

(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
2019 年 5 月 31 日 (31.05.2019)



(10) 国际公布号
WO 2019/101083 A1

- (51) 国际专利分类号:
G10L 15/02 (2006.01) *G10L 15/22* (2006.01)
G10L 15/06 (2013.01)
- (21) 国际申请号: PCT/CN2018/116590
- (22) 国际申请日: 2018 年 11 月 21 日 (21.11.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201711191651.3 2017 年 11 月 24 日 (24.11.2017) 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: 马龙(MA, Long); 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)。李俊(LI, Jun); 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)。张力(ZHANG, Li); 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: 北京德琦知识产权代理有限公司 (DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路 1 号学院国际大厦 7 层, Beijing 100083 (CN)。

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

(54) Title: VOICE DATA PROCESSING METHOD, VOICE-BASED INTERACTIVE DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种语音数据处理方法、语音交互设备及存储介质

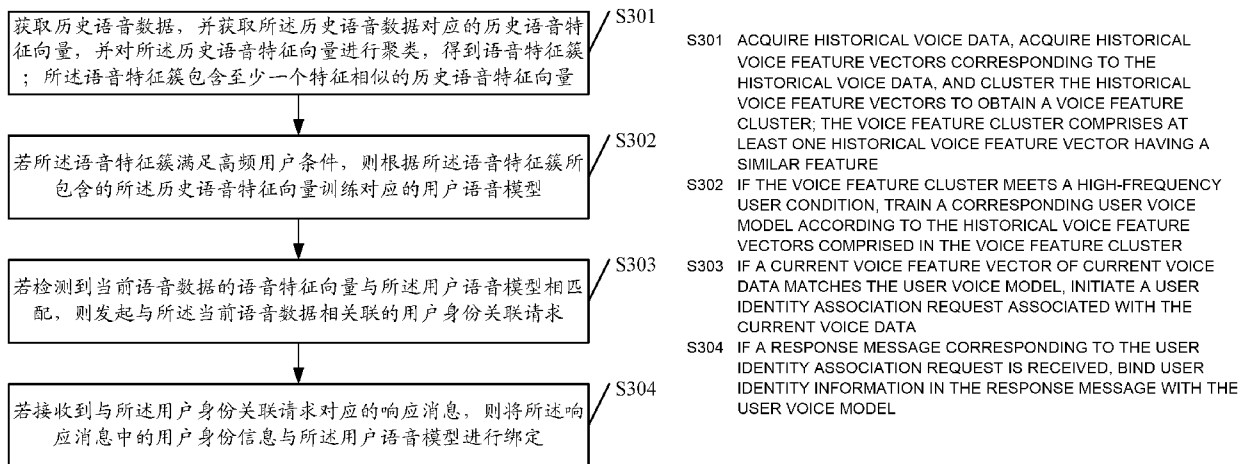


图 3

(57) Abstract: Embodiments of the application disclose a voice data processing method, a voice-based interactive device, and a storage medium. The method comprises: acquiring historical voice data, acquiring historical voice feature vectors corresponding to the historical voice data, and clustering the historical voice feature vectors to obtain a voice feature cluster, the voice feature cluster comprising at least one historical voice feature vector having a similar feature; if the voice feature cluster meets a high-frequency user condition, training a corresponding user voice model according to the historical voice feature vectors comprised in the voice feature cluster; if a current voice feature vector of current voice data matches the user voice model, initiating a user identity association request associated with the current voice data; and if a response message corresponding to the user identity association request is received, binding user identity information in the response message with the user voice model.

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: 本申请实施例公开了一种语音数据处理方法、语音交互设备及存储介质, 其中方法包括: 获取历史语音数据, 并获取历史语音数据对应的历史语音特征向量, 并对历史语音特征向量进行聚类, 得到语音特征簇; 语音特征簇包含至少一个特征相似的历史语音特征向量; 若语音特征簇满足高频用户条件, 则根据语音特征簇所包含的历史语音特征向量训练对应的用户语音模型; 若检测到当前语音数据的当前语音特征向量与用户语音模型相匹配, 则发起与当前语音数据相关联的用户身份关联请求; 若接收到与用户身份关联请求对应的响应消息, 则将响应消息中的用户身份信息与用户语音模型进行绑定。

一种语音数据处理方法、语音交互设备及存储介质

本申请要求于2017年11月24日提交中国专利局、申请号为201711191651.3、发明名称为“一种语音数据处理方法、装置以及语音交互设备”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

本申请涉及计算机技术领域，尤其涉及一种语音数据处理方法、语音交互设备及存储介质。

发明背景

10 随着语音识别技术的发展，有越来越多的产品使用到了语音识别技术，例如可以音控的汽车、音箱、电视等等，即语音交互设备可以对说话人的语音进行识别并根据识别的内容实现自动化控制。

可进行语音识别的语音交互设备可以针对不同说话人的语音特征进行个性化服务，在此之前，说话人需要主动对语音交互设备进行语音注册，以注册该说话人的语音特征和该说话人的说话人信息之间的关系，从而在后续识别出某语音与该说话人的语音特征相匹配后，可以提供与该说话人的说话人信息对应的使用权限。但是目前的语音注册过程通常都需要说话人对着语音交互设备重复且清晰地说出许多遍的固定句子，以提取说话人的语音特征，由此可见，目前的语音注册方式是需要由说话人主动发起，且注册时间可能会花费较长时间，导致语音注册效率低下；而且在语音注册过程中，说话人很容易因一时粗心导致说话人的语音内容与系统提供的固定句子不同，进而导致语音注册失败，从而降低了语音注册的成功率。

发明内容

本申请实施例提供一种语音数据处理方法、语音交互设备及存储介质，可提高语音注册效率，且可以提高语音注册的成功率。

本申请的一方面提供了一种语音数据处理方法，由语音交互设备执行，包括：

获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇；所述语音特征簇包含至少一个特征相似的历史语音特征向量；

若所述语音特征簇满足高频用户条件，则根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型；

若检测到当前语音数据的当前语音特征向量与所述用户语音模型相匹配，则发起与所述当前语音数据相关联的用户身份关联请求；

若接收到与所述用户身份关联请求对应的响应消息，则将所述响应消息中的用户身份信息与所述用户语音模型进行绑定。

本申请的另一方面提供了一种语音交互设备，包括：处理器、存储器；所述处理器与存储器相连，其中，所述存储器用于存储程序代码，所述处理器用于调用所述程序代码，以执行以下操作：

获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇；所述语音特征簇包含至少一个特征相似的历史语音特征向量；

若所述语音特征簇满足高频用户条件，则根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型；

若检测到当前语音数据的当前语音特征向量与所述用户语音模型相匹配，则发起与所述当前语音数据相关联的用户身份关联请求；

若接收到与所述用户身份关联请求对应的响应消息，则将所述响应消息中的用户身份信息与所述用户语音模型进行绑定。

本申请的另一方面提供了一种计算机存储介质，所述计算机存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时，执行上述语音数据处理方法。

附图简要说明

- 5 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术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他的附图。
- 10 图 1 是本申请实施例提供的一种系统架构的示意图；
图 2a 是本申请实施例提供的一种语音数据处理方法的场景示意图；
图 2b 是本申请实施例提供的另一种语音数据处理方法的场景示意图；
图 2c 是本申请实施例提供的又一种语音数据处理方法的场景示意图；
15 图 3 是本申请实施例提供的一种语音数据处理方法的流程示意图；
图 4 是本申请实施例提供的另一种语音数据处理方法的流程示意图；
图 5 是本申请实施例提供的一种参数更新方法的场景示意图；
图 6 是本申请实施例提供的一种性能验证结果的示意图；
20 图 7 是本申请实施例提供的另一种性能验证结果的示意图；
图 8 是本申请实施例提供的一种语音数据处理装置的结构示意图；
图 9 是本申请实施例提供的一种语音交互设备的结构示意图。

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作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。

5 请参见图 1，是本申请实施例提供的一种系统架构的示意图。如图 1 所示，该系统架构可以包括语音交互设备 100a 和后台服务器 100b，语音交互设备 100a 可以通过互联网连接后台服务器 100b；语音交互设备 100a 可以包括可进行语音识别的智能设备，如智能音箱、智能手机、电脑、智能电视、智能空调等。语音交互设备 100a 可以接收用户的语音数据，并将语音数据发送到后台服务器 100b，使得后台服务器 100b 可以对语音数据进行语音识别并根据识别出的语义生成控制指令，语音交互设备 100a 接收后台服务器 100b 发送的控制指令以执行对应的控制操作。例如，用户对着语音交互设备 100a 说“播放歌曲 A”，则语音交互设备 100a 可以将语音数据“播放歌曲 A”发送到后台服务器 100b，后台服务器 100b 对语音数据“播放歌曲 A”进行语音识别并生成对应的控制指令，语音交互设备 100a 接收该控制指令并根据该控制指令播放歌曲 A。

其中，后台服务器 100b 还可以根据语音交互设备 100a 所发送的语音数据进行高频用户的发现以及主动对高频用户发起身份注册，该过程可以一并参见图 2a-图 2c，均为本申请实施例提供的一种语音数据处理方法的场景示意图。如图 2a 所示，语音交互设备 100a 可以接收多个用户的语音数据，每个语音数据都可以转发到后台服务器 100b（其中根据语音数据进行语音控制的过程请参见上述图 1 对应实施例），因此，后台服务器 100b 可以存储大量语音数据，形成历史语音数据；如图 2a 所示，后台服务器 100b 可以对所有历史语音数据进行聚类，得到语音特征簇 1、语音特征簇 2、语音特征簇 3、语音特征簇 4，每个语音特征簇包含至少一个特征相似的历史语音特征向量，历史语音特征向量可以是

指历史语音数据对应的 i-Vector。如图 2a 所示，后台服务器 100b 可以进一步基于所有语音特征簇进行高频用户发现，具体可以通过分析每个语音特征簇中的历史语音特征向量的数量和历史语音特征向量的分布密度，来确定语音特征簇所对应的用户是否为高频用户，在图 2a 中，后台服务器 100b 分析出语音特征簇 1 和语音特征簇 4 所对应的用户属于高频用户(经常向语音交互设备 100a 发送语音数据的用户可以确定为高频用户)，后台服务器 100b 再进一步为语音特征簇 1 创建对应的用户语音模型 1，并为语音特征簇 4 创建对应的用户语音模型 4。其中，用户语音模型 1 和用户语音模型 4 均属于未绑定用户身份信息的用户语音模型，即未注册的用户语音模型。

进一步的，再请一并参见图 2b，在图 2b 中，语音交互设备 100a 可以将当前时刻的用户 1 的语音数据转发至后台服务器 100b(其中根据该语音数据进行语音控制的过程请参见上述图 1 对应实施例)，后台服务器 100b 可以对该语音数据进行模型匹配，具体是将该语音数据对应的 i-Vector 分别与图 2a 中的用户语音模型 1 和用户语音模型 4 进行比较，如图 2b 所示，用户语音模型 1 与该语音数据相匹配，此时，可以向语音交互设备 100a 发起与用户语音模型 1 对应的用户身份关联请求，语音交互设备 100a 中的身份关联模块可以向用户 1 发出用户注册提示音(例如，该提示音为“请输入您的身份信息”)，用户 1 根据该提示音可以通过语音或客户端的方式将用户 1 的身份信息发送给语音交互设备 100a，语音交互设备 100a 再将用户 1 的身份信息转发给后台服务器 100b，后台服务器 100b 可以对用户 1 的身份信息进行用户身份注册，该注册过程即为将用户语音模型 1 与用户 1 的身份信息进行绑定。其中，用户语音模型 1 属于已绑定用户身份信息的用户语音模型，即为已注册的用户语音模型；用户语音模型 4 仍属于未绑定用户身份信息的用户语音模型。由此可见，可以避免用户需要重复多次发出固定句子的语音内容才能实

现语音注册，即用户只需响应语音交互设备所发起的用户身份关联请求即可完成语音注册，进而可以提高语音注册的效率。

在对用户语音模型 1 进行身份信息绑定后，再一并参见图 2c，如图 2c 所示，后台服务器 100b 可以持续接收语音交互设备 100a 发送的语音数据，以形成更多的历史语音数据，为了保证后台服务器 100b 可以继续发现新的高频用户，后台服务器 100b 可以定时或定量的对历史语音数据进行重新聚类。在图 2c 中，后台服务器 100b 可以对所有历史语音数据中除了与用户语音模型 1 相匹配的历史语音数据以外的其他历史语音数据进行聚类（由于用户语音模型 1 已完成注册，所以无需再对与用户语音模型 1 相匹配的历史语音数据进行聚类），得到语音特征簇 2、语音特征簇 3、语音特征簇 4、语音特征簇 5；图 2c 中的语音特征簇 2 可以包含图 2a 中的语音特征簇 2，且还可以包含一些新增的历史语音特征向量；图 2c 中的语音特征簇 3 可以包含图 2a 中的语音特征簇 3，且还可以包含一些新增的历史语音特征向量；图 2c 中的语音特征簇 4 可以包含图 2a 中的语音特征簇 4，且还可以包含一些新增的历史语音特征向量；图 2c 中的语音特征簇 5 为新增加的语音特征簇。后台服务器 100b 进一步在语音特征簇 2、语音特征簇 3、语音特征簇 4、语音特征簇 5 中进行高频用户发现，进而分析出语音特征簇 3、语音特征簇 4 分别对应的用户属于高频用户，由于语音特征簇 4 已有对应的用户语音模型 4，所以只需要创建语音特征簇 3 对应的用户语音模型 3；其中，还可以通过图 2c 中的语音特征簇 4 中的所有历史语音特征向量对已有的用户语音模型 4 进行更新。如图 2c 所示，此时的后台服务器 100b 中的用户语音模型包括用户语音模型 1、用户语音模型 3、用户语音模型 4，用户语音模型 1 属于已绑定用户身份信息的用户语音模型；用户语音模型 3、用户语音模型 4 属于未绑定用户身份信息的用户语音模型，因此，后续在检测到语音数据与用户语音模型 3 或用户语音模型 4 相匹配时，可以发起用

户身份注册。随着用户使用量的增加，可以增加更多的未绑定用户身份信息的用户语音模型，且基于用户身份自动注册的机制，可以逐渐将未绑定用户身份信息的用户语音模型转换为已绑定用户身份信息的用户语音模型，即逐渐完成每个高频用户的身份注册。

- 5 后台服务器 100b 的所有功能均可以集成到语音交互设备 100a 中，即语音交互设备 100a 可以直接对语音数据进行语音识别以实现语音控制，语音交互设备 100a 也可以直接根据所接收到的语音数据进行高频用户发现以及主动对高频用户发起身份注册。

10 以下图 3-图 9 对应的实施例，以后台服务器 100b 集成到语音交互设备 100a 为例，对高频用户发现以及主动对高频用户发起身份注册的具体过程进行详细描述。

请参见图 3，是本申请实施例提供的一种语音数据处理方法的流程示意图，所述方法可以包括：

- 15 S301，获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇；所述语音特征簇包含至少一个特征相似的历史语音特征向量；

具体的，在启动语音交互设备（该语音交互设备可以具体为上述图 1 对应实施例中的集成有后台服务器 100b 的所有功能的语音交互设备 100a）后，语音交互设备可直接对所获取到的用户语音进行语义识别，
20 进而执行与语义相关联的控制操作。其中，语音交互设备可以包括音响/音箱、电视、汽车、手机、VR（Virtual Reality，虚拟现实）设备等可进行语音交互识别和控制的智能设备。例如，用户对语音交互设备说“播下一首歌”，则语音交互设备在分析出该语义后，语音交互设备即可将当前歌曲切换到下一首歌进行播放。因此，语音交互设备可以无需等待
25 用户完成语音注册后才能启动语音控制功能，即在将用户的语音特征与用户身份信息进行绑定之前，语音交互设备即可根据用户的语音内容执

行相关联的控制操作。而且多个不同的用户均可以向该语音交互设备说出用户语音，使得语音交互设备可以根据各用户语音指令执行相关联的控制操作，语音交互设备还可以记录和保存各用户语音，并将所保存的每条用户语音均确定为历史语音数据。

- 5 当所保存的历史语音数据的数量达到第一数量阈值时，获取所有历史语音数据，并确定每个历史语音数据分别对应的历史语音特征向量。历史语音特征向量可以为 i-vector (identity-vector)。其中获取 i-vector 的过程可以为：首先利用所有历史语音数据训练得到高阶的高斯混合模型 (GMM, Gaussian Mixture Model)，GMM 可以用于刻画说话人的语音特征空间，这个模型通常被称作通用背景模型 (UBM, Universal Background Model)，即 GMM-UBM 模型；再利用 GMM-UBM 模型对
- 10 每条历史语音数据进行参数估计，以确定高斯混合模型各分量的混合权重、均值向量、方差矩阵分别对应的零阶、一阶和二阶 Baum-Welch 统计量，然后再利用 EM (Expectation Maximization Algorithm, 期望最大化算法) 算法迭代得到全局差异矩阵 T；通过矩阵 T，可以把分别把每条历史语音数据中隐含在高维说话人语音空间中的说话人及通道相关特性，投影到低维空间中，从而获得每条历史语音数据的历史语音特征向量，也就是 i-Vector。即确定 i-Vector 的具体公式可以为： $M=m+Tw$ ，其中，M 代表高斯混合模型的高维均值超矢量，m 代表与说话人信息和
- 15 信道信息无关的一个超矢量，T 为全局差异空间，w 是包含整段语音中的说话人信息和信道信息的一个全差异因子 (即 i-Vector)。全局差异矩阵 T 也可以基于深度神经网络训练而成。
- 20

- 语音交互设备可以进一步对历史语音特征向量进行降维，并根据目标聚类模型参数对降维后的历史语音特征向量进行聚类，得到语音特征簇。其中，对历史语音特征向量进行降维的过程可以为：采用 PCA (Principal Component Analysis, 主成份分析)、tSNE (t-distributed
- 25

stochastic neighbor embedding, t 分布领域嵌入) 及 LDA (Linear Discriminant Analysis, 线性判别分析) 等算法, 对获取到的历史语音特征向量 (即 i-vector) 进行数据降维处理, 去除数据中冗余的多重共线成分, 减小聚类的计算量。其中, 利用 PCA 和 tSNE 的降维是非监督的, 5 即毋须预先训练模型, 可以直接应用于对 i-vector 进行降维; 其中, 利用 LDA 降维需要预先使用带实际标签的 i-vector 数据训练出最优的投影方向, 然后将其应用于对 i-vector 进行降维。

其中, 对所有历史语音特征向量进行聚类的具体过程可以为: 采用 DBSCAN (Density-Based Spatial Clustering of Applications with Noise, 基
10 于密度的带噪声应用的空间聚类) 聚类算法, 使用欧式距离作为样本的距离度量, 将降维后的历史语音特征向量聚类成簇 (即语音特征簇)。DBSCAN 聚类算法可以找出特征空间中形状不规则的簇, 且聚类时毋须事先设定簇的数量, 从而可以满足本申请实施例中说话人数量事先未知的场景需求。

15 S302, 若所述语音特征簇满足高频用户条件, 则根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型;

具体的, 在生成至少一个语音特征簇后, 针对每一个语音特征簇, 均可以根据语音特征簇中的历史语音特征向量的数量以及该语音特征簇中的历史语音特征向量的分布情况, 确定该语音特征簇是否满足高频
20 用户条件。例如, 语音特征簇中的历史语音特征向量的数量超过预设的数量阈值以及该语音特征簇中的历史语音特征向量的分布密度也超过预设的密度阈值, 则可以确定该语音特征簇满足高频用户条件, 也即说明该语音特征簇所对应的说话人为经常与语音交互设备进行语音交互的用户。

25 在确定出某语音特征簇满足高频用户条件后, 可以根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型。其中,

训练用户语音模型的过程可以为：获取满足高频用户条件的语音特征簇中的所有历史语音特征向量，并对所获取的这些历史语音特征向量进行均值计算或插值计算，得到目标历史语音特征向量，并将所述目标历史语音特征向量作为所述语音特征簇对应的用户语音模型的模型参数。其中，对语音特征簇中的历史语音特征向量进行均值计算的方式可以为：将语音特征簇中的各历史语音特征向量进行相加，再除以该语音特征簇中的历史语音特征向量的数量，得到目标历史语音特征向量。或者，对语音特征簇中的历史语音特征向量进行均值计算的方式可以为：根据权重系数对语音特征簇中的各历史语音特征向量进行权重相加，再除以该语音特征簇中的历史语音特征向量的数量，得到目标历史语音特征向量。

S303，若检测到当前语音数据的语音特征向量与所述用户语音模型相匹配，则发起与所述当前语音数据相关联的用户身份关联请求；

具体的，语音交互设备每接收到一条语音数据，都会获取该语音数据的语音特征向量（即 i-vector），然后将该语音数据的语音特征向量与各个已创建的用户语音模型进行比较。如果当前接收到的一条语音数据（即当前语音数据）与某一个用户语音模型相匹配，且该用户语音模型未与用户身份信息进行绑定，则语音交互设备可以发起与当前语音数据相关联的用户身份关联请求，该用户身份关联请求的具体形式可以为：用于进行用户身份关联的语音提示（如语音交互设备发送语音“请绑定您的身份信息”），或者是向用户终端发送的用于进行用户身份关联的注册界面（如该注册界面可以在用户手机上显示，用户通过在注册界面上填写自己的身份信息，或也可以在注册界面上与用户账号进行绑定，以完成语音注册）。其中，可以通过欧氏距离分别确定当前语音数据的 i-vector 与每个用户语音模型中的 i-vector 之间的向量距离，将向量距离小于距离阈值的用户语音模型确定为与当前语音数据相匹配的用户语音模型。

若与当前语音数据相匹配的用户语音模型为已绑定用户身份信息
的用户语音模型，则将当前语音数据对应的语音特征向量保存至相匹配
的用户语音模型对应的语音特征簇中，以便于用于后续更新该用户语音
模型以提高该用户语音模型的准确性，同时，语音交互设备也可以根据
5 该用户语音模型向当前语音数据对应的说话人提供相应的个性化服务，
例如，若当前语音数据为“播放 A 歌曲”，则语音交互设备可以根据该
用户语音模型所绑定的用户身份信息，获取与该用户身份信息对应的用
户习惯参数（如该用户所喜欢的音调、音量等参数），并根据用户习惯
参数对 A 歌曲的音频参数进行调整和播放；或者，若该用户语音模型所
10 绑定的用户身份信息为管理员身份信息，则语音交互设备可以向当前语
音数据对应的说话人开放系统管理权限。

S304，若接收到与所述用户身份关联请求对应的响应消息，则将所
述响应消息中的用户身份信息与所述用户语音模型进行绑定；

具体的，语音交互设备发起用户身份关联请求后，当前语音数据对
15 应的说话人可以通过语音反馈对应的响应消息以完成语音注册。例如，
说话人可以说出响应消息“我的身份信息为 XXXX”，则语音交互设备
可以通过语音识别得知该响应消息中的用户身份信息为“XXXX”，进行
将用户身份信息“XXXX”与当前语音数据相匹配的用户语音模型进行
绑定。

20 或者，若用户身份关联请求的具体形式是向用户终端发送的用于进
行用户身份关联的注册界面，则当前语音数据对应的说话人可以通过该
注册界面输入对应的响应消息以完成语音注册。例如，注册界面中包含
用户名称输入框、密码输入框、用户兴趣爱好输入框等等，说话人可以
在注册界面中的各个输入框中输入相应的数据，并在点击提交后，用户
25 终端可以将注册界面中所输入的数据封装为响应消息，并将该响应消息
发送给语音交互设备，使得语音交互设备将响应消息中的用户身份信息

(如包含注册界面中所输入的用户名称、密码、用户兴趣爱好等信息)与当前语音数据相匹配的用户语音模型进行绑定。

请参见图 4, 是本申请实施例提供的另一种语音数据处理方法的流程示意图, 所述方法可以包括:

5 S401, 获取所有历史语音数据, 并根据所述所有历史语音数据训练高斯混合模型和全局差异空间矩阵;

S402, 根据所述高斯混合模型和所述全局差异空间矩阵将所述所有历史语音数据投影至向量空间, 生成每个历史语音数据分别对应的历史语音特征向量, 并对所述历史语音特征向量进行降维;

10 其中, S401-S402 步骤的具体实现方式可以参见上述图 3 对应实施例中的 S301, 这里不再进行赘述。

S403, 根据目标聚类模型参数对降维后的历史语音特征向量进行聚类, 得到所述语音特征簇;

具体的, 语音交互设备 (该语音交互设备可以具体为上述图 1 对应
15 实施例中的集成有后台服务器 100b 的所有功能的语音交互设备 100a) 可以基于 DBSCAN 聚类算法对降维后的历史语音特征向量进行聚类, DBSCAN 聚类算法可以假设聚类结构能够通过样本分布的紧密程度来确定, 该紧密程度可由一对参数(Eps, MinPts)刻画, Eps 为定义密度时的邻域半径, MinPts 为定义核心样本时的阈值, 即目标聚类模型参数可以包括: Eps (即密度领域半径) 和 MinPts (即核心样本阈值)。基于
20 DBSCAN 聚类算法可以以所有降维后的历史语音特征向量为样本点生成包含所述样本点的样本数据集, 并根据所述密度领域半径和所述核心样本阈值在所述样本数据集中查找所有为核心点的样本点; 在所有核心点中确定任意一个核心点为出发点, 并在所述样本数据集中查找与所述
25 出发点具有密度可达关系的所有样本点, 作为可达样本点 (为了区别其他的样本点, 所以这里将与所述出发点具有密度可达关系的样本点定义

为可达样本点), 并生成包含所述出发点和所有所述可达样本点的语音特征簇, 并将所有核心点中的下一个核心点确定为所述出发点, 重复执行本步骤, 直至所有核心点均被确定为所述出发点。

例如, 假设将所有历史语音数据对应的历史语音特征向量确定为样本数据集 $D = \{x_1, x_2, \dots, x_m\}$, 其中每个样本点 x_j 即为一条历史语音数据对应的历史语音特征向量, 设任意两个样本点的距离函数为 $\text{dist}()$ 。其中, Eps-邻域的定义为: 对 $x_j \in D$, 其 Eps-邻域包含 D 中与 x_j 的距离不大于 Eps 的样本点, 即 $N_{\text{Eps}}(x_j) = \{x_i \in D \mid \text{dist}(x_i, x_j) \leq \text{Eps}\}$ 。密度直达的定义为: 若 x_j 位于 x_i 的 Eps-邻域中, 且 x_i 是核心点, 则称 x_j 由 x_i 密度直达。密度可达的定义为: 对 x_i 与 x_j , 若存在样本点序列 $p_1, p_2, \dots, p_n$, 其中 $p_1 = x_i$, $p_n = x_j$, 且 p_{i+1} 由 p_i 密度直达, 则称 x_j 由 x_i 密度可达。密度相连的定义为: 对 x_i 与 x_j , 若存在 x_k 使得 x_i 与 x_j 均由 x_k 密度可达, 则称 x_i 与 x_j 密度相连。基于上述概念, DBSCAN 聚类算法可以将样本数据集 D 中的所有样本点分为三类: 核心点、边界点以及噪声点; 其中, 核心点为: 在半径 Eps 内含有不少于 MinPts 个样本点的样本点; 边界点为: 在半径 Eps 内样本点数量小于 MinPts 且落在某核心点邻域内的样本点; 噪声点为: 既非核心点也非边界点的样本点。因此, 一个聚类簇定义为一个具有密度相连关系的核心点与边界点的集合。DBSCAN 聚类算法首先根据参数(Eps, MinPts)找出样本数据集 D 中所有的核心点, 然后以任意核心点为出发点找出所有由其密度可达的样本点, 作为可达样本点, 并生成包含出发点和所有可达样本点的语音特征簇, 直到所有的核心点均被访问为止, 即每个语音特征簇均可以包含至少一个特征相似的历史语音特征向量。生成语音特征簇的具体算法流程描述如下:

输入: 样本数据集 $D = \{x_1, x_2, \dots, x_m\}$; 参数(Eps, MinPts);

输出: 语音特征簇划分 $C = \{C_1, C_2, \dots, C_k\}$;

过程: 初始化核心点集合: $\Omega = \emptyset$,

```

for j = 1,2,...,m, do
    确定样本点  $x_j$  的 Eps-邻域  $N_{Eps}(x_j)$ ,
    if  $|N_{Eps}(x_j)| \geq \text{MinPts}$  then 将样本点  $x_j$  加入核心点集合  $\Omega$ ,
    end if,
5   end for,
    初始化语音特征簇数:  $k = 0$ ,
    初始化未访问样本数据集合  $\Gamma = D$ ,
    while  $\Omega \neq \emptyset$  do
        记录当前未访问样本数据集合:  $\Gamma_{old} = \Gamma$ ,
10        随机选取一个核心点  $\omega \in \Omega$ , 初始化队列  $Q = \langle \omega \rangle$ ,
         $\Gamma = \Gamma \setminus \{\omega\}$ ,
        while  $Q \neq \emptyset$  do
            取出队列  $Q$  中的首个样本点  $q$ ,
            if  $|N_{Eps}(q)| \geq \text{MinPts}$  then
15                令  $\Delta = N_{Eps}(q) \cap \Gamma$ ,
                将  $\Delta$  中的样本点加入队列  $Q$ ,
                 $\Gamma = \Gamma \setminus \Delta$ ,
            end if,
        end while,
    end while,
20     $k = k + 1$ , 生成语音特征簇  $C_k = \Gamma_{old} \setminus \Gamma$ ,
     $\Omega = \Omega \setminus C_k$ ,
end while;
```

由上述内容可知, DBSCAN 聚类算法中两个关键参数 Eps 和 MinPts 直接决定着聚类的性能。

25 S404, 获取所述语音特征簇所包含的所述历史语音特征向量的数量, 并根据所述语音特征簇所包含的所述历史语音特征向量的数量, 以及所

述语音特征簇所包含的所述历史语音特征向量，确定所述语音特征簇对应的类内散度；

具体的，若该语音特征簇所包含的历史语音特征向量的数量大于系统数量阈值，且类内散度小于系统类内散度阈值，则确定该语音特征簇满足高频用户条件，进而可以分析出哪些语音特征簇满足高频用户条件。其中，确定类内散度 divergence 的公式为：

$$\text{divergence} = \frac{2}{|C|(|C|-1)} \sum_{1 \leq i < j \leq |C|} \|x_i - x_j\|_2,$$

其中， $|C|$ 表示某语音特征簇内的样本数量（即语音特征簇所包含的历史语音特征向量的数量）， x_i 和 x_j 是该语音特征簇的两个样本点（即该语音特征簇中的两个历史语音特征向量）， $\|\cdot\|_2$ 表示计算代数式的 2-范数。其中，系统数量阈值可以为 base_frequency，系统类内散度阈值可以为 base_divergence，即若该语音特征簇内的样本数量大于 base_frequency 且该语音特征簇的类内散度小于 base_divergence，则可以确定该语音特征簇满足高频用户条件。其中，base_frequency 和 base_divergence 为由系统设置的超参数。

S405，若所述语音特征簇所包含的所述历史语音特征向量的数量大于系统数量阈值，且所述类内散度小于系统类内散度阈值，则确定所述语音特征簇满足高频用户条件，并根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型；

具体的，在确定出某语音特征簇满足高频用户条件后，可以根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型。其中，训练用户语音模型的过程可以为：获取满足高频用户条件的语音特征簇中的所有历史语音特征向量，并对所获取的这些历史语音特征向量进行均值计算或插值计算，得到目标历史语音特征向量，并将所述目标历史语音特征向量作为所述语音特征簇对应的用户语音模型的

模型参数。

S406, 若检测到当前语音数据的当前语音特征向量与所述用户语音模型相匹配, 则发起与所述当前语音数据相关联的用户身份关联请求;

S407, 若接收到与所述用户身份关联请求对应的响应消息, 则将所述响应消息中的用户身份信息与所述用户语音模型进行绑定;

其中, S406-S407 的具体实现方式可以参见上述图 3 对应实施例中的 S303-S304, 这里不再进行赘述。

S408, 若聚类后所累计新增的历史语音数据的数量达到第一数量阈值, 或聚类后所累计时长达到第一时长阈值, 则更新当前的聚类模型参数, 得到所述目标聚类模型参数;

具体的, 由于随着历史语音数据的增加, 可能会新增一些高频用户, 因此, 需要定时进行重新聚类, 以划分出新的语音特征簇, 并且在新的语音特征簇满足高频用户条件时可以进一步训练对应的用户语音模型以及绑定相应的用户身份信息。而 DBSCAN 聚类算法中两个关键参数 Eps 和 MinPts 直接决定着聚类的性能, 为了逐步提高聚类算法的性能, 可以定时对 Eps 和 MinPts 进行更新, 即 Eps 和 MinPts 越准确, 则聚类出的新的语音特征簇就越精确。因此, 若聚类后所累计新增的历史语音数据的数量达到第一数量阈值, 或聚类后所累计时长达到第一时长阈值, 则获取与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量, 作为第一历史语音特征向量; 根据聚类算法性能参数最大化条件、所述第一历史语音特征向量与已绑定的所述用户身份信息之间的对应关系, 更新当前的聚类模型参数, 得到所述目标聚类模型参数。其中, 聚类算法性能参数可以包括两个外部指标, 即 Jaccard 系数 (杰卡德系数, JC) 和 Rand 指数 (兰德指数, RI), 通过 JC 和 RI 可以衡量聚类算法的性能, 即当聚类性能提升时, JC 和 RI 也会随之增大; 其中, 聚类算法性能参数最大化条件可以是指 JC 最大化的条件。其中, $JC = SS /$

(SS+SD+DS), $RI = (SS+DD) / (SS+SD+DS+DD)$, 其中, SS 表示实际标签相同且聚类标签也相同的样本点对的数量, SD 表示实际标签相同但聚类标签不同的样本点对的数量, DS 表示实际标签不同但聚类标签相同的样本点对的数量, DD 表示实际标签不同且聚类标签也不同的样本点对的数量 (这里的标签可以是指说话人的身份信息)。

例如, 更新当前的聚类模型参数的具体过程可以为: 获取与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量, 作为第一历史语音特征向量, 并将 70% 的第一历史语音特征向量作为训练集, 并将剩余的 30% 的第一历史语音特征向量作为验证集; 使用训练集训练一个 DBSCAN 聚类模型, 训练目标为最大化 JC; 为避免训练过拟合, 在训练过程中确定该聚类模型在验证集上的 JC, 选择使验证集 JC 值最大的 Eps 和 MinPts 参数作为优化的模型参数 (即目标聚类模型参数)。之后可以继续定时或定量的更新目标聚类模型参数, 使得可以逐渐优化目标聚类模型参数。

也可以在生成目标聚类模型参数后累计新增的历史语音数据的数量, 且该数量达到第一数量阈值时, 执行 S408 步骤; 或者, 在生成目标聚类模型参数后开始累计时长, 且该累计时长达到第一时长阈值时, 执行 S408 步骤。

S409, 若聚类后所累计新增的历史语音数据的数量达到第二数量阈值, 或聚类后所累计时长达到第二时长阈值, 则对未绑定所述用户身份信息的用户语音模型所对应的语音特征簇进行更新, 并对未满足所述高频用户条件的语音特征簇进行替换;

具体的, 由于随着历史语音数据的增加, 可能会新增一些高频用户, 因此, 需要定时进行重新聚类, 以划分出新的语音特征簇, 进而可以通过新的语音特征簇来发现新的高频用户。因此, 若聚类后所累计新增的历史语音数据的数量达到第二数量阈值, 或聚类后所累计时长达到第二

时长阈值，则获取与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，以及与所有用户语音模型均不匹配的历史语音特征向量（即不属于高频用户的历史语音特征向量），作为第二历史语音特征向量，并对所述第二历史语音特征向量进行聚类，得到当前生成的语音特征簇；其中，对第二历史语音特征向量进行聚类的过程可以参见上述步骤 S403，这里不再进行赘述。其中，对第二历史语音特征向量进行聚类之前，可以先对还未降维的第二历史语音特征向量进行降维处理。

再根据所述当前生成的语音特征簇对未绑定所述用户身份信息的用户语音模型所对应的语音特征簇进行更新，该更新过程可以具体为：检测每个当前生成的语音特征簇是否满足高频用户条件，并将当前生成的语音特征簇中满足高频用户条件的语音特征簇确定为待更新语音特征簇，训练待更新语音特征簇对应的用户语音模型，并将待更新语音特征簇对应的用户语音模型与重新聚类之前已存在的未绑定所述用户身份信息的用户语音模型进行比较，若存在某个待更新语音特征簇对应的用户语音模型与某个未绑定所述用户身份信息的用户语音模型相近似（如两个用户语音模型的 i-Vector 之间的向量距离小于预设距离阈值），则可以将该待更新语音特征簇中的用户画像数据传导和继承到与之具有相近似的用户语音模型的语音特征簇中，以完成对该未绑定所述用户身份信息的用户语音模型所对应的语音特征簇的更新。

再根据所有当前生成的语音特征簇中除了已用于传导和继承用户画像数据的待更新语音特征簇以外的语音特征簇，对未满足所述高频用户条件的语音特征簇进行替换，即将重新聚类之前已存在的未满足所述高频用户条件的语音特征簇删除，并保留所有当前生成的语音特征簇中除了已用于传导和继承用户画像数据的待更新语音特征簇以外的语音特征簇。例如，若重新聚类之前存在语音特征簇 a1（未满足高频用户

条件)、语音特征簇 a2 (未满足高频用户条件)、语音特征簇 a3 (具有未绑定用户身份信息的用户语音模型)、语音特征簇 a4 (具有未绑定用户身份信息的用户语音模型)、语音特征簇 a5 (具有已绑定用户身份信息的用户语音模型); 重新聚类后, 得到当前生成的语音特征簇 b1、语音特征簇 b2、语音特征簇 b3、语音特征簇 b4, 其中, 语音特征簇 b1、语音特征簇 b2 均未满足高频用户条件, 语音特征簇 b3 和语音特征簇 b4 均满足高频用户条件, 并进一步训练语音特征簇 b3 对应的用户语音模型和语音特征簇 b4 对应的用户语音模型, 其中, 语音特征簇 b3 对应的用户语音模型与语音特征簇 a4 对应的用户语音模型相近似, 因此, 可以将语音特征簇 b3 中的用户画像数据传导和继承到语音特征簇 a4 中, 以完成对语音特征簇 a4 的更新; 其中, 语音特征簇 b4 对应的用户语音模型与语音特征簇 a4 对应的用户语音模型、语音特征簇 a3 对应的用户语音模型均不相似, 进而可以保留语音特征簇 b4、语音特征簇 b1、语音特征簇 b2, 并删除语音特征簇 a1 和语音特征簇 a2, 此时, 语音交互设备中的所有语音特征簇包括: 语音特征簇 b4、语音特征簇 b1、语音特征簇 b2、更新后的语音特征簇 a4、语音特征簇 a3 以及语音特征簇 a5。

S410, 若聚类后所累计新增的历史语音数据的数量达到第三数量阈值, 或聚类后所累计时长达到第三时长阈值, 则更新用户语音模型;

具体的, 若聚类后所累计新增的历史语音数据的数量达到第三数量阈值, 或聚类后所累计时长达到第三时长阈值, 则获取与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量, 作为第三历史语音特征向量, 并根据所述第三历史语音特征向量更新已绑定所述用户身份信息的用户语音模型; 其中, 某个已绑定所述用户身份信息的用户语音模型对应的第三历史语音特征向量可以包括: 已有的历史语音特征向量和聚类后所累计新增的历史语音特征向量; 该用户语音模型的模型参数 (模型参数也为 i-Vector) 是根据已有的历史语音特征向量生

成的，因此，更新该用户语音模型的过程可以为：对该用户语音模型的模型参数和聚类后所累计新增的历史语音特征向量进行均值计算或插值计算，得到更新后的历史语音特征向量，并用该更新后的历史语音特征向量替换该用户语音模型的模型参数，以完成对该用户语音模型的更新。以均值计算的方式更新用户语音模型为例，某个已绑定用户身份信息

5 的用户语音模型 A 包含模型参数 a1, 且聚类后所新增的且与该用户语音模型 A 相匹配的历史语音特征向量包括：历史语音特征向量 b1、历史语音特征向量 b2、.....、历史语音特征向量 bn，则更新后的用户语音模型 A 所包含的模型参数= (a*a1+b* (b1+b2+.....+bn)) / (n+1)，

10 其中，a 和 b 为权重值。

在更新已绑定所述用户身份信息

的用户语音模型的同时，还可以获取与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第四历史语音特征向量，并根据所述第四历史语音特征向量更新未绑定所述用户身份信息的用户语音模型。其中，更新未绑定

15 所述用户身份信息的用户语音模型的具体过程与更新已绑定所述用户身份信息的用户语音模型的过程相同，这里不再进行赘述。

还可以在更新某用户语音模型后累计与该用户语音模型相匹配的新增历史语音数据的数量，且该数量达到第三数量阈值时，执行 S410 步骤；或者，在更新某用户语音模型后开始累计时长，且该累计时长达到第三时长阈值时，执行 S410 步骤。

20

其中，S408 的步骤可以在 S401-S407 之间的任一时刻或 S401 之前或 S407 之后执行，即每次聚类后均可以定时或定量的更新当前的聚类模型参数，因此，不对 S408 的步骤执行顺序进行限定。S409 的步骤可以在 S401-S407 之间的任一时刻或 S401 之前或 S407 之后执行，即可以

25 定时或定量的进行重新聚类，以更新或替换相应的语音特征簇，因此，不对 S409 的步骤执行顺序进行限定。S410 的步骤可以在 S401-S407 之

间的任一时刻或 S401 之前或 S407 之后执行，即每次聚类后均可以定时或定量的更新相应的用户语音模型，因此，不对 S410 的步骤执行顺序进行限定。

其中，第一数量阈值、第二数量阈值、第三数量阈值之间可以相同或不同，第一时长阈值、第二时长阈值、第三时长阈值之间也可以相同或不同，这里不对其进行限定。若为了保证语音交互设备的工作效率，可以设置第一数量阈值略小于第二数量阈值（两个数量阈值之间的差值很小），或第一时长阈值略小于第二时长阈值（两个时长阈值之间的差值很小），以保证在每次聚类之前都先更新目标聚类模型参数，使得每次聚类都可以基于更新后的目标聚类模型参数进行聚类，以提高每次聚类的准确性；且可以设置第一数量阈值和第二数量阈值均大于第三数量阈值，或设置第一时长阈值和第二时长阈值均大于第三时长阈值，以避免过于频繁的更新目标聚类模型参数和语音特征簇，因此，过于频繁的更新容易导致更新前后的两个目标聚类模型参数过于相似，进而导致系统资源的浪费，且过于频繁的更新也容易导致更新前后的语音特征簇没有太大的变化，进而导致系统资源的浪费；而对于用户语音模型，可以较为频繁的更新，以保证用户语音模型的准确性，使得用户的语音可以更快、更准确地匹配到正确的用户语音模型。

为了提高计算每个语音数据对应的 i-vector 的准确性，也可以定时或定量的更新 GMM，随着时间的推移，所累计的历史语音数据越来越多，进而根据数量增加后的所有历史语音数据训练 GMM，可以提高 GMM 的准确性，进而在更新 GMM 后，可以提高所计算出的 i-vector 的准确性。

在 S401 的步骤之前（如语音交互设备在出厂之前的阶段，即还未接到任何用户的语音），语音交互设备可以获取样本语音数据，并为所述样本语音数据设置对应的样本用户身份标签（即已知每一条样本语音

数据对应的说话人信息), 再根据聚类算法性能参数最大化条件、所述样本语音数据与所述样本用户身份标签之间的对应关系, 训练初始聚类模型参数, 并将训练后的初始聚类模型参数确定为所述目标聚类模型参数。训练初始聚类模型参数的具体过程可以参见上述 S408 步骤中对当前的聚类模型参数进行更新的过程, 这里不再进行赘述。在得到初始聚类模型参数后, 可以根据初始聚类模型参数进行第一次聚类, 并将初始聚类模型参数确定为目标聚类模型参数, 此后即可定时或定量地对目标聚类模型参数进行更新。例如, 获取 20 组含说话人实际身份标签 (即样本用户身份标签) 的唤醒词语音数据 (即样本语音数据), 每组包含 10 个说话人, 每个说话人含 10 条唤醒词语音数据, 从每组中随机选取 7 个说话人的唤醒词语音数据作为训练集, 剩余 3 个说话人的唤醒词语音数据作为验证集; 对于每组数据, 提取唤醒词语音数据的 i-vector 并降维后, 使用训练集训练一个 DBSCAN 聚类模型, 训练目标为最大化 JC; 为避免训练过拟合, 在训练过程中确定该聚类模型在验证集上的 JC, 选择使验证集 JC 值最大的 Eps 和 MinPts 参数作为初始聚类模型参数。

进一步的, 请一并参见图 5, 是本申请实施例提供的一种参数更新方法的场景示意图。如图 5 所示, 语音交互设备在首次聚类之前, 可以先获取样本语音数据, 并生成样本语音数据对应的 i-vector (这里可以为降维后的 i-vector), 并根据样本语音数据对应的 i-vector 训练一个 DBSCAN 聚类模型, 训练目标为最大化 JC; 为避免训练过拟合, 在训练过程中确定该聚类模型在验证集上的 JC, 选择使验证集 JC 值最大的 Eps 和 MinPts 参数作为初始聚类模型参数, 即初始化的 Eps 和 MinPts。如图 5 所示, 语音交互设备在首次聚类之前, 可以先生成历史语音数据对应的 i-vector (这里可以为降维后的 i-vector), 并根据初始化的 Eps 和 MinPts 对历史语音数据对应的 i-vector 进行 DBSCAN 聚类, 进而可以根据聚类后所得到的语音特征簇进行高频用户发现和用户身份自动注册

(具体可以参见上述图 4 对应实施例中的 S401-S407)。如图 5 所示, 语音交互设备所得到的已绑定用户身份信息的用户语音模型可以包括用户语音模型 a、用户语音模型 b、用户语音模型 c。语音交互设备还可以定时或定量的根据已绑定用户身份信息的用户语音模型对应的语音特征簇 (如图 5 中的语音特征簇 a、语音特征簇 b、语音特征簇 c), 训练一个 DBSCAN 聚类模型, 训练目标为最大化 JC; 为避免训练过拟合, 在训练过程中确定该聚类模型在验证集上的 JC, 选择使验证集 JC 值最大的 Eps 和 MinPts 参数作为更新后的聚类模型参数, 即更新后的 Eps 和 MinPts (具体可以参见上述图 4 对应实施例中的 S408)。进而语音交互设备在下一次聚类时, 可以根据更新后的 Eps 和 MinPts 对历史语音数据 (包含新增的历史语音数据) 对应的 i-vector 进行 DBSCAN 聚类, 得到如图 5 所示的语音特征簇 1、语音特征簇 2、.....、语音特征簇 n, 并根据语音特征簇 1、语音特征簇 2、.....、语音特征簇 n, 对未绑定所述用户身份信息的用户语音模型所对应的语音特征簇进行更新, 并对未满足所述高频用户条件的语音特征簇进行替换 (具体可以参见上述图 4 对应实施例中的 S409)。此后, 可以定时或定量根据已绑定用户身份信息的用户语音模型所对应的语音特征簇对 Eps 和 MinPts 进行更新, 而且随着已绑定用户身份信息的用户语音模型的增加, 可以逐步训练出更准确、合理的 Eps 和 MinPts。其中, 初始化的 Eps 和 MinPts 仅用于第一次聚类使用, 此后的每一次聚类都使用最近一次更新后的 Eps 和 MinPts。

以语音交互设备为智能音箱为例, 本申请实施例对上述方案作了技术可行性验证。智能音箱通常不归属于某个特定的用户, 由多个用户共同使用, 但是其用户规模又十分有限。比如在家庭中使用的音箱设备, 用户数目通常不超过 10 人; 且在家庭中的成员, 由于年龄、性别等方面的差异, 其声纹特征的区别性比较明显。

首先, 利用大规模的数据集合, 随机从 600 人中不重复地抽取 10

人作为一组，每人提供 10 句内容完全一样的唤醒词作为语音样本。本申请实施例组织了两组实验，分别用于验证上述的聚类方法的可行性以及高频用户发现的可行性。

其中，聚类方法的可行性验证过程可以为：随机生成 10 组数据（每一组数据包括不重复的 10 个人分别提供的 10 句内容完全一样的语音样本）作为训练集，每组中随机选取 7 人的语音数据用于训练模型参数(Eps, MinPts)，训练目标为最大化 JC，其余 3 人的语音数据用于验证以减轻模型过拟合；随机生成 10 组数据作为测试集，测试训练得到的聚类模型的性能，具体基于 JC 和 RI 衡量聚类模型的性能。请一并参见图 6，
是本申请实施例提供的一种性能验证结果的示意图，如图 6 所示，10 组测试集（如图 6 中的 group1-group10）中的 JC 和 RI 都较高，即表明聚类模型具有较高的性能，因此，本申请实施例中的聚类方法具有可行性。

其中，高频用户发现的可行性验证过程可以为：首先，获取上述聚类方法的可行性验证过程中的 10 组测试集，对于每组测试集，在聚类及高频用户发现完成后，将发现到的高频用户所在的语音特征簇的类别设定为所在语音特征簇中出现次数最多的语音样本的类别。此时，对于每一组测试集，均可以确定该测试集中每一个发现的满足高频用户条件的语音特征簇的查准率（Precision）和查全率（Recall），以所有满足高频用户条件的语音特征簇的查准率和查全率的均值来表示高频用户发现算法在该测试集上的性能；其中，查准率和查全率越高，表明所发现的高频簇越精确。请一并参见图 7，是本申请实施例提供的另一种性能验证结果的示意图，如图 7 所示，10 个测试集（如图 7 中的 group1-group10）中的 Precision 和 Recall 都较高，即表明高频用户发现算法具有较高的性能，因此，本申请实施例中的高频用户发现具有可行性。

请参见图 8，是本申请实施例提供的一种语音数据处理装置的结构示意图。如图 8 所示，该语音数据处理装置 1 可以应用于上述图 3 或图

4 对应实施例中的语音交互设备，该语音数据处理装置 1 可以包括：聚类模块 10、第一训练模块 20、请求发起模块 30、绑定模块 40；

聚类模块 10，获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇；所述语音特征簇包含至少一个特征相似的历史语音特征向量；

第一训练模块 20，用于若所述语音特征簇满足高频用户条件，则根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型；

其中，所述第一训练模块 20 可以具体用于对所述语音特征簇所包含的所述历史语音特征向量进行均值计算或插值计算，得到目标历史语音特征向量，并将所述目标历史语音特征向量作为所述语音特征簇对应的用户语音模型的模型参数。

请求发起模块 30，用于若检测到当前语音数据的当前语音特征向量与所述用户语音模型相匹配，则发起与所述当前语音数据相关联的用户身份关联请求；

绑定模块 40，用于若接收到与所述用户身份关联请求对应的响应消息，则将所述响应消息中的用户身份信息与所述用户语音模型进行绑定。

其中，聚类模块 10、第一训练模块 20、请求发起模块 30、绑定模块 40 的具体功能实现方式可以参见上述图 3 对应实施例中的 S301-S304，这里不再进行赘述。

如图 8 所示，聚类模块 10 可以包括：获取训练单元 101、向量处理单元 102、聚类单元 103；

获取训练单元 101，用于获取所有历史语音数据，并根据所述所有历史语音数据训练高斯混合模型和全局差异空间矩阵；

向量处理单元 102，用于根据所述高斯混合模型和所述全局差异空间矩阵将所述所有历史语音数据投影至向量空间，生成每个历史语音数

据分别对应的历史语音特征向量，并对所述历史语音特征向量进行降维；

聚类单元 103，用于根据目标聚类模型参数对降维后的历史语音特征向量进行聚类，得到所述语音特征簇。

其中，所述目标聚类模型参数包括：密度领域半径和核心样本阈值。

5 其中，获取训练单元 101、向量处理单元 102、聚类单元 103 的具体功能实现方式可以参见上述图 4 对应实施例中的 S401-S403，这里不再进行赘述。

进一步的，如图 8 所示，所述聚类单元 103 可以包括：查找子单元 1031、聚类子单元 1032、通知子单元 1033；

10 查找子单元 1031，用于以所有降维后的历史语音特征向量为样本点生成包含所述样本点的样本数据集，并根据所述密度领域半径和所述核心样本阈值在所述样本数据集中查找所有为核心点的样本点；

聚类子单元 1032，用于在所有核心点中确定任意一个核心点为出发点，并在所述样本数据集中查找与所述出发点具有密度可达关系的所有
15 样本点，作为可达样本点，并生成包含所述出发点和所有所述可达样本点的语音特征簇；

通知子单元 1033，用于将所有核心点中的下一个核心点确定为所述出发点，并通知所述聚类子单元 1032 生成所述出发点对应的所述语音特征簇，直至所有核心点均被确定为所述出发点。

20 其中，查找子单元 1031、聚类子单元 1032、通知子单元 1033 的具体功能实现方式可以参见上述图 4 对应实施例中的 S403，这里不再进行赘述。

如图 8 所示，该语音数据处理装置 1 还可以包括：获取计算模块 50、条件确定模块 60、样本设置模块 70、第二训练模块 80、第一更新模块
25 90、第二更新模块 100、第三更新模块 110；

获取计算模块 50，用于获取所述语音特征簇所包含的所述历史语音

特征向量的数量，并根据所述语音特征簇所包含的所述历史语音特征向量的数量，以及所述语音特征簇所包含的所述历史语音特征向量，确定所述语音特征簇对应的类内散度；

条件确定模块 60，用于若所述语音特征簇所包含的所述历史语音特征向量的数量大于系统数量阈值，且所述类内散度小于系统类内散度阈值，则确定所述语音特征簇满足高频用户条件。

其中，获取计算模块 50、条件确定模块 60 的具体功能实现方式可以参见上述图 4 对应实施例中的 S404-S405，这里不再进行赘述。

样本设置模块 70，用于获取样本语音数据，并为所述样本语音数据设置对应的样本用户身份标签；

第二训练模块 80，用于根据聚类算法性能参数最大化条件、所述样本语音数据与所述样本用户身份标签之间的对应关系，训练初始聚类模型参数，并将训练后的初始聚类模型参数确定为所述目标聚类模型参数。

其中，样本设置模块 70、第二训练模块 80 的具体功能实现方式可以参见上述图 4 对应实施例中对初始化聚类模型参数的过程，这里不再进行赘述。

第一更新模块 90，用于若聚类后所累计新增的历史语音数据的数量达到第一数量阈值，或聚类后所累计时长达到第一时长阈值，则获取与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第一历史语音特征向量，并根据聚类算法性能参数最大化条件、所述第一历史语音特征向量与已绑定的所述用户身份信息之间的对应关系，更新当前的聚类模型参数，得到所述目标聚类模型参数。

其中，第一更新模块 90 的具体功能实现方式可以参见上述图 4 对应实施例中的 S408，这里不再进行赘述。

第二更新模块 100，用于若聚类后所累计新增的历史语音数据的数量达到第二数量阈值，或聚类后所累计时长达到第二时长阈值，则获取

与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，以及与所有用户语音模型均不匹配的历史语音特征向量，作为第二历史语音特征向量，并对所述第二历史语音特征向量进行聚类，得到当前生成的语音特征簇，并根据所述当前生成的语音特征簇对未绑定
5 所述用户身份信息的用户语音模型所对应的语音特征簇进行更新，并对未满足所述高频用户条件的语音特征簇进行替换。

其中，第二更新模块 100 的具体功能实现方式可以参见上述图 4 对应实施例中的 S409，这里不再进行赘述。

第三更新模块 110，用于若聚类后所累计新增的历史语音数据的数量达到第三数量阈值，或聚类后所累计时长达到第三时长阈值，则获取
10 与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第三历史语音特征向量，并根据所述第三历史语音特征向量更新已绑定所述用户身份信息的用户语音模型；

所述第三更新模块 110，还用于获取与未绑定所述用户身份信息
15 的用户语音模型相匹配的所有历史语音特征向量，作为第四历史语音特征向量，并根据所述第四历史语音特征向量更新未绑定所述用户身份信息的用户语音模型。

其中，第三更新模块 110 的具体功能实现方式可以参见上述图 4 对应实施例中的 S410，这里不再进行赘述。

20 请参见图 9，是本申请实施例提供的一种语音交互设备的结构示意图。如图 9 所示，所述语音交互设备 1000 可以为上述图 3 或图 4 对应实施例中的语音交互设备，所述语音交互设备 1000 可以包括：处理器 1001，网络接口 1004 和存储器 1005，此外，所述语音交互设备 1000 还可以包括：用户接口 1003，和至少一个通信总线 1002。其中，通信总线 1002 用于实现这些组件之间的连接通信。其中，用户接口 1003 可以
25 包括显示屏（Display）、键盘（Keyboard），可选用户接口 1003 还可以

包括标准的有线接口、无线接口。网络接口 1004 可以包括标准的有线接口、无线接口（如 WI-FI 接口）。存储器 1005 可以是高速 RAM 存储器，也可以是非不稳定的存储器（non-volatile memory），例如至少一个磁盘存储器。存储器 1005 还可以是至少一个位于远离前述处理器 1001 的存储装置。如图 9 所示，作为一种计算机存储介质的存储器 1005 中可以包括操作系统、网络通信模块、用户接口模块以及设备控制应用程序。

在图 9 所示的语音交互设备 1000 中，网络接口 1004 可提供网络通讯功能；而用户接口 1003 主要用于为用户提供输入的接口；而处理器 1001 可以用于调用存储器 1005 中存储的设备控制应用程序，以实现图 3 至图 4 所述的语音数据处理方法。确定确定确定

应当理解，本申请实施例中所描述的语音交互设备 1000 可执行前文图 3 到图 4 所对应实施例的所述语音数据处理方法，也可包括前文图 8 对应实施例中的所述语音数据处理装置 1，在此不再赘述。

此外，这里需要指出的是：本申请实施例还提供了一种计算机存储介质，且所述计算机存储介质中存储有前文提及的语音数据处理装置 1 所执行的计算机程序，且所述计算机程序包括程序指令，当所述处理器执行所述程序指令时，能够执行前文图 3 到图 4 所对应实施例中的所述语音数据处理方法，因此，这里将不再进行赘述。另外，对采用相同方法的有益效果描述，也不再进行赘述。对于本申请所涉及的计算机存储介质实施例中未披露的技术细节，请参照本申请方法实施例的描述。

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，所述的程序可存储于一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，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）或随机存储记忆体（Random

Access Memory, RAM) 等。

以上所揭露的仅为本申请较佳实施例而已，当然不能以此来限定本申请之权利范围，因此依本申请权利要求所作的等同变化，仍属本申请所涵盖的范围。

权利要求书

1、一种语音数据处理方法，由语音交互设备执行，包括：

获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇；所述语音特征簇包含至少一个特征相似的历史语音特征向量；

若所述语音特征簇满足高频用户条件，则根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型；

若检测到当前语音数据的当前语音特征向量与所述用户语音模型相匹配，则发起与所述当前语音数据相关联的用户身份关联请求；

若接收到与所述用户身份关联请求对应的响应消息，则将所述响应消息中的用户身份信息与所述用户语音模型进行绑定。

2、如权利要求 1 所述的方法，还包括：

获取所述语音特征簇所包含的所述历史语音特征向量的数量，并根据所述语音特征簇所包含的所述历史语音特征向量的数量，以及所述语音特征簇所包含的所述历史语音特征向量，确定所述语音特征簇对应的类内散度；

若所述语音特征簇所包含的所述历史语音特征向量的数量大于系统数量阈值，且所述类内散度小于系统类内散度阈值，则确定所述语音特征簇满足高频用户条件。

3、如权利要求 1 所述的方法，所述获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇，包括：

获取所有历史语音数据，并根据所述所有历史语音数据训练高斯混合模型和全局差异空间矩阵；

根据所述高斯混合模型和所述全局差异空间矩阵将所述所有历史语音数据投影至向量空间，生成每个历史语音数据分别对应的历史语音特征向量，并对所述历史语音特征向量进行降维；

根据目标聚类模型参数对降维后的历史语音特征向量进行聚类，得到所述语音特征簇。

4、如权利要求 3 所述的方法，所述目标聚类模型参数包括：密度领域半径和核心样本阈值；

- 5 所述根据目标聚类模型参数对降维后的历史语音特征向量进行聚类，得到所述语音特征簇，包括：

以所有降维后的历史语音特征向量为样本点生成包含所述样本点的样本数据集，并根据所述密度领域半径和所述核心样本阈值在所述样本数据集中查找所有为核心点的样本点；

- 10 在所有核心点中确定任意一个核心点为出发点，并在所述样本数据集中查找与所述出发点具有密度可达关系的所有样本点，作为可达样本点，并生成包含所述出发点和所有所述可达样本点的语音特征簇，并将所有核心点中的下一个核心点确定为所述出发点，重复执行本步骤，直至所有核心点均被确定为所述出发点。

- 15 5、如权利要求 1 所述的方法，所述根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型，具体包括：

对所述语音特征簇所包含的所述历史语音特征向量进行均值确定或插值确定，得到目标历史语音特征向量，并将所述目标历史语音特征向量作为所述语音特征簇对应的用户语音模型的模型参数。

- 20 6、如权利要求 3 所述的方法，还包括：

获取样本语音数据，并为所述样本语音数据设置对应的样本用户身份标签；

- 25 根据聚类算法性能参数最大化条件、所述样本语音数据与所述样本用户身份标签之间的对应关系，训练初始聚类模型参数，并将训练后的初始聚类模型参数确定为所述目标聚类模型参数。

- 7、如权利要求 3 所述的方法，还包括：

若聚类后所累计新增的历史语音数据的数量达到第一数量阈值，或聚类后所累计时长达到第一时长阈值，则获取与已绑定所述用户身份信

息的用户语音模型相匹配的所有历史语音特征向量，作为第一历史语音特征向量；

- 根据聚类算法性能参数最大化条件、所述第一历史语音特征向量与已绑定的所述用户身份信息之间的对应关系，更新当前的聚类模型参数，
- 5 得到所述目标聚类模型参数。

8、如权利要求 3 所述的方法，还包括：

- 若聚类后所累计新增的历史语音数据的数量达到第二数量阈值，或聚类后所累计时长达到第二时长阈值，则获取与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，以及与所有用户语音模型均不匹配的历史语音特征向量，作为第二历史语音特征向量，并
- 10 对所述第二历史语音特征向量进行聚类，得到当前生成的语音特征簇；

根据所述当前生成的语音特征簇对未绑定所述用户身份信息的用户语音模型所对应的语音特征簇进行更新，并对未满足所述高频用户条件的语音特征簇进行替换。

- 15 9、如权利要求 3 所述的方法，还包括：

- 若聚类后所累计新增的历史语音数据的数量达到第三数量阈值，或聚类后所累计时长达到第三时长阈值，则获取与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第三历史语音特征向量，并根据所述第三历史语音特征向量更新已绑定所述用户身份
- 20 信息的用户语音模型；

获取与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第四历史语音特征向量，并根据所述第四历史语音特征向量更新未绑定所述用户身份信息的用户语音模型。

10、一种语音交互设备，包括：处理器、存储器；

- 25 所述处理器与存储器相连，其中，所述存储器用于存储程序代码，所述处理器用于调用所述程序代码，以执行以下操作：

获取历史语音数据，并获取所述历史语音数据对应的历史语音特征向量，并对所述历史语音特征向量进行聚类，得到语音特征簇；所述语

音特征簇包含至少一个特征相似的历史语音特征向量；

若所述语音特征簇满足高频用户条件，则根据所述语音特征簇所包含的所述历史语音特征向量训练对应的用户语音模型；

若检测到当前语音数据的当前语音特征向量与所述用户语音模型
5 相匹配，则发起与所述当前语音数据相关联的用户身份关联请求；

若接收到与所述用户身份关联请求对应的响应消息，则将所述响应消息中的用户身份信息与所述用户语音模型进行绑定。

11、如权利要求 10 所述的语音交互设备，所述处理器进一步用于执行以下操作：

10 获取所述语音特征簇所包含的所述历史语音特征向量的数量，并根据所述语音特征簇所包含的所述历史语音特征向量的数量，以及所述语音特征簇所包含的所述历史语音特征向量，确定所述语音特征簇对应的类内散度；

若所述语音特征簇所包含的所述历史语音特征向量的数量大于系统数量阈值，且所述类内散度小于系统类内散度阈值，则确定所述语音
15 特征簇满足高频用户条件。

12、如权利要求 10 所述的语音交互设备，所述处理器进一步用于执行以下操作：

获取所有历史语音数据，并根据所述所有历史语音数据训练高斯混
20 合模型和全局差异空间矩阵；

根据所述高斯混合模型和所述全局差异空间矩阵将所述所有历史语音数据投影至向量空间，生成每个历史语音数据分别对应的历史语音特征向量，并对所述历史语音特征向量进行降维；

根据目标聚类模型参数对降维后的历史语音特征向量进行聚类，得
25 到所述语音特征簇。

13、如权利要求 12 所述的语音交互设备，所述目标聚类模型参数包括：密度领域半径和核心样本阈值；

所述处理器进一步用于执行以下操作：

以所有降维后的历史语音特征向量为样本点生成包含所述样本点的样本数据集，并根据所述密度领域半径和所述核心样本阈值在所述样本数据集中查找所有为核心点的样本点；

- 5 在所有核心点中确定任意一个核心点为出发点，并在所述样本数据集中查找与所述出发点具有密度可达关系的所有样本点，作为可达样本点，并生成包含所述出发点和所有所述可达样本点的语音特征簇；

将所有核心点中的下一个核心点确定为所述出发点，并通知所述聚类子单元生成所述出发点对应的所述语音特征簇，直至所有核心点均被确定为所述出发点。

- 10 14、如权利要求 10 所述的语音交互设备，所述处理器进一步用于执行以下操作：

对所述语音特征簇所包含的所述历史语音特征向量进行均值计算或插值计算，得到目标历史语音特征向量，并将所述目标历史语音特征向量作为所述语音特征簇对应的用户语音模型的模型参数。

- 15 15、如权利要求 12 所述的语音交互设备，所述处理器进一步用于执行以下操作：

获取样本语音数据，并为所述样本语音数据设置对应的样本用户身份标签；

- 20 根据聚类算法性能参数最大化条件、所述样本语音数据与所述样本用户身份标签之间的对应关系，训练初始聚类模型参数，并将训练后的初始聚类模型参数确定为所述目标聚类模型参数。

16、如权利要求 12 所述的语音交互设备，所述处理器进一步用于执行以下操作：

- 25 若聚类后所累计新增的历史语音数据的数量达到第一数量阈值，或聚类后所累计时长达到第一时长阈值，则获取与已绑定所述用户身份信息的所有历史语音特征向量，作为第一历史语音特征向量；

根据聚类算法性能参数最大化条件、所述第一历史语音特征向量与

已绑定的所述用户身份信息之间的对应关系，更新当前的聚类模型参数，得到所述目标聚类模型参数。

17、如权利要求 12 所述的语音交互设备，所述处理器进一步用于执行以下操作：

- 5 若聚类后所累计新增的历史语音数据的数量达到第二数量阈值，或聚类后所累计时长达到第二时长阈值，则获取与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，以及与所有用户语音模型均不匹配的历史语音特征向量，作为第二历史语音特征向量，并对所述第二历史语音特征向量进行聚类，得到当前生成的语音特征簇；
- 10 根据所述当前生成的语音特征簇对未绑定所述用户身份信息的用户语音模型所对应的语音特征簇进行更新，并对未满足所述高频用户条件的语音特征簇进行替换。

18、如权利要求 12 所述的语音交互设备，所述处理器进一步用于执行以下操作：

- 15 若聚类后所累计新增的历史语音数据的数量达到第三数量阈值，或聚类后所累计时长达到第三时长阈值，则获取与已绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第三历史语音特征向量，并根据所述第三历史语音特征向量更新已绑定所述用户身份信息的用户语音模型；
- 20 获取与未绑定所述用户身份信息的用户语音模型相匹配的所有历史语音特征向量，作为第四历史语音特征向量，并根据所述第四历史语音特征向量更新未绑定所述用户身份信息的用户语音模型。

- 19、一种计算机存储介质，所述计算机存储介质存储有计算机程序，所述计算机程序包括程序指令，当处理器执行所述程序指令时执行如权
- 25 利要求 1-9 任一项所述的方法。

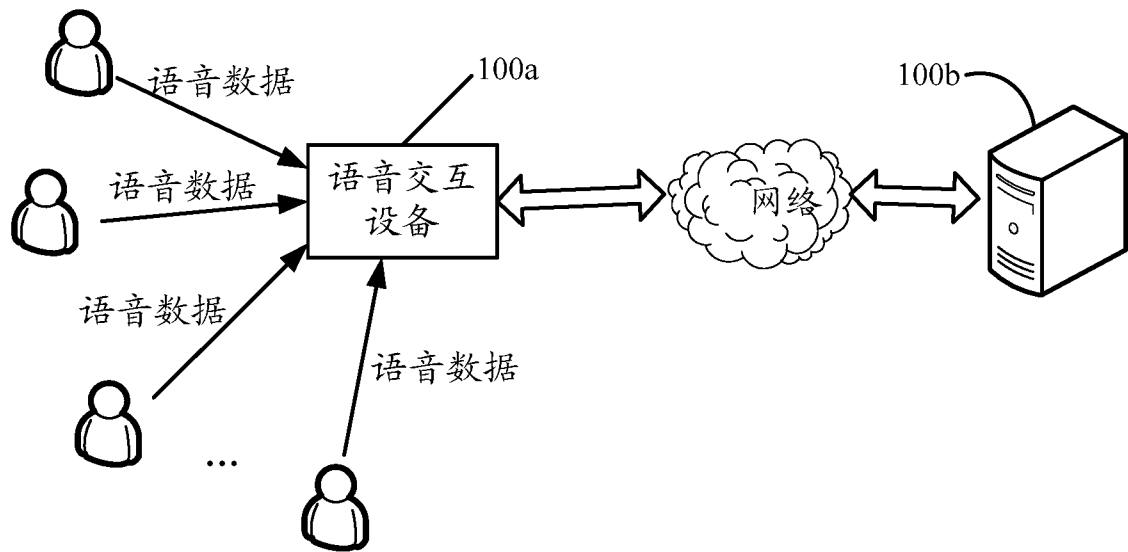


图 1

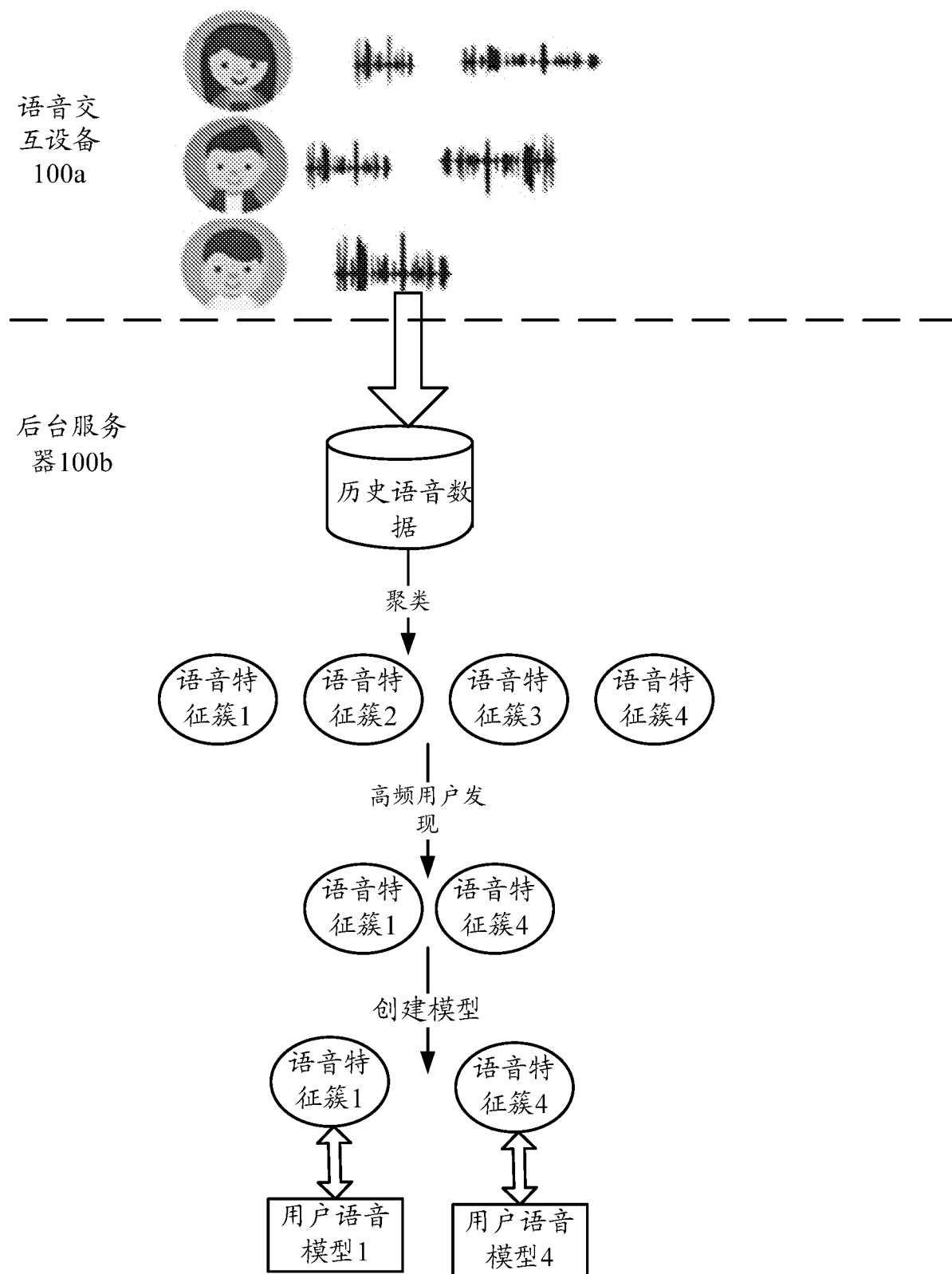


图 2a

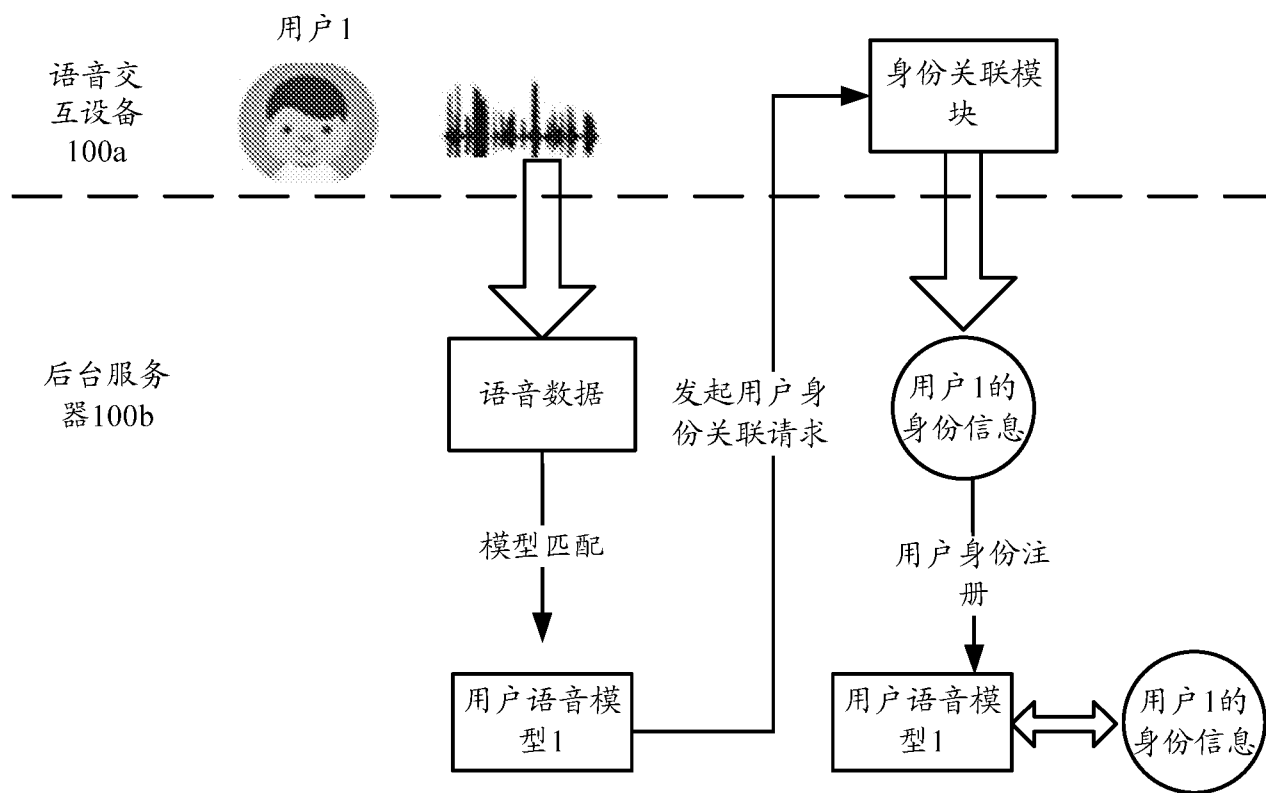


图 2b

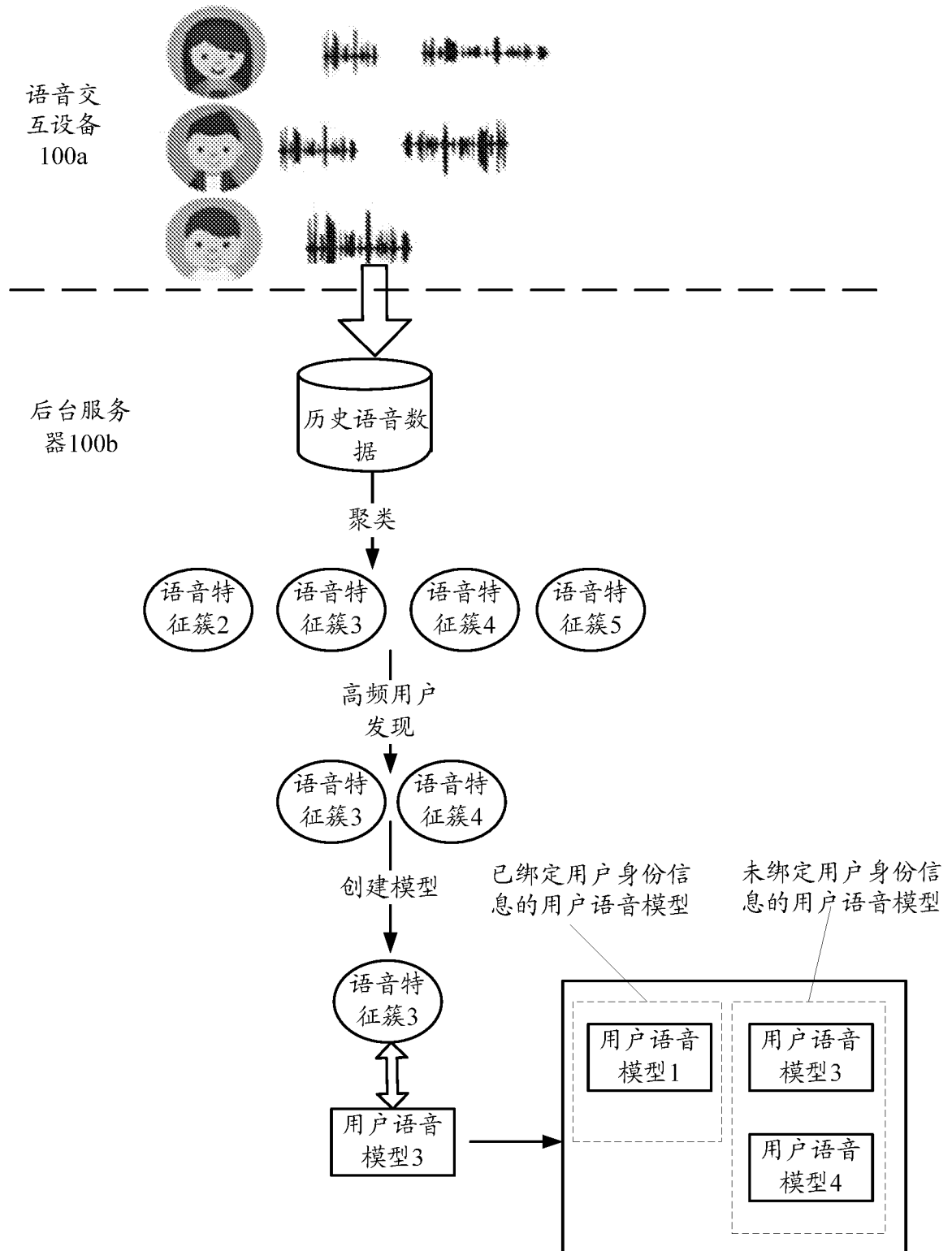


图 2c

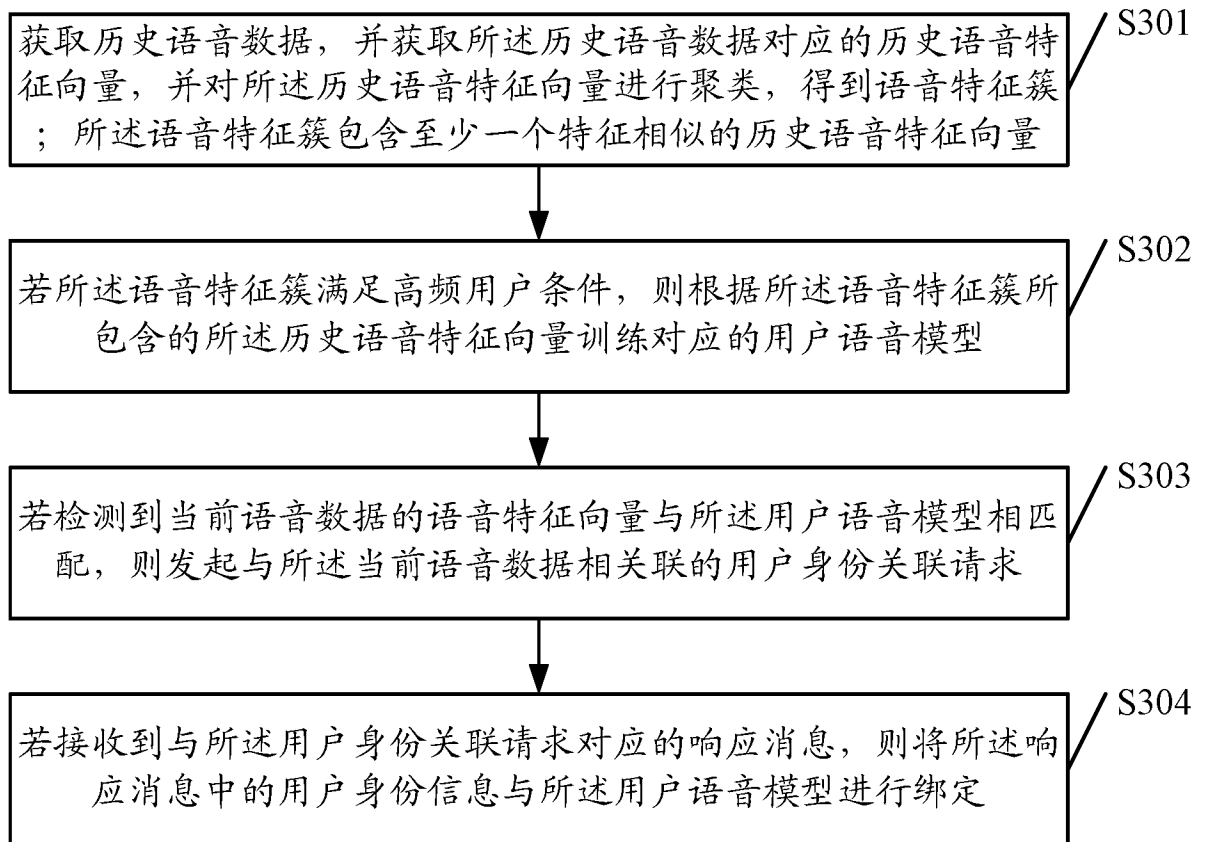


图 3

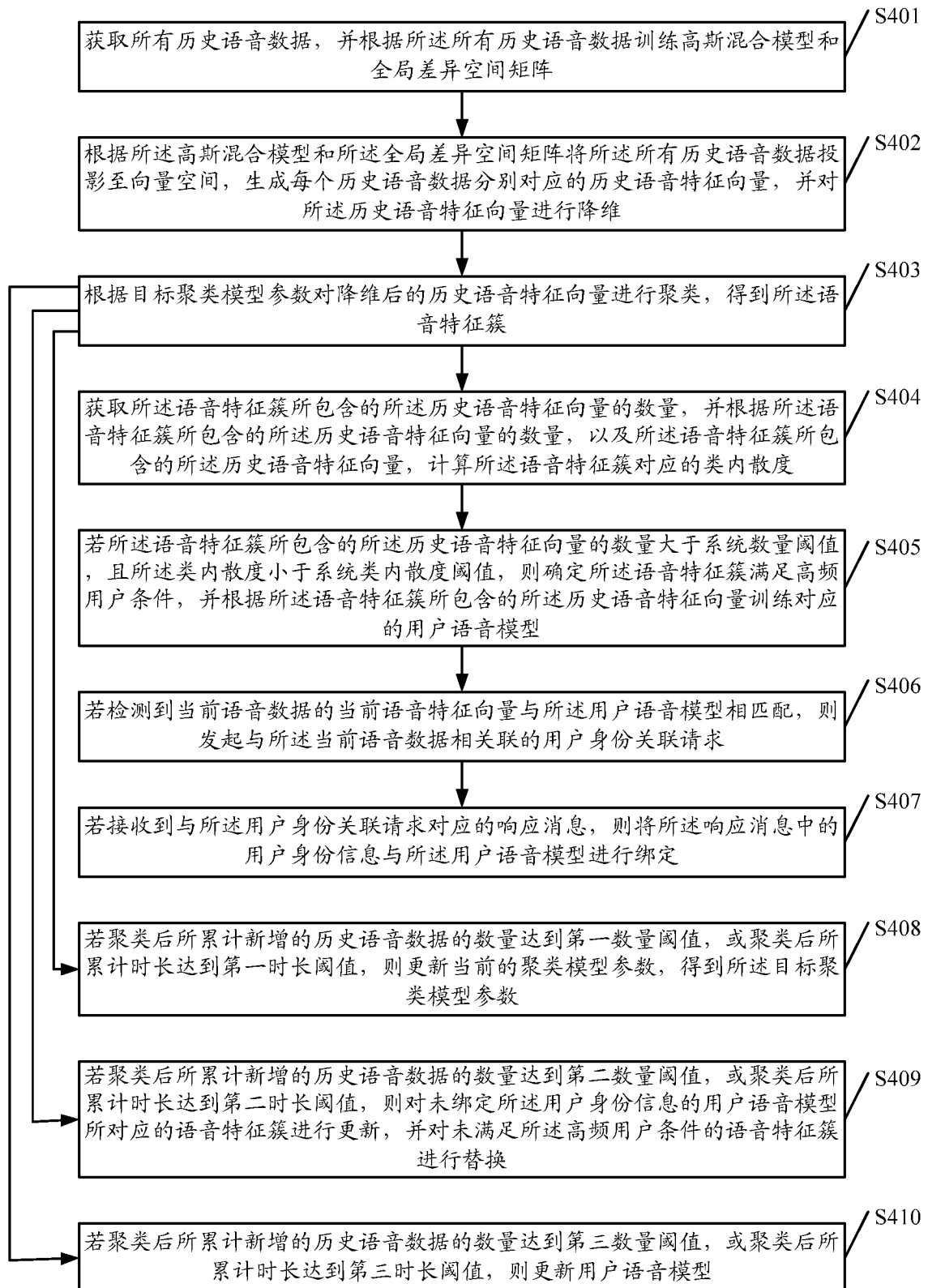


图 4

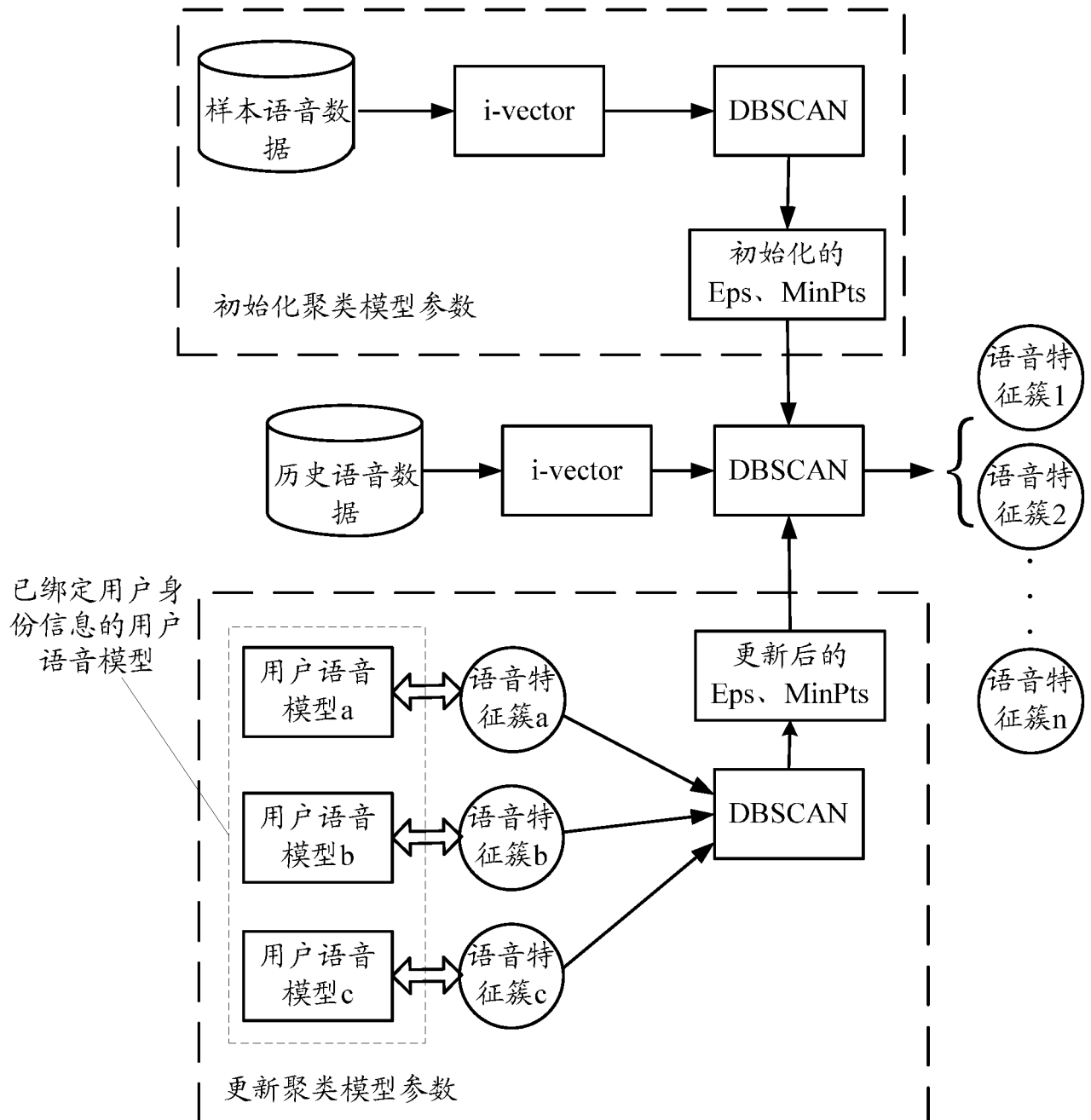


图 5

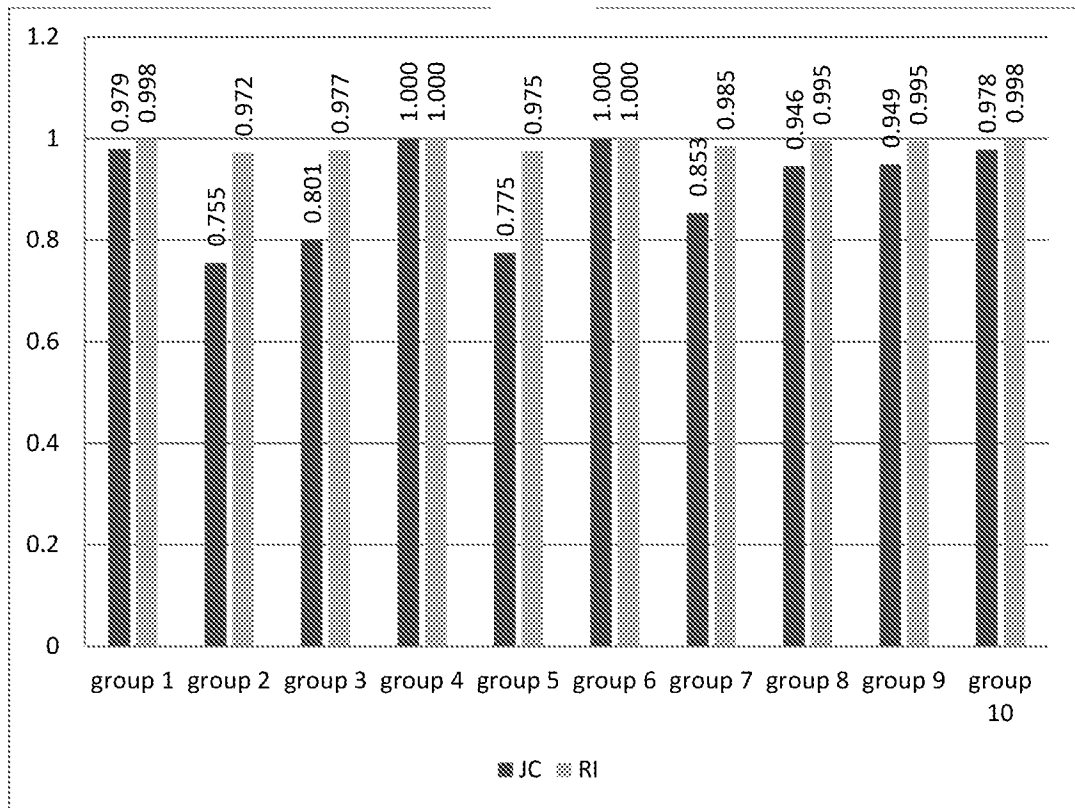


图 6

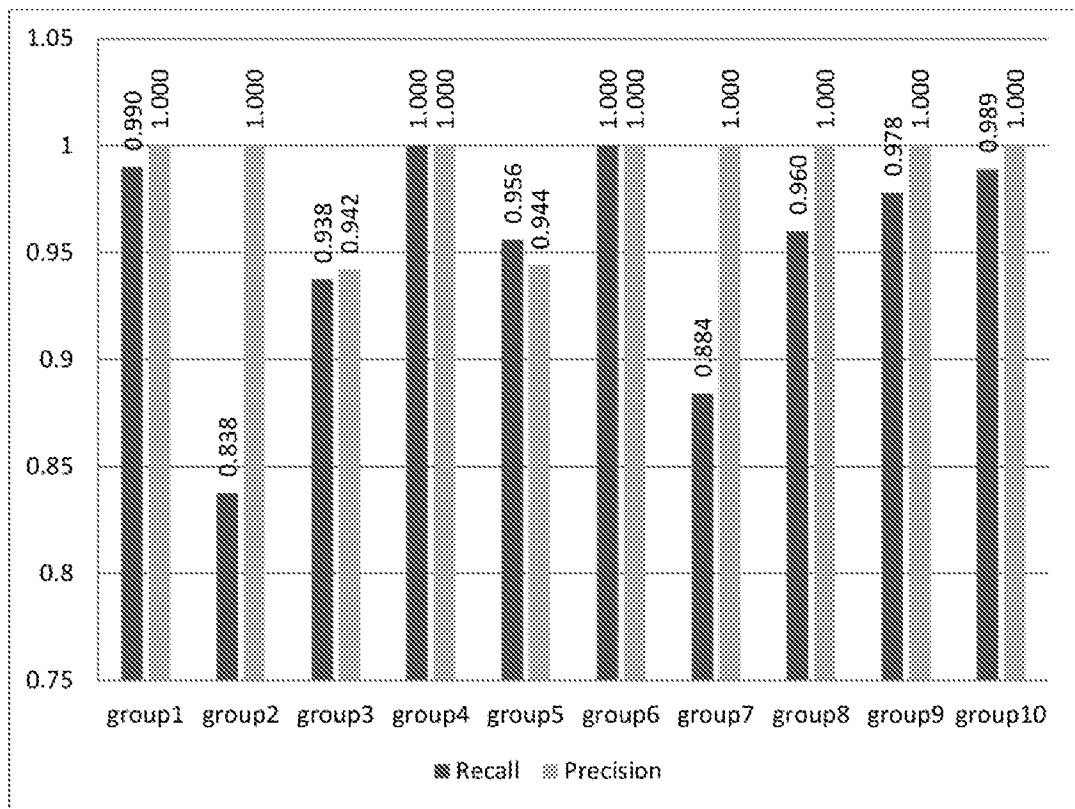


图 7

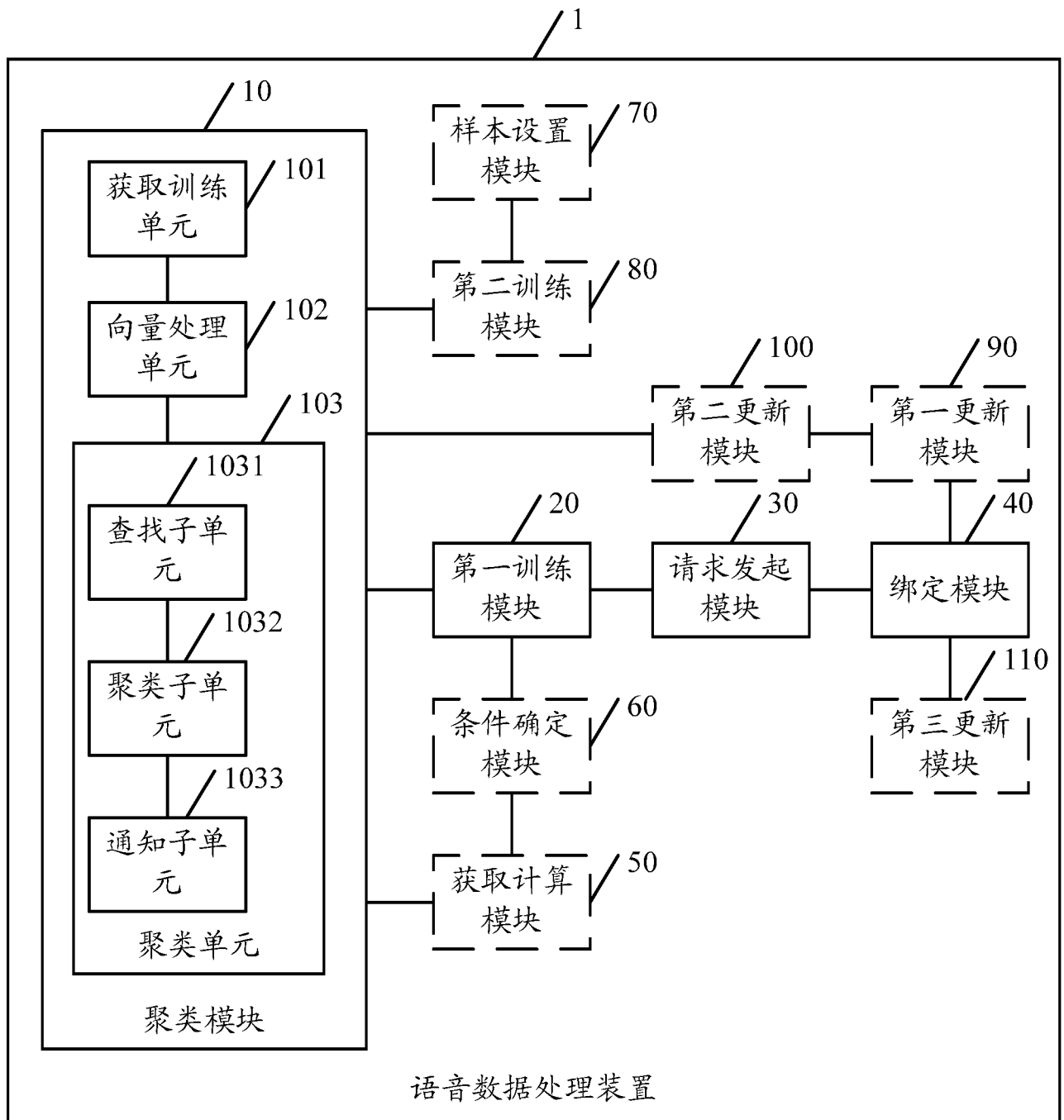


图 8

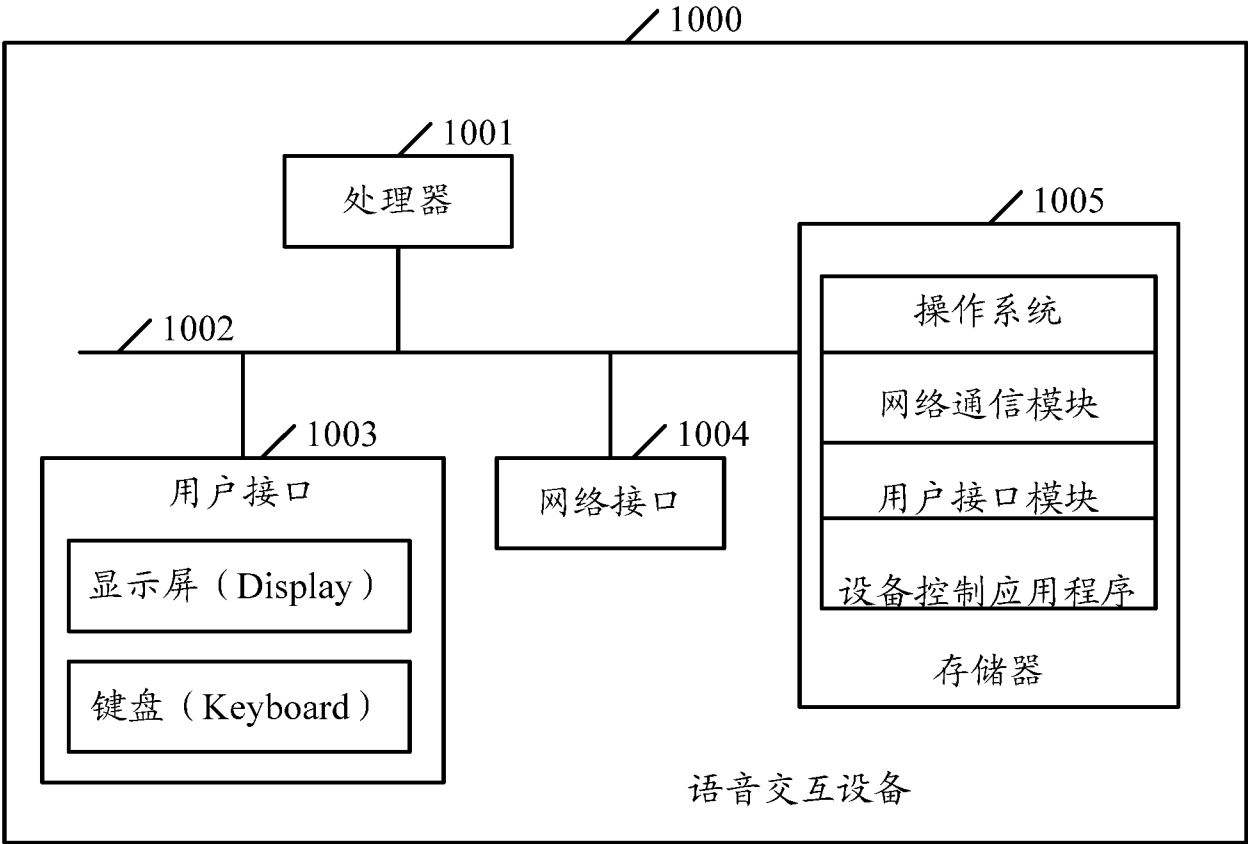


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/116590

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02(2006.01)i; G10L 15/06(2013.01)n; G10L 15/22(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

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

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

CNABS; CNTXT; VEN; USTXT; WOTXT; EPTXT; CNKI: 语音, 特征, 向量, 聚类, 特征簇, 高频, 身份, 绑定, 注册, 内散度, 阈值, 降维, voice, speech, character, cluster, high, frequency, identity, login, threshold, bind, dimensionality, reduction

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107978311 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 01 May 2018 (2018-05-01) entire document	1-19
A	CN 103337241 A (BEIJING UNISOUND INFORMATION TECHNOLOGY CO., LTD.) 02 October 2013 (2013-10-02) entire document	1-19
A	CN 103903621 A (LENOVO (BEIJING) LTD.) 02 July 2014 (2014-07-02) entire document	1-19
A	CN 105869645 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 17 August 2016 (2016-08-17) entire document	1-19
A	EP 3057092 A1 (HAND HELD PROD INC.) 17 August 2016 (2016-08-17) entire document	1-19



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

17 January 2018

Date of mailing of the international search report

19 February 2019

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

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/116590

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107978311	A	01 May 2018	None			
CN	103337241	A	02 October 2013	CN	103337241	B	24 June 2015
CN	103903621	A	02 July 2014	None			
CN	105869645	A	17 August 2016	WO	2017162017	A1	28 September 2017
				IN	201847019837	A	01 June 2018
				US	2018286410	A1	04 October 2018
EP	3057092	A1	17 August 2016	EP	3057092	B1	26 July 2017
				US	2016232891	A1	11 August 2016
				US	10121466	B2	06 November 2018

国际检索报告

国际申请号

PCT/CN2018/116590

A. 主题的分类

G10L 15/02(2006.01)i; G10L 15/06(2013.01)n; G10L 15/22(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;VEN;USTXT;WOTXT;EPTXT;CNKI: 语音, 特征, 向量, 聚类, 特征簇, 高频, 身份, 绑定, 注册, 内散度, 阈值, 降维, voice, speech, character, cluster, high, frequency, identity, login, threshold, bind, dimensionality, reduction

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 107978311 A (腾讯科技深圳有限公司) 2018年 5月 1日 (2018 - 05 - 01) 全文	1-19
A	CN 103337241 A (北京云知声信息技术有限公司) 2013年 10月 2日 (2013 - 10 - 02) 全文	1-19
A	CN 103903621 A (联想北京有限公司) 2014年 7月 2日 (2014 - 07 - 02) 全文	1-19
A	CN 105869645 A (腾讯科技深圳有限公司) 2016年 8月 17日 (2016 - 08 - 17) 全文	1-19
A	EP 3057092 A1 (HAND HELD PROD INC) 2016年 8月 17日 (2016 - 08 - 17) 全文	1-19

☐ 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

2018年 1月 17日

国际检索报告邮寄日期

2019年 2月 19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

刘剑锋

电话号码 86-(0512)-88996610

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/116590

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	107978311	A	2018年 5月 1日	无			
CN	103337241	A	2013年 10月 2日	CN	103337241	B	2015年 6月 24日
CN	103903621	A	2014年 7月 2日	无			
CN	105869645	A	2016年 8月 17日	WO	2017162017	A1	2017年 9月 28日
				IN	201847019837	A	2018年 6月 1日
				US	2018286410	A1	2018年 10月 4日
EP	3057092	A1	2016年 8月 17日	EP	3057092	B1	2017年 7月 26日
				US	2016232891	A1	2016年 8月 11日
				US	10121466	B2	2018年 11月 6日