



(12) 发明专利申请

(10) 申请公布号 CN 115547343 A

(43) 申请公布日 2022. 12. 30

(21) 申请号 202211204613.8

(22) 申请日 2022.09.29

(71) 申请人 中国电信股份有限公司

地址 100033 北京市西城区金融大街31号

(72) 发明人 王磊 梁伟 赵德欣 林悦

(74) 专利代理机构 北京律智知识产权代理有限公司 11438

专利代理师 孙宝海

(51) Int. Cl.

G10L 17/02 (2013.01)

G10L 15/34 (2013.01)

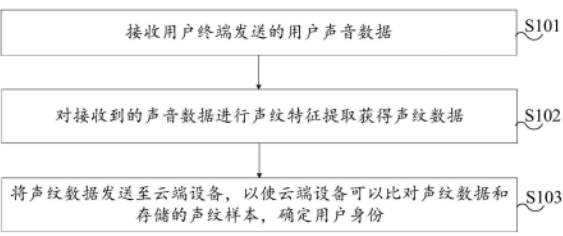
权利要求书2页 说明书11页 附图6页

(54) 发明名称

声纹识别方法、装置、边缘端设备及声纹识别系统

(57) 摘要

本申请涉及人工智能技术,揭示一种声纹识别方法、声纹识别装置、边缘端设备、计算机可读存储介质及声纹识别系统,该方案在用户终端获取用户的语音数据,在边缘端设备对语音数据进行声纹特征提取,获得声纹数据,在云端设备比对声纹数据和存储的声纹样本,从而确定用户身份。用户的语音内容不出边缘侧,对语音内容有更安全的保护;而且,通过传输声纹数据,对于语音内容不进行传输,减少了网络带宽的压力,降低网络延迟。



1. 一种声纹识别方法,用于边缘端设备,所述边缘端设备用于与用户终端和云端设备连接,其特征在于,所述声纹识别方法包括:

接收用户终端发送的用户声音数据;

对所述声音数据进行声纹特征提取获得声纹数据;

将所述声纹数据发送至云端设备,以使所述云端设备可以比对所述声纹数据和存储的声纹样本,确定用户身份。

2. 一种声纹识别装置,用于边缘端设备,所述边缘端设备用于与用户终端和云端设备连接,其特征在于,所述声纹识别装置包括:

接收模块,用于接收用户终端发送的用户声音数据;

特征提取模块,用于对所述声音数据进行声纹特征提取获得声纹数据;

发送模块,用于将所述声纹数据发送至所述云端设备,以使所述云端设备可以比对所述声纹数据和存储的声纹样本,确定用户身份。

3. 一种边缘端设备,用于与用户终端和云端设备连接,其特征在于,所述边缘端设备包括:

一个或多个处理器;

存储器,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述边缘端设备实现如权利要求1所述的声纹识别方法。

4. 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存储有计算机可读指令,当所述计算机可读指令被计算机的处理器执行时,使所述计算机执行如权利要求1所述的声纹识别方法。

5. 一种声纹识别方法,用于声纹识别系统,所述声纹识别系统包括用户终端、与所述用户终端连接的边缘端设备以及与所述边缘端设备连接的云端设备;其特征在于,所述声纹识别方法包括:

通过用户终端获取用户的声音数据,并将所述声音数据发送至边缘端设备;

通过所述边缘端设备对所述声音数据进行声纹特征提取获得声纹数据,并将所述声纹数据发送至云端设备;

通过所述云端设备比对所述声纹数据和存储的声纹样本,确定用户身份。

6. 根据权利要求5所述的声纹识别方法,其特征在于,所述通过用户终端获取用户的声音数据包括:

通过用户终端获取通话过程中包含通话双方的声音数据的语音流;所述通话双方包括所述用户和验证客服,所述用户的语音数据中包含有所述用户的IP地址,所述验证客服的语音数据中包含有所述验证客服的IP地址,且所述用户的IP地址与所述验证客服的IP地址不同;

根据IP地址,从所述语音流中获取用户的语音数据。

7. 根据权利要求5所述的声纹识别方法,其特征在于,所述云端设备根据地理区域建立多个声纹库,所述多个声纹库中的至少一者存储有用户对应于所述地理区域的声纹样本;

所述通过所述云端设备比对所述声纹数据和存储的声纹样本,确定用户身份,包括:

所述云端设备根据所述声纹数据对应的IP地址,确定用户所处的地理区域;

根据用户所处的地理区域,选择对应于所述地理区域的语音的声纹库;

比对所述声纹数据和所述声纹库的声纹样本,确定用户身份。

8. 根据权利要求5所述的声纹识别方法,其特征在于,在通过所述云端设备比对所述声纹数据和存储的声纹样本之前,所述声纹识别方法还包括:

通过用户终端采集各个用户的多条声音数据,并将多条所述声音数据发送至边缘端设备;

通过所述边缘端设备对多条所述声音数据进行声纹特征提取,并将多条所述声音数据对应的声纹数据发送至云端设备;

通过所述云端设备根据多条所述声音数据对应的声纹数据,确定各个用户所对应的声纹样本。

9. 根据权利要求5所述的声纹识别方法,其特征在于,在通过所述边缘端设备对所述声音数据进行声纹特征提取获得声纹数据之前,所述声纹识别方法还包括:

在所述云端设备训练神经网络模型,获得损失函数符合预设要求的神经网络模型;

将损失函数符合预设要求的神经网络模型下发至各所述边缘端设备,以使所述边缘端设备基于所述神经网络模型对所述声音数据进行声纹特征提取,获得声纹数据。

10. 一种声纹识别系统,其特征在于,包括用户终端、与所述用户终端连接的边缘端设备以及与所述边缘端设备连接的云端设备;

所述用户终端用于获取用户的声音数据,并将所述声音数据发送至所述边缘端设备;

所述边缘端设备用于对所述声音数据进行声纹特征提取获得声纹数据,并将所述声纹数据发送至所述云端设备;

所述云端设备用于比对所述声纹数据和存储的声纹样本,确定用户身份。

声纹识别方法、装置、边缘端设备及声纹识别系统

技术领域

[0001] 本申请涉及人工智能技术领域,特别涉及一种声纹识别方法、声纹识别装置、计算机可读存储介质、边缘端设备及声纹识别系统。

背景技术

[0002] 声纹是一种非接触生物特征,同人脸、虹膜、指纹、指静脉、掌纹等人体生物特征一样,具有身份唯一性,即通过说一句话可以唯一确定一个身份。

[0003] 声纹识别,相比较其他认证技术,更加方便,更加安全,因此,基于声纹的多模态的智能身份验证目前在智能对话等许多领域广泛使用。然而,现有技术中,往往是通过网络直接将用户的语音数据发送至例如云端服务器等具有强大的数据运算能力的设备,以提取用户的声纹特征并进行声纹比对。

[0004] 但是,实际应用当中经常碰到用户以及监管机构提出的数据隐私保护的实际需求,例如:金融领域基于声纹的身份认证,并不允许把用户的语音数据内容直接提交给验证系统进行无感的声纹身份注册和验证。

[0005] 因此,亟需提供一种声音数据隐私保护更佳的声纹识别方案,以解决上述现有技术存在的问题。

发明内容

[0006] 为了解决声音数据隐私保护的问题,本申请提供了一种声纹识别方法、声纹识别装置、计算机可读存储介质、边缘端设备及声纹识别系统。

[0007] 根据本申请实施例的一方面,公开了一种声纹识别方法,用于边缘端设备,所述边缘端设备用于与用户终端和云端设备连接。所述声纹识别方法包括:

[0008] 接收用户终端发送的用户声音数据;

[0009] 对所述声音数据进行声纹特征提取获得声纹数据;

[0010] 将所述声纹数据发送至云端设备,以使所述云端设备可以比对所述声纹数据和存储的声纹样本,确定用户身份。

[0011] 根据本申请实施例的一方面,公开了一种声纹识别装置,用于边缘端设备,所述边缘端设备用于与用户终端和云端设备连接。所述声纹识别装置包括接收模块、特征提取模块以及发送模块,其中,所述接收模块用于接收用户终端发送的用户声音数据;所述特征提取模块用于对所述声音数据进行声纹特征提取获得声纹数据;所述发送模块用于将所述声纹数据发送至所述云端设备,以使所述云端设备可以比对所述声纹数据和存储的声纹样本,确定用户身份。

[0012] 根据本申请实施例的一方面,公开了一种边缘端设备,用于与用户终端和云端设备连接。所述边缘端设备包括一个或多个处理器及存储器,所述存储器用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述边缘端设备实现如上所述的声纹识别方法。

[0013] 根据本申请实施例的一方面,公开了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机可读指令,当所述计算机可读指令被计算机的处理器执行时,使所述计算机执行如上所述的声纹识别方法。

[0014] 根据本申请实施例的一方面,公开了一种声纹识别方法,用于声纹识别系统,所述声纹识别系统包括用户终端、与所述用户终端连接的边缘端设备以及与所述边缘端设备连接的云端设备。所述声纹识别方法包括:

[0015] 通过用户终端获取用户的语音数据,并将所述语音数据发送至边缘端设备;

[0016] 通过所述边缘端设备对所述语音数据进行声纹特征提取获得声纹数据,并将所述声纹数据发送至云端设备;

[0017] 通过所述云端设备比对所述声纹数据和存储的声纹样本,确定用户身份。

[0018] 在一种示例性实施例中,所述通过用户终端获取用户的语音数据包括:

[0019] 通过用户终端获取通话过程中包含通话双方的语音数据的语音流;所述通话双方包括所述用户和验证客服,所述用户的语音数据中包含有所述用户的IP地址,所述验证客服的语音数据中包含有所述验证客服的IP地址,且所述用户的IP地址与所述验证客服的IP地址不同;

[0020] 根据IP地址,从所述语音流中获取用户的语音数据。

[0021] 在一种示例性实施例中,所述云端设备根据地理区域建立多个声纹库,所述多个声纹库中的至少一者存储有用户对应于所述地理区域的声纹样本。

[0022] 在一种示例性实施例中,所述通过所述云端设备比对所述声纹数据和存储的声纹样本,确定用户身份,包括:

[0023] 所述云端设备根据所述声纹数据对应的IP地址,确定用户所处的地理区域;

[0024] 根据用户所处的地理区域,选择对应于所述地理区域的语音的声纹库;

[0025] 比对所述声纹数据和所述声纹库的声纹样本,确定用户身份。

[0026] 在一种示例性实施例中,在通过所述云端设备比对所述声纹数据和存储的声纹样本之前,所述声纹识别方法还包括:

[0027] 通过用户终端采集各个用户的多条语音数据,并将多条所述语音数据发送至边缘端设备;

[0028] 通过所述边缘端设备对多条所述语音数据进行声纹特征提取,并将多条所述语音数据对应的声纹数据发送至云端设备;

[0029] 通过所述云端设备根据多条所述语音数据对应的声纹数据,确定各个用户所对应的声纹样本。

[0030] 在一种示例性实施例中,在通过所述边缘端设备对所述语音数据进行声纹特征提取获得声纹数据之前,所述声纹识别方法还包括:

[0031] 在所述云端设备训练神经网络模型,获得损失函数符合预设要求的神经网络模型;

[0032] 将损失函数符合预设要求的神经网络模型下发至各所述边缘端设备,以使所述边缘端设备基于所述神经网络模型对所述语音数据进行声纹特征提取,获得声纹数据。

[0033] 根据本申请实施例的一方面,公开了一种声纹识别系统,包括用户终端、与所述用户终端连接的边缘端设备以及与所述边缘端设备连接的云端设备;其中,所述用户终端用

于获取用户的声音数据,并将所述声音数据发送至所述边缘端设备;所述边缘端设备用于对所述声音数据进行声纹特征提取获得声纹数据,并将所述声纹数据发送至所述云端设备;所述云端设备用于比对所述声纹数据和存储的声纹样本,确定用户身份。

[0034] 本申请的实施例提供的技术方案至少包括以下有益效果:

[0035] 本申请提供的技术方案,在用户终端获取用户的声音数据,在边缘端设备对声音数据进行声纹特征提取,获得声纹数据,在云端设备比对声纹数据和存储的声纹样本,从而确定用户身份。用户的声音内容不出边缘侧,对声音内容有更安全的保护;而且,通过传输声纹数据,对于声音内容不进行传输,减少了网络带宽的压力,降低网络延迟。

[0036] 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的,并不能限制本申请。

附图说明

[0037] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。

[0038] 图1是根据一示例性实施例示出的声纹识别系统示意图;

[0039] 图2是根据一示例性实施例示出的一种声纹识别方法的流程图;

[0040] 图3是图2对应实施例中步骤S102的细节流程图;

[0041] 图4是图2对应实施例中步骤S102的细节逻辑框图;

[0042] 图5是根据一示例性实施例示出的一种声纹识别装置的框图;

[0043] 图6是根据一示例性实施例示出的一种边缘端设备的框图;

[0044] 图7是根据一示例性实施例示出的一种计算机系统的框图;

[0045] 图8是根据一示例性实施例示出的一种声纹识别方法的流程图;

[0046] 图9是图8对应实施例中步骤S201的细节流程图;

[0047] 图10是图9对应实施例中步骤S2012的细节流程图;

[0048] 图11是图8对应实施例中步骤S203的细节流程图;

[0049] 图12是根据一示例性实施例示出的一种声纹样本采集流程图;

[0050] 图13是根据另一示例性实施例示出的一种声纹识别方法的流程图。

[0051] 附图标记说明如下:

[0052] 100、声纹识别系统;101、用户终端;102、边缘端设备;103、云端设备;104、UPF;105、边缘端设备;106、网络;200、声纹识别装置;201、接收模块;202、特征提取模块;203、发送模块;300、边缘端设备;301、处理器;302、存储器;400、计算机系统;401、CPU;402、ROM;403、存储部分;404、RAM;405、总线;406、I/O接口;407、输入部分;408、输出部分;409、通信部分;410、驱动器;411、可拆卸介质。

具体实施方式

[0053] 尽管本申请可以容易地表现为不同形式的实施方式,但在附图中示出并且在本说明书中将详细描述的仅仅是其中一些具体实施方式,同时可以理解的是本说明书应视为是本申请原理的示范性说明,而并非旨在将本申请限制到在此所说明的那样。

[0054] 此外,本申请的描述中所提到的术语“包括”、“包含”、“具有”以及它的任何变形,

意图在于覆盖不排他的包含。例如包含了一系列步骤或模块的过程、方法、系统、产品或设备,没有限定于已列出的步骤或模块,而是可选的还包括其它没有列出的步骤或模块,或可选的还包括对于这些过程、方法、产品或设备固有的其它步骤或模块。

[0055] 此外,术语“第一”、“第二”、“第三”仅用于描述目的,而不能理解为指示或暗示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”、“第三”的特征可以明示或者隐含地包括一个或者更多个特征。

[0056] 在本申请的描述中,除非另有说明,“多个”的含义是指两个或两个以上。

[0057] 需要说明的是,本申请实施例中,“示例性”或者“举例地”等词用于表示作例子、例证或说明。本申请实施例中描述为“示例性”或者“举例地”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言,使用“示例性”或者“举例地”等词旨在以具体方式呈现相关概念。

[0058] 以下将详细地对示例性实施例进行说明。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反,它们仅是与发明内容中所详述的、本申请的一些方面相一致的装置和方法的例子。

[0059] 本申请示例性实施例提供了一种声纹识别方法、声纹识别装置、边缘端设备、计算机可读存储介质及声纹识别系统。图1示出了可以执行本申请示例性实施例的声纹识别方法的一种声纹识别系统示意图。如图1所示,声纹识别系统100可以包括用户终端101、边缘端设备102以及云端设备103,其中,用户终端101与边缘端设备102通讯连接,以实现用户终端101与边缘端设备102之间的信息交互,云端设备103与边缘端设备102通讯连接,以实现云端设备103与边缘端设备102之间的信息交互。

[0060] 详细地,用户终端101与边缘端设备102可以是通过有线方式连接,例如,通过RS485总线、USB线、以太网等连接。用户终端101与边缘端设备102也可以是通过无线方式连接,例如,通过WI-FI、蓝牙、ZigBee、LoRa (Long Range Radio,远距离无线电)等连接。

[0061] 详细地,边缘端设备102与云端设备103可以通过4G通讯、5G通讯等方式通讯连接。

[0062] 可以理解地,用户终端101可以例如是智能手机、电话手表、支持电话功能的平板电脑等。

[0063] 可以理解地,边缘端设备102可以是任意具有数据交互能力和数据处理能力的电子设备,例如安装有声纹特征提取程序的计算机等。同样地,云端设备103可以是任意具有信息交互能力和数据处理能力的电子设备,例如安装有声纹库和声纹比对程序的计算机等。

[0064] 本申请实施例提供一种声纹识别方法、声纹识别装置、边缘端设备、计算机可读存储介质及声纹识别系统,可以用于智能身份验证等场景。由于本申请中用户的语音内容不出边缘侧,对语音内容有更安全的保护,隐私保护更佳。本申请实施例提供的声纹识别方法、声纹识别装置、边缘端设备、计算机可读存储介质及声纹识别系统,具体通过如下实施例进行说明。

[0065] 参阅图2所示,本实施例提供了一种声纹识别方法,该声纹识别方法用于边缘端设备。该声纹识别方法包括以下步骤S101~S103。

[0066] S101,接收用户终端发送的用户声音数据。

[0067] S102,对接收到的声音数据进行声纹特征提取获得声纹数据。

[0068] 详细地,对接收到的声音数据进行声纹特征提取,包括对接收到的声音数据进行解码,再在解码之后的基础上进一步进行离散傅里叶变换等操作,进而提取出声纹特征。

[0069] 更详细地,如图3所示,步骤S102,包括以下步骤S1021~S1025。

[0070] S1021,对接收到的声音数据进行解码。

[0071] 可以理解地,在传输声音数据的过程中,会对声音数据进行编码,因此,在收到用户终端发送的声音数据之后,需要先对声音数据进行解码,再进行下面的声纹特征提取操作。

[0072] S1022,对解码后的声音数据进行预处理,获得第一数据。

[0073] 详细地,在步骤S1022中,对声音数据进行预处理,包括预加重、分帧、加窗。如图4所示,首先,对解码后的声音数据进行预加重;然后,对预加重之后的声音数据进行分帧;然后,对经过分帧的声音数据执行加窗操作。由于声音数据的很多特征都集中在高频段,而频率越高时,谱值反而越小,因此首先对声音数据的高频部分进行预加重处理,从而提高声音数据的高频分辨率。分帧可以获得声音数据的短时特征,而进一步对经过分帧的声音数据执行加窗操作,可以减少帧间的不连续性。

[0074] 可以理解地,对声音数据进行预处理,不限于是进行预加重、分帧、加窗操作。

[0075] S1023,对步骤S1022中获得的第一数据进行离散傅里叶变换(Discrete Fourier Transform,简称DFT),获得第二数据。

[0076] 举例地,对加窗操作后的声音数据逐层进行离散傅里叶变换,输出比特置换的复数数据,即第二数据。

[0077] S1024,对步骤S1023中获得的第二数据进行梅尔滤波,获得第三数据。

[0078] 举例地,对第二数据的能量值进行奇偶分离处理,对奇数时间索引或偶数时间索引的能量值进行多阶梅尔滤波,得到当前帧的多阶梅尔值,即第三数据。

[0079] S1025,对步骤S1024中获得的第三数据进行取对数操作,获得声纹数据。

[0080] S103,将声纹数据发送至云端设备,以使云端设备可以比对声纹数据和存储的声纹样本,确定用户身份。

[0081] 参阅图5所示,本实施例提供了一种声纹识别装置200,该声纹识别装置200用于边缘端设备,其包括接收模块201、特征提取模块202以及发送模块203。其中,接收模块201用于接收用户终端发送的用户声音数据;特征提取模块202用于对声音数据进行声纹特征提取获得声纹数据;发送模块203用于将声纹数据发送至云端设备,以使云端设备可以比对声纹数据和存储的声纹样本,确定用户身份。

[0082] 详细地,特征提取模块202首先对接收到的声音数据进行解码;再对解码后的声音数据进行预处理,获得第一数据,预处理包括预加重、分帧、加窗;之后再对第一数据进行离散傅里叶变换,获得第二数据;然后对第二数据进行梅尔滤波,获得第三数据;最后对第三数据进行取对数操作,获得声纹数据。

[0083] 上述声纹识别装置200中各个模块的功能和作用的实现过程具体详见上述声纹识别方法中对应步骤的实现过程,在此不再赘述。

[0084] 参阅图6所示,本实施例提供了一种边缘端设备300,该边缘端设备300包括一个或多个处理器301以及存储器302,存储器302用于存储一个或多个程序,当一个或多个程序被一个或多个处理器301执行时,使得边缘端设备300实现本申请的声纹识别方法。

[0085] 图7示意性地示出了用于实现本申请的一些实施例的计算机系统结构框图。需要说明的是,图7示出的计算机系统仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0086] 如图7所示,计算机系统400包括中央处理器401 (Central Processing Unit, CPU),其可以根据存储在只读存储器402 (Read-Only Memory, ROM) 中的程序或者从存储部分403加载到随机访问存储器404 (Random Access Memory, RAM) 中的程序而执行各种适当的动作和处理。在随机访问存储器404中,还存储有装置操作所需的各种程序和数据。中央处理器401、在只读存储器402以及随机访问存储器404通过总线405彼此相连。输入/输出接口406 (Input/Output接口,即I/O接口) 也连接至总线405。

[0087] 以下部件连接至输入/输出接口406:包括键盘、鼠标等的输入部分407;包括诸如阴极射线管 (Cathode Ray Tube, CRT)、液晶显示器 (Liquid Crystal Display, LCD) 等以及扬声器等的输出部分408;包括硬盘等的存储部分403;以及包括诸如局域网卡、调制解调器等网络接口卡的通信部分409。通信部分409经由诸如因特网的网络执行通信处理。驱动器410也根据需要连接至输入/输出接口406。可拆卸介质411,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器410上,以便于从其上读出的计算机程序根据需要被安装入存储部分403。

[0088] 特别地,根据本申请的实施例,各个方法流程图中所描述的过程可以被实现为计算机软件程序。例如,本申请的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分409从网络上被下载和安装,和/或从可拆卸介质411被安装。在该计算机程序被中央处理器401执行时,执行本申请的装置中限定的各种功能。

[0089] 需要说明的是,本申请实施例所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的装置、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (Erasable Programmable Read Only Memory, EPROM)、闪存、光纤、便携式紧凑磁盘只读存储器 (Compact Disc Read-Only Memory, CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中,计算机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装置、装置或者器件使用或者与其结合使用。而在本申请中,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,其中承载了计算机可读的程序代码。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。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行装置、装置或者器件使用或者与其结合使用的程序。计算机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、有线等等,或者上述的任意合适的组合。

[0090] 接下来参阅图8所示,本申请一示例性实施例提供的声纹识别方法,包括以下步骤

S201~S203。

[0091] S201,通过用户终端获取用户的声音数据,并将声音数据发送至边缘端设备。

[0092] 在通话场景中,由于通话双方的声音数据会混合在一起,因此,需要进行角色分离,将通话双方的声音数据区分开。举例地,用户作为接听者,验证客服为拨打者,需要获取用户(接听者)的声音数据,以进行进一步的处理从而验证用户的身份。

[0093] 在一示例性实施例中,用户的声音数据中包含有用户的IP地址,验证客服的声音数据中包含有验证客服的IP地址,且用户的IP地址与验证客服的IP地址不同,使用IP地址来识别用户。该方式无需通过人工智能算法进行角色分离,效率更高;且通过IP地址进行区分,角色划分也更加精准。

[0094] 在一个示例性实施例中,如图9所示,步骤S201,包括以下步骤S2011~S2013。

[0095] S2011,通过用户终端获取通话过程中包含用户和验证客服的声音数据的语音流。

[0096] 其中,用户的语音数据中包含有用户的IP地址,验证客服的语音数据中包含有验证客服的IP地址,且用户的IP地址与验证客服的IP地址不同。

[0097] S2012,根据IP地址,从语音流中获取用户的语音数据。

[0098] 详细地,每一条语音数据中包括有IP地址和语音内容,通过识别IP地址,即可区分用户和验证客服。

[0099] 举例地,可以采用socket(套接字)、scapy等实现用户语音数据的获取。

[0100] 更详细地,验证客服的IP地址相对固定,而用户是不断变化的,不同的用户,IP地址则不同。举例地,IP地址包括有四段,验证客服的IP地址的前三段相同,仅第四段因为不同验证客服席位不同而有所区别,在一个示例性实施例中,如图10所示,根据IP地址,从语音流中获取用户的语音数据包括以下步骤S20121~S20122。

[0101] S20121,识别语音流中IP地址与验证客服的IP地址前三段一致的语音数据。

[0102] S20122,将IP地址与验证客服的IP地址前三段一致的语音数据以外的语音数据作为当前用户的语音数据。

[0103] 通过识别验证客服的IP地址前三段,可以快速实现用户与验证客服的角色分离,从语音流中获取用户的语音数据;且,角色分离更加准确。

[0104] S2013,将用户的语音数据发送至边缘端设备。

[0105] 详细地,一个声纹识别系统中可以包含有多个边缘端设备,同一用户终端可以与多个边缘端设备建立通讯连接,在步骤S2013中,可以是将用户的语音数据发送至多个边缘端设备当中与当前用户终端在地理位置上距离最近的一边缘端设备,以减小语音数据的传输距离,从而提高语音数据的传输效率和确保传输质量。

[0106] 可以理解地,不限于是通过IP地址区分用户和验证客服,并进一步基于IP地址获取用户的语音数据,在其他实施例中,也可以采用其他方式从包含通话双方语音数据的语音流中获取用户的语音数据。

[0107] 前述实施例以双方通话场景进行说明,可以理解地,本申请并非仅仅适用于双方通话场景。

[0108] S202,通过边缘端设备对语音数据进行声纹特征提取获得声纹数据,并将声纹数据发送至云端设备。

[0109] 详细地,边缘端设备首先接收用户终端发送的语音数据,并对语音数据进行解码;

然后,对解码后的声音数据进行预处理,获得第一数据;再对获得的第一数据进行离散傅里叶变换,获得第二数据;之后对获得的第二数据进行梅尔滤波,获得第三数据;最后对获得的第三数据进行取对数操作,获得声纹数据。

[0110] S203,通过云端设备比对声纹数据和存储的声纹样本,确定用户身份。

[0111] 由于用户使用不同的语言说话时,声纹特征可能会有差异,因此,在一个示例性实施例中,根据地理区域在云端设备建立多个声纹库。在这一个示例性实施例中,多个声纹库中,至少有其中一个存储有用户对应于地理区域的声纹样本。

[0112] 可以理解地,同一声纹库可以存储有用户分别使用不同的语言说话时的声纹样本,也即,同一声纹库可以包含有多种语言的声纹样本。同样地,同一用户的声纹样本,可以是存储在多个声纹库中。

[0113] 举例地,对于以普通话为日常交流语言的地区,该地区的声纹库存储用户使用普通话说话时,对应的声纹特征;对于以方言为日常交流语言的地区,该地区的声纹库存储用户使用方言说话时,对应的声纹特征;对于以英语为日常交流语言的地区,该地区的声纹库存储用户使用英语说话时,对应的声纹特征;对于以普通话和方言为日常交流语言的地区,该地区的声纹库分别存储用户使用普通话说话时和使用方言说话时,对应的声纹特征。

[0114] 在一个示例性实施例中,如图11所示,步骤S203,包括以下步骤S2031~S2033。

[0115] S2031,云端设备根据声纹数据对应的IP地址,确定用户所处的地理区域。

[0116] S2032,根据用户所处的地理区域,选择对应于地理区域的语言的声纹库。

[0117] 可以理解地,地理区域的语言可以包括通用语言和除了通用语言以外的方言,其中,通用语言包括中国通用语言(如,普通话)、外国通用语言(如,英语)等;方言可以例如是粤语、客家话等。

[0118] S2033,比对声纹数据和声纹库的声纹样本,确定用户身份。

[0119] 可以理解地,若同一声纹库中存储有同一用户使用不同语言说话时的声纹样本,在步骤S2033中,还可以是结合用户的声纹数据与使用不同语言说话时对应的声纹样本的比对结果,来确定用户身份。

[0120] 接下来请参阅图12,图12示出了一示例性实施例声纹样本采集的流程图。在执行步骤S203通过云端设备比对声纹数据和存储的声纹样本,以确定用户身份之前,还需先采集声纹样本。如图12所示,在一示例性实施例中,采集声纹样本包括以下步骤S301~S303。

[0121] S301,通过用户终端采集各个用户的多条声音数据,并将多条声音数据发送至边缘端设备。

[0122] S302,通过边缘端设备对多条声音数据进行声纹特征提取,并将多条声音数据对应的声纹数据发送至云端设备。

[0123] S303,通过云端设备根据多条声音数据对应的声纹数据,确定各个用户所对应的声纹样本。

[0124] 通常地,是在用户的注册阶段采集用户的多条声音数据,多条声音数据作为注册声音数据,作为用户后续操作时身份验证的依据。一般地,注册阶段每一条声音数据的时长为6-8秒。

[0125] 由于用户在不同时间段或不同状态下说话的声纹特征有所不同,因此,通过用户终端采集各个用户的多条声音数据,获得的声纹样本更能准确反映用户的声纹特征,声纹

识别的泛化能力也更强。

[0126] 举例地,在一个示例性实施例中,在步骤S303中,通过云端设备获取各个用户的多条声音数据对应的声纹数据的均值,将均值作为各个用户所对应的声纹样本。当然,不限于是对多条声音数据对应的声纹数据进行求均值,来获得各个用户所对应的声纹样本。

[0127] 可以理解地,在执行步骤S202通过边缘端设备对声音数据进行声纹特征提取获得声纹数据之前,还需在边缘端设备设置神经网络模型,而神经网络模型需要先经过训练,以使得神经网络模型能够实现本申请中的声纹特征提取。参阅图13所示,在一个示例实施例中,声纹识别方法包括以下步骤S401~S405。

[0128] S401,在云端设备训练神经网络模型,获得损失函数符合预设要求的神经网络模型。

[0129] 详细地,损失函数采用AM-softmax函数,有更强的鲁棒性和泛化性。

[0130] S402,将损失函数符合预设要求的神经网络模型下发至各边缘端设备。

[0131] S403,通过用户终端获取用户的语音数据,并将声音数据发送至边缘端设备。

[0132] S404,通过边缘端设备采用神经网络模型对声音数据进行声纹特征提取获得声纹数据,并将声纹数据发送至云端设备。

[0133] S405,通过云端设备比对声纹数据和存储的声纹样本,确定用户身份。

[0134] 通过在云端设备统一训练神经网络模型,再分发更新到各个边缘端设备,可以保持神经网络模型的统一和实效性。

[0135] 综上,本申请采用基于云边端协同的声纹特征提取和验证方式,在用户的语音内容不出边缘侧的情况下实现了声纹验证;同时,无需通过安装插件监听网卡来抓取所有语音流,无需监听网卡及流经网卡的所有流量,能够适用于对信息隐私要求高的应用场景;再者,在云端设备上做声纹库的集约和声纹特征的比对,对声纹有统一安全的管控;此外,由于边缘端设备传输的是用户的声纹数据,对于语音内容不进行传输,减少了网络带宽的压力,降低网络延迟。

[0136] 接下来再请参阅图1,图1是根据一示例性实施例示出的声纹识别系统的示意图。如图1所示,声纹识别系统100包括用户终端101、边缘端设备102及云端设备103,用户终端101与边缘端设备102通讯连接,边缘端设备102与云端设备103通讯连接。

[0137] 详细地,用户终端101用于获取用户的语音数据,并将声音数据发送至边缘端设备102。边缘端设备102用于对声音数据进行声纹特征提取获得声纹数据,并将声纹数据发送至云端设备103。云端设备103用于比对声纹数据和存储的声纹样本,确定用户身份。

[0138] 可以理解地,除了语音验证业务之外,声纹识别系统100还可以涉及其他业务,如正常的通话业务等。声纹识别系统100中还设有UPF (User Plane Function,用户面功能) 104,除了用于声纹特征提取的边缘端设备102之外,还有用于其他业务处理的边缘端设备105,各个用户终端101与UPF104通讯连接,各个用户终端101发出的数据发送至UPF104,UPF104进行分流,将语音验证业务的语音数据转发至相应的边缘端设备102,而其他数据则转发至边缘端设备105,边缘端设备102再将数据经过网络106转发至云端设备103,边缘端设备105。

[0139] 详细地,用户终端101获取通话过程中包含用户和验证客服的语音数据的语音流,用户的语音数据中包含有用户的IP地址,验证客服的语音数据中包含有验证客服的IP地

址,且用户的IP地址与验证客服的IP地址不同;用户终端101根据IP地址,从声音流中获取用户的语音数据。

[0140] 详细地,边缘端设备102接收用户终端101发送的声音数据,并对声音数据进行解码;对解码后的声音数据进行预处理,获得第一数据,预处理包括预加重、分帧、加窗;对第一数据进行离散傅里叶变换,获得第二数据;对第二数据进行梅尔滤波,获得第三数据;对第三数据进行取对数操作,获得声纹数据。

[0141] 详细地,云端设备103根据地理区域建立多个声纹库,多个声纹库中的至少一者存储有用户对应于地理区域的声纹样本。

[0142] 详细地,云端设备103根据声纹数据对应的IP地址,确定用户所处的地理区域;根据用户所处的地理区域,选择对应于地理区域的语言的声纹库;比对声纹数据和声纹库的声纹样本,确定用户身份。

[0143] 详细地,用户终端101采集各个用户的多条声音数据,并将多条声音数据发送至边缘端设备102;边缘端设备102对多条声音数据进行声纹特征提取,并将多条声音数据对应的声纹数据发送至云端设备103;云端设备103根据多条声音数据对应的声纹数据,确定各个用户所对应的声纹样本。

[0144] 更详细地,云端设备103获取各个用户的多条声音数据对应的声纹数据的均值,将均值作为各个用户所对应的声纹样本。

[0145] 详细地,云端设备103还用于训练神经网络模型,获得损失函数符合预设要求的神经网络模型;并将损失函数符合预设要求的神经网络模型下发至各边缘端设备102,以使边缘端设备102基于神经网络模型对声音数据进行声纹特征提取,获得声纹数据。

[0146] 上述声纹识别系统100中各个组成部分的功能和作用更详细的实现过程具体详见上述声纹识别方法中对应步骤的实现过程,在此不再赘述。

[0147] 本申请提供的技术方案,在用户终端获取用户的语音数据,在边缘端设备对声音数据进行声纹特征提取,获得声纹数据,在云端设备比对声纹数据和存储的声纹样本,从而确定用户身份。用户的语音内容不出边缘侧,对语音内容有更安全的保护;而且,通过传输声纹数据,对于语音内容不进行传输,减少了网络带宽的压力,降低网络延迟。

[0148] 本领域技术人员应该可以意识到,在上述一个或多个示例中,本申请所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时,可以将这些功能存储在计算机可读存储介质中或者作为计算机可读存储介质上的一个或多个指令或代码进行传输。

[0149] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将装置的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。

[0150] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,模块地划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。例如多个单元或组件可以结合或者可以集成到另一个装置,或一些特征可以忽略,或不执行。

[0151] 应当理解的是,本申请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执行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

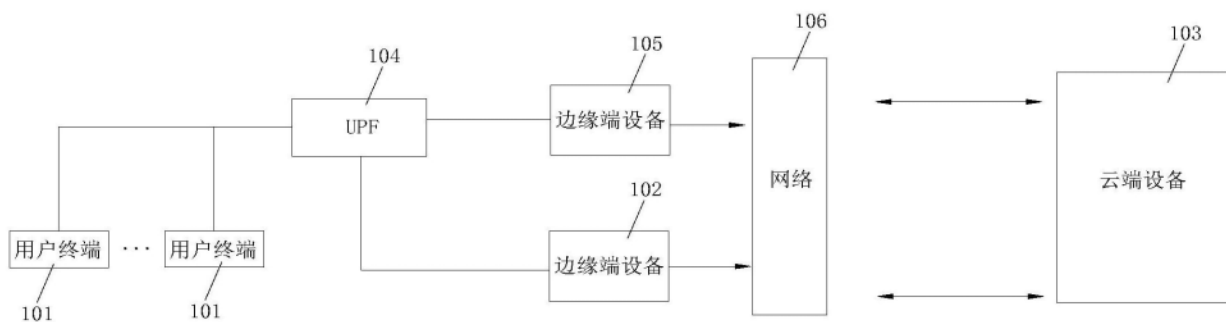


图1

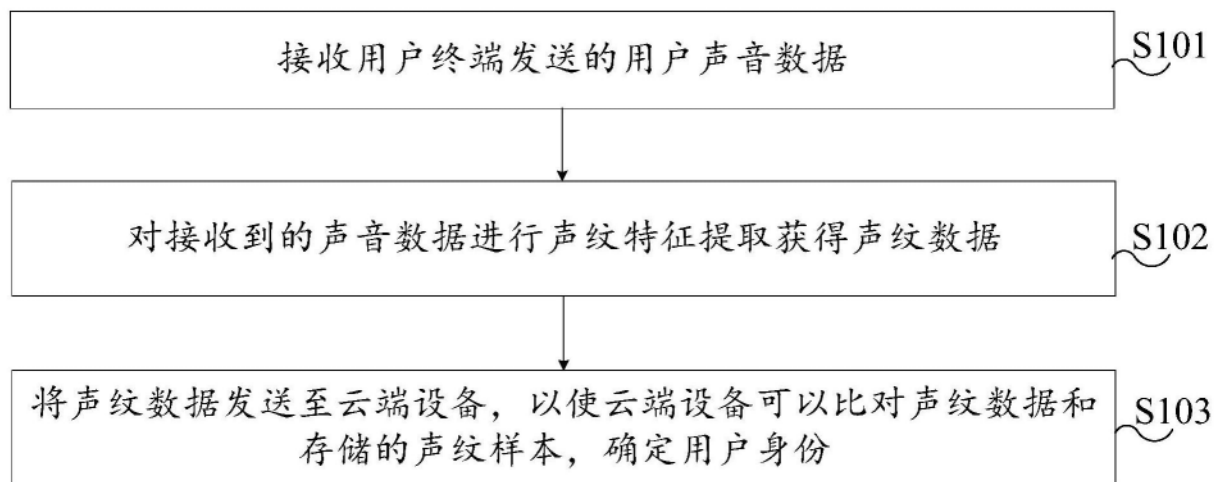


图2

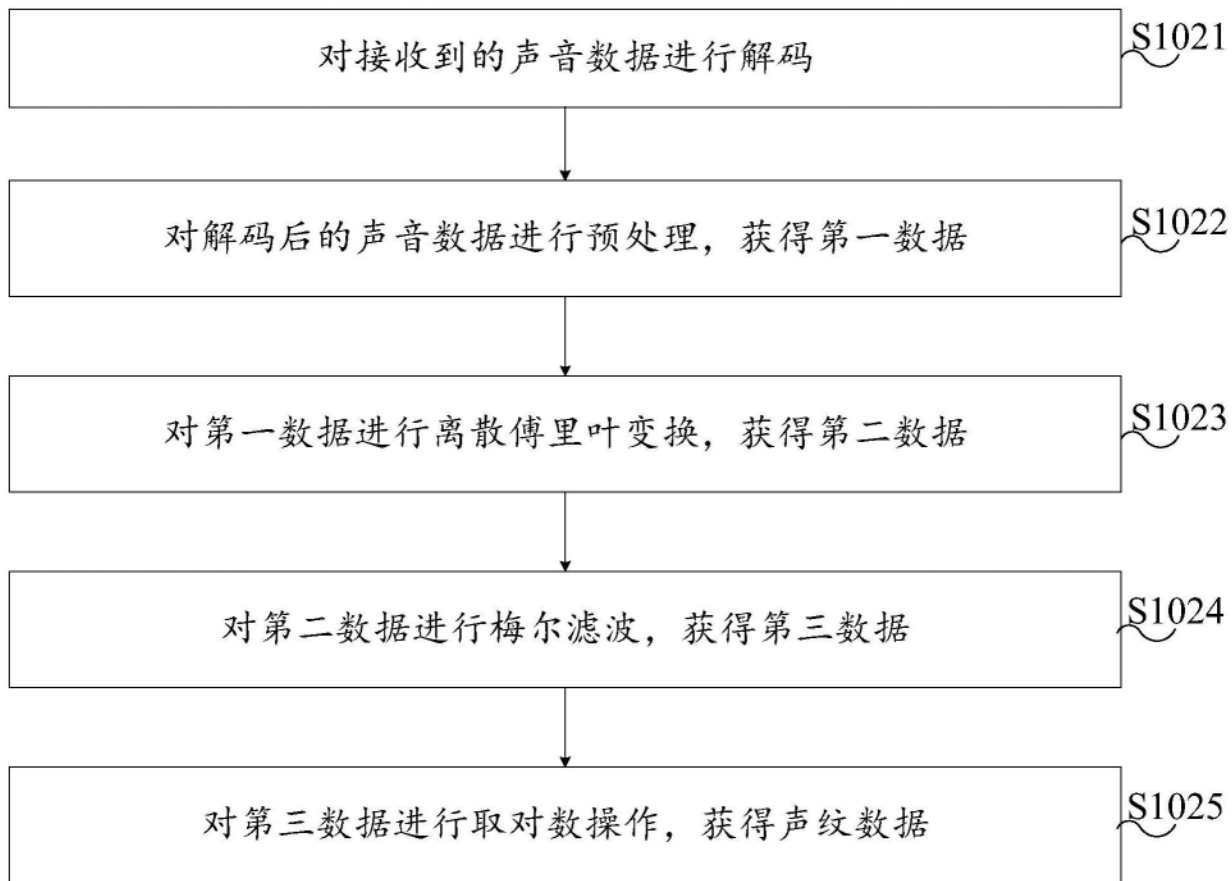


图3

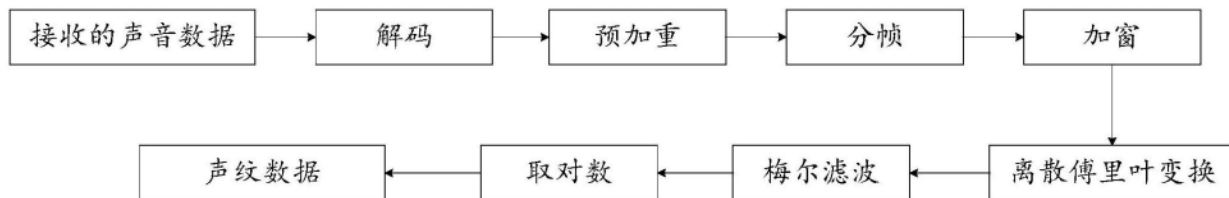


图4

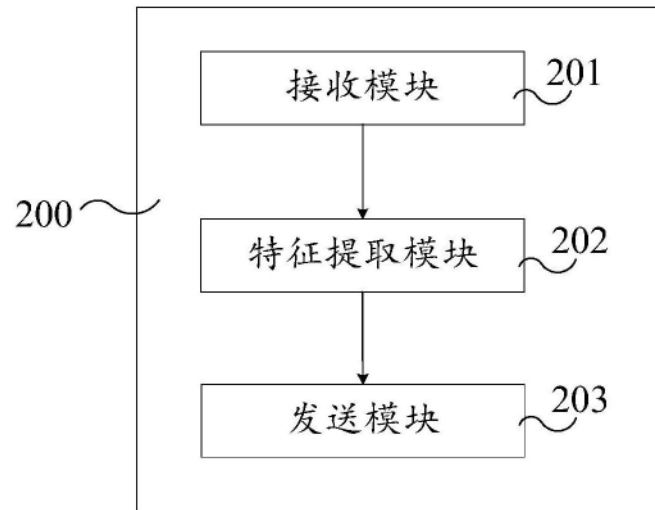


图5

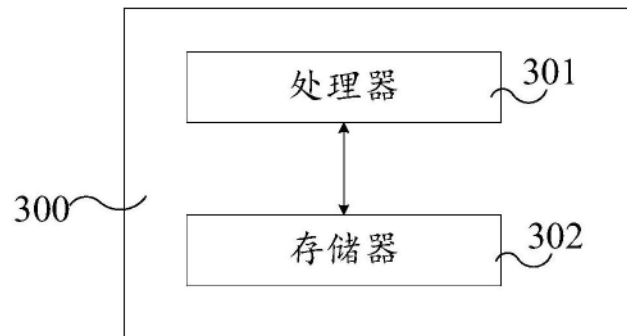


图6

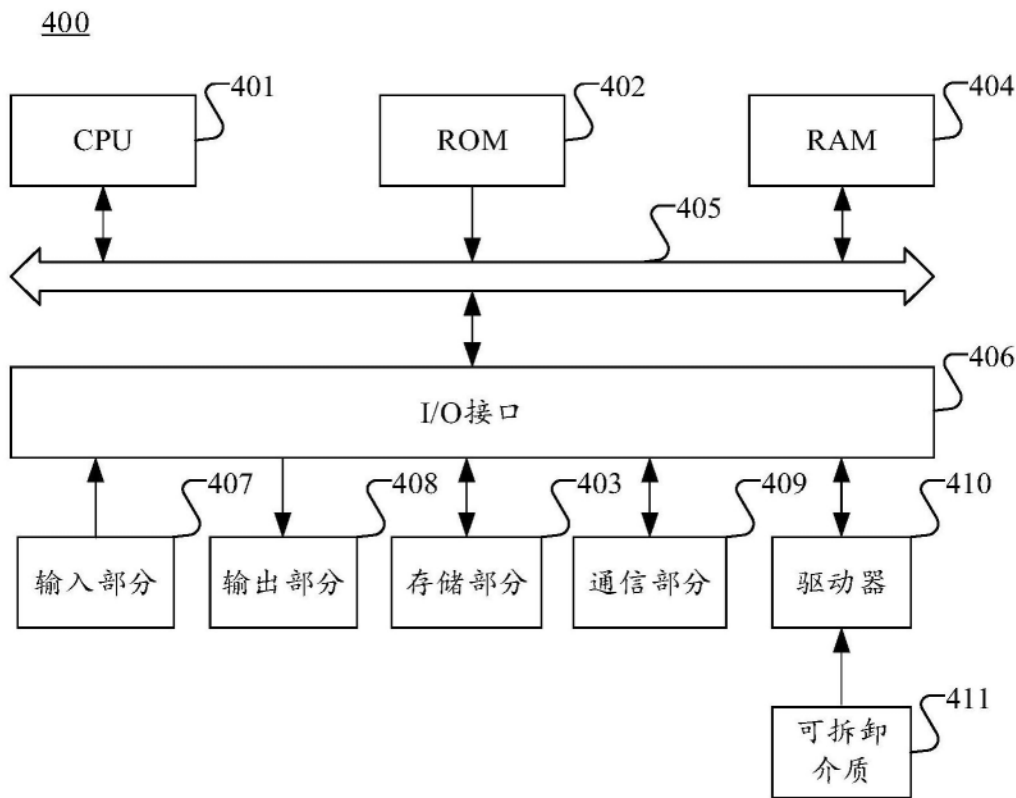


图7

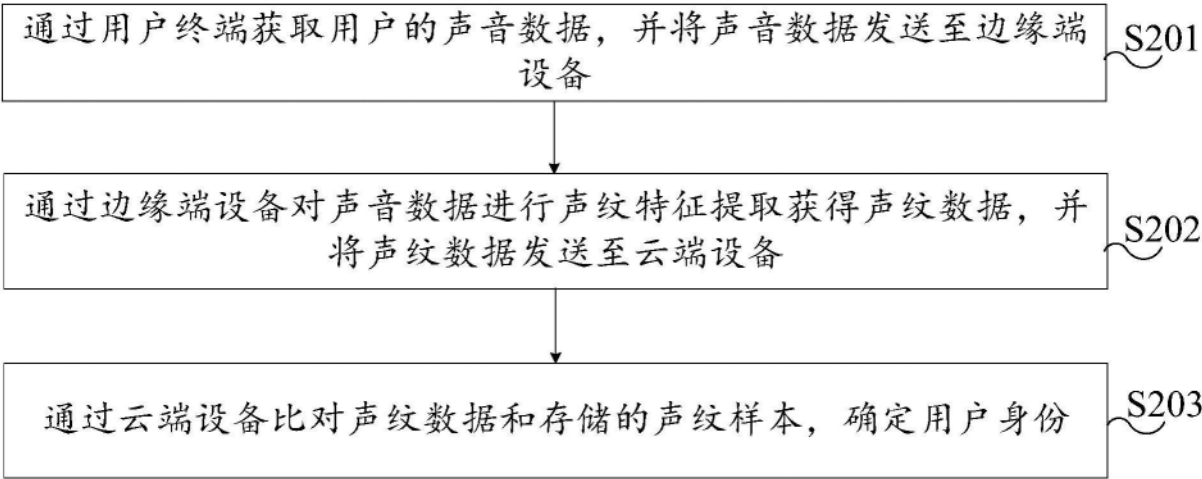


图8

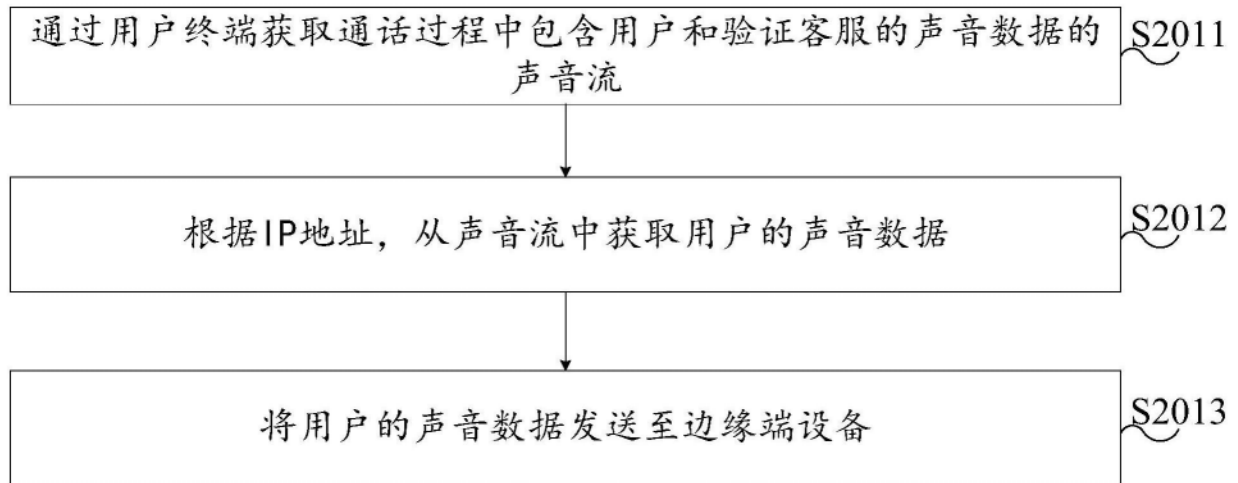


图9

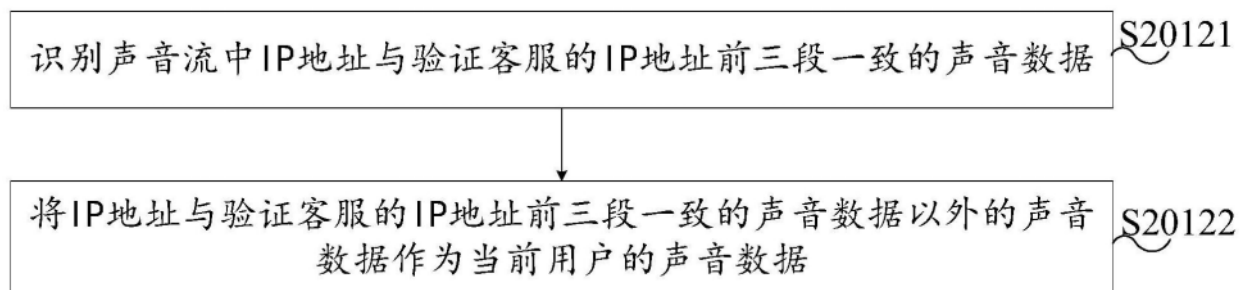


图10

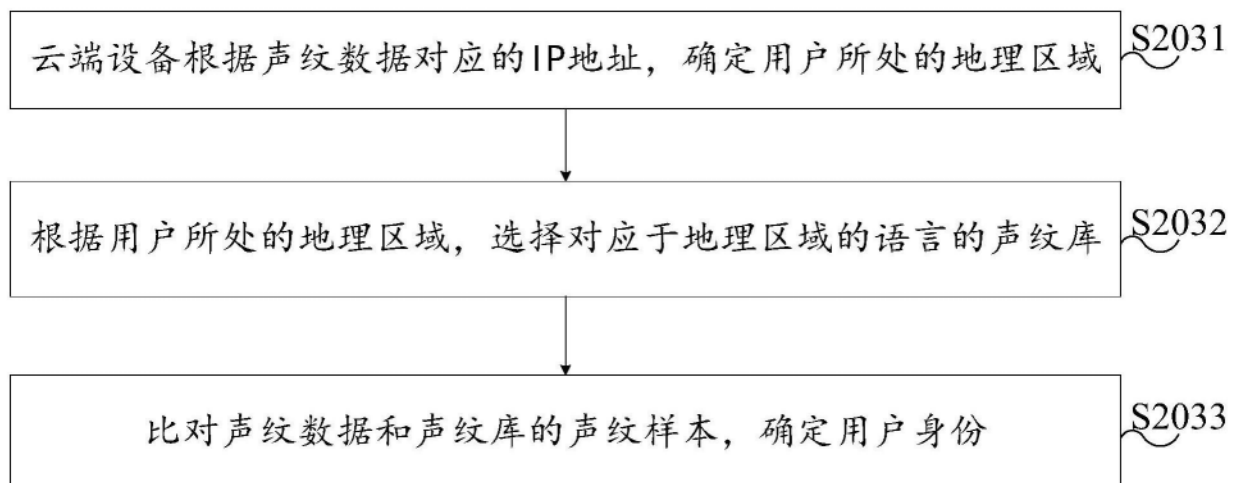


图11

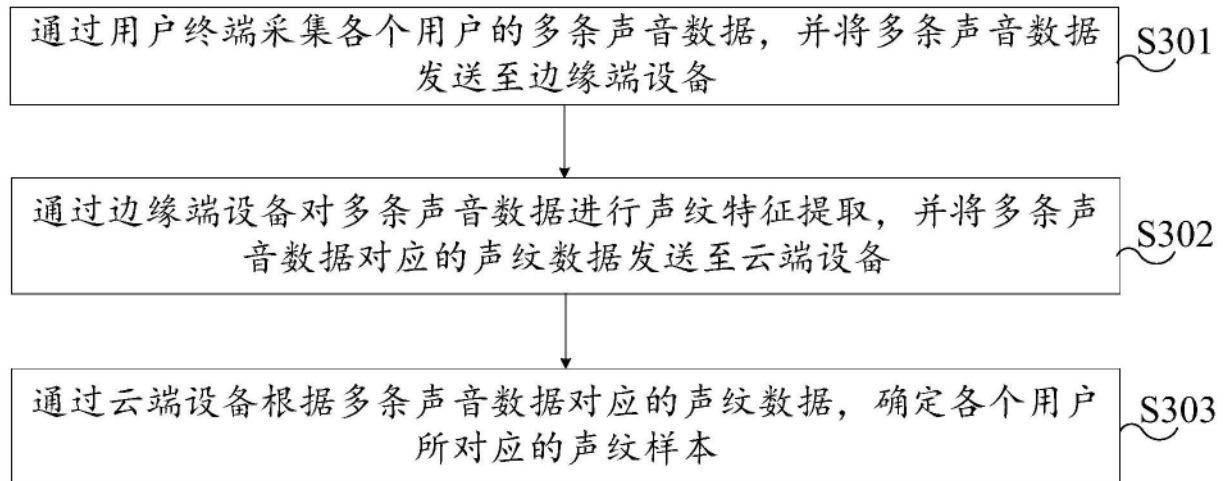


图12

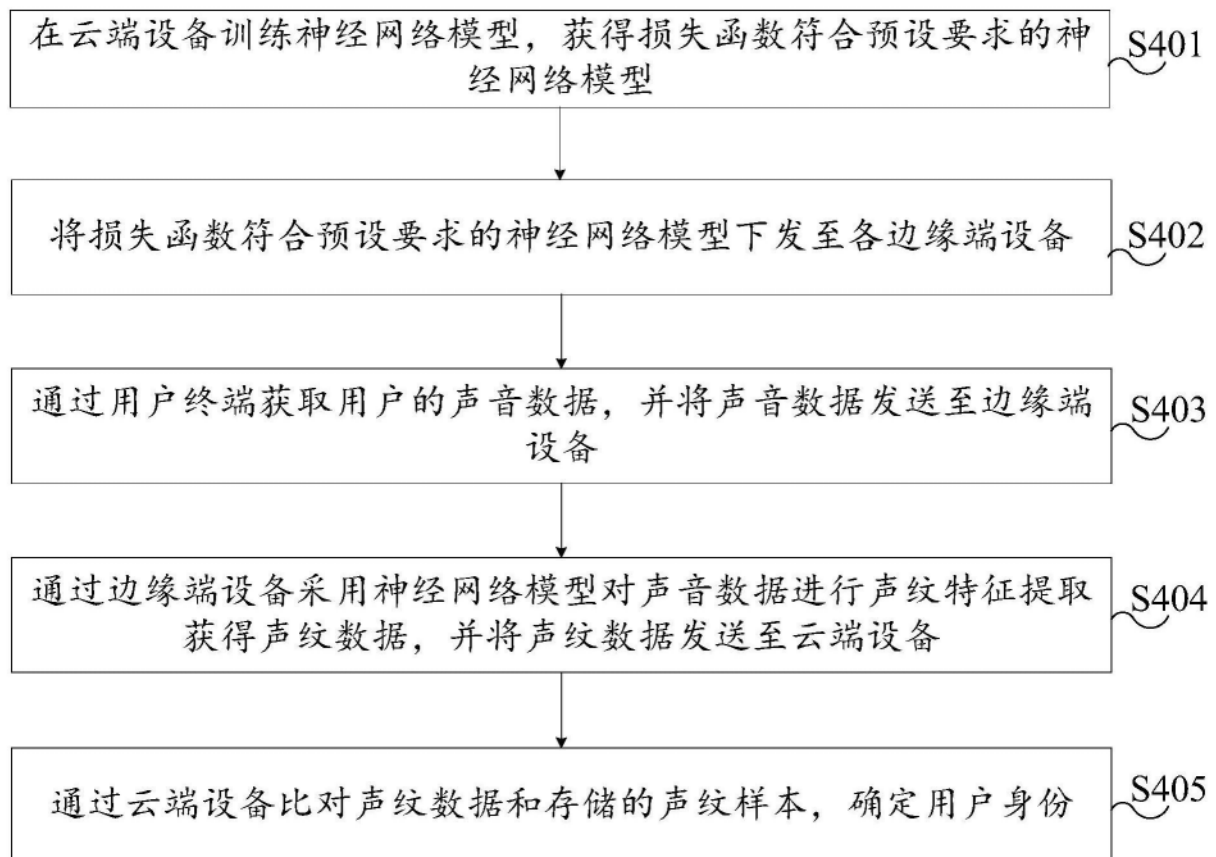


图13