[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G10L 15/00
//G10L101：00

[12] 发明专利说明书

[21] ZL 专利号 99118389.4

[45] 授权公告日 2005 年 2 月 9 日

[11] 授权公告号 CN 1188828C

[22] 申请日 1999.9.3 [21] 申请号 99118389.4

[30] 优先权

[32] 1998. 9. 4 [33] US [31] 09/148，911

[71] 专利权人 松下电器产业株式会社

地址 日本大阪府

[72] 发明人 罗兰德·库恩 帕特里克·贵恩
吉恩－克劳德·琼克瓦
罗伯特·博曼

审查员 杨 叁

[74] 专利代理机构 中国国际贸易促进委员会专利商标事务所

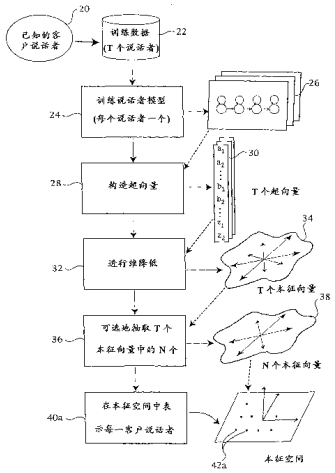
代理人 于 静

权利要求书 2 页 说明书 14 页 附图 6 页

[54] 发明名称 基于本征话音的说话者检验和说话者识别

[57] 摘要

对已知的客户说话者(在说话者检验的情形下，还对冒名顶替说话者)构造并训练语音模型。连接来自这些模型的参数以定义超向量，并对这些超向量作线性变换其结果为维数降低，产生称为本征空间的低维空间。 然后训练说话者被表示为本征空间中的点或分布。 此后把来自测试说话者新的语音数据通过类似的线性变换放置在本征空间中，并且测试说话者对训练说话者在本征空间中的接近程度用来鉴别或识别测试说话者。



ISSN1008-4274

1. 用于对预定客户说话者执行说话者检验或说话者识别的方法，包括：

对来自多个训练说话者的语音训练一组语音模型，该多个训练说话者包括至少一个客户说话者；

通过对所述语音模型组进行维数降低来构造表示所述多个训练说话者的本征空间，以便产生定义所述本征空间的一组基向量；

把所述客户说话者表示为所述本征空间中的第一位置；

根据来自新说话者的语音输入数据训练新的语音模型；

对所述新的语音模型进行维数降低，以产生所述新的说话者作为本征空间中第二位置的表示；

估计所述第一和第二位置之间的接近程度，并使用所述估计作为新的说话者是否为客户说话者的指示。

2. 根据权利要求 1 的说话者检验或识别方法，其中，所述多个训练说话者包括多个不同的客户说话者，并且其中，所述方法还包括：

把所述多个客户说话者中的每一个表示为所述本征空间中的训练说话者位置，以及

估计所述第二位置和所述训练说话者位置之间的接近程度，并至少部分地基于所述接近程度的估计，把所述新的说话者识别为所述多个客户说话者中选择的一个。

3. 根据权利要求 1 的说话者检验或识别方法，其中，所述多个训练说话者包括被表示为本征空间中第三位置的至少一个冒名顶替说话者。

4. 根据权利要求 3 的说话者检验或识别方法，还包括附加估计所述第二和第三位置之间的接近程度，并使用所述附加估计作为对新的说话者是否为客户说话者的进一步指示。

5. 权利要求 1 的说话者检验或识别方法，其中，估计接近程度

的所述步骤通过确定所述第一和第二位置之间的距离进行。

6. 权利要求 1 的说话者检验或识别方法，其中，所述训练说话者被表示为所述本征空间中的位置。

7. 权利要求 1 的说话者检验或识别方法，其中，所述训练说话者被表示为所述本征空间中的点。

8. 权利要求 1 的说话者检验或识别方法，其中，所述训练说话者被表示为所述本征空间中的分布。

9. 权利要求 1 的说话者检验或识别方法，其中，执行维数降低还包括使用所述输入数据产生一概率函数，并且然后使所述概率函数极大化以确定位于所述本征空间内的一个极大似然向量。

10. 权利要求 1 的说话者检验或识别方法，其中，所述多个训练说话者包括多个客户说话者和至少一个冒名顶替说话者。

11. 权利要求 1 的说话者检验或识别方法，还包括周期地估计所述第一和第二位置之间的接近程度，并使用所述估计作为新的说话者是否为客户说话者的指示，以便确定所述新的说话者身份是否有变化。

基于本征话音的说话者检验和说话者识别

技术领域

本发明一般涉及语音技术，并特别涉及用于进行说话者检验或说话者识别的系统和方法。

背景技术

授权问题处于几乎每一项交易的核心。成百万的人通过电话进行保密的金融交易，诸如访问他们的银行帐户或使用他们的信用卡。当前实际进行的授权远非完全安全的。各方面交换认为秘密的信息的某种形式，诸如社会保险号码，母亲未婚前娘家的姓等。显然，这种信息可能受到侵犯，其结果是伪冒的授权。

本发明的一方面是要通过提供用于进行说话者检验的系统和方法解决上述问题。说话者检验涉及确定给定的话音是属于一定说话者(这里称为“客户”)还是冒名顶替者(客户以外的任何人)。

与说话者检验相关的问题是说话者识别问题。说话者识别涉及把给定的话音与一组已知的话音之一匹配。类似于说话者检验，说话者识别具有一些有吸引力的应用。例如，说话者识别系统可用于对话音样本可得的一组说话者发出的话音邮件进行分类。这种功能允许计算机实现的电话系统在计算机屏幕上显示已经在话音邮件系统上留言的呼叫者的身份。

虽然说话者检验和说话者识别的应用实际上是无限的，但迄今进行这两个任务的解决方法证明是困难的。识别人类语音、特别是从其它说话者鉴别一说话者是一个复杂的问题。由于人类语音是如此产生的，即使是单独一个词一个人很少以相同的方式说出两次。

人类语音是空气在压力下从肺脏用力通过声带的产物，并受到声门的调制产生声波，然后该声波在由舌头、颌部、牙齿和嘴唇清晰发音之前，在口腔和鼻腔中共鸣。许多因素影响这些声音产生机制如何相互作用。例如，通常的感冒就会大大改变鼻腔的共鸣以及声带音调的质量。

由于人类产生语音的复杂性和多变性，通过比较新的说话者与先

前的记录语音样本并不能容易地进行说话者检验和说话者识别。为了排除冒名顶替者而采用高相似性阈值，但当他或她患感冒时，可能会排除授权的说话者。另一方面，采用低相似性阈值能够使系统倾向于作出错误的检验。

发明内容

本发明提供用于对预定客户说话者执行说话者检验或说话者识别的方法，包括：对来自多个训练说话者的语音训练一组语音模型，该多个训练说话者包括至少一个客户说话者；通过对所述语音模型组进行维数降低来构造表示所述多个训练说话者的本征空间，以便产生定义所述本征空间的一组基向量；把所述客户说话者表示为所述本征空间中的第一位置；根据来自新说话者的语音输入数据训练新的语音模型；对所述新的语音模型进行维数降低，以产生所述新的说话者作为本征空间中第二位置的表示；估计所述第一和第二位置之间的接近程度，并使用所述估计作为新的说话者是否为客户说话者的指示。

本发明对说话者检验和说话者识别使用基于模型的分析方法。对已知的客户说话者的语音(并在说话者检验的情形下还对一个或多个冒名顶替者的语音)构造模型并进行训练。这些说话者模型一般采用复合参数(诸如隐藏马尔科夫模型参数)。不是直接使用这些参数，而是把参数连接在一起形成超向量。这些超向量，每个说话者一个，表示整个训练数据的说话者分布。

对超训练进行结果为维数降低的线性变换，这产生我们称为本征空间的低维空间。这一本征空间的基向量我们称为“本征话音”向量或“本征向量”。如果需要，能够通过抛弃某些本征向量项在维数上进一步降低本征空间。

然后，在本征空间中表示出包含训练数据的每一说话者，或者作为本征空间中的一个点，或者作为本征空间中的概率分布。前者精确度稍低，在于这方法把来自每一说话者的语音相对不变地处理。后者反映出每一说话者的语音将随发音发生变化。

在本征空间中对每一说话者表示出训练数据后，系统可用于进行说话者检验或说话者识别。

获得新的说话者数据并用于构造超向量，然后其维数降低并在本征空间中表示。估计新的说话者数据对本征空间中先前数据的近似程度，进行说话者检验或说话者识别。如果其在本征空间内的对应

点或分布在对该客户说话者的训练数据的阈值近似度之内，则确认来自说话者的新的语音。如果其放置在本征空间中落在比较接近冒名顶替者语音，则系统在进行授权时可能会拒绝新的语音。

说话者识别以类似的方式进行。新的语音数据放置在本征空间中，并对分布的本征向量点最接近的训练说话者进行识别。

估计本征空间中新的语音数据和训练数据的近似程度具有数个优点。首先，本征空间以精确的低维方式表示出每一完整的说话者，不只是选择每一说话者少量特性。在本征空间中进行近似程度的计算能够相当快速地进行，因为与原始说话者模型空间或特征向量空间中相比，在本征空间中一般只需对相当少的维数进行处理。而且系统不需要新语音数据每一包含在构造原始训练数据所使用的每一例子或发音。通过这里所述的技术，能够对超向量进行维数降低，抛弃其某些成分。然而这样产生的分布在本征空间的点将能非常好地表示说话者。

为了完整地理解本发明、其目的和优点，请参见以下说明及附图。

附图说明

图 1 表示理解本发明使用的示例性的隐藏马尔科夫模型(HMM)；

图 2 是表示如何构造本征空间以实现说话者识别系统的流程图，其中已知的客户说话者表示为本征空间中的一个点；

图 3 是表示如何构造本征空间以实现说话者检验系统的流程图，其中客户说话者和潜在的冒名顶替者按本征空间中的分布来表示；

图 4 是表示使用在训练期间形成的本征空间可进行说话者识别或说话者检验的过程的流程图；

图 5 是如何实施极大似然技术的示意图；

图 6 是表示如何基于极大似然运算将来自说话者的观测数据放置到本征空间中的数据结构图。

具体实施方式

本发明所采用的本征话音技术将对许多不同的语音模型有效。我们说明与隐藏马尔科夫模型识别器相关的优选实施例，因为其在当今的语音识别技术中具有普遍性。然而应当理解，本发明能够使用任何其它类型的基于模型的识别器实现，诸如音素相似性识别器。

为了更好地理解本发明的说话者识别和检验技术，对语音识别

系统基本的理解是有帮助的。迄今当前大多数语音识别器采用隐藏马尔科夫模型(HMM)表示语音，这里将说明 HMM 技术使读者熟悉。

隐藏马尔科夫模型是涉及状态图的建模方法。任何语音单元(诸如短语、词、半词、音素等)都能够以包含在该模型中的所有知识源被建模。HMM 表示产生离散区间可观测的输出序列一种未知的过程，输出是某些有限的字母表成员(对应于语音单元预定的集合)。这些模型被称为“隐藏的”，因为产生可观测输出的状态序列是未知的。

如图 1 所示，HMM 10 由状态集合($S_1, S_2, \cdots S_5$)、定义图 1 中箭头所示的某些状态对之间的转移的向量以及概率数据集合来表示。特别地，隐藏马尔科夫模型包括与转移向量相关的转移概率集合 12 以及与每一状态观测的输出相关的向量和输出概率集合 14。对模型从一个状态到另一状态按规则间隔、离散区间定时。按时钟时间，模型可以从其当前状态变为对其转移向量存在的任何状态。如图所示，转移可从给定的状态返回到自身。

转移概率表示当对模型计时时从一个状态向另一状态转移将发生的似然率。于是如图 1 所示，每一转移与一概率值(0 与 1 之间)相联系。处于任意状态的所有概率之和等于 1。举例来说，在转移概率表格 12 中给出了示例性转移概率值集合。应当理解，在一有效的实施例中，这些值将由训练数据产生，其限制是处于任意状态的所有概率之和等于 1。

每次进行转移时，可以把模型设想为发出或输出其字母表的一个成员。在图 1 所示的实施例中，假设基于音素的语音单元。这样在输出概率表 14 中定义的符号对应于标准英语中找到的音素。在每一转移时发出哪一个字母表成员取决于输出概率值或训练期间学习的函数。这样发出的输出表示观测的序列(基于训练数据)，并且每一字母表成员有一被发出的概率。

在对语音建模中，通常实际的作法是把输出作为与离散字母表

符号序列相对的连续向量序列。这需要输出概率表示为与单个数值相对的连续概率函数。这样，HMM 常常基于包括一个或多个高斯分布的概率函数。当使用多个高斯函数时，如在 16 处所示，它们一般相加地混合在一起以定义一复合的概率分布。

无论表示为单一高斯函数还是表示为高斯函数的混合，概率分布能够由多个参数描述。如同转移概率值(表 12)那样，这些输出概率参数可能包含浮点数。参数表 18 标识一般用来基于来自训练说话者的观测数据表示概率密度函数(pdf)。由图 1 中高斯函数 16 的方程式所示，要进行建模的观测向量 O 的概率密度函数是乘以由高斯密度 N 的每一混合分量的混合系数的叠代和，其中高斯密度具有平均向量 u_j，以及从倒谱或滤波器组系数语音参数计算的协方差矩阵 U_j。

隐藏马尔科夫模型识别器实现的细节从一个应用到另一应用可以有很大变化。图 1 所示的 HMM 例子只是要解释隐藏马尔科夫模型是如何构造的，并不是作为对本发明范围的限制。就此而言，有许多各种不同的隐藏马尔科夫建模概念。正如从以下说明能够更允分理解那样，本发明的本征话音适应技术能够很好地适用于每一种不同的隐藏马尔科夫模型变形，以及其它基于参数的语音建模系统。

图 2 和 3 分别表示，使用本发明的技术如何进行说话者识别和说话者检验。作为进行说话者识别或说话者检验的第一步，要构造本征空间。要构造的具体的本征空间与应用有关。在图 2 所示的说话者识别的情形下，使用一组已知的客户说话者 20 提供对其生成本征空间的训练数据 22。另一方面，对于如图 3 所示的说话者检验，从希望对其进行检验的客户说话者 21a，以及还从一个或多个潜在的冒名顶替者 21b 提供训练数据。对说话者识别和说话者检验应用，除了训练数据源的这一区别外，用于产生本征空间的过程基本上相同。于是对图 2 和 3 使用了类似的标号。

参照图 2 和 3，通过对在训练数据 22 中表示的每一说话者形成

并训练说话者模型而构造本征空间。这一步骤示于24，并对每一说话者产生一组模型26。虽然这里已经解释隐藏马尔科夫模型，但是本发明不限于隐藏马尔科夫模型。而是可使用具有适于连接的参数的任何语音模型。模型26最好使用足够的训练数据训练，使得由模型所定义的所有声音单元由每一说话者实际的语音至少一个例子训练。虽然在图2和3中没有明显示出，但模型训练步骤24能够包含适当的辅助说话者适应处理，以便加细模型。这种辅助处理的例子包括极大A后验估计(MAP)及其它基于变换的方法，诸如极大似然线性回归(MLLR)。生成说话者模型26的目的是要精确地表示训练数据语料库，因为这个语料库要用来定义每一训练说话者被放置在其中，并对其测试每一新的语音发音的本征空间的界线和边界。

在构造模型26之后，在步骤28使用每一说话者的模型构造超向量。30处所示的超向量可通过连接每一说话者模型的参数形成。在使用隐藏马尔科夫模型时，每一说话者的超向量可组成参数(一般为浮点数)的一有序列表，这些参数对应于该说话者隐藏马尔科夫模型的至少一部分参数。对应于每一声音单元的参数包含在给定的说话者超向量中。这些参数可以任何方便的顺序组织起来。顺序不是重要的；然而一旦采用一种顺序，则对所有的训练说话者必须遵从。

用来构造超向量的模型参数的选择将取决于计算机系统可用的处理能力。当使用隐藏马尔科夫模型参数时，我们通过从高斯均值构造超向量而获得了良好的结果。如果可使用更大的处理能力，超向量还可包括其它的参数，诸如转移概率(图1表12)，或协方差矩阵参数(图1参数18)。如果隐藏马尔科夫模型产生离散输出(与概率密度相反)，则这些输出值可用来组成超向量。

在构造超向量之后，在步骤32进行维数降低操作。维数降低能够通过把原来的高维超向量降低为基向量的任何线性变换实现。例子的非穷尽列表包括：

主成分分析(PCA)，独立成分分析(ICA)，

线性鉴别分析(LDA)，因素分析(FA)，单值分解(SVD)。

具体来说，在实现本发明中使用的维数降低技术的分类定义如下。考虑从用于语音识别的说话者相关模型获得的一组 T 个训练超向量。设这些超向量的每一个具有维数 V；这样，我们能够把每一超向量标记为 X=[x1, x2, …, xV]^T(V*1 向量)。考虑能够施加到超向量(即施加到维数 V 的任何向量)以产生新的维 E 的向量(E 小于或等于训练超向量数目 T)；每一变换后的向量可标记为 W=[w1, w2, …, wV]^T。以某种方式从 T 个训练超向量的组计算 M 的参数值。

这样，我们具有线性变换 W=M*X。M 有维数 E*V，且 W 具有维数 E*1，其中 E<=T;对于特定的训练超向量组，M 将是固定不变的。可使用几种维数降低技术从 T 个训练超向量的组计算线性变换 M，使 W 具有维数 E<=T。

例子包括主成分分析、独立成分分析、线性鉴别分析、因素分析、单值分解。在输入向量为从说话者相关建模导出的训练超向量、并且 M 用来实施上述技术的具体情形下，可使用任何用于找出这种固定线性变换 M 的方法(不仅是那些列出的)实现本发明。

在步骤 32 产生的基向量定义由本征向量覆盖的一本征空间。维数降低对每一训练的说话者产生一本征向量。这样，如果有 T 个训练说话者，则维数降低步骤 32 产生 T 个本征向量。这些本征向量定义了所谓本征话音空间或本征空间。

如 34 处所示，形成本征话音空间的本征向量每一表示可通过其区分不同说话者的不同维。原始训练集中每一超向量可被表示为这些本征向量的线性组合。本征向量按它们在对数据建模中的的重要性排序：第一本征向量比第二本征向量重要，第二本征向量比第三本征向量重要等等。至此我们对这一技术的经验表明，第一本征向量似乎对应于性别维。

虽然在步骤 32 产生的极大 T 个本征向量，实际上能够抛弃这些向量的几个，仅保留前 N 个本征向量。这样在步骤 36 我们可选地抽取 T 个本征向量的 N 个，在步骤 38 组成降低的参数本征空间。较高阶的本征向量可被抛弃，因为它们一般包含用于在说话者之间进行

鉴别的次要信息。把本征话音空间降低到少于训练说话者总数就提供了本质的数据压缩，这在以有限的存储器和处理器资源构造实际系统时能够有帮助。

在从训练数据产生了本征向量之后，在本征空间中表示出训练数据中的每一说话者。在说话者识别的情形下，如步骤 40a 所示及 42a 处图示，在本征空间中表示出每一已知的客户说话者。在说话者检验的情形下，如步骤 40b 所示及 42b 处所示，在本征空间中表示出客户说话者和潜在的冒名顶替说话者。说话者可以表示为本征空间中的点(如图 2 中 42a 处所示)或表示为本征空间中的概率分布(如图 3 中 42b 处所示)。

使用说话者识别或说话者检验系统

寻求说话者识别或检验的用户在 44 提供新的语音数据，且如步骤 46 处所示，这些语音数据用来训练说话者相关模型。然后在步骤 50 使用模型 48 构造超向量 52。注意，新的语音数据可能不需要包含每一声音单元的例子。例如，新的语音发音可能太短而不能包含所有声音单元的例子。如以下将充分说明的，系统将处理这种情形。

在步骤 54 对超向量 52 进行维数降低，其结果是如步骤 56 所示及 58 处所示可在本征空间中表示的新的数据点。在 58 的图示中本征空间(基于训练说话者)中先前所需的点表示为圆点，而新的语音数据点表示为星号。

把新的数据点放置到本征空间之后，现在可以估计其对其它先前的数据点逼近程度，或对应于训练说话者的数据分布。图 4 示出说话者识别和说话者检验的两者的一示例性实施例。

对于说话者识别，把新的语音数据指定给本征空间中最接近的训练说话者,步骤 62 图示在 64 处。这样系统将把新的语音标识为其数据点或数据分布在本征空间中最接近新的语音的先前的训练说话者的语音。

对于说话者检验，系统在步骤 66 测试新的数据点以确定它是否与本征空间中客户说话者处于预定的阈值接近程度。如果新的说话

者数据在本征空间中更为接近冒名顶替者而不是客户说话者，则作为安全措施在步骤 68，系统可以拒绝新的说话者数据。这图示在 69 处，其中描绘出对客户说话者的接近程度和对最接近的冒名顶替者的接近程度。

极大似然本征空间分解(MLED)技术

一个用于把新的说话者放置在本征空间内的简单的技术是使用简单的投影运算。投影运算寻找尽可能接近对应于新的说话者输入语音本征空间之外的点的本征空间内的点。请记住，这些点实际上是从其能够重新构造一组 HMM 的超向量。

投影运算是比较粗糙的技术，它不能保证本征空间内的点对新的说话者最优。此外，投影运算要求对新的说话者超向量包含完整的数据集，以表示对该说话者整个的 HMM 组。这一要求引起实施上相当大的限制。当使用投影把新的说话者约束到本征空间时，说话者必须提供足够的输入语音，使所有的语音单元能在数据中表示。例如，如果隐藏马尔科夫模型指定表示英语中所有的音素，则在使用简单投影技术之前，训练说话者必须提供所有音素的例子。在许多应用中，这一约束简直是不实际的。

本发明的极大似然技术要解决简单投影的上述两个缺陷。本发明的极大似然技术寻求本征空间内的一点，该点表示对应于具有产生由新说话者提供的语音的最大概率的一组隐藏马尔科夫模型的超向量。

简单的投影运算把所有的超向量成员作为具有同等重要性对待，而最大似然技术是基于从实际适应数据引起的概率的，这样更可能的数据权重越重。与简单投影技术不同，即使新的说话者没有提供完全的训练数据集合(即对某些声音单元的数据缺失)，极大似然技术仍将有效。实际上，极大似然技术把构造超向量的场合考虑在内，即从涉及一定模型比另外的模型更可能产生由新说话者提供的输入语音的概率的隐藏马尔科夫模型进行构造。

实际上，极大似然技术将在本征空间内选择与新的说话者输入

语音最一致的超向量，而不论实际上究竟有多少输入语音可得。为了说明，假设新的说话者是 Alabama 当地人的年轻女性。在收到来自这一说话者发出的一些音节时，极大似然技术将在本征空间内选择表示与说话者的当地 Alabama 女性口音一致的所有音素(即使那些在输入语音中还没有表示的音素)的点。

图 5 表示极大似然技术如何工作。来自新说话者的语音输入用来构造超向量 70。如上所述，超向量包括对应于倒谱系数等语音参数的连接列表。在所示的实施例中，这些参数是表示从对应于新说话者的隐藏马尔科夫模型集合抽取的高斯均值的浮点数。其它的 HMM 参数也可使用。在图示中，这些 HMM 均值作为如 72 处的圆点所示。当以数据完全分布时，超向量 70 将对每一 HMM 均值包含对应于由 HMM 模型表示的每一声音单元的浮点数。为了进行说明，这里假设音素“ah”的参数出现，而音素“iy”的参数缺失。

本征空间 38 由本征向量 74、76 和 78 的集合表示。对应于来自新说话者的观测数据的超向量 70 可在本征空间中由每一本征向量乘以标记为 $W_1, W_2, \cdots W_n$ 的对应的本征值表示。这些本征值起初是未知的。极大似然技术寻找这些未知本征值的值。如将以下更充分说明那样，通过寻找将能在本征空间中最佳表示新说话者的优化解而选择这些值。

在使本征值与对应的本征空间 38 的本征向量相乘并对结果乘积求和之后，产生一个适应模型 80。由于输入语音的超向量(超向量 70)可能已有某些缺失的参数值(例如“iy”参数)，表示适应模型的超向量 80 以数值完全分布。此即本发明的一个好处。此外，超向量 80 中的值表示优化解，即它在本征空间中具有表示新说话者的极大似然。

各本征值 $W_1, W_2, \cdots W_n$ 可看作为构成极大似然向量，这里称为极大似然向量。图 5 在 82 处图示出向量。如图示所示，极大似然向量 82 组成本征值 $W_1, W_2, \cdots W_n$ 的集合。

图 6 中示出使用极大似然技术进行适应的过程。来自新说话者

组成观测数据的语音用来构造如 100 处所示的 HMM 集合。然后 HMM 集合 102 用于构成如 104 处所示的超向量。如图所示，超向量 106 构成从 HMM 模型 102 抽取的 HMM 参数的连接的列表。

使用超向量 106，在 108 构造概率函数 Q。当前优选的实施例采用一种概率函数，该函数表示对 HMM 模型 102 的预定集合产生被观测数据的概率。如果函数包含的不只是概率项 P，而且还有这项的对数 logP，则易于进行概率函数 Q 的后继操作。

然后在步骤 110 通过分别对每一本征值 $W_1, W_2, \cdots W_n$ 取概率函数的导数，得到概率函数最大值。例如，如果本征空间维数为 100，这一系统计算概率函数 Q 的 100 个导数，置每一个为零并对各个 W 求解。虽然这好象是很大的计算量，但是比传统的 MAP 或 MLLR 技术进行一般所需的成千次的计算在计算耗费上要小得多。

这样获得的 Ws 结果集合表示标识本征空间中对应于极大似然点的点所需的本征值。这样，Ws 的集合构成本征空间中极大似然向量。就此而言，每一本征向量(图 5 中的本征向量 74、76 和 78)定义了一组正交向量或坐标，本征值乘以该坐标而定义约束在本征空间内的点。在 112 示出的这一极大似然向量用来构造对应于本征空间中最优点(图 4 中的点 66)的超向量 114。然后在步骤 116 超向量 114 可用来构造对新说话者的适应模型 118。

在本发明的极大似然结构的场合中，我们希望使观测 $O=o_1 \cdots o_T$ 的似然关于模型λ最大化。这可通过叠代求辅助函数 Q(以下)的最大值进行，其中λ是叠代处的当前模型，而λ^是估计的模型。我们有：

$$Q(\lambda,\hat{\lambda})=\sum_{\theta\in states} P(O,\theta\mid\lambda)\log$$

作为最初的逼近，我们可希望只对均值进行最大化。在概率 P 由 HMM 集合给出的场合下，我们获得以下结果：

$$Q(\lambda,\hat{\lambda})=const-\frac{1}{2}P(O\mid\lambda)\sum_{\substack{states\\ in\ \lambda}}^{S_\lambda}\sum_{\substack{mixt\\ gauss\\ in S}}^{M_s}\sum_{\substack{time\\ t}}^{T}\left\{\gamma_m^{(s)}(t)[n\log(2\pi)+\log|C_m^{(s)}|+h(o_t,m,s)]\right\}$$

其中:

$$h(o_t, m, s) = (o_t - \hat{\mu}_m^{(s)})^T C_m^{(s)-1} (o_t - \hat{\mu}_m^{(s)})$$

并设:

o_t 为时间 t 处的特征向量

$C_m^{(S)-1}$ 为状态 s 的混合高斯逆协方差

$\mu\hat{}_m^{(S)}$ 为对状态 s 的逼近的适应均值，混合分量 m

$\gamma_m^{(S)}(t)$ 为 P(使用混合高斯 m| $\lambda_s o_t$)

设新说话者的 HMM 的高斯均值位于本征空间中。设这一空间是由均值超向量 $\bar{\mu}_j$ 覆盖的空间，j=1…E,

(原文 P20 公式 1)

$$\bar{\mu}_j = \begin{bmatrix} \bar{\mu}_1^{(1)}(j) \\ \bar{\mu}_2^{(1)}(j) \\ \cdot \\ \cdot \\ \bar{\mu}_m^{(s)}(j) \\ \bar{\mu}_{Ms_\lambda}^{(S_\lambda)}(j) \end{bmatrix}$$

其中 $\bar{\mu}_m^{(s)}(j)$表示在本征向量(本征模型)j 的状态 s 下混合高斯 m 的均值向量。

然后我需要:

$$\hat{\mu} = \sum_{j=1}^{E} w_j \bar{\mu}_j$$

$\bar{\mu}_j$ 为正交的，且 W_j 是我们的说话者模型的本征值。这里我们假设，可对任何新的说话者建模为被观测的说话者的数据库的线性组合。然后

$$\hat{\mu}_m^{(s)} = \sum_{j=1}^{E} w_j \bar{\mu}_m^{(s)}(j)$$

s是M的混合高斯值中的λ、m的状态。

由于我们需要使Q最大化，我们只需设定

$$\frac{\partial Q}{\partial w_e}=0, \qquad e=1..E.$$

(注意，因为本征向量是正交的，故$\frac{\partial w_i}{\partial w_j}=0,\ i\neq j..$)

因而我们有

$$\frac{\partial Q}{\partial w_e}=0=\sum_{\substack{states\\ in\ \lambda}}^{S_\lambda}\sum_{\substack{mixt\\ gauss\\ in S}}^{M_s}\sum_{time\ t}^{T}\{\frac{\partial}{\partial w_e}\gamma_m^{(s)}(t)h(o_t,s)\},\ e=1...E.$$

计算以上的导数，我们有:

$$0=\sum_s\sum_m\sum_t\gamma_m^{(s)}(t)\left\{-\overline{\mu}_m^{(s)T}(e)C_m^{(s)-1}o_t+\sum_{j=1}^{E}w_j\overline{\mu}_m^{(s)T}(j)C_m^{(s)-1}\overline{\mu}_m^{(s)}(e)\right\}$$

由此我们求得线性方程式组

$$\sum_s\sum_m\sum_t\gamma_m^{(s)}(t)\overline{\mu}_m^{(s)T}(e)C_m^{(s)-1}o_t=\sum_s\sum_m\sum_t\gamma_m^{(s)}(t)\sum_{j=1}^{E}w_j\overline{\mu}_m^{(s)T}(j)C_m^{(s)-1}\overline{\mu}_m^{(s)}(e),\ e=1..E.$$

估计本征空间中的接近程度

当把说话者表示为本征空间中的点时，能够使用简单的几何距离计算识别哪一个训练数据说话者最靠近新的说话者。当把说话者表示为本征空间中的分布时，通过把新的说话者数据作为观测O并然后通过测试每一分布候选项(表示训练说话者)估计接近程度，以确定候选项产生观测数据的概率如何。具有最高概率的候选项被估计为具有最接近的程度。在某些高度安全的应用中，如最可能的候选项具有低于预定阈值的概率，可能希望拒绝认证。这样可使用一价值函数区分出缺乏高度确定性的候选项。

如以上所述，估计新的说话者对训练说话者的接近程度可完全在本征空间内进行。另外，可对更高精确性情形使用贝叶斯估计技

术。

为了使用贝叶斯估计强化接近程度的估计，本征空间内训练说话者高斯密度乘以正交互补空间中，表示通过维数降低而被抛弃的说话者数据的估计的边际密度。就此而言，要认识到，对说话者模型进行维数降低的结果是从高维空间向低维空间显著的数据压缩。虽然维数降低保留了大部分重要的基向量，但某些抛弃了某些较高阶的信息。贝叶斯估计技术估计对应于这一被抛弃信息的边际高斯密度。

为了说明，假设原始的本征空间是通过维数降低过程由超向量的线性变换构造的，从而从所有分量较大的数目 N 中抽取 M 个分量。较小的所抽取的 M 个分量表示对应于极大本征值的变换基的较低维子空间。这样，本征空间由分量 i=1…M 定义，其中抛弃的次要分量对应于 i=M+1…N。这两组分量定义了两个相互排斥并互补的子空间，主子空间表示有用的本征空间，而其正交分量表示通过维数降低被抛弃的数据。

我们可以通过以下方程式作为这两个彼此正交的空间中的高斯密度的乘积计算似然估计：

$$\hat{P}(x \mid \Omega) = P_E(x \mid \Omega) * P_{\bar{E}}(x \mid \Omega)$$

在以上方程式中，第一项是本征空间 E 中单一高斯密度，而第二项是与本征空间正交的空间中单一高斯密度。由此得出，只使用到本征空间的投影和残值即可从训练数据向量集合完全估计这两项。

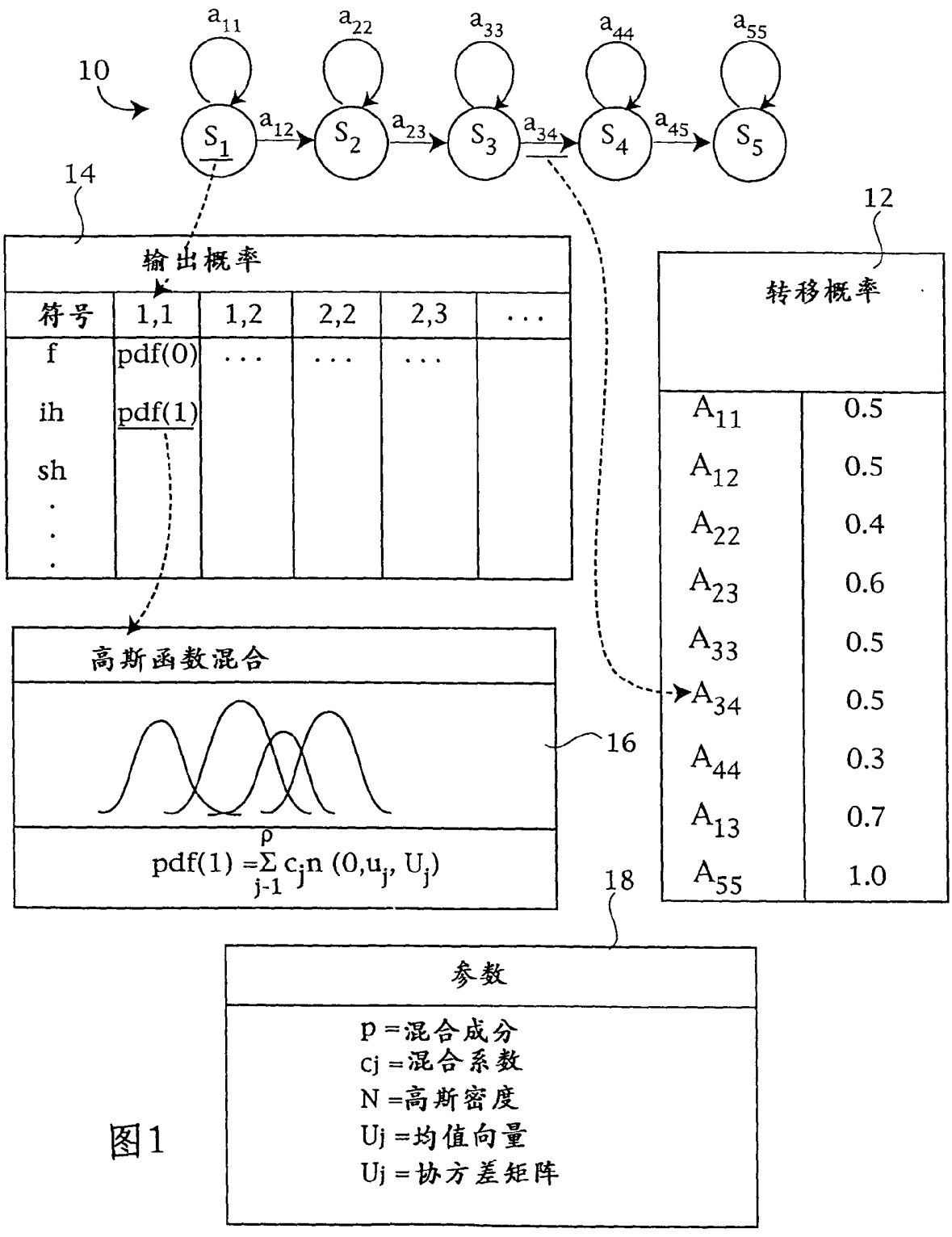

10
a_{11}
a_{22}
a_{33}
a_{44}
a_{55}
a_{12}
a_{23}
a_{34}
a_{45}
S_1
S_2
S_3
S_4
S_5
14
输出概率
符号
1,1
1,2
2,2
2,3
...
f
ih
sh
pdf(0)
pdf(1)
12
转移概率
A_{11} 0.5
A_{12} 0.5
A_{22} 0.4
A_{23} 0.6
A_{33} 0.5
A_{34} 0.5
A_{44} 0.3
A_{13} 0.7
A_{55} 1.0
高斯函数混合
16
$pdf(1) = \sum_{j-1}^{\rho} c_j n(0, u_j, U_j)$
18
参数
p =混合成分
cj =混合系数
N =高斯密度
Uj =均值向量
Uj =协方差矩阵

图1

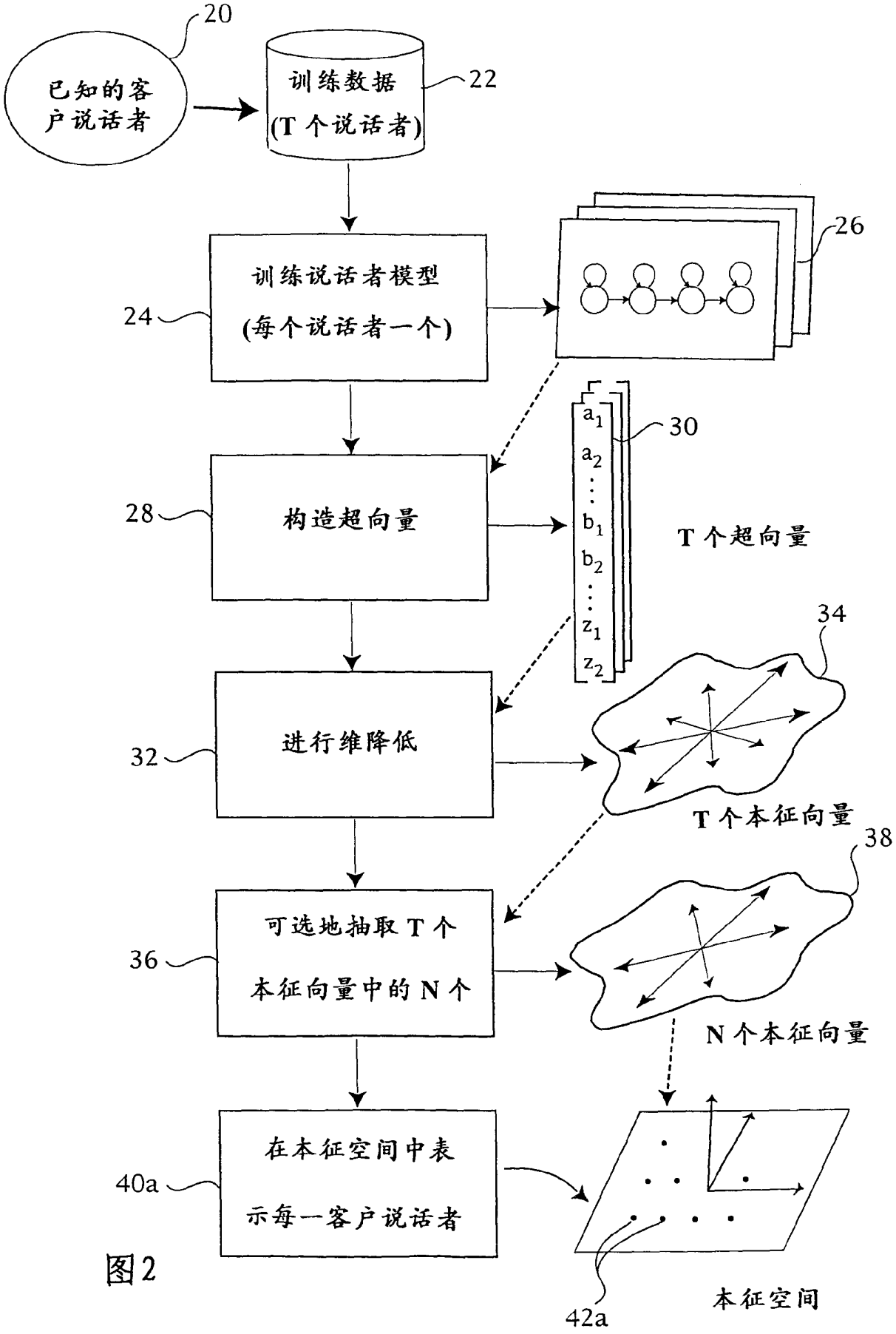

20
已知的客户说话者
训练数据(T个说话者)
22
训练说话者模型(每个说话者一个)
24
26
构造超向量
28
a1
a2
b1
b2
z1
z2
30
T个超向量
进行维降低
32
34
T个本征向量
可选地抽取T个本征向量中的N个
36
38
N个本征向量
在本征空间中表示每一客户说话者
40a
42a
本征空间

图2

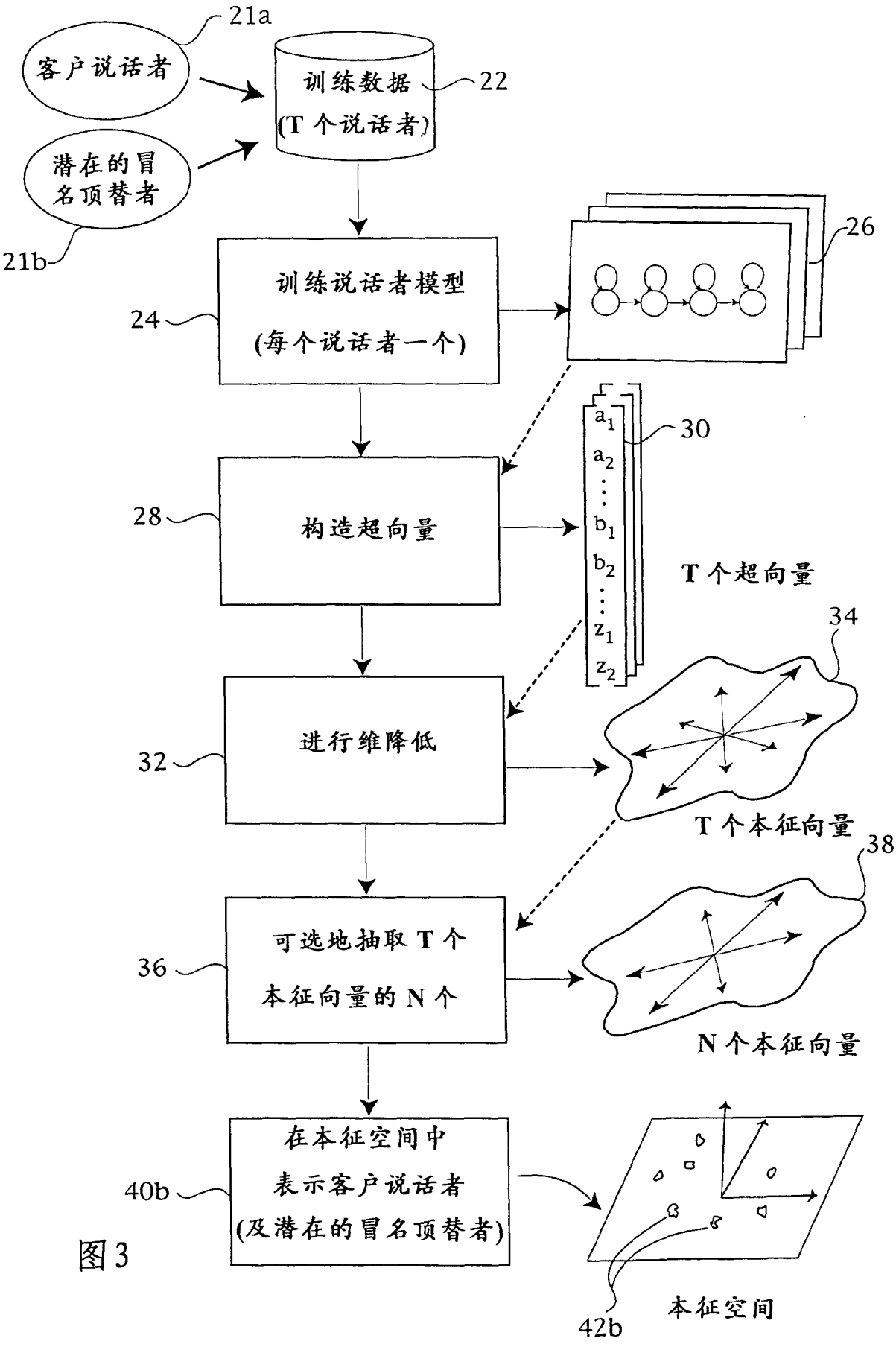

21a
客户说话者
训练数据
(T 个说话者)
22
潜在的冒名顶替者
21b
训练说话者模型
(每个说话者一个)
24
26
构造超向量
28
30
T 个超向量
进行维降低
32
34
T 个本征向量
可选地抽取 T 个
本征向量的 N 个
36
38
N 个本征向量
在本征空间中
表示客户说话者
(及潜在的冒名顶替者)
40b
本征空间
42b

图 3

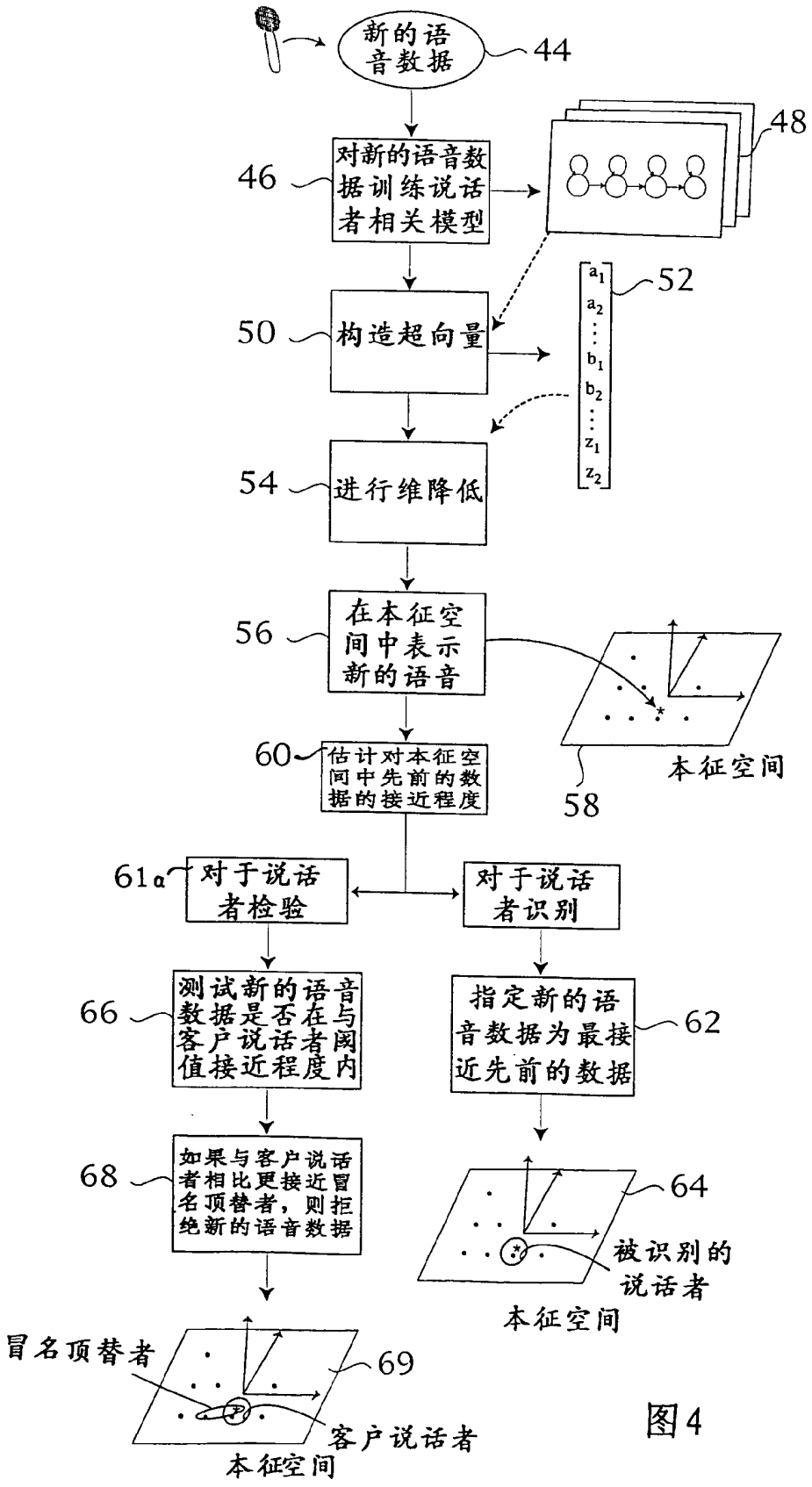

新的语音数据
44
对新的语音数据训练说话者相关模型
46
48
构造超向量
50
52
进行维降低
54
在本征空间中表示新的语音
56
本征空间
58
估计对本征空间中先前的数据的接近程度
60
对于说话者检验
61a
对于说话者识别
测试新的语音数据是否在与客户说话者阈值接近程度内
66
指定新的语音数据为最接近先前的数据
62
如果与客户说话者相比更接近冒名顶替者，则拒绝新的语音数据
68
64
被识别的说话者
本征空间
冒名顶替者
69
客户说话者
本征空间

图4

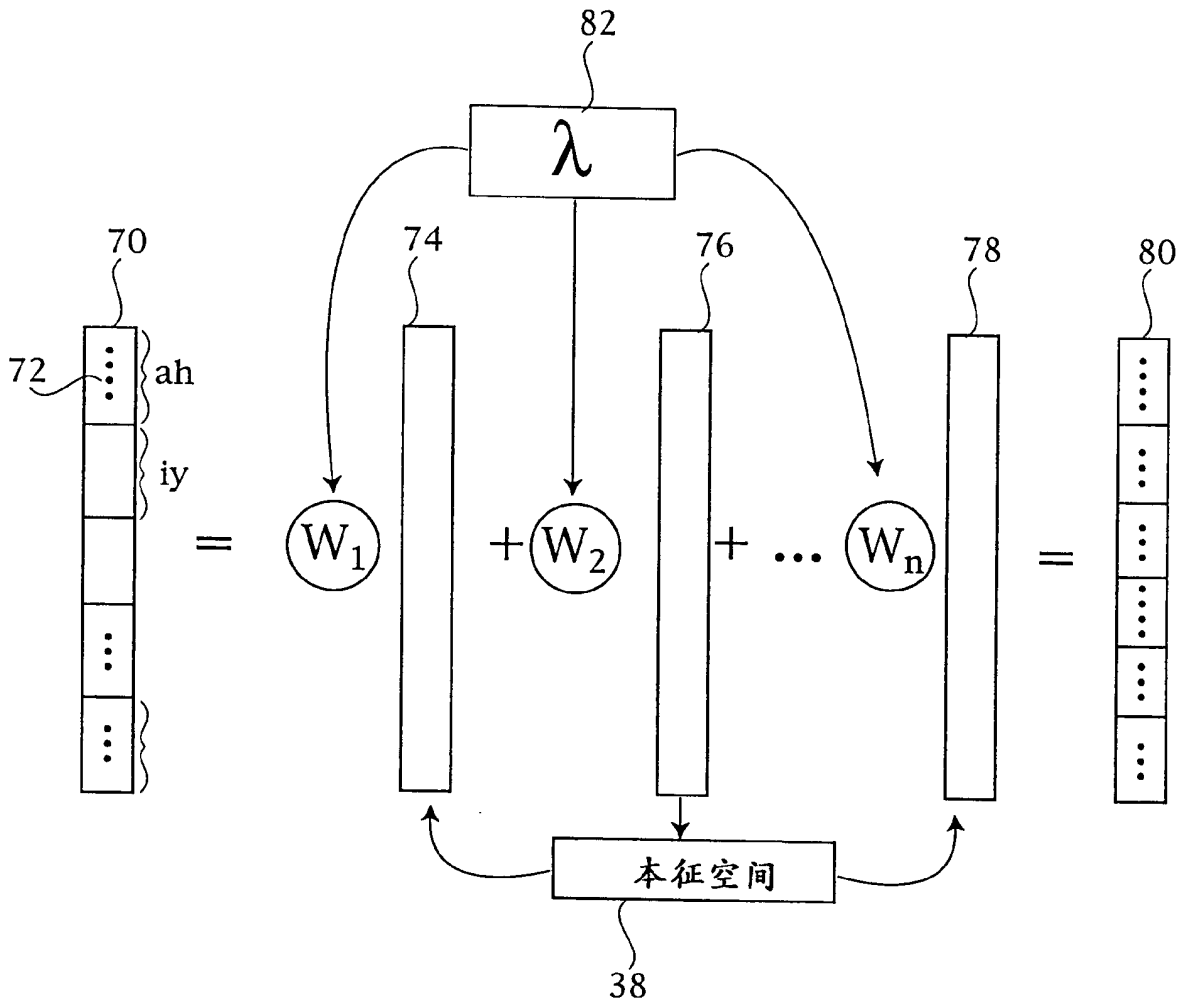


图5

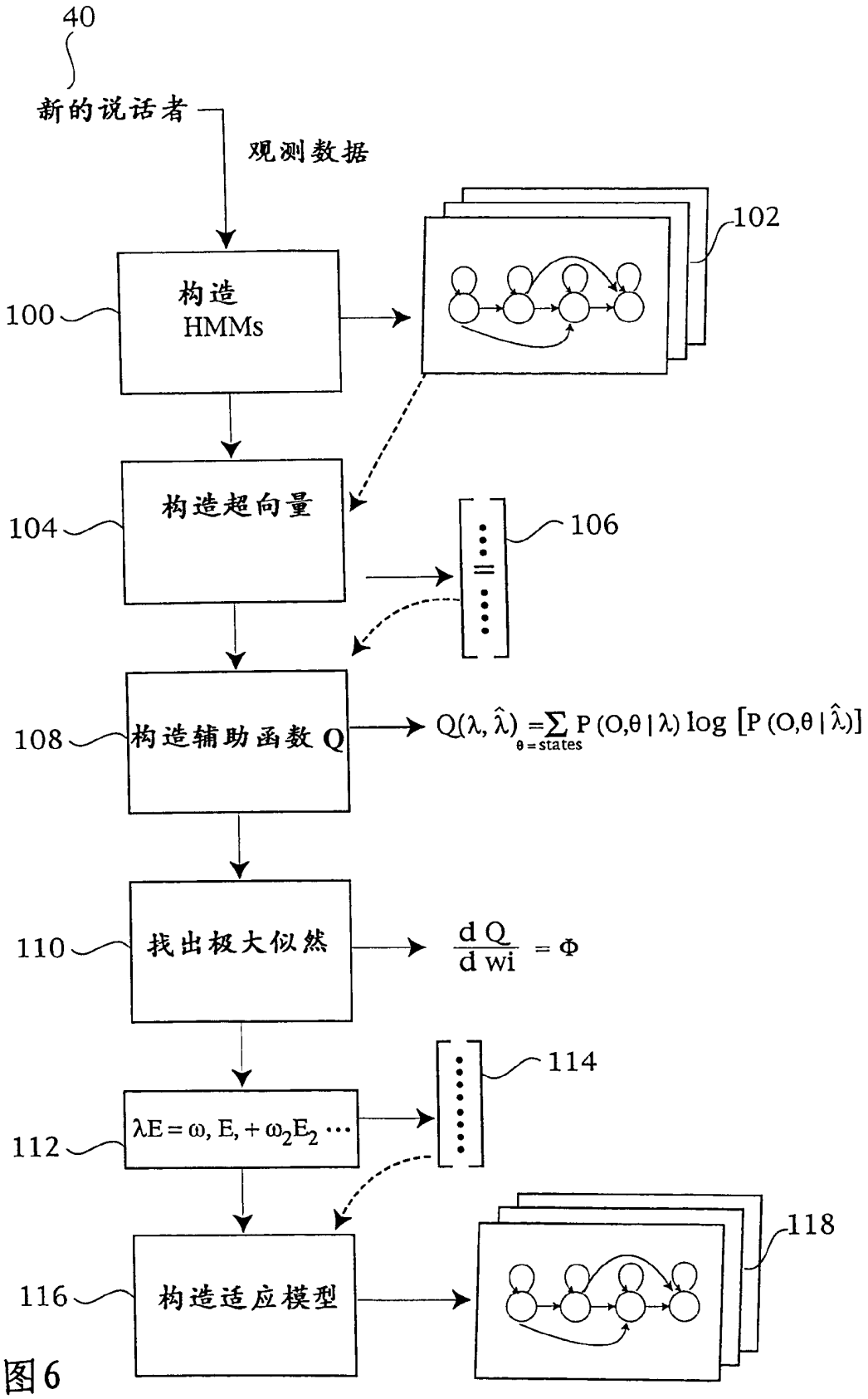

40
新的说话者
观测数据
构造
HMMs
100
102
构造超向量
104
106
构造辅助函数 Q
108
$Q(\lambda,\hat{\lambda}) = \sum_{\theta=\text{states}} P(O,\theta|\lambda)\log[P(O,\theta|\hat{\lambda})]$
找出极大似然
110
$\frac{dQ}{d\,wi} = \Phi$
$\lambda E = \omega, E, + \omega_2 E_2 \cdots$
112
114
构造适应模型
116
118

图6