

语音识别方法及装置、存储介质及电子设备

相关申请的交叉引用

本申请要求于 2021 年 1 月 18 日提交的申请号为 202110062350.0 名称为“语音识别方法及装置、存储介质及电子设备”的中国专利申请的优先权，该中国专利申请的全部内容通过引用全部并入本文。

技术领域

本申请涉及语音识别技术领域，具体而言，涉及一种语音识别方法、语音识别装置、计算机可读存储介质和电子设备。

背景技术

语音识别技术是一种将人的语音转换为文本的技术，广泛应用于各类人工智能产品中，比如智能对话机器人、智能音箱、智能翻译设备等，并主要通过人机对话的方式来实现信息的交换。

在日常的人机对话过程中，用户通常会通过回答例如“是/不是”、“对/不对”、“买/没买”等短字词来表明自己的意图，因此，对这些短字词识别的准确率也显得尤为重要。

现有的大规模连续语音识别框架所使用的模型训练语料中，上述短字词的占比非常小，并没有专门针对上述含有前置字的短字词进行识别的方法。

发明内容

根据本申请的一方面，提供一种语音识别方法，包括：获取样本语音信号，对所述样本语音信号进行解码，获得解码结果，从所述解码结果中提取第一特征，所述第一特征包含所述样本语音信号的前置字信息；从所述样本语音信号中抽取目标语音片段，获取所述目标语音片段的对数幅度谱；根据所述对数幅度谱确定第二特征，所述第二特征为所述样本语音信号的二维时频域特征；将所述第一特征和所述第二特征结合，得到第三特征；使用所述第三特征对未经训练的分类器进行训练，获得经训练的分类器；获取待识别语音信号的待识别第三特征，使用所述经训练的分类器对所述待识别第三特征进行分类，以确定所述待识别第三特征中是否含有所述前置字。

根据本申请的一方面，提供一种语音识别装置，包括：第一特征获取模块，用于获取样本语音信号，对所述样本语音信号进行解码，获得解码结果，从所述解码结果中提取第一特征，所述第一特征包含所述样本语音信号的前置字信息；第二特征获取模块，用于从所述样本语音信号中抽取目标语音片段，获取所述目标语音片段的对数幅度谱；根据所述对数幅度谱确定第二特征，所述第二特征为所述样本语音信号的二维时频域特征；第三特征获取模块，用于将所述第一特征和所述第二特征结合，得到第三特征；分类器训练模

块，用于使用所述第三特征对未经训练的分类器进行训练，获得经训练的分类器；分类识别模块，用于获取待识别语音信号的待识别第三特征，使用所述经训练的分类器对所述待识别第三特征进行分类，以确定所述待识别第三特征中是否含有所述前置字。

根据本申请的一方面，提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述的语音识别方法。

根据本申请的一方面，提供一种电子设备，包括处理器；存储器，用于存储一个或多个程序，当一个或多个程序被处理器执行时，使得所述处理器实现上述的语音识别方法。

本申请提供的技术方案可以包括以下有益效果：

本申请的示例性实施例中的语音识别方法及装置、存储介质及电子设备，一方面，通过10 将前置字信息和二维时频域特征作为样本语音信号的特征进行提取，并对分类器进行训练，以获得可以对前置字进行识别的经训练的分类器，使用该经训练的分类器对待识别语音信号进行识别，以确定待识别语音信号中是否含有前置字。一方面，该语音识别方法无需大量训练通用的语料，降低了算法的复杂度。另一方面，通过提出一种新的信号二维时频域特征，可以对语音信号时间轴和频率轴两个方向进行统计建模，针对短字词识别的特定场景，可以降低生活化的表达方式中的短语粘连现象、环境噪声以及远场等的影响，从而15 可以提高短字词识别的准确率。又一方面，针对含有前置字的短字词，提供了一种正确率更高的语音识别方法。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本申请。

20

附图说明

此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本申请的实施例，并与说明书一起用于解释本申请的原理。显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以25 根据这些附图获得其他的附图。

图 1 示意性示出了根据本申请的示例性实施方式的语音识别方法的流程图；

图 2 示意性示出了根据本申请的示例性实施方式的语音识别方法的步骤流程图；

图 3 示意性示出了根据本申请的示例性实施方式的语音识别方法中第一特征的方框图；

30 图 4 示意性示出了根据本申请的示例性实施方式的语音识别方法中获取第二特征的步骤流程图；

图 5 示意性示出了根据本申请的示例性实施方式的语音识别装置的方框图；

图 6 示意性示出了根据本申请的示例性实施方式的电子设备的模块示意图；

图 7 示意性示出了根据本申请的示例性实施方式的程序产品示意图。

具体实施方式

现在将参考附图更全面地描述示例实施例。然而，示例实施例能够以多种形式实施，
5 且不应被理解为限于在此阐述的范例；相反，提供这些实施例使得本申请将更加全面和完整，并将示例实施例的构思全面地传达给本领域的技术人员。图中相同的附图标记表示相同或类似的结构，因而将省略它们的详细描述。

虽然本说明书中使用相对性的用语，例如“上”“下”来描述图标的一个组件对于另一组件的相对关系，但是这些术语用于本说明书中仅出于方便，例如根据附图中所述的示例的方向。能理解的是，如果将图标的装置翻转使其上下颠倒，则所叙述在“上”的组件将会成为在“下”的组件。其他相对性的用语，例如“高”“低”“顶”“底”“左”“右”等也作具有类似含义。当某结构在其它结构“上”时，有可能是指某结构一体形成于其它结构上，或指某结构“直接”设置在其它结构上，或指某结构通过另一结构“间接”设置在其它结构上。

用语“一个”、“一”、“所述”用以表示存在一个或多个要素/组成区分/等；用语“包括”
15 和“具有”用以表示开放式的包括在内的意思并且是指除了列出的要素/组成区分/等之外还可存在另外的要素/组成区分/等。

人类用于交流的语言通常包括语音和文字两种表现方式。随着信息技术的发展，人和机器之间也需要进行大量的信息交换，目前计算机已经开始模拟人类交流信息的过程。

具体的，人类交流信息的过程包括：1、自然语言生成：将大脑产生的思想转换成语言；
20 2、语音合成：将语言转换成语音；3、语音识别：识别表达语言的语音内容；4、自然语言理解：理解语音所表达的语言意义。其中前两点是说者执行过程，后两点是听者执行过程。其中，语音识别是上述过程中的“识别表达语言的语音内容”，对于设备而言就是：识别人类所说的语音，并将其转化成文字。

语音识别是一种模式识别系统，语音识别主要包含如下步骤：1、语言输入；2、预处理；
25 3、特征提取，作为两个分支分别进入 4 的训练类聚以及 5~7 的识别操作；其中，5 是使用模板库的参考模式执行相似度比较；6 是执行识别过程中对 5 中结果进行失真检测，然后进入 7；7 是输出识别结果。

上述的预处理部分包括语音信号的采样、滤波等处理，特征提取的作用是从语音信号中提取出几组能描述其信号特征的特征参数，如能量、共振峰、倒谱系数等，以便于训练和识别。
30 建立语音识别系统的过程是，先用大量语音进行训练得到模板库，然后读取模板库的模板，和待识别语音进行相似度比较，得到识别结果。

语音识别技术中，声学模型和语言模型是自动语音识别技术的根本。声学模型用以从声音信号中识别出“音”，而语言模型用以把音转换成“字”。大规模非特定人的连续语音识别依赖于大规模声学和本地训练语料。为了达到良好的性能，训练过程中的声学模型

需要适配实际应用场景中不同种类的口音、噪声、语气变化、信道传输压缩和失真等；语言模型需要适配不同领域的专有名词、话术等。

然而，传统的孤立词或命令词识别技术和短字词识别相关，但是由于该类识别技术通常采用有限解码搜索空间，无法用于大规模非特定人连续语音识别的应用场景中。在大规模连续语音识别框架内，针对“是/不是”，“对/不对”，“在/不在”，“买/没买”类别的短字词识别，缺少特定的优化；并且这类短字词的占比较小，采用大规模语料重新训练模型的人力成本和时间成本均较高，且识别的提升效果不明显。

基于上述问题，本申请的示例性实施方式中提供了一种语音识别方法和装置，旨在提升在各业务领域和应用场景中频繁出现的短字词识别准确率的问题。该语音识别方法和装置广泛应用于各种设备中，例如：手机、电脑等各种设备中，本示例性实施方式对具体使用本申请实施例方案的设备不予限定。

图1示意性示出了本申请的示例性实施方式的语音识别方法的流程图。参考图1，语音识别方法可以包括以下步骤：

步骤S110、获取样本语音信号，对样本语音信号进行解码，获得解码结果，从解码结果中提取第一特征，第一特征包含样本语音信号的前置字信息。

在本申请的示例性实施方式中，样本语音信号是供分类器进行训练的信号样本。该信号样本可以是一段包含前置字的连续长语音信号或短语音信号，也可以是不包含前置字的连续长语音信号或短语音信号。在解码结果中包含前置字时，前置字信息为1，在解码结果中不包含前置字时，前置字信息为0。

在实际应用中，前置字可以是“不”、或“没”等，也可以是其它字，本示例性实施方式提供的语音识别方法对于含有任何一种前置字的短字词均可以进行识别。因此，本申请的示例性实施方式对于具体的前置字不作特殊限定。

本申请的示例性实施方式中，对获取的样本语音信号进行解码的过程中，可以使用现有经训练的声学模型和经训练的语言模型进行解码。通常在解码之后，会获得多个解码候选结果。

为了提高识别结果的准确性，本示例性实施方式中，可以抽取解码候选结果中排名前三的候选结果作为解码结果，以考虑准确率的同时，减少复杂度。

在实际应用中，也可以根据实际需要抽取排名前二或排名前四的候选结果，本示例性实施方式对此不作限定。

对于候选结果，以样本语音信号“京东”为例，解码后的候选结果按照匹配度排名可能是：“京东”、“叮咚”、“滴咚”等，那么就将这三个词作为解码结果来提取第一特征。

本申请的示例性实施方式中，从解码结果中提取第一特征具体可以包括：获取前置字信息，以及获取解码结果的声学模型得分和语言模型得分，对声学模型得分和语言模型得分进行归一化，得到归一化声学模型得分和归一化语言模型得分，作为第一特征。其中，

得分指的是一个字在解码结果中出现的概率，出现的概率越大，得分越高，也说明，识别出是这个字的可能性越大。

通过获取包含归一化声学模型得分、归一化语言模型得分，以及前置字信息的第一特征，通过该第一特征可以初步对含有前置字的或不含有前置字的样本语音信号进行表征。

5 本示例性实施方式中，对声学模型得分和语言模型得分进行归一化具体可以包括：将声学模型得分与最优解码结果的声学模型得分相除，获得归一化声学模型得分，将语言模型得分与最优解码结果的语言模型得分相除，获得归一化语言模型得分。其中，最优解码结果为解码候选结果中排名第一的候选结果。通过归一化可以将声学模型得分和语言模型得分统一到 0-1 区间内。

10 需要说明的是，在从解码结果中提取第一特征之前，本申请示例性实施方式提供的语音识别方法还需要确定包含前置字 A 的敏感词集合 $\Omega = \{AB, B, AC, C, \dots\}$ ，其中 B, C 表示和 A 组合的常见短语。例如对于“不”，B, C 通常为“在”、“是”、“对”等。当解码结果中包含敏感词集合 Ω 中的任一元素时，才从解码结果中提取第一特征。

步骤 S120、从样本语音信号中抽取目标语音片段，获取目标语音片段的对数幅度谱；
15 根据对数幅度谱确定第二特征，第二特征为样本语音信号的二维时频域特征。

在实际应用中，如果步骤 S110 解码结果中包含敏感词集合 Ω 中某一元素，则根据该解码结果的时间信息，可以确定出上述元素所对应的时间起始点和时间终止点。并且根据该对应的时间起始点和时间终止点，可以从样本语音信号中抽取该时间起始点和该时间终止点之间的语音片段作为目标语音片段，即含有敏感词集合 Ω 中某一元素的目标语音片段。

20 在本申请示例性实施方式中，还可以再将目标语音片段划分成预设个数 t 的子片段，对每个子片段都进行预设点数 f 的短时傅里叶变换，以获得语谱图。从而可以将一维信号变换成二维图，例如，当 $t=120$, $f=128$ 的时候，获得的语谱图就是一个 120×128 的二维图。对于一个短字的目标语音片段而言，其时间区间本身就很短，再划分成 t 个子片段后，每个子片段内的信号已经趋于平稳，因此，采用短时傅里叶变换在满足分辨率需求的情况下，
25 下，可以提高计算效率。

需要说明的是，上述短时傅里叶变换的帧长、窗口类型，以及帧移需要与解码时采用的帧长、窗口类型，以及帧移保持一致，从而可以保证所提取的第一特征和第二特征所表征信息的一致性。

接着，可以绘制语谱图的对数幅度谱，为了便于分析计算，需要将对数幅度谱归一化到 0 到 1 区间，得到归一化对数幅度谱 $Y(t, f)$ ；再从该归一化对数幅度谱 $Y(t, f)$ 中抽取第二特征。
30

本示例性实施方式中，在时间方向上，从归一化对数幅度谱 $Y(t, f)$ 中抽取第二特征具体可以包括：将归一化对数幅度谱 $Y(t, f)$ 划分为多个子频带；对子频带的子频带能量在时间方向上进行平滑操作，获得时间平滑子频带能量值 $P_T(t, f)$ 如公式 (1) 所示：

$$P_T(t, f) = \frac{1}{5} \sum_{i=-2}^2 Y(t+i, f) Y^*(t+i, f) \quad (1)$$

其中,时间平滑子频带能量值 $P_T(t, f)$ 是当前时刻的子频带能量与相邻时刻的子频带能量进行平滑操作得到的,例如通过将当前时刻的子频带能量与相邻时刻的子频带能量进行平均操作,可以获得时间平滑子频带能量值;式(1)中选择的相邻时刻是当前时刻的前2个时刻和后2个时刻的子频带能量。子频带能量是归一化对数幅度谱 $Y(t, f)$ 共轭复数 $Y(t+i, f)$ 和 $Y^*(t+i, f)$ 的乘积。

通过对上述子频带能量进行平滑操作,一方面可以抑制环境噪声的影响,另一方面可以得到更稳定的频带能量变化特性。

在本示例性实施方式中,在获得时间平滑子频带能量值 $P_T(t, f)$ 之后,可以根据时间平滑子频带能量值 $P_T(t, f)$,计算得到子频带能量时间跳变比 $r_T(t, f)$ 如公式(2)所示:

$$r_T(t, f) = P_T(t+5, f) / P_T(t, f) \quad (2)$$

其中,子频带能量时间跳变比 $r_T(t, f)$ 是预设时刻对应的时间平滑子频带能量值 $P_T(t+5, f)$ 与当前时刻对应的时间平滑子频带能量值 $P_T(t, f)$ 的商。

在实际应用中,预设时刻可以根据实际情况确定,例如,预设时刻与当前时刻间隔5个时刻等,本申请示例性实施方式对此不作限定。

由于前置字与后续字之间常存在短时间的间隔,当 t 处于间隔区时, $P_T(t, f)$ 较小,从而导致时间跳变比取得较大值。因此,通过计算时间跳变比的最大值,可以有效找出子频带中是否包含短时间间隔。

短时间间隔的一个重要特征是在各个子频带的 t 基本相等。因此,可以对每一个时刻 t 对应的多个子频带的子频带能量时间跳变比 $r_T(t, f)$ 进行平均,获得时刻 t 对应的全频带时间跳变比 $r_T(t) = \frac{1}{F} \sum_f r_T(t, f)$ 。

当短字词中存在短时间间隔时,一个明显特征是在抽取的时间范围内,存在较大值可以作为是否存在前置字的参考特征。基于此,我们得到 $\{r_T(1), r_T(2), \dots, r_T(N)\}$, N 为总帧数,并且可以统计出上述集合中的{最大值,平均值,标准差},即时间跳变比最大值、时间跳变比平均值和时间跳变比标准差。并且,可以将多个时刻对应的多个全频带时间跳变比 $r_T(t)$ 中获取的时间跳变比最大值、时间跳变比平均值和时间跳变比标准差,作为第二特征。

类似地,在频率方向上,从归一化对数幅度谱 $Y(t, f)$ 中抽取第二特征具体可以包括:对子频带的子频带能量在频率方向上进行平滑操作,获得频率平滑子频带能量值 $P_F(t, f)$;如公式(3)所示:

$$P_F(t, f) = \frac{1}{5} \sum_{i=-2}^2 Y(t, f+i) Y^*(t, f+i) \quad (3)$$

其中,频率平滑子频带能量值 $P_F(t, f)$ 是当前频率的子频带能量与相邻频率的子频带能量进行平滑操作得到的,例如通过将当前频率的子频带能量与相邻频率的子频带能量进行平均操作,可以获得频率平滑子频带能量值;式(3)中选择的相邻频率是当前频率的前

2 个频率和后 2 个频率的子频带能量。子频带能量是归一化对数幅度谱 $Y(t, f)$ 共轭复数 $Y(t, f + i)$ 和 $Y^*(t, f + i)$ 的乘积。

通过对上述子频带能量进行平滑操作，一方面可以抑制环境噪声的影响，另一方面可以得到更稳定的频带能量变化特性。

- 5 在本示例性实施方式中，在获得频率平滑子频带能量值 $P_F(t, f)$ 之后，可以根据频率平滑子频带能量值 $P_F(t, f)$ ，计算得到子频带能量频率跳变比 $r_F(t, f)$ 如公式 (4) 所示：

$$r_F(t, f) = P_F(t, f + 5) / P_F(t, f) \quad (4)$$

其中，子频带能量频率跳变比 $r_T(t, f)$ 是预设频率对应的频率平滑子频带能量值 $P_F(t, f + 5)$ 与当前频率对应的频率平滑子频带能量值 $P_F(t, f)$ 的商。

- 10 在实际应用中，预设频率可以根据实际情况确定，例如，预设频率与当前频率间隔 5 个频率等，本申请示例性实施方式对此不作限定。

由于前置字与后续字之间常存在短时间的间隔，当 t 处于间隔区时，由于处于静音区，各个频带之间的能力差异较小，使得子频带能量频率跳变比 $r_F(t, f)$ 的取值较小。相反，在语音段，子频带能量频率跳变比 $r_F(t, f)$ 将产生较大的值。

- 15 仍然对每一个时刻 t 对应的多个子频带的子频带能量频率跳变比 $r_F(t, f)$ 进行平均，获得时刻 t 对应的全频带频率跳变比 $r_F(t) = \frac{1}{F} \sum_f r_F(t, f)$ 。

当短字词中存在短时间间隔时，另一个明显特征是存在较小的 $r_F(t, f)$ 。基于此，我们得到 $\{r_F(1), r_F(2), \dots, r_F(N)\}$ ， N 为总帧数，并且可以统计出上述集合中的 {最小值，平均值，标准差}，即频率跳变比最小值、频率跳变比平均值和频率跳变比标准差。并且，可以

20 可以将多个时刻对应的多个全频带频率跳变比 $r_F(t)$ 中获取的频率跳变比最小值、频率跳变比平均值和频率跳变比标准差，作为第二特征。

至此，作为样本语音信号的二维时频域特征的第二特征就完全获得了，分别是：时间跳变比最大值、时间跳变比平均值、时间跳变比标准差、频率跳变比最小值、频率跳变比平均值和频率跳变比标准差。

- 25 步骤 S130、将第一特征和第二特征结合，得到第三特征。

在本示例性实施方式中，第一特征包含归一化声学模型得分、归一化语言模型得分，以及前置字信息；第二特征包含时间跳变比最大值、时间跳变比平均值、时间跳变比标准差、频率跳变比最小值、频率跳变比平均值和频率跳变比标准差；将上述 9 个特征结合拼接成一个第三特征。在实际应用中，第三特征可以以向量的形式来表示。

- 30 步骤 S140、使用第三特征对未经训练的分类器进行训练，获得经训练的分类器。

在本示例性实施方式中，使用第三特征对未经训练的朴素贝叶斯分类器进行训练，获得经训练的朴素贝叶斯分类器。其中，该朴素贝叶斯分类器可以是基于高斯分布的朴素贝叶斯分类器。由于朴素贝叶斯分类器的复杂度较低，在满足第三特征训练需求的情况下，可以提高训练的效率，达到无需重新训练语音识别的声学模型和语言模型的目的，并且可

以提高对短字词识别的正确率。此处对于具体的训练方法不作赘述。

步骤 S150、获取待识别语音信号的待识别第三特征，使用经训练的分类器对待识别第三特征进行分类，以确定待识别第三特征中是否含有前置字。

- 识别(Recognition)通常是将待识语音信号经过与训练时相同的分析，得到语音参数，
5 即待识别第三特征，并将该待识别第三特征输入经训练的朴素贝叶斯分类器中，得出判别结果，即待识别第三特征中是否含有前置字。

- 综上所述，基于本申请示例性实施方式的语音识别方法，通过将前置字信息和二维时频域特征作为样本语音信号的第三特征进行提取，并对朴素贝叶斯分类器进行训练，以获得可以对前置字进行识别的经训练的朴素贝叶斯分类器，使用该经训练的朴素贝叶斯分类器对待识别语音信号进行识别，以确定待识别语音信号中是否含有前置字。一方面，由于朴素贝叶斯分类器的复杂度较低，在满足第三特征训练需求的情况下，可以提高训练的效率，达到无需重新训练语音识别的声学模型和语言模型的目的。另一方面，通过提出一种新的信号二维时频域特征，可以对语音信号时间轴和频率轴两个方向进行统计建模，针对短字词识别的特定场景，可以降低生活化的表达方式中的短语粘连现象、环境噪声以及远场等的影响，从而可以提高短字词识别的准确率。又一方面，针对含有前置字的短字词，
10 本申请示例性实施方式提供了一种正确率更高的语音识别方法。
15

下面将参考图 2 对本申请示例性实施方式的语音识别方法的流程进行说明：

- 在步骤 S201 中，获取样本语音信号；在步骤 S202 中，对样本语音信号进行解码，获得解码结果；在步骤 S203 中，进入判断条件，判断解码结果中是否包含敏感词集合中的任一元素，该敏感词集合中包含有前置字；如果否，则结束。如果是，即存在敏感词集合元素，则执行步骤 S204，从界面结果中提取第一特征；另外，执行步骤 S205，确定敏感词集合元素对应的时间起始点和时间终止点，简称时间起止点；在步骤 S206 中，根据时间起止点获取目标语音片段；再执行步骤 S207，对目标语音片段进行短时傅里叶变换，获得对数幅度谱；接着，执行步骤 S208，对对数幅度谱进行归一化，获得归一化对数幅度谱；然后，执行步骤 S209，从归一化对数幅度谱中获取二维时频域特征作为第二特征；
20 在步骤 S210 中，将第一特征和第二特征结合，获得第三特征；在步骤 S211 中，使用第三特征对未经训练的朴素贝叶斯分类器进行训练，获得经训练的朴素贝叶斯分类器；在步骤 S212 中，通过经训练的朴素贝叶斯分类器对待识别语音信号的待识别第三特征进行分类，以确定待识别第三特征中是否含有前置字，简称识别待识别语音信号中是否含有前置字。
25

- 其中，如图 3 所示，第一特征 300 包含归一化声学模型得分 310、归一化语言模型得分 320，以及前置字信息 330。图 4 示出了从归一化对数幅度谱中获取二维时频域特征作为第二特征的过程，即第二特征的获取过程如下：
30

- 在步骤 S401 中，获得归一化对数幅度谱；在步骤 S402 中，从归一化对数幅度谱中获取时间平滑子频带能量值；接着，在步骤 S403 中，根据时间平滑子频带能量值，计算子频带能量时间跳变比；在步骤 S404 中，从归一化对数幅度谱中获取频率平滑子频带能量
35

值；接着，在步骤 S405 中，根据频率平滑子频带能量值，计算子频带能量频率跳变比；最后，在步骤 S406 中，从子频带能量时间跳变比获取时间跳变比最大值、时间跳变比平均值和时间跳变比标准差，从子频带能量频率跳变比中获取频率跳变比最小值、频率跳变比平均值和频率跳变比标准差，分别作为所述第二特征。

- 5 应当注意，尽管在附图中以特定顺序描述了本申请中方法的各个步骤，但是，这并非要求或者暗示必须按照该特定顺序来执行这些步骤，或是必须执行全部所示的步骤才能实现期望的结果。附加的或备选的，可以省略某些步骤，将多个步骤合并为一个步骤执行，以及/或者将一个步骤分解为多个步骤执行等。

进一步的，本示例实施方式中还提供了一种语音识别装置。

- 10 图 5 示意性示出了本申请的示例性实施方式的语音识别装置的方框图。参考图 5，根据本申请的示例性实施方式的语音识别装置 500 可以包括：第一特征获取模块 510、第二特征获取模块 520、第三特征获取模块 530、分类器训练模块 540、分类识别模块 550。

- 具体的，第一特征获取模块 510，可以用于获取样本语音信号，对样本语音信号进行解码，获得解码结果，从解码结果中提取第一特征，第一特征包含样本语音信号的前置字信息；第二特征获取模块 520，可以用于从样本语音信号中抽取目标语音片段，获取目标语音片段的对数幅度谱；根据对数幅度谱确定第二特征，第二特征为样本语音信号的二维时频域特征；第三特征获取模块 530，可以用于将第一特征和第二特征结合，得到第三特征；分类器训练模块 540，可以用于使用第三特征对未经训练的分类器进行训练，获得经训练的分类器；分类识别模块 550，可以用于获取待识别语音信号的待识别第三特征，使用经训练的分类器对待识别第三特征进行分类，以确定待识别第三特征中是否含有前置字。

20 由于本申请实施方式的语音识别装置的各个功能模块与上述方法实施方式中相同，在此不再赘述。

- 通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例性实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本申请实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、终端装置、或者网络设备等）执行根据本申请实施方式的方法。

- 30 此外，上述附图仅是根据本申请示例性实施例的方法所包括的处理的示意性说明，而不是限制目的。易于理解，上述附图所示的处理并不表明或限制这些处理的时间顺序。另外，也易于理解，这些处理可以是例如在多个模块中同步或异步执行的。

- 应当注意，尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元，但是这种划分并非强制性的。实际上，根据本申请的实施方式，上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之，上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

在本申请的示例性实施例中，还提供了一种能够实现上述方法的电子设备。

所属技术领域的技术人员能够理解，本发明的各个方面可以实现为系统、方法或程序产品。因此，本发明的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、微代码等），或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。

下面参照图 6 来描述根据本发明的这种实施方式的电子设备 600。图 6 显示的电子设备 600 仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

如图 6 所示，电子设备 600 以通用计算设备的形式表现。电子设备 600 的组件可以包括但不限于：上述至少一个处理单元 610、上述至少一个存储单元 620、连接不同系统组件（包括存储单元 620 和处理单元 610）的总线 630、显示单元 640。

其中，所述存储单元 620 存储有程序代码，所述程序代码可以被所述处理单元 610 执行，使得所述处理单元 610 执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的步骤。例如，所述处理单元 610 可以执行如图 1 中所示的步骤 S110、获取样本语音信号，对样本语音信号进行解码，获得解码结果，从解码结果中提取第一特征，第一特征包含样本语音信号的前置字信息；步骤 S120、从样本语音信号中抽取目标语音片段，获取目标语音片段的对数幅度谱；根据对数幅度谱确定第二特征，第二特征为样本语音信号的二维时频域特征；步骤 S130、将第一特征和第二特征结合，得到第三特征；步骤 S140、使用第三特征对未经训练的分类器进行训练，获得经训练的分类器；步骤 S150、获取待识别语音信号的待识别第三特征，使用经训练的分类器对待识别第三特征进行分类，以确定待识别第三特征中是否含有前置字。

存储单元 620 可以包括易失性存储单元形式的可读介质，例如随机存取存储单元（RAM）6201 和/或高速缓存存储单元 6202，还可以进一步包括只读存储单元（ROM）6203。

存储单元 620 还可以包括具有一组（至少一个）程序模块 6205 的程序/实用工具 6204，这样的程序模块 6205 包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

总线 630 可以为表示几类总线结构中的一种或多种，包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

电子设备 600 也可以与一个或多个外部设备 670（例如键盘、指向设备、蓝牙设备等）通信，还可与一个或者多个使得用户能与该电子设备 600 交互的设备通信，和/或与使得该电子设备 600 能与一个或多个其它计算设备进行通信的任何设备（例如路由器、调制解调器等等）通信。这种通信可以通过输入/输出（I/O）接口 650 进行。并且，电子设备 600 还可以通过网络适配器 660 与一个或者多个网络（例如局域网（LAN），广域网（WAN）和/或公共网络，例如因特网）通信。如图所示，网络适配器 660 通过总线 630 与电子设

备 600 的其它模块通信。应当明白，尽管图中未示出，可以结合电子设备 600 使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储系统等。

通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式
5 可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本申请实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、终端装置、或者网络设备等）执行根据本申请实施方式的方法。

10 在本申请的示例性实施例中，还提供了一种计算机可读存储介质，其上存储有能够实现本说明书上述方法的程序产品。在一些可能的实施方式中，本发明的各个方面还可以实现为一种程序产品的形式，其包括程序代码，当所述程序产品在终端设备上运行时，所述程序代码用于使所述终端设备执行本说明书上述“示例性方法”部分中描述的根据本发明各种示例性实施方式的步骤。

15 参考图 7 所示，描述了根据本发明的实施方式的用于实现上述方法的程序产品 700，其可以采用便携式紧凑盘只读存储器（CD-ROM）并包括程序代码，并可以在终端设备，例如个人电脑上运行。然而，本发明的程序产品不限于此，在本文件中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

20 所述程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑
25 盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。

计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结
30 合使用的程序。

可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF 等等，或者上述的任意合适的组合。

可以以一种或多种程序设计语言的任意组合来编写用于执行本发明操作的程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、C++等，还包括常规的过程
35 式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计

算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网（LAN）或广域网（WAN），连接到用户计算设备，或者，可以连接到外部计算设备（例如利用因特网服务提供商来通过因特网连接）。

此外，上述附图仅是根据本发明示例性实施例的方法所包括的处理的示意性说明，而不是限制目的。易于理解，上述附图所示的处理并不表明或限制这些处理的时间顺序。另外，也易于理解，这些处理可以是例如在多个模块中同步或异步执行的。

本领域技术人员在考虑说明书及实践这里公开的内容后，将容易想到本申请的其他实施例。本申请旨在涵盖本申请的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本申请的真正范围和精神由权利要求指出。

应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限定。

权利要求

1. 一种语音识别方法，包括：

获取样本语音信号，对所述样本语音信号进行解码，获得解码结果，从所述解码结果中提取第一特征，所述第一特征包含所述样本语音信号的前置字信息；

5 从所述样本语音信号中抽取目标语音片段，获取所述目标语音片段的对数幅度谱；根据所述对数幅度谱确定第二特征，所述第二特征为所述样本语音信号的二维时频域特征；

将所述第一特征和所述第二特征结合，得到第三特征；

使用所述第三特征对未经训练的分类器进行训练，获得经训练的分类器；

10 获取待识别语音信号的待识别第三特征，使用所述经训练的分类器对所述待识别第三特征进行分类，以确定所述待识别第三特征中是否含有所述前置字。

2. 根据权利要求1所述的方法，其中，对所述样本语音信号进行解码，获得解码结果包括：

使用经训练的声学模型和经训练的语言模型对所述样本语音信号进行解码，

15 抽取解码候选结果中排名前三的候选结果作为所述解码结果。

3. 根据权利要求1或2所述的方法，其中，从所述解码结果中提取第一特征还包括：

获取所述解码结果的声学模型得分和语言模型得分，对所述声学模型得分和所述语言模型得分进行归一化，得到归一化声学模型得分和归一化语言模型得分，

20 作为所述第一特征。

4. 根据权利要求3所述的方法，其中，对所述声学模型得分和所述语言模型得分进行归一化，得到归一化声学模型得分和归一化语言模型得分包括：

将所述声学模型得分与最优解码结果的声学模型得分相除，获得所述归一化声学模型得分，所述最优解码结果为解码候选结果中排名第一的候选结果；

25 将所述语言模型得分与所述最优解码结果的语言模型得分相除，获得所述归一化语言模型得分。

5. 根据权利要求1所述的方法，其中，还包括：

在所述解码结果包含前置字时，所述前置字信息为1；

在所述解码结果不包含所述前置字时，所述前置字信息为0。

30 6. 根据权利要求5所述的方法，其中，还包括：

所述前置字为不、没中的至少一个。

7. 根据权利要求1所述的方法，其中，从所述解码结果中提取第一特征之前，

所述方法还包括：

确定包含前置字的敏感词集合，当所述解码结果中包含所述敏感词集合中的任一元素时，从所述解码结果中提取所述第一特征。

- 5 8. 根据权利要求 7 所述的方法，其中，从所述样本语音信号中抽取语音片段包括：

根据所述解码结果的时间信息，确定所述元素对应的时间起始点和时间终止点；

从所述样本语音信号中抽取所述时间起始点和所述时间终止点之间的语音片段作为所述目标语音片段。

- 10 9. 根据权利要求 1 或 8 所述的方法，其中，获取所述目标语音片段的对数幅度谱包括：

将所述目标语音片段划分成预设个数的子片段，对每个所述子片段进行预设点数的短时傅里叶变换，获得语谱图；

根据所述语谱图获取所述对数幅度谱。

- 15 10. 根据权利要求 9 所述的方法，其中，根据所述对数幅度谱确定第二特征包括：

将所述对数幅度谱归一化到 0 到 1 区间，得到归一化对数幅度谱；

从所述归一化对数幅度谱中抽取所述第二特征。

- 20 11. 根据权利要求 10 所述的方法，其中，从所述归一化对数幅度谱中抽取所述第二特征包括：

将所述归一化对数幅度谱划分为多个子频带；

对所述子频带的子频带能量在时间方向上进行平滑操作，获得时间平滑子频带能量值；

根据所述时间平滑子频带能量值，计算子频带能量时间跳变比；

- 25 对每一个时刻对应的多个所述子频带的所述子频带能量时间跳变比进行平均，获得所述时刻对应的全频带时间跳变比；

将多个时刻对应的多个所述全频带时间跳变比中获取的时间跳变比最大值、时间跳变比平均值和时间跳变比标准差，作为所述第二特征；

- 30 对所述子频带的子频带能量在频率方向上进行平滑操作，获得频率平滑子频带能量值；

根据所述频率平滑子频带能量值，计算子频带能量频率跳变比；

对每一个时刻对应的多个所述子频带的所述子频带能量频率跳变比进行平均，

获得所述时刻对应的全频带频率跳变比；

将多个时刻对应的多个所述全频带频率跳变比中获取频率跳变比最小值、频率跳变比平均值和频率跳变比标准差，作为所述第二特征。

12. 根据权利要求 11 所述的方法，其中，对所述子频带的子频带能量在时间方向上进行平滑操作，获得时间平滑子频带能量值包括：

将当前时刻的子频带能量与相邻时刻的子频带能量进行平均，获得所述时间平滑子频带能量值；

对所述子频带的子频带能量在频率方向上进行平滑操作，获得频率平滑子频带能量值包括：

10 将当前频率的子频带能量与相邻频率的子频带能量进行平均，获得所述频率平滑子频带能量值。

13. 根据权利要求 11 所述的方法，其中，根据所述时间平滑子频带能量值，计算子频带能量时间跳变比包括：

15 获取预设时刻对应的所述时间平滑子频带能量值与当前时刻对应的所述时间平滑子频带能量值的商，作为所述子频带能量时间跳变比；

根据所述频率平滑子频带能量值，计算子频带能量频率跳变比包括：

获取预设频率对应的所述频率平滑子频带能量值与当前频率对应的所述频率平滑子频带能量值的商，作为所述子频带能量频率跳变比。

14. 根据权利要求 13 所述的方法，其中，所述预设时刻与所述当前时刻间隔 20 5 个时刻，所述预设频率与所述当前频率间隔 5 个频率。

15. 根据权利要求 1 所述的方法，其中，使用所述第三特征对未经训练的分类器进行训练，获得经训练的分类器包括：

使用所述第三特征对未经训练的朴素贝叶斯分类器进行训练，获得经训练的朴素贝叶斯分类器。

25 16. 一种语音识别装置，包括：

第一特征获取模块，用于获取样本语音信号，对所述样本语音信号进行解码，获得解码结果，从所述解码结果中提取第一特征，所述第一特征包含所述样本语音信号的前置字信息；

30 第二特征获取模块，用于从所述样本语音信号中抽取目标语音片段，获取所述目标语音片段的对数幅度谱；根据所述对数幅度谱确定第二特征，所述第二特征为所述样本语音信号的二维时频域特征；

第三特征获取模块，用于将所述第一特征和所述第二特征结合，得到第三特

征；

分类器训练模块，用于使用所述第三特征对未经训练的分类器进行训练，获得经训练的分类器；

- 5 分类识别模块，用于获取待识别语音信号的待识别第三特征，使用所述经训练的分类器对所述待识别第三特征进行分类，以确定所述待识别第三特征中是否含有所述前置字。

17. 一种计算机可读存储介质，其上存储有计算机程序，所述程序被处理器执行时实现如权利要求 1 至 15 中任一项所述的语音识别方法。

- 10 18. 一种电子设备，包括：
处理器；

存储器，用于存储一个或多个程序，当所述一个或多个程序被所述处理器执行时，使得所述处理器实现如权利要求 1 至 15 中任一项所述的语音识别方法。