(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108288470 A
(43)申请公布日 2018.07.17

(21)申请号 201710019531.9

(22)申请日 2017.01.10

(71)申请人 富士通株式会社
地址 日本神奈川县

(72)发明人 石自强 刘柳 刘汝杰

(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 康建峰 江河清

(51)Int.Cl.
G10L 17/04(2013.01)
G10L 17/18(2013.01)

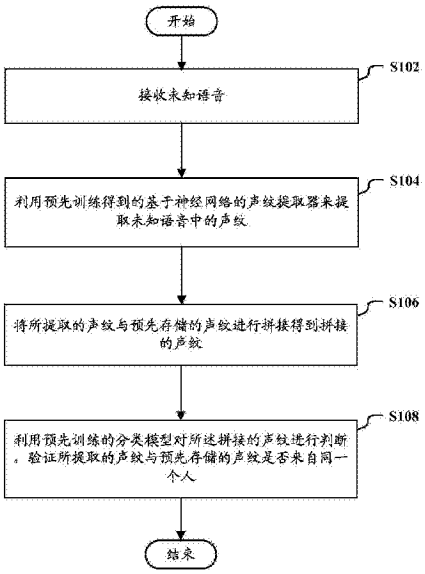
权利要求书2页 说明书9页 附图4页

(54)发明名称

基于声纹的身份验证方法和装置

(57)摘要

本发明涉及基于声纹的身份验证方法和装置。该方法包括:一种基于声纹的身份验证方法,包括:接收未知语音;利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹;将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹;以及利用预先训练的分类模型对所述拼接的声纹进行判断,验证所提取的声纹与预先存储的声纹是否来自同一个人。根据本发明的身份验证方法和装置,可以从较短的语音中提取说话者的全息声纹,使得验证结果更加鲁棒。



CN 108288470 A

1.一种基于声纹的身份验证方法,包括:

接收未知语音;

利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹;

将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹;以及

利用预先训练的分类模型对所述拼接的声纹进行判断,验证所提取的声纹与预先存储的声纹是否来自同一个人。

2.根据权利要求1所述的身份验证方法,其中,通过下述步骤来训练得到基于神经网络的声纹提取器:

采集来自不同说话者的语音;

将说话者的辅助信息作为神经网络的分组标签进行标记,将标记过的语音作为训练样本输入神经网络;以及

进行深度学习来得到所述基于神经网络的声纹提取器。

3.根据权利要求2所述的身份验证方法,其中,所述说话者的辅助信息包括:说话者的语种、音素序列、发声通道、情感、年龄、生活区域以及性别中的一项或多项。

4.根据权利要求2所述的身份验证方法,其中,进行深度学习来得到所述基于神经网络的声纹提取器包括:

采用层次神经网络将不同的分组标签分别放置在不同层进行深度学习得到层次网络提取器;以及

采用扁平神经网络将全部分组标签放置在输出层进行深度学习得到扁平网络提取器。

5.根据权利要求4所述的身份验证方法,其中,利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹的步骤包括:

分别利用所述层次网络提取器和所述扁平网络提取器提取来提取所述未知语音中的声纹;以及

将所述层次网络提取器所提取的声纹和所述扁平网络提取器所提取的声纹拼接在一起作为所提取的所述未知语音中的声纹。

6.根据权利要求4所述的身份验证方法,其中,所述分组标签包含的信息越简单,则该分组标签在层次神经网络中放置的位置越靠前。

7.根据权利要求2所述的身份验证方法,其中,利用随机梯度下降方法进行所述深度学习,所述随机梯度方法使用以下公式来计算神经网络的权重:

$w_{t+1}=\theta_t+(a_t-1)(\theta_t-\theta_{t-1})/a_{t+1}+a_t(\theta_t-w_t)/a_{t+1}$

其中,w_t是权重,$\theta_t = w_t - \epsilon_t \nabla f(w_t)$,$\in_t$是学习率,$\nabla f(w_t)$是损失函数f的梯度,$a_0=1$,$a_{t+1} = (1+\sqrt{1+a_t^2})/2$。

8.根据权利要求1所述的身份验证方法,其中,利用预先训练的分类模型对所述拼接的声纹进行判断,验证所提取的声纹与预先存储的声纹是否来自同一个人的步骤包括:

训练第一高斯混合模型和第二高斯混合模型,其中,所述第一高斯混合模型对应两个声纹属于同一个人的情况,所述第二高斯混合模型对应两个声纹不属于同一个人的情况;

计算所述拼接的声纹分别在所述第一高斯混合模型上的第一概率和所述第二高斯混

合模型上的第二概率;以及

当所述第一概率大于所述第二概率,则所提取的声纹与预先存储的声纹是来自同一个人,否则,则是来自不同的人。

9.根据权利要求2所述的身份验证方法,其中,所提取的所述未知语音中的声纹包含所述说话者的辅助信息。

10.一种基于声纹的身份验证装置,包括:

语音接收单元,被配置为接收未知语音;

声纹提取单元,被配置为利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹;

声纹拼接单元,被配置为将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹;以及

身份验证单元,被配置为利用预先训练的分类模型对所述拼接的声纹进行判断,验证所提取的声纹与预先存储的声纹是否来自同一个人。

基于声纹的身份验证方法和装置

技术领域

[0001] 本发明涉及语音处理领域，更具体地涉及一种基于声纹的身份验证方法和装置。

背景技术

[0002] 声纹已在许多领域有着广泛的应用，包括用户接口，国土安全，电话银行等。传统的方法通过联合因子分析(joint factor analysis)将语音片段映射到某空间，得到一种i-vector作为声纹。但是这种方法有两种缺陷：1，为了得到满意的性能，必须采用较长(20-30秒)的语音段来提取i-vector；2，说话者的其他信息，例如年龄、性别、语种信息是有助于身份确认的，但是目前这种框架没有办法或者很难加入说话者的其他信息。

[0003] 因此，希望提供一种能够基于较短的、包含说话者的多种信息的声纹来进行身份验证的方法和装置。

发明内容

[0004] 在下文中给出关于本发明的简要概述，以便提供关于本发明的某些方面的基本理解。应当理解，这个概述并不是关于本发明的穷举性概述。它并不是意图确定本发明的关键或重要部分，也不是意图限定本发明的范围。其目的仅仅是以简化的形式给出某些概念，以此作为稍后论述的更详细描述的前序。

[0005] 本发明的一个主要目的在于，提供了一种基于声纹的身份验证方法，包括：接收未知语音；利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹；将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹；以及利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人。

[0006] 根据本发明的一个方面，提供一种基于声纹的身份验证装置，包括：语音接收单元，被配置为接收未知语音；声纹提取单元，被配置为利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹；声纹拼接单元，被配置为将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹；以及身份验证单元，被配置为利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人。

[0007] 另外，本发明的实施例还提供了用于实现上述方法的计算机程序。

[0008] 此外，本发明的实施例还提供了至少计算机可读介质形式的计算机程序产品，其上记录有用于实现上述方法的计算机程序代码。

[0009] 通过以下结合附图对本发明的最佳实施例的详细说明，本发明的这些以及其他优点将更加明显。

附图说明

[0010] 参照下面结合附图对本发明实施例的说明，会更加容易地理解本发明的以上和其它目的、特点和优点。附图中的部件只是为了示出本发明的原理。在附图中，相同的或类似

的技术特征或部件将采用相同或类似的附图标记来表示。

[0011] 图1示出了根据本发明的一个实施例的基于声纹的身份验证方法100的示例性过程的流程图；

[0012] 图2示出了训练得到基于神经网络的声纹提取器的示例性过程的流程图；

[0013] 图3A示出了用于提取声纹的层次神经网络；

[0014] 图3B示出了用于提取声纹的扁平神经网络；

[0015] 图4是示出根据本发明的另一个实施例的基于声纹的身份验证装置400的示例性配置的框图；

[0016] 图5是示出用于训练声纹提取器的声纹提取器训练装置500的一种示例性配置的框图；以及

[0017] 图6是示出可以用于实施本发明的基于声纹的身份验证方法和装置的计算设备的示例性结构图。

具体实施方式

[0018] 在下文中将结合附图对本发明的示范性实施例进行描述。为了清楚和简明起见，在说明书中并未描述实际实施方式的所有特征。然而，应该了解，在开发任何这种实际实施例的过程中必须做出很多特定于实施方式的决定，以便实现开发人员的具体目标，例如，符合与系统及业务相关的那些限制条件，并且这些限制条件可能会随着实施方式的不同而有所改变。此外，还应该了解，虽然开发工作有可能是非常复杂和费时的，但对得益于本公开内容的本领域技术人员来说，这种开发工作仅仅是例行的任务。

[0019] 在此，还需要说明的一点是，为了避免因不必要的细节而模糊了本发明，在附图中仅仅示出了与根据本发明的方案密切相关的设备结构和/或处理步骤，而省略了与本发明关系不大的其他细节。

[0020] 本发明提出了一种基于声纹的身份验证方法，该方法不仅可以从较短(例如3-5秒)的语音中提取说话者的身份信息，还可以同时提取其他的说话者信息，包括发声通道、音素序列、性别、年龄、语种、生活区域、情感等，这种信息被称为全息声纹。

[0021] 下面结合附图详细说明根据本发明的实施例的基于声纹的身份验证方法和装置。下文中的描述按如下顺序进行：

[0022] 1.基于声纹的身份验证方法

[0023] 2.基于声纹的身份验证装置

[0024] 3.用以实施本申请的方法和装置的计算设备

[0025] [1.基于声纹的身份验证方法]

[0026] 图1示出了根据本发明的一个实施例的基于声纹的身份验证方法100的示例性过程的流程图。

[0027] 首先，在步骤S102中，接收未知语音。

[0028] 接着，在步骤S104中，利用预先训练得到的基于神经网络的声纹提取器来提取未知语音中的声纹。

[0029] 图2示出了训练得到基于神经网络的声纹提取器的示例性过程的流程图。

[0030] 如图2所示，在步骤S1042中，采集来自不同说话者的语音。

[0031] 接着，在步骤S1044中，将说话者的辅助信息作为神经网络的分组标签进行标记，将标记过的语音作为训练样本输入神经网络。

[0032] 具体地，需要收集不同说话者的大量语音，并对这些语音进行标记。可以包括说话者身份(ID)、语种、音素序列、发声通道、情感、年龄、生活区域、性别等信息。在本发明中，使得提取的声纹包含以上这些信息，从而根据本发明的身份验证方法更加鲁棒。用这些标记过的语音作为训练样本输入神经网络。

[0033] 通常采用下述方法对收集的语音数据进行处理。将标记的语音数据分割成帧长25毫秒、帧移10毫秒的信号，提取13维的梅尔频率倒谱系数(MFCCs)，以及该系数的一阶差分和二阶差分连接起来共39维做为特征。同时联合上下文共39帧(左25帧，右13帧)作为最终的特征共1521维(39*39)。

[0034] 本领域技术人员可以理解，对语音数据的处理也可以采用本领域公知的其它方法，在此不做赘述。

[0035] 其中，说话者的辅助信息包括：说话者的语种、音素序列、发声通道、情感、年龄、生活区域以及性别中的一项或多项。

[0036] 最后，在步骤S1046中，进行深度学习来得到所述基于神经网络的声纹提取器。

[0037] 在本发明中，分别采用了两种神经网络、即层次神经网络和扁平神经网络进行深度学习。图3A示出了用于提取声纹的层次神经网络，图3B示出了用于提取声纹的扁平神经网络。这两种神经网络的区别在于在何处放置分组标签。如图3A所示，层次神经网络是将不同的分组标签分别放置在神经网络的不同层进行深度学习，可以得到用于提取声纹的层次网络提取器；如图3B所示，扁平神经网络是将全部分组标签放置在神经网络的输出层进行深度学习，可以得到用于提取声纹的扁平网络提取器。

[0038] 两种神经网络的输入为在步骤S2042中得到的39*39的语音特征，将输出神经元分成若干个块，每一块关联说话者的一种信息，例如第一块和说话者ID相关，第二块和年龄相关等。每一块神经元的数目为说话者的数目、年龄的跨度、性别的种类、语种的数目、文本的类别、信道的种类(移动电话、固定电话)、情感的类别等等。输出的标签是独热(one-hot)向量，其中的非零元素对应说话者的身份、性别、年龄等。

[0039] 步骤S104(利用预先训练得到的基于神经网络的声纹提取器来提取未知语音中的声纹)具体地包括：分别利用所述层次网络提取器和所述扁平网络提取器提取来提取未知语音中的声纹，再将所述层次网络提取器所提取的声纹和所述扁平网络提取器所提取的声纹拼接在一起作为所提取的未知语音中的声纹。

[0040] 优选地，在层次神经网络中，分组标签包含的信息越简单，则该分组标签在神经网络中放置的位置越靠前。

[0041] 例如，与性别相关的神经元的数目为2，则将性别标签放置在层次神经网络较靠前的层；而音素序列和情感等包含的神经元的数目较多，则将其放置在层次神经网络较靠后的层。

[0042] 在一个示例中，可以利用随机梯度下降方法对上述两种神经网络进行深度学习。随机梯度下降方法可以使用以下公式来计算神经网络的权重：

[0043] $w_{t+1}=\theta_t+(a_t-1)(\theta_t-\theta_{t-1})/a_{t+1}+a_t(\theta_t-w_t)/a_{t+1}$。

[0044] 其中，wt是权重，$\theta_t = w_t - \epsilon_t \nabla f(w_t)$，$\epsilon_t'$是学习率，$\nabla f(w_t)$是损失函数f的梯度，a0＝1，$a_{t+1} = (1 + \sqrt{1 + a_t^2})/2$。

[0045] 利用该公式进行来计算神经网络的权重，可以使得随机梯度下降的收敛速度更快。

[0046] 接着，在步骤S106中，将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹。

[0047] 在需要身份验证的设备中，通常预先存储有该设备的合法用户的声纹。在进行身份验证时，将所提取的未知语音中的声纹与预先存储的声纹拼接在一起，进行下面的判断。

[0048] 最后，在步骤S108中，利用预先训练的分类模型对拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人。

[0049] 具体地，可以训练两种高斯混合模型分别对应两种情况。第一种高斯混合模型对应两个声纹属于同一个人的情况，第二种高斯混合模型对应两个声纹不属于同一个人的情况。

[0050] 在进行判断的时候，分别计算拼接的声纹在第一高斯混合模型上的第一概率和在第二高斯混合模型上的第二概率。当第一概率大于第二概率，则所提取的声纹与预先存储的声纹是来自同一个人，否则，则是来自不同的人。

[0051] 此外，还可以利用判别模型对所提取的声纹与预先存储的声纹是否来自同一个人进行判断。例如，采用支持向量机、神经网络等来构建表面，从而区分声纹是否来自同一个人。

[0052] 利用根据本发明的训练方法得到的基于神经网络的声纹提取器所提取的语音中的声纹包含说话者的辅助信息。

[0053] 本发明具有以下优势：

[0054] 1)相比现有技术，可以从较短(例如3-5秒)的语音中提取说话者的声纹；

[0055] 2)设计了两种提取声纹的神经网络(一种是层次神经网络，另一种是扁平神经网络)，可以提取说话者的全息声纹，使得根据本发明的身份验证方法更加鲁棒；

[0056] 3)采用了一种新的神经网络的训练算法，可以加快收敛速度；

[0057] 4)提出的两种声纹确认方法。

[0058] 2.基于声纹的身份验证装置

[0059] 图4是示出根据本发明的另一个实施例的基于声纹的身份验证装置400的示例性配置的框图。

[0060] 如图4所示，基于声纹的身份验证装置400包括语音接收单元402、声纹提取单元404、声纹拼接单元406和身份验证单元408。

[0061] 其中，语音接收单元402被配置为接收未知语音；声纹提取单元404被配置为利用声纹提取器训练装置训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹；声纹拼接单元406被配置为将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹；以及身份验证单元408被配置为利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人。

[0062] 图5是示出用于训练声纹提取器的声纹提取器训练装置500的一种示例性配置的框图。如图5所示，声纹提取器训练装置500包括：语音采集单元502、标记单元504和训练单元506。

[0063] 其中，语音采集单元502被配置为采集来自不同说话者的语音；标记单元504被配置为将说话者的辅助信息作为神经网络的分组标签进行标记，将标记过的语音作为训练样本输入神经网络；以及训练单元506被配置为进行深度学习来得到所述基于神经网络的声纹提取器。

[0064] 其中，所述说话者的辅助信息包括：说话者的语种、音素序列、发声通道、情感、年龄、生活区域以及性别中的一项或多项。

[0065] 其中，训练单元进一步被配置为：采用层次神经网络将不同的分组标签分别放置在不同层进行深度学习得到层次网络提取器；以及采用扁平神经网络将全部分组标签放置在输出层进行深度学习得到扁平网络提取器。

[0066] 其中，声纹提取单元进一步被配置为：分别利用所述层次网络提取器和所述扁平网络提取器提取来提取所述未知语音中的声纹；以及将所述层次网络提取器所提取的声纹和所述扁平网络提取器所提取的声纹拼接在一起作为所提取的所述未知语音中的声纹。

[0067] 其中，所述分组标签包含的信息越简单，则该分组标签在层次神经网络中放置的位置越靠前。

[0068] 深度学习可以利用随机梯度下降方法进行，所述随机梯度方法可以使用以下公式来计算神经网络的权重：

[0069] $w_{t+1}=\theta_t+(a_t-1)(\theta_t-\theta_{t-1})/a_{t+1}+a_t(\theta_t-w_t)/a_{t+1}$

[0070] 其中，w_t是权重，$\theta_t = w_t - \epsilon_t \nabla f(w_t)$，$\epsilon_t'$是学习率，$\nabla f(w_t)$是损失函数f的梯度，$a_0=1$，$a_{t+1} = (1+\sqrt{1+a_t^2})/2$。

[0071] 其中，身份验证单元进一步被配置为：训练第一高斯混合模型和第二高斯混合模型，其中，所述第一高斯混合模型对应两个声纹属于同一个人的情况，所述第二高斯混合模型对应两个声纹不属于同一个人的情况；计算所述拼接的声纹分别在所述第一高斯混合模型上的第一概率和所述第二高斯混合模型上的第二概率；以及当所述第一概率大于所述第二概率，则所提取的声纹与预先存储的声纹是来自同一个人，否则，则是来自不同的人。

[0072] 其中，所提取的未知语音中的声纹包含所述说话者的辅助信息。

[0073] 关于基于声纹的身份验证装置400和声纹提取器训练装置500的各个部分的操作和功能的细节可以参照结合图1-3描述的本发明的基于声纹的身份验证方法的实施例，这里不再详细描述。

[0074] 在此需要说明的是，图4-5所示的基于声纹的身份验证装置400和声纹提取器训练装置500及其组成单元的结构仅仅是示例性的，本领域技术人员可以根据需要对图4-5所示的结构框图进行修改。

[0075] 本发明提出了一种基于声纹的身份验证方法和装置，可以从较短(例如3-5秒)的语音中提取说话者的全息声纹，使得根据本发明的身份验证方法更加鲁棒。

[0076] [3.用以实施本申请的方法和装置的计算设备]

[0077] 以上结合具体实施例描述了本发明的基本原理，但是，需要指出的是，对本领域的普通技术人员而言，能够理解本发明的方法和装置的全部或者任何步骤或者部件，可以在任何计算装置(包括处理器、存储介质等)或者计算装置的网络中，以硬件、固件、软件或者它们的组合加以实现，这是本领域普通技术人员在阅读了本发明的说明的情况下运用他们的基本编程技能就能实现的。

[0078] 因此，本发明的目的还可以通过在任何计算装置上运行一个程序或者一组程序来实现。所述计算装置可以是公知的通用装置。因此，本发明的目的也可以仅仅通过提供包含实现所述方法或者装置的程序代码的程序产品来实现。也就是说，这样的程序产品也构成本发明，并且存储有这样的程序产品的存储介质也构成本发明。显然，所述存储介质可以是任何公知的存储介质或者将来所开发出来的任何存储介质。

[0079] 在通过软件和/或固件实现本发明的实施例的情况下，从存储介质或网络向具有专用硬件结构的计算机，例如图6所示的通用计算机600安装构成该软件的程序，该计算机在安装有各种程序时，能够执行各种功能等等。

[0080] 在图6中，中央处理单元(CPU)601根据只读存储器(ROM)602中存储的程序或从存储部分608加载到随机存取存储器(RAM)603的程序执行各种处理。在RAM 603中，也根据需要存储当CPU 601执行各种处理等等时所需的数据。CPU 601、ROM 602和RAM 603经由总线604彼此链路。输入/输出接口605也链路到总线604。

[0081] 下述部件链路到输入/输出接口605：输入部分606(包括键盘、鼠标等等)、输出部分607(包括显示器，比如阴极射线管(CRT)、液晶显示器(LCD)等，和扬声器等)、存储部分608(包括硬盘等)、通信部分609(包括网络接口卡比如LAN卡、调制解调器等)。通信部分609经由网络比如因特网执行通信处理。根据需要，驱动器610也可链路到输入/输出接口605。可拆卸介质611比如磁盘、光盘、磁光盘、半导体存储器等等根据需要被安装在驱动器610上，使得从中读出的计算机程序根据需要被安装到存储部分608中。

[0082] 在通过软件实现上述系列处理的情况下，从网络比如因特网或存储介质比如可拆卸介质611安装构成软件的程序。

[0083] 本领域的技术人员应当理解，这种存储介质不局限于图6所示的其中存储有程序、与设备相分离地分发以向用户提供程序的可拆卸介质611。可拆卸介质611的例子包含磁盘(包含软盘(注册商标))、光盘(包含光盘只读存储器(CD-ROM)和数字通用盘(DVD))、磁光盘(包含迷你盘(MD)(注册商标))和半导体存储器。或者，存储介质可以是ROM 602、存储部分608中包含的硬盘等等，其中存有程序，并且与包含它们的设备一起被分发给用户。

[0084] 本发明还提出一种存储有机器可读取的指令代码的程序产品。指令代码由机器读取并执行时，可执行上述根据本发明实施例的方法。

[0085] 相应地，用于承载上述存储有机器可读取的指令代码的程序产品的存储介质也包括在本发明的公开中。存储介质包括但不限于软盘、光盘、磁光盘、存储卡、存储棒等。

[0086] 本领域的普通技术人员应理解，在此所例举的是示例性的，本发明并不局限于此。

[0087] 在本说明书中，“第一”、“第二”以及“第N个”等表述是为了将所描述的特征在文字上区分开，以清楚地描述本发明。因此，不应将其视为具有任何限定性的含义。

[0088] 作为一个示例，上述方法的各个步骤以及上述设备的各个组成模块和/或单元可以实施为软件、固件、硬件或其组合，并作为相应设备中的一部分。上述装置中各个组成模

块、单元通过软件、固件、硬件或其组合的方式进行配置时可使用的具体手段或方式为本领域技术人员所熟知，在此不再赘述。

[0089] 作为一个示例，在通过软件或固件实现的情况下，可以从存储介质或网络向具有专用硬件结构的计算机(例如图6所示的通用计算机600)安装构成该软件的程序，该计算机在安装有各种程序时，能够执行各种功能等。

[0090] 在上面对本发明具体实施例的描述中，针对一种实施方式描述和/或示出的特征可以以相同或类似的方式在一个或更多个其他实施方式中使用，与其他实施方式中的特征相组合，或替代其他实施方式中的特征。

[0091] 应该强调，术语“包括/包含”在本文使用时指特征、要素、步骤或组件的存在，但并不排除一个或更多个其他特征、要素、步骤或组件的存在或附加。

[0092] 此外，本发明的方法不限于按照说明书中描述的时间顺序来执行，也可以按照其他的时间顺序地、并行地或独立地执行。因此，本说明书中描述的方法的执行顺序不对本发明的技术范围构成限制。

[0093] 本发明及其优点，但是应当理解在不超出由所附的权利要求所限定的本发明的精神和范围的情况下可以进行各种改变、替代和变换。而且，本发明的范围不仅限于说明书所描述的过程、设备、手段、方法和步骤的具体实施例。本领域内的普通技术人员从本发明的公开内容将容易理解，根据本发明可以使用执行与在此的相应实施例基本相同的功能或者获得与其基本相同的结果的、现有和将来要被开发的过程、设备、手段、方法或者步骤。因此，所附的权利要求旨在在它们的范围内包括这样的过程、设备、手段、方法或者步骤。

[0094] 基于以上的说明，可知公开至少公开了以下技术方案：

[0095] 1、一种基于声纹的身份验证方法，包括：

[0096] 接收未知语音；

[0097] 利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹；

[0098] 将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹；以及

[0099] 利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人。

[0100] 2、根据附记1所述的身份验证方法，其中，通过下述步骤来训练得到基于神经网络的声纹提取器：

[0101] 采集来自不同说话者的语音；

[0102] 将说话者的辅助信息作为神经网络的分组标签进行标记，将标记过的语音作为训练样本输入神经网络；以及

[0103] 进行深度学习来得到所述基于神经网络的声纹提取器。

[0104] 3、根据附记2所述的身份验证方法，其中，所述说话者的辅助信息包括：说话者的语种、音素序列、发声通道、情感、年龄、生活区域以及性别中的一项或多项。

[0105] 4、根据附记2所述的身份验证方法，其中，进行深度学习来得到所述基于神经网络的声纹提取器包括：

[0106] 采用层次神经网络将不同的分组标签分别放置在不同层进行深度学习得到层次网络提取器；以及

[0107] 采用扁平神经网络将全部分组标签放置在输出层进行深度学习得到扁平网络提

取器。

[0108] 5、根据附记4所述的身份验证方法，其中，利用预先训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹的步骤包括：

[0109] 分别利用所述层次网络提取器和所述扁平网络提取器提取来提取所述未知语音中的声纹；以及

[0110] 将所述层次网络提取器所提取的声纹和所述扁平网络提取器所提取的声纹拼接在一起作为所提取的所述未知语音中的声纹。

[0111] 6、根据附记4所述的身份验证方法，其中，所述分组标签包含的信息越简单，则该分组标签在层次神经网络中放置的位置越靠前。

[0112] 7、根据附记2所述的身份验证方法，其中，利用随机梯度下降方法进行所述深度学习，所述随机梯度方法使用以下公式来计算神经网络的权重：

[0113] $w_{t+1}=\theta_t+(a_t-1)(\theta_t-\theta_{t-1})/a_{t+1}+a_t(\theta_t-w_t)/a_{t+1}$

[0114] 其中，w_t是权重，$\theta_t = w_t - \epsilon_t \nabla f(w_t)$，$\epsilon_t'$是学习率，$\nabla f(w_t)$是损失函数f的梯度，$a_0=1$，$a_{t+1} = (1+\sqrt{1+a_t^2})/2$。

[0115] 8、根据附记1所述的身份验证方法，其中，利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人的步骤包括：

[0116] 训练第一高斯混合模型和第二高斯混合模型，其中，所述第一高斯混合模型对应两个声纹属于同一个人的情况，所述第二高斯混合模型对应两个声纹不属于同一个人的情况；

[0117] 计算所述拼接的声纹分别在所述第一高斯混合模型上的第一概率和所述第二高斯混合模型上的第二概率；以及

[0118] 当所述第一概率大于所述第二概率，则所提取的声纹与预先存储的声纹是来自同一个人，否则，则是来自不同的人。

[0119] 9、根据附记2所述的身份验证方法，其中，所提取的所述未知语音中的声纹包含所述说话者的辅助信息。

[0120] 10、一种基于声纹的身份验证装置，包括：

[0121] 语音接收单元，被配置为接收未知语音；

[0122] 声纹提取单元，被配置为利用声纹提取器训练装置训练得到的基于神经网络的声纹提取器来提取所述未知语音中的声纹；

[0123] 声纹拼接单元，被配置为将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹；以及

[0124] 身份验证单元，被配置为利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人。

[0125] 11、根据附记10所述的身份验证装置，其中，所述声纹提取器训练装置包括：

[0126] 语音采集单元，被配置为采集来自不同说话者的语音；

[0127] 标记单元，被配置为将说话者的辅助信息作为神经网络的分组标签进行标记，将标记过的语音作为训练样本输入神经网络；以及

[0128] 训练单元，被配置为进行深度学习来得到所述基于神经网络的声纹提取器。

[0129] 12、根据附记11所述的身份验证方法，其中，所述说话者的辅助信息包括：说话者的语种、音素序列、发声通道、情感、年龄、生活区域以及性别中的一项或多项。

[0130] 13、根据附记11所述的身份验证方法，其中，所述训练单元进一步被配置为：

[0131] 采用层次神经网络将不同的分组标签分别放置在不同层进行深度学习得到层次网络提取器；以及

[0132] 采用扁平神经网络将全部分组标签放置在输出层进行深度学习得到扁平网络提取器。

[0133] 14、根据附记13所述的身份验证方法，其中，所述声纹提取单元进一步被配置为：

[0134] 分别利用所述层次网络提取器和所述扁平网络提取器提取来提取所述未知语音中的声纹；以及

[0135] 将所述层次网络提取器所提取的声纹和所述扁平网络提取器所提取的声纹拼接在一起作为所提取的所述未知语音中的声纹。

[0136] 15、根据附记13所述的身份验证方法，其中，所述分组标签包含的信息越简单，则该分组标签在层次神经网络中放置的位置越靠前。

[0137] 16、根据附记11所述的身份验证方法，其中，所述深度学习利用随机梯度下降方法进行，所述随机梯度方法使用以下公式来计算神经网络的权重：

[0138] $w_{t+1}=\theta_t+(a_t-1)(\theta_t-\theta_{t-1})/a_{t+1}+a_t(\theta_t-w_t)/a_{t+1}$

[0139] 其中，w_t是权重，$\theta_t = w_t - \epsilon_t \nabla f(w_t)$，$\in_t'$是学习率，$\nabla f(w_t)$是损失函数f的梯度，$a_0=1$，$a_{t+1} = (1+\sqrt{1+a_t^2})/2$。

[0140] 17、根据附记10所述的身份验证方法，其中，所述身份验证单元进一步被配置为：

[0141] 训练第一高斯混合模型和第二高斯混合模型，其中，所述第一高斯混合模型对应两个声纹属于同一个人的情况，所述第二高斯混合模型对应两个声纹不属于同一个人的情况；

[0142] 计算所述拼接的声纹分别在所述第一高斯混合模型上的第一概率和所述第二高斯混合模型上的第二概率；以及

[0143] 当所述第一概率大于所述第二概率，则所提取的声纹与预先存储的声纹是来自同一个人，否则，则是来自不同的人。

[0144] 18、根据附记11所述的身份验证方法，其中，所提取的所述未知语音中的声纹包含所述说话者的辅助信息。

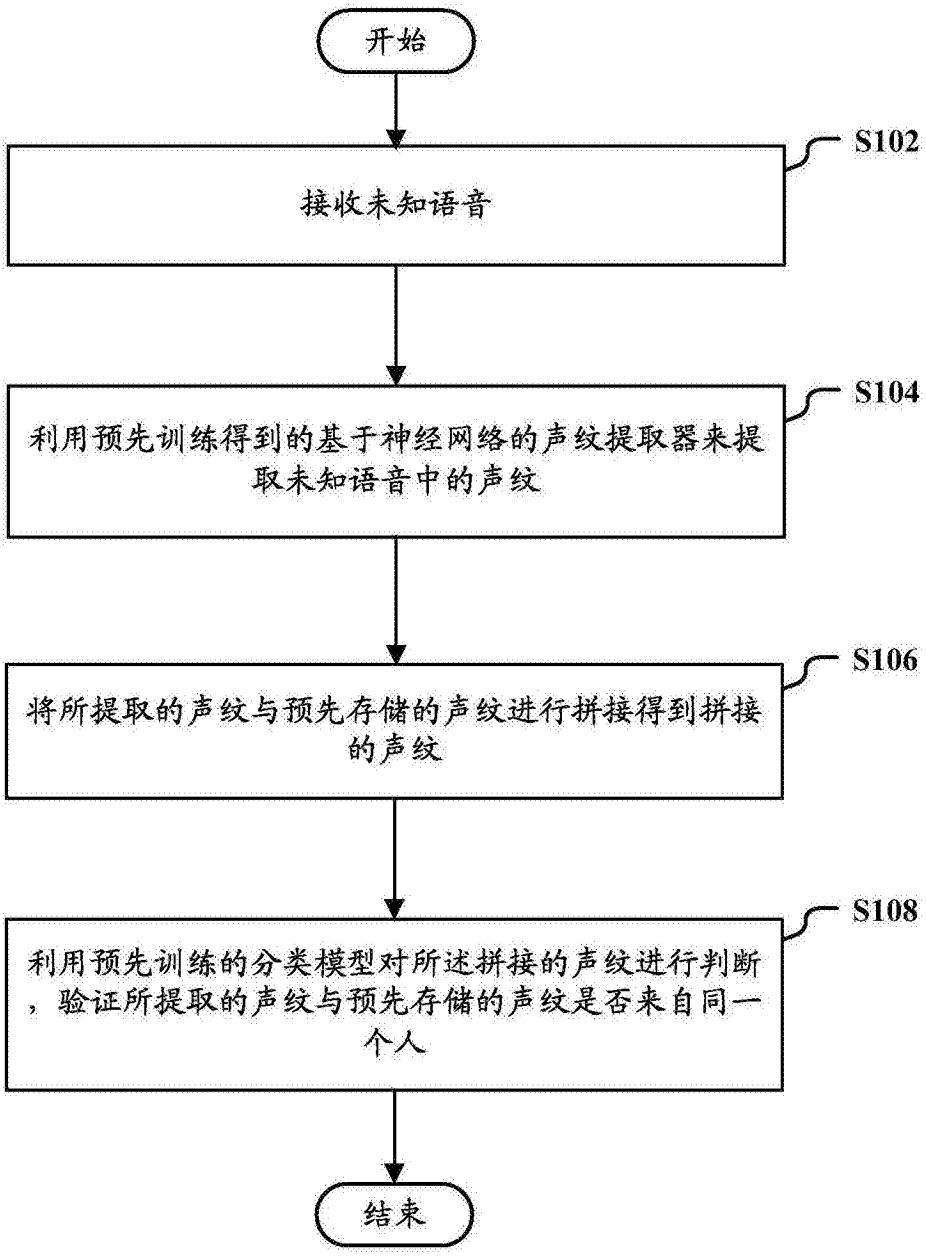

开始
S102
接收未知语音
S104
利用预先训练得到的基于神经网络的声纹提取器来提取未知语音中的声纹
S106
将所提取的声纹与预先存储的声纹进行拼接得到拼接的声纹
S108
利用预先训练的分类模型对所述拼接的声纹进行判断，验证所提取的声纹与预先存储的声纹是否来自同一个人
结束

图1

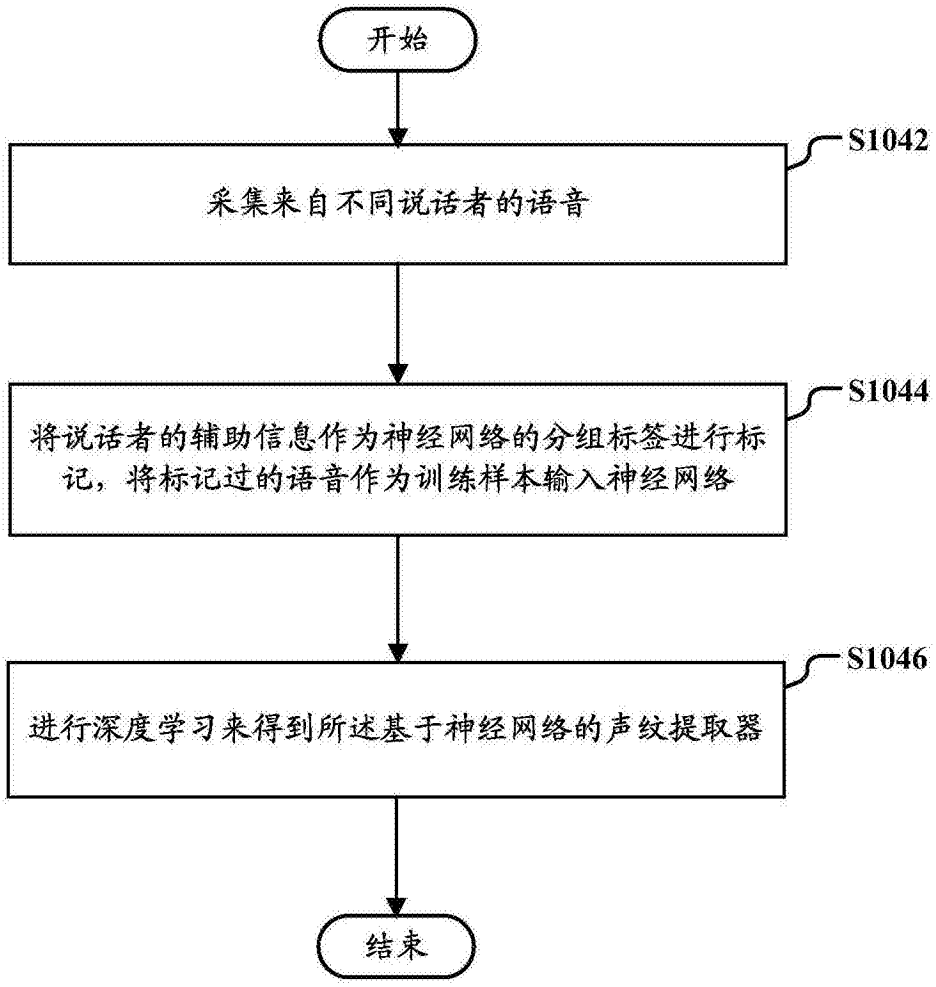

开始
S1042
采集来自不同说话者的语音
S1044
将说话者的辅助信息作为神经网络的分组标签进行标记，将标记过的语音作为训练样本输入神经网络
S1046
进行深度学习来得到所述基于神经网络的声纹提取器
结束

图2

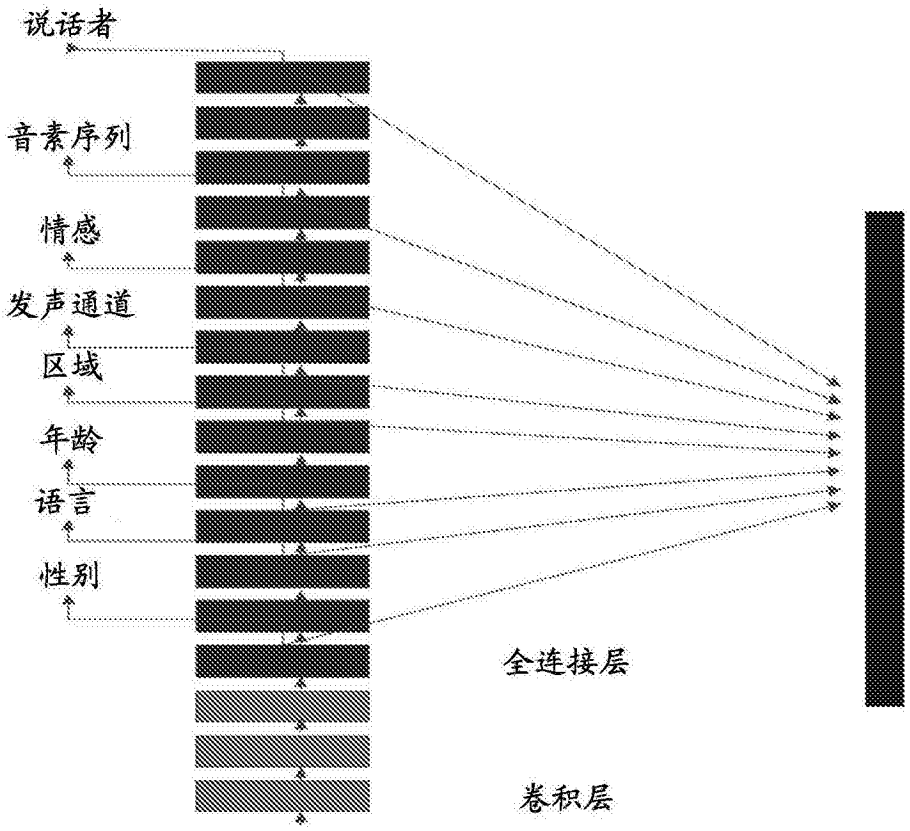

说话者
音素序列
情感
发声通道
区域
年龄
语言
性别
全连接层
卷积层

图3A

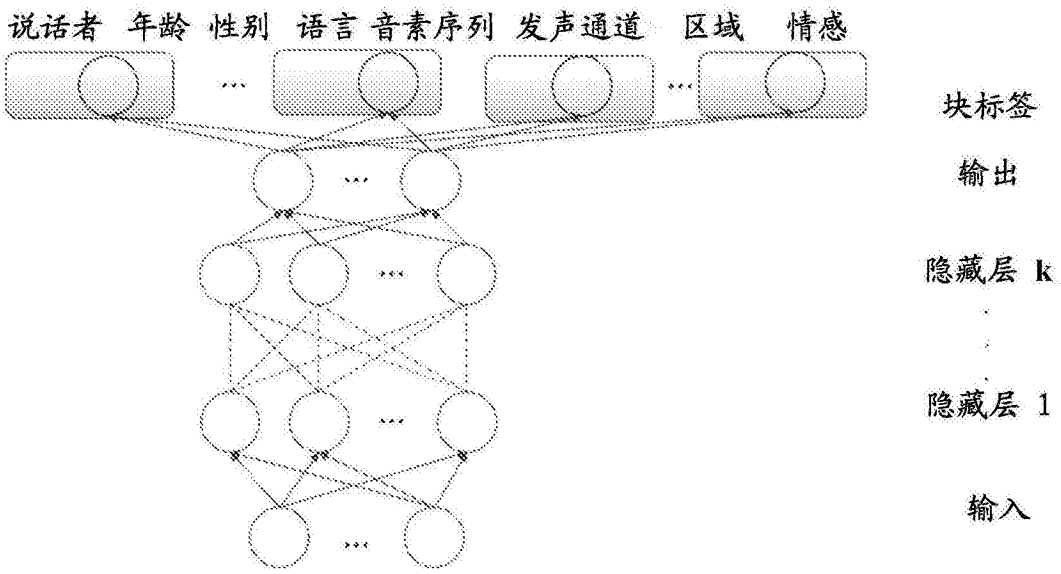

说话者 年龄 性别 语言 音素序列 发声通道 区域 情感
块标签
输出
隐藏层 k
隐藏层 1
输入

图3B

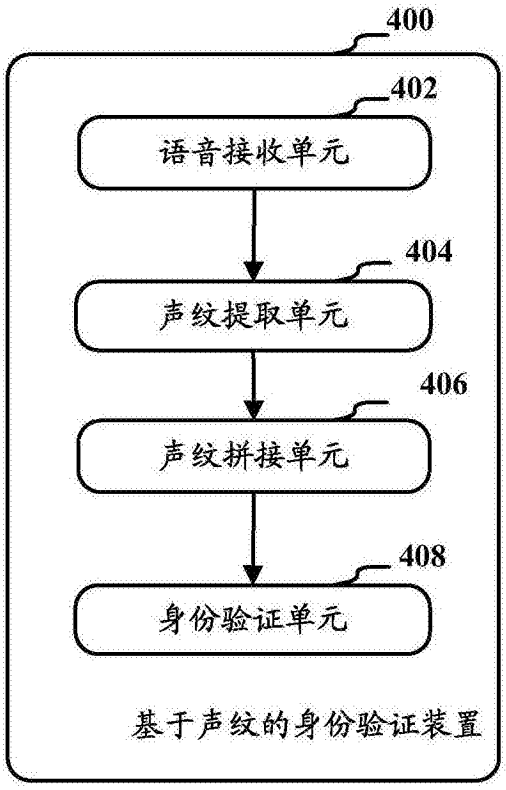

400
402
语音接收单元
404
声纹提取单元
406
声纹拼接单元
408
身份验证单元
基于声纹的身份验证装置

图4

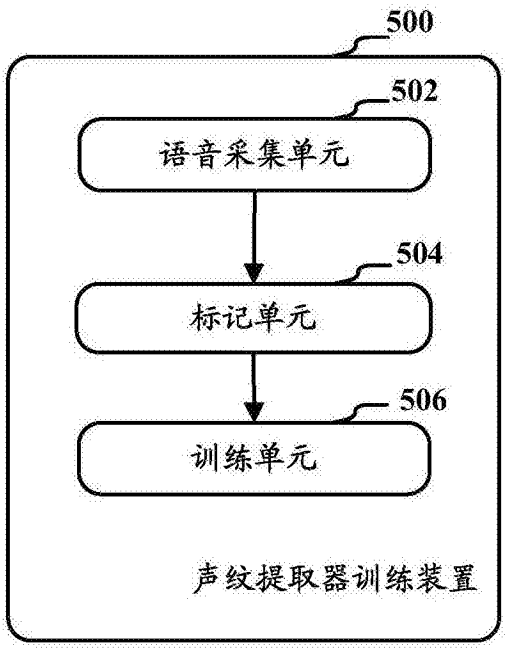

500
502
语音采集单元
504
标记单元
506
训练单元
声纹提取器训练装置

图5

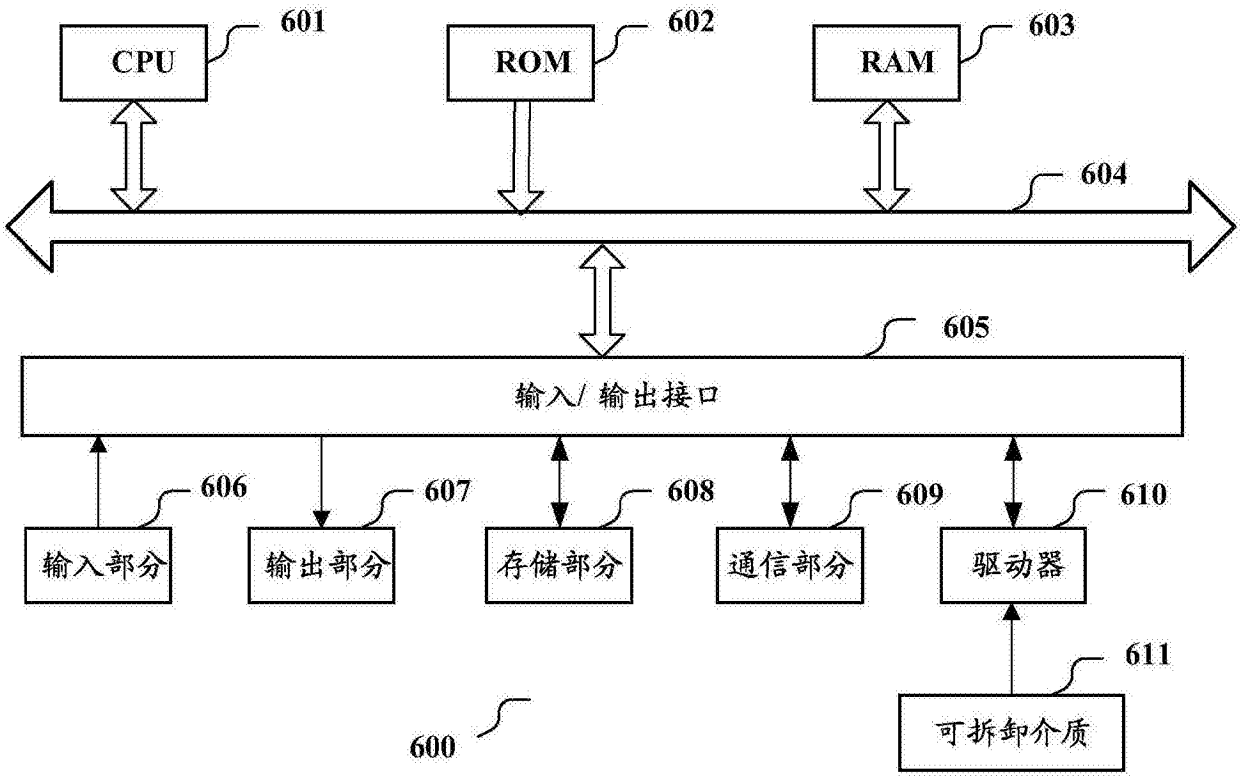

601
CPU
602
ROM
603
RAM
604
605
输入/ 输出接口
606
输入部分
607
输出部分
608
存储部分
609
通信部分
610
驱动器
611
可拆卸介质
600

图6