(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107527620 B
(45)授权公告日 2019.03.26

(21)申请号 201710614649.6

(22)申请日 2017.07.25

(65)同一申请的已公布的文献号
申请公布号 CN 107527620 A

(43)申请公布日 2017.12.29

(73)专利权人 平安科技(深圳)有限公司
地址 518000 广东省深圳市福田区八卦岭工业区平安大厦六楼

(72)发明人 王健宗 郭卉 肖京

(74)专利代理机构 深圳市沃德知识产权代理事务所(普通合伙) 44347
代理人 于志光 高杰

(51)Int.Cl.
G10L 17/02(2013.01)
G10L 17/04(2013.01)
G10L 17/18(2013.01)
G10L 25/24(2013.01)
G06F 21/32(2013.01)

(56)对比文件
CN 106448684 A,2017.02.22,
CN 104008751 A,2014.08.27,
CN 105788592 A,2016.07.20,
CN 105869644 A,2016.08.17,
CN 106710599 A,2017.05.24,
US 2016293167 A1,2016.10.06,

审查员 王玥

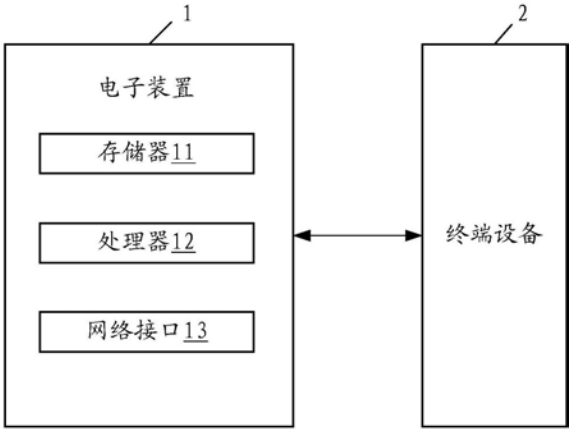
权利要求书3页 说明书10页 附图1页

(54)发明名称

电子装置、身份验证的方法及计算机可读存储介质

(57)摘要

本发明涉及一种电子装置、身份验证的方法及计算机可读存储介质，电子装置包括存储器及处理器，存储器中存储有身份验证的系统，身份验证的系统被处理器执行时实现：在接收到待进行身份验证的目标用户的当前语音数据后，对当前语音数据按照预设的分帧参数进行分帧处理，以获得多个语音帧；利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成当前语音数据对应的多个观测特征单元；将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元；将多组配对后的观测特征单元输入预先训练生成的预设类型身份验证模型，以对该目标用户进行身份验证。本发明能够降低短语音识别的错误率。



CN 107527620 B

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身份验证的系统,所述身份验证的系统被所述处理器执行时实现如下步骤:

S1,在接收到待进行身份验证的目标用户的当前语音数据后,对所述当前语音数据按照预设的分帧参数进行分帧处理,以获得多个语音帧;

S2,利用预定的滤波器提取各个语音帧中预设类型的声学特征,根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元;

S3,将各个观测特征单元分别与预存的观测特征单元进行两两配对,以获得多组配对后的观测特征单元;

S4,将多组配对后的观测特征单元输入预先训练生成的深度卷积神经网络模型,并获取输出的身份验证结果,以对该目标用户进行身份验证,所述深度卷积神经网络模型采用识别函数进行身份验证,所述识别函数为:

obj=-∑In(P(x,y))-K∑In(1-P(x,y)),

其中,$P(x,y)=\frac{1}{1+e^{-L(x,y)}}$,$L(x,y)=x^{T}Uy-x^{T}Vx-y^{T}Vy+b$;

P(x,y)为计算一组观测特征单元中的各个观测特征单元属于同一用户的概率,L(x,y)为计算一组观测特征单元中的各个观测特征单元的相似度L,x为一组观测特征单元中的一个观测特征单元输入深度卷积神经网络模型的归一化层得到的用户特征,y为该组观测特征单元中另一个观测特征单元输入所述归一化层得到的用户特征,K为常量,U为用户的类内关系矩阵,V为用户类间关系矩阵,b为偏置量,T为矩阵转置。

2.根据权利要求1所述的电子装置,其特征在于,所述预定的滤波器为梅尔滤波器,所述利用预定的滤波器提取各个语音帧中预设类型的声学特征的步骤包括:

对所述语音帧进行加窗处理;

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;

在梅尔频谱上面进行倒谱分析以获得梅尔频率倒谱系数MFCC,以所述梅尔频率倒谱系数MFCC作为该语音帧的声学特征。

3.根据权利要求2所述的电子装置,其特征在于,所述根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元的步骤包括:

以所述当前语音数据中的每个录音数据中的语音帧为一语音帧集合,将所述语音帧集合中的每个语音帧的20维梅尔频率倒谱系数MFCC按对应语音帧的分帧时间的先后顺序拼接,生成对应的(20,N)维矩阵的观测特征单元,所述N为该语音帧集合的总帧数。

4.一种身份验证的方法,其特征在于,所述身份验证的方法包括:

S1,在接收到待进行身份验证的目标用户的当前语音数据后,对所述当前语音数据按照预设的分帧参数进行分帧处理,以获得多个语音帧;

S2,利用预定的滤波器提取各个语音帧中预设类型的声学特征,根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元;

S3,将各个观测特征单元分别与预存的观测特征单元进行两两配对,以获得多组配对后的观测特征单元;

S4,将多组配对后的观测特征单元输入预先训练生成的深度卷积神经网络模型,并获取输出的身份验证结果,以对该目标用户进行身份验证,所述深度卷积神经网络模型采用识别函数进行身份验证,所述识别函数为:

obj=-∑In(P(x,y))-K∑In(1-P(x,y)),

其中,$P(x,y)=\frac{1}{1+e^{-L(x,y)}}$,L(x,y)=$x^{T}Uy-x^{T}Vx-y^{T}Vy+b$;

P(x,y)为计算一组观测特征单元中的各个观测特征单元属于同一用户的概率,L(x,y)为计算一组观测特征单元中的各个观测特征单元的相似度L,x 为一组观测特征单元中的一个观测特征单元输入深度卷积神经网络模型的归一化层得到的用户特征,y为该组观测特征单元中另一个观测特征单元输入所述归一化层得到的用户特征,K为常量,U为用户的类内关系矩阵,V为用户类间关系矩阵,b为偏置量,T为矩阵转置。

5.根据权利要求4所述的身份验证的方法,其特征在于,所述预定的滤波器为梅尔滤波器,所述利用预定的滤波器提取各个语音帧中预设类型的声学特征的步骤包括:

对所述语音帧进行加窗处理;

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;

在梅尔频谱上面进行倒谱分析以获得梅尔频率倒谱系数MFCC,以所述梅尔频率倒谱系数MFCC作为该语音帧的声学特征。

6.根据权利要求5所述的身份验证的方法,其特征在于,所述根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元的步骤包括:

以所述当前语音数据中的每个录音数据中的语音帧为一语音帧集合,将所述语音帧集合中的每个语音帧的20维梅尔频率倒谱系数MFCC按对应语音帧的分帧时间的先后顺序拼接,生成对应的(20,N)维矩阵的观测特征单元,所述N为该语音帧集合的总帧数。

7.根据权利要求4所述的身份验证的方法,其特征在于,所述步骤S4之前还包括:

获取同一用户第一预设数量的语音对,并获取不同用户第二预设数量的语音对,分别对各个语音对中的语音按照预设的分帧参数进行分帧处理,以获得各个语音对对应的多个语音帧;

利用预定的滤波器提取各个语音帧中预设类型的声学特征,根据所提取的声学特征生成每个语音对的多个观测特征单元;

将属于同一用户及属于不同用户的两个语音对应的观测特征单元进行两两配对,以获得多组配对的观测特征单元;

将各个语音对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之和小于或者等于1;

利用训练集中的各个语音对的各组观测特征单元对深度卷积神经网络模型进行训练,并在训练完成后利用验证集对训练后的深度卷积神经网络模型的准确率进行验证;

若所述准确率大于预设阈值,则训练结束,以训练后的深度卷积神经网络模型为所述步骤S4中的深度卷积神经网络模型,或者,若所述准确率小于或者等于所述预设阈值,则增加进行训练的语音对的数量,以重新进行训练。

8.一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身份验

证的系统,所述身份验证的系统被处理器执行时实现如权利要求4至7中任一项所述的身份验证的方法的步骤。

电子装置、身份验证的方法及计算机可读存储介质

技术领域

[0001] 本发明涉及通信技术领域，尤其涉及一种电子装置、身份验证的方法及计算机可读存储介质。

背景技术

[0002] 声纹识别是一种通过对目标语音进行计算机仿真判别的身份认证技术，可广泛应用在互联网、银行系统、公安司法等领域。目前，传统的声纹识别方案采用的是基于高斯混合模型建模的通用背景模型对说话人录音，并进行差异分析，然后根据差异提取声纹特征，并通过相似性测度进行打分，给出识别结果。这种声纹识别方案对于长录音(例如，30秒及以上时长的录音)的识别错误率较低，识别效果好，但是针对不同业务场景中广泛出现的短录音(例如，小于30秒时长的录音)，由于参数有限，利用通用背景模型框架无法很好地对录音中的细微差异进行建模，造成对短语音识别的性能不佳，识别错误率高。

发明内容

[0003] 本发明的目的在于提供一种电子装置、身份验证的方法及计算机可读存储介质，旨在降低短语音识别的错误率。

[0004] 为实现上述目的，本发明提供一种电子装置，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身份验证的系统，所述身份验证的系统被所述处理器执行时实现如下步骤：

[0005] S1，在接收到待进行身份验证的目标用户的当前语音数据后，对所述当前语音数据按照预设的分帧参数进行分帧处理，以获得多个语音帧；

[0006] S2，利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元；

[0007] S3，将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元；

[0008] S4，将多组配对后的观测特征单元输入预先训练生成的深度卷积神经网络模型，并获取输出的身份验证结果，以对该目标用户进行身份验证，所述深度卷积神经网络模型采用识别函数进行身份验证，所述识别函数为：

[0009] $Obj=-\sum In(P(x,y))-K\sum In(1-P(x,y))$，

[0010] 其中，$P(x,y)=\frac{1}{1+e^{-L(x,y)}}$，$L(x,y)=x^{T}Uy-x^{T}Vx-y^{T}Vy+b$；

[0011] P(x,y)为计算一组观测特征单元中的各个观测特征单元属于同一用户的概率，L(x,y)为计算一组观测特征单元中的各个观测特征单元的相似度L，x为一组观测特征单元中的一个观测特征单元输入深度卷积神经网络模型的归一化层得到的用户特征，y为该组观测特征单元中另一个观测特征单元输入所述归一化层得到的用户特征，K为常量，U为用户的类内关系矩阵，V为用户类间关系矩阵，b为偏置量，T为矩阵转置。

[0012] 优选地，所述预定的滤波器为梅尔滤波器，所述利用预定的滤波器提取各个语音帧中预设类型的声学特征的步骤包括：

[0013] 对所述语音帧进行加窗处理；

[0014] 对每一个加窗进行傅立叶变换得到对应的频谱；

[0015] 将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

[0016] 在梅尔频谱上面进行倒谱分析以获得梅尔频率倒谱系数MFCC，以所述梅尔频率倒谱系数MFCC作为该语音帧的声学特征。

[0017] 优选地，所述根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元的步骤包括：

[0018] 以所述当前语音数据中的每个录音数据中的语音帧为一语音帧集合，将所述语音帧集合中的每个语音帧的20维梅尔频率倒谱系数MFCC按对应语音帧的分帧时间的先后顺序拼接，生成对应的(20,N)维矩阵的观测特征单元，所述N为该语音帧集合的总帧数。

[0019] 为实现上述目的，本发明还提供一种身份验证的方法，所述身份验证的方法包括：

[0020] S1，在接收到待进行身份验证的目标用户的当前语音数据后，对所述当前语音数据按照预设的分帧参数进行分帧处理，以获得多个语音帧；

[0021] S2，利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元；

[0022] S3，将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元；

[0023] S4，将多组配对后的观测特征单元输入预先训练生成的深度卷积神经网络模型，并获取输出的身份验证结果，以对该目标用户进行身份验证，所述深度卷积神经网络模型采用识别函数进行身份验证，所述识别函数为：

[0024] $Obj=-\sum In(P(x,y))-K\sum In(1-P(x,y))$，

[0025] 其中，$P(x,y)=\frac{1}{1+e^{-L(x,y)}}$，$L(x,y)=x^{T}Uy-x^{T}Vx-y^{T}Vy+b$；

[0026] P(x,y)为计算一组观测特征单元中的各个观测特征单元属于同一用户的概率，L(x,y)为计算一组观测特征单元中的各个观测特征单元的相似度L，x为一组观测特征单元中的一个观测特征单元输入深度卷积神经网络模型的归一化层得到的用户特征，y为该组观测特征单元中另一个观测特征单元输入所述归一化层得到的用户特征，K为常量，U为用户的类内关系矩阵，V为用户类间关系矩阵，b为偏置量，T为矩阵转置。

[0027] 优选地，所述预定的滤波器为梅尔滤波器，所述利用预定的滤波器提取各个语音帧中预设类型的声学特征的步骤包括：

[0028] 对所述语音帧进行加窗处理；

[0029] 对每一个加窗进行傅立叶变换得到对应的频谱；

[0030] 将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

[0031] 在梅尔频谱上面进行倒谱分析以获得梅尔频率倒谱系数MFCC，以所述梅尔频率倒谱系数MFCC作为该语音帧的声学特征。

[0032] 优选地，所述根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元的步骤包括：

[0033] 以所述当前语音数据中的每个录音数据中的语音帧为一语音帧集合，将所述语音帧集合中的每个语音帧的20维梅尔频率倒谱系数MFCC按对应语音帧的分帧时间的先后顺序拼接，生成对应的(20，N)维矩阵的观测特征单元，所述N为该语音帧集合的总帧数。

[0034] 优选地，所述步骤S4之前还包括：

[0035] 获取同一用户第一预设数量的语音对，并获取不同用户第二预设数量的语音对，分别对各个语音对中的语音按照预设的分帧参数进行分帧处理，以获得各个语音对对应的多个语音帧；

[0036] 利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成每个语音对的多个观测特征单元；

[0037] 将属于同一用户及属于不同用户的两个语音对应的观测特征单元进行两两配对，以获得多组配对的观测特征单元；

[0038] 将各个语音对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之和小于或者等于1；

[0039] 利用训练集中的各个语音对的各组观测特征单元对深度卷积神经网络模型，并在训练完成后利用验证集对训练后的深度卷积神经网络模型的准确率进行验证；

[0040] 若所述准确率大于预设阈值，则训练结束，以训练后的深度卷积神经网络模型为所述步骤S4中的深度卷积神经网络模型，或者，若所述准确率小于或者等于所述预设阈值，则增加进行训练的语音对的数量，以重新进行训练。

[0041] 本发明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身份验证的系统，所述身份验证的系统被处理器执行时实现上述的身份验证的方法的步骤。

[0042] 本发明的有益效果是：本发明首先对当前语音数据分帧处理以获得多个语音帧，利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元，将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元，将多组配对后的观测特征单元输入预设类型身份验证模型中，获取输出的身份验证结果，以对该目标用户进行身份验证，本发明对于多种业务场景中出现的短录音进行身份认证时，对短录音进行分帧、提取声学特征并将其转化为观测特征单元，最终将配对后的观测特征单元输入至身份验证模型中进行身份验证，对短语音识别的性能较佳，能够降低识别错误率。

附图说明

[0043] 图1为本发明各个实施例一可选的应用环境示意图；

[0044] 图2为本发明身份验证的方法一实施例的流程示意图。

具体实施方式

[0045] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明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发明，并不用于限定本发明。基于本发明中的实施例，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，都属于本发明保护的范围。

[0046] 需要说明的是，在本发明中涉及“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，也不在本发明要求的保护范围之内。

[0047] 参阅图1所示，是本发明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2。电子装置1可以通过网络、近场通信技术等适合的技术与终端设备2进行数据交互。

[0048] 所述终端设备2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，该电子产品可以利用语音采集装置(例如麦克风)采集用户的语音数据，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、笔记本、服务器等等的固定终端。

[0049] 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，由一群松散耦合的计算机集组成的一个超级虚拟计算机。

[0050] 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。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，可以替代的实施更多或者更少的组件。

[0051] 其中，存储设备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(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，存储设备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，例如本发明一实施例中的身份验证的系统的程序代码等。此外，存储设备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。

[0052] 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2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，例如运行身份验证的系统等。

[0053] 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2相连，在电子装置1与一个或多个终端设备2之间建立数据传输通道和通信连接。

[0054] 所述身份验证的系统存储在存储器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。

[0055] 在一实施例中，上述身份验证的系统被所述处理器12执行时实现如下步骤：

[0056] S1，在接收到待进行身份验证的目标用户的当前语音数据后，对所述当前语音数据按照预设的分帧参数进行分帧处理，以获得多个语音帧；

[0057] 本实施例中，在多种业务场景下需要用户进行录音，录音过程中接收到的当前语音数据为一段的录音数据，该一段的录音数据为短语音。

[0058] 在进行录音时，应尽量防止环境噪声和语音采集设备的干扰。录音设备与用户保持适当距离，且尽量不用失真大的录音设备，电源优选使用市电，并保持电流稳定；在进行电话录音时应使用传感器。在进行分帧处理之前，可以对语音数据进行去噪音处理，以进一步减少干扰。

[0059] 其中，对当前语音数据中每段录音数据按照预设的分帧参数进行分帧处理时，预设的分帧参数例如为每隔25毫秒分帧、帧移10毫秒，分帧处理后每段录音数据得到多个语音帧。当然，本实施例不限定上述的这种分帧处理方式，可以采用其他的分帧参数进行分帧处理的方式，其均在本实施例的保护范围内。

[0060] S2，利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元；

[0061] 本实施例中，预定的滤波器优选为梅尔滤波器，声学特征即为声纹特征，声纹特征包括多种类型，例如宽带声纹、窄带声纹、振幅声纹等，本实施例的声纹特征为优选地为梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient，MFCC)。

[0062] 在根据声学特征生成所述当前语音数据对应的多个观测特征单元时，根据梅尔频率倒谱系数组成特征数据矩阵，具体地，根据每段录音数据的梅尔频率倒谱系数组成特征数据矩阵，多段录音数据对应的特征数据矩阵即为当前语音数据对应的多个观测特征单元。

[0063] S3，将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元；

[0064] S4，将多组配对后的观测特征单元输入预先训练生成的预设类型身份验证模型，并获取输出的身份验证结果，以对该目标用户进行身份验证。

[0065] 本实施例中，预先存储较多数量的用户的观测特征单元，在生成当前语音数据对应的多个观测特征单元后，将生成的多个观测特征单元与预先存储的观测特征单元进行来两两配对，配对后得到多组观测特征单元。

[0066] 其中，预设类型身份验证模型优选地为深度卷积神经网络模型，深度卷积神经网络模型由1个输入层，4个卷积层，1个池化层，2个全连接层，1个归一化层，1个分类层构成。所述深度卷积神经网络模型的详细结构如上述表1所示：

[0067]

Layer Name	Batch Size	Kernel Size	Stride Size	Filter Size
Input	128	20*9	1*1	512
Conv1	128	1*1	1*1	512
Conv2	128	1*1	1*1	512
Conv3	128	1*1	1*1	512
Conv4	128	1*1	1*1	512
Mean_std_pooling	128	\	\	\
Full connected	128	1024*512	\	\
Full connected	128	512*300	\	\
Normalize Wrap	128	\	\	\
Scoring:	128	300*300,300*300	\	\

[0068] 表1

[0069] 其中，Layer Name列表示每一层的名称，Input表示输入层，Conv表示卷积层，Conv1表示第1个卷积层，Mean_std_pooling表示池化层，Full connected表示全连接层，Normalize Wrap表示归一化层，Scoring表示分类层。Batch Size表示当前层的输入的观测特征单元的数目，Kernel Size表示当前层卷积核的尺度(例如，Kernel Size可以等于3，表示卷积核的尺度为3x 3)，Stride Size表示卷积核的移动步长，即做完一次卷积之后移动到下一个卷积位置的距离。Filter size指每层输出的通道，如在Input层的输入语音通道为1(即原始数据)，经过Input层通道变成512。具体地，输入层表示对输入的观测特征单元进行采样，Conv层的卷积核1*1可以对输入进行缩放及特征组合，Normalize Wrap层对输入进行方差归一化，Scoring层训练用户的类内关系矩阵U及用户的类间关系矩阵V，其维度均为300*300。

[0070] 本实施例中，将多组配对后的观测特征单元输入深度卷积神经网络模型后，并获取输出的身份验证结果，输出的身份验证结果包括验证通过及验证失败，以对该目标用户进行身份验证。

[0071] 与现有技术相比，本实施例首先对当前语音数据分帧处理以获得多个语音帧，利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元，将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元，将多组配对后的观测特征单元输入预设类型身份验证模型中，获取输出的身份验证结果，以对该目标用户进行身份验证，本实施例对于多种业务场景中出现的短录音进行身份认证时，对短录音进行分帧、提取声学特征并将其转化为观测特征单元，最终将配对后的观测特征单元输入至身份验证模型中进行身份验证，对短语音识别的性能较佳，能够降低识别错误率。

[0072] 在一优选的实施例中，在上述图1的实施例的基础上，所述利用预定的滤波器提取各个语音帧中预设类型的声学特征的步骤包括：

[0073] 对所述语音帧进行加窗处理；

[0074] 对每一个加窗进行傅立叶变换得到对应的频谱；

[0075] 将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

[0076] 在梅尔频谱上面进行倒谱分析以获得梅尔频率倒谱系数MFCC,以所述梅尔频率倒谱系数MFCC作为该语音帧的声学特征。

[0077] 其中,在对语音数据进行分帧之后,每一帧数据都当成平稳信号来处理。由于后续需要利用傅里叶展开每一项以获取Mel频谱特征,因此会出现吉布斯效应,即将具有不连续点的周期函数(如矩形脉冲)进行傅立叶级数展开后,选取有限项进行合成,当选取的项数越多,在所合成的波形中出现的峰起越靠近原信号的不连续点,当选取的项数很大时,该峰起值趋于一个常数,大约等于总跳变值的9%。为了避免吉布斯效应,则需要对语音帧进行加窗处理,以减少语音帧起始和结束的地方信号的不连续性问题。

[0078] 其中,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系数。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语音帧的声学特征。

[0079] 在一优选的实施例中,在上述的实施例的基础上,所述根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元的步骤包括:

[0080] 以所述当前语音数据中的每个录音数据中的全部语音帧为一语音帧集合,将所述语音帧集合中的每个语音帧的20维梅尔频率倒谱系数MFCC(即声学特征)按对应语音帧的分帧时间的先后顺序拼接,生成对应的(20,N)维矩阵的观测特征单元,所述N为该语音帧集合的总帧数。

[0081] 在一优选的实施例中,在上述的实施例的基础上,深度卷积神经网络模型采用识别函数进行身份验证,所述识别函数为:

[0082] $Obj=-\sum In(P(x,y))-K\sum In(1-P(x,y))$;

[0083] 其中,$P(x,y)=\frac{1}{1+e^{-L(x,y)}}$;

[0084] $L(x,y)=x^{T}Uy-x^{T}Vx-y^{T}Vy+b$;

[0085] Obj也称作深度卷积神经网络模型的目标函数,通过最大化该目标函数,使深度卷积神经网络模型给出正确判别的概率增大到收敛,由此对目标的身份进行验证。P(x,y)为计算一组观测特征单元中的各个观测特征单元属于同一用户的概率,L(x,y)为计算一组观测特征单元中的各个观测特征单元的相似度L,x为一组观测特征单元中其中一个观测特征单元在所述归一化层得到的用户特征,y为该组观测特征单元中另一个观测特征单元在所述归一化层得到的用户特征,K为常量,U为用户的类内关系矩阵,V为用户类间关系矩阵,b为偏置量,T为矩阵转置。

[0086] 如图2所示,图2为本发明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:

[0087] 步骤S1,在接收到待进行身份验证的目标用户的当前语音数据后,对所述当前语音数据按照预设的分帧参数进行分帧处理,以获得多个语音帧;

[0088] 本实施例中,在多种业务场景下需要用户进行录音,录音过程中接收到的当前语音数据为一段段的录音数据,该一段段的录音数据为短语音。

[0089] 在进行录音时,应尽量防止环境噪声和语音采集设备的干扰。录音设备与用户保持适当距离,且尽量不用失真大的录音设备,电源优选使用市电,并保持电流稳定;在进行

电话录音时应使用传感器。在进行分帧处理之前，可以对语音数据进行去噪音处理，以进一步减少干扰。

[0090] 其中，对当前语音数据中每段录音数据按照预设的分帧参数进行分帧处理时，预设的分帧参数例如为每隔25毫秒分帧、帧移10毫秒，分帧处理后每段录音数据得到多个语音帧。当然，本实施例不限定上述的这种分帧处理方式，可以采用其他的分帧参数进行分帧处理的方式，其均在本实施例的保护范围内。

[0091] 步骤S2，利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元；

[0092] 本实施例中，预定的滤波器优选为梅尔滤波器，声学特征即为声纹特征，声纹特征包括多种类型，例如宽带声纹、窄带声纹、振幅声纹等，本实施例的声纹特征为优选地为梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient，MFCC)。

[0093] 在根据声学特征生成所述当前语音数据对应的多个观测特征单元时，根据梅尔频率倒谱系数组成特征数据矩阵，具体地，根据每段录音数据的梅尔频率倒谱系数组成特征数据矩阵，多段录音数据对应的特征数据矩阵即为当前语音数据对应的多个观测特征单元。

[0094] 步骤S3，将各个观测特征单元分别与预存的观测特征单元进行两两配对，以获得多组配对后的观测特征单元；

[0095] 步骤S4，将多组配对后的观测特征单元输入预先训练生成的预设类型身份验证模型，并获取输出的身份验证结果，以对该目标用户进行身份验证。

[0096] 本实施例中，预先存储较多数量的用户的观测特征单元，在生成当前语音数据对应的多个观测特征单元后，将生成的多个观测特征单元与预先存储的观测特征单元进行来两两配对，配对后得到多组观测特征单元。

[0097] 其中，预设类型身份验证模型优选地为深度卷积神经网络模型，深度卷积神经网络模型由1个输入层，4个卷积层，1个池化层，2个全连接层，1个归一化层，1个分类层构成。所述深度卷积神经网络模型的详细结构如上述表1所示，此处不再赘述。

[0098] 其中，Layer Name列表示每一层的名称，Input表示输入层，Conv表示卷积层，Conv1表示第1个卷积层，Mean_std_pooling表示池化层，Full connected表示全连接层，Normalize Wrap表示归一化层，Scoring表示分类层。Batch Size表示当前层的输入的观测特征单元的数目，Kernel Size表示当前层卷积核的尺度(例如，Kernel Size可以等于3，表示卷积核的尺度为3x 3)，Stride Size表示卷积核的移动步长，即做完一次卷积之后移动到下一个卷积位置的距离。Filter size指每层输出的通道，如在Input层的输入语音通道为1(即原始数据)，经过Input层通道变成512。具体地，输入层表示对输入的观测特征单元进行采样，Conv层的卷积核1*1可以对输入进行缩放及特征组合，Normalize Wrap层对输入进行方差归一化，Scoring层训练用户的类内关系矩阵U及用户的类间关系矩阵V，其维度均为300*300。

[0099] 本实施例中，将多组配对后的观测特征单元输入深度卷积神经网络模型后，并获取输出的身份验证结果，输出的身份验证结果包括验证通过及验证失败，以对该目标用户进行身份验证。

[0100] 在一优选的实施例中，在上述图2的实施例的基础上，在上述步骤S2中利用预定的

滤波器提取各个语音帧中预设类型的声学特征的步骤包括：

[0101] 对所述语音帧进行加窗处理；

[0102] 对每一个加窗进行傅立叶变换得到对应的频谱；

[0103] 将所述频谱输入梅尔滤波器以输出得到梅尔频谱；

[0104] 在梅尔频谱上面进行倒谱分析以获得梅尔频率倒谱系数MFCC，以所述梅尔频率倒谱系数MFCC作为该语音帧的声学特征。

[0105] 其中，在对语音数据进行分帧之后，每一帧数据都当成平稳信号来处理。由于后续需要利用傅里叶展开每一项以获取Mel频谱特征，因此会出现吉布斯效应，即将具有不连续点的周期函数(如矩形脉冲)进行傅立叶级数展开后，选取有限项进行合成，当选取的项数越多，在所合成的波形中出现的峰起越靠近原信号的不连续点，当选取的项数很大时，该峰起值趋于一个常数，大约等于总跳变值的9％。为了避免吉布斯效应，则需要对语音帧进行加窗处理，以减少语音帧起始和结束的地方信号的不连续性问题。

[0106] 其中，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系数。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语音帧的声学特征。

[0107] 在一优选的实施例中，在上述的实施例的基础上，在上述步骤S2中根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元的步骤包括：以所述当前语音数据中的每个录音数据中的全部语音帧为一语音帧集合，将所述语音帧集合中的每个语音帧的20维梅尔频率倒谱系数MFCC(即声学特征)按对应语音帧的分帧时间的先后顺序拼接，生成对应的(20,N)维矩阵的观测特征单元，所述N为该语音帧集合的总帧数。

[0108] 在一优选的实施例中，在上述的实施例的基础上，深度卷积神经网络模型采用识别函数进行身份验证，所述识别函数包括：

[0109] $Obj=-\sum In(P(x,y))-K\sum In(1-P(x,y))$；

[0110] 其中，$P(x,y)=\frac{1}{1+e^{-L(x,y)}}$；

[0111] $L(x,y)=x^{T}Uy-x^{T}Vx-y^{T}Vy+b$；

[0112] Obj也称作所述深度卷积神经网络模型的目标函数，通过最大化该目标函数，使深度卷积神经网络模型给出正确判别的概率增大到收敛，由此对目标的身份进行验证。P(x,y)为计算一组观测特征单元中的各个观测特征单元属于同一用户的概率，L(x,y)为计算一组观测特征单元中的各个观测特征单元的相似度L，x为一组观测特征单元中其中一个观测特征单元在所述归一化层得到的用户特征，y为该组观测特征单元中另一个观测特征单元在所述归一化层得到的用户特征，K为常量，U为用户的类内关系矩阵，V为用户类间关系矩阵，b为偏置量，T为矩阵转置。

[0113] 在一优选的实施例中，在上述的实施例的基础上，所述步骤S4之前还包括：

[0114] 获取同一用户第一预设数量的语音对，例如，获取1000个用户，每个用户获取1000对语音对，每一对语音对由同一用户对应两个不同发音内容的两段语音组成；并获取不同用户第二预设数量的语音对，例如，获取1000个用户，将各个用户进行两两配对，每对用户对应一个相同发音内容得到一对语音对。分别对各个语音对中的语音按照预设的分帧参数

进行分帧处理，例如，所述预设的分帧参数为每隔25毫秒分帧，帧移10毫秒，以获得各个语音对对应的多个语音帧；

[0115] 利用预定的滤波器(例如，梅尔滤波器)提取各个语音帧中预设类型的声学特征(例如，20维梅尔频率倒谱系数MFCC频谱特征)，根据所提取的声学特征生成每个语音对的多个观测特征单元，即根据梅尔频率倒谱系数组成多个特征数据矩阵，该特征数据矩阵即为观测特征单元；

[0116] 将属于同一用户及属于不同用户的两个语音对应的观测特征单元进行两两配对，以获得多组配对的观测特征单元；

[0117] 将各个语音对分为第一百分比(例如70％)的训练集和第二百分比(例如20％)的验证集，所述第一百分比和第二百分比之和小于或者等于1；

[0118] 利用训练集中的各个语音对的各组观测特征单元对深度卷积神经网络模型，并在训练完成后利用验证集对训练后的深度卷积神经网络模型的准确率进行验证；

[0119] 若所述准确率大于预设阈值(例如，98.5％)，则训练结束，以训练后的深度卷积神经网络模型为所述步骤S4中的深度卷积神经网络模型，或者，若所述准确率小于或者等于所述预设阈值，则增加进行训练的语音对的数量，重新执行上述步骤，以重新进行训练，直至训练后的深度卷积神经网络模型的准确率大于预设阈值。

[0120] 本发明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身份验证的系统，所述身份验证的系统被处理器执行时实现上述的身份验证的方法的步骤。

[0121] 上述本发明实施例序号仅仅为了描述，不代表实施例的优劣。

[0122] 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，本发明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)执行本发明各个实施例所述的方法。

[0123] 以上仅为本发明的优选实施例，并非因此限制本发明的专利范围，凡是利用本发明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，均同理包括在本发明的专利保护范围内。

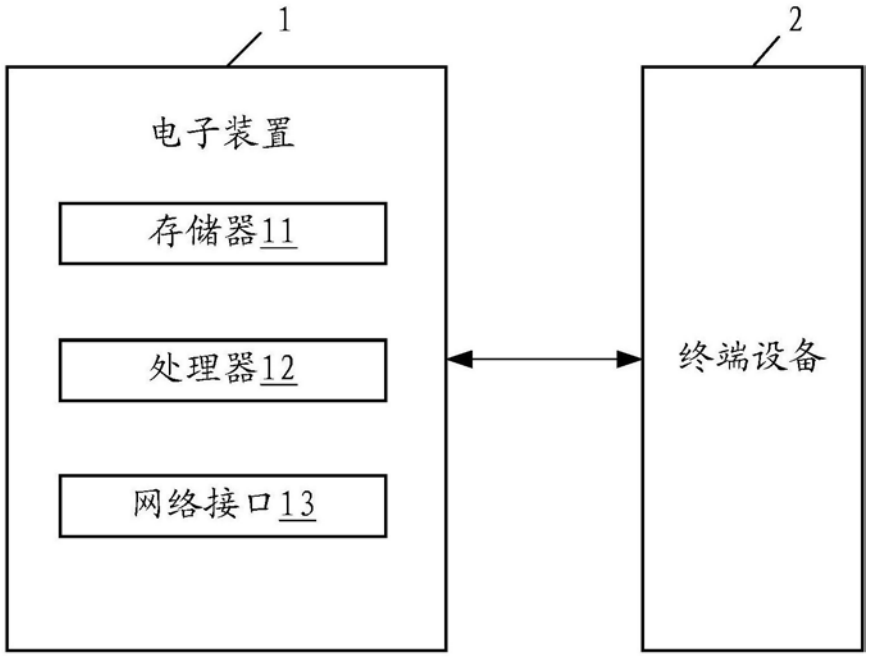

1
2
电子装置
存储器11
处理器12
网络接口13
终端设备

图1

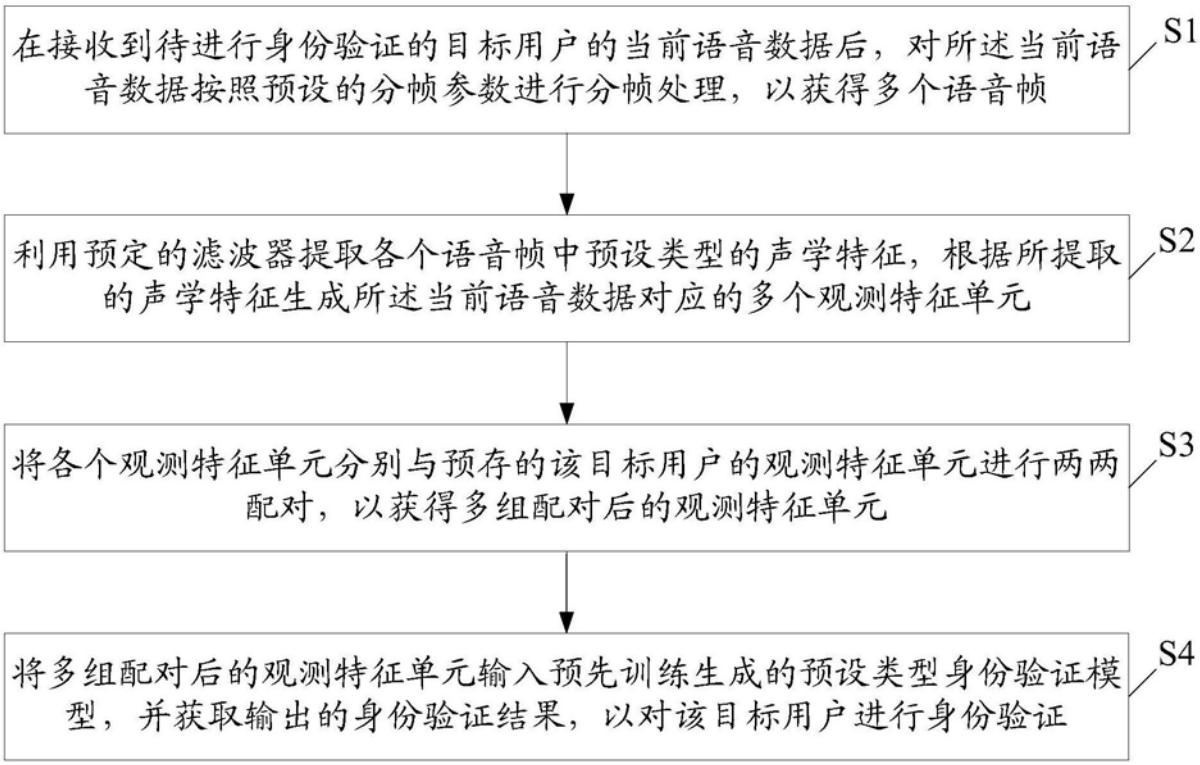

在接收到待进行身份验证的目标用户的当前语音数据后，对所述当前语音数据按照预设的分帧参数进行分帧处理，以获得多个语音帧
S1
利用预定的滤波器提取各个语音帧中预设类型的声学特征，根据所提取的声学特征生成所述当前语音数据对应的多个观测特征单元
S2
将各个观测特征单元分别与预存的该目标用户的观测特征单元进行两两配对，以获得多组配对后的观测特征单元
S3
将多组配对后的观测特征单元输入预先训练生成的预设类型身份验证模型，并获取输出的身份验证结果，以对该目标用户进行身份验证
S4

图2