用以存储媒体的存储装置,且可进一步包括与天线系统和媒体处理系统耦合的射频(RF)收发器(例如,用于蜂窝式电话的 RF 收发器)。在某些实施例中,存储在远程存储装置上的媒体可通过 RF 收发器发射到媒体播放器。媒体可为(例如)一个或一个以上音乐或其它音频、静止图片或运动图片。

[0098] 便携式媒体播放器的实例在第 7,345,671 号公开美国专利和第 2004/0224638 号

美国公开专利申请案中描述,所述两者均以引用的方式并入本文中。

[0099] 可在一些实施例中使用一个或一个以上应用程序编程接口 (API)。API 是由程序代码组件或硬件组件 (下文中称为“API 实施组件”) 实施的接口,其允许不同的程序代码组件或硬件组件 (下文中称为“API 调用组件”) 接入并使用由 API 实施组件所提供的一个或一个以上函数、方法、程序、数据结构、类别和 / 或其它服务。API 可界定在 API 调用组件与 API 实施组件之间传递的一个或一个以上参数。

[0100] API 允许 API 调用组件的开发者 (其可为第三方开发者) 利用由 API 实施组件所提供的指定特征。可存在一个 API 调用组件或可存在一个以上此类组件。API 可为计算机系统或程序库提供以便支持请求来自应用程序的服务的源代码接口。操作系统 (OS) 可具有多个 API 来允许在 OS 上运行的应用程序调用那些 API 中的一者或一者以上,且服务 (例如程序库) 可具有多个 API 来允许使用所述服务的应用程序调用那些 API 中的一者或一者以上。API 可依据可在构建应用程序时解释或编译的编程语言来指定。

[0101] 在一些实施例中,API 实施组件可提供一个以上 API,其各自提供由 API 实施组件实施的功能性的不同视图或具有接入由 API 实施组件实施的功能性的不同方面的不同方面。举例来说,API 实施组件的一个 API 可提供第一组功能且可暴露到第三方开发者,且 API 实施组件的另一 API 可被隐蔽 (不暴露) 且提供第一组功能的子组并还提供另一组功能,例如测试或调试不在第一组功能中的功能。在其它实施例中,API 实施组件可自身经由基础 API 调用一个或一个以上其它组件且因此为 API 调用组件和 API 实施组件两者。

[0102] API 界定 API 调用组件在接入并使用 API 实施组件的指定特征时使用的语言和参数。举例来说,API 调用组件通过由 API 暴露的一个或一个以上 API 调用或引用 (例如由函数或方法调用来体现) 接入 API 实施组件的指定特征且经由 API 调用或引用使用参数传递数据和控制信息。API 实施组件可响应于来自 API 调用组件的 API 调用而通过 API 返回值。尽管 API 界定 API 调用的语法和结果 (例如,如何调用 API 调用以及 API 调用进行什么),但 API 可不展现 API 调用如何完成所述 API 调用所指定的函数。经由调用 (API 调用组件) 与 API 实施组件之间的一个或一个以上应用程序编程接口来传送各种 API 调用。传送 API 调用可包括发布、起始、引用、调用、接收、返回或响应于所述函数调用或消息;换句话说,传送可描述 API 调用组件或 API 实施组件中的任一者所进行的动作。API 的函数调用或其它引用可通过参数列表或其它结构发送或接收一个或一个以上参数。参数可为常数、关键字、数据结构、对象、对象类别、变量、数据类型、指针、阵列、列表或指向函数或方法的指针或者用以提及将经由 API 传递的数据或其它项目的另一方式。

[0103] 此外,数据类型或类别可由 API 提供且由 API 实施组件实施。因此,API 调用组件可声明变量、使用指针、通过使用在 API 中所提供的定义来使用或例示所述类型或类别的常数值。

[0104] 一般来说,API 可用以接入由 API 实施组件提供的服务或数据,或起始由 API 实施组件所提供的操作或计算的执行。借助于实例,API 实施组件和 API 调用组件可各自为操作系统、库、装置驱动器、API、应用程序或其它模块中的任一者 (应理解,API 实施组件和 API 调用组件可为彼此相同或不同类型的模块)。API 实施组件可在一些情况下至少部分地以固件、微码或其它硬件逻辑来体现。在一些实施例中,API 可允许客户端程序使用由软件开发套件 (SDK) 库所提供的服务。在其它实施例中,应用程序或其它客户端程序可使用由

应用程序框架所提供的 API。在这些实施例中,应用程序或客户端程序可并入对由 SDK 提供和由 API 提供的函数或方法的调用,或使用在 SDK 中界定和由 API 提供的数据类型或对象。应用程序框架可在这些实施例中为对所述框架所界定的各种事件做出响应的程序提供主要事件环路。API 允许应用程序使用应用程序框架指定所述事件以及对事件的响应。在一些实施方案中,API 调用可向应用程序报告硬件装置的能力或状态,包括与例如输入能力和状态、输出能力和状态、处理能力、功率状态、存储容量和状态、通信能力等方面相关的那些,且 API 可部分地通过固件、微码或部分地在硬件组件上执行的其它低级逻辑来实施。

[0105] API 调用组件可为本地组件(即,在与 API 实施组件相同的数据处理系统上)或远程组件(即,在与 API 实施组件不同的数据处理系统上),其经由网络通过 API 而与 API 实施组件进行通信。应理解,API 实施组件也可充当 API 调用组件(即,其可对由不同 API 实施组件暴露的 API 做出 API 调用),且 API 调用组件也可通过实施暴露到不同 API 调用组件的 API 来充当 API 实施组件。

[0106] API 可允许以不同编程语言编写的多个 API 调用组件与 API 实施组件通信(因此,API 可包括用于翻译 API 实施组件与 API 调用组件之间的调用和返回的特征);然而,API 可依据特定编程语言来实施。API 调用组件可在一个实施例中调用来自不同提供者的 API,例如来自 OS 提供者的一组 API 和来自插件提供者的另一组 API 和来自另一提供者(例如,软件库的提供者)或另一组 API 的创建者的另一组 API。

[0107] 图 11 为说明可在本发明的一些实施例中使用的示范性 API 架构的框图。如图 11 中所展示,API 架构 1100 包括实施 API 1120 的 API 实施组件 1110(例如,操作系统、库、装置驱动器、API、应用程序、软件或其它模块)。API 1120 指定可由 API 调用组件 1130 使用的 API 实施组件的一个或一个以上函数、方法、类别、对象、协议、数据结构、格式和/或其它特征。API 1120 可指定至少一个调用协定,其指定 API 实施组件中的函数如何从 API 调用组件接收参数且所述函数如何将结果返回到 API 调用组件。API 调用组件 1130(例如,操作系统、库、装置驱动器、API、应用程序、软件或其它模块)通过 API 1120 做出 API 调用以接入并使用 API 实施组件 1110 的由 API 1120 指定的特征。API 实施组件 1110 可响应于 API 调用而通过 API 1120 将值返回到 API 调用组件 1130。

[0108] 将了解,API 实施组件 1110 可包括未通过 API 1120 指定且不可由 API 调用组件 1130 获得的额外函数、方法、类别、数据结构和/或其它特征。应理解,API 调用组件 1130 可位于与 API 实施组件 1110 相同的系统上或可远程地定位且经由网络使用 API 1120 接入 API 实施组件 1110。尽管图 11 说明单个 API 调用组件 1130 与 API 1120 交互,但其理解,可以与 API 调用组件 1130 不同的语言(或相同的语言)编写的其它 API 调用组件也可使用 API 1120。

[0109] API 实施组件 1110、API 1120 和 API 调用组件 1130 可存储在机器可读非暂时存储媒体中,所述机器可读非暂时存储媒体包括可由机器(例如,计算机或其它数据处理系统)读取的呈有形形式的用于存储信息的任何机制。举例来说,机器可读媒体包括磁盘、光盘、随机存取存储器、只读存储器、快闪存储器装置等,且可为本地存储媒体或位于通过一个或一个以上网络耦合到客户端装置的远程装置上的存储媒体。

[0110] 在图 10(“软件堆叠”)(示范性实施例)中,应用程序可使用若干服务 API 对服务 1 或 2 且使用若干 OS API 对操作系统(OS)做出调用。服务 1 和 2 可使用若干 OS API 对

OS 做出调用。

[0111] 请注意,服务 2 具有两个 API,其中一者(服务 2 API 1)从应用程序 1 接收调用且将值返回到应用程序 1,且另一者(服务 2 API 2)从应用程序 2 接收调用且将值返回到应用程序 2。服务 1(其可例如为软件库)对 OS API 1 做出调用且从 OS API 1 接收返回值,且服务 2(其可例如为软件库)对 OS API 1 和 OS API 2 两者做出调用且从 OS API 1 和 OS API 2 两者接收返回值。应用程序 2 对 OS API 2 做出调用且从 OS API 2 接收返回值。

[0112] 说明书中对“一个实施例”或“一实施例”的参考意味着结合所述实施例描述的特定特征、结构或特性包括在本发明的至少一个实施例中。在说明书的各种地方出现短语“在一个实施例中”不必全部指代同一实施例。

[0113] 在前述说明书中,已经参考本发明的特定示范性实施例来描述了本发明。将明显的是,可在不脱离如所附权利要求书中所陈述的本发明的较广精神和范围的情况下对本发明做出各种修改。因此,应在说明性意义而非限制性意义上看待说明书和图式。

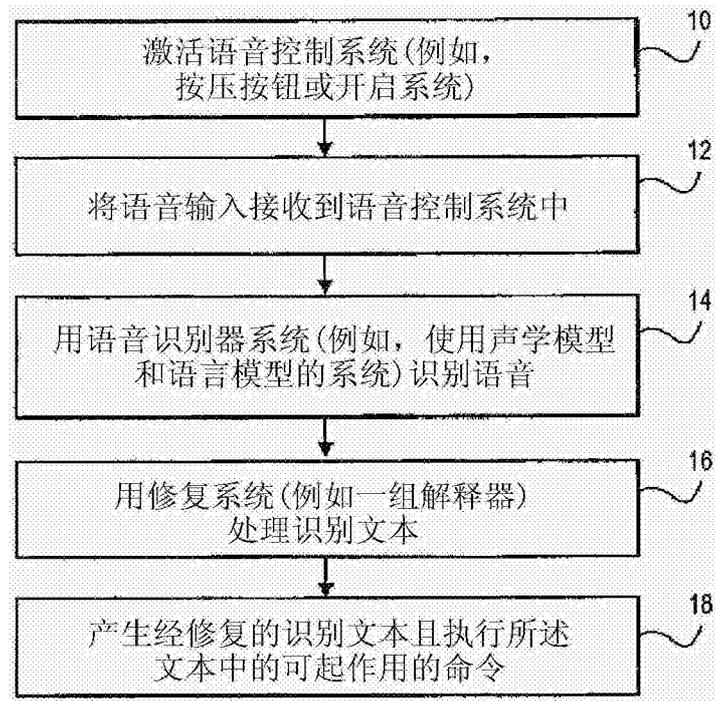


图 1

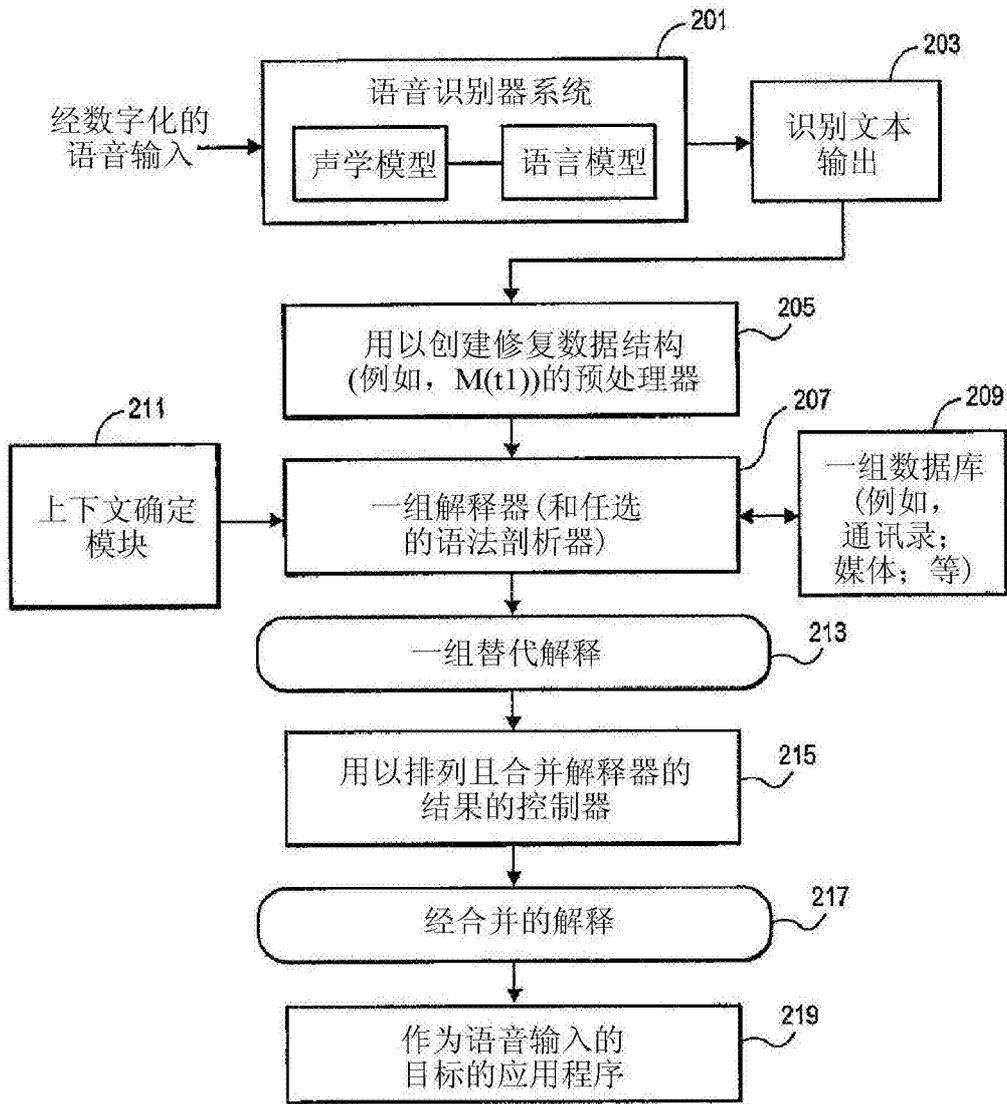


图 2

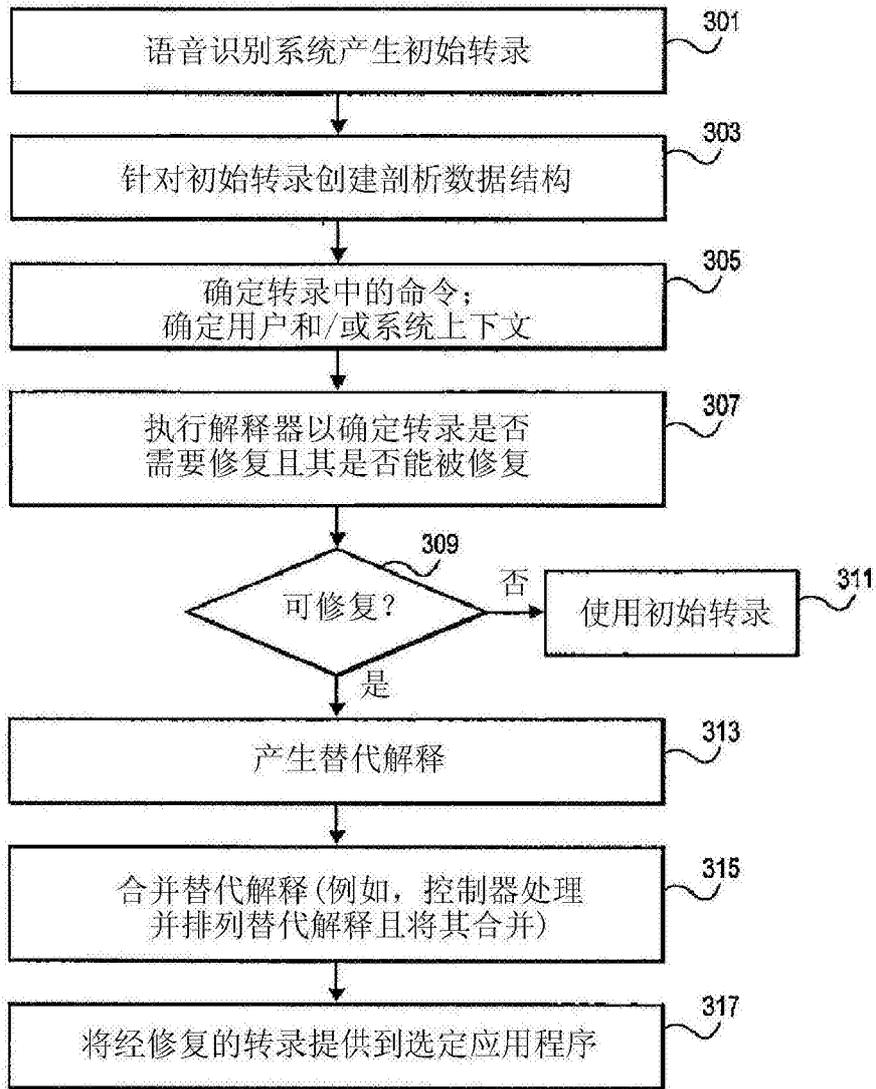


图 3

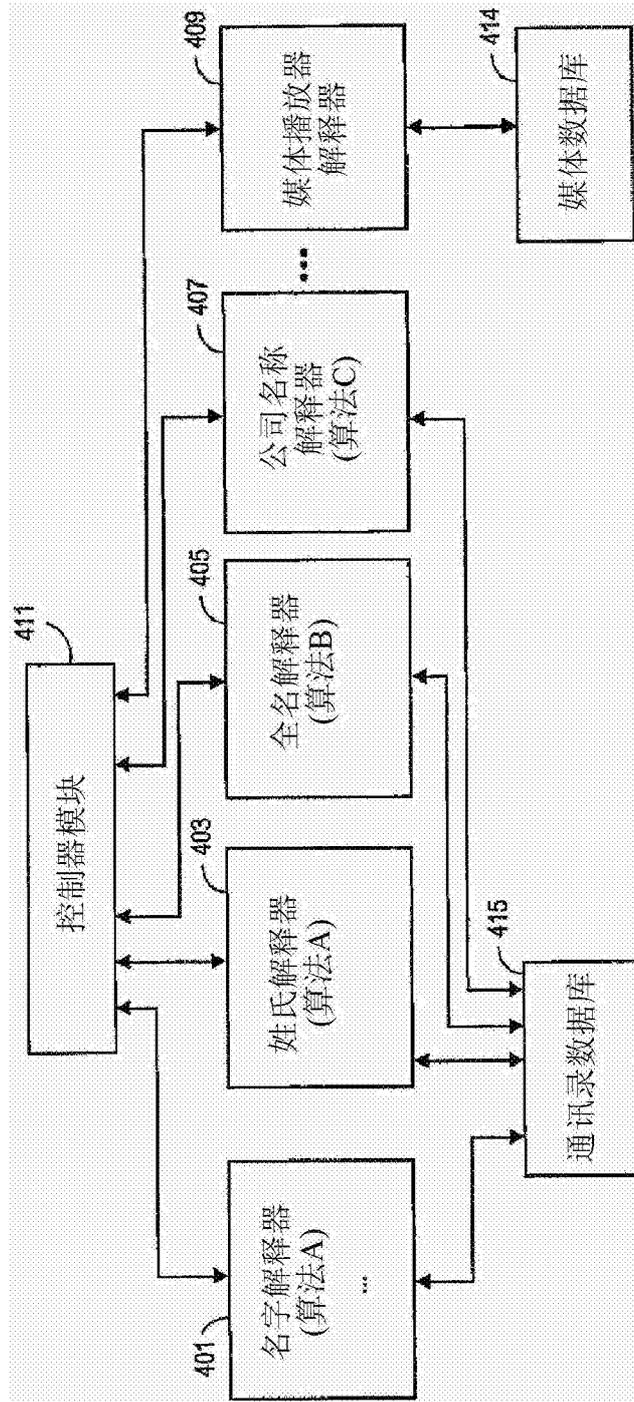


图 4

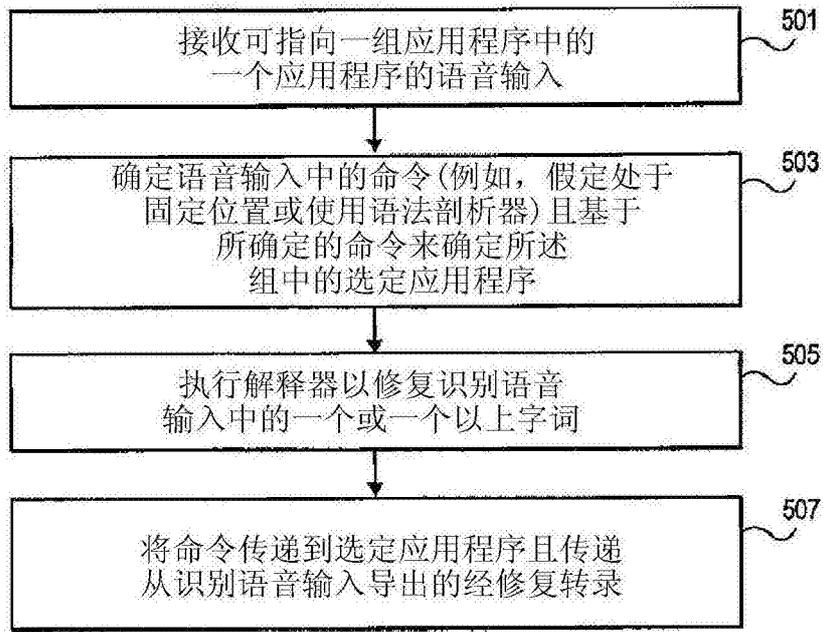


图 5A

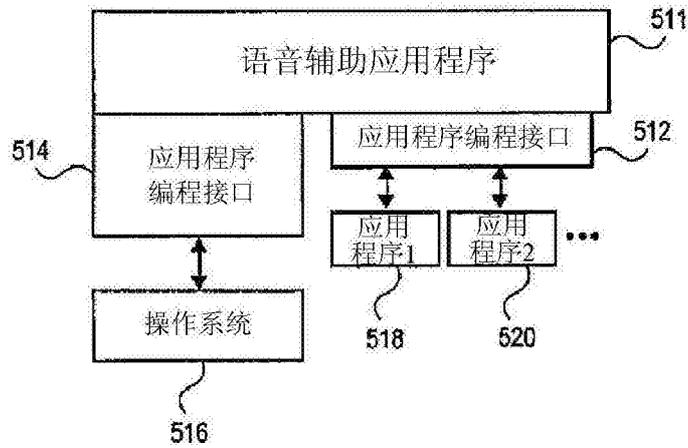


图 5B

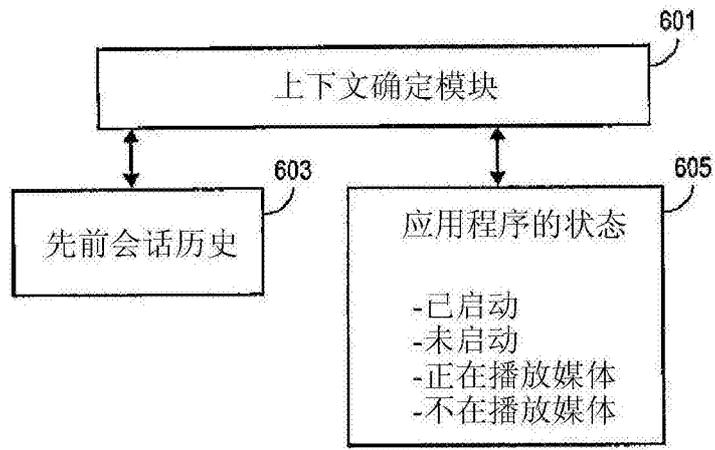


图 6

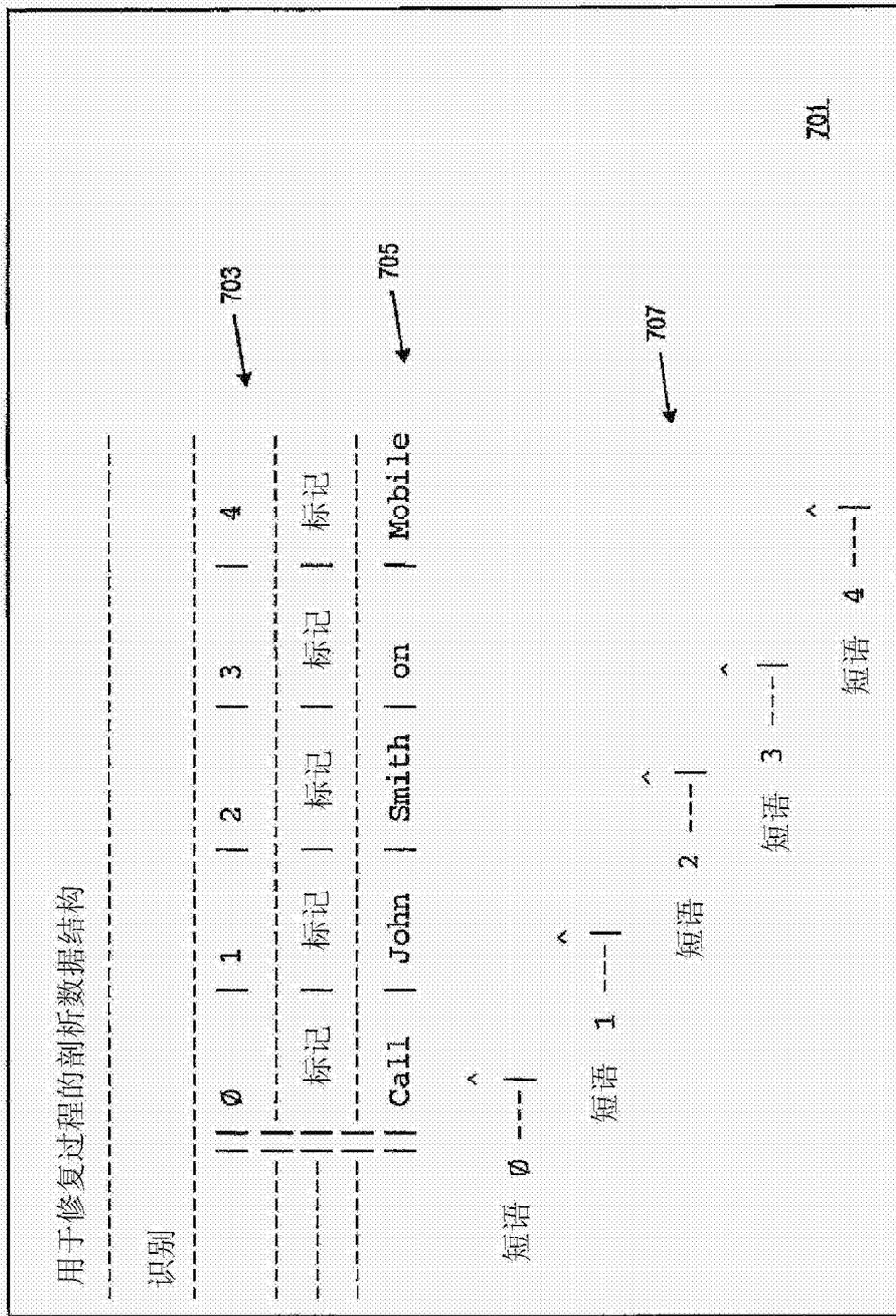


图 7

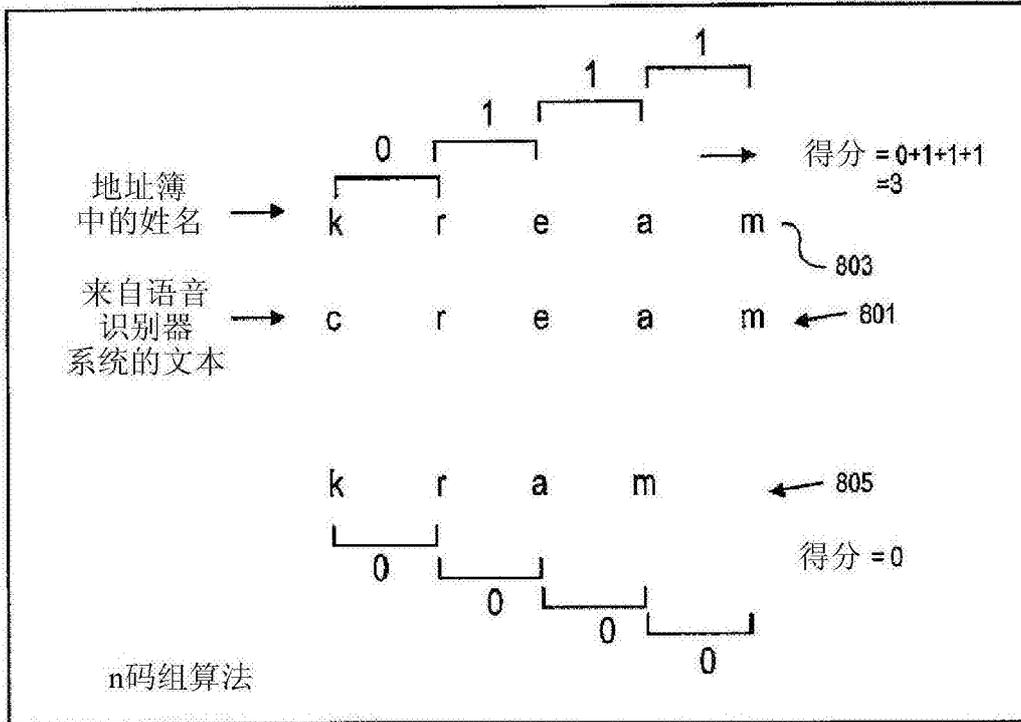


图 8

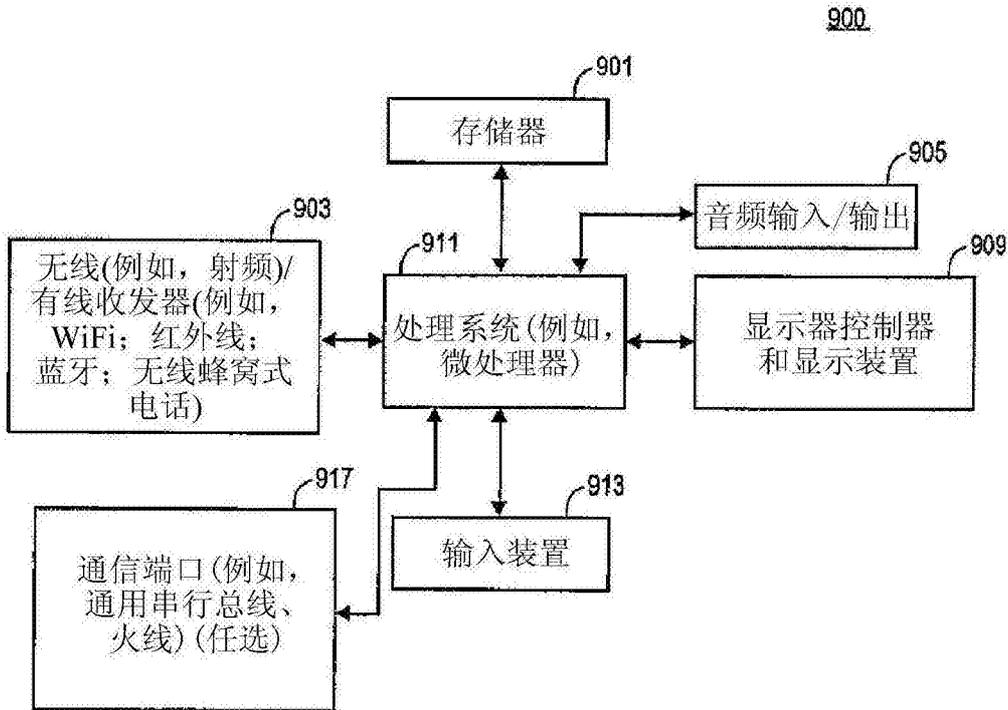


图 9

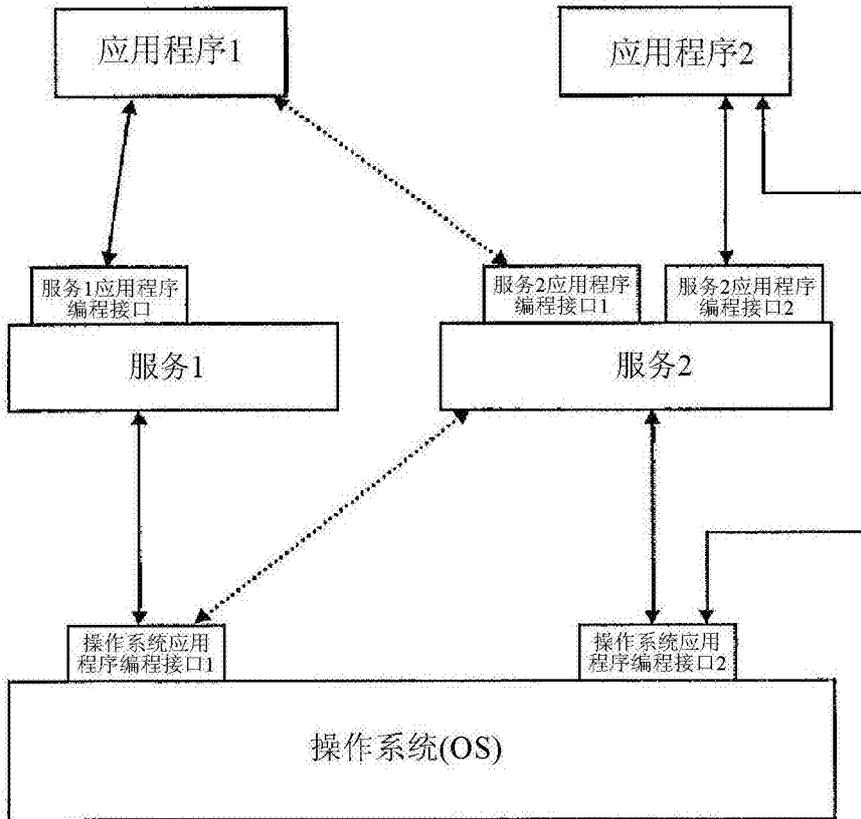


图 10

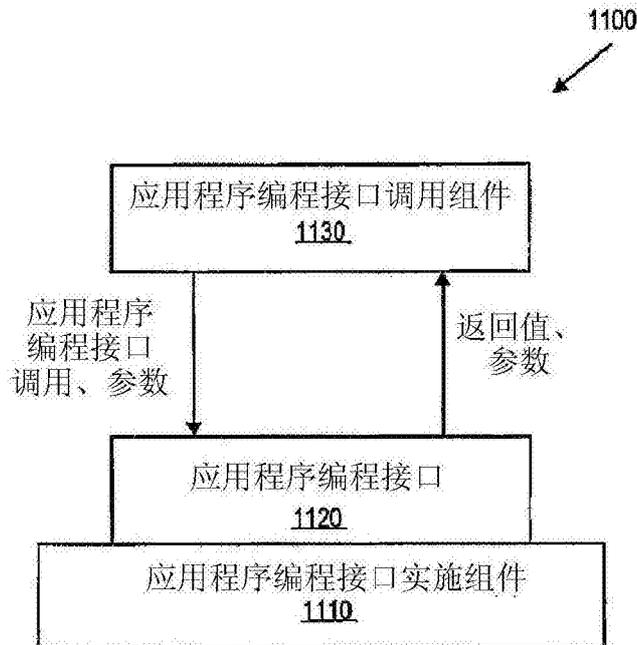


图 11