

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 2 月 13 日 (13.02.2020)



(10) 国际公布号
WO 2020/029404 A1

(51) 国际专利分类号:

G10L 15/02 (2006.01) *G10L 25/18* (2013.01)
G10L 15/14 (2006.01) *G10L 25/24* (2013.01)
G10L 15/26 (2006.01) *G10L 25/51* (2013.01)
G10L 19/02 (2013.01)

(21) 国际申请号: PCT/CN2018/108190

(22) 国际申请日: 2018 年 9 月 28 日 (28.09.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

201810897646.2 2018 年 8 月 8 日 (08.08.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

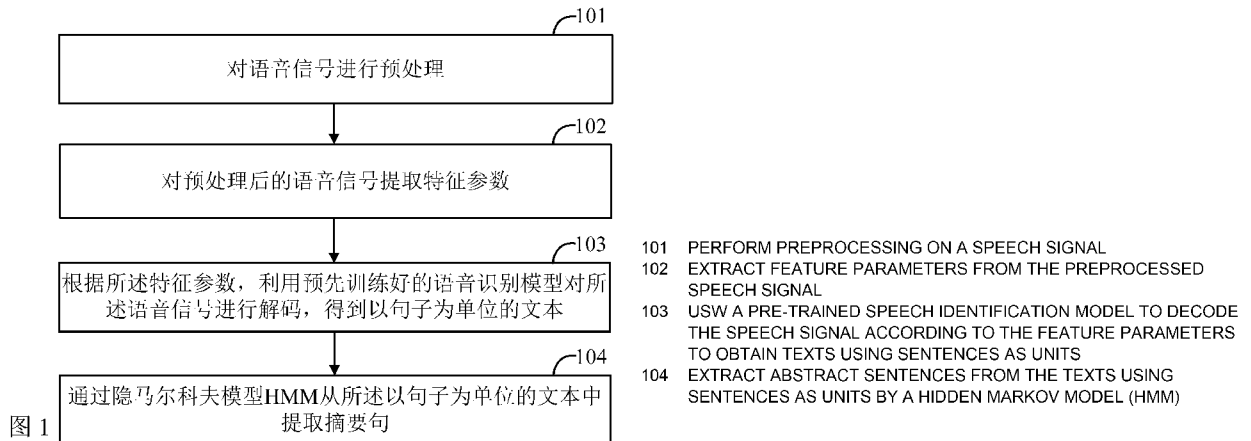
(72) 发明人: 王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。王珏(WANG, Jue); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市赛恩倍吉知识产权代理有限公司(SHENZHEN SCIENBIZIP INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市龙华新区龙观东路 83 号荣群大厦 9 楼, Guangdong 518109 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

(54) Title: SPEECH PROCESSING METHOD AND DEVICE, COMPUTER DEVICE AND READABLE STORAGE MEDIUM

(54) 发明名称: 语音处理方法及装置、计算机装置及可读存储介质



(57) Abstract: Disclosed is a speech processing method. The method comprises: performing preprocessing on a speech signal; extracting feature parameters from the preprocessed speech signal; using a pre-trained speech identification model to decode the speech signal according to the feature parameters to obtain texts using sentences as units; extracting abstract sentences from the texts using sentences as units by a Hidden Markov Model (HMM). The present application further provides a speech processing device, a computer device and a nonvolatile readable storage medium. According to the present application, the speech can be identified and useless information can be removed from the speech identification result.

(57) 摘要: 一种语音处理方法, 所述方法包括: 对语音信号进行预处理; 对预处理后的语音信号提取特征参数; 根据所述特征参数, 利用预先训练好的语音识别模型对所述语音信号进行解码, 得到以句子为单位的文本; 通过隐马尔科夫模型HMM从所述以句子为单位的文本中提取摘要句。本申请还提供一种语音处理装置、计算机装置及非易失性可读存储介质。本申请可以对语音进行识别, 并从语音识别结果中去除无用信息。

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

语音处理方法及装置、计算机装置及可读存储介质

本申请要求于 2018 年 08 月 8 日提交中国专利局, 申请号为 201810897646.2 发明名称为“语音处理方法及装置、计算机装置及可读存储介质”的中国专利申请的优先权, 其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机听觉技术领域, 具体涉及一种语音处理方法及装置、计算机装置和非易失性可读存储介质。

背景技术

在智能会议系统中, 语音识别技术是一项关键技术, 其可以将人的说话信号转换为可被计算机识别的文字信息作为输出。

然而, 现有的智能会议系统只是实现语音到文字的转换, 而不能对识别出来的文字信息做进一步的处理, 根据语音直接转换得到的文字信息可能会包含无用的信息, 例如一些与会议内容无关的句子。

发明内容

鉴于以上内容, 有必要提出一种语音处理方法及装置、计算机装置和非易失性可读存储介质, 其可以对语音进行识别, 并从语音识别结果中去除无用信息。

本申请的第一方面提供一种语音处理方法, 所述方法包括:

对语音信号进行预处理;

对预处理后的语音信号提取特征参数;

根据所述特征参数, 利用预先训练好的语音识别模型对所述语音信号进行解码, 得到以句子为单位的文本;

通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

本申请的第二方面提供一种语音处理装置, 所述装置包括:

预处理单元, 用于对语音信号进行预处理;

特征提取单元, 用于对预处理后的语音信号提取特征参数;

解码单元, 用于根据所述特征参数, 利用预先训练好的语音识别模型对所述语音信号进行解码, 得到以句子为单位的文本;

摘要提取单元, 用于通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

本申请的第三方面提供一种计算机装置, 所述计算机装置包括处理器, 所述处理器用于执行存储器中存储的计算机可读指令时实现所述语音处理方

法。

本申请的第四方面提供一种非易失性可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现所述语音处理方法。

本申请对语音信号进行预处理；对预处理后的语音信号提取特征参数；根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。本申请不仅将语音信息转化为文字，还提取文字中的摘要句进行输出，剔除了由语音识别结果中的无用信息，获得更好的语音处理结果。

附图说明

图 1 是本申请实施例提供的语音处理方法的流程图。

图 2 是本申请实施例提供的语音处理装置的结构图。

图 3 是本申请实施例提供的计算机装置的示意图。

具体实施方式

为了能够更清楚地理解本申请的上述目的、特征和优点，下面结合附图和具体实施例对本申请进行详细描述。需要说明的是，在不冲突的情况下，本申请的实施例及实施例中的特征可以相互组合。

在下面的描述中阐述了很多具体细节以便于充分理解本申请，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中在本申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在于限制本申请。

优选地，本申请的语音处理方法应用在一个或者多个计算机装置中。所述计算机装置是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程门阵列(Field Programmable Gate Array, FPGA)、数字处理器(Digital Signal Processor, DSP)、嵌入式设备等。

所述计算机装置可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机装置可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

实施例一

图 1 是本申请实施例一提供的语音处理方法的流程图。所述语音处理方法应用于计算机装置。所述语音处理方法从语音信号中识别出以句子为单位的文本，从以句子为单位的文本中提取出摘要句。

如图 1 所示，所述语音处理方法具体包括以下步骤：

步骤 101, 对语音信号进行预处理。

所述语音信号可以是模拟语音信号, 也可以是数字语音信号。若所述语音信号是模拟语音信号, 则将所述模拟语音信号进行模数变换, 转换为数字语音信号。

本申请用于连续语音识别, 即对连续的音频流进行处理。在本申请的一个实施例中, 所述语音处理方法应用于智能会议系统中, 所述语音信号是通过语音输入设备(例如麦克风、手机话筒等)输入到智能会议系统的发言者的语音信号。

对所述语音信号进行预处理可以包括对所述语音信号进行预加重。

预加重的目的是提升语音的高频分量, 使信号的频谱变得平坦。语音信号由于受声门激励和口鼻辐射的影响, 能量在高频端明显减小, 通常是频率越高幅值越小。当频率提升两倍时, 功率谱幅度按 6dB/oct 跌落。因此, 在对语音信号进行频谱分析或声道参数分析前, 需要对语音信号的高频部分进行频率提升, 即对语音信号进行预加重。预加重一般利用高通滤波器实现, 高通滤波器的传递函数可以为:

$$H(z) = 1 - \kappa z^{-1}, \quad 0.9 \leq \kappa \leq 1.0.$$

其中, κ 为预加重系数, 优选取值在 0.94-0.97 之间。

对所述语音信号进行预处理还可以包括对所述语音信号进行加窗分帧。

语音信号是一种非平稳的时变信号, 主要分为浊音和清音两大类。浊音的基音周期、请浊音信号幅度和声道参数等都随时间而缓慢变化, 但通常在 10ms-30ms 的时间内可以认为具有短时平稳性。语音信号处理中可以把语音信号分成一些短段(即获得短时平稳信号)来进行处理, 这个过程称为分帧, 得到的短段的语音信号称为语音帧。分帧是通过对语音信号进行加窗处理来实现的。为了避免相邻两帧的变化幅度过大, 帧与帧之间需要重叠一部分。在本申请的一个实施例中, 每个语音帧为 25 毫秒, 相邻两个语音帧之间存在 15 毫秒重叠, 也就是每隔 10 毫秒取一个语音帧。

常用的窗函数有矩形窗、汉明窗和汉宁窗, 矩形窗函数为:

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases};$$

汉明窗函数为:

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left[\frac{2\pi n}{N-1}\right], & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases};$$

汉宁窗函数为:

$$w(n) = \begin{cases} 0.5[1 - \cos(2\pi n/(N-1))], & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases}.$$

其中, N 为一个语音帧所包含的采样点的个数。

对所述语音信号进行预处理还可以包括检测所述语音信号中的有效语音。检测有效语音的目的是从语音信号中剔除非有效语音(即非语音段),

获得有效语音（即语音段），以降低特征提取的计算量和准确度，缩短语音识别的时间，提高识别率。可以根据语音信号的短时能量和短时过零率等进行有效语音检测。

在一实施例中，假设语音信号中第 n 个语音帧为 $x_n(m)$ ，则短时能量为：

$$E_n = \sum_{m=0}^{N-1} x_n^2(m)。$$

短时过零率为：

$$Z_n = \frac{1}{2N} \sum_{m=0}^{N-1} |sgn[x_n(m)] - sgn[x_n(m-1)]|。$$

其中， $sgn[.]$ 为符号函数，表达式为：

$$sgn[x] = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}。$$

可以采用两级判断法检测所述语音信号中有效语音的起点和终点。两级判断法为本领域的公知技术，此处不再赘述。

在另一实施例中，可以通过下述方法检测所述语音信号中的有效语音：

(1) 对所述语音信号进行加窗分帧，得到所述语音信号的语音帧 $x(n)$ 。在一个具体实施例中，可以对所述语音信号加汉明窗，每帧 20ms，帧移 10ms。若预处理过程中已对语音信号加窗分帧，则该步骤省略。

(2) 对所述语音帧 $x(n)$ 进行离散傅里叶变换 (Discrete Fourier Transform, DFT)，得到所述语音帧 $x(n)$ 的频谱：

$$x_a(k) = \sum_{n=0}^{N-1} x(n) * e^{-\frac{j2\pi nk}{N}}; 0 \leq k \leq N。$$

(3) 根据所述语音帧 $x(n)$ 的频谱计算各个频带的累计能量：

$$E(m) = \sum_{k=m_1}^{m_2} x_a(k); 0 \leq k \leq N。$$

其中 $E(m)$ 表示第 m 个频带的累计能量， (m_1, m_2) 表示第 m 个频带的起始频带点。

(4) 对所述各个频带的累计能量进行对数运算，得到所述各个频带的累计能量对数值。

(5) 将所述各个频带的累计能量对数值与预设阈值比较，得到所述有效语音。若一个频带的累计能量对数值高于预设阈值，则所述频带对应的语音为有效语音。

步骤 102，对预处理后的语音信号提取特征参数。

特征参数提取是对语音信号进行分析，提取出反映语音本质特征的声学参数序列。

提取的特征参数可以包括短时平均能量、短时平均过零率、共振峰、基音周期等时域参数，还可以包括线性预测系数 (Linear Prediction Coefficient, LPC)、线性预测倒谱系数 (Linear Prediction Cepstrum Coefficient, LPCC)、梅尔倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC)、感知线性预测 (Perceptual Linear Predictive, PLP) 等变换域参数。

在本申请的一个实施例中，可以提取语音信号的 MFCC 特征参数。提取 MFCC 特征参数的步骤如下：

(1) 对每个语音帧进行离散傅里叶变换 (Discrete Fourier Transform, DFT, 可以是快速傅里叶变换), 得到该语音帧的频谱。

(2) 求该语音帧的频谱幅度的平方, 得到该语音帧的离散能量谱。

(3) 将该语音帧的离散能量谱通过一组 Mel 频率上均匀分布的三角滤波器 (即三角滤波器组), 得到各个三角滤波器的输出。该组三角滤波器的中心频率在 Mel 频率刻度上均匀排列, 且每个三角滤波器的三角形两个底点的频率分别等于相邻的两个三角滤波器的中心频率。三角滤波器的中心频率为:

$$f(m) = \left(\frac{N}{F_s}\right) B^{-1} \left(B(f_1) + m \frac{B(f_h) - B(f_1)}{M+1} \right).$$

三角滤波器的频率响应为:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) < k < f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k < f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (0 \leq m \leq N = M),$$

其中, f_h 、 f_1 为三角滤波器的高频和低频; N 为傅里叶变换的点数; F_s 为采样频率; M 为三角滤波器的个数; $B^{-1} = 700 (e^{b/1125} - 1)$ 是 f_{mel} 的逆函数。

(4) 对所有三角滤波器的输出做对数运算, 得到该语音帧的对数功率谱 $S(m)$ 。

(5) 对 $S(m)$ 做离散余弦变换 (Discrete Cosine Transform, DCT), 得到该语音帧的初始 MFCC 特征参数。离散余弦变换为:

$$C_n = \sum_{m=1}^M S(m) \cos\left[\frac{\pi n(m+1/2)}{M}\right] \quad (0 \leq m \leq M).$$

MFCC 中引入了三角滤波器组, 且三角滤波器在低频段分布较密, 在高频段分布较疏, 符合人耳听觉特性, 在噪声环境下仍具有较好的识别性能。

提取 MFCC 特征参数的步骤还可以包括:

(6) 根据语音帧的初始 MFCC 特征参数提取语音帧的动态差分 MFCC 特征参数。初始 MFCC 特征参数只反映了语音参数的静态特性, 语音的动态特性可通过静态特征的差分谱来描述, 动静态结合可以有效提升系统的识别性能, 通常使用一阶和/或者二阶差分 MFCC 特征参数。

在一具体实施例中, 提取的 MFCC 特征参数为 39 维的特征矢量, 包括 13 维初始 MFCC 特征参数、13 维一阶差分 MFCC 特征参数和 13 维二阶差分 MFCC 特征参数。

在本申请的一个实施中, 在对预处理后的语音信号提取特征参数之后, 还可以对提取的特征参数进行降维处理, 得到降维后的特征参数。例如, 采用分段均值数据降维算法对所述特征参数 (例如 MFCC 特征参数) 进行降维处理, 得到降维后的特征参数。降维后的特征参数将用于后续的步骤。

步骤 103, 根据所述特征参数, 利用预先训练好的语音识别模型对所述语音信号进行解码, 得到以句子为单位的文本。

所述语音识别模型可以包括动态时间规整模型、隐马尔科夫模型、人工

神经网络模型、支持向量机分类模型等。所述语音识别模型也可以是两种或两种以上所述模型的组合。

在本申请的一个实施例中,所述语音识别模型为隐马尔科夫模型(HMM)。所述 HMM 包括声学模型和语音模型。

声学模型 (Acoustic Model): 采用隐马尔科夫模型对音素建模。在语音领域中,并不是以单词,而是以子词为识别单位,子词是声学模型基本的声学单元。在英语中子词为音素,对于某个特定的单词,相应的声学模型由多个音素通过查找发音字典的语法规则拼接得成。在汉语中子词为声母和韵母。每个子词可以用包括多个状态的 HMM 建模。举例来说,每一个音素可以用包含最多 6 个状态的 HMM 建模,每个状态可以用高斯混合模型 (GMM) 拟合对应的观测帧,观测帧按时序组合成观测序列。而每个声学模型可以生成长短不一的观测序列,即一对多映射。

语言模型 (Language Model): 是为了在语音识别的过程中有效地结合语法和语义的知识,提高识别率,减少搜索的范围。由于很难准确确定词的边界,以及声学模型描述语音变异性的能力有限,识别时将产生很多概率得分相似的词的序列。因此在实用的语音识别系统中通常使用语言模型 $P(W)$ 从诸多候选结果中选择最有可能的次序列来补充声学模型的不足。

在本实施例中,采用基于规则的语言模型。基于规则的语言模型可以总结出语法规则乃至语义规则,然后用这些规则排除声学识别中不合语法规则或语义规则的结果。统计语言模型通过统计概率描述词与词之间的依赖关系,间接地对语法或语义规则进行编码。

解码就是在状态网络中搜索一条最佳路径,语音对应这条路径的概率最大。在本实施例中,利用动态规划算法(即 Viterbi 算法)寻找全局最优路径。

假设对语音信号提取的特征参数为特征向量 Y ,通过解码算法寻找最有可能生成 Y 的词序列 $w_{1:L} = w_1, w_2 \dots w_L$ 。

解码算法是求解使得后验概率 $P(w|Y)$ 最大所对应的参数 w , 即:

$$w_{best} = \operatorname{argmax}\{p(w|Y)\}$$

由贝叶斯定理将上式转化为:

$$w_{best} = \operatorname{argmax}\left\{\frac{p(Y|w)p(w)}{p(Y)}\right\}$$

由于观测概率 $P(Y)$ 在给定观测序列下是常数,上式可进一步简化为:

$$w_{best} = \operatorname{argmax}\{p(Y|w)p(w)\}$$

其中先验概率 $P(W)$ 由语言模型决定,似然概率 $p(Y|w)$ 由声学模型决定。通过以上计算即可得出后验概率 $P(w|Y)$ 最大所对应的参数 w 。

步骤 104,通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

经过步骤 103,语音信号被解码为以句子为单位的文本,在常规的语音识别系统中,语音识别工作已经完成。本方法进一步从识别出来的以句子为单位的文本中提取出摘要句。

提取摘要句的目的是从语音中抽取重要的信息,剔除无用信息。

本方法通过 HMM 模型提取摘要句。此时，HMM 模型的双重随机关系可以描述为：一重随机关系为句子序列的释放，是可观察的；另一重随机关系为该句子是否应该被归为摘要句的性质，是不可观察的。所以用 HMM 模型来提取摘要句的过程可以描述为给定句子序列 $O = \{O_1, O_2 \dots O_n\}$ ，以确定句子是否为摘要句的最大可能性。主要步骤如下：

(1) 获得以句子为单位的文本的观察状态序列 $O = \{O_1, O_2 \dots O_n\}$ ；

(2) 确定 HMM 隐含状态。可以设置 5 个隐含状态。可以把隐含状态设置为“1”-符合，“2”-较符合，“3”-一般，“4”-较不符合，“5”-不符合，用来依次表示句子符合摘要句的程度。

(3) 进行 HMM 参数估计。首先随机产生初始的概率参数，经过不断地迭代，当达到设定的阈值时，停止计算，得到适合的 HMM 参数。

(4) 根据训练好的 HMM，通过 Viterbi 算法对句子进行标记，得到各个句子符合摘要句的符合度。

(5) 将满足预设符合度的句子（例如至少较符合的句子）从所述以句子为单位的文本中提取出来，得到所述以句子为单位的文本中的摘要句。

实施例一的语音处理方法对语音信号进行预处理；对预处理后的语音信号提取特征参数；根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；通过隐马尔科夫模型 HMM 从以句子为单位的文本中提取摘要句。实施例一不仅将语音信息转化为文字，还提取文字中的摘要句进行输出，剔除了由语音识别结果中的无用信息，获得更好的语音处理结果。

在另一实施例中，在提取 MFCC 特征参数时，可以进行声道长度归一化（Vocal Tract Length Normalization, VTLN），得到声道长度归一化的 MFCC 特征参数。

声道可以表示为级联声管模型，每个声管都可以看成是一个谐振腔，它们的共振频率取决于声管的长度和形状。因此，说话人之间的部分声学差异是由于说话人的声道长度不同。例如，声道长度的变化范围一般从 13cm（成年女性）变化到 18cm（成年男性），因此，不同性别的人说同一个元音的共振峰频率相差很大。VTLN 就是为了消除男、女声道长度的差异，使口音识别的结果不受性别的干扰。

VTLN 可以通过弯折和平移频率坐标来使各说话人的共振峰频率相匹配。在本实施例中，可以采用基于双线性变换的 VTLN 方法。该基于双线性变换的 VTLN 方法并不直接对语音信号的频谱进行折叠，而是采用双线性变换低通滤波器截止频率的映射公式，计算对齐不同说话人平均第三共振峰的频率弯折因子；根据所述频率弯折因子，采用双线性变换对三角滤波器组的位置（例如三角滤波器的起点、中间点和结束点）和宽度进行调整；根据调整后的三角滤波器组计算声道归一化的 MFCC 特征参数。例如，若要对语音信号进行频谱压缩，则对三角滤波器的刻度进行拉伸，此时三角滤波器组向左扩展和移动。若要对语音信号进行频谱拉伸，则对三角滤波器的刻度进行压缩，此时三角滤波器组向右压缩和移动。采用该基于双线性变换的 VTLN 方法对

特定人群或特定人进行声道归一化时，仅需要对三角滤波器组系数进行一次变换即可，无需每次在提取特征参数时都对信号频谱折叠，从而大大减小了计算量。并且，该基于双线性变换的 VTLN 方法避免了对频率因子线性搜索，减小了运算复杂度。同时，该基于双线性变换的 VTLN 方法利用双线性变换，使弯折的频率连续且无带宽改变。

实施例二

图 2 为本申请实施例二提供的语音处理装置的结构图。如图 2 所示，所述语音处理装置 10 可以包括：预处理单元 201、特征提取单元 202、解码单元 203、摘要提取单元 204。

预处理单元 201，用于对语音信号进行预处理。

所述语音信号可以是模拟语音信号，也可以是数字语音信号。若所述语音信号是模拟语音信号，则将所述模拟语音信号进行模数变换，转换为数字语音信号。

本申请用于连续语音识别，即对连续的音频流进行处理。在本申请的一个实施例中，所述语音处理方法应用于智能会议系统中，所述语音信号是通过语音输入设备（例如麦克风、手机话筒等）输入到智能会议系统的发言者的语音信号。

对所述语音信号进行预处理可以包括对所述语音信号进行预加重。

预加重的目的是提升语音的高频分量，使信号的频谱变得平坦。语音信号由于受声门激励和口鼻辐射的影响，能量在高频端明显减小，通常是频率越高幅值越小。当频率提升两倍时，功率谱幅度按 6dB/oct 跌落。因此，在对语音信号进行频谱分析或声道参数分析前，需要对语音信号的高频部分进行频率提升，即对语音信号进行预加重。预加重一般利用高通滤波器实现，高通滤波器的传递函数可以为：

$$H(z) = 1 - \kappa z^{-1}, \quad 0.9 \leq \kappa \leq 1.0.$$

其中， κ 为预加重系数，优选取值在 0.94-0.97 之间。

对所述语音信号进行预处理还可以包括对所述语音信号进行加窗分帧。

语音信号是一种非平稳的时变信号，主要分为浊音和清音两大类。浊音的基音周期、清浊音信号幅度和声道参数等都随时间而缓慢变化，但通常在 10ms-30ms 的时间内可以认为具有短时平稳性。语音信号处理中可以把语音信号分成一些短段（即获得短时平稳信号）来进行处理，这个过程称为分帧，得到的短段的语音信号称为语音帧。分帧是通过通过对语音信号进行加窗处理来实现的。为了避免相邻两帧的变化幅度过大，帧与帧之间需要重叠一部分。在本申请的一个实施例中，每个语音帧为 25 毫秒，相邻两个语音帧之间存在 15 毫秒重叠，也就是每隔 10 毫秒取一个语音帧。

常用的窗函数有矩形窗、汉明窗和汉宁窗，矩形窗函数为：

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases};$$

汉明窗函数为：

$$w(n) = \begin{cases} 0.54 - 0.46 \cos \left[\frac{2\pi n}{N-1} \right], & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases};$$

汉宁窗函数为:

$$w(n) = \begin{cases} 0.5[1 - \cos(2\pi n/(N-1))], & 0 \leq n \leq N-1 \\ 0, & \text{其他} \end{cases}。$$

其中, N 为一个语音帧所包含的采样点的个数。

对所述语音信号进行预处理还可以包括检测所述语音信号中的有效语音。

检测有效语音的目的是从语音信号中剔除非有效语音 (即非语音段), 获得有效语音 (即语音段), 以降低特征提取的计算量和准确度, 缩短语音识别的时间, 提高识别率。可以根据语音信号的短时能量和短时过零率等进行有效语音检测。

在一实施例中, 假设语音信号中第 n 个语音帧为 $x_n(m)$, 则短时能量为:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m)。$$

短时过零率为:

$$Z_n = \frac{1}{2N} \sum_{m=0}^{N-1} |sgn[x_n(m)] - sgn[x_n(m-1)]|。$$

其中, $sgn[.]$ 为符号函数, 表达式为:

$$sgn[x] = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}。$$

可以采用两级判断法检测所述语音信号中有效语音的起点和终点。两级判断法为本领域的公知技术, 此处不再赘述。

在另一实施例中, 可以通过下述方法检测所述语音信号中的有效语音:

(1) 对所述语音信号进行加窗分帧, 得到所述语音信号的语音帧 $x(n)$ 。在一个具体实施例中, 可以对所述语音信号加汉明窗, 每帧 20ms, 帧移 10ms。若预处理过程中已对语音信号加窗分帧, 则该步骤省略。

(2) 对所述语音帧 $x(n)$ 进行离散傅里叶变换 (Discrete Fourier Transform, DFT), 得到所述语音帧 $x(n)$ 的频谱:

$$x_a(k) = \sum_{n=0}^{N-1} x(n) * e^{-\frac{j2\pi nk}{N}}; 0 \leq k \leq N。$$

(3) 根据所述语音帧 $x(n)$ 的频谱计算各个频带的累计能量:

$$E(m) = \sum_{k=m_1}^{m_2} x_a(k); 0 \leq k \leq N。$$

其中 $E(m)$ 表示第 m 个频带的累计能量, (m_1, m_2) 表示第 m 个频带的起始频带点。

(4) 对所述各个频带的累计能量进行对数运算, 得到所述各个频带的累计能量对数值。

(5) 将所述各个频带的累计能量对数值与预设阈值比较, 得到所述有效语音。若一个频带的累计能量对数值高于预设阈值, 则所述频带对应的语音为有效语音。

特征提取单元 202, 用于对预处理后的语音信号提取特征参数。

特征参数提取是对语音信号进行分析, 提取出反映语音本质特征的声学

参数序列。

提取的特征参数可以包括短时平均能量、短时平均过零率、共振峰、基音周期等时域参数，还可以包括线性预测系数（Linear Prediction Coefficient, LPC）、线性预测倒谱系数（Linear Prediction Cepstrum Coefficient, LPCC）、梅尔倒谱系数（Mel Frequency Cepstrum Coefficient, MFCC）、感知线性预测（Perceptual Linear Predictive, PLP）等变换域参数。

在本申请的一个实施例中，可以提取语音信号的 MFCC 特征参数。提取 MFCC 特征参数的步骤如下：

（1）对预处理单元 201 得到的每一个语音帧进行离散傅里叶变换（Discrete Fourier Transform, DFT，可以是快速傅里叶变换），得到该语音帧的频谱。

（2）求该语音帧的频谱幅度的平方，得到该语音帧的离散能量谱。

（3）将该语音帧的离散能量谱通过一组 Mel 频率上均匀分布的三角滤波器（即三角滤波器组），得到各个三角滤波器的输出。该组三角滤波器的中心频率在 Mel 频率刻度上均匀排列，且每个三角滤波器的三角形两个底点的频率分别等于相邻的两个三角滤波器的中心频率。三角滤波器的中心频率为：

$$f(m) = \left(\frac{N}{F_s}\right) B^{-1} \left(B(f_1) + m \frac{B(f_h) - B(f_1)}{M+1} \right).$$

三角滤波器的频率响应为：

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) < k < f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k < f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (0 \leq m \leq N = M),$$

其中， f_h 、 f_1 为三角滤波器的高频和低频； N 为傅里叶变换的点数； F_s 为采样频率； M 为三角滤波器的个数； $B^{-1} = 700 \left(e^{b/1125} - 1 \right)$ 是 f_{mel} 的逆函数。

（4）对所有三角滤波器的输出做对数运算，得到该语音帧的对数功率谱 $S(m)$ 。

（5）对 $S(m)$ 做离散余弦变换（Discrete Cosine Transform, DCT），得到该语音帧的初始 MFCC 特征参数。离散余弦变换为：

$$C_n = \sum_{m=1}^M S(m) \cos \left[\frac{\pi n(m+1/2)}{M} \right] \quad (0 \leq m \leq M).$$

MFCC 中引入了三角滤波器组，且三角滤波器在低频段分布较密，在高频段分布较疏，符合人耳听觉特性，在噪声环境下仍具有较好的识别性能。

提取 MFCC 特征参数的步骤还可以包括：

（6）提取语音帧的动态差分 MFCC 特征参数。初始 MFCC 特征参数只反映了语音参数的静态特性，语音的动态特性可通过静态特征的差分谱来描述，动静态结合可以有效提升系统的识别性能，通常使用一阶和/或者二阶差分 MFCC 特征参数。

在一具体实施例中，提取的 MFCC 特征参数为 39 维的特征矢量，包括

13 维初始 MFCC 特征参数、13 维一阶差分 MFCC 特征参数和 13 维二阶差分 MFCC 特征参数。

在本申请的一个实施中，在对预处理后的语音信号提取特征参数之后，还可以对提取的特征参数进行降维处理，得到降维后的特征参数。例如，采用分段均值数据降维算法对所述特征参数（例如 MFCC 特征参数）进行降维处理，得到降维后的特征参数。降维后的特征参数将用于后续的步骤。

解码单元 203，用于根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本。

所述语音识别模型可以包括动态时间规整模型、隐马尔科夫模型、人工神经网络模型、支持向量机分类模型等。所述语音识别模型也可以是两种或两种以上所述模型的组合。

在本申请的一个实施例中，所述语音识别模型为隐马尔科夫模型(HMM)。所述 HMM 包括声学模型和语言模型。

声学模型 (Acoustic Model)：采用隐马尔科夫模型对音素建模。在语音领域中，并不是以单词，而是以子词为识别单位，子词是声学模型基本的声学单元。在英语中子词为音素，对于某个特定的单词，相应的声学模型由多个音素通过查找发音字典的语法规则拼接得成。在汉语中子词为声母和韵母。每个子词可以用包括多个状态的 HMM 建模。举例来说，每一个音素可以用包含最多 6 个状态的 HMM 建模，每个状态可以用高斯混合模型 (GMM) 拟合对应的观测帧，观测帧按时序组合成观测序列。而每个声学模型可以生成长短不一的观测序列，即一对多映射。

语言模型 (Language Model)：是为了在语音识别的过程中有效地结合语法和语义的知识，提高识别率，减少搜索的范围。由于很难准确确定词的边界，以及声学模型描述语音变异性的能力有限，识别时将产生很多概率得分相似的词的序列。因此在实用的语音识别系统中通常使用语言模型 $P(W)$ 从诸多候选结果中选择最有可能的次序列来补充声学模型的不足。

在本实施例中，采用基于规则的语言模型。基于规则的语言模型可以总结出语法规则乃至语义规则，然后用这些规则排除声学识别中不合语法规则或语义规则的结果。统计语言模型通过统计概率描述词与词之间的依赖关系，间接地对语法或语义规则进行编码。

解码就是在状态网络中搜索一条最佳路径，语音对应这条路径的概率最大。在本实施例中，利用动态规划算法（即 Viterbi 算法）寻找全局最优路径。

假设特征提取单元 202 提取的特征参数为特征向量 Y ，通过解码算法寻找最有可能生成 Y 的词序列 $w_{1:L} = w_1, w_2 \dots w_L$ 。

解码算法是求解使得后验概率 $P(w|Y)$ 最大所对应的参数 w ，即：

$$w_{best} = \operatorname{argmax}\{p(w|Y)\}$$

由贝叶斯定理将上式转化为：

$$w_{best} = \operatorname{argmax}\left\{\frac{p(Y|w)p(w)}{p(Y)}\right\}$$

由于观测概率 $P(Y)$ 在给定观测序列下是常数，上式可进一步简化为：

$$w_{best} = \operatorname{argmax}\{p(Y|w)p(w)\}$$

其中先验概率 $P(W)$ 由语言模型决定，似然概率 $p(Y|w)$ 由声学模型决定。通过以上计算即可得出后验概率 $P(w|Y)$ 最大所对应的参数 w 。

摘要提取单元 204，用于从所述以句子为单位的文本中提取摘要句。

解码单元 203 将语音信号解码为以句子为单位的文本，在常规的语音识别系统中，语音识别工作已经完成。在本申请中，摘要提取单元 204 从识别出来的以句子为单位的文本中提取出摘要句。

提取摘要句的目的是从语音中抽取重要的信息，剔除无用信息。

摘要提取单元 204 通过 HMM 模型提取摘要句。此时，HMM 模型的双重随机关系可以描述为：一重随机关系为句子序列的释放，是可观察的；另一重随机关系为该句子是否应该被归为摘要句的性质，是不可观察的。所以用 HMM 模型来提取摘要句的过程可以描述为给定句子序列 $O = \{O_1, O_2 \dots O_n\}$ ，以确定句子是否为摘要句的最大可能性。主要步骤如下：

(1) 获得以句子为单位的文本的观察状态序列 $O = \{O_1, O_2 \dots O_n\}$ ；

(2) 确定 HMM 隐含状态。可以设置 5 个隐含状态。可以把隐含状态设置为“1”-符合，“2”-较符合，“3”-一般，“4”-较不符合，“5”-不符合，用来依次表示句子符合摘要句的程度。

(3) 进行 HMM 参数估计。首先随机产生初始的概率参数，经过不断地迭代，当达到设定的阈值时，停止计算，得到适合的 HMM 参数。

(4) 根据训练好的 HMM，通过 Viterbi 算法对句子进行标记，得到各个句子符合摘要句的符合度。

(5) 将满足预设符合度的句子（例如至少较符合的句子）从所述以句子为单位的文本中提取出来，得到所述以句子为单位的文本中的摘要句。

实施例二的语音处理装置 10 对语音信号进行预处理；对预处理后的语音信号提取特征参数；根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；通过隐马尔科夫模型 HMM 从以句子为单位的文本中提取摘要句。实施例二不仅将语音信息转化为文字，还提取文字中的摘要句进行输出，剔除了由语音识别结果中的无用信息，获得更好的语音处理结果。

在另一实施例中，特征提取单元 202 在提取 MFCC 特征参数时，可以进行声道长度归一化（Vocal Tract Length Normalization, VTLN），得到声道长度归一化的 MFCC 特征参数。

声道可以表示为级联声管模型，每个声管都可以看成是一个谐振腔，它们的共振频率取决于声管的长度和形状。因此，说话人之间的部分声学差异是由于说话人的声道长度不同。例如，声道长度的变化范围一般从 13cm（成年女性）变化到 18cm（成年男性），因此，不同性别的人说同一个元音的共振峰频率相差很大。VTLN 就是为了消除男、女声道长度的差异，使口音识别的结果不受性别的干扰。

VTLN 可以通过弯折和平移频率坐标来使各说话人的共振峰频率相匹配。在本实施例中，可以采用基于双线性变换的 VTLN 方法。该基于双线性变换

的 VTLN 方法并不直接对语音信号的频谱进行折叠，而是采用双线性变换低通滤波器截止频率的映射公式，计算对齐不同说话人平均第三共振峰的频率弯折因子；根据所述频率弯折因子，采用双线性变换对三角滤波器组的位置（例如三角滤波器的起点、中间点和结束点）和宽度进行调整；根据调整后的三角滤波器组计算声道归一化的 MFCC 特征参数。例如，若要对语音信号进行频谱压缩，则对三角滤波器的刻度进行拉伸，此时三角滤波器组向左扩展和移动。若要对语音信号进行频谱拉伸，则对三角滤波器的刻度进行压缩，此时三角滤波器组向右压缩和移动。采用该基于双线性变换的 VTLN 方法对特定人群或特定人进行声道归一化时，仅需要对三角滤波器组系数进行一次变换即可，无需每次在提取特征参数时都对信号频谱折叠，从而大大减小了计算量。并且，该基于双线性变换的 VTLN 方法避免了对频率因子线性搜索，减小了运算复杂度。同时，该基于双线性变换的 VTLN 方法利用双线性变换，使弯折的频率连续且无带宽改变。

实施例三

本实施例提供一种非易失性可读存储介质，该非易失性可读存储介质上存储有计算机可读指令，该计算机可读指令被处理器执行时实现上述语音处理方法实施例中的步骤，例如图 1 所示的步骤 101-104：

步骤 101，对语音信号进行预处理；

步骤 102，对预处理后的语音信号提取特征参数；

步骤 103，根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；

步骤 104，通过隐马尔科夫模型 HMM 从以句子为单位的文本中提取摘要句。

或者，该计算机可读指令被处理器执行时实现上述装置实施例中各模块/单元的功能，例如图 2 中的单元 201-204：

预处理单元 201，用于对语音信号进行预处理；

特征提取单元 202，用于对预处理后的语音信号提取特征参数；

解码单元 203，用于根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；

摘要提取单元 204，用于通过隐马尔科夫模型 HMM 从以句子为单位的文本中提取摘要句。

实施例四

图 3 为本申请实施例三提供的计算机装置的示意图。所述计算机装置 1 包括存储器 20、处理器 30 以及存储在所述存储器 20 中并可在所述处理器 30 上运行的计算机可读指令 40，例如语音处理程序。所述处理器 30 执行所述计算机可读指令 40 时实现上述语音处理方法实施例中的步骤，例如图 1 所示的步骤 101-104：

步骤 101，对语音信号进行预处理；

步骤 102，对预处理后的语音信号提取特征参数；

步骤 103，根据所述特征参数，利用预先训练好的语音识别模型对所述

语音信号进行解码，得到以句子为单位的文本；

步骤 104，通过隐马尔科夫模型 HMM 从以句子为单位的文本中提取摘要句。

或者，所述处理器 30 执行所述计算机可读指令 40 时实现上述装置实施例中各模块/单元的功能，例如图 2 中的单元 201-204：

预处理单元 201，用于对语音信号进行预处理；

特征提取单元 202，用于对预处理后的语音信号提取特征参数；

解码单元 203，用于根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；

摘要提取单元 204，用于通过隐马尔科夫模型 HMM 从以句子为单位的文本中提取摘要句。

示例性的，所述计算机可读指令 40 可以被分割成一个或多个模块/单元，所述一个或者多个模块/单元被存储在所述存储器 20 中，并由所述处理器 30 执行，以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令段，该指令段用于描述所述计算机可读指令 40 在所述计算机装置 1 中的执行过程。例如，所述计算机可读指令 40 可以被分割成图 2 中的预处理单元 201、特征提取单元 202、解码单元 203、摘要提取单元 204，各单元具体功能参见实施例二。

所述计算机装置 1 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。本领域技术人员可以理解，所述示意图 3 仅仅是计算机装置 1 的示例，并不构成对计算机装置 1 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，例如所述计算机装置 1 还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 30 可以是中央处理单元(Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器 30 也可以是任何常规的处理器等，所述处理器 30 是所述计算机装置 1 的控制中心，利用各种接口和线路连接整个计算机装置 1 的各个部分。

所述存储器 20 可用于存储所述计算机可读指令 40 和/或模块/单元，所述处理器 30 通过运行或执行存储在所述存储器 20 内的计算机可读指令和/或模块/单元，以及调用存储在存储器 20 内的数据，实现所述计算机装置 1 的各种功能。所述存储器 20 可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序（比如声音播放功能、图像播放功能等）等；存储数据区可存储根据计算机装置 1 的使用所创建的数据（比如音频数据、电话本等）等。此外，存储器 20 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如硬盘、内存、插接式硬盘，智能存储卡 (Smart Media Card, SMC)，安全数字 (Secure Digital, SD) 卡，闪存卡 (Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固

态存储器件。

所述计算机装置 1 集成的模块/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，可以存储在一个非易失性可读取存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，也可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性可读取存储介质中，该计算机可读指令在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机可读指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述非易失性可读取介质可以包括：能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号、电信信号以及软件分发介质等。需要说明的是，所述非易失性可读取介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减，例如在某些司法管辖区，根据立法和专利实践，非易失性可读取介质不包括电载波信号和电信信号。

在本申请所提供的几个实施例中，应该理解到，所揭露的计算机装置和方法，可以通过其它的方式实现。例如，以上所描述的计算机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。

另外，在本申请各个实施例中的各功能单元可以集成在相同处理单元中，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相同单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

对于本领域技术人员而言，显然本申请不限于上述示范性实施例的细节，而且在不背离本申请的精神或基本特征的情况下，能够以其他的具体形式实现本申请。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本申请的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外，显然“包括”一词不排除其他单元或步骤，单数不排除复数。计算机装置权利要求中陈述的多个单元或计算机装置也可以由同一个单元或计算机装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

权 利 要 求 书

1. 一种语音处理方法，其特征在于，所述方法包括：

对语音信号进行预处理；

对预处理后的语音信号提取特征参数；

根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；

通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

2. 如权利要求 1 所述的方法，其特征在于，所述通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句，具体包括：

获得所述以句子为单位的文本的观察状态序列 $O = \{O_1, O_2 \dots O_n\}$ ；

确定 HMM 的隐含状态；

进行 HMM 参数估计，得到训练好的 HMM；

根据所述训练好的 HMM，通过 Viterbi 算法对所述句子进行标记，得到各个句子符合摘要句的符合度；

将满足预设符合度的句子从所述以句子为单位的文本中提取出来，得到所述以句子为单位的文本中的摘要句。

3. 如权利要求 1 所述的方法，其特征在于，所述对语音信号进行预处理包括检测所述语音信号中的有效语音，具体包括：

对所述语音信号进行加窗分帧，得到所述语音信号的语音帧；

对所述语音帧进行离散傅里叶变换，得到所述语音帧的频谱；

根据所述语音帧的频谱计算各个频带的累计能量；

对所述各个频带的累计能量进行对数运算，得到所述各个频带的累计能量对数值；

将所述各个频带的累计能量对数值与预设阈值进行比较，得到所述有效语音。

4. 如权利要求 1 所述的方法，其特征在于，所述特征参数包括初始梅尔倒谱系数 MFCC 特征参数、一阶差分 MFCC 特征参数和二阶差分 MFCC 特征参数。

5. 如权利要求 1 所述的方法，其特征在于，所述方法还包括：

对所述特征参数进行降维处理，得到降维后的特征参数。

6. 如权利要求 1 所述的方法，其特征在于，所述对预处理后的语音信号提取特征参数包括对预处理后的语音信号提取梅尔倒谱系数 MFCC 特征参数，具体包括：

采用双线性变换低通滤波器截止频率的映射公式，计算对齐不同说话人平均第三共振峰的频率弯折因子；

根据所述频率弯折因子，采用双线性变换对 MFCC 特征参数提取所使用的三角滤波器组的位置和宽度进行调整；

根据调整后的三角滤波器组计算声道归一化的 MFCC 特征参数。

7. 如权利要求 1 所述的方法, 其特征在于, 所述对预处理后的语音信号提取特征参数包括对预处理后的语音信号提取梅尔倒谱系数 MFCC 特征参数, 具体包括:

对每个语音帧进行离散傅里叶变换 DFT, 得到该语音帧的频谱;

求该语音帧的频谱幅度的平方, 得到该语音帧的离散能量谱;

将该语音帧的离散能量谱通过 Mel 频率上均匀分布的三角滤波器组, 得到各个三角滤波器的输出;

对所有三角滤波器的输出做对数运算, 得到该语音帧的对数功率谱;

对所述对数功率谱离散余弦变换 DCT, 得到该语音帧的初始 MFCC 特征参数。

8. 一种语音处理装置, 其特征在于, 所述装置包括:

预处理单元, 用于对语音信号进行预处理;

特征提取单元, 用于对预处理后的语音信号提取特征参数;

解码单元, 用于根据所述特征参数, 利用预先训练好的语音识别模型对所述语音信号进行解码, 得到以句子为单位的文本;

摘要提取单元, 用于通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

9. 一种计算机装置, 其特征在于, 所述计算机装置包括处理器和存储器, 所述存储器用于存储至少一个计算机可读指令, 所述处理器用于执行所述至少一个计算机可读指令以实现以下步骤:

对语音信号进行预处理;

对预处理后的语音信号提取特征参数;

根据所述特征参数, 利用预先训练好的语音识别模型对所述语音信号进行解码, 得到以句子为单位的文本;

通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

10. 如权利要求 9 所述的计算机装置, 其特征在于, 所述通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句, 具体包括:

获得所述以句子为单位的文本的观察状态序列 $O = \{O_1, O_2 \dots O_n\}$;

确定 HMM 的隐含状态;

进行 HMM 参数估计, 得到训练好的 HMM;

根据所述训练好的 HMM, 通过 Viterbi 算法对所述句子进行标记, 得到各个句子符合摘要句的符合度;

将满足预设符合度的句子从所述以句子为单位的文本中提取出来, 得到所述以句子为单位的文本中的摘要句。

11. 如权利要求 9 所述的计算机装置, 其特征在于, 所述对语音信号进行预处理包括检测所述语音信号中的有效语音, 具体包括:

对所述语音信号进行加窗分帧, 得到所述语音信号的语音帧;

对所述语音帧进行离散傅里叶变换, 得到所述语音帧的频谱;

根据所述语音帧的频谱计算各个频带的累计能量;

对所述各个频带的累计能量进行对数运算, 得到所述各个频带的累计能

量对数值；

将所述各个频带的累计能量对数值与预设阈值进行比较，得到所述有效语音。

12. 如权利要求 9 所述的计算机装置，其特征在于，所述处理器还用于执行所述至少一个计算机可读指令以实现以下步骤：

对所述特征参数进行降维处理，得到降维后的特征参数。

13. 如权利要求 9 所述的计算机装置，其特征在于，所述对预处理后的语音信号提取特征参数包括对预处理后的语音信号提取梅尔倒谱系数 MFCC 特征参数，具体包括：

采用双线性变换低通滤波器截止频率的映射公式，计算对齐不同说话人平均第三共振峰的频率弯折因子；

根据所述频率弯折因子，采用双线性变换对 MFCC 特征参数提取所使用的三角滤波器组的位置和宽度进行调整；

根据调整后的三角滤波器组计算声道归一化的 MFCC 特征参数。

14. 如权利要求 9 所述的计算机装置，其特征在于，所述对预处理后的语音信号提取特征参数包括对预处理后的语音信号提取梅尔倒谱系数 MFCC 特征参数，具体包括：

对每个语音帧进行离散傅里叶变换 DFT，得到该语音帧的频谱；

求该语音帧的频谱幅度的平方，得到该语音帧的离散能量谱；

将该语音帧的离散能量谱通过 Mel 频率上均匀分布的三角滤波器组，得到各个三角滤波器的输出；

对所有三角滤波器的输出做对数运算，得到该语音帧的对数功率谱；

对所述对数功率谱离散余弦变换 DCT，得到该语音帧的初始 MFCC 特征参数。

15. 一种非易失性可读存储介质，所述非易失性可读存储介质上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现以下步骤：

对语音信号进行预处理；

对预处理后的语音信号提取特征参数；

根据所述特征参数，利用预先训练好的语音识别模型对所述语音信号进行解码，得到以句子为单位的文本；

通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句。

16. 如权利要求 15 所述的存储介质，其特征在于，所述通过隐马尔科夫模型 HMM 从所述以句子为单位的文本中提取摘要句，具体包括：

获得所述以句子为单位的文本的观察状态序列 $O = \{O_1, O_2 \dots O_n\}$ ；

确定 HMM 的隐含状态；

进行 HMM 参数估计，得到训练好的 HMM；

根据所述训练好的 HMM，通过 Viterbi 算法对所述句子进行标记，得到各个句子符合摘要句的符合度；

将满足预设符合度的句子从所述以句子为单位的文本中提取出来，得到所述以句子为单位的文本中的摘要句。

17. 如权利要求 15 所述的存储介质, 其特征在于, 所述对语音信号进行预处理包括检测所述语音信号中的有效语音, 具体包括:

对所述语音信号进行加窗分帧, 得到所述语音信号的语音帧;

对所述语音帧进行离散傅里叶变换, 得到所述语音帧的频谱;

根据所述语音帧的频谱计算各个频带的累计能量;

对所述各个频带的累计能量进行对数运算, 得到所述各个频带的累计能量对数值;

将所述各个频带的累计能量对数值与预设阈值进行比较, 得到所述有效语音。

18. 如权利要求 15 所述的存储介质, 其特征在于, 所述至少一个计算机可读指令被所述处理器执行时还实现以下步骤:

对所述特征参数进行降维处理, 得到降维后的特征参数。

19. 如权利要求 15 所述的存储介质, 其特征在于, 所述对预处理后的语音信号提取特征参数包括对预处理后的语音信号提取梅尔倒谱系数 MFCC 特征参数, 具体包括:

采用双线性变换低通滤波器截止频率的映射公式, 计算对齐不同说话人平均第三共振峰的频率弯折因子;

根据所述频率弯折因子, 采用双线性变换对 MFCC 特征参数提取所使用的三角滤波器组的位置和宽度进行调整;

根据调整后的三角滤波器组计算声道归一化的 MFCC 特征参数。

20. 如权利要求 15 所述的存储介质, 其特征在于, 所述对预处理后的语音信号提取特征参数包括对预处理后的语音信号提取梅尔倒谱系数 MFCC 特征参数, 具体包括:

对每个语音帧进行离散傅里叶变换 DFT, 得到该语音帧的频谱;

求该语音帧的频谱幅度的平方, 得到该语音帧的离散能量谱;

将该语音帧的离散能量谱通过 Mel 频率上均匀分布的三角滤波器组, 得到各个三角滤波器的输出;

对所有三角滤波器的输出做对数运算, 得到该语音帧的对数功率谱;

对所述对数功率谱离散余弦变换 DCT, 得到该语音帧的初始 MFCC 特征参数。

1/2

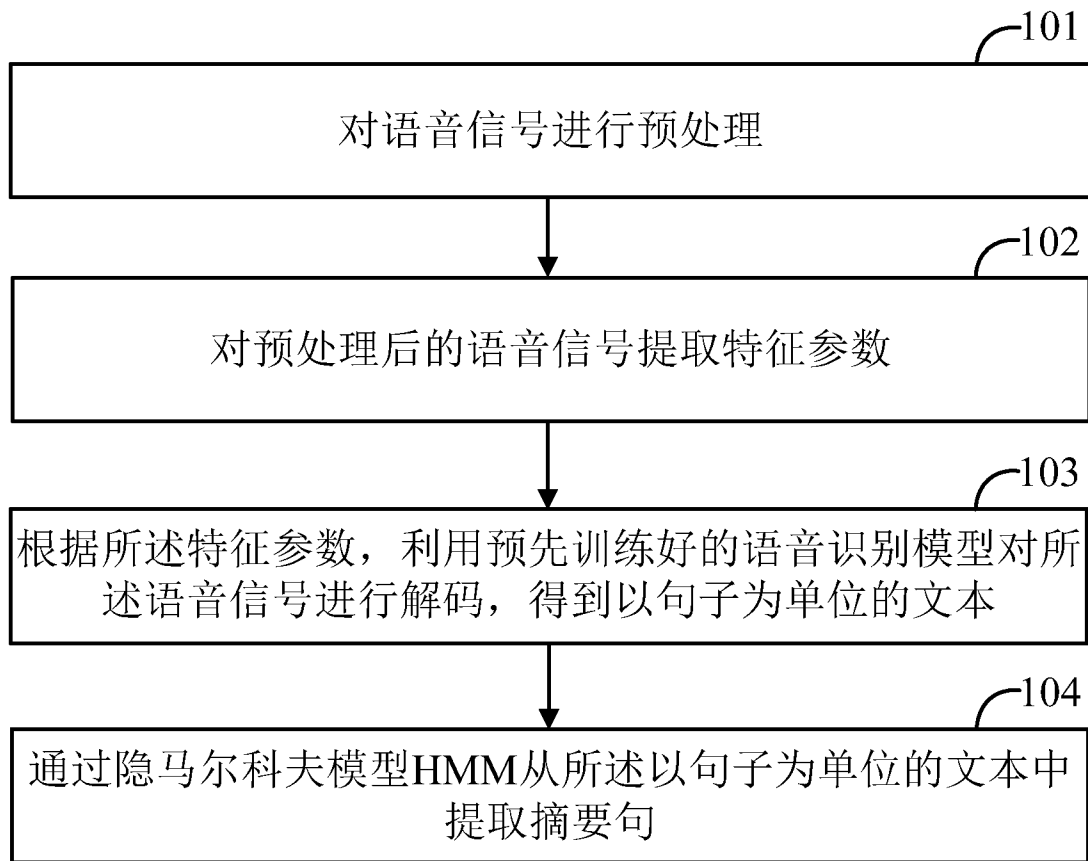


图 1

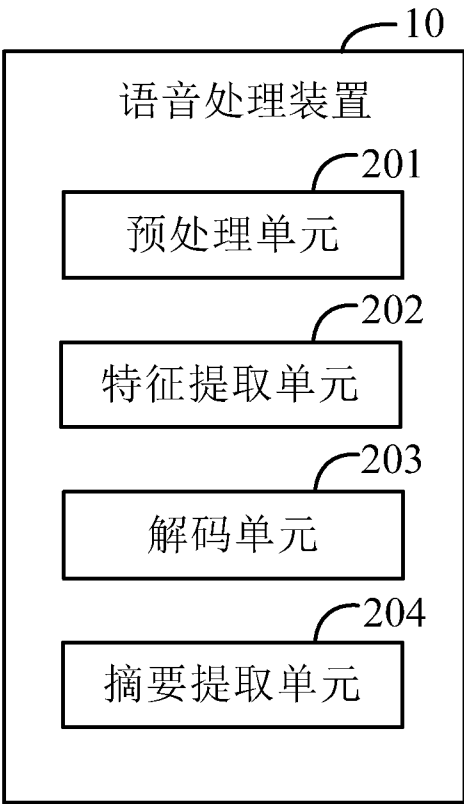


图 2

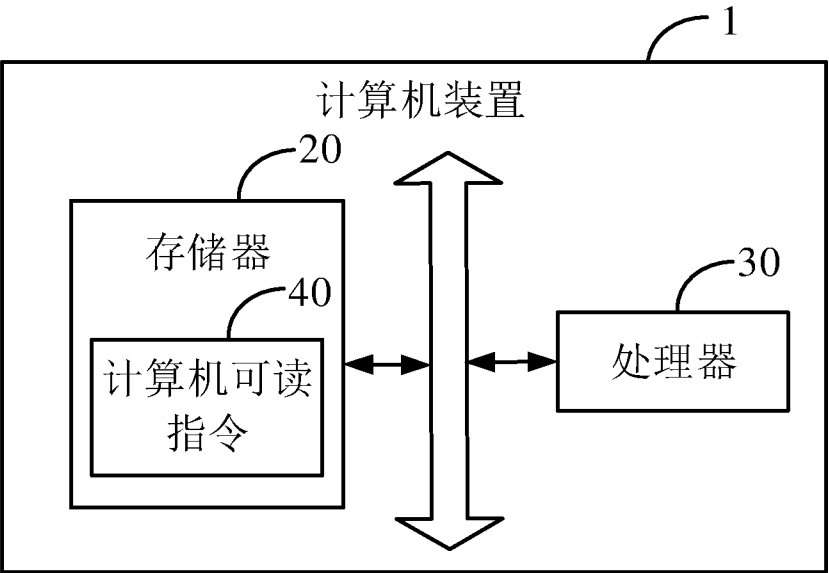


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/108190

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02(2006.01)i; G10L 15/14(2006.01)i; G10L 15/26(2006.01)i; G10L 19/02(2013.01)i; G10L 25/18(2013.01)i;
G10L 25/24(2013.01)i; G10L 25/51(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; VEN; USTXT; EPTXT; WOTXT; 3GPP: 语音, 识别, 剔除, 无关, 无用, 摘要, 文本, 隐马尔可夫, 隐马尔科夫, 提取, voice, identify, select, abstract, eliminate, irrelevant, redundancy, markov, viterbi, HMM, text

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102436809 A (SOUTHEAST UNIVERSITY) 02 May 2012 (2012-05-02) description, paragraphs [0090]-[0161]	1-20
Y	CN 101464898 A (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LTD.) 24 June 2009 (2009-06-24) description, page 5, lines 4-14	1-20
A	CN 103646094 A (SHANGHAI ZIZHU DIGITAL CREATIVE HARBOR CO., LTD.) 19 March 2014 (2014-03-19) entire document	1-20
A	CN 108305632 A (SHENZHEN EAGLESOUL TECHNOLOGY CO., LTD.) 20 July 2018 (2018-07-20) entire document	1-20
A	US 2014180694 A1 (SPANSION LLC.) 26 June 2014 (2014-06-26) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

15 March 2019

Date of mailing of the international search report

11 April 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/108190

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102436809	A	02 May 2012	CN	102436809	B	24 April 2013
CN	101464898	A	24 June 2009	CN	101464898	B	21 September 2011
CN	103646094	A	19 March 2014	CN	103646094	B	31 May 2017
CN	108305632	A	20 July 2018	None			
US	2014180694	A1	26 June 2014	US	9514739	B2	06 December 2016
				WO	2014100195	A3	21 August 2014
				WO	2014100195	A2	26 June 2014

国际检索报告

国际申请号

PCT/CN2018/108190

A. 主题的分类

G10L 15/02(2006.01)i; G10L 15/14(2006.01)i; G10L 15/26(2006.01)i; G10L 19/02(2013.01)i; G10L 25/18(2013.01)i; G10L 25/24(2013.01)i; G10L 25/51(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;CNKI;VEN;USTXT;EPTXT;WOTXT;3GPP: 语音, 识别, 剔除, 无关, 无用, 摘要, 文本, 隐马尔可夫, 隐马尔科夫, 提取, voice, identify, select, abstract, eliminate, irrelevant, redundancy, markov, viterbi, HMM, text

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	CN 102436809 A (东南大学) 2012年 5月 2日 (2012 - 05 - 02) 说明书第[0090]-[0161]段	1-20
Y	CN 101464898 A (腾讯科技深圳有限公司) 2009年 6月 24日 (2009 - 06 - 24) 说明书第5页第4-14行	1-20
A	CN 103646094 A (上海紫竹数字创意港有限公司) 2014年 3月 19日 (2014 - 03 - 19) 全文	1-20
A	CN 108305632 A (深圳市鹰硕技术有限公司) 2018年 7月 20日 (2018 - 07 - 20) 全文	1-20
A	US 2014180694 A1 (SPANSION LLC) 2014年 6月 26日 (2014 - 06 - 26) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 3月 15日

国际检索报告邮寄日期

2019年 4月 11日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

马小瑜

电话号码 86-(512)-88996083

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/108190

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	102436809	A	2012年 5月 2日	CN	102436809	B	2013年 4月 24日
CN	101464898	A	2009年 6月 24日	CN	101464898	B	2011年 9月 21日
CN	103646094	A	2014年 3月 19日	CN	103646094	B	2017年 5月 31日
CN	108305632	A	2018年 7月 20日	无			
US	2014180694	A1	2014年 6月 26日	US	9514739	B2	2016年 12月 6日
				WO	2014100195	A3	2014年 8月 21日
				WO	2014100195	A2	2014年 6月 26日

表 PCT/ISA/210 (同族专利附件) (2015年1月)