



(12)发明专利

(10)授权公告号 CN 107221333 B

(45)授权公告日 2019.11.08

(21)申请号 201610162027.X

G10L 15/14(2006.01)

(22)申请日 2016.03.21

G10L 15/02(2006.01)

(65)同一申请的已公布的文献号

H04L 9/32(2006.01)

申请公布号 CN 107221333 A

(43)申请公布日 2017.09.29

(73)专利权人 中兴通讯股份有限公司

地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦法务部

(56)对比文件

CN 102543084 A, 2012.07.04,

US 2006294390 A1, 2006.12.28,

US 6671672 B1, 2003.12.30,

张慧珊. 基于声纹识别和动态密码的双因素身份认证系统的研究和实现.《武汉大学博士论文》.2014,

(72)发明人 朱长宝 李欢欢 袁浩 王金明

审查员 游晓梅

(74)专利代理机构 北京品源专利代理有限公司

11332

代理人 孟金喆

(51)Int.Cl.

G10L 17/24(2013.01)

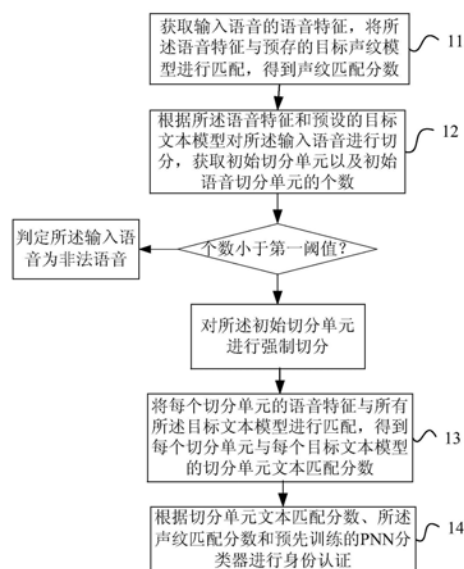
权利要求书5页 说明书14页 附图4页

(54)发明名称

一种身份认证的方法和装置

(57)摘要

一种身份认证的方法,包括:获取输入语音的语音特征,将语音特征与预存的目标声纹模型进行匹配得到声纹匹配分数;根据所述语音特征和目标文本模型对输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如初始语音切分单元的个数大于或等于第一阈值,则对初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;将每个切分单元的语音特征与所有目标文本模型进行匹配,得到每个切分单元与每个目标文本模型的切分单元文本匹配分数;根据所述切分单元文本匹配分数、声纹匹配分数和预先训练的概率神经网络PNN分类器进行身份认证,本方案实现了对用户进行双重验证的目的,提高了系统的安全性、可靠性和准确性。



1. 一种身份认证的方法,包括:

获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于第一阈值,则判定所述输入语音为非法语音;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的概率神经网络PNN分类器进行身份认证。

2. 根据权利要求1所述的方法,其特征在于,所述PNN分类器是通过以下方式进行训练的:

将目标语音与所述目标文本模型和目标声纹模型进行匹配分别得到第一文本打分和第一声纹打分,将所述第一文本打分和第一声纹打分组合成为所述判决分类器的接受特征信息;

将非目标语音与所述目标文本模型和目标声纹模型进行匹配分别得到第二文本打分和第二声纹打分,将所述第二文本打分和第二声纹打分组合成为所述判决分类器的拒绝特征信息;

根据所述接受特征信息和所述拒绝特征信息对所述PNN分类器进行训练。

3. 根据权利要求2所述的方法,其特征在于,在根据所述接受特征信息和所述拒绝特征信息对所述PNN分类器进行训练之前,还包括对所述目标语音和所述非目标语音的声纹打分和文本打分进行得分规整,包括:

依次选取所述目标文本模型,取非目标文本的语音特征与对应的所述目标文本模型匹配,得到冒认文本打分,获取所述目标文本模型对应的冒认文本打分的均值及标准差;

将所述第一文本打分和所述第二文本打分分别减去对应的所述冒认文本打分的均值且除以所述标准差,分别得到规整后的文本打分;

合并规整后的第一文本打分和所述第一声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第一文本打分和所述第一声纹打分进行归一化,作为所述PNN分类器的接受特征信息;

合并规整后的第二文本打分和所述第二声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第二文本打分和所述第二声纹打分进行归一化,作为所述PNN分类器的拒绝特征信息。

4. 根据权利要求1所述的方法,其特征在于,所述根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元,包括:

根据目标密码中的目标文本序列,将对应的目标文本隐马尔可夫模型HMM组合成第一复合HMM;

将所述语音特征作为所述第一复合HMM的输入进行维特比解码,得到第一状态输出序列,将所述第一状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置

作为初始切分点；

依次选取所述相邻两个初始切分点作为区间起止点，在所述区间内，以指定帧为单位计算平均能量，寻找平均能量连续指定次增大的点，并将开始增大的点作为新的初始切分点，由所述初始切分点分割成的所述初始切分单元。

5. 根据权利要求4所述方法，其特征在于，将对应的目标文本HMM组合成第一复合HMM，包括：

所述第一复合HMM的状态数为单个目标文本HMM的状态数总和；所述第一复合HMM的每个状态的高斯混合模型参数与所述单个目标文本HMM模型每个状态的高斯混合模型参数相同；

将所述单个目标文本HMM的状态转移矩阵中的最后一个状态自身转移概率设为0，转移到下一个状态的状态转移概率设为1；所述目标文本的最后一个单个目标文本HMM的状态转移概率矩阵不作改变；

将所述单个目标文本HMM的状态转移概率矩阵按照所述目标文本的单个目标文本排列顺序合并，得到所述复合HMM的状态转移概率矩阵。

6. 根据权利要求1所述方法，其特征在于，所述对所述初始切分单元进行强制切分，使得切分单元的总个数与预设的目标文本的个数相同，包括：

选择特征段最长的所述初始切分单元进行强制切分，使得强制切分后的所有切分单元的总个数与预设的目标文本的个数相同。

7. 根据权利要求1所述方法，其特征在于，所述对所述初始切分单元进行强制切分，使得切分单元的总个数与预设的目标文本的个数相同，包括：

按照所述初始切分单元的长度从大到小的顺序开始强制拆分，每次将一个所述初始切分单元平均切分成两个段，直至切分后的切分单元总个数等于所述目标文本的个数为止；

若强制切分的次数大于等于第二阈值，则强制切分结束；若强制切分的次数小于所述第二阈值，则将当前每个切分单元分别与每个目标文本隐马尔可夫模型HMM进行匹配打分，分别选定最高打分对应的所述目标文本HMM，将所选定的所述目标文本HMM组合成第二复合HMM；将所述语音特征作为所述第二复合HMM的输入进行维特比解码，得到第二状态输出序列，将所述第二状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为切分点，由该切分点对所述语音特征分割得到的不同单元为所述切分单元，若当前的所述切分单元个数小于第三阈值，则将当前切分后的切分单元作为所述初始切分单元继续进行强制切分，当前的所述切分单元个数大于或等于所述第三阈值，则强制切分结束。

8. 根据权利要求1所述方法，其特征在于，所述将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配，得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数，包括：

将每个所述切分单元的语音特征作为每个目标文本隐马尔可夫模型HMM的输入，将根据维特比算法获得的输出概率作为对应的切分单元文本匹配分数。

9. 根据权利要求1-8任一项所述方法，其特征在于，所述根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的概率神经网络PNN分类器进行身份认证，包括：

取每个所述切分单元对应的所述切分单元文本匹配分数中m个最高分数对应的文本作为待选文本，若所述待选文本中包含所述切分单元对应的目标文本，则所述切分单元认证

通过,计算通过的切分单元的总数,若通过的切分单元总数小于或等于第四阈值,则文本认证不通过,身份认证不通过;若通过的切分单元总数大于所述第四阈值,则所述输入语音的文本认证通过;

判断所述声纹匹配分数是否大于第五阈值,如是,则声纹认证通过,身份认证通过;如不是,则将每个所述切分单元与对应目标文本模型的文本打分以及所述声纹匹配分数进行得分规整,将规整后的打分作为所述概率神经网络PNN分类器的输入进行身份认证。

10. 一种身份认证的装置,包括概率神经网络PNN分类器,其特征在于,包括:

声纹匹配模块,设置为获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

切分模块,设置为根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于阈值,则判定所述输入语音为非法语音;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

文本匹配模块,设置为将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

认证模块,设置为根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的所述PNN分类器进行身份认证。

11. 根据权利要求10所述的装置,其特征在于,还包括处理模块,

所述声纹匹配模块,是设置为将目标语音与目标声纹模型进行匹配得到第一声纹打分,将非目标语音与所述目标声纹模型进行匹配得到第二声纹打分;

所述文本匹配模块,是设置为将所述目标语音与所述目标文本模型进行匹配得到第一文本打分,将所述非目标语音与所述目标文本模型进行匹配得到第二文本打分;

所述处理模块,设置为将所述第一文本打分和第一声纹打分组合成为所述PNN分类器的接受特征信息,将所述第二文本打分和第二声纹打分组合成为所述PNN分类器的拒绝特征信息;

所述PNN分类器,根据所述接受特征信息和所述拒绝特征信息进行训练。

12. 根据权利要求11所述的装置,其特征在于,

所述处理模块,还设置为依次选取所述目标文本模型,取非目标文本的语音特征与对应的所述目标文本模型匹配,得到冒认文本打分,获取所述目标文本模型对应的冒认文本打分的均值及标准差;将所述第一文本打分和所述第二文本打分分别减去对应的所述冒认文本打分的均值且除以所述标准差,分别得到规整后的文本打分;合并规整后的第一文本打分和所述第一声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第一文本打分和所述第一声纹打分进行归一化,作为所述PNN分类器的接受特征信息;合并规整后的第二文本打分和所述第二声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第二文本打分和所述第二声纹打分进行归一化,作为所述PNN分类器的拒绝特征信息。

13. 根据权利要求10所述的装置,其特征在于,

所述切分模块,根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,

获取初始切分单元,包括:根据目标密码中的目标文本序列,将对应的目标文本隐马尔可夫模型HMM组合成第一复合HMM;将所述语音特征作为所述第一复合HMM的输入进行维特比解码,得到第一状态输出序列,将所述第一状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为初始切分点;依次选取所述相邻两个初始切分点作为区间起止点,在所述区间内,以指定帧为单位计算平均能量,寻找平均能量连续指定次增大的点,并将开始增大的点作为新的初始切分点,由所述初始切分点分割成的所述初始切分单元。

14. 根据权利要求13所述的装置,其特征在于,

所述切分模块,将对应的目标文本HMM组合成第一复合HMM,包括:所述第一复合HMM的状态数为单个目标文本HMM的状态数总和;所述第一复合HMM的每个状态具有的高斯混合模型参数与所述单个目标文本HMM的每个状态具有的高斯混合模型参数相同;将所述单个目标文本HMM的状态转移矩阵中的最后一个状态自身转移概率设为0,转移到下一个状态的状态转移概率设为1;所述目标文本的最后一个单个目标文本HMM的状态转移概率矩阵不作改变;将所述单个目标文本HMM的状态转移概率矩阵按照所述目标文本的单个目标文本排列顺序合并,得到所述复合HMM的状态转移概率矩阵。

15. 根据权利要求10所述的装置,其特征在于,

所述切分模块,对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:选择特征段最长的所述初始切分单元进行强制切分,使得强制切分后的所有切分单元的总个数与预设的目标文本的个数相同。

16. 根据权利要求10所述的装置,其特征在于,

所述切分模块,对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:按照所述初始切分单元的长度从大到小的顺序开始拆分,每次将一个所述初始切分单元平均切分成两个段,直至切分后的单元总个数等于所述目标文本的个数为止;若强制切分的次数大于等于第二阈值,则强制切分结束;若强制切分的次数小于所述第二阈值,则将当前每个切分的单元分别与每个目标文本隐马尔可夫模型HMM进行匹配打分,分别选定最高打分对应的所述目标文本HMM,将所选定的所述目标文本HMM组合成第二复合HMM;将所述语音特征作为所述第二复合HMM的输入进行维特比解码,得到第二状态输出序列,将所述第二状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为切分点,由该切分点对所述语音特征分割得到的不同单元为所述切分单元,若当前的所述切分单元个数小于第三阈值,则将当前切分后的切分单元作为所述初始切分单元继续进行强制切分,若当前的所述切分单元个数大于或等于所述第三阈值,则强制切分结束。

17. 根据权利要求10所述的装置,其特征在于,

所述文本匹配模块,将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数,包括:将每个所述切分单元的语音特征作为每个目标文本隐马尔可夫模型HMM的输入,将根据维特比算法获得的输出概率作为对应的切分单元文本匹配分数。

18. 根据权利要求10-17任一项所述的装置,其特征在于,

所述认证模块,根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的所述PNN分类器进行身份认证,包括:取每个所述切分单元对应的所述切分单元文本匹配分数

中 m 个最高分数对应的文本作为待选文本,若所述待选文本中包含所述切分单元对应的目标文本,则所述切分单元认证通过,计算通过的切分单元的总数,若通过的切分单元总数小于或等于第四阈值,则文本认证不通过,身份认证不通过;若通过的切分单元总数大于所述第四阈值,则所述输入语音的文本认证通过;判断所述声纹匹配分数是否大于第五阈值,如是,则声纹认证通过,身份认证通过;如不是,则将每个所述切分单元与对应目标文本模型的文本打分以及所述声纹匹配分数进行得分规整,将规整后的打分作为所述PNN分类器的输入进行身份认证。

一种身份认证的方法和装置

技术领域

[0001] 本文涉及但不限于生物安全动态认证技术领域,尤指一种身份认证的方法和装置。

背景技术

[0002] 随着互联网信息技术的不断发展,网上业务、电子商务等日益繁荣,人们与计算机网络的联系越来越紧密,各种网络安全威胁也随之而来,保护用户个人信息成为人们急需解决的问题。动态声纹密码识别技术结合了说话人识别和语音识别两重身份认证技术,从而可以有效地防止了录音攻击,极大地增强了系统的安全性。通常,在接收到用户含有密码的语音后,系统首先对声纹和动态密码分别计算得分,然后分别比较两种得分与阈值大小,或者将两种得分融合后判断其与综合阈值的大小,若大于事先设定的阈值,则请求人进入被保护系统,否则,拒绝其进入。但在实际应用时,受环境的影响,说话人声纹匹配分数分布和文本匹配分数分布往往各不相同,而仅仅利用预先设定的阈值来判断则有失准确性。

发明内容

[0003] 以下是对本文详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

[0004] 本发明实施例提供一种身份认证的方法,包括:

[0005] 获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

[0006] 根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于第一阈值,则判定所述输入语音为非法语音;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

[0007] 将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

[0008] 根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的概率神经网络PNN分类器进行身份认证。

[0009] 可选地,所述PNN分类器是通过以下方式进行训练的:

[0010] 将目标语音与所述目标文本模型和目标声纹模型进行匹配分别得到第一文本打分和第一声纹打分,将所述第一文本打分和第一声纹打分组合成为所述判决分类器的接受特征信息;

[0011] 将非目标语音与所述目标文本模型和目标声纹模型进行匹配分别得到第二文本打分和第二声纹打分,将所述第二文本打分和第二声纹打分组合成为所述判决分类器的拒绝特征信息;

[0012] 根据所述接受特征信息和所述拒绝特征信息对所述PNN分类器进行训练。

[0013] 可选地,在根据所述接受特征信息和所述拒绝特征信息对所述PNN分类器进行训练之前,还包括对所述目标语音和所述非目标语音的声纹打分和文本打分进行得分规整,包括:

[0014] 依次选取所述目标文本模型,取非目标文本的语音特征与对应的所述目标文本模型匹配,得到冒认文本打分,获取所述目标文本模型对应的冒认文本打分的均值及标准差;

[0015] 将所述第一文本打分和所述第二文本打分分别减去对应的所述冒认文本打分的均值且除以所述标准差,分别得到规整后的文本打分;

[0016] 合并规整后的第一文本打分和所述第一声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第一文本打分和所述第一声纹打分进行归一化,作为所述PNN分类器的接受特征信息;

[0017] 合并规整后的第二文本打分和所述第二声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第二文本打分和所述第二声纹打分进行归一化,作为所述PNN分类器的拒绝特征信息。

[0018] 可选地,所述根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元,包括:

[0019] 根据目标密码中的目标文本序列,将对应的目标文本隐马尔可夫模型HMM组合成第一复合HMM;

[0020] 将所述语音特征作为所述第一复合HMM的输入进行维特比解码,得到第一状态输出序列,将所述第一状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为初始切分点;

[0021] 依次选取所述相邻两个初始切分点作为区间起止点,在所述区间内,以指定帧为单位计算平均能量,寻找平均能量连续指定次增大的点,并将开始增大的点作为新的初始切分点,由所述初始切分点分割成的所述初始切分单元。

[0022] 可选地,将对应的目标文本HMM组合成第一复合HMM,包括:

[0023] 所述第一复合HMM的状态数为单个目标文本HMM的状态数总和;所述第一复合HMM的每个状态的高斯混合模型参数与所述单个目标文本HMM模型每个状态的高斯混合模型参数相同;

[0024] 将所述单个目标文本HMM的状态转移矩阵中的最后一个状态自身转移概率设为0,转移到下一个状态的状态转移概率设为1;所述目标文本的最后一个单个目标文本HMM的状态转移概率矩阵不作改变;

[0025] 将所述单个目标文本HMM的状态转移概率矩阵按照所述目标文本的单个目标文本排列顺序合并,得到所述复合HMM的状态转移概率矩阵。

[0026] 可选地,所述对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:

[0027] 选择特征段最长的所述初始切分单元进行强制切分,使得强制切分后的所有切分单元的总个数与预设的目标文本的个数相同。

[0028] 可选地,所述对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:

[0029] 按照所述初始切分单元的长度从大到小的顺序开始强制拆分,每次将一个所述初始切分单元平均切分成两个段,直至切分后的切分单元总个数等于所述目标文本的个数为止;

[0030] 若强制切分的次数大于等于第二阈值,则强制切分结束;若强制切分的次数小于所述第二阈值,则将当前每个切分单元分别与每个目标文本隐马尔可夫模型HMM进行匹配打分,分别选定最高打分对应的所述目标文本HMM,将所选定的所述目标文本HMM组合成第二复合HMM;将所述语音特征作为所述第二复合HMM的输入进行维特比解码,得到第二状态输出序列,将所述第二状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为切分点,由该切分点对所述语音特征分割得到的不同单元为所述切分单元,若当前的所述切分单元个数小于第三阈值,则将当前切分后的切分单元作为所述初始切分单元继续进行强制切分,当前的所述切分单元个数大于或小于所述第三阈值,则强制切分结束。

[0031] 可选地,所述将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数,包括:

[0032] 将每个所述切分单元的语音特征作为每个目标文本隐马尔可夫模型HMM的输入,将根据维特比算法获得的输出概率作为对应的切分单元文本匹配分数。

[0033] 可选地,所述根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的判决分类器进行身份认证,包括:

[0034] 取每个所述切分单元对应的所述切分单元文本匹配分数中m个最高分数对应的文本作为待选文本,若所述待选文本中包含所述切分单元对应的目标文本,则所述切分单元认证通过,计算通过的切分单元的总数,若通过的切分单元总数小于或等于第四阈值,则文本认证不通过,身份认证不通过;若通过的切分单元总数大于所述第四阈值,则所述输入语音的文本认证通过;

[0035] 判断所述声纹匹配分数是否大于第五阈值,如是,则声纹认证通过,身份认证通过;如不是,则将每个所述切分单元与对应目标文本模型的文本打分以及所述声纹匹配分数进行得分规整,将规整后的打分作为所述判决分类器的输入进行身份认证。

[0036] 本发明实施例还提供了一种身份认证的装置,包括概率神经网络PNN分类器,包括:

[0037] 声纹匹配模块,设置为获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

[0038] 切分模块,设置为根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于阈值,则判定所述输入语音为非法语音;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

[0039] 文本匹配模块,设置为将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

[0040] 认证模块,设置为根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的所述PNN分类器进行身份认证。

[0041] 可选地,所述装置还包括处理模块,

[0042] 所述声纹匹配模块,是设置为将目标语音与目标声纹模型进行匹配得到第一声纹打分,将非目标语音与所述目标声纹模型进行匹配得到第二声纹打分;

[0043] 所述文本匹配模块,是设置为将所述目标语音与所述目标文本模型进行匹配得到第一文本打分,将所述非目标语音与所述目标文本模型进行匹配得到第二文本打分;

[0044] 所述处理模块,设置为将所述第一文本打分和第一声纹打分组合成为所述PNN分类器的接受特征信息,将所述第二文本打分和第二声纹打分组合成为所述PNN分类器的拒绝特征信息;

[0045] 所述PNN分类器,根据所述接受特征信息和所述拒绝特征信息进行训练。

[0046] 可选地,所述处理模块,还设置为依次选取所述目标文本模型,取非目标文本的语音特征与对应的所述目标文本模型匹配,得到冒认文本打分,获取所述目标文本模型对应的冒认文本打分的均值及标准差;将所述第一文本打分和所述第二文本打分分别减去对应的所述冒认文本打分的均值且除以所述标准差,分别得到规整后的文本打分;合并规整后的第一文本打分和所述第一声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第一文本打分和所述第一声纹打分进行归一化,作为所述PNN分类器的接受特征信息;合并规整后的第二文本打分和所述第二声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第二文本打分和所述第二声纹打分进行归一化,作为所述PNN分类器的拒绝特征信息。

[0047] 可选地,所述切分模块,根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元,包括:根据目标密码中的目标文本序列,将对应的目标文本隐马尔可夫模型HMM组合成第一复合HMM;将所述语音特征作为所述第一复合HMM的输入进行维特比解码,得到第一状态输出序列,将所述第一状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为初始切分点;依次选取所述相邻两个初始切分点作为区间起止点,在所述区间内,以指定帧为单位计算平均能量,寻找平均能量连续指定次增大的点,并将开始增大的点作为新的初始切分点,由所述初始切分点分割成的所述初始切分单元。

[0048] 可选地,所述切分模块,将对应的目标文本HMM组合成第一复合HMM,包括:所述第一复合HMM的状态数为单个目标文本HMM的状态数总和;所述第一复合HMM的每个状态具有的高斯混合模型参数与所述单个目标文本HMM的每个状态具有的高斯混合模型参数相同;将所述单个目标文本HMM的状态转移矩阵中的最后一个状态自身转移概率设为0,转移到下一个状态的状态转移概率设为1;所述目标文本的最后一个单个目标文本HMM的状态转移概率矩阵不作改变;将所述单个目标文本HMM的状态转移概率矩阵按照所述目标文本的单个目标文本排列顺序合并,得到所述复合HMM的状态转移概率矩阵。

[0049] 可选地,所述切分模块,对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:选择特征段最长的所述初始切分单元进行强制切分,使得强制切分后的所有切分单元的总个数与预设的目标文本的个数相同。

[0050] 可选地,所述切分模块,对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:按照所述初始切分单元的长度从大到小的顺序开始拆分,每次将一个所述初始切分单元平均切分成两个段,直至切分后的单元总个数等于

所述目标文本的个数为止；若强制切分的次数大于等于第二阈值，则强制切分结束；若强制切分的次数小于所述第二阈值，则将当前每个切分的单元分别与每个目标文本隐马尔可夫模型HMM进行匹配打分，分别选定最高打分对应的所述目标文本HMM，将所选定的所述目标文本HMM组合成第二复合HMM；将所述语音特征作为所述第二复合HMM的输入进行维特比解码，得到第二状态输出序列，将所述第二状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为切分点，由该切分点对所述语音特征分割得到的不同单元为所述切分单元，若当前的所述切分单元个数小于第三阈值，则将当前切分后的切分单元作为所述初始切分单元继续进行强制切分，若当前的所述切分单元个数大于或等于第三阈值，则强制切分结束。

[0051] 可选地，所述文本匹配模块，将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配，得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数，包括：将每个所述切分单元的语音特征作为每个目标文本隐马尔可夫模型HMM的输入，将根据维特比算法获得的输出概率作为对应的切分单元文本匹配分数。

[0052] 可选地，所述认证模块，根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的判决分类器进行身份认证，包括：取每个所述切分单元对应的所述切分单元文本匹配分数中m个最高分数对应的文本作为待选文本，若所述待选文本中包含所述切分单元对应的目标文本，则所述切分单元认证通过，计算通过的切分单元的总数，若通过的切分单元总数小于或等于第四阈值，则文本认证不通过，身份认证不通过；若通过的切分单元总数大于所述第四阈值，则所述输入语音的文本认证通过；判断所述声纹匹配分数是否大于第五阈值，如是，则声纹认证通过，身份认证通过；如不是，则将每个所述切分单元与对应目标文本模型的文本打分以及所述声纹匹配分数进行得分规整，将规整后的打分作为所述PNN分类器的输入进行身份认证。

[0053] 本发明实施例还提供了一种计算机可读存储介质，存储有计算机可执行指令，所述计算机可执行指令用于上述的一种身份认证的方法。

[0054] 综上，本发明实施例提供一种身份认证的方法及装置，将声纹与动态密码认证两者相结合，实现了对用户进行双重验证的目的，提高了系统的安全性、可靠性和准确性。

附图说明

[0055] 图1是本发明实施例提供的一种身份认证的方法的流程图；

[0056] 图2是本发明实施例的训练PNN分类器的方法的流程图；

[0057] 图3是本发明实施例一的一种身份认证的方法的流程图；

[0058] 图4是本发明实施例一的语音信号初始切分的方法的流程图；

[0059] 图5是本发明实施例一的声纹与文本初步认证的方法的流程图；

[0060] 图6是本发明实施例一的得分规整的方法的流程图；

[0061] 图7是本发明实施例二的一种身份认证的方法的流程图；

[0062] 图8是本发明实施例二的语音信号初始切分的方法的流程图；

[0063] 图9为本发明实施例的一种身份认证的装置的示意图。

具体实施方式

[0064] 下文中将结合附图对本发明的实施例进行详细说明。需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互任意组合。

[0065] 图1是本发明实施例提供的一种身份认证的方法的流程图,如图1所示,本实施例的方法包括以下步骤:

[0066] 步骤11、获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

[0067] 步骤12、根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于阈值,则判定所述输入语音为非法语音,结束流程;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

[0068] 步骤13、将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

[0069] 步骤14、根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的PNN (Probabilistic neural networks,概率神经网络)分类器进行身份认证。

[0070] 本发明实施例提供的一种身份认证方法,将声纹与动态密码认证两者相结合,实现了对用户进行双重验证的目的,提高了系统的安全性、可靠性和准确性。

[0071] 本实施例中,需要预先对PNN分类器进行训练,根据已有的语音获取目标文本模型和目标声纹模型;将已有语音与所述目标文本模型和目标声纹模型进行匹配得到文本打分和声纹打分,根据所述声纹打分和文本打分组合成接受特征信息和拒绝特征信息,将所述信息接受特征和所述拒绝特征信息作为综合PNN判决分类器的输入进行训练,得到最终的综合判决分类器;实现方式如下:

[0072] 将目标语音与所述目标文本模型和目标声纹模型进行匹配分别得到第一文本打分和第一声纹打分,将所述第一文本打分和第一声纹打分组合成为所述判决分类器的接受特征信息;

[0073] 将非目标语音与所述目标文本模型和目标声纹模型进行匹配分别得到第二文本打分和第二声纹打分,将所述第二文本打分和第二声纹打分组合成为所述判决分类器的拒绝特征信息;

[0074] 根据所述接受特征信息和所述拒绝特征信息对所述PNN分类器进行训练。

[0075] 所述目标语音为所述目标话者读取所述目标文本的语音,所述非目标语音为所述目标话者读取非目标文本的语音以及非目标话者的语音。

[0076] 可选地,在训练所述综合分类器之前对所述声纹打分和文本打分进行得分规整,例如包括以下步骤:

[0077] a.依次选取目标文本模型,取非目标文本语音特征与该目标文本模型匹配,得到冒认文本打分;

[0078] b.求所述目标文本模型对应的所述冒认文本打分均值及标准差;

[0079] c.将所述第一文本打分和所述第二文本打分分别减去对应的所述冒认文本打分的均值且除以所述标准差,分别得到规整后的文本打分;

[0080] d.合并所述声纹打分和规整后的文本打分,求得每一目标文本对应的最大值和最小值,利用步骤d中的所述最大值和最小值将所述声纹打分和文本打分进行归一化;例如:

[0081] 合并规整后的第一文本打分和所述第一声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第一文本打分和所述第一声纹打分进行归一化,作为所述PNN分类器的接受特征信息;

[0082] 合并规整后的第二文本打分和所述第二声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第二文本打分和所述第二声纹打分进行归一化,作为所述PNN分类器的拒绝特征信息。

[0083] 本实施例中为方便描述,做以下定义:

[0084] 目标文本:事先选定作为备选密码的文本,如0~9数字;

[0085] 目标话者:系统受信任的话者,在声纹认证时需要让其通过的话者;

[0086] 冒认话者:系统非受信任话者,在声纹认证时需要拒绝其进入的话者;

[0087] 目标密码:系统受信任的目标文本组合,在文本认证时需要让其通过;

[0088] 冒认密码:系统不受信任的文本组合,在文本认证时需要拒绝其进入的文本。

[0089] 系统进行认证之前,需要选择目标文本集,并针对目标文本集中的每个目标文本进行训练,得到目标文本模型集。以下实施例目标文本集选择为:0~9十个数字,目标模型集由0~9十个数字训练出来的模型组成,目标模型种类可以为HMM(Hidden Markov Model,隐马尔可夫模型)。为方便描述,动态密码均由0~9十个数字中的8个组成,即系统选择8个目标文本,作为目标密码。同时在系统进行认证之前,需要注册目标话者的声纹信息,通过训练生成声纹模型,并通过声纹模型和目标模型训练综合判决分类器,如图2所示包括如下步骤:

[0090] 步骤001:训练目标文本模型:使用0~9的数字录音训练单个数字的HMM,每个数字的模型称为目标文本模型,训练方法可使用现有的训练方法;

[0091] HMM是一个双重随机过程,一个过程用来描述短时平稳信号的时变性,另一个过程用来描述HMM模型的状态数与特征序列之间的对应关系。两个过程相互作用,不仅能够描述语音信号的动态特性,而且可以解决短时平稳信号之间的过渡问题。

[0092] 步骤002:注册目标话者声纹模型:系统在使用之前,事先注册目标话者声纹模型,目标话者即为系统受信任的话者,在认证时需要让其通过;

[0093] 步骤003:求接受特征:使用目标话者的目标文本对应的语音与其对应的HMM进行匹配,得到目标文本接受打分;使用目标话者的目标文本对应的语音与目标话者声纹模型进行打分,得到目标话者声纹接受打分;一系列的目标话者声纹接受打分和目标文本接受打分组成综合分类器的接受特征,对应综合分类器输出为1;

[0094] 步骤004:求拒绝特征:使用目标文本对应的语音与非对应的HMM模型进行匹配,得到冒认文本的拒绝打分;使用冒认话者与目标声纹模型进行打分,得到冒认声纹拒绝打分,由一系列的冒认文本拒绝打分和冒认声纹拒绝打分组成综合分类器的拒绝特征,对应综合分类器输出为0;

[0095] 步骤005:训练分类器:合并综合分类器的接受特征和拒绝特征,将合并后的特征进行得分规整(详见步骤109)后作为分类器的训练输入,根据现有训练算法(如梯度下降算法)可得到综合分类器。

[0096] 实施例一：

[0097] 如图3所示，包括以下步骤：

[0098] 步骤101、预处理：根据短时能量和短时过零率，对用户输入的测试语音进行预处理，去掉语音中的非语音段；

[0099] 步骤102、特征参数提取：对预处理后的测试语音进行特征参数提取，该系统可以采用12维梅尔频域倒谱系数(Mel Frequency Cepstrum Coefficient,简称MFCC)和其一阶差分系数作为特征参数，共24维；

[0100] 步骤103、计算声纹匹配分数：将测试语音特征与目标话者的声纹模型进行匹配，得到声纹匹配分数；

[0101] 步骤104、对语音特征初始切分：通过对测试语音特征的初始切分，获得初始切分单元以及初始切分单元个数。

[0102] 本实施例中，根据目标密码中的目标文本序列，将对应的目标文本HMM组合成复合HMM；

[0103] 将所述语音特征作为所述复合HMM的输入进行Viterbi(维特比)解码，得到第一状态输出序列，将所述第一状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为初始切分点；

[0104] 依次选取所述相邻两个初始切分点作为区间起止点，在所述区间内，以指定帧为单位计算平均能量，寻找平均能量连续指定次增大的点，并将开始增大的点作为新的初始切分点，否则，不更新初始切分点，由所述初始切分点分割成的所述初始切分单元。

[0105] 其中，所述复合HMM的状态数为单个目标文本HMM的状态数总和；所述复合HMM的每个状态具有的高斯混合模型参数与所述单个目标文本HMM的每个状态具有的高斯混合模型参数相同，

[0106] 将所述单个目标文本HMM的状态转移矩阵中的最后一个状态自身转移概率设为0，转移到下一个状态的状态转移概率设为1；所述目标文本的最后一个单个目标文本HMM的状态转移概率矩阵不作改变；

[0107] 将所述单个目标文本HMM的状态转移概率矩阵按照所述目标文本的单个目标文本排列顺序合并，得到所述复合HMM的状态转移概率矩阵。

[0108] 对语音特征初始切分的方法如图4所示，包括步骤如下：

[0109] 步骤104a、复合HMM模型的组合：按照目标密码中目标文本序列，将对应的单个目标文本HMM组合为复合HMM模型。

[0110] 假设每个数字的HMM模型有8个状态数，每个状态由3个高斯函数拟合，那么，复合HMM模型的状态数为单个目标文本HMM模型状态数之和，每个状态仍由3个高斯函数拟合，且其高斯混合模型参数与单个HMM模型每个状态的高斯混合模型参数相同，复合HMM的状态转移概率矩阵参数的变化以3个单个目标文本HMM模型连接成一个复合型HMM为例进行说明，该例中单个目标文本HMM模型状态数为3，如下式所示：

$$[0111] \quad trans1 = \begin{bmatrix} 0.95 & 0.05 & \\ & 0.92 & 0.08 \\ & & 1 \end{bmatrix} \quad (1)$$

$$[0112] \quad trans2 = \begin{bmatrix} 0.93 & 0.07 & & \\ & 0.91 & 0.09 & \\ & & 1 & \end{bmatrix} \quad (2)$$

$$[0113] \quad trans3 = \begin{bmatrix} 0.94 & 0.06 & & \\ & 0.93 & 0.07 & \\ & & 1 & \end{bmatrix} \quad (3)$$

[0114] 组合成复合HMM模型时,每个状态矩阵将改写成如下形式:

$$[0115] \quad trans1 = \begin{bmatrix} 0.95 & 0.05 & & & \\ & 0.92 & 0.08 & & \\ & & 0 & 1 & \end{bmatrix} \quad (4)$$

$$[0116] \quad trans2 = \begin{bmatrix} 0.93 & 0.07 & & & \\ & 0.91 & 0.09 & & \\ & & 0 & 1 & \end{bmatrix} \quad (5)$$

$$[0117] \quad trans3 = \begin{bmatrix} 0.94 & 0.06 & & & \\ & 0.93 & 0.97 & & \\ & & 1 & & \end{bmatrix} \quad (6)$$

[0118] 于是复合HMM模型的状态转移概率矩阵为:

$$[0119] \quad trans = \begin{bmatrix} 0.95 & 0.05 & & & & & & \\ & 0.92 & 0.08 & & & & & \\ & & 0 & 1 & & & & \\ & & & 0.93 & 0.07 & & & \\ & & & & 0.91 & 0.09 & & \\ & & & & & 0 & 1 & \\ & & & & & & 0.94 & 0.06 \\ & & & & & & & 0.93 & 0.07 \\ & & & & & & & & 1 \end{bmatrix} \quad (7)$$

[0120] 步骤104b、Viterbi (维特比) 解码: 利用Viterbi解码将步骤102中得到的特征序列与步骤104a中得到的复合HMM模型匹配, 得到一个最佳状态输出序列, 使每一帧特征都有其对应的状态;

[0121] 步骤104c、寻找初始切分点: 由步骤104a可知单个数字HMM模型的状态数为8, 在步骤104b中所得最佳状态输出序列中寻找对应状态为8的整数倍的位置作为初始切分点P(i);

[0122] 步骤104d、更新初始切分点: 依次选取步骤104c中相邻的两个初始切分点P(i-1)和P(i), 并分别作为区间的起始点和终止点。在该区间内, 每K帧组成一段, 共L段, 每段平均

能量为 $E(n)$, n 为段索引号, 计算 $S(n-1) = E(n) - E(n-1)$, $n=2 \cdots L$, 从 $S(n_1) > 0$, $n_1=1 \cdots L-1$ 的索引号开始向后搜索, 若 $S(n_1+1), S(n_1+2), \cdots, S(n_1+q)$ 均大于0, 其中 q 是一个大于1的常数, 则将 n_1 段的起始点作为新的初始切分点代替 $P(i-1)$; 若无该类索引号, 则不更新初始切分点。由初始切分点分割成的不同单元即初始切分单元, 假设初始切分单元个数为 M , 由于最佳状态序列的最大状态为64, 所以初始切分单元个数小于等于8个(该更新过程并未改变初始切分点个数);

[0123] 步骤105、初始切分单元个数判决: 步骤104将语音切分后得到若干个初始切分单元, 对于目标密码语音, 其初始切分单元个数一般近似等于目标密码中目标文本个数; 对于冒认密码语音, 其切分单元个数往往远小于目标密码中目标文本个数。由步骤104可知测试语音初始切分单元数为 M , 假设最少切分单元个数为 T , 当 $M < T$ 时, 系统直接拒绝该请求人, 判决结束, 否则, 执行步骤106;

[0124] 步骤106、强制切分: 当 $8-M > 0$ 时, 取初始切分单元中对应特征段最长的切分单元, 并将该特征段平均切分为 $(8-M+1)$ 份, 强制切分后的切分单元总数变为8;

[0125] 步骤107、计算文本匹配分数: 将步骤106中得到的切分单元对应特征序列与0~9十个目标文本的目标模型HMM进行匹配, 每个切分单元对应10个匹配打分, 假设该打分为 $\text{word_score}(i, j)$, 该变量表示动态密码中第 i 个切分单元与数字 j 的模型的文本匹配分数;

[0126] 步骤108、声纹与文本初步认证:

[0127] 取每个所述切分单元对应的所述切分单元文本匹配分数中 m 个最高分数对应的文本作为待选文本, 若所述待选文本中包含所述切分单元对应的目标文本, 则所述切分单元认证通过, 计算通过的切分单元的总数, 若通过的切分单元总数小于或等于第四阈值, 则文本认证不通过, 身份认证不通过, 判决结束; 若通过的切分单元总数大于所述第四阈值, 则所述输入语音的文本认证通过;

[0128] 判断所述声纹匹配分数是否大于第五阈值, 如是, 则声纹认证通过, 身份认证通过, 判决结束; 如不是, 则将每个所述切分单元与对应目标文本模型的文本打分以及所述声纹匹配分数进行得分规整, 将规整后的打分作为所述判决分类器的输入进行身份认证。

[0129] 如图5所示, 其实施方法如下:

[0130] 步骤108a、每个切分单元各取 m 个最高得分: 由上述步骤106可知, 每个切分单元对应应有10个得分, 各取 m (一般为2或3)个最高打分, 分别对应 m 个待匹配文本;

[0131] 步骤108b、切分单元文本认证: 对每个切分单元进行文本认证, 若切分单元对应的 m 个待匹配文本中包含该切分单元对应的目标文本, 则该切分单元的文本认证通过, 反之, 认证不通过;

[0132] 步骤108c、计算切分单元文本认证通过的总个数 W ;

[0133] 步骤108d、测试语音文本认证: 假设测试语音切分单元文本认证通过的最小数为 p , 当 W 大于 p 时, 则判定该语音文本认证通过, 并转至步骤108e, 否则, 文本认证不通过, 身份认证不通过, 判决结束;

[0134] 步骤108e、测试语音声纹认证: 设置一个较大的声纹阈值, 以保证系统的严格性, 当声纹匹配分数大于阈值时, 声纹认证通过, 该测试语音身份认证通过, 否则, 转至步骤109;

[0135] 步骤109、得分规整: 首先求得大量冒认密码语音对应目标文本模型的打分均值与

方差,在得到测试语音中每个切分单元对应的文本打分后减去冒认得分均值并除以标准差。如图6所示,其实施方法如下:

[0136] 步骤109a、求大量冒认文本打分:依次取0~9的单个数字模型HMM,假设取数字1的模型HMM¹,根据Viterbi算法,取大量非1的冒认语音特征作为模型HMM¹的输入,得到大量冒认文本打分;

[0137] 步骤109b、求均值与标准差:计算每个文本对应的冒认文本打分均值与标准差;

[0138] 步骤109c、零归整及归一化:在步骤107计算文本匹配分数的基础上,找出每个切分单元与其对应目标文本模型的打分,此时每个切分单元对应一个文本打分。根据零归整方法,将每个文本打分分别减去对应文本的冒认打分均值并除以标准差,得到规整后的文本匹配分数,将步骤103中得到的声纹匹配分数与规整后的8个文本匹配分数合并组成一个9维的特征向量score(得分)。由于该特征向量中的声纹匹配分数不论是目标话者还是冒认话者的声纹打分,其打分一般远大于文本匹配分数,因此,又对特征向量增加了归一化处理,使得声纹匹配分数与文本匹配分数均在[0,1]之间。假设该特征向量的最大值和最小值分别为max_score和min_score,对特征向量作线性变换,得到一个新的特征向量new_score = (score-min_score)/(max_score-min_score);

[0139] 步骤110综合判决:利用综合判决分类器对输入特征向量new_score进行判决,对于每一个输入,其输出为1或0,当输出为1时表示测试语音判决通过,输出为0时拒绝测试语音通过。

[0140] 实施例二:

[0141] 针对第一种实施方式中步骤104的对于语音特征初始切分、步骤105的切分单元个数判决,以及步骤106的强制切分,本实施例中采用以下方法进行切分和判决:

[0142] 步骤201,语音信号初始切分;

[0143] 本实施例中,按照所述初始切分单元的长度从大到小的顺序开始拆分,每次将一个所述初始切分单元平均切分成两个段,直至切分单元的总个数等于所述目标文本的个数为止;

[0144] 若强制切分的次数大于等于第二阈值,则强制切分结束;若强制切分的次数小于所述第二阈值,则将当前每个切分的单元分别与每个目标文本HMM进行匹配打分,分别选定最高打分对应的所述目标文本HMM,将所选定的所述目标文本HMM组合成第二复合HMM;

[0145] 将所述语音特征作为所述第二复合HMM的输入进行维特比解码,得到第二状态输出序列,将所述第二状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为切分点,由该切分点对所述语音特征分割得到的不同单元即为所述切分单元,若当前的切分单元个数小于第三阈值,则将当前切分后的切分单元作为所述初始切分单元继续进行强制切分,若当前的切分单元个数大于或等于第三阈值,则强制切分结束,并将上述强制切分后的切分单元作为最终切分单元。如图8所示,包括如下步骤:

[0146] 步骤201a、初始分割:计算语音信号包络,选择8个极大包络处附近区域作为初始分割结果;

[0147] 步骤201b、根据打分对初始分割段判决:将每个分割段对0~9十个数字模型进行打分,每个分割段取最高得分对应的数字,并作为该分割段的判决结果;

[0148] 步骤201c、复合HMM模型的组合:根据上述步骤201b中的分割判决结果,选择相应

的HMM模型,组合成复合HMM模型,该组合过程可参见第一种实施方式中的步骤104a;

[0149] 步骤201d、根据Viterbi解码作进一步分割:根据步骤201c输出的组合模型对输入信号进行Viterbi解码,根据最佳状态序列对信号做进一步分割,该分割过程可参见第一种实施方式中的步骤104c。

[0150] 步骤202、强制切分:将分割段长度大小排序,按大小顺序平均分割为两个,直至分割为8段为止。

[0151] 步骤203、初始切分判决:若步骤201d的分割段个数小于X(相当于第三阈值, $X < 8$)个,则转至步骤201b,将步骤202的输出结果作为步骤201b的输入,继续进行分割;若分割段个数大于或等于X,则分割结束。设定一个最大迭代次数D(相当于第二阈值),若该过程迭代次数等于D时,步骤201b的分割段数仍小于X个,则停止迭代,并拒绝该语音;若该过程迭代次数小于D或等于D时分割段数大于等于X,则继续进行判决,执行第一种实施方式中的步骤107及后续步骤。

[0152] 图9为本发明实施例的一种身份认证的装置的示意图,本实施例的装置包括PNN分类器,如图9所示,本实施例的装置包括:

[0153] 声纹匹配模块,设置为获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

[0154] 切分模块,设置为根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于阈值,则判定所述输入语音为非法语音;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

[0155] 文本匹配模块,设置为将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

[0156] 认证模块,设置为根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的所述PNN分类器进行身份认证。

[0157] 在一可选实施例中,所述装置还包括处理模块,

[0158] 所述声纹匹配模块,是设置为将目标语音与目标声纹模型进行匹配得到第一声纹打分,将非目标语音与所述目标声纹模型进行匹配得到第二声纹打分;

[0159] 所述文本匹配模块,是设置为将所述目标语音与所述目标文本模型进行匹配得到第一文本打分,为将所述非目标语音与所述目标文本模型进行匹配得到第二文本打分;

[0160] 所述处理模块,设置为将所述第一文本打分和第一声纹打分组合成为所述PNN分类器的接受特征信息,将所述第二文本打分和第二声纹打分组合成为所述PNN分类器的拒绝特征信息;

[0161] 所述PNN分类器,根据所述接受特征信息和所述拒绝特征信息进行训练。

[0162] 在一可选实施例中,所述处理模块,还设置为依次选取所述目标文本模型,取非目标文本的语音特征与对应的所述目标文本模型匹配,得到冒认文本打分,获取所述目标文本模型对应的冒认文本打分的均值及标准差;将所述第一文本打分和所述第二文本打分分别减去对应的所述冒认文本打分的均值且除以所述标准差,分别得到规整后的文本打分;合并规整后的第一文本打分和所述第一声纹打分,获取每一目标文本对应的最大值和最小

值;利用该最大值和最小值将规整后的第一文本打分和所述第一声纹打分进行归一化,作为所述PNN分类器的接受特征信息;合并规整后的第二文本打分和所述第二声纹打分,获取每一目标文本对应的最大值和最小值;利用该最大值和最小值将规整后的第二文本打分和所述第二声纹打分进行归一化,作为所述PNN分类器的拒绝特征信息。

[0163] 在一可选实施例中,所述切分模块,根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元,包括:根据目标密码中的目标文本序列,将对应的目标文本隐马尔可夫模型HMM组合成第一复合HMM;将所述语音特征作为所述第一复合HMM的输入进行维特比解码,得到第一状态输出序列,将所述第一状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为初始切分点;依次选取所述相邻两个初始切分点作为区间起止点,在所述区间内,以指定帧为单位计算平均能量,寻找平均能量连续指定次增大的点,并将开始增大的点作为新的初始切分点,由所述初始切分点分割成的所述初始切分单元。

[0164] 在一可选实施例中,所述切分模块,将对应的目标文本HMM组合成第一复合HMM,包括:所述第一复合HMM的状态数为单个目标文本HMM的状态数总和;所述第一复合HMM的每个状态具有的高斯混合模型参数与所述单个目标文本HMM的每个状态具有的高斯混合模型参数相同;将所述单个目标文本HMM的状态转移矩阵中的最后一个状态自身转移概率设为0,转移到下一个状态的状态转移概率设为1;所述目标文本的最后一个单个目标文本HMM的状态转移概率矩阵不作改变;将所述单个目标文本HMM的状态转移概率矩阵按照所述目标文本的单个目标文本排列顺序合并,得到所述复合HMM的状态转移概率矩阵。

[0165] 在一可选实施例中,所述切分模块,对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:选择特征段最长的所述初始切分单元进行强制切分,使得强制切分后的所有切分单元的总个数与预设的目标文本的个数相同。

[0166] 在一可选实施例中,所述切分模块,对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同,包括:按照所述初始切分单元的长度从大到小的顺序开始拆分,每次将一个所述初始切分单元平均切分成两个段,直至切分单元的总个数等于所述目标文本的个数为止;若强制切分的次数大于等于第二阈值,则强制切分结束;若强制切分的次数小于所述第二阈值,则将当前每个切分的单元分别与每个目标文本HMM进行匹配打分,分别选定最高打分对应的所述目标文本HMM,将所选定的所述目标文本HMM组合成第二复合HMM;将所述语音特征作为所述第二复合HMM的输入进行维特比解码,得到第二状态输出序列,将所述第二状态输出序列中为单个目标文本HMM的状态数的整数倍的状态对应的位置作为切分点,由该切分点对所述语音特征分割得到的不同单元即为切分单元,若当前的切分单元个数小于第三阈值,则将当前切分后的切分单元作为所述初始切分单元继续进行强制切分,若当前的切分单元个数大于或等于第三阈值,强制切分结束,并将上述强制切分后的切分单元作为最终切分单元。

[0167] 在一可选实施例中,所述文本匹配模块,将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数,包括:将每个所述切分单元的语音特征作为每个目标文本隐马尔可夫模型HMM的输入,将根据维特比算法获得的输出概率作为对应的切分单元文本匹配分数。

[0168] 在一可选实施例中,所述认证模块,根据所述切分单元文本匹配分数、所述声纹匹

配分数和预先训练的判决分类器进行身份认证,包括:取每个所述切分单元对应的所述切分单元文本匹配分数中 m 个最高分数对应的文本作为待选文本,若所述待选文本中包含所述切分单元对应的目标文本,则所述切分单元认证通过,计算通过的切分单元的总数,若通过的切分单元总数小于或等于第四阈值,则文本认证不通过,身份认证不通过,判决结束;若通过的切分单元总数大于所述第四阈值,则所述输入语音的文本认证通过;判断所述声纹匹配分数是否大于第五阈值,如是,则声纹认证通过,身份认证通过,判决结束;如不是,则将每个所述切分单元与对应目标文本模型的文本打分以及所述声纹匹配分数进行得分规整,将规整后的打分作为所述PNN分类器的输入进行身份认证。

[0169] 本发明实施例还提供了一种计算机可读存储介质。可选地,在本实施例中,上述存储介质可以被设置为存储由处理器执行的程序代码,程序代码的步骤如下:

[0170] S1、获取输入语音的语音特征,将所述语音特征与预存的目标声纹模型进行匹配,得到声纹匹配分数;

[0171] S2、根据所述语音特征和预设的目标文本模型对所述输入语音进行切分,获取初始切分单元以及初始语音切分单元的个数,如所述初始语音切分单元的个数小于第一阈值,则判定所述输入语音为非法语音;如所述初始语音切分单元的个数大于或等于第一阈值,则对所述初始切分单元进行强制切分,使得切分单元的总个数与预设的目标文本的个数相同;

[0172] S3、将每个所述切分单元的语音特征与所有所述目标文本模型进行匹配,得到每个所述切分单元与每个所述目标文本模型的切分单元文本匹配分数;

[0173] S4、根据所述切分单元文本匹配分数、所述声纹匹配分数和预先训练的概率神经网络PNN分类器进行身份认证。

[0174] 可选地,在本实施例中,上述存储介质可以包括但不限于:U盘、只读存储器(Read-Only Memory,简称为ROM)、随机存取存储器(Random Access Memory,简称为RAM)、移动硬盘、磁碟或者光盘等每种可以存储程序代码的介质。

[0175] 本领域普通技术人员可以理解上述方法中的全部或部分步骤可通过程序来指令相关硬件完成,所述程序可以存储于计算机可读存储介质中,如只读存储器、磁盘或光盘等。可选地,上述实施例的全部或部分步骤也可以使用一个或多个集成电路来实现。相应地,上述实施例中的各模块/单元可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。本发明不限制于任何特定形式的硬件和软件的结合。

[0176] 以上仅为本发明的优选实施例,当然,本发明还可有其他多种实施例,在不背离本发明精神及其实质的情况下,熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形,但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

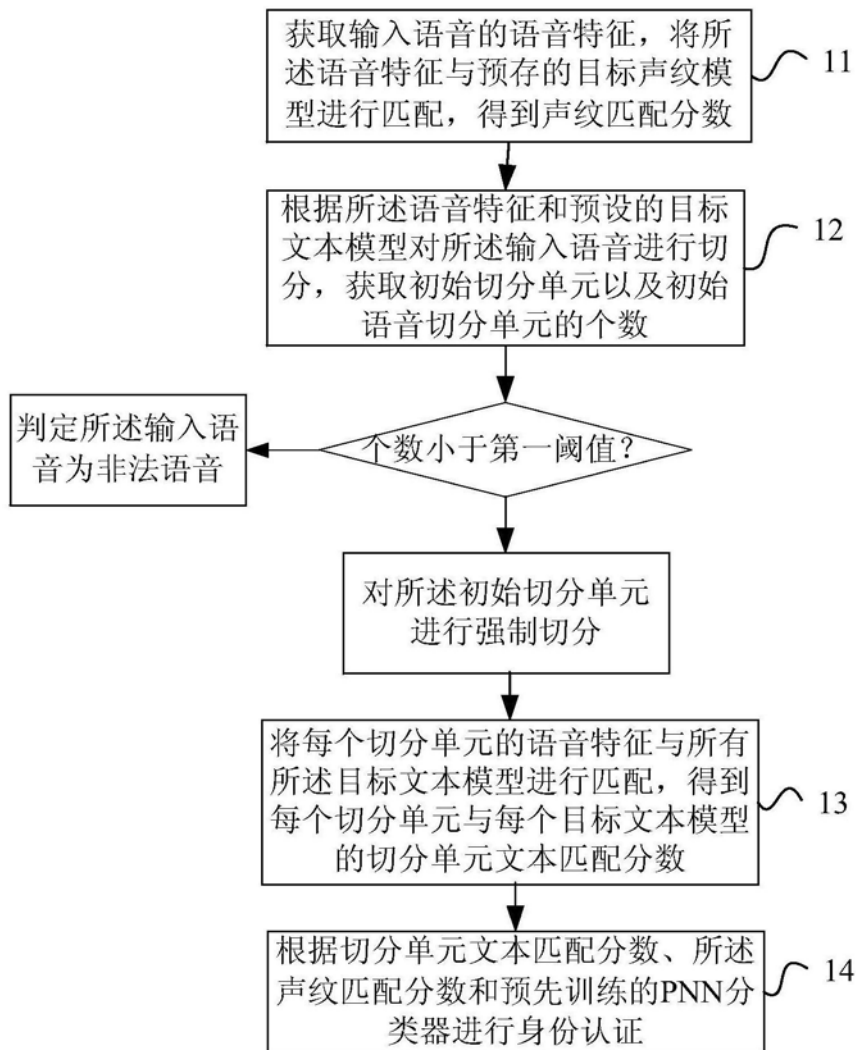


图1

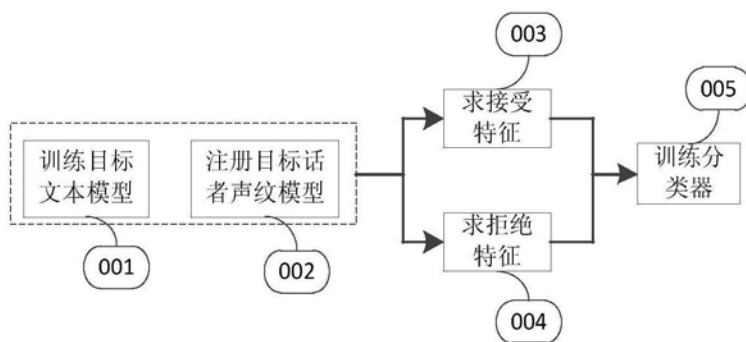


图2

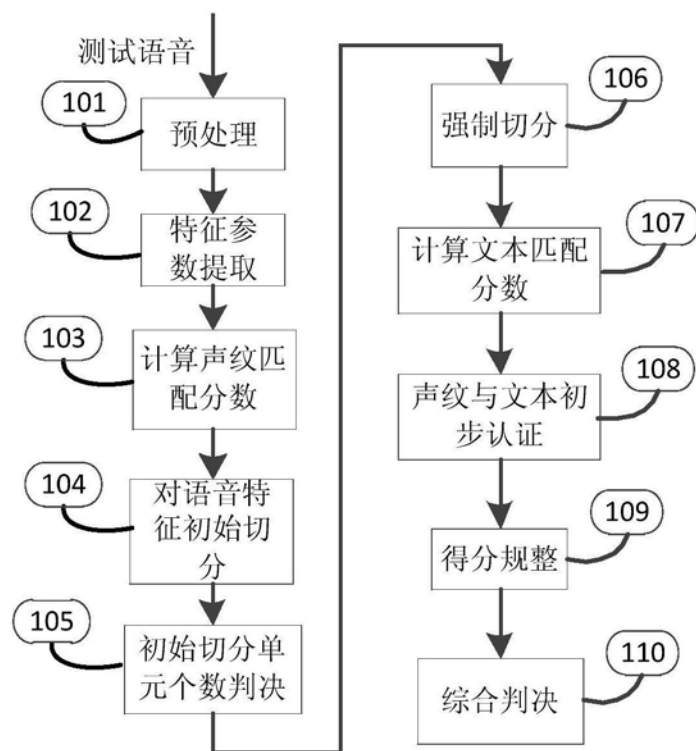


图3

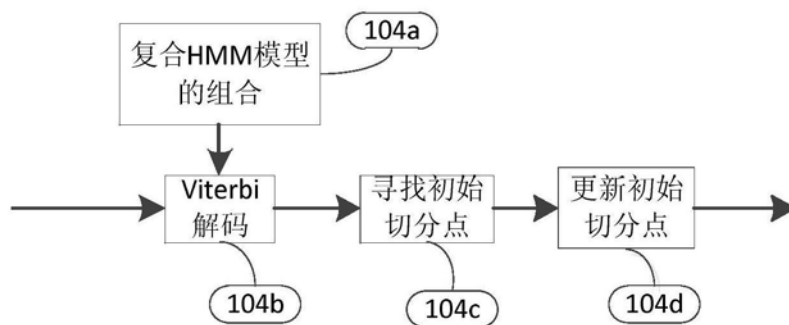


图4

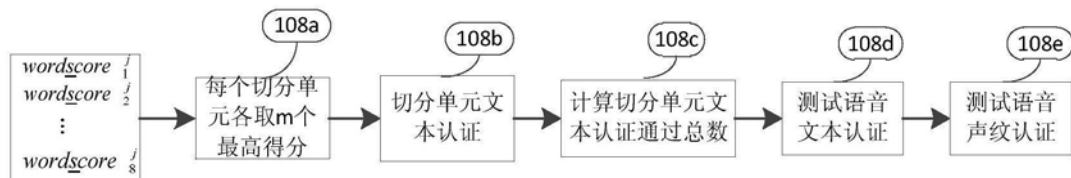


图5

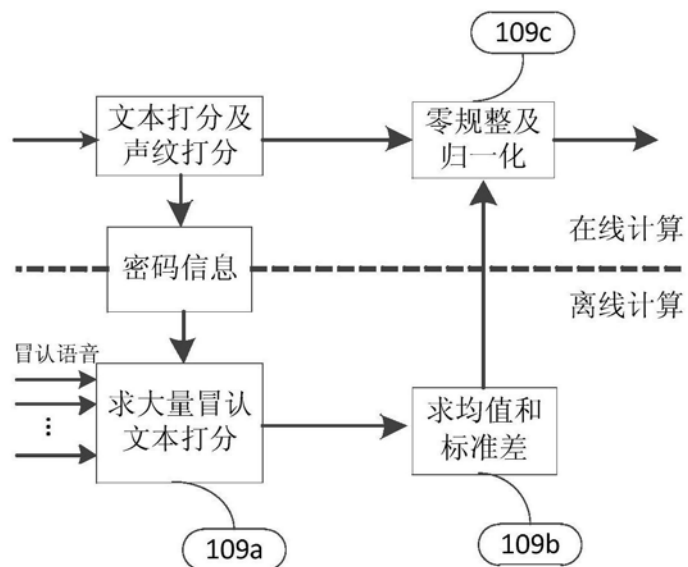


图6

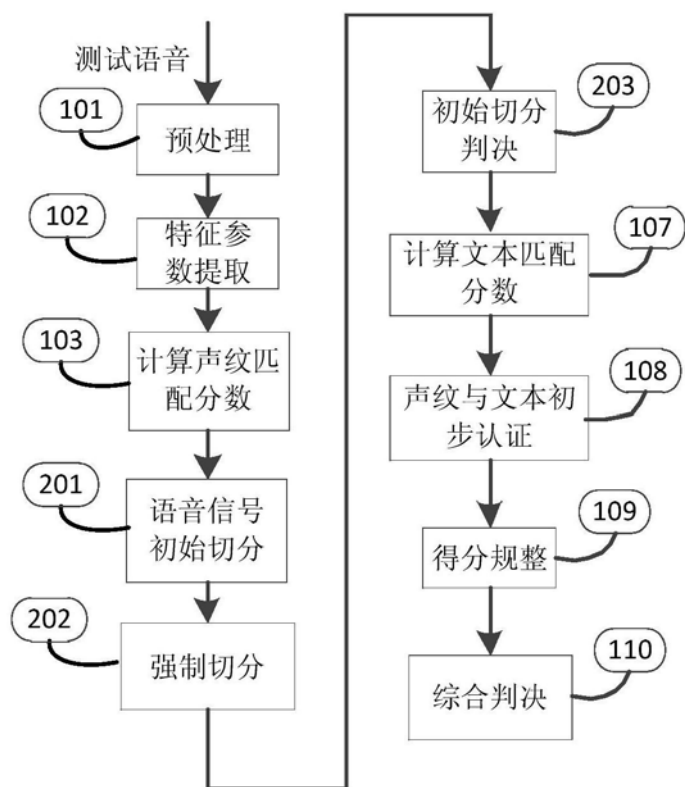


图7

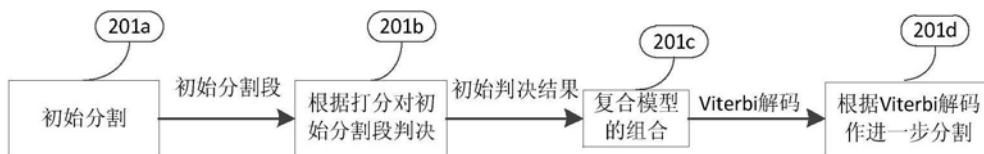


图8

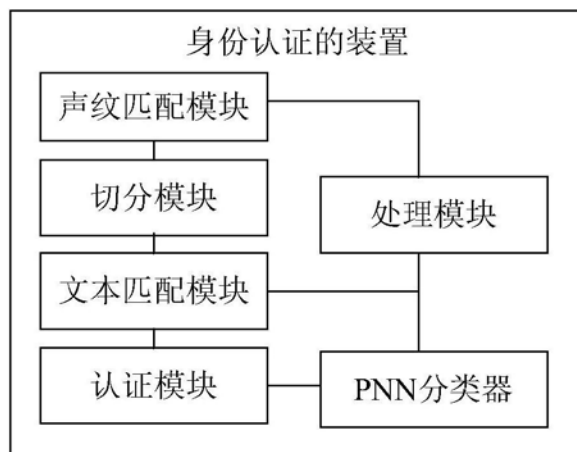


图9