(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 102238189 B
(45)授权公告日 2013.12.11

(21)申请号 201110218042.9

(22)申请日 2011.08.01

(73)专利权人 安徽科大讯飞信息科技股份有限公司
地址 230088 安徽省合肥市高新开发区黄山路616号

(72)发明人 何婷婷 胡国平 胡郁 王智国 刘庆峰

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 赵景平 逯长明

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 9/32(2006.01)
G10L 17/08(2013.01)

(56)对比文件
US 7386448 B1,2008.06.10,
EP 1526505 A1,2005.04.27,
US 2009171660 A1,2009.07.02,
CN 101124623 A,2008.02.13,
CN 1547191 A,2004.11.17,

审查员 张华晶

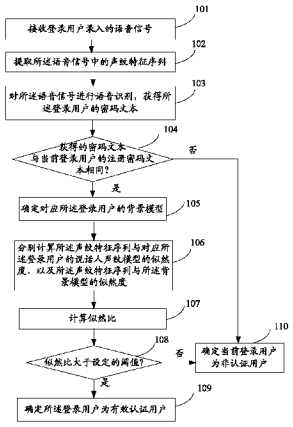
权利要求书3页 说明书12页 附图5页

(54)发明名称

声纹密码认证方法及系统

(57)摘要

本发明公开了一种声纹密码认证方法及系统，该方法包括：接收登录用户录入的语音信号；提取所述语音信号中的声纹特征序列；对所述语音信号进行语音识别，获得所述登录用户的密码内容；如果获得的密码内容与对应所述登录用户的注册密码文本不同，则确定所述登录用户为非认证用户；否则，分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型和为所述登录用户选择的背景模型的似然度，背景模型包括：与文本无关的通用背景模型、以及与文本相关的优化背景模型；根据得到的似然度，计算似然比；如果所述似然比大于设定的阈值，则确定所述登录用户为有效认证用户，否则确定所述登录用户为非认证用户。本发明可以提高声纹密码认证的准确率。



CN 102238189 B

1. 一种声纹密码认证方法，其特征在于，包括：

接收登录用户录入的语音信号；

提取所述语音信号中的声纹特征序列；

对所述语音信号进行语音识别，获得所述登录用户的密码文本；

如果获得的密码文本与对应所述登录用户的注册密码文本不同，则确定所述登录用户为非认证用户；

如果获得的密码文本与对应所述登录用户的注册密码文本相同，

则确定对应所述登录用户的背景模型，所述背景模型包括：与文本无关的通用背景模型、以及与文本相关的优化背景模型；

分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述背景模型的似然度；

根据所述声纹特征序列与说话人声纹模型的似然度、以及所述声纹特征序列与背景模型的似然度，计算似然比；

如果所述似然比大于设定的阈值，则确定所述登录用户为有效认证用户，否则确定所述登录用户为非认证用户；

其中，所述似然比具体为：所述声纹特征序列与说话人声纹模型的似然度和所述声纹特征序列与背景模型的似然度之比。

2. 如权利要求 1 所述的方法，其特征在于，所述确定对应所述登录用户的背景模型包括：

如果有与所述登录用户的密码文本对应的优化背景模型，则选择该优化背景模型作为对应所述登录用户的背景模型；否则选择所述通用背景模型作为对应所述登录用户的背景模型。

3. 如权利要求 1 所述的方法，其特征在于，所述方法还包括：

将登录用户录入的语音信号或者从登录用户录入的语音信号中提取的声纹特征序列写入与所述登录用户录入的语音信号相应的密码文本对应的缓存区；

接收注册用户录入的注册语音信号；

对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本；

将所述注册语音信号或者从所述注册语音信号中提取的声纹特征序列写入与该注册语音信号相应的密码文本对应的缓存区；

根据所述注册用户录入的注册语音信号训练对应所述注册用户的说话人声纹模型；

实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型。

4. 如权利要求 3 所述的方法，其特征在于，所述实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型：

如果一个缓存区中存储的数据量达到第一预设值，并且当前没有与该缓存区对应密码文本相关的优化背景模型，则以所述通用背景模型为初始模型，根据该缓存区中的数据生成与该缓存区对应密码文本相关的优化背景模型，并删除该缓存区中存储的数据；如果一个缓存区中存储的数据量达到第一预设值，并且当前有与该缓存区对应密码文本相关的优化背景模型，则以该优化背景模型为初始模型，根据该缓存区中的数据更新该优化背景模

型，并删除该缓存区中存储的数据。

5. 如权利要求 3 所述的方法，其特征在于，所述实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型：

如果一个缓存区中存储的数据量达到第二预设值的整数倍，则以所述通用背景模型为初始模型，根据该缓存区中的数据重新生成与该缓存区对应密码文本相关的优化背景模型。

6. 如权利要求 3 至 5 任一项所述的方法，其特征在于，所述注册用户录入的注册语音信号重复多次；

所述对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本包括：

分别对每次录入的注册语音信号进行语音识别，得到多个识别结果及与各识别结果对应的识别似然度得分；

选择具有最高似然度得分的识别结果作为所述注册用户的注册密码文本。

7. 一种声纹密码认证系统，其特征在于，包括：

接收单元，用于在用户登录时，接收登录用户录入的语音信号；

声纹特征提取单元，用于提取所述语音信号中的声纹特征序列；

语音识别单元，用于对所述语音信号进行语音识别，获得所述登录用户的密码文本；

判断单元，用于判断所述语音识别单元获得的密码文本与对应所述登录用户的注册密码是否相同；

认证结果单元，用于在所述判断单元的判断结果是所述语音识别单元获得的密码文本与对应所述登录用户的注册密码文本不同时，确定所述登录用户为非认证用户；

模型确定单元，用于在所述判断单元的判断结果是所述语音识别单元获得的密码文本与所述登录用户的注册密码文本相同时，确定对应所述登录用户的背景模型，所述背景模型包括：与文本无关的通用背景模型、以及与文本相关的优化背景模型；

第一计算单元，用于分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述模型确定单元确定的背景模型的似然度；

第二计算单元，用于根据所述声纹特征序列与说话人声纹模型的似然度、以及所述声纹特征序列与背景模型的似然度，计算似然比；其中，所述似然比具体为：所述声纹特征序列与说话人声纹模型的似然度和所述声纹特征序列与背景模型的似然度之比；

所述判断单元，还用于判断所述第二计算单元计算得到的似然比是否大于设定的阈值；

所述认证结果单元，还用于在所述判断单元的判断结果是所述第二计算单元计算得到的似然比大于设定的阈值时，确定所述登录用户为有效认证用户，否则确定所述登录用户为非认证用户。

8. 如权利要求 7 所述的系统，其特征在于，所述系统还包括：

检查单元，用于检查是否存在与所述登录用户的注册密码文本对应的优化背景模型；

所述模型确定单元，具体用于在所述检查单元的检查结果是有与所述登录用户的注册密码文本对应的优化背景模型时，选择该优化背景模型作为对应所述登录用户的背景模型；否则选择所述通用背景模型作为对应所述登录用户的背景模型。

9. 如权利要求 8 所述的系统，其特征在于，

所述语音识别单元，还用于将登录用户录入的语音信号或者从登录用户录入的语音信号中提取的声纹特征序列写入与所述登录用户录入的语音信号相应的密码文本对应的缓存区；

所述接收单元，还用于接收注册用户录入的注册语音信号；

所述语音识别单元，还用于对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本；

所述系统还包括：

说话人声纹模型构建单元，用于根据所述注册用户录入的注册语音信号训练对应所述注册用户的说话人声纹模型；

背景模型构建单元，用于实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型。

10. 如权利要求 9 所述的系统，其特征在于，

所述背景模型构建单元，具体用于在一个缓存区中存储的数据量达到第一预设值，并且当前没有与该缓存区对应密码文本相关的优化背景模型时，以所述通用背景模型为初始模型，根据该缓存区中的数据生成与该缓存区对应密码文本相关的优化背景模型，并删除该缓存区中存储的数据；在一个缓存区中存储的数据量达到第一预设值，并且当前有与该缓存区对应密码文本相关的优化背景模型时，以该优化背景模型为初始模型，根据该缓存区中的数据更新该优化背景模型，并删除该缓存区中存储的数据。

11. 如权利要求 9 所述的系统，其特征在于，

所述背景模型构建单元，具体用于在一个缓存区中存储的数据量达到第二预设值的整数倍，则以所述通用背景模型为初始模型，根据该缓存区中的数据重新生成与该缓存区对应密码文本相关的优化背景模型。

12. 如权利要求 9 至 11 任一项所述的系统，其特征在于，所述注册用户录入的注册语音信号重复多次；

所述语音识别单元分别对每次录入的注册语音信号进行语音识别，得到多个识别结果及与各识别结果对应的识别似然度得分；

所述系统还包括：

密码确定单元，用于从所述语音识别单元得到的多个识别结果中选择选择具有最高似然度得分的识别结果作为所述注册用户的注册密码文本。

声纹密码认证方法及系统

技术领域

[0001] 本发明涉及密码认证技术领域，特别涉及一种声纹密码认证方法及系统。

背景技术

[0002] 声纹识别 (Voiceprint Recognition，VPR) 也称为说话人识别，有两类，即说话人辨认和说话人确认。前者用以判断某段语音是若干人中的哪一个所说的，是“多选一”问题；而后者用以确认某段语音是否是指定的某个人所说的，是“一对一判别”问题。不同的任务和应用会使用不同的声纹识别技术。

[0003] 声纹认证是指根据采集到的语音信号确认说话人身份，属于“一对一”的判别问题。现今主流的声纹认证系统采用了基于假设检验的框架，通过分别计算声纹信号相对于说话人声纹模型以及背景模型的似然度并比较它们的似然比和预先根据经验设置的阈值大小来确认。显然背景模型和说话人声纹模型的精确度将直接影响到声纹认证效果，在基于数据驱动的统计模型设定下训练数据量越大则模型效果越好。

[0004] 声纹密码认证是一种文本相关的说话人身份认证方法。该方法要求用户语音输入确定密码文本，并据此确认说话人身份。在该应用中用户注册及身份认证均采用确定密码文本的语音输入，因而其声纹往往较为一致，相应的可取得相比于文本无关的说话人确认更好的认证效果。

[0005] 在声纹密码认证系统中，用户以语音输入信号替代传统的字串密码输入，相应的认证系统以说话人声纹模型的形式保存用户的声纹密码。现有的声纹密码认证系统大都是采用计算声纹信号相对于说话人声纹模型及背景模型的似然度，并比较其似然度比和预设的阈值大小来确认用户身份。因此，背景模型和说话人声纹模型的精确程度将直接影响到声纹密码认证的效果。

[0006] 在现有技术中，声纹密码认证系统普遍采用通用背景模型，用于模拟文本无关的用户声纹特性，具体是在采集的多说话人数据上以离线方式训练得到单一的通用背景模型。这种通用背景模型虽然有较好的普适性，但模型描述不够精确，区分度较低，在一定程度上影响了密码认证的准确性。

发明内容

[0007] 本发明实施例提供一种声纹密码认证方法及系统，以提高基于声纹密码进行身份认证的准确率。

[0008] 一种声纹密码认证方法，包括：

[0009] 接收登录用户录入的语音信号；

[0010] 提取所述语音信号中的声纹特征序列；

[0011] 对所述语音信号进行语音识别，获得所述登录用户的密码文本；

[0012] 如果获得的密码文本与对应所述登录用户的注册密码文本不同，则确定所述登录用户为非认证用户；

[0013] 如果获得的密码文本与对应所述登录用户的注册密码文本相同，

[0014] 则确定对应所述登录用户的背景模型，所述背景模型包括：与文本无关的通用背景模型、以及与文本相关的优化背景模型；

[0015] 分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述背景模型的似然度；

[0016] 根据所述声纹特征序列与说话人声纹模型的似然度、以及所述声纹特征序列与背景模型的似然度，计算似然比；

[0017] 如果所述似然比大于设定的阈值，则确定所述登录用户为有效认证用户，否则确定所述登录用户为非认证用户。

[0018] 优选地，所述确定对应所述登录用户的背景模型包括：

[0019] 如果有与所述登录用户的密码文本对应的优化背景模型，则选择该优化背景模型作为对应所述登录用户的背景模型；否则选择所述通用背景模型作为对应所述登录用户的背景模型。

[0020] 优选地，所述方法还包括：

[0021] 将登录用户录入的语音信号或者从登录用户录入的语音信号中提取的声纹特征序列写入与所述登录用户录入的语音信号相应的密码文本对应的缓存区；

[0022] 接收注册用户录入的注册语音信号；

[0023] 对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本；

[0024] 将所述注册语音信号或者从所述注册语音信号中提取的声纹特征序列写入与该注册语音信号相应的密码文本对应的缓存区；

[0025] 根据所述注册用户录入的注册语音信号训练对应所述注册用户的说话人声纹模型；

[0026] 实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型。

[0027] 可选地，所述实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型：

[0028] 如果一个缓存区中存储的数据量达到第一预设值，并且当前没有与该缓存区对应密码文本相关的优化背景模型，则以所述通用背景模型为初始模型，根据该缓存区中的数据生成与该缓存区对应密码文本相关的优化背景模型，并删除该缓存区中存储的数据；如果一个缓存区中存储的数据量达到第一预设值，并且当前有与该缓存区对应密码文本相关的优化背景模型，则以该优化背景模型为初始模型，根据该缓存区中的数据更新该优化背景模型，并删除该缓存区中存储的数据。

[0029] 可选地，所述实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型：

[0030] 如果一个缓存区中存储的数据量达到第二预设值的整数倍，则以所述通用背景模型为初始模型，根据该缓存区中的数据重新生成与该缓存区对应密码文本相关的优化背景模型。

[0031] 优选地，所述注册用户录入的注册语音信号重复多次；

[0032] 所述对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本包

括：

[0033] 分别对每次录入的注册语音信号进行语音识别，得到多个识别结果及与各识别结果对应的识别似然度得分；

[0034] 选择具有最高似然度得分的识别结果作为所述注册用户的注册密码文本。

[0035] 一种声纹密码认证系统，包括：

[0036] 接收单元，用于在用户登录时，接收登录用户录入的语音信号；

[0037] 声纹特征提取单元，用于提取所述语音信号中的声纹特征序列；

[0038] 语音识别单元，用于对所述语音信号进行语音识别，获得所述登录用户的密码文本；

[0039] 判断单元，用于判断所述语音识别单元获得的密码文本与对应所述登录用户的注册密码是否相同；

[0040] 认证结果单元，用于在所述判断单元的判断结果是所述语音识别单元获得的密码文本与对应所述登录用户的注册密码文本不同时，确定所述登录用户为非认证用户；

[0041] 模型确定单元，用于在所述判断单元的判断结果是所述语音识别单元获得的密码文本与所述登录用户的注册密码文本相同时，确定对应所述登录用户的背景模型，所述背景模型包括：与文本无关的通用背景模型、以及与文本相关的优化背景模型；

[0042] 第一计算单元，用于分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述模型确定单元确定的背景模型的似然度；

[0043] 第二计算单元，用于根据所述声纹特征序列与说话人声纹模型的似然度、以及所述声纹特征序列与背景模型的似然度，计算似然比；

[0044] 所述判断单元，还用于判断所述第二计算单元计算得到的似然比是否大于设定的阈值；

[0045] 所述认证结果单元，还用于在所述判断单元的判断结果是所述第二计算单元计算得到的似然比大于设定的阈值时，确定所述登录用户为有效认证用户，否则确定所述登录用户为非认证用户。

[0046] 优选地，所述系统还包括：

[0047] 检查单元，用于检查是否存在与所述登录用户的注册密码文本对应的优化背景模型；

[0048] 所述模型确定单元，具体用于在所述检查单元的检查结果是有与所述登录用户的注册密码文本对应的优化背景模型时，选择该优化背景模型作为对应所述登录用户的背景模型；否则选择所述通用背景模型作为对应所述登录用户的背景模型。

[0049] 优选地，所述语音识别单元，还用于将登录用户录入的语音信号或者从登录用户录入的语音信号中提取的声纹特征序列写入与所述登录用户录入的语音信号相应的密码文本对应的缓存区；

[0050] 所述接收单元，还用于接收注册用户录入的注册语音信号；

[0051] 所述语音识别单元，还用于对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本；

[0052] 所述系统还包括：

[0053] 说话人声纹模型构建单元，用于根据所述注册用户录入的注册语音信号训练对应

所述注册用户的说话人声纹模型；

[0054] 背景模型构建单元，用于实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型。

[0055] 可选地，所述背景模型构建单元，具体用于在一个缓存区中存储的数据量达到第一预设值，并且当前没有与该缓存区对应密码文本相关的优化背景模型时，以所述通用背景模型为初始模型，根据该缓存区中的数据生成与该缓存区对应密码文本相关的优化背景模型，并删除该缓存区中存储的数据；在一个缓存区中存储的数据量达到第一预设值，并且当前有与该缓存区对应密码文本相关的优化背景模型时，以该优化背景模型为初始模型，根据该缓存区中的数据更新该优化背景模型，并删除该缓存区中存储的数据。

[0056] 可选地，所述背景模型构建单元，具体用于在一个缓存区中存储的数据量达到第二预设值的整数倍，则以所述通用背景模型为初始模型，根据该缓存区中的数据重新生成与该缓存区对应密码文本相关的优化背景模型。

[0057] 优选地，所述注册用户录入的注册语音信号重复多次；

[0058] 所述语音识别单元分别对每次录入的注册语音信号进行语音识别，得到多个识别结果及与各识别结果对应的识别似然度得分；

[0059] 所述系统还包括：

[0060] 密码确定单元，用于从所述语音识别单元得到的多个识别结果中选择选择具有最高似然度得分的识别结果作为所述注册用户的注册密码文本。

[0061] 本发明实施例提供的声纹密码认证方法及系统，在进行用户身份识别时，不仅对用户登录时录入的语音信号进行语音识别，确定其密码内容，而且对其进行声纹认证，在进行声纹认证时，基于多背景模型，即与文本无关的通用背景模型及与文本相关的优化背景模型，通过选择合适的背景模型实现精确匹配，有效地提高了基于声纹密码进行身份认证的准确率。

附图说明

[0062] 为了更清楚地说明本发明实施的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，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。

[0063] 图 1 是本发明实施例声纹密码认证方法的流程图；

[0064] 图 2 是本发明实施例中与文本无关的通用背景模型的构建流程图；

[0065] 图 3 是本发明实施例中构建与文本相关的优化背景模型的一种流程图；

[0066] 图 4 是本发明实施例中对注册用户录入的注册语音信号进行语音识别的流程图；

[0067] 图 5 是本发明实施例声纹密码认证系统的一种结构示意图；

[0068] 图 6 是本发明实施例声纹密码认证系统的另一种结构示意图；

[0069] 图 7 是本发明实施例声纹密码认证系统的另一种结构示意图。

具体实施方式

[0070] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于

本发明中的实施例，本领域普通技术人员在没有作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。

[0071] 如图 1 所示，是本发明实施例声纹密码认证方法的流程图，包括以下步骤：

[0072] 步骤 101，接收登录用户录入的语音信号。

[0073] 步骤 102，提取所述语音信号中的声纹特征序列。

[0074] 所述声纹特征序列包含一组声纹特征，可以有效地区分不同的说话人，且对同一说话人的变化保持相对稳定。

[0075] 比如，所述声纹特征主要有：谱包络参数语音特征，基音轮廓、共振峰频率带宽特征，线性预测系数，倒谱系数等。考虑到上述声纹特征的可量化性、训练样本的数量和系统性能的评价等问题，可以选用 MFCC(Mel Frequency Cepstrum Coefficient，Mel 频率倒谱系数）特征，对窗长 25ms 帧移 10ms 的每帧语音数据做短时分析得到 MFCC 参数及其一阶二阶差分，共计 39 维。这样，每句语音信号可以量化为一个 39 维声纹特征序列 X。

[0076] 步骤 103，对所述语音信号进行语音识别，获得登录用户的密码文本。

[0077] 具体的语音识别的处理方式可以采用一些现有的方式，在此不再详细说明。

[0078] 步骤 104，判断获得的密码文本与当前登录用户的注册密码文本是否相同；如果是，则执行步骤 105；否则，执行步骤 110。

[0079] 步骤 105，确定对应所述登录用户的背景模型。

[0080] 其中，说话人声纹模型用于模拟已注册用户在确定密码文本上的发音特点，背景模型用于模拟多说话人的发音共性。

[0081] 在本发明实施例中，说话人声纹模型可以在用户注册时根据用户录入的注册语音信号构建，具体可以采用现有技术中的一些构建方式。背景模型的构建可以综合采用两种方式分别构建与文本无关的通用背景模型和与文本相关的优化背景模型，其中，与文本无关的通用背景模型可以通过预先采集的多说话人数据以离线方式训练得到，具体的训练过程可以参照现有技术中的一些处理方式，对此本发明实施例不做限定；与文本相关的优化背景模型可以根据记录的用户注册和登录时录入的语音信号中提取的声纹特征序列以在线方式训练得到。

[0082] 相应地，在本步骤中，可以根据需要，有多种不同的方式来选择对应所述登录用户的背景模型，对此将在后面详细说明。

[0083] 步骤 106，分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述背景模型的似然度。

[0084] 上述说话人声纹模型可以在用户注册时根据注册语音信号在线训练得到。比如，以通用背景模型为初始模型通过各种自适应方法根据少量说话人数据调整模型部分参数，如目前最为常用的基于最大后验概率 (Maximum A Posterior，MAP) 的自适应算法等，将用户声纹共性自适应为当前说话人个性。当然，还可以采用其他方式训练得到说话人声纹模型，对此本发明实施例不做限定。

[0085] 假设得到帧数为 T 的声纹特征序列 X，则其相应于背景模型的似然度为：

[0086] $$p(X \mid UBM) = \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M} c_m N(X_t; \mu_m, \Sigma_m) \qquad (1)$$

[0087] 其中，c_m 是第 m 个高斯的加权系数，满足 $\sum_{m=1}^{M} c_m = 1$。μ_m 以及 Σ_m 分别是第 m 个高斯的均值和方差。其中 N(.) 满足正态分布，用于计算 t 时刻的声纹特征矢量 X_t 在单高斯分量上的似然度：

[0088] $$N(X_t; \mu_m, \Sigma_m) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_m|}} e^{-\frac{1}{2}(X_t-\mu_m)\Sigma_m^{-1}(X_t-\mu_m)} \quad (2)$$

[0089] 所述声纹特征序列 X 相应于说话人声纹模型的似然度的计算与上述类似，在此不再详细说明。

[0090] 步骤 107，根据所述声纹特征序列与说话人声纹模型的似然度、以及所述声纹特征序列与背景模型的似然度，计算似然比。

[0091] 似然比为：$$p = \frac{p(X|U)}{p(X|UBM)} \quad (3)$$

[0092] 其中，p(X|U) 为所述声纹特征与说话人声纹模型的似然度，p(X|UBM) 为所述声纹特征与背景模型的似然度。

[0093] 步骤 108，判断所述似然比是否大于设定的阈值；如果是，则执行步骤 109；否则执行步骤 110。

[0094] 上述阈值可以由系统预先设定，一般来说，该阈值越大，则系统的灵敏度越高，要求用户在登录时尽可能按照注册时录入的语音信号（即密码）的发音，反之，则系统的灵敏度较低，允许用户登录时录入的语音信号的发音与注册时的发音存在一定的变化。

[0095] 步骤 109，确定登录用户为有效认证用户。

[0096] 步骤 110，确定登录用户为非认证用户。

[0097] 需要说明的是，为了提高系统的鲁棒性，在上述步骤 101 和步骤 102 之间，还可以对所述语音信号进行降噪处理，比如，首先通过对语音信号的短时能量和短时过零率分析，将连续的语音信号分割成独立的语音片断和非语音片断。然后通过前端降噪处理减少信道噪音及背景噪音的干扰，提高语音信噪比，为后续系统处理提供干净的信号。

[0098] 前面提到，在本发明实施例中，背景模型可以包括：与文本无关的通用背景模型、以及与文本相关的优化背景模型，而且可以根据需要，有多种不同的方式来选择对应所述登录用户的背景模型，比如，可以在系统初始化阶段（比如，可以设定一定的时间段），选择与文本无关的通用背景模型，以适应用户录入的各种不同声纹密码；而随着系统的运行，采集到的与特定密码文本相关的用户数据不断增加，可以根据这些用户数据训练得到与该密码文本相关的优化背景模型。此后，可以根据上述步骤 103 获得的当前登录用户的密码文本来选择相应的背景模型。当然，为了简化实现上的复杂度，也可以从系统启动开始，就根据当前登录用户的密码文本来选择相应的背景模型。

[0099] 上述与文本无关的通用背景模型可以采用现有技术中的一些方式，比如采用 1024 或者更大高斯数的混合高斯模型来构建，其模型参数训练过程如图 2 所示。

[0100] 步骤 201，从多说话人训练语音信号中分别提取声纹特征，每个声纹特征作为一个特征矢量。

[0101] 步骤 202，利用聚类算法对上述特征矢量进行聚类，得到 K 个高斯的初始化均值，K

是预先设置的混合高斯模型个数。

[0102] 比如，可以采用传统的 LBG(Linde，Buzo，Gray) 聚类算法，通过训练矢量集和一定的迭代算法来逼近最优的再生码本。

[0103] 步骤 203，利用 EM(Expectation Maximization) 算法迭代更新上述均值、方差及各高斯对应的加权系数，得到与文本无关的通用背景模型。

[0104] 具体的迭代更新过程与现有技术相同，在此不再详细描述。

[0105] 当然，还可以采用其他方式构建上述与文本无关的通用背景模型，对此本发明实施例不做限定。

[0106] 在本发明实施例中，不论用户是处于登录模式还是注册模式，都可以将用户录入的语音信号或者从该语音信号中提取的声纹特征写入对该语音信号识别出的密码文本对应的缓存区中，并且根据缓存区中的数据实时构建或更新与相应的密码文本相关的优化背景模型。这样，可以快速收集针对特定密码文本的相关数据，从而使所述优化背景模型得到快速优化，提高声纹识别的效率及准确性。

[0107] 当然，在实际应用中，为了减少系统的运算量，也可以只在注册模式或登录模式下构建或更新与相应的密码文本相关的优化背景模型。对此本发明实施例不做限定。

[0108] 因此，在上述图 1 所示流程中，还可进一步包括以下步骤：将登录用户录入的语音信号或者从登录用户录入的语音信号中提取的声纹特征序列写入与所述密码文本对应的缓存区。在注册状态，接收注册用户录入的注册语音信号；对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本；将所述注册语音信号或者从所述注册语音信号中提取的声纹特征序列写入与该注册语音信号相应的密码文本对应的缓存区。另外，需要根据所述注册用户录入的注册语音信号训练对应所述注册用户的说话人声纹模型，还需要实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型。

[0109] 在本发明实施例中，可以为每个密码文本设立一个对应的缓存区，不同密码文本对应不同的缓存区，在该缓存区中存储对应同一密码文本的语音信号或者从所述语音信号中提取的声纹特征序列，上述语音信号不仅包括登录用户录入的语音信号，也包括注册用户录入的注册语音信号，当然，在一个缓存区中存储的来自不同用户的语音信号都对应了同一个密码文本。

[0110] 在实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型时，可以在每次有新数据加入所述缓存区后，即对当前与所述密码文本相关的优化背景模型进行更新。当然，为了减少系统开销及运算工作量，还可以在对应一个密码文本的缓存区中存储的数据满足一定预定条件时，根据所述缓存区中的数据构建或更新相应的优化背景模型。在具体应用时，上述预设条件及相应的构建或更新优化背景模型的方式可以有多种，比如：

[0111] 一种方式是：如果一个缓存区中存储的数据量达到第一预设值（比如 500、或 600 等），并且当前没有与该缓存区对应密码文本相关的优化背景模型，则以所述通用背景模型为初始模型，根据该缓存区中的数据生成与该缓存区对应密码文本相关的优化背景模型，并删除该缓存区中存储的数据；如果一个缓存区中存储的数据量达到第一预设值，并且当前有与该缓存区对应密码文本相关的优化背景模型，则以该优化背景模型为初始模型，根据该缓存区中的数据更新该优化背景模型，并删除该缓存区中存储的数据。

[0112] 在这种方式下，每次构建或更新优化背景模型时依据的数据量相同，而且在构建优化背景模型时，采用的初始模型是前面提到的通用背景模型，在更新优化背景模型时，采用的初始模型是当前的优化背景模型。另外，在这种方式下，无论是构建优化模型还是更新当前优化背景模型，之后都需要清除相应缓存区中的数据，以便采集下一组数据。这种方式可以降低对缓存区存储空间的需求。

[0113] 另一种方式是：如果一个缓存区中存储的数据量达到第二预设值（比如 500、或 600 等）的整数倍，则以所述通用背景模型为初始模型，根据该缓存区中的数据重新生成与该缓存区对应密码文本相关的优化背景模型。

[0114] 在这种方式下，每次构建或更新优化背景模型时依据的数据量不同，而且在构建和更新当前优化背景模型时，采用的初始模型均是前面提到的通用背景模型。另外，在这种方式下，无需在每次构建或更新当前优化背景模型后都要清除相应缓存区中的数据，但对缓存空间的需求较大，可以应用在具有海量缓存空间的环境下。当然，也可以采用与上述第一种类似的处理方式，在缓存区中的数据量达到一定程度（比如 50000）时，清除该缓存区中的数据，为了保证优化背景模型的特性，在该缓存区中的数据量重新达到上述第二预设值时，不是以通用背景模型为初始模型进行更新过程，而是以当前的优化背景模型为初始模型进行更新过程，然后在后续缓存区中的数据量再次达到进行更新的条件时，再继续以通用背景模型为初始模型进行更新过程。

[0115] 如图 3 所示，是本发明实施例中构建或更新优化背景模型的一种流程图，包括以下步骤：

[0116] 步骤 301，利用缓存区内所有声纹特征序列自适应更新通用背景模型混合高斯的均值 μ_m。

[0117] 具体地，新高斯均值$\hat{\mu}_m$计算为样本统计量和原始高斯均值的加权平均，即：

[0118] $$\hat{\mu}_m = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\gamma_m(x_t)x_t + \tau\mu_m}{\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_m(x_t) + \tau} \qquad (4)$$

[0119] 其中，N 是声纹特征序列总数，T_i 是第 i 句声纹特征序列的总帧长，x_t 表示第 t 帧声纹特征，$\gamma_m(x_t)$ 表示第 t 帧声纹特征落于第 m 个高斯的概率，τ 是遗忘因子，用于平衡历史均值以及样本对新均值的更新力度。一般来说，τ 值越大，则新均值主要受原始均值制约。而若 τ 值较小，则新均值主要由样本统计量决定，更多的体现了新样本分布的特点。τ 值可以由系统预先确定，也可以选择随时间逐渐变化的参数值，以不断提升新样本数据的作用

[0120] 步骤 302，复制通用背景模型方差作为与密码文本相关的优化背景模型方差。

[0121] 步骤 303，生成与密码文本相关的优化背景模型。

[0122] 根据缓存区中的数据更新与注册密码文本相关的优化背景模型的过程与上述类似，在此不再赘述。

[0123] 需要说明的是，在本发明实施例中，注册用户录入的注册语音信号可以是录入一次，也可以是重复录入多次，以保证注册密码的准确性。

[0124] 如果是重复录入多次，相应地，在通过语音识别确定所述注册用户的注册密码文

本时，可以分别对每次录入的注册语音信号进行语音识别，得到多个识别结果及与各识别结果对应的识别似然度得分；然后选择具有最高似然度得分的识别结果作为所述注册用户的注册密码文本。

[0125] 下面结合语音识别的具体过程对此进行简单说明。

[0126] 假定系统可以支持用户任意定义密码内容，如图 4 所示，是本发明实施例中对注册用户录入的注册语音信号进行语音识别的流程图，包括以下步骤：

[0127] 步骤 401，获取当前需要识别的语音信号。

[0128] 步骤 402，从所述语音信号中提取声学特征序列。

[0129] 步骤 403，在大词汇量连续语音识别的搜索网络中搜索对应于步骤 302 的最优路径，并记录其路径历史累计概率（即上述似然度得分），具体过程与现有技术类似，在此不再详细描述。

[0130] 考虑到汉语字符过多，对每个字符建模容易导致内存过大，因而可以选择对更小的语音单元，如 400 余个音节或 1300 余个带调的音节等，并据此构建搜索网络。

[0131] 需要说明的是，在本发明实施例中，还可以预先设定密码文本选择范围，如常用成语，常用口令等供用户挑选使用。在这种情况下，本发明实施例中对注册用户录入的注册语音信号进行语音识别可以按照命令词识别方式（即按照密码文本构建上述搜索网络）进行，以提高解码效率。

[0132] 当然，在实际应用中，还可以由用户选择或者自定义密码文本。

[0133] 需要说明的是，如果注册用户在注册时多次录入的注册语音信号，也可以将每次录入的注册语音信号或者从每次录入的注册语音信号中提取的声纹特征序列写入该语音信号对应的密码文本相应的存储区，以增加相应密码文本的用户数据，为细化与该密码文本相关的背景模型提供足够的数据。

[0134] 本发明实施例提供的声纹密码认证方法，在进行用户身份识别时，不仅对用户登录时录入的语音信号进行语音识别，确定其密码内容，而且对其进行声纹认证，在进行声纹认证时，基于多背景模型，即与文本无关的通用背景模型及与文本相关的优化背景模型，通过选择合适的背景模型实现精确匹配，有效地提高了基于声纹密码进行身份认证的准确率。

[0135] 在本发明实施例中，利用用户注册及登录数据训练优化背景模型，使得系统从初始单一的通用背景模型，不断细化得到对应于不同密码文本的多背景模型，从而为用户不同密码提供了具有较强针对性的背景模型，提高了模型之间的区分性，进而提高了语音识别的准确率和识别效率。

[0136] 相应地，本发明实施例还提供一种声纹密码认证系统，如图 5 所示，是该系统的一种结构示意图。

[0137] 在该实施例中，所述声纹密码认证系统包括：

[0138] 接收单元 501，用于在用户登录时，接收登录用户录入的语音信号；

[0139] 声纹特征提取单元 502，用于提取所述语音信号中的声纹特征序列；

[0140] 所述声纹特征序列包含一组声纹特征，可以有效地区分不同的说话人，且对同一说话人的变化保持相对稳定。比如，所述声纹特征主要有：谱包络参数语音特征，基音轮廓、共振峰频率带宽特征，线性预测系数，倒谱系数等；考虑到上述声纹特征的可量化

性、训练样本的数量和系统性能的评价等问题，可以选用 MFCC(Mel Frequency Cepstrum Coefficient，Mel 频率倒谱系数）特征，对窗长 25ms 帧移 10ms 的每帧语音数据做短时分析得到 MFCC 参数及其一阶二阶差分，共计 39 维。这样，每句语音信号可以量化为一个 39 维声纹特征序列 X ；

[0141] 语音识别单元 503，用于对所述语音信号进行语音识别，获得所述登录用户的密码文本，具体的语音识别的处理方式可以采用一些现有的方式，在此不再详细说明 ；

[0142] 判断单元 504，用于判断语音识别单元 503 获得的密码文本与对应所述登录用户的注册密码是否相同 ；

[0143] 认证结果单元 505，用于在判断单元 504 的判断结果是所述语音识别单元 503 获得的密码文本与所述登录用户的注册密码文本不同时，确定所述登录用户为非认证用户 ；

[0144] 模型确定单元 506，用于在所述判断单元 504 的判断结果是所述语音识别单元 503 获得的密码文本与所述登录用户的注册密码文本相同时，确定对应所述登录用户的背景模型，所述背景模型包括 ：与文本无关的通用背景模型、以及与文本相关的优化背景模型，而且，在实际应用中，模型确定单元 506 可以根据需要，有多种不同的方式来确定对应所述登录用户的背景模型，具体可参考前面的描述 ；

[0145] 第一计算单元 507，用于分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述背景模型的似然度 ；

[0146] 第二计算单元 508，用于根据所述声纹特征序列与说话人声纹模型的似然度、以及所述声纹特征序列与背景模型的似然度，计算似然比 ；

[0147] 上述第一计算单元 507 和第二计算单元 508 的具体计算过程可参照前面本发明声纹密码认证方法实施例中的描述，在此不再详细说明。

[0148] 在该实施例中，上述判断单元 504 还用于判断所述第二计算单元 508 计算得到的似然比是否大于设定的阈值 ；相应地，上述认证结果单元 505 还用于在判断单元 504 的判断结果是第二计算单元 508 计算得到的似然比大于设定的阈值时，确定所述登录用户为有效认证用户，否则确定所述登录用户为非认证用户。

[0149] 上述阈值可以由系统预先设定，一般来说，该阈值越大，则系统的灵敏度越高，要求用户在登录时尽可能按照注册时录入的语音信号（即密码）的发音，反之，则系统的灵敏度较低，允许用户登录时录入的语音信号的发音与注册时的发音存在一定的变化。

[0150] 如图 6 所示，是本发明实施例声纹密码认证系统的另一种结构示意图。

[0151] 与图 5 所示实施例不同的是，在该实施例中，所述系统还包括 ：

[0152] 检查单元 601，用于检查是否存在与所述登录用户的注册密码文本对应的优化背景模型。

[0153] 相应地，模型确定单元 506 可以在所述检查单元 601 的检查结果是有与所述登录用户的注册密码文本对应的优化背景模型时，选择该优化背景模型作为对应所述登录用户的背景模型 ；否则选择所述通用背景模型作为对应所述登录用户的背景模型。

[0154] 当然，本发明实施例声纹密码认证系统中，模型确定单元 506 还可以根据需要，有多种不同的方式来选择对应所述登录用户的背景模型，比如，可以在系统初始化阶段（比如，可以设定一定的时间段），选择与文本无关的通用背景模型，以适应用户录入的各种不同声纹密码 ；而随着系统的运行，采集到的特定密码相关的用户数据不断增加，可以根据这

些用户数据训练得到与文本相关的优化背景模型，该优化背景模型是与用户密码文本相关的模型，此后，可以根据当前登录用户的密码文本来选择相应的背景模型。

[0155] 如图 7 所示，是本发明实施例声纹密码认证系统的另一种结构示意图。

[0156] 与图 6 所示实施例不同的是，在该实施例中，所述系统还包括：背景模型构建单元 701 和说话人声纹模型构建单元 702。

[0157] 另外，在该实施例中，语音识别单元 503 还用于将登录用户录入的语音信号或者从登录用户录入的语音信号中提取的声纹特征序列写入与所述密码文本对应的缓存区。

[0158] 接收单元 501 还用于接收注册用户录入的注册语音信号，相应地，语音识别单元 503 还用于对所述注册语音信号进行语音识别，得到所述注册用户的注册密码文本。

[0159] 背景模型构建单元 701 用于实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型。

[0160] 说话人声纹模型构建单元 702，用于根据所述注册用户录入的注册语音信号训练对应所述注册用户的说话人声纹模型。

[0161] 当然，在实际应用中，也可以由声纹特征提取单元 502 根据语音识别单元 503 识别出的语音信号（包括登录用户录入的语音信号和注册用户录入的注册语音信号）对应的密码文本，将所述语音信号写入与该密码文本对应的缓存区，对此本发明实施例不做限定。

[0162] 在本发明实施例的系统中，可以为每个密码文本设立一个对应的缓存区，不同密码文本对应不同的缓存区，在该缓存区中存储对应同一密码文本的语音信号或者从所述语音信号中提取的声纹特征序列，上述语音信号不仅包括登录用户录入的语音信号，也包括注册用户录入的注册语音信号，当然，在一个缓存区中存储的来自不同用户的语音信号都对应了同一个密码文本。

[0163] 背景模型构建单元 701 实时根据每个缓存区中的数据构建或更新与所述缓存区对应密码文本相关的优化背景模型，可以是在每次有新数据加入所述缓存区后，即对当前与所述密码文本相关的优化背景模型进行更新。当然，为了减少系统开销及运算工作量，还可以是在对应一个密码文本的缓存区中存储的数据满足一定预定条件后，根据所述缓存区中的数据构建或更新相应的优化背景模型。在具体应用时，上述预设条件及相应的构建或更新优化背景模型的方式可以有多种，比如：在一种实施例中，背景模型构建单元 701 可以在一个缓存区中存储的数据量达到第一预设值，并且当前没有与该缓存区对应密码文本相关的优化背景模型时，以所述通用背景模型为初始模型，根据该缓存区中的数据生成与该缓存区对应密码文本相关的优化背景模型，并删除该缓存区中存储的数据；在一个缓存区中存储的数据量达到第一预设值，并且当前有与该缓存区对应密码文本相关的优化背景模型时，以该优化背景模型为初始模型，根据该缓存区中的数据更新该优化背景模型，并删除该缓存区中存储的数据。

[0164] 在另一种实施例中，背景模型构建单元 701 可以在一个缓存区中存储的数据量达到第二预设值的整数倍，则以所述通用背景模型为初始模型，根据该缓存区中的数据重新生成与该缓存区对应密码文本相关的优化背景模型。

[0165] 上述两个实施例中背景模型构建单元 701 构建或更新与密码文本相关的优化背景模型的具体过程可参见前面本发明方法实施例中的描述，在此不再赘述。

[0166] 需要说明的是，在具体应用时，所述注册用户录入的注册语音信号可以是录入一

次，也可以是重复录入多次，如果是重复录入多次，相应地，所述语音识别单元 503 可以分别对每次录入的注册语音信号进行语音识别，得到多个识别结果及与各识别结果对应的识别似然度得分。

[0167] 相应地，所述系统还可进一步包括：密码确定单元（未图示），用于从所述语音识别单元 503 得到的多个识别结果中选择具有最高似然度得分的识别结果作为所述注册用户的注册密码文本。具体过程可参照前面的描述，在此不再赘述。

[0168] 本发明实施例提供的声纹密码认证系统，在进行用户身份识别时，不仅对用户登录时录入的语音信号进行语音识别，确定其密码内容，而且对其进行声纹认证，在进行声纹认证时，基于多背景模型，即与文本无关的通用背景模型及与文本相关的优化背景模型，通过选择合适的背景模型实现精确匹配，有效地提高了基于声纹密码进行身份认证的准确率。

[0169] 在本发明实施例中，利用用户注册及登录数据训练优化背景模型，使得系统从初始单一的通用背景模型，不断细化得到对应于不同密码文本的多背景模型，从而为用户不同密码提供了具有较强针对性的背景模型，提高了模型之间的区分性，进而提高了语音识别的准确率和识别效率。

[0170] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的系统实施例仅仅是示意性的，其中所述作为分离部件说明的单元及模块可以是或者也可以不是物理上分开的。另外，还可以根据实际的需要选择其中的部分或者全部单元和模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

[0171] 以上公开的仅为本发明的优选实施方式，但本发明并非局限于此，任何本领域的技术人员能思之的没有创造性的变化，以及在不脱离本发明原理前提下所作的若干改进和润饰，都应落在本发明的保护范围内。

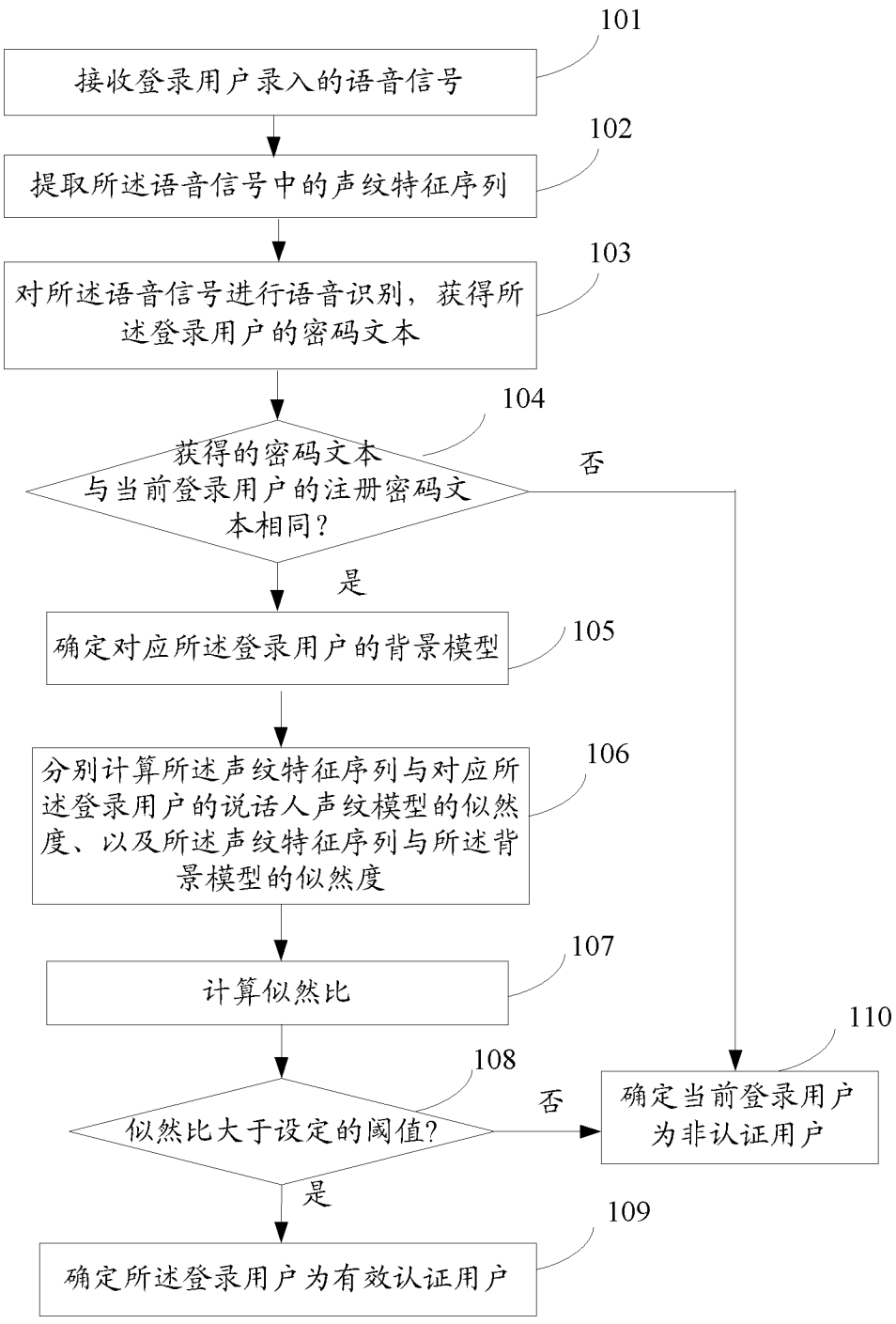

101
接收登录用户录入的语音信号
102
提取所述语音信号中的声纹特征序列
103
对所述语音信号进行语音识别，获得所述登录用户的密码文本
104
获得的密码文本与当前登录用户的注册密码文本相同？
否
是
105
确定对应所述登录用户的背景模型
106
分别计算所述声纹特征序列与对应所述登录用户的说话人声纹模型的似然度、以及所述声纹特征序列与所述背景模型的似然度
107
计算似然比
108
似然比大于设定的阈值？
否
是
110
确定当前登录用户为非认证用户
109
确定所述登录用户为有效认证用户

图 1

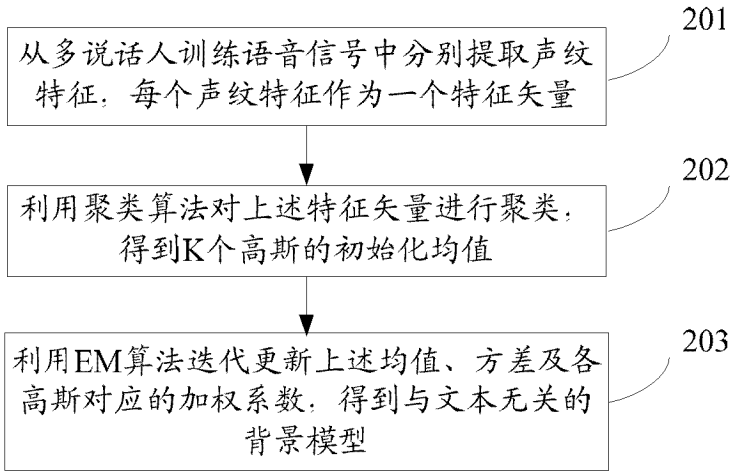

201
从多说话人训练语音信号中分别提取声纹特征，每个声纹特征作为一个特征矢量
202
利用聚类算法对上述特征矢量进行聚类，得到K个高斯的初始化均值
203
利用EM算法迭代更新上述均值、方差及各高斯对应的加权系数，得到与文本无关的背景模型

图 2

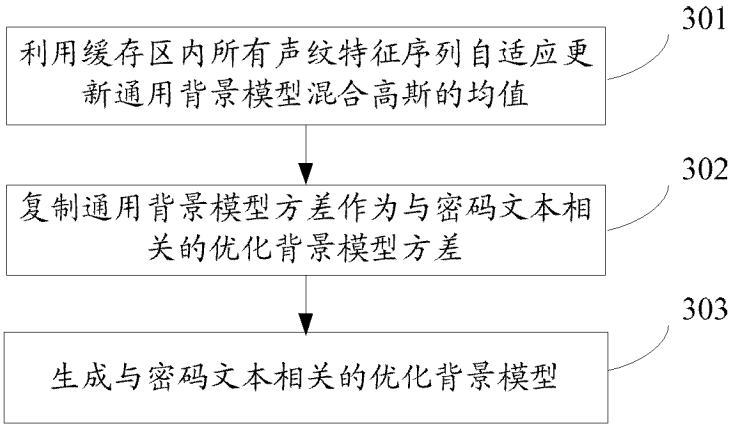

301
利用缓存区内所有声纹特征序列自适应更新通用背景模型混合高斯的均值
302
复制通用背景模型方差作为与密码文本相关的优化背景模型方差
303
生成与密码文本相关的优化背景模型

图 3

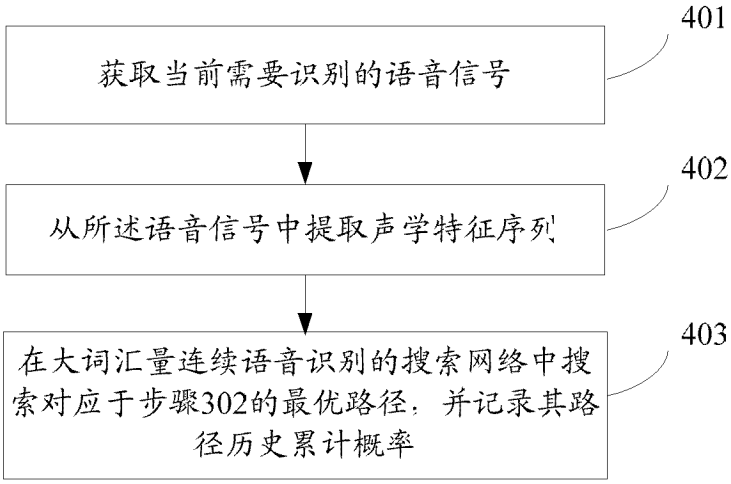

401
获取当前需要识别的语音信号
402
从所述语音信号中提取声学特征序列
403
在大词汇量连续语音识别的搜索网络中搜索对应于步骤302的最优路径，并记录其路径历史累计概率

图 4

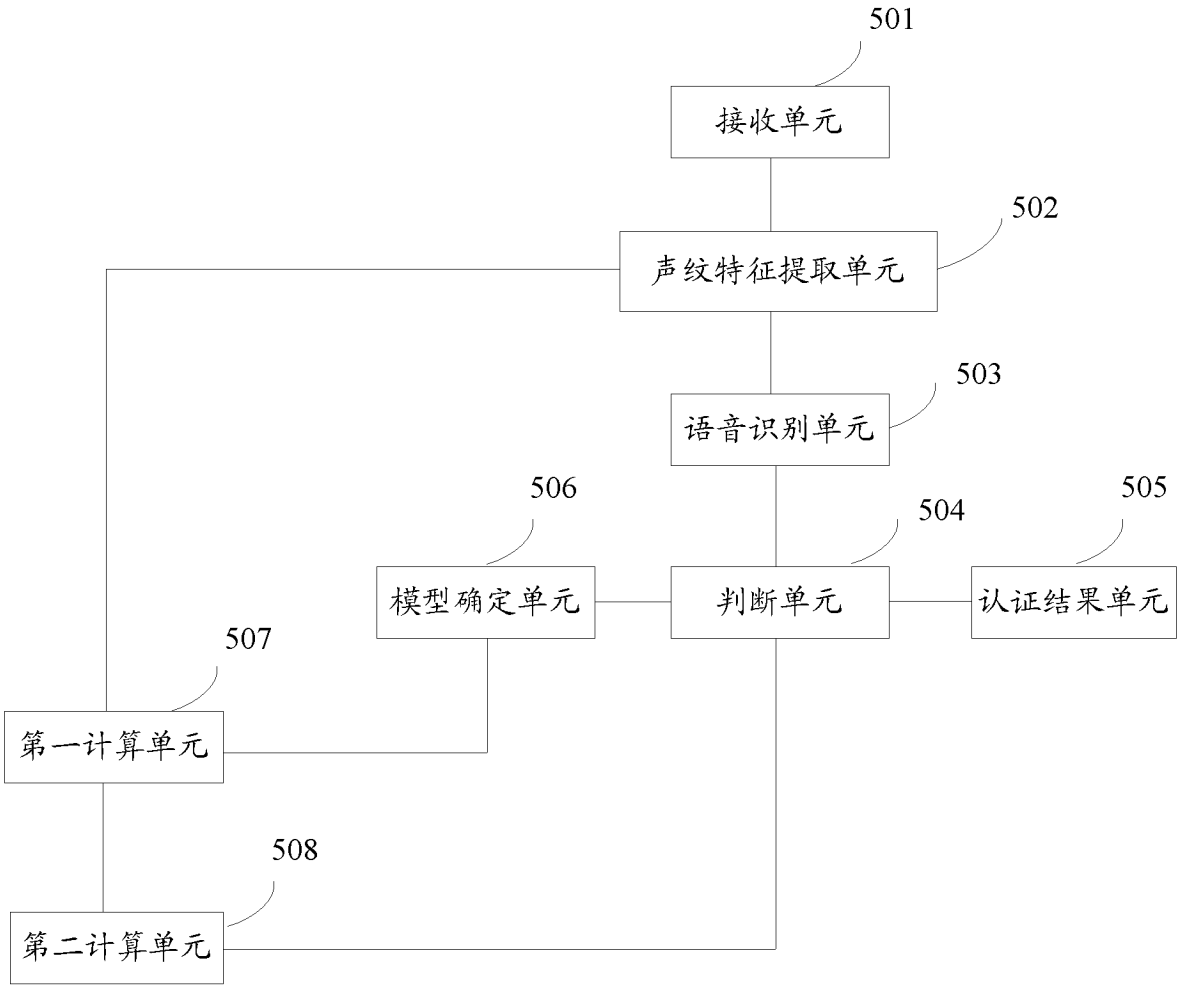

501
接收单元
502
声纹特征提取单元
503
语音识别单元
506
504
505
模型确定单元
判断单元
认证结果单元
507
第一计算单元
508
第二计算单元

图 5

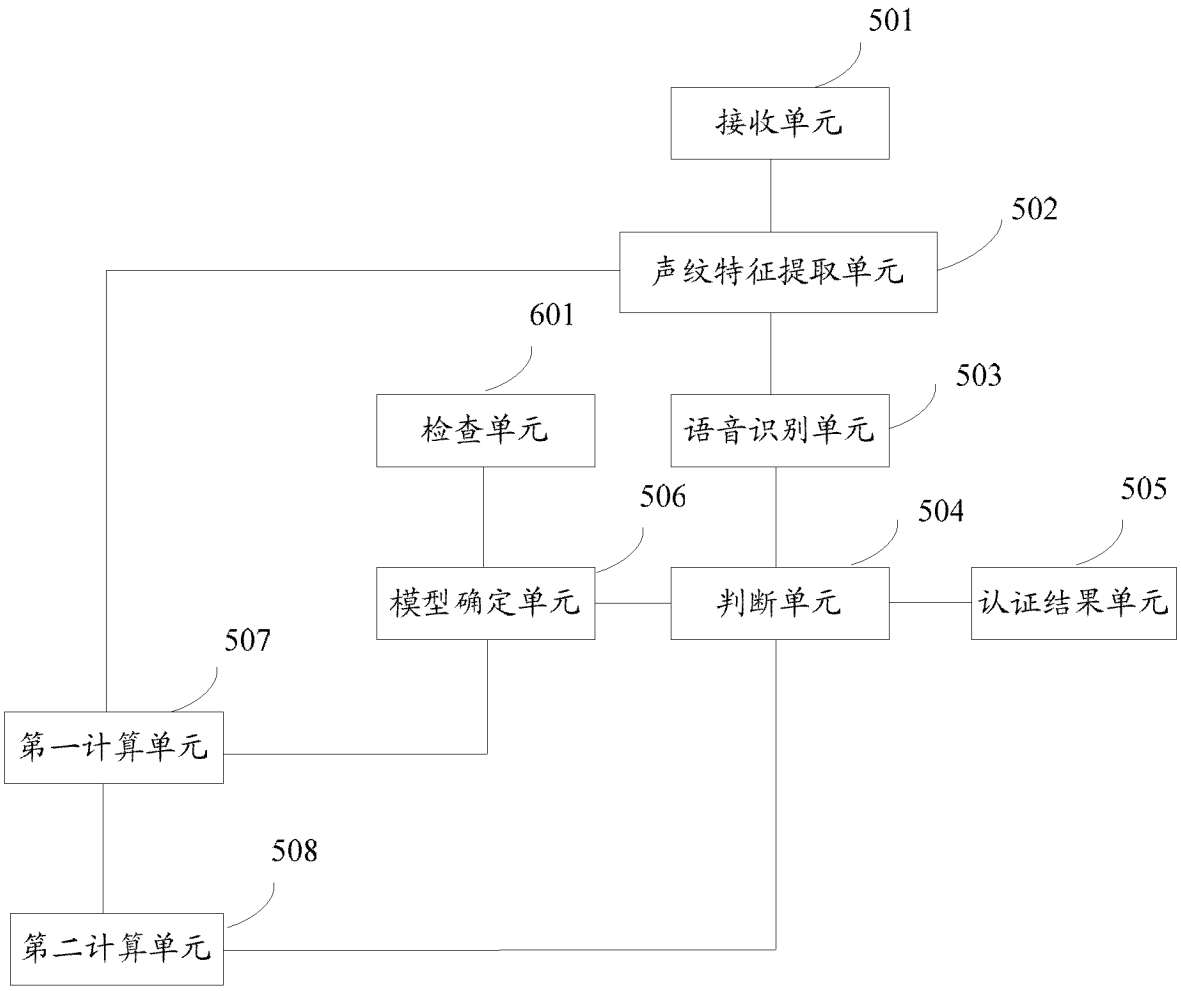

501
接收单元
502
声纹特征提取单元
601
检查单元
503
语音识别单元
506
504
505
模型确定单元
判断单元
认证结果单元
507
第一计算单元
508
第二计算单元

图 6

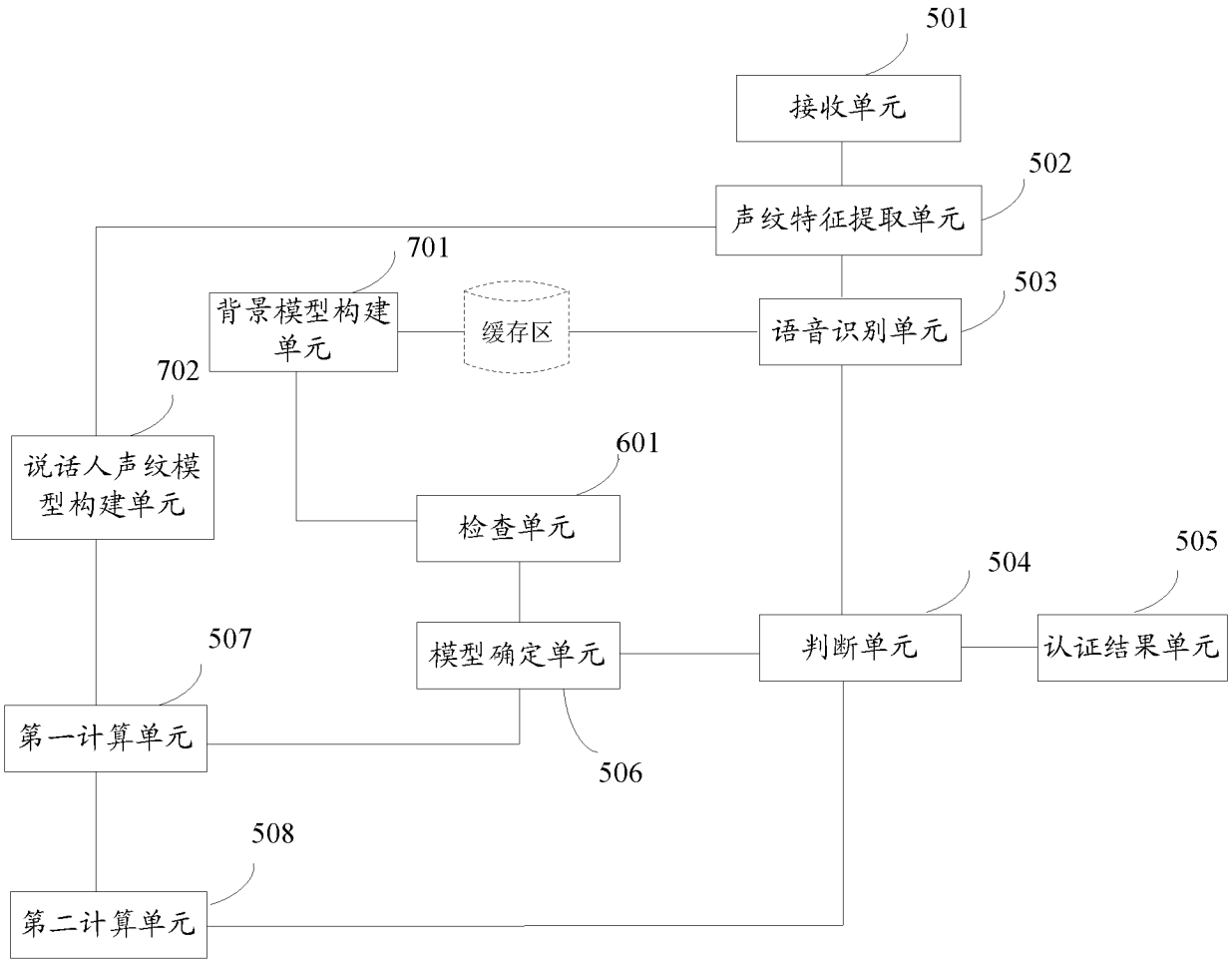

501
接收单元
502
声纹特征提取单元
701
背景模型构建单元
缓存区
503
语音识别单元
702
说话人声纹模型构建单元
601
检查单元
504
505
507
模型确定单元
判断单元
认证结果单元
第一计算单元
506
508
第二计算单元

图 7