(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 109791767 B
(45)授权公告日 2023.09.05

(21)申请号 201780060607.0
(22)申请日 2017.09.14
(65)同一申请的已公布的文献号
申请公布号 CN 109791767 A
(43)申请公布日 2019.05.21
(30)优先权数据
15/281973 2016.09.30 US
(85)PCT国际申请进入国家阶段日
2019.03.29
(86)PCT国际申请的申请数据
PCT/EP2017/073162 2017.09.14
(87)PCT国际申请的公布数据
WO2018/059957 EN 2018.04.05
(73)专利权人 罗伯特·博世有限公司
地址 德国斯图加特
(72)发明人 Z.周 Z.冯
(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 毕铮 申屠伟进
(51)Int.Cl.
G10L 15/32(2006.01)
G10L 15/22(2006.01)

(56)对比文件
CN 101093478 A,2007.12.26
EP 1562178 A1,2005.08.10
US 2016104478 A1,2016.04.14
US 6006182 A,1999.12.21
WO 2014180218 A1,2014.11.13
US 2013132089 A1,2013.05.23
CN 1351744 A,2002.05.29
WO 2015079568 A1,2015.06.04
US 2008208585 A1,2008.08.28
CA 2814109 A1,2013.10.30
AU PR082400 A0,2000.11.09
CN 101082836 A,2007.12.05
JP 2016099501 A,2016.05.30
CA 2507999 A1,2004.07.08
US 2002116196 A1,2002.08.22
GB 0426347 D0,2005.01.05
CA 1246229 A,1988.12.06
US 2007038449 A1,2007.02.15 (续)

审查员 毛健

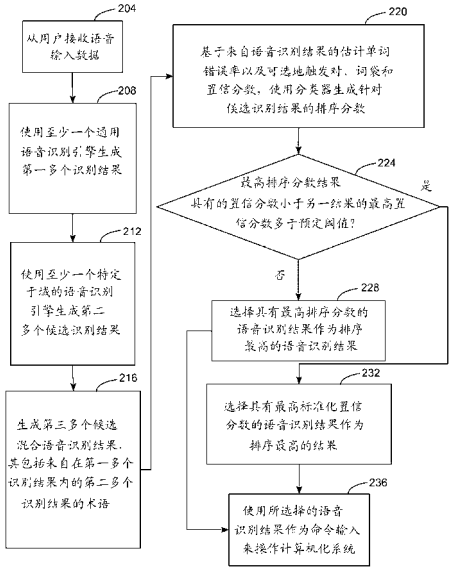
权利要求书4页 说明书13页 附图4页

(54)发明名称

用于语音识别的系统和方法

(57)摘要

用于自动化语音识别的方法包括分别使用第一通用语音识别引擎和第二特定于域的语音识别引擎生成对应于音频输入数据的第一和第二多个候选语音识别结果。所述方法还包括:生成第三多个候选语音识别结果,其包括第一多个语音识别结果中的一个中包括的多个单词和第二多个语音识别结果中的另一个中包括的至少一个单词;使用成对排序器对第三多个候选语音识别结果进行排序,以标识排序最高的候选语音识别结果;以及使用排序最高的语音识别结果作为来自用户的输入来操作自动化系统。



CN 109791767 B

[转续页]

[接上页]

(56)对比文件

郑铁然等.汉语语音文档检索中后验概率的索引方法.《哈尔滨工业大学学报》.2009,(第08期),

1.一种用于自动化系统中的语音识别的方法,包括:

利用音频输入设备生成对应于来自用户的语音输入的音频输入数据;

利用控制器,使用第一通用语音识别引擎生成对应于音频输入数据的第一多个候选语音识别结果;

利用控制器,使用第一特定于域的语音识别引擎生成对应于音频输入数据的第二多个候选语音识别结果;

利用控制器生成第三多个候选语音识别结果,第三多个候选语音识别结果中的每个候选语音识别结果包括第一多个候选语音识别结果中的一个中包括的多个单词和第二多个候选语音识别结果中的另一个中包括的至少一个单词;

利用控制器,使用成对排序器对至少第三多个候选语音识别结果进行排序,以标识排序最高的候选语音识别结果;以及

利用控制器,使用排序最高的候选语音识别结果作为来自用户的输入来操作所述自动化系统。

2.根据权利要求1所述的方法,第三多个候选语音识别结果中的至少一个候选语音识别结果的生成还包括:

利用控制器标识第一多个候选语音识别结果中的第一候选语音识别结果的第一多个单词中的第一单词,所述第一单词对应于第二多个候选语音识别结果中的第二候选语音识别结果中的第二多个单词中的第二单词,第二单词与第一单词不同;以及

利用控制器生成针对第三多个候选语音识别结果的候选语音识别结果,所述候选语音识别结果包括来自第一候选语音识别结果的第一多个单词与来自第二候选语音识别结果的代替来自第一候选语音识别结果的第一单词的第二单词。

3.根据权利要求2所述的方法,还包括:

利用控制器,基于第二多个单词中也存在于第一多个单词中的至少一个单词的位置,将第二候选语音识别结果中的第二多个单词与第一候选语音识别结果中的第一多个单词对齐;以及

利用控制器,标识第一多个语音识别中的第一候选语音识别结果的第一多个单词中的第一单词,所述第一单词在与第二多个单词对齐的第一多个单词中的单词位置处对应于第二候选语音识别结果中的第二多个单词中的第二单词。

4.根据权利要求1所述的方法,排序还包括:

利用控制器,使用成对排序器来在从第三多个候选语音识别结果中选择的多对候选语音识别结果之间使用成对排序过程、基于排序分数来标识排序最高的候选语音识别结果,每对候选语音识别结果的排序还包括:

利用控制器,使用成对排序器估计第三多个候选语音识别结果中的第一候选语音识别结果的第一单词错误率;

利用控制器,使用成对排序器估计第三多个候选语音识别结果中的第二候选语音识别结果的第二单词错误率;

利用控制器,响应于第一单词错误率小于第二单词错误率而增加与第一候选语音识别结果相关联的排序分数;以及

利用控制器,响应于第一单词错误率大于第二单词错误率而增加与第二候选语音识别

结果相关联的另一排序分数。

5.根据权利要求4所述的方法，还包括：

利用控制器，参考存储在存储器中的多个预定触发对，生成包括对应于至少一个触发对的特征的第一特征向量，所述至少一个触发对包括第一候选语音识别结果内的两个预定触发单词；

利用控制器，参考多个预定触发对，生成包括对应于至少一个触发对的特征的第二特征向量，所述至少一个触发对包括第二候选语音识别结果内的两个预定触发单词；

利用控制器，基于第一特征向量和第二特征向量之间的差异生成第三特征向量；以及

利用控制器，使用成对排序器基于第三特征向量来估计第一候选语音识别结果中的第一单词错误率和第二候选语音识别结果中的第二单词错误率。

6.根据权利要求4所述的方法，还包括：

利用控制器生成包括对应于具有衰减的词袋值的特征的第一特征向量，所述具有衰减的词袋值对应于第一候选语音识别结果中的至少一个单词；

利用控制器生成包括对应于具有衰减的词袋值的特征的第二特征向量、第一特征向量，所述具有衰减的词袋值对应于第二候选语音识别结果中的至少一个单词；

利用控制器，基于第一特征向量和第二特征向量之间的差异生成第三特征向量；以及

利用控制器，使用成对排序器基于第三特征向量来估计第一候选语音识别结果中的第一单词错误率和第二候选语音识别结果中的第二单词错误率。

7.根据权利要求4所述的方法，排序还包括：

利用控制器，响应于第三多个候选语音识别结果中具有最高排序分数的一个候选语音识别结果的置信分数在第三多个候选语音识别结果中具有最高置信分数的另一个候选语音识别结果的预定阈值内，将排序最高的候选语音识别结果标识为具有最高排序分数的所述一个候选语音识别结果；以及

利用控制器，响应于最高置信分数大于具有最高排序分数的所述一个候选语音识别结果的置信分数多于预定阈值，将第三多个候选语音识别结果中具有最高置信分数的所述另一个候选语音识别结果标识为排序最高的候选语音识别结果。

8.根据权利要求4所述的方法，排序还包括：

利用控制器，将排序最高的候选语音识别结果标识为第三多个候选语音识别结果中具有最高排序分数的一个候选语音识别结果。

9.根据权利要求1所述的方法，排序还包括：

利用控制器，使用成对排序器对第一多个候选语音识别结果和第三多个候选语音识别结果进行排序，以标识排序最高的候选语音识别结果。

10.一种具有语音输入控制的自动化系统，包括：

音频输入设备，其被配置为生成对应于来自用户的语音输入的音频输入数据；以及

控制器，其操作性地连接到音频输入设备和存储器，控制器被配置为：

从音频输入设备接收音频输入数据；

使用第一通用语音识别引擎生成对应于音频输入数据的第一多个候选语音识别结果；

使用第一特定于域的语音识别引擎生成对应于音频输入数据的第二多个候选语音识别结果；

生成第三多个候选语音识别结果，第三多个候选语音识别结果中的每个候选语音识别结果包括第一多个候选语音识别结果中的一个中包括的多个单词和第二多个候选语音识别结果中的另一个中包括的至少一个单词；

使用成对排序器对至少第三多个候选语音识别结果进行排序，以标识排序最高的候选语音识别结果；以及

使用排序最高的候选语音识别结果作为来自用户的输入来操作所述自动化系统。

11.根据权利要求10所述的自动化系统，控制器还被配置为：

标识第一多个候选语音识别结果中的第一候选语音识别结果的第一多个单词中的第一单词，所述第一单词对应于第二多个候选语音识别结果中的第二候选语音识别结果中的第二多个单词中的第二单词，第二单词与第一单词不同；以及

生成针对第三多个候选语音识别结果的候选语音识别结果，所述候选语音识别结果包括来自第一候选语音识别结果的第一多个单词与来自第二候选语音识别结果的代替来自第一候选语音识别结果的第一单词的第二单词。

12.根据权利要求11所述的自动化系统，控制器还被配置为：

基于第二多个单词中也存在于第一多个单词中的至少一个单词的位置，将第二候选语音识别结果中的第二多个单词与第一候选语音识别结果中的第一多个单词对齐；以及

标识第一多个候选语音识别中的第一候选语音识别结果的第一多个单词中的第一单词，所述第一单词在与第二多个单词对齐的第一多个单词中的单词位置处对应于第二候选语音识别结果中的第二多个单词中的第二单词。

13.根据权利要求10所述的自动化系统，控制器还被配置为：

使用成对排序器来在从第三多个候选语音识别结果中选择的多对候选语音识别结果之间使用成对排序过程、基于排序分数来标识排序最高的候选语音识别结果，每对候选语音识别结果的排序还包括控制器被配置为：

使用成对排序器估计第三多个候选语音识别结果中的第一候选语音识别结果的第一单词错误率；

使用成对排序器估计第三多个候选语音识别结果中的第二候选语音识别结果的第二单词错误率；

利用控制器，响应于第一单词错误率小于第二单词错误率而增加与第一候选语音识别结果相关联的排序分数；以及

响应于第一单词错误率大于第二单词错误率而增加与第二候选语音识别结果相关联的另一排序分数。

14.根据权利要求13所述的自动化系统，控制器还被配置为：

参考存储在存储器中的多个预定触发对，生成包括对应于至少一个触发对的特征的第一特征向量，所述至少一个触发对包括第一候选语音识别结果内的两个预定触发单词；

参考多个预定触发对，生成包括对应于至少一个触发对的特征的第二特征向量，所述至少一个触发对包括第二候选语音识别结果内的两个预定触发单词；

基于第一特征向量和第二特征向量之间的差异生成第三特征向量；以及

使用成对排序器基于第三特征向量，来估计第一候选语音识别结果中的第一单词错误率和第二候选语音识别结果中的第二单词错误率。

15.根据权利要求13所述的自动化系统，控制器还被配置为：

生成包括对应于具有衰减的词袋值的特征的第一特征向量，所述具有衰减的词袋值对应于第一候选语音识别结果中的至少一个单词；

生成包括对应于具有衰减的词袋值的特征的第二特征向量、第一特征向量，所述具有衰减的词袋值对应于第二候选语音识别结果中的至少一个单词；

基于第一特征向量和第二特征向量之间的差异生成第三特征向量；以及

使用成对排序器基于第三特征向量，来估计第一候选语音识别结果中的第一单词错误率和第二候选语音识别结果中的第二单词错误率。

16.根据权利要求13所述的自动化系统，控制器还被配置为：

响应于第三多个候选语音识别结果中具有最高排序分数的一个候选语音识别结果的置信分数在第三多个候选语音识别结果中具有最高置信分数的另一个候选语音识别结果的预定阈值内，将排序最高的候选语音识别结果标识为具有最高排序分数的所述一个候选语音识别结果；以及

响应于最高置信分数大于具有最高排序分数的所述一个候选语音识别结果的置信分数多于预定阈值，将第三多个候选语音识别结果中具有最高置信分数的所述另一个候选语音识别结果标识为排序最高的候选语音识别结果。

17.根据权利要求13所述的自动化系统，控制器还被配置为：

将排序最高的候选语音识别结果标识为第三多个语音识别结果中具有最高排序分数的一个候选语音识别结果。

18.根据权利要求10所述的自动化系统，控制器还被配置为：

使用成对排序器对第一多个候选语音识别结果和第三多个候选语音识别结果进行排序，以标识排序最高的候选语音识别结果。

用于语音识别的系统和方法

技术领域

[0001] 本公开总地涉及自动化语音识别领域，并且更具体地，涉及改进利用多个语音识别引擎的语音识别系统的操作的系统和方法。

背景技术

[0002] 自动化语音识别是在宽范围的应用中实现人机接口(HMI)的重要技术。具体地，语音识别在这样的情形下是有用的：人类用户需要聚焦于执行其中使用诸如鼠标和键盘之类的传统输入设备将是不方便或不实际的任务。例如，车载“信息娱乐”系统、家庭自动系统以及诸如智能电话、平板计算机和可穿戴计算机的小电子移动设备的很多用途可以采用语音识别来接收来自用户的语音命令和其他输入。

[0003] 大多数现有技术的语音识别系统使用经训练的语音识别引擎将来自用户的记录的口头输入转换成适合于在计算机化系统中处理的数字数据。本领域已知的各种语音引擎执行自然语言理解技术以识别用户说出的单词并从单词中提取语义含义来控制计算机化系统的操作。

[0004] 在一些情形下，当用户执行不同任务时，单个语音识别引擎对于识别来自用户的语音而言不一定是最佳的。现有技术解决方案试图组合多个语音识别系统以改进语音识别的准确度，包括基于预定的排序过程从声学模型不同语音识别模型中选择低级输出或从不同语音识别引擎中选择整组输出。然而，从不同语音识别引擎中挑选输出的现有技术通常不适合于在特定任务中使用，在所述特定任务中用户通常采用来自自然语言的一些语音但是将自然语言语音命令与用于特定目的的单词和句子组合。例如，在车载信息娱乐系统中，来自车辆操作者的语音输入可能包括与语音识别引擎未很好识别的特定单词和短语组合的诸如英语或中文的自然语言，并且仅选择每个高概率包括错误的不同语音识别引擎的输出不会增大语音识别的整体准确度。此外，仅组合诸如声学模型输出之类的低级输出或来自多个语音识别引擎的其他低级特征的现有语音识别系统不能使用较高级语言学特征来评估不同语音识别引擎的输出。因此，对自动化系统的操作的改进以增大使用多个语音识别引擎的语音识别的准确度将是有益的。

发明内容

[0005] 在一个实施例中，已经开发了一种用于使用混合语音识别结果来执行语音识别的方法。所述方法包括：利用音频输入设备生成对应于来自用户的语音输入的音频输入数据；利用控制器，使用第一通用语音识别引擎生成对应于音频输入数据的第一多个候选语音识别结果；利用控制器，使用第一特定于域的语音识别引擎生成对应于音频输入数据的第二多个候选语音识别结果；利用控制器生成第三多个候选语音识别结果，第三多个候选语音识别结果中的每个候选语音识别结果包括第一多个候选语音识别结果中的一个中包括的多个单词和第二多个候选语音识别结果中的另一个中包括的至少一个单词；利用控制器，使用成对排序器对至少第三多个语音识别结果进行排序，以标识排序最高的候选语音识别

结果;以及利用控制器,使用排序最高的候选语音识别结果作为来自用户的输入来操作自动化系统。

[0006] 在另一个实施例中,已经开发了一种使用混合语音识别结果执行语音识别的自动化系统。所述系统包括:音频输入设备,其被配置为生成对应于来自用户的语音输入的音频输入数据;以及控制器,其操作性地连接到音频输入设备和存储器。控制器被配置为:从音频输入设备接收音频输入数据;使用第一通用语音识别引擎生成对应于音频输入数据的第一多个候选语音识别结果;使用第一特定于域的语音识别引擎生成对应于音频输入数据的第二多个候选语音识别结果;生成第三多个候选语音识别结果,第三多个候选语音识别结果中的每个候选语音识别结果包括第一多个候选语音识别结果中的一个中包括的多个单词和第二多个候选语音识别结果中的另一个中包括的至少一个单词;使用成对排序器对至少第三多个候选语音识别结果进行排序,以标识排序最高的候选语音识别结果;以及使用排序最高的候选语音识别结果作为来自用户的输入来操作所述自动化系统。

附图说明

[0007] 图1是如在车辆的乘客室中的车载信息系统中所体现的计算机化系统的组件的示意性视图,所述计算机化系统接收来自用户的语音输入命令。

[0008] 图 2 是用于使用多个语音识别引擎和成对排序器执行语音识别的过程的框图。

[0009] 图3是用于基于语音识别结果中的触发单词序列生成特征向量的过程的框图。

[0010] 图4是描绘针对单个语音输入的两个不同语音识别结果和组合来自两个语音识别结果的单词的混合语音识别结果的图。

具体实施方式

[0011] 出于促进对本文公开的实施例的原理的理解的目的,现在参考附图和以下撰写的说明书中的描述。参考没有对本主题范围的限制的意图。如本公开所涉及领域的技术人员通常会想到的,本公开还包括对说明的实施例的任何更改和修改,并且包括所公开的实施例的原理的另外应用。

[0012] 如本文所使用的,术语“语音识别引擎”指代数据模型和可执行程序代码,所述数据模型和可执行程序代码使计算机化系统能够基于经由麦克风或其他音频输入设备接收的口头单词的记录音频输入数据来标识来自操作者的口头单词。语音识别系统通常包括识别声音记录中人类语音的各个声音的较低级声学模型,以及基于来自用于预定语言的声学模型的声音序列来识别单词和句子的较高级语言模型。本领域已知的语音识别引擎典型地实现一个或多个统计模型,诸如例如隐马尔可夫模型(HMM)、支持向量机(SVM)、经训练的神经网络或使用多个经训练参数生成针对记录的人类语音的统计预测的另一统计模型,所述多个经训练参数被应用于对应于人类语音的输入数据的特征向量。语音识别引擎使用例如本领域已知的提取所记录的语音信号的属性(“特征”)并将特征组织成一维或多维向量的各种信号处理技术来生成特征向量,所述一维或多维向量可以使用统计模型来处理以标识包括各个单词和句子的语音的各种部分。语音识别引擎可以产生针对对应于各个口头音素和更复杂的声音模式的语音输入的结果,所述更复杂的声音模式包括口头单词和包括相关单词序列的句子。

[0013] 如本文所使用的，术语“语音识别结果”指代语音识别引擎针对给定输入生成的机器可读输出。结果可以例如是以机器可读格式编码的文本或充当控制自动化系统的操作的输入的另一组编码数据。归因于语音识别引擎的统计特性，在一些配置中，语音引擎针对单个输入生成多个潜在的语音识别结果。语音引擎还针对每个语音识别结果生成“置信分数”，其中置信分数是基于语音识别引擎的经训练统计模型对每个语音识别结果为准确的可能性的统计估计。如下面更详细描述的，混合语音识别系统使用由多个语音识别引擎产生的语音识别结果，生成附加的混合语音识别结果，并最终基于多个先前生成的语音识别结果产生至少一个输出语音识别结果。如本文所使用的，术语“候选语音识别结果”或更简单地“候选结果”指代作为来自混合语音识别系统的最终语音识别结果的候选的语音识别结果，所述混合语音识别系统产生多个候选结果并且仅选择结果的子集(典型为一个)作为最终语音识别结果。在各种实施例中，候选语音识别结果包括来自通用和特定于域的语音识别引擎的语音识别结果以及混合语音识别结果这两者，系统100使用来自多个候选语音识别结果的单词生成所述混合语音识别结果。

[0014] 如本文所使用的，术语“通用语音识别引擎”指代一种类型的语音识别引擎，其被训练为识别来自诸如英语或中文的自然人类语言的宽广范围的语音。通用语音识别引擎基于经训练单词的宽广词汇和对应于自然语言中广泛使用的语音模式的经训练的语法模型来生成语音识别结果。如本文所使用的，术语“特定于域的语音识别引擎”指代一种类型的语音识别引擎，其被训练为识别特定使用区域或“域”中的语音输入，所述特定使用区域或“域”通常包括与较宽广的自然语言稍微不同的词汇和潜在不同的预期语法结构。用于特定域的词汇典型地包括来自较宽广自然语言的一些术语，但可能包括较窄的整体词汇，并且在一些实例中包括未被官方识别为自然语言中的官方单词但是对于特定域是众所周知的专门化术语。例如，在导航应用中，特定于域的语音识别可以识别针对道路、城镇或其他地理名的术语，这些术语在更通用语言中典型地不被识别为专有名称。在其他配置中，特定域使用对特定域有用但可能未在较宽广语言中很好地识别的一组特定行话。例如，飞行员官方使用英语作为用于交流的语言，但也采用不是标准英语的部分的大量特定于域的行话单词和其他缩写。

[0015] 如本文所使用的，术语“触发对”指代两个单词，其中每个单词可以是单词(例如，“播放”)或表示落入预定类(例如，<歌曲名称>)的单词序列(例如，“Poker Face”)的预定类，所述预定类诸如歌曲的专有名称、人、位置名称等。触发对中的单词当在语音识别结果的句子文本内容中的单词内以特定顺序出现时，对于A→B的触发对，在音频输入数据中观察到早先单词A的情形下，在稍后单词B的出现之间具有高水平的相关性。在很多实例中，触发对包括两个单词在触发对中，虽然触发对可以包括具有多于两个单词的序列。如下面更详细描述的，在经由训练过程标识一组触发对之后，在候选语音识别结果的文本中触发单词对的发生形成针对每个候选结果的特征向量的部分，排序过程使用所述特征向量的部分来对不同的候选语音识别结果进行排序。

[0016] 图1描绘了车载信息系统100，其包括平视显示器(HUD)120、一个或多个控制台LCD面板124、一个或多个输入麦克风128、以及一个或多个输出扬声器132。LCD显示器124和HUD 120至少部分地基于系统100从车辆的操作者或其他乘员接收的语音输入命令而从系统100生成视觉输出响应。控制器148操作性地连接到车载信息系统100中的每个组件。在一些实

施例中，控制器148连接到或并入诸如全球定位系统(GPS)接收器152和无线网络设备154之类的附加组件，以提供导航以及与外部数据网络和计算设备的通信。

[0017] 在一些操作模式中，车载信息系统100独立操作，而在其他操作模式中，车载信息系统100与诸如智能电话170、平板计算机、笔记本计算机或其他电子设备之类的移动电子设备交互。车载信息系统使用诸如USB的有线接口或诸如蓝牙的无线接口与智能电话170通信。车载信息系统100提供语音识别用户接口，所述语音识别用户接口使操作者能够在操作车辆时使用减少分心的口头命令来控制智能电话170或另一移动电子通信设备。例如，车载信息系统100提供语音接口以使车辆操作者能够利用智能电话170进行电话呼叫或发送文本消息，而不要求操作者握住或看向智能电话170。在一些实施例中，智能电话170包括诸如GPS和无线联网设备的各种设备，所述设备补充或代替容纳在车辆中的设备的功能性。

[0018] 麦克风128根据从车辆操作者或另一车辆乘客接收的口头输入生成音频数据。控制器148包括处理音频数据的硬件(诸如DSP)，以及将来自麦克风128的输入信号转换成音频输入数据的软件组件。如下面阐述的，控制器148使用至少一个通用和至少一个特定于域的语音识别引擎来基于音频输入数据生成候选语音识别结果，并且控制器148还使用成对排序器来改进最终语音识别结果输出的准确度。此外，控制器148包括使得能够通过扬声器132生成合成语音或其他音频输出的硬件和软件组件。

[0019] 车载信息系统100使用LCD面板124、投影到挡风玻璃102上的HUD 120并通过定位在仪表板108中的仪表、指示灯或附加LCD面板，来向车辆操作者提供视觉反馈。当车辆在运动中时，控制器148可选地停用LCD面板124或仅通过LCD面板124显示简化的输出以减少对车辆操作者的分心。控制器148使用HUD 120显示视觉反馈，以使操作者能够在接收视觉反馈的同时查看车辆周围的环境。控制器148典型地在HUD 120上在对应于车辆操作者的周界视力的区中显示简化数据，以确保车辆操作者具有车辆周围的道路和环境的无障碍视野。

[0020] 如上面所描述的，HUD 120在挡风玻璃120的部分上显示视觉信息。如本文所使用的，术语“HUD”一般指代宽范围的平视显示设备，包括但不限于包括单独的组合器元件的组合平视显示器(CHUD)等。在一些实施例中，HUD 120显示单色文本和图形，而其他HUD实施例包括多色显示。虽然HUD 120被描绘为在挡风玻璃102上显示，但是在可替换实施例中，平视单元与操作者在操作期间穿戴的眼镜、头盔遮阳板或分划板集成在一起。

[0021] 控制器148包括被配置为中央处理单元(CPU)、微控制器、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、数字信号处理器(DSP)或任何其他合适的数字逻辑设备的一个或多个集成电路。控制器148还包括存储用于车载信息系统100的操作的编程指令的存储器，诸如固态或磁性数据存储设备。

[0022] 在操作期间，车载信息系统100从多个输入设备接收输入请求，所述输入请求包括通过麦克风128接收的语音输入命令。具体地，控制器148经由麦克风128接收对应于来自用户的语音的音频输入数据。

[0023] 控制器148包括被配置为中央处理单元(CPU)、微控制器、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、数字信号处理器(DSP)或任何其他合适的数字逻辑设备的一个或多个集成电路。控制器148还操作性地连接到存储用于车载信息系统100的操作的编程指令的存储器160，诸如固态或磁性数据存储设备。存储器160存储模型数据和可执行程序指令代码以实现至少一个通用语音识别引擎和至少一个特定于域的语音识别引擎162、混

合语音识别结果生成引擎163、对来自语音识别引擎162的候选语音识别结果和来自混合语音识别结果生成引擎163的候选混合语音识别结果进行排序的成对排序器164、以及成对排序器164作为排序过程的部分使用的多个预定触发对166。使用预定训练过程训练语音识别引擎162，并且语音识别引擎162以其他方式在本领域已知。尽管图1的实施例包括存储在机动车辆内的系统100的存储器160内的元件，但是在一些实施例中，外部计算设备(诸如网络连接的服务器)实现系统100中描绘的一些或所有特征。因此，本领域技术人员将认识到，在系统100的可替换实施例中，对包括控制器148和存储器160的系统100的操作的任何引用还应当包括服务器计算设备和其他分布式计算组件的操作。

[0024] 在图1的实施例中，混合语音识别结果生成引擎163生成附加语音识别，所述附加语音识别包括来自语音识别引擎162在系统100的操作期间产生的至少两组不同语音识别结果的单词。如下面更详细描述的，混合语音识别结果生成引擎163将来自通用语音识别引擎的语音识别结果的单词与来自特定于域的语音识别结果的所选择单词组合，以产生不是由任何个体语音识别引擎162产生的新的语音识别结果。如本文所使用的，术语“混合”语音识别结果指代包括来自由通用和特定于域的语音识别引擎162产生的至少两个语音识别结果的单词的语音识别结果。混合语音识别结果生成引擎163不是传统的语音识别引擎。取而代之的是，混合语音识别结果生成引擎163使用语言模型来标识来自特定于域的语音识别结果在特定于域的语音识别域中语言学上重要的单词，并使用来自特定于域的语音识别结果的单词来代替来自通用语音识别引擎的语音识别结果中的单词。混合语音识别结果生成引擎163还将针对每个混合语音识别结果的置信分数生成为针对形成混合结果的每个原始语音识别结果的来自语音识别引擎162的置信分数的平均。

[0025] 成对排序器164是使用用于训练语音识别引擎162的相同训练数据组来训练的随机森林成对排序器。然而，成对排序器164不是传统的语音识别引擎。取而代之的是，成对排序器被训练为使用成对排序过程对语音识别引擎162的候选语音识别结果和来自混合语音识别结果生成引擎163的候选混合语音识别结果进行排序，所述成对排序过程在一对输入语音识别结果中选择具有最低估计单词错误率的一个语音识别结果作为针对每对语音识别结果组合的“获胜者”。在训练过程期间，成对排序器164被训练为基于对应于每个候选语音识别结果的特征向量输入对语音识别结果进行排序以估计单词错误率，其中对于给定对，具有最低估计单词错误率的语音识别输入为“获胜者”。使用不同的语音识别结果训练成对排序器，所述不同的语音识别结果使用具有预定正确值作为基线的训练输入来做出关于来自多个语音识别引擎162的语音识别结果的准确度的估计。在一些实施例中，还使用来自语音识别结果的附加数据来训练成对排序器164，所述附加数据诸如标识预定触发对166的特征向量和每个语音识别引擎162利用语音识别结果产生的置信分数。此外，如下面描述的，控制器148生成混合语音识别结果，所述混合语音识别结果用来自特定于域的语音识别引擎的结果的单词代替通用语音识别引擎的结果中的所选择单词，以产生成对排序器164用作输入的多个混合语音识别结果。

[0026] 例如，给定针对两个候选语音识别结果*h1*和*h2*生成的特征向量作为输入，如果针对*h1*的特征向量输入具有比*h2*更低的估计单词错误率(这指示*h1*比*h2*“更好”)，则控制器148执行成对排序器164生成意味着*h1*获胜的第一“正”输出。否则，成对排序器164生成第二“负”输出以指示*h2*的估计单词错误率低于*h1*。在处理每对候选语音识别结果之后，系统100

将来自成对排序器164的具有最大获胜数的候选语音识别结果标识为排序最高的候选语音识别结果。例如，对于假设列表“*h1*，*h2*，*h3*”，如果*h2*在假设对(*h1*，*h2*)中获胜，*h1*在(*h1*，*h3*)中获胜并且*h2*在(*h2*，*h3*)中获胜，则*h1*、*h2*、*h3*分别获胜1次、2次和0次。由于*h2*获胜最多次数，因此系统100将*h2*标识为排序最高的候选语音识别结果。成对排序器164的可替换实施例使用其他分类技术而不是随机森林方案来对候选语音识别结果进行排序。在一些实施例中，附加于触发对相关特征，还使用诸如置信分数相关特征和“具有衰减的词袋”相关特征之类的其他分类特征来训练成对排序器164。使用某种方案基于候选假设的句子级置信分数来计算置信分数相关特征。在成对排序器164的步骤1中生成的候选句子假设列表中，作为来自语音识别引擎的原始识别结果的那些假设具有每个语音识别引擎162利用语音识别结果产生的句子级置信分数、以及针对混合语音识别结果的置信分数。基于候选假设的文本内容(即，单词序列)计算“具有衰减的词袋”相关特征。

[0027] 在系统100中，触发对166每个包括预定的一组两个或更多个单词，所述预定的一组两个或更多个单词先前已被标识为在来自表示预期语音输入的结构的训练语料库的语音输入序列中具有强相关性。在语音输入中，第一触发单词具有被触发对中的第二触发单词跟随的强统计可能性，虽然在不同语音输入中单词可能被不确定数量的中间单词分离。因此，如果语音识别结果包括触发单词，则归因于第一和第二触发单词之间的统计相关性，语音识别结果中的那些触发单词是准确的可能性相当高。在系统100中，使用本领域已知的统计方法基于互信息分数生成触发单词166。存储器160基于具有高互信息分数的触发单词组，在特征向量中存储预定的一组N个触发对元素，所述预定的一组N个触发对元素对应于在触发单词序列中的第一单词和一个或多个后续单词之间具有高相关性水平的触发对。如下面描述的，触发单词序列166向成对排序器164提供语音识别结果的附加特征，所述附加特征使成对排序器164能够使用超出语音识别结果中存在的单词的语音识别结果的附加特征对语音识别结果进行排序。

[0028] 如下面以附加细节描述的，系统100使用麦克风128接收音频输入数据，并使用多个语音引擎162来生成多个语音识别结果。控制器148还将来自特定于域的语音识别引擎结果的所选择术语与来自通用语音引擎的语音引擎结果组合以生成混合语音识别结果。控制器148使用成对排序器164对混合语音识别结果进行排序，并使用排序最高的结果来控制车载信息系统100或可替换实施例中的任何其他自动化系统的操作。作为排序过程的部分，成对排序器164标识语音识别结果中预定触发对166的发生，并基于所标识的触发对生成特征向量，以向成对排序器164提供附加的高级语言学信息。

[0029] 虽然图1描绘了车载信息系统100作为执行语音识别以接收和执行来自用户的命令的自动化系统的说明性示例，但可以在其他上下文中实现类似的语音识别过程。例如，诸如智能电话170或其他合适设备的移动电子设备典型地包括可以实现语音识别引擎的处理器和一个或多个麦克风、成对排序器、存储的触发对以及实现语音识别和控制系统的其他组件。在另一个实施例中，家庭自动系统使用至少一个计算设备来控制房屋中的HVAC和装置，所述计算设备接收来自用户的语音输入并使用多个语音识别引擎执行语音识别来控制房屋中的各种自动化系统的操作。在每个实施例中，所述系统可选地被配置为使用针对不同自动化系统的特定应用和操作而定制的不同组特定于域的语音识别引擎。

[0030] 图2描绘了用于使用多个语音识别引擎和成对排序器来执行语音识别的过程200。

在下面的描述中，对过程200执行功能或动作的引用指代控制器执行存储的程序指令以使用经由语音识别接口接收命令输入的自动化系统的一个或多个组件来实现所述功能或动作的操作。出于说明性目的，结合图1的系统100来描述过程200。

[0031] 过程200以系统100从用户接收音频输入数据而开始(块204)。诸如麦克风128的音频输入设备生成对应于来自用户的语音输入的音频输入数据。控制器148接收以数字格式的音频输入数据，并且可选地执行滤波或其他数字信号处理操作以从音频输入数据中移除噪声。

[0032] 过程200以系统100基于音频输入数据使用第一通用语音识别引擎生成对应于音频输入数据的第一多个候选语音识别结果而继续(块208)。系统100还使用至少一个特定于域的语音识别引擎生成第二多个候选语音识别结果(块212)。在系统100中，控制器148使用一个或多个通用语音识别引擎162来生成第一多个结果，并使用一个或多个特定于域的语音识别引擎162来生成第二多个候选结果。在一些实施例中，控制器148从每个语音识别引擎中选择每个语音识别引擎指示具有最高置信分数值的预定数量的语音识别结果，以形成各自多个语音识别结果。例如，在一个实施例中，控制器148生成五个候选语音识别结果，其具有来自每个语音识别引擎162的最高置信分数值。在包括并行处理硬件(诸如多个处理器核)的控制器148的实施例中，第一、第二多个候选语音识别结果的生成以任何顺序发生或并发发生。

[0033] 过程200以控制器148基于第一多个候选语音识别结果和第二多个候选语音识别结果生成第三多个候选混合语音识别结果而继续(块216)。第三多个候选语音识别结果也被称为混合语音识别结果，这是因为这些结果组合来自由两个或更多个语音识别引擎产生的语音识别结果的单词。控制器148生成第三多个候选语音识别结果中的每个语音识别结果，每个语音识别结果包括来自通用语音识别的第一多个候选语音识别结果中的一个中包括的多个单词和来自特定于域的语音识别引擎的第二多个候选语音识别结果中的另一个中包括的至少一个单词。控制器148标识两个语音引擎的候选语音识别结果中的共同单词，并用来自特定于域的语音引擎结果与通用语音识别结果不同的对应单词来替换来自通用语音引擎结果的单词。

[0034] 为了生成第三多个语音识别结果中的每个候选混合语音识别结果，控制器148首先使用本领域已知的技术(诸如使用动态编程过程)来对齐每个语音识别结果中的共同单词，来以最小化单词序列之间差异的“最小编辑距离”对齐单词。然后，控制器148从特定于域的语音识别引擎的候选语音识别结果中选择与第一语音识别结果中的不同单词对齐的单词，并将那些单词替换到通用语音识别引擎的候选语音识别结果中，以产生第三多个候选语音识别结果中的混合候选语音识别结果。如果来自特定于域的语音识别引擎的候选语音识别结果包括未在通用语音识别引擎的对应候选语音识别结果中出现的特定于域的单词t_1，t_2……t_k，则控制器148将来自特定于域的语音识别引擎的各个单词的排列和各个单词的组合替换到来自通用语音识别引擎的候选语音识别结果中，以产生包括来自特定于域的语音识别引擎的不同单词的排列的多个候选混合语音识别结果。

[0035] 图4描绘了两个候选语音识别结果404和408以及混合候选语音识别结果450的示例。在图4的示例中，语音输入基于从用户到车载导航系统的导航请求。通用语音识别引擎基于一般英语语言模型生成语音识别结果404。特定于域的语音识别引擎聚焦于导航，并包

括针对道路和位置名称的附加术语，所述附加术语包括不是官方英语语言词汇的部分的术语。在系统100中，控制器148执行混合语音识别结果生成引擎163中的程序代码，以使用来自由通用和特定于域的语音识别引擎162生成的至少两个语音识别结果的单词来生成混合语音识别结果。每个语音识别结果包括形成句子的单词序列。控制器148基于对两个序列共同的单词来对齐两个语音识别结果404和408，所述共同的单词诸如示例单词“boulevard(林荫大道)”，其在第一语音识别结果中被示出为共同单词406并在第二语音识别结果中被示出为410。第一语音识别结果404中的单词“mope(忧郁)”420与语音识别结果408中的单词“Mopac(莫帕克)”424对齐。然后，控制器148标识来自特定于域的语音识别引擎的第二语音识别408中在对齐的语音识别结果404中不存在的单词。在图4的示例中，第二识别结果404中的术语“Mopac”424是特定于域的语音识别引擎在音频输入数据中识别的用于描述德克萨斯州奥斯汀的主要林荫大道的口语术语。然而，通用语音识别引擎将单词Mopac误标识为“mope”，这是因为通用语音识别引擎被训练为识别宽广范围的英语单词。此外，第二语音识别结果408包括较窄的术语组，这是因为特定于域的语音识别引擎不直接识别来自音频输入数据的一些单词。

[0036] 控制器148使用来自第一候选语音识别结果404的单词作为基础并用术语“Mopac”替换到其中以代替单词“mope”以并入来自第二候选语音识别结果408的不同术语，来生成混合候选语音识别输出450。控制器148可选地用来自特定于域的语音识别引擎的不同单词代替来自通用语音识别引擎的语音识别结果的多个术语，以形成混合语音识别结果。在过程200期间，控制器148针对多组通用语音识别结果和特定于域的语音识别结果执行上面描述的过程，以生成包括来自一个或多个通用语音识别引擎和特定于域的语音识别引擎两者的单词的第三多个混合语音识别结果。

[0037] 在一些实施例中，控制器148仅将在特定于域的语音识别引擎的候选语音识别结果中具有特定语义意义的单词替换到通用语音识别引擎的语音识别结果中。例如，在图4中，特定于域的语音识别引擎162被特别训练为以比通用语音识别引擎更高的准确度识别街道名称和其他地理术语。因此，在图4中，控制器148用“Mopac”替换单词“mope”，这是因为术语“Mopac”在特定于域的语音识别引擎中具有作为道路名称的语义意义。然而，如果特定于域的语音识别引擎针对另一个单词(诸如常用的英语动词或代词)生成不同的结果，则控制器148继续依靠来自通用语音识别引擎的结果，所述通用语音识别引擎可论证地针对更典型的自然语言模式产生更准确的结果。控制器148基于来自每个语音识别引擎162中的语言模型的信息来标识特定单词的语义分类。

[0038] 在过程200期间，控制器148还使用混合语音识别结果生成引擎163来针对每个候选混合语音识别结果产生置信分数。在系统100中，控制器148生成来自通用和特定于域的语音识别引擎162的语音识别结果的置信分数的平均值作为针对混合语音识别结果的置信值。如下面更详细描述的，控制器148使用线性回归过程来标准化两个或更多个不同语音识别引擎的置信分数，并且在一个实施例中，控制器148标准化来自形成混合语音识别结果的基础的原始语音识别结果的置信分数，以生成针对混合语音识别结果的置信分数。在过程200期间，成对排序器164接收针对混合语音识别结果的标准化置信分数作为输入特征向量中的一个特征，以对候选语音识别结果对进行排序。

[0039] 再次参考图2，过程200以以下操作继续：控制器148使用所使用成对排序器164基

于来自语音识别结果的估计单词错误率与可选地语音识别结果中的所标识的单词触发对和词袋特征以及针对语音识别结果的置信分数，来生成针对第三多个候选混合语音结果识别结果的排序分数(块220)。如上面指出的，控制器148使用成对排序器164来使用成对过程对语音识别结果进行排序，在所述成对过程中成对排序器164接收两个语音识别结果并将“获胜”结果标识为该对中具有最低估计单词错误率的语音识别结果。此外，在一些实施例中，附加于第三多个候选混合语音识别结果，系统100还对来自通用语音识别引擎的第一多个候选语音识别结果或来自特定于域的语音识别引擎的第二多个候选语音识别结果中的一些或全部、或者第一和第二多个候选语音识别结果两者进行排序。

[0040] 在系统100中，成对排序器164是随机森林排序系统，其接收两个语音识别结果作为输入，并基于针对每个语音识别结果的估计单词错误率对该对语音识别进行排序，其中较低的估计单词错误率产生较高的排序。在过程200期间，控制器148将每对语音识别结果组合供给到成对排序器164，以确定不同对语音识别结果的相对排序。控制器148增加与利用成对排序器的每次比较时在具有最低估计单词错误率方面“获胜”的第一或第二候选语音识别结果相关联的排序分数。然后，控制器148在成对排序器164标识每对语音识别结果之间的最低单词错误率之后，将最高排序结果标识为具有最高排序分数的候选语音识别结果。

[0041] 在操作期间，成对排序器164接收以预定特征向量格式的语音识别结果，并且成对排序器中的经训练随机森林模型基于每个语音识别结果中的单词、单词的结构、以及产生每个语音识别结果的语音识别引擎的身份来生成每个语音识别结果中的单词错误率的估计。具体地，在使用多于一个通用或专用语音识别引擎的系统100的配置中，经训练随机森林成对排序器可以基于生成每个语音识别结果的语音识别引擎的身份来生成针对单词错误率的不同估计，这是因为例如一些语音识别引擎在标识特定的单词或短语组时更准确。成对排序器164被训练为基于在过程200之前发生的训练过程期间基于一组预定训练数据从每个语音识别引擎162观察到的单词错误率来估计单词错误率。如下面更详细描述的，附加于基于每个结果的实际内容对语音识别结果进行排序，成对排序器164还可选地使用与候选语音识别结果的特征向量相关联的单词触发对、候选语音识别结果置信分数值和具有衰减的词袋特征，来估计最低单词错误率并针对每对候选语音识别结果产生排序。

[0042] 图3描绘了用于对应于在一个或多个语音识别结果中存在的触发对的特征向量的生成的过程300。在下面的描述中，对过程300执行功能或动作的引用指代控制器执行存储的程序指令以使用经由语音识别接口接收命令输入的自动化系统的一个或多个组件来实现所述功能或动作的操作。出于说明性目的，结合图1的系统100和图2的过程200来描述过程300。

[0043] 过程300以以下操作开始：控制器148在对应于一对候选语音识别结果的文本数据中标识包括触发对、置信分数和具有衰减的词袋特征中的至少一个的特征(块304)。例如，通过使用图4的示例语音识别结果450，如果存储在存储器160中的触发单词序列166中的一个包括触发对(“Shops(商店)”，“around(周围)”)，则控制器148将单词“Shops”标识为触发对中的第一触发术语，并解析语音识别结果中的任何后续单词以标识触发对中的附加单词，诸如单词“around”。在一些实例中，控制器148标识单个语音识别结果中的多组触发单词序列。

[0044] 过程300以以下操作继续:控制器148生成特征向量,所述特征向量包括在语音识别结果中标识的触发对、置信分数和具有衰减的词袋特征中每个的值(块308)。控制器148生成具有预定数量的N个元素的特征向量,每个元素对应于存储在存储器160中的N个触发单词序列166中的一个。因此,特征向量中的每个索引在多个语音识别结果之间以一致的方式对应于一个触发短语。在典型的实例中,大多数(有时是全部)触发单词序列不存在于语音识别结果中,并且控制器148可选地将特征向量生成为稀疏向量,所述稀疏向量仅包括针对实际存在于语音识别结果内的触发单词序列的非平凡条目。

[0045] 在成对排序过程中,控制器148抵消均包括相同触发对的两个特征向量中的每个触发对的发生。例如,如果两个候选语音识别结果均包括触发对(“Shops”,“around”),则控制器148从两个候选语音识别结果的特征向量中移除该条目,这是因为触发对在两个候选结果中均发生并在成对比较过程中被有效抵消。然而,如果仅一个候选结果包括触发对,则针对该候选结果的特征向量包括指示触发对仅存在于该候选语音识别结果中的值。然后,控制器148基于两个候选结果的两个特征向量之间的差异,针对成对排序器164生成对应于该对候选语音识别结果的输入特征向量。在过程200期间,控制器148基于第一特征向量和第二特征向量针对每个成对比较生成第三特征向量。第三特征向量形成到成对排序器164的输入,所述输入包括两个原始特征向量的结果之间的编码差异。第三特征向量包括负值和正值,所述负值和正值使成对排序器164能够标识包括每个特征的特定语音识别结果。例如,通过使用对应于触发对、具有衰减的词袋值、置信分数或其他特征的简化特征向量阵列,第一候选语音识别结果包括[0,0,1,1,0,0],第二候选语音识别结果包括[0,1,1,0,0,0],并且控制器148基于从第一特征向量中减去第二特征向量,针对成对排序器164生成最终语音识别结果:[0,-1,0,1,0,0]。在该示例中,负值指示特征仅存在于第二特征向量中或者第二特征向量针对给定特征具有更高数值特征值。在上面的示例中,针对第一和第二特征向量两者的第三索引值是“1”,并且最终特征向量在第三索引中包括平凡值“0”,这是由于两个输入特征向量包括相同的特征,所述相同特征由于该特征在两个候选语音识别结果之间没有区分而被成对排序器164忽略。

[0046] 在一些实施例中,控制器148标识每对候选语音识别结果中每个触发对的发生频率。如果候选语音识别结果每个包括以相同频率发生的相同触发对,则控制器148从两个候选语音识别结果的特征向量中移除该触发对,这是由于该触发对的发生不为成对排序器164提供执行排序过程的附加信息。然而,如果候选语音识别结果中的一个比该对中的另一个语音识别结果更频繁地包括该触发对,则控制器148在针对该对候选语音识别结果生成的最终特征向量中包括频率差异作为对应于触发短语的值。

[0047] 过程300以以下操作结束:如上面参考图2中的块220的处理描述的,控制器148使用对应于触发单词对、置信分数和具有衰减的词袋特征的特征向量数据作为用于对语音识别结果进行排序的过程的部分(块312)。在过程200期间,控制器148执行过程300以生成针对每个语音识别结果的附加特征向量数据。控制器148使用成对排序器164至少部分地基于特征向量来估计每个语音识别结果中的单词错误率,所述特征向量编码关于触发对中的非相邻单词之间的关系的附加信息。特征向量中的附加信息编码关于触发单词序列的较高级语言学特征,所述触发单词序列典型地不包括到成对排序器164的句子中的相邻单词,这改进排序过程的准确度。

[0048] 附加于针对每个候选语音识别结果生成包括触发对元素的特征向量，成对排序器164还可选地将基于候选句子假设所计算的“具有衰减的词袋”特征添加到特征向量。本文使用的术语“具有衰减的词袋”特征指代控制器148基于候选语音识别结果内单词的位置以及单词多么频繁地发生而分配给结果中存在的每个单词的数值分数。控制器148针对候选语音识别结果中存在的每个识别的字典单词生成具有衰减的词袋分数。在系统100中，字典数据与例如语音识别引擎模型数据162相关联地存储在存储器160中。对于预定字典中的给定单词w_i，具有衰减的词袋分数是：$bow_i = \sum_{p \in P'(w_i)} \gamma^p$，其中$P'(w_i)$是候选语音识别结果中单词$w_i$发生所处的一组位置，并且项γ是在(0,1.0)的范围中的预定数值衰减因子，所述预定数值衰减因子例如在系统100的说明性实施例中被设置为0.9。

[0049] 在过程200期间，控制器148生成包括具有衰减的词袋值的特征向量，以增补或代替指示在候选语音识别结果中来自预定字典的每个单词的存在或不存在的特征向量值。以与触发对特征向量值类似的方式，控制器148针对所述对中的每个候选语音识别结果生成各个具有衰减的词袋特征值，并且后续生成两个特征向量之间的差异作为提供给成对排序器164的最终特征向量值。因此，控制器148仅当所述对中的两个语音识别结果针对单词具有不同的具有衰减的词袋分数时才针对该单词生成具有非平凡条目的特征向量，并且针对或是未出现在该对中的两个候选语音识别结果中的每个中或是针对该对中的两个候选语音识别结果具有相同的具有衰减的词袋分数的每个单词包括零值特征向量条目。

[0050] 附加于针对每对候选语音识别结果生成包括触发对元素和具有衰减的词袋特征的特征向量，成对排序器164还可选地将置信分数特征添加为提供给成对排序器的特征向量中的一个附加特征。置信分数特征被计算为该对中两个语音识别结果的置信分数之间的差异。

[0051] 再次参考图2，过程200基于上面描述的特征向量输入，针对第三多个候选混合语音识别结果中的每对结果生成排序分数。在一些配置中，控制器148还针对来自通用语音识别引擎的第一多个语音识别结果和来自特定于域的语音识别引擎的第二多个语音识别结果中的任一个或两者生成排序分数。

[0052] 在图2的实施例中，针对每个语音识别结果生成的排序分数不是控制器148用于标识排序最高的语音识别结果的唯一度量。在排序过程中，控制器148还使用每个语音识别引擎与包括混合语音识别结果的每个语音识别结果相关联地生成的置信分数。在过程200期间，控制器148标识具有最高置信分数的候选语音识别结果，并将最高置信分数与具有最高排序分数的语音识别结果的置信分数进行比较。如果具有最高排序分数的候选语音识别结果的置信分数在最高整体置信分数的预定阈值范围内(块224)，则控制器148选择具有最高排序分数的候选语音识别结果作为从候选语音识别结果中选择的排序最高的输出语音识别结果(块228)。例如，如果置信分数的差异关于在下面更详细描述的标准化置信分数范围的15％内，则控制器148选择具有最高排序分数的语音识别结果作为整体排序最高的语音识别结果。然而，如果另一语音识别结果的最高置信分数超过具有最高排序分数的语音识别结果的置信分数多于预定阈值(块224)，则控制器148选择具有最高置信分数的语音识别结果作为输出语音识别结果(块232)。在块224-232中描绘的处理的可替换实施例中，控制器148选择具有最高排序分数的候选语音识别结果作为最终输出语音识别结果，而不执行附加处理来比较候选语音识别结果的置信分数。

[0053] 当然，在很多实例中，具有最高排序分数的语音识别结果的置信分数也是所有语音识别结果当中的最高置信分数或非常高的置信分数，并且控制器148将具有最高排序分数的语音识别结果标识为排序最高的语音识别结果。然而，在其他情形下，如果具有最高排序分数的语音识别结果具有低得多的置信分数，则控制器148选择具有最高置信分数的语音识别结果。在另一配置中，控制器148将排序分数和置信分数组合成复合分数，以标识排序最高的语音识别结果。例如，在一些情形下，语音识别结果可能具有高排序分数和高置信分数，但可能不具有所有语音识别结果当中的最高排序分数或置信分数。控制器148使用诸如排序分数和置信分数的加权平均或其他组合之类的复合分数将该语音识别结果标识为具有最高排序。

[0054] 如上面所描述的，控制器148部分地基于与每个语音识别结果相关联的置信分数来标识排序最高的语音识别结果。置信分数是语音识别引擎162与语音识别结果相关联地生成的针对每个语音识别结果的准确度(置信度)估计的统计值。然而，针对一个语音识别引擎的数值置信分数范围典型不转化为另一语音识别引擎，这增大比较来自多个语音识别引擎的语音识别结果的置信分数的难度。例如，第一语音识别引擎A在1-100的标度上生成置信分数，而第二语音识别引擎B在1-1000的标度上生成置信分数。然而，仅仅缩放引擎A的数值结果以匹配引擎B中的置信分数范围(或反之亦然)不足以使置信分数可比较。这是因为对应于特定置信分数的实际准确度估计在两个不同的语音识别引擎之间典型是不相同的。例如，对于引擎A的标准化标度上330的任意置信分数可以对应于75％的估计准确度，但是对于引擎B的相同分数可以对应于84％的估计准确度，给定高质量语音识别引擎中预期的准确度水平范围，这可能是大量差异。

[0055] 在系统100中，控制器148使用线性回归过程来标准化不同语音识别引擎之间的置信分数。控制器148首先将置信分数范围细分为预定数量的细分或“箱”，诸如针对两个语音识别引擎A和B为二十个独特箱。然后，控制器148基于观察到的语音识别结果和在过程200之前的训练过程期间使用的实际根本输入，来标识针对各种语音识别结果对应于每个分数箱的实际准确率。控制器148在“边缘”周围的预定数值窗口内执行置信分数的聚类操作，并标识对应于每个边缘置信分数值的平均准确度分数，所述“边缘”分离针对来自不同语音识别引擎的每组结果的箱。“边缘”置信分数沿着每个语音识别引擎的置信分数范围均匀分布，并提供预定数量的比较点以执行线性回归，所述线性回归将第一语音识别引擎的置信分数映射到另一语音识别引擎具有类似准确率的置信分数。控制器148使用针对每个边缘分数所标识的准确度数据来执行线性回归映射，所述线性回归映射使控制器148能够将来自第一语音识别引擎的置信分数转换为对应于来自第二语音识别引擎的等同置信分数的另一置信分数值。来自第一语音识别引擎的一个置信分数到来自另一语音识别的另一置信分数的映射也称为分数对齐过程，并且在一些实施例中，控制器148使用以下等式确定来自第一语音识别引擎的置信分数与第二语音识别引擎的对齐：

[0056] $$x' = e_i' + \frac{(x-e_i)}{(e_{i+1}-e_i)}(e_{i+1}' - e_i')$$

[0057] 其中x是来自第一语音识别引擎的分数，x’是第二语音识别引擎的置信分数范围内x的等同值，值e_i和e_{i+1}对应于针对最接近针对第一语音识别引擎的值$x_$的不同边缘值的估计准确度分数(例如，针对在22的置信分数周围的边缘值20和25的估计准确度分数)，

并且值e_i'和e_{i+1}'对应于针对第二语音识别引擎的相同相对边缘值处的估计准确度分数。

[0058] 在一些实施例中，控制器148在存储器160中将线性回归的结果存储为查找表或其他合适的数据结构，以使得能够高效标准化不同语音识别引擎162之间的置信分数，而不必针对每次比较重新生成线性回归。

[0059] 再次参考图2，过程200以以下操作继续：控制器148使用所选择的排序最高的语音识别结果作为来自用户的输入来控制自动化系统(块236)。在图1的车载信息系统100中，控制器148操作包括例如车辆导航系统的各种系统，所述车辆导航系统使用GPS 152、无线网络设备154和LCD显示器124或HUD 120以响应于来自用户的语音输入而执行车辆导航操作。在另一配置中，控制器148响应于语音命令而通过音频输出设备132播放音乐。在又一配置中，系统100使用智能电话170或另一网络连接设备来基于来自用户的语音输入发出免提电话呼叫或传输文本消息。虽然图1描绘了车载信息系统实施例，但是其他实施例采用使用音频输入数据来控制各种硬件组件和软件应用的操作的自动化系统。

[0060] 将领会，上面公开的特征和功能的变体及其他特征和功能或它们的替换物可以合期望地组合成很多其他不同的系统、应用或方法。本领域技术人员后续可以做出也意图被所附权利要求涵盖的各种目前未预见或未预料到的替换物、修改、变型或改进。

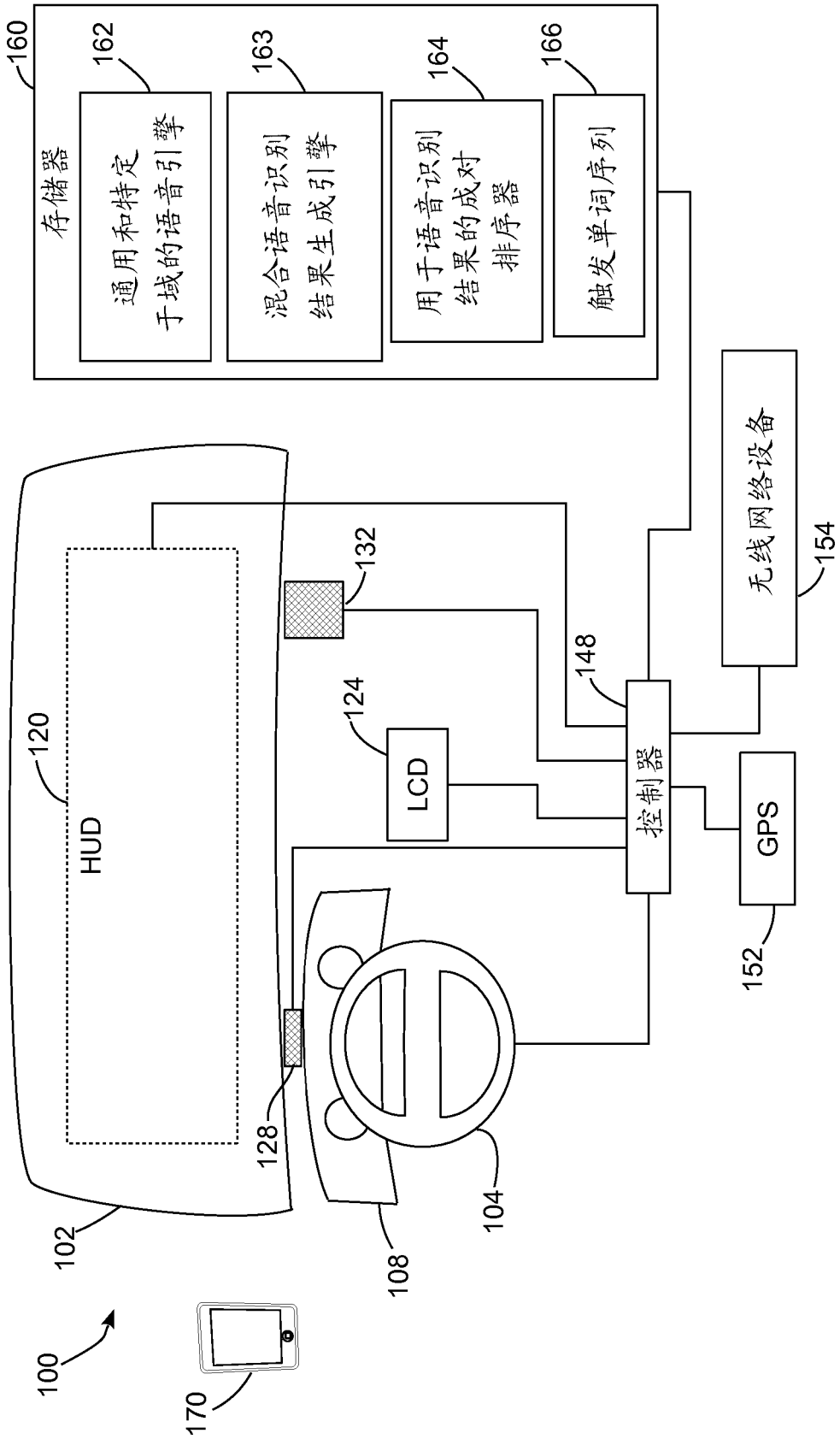

160
存储器
162
通用和特定
于域的语音引擎
163
混合语音识别
结果生成引擎
164
用于语音识别
结果的成对
排序器
166
触发单词序列
无线网络设备
154
132
148
124
LCD
控制器
GPS
152
120
HUD
128
104
102
108
100
170

图 1

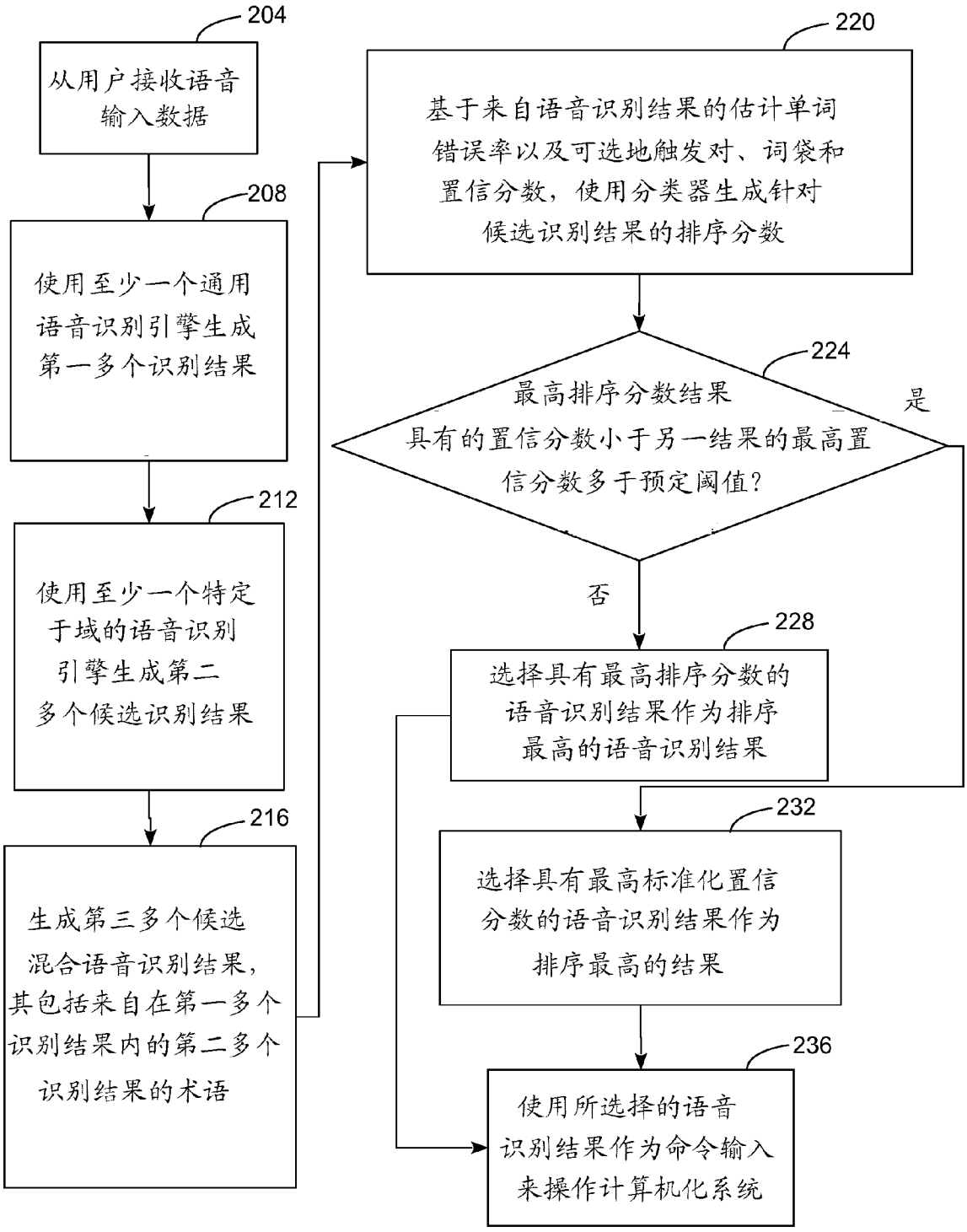

204
从用户接收语音输入数据
208
使用至少一个通用语音识别引擎生成第一多个识别结果
212
使用至少一个特定于域的语音识别引擎生成第二多个候选识别结果
216
生成第三多个候选混合语音识别结果，其包括来自在第一多个识别结果内的第二多个识别结果的术语
220
基于来自语音识别结果的估计单词错误率以及可选地触发对、词袋和置信分数，使用分类器生成针对候选识别结果的排序分数
224
最高排序分数结果具有的置信分数小于另一结果的最高置信分数多于预定阈值？
是
否
228
选择具有最高排序分数的语音识别结果作为排序最高的语音识别结果
232
选择具有最高标准化置信分数的语音识别结果作为排序最高的结果
236
使用所选择的语音识别结果作为命令输入来操作计算机化系统

图 2

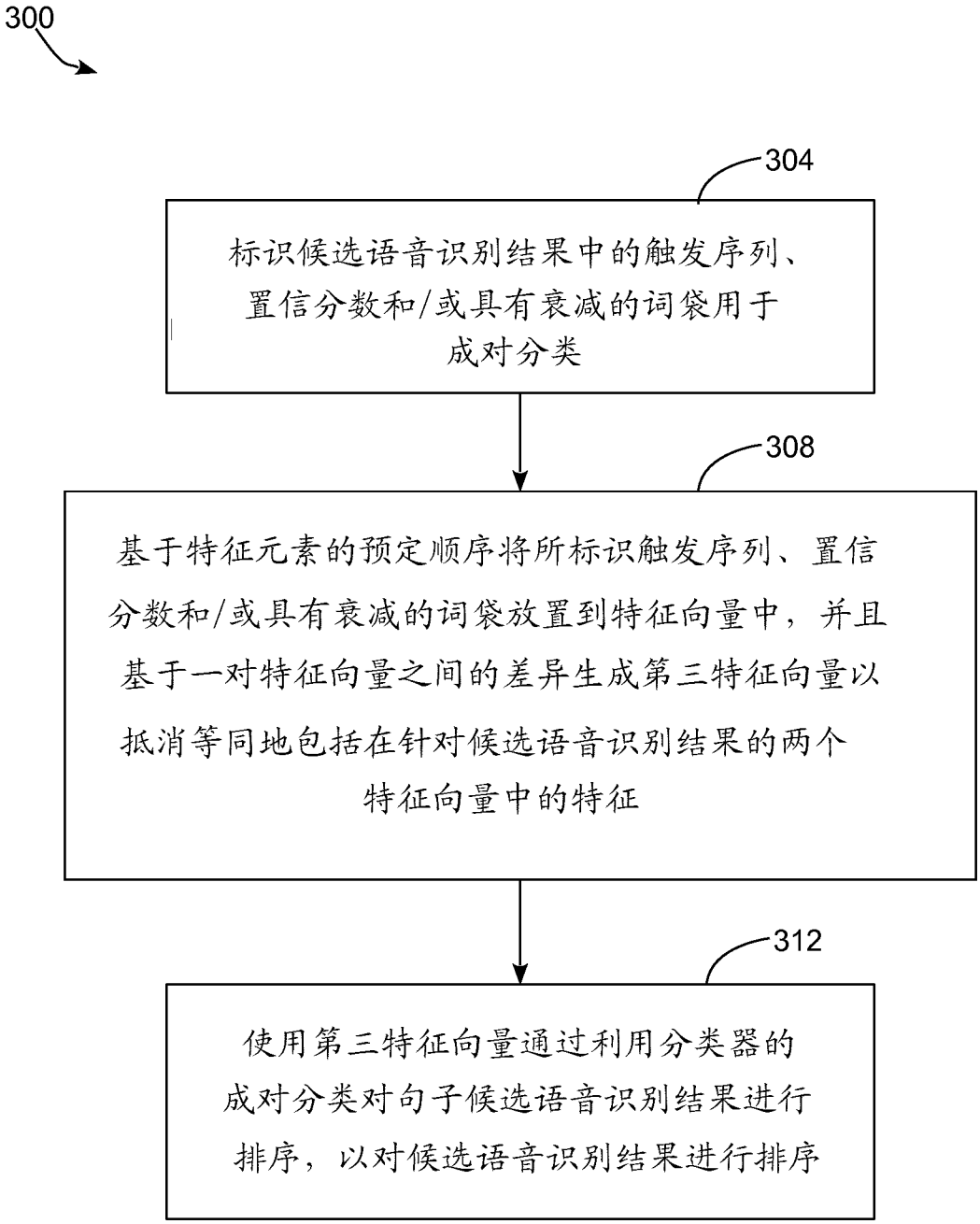

300
304
标识候选语音识别结果中的触发序列、置信分数和/或具有衰减的词袋用于成对分类
308
基于特征元素的预定顺序将所标识触发序列、置信分数和/或具有衰减的词袋放置到特征向量中，并且基于一对特征向量之间的差异生成第三特征向量以抵消等同地包括在针对候选语音识别结果的两个特征向量中的特征
312
使用第三特征向量通过利用分类器的成对分类对句子候选语音识别结果进行排序，以对候选语音识别结果进行排序

图 3

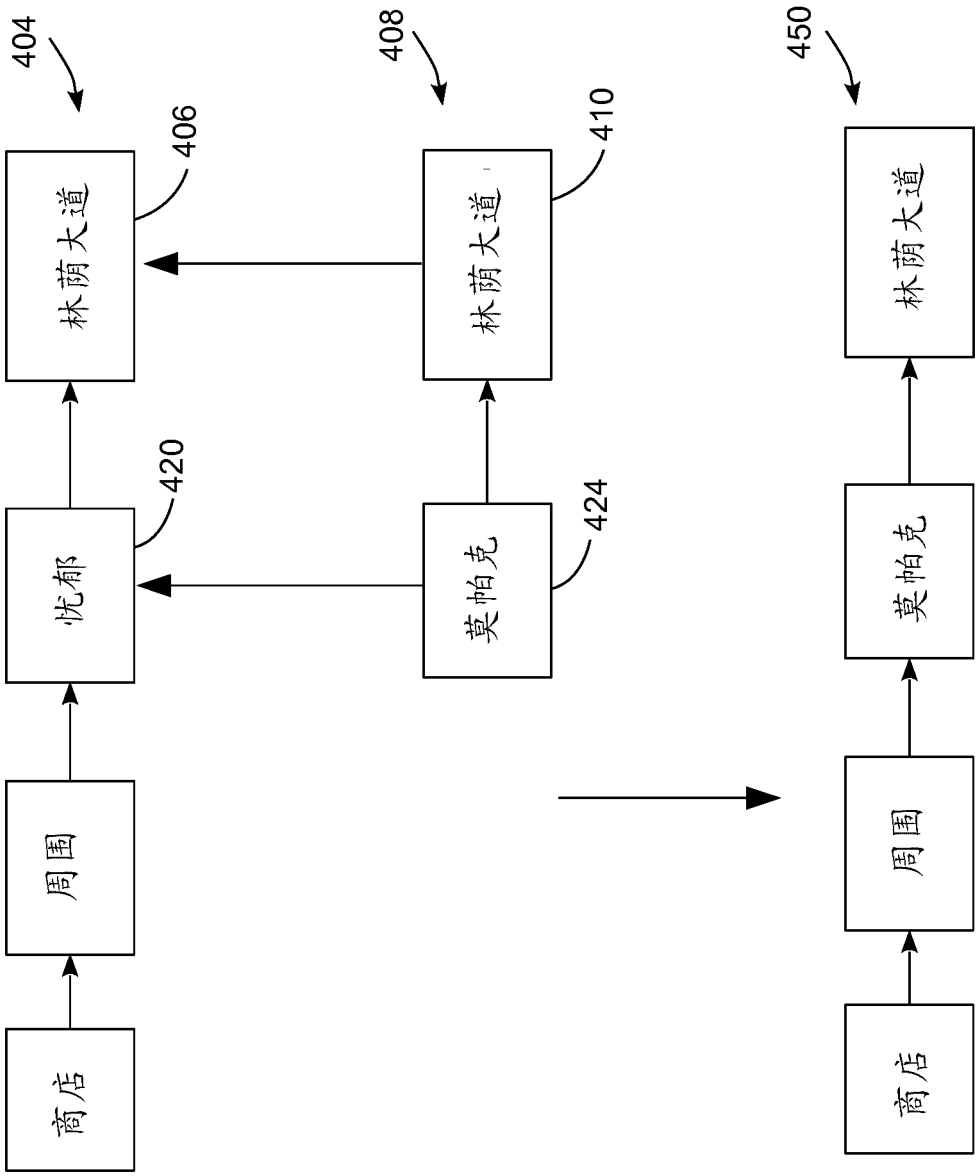


图 4