(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
2018 年 1 月 18 日 (18.01.2018)

WIPO | PCT

(10) 国际公布号
WO 2018/010683 A1

(51) 国际专利分类号:
G10L 17/02 (2013.01)

(21) 国际申请号: PCT/CN2017/092892

(22) 国际申请日: 2017 年 7 月 14 日 (14.07.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610560366.3 2016年7月15日 (15.07.2016) CN

(71) 申请人: 腾讯科技（深圳）有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。

(72) 发明人: 李为 (LI, Wei); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 钱柄桦 (QIAN, Binghua); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 金星明 (JIN, Xingming); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 李科 (LI, Ke); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 吴富章 (WU, Fuzhang); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 吴永坚 (WU, Yongjian); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 黄飞跃 (HUANG, Feiyue); 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区花城大道85号3901房, Guangdong 510623 (CN)。

(54) **Title:** IDENTITY VECTOR GENERATING METHOD, COMPUTER APPARATUS AND COMPUTER READABLE STORAGE MEDIUM

(54) **发明名称**: 身份向量生成方法、计算机设备和计算机可读存储介质

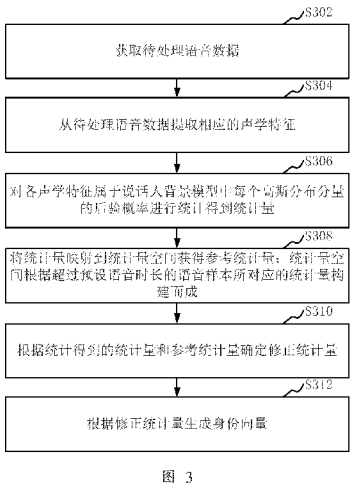

图 3

S302 Acquire voice data to be processed
S304 Extract corresponding acoustic features from the voice data to be processed
S306 Perform statistical calculations on the posterior probability for the acoustic features belonging to each Gaussian distribution component in the speaker's background model so as to obtain a statistic
S308 Map the statistics onto a statistics space to acquire reference statistics; the statistics space being built according to the statistics that the voice samples exceeding a preset voice duration correspond to
S310 Determine correction statistics according to the statistics obtained by performing statistical calculations and the reference statistic
S312 Generate an i-vector according to the correction statistics

(57) Abstract: An identity vector (i-vector) generating method, comprises: acquiring voice data to be processed (S302, S402); extracting corresponding acoustic features from the voice data to be processed (S304, S404); performing statistical calculations on the posterior probability for the acoustic features belonging to each Gaussian distribution component in the speaker's background model so as to obtain a statistic (S306); mapping the statistic onto a statistic space to acquire a reference statistic; the statistic space being built according to the statistic that the voice samples exceeding a preset voice duration correspond to (S308); determining a correction statistic according to the statistic obtained by performing statistical calculations and the reference statistic (S310); and generating an i-vector according to the correction statistic (S312). The method can compensate for the estimated deviation of the statistic which results from cases where the voice duration for the voice data to be processed is too short and the voice is sparse, thereby improving the identity recognition performance of the i-vector.

(57) 摘要: 一种身份向量生成方法，包括：获取待处理语音数据（S302，S402）；从所述待处理语音数据提取相应的声学特征（S304，S404）；对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量（S306）；将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成（S308）；根据统计得到的所述统计量和所述参考统计量确定修正统计量（S310）；及根据所述修正统计量生成身份向量（S312）。该方法能够补偿因待处理语音数据的语音时长过短和语音稀疏的情况下导致的统计量偏估，提高身份向量的身份识别性能。

WO 2018/010683 A1

[见续页]

(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))。

身份向量生成方法、计算机设备和计算机可读存储介质

本申请要求于 2016 年 7 月 15 日提交中国专利局，申请号为 201610560366.3，发明名称为“身份向量生成方法和装置”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

本申请涉及计算机技术领域，特别是涉及一种身份向量生成方法、计算机设备和计算机可读存储介质。

背景技术

说话人身份识别是一种重要的身份识别手段，采集用户说出一段语音，并将采集的语音进行预处理、特征提取、建模和参数估计等一系列操作后，将语音映射为一段定长的可以表达说话人语音特征的向量，该向量称为身份向量(i-vector)。身份向量可以良好地表达相应语音中包括的说话人身份信息。

目前在生成语音数据的身份向量时，需要提取出其声学特征，并基于高斯混合模型形式的说话人背景模型，统计各声学特征属于说话人背景模型中每个高斯分布分量的后验概率的统计量，进而基于该统计量生成身份向量。

然而，目前生成身份向量的方式，在语音数据语音长度比较短或者语音比较稀疏的情况下，会导致身份向量的身份识别性能降低。

发明内容

根据本申请的各种实施例，提供一种身份向量生成方法、计算机设备和计算机可读存储介质。

一种身份向量生成方法，包括:

获取待处理语音数据;

从所述待处理语音数据提取相应的声学特征;

对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；

将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；

根据统计得到的所述统计量和所述参考统计量确定修正统计量；及

根据所述修正统计量生成身份向量。

一种计算机设备，包括存储器和处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

获取待处理语音数据；

从所述待处理语音数据提取相应的声学特征；

对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；

将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；

根据统计得到的所述统计量和所述参考统计量确定修正统计量；及

根据所述修正统计量生成身份向量。

一个或多个存储有计算机可读指令的非易失性的计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理语音数据；

从所述待处理语音数据提取相应的声学特征；

对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；

将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；

根据统计得到的所述统计量和所述参考统计量确定修正统计量；及

根据所述修正统计量生成身份向量。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例的技术方案，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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其它的附图。

图 1 为一个实施例中说话人识别系统的应用环境图；

图 2A 为一个实施例中服务器的内部结构示意图；

图 2B 为一个实施例中终端的内部结构示意图；

图 3 为一个实施例中身份向量生成方法的流程示意图；

图 4 为另一个实施例中身份向量生成方法的流程示意图；

图 5 为一个实施例中构建统计量空间的步骤的流程示意图；

图 6 为一个实施例中计算机设备的结构框图；

图 7 为一个实施例中统计量生成模块的结构框图；

图 8 为另一个实施例中计算机设备的结构框图；

图 9 为再一个实施例中计算机设备的结构框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

可以理解，本申请所使用的术语“第一”、“第二”等可在本文中用于描述各种元件，但这些元件不受这些术语限制。这些术语仅用于将第一个元件与另一个元件区分。第一零阶统计量和第二零阶统计量两者都是零阶统计量，

但其不是同一零阶统计量。

图 1 为一个实施例中说话人识别系统的应用环境图。如图 1 所示，该系统包括通过网络连接的终端 110 和服务器 120。终端 110 可用于采集待验证语音数据，并采用本申请中的身份向量生成方法生成待验证身份向量，并将待验证身份向量发送到服务器 120。服务器 120 可收集目标说话人类别的语音数据，并采用本申请中的身份向量生成方法生成目标说话人身份向量。服务器 120 可用于计算待验证身份向量和目标说话人身份向量的相似度；根据相似度进行说话人身份验证。服务器 120 可用于向终端 110 反馈身份验证结果。

图2A为一个实施例中服务器的内部结构示意图。如图2A所示，该服务器包括通过系统总线连接的处理器、非易失性存储介质、内存储器和网络接口。其中，该服务器的非易失性存储介质存储有操作系统、数据库和计算机可读指令，该计算机可读指令被处理器执行时，可使得处理器实现一种身份向量生成方法。该服务器的处理器用于提供计算和控制能力，支撑整个服务器的运行。该服务器的内存储器中可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，可使得处理器执行一种身份向量生成方法。该服务器的网络接口用于与终端连接通信。服务器可以用独立的服务器或者是多个服务器组成的服务器集群来实现。本领域技术人员可以理解，图2A中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的服务器的限定，具体的服务器可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

图2B为一个实施例中终端的内部结构示意图。如图2B所示，该终端包括通过系统总线连接的处理器、非易失性存储介质、内存储器、网络接口和声音采集装置。其中，终端的非易失性存储介质存储有操作系统，还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，可使得处理器实现一种身份向量生成方法。该处理器用于提供计算和控制能力，支撑整个终端的运行。终端中的内存储器中可储存有计算机可读指令，该计算机可读指令被处

理器执行时，可使得处理器执行一种身份向量生成方法。网络接口用于与服务器进行网络通信。该终端可以是手机、平板电脑或者个人数字助理或穿戴式设备等。本领域技术人员可以理解，图2B中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的终端的限定，具体的终端可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

图 3 为一个实施例中身份向量生成方法的流程示意图。本实施例以该方法应用于服务器 120 来举例说明。参照图 3，该方法具体包括如下步骤：

S302，获取待处理语音数据。

其中，待处理语音数据是指需要对其进行一系列处理以生成相应的身份向量的语音数据。语音数据是在说话人将语音说出后由声音采集设备所采集的声音进行保存而形成的数据。待处理语音数据可以包括待验证语音数据和目标说话人类别的语音数据，其中待验证语音数据是指未知说话人类别并需要判断是否属于目标说话人类别的语音数据；目标说话人类别是已知的说话人类别，是目标说话人说话形成的语音数据所构成的类别。

S304，从待处理语音数据提取相应的声学特征。

具体地，服务器可以对待处理语音数据进行预处理，比如滤除噪声或者统一语音格式等，再从经过预处理的待处理语音数据提取相应的声学特征向量。声学特征向量是指反映声学特性的声学特征所构成的向量。声学特征向量包括一系列的声学特征，该声学特征可以是梅尔倒谱系数（MFCC，Mel Frequency Cepstrum Coefficient）或者线性预测倒谱系数（LPCC）。

S306，对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量。

其中，说话人背景模型是采用一系列的语音样本训练得到的高斯混合模型，用来训练表示与说话人无关的特征分布。其中高斯混合模型是固定数量的高斯分布分量叠加而成的数学模型。说话人背景模型可通过 EM 算法（Expectation Maximization Algorithm，译为期望最大化算法）训练得到。说

话人背景模型可采用 GMM-UBM（Gaussian Mixture Model-Universal Background Model，高斯混合模型-通用背景模型）。

在一个实施例中，说话人背景模型可用如下公式（1）表示：

$$P(x) = \sum_{c=1}^{C} a_c N(x|\mu_c, \Sigma_c) \qquad \text{公式（1）}$$

其中，x表示语音样本；C 是高斯混合模型所包括高斯分布分量的总数，c 表示高斯混合模型所包括的高斯分布分量的序号；$N(x|\mu_c, \Sigma_c)$表示第 c 个高斯分布分量；a_c是第 c 个高斯分布分量的系数；μ_c是第 c 个高斯分布分量的均值；Σ_c是第 c 个高斯分布分量的方差。

在一个实施例中，声学特征向量可表达为：$\{y_1, y_2 \ldots y_L\}$。该声学特征向量包括 L 个声学特征，每个声学特征可表示为y_t，其中，$t \in [1, L]$。在一个实施例中，声学特征向量中各声学特征属于说话人背景模型中每个高斯分布分量的后验概率可表示为：$P(c|y_t, \Omega)$。其中，Ω表示说话人背景模型。$P(c|y_t, \Omega)$表示在说话人背景模型Ω和声学特征y_t已观测到的情况下声学特征y_t属于第c个高斯分布分量的后验概率。服务器可基于后验概率$P(c|y_t, \Omega)$进行统计得到统计量。

S308，将统计量映射到统计量空间获得参考统计量；统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成。

其中，统计量空间是一种向量空间，统计量空间根据语音样本所对应的与上述统计得到的统计量同类型的统计量构建而成，该用来构建统计量空间的语音样本的语音时长超过预设语音时长，预设语音时长比如 30 秒。用来构建统计量空间的语音样本可以是从用于训练说话人背景模型的语音样本中筛选出的超过预设语音时长的语音样本。将统计得到的统计量映射到统计量空间后得到参考统计量，该参考统计量是根据超过预设语音时长的语音样本所对应的统计量确定的先验统计量。

S310，根据统计得到的统计量和参考统计量确定修正统计量。

其中，修改统计量是利用参考统计量修正统计得到的统计量后得到的统计量，该统计量结合了先验的统计量和后验的统计量。

S312，根据修正统计量生成身份向量。

具体地，在得到修正统计量后，可以利用修正统计量并采用常规的生成身份向量的方式来生成身份向量。

上述身份向量生成方法，统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成，在对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量后，将该统计量映射到该统计量空间中，得到的参考统计量是先验统计量。利用先验统计量来对统计得到的统计量进行修正得到修正统计量，该修正统计量能够补偿因待处理语音数据的语音时长过短和语音稀疏的情况下导致的统计量偏估，提高身份向量的身份识别性能。

图 4 为另一个实施例中身份向量生成方法的流程示意图。如图 4 所示，该身份向量生成方法包括如下步骤：

S402，获取待处理语音数据。

S404，从待处理语音数据提取相应的声学特征。

S406，对应于说话人背景模型中的每个高斯分布分量，分别统计各声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量。

具体地，对应于说话人背景模型Ω中的每个高斯分布分量c，分别统计各声学特征$\boldsymbol{y}_t$属于相应高斯分布分量c的后验概率$P(c|\boldsymbol{y}_t,\Omega)$的总和，将该总和作为相应高斯分布分量$c$所对应的第一零阶统计量。

更具体地，可采用如下公式（2）计算对应于高斯分布分量c的第一零阶统计量$N_c(u)$：

$$N_c(u)=\sum_{t=1}^{L}P(c|\boldsymbol{y}_t,\Omega) \qquad \text{公式（2）}$$

其中，u表示待处理语音数据；$N_c(u)$表示待处理语音数据u对应于高斯分布分量c的第一零阶统计量；$\boldsymbol{y}_t$表示声学特征向量的 L 个声学特征中第 t 个声学特征；$P(c|\boldsymbol{y}_t,\Omega)$表示在说话人背景模型$\Omega$和声学特征$\boldsymbol{y}_t$已观测到的情况下声学特征$\boldsymbol{y}_t$属于第$c$个高斯分布分量的后验概率。

S408，对应于说话人背景模型中的每个高斯分布分量，分别将各声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

其中，S404 和 S406 包括于上述步骤 S304。具体地，对应于说话人背景模型中的每个高斯分布分量c，分别将各声学特征$\boldsymbol{y}_t$以该声学特征$\boldsymbol{y}_t$属于相应高斯分布分量c的后验概率$P(c|\boldsymbol{y}_t,\Omega)$为权重计算加权和，将该加权和作为应高斯分布分量$c$所对应的第一一阶统计量。

更具体地，可采用如下公式（3）计算对应于高斯分布分量c的第一一阶统计量$F_c(u)$:

$$F_c(u)=\sum_{t=1}^{L}P(c|\boldsymbol{y}_t,\Omega)\boldsymbol{y}_t \qquad \text{公式（3）}$$

其中，u表示待处理语音数据；$F_c(u)$表示待处理语音数据u对应于高斯分布分量c的第一一阶统计量；$\boldsymbol{y}_t$表示声学特征向量的 L 个声学特征中第 t 个声学特征；$P(c|\boldsymbol{y}_t,\Omega)$表示在说话人背景模型$\Omega$和声学特征$\boldsymbol{y}_t$已观测到的情况下声学特征$\boldsymbol{y}_t$属于第$c$个高斯分布分量的后验概率。

S410，将第一零阶统计量和第一一阶统计量映射到统计量空间，获得对应说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成。

具体地，将第一零阶统计量$N_c(u)$和第一一阶统计量$F_c(u)$映射到统计量空间 **H**，得到对应说话人背景模型中每个高斯分布分量c的参考一阶统计量$\boldsymbol{F}_c^{ref}(u)$和相应参考零阶统计量$N_c^{ref}(u)$的第二商：$\boldsymbol{F}_c^{ref}(u)/N_c^{ref}(u)$。

S412，将第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的第二商加权求和，得到对应说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

具体地，可采用如下公式（4）计算对应于高斯分布分量c的修正统计量:

$$\frac{\widetilde{F_c}(u)}{\widetilde{N_c}(u)}=R1*\frac{F_c^{ref}(u)}{N_c^{ref}(u)}+R2*\frac{F_c(u)}{N_c(u)} \qquad \text{公式（4）}$$

其中，$\widetilde{F_c}(u)$表示对应于高斯分布分量c的修正一阶统计量；$\widetilde{N_c}(u)$表示对

应于高斯分布分量c的修正零阶统计量；$R1$和$R2$是权重；$\frac{F_c^{ref}(u)}{N_c^{ref}(u)}$表示对应于高斯分布分量$c$的第二商；$\frac{F_c(u)}{N_c(u)}$表示对应于高斯分布分量$c$的第三商。可限定$R1$和$R2$的和为 1。

在一个实施例中，加权求和中，第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，第二商的权重为可调参数除以相应高斯分布分量的第一零阶统计量与可调参数的和。

具体地，可采用如下公式（5）计算对应于高斯分布分量c的修正统计量：

$$\frac{\widetilde{F_c}(u)}{\widetilde{N_c}(u)} = \frac{q}{N_c(u)+q} * \frac{F_c^{ref}(u)}{N_c^{ref}(u)} + \frac{N_c(u)}{N_c(u)+q} * \frac{F_c(u)}{N_c(u)} \qquad \text{公式（5）}$$

其中，第三商$\frac{F_c(u)}{N_c(u)}$的权重为$\frac{N_c(u)}{N_c(u)+q}$，是相应高斯分布分量 c 的第一零阶统计量$N_c(u)$除以相应的第一零阶统计量$N_c(u)$与可调参数q的和；第二商$\frac{F_c^{ref}(u)}{N_c^{ref}(u)}$的权重为$\frac{q}{N_c(u)+q}$，是可调参数$q$除以相应高斯分布分量 c 的第一零阶统计量$N_c(u)$与可调参数q的和。q取 0.4~1 时可达到很好的效果。本实施例中，通过调整可调参数，可以针对不同环境进行差异性调整，增加鲁棒性。

S414，根据修正统计量生成身份向量。

具体地，当$\widetilde{N_c}(u)$=$N_c(u)$时可求得$\widetilde{F_c}(u)$。

按照如下公式（6）定义说话人背景模型的均值超向量$\boldsymbol{m}$：

$$\mathbf{m} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_C \end{pmatrix} \qquad \text{公式（6）}$$

其中，$\boldsymbol{\mu}_1$、$\boldsymbol{\mu}_2$……$\boldsymbol{\mu}_C$分别是说话人背景模型各高斯分布分量的均值。

按照如下公式（7）定义对角矩阵形式的修正零阶统计量矩阵$\widetilde{\boldsymbol{N}}(u)$：

$$\widetilde{\boldsymbol{N}}(u) = \begin{bmatrix} \widetilde{N}_1(u) & & & \\ & \widetilde{N}_2(u) & & \\ & & \ddots & \\ & & & \widetilde{N}_C(u) \end{bmatrix} \qquad \text{公式（7）}$$

其中，$\widetilde{N}_1(u)$、$\widetilde{N}_2(u)$……$\widetilde{N}_C(u)$分别是对应于说话人背景模型各高斯分布

分量的修正零阶统计量。

按照如下公式（8）定义修正一阶统计量矩阵$\widetilde{\boldsymbol{F}}(u)$:

$$\widetilde{\boldsymbol{F}}(u)=\begin{pmatrix}\widetilde{\boldsymbol{F}}_1(u)\\ \widetilde{\boldsymbol{F}}_2(u)\\ \vdots\\ \widetilde{\boldsymbol{F}}_C(u)\end{pmatrix} \qquad \text{公式（8）}$$

其中，$\widetilde{\boldsymbol{F}}_1(u)$、$\widetilde{\boldsymbol{F}}_2(u)\ldots\ldots\widetilde{\boldsymbol{F}}_C(u)$分别是对应于说话人背景模型各高斯分布分量的修正一阶统计量。

在一个实施例中，可根据如下公式（9）计算身份向量$\tilde{\boldsymbol{\omega}}(u)$:

$$\tilde{\boldsymbol{\omega}}(u)=\left(\boldsymbol{I}+\boldsymbol{T}^t\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{N}}(u)\boldsymbol{T}\right)^{-1}\boldsymbol{T}^t\boldsymbol{\Sigma}^{-1}\left(\widetilde{\boldsymbol{F}}(u)-\widetilde{\boldsymbol{N}}(u)\boldsymbol{m}\right) \qquad \text{公式（9）}$$

其中，$\boldsymbol{I}$表示单位矩阵；$\boldsymbol{T}$表示已知的全因子矩阵（Total Factor Matrix）；t表示转置；$\boldsymbol{\Sigma}$表示对角矩阵形式的协方差矩阵，$\boldsymbol{\Sigma}$的对角元素是各高斯分布分量的协方差；$\boldsymbol{m}$表示说话人背景模型的均值超向量；$\widetilde{\boldsymbol{N}}(u)$表示修正零阶统计量矩阵；$\widetilde{\boldsymbol{F}}(u)$表示修正一阶统计量矩阵。

在一个实施例中，可对上述公式（9）进行变换，将涉及矩阵$\widetilde{\boldsymbol{F}}(u)$和$\widetilde{\boldsymbol{N}}(u)$的计算变换为涉及$\frac{\widetilde{F_c}(u)}{\widetilde{N_c}(u)}$和$\widetilde{N_c}(u)$的计算，而$\widetilde{N_c}(u)$=$N_c(u)$。本实施例中在得到$\frac{\widetilde{F_c}(u)}{\widetilde{N_c}(u)}$后可直接用来计算身份向量，不必构建矩阵$\widetilde{\boldsymbol{F}}(u)$和$\widetilde{\boldsymbol{N}}(u)$，简化计算。

本实施例中，利用第一一阶统计量和第一零阶统计量可以更加准确地反映声学特征的特性，便于计算出准确的修正统计量。由于一阶统计量与相应零阶统计量的商基本保持在稳定的范围内，可以在确定修正统计量时直接进行线性加和，减少计算量。

图 5 为一个实施例中构建统计量空间的步骤的流程示意图。参照图 5，构建统计量空间的步骤具体包括如下步骤

S502，获取超过预设语音时长的语音样本。

具体地，可从用于训练说话人背景模型的语音样本中筛选出语音时长超过预设语音时长的语音样本。

S504，按照语音样本中说话人类别统计对应于说话人背景模型中的每个

高斯分布分量的第二零阶统计量和第二一阶统计量。

具体地，若获取的语音样本共有 S 个说话人类别，对于第 s 个说话人类别，参照上述公式（2）和（3），分别统计对应于每个高斯分布分量 c 的第二零阶统计量$\widetilde{N_c(s)}$和第二一阶统计量$\widetilde{F_c(s)}$。

S506，计算第二一阶统计量和相应的第二零阶统计量的第一商。

具体地，对于每个说话类别 s，分别计算对应于说话人背景模型中每个高斯分布分量 c 的第二一阶统计量$\widetilde{F_c(s)}$和相应的第二零阶统计量$\widetilde{N_c(s)}$的第一商$\widetilde{F_c(s)}/\widetilde{N_c(s)}$。

S508，根据计算出的第一商构建统计量空间。

具体地，可将对于每个说话类别 s 且对应于说话人背景模型中每个高斯分布分量 c 的第一商，按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

本实施例中，基于第二一阶统计量和相应的第二零阶统计量的第一商建立统计量空间，由于一阶统计量与相应零阶统计量的商基本保持在稳定的范围内，便于将第一零阶统计量和第一一阶统计量映射到统计量空间的计算，提高计算效率。

在一个实施例中，S508 包括：将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

具体地，可按照如下公式（10）确定表征统计量空间的矩阵$\mathbf{H}$:

$$\mathbf{H}=\left[\left(\frac{\widetilde{F(1)}}{\widetilde{N(1)}}-\boldsymbol{m}\right),\left(\frac{\widetilde{F(2)}}{\widetilde{N(2)}}-\boldsymbol{m}\right),\ldots,\left(\frac{\widetilde{F(S)}}{\widetilde{N(S)}}-\boldsymbol{m}\right)\right] \qquad \text{公式（10）}$$

其中，$\boldsymbol{m}$表示说话人背景模型的均值超向量；$\widetilde{F(s)}$，$s\in[1,S]$，表示第 s 个说话人类别对应的第二一阶统计量矩阵，$\widetilde{N(s)}$表示各第 s 个说话人类别的对应于说话人背景模型各高斯分布分量 c 的第二零阶统计量。

$\frac{\widetilde{F(s)}}{\widetilde{N(s)}}$可表示为如下形式：

$$\overline{\frac{F(s)}{N(s)}} = \begin{pmatrix} \overline{F_1(s)}/\overline{N_1(s)} \\ \overline{F_2(s)}/\overline{N_2(s)} \\ \vdots \\ \overline{F_C(s)}/\overline{N_C(s)} \end{pmatrix}$$

因此，上述公式（10）可变形为如下公式（11）

$$\mathbf{H} = \begin{bmatrix} (\overline{F_1(1)}/\overline{N_1(1)} - \mu_1), (\overline{F_1(2)}/\overline{N_1(2)} - \mu_1), \ldots, (\overline{F_1(S)}/\overline{N_1(S)} - \mu_1) \\ (\overline{F_2(1)}/\overline{N_2(1)} - \mu_2), (\overline{F_2(2)}/\overline{N_2(2)} - \mu_2), \ldots, (\overline{F_2(S)}/\overline{N_2(S)} - \mu_2) \\ \vdots \\ (\overline{F_C(1)}/\overline{N_C(1)} - \mu_C), (\overline{F_C(2)}/\overline{N_C(2)} - \mu_C), \ldots, (\overline{F_C(S)}/\overline{N_C(S)} - \mu_C) \end{bmatrix} \quad \text{公式（11）}$$

本实施例中，将计算出的第一商减去相应高斯分布分量的均值得到相应的差值，从而将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵，使得构建出的统计量空间中心大致在统计量空间的原点处，便于计算，提高计算效率。

在一个实施例中，步骤 S410 具体包括：获取统计量空间的正交基向量；求取正交基向量的映射系数，正交基向量与映射系数的乘积加上相应高斯分布分量的均值后，与相应高斯分布分量的第三商之间的二范数距离最小化；将正交基向量乘以映射系数后加上相应高斯分布分量的均值，得到对应说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商。

具体地，统计量空间可通过特征值分解得到统计量空间的一组正交基向量F^{eigen}。可定义如下公式（12）的优化函数：

$$\min_{\varphi(u)} N_c(u) \left\| \frac{F_c(u)}{N_c(u)} - \mu_c - F^{eigen} * \varphi(u) \right\|_2^2 \qquad \text{公式（12）}$$

其中，$N_c(u)$表示对应于高斯分布分量c的第一零阶统计量；$F_c(u)$表示对应于高斯分布分量c的第一一阶统计量；$\frac{F_c(u)}{N_c(u)}$表示对应于高斯分布分量c的第三商；μ_c表示对应于高斯分布分量c的均值；F^{eigen}表示统计量空间 H 的正交基向量；$\varphi(u)$表示映射系数。

优化如公式（12）的优化函数，得到的最优的映射系数$\boldsymbol{\varphi}(u)$如下公式（13）：

$$\boldsymbol{\varphi}(u)=\left[\sum_{c=1}^{C}N_c(u)\left(\boldsymbol{F}^{eigen}\right)^t\boldsymbol{F}^{eigen}\right]^{-1}\sum_{c=1}^{C}\left(\boldsymbol{F}^{eigen}\right)^t\left(\boldsymbol{F}_c(u)-N_c(u)\boldsymbol{\mu}_c\right) \quad \text{公式（13）}$$

进一步地，按照如下公式（14）计算对应说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商：

$$\frac{{F_c}^{ref}(u)}{{N_c}^{ref}(u)}=\boldsymbol{F}^{eigen}\boldsymbol{\varphi}(u)+\boldsymbol{\mu}_c \quad \text{公式（14）}$$

本实施例中，可实现准确地将第一零阶统计量和第一一阶统计量映射到统计量空间。

在一个实施例中，待处理语音数据包括待验证语音数据和目标说话人类别的语音数据；步骤S312包括：根据与待验证语音数据对应的修正统计量生成待验证身份向量；根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身份向量。该身份向量生成方法还包括：计算待验证身份向量和目标说话人身份向量的相似度；根据相似度进行说话人身份验证。

具体地，说话人身份识别可以应用于多种需要认证未知用户身份的场景。说话人身份识别分为线下（off-line）和线上（on-line）两个阶段：线下阶段需要收集大量的非目标说话人类别的语音样本用于训练说话人身份识别系统，说话人身份识别系统包括身份向量提取模块与身份向量规整模块。

线上阶段又分为两个阶段：注册阶段与识别阶段。在注册阶段中，需要获取目标说话人的语音数据，将该语音数据进行预处理、特征提取与模型训练后，映射为一段定长的身份向量，该已知身份向量即是表征目标说话人身份的一个模型。而在识别阶段中，获取一段身份未知的待验证语音，将该待验证语音同样经过预处理、特征提取与模型训练后，映射为一段待验证身份向量。

目标说话人类别的身份向量与识别阶段的待验证身份向量接下来在相似度计算模块中计算相似度，将相似度与预先人工设定的一个门限值进行比较，若相似度大于等于门限值，则可判定待验证语音对应的身份与目标说话人身份匹配，身份验证通过。若相似度小于门限值，则可判定待验证语音对应的

身份与目标说话人身份不匹配，身份验证未通过。相似度可采用余弦相似度、皮尔森相关系数或者欧氏距离等。

本实施例中，即使是语音时长很短的语音数据，通过本实施例的身份向量生成方法，依然可以生成身份识别性能较高的身份向量，不需要说话人说出太长的语音，使得短时文本无关说话人识别能够广泛推广。

图 6 为一个实施例中计算机设备 600 的结构框图。计算机设备 600 可用作服务器，也可以用作终端。服务器的内部结构可对应于如图 2A 所示的结构，终端的内部结构可对应于如图 2B 所示的结构。下述每个模块可全部或部分通过软件、硬件或其组合来实现。

如图 6 所示，计算机设备 600 包括声学特征提取模块 610、统计量生成模块 620、映射模块 630、修正统计量确定模块 640 和身份向量生成模块 650。

声学特征提取模块 610，用于获取待处理语音数据；从待处理语音数据提取相应的声学特征。

统计量生成模块 620，用于对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量。

映射模块 630，用于将统计量映射到统计量空间获得参考统计量；统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成。

修正统计量确定模块 640，用于根据统计得到的统计量和参考统计量确定修正统计量。

身份向量生成模块 650，用于根据修正统计量生成身份向量。

上述计算机设备 600，统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成，在对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量后，将该统计量映射到该统计量空间中，得到的参考统计量是先验统计量。利用先验统计量来对统计得到的统计量进行修正得到修正统计量，该修正统计量能够补偿因待处理语音数据的语音时长过短和语音稀疏的情况下导致的统计量偏估，提高身份向量的身份识别性能。

图 7 为一个实施例中统计量生成模块 620 的结构框图。本实施例中，统计得到的统计量包括第一零阶统计量和第一一阶统计量；统计量生成模块 620 包括：第一零阶统计量生成模块 621 和第一一阶统计量生成模块 622。

第一零阶统计量生成模块 621，用于对应于说话人背景模型中的每个高斯分布分量，分别统计各声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量。

第一一阶统计量生成模块 622，用于对应于说话人背景模型中的每个高斯分布分量，分别将各声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

图 8 为另一个实施例中计算机设备 600 的结构框图。计算机设备 600 还包括：统计量统计模块 660 和统计量空间构建模块 670。

统计量统计模块 660，用于获取超过预设语音时长的语音样本；按照语音样本中说话人类别统计对应于说话人背景模型中的每个高斯分布分量的第二零阶统计量和第二一阶统计量。

统计量空间构建模块 670，用于计算第二一阶统计量和相应的第二零阶统计量的第一商；根据计算出的第一商构建统计量空间。

本实施例中，基于第二一阶统计量和相应的第二零阶统计量的第一商建立统计量空间，由于一阶统计量与相应零阶统计量的商基本保持在稳定的范围内，便于将第一零阶统计量和第一一阶统计量映射到统计量空间的计算，提高计算效率。

在一个实施例中，统计量空间构建模块 670 还用于将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

本实施例中，将计算出的第一商减去相应高斯分布分量的均值得到相应的差值，从而将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵，使得构建出的统计量空间中心大致在统计量空间的原点处，便于计算，提高计算效率。

在一个实施例中，参考统计量包括对应说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；修正统计量确定模块 640 还用于将第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的第二商加权求和，得到对应说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

在一个实施例中，修正统计量确定模块 640 用于加权求和时，第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，第二商的权重为可调参数除以相应高斯分布分量的第一零阶统计量与可调参数的和。本实施例中，通过调整可调参数，可以针对不同环境进行差异性调整，增加鲁棒性。

在一个实施例中，映射模块 630 还用于获取统计量空间的正交基向量；求取正交基向量的映射系数，正交基向量与映射系数的乘积加上相应高斯分布分量的均值后，与相应高斯分布分量的第三商之间的二范数距离最小化；将正交基向量乘以映射系数后加上相应高斯分布分量的均值，得到对应说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商。

在一个实施例中，待处理语音数据包括待验证语音数据和目标说话人类别的语音数据；身份向量生成模块 650 还用于根据与待验证语音数据对应的修正统计量生成待验证身份向量；根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身份向量。

图 9 为再一个实施例中计算机设备 600 的结构框图。本实施例中计算机设备 600 还包括：说话人身份验证模块 680，用于计算待验证身份向量和目标说话人身份向量的相似度；根据相似度进行说话人身份验证。

本实施例中，即使是语音时长很短的语音数据，通过本实施例的身份向量生成方法，依然可以生成身份识别性能较高的身份向量，不需要说话人说出太长的语音，使得短时文本无关说话人识别能够广泛推广。

在一个实施例中，提供了一种计算机设备，包括存储器和处理器，所述

存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：获取待处理语音数据；从所述待处理语音数据提取相应的声学特征；对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；根据统计得到的所述统计量和所述参考统计量确定修正统计量；及根据所述修正统计量生成身份向量。

在一个实施例中，统计得到的所述统计量包括第一零阶统计量和第一一阶统计量；所述对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量包括：对应于说话人背景模型中的每个高斯分布分量，分别统计各所述声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量；及对应于说话人背景模型中的每个高斯分布分量，分别将各所述声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

在一个实施例中，计算机可读指令还使得处理器执行以下步骤：获取超过预设语音时长的语音样本；按照所述语音样本中说话人类别统计对应于说话人背景模型中的每个高斯分布分量的第二零阶统计量和第二一阶统计量；计算所述第二一阶统计量和相应的第二零阶统计量的第一商；及根据计算出的第一商构建统计量空间。

在一个实施例中，所述根据计算出的第一商构建统计量空间包括：将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；及将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

在一个实施例中，所述参考统计量包括对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；所述根据统计得到的所述统计量和所述参考统计量确定修正统计量包括：将所述第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的所述第

二商加权求和，得到对应所述说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

在一个实施例中，所述加权求和中，所述第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，所述第二商的权重为所述可调参数除以所述相应高斯分布分量的第一零阶统计量与所述可调参数的和。

在一个实施例中，所述将所述统计量映射到统计量空间获得参考统计量包括：获取所述统计量空间的正交基向量；求取所述正交基向量的映射系数，所述正交基向量与所述映射系数的乘积加上相应高斯分布分量的均值后，与相应高斯分布分量的第三商之间的二范数距离最小化；及将所述正交基向量乘以所述映射系数后加上相应高斯分布分量的均值，得到对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商。

在一个实施例中，所述待处理语音数据包括待验证语音数据和目标说话人类别的语音数据；所述根据所述修正统计量生成身份向量包括：根据与所述待验证语音数据对应的修正统计量生成待验证身份向量；及根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身份向量；所述计算机可读指令还使得所述处理器执行以下步骤：计算所述待验证身份向量和所述目标说话人身份向量的相似度；及根据所述相似度进行说话人身份验证。

上述计算机设备，统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成，在对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量后，将该统计量映射到该统计量空间中，得到的参考统计量是先验统计量。利用先验统计量来对统计得到的统计量进行修正得到修正统计量，该修正统计量能够补偿因待处理语音数据的语音时长过短和语音稀疏的情况下导致的统计量偏估，提高身份向量的身份识别性能。

在一个实施例中，提供了一个或多个存储有计算机可读指令的非易失性

的计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：获取待处理语音数据；从所述待处理语音数据提取相应的声学特征；对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；根据统计得到的所述统计量和所述参考统计量确定修正统计量；及根据所述修正统计量生成身份向量。

在一个实施例中，统计得到的所述统计量包括第一零阶统计量和第一一阶统计量；所述对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量包括：对应于说话人背景模型中的每个高斯分布分量，分别统计各所述声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量；及对应于说话人背景模型中的每个高斯分布分量，分别将各所述声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

在一个实施例中，计算机可读指令还使得处理器执行以下步骤：获取超过预设语音时长的语音样本；按照所述语音样本中说话人类别统计对应于说话人背景模型中的每个高斯分布分量的第二零阶统计量和第二一阶统计量；计算所述第二一阶统计量和相应的第二零阶统计量的第一商；及根据计算出的第一商构建统计量空间。

在一个实施例中，所述根据计算出的第一商构建统计量空间包括：将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；及将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

在一个实施例中，所述参考统计量包括对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；所述根据统计得到的所述统计量和所述参考统计量确定修正统计量包括：将所述第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的所述第

二商加权求和，得到对应所述说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

在一个实施例中，所述加权求和中，所述第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，所述第二商的权重为所述可调参数除以所述相应高斯分布分量的第一零阶统计量与所述可调参数的和。

在一个实施例中，所述将所述统计量映射到统计量空间获得参考统计量包括：获取所述统计量空间的正交基向量；求取所述正交基向量的映射系数，所述正交基向量与所述映射系数的乘积加上相应高斯分布分量的均值后，与相应高斯分布分量的第三商之间的二范数距离最小化；及将所述正交基向量乘以所述映射系数后加上相应高斯分布分量的均值，得到对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商。

在一个实施例中，所述待处理语音数据包括待验证语音数据和目标说话人类别的语音数据；所述根据所述修正统计量生成身份向量包括：根据与所述待验证语音数据对应的修正统计量生成待验证身份向量；及根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身份向量；所述计算机可读指令还使得所述处理器执行以下步骤：计算所述待验证身份向量和所述目标说话人身份向量的相似度；及根据所述相似度进行说话人身份验证。

上述计算机可读存储介质，统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成，在对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量后，将该统计量映射到该统计量空间中，得到的参考统计量是先验统计量。利用先验统计量来对统计得到的统计量进行修正得到修正统计量，该修正统计量能够补偿因待处理语音数据的语音时长过短和语音稀疏的情况下导致的统计量偏估，提高身份向量的身份识别性能。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，

是可以通过计算机程序来指令相关的硬件来完成，该程序可存储于一非易失性计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，该存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory，ROM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种身份向量生成方法，包括：

获取待处理语音数据；

从所述待处理语音数据提取相应的声学特征；

对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；

将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；

根据统计得到的所述统计量和所述参考统计量确定修正统计量；及

根据所述修正统计量生成身份向量。

2、根据权利要求 1 所述的方法，其特征在于，统计得到的所述统计量包括第一零阶统计量和第一一阶统计量；所述对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量包括：

对应于说话人背景模型中的每个高斯分布分量，分别统计各所述声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量；及

对应于说话人背景模型中的每个高斯分布分量，分别将各所述声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

3、根据权利要求 2 所述的方法，其特征在于，还包括：

获取超过预设语音时长的语音样本；

按照所述语音样本中说话人类别统计对应于说话人背景模型中的每个高斯分布分量的第二零阶统计量和第二一阶统计量；

计算所述第二一阶统计量和相应的第二零阶统计量的第一商；及

根据计算出的第一商构建统计量空间。

4、根据权利要求 3 所述的方法，其特征在于，所述根据计算出的第一商构建统计量空间包括：

将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；及

将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

5、根据权利要求 2 所述的方法，其特征在于，所述参考统计量包括对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；所述根据统计得到的所述统计量和所述参考统计量确定修正统计量包括：

将所述第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的所述第二商加权求和，得到对应所述说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

6、根据权利要求 5 所述的方法，其特征在于，所述加权求和中，所述第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，所述第二商的权重为所述可调参数除以所述相应高斯分布分量的第一零阶统计量与所述可调参数的和。

7、根据权利要求 5 所述的方法，其特征在于，所述将所述统计量映射到统计量空间获得参考统计量包括：

获取所述统计量空间的正交基向量；

求取所述正交基向量的映射系数，所述正交基向量与所述映射系数的乘积加上相应高斯分布分量的均值后，与相应高斯分布分量的第三商之间的二范数距离最小化；及

将所述正交基向量乘以所述映射系数后加上相应高斯分布分量的均值，得到对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商。

8、根据权利要求 1 所述的方法，其特征在于，所述待处理语音数据包括待验证语音数据和目标说话人类别的语音数据；所述根据所述修正统计量生成身份向量包括：

根据与所述待验证语音数据对应的修正统计量生成待验证身份向量；及

根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身

份向量；

所述方法还包括：

计算所述待验证身份向量和所述目标说话人身份向量的相似度；及

根据所述相似度进行说话人身份验证。

9、一种计算机设备，包括存储器和处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述处理器执行以下步骤：

获取待处理语音数据；

从所述待处理语音数据提取相应的声学特征；

对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；

将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；

根据统计得到的所述统计量和所述参考统计量确定修正统计量；及

根据所述修正统计量生成身份向量。

10、根据权利要求 9 所述的计算机设备，其特征在于，统计得到的所述统计量包括第一零阶统计量和第一一阶统计量；所述对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量包括：

对应于说话人背景模型中的每个高斯分布分量，分别统计各所述声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量；及

对应于说话人背景模型中的每个高斯分布分量，分别将各所述声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

11、根据权利要求 10 所述的计算机设备，其特征在于，所述计算机可读指令还使得所述处理器执行以下步骤：

获取超过预设语音时长的语音样本；

按照所述语音样本中说话人类别统计对应于说话人背景模型中的每个高

斯分布分量的第二零阶统计量和第二一阶统计量；

计算所述第二一阶统计量和相应的第二零阶统计量的第一商；及

根据计算出的第一商构建统计量空间。

12、根据权利要求 11 所述的计算机设备，其特征在于，所述根据计算出的第一商构建统计量空间包括：

将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；及

将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

13、根据权利要求 10 所述的计算机设备，其特征在于，所述参考统计量包括对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；所述根据统计得到的所述统计量和所述参考统计量确定修正统计量包括：

将所述第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的所述第二商加权求和，得到对应所述说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

14、根据权利要求 13 所述的计算机设备，其特征在于，所述加权求和中，所述第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，所述第二商的权重为所述可调参数除以所述相应高斯分布分量的第一零阶统计量与所述可调参数的和。

15、根据权利要求 13 所述的计算机设备，其特征在于，所述将所述统计量映射到统计量空间获得参考统计量包括：

获取所述统计量空间的正交基向量；

求取所述正交基向量的映射系数，所述正交基向量与所述映射系数的乘积加上相应高斯分布分量的均值后，与相应高斯分布分量的第三商之间的二范数距离最小化；及

将所述正交基向量乘以所述映射系数后加上相应高斯分布分量的均值，得到对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应

参考零阶统计量的第二商。

16、根据权利要求 9 所述的计算机设备，其特征在于，所述待处理语音数据包括待验证语音数据和目标说话人类别的语音数据；所述根据所述修正统计量生成身份向量包括：

根据与所述待验证语音数据对应的修正统计量生成待验证身份向量；及

根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身份向量；

所述计算机可读指令还使得所述处理器执行以下步骤：

计算所述待验证身份向量和所述目标说话人身份向量的相似度；及

根据所述相似度进行说话人身份验证。

17、一个或多个存储有计算机可读指令的非易失性的计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取待处理语音数据；

从所述待处理语音数据提取相应的声学特征；

对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量；

将所述统计量映射到统计量空间获得参考统计量；所述统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成；

根据统计得到的所述统计量和所述参考统计量确定修正统计量；及

根据所述修正统计量生成身份向量。

18、根据权利要求 17 所述的计算机可读存储介质，其特征在于，统计得到的所述统计量包括第一零阶统计量和第一一阶统计量；所述对各所述声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量包括：

对应于说话人背景模型中的每个高斯分布分量，分别统计各所述声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量；及

对应于说话人背景模型中的每个高斯分布分量，分别将各所述声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量。

19、根据权利要求 18 所述的计算机可读存储介质，其特征在于，所述计算机可读指令还使得所述处理器执行以下步骤：

获取超过预设语音时长的语音样本；

按照所述语音样本中说话人类别统计对应于说话人背景模型中的每个高斯分布分量的第二零阶统计量和第二一阶统计量；

计算所述第二一阶统计量和相应的第二零阶统计量的第一商；及

根据计算出的第一商构建统计量空间。

20、根据权利要求 19 所述的计算机可读存储介质，其特征在于，所述根据计算出的第一商构建统计量空间包括：

将计算出的第一商减去相应高斯分布分量的均值得到相应的差值；及

将得到的差值按照说话人类别和对应的高斯分布分量依次排布形成表征统计量空间的矩阵。

21、根据权利要求 18 所述的计算机可读存储介质，其特征在于，所述参考统计量包括对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；所述根据统计得到的所述统计量和所述参考统计量确定修正统计量包括：

将所述第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的所述第二商加权求和，得到对应所述说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量。

22、根据权利要求 21 所述的计算机可读存储介质，其特征在于，所述加权求和中，所述第三商的权重为相应高斯分布分量的第一零阶统计量除以相应的第一零阶统计量与可调参数的和，所述第二商的权重为所述可调参数除以所述相应高斯分布分量的第一零阶统计量与所述可调参数的和。

23、根据权利要求 21 所述的计算机可读存储介质，其特征在于，所述将

所述统计量映射到统计量空间获得参考统计量包括:

获取所述统计量空间的正交基向量;

求取所述正交基向量的映射系数,所述正交基向量与所述映射系数的乘积加上相应高斯分布分量的均值后,与相应高斯分布分量的第三商之间的二范数距离最小化;及

将所述正交基向量乘以所述映射系数后加上相应高斯分布分量的均值,得到对应所述说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商。

24、根据权利要求 17 所述的计算机可读存储介质,其特征在于,所述待处理语音数据包括待验证语音数据和目标说话人类别的语音数据;所述根据所述修正统计量生成身份向量包括:

根据与所述待验证语音数据对应的修正统计量生成待验证身份向量;及

根据与目标说话人类别的语音数据对应的修正统计量生成目标说话人身份向量;

所述计算机可读指令还使得所述处理器执行以下步骤:

计算所述待验证身份向量和所述目标说话人身份向量的相似度;及

根据所述相似度进行说话人身份验证。

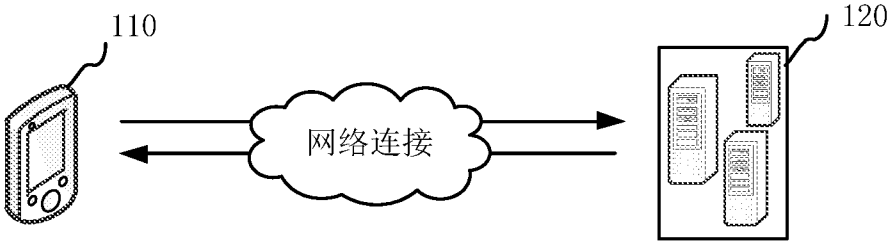

图 1

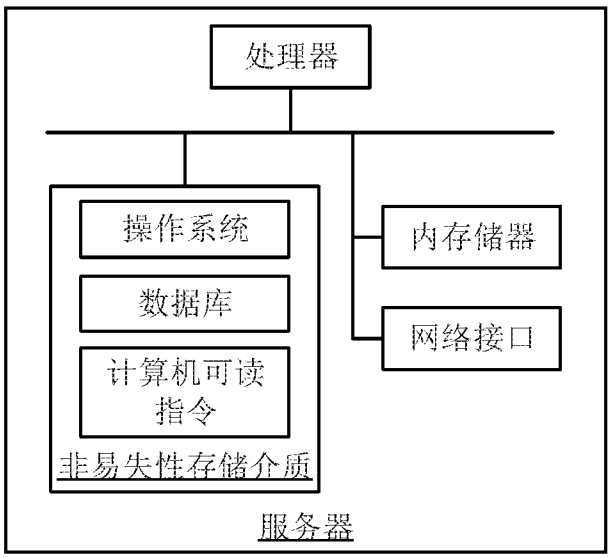

图 2A

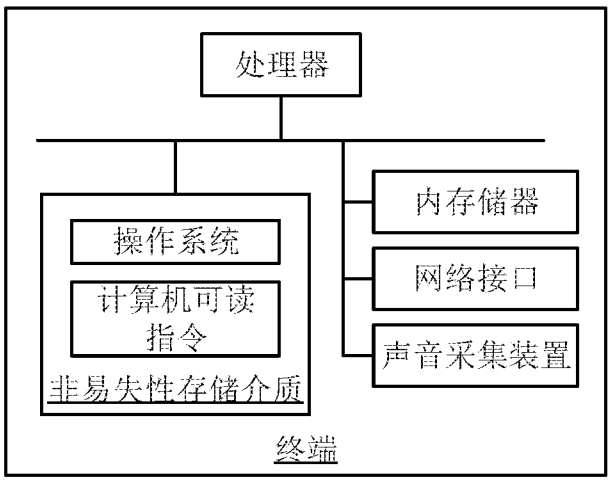
处理器
操作系统
计算机可读指令
非易失性存储介质
内存储器
网络接口
声音采集装置
终端

图 2B

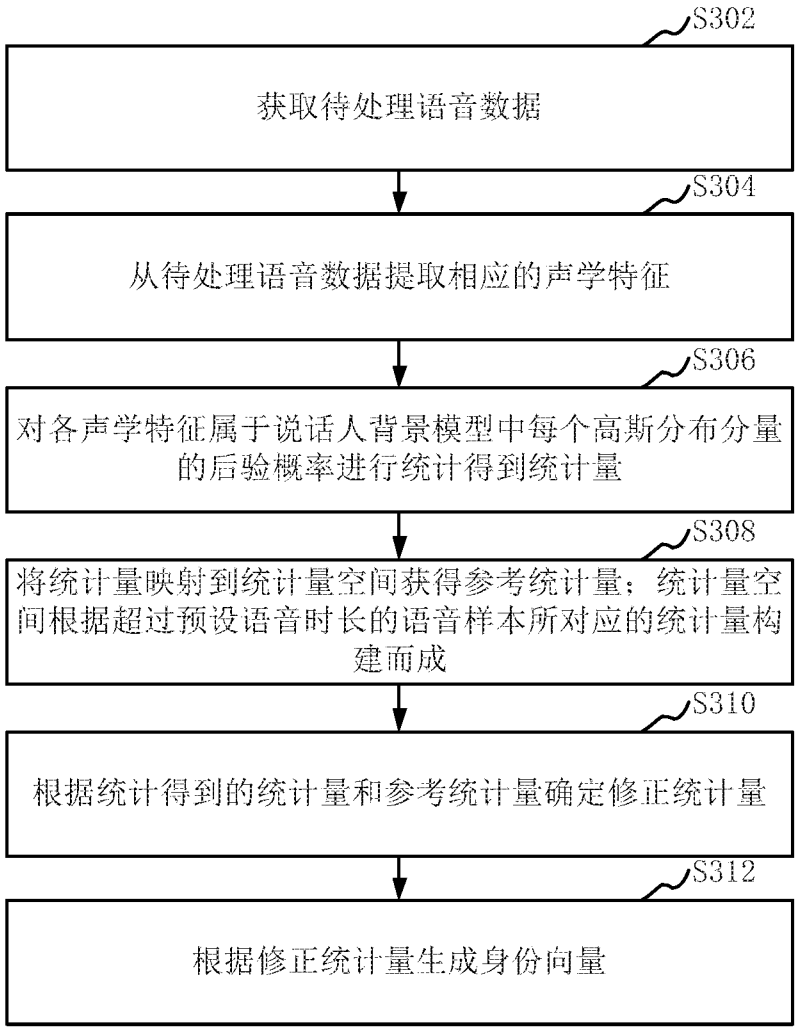
S302
获取待处理语音数据
S304
从待处理语音数据提取相应的声学特征
S306
对各声学特征属于说话人背景模型中每个高斯分布分量的后验概率进行统计得到统计量
S308
将统计量映射到统计量空间获得参考统计量；统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成
S310
根据统计得到的统计量和参考统计量确定修正统计量
S312
根据修正统计量生成身份向量

图 3

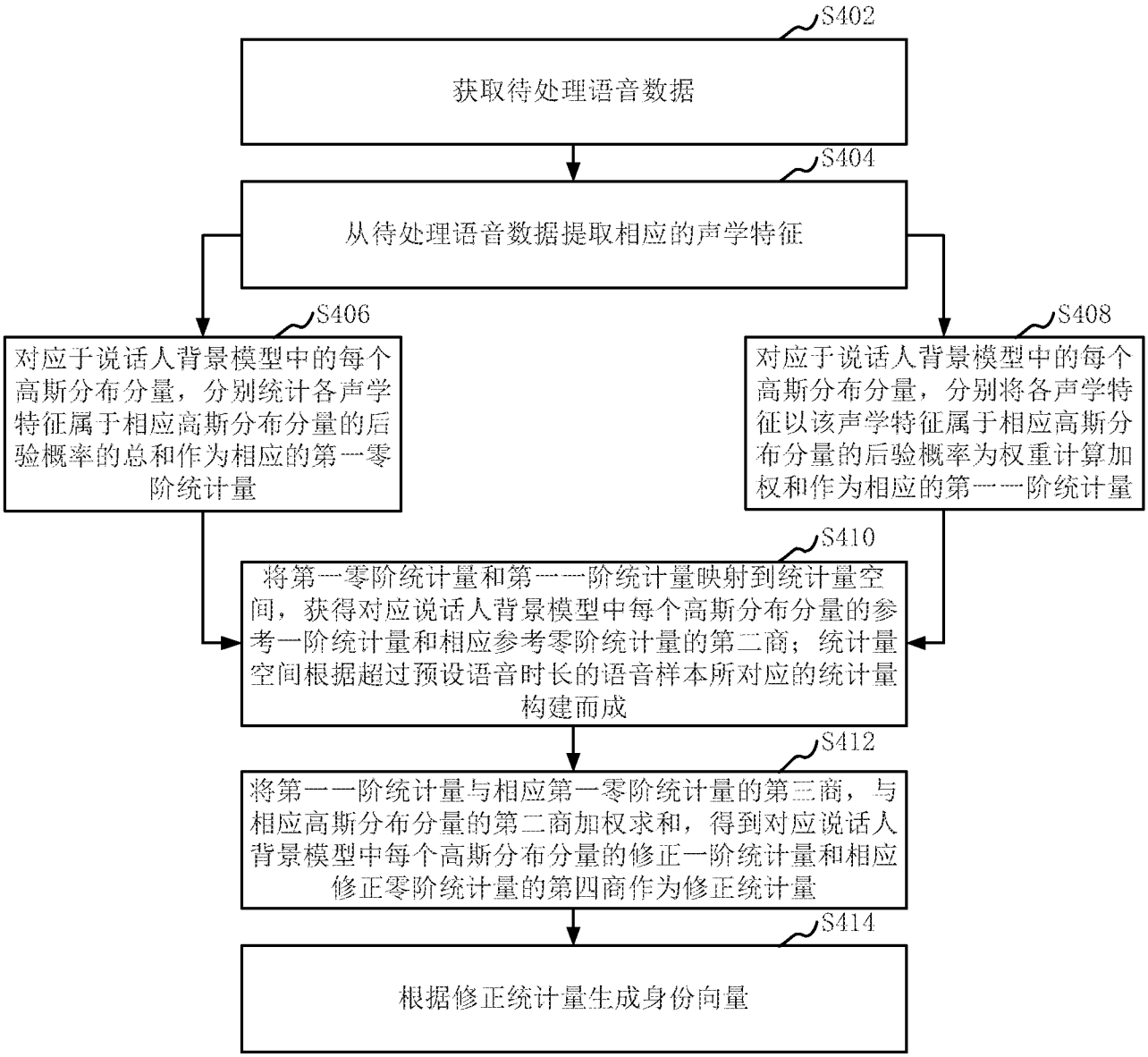
S402
获取待处理语音数据
S404
从待处理语音数据提取相应的声学特征
S406
对应于说话人背景模型中的每个高斯分布分量，分别统计各声学特征属于相应高斯分布分量的后验概率的总和作为相应的第一零阶统计量
S408
对应于说话人背景模型中的每个高斯分布分量，分别将各声学特征以该声学特征属于相应高斯分布分量的后验概率为权重计算加权和作为相应的第一一阶统计量
S410
将第一零阶统计量和第一一阶统计量映射到统计量空间，获得对应说话人背景模型中每个高斯分布分量的参考一阶统计量和相应参考零阶统计量的第二商；统计量空间根据超过预设语音时长的语音样本所对应的统计量构建而成
S412
将第一一阶统计量与相应第一零阶统计量的第三商，与相应高斯分布分量的第二商加权求和，得到对应说话人背景模型中每个高斯分布分量的修正一阶统计量和相应修正零阶统计量的第四商作为修正统计量
S414
根据修正统计量生成身份向量

图 4

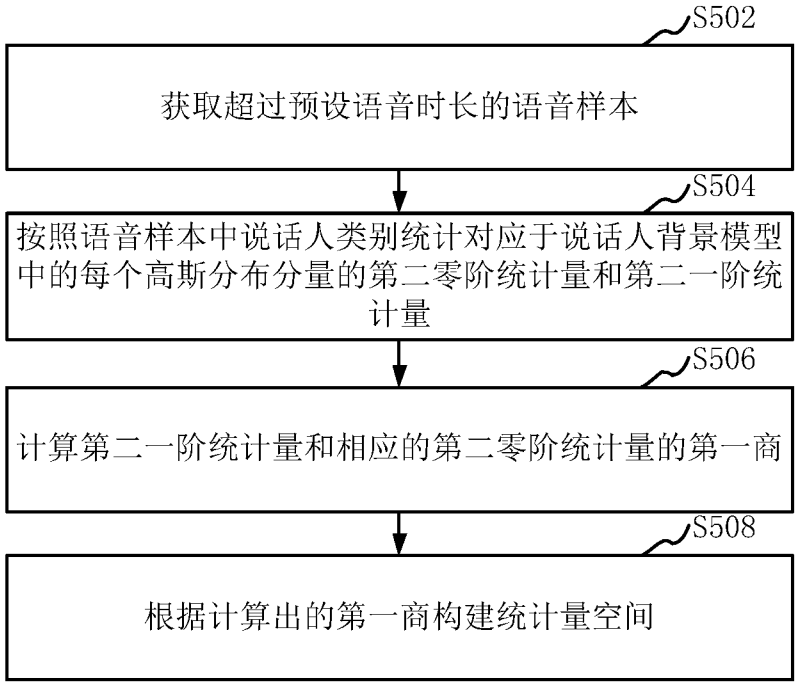
S502
获取超过预设语音时长的语音样本
S504
按照语音样本中说话人类别统计对应于说话人背景模型中的每个高斯分布分量的第二零阶统计量和第二一阶统计量
S506
计算第二一阶统计量和相应的第二零阶统计量的第一商
S508
根据计算出的第一商构建统计量空间

图 5

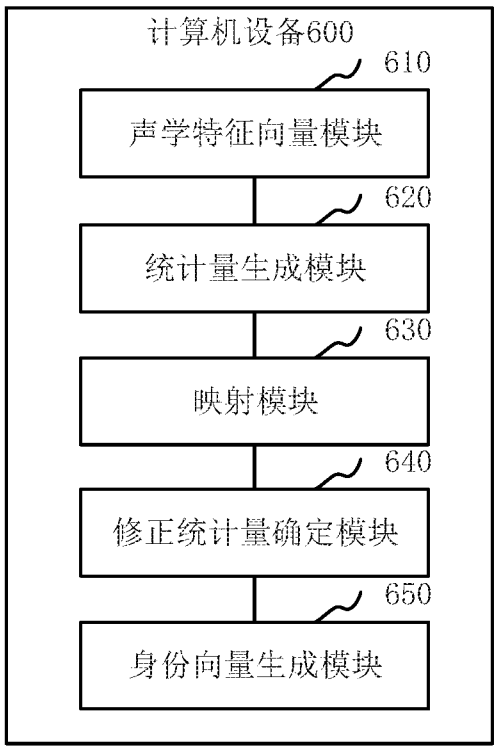
计算机设备600
610
声学特征向量模块
620
统计量生成模块
630
映射模块
640
修正统计量确定模块
650
身份向量生成模块

图 6

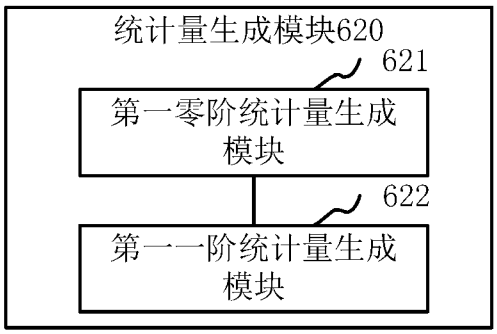
统计量生成模块620
621
第一零阶统计量生成模块
622
第一一阶统计量生成模块

图 7

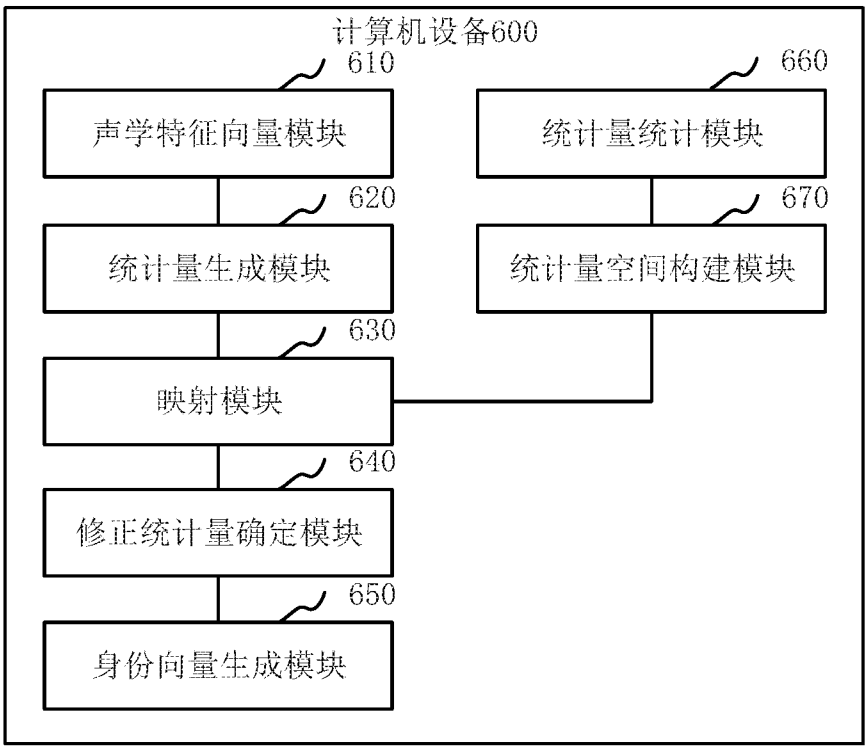
计算机设备600
610
声学特征向量模块
620
统计量生成模块
630
映射模块
640
修正统计量确定模块
650
身份向量生成模块
660
统计量统计模块
670
统计量空间构建模块

图 8

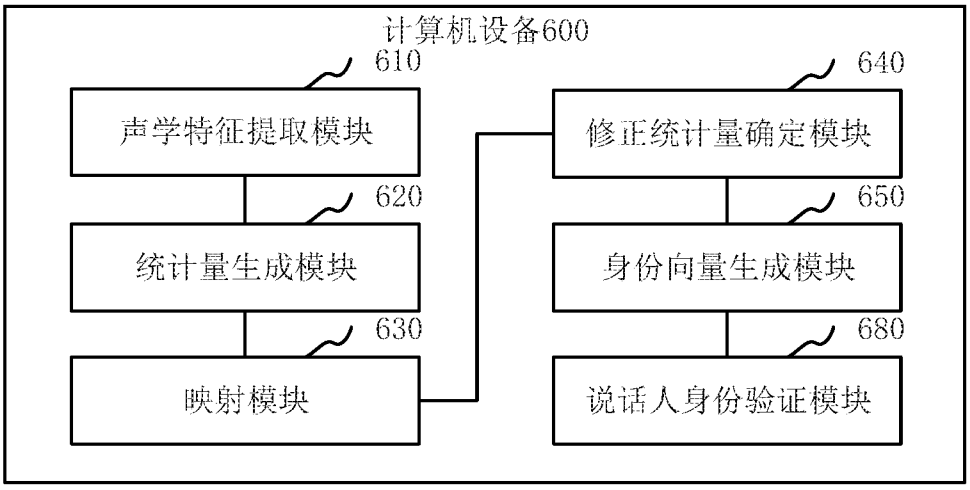
计算机设备600
610
声学特征提取模块
620
统计量生成模块
630
映射模块
640
修正统计量确定模块
650
身份向量生成模块
680
说话人身份验证模块

图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/092892

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; EPODOC; WPI; GOOGLE; CNKI: voiceprint, sound, user, gaussian mixture, time duration, exceed, greater than, pre-set, predetermine, voice, speech, i-vector, recognition, identification, speaker, feature, GMM, model, background, space, map, duration, time, longer, threshold, probability

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106169295 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 30 November 2016 (30.11.2016), claims 1-16, and description, paragraphs [0033], [0034] and [0147]	1-24
A	CN 102024455 A (SONY CORPORATION), 20 April 2011 (20.04.2011), description, paragraphs [0009]-[0031] and [0093]	1-24
A	US 2008010065 A1 (BRATT, H. et al.), 10 January 2008 (10.01.2008), the whole document	1-24
A	CN 102820033 A (NANJING UNIVERSITY), 12 December 2012 (12.12.2012), the whole document	1-24
A	YUN, Lei et al., "A Noise Robust I-vector Extractor Using Vector Taylor Series for Speaker Recognition", ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 2013 IEEE INTERNATIONAL CONFERENCE ON, 21 October 2013 (21.10.2013), ISSN: 1520-6149, the whole document	1-24

☒ 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	Date of mailing of the international search report
11 September 2017 (11.09.2017)	**27 September 2017 (27.09.2017)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **OU, Xiaodan** Telephone No.: (86-10) **82246933**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/092892

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MD, J.A. et al., "Multi-taper MFCC Features for Speaker Verification Using I-vectors", AUTOMATIC SPEECH RECOGNITION AND UNDERSTANDING (ASRU), 2011 IEEE WORKSHOP ON, 05 March 2012 (05.03.2012), the whole document	1-24

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/092892

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106169295 A	30 November 2016	None	
CN 102024455 A	20 April 2011	CN 102024455 B	17 September 2014
US 2008010065 A1	10 January 2008	None	
CN 102820033 A	12 December 2012	CN 102820033 B	04 December 2013

国际检索报告

国际申请号
PCT/CN2017/092892

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;EPODOC;WPI;GOOGLE; CNKI：语音，声纹，声音，识别，说话人，说话者，用户，特征，高斯混合，模型，背景，空间，映射，时间长度，时长，超过，超出，大于，阈值，预设，预定，概率，voice，speech，i-vector，recognition，identification，speaker，feature，GMM，model，background，space，map，duration，time，longer，threshold，probability

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
PX	CN 106169295 A (腾讯科技深圳有限公司) 2016年 11月 30日 (2016-11-30) 权利要求1-16、说明书第[0033]，[0034]，[0147]段	1-24
A	CN 102024455 A (索尼株式会社) 2011年 4月 20日 (2011-04-20) 说明书第[0009]-[0031]，[0093]段	1-24
A	US 2008010065 A1 (BRATT, HARRY等) 2008年 1月 10日 (2008-01-10) 全文	1-24
A	CN 102820033 A (南京大学) 2012年 12月 12日 (2012-12-12) 全文	1-24
A	YUN, Lei et al. "A Noise Robust I-vector Extractor Using Vector Taylor Series for Speaker Recognition" Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013年 10月 21日 (2013-10-21), ISSN：1520-6149, 全文	1-24

☑ 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	国际检索报告邮寄日期
2017年 9月 11日	2017年 9月 27日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(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 欧晓丹 电话号码 (86-10)82246933

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
A	MD, Jahangir Alam et al. "Multi-taper MFCC Features for Speaker Verification Using I-vectors" Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on, 2012年 3月 5日 (2012-03-05), 全文	1-24

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/092892

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	106169295	A	2016年 11月 30日	无			
CN	102024455	A	2011年 4月 20日	CN	102024455	B	2014年 9月 17日
US	2008010065	A1	2008年 1月 10日	无			
CN	102820033	A	2012年 12月 12日	CN	102820033	B	2013年 12月 4日