



(12) 发明专利

(10) 授权公告号 CN 109378006 B

(45) 授权公告日 2022.09.16

(21) 申请号 201811620418.7

G10L 17/22 (2013.01)

(22) 申请日 2018.12.28

G10L 17/04 (2013.01)

G10L 15/22 (2006.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 109378006 A

(56) 对比文件

(43) 申请公布日 2019.02.22

CN 108288470 A, 2018.07.17

CN 108492830 A, 2018.09.04

(73) 专利权人 三星电子(中国)研发中心  
地址 210012 江苏省南京市雨花台区安德  
门大街57号6幢5-12楼

CN 108615537 A, 2018.10.02

CN 101047508 A, 2007.10.03

CN 108259280 A, 2018.07.06

专利权人 三星电子株式会社

US 2018/0286395 A1, 2018.10.04

CN 104639517 A, 2015.05.20

CN 105321520 A, 2016.02.10

(72) 发明人 武永超 陈洁

审查员 姚宇鹤

(74) 专利代理机构 北京德琦知识产权代理有限  
公司 11018

专利代理师 蒋欢 王琦

(51) Int. Cl.

G10L 17/24 (2013.01)

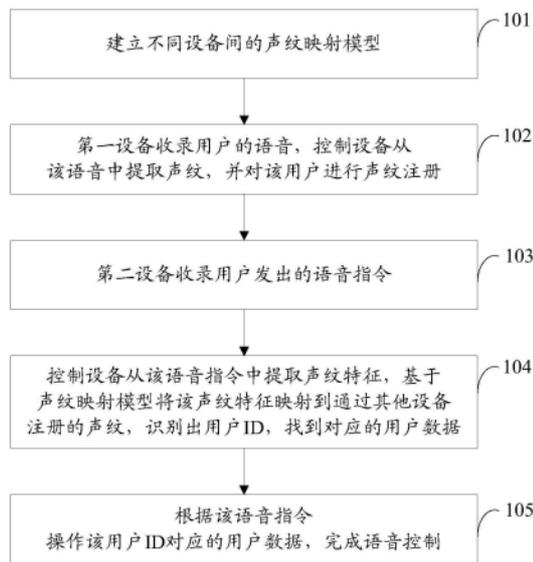
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种跨设备声纹识别方法及系统

(57) 摘要

本申请公开了一种跨设备声纹识别方法和系统,包括:建立不同设备之间的声纹映射模型;从第一设备收录的语音中提取声纹,进行声纹注册;从第二设备收录的语音指令中提取声纹特征,基于所建立的声纹映射模型将所述声纹特征映射到通过其他设备注册的声纹,识别出对应的用户ID;根据所述语音指令操作所述用户ID对应的用户数据,完成语音控制。应用本申请公开的技术方案,能够提高跨设备声纹识别的准确率。



1. 一种跨设备声纹识别方法,应用于控制设备,其特征在于,包括:  
建立不同设备之间的声纹映射模型;  
从第一设备收录的语音中提取声纹,进行声纹注册;  
从第二设备收录的语音指令中提取声纹特征,基于所建立的声纹映射模型将所述声纹特征映射到通过所述第一设备注册的声纹,识别出对应的用户ID;  
根据所述语音指令操作所述用户ID对应的用户数据,完成语音控制;  
所述建立不同设备之间的声纹映射模型包括:从各设备收录的语音中提取声纹,通过提取声纹特征进行建模,或者通过深度学习的方法进行建模,构建得到不同设备所收录语音的声纹之间映射关系的声纹映射模型。
2. 根据权利要求1所述的方法,其特征在于:  
所述进行声纹注册包括:将所述声纹与现有的用户ID进行绑定,或者将所述声纹作为用户ID。
3. 根据权利要求1所述的方法,其特征在于:  
所述第一设备收录的语音为:用户朗读的一段预设的提示文字。
4. 根据权利要求2所述的方法,其特征在于:  
在进行声纹注册之后,还包括:将播放历史数据导入对应的用户ID,  
在识别出对应的用户ID之后,还包括:读取所述用户ID的播放历史数据,根据所述播放历史数据进行内容推荐。
5. 一种跨设备声纹识别系统,其特征在于,包括:控制设备和至少一个具有麦克风的设备;  
所述至少一个具有麦克风的设备,用于收录语音,并发送给所述控制设备;其中,第一设备和第二设备是所述具有麦克风的设备;  
所述控制设备,用于建立不同设备之间的声纹映射模型;并用于从第一设备收录的语音中提取声纹,进行声纹注册;还用于从第二设备收录的语音指令中提取声纹特征,基于所建立的声纹映射模型将所述声纹特征映射到通过所述第一设备注册的声纹,识别出对应的用户ID;还用于根据所述语音指令操作所述用户ID对应的用户数据,完成语音控制;  
所述建立不同设备之间的声纹映射模型包括:从各设备收录的语音中提取声纹,通过提取声纹特征进行建模,或者通过深度学习的方法进行建模,构建得到不同设备所收录语音的声纹之间映射关系的声纹映射模型。
6. 根据权利要求5所述的系统,其特征在于,所述控制设备包括:声纹映射管理模块、声纹映射模型模块、声纹特征提取模块、用户系统模块和模型生成管理模块,其中:  
所述声纹映射管理模块,用于根据预先建立的不同设备之间的声纹映射模型,将当前输入的声纹映射成其他设备的声纹;  
所述声纹映射模型是预先生成的设备间的声纹映射模型;  
所述声纹特征提取模块,用于根据输入的语音数据,提取声纹特征;  
所述用户系统模块,用于进行声纹注册和用户ID管理;  
所述模型生成管理模块,用于构建不同设备之间的声纹映射模型。
7. 根据权利要求6所述的系统,其特征在于:  
所述声纹特征提取模块所提取的声纹特征是基于输入语音提取的所有声纹数据或者

部分声纹数据。

8. 根据权利要求6所述的系统,其特征在于:

所述声纹特征提取模块从各设备收录的语音中提取声纹,所述模型生成管理模块通过提取的声纹特征进行建模,或者通过深度学习的方法进行建模,构建得到不同设备所收录语音的声纹之间映射关系的声纹映射模型。

9. 根据权利要求6所述的系统,其特征在于:

所述用户系统模块具体包括:声纹注册子模块、声纹匹配子模块、用户ID子模块和用户数据子模块;其中:

所述声纹注册子模块,用于将新注册用户的声纹与用户系统中现有的用户ID相关联;

所述声纹匹配子模块,用于对接收到的声纹与已注册的声纹进行匹配,以确定对应的用户ID;

所述用户ID子模块,用于维护用户ID,其中,用户ID是用户的唯一标志;

所述用户数据子模块,用于维护用户数据,其中,用户数据是与用户相关的私有数据。

## 一种跨设备声纹识别方法及系统

### 技术领域

[0001] 本申请涉及声纹识别技术领域,特别涉及一种跨设备声纹识别方法及系统。

### 背景技术

[0002] 声纹(voiceprint):是用电声学仪器显示的携带言语信息的声波频谱。现代科学研究表明,声纹不仅具有特定性,而且有相对稳定性的特点。成年以后,人的声音可保持长期相对稳定不变。实验证明,无论讲话者是故意模仿他人的声音和语气,还是耳语轻声讲话,即使模仿得惟妙惟肖,两者的声纹却始终不相同。

[0003] 声纹识别:就是把声信号转换成电信号,再用计算机进行识别。

[0004] 声纹注册:用户注册一段语音,后台从这段语音中提取声纹作为用户ID。当用户说话时,提取其中的声纹,对应到相应的ID,从而实现用户区分。

[0005] 现有技术的主要问题在于:由于麦克风设备硬件存在差异,使得不同设备的声音采集质量参差不齐,进而导致声纹识别的准确率较低。例如:用户使用手机1注册的声纹,用手机2去做语音控制时,可能识别不出是同一个人。

### 发明内容

[0006] 本申请提供了一种跨设备声纹识别方法及系统,以提高跨设备声纹识别的准确率。

[0007] 本申请提供了一种跨设备声纹识别方法,应用于控制设备,包括:

[0008] 建立不同设备之间的声纹映射模型;

[0009] 从第一设备收录的语音中提取声纹,进行声纹注册;

[0010] 从第二设备收录的语音指令中提取声纹特征,基于所建立的声纹映射模型将所述声纹特征映射到通过其他设备注册的声纹,识别出对应的用户ID;

[0011] 根据所述语音指令操作所述用户ID对应的用户数据,完成语音控制。

[0012] 较佳的,所述建立不同设备之间的声纹映射模型包括:从各设备收录的语音中提取声纹,通过提取声纹特征进行建模,或者通过深度学习的方法进行建模,构建得到不同设备所收录语音的声纹之间映射关系的声纹映射模型。

[0013] 较佳的,所述进行声纹注册包括:将所述声纹与现有的用户ID进行绑定,或者将所述声纹作为用户ID。

[0014] 较佳的,所述第一设备收录的语音为:用户朗读的一段预设的提示文字。

[0015] 较佳的,在进行声纹注册之后,还包括:将播放历史数据导入对应的用户ID

[0016] 在识别出对应的用户ID之后,还包括:读取所述用户ID的播放历史数据,根据所述播放历史数据进行内容推荐。

[0017] 本申请还公开了一种跨设备声纹识别系统,包括:控制设备和至少一个具有麦克风的设备;

[0018] 所述至少一个具有麦克风的设备,用于收录语音,并发送给所述控制设备;其中,

第一设备和第二设备是所述具有麦克风的设备；

[0019] 所述控制设备,用于建立不同设备之间的声纹映射模型;并用于从第一设备收录的语音中提取声纹,进行声纹注册;还用于从第二设备收录的语音指令中提取声纹特征,基于所建立的声纹映射模型将所述声纹特征映射到通过其他设备注册的声纹,识别出对应的用户ID;还用于根据所述语音指令操作所述用户ID对应的用户数据,完成语音控制。

[0020] 较佳的,所述控制设备包括:声纹映射管理模块、声纹映射模型模块、声纹特征提取模块、用户系统模块和模型生成管理模块,其中:

[0021] 所述声纹映射管理模块,用于根据预先建立的不同设备之间的声纹映射模型,将当前输入的声纹映射成其他设备的声纹;

[0022] 所述声纹映射模型是预先生成的设备间的声纹映射模型;

[0023] 所述声纹特征提取模块,用于根据输入的语音数据,提取声纹特征;

[0024] 所述用户系统模块,用于进行声纹注册和用户ID管理;

[0025] 所述模型生成管理模块,用于构建不同设备之间的声纹映射模型。

[0026] 较佳的,所述声纹特征提取模块所提取的声纹特征是基于输入语音提取的所有声纹数据或者部分声纹数据。

[0027] 较佳的,所述声纹特征提取模块从各设备收录的语音中提取声纹,所述模型生成管理模块通过提取的声纹特征进行建模,或者通过深度学习的方法进行建模,构建得到不同设备所收录语音的声纹之间映射关系的声纹映射模型。

[0028] 较佳的,所述用户系统模块具体包括:声纹注册子模块、声纹匹配子模块、用户ID子模块和用户数据子模块;其中:

[0029] 所述声纹注册子模块,用于将新注册用户的声纹与用户系统中现有的用户ID相关联;

[0030] 所述声纹匹配子模块,用于对接收到的声纹与已注册的声纹进行匹配,以确定对应的用户ID;

[0031] 所述用户ID子模块,用于维护用户ID,其中,用户ID是用户的唯一标志;

[0032] 所述用户数据子模块,用于维护用户数据,其中,用户数据是与用户相关的私有数据。

[0033] 由上述技术方案可见,采用本发明提供的技术方案,能够屏蔽不同语音设备之间的差异,提高跨设备声纹识别的准确率,从而通过不同设备实现对同一设备的语音操作,给用户提供更好的体验。

## 附图说明

[0034] 图1为本发明跨设备声纹识别方法的流程示意图;

[0035] 图2为本发明跨设备声纹识别系统的组成结构示意图;

[0036] 图3为本发明实施例一跨麦克风设备进行音乐播放控制的流程示意图。

## 具体实施方式

[0037] 为使本申请的目的、技术方案及优点更加清楚明白,以下参照附图并举实施例,对本申请作进一步详细说明。

[0038] 本发明提出一种跨设备声纹识别方法及系统,通过预先建立不同设备之间的声纹映射模型,使得用户使用任何一种设备发出语音指令时,都能映射到该用户已注册的声纹,进而提高用户声纹的识别率。

[0039] 本发明提出的跨设备声纹识别方法流程图如图1所示,包含如下关键步骤:

[0040] 步骤101:建立不同设备(如电视语音遥控器、手机、录音笔等)之间的声纹映射模型。

[0041] 建立声纹映射模型的具体过程为:

[0042] 首先,采集同一个人不同设备上的语音数据,提取声纹,建立一条数据记录;

[0043] 然后,以上述方法采集多人的语音数据,并提取声纹,建立大的数据集;

[0044] 最后,根据建立的大数据集,以深度学习的方式,训练映射模型,将同一个人设备A上的声纹,映射到设备B上。反过来,用户使用不同的设备发出语音,根据该模型,也能判断是否为同一个人。

[0045] 假设 $f(\text{device}, \text{voiceprint})$ 是本发明声纹映射模型的映射函数,输入声纹 $\text{voiceprint}$ 和设备名称 $\text{device}$ ,就能得到基于 $\text{device}$ 设备的声纹数据。

[0046] 步骤102:第一设备收录用户的语音(例如:用户可以朗读一段预设的提示文字),控制设备从该语音中提取声纹,并对该用户进行声纹注册。

[0047] 这里,控制设备是指:在一个小型家庭网络中,用于进行后台控制和处理的设备。

[0048] 其中,声纹注册可以是将该声纹与现有的用户ID进行绑定,或者直接将该声纹作为用户ID。

[0049] 步骤103:第二设备收录用户发出的语音指令。

[0050] 步骤104:控制设备从该语音指令中提取声纹特征,基于所建立的声纹映射模型,将该声纹特征映射到通过其他设备注册的声纹,识别出对应的用户ID,从而找到对应的用户数据。

[0051] 步骤105:根据该语音指令操作该用户ID对应的用户数据,完成语音控制。

[0052] 对应于上述方法,本发明还提出了一种跨设备声纹识别系统,其组成结构如图2所示,该系统中除麦克风设备之外的其他模块都是控制设备中的模块,具体而言:

[0053] 1) 声纹映射管理模块

[0054] 用于根据预先建立的不同设备之间的声纹映射模型,将当前输入的声纹映射成其他设备的声纹。

[0055] 2) 声纹映射模型模块

[0056] 声纹映射模型是预先生成的设备间的声纹映射模型,声纹映射模型可以通过提取声纹特征建模生成,也可以通过深度学习训练生成,或者采用其他方式生成。

[0057] 3) 声纹特征提取模块

[0058] 用于根据输入的语音数据,提取声纹特征。声纹特征可以是基于输入语音提取的所有声纹数据,也可以是基于输入语音提取的部分声纹数据。

[0059] 4) 用户系统模块,具体包括:声纹注册子模块、声纹匹配子模块、用户ID子模块和用户数据子模块;其中:

[0060] A) 声纹注册子模块:用于将新注册用户的声纹与用户系统中现有的用户ID相关联。

[0061] B) 声纹匹配子模块:用于对接收到的声纹与已注册的声纹进行匹配,从而确定对应的用户ID。

[0062] C) 用户ID子模块:用于维护用户ID,其中,用户ID是用户的唯一标志。

[0063] D) 用户数据子模块:用于维护用户数据,其中,用户数据是与用户相关的私有数据,如访问历史、收藏夹等。

[0064] 5) 模型生成管理模块

[0065] 用于构建不同设备之间的声纹映射模型。

[0066] 下面通过两个较佳实施例对本申请上述技术方案进行进一步详细说明。

[0067] 实施例一:

[0068] 本实施例以跨麦克风设备进行音乐播放控制为例说明本发明实现流程,本实施例的流程示意图如图3所示。

[0069] 该流程的预处理部分包括:控制设备从各语音设备收录的语音中提取声纹,通过提取声纹特征进行建模,或者通过深度学习等方法进行建模,构建得到不同设备所收录语音的声纹之间映射关系的声纹映射模型。其中,可以由如图2所示的模型生成管理模块负责该预处理工作。

[0070] 图3所示跨麦克风设备进行音乐播放控制的实施流程如下:

[0071] 1) 用户对手机麦克风朗读一段提示文字,声纹注册模块将用户的声纹提取出来,并关联用户ID,或者直接将该声纹作为用户ID。

[0072] 2) 该用户在使用电视机时,用电视遥控器说“播放我的音乐”。此时,麦克风硬件已经更换,由于设备硬件参数不同,声纹数据有所不同。

[0073] 3) 声纹特征提取模块根据当前的电视遥控器录入的声音,提取对应的声纹。但是用户系统内的注册声纹为手机麦克风采集的声纹,所以需要进一步进行声纹匹配。

[0074] 4) 用户系统的声纹匹配模块,通过声纹映射管理模块读取声纹映射模型,产生相同声音在其他设备上的声纹,并与系统内已经注册的声纹进行匹配。

[0075] 5) 在用户系统内,寻找对应的用户ID。如果该用户ID不存在,则匹配失败,给出相应的提示;如果匹配成功,则进行下一步处理。

[0076] 采用本发明进行上述处理后,如果寻找到对应的用户ID,接下来可以按照现有技术进行后续的语音指令解析和设备控制,具体包括:

[0077] 6) 自然语音处理模块解析用户的语音指令,将其翻译为设备可以执行的命令,其他执行单元根据该命令执行对应的操作。

[0078] 7) 执行单元访问该用户的收藏音乐数据并播放。

[0079] 实施例二:

[0080] 本实施例以推荐用户喜欢的影视剧节目为例说明本发明实现流程,如下:

[0081] 预处理部分与实施例一相同,在此不再赘述。

[0082] 本实施例的实施流程如下:

[0083] 1) 用户1通过语音播放武侠类电视剧,例如说“播放金庸的神雕侠侣”。

[0084] 2) 声纹提取模块提取用户1的声纹,并与播放历史数据相关联。如果用户系统中不存在该声纹,则创建新的用户ID;如果用户系统中存在该声纹,则将播放历史数据导入用户系统中对应的用户ID。

[0085] 3) 用户2通过语音播放宫廷剧,例如说“我要看甄嬛传”。

[0086] 4) 声纹提取模块提取用户2的声纹,与该用户的播放历史数据相关联,并对用户2执行与上述步骤2) 相同的操作。

[0087] 5) 假设本实施例中的电视安装了基于用户播放历史的影视剧推荐系统,该推荐系统可以根据用户看过什么来猜测用户喜好,并给出推荐的影视剧。

[0088] 6) 用户1在初始状态下,说“播放我喜欢的影视剧”。

[0089] 7) 声纹特征提取模块根据当前的声音,生成对应的声纹。

[0090] 8) 用户系统的声纹匹配模块,通过声纹映射管理模块读取声纹映射模型,产生相同声音在其他设备上的声纹,并与系统内已经注册的声纹进行匹配,找到用户1对应的用户ID。

[0091] 9) 推荐系统读取用户1的历史播放数据“神雕侠侣”,给出推荐影视剧“射雕英雄传”、“侠客行”等。

[0092] 以上所述仅为本申请的较佳实施例而已,并不用以限制本申请,凡在本申请的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本申请保护的范围之内。

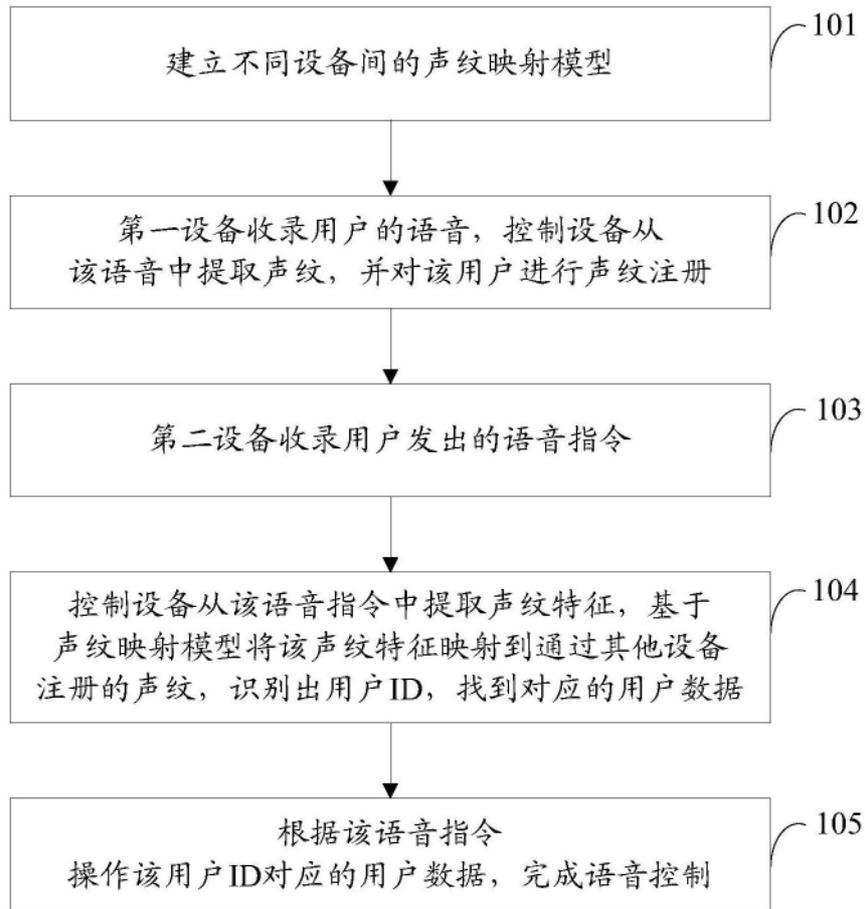


图1

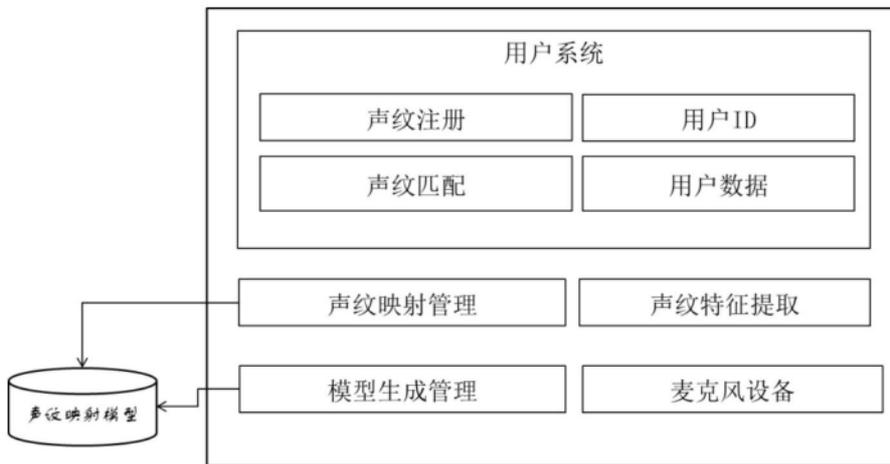


图2

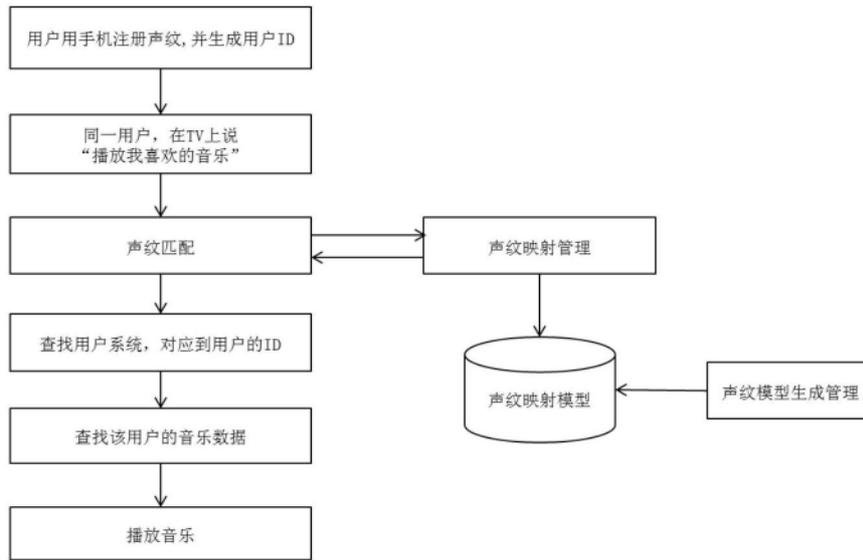


图3