(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113012683 A
(43)申请公布日 2021.06.22

(21)申请号 202110150327.7

(22)申请日 2021.02.02

(71)申请人 虫洞创新平台(深圳)有限公司
地址 518000 广东省深圳市光明区凤凰街道观光路3009号招商局光明科技园A6栋2C单元

(72)发明人 陈文明 冯兵兵 邓高锋 张世明

(74)专利代理机构 深圳市恒程创新知识产权代理有限公司 44542
代理人 张小容

(51)Int.Cl.
G10L 15/02(2006.01)
G10L 15/06(2013.01)
G10L 15/22(2006.01)
G10L 15/26(2006.01)

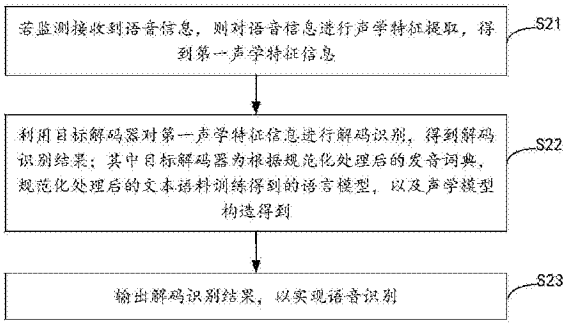
权利要求书2页 说明书10页 附图5页

(54)发明名称

语音识别方法及装置、设备、计算机可读存储介质

(57)摘要

本发明涉及语音识别技术领域，公开了一种语音识别方法及装置、设备、计算机可读存储介质。本发明通过若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息，进而利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果，其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到，再输出解码识别结果，以实现语音识别；解决了相关技术中语音识别的准确率差的问题。



CN 113012683 A

1.一种语音识别方法,其特征在于,所述语音识别方法包括以下步骤:

若监测接收到语音信息,则对所述语音信息进行声学特征提取,得到第一声学特征信息;

利用目标解码器对所述第一声学特征信息进行解码识别,得到解码识别结果;其中,所述目标解码器为根据规范化处理后的发音词典,规范化处理后的文本语料训练得到的语言模型,以及声学模型构造得到;

输出所述解码识别结果,以实现语音识别。

2.如权利要求1所述的语音识别方法,其特征在于,所述利用目标解码器对所述第一声学特征信息进行解码,得到解码识别结果的步骤之前,所述语音识别方法还包括:

对发音词典进行规范化处理,得到规范化发音词典;

对文本语料进行规范化处理,得到规范化文本语料,并对所述规范化文本语料进行训练,得到语言模型;

根据所述规范化处理后的发音词典,所述语言模型,以及声学模型构造解码器,以得到目标解码器。

3.如权利要求2所述的语音识别方法,其特征在于,所述根据所述规范化处理后的发音词典,所述语言模型,以及声学模型构造解码器,以得到目标解码器的步骤之前,还包括:

对语音语料进行声学特征提取,得到第二声学特征信息;

对所述第二声学特征信息进行训练,得到声学模型。

4.如权利要求2所述的语音识别方法,其特征在于,所述根据所述规范化处理后的发音词典,所述语言模型,以及声学模型构造解码器,以得到目标解码器的步骤之前,还包括:

获取声学模型和所述语言模型;

根据所述声学模型中的音素和所述语言模型中的中文词,建立音素与中文词的映射关系,以及根据所述声学模型中的音素和所述语言模型中的单词,建立音素与单词的映射关系;

根据所述音素与中文词的映射关系以及音素与单词的映射关系,得到发音词典。

5.如权利要求2-4中任一项所述的语音识别方法,其特征在于,所述对发音词典进行规范化处理,得到规范化发音词典的步骤,包括:

对所述发音词典进行训练得到词转音素模型;

根据所述词转音素模型,生成补充发音词典;其中,所述补充发音词典中包含的音素不在所述发音词典中,所述补充发音词典中包含的音素对应的中文词和音素对应的单词在所述语言模型中;

根据所述补充发音词典和所述发音词典,得到组合发音词典;

根据所述组合发音词典,得到规范化发音词典。

6.如权利要求5所述的语音识别方法,其特征在于,所述根据所述组合发音词典,得到规范化发音词典的步骤,包括:

对所述组合发音词典中包含的音素统一大小写;

根据预设专有名词大小写规则,对所述组合发音词典中包含的音素进行大小写处理;

将静音词和/或噪声词和/或集外词对应的音素,添加至所述组合发音词典中,得到规范化发音词典。

7.如权利要求2-4中任一项所述的语音识别方法，其特征在于，所述对文本语料进行规范化处理，得到规范化文本语料的步骤，包括；

从多个领域中采集文本语料；

对所述文本语料进行规范化处理，得到规范化文本语料；

所述对所述规范化文本语料进行训练，得到语言模型的步骤，包括：

从所述规范化文本语料中获取使用频率高于预设阈值的中文词和/或单词；

根据所述使用频率高于预设阈值的中文词和/或单词，生成构造词汇表；

对所述规范化文本语料和所述构造词汇表进行训练，得到语言模型。

8.如权利要求7所述的语音识别方法，其特征在于，所述对所述文本语料进行规范化处理，得到规范化文本语料的步骤，包括：

根据预设删除规则对所述文本语料中的字符和/或字符串进行删除；

和/或，将所述文本语料中的非ASCII字符转换为ASCII字符；

和/或，将所述文本语料中的罗马数字转换为十进制；

和/或，对所述文本语料中的文本进行语义分割；

和/或，对所述文本语料中的单词统一大小写；

和/或，对所述文本语料中的中文词和/或单词进行纠错处理；

和/或，构造复合词和/或缩略词和/或人名和/或地名，并将其添加至所述文本语料中。

9.一种语音识别装置，其特征在于，所述语音识别装置包括：

提取模块，用于若监测接收到语音信息，则对所述语音信息进行声学特征提取，得到第一声学特征信息；

解码模块，用于利用目标解码器对所述第一声学特征信息进行解码，得到解码识别结果；其中，所述目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到；

输出模块，用于输出所述解码识别结果，以实现语音识别。

10.一种设备，其特征在于，所述设备包括：存储器、处理器及存储在所述存储器上并在所述处理器上运行语音识别程序，所述语音识别程序被所述处理器执行时实现如权利要求1-8中任一项所述的语音识别方法的步骤。

11.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有语音识别程序，所述语音识别程序被处理器执行时实现如权利要求1-8中任一项所述的语音识别方法的步骤。

语音识别方法及装置、设备、计算机可读存储介质

技术领域

[0001] 本发明涉及语音识别技术领域，尤其涉及一种语音识别方法及装置、设备、计算机可读存储介质。

背景技术

[0002] 随着计算机技术和信号处理技术的快速发展，健壮性语音识别已达到真正意义上的应用，能够实现自由的人机交互；但是，目前的语音识别准确率较低，例如在识别专有名词复合词如Editor-in-Chief、缩略词如UFO、人名如Jessie、地名如Beijing等的过程中识别准确率都较低，由此大大降低了用户的使用体验。

[0003] 因此，如何提升语音识别的准确率是亟待解决的问题。

发明内容

[0004] 本发明的主要目的在于提供语音识别方法及装置、设备、计算机可读存储介质，旨在提升语音识别的准确率。

[0005] 为实现上述目的，本发明提供一种语音识别方法，所述语音识别方法包括以下步骤：

[0006] 若监测接收到语音信息，则对所述语音信息进行声学特征提取，得到第一声学特征信息；

[0007] 利用目标解码器对所述第一声学特征信息进行解码识别，得到解码识别结果；其中，所述目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到；

[0008] 输出所述解码识别结果，以实现语音识别。

[0009] 可选的，所述利用目标解码器对所述第一声学特征信息进行解码，得到解码识别结果的步骤之前，所述语音识别方法还包括：

[0010] 对发音词典进行规范化处理，得到规范化发音词典；

[0011] 对文本语料进行规范化处理，得到规范化文本语料，并对所述规范化文本语料进行训练，得到语言模型；

[0012] 根据所述规范化处理后的发音词典，所述语言模型，以及声学模型构造解码器，以得到目标解码器。

[0013] 可选的，所述根据所述规范化处理后的发音词典，所述语言模型，以及声学模型构造解码器，以得到目标解码器的步骤之前，还包括：

[0014] 对语音语料进行声学特征提取，得到第二声学特征信息；

[0015] 对所述第二声学特征信息进行训练，得到声学模型。

[0016] 可选的，所述根据所述规范化处理后的发音词典，所述语言模型，以及声学模型构造解码器，以得到目标解码器的步骤之前，还包括：

[0017] 获取声学模型和所述语言模型；

[0018] 根据所述声学模型中的音素和所述语言模型中的中文词，建立音素与中文词的映射关系，以及根据所述声学模型中的音素和所述语言模型中的单词，建立音素与单词的映射关系；

[0019] 根据所述音素与中文词的映射关系以及音素与单词的映射关系，得到发音词典。

[0020] 可选的，所述对发音词典进行规范化处理，得到规范化发音词典的步骤，包括：

[0021] 对所述发音词典进行训练得到词转音素模型；

[0022] 根据所述词转音素模型，生成补充发音词典；其中，所述补充发音词典中包含的音素不在所述发音词典中，所述补充发音词典中包含的音素对应的中文词和音素对应的单词在所述语言模型中；

[0023] 根据所述补充发音词典和所述发音词典，得到组合发音词典；

[0024] 根据所述组合发音词典，得到规范化发音词典。

[0025] 可选的，所述根据所述组合发音词典，得到规范化发音词典的步骤，包括：

[0026] 对所述组合发音词典中包含的音素统一大小写；

[0027] 根据预设专有名词大小写规则，对所述组合发音词典中包含的音素进行大小写处理；

[0028] 将静音词和/或噪声词和/或集外词对应的音素，添加至所述组合发音词典中，得到规范化发音词典。

[0029] 可选的，所述对文本语料进行规范化处理，得到规范化文本语料的步骤，包括：

[0030] 从多个领域中采集文本语料；

[0031] 对所述文本语料进行规范化处理，得到规范化文本语料；

[0032] 所述对所述规范化文本语料进行训练，得到语言模型的步骤，包括：

[0033] 从所述规范化文本语料中获取使用频率高于预设阈值的中文词和/或单词；

[0034] 根据所述使用频率高于预设阈值的中文词和/或单词，生成构造词汇表；

[0035] 对所述规范化文本语料和所述构造词汇表进行训练，得到语言模型。

[0036] 可选的，所述对所述文本语料进行规范化处理，得到规范化文本语料的步骤，包括：

[0037] 根据预设删除规则对所述文本语料中的字符和/或字符串进行删除；

[0038] 和/或，将所述文本语料中的非ASCII字符转换为ASCII字符；

[0039] 和/或，将所述文本语料中的罗马数字转换为十进制；

[0040] 和/或，对所述文本语料中的文本进行语义分割；

[0041] 和/或，对所述文本语料中的单词统一大小写；

[0042] 和/或，对所述文本语料中的中文词和/或单词进行纠错处理；

[0043] 和/或，构造复合词和/或缩略词和/或人名和/或地名，并将其添加至所述文本语料中。

[0044] 此外，为实现上述目的，本发明还提供一种语音识别装置，语音识别装置包括：

[0045] 提取模块，用于若监测接收到语音信息，则对所述语音信息进行声学特征提取，得到第一声学特征信息；

[0046] 解码模块，用于利用目标解码器对所述第一声学特征信息进行解码，得到解码识别结果；其中，所述目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料

训练得到的语言模型，以及声学模型构造得到；

[0047] 输出模块，用于输出所述解码识别结果，以实现语音识别。

[0048] 此外，为实现上述目的，本发明还提供一种设备，所述设备包括：存储器、处理器及存储在所述存储器上并在所述处理器上运行语音识别程序，所述语音识别程序被所述处理器执行时实现如上文的语音识别方法的步骤。

[0049] 此外，为实现上述目的，本发明还提供一种计算机可读存储介质，所述计算机可读存储介质上存储有语音识别程序，语音识别程序被处理器执行时实现如上文的语音识别方法的步骤。

[0050] 本发明提供的技术方案，通过若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息，进而利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果，其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到，再输出解码识别结果，以实现语音识别；解决了相关技术中语音识别的准确率差的问题。

[0051] 也即本发明提供的技术方案，通过在利用目标解码器对提取的第一声学特征信息进行解码识别的过程中，目标解码器是根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到的，由于发音词典和语言模型经过规范化处理，相应地，由其构造得到的目标解码器更为规范，使得利用该更为规范的目标解码器解码识别的准确率更高，提升了语音识别准确率，进而提升了用户的使用体验满意度。

附图说明

[0052] 为了更清楚地说明本发明实施例的技术方案，下面将对实施例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示出的结构获得其他的附图。

[0053] 图1是本发明实施例方案涉及的硬件运行环境的设备结构示意图；

[0054] 图2为本发明语音识别方法第一实施例的流程示意图；

[0055] 图3为本发明语音识别方法第二实施例的流程示意图；

[0056] 图4为本发明语音识别方法第三实施例的流程示意图；

[0057] 图5为本发明语音识别方法第四实施例的流程示意图；

[0058] 图6为本发明语音识别装置第一实施例的结构框图；

[0059] 图7为本发明语音识别装置第一实施例执行语音识别方法的示意图。

[0060] 本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

[0061] 应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0062] 请参见图1所示，图1为本发明实施例方案涉及的硬件运行环境的设备结构示意图。

[0063] 设备包括：至少一个处理器101、存储器102以及存储在存储器上并可在处理器上运行的语音识别程序，语音识别程序配置为实现如下任一实施例的语音识别方法的步骤。

[0064] 处理器101可以包括一个或多个处理核心，比如4核心处理器、8核心处理器等。处理器101可以采用DSP(Digital Signal Processing，数字信号处理)、FPGA(Field－Programmable Gate Array，现场可编程门阵列)、PLA(Programmable Logic Array，可编程逻辑阵列)中的至少一种硬件形式来实现。处理器101也可以包括主处理器和协处理器，主处理器是用于对在唤醒状态下的数据进行处理的处理器，也称CPU(Central ProcessingUnit，中央处理器)；协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中，处理器101可以在集成有GPU(Graphics Processing Unit，图像处理器)，GPU用于负责显示屏所需要显示的内容的渲染和绘制。处理器101还可以包括AI(Artificial Intelligence，人工智能)处理器，该AI处理器用于处理有关语音识别方法操作，使得语音识别方法模型可以自主训练学习，提高效率和准确度。

[0065] 存储器102可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器102还可包括高速随机存取存储器，以及非易失性存储器，比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中，存储器102中的非暂态的计算机可读存储介质用于存储至少一个指令，该至少一个指令用于被处理器101所执行以实现本申请中方法实施例提供的语音识别方法。

[0066] 在一些实施例中，设备还可选包括有：通信接口103和至少一个外围设备。处理器101、存储器102和通信接口103之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与通信接口103相连。具体地，外围设备包括：射频电路104、显示屏105和电源106中的至少一种。

[0067] 通信接口103可被用于将I/O(Input/Output，输入/输出)相关的至少一个外围设备连接到处理器101和存储器102。在一些实施例中，处理器101、存储器102和通信接口103被集成在同一芯片或电路板上；在一些其他实施例中，处理器101、存储器102和通信接口103中的任意一个或两个可以在单独的芯片或电路板上实现，本实施例对此不加以限定。

[0068] 射频电路104用于接收和发射RF(Radio Frequency，射频)信号，也称电磁信号。射频电路104通过电磁信号与通信网络以及其他通信设备进行通信。射频电路104将电信号转换为电磁信号进行发送，或者，将接收到的电磁信号转换为电信号。可选地，射频电路104包括：天线系统、RF收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。射频电路104可以通过至少一种无线通信协议来与其它终端进行通信。该无线通信协议包括但不限于：城域网、各代移动通信网络(2G、3G、4G及5G)、无线局域网和/或WiFi(Wireless Fidelity，无线保真)网络。在一些实施例中，射频电路104还可以包括NFC(Near Field Communication，近距离无线通信)有关的电路，本申请对此不加以限定。

[0069] 显示屏105用于显示UI(User Interface，用户界面)。该UI可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏105是触摸显示屏时，显示屏105还具有采集在显示屏105的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器101进行处理。此时，显示屏105还可以用于提供虚拟按钮和/或虚拟键盘，也称软按钮和/或软键盘。在一些实施例中，显示屏105可以为一个，设备的前面板；在另一些实施例中，显示屏105可以为至少两个，分别设置在设备的不同表面或呈折叠设计；在一些实施例中，显示屏105可以是柔性显示屏，设置在设备的弯曲表面上或折叠面上。甚至，显示屏105还可以设

置成非矩形的不规则图形，也即异形屏。显示屏105可以采用LCD(LiquidCrystal Display，液晶显示屏)、OLED(Organic Light-Emitting Diode，有机发光二极管)等材质制备。

[0070] 电源106用于为设备中的各个组件进行供电。电源106可以是交流电、直流电、一次性电池或可充电电池。当电源106包括可充电电池时，该可充电电池可以支持有线充电或无线充电。该可充电电池还可以用于支持快充技术。

[0071] 本领域技术人员可以理解，图1中示出的结构并不构成对设备的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。

[0072] 基于上述硬件结构，提出本发明的各实施例。

[0073] 请参见图2所示，图2为本发明语音识别方法第一实施例的流程示意图，语音识别方法包括以下步骤：

[0074] 步骤S21：若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息。

[0075] 本实施例中接收到的语音信息可以是用户实时下发的，也可以是用户提前录制好上传的，还可以是从网上等下载并上传的；在实际应用中，可以根据具体应用场景做灵活调整。

[0076] 可以理解的是，本实施例中对语音信息进行提取的声学特征，指的是表示语音声学特性的物理量，其是声音诸要素声学表现的统称，例如表示音色的能量集中区、共振峰频率、共振峰强度和带宽，以及表示语音韵律特性的时长、基频、平均语声功率等。

[0077] 在一些示例中，对语音信息进行声学特征提取得到第一声学特征信息，可以是通过梅尔倒谱系数MFCC特征提取；具体地，MFCC特征提取包括A/D转换、预加重、加窗分帧、DFT+取平方、Mel滤波、取对数、IDFT、动态特征等步骤。

[0078] 在一些示例中，对语音信息进行声学特征提取得到第一声学特征信息，可以是通过深度学习特征提取；具体地，深度学习特征提取包括采样、分帧、傅里叶变换、识别字符、获取映射图等步骤。

[0079] 在一些示例中，可以每隔预设时长监测是否接收到语音信息，例如每隔10s监测是否接收到语音信息，进而在监测接收到语音信息时，对语音信息进行声学特征提取，得到第一声学特征信息；这样能够在一定程度上减少系统消耗，节省电量。

[0080] 在一些示例中，可以持续监测是否接收到语音信息，进而在监测接收到语音信息时，对语音信息进行声学特征提取，得到第一声学特征信息；这样能够在一定程度上提升监测准确率，并在第一时间获取到语音信息，提升语音识别速率。

[0081] 步骤S22：利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果；其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到。

[0082] 本实施例中对接收到的语音信息进行声学特征提取得到第一声学特征信息之后，需要利用目标解码器对第一声学特征信息进行解码识别，从而得到解码识别结果。需要说明的是，本实施例中的目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到，即本实施例中会先对发音词典和语言模型都是进行规范化处理，进而利用规范化处理后的发音词典和语言模型，以及声学模型来构造得到目标解码器；这样根据经过规范化处理后的发音词典和语言模型，以及声学模

型来构造得到目标解码器更为规范，使得利用该更为规范的目标解码器解码识别的准确率更高，从而提升了语音识别准确率。

[0083] 其中，本实施例中的语言模型包含多个词串，它们在文本语料库中出现的概率大小，例如不合语法的词串概率接近0，很合乎语法的词串概率大。可以理解的是，语言模型是由采集到的各个领域的文本语料训练而成；具体地，采集到的各个领域的文本语料可以是通用领域涵盖政治、经济、社会、文化、宗教、体育、格式文档、目录等新闻类文本语料，也可以是特定领域涵盖消息类等对话类文本语料等；通常情况下，采集到的文本语料来自更为广泛的领域，从而训练得到的语言模型更为准确。

[0084] 其中，本实施例中的声学模型包含多个识别单个音素的模型，例如音素a的模型可以判定小段语音是否是a，音素b的模型可以判定小段语音是否是b等。可以理解的是，声学模型是由采集到的大批量的语音语料训练而成；具体地，采集到的大批量的语音语料可以是涉及各地口音、不同年龄、不同性别、不同语速、不同声音大小等语音语料；通常情况下，采集到的语音语料来自涉及的因素越多，从而训练得到的声学模型更为准确。

[0085] 其中，本实施例中的发音词典包含系统所能处理的单词的集合，并标明了其发音，通过发音字典得到声学模型的建模单元和语言模型建模单元间的映射关系，从而把声学模型和语言模型连接起来，组成一个搜索的状态空间以用于解码器进行解码工作；具体地，映射关系是音素与中文词的映射关系，以及音素与单词的映射关系，可以理解的是，音素与中文词的映射关系，最初是拼音与中文词的映射关系，经过预设拼音与音素的转换规则处理，从而得到音素与中文词的映射关系。

[0086] 可以理解的是，本实施例中利用目标解码器对第一声学特征信息进行解码识别，具体地，是根据语言模型确定出第一声学特征信息对应的所有可能的词串，并根据发音词典展开为音素串，再根据声学模型得到解码图，然后在解码图上实施Viterbi算法，得到最佳序列，进而得到识别结果。

[0087] 步骤S23：输出解码识别结果，以实现语音识别。

[0088] 本实施例中利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果之后，将解码识别结果进行输出；可以理解的是，将解码识别结果进行输出的方式包括但不限于输出在终端屏幕上，其中，该终端可以是与接收语音信息相同的终端，也可以是与接收语音信息不同的终端；在实际应用中，可以根据具体应用场景做灵活调整。

[0089] 本实施例中，在利用目标解码器对提取的第一声学特征信息进行解码识别的过程中，目标解码器是根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到的，由于发音词典和语言模型经过规范化处理，相应地，由其构造得到的目标解码器更为规范，使得利用该更为规范的目标解码器解码识别的准确率更高，提升了语音识别准确率，进而提升了用户的使用体验满意度。

[0090] 基于上述实施例，提出本发明语音识别方法的第二实施例，请参见图3所示，图3为本发明语音识别方法的第二实施例的流程示意图。

[0091] 本实施例中，步骤S22利用目标解码器对第一声学特征信息进行解码，得到解码识别结果的步骤之前，语音识别方法还可以包括以下步骤：

[0092] 步骤S31：对发音词典进行规范化处理，得到规范化发音词典。

[0093] 本实施例中对发音词典进行规范化处理，得到规范化发音词典的步骤，可以包括

以下步骤：

[0094] 首先，对发音词典进行训练得到词转音素模型；

[0095] 然后，根据词转音素模型，生成补充发音词典；其中补充发音词典中包含的音素不在发音词典中，补充发音词典中包含的音素对应的中文词和音素对应的单词在语言模型中；

[0096] 其次，根据补充发音词典和发音词典，得到组合发音词典；

[0097] 进而，根据组合发音词典，得到规范化发音词典。

[0098] 应当明确的是，本实施例中是对获取到的发音词典进行训练得到词转音素(Grapheme-to-Phoneme，G2P)模型，然后根据G2P模型，生成补充发音词典；其中补充发音词典中包含的音素不在发音词典中，但是补充发音词典中包含的音素对应的中文词和音素对应的单词在语言模型中，即在原有发音词典的基础之上，另外生成了一个补充发音词典，其包含的音素在原有发音词典中是没有的，以实现补充发音词典对原有发音词典的补充，再根据补充发音词典和原有发音词典得到组合发音词典，这样得到的组合发音词典中包含的音素是全面、完善的；进而再根据该全面、完善的组合发音词典，得到规范化发音词典。

[0099] 本实施例中根据组合发音词典，得到规范化发音词典的步骤，可以包括以下步骤：

[0100] 首先，对组合发音词典中包含的音素统一大小写。

[0101] 可以理解的是，本实施例中得到组合发音词典之后，可以通过首先对组合发音词典中包含的音素统一大小写；具体地，可以将发音词典中包含的音素统一为大写，或者可以将发音词典中包含的音素统一为小写，这样先进行统一大小写处理更加便于后续操作。

[0102] 然后，根据预设专有名词大小写规则，对组合发音词典中包含的音素进行大小写处理。

[0103] 可以理解的是，不同专有名词有对应的大小写规则，例如人名或地名的首字母大写，因此，本实施例中在对组合发音词典中包含的音素统一大小写之后，可以进一步根据预设专有名词大小写规则，对组合发音词典中包含的音素进行大小写处理；例如若将发音词典中包含的音素统一为大写之后，进一步将人名或地名除首字母大写之外，调整为小写，若将发音词典中包含的音素统一为小写之后，进一步将人名或地名首字母调整为大写。

[0104] 其次，将静音词和/或噪声词和/或集外词对应的音素，添加至组合发音词典中，得到规范化发音词典。

[0105] 可以理解的是，为了使得组合发音词典中包含的音素更加全面、完善，因此，本实施例中可以预先构造好静音词和/或噪声词和/或集外词对应的音素，进而将预先构造好的静音词和/或噪声词和/或集外词对应的音素添加至组合发音词典中，从而得到更为全面、完善的组合发音词典；其中，静音词指的是停顿相关词，噪声词指的是语气词如啊、呀、呢等，集外词指的是未在发音词典中的词。

[0106] 需要说明的是，上述对组合发音词典进行一些列步骤处理之后，组合发音词典更加规范、全面、完善；因而为了区分，将其称之为规范化发音词典。

[0107] 步骤S32：对文本语料进行规范化处理，得到规范化文本语料，并对规范化文本语料进行训练，得到语言模型。

[0108] 本实施例中对文本语料进行规范化处理，得到规范化文本语料的步骤，可以包括以下步骤；

[0109] 首先，从多个领域中采集文本语料；

[0110] 然后，对文本语料进行规范化处理，得到规范化文本语料。

[0111] 应当明确的是，本实施例中是从多个领域中采集文本语料，其中领域可以是通用领域涵盖政治、经济、社会、文化、宗教、体育、格式文档、目录等新闻类文本语料，也可以是特定领域涵盖消息类等对话类文本语料等，这样从多个领域中采集得到的文本语料更加全面、完善；进而对采集到的文本语料进行规范化处理，得到规范化文本语料。

[0112] 本实施例中对文本语料进行规范化处理，得到规范化文本语料的步骤，可以包括以下步骤：

[0113] 根据预设删除规则对文本语料中的字符和/或字符串进行删除；

[0114] 和/或，将文本语料中的非ASCII字符转换为ASCII字符；

[0115] 和/或，将文本语料中的罗马数字转换为十进制；

[0116] 和/或，对文本语料中的文本进行语义分割；

[0117] 和/或，对文本语料中的单词统一大小写；

[0118] 和/或，对文本语料中的中文词和/或单词进行纠错处理；

[0119] 和/或，构造复合词和/或缩略词和/或人名和/或地名，并将其添加至文本语料中。

[0120] 相应地，本实施例中对规范化文本语料进行训练，得到语言模型的步骤，可以包括以下步骤：

[0121] 首先，从规范化文本语料中获取使用频率高于预设阈值的中文词和/或单词；

[0122] 然后，根据使用频率高于预设阈值的中文词和/或单词，生成构造词汇表；

[0123] 其次，对规范化文本语料和构造词汇表进行训练，得到语言模型。

[0124] 应当明确的是，本实施例中得到规范化文本语料之后，可以首先从规范化文本语料中获取使用频率高于预设阈值的中文词和/或单词，然后根据使用频率高于预设阈值的中文词和/或单词，生成构造词汇表，进而根据规范化文本语料和构造词汇表进行训练，得到语言模型；即会根据规范化文本语料中使用频率高于预设阈值的中文词和/或单词生成的构造词汇表，和规范化文本语料两者来共同训练得到语言模型，提升了训练得到的语言模型的准确率，从而进一步提升语音识别准确率。

[0125] 步骤S33：根据规范化处理后的发音词典，语言模型，以及声学模型构造解码器，以得到目标解码器。

[0126] 应当明确的是，本实施例中对发音词典进行规范化处理后，以及对文本语料进行规范化处理后训练得到语言模型，进而便可以利用该规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造解码器，以得到目标解码器。

[0127] 本实施例中，利用规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造解码器，得到目标解码器，这样得到的目标解码器更为规范、全面、完善，从而利用该目标解码器进行声学特征信息的解码识别，得到的解码识别结果更为准确，即提升了语音识别准确率；且整个流程简单、易于开发实现。

[0128] 基于上述实施例，提出本发明语音识别方法的第三实施例，请参见图4所示，图4为本发明语音识别方法的第三实施例的流程示意图。

[0129] 本实施例中，步骤S33根据规范化处理后的发音词典，语言模型，以及声学模型构造解码器，以得到目标解码器的步骤之前，还可以包括以下步骤：

[0130] 步骤S41:对语音语料进行声学特征提取,得到第二声学特征信息;

[0131] 步骤S42:对第二声学特征信息进行训练,得到声学模型。

[0132] 应当明确的是,本实施例中可以通过首先采集大批量的语音语料,其中语音语料可以是涉及各地口音、不同年龄、不同性别、不同语速、不同声音大小等语音语料,这样从不同因素考虑采集到的语音语料更加全面、完善;然后对采集到的语音语料进行声学特征提取,得到第二声学特征信息,进而再对第二声学特征信息进行训练,得到声学模型。

[0133] 本实施例中,通过从不同因素考虑采集语音语料,使得采集到的语音语料更加全面、完善,从而从语音语料中提取出的声学特征信息也更加全面、完善,提升了训练得到的声学模型的准确率,从而进一步提升语音识别准确率。

[0134] 基于上述实施例,提出本发明语音识别方法的第四实施例,请参见图5所示,图5为本发明语音识别方法的第四实施例的流程示意图。

[0135] 本实施例中,步骤S33根据规范化处理后的发音词典,语言模型,以及声学模型构造解码器,以得到目标解码器的步骤之前,还可以包括以下步骤:

[0136] 步骤S51:获取声学模型和语言模型;

[0137] 步骤S52:根据声学模型中的音素和语言模型中的中文词,建立音素与中文词的映射关系,以及根据声学模型中的音素和语言模型中的单词,建立音素与单词的映射关系;

[0138] 步骤S53:根据音素与中文词的映射关系以及音素与单词的映射关系,得到发音词典。

[0139] 应当明确的是,本实施例中可以通过首先获取声学模型和语言模型,其中语言模型是根据规范化处理后的文本语料训练得到的,然后根据声学模型中的音素和语言模型中的中文词,建立音素与中文词的映射关系,以及根据声学模型中的音素和语言模型中的单词,建立音素与单词的映射关系,进而再根据建立的音素与中文词的映射关系以及音素与单词的映射关系,得到发音词典。

[0140] 本实施例中,发音词典是根据声学模型和规范化处理后的文本语料训练得到的语言模型得到,因此,在此基础上得到的发音词典也是较为规范、全面、完善的,并对得到的该发音词典本身进行规范化处理,进一步使得发音词典更加规范、全面、完善;且由于在前得到的发音词典已经比较规范、全面、完善,再对该发音词典本身进行规范化处理,能够减少规范化处理时间,提升了对该发音词典规范化处理效率。

[0141] 此外,请参见图6所示,本发明实施例在上述语音识别方法的基础上,还提出一种语音识别装置,语音识别装置包括:

[0142] 提取模块601,用于若监测接收到语音信息,则对语音信息进行声学特征提取,得到第一声学特征信息;

[0143] 解码模块602,用于利用目标解码器对第一声学特征信息进行解码,得到解码识别结果;其中,目标解码器为根据规范化处理后的发音词典,规范化处理后的文本语料训练得到的语言模型,以及声学模型构造得到;

[0144] 输出模块603,用于输出解码识别结果,以实现语音识别。

[0145] 需要说明的是,本实施例中语音识别装置还可选的包括有对应的其他模块,以实现上述语音识别方法的步骤;其中,为了更好地理解,请参见图7所示,为本发明语音识别装置实现上述语音识别方法步骤的示意图。

[0146] 本发明的语音识别装置采用了上述所有实施例的全部技术方案，因此至少具有上述实施例的技术方案所带来的所有有益效果，在此不再一一赘述。

[0147] 此外，本发明实施例还提出一种计算机可读存储介质，计算机可读存储介质上存储有语音识别程序，语音识别程序被处理器执行时实现如前述语音识别方法的步骤。

[0148] 该计算机可读存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、计算机程序模块或其他数据)的任何方法或技术中实施的易失性或非易失性、可移除或不可移除的介质。计算机可读存储介质包括但不限于RAM(Random Access Memory，随机存取存储器)，ROM(Read-Only Memory，只读存储器)，EEPROM(Electrically Eraable Programmable read only memory，带电可擦可编程只读存储器)、闪存或其他存储器技术、CD-ROM(Compact Disc Read-Only Memory，光盘只读存储器)，数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。

[0149] 可见，本领域的技术人员应该明白，上文中所公开方法中全部或某些步骤、系统、设备中功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。在硬件实施方式中，在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分；例如，一个物理组件可以具有多个功能，或者一个功能或步骤可以由若干物理组件合作执行。某些物理组件或所有物理组件可以被实施为由处理器，如中央处理器、数字信号处理器或微处理器执行的软件，或者被实施为硬件，或者被实施为集成电路，如专用集成电路。

[0150] 以上仅为本发明的优选实施例，并非因此限制本发明的专利范围，凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本发明的专利保护范围内。

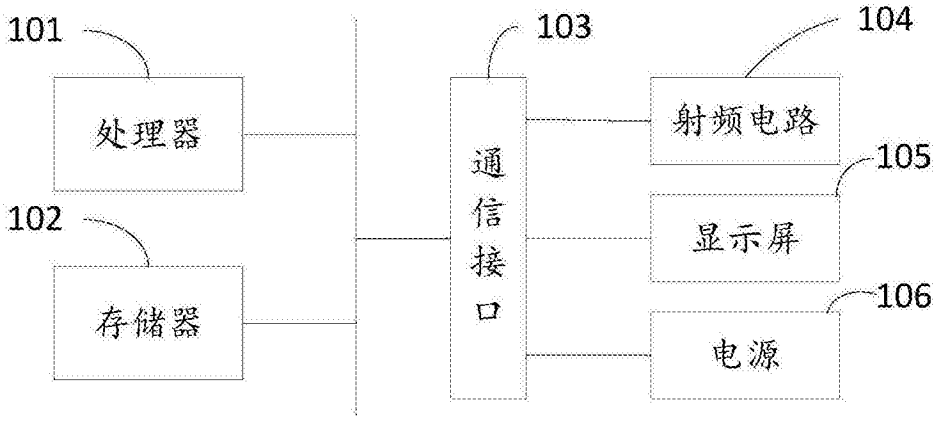

101
处理器
102
存储器
103
通信接口
104
射频电路
105
显示屏
106
电源

图1

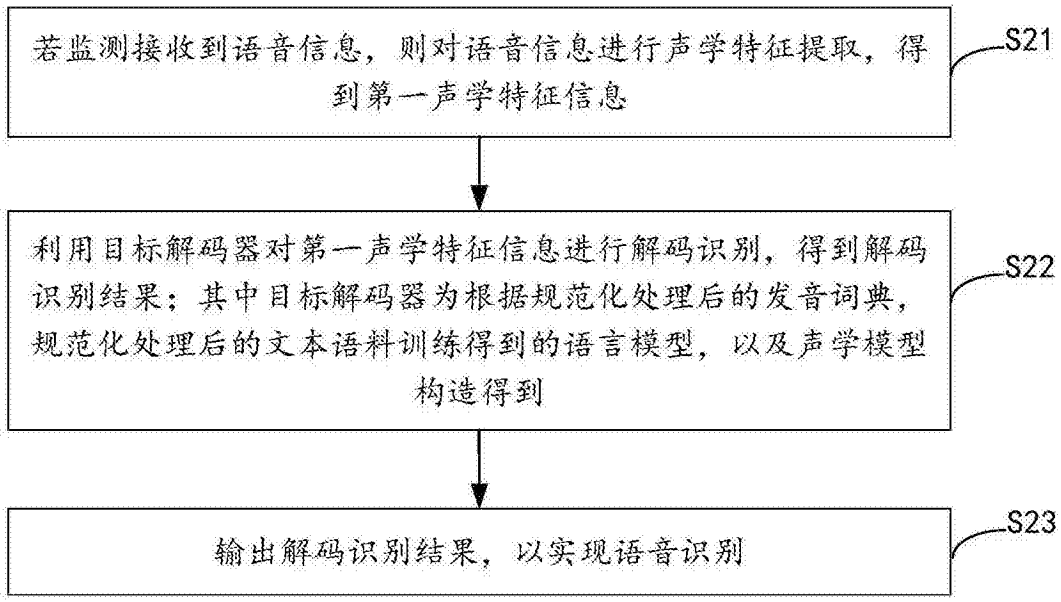

若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息
S21
利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果；其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到
S22
输出解码识别结果，以实现语音识别
S23

图2

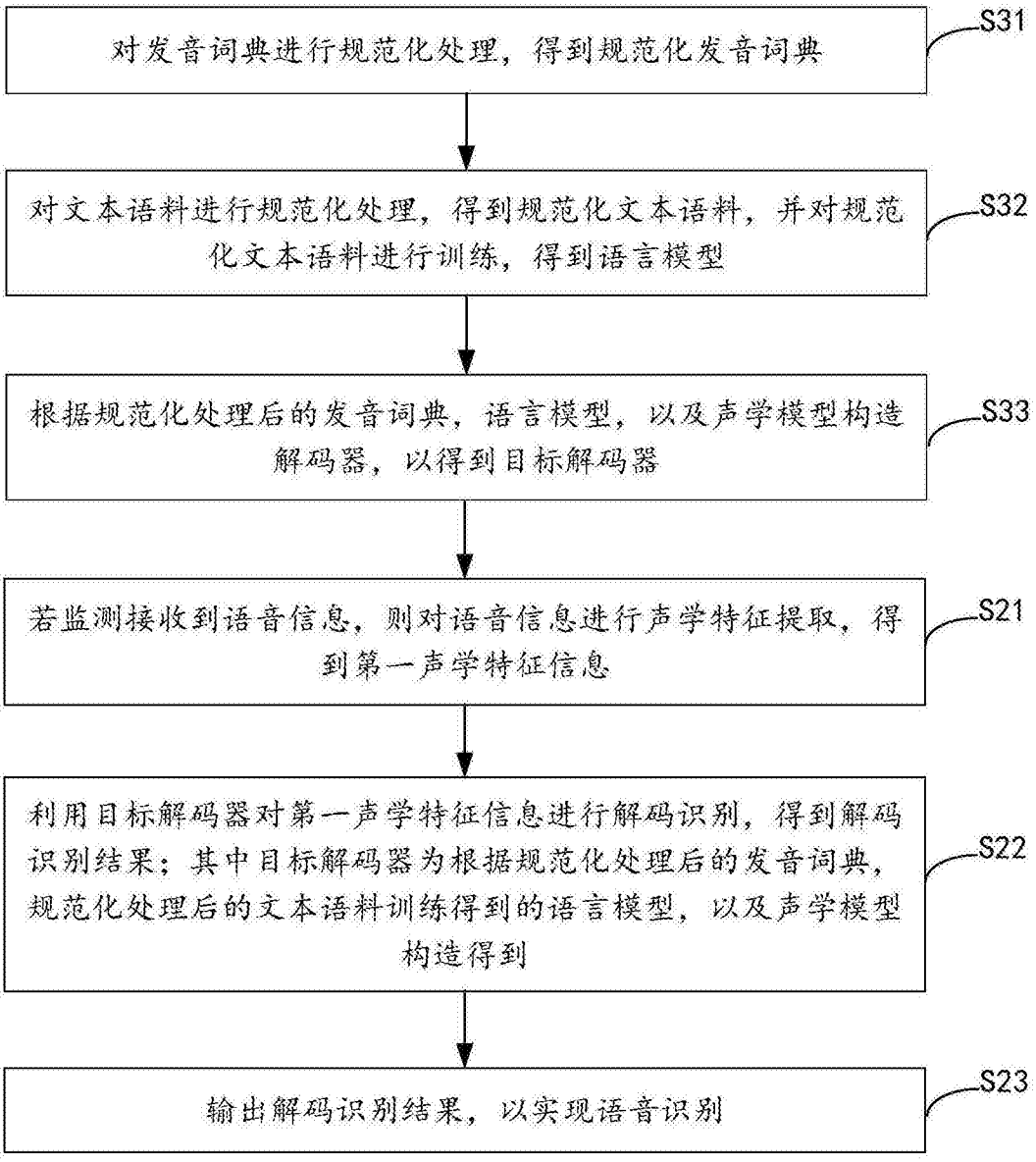

对发音词典进行规范化处理，得到规范化发音词典
S31
对文本语料进行规范化处理，得到规范化文本语料，并对规范化文本语料进行训练，得到语言模型
S32
根据规范化处理后的发音词典，语言模型，以及声学模型构造解码器，以得到目标解码器
S33
若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息
S21
利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果；其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到
S22
输出解码识别结果，以实现语音识别
S23

图3

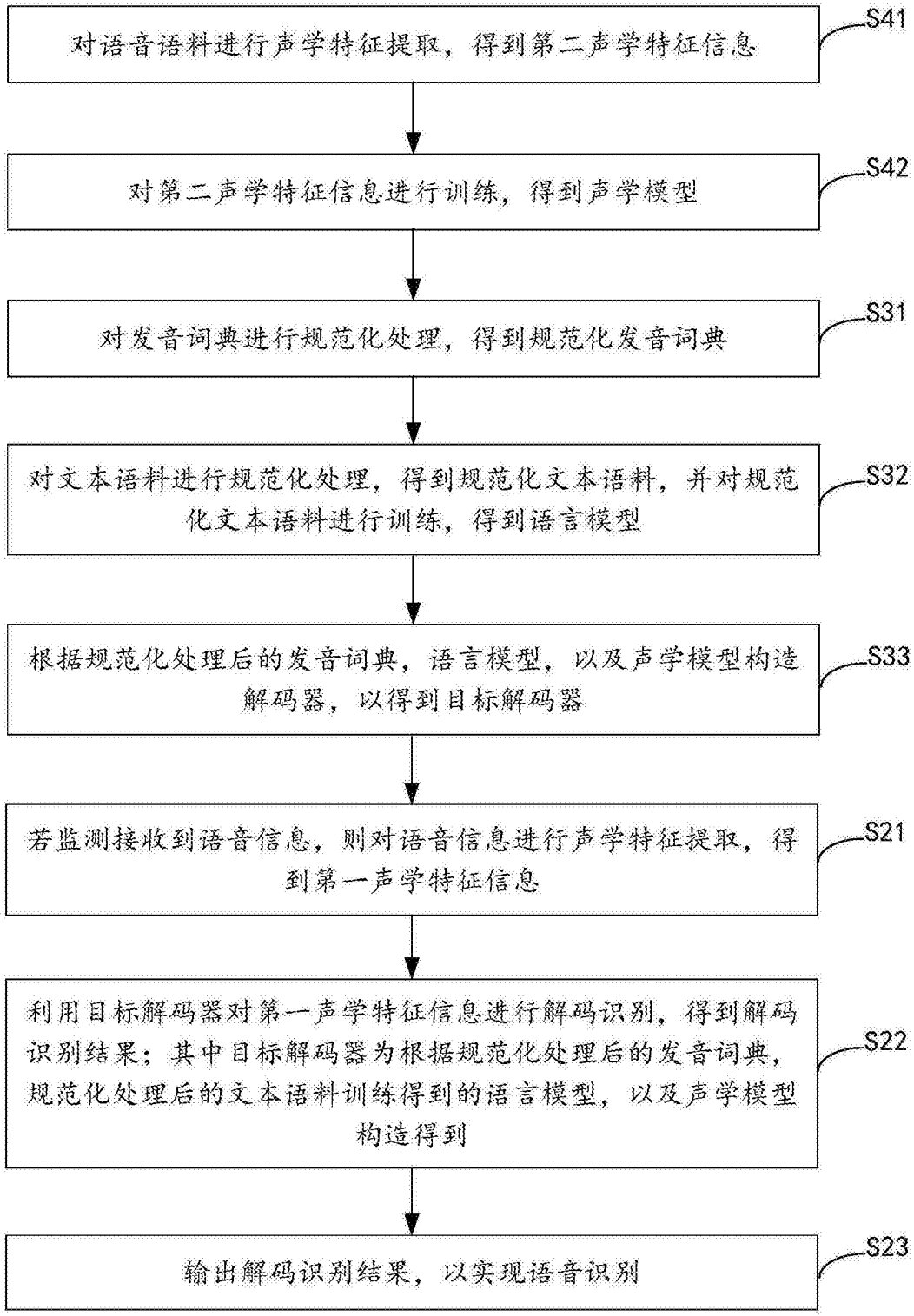

对语音语料进行声学特征提取，得到第二声学特征信息
S41
对第二声学特征信息进行训练，得到声学模型
S42
对发音词典进行规范化处理，得到规范化发音词典
S31
对文本语料进行规范化处理，得到规范化文本语料，并对规范化文本语料进行训练，得到语言模型
S32
根据规范化处理后的发音词典，语言模型，以及声学模型构造解码器，以得到目标解码器
S33
若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息
S21
利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果；其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到
S22
输出解码识别结果，以实现语音识别
S23

图4

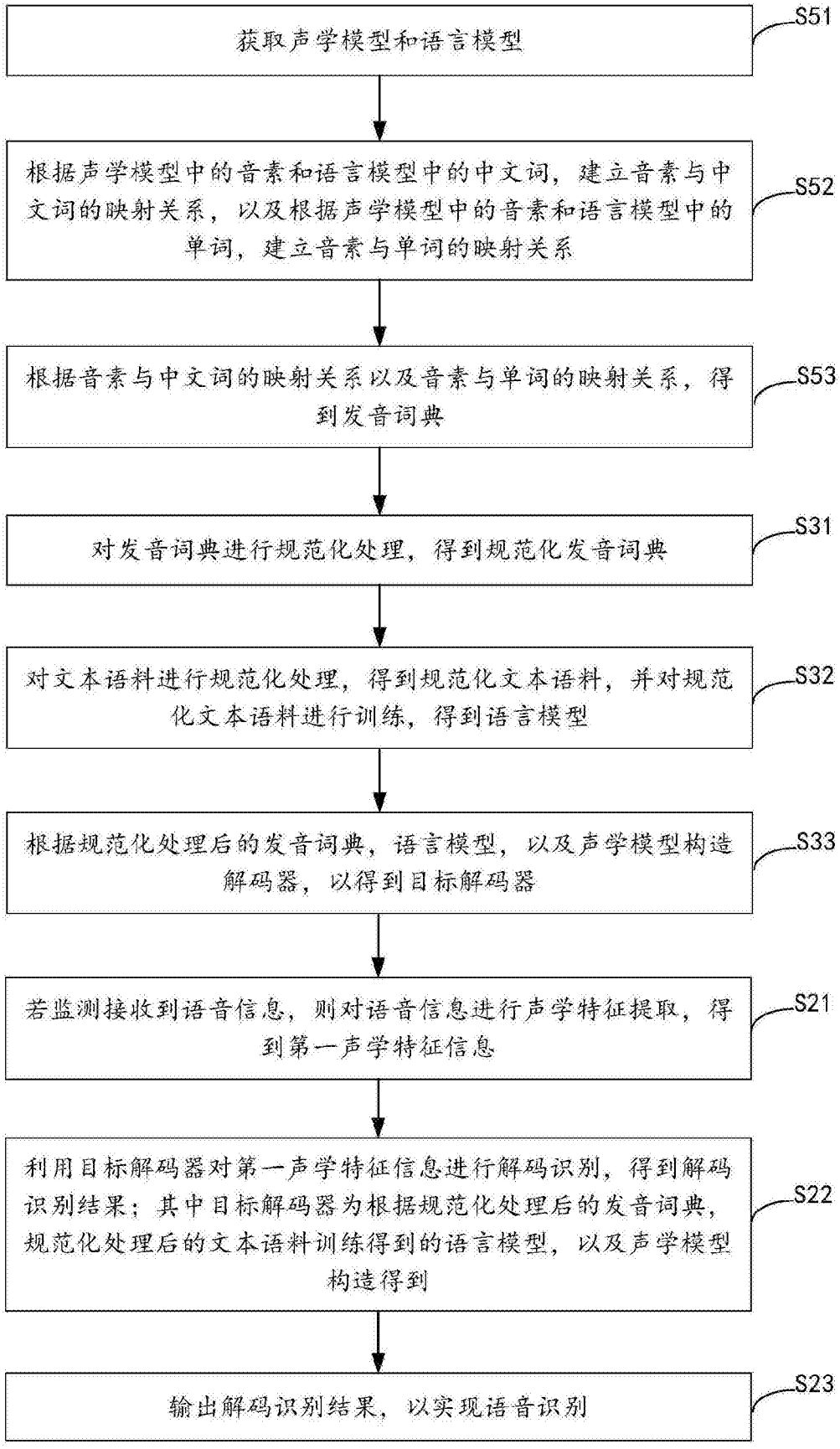

获取声学模型和语言模型
S51
根据声学模型中的音素和语言模型中的中文词，建立音素与中文词的映射关系，以及根据声学模型中的音素和语言模型中的单词，建立音素与单词的映射关系
S52
根据音素与中文词的映射关系以及音素与单词的映射关系，得到发音词典
S53
对发音词典进行规范化处理，得到规范化发音词典
S31
对文本语料进行规范化处理，得到规范化文本语料，并对规范化文本语料进行训练，得到语言模型
S32
根据规范化处理后的发音词典，语言模型，以及声学模型构造解码器，以得到目标解码器
S33
若监测接收到语音信息，则对语音信息进行声学特征提取，得到第一声学特征信息
S21
利用目标解码器对第一声学特征信息进行解码识别，得到解码识别结果；其中目标解码器为根据规范化处理后的发音词典，规范化处理后的文本语料训练得到的语言模型，以及声学模型构造得到
S22
输出解码识别结果，以实现语音识别
S23

图5

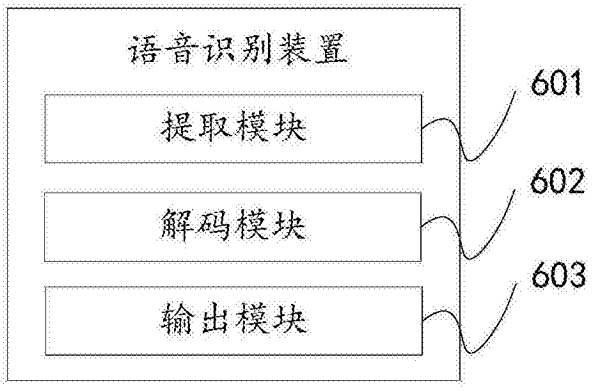


图6

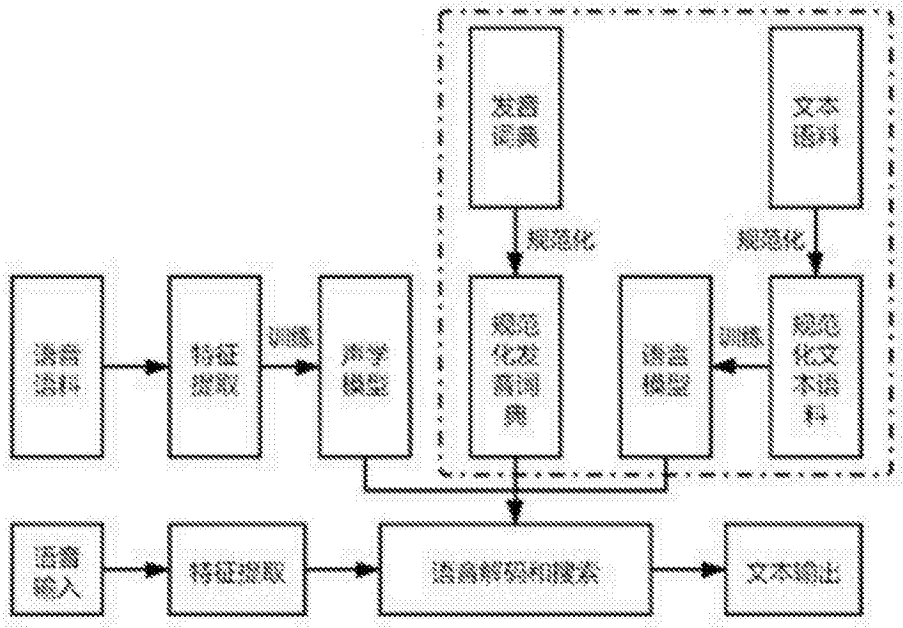


图7