(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 6 月 18 日 (18.06.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/119448 A1

(51) 国际专利分类号:
G10L 17/24 (2013.01) *G10L 17/18* (2013.01)

(21) 国际申请号: PCT/CN2019/120972

(22) 国际申请日: 2019 年 11 月 26 日 (26.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811527293.3 2018年12月13日 (13.12.2018) CN

(71) 申请人: 北京三快在线科技有限公司 (**BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD**) [CN/CN]; 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。

(72) 发明人: 陈欢 (**CHEN, Huan**); 中国北京市海淀区北四环西路9号2106-030, Beijing 100080 (CN)。

(74) 代理人: 北京曼威知识产权代理有限公司 (**BEIJING MAVAIP INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 411 室, Beijing 100085 (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,

(54) **Title:** VOICE INFORMATION VERIFICATION

(54) 发明名称: 语音信息验证

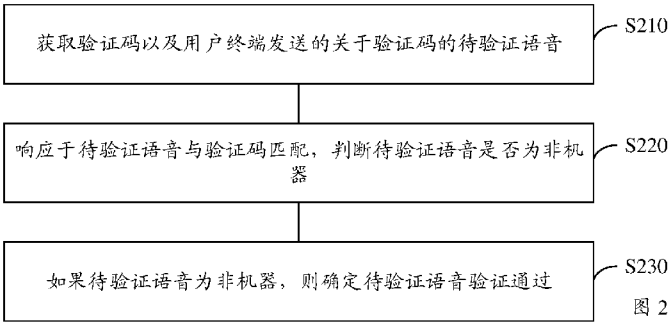


图 2

S210 Obtain a verification code and a voice to be verified, said voice being sent by a user terminal and related to the verification code

S220 In response to said voice being matched with the verification code, determine whether said voice is not made by a machine

S230 If said voice is not made by a machine, determine that the verification of said voice is passed

(57) **Abstract:** A method and apparatus for voice information verification, and an electronic device and a computer-readable storage medium. The method comprises: obtaining a verification code and a voice to be verified, said voice being sent by a user terminal and related to the verification code (S210); in response to said voice being matched with the verification code, determining whether said voice is not made by a machine (S220); and if said voice is not made by a machine, determining that the verification of said voice is passed (S230).

(57) 摘要: 一种语音信息验证方法及装置、电子设备、计算机可读存储介质。方法包括：获取验证码以及用户终端发送的关于验证码的待验证语音(S210)；响应于待验证语音与验证码匹配，判断待验证语音是否为非机器(S220)；如果待验证语音为非机器，则确定待验证语音验证通过(S230)。

WO 2020/119448 A1

[见续页]

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))。

语音信息验证

技术领域

[01]本公开涉及人工智能技术领域，尤其涉及一种语音信息验证方法、语音信息验证装置、电子设备及计算机可读存储介质。

背景技术

[02]随着计算机技术的发展，在很多 App（Application，应用程序）及网站上出现了盗取账户、虚假账户、网络诈骗等恶意行为，对 App 及网站的正常运营造成了安全隐患。因此很有必要对各种账户进行身份验证，以阻止盗取账户或虚假账户进行各种网络行为，保障真实用户的利益。

[03]需要说明的是，在上述背景技术部分公开的信息仅用于加强对本公开的背景的理解，因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

发明内容

[04]本公开提供了一种语音信息验证方法、语音信息验证装置、电子设备及计算机可读存储介质，进而至少在一定程度上克服由于现有技术的缺陷而导致的身份验证安全性较低的问题。

[05]本公开的其他特性和优点将通过下面的详细描述变得显然，或部分地通过本公开的实践而习得。

[06]根据本公开的第一方面，提供一种语音信息验证方法，包括：获取验证码以及用户终端发送的关于所述验证码的待验证语音；响应于所述待验证语音与所述验证码匹配，则判断所述待验证语音是否为非机器；如果所述待验证语音为非机器，则判定所述待验证语音验证通过。

[07]根据本公开的第二一方面，提供一种语音信息验证装置，包括：信息获取模块，用于获取验证码以及用户终端发送的关于所述验证码的待验证语音；人机判断模块，用于响应于所述待验证语音与所述验证码匹配，判断所述待验证语音是否为非机器；语音验证模块，用于如果所述待验证语音为非机器，则确定所述待验证语音验证通过。

[08]根据本公开的第三方面，提供一种电子设备，包括：处理器；以及存储器，用于存

储所述处理器的可执行指令；其中，当执行所述机器可执行指令时，所述处理器配置为执行上述第一方面所述的方法。

[09]根据本公开的第四方面，提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述第一方面所述的方法。

[10]应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本公开。

附图说明

[11]此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本公开的实施例，并与说明书一起用于解释本公开的原理。显而易见地，下面描述中的附图仅仅是本公开的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[12]图 1 示意性示出示例性实施例运行环境的一种系统架构示意图；

[13]图 2 示意性示出示例性实施例中一种语音信息验证方法的步骤图；

[14]图 3 示意性示出示例性实施例中一种语音信息验证方法的流程图；

[15]图 4 示意性示出示例性实施例中一种获取验证码方法的流程图；

[16]图 5 示意性示出示例性实施例中另一种语音信息验证方法的流程图；

[17]图 6 示意性示出示例性实施例中一种语音信息验证方法的流程图；

[18]图 7 示意性示出示例性实施例中一种语音信息验证装置的结构框图；

[19]图 8 示意性示出示例性实施例中一种用于实现上述方法的电子设备；

[20]图 9 示意性示出示例性实施例中一种用于实现上述方法的计算机可读存储介质。

具体实施方式

[21]现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的范例；相反，提供这些实施方式使得本公开将更加全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。

[22]在相关技术的一种方案中，可通过用户自身的声纹特征匹配以进行语音验证。需要用户在注册时输入语音注册信息，从中提取声纹特征信息，以便于之后在用户登录时，对语音验证信息进行声纹特征匹配。但是，在该方案中，如果语音注册信息或语音验证信息是由机器合成的，例如爬取用户的日常语音信息后，利用其原声合成的语音验证信息，则服务器难以识别，且需要存储用户的声纹特征信息，增加了系统的资源占用。

[23]鉴于上述问题，本公开的示例性实施例首先提供了一种语音信息验证方法，可以应用于用户在登录 App、网页或进行支付确认、修改密码等敏感性操作时，通过验证码对用户进行身份验证的场景。

[24]图 1 示出了本示例性实施例的运行环境的一种系统架构示意图。参考图 1 所示，该系统 110 可以包括用户终端 111，网络 112 及服务器 113。其中，用户终端 111 可以是用户使用的各种个人计算机、平板电脑、智能手机或可穿戴设备等终端设备，采集到关于验证码的待验证语音后，通过网络 112 发送至服务器 113；服务器 113 可以从用户终端 111 获取待验证语音，并对其进行语音信息验证。

[25]应当理解，图 1 中所示各装置的数目仅为示例性的，根据实际需要，可以设置任意数目的用户终端 111 或网络 112，服务器 113 也可以是由多个服务器组成的服务器集群。

[26]基于上述说明，本示例性实施例中的方法可以应用于图 1 所示的服务器 113 上。

[27]下面结合附图 2 对本示例性实施例做进一步说明，如图 2 所示，语音信息验证方法可以包括以下步骤 S210~S230:

[28]步骤 S210，获取验证码以及用户终端发送的关于该验证码的待验证语音。

[29]其中，验证码可以是文本验证码、语音验证码、图像验证码等，可以由服务器的特定程序生成，也可以从其他服务器获取。待验证语音是用户根据验证码的内容发送的语音，例如用户朗读验证码的内容而录制并上传的语音。用户终端可以通过用户的触发操作来采集关于验证码的待验证语音，然后发送至服务器，例如在输入待验证语音时，用户可以点击特定控件录入待验证语音，或者选择将已有的录音文件作为待验证语音等。

[30]步骤 S220，响应于待验证语音与验证码匹配，判断待验证语音是否为非机器。

[31]在本示例性实施例中，首先，可以检测待验证语音与验证码是否匹配，对于不同类型的验证码，待验证语音与验证码可以有多种匹配方式，例如：如果验证码为文本验证码，则可以先将待验证语音转换为待验证文本，再将待验证文本与文本验证码对比来进行匹配；如果验证码为语音验证码，则可以通过对比语音验证码与待验证语音的声纹特

征等来进行匹配等等。在步骤 S220 中，对待验证语音进行验证码匹配相当于第一次验证，主要用于验证身份一致性，可以识别出用户账户被窃取而在非绑定终端上登录或操作等情况。

[32]在本示例性实施例中，如果上述第一次验证通过，则对待验证语音继续进行人机判断，为第二次验证。在判断待验证语音为机器或非机器时，可以对待验证语音的声纹特征进行分析匹配，比较其声纹特征是否与机器合成声音的声纹特征相同或相近，也可以比较其声纹特征是否与用户声音数据库中的声纹特征匹配等。

[33]步骤 S230，如果待验证语音为非机器，则确定待验证语音验证通过。

[34]通过对待验证语音进行第二次验证，可以排除机器合成语音的情况，从而确定了发送待验证语音的用户收到了正确的验证码，并且为真实用户，可以确定待验证语音最终通过验证。

[35]在一示例性实施例中，步骤 S220 可以包括以下步骤:

如果待验证语音与验证码匹配，则将待验证语音转换为目标频谱图;

通过卷积神经网络模型对目标频谱图进行分析，得到目标频谱图的人机分类结果;

基于人机分类结果，确定待验证语音是否为非机器。

[36]目标频谱图是待验证语音所对应的频谱图。频谱图通常有两种：一种是以频率为横坐标，以信号能量为纵坐标的瞬时频谱图，待验证语音可以转换为多幅瞬时频谱图的序列；一种是以时间为横坐标，叠加频率为纵坐标的连续频谱图，待验证语音可以转换一幅连续频谱图。在服务器获取到待验证语音后，可以将其转换为合适格式的语音文件，对其进行时域分析和频域分析，绘制出目标频谱图。可见，目标频谱图包含了待验证语音的频谱特征，而频谱特征又能够体现出待验证语音的声纹特征，因此通过卷积神经网络模型处理，可以识别其中的声纹特征，并将其与模型训练时学习到的机器合成语音或人类真实语音的声纹特征进行比较，判断目标频谱图为机器或非机器，这样得到了待验证语音的人机分类结果，从而可以完成第二次验证的过程。

[37]图 3 示出了本示例性实施例中一种语音信息验证方法的流程图。如图 3 所示，获取待检测语音后，可以首先进行验证码匹配，如果匹配失败，则可以直接输出验证失败的结果，如果匹配成功，则将待验证语音转换为目标频谱图，再根据卷积神经网络输出目标频谱图的机器判定概率，并判断机器判定概率是否大于概率阈值，如果大于概率阈值，则验证失败，如果未大于概率阈值，则可以输出最终验证通过的结果。

[38]基于上述说明，本示例性实施例对验证码及用户终端发送的关于验证码的待验证语音进行匹配，并判断匹配成功的将待验证语音是否为非机器，如果判断结果为非机器，则得到待验证语音的验证结果。一方面，对待验证语音进行验证码匹配以及人机判断的验证，能够在验证用户身份一致性的同时，减少机器合成语音对系统进行的攻击，提高了语音信息验证方法的安全性。另一方面，本示例性实施例中，不需要用户预先输入注册语音信息，也不需要存储用户的声纹特征信息，从而降低了语音信息验证方法的使用成本，简化了用户的操作流程，减少了系统的资源占用，提高了效率。

[39]在一示例性实施例中，上述卷积神经网络模型可以通过以下步骤实现训练:

获取多个样本语音以及样本语音的分类标签;

将各样本语音转换为样本频谱图;

利用样本频谱图以及分类标签训练并得到卷积神经网络模型。

[40]其中，样本语音可以是历史验证语音，也可以包括一部分机器合成语音；分类标签可以指人工标注的每个样本语音属于“机器”或“非机器”的标签。

[41]卷积神经网络模型以样本频谱图为输入，输出各样本频谱图的分类结果，通过调整模型参数，可以使输出的分类结果越来越接近分类标签。在训练中，也可以将样本频谱图及分类标签划分为训练集与验证集（例如 8:2 划分）。其中，训练集用于训练模型，设定初始学习率，以及每学习预设次数学习率降低的百分比，可以使模型更快的收敛;验证集用于验证模型的训练效果，如果在验证集上模型运行的准确率达到一定标准，则可以认为训练完成。

[42]利用训练完成的卷积神经网络模型对目标频谱图进行分析，可以得到目标频谱图的人机分类结果，即目标频谱图属于机器或非机器。

[43]利用卷积神经网络模型对待验证语音的目标频谱图进行人机分类，相当于对待验证语音进行了第二次验证，主要用于识别机器合成语音的情况。因此，当目标频谱图的人机分类结果为非机器时，可以认为待验证语音为真实的人类语音，从而最终判定待验证语音验证通过。

[44]在一示例性实施例中，对目标频谱图的人机分类可以通过卷积神经网络模型输出的机器判定概率实现。卷积神经网络模型分析目标频谱图时，可以根据目标频谱图与机器合成语音的特征频谱图相似程度，输出机器判定概率，概率越高表示目标频谱图属于机器的可能性越大。如果机器判定概率大于概率阈值，则可以认为目标频谱图的人机分类

结果为机器。可以根据训练及实际应用的情况设定并调整概率阈值，使其可以准确区分机器与非机器的目标频谱图。

[45]在一示例性实施例中，步骤 S210 中获取验证码可以包括：获取经过扰动处理的验证码。

[46]其中，扰动处理可以是添加对验证码的识别具有干扰性的信息，从而增加机器识别的难度。例如：如果验证码是文本验证码，扰动处理可以是在文本验证码中加入干扰字符等；如果验证码是语音验证码，扰动处理可以是在语音验证码中加入干扰音乐、特定频率的声波等；如果验证码是图像验证码，扰动处理可以是改变图像验证码的图像颜色、降低图像分辨率、打乱图像验证码的图像组成顺序等；如果验证码是短信验证码，扰动处理可以是在短信中添加乱序字母或数字串等等。

[47]在一示例性实施例中，获取添加扰动信息的验证码可以包括以下步骤：

获取预设文本，并将预设文本转换为目标图片；

通过对目标图片进行变形、变色、模糊化及增加噪点中的一种或多种处理，生成经过扰动处理的图像验证码。

[48]图 4 示出了上述获取验证码的流程图。如图 4 所示，首先获取预设文本，预设文本可以从文本库中获取，也可以是生成的随机数等，将获得的预设文本转换为目标图片，目标图片可以是包含了验证码文本信息的图片，例如包含文本验证码字符的图片，或者图像验证码碎片的图片等。在转换目标图片时，可以生成验证码的艺术字，并嵌入到特定的背景中以得到目标图片；也可以如图 4 所示，将验证码的字符拆分，每个字符生成单字图片，然后打乱顺序并拼接为一幅目标图片。对于生成的目标图片，可以根据需要进行剪裁，然后进行扰动处理。其中，变形处理可以是将图片进行拉伸、旋转、液化等处理；变色处理是指改变目标图片的色阶参数，例如将目标图片转换为灰度图、调整明暗分布等；模糊化处理是指降低目标图片的分辨率，例如对目标图片进行局部像素压缩等；增加噪点处理可以是对目标图片添加影响辨认的颗粒等。另外，还可以在目标图片中随机添加干扰性元素，例如，在目标图片上添加横线或竖线等。本领域技术人员容易想到的其他干扰处理方式也应当包含在本公开的保护范围内。

[49]在一示例性实施例中，在响应于待验证语音与验证码匹配之前，语音信息验证方法还可以包括以下步骤：

检测待验证语音的长度；

如果待验证语音的长度小于一预设长度，则判定待验证语音验证失败，并向用户终端返回失败提示信息；

如果待验证语音的长度大于等于预设长度，则将待验证语音转换为待验证文本，并将待验证文本与验证码进行匹配。

[50]其中，待验证语音的长度可以是其时长，也可以是其文件大小等。考虑到过短的待验证语音可能是由误输入造成的，导致无意义的验证，影响服务器的工作效率，可以设置一个预设长度作为待验证语音长度的判断标准，以滤除上述情况，例如将预设长度设置为两秒，不足两秒的语音将被服务器自动过滤，并返回失败提示信息；又例如将预设长度设置为 5KB，则小于 5KB 的待验证语音也将自动被服务器过滤等。

[51]在检测到待验证语音的长度大于等于预设长度之后，可以将待验证语音转换为待验证文本，其中，待验证文本是与待验证语音内容对应的文本形式的待验证信息，可以利用语音文本转换工具实现转换。

[52]上述预设长度可视为待验证语音的长度下限，需要补充的是，可能存在用户误输入或发送错误的录音文件，使得服务器接收的待验证语音过于复杂或冗长的情况，则可能造成服务器处理困难，增加无意义的工作量。因此，还可以设置待验证语音的长度上限，以滤除上述过于复杂或冗长的待验证语音。当语音长度超过语音长度上限时，可以输出验证失败的结果，返回失败提示信息。此外，可以在客户端程序内设置待验证语音的录制时间上限，当超过上限时自动终止语音的录入，并将已录入的语音发送到服务器以进行语音信息验证。

[53]进一步的，将待验证语音转换为待验证文本可以通过以下步骤实现：

对待验证语音进行声道转换、预加重、语音增强、去除空白中的一种或多种预处理；

利用时延神经网络模型将预处理后的待验证语音转换为待验证文本。

[54]考虑到服务器获取的待验证语音可能包含了其他噪声等影响因素，对待验证语音的处理和识别造成一定的障碍。在将待验证语音转换为待验证文本时，可以对待验证语音进行预处理。其中，声道转换处理是指如果所要提取的待验证语音的语音特征不区分声道，则可以将多声道的待验证语音转换为单声道；预加重处理可以是保留待验证语音一定频率范围内的语音信号，有利于服务器对待验证语音信息的分析；语音增强处理可以是从待验证语音中滤除噪音，以提取纯净的语音信号；去除空白处理是指去除待验证语音中无实际信号的片段，例如用户在输入待验证语音的过程中被打断或思考的情况下，

待验证语音可能存在噪声或其他空白的无效语音片段，去除空白处理也可以减小待验证语音的时长与文件大小，降低服务器的处理量。当外界噪声过大时，例如用户在人流量较多或受其他音响设备影响的情况下，对待验证语音进行处理之前，还可以通过语音活动检测来判断待验证语音中是否存在可以验证的语音信息，如果不存在则可以输出验证失败的结果。此外，本领域技术人员容易想到的其他预处理方式也应当包含在本公开的保护范围内。

[55]在对待验证语音进行预处理后，可以将其输入到时延神经网络模型中，时延神经网络模型可以对待验证语音进行分帧识别，最终转换为对应的待验证文本。时延神经网络模型的训练和使用可以参考本领域技术人员熟知的方法，在此不再详述。

[56]在一示例性实施例中，上述验证码可以包括文本验证码，将待验证文本与验证码进行匹配可以包括以下步骤：将待验证文本与文本验证码匹配，以得到待验证文本的错字比例。

[57]相应的，步骤 S220 可以包括以下步骤：如果错字比例低于匹配阈值，则判断待验证语音是否为非机器。

[58]其中，错字比例可以是未匹配成功的字符数占总字符数的百分比。匹配可以有多种方式，例如正向匹配、逆向匹配或双向匹配等。匹配结果可能由于选择的匹配方式不同而不同。例如：如果文本验证码为“今天天气真好”，待验证文本为“今天气真好”，采用正向匹配的方式，其匹配结果的错字比例将会高于采用双向匹配得到的错字比例。通常情况下，采用双向匹配的方式能够提高服务器匹配计算的准确率，但也增加了服务器的计算任务，对服务器的配置具有较高的要求。正向匹配及逆向匹配对服务器的计算能力要求较低，处理量较少。可以根据实际情况选择合适的匹配方式。

[59]匹配阈值可以是设定的一个可接受的匹配错字比例的上限。考虑到服务器在将待验证语音转换待验证文本，以及匹配待验证文本与文本验证码时可能出现的误差，可以允许待验证文本与文本验证码之间存在一定的不一致。匹配阈值可以根据实际情况设定，例如验证码较短或干扰信息较少时，匹配阈值可以设置的较高，验证码较长或内容复杂时，匹配阈值可以设置的较低等，本实施例对此不做具体限定。当错字比例低于匹配阈值时，可以认为待验证文本与验证码匹配成功，也即待验证语音与验证码匹配成功，通过第一次验证，可以继续进行第二次验证，即对待验证语音进行人机判断。

[60]图 5 示出了本示例性实施例中一种语音信息验证方法的流程图。如图 5 所示，获取

待验证语音后，可以先判断获取的待验证语音的长度是否达到预设长度，如果不足预设长度，可以直接输出验证失败的结果，如果达到预设长度，则可以对待验证语音进行预处理，并通过时延神经网络模型转换为待验证文本。然后对待验证文本和验证码进行匹配，判断错字比例是否低于匹配阈值，如果错字比例未低于匹配阈值，则可以输出验证失败的结果，如果错字比例低于匹配阈值，则可以将待验证语音转换为目标频谱图，以进行后续的验证。

[61]在一示例性实施例中，将待验证语音转换为目标频谱图可以包括：通过短时傅里叶变换将待验证语音转换为目标频谱图。

[62]短时傅里叶变换可以将复杂的声音信号转换到频域中，然后通过频谱特性分析待验证语音的时域信号特征。举例而言，可以从待验证语音中提取多个瞬时片段，并按照时序排列，每个片段转换为频率-能量图像，从而得到多个目标频谱图排成的序列，可以在后续利用卷积神经网络模型进行处理。

[63]图 6 示出了本示例性实施例中一种语音信息验证方法的流程图。如图 6 所示，服务器可以由文本库获取预设文本并经过干扰处理得到验证码，由用户终端获取用户输入的待验证语音，并通过对待验证语音进行预处理后，利用时延神经网络模型将待验证语音转换为待验证文本，通过计算错字比例来匹配待验证文本和验证码。如果匹配成功，可以将待验证语音通过短时傅里叶变换转换为目标频谱图，进一步，再通过训练和测试完成的卷积神经网络模型对目标频谱图进行分析，得到人机分类结果，如果人机分类结果为非机器，则可以输出最终验证通过的结果。

[64]本公开的实例性实施例还提供了一种语音信息验证装置。参考图 7 所示，该装置 700 可以包括：信息获取模块 710，用于获取验证码以及用户终端发送的关于验证码的待验证语音；人机判断模块 720，用于响应于待验证语音与验证码匹配，判断待验证语音是否为非机器；语音验证模块 730，用于如果待验证语音为非机器，则确定待验证语音验证通过。

[65]在一示例性实施例中，人机判断模块 720 可以包括：频谱图转换单元，用于如果待验证语音与验证码匹配，则将待验证语音转换为目标频谱图；频谱图分析单元，用于通过卷积神经网络模型对目标频谱图进行分析，得到目标频谱图的人机分类结果；语音判断单元，用于基于人机分类结果，确定待验证语音是否为非机器。

[66]在一示例性实施例中，信息获取模块 710 还可以用于获取经过扰动处理的验证码。

[67]在一示例性实施例中，信息获取模块 710 可以包括：文本获取单元，用于获取预设文本，并将预设文本转换为目标图片；图片处理单元，用于通过对目标图片进行变形、增加噪点、变色及模糊化中的一种或多种处理，生成经过扰动处理的图像验证码。

[68]在一示例性实施例中，语音信息验证装置 700 还可以包括：语音长度检测单元，用于检测待验证语音的长度；预设长度判定单元，用于在待验证语音的长度小于一预设长度时，判定待验证语音验证失败，并向用户终端返回失败提示信息，以及在待验证语音的长度大于等于预设长度时，将待验证语音转换为待验证文本，并将待验证文本与验证码进行匹配。

[69]在一示例性实施例中，预设长度判定单元还可以包括：预处理子单元，用于对待验证语音进行声道转换、预加重、语音增强、去除空白中的一种或多种预处理；模型处理子单元，用于利用时延神经网络模型将预处理后的待验证语音转换为待验证文本。

[70]在一示例性实施例中，验证码可以包括文本验证码，预设长度判定单元还可以包括：文本匹配单元，用于将待验证文本与文本验证码匹配，以得到待验证文本的错字比例。在这种情况下，人机判断模块 720 可以用于如果错字比例低于匹配阈值，则判断待验证语音是否为非机器。

[71]上述各模块/单元的具体细节已经在对应的方法部分实施例中进行了详细的描述，因此此处不再赘述。

[72]本公开的示例性实施例还提供了一种能够实现上述方法的电子设备。

[73]所属技术领域的技术人员能够理解，本公开的各个方面可以实现为系统、方法或程序产品。因此，本公开的各个方面可以具体实现为以下形式，即：完全的硬件实施方式、完全的软件实施方式（包括固件、微代码等），或硬件和软件方面结合的实施方式，这里可以统称为“电路”、“模块”或“系统”。

[74]下面参照图 8 来描述根据本公开的这种示例性实施例的电子设备 800。图 8 显示的电子设备 800 仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

[75]如图 8 所示，电子设备 800 以通用计算设备的形式表现。电子设备 800 的组件可以包括但不限于：上述至少一个处理器 810、上述至少一个存储器 820、连接不同系统组件（包括存储器 820 和处理器 810）的总线 830、显示器 840。

[76]其中，存储器 820 存储有程序代码，程序代码可以被处理器 810 执行，使得处理器 810 执行本说明书上述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步

骤。例如，处理器 810 可以执行图 2 所示的步骤 S210~S230 等。

[77]存储器 820 可以包括易失性存储器形式的可读介质，例如随机存取存储器（Random Access Memory，RAM）821 和/或高速缓存存储器 822，还可以进一步包括只读存储器（Read-Only，ROM）823。

[78]存储器 820 还可以包括具有一组（至少一个）程序模块 825 的程序/实用工具 824，这样的程序模块 825 包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

[79]总线 830 可以为表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器、外围总线、图形加速端口、处理器或者使用多种总线结构中的任意总线结构的局域总线。

[80]电子设备 800 也可以与一个或多个外部设备 1000（例如键盘、指向设备、蓝牙设备等）通信，还可与一个或者多个使得用户能与该电子设备 800 交互的设备通信，和/或与使得该电子设备 800 能与一个或多个其它计算设备进行通信的任何设备（例如路由器、调制解调器等等）通信。这种通信可以通过输入/输出（Input/Output，I/O）接口 850 进行。并且，电子设备 800 还可以通过网络适配器 860 与一个或者多个网络（例如局域网（Local Area Network，LAN），广域网（Wide Area Network，WAN）和/或公共网络，例如因特网）通信。如图 8 所示，网络适配器 860 通过总线 830 与电子设备 800 的其它模块通信。应当明白，尽管图中未示出，可以结合电子设备 800 使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理器、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储系统等。

[81]通过以上的实施方式的描述，本领域的技术人员易于理解，这里描述的示例实施方式可以通过软件实现，也可以通过软件结合必要的硬件的方式来实现。因此，根据本公开实施方式的技术方案可以以软件产品的形式体现出来，该软件产品可以存储在一个非易失性存储介质（可以是 CD-ROM，U 盘，移动硬盘等）中或网络上，包括若干指令以使得一台计算设备（可以是个人计算机、服务器、终端装置、或者网络设备等）执行根据本公开示例性实施例的方法。

[82]本公开的示例性实施例还提供了一种计算机可读存储介质，其上存储有能够实现本说明书上述方法的程序产品。在一些可能的实施方式中，本公开的各个方面还可以实现为一种程序产品的形式，其包括程序代码，当程序产品在终端设备上运行时，程序代码

用于使终端设备执行本说明书上述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。

[83]参考图 9 所示，描述了根据本公开的示例性实施例的用于实现上述方法的程序产品900，其可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码，并可以在终端设备，例如个人电脑上运行。然而，本公开的程序产品不限于此，在本文件中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[84]程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。可读存储介质包括，但不限于：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（Erasable Programmable Read-only Memory，EPROM或闪存）、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[85]计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[86]可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF 等等，或者上述的任意合适的组合。

[87]可以以一种或多种程序设计语言的任意组合来编写用于执行本公开操作的程序代码，程序设计语言包括面向对象的程序设计语言—诸如 Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网（LAN）或广域网（WAN），连接到用户计算设备，或者，可以连接到外部计算设备（例如利用因特网服务提供商来通过因特网连接）。

[88]此外，上述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明，而不是限制目的。易于理解，上述附图所示的处理并不表明或限制这些处理的时间顺序。另外，也易于理解，这些处理可以是例如在多个模块中同步或异步执行的。

[89]应当注意，尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元，但是这种划分并非强制性的。实际上，根据本公开的示例性实施例，上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之，上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[90]本领域技术人员在考虑说明书及实践这里公开的发明后，将容易想到本公开的其他实施例。本申请旨在涵盖本公开的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本公开的真正范围和精神由权利要求指出。

[91]应当理解的是，本公开并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限。

权利要求书

1、一种语音信息验证方法，包括:

获取验证码以及用户终端发送的关于所述验证码的待验证语音;

响应于所述待验证语音与所述验证码匹配，判断所述待验证语音是否为非机器;

如果所述待验证语音为非机器，则确定所述待验证语音验证通过。

2、根据权利要求 1 所述的方法，其特征在于，判断所述待验证语音是否为非机器包括:

将所述待验证语音转换为目标频谱图;

通过卷积神经网络模型对所述目标频谱图进行分析，得到所述目标频谱图的人机分类结果;

基于所述人机分类结果，确定所述待验证语音是否为非机器。

3、根据权利要求 1 所述的方法，其特征在于，获取所述验证码包括:

获取经过扰动处理的验证码。

4、根据权利要求 3 所述的方法，其特征在于，获取所述经过扰动处理的验证码包括:

获取预设文本;

将所述预设文本转换为目标图片;

通过对所述目标图片进行变形、变色、模糊化及增加噪点中的一种或多种处理，生成所述经过扰动处理的图像验证码。

5、根据权利要求 1 所述的方法，其特征在于，在响应于所述待验证语音与所述验证码匹配之前，所述方法还包括:

检测所述待验证语音的长度;

如果所述待验证语音的长度小于预设长度，则判定所述待验证语音验证失败，并向所述用户终端返回失败提示信息;

如果所述待验证语音的长度大于等于所述预设长度，则将所述待验证语音转换为待验证文本，并将所述待验证文本与所述验证码进行匹配。

6、根据权利要求 5 所述的方法，其特征在于，将所述待验证语音转换为所述待验证文本包括:

对所述待验证语音进行声道转换、预加重、语音增强、去除空白中的一种或多种预处理;

利用时延神经网络模型将预处理后的所述待验证语音转换为所述待验证文本。

7、根据权利要求 5 所述的方法，其特征在于，所述验证码包括文本验证码，将所述待验证文本与所述验证码进行匹配包括:

将所述待验证文本与所述文本验证码匹配，以得到所述待验证文本的错字比例;

响应于所述待验证语音与所述验证码匹配，判断所述待验证语音是否为非机器包括:

如果所述错字比例低于匹配阈值，则判断所述待验证语音是否为非机器。

8、根据权利要求 2 所述的方法，其特征在于，将所述待验证语音转换为所述目标频谱图包括:

通过短时傅里叶变换将所述待验证语音转换为所述目标频谱图。

9、一种语音信息验证装置，包括:

信息获取模块，用于获取验证码以及用户终端发送的关于所述验证码的待验证语音;

人机判断模块，用于响应于所述待验证语音与所述验证码匹配，判断所述待验证语音是否为非机器;

语音验证模块，用于如果所述待验证语音为非机器，则确定所述待验证语音验证通过。

10、一种电子设备，包括:

处理器；以及

存储器，用于存储所述处理器可执行的可执行指令;

其中，当读取所述可执行指令时，所述处理器被促使执行权利要求 1-8 任一项所述的方法。

11、一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现权利要求 1-8 任一项所述的方法。

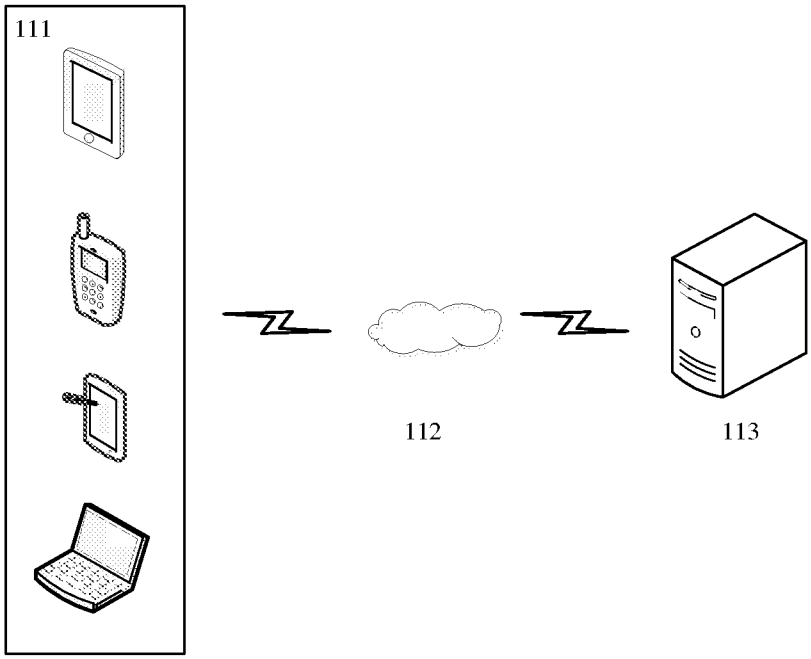

111
112
113

图 1

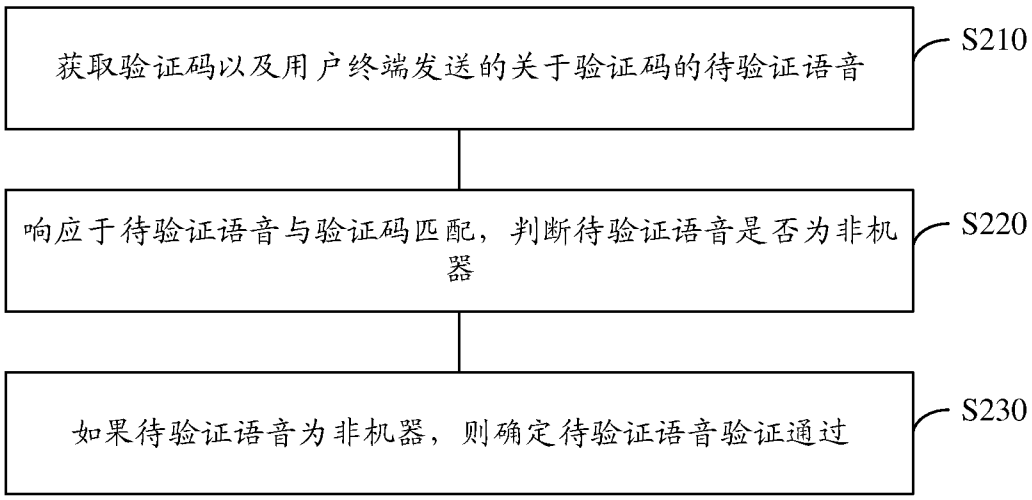

获取验证码以及用户终端发送的关于验证码的待验证语音
S210
响应于待验证语音与验证码匹配，判断待验证语音是否为非机器
S220
如果待验证语音为非机器，则确定待验证语音验证通过
S230

图 2

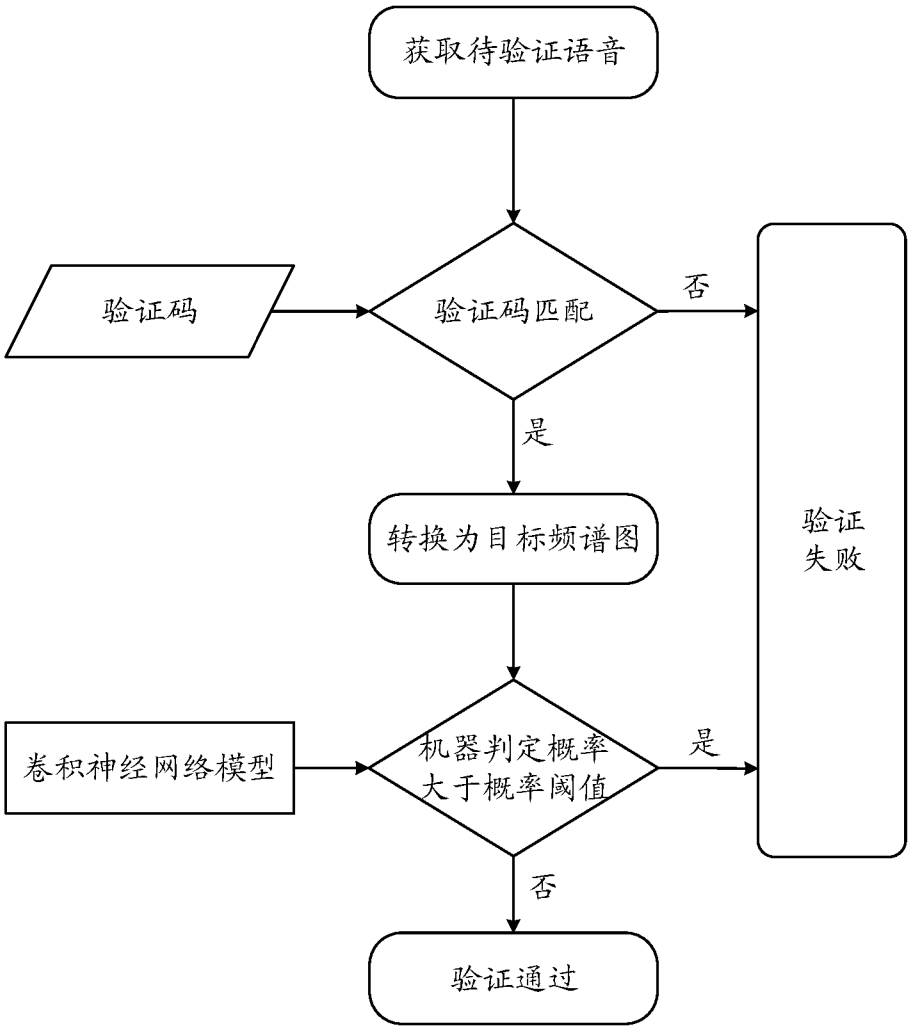

获取待验证语音
验证码
验证码匹配
否
是
转换为目标频谱图
验证
失败
卷积神经网络模型
机器判定概率
大于概率阈值
是
否
验证通过

图 3

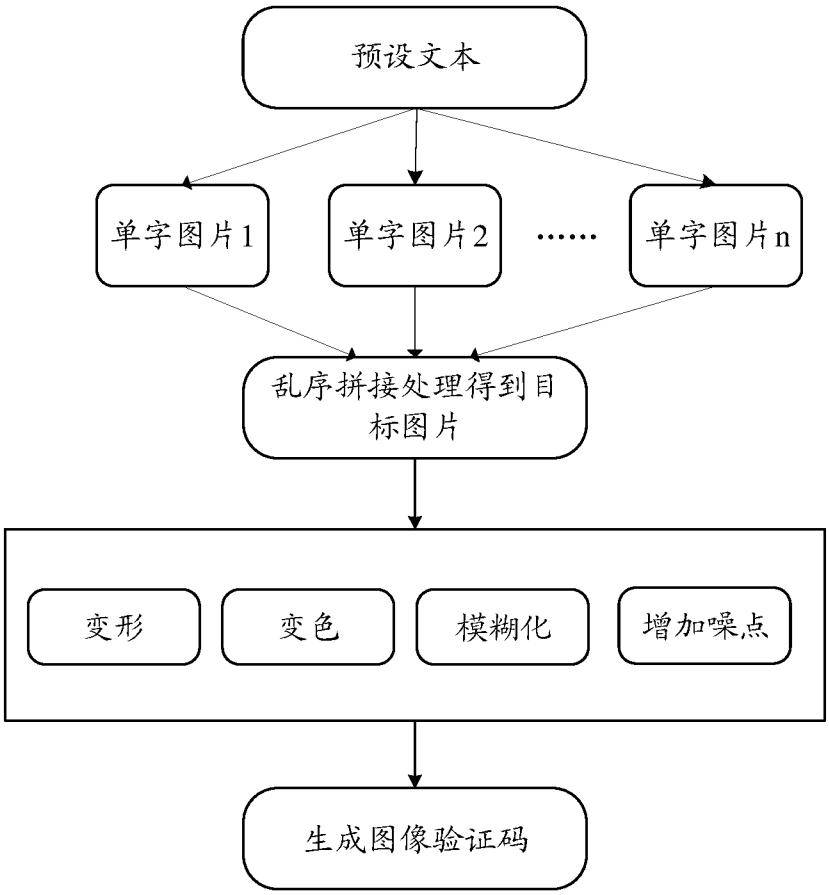


图 4

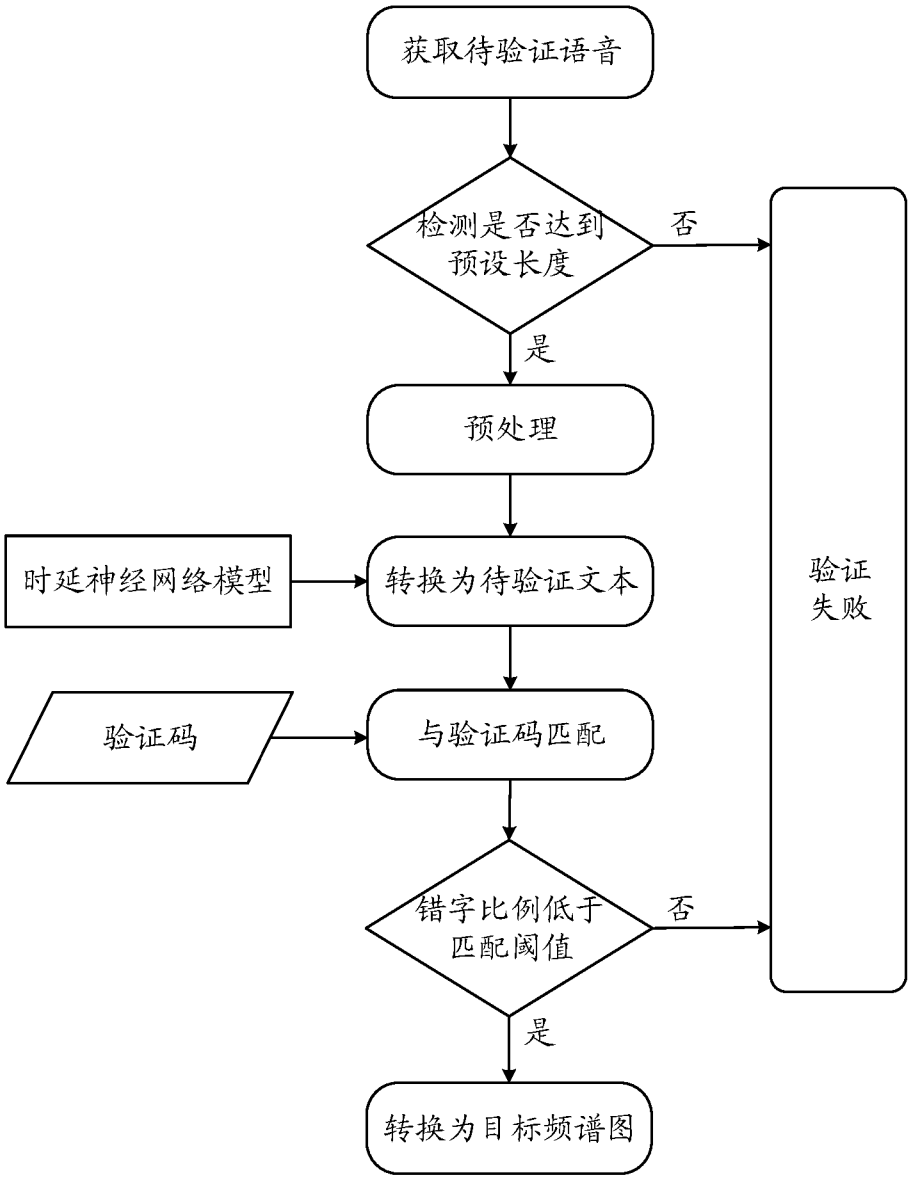

获取待验证语音
检测是否达到预设长度
否
是
预处理
时延神经网络模型
转换为待验证文本
验证码
与验证码匹配
错字比例低于匹配阈值
否
是
转换为目标频谱图
验证失败

图 5

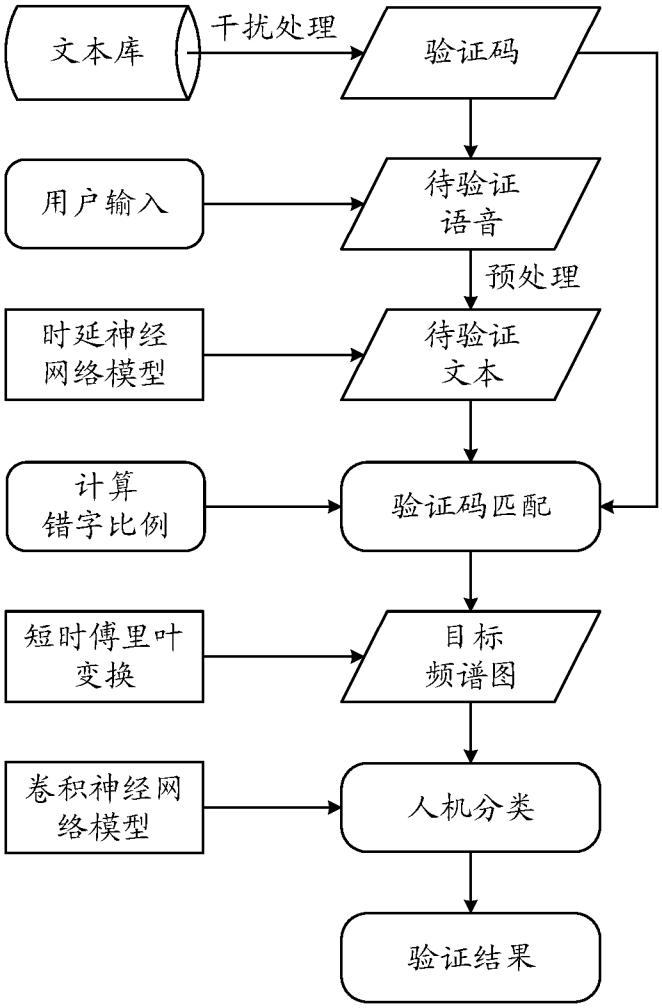

文本库
干扰处理
验证码
用户输入
待验证
语音
预处理
时延神经
网络模型
待验证
文本
计算
错字比例
验证码匹配
短时傅里叶
变换
目标
频谱图
卷积神经网
络模型
人机分类
验证结果

图 6

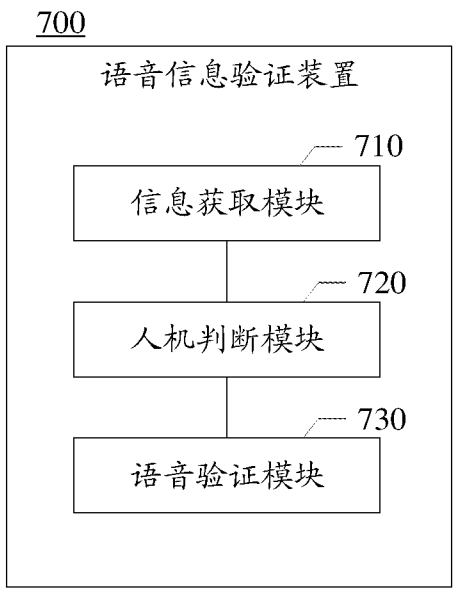

700
语音信息验证装置
710
信息获取模块
720
人机判断模块
730
语音验证模块

图 7

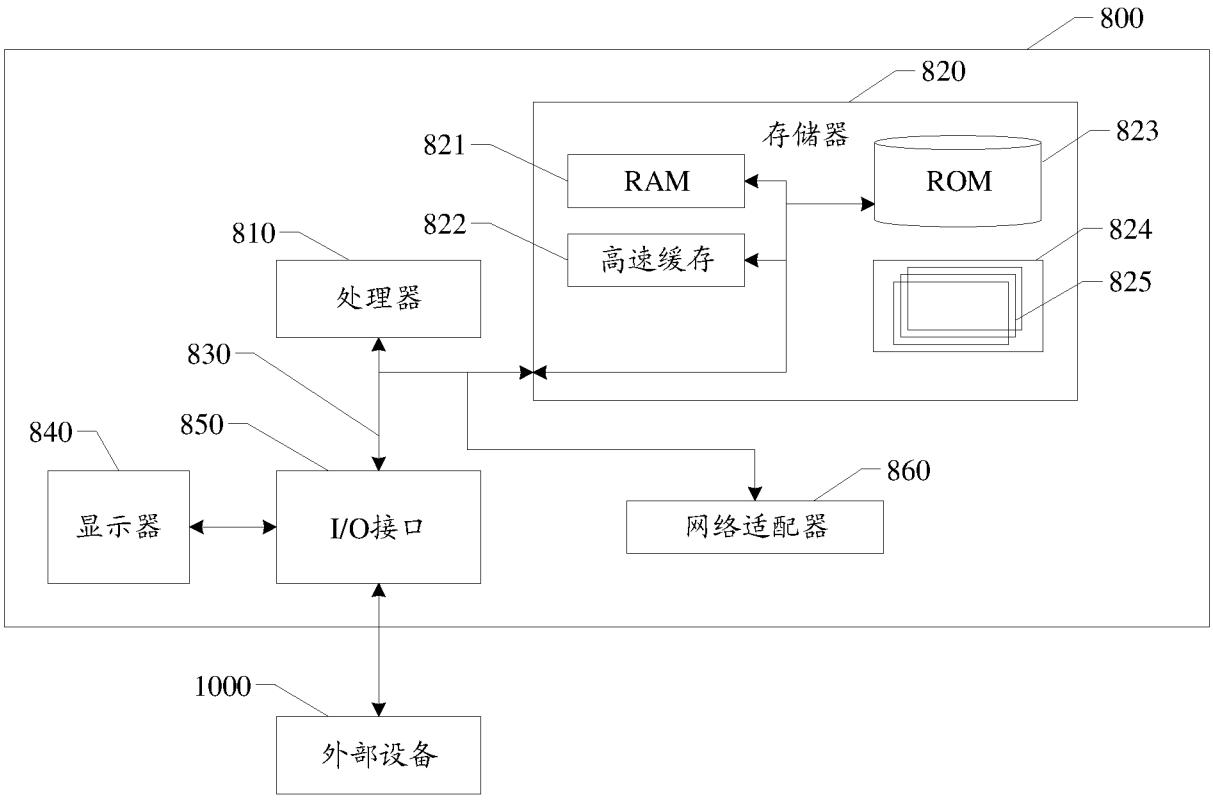

800
820
存储器
821
RAM
823
ROM
822
高速缓存
824
825
810
处理器
830
840
850
860
显示器
I/O接口
网络适配器
1000
外部设备

图 8

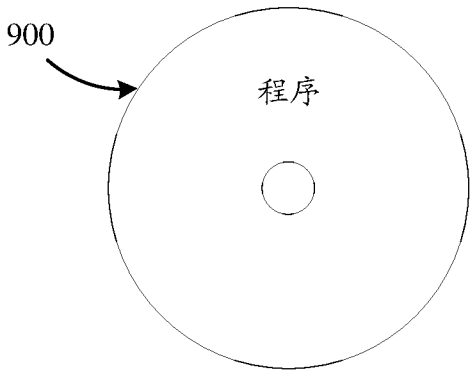

900
程序

图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/120972

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/24(2013.01)i; G10L 17/18(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)

G10L17

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)

CNKI, CNPAT, EPODOC, WPI: 验证码, 校验码, 人机, 机器, 语音, 声音, 声纹, 神经网络, 时延, 卷积, 长度, 错字, 匹配, 比例, 阈值, 频谱, 阈值, 频谱图, verification code, check code, man-machine, machine, voice, sound, voiceprint, neural network, convolution, length, match, scale, threshold, spectrum, threshold, spectrogram, CAPTCHA.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109493872 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 19 March 2019 (2019-03-19) description, paragraphs [0004]-[0113]	1-11
X	CN 104036780 A (ALIBABA GROUP HOLDING LIMITED) 10 September 2014 (2014-09-10) description, paragraphs [0006]-[0062]	1, 3-6, 9-11
X	CN 104104664 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 15 October 2014 (2014-10-15) description, paragraphs [0085]-[0104]	1, 3-6, 9-11
X	CN 106330915 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 11 January 2017 (2017-01-11) description, paragraphs [0048]-[0098]	1, 3-6, 9-11
X	CN 108877813 A (ALIBABA GROUP HOLDING LIMITED) 23 November 2018 (2018-11-23) description, paragraphs [0081]-[0108]	1, 3-6, 9-11

☑ 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	Date of mailing of the international search report
11 February 2020	**24 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.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2014307876 A1 (GOOGLE INC.) 16 October 2014 (2014-10-16) entire document	1-11
A	US 2016072803 A1 (YAHOO! INC.) 10 March 2016 (2016-03-10) entire document	1-11

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/120972

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109493872	A	19 March 2019	None			
CN	104036780	A	10 September 2014	CN	104036780	B	24 May 2017
				US	9571490	B2	14 February 2017
				EP	2965249	A2	13 January 2016
				TW	I592820	B	21 July 2017
				JP	2016511475	A	14 April 2016
				TW	I637285	B	01 October 2018
				TW	201730803	A	01 September 2017
				US	9258302	B2	09 February 2016
				US	2014259138	A1	11 September 2014
				WO	2014138129	A2	12 September 2014
				EP	2965249	B1	06 November 2019
				US	2016191516	A1	30 June 2016
				TW	201435640	A	16 September 2014
CN	104104664	A	15 October 2014	WO	2014166362	A1	16 October 2014
				US	9979721	B2	22 May 2018
				US	2016014120	A1	14 January 2016
CN	106330915	A	11 January 2017	None			
CN	108877813	A	23 November 2018	None			
US	2014307876	A1	16 October 2014	US	9263055	B2	16 February 2016
US	2016072803	A1	10 March 2016	US	9843583	B2	12 December 2017

国际检索报告

国际申请号 PCT/CN2019/120972

A. 主题的分类

G10L 17/24(2013.01)i; G10L 17/18(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)

G10L17

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））

CNKI,CNPAT,EPODOC,WPI:验证码,校验码,人机,机器,语音,声音,声纹,神经网络,时延,卷积,长度,错字,匹配,比例,阈值,频谱,阈值,频谱图,verification code, check code, man-machine, machine, voice, sound, voiceprint, neural network , convolution, length, match, scale, threshold, spectrum, threshold, spectrogram，CAPTCHA.

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 109493872 A (北京三快在线科技有限公司) 2019年 3月 19日 (2019-03-19) 说明书第[0004]-[0113]段	1-11
X	CN 104036780 A (阿里巴巴集团控股有限公司) 2014年 9月 10日 (2014-09-10) 说明书第[0006]-[0062]段	1，3-6，9-11
X	CN 104104664 A (腾讯科技深圳有限公司) 2014年 10月 15日 (2014-10-15) 说明书第[0085]-[0104]段	1，3-6，9-11
X	CN 106330915 A (百度在线网络技术北京有限公司) 2017年 1月 11日 (2017-01-11) 说明书第[0048]-[0098]段	1，3-6，9-11
X	CN 108877813 A (阿里巴巴集团控股有限公司) 2018年 11月 23日 (2018-11-23) 说明书第[0081]-[0108]段	1，3-6，9-11
A	US 2014307876 A1 (GOOGLE INC.) 2014年 10月 16日 (2014-10-16) 全文	1-11
A	US 2016072803 A1 (YAHOO! INC.) 2016年 3月 10日 (2016-03-10) 全文	1-11

☐ 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	国际检索报告邮寄日期
2020年 2月 11日	2020年 2月 24日

ISA/CN的名称和邮寄地址	受权官员
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	王玥
传真号 (86-10)62019451	电话号码 86-（10）- 53962545

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/120972

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	109493872	A	2019年 3月 19日	无			
CN	104036780	A	2014年 9月 10日	CN	104036780	B	2017年 5月 24日
				US	9571490	B2	2017年 2月 14日
				EP	2965249	A2	2016年 1月 13日
				TW	I592820	B	2017年 7月 21日
				JP	2016511475	A	2016年 4月 14日
				TW	I637285	B	2018年 10月 1日
				TW	201730803	A	2017年 9月 1日
				US	9258302	B2	2016年 2月 9日
				US	2014259138	A1	2014年 9月 11日
				WO	2014138129	A2	2014年 9月 12日
				EP	2965249	B1	2019年 11月 6日
				US	2016191516	A1	2016年 6月 30日
				TW	201435640	A	2014年 9月 16日
CN	104104664	A	2014年 10月 15日	WO	2014166362	A1	2014年 10月 16日
				US	9979721	B2	2018年 5月 22日
				US	2016014120	A1	2016年 1月 14日
CN	106330915	A	2017年 1月 11日	无			
CN	108877813	A	2018年 11月 23日	无			
US	2014307876	A1	2014年 10月 16日	US	9263055	B2	2016年 2月 16日
US	2016072803	A1	2016年 3月 10日	US	9843583	B2	2017年 12月 12日