



(12) 发明专利

(10) 授权公告号 CN 103000176 B

(45) 授权公告日 2014. 12. 10

(21) 申请号 201210585431. X

CN 1979638 A, 2007. 06. 13,

(22) 申请日 2012. 12. 28

WO 2007/053294 A1, 2007. 05. 10,

CN 101295293 A, 2008. 10. 29,

(73) 专利权人 安徽科大讯飞信息科技股份有限公司

审查员 陈斌

地址 230088 安徽省合肥市高新开发区黄山路 616 号

(72) 发明人 何婷婷 胡郁 胡国平 刘庆峰

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 王宝筠

(51) Int. Cl.

G10L 15/26 (2006. 01)

(56) 对比文件

CN 1282072 A, 2001. 01. 31,

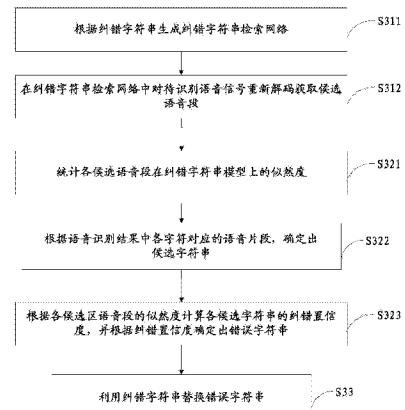
权利要求书2页 说明书11页 附图7页

(54) 发明名称

语音识别方法和系统

(57) 摘要

本发明实施例公开了语音识别方法和系统，以解决上述用户人工进行定位存在的定位不便的问题。该语音识别方法包括：对用户输入的语音信号进行语音识别，获得语音识别结果以及语音识别结果中各字符对应的语音片段；接收用户单独输入的纠错信息并生成相应的纠错字符串；获取纠错字符串对应的至少一个语音段作为候选语音段，并分别统计各候选语音段在纠错字符串模型上的似然度；根据语音识别结果中各字符对应的语音片段，分别确定上述各候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串；根据候选语音段的似然度计算与之对应的候选字符串的纠错置信度，并根据纠错置信度确定出错误字符串；利用纠错字符串替换错误字符串。



1. 一种语音识别方法,其特征在于,包括:

对用户输入的语音信号进行语音识别,获得第一最优解码路径,所述第一最优解码路径包括语音识别结果以及所述语音识别结果中各字符对应的语音片段;

接收用户单独输入的纠错信息并生成相应的纠错字符串,所述纠错信息通过非语音方式或语音方式输入;

获取所述纠错字符串在所述用户输入的语音信号中对应的至少一个语音段作为候选语音段,并分别统计各候选语音段在所述纠错字符串模型上的似然度;

根据所述语音识别结果中各字符对应的语音片段,分别确定上述各候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串;

根据所述候选语音段的似然度计算与之对应的候选字符串的纠错置信度,并根据纠错置信度确定出错误字符串;

利用所述纠错字符串替换所述错误字符串。

2. 如权利要求 1 所述的方法,其特征在于,所述根据所述候选语音段对应的似然度计算与之对应的候选字符串的纠错置信度包括:

在满足预设条件时,按照候选字符串所对应候选语音段的似然度从大到小的顺序,依次将候选字符串设为当前考察对象;

在所述纠错字符串和当前考察对象相同时,将所述当前考察对象的纠错置信度设为第一预设值;

在所述纠错字符串和当前考察对象不相同,并且,当前考察对象为第一个候选字符串时,将所述当前考察对象的纠错置信度设为第二预设值;

在所述纠错字符串和当前考察对象不相同,并且,当前考察对象不为第一个候选字符串时,按照预设算法计算所述当前考察对象对应的纠错置信度。

3. 如权利要求 2 所述的方法,其特征在于,所述预设条件包括存在尚未考察的候选字符串,并且已考察的候选字符串总数不大于预设的最大候选总数。

4. 如权利要求 3 所述的方法,其特征在于,所述按照预设算法计算所述当前考察对象对应的纠错置信度具体包括:

当所述当前考察对象所对应语音段的似然度不大于第一预设阈值  $T$  时,将所述当前考察对象的纠错置信度设为所述第一预设值;

当所述当前考察对象所对应语音段的似然度大于所述  $T$  并且差值  $\Delta S$  大于第二预设阈值时,将所述当前考察对象的纠错置信度设为所述第二预设值;

当所述当前考察对象所对应语音段的似然度大于所述  $T$  并且差值  $\Delta S$  小于第二预设阈值时,将所述当前考察对象的纠错置信度设为所述第一预设值;

所述差值  $\Delta S$  为当前考察对象所对应语音段的似然度与最大似然度之间的差值,所述最大似然度为所有候选语音段所对应的似然度中的最大值。

5. 如权利要求 4 所述的方法,其特征在于,所述根据纠错置信度确定出错误字符串具体包括:

将所有纠错置信度等于所述第二预设值的候选字符串作为错误字符串。

6. 如权利要求 1 所述的方法,其特征在于,所述根据所述纠错字符串确定所述用户输入的语音信号中候选语音段,包括:

根据所述纠错字符串生成纠错字符串检索网络,所述纠错字符串检索网络包括所述纠错字符串对应的纠错字符串模型以及预置的吸收模型,所述吸收模型是由系统预先在海量语音数据训练得到的背景模型;

在所述纠错字符串检索网络内搜索相应于所述用户输入的语音信号的第二最优解码路径,所述第二最优解码路径包括所述纠错字符串模型对应的语音段作为所述候选语音段;

确定所述候选语音段在所述用户输入的语音信号中对应的起始位置和结束位置。

7. 如权利要求 6 所述的方法,其特征在于,所述根据所述纠错字符串生成纠错字符串检索网络,包括:

获取所述纠错字符串对应的纠错字符串模型;

获取预置的吸收模型;

根据获取的纠错字符串模型以及吸收模型生成所述纠错字符串检索网络。

8. 如权利要求 1 至 7 任一项所述的方法,其特征在于,所述利用所述纠错字符串替换所述错误字符串具体包括:

在所述错误字符串的数目等于 1 时,直接利用所述纠错字符串替换所述错误字符串;

在所述错误字符串的数目大于 1 时,利用所述纠错字符串替换用户指定的错误字符串。

9. 如权利要求 8 所述的方法,其特征在于,所述利用所述纠错字符串替换用户指定的错误字符串具体包括:

在所述语音识别结果中突出显示所有错误字符串;

接受用户选择,利用所述纠错字符串更新用户选定的错误字符串。

10. 一种语音识别系统,其特征在于,包括:

语音识别单元,用于对用户输入的语音信号进行语音识别,获得第一最优解码路径,所述第一最优解码路径包括语音识别结果以及所述语音识别结果中各字符对应的语音片段;

纠错字符串生成单元,用于接收用户单独输入的纠错信息并生成相应的纠错字符串,所述纠错信息通过非语音方式或语音方式输入;

自动纠错单元,用于根据所述纠错字符串确定所述用户输入的语音信号中候选语音段;统计所述候选语音段在所述纠错字符串模型上的似然度;根据所述语音识别结果中各字符对应的语音片段,确定上述候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串;根据所述候选语音段对应的似然度计算与之对应的候选字符串的纠错置信度,并根据纠错置信度确定出错误字符串;利用所述纠错字符串替换所述错误字符串。

## 语音识别方法和系统

### 技术领域

[0001] 本发明涉及语音识别技术领域,更具体地说,涉及语音识别方法和系统。

### 背景技术

[0002] 语音识别技术是一种对用户录入的语音信号进行识别,最终转化为文本/字符串(也即识别结果为文本)的技术,其为自然人性的人机交互提供了便利。以采用语音识别技术的移动设备为例,在语音识别技术的支持下,用户只要对着移动设备说话,经过语音识别系统识别后就会自动形成文字,大大提高了用户的输入效率。

[0003] 但是,在大词汇量随意说的应用环境下,语音识别技术依然不能达到百分百正确的识别率,需要人工对识别结果进行修正编辑。移动设备(语音识别系统)将语音识别结果显示到屏幕的文本输入区后,用户如想对语音识别结果进行修改编辑,则首先需要在语音识别结果中定位需要修正(也可称为待修改)的字符。

[0004] 而在移动设备上,特别是小屏幕的指触屏设备上,由于屏幕尺寸受限,用户在从连续大段文本中对某个确定的字符进行定位时,特别是在相邻两字符间插入编辑光标时,存在定位不便的问题。

### 发明内容

[0005] 有鉴于此,本发明实施例目的在于提供语音识别方法和系统,以解决上述用户人工进行定位存在的定位不便的问题。

[0006] 为实现上述目的,本发明实施例提供如下技术方案:

[0007] 根据本发明实施例的一个方面,提供一种语音识别方法,包括:

[0008] 对用户输入的语音信号进行语音识别,获得第一最优解码路径,所述第一最优解码路径包括语音识别结果以及所述语音识别结果中各字符对应的语音片段;

[0009] 接收用户单独输入的纠错信息并生成相应的纠错字符串,所述纠错信息通过非语音方式或语音方式输入;

[0010] 获取所述纠错字符串在所述用户输入的语音信号中对应的至少一个语音段作为候选语音段,并分别统计各候选语音段在所述纠错字符串模型上的似然度;

[0011] 根据所述语音识别结果中各字符对应的语音片段,分别确定上述各候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串;

[0012] 根据所述候选语音段的似然度计算与之对应的候选字符串的纠错置信度,并根据纠错置信度确定出错误字符串;

[0013] 利用所述纠错字符串替换所述错误字符串。

[0014] 根据本发明实施例的另一个方面,提供一种语音识别系统,包括:

[0015] 语音识别单元,用于对用户输入的语音信号进行语音识别,获得第一最优解码路径,所述第一最优解码路径包括语音识别结果以及所述语音识别结果中各字符对应的语音片段;

[0016] 纠错字符串生成单元,用于接收用户单独输入的纠错信息并生成相应的纠错字符串,所述纠错信息通过非语音方式或语音方式输入;

[0017] 自动纠错单元,用于根据所述纠错字符串确定所述用户输入的语音信号中候选语音段;统计所述候选语音段在所述纠错字符串模型上的似然度;根据所述语音识别结果中各字符对应的语音片段,确定上述候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串;根据所述候选语音段对应的似然度计算与之对应的候选字符串的纠错置信度,并根据纠错置信度确定出错误字符串;利用所述纠错字符串替换所述错误字符串。

[0018] 从上述的技术方案可以看出,本发明实施例所公开的技术方案根据用户单独输入的纠错信息所生成的纠错字符串来确定候选语音段,再通过候选语音段找到其在语音识别结果中所对应的错误字符串,实现了用户所输入的纠错字符串与错误字符串的对应,进而实现了对语音识别结果中错误字符串的自动定位,解决了用户人工进行定位存在的定位不便的问题。

### 附图说明

[0019] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,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,还可以根据这些附图获得其他的附图。

[0020] 图 1 为本发明实施例提供的语音识别方法流程图;

[0021] 图 2 为本发明实施例提供的手写输入识别流程图;

[0022] 图 3 为本发明实施例提供字符覆盖的最小区域示意图;

[0023] 图 4 为本发明实施例提供的自动纠错过程流程图;

[0024] 图 5 为本发明实施例提供的纠错字符串检索网络结构示意图;

[0025] 图 6 为本发明实施例提供的自动纠错过程流程图;

[0026] 图 7 为本发明实施例提供的确定出错误字符串流程图;

[0027] 图 8 为本发明实施例提供的纠错置信度计算流程图;

[0028] 图 9 为本发明实施例提供的语音识别系统结构示意图。

### 具体实施方式

[0029] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0030] 作为一种简单方便而又高效的输入方式,语音识别改变了传统的基于复杂编码或拼音输入的键盘模式,为自然人性的人机交互提供了便利条件。特别是近年来随着科技的发展和无线通讯网络的创新普及,各种在线语音识别应用,如发微博,写短信,网络即时通讯等受到了越来越多的关注。在语音识别技术的支持下,用户只要对着移动设备说话,经过系统识别后就会自动形成文字,大大提高了用户的输入效率。

[0031] 但是,在大词汇量随意说的应用环境下,语音识别技术依然不能达到百分百正确的识别率,需要人工对识别结果进行修正编辑。移动设备(语音识别系统)将语音识别结果显示到屏幕的文本输入区后,用户如想对语音识别结果进行修改编辑,则需要在识别结果中定位需要修正(也可称为待修改)的字符。

[0032] 而在移动设备上,特别是小屏幕的指触屏设备上,由于屏幕尺寸受限,用户在从连续大段文本中对某个确定的字符进行定位时,特别是在相邻两字符间插入编辑光标时,存在定位不准的问题。

[0033] 为便于理解,现对语音识别介绍如下:

[0034] 如果将待识别的一段语音信号记作  $S$ ,对  $S$  进行一系列处理后得到与之相对应的语音特征序列  $O$ ,记作  $O = \{O_1, O_2, \dots, O_i, \dots, O_T\}$ ,其中  $O_i$  是第  $i$  个语音特征, $T$  为语音特征总个数。语音信号  $S$  对应的句子可看作是由许多词组成的一个词串,记作  $W = \{w_1, w_2, \dots, w_n\}$ 。语音识别的任务就是根据已知的语音特征序列  $O$ ,求出最可能的词串  $W'$ 。

[0035] 在语音识别的具体过程中,一般首先提取语音信号对应的语音特征参数,随后在由预置的声学模型和语言模型构成的网络搜索空间中,根据预设的搜索算法(比如 Viterbi 算法),搜索相对于所提取得语音特征参数的最优路径(也即最优解码路径)。

[0036] 在了解了语音识别的一些概念后,现对本发明实施例的技术方案介绍如下。

[0037] 为解决上述定位不便的问题,本发明实施例所提供的语音识别方法至少包括如下步骤:

[0038] 语音识别过程:对用户输入的语音信号进行语音识别,获得最优解码路径,其中,最优解码路径包括语音识别结果以及语音识别结果中各字符对应的语音片段;

[0039] 纠错字符串生成过程:接收用户单独输入的纠错信息并生成相应的纠错字符串,上述纠错信息允许通过非语音方式或语音方式输入;

[0040] 自动纠错过程:根据纠错字符串确定用户输入的语音信号中候选语音段;根据语音识别结果中各字符对应的语音片段,确定上述候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串,从候选字符串中确定出错误字符串;并利用纠错字符串替换所述错误字符串。下面对各个过程进行一一介绍。

[0041] 一,语音识别过程

[0042] 为了最大可能的满足用户日常交互需求,本发明实施例采用大词汇量连续语音识别技术,以实现任意说语音的文本转换。

[0043] 其中,参见图 1,上述语音识别过程具体包括:

[0044] S11、跟踪采集用户输入的语音信号(也即上述待识别的一段语音信号);

[0045] 在本发明其他实施例中,可将上述语音信号存入数据缓存区;

[0046] S12、对上述语音信号进行预处理,以得到经过预处理后的语音数据;

[0047] 上述预处理可包括语音信号采样、反混叠带通滤波、分帧处理、去除个体发音差异和设备、环境引起的噪声影响,端点检测。为了提高语音识别系统的鲁棒性上述预处理具体还可包括前端降噪处理,以为后续语音处理提供较为纯净的语音。

[0048] S13、对上述经过预处理的语音数据中每帧语音数据分别进行特征提取,以获取特征矢量序列。

[0049] 在步骤 S13 中,对每帧语音数据进行特征提取后可提取出有效语音特征(或者特

征矢量)。这样,经特征提取后,每一帧语音数据形成一个特征矢量,相应的,上述语音数据即可用一特征矢量序列表示;

[0050] 本领域技术人员可以理解的是,如果对经过预处理后的语音数据包括 30 帧语音数据的话,那么这 30 帧语音数据就可提取出 30 个特征矢量,而这 30 个特征矢量按照时间先后顺序即可组成上述特征矢量序列。

[0051] 在本发明其他实施例中,上述有效语音特征可为线性预测倒谱或 MFCC (Mel 倒谱)特征。具体的,以 MFCC 特征为例,可对窗长 25ms 帧移 10ms 的每帧语音数据通过短时分析得到 MFCC 参数和 / 或 MFCC 参数的一阶 / 二阶差分,共计 39 维。这样,每帧语音数据经过特征提取可得到一个 39 维的特征矢量。

[0052] 在本发明其他实施例中,可将上述语音特征 / 语音特征矢量序列存入特征缓存区内。

[0053] S14、在预先构建的检索网络中对上述特征矢量序列进行最优路径搜索(上述检索网络主要由系统预设的声学模型,词典,语言模型等构成),以获取与上述特征矢量序列具有最大模型似然概率的模型串作为语音识别结果输出。

[0054] 在具体实施时,可采用业内主流的基于动态规划思想的 Viterbi 搜索算法,对每一特征矢量遍历检索网络中满足预设条件的活跃节点计算累计历史路径概率并保留满足预设条件的历史路径作为后续搜索网络的活跃节点,最后通过对具有最大历史路径概率的路径(也即上述第一最优解码路径)回溯实现对输入语音的识别解码。在解码中第一最优解码路径对每帧语音数据均保留其所对应的识别单元模型,进而对语音识别结果中每一字符都可获取其所对应的语音片段,当然,也可获取每一字符所对应语音片段的起始位置信息和结束位置信息。

[0055] 需要说明的是,上述语音片段既可为语音信号中的语音片段,也可为经过预处理后的语音数据中的至少一帧语音数据,还可为特征矢量序列中的特征矢量子序列。为称呼方便,本文后续将语音信号、经过预处理后的语音数据以及特征矢量序列统称为待识别语音信号。也即,本文下述提及的待识别语音信号具体可为用户输入的语音信号、经过预处理后的语音数据或特征矢量序列。而本文下述提及的语音片段具体可为用户输入的语音信号中的语音片段、至少一帧语音数据或特征矢量子序列。

[0056] 也就是说,我们可将步骤 S11 中的语音信号或者步骤 S12 中经过预处理后的语音数据或者步骤 S13 中的特征矢量序列划分成与语音识别结果中的字符相对应的语音片段,从而令语音识别结果中的每个字符对应一个确切的语音片段。

[0057] 举例来讲,如果语音识别结果为“我们去爬山”这一字符串,该字符串对应的解码路径信息可保存为:(0000000 2200000),(2200000 3600000),(3600000 4300000),(4300000 5000000),(5000000 7400000)。

[0058] 其中,(0000000 2200000)指示了“我”这个字符所对应的语音片断的起始位置信息和结束位置信息。其中,0000000 是“我”所对应的语音片断在待识别语音信号中的起始位置(时刻),而 2200000 是“我”所对应的语音片断在待识别语音信号中的结束位置(时刻)。

[0059] 二、纠错字符串生成过程

[0060] 本发明实施例支持用户以非语音方式或语音方式输入纠错信息并生成纠错字符串。

[0061] 在采用语音方式输入纠错信息时,所输入的纠错信息具体为语音信号,由于与语音识别过程一样是以语音方式输入,则系统可能无法确定当前的语音输入是为了继续新文本的语音输入,还是为了对原始文本进行语音纠错输入。因此,可以设置单独的纠错信息输入控制按键,控制从新文本的语音输入切换至对原始文本的语音纠错输入。在语音方式输入纠错信息的模式下,由于纠错信息为语音信号,在将其转化成纠错字符串时的处理过程同上述语音识别过程相同,在此不作赘述,并且,还可提供多个识别候选字符串供用户选择以提高生成纠错字符串的准确率。

[0062] 此外,本发明实施例还支持用户以按键输入(比如拼音输入、笔划输入、区位码输入等等)、手写输入等非语音方式输入纠错信息,此时,如以按键输入,所输入的纠错信息具体为按键序列,如以手写输入,所输入的纠错信息具体为书写笔迹。

[0063] 现以拼音输入和手写输入为例,对非语音方式输入过程进行介绍。

[0064] 其具体流程仍请参见图 1:

[0065] S21、判断用户的输入方式,如是拼音按键输入转入步骤 S22,如果是手写输入转入步骤 S23。

[0066] S22、将用户输入的按键序列转换成候选纠错字符串。

[0067] 其中,步骤 S22 具体可包括:

[0068] S221,跟踪采集用户的按键序列,将其对应成字母串序列;

[0069] S222,将采集到的字母串序列和预置的拼音词典匹配以找到候选纠错字符串,并显示。

[0070] 比如用户在输入 qinghua 后,系统可能显示清华、青花、亲华等多个候选纠错字符串供用户选择。

[0071] S23、识别用户输入的书写笔迹,将用户输入的书写笔迹转化为至少一个候选纠错字符串;

[0072] 其中,参见图 2,步骤 S23 可具体包括:

[0073] S231,跟踪用户输入的书写笔迹,并将采集到的书写笔迹保存在笔迹数据缓存区内;

[0074] 在联机手写识别系统中,用户的书写笔迹通常用一序列的二维(位置坐标)或三维点(位置坐标和抬笔/落笔状态)坐标表示,用以描述字符书写的空间和时间信息。

[0075] S232,对上述书写笔迹进行预处理。

[0076] 由于采集设备或用户在书写时抖动等原因,原始采集到的书写笔迹中可能存在各种噪音干扰。为了提高系统的鲁棒性,可对采集到的笔迹进行预处理。具体的,可通过字符大小归一化、野点去除、平滑,重采样等处理方式加以组合,以尽可能减少噪音干扰带来的识别率下降的问题。

[0077] S233,对经过预处理的书写笔迹进行笔迹特征提取。

[0078] 和语音识别相类似,在手写识别中,也需要从原始的笔迹轨迹上提取反映字符特点的字符特征。

[0079] 具体的,本实施例提取手写识别领域常用的八方向特征,并通过 LDA 等技术提高笔迹特征的可区分性。

[0080] S234,将提取的字符特征与预置模型进行匹配,计算相似度。



[0081] S235、选取与上述字符特征具有最高相似度的至少一个预置模型作为候选纠错字符串,并显示。

[0082] 考虑到拼音输入以及手写识别技术的准确率往往很好,因而通常上述候选纠错字符串的个数可以选择 3 到 5。

[0083] 当然,本领域技术人员可以理解的是,当用户的非语音输入足够长时,也可能只有一个候选纠错字符串。

[0084] S25、从候选纠错字符串中确定纠错字符串。

[0085] 步骤 S25 可具体包括:

[0086] 接受用户的选择指定,从至少一个候选纠错字符串中确定唯一的纠错字符串。

[0087] S25 可以单独列出,作为对纠错字符串的进一步确认,以兼容语音输入和非语音输入方式。

[0088] 三,自动纠错过程

[0089] 考虑到用户输入的纠错字符串和语音识别结果中错误字符串所对应的语音段往往具有一致性,本发明实施例自动纠错的核心思想是:将用户单独输入的纠错信息所生成的纠错字符串映射到语音段上,再通过该语音段找到其在用户输入的语音信号对应的语音识别结果中所对应的字词(也即错误字符串),从而实现了纠错字符串与错误字符串的对应。这样,就实现了对语音识别结果中错误字符串的自动定位,解决了用户人工进行定位存在的定位不便的问题。

[0090] 具体来说,该方法首先在待识别语音信号中找到对应于上述纠错字符串的语音段。随后在语音识别结果中定位与该语音段相对应的字符串作为“错误字符串”。上述“错误字符串”是在步骤 S14 中得到的模型串中的子串,该子串在待识别语音信号中所对应的语音段的起始时刻和结束时刻,与上述纠错字符串在待识别语音信号中所对应的语音段的起始时刻和结束时刻具有一致性。

[0091] 自动纠错过程的流程请仍参见图 1,包括:

[0092] S31、根据纠错字符串确定待识别语音信号中候选语音段;

[0093] S32、根据语音识别结果中各字符对应的语音片段,确定上述候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串,从候选字符串中确定出错误字符串;

[0094] S33、利用纠错字符串替换上述错误字符串。

[0095] 在本发明其他实施例中,步骤 S33 可包括如下步骤:

[0096] 在错误字符串的数目 $\hat{k}$ 等于 1 时,直接利用纠错字符串替换该错误字符串;

[0097] 在错误字符串的数目 $\hat{k}$ 大于 1 时,利用纠错字符串替换用户指定的错误字符串。

[0098] 本发明一些实施例可接受用户主动参与选择,因此,上述“利用纠错字符串替换用户指定的错误字符串”的具体流程可包括:

[0099] a,在语音识别结果中突出显示所有错误字符串。

[0100] 在本发明其他实施例中,除突出显示所有错误字符串外,还可以设置除错误字符串外的其它识别结果为非活跃状态,以提高定位精确度;

[0101] b,接受用户的选择指定,利用上述纠错字符串更新用户所选定的错误字符串。

[0102] 此外,在本发明其他实施例中,还可支持用户的模糊选择指定——即并不要求用

户精准定位错误字符串,而是通过近邻方式进行定位:当手写笔的落笔点落入错误字符串近邻区域时,自动将其定位到对应的错误字符串上。

[0103] 具体来说,计算落笔点距每个错误字符串所覆盖的最小区域的最短距离,选择具有最小“最短距离”的错误字符串作为用户选定的错误字符串。例如,参见图 3,可设定一个字符(我)所覆盖的最小区域的高度  $H$  为该字符字高  $h$  的  $A$  倍,而一个字符所覆盖的最小区域的宽度  $W$  为该字符字宽  $w$  的  $B$  倍, $A$  和  $B$  可为大于等于 1 的任意正数。那么,错误字符串所覆盖的最小区域则为组成该错误字符串中所有字符所覆盖的最小区域的总和。

[0104] 参见图 4,在本发明其他实施例中,上述步骤 S31 可具体包括如下步骤:

[0105] S311,根据上述纠错字符串生成纠错字符串检索网络。

[0106] 请参见图 5,上述纠错字符串检索网络包括纠错字符串模型以及预置的吸收模型。

[0107] 其中,纠错字符串模型由纠错字符串生成:通过预置的词典将纠错字符串扩展为相应的模型序列得到对应的纠错字符串模型。由于用户每次输入的纠错信息并生成的纠错字符串都不尽相同,因此,纠错字符串网络中的纠错字符串模型需要实时更新。

[0108] 因此,上述步骤 S31 又可具体包括:

[0109] 获取纠错字符串对应的纠错字符串模型;

[0110] 获取预置的吸收模型;

[0111] 根据获取的纠错字符串模型以及吸收模型生成纠错字符串检索网络。

[0112] 需要说明的是,如果语音识别结果中存在不相邻且不相干的多处识别错误,比如语音识别结果中存在“清华”和“西站”两处识别错误,则需要多次通过语音或非语音方式输入纠错信息生成纠错字符串。而对每次输入的纠错字符串,不管其包含多少字词,都将其看作一个独立的纠错字符串。比如,用户在某次输入纠错字符串时,共输入了 3 个汉字,则纠错字符串包括 3 个汉字,随后通过字典将包括该 3 个汉字的纠错字符串扩展成对应的纠错字符串模型。

[0113] 在将纠错字符串扩展成纠错字符串模型时,根据预置的声学模型的不同可采用不同的扩展方式。比如,可基于音节模型单元的声学模型(如基于音节模型单元的声学模型,单个汉字由 1 个音节构成),也可基于音素模型单元的声学模型(如基于音素模型单元的声学模型,单个汉字由 2 个音素构成),具体由在进行语音识别时所采用的模型单元所决定。因此,如对上述包括 3 个汉字的纠错字符串进行扩展,可扩展得到由 3 个音节模型单元相串联的纠错字符串模型或者由 6 个音素模型单元相串联的纠错字符串模型。

[0114] 至于吸收模型则是由系统预先在海量语音数据训练得到的背景模型,也可采用多个吸收模型以提高复杂语音匹配的准确性。需要注意的是,多个单独吸收模型是并联的。

[0115] S312,在纠错字符串检索网络中对待识别语音信号重新解码获取第二最优解码路径。

[0116] 其中,第二最优解码路径包括纠错字符串模型对应的语音段,纠错字符串模型对应的语音段即为候选语音段。

[0117] 具体的,上述纠错字符串模型所对应的语音段可为用户输入的语音信号中的语音段,也可为经过预处理后的语音数据中的语音段,还可为特征矢量序列中的特征矢量子序列。为简单化起见,可选用纠错字符串模型所对应的特征矢量子序列作为候选语音段。则步骤 S312 可具体包括:

[0118] 在纠错字符串检索网络中搜索相应于特征矢量序列的最优路径(即第二最优路径),得到上述纠错字符串模型所对应的特征矢量子序列在整个特征矢量序列中的起始位置和结束位置。

[0119] 步骤 S312 中的解码,与上述步骤 S14 相类似,二者的不同在于,步骤 S312 所利用的网络是根据纠错字符串生成的纠错字符串检索网络,而步骤 S14 所利用的检索网络的范围要大于上述纠错字符串检索网络。因此,步骤 S312 的解码,仍可采用业内主流的基于动态规划思想的 Viterbi 搜索算法,对每帧特征矢量遍历纠错字符串检索网络中满足预设条件的活跃节点并保留满足预设条件的历史路径作为后续搜索网络的活跃节点,最后通过对具有最大历史路径概率的路径(即第二最优解码路径)获得纠错字符串模型对应的语音段,从而确定了候选语音段。

[0120] 考虑到,纠错字符串模型可能对应多个语音段,而由于种种原因,实际上,这多个语音段中可能有一些并未发生识别错误,因此,需要对这多个语音段对应的字符串进行甄选,也即,需要对上述候选字符串进行甄选。

[0121] 基于此,参见图 6,步骤 S32 可具体包括:

[0122] S321、将纠错字符串模型 M 对应的候选语音段的集合,记为  $\{X_1, X_2, \dots, X_i, \dots, X_k\}$  或  $\{X_i\}$ ,并分别统计各候选语音段在纠错字符串模型 M 上的似然度,记为  $p(X_i|M)$ 。其中  $(i=1, 2, \dots, k)$ ,  $k$  表示候选语音段的总数量。

[0123] 所述纠错字符串模型 M 可以通过拼接系统预设的基本语音单元,如音节单元得到,在获取了候选语音段  $X_i$  后,直接就可以计算  $X_i$  相应于纠错字符串模型的似然度。

[0124] S322、根据语音识别结果中各字符对应的语音片段,确定上述候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串。

[0125] 将候选语音段对应的候选字符串的集合,记为  $\{H_1, H_2, \dots, H_i, \dots, H_k\}$  或  $\{H_i\}$ ,则可知,  $X_i$  与  $H_i$  对应。

[0126] 由于在步骤 S312 中,已经获取了候选语音段的起始位置(时刻)和结束位置(时刻),因此,可根据语音识别结果中各字符对应的语音片段,确定候选语音段的起始位置在语音识别结果中所对应的起始字符。同理,可确定候选语音段的结束位置在语音识别结果中所对应的结束字符,在确定了起始字符和结束字符后,就可确定出候选语音段在语音识别结果中所对应的字符串了。

[0127] 更具体的,可通过如下方式来确定起始字符:

[0128] 将起始位置对应的字符作为第一字符,并将该第一字符所对应的语音片断作为第一语音片断;

[0129] 若上述起始位置位于第一语音片断的前部,则将该第一字符作为起始字符,否则选择语音识别结果中的下一字符作为起始字符。

[0130] 而在确定结束字符时,可通过如下方式:

[0131] 将结束位置对应的字符作为第二字符,将第二字符所对应的语音片断作为第二语音片断;

[0132] 若结束位置位于第二语音片断的前部时,选择语音识别结果中的上一字符作为结束字符,否则,将第二字符作为结束字符。

[0133] 仍以前述的“我们去爬山”这一语音识别结果为例,前已述及,该语音识别结

果中各个字符所对应的语音片段的起始位置和结束位置分别为：(0000000 2200000)，(2200000 3600000)，(3600000 4300000)，(4300000 5000000)，(5000000 7400000)。

[0134] 假设，候选语音段  $X_i$  的起始和终点位置为(0000050 3600000)，由于起始位置 0000050 在(0000000 2200000)的前部，可确定“我”作为起始字符，而结束位置 3600000 在(2200000 3600000)的后部，可确定“们”为结束字符，由此可确定， $X_i$  与“我们”这个字符串相对应。因此，将“我们”作为与候选语音段  $X_i$  对应的候选字符串  $H_i$ 。

[0135] S323、根据各个候选语音段的似然度计算各候选字符串的纠错置信度，分别记为  $\{CM_1, CM_2, \dots, CM_i, \dots, CM_k\}$  或  $\{CM_i\}$ ，并根据纠错置信度确定出错误字符串。

[0136] 更具体的，参见图 7，步骤 S323 包括：

[0137] S3231，初始化，设置当前统计错误字符串总数  $\hat{k}=0$ ；(也可不设置步骤 S3231)

[0138] S3232，按照似然度从大到小的顺序依次将候选字符串设为当前考察对象；

[0139] 换句话说，在多个尚未考察的候选语音段中选择具有最大似然度的候选语音段所对应的字符串作为当前考察对象。

[0140] S3233，判断纠错字符串和当前考察对象是否相同，若相同则转入 S3238，否则转入 S3234；

[0141] S3234，判断当前考察对象是否为第一个候选字符串，若是转入 S3239，否则转入 S3235；

[0142] S3235，计算当前考察对象的纠错置信度；

[0143] S3236，判断是否满足预设条件，若是转入 S3232，否则转入 S3237；

[0144] 上述预设条件包括存在尚未考察的候选字符串，并且已考察的候选字符串总数不大于预设的最大候选总数；

[0145] S3237，将所有纠错置信度等于第二预设值的候选字符串作为错误字符串，并将错误字符串的数量赋值给  $\hat{k}$ 。

[0146] S3238，将纠错置信度设为第一预设值转步骤 S3236；

[0147] S3239，将纠错置信度设为第二预设值转步骤 S3236。

[0148] 在本发明其他实施例中，参见图 8，步骤 S3235 还可包括如下步骤：

[0149] A，判断当前考察对象所对应候选语音段的似然度是否大于第一预设阈值 T，若是(则说明当前考察对象和纠错字符串的匹配程度较高)，转入步骤 B，否则转入步骤 E；

[0150] 需要说明的是，T 由系统在开发集上得到，在取 log 域的似然度计算时，T 往往被设置成是一个负数，如 -50。

[0151] B，计算当前考察对象所对应候选语音段的似然度与语音段集合中似然度最大值之间的差值  $\Delta S$ ；

[0152] C，判断是否满足  $\Delta S > \Delta T$ ，如果是，转步骤 D，否则转步骤 E；其中  $\Delta T$  是第二预设阈值。 $\Delta T$  用于衡量可能的候选和不可能的候选的差异；

[0153] 同样在取 log 域的似然度计算时， $\Delta T$  也可设置为一个负值，如 -200 等。

[0154] D，设置当前考察对象的纠错置信度为第二预设值，比如可将第二预设值设定为 1；

[0155] E，设置当前考察对象的纠错置信度为第一预设值，比如可将第一预设值设定为 0。

[0156] 与上述方法相对应,本发明实施例还提供语音识别系统。图 9 示出了上述系统的一种结构,包括:

[0157] 语音识别单元 1,用于对用户输入的语音信号进行语音识别,获得第一最优解码路径,其中,第一最优解码路径包括语音识别结果以及所述语音识别结果中各字符对应的语音片段;

[0158] 更具体的,语音识别单元可包括处理器,由处理器对用户输入的语音信号进行语音识别。

[0159] 纠错字符串生成单元 2,用于接收用户单独输入的纠错信息并生成相应的纠错字符串;

[0160] 更具体的,如以语音方式输入纠错信息,则纠错字符串生成单元仍可包括上述处理器,由处理器对纠错信息进行语音识别生成纠错字符串;

[0161] 如以按键输入方式输入纠错信息,则纠错字符串生成单元至少可包括键盘和处理器,由处理器对将用户输入的按键序列转换成候选纠错字符串,并接受用户的选择指定,从至少一个候选纠错字符串中确定唯一的纠错字符串。当然也可由另一独立的芯片或处理器来将用户输入的按键序列转换成候选纠错字符串,并接受用户的选择指定,从至少一个候选纠错字符串中确定唯一的纠错字符串。

[0162] 如以手写输入方式输入纠错信息,则纠错字符串生成单元至少可包括手写笔、触摸屏和处理器,由处理器对将用户输入的书写笔迹转换成候选纠错字符串,并接受用户的选择指定,从至少一个候选纠错字符串中确定唯一的纠错字符串。当然也可由另一独立的芯片或处理器来将用户输入的书写笔迹转换成候选纠错字符串,并接受用户的选择指定,从至少一个候选纠错字符串中确定唯一的纠错字符串。

[0163] 当然,为了保证用户可采用多种方式输入纠错信息,纠错字符串生成单元也可同时包括上述多种器件。

[0164] 自动纠错单元 3,用于获取纠错字符串在用户输入的语音信号(或待识别语音信号)中的候选语音段;统计候选语音段在纠错字符串模型上的似然度;根据语音识别结果中各字符对应的语音片段,确定上述候选语音段在第一最优解码路径的语音识别结果中所对应的字符串作为候选字符串;根据候选语音段对应的似然度计算与之对应的候选字符串的纠错置信度,并根据纠错置信度确定出错误字符串;利用纠错字符串替换错误字符串。

[0165] 更具体的,自动纠错单元 3 的功能也可通过上述处理器或其他独立的芯片或处理器实现。

[0166] 上述各单元的更详尽的功能可参见前述方法记载,在此不作赘述。

[0167] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本发明的范围。

[0168] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0169] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0170] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0171] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0172] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U 盘、移动硬盘、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0173] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

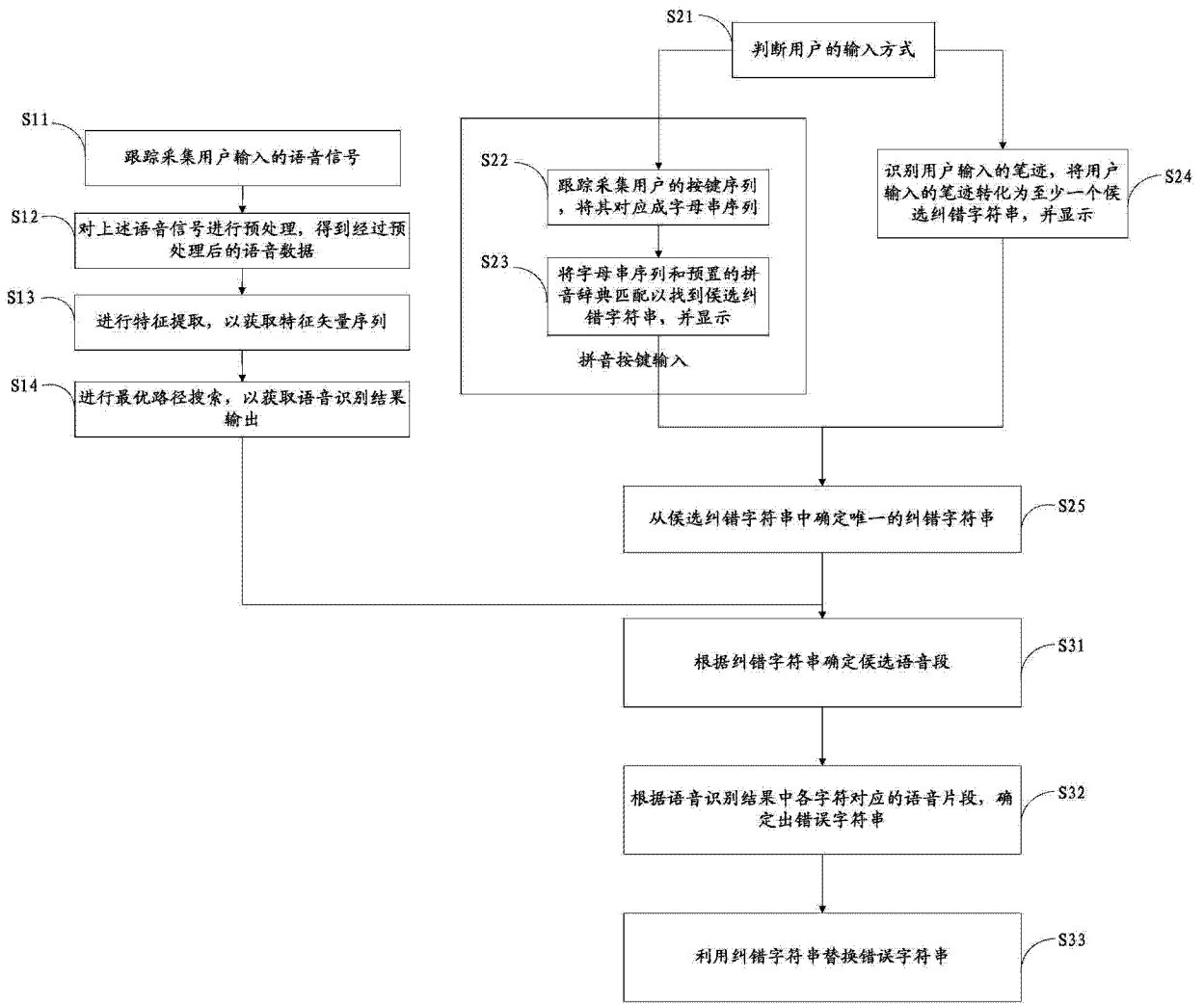


图 1

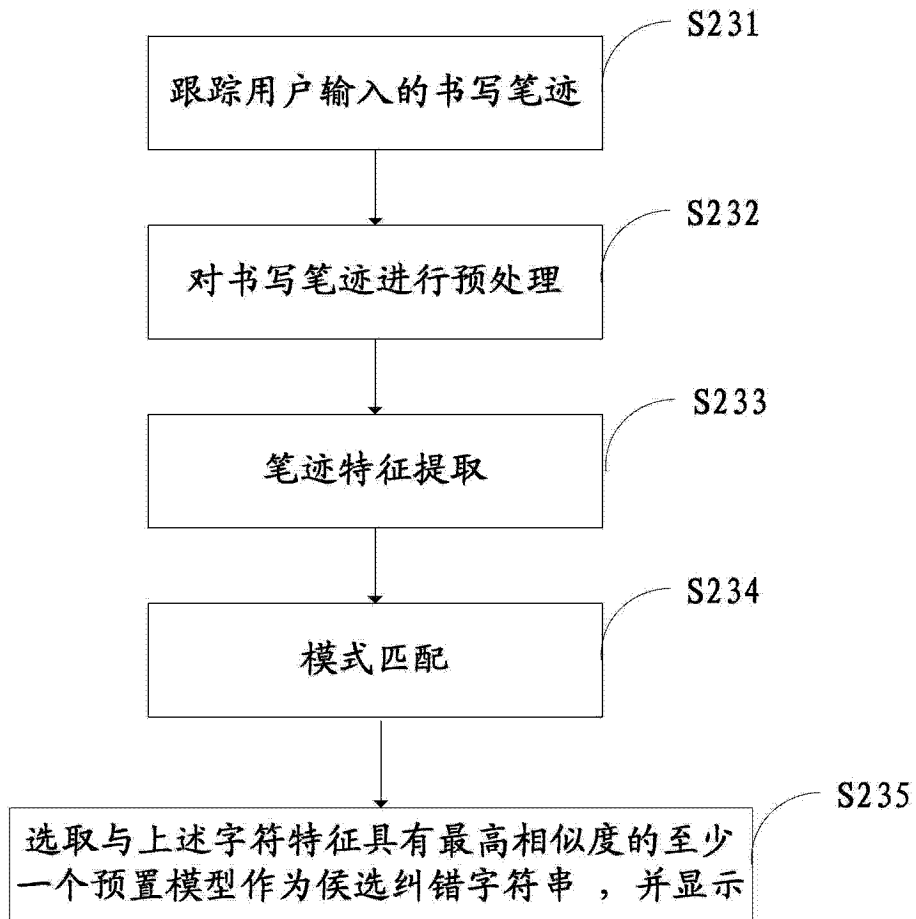


图 2

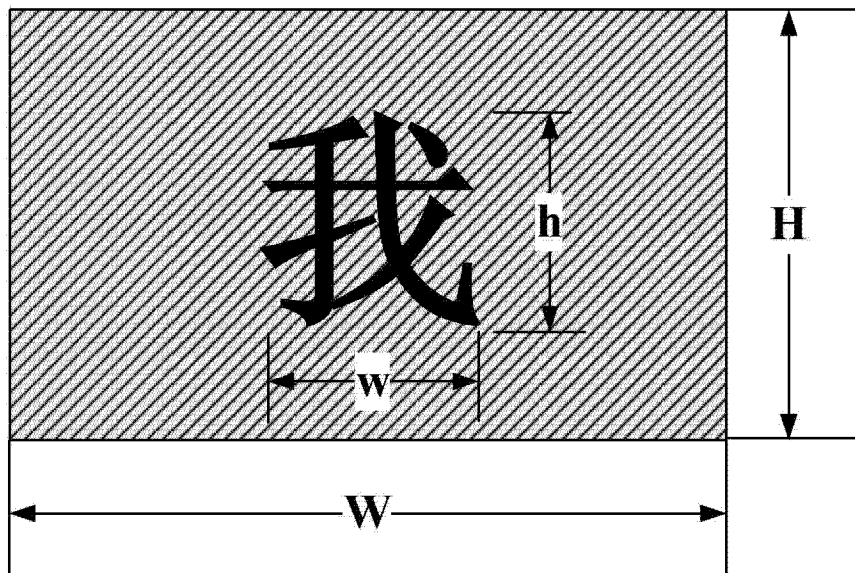


图 3



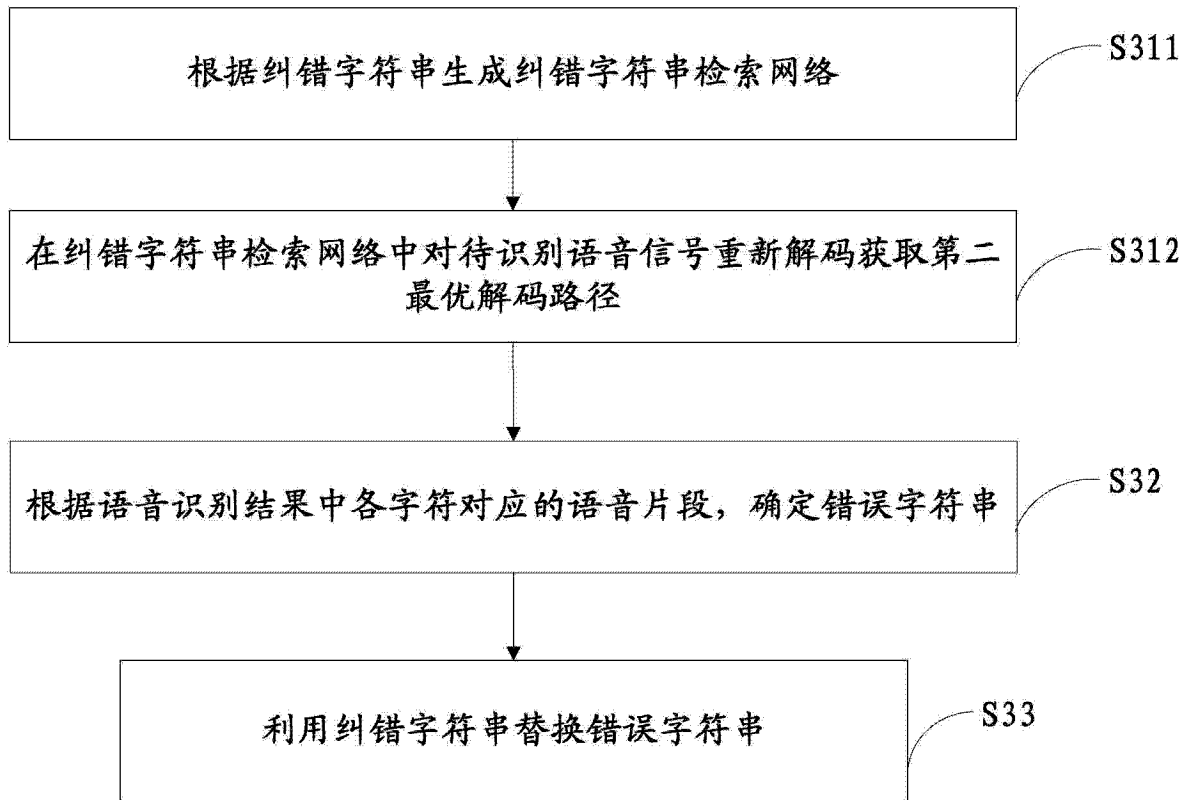


图 4

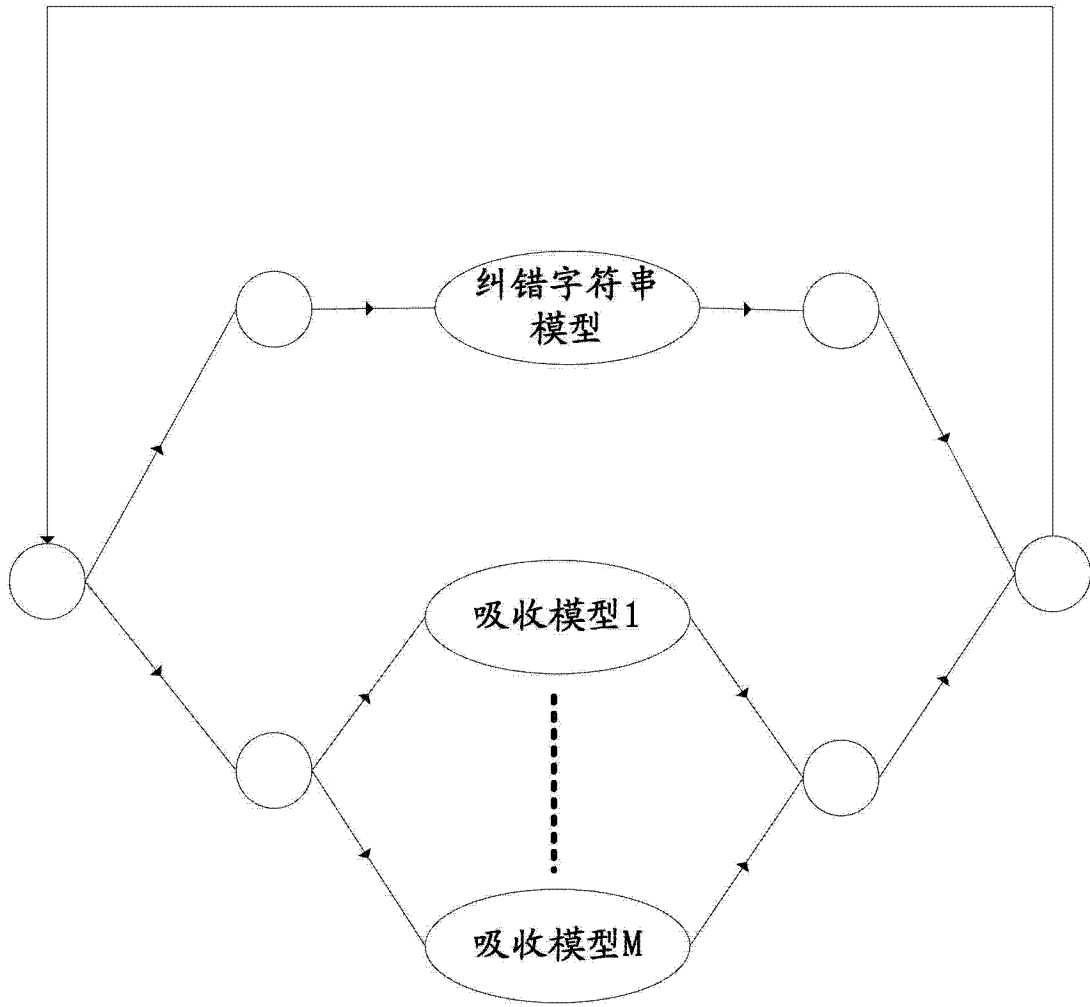


图 5

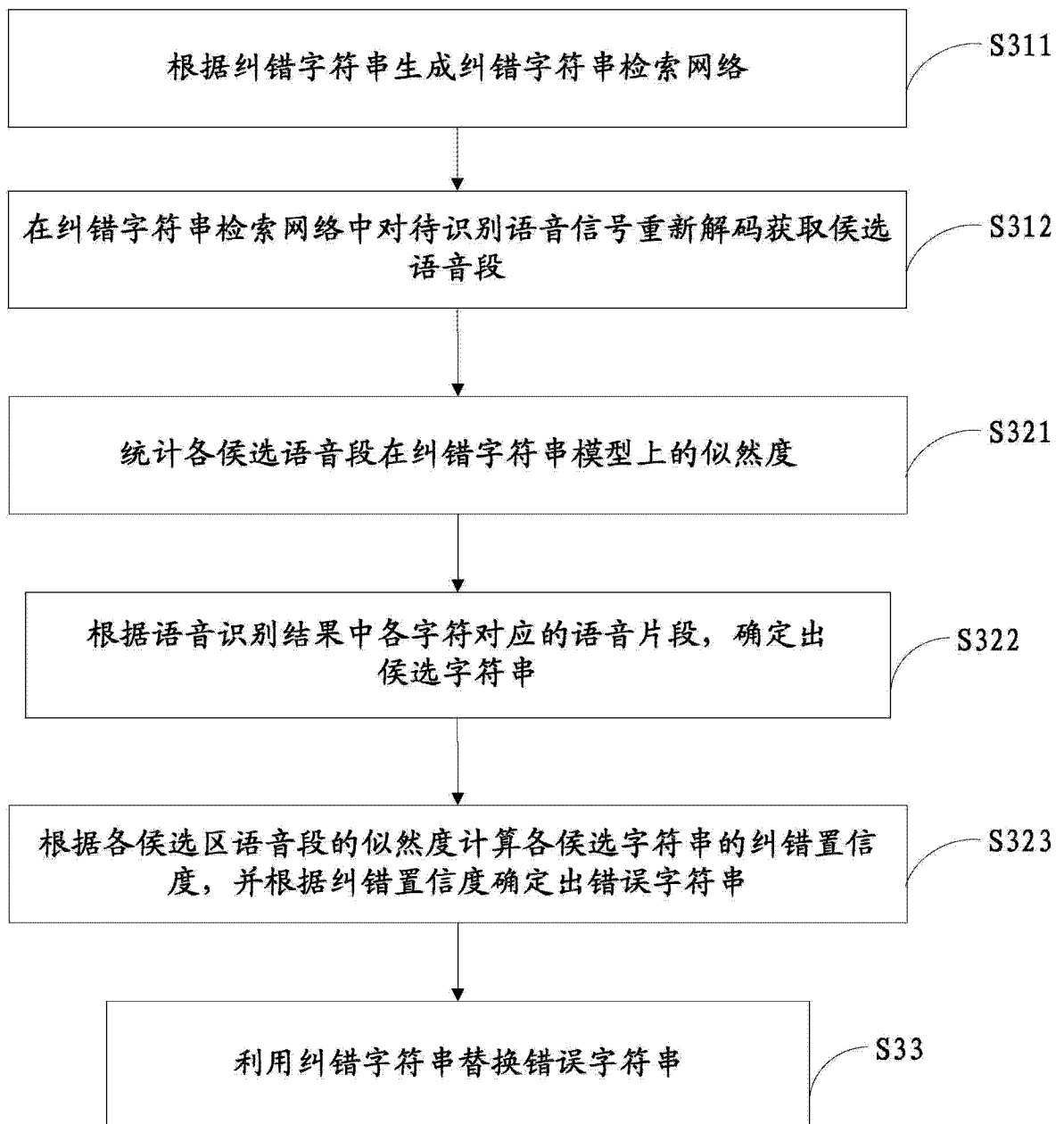


图 6

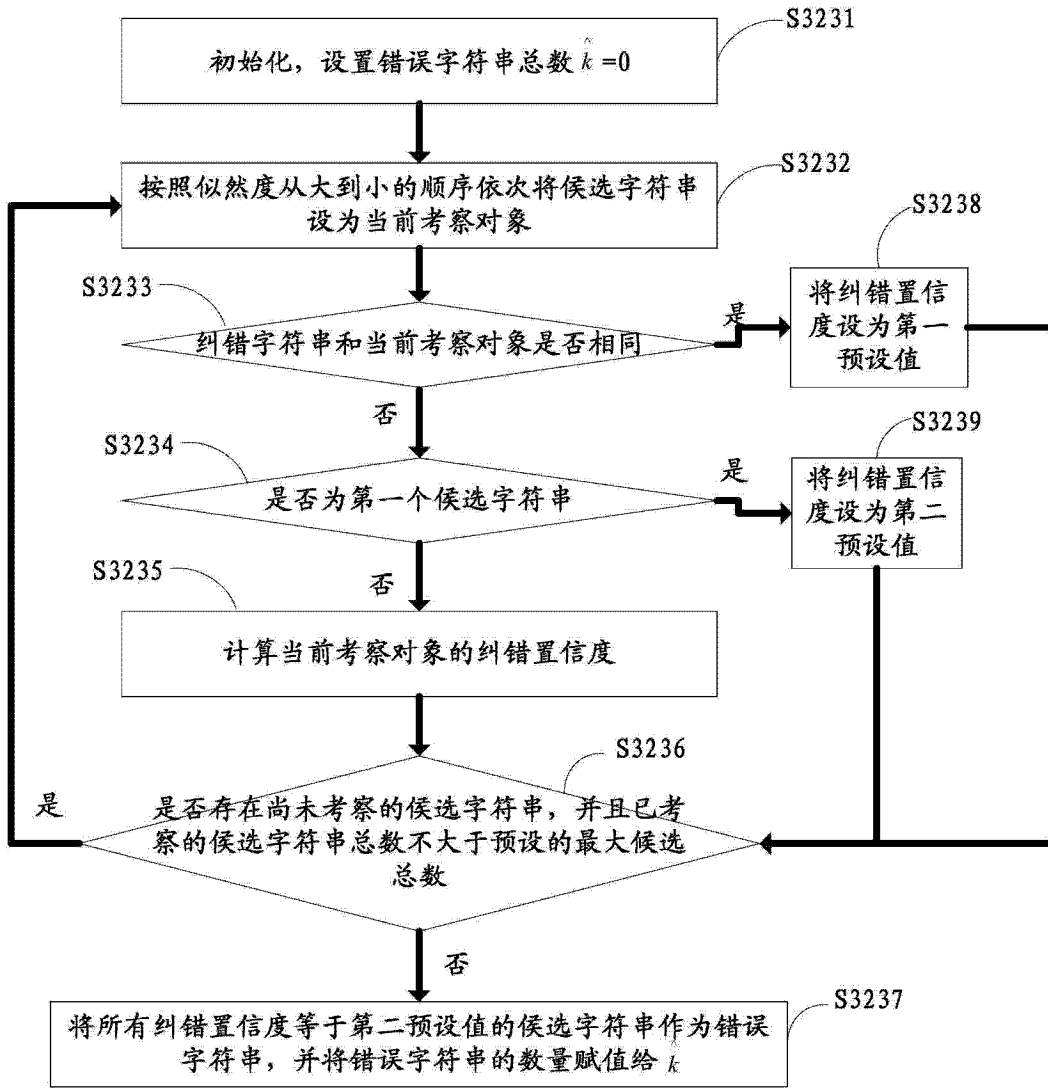


图 7

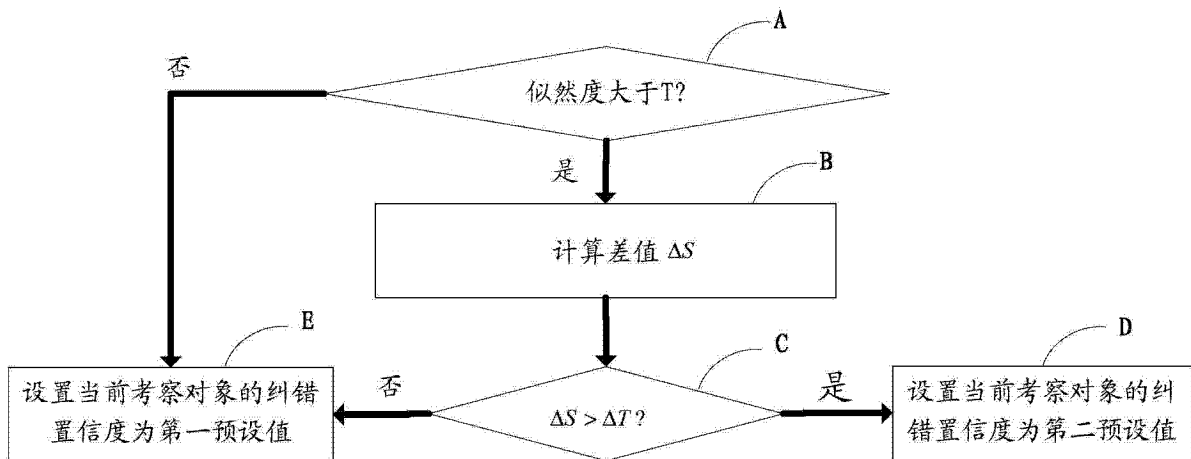


图 8

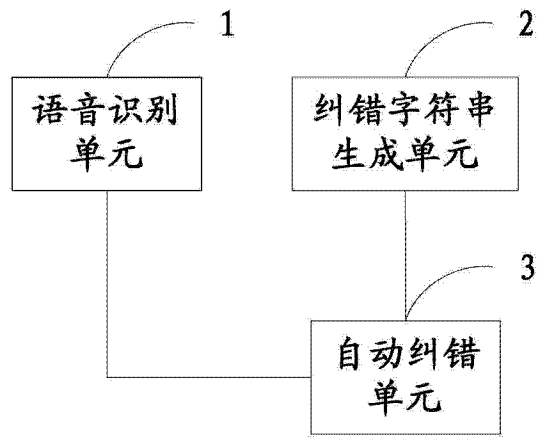


图 9