

(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 4 月 16 日 (16.04.2020)



(10) 国际公布号
WO 2020/073518 A1

(51) 国际专利分类号:
G10L 17/00 (2013.01) *G10L 21/0208* (2013.01)
G10L 17/04 (2013.01)

(21) 国际申请号: PCT/CN2018/124401

(22) 国际申请日: 2018 年 12 月 27 日 (27.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811184693.9 2018 年 10 月 11 日 (11.10.2018) 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: 杨翹楚(YANG, Qiaochu); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
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: 深圳市明日今典知识产权代理事务所(普通合伙)(SHENZHEN MING-RIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM(GENERAL)); 中国广东省深圳市南山区南头街道智恒新兴产业园 E 区 01B 栋 405 室, Guangdong 518000 (CN)。

(54) Title: VOICEPRINT VERIFICATION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 声纹验证的方法、装置、计算机设备和存储介质

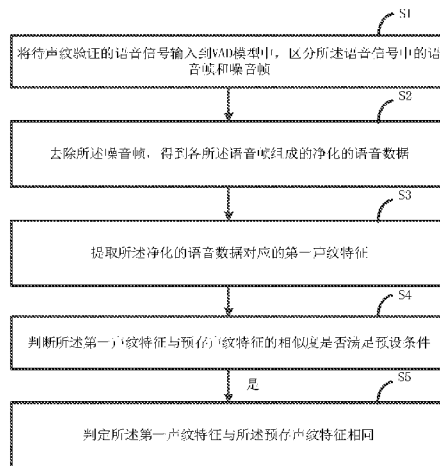


图 1

S1 Input into a VAD model a voice signal to be subject to voiceprint verification, and distinguish voice frames and noise frames in the voice signal
S2 Remove the noise frames to obtain purified voice data composed of each voice frame
S3 Extract a first voiceprint feature corresponding to the purified voice data
S4 Determine whether the similarity between the first voiceprint feature and a prestored voiceprint feature meets a preset condition
S5 Determine that the first voiceprint feature is the same as the pre-stored voiceprint feature
AA Yes

(57) Abstract: Disclosed is a voiceprint verification method, comprising: inputting into a VAD model a voice signal to be subject to voiceprint verification, and distinguishing voice frames and noise frames in the voice signal; removing the noise frames to obtain purified voice data composed of each voice frame; extracting a first voiceprint feature corresponding to the purified voice data; determining whether the similarity between the first voiceprint feature and a prestored voiceprint feature meets a preset condition; if yes, determining that the first voiceprint feature is the same as the pre-stored voiceprint feature, or otherwise, the first voiceprint feature is different from the pre-stored voiceprint feature.

(57) 摘要: 本申请揭示了声纹验证的方法, 包括: 将待声纹验证的语音信号输入到 VAD 模型中, 区分语音信号中的语音帧和噪音帧; 去除噪音帧, 得到各语音帧组成的净化的语音数据; 提取净化的语音数据对应的第一声纹特征; 判断第一声纹特征与预存声纹特征的相似度是否满足预设条件; 若满足则判定第一声纹特征与预存声纹特征相同, 否则不相同。

(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))。

声纹验证的方法、装置、计算机设备和存储介质

本申请要求于2018年10月11日提交中国专利局、申请号为2018111846939，发明名称为“声纹验证的方法、装置、计算机设备和存储介质”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

5 本申请涉及到声纹验证领域，特别是涉及到声纹验证的方法、装置、计算机设备和存储介质。

背景技术

10 目前，很多大型金融公司的业务范围涉及保险、银行、投资等多个业务范畴，而每个业务范畴通常都需要同客户进行沟通，且都需要进行反欺诈识别，因此，对客户的身

份验证及反欺诈识别也就成为保证业务安全的重要组成部分。在客户身份验证环节中，声纹验证由于其具有的实时性和易便性而被许多公司采用。发明人意识到在实际应用中，受说话人在身份注册或身份验证环节所处的环境因素影响，采集到的语音数据常常带有非来自说话人的背景噪音，这一因素成为影响声纹验证成功率的主要因素之一。

技术问题

15 本申请的主要目的为提供一种声纹验证的方法，旨在解决现有语音数据中的噪音对声纹验证效果产生不良影响的技术问题。

技术方案

本申请提出一种声纹验证的方法，包括：

20 将待声纹验证的语音信号输入到 VAD 模型中，区分所述语音信号中的语音帧和噪音帧；

去除所述噪音帧，得到各所述语音帧组成的净化的语音数据；

提取所述净化的语音数据对应的第一声纹特征；

判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件；

25 若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

本申请还提供了一种声纹验证的装置，包括：

区分模块，用于将待声纹验证的语音信号输入到 VAD 模型中，区分所述语音信号中的语音帧和噪音帧；

30 去除模块，用于去除所述噪音帧，得到各所述语音帧组成的净化的语音数据；

提取模块，用于提取所述净化的语音数据对应的第一声纹特征；

判断模块,用于判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件;

判定模块,用于若满足预设条件,则判定所述第一声纹特征与所述预存声纹特征相同,否则不相同。

- 5 本申请还提供了一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现上述方法的步骤。
- 本申请还提供了一种计算机非易失性可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现上述的方法的步骤。

有益效果

- 10 本申请通过识别语音信号中的噪音数据,并去除噪音数据得到净化的语音数据,然后依据净化后的语音数据进行声纹识别,提高声纹验证的准确性。本申请通过GMM-VAD模型,结合局部判决和全局判决,实现对噪音数据和语音数据的准确区分,以提高净化语音信号的程度,进一步提高声纹验证的准确性。
- 15 本申请基于GMM-UBM实现将各所述声纹特征向量分别映射为低维度的声纹鉴别向量l-vector,降低声纹特征提取过程中的计算成本,降低声纹验证的使用成本。本申请在声纹验证过程中通过与多人的预存数据进行比较分析,降低声纹验证的等错率,降低声纹验证的模型误差带来的声纹验证精准度的误差。

附图说明

- 图1 本申请一实施例的声纹验证的方法流程示意图;
- 20 图2 本申请一实施例的声纹验证的装置结构示意图;
- 图3 本申请一实施例的计算机设备内部结构示意图。

本发明的最佳实施方式

参照图1,本申请一实施例的一种声纹验证的方法,包括:

- 25 S1:将待声纹验证的语音信号输入到VAD模型中,区分所述语音信号中的语音帧和噪音帧。

- 本实施例的VAD模型,又称语音端点检测器,用于在噪声环境中检测是否存在人声的语音数据。VAD模型通过对输入的每一帧语音信号进行打分,即该帧语音信号是语音帧或噪音帧的概率,当语音帧的概率值大于预先设定的判决门限,则判定为语音帧,否则为噪音帧。VAD模型根据上述判决结果对语音帧和噪音帧进行区分,以便去除语音信号中的噪音帧。本实施例的判决门限采用了WebRTC源代码中默认的判决门限,该判决门限是WebRTC技术开发时通过分
- 30

析大量数据得来的,以提高区分的效果和准确度,并同时减少 VAD 模型的模型训练工作量。

S2: 去除所述噪音帧,得到各所述语音帧组成的净化的语音数据。

本实施例根据上述区分结果,将标记为噪音帧的数据剪切掉,将剩余的各所述语音帧按照原排布时间顺序依次连续排布,形成各所述语音帧组成的所述净化的语音数据。本申请其他实施例也可通过上述区分结果,筛选标记为语音帧的数据进行提取保存,将提取保存的各所述语音帧按照原排布时间顺序依次连续排布,形成各所述语音帧组成的所述净化的语音数据。本实施例通过将在身份注册或身份验证环节所处的环境的非来自说话人的背景噪音数据去除掉,降低语音信号中的噪音数据对声纹验证效果的影响,以提高声纹验证成功率。

S3: 提取所述净化的语音数据对应的第一声纹特征。

本实施例通过只分析净化的语音数据对应的第一声纹特征,以减少声纹验证中的计算量,并同时提高声纹验证的有效性、针对性以及时效性。

S4: 判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件。

本实施例的预设条件包括指定的预设阈值范围、或指定的排序等,可根据具体的应用场景进行自定义设定,以更广泛地满足个性化使用需求。

S5: 若满足,则判定第一声纹特征与预存声纹特征相同,否则不相同。

本实施例将判定所述第一声纹特征与所述预存声纹特征相同,则反馈验证通过的结果到客户端,否则,反馈验证失败的结果到客户端,以便客户端根据反馈结果进行进一步的应用操作。举例地,验证通过后控制智能门打开等。再举例地,验证失败指定次数后控制安全系统进行锁屏,以防犯罪分子进一步破坏电子银行系统。

进一步地,本实施例的 VAD 模型中包括傅里叶变换,高斯混合分布的 GMM-NOISE 和 GMM-SPEECH,步骤 S1,包括:

S100: 将所述语音信号输入到 VAD 模型中的傅里叶变换中,将所述语音信号从时域信号形式转变为频域信号形式。

本实施例通过 VAD 模型中的傅里叶变换将时域信号形式一一对应地转换成频域信号形式,进行分析各帧语音信号的属性,方便区分语音帧和噪音帧。

S101: 将频域信号形式的语音信号的每一帧数据分别输入到 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决,以区分语音信号中的语音帧和噪音帧。

本实施例优选基于混合高斯 GMM 的 VAD 模型,其对输入的每一帧频域信号

形式的语音信号在 6 个频段上进行能量提取, 作为该帧语音信号的特征向量, 并对噪音和语音在 6 个频段上分别进行高斯混合分布 GMM 建模, 每个频段上都有含有两个高斯分量的噪音 GMM-NOISE 和含有两个高斯分量的语音 GMM-SPEECH。上述 6 个频段根据 Webrtc 技术基于噪音和语音的频谱差异进行设置, 以便提高分析准确度以及与 Webrtc 技术的匹配性。本申请其他实施例的分析频段不一定必须是 6 个, 可根据实际需求进行设定。而且基于我国交流电标准是 220V、50Hz, 电源 50Hz 的干扰会混入采集语音信号的麦克风中, 采集到的干扰信号以及物理震动均会带来影响, 本实施例优选采集 80Hz 以上的语音信号, 以减少交流电的干扰, 而语音能达到的最高频率是 4kHz, 所以本实施例优选在 80Hz 至 4kHz 范围内的频谱波谷处划分界限。本实施例的 VAD 判决包括局部判决 (Local Decision) 和全局判决 (Global Decision)。

进一步地, 本实施例的步骤 S101, 包括:

S1010: 将频域信号形式的语音信号的各帧数据, 分别输入到 GMM-NOISE 和 GMM-SPEECH 中, 分别得到各帧数据的噪音帧概率 $P_{local}^{(x|noise)}$ 和语音帧概率 $P_{local}^{(x|speech)}$ 。

本实施例通过将预分析是语音帧还是噪音帧的语音信号各帧数据, 分别输入到 GMM-NOISE 和 GMM-SPEECH 中, 获取 GMM-NOISE 和 GMM-SPEECH 分别分析各帧数据的噪音帧概率值和语音帧概率值, 以便通过比较噪音帧概率值和语音帧概率值的大小, 以便确定是噪音帧还是语音帧。

S1011: 根据 $\log \text{likelihood ratio}_{local,i} = \log_2 \left(\frac{P_{local}^{(x|speech_i)}}{P_{local}^{(x|noise_i)}} \right), i = 1, 2, \dots, n$,

计算局部对数似然比。

本实施例优选基于混合高斯 GMM 的 VAD 模型, 其对输入的每一帧频域信号形式的语音信号, 在 6 个频段上进行能量提取, 作为该帧语音信号的特征向量, 所以本实施例中 n 取值为 6, 对每一帧进行判断的时候, 都会进行 6 次局部判决, 即在 6 个频段上分别进行局部判决, 只要有一次认为该帧为语音帧, 即保留这一帧。

S1012: 判断所述局部对数似然比是否高于局部门限值。

本实施例通过局部判决, 实现对语音帧和噪音帧的区分, 本实施例的局部判决在每个频段上做一次, 一共 6 次。似然比是反映真实性的一种指标, 属于

同时反映灵敏度和特异度的复合指标，提高概率估算准确度，本实施例在确保语音帧概率值大于噪音帧概率值的情况下，进一步通过比较局部对数似然比是否高于局部门限值，以确保判定为该语音信号为语音帧的准确性。

S1013: 若是，则判定局部对数似然比高于局部门限值的帧数据为语音帧。

5 本实施例的 GMM 的参数具有自适应更新能力，在每一帧语音信号被判断为语音帧或者噪声帧之后，会根据该帧的特征值来更新其对应模型的参数。例如，如果该帧被判断为语音帧，则 GMM-SPEECH 的期望值、标准差和高斯分量权重值就根据该帧的特征值进行一次更新，在越来越多的语音帧输入 GMM-SPEECH 之后，GMM-SPEECH 会越来越适应此通语音信号的说话人的声纹特征，给出的
10 分析结论会更加准确。

进一步地，本申请另一实施例的步骤 S1012 之后，包括：

S1014：若局部对数似然比不高于局部门限值，则根据
$$\text{sum log likelihood ratio} = \sum_{i=1}^n \text{weight}_i * \log \text{likelihood ratio}_{\text{local},i}$$
，计算全局对数似然比。

本实施例先进行局部判决，再进行全局判决，全局判决是基于局部判决结果的基础上进行各频段加权值的计算，提高区分语音帧和噪音帧的准确度。
15

S1015：判断所述全局对数似然比是否高于全局门限值。

本实施例的全局判决中将全局对数似然比与全局门限值相比，以进一步提高筛选语音帧的准确性。

S1016：若全局对数似然比高于全局门限值，则判定全局对数似然比高于
20 全局门限值的帧数据为语音帧。

本实施例可以先根据局部判决结果有语音存在，则不进行全局判决，以提高声纹验证的效率，且尽量将所有的语音帧都能识别到，以免语音失真。本申请其他实施例也可在局部判决结果有语音存在，再进行全局判决，以进一步核实和确认语音的存在，提高区分语音帧和噪音帧的准确度。

25 进一步地，本实施例的步骤 S3，包括：

S30：提取净化的语音数据中各语音帧分别对应的 MFCC 类型声纹特征。

本实施例提取 MFCC (Mel Frequency Cepstrum Coefficient, 梅尔频率倒谱系数) 类型声纹特征的过程如下：先采样和量化，将净化的语音数据的连续模拟语音信号以一定的采样周期采样，转化为离散信号，并根据一定的编码规

则将离散信号量化为数字信号；然后预加重，由于人体的生理特性，语音信号的高频成分往往被压抑，预加重的作用是补偿高频成分；接着分帧处理，由于语音信号的“瞬时平稳性”，在进行频谱分析时对一段语音信号进行分帧处理（一般为 10 至 30 毫秒一帧），然后以帧为单位进行特征提取；接着加窗处理，作用是减少帧起始和帧结束对应信号的不连续性问题，采用汉明窗进行加窗处理；接着对帧信号进行 DFT，将信号从时域转换到频域，然后再利用如下公式将信号从线性频谱域映射到梅尔频谱域： $f_{Mel} = 2595 \times \log_{10}(1 + \frac{f}{700})$ ；将转化后的帧信号输入到一组梅尔三角滤波器组，计算每个频段的滤波器输出的信号对数能量，得到一个对数能量序列；对上一步得到的对数能量序列做离散余弦变换（DCT, Discrete Cosine Transform）即可得到该帧语音信号的 MFCC 类型声纹特征。

S31：根据各 MFCC 类型声纹特征构建各语音帧分别对应的声纹特征向量。

MFCC 类型声纹特征具有非线性特征，使各频段上的分析结果更贴近人体发出的真实语音的特征，使声纹特征提取更加准确，提高声纹验证的效果。

S32：将各所述声纹特征向量分别映射为低维度的声纹鉴别向量 l-vector，以得到净化的语音数据中各所述语音帧分别对应的第一声纹特征。

本实施例基于 GMM-UBM (Gaussian Mixture Model-Universal Background Model, 高斯混合模型-背景模型) 实现将各所述声纹特征向量分别映射为低维度的声纹鉴别向量 l-vector，降低声纹特征提取过程中的计算成本，降低声纹验证的使用成本。本实施例的 GMM-UBM 的训练过程如下：B1：获取预设数量（例如，10 万个）的语音数据样本，每个语音数据样本对应一个声纹鉴别向量，每个语音数据样本可以采集自不同的人在不同环境中的语音形成，这样的语音数据样本用来训练能够表征一般语音特性的通用背景模型（GMM-UBM）；B2、分别对各个语音数据样本进行处理以提取出各个语音数据样本对应的预设类型声纹特征，并基于各个语音数据样本对应的预设类型声纹特征构建各个语音数据样本对应的声纹特征向量；B3、将构建出的所有预设类型声纹特征向量分为第一百分比的训练集和第二百分比的验证集，所述第一百分比和第二百分比小于或等于 100%；B4、利用训练集中的声纹特征向量对所述第二模型进行训练，并在训练完成之后利用验证集对训练的所述第二模型的准确率进行验证；B5、若准确率大于预设准确率（例如，98.5%），则模型训练结束，否则，

增加语音数据样本的数量，并基于增加后的语音数据样本重新执行上述步骤 B2、B3、B4、B5。

本实施例的声纹鉴别向量采用 l-vector 表达，i-vector 是一个向量，相对于高斯空间的维度来讲，i-vector 维度更低，便于降低计算成本，而提取低维度的 i-vector 的过程是通过下述计算公式，将低维度的向量 w 与一个转换矩阵 T 相乘映射到维度较高的高斯空间。l-vector 的提取包括如下步骤：将来自某位目标说话人的训练语音数据处理后，提取得到的预设类型声纹特征向量（例如，MFCC）输入到 GMM-VAD 模型，得到一个表征该段语音数据在各个高斯分量上的概率分布的高斯超向量；利用如下公式可以计算得到该段语音对应的较低维度的声纹鉴别向量 l-vector: $m_r = \mu + T\omega_r$ ，其中 ω_r 为代表该段语音的高斯超向量， μ 为所述第二模型的均值超向量， T 为将低维度的 l-vector， ω_r 映射到高维度的高斯空间的转换矩阵， T 的训练采用 EM 算法。

进一步地，本实施例的步骤 S4，包括：

S40：分别在预存的多个人的声纹特征数据中获取各自对应的预存声纹特征，其中，多个人的声纹特征数据中包括目标人的预存声纹特征。

本实施例将预存的包括目标人的多人的声纹特征数据，同时用于判断当前采集的语音信号的声纹特征是否与目标人的声纹特征相同，以提高判断准确性。

S41：分别计算各所述预存声纹特征与第一声纹特征之间的相似度值。

本实施例的相似度值表征了预存声纹特征与所述第一声纹特征之间的相似度，相似度值越大，则上述两者越相似。本实施例的相似度值的获取方法包括通过比较预存声纹特征与所述第一声纹特征之间的特征距离值得到，上述特征距离值包括余弦距离值、欧式距离值等。

S42：将各所述相似度值按照从大到小的顺序进行排序。

本实施例通过将各所述预存声纹特征与所述第一声纹特征之间的相似度值进行从大到小排序，以便更准确地分析第一声纹特征与各预存声纹特征的相似度分布状态，以便更准确地获得对第一声纹特征的验证。

S43：判断排序在前的预设数量的相似度值中，是否包括所述目标人的预存声纹特征对应的相似度值。

本实施例通过排序在前的预设数量的相似度值，包括目标人的预存声纹特征对应的相似度值，则判定第一声纹特征与预存的目标人的声纹特征相同，以

减小模型误差带来的识别等错率，等错率为“应验证通过时发生的验证未通过的频率，与应验证未通过时发生的验证通过的频率相等”。本实施例的预设数量的相似度值包括1个、2个或3个等，可根据使用需求进行自设定。

5 S44：若是，则判定所述第一声纹特征与预存声纹特征的相似度满足预设条件，否则不满足预设条件。

本申请其他实施例通过设定第一声纹特征与目标用户的预存声纹特征的距离阈值，实现有效的声纹验证。举例地，预设阈值为0.6，若计算第一声纹特征与目标用户的预存声纹特征的余弦距离小于或等于预设阈值，则确定第一声纹特征与目标用户的预存声纹特征相同，则验证通过；若计算第一声纹特征
10 与目标用户的预存声纹特征的余弦距离大于预设阈值，则确定第一声纹特征与目标用户的预存声纹特征不相同，则验证失败。

进一步地，本实施例的步骤S41，包括：

S410：分别通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$ ，计算各所述预存声纹特征与所述第一声纹特征之间的余弦距离值，其中，x代表各预存声纹鉴别向量，
15 y代表第一声纹特征的声纹鉴别向量。

本实施例通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$ ，表示各所述预存声纹特征与所述第一声纹特征之间的相似度，其中余弦距离的距离值越小，表明两声纹特征更接近或相同。

S411：将所述余弦距离值转换成所述相似度值，其中，最小的所述余弦距离值对应最大的相似度值。
20

本实施例可通过将余弦距离值按照携带指定反比系数的反比例公式，将余弦距离值转换成相似度值。

本实施例通过识别语音信号中的噪音数据，并去除噪音数据得到净化的语音数据，然后依据净化后的语音数据进行声纹识别，提高声纹验证的准确性。
25 本实施例通过 GMM-VAD 模型，结合局部判决和全局判决，实现对噪音数据和语音数据的准确区分，以提高净化语音信号的程度，进一步提高声纹验证的准确性。本实施例基于 GMM-UBM 实现将各所述声纹特征向量分别映射为低维度的声纹鉴别向量 l-vector，降低声纹特征提取过程中的计算成本，降低声纹验证的使用成本。本实施例在声纹验证过程中通过与多人的预存数据进行比较分

析,降低声纹验证的等错率,降低声纹验证的模型误差带来的声纹验证精准度的误差。

参照图2,本申请一实施例的一种声纹验证的装置,包括:

区分模块1,用于将待声纹验证的语音信号输入到VAD模型中,区分所述
5 语音信号中的语音帧和噪音帧。

本实施例的VAD模型,又称语音端点检测器,用于在噪声环境中检测是否存在人声的语音数据。VAD模型通过对输入的每一帧语音信号进行打分,即该帧语音信号是语音帧或噪音帧的概率,当语音帧的概率值大于预先设定的判决门限,则判定为语音帧,否则为噪音帧。VAD模型根据上述判决结果对语音帧
10 和噪音帧进行区分,以便去除语音信号中的噪音帧。本实施例的判决门限采用了Webrtc源代码中默认的判决门限,该判决门限是Webrtc技术开发时通过分析大量数据得来的,以提高区分的效果和准确度,并同时减少VAD模型的模型训练工作量。

去除模块2,用于去除噪音帧,得到各所述语音帧组成的净化的语音数据。

15 本实施例根据上述区分结果,将标记为噪音帧的数据剪切掉,将剩余的各所述语音帧按照原排布时间顺序依次连续排布,形成各所述语音帧组成的所述净化的语音数据。本申请其他实施例也可通过上述区分结果,筛选标记为语音帧的数据进行提取保存,将提取保存的各所述语音帧按照原排布时间顺序依次连续排布,形成各所述语音帧组成的所述净化的语音数据。本实施例通过将在
20 身份注册或身份验证环节所处的环境的非来自说话人的背景噪音数据去除掉,降低语音信号中的噪音数据对声纹验证效果的影响,以提高声纹验证成功率。

提取模块3,用于提取所述净化的语音数据对应的第一声纹特征。

本实施例通过只分析净化的语音数据对应的第一声纹特征,以减少声纹验证中的计算量,并同时提高声纹验证的有效性、针对性以及时效性。

25 判断模块4,用于判断第一声纹特征与预存声纹特征的相似度是否满足预设条件。

本实施例的预设条件包括指定的预设阈值范围、或指定的排序等,可根据具体的应用场景进行自定义设定,以更广泛地满足个性化使用需求。

判定模块5,用于若满足预设条件,则判定所述第一声纹特征与所述预存
30 声纹特征相同,否则不相同。

本实施例将判定所述第一声纹特征与所述预存声纹特征相同,则反馈验证

通过的结果到客户端，否则，反馈验证失败的结果到客户端，以便客户端根据反馈结果进行进一步的应用操作。举例地，验证通过后控制智能门打开等。再举例地，验证失败指定次数后控制安全系统进行锁屏，以防犯罪分子进一步破坏电子银行系统。

5 进一步地，本实施例的 VAD 模型中包括傅里叶变换，高斯混合分布的 GMM-NOISE 和 GMM-SPEECH，上述区分模块 1，包括：

转变单元，用于将所述语音信号输入到 VAD 模型中的傅里叶变换中，将所述语音信号从时域信号形式转变为频域信号形式。

10 本实施例通过 VAD 模型中的傅里叶变换将时域信号形式一一对应地转换成频域信号形式，进行分析各帧语音信号的属性，方便区分语音帧和噪音帧。

区分单元，用于将频域信号形式的语音信号的每一帧数据分别输入到 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧。

15 本实施例优选基于混合高斯 GMM 的 VAD 模型，其对输入的每一帧频域信号形式的语音信号在 6 个频段上进行能量提取，作为该帧语音信号的特征向量，并对噪音和语音在 6 个频段上分别进行高斯混合分布 GMM 建模，每个频段上都有含有两个高斯分量的噪音 GMM-NOISE 和含有两个高斯分量的语音 GMM-SPEECH。上述 6 个频段根据 WebRTC 技术基于噪音和语音的频谱差异进行设置，以便提高分析准确度以及与 WebRTC 技术的匹配性。本申请其他实施例的分析频段不一定必须是 6 个，可根据实际需求进行设定。而且本实施例基于我国交流电标准是 220V、50Hz，电源 50Hz 的干扰会混入采集语音信号的麦克风中，采集到的干扰信号以及物理震动均会带来影响，本实施例优选采集 80Hz 以上的语音信号，以减少交流电的干扰，而语音能达到的最高频率是 4kHz，所以本实施例优选在 80Hz 至 4kHz 范围内的频谱波谷处划分界限。本实施例的
25 VAD 判决包括局部判决（Local Decision）和全局判决（Global Decision）。

进一步地，本实施例的区分单元，包括：

输入子单元，用于将频域信号形式的语音信号各帧数据，分别输入到 GMM-NOISE 和 GMM-SPEECH 中，分别得到各帧数据的噪音帧概率 $P_{local}^{(x|noise)}$ 和语音帧概率 $P_{local}^{(x|speech)}$ 。

30 本实施例通过将预分析是语音帧还是噪音帧的语音信号各帧数据，分别输入到 GMM-NOISE 和 GMM-SPEECH 中，获取 GMM-NOISE 和 GMM-SPEECH 分别分析

各帧数据的噪音帧概率值和语音帧概率值,以便通过比较噪音帧概率值和语音帧概率值的大小,以便确定是噪音帧还是语音帧。

第一计算子单元,用于根据 $\log \text{likelihood ratio}_{local,i} = \log_2 \left(\frac{P_{local}^{(x|speech_i)}}{P_{local}^{(x|noise_i)}} \right), i=1,2,\dots,n$,

计算局部对数似然比。

5 本实施例优选基于混合高斯 GMM 的 VAD 模型,其对输入的每一帧频域信号形式的语音信号,在 6 个频段上进行能量提取,作为该帧语音信号的特征向量,所以本实施例中 n 取值为 6,对每一帧进行判断的时候,都会进行 6 次局部判决,即在 6 个频段上分别进行局部判决,只要有一次认为该帧为语音帧,即保留这一帧。

10 第一判断子单元,用于判断所述局部对数似然比是否高于局部门限值。

本实施例通过局部判决,实现对语音帧和噪音帧的区分,本实施例的局部判决在每个频段上做一次,一共 6 次。似然比是反映真实性的一种指标,属于同时反映灵敏度和特异度的复合指标,提高概率估算准确度,本实施例在确保语音帧概率值大于噪音帧概率值的情况下,进一步通过比较局部对数似然比是否高于局部门限值,以确保判定为该语音信号为语音帧的准确性。

15 第一判定子单元,用于若局部对数似然比高于局部门限值,则判定局部对数似然比高于局部门限值的帧数据为语音帧。

本实施例的 GMM 的参数具有自适应更新能力,在每一帧语音信号被判断为语音帧或者噪声帧之后,会根据该帧的特征值来更新其对应模型的参数。例如,20 如果该帧被判断为语音帧,则 GMM-SPEECH 的期望值、标准差和高斯分量权重值就根据该帧的特征值进行一次更新,在越来越多的语音帧输入 GMM-SPEECH 之后, GMM-SPEECH 会越来越适应此通语音信号的说话人的声纹特征,给出的分析结论会更加准确。

进一步地,本申请另一实施例的区分单元,包括:

25 第二计算子单元,用于若局部对数似然比不高于局部门限值,则根据 $\sum_{i=1}^n \text{weight}_i * \log \text{likelihood ratio}_{local,i}$, 计算全局对数似然比。

本实施例进行局部判决,再进行全局判决,全局判决是基于局部判决结果的基础上进行各频段加权求和的计算,以便提高区分语音帧和噪音帧的准确度。

第二判断子单元,用于判断所述全局对数似然比是否高于全局门限值。

30 本实施例的全局判决中将全局对数似然比与全局门限值相比,以进一步提

高筛选语音帧的准确性。

第二判定子单元,用于若全局对数似然比高于全局门限值,则判定全局对数似然比高于全局门限值的帧数据为语音帧。

本实施例可以先根据局部判决结果有语音存在,则不进行全局判决,以提高声纹验证的效率,且尽量将所有的语音帧都能识别到,以免语音失真。本申请其他实施例也可在局部判决结果有语音存在,再进行全局判决,以进一步核实和确认语音的存在,提高区分语音帧和噪音帧的准确度。

进一步地,本实施例的提取模块3,包括:

提取单元,用于提取所述净化的语音数据中各所述语音帧分别对应的 MFCC 类型声纹特征。

本实施例提取 MFCC (Mel Frequency Cepstrum Coefficient, 梅尔频率倒谱系数) 类型声纹特征的过程如下: 先采样和量化,将净化的语音数据的连续模拟语音信号以一定的采样周期采样,转化为离散信号,并根据一定的编码规则将离散信号量化为数字信号;然后预加重,由于人体的生理特性,语音信号的高频成分往往被压抑,预加重的作用是补偿高频成分;接着分帧处理,由于语音信号的“瞬时平稳性”,在进行频谱分析时对一段语音信号进行分帧处理(一般为 10 至 30 毫秒一帧),然后以帧为单位进行特征提取;接着加窗处理,作用是减少帧起始和帧结束对应信号的不连续性问题,采用汉明窗进行加窗处理;接着对帧信号进行 DFT,将信号从时域转换到频域,然后再利用如下公式将信号从线性频谱域映射到梅尔频谱域: $f_{Mel} = 2595 \times \log_{10}(1 + \frac{f}{700})$; 将转化后的帧信号输入到一组梅尔三角滤波器组,计算每个频段的滤波器输出的信号对数能量,得到一个对数能量序列;对上一步得到的对数能量序列做离散余弦变换(DCT, Discrete Cosine Transform)即可得到该帧语音信号的 MFCC 类型声纹特征。

构建单元,用于根据各所述 MFCC 类型声纹特征构建各所述语音帧分别对应的声纹特征向量。

MFCC 类型声纹特征具有非线性特征,使各频段上的分析结果更贴近人体发出的真实语音的特征,使声纹特征提取更加准确,提高声纹验证的效果。

映射单元,用于将各声纹特征向量分别映射为低维度的声纹鉴别向量

l-vector, 以得到净化的语音数据中各所述语音帧分别对应的第一声纹特征。

本实施例基于 GMM-UBM (Gaussian Mixture Model-Universal Background Model, 高斯混合模型-背景模型) 实现将各所述声纹特征向量分别映射为低维度的声纹鉴别向量 l-vector, 降低声纹特征提取过程中的计算成本, 降低声纹验证的使用成本。本实施例的 GMM-UBM 的训练过程如下: B1: 获取预设数量 (例如, 10 万个) 的语音数据样本, 每个语音数据样本对应一个声纹鉴别向量, 每个语音数据样本可以采集自不同的人在不同环境中的语音形成, 这样的语音数据样本用来训练能够表征一般语音特性的通用背景模型 (GMM-UBM); B2、分别对各个语音数据样本进行处理以提取出各个语音数据样本对应的预设类型声纹特征, 并基于各个语音数据样本对应的预设类型声纹特征构建各个语音数据样本对应的声纹特征向量; B3、将构建出的所有预设类型声纹特征向量分为第一百分比的训练集和第二百分比的验证集, 所述第一百分比和第二百分比小于或等于 100%; B4、利用训练集中的声纹特征向量对所述第二模型进行训练, 并在训练完成之后利用验证集对训练的所述第二模型的准确率进行验证; B5、若准确率大于预设准确率 (例如, 98.5%), 则模型训练结束, 否则, 增加语音数据样本的数量, 并基于增加后的语音数据样本重新执行上述步骤 B2、B3、B4、B5。

本实施例的声纹鉴别向量采用 l-vector 表达, i-vector 是一个向量, 相对于高斯空间的维度来讲, i-vector 维度更低, 便于降低计算成本, 而提取低维度的 i-vector 的过程是通过下述计算公式, 将低维度的向量 w 与一个转换矩阵 T 相乘映射到维度较高的高斯空间。l-vector 的提取包括如下步骤: 将来自某位目标说话人的训练语音数据处理后, 提取得到的预设类型声纹特征向量 (例如, MFCC) 输入到 GMM-VAD 模型, 得到一个表征该段语音数据在各个高斯分量上的概率分布的高斯超向量; 利用如下公式可以计算得到该段语音对应的较低维度的声纹鉴别向量 l-vector: $m_r = \mu + T\omega_r$, 其中 ω_r 为代表该段语音的高斯超向量, μ 为所述第二模型的均值超向量, T 为将低维度的 l-vector, ω_r 映射到高维度的高斯空间的转换矩阵, T 的训练采用 EM 算法。

进一步地, 本实施例的判断模块 4, 包括:

获取单元, 用于分别在预存的多个人的声纹特征数据中获取各自对应的预存声纹特征, 其中, 多个人的声纹特征数据中包括目标人的预存声纹特征。

本实施例通过将预存的包括目标人的多人的声纹特征数据, 同时用于判断

当前采集的语音信号的声纹特征是否与目标人的声纹特征相同,以提高判断准确性。

计算单元,用于分别计算各预存声纹特征与第一声纹特征之间的相似度值。

5 本实施例的相似度值表征了预存声纹特征与所述第一声纹特征之间的相似度,相似度值越大,则上述两者越相似。本实施例的相似度值的获取方法包括通过比较预存声纹特征与所述第一声纹特征之间的特征距离值得到,上述特征距离值包括余弦距离值、欧式距离值等。

排序单元,用于将各所述相似度值按照从大到小的顺序进行排序。

10 本实施例通过将各所述预存声纹特征与所述第一声纹特征之间的相似度值进行从大到小排序,以便更准确地分析第一声纹特征与各预存声纹特征的相似度分布状态,以便更准确地获得对第一声纹特征的验证。

判断单元,用于判断排序在前的预设数量的相似度值中,是否包括所述目标人的预存声纹特征对应的相似度值。

15 本实施例通过排序在前的预设数量的相似度值中,包括目标人的预存声纹特征对应的相似度值,则判定第一声纹特征与预存的目标人的声纹特征相同,以减小模型误差带来的识别等错率,等错率为“应验证通过时发生的验证未通过的频率,与应验证未通过时发生的验证通过的频率相等”。本实施例的预设数量的相似度值包括1、2或3个等,可根据使用需求进行自设定。

20 判定单元,用于若包括目标人的预存声纹特征对应的相似度值,则判定第一声纹特征与预存声纹特征的相似度满足预设条件,否则不满足预设条件。

本申请其他实施例通过设定第一声纹特征与目标用户的预存声纹特征的距离阈值,实现有效的声纹验证。举例地,预设阈值为0.6,若计算第一声纹特征与目标用户的预存声纹特征的余弦距离小于或等于预设阈值,则确定第一
25 声纹特征与目标用户的预存声纹特征相同,则验证通过;若计算第一声纹特征与目标用户的预存声纹特征的余弦距离大于预设阈值,则确定第一声纹特征与目标用户的预存声纹特征不相同,则验证失败。

进一步地,本实施例的计算单元,包括:

第三计算子单元,用于分别通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$, 计算各

30 所述预存声纹特征与所述第一声纹特征之间的余弦距离值,其中,x代表各预

存声纹鉴别向量， y 代表第一声纹特征的声纹鉴别向量。

本实施例通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$ ，表示各所述预存声纹特征与所述第一声纹特征之间的相似度，其中余弦距离的距离值越小，表明两声纹特征更接近或相同。

- 5 转换子单元，用于将所述余弦距离值转换成所述相似度值，其中，最小的所述余弦距离值对应最大的相似度值。

本实施例可通过将余弦距离值按照携带指定反比系数的反比例公式，将余弦距离值转换成相似度值。

- 10 参照图 3，本申请实施例中还提供一种计算机设备，该计算机设备可以是服务器，其内部结构可以如图 3 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机
15 设备的数据库用于存储声纹验证等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令在执行时，执行如上述各方法的实施例的流程。本领域技术人员可以理解，图 3 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定。

- 20 本申请一实施例还提供一种计算机非易失性可读存储介质，其上存储有计算机可读指令，该计算机可读指令在执行时，执行如上述各方法的实施例的流程。以上所述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、一种声纹验证的方法，其特征在于，包括：

将待声纹验证的语音信号输入到 VAD 模型中，区分所述语音信号中的语音帧和噪音帧；

5 去除所述噪音帧，得到各所述语音帧组成的净化的语音数据；

提取所述净化的语音数据对应的第一声纹特征；

判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件；

若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

2、根据权利要求 1 所述的声纹验证的方法，其特征在于，所述 VAD 模型
10 中包括傅里叶变换、高斯混合分布的 GMM-NOISE 和 GMM-SPEECH，所述将所述语音信号输入到 VAD 模型中，区分语音信号中的语音帧和噪音帧的步骤，包括：

将所述语音信号输入到 VAD 模型中的傅里叶变换中，将所述语音信号从时域信号形式转变为频域信号形式；

15 将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧。

3、根据权利要求 2 所述的声纹验证的方法，其特征在于，所述将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧的步骤，包括：

20 将频域信号形式的语音信号各帧数据，分别输入到 GMM-NOISE 和 GMM-SPEECH 中，分别得到各帧数据的噪音帧概率 $P_{local}^{(x|noise)}$ 和语音帧概率 $P_{local}^{(x|speech)}$ ；

根据 $\log \text{ likelihood ratio}_{local,i} = \log_2 \left(\frac{P_{local}^{(x|speech_i)}}{P_{local}^{(x|speech_i)}}$), $i = 1, 2, \dots, n$ ，计算局部

对数似然比；

判断所述局部对数似然比是否高于局部门限值；

25 若是，则判定所述局部对数似然比高于局部门限值的帧数据为语音帧。

4、根据权利要求 3 所述的声纹验证的方法，其特征在于，所述判断所述对数似然比是否高于局部门限值的步骤之后，包括：

若局部对数似然比不高于局部门限值，则根据 $\text{sum log likelihood ratio} = \sum_{i=1}^n \text{weight}_i * \log \text{ likelihood ratio}_{local,i}$ ，计算全局对数似然

比；

判断所述全局对数似然比是否高于全局门限值；

若全局对数似然比高于全局门限值，则判定所述全局对数似然比高于全局门限值的帧数据为语音帧。

5 5、根据权利要求 1 所述的声纹验证的方法，其特征在于，所述提取所述净化的语音数据对应的第一声纹特征的步骤，包括：

提取所述净化的语音数据中各所述语音帧分别对应的 MFCC 类型声纹特征；

10 根据各所述 MFCC 类型声纹特征构建各所述语音帧分别对应的声纹特征向量；

将各所述声纹特征向量分别映射为低维度的声纹鉴别向量 l-vector，以得到所述净化的语音数据中各所述语音帧分别对应的第一声纹特征。

6、根据权利要求 5 所述的声纹验证的方法，其特征在于，所述判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件的步骤，包括：

15 分别在预存的多个人的声纹特征数据中获取各自对应的预存声纹特征，其中，多个人的声纹特征数据中包括目标人的预存声纹特征；

分别计算各所述预存声纹特征与所述第一声纹特征之间的相似度值；

将各所述相似度值按照从大到小的顺序进行排序；

20 判断排序在前的预设数量的相似度值中，是否包括所述目标人的预存声纹特征对应的相似度值；

若是，则判定所述第一声纹特征与预存声纹特征的相似度满足预设条件，否则不满足预设条件。

7、根据权利要求 6 所述的声纹验证的方法，其特征在于，所述分别计算各所述预存声纹特征与所述第一声纹特征之间的相似度值的步骤，包括：

25 分别通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$ ，计算各所述预存声纹特征与所

述第一声纹特征之间的余弦距离值，其中，x 代表各预存声纹鉴别向量，y 代表第一声纹特征的声纹鉴别向量；

将所述余弦距离值转换成所述相似度值，其中，最小的所述余弦距离值对应最大的相似度值。

30 8、一种声纹验证的装置，其特征在于，包括：

区分模块,用于将待声纹验证的语音信号输入到 VAD 模型中,区分所述语音信号中的语音帧和噪音帧;

去除模块,用于去除所述噪音帧,得到各所述语音帧组成的净化的语音数据;

5 提取模块,用于提取所述净化的语音数据对应的第一声纹特征;

判断模块,用于判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件;

判定模块,用于若满足预设条件,则判定所述第一声纹特征与所述预存声纹特征相同,否则不相同。

10 9、根据权利要求 8 所述的声纹验证的装置,其特征在于,所述 VAD 模型中包括傅里叶变换、高斯混合分布的 GMM-NOISE 和 GMM-SPEECH,所述区分模块,包括:

转变单元,用于将所述语音信号输入到 VAD 模型中的傅里叶变换中,将所述语音信号从时域信号形式转变为频域信号形式;

15 区分单元,用于将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决,以区分语音信号中的语音帧和噪音帧。

10、根据权利要求 9 所述的声纹验证的装置,其特征在于,所述区分单元,包括:

20 输入子单元,用于将频域信号形式的语音信号的各帧数据,分别输入到 GMM-NOISE 和 GMM-SPEECH 中,分别得到各帧数据的噪音帧概率 $P_{local}^{(x|noise)}$ 和语音帧概率 $P_{local}^{(x|speech)}$;

第一计算子单元,用于根据
$$\log \text{ likelihood ratio}_{local,i} = \log_2 \left(\frac{P_{local}^{(x|speech_i)}}{P_{local}^{(x|noise_i)}} \right), i = 1, 2, \dots, n$$
 , 计算局部对数似然

25 比;

第一判断子单元,用于判断所述局部对数似然比是否高于局部门限值;

第一判定子单元,用于若局部对数似然比高于局部门限值,则判定所述局部对数似然比高于局部门限值的帧数据为语音帧。

11、根据权利要求 10 所述的声纹验证的装置,其特征在于,所述区分单

元, 包括:

第二计算子单元, 用于若局部对数似然比不高于局部门限值, 则根据 $\text{sum log likelihood ratio} = \sum_{i=1}^n \text{weight}_i * \log \text{likelihood ratio}_{\text{local}, i}$, 计算全局对数似然比;

5 第二判断子单元, 用于判断所述全局对数似然比是否高于全局门限值;

第二判定子单元, 用于若全局对数似然比高于全局门限值, 则判定所述全局对数似然比高于全局门限值的帧数据为语音帧。

12、根据权利要求 8 所述的声纹验证的装置, 其特征在于, 所述提取模块, 包括:

10 提取单元, 用于提取所述净化的语音数据中各所述语音帧分别对应的 MFCC 类型声纹特征;

构建单元, 用于根据各所述 MFCC 类型声纹特征构建各所述语音帧分别对应的声纹特征向量;

15 映射单元, 用于将各所述声纹特征向量分别映射为低维度的声纹鉴别向量 l-vector, 以得到所述净化的语音数据中各所述语音帧分别对应的第一声纹特征。

13、根据权利要求 12 所述的声纹验证的装置, 其特征在于, 所述判断模块, 包括:

20 获取单元, 用于分别在预存的多个人的声纹特征数据中获取各自对应的预存声纹特征, 其中, 多个人的声纹特征数据中包括目标人的预存声纹特征;

计算单元, 用于分别计算各所述预存声纹特征与所述第一声纹特征之间的相似度值;

排序单元, 用于将各所述相似度值按照从大到小的顺序进行排序;

25 判断单元, 用于判断排序在前的预设数量的相似度值中, 是否包括所述目标人的预存声纹特征对应的相似度值;

判定单元, 用于若包括所述目标人的预存声纹特征对应的相似度值, 则判定所述第一声纹特征与预存声纹特征的相似度满足预设条件, 否则不满足预设条件。

30 14、根据权利要求 13 所述的声纹验证的装置, 其特征在于, 所述计算单元, 包括:

第三计算子单元，用于分别通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$ ，计算各所述预存声纹特征与所述第一声纹特征之间的余弦距离值，其中，x 代表各预存声纹鉴别向量，y 代表第一声纹特征的声纹鉴别向量；

转换子单元，用于将所述余弦距离值转换成所述相似度值，其中，最小的所述余弦距离值对应最大的相似度值。

15、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，其特征在于，所述处理器执行所述计算机程序时实现声纹验证的方法，声纹验证的方法包括：

将待声纹验证的语音信号输入到 VAD 模型中，区分所述语音信号中的语音帧和噪音帧；

去除所述噪音帧，得到各所述语音帧组成的净化的语音数据；

提取所述净化的语音数据对应的第一声纹特征；

判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件；

若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

16、根据权利要求 15 所述的计算机设备，其特征在于，所述 VAD 模型中包括傅里叶变换、高斯混合分布的 GMM-NOISE 和 GMM-SPEECH，所述将所述语音信号输入到 VAD 模型中，区分语音信号中的语音帧和噪音帧的步骤，包括：

将所述语音信号输入到 VAD 模型中的傅里叶变换中，将所述语音信号从时域信号形式转变为频域信号形式；

将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧。

17、根据权利要求 16 所述的计算机设备，其特征在于，所述将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和 GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧的步骤，包括：

将频域信号形式的语音信号的各帧数据，分别输入到 GMM-NOISE 和 GMM-SPEECH 中，分别得到各帧数据的噪音帧概率 $P_{local}^{(x|noise)}$ 和语音帧概率 $P_{local}^{(x|speech)}$ ；

根据 $\log \text{likelihood ratio}_{local,i} = \log_2 \left(\frac{P_{local}^{(x|speech_i)}}{P_{local}^{(x|speech_i)}} \right), i = 1, 2, \dots, n$ ，计算局部

对数似然比；

判断所述局部对数似然比是否高于局部门限值；

若是，则判定所述局部对数似然比高于局部门限值的帧数据为语音帧。

18、一种计算机非易失性可读存储介质，其上存储有计算机程序，其特征
5 在于，所述计算机程序被处理器执行时实现声纹验证的方法，声纹验证的方法包括：

将待声纹验证的语音信号输入到 VAD 模型中，区分所述语音信号中的语音
帧和噪音帧；

去除所述噪音帧，得到各所述语音帧组成的净化的语音数据；

10 提取所述净化的语音数据对应的第一声纹特征；

判断所述第一声纹特征与预存声纹特征的相似度是否满足预设条件；

若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

19、根据权利要求 18 所述的计算机非易失性可读存储介质，其特征在于，
所述 VAD 模型中包括傅里叶变换、高斯混合分布的 GMM-NOISE 和 GMM-SPEECH，
15 所述将所述语音信号输入到 VAD 模型中，区分语音信号中的语音帧和噪音帧的
步骤，包括：

将所述语音信号输入到 VAD 模型中的傅里叶变换中，将所述语音信号从时
域信号形式转变为频域信号形式；

20 将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和
GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧。

20、根据权利要求 19 所述的计算机非易失性可读存储介质，其特征在于，
所述将频域信号形式的语音信号的每一帧数据分别输入到所述 GMM-NOISE 和
GMM-SPEECH 中进行 VAD 判决，以区分语音信号中的语音帧和噪音帧的步骤，
包括：

25 将频域信号形式的语音信号各帧数据，分别输入到 GMM-NOISE 和
GMM-SPEECH 中，分别得到各帧数据的噪音帧概率 $P_{local}^{(x|noise)}$ 和语音帧概率
 $P_{local}^{(x|speech)}$ ；

根据 $\log \text{likelihood ratio}_{local,i} = \log_2 \left(\frac{P_{local}^{(x|speech_i)}}{P_{local}^{(x|speech_i)}} \right), i = 1, 2, \dots, n$ ，计算局部

对数似然比；

判断所述局部对数似然比是否高于局部门限值；

若是，则判定所述局部对数似然比高于局部门限值的帧数据为语音帧。

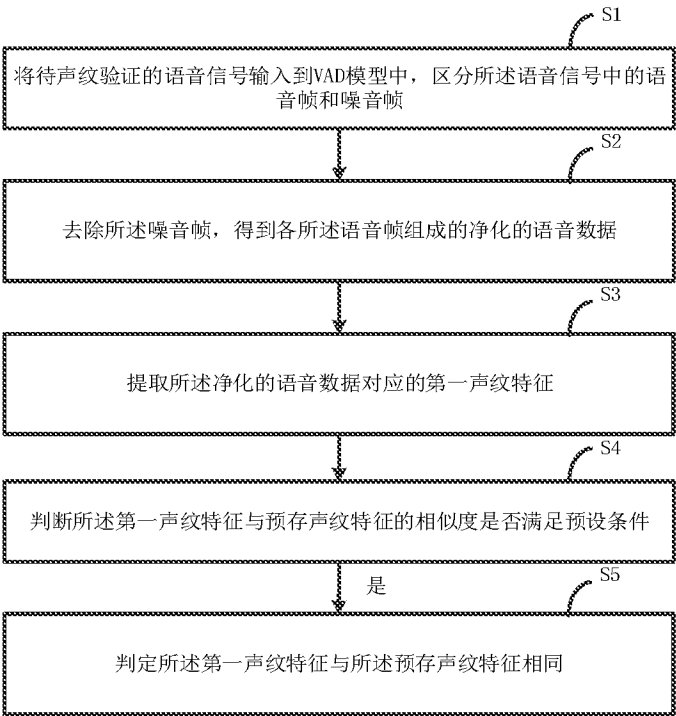


图 1

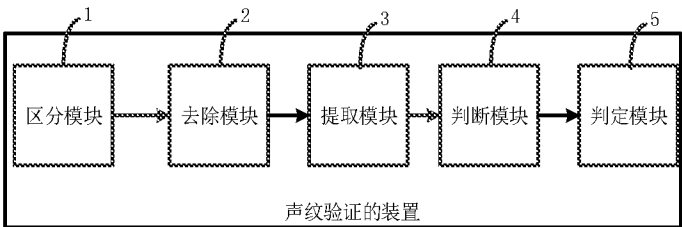


图 2

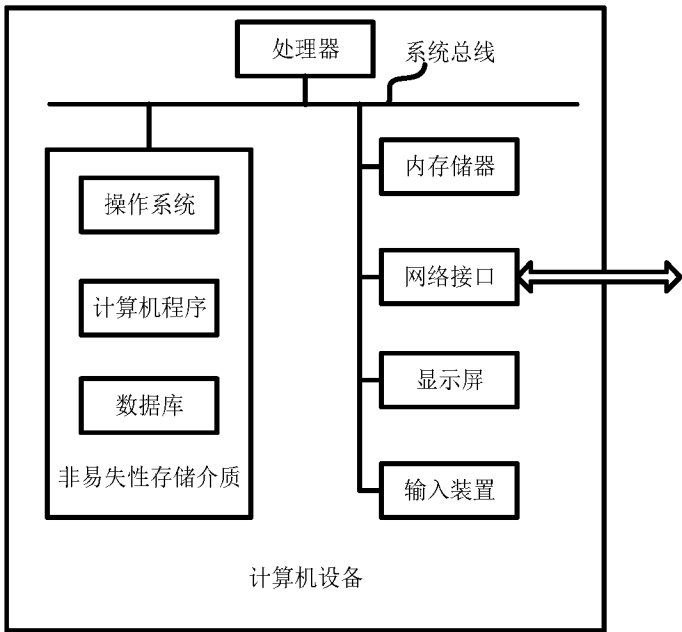


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/124401

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/00(2013.01)i; G10L 17/04(2013.01)i; G10L 21/0208(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, VEN: 声纹, 声音, 语音, 音频, 验证, 识别, 检测, 噪音, 特征, 提取, VAD, GMM, 似然比, 相似度, MFCC, 向量, 矢量, 距离, voiceprint, soundwave, audio, acoustic, verification, recognize, detect, noise, feature, extract, probability, likelihood, vector, distance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102479511 A (SHENGLE INFORMATION TECHNOLOGY (SHANGHAI) CO., LTD.) 30 May 2012 (2012-05-30) description, paragraphs 27-50	1-20
Y	CN 108428456 A (ZHEJIANG KAICHI ELECTRONIC TECHNOLOGY CO., LTD.) 21 August 2018 (2018-08-21) description, paragraphs 43-75	1-20
Y	CN 108109612 A (SOOCHOW UNIVERSITY) 01 June 2018 (2018-06-01) description, paragraphs 51-54	2-4, 9-11, 16-17, 19-20
A	CN 103236260 A (BOE TECHNOLOGY GROUP CO., LTD. ET AL.) 07 August 2013 (2013-08-07) entire document	1-20
A	WO 0111606 A1 (ERICSSON, INC.) 15 February 2001 (2001-02-15) entire document	1-20

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

21 June 2019

Date of mailing of the international search report

28 June 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/CN2018/124401

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102479511	A	30 May 2012	None			
CN	108428456	A	21 August 2018	None			
CN	108109612	A	01 June 2018	None			
CN	103236260	A	07 August 2013	CN	103236260	B	12 August 2015
				US	2015340027	A1	26 November 2015
				WO	2014153800	A1	02 October 2014
WO	0111606	A1	15 February 2001	US	6349278	B1	19 February 2002
				AU	6090900	A	05 March 2001
				MY	130355	A	29 June 2007

国际检索报告

国际申请号

PCT/CN2018/124401

A. 主题的分类 G10L 17/00(2013.01)i; G10L 17/04(2013.01)i; G10L 21/0208(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, VEN: 声纹, 声音, 语音, 音频, 验证, 识别, 检测, 噪音, 特征, 提取, VAD, GMM, 似然比, 相似度, MFCC, 向量, 矢量, 距离, voiceprint, soundwave, audio, acoustic, verification, recognize, detect, noise, feature, extract, probability, likelihood, vector, distance																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>CN 102479511 A (盛乐信息技术上海有限公司) 2012年 5月 30日 (2012 - 05 - 30) 说明书第27-50段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 108428456 A (浙江凯池电子科技有限公司) 2018年 8月 21日 (2018 - 08 - 21) 说明书第43-75段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 108109612 A (苏州大学) 2018年 6月 1日 (2018 - 06 - 01) 说明书第51-54段</td> <td>2-4, 9-11, 16-17, 19-20</td> </tr> <tr> <td>A</td> <td>CN 103236260 A (京东方科技集团股份有限公司等) 2013年 8月 7日 (2013 - 08 - 07) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>WO 0111606 A1 (ERICSSON, INC.) 2001年 2月 15日 (2001 - 02 - 15) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 102479511 A (盛乐信息技术上海有限公司) 2012年 5月 30日 (2012 - 05 - 30) 说明书第27-50段	1-20	Y	CN 108428456 A (浙江凯池电子科技有限公司) 2018年 8月 21日 (2018 - 08 - 21) 说明书第43-75段	1-20	Y	CN 108109612 A (苏州大学) 2018年 6月 1日 (2018 - 06 - 01) 说明书第51-54段	2-4, 9-11, 16-17, 19-20	A	CN 103236260 A (京东方科技集团股份有限公司等) 2013年 8月 7日 (2013 - 08 - 07) 全文	1-20	A	WO 0111606 A1 (ERICSSON, INC.) 2001年 2月 15日 (2001 - 02 - 15) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
Y	CN 102479511 A (盛乐信息技术上海有限公司) 2012年 5月 30日 (2012 - 05 - 30) 说明书第27-50段	1-20																		
Y	CN 108428456 A (浙江凯池电子科技有限公司) 2018年 8月 21日 (2018 - 08 - 21) 说明书第43-75段	1-20																		
Y	CN 108109612 A (苏州大学) 2018年 6月 1日 (2018 - 06 - 01) 说明书第51-54段	2-4, 9-11, 16-17, 19-20																		
A	CN 103236260 A (京东方科技集团股份有限公司等) 2013年 8月 7日 (2013 - 08 - 07) 全文	1-20																		
A	WO 0111606 A1 (ERICSSON, INC.) 2001年 2月 15日 (2001 - 02 - 15) 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 6月 21日		国际检索报告邮寄日期 2019年 6月 28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 王博 电话号码 86-010-62089145																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/124401

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	102479511	A	2012年 5月 30日	无			
CN	108428456	A	2018年 8月 21日	无			
CN	108109612	A	2018年 6月 1日	无			
CN	103236260	A	2013年 8月 7日	CN	103236260	B	2015年 8月 12日
				US	2015340027	A1	2015年 11月 26日
				WO	2014153800	A1	2014年 10月 2日
WO	0111606	A1	2001年 2月 15日	US	6349278	B1	2002年 2月 19日
				AU	6090900	A	2001年 3月 5日
				MY	130355	A	2007年 6月 29日