



(12)发明专利

(10)授权公告号 CN 103544955 B

(45)授权公告日 2018.09.25

(21)申请号 201310301732.X

(22)申请日 2013.07.15

(65)同一申请的已公布的文献号

申请公布号 CN 103544955 A

(43)申请公布日 2014.01.29

(30)优先权数据

10-2012-0076809 2012.07.13 KR

(73)专利权人 三星电子株式会社

地址 韩国京畿道水原市

(72)发明人 李在原 陆东锡 林贤泽 金兑润

(74)专利代理机构 北京铭硕知识产权代理有限公司 11286

代理人 张云珠 韩明星

(51)Int.Cl.

G10L 15/26(2006.01)

G10L 15/02(2006.01)

(56)对比文件

GB 2240203 A, 1991.07.24,

GB 2240203 A, 1991.07.24,

CN 101326572 A, 2008.12.17,

CN 1465043 A, 2003.12.31,

WO 00/31723 A1, 2000.06.02,

US 7587321 B2, 2009.09.08,

CN 101120397 A, 2008.02.06,

US 2007/0271241 A1, 2007.11.22,

审查员 可杨

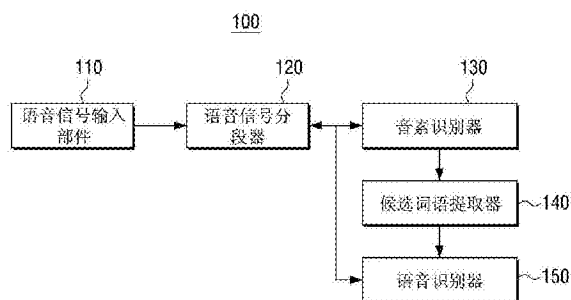
权利要求书1页 说明书7页 附图4页

(54)发明名称

识别语音的方法及其电子装置

(57)摘要

提供一种识别语音的方法及其电子装置。所述方法包括:以预设时间间隔将语音信号分段为多个片段;通过使用第一声学模型来执行针对语音信号的多个片段中的一个片段的音素识别;通过使用音素识别结果来提取所述语音信号的多个片段中的所述一个片段的候选词语;通过使用候选词语来执行针对所述语音信号的多个片段中的所述一个片段的语音识别。



1. 一种在电子装置中识别语音的方法,所述方法包括:
如果语音信号被输入,则以预设时间间隔将语音信号分段为多个片段;
通过使用第一声学模型来执行针对多个片段中的一个片段的语音信号的音素识别;
通过使用音素识别结果来提取所述一个片段的语音信号的候选词语;
通过使用候选词语来执行针对所述一个片段的语音信号的语音识别,
其中,执行音素识别的步骤还包括:
通过使用分段的维特比算法来删除所述一个片段的语音信号的多个音素中的至少一个最后音素,
其中,删除的至少一个音素被用于执行针对下一片段的语音信号的音素识别。
2. 如权利要求1的所述方法,其中,提取的步骤包括:
提取与识别的音素发音相似的相似音素;
通过使用所述相似音素来产生用于提取所述一个片段的候选词语的词语图表。
3. 如权利要求2的所述方法,其中,执行语音识别的步骤包括:
通过使用第二声学模型来计算所述一个片段的语音信号的高斯概率;
通过使用第二声学模型和语言模型来输出在词语图表中具有最高概率的词语串。
4. 如权利要求3所述的方法,其中,第一声学模型和第二声学模型彼此不同。
5. 如权利要求1所述的方法,其中,通过不同的核并行执行所述执行音素识别的步骤、所述提取步骤和所述执行语音识别的步骤。
6. 一种电子装置,包括:
语音信号输入部件,接收语音信号;
语音信号分段器,以预设时间间隔将通过语音信号输入部件输入的语音信号分段为多个片段;
音素识别器,通过使用第一声学模型来执行针对多个片段中的一个片段的语音信号的音素识别;
候选词语提取器,通过使用音素识别结果来提取所述一个片段的语音信号的候选词语;
语音识别器,通过使用候选词语来执行针对所述一个片段的语音信号的语音识别,
其中,音素识别器通过使用分段的维特比算法来删除所述一个片段的语音信号的多个音素中的至少一个最后音素,以执行音素识别,
其中,删除的至少一个音素被用于执行针对下一片段的语音信号的音素识别。
7. 如权利要求6所述的电子装置,其中,候选词语提取器提取与识别的音素发音相似的相似音素,并通过使用所述相似音素来产生用于提取所述片段的候选词语的词语图表。
8. 如权利要求7所述的电子装置,其中,语音识别器通过使用第二声学模型来计算所述一个片段的语音信号的高斯概率,并通过使用第二声学模型和语言模型来输出在词语图表中具有最高概率的词语串,以执行语音识别。
9. 如权利要求8所述的电子装置,其中,音素识别器的第一声学模型和语音识别器的第二声学模型彼此不同。
10. 如权利要求6所述的电子装置,其中,音素识别器、候选词语提取器和语音识别器被实现为不同的核。

识别语音的方法及其电子装置

[0001] 本申请要求于2012年7月13日提交到韩国知识产权局的第10-2012-0076809号韩国专利申请的优先权,其公开通过引用全部合并于此。

技术领域

[0002] 与示例性实施例一致的方法和设备总体涉及识别语音,更具体地讲,涉及通过使用声学模型、语言模型和发音字典识别用户发出的语音。

背景技术

[0003] 通过使用电子装置(诸如,智能电话、导航系统等)来控制语音识别。随着电子装置的硬件性能的提高以及用户对于语音识别的需求程度的增加,用户环境已经从用许多现有的命令识别用户的语音的孤立词语识别方法改变为识别多种自然语言的连续语音识别方法。

[0004] 连续语音识别方法从数十万到数百万词语中识别包括至少一个词语的词语串,并针对所有可用词语形成搜索空间。连续语音识别方法通过使用包括声学模型、语言模型和发音字典的信息来计算概率,以确定相应的发音是否是句子的类型,并根据计算结果获得识别的句子。

[0005] 然而,在连续语音识别方法中,搜索空间变得更大,因此,存储需求增加。由于计算次数的增加,语音识别是不可行的或者可能被极大地降低。

[0006] 因此,需要快速识别多种自然语言的语音识别方法。

发明内容

[0007] 示例性实施例解决至少上述问题和/或缺点以及上面未描述的其它缺点。另外,示例性实施例不需要克服上述缺点,并且示例性实施例可不克服任何上述问题。

[0008] 示例性实施例提供一种进一步快速识别多种自然语言的语音识别方法及其电子装置。

[0009] 根据示例性实施例的一方面,提供一种在电子装置中识别语音的方法。所述方法可包括:以预设时间间隔将语音信号分段为多个片段;基于第一声学模型来执行针对语音信号的多个片段中的一个片段的音素识别;基于音素识别的结果来提取所述语音信号的多个片段中的所述一个片段的候选词语;基于候选词语来执行针对所述多个片段中的所述一个片段的语音识别。

[0010] 执行音素识别的步骤可还包括:基于分段的维特比算法来删除所述语音信号的多个片段中的所述一个片段的多个音素的至少一个最后音素。删除的至少一个音素可被用于执行针对所述多个片段中的所述一个片段之后的语音信号的下一片段的音素识别。

[0011] 提取的步骤可包括:提取与识别的音素发音相似的相似音素;基于所述相似音素来产生用于提取所述多个片段中的所述一个片段的候选词语的词语图表。

[0012] 执行语音识别的步骤可还包括:基于第二声学模型来计算所述多个片段中的所述

一个片段的语音信号的高斯概率;基于第二声学模型和语言模型来输出在词语图表中具有最高概率的词语串。

[0013] 第一声学模型和第二声学模型可彼此不同。

[0014] 可通过不同的核并行执行音素识别的执行、提取和语音识别的执行。

[0015] 根据另一示例性实施例的一方面,提供一种电子装置,包括:语音信号输入部件,被配置为接收语音信号;语音信号分段器,被配置为以预设时间间隔将通过语音信号输入部件输入的语音信号分段为多个片段;音素识别器,被配置为基于第一声学模型来执行针对语音信号的多个片段中的一个片段的音素识别;候选词语提取器,被配置为基于音素识别的结果来提取所述语音信号的多个片段中的所述一个片段的候选词语;语音识别器,被配置为基于候选词语来执行针对所述多个片段中的所述一个片段的语音识别。

[0016] 音素识别器被配置为基于分段的维特比算法来删除所述语音信号的多个片段中的所述一个片段的多个音素中的至少一个最后音素,以执行音素识别。删除的至少一个音素可被用于执行针对所述多个片段中的所述一个片段之后的语音信号的下一片段的音素识别。

[0017] 候选词语提取器被配置为提取与识别的音素发音相似的相似音素,并基于所述相似音素来产生用于提取所述多个片段中的所述一个片段的候选词语的词语图表。

[0018] 语音识别器被配置为基于第二声学模型来计算所述多个片段中的所述一个片段的语音信号的高斯概率,并基于第二声学模型和语言模型来输出在词语图表中具有最高概率的词语串,以执行语音识别。

[0019] 音素识别器的第一声学模型和语音识别器的第二声学模型可彼此不同。

[0020] 音素识别器、候选词语提取器和语音识别器可被实现为电子装置的不同的核。

附图说明

[0021] 通过参照附图对特定示例性实施例的描述,上述和/或其它方面将更加清楚,其中:

[0022] 图1是示出根据示例性实施例的用于执行语音识别的电子装置的结构示意性框图;

[0023] 图2是示出根据示例性实施例的用于识别语音的图1的电子装置的详细结构的框图;

[0024] 图3是示出根据示例性实施例的处理并行语音识别的方法的示意图;

[0025] 图4是示出根据示例性实施例的识别语音的方法的流程图。

具体实施方式

[0026] 参照附图更加详细地描述示例性实施例。

[0027] 在下面的描述中,即使在不同的附图中,相同的附图参考标号也被用于相同的元件。提供在描述中被限定的内容(诸如详细的结构和元件)以帮助对示例性实施例的全面理解。因此,明显的是,在没有那些专门定义的内容的情况下,示例性实施例可被执行。此外,因为公知功能或结构会用不必要的细节模糊示例性实施例,所以不对公知功能或结构进行详细描述。

[0028] 图1是示出根据示例性实施例的用于执行语音识别的电子装置100的结构的示意性框图。参照图1,电子装置100包括:语音信号输入部件110、语音信号分段器120、音素识别器130、候选词语提取器140和语音识别器150。根据本示例性实施例的电子装置100可被实现为各种类型的电子装置(诸如,智能电话、智能电视(TV)、台式个人计算机(PC)、平板PC等)。因此,电子装置的上述元件可采取全硬件实施例(诸如,处理器或电路)、全软件实施例(包括固件、常驻软件、微码等)或组合软件和硬件的实施例的形式。

[0029] 语音信号输入部件110接收与用户发出的语音相应的语音信号。语音信号输入部件110可包括麦克风和用于放大接收到的语音的放大器。然而,通过使用麦克风实时接收语音信号仅是示例性实施例,因此,语音信号输入部件110可通过预存文件接收语音信号。

[0030] 语音信号分段器120将语音信号分段多个片段。详细地,语音信号分段器120可按照预设时间间隔(例如,0.1秒)将语音信号分段为多个片段。

[0031] 音素识别器130识别由语音信号分段器120分段的多个片段中的一个片段的语音信号的音素。详细地,音素识别器130可通过使用用于音素识别的第一声学模型来计算与所述一个片段的语音信号相应的特征向量的高斯概率分布,并选择最佳音素。

[0032] 音素识别器130可通过使用分段的维特比算法来删除所述一个片段的语音信号的多个音素中的至少最后一个音素。详细地,由于语音信号分段器120以时间为单位而不是以音素为单位分段语音信号,因此音素识别器130可能不能正确地识别位于所述一个片段的尾部的音素。因此,音素识别器130删除位于所述一个片段的尾部的至少一个音素,并将删除的至少一个音素输出到语音信号分段器120,以使用删除的至少一个音素用于识别下一片段的音素。

[0033] 候选词语提取器140通过使用从音素识别器130输出的音素识别结果来提取识别的音素的候选词语。详细地,候选词语提取器140提取与从音素识别器130输出的音素发音相似的相似音素,并通过使用所述相似音素来产生所述一个片段的语音信号的用于提取候选词语的词语图表。然而,产生所述一个片段的语音信号的词语图表来提取候选词语仅是示例性的,因此,可产生所述一个片段的语音信号的候选词语列表。候选词语提取器140将所述一个片段的语音信号的词语图表输出到语音识别器150。

[0034] 语音识别器150通过使用候选词语提取器140提取的候选词语来执行针对所述一个片段的语音信号的语音识别。详细地,语音识别器150可在从候选词语提取器140输出的词语图表中搜索从语音信号分段器120输出的所述一个片段的语音信号的最佳路径,以执行语音识别。

[0035] 音素识别器130、候选词语提取器140和语音识别器150可在处理器的不同的核或不同处理器中并行操作。换句话说,如果音素识别器130执行针对第一片段的语音信号的音素识别,则语音识别器150将第一片段的识别的结果发送到候选词语提取器140,并且执行针对第二片段的语音信号的音素识别。候选词语提取器140基于第一片段的音素识别结果来提取候选词语,将提取的候选词语输出到语音识别器150,并通过使用从音素识别器130输出的第二片段的语音识别的音素识别结果来提取候选词语。语音识别器150通过使用由候选词语提取器140提取的第一片段的语音信号的候选词语来执行针对第一片段的语音信号的语音识别,并通过使用由候选词语提取器140提取的第二片段的语音信号的候选词语来执行针对第二片段的语音信号的语音识别。

[0036] 上述电子装置100用相对少量的计算快速执行音素识别,基于音素识别的结果提取少量的候选词语,通过使用比识别多个连续词语的现有方法显著更小数量的候选词语来执行语音识别。另外,电子装置100并行执行音素识别、候选词语的提取和语音识别,来允许用户进一步快速执行语音识别。

[0037] 图2是示出根据示例性实施例的用于识别语音的电子装置100的详细结构的框图。参照图2,电子装置100包括语音信号输入部件110、语音信号分段器120、音素识别器130、候选词语提取器140和语音识别器150。

[0038] 语音信号输入部件110接收与用户的语音相应的语音信号。语音信号输入部件110可从语音输入装置(诸如,麦克风)实时接收语音信号。然而,这仅是示例,语音信号输入部件110可从存储在电子装置100的存储器(未示出)中的文件中接收语音信号。

[0039] 语音信号分段器120按照预设时间间隔将语音信号分段为多个片段。这里,语音信号分段器120包括片段分段器121、预处理器122和特征向量提取器123。

[0040] 片段分段器121按照预设时间间隔(例如,0.1秒)对从语音信号输入部件110输出的语音信号进行分段。

[0041] 预处理器122执行针对多个片段中的一个片段的语音信号的信号处理(诸如,噪声去除)。

[0042] 特征向量提取器123从预处理的所述一个片段的语音信号中提取特征向量。特征向量提取器123将所述一个片段的语音信号的特征向量输出到音素识别器130和语音识别器150。

[0043] 音素识别器130通过使用由特征向量提取器123提取的特征向量来执行语音识别。这里,音素识别器130包括第一高斯概率计算器131、第一声学模型132、最佳候选搜索器133和片段分段错误校正器134。

[0044] 第一高斯概率计算器131通过使用第一声学模型132来计算所述一个片段的语音信号的特征向量的高斯概率。

[0045] 第一声学模型132是用于音素识别的声学模型,并且在韩语的情况下,存储关于40到50个音素的信息。第一声学模型132可以是隐型马尔可夫模型(HMM)声学模型。具体地,第一声学模型132比应用于识别多个连续词语的现有方法的声学模型更容易实现,使得能够快速语音识别。

[0046] 最佳候选搜索器133基于第一声学模型132和第一高斯概率计算器131的计算结果来选择包括在所述一个片段的语音信号中的最佳音素。

[0047] 片段分段错误校正器134删除由最佳候选搜索器133选择的多个音素中的至少最后一个音素。详细地,根据本示例性实施例的语音信号分段器120基于时间而不是基于音素对语音信号进行分段。因此,输入到音素识别器130的所述一个片段的语音信号的最后音素的所有数据可不被输入,因此,由最佳候选搜索器133选择的所述多个音素中的至少最后一个音素可能是错误选择的音素。因此,片段分段错误校正器134删除由最佳候选搜索器133选择的所述多个音素中的至少最后一个音素,并将未被删除的音素输出到候选词语提取器140。片段分段错误校正器134将删除的至少一个音素输出到片段分段器121,以在下一片段中识别删除的至少一个音素。

[0048] 根据本示例性实施例的音素识别器130删除由最佳候选搜索器133选择的所述多

个音素中的至少最后一个音素,以通过第二分段错误校正器134校正片段分段错误。然而,这仅是示例,音素识别器130可通过使用HMM状态位置检查或信号处理技术来搜索音素的尾部,以最小化片段分段错误。

[0049] 候选词语提取器140基于由音素识别器130识别的所述一个片段的语音信号的音素来提取候选词语。候选词语提取器140包括相似度计算器141和片段词语图表产生器142。

[0050] 相似度计算器141通过使用发音字典来计算所述一个片段的语音信号的音素和其它音素之间的发音相似度,以提取与所述一个片段的语音信号的音素发音相似的相似音素。

[0051] 片段词语图表产生器142基于提取的相似音素产生用于产生候选词语的片段词语图表。这里,片段词语图表可以是识别的音素被连接到相似音素的网络类型图表。片段词语图表产生器142将用于提取所述一个片段的语音信号的候选词语的片段词语图表输出到最佳词语图表路径搜索器153。

[0052] 在上述示例性实施例中,候选词语提取器140产生片段词语图表,但是这仅是示例。因此,候选词语提取器140可提取候选词语来产生候选词语列表。

[0053] 语音识别器150通过使用从候选词语提取器140输出的候选词语来执行针对一个片段的语音识别。语音识别器150包括第二高斯概率计算器151、第二声学模型152、最佳词语图表路径搜索器153、语言模型154和语音识别输出部件155。

[0054] 第二高斯概率计算器151通过使用第二声学模型152来计算所述一个片段的语音信号的高斯概率分布。

[0055] 这里,第二声学模型152是在识别多个连续词语的一般方法中使用的声学模型,并可以是使用三音素的声学模型。具体地,为了执行复杂的语音识别,第二声学模型152存储比第一声学模型132更多的数据量。

[0056] 通过使用第二声学模型152来计算一个片段的语音信号的的高斯概率,并且通过使用第二声学模型152和语言模型154来输出在词语图表中具有最高概率的词语串。具体地,最佳词语图表路径搜索器153基于第二高斯概率计算器151的计算结果来搜索与从片段词语图表产生器142输出的片段词语图表的语音信号相应的最佳路径,其中,所述最佳路径与在词语图表中具有最高概率的词语串相应。这里,为了进一步准确识别句子,最佳词语路径搜索器153可通过使用存储语法和句子结构的语言模型154来执行语音识别。换句话说,第一声学模型132可以是专门用于高速语音识别的声学模型,第二声学模型152可以是用于提高连续词语语音识别的性能的精细的声学模型。

[0057] 语音识别输出部件155输出由最佳词语图表路径搜索器153搜索的最佳路径产生的词语串(句子)。

[0058] 音素识别器130、候选词语提取器140和语音识别器150可被形成为通过不同的核并行操作的流水线形状。详细地,如图3中所示,语音信号分段器120将语音信号分段为N个片段,并将N个片段的语音信号发送到音素识别器130。音素识别器130在时间t1执行针对第一片段的音素识别。在时间t2,音素识别器130执行针对第二片段的音素识别,并且候选词语提取器140提取第一片段的候选词语。在时间t3,音素识别器130执行针对第三片段的音素识别,候选词语提取器140提取第二片的候选词语,并且语音识别器150执行针对第一片段的语音识别。根据这种方法,音素识别器130、候选词语提取器140和语音识别器150每次

并行操作。语音识别器150在从用户结束发音的时间起的短时间 $t_{N+2}-t_N$ 之后,执行并输出针对所有片段的语音信号的语音识别。

[0059] 如上所述,电子装置100并行执行音素识别操作,使用音素识别的候选词语提取操作和使用候选词语的语音识别操作。因此,电子装置100比识别多个连续词语的现有方法更快速地执行语音识别。

[0060] 现在将参照图4描述根据示例性实施例的电子装置100的语音识别方法。

[0061] 参照图4,在操作S410中,电子装置100确定语音信号是否被输入。语音信号可通过语音输入装置(诸如,麦克风)或通过预存文件被实时输入。

[0062] 如果在操作S410中确定语音信号被输入,则在操作S420中,电子装置100以预设时间间隔将语音信号分段为多个片段。详细地,电子装置100以预设时间间隔(例如,0.1秒)将输入语音信号分段为多个片段,并执行针对多个片段中的一个片段的语音信号的信号处理来提取特征向量。

[0063] 在操作S430中,电子装置100识别所述一个片段的语音信号的音素。详细地,电子装置100通过使用第一声学模型来识别所述一个片段的语音信号的音素。为了进一步准确地识别音素,电子装置100删除多个识别的音素中的至少一个最后音素,并使用删除的至少一个音素来识别下一片段的语音信号的音素。

[0064] 在操作S440中,电子装置100通过使用音素识别结果来提取所述一个片段的语音信号的候选词语。详细地,电子装置100提取多个识别的音素的相似音素,并产生用于提取候选词语的词语图表。这里,词语图表是在其上识别的音素被分别连接到相似音素的网络类型图表。

[0065] 在操作S450中,电子装置100通过使用候选词语来执行针对所述一个片段的语音信号的语音识别。详细地,电子装置100通过使用第二声学模型和在操作S440中提取的候选词语(词语图表)的语言模型来执行针对所述一个片段的语音信号的语音识别。

[0066] 电子装置100可针对下一片段的语音信号重复执行操作S430到S450。电子装置100可通过使用处理器的不同核重复并行执行操作S430到S450。

[0067] 如上所述,根据语音识别方法,电子装置可比识别多个连续词语的现有方法更快速并且准确地执行语音识别。

[0068] 如本领域技术人员将理解的,本示例性实施例的多个方面可被实现为设备、系统、方法或计算机程序产品。因此,示例性实施例的多个方面可采取在这里一般全部可被称为“电路”、“模块”或“系统”的全硬件实施例、全软件实施例(包括固件、常驻软件、微码等)、或组合软件和硬件方面的实施例的形式。此外,示例性实施例的多个方面可采用计算机程序产品的形式,并由硬件处理器来执行,其中,所述计算机程序产品实现在包含计算机可读程序代码的一个或多个计算机可读介质中。

[0069] 一个或多个计算机可读介质的任意组合可被利用。计算机可读介质可以是计算机可读信号介质或计算机可读存储介质。计算机可读存储介质可以是,例如,但不限于,电子、磁、光学、电磁、红外或半导体系统、设备或装置、或者是上述的任何适当组合。计算机可读存储介质的更具体的示例(非穷举列表)将包括以下:具有一条或多条导线的电连接器、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或闪存)、光纤、便携式光盘只读存储器(CD-ROM)、光学存储装置、磁存储装置或上

述的任何适当组合。在本公开的上下文中,计算机可读存储介质可以是可包含或存储用于由指令执行系统、设备或装置使用的程序的任何有形介质,或是与指令执行系统、设备或装置连接的任何有形介质。

[0070] 上述示例性实施例仅仅是示例性的,而不应被解释为限制。本教导可被容易地应用到其它类型的设备。此外,示例性实施例的描述意在示出,而不是限制权利要求的范围,并且很多替代物、修改和变化对于本领域的技术人员将是清楚的。

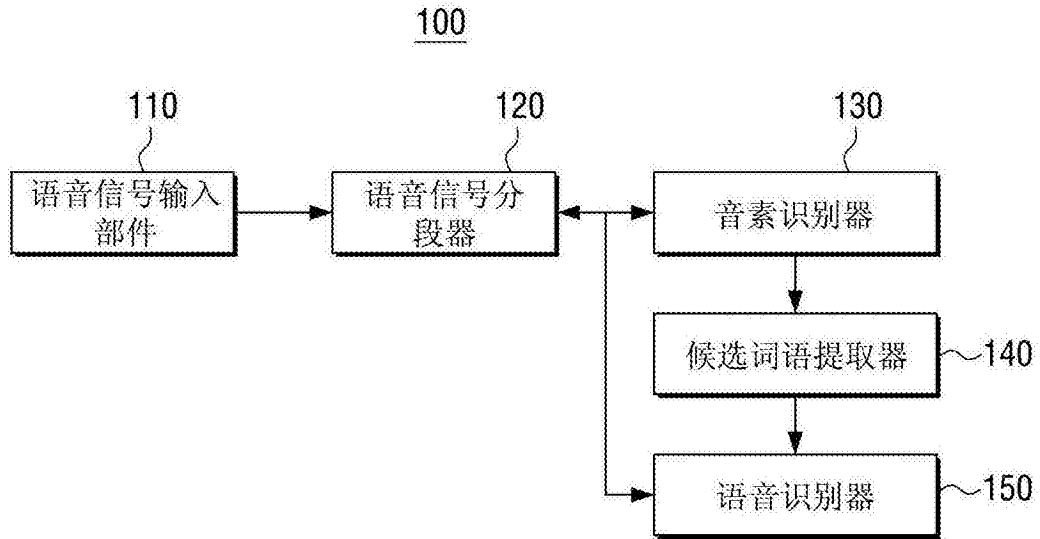


图1

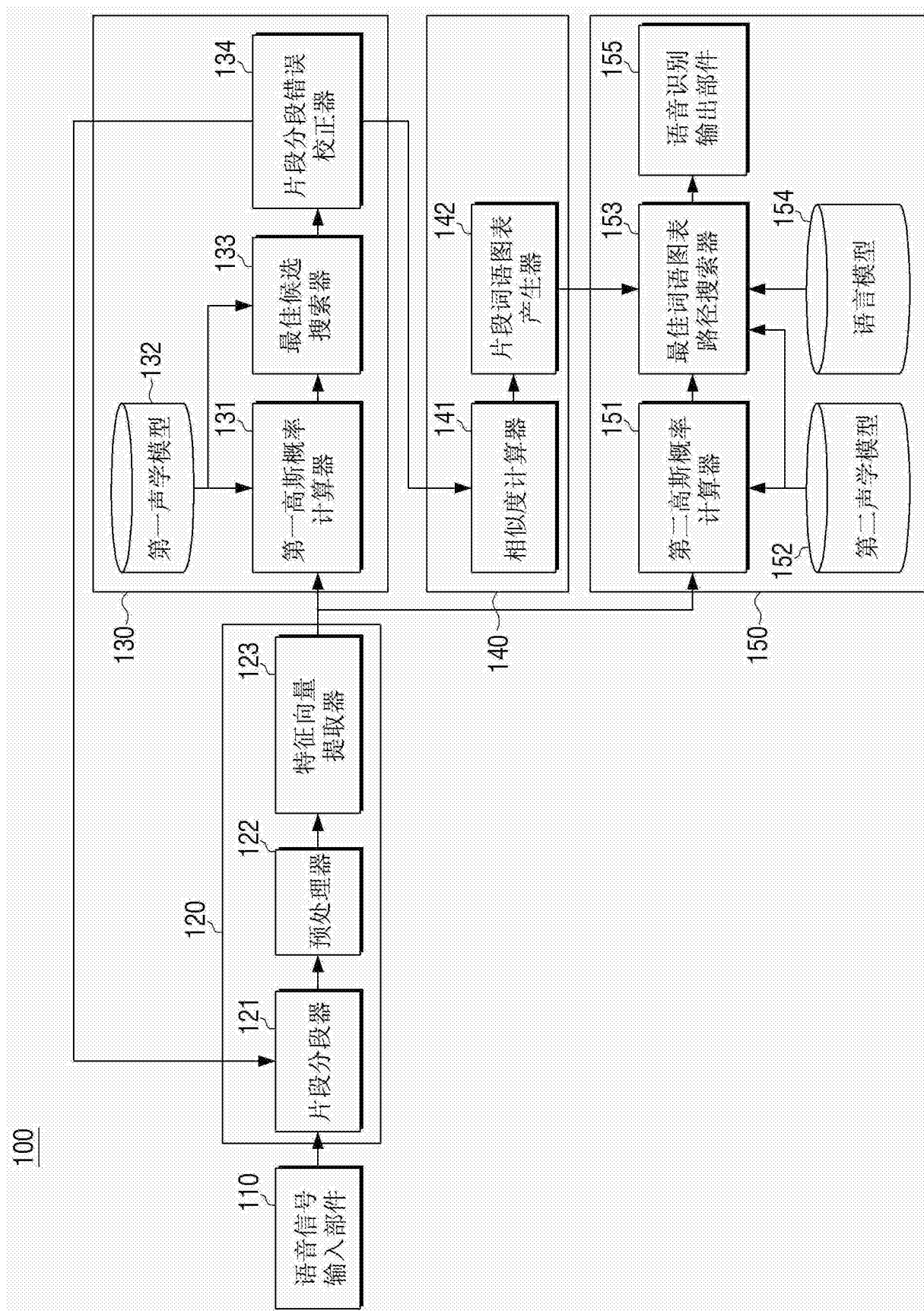


图2

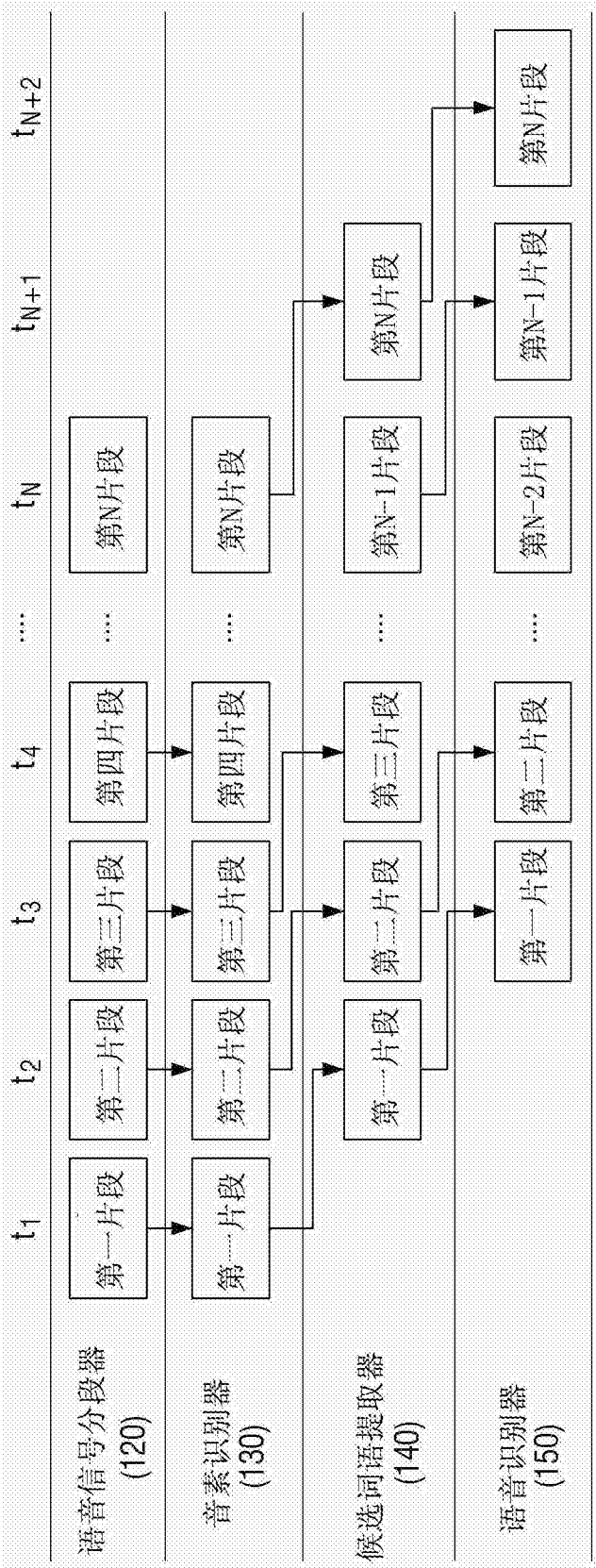


图3

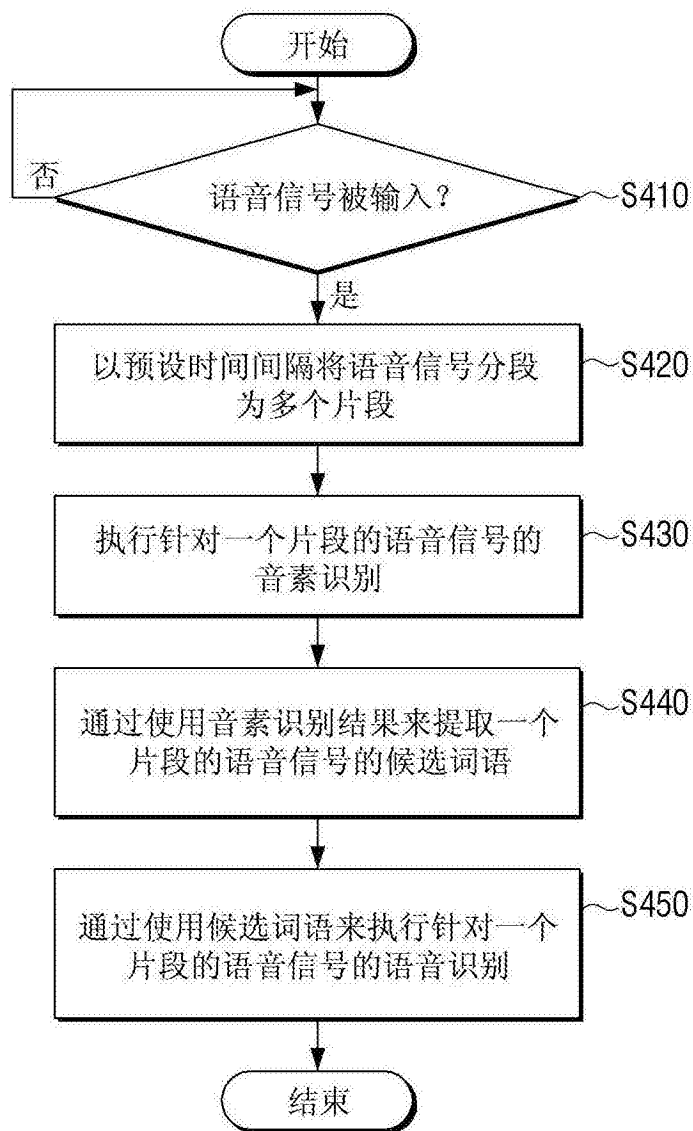


图4