(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 112951237 A
(43)申请公布日 2021.06.11

(21)申请号 202110293229.9

(22)申请日 2021.03.18

(71)申请人 深圳奇实科技有限公司
地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室

(72)发明人 张子奇 聂鹏

(74)专利代理机构 北京广技专利代理事务所(特殊普通合伙) 11842
代理人 崔征

(51)Int.Cl.
G10L 15/26(2006.01)
G10L 15/02(2006.01)
G10L 15/16(2006.01)
G10L 15/06(2013.01)
G10L 15/08(2006.01)
G10L 15/18(2013.01)
G10L 21/0208(2013.01)
G10L 25/57(2013.01)

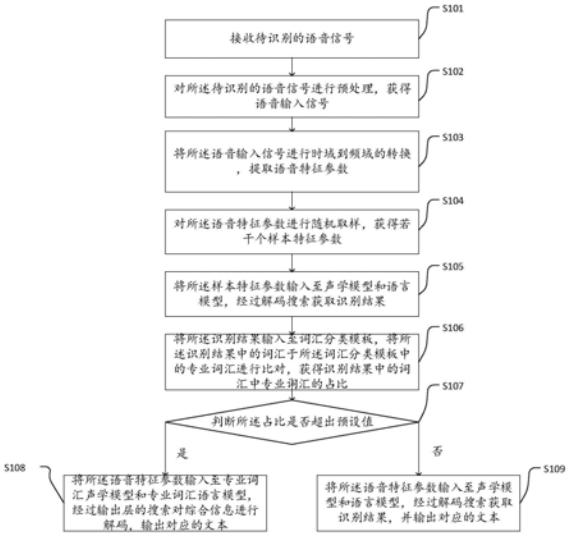
权利要求书4页 说明书12页 附图2页

(54)发明名称

一种基于人工智能的自动语音识别方法及系统

(57)摘要

本发明公开了一种基于人工智能的自动语音识别方法及系统，该方法包括应用词汇分类模板，将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对，获得识别结果中的词汇中专业词汇的占比，结合占比判断是否需要专业词汇的语音识别，采用本发明提供的方案可以提高对专业词汇识别的精确度和准确率，特别是增强专业领域中视频会议记录的准确性、精准性，特别是专业性，提高企业在相关专业领域的专业性，更重要的是减少因为对专业词汇的自动识别语音识别时造成的词汇识别的误解，防止因为语音识别造成误解进而造成重大损失。同时，由于以词汇分类模板做基础，提高专业词汇的搜索速率，进而提高了针对专业词汇的自动语音的识别效率。



CN 112951237 A

1.一种基于人工智能的自动语音识别方法,其特征在于,包括:

接收待识别的语音信号;

对所述待识别的语音信号进行预处理,获得语音输入信号;

将所述语音输入信号进行时域到频域的转换,提取语音特征参数;

对所述语音特征参数进行随机取样,获得若干个样本特征参数;

将所述样本特征参数输入至声学模型和语言模型,经过解码搜索获取识别结果;

将所述识别结果输入至词汇分类模板,将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对,获得识别结果中的词汇中专业词汇的占比;

判断所述占比是否超出预设值,若是,将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型,经过输出层的搜索对综合信息进行解码,输出对应的文本;所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配,提高获得专业词汇的概率;

若否,将所述语音特征参数输入至声学模型和语言模型,经过解码搜索获取识别结果,并输出对应的文本。

2.根据权利要求1所述的基于人工智能的自动语音识别方法,其特征在于,在所述并输出对应的文本之后,包括:

将输出的所述文本输入至拼写纠错模型,获得纠错后文本;

将纠错后的文本作为最终文本输出。

3.根据权利要求1所述的基于人工智能的自动语音识别方法,其特征在于,所述词汇分类模板构建方法包括:

获取大量分属于不同行业的专业词汇;

将所述专业词汇采用卷积神经网络按照专业词汇所属的行业进行分类训练;

获得分类结果,并将所述分类结果存储于分类数据库中,构成词汇分类模板。

4.根据权利要求3所述的基于人工智能的自动语音识别方法,其特征在于,所述专业词汇声学模型构建方法包括:

将词汇分类模板中的分类数据库设置为专业词汇字典;

基于音素或其组合优先从所述专业词汇字典中进行映射;

若所述专业词汇字典中无映射内容,则基于声学模型中的字典进行映射;

根据上述映射结果获取相应音素及其组合的声学得分。

5.根据权利要求3所述的基于人工智能的自动语音识别方法,其特征在于,所述专业词汇语言模型的构建方法包括:

基于词汇分类模板中分类数据库中存储的专业词汇,结合词典获取专业词汇的词序及连接词;所述词序及连接词的概率值排序为前五位;

将所述获取到的词序及连接词结合专业词汇记录在专业词汇语言数据库中;

基于所述声学得分及所述专业词汇语言数据库,确定语言得分。

6.根据权利要求1所述的基于人工智能的自动语音识别方法,其特征在于,所述对所述待识别的语音信号进行预处理的方法包括:

A1,获取环境中有规律噪声的频谱;

A2,获取收音装置噪声的频谱;

A3,基于环境噪声的频谱和收音装置噪声的频谱,结合最小方差无畸变响应滤波器增强后的信号采用下述公式确定:

$$S(f,t,n)=\sum_{i=1}^{P}w_i(f)\left|\sum(R_i*x_i(f,t,n))+\varepsilon\right|^2$$

$$+\sum_{t=1}^{T}\sum_{i=1}^{P}\frac{1}{T}[N_T(f,t,n)+N_i(f,t,n)]^2w_i(f)$$

其中,$N_T(f,t,n)$为环境中有规律噪声的频谱;$N_i(f,t,n)$为收音装置噪声的频谱;$Y_i(f,t,n)$为包含噪声的语音信号;$w_i(f)$为滤波器的加权系数;$S(f,t,n)$是获得的语音输入信号;$x_i(f,t,n)$为待降噪的信号;

f为当前频率,t为当前时间,n为当前帧,P为收音装置的数量,i=1,2...P,T为有规律噪声出现的次数,t=1,2...T,R_i是训练误差取最小值时对应的初始系数,ε表示训练误差的最小值;

A4,基于所述获得的语音输入信号,采用下述公式对所述语音输入信号的噪声判定值,若每个信号数据的噪声判定值G大于预设的判定阈值,则将该信号数据判定为噪声点,其中G值的计算公式为:

$$G_i=\frac{1}{\sqrt{2\times\pi}}exp\left(\frac{\frac{|a_i-a|}{\sum_{j=1}^{N}|a_j-a|}*\frac{\left|a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right|}{\sqrt{\frac{1}{N}*\sum_{k=1}^{N}\left(a_k-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}}}{\sqrt{\frac{\sum_{i=1}^{N}\left(a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}{N-1}}}\right)$$

其中,a_k信号数据集合M中的第k个信号数据;a_i代表信号据集合M中的第i个信号数据,a_j代表信号数据集合M中的第j个信号数据,i=1,2,3...N,j=1,2,3...N;G_i代表信号数据集合M中第i个信号数据的噪声判定值,π代表自然常数,exp代表指数函数,a代表信号数据集合M中信号数据的中值;

A5,将数据集合M中的每个信号数据都进行一一判定,当为噪声点时,则进行剔除,不为噪声点时,则进行保留,将保留后的信号数据形成最后处理后的信号。

7.一种基于人工智能的自动语音识别系统,其特征在于,包括:

接收装置,用于接收待识别的语音信号;

预处理装置,用于对所述待识别的语音信号进行预处理,获得语音输入信号;

提取装置,用于将所述语音输入信号进行时域到频域的转换,提取语音特征参数;

抽样装置,用于对所述语音特征参数进行随机取样,获得若干个样本特征参数;

结果获取装置,用于将所述样本特征参数输入至声学模型和语言模型,经过解码搜索获取识别结果;

专业词汇设置装置,用于将所述识别结果输入至词汇分类模板,将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对,获得识别结果中的词汇中专业词汇的占

比；

判断装置，用于判断所述占比是否超出预设值；

第一输出装置，用于当判断装置的判断结果为是时，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率；

第二输出装置，用于当判断装置的判断结果为否时，将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

8.根据权利要求7所述的基于人工智能的自动语音识别系统，其特征在于，所述专业词汇设置装置中词汇分类模板包括：

获取子装置，用于获取大量分属于不同行业的专业词汇；

训练子装置，用于将所述专业词汇采用卷积神经网络按照专业词汇所属的行业进行分类训练；

分类结果获取子装置，用于获得分类结果，并将所述分类结果存储于分类数据库中，构成词汇分类模板。

9.根据权利要求7所述的基于人工智能的自动语音识别系统，其特征在于，所述第一输出装置中所述专业词汇声学模型包括：

分类子装置，用于将词汇分类模板中的分类数据库设置为专业词汇字典；

第一映射子装置，用于基于音素或其组合优先从所述专业词汇字典中进行映射；

第二映射子装置，用于当第一映射子装置中所述专业词汇字典中无映射内容时，则基于声学模型中的字典进行映射；

声学得分子装置，用于根据上述映射结果获取相应音素及其组合的声学得分。

10.根据权利要求1所述的基于人工智能的自动语音识别系统，其特征在于，所述预处理装置包括：

第一噪声频谱获取子装置，用于获取环境中有规律噪声的频谱；

第二噪声频谱获取子装置，用于获取收音装置噪声的频谱；

信号确定子装置，用于基于环境噪声的频谱和收音装置噪声的频谱，结合最小方差无畸变响应滤波器增强后的信号采用下述公式确定：

$$S(f,t,n)=\sum_{i=1}^{P}w_i(f)\left|\sum(R_i*x_i(f,t,n))+\varepsilon\right|^2$$

$$+\sum_{t=1}^{T}\sum_{i=1}^{P}\frac{1}{T}[N_T(f,t,n)+N_i(f,t,n)]^2w_i(f)$$

其中，$N_T(f,t,n)$为环境中有规律噪声的频谱；$N_i(f,t,n)$为收音装置噪声的频谱；$Y_i(f,t,n)$为包含噪声的语音信号；$w_i(f)$为滤波器的加权系数；$S(f,t,n)$是获得的语音输入信号；$x_i(f,t,n)$为待降噪的信号；

f为当前频率，t为当前时间，n为当前帧，P为收音装置的数量，i＝1,2...P，T为有规律噪声出现的次数，t＝1,2...T。R_i是训练误差取最小值时对应的初始系数，ε表示训练误差

的最小值；

判定值确定子装置，用于基于所述获得的语音输入信号，采用下述公式对所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，其中G值的计算公式为：

$$G_i = \frac{1}{\sqrt{2\times\pi}} exp\left(\frac{\frac{|a_i - a|}{\sum_{j=1}^{N}|a_j - a|} * \frac{\left|a_i - \frac{1}{N} * \sum_{j=1}^{N} a_j\right|}{\sqrt{\frac{1}{N} * \sum_{k=1}^{N}\left(a_k - \frac{1}{N} * \sum_{j=1}^{N} a_j\right)^2}}}{\sqrt{\frac{\sum_{i=1}^{N}\left(a_i - \frac{1}{N} * \sum_{j=1}^{N} a_j\right)^2}{N-1}}} \right)$$

其中，a_k信号数据集合M中的第k个信号数据；a_i代表信号据集合M中的第i个信号数据，a_j代表信号数据集合M中的第j个信号数据，i＝1,2,3...N,j＝1,2,3...N;G_i代表信号数据集合M中第i个信号数据的噪声判定值，π代表自然常数，exp代表指数函数，a代表信号数据集合M中信号数据的中值；

判定子装置，用于将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。

一种基于人工智能的自动语音识别方法及系统

技术领域

[0001] 本发明涉及人工智能技术领域，具体涉及一种基于人工智能的自动语音识别方法及系统。

背景技术

[0002] 自动语音识别技术(Automatic Speech Recognition，简称“ASR”)是一种将人的语音转换为文本的技术。语音识别是一个多学科交叉的领域，它与声学、语音学、语言学、数字信号处理理论、信息论、计算机科学等众多学科紧密相连。由于语音信号的多样性和复杂性，语音识别系统只能在一定的限制条件下获得满意的性能，或者说只能应用于某些特定的场合。

[0003] 自动语音识别技术的目标是让计算机能够“听写”出不同人所说出的连续语音，也就是俗称的“语音听写机”，是实现“声音”到“文字”转换的技术。

[0004] 但是，现有的自动语音识别技术应用在包含专业词汇的语音识别中存在一定的问题，由于专业词汇的特殊性及应用专业词汇的人员的特定性，具有相应专业领域知识的人员可能辨识某些词汇的含义，因此，通过普通的自动语音识别技术可能存在识别专业词汇不准确的情况，或者针对专业词汇的识别效率低的情况。

[0005] 因此，亟需一种自动语音识别方法可以解决上述技术问题。

发明内容

[0006] 本发明提供一种基于人工智能的自动语音识别方法及系统，用以解决现有技术中存在的识别专业词汇不准确，以及专业词汇的识别效率低的问题。

[0007] 本发明提供一种基于人工智能的自动语音识别方法，该方法包括：

[0008] 接收待识别的语音信号；

[0009] 对所述待识别的语音信号进行预处理，获得语音输入信号；

[0010] 将所述语音输入信号进行时域到频域的转换，提取语音特征参数；

[0011] 对所述语音特征参数进行随机取样，获得若干个样本特征参数；

[0012] 将所述样本特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果；

[0013] 将所述识别结果输入至词汇分类模板，将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对，获得识别结果中的词汇中专业词汇的占比；

[0014] 判断所述占比是否超出预设值，若是，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率；

[0015] 若否，将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

[0016] 可选的，在所述并输出对应的文本之后，包括：

[0017] 将输出的所述文本输入至拼写纠错模型，获得纠错后文本；

[0018] 将纠错后的文本作为最终文本输出。

[0019] 可选的，所述词汇分类模板构建方法包括：

[0020] 获取大量分属于不同行业的专业词汇；

[0021] 将所述专业词汇采用卷积神经网络按照专业词汇所属的行业进行分类训练；

[0022] 获得分类结果，并将所述分类结果存储于分类数据库中，构成词汇分类模板。

[0023] 可选的，所述专业词汇声学模型构建方法包括：

[0024] 将词汇分类模板中的分类数据库设置为专业词汇字典；

[0025] 基于音素或其组合优先从所述专业词汇字典中进行映射；

[0026] 若所述专业词汇字典中无映射内容，则基于声学模型中的字典进行映射；

[0027] 根据上述映射结果获取相应音素及其组合的声学得分。

[0028] 可选的，所述专业词汇语言模型的构建方法包括：

[0029] 基于词汇分类模板中分类数据库中存储的专业词汇，结合词典获取专业词汇的词序及连接词；所述词序及连接词的概率值排序为前五位；

[0030] 将所述获取到的词序及连接词结合专业词汇记录在专业词汇语言数据库中；

[0031] 基于所述声学得分及所述专业词汇语言数据库，确定语言得分。

[0032] 可选的，所述对所述语音信号进行预处理的方法包括：

[0033] A1，获取环境中有规律噪声的频谱；

[0034] A2，获取收音装置噪声的频谱；

[0035] A3，基于环境噪声的频谱和收音装置噪声的频谱，结合最小方差无畸变响应滤波器增强后的信号采用下述公式确定：

[0036]
$$S(f,t,n)=\sum_{i=1}^{P}w_i(f)\left|\sum(R_i*x_i(f,t,n))+\varepsilon\right|^2$$
$$+\sum_{t=1}^{T}\sum_{i=1}^{P}\frac{1}{T}[N_T(f,t,n)+N_i(f,t,n)]^2w_i(f)$$

[0037] 其中，$N_T(f,t,n)$为环境中有规律噪声的频谱；$N_i(f,t,n)$为收音装置噪声的频谱；$Y_i(f,t,n)$为包含噪声的语音信号；$w_i(f)$为滤波器的加权系数；$S(f,t,n)$是获得的语音输入信号；$x_i(f,t,n)$为待降噪的信号；

[0038] f为当前频率，t为当前时间，n为当前帧，P为收音装置的数量，i＝1,2...P,T为有规律噪声出现的次数，t＝1,2...T。R_i是训练误差取最小值时对应的初始系数，ε表示训练误差的最小值；

[0039] A4，基于所述获得的语音输入信号，采用下述公式对所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，其中G值的计算公式为：

[0040] $$G_i=\frac{1}{\sqrt{2\times\pi}}exp\left(\frac{\frac{|a_i-a|}{\sum_{j=1}^{N}|a_j-a|}*\frac{\left|a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right|}{\sqrt{\frac{1}{N}*\sum_{k=1}^{N}\left(a_k-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}}}{\sqrt{\frac{\sum_{i=1}^{N}\left(a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}{N-1}}}\right)$$

[0041] 其中，a_k信号数据集合M中的第k个信号数据；a_i代表信号据集合M中的第i个信号数据，a_j代表信号数据集合M中的第j个信号数据，i＝1,2,3...N,j＝1,2,3...N;G_i代表信号数据集合M中第i个信号数据的噪声判定值，π代表自然常数，exp代表指数函数，a代表信号数据集合M中信号数据的中值；

[0042] A5，将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。

[0043] 本发明提供一种基于人工智能的自动语音识别系统，该系统包括：

[0044] 接收装置，用于接收待识别的语音信号；

[0045] 预处理装置，用于对所述待识别的语音信号进行预处理，获得语音输入信号；

[0046] 提取装置，用于将所述语音输入信号进行时域到频域的转换，提取语音特征参数；

[0047] 抽样装置，用于对所述语音特征参数进行随机取样，获得若干个样本特征参数；

[0048] 结果获取装置，用于将所述样本特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果；

[0049] 专业词汇设置装置，用于将所述识别结果输入至词汇分类模板，将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对，获得识别结果中的词汇中专业词汇的占比；

[0050] 判断装置，用干判断所述占比是否超出预设值；

[0051] 第一输出装置，用于当判断装置的判断结果为是时，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率；

[0052] 第二输出装置，用于当判断装置的判断结果为否时，将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

[0053] 可选的，所述专业词汇设置装置中词汇分类模板包括：

[0054] 获取子装置，用于获取大量分属于不同行业的专业词汇；

[0055] 训练子装置，用于将所述专业词汇采用卷积神经网络按照专业词汇所属的行业进行分类训练；

[0056] 分类结果获取子装置，用于获得分类结果，并将所述分类结果存储于分类数据库中，构成词汇分类模板。

[0057] 可选的，所述第一输出装置中所述专业词汇声学模型包括：

[0058] 分类子装置，用于将词汇分类模板中的分类数据库设置为专业词汇字典；

[0059] 第一映射子装置，用于基于音素或其组合优先从所述专业词汇字典中进行映射；

[0060] 第二映射子装置，用于当第一映射子装置中所述专业词汇字典中无映射内容时，则基于声学模型中的字典进行映射；

[0061] 声学得分子装置，用于根据上述映射结果获取相应音素及其组合的声学得分。

[0062] 可选的，所述预处理装置包括：

[0063] 第一噪声频谱获取子装置，用于获取环境中有规律噪声的频谱；

[0064] 第二噪声频谱获取子装置，用于获取收音装置噪声的频谱；

[0065] 信号确定子装置，用于基于环境噪声的频谱和收音装置噪声的频谱，结合最小方差无畸变响应滤波器增强后的信号采用下述公式确定：

[0066]
$$S(f,t,n)=\sum_{i=1}^{P}w_i(f)\left|\sum(R_i*x_i(f,t,n))+\varepsilon\right|^2$$
$$+\sum_{t=1}^{T}\sum_{i=1}^{P}\frac{1}{T}[N_T(f,t,n)+N_i(f,t,n)]^2w_i(f)$$

[0067] 其中，$N_T(f,t,n)$为环境中有规律噪声的频谱；$N_i(f,t,n)$为收音装置噪声的频谱；$Y_i(f,t,n)$为包含噪声的语音信号；$w_i(f)$为滤波器的加权系数；$S(f,t,n)$是获得的语音输入信号；$x_i(f,t,n)$为待降噪的信号；

[0068] f为当前频率，t为当前时间，n为当前帧，P为收音装置的数量，i＝1,2...P,T为有规律噪声出现的次数，t＝1,2...T。R_i是训练误差取最小值时对应的初始系数，ε表示训练误差的最小值；

[0069] 判定值确定子装置，用于基于所述获得的语音输入信号，采用下述公式对所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，其中G值的计算公式为：

[0070]
$$G_i=\frac{1}{\sqrt{2\times\pi}}exp\left(\frac{\frac{|a_i-a|}{\sum_{j=1}^{N}|a_j-a|}*\frac{\left|a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right|}{\sqrt{\frac{1}{N}*\sum_{k=1}^{N}\left(a_k-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}}}{\sqrt{\frac{\sum_{i=1}^{N}\left(a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}{N-1}}}\right)$$

[0071] 其中，a_k信号数据集合M中的第k个信号数据；a_i代表信号据集合M中的第i个信号数据，a_j代表信号数据集合M中的第j个信号数据，i＝1,2,3...N,j＝1,2,3...N;G_i代表信号数据集合M中第i个信号数据的噪声判定值，π代表自然常数，exp代表指数函数，a代表信号数据集合M中信号数据的中值；

[0072] 判定子装置，用于将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。

[0073] 本发明提供一种基于人工智能的自动语音识别方法，采用本发明提供的方案可以提高对专业词汇识别的精确度和准确率，特别是增强专业领域中视频会议记录的准确性、精准性，特别是专业性，提高企业在相关专业领域的专业性，更重要的是减少因为对专业词

汇的自动识别语音识别造成的专业误解，防止因为语音识别造成误解进而造成重大损失。同时，由于以词汇分类模板做基础，提高专业词汇的搜索速率，进而提高了针对专业词汇的自动语音的识别效率。

[0074] 本发明的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本发明而了解。本发明的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

[0075] 下面通过附图和实施例，对本发明的技术方案做进一步的详细描述。

附图说明

[0076] 附图用来提供对本发明的进一步理解，并且构成说明书的一部分，与本发明的实施例一起用于解释本发明，并不构成对本发明的限制。在附图中：

[0077] 图1为本发明实施例中一种基于人工智能的自动语音识别方法的流程图；

[0078] 图2为本发明实施例中一种基于人工智能的自动语音识别系统的结构示意图。

具体实施方式

[0079] 以下结合附图对本发明的优选实施例进行说明，应当理解，此处所描述的优选实施例仅用于说明和解释本发明，并不用于限定本发明。

[0080] 实施例1：

[0081] 本发明实施例提供了一种基于人工智能的自动语音识别方法，图1为本发明实施例中一种基于人工智能的自动语音识别方法的流程图，请参照图1，该方法包括以下步骤：

[0082] 步骤S101，接收待识别的语音信号；

[0083] 步骤S102，对所述待识别的语音信号进行预处理，获得语音输入信号；

[0084] 步骤S103，将所述语音输入信号进行时域到频域的转换，提取语音特征参数；

[0085] 步骤S104，对所述语音特征参数进行随机取样，获得若干个样本特征参数；

[0086] 步骤S105，将所述样本特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果；

[0087] 步骤S106，将所述识别结果输入至词汇分类模板，将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对，获得识别结果中的词汇中专业词汇的占比；

[0088] 步骤S107，判断所述占比是否超出预设值；若判断结果为是，则执行步骤S108，若判断结果为否，则执行步骤S109。

[0089] 步骤S108，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率；

[0090] 步骤S109，将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

[0091] 上述技术方案的工作原理为：本实施例采用的方案是通过对待识别语音信号提取的语音特征参数进行随机取样，并对采样的参数基于声学模型和语言模型获取识别结果，再次对所述识别结果基于词汇分类模板判断是否属于涉及专业词汇的语音识别，若是，则说明待识别的语音信号是与专业方面相关的语音，而对此类的语音的识别需要相对专业的

词汇库提供基础支持。因此，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率。而对于判断不属于专业词汇的语音识别时，则进行普通的自动语音识别技术，即将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

[0092] 需要说明的是，将所述语音输入信号进行时域到频域的转换，提取语音特征参数，可采用的方式包括梅尔频率倒谱的方式提取语音特征，通过梅尔频率倒谱获得声谱，然后将声谱通过滤波器进行滤波处理。

[0093] 另外，所述语音特征的提取还可以采用深度卷积神经网络的原理进行语音特征的提取，获得语音特征参数。

[0094] 此外，对词汇分类模板进行简单的介绍和说明，所述词汇分类模板中包含有专业词汇数据库，并且不同行业的专业词汇也不同，因此，可以按照行业的不同分别设置不同类别的专业词汇，根据所需专业词汇分类的不同，可以从不同的分类的数据库中进行相关词汇的搜索。

[0095] 本实施例提供的方案可以应用的范围很广，例如涉及到专业度高的行业会议的会议记录，某些产品的现场展示视频中语音的自动识别等，如果涉及到专业领域的视频或录音等需要自动语音识别的场合都可采用本实施例提供的方案。

[0096] 上述技术方案的有益效果为：采用本实施例提供的方案可以提高对专业词汇识别的精确度和准确率，特别是增强专业领域中视频会议记录的准确性、精准性，特别是专业性，提高企业在相关专业领域的专业性，更重要的是减少因为对专业词汇的自动识别语音识别造成的专业误解，防止因为语音识别造成误解进而造成重大损失。同时，由于以词汇分类模板做基础，提高专业词汇的搜索速率，进而提高了针对专业词汇的自动语音的识别效率。

[0097] 实施例2：

[0098] 在实施例1的基础上，在所述并输出对应的文本之后，包括：

[0099] 将输出的所述文本输入至拼写纠错模型，获得纠错后文本；

[0100] 将纠错后的文本作为最终文本输出。

[0101] 上述技术方案的工作原理及有益效果为：本实施例采用的方案是对输入的文本进行拼写纠错的过程，通过声学模型和语言模型之后，其输出的文本可能会存在拼写的错误等形式的问题，为了保证自动语音识别的精准性和专业性，需要后续对输出的文本的拼写进行纠错，通过设置拼写纠错模型保证输出的最终文本没有形式的拼写错误，提高自动语音识别的精准度。

[0102] 实施例3：

[0103] 在实施例1的基础上，所述词汇分类模板构建方法包括：

[0104] 获取大量分属于不同行业的专业词汇；

[0105] 将所述专业词汇采用卷积神经网络按照专业词汇所属的行业进行分类训练；

[0106] 获得分类结果，并将所述分类结果存储于分类数据库中，构成词汇分类模板。

[0107] 上述技术方案的工作原理为：本实施例采用的方案是对词汇分类模板的构建方法的描述。通过获取大量的不同行业的专业词汇，采用卷积神经网络对上述专业词汇按照以

行业为基准进行分类训练，也就是不同的行业包含的专业词汇是不同的，通过词汇分类模板将不同专业词汇进行分类，并将分类结果存储于分类数据库中，便于后续过程中对相应的专业词汇进行查询。

[0108] 上述技术方案的有益效果为：采用本实施例提供的方案将专业词汇进行汇总并分类，由于以词汇分类模板做基础，提高专业词汇的搜索速率，进而提高了针对专业词汇的自动语音的识别效率。另外，采用本实施例提供的方案可以提高对专业词汇识别的精确度和准确率，特别是增强专业领域中视频会议记录的准确性、精准性，特别是专业性，提高企业在相关专业领域的专业性，更重要的是减少因为对专业词汇的自动识别语音识别造成的专业误解，防止因为语音识别造成误解进而造成重大损失。

[0109] 实施例4：

[0110] 在实施例3的基础上，所述专业词汇声学模型构建方法包括：

[0111] 将词汇分类模板中的分类数据库设置为专业词汇字典；

[0112] 基于音素或其组合优先从所述专业词汇字典中进行映射；

[0113] 若所述专业词汇字典中无映射内容，则基于声学模型中的字典进行映射；

[0114] 根据上述映射结果获取相应音素及其组合的声学得分。

[0115] 上述技术方案的工作原理为：本实施例采用的方案是对专业词汇声学模型的构建方法的描述。通过将词汇分类模板中的分类数据库设置为专业词汇字典，结合声学模型中的字典共同对待识别语音信号的拆解出的因素或其契合进行映射，其映射顺序为先从所述专业词汇字典中进行映射，当所述专业词汇字典中没有映射内容，则基于声学模型中的字典进行映射，综上所述映射以获得相应音素及其组合的声学得分，进而构成所述专业词汇声学模型。

[0116] 上述技术方案的有益效果为：采用本实施例提供的方案构建的专业词汇声学模型，优先采用词汇分了模板中的分类数据库作为字典进行搜索，当分类数据库有对应的专业词汇时，直接从该数据库读取映射，当该分类数据库没有相应的词汇时，则采用声学模型中的字典进行词汇映射。因此，采用本实施例的方案一方面提高专业词汇的识别准确性，另一方面，由于以分类数据库中的专业词汇为基础，采用本实施例方案也可以提高专业词汇的识别效率。

[0117] 实施例5：

[0118] 在实施例3的基础上，所述专业词汇语言模型的构建方法包括：

[0119] 基于词汇分类模板中分类数据库中存储的专业词汇，结合词典获取专业词汇的词序及连接词；所述词序及连接词的概率值排序为前五位；

[0120] 将所述获取到的词序及连接词结合专业词汇记录在专业词汇语言数据库中；

[0121] 基于所述声学得分及所述专业词汇语言数据库，确定语言得分。

[0122] 上述技术方案的工作原理及工作原理为：本实施例采用的方案是构建专业词汇语言模型的方法，基于词汇分类模板中分类数据库中存储的专业词汇，结合词典获取专业词汇的词序及连接词，并将专业词汇的词序及连接词按照概率值由高到低进行排序，并提取概率值排序前五位的词序及连接词，将排序后的词序及连接词结合专业词汇记录在专业词汇语言数据库中，最后再结合声学得分及专业词汇语言数据库，确定语言得分，最终通过本实施例构建成所述专业词汇语言模型。

[0123] 因此，采用本实施例的方案一方面提高专业词汇的识别准确性，另一方面，由于以分类数据库中的专业词汇为基础，采用本实施例方案也可以提高专业词汇的识别效率。

[0124] 实施例6：

[0125] 在实施例1的基础上，所述对所述待识别的语音信号进行预处理的方法包括：

[0126] A1，获取环境中有规律噪声的频谱；

[0127] A2，获取收音装置噪声的频谱；

[0128] A3，基于环境噪声的频谱和收音装置噪声的频谱，结合最小方差无畸变响应滤波器增强后的信号采用下述公式确定：

[0129]
$$S(f,t,n)=\sum_{i=1}^{P}w_i(f)\left|\sum(R_i*x_i(f,t,n))+\varepsilon\right|^2$$
$$+\sum_{t=1}^{T}\sum_{i=1}^{P}\frac{1}{T}[N_T(f,t,n)+N_i(f,t,n)]^2w_i(f)$$

[0130] 其中，$N_T(f,t,n)$为环境中有规律噪声的频谱；$N_i(f,t,n)$为收音装置噪声的频谱；$Y_i(f,t,n)$为包含噪声的语音信号；$w_i(f)$为滤波器的加权系数；$S(f,t,n)$是获得的语音输入信号；$x_i(f,t,n)$为待降噪的信号；

[0131] f为当前频率，t为当前时间，n为当前帧，P为收音装置的数量，i＝1,2...P，T为有规律噪声出现的次数，t＝1,2...T。R_i是训练误差取最小值时对应的初始系数，ε表示训练误差的最小值；

[0132] A4，基于所述获得的语音输入信号，采用下述公式对所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，其中G值的计算公式为：

[0133]
$$G_i=\frac{1}{\sqrt{2\times\pi}}exp\left(\frac{\frac{|a_i-a|}{\sum_{j=1}^{N}|a_j-a|}*\frac{\left|a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right|}{\sqrt{\frac{1}{N}*\sum_{k=1}^{N}\left(a_k-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}}}{\sqrt{\frac{\sum_{i=1}^{N}\left(a_i-\frac{1}{N}*\sum_{j=1}^{N}a_j\right)^2}{N-1}}}\right)$$

[0134] 其中，a_k信号数据集合M中的第k个信号数据；a_i代表信号据集合M中的第i个信号数据，a_j代表信号数据集合M中的第j个信号数据，i＝1,2,3...N，j＝1,2,3...N；G_i代表信号数据集合M中第i个信号数据的噪声判定值，π代表自然常数，exp代表指数函数，a代表信号数据集合M中信号数据的中值；

[0135] A5，将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。

[0136] 上述技术方案的工作原理及有益效果为：本实施例采用的方案是对所述待识别的语音信号进行预处理的方法。该方法是对环境中有规律噪声以及收音装置的噪声进行降噪处理的过程。首先有规律噪声例如需要对视频会议做语音识别的会议记录，当视频会议过

程中，需要在电脑上敲击键盘，敲击键盘造成的噪声，以及为了提醒会议进度设置的定时计时铃声等，这些均属于有规律的噪声。另外，针对视频会议中麦克风装置本身收声时产生的噪声等也需要进行噪声的消除。由于上述两类噪声出现的频率较高，因此，将上述两类造成作为降噪的主动对象。采用本实施例提供的公式进行降噪处理，可以最大程度的降低上述两类噪声对输入的语音信号的干扰，通过降噪处理，以提高待识别的语音信号的质量，保证自动语音识别的准确性。

[0137] 另外，通过设置所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。进一步通过噪声判定值的方式减少噪声点，提高所述语音输入信号的质量，进而提供高自动语音识别的准确性。

[0138] 实施例7：

[0139] 本实施例提供一种基于人工智能的自动语音识别系统，图2为本发明实施例中一种基于人工智能的自动语音识别系统的结构示意图，请参照图2，该系统包括以下装置：

[0140] 接收装置201，用于接收待识别的语音信号；

[0141] 预处理装置202，用于对所述待识别的语音信号进行预处理，获得语音输入信号；

[0142] 提取装置203，用于将所述语音输入信号进行时域到频域的转换，提取语音特征参数；

[0143] 抽样装置204，用于对所述语音特征参数进行随机取样，获得若干个样本特征参数；

[0144] 结果获取装置205，用于将所述样本特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果；

[0145] 专业词汇设置装置206，用于将所述识别结果输入至词汇分类模板，将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对，获得识别结果中的词汇中专业词汇的占比；

[0146] 判断装置207，用于判断所述占比是否超出预设值；

[0147] 第一输出装置208，用于当判断装置的判断结果为是时，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率；

[0148] 第二输出装置209，用于当判断装置的判断结果为否时，将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

[0149] 上述技术方案的工作原理为：本实施例采用的方案是通过对待识别语音信号提取的语音特征参数进行随机取样，并对采样的参数基于声学模型和语言模型获取识别结果，再次对所述识别结果基于词汇分类模板判断是否属于涉及专业词汇的语音识别，若是，则说明待识别的语音信号是与专业方面相关的语音，而对此类的语音的识别需要相对专业的词汇库提供基础支持。因此，将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本；所述专业词汇声学模型和专业词汇语言模型中对专业词汇的权重进行了重新匹配，提高获得专业词汇的概率。而

对于判断不属于专业词汇的语音识别时，则进行普通的自动语音识别技术，即将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本。

[0150] 需要说明的是，将所述语音输入信号进行时域到频域的转换，提取语音特征参数，可采用的方式包括梅尔频率倒谱的方式提取语音特征，通过梅尔频率倒谱获得声谱，然后将声谱通过滤波器进行滤波处理。

[0151] 另外，所述语音特征的提取还可以采用深度卷积神经网络的原理进行语音特征的提取，获得语音特征参数。

[0152] 此外，对词汇分类模板进行简单的介绍和说明，所述词汇分类模板中包含有专业词汇数据库，并且不同行业的专业词汇也不同，因此，可以按照行业的不同分别设置不同类别的专业词汇，根据所需专业词汇分类的不同，可以从不同的分类的数据库中进行相关词汇的搜索。

[0153] 本实施例提供的方案可以应用的范围很广，例如涉及到专业度高的行业会议的会议记录，某些产品的现场展示视频中语音的自动识别等，如果涉及到专业领域的视频或录音等需要自动语音识别的场合都可采用本实施例提供的方案。

[0154] 上述技术方案的有益效果为：采用本实施例提供的方案可以提高对专业词汇识别的精确度和准确率，特别是增强专业领域中视频会议记录的准确性、精准性，特别是专业性，提高企业在相关专业领域的专业性，更重要的是减少因为对专业词汇的自动识别语音识别造成的专业误解，防止因为语音识别造成误解进而造成重大损失。同时，由于以词汇分类模板做基础，提高专业词汇的搜索速率，进而提高了针对专业词汇的自动语音的识别效率。

[0155] 实施例8：

[0156] 在实施例7的基础上，所述专业词汇设置装置中词汇分类模板包括：

[0157] 获取子装置，用于获取大量分属于不同行业的专业词汇；

[0158] 训练子装置，用于将所述专业词汇采用卷积神经网络按照专业词汇所属的行业进行分类训练；

[0159] 分类结果获取子装置，用于获得分类结果，并将所述分类结果存储于分类数据库中，构成词汇分类模板。

[0160] 上述技术方案的工作原理为：本实施例采用的方案是对词汇分类模板的构建方法的描述。通过获取大量的不同行业的专业词汇，采用卷积神经网络对上述专业词汇按照以行业为基准进行分类训练，也就是不同的行业包含的专业词汇是不同的，通过词汇分类模板将不同专业词汇进行分类，并将分类结果存储于分类数据库中，便于后续过程中对相应的专业词汇进行查询。

[0161] 上述技术方案的有益效果为：采用本实施例提供的方案将专业词汇进行汇总并分类，由于以词汇分类模板做基础，提高专业词汇的搜索速率，进而提高了针对专业词汇的自动语音的识别效率。另外，采用本实施例提供的方案可以提高对专业词汇识别的精确度和准确率，特别是增强专业领域中视频会议记录的准确性、精准性，特别是专业性，提高企业在相关专业领域的专业性，更重要的是减少因为对专业词汇的自动识别语音识别造成的专业误解，防止因为语音识别造成误解进而造成重大损失。

[0162] 实施例9：

[0163] 在实施例7的基础上，所述第一输出装置中所述专业词汇声学模型包括：

[0164] 分类子装置，用于将词汇分类模板中的分类数据库设置为专业词汇字典；

[0165] 第一映射子装置，用于基于音素或其组合优先从所述专业词汇字典中进行映射；

[0166] 第二映射子装置，用于当第一映射子装置中所述专业词汇字典中无映射内容时，则基于声学模型中的字典进行映射；

[0167] 声学得分子装置，用于根据上述映射获结果取相应音素及其组合的声学得分。

[0168] 上述技术方案的工作原理为：本实施例采用的方案是对专业词汇声学模型的构建方法的描述。通过将词汇分类模板中的分类数据库设置为专业词汇字典，结合声学模型中的字典共同对待识别语音信号的拆解出的因素或其契合进行映射，其映射顺序为先从所述专业词汇字典中进行映射，当所述专业词汇字典中没有映射内容，则基于声学模型中的字典进行映射，综上所述映射以获得相应音素及其组合的声学得分，进而构成所述专业词汇声学模型。

[0169] 上述技术方案的有益效果为：采用本实施例提供的方案构建的专业词汇声学模型，优先采用词汇分了模板中的分类数据库作为字典进行搜索，当分类数据库有对应的专业词汇时，直接从该数据库读取映射，当该分类数据库没有相应的词汇时，则采用声学模型中的字典进行词汇映射。因此，采用本实施例的方案一方面提高专业词汇的识别准确性，另一方面，由于以分类数据库中的专业词汇为基础，采用本实施例方案也可以提高专业词汇的识别效率。

[0170] 实施例10：

[0171] 在实施例1的基础上，所述预处理装置包括：

[0172] 第一噪声频谱获取子装置，用于获取环境中有规律噪声的频谱；

[0173] 第二噪声频谱获取子装置，用于获取收音装置噪声的频谱；

[0174] 信号确定子装置，用于基于环境噪声的频谱和收音装置噪声的频谱，结合最小方差无畸变响应滤波器增强后的信号采用下述公式确定：

[0175]
$$S(f,t,n)=\sum_{i=1}^{P}w_i(f)\left|\sum(R_i*x_i(f,t,n))+\varepsilon\right|^2$$
$$+\sum_{t=1}^{T}\sum_{i=1}^{P}\frac{1}{T}[N_T(f,t,n)+N_i(f,t,n)]^2w_i(f)$$

[0176] 其中，$N_T(f,t,n)$为环境中有规律噪声的频谱；$N_i(f,t,n)$为收音装置噪声的频谱；$Y_i(f,t,n)$为包含噪声的语音信号；$w_i(f)$为滤波器的加权系数；$S(f,t,n)$是获得的语音输入信号；$x_i(f,t,n)$为待降噪的信号；

[0177] f为当前频率，t为当前时间，n为当前帧，P为收音装置的数量，i＝1,2...P，T为有规律噪声出现的次数，t＝1,2...T。R_i是训练误差取最小值时对应的初始系数，ε表示训练误差的最小值；

[0178] 判定值确定子装置，用于基于所述获得的语音输入信号，采用下述公式对所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，其中G值的计算公式为：

[0179] $$G_i = \frac{1}{\sqrt{2\times\pi}} exp\left(\frac{\frac{|a_i - a|}{\sum_{j=1}^{N}|a_j - a|} * \frac{\left|a_i - \frac{1}{N} * \sum_{j=1}^{N} a_j\right|}{\sqrt{\frac{1}{N} * \sum_{k=1}^{N}\left(a_k - \frac{1}{N} * \sum_{j=1}^{N} a_j\right)^2}}}{\sqrt{\frac{\sum_{i=1}^{N}\left(a_i - \frac{1}{N} * \sum_{j=1}^{N} a_j\right)^2}{N-1}}} \right)$$

[0180] 其中，a_k信号数据集合M中的第k个信号数据；a_i代表信号据集合M中的第i个信号数据，a_j代表信号数据集合M中的第j个信号数据，i＝1,2,3...N,j＝1,2,3...N;G_i代表信号数据集合M中第i个信号数据的噪声判定值，π代表自然常数，exp代表指数函数，a代表信号数据集合M中信号数据的中值；

[0181] 判定子装置，用于将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。

[0182] 上述技术方案的工作原理及有益效果为：本实施例采用的方案是对所述待识别的语音信号进行预处理的方法。该方法是对环境中有规律噪声以及收音装置的噪声进行降噪处理的过程。首先有规律噪声例如需要对视频会议做语音识别的会议记录，当视频会议过程中，需要在电脑上敲击键盘，敲击键盘造成的噪声，以及为了提醒会议进度设置的定时计时铃声等，这些均属于有规律的噪声。另外，针对视频会议中麦克风装置本身收声时产生的噪声等也需要进行噪声的消除。由于上述两类噪声出现的频率较高，因此，将上述两类造成作为降噪的主动对象。采用本实施例提供的公式进行降噪处理，可以最大程度的降低上述两类噪声对输入的语音信号的干扰，通过降噪处理，以提高待识别的语音信号的质量，保证自动语音识别的准确性。

[0183] 另外，通过设置所述语音输入信号的噪声判定值，若每个信号数据的噪声判定值G大于预设的判定阈值，则将该信号数据判定为噪声点，将数据集合M中的每个信号数据都进行一一判定，当为噪声点时，则进行剔除，不为噪声点时，则进行保留，将保留后的信号数据形成最后处理后的信号。进一步通过噪声判定值的方式减少噪声点，提高所述语音输入信号的质量，进而提供高自动语音识别的准确性。

[0184] 显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

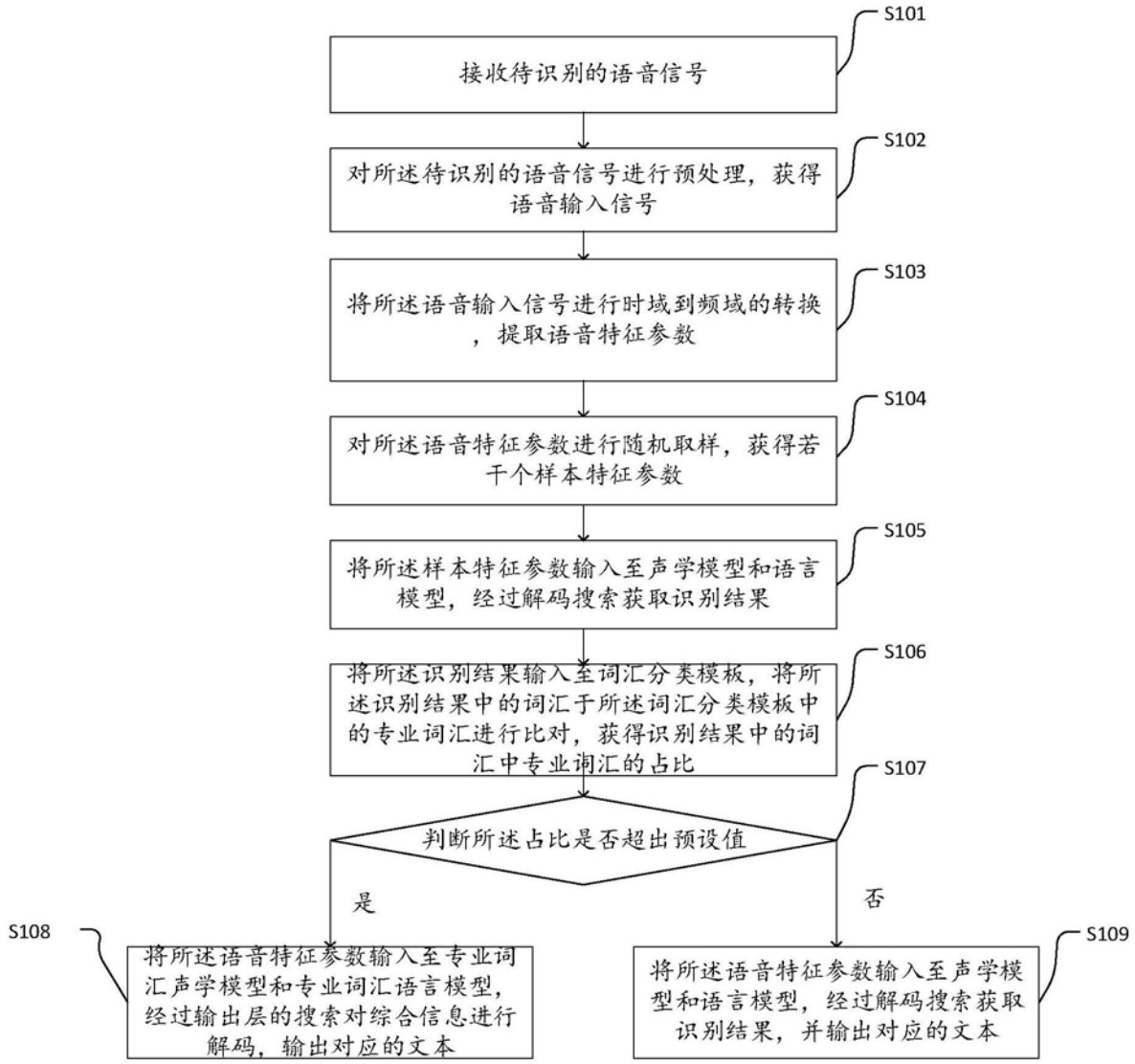

S101
接收待识别的语音信号
S102
对所述待识别的语音信号进行预处理，获得语音输入信号
S103
将所述语音输入信号进行时域到频域的转换，提取语音特征参数
S104
对所述语音特征参数进行随机取样，获得若干个样本特征参数
S105
将所述样本特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果
S106
将所述识别结果输入至词汇分类模板，将所述识别结果中的词汇于所述词汇分类模板中的专业词汇进行比对，获得识别结果中的词汇中专业词汇的占比
S107
判断所述占比是否超出预设值
是
否
S108
将所述语音特征参数输入至专业词汇声学模型和专业词汇语言模型，经过输出层的搜索对综合信息进行解码，输出对应的文本
S109
将所述语音特征参数输入至声学模型和语言模型，经过解码搜索获取识别结果，并输出对应的文本

图1

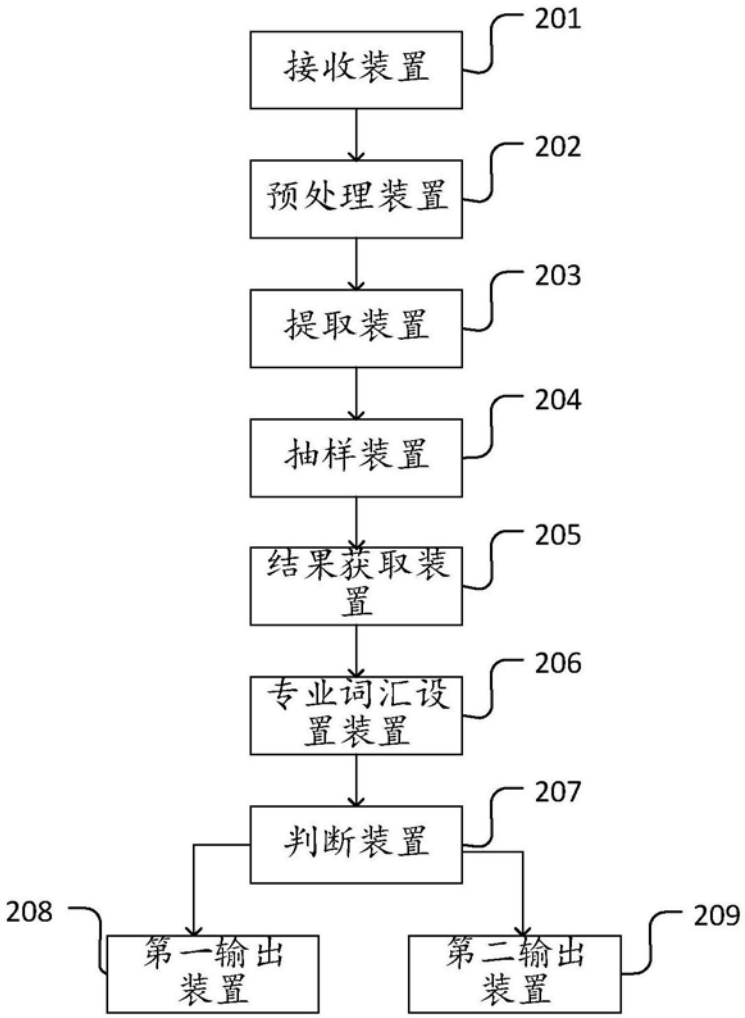

201
接收装置
202
预处理装置
203
提取装置
204
抽样装置
205
结果获取装置
206
专业词汇设置装置
207
判断装置
208
第一输出装置
209
第二输出装置

图2