

(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 20 日 (20.02.2020)



(10) 国际公布号
WO 2020/034628 A1

(51) 国际专利分类号:
G10L 17/02 (2013.01)

(21) 国际申请号: PCT/CN2019/077512

(22) 国际申请日: 2019 年 3 月 8 日 (08.03.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810922056.0 2018 年 8 月 14 日 (14.08.2018) 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: 张丝潞(ZHANG, Siying); 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)。
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)。

(54) Title: ACCENT IDENTIFICATION METHOD AND DEVICE, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 口音识别方法、装置、计算机装置及存储介质

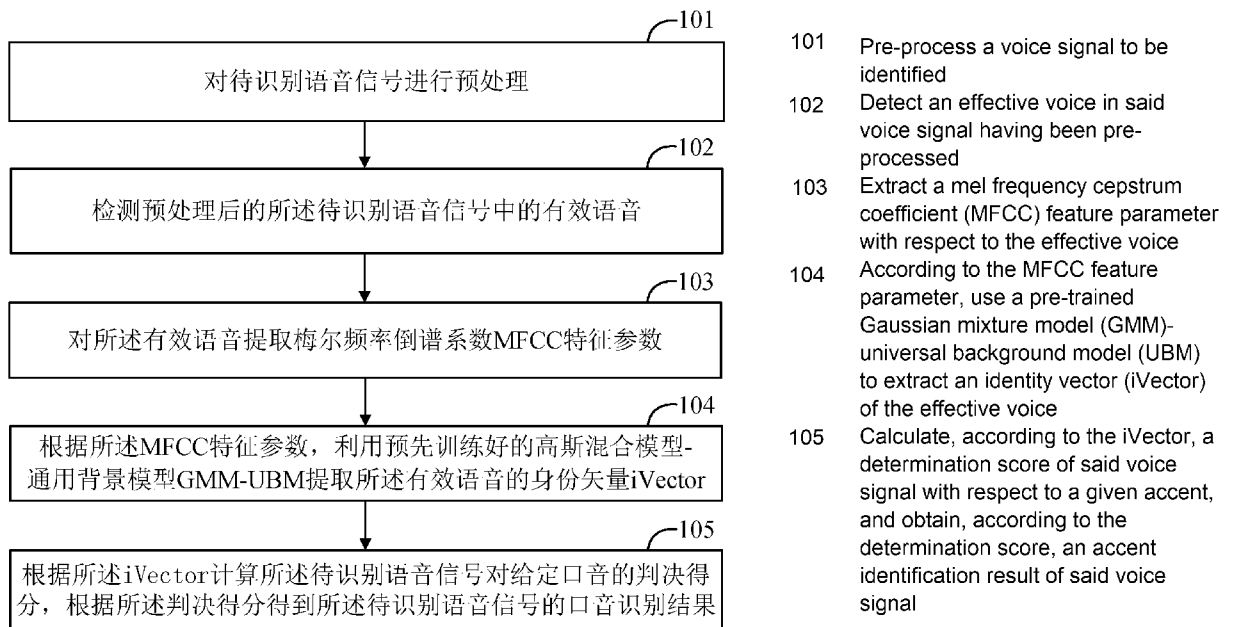


图 1

(57) Abstract: An accent identification method, comprising: pre-processing a voice signal to be identified (101); detecting an effective voice in said voice signal having been pre-processed (102); extracting a mel frequency cepstrum coefficient (MFCC) feature parameter with respect to the effective voice (103); according to the MFCC feature parameter, using a pre-trained Gaussian mixture model (GMM)-universal background model (UBM) to extract an identity vector (iVector) of the effective voice (104); and calculating, according to the iVector, a determination score of said voice signal with respect to a given accent, and obtaining, according to the determination score, an accent identification result of said voice signal (105).

(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, 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))。

(57) 摘要: 一种口音识别方法, 包括: 对待识别语音信号进行预处理 (101); 检测预处理后的所述待识别语音信号中的有效语音 (102); 对所述有效语音提取梅尔频率倒谱系数MFCC特征参数 (103); 根据所述MFCC特征参数, 利用预先训练好的高斯混合模型-通用背景模型GMM-UBM提取所述有效语音的身份矢量iVector (104); 根据所述iVector计算所述待识别语音信号对给定口音的判决得分, 根据所述判决得分得到所述待识别语音信号的口音识别结果 (105)。

口音识别方法、装置、计算机装置及存储介质

本申请要求于2018年08月14日提交中国专利局，申请号为201810922056.0发明名称为“口音识别方法、装置、计算机装置及计算机可读存储介质”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

本申请计算机听觉技术领域，具体涉及一种口音识别方法及装置、计算机装置和存储介质。

背景技术

随着各类智能身份认证的不断出现和落地应用，诸如人脸识别、声纹识别已经获得了较为成熟的发展，但识别的准确性依然有提升的空间，诸如在声纹识别方向仍然可找到突破点以得到更为准确的识别结果，口音因素就是一个。由于说话人所生活的地域不同，即使在都讲普通话的情况下或多或少依然会有口音的差别，若能在现有的声纹识别中加入口音识别作为补充，应用场景将会有进一步的扩展，最为直接的应用为在声纹识别前识别出该说话人所处地域的范围，进而缩小后续识别的对象范围。然而，现有的口音识别效果并不理想，识别速度较慢且准确度不高。

发明内容

鉴于以上内容，有必要提出一种口音识别方法及装置、计算机装置和存储介质，其可以实现快速准确的口音识别。

本申请的第一方面提供一种口音识别方法，所述方法包括：

对待识别语音信号进行预处理；

检测预处理后的所述待识别语音信号中的有效语音；

对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；

根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；

根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

本申请的第二方面提供一种口音识别装置，所述装置包括：

预处理单元，用于对待识别语音信号进行预处理；

检测单元，用于检测预处理后的所述待识别语音信号中的有效语音；

第一提取单元，用于对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；

第二提取单元，用于根据所述 MFCC 特征参数，利用预先训练好的高斯

混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；识别单元，用于根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

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口音识别方法。

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口音识别方法。

本申请对待识别语音信号进行预处理；检测预处理后的所述待识别语音信号中的有效语音；对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。本申请可以通过数据库层面发现问题，而无需测试人员通过复杂、大量的功能测试去发现问题。本申请可以实现快速准确的口音识别。

附图说明

图 1 是本申请实施例提供的口音识别方法的流程图。

图 2 是本申请实施例提供的口音识别装置的结构图。

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，本申请的口音识别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口音识别方法的流程图。所述口音识别方法应用于计算机装置。所述口音识别方法检测出测试库的失效对象，以便开发人员做代码修复，完成失效对象整改。

如图1所示，所述口音识别方法具体包括以下步骤：

步骤101，对待识别语音信号进行预处理。

所述待识别语音信号可以是模拟语音信号，也可以是数字语音信号。若所述待识别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。

所述待识别语音信号可以通过语音输入设备（例如麦克风、手机话筒等）采集到的语音信号。

对所述待识别语音信号进行预处理可以包括对所述待识别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，在对待识别语音信号进行频谱分析或声道参数分析前，需要对待识别语音信号的高频部分进行频率提升，即对待识别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。

对所述待识别语音信号进行预处理还可以包括对所述待识别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。为了获得短时平稳信号，语音信号处理中可以把语音信号分成一些短段来进行处理，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20毫秒，相邻两个语音帧之间存在10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。

步骤 102, 检测预处理后的所述待识别语音信号中的有效语音。

可以根据预处理后的所述待识别语音信号的短时能量和短时过零率等进行端点检测, 以确定所述待识别语音信号中的有效语音。

在本实施例中, 可以通过下述方法检测预处理后的所述待识别语音信号中的有效语音:

(1) 对预处理后的所述待识别语音信号进行加窗分帧, 得到所述待识别语音信号的语音帧 $x(n)$ 。在一个具体实施例中, 可以对预处理后的所述待识别语音信号加汉明窗, 每帧 20ms, 帧移 10ms。若预处理过程中已对待识别语音信号加窗分帧, 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 103, 对所述有效语音提取梅尔频率倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC) 特征参数。

提取 MFCC 特征参数的流程如下:

(1) 对每一个语音帧进行离散傅里叶变换 (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_l 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:

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

(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。

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。

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

步骤 104, 根据所述 MFCC 特征参数, 利用预先训练好的高斯混合模型 (Gaussian Mixture Model, GMM)-通用背景模型 (Universal Background Model, UBM) 提取所述有效语音的身份矢量 (identity-vector, iVector)。

提取 iVector 之前, 首先要用大量属于不同口音的训练数据训练出通用背景模型。通用背景模型实际上是一种高斯混合模型 (GMM), 旨在解决实际场景数据量稀缺的问题。GMM 是一种参数化的生成性模型, 具备对实际数据极强的表征力 (基于高斯分量实现)。高斯分量越多, GMM 表征力越强, 规模也越庞大, 此时负面效应逐步凸显——若想获得一个泛化能力较强的 GMM 模型, 则需要足够的数据来驱动 GMM 的参数训练, 然而实际场景中获取的语音数据甚至连分钟级都很难企及。UBM 正是解决了训练数据不足的问题。UBM 是利用大量属于不同口音的训练数据 (无关乎说话人、地域) 混合起来充分训练, 得到一个可以对语音共通特性进行表征的全局 GMM, 可大大缩减从头计算 GMM 参数所消耗的资源。通用背景模型训练完成后,

只需利用单独属于每个口音的训练数据，分别对 UBM 的参数进行微调（例如通过 UBM 自适应），得到属于各个口音的 GMM。

在一个实施例中，不同口音可以是属于不同地域的口音。所述地域可以是按照行政区域来划分，例如辽宁、北京、天津、上海、河南、广东等。所述地域也可以是按照普遍的经验以口音对地区来划分，例如闽南、客家等。

提取 iVector 是基于全差异空间建模（TV）方法将 UBM 训练得出的高维 GMM 映射至低维度的全变量子空间，可突破随语音信号时长过长而提取的向量维度过大不便计算的限制，并能提升计算速度，表达出更全面的特征。GMM-UBM 中的 GMM 超矢量可以包括跟说话人本身有关的矢量特征和跟信道以及其他变化有关的矢量特征的线性叠加。

TV 模型的子空间建模形式为：

$$M = m + Tw$$

其中，M 表示语音的 GMM 超矢量，即所述 MFCC 特征参数，m 表示口音无关的 GMM 超矢量，T 表示描述差异的空间的载荷矩阵，w 表示 GMM 超矢量 M 在载荷矩阵空间下对应的低维因子表示，即 iVector。

在本实施例中，可以对提取的 iVector 行噪声补偿。在一实施例中，可以采用线性判别分析（Linear Discriminative Analysis, LDA）和类内协方差规整（Within Class Covariance Normalization, WCCN）对提取的 iVector 进行噪声补偿。

步骤 105，根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

给定口音可以是一个也可以是多个。例如，若给定口音为一个，可以根据所述 iVector 计算所述待识别语音信号对该给定口音的判决得分，根据所述待识别语音信号对该给定口音的判决得分判断所述待识别语音信号是否为该给定口音。可以判断所述判决得分是否大于预设得分（例如 9 分），若所述判决得分大于预设得分，则判断所述待识别语音信号为该给定口音。

若给定口音为多个，可以根据所述 iVector 计算所述待识别语音信号对每个给定口音的判决得分，根据所述待识别语音信号对每个给定口音的判决得分判断所述语音为多个给定口音中的哪一个。可以确定对多个给定口音的判决得分中的最高得分，将所述最高得分对应的给定口音作为所述待识别语音信号所属的口音。

在本实施例中，可以利用逻辑回归（Logistic Regression）模型计算所述待识别语音信号对给定口音的判决得分。逻辑回归模型作为一个分类器，可根据待识别语音信号的 iVector 对待识别语音信号进行打分。特别地，在一具体实施例中，可以使用多类逻辑回归模型计算所述待识别语音信号对给定口音的判决得分。

假设给定口音包括口音 1、口音 2、... 口音 N 共 N 种口音，则利用 N 类逻辑回归模型计算所述待识别语音信号对给定口音的判决得分。将待识别语音信号的 iVector（记为 x_t ）输入所述 N 类逻辑回归模型，得到 N 个判决得分 s_{it} （即所述待识别语音信号对 N 种给定口音的判决得分）， $s_{it} = w_i * x_t + k_i$ ，

$i = 1, \dots, N$ 。求取 N 个判决得分 s_{it} , $i = 1, \dots, N$ 中的最高得分 s_{jt} , 最高得分 s_{jt} 对应的口音 j 为所述待识别语音信号所属的口音。其中, w_i 、 k_i 是 N 类逻辑回归模型的参数, w_i 为回归系数, k_i 为常数, 针对每个给定口音均会有对应的 w_i 和 k_i , w_i 、 k_i 组成 N 类逻辑回归模型的参数向量, $M = \{(w_1, k_1), (w_2, k_2), \dots, (w_N, k_N)\}$ 。

实施例一的口音识别方法对待识别语音信号进行预处理; 检测预处理后的所述待识别语音信号中的有效语音; 对所述有效语音提取 MFCC 特征参数; 根据所述 MFCC 特征参数, 利用预先训练好的 GMM-UBG 提取所述有效语音的 iVector; 根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分, 根据所述判决得分得到所述待识别语音信号的口音识别结果。实施例一可以实现快速准确的口音识别。

在其他的实施例中, 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 方法并不直接对待识别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, 若要对待识别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。若要对待识别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。

在另一实施例中, 所述口音识别方法还可以包括: 根据所述口音识别结果进行声纹识别。由于说话人所生活的地域不同, 即使在都讲普通话的情况下或多或少依然会有口音的差别, 将口音识别应用到声纹识别中, 可以缩小后续声纹识别的对象范围, 得到更为准确的识别结果。

实施例二

图 2 为本申请实施例二提供的口音识别装置的结构图。如图 2 所示，所述口音识别装置 10 可以包括：预处理单元 201、检测单元 202、第一提取单元 203、第二提取单元 204、识别单元 205。

预处理单元 201，用于对待识别语音信号进行预处理。

所述待识别语音信号可以是模拟语音信号，也可以是数字语音信号。若所述待识别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。

所述待识别语音信号可以通过语音输入设备（例如麦克风、手机话筒等）采集到的语音信号。

对所述待识别语音信号进行预处理可以包括对所述待识别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，在对待识别语音信号进行频谱分析或声道参数分析前，需要对待识别语音信号的高频部分进行频率提升，即对待识别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。

对所述待识别语音信号进行预处理还可以包括对所述待识别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。为了获得短时平稳信号，语音信号处理中可以把语音信号分成一些短段来进行处理，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 20 毫秒，相邻两个语音帧之间存在 10 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。

检测单元 202, 用于检测预处理后的所述待识别语音信号中的有效语音。

可以根据预处理后的所述待识别语音信号的短时能量和短时过零率等进行端点检测, 以确定所述待识别语音信号中的有效语音。

在本实施例中, 可以通过下述方法检测预处理后的所述待识别语音信号中的有效语音:

(1) 对预处理后的所述待识别语音信号进行加窗分帧, 得到所述待识别语音信号的语音帧 $x(n)$ 。在一个具体实施例中, 可以对预处理后的所述待识别语音信号加汉明窗, 每帧 20ms, 帧移 10ms。若预处理过程中已对待识别语音信号加窗分帧, 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。

第一提取单元 203, 用于对所述有效语音提取梅尔频率倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC) 特征参数。

提取 MFCC 特征参数的流程如下:

(1) 对每一个语音帧进行离散傅里叶变换 (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_l 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:

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

(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。

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。

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

第二提取单元 204, 用于根据所述 MFCC 特征参数, 利用预先训练好的高斯混合模型 (Gaussian Mixture Model, GMM) - 通用背景模型 (Universal Background Model, UBM) 提取所述有效语音的身份矢量 (identity-vector, iVector)。

提取 iVector 之前, 首先要用大量属于不同口音的训练数据训练出通用背景模型。通用背景模型实际上是一种高斯混合模型 (GMM), 旨在解决实际场景数据量稀缺的问题。GMM 是一种参数化的生成性模型, 具备对实际数据极强的表征力 (基于高斯分量实现)。高斯分量越多, GMM 表征力越强, 规模也越庞大, 此时负面效应逐步凸显——若想获得一个泛化能力较强的 GMM 模型, 则需要足够的数据来驱动 GMM 的参数训练, 然而实际场景中获取的语音数据甚至连分钟级都很难企及。UBM 正是解决了训练数据不足的问题。UBM 是利用大量属于不同口音的训练数据 (无关乎说话人、地域) 混合起来充分训练, 得到一个可以对语音共通特性进行表征的全局 GMM,

可大大缩减从头计算 GMM 参数所消耗的资源。通用背景模型训练完成后，只需利用单独属于每个口音的训练数据，分别对 UBM 的参数进行微调（例如通过 UBM 自适应），得到属于各个口音的 GMM。

在一个实施例中，不同口音可以是属于不同地域的口音。所述地域可以是按照行政区域来划分，例如辽宁、北京、天津、上海、河南、广东等。所述地域也可以是按照普遍的经验以口音对地区来划分，例如闽南、客家等。

提取 iVector 是基于全差异空间建模（TV）方法将 UBM 训练得出的高维 GMM 映射至低维度的全变量子空间，可突破随语音信号时长过长而提取的向量维度过大不便计算的限制，并能提升计算速度，表达出更全面的特征。GMM-UBM 中的 GMM 超矢量可以包括跟说话人本身有关的矢量特征和跟信道以及其他变化有关的矢量特征的线性叠加。

TV 模型的子空间建模形式为：

$$M = m + Tw$$

其中，M 表示语音的 GMM 超矢量，即所述 MFCC 特征参数，m 表示口音无关的 GMM 超矢量，T 表示描述差异的空间的载荷矩阵，w 表示 GMM 超矢量 M 在载荷矩阵空间下对应的低维因子表示，即 iVector。

在本实施例中，可以对提取的 iVector 行噪声补偿。在一实施例中，可以采用线性判别分析（Linear Discriminative Analysis, LDA）和类内协方差规整（Within Class Covariance Normalization, WCCN）对提取的 iVector 进行噪声补偿。

识别单元 205，用于根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

给定口音可以是一个也可以是多个。例如，若给定口音为一个，可以根据所述 iVector 计算所述待识别语音信号对该给定口音的判决得分，根据所述待识别语音信号对该给定口音的判决得分判断所述待识别语音信号是否为该给定口音。可以判断所述判决得分是否大于预设得分（例如 9 分），若所述判决得分大于预设得分，则判断所述待识别语音信号为该给定口音。

若给定口音为多个，可以根据所述 iVector 计算所述待识别语音信号对每个给定口音的判决得分，根据所述待识别语音信号对每个给定口音的判决得分判断所述语音为多个给定口音中的哪一个。可以确定对多个给定口音的判决得分中的最高得分，将所述最高得分对应的给定口音作为所述待识别语音信号所属的口音。

在本实施例中，可以利用逻辑回归（Logistic Regression）模型计算所述待识别语音信号对给定口音的判决得分。逻辑回归模型作为一个分类器，可根据待识别语音信号的 iVector 对待识别语音信号进行打分。特别地，在一具体实施例中，可以使用多类逻辑回归模型计算所述待识别语音信号对给定口音的判决得分。

假设给定口音包括口音 1、口音 2、... 口音 N 共 N 种口音，则利用 N 类逻辑回归模型计算所述待识别语音信号对给定口音的判决得分。将待识别语音信号的 iVector（记为 x_t ）输入所述 N 类逻辑回归模型，得到 N 个判决得分

s_{it} (即所述待识别语音信号对 N 种给定口音的判决得分), $s_{it}=w_i * x_t+k_i$, $i=1, \dots, N$ 。求取 N 个判决得分 s_{it} , $i=1, \dots, N$ 中的最高得分 s_{jt} , 最高得分 s_{jt} 对应的口音 j 为所述待识别语音信号所属的口音。其中, w_i 、 k_i 是 N 类逻辑回归模型的参数, w_i 为回归系数, k_i 为常数, 针对每个给定口音均会有对应的 w_i 和 k_i , w_i 、 k_i 组成 N 类逻辑回归模型的参数向量, $M=\{(w_1, k_1), (w_2, k_2), \dots, (w_N, k_N)\}$ 。

实施例二的口音识别装置 10 对待识别语音信号进行预处理; 检测预处理后的所述待识别语音信号中的有效语音; 对所述有效语音提取 MFCC 特征参数; 根据所述 MFCC 特征参数, 利用预先训练好的 GMM-UBG 提取所述有效语音的 iVector; 根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分, 根据所述判决得分得到所述待识别语音信号的口音识别结果。实施例二可以实现快速准确的口音识别。

在其他的实施例中, 第一提取单元 203 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 方法并不直接对待识别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, 若要对待识别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。若要对待识别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。

在另一实施例中, 所述识别单元 205 还可以用于根据所述口音识别结果进行声纹识别。由于说话人所生活的地域不同, 即使在都讲普通话的情况下或多或少依然会有口音的差别, 将口音识别应用到声纹识别中, 可以缩小后

续声纹识别的对象范围，得到更为准确的识别结果。

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口音识别方法实施例中的步骤，例如图 1 所示的步骤 101-105：

步骤 101，对待识别语音信号进行预处理；

步骤 102，检测预处理后的所述待识别语音信号中的有效语音；

步骤 103，对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；

步骤 104，根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；

步骤 105，根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

所述检测预处理后的所述待识别语音信号中的有效语音可以包括：

对所述待识别语音信号进行加窗分帧，得到所述待识别语音信号的语音帧；

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。

所述对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数可以包括：

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。

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 201-205：

预处理单元 201，用于对待识别语音信号进行预处理；

检测单元 202，用于检测预处理后的所述待识别语音信号中的有效语音；

第一提取单元 203，用于对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；

第二提取单元 204，用于根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；

识别单元 205，用于根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

所述检测单元 202 具体可以用于：

对所述待识别语音信号进行加窗分帧，得到所述待识别语音信号的语音帧；

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。

所述第一提取单元 203 具体可以用于：

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。

实施例四

图 3 为本申请实施例四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，例如口音识别程序。所述处理器 30 执行所述计算机可读指令 40 时实现上述口音识别方法实施例中的步骤，例如图 1 所示的步骤 101-105：

步骤 101，对待识别语音信号进行预处理；

步骤 102，检测预处理后的所述待识别语音信号中的有效语音；

步骤 103，对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；

步骤 104，根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；

步骤 105，根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

所述检测预处理后的所述待识别语音信号中的有效语音可以包括：

对所述待识别语音信号进行加窗分帧，得到所述待识别语音信号的语音帧；

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。

所述对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数可以包括：

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。

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 201-205:

预处理单元 201,用于对待识别语音信号进行预处理;

检测单元 202,用于检测预处理后的所述待识别语音信号中的有效语音;

第一提取单元 203,用于对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数;

第二提取单元 204,用于根据所述 MFCC 特征参数,利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector;

识别单元 205,用于根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分,根据所述判决得分得到所述待识别语音信号的口音识别结果。

所述检测单元 202 具体可以用于:

对所述待识别语音信号进行加窗分帧,得到所述待识别语音信号的语音帧;

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。

所述第一提取单元 203 具体可以用于:

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。

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、检测单元 202、第一提取单元 203、第二提取单元 204、识别单元 205,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包括计算机可读指令代码, 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. 一种口音识别方法，其特征在于，所述方法包括：
对待识别语音信号进行预处理；
检测预处理后的所述待识别语音信号中的有效语音；
对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；
根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；
根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。
2. 如权利要求 1 所述的方法，其特征在于，所述检测预处理后的所述待识别语音信号中的有效语音包括：
对预处理后的所述待识别语音信号进行加窗分帧，得到所述待识别语音信号的语音帧；
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。
3. 如权利要求 1 所述的方法，其特征在于，所述 MFCC 特征参数包括初始 MFCC 特征参数、一阶差分 MFCC 特征参数和二阶差分 MFCC 特征参数。
4. 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 iVector 进行噪声补偿。
5. 如权利要求 1 所述的方法，其特征在于，所述根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分包括：
将所述 iVector 输入逻辑回归模型，得到所述待识别语音信号对给定口音的判决得分。
6. 如权利要求 1 所述的方法，其特征在于，所述对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数包括：
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，所述对待识别语音信号进行预处理包括：
对所述待识别语音信号进行预加重；和
对所述待识别语音信号进行加窗分帧。

8. 一种口音识别装置，其特征在于，所述装置包括：
预处理单元，用于对待识别语音信号进行预处理；
检测单元，用于检测预处理后的所述待识别语音信号中的有效语音；
第一提取单元，用于对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；
第二提取单元，用于根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；
识别单元，用于根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。
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，所述处理器用于执行存储器中存储的计算机可读指令以实现以下步骤：
对待识别语音信号进行预处理；
检测预处理后的所述待识别语音信号中的有效语音；
对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；
根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；
根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。
10. 如权利要求 9 所述的计算机装置，其特征在于，所述检测预处理后的所述待识别语音信号中的有效语音包括：
对预处理后的所述待识别语音信号进行加窗分帧，得到所述待识别语音信号的语音帧；
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。
11. 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计算机可读指令以实现以下步骤：
对所述 iVector 进行噪声补偿。
12. 如权利要求 9 所述的计算机装置，其特征在于，所述根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分包括：
将所述 iVector 输入逻辑回归模型，得到所述待识别语音信号对给定口音的判决得分。
13. 如权利要求 9 所述的计算机装置，其特征在于，所述对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数包括：
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，所述对待识别语音信号进行预处理包括：

对所述待识别语音信号进行预加重；和

对所述待识别语音信号进行加窗分帧。

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：

对待识别语音信号进行预处理；

检测预处理后的所述待识别语音信号中的有效语音；

对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数；

根据所述 MFCC 特征参数，利用预先训练好的高斯混合模型-通用背景模型 GMM-UBM 提取所述有效语音的身份矢量 iVector；

根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分，根据所述判决得分得到所述待识别语音信号的口音识别结果。

16. 如权利要求 15 所述的存储介质，其特征在于，所述检测预处理后的所述待识别语音信号中的有效语音包括：

对预处理后的所述待识别语音信号进行加窗分帧，得到所述待识别语音信号的语音帧；

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。

17. 如权利要求 15 所述的存储介质，其特征在于，所述计算机可读指令被处理器执行时还用以实现以下步骤：

对所述 iVector 进行噪声补偿。

18. 如权利要求 15 所述的存储介质，其特征在于，所述根据所述 iVector 计算所述待识别语音信号对给定口音的判决得分包括：

将所述 iVector 输入逻辑回归模型，得到所述待识别语音信号对给定口音的判决得分。

19. 如权利要求 15 所述的存储介质，其特征在于，所述对所述有效语音提取梅尔频率倒谱系数 MFCC 特征参数包括：

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，所述对待识别语音信号

进行预处理包括：

对所述待识别语音信号进行预加重；和
对所述待识别语音信号进行加窗分帧。

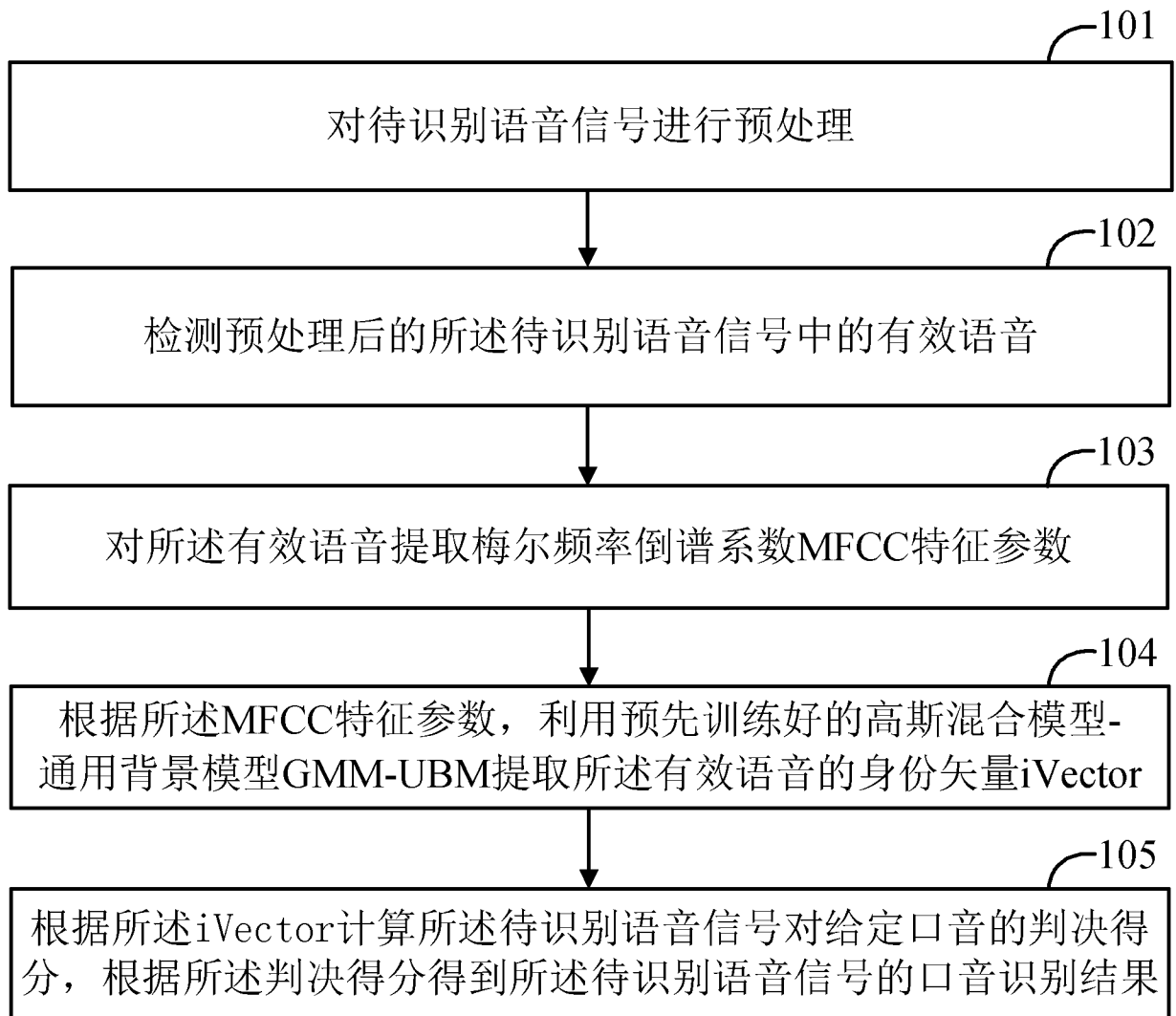


图 1

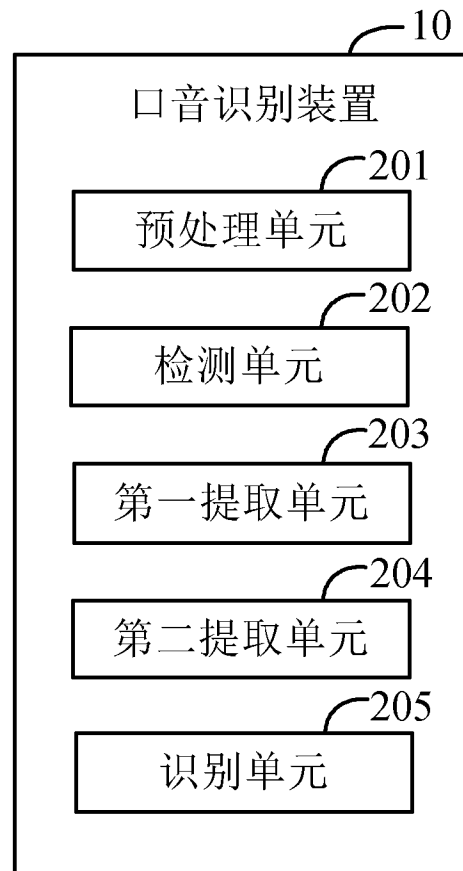


图 2

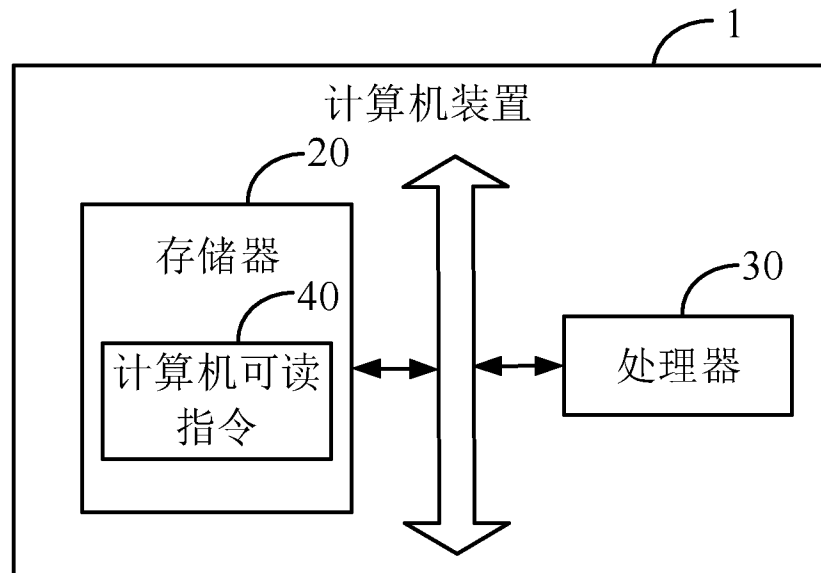


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/077512

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/02(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)

CNPAT, CNKI, WPI, EPODOC, IEEE, GOOGLE: 口音, 语音, 识别, 参数, 矢量, 得分, 梅尔频率倒谱系数, 高斯混合模型, 通用背景模型, accent, voice, recognition, parameter, vector, score, MFCC, GMM, UBM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109036437 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 18 December 2018 (2018-12-18) claims 1-10	1-10
Y	CN 103236260 A (BOE TECHNOLOGY GROUP CO., LTD. ET AL.) 07 August 2013 (2013-08-07) description, paragraphs [0005]-[0024]	1-20
Y	CN 105976819 A (GUANGZHOU SPEAKIN NETWORK TECHNOLOGY CO., LTD.) 28 September 2016 (2016-09-28) description, paragraphs [0010]-[0088]	1-20
A	CN 108369813 A (SHENZHEN H&T SMART HOME TECHNOLOGY CO., LTD.) 03 August 2018 (2018-08-03) entire document	1-20
A	CN 105679321 A (YULONG COMPUTER TELECOMMUNICATION SCIENTIFIC (SHENZHEN) CO., LTD.) 15 June 2016 (2016-06-15) 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

16 May 2019

Date of mailing of the international search report

29 May 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
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/CN2019/077512

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109036437	A	18 December 2018	None			
CN	103236260	A	07 August 2013	US	2015340027	A1	26 November 2015
				WO	2014153800	A1	02 October 2014
CN	105976819	A	28 September 2016	None			
CN	108369813	A	03 August 2018	WO	2019023877	A1	07 February 2019
CN	105679321	A	15 June 2016	None			

国际检索报告

国际申请号

PCT/CN2019/077512

A. 主题的分类 G10L 17/02 (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)) CNPAT, CNKI, WPI, EP000C, IEEE, GOOGLE: 口音, 语音, 识别, 参数, 矢量, 得分, 梅尔频率倒谱系数, 高斯混合模型, 通用背景模型, accent, voice, recognition, parameter, vector, score, MFCC, GMM, UBM																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 109036437 A (平安科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 权利要求1-10</td> <td>1-10</td> </tr> <tr> <td>Y</td> <td>CN 103236260 A (京东方科技集团股份有限公司 等) 2013年 8月 7日 (2013 - 08 - 07) 说明书第[0005]-[0024]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 105976819 A (广州势必可赢网络科技有限公司) 2016年 9月 28日 (2016 - 09 - 28) 说明书第[0010]-[0088]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108369813 A (深圳和而泰智能家居科技有限公司) 2018年 8月 3日 (2018 - 08 - 03) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 105679321 A (宇龙计算机通信科技深圳有限公司) 2016年 6月 15日 (2016 - 06 - 15) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 109036437 A (平安科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 权利要求1-10	1-10	Y	CN 103236260 A (京东方科技集团股份有限公司 等) 2013年 8月 7日 (2013 - 08 - 07) 说明书第[0005]-[0024]段	1-20	Y	CN 105976819 A (广州势必可赢网络科技有限公司) 2016年 9月 28日 (2016 - 09 - 28) 说明书第[0010]-[0088]段	1-20	A	CN 108369813 A (深圳和而泰智能家居科技有限公司) 2018年 8月 3日 (2018 - 08 - 03) 全文	1-20	A	CN 105679321 A (宇龙计算机通信科技深圳有限公司) 2016年 6月 15日 (2016 - 06 - 15) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 109036437 A (平安科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 权利要求1-10	1-10																		
Y	CN 103236260 A (京东方科技集团股份有限公司 等) 2013年 8月 7日 (2013 - 08 - 07) 说明书第[0005]-[0024]段	1-20																		
Y	CN 105976819 A (广州势必可赢网络科技有限公司) 2016年 9月 28日 (2016 - 09 - 28) 说明书第[0010]-[0088]段	1-20																		
A	CN 108369813 A (深圳和而泰智能家居科技有限公司) 2018年 8月 3日 (2018 - 08 - 03) 全文	1-20																		
A	CN 105679321 A (宇龙计算机通信科技深圳有限公司) 2016年 6月 15日 (2016 - 06 - 15) 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 5月 16日		国际检索报告邮寄日期 2019年 5月 29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		授权官员 田民丽 电话号码 86-(10)-53961355																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/077512

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	109036437	A	2018年 12月 18日	无			
CN	103236260	A	2013年 8月 7日	US	2015340027	A1	2015年 11月 26日
				WO	2014153800	A1	2014年 10月 2日
CN	105976819	A	2016年 9月 28日	无			
CN	108369813	A	2018年 8月 3日	WO	2019023877	A1	2019年 2月 7日
CN	105679321	A	2016年 6月 15日	无			