

(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
2021 年 11 月 4 日 (04.11.2021)



(10) 国际公布号
WO 2021/218136 A1

- (51) 国际专利分类号:
H04M 3/51 (2006.01) *G10L 15/22* (2006.01)
G10L 15/02 (2006.01) *G10L 15/24* (2013.01)
G10L 15/06 (2013.01)
- (21) 国际申请号: PCT/CN2020/131612
- (22) 国际申请日: 2020 年 11 月 26 日 (26.11.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202010345904.3 2020 年 4 月 27 日 (27.04.2020) 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: 赵婧 (ZHAO, 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)。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)。

(74) 代理人: 深圳市精英专利事务所 (SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路 6009 号绿景广场 B 栋 20 层 B, Guangdong 518000 (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,

(54) Title: VOICE-BASED USER GENDER AND AGE RECOGNITION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 基于语音的用户性别年龄识别方法、装置、计算机设备及存储介质

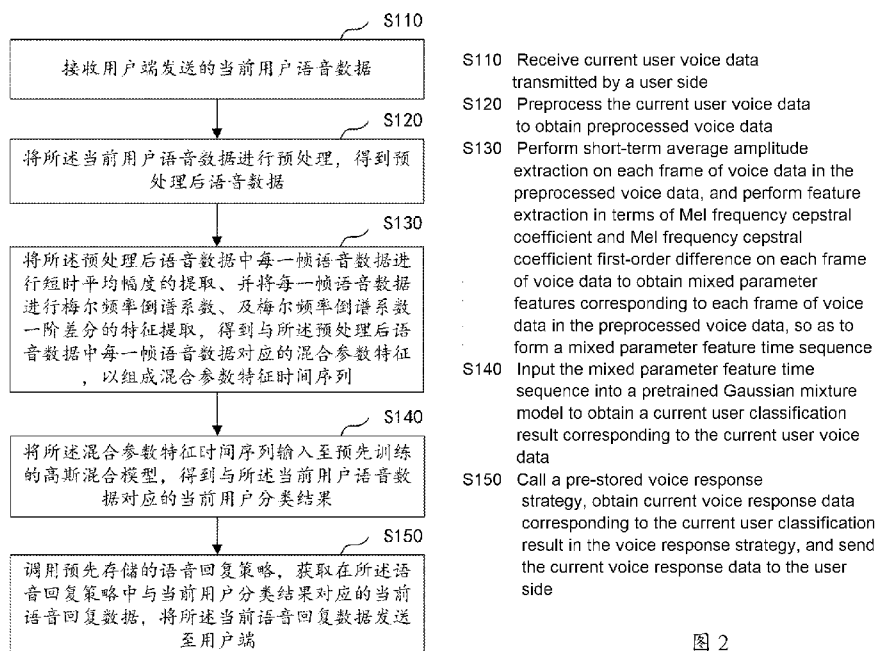


图 2

(57) Abstract: The present application relates to the technical field of voice classification in artificial intelligence, and provides a voice-based user gender and age recognition method and apparatus, a computer device, and a storage medium. The method comprises: preprocessing received current user voice data transmitted by a user side to obtain preprocessed voice data; performing feature extraction in terms of short-time average amplitude, Mel-frequency cepstral coefficient, and Mel-frequency cepstral coefficient first-order difference on each frame of voice data to obtain corresponding mixed parameter features, so as to form a mixed parameter feature time

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, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

sequence; inputting the mixed parameter feature time sequence into a Gaussian mixture model to obtain a corresponding current user classification result; and calling a voice response strategy, obtaining corresponding current voice response data, and sending the current voice response data to the user side. Accurate gender and age recognition based on user voice is implemented.

(57) 摘要: 本申请公开了基于语音的用户性别年龄识别方法、装置、计算机设备及存储介质, 涉及人工智能中的语音分类技术领域, 包括将所接收由用户端发送的当前用户语音数据进行预处理, 得到预处理后语音数据; 将其中每一帧语音数据进行短时平均幅度、梅尔频率倒谱系数及梅尔频率倒谱系数一阶差分的特征提取, 得到对应的混合参数特征以组成混合参数特征时间序列; 将混合参数特征时间序列输入至高斯混合模型得到对应的当前用户分类结果; 调用语音回复策略, 获取对应的当前语音回复数据, 将当前语音回复数据发送至用户端。实现了基于用户语音对性别和年龄的精准识别。

说明书

基于语音的用户性别年龄识别方法、装置、计算机设备及存储介质

本申请要求于2020年4月27日提交中国专利局、申请号为202010345904.3，发明名称为“基于语音的用户性别年龄识别方法、装置及计算机设备”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

本申请涉及人工智能中的语音分类技术领域，尤其涉及一种基于语音的用户性别年龄识别方法、装置、计算机设备及存储介质。

背景技术

目前，发明人意识到，智能电话外呼系统在自动根据待外呼用户清单中的用户信息对各用户进行电话外呼时，均是根据用户信息中的年龄和性别来确定外呼坐席声音的类型和外呼流程。

例如根据用户信息获知该用户为中年男性时，则智能电话外呼系统则调用女性坐席录音以实现外呼。但是若发生接电话的用户不是本人时，导致性别播报准确率较低。

发明内容

本申请实施例提供了一种基于语音的用户性别年龄识别方法、装置、计算机设备及存储介质，旨在解决现有技术智能电话外呼系统在自动根据待外呼用户清单中的用户信息对各用户进行电话外呼时，若接电话的用户不是本人，易导致性别播报准确率较低的问题。

第一方面，本申请实施例提供了一种基于语音的用户性别年龄识别方法，其包括：

接收用户端发送的当前用户语音数据；

将所述当前用户语音数据进行预处理，得到预处理后语音数据；

将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征，以组成混合参数特征时间序列；

将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果；其中，所述当前用户分类结果包括性别参数和预估年龄参数；以及

调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

第二方面，本申请实施例提供了一种基于语音的用户性别年龄识别装置，其包括：

语音数据接收单元，用于接收用户端发送的当前用户语音数据；

语音预处理单元，用于将所述当前用户语音数据进行预处理，得到预处理后语音数据；

混合参数序列获取单元，用于将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征，以组成混合参数特征时间序列；

用户分类单元，用于将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果；其中，所述当前用户分类结果包括性别参数和预估年龄参数；以及

回复数据发送单元，用于调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

第三方面，本申请实施例又提供了一种计算机设备，其包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现

以下步骤:

接收用户端发送的当前用户语音数据;

将所述当前用户语音数据进行预处理, 得到预处理后语音数据;

将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取, 得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列;

将所述混合参数特征时间序列输入至预先训练的高斯混合模型, 得到与所述当前用户语音数据对应的当前用户分类结果; 其中, 所述当前用户分类结果包括性别参数和预估年龄参数; 以及

调用预先存储的语音回复策略, 获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据, 将所述当前语音回复数据发送至用户端。

第四方面, 本申请实施例还提供了一种计算机可读存储介质, 其中所述计算机可读存储介质存储有计算机程序, 所述计算机程序当被处理器执行时使所述处理器执行以下操作:

接收用户端发送的当前用户语音数据;

将所述当前用户语音数据进行预处理, 得到预处理后语音数据;

将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取, 得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列;

将所述混合参数特征时间序列输入至预先训练的高斯混合模型, 得到与所述当前用户语音数据对应的当前用户分类结果; 其中, 所述当前用户分类结果包括性别参数和预估年龄参数; 以及

调用预先存储的语音回复策略, 获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据, 将所述当前语音回复数据发送至用户端。

本申请实施例提供了一种基于语音的用户性别年龄识别方法、装置、计算机设备及存储介质, 包括接收用户端发送的当前用户语音数据; 将所述当前用户语音数据进行预处理, 得到预处理后语音数据; 将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取, 得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列; 将所述混合参数特征时间序列输入至预先训练的高斯混合模型, 得到与所述当前用户语音数据对应的当前用户分类结果; 其中, 所述当前用户分类结果包括性别参数和预估年龄参数; 以及调用预先存储的语音回复策略, 获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据, 将所述当前语音回复数据发送至用户端。该方法综合考虑了短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分等特征对性别识别的影响, 实现了基于用户语音对性别和年龄的精准识别。

附图说明

为了更清楚地说明本申请实施例技术方案, 下面将对实施例描述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他的附图。

图1为本申请实施例提供的基于语音的用户性别年龄识别方法的应用场景示意图;

图2为本申请实施例提供的基于语音的用户性别年龄识别方法的流程示意图;

图3为本申请实施例提供的基于语音的用户性别年龄识别方法的子流程示意图;

图4为本申请实施例提供的基于语音的用户性别年龄识别方法的另一子流程示意图;

图5为本申请实施例提供的基于语音的用户性别年龄识别装置的示意性框图;

图6为本申请实施例提供的基于语音的用户性别年龄识别装置的子单元示意性框图;

图7为本申请实施例提供的基于语音的用户性别年龄识别装置的另一子单元示意性框图;

图 8 为本申请实施例提供的计算机设备的示意性框图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例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。

应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

还应当理解，在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

还应当进一步理解，在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

请参阅图 1 和图 2，图 1 为本申请实施例提供的基于语音的用户性别年龄识别方法的应用场景示意图；图 2 为本申请实施例提供的基于语音的用户性别年龄识别方法的流程示意图，该基于语音的用户性别年龄识别方法应用于服务器中，该方法通过安装于服务器中的应用软件进行执行。

如图 2 所示，该方法包括步骤 S110~S150。

S110、接收用户端发送的当前用户语音数据。

在本实施例中，当服务器中部署的智能语音系统需对用户语音进行性别和年龄识别时，初始需接收用户端上传的当前用户语音数据，从而进行后续的语音预处理和分类识别过程。

S120、将所述当前用户语音数据进行预处理，得到预处理后语音数据。

在本实施例中，由于实际的语音信号（例如本申请中采集的当前用户语音数据）是模拟信号，因此在对语音信号进行数字处理之前，首先要将所述当前用户语音数据（将所述当前用户语音数据记为 $s(t)$ ）以采样周期 T 采样，将其离散化为 $s(n)$ ，采用周期的选取应根据当前用户语音数据的带宽（依奈奎斯特采样定理）来确定，以避免信号的频域混叠失真。在对离散后的语音信号进行量化处理过程中会带来一定的量化噪声和失真。有了初始的所述当前用户语音数据后，对其进行语音的预处理包括：预加重和加窗分帧等步骤。

在一实施例中，如图 3 所示，步骤 S120 包括：

S121、调用预先存储的采样周期将所述当前用户语音数据进行采样，得到当前离散语音信号；

S122、调用预先存储的一阶 FIR 高通数字滤波器对所述当前离散语音信号进行预加重，得到当前预加重语音信号；

S123、调用预先存储的汉明窗对所述当前预加重语音信息进行加窗，得到加窗后语音数据；

S124、调用预先存储的帧移和帧长对所述加窗后语音数据进行分帧，得到预处理后语音数据。

在本实施例中，在对语音信号进行数字处理之前，首先要将所述当前用户语音数据（将所述当前用户语音数据记为 $s(t)$ ）以采样周期 T 采样，将其离散化为 $s(n)$ 。

然后，调用预先存储的一阶 FIR 高通数字滤波器时，一阶 FIR 高通数字滤波器即为一阶非递归型高通数字滤波器，其传递函数如下式 (1)：

$$H(z) = 1 - az^{-1} \quad (1)$$

具体实施时， a 的取值为 0.98。例如，设 n 时刻的所述当前离散语音信号的采样值为 $x(n)$ ，经过预加重处理后的当前预加重语音信号中与 $x(n)$ 对应的采样值为 $y(n)=x(n)-ax(n-1)$ 。

之后, 所调用的汉明窗的函数如下式 (2):

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

通过汉明窗对所述当前预加重语音信息进行加窗, 得到的加窗后语音数据可以表示为:
 $Q(n) = y(n) * \omega(n)$ 。

最后, 调用预先存储的帧移和帧长对所述加窗后语音数据进行分帧时, 例如所述加窗后语音数据对应的时域信号为 $x(l)$, 加窗分帧处理后的预处理后语音数据中第 n 帧语音数据为 $x_n(m)$, 且 $x_n(m)$ 满足式 (3):

$$x_n(m) = \omega(n) * x(n+m), \quad 0 \leq m \leq N-1 \quad (3)$$

其中, $n=0, 1T, 2T, \dots, N$ 是帧长, T 是帧移, $\omega(n)$ 是汉明窗的函数。

通过对所述当前用户语音数据进行预处理, 能有效用于后续的声音参数提取。

S130、将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取, 得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列。

在本实施例中, 对所述预处理后语音数据中进行重要参数提取时, 一般是提取短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分, 之后由所提取的参数组成与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列。通过这一方式, 得到了由所述预处理后语音数据中提取的重要参数, 结合这些重要参数能更加准确的进行用户类型分类 (主要是年龄和性别的分类)。

其中, 将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取时, 具体的根据 $M_n = \sum_{m=0}^{N-1} |x_n(m)|$ 计算预处理后语音数据中第 n 帧语音数据的短时平均幅度; 其中, M_n 表示预处理后语音数据中第 n 帧语音数据的短时平均幅度, 预处理后语音数据中第 n 帧语音数据为 $x_n(m)$, $0 \leq m \leq N-1$, N 是帧长。

在一实施例中, 如图 4 所示, 步骤 S130 包括:

S131、将所述预处理后语音数据依次进行傅里叶变换, 得到频域语音数据;

S132、将所述频域语音数据取绝对值, 得到取绝对值后语音数据;

S133、将所述取绝对值后语音数据通过梅尔滤波, 得到梅尔滤波后语音数据;

S134、将所述梅尔滤波后语音数据依次进行取对数运算和离散余弦变换, 得到与所述预处理后语音数据对应的梅尔频率倒谱系数;

S135、获取所述梅尔频率倒谱系数中连续相邻两项之差, 以得到梅尔频率倒谱系数一阶差分。

在本实施例中, 由于所述预处理后语音数据往往是时域上的语音信号, 要想将其映射到线性频率上, 就必须用 DFT (DFT 即离散傅里叶变换) 或者 FFT (FFT 即傅里叶变换), 以实现时域到频域的转换。对 N 点的信号, 若 $N/2$ 为整数, 可以使用 FFT, 以加快算法的处理速度。若 $N/2$ 不为整数, 就只能使用 DFT, 算法速度会随着点数的增加而下降。所以在分帧时, 点数必须为 2 的整数倍。

由于 FFT 出来结果的是复数, 有实部和虚部, 对其取绝对值, 得到复数的模, 而去掉相位。模反应的是声音的幅值, 幅值包含有用的信息。人耳对声音的相位并不敏感, 可以忽略相位。

通过梅尔滤波器组对所述取绝对值后语音数据通过梅尔滤波, 得到梅尔滤波后语音数据。梅尔滤波器组的具体参数如下:

设置梅尔滤波器组的采样率 $f_s=8000\text{Hz}$, 滤波器频率范围的最低频率 $f_l=0$, 滤波器频率范围的最高频率 $f_h=f_s/2=8000/2=4000$; 设置滤波器个数 $M=24$, FFT 的长度 $N=256$ 。对所述取绝对值后语音数据通过梅尔滤波后, 是对线性的频率进行梅尔滤波, 反应了人耳的听觉特性。

将所述梅尔滤波后语音数据依次进行取对数运算和离散余弦变换时, 离散余弦变换即

DCT 变换, 时域信号变换到频域, 取对数, 再经过 DCT 变换, 得到的是倒谱系数。若在频域之后增加 Mel 滤波 (即梅尔滤波), 则最终得到 MFCC (MFCC 即梅尔频率倒谱系数)。

一阶差分就是离散函数中连续相邻两项之差。当自变量从 x 变到 $x+1$ 时, 函数 $y=y(x)$ 的改变量 $\Delta y_x=y(x+1)-y(x)$, ($x=0, 1, 2, \dots$) 称为函数 $y(x)$ 在点 x 的一阶差分, 记为 $\Delta y_x=y_{x+1}-y_x$, ($x=0, 1, 2, \dots$)。

由于所述预处理后语音数据中每一帧语音数据都可以对应获取上述三个特征参数 (即短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分), 也就是一帧语音数据对应一个 1×3 的行向量, 且所述预处理后语音数据中包括 M 帧语音数据, 每一帧语音数据对应的一个 1×3 的行向量按照时序串接后, 得到一个 $1 \times 3M$ 的行向量, 该 $1 \times 3M$ 的行向量为与所述预处理后语音数据对应的混合参数特征时间序列。

具体实施时, 除了对所述预处理后语音数据中每一帧语音数据都可以对应获取短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分, 还可以对所述预处理后语音数据中每一帧语音数据都对应获取基频、语速、声压级这三个参数, 从而组成参数维度更多的混合参数特征时间序列。

S140、将所述混合参数特征时间序列输入至预先训练的高斯混合模型, 得到与所述当前用户语音数据对应的当前用户分类结果; 其中, 所述当前用户分类结果包括性别参数和预估年龄参数。

在本实施例中, 预先训练的高斯混合模型时, 需要分别训练若干个子高斯混合模型, 例如用于识别 18-20 岁男性的第一子高斯混合模型、用于识别 21-30 岁男性的第二子高斯混合模型、用于识别 31-49 岁男性的第三子高斯混合模型、用于识别 41-50 岁男性的第四子高斯混合模型、用于识别 51-70 岁男性的第五子高斯混合模型、用于识别 18-20 岁女性的第六子高斯混合模型、用于识别 21-30 岁女性的第七子高斯混合模型、用于识别 31-49 岁女性的第八子高斯混合模型、用于识别 41-50 岁女性的第九子高斯混合模型、用于识别 51-70 岁女性的第十子高斯混合模型。

高斯混合模型 (即 Gaussian mixture model, 简记为 GMM) 是指具有如下式 (4) 的概率分布模型:

$$P(y|\theta) = \sum_{k=1}^K \alpha_k \phi(y|\theta_k) \quad (4)$$

其中, α_k 是系数且 $\alpha_k \geq 0$, $\sum_{k=1}^K \alpha_k = 1$; $\phi(y|\theta_k)$ 是高斯分布密度, $\theta_k = (\mu_k, \sigma_k^2)$,

其中, $\phi(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right)$ 成为第 k 个子模型。

在一实施例中, 步骤 S140 中所述高斯混合模型中包括多个子高斯混合模型; 其中, 多个子高斯混合模型中的其中一个记为第一子高斯混合模型, 所述第一子高斯混合模型为用于识别 18-20 岁男性的识别模型。以训练用于识别 18-20 岁男性的第一子高斯混合模型为例来说明, 步骤 S140 之前还包括:

获取第一样本数据; 其中, 第一样本数据中为多个 18-20 岁男性的语音数据对应的混合参数特征时间序列;

通过第一样本数据对待训练第一子高斯混合模型进行训练, 得到用于识别 18-20 岁男性的第一子高斯混合模型;

将训练后的第一子高斯混合模型存储至区块链网络。

区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain), 本质上是一个去中心化的数据库, 是一串使用密码学方法相关联产生的数据块, 每一个数据块中包含了一批次网络交易的信息, 用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层。

在本实施例中, 第一样本数据中获取 18-20 岁男性的语音数据对应的混合参数特征时间序列的方式, 可以参考步骤 S110 至步骤 S130 中获取当前用户语音数据对应的混合参数特征时间序列的具体过程。对待训练第一子高斯混合模型进行训练的过程就是输入多组混合参数

特征时间序列，并通过EM算法（EM算法即最大期望算法）求解待训练第一子高斯混合模型中的参数，从而得到第一子高斯混合模型。

在服务器中的所述训练后的第一子高斯混合模型可以上链存储至区块链网络（该区块链网络较佳的是私有链，以供企业的各子公司使用该私有链调用第一子高斯混合模型），除了所述高斯混合模型中包括的第一子高斯混合模型可以上链存储至区块链网络，所述高斯混合模型中的其他子高斯混合模型也可以上链存储至区块链网络。所述高斯混合模型中每一个子高斯混合模型中包括的各参数值（如 α_k ， $\phi(y|\theta_k)$ 对应的参数值）均存储至区块链网络。在此过程中，服务器视为区块链网络中的一个区块链节点设备，其具备上传数据至区块链网络的权限。当服务器需从区块链网络中获取所述第一子高斯混合模型时，对服务器是否具备区块链节点设备的权限进行验证，若服务器具备区块链节点设备的权限，则获取所述第一子高斯混合模型，并在区块链网络中进行广播以告知区块链节点设备服务器已获取了所述第一子高斯混合模型。

S150、调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

在本实施例中，服务器中存储的语音回复策略中包括多种语音风格模板数据，每一种语音风格模板数据对应一种语音回复数据，且每一语音风格模板数据所使用的发音者性别、发音者风格、话术流程均是预先设置好的。

例如，获取了当前用户分类结果为18-20岁男性，在所述语音回复策略中与18-20岁男性这一用户分类结果对应的当前语音回复数据为女性甜美风格的、活泼话术流程。也即对于识别出了男性客户的时候自动调用甜美女性坐席录音，并且在话术流程中称呼对方为先生，增加礼貌。在女性客户接听电话的时候自动调用磁性男声坐席录音，并称呼其为女士以表礼貌。对于年轻的客户调用轻松活泼的话术流程，对于年长客户调用成熟稳重的话术流程。

在一实施例中，步骤S150之后还包括：

通过预先训练的N-gram模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码。

在本实施例中，通过所述N-gram模型（即多元模型）对所述当前用户语音数据进行识别，识别得到的是一整句话，例如“我叫张三，性别男，年龄25，今天需要办理A业务。”，通过N-gram模型能对所述当前用户语音数据进行有效识别，得到识别概率最大的语句作为识别结果。

由于此时已将当前用户语音数据转化为识别结果这一文本数据，此时定位识别结果中的几个关键字符串，即可获取所述识别结果中与用户年龄字段及用户性别字段分别对应的用户年龄取值和用户性别取值。同时也可获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码，该用户身份唯一识别码最佳为用户身份证号。

在一实施例中，所述通过预先训练的N-gram模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码之后，还包括：

根据所述用户身份唯一识别码，获取与用户端对应的用户真实年龄值和用户真实性别取值，判断所述预估年龄参数的取值是否等于所述用户真实年龄值，且判断所述性别参数的取值是否等于所述用户真实性别取值；

若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，将所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

在本实施例中，当获取了所述用户身份唯一识别码（也即用户的身份证号）之后，可以通过用户身份唯一识别码来获取该用户的真实年龄和性别。而通过高斯混合模型以对当前用户语音数据的分类，得到了包括性别参数和预估年龄参数的当前用户分类结果。此时将所述预估年龄参数的取值与用户真实年龄值相比较而判断是否相等，同时将所述性别参数的取值与用户真实性别取值相比较而判断是否相等。通过上述比较，即可判断通过高斯混合模

型以对当前用户语音数据的分类是否正确。

若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，则表示当前用户分类结果中的性别参数的取值和/或预估年龄参数的取值不准确，此时根据当前用户分类结果对应获得的当前语音回复数据并不适合于当前用户，故将所有分类不准确的所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

在服务器中的第一存储区域中记录智能识别性别年龄结果不准确的数据，作为客户的历史记录，以便于后续改进高斯混合模型。

若所述预估年龄参数的取值等于所述用户真实年龄值，且所述性别参数的取值等于所述用户真实性别取值，则表示当前用户分类结果中的性别参数的取值和预估年龄参数的取值均是准确的，此时根据当前用户分类结果对应获得的当前语音回复数据是适合于当前用户，此时无需对针对当前用户分类结果对应的当前用户语音数据进行调整。

该方法综合考虑了短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分等特征对性别识别的影响，实现了基于用户语音对性别和年龄的精准识别。

本申请实施例还提供一种基于语音的用户性别年龄识别装置，该基于语音的用户性别年龄识别装置用于执行前述基于语音的用户性别年龄识别方法的任一实施例。具体地，请参阅图 5，图 5 是本申请实施例提供的基于语音的用户性别年龄识别装置的示意性框图。该基于语音的用户性别年龄识别装置 100 可以配置于服务器中。

如图 5 所示，基于语音的用户性别年龄识别装置 100 包括：语音数据接收单元 110、语音预处理单元 120、混合参数序列获取单元 130、用户分类单元 140、回复数据发送单元 150。

语音数据接收单元 110，用于接收用户端发送的当前用户语音数据。

在本实施例中，当服务器中部署的智能语音系统需对用户语音进行性别和年龄识别时，初始需接收用户端上传的当前用户语音数据，从而进行后续的语音预处理和分类识别过程。

语音预处理单元 120，用于将所述当前用户语音数据进行预处理，得到预处理后语音数据。

在本实施例中，由于实际的语音信号（例如本申请中采集的当前用户语音数据）是模拟信号，因此在对语音信号进行数字处理之前，首先要将所述当前用户语音数据（将所述当前用户语音数据记为 $s(t)$ ）以采样周期 T 采样，将其离散化为 $s(n)$ ，采用周期的选取应根据当前用户语音数据的带宽（依奈奎斯特采样定理）来确定，以避免信号的频域混叠失真。在对离散后的语音信号进行量化处理过程中会带来一定的量化噪声和失真。有了初始的所述当前用户语音数据后，对其进行语音的预处理包括：预加重和加窗分帧等步骤。

在一实施例中，如图 6 所示，语音预处理单元 120 包括：

语音数据采样单元 121，用于调用预先存储的采样周期将所述当前用户语音数据进行采样，得到当前离散语音信号；

预加重单元 122，用于调用预先存储的一阶 FIR 高通数字滤波器对所述当前离散语音信号进行预加重，得到当前预加重语音信号；

加窗单元 123，用于调用预先存储的汉明窗对所述当前预加重语音信息进行加窗，得到加窗后语音数据；

分帧单元 124，用于调用预先存储的帧移和帧长对所述加窗后语音数据进行分帧，得到预处理后语音数据。

在本实施例中，在对语音信号进行数字处理之前，首先要将所述当前用户语音数据（将所述当前用户语音数据记为 $s(t)$ ）以采样周期 T 采样，将其离散化为 $s(n)$ 。

然后，调用预先存储的一阶 FIR 高通数字滤波器时，一阶 FIR 高通数字滤波器即为一阶非递归型高通数字滤波器，其传递函数如上式（1）。

例如，设 n 时刻的所述当前离散语音信号的采样值为 $x(n)$ ，经过预加重处理后的当前预加重语音信号中与 $x(n)$ 对应的采样值为 $y(n)=x(n)-ax(n-1)$ 。

之后, 所调用的汉明窗的函数如上式 (2), 通过汉明窗对所述当前预加重语音信息进行加窗, 得到的加窗后语音数据可以表示为: $Q(n)=y(n)*\omega(n)$ 。

最后, 调用预先存储的帧移和帧长对所述加窗后语音数据进行分帧时, 例如所述加窗后语音数据对应的时域信号为 $x(l)$, 加窗分帧处理后的预处理后语音数据中第 n 帧语音数据为 $x_n(m)$, 且 $x_n(m)$ 满足上式 (3)。通过对所述当前用户语音数据进行预处理, 能有效用于后续的声音参数提取。

混合参数序列获取单元 130, 用于将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取, 得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列。

在本实施例中, 对所述预处理后语音数据中进行重要参数提取时, 一般是提取短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分, 之后由所提取的参数组成与所述预处理后语音数据中每一帧语音数据对应的混合参数特征, 以组成混合参数特征时间序列。通过这一方式, 得到了由所述预处理后语音数据中提取的重要参数, 结合这些重要参数能更加准确的进行用户类型分类 (主要是年龄和性别的分类)。

其中, 将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取时, 具体的根据 $M_n = \sum_{m=0}^{N-1} |x_n(m)|$ 计算预处理后语音数据中第 n 帧语音数据的短时平均幅度; 其中, M_n 表示预处理后语音数据中第 n 帧语音数据的短时平均幅度, 预处理后语音数据中第 n 帧语音数据为 $x_n(m)$, $0 \leq m \leq N-1$, N 是帧长。

在一实施例中, 如图 7 所示, 混合参数序列获取单元 130 包括:

傅里叶变换单元 131, 用于将所述预处理后语音数据依次进行傅里叶变换, 得到频域语音数据;

取绝对值单元 132, 用于将所述频域语音数据取绝对值, 得到取绝对值后语音数据;

梅尔滤波单元 133, 用于将所述取绝对值后语音数据通过梅尔滤波, 得到梅尔滤波后语音数据;

梅尔频率倒谱系数获取单元 134, 用于将所述梅尔滤波后语音数据依次进行取对数运算和离散余弦变换, 得到与所述预处理后语音数据对应的梅尔频率倒谱系数;

一阶差分获取单元 135、获取所述梅尔频率倒谱系数中连续相邻两项之差, 以得到梅尔频率倒谱系数一阶差分。

在本实施例中, 由于所述预处理后语音数据往往是时域上的语音信号, 要想将其映射到线性频率上, 就必须用 DFT (DFT 即离散傅里叶变换) 或者 FFT (FFT 即傅里叶变换), 以实现时域到频域的转换。对 N 点的信号, 若 $N/2$ 为整数, 可以使用 FFT, 以加快算法的处理速度。若 $N/2$ 不为整数, 就只能使用 DFT, 算法速度会随着点数的增加而下降。所以在分帧时, 点数必须为 2 的整数倍。

由于 FFT 出来结果的是复数, 有实部和虚部, 对其取绝对值, 得到复数的模, 而去掉相位。模反应的是声音的幅值, 幅值包含有用的信息。人耳对声音的相位并不敏感, 可以忽略相位。

通过梅尔滤波器组对所述取绝对值后语音数据通过梅尔滤波, 得到梅尔滤波后语音数据。梅尔滤波器组的具体参数如下:

设置梅尔滤波器组的采样率 $f_s=8000\text{Hz}$, 滤波器频率范围的最低频率 $f_l=0$, 滤波器频率范围的最高频率 $f_h=f_s/2=8000/2=4000$; 设置滤波器个数 $M=24$, FFT 的长度 $N=256$ 。对所述取绝对值后语音数据通过梅尔滤波后, 是对线性的频率进行梅尔滤波, 反应了人耳的听觉特性。

将所述梅尔滤波后语音数据依次进行取对数运算和离散余弦变换时, 离散余弦变换即 DCT 变换, 时域信号变换到频域, 取对数, 再经过 DCT 变换, 得到的是倒谱系数。若在频域之后增加 Mel 滤波 (即梅尔滤波), 则最终得到 MFCC (MFCC 即梅尔频率倒谱系数)。

一阶差分就是离散函数中连续相邻两项之差。当自变量从 x 变到 $x+1$ 时, 函数 $y=y(x)$ 的改变量 $\Delta yx=y(x+1)-y(x)$, ($x=0, 1, 2, \dots$) 称为函数 $y(x)$ 在点 x 的一阶差分, 记为 $\Delta yx=yx+1-yx$, ($x=0, 1, 2, \dots$)。

由于所述预处理后语音数据中每一帧语音数据都可以对应获取上述三个特征参数 (即短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分), 也就是一帧语音数据对应一个 1×3 的行向量, 且所述预处理后语音数据中包括 M 帧语音数据, 每一帧语音数据对应的一个 1×3 的行向量按照时序串接后, 得到一个 $1 \times 3M$ 的行向量, 该 $1 \times 3M$ 的行向量为与所述预处理后语音数据对应的混合参数特征时间序列。

具体实施时, 除了对所述预处理后语音数据中每一帧语音数据都可以对应获取短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分, 还可以对所述预处理后语音数据中每一帧语音数据都对应获取基频、语速、声压级这三个参数, 从而组成参数维度更多的混合参数特征时间序列。

用户分类单元 140, 用于将所述混合参数特征时间序列输入至预先训练的高斯混合模型, 得到与所述当前用户语音数据对应的当前用户分类结果; 其中, 所述当前用户分类结果包括性别参数和预估年龄参数。

在本实施例中, 预先训练的高斯混合模型时, 需要分别训练若干个子高斯混合模型, 例如用于识别 18-20 岁男性的第一子高斯混合模型、用于识别 21-30 岁男性的第二子高斯混合模型、用于识别 31-49 岁男性的第三子高斯混合模型、用于识别 41-50 岁男性的第四子高斯混合模型、用于识别 51-70 岁男性的第五子高斯混合模型、用于识别 18-20 岁女性的第六子高斯混合模型、用于识别 21-30 岁女性的第七子高斯混合模型、用于识别 31-49 岁女性的第八子高斯混合模型、用于识别 41-50 岁女性的第九子高斯混合模型、用于识别 51-70 岁女性的第十子高斯混合模型。

高斯混合模型 (即 Gaussian mixture model, 简记为 GMM) 是指具有如上式 (4) 的概率分布模型。

在一实施例中, 用户分类单元 140 所述高斯混合模型中包括多个子高斯混合模型; 其中, 多个子高斯混合模型中的其中一个记为第一子高斯混合模型, 所述第一子高斯混合模型为用于识别 18-20 岁男性的识别模型。以训练用于识别 18-20 岁男性的第一子高斯混合模型为例来说明, 基于语音的用户性别年龄识别装置 100 还包括:

第一样本获取单元, 用于获取第一样本数据; 其中, 第一样本数据中为多个 18-20 岁男性的语音数据对应的混合参数特征时间序列;

第一子模型训练单元, 用于通过第一样本数据对待训练第一子高斯混合模型进行训练, 得到用于识别 18-20 岁男性的第一子高斯混合模型;

子模型上链单元, 用于将训练后的第一子高斯混合模型存储至区块链网络。

在本实施例中, 第一样本数据中获取 18-20 岁男性的语音数据对应的混合参数特征时间序列的方式, 可以参考获取当前用户语音数据对应的混合参数特征时间序列的具体过程。对待训练第一子高斯混合模型进行训练的过程就是输入多组混合参数特征时间序列, 并通过 EM 算法 (EM 算法即最大期望算法) 求解待训练第一子高斯混合模型中的参数, 从而得到第一子高斯混合模型。

回复数据发送单元 150, 用于调用预先存储的语音回复策略, 获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据, 将所述当前语音回复数据发送至用户端。

在本实施例中, 服务器中存储的语音回复策略中包括多种语音风格模板数据, 每一种语音风格模板数据对应一种语音回复数据, 且每一语音风格模板数据所使用的发音者性别、发音者风格、话术流程均是预先设置好的。

例如, 获取了当前用户分类结果为 18-20 岁男性, 在所述语音回复策略中与 18-20 岁男性这一用户分类结果对应的当前语音回复数据为女性甜美风格的、活泼话术流程。也即对于识别出了男性客户的时候自动调用甜美女性坐席录音, 并且在话术流程中称呼对方为先生,

增加礼貌。在女性客户接听电话的时候自动调用磁性男声坐席录音，并称呼其为女士以表礼貌。对于年轻的客户调用轻松活泼的话术流程，对于年长客户调用成熟稳重的话术流程。

在一实施例中，基于语音的用户性别年龄识别装置 100 还包括：

身份唯一识别码获取单元，用于通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码。

在本实施例中，通过所述 N-gram 模型（即多元模型）对所述当前用户语音数据进行识别，识别得到的是一整句话，例如“我叫张三，性别男，年龄 25，今天需要办理 A 业务。”，通过 N-gram 模型能对所述当前用户语音数据进行有效识别，得到识别概率最大的语句作为识别结果。

由于此时已将当前用户语音数据转化为识别结果这一文本数据，此时定位识别结果中的几个关键字串，即可获取所述识别结果中与用户年龄字段及用户性别字段分别对应的用户年龄取值和用户性别取值。同时也可获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码，该用户身份唯一识别码最佳为用户身份证号。

在一实施例中，基于语音的用户性别年龄识别装置 100，还包括：

性别年龄比对单元，用于根据所述用户身份唯一识别码，获取与用户端对应的用户真实年龄值和用户真实性别取值，判断所述预估年龄参数的取值是否等于所述用户真实年龄值，且判断所述性别参数的取值是否等于所述用户真实性别取值；

误差数据存储单元，用于若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，将所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

在本实施例中，当获取了所述用户身份唯一识别码（也即用户的身份证号）之后，可以通过用户身份唯一识别码来获取该用户的真实年龄和性别。而通过高斯混合模型以对当前用户语音数据的分类，得到了包括性别参数和预估年龄参数的当前用户分类结果。此时将所述预估年龄参数的取值与用户真实年龄值相比较而判断是否相等，同时将所述性别参数的取值与所述用户真实性别取值相比较而判断是否相等。通过上述比较，即可判断通过高斯混合模型以对当前用户语音数据的分类是否正确。

若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，则表示当前用户分类结果中的性别参数的取值和/或预估年龄参数的取值不准确，此时根据当前用户分类结果对应获得的当前语音回复数据并不适合于当前用户，故将所有分类不准确的所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

在服务器中的第一存储区域中记录智能识别性别年龄结果不准确的数据，作为客户的历史记录，以便于后续改进高斯混合模型。

若所述预估年龄参数的取值等于所述用户真实年龄值，且所述性别参数的取值等于所述用户真实性别取值，则表示当前用户分类结果中的性别参数的取值和预估年龄参数的取值均是准确的，此时根据当前用户分类结果对应获得的当前语音回复数据是适合于当前用户，此时无需对针对当前用户分类结果对应的当前用户语音数据进行调整。

该装置综合考虑了短时平均幅度、梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分等特征对性别识别的影响，实现了基于用户语音对性别和年龄的精准识别。

上述基于语音的用户性别年龄识别装置可以实现为计算机程序的形式，该计算机程序可以在如图 8 所示的计算机设备上运行。

请参阅图 8，图 8 是本申请实施例提供的计算机设备的示意性框图。该计算机设备 500 是服务器，服务器可以是独立的服务器，也可以是多个服务器组成的服务器集群。

参阅图 8，该计算机设备 500 包括通过系统总线 501 连接的处理器 502、存储器和网络接口 505，其中，存储器可以包括非易失性存储介质 503 和内存 504。

该非易失性存储介质 503 可存储操作系统 5031 和计算机程序 5032。该计算机程序 5032

被执行时,可使得处理器 502 执行基于语音的用户性别年龄识别方法。

该处理器 502 用于提供计算和控制能力,支撑整个计算机设备 500 的运行。

该内存储器 504 为非易失性存储介质 503 中的计算机程序 5032 的运行提供环境,该计算机程序 5032 被处理器 502 执行时,可使得处理器 502 执行基于语音的用户性别年龄识别方法。

该网络接口 505 用于进行网络通信,如提供数据信息的传输等。本领域技术人员可以理解,图 8 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计算机设备 500 的限定,具体的计算机设备 500 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。

其中,所述处理器 502 用于运行存储在存储器中的计算机程序 5032,以实现本申请实施例公开的基于语音的用户性别年龄识别方法。

本领域技术人员可以理解,图 8 中示出的计算机设备的实施例并不构成对计算机设备具体构成的限定,在其他实施例中,计算机设备可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。例如,在一些实施例中,计算机设备可以仅包括存储器及处理器,在这样的实施例中,存储器及处理器的结构及功能与图 8 所示实施例一致,在此不再赘述。

应当理解,在本申请实施例中,处理器 502 可以是中央处理单元 (Central Processing Unit, CPU),该处理器 502 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。其中,通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

在本申请的另一实施例中提供计算机可读存储介质。该计算机可读存储介质可以是非易失性,也可以是易失性。该计算机可读存储介质存储有计算机程序,其中计算机程序被处理器执行时实现本申请实施例公开的基于语音的用户性别年龄识别方法。

所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,上述描述的设备、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

在本申请所提供的几个实施例中,应该理解到,所揭露的设备、装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为逻辑功能划分,实际实现时可以有另外的划分方式,也可以将具有相同功能的单元集合成一个单元,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口、装置或单元的间接耦合或通信连接,也可以是电的,机械的或其它的形式连接。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本申请实施例方案的目的。

另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以是两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

所述集成的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,可以存储在一个存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分,或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件

产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权利要求书

1. 一种基于语音的用户性别年龄识别方法，其中，包括：

接收用户端发送的当前用户语音数据；

将所述当前用户语音数据进行预处理，得到预处理后语音数据；

将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征，以组成混合参数特征时间序列；

将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果；其中，所述当前用户分类结果包括性别参数和预估年龄参数；以及

调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

2. 根据权利要求1所述的基于语音的用户性别年龄识别方法，其中，所述调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端之后，还包括：

通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码。

3. 根据权利要求2所述的基于语音的用户性别年龄识别方法，其中，所述通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码之后还包括：

根据所述用户身份唯一识别码，获取与用户端对应的用户真实年龄值和用户真实性别取值，判断所述预估年龄参数的取值是否等于所述用户真实年龄值，且判断所述性别参数的取值是否等于所述用户真实性别取值；

若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，将所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

4. 根据权利要求1所述的基于语音的用户性别年龄识别方法，其中，所述将所述当前用户语音数据进行预处理，得到预处理后语音数据，包括：

调用预先存储的采样周期将所述当前用户语音数据进行采样，得到当前离散语音信号；

调用预先存储的一阶 FIR 高通数字滤波器对所述当前离散语音信号进行预加重，得到当前预加重语音信号；

调用预先存储的汉明窗对所述当前预加重语音信息进行加窗，得到加窗后语音数据；

调用预先存储的帧移和帧长对所述加窗后语音数据进行分帧，得到预处理后语音数据。

5. 根据权利要求1所述的基于语音的用户性别年龄识别方法，其中，所述将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，包括：

将所述预处理后语音数据依次进行傅里叶变换，得到频域语音数据；

将所述频域语音数据取绝对值，得到取绝对值后语音数据；

将所述取绝对值后语音数据通过梅尔滤波，得到梅尔滤波后语音数据；

将所述梅尔滤波后语音数据依次进行取对数运算和离散余弦变换，得到与所述预处理后语音数据对应的梅尔频率倒谱系数；

获取所述梅尔频率倒谱系数中连续相邻两项之差，以得到梅尔频率倒谱系数一阶差分。

6. 根据权利要求1所述的基于语音的用户性别年龄识别方法，其中，所述高斯混合模型中包括多个子高斯混合模型；其中，多个子高斯混合模型中的其中一个记为第一子高斯混合模型，所述第一子高斯混合模型为用于识别 18-20 岁男性的识别模型；

所述将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果之前，还包括：

获取第一样本数据；其中，第一样本数据中为多个 18-20 岁男性的语音数据对应的混合参数特征时间序列；

通过第一样本数据对待训练第一子高斯混合模型进行训练，得到用于识别 18-20 岁男性的第一子高斯混合模型；

将训练后的第一子高斯混合模型存储至区块链网络。

7. 根据权利要求 1 所述的基于语音的用户性别年龄识别方法，其中，所述将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取，包括：

根据 $M_n = \sum_{m=0}^{N-1} |x_n(m)|$ 计算预处理后语音数据中第 n 帧语音数据的短时平均幅度；其中， M_n 表示预处理后语音数据中第 n 帧语音数据的短时平均幅度，预处理后语音数据中第 n 帧语音数据为 $x_n(m)$ ， $0 \leq m \leq N-1$ ， N 是帧长。

8. 根据权利要求 5 所述的基于语音的用户性别年龄识别方法，其中，所述将所述取绝对值后语音数据通过梅尔滤波，得到梅尔滤波后语音数据中通过梅尔滤波器组对所述取绝对值后语音数据通过梅尔滤波，得到梅尔滤波后语音数据；其中，梅尔滤波器组的采样率 $f_s=8000\text{Hz}$ ，滤波器频率范围的最低频率 $f_l=0$ ，滤波器频率范围的最高频率 $f_h=f_s/2=8000/2=4000$ ；设置滤波器个数 $M=24$ ，FFT 的长度 $N=256$ 。

9. 一种基于语音的用户性别年龄识别装置，其中，包括：

语音数据接收单元，用于接收用户端发送的当前用户语音数据；

语音预处理单元，用于将所述当前用户语音数据进行预处理，得到预处理后语音数据；

混合参数序列获取单元，用于将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征，以组成混合参数特征时间序列；

用户分类单元，用于将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果；其中，所述当前用户分类结果包括性别参数和预估年龄参数；以及

回复数据发送单元，用于调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

10. 一种计算机设备，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，其中，所述处理器执行所述计算机程序时实现以下步骤：

接收用户端发送的当前用户语音数据；

将所述当前用户语音数据进行预处理，得到预处理后语音数据；

将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征，以组成混合参数特征时间序列；

将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果；其中，所述当前用户分类结果包括性别参数和预估年龄参数；以及

调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

11. 根据权利要求 10 所述的计算机设备，其中，所述调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端之后，还包括：

通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码。

12. 根据权利要求 11 所述的计算机设备，其中，所述通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的

用户身份唯一识别码之后还包括：

根据所述用户身份唯一识别码，获取与用户端对应的用户真实年龄值和用户真实性别取值，判断所述预估年龄参数的取值是否等于所述用户真实年龄值，且判断所述性别参数的取值是否等于所述用户真实性别取值；

若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，将所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

13. 根据权利要求10所述的基于语音的计算机设备，其中，所述将所述当前用户语音数据进行预处理，得到预处理后语音数据，包括：

调用预先存储的采样周期将所述当前用户语音数据进行采样，得到当前离散语音信号；

调用预先存储的一阶FIR高通数字滤波器对所述当前离散语音信号进行预加重，得到当前预加重语音信号；

调用预先存储的汉明窗对所述当前预加重语音信息进行加窗，得到加窗后语音数据；

调用预先存储的帧移和帧长对所述加窗后语音数据进行分帧，得到预处理后语音数据。

14. 根据权利要求10所述的基于语音的计算机设备，其中，所述将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，包括：

将所述预处理后语音数据依次进行傅里叶变换，得到频域语音数据；

将所述频域语音数据取绝对值，得到取绝对值后语音数据；

将所述取绝对值后语音数据通过梅尔滤波，得到梅尔滤波后语音数据；

将所述梅尔滤波后语音数据依次进行取对数运算和离散余弦变换，得到与所述预处理后语音数据对应的梅尔频率倒谱系数；

获取所述梅尔频率倒谱系数中连续相邻两项之差，以得到梅尔频率倒谱系数一阶差分。

15. 根据权利要求10所述的基于语音的计算机设备，其中，所述高斯混合模型中包括多个子高斯混合模型；其中，多个子高斯混合模型中的其中一个记为第一子高斯混合模型，所述第一子高斯混合模型为用于识别18-20岁男性的识别模型；

所述将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果之前，还包括：

获取第一样本数据；其中，第一样本数据中为多个18-20岁男性的语音数据对应的混合参数特征时间序列；

通过第一样本数据对待训练第一子高斯混合模型进行训练，得到用于识别18-20岁男性的第一子高斯混合模型；

将训练后的第一子高斯混合模型存储至区块链网络。

16. 根据权利要求10所述的计算机设备，其中，所述将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取，包括：

根据 $M_n = \sum_{m=0}^{N-1} |x_n(m)|$ 计算预处理后语音数据中第n帧语音数据的短时平均幅度；其中， M_n 表示预处理后语音数据中第n帧语音数据的短时平均幅度，预处理后语音数据中第n帧语音数据为 $x_n(m)$ ， $0 \leq m \leq N-1$ ，N是帧长。

17. 根据权利要求14所述的计算机设备，其中，所述将所述取绝对值后语音数据通过梅尔滤波，得到梅尔滤波后语音数据中通过梅尔滤波器组对所述取绝对值后语音数据通过梅尔滤波，得到梅尔滤波后语音数据；其中，梅尔滤波器组的采样率 $f_s=8000\text{Hz}$ ，滤波器频率范围的最低频率 $f_l=0$ ，滤波器频率范围的最高频率 $f_h=f_s/2=8000/2=4000$ ；设置滤波器个数 $M=24$ ，FFT的长度 $N=256$ 。

18. 一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行以下操作：

接收用户端发送的当前用户语音数据；

将所述当前用户语音数据进行预处理，得到预处理后语音数据；

将所述预处理后语音数据中每一帧语音数据进行短时平均幅度的提取、并将每一帧语音数据进行梅尔频率倒谱系数、及梅尔频率倒谱系数一阶差分的特征提取，得到与所述预处理后语音数据中每一帧语音数据对应的混合参数特征，以组成混合参数特征时间序列；

将所述混合参数特征时间序列输入至预先训练的高斯混合模型，得到与所述当前用户语音数据对应的当前用户分类结果；其中，所述当前用户分类结果包括性别参数和预估年龄参数；以及

调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端。

19. 根据权利要求 18 所述的计算机可读存储介质，其中，所述调用预先存储的语音回复策略，获取在所述语音回复策略中与当前用户分类结果对应的当前语音回复数据，将所述当前语音回复数据发送至用户端之后，还包括：

通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码。

20. 根据权利要求 19 所述的计算机可读存储介质，其中，所述通过预先训练的 N-gram 模型对所述当前用户语音数据进行识别得到识别结果，获取所述识别结果中与用户识别码字段对应的用户身份唯一识别码之后还包括：

根据所述用户身份唯一识别码，获取与用户端对应的用户真实年龄值和用户真实性别取值，判断所述预估年龄参数的取值是否等于所述用户真实年龄值，且判断所述性别参数的取值是否等于所述用户真实性别取值；

若所述预估年龄参数的取值不等于所述用户真实年龄值，或者所述性别参数的取值不等于所述用户真实性别取值，将所述当前用户分类结果及所述当前用户语音数据存储至预先创建的第一存储区域。

说明书附图

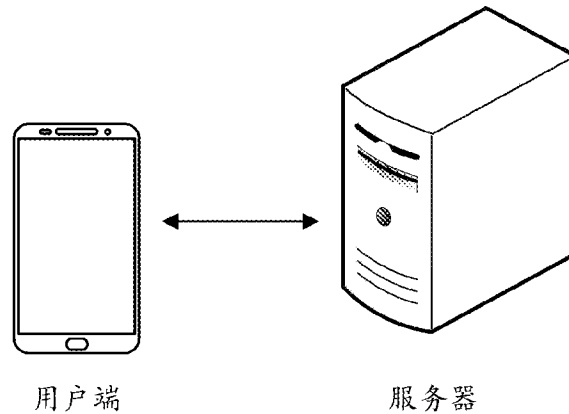


图 1

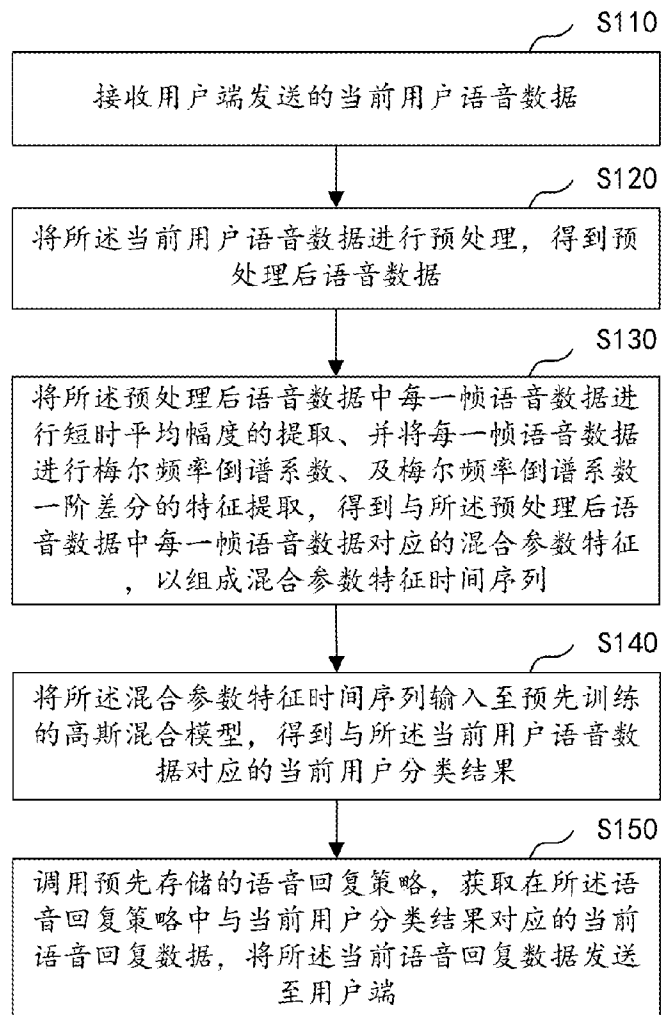


图 2

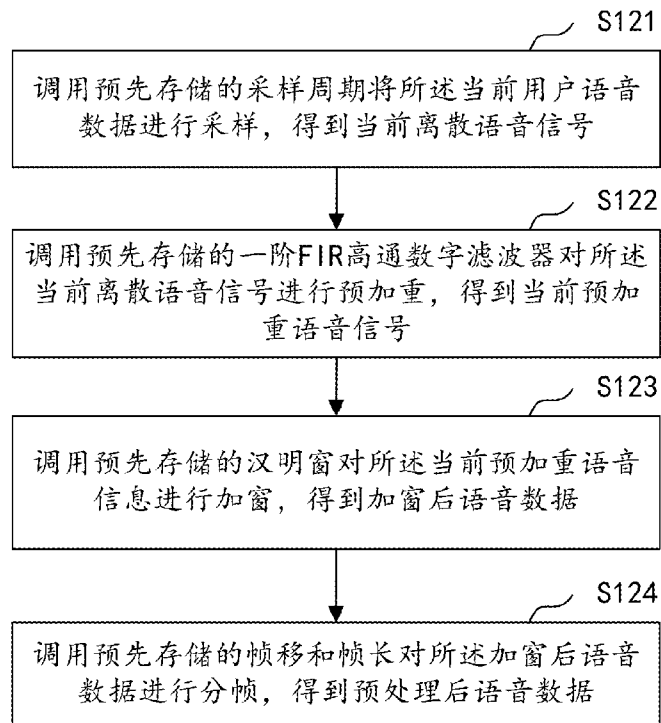


图 3

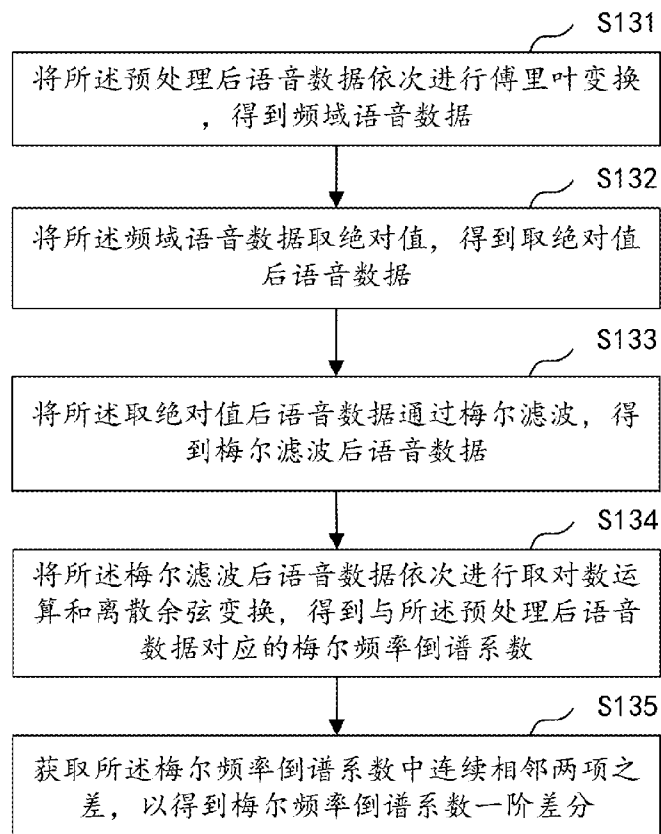


图 4

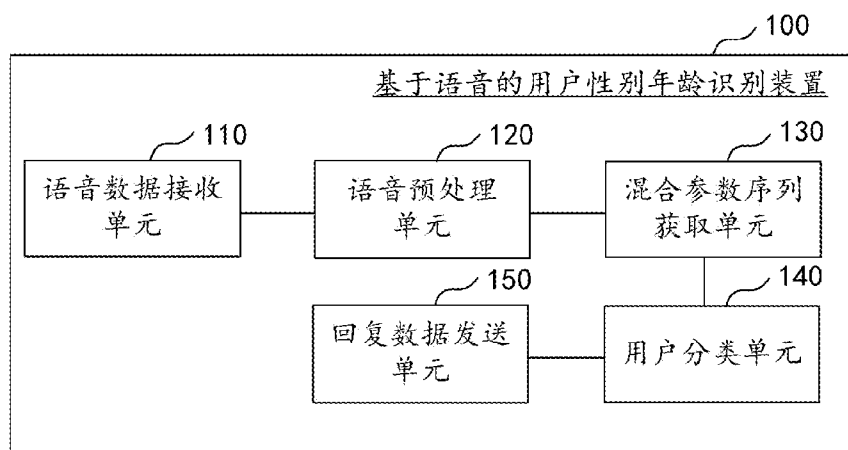


图 5

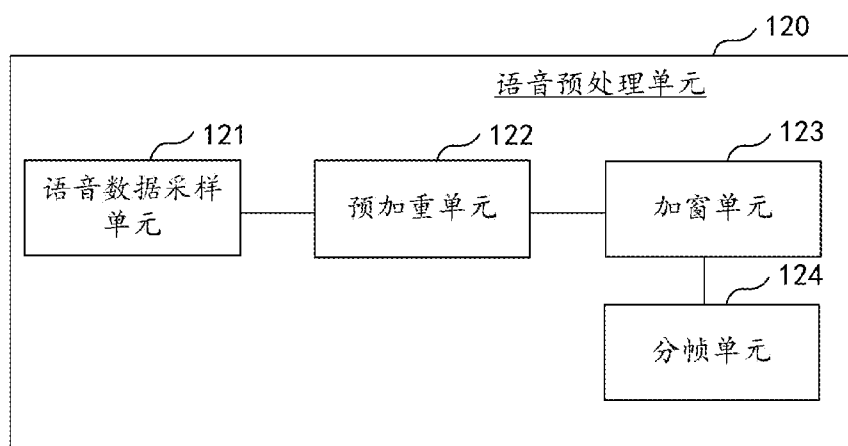


图 6

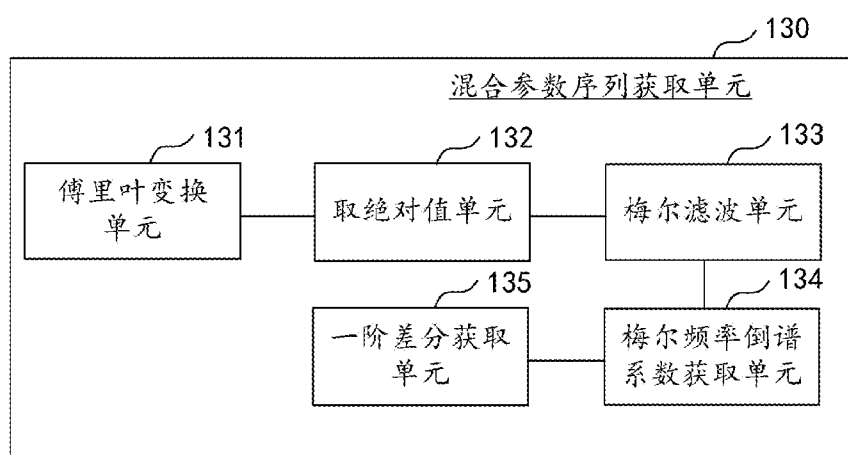


图 7

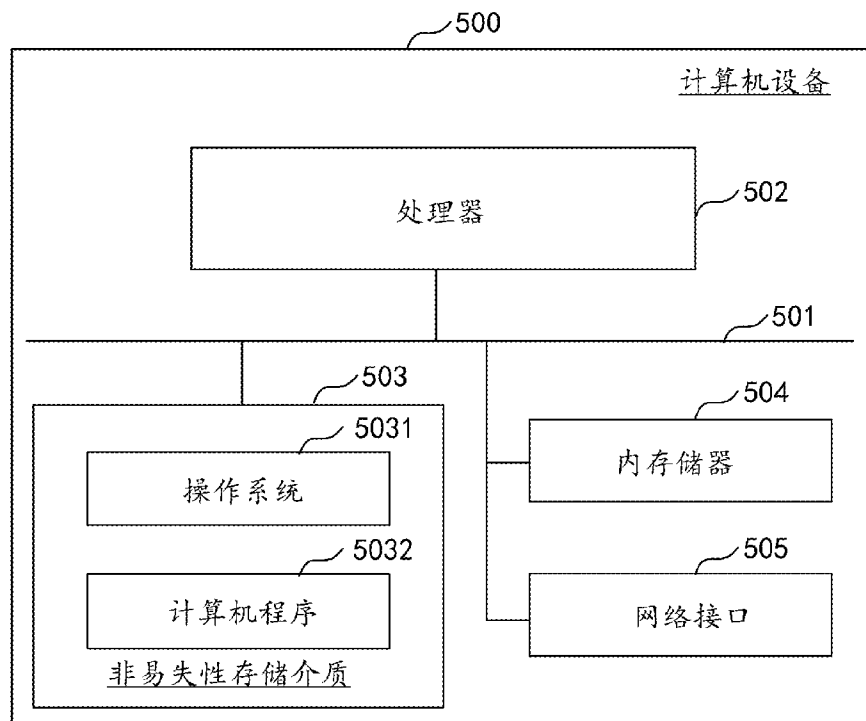


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/131612

A. CLASSIFICATION OF SUBJECT MATTER

H04M 3/51(2006.01)i; G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G10L 15/24(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)

H04M; 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, WPI, EPODOC, CNKI, IEEE: 用户, 性别, 年龄, 客服, 坐席, 语音回复, 平均幅度, 梅尔, 高斯混合模型, sex, gender, age, MFCC, MEL, GMM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111683181 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 18 September 2020 (2020-09-18) description, paragraphs [0033]-[0171]	1-20
Y	CN 110648672 A (SHENZHEN ZHUIYI TECHNOLOGY CO., LTD.) 03 January 2020 (2020-01-03) description, paragraphs [0044]-[0133], figure 6	1-20
Y	CN 104700843 A (HISENSE GROUP CO., LTD.) 10 June 2015 (2015-06-10) description, paragraphs [0062], [0088], [0116]-[0124], figure 7	1-20
Y	CN 103236259 A (LG ELECTRONICS RESEARCH & DEVELOPMENT CENTER (SHANGHAI) CO., LTD.) 07 August 2013 (2013-08-07) description, paragraphs [0036]-[0088], and figures 2-6	1-20
A	CN 108694954 A (SPEAKIN TECHNOLOGIES CO., LTD.) 23 October 2018 (2018-10-23) entire document	1-20
A	US 8160877 B1 (NARUS, INC.) 17 April 2012 (2012-04-17) 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

02 February 2021

Date of mailing of the international search report

03 March 2021

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/CN2020/131612

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	111683181	A	18 September 2020	None	
CN	110648672	A	03 January 2020	None	
CN	104700843	A	10 June 2015	None	
CN	103236259	A	07 August 2013	None	
CN	108694954	A	23 October 2018	None	
US	8160877	B1	17 April 2012	None	

国际检索报告

国际申请号

PCT/CN2020/131612

A. 主题的分类

H04M 3/51(2006.01)i; G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G10L 15/24(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)

H04M; 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, WPI, EPDOC, CNKI, IEEE: 用户, 性别, 年龄, 客服, 坐席, 语音回复, 平均幅度, 梅尔, 高斯混合模型, sex, gender, age, MFCC, MEL, GMM

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 111683181 A (平安科技深圳有限公司) 2020年 9月 18日 (2020 - 09 - 18) 说明书第[0033]-[0171]段	1-20
Y	CN 110648672 A (深圳追一科技有限公司) 2020年 1月 3日 (2020 - 01 - 03) 说明书第[0044]-[0133]段, 图6	1-20
Y	CN 104700843 A (海信集团有限公司) 2015年 6月 10日 (2015 - 06 - 10) 说明书第[0062], [0088], [0116]-[0124]段, 图7	1-20
Y	CN 103236259 A (乐金电子研发中心上海有限公司) 2013年 8月 7日 (2013 - 08 - 07) 说明书第[0036]-[0088]段, 图2-6	1-20
A	CN 108694954 A (广州势必可赢网络科技有限公司) 2018年 10月 23日 (2018 - 10 - 23) 全文	1-20
A	US 8160877 B1 (NARUS, INC.) 2012年 4月 17日 (2012 - 04 - 17) 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

2021年 2月 2日

国际检索报告邮寄日期

2021年 3月 3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

李文娟

电话号码 86-(10)-53961585

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/131612

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	111683181	A	2020年 9月 18日	无	
CN	110648672	A	2020年 1月 3日	无	
CN	104700843	A	2015年 6月 10日	无	
CN	103236259	A	2013年 8月 7日	无	
CN	108694954	A	2018年 10月 23日	无	
US	8160877	B1	2012年 4月 17日	无	