

(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



(10) 国际公布号

WO 2019/196305 A1

(43) 国际公布日

2019 年 10 月 17 日 (17.10.2019)

WIPO | PCT

(51) 国际专利分类号:

G10L 17/24 (2013.01) G10L 17/16 (2013.01)

G10L 17/08 (2013.01) G10L 17/18 (2013.01)

G10L 17/04 (2013.01) G10L 25/24 (2013.01)

(21) 国际申请号: PCT/CN2018/102208

(22) 国际申请日: 2018 年 8 月 24 日 (24.08.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

201810311721.2 2018年4月9日 (09.04.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 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)。于夕畔(YU, Xipan); 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)。李瑾瑾(LI, Jinjin); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市沃德知识产权代理事务所(普通合伙)(SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市福田区园岭街道八卦四路 10 号中浩大厦 1528-1530 室于志光, 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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

(54) Title: ELECTRONIC DEVICE, IDENTITY VERIFICATION METHOD, AND STORAGE MEDIUM

(54) 发明名称: 电子装置、身份验证的方法及存储介质

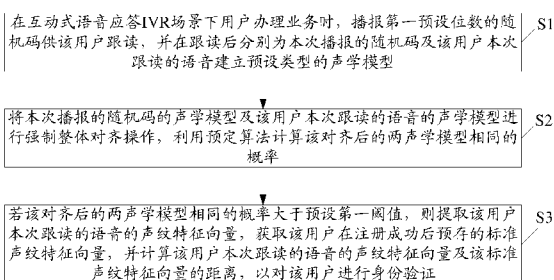


图 2

S1 When a user performs a service in an interactive voice response (IVR) scenario, broadcast a random code of a first preset number of bits for the user to read, and upon reading, establish acoustic models of preset types for the broadcasted random code and the voice read out by the user, respectively

S2 Perform a forced overall alignment operation on the acoustic model of the broadcasted random code and the acoustic model of the voice read out by the user, and use a predetermined algorithm to calculate the probability that the two aligned acoustic models are the same

S3 If the probability that the two aligned acoustic models are the same is greater than a preset first threshold, extract a voiceprint feature vector of the voice read out by the user, acquire a standard voiceprint feature vector of the user pre-stored after successful registration, and calculate the distance between the voiceprint feature vector of the voice read out by the user and that standard voiceprint feature vector, so as to perform identity verification on the user

(57) Abstract: The present application relates to an electronic device, an identity verification method, and a storage medium. The method comprises: when a user performs transactions in an IVR scenario, broadcasting a random code of a first preset number of bits for the user to read, and upon reading, establishing acoustic models of preset types for the broadcasted random code and the voice read out by the user, respectively; performing a forced overall alignment operation on the acoustic model of the broadcasted random code and the acoustic model of the voice read out by the user, and using a predetermined algorithm to calculate the probability that the two aligned acoustic models are the same; if the probability is greater than a preset first threshold, extracting a voiceprint feature vector of the voice read out by the user, acquiring a standard voiceprint feature vector of the user pre-stored after successful registration, and calculating the distance between the voiceprint feature vector of the voice read out by the user and that standard voiceprint feature vector, so as to perform identity verification on the user. The present application performs dual verification on the identity of a user, and can accurately determine the identity of the user.

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本申请涉及一种电子装置、身份验证的方法及存储介质, 该方法包括: 在IVR场景下用户办理业务时, 播报第一预设位数的随机码供该用户跟读, 并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型; 将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作, 利用预定算法计算该对齐后的两声学模型相同的概率; 若该概率大于预设第一阈值, 则提取该用户本次跟读的语音的声纹特征向量, 获取该用户在注册成功后预存的标准声纹特征向量, 并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离, 以对该用户进行身份验证。本申请对用户身份进行双重验证, 能够准确确认用户身份。

电子装置、身份验证的方法及存储介质

优先权申明

本申请基于巴黎公约申明享有 2018 年 04 月 09 日递交的申请号为 CN
5 2018103117212、名称为“电子装置、身份验证的方法及存储介质”中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

10 本申请涉及通信技术领域，尤其涉及一种电子装置、身份验证的方法及存储介质。

背景技术

目前，在交互式语音应答 IVR(Interactive Voice Response)场景中，提供
15 了将交互式语音应答 IVR 与声纹识别结合，以对客户进行身份验证的方案，例如，客户收到信用卡后使用电话进行信用卡激活或修改密码时，需要验证客户身份的场景。现有技术交互式语音应答 IVR(Interactive Voice Response)场景中，鉴于远程声纹验证双方不是当面进行验证，因此，可能会存在客户利用预先准备的合成音的欺诈行为，不能准确确认客户身份，身份验证的安全
20 性低。

发明内容

本申请的目的在于提供一种电子装置、身份验证的方法及存储介质，旨在对用户身份进行双重验证，能够准确确认用户身份。

25 为实现上述目的，本申请提供一种电子装置，所述电子装置包括存储器及与所述存储器连接的处理器，所述存储器中存储有可在所述处理器上运行的处理系统，所述处理系统被所述处理器执行时实现如下步骤：

声学模型建立步骤，在互动式语音应答 IVR 场景下用户办理业务时，播报第一预设位数的随机码供该用户跟读，并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型；

5 强制整体对齐步骤，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算该对齐后的两声学模型相同的概率；

身份验证步骤，若该对齐后的两声学模型相同的概率大于预设第一阈值，则提取该用户本次跟读的语音的声纹特征向量，获取该用户在注册成功后预存的标准声纹特征向量，并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离，以对该用户进行身份验证。

10

为实现上述目的，本申请还提供一种身份验证的方法，所述身份验证的方法包括：

S1，在互动式语音应答 IVR 场景下用户办理业务时，播报第一预设位数的随机码供该用户跟读，并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型；

15

S2，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算该对齐后的两声学模型相同的概率；

S3，若该对齐后的两声学模型相同的概率大于预设第一阈值，则提取该用户本次跟读的语音的声纹特征向量，获取该用户在注册成功后预存的标准声纹特征向量，并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离，以对该用户进行身份验证。

20

本申请还提供一种计算机可读存储介质，所述计算机可读存储介质上存储有处理系统，所述处理系统被处理器执行时实现上述的身份验证的方法的步骤。

25

本申请的有益效果是：本申请在互动式语音应答 IVR 场景下进行身份识别时，利用随机码供用户跟读能够有效防止了预先准备的合成音进行欺诈，将随机码与声纹识别结合，实现了对用户身份的双重验证，能够准确确认用户身份，提高互动式语音应答 IVR 场景下身份验证的安全性，此外，对播报的随机码的声学模型及该用户跟读的语音的声学模型进行强制整体对齐操作，能够降低计算量，提高了身份识别效率。

附图说明

图 1 为本申请各个实施例一可选的应用环境示意图；

图 2 为本申请身份验证的方法一实施例的流程示意图。

具体实施方式

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

需要说明的是，在本申请中涉及“第一”、“第二”等的描述仅用于描述目的，而不能理解为指示或暗示其相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。另外，各个实施例之间的技术方案可以相互结合，但是必须是以本领域普通技术人员能够实现为基础，当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在，也不在本申请要求的保护范围之内。

参阅图 1 所示，是本申请身份验证的方法的较佳实施例的应用环境示意

图。该应用环境示意图包括电子装置 1 及终端设备。电子装置 1 可以通过网络、近场通信技术等技术适合的与终端设备进行数据交互。本实施例中，用户通过终端设备登录电子装置 1 的互动式语音应答 IVR 系统，以执行声纹注册及声纹识别的操作。

5 所述终端设备包括，但不限于，任何一种可与用户通过键盘、鼠标、遥控器、触摸板或者声控设备等方式进行人机交互的电子产品，例如，个人计算机、平板电脑、智能手机、个人数字助理 (Personal Digital Assistant, PDA)、游戏机、交互式网络电视 (Internet Protocol Television, IPTV)、智能式穿戴式设备、导航装置等等的可移动设备，或者诸如数字 TV、台式计算机、笔
10 记本、服务器等等的固定终端。

所述电子装置 1 是一种能够按照事先设定或者存储的指令，自动进行数值计算和/或信息处理的设备。所述电子装置 1 可以是计算机、也可以是单个网络服务器、多个网络服务器组成的服务器组或者基于云计算的由大量主机或者网络服务器构成的云，其中云计算是分布式计算的一种，由一群松散耦
15 合的计算机集组成的一个超级虚拟计算机。

在本实施例中，电子装置 1 可包括，但不仅限于，可通过系统总线相互通信连接的存储器 11、处理器 12、网络接口 13，存储器 11 存储有可在处理器 12 上运行的处理系统。需要指出的是，图 1 仅示出了具有组件 11-13 的电子装置 1，但是应理解的是，并不要求实施所有示出的组件，可以替代的实
20 施更多或者更少的组件。

其中，存储器 11 包括内存及至少一种类型的可读存储介质。内存为电子装置 1 的运行提供缓存；可读存储介质可为如闪存、硬盘、多媒体卡、卡型存储器（例如，SD 或 DX 存储器等）、随机访问存储器 (RAM)、静态随机访问存储器 (SRAM)、只读存储器 (ROM)、电可擦除可编程只读存储器
25 (EEPROM)、可编程只读存储器 (PROM)、磁性存储器、磁盘、光盘等的

非易失性存储介质。在一些实施例中，可读存储介质可以是电子装置 1 的内部存储单元，例如该电子装置 1 的硬盘；在另一些实施例中，该非易失性存储介质也可以是电子装置 1 的外部存储设备，例如电子装置 1 上配备的插接式硬盘，智能存储卡(Smart Media Card, SMC)，安全数字(Secure Digital, SD)卡，闪存卡(Flash Card)等。本实施例中，存储器 11 的可读存储介质通常用于存储安装于电子装置 1 的操作系统和各类应用软件，例如存储本申请一实施例中的处理系统的程序代码等。此外，存储器 11 还可以用于暂时地存储已经输出或者将要输出的各类数据。

所述处理器 12 在一些实施例中可以是中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器 12 通常用于控制所述电子装置 1 的总体操作，例如执行与所述终端设备进行数据交互或者通信相关的控制和处理等。本实施例中，所述处理器 12 用于运行所述存储器 11 中存储的程序代码或者处理数据，例如运行处理系统等。

所述网络接口 13 可包括无线网络接口或有线网络接口，该网络接口 13 通常用于在所述电子装置 1 与其他电子设备之间建立通信连接。本实施例中，网络接口 13 主要用于将电子装置 1 与一个或多个终端设备相连，在电子装置 1 与一个或多个终端设备之间建立数据传输通道和通信连接。

所述处理系统存储在存储器 11 中，包括至少一个存储在存储器 11 中的计算机可读指令，该至少一个计算机可读指令可被处理器 12 执行，以实现本申请各实施例的方法；以及，该至少一个计算机可读指令依据其各部分所实现的功能不同，可被划为不同的逻辑模块。

在一实施例中，上述处理系统被所述处理器 12 执行时实现如下步骤：

声学模型建立步骤，在互动式语音应答 IVR 场景下用户办理业务时，播报第一预设位数的随机码供该用户跟读，并在跟读后分别为本次播报的随机

码及该用户本次跟读的语音建立预设类型的声学模型；

在互动式语音应答 IVR 场景下，用户请求办理业务时发送身份识别码，例如身份证号，在接收到用户的请求后，分析用户所办理的业务是否需要进一步的身份验证，且根据用户的身份识别码分析该用户是否已注册有声纹，
5 若需要进一步的身份验证且该用户已注册有声纹，则生成第一预设位数的随机码并采用语音合成技术以语音形式播报该随机码，引导用户进行跟读，该第一预设位数例如为 8 位。

在用户跟读后，为本次播报的随机码的语音建立预设类型的声学模型、为该用户本次跟读的语音建立预设类型的声学模型。在一优选实施例中，该
10 预设类型的声学模型为深度神经网络-隐马尔可夫声学模型，即 DNN-HMM 声学模型。在其他实施例中，该预设类型的声学模型也可以为其他的声学模型，例如为隐马尔可夫声学模型等。

在一具体的实例中，以 DNN-HMM 声学模型为例，其中，HMM 用来描述语音信号的动态变化，利用 DNN 的每个输出节点来估计连续密度 HMM
15 的某个状态的后验概率，即可得到 DNN-HMM 模型。本次播报的随机码的语音及该用户本次跟读的语音都是一连串的音节，若要辨识成的文字，则是一连串的字符。本实施例在建立 DNN-HMM 声学模型时，基于预定的字符语音库，通过全局字符声学自适应训练得到本次播报的随机码的语音的 DNN-HMM 声学模型、该用户本次跟读的语音的 DNN-HMM 声学模型。

20 强制整体对齐步骤，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算该对齐后的两声学模型相同的概率；

其中，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐（Force Alignment）操作，相比于传统的采取逐字对
25 比的方法，本实施例能够大大降低计算量，有利提高身份识别的效率。

其中，预定算法在一实施例中为前验后验概率算法，在其他实施例中，还可以是相似度算法，例如该相似度算法为计算对齐后的两声学模型中字符的编辑距离，编辑距离越小则对齐后的两声学模型相同的概率越大；该相似度算法还可以是最长公共子序列算法，若得到的最长公共子序列均与对齐后的两声学模型中字符的长度相差越小，则对齐后的两声学模型相同的概率越大。

身份验证步骤，若该对齐后的两声学模型相同的概率大于预设第一阈值，则提取该用户本次跟读的语音的声纹特征向量，获取该用户在注册成功后预存的标准声纹特征向量，并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离，以对该用户进行身份验证。

本实施例中，若该对齐后的两声学模型相同的概率大于预设第一阈值，例如预设第一阈值为 0.985，则认为用户本次跟读的字符与本次播报的随机码一致。由于播报的是随机码，因此可以有效防止了用户预先准备的合成音进行欺诈，提升身份识别的安全性。

在一实施例中，提取该用户本次跟读的语音的声纹特征向量的步骤包括：对该用户本次跟读的语音进行预加重及加窗处理，对每一个加窗进行傅立叶变换得到对应的频谱，将所述频谱输入梅尔滤波器以输出得到梅尔频谱；在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC，基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

其中，对该用户本次跟读的语音进行分帧，然后对分帧后的语音数据进行预加重处理，预加重处理实际是高通滤波处理，滤除低频数据，使得该语音数据中的高频特性更加突显，具体地，高通滤波的传递函数为：

$H(Z)=1-\alpha Z^{-1}$ ，其中， Z 为语音数据， α 为常量系数，优选地， α 的取值为 0.97；由于语音在分帧之后在一定程度上背离原始语音，因此，需要对该语音数据进行加窗处理。

本实施例中，在梅尔频谱上进行倒谱分析例如为取对数、做逆变换，逆变换一般是通过 DCT 离散余弦变换来实现，取 DCT 后的第 2 个到第 13 个系数作为梅尔频率倒谱系数 MFCC。梅尔频率倒谱系数 MFCC 即为这帧语音数据的声纹特征，将每帧的梅尔频率倒谱系数 MFCC 组成特征数据矩阵，
5 该特征数据矩阵即为该用户本次跟读的语音的声纹特征向量。

本实施例取语音数据的梅尔频率倒谱系数 MFCC 组成对应的声纹特征向量，由于其比用于正常的对数倒频谱中的线性间隔的频带更能近似人类的听觉系统，因此能够提高身份验证的准确性。

在一实施例中，计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离为计算两者的余弦距离，包括：

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \times \vec{y}}{|\vec{x}| |\vec{y}|}$$

，其中，所述 $\vec{x}$ 为标准声纹特征向量，所述 $\vec{y}$ 为该用户本次跟读的语音的声纹特征向量。

若余弦距离小于或者等于预设的距离阈值，则身份验证通过；若余弦距离大于预设的距离阈值，则身份验证不通过。

15 在一实施例中，在用户注册成功后预存的标准声纹特征向量，该注册声纹的步骤包括：

在互动式语音应答 IVR 场景下用户进行声纹注册时，播报第二预设位数的随机码供用户跟读预设次，在每次跟读后分别为播报的随机码及用户跟读的语音建立所述预设类型的声学模型；

20 分别将每次播报的随机码的声学模型及对应的用户跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算对齐后的两声学模型相同的概率；

若对齐后的两声学模型相同的概率均大于预设第二阈值，则提取每次用户跟读的语音的声纹特征向量，计算两两声纹特征向量的距离，以分析每次

跟读的用户是否为同一用户；

若是，则以该声纹特征向量作为该用户的标准声纹特征向量进行存储；

若否，则提示用户重新录入，再次进行注册声纹的步骤。

其中，在互动式语音应答 IVR 场景下，用户请求注册时发送身份识别码，

5 例如身份证号，在接收到用户的请求后，生成第二预设位数的随机码并采用语音合成技术以语音形式播报该随机码，引导用户进行跟读预设次（例如 3 次），该第二预设位数例如为 8 位。

在用户跟读后，为每次播报的随机码的语音建立预设类型的声学模型、为该用户每次跟读的语音建立预设类型的声学模型。在一优选实施例中，该
10 预设类型的声学模型为深度神经网络-隐马尔可夫声学模型，即 DNN-HMM 声学模型。在其他实施例中，该预设类型的声学模型也可以为其他的声学模型，例如为隐马尔可夫声学模型等。具体的实例可以参考上述的实施例，此处不再赘述。

在一具体的实例中，以 DNN-HMM 声学模型为例，其中，HMM 用来描
15 述语音信号的动态变化，利用 DNN 的每个输出节点来估计连续密度 HMM 的某个状态的后验概率，即可得到 DNN-HMM 模型。每次播报的随机码的语音及该用户跟读的语音都是一连串的音节，若要辨识成的文字，则是一连串的字符。本实施例在建立 DNN-HMM 声学模型时，基于预定的字符语音库，通过全局字符声学自适应训练得到播报的随机码的语音的 DNN-HMM
20 声学模型、该用户跟读的语音的 DNN-HMM 声学模型。

其中，将每次播报的随机码的声学模型及该用户跟读的语音的声学模型进行强制整体对齐（Force Alignment）操作，相比于传统的采取逐字对比的方法，本实施例能够大大降低计算量，有利提高身份识别的效率。

其中，预定算法在一实施例中为前验后验概率算法，在其他实施例中，
25 还可以是相似度算法，具体的实例可以参考上述的实施例，此处不再赘述。

本实施例中，若对齐后的两声学模型相同的概率均大于预设第二阈值，例如预设第二阈值为 0.985，则认为用户每次跟读的字符与所播报的随机码一致。由于播报的是随机码，因此可以有效防止了用户预先准备的合成音进行欺诈，提升身份识别的安全性。

5 在一实施例中，提取每次用户跟读的语音的声纹特征向量的步骤与上述实施例的提取语音的声纹特征向量的方法基本相同，此处不再赘述。

在一实施例中，计算两两声纹特征向量的距离的步骤，与上述计算余弦距离的步骤基本相同，此处不再赘述。

若余弦距离小于或者等于预设的距离阈值，则每次跟读的用户为同一用户，此时以该声纹特征向量作为该用户的标准声纹特征向量进行存储；若余弦距离大于预设的距离阈值，则每次跟读的用户不为同一用户，提示用户重新注册。

与现有技术相比，本申请在互动式语音应答 IVR 场景下进行身份识别时，利用随机码供用户跟读能够有效防止了预先准备的合成音进行欺诈，将随机码与声纹识别结合，实现了对用户身份的双重验证，能够准确确认用户身份，提高互动式语音应答 IVR 场景下身份验证的安全性，此外，对播报的随机码的声学模型及该用户跟读的语音的声学模型进行强制整体对齐操作，能够降低计算量，提高了身份识别效率。

如图 2 所示，图 2 为本申请身份验证的方法一实施例的流程示意图，该身份验证的方法包括以下步骤：

步骤 S1，在互动式语音应答 IVR 场景下用户办理业务时，播报第一预设位数的随机码供该用户跟读，并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型；

在互动式语音应答 IVR 场景下，用户请求办理业务时发送身份识别码，例如身份证号，在接收到用户的请求后，分析用户所办理的业务是否需要进

一步的身份验证，且根据用户的身份识别码分析该用户是否已注册有声纹，若需要进一步的身份验证且该用户已注册有声纹，则生成第一预设位数的随机码并采用语音合成技术以语音形式播报该随机码，引导用户进行跟读，该第一预设位数例如为 8 位。

5 在用户跟读后，为本次播报的随机码的语音建立预设类型的声学模型、为该用户本次跟读的语音建立预设类型的声学模型。在一优选实施例中，该预设类型的声学模型为深度神经网络-隐马尔可夫声学模型，即 DNN-HMM 声学模型。在其他实施例中，该预设类型的声学模型也可以为其他的声学模型，例如为隐马尔可夫声学模型等。

10 在一具体的实例中，以 DNN-HMM 声学模型为例，其中，HMM 用来描述语音信号的动态变化，利用 DNN 的每个输出节点来估计连续密度 HMM 的某个状态的后验概率，即可得到 DNN-HMM 模型。本次播报的随机码的语音及该用户本次跟读的语音都是一连串的音节，若要辨识成的文字，则是一连串的字符。本实施例在建立 DNN-HMM 声学模型时，基于预定的字符
15 语音库，通过全局字符声学自适应训练得到本次播报的随机码的语音的 DNN-HMM 声学模型、该用户本次跟读的语音的 DNN-HMM 声学模型。

步骤 S2，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算该对齐后的两声学模型相同的概率；

20 其中，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐（Force Alignment）操作，相比于传统的采取逐字对比的方法，本实施例能够大大降低计算量，有利提高身份识别的效率。

其中，预定算法在一实施例中为前验后验概率算法，在其他实施例中，还可以是相似度算法，例如该相似度算法为计算对齐后的两声学模型中字符
25 的编辑距离，编辑距离越小则对齐后的两声学模型相同的概率越大；该相似

度算法还可以是最长公共子序列算法，若得到的最长公共子序列均与对齐后的两声学模型中字符的长度相差越小，则对齐后的两声学模型相同的概率越大。

步骤 S3，若该对齐后的两声学模型相同的概率大于预设第一阈值，则
5 提取该用户本次跟读的语音的声纹特征向量，获取该用户在注册成功后预存的标准声纹特征向量，并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离，以对该用户进行身份验证。

本实施例中，若该对齐后的两声学模型相同的概率大于预设第一阈值，例如预设第一阈值为 0.985，则认为用户本次跟读的字符与本次播报的随机
10 码一致。由于播报的是随机码，因此可以有效防止了用户预先准备的合成音进行欺诈，提升身份识别的安全性。

在一实施例中，提取该用户本次跟读的语音的声纹特征向量的步骤包括：对该用户本次跟读的语音进行预加重及加窗处理，对每一个加窗进行傅立叶变换得到对应的频谱，将所述频谱输入梅尔滤波器以输出得到梅尔频
15 谱；在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC，基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

其中，对该用户本次跟读的语音进行分帧，然后对分帧后的语音数据进行预加重处理，预加重处理实际是高通滤波处理，滤除低频数据，使得该语音数据中的高频特性更加突显，具体地，高通滤波的传递函数为：
20 $H(Z)=1-\alpha Z^{-1}$ ，其中，Z 为语音数据， α 为常量系数，优选地， α 的取值为 0.97；由于语音在分帧之后在一定程度上背离原始语音，因此，需要对该语音数据进行加窗处理。

本实施例中，在梅尔频谱上进行倒谱分析例如为取对数、做逆变换，逆变换一般是通过 DCT 离散余弦变换来实现，取 DCT 后的第 2 个到第 13 个
25 系数作为梅尔频率倒谱系数 MFCC。梅尔频率倒谱系数 MFCC 即为这帧语

音数据的声纹特征，将每帧的梅尔频率倒谱系数 MFCC 组成特征数据矩阵，该特征数据矩阵即为该用户本次跟读的语音的声纹特征向量。

本实施例取语音数据的梅尔频率倒谱系数 MFCC 组成对应的声纹特征向量，由于其比用于正常的对数倒频谱中的线性间隔的频带更能近似人类的听觉系统，因此能够提高身份验证的准确性。

在一实施例中，计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离为计算两者的余弦距离，包括：

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \times \vec{y}}{|\vec{x}| |\vec{y}|}, \text{ 其中, 所述 } \vec{x} \text{ 为标准声纹特征向量, 所述 } \vec{y} \text{ 为该用户本}$$

次跟读的语音的声纹特征向量。

若余弦距离小于或者等于预设的距离阈值，则身份验证通过；若余弦距离大于预设的距离阈值，则身份验证不通过。

在一实施例中，在用户注册成功后预存的标准声纹特征向量，该注册声纹的步骤包括：

在互动式语音应答 IVR 场景下用户进行声纹注册时，播报第二预设位数的随机码供用户跟读预设次，在每次跟读后分别为播报的随机码及用户跟读的语音建立所述预设类型的声学模型；

分别将每次播报的随机码的声学模型及对应的用户跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算对齐后的两声学模型相同的概率；

若对齐后的两声学模型相同的概率均大于预设第二阈值，则提取每次用户跟读的语音的声纹特征向量，计算两两声纹特征向量的距离，以分析每次跟读的用户是否为同一用户；

若是，则以该声纹特征向量作为该用户的标准声纹特征向量进行存储；

若否，则提示用户重新录入，再次进行注册声纹的步骤。

其中,在互动式语音应答 IVR 场景下,用户请求注册时发送身份识别码,例如身份证号,在接收到用户的请求后,生成第二预设位数的随机码并采用语音合成技术以语音形式播报该随机码,引导用户进行跟读预设次(例如 3 次),该第二预设位数例如为 8 位。

5 在用户跟读后,为每次播报的随机码的语音建立预设类型的声学模型、为该用户每次跟读的语音建立预设类型的声学模型。在一优选实施例中,该预设类型的声学模型为深度神经网络-隐马尔可夫声学模型,即 DNN-HMM 声学模型。在其他实施例中,该预设类型的声学模型也可以为其他的声学模型,例如为隐马尔可夫声学模型等。具体的实例可以参考上述的实施例,此
10 处不再赘述。

在一具体的实例中,以 DNN-HMM 声学模型为例,其中,HMM 用来描述语音信号的动态变化,利用 DNN 的每个输出节点来估计连续密度 HMM 的某个状态的后验概率,即可得到 DNN-HMM 模型。每次播报的随机码的语音及该用户跟读的语音都是一连串的音节,若要辨识成的文字,则是一连
15 串字符。本实施例在建立 DNN-HMM 声学模型时,基于预定的字符语音库,通过全局字符声学自适应训练得到播报的随机码的语音的 DNN-HMM 声学模型、该用户跟读的语音的 DNN-HMM 声学模型。

其中,将每次播报的随机码的声学模型及该用户跟读的语音的声学模型进行强制整体对齐(Force Alignment)操作,相比于传统的采取逐字对比的方法,本实施例能够大大降低计算量,有利提高身份识别的效率。
20 方法,本实施例能够大大降低计算量,有利提高身份识别的效率。

其中,预定算法在一实施例中为前验后验概率算法,在其他实施例中,还可以是相似度算法,具体的实例可以参考上述的实施例,此处不再赘述。

本实施例中,若对齐后的两声学模型相同的概率均大于预设第二阈值,例如预设第二阈值为 0.985,则认为用户每次跟读的字符与所播报的随机码
25 一致。由于播报的是随机码,因此可以有效防止了用户预先准备的合成音进

行欺诈，提升身份识别的安全性。

在一实施例中，提取每次用户跟读的语音的声纹特征向量的步骤与上述实施例的提取语音的声纹特征向量的方法基本相同，此处不再赘述。

在一实施例中，计算两两声纹特征向量的距离的步骤，与上述计算余弦距离的步骤基本相同，此处不再赘述。

若余弦距离小于或者等于预设的距离阈值，则每次跟读的用户为同一用户，此时以该声纹特征向量作为该用户的标准声纹特征向量进行存储；若余弦距离大于预设的距离阈值，则每次跟读的用户不为同一用户，提示用户重新注册。

10 本申请还提供一种计算机可读存储介质，所述计算机可读存储介质上存储有处理系统，所述处理系统被处理器执行时实现上述的身份验证的方法的步骤。

上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述
15 实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。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（如 ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端设备（可以是手机，计算机，服务器，空调器，或者网络设备等）执行本申请各个实施例所述的方法。
20

以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

1.一种电子装置，其特征在于，所述电子装置包括存储器及与所述存储器连接的处理器，所述存储器中存储有可在所述处理器上运行的处理系统，所述处理系统被所述处理器执行时实现如下步骤：

5 声学模型建立步骤，在互动式语音应答 IVR 场景下用户办理业务时，播报第一预设位数的随机码供该用户跟读，并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型；

 强制整体对齐步骤，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算该对齐后的两
10 声学模型相同的概率；

 身份验证步骤，若该对齐后的两声学模型相同的概率大于预设第一阈值，则提取该用户本次跟读的语音的声纹特征向量，获取该用户在注册成功后预存的标准声纹特征向量，并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离，以对该用户进行身份验证。

15 2.根据权利要求 1 所述的电子装置，其特征在于，所述处理系统被所述处理器执行时，还实现如下步骤：

 在互动式语音应答 IVR 场景下用户进行声纹注册时，播报第二预设位数的随机码供用户跟读预设次，在每次跟读后分别为播报的随机码及用户跟读的语音建立所述预设类型的声学模型；

20 分别将每次播报的随机码的声学模型及对应的用户跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算对齐后的两声学模型相同的概率；

 若对齐后的两声学模型相同的概率均大于预设第二阈值，则提取每次用户跟读的语音的声纹特征向量，计算两两声纹特征向量的距离，以分析每次
25 跟读的用户是否为同一用户；

若是，则以该声纹特征向量作为该用户的标准声纹特征向量进行存储。

3.根据权利要求 1 所述的电子装置，其特征在于，所述预设类型的声学模型为深度神经网络-隐马尔可夫模型。

4.根据权利要求 2 所述的电子装置，其特征在于，所述预设类型的声学模型为深度神经网络-隐马尔可夫模型。

5.根据权利要求 1 所述的电子装置，其特征在于，所述提取该用户本次跟读的语音的声纹特征向量的步骤包括：

对该用户本次跟读的语音进行预加重及加窗处理，对每一个加窗进行傅立叶变换得到对应的频谱，将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC，基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

6.根据权利要求 2 所述的电子装置，其特征在于，所述提取该用户本次跟读的语音的声纹特征向量的步骤包括：

对该用户本次跟读的语音进行预加重及加窗处理，对每一个加窗进行傅立叶变换得到对应的频谱，将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC，基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

7.根据权利要求 1 或 2 所述的电子装置，其特征在于，所述计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离的步骤包括：

$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \times \vec{y}}{|\vec{x}| |\vec{y}|}$ ，其中，所述 $\vec{x}$ 为标准声纹特征向量，所述 $\vec{y}$ 为该用户本次跟读的语音的声纹特征向量。

8.一种身份验证的方法，其特征在于，所述身份验证的方法包括：

S1, 在互动式语音应答 IVR 场景下用户办理业务时, 播报第一预设位数的随机码供该用户跟读, 并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型;

5 S2, 将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作, 利用预定算法计算该对齐后的两声学模型相同的概率;

S3, 若该对齐后的两声学模型相同的概率大于预设第一阈值, 则提取该用户本次跟读的语音的声纹特征向量, 获取该用户在注册成功后预存的标准声纹特征向量, 并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离, 以对该用户进行身份验证。

9. 根据权利要求 8 所述的身份验证的方法, 其特征在于, 所述步骤 S1 之前, 还包括:

S01, 在互动式语音应答 IVR 场景下用户进行声纹注册时, 播报第二预设位数的随机码供用户跟读预设次, 在每次跟读后分别为播报的随机码及用户跟读的语音建立所述预设类型的声学模型;

S02, 分别将每次播报的随机码的声学模型及对应的用户跟读的语音的声学模型进行强制整体对齐操作, 利用预定算法计算对齐后的两声学模型相同的概率;

20 S03, 若对齐后的两声学模型相同的概率均大于预设第二阈值, 则提取每次用户跟读的语音的声纹特征向量, 计算两两声纹特征向量的距离, 以分析每次跟读的用户是否为同一用户;

S04, 若是, 则以该声纹特征向量作为该用户的标准声纹特征向量进行存储。

25 10. 根据权利要求 8 所述的身份验证的方法, 其特征在于, 所述预设类型的声学模型为深度神经网络-隐马尔可夫模型。

11.根据权利要求 9 所述的身份验证的方法,其特征在于,所述预设类型的声学模型为深度神经网络-隐马尔可夫模型。

12.根据权利要求 8 所述的身份验证的方法,其特征在于,所述提取该用户本次跟读的语音的声纹特征向量的步骤包括:

5 对该用户本次跟读的语音进行预加重及加窗处理,对每一个加窗进行傅立叶变换得到对应的频谱,将所述频谱输入梅尔滤波器以输出得到梅尔频谱;

在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC,基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

10 13.根据权利要求 9 所述的身份验证的方法,其特征在于,所述提取该用户本次跟读的语音的声纹特征向量的步骤包括:

对该用户本次跟读的语音进行预加重及加窗处理,对每一个加窗进行傅立叶变换得到对应的频谱,将所述频谱输入梅尔滤波器以输出得到梅尔频谱;

15 在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC,基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

14.根据权利要求 8 或 9 所述的身份验证的方法,其特征在于,所述计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离的步骤包括:

20 $\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \times \vec{y}}{|\vec{x}| |\vec{y}|}$, 其中, 所述 $\vec{x}$ 为标准声纹特征向量, 所述 $\vec{y}$ 为该用户本次跟读的语音的声纹特征向量。

15.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有处理系统,所述处理系统被处理器执行时实现步骤:

声学模型建立步骤,在互动式语音应答 IVR 场景下用户办理业务时,播

报第一预设位数的随机码供该用户跟读，并在跟读后分别为本次播报的随机码及该用户本次跟读的语音建立预设类型的声学模型；

5 强制整体对齐步骤，将本次播报的随机码的声学模型及该用户本次跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算该对齐后的两声学模型相同的概率；

身份验证步骤，若该对齐后的两声学模型相同的概率大于预设第一阈值，则提取该用户本次跟读的语音的声纹特征向量，获取该用户在注册成功后预存的标准声纹特征向量，并计算该用户本次跟读的语音的声纹特征向量及该标准声纹特征向量的距离，以对该用户进行身份验证。

10 16.根据权利要求 15 所述的计算机可读存储介质，其特征在于，所述处理系统被所述处理器执行时，还实现如下步骤：

在交互式语音应答 IVR 场景下用户进行声纹注册时，播报第二预设位数的随机码供用户跟读预设次，在每次跟读后分别为播报的随机码及用户跟读的语音建立所述预设类型的声学模型；

15 分别将每次播报的随机码的声学模型及对应的用户跟读的语音的声学模型进行强制整体对齐操作，利用预定算法计算对齐后的两声学模型相同的概率；

20 若对齐后的两声学模型相同的概率均大于预设第二阈值，则提取每次用户跟读的语音的声纹特征向量，计算两两声纹特征向量的距离，以分析每次跟读的用户是否为同一用户；

若是，则以该声纹特征向量作为该用户的标准声纹特征向量进行存储。

17.根据权利要求 15 所述的计算机可读存储介质，其特征在于，所述预设类型的声学模型为深度神经网络-隐马尔可夫模型。

25 18.根据权利要求 16 所述的计算机可读存储介质，其特征在于，所述预设类型的声学模型为深度神经网络-隐马尔可夫模型。

19.根据权利要求 15 所述的计算机可读存储介质，其特征在于，所述提取该用户本次跟读的语音的声纹特征向量的步骤包括：

5 对该用户本次跟读的语音进行预加重及加窗处理，对每一个加窗进行傅立叶变换得到对应的频谱，将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC，基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

20.根据权利要求 16 所述的计算机可读存储介质，其特征在于，所述提取该用户本次跟读的语音的声纹特征向量的步骤包括：

10 对该用户本次跟读的语音进行预加重及加窗处理，对每一个加窗进行傅立叶变换得到对应的频谱，将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

在梅尔频谱上进行倒谱分析以获得梅尔频率倒谱系数 MFCC，基于所述梅尔频率倒谱系数 MFCC 组成该用户本次跟读的语音的声纹特征向量。

15

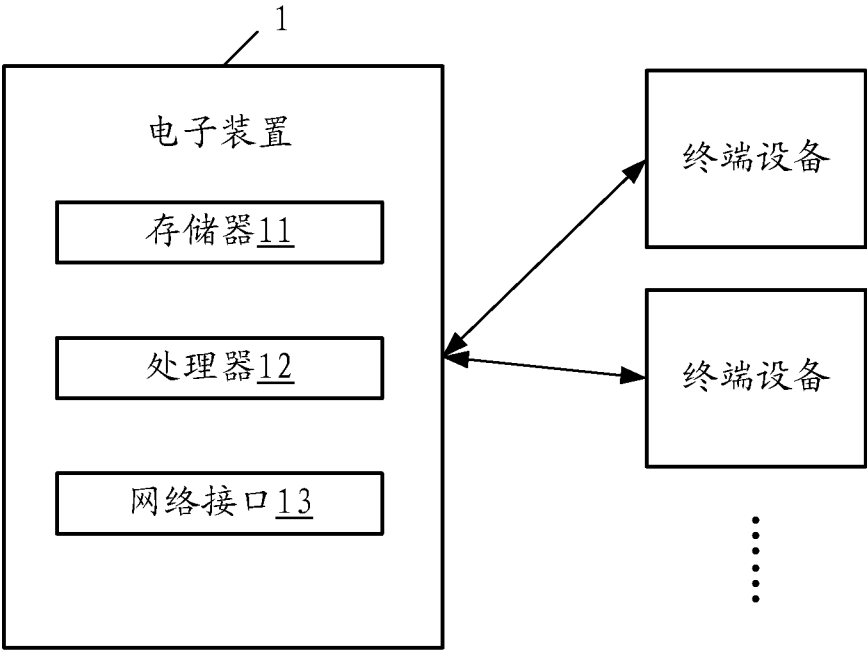


图 1

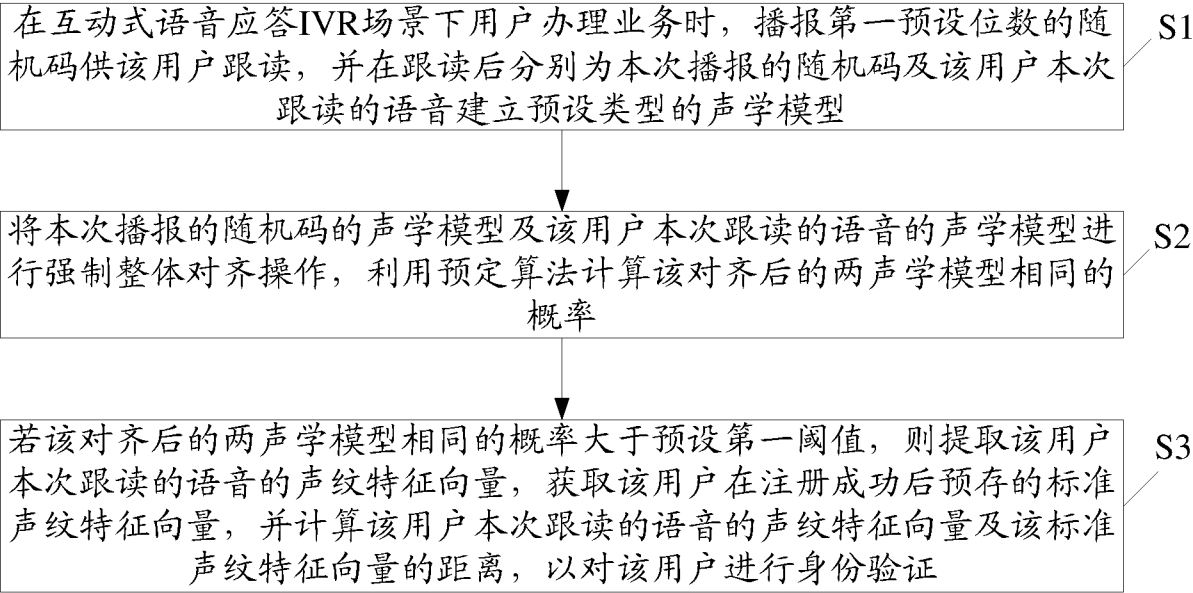


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/102208

A. CLASSIFICATION OF SUBJECT MATTER G10L 17/24(2013.01)i; G10L 17/08(2013.01)i; G10L 17/04(2013.01)i; G10L 17/16(2013.01)i; G10L 17/18(2013.01)i; G10L 25/24(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) 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) CPRSABS, CNTXT, DWPI, SIPOABS, USTXT, CNKI, GOOGLE SCHOLAR: 语音, 声纹, 随机, 对齐, 验证, 识别, 特征, 向量, 距离, sound, voice, vocal print, random, stochastic, alignment, validate, recognize, feature, vector, distance														
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <thead> <tr> <th style="width: 10%;">Category*</th> <th style="width: 70%;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="width: 20%;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">X</td> <td>CN 103986725 A (AGRICULTURAL BANK OF CHINA) 13 August 2014 (2014-08-13) description, paragraphs 35-78</td> <td style="text-align: center;">1-20</td> </tr> <tr> <td style="text-align: center;">X</td> <td>CN 107517207 A (PING AN TECHNOLOGY SHENZHEN CO., LTD.) 26 December 2017 (2017-12-26) description, paragraphs 52-175</td> <td style="text-align: center;">1-20</td> </tr> <tr> <td style="text-align: center;">A</td> <td>CN 103680497 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 26 March 2014 (2014-03-26) entire document</td> <td style="text-align: center;">1-20</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	CN 103986725 A (AGRICULTURAL BANK OF CHINA) 13 August 2014 (2014-08-13) description, paragraphs 35-78	1-20	X	CN 107517207 A (PING AN TECHNOLOGY SHENZHEN CO., LTD.) 26 December 2017 (2017-12-26) description, paragraphs 52-175	1-20	A	CN 103680497 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 26 March 2014 (2014-03-26) entire document	1-20
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.												
X	CN 103986725 A (AGRICULTURAL BANK OF CHINA) 13 August 2014 (2014-08-13) description, paragraphs 35-78	1-20												
X	CN 107517207 A (PING AN TECHNOLOGY SHENZHEN CO., LTD.) 26 December 2017 (2017-12-26) description, paragraphs 52-175	1-20												
A	CN 103680497 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 26 March 2014 (2014-03-26) entire document	1-20												
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.														
<table style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> * 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 </td> <td style="width: 50%; vertical-align: top;"> "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 </td> </tr> </table>			* 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										
* 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 05 January 2019		Date of mailing of the international search report 15 January 2019												
Name and mailing address of the ISA/CN State Intellectual Property Office of the P. R. China (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/CN2018/102208

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	103986725	A	13 August 2014	None			
CN	107517207	A	26 December 2017	TW	201833810	A	16 September 2018
				WO	2018166187	A1	20 September 2018
				CN	107068154	A	18 August 2017
				WO	2018166112	A1	20 September 2018
CN	103680497	A	26 March 2014	CN	103680497	B	15 March 2017

国际检索报告

国际申请号

PCT/CN2018/102208

A. 主题的分类 G10L 17/24(2013.01)i; G10L 17/08(2013.01)i; G10L 17/04(2013.01)i; G10L 17/16(2013.01)i; G10L 17/18(2013.01)i; G10L 25/24(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) 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)) CPRSABS, CNTXT, DWPI, SIPOABS, USTXT, CNKI, GOOGLE SCHOLAR: 语音, 声纹, 随机, 对齐, 验证, 识别, 特征, 向量, 距离, sound, voice, vocal print, random, stochastic, alignment, validate, recognize, feature, vector, distance		
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
X	CN 103986725 A (中国农业银行股份有限公司) 2014年 8月 13日 (2014 - 08 - 13) 说明书第35-78段	1-20
X	CN 107517207 A (平安科技深圳有限公司) 2017年 12月 26日 (2017 - 12 - 26) 说明书第52-175段	1-20
A	CN 103680497 A (百度在线网络技术北京有限公司) 2014年 3月 26日 (2014 - 03 - 26) 全文	1-20
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2019年 1月 5日		国际检索报告邮寄日期 2019年 1月 15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 王雪莲 电话号码 62411747

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/102208

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	103986725	A	2014年 8月 13日	无			
CN	107517207	A	2017年 12月 26日	TW	201833810	A	2018年 9月 16日
				WO	2018166187	A1	2018年 9月 20日
				CN	107068154	A	2017年 8月 18日
				WO	2018166112	A1	2018年 9月 20日
CN	103680497	A	2014年 3月 26日	CN	103680497	B	2017年 3月 15日