

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 5 月 22 日 (22.05.2020)



(10) 国际公布号
WO 2020/098523 A1

(51) 国际专利分类号:
G10L 15/22 (2006.01) *G10L 17/22* (2013.01)

(21) 国际申请号: PCT/CN2019/115308

(22) 国际申请日: 2019 年 11 月 4 日 (04.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

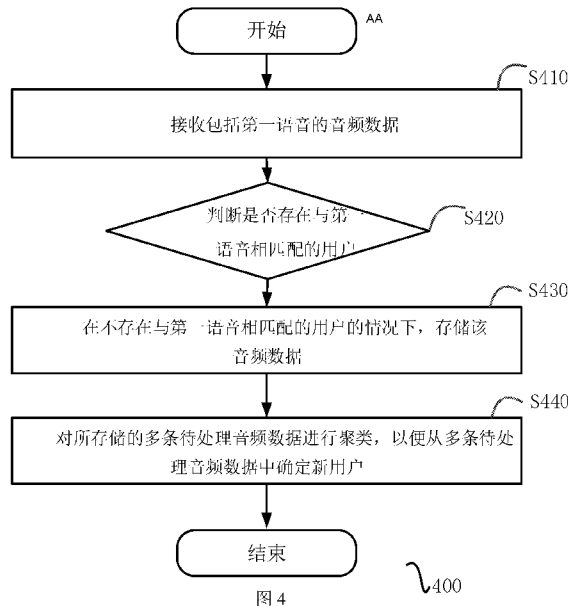
(30) 优先权:
201811340092.2 2018 年 11 月 12 日 (12.11.2018) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 赵情恩 (ZHAO, Qing'en); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 索宏彬 (SU, Hongbin); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 刘刚 (LIU, Gang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 卓著 (ZHUO, Zhu); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 雷赞 (LEI, Yun); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 张平 (ZHANG, Ping); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 孙尧 (SUN, Yao); 中国浙

(54) Title: VOICE RECOGNITION METHOD AND DEVICE AND COMPUTING DEVICE

(54) 发明名称: 一种语音识别方法、装置及计算设备



AA Start
S410 Receiving audio data including a first voice
S420 Determining whether there is a user matching the first voice
S430 If there is no user matching the first voice, storing the audio data
S440 Clustering multiple pieces of stored audio data so as to determine a new user therefrom
400 End

图 4

(57) Abstract: Disclosed is a voice recognition method, comprising the steps of: receiving audio data including a first voice (S410); determining whether there is a user matching the first voice (S420); if there is no user matching the first voice, storing the audio data (S430); and clustering multiple pieces of stored audio data so as to determine a new user therefrom (S440). Further disclosed are a corresponding voice recognition device and system as well as a computing device.

(57) 摘要: 一种语音识别方法, 包括步骤: 接收包括第一语音的音频数据 (S410); 判断是否存在与第一语音相匹配的用户 (S420); 在不存在与第一语音相匹配的用户的情况下, 存储该音频数据 (S430); 对所存储的多条音频数据进行聚类, 以便从多条音频数据中确定新用户 (S440)。还公开了相应的语音识别装置、系统和计算设备。

江省杭州市余杭区文一西路969号3号楼5楼阿里
巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北 京 三 友 知 识 产 权 代 理
有 限 公 司 (BEIJING SANYOU INTELLECTUAL
PROPERTY AGENCY LTD.); 中 国 北 京 市 金
融 街 35 号 国 际 企 业 大 厦 A 座 16 层,
Beijing 100033 (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,
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))。

一种语音识别方法、装置及计算设备

本申请要求 2018 年 11 月 12 日递交的申请号为 2018113400922、发明名称为“一种语音识别方法、装置及计算设备”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。

5

技术领域

本发明涉及语音识别技术领域，尤其是一种语音识别方法、装置及计算设备。

背景技术

10 随着诸如移动终端和智能音箱之类终端设备的广泛使用，人们越来越习惯于使用语音来与这些终端设备进行交互。其中，终端设备可以采用声纹识别技术来识别用户身份。

声纹识别（Voiceprint Identification），又称说话人识别（Speaker Identification），该技术是从说话人发出的语音信号中提取语音特征，并据此对说话人进行身份验证的生物识别技术。其中，声纹是指人类语音中携带言语信息的声波频谱。同指纹一样，声纹
15 具备独特的生物学特征，具有身份识别的作用，不仅具有特定性，而且具有相对的稳定性。

通常地，说话者需要预先在终端设备上注册声纹，而后终端设备通过声纹来识别该用户，从而可以通过分析该用户的语音对应的指令来分析用户行为，以便为该用户提供个性化、定制化服务，例如歌曲推荐等。

20 由于目前大部分终端设备的用户没有主动注册声纹，所以无法准确识别用户，从而无法分析用户行为向其提供个性化服务，或者向其提供的个性化服务很难达到比较好的效果。

因此，需要提供一种更优越的语音识别方案，以便为用户提供服务。

25 发明内容

为此，本发明实施例提供了一种语音识别方法、装置及计算设备，以力图解决或者至少缓解上面存在的至少一个问题。

根据本发明实施例的一个方面，提供了一种语音识别方法，包括步骤：接收包括第一语音的音频数据；判断是否存在与第一语音相匹配的用户；在不存在与第一语音相匹配的用户的情况下，存储该音频数据；对所存储的多条音频数据进行聚类，以便从多条
30

音频数据中确定新用户。

可选地，在根据本发明实施例的语音识别方法中，用户对应有用户简档，用户简档包括用户的声纹，判断是否存在与第一语音相匹配的用户的步骤包括：判断第一语音是否与用户的声纹相匹配，以判断是否存在与第一语音相匹配的用户。

5 可选地，在根据本发明实施例的语音识别方法中，对所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户的步骤包括：基于多条音频数据中两两之间的相似评分，将多条音频数据划分为多个集合；基于集合的样本密度和样本数量确定至少一个目标集合，目标集合对应于新用户；为目标集合对应的新用户创建用户简档，并使用目标集合中的至少部分音频数据来生成该新用户的声纹。

10 可选地，在根据本发明实施例的语音识别方法中，使用目标集合中的至少部分音频数据来生成该新用户的声纹的步骤包括：根据到目标集合的质心的距离来确定目标集合中用于生成新用户的声纹的音频数据。

可选地，在根据本发明实施例的语音识别方法中，用户简档包括指示用户是否为主动注册的用户标记，为目标集合对应的新用户创建用户简档的步骤包括：将为目标集合对应的新用户所创建的用户简档中的用户标识置为非主动注册；以及方法还包括步骤：
15 在存在与第一语音相匹配的用户且对应的用户标记指示用户为非主动注册的情况下，记录来自用户的音频数据条数。

可选地，在根据本发明实施例的语音识别方法中，还包括步骤：在记录来自用户的音频数据条数之后，判断音频数据条数是否在特定时间段内达到特定数量；若否，删除
20 该用户对应的用户简档。

可选地，在根据本发明实施例的语音识别方法中，用户简档还包括与用户相关联的终端设备的设备标识，该方法包括步骤：接收发送音频数据的终端设备的设备标识；基于设备标识判断是否存在与终端设备相关联的用户；如果不存在，则存储该音频数据。

可选地，在根据本发明实施例的语音识别方法中，还包括步骤：在存在与第一语音
25 相匹配的用户的情况下，将第一语音对应的指令与用户相关联地存储。

可选地，在根据本发明实施例的语音识别方法中，还包括步骤：接收包括第二语音的音频数据，第二语音用于主动注册新用户；为主动注册的新用户创建用户简档，并使用包括第二语音的音频数据来生成新用户的声纹；以及将为主动注册的新用户所创建的用户简档中的用户标识置为主动注册。

30 可选地，在根据本发明实施例的语音识别方法中，还包括步骤：接收发送包括第二

语音的音频数据的终端设备的设备标识；将设备标识与主动注册的新用户相关联地存储至对应的用户简档。

可选地，在根据本发明实施例的语音识别方法中，判断第一语音是否与用户的声纹相匹配的步骤包括：根据包括第一语音的音频数据，提取第一语音的语音特征；基于所述第一语音的语音特征得到第一语音与用户的声纹之间的相似评分；根据相似评分来确定第一语音是否与用户的声纹相匹配。

根据本发明实施例的另一方面，提供了一种用户识别方法，包括步骤：接收包括第一语音的音频数据；判断是否存在与第一语音相匹配的用户；在不存在与第一语音相匹配的用户的情况下，存储音频数据；对所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户，并对新用户进行行为分析。

根据本发明实施例的另一方面，提供了一种语音识别装置，包括：通信模块，适于接收包括第一语音的音频数据；语音识别模块，适于判断是否存在与第一语音相匹配的用户；在不存在与第一语音相匹配的用户的情况下，将该音频数据存储至音频存储模块；音频存储模块，适于存储音频数据；以及用户发现模块，适于对音频存储模块所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户。

根据本发明实施例的另一方面，提供了一种用户识别装置，包括：通信模块，适于接收包括第一语音的音频数据；语音识别模块，适于判断是否存在与第一语音相匹配的用户；在不存在与第一语音相匹配的用户的情况下，将该音频数据存储至音频存储模块；音频存储模块，适于存储音频数据；以及用户发现模块，适于对音频存储模块所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户，并对新用户进行行为分析。

根据本发明实施例的另一方面，提供了一种语音识别系统，包括终端设备和服务器，其中终端设备适于接收说话人的语音，并将包括语音的音频数据发送至服务器；服务器驻留有根据本发明的语音识别装置。

根据本发明实施例的又一方面，提供了一种计算设备，包括：至少一个处理器；和存储有程序指令的存储器，其中，程序指令被配置为适于由至少一个处理器执行，程序指令包括用于执行根据本发明的语音识别方法的指令。

根据本发明实施例的语音识别方案，通过对所存储的多条音频数据进行聚类来从中确定新用户，整个新用户确定过程用户是无感知的，省去了用户的主动注册操作，提高了用户的使用体验。

附图说明

为了实现上述以及相关目的，本文结合下面的描述和附图来描述某些说明性方面，这些方面指示了可以实践本文所公开的原理的各种方式，并且所有方面及其等效方面旨在落入所要求保护的主题的范围内。通过结合附图阅读下面的详细描述，本公开的上述以及其它目的、特征和优势将变得更加明显。遍及本公开，相同的附图标记通常指代相同的部件或元素。

图 1 示出了根据本发明一个实施例的语音识别系统 100 的示意图；

图 2 示出了根据本发明一个实施例的语音识别装置 200 的架构图；

图 3 示出了根据本发明一个实施例的计算设备 300 的示意图；以及

图 4 示出了根据本发明一个实施例的语音识别方法 400 的结构框图。

具体实施方式

下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

图 1 示出了根据本发明一个实施例的语音识别系统 100 的示意图。如图 1 所示，语音识别系统 100 包括终端设备 102 和服务器 106。

终端设备 102 是任何说话人语音的接收方。说话人可以使用语音经由终端设备 102 与服务器 106 交互。终端设备 102 可以通过诸如局域网（LAN）或者如因特网的广域网（WAN）的一个或多个网络 105 耦合至服务器 106 的计算设备。例如，终端设备 102 可以是桌面型计算设备、膝上型计算设备、平板型计算设备、移动电话计算设备、音箱计算设备、车辆的计算设备（例如，车载通信系统、车载娱乐系统、车载导航系统）、包括计算设备的可穿戴装置（例如，具有计算设备的手表、具有计算设备的眼镜）或者包括计算设备的家居装置（例如，具有计算设备的音箱、具有计算设备的电视、具有计算设备的洗衣机）。尽管说话人有可能会操作多个计算设备，但为简洁起见，本公开中的示例将针对说话人操作终端设备 102。

终端设备 102 可以操作一个或多个应用和/或组件，它们可以涉及向说话人提供通知

以及提供各种类型的信号。这些应用和/或组件可以包括但不限于麦克风 103、输出设备 104、诸如全球定位系统(“GPS”)组件(图 1 未示出)的位置坐标组件等等。在一些实施例中, 这些应用和/或组件中的一个或多个可以在由说话人操作的多个终端设备上运行。图 1 中并未示出的终端设备 102 的其他组件包括但不限于气压计、相机、光线传感器、存在传感器、温度计、健康传感器(例如, 心率监视器、血糖仪、血压计)、加速计、陀螺仪等等。

在一些实施方式中, 输出设备 104 可以包括扬声器(多个扬声器)、屏幕、触摸屏、一个或多个通知灯(例如, 发光二极管)、打印机等等中的一个或多个。在一些实施方式中, 输出设备 104 可以被用于基于响应于说话人语音而调用的一个或多个操作(诸如打开程序、播放歌曲、发送电子邮件或者文本消息、拍照等操作)提供输出。

终端设备 102 包括用于存储数据和软件应用的一个或多个存储器、用于访问数据和执行应用的一个或多个处理器以及促进通过网络的通信的其他组件。在一些实施方式中, 终端设备 102 可以被配置成例如使用麦克风 103 来感测一个或多个可听声音(例如, 说话人口述的语音), 并且可以基于所感测到的一个或多个可听声音(也被称作“音频输入”)而将音频数据提供至各种其他计算设备。那些其他计算设备(其示例将会在下文更详细地描述)可以基于音频数据而执行各种操作以识别匹配音频数据。在各种实施方式中, 音频数据可以包括: 一个或多个说话人口述语音的原始记录; 记录的压缩版本; 经由终端设备 102 的麦克风 103 所获得的音频输入的一个或多个特征的指示, 诸如音高、音调、音频和/或音量; 和/或经由麦克风 103 所获得的音频输入的转录等等。

在一些实施方式中, 终端设备 102 将包括说话人语音的音频数据发送至服务器 106。服务器 106 中驻留有语音识别装置 200。当然, 在另一些实施方式中, 终端设备 102 中也可以驻留有语音识别装置 200。也就是说, 直接在终端设备 102 上执行下述处理。

图 2 示出了根据本发明一个实施例的语音识别装置 200 的结构框图。如图 2 所示, 语音识别装置 200 包括通信模块 210、语音识别模块 220、音频存储模块 230 和用户发现模块 240。

通信模块 210 可以从终端设备 102 接收到包括第一语音的音频数据, 这里第一语音通常用于指示终端设备 102 执行操作。

语音识别模块 220 对该音频数据进行语音识别, 以得到第一语音对应的指令。而后语音识别模块 220 经由通信模块 210 向终端设备 102 返回对该指令的响应结果, 以便终

端设备 102 至少根据该响应结果执行相应操作。

例如，在一种实施方式中，终端设备 102 可以实现为具有计算设备的音箱。音箱接收到说话人所说的语音——“播放歌曲青花瓷”，将包括该语音的音频数据发送至服务器 106。服务器 106 向音箱返回相应的响应结果——《青花瓷》的音频文件。音箱根据
5 响应结果执行相应操作——播放该音频文件。

当然，对音频数据进行语音识别以得到指令的过程也可以在终端设备 102 上进行。即，终端设备 102 对音频数据进行语音识别，而后将音频数据和识别得到的指令发送至语音识别装置 200。

语音识别模块 220 还判断是否存在与第一语音相匹配的用户。通常地，用户指的是
10 语音识别系统标识其身份的说话人。根据一种实施方式，用户对应有记录与用户相关的数据的用户简档，这些用户简档可以存储在与语音识别装置 200 相耦接的用户数据存储设备中，也可以存储在语音识别装置 200 所包括的用户数据存储模块（图 2 未示出）中。

通常地，可以采用用户的诸如指纹、声纹和虹膜之类的生物特征来唯一标识用户。在本发明的一些实施方式中，可以采用声纹来唯一标识用户，声纹是指说话人语音中携
15 带言语信息的声波频谱，可以唯一标识说话人。语音识别模块 220 可以采用各种声纹识别技术来判断是否存在与第一语音相匹配的用户。

具体地，在各种实施方式中，用户简档可以包括用户的声纹。语音识别模块 220 可以通过判断第一语音是否与用户的声纹相匹配来判断是否存在与第一语音相匹配的用户。

20 下面将详细介绍判断第一语音是否与用户的声纹相匹配的过程。

在一些实施方式中，音频数据可以在被由语音识别模块 220 匹配至用户之前经受不同层次的预处理。在一些实施例中，这种预处理可以促进语音识别模块 220 进行更加高效的语音识别。在各种实施方式中，预处理可以由终端设备 102 或者由另一组件来执行，诸如语音识别装置 200 的组件。在一些实施方式中，语音识别模块 220 本身可以预处理
25 音频数据。

作为预处理的非限制性示例，音频数据可以最初例如由终端设备 102 的麦克风 103 捕捉，作为原始数据（例如，以诸如 wav 文件的“无损”形式或者诸如 MP3 文件的“有损”形式）。这种原始数据可以例如由终端设备 102 或者语音识别装置 200 的一个或多个组件

进行预处理，以促进语音识别。在各种实施方式中，预处理可以包括：采样；量化；去除非语音的音频数据和静默的音频数据；对包括语音的音频数据进行分帧、加窗，以供后续处理等等。

5 经过预处理之后，语音识别模块 220 可以根据包括第一语音的音频数据来提取第一语音的语音特征，并基于第一语音的语音特征将第一语音与用户的声纹进行匹配。

在一些实施方式中，语音特征可以是滤波器组 FBank (Filter Bank)、梅尔频率倒谱系数 MFCC (Mel Frequency Cepstral Coefficients)、感知线性预测系数 PLP、深度特征 Deep Feature、以及能量规整谱系数 PNCC 等特征中的一种或者多种的组合。在一种实施例中，语音识别模块 220 还可以对提取得到的语音特征进行归一化处理。

10 而后，语音识别模块 220 基于第一语音的语音特征，将第一语音与用户的声纹进行匹配，以得到第一语音与用户的声纹之间的相似评分，并根据该相似评分来确定与第一语音相匹配的用户。

具体地，在一些实施方式中，用户的声纹以声纹模型来描述，例如隐马尔可夫模型 (HMM 模型)、高斯混合模型 (GMM 模型) 等等。用户的声纹模型以语音特征为特征，利用包括用户语音的音频数据 (后文简称为用户的音频数据) 训练得到。语音识别模块 220 可以采用匹配运算函数来计算第一语音与用户的声纹之间的相似度。例如可以计算第一语音的语音特征与用户的声纹模型相匹配的后验概率来作为相似评分，也可以计算第一语音的语音特征与用户的声纹模型之间的似然度来作为相似评分。

但由于训练好用户的声纹模型需要大量该用户的音频数据，因此在一些实施方式中，20 用户的声纹模型可以基于与用户无关的通用背景模型，利用少量用户的音频数据训练得到 (同样以语音特征为特征)。例如，可以先使用与用户无关的、多个说话人的音频数据，通过期望最大化算法 EM 训练得到通用背景模型 (Universal Background Model, UBM)，以表征用户无关的特征分布。再基于该 UBM 模型，利用少量的用户的音频数据通过自适应算法 (如最大后验概率 MAP，最大似然线性回归 MLLR 等) 训练得到 GMM 模型 (这样得到的 GMM 模型称之为 GMM-UBM 模型)，以表征用户的特征分布。该 GMM-UBM 模型即为用户25 的声纹模型。此时，语音识别模块 220 可以基于第一语音的语音特征，分别将第一语音与用户的声纹模型和通用背景模型进行匹配，以得到第一语音与用户的声纹之间的相似评分。例如，分别计算第一语音的语音特征与上述 UBM 模型和 GMM-UBM 模型之间的似然度，然后将这两个似然度相除后取对数，将得到的值作为第一语音与用户的声纹之间的

相似评分。

在另一些实施方式中，用户的声纹以声纹向量来描述，例如 i-vector、d-vector、x-vector 和 j-vector 等等。语音识别模块 220 可以至少基于第一语音的语音特征，提取第一语音的声纹向量。

- 5 根据一种实施例，可以先利用第一语音的语音特征训练第一语音说话人的声纹模型。如前文类似地，可以基于预先训练好的与用户无关的上述通用背景模型，利用第一语音的语音特征训练得到第一语音说话人的声纹模型。

- 在得到第一语音说话人的声纹模型之后，可以根据该声纹模型提取第一语音的均值超矢量。例如，可以将第一语音说话人的 GMM-UBM 模型的各个 GMM 分量的均值进行拼接，
10 得到第一语音说话人的 GMM-UBM 模型的均值超矢量，即第一语音的均值超矢量。

之后，可以采用联合因子分析法（JFA）或者简化的联合因子分析法，从第一语音的均值超矢量中提取得到低维的声纹向量。

- 以 i-vector 为例，在训练得到与用户无关的上述通用背景模型（UBM 模型）之后，可以提取该通用背景模型的均值超矢量，并估计全局差异空间（Total Variability Space, T）矩阵。而后基于第一语音的均值超矢量、T 矩阵、通用背景模型的均值超矢量来计算第一语音的 i-vector。
15

具体地，i-vector 可以根据以下公式计算得到：

$$M_{s,h} = m_u + T \omega_{s,h}$$

- 其中， $M_{s,h}$ 是从说话人 s 的语音 h 中得到的均值超矢量， m_u 是通用背景模型的均值超矢量， T 是全局差异空间矩阵， $\omega_{s,h}$ 是全局差异因子，也就是 i-vector。
20

根据另一种实施例，还可以利用训练好的深度神经网络（Deep Neural Network, DNN）来得到第一语音的声纹向量。以 d-vector 为例，DNN 可以包括输入层、隐层和输出层。可以先将第一语音的 FBank 特征输入到 DNN 输入层，DNN 最后一个隐层的输出即为 d-vector。

- 25 在得到第一语音的声纹向量之后，语音识别模块 220 可以基于第一语音的声纹向量和用户的声纹向量，来计算第一语音与用户的声纹之间的相似评分。其中，可以采用支持向量机（SVM）、LDA（Linear Discriminant Analysis，线性判别分析）、PLDA（Probabilistic Linear Discriminant Analysis，概率线性判别分析）、似然度和余

弦距离 (Cosine Distance) 等算法来计算第一语音与用户的声纹之间的相似评分。

以 PLDA 算法为例, 假设语音由 I 个说话人的语音组成, 其中每个说话人有 J 段不一样的语音, 并且定义第 i 个说话人的第 j 段语音为 Y_{ij} 。那么, 定义 Y_{ij} 的生成模型为:

$$Y_{ij} = \mu + Fh_i + Gw_{ij} + \varepsilon_{ij}$$

5 其中, μ 是声纹向量的均值, F 、 G 是空间特征矩阵, 各自代表说话人类间特征空间和类内特征空间。 F 的每一列, 相当于类间特征空间的特征向量, G 的每一列, 相当于类内特征空间的特征向量。向量 h_i 和 w_{ij} 可以看作是該语音分别在各自空间的特征表示, ε_{ij} 则是噪声协方差。如果两条语音的 h_i 特征相同的似然度越大, 即相似评分越高, 那么它们来自同一个说话人的可能性就越大。

10 PLDA 的模型参数包括 4 个, 即 μ 、 F 、 G 和 ε_{ij} , 是采用 EM 算法迭代训练而成。通常地, 可以采用简化版的 PLDA 模型, 忽略类内特征空间矩阵 G 的训练, 只训练类间特征空间矩阵 F , 即:

$$Y_{ij} = \mu + Fh_i + \varepsilon_{ij}$$

15 语音识别模块 220 可以基于第一语音的声纹向量, 参照上述公式得到第一语音的 h_i 特征。同样地, 基于用户的声纹向量, 参照上述公式得到用户语音的 h_i 特征。而后, 可以计算两个 h_i 特征的对数似然比或余弦距离来作为第一语音与用户的声纹之间的相似评分。

20 应当注意的是, 声纹并不限于上述声纹向量 (i-vector、d-vector 和 x-vector 等等) 和上述声纹模型 (HMM 模型和 GMM 模型等等), 相应的相似评分算法也可依据所选定的声纹来任意选取, 本发明对此不做限制。

在各种实施方式中, 如果得到的相似评分超过相似阈值, 则语音识别模块 220 确定第一语音与该用户的声纹相匹配, 也就是确定第一语音与该声纹对应的用户相匹配。否则语音识别模块 220 确定第一语音不与该用户的声纹相匹配。

25 语音识别模块 220 可以将第一语音与每个用户的声纹相匹配, 以判断是否存在与第一语音相匹配的用户。在存在与第一语音相匹配的用户的情况下, 语音识别模块 220 在对音频数据进行语音识别以得到指令之外, 可以将第一语音对应的指令与匹配到的用户相关联地存储, 例如存储至该用户的用户简档。这样, 语音识别装置 200 后续可以根据来自该用户的所有指令分析该用户的行为偏好, 从而为该用户提供个性化和定制化的服务。例如, 根据用户所有与播放歌曲相关的指令来分析用户的歌曲偏好, 从而可以为用

户推荐符合其偏好的歌曲。

在不存在与第一语音相匹配的用户的情况下，语音识别模块 220 可以将该（条）包括第一语音的音频数据存储至音频存储模块 230。音频存储模块 230 适于存储音频数据。

5 用户发现模块 240 可以对音频存储模块 230 所存储的多条音频数据进行聚类，以便从这多条音频数据中确定新用户。这样，对于后续接收到的包括该新用户语音的音频数据，语音识别装置 200 就可以匹配到该新用户，并将对应指令与该新用户相关联地存储，以便后续可以根据来自该新用户的所有指令分析该新用户的行为偏好，从而为该新用户提供个性化服务。在一些实施方式中，用户发现模块 240 可以每隔预定周期就提取所存储的多条音频数据（例如固定数量条音频数据）来进行聚类。

10 具体地，用户发现模块 240 先基于多条音频数据中两两之间的相似评分，将这多条音频数据划分为多个集合。可以认为每个集合所包含的音频数据彼此相似。在一种实施例中，可以采用聚类算法来实现集合的划分。

其中，相似评分的计算已在上述对第一语音与用户的声纹之间的相似评分的计算过程的描述中详细介绍，此处不再赘述。

15 而后，用户发现模块 240 基于集合的样本特征来确定至少一个目标集合，每个目标集合即对应于一个新用户。其中，样本特征可以包括样本密度、样本数量等等，样本则指的是音频数据。在一种实施例中，对于每个集合，可以计算该集合的样本密度和样本数量。而后，选择样本密度和样本数量满足预定条件的集合作为目标集合。预定条件例如可以是：样本密度超过预定密度；样本数量超过预定数量等等。预定条件可以根据所
20 需要确定的目标集合个数来配置，本发明对此不做限制。

在确定目标集合（即发现新用户）之后，用户发现模块 240 为该目标集合对应的新用户创建用户简档，并使用该目标集合中的至少部分音频数据来生成该新用户的声纹。声纹可以是声纹模型或者声纹向量。例如，可以以这些音频数据的语音特征为特征来训练一个 GMM 模型或者 GMM-UBM 模型，作为新用户的声纹。也可以基于这些音频数据的语音特征来提取声纹向量，作为新用户的声纹。具体的声纹生成过程可以参考前文关于声纹的描述，此处不再赘述。
25

其中，可以随机选择目标集合中的至少部分音频数据来生成声纹。也可以根据到目标集合的质心的距离来确定目标集合中用于生成新用户的声纹的音频数据。例如，先确定目标集合的质心，再计算目标集合中各样本到目标集合的质心的距离，选择距离较小

的那些样本作为用于生成新用户的声纹的音频数据。质心的计算为本领域常规技术，此处不再赘述。

如果没有确定目标集合（即没有发现新用户），例如所有集合中没有满足预定条件的目标集合，那么，用户发现模块 240 可以删除这些音频数据，也就是之前从音频存储模块 230 所提取的多条音频数据。

可以理解地，创建用户简档可以认为是用户的注册过程。通常地，用户可以主动提供包括用户语音的音频数据（例如，经由终端设备向服务器发送主动注册请求，并根据相应注册提示针对特定文本主动录入语音），以便根据这些主动提供的音频数据来生成用户的声纹。这一用户主动操作的过程可以认为是主动注册过程。而通过聚类来发现新用户、为其创建用户简档并生成声纹的过程，用户并无感知，因此，该过程可以认为是非主动注册过程。

根据本发明的实施方式，用户简档还可以包括指示用户是否为主动注册的用户标记。如图 2 所示，语音识别装置 200 还可以包括用户注册模块 250。通信模块 210 可以接收包括第二语音的音频数据，第二语音通常用于主动注册新用户，例如，可以是按照终端设备 102 的注册提示而录入的语音。用户注册模块 250 则可以为主动注册的新用户创建用户简档，并使用该包括第二语音的音频数据来生成新用户的声纹，以及将该主动注册的新用户的用户简档中的用户标识置为主动注册。

相应地，对于通过聚类发现的新用户，用户发现模块 240 在为该目标集合对应的新用户创建用户简档时，可以将所创建的用户简档中的用户标识置为非主动注册。

这样，语音识别模块 220 就可以在确定存在与第一语音相匹配的用户之后，基于对应用户简档中的用户标识来判断该用户是否为主动注册。如果用户标记指示用户为非主动注册，语音识别模块 220 可以记录来自该用户的音频数据条数。具体地，用户简档可以包括来自用户的音频数据条数。每接收一条来自该用户的音频数据，语音识别模块 220 将来自用户的音频数据条数加一。相应地，用户发现模块 240 在为非主动注册的新用户创建用户简档时，可以将所创建的用户简档中来自用户的音频数据条数置为初始值。初始值通常可以为 0。

语音识别模块 220 还可以判断来自该用户的音频数据条数是否在特定时段内达到特定数量（例如在自注册以来的 1 个月内达到特定数量）。如果没达到，语音识别模块 220 可以删除该用户对应的用户简档，也就是说，注销该用户。如果达到，则可以不

任何操作。

根据本发明的另一个实施方式，用户简档还可以包括与用户相关联的终端设备的设备标识。例如，在主动注册过程中，通信模块 210 可以接收发送包括第二语音的音频数据的终端设备的设备标识，用户注册模块 250 可以将该设备标识与主动注册的新用户相关联地存储至对应的用户简档。这样，语音识别模块 220 可以在接收音频数据时也接收发送该音频数据的终端设备的设备标识，并在判断是否存在与第一语音相匹配的用户之前，先基于该设备标识判断是否存在与对应终端设备相关联的用户，也就是查找是否存在包括该设备标识的用户简档。

如果不存在与该终端设备相关联的用户，则语音识别模块 220 可以将该音频数据存储至音频存储模块 230。如果存在与该终端设备相关联的用户，则语音识别模块 220 判断是否存在与第一语音相匹配的用户。

此外，本发明的实施例还提供了一种用户识别装置。该用户识别装置包括通信模块、语音识别模块、音频存储模块和用户发现模块。通信模块接收包括第一语音的音频数据，语音识别模块可以判断是否存在与第一语音相匹配的用户，在不存在与第一语音相匹配的用户的情况下，将音频数据存储至音频存储模块。音频存储模块存储音频数据。用户发现模块则可以对音频存储模块所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户，并对新用户进行行为分析。例如，可以根据该新用户的语音对应的指令来分析该用户的行为偏好，从而为该新用户提供个性化服务。

其中，用户识别装置中各模块的处理例如可以与上文中结合图 1 和图 2 所描述的语音识别装置 200 中各模块的处理相同，并能够达到相类似的技术效果，在此不再赘述。

在下文中将结合附图描述在上文中提及的各个模块和装置等的具体结构以及对应的处理方法。

根据本发明的实施方式，上述语音识别装置 200（和上述用户识别装置）中的各种部件，如各种模块等均可以通过如下所述的计算设备 300 来实现。图 3 示出了根据本发明一个实施例的计算设备 300 的示意图。

如图 3 所示，在基本的配置 302 中，计算设备 300 典型地包括系统存储器 306 和一个或者多个处理器 304。存储器总线 308 可以用于在处理器 304 和系统存储器 306 之间的通信。

取决于期望的配置，处理器 304 可以是任何类型的处理，包括但不限于：微处理器（ μ P）、微控制器（ μ C）、数字信息处理器（DSP）或者它们的任何组合。处理器 304 可以包括诸如一级高速缓存 310 和二级高速缓存 312 之类的一个或者多个级别的高速缓存、处理器核心 314 和寄存器 316。示例的处理器核心 314 可以包括运算逻辑单元（ALU）、浮点数单元（FPU）、数字信号处理核心（DSP 核心）或者它们的任何组合。示例的存储器控制器 318 可以与处理器 304 一起使用，或者在一些实现中，存储器控制器 318 可以是处理器 304 的一个内部部分。

取决于期望的配置，系统存储器 306 可以是任意类型的存储器，包括但不限于：易失性存储器（诸如 RAM）、非易失性存储器（诸如 ROM、闪存等）或者它们的任何组合。系统存储器 306 可以包括操作系统 320、一个或者多个应用 322 以及程序数据 324。在一些实施方式中，应用 322 可以布置为在操作系统上由一个或多个处理器 304 利用程序数据 324 执行指令。

计算设备 300 还可以包括有助于从各种接口设备（例如，输出设备 342、外设接口 344 和通信设备 346）到基本配置 302 经由总线/接口控制器 330 的通信的接口总线 340。示例的输出设备 342 包括图形处理单元 348 和音频处理单元 350。它们可以被配置为有助于经由一个或者多个 A/V 端口 352 与诸如显示器或者扬声器之类的各种外部设备进行通信。示例外设接口 344 可以包括串行接口控制器 354 和并行接口控制器 356，它们可以被配置为有助于经由一个或者多个 I/O 端口 358 和诸如输入设备（例如，键盘、鼠标、笔、语音输入设备、触摸输入设备）或者其他外设（例如打印机、扫描仪等）之类的外部设备进行通信。示例的通信设备 346 可以包括网络控制器 360，其可以被布置为便于经由一个或者多个通信端口 364 与一个或者多个其他计算设备 362 通过网络通信链路的通信。

网络通信链路可以是通信介质的一个示例。通信介质通常可以体现为在诸如载波或者其他传输机制之类的调制数据信号中的计算机可读指令、数据结构、程序模块，并且可以包括任何信息递送介质。“调制数据信号”可以是这样的信号，它的数据集中的一个或者多个或者它的改变可以在信号中编码信息的方式进行。作为非限制性的示例，通信介质可以包括诸如有线网络或者专线网络之类的有线介质，以及诸如声音、射频（RF）、微波、红外（IR）或者其他无线介质在内的各种无线介质。这里使用的术语计算机可读介质可以包括存储介质和通信介质二者。

计算设备 300 可以实现为服务器，例如数据库服务器、应用程序服务器和 WEB 服务器等，也可以实现为包括桌面计算机和笔记本计算机配置的个人计算机。当然，计算设备 300 也可以实现为小尺寸便携（或者移动）电子设备的一部分。

5 在根据本发明的实施例中，计算设备 300 被实现为语音识别装置 200，并被配置为执行根据本发明实施例的语音识别方法 400。其中，计算设备 300 的应用 322 中包含执行根据本发明实施例的语音识别方法 400 的多条程序指令，而程序数据 324 还可以存储语音识别系统 100 的配置信息等。

图 4 示出了根据本发明一个实施例的语音识别方法 400。如图 4 所示，语音识别方法 400 始于步骤 S410。

10 在步骤 S410 中，接收包括第一语音的音频数据。如前所述，第一语音通常是指示终端设备 102 执行操作的语音。因此，根据本发明的实施方式，可以对该音频数据进行语音识别而得到第一语音对应的指令，而后向终端设备 102 返回对该指令的响应结果，以便终端设备 102 至少根据该响应结果来执行相应操作。

15 随后在步骤 S420 中，可以判断是否存在与第一语音相匹配的用户。根据一种实施方式，用户均会对应有记录与用户相关的数据的用户简档，这些用户简档可以存储在与语音识别装置 200 相耦接的用户数据存储设备中，也可以存储在语音识别装置 200 所包括的用户数据存储模块中。

通常地，可以采用用户的诸如指纹、声纹和虹膜之类的生物特征来唯一标识用户。在本发明的一种实施方式中，可以采用声纹来唯一标识用户，用户简档包括用户的声纹，20 可以判断第一语音是否与用户的声纹相匹配，以判断是否存在与第一语音相匹配的用户。

具体地，可以先根据包括第一语音的音频数据，提取第一语音的语音特征。在一些实施方式中，语音特征可以是滤波器组 FBank (Filter Bank)、梅尔频率倒谱系数 MFCC (Mel Frequency Cepstral Coefficients)、感知线性预测系数 PLP、深度特征 Deep Feature、以及能量规整谱系数 PNCC 等特征中的一种或者多种的组合。

25 而后基于第一语音的语音特征得到第一语音与用户的声纹之间的相似评分，根据该相似评分来确定第一语音是否与用户的声纹相匹配。如果得到的相似评分超过相似阈值，则确定第一语音与该用户的声纹相匹配，否则确定第一语音不与该用户的声纹相匹配。

如果存在与第一语音相匹配的用户，则可以将第一语音对应的指令与匹配到的用户

相关联地存储。如果不存在与第一语音相匹配的用户，那么在步骤 S430 中，存储该（条）音频数据。

而后在步骤 S440 中，对所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户。具体地，可以先基于多条音频数据中两两之间的相似评分，将多条音频数据划分为多个集合。再基于集合的样本密度和样本数量确定至少一个目标集合，目标集合对应于新用户。最后为目标集合对应的新用户创建用户简档，并使用目标集合中的至少部分音频数据来生成该新用户的声纹。

在一种实施例中，可以根据到目标集合的质心的距离来确定目标集合中用于生成新用户的声纹的音频数据。例如，先确定目标集合的质心，再计算目标集合中各样本到目标集合的质心的距离，选择距离较小的那些样本作为用于生成新用户的声纹的音频数据。

如果没有确定目标集合，则可以删除这些音频数据，也就是之前的多条音频数据。

根据本发明的一种实施方式，用户简档还可以包括指示用户是否为主动注册的用户标记，在为目标集合对应的新用户创建用户简档时，可以将该用户简档中的用户标识置为非主动注册。在存在与第一语音相匹配的用户且对应的用户标记指示用户为非主动注册的情况下，还可以记录来自用户的音频数据条数，判断音频数据条数是否在特定时间段内达到特定数量。若否，可以删除该用户对应的用户简档。

根据本发明的一种实施方式，语音识别方法 400 还可以包括步骤：接收包括第二语音的音频数据，第二语音通常用于主动注册新用户。为主动注册的新用户创建用户简档，并使用该包括第二语音的音频数据来生成新用户的声纹，以及将该主动注册的新用户的用户简档中的用户标识置为主动注册。

根据本发明的一种实施方式，用户简档还可以包括与用户相关联的终端设备的设备标识，语音识别方法 400 还可以包括步骤：接收发送上述音频数据的终端设备的设备标识，基于该设备标识判断是否存在与该终端设备相关联的用户。如果不存在，则存储上述音频数据。

语音识别方法 400 的具体步骤以及实施例，在结合图 1~图 3 对语音识别系统 100 的描述中已经详细公开，此处不再赘述。

此外，本发明实施例还提供了一种用户识别方法，包括步骤：接收包括第一语音的音频数据；判断是否存在与第一语音相匹配的用户；在不存在与第一语音相匹配的用户

的情况下，存储该音频数据；对所存储的多条音频数据进行聚类，以便从多条音频数据中确定新用户，并对新用户进行行为分析。其中，用户识别方法中各步骤的处理例如可以与上文中结合图 4 所描述的语音识别方法 400 中各步骤的处理相同，并能够达到相类似的技术效果，在此不再赘述。

5 综上所述，根据本发明实施例的语音识别方案，通过对所存储的多条音频数据进行聚类来从中确定新用户以及生成该新用户的声纹，以便后续可以根据声纹识别该用户，并根据来自该用户的指令分析该用户的行为偏好，从而可以为该用户提供更精准的个性化服务。并且，整个新用户确定和声纹生成过程用户是无感知的，省去了用户的主动注册操作，提高了用户的使用体验。

10 应当理解，为了精简本公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释成反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，
15 遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本发明的单独实施例。

本领域那些技术人员应当理解在本文所公开的示例中的设备的模块或单元或组件可以布置在如该实施例中所描述的设备中，或者可替换地可以定位在与该示例中的设备不同的一个或多个设备中。前述示例中的模块可以组合为一个模块或者此外可以分成多个
20 子模块。

本领域那些技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外，可以采用
25 任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中

所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式使用。

此外，所述实施例中的一些在此被描述成可以由计算机系统的处理器或者由执行所述功能的其它装置实施的方法或方法元素的组合。因此，具有用于实施所述方法或方法元素的必要指令的处理器形成用于实施该方法或方法元素的装置。此外，装置实施例的在此所述的元素是如下装置的例子：该装置用于实施由为了实施该发明的目的的元素所执行的功能。

如在此所使用的那样，除非另行规定，使用序数词“第一”、“第二”、“第三”等等来描述普通对象仅仅表示涉及类似对象的不同实例，并且并不意图暗示这样被描述的对象必须具有时间上、空间上、排序方面或者以任意其它方式的给定顺序。

尽管根据有限数量的实施例描述了本发明，但是受益于上面的描述，本技术领域内的技术人员明白，在由此描述的本发明的范围内，可以设想其它实施例。此外，应当注意，本说明书中使用的语言主要是为了可读性和教导的目的而选择的，而不是为了解释或者限定本发明的主题而选择的。因此，在不偏离所附权利要求书的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。对于本发明的范围，对本发明所做的公开是说明性的，而非限制性的，本发明的范围由所附权利要求书限定。

权 利 要 求 书

1、一种语音识别方法，包括步骤：

接收包括第一语音的音频数据；

判断是否存在与所述第一语音相匹配的用户；

- 5 在不存在与所述第一语音相匹配的用户的情况下，存储所述音频数据；
对所存储的多条音频数据进行聚类，以便从所述多条音频数据中确定新用户。

2、如权利要求 1 所述的方法，其中，所述用户对应有用户简档，所述用户简档包括所述用户的声纹，所述判断是否存在与所述第一语音相匹配的用户的步骤包括：

- 10 判断所述第一语音是否与所述用户的声纹相匹配，以判断是否存在与所述第一语音相匹配的用户。

3、如权利要求 2 所述的方法，其中，所述对所存储的多条音频数据进行聚类，以便从所述多条音频数据中确定新用户的步骤包括：

基于所述多条音频数据中两两之间的相似评分，将所述多条音频数据划分为多个集合；

- 15 基于所述集合的样本密度和样本数量确定至少一个目标集合，所述目标集合对应于所述新用户；

为所述目标集合对应的新用户创建用户简档，并使用所述目标集合中的至少部分音频数据来生成该新用户的声纹。

- 20 4、如权利要求 3 所述的方法，其中，所述使用所述目标集合中的至少部分音频数据来生成该新用户的声纹的步骤包括：

根据到所述目标集合的质心的距离来确定所述目标集合中用于生成新用户的声纹的音频数据。

5、如权利要求 3 所述的方法，其中，所述用户简档包括指示用户是否为主动注册的用户标记，所述为所述目标集合对应的新用户创建用户简档的步骤包括：

- 25 将为目标集合对应的新用户所创建的用户简档中的用户标识置为非主动注册；以及
所述方法还包括步骤：

在存在与所述第一语音相匹配的用户且对应的用户标记指示所述用户为非主动注册的情况下，记录来自所述用户的音频数据条数。

6、如权利要求 5 所述的方法，其中，还包括步骤：

- 30 在记录来自所述用户的音频数据条数之后，判断所述音频数据条数是否在特定时间

段内达到特定数量；若否，删除所述用户对应的用户简档。

7、如权利要求 2 所述的方法，其中，所述用户简档还包括与用户相关联的终端设备的设备标识，所述方法包括步骤：

接收发送所述音频数据的终端设备的设备标识；

- 5 基于所述设备标识判断是否存在与所述终端设备相关联的用户；
 如果不存在，则存储所述音频数据。

8、如权利要求 1 所述的方法，其中，还包括步骤：

在存在与所述第一语音相匹配的用户的情况下，将所述第一语音对应的指令与所述用户相关联地存储。

- 10 9、如权利要求 1-8 中任一项所述的方法，其中，还包括步骤：

接收包括第二语音的音频数据，所述第二语音用于主动注册新用户；

为主动注册的新用户创建用户简档，并使用所述包括第二语音的音频数据来生成所述新用户的声纹；以及

将为主动注册的新用户所创建的用户简档中的用户标识置为主动注册。

- 15 10、如权利要求 9 所述的方法，其中，还包括步骤：

接收发送包括第二语音的音频数据的终端设备的设备标识；

将所述设备标识与所述主动注册的新用户相关联地存储至对应的用户简档。

11、如权利要求 2-10 中任一项所述的方法，其中，所述判断所述第一语音是否与
所述用户的声纹相匹配的步骤包括：

- 20 根据所述包括第一语音的音频数据，提取第一语音的语音特征；

基于所述第一语音的语音特征得到所述第一语音与用户的声纹之间的相似评分；

根据所述相似评分来确定第一语音是否与用户的声纹相匹配。

12、一种用户识别方法，包括步骤：

接收包括第一语音的音频数据；

- 25 判断是否存在与所述第一语音相匹配的用户；

在不存在与所述第一语音相匹配的用户的情况下，存储所述音频数据；

对所存储的多条音频数据进行聚类，以便从所述多条音频数据中确定新用户，并对
所述新用户进行行为分析。

13、一种语音识别装置，包括：

- 30 通信模块，适于接收包括第一语音的音频数据；

语音识别模块，适于判断是否存在与所述第一语音相匹配的用户；在不存在与所述第一语音相匹配的用户的情况下，将所述音频数据存储至音频存储模块；

音频存储模块，适于存储所述音频数据；以及

用户发现模块，适于对所述音频存储模块所存储的多条音频数据进行聚类，以便从
5 所述多条音频数据中确定新用户。

14、如权利要求 13 所述的装置，其中，所述用户对应有用户简档，所述用户简档包括所述用户的声纹，所述语音识别模块适于

判断所述第一语音是否与所述用户的声纹相匹配，以判断是否存在与所述第一语音相匹配的用户。

10 15、如权利要求 14 所述的装置，其中，所述用户发现模块适于
基于所述多条音频数据中两两之间的相似评分，将所述多条音频数据划分为多个集合；

基于所述集合的样本密度和样本数量确定至少一个目标集合，所述目标集合对应于所述新用户；

15 为所述目标集合对应的新用户创建用户简档，并使用所述目标集合中的至少部分音频数据来生成该新用户的声纹。

16、如权利要求 15 所述的装置，其中，所述用户发现模块适于
根据到所述目标集合的质心的距离来确定所述目标集合中用于生成新用户的声纹的音频数据。

20 17、如权利要求 15 所述的装置，其中，所述用户简档包括指示用户是否为主动注册的用户标记，所述用户发现模块适于

将为目标集合对应的新用户所创建的用户简档中的用户标识置为非主动注册；以及
所述语音识别模块还适于

25 在存在与所述第一语音相匹配的用户且对应的用户标记指示所述用户为非主动注册的情况下，记录来自所述用户的音频数据条数。

18、如权利要求 17 所述的装置，其中，所述语音识别模块适于
在记录来自所述用户的音频数据条数之后，判断所述音频数据条数是否在特定时间段内达到特定数量；若否，删除所述用户对应的用户简档。

19、如权利要求 14 所述的装置，其中，所述用户简档包括与用户相关联的终端设
30 备的设备标识，

所述通信模块还适于接收发送所述音频数据的终端设备的设备标识；以及所述语音识别模块还适于

基于所述设备标识判断是否存在与所述终端设备相关联的用户；

如果不存在，则将所述音频数据存储至所述音频存储模块。

5 20、如权利要求 13 所述的装置，其中，所述语音识别模块还适于

在存在与所述第一语音相匹配的用户的情况下，将所述第一语音对应的指令与所述用户相关联地存储。

21、如权利要求 13-19 中任一项所述的装置，其中，所述通信模块还适于接收包括第二语音的音频数据，所述第二语音用于主动注册新用户；所述装置还包括：

10 用户注册模块，适于为主动注册的新用户创建用户简档，并使用所述包括第二语音的音频数据来生成所述新用户的声纹；以及将为主动注册的新用户所创建的用户简档中的用户标识置为主动注册。

22、如权利要求 21 所述的装置，其中，所述通信模块还适于接收发送包括第二语音的音频数据的终端设备的设备标识；所述用户注册模块还适于将所述设备标识与所述
15 主动注册的新用户相关联地存储至对应的用户简档。

23、如权利要求 14-22 中任一项所述的装置，其中，所述语音识别模块还适于

根据所述包括第一语音的音频数据，提取第一语音的语音特征；

基于所述第一语音的语音特征得到所述第一语音与用户的声纹之间的相似评分；

根据所述相似评分来确定第一语音是否与用户的声纹相匹配。

20 24、如权利要求 14-23 中任一项所述的装置，该装置驻留在终端设备中，所述终端设备为音箱、电视机或洗衣机。

25、一种用户识别装置，包括：

通信模块，适于接收包括第一语音的音频数据；

25 语音识别模块，适于判断是否存在与所述第一语音相匹配的用户；在不存在与所述第一语音相匹配的用户的情况下，将所述音频数据存储至音频存储模块；

音频存储模块，适于存储所述音频数据；以及

用户发现模块，适于对所述音频存储模块所存储的多条音频数据进行聚类，以便从所述多条音频数据中确定新用户，并对所述新用户进行行为分析。

26、一种语音识别系统，包括终端设备和服务器，其中

30 所述终端设备适于接收说话人的语音，并将包括语音的音频数据发送至所述服务

器；所述服务器驻留有如权利要求 13-24 中任一项所述的语音识别装置。

27、一种计算设备，包括：

至少一个处理器；和

5 存储有程序指令的存储器，其中，所述程序指令被配置为适于由所述至少一个处理器执行，所述程序指令包括用于执行如权利要求 1-11 中任一项所述的语音识别方法的指令。

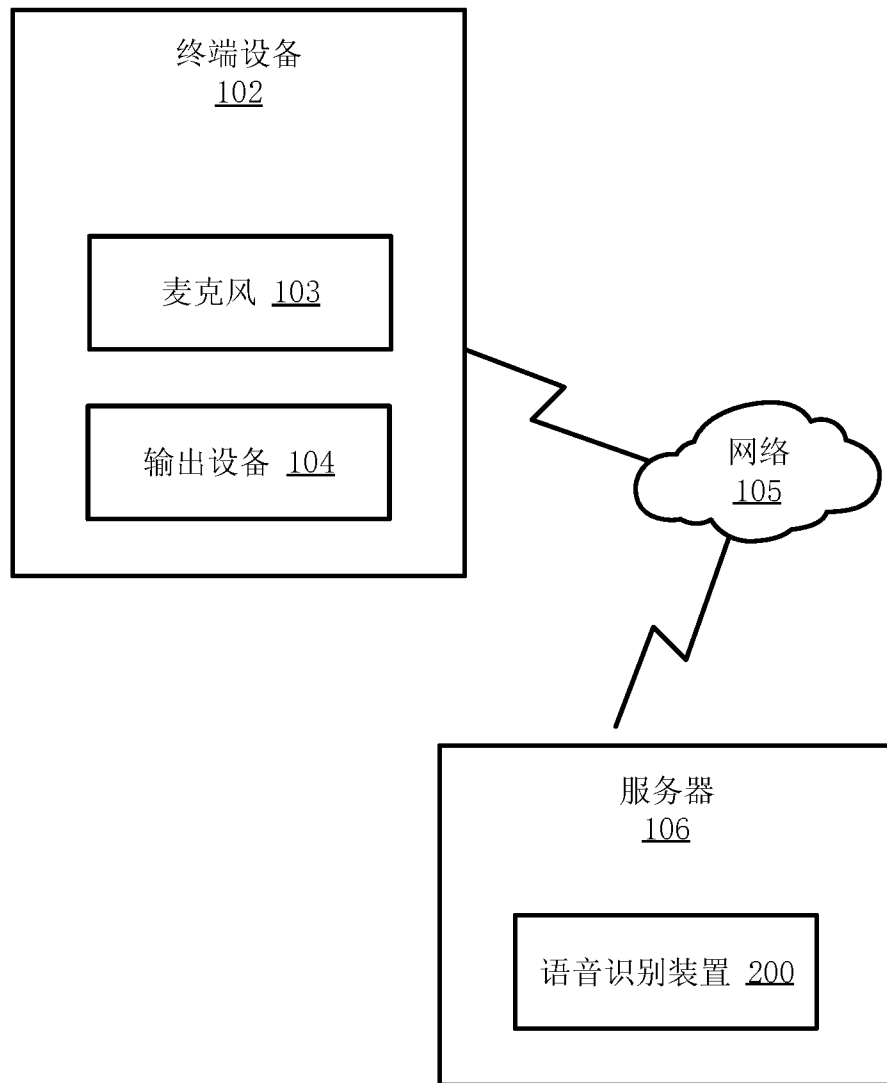


图 1

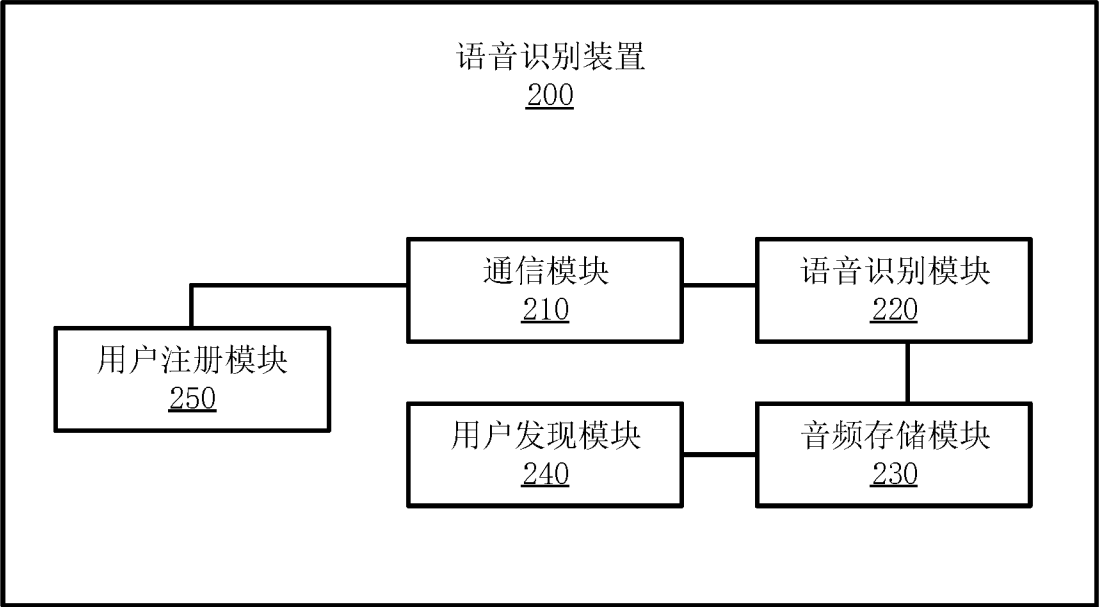


图 2

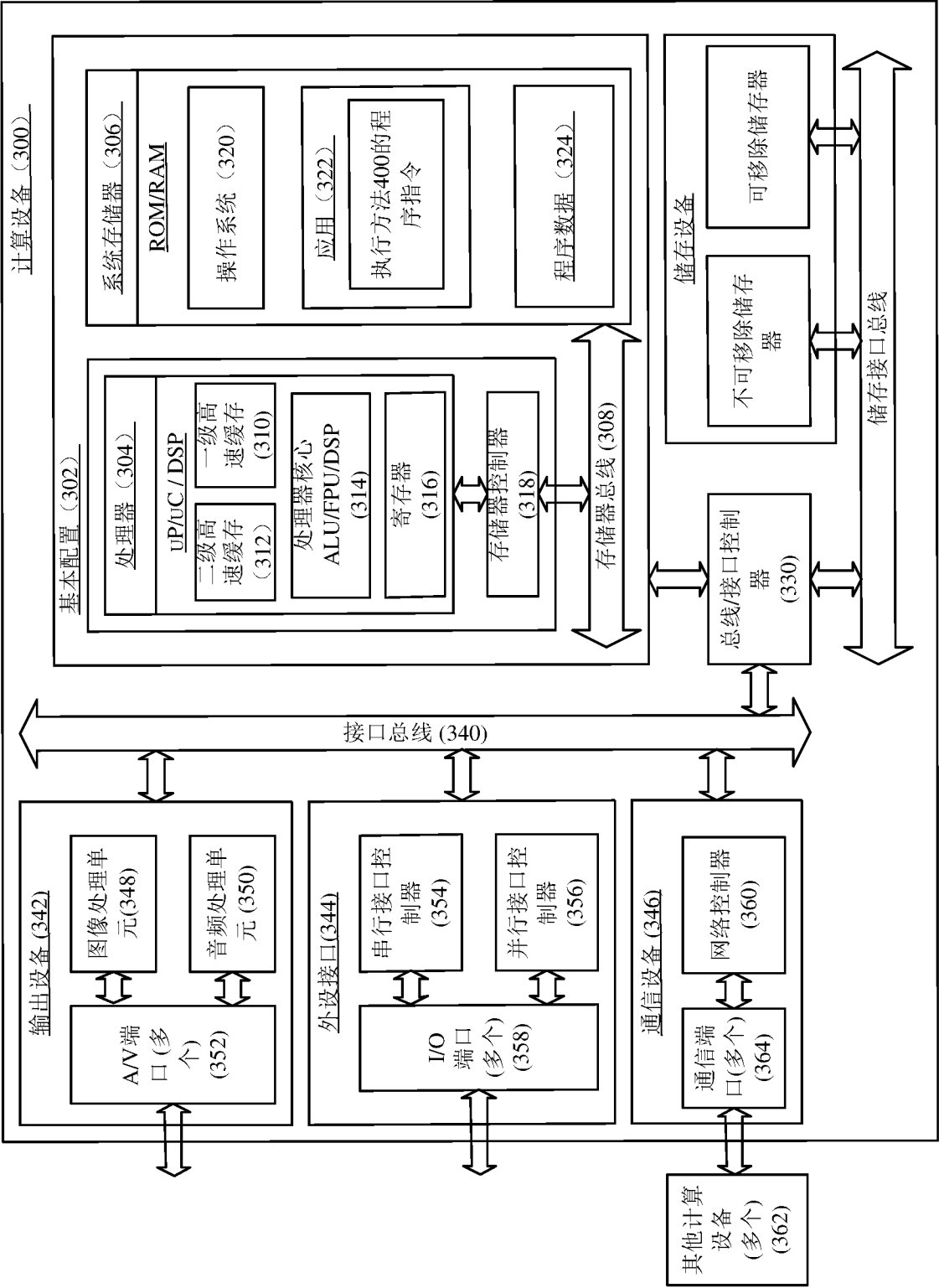


图3

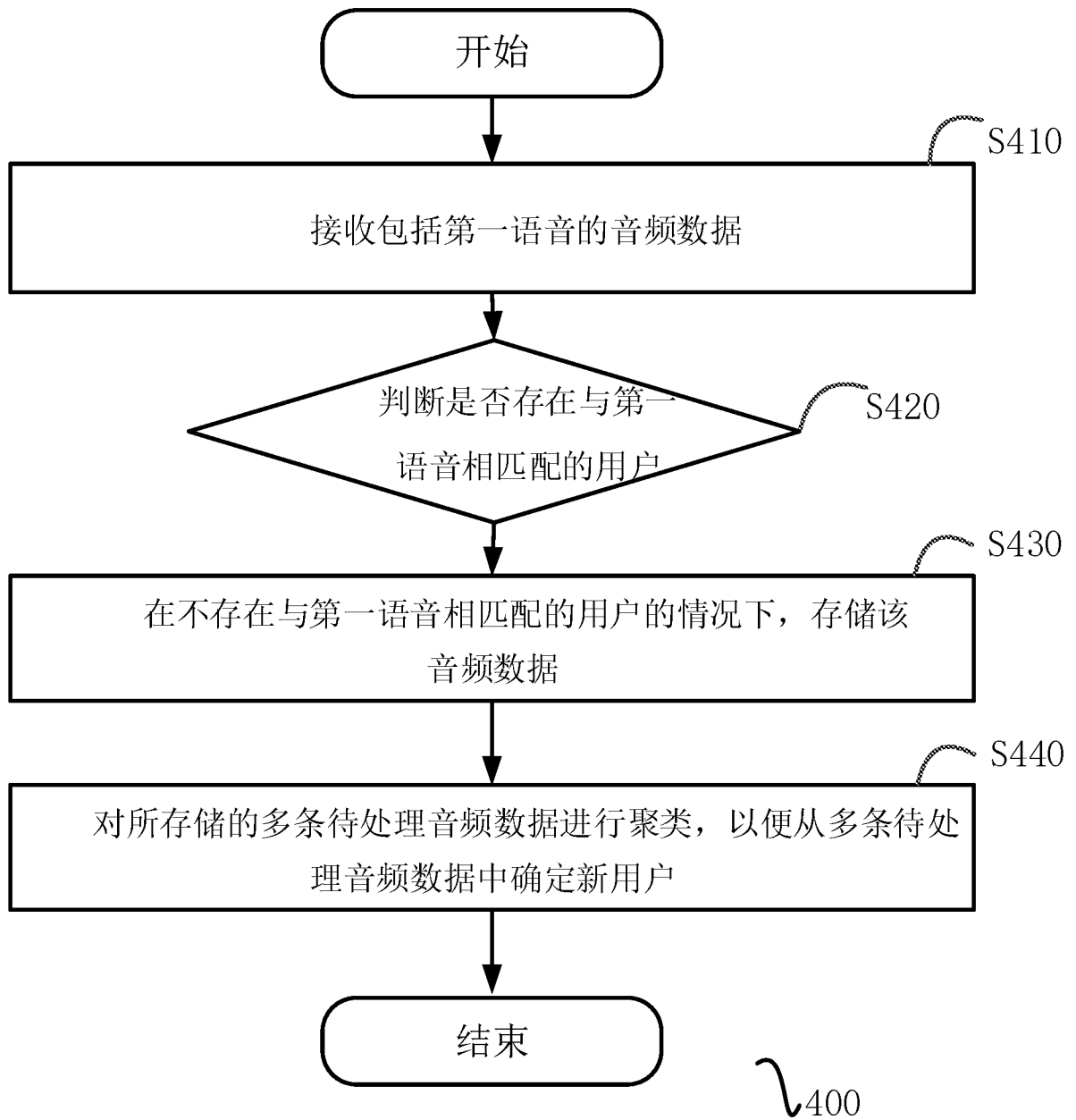


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/115308

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/22(2006.01)i; G10L 17/22(2013.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

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

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

CNABS, CNTXT, CNKI, DWPI, SIPOABS, VEN, 阿里巴巴, 赵情恩, 索宏彬, 刘刚, 卓著, 雷赞, 张平, 孙尧, 语音, 音频, 声音, 匹配, 声纹, 特征, 声码, 新, 注册, 用户, 新注册, 未注册, 相似, 集, 组, 集合, 簇, 聚类, 数量, 密度, voice, sound, voiceprint, characteristic, feature, match, user, identity, new, register, cluster

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 107863108 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 30 March 2018 (2018-03-30) description, paragraphs 0005-0015, 0055, 0057-0059 and 0077	1-27
Y	CN 106782564 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 31 May 2017 (2017-05-31) description, paragraphs 0012 and 0023	1-27
A	CN 108075892 A (ALIBABA GROUP HOLDING LIMITED) 25 May 2018 (2018-05-25) entire document	1-27
A	CN 107147618 A (BEIJING ORION STAR TECHNOLOGY CO., LTD.) 08 September 2017 (2017-09-08) entire document	1-27
A	CN 108597525 A (SICHUAN YUANJIAN TECHNOLOGY CO., LTD.) 28 September 2018 (2018-09-28) entire document	1-27



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

04 February 2020

Date of mailing of the international search report

10 February 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/115308**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107978311 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 01 May 2018 (2018-05-01) entire document	1-27
A	US 2009187405 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 23 July 2009 (2009-07-23) entire document	1-27

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/115308

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107863108	A	30 March 2018	JP	2019091419	A	13 June 2019
				US	2019147885	A1	16 May 2019
CN	106782564	A	31 May 2017	JP	2018081297	A	24 May 2018
				JP	6429945	B2	28 November 2018
				CN	106782564	B	11 September 2018
				US	2018144742	A1	24 May 2018
				US	10140984	B2	27 November 2018
				US	2019066665	A1	28 February 2019
CN	108075892	A	25 May 2018	HK	1255817	A0	30 August 2019
CN	107147618	A	08 September 2017	WO	2018188586	A1	18 October 2018
CN	108597525	A	28 September 2018	CN	108597525	B	03 May 2019
CN	107978311	A	01 May 2018	WO	2019101083	A1	31 May 2019
US	2009187405	A1	23 July 2009	US	8140340	B2	20 March 2012

国际检索报告

国际申请号

PCT/CN2019/115308

A. 主题的分类

G10L 15/22 (2006.01) i; G10L 17/22 (2013.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNABS, CNTXT, CNKI, DWPI, SIPOABS, VEN, 阿里巴巴, 赵情恩, 索宏彬, 刘刚, 卓著, 雷赞, 张平, 孙尧, 语音, 音频, 声音, 匹配, 声纹, 特征, 声码, 新, 注册, 用户, 新注册, 未注册, 相似, 集, 组, 集合, 簇, 聚类, 数量, 密度, voice, sound, voiceprint, characteristic, feature, match, user, identity, new, register, cluster

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	CN 107863108 A (百度在线网络技术北京有限公司) 2018年 3月 30日 (2018 - 03 - 30) 说明书0005-0015、0055、0057-0059、0077段	1-27
Y	CN 106782564 A (百度在线网络技术北京有限公司) 2017年 5月 31日 (2017 - 05 - 31) 说明书0012、0023段	1-27
A	CN 108075892 A (阿里巴巴集团控股有限公司) 2018年 5月 25日 (2018 - 05 - 25) 全文	1-27
A	CN 107147618 A (北京猎户星空科技有限公司) 2017年 9月 8日 (2017 - 09 - 08) 全文	1-27
A	CN 108597525 A (四川远鉴科技有限公司) 2018年 9月 28日 (2018 - 09 - 28) 全文	1-27
A	CN 107978311 A (腾讯科技深圳有限公司) 2018年 5月 1日 (2018 - 05 - 01) 全文	1-27
A	US 2009187405 A1 (INT. BUSINESS MACHINES CORP.) 2009年 7月 23日 (2009 - 07 - 23) 全文	1-27

☐ 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

2020年 2月 4日

国际检索报告邮寄日期

2020年 2月 10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

吴琼

电话号码 (86-10)62089510

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/115308

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	107863108	A	2018年 3月 30日	JP	2019091419	A	2019年 6月 13日
				US	2019147885	A1	2019年 5月 16日
CN	106782564	A	2017年 5月 31日	JP	2018081297	A	2018年 5月 24日
				JP	6429945	B2	2018年 11月 28日
				CN	106782564	B	2018年 9月 11日
				US	2018144742	A1	2018年 5月 24日
				US	10140984	B2	2018年 11月 27日
				US	2019066665	A1	2019年 2月 28日
CN	108075892	A	2018年 5月 25日	HK	1255817	A0	2019年 8月 30日
CN	107147618	A	2017年 9月 8日	WO	2018188586	A1	2018年 10月 18日
CN	108597525	A	2018年 9月 28日	CN	108597525	B	2019年 5月 3日
CN	107978311	A	2018年 5月 1日	WO	2019101083	A1	2019年 5月 31日
US	2009187405	A1	2009年 7月 23日	US	8140340	B2	2012年 3月 20日