



(12) 发明专利

(10) 授权公告号 CN 107112010 B

(45) 授权公告日 2021.06.01

(21) 申请号 201580073696.3
 (22) 申请日 2015.01.16
 (65) 同一申请的已公布的文献号
 申请公布号 CN 107112010 A
 (43) 申请公布日 2017.08.29
 (85) PCT国际申请进入国家阶段日
 2017.07.17
 (86) PCT国际申请的申请数据
 PCT/KR2015/000486 2015.01.16
 (87) PCT国际申请的公布数据
 W02016/114428 KO 2016.07.21
 (73) 专利权人 三星电子株式会社
 地址 韩国京畿道
 (72) 发明人 朴致衍 金逸桓 李暎慤 金南勋
 李在原
 (74) 专利代理机构 北京市柳沈律师事务所
 11105
 代理人 邵亚丽
 (51) Int.Cl.
 G10L 15/197 (2013.01) (续)

(56) 对比文件
 US 2011/231183 A1, 2011.09.22
 CN 1223739 A, 1999.07.21
 CN 102023995 A, 2011.04.20
 CN 1409842 A, 2003.04.09
 CN 101558442 A, 2009.10.14
 US 2010/049514 A1, 2010.02.25
 CN 102084417 A, 2011.06.01
 JP 特开2003-202890 A, 2003.07.18
 CN 102968989 A, 2013.03.13
 Oh-Wook Kwon. "Korean large vocabulary continuous speech recognition with morpheme-based recognition units".《Speech Communication》.2003,
 P.Geutner. "ADAPTIVE VOCABULARIES FOR TRANSCRIBING MULTILINGUAL BROADCAST NEWS".《IEEE》.1998,
 Sarah E. Schwarm. "Adaptive Language Modeling With Varied Sources to Cover New Vocabulary Items".《IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING》.2004, 第12卷 (第3期), (续)

审查员 董小东

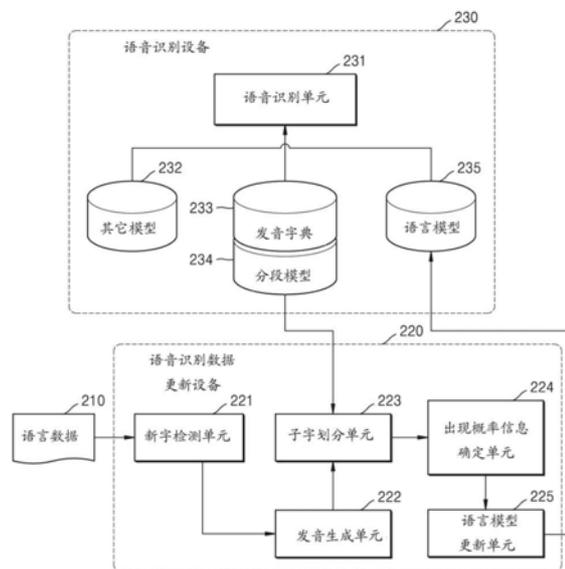
权利要求书1页 说明书40页 附图31页

(54) 发明名称

用于使用语法模型执行语音识别的方法和设备

(57) 摘要

本申请公开了一种更新包括用于语音识别的语言模型的语音识别数据的方法,所述方法包括:获取包括至少一个字的语言数据;从至少一个字当中检测不存在于语言模型中的字;获取关于检测到的字的至少一个音素序列;通过将至少一个音素序列划分成预定单元分量来获取构成至少一个音素序列的分量;确定关于在语音识别期间构成至少一个音素序列中的每一个的各个分量出现的概率的信息;以及基于所确定的概率信息来更新语言模型。



[转续页]

CN 107112010 B

[接上页]

(51) Int.Cl.

G10L 15/28 (2013.01)

(56) 对比文件

MATHIAS CREUTZ. "Morph-Based Speech

Recognition and Modeling of Out-of-Vocabulary Words Across Languages".《ACM Transactions on Speech and Language Processing》.2007,

1. 一种更新包括用于语音识别的语言模型的语音识别数据的方法,所述方法包括:
获取包括至少一个字的语言数据;
从所述至少一个字当中检测不存在于所述语言模型中的字;
获取关于检测到的字的至少一个音素序列;
通过将所述至少一个音素序列划分成预定单元分量来获取构成所述至少一个音素序列的分量;
确定关于代表在所述语音识别期间构成所述至少一个音素序列中的每一个的各个分量出现概率的概率值的信息;以及
基于所确定的关于所述概率值的信息来更新所述语言模型。
2. 如权利要求1所述的方法,其中,所述语言模型包括第一语言模型和包括至少一个语言模型的第二语言模型,并且
所述语言模型的更新包括基于所述所确定的关于所述概率值的信息更新第二语言模型。
3. 如权利要求2所述的方法,进一步包括:
基于包括在所述第二语言模型中的至少一个出现概率信息来更新所述第一语言模型;
以及
基于所述检测到的字的所述至少一个音素序列来更新包括关于字的音素序列的信息的发音字典。
4. 如权利要求1所述的方法,其中,所述出现概率信息包括关于在字或者另一分量在所述分量之前出现的条件下,相应的分量中的每一个的出现概率的信息。
5. 如权利要求1所述的方法,其中,确定所述出现概率信息包括:
获取关于与所述检测到的字相对应的周围情形的情形信息;以及
基于所述情形信息,选择要添加关于所述检测到的字的出现概率信息的语言模型。
6. 如权利要求5所述的方法,其中,所述语言模型的更新包括基于所确定的出现概率信息来更新关于与所述情形信息相对应的模块的第二语言模型。
7. 一种用于在语音识别期间更新包括关于各个字的出现概率信息的语言模型的设备,所述设备包括:
控制器,所述控制器获取包括至少一个字的语言数据、从所述至少一个字当中检测不存在于所述语言模型中的字、获取关于检测到的字的至少一个音素序列、通过将所述至少一个音素序列划分成预定单元分量来获取构成所述至少一个音素序列的分量、确定关于代表在语音识别期间构成所述至少一个音素序列中的每一个的各个分量出现概率的概率值的信息、以及基于所确定的关于所述概率值的信息更新所述语言模型;以及
存储器,所述存储器存储所更新的语言模型。
8. 一种存储用于实现权利要求1-6所述的方法的程序的计算机可读记录介质。

用于使用语法模型执行语音识别的方法和设备

技术领域

[0001] 本发明涉及一种用于使用语言模型来执行语音识别的方法和设备。

背景技术

[0002] 语音识别是一种用于从用户接收语音的输入、将语音自动地转换成文本、并且识别该文本的技术。最近,语音识别被用作一种用于替代用于智能电话或者TV的键盘输入的接口技术。

[0003] 语音识别系统可以包括用于接收语音信号的客户端和用于从语音信号中识别语音的自动语音识别(automatic speech recognition,ASR)引擎,其中客户端和ASR引擎可以被独立地设计。

[0004] 一般地,语音识别系统可以通过使用声学模型、语言模型和发音字典来执行语音识别。有必要预先为语音识别系统建立关于预定字的语言模型和发音字典,以从语音信号中语音识别出预定字。

发明内容

[0005] 技术方案

[0006] 本发明提供了一种用于使用语言模型执行语音识别的方法和设备,并且更具体地,提供了一种用于为新字的语音识别建立语言模型和针对包括新字的语音执行语音识别的方法和装置。

[0007] 有益效果

[0008] 根据以上示例性实施例中的一个或者更多,可以通过更新包括相对小数量的概率的语言模型而不是更新包括相对大数量的概率的语言模型来最小化用于更新语言模型所消耗的时间段。

附图说明

[0009] 图1是例示根据实施例的执行语音识别的设备100的框图;

[0010] 图2是示出根据实施例的语音识别设备230和用于更新语音识别数据的语音识别数据更新设备220的框图;

[0011] 图3是示出根据实施例的更新用于新字的识别的语音识别数据的方法的流程图;

[0012] 图4是示出根据实施例的用于添加新字的系统的示例的框图;

[0013] 图5和图6是示出根据实施例的添加新字的示例的流程图;

[0014] 图7是示出根据实施例的新字与子字之间的对应关系的示例的表;

[0015] 图8是示出根据实施例的在语音识别期间关于新字的出现概率信息的示例的表;

[0016] 图9是示出根据实施例的用于更新用于识别新字的语音识别数据的系统的框图;

[0017] 图10是示出根据实施例的更新用于识别新字的语言数据的方法的流程图;

[0018] 图11是示出根据实施例的执行语音识别的语音识别设备的框图;

- [0019] 图12是示出根据实施例的执行语音识别的方法的流程图；
- [0020] 图13是示出根据实施例的执行语音识别的方法的流程图；
- [0021] 图14是示出根据实施例的基于以情形信息为基础执行的语音识别的结果来执行模块的语音识别系统的框图；
- [0022] 图15是示出根据实施例的关于模块的情形信息的示例的示意图；
- [0023] 图16是示出根据实施例的执行语音识别的方法的示例的流程图；
- [0024] 图17是示出根据实施例的执行语音识别的方法的示例的流程图；
- [0025] 图18是示出根据实施例的根据基于情形信息执行的语音识别的结果来执行多个模块的语音识别系统的框图；
- [0026] 图19是示出根据实施例的针对多个设备的话音命令的示例的示意图；
- [0027] 图20是示出根据实施例的语音识别设备的示例的框图；
- [0028] 图21是示出根据实施例的在显示设备上执行语音识别的示例的框图；
- [0029] 图22是示出根据实施例的考虑到情形信息来更新语言模型的示例的框图；
- [0030] 图23是示出根据实施例的包括与各个应用相对应的语言模型的语音识别系统的示例的框图；
- [0031] 图24是示出根据实施例的基于语音识别的结果发送执行任务的请求的用户设备的示例的示意图；
- [0032] 图25是示出根据实施例的生成关于语音数据的类别的个人优选内容列表的方法的框图；
- [0033] 图26是示出根据实施例的确定语音数据的类别的示例的示意图；
- [0034] 图27是示出根据实施例的根据语音数据的类别来更新语音识别数据的方法的流程图；
- [0035] 图28和图29是示出根据实施例的可以被分类的声学数据的示例的示意图；
- [0036] 图30和图31是示出根据实施例的执行个性化语音识别方法的示例的框图；
- [0037] 图32是示出根据实施例的语音识别数据更新设备的内部配置的框图；
- [0038] 图33是示出根据实施例的语音识别设备的内部配置的框图；
- [0039] 图34是用于描述根据实施例的用户设备3400的配置的框图。

具体实施方式

[0040] 根据本发明的一个方面,提供了一种更新包括用于语音识别的语言模型的语音识别数据的方法,所述方法包括:获取包括至少一个字的语言数据;从至少一个字当中检测语言模型中不存在的字;获取关于检测到的字的至少一个音素序列;通过将至少一个音素序列划分成预定单元分量来获取构成至少一个音素序列的分量;确定关于在语音识别期间构成至少一个音素序列中的每一个的各个分量出现的概率的信息;以及基于所确定的概率信息来更新语言模型。

[0041] 此外,语言模型包括第一语言模型和包括至少一个语言模型的第二语言模型,并且语言模型的更新包括基于所确定的概率信息来更新第二语言模型。

[0042] 此外,所述方法进一步包括:基于包括在第二语言模型中的至少一个出现概率信息来更新第一语言模型;以及基于检测到的字的音素序列来更新包括关于字的音素序列的

信息的发音字典。

[0043] 此外,在字或者另一分量在相应的分量之前出现的条件下,出现概率信息包括关于分量中的每一个的出现概率的信息。

[0044] 此外,出现概率信息的确定包括:获取关于与检测到的字相对应的周围情形的情形信息;以及基于该情形信息选择语言模型以添加关于检测出的字的出现概率信息。

[0045] 此外,语言模型的更新包括基于所确定的出现概率信息来更新关于与情形信息相对应的模块的第二语言模型。

[0046] 根据本发明的另一方面,提供了一种执行语音识别的方法,所述方法包括:获取用于执行语音识别的语音数据;从语音数据中获取至少一个音素序列;获取关于构成至少一个音素序列的预定单元分量出现的概率的信息;基于关于预定单元分量出现的概率的信息来确定至少一个音素序列中的一个;以及基于用于将包括在所确定的音素序列中的预定单元分量转换成字的分段信息获取与所确定的音素序列相对应的字。

[0047] 此外,至少一个音素序列的获取包括获取以下各项:对于音素序列而言,关于与该音素序列相对应的字的信息存在于包括关于字的音素序列的信息的发音字典中的音素序列;以及对于音素序列而言,关于与音素序列相对应的字的信息不存在于发音字典中的音素序列。

[0048] 此外,关于分量的出现概率信息的获取包括:确定包括关于分量的出现概率信息的多个语言模型;确定针对多个所确定的语言模型的权重;从多个语言模型中获取关于分量的至少一个出现概率信息;以及根据各个出现概率信息所属于的语言模型,通过将所确定的权重应用到所获取的出现概率信息,来获取关于分量的出现概率信息。

[0049] 此外,关于分量的出现概率信息的获取包括:获取关于语音数据的情形信息;基于情形信息确定至少一个第二语言模型;以及从至少一个所确定的第二语言模型中获取关于分量的出现概率信息。

[0050] 此外,至少一个第二语言模型与模块或者包括至少一个模块的组相对应,并且如果所获取的情形信息包括模块的标识符,则至少一个第二语言模型的确定包括确定与标识符相对应的至少一个第二语言模型。

[0051] 此外,情形信息包括个性化模型信息,所述个性化模型信息包括按类别的声学信息和按类别的关于优选语言的信息中的至少一个,并且第二语言模型的确定包括:基于声学信息和关于按类别的优选的语言的信息中的至少一个来确定关于语音数据的类别;以及基于所确定的类别来确定第二语言模型。

[0052] 此外,所述方法进一步包括:获取语音数据和作为语音数据的语音识别的结果的文本;从文本或者情形信息中检测关于内容的信息;从语音数据中检测声学信息;确定与关于内容和声学信息的信息相对应的类别;以及基于关于内容的信息和情形信息中的至少一个来更新关于与所确定的类别相对应的语言模型的信息。

[0053] 根据本发明的另一方面,提供了一种用于在语音识别期间更新包括关于各个字的出现概率信息的语言模型的设备,该设备包括控制器和存储器,所述控制器获取包括至少一个字的语言数据、从至少一个字当中检测语言模型中不存在的字、获取关于检测到的字的至少一个音素序列、通过将至少一个音素序列划分成预定单元分量来获取构成至少一个音素序列的分量、确定在语音识别期间关于构成至少一个音素序列中的每一个的各个分量

出现的概率的信息、以及基于所确定的概率信息来更新语言模型,所述存储器存储所更新的语言模型。

[0054] 根据本发明的另一方面,提供了一种用于执行语音识别的设备,该设备包括用户输入器和控制器,所述用户输入器获取用于执行语音识别的语音数据,所述控制器从语音数据中获取至少一个音素序列、获取关于构成至少一个音素序列的预定单元分量出现的概率的信息、基于关于预定单元分量出现的概率的信息来确定至少一个音素序列中的一个、以及基于用于将包括在所确定的音素序列中的预定单元分量转换成字的分段信息来获取与所确定的音素序列相对应的字。

[0055] 现在将参考附图更充分地描述本发明,附图中示出了本发明的示例性实施例。在本发明的描述中,当认为对相关技术的某些详细说明可能不必要地模糊本发明的实质时,省略该详细说明。附图中相同的参考标号始终表示相同的元素。

[0056] 以下参考附图详细地描述本发明的优选实施例。在描述实施例之前,在说明书和权利要求书中使用的词语和术语不应该被解释为常见或者字典的含义,而是在(多个)发明人可以适当地定义术语的概念来以最佳方法解释本发明的原则下,被解释为符合本发明的精神的含义和概念。因此,在说明书中描述的实施例和在图中所示的配置不超过本发明的最优选实施例,并且不充分覆盖本发明的精神。因此,应当理解,当本申请被提交时,可以有能够替代它们的各种等价和修改。

[0057] 在附图中,一些元素被夸大、省略或者简化,并且各个元素的尺寸不充分代表其实际尺寸。本发明不限于附图中所示的相对尺寸或者距离。

[0058] 另外,除非明确地相反描述,否则词语“包括”以及诸如“包括着”或者“包括了”的变形形式将理解为:暗示所陈述元素的包含,但是不暗示对任何其它元素的排除。另外,在说明书中描述的术语“单元”意思是用于处理至少一个功能和操作并能够由诸如FPGA或者ASIC的软件组件或者硬件组件来实现的单元。然而,“单元”不限于软件组件或者硬件组件。“单元”可以在记录介质上具体化,并且可以被配置为操作一个或者更多处理器。

[0059] 因此,例如,“单元”可以包括:组件(诸如软件组件、面向对象的软件组件、类组件和任务组件)、进程、功能、属性、程序、子程序、程序代码段、驱动器、固件、微代码、电路、数据、数据库、数据结构、表、数组和变量。“单元”中提供的组件和功能可以组合成更小数量的组件和“单元”,或者可以进一步划分成更大数量的组件和“单元”。

[0060] 现在将参考附图更充分地描述本发明,在附图中示出了本发明的示例性实施例。然而,本发明可以以许多不同的形式来具体化,并且不应该被解释为限于在此所阐述的实施例;而是提供这些实施例,以便本公开将是彻底的和完整的,并且将向本领域技术人员充分地传达本发明的概念。在本发明的描述中,当认为对相关技术的某些详细说明可能不必要地模糊本发明的实质时,省略该详细说明。附图中相同的参考标号表示相同的元素,并且因此将省略它们的描述。

[0061] 以下,将参考附图通过解释本发明的优选实施例来详细地描述本发明。

[0062] 图1是例示根据实施例的执行语音识别的设备100的框图。

[0063] 参考图1,设备100可以包括以下各项作为用于执行语音识别的组件:特征提取单元110、候选音素序列检测单元120、和字选择单元140。特征提取单元110提取关于输入语音信号的特征信息。候选音素序列检测单元120从所提取的特征信息中检测至少一个候选音

素序列。字选择单元140基于关于各个候选音素序列的出现概率信息来选择最终的语音识别出的字。关于字的出现概率信息指的是表明在语音识别期间在语音识别出的字中该字出现的概率的信息。以下,将详细地描述设备100的组件。

[0064] 当接收到话音信号时,设备100可以检测由讲话者实际讲的话音部分,并且提取表明话音信号的特征的信息。表明话音信号的特征的信息可以包括基于与话音信号相对应的波形表明嘴的形状或舌头的位置的信息。

[0065] 候选音素序列检测单元120可以通过使用所提取的关于话音信号的特征信息和声学模型130,来检测可以与话音信号相匹配的至少一个候选音素序列。可以根据话音信号提取多个候选音素序列。例如,由于发音“jyeo”和“jeo”彼此相似,针对相同的话音信号可以检测包括发音“jyeo”和“jeo”的多个候选音素序列。可以逐字地检测候选音素序列。然而,本发明不限于此,还可以以各种单位中的任一个(诸如以音素为单位)来检测候选音素序列。

[0066] 声学模型130可以包括用于从关于话音信号的特征信息中检测候选音素序列的信息。此外,声学模型130可以通过使用统计方法基于大量语音数据生成、可以基于关于未指定的讲话者的发音数据生成、或者可以基于关于特定的讲话者的发音数据生成。因此,声学模型130可以根据特定的讲话者被独立地应用于语音识别。

[0067] 字选择单元140可以通过使用发音字典150和语言模型160,来获取关于由候选音素序列检测单元120检测到的各个候选音素序列的出现概率信息。接下来,字选择单元140基于关于各个候选音素序列的出现概率信息选择最终的语音识别出的字。详细地,字选择单元140可以通过使用发音字典150确定与各个候选音素序列相对应的字,并且通过使用语言模型160获取关于所确定的字的各个出现概率。

[0068] 发音字典150可以包括用于获取与由候选音素序列检测单元120检测到的候选音素序列相对应的字的信息。发音字典150可以基于候选音素序列被建立,所述候选音素序列是基于各个字的音素的变化被获取的。

[0069] 字的发音不是始终如一的,因为字的发音可以基于该字前后的字、该字在句子中的位置、或者讲话者的特点而变化。此外,关于字的出现概率指的是该字可能出现的概率或者该字可能与特定的字一起出现的概率。设备100可以基于出现概率,考虑到上下文来执行语音识别。设备100可以通过获取与候选音素序列相对应的字(通过使用发音字典150)以及通过获取关于各个字的出现概率的信息(通过使用语言模型160)来执行语音识别。然而,本发明不限于此,设备100可以不获取与候选音素序列相对应的字,通过使用候选音素序列从语言模型160中获取出现概率。

[0070] 例如,在韩语的情况下,当检测到候选音素序列“hakkkyo”时,字选择单元140可以通过使用发音字典150获取字“hakgyo”作为与检测到的候选音素序列“hakkkyo”相对应的字。另一示例中,在英语的情况下,当检测到候选音素序列“skul”时,字选择单元140可以通过使用发音字典150获取字“school”作为与检测到的候选音素序列“skul”相对应的字。

[0071] 语言模型160可以包括关于字的出现概率信息。可以有与关于每个字的出现概率相关的信息。设备100可以从语言模型160中获取关于包括在各个候选音素序列中的字的出现概率信息。

[0072] 例如,如果字A在当前字B出现之前出现,则语言模型160可以包括关于出现概率P

(B|A)的信息,所述出现概率 $P(B|A)$ 为当前字B可能出现的概率。换句话说,关于字B的出现概率 $P(B|A)$ 可能受制于在字B的出现之前的字A的出现。另一示例中,语言模型160可以包括出现概率 $P(B|A C)$,所述出现概率 $P(B|A C)$ 受制于字A和字C的出现,即在字B的出现之前的多个字的出现。换句话说,出现概率 $P(B|A C)$ 可能受制于在字B的出现之前的字A和字C两者的出现。另一示例中,语言模型160可以包括关于字B的出现概率 $P(B)$,而不是条件概率。出现概率 $P(B)$ 指的是在语音识别期间字B可能出现的概率。

[0073] 通过使用语言模型160,设备100可以基于关于与通过字选择单元140确定的各个候选音素序列相对应的字的出现概率,来最终确定语音识别出的字。换句话说,设备100可以将与最高出现概率相对应的字最终确定为语音识别出的字。字选择单元140可以将语音识别出的字输出为文本。

[0074] 虽然本发明不限于更新语言模型或者逐字地执行语音识别,并且这样的操作可以被逐个序列地执行,为方便说明,以下将描述更新语言模型或者逐字执行语音识别的方法。

[0075] 以下,参考图2至图9,将详细地描述更新用于新字的语音识别的语音识别数据的方法。

[0076] 图2是示出根据实施例的语音识别设备230和用于更新语音识别数据的语音识别数据更新设备220的框图。

[0077] 虽然图2示出语音识别数据更新设备220和语音识别设备230是单独的设备,但是它仅仅是实施例,并且语音识别数据更新设备220和语音识别设备230可以具体化为单一设备,例如,语音识别数据更新设备220可以包括在语音识别设备230中。在附图和以下描述的实施例中,包括在语音识别数据更新设备220和语音识别设备230中的组件可以是物理上或逻辑上分散的或相互集成的。

[0078] 语音识别设备230可以是通过使用从设备接收的语音数据来执行语音识别并输出语音识别出的字的自动语音识别(ASR)服务器。

[0079] 语音识别设备230可以包括执行语音识别的语音识别单元231和用于执行语音识别的语音识别数据232、233、和235。语音识别数据232、233和235可以包括其它模型232、发音字典233、和语言模型235。此外,根据实施例的语音识别设备230可以进一步包括用于更新语音识别数据232、233、和235的分段模型234。

[0080] 图1中的设备100可以与图2中的语音识别单元231相对应,并且图2中的语音识别数据232、233、和235可以分别与图1中的声学模型130、发音字典150、和语言模型160相对应。

[0081] 发音字典233可以包括关于候选音素序列和至少一个字之间的至少对应(correspondence)的信息。语言模型235可以包括关于字的出现概率信息。其它模型232可以包括可以用于语音识别的其它模型。例如,其它模型232可以包括用于从关于语音信号的特征信息中检测候选音素序列的声学模型。

[0082] 根据实施例的语音识别设备230可以进一步包括用于通过反映新字来更新语言模型235的分段模型234。根据实施例,分段模型234包括可以用于通过使用新字来更新语音识别数据的信息。详细地,分段模型234可以包括用于将包括在收集的语言数据中的新字划分成预定单元分量的信息。例如,如果将新字划分成子字(subword)的单元,则分段模型234可以包括子字文本,诸如“ga gya ah re pl tam”。然而,本发明不限于此,分段模型234可以

包括划分成预定单元分量的字,并且可以根据预定单元分量来划分新字。子字指的是可以独立地清晰发音的话音单元。

[0083] 图2中的分段模型234被包括在语音识别设备230中。然而,本发明不限于此,分段模型234可以被包括在语音识别数据更新设备220中,或者可以被包括在另一外部设备中。

[0084] 语音识别数据更新设备220可以更新用于语音识别的语音识别数据232、233、和235中的至少一个。语音识别数据更新设备220可以包括以下各项作为用于更新语音识别数据的组件:新字检测单元221、发音生成单元222、子字划分单元223、出现概率信息确定单元224、和语言模型更新单元225。

[0085] 语音识别数据更新设备220可以收集包括至少一个字的语言数据210,并且通过使用包括在语言数据210中的新字来更新语音识别数据232、233、和235中的至少一个。

[0086] 语音识别数据更新设备220可以周期性地或者当事件发生时收集语言数据210并更新语音识别数据。例如,当将用户设备的显示单元上的屏幕图像切换到另一屏幕图像时,语音识别数据更新设备220可以收集包括在切换的屏幕图像中的语言数据210,并且基于所收集的语言数据210来更新语音识别数据。语音识别数据更新设备220可以通过从用户设备接收包括在显示单元上的屏幕图像中的语言数据210,来收集语言数据210。

[0087] 可替换地,如果语音识别数据更新设备220是用户设备,则可以根据内部算法获取包括在显示单元上的屏幕图像中的语言数据210。用户设备可以是与语音识别设备230、或者语音识别数据更新设备220、或者外部设备相同的设备。

[0088] 当语音识别数据通过语音识别数据更新设备220被更新时,语音识别设备230可以针对与新字相对应的话音信号执行语音识别。

[0089] 语言数据210可以以文本的形式被收集。例如,语言数据210可以包括文本,所述文本包括在内容或者网页中。如果文本被包括在图像文件中,则该文本可以经由光学字符识别(optical character recognition,OCR)获取。语言数据210可以包括以包括多个字的句子或者段落的形式文本。

[0090] 新字检测单元221可以从所收集的语言数据210中检测不包括在语言模型235中的新字。当语音识别设备230执行语音识别时,针对不包括在语言模型235中的字,关于出现概率的信息不能被获取,因此不包括在语言模型235中的字不能被输出为语音识别出的字。根据实施例的语音识别数据更新设备220可以通过检测不包括在语言模型235中的新字并将关于新字的出现概率信息添加到语言模型235中,来更新语音识别数据。接下来,语音识别设备230可以基于关于新字的出现概率来将新字输出为语音识别出的字。

[0091] 语音识别数据更新设备220可以将新字划分成子字,并且将关于新字的各个子字的出现概率信息添加到语言模型235中。由于根据实施例的语音识别数据更新设备220可以不更新发音字典233和其它模型232而仅仅通过更新语言模型235来更新用于识别新字的语音识别数据,因此语音识别数据可以被快速地更新。

[0092] 发音生成单元222可以根据标准发音规则或者反映讲话者的特点的发音规则,将通过新字检测单元221检测到的新字转换为至少一个音素序列。

[0093] 另一示例中,可以基于用户输入确定关于新字的音素序列,而不是经由发音生成单元222生成音素序列。此外,不限于以上陈述的实施例的发音规则,可以基于与诸如关于新字或者时间的讲话者的特点和位置特点的各种情形相对应的条件来确定音素序列。例

如,可以基于相同字符可能根据讲话者的情形,例如,在早晨和晚上不同的话音或者讲话者的语言行为的变化,而不同地发音的事实来确定音素序列。

[0094] 子字划分单元223可以基于分段模型234将由发音生成单元222转换的音素序列划分成预定单元分量。

[0095] 例如,在韩语的情况下,发音生成单元222可以将新字“gim yeon a”转换为音素序列“gi myeo na”。接下来,子字划分单元223可以参考包括在分段模型234中的子字信息并将音素序列“gi myeo na”划分成子字分量“gi”、“myeo”、和“na”。详细地,子字划分单元223可以从包括在分段模型234中的子字当中提取与音素序列“gi myeo na”的子字分量相对应的“gi”、“myeo”、和“na”。子字划分单元223通过使用检测出的子字可以将音素序列“gi myeo na”划分成子字分量“gi”、“myeo”、和“na”。

[0096] 在英语的情况下,发音生成单元222可以将被识别为新字的字“texas”转换为音素序列“teksəs”。接下来,参考包括在分段模型234中的子字信息,子字划分单元223可以将“teksəs”划分成子字“teks”和“əs”,根据实施例,基于分段模型234的用于划分的预定单元不仅可以包括子字,而且还包括其它话音单元,诸如分段。

[0097] 在韩语的情况下,子字可以包括四种类型:仅仅元音,元音和辅音的组合,辅音和元音的组合,以及辅音、元音和辅音的组合。如果音素序列被划分成子字,则分段模型234可以包括成千上万的子字信息,例如,ga、gya、gan、gal、nam、nan、un、hu等。

[0098] 子字划分单元223可以将可以是日语字或者汉语字的新字转换为通过使用音标(例如,拉丁字母、片假名、谚文等)表明的音素序列,并且所转换的音素序列可以被划分成子字。

[0099] 在除了上述语言的语言的情况下,对于语言中的每一种,分段模型234可以包括用于将新字划分成预定单元分量的信息。此外,子字划分单元223可以基于分段模型234将新字的音素序列划分成预定单元分量。

[0100] 出现概率信息确定单元224可以确定关于构成新字的音素序列的预定单元分量的出现概率信息。如果新字被包括在语言数据的句子中,则出现概率信息确定单元224可以通过使用包括在除了新字的句子中的字来获取出现概率信息。

[0101] 例如,在句子“oneul gim yeon a boyeojyo”中,如果字“gimyeona”被检测为新字,则出现概率信息确定单元224可以确定关于子字“gi”、“myeo”、和“na”的出现概率。例如,出现概率信息确定单元224可以通过使用关于包括在句子中的字“oneul”的出现概率信息来确定出现概率 $P(\text{gi}/\text{oneul})$ 。此外,如果“texas”被检测为新字,则可以针对各个子字“teks”和“əs”来确定出现概率信息。

[0102] 如果假设至少一个特定子字或者字在当前子字之前出现,则关于子字的出现概率信息可以包括关于在语音识别期间当前子字可能出现的概率的信息。此外,关于子字的出现概率信息可以包括关于在语音识别期间当前子字可能出现的无条件概率的信息。

[0103] 语言模型更新单元225可以通过使用针对各个子字所确定的出现概率信息,来更新分段模型234。语言模型更新单元225可以更新语言模型235,使得在特定子字或者字在当前字或者子字之前出现的条件下,所有概率的和为1。

[0104] 详细地,如果针对各个子字确定的出现概率信息中的一个为 $P(B|A)$,则在A出现在

当前字或者子字之前的条件下,语言模型更新单元225可以获取包括在语言模型235中的概率 $P(C|A)$ 和 $P(D|A)$ 。接下来,语言模型更新单元225可以重新确定概率 $P(B|A)$ 、 $P(C|A)$ 、和 $P(D|A)$ 的值,使得 $P(B|A)+P(C|A)+P(D|A)$ 为1。

[0105] 当语言模型被更新时,语言模型更新单元225可以重新确定关于包括在语言模型235中的其它字或者子字的概率,并且用于更新语言模型所消耗的时间段可以随着包括在语言模型235中的概率的数量的增加而增加。因此,根据实施例的语言模型更新单元225可以通过更新包括相对小数量的概率的语言模型而不是更新包括相对大数量的概率的语言模型,来最小化用于更新语言模型所消耗的时间段。

[0106] 在上述语音识别过程中,语音识别设备230可以将声学模型、发音字典、和语言模型一起使用,以识别包括在语音信号中的单一字。因此,当语音识别数据被更新时,有必要将声学模型、发音字典、和语言模型一起更新,使得新字可以被语音识别出。然而,为了将声学模型、发音字典、和语言模型一起更新以语音识别出新字,还有必要更新关于一起存在的字的信息,因此1小时或者更长的时间段是必要的。因此,对于语音识别设备230,当新字被收集时立即执行关于新字的语音识别是困难的。

[0107] 对于根据实施例的语音识别数据更新设备220,基于新字的特点更新其它模型232和发音字典233是不必要的。语音识别数据更新设备220可以基于针对构成新字的各个子字分量所确定的出现概率信息,仅仅更新语言模型235。因此,在根据实施例的更新语言模型的方法中,语言模型可以针对新字在几秒内被更新,并且语音识别设备230可以在语音识别中实时地反映新字。

[0108] 图3是示出根据实施例的更新用于新字的识别的语音识别数据的方法的流程图。

[0109] 参考图3,在操作S301中,语音识别数据更新设备220可以获取包括至少一个字的语言数据。语言数据可以包括文本,所述文本包括在正在被用户或者设备的模块使用的设备的显示屏幕上正在显示的内容或者网页中。

[0110] 在操作S303中,语音识别数据更新设备220可以从至少一个字当中检测语言数据中不存在的字。语言数据中不存在的字是没有关于其出现概率的信息的字,并且不能被检测为语音识别出的字。因此,语音识别数据更新设备220可以将语言数据中不存在的字检测为用于更新语音识别数据的新字。

[0111] 在操作S305中,语音识别数据更新设备220可以获取与在操作S303中检测到的新字相对应的至少一个音素序列。与字相对应的多个音素序列可以基于包括发音规则或者讲话者的特点的各种条件而存在。此外,数字或者符号可以与各种发音规则相对应,因此针对数字或者符号可以存在多个相应的音素序列。

[0112] 在操作S307中,语音识别数据更新设备220可以将操作S305中获取的至少一个音素序列中的每一个划分成预定单元分量,并且获取构成至少一个音素序列中的每一个的分量。详细地,语音识别数据更新设备220可以基于包括在分段模型234中的子字信息将音素序列中的每一个划分成子字,从而获取构成新字的音素序列中的每一个的分量。

[0113] 在操作S309中,语音识别数据更新设备220可以在语音识别期间确定关于在操作S307中获取的分量中的每一个的出现概率的信息。关于出现概率的信息可以包括条件概率,并且可以包括关于在特定子字或字在当前子字之前出现的条件下当前子字的出现概率的信息。然而,本发明不限于此,并且关于出现概率的信息可以包括关于当前子字的无条件

出现概率。

[0114] 语音识别数据更新设备220可以通过使用在操作S301中获取的语言数据,来确定关于预定分量的出现概率信息。语音识别数据更新设备220可以通过使用新字的音素序列的子字分量所属于的句子或者段落,来确定关于各个分量的出现概率,并且可以确定关于各个分量的出现概率信息。此外,语音识别数据更新设备220可以通过使用在操作S305中获取的至少一个音素序列与分量所属于的句子或者段落一起确定关于各个分量的出现概率信息。下面将参考图16和图17给出其详细的描述。

[0115] 关于可以在操作S309中确定的出现概率的信息可以不仅包括条件概率,而且还包括无条件概率。

[0116] 在操作S311中,语音识别数据更新设备220可以通过使用在操作S309中确定的出现概率信息来更新语言模型。例如,语音识别数据更新设备220可以通过使用针对各个子字确定的出现概率信息来更新语言模型235。详细地,语音识别数据更新设备220可以更新语言模型235,使得在特定子字或者字在当前字或者子字之前出现的条件下,包括在语言模型235中的至少一个概率的和为1。

[0117] 图4是示出根据实施例的用于添加新字的系统的示例的框图。

[0118] 参考图4,根据实施例,系统可以包括用于添加新字的语音识别数据更新设备420和用于执行语音识别的语音识别设备430。不同于图2的语音识别设备230,图4的语音识别设备430可以进一步包括分段信息438、语言模型组合单元435、第一语言模型436、和第二语言模型437。图4的语音识别数据更新设备420和语音识别设备430可以与图2的语音识别数据更新设备220和语音识别设备230相对应,并且将省略对其的重复的描述。

[0119] 当语音识别被执行时,不同于图2的语言模型235,图4的语言模型组合单元435可以通过组合多个语言模型来确定关于各个字的出现概率。换句话说,语言模型组合单元435可以获取关于包括在多个语言模型中的字的出现概率,并且通过组合多个所获取的关于该字的出现概率来获取关于该字的出现概率。参考图4,语言模型组合单元435可以通过组合第一语言模型436和第二语言模型437来获取关于各个字的出现概率。

[0120] 第一语言模型436是预先包括在语音识别设备430中的语言模型,并且可以包括可以用于一般语音识别系统的一般用途的语言数据。第一语言模型436可以包括关于字或者基于大量语言数据(例如,包括在网页、内容等中的成千上万个句子)所确定的预定单元的出现概率。因此,由于第一语言模型436是基于大量的样本数据获取的,所以基于第一语言模型436的语音识别可以保证高效率 and 稳定性。

[0121] 第二语言模型437是包括关于新字的出现概率的语言模型。不同于第一语言模型436,第二语言模型437可以基于情形选择性地被应用,并且可以存在可以基于情形选择性地被应用的至少一个第二语言模型437。

[0122] 根据实施例,第二语言模型437是包括关于新字的出现概率信息的语言模型。不同于第一语言模型436,第二语言模型437可以根据不同的情形选择性地被应用,并且可以存在可以根据情形选择性地被应用的至少一个第二语言模型437。

[0123] 第二语言模型437可以由语音识别数据更新设备420实时更新。当语言模型被更新时,语音识别数据更新设备420可以通过使用关于新字的出现概率来重新确定包括在语言模型中的出现概率。由于第二语言模型437包括相对小数量的出现概率信息,所以要被考虑

用于更新第二语言模型437的出现概率信息的数量是相对小的。因此,用于识别新字的第二语言模型437的更新可以更快地被执行。

[0124] 在语音识别期间语言模型组合单元435通过组合第一语言模型436和第二语言模型437来获取关于字或者子字的出现概率的方法的详细描述将参考图11和图12在以下给出,在所述图11和图12中示出了根据实施例的执行语音识别的方法。

[0125] 不同于语音识别设备230,图4的语音识别设备430可以进一步包括分段信息438。

[0126] 分段信息438可以包括关于新字与通过划分新字获取的子字分量之间的对应关系的信息。如图4中所示,当新字的音素序列基于分段模型434被划分成子字时,可以由语音识别数据更新设备420生成分段信息438。

[0127] 例如,如果新字是“gim yeon a”,并且其子字是“gi”、“myeo”和“na”,则分段信息438可以包括表明新字“gim yeon a”与子字“gi”、“myeo”和“na”彼此相对应的信息。另一示例中,如果新字是“texas”,并且其子字是“teks”和“əs”,则分段信息438可以包括表明新字“texas”与子字“teks”和“əs”彼此相对应的信息。

[0128] 在执行语音识别的方法中,与基于声学模型确定的音素序列相对应的字可以从发音字典433中获取。然而,如果根据实施例更新语音识别设备430的第二语言模型437,而不更新发音字典433,因此发音字典433不包括关于新字的信息。

[0129] 因此,语音识别设备430可以获取关于与通过使用分段信息438划分的预定单元分量相对应的字的信息,并且以文本的形式输出最终的语音识别结果。

[0130] 通过使用分段信息438来执行语音识别的方法的详细描述将在以下参考涉及执行语音识别的方法的图12至图14给出。

[0131] 图5和图6是示出根据实施例的添加新字的示例的流程图。

[0132] 参考图5,在操作510中,语音识别数据更新设备220可以以文本数据的形式获取包括句子“oneul 3:10tu yuma eonje hae?”的语言数据。

[0133] 在操作530中,通过使用包括第一语言模型和第二语言模型中的至少一个的语言模型520,语音识别数据更新设备220可以检测不存在于语言模型520中的字“3:10”和“yuma”。

[0134] 在操作540中,语音识别数据更新设备220可以通过使用分段模型550和发音生成单元422获取与检测到的字相对应的音素序列,并且将音素序列中的每一个划分成预定单元分量。在操作541和542中,语音识别数据更新设备220可以获取与字“3:10”相对应的音素序列“ssuriten”、“samdaesip”和“sesisippun”以及与字“yuma”相对应的音素序列“yuma”。接下来,语音识别数据更新设备220可以将音素序列中的每一个划分成子字分量。

[0135] 在操作560中,语音识别数据更新设备220可以合成包括在操作541和542中获取的音素序列的句子。由于获取了与“3:10”相对应的三个音素序列,所以可以合成三个句子。

[0136] 在操作570中,语音识别数据更新设备220可以确定关于在操作560中合成的句子中的每一个中的预定单元分量的出现概率信息。

[0137] 例如,关于第一个句子的“ssu”的概率 $P(ssu|oneul)$ 可以具有值 $1/3$,因为,当“oneul”出现时,“ssu”、第二句的“sam”、或者第三句的“se”可能跟随。在相同的方面,概率 $P(sam|oneul)$ 和概率 $P(se|oneul)$ 可以具有值 $1/3$ 。由于只当在三个句子中“ssu”出现之后

“ri”出现时关于“ri”的概率 $P(ri|ssu)$ 存在,因此概率 $P(ri|ssu)$ 可以具有值1。在相同的方面,概率 $P(ten|ri)$ 、概率 $P(yu|tu)$ 、概率 $P(ma|yu)$ 、概率 $P(dae|sam)$ 、概率 $P(sip|dae)$ 、概率 $P(si|se)$ 、和概率 $P(sip|si)$ 可以具有值1。在概率 $P(ppun|sip)$ 的情况下,当“sip”出现时“tu”或者“ppun”可能出现,因此概率 $P(ppun|sip)$ 可以具有值1/2。

[0138] 在操作580中,语音识别数据更新设备220可以基于在操作570中确定的出现概率信息来更新第一语言模型中的一个或者更多和至少一个第二语言模型。在更新用于新字的语音识别的语言模型的情况下,语音识别数据更新设备220可以基于关于已经包括在语言模型中的其它字或者子字的出现概率来更新语言模型。

[0139] 例如,在“oneul”首先出现的条件下,考虑到已经包括在语言模型中的概率,概率 $P(X|oneul)$,例如,概率 $P(ssu|oneul)$ 、概率 $P(sam|oneul)$ 、以及概率 $P(se|oneul)$,则已经包括在语言模型中的概率 $P(X|oneul)$ 可以被重新确定。例如,如果已经包括在语言模型中的概率中存在 $P(du|oneul) = P(tu|oneul) = 1/2$ 的概率,则语音识别数据更新设备220可以基于语言模型中已经存在的概率和在操作570中获取的概率,来重新确定概率 $P(X|oneul)$ 。详细地,由于共有5个“oneul”出现的情况,所以关于各个子字的出现概率中的每一个都为1/5,因此概率 $P(X|oneul)$ 中的每一个可以具有值1/5。因此,语音识别数据更新设备220可以基于包括在相同语言模型中的相同条件来重新确定条件出现概率,使得出现概率的值之和为1。

[0140] 参考图6,在操作610中,语音识别数据更新设备220可以以文本数据的形式获取包括句子“oneul gim yeon a boyeojyo”的语言数据。

[0141] 在操作630中,语音识别数据更新设备220可以通过使用第一语言模型和第二语言模型中的至少一个来检测语言模型620中不存在的字“gim yeon a”和“boyeojyo”。

[0142] 在操作640中,语音识别数据更新设备220可以通过使用分段模型650和发音生成单元622获取与检测到的字相对应的音素序列,并且将音素序列中的每一个划分为预定单元分量。在操作641和642中,语音识别数据更新设备220可以获取与字“gim yeon a”相对应的音素序列“gi myeo na”和与字“boyeojyo”相对应的音素序列“boyeojyo”和“boyeojeo”。接下来,语音识别数据更新设备220可以将音素序列中的每一个划分成子字分量。

[0143] 在操作660中,语音识别数据更新设备220可以合成包括在操作641和642中获取的音素序列的句子。由于获取了与字“boyeojyo”相对应的两个音素序列,所以可以合成两个句子。

[0144] 在操作670中,语音识别数据更新设备220可以确定关于在操作660中合成的句子中的每一个中的预定单元分量的出现概率信息。

[0145] 例如,关于第一个句子中的“gi”的概率 $P(gi|oneul)$ 可以具有值1,因为在“oneul”出现的两个句子中“gi”跟随。在相同的方面,概率 $P(myeo|gi)$ 、概率 $P(na|myeo)$ 、概率 $P(bo|na)$ 、和概率 $P(yeo|bo)$ 可以具有值1,因为在每个条件下仅仅存在一次情况。在概率 $P(jyo|yeo)$ 和概率 $P(jeo|yeo)$ 的情况,在两个句子中当“yeo”出现时,“jyo”或者“jeo”可能出现,因此概率 $P(jyo|yeo)$ 和概率 $P(jeo|yeo)$ 两者都可以具有值1/2。

[0146] 在操作680中,语音识别数据更新设备220可以基于在操作670中确定的出现概率信息来更新第一语言模型中的一个或者更多和至少一个第二语言模型。

[0147] 图7是示出根据实施例的新字与子字之间的对应关系的示例的表。

[0148] 参考图7,如果字“gim yeon a”被检测为新字,则可以将“gi”、“myeo”、和“na”确定为与字“gim yeon a”相对应的子字,如710中所示。在相同的方面,如果字“boyeojyo”被检测为新字,则可以将“bo”、“yeo”和“jyo”与“bo”、“yeo”和“jeo”确定为与字“boyeojyo”相对应的子字,如720和730中所示。

[0149] 如图7中所示的关于新字与子字之间的对应关系的信息,可以被存储为分段信息438并在语音识别期间被利用。

[0150] 图8是示出根据实施例的在语音识别期间关于新字的出现概率信息的示例的表。

[0151] 参考图8,关于出现概率的信息可以包括关于无条件出现概率的信息和关于在先前出现的字的条件下的出现概率的信息中的至少一个。

[0152] 关于无条件出现概率810的信息可以包括关于与字或者子字相关的诸如概率P(oneul)、概率P(gi)、和概率P(jeo)的无条件出现概率的信息。

[0153] 关于在先前出现的字的条件下的出现概率820的信息可以包括关于在先前出现的字的条件下字或者子字的诸如概率P(gi|oneul)、概率P(myeo|gi)、和概率P(jyo|yeo)的出现概率信息。关于如图8中所示的“oneul gi”、“gi myeo”、和“yeo jyo”的出现概率,可以分别与概率P(gi|oneul)、概率P(myeo|gi)、和概率P(jyo|yeo)相对应。

[0154] 图9是示出根据实施例的用于更新用于识别新字的语音识别数据的系统的框图。

[0155] 图9中所示的语音识别数据更新设备920可以包括用于更新其它模型932、发音字典933、和第一语言模型935中的至少一个的新字信息922和语音识别数据更新单元923。

[0156] 图9的语音识别数据更新设备920和语音识别设备930可以与图2和图4的语音识别数据更新设备220和420,以及图2和图4的语音识别设备230和430相对应,并且将省略对其的重复的描述。此外,图9的语言模型更新单元921可以与包括在图2和图4的语音识别数据更新设备220和420中的组件221至225和421至425相对应,并且将省略对其的重复的描述。

[0157] 图9的新字信息922可以包括关于由语音识别数据更新设备920识别为新字的字的信息。新字信息922可以包括关于用于更新其它模型932、发音字典933和第一语言模型935中的至少一个的新字的信息。详细地,新字信息922可以包括关于与由语音识别数据更新设备920添加到第二语言模型936的出现概率相对应的字的信息。例如,新字信息922可以包括新字的音素序列、关于通过划分新字的音素序列获取的预定单元分量的信息、以及关于新字的各个分量的出现概率信息中的至少一个。

[0158] 语音识别数据更新单元923可以通过使用新字信息922来更新语音识别设备930的其它模型932、发音字典933、和第一语言模型935中的至少一个。详细地,语音识别数据更新单元923可以通过使用关于新字的音素序列的信息来更新声学模型、其它模型932和发音字典933。此外,语音识别数据更新单元923可以通过使用关于通过划分新字的音素序列获取的预定单元分量的信息和关于新字的各个分量的出现概率信息,来更新第一语言模型935。

[0159] 不同于关于包括在第二语言模型936中的出现概率的信息,关于包括在由语音识别数据更新单元923更新的第一语言模型935中的新字的出现概率信息可以包括关于没有被划分成预定单元分量的新字的出现概率信息。

[0160] 例如,如果新字信息922包括关于“gim yeon a”的信息,则语音识别数据更新单元923可以通过使用与“gim yeon a”相对应的音素序列“gi myeo na”,来更新声学模型和发音字典933。声学模型可以包括关于与“gi myeo na”相对应的话音信号的特征信息。发音字

典933可以包括与“gim yeon a”相对应的音素序列信息“gi myeo na”。此外,语音识别数据更新单元923可以通过重新确定包括在第一语言模型935中的出现概率信息(通过使用关于“gim yeon a”的出现概率信息),来更新第一语言模型935。

[0161] 包括在第一语言模型935中的出现概率信息是基于关于句子的大量信息被获取的,因此包括大数量的出现概率信息。因此,由于有必要基于关于新字的信息来重新确定包括在第一语言模型935中的出现概率信息,以更新第一语言模型935,因此更新第一语言模型935比更新第二种语言模型936可能花费明显更长的时间。语音识别数据更新设备920可以通过收集语言数据来实时地更新第二语言模型936,而语音识别数据更新设备920可能以长的时间段间隔(例如,一周一次或者一月一次)周期性地更新第一语言模型935。

[0162] 如果语音识别设备930通过使用第二语言模型936执行语音识别,则有必要在最终选择语音识别出的语言之后,通过使用分段信息进一步执行与预定单元分量相对应的文本的恢复。其原因在于,由于使用了关于预定单元分量的出现概率信息,最终选择的语音识别出的语言包括通过将新字划分为单元分量获取的音素序列。此外,包括在第二语言模型936中的出现概率信息不是基于关于句子的大量的信息获取的,而是基于包括新字的句子或者包括在第二语言模型936中的有限量的出现概率信息获取的。因此,包括在第一语言模型935中的出现概率信息可以比包括在第二语言模型936中的出现概率信息更准确。

[0163] 换句话说,在效率和稳定性方面,对于语音识别设备930,通过使用第一语言模型935可能比通过使用第二语言模型936执行语音识别更高效。因此,根据实施例的语音识别数据更新单元923可以周期性地更新第一语言模型935、发音字典933、和声学模型。

[0164] 图10是示出根据实施例的更新用于识别新字的语言数据的方法的流程图。

[0165] 不同于图3中所示的方法,图10中所示的方法可以进一步包括用于基于情形信息来选择至少一个或者更多第二语言模型中的一个并更新所选择的第二语言模型的操作。此外,图10中所示的方法可以进一步包括用于基于关于新字的信息来更新第一语言模型的操作,所述信息用于更新第二语言模型。

[0166] 参考图10,在操作S1001中,语音识别数据更新设备420可以获取包括字的语言数据。操作S1001可以与图3中的操作S301相对应。语言数据可以包括文本,所述文本包括在正在被用户或设备的模块使用的设备的显示屏幕上显示的内容或者网页中。

[0167] 在操作S1003中,语音识别数据更新设备420可以检测语言数据中不存在的字。换句话说,语音识别数据更新设备420可以从包括在语言数据中的至少一个字当中检测字,关于该字,在第一语言模型或者第二语言模型中关于出现概率的信息不存在。操作S1003可以与图3的操作S303相对应。

[0168] 由于第二语言模型包括关于通过将字划分为预定单元分量获取的各个分量的出现概率信息,因此根据实施例的第二语言模型不包括关于整个字的出现概率信息。语音识别数据更新设备420可以通过使用包括关于字和通过将字划分成预定单元分量获取的各个分量之间的对应关系的信息的分段信息来检测字,针对该检测到的字在第二语言模型中关于出现概率的信息不存在。

[0169] 在操作S1005中,语音识别数据更新设备420可以获取与在操作S1003中检测到的新字相对应的至少一个音素序列。基于包括发音规则或者讲话者的特点的各种条件,与字相对应的多个音素序列可以存在。操作S1005可以与图3的操作S305相对应。

[0170] 在操作S1007中,语音识别数据更新设备420可以将将在操作S1005中获取的至少一个音素序列中的每一个划分成预定单元分量,并且获取构成该至少一个音素序列中的每一个的分量。详细地,语音识别数据更新设备420可以基于包括在分段模型434中的子字信息将音素序列中的每一个划分成子字,从而获取构成新字的音素序列中的每一个的分量。操作S1007可以与图3的操作S307相对应。

[0171] 在操作S1009中,语音识别数据更新设备420可以获取与在操作S1003中检测到的字相对应的情形信息。情形信息可以包括关于检测到的新字的情形信息。

[0172] 根据实施例的情形信息可以包括关于用户的信息、模块标识信息、关于设备的位置的信息、和关于获取新字的位置的信息中的至少一个。例如,当新字在特定模块处或者正在执行模块的同时被获取时,情形信息可以包括特定模块或者关于正在执行的模块的信息。如果新字在特定讲话者正在使用语音识别数据更新设备420的同时被获取或者新字与特定说话者有关时,则关于新字的情形信息可以包括关于特定讲话者的信息。

[0173] 在操作S1011中,语音识别数据更新设备420可以基于在操作S1009中获取的情形信息来选择第二语言模型。语音识别数据更新设备420可以通过将关于新字的分量的出现概率信息添加到所选择的第二语言模型,来更新第二语言模型。

[0174] 根据实施例,语音识别设备430可以包括多个独立的第二语言模型。详细地,第二语言模型可以包括基于特定模块、模块、或者讲话者可以被选择性地应用的多个独立的语言模型。在操作S1011中,语音识别数据更新设备420可以从多个独立的语言模型当中选择与情形信息相对应的第二语言模型。在语音识别期间,语音识别设备430可以收集情形信息,并且通过使用与情形信息相对应的第二语言模型来执行语音识别。因此,根据实施例,可以基于情形信息执行自适应语音识别,因此可以提高语音识别效率。

[0175] 在操作S1013中,在语音识别期间语音识别数据更新设备420可以确定关于在操作S1007中获取的分量中的每一个的出现概率的信息。例如,语音识别数据更新设备420可以通过使用包括在语言数据中的字的分量所属于的句子或者段落,来确定关于各个子字分量的出现概率。操作S1013可以与图3的操作S309相对应。

[0176] 在操作S1015中,语音识别数据更新设备420可以通过使用在操作S1013中确定的出现概率信息来更新第二语言模型。语音识别数据更新设备420可以将关于新字的分量的出现概率信息简单地添加到第二语言模型。可替换地,语音识别数据更新设备420可以将关于新字的分量的出现概率信息添加到在操作S1011中所选择的语言模型,并且重新确定包括在操作S1011中被选择的语言模型中的出现概率信息,从而更新第二语言模型。操作S1015可以与图3的操作S311相对应。

[0177] 在操作S1017中,语音识别数据更新设备420可以生成用于将在操作S1003中检测到的字添加到第一语言模型的新字信息。详细地,新字信息可以包括关于通过划分用于更新第二语言模型的新字获取的分量的信息、关于音素序列的信息、情形信息、和关于各个分量的出现概率中的至少一个。如果第二语言模型被重复地更新,则新字信息可以包括关于多个新字的信息。

[0178] 在操作S1019中,语音识别数据更新设备420可以确定是否更新其它模型、发音字典、和第一语言模型中的至少一个。接下来,在操作S1019中,语音识别数据更新设备420可以通过使用在操作S1017中生成的新字信息来更新其它模型、发音字典、和第一语言模型中

的至少一个。其它模型可以包括声学模型,所述声学模型包括用于获取与话音信号相对应的音素序列的信息。显著的 (significant) 时间段可以被消耗来用于更新其它模型、发音字典、和第一语言模型中的至少一个,因为有必要基于关于新字的信息来重新确定包括在各个模型中的数据。因此,语音识别数据更新设备420可以在空闲时隙或者以每周或者每月的间隔来更新整个模型。

[0179] 根据实施例的语音识别数据更新设备420可以实时地更新用于被检测为新字的字的语音识别的第二语言模型。由于在第二语言模型中包括小数量的概率信息,所以更新第二语言模型可以比更新第一语言模型更快,语音识别数据可以被实时地更新。

[0180] 然而,在识别结果的效率和稳定性方面,与通过使用第一语言模型的语音识别相比,通过使用第二语言模型来执行语音识别不是优选的。因此,语音识别数据更新设备420可以通过使用包括在第二语言模型中的出现概率信息来周期性地更新第一语言模型,使得新字可以通过使用第一语言模型被识别。

[0181] 以下,将更详细地描述根据实施例的基于所更新的语音识别数据执行语音识别的方法。

[0182] 图11是示出根据实施例的执行语音识别的语音识别设备的框图。

[0183] 参考图11,根据实施例的语音识别设备1130可以包括语音识别器1131、其它模型1132、发音字典1133、语言模型组合单元1135、第一语言模型1136、第二语言模型1137、和文本恢复单元1138。图11的语音识别设备1130可以与图1、图2、图4、和图9中的语音识别设备100、230、430、和930相对应,其中将省略重复的描述。

[0184] 此外,图11的语音识别器1131、其它模型1132、发音字典1133、语言模型组合单元1135、第一语言模型1136、和第二语言模型1137可以与图1、图2、图4、和图9中以下各项相对应:语音识别单元100、231、431、和931;其它模型232、432、和932;发音字典150、233、433、和933;语言模型组合单元435和935;第一语言模型436和936;以及第二语言模型437和937,其中将省略重复的描述。

[0185] 不同于图1、图2、图4、和图9的语音识别设备100、230、430、和930,图11中所示的语音识别设备1130进一步包括文本恢复单元1138,并且可以在语音识别期间执行文本恢复。

[0186] 语音识别器1131可以获取用于执行语音识别的语音数据1110。语音识别器1131可以通过使用其它模型1132、发音字典1133、和语言模型组合单元1135来执行语音识别。详细地,语音识别器1131可以提取关于话音数据信号的特征信息,并且通过使用声学模型获取与所提取的特征信息相对应的候选音素序列。接下来,语音识别器1131可以从发音字典1133中获取与各个候选音素序列相对应的字。语音识别器1131可以基于从语言模型组合单元1135中获取的关于各个字的出现概率,最终选择与最高出现概率相对应的字并输出语音识别出的语言。

[0187] 文本恢复单元1138可以基于关于构成字的各个分量的出现概率是否用于语音识别,来确定是否执行文本恢复。根据实施例,文本恢复指的是将包括在由语音识别器1131语音识别出的语言中的预定单元分量的字符转换为相应的字。

[0188] 例如,可以基于表明在语音识别期间针对各个子字出现概率被使用的信息、由语音识别器1131生成的信息来确定是否执行文本恢复。另一示例中,文本恢复单元1138可以基于分段信息1126或者发音字典1133,通过从语音识别出的语言中检测子字分量,来确定

是否执行文本恢复。然而,本发明不限于此,并且文本恢复单元1138可以确定是否执行文本恢复和用于针对语音识别出的语言执行文本恢复的部分。

[0189] 在执行文本恢复的情况下,文本恢复单元1138可以基于分段信息1126恢复子字符。例如,如果由语音识别器1131语音识别出的句子是“oneul gi myeo na bo yeo jyo”,则文本恢复单元1138可以确定为了语音识别该句子,针对子字中的每一个出现概率信息是否被使用。此外,文本恢复单元1138可以确定在语音识别出的句子中,出现概率被用于各个子字的部分,即用于文本恢复的部分。文本恢复单元1138可以将“gi”、“myeo”、“na”、“bo”、“yeo”、和“jyo”确定为出现概率被用于各个子字的部分。此外,文本恢复单元1138可以参考存储在分段信息1126中的子字和字之间的对应关系,并通过将“gi myeo na”转换成“giyeon a”以及将“bo yeo jyo”转换成“boyeojyo”来执行文本恢复。文本恢复单元1138可以最终输出包括恢复的文本的语音识别出的语言1140。

[0190] 图12是示出根据实施例的执行语音识别的方法的流程图。

[0191] 参考图12,在操作S1210中,语音识别设备100可以获取用于执行语音识别的语音数据。

[0192] 在操作S1220中,语音识别设备100可以获取包括在语音数据中的至少一个音素序列。详细地,语音识别设备100可以检测关于语音数据的特征信息,并且通过使用声学模型从特征信息中获取音素序列。从特征信息中可以获取至少一个或者更多音素序列。如果基于声学模型从相同的语音数据中获取了多个音素序列,则语音识别设备100可以通过获取关于与多个音素序列相对应的字的出现概率来最终确定语音识别出的字。

[0193] 在操作S1230中,语音识别设备100可以获取关于构成至少一个音素序列的预定单元分量的出现概率信息。详细地,语音识别设备100可以获取关于包括在语言模型中的预定单元分量的出现概率信息。

[0194] 如果不能从语言模型中获取关于构成音素序列的预定单元分量的出现概率信息,则语音识别设备100不能获取关于与相应的音素序列相对应的字的信息。因此,语音识别设备100可以确定不能语音识别出的相应的音素序列,并且针对在操作S1220中获取的关于相同的语音数据的其它音素序列,来执行语音识别。如果不能针对其它音素序列执行语音识别,则语音识别设备100可以确定不能语音识别出该语音数据。

[0195] 在操作S1240中,语音识别设备100可以基于关于构成音素序列的预定单元分量的出现概率信息,来选择至少一个音素序列中的至少一个。例如,语音识别设备100可以基于与构成候选音素序列的子字分量相对应的出现概率信息,从至少一个候选音素序列当中选择与最高概率相对应的音素序列。

[0196] 在操作S1250中,语音识别设备100可以基于包括关于与至少一个预定单元分量相对应的字的信息的分段信息,来获取与在操作S1240中选择的音素序列相对应的字。根据实施例的分段信息可以包括关于与字相对应的预定单元分量的信息。因此,语音识别设备100可以基于分段信息将构成音素序列的子字分量转换为相应的字。语音识别设备100可以将基于分段信息转换的字输出为语音识别的结果。

[0197] 图13是示出根据实施例的执行语音识别的方法的流程图。不同于图12中所示的方法,图13中所示的执行语音识别的方法可以用于基于关于语音数据的情形信息来执行语音识别。图13中所示的方法的操作中的一些可以与图12中所示的方法的操作中的一些相对

应,其中将省略重复的描述。

[0198] 参考图13,在操作S1301中,语音识别设备430可以获取用于执行语音识别的语音数据。操作S1301可以与图12中的操作S1210相对应。

[0199] 在操作S1303中,语音识别设备430可以获取与语音数据相对应的至少一个音素序列。详细地,语音识别设备430可以检测关于语音数据的特征信息,并且通过使用声学模型从特征信息中获取音素序列。如果获取了多个音素序列,则语音识别设备430可以通过基于关于与各个音素序列相对应的子字或者字的出现概率最终确定一个子字或者字来执行语音识别。

[0200] 在操作S1305中,语音识别设备430可以获取关于语音数据的情形信息。语音识别设备430可以考虑到关于语音数据的情形信息,基于关于语音数据的情形信息,通过选择要在语音识别期间应用的语言模型来执行语音识别。

[0201] 根据实施例,关于语音数据的情形信息可以包括关于用户的信息、模块标识信息、和关于设备的位置的信息中的至少一个。在语音识别期间可以选择的语言模型可以包括关于字或者子字的出现概率信息,并且可以与至少一个情形信息相对应。

[0202] 在操作S1307中,语音识别设备430可以确定在发音字典中是否存在关于与在操作S1303中获取的各个音素序列相对应的字的信息。在发音字典中存在关于与音素序列相对应的字的信息的情况下,语音识别设备430可以基于与相应的音素序列相对应的字来针对相应的音素序列执行语音识别。在发音字典中不存在关于与音素序列相对应的字的信息的情况下,语音识别设备430可以基于构成相应的音素序列的子字分量来针对相应的音素序列执行语音识别。发音字典中不存在的字可以是不能够语音识别出的字或者是根据实施例在更新语音识别数据时添加到语言模型中的新字。

[0203] 在与发音字典中存在的信息相对应的音素序列的情况下,语音识别设备100可以通过使用发音字典获取与音素序列相对应的字,并且基于关于字的出现概率信息来最终确定语音识别出的字。

[0204] 在与发音字典中存在的信息相对应的音素序列的情况下,语音识别设备100还可以将音素序列划分成预定单元分量,并且确定关于分量的出现概率信息。换句话说,可以针对与发音字典中存在的信息相对应的音素序列执行操作S1307至S1311和操作S1317至S1319的全部。如果针对音素序列获取了多个出现概率信息,则如下所述,语音识别设备100可以通过组合从多个语言模型中获取的出现概率,来获取关于音素序列的出现概率。

[0205] 以下将在操作S1317至S1321的描述中,详细地描述在发音字典包括关于与音素序列相对应的字的信息的情况下,针对该音素序列执行语音识别的方法。此外,以下将在操作S1309至S1315的描述中,详细地描述在发音字典不包括关于与音素序列相对应的字的信息的情况下,针对该音素序列执行语音识别的方法。

[0206] 在发音字典包括关于与音素序列相对应的字的信息的音素序列的情况下,语音识别设备430可以从在操作S1317中从发音字典获取与各个音素序列相对应的字。发音字典可以包括关于可以与字相对应的至少一个音素序列的信息。可以存在与字相对应的多个音素序列。另一方面,可以存在与音素序列相对应的多个字。关于可以与字相对应的音素序列的信息一般可以基于发音规则来确定。然而,本发明不限于此,并且关于可以与字相对应的音素序列的信息也可以基于用户输入或者学习多个语音数据的结果来确定。

[0207] 在操作S1319中,语音识别设备430可以从第一语言模型中获取关于在操作S1317中获取的字的出现概率信息。第一语言模型可以包括可以用于一般语音识别的一般用途语言模型。此外,第一语言模型可以包括关于包括在发音字典中的字的出现概率信息。

[0208] 如果第一语言模型包括与情形信息相对应的至少一个语言模型,则语音识别设备430可以基于在操作S1305中获取的情形信息,来确定包括在第一语言模型中的至少一个语言模型。接下来,语音识别设备430可以从所确定的语言模型中获取关于在操作S1317中获取的字的出现概率信息。因此,即使在应用第一语言模型的情况下,语音识别设备430也可以通过选择与情形信息相对应的语言模型,基于情形信息来执行自适应语音识别。

[0209] 如果多个语言模型被确定,并且关于字的出现概率信息被包括在所确定的语言模型中的两个或者更多中,则语音识别设备430可以通过组合语言模型来获取关于字的出现概率信息。以下将在操作S1313的描述中给出其详细描述。

[0210] 在操作S1321中,语音识别设备430可以基于关于在操作S1319中获取的出现概率的信息来最终确定语音识别出的字。如果存在可以与相同的语音数据相对应的多个字,则语音识别设备430可以基于关于各个字的出现概率来最终确定并输出语音识别出的字。

[0211] 在发音字典不包括关于与音素序列相对应的字的信息的音素序列的情况下,在操作S1309中,语音识别设备430可以基于在操作S1305中获取的情形信息来确定第二语言模型中的至少一个。语音识别设备430可以包括可以基于情形信息在语音识别期间应用的至少一个独立的第二语言模型。语音识别设备430可以基于情形信息确定多个语言模型。此外,可以在操作S1309中确定的第二语言模型可以包括关于构成音素序列的预定单元分量的出现概率信息。

[0212] 在操作S1311中,语音识别设备430可以确定在操作S1309中确定的第二语言模型是否包括关于构成音素序列的预定单元分量的出现概率信息。如果第二语言模型不包括关于分量的出现概率信息,则关于音素序列的出现概率信息不能被获取,因此语音识别不再能够被执行。如果存在与相同的语音数据相对应的多个音素序列,则语音识别设备430可以在操作S1307中确定与除了关于不能获取与那些音素序列的出现概率有关的信息的那些音素序列的音素序列相对应的字是否在发音字典中存在。

[0213] 在操作S1313中,语音识别设备430可以基于关于包括在操作S1309中确定的第二语言模型中的预定单元分量的出现概率信息,来确定至少一个音素序列中的一个。详细地,语音识别设备430可以从第二语言模型中获取关于构成音素序列的预定单元分量的出现概率信息。接下来,语音识别设备430可以基于关于预定单元分量的出现概率信息,来确定与最高出现概率相对应的音素序列。

[0214] 当多个语言模型在操作S1309或者操作S1319中被选择时,关于预定单元分量或者字的出现概率信息可以被包括在两个或者更多语言模型中。可以被选择的多个语言模型可以包括第一语言模型和第二语言模型中的至少一个。

[0215] 例如,当语音识别数据被更新时,如果基于情形信息将新字添加到两个或者更多语言模型中,则关于相同的字或者子字的出现概率信息可以被添加到两个或者更多语言模型。另一示例中,当语音识别数据被周期性地更新时,如果将仅仅在第二语言模型中存在的字添加到第一语言模型,则关于相同字或者子字的出现概率信息可以被包括在第一语言模型和第二语言模型中。语音识别设备430可以通过组合语言模型来获取关于预定单元分量

或者字的出现概率。

[0216] 当有关于单一字或者分量的多个出现概率信息作为多个语言模型被选择时,语音识别设备430的语言模型组合单元435可以获取单一出现概率。

[0217] 例如,如下方程式1所示,语言模型组合单元435可以通过获取关于各个出现概率的权重的和来获取单一出现概率。

[0218] [方程式1]

$$P(a|b) = \omega_1 P_1(a|b) + \omega_2 P_2(a|b) \quad (\omega_1 + \omega_2 = 1)$$

[0220] 在方程式1中, $P(a|b)$ 表示在b在a之前出现的条件下关于a的出现概率。 P_1 和 P_2 分别表示关于a包括在第一语言模型和第二语言模型中的出现概率。 ω_1 和 ω_2 分别表示可应用到 P_1 和 P_2 的权重。方程式1的右侧分量的数量可以根据包括关于a的出现概率信息的语言模型的数量而增加。

[0221] 可以应用到各个出现概率的权重可以基于情形信息或者各种其它条件(例如,关于用户、区域、命令历史、正在执行的模块的信息等)来确定。

[0222] 根据方程式1,随着关于出现概率的信息被包括在更多语言模型中,出现概率可以增加。相反地,随着关于出现概率的信息包括在更少的语言模型中,出现概率可以减小。因此,在根据方程式1确定出现概率的情况下,优选的出现概率可以不被确定。

[0223] 语言模型组合单元435可以基于贝叶斯插值根据方程式2来获取关于字或者子字的出现概率。在根据方程式2确定出现概率的情况下,根据包括出现概率信息的语言模型的数量,出现概率可以不增加或者减小。在出现概率仅包括在第一语言模型或者第二语言模型中的情况下,出现概率可以不减小并且可以根据方程式2被保持。

[0224] [方程式2]

$$P(a|b) = \frac{\omega_1 P_1(b) P_1(a|b) + \omega_2 P_2(b) P_2(a|b)}{\omega_1 P_1(b) + \omega_2 P_2(b)} \quad (\omega_1 + \omega_2 = 1)$$

[0226] 此外,语言模型组合单元435可以根据方程式3获取出现概率。根据方程式3,出现概率信息可以是包括在各个语言模型中的出现概率当中最大的一个。

[0227] [方程式3]

$$P(a|b) = \max \{P_1(a|b), P_2(a|b)\}$$

[0229] 在根据方程式3确定出现概率的情况下,出现概率可以是出现概率当中最大的一个,因此关于包括在语言模型中的每一个中一次或者多次的字或者子字的出现概率可以具有相对大的值。因此,根据方程式3,根据实施例的关于作为新字添加到语言模型中的字的出现概率可能被错误地减小。

[0230] 在操作S1315中,语音识别设备430可以基于分段信息来获取与在操作S1313中确定的音素序列相对应的字。分段信息可以包括关于构成音素序列的至少一个单元分量和字之间的对应关系的信息。如果新字按照根据实施例的更新语音识别数据的方法被检测出,则关于每个字的信息可以被生成为关于新字的信息。如果音素序列基于概率信息被确定为语音识别的结果,则语音识别设备430可以基于分段信息将音素序列转换为字,因此语音识别的结果可以被输出为字。

[0231] 图14是示出根据实施例的基于以情形信息为基础执行的语音识别的结果来执行模块的语音识别系统的框图。

[0232] 参考图14,语音识别系统1400可以包括语音识别数据更新设备1420、语音识别设备1430、和用户设备1450。如图14中所示,语音识别数据更新设备1420、语音识别设备1430、和用户设备1450可以作为独立的设备存在。然而,本发明不限于此,并且语音识别数据更新设备1420、语音识别设备1430、和用户设备1450可以被包括在单一设备中作为该设备的组件。图14的语音识别数据更新设备1420和语音识别设备1430可以与上述参考图13的语音识别数据更新设备220和420,以及语音识别设备230和430相对应,其中将省略重复的描述。

[0233] 首先,将会通过使用图14中所示的语音识别系统1400描述考虑到情形信息来更新语音识别数据的方法。

[0234] 语音识别数据更新设备1420可以获取用于更新语音识别数据的语言数据1410。语言数据1410可以从各种设备中获取,并且可以被发送到语音识别数据更新设备1420。例如,语言数据1410可以由用户设备1450获取,并且被发送到语音识别数据更新设备1420。

[0235] 此外,用户设备1450的情形信息管理单元1451可以获取与语言数据1410相对应的情形信息,并且将获取到的情形信息发送到语音识别数据更新设备1420。语音识别数据更新设备1420可以基于从情形信息管理单元1451接收的情形信息,来确定要添加包括在语言数据1410中的新字的语言模型。如果没有与情形信息相对应的语言模型存在,则语音识别数据更新设备1420可以生成新的语言模型,并且将关于新字的出现概率信息添加到新生成的语言模型中。

[0236] 语音识别数据更新设备1420可以检测包括在语言数据1410中的新字“Let it go”和“bom bom bom”。与语言数据1410相对应的情形信息可以包括用于音乐回放的应用A。情形信息可以针对语言数据1410被确定,或者也可以针对包括在语言数据1410中的新字中的每一个被确定。

[0237] 语音识别数据更新设备1420可以将关于“Let it go”和“bom bom bom”的出现概率信息添加到与应用A相对应的至少一个语言模型中。语音识别数据更新设备1420可以通过将关于新字的出现概率信息添加到与情形信息相对应的语言模型来更新语音识别数据。语音识别数据更新设备1420可以通过重新确定包括在关于新字的出现概率信息被添加到语言模型中的出现概率信息来更新语音识别数据。出现概率信息可以添加到语言模型可以与一个应用或者包括至少一个应用的组相对应。

[0238] 语音识别数据更新设备1420可以基于用户输入实时地更新语言模型。涉及根据实施例的语音识别设备1430,用户可以根据由用户定义的语言向应用或者应用组发出话音命令。如果在语言模型中仅仅关于命令“播放[歌曲]”的出现概率存在,则可以基于用户定义将关于命令“让我听[歌曲]”的出现概率信息添加到语言模型中。

[0239] 然而,如果能够基于用户定义确定语言,则当由另一用户定义的语言被应用时,意料外的话音命令可以被执行。因此,语音识别数据更新设备1420可以为语言模型的应用设置应用或者时间,以作为用于应用基于用户定义确定的语言模型的范围。

[0240] 语音识别数据更新设备1420可以基于从用户设备1450的情形信息管理单元1451接收的情形信息,来实时地更新语音识别数据。如果用户设备1450位于电影院附近,则用户设备1450可以将关于相应的电影院的信息作为情形信息发送到语音识别数据更新设备1420。关于电影院的信息可以包括关于在相应的电影院正在播放的电影的信息、关于电影院附近的餐馆的信息、交通信息等。语音识别数据更新设备1420可以经由网络信息采集或

者从内容提供者来收集关于相应的电影院的信息。接下来,语音识别数据更新设备1420可以基于所收集的信息来更新语音识别数据。因此,由于语音识别设备1430可以考虑到用户设备1450的位置来执行语音识别,所以语音识别效率可以被进一步提高。

[0241] 其次,将描述在语音识别系统1400处执行语音识别和基于语音识别的结果来执行模块的方法。

[0242] 用户设备1450可以包括可以由用户使用的各种类型的终端设备。例如,用户设备1450可以是移动电话、智能电话、膝上型计算机、平板型计算机、电子书终端、数字广播设备、个人数字助理(personal digital assistant,PDA)、可携带的多媒体播放器(portable multimedia player,PMP)、导航设备、MP3播放器、数码相机、或者可穿戴设备(例如,眼镜、手表、戒指等)。然而,本发明不限于此。

[0243] 根据实施例的用户设备1450可以收集涉及语音数据1440和用户设备1450的情形信息中的至少一个,并且基于以情形信息为基础的语音识别出的语音识别的字来执行所确定的任务。

[0244] 用户设备1450可以包括情形信息管理单元1451、模块选择和指示单元1452、以及用于基于语音识别结果执行任务的应用A 1453。

[0245] 情形信息管理单元1451可以在语音识别设备1430处收集在语音识别期间用于选择语言模型的情形信息,并且将情形信息发送到语音识别设备1430。

[0246] 情形信息可以包括关于在用户设备1450上当前正执行的模块的信息、使用模块的历史、语音命令的历史、关于可以在用户设备1450上执行并且与现有的语言模型相对应的应用的信息、关于当前正在使用用户设备1450的用户的信息等。使用模块的历史和语音命令的历史可以分别包括关于使用各个模块的时间点和接收各个语音命令的时间点的信息。

[0247] 根据实施例的情形信息可以如下表1中所示来配置。

[0248] **【表1】**

	情形信息
当前使用的模块	电影播放器模块 1
模块使用的历史	音乐播放器模块器 1/1 天前 有线广播/ 1 小时前 音乐播放器模块 1/30 分钟前
[0249] 语音命令的历史	家庭影院播放[歌手 1]歌曲/ 10 分钟前 音乐播放器模块 1/30 分钟前
具有语言的应用	广播 音乐播放器模块 1 电影播放器模块 1 音乐播放器模块 2

[0250] 语音识别设备1430可以基于情形信息来选择要在语音识别期间使用的至少一个语言模型。如果情形信息表明在应用A被执行的同时从用户设备1450获取到语音数据1440,

则语音识别设备1430可以选择与应用A和用户设备1450中的至少一个相对应的语言模型。

[0251] 模块选择和指示单元1452可以基于由语音识别设备1430执行的语音识别的结果来选择模块,并且向所选择的模块发送执行任务的命令。首先,模块选择和指示单元1452可以确定语音识别的结果是否包括模块的标识符和用于命令的关键字。用于命令的关键词可以包括表明用于请求模块执行各个任务的命令(例如,播放、暂停、下一个等)的标识符。

[0252] 如果模块标识符被包括在语音识别结果中,则模块选择和指示单元1452可以选择与模块标识符相对应的模块,并且向所选择的模块发送命令。

[0253] 如果模块标识符没有被包括在语音识别结果中,则模块选择和指示单元1452可以获取包括在语音识别结果中的用于命令的关键字和与语音识别结果相对应的情形信息中的至少一个。基于用于命令的关键字和情形信息中的至少一个,模块选择和指示单元1452可以根据语音识别的结果来确定用于执行任务的模块。

[0254] 详细地,模块选择和指示单元1452可以基于用于命令的关键字来确定用于执行任务的模块。此外,模块选择和指示单元1452可以基于情形信息确定最适合于执行任务的模块。例如,模块选择和指示单元1452可以基于执行频率或者相应的模块是否是最近执行的模块来确定模块。

[0255] 可以由模块选择和指示单元1452收集的情形信息可以包括关于在用户设备1450上当前正执行的模块的信息、使用模块的历史、语音命令的历史、关于与现有的模块相对应的应用的信息等。使用模块的历史和语音命令的历史可以包括关于使用模块的时间点的信息和接收语音命令的时间点的信息。

[0256] 即使语音识别的结果包括模块标识符,相应的模块也可能不能根据命令来执行任务。如同在语音识别的结果不包括模块标识符的情况下,模块选择和指示单元1452可以确定要执行任务的模块。

[0257] 参考图14,模块选择和指示单元1452可以从语音识别设备1430接收“让我听Let it go”作为语音识别的结果。由于语音识别的结果不包括应用标识符,所以用于基于语音识别的结果执行任务的应用A可以基于情形信息或者用于命令的关键字被确定。模块选择和指示单元1452可以请求应用A回放歌曲“Let it go”。

[0258] 图15是示出根据实施例的关于模块的情形信息的示例的示意图。

[0259] 参考图15,示出了用于基于语音命令执行任务的音乐播放器程序1510的命令的示例。语音识别数据更新设备1520可以与图14的语音识别数据更新设备1420相对应。

[0260] 语音识别数据更新设备1520可以从用户设备1450接收关于音乐播放器程序1510的情形信息,并且基于接收到的情况信息来更新语音识别数据。

[0261] 如图15中所示,关于音乐播放器程序1510的情形信息可以包括标头1511、命令语言1512、和音乐信息1513。

[0262] 标头1511可以包括用于标识音乐播放器程序1510的信息,并且可以包括关于音乐播放器程序1510的类型、存储位置、和名称的信息。

[0263] 命令语言1512可以包括关于音乐播放器程序1510的命令的示例。当接收到像命令语言1512的语音识别出的句子时,音乐播放器程序1510可以执行任务。命令语言1512的命令也可以由用户设置。

[0264] 音乐信息1513可以包括关于可以由音乐播放器程序1510回放的音乐的信息。例

如,音乐信息1513可以包括关于可以由音乐播放器程序1510回放的音乐文件的标识信息及其分类信息(诸如关于专辑和歌手的信息)。

[0265] 语音识别数据更新设备1520可以通过使用命令语言1512的句子和包括在音乐信息1513中的字,来更新关于音乐播放器程序1510的第二语言模型。例如,语音识别数据更新设备1520可以通过将包括在音乐信息1513中的字包括在命令语言1512的句子中来获取出现概率信息。

[0266] 当新的应用被安装时,根据实施例的用户设备1450可以向语音识别数据更新设备1520发送关于应用的信息,所述信息包括标头1511、命令语言1512、和音乐信息1513。此外,当关于应用的新事件发生时,用户设备1450可以更新关于应用的信息,所述信息包括标头1511、命令语言1512、和音乐信息1513,并且将更新的信息发送到语音识别数据更新设备1520。因此,语音识别数据更新设备1520可以基于关于应用的最新信息来更新语言模型。

[0267] 当语音识别设备1430执行语音识别时,用户设备1450可以向语音识别设备1430发送用于执行语音识别的情形信息。情形信息可以包括图15中所示的关于音乐播放器程序的信息。

[0268] 情形信息可以如下表2中所示来配置。

[0269] 【表2】

	情形信息
当前使用的模块	备忘录
命令历史	音乐播放器模块 3 播放[歌曲标题]1/10 分钟前 音乐播放器模块 3 播放[歌手 1]歌曲/15 分钟前
同时模块使用的历史	备忘录-音乐播放器模块 3 / 1 天前 备忘录-音乐播放器模块 3 / 2 天前
模块信息	音乐播放器模块 1 [歌手 1-3] N 首歌 音乐播放器模块 2 [歌手 3-6] N 首歌 音乐播放器模块 3 [歌手 6-8] N 首歌
SNS 历史	音乐播放器模块 1 陈述一次 音乐播放器模块 2 陈述四次 音乐播放器模块 N 陈述两次

[0270] 语音识别设备1430可以基于从表2中所示的情形信息当中的同时模块使用的历史,来确定对与各个音乐播放器程序相对应的语言模型可应用的权重。如果备忘录程序当前正在被执行,则语音识别设备1430可以通过将权重应用到与已经跟备忘录程序同时使用的音乐播放器程序相对应的语言模块,来执行语音识别。

[0271] 当从用户接收到语音输入时,如果通过语音识别设备1430执行的语音识别的结果被输出为“播放所有[歌手3]歌曲”,则模块选择和指示单元1452可以确定要执行相应任务的模块。由于语音识别出的命令不包括模块标识符,所以模块选择和指示单元1452可以基于命令和情形信息来确定要执行相应任务的模块。详细地,考虑到包括同时模块使用的历史、最近模块使用的历史、和包括在情形信息中的SNS使用的历史的各种信息,模块选择指

示单元1452可以根据命令来选择回放音乐的模块。参考表1,从能够回放[歌手3]的歌曲的音乐播放器模块1和2之间,在SNS上提到的音乐播放器模块2的次数大于音乐播放器模块1的次数,模块选择和指示单元1452可以选择音乐播放器模块2。由于命令不包括模块标识符,所以模块选择和指示单元1452可以基于用户输入来最终决定是否通过使用所选择的音乐播放器模块2播放音乐。

[0273] 模块选择和指示单元1452可以根据语音识别出的命令来针对多个模块请求执行多个任务。假设情形信息如下表3所示来配置。

[0274] 【表3】

	情形信息
当前使用的模块	主屏幕
[0275] 命令历史	音乐播放器模块 3 播放[歌曲] / 10 分钟前 我将写备忘录 / 20 分钟前
对使用模块的使用设置的历史	电影播放器模块-音量 1/1 天前 电影播放器模块-增加亮度 / 1 天前

[0276] 如果语音识别出的命令是“给我显示[电影]”,则模块选择和指示单元1452可以选择能够回放[电影]的电影播放器模块作为执行相应任务的模块。模块选择和指示单元1452可以基于从情形信息当中的关于对使用模块的使用设置的历史的信息,来确定除了电影播放器模块的执行命令的多个模块。

[0277] 详细地,模块选择和指示单元1452可以基于关于对使用模块的使用设置的历史的信息,来选择用于调节音量和照度的音量调节模块和照度调节模块。接下来,模块选择和指示单元1452可以基于关于对使用模块的使用设置的历史的信息,向所选择的模块发送对调整音量和照度的请求。

[0278] 图16是示出根据实施例的执行语音识别的方法的示例的流程图。

[0279] 参考图16,在操作1610中,语音识别设备1430可以获取要执行语音识别的语音数据。

[0280] 在操作1620中,语音识别设备1430可以获取关于语音数据的情形信息。如果用于音乐回放的应用A在获取语音数据的用户设备1450上正在被执行,则情形信息可以包括表明应用A正在被执行的情形信息。

[0281] 在操作1630中,语音识别设备1430可以基于在操作1620中获取的情形信息来确定至少一个语言模型。

[0282] 在操作1640和1670中,语音识别设备1430可以获取与语音数据的相对应音素序列。与包括语音“Let it go”的语音数据相对应的音素序列可以包括音素序列“leritgo”和“naerigo”。此外,与包括语音“dulryojoyo”的语音数据相对应的音素序列可以包括音素序列“dulryojoyo”和“dulyeojoyo”。

[0283] 如果在所获取的音素序列中存在与发音字典相对应的字,则语音识别设备1430可以将音素序列转换为字。此外,没有与发音字典相对应的字的音素序列可以被划分成预定单元分量。

[0284] 从音素序列当中,由于在发音字典中不存在与音素序列“leritgo”相对应的字,所

以音素序列“leritgo”可以被划分成预定单元分量。此外,关于来自音素序列当中的音素序列“naerigo”,发音字典中的对应字“naerigo”和预定单元分量“nae ri go”可以被获取。

[0285] 由于在发音字典中存在与音素序列“dulryojyo”和“dulyeojyo”相对应的字,因此音素序列“dulryojyo”和“dulyeojyo”可以被获取。

[0286] 在操作1650中,语音识别设备1430可以基于出现概率信息从“le rit go”、“naerigo”、and“nae ri go”当中确定“le rit go”。此外,在操作1680中,语音识别设备1430可以基于出现概率信息从“dulryojyo”和“dulyeojyo”之间确定“dulryojyo”。

[0287] 从音素序列当中,有两个关于音素序列“naerigo”的出现概率信息,因此关于音素序列“naerigo”的出现概率可以通过组合如上所述的语言模型来确定。

[0288] 在操作1660中,语音识别设备1430可以基于分段信息将“le rit go”恢复为原始字“Let it go”。由于“dulryojyo”不是划分的字,并且分段信息不包括关于“dulryojyo”的信息,所以不可以在其上执行像操作1660的操作。

[0289] 在操作1690中,语音识别设备1430可以输出“Let it go dulryojyo”作为语音识别的最终结果。

[0290] 图17是示出根据实施例的执行语音识别的方法的示例的流程图。

[0291] 参考图17,在操作1710中,语音识别设备1430可以获取要执行语音识别的语音数据。

[0292] 在操作1703中,语音识别设备1430可以获取关于语音数据的情形信息。在操作1730中,语音识别设备1430可以基于在操作1720中获取的情形信息来确定至少一个语言模型。

[0293] 在操作1707、1713、和1719中,语音识别设备1430可以获取与语音数据相对应的音素序列。与包括语音“oneul”和“gim yeon a”的语音数据相对应的音素序列可以分别包括“oneul”和“gi myeo na”。此外,与包括语音“boyeojyo”的语音数据相对应的音素序列可以包括“boyeojeo”和“boyeojyo”。然而,不限于上述音素序列,根据语音数据可以获取与示例不同的音素序列。

[0294] 在操作1707中,语音识别设备1430可以通过使用发音字典来获取与音素序列“oneul”相对应的字“oneul”。在操作1713中,语音识别设备1430可以通过使用发音字典来获取与音素序列“gi myeo na”相对应的字“gim yeon a”。

[0295] 此外,在操作1713和1719中,语音识别设备1430可以将“gimyeona”、“boyeojyo”、和“boyeojeo”划分成指定的单元分量,并且分别获取“gi myeo na”、“bo yeo jyo”和“bo yeo jeo”。

[0296] 在操作1709、1715、和1721中,语音识别设备1430可以基于出现概率信息来确定“oneul”、“gi myeo na”、和“bo yeo jeo”。从音素序列当中,可以存在涉及“gi myeo na”的两个出现概率信息,因此关于“gi myeo na”的出现概率可以通过组合如上所述的语言模型被确定。

[0297] 在操作1717和1723中,语音识别设备1430可以基于分段信息来恢复原始字“gimyeona”和“boyeojyo”。由于“oneul”不是划分成预定单元分量的字,并且分段信息不包括“oneul”,所以不可以执行恢复操作。

[0298] 在操作1725中,语音识别设备1430可以输出“oneul gimyeona boyeojyo”作为语

音识别的最终结果。

[0299] 图18是示出根据实施例的基于情形信息根据执行的语音识别的结果来执行多个模块的语音识别系统的框图。

[0300] 参考图18,语音识别系统1800可以包括语音识别数据更新设备1820、语音识别设备1830、用户设备1850、以及外部设备1860和1870。语音识别数据更新设备1820、语音识别设备1830、和用户设备1850可以被具体化为如图18中所示的独立设备。然而,本发明不限于此,并且语音识别数据更新设备1820、语音识别设备1830、和用户设备1850可以作为设备的组件被嵌入在单一设备中。图18的语音识别数据更新设备1820和语音识别设备1830可以与上述参考图1至17的语音识别数据更新设备220和420、以及语音识别设备230和430相对应,其中将在下面省略对其的重复的描述。

[0301] 首先,将描述通过使用图18中所示的语音识别系统1800考虑到情形信息来更新语音识别数据的方法。

[0302] 语音识别数据更新设备1820可以获取用于更新语音识别数据的语言数据1810。此外,用户设备1850的情形信息管理单元1851可以获取关于与语言数据1810相对应的信息,并且将所获取的情形信息发送到语音识别数据更新设备1820。语音识别数据更新设备1820可以基于从情形信息管理单元1851接收的情形信息,来确定添加包括在语言数据1810中的新字的语言模型。

[0303] 语音识别数据更新设备1820可以检测包括在语言数据1810中的新字“冬季王国”和“5.1频道”。关于字“冬季王国”的情形信息可以包括关于涉及用于电影回放的数字通用盘(digital versatile disc,DVD)播放器设备1860的信息。此外,关于字“5.1频道”的情形信息可以包括关于用于音频输出的家庭影院设备1870的信息。

[0304] 语音识别数据更新设备1820可以将关于“冬季王国”和“5.1频道”的出现概率信息添加到分别与DVD播放器设备1860和家庭影院设备1870相对应的至少一个或者更多语言模型。

[0305] 其次,将描述图18中所示的语音识别系统1800执行语音识别以及每个设备基于语音识别的结果执行任务的方法。

[0306] 用户设备1850可以包括可以由用户使用的各种类型的终端。

[0307] 根据实施例的用户设备1850可以收集语音数据1840和关于用户设备1850的情形信息中的至少一个。接下来,用户设备1850可以基于情形信息请求至少一个设备执行根据语音识别出的语言确定的任务。

[0308] 用户设备1850可以包括情形信息管理单元1851以及模块选择和指示单元1852。

[0309] 情形信息管理单元1851可以收集用于选择用于由语音识别设备1830执行的语音识别的语言模型的情形信息,并且将情形信息发送到语音识别设备1830。

[0310] 语音识别设备1830可以基于情形信息来选择要用于语音识别的至少一个语言模型。如果情形信息包括表明DVD播放器设备1860和家庭影院设备1870可以被使用的信息,则语音识别设备1830可以选择与DVD播放器设备1860和家庭影院设备1870相对应的语言模型。可替换地,如果语音信号包括模块标识符,则语音识别设备1830可以选择与模块标识符相对应的语言模型并且执行语音识别。模块标识符可以包括用于识别不仅模块还有模块组或者模块类型的信息。

[0311] 模块选择和指示单元1852可以基于由语音识别设备1830执行的语音识别的结果，来确定向其发送命令的至少一个设备，并且向所确定的设备发送命令。

[0312] 如果语音识别的结果包括用于标识设备的信息，则模块选择和指示单元1852可以向与标识信息相对应的设备发送命令。

[0313] 如果语音识别的结果不包括用于标识设备的信息，则模块选择和指示单元1852可以获取包括在语音识别结果中的用于命令的关键字和情形信息中的至少一个。模块选择和指示单元1852可以基于用于命令的关键字和情形信息中的至少一个，来确定用于向其发送命令的至少一个设备。

[0314] 参考图18，模块选择和指示单元1852可以从语音识别设备1830接收“给我显示5.1频道中的冬王国”作为语音识别的结果。由于语音识别的结果不包括设备标识符或者应用标识符，所以向其发送命令的DVD播放器设备1860和家庭影院设备1870可以基于情形信息或者用于命令的关键字确定。

[0315] 详细地，模块选择和指示单元1852可以从当前可用的设备当中确定能够在5.1频道中输出声音并能够输出运动图像的多个设备。模块选择和指示单元1852可以基于情形信息（诸如各个设备的使用的历史），从多个所确定的设备当中最终确定用于执行命令的设备。

[0316] 可以由情形信息管理单元1851获取的情形信息可以如下表4中所示来配置。

[0317] 【表4】

	情形信息
当前使用的模块	TV 广播模块
同时模块使用的历史	TV 广播模块-家庭影院设备/ 20 分钟前 DVD 播放器设备-家庭影院设备/ 1 天前
[0318] 语音命令的历史	家庭影院播放[歌手 1]歌曲/ 10 分钟前 DVD 播放器播放[电影 1]/ 1 天前
具有语言的应用	TV 广播模块 DVD 播放器设备 电影播放器模块 1 家庭影院设备

[0319] 接下来，模块选择和指示单元1852可以向最终确定的设备发送命令。详细地，基于语音“给我显示5.1频道中的冬季王国”的识别的结果，模块选择和指示单元1852可以向DVD播放设备1860发送请求回放“冬季王国”的命令。此外，模块选择和指示单元1852可以向家庭影院设备1870发送请求输出5.1频道中的“冬季王国”的声音信号命令。

[0320] 因此，根据实施例，基于语音识别的单一结果，命令可以被发送到多个设备或者模块，并且多个设备或者模块可以同时执行任务。此外，即使语音识别的结果不包括模块和设备标识符，根据实施例的模块选择和指示单元1852也可以基于用于命令的关键字和情形信息来确定用于执行任务的最合适的模块或者设备。

[0321] 图19是示出根据实施例的针对多个设备的话音命令的示例的示意图。

[0322] 参考图19，基于模块选择和指示单元1922，示出了用于能够根据话音命令执行任

务的设备的命令的示例。模块选择和指示单元1922可以与图18的模块选择和指示单元1852相对应。此外,DVD播放器设备1921和家庭影院设备1923可以分别与图18中的DVD播放器设备1860和家庭影院设备1870相对应。

[0323] 语音指令1911是根据实施例的可以基于语音识别输出的语音识别的结果的示例。如果语音指令1911包括视频的名称和5.1频道,则模块选择和指示单元1922可以选择能够回放视频的DVD播放器设备1921和家庭影院设备1923作为用于向其发送命令的设备。

[0324] 如图19中所示,模块选择和指示单元1922可以在关于DVD播放器设备1921和家庭影院设备1923的信息中包括标头1931和1934、命令语言1932和1935、视频信息1933、以及声音预设1936。

[0325] 标头1931和1934可以分别包括用于标识DVD播放器设备1921和家庭影院设备1923的信息。标头1931和1934可以包括信息,所述信息包括各个设备的类型、位置、和名称。

[0326] 命令语言1932和1935可以包括针对设备1921和1923的命令的示例。当接收到与命令语言1932和1935相同的话音时,各个设备1921和1923可以执行与接收到的命令相对应的任务。

[0327] 视频信息1933可以包括关于可以由DVD播放器设备1921回放的视频的信息。例如,视频信息1933可以包括关于可以由DVD播放器设备1921回放的视频文件的标识信息和详细信息。

[0328] 声音预设1936可以包括关于与家庭影院设备1923的声音输出有关的可用设置的信息。如果家庭影院设备1923可以被设置为7.1频道、5.1频道、和2.1频道,则声音预设1936可以包括7.1频道、5.1频道、和2.1频道,作为关于与家庭影院设备1923的频道有关的可用设置的信息。除了频道,声道预设1936可以包括均衡器设置、音量设置等,并且基于用户设置可以进一步包括关于针对家庭影院设备1923的各种可用设置的信息。

[0329] 模块选择和指示单元1922可以将关于DVD播放器设备1921和家庭影院设备1923的信息1931至1936发送到语音识别数据更新设备1820。语音识别数据更新设备1820可以基于接收到的信息1931至1936更新与各个设备1921和1923相对应的第二语言模型。

[0330] 语音识别数据更新设备1820可以通过使用包括在命令语言1932和1935的句子、视频信息1933、或者声音预设1936中的字,来更新与各个设备1921和1923相对应的语言模型。例如,语音识别数据更新设备1820可以将包括在视频信息1933或者声音预设1936中的字包括在命令语言1932和1935中,并且获取关于所述字的出现概率信息。

[0331] 图20是示出根据实施例的语音识别设备的示例的框图。

[0332] 参考图20,语音识别设备2000可以包括前端引擎2010和语音识别引擎2020。

[0333] 前端引擎2010可以从语音识别设备2000接收语音数据或者语言数据,并且输出关于语音数据的语音识别的结果。此外,前端引擎2010可以针对所接收的语音数据或者语言数据执行预处理,并且将预处理的语音数据或者语言数据发送到语音识别引擎2020。

[0334] 前端引擎2010可以与参考图1至图17的上述语音识别数据更新设备220和420相对应。语音识别引擎2020可以与参考图1至图18的上述语音识别设备230和430相对应。

[0335] 由于更新语音识别数据和语音识别可以由独立的引擎分别执行,所以语音识别和更新语音识别可以在语音识别设备2000中同时执行。

[0336] 前端引擎2010可以包括用于接收语音数据并将语音数据发送到语音识别器2022

的语音缓冲器2011、以及用于更新语音识别的语言模型更新单元2012。此外,根据实施例,前端引擎2010可以包括分段信息2013,所述分段信息包括用于将语音识别出的子字恢复成字的信息。前端引擎2010可以通过使用分段信息2013来将由语音识别器2022语音识别出的子字恢复成字,并且将包括恢复出的字的语音识别出的语言2014输出为语音识别的结果。

[0337] 语音识别引擎2020可以包括由语言模型更新单元2012更新的语言模型2021。此外,语音识别引擎2020可以包括能够基于从语音缓冲器2011接收的语音数据和语言模型2021来执行语音识别的语音识别器2022。

[0338] 当随着执行记录而输入语音数据时,语音识别设备2000可以同时收集包括新字的语言数据。接下来,由于包括记录的语音的语音数据存储于语音缓冲器2011中,所以语言模型更新单元2012可以通过使用新字来更新语言模型2021的第二语言模型。当第二语言模型被更新时,语音识别器2022可以接收存储在语音缓冲器2011中的语音数据,并且执行语音识别。语音识别出的语言可以被发送到前端引擎2010,并且基于分段信息2013来恢复。前端引擎2010可以输出包括恢复出的字的语音识别的结果。

[0339] 图21是示出根据实施例的在显示设备上执行语音识别的示例的框图。

[0340] 参考图21,显示设备2110可以从用户接收语音数据、将语音数据发送到语音识别服务器2120、从语音识别服务器2120接收语音识别的结果、并且输出语音识别的结果。显示设备2110可以基于语音识别的结果执行任务。

[0341] 显示设备2110可以包括用于生成用于在语音识别服务器2120上更新语音识别数据的语言数据的语言数据生成单元2114。语言数据生成单元2114可以从当前显示在显示设备2110上的信息或者涉及当前显示在显示设备2110上的信息的内容信息来生成语言数据,并且将语言数据发送到语音识别服务器2120。例如,语言数据生成单元2114可以从包括在当前显示的、先前显示的、或者将会显示的内容中的文本2111和当前广播的信息2112生成语言数据。此外,语言数据生成单元2114可以从会话管理单元2113接收关于显示在显示设备2110上的会话的信息,并且通过使用接收到的信息生成语言数据。可以从会话管理单元2113接收的信息可以包括:包括在社交网络服务(social network service, SNS)中的文本、包括在短消息服务(short message service, SMS)中的文本、包括在多媒体消息服务(multimedia message service, MMS)中的文本、以及关于在显示设备2110和用户之间的会话的信息。

[0342] 语言模型更新单元2121可以通过使用从显示设备2110的语言数据生成单元2114接收的语言数据来更新语言模型。接下来,语音识别单元2122可以基于更新的语言模型来执行语音识别。如果语音识别出的语言包括子字,则根据实施例,文本恢复单元2123可以基于分段信息执行文本恢复。语音识别服务器2120可以将文本恢复出和语音识别出的语言发送到显示设备2110,并且显示设备2110可以输出语音识别出的语言。

[0343] 在根据实施例的通过将新字划分成预定单元分量来更新语音识别数据的情况下,显示设备2110可以在几毫秒内更新语音识别数据。因此,语音识别服务器2120可以立即将显示在显示设备2110上的文本中的新字添加到语言模型。

[0344] 用户不仅可以讲出设置命令,还可以讲出正在广播的广播节目的名称或者在显示设备2110上显示的文本。因此,根据实施例的语音识别服务器2120可以接收显示设备2110上显示的文本或者很可能被讲出的关于显示设备2110上显示的内容的信息。接下来,语音

识别服务器2120可以基于接收到的信息来更新语音识别数据。由于语音识别服务器2120能够在几毫秒到几秒内更新语言模型,所以一旦新字被获取,很可能被讲出的新字就可以被处理以被识别。

[0345] 图22是示出根据实施例的考虑到情形信息来更新语言模型的示例的框图。

[0346] 图22的语音识别数据更新设备2220和语音识别设备2240可以分别与图2至17所示的语音识别数据更新设备220和420、以及语音识别设备23和430相对应。

[0347] 参考图22,语音识别数据更新设备2220可以从用户设备2210或者服务提供服务器2230中获取个性化信息2221。

[0348] 语音识别数据更新设备2220可以包括关于来自用户设备2210的用户的信息,所述信息包括:地址簿2211、安装的应用列表2212、和存储的专辑列表2213。然而,本发明不限于此,并且语音识别数据更新设备2220可以从用户设备2210接收关于用户设备2210的各种信息。

[0349] 由于个体用户彼此具有不同的发音模式,所以语音识别数据更新设备2220可以周期性地接收用于为用户中的每一个执行语音识别的信息,并且将所述信息存储在个性化信息2221中。此外,语音识别数据更新设备2220的语言模型更新单元2222可以基于各个用户的个性化信息2221来更新语言模型。此外,语音识别数据更新设备2220可以从服务提供服务器2230收集关于涉及各个用户的收集到的服务使用的信息,并且将所述信息存储在个性化信息2221中。

[0350] 服务提供服务器2230可以包括对于每个用户的优选频道列表2231、经常观看的视频点播(video-on-demand,VOD) 2232、会话历史2233、和语音识别结果历史2234。换句话说,服务提供服务器2230可以存储关于提供给用户设备2210的服务,例如,广播节目提供服务、VOD服务、SNS服务、语音识别服务等,的信息。可收集的信息仅仅是示例,并不限于此。服务提供服务器2230可以收集关于用户中的每一个的各种信息,并且将所收集的信息发送到语音识别数据更新设备2220。语音识别结果历史2234可以包括关于由语音识别设备2240针对各个用户执行的语音识别的结果的信息。

[0351] 详细地,语言模型更新单元2222可以确定与每个用户相对应的第二语言模型2223。在语音识别数据更新设备2220中,可以存在与每个用户相对应的至少一个第二语言模型2223。如果没有与用户相对应的第二语言模型2223,则语言模型更新单元2222可以新生成与该用户相对应的第二语言模型2223。接下来,语言模型更新单元2222可以基于个性化信息2221更新与各个用户相对应的语言模型。详细地,语言模型更新单元2222可以从个性化信息2221中检测新字,并且通过使用检测到的新字来更新与各个用户相对应的第二语言模型2223。

[0352] 语音识别设备2240的话音识别器2241可以通过使用针对各个用户建立的第二语言模型2223来执行语音识别。当包括话音命令的语音数据被接收时,话音识别器2241可以通过使用与正在发出话音命令的用户相对应的第二语言模型2223来执行语音识别。

[0353] 图23是示出根据实施例的包括与各个应用相对应的语言模型的语音识别系统的示例的框图。

[0354] 参考图23,可以基于关于安装在用户设备2310上的至少一个应用的设备信息2321来更新或者生成语音识别数据更新设备2320的第二语言模型2323。因此,安装在用户设备

2310中的应用中的每一个可以不由自身执行语音识别,并且可以在用于语音识别的单独的平台执行语音识别。接下来,基于在用于语音识别的平台执行语音识别的结果,可以向至少一个应用请求任务。

[0355] 用户设备2310可以包括其中至少一个应用可以被安装在其上的可以由用户使用的各种类型的终端设备。安装在用户设备2310上的应用2311可以包括关于可以根据命令被执行的任务的信息。例如,应用2311可以包括“播放”、“暂停”、和“停止”,作为关于与命令“播放”、“暂停”、和“停止”相对应的任务的信息。此外,应用2311可以包括关于可以被包括在命令中的文本的信息。用户设备2310可以将关于可以基于命令被执行的应用2311的任务的信息和关于可以包括在命令中的文本的信息中的至少一个,发送到语音识别数据更新设备2320。语音识别数据更新设备2320可以基于从用户设备2310接收到的信息来执行语音识别。

[0356] 语音识别数据更新设备2320可以包括设备信息2321、语言模型更新单元2322、第二语言模型2323、和分段信息2324。语音识别数据更新设备2320可以与图2至20所示的语音识别数据更新设备220和420相对应。

[0357] 设备信息2321可以包括关于应用2311的信息、从用户设备2310接收的信息。语音识别数据更新设备2320可以从用户设备2310接收关于可以基于命令被执行的应用2311的任务的信息和关于可以被包括在命令中的文本的信息中的至少一个。语音识别数据更新设备2320可以将从用户设备2310接收的关于应用2311的信息中的至少一个存储为设备信息2321。语音识别数据更新设备2320可以为用户设备2310中的每一个存储设备信息2321。

[0358] 语音识别数据更新设备2320可以周期性地或者当关于应用2311的新事件发生时,从用户设备2310接收关于应用2311的信息。可替换地,当语音识别设备2330开始执行语音识别时,语音识别数据更新设备2320可以向用户设备2310请求关于应用2311的信息。此外,语音识别数据更新设备2320可以将接收的信息存储为设备信息2321。因此,语音识别数据更新设备2320可以基于关于应用2311的最新信息来更新语言模型。

[0359] 语言模型更新单元2322可以基于设备信息2321更新可以被用于执行语音识别的语言模型。可以基于设备信息2321被更新的语言模型可以包括来自所述至少一个第二语言模型2323当中的与用户设备2310相对应的第二语言模型。此外,可以基于设备信息2321更新的语言模型可以包括来自至少一个第二语言模型2323当中的与应用2311相对应的第二语言模型。

[0360] 第二语言模型2323可以包括可以基于情形信息选择性地被应用的至少一个独立的语言模型。语音识别设备2330可以基于情形信息来选择第二语言模型2323中的至少一个,并且通过使用所选择的第二语言模型2323来执行语音识别。

[0361] 根据实施例,分段信息2324可以包括关于当语音识别数据被更新时可以生成的新字的预定单元分量的信息。语音识别数据更新设备2320可以将新字划分为子字,并且根据实施例更新语音识别数据,以实时地将新字添加到第二语言模型2323。因此,当划分成子字的新字被语音识别出时,它的语音识别的结果可以包括子字。如果由语音识别设备2330执行语音识别,则分段信息2324可以被用于将语音识别出的子字恢复为原始字。

[0362] 语音识别设备2330可以包括针对接收到的语音命令执行语音识别的语音识别单元2331和将子字恢复为原始字的文本恢复设备2332。文本恢复设备2332可以将语音识别出

的子字恢复为原始字,并且输出语音识别的最终结果。

[0363] 图24是示出根据实施例的基于语音识别的结果发送执行任务的请求的用户设备的示例的示意图。用户设备2410可以与图18、图22、或者图21的用户设备1850、2210、和2310相对应。

[0364] 参考图24,如果用户设备2410是电视机(TV),则基于语音识别的结果的命令可以经由用户设备2410被发送到包括用户设备2410的外部设备,即空调2420、清洁器2430、和洗衣机2450。

[0365] 当用户在位置a 2440发出话音命令时,语音数据可以被空调2420、清洁器2430、和用户设备2410收集。用户设备2410可以在信噪比(signal-to-noise ratio,SNR)或者音量方面,将由用户设备2410收集的语音数据与由空调2420和清洁器2430收集的语音数据进行比较。作为比较的结果,用户设备2410可以选择最高质量的语音数据,并且将所选择的语音数据发送到用于执行语音识别的语音识别设备。参考图24,由于用户位于最靠近清洁器2430的位置,所以由清洁器2430收集的语音数据可以是最高质量的语音数据。

[0366] 根据实施例,可以通过使用多个设备收集语音数据,因此即使用户远离用户设备2410也可以收集到高质量的语音数据。因此,根据用户与用户设备2410之间的距离的成功率的变化可以被减小。

[0367] 此外,即使用户位于远离用户设备2410所位于的起居室的洗衣房中的位置2460处,包括用户的话音命令的语音数据也可以由洗衣机2450来收集。洗衣机2450可以将所收集的语音数据发送到用户设备2410,并且用户设备2410可以基于接收到的语音数据执行任务。因此,使用各种设备,无论与用户设备2410的距离如何,用户都可以以高成功率发出话音命令。

[0368] 以下,将更详细地描述关于每个用户执行语音识别的方法。

[0369] 图25是示出根据实施例的生成关于语音数据的类别的个人优选内容列表的方法的框图。

[0370] 参考图25,语音识别设备230可以从语音数据和文本数据2510接收声学数据2520和内容信息2530。文本数据和声学数据2520可以彼此相对应,其中内容信息2530可以从文本数据获取,并且声学数据2520可以从语音数据获取。文本数据可以从对语音数据执行语音识别的结果获取。

[0371] 声学数据2520可以包括用于区分不同人的话音的话音特征信息。语音识别设备230可以基于声学数据2520来区分类别,并且,如果由于根据时间间隙的不同的话音特征,针对同一用户的声学数据2520不同,则该声学数据2520可以被分类为不同类别。声学数据2520可以包括关于语音数据的特征信息,诸如表明声音多高或者多低的音高的平均值、方差、抖动(声带的振动的变化)、振幅微扰(话音波形的规律性)、持续时间、梅尔频率倒谱系数(Mel frequency cepstral coefficients,MFCC)的平均值和方差。

[0372] 内容信息2530可以基于包括在文本数据中的标题信息被获取。内容信息2530可以包括原样文本数据中包括的标题。此外,内容信息2530可以进一步包括涉及标题的字。

[0373] 例如,如果包括在文本数据中的标题是“天气”和“职业棒球比赛结果”,则涉及“天气”的“天气信息”、涉及“新闻”和“职业棒球比赛结果”的“体育新闻”和“职业棒球重播”可以被获取作为内容信息2540。

[0374] 语音识别设备230可以基于声学数据2520和从文本数据获取的内容信息2540来确定涉及语音数据的类别。类别可以包括声学数据和与各个类别相对应的个人优选内容列表。语音识别设备230可以基于声学数据和关于相应的类别的个人优选内容列表来确定关于语音数据的类别。

[0375] 由于在语音数据被初始分类和被初始化之前不存在个人优选内容列表,所以语音识别设备230可以基于声学数据来分类语音数据。接下来,语音识别设备230可以从与各个分类的语音数据相对应的文本数据中提取内容信息2540,并且生成与各个类别相对应的个人优选内容列表。接下来,通过在后续的语音识别期间,将所提取的内容信息2540添加到个人优选内容列表,可以逐渐增加在分类期间被应用到个人优选内容列表的权重。

[0376] 可以基于以下方程式4执行更新类别的方法。

[0377] [方程式4]

$$[0378] \text{Class}_{\text{similarity}} = W_a A_v + W_1 L_v$$

[0379] 在方程式4中, A_v 和 W_a 分别表示基于语音数据的声学数据的类别和关于其的权重,而 L_v 和 W_1 分别表示基于个人优选内容列表的类别和关于其的权重。

[0380] 最初, W_1 的值可以为0,并且随着个人优选内容列表被更新, W_1 的值可以增大。

[0381] 此外,语音识别设备230可以基于各个类别的个人优选内容列表和各个类别的语音识别历史来生成与各个类别相对应的语言模型。此外,语音识别设备230可以通过应用讲话者-自适应算法(例如,最大似然线性回归(maximum likelihood linear regression, MLLR)、最大A后验(maximum A posterior, MAP)等),基于与各个类别相对应的语音数据和全局的声学模型,来生成用于各个类别的个性化声学模型。

[0382] 在语音识别期间,语音识别设备230可以从语音数据中标识类别,并且确定与所标识的类别相对应的语言模型或者声学模型。语音识别设备230可以通过使用所确定的语言模型或者声学模型来执行语音识别。

[0383] 在执行语音识别之后,语音识别数据更新设备220可以通过使用语音识别的结果,来更新语音识别出的语音数据和文本数据分别属于的语言模型和声学模型。

[0384] 图26是示出根据实施例的确定语音数据的类别的示例的示意图。

[0385] 参考图26,每个声学数据可以具有包括声学信息和内容信息的特征信息。每个声学数据可以由图表明,其中x轴表明声学信息,y轴表明内容信息。通过使用K-均值聚类方法,声学数据可以基于声学信息和内容信息被分类为n个类别。

[0386] 图27是示出根据实施例的根据语音数据的类别来更新语音识别数据的方法的流程图。

[0387] 参考图27,在操作S2701中,语音识别数据更新设备220可以获取语音数据和与语音数据相对应的文本。语音识别数据更新设备220可以获取与语音数据相对应的文本作为由语音识别设备230执行的语音识别的结果。

[0388] 在操作S2703中,语音识别数据更新设备220可以检测在操作S2701中获取到的文本或者涉及文本的内容信息。例如,内容信息可以进一步包括涉及文本的字。

[0389] 在操作S2705中,语音识别数据更新设备220可以从在操作S2701中获取到的语音数据中提取声学信息。可以在操作S2705中提取的声学信息可以包括关于语音数据的声学特征的信息,并且可以包括音高、抖动、和振幅微扰之类的上述特征信息。

[0390] 在操作S2707中,语音识别数据更新设备220可以确定与在操作S2703和操作S2705中检测到的内容信息和声学信息相对应的类别。

[0391] 在操作S2709中,语音识别数据更新设备220可以基于内容信息和声学信息来更新与在操作S2707中所确定的类别相对应的语言模型或者声学模型。语音识别数据更新设备220可以通过检测包括在内容信息中的新字来更新语言模型。此外,语音识别数据更新设备220可以通过应用声学信息、全局声学模型和讲话者-自适应算法来更新声学模型。

[0392] 图28和图29是示出根据实施例的可以被分类的声学数据的示例的示意图。

[0393] 参考图28,关于多个用户的语音数据可以被分类为单一类别。没有必要将具有相似的声学特点和相似的内容偏好的用户分类到不同的类别中,因此这样的用户可以被分类为单一类别。

[0394] 参考图29,基于各个语音数据的特点,关于同一用户的语音数据可以被分类为不同的类别。在早晨和晚上用户的话音不同的情况下,可以有区别地检测关于语音数据的声学信息,因此关于早晨的话音的语音数据和关于晚上的话音的语音数据可以被分类为不同的类别。

[0395] 此外,如果关于同一用户的语音数据的内容信息不同,则语音数据可以被分类为不同的类别。例如,同一用户可以使用“婴儿相关的”内容用于护理婴儿。因此,如果语音数据的内容信息不同,则包括同一用户的话音的语音数据可以被分类为不同的类别。

[0396] 根据实施例,语音识别设备230可以通过使用为各个用户确定的第二语言模型来执行语音识别。此外,在使用同一设备ID而不能用设备ID区分用户的情况下,可以基于语音数据的声学信息和内容信息来将用户分类。语音识别设备230可以基于所确定的类别来确定声学模型或者语言模型,并且可以执行语音识别。

[0397] 此外,如果仅仅由于用户(例如,兄弟、家庭成员等)的话音的相似性而不能基于声学信息来区分用户,则语音识别设备230可以通过进一步考虑内容信息来区分类别,从而执行讲话者-自适应的语音识别。

[0398] 图30和图31是示出根据实施例的执行个性化语音识别方法的示例的框图。

[0399] 参考图30和图31,用于为各个类别执行个性化语音识别的信息可以包括:基于包括关于个体的信息的个性化信息3021、3031、3121、和3131来更新第二语言模型3023、3033、3123、和3133的语言模型更新单元3022、3032、3122、和3132;以及当更新第二语言模型3023、3033、3123、和3133时可以生成的分段信息3024、3034、3124、和3134。用于为各个类别执行个性化语音识别的信息可以被包括在执行语音识别的语音识别设备3010或者语音识别数据更新设备220中。

[0400] 当多个人在清晰发音时,语音识别设备3010可以为用于语音识别的各个个体插入语言模型。

[0401] 参考图30,使用多个语言模型的插入方法可以是参考方程式1至3的如上所述的方法。例如,语音识别设备3010可以对与持有麦克风的人相对应的语言模型应用更高的权重。根据方程式1,如果多个语言模型被使用,则包括在语言模型中的字通常可以具有高概率。根据方程式2和3,包括在用于各个个体的语言模型中的字可以被简单地组合。

[0402] 参考图30,如果用于各个个体的语言模型的尺寸不大,则语音识别可以基于单一语言模型3141被执行,所述单一语言模型3141是对于多个人的语言模型的组合。随着语言

模型被组合,用于语音识别的要计算的概率的量可以被减少。然而,在组合语言模型的情况下,有必要通过重新确定各个概率来生成组合的语言模型。因此,如果用于各个个体的语言模型的尺寸小,则组合语言模型是有效率的。如果由多个个体组成的组可以被预先设立,则语音识别设备3010可以在执行语音识别的时间点之前获取关于该组的组合的语言模型。

[0403] 图32是示出根据实施例的语音识别数据更新设备的内部配置的框图。图32的语音识别数据更新设备可以与图2至图23的语音识别数据更新设备相对应。

[0404] 语音识别数据更新设备3200可以包括可以由用户或者可以经由网络连接到用户设备的服务器设备使用的各种类型的设备。

[0405] 参考图32,语音识别数据更新设备3200可以包括控制器3210和存储器3220。

[0406] 控制器3210可以检测包括在所收集的语言数据中的新字,并且更新在语音识别期间可以使用的语言模型。详细地,控制器3210可以将新字转换为音素序列,将音素序列中的每一个划分成预定单元分量,并且确定关于音素序列的分量的出现概率信息。此外,控制器3210可以通过使用出现概率信息来更新语言模型。

[0407] 存储器3220可以存储由控制器3210更新的语言模型。

[0408] 图33是示出根据实施例的语音识别设备的内部配置的框图。图33中的语音识别设备可以与图2至图31的语音识别设备相对应。

[0409] 语音识别设备3300可以包括可以由用户或者可以经由网络连接到用户设备的服务器设备使用的各种类型的设备。

[0410] 参考图33,语音识别设备3300可以包括控制器3310和通信单元3320。

[0411] 控制器3310可以通过使用语音数据来执行语音识别。详细地,控制器3310可以从语音数据中获取至少一个音素序列,并且获取关于通过划分音素序列获取的预定单元分量的出现概率。接下来,控制器3310可以基于出现概率获取一个音素序列,并且基于关于所获取的音素序列的分段信息,将与音素序列相对应的字输出为语音识别出的字。

[0412] 通信单元3320可以根据用户输入接收包括用户的清晰发音的语音数据。如果语音识别设备3300是服务器设备,则语音识别设备3300可以从用户设备接收语音数据。接下来,通信单元3320可以将由控制器3310语音识别出的字发送到用户设备。

[0413] 图34是用于描述根据实施例的用户设备3400的配置的框图。

[0414] 如图34中所示,用户设备3400可以包括可以由用户使用的各种类型的设备,例如,移动电话、平板型计算机、PDA、MP3播放器、自助服务亭、电子框架、导航设备、数字TV、以及诸如手表或者头戴式显示器(head mounted display,HMD)的可穿戴设备。

[0415] 用户设备3400可以与图2至图24的用户设备相对应、可以接收用户的清晰发音、将用户的清晰发音发送到语音识别设备、从语音识别设备接收语音识别出的语言、并且输出语音识别出的语言。

[0416] 例如,如图34中所示,根据实施例的用户设备3400可以包括不仅显示单元3410和控制器3470,而且存储器3420、GPS芯片3425、通信单元3430、视频处理器3435、音频处理器3440、用户输入器3445、麦克风单元3450、图像拾取单元3455、扬声器单元3460、和移动检测单元3465。

[0417] 以下将给出上述组件的详细说明。

[0418] 显示单元3410可以包括显示面板3411和用于控制显示面板3411的控制器(未示

出)。显示面板3411可以被具体化为各种类型的显示面板中的任一种,诸如液晶显示器(liquid crystal display,LCD)面板、有机发光二极管(organic light emitting diode,OLED)显示面板、有源矩阵有机发光二极管(active-matrix organic light emitting diode,AM-OLED)面板、和等离子体显示面板(plasma display panel,PDP)。显示面板3411可以被具体化为柔性的、透明的、或者可穿戴的。显示单元3410可以与用户输入器3445的触摸面板3447组合,并且作为触摸屏被提供。例如,触摸屏可以包括其中显示面板3411和触摸面板3447以堆栈结构相互组合的集成模块。

[0419] 根据实施例的显示单元3410可以在控制器3470的控制下显示语音识别的结果。

[0420] 存储器3420可以包括内部存储器(未示出)和外部存储器(未示出)中的至少一个。

[0421] 例如,内部存储器可以包括以下各项中的至少一个:易失性存储器(例如,动态随机存取存储器(dynamic random access memory,DRAM)、静态RAM(static RAM,SRAM)、同步动态RAM(synchronous dynamic RAM,SDRAM)等)、非易失性存储器(例如,一次可编程只读存储器(one time programmable read-only memory,OTPROM)、可编程ROM(programmable ROM,PROM)、可擦除/可编程ROM(erasable/programmable ROM,EPRM)、电可擦除/可编程ROM(electrically erasable/programmable ROM,EEPROM)、掩模ROM、闪存ROM等)、硬盘驱动器(hard disk drive,HDD)、或者固态硬盘(solid state disk,SSD)。根据实施例,控制器3470可以将命令或者从非易失性存储器或者其它组件中的至少一个接收的数据加载到易失性存储器,并且对其进行处理。此外,控制器3470可以将从其它组件接收到的或者由其它组件生成的数据存储到非易失性存储器中。

[0422] 外部存储器可以包括以下各项中的至少一个:紧凑闪存(compact flash,CF)、安全数字(secure digital,SD)、微型安全数字(micro secure digital,Mini-SD)、迷你安全数字(mini secure digital,Mini-SD)、极限数字(extreme digital,xD)、和记忆棒。

[0423] 存储器3420可以存储用于用户设备3400的操作的各种程序和数据。例如,存储器3420可以临时或者永久地存储包括用户的清晰发音的语音数据和基于语音数据的语音识别的结果数据中的至少一个。

[0424] 控制器3470可以控制显示单元3410以在显示单元3410上显示存储在存储器3420中的一部分信息。换句话说,控制器3470可以在显示单元3410上显示存储在3420中的语音识别的结果。可替换地,当在显示单元3410的区域执行用户手势时,控制器3470可以执行与用户手势相对应的控制操作。

[0425] 控制器3470可以包括RAM 3471、ROM 3472、CPU 3473、图形处理单元(GPU) 3474、和总线3475中的至少一个。RAM 3471、ROM 3472、CPU 3473、和GPU 3474可以经由总线3475相互连接。

[0426] CPU 3473访问存储器3420,并且通过使用存储在存储器3420中的OS执行启动操作。接下来,CPU 3473通过使用各种程序、内容、和存储在存储器3420中的数据来执行各种操作。

[0427] 用于启动系统的命令集存储在ROM 3472中。例如,当输入开启命令并且向用户设备3400供电时,CPU 3473可以根据存储在ROM3472中的命令将存储在存储器3420中的OS复制到RAM 3471,执行OS,并且启动系统。当用户设备3400被启动时,CPU 3473复制存储在存储器3420中的各种程序,并且通过执行复制到RAM 3471的程序来执行各种操作。当用户设

备3400被启动时, GPU 3474在显示单元3410的区域中显示UI屏幕图像。详细地, GPU 3474可以生成其中显示包括诸如内容、图标、和菜单的各种对象的电子文档的屏幕图像。GPU 3474基于屏幕图像的布局计算各个对象的坐标、形状、尺寸、和颜色之类的属性值。接下来, GPU 3474可以基于所计算的属性值来生成包括对象的各种布局的屏幕图像。可以将由GPU 3474生成的屏幕图像提供给显示单元3410, 并且在显示单元3410的各个区域中显示。

[0428] GPS芯片3425可以从全球定位系统(GPS)卫星接收GPS信号, 并且计算用户设备3400的当前位置。当为了使用导航程序或者其它目的需要用户的当前位置时, 控制器3470可以通过使用GPS芯片3425计算用户的当前位置。例如, 控制器3470可以将包括通过使用GPS芯片3425计算出的用户的位置的情形信息发送到语音识别设备或者语音识别数据更新设备。基于情形信息, 可以由语音识别设备或者语音识别数据更新设备更新语言模型或者执行语音识别。

[0429] 通信单元3430可以经由各种形式的通信协议来执行与各种类型的外部设备的通信。通信单元3430可以包括Wi-Fi芯片3431、蓝牙芯片3432、无线通信芯片3433、和NFC芯片3434中的至少一个。控制器3470可以通过使用通信单元3430执行与各种外部设备的通信。例如, 控制器3470可以通过使用通信单元3430接收对控制显示单元3410上显示的备忘录的请求, 并且基于接收到的请求将结果发送到外部设备。

[0430] Wi-Fi芯片3431和蓝牙芯片3432可以经由Wi-Fi协议和蓝牙协议执行通信。在使用Wi-Fi芯片3431或者蓝牙芯片3432的情况下, 首先发送和接收诸如服务集标识符(service set identifier, SSID)和会话密钥的各种连接信息, 通过使用上述各种连接信息建立通信, 然后可以发送和接收各种信息。无线通信芯片3433指的是经由诸如IEEE、Zigbee、第三代(3rd generation, 3G)、第三代合作项目(3rd generation partnership project, 3GPP)、和长期演进(long term evolution, LTE)的各种通信规范执行通信的芯片。NFC芯片3434指的是根据使用来自各种RF-ID频带当中的13.56MHz频带的近场通信(near field communication, NFC)协议进行操作的芯片; 例如, 135kHz频带、13.56MHz频带、433MHz频带、860~960MHz频带、和2.45GHz频带。

[0431] 视频处理器3435可以处理经由通信单元3430接收到的内容或者包括在存储于存储器3420的内容中的视频数据。视频处理器3435可以针对视频数据执行各种图像处理操作, 例如, 解码、缩放、噪声滤波、帧速率转换、分辨率转换等。

[0432] 音频处理器3440可以处理包括在经由通信单元3430接收的内容中或者包括在存储于存储器3420的内容中的音频数据。音频处理器3440可以针对音频数据执行各种音频处理操作, 例如, 解码、放大、噪声滤波等。例如, 音频处理器3440可以回放包括用户的清晰发音的语音数据。

[0433] 当执行用于回放多媒体内容的程序时, 控制器3470可以操作用户输入器3445和音频处理器3440, 并且回放相应的内容。扬声器单元3460可以输出通过音频处理器3440生成的音频数据。

[0434] 用户输入器3445可以接收由用户输入的各种命令。用户输入器3445可以包括按键3446、触摸面板3447、和笔识别面板3448中的至少一个。用户设备3400可以基于从按键3446、触摸面板3447、和笔识别面板3448中的至少一个接收到的用户输入, 来显示各种内容或者用户界面。

[0435] 按键3446可以包括在用户设备3400的外表面的各个区域(诸如前表面、侧表面、或者后表面)上形成的各种类型的按键,诸如机械按钮或者轮。

[0436] 触摸面板3447可以检测用户的触摸,并且输出与检测到的触摸信号相对应的触摸事件值。如果通过组合触摸面板3447与显示面板3411来形成触摸屏(未示出),则触摸屏可以具体化为诸如电容型、电阻型、和压电型的各种类型的触摸传感器中的任一种。当用户的身体部分触摸到电容型触摸屏的表面时,通过检测由用户的身体部分引起的微电流来计算触摸的坐标。电阻型触摸屏包括布置在触摸屏内部的两个电极板,并且当用户触摸触摸屏时,通过检测在触摸位置上的上板和下板相互接触时流过的电流来计算触摸的坐标。在触摸屏上发生的触摸事件通常可以由人的手指生成,但触摸事件也可以由用于应用电容变化的导电材料形成的物体产生。

[0437] 笔识别面板3448可以检测由用户操作的触摸笔(例如,触控笔(stylus pen)或者数字化笔)的接近笔(proximity pen)输入或者触摸笔输入,并且输出检测到的笔接近事件或者笔触摸事件。例如,笔识别面板3448可以具体化为电磁共振(electro-magnetic resonance, EMR)型面板,并且能够基于由于笔的接近或者触摸引起的电磁场的强度的变化来检测触摸输入或者接近输入。详细地,笔识别面板3448可以包括具有网格结构的电磁感应线圈传感器(未示出)和电磁信号处理单元(未示出),所述电磁信号处理单元顺序地向电磁感应线圈传感器的各个环形线圈提供具有预定频率的交替信号。当在笔识别面板3448的环形线圈附近存在包括谐振电路的笔时,由相应的环形线圈发送的磁场基于互电磁感应在笔内的谐振电路中生成电流。基于电流,由构成笔内的谐振电路的线圈生成感应磁场,并且笔识别面板3448在信号接收模式中检测环形线圈处的感应磁场,因此可以检测笔的接近位置或者触摸位置。笔识别面板3448可以被布置为占据显示面板3411下方的预定区域,例如,足以覆盖显示面板3411的显示区域的区域。

[0438] 麦克风单元3450可以接收用户的语音或其它声音,并且将其转换为音频数据。控制器3470可以使用经由用于电话呼叫操作的麦克风单元3450输入的用户语音,或者可以将用户的语音转换为音频数据,并且将其存储在存储器3420中。例如,控制器3470可以将经由麦克风单元3450输入的用户语音转换成音频数据,将转换的音频数据包括在备忘录中,并且存储包括音频数据的备忘录。

[0439] 图像拾取单元3455可以在用户的控制下拾取静止图像或者运动图像。图像拾取单元3455可以具体化为多个单元,诸如前置摄像头和后置摄像头。

[0440] 如果布置了图像拾取单元3455和麦克风单元3450,则控制器3470可以基于经由麦克风单元3450输入的用户语音或者由图像拾取单元3455识别的用户的动作来执行控制操作。例如,用户设备3400可以以动作控制模式或者语音控制模式来操作。如果用户设备3400在动作控制模式下操作,则控制器3470可以激活图像拾取单元3455、拾取用户的图像、跟踪用户的动作的变化、并且执行与之相对应的控制操作。例如,控制器3470可以基于由图像拾取单元3455检测到的用户的动作输入来显示备忘录或者电子文档。如果用户设备3400在语音控制模式下操作,则控制器3470可以在语音识别模式下操作以分析经由麦克风单元3450输入的用户语音,并且根据所分析的用户语音来执行控制操作。

[0441] 动作检测单元3465可以检测用户设备3400的主体的动作。用户设备3400可以在各个方向上旋转或者倾斜。这里,动作检测单元3465可以通过使用各种传感器(诸如地磁传感

器、陀螺传感器、和加速度传感器)中的至少一个来检测动作特点(诸如旋转方向、旋转角度、和倾斜角度)。例如,动作检测单元3465可以通过检测用户设备3400的主体的动作来接收用户的输入,并且基于接收到的输入显示备忘录或者电子文档。

[0442] 此外,尽管在图34中未示出,根据实施例,用户设备3400可以进一步包括:USB连接器可以经由其连接到用户设备3400的USB端口、要连接到各种外部终端的各种外部输入端口(诸如耳机、鼠标、和LAN)、用于接收和处理DMB信号的数字多媒体广播(digital multimedia broadcasting, DMB)芯片、以及各种其它传感器。

[0443] 用户设备3400的上述组件的名称可以变化。此外,根据本实施例的用户设备3400可以包括上述组件中的至少一个,其中可以省略一些组件或者可以进一步包括附加组件。

[0444] 本发明还能够具体化为计算机可读记录介质上的计算机可读代码。计算机可读记录介质是能够存储能够由计算机系统随后读取的数据的任何数据存储设备。计算机可读记录介质的示例包括只读存储器(ROM)、随机存取存储器(RAM)、CD-ROM、磁带、软盘、光学数据存储设备等。

[0445] 虽然参考本发明的优选实施例已经具体地示出和描述了本发明,但是本领域技术人员将理解,可以在这里进行各种形式和细节上的改变,而不脱离如所附权利要求所定义的本发明的精神和范围。优选的实施例应该被认为仅仅是描述性意义,而不是为了限制的目的。因此,本发明的范围不是由本发明的详细描述定义,而是由所附权利要求定义,并且该范围内的所有差异将被理解为包括在本发明中。

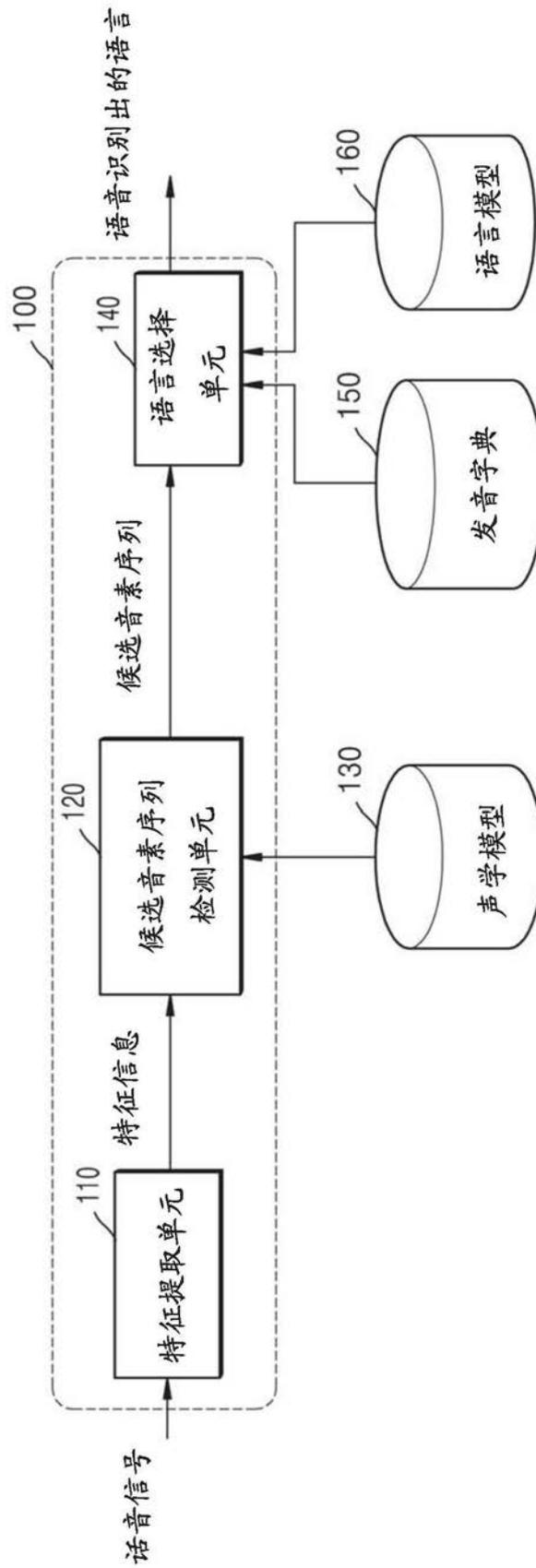


图1

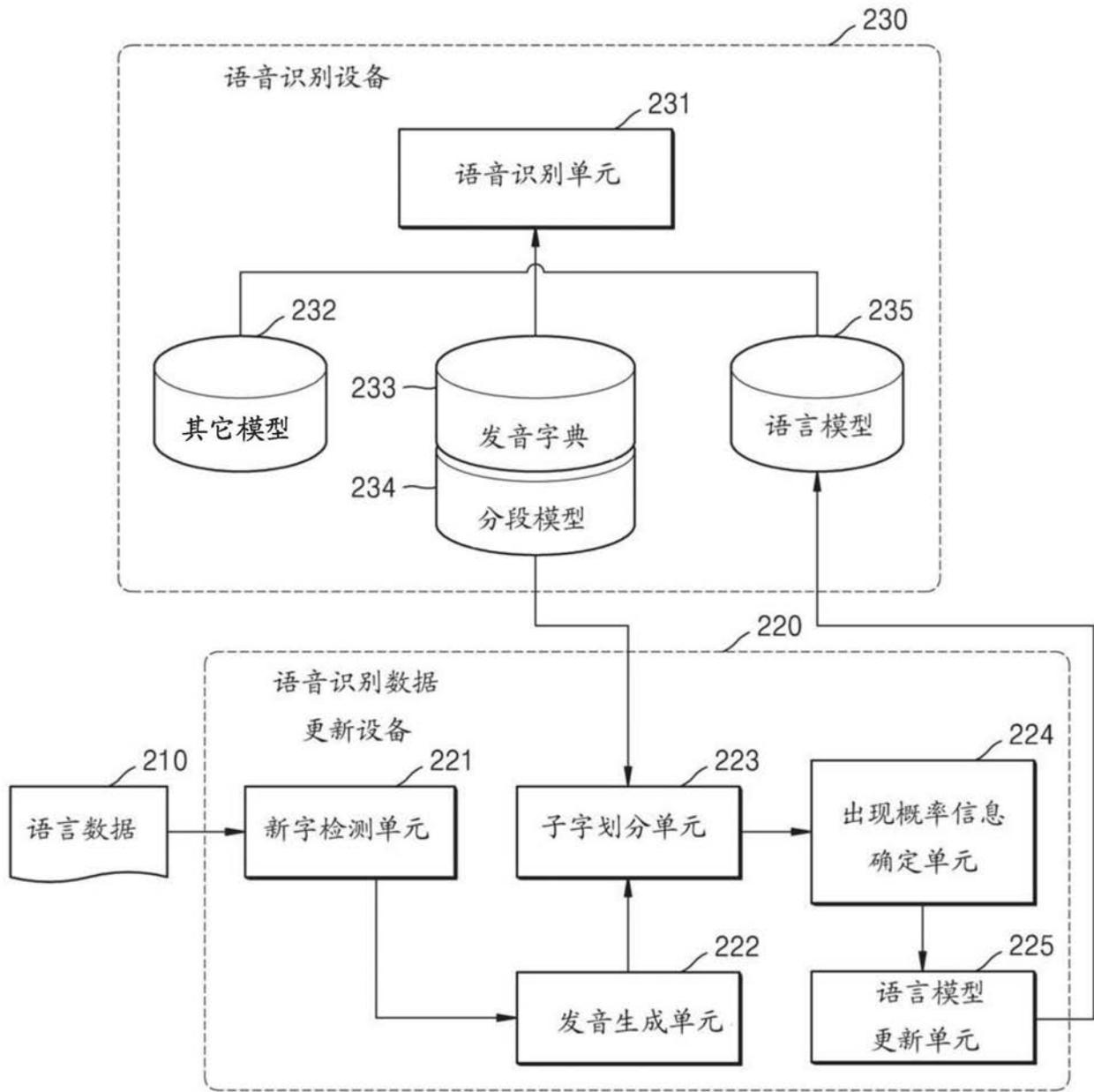


图2

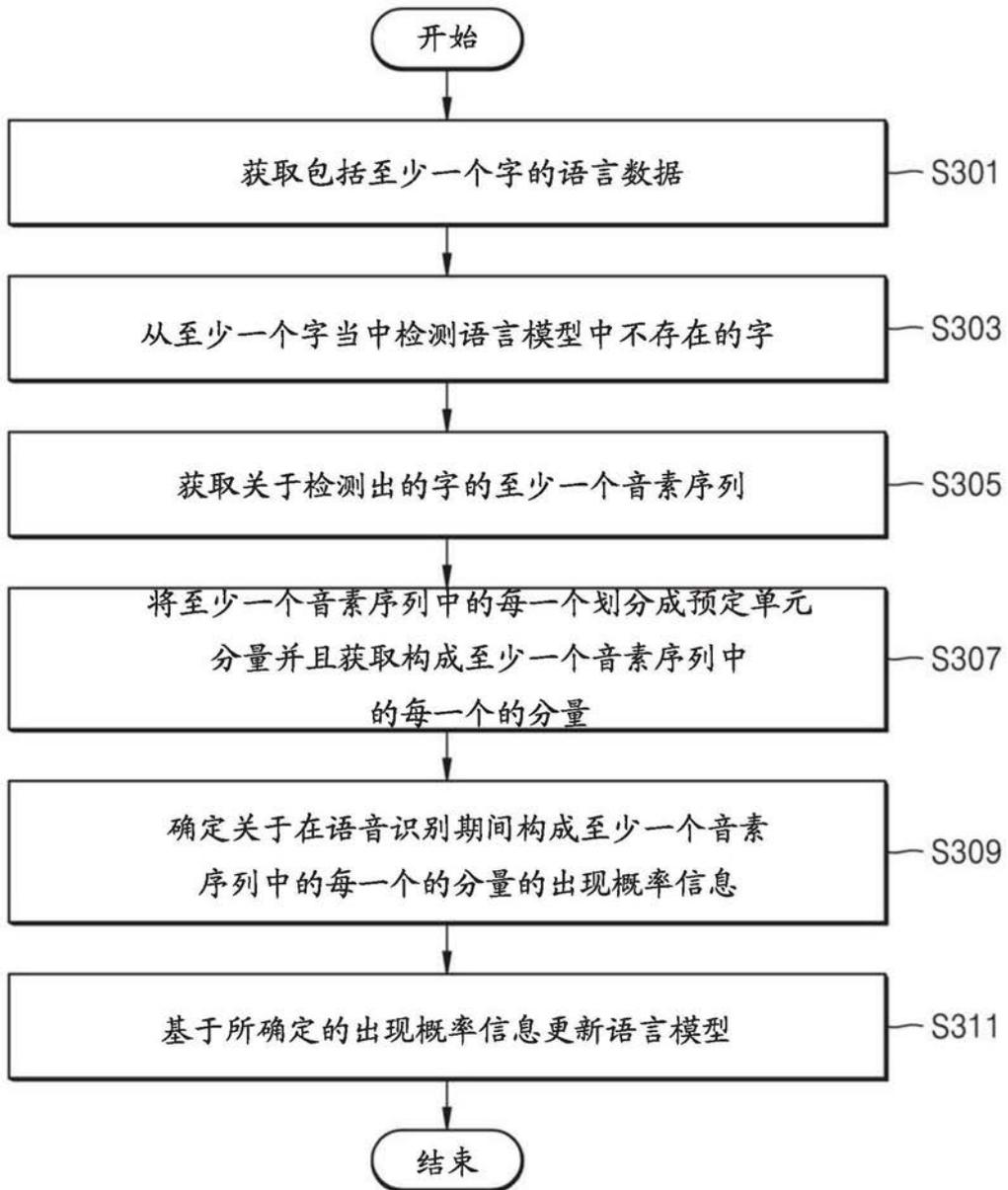


图3

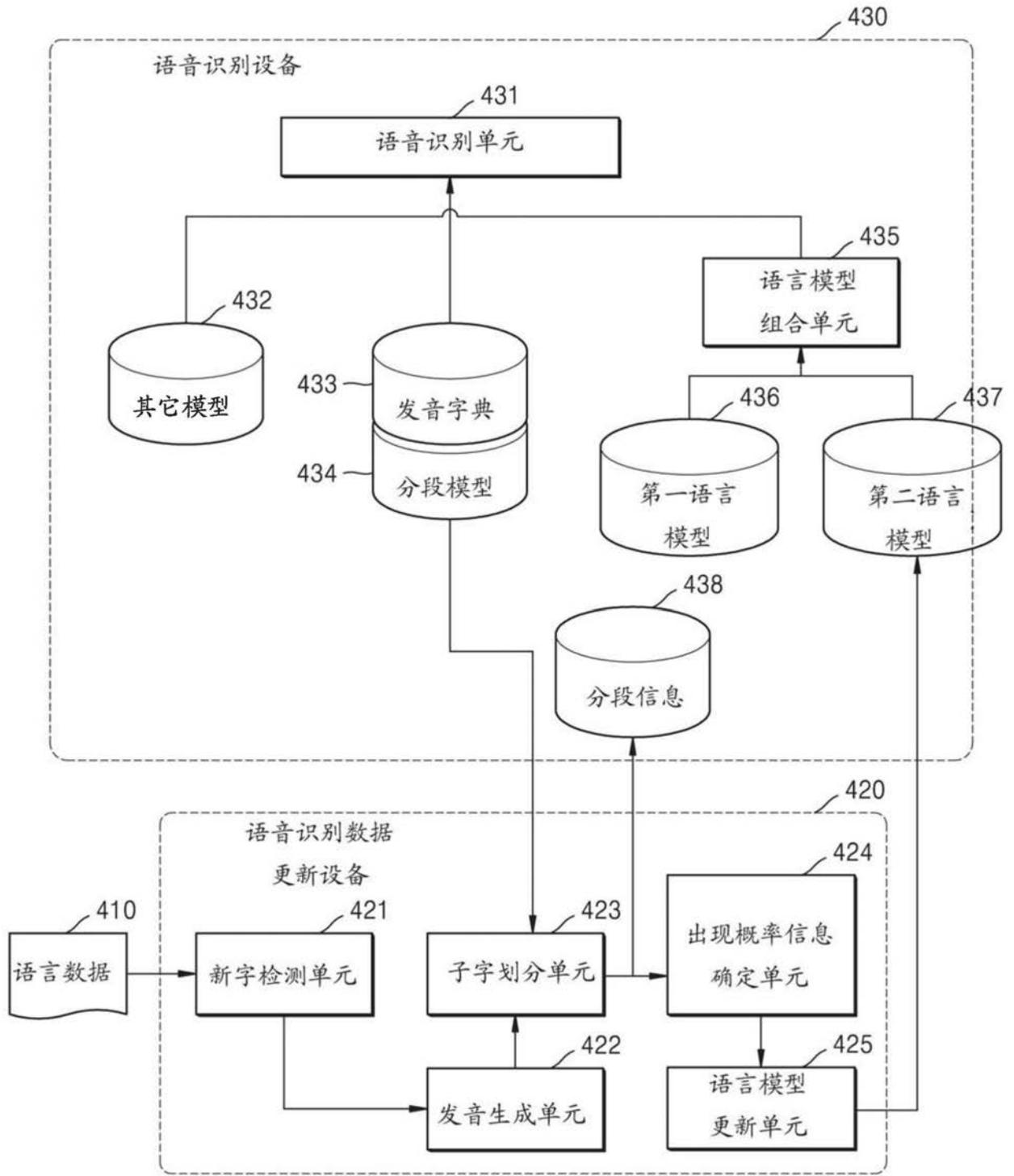


图4

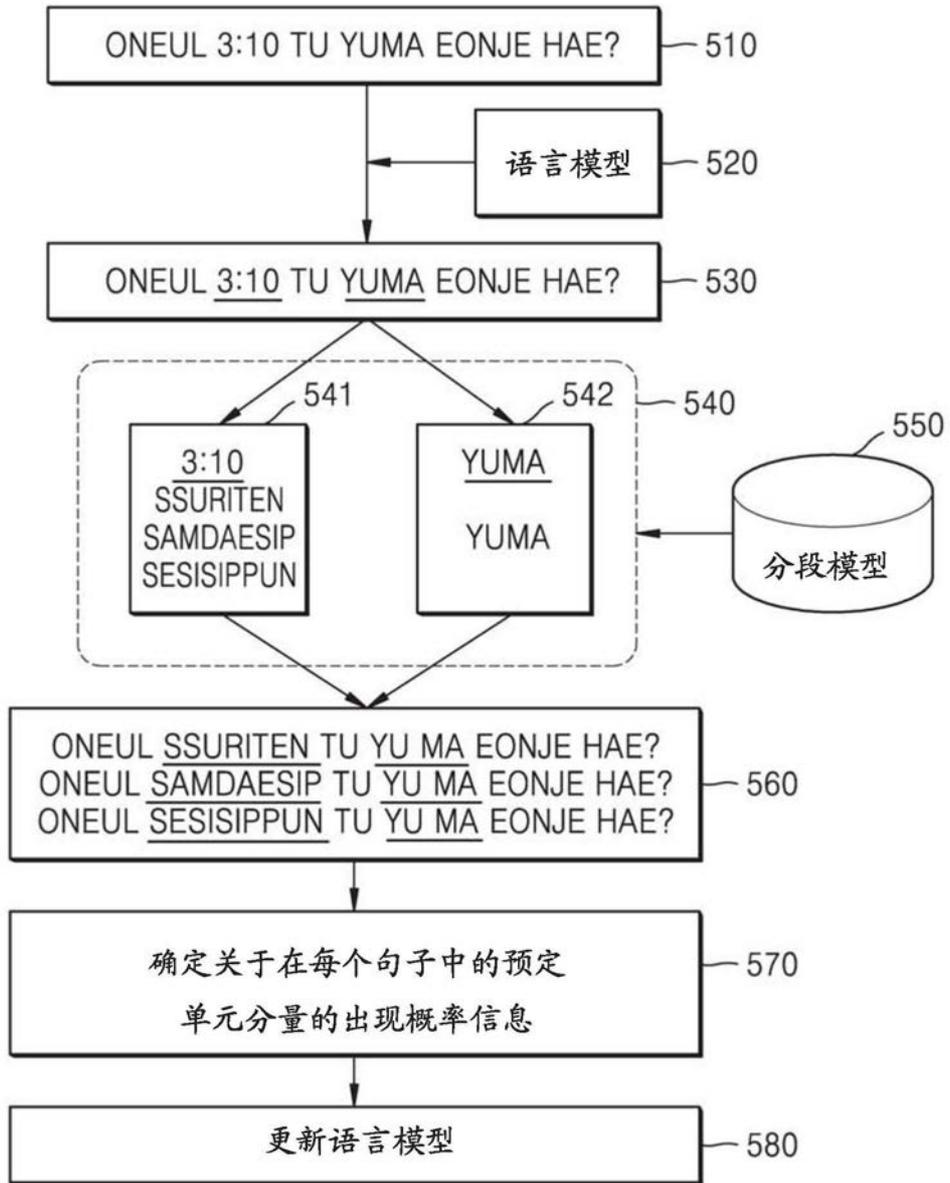


图5

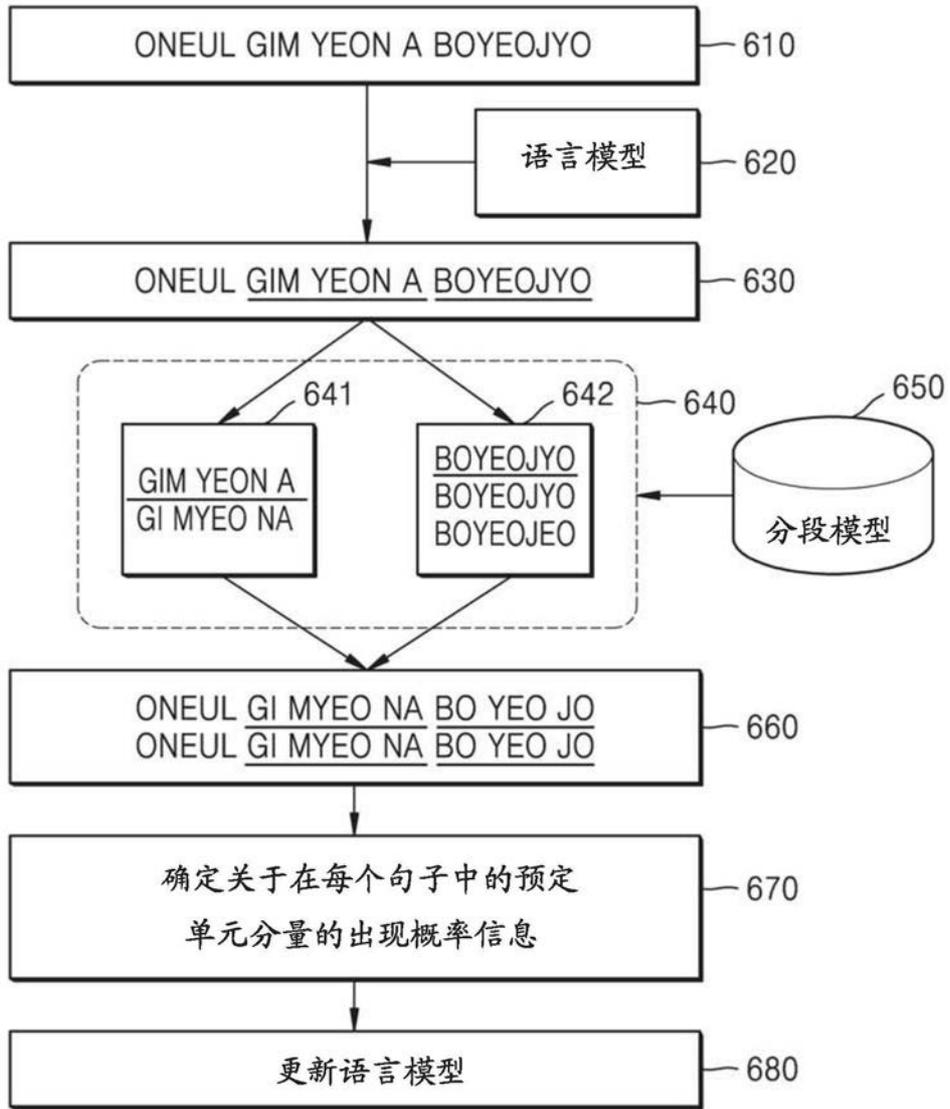


图6

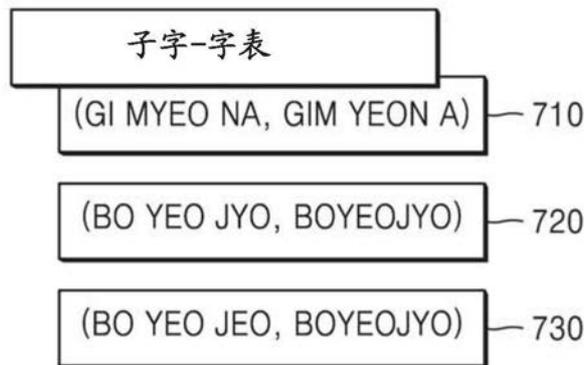


图7

字出现概率	
一个字	概率
ONEUL	0.143
GI	0.143
JEO	0.071
-	

810

两个字	概率
ONEUL GI	0.166
GI MYEO	0.166
YEO JYO	0.083
-	

820

图8

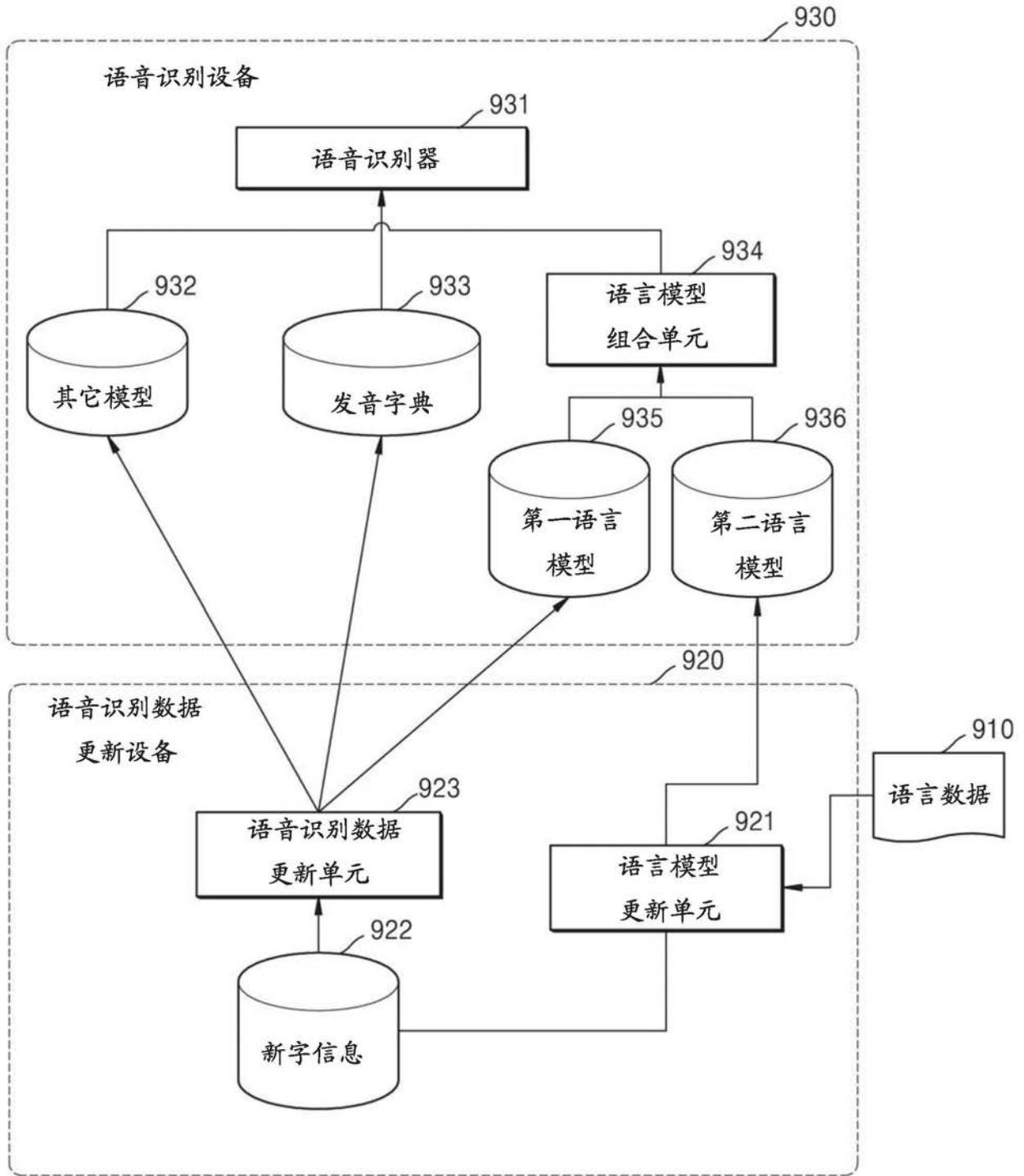


图9

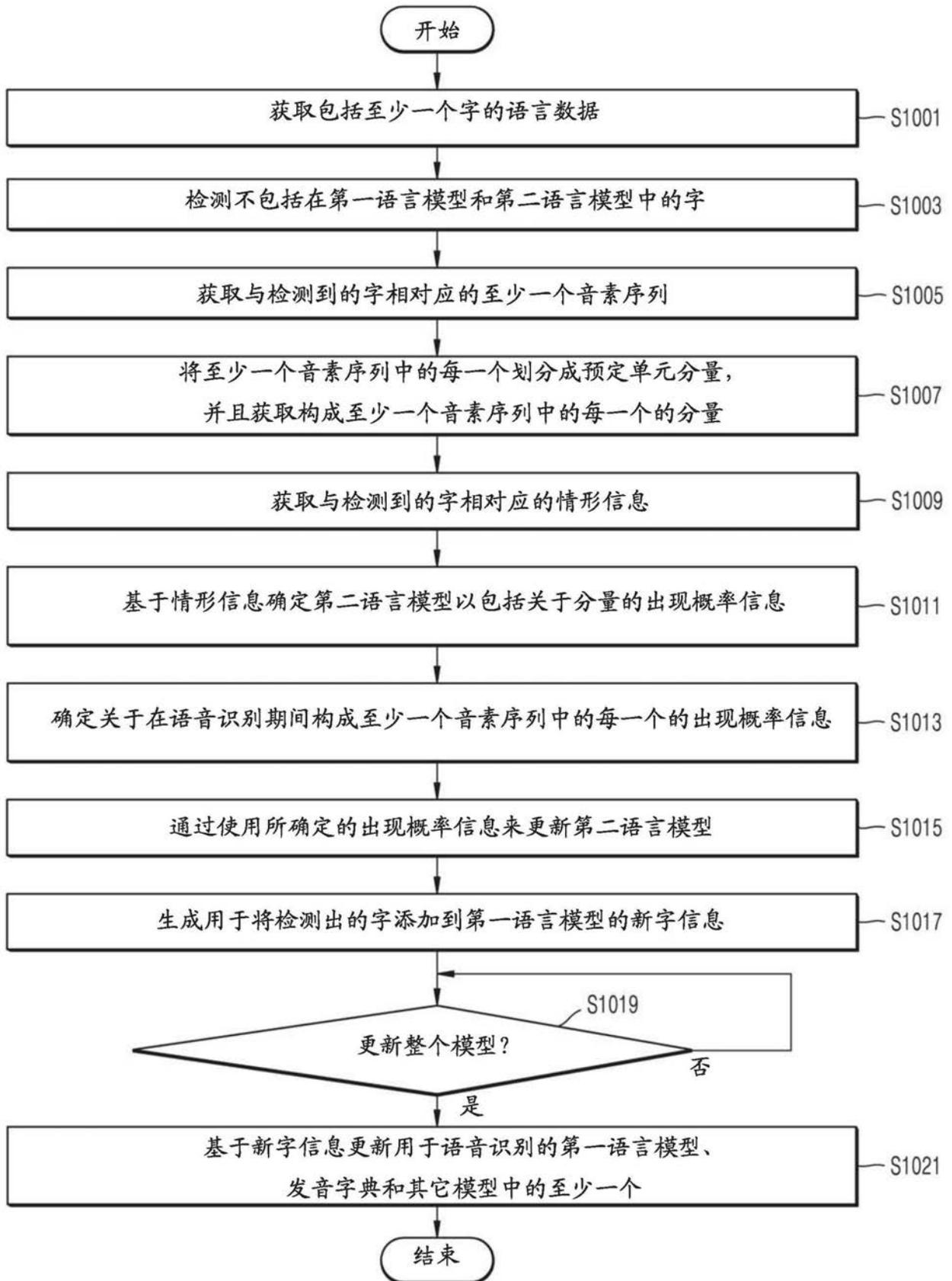


图10

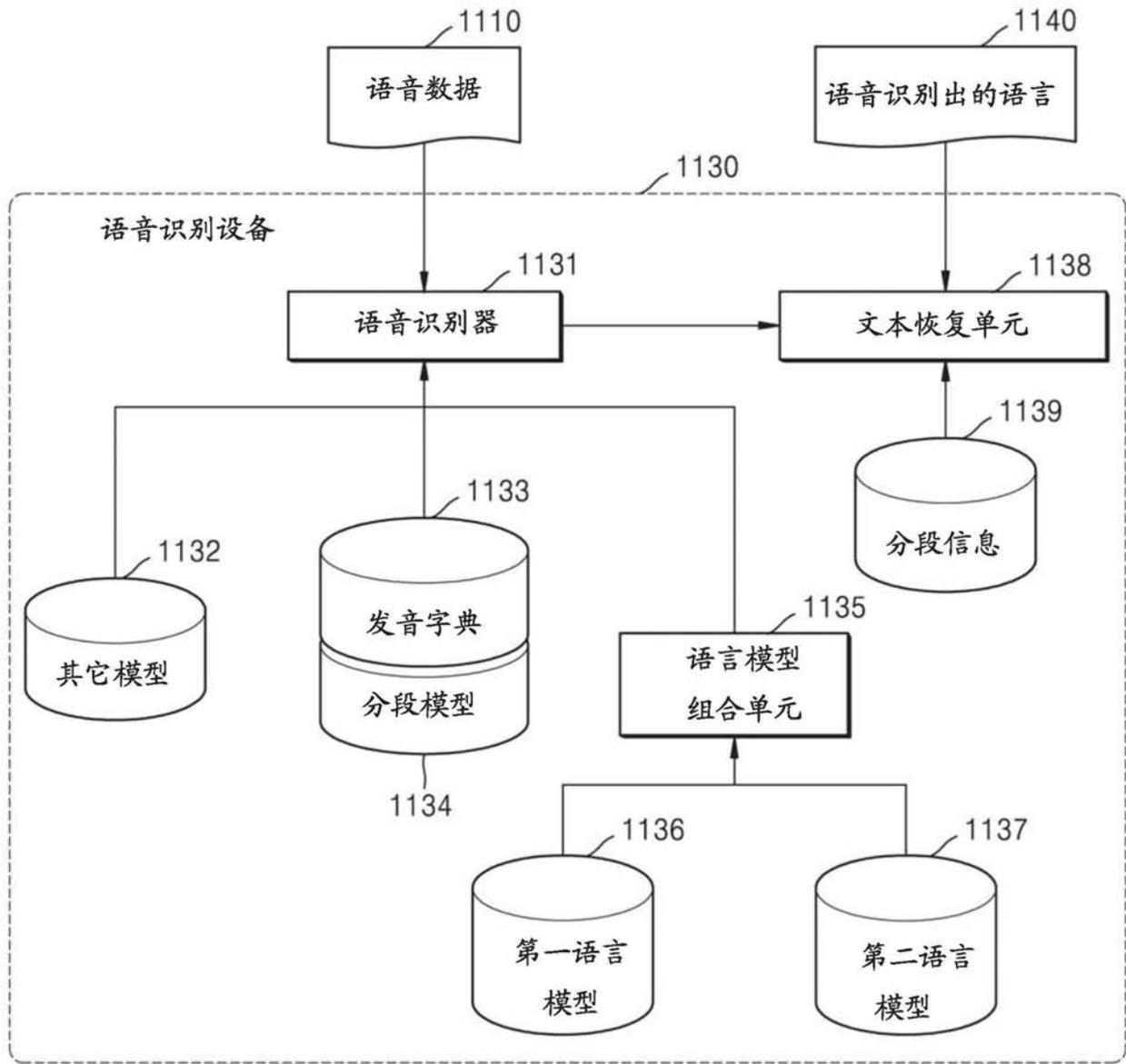


图11

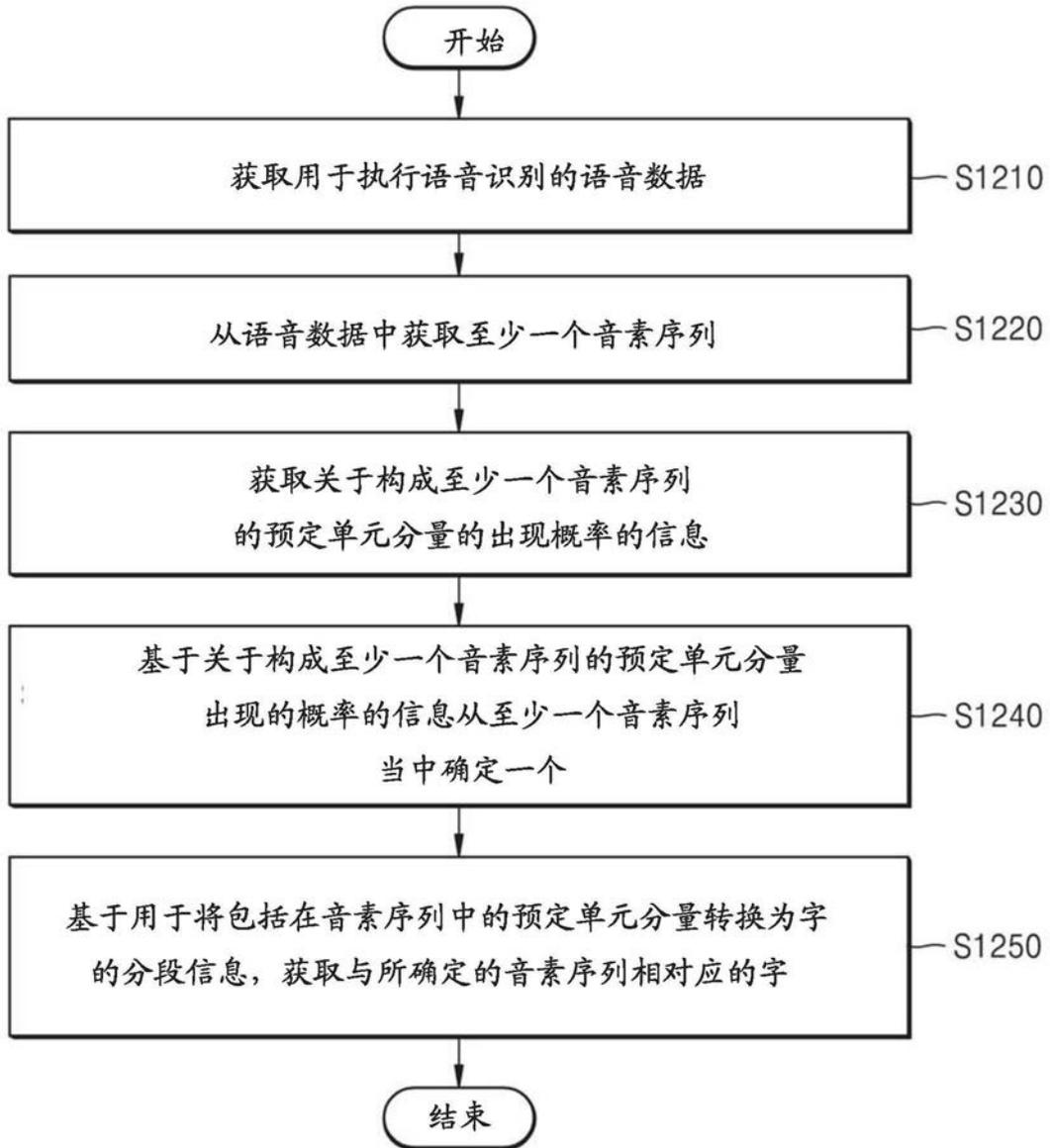


图12

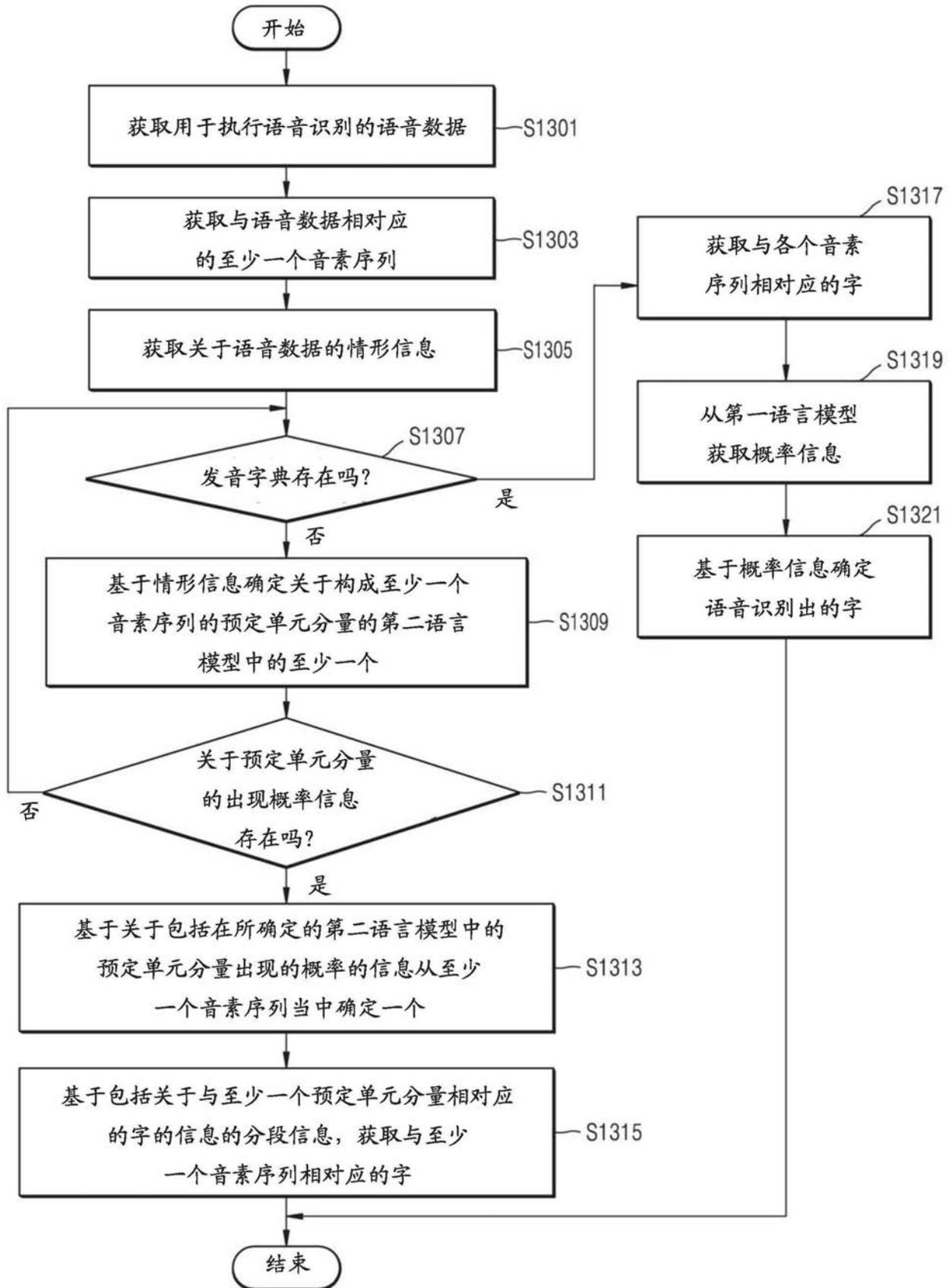


图13

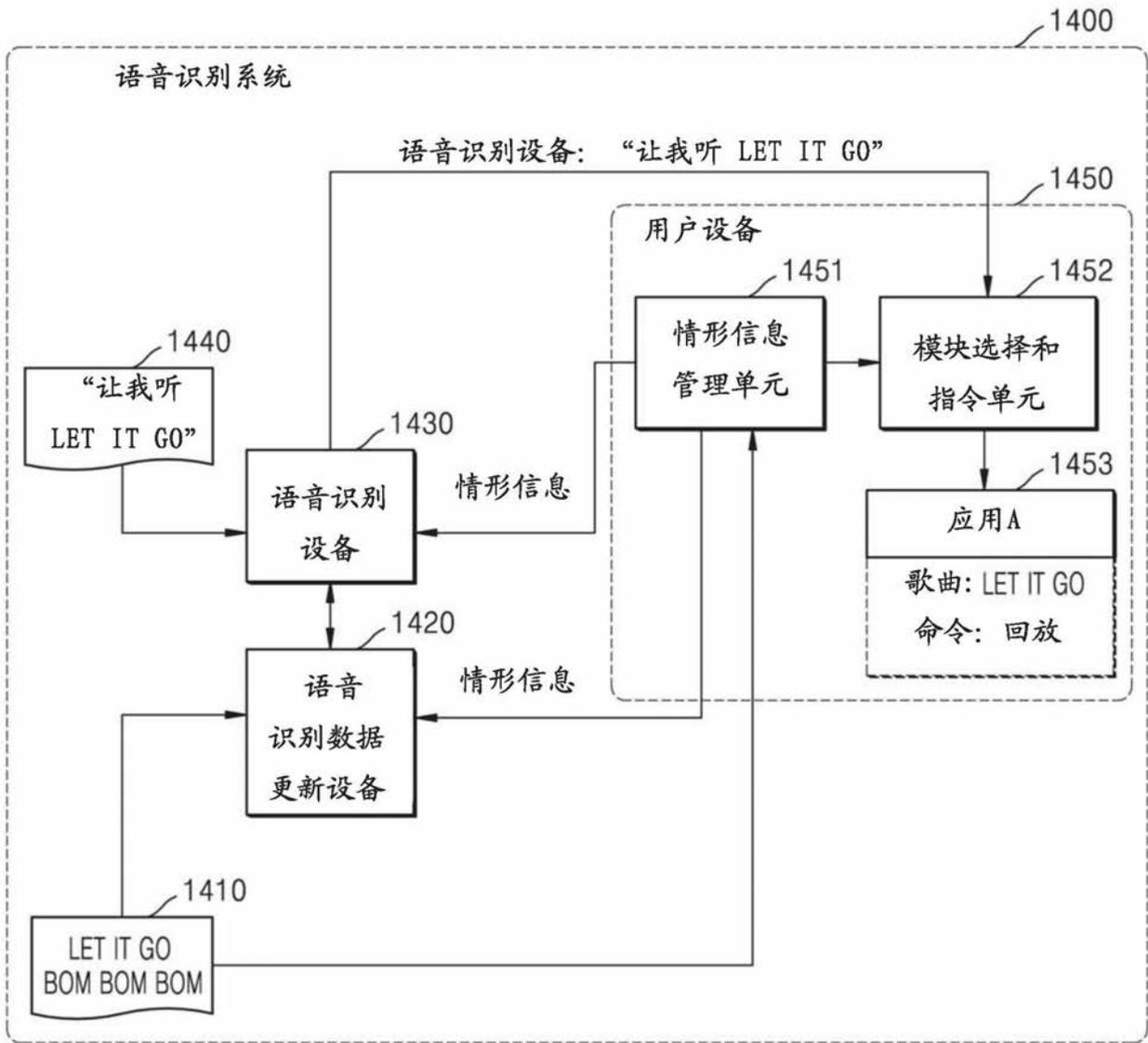


图14

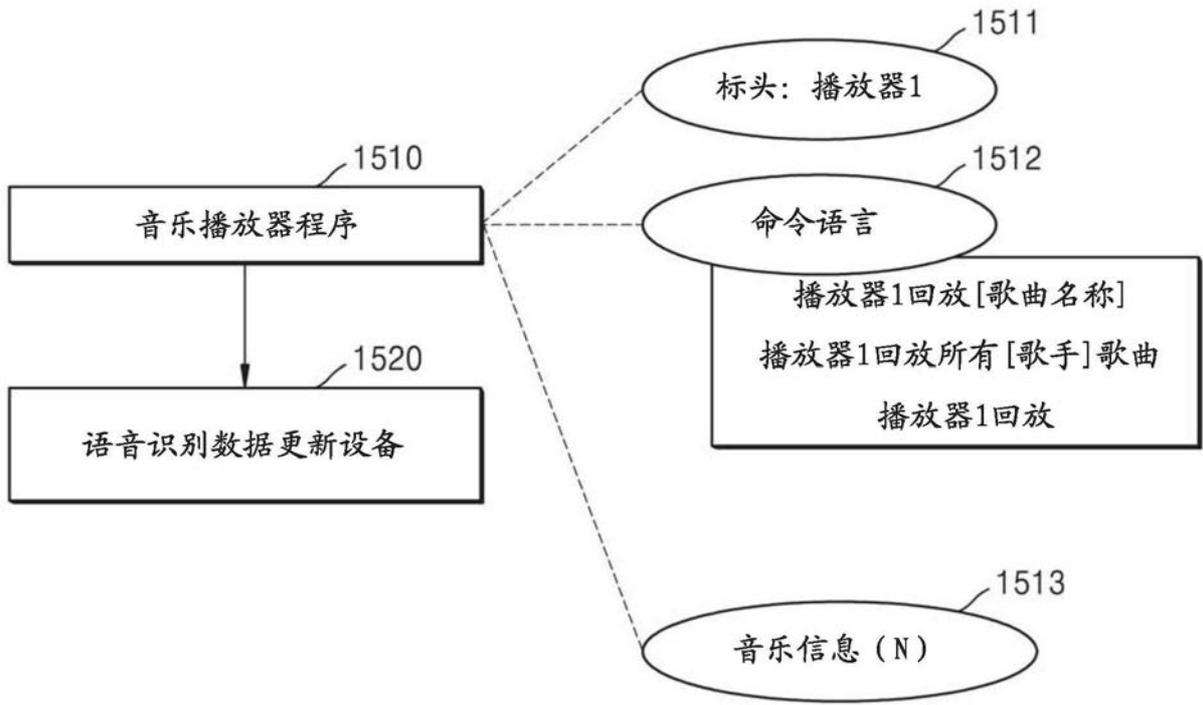


图15

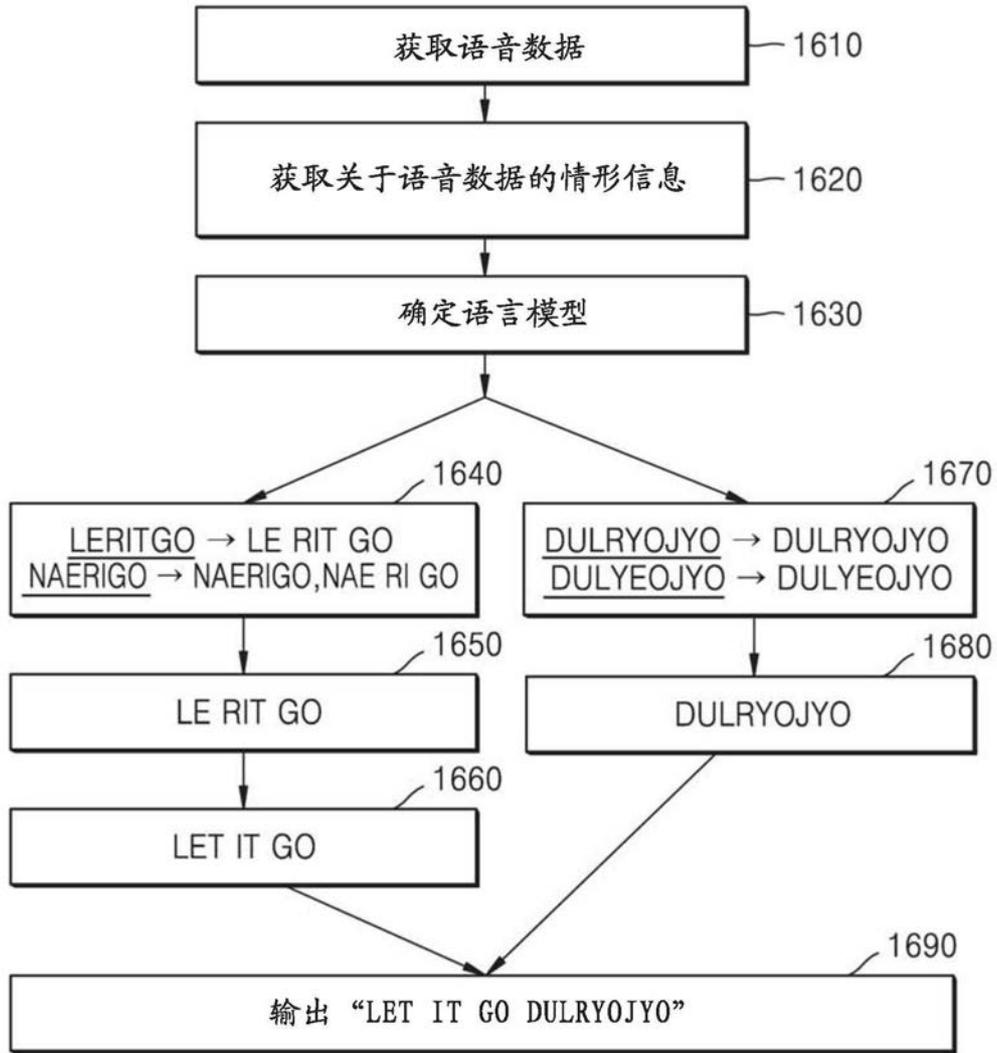


图16

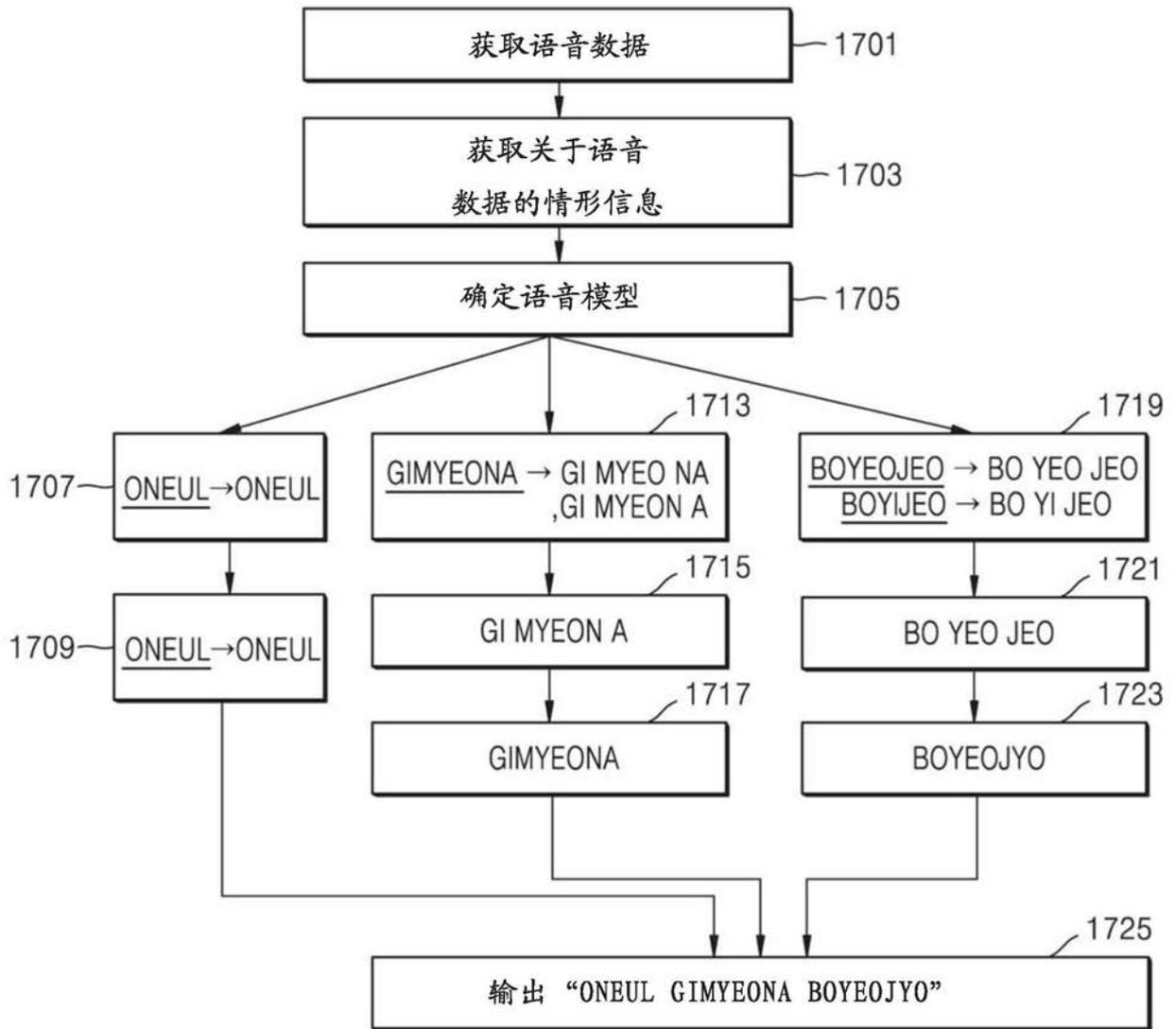


图17

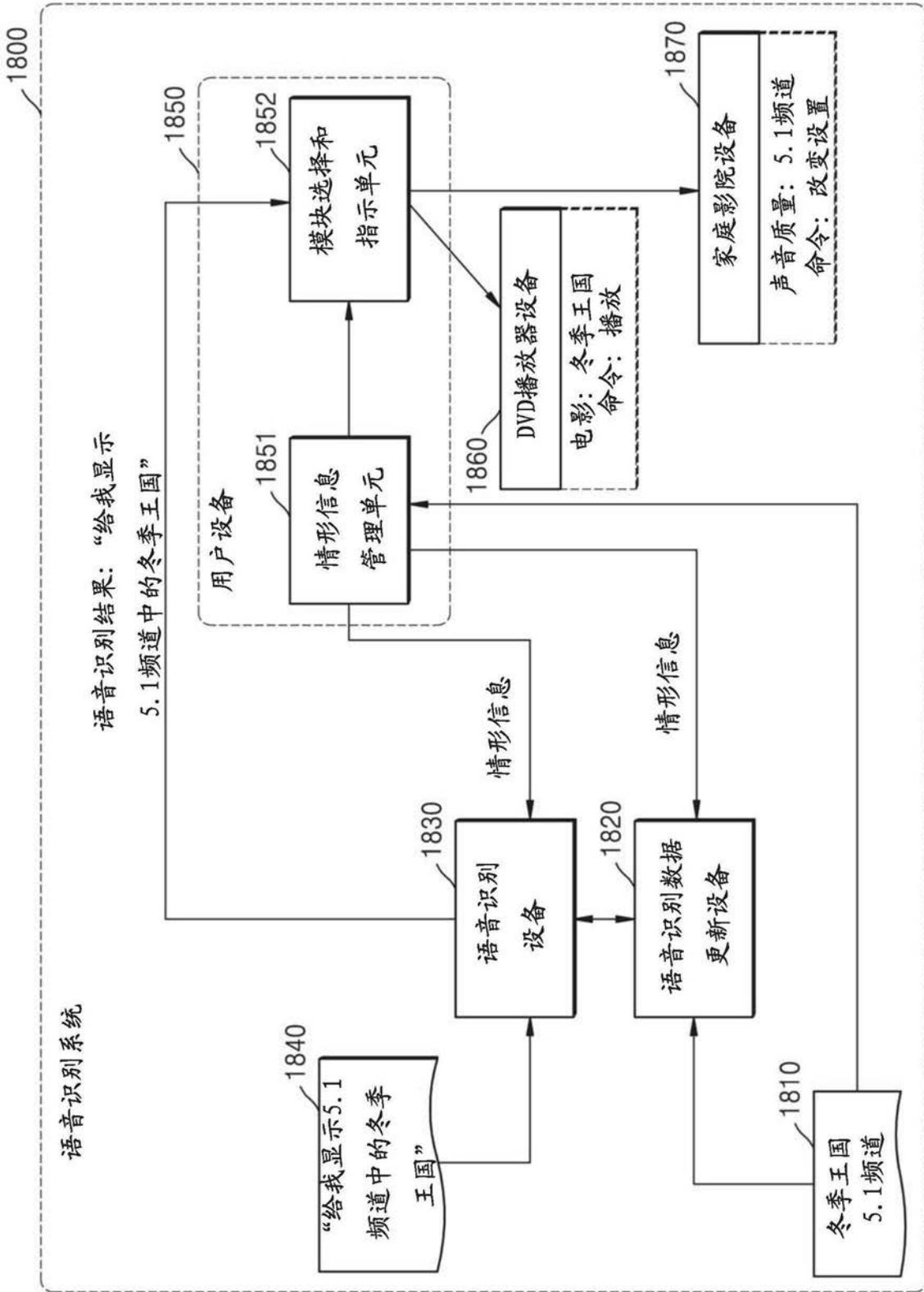


图18

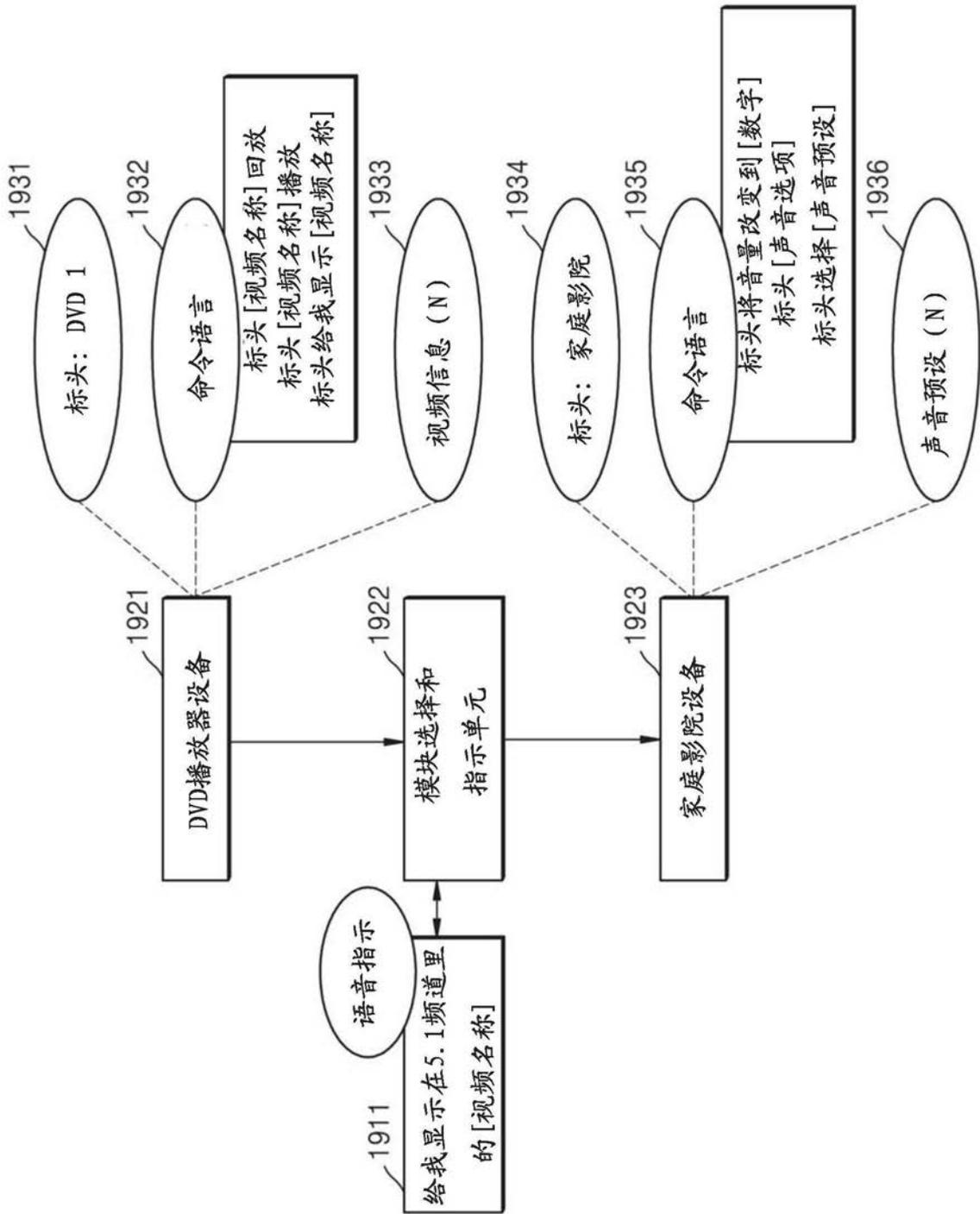


图19

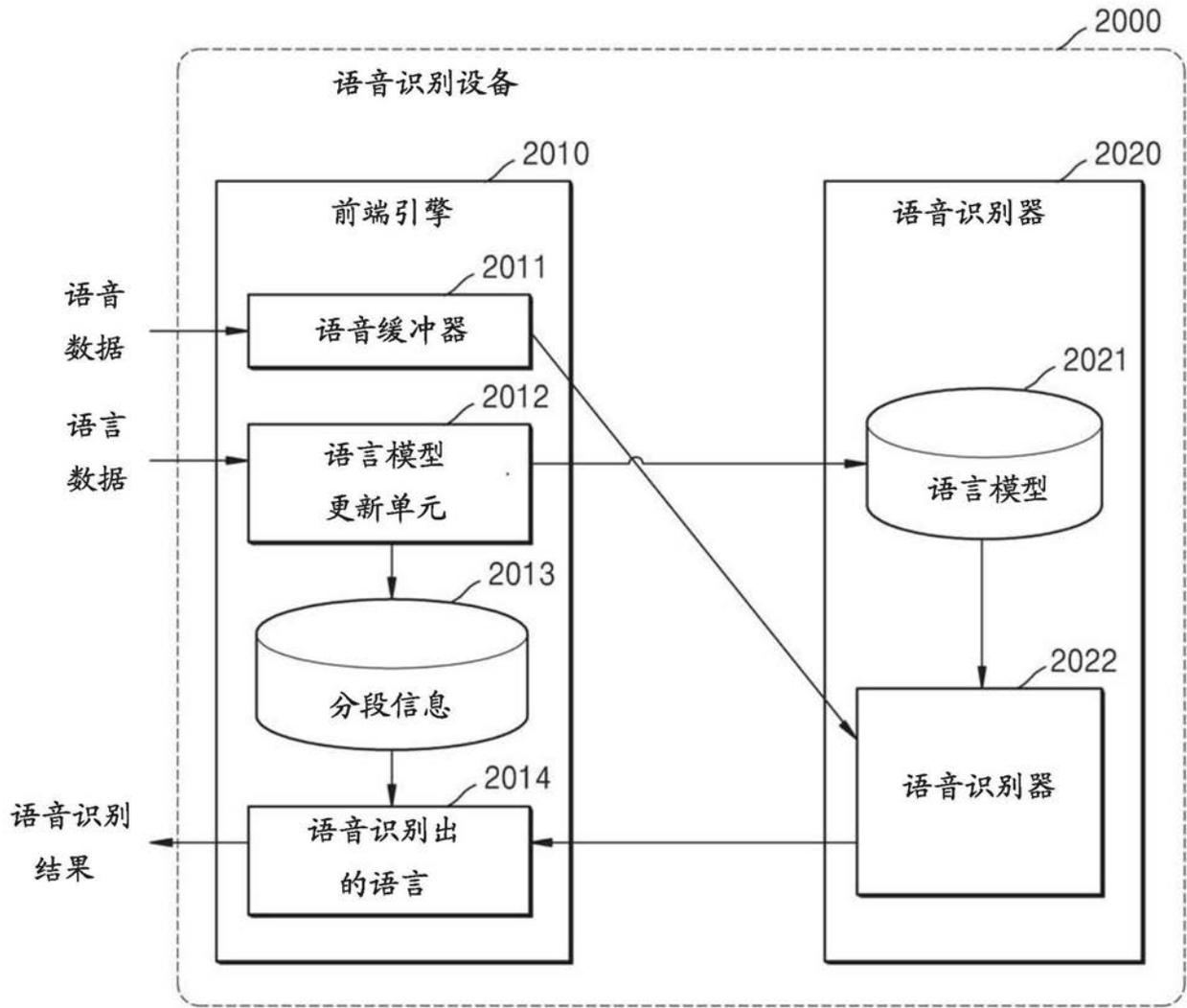


图20

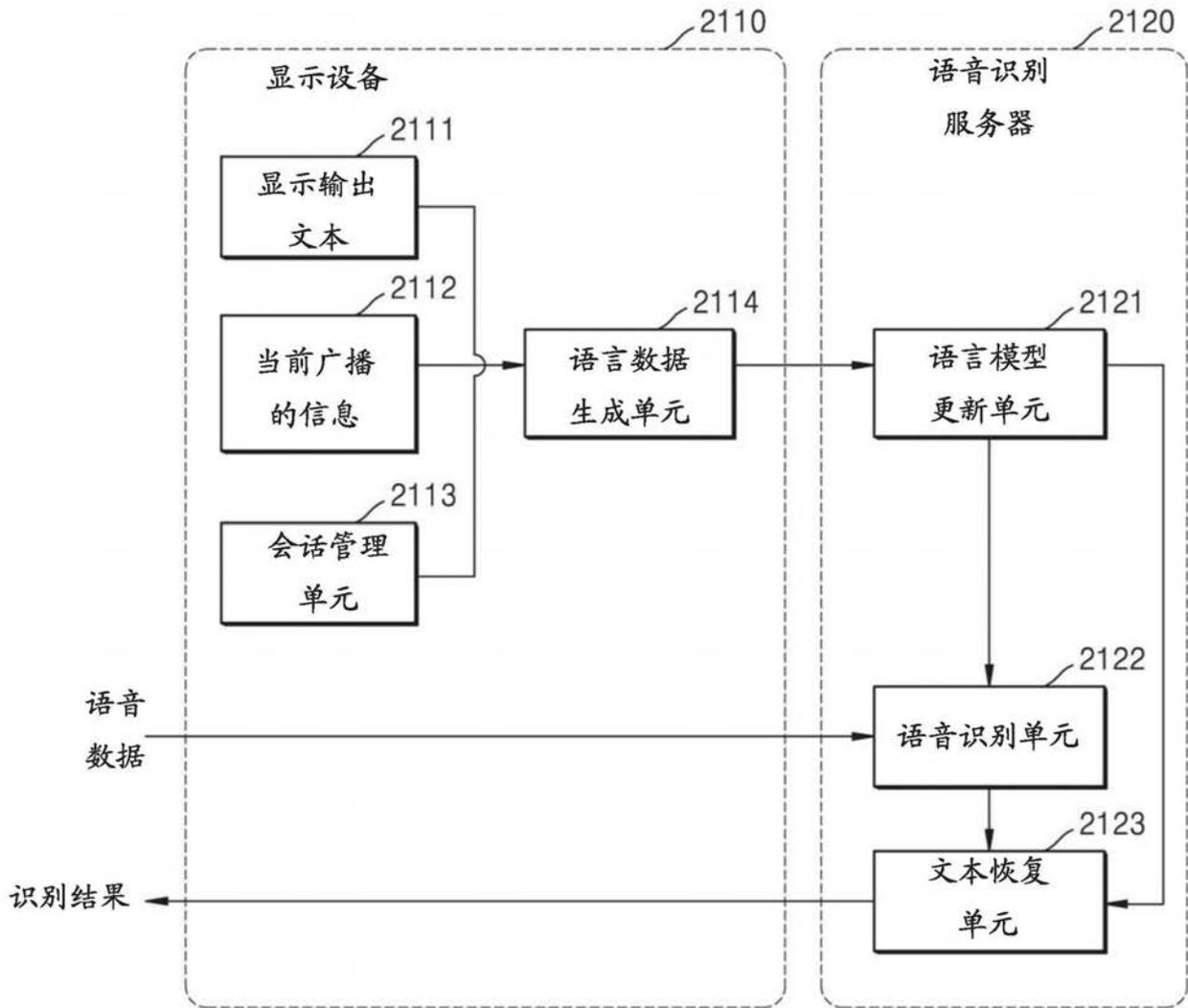


图21

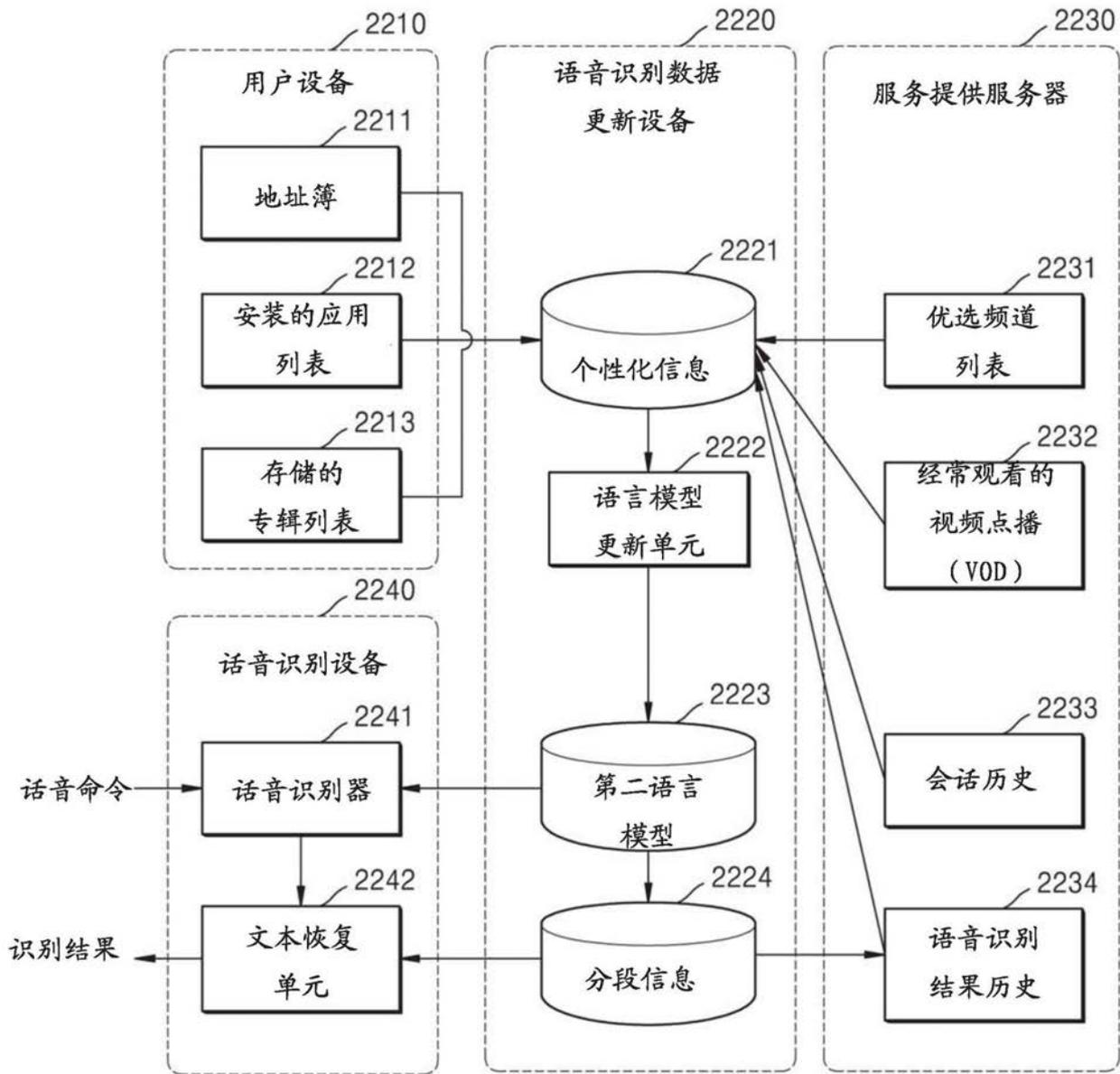


图22

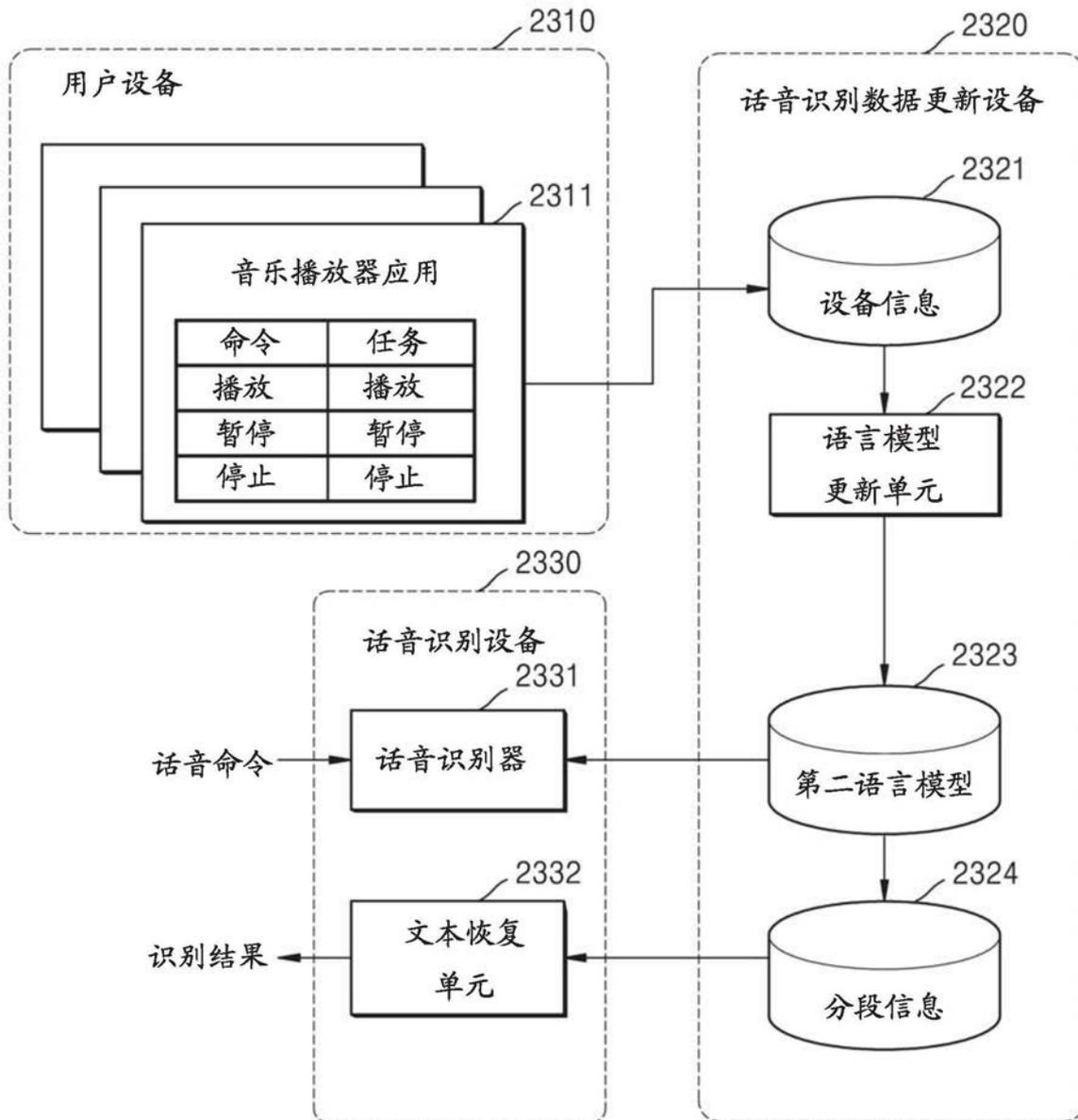


图23

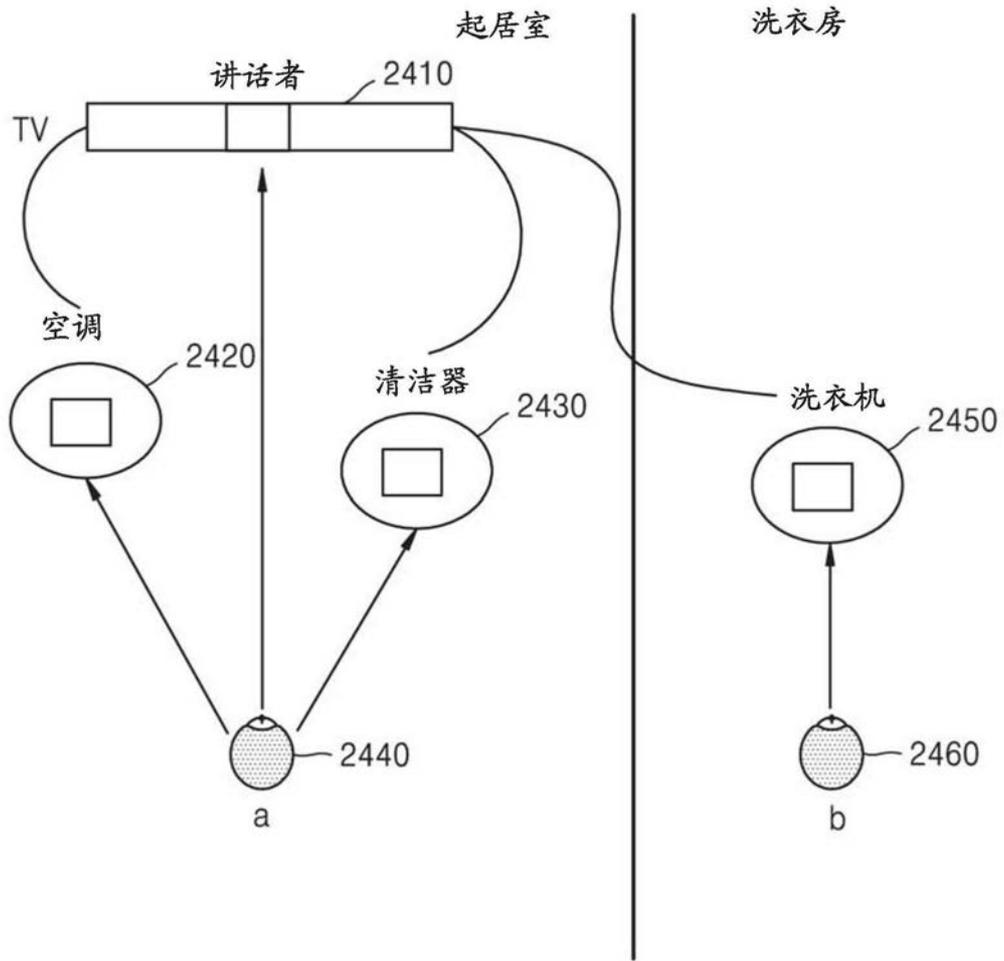


图24

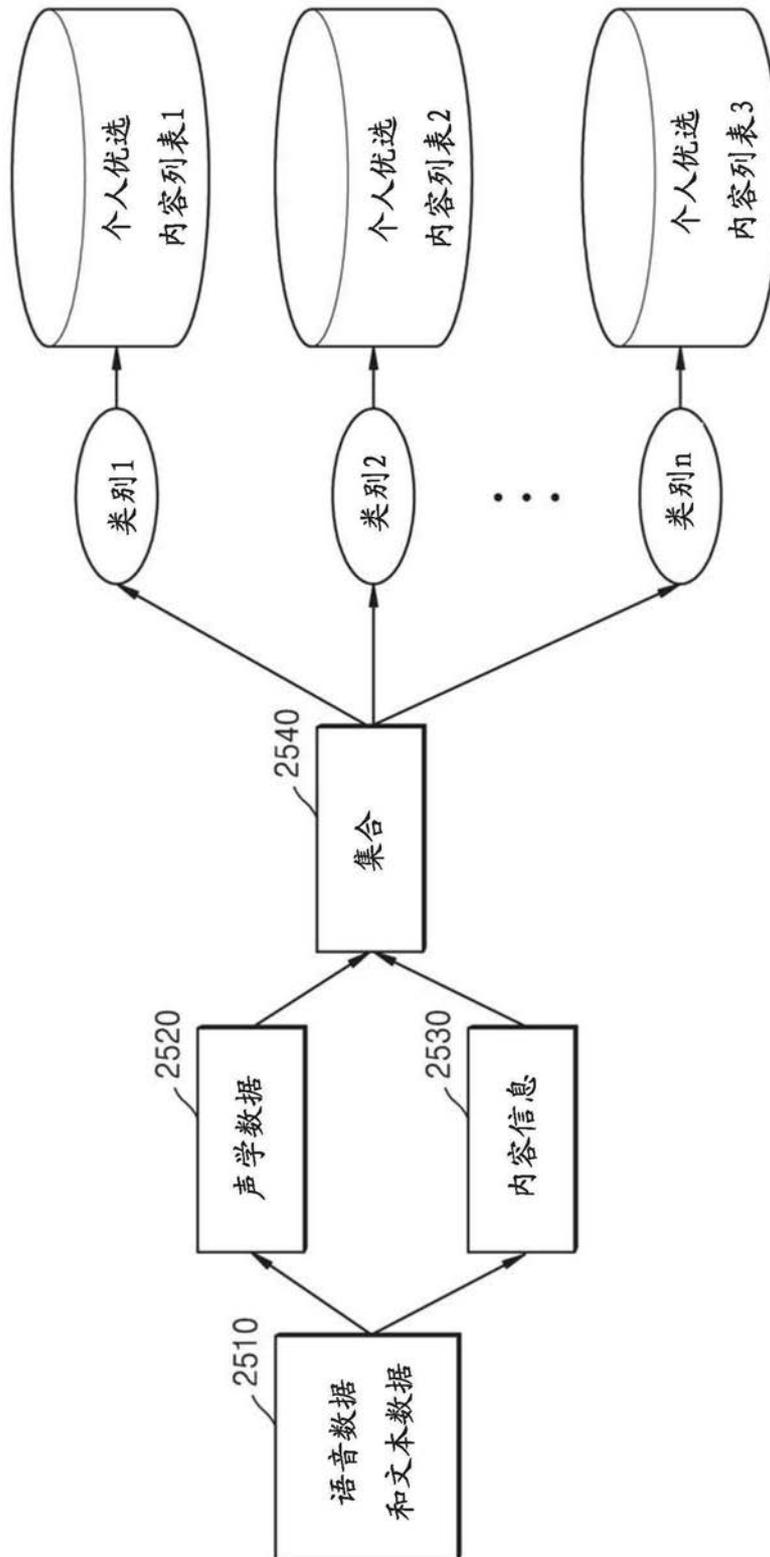


图25

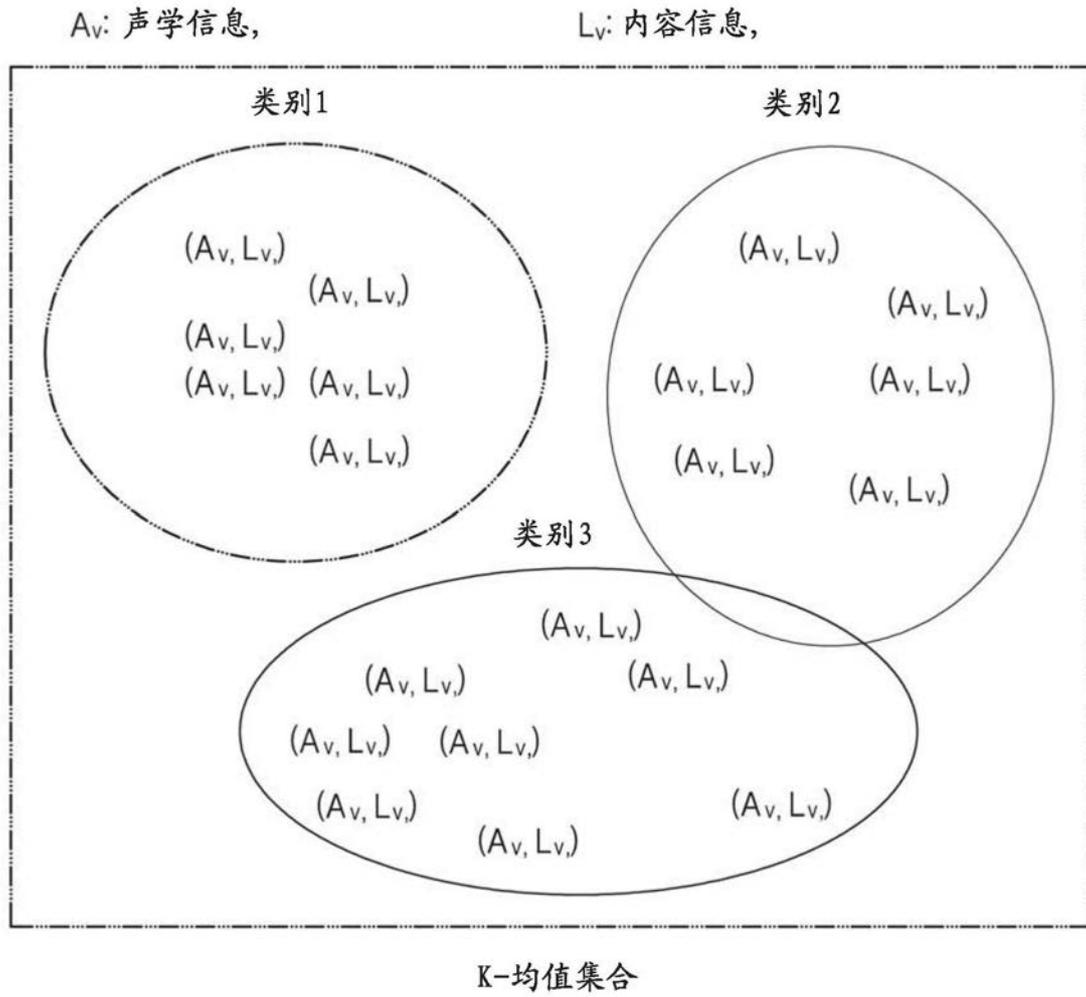


图26

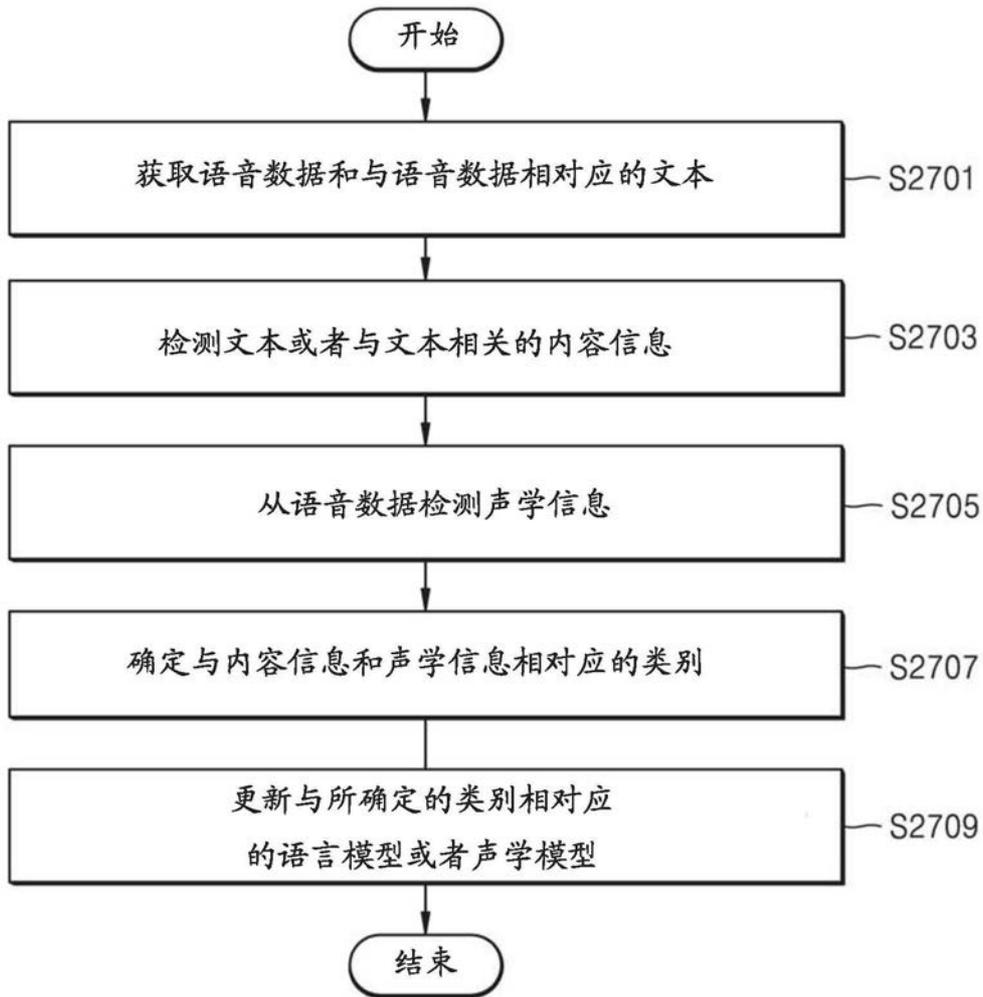


图27

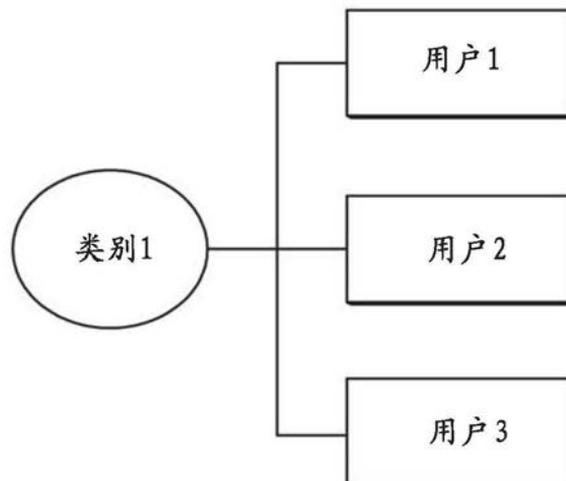


图28

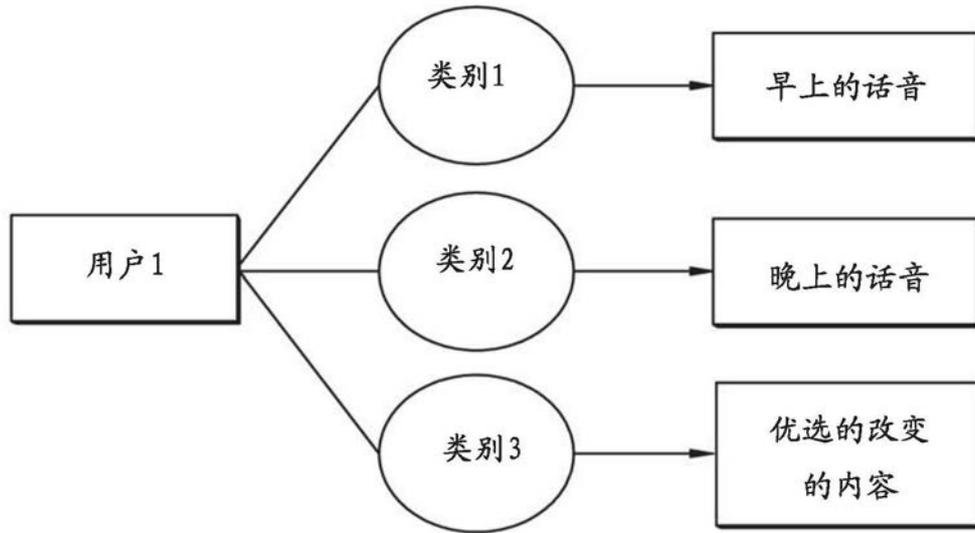


图29

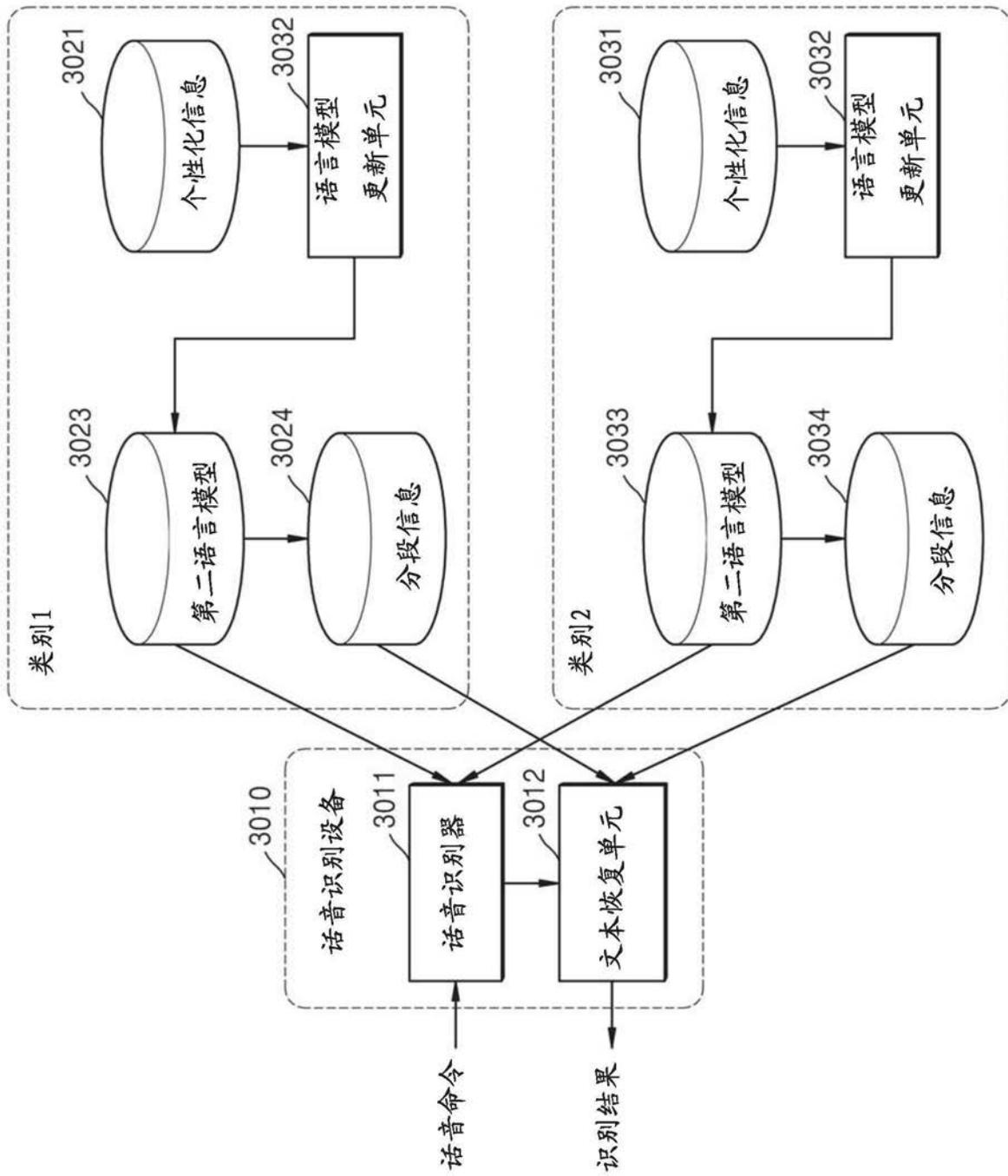


图30

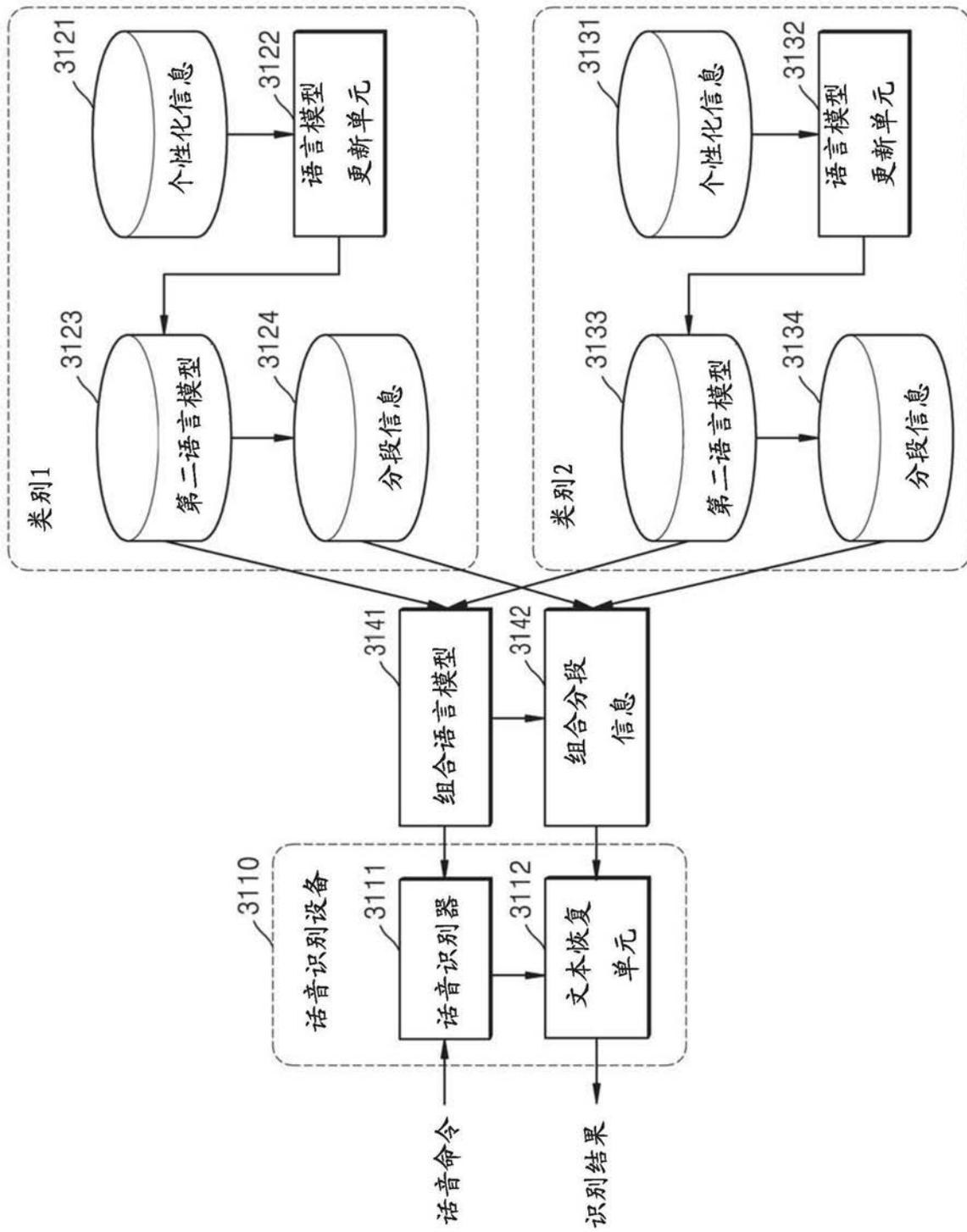


图31

3200

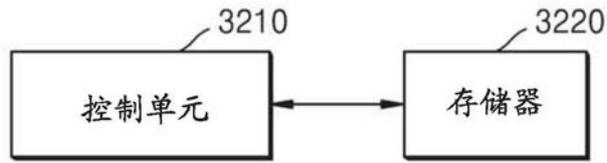


图32

3300

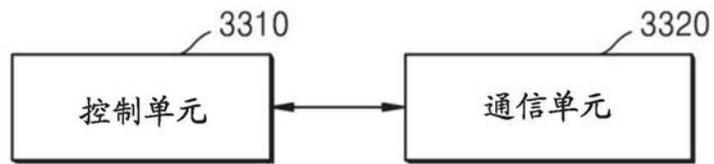


图33

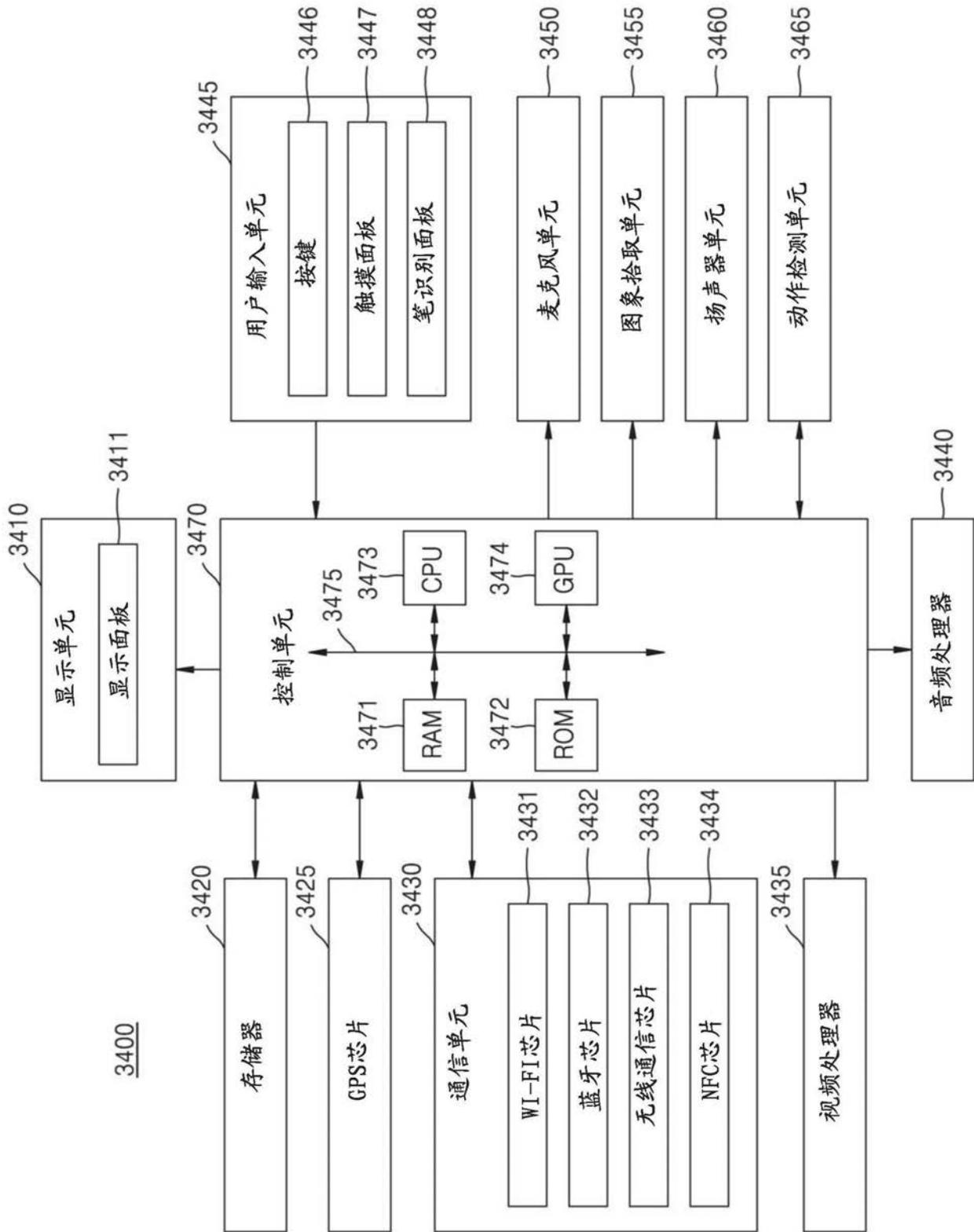


图34