

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G10L 15/00

G10L 15/20



[12] 发明专利申请公开说明书

[21] 申请号 03120433.3

[43] 公开日 2003 年 10 月 1 日

[11] 公开号 CN 1445752A

[22] 申请日 2003.3.14 [21] 申请号 03120433.3

[30] 优先权

[32] 2002. 3. 15 [33] US [31] 10/099,305

[71] 申请人 松下电器产业株式会社

地址 日本大阪府

[72] 发明人 扬尼斯·索尔米 路卡·里加兹奥
帕特里克·尼古伊恩
让-克劳德·詹卡

[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

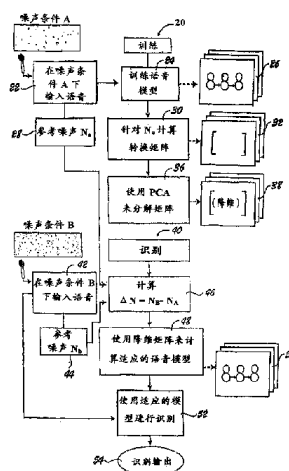
代理人 吴丽丽

权利要求书 5 页 说明书 21 页 附图 8 页

[54] 发明名称 用于特征域中信道与加性噪声联合
补偿的方法与装置

[57] 摘要

一种方法，用于对输入到一个语音识别系统的目标语音信号进行噪声适应，所述目标语音信号含有加性噪声和卷积噪声，本方法包括：在所述目标语音信号中，对加性噪声偏差和卷积噪声偏差作出估计；以及在特征域中，对目标语音信号中的加性噪声偏差和卷积噪声偏差进行联合补偿。



1. 一种方法，用于对输入到一个语音识别系统的目标语音信号进行噪声适应，所述目标语音信号含有加性噪声和卷积噪声，所述方法包括：

在所述目标语音信号中，对加性噪声和卷积噪声作出估计；以及在特征域中，对目标语音信号中的加性噪声偏差和卷积噪声偏差进行联合补偿。

2. 根据权利要求1所述的方法，其中，所述目标语音信号包括多个帧，并且其中，所述对卷积噪声偏差进行估计包括

确定部分地适应于在加性噪声中的测试环境的语音信号；以及利用该部分地适应的语音信号估计卷积噪声偏差。

3. 根据权利要求2所述的方法，其中，所述确定部分地适应于在加性噪声中的测试环境的语音信号的步骤包括确定如下式所示的一阶近似：

$$\begin{aligned} C(H_{tar} \cdot S + N_{ref}) &= C(H_{tar} \cdot S + N_{tar}) + \left. \frac{\partial C(H_{tar} \cdot S + N)}{\partial N} \right|_{N=N_{tar}} (N_{ref} - N_{tar}), \\ &= C(H_{tar} \cdot S + N_{tar}) + J_N \Delta N, \end{aligned}$$

其中：

$C()$ 为倒频谱运算符；

H_{tar} 为针对测试环境的信道频率响应；

N_{ref} 为针对训练环境的加性噪声；

N_{tar} 为针对测试环境的加性噪声；

S 表示目标语音信号的一种无噪声样式；

J_N 为 $\left. \frac{\partial C(H_{tar} \cdot S + N)}{\partial N} \right|_{N=N_{tar}}$ ，一个噪声补偿矩阵；

ΔN 为 $(N_{ref} - N_{tar})$ ；以及

$C(H_{tar} \cdot S + N_{ref})$ 为所述部分适应的语音信号。

4. 根据权利要求3所述的方法还包括确定如下式所示的噪声补

偿矩阵:

$$J_N = F' \frac{1}{H_{tar} \cdot S + N_{tar}}$$

式中, F' 为一个解相关转换, $H_{tar} \cdot S + N_{tar}$ 为带噪声的目标语音信号。

5. 根据权利要求 3 所述的方法, 其中, 用以估计所述目标语音信号的所述帧之一的卷积噪声偏差的所述部分适应的语音信号是从一个先前的所述信号帧确定的一个部分适应的语音信号。

6. 根据权利要求 5 所述的方法, 其中, 所述估计卷积噪声偏差的步骤还包括确定在频谱域中对应于所述目标语音信号的一个模型 $H_{ref} \cdot S + N_{ref}$, 其中, H_{ref} 为针对训练环境的信道频率响应。

7. 根据权利要求 6 所述的方法, 其中, 所述确定在频谱域中对应于所述目标语音信号的一个模型的步骤包括利用 Viterbi 解码来估计一个最可能的部分路径的步骤。

8. 根据权利要求 7 所述的方法, 其中, 所述卷积噪声偏差 $\Delta H = H_{ref} - H_{tar}$ 被确定为:

$$\Delta H = J_P^{-1} [C(H_{ref} \cdot S + N_{ref}) - C(H_{tar} \cdot S + N_{ref})]$$

$$\text{其中: } J_P^{-1} \text{ 是一个逆矩阵, 在其中, } J_P = \frac{\partial C(H \cdot S + N_{ref})}{\partial H} \Big|_{H=H_{ref}}$$

$$= F' \frac{S}{H_{ref} \cdot S + N_{ref}};$$

并且 F' 为一个解相关转换。

9. 根据权利要求 8 所述的方法, 其中, F' 为离散余弦变换。

10. 根据权利要求 2 所述的方法, 其中, 用以估计所述目标语音信号的所述帧之一的卷积噪声偏差的所述部分适应的语音信号是从一个先前的所述信号帧确定的一个部分适应的语音信号。

11. 根据权利要求 10 所述的方法, 其中, 所述估计卷积噪声偏差的步骤还包括确定在频谱域中对应于所述目标语音信号的一个模型的步骤。

12. 根据权利要求 11 所述的方法, 其中, 所述确定在频谱域中对应于所述目标语音信号的一个模型的步骤包括利用 Viterbi 解码来

估计一个最可能的部分路径的步骤。

13. 根据权利要求 12 所述的方法还包括利用倒频谱运算符 $C^{-1}(\cdot)$ 从倒频谱域中的一个模型来确定在频谱域中的所述模型的步骤, 其中, $C^{-1}(x) = \exp(F^{-1}x)$, F 是一个矩阵, F' 是 F 的解相关转换和逆共轭矩阵, 并且 $FF' = I$, 其中 I 是一个单位矩阵。

14. 一个语音识别装置, 包括对输入到所述装置的一个目标语音信号的噪声适应, 所述目标语音信号含有加性噪声和卷积噪声二者, 所述装置被配置为:

对在所述目标语音信号中的加性噪声偏差和卷积噪声偏差作出估计; 以及

在特征域中, 联合地补偿目标语音信号中的加性噪声偏差和卷积噪声偏差。

15. 根据权利要求 14 所述的装置, 其中, 所述目标语音信号包括多个帧, 并且其中为对一个卷积噪声偏差进行估计, 所述装置被配置为:

确定部分地适应于在加性噪声中的测试环境的语音信号; 以及
利用部分地适应的语音信号估计卷积噪声偏差。

16. 根据权利要求 15 所述的装置, 其中, 为确定部分地适应于在加性噪声中的测试环境的语音信号, 所述装置被配置为确定如下式所示的一阶近似:

$$\begin{aligned} C(H_{tar} \cdot S + N_{ref}) &= C(H_{tar} \cdot S + N_{tar}) + \left. \frac{\partial C(H_{tar} \cdot S + N)}{\partial N} \right|_{N=N_{tar}} (N_{ref} - N_{tar}), \\ &= C(H_{tar} \cdot S + N_{tar}) + J_N \Delta N, \end{aligned}$$

其中:

$C(\cdot)$ 为倒频谱运算符;

H_{tar} 为针对测试环境的信道频率响应;

N_{ref} 为针对训练环境的加性噪声;

N_{tar} 为针对测试环境的加性噪声;

S 表示目标语音信号的一种无噪声样式;

J_N 为 $\frac{\partial C(H_{tar} \cdot S + N)}{\partial N} |_{N=N_{tar}}$, 一个噪声补偿矩阵;

ΔN 为 $(N_{ref} - N_{tar})$; 以及

$C(H_{tar} \cdot S + N_{ref})$ 为所述部分适应的语音信号。

17. 根据权利要求 16 所述的装置还被配置为确定如下式所示的噪声补偿矩阵:

$$J_N = F' \frac{1}{H_{tar} \cdot S + N_{tar}}$$

式中, F' 为一个解相关转换, $H_{tar} \cdot S + N_{tar}$ 为带噪声的目标语音信号。

18. 根据权利要求 16 所述的装置, 其中, 用以估计所述目标语音信号的所述帧之一的卷积噪声偏差的所述部分适应的语音信号是从一个先前的所述信号帧确定的一个部分适应的语音信号。

19. 根据权利要求 18 所述的装置, 其中, 为估计卷积噪声偏差, 所述装置还被配置为确定在频谱域中对应于所述目标语音信号的一个模型 $H_{ref} \cdot S + N_{ref}$, 其中, H_{ref} 为针对训练环境的信道频率响应。

20. 根据权利要求 19 所述的装置, 其中, 为确定在频谱域中对应于所述目标语音信号的一个模型, 所述装置被配置为利用 Viterbi 解码来估计一个最可能的部分路径。

21. 根据权利要求 20 所述的装置被配置为将所述卷积噪声偏差 $\Delta H = H_{ref} - H_{tar}$ 确定为:

$$\Delta H = J_P^{-1} [C(H_{ref} \cdot S + N_{ref}) - C(H_{tar} \cdot S + N_{ref})]$$

其中:

$$J_P^{-1} \text{ 是一个逆矩阵, 在其中, } J_P = \frac{\partial C(H \cdot S + N_{ref})}{\partial H} |_{H=H_{ref}} = F' \frac{S}{H_{ref} \cdot S + N_{ref}};$$

并且 F' 为一个解相关转换。

22. 根据权利要求 21 所述的装置, 其中, F' 为离散余弦变换。

23. 根据权利要求 22 所述的装置, 其中, 用以估计所述目标语音信号的所述帧之一的卷积噪声偏差的所述部分适应的语音信号是从

一个先前的所述信号帧确定的一个部分适应的语音信号。

24. 根据权利要求 23 所述的装置, 其中, 为估计卷积噪声偏差, 所述装置还被配置为确定在频谱域中对应于所述目标语音信号的一个模型。

25. 根据权利要求 24 所述的装置, 其中, 为确定在频谱域中对应于所述目标语音信号的一个模型, 所述装置被配置为利用 Viterbi 解码来估计一个最可能的部分路径。

26. 根据权利要求 25 所述的装置还被配置为利用倒频谱运算符 $C^{-1}(\cdot)$ 从倒频谱域中的一个模型来确定在频谱域中的所述模型, 其中, $C^{-1}(x) = \exp(F^{-1}x)$, F 是一个矩阵, F' 是 F 的解相关转换和逆共轭矩阵, 并且 $FF' = I$, 其中 I 是一个单位矩阵。

用于特征域中信道与 加性噪声联合补偿的方法与装置

相关申请的交叉参考

本申请是 2000 年 7 月 31 日提交的美国专利申请第 09/628,376 号的继续申请，它本身又是 2000 年 4 月 18 日提交的美国专利申请第 09/551,001 号的继续申请，二者都作为参考文献，以全文的形式被收入本文。

技术领域

本发明涉及自动语音识别系统。更具体地说，本发明涉及一种技术，用于适应识别器，使之在出现噪声时能更好地进行工作。

背景技术

当前的自动语音识别系统在实验室条件下工作得很好，但是当在现实世界应用项目中使用时，很快就会出现性能下降。在现实世界应用项目中影响识别器性能的一个重要因素就是损坏语音信号的环境噪声的存在。为了解决噪声问题，人们已经研究开发了多种方法，诸如频谱相减法或并行模型组合法。然而，这些解决方案不是过于局限，就是计算开销太大。

最近，已经提出一种 **Jacobian** 适应方法，用以处理在噪声从噪声 A 改变为噪声 B 的条件下的加性噪声。例如，授予 Yamaguchi 的美国专利第 6,026,359 号就描述了这样一种方案，它在模式识别中，根据存储一个表示模型参数的 **Taylor** 展开式的 **Jacobian** 矩阵来进行模型适应。然而，为了使这个模型工作得很好，必须令噪声 A 和噪声 B 在特性和电平方面互相接近。例如，当噪声 A 是在以时速 30 英里行驶在平滑路面上的一辆给定的汽车的乘客车厢内测量时，而噪声 B

也具有类似的特性,例如在以时速 45 英里行驶在相同路面上的相同汽车内测量的噪声时, **Jacobian** 适应技术可能工作得很好。

当噪声 A 和噪声 B 彼此相差甚远时,例如当噪声 A 是在以时速 30 英里行驶在平滑路面上的一辆给定的汽车的乘客车厢内测量、而噪声 B 是在关上窗口的汽车或者以时速 60 英里行驶时,已知的 **Jacobian** 适应技术就开始失灵了。

由于在训练阶段,通常难以预测在测试阶段(当系统在使用中时)可能出现的噪声,所以已知的 **Jacobian** 适应技术的这种缺点限制了在许多实际的应用场合中的用途。同样,由于所需的计算开销(处理时间和/或对存储器的需求)使得它们行不通,所以在许多实际的应用场合中, **Jacobian** 适应技术的改进都受到限制。

另一方面的考虑涉及卷积噪声的补偿。由于卷积噪声来自语音信道,所以就能将卷积噪声从上述加性噪声中区分出来。例如,从讲话者到传声器之间的距离发生改变,传声器的不够完善,甚至在其上传输信号的电话线全都对卷积噪声作出贡献。另一方面,加性噪声典型地来自讲话者在其中讲话的环境。

卷积噪声的一个重要特性就是,当语音信号处于频谱域之中时,它是相乘的,而加性噪声在频谱域之中是相加的。这些特性给噪声补偿带来特殊的困难。事实上,大多数常规的方案仅处理卷积噪声或加性噪声其中之一,而不是同时处理二者。

发明内容

因此,本发明的一种配置提供了一种对输入到一个语音识别系统的目标语音信号进行噪声适应的方法,上述目标语音信号含有加性噪声和卷积噪声二者。本方法包括在目标语音信号中对加性噪声偏差以及卷积噪声偏差作出估计;以及在特征域(feature domain)中,对目标语音信号的卷积噪声偏差和加性噪声偏差进行联合补偿。

本发明的另一种配置提供了一种语音识别装置,它包括对输入到该装置的目标语音信号的噪声适应,上述目标语音信号含有加性噪声

和卷积噪声二者。本装置被配置为对在目标语音信号中的加性噪声偏差以及卷积噪声偏差作出估计；以及在特征域中，对目标语音信号的卷积噪声偏差和加性噪声偏差进行联合补偿。

本发明的这些配置不易受到非静止噪声和信道的影响，在每一次迭代中，需要修改的参数很少，并且仅修改信号的特征，而不是所有的模型特征。因此，本发明的这些配置使用较小的存储器（例如，随机存取存储器或 RAM），这对于小型的嵌入式系统来说是有利的。而且，由于不修改模型的特征，所以已适应的模型的特征不需要存储在存储器之中。

从下文所提供的详细说明中，本发明的进一步的适用性将变得更加明显。应当理解，在表示本发明的优选实施例的同时，作者打算将详细的说明以及特定的实例仅用于说明目的，并且不打算用来限制本发明的范围。

附图说明

通过详细的说明以及诸附图，将使本发明变得更易于理解。

图 1 以图解方式表示不同的噪声条件，这对理解本发明是有用的。

图 2 是在训练和识别阶段的数据流图，说明改进的转移矩阵适应的一个当前的优选实施例。

图 3 是对数频谱图，该图将常规的 Jacobian 适应跟并行模型组合（PMC）适应进行比较。

图 4 和 5 是 α 适应参数曲线，表示不同的 α 数值对识别精度的影响。

图 6 是本发明的噪声适应系统的一种配置的方框图。

图 7 是本发明的特征 Jacobian 适应模块的一种配置的方框图。

图 8 是使用图 7 的特征 Jacobian 适应模块的一个卷积噪声模块的一种配置的方框图。

具体实施方式

以下关于优选实施例的说明从性质上来说仅仅是示例性的，并且无论如何作者都不打算用来限制本发明及其应用或用途。

本发明的至少一种配置可应用于在图 1 中所图解的问题。如在 10 中所示，假定，例如，自动语音识别系统必须在一个有噪声的环境中进行工作，例如处于行驶中的机动车的乘客车厢里面。随着车辆的加速，在乘客车厢里面测得的噪声水平典型地从噪声 A 增加到噪声 A'。虽然噪声水平典型地从噪声 A 增加到噪声 A'，但是噪声的特性或性质在很大程度上保持相同。在一部正在行驶的汽车中，例如，随着汽车的加速，噪声频谱典型地以一种可预测的方式发生改变。风的噪声在幅度上将有所增加，但在很大程度上保持其随机白噪声或粉噪声的特性。路面噪声（轮胎在路面上滚动的声音）在频率上的增加正比于噪声的增加。

不幸的是，在许多现实世界的应用中，环境噪声的特性和性质不可能事先预测，像图 1 的 10 中所隐含的情况那样。例如，考虑移动蜂窝电话。蜂窝电话可以应用于行驶中的车辆，在那里，它经受着在 10 中所示的噪声范围；或者它可以被应用于充满完全不同的交通声音的一个街角处；或者它可以被应用于具有完全不同的噪声性质的一家商厦。在图 1 的 12 中，示出了不同噪声性质的宽广的多样性，在其中，3 种不同的噪声模式以图解方式被描绘为噪声 A，噪声 B 和噪声 C。对于必须在这些变化的噪声环境中进行工作的自动语音识别系统来说，噪声性质的不可预测性提出了严重的挑战。

图 2 表示本发明在语音识别应用中的一个示例性的实施例。语音识别应用程序使用一个基于模型的识别器。该模型是在训练阶段中被研发出来的，并且在随后的识别阶段中被使用。在图 2 中，系统的训练阶段一般地被表示为 20，识别阶段被表示为 40。在训练阶段中，在噪声状态 A 下提供输入语音，如 22 所示。输入语音被用来对语音模型进行训练，如步骤 24 所示，示例性的语音模型以图解方式示于 26。在典型的输入语音信号中，将会有不出现语音的时间，例如在语音开

始之前或语音结束之后。这些没有语音的部分可以被用来记录表示跟噪声状态 A 有关的参考噪声 N_a 的数据。在图 2 中, 参考噪声 N_a 被存储在方块 28 之中。必要时, 可以使用用以建立语音模型 26 的相同的训练操作来建立该噪声的模型 (背景模型)。

在对语音模型进行训练之后, 在步骤 30, 计算一组转换矩阵, 并将其存储在 32 处。在识别阶段, 这些矩阵被用来适应语音模型, 因此, 在识别阶段中, 在存在噪声状态的情况下, 他们将表现得更好。基本的 Jacobian 适应过程假定在识别阶段的噪声性质大致上跟训练阶段相同。否则, 古典的 Jacobian 适应可能产生不出最佳的结果。

改进的适应技术基于使用针对初始噪声状态 N_a 而产生的一组转换矩阵。通过向一组针对初始噪声状态 N_a 而开发的 Jacobian 矩阵实施线性和非线性变换, 来计算转换矩阵。现时的优选实施例通过实施在下一节中提出的 α 适应因子, 来进行 (线性和非线性) 变换。尽管 α 适应因子当前是优选的, 但还可以使用神经网络或者其他人工智能部件来实现这种变换。

矩阵的操作可能是一种计算开销很大的处理过程。一个值得考虑的成本因素就是为存储所有的矩阵所需的存储器空间。在一个典型的实施例中, 在词典中针对每一个条目的语音模型可以使用多个隐藏马尔柯夫模型状态, 其中的每一种状态又跟多种高斯密度有关。因此, 对于每一种状态中的每一种高斯密度, 都有一个矩阵。这将导致需要存储数百个矩阵。

该优选实施例执行了矩阵分解步骤 36, 以产生 Jacobian 矩阵 38 的一个降低复杂性的 (reduced-complexity) 集合, 如将在下面充分讨论的那样, 现行的优选分解技术使用主成分分析 (PCA) 来构建降低复杂性的转换矩阵。

在识别阶段, 在步骤 42 中提供来自用户的输入语音。该输入语音与一个如步骤 44 所示的噪声条件 B (也可被称为目标噪声 N_b) 相关联。如前所述, 若噪声条件 B 跟在训练阶段所使用的噪声条件 A 在性质上不同, 则常规的 Jacobian 适应技术就不能产生最好的结果。然

而，我们发现 α -适应过程（当在训练阶段中定义了 Jacobian 矩阵时使用）大大地改进了在不利的噪声条件下的识别结果。在下文所示的实例中，提供了我们的测试结果。

在图 2 的步骤 44 中所示的目标噪声 N_b 是从输入语音 42 中提取出来的，并且被用来计算其与参考噪声 N_a 的差值，如步骤 46 所示。利用这个噪声差值以及在训练阶段开发的降低了复杂性的转换矩阵来计算新的适应语音模型，如步骤 48 所示。然后在步骤 52 中，使用所得到的适应语音模型 50，对输入语音 42 进行语音识别，以便提供识别结果 54。

α -适应

为了更好的理解我们的转换矩阵是如何工作的，有必要先了解常规的 Jacobian 适应。常规的 Jacobian 适应是与另外一种形式的适应——并行模型组合（PMC）相关的。从传统上来说，Jacobian 适应被用作 PMC 的一种近似，以便降低 PMC 所要求的计算负荷。由于对语音模型的每一种密度来说，必须将平均矢量转换到频谱域，所以使用 PMC 的计算开销非常大。然后，在将平均矢量添加到目标噪声之后，结果矢量必须再转换到倒频谱域（cepstral domain）。这种双重转换使用矩阵乘法运算和双非线性函数，对嵌入式系统来说通常是很耗费的。

从传统上来说，Jacobian 适应在倒频谱域中被用作 PMC 的一种近似。为了比较的目的，再现于下面的等式（1）描述了 PMC 计算，其中大写字母 F 代表离散余弦变换（DCT）的矩阵。等式（2）代表常规的 Jacobian 适应计算，被用来作为开销更大的 PMC 计算的一种近似。

$$C(S+N) = F \cdot \log(\exp(F^{-1} \cdot C(S))) + \exp(F^{-1} \cdot C(N)) \quad (1)$$

$$\Delta C(S+N) = \frac{\partial C(S+N)}{\partial C(N)} \cdot \Delta C(N) = F \cdot \frac{N}{S+N} \cdot F^{-1} \cdot \Delta C(N) \quad (2)$$

Jacobian 适应近似法的问题在于它仅在目标噪声（在使用系统时出现）在性质上跟参照噪声（在系统进行训练时出现）相似时才成立。图 3 分析了问题的本质，图中描绘了在对数频谱域中，当噪声增大时

带有噪声的语音的参数估计。具体地说，这些图表示了对数频谱域中，PMC 适应跟常规的 Jacobian 适应是如何进行比较的。图 3 中左边的阴影区域对应于语音信号远远大于噪声时的状况，而右边的区域则对应于噪声远远大于语音信号的状况。如果训练和测试环境处于这两个区域的相同区域中，那么 Jacobian 适应就跟 PMC 计算相似。然而，如果两个环境的其中之一处于中间区域，或者它处于另一个环境以外的其它区域，那么 Jacobian 适应就不同于 PMC，而且实际上通常低估了模型的适应。

我们已经发现，通过对 Jacobian 矩阵的线性或非线性转换，可以使常规的 Jacobian 适应得到重大的改进。为了实现线性或非线性转换，本优选实施例使用了一个称为 α —适应参数的参数。下面的等式 (3) 说明了优选使用 α —适应参数来实现非线性转换。等式 (4) 表示参数的另一种使用法，用以实现线性转换。如上面所提到的，虽然目前优选了 α —适应参数来实现转换，但采用其他的转换技术也是可能的。例如，可以使用神经网络或者其他人工智能部件来转换针对初始噪声条件的 Jacobian 矩阵。另一种转换技术涉及将第一 α —适应参数或因子应用到输入的语音上，同时将第二 α —适应参数或因子应用到噪声上。其他的变动也是可能的。

$$\Delta C(S+N) = F \cdot \frac{\alpha N}{S+\alpha N} \cdot F^{-1} \cdot \Delta C(N) \quad (3)$$

$$\text{JA: } \Delta C(S+N) = \frac{\partial C(S+N)}{\partial C(N)} \cdot \alpha \Delta C(N) \quad (4)$$

参照等式 (3)， α —适应参数的功能如下。若参考噪声接近于 0，而且 α 的数值不是太大，则两条切线（分别在 X 坐标 N 和 αN 处进行计算）就都是水平的。若参考噪声非常重要，则两条切线将对应于直线 $y = x$ 。若参考噪声属于在图 3 中的中央区域，则切线的新斜率将大于常规的 Jacobian 适应曲线所产生的斜率。

等式 (3) 中 α —适应参数的使用导致矩阵的非线性转换。分子与分母都与该参数相乘。这样就产生一种非线性转换效果。在等式 (4) 中， α —适应参数与所得到的分子和分母的商相乘，这样就产生了一

种线性转换效果。

在两种情况下， α —适应参数的主要效果都是为了增大被添加到语音模型上的适应偏差。由于它纠正了常规的 Jacobian 适应低估噪声影响的缺点，所以这是有用的。在本文件的后面部分我们将陈述我们的实验结果，该结果表示使用 α —适应参数可能得到的改进。

α —适应参数的选择

从理论上来说， α —适应参数的最佳值依环境而定。当目标噪声和参考噪声之间的失配较大时， α 的数值也应当是较大的。然而，我们已经发现， α —适应参数比理论上所预测的要稳定得多。当使用 α —适应参数代替常规的 Jacobian 适应来产生转换矩阵时，对于小的 α 数值来说，语音识别的精度的变化很小，当 α 的数值为中等时，精度的变化有所增加，而当 α 的数值超过一定的点之后，精度的变化又再次降低。这种现象的产生归因于图 3 中曲线的形状。具体地说，不管 α 的数值如何，切线的斜率仅在 0 和 1 之间发生变化。

为了说明这一点，我们已经完成了一组在不利环境中进行的数字识别的实验。首先构建 12 种与环境无关的数字模型：数字从 1 到 9 加上用于“0”和“零”的模型，再加上一个用于无声的模型。无声的情形由具有 5 种状态的隐藏马尔科夫模型（HMM）来建模。剩下的模型使用了 15 种状态。所有 HMM 的每一种状态都使用四种高斯密度。用来训练模型的训练集合包括 3803 个数字序列，由 80 名讲话者来进行朗读。训练集合在没有噪声的实验室条件下进行录制。图 4 和图 5 表示当 α 的数值在 1 到 4 之间变化时精度的改变。该数据是在 6 种不同的声学环境的基础上产生的：

在无噪声条件下记录的验证资料。

相同的资料，但添加了汽车噪声，信噪比为 10 dB。

相同的资料，但添加了汽车噪声，信噪比为 0 dB。

相同的资料，但添加了白噪声，信噪比为 15 dB。

测试资料，在一部时速为 30 英里的汽车中记录。

另一份测试资料，在一部时速为 60 英里的汽车中记录。

参照图 4 和 5, 就能看出不管声学环境如何, 在 $\alpha = 2.4$ 至 $\alpha = 3.6$ 的范围内, 针对不同的 α 数值的精度变化是很小的。这表明 α 有一个适于在本发明的一个实际的实施例中使用的稳定的范围。尽管我们现在优选一个介于大约 2.4 至 3.6 之间的 α 适应参数, 但应当理解, 这仅代表一种可能的稳定的范围。一般来说, 使用其他的 α 数值也可能得到有利的结果。换句话说, 在 α 的真正的“最佳”值以及在一个被考虑的范围内(例如 2.4 至 3.6)可选择的 α 的任何其他数值之间, 精度的降低是很小的。我们的数据表明, 从“最佳”点算起的精度降低小于 3 %。这使得我们的改进的 Jacobian 适应成为一种十分牢靠的方法。

为了减小计算开销而降低维数

如上所述, 虽然 Jacobian 适应的计算开销比 PMC 来得小, 但是它仍然成为识别系统的一种相当沉重的负担, 特别是对嵌入式系统来说, 尤其是这样。

确实, 我们已经看到, 每一个转换矩阵都可以表示为下列的等式 (5):

$$\frac{\partial C(S+N)}{\partial C(N)} = F \cdot \frac{S+\alpha N}{\alpha N} \cdot F^{-1} \quad (5)$$

式中, $\frac{S+\alpha N}{\alpha N}$ 是一个具有维数 $N_{filt} \times N_{filt}$ 的对角线矩阵, 其中, N_{filt} 是在频谱滤波器组中所使用的滤波器的数目。

因此, 每一个转换矩阵都可以被表示为 N_{filt} 个正则矩阵的加权和, 正则矩阵事实上就是 Jacobian 矩阵所属的空间的基 (base)。这些正则阵由 $J_i = F \cdot \text{diag}(i) \cdot F^{-1}$ 给出定义, 式中, $\text{diag}(i)$ 指的是一个 $N_{filt} \times N_{filt}$ 的对角线矩阵, 在该矩阵的位置 i 处的矩阵元素为 1, 而在其余各处, 矩阵元素均为 0。

因此, 每一个转换矩阵都可以被表示为:

$$\frac{\partial C(S+N)}{\partial C(N)} = \sum_{i=1}^{N_{filt}} \gamma_i \cdot J_i \quad (6)$$

这样一来, 取代存储 N_d 个矩阵 (这里 N_d 在所有的语音模型中

的密度的总数)，存储 N_{filt} 个正则矩阵加上 N_d 乘以 N_{filt} 系数 γ_i 就足够了。这相当可观地降低了对存储器的需求。

然而，由于这种解决方案增加了算法的时间复杂性，所以对它还可以作出进一步的改进。确实，当所有的转换矩阵都被存储时，等式 (2) 就可以直接地应用于所有的密度，共需进行 N_d 次矩阵乘法运算。

若选择第二解决方案，则等式 (2) 的右半部分变为：

$$\frac{\partial C(S+N)}{\partial C(N)} \cdot \Delta C(N) = \left(\sum_{i=1}^{N_{filt}} \gamma_i \cdot J_i \right) \cdot \Delta C(N) = \sum_{i=1}^{N_{filt}} \gamma_i \cdot (J_i \cdot \Delta C(N)).$$

在这个等式中，代价是进行 N_{filt} 次矩阵加法运算，以及 N_{filt} 次矩阵乘法（乘以标量）运算：这种运算应当针对每一种密度重复地进行。因此，总的代价是 $2 * N_d * N_{filt}$ 次矩阵运算。

若我们不希望使用额外的计算时间，则必须减少正则矩阵的数目。

目前，用以降低一个空间的维数的优选技术就是对属于该空间的元素集合实行主成分分析。因此我们首先必须计算所有的矢量 $\frac{S+\alpha N}{\alpha N}$ 并对这个矢量集合进行单值分解。所得到的正则矢量已经被用来计算按照它们的特征值的降序进行排序的 N_{filt} 个正则 Jacobian 矩阵 $F \cdot \frac{S+\alpha N}{\alpha N} \cdot F^{-1}$ 。

如上所述，使用主成分分析在减轻计算负担方面可以产生相当可观的改进。实验已经表明，有可能将有用的正则矩阵的数目减少到 5 个矩阵。甚至进一步的减少也是可能的。减少矩阵的数目降低了对空间的需求以及为进行适应所需的计算时间。为了更好地理解通过降维而获得的改进，表 I 对使用与不使用主成分分析的两种转换矩阵适应过程进行比较。

[表 I]

	维数	α 数值	无噪声	每小时 30 英里	每小时 60 英里	计算时间
不采用 PCA	32	2	98.74%	95.29%	90.21%	2439.28 μ s
采用 PCA	7	2.4	98.56%	95.70%	90.63%	1932.42 μ s

在上面的表 I 中，第一列表示维数，即正则矩阵的数目。第二列表示所使用的 α 适应因子的数值。其余各列给出针对 3 种环境条件的识别精度以及所需的相关的计算时间（对整个数据库进行适应的累计时间，以微秒表示），上述 3 种环境条件分别是：安静（无噪声），时速为 30 英里的汽车以及时速为 60 英里的汽车。

实验结果

在不同噪声条件下对上述噪声适应技术进行了测试。在本节中将再现我们的测试结果。为了测试适应系统，使用一个用于汽车导航系统的语音识别器。当然，本文所描述的适应技术并不局限于汽车导航或任何其他的识别任务。由于行驶中的汽车的噪声条件在不同的车速下可能出现大范围的改变，所以汽车导航被选择用于我们的测试。因此，在汽车环境中的适应系统的测试被选择为适应系统的能力的一种良好的度量。

这些实验的实验设置跟前面所述的相同。建立了 3 个测试组：（1）包括一个验证组，它由 20 名讲话者朗读的 462 个数字序列组成（不同于训练组），在与训练组相同的条件下进行记录；（2）由不同的讲话者朗读的 947 个数字序列组成，并且在时速为 30 英里的汽车中进行记录；（3）由相同的讲话者朗读的 475 个 5 位数的序列组成，并且在时速为 60 英里的汽车中进行记录。

使用一种简单的循环语法来进行识别，对所有的数字和无声状态来说，它们都具有相等的转移概率（“0”和“零”都是同一数字的模型）。在从已识别的句子中除去无声的情况之后，对 10 组数字进行精

度计算。

对这些第一组实验来说，信号被编码为 9 个 PLP 系数（包括残差）加上 9 个 δ 系数的矢量序列。适应，如果进行的话，仅被施加到前 9 个静态系数的平均值。为了适应，使用每一个句子的前 30 个帧来计算目标噪声。

结果再现于下面的表 II。在表中将不采用适应技术的隐藏马尔柯夫模型（HMM）的性能跟使用并行模型组合（PMC）以及传统的 Jacobian 适应（JA）所获得的结果进行比较。因此，表 II 说明并行模型组合以及 Jacobian 适应在出现噪声时如何改进识别性能。然而，表 II 没有说明使用 α 适应的改进的转换矩阵适应的性能。提供这份表是用来作为一个参考基准，参照于它，就能更好地理解改进的转换矩阵适应技术。

[表 II]

系统	无噪声验证	每小时 30 英里	每小时 60 英里
HMM（无适应）	98.84%	56.27%	35.83%
PMC	95.78%	91.72%	89.60%
JA	98.66%	83.76%	70.02%

[表 III]

适应技术	α 数值	无噪声	每小时 30 英里	每小时 60 英里
无		98.84%	56.27%	35.83%
PMC		95.78%	91.72%	89.60%
JA		98.66%	83.76%	70.02%
α -PMC	1.3	96.03%	91.67%	89.81%
α -TM	3	98.73%	95.24%	89.81%

表 III 表示在具有和没有 α 因子的情况下，并行模型组合以及 Jacobian 适应两者的性能比较。在表 III 中，具有 α 适应的转换矩阵适

应被表示为“ α -TM”。为了比较的目的， α 因子也被施加到并行模型组合技术，如在表示为“ α -PMC”的行中所示。

对表III中的结果进行比较，要注意的是，在出现噪声的情况下，改进的转换矩阵适应（ α -TM）的表现显著地优于标准的 Jacobian 适应（JA）。尽管 α 因子没有实质上降低 PMC 适应的性能，但它也没有提供显著的改进。

我们的实验结果表明，使用 α 适应因子的改进的转换矩阵适应技术给出大大优于标准的 Jacobian 适应的结果。而且，由于转换矩阵适应技术的计算开销固有地小于 PMC，所以它成为没有大量的处理能力或存储器的嵌入式识别系统的理想的候选者。这样的应用包括：例如，蜂窝电话识别系统、汽车导航系统以及其他消费类产品。

此外，通过使用本文所说明的降维技术，还可以在系统性能方面得到进一步的改进。当跟转换矩阵适应技术相结合时，其结果将是一种能很好地用于多种识别应用的紧凑的、有效的、和牢靠的适应系统。

进一步的实施例

人们将理解，尽管以上的讨论说明了 α -Jacobian 适应对加性噪声的优点，但它对卷积噪声也能获得类似的优点。一开始，在实践中，重要的是注意到所有的语音信号不仅被加性噪声、而且被卷积（或信道）噪声所恶化。在频谱域中表示结果信号的古典等式为 $S' = HS + N$ ，式中， S 为语音， S' 为结果信号， H 为信道噪声以及 N 为加性噪声。写在倒频谱域中，这个等式变为 $C(S') = C(HS + N)$ 。

让我们将 H_{tar} 和 H_{tar}' 分别定义为目标卷积噪声和目标加性噪声（即，在测试阶段所估计的），将 H_{ref} 和 H_{ref}' 分别定义为参考卷积噪声和参考加性噪声（即，在训练阶段所估计的）。然后，我们可以计算：

$$C(H_{tar}S+N) = C\left(\frac{H_{tar}}{H_{ref}}\left(H_{ref}S + \frac{H_{ref}}{H_{tar}}N_{tar}\right)\right)$$

$$C(H_{tar}S+N_{tar}) = C\left(\frac{H_{tar}}{H_{ref}}\right) + C\left(H_{ref}S + \frac{H_{ref}}{H_{tar}} N_{tar}\right)$$

然后，可以将 Jacobian 近似（或 α -Jacobian 近似）施加到总和的第二项：

$$\begin{aligned} C(H_{ref}S + \frac{H_{ref}}{H_{tar}} N_{tar}) &= C(H_{ref}S + N_{ref}) + \frac{\partial C(H_{ref}S + N_{ref})}{\partial C(N_{ref})} (C(\frac{H_{ref}}{H_{tar}} N_{tar}) - C(N_{ref})) \\ C(H_{ref}S + \frac{H_{ref}}{H_{tar}} N_{tar}) &= C(H_{tar}S + N_{ref}) + J_c(C(N_{tar}) - C(N_{ref})) - J_c(C(H_{tar}) - C(H_{ref})) \end{aligned}$$

式中， $J_c = \frac{\partial C(H_{ref}S + N_{ref})}{\partial C(N_{ref})}$ 表示语音矢量 S 的噪声适应矩阵。

因此，

$$C(H_{tar}S + N_{tar}) = (1 - J_c)(C(H_{tar}) - C(H_{ref})) + C(H_{ref}S + N_{ref}) + J_c(C(N_{tar}) - C(N_{ref})) \quad (7)$$

第一项 $(1 - J_c)(C(H_{tar}) - C(H_{ref}))$ 表示由信道适应矩阵 $(1 - J_c)$ 转换的卷积偏差 $(C(H_{tar}) - C(H_{ref}))$ 。重要的是要注意，这种转换属于卷积偏差，如在倒频谱域中表示的。第三项 $J_c(C(N_{tar}) - C(N_{ref}))$ 表示由噪声适应矩阵 J_c 转换的加性偏差。第二项 $C(H_{ref}S + N_{ref})$ 表示在倒频谱域中参考信号的一个模型。

现在转到图 6，图中示出了在一个语音识别系统（未示出）中，能进行噪声适应的一个噪声适应系统 60。一般来说，噪声适应系统 60 使用在识别环境 66 中从讲话者 62 发出的训练语音信号 75，以及在识别环境 66 中从讲话者 62 发出的目标语音信号 79（也被称为“输入语音信号”）。训练语音信号 75 被表示为具有被纳入到相加模块 74 之中的来自训练信道 72 的一个卷积噪声分量 H_{ref} 以及加性噪声分量 N_{ref} 。类似地，目标噪声信号 79 具有被纳入到相加器 78 之中的来自识别信道 76 的卷积噪声分量 H_{tar} 以及加性噪声分量 N_{tar} 。

信道 72、76 将典型地包括各种信号处理装置，诸如传声器、电话线路，等等。当识别信道 76 具有不同于训练信道 72 的特性时，在噪声适应方面将会出现困难。在这种情况下，目标卷积噪声分量 H_{tar}

将不同于参考卷积噪声分量 H_{ref} 。

在本发明的一种配置中，提供了一种方法，用以在特征域中的一个被观测的信号（或目标信号）的每一帧中，联合地补偿加性噪声和卷积噪声二者。使用一阶展开来近似非线性补偿函数：

$$\phi(C(H_{\text{ref}} \cdot S + N_{\text{ref}}), C(H_{\text{tar}} \cdot S + N_{\text{tar}})) \quad (8)$$

式中， $C(\cdot)$ 为倒频谱运算符， S 表示目标语音信号的一种无噪声的样式， H_{ref} 为针对训练环境的信道频率响应， H_{tar} 为测试环境的信道频率响应， N_{ref} 为针对训练环境的加性噪声，以及 N_{tar} 为针对测试环境的加性噪声（这里所使用的术语“测试环境”仅仅是按照常规来使用）。术语“测试环境”还用来包含其他识别环境，例如，由图 1 所说明的环境）。

ϕ 的一阶线性近似为：

$$\begin{aligned} \hat{C}(H_{\text{ref}} S + N_{\text{ref}}) = C(H_{\text{tar}} S + N_{\text{tar}}) &+ \left. \frac{\partial C(H_{\text{tar}} S + N)}{\partial N} \right|_{N=N_{\text{tar}}} (N_{\text{ref}} - N_{\text{tar}}) \\ &+ \left. \frac{\partial C(H S + N_{\text{tar}})}{\partial H} \right|_{H=H_{\text{tar}}} (H_{\text{ref}} - H_{\text{tar}}). \end{aligned} \quad (9)$$

令：

$$J_N = \left. \frac{\partial C(H_{\text{tar}} \cdot S + N)}{\partial N} \right|_{N=N_{\text{tar}}} = F^\dagger \frac{1}{H_{\text{tar}} \cdot S + N_{\text{tar}}} \quad (10)$$

$$J_H = \left. \frac{\partial C(H \cdot S + N_{\text{tar}})}{\partial H} \right|_{H=H_{\text{tar}}} = F^\dagger \frac{S}{H_{\text{tar}} \cdot S + N_{\text{tar}}}, \quad (11)$$

式中， $F^\dagger$ 是一个解相关（decorrelating）转换矩阵，例如离散余弦变换（矩阵 $F^\dagger$ 可以是复数，另一个矩阵 F 也可能是复数，它们可以容易地被确定，使得 $FF^\dagger = I$ ，为单位矩阵，因此， $F^\dagger$ 是矩阵 F 的逆共轭矩阵）。因此，本发明的一种配置对加性噪声偏差 $\Delta N = N_{\text{ref}} - N_{\text{tar}}$ 以及信道偏差 $\Delta H = H_{\text{ref}} - H_{\text{tar}}$ 作出估计。在一种配置中，等式 10 允许从有噪声的目标语音信号 $H_{\text{tar}} \cdot S + N_{\text{tar}}$ 中，直接地将 J_N 确定为 $J_N = F^\dagger \frac{1}{H_{\text{tar}} \cdot S + N_{\text{tar}}}$ 。

在本发明的一个特征 Jacobian 适应模块 100 的一种配置中，并且参照于图 7，在语音信号的无声区段，作出加性噪声偏差 $\Delta N = N_{\text{ref}} - N_{\text{tar}}$ 的一种常规的估计。这个加性噪声偏差被送往适应噪声补偿模块 102。为了对信道偏差 ΔH 作出估计，对一个部分地适应于具有加性噪声的测试环境的语音信号进行计算。例如，由噪声补偿模块 102 使用下式来确定一阶近似：

$$\begin{aligned} C(H_{\text{tar}} \cdot S + N_{\text{ref}}) &= C(H_{\text{tar}} \cdot S + N_{\text{tar}}) + \left. \frac{\partial C(H_{\text{tar}} \cdot S + N)}{\partial N} \right|_{N=N_{\text{tar}}} (N_{\text{ref}} - N_{\text{tar}}) \\ &= C(H_{\text{tar}} \cdot S + N_{\text{tar}}) + J_N \Delta N, \end{aligned} \quad (12)$$

式中， J_N 是一个噪声补偿矩阵，由于它仅依赖于被观测的信号（即，目标语音信号，见等式 10），所以它是可以容易地被确定的， $C(H_{\text{tar}} \cdot S + N_{\text{tar}})$ 是被观测信号的倒频谱。

接下来，由信道估计模块 104 使用“部分地”适应信号：

$$C(H_{\text{tar}} \cdot S + N_{\text{ref}}) = C(H_{\text{ref}} \cdot S + N_{\text{ref}}) + \left. \frac{\partial C(H \cdot S + N_{\text{ref}})}{\partial H} \right|_{H=H_{\text{ref}}} (\Delta H) \quad (13)$$

对卷积噪声偏差 $\Delta H = H_{\text{ref}} - H_{\text{tar}}$ 作出估计。式中， $C(H_{\text{tar}} \cdot S + N_{\text{ref}})$ 是在前一个步骤中计算出来的“部分地”适应的信号，并且， $C(H_{\text{ref}} \cdot S + N_{\text{ref}})$ 是对应于在频谱域中的被观测信号的倒频谱模型。然而，用于当前被观测的信号帧的模型对估计模块 104 不可用，因为它尚未被确定。因此，估计模块 104 使用这样的假设：在两个连续的帧以内，该信道是静止的，即，由信道估计模块 104 使用先前的帧信息对 ΔH 作出估计，由卷积偏差模块 106 进行输出。

更具体地说，信道估计模块 104 使用由适应噪声补偿模块 102 从先前的信号帧中输出的 $C(H_{\text{tar}} \cdot S + N_{\text{ref}})$ ，使用最可能的部分路径，上溯到前一个已解码的帧，来对 $H_{\text{ref}} \cdot S + N_{\text{ref}}$ 作出估计。Viterbi 解码或隐藏马尔柯夫模型（HMM）在业界中是众所周知的，并且在一种配置中被用来确定最相似的部分路径或者“部分追溯”。利用逆倒频谱运算符 $C^{-1}(\cdot)$ 从倒频谱域中的模型来确定频谱域中的模型，其中，

$C^{-1}(x) = \exp(F^{-1}x)$, F 是一个矩阵, F' 是 F 的解相关转换和逆共轭矩阵, 并且 $FF' = I$, 其中 I 是一个单位矩阵。令:

$$J_p = \left. \frac{\partial C(H \cdot S + N_{ref})}{\partial H} \right|_{H=H_{ref}} = F' \frac{S}{H_{ref} \cdot S + N_{ref}} \quad (14)$$

然后, 在一种配置中, 使用逆矩阵 J_p^{-1} 对 ΔH 作出估计:

$$\Delta H = J_p^{-1} [C(H_{ref} \cdot S + N_{ref}) - C(H_{tar} \cdot S + N_{ref})] \quad (15)$$

联合噪声与信道估计模块 106 使用由等式 (9) 给出的 ϕ 的线性近似来补偿被观测信号中的加性与卷积噪声。由信道估计模块 104 产生的数值 ΔH , 在语音信号的无声区段中产生的 $\Delta N = N_{ref} - N_{tar}$ 以及输入数值 $C(H_{tar} \cdot S + N_{tar})$ 被用来产生估计 $\hat{C}(H_{ref}S + N_{ref})$ 。

由于每一帧都进行适应, 所以当存在对加性噪声的“精细”估计时, 本发明的配置对非静止噪声和信道来说就显得更加牢靠。此外, 虽然以高速率进行修改, 也只有很少的参数被修改, 并且仅对信号的特征, 而不是对所有的模型特征进行修改。因此, 本发明的配置使用较小的存储器 (例如随机存取存储器或 RAM), 这对小的嵌入式系统来说是有利的。而且, 由于模型的特征不用修改, 所以适应的模型特征不需要存储在存储器之中。

在本发明的一种配置中, 并且参照图 8, 利用特征 Jacobian 适应模块 100 对倒频谱域中的一个参考模型的卷积噪声进行补偿 (作为联合补偿的一部分)。卷积模块 82 提供卷积偏差的估计并且令参考模型跟这个估计相适应。结果得到一个卷积适应模型 96。然后, 由加性噪声模块 80 对卷积适应模型 96 进行运算, 以便在加性与卷积噪声方面对参考模型充分地进行补偿。特别是, 卷积噪声模块 82 的一种配置包括一个卷积偏差模块 84, 用于对介于参考模型 70 以及目标语音信号 79 之间的卷积偏差作出估计。一个适应矩阵模块 86 根据训练语音信号来产生一个信道适应矩阵 88。信道适应矩阵 88 基本上是一个修改后的噪声适应矩阵 $1 - J_c$ 。矩阵 J_c 可能是一个标准的 Jacobian 矩阵, 或者是一个 α -Jacobian 矩阵。卷积偏差模块 84 用信道适应矩阵 88

来转换已估计的卷积偏差。卷积噪声模块 82 还包括一个被连接到卷积偏差模块 84 的相加模块 90，用以在倒频谱域中将已转换的卷积偏差添加到参考模型 70 中去。在由倒频谱分析模块 108 转换到倒频谱域之后，通过特征 Jacobian 适应模块 100 在特征域中对目标语音信号 79 的信道与加性噪声进行联合补偿，以产生补偿后的信号 97。

适应矩阵模块 86 基于训练语音信号来产生一个信道适应矩阵 88。信道适应矩阵 88 基本上是一个修改后的噪声适应矩阵 $1 - J_c$ 。重要的是要注意到，如上所述，矩阵 J_c 可能是一个标准的 Jacobian 矩阵，或者是一个 α -Jacobian 矩阵。卷积偏差模块 84 用信道适应矩阵 88 来转换已估计的卷积偏差。卷积噪声模块 82 还包括一个被连接到卷积偏差模块 84 的相加模块 90，用以在倒频谱域中将已转换的卷积偏差添加到参考模型 70 中去。

卷积偏差的计算

为了对卷积偏差作出估计，让我们假设介于参考模型以及信号之间的高斯对准 (Gaussian alignment) 为已知。基于这样的假设，我们将说明计算卷积偏差的两种可能的方案：

I. 第一方案

若我们假设 $H_{tar}S \gg N_{tar}$ ，则可能求出对应于语音单元的信号区段的平均值，以便获得：

$$\hat{C}(S_{tar}) = \frac{1}{T} \sum_{t=1}^T (C(H_{tar} \cdot S_t)) = C(H_{tar}) + \frac{1}{T} \sum_{t=1}^T C(S_t).$$

类似地，通过对相同的参考（即，无噪声）信号求平均值来计算项 $\hat{C}(S_{ref})$ 也应当是可能的。然而，由于只有带噪声的（目标）信号为已知，所以这个信号通常是得不到的。然后，我们必须对参考信号作出估计。通过使用介于模型与信号之间的高斯对准，就能获得这样一种估计。因此，对每一帧来说，对准于目标帧 $C(H_{tar} S_1)$ 的高斯密度代表对应的参考帧 $C(H_{ref} S_t)$ 。然后，有可能计算：

$$\hat{C}(S_{ref}) = \frac{1}{T} \sum_{t=1}^T (C(H_{ref} \cdot S)) = C(H_{ref}) + \frac{1}{T} \sum_{t=1}^T C(S_t).$$

通过将前面两个等式相减，我们得到：

$$C(H_{tar}) - C(H_{ref}) = \hat{C}(S_{tar}) - \hat{C}(S_{ref}). \quad (16)$$

在实际中，为了计算由上述等式表示的信道偏差，我们需要由识别器给出的当前句子的高斯对准。为了避免使用一个双通（two-pass）识别系统，我们已经决定令当前的句子适应于对前一个句子计算出来的适应偏差。这使得系统对环境的变化（例如，当一名不同的讲话者对先前的和当前的句子进行朗读时）更加敏感。然而，这种近似的优点在于，这种适应仍然是很快的，并且仅需要一次单独的识别通过。

卷积噪声模块 82 还可以包括一个高斯对准模块 92，它介于特征 Jacobian 适应模块 100 以及卷积偏差模块 84 之间，用以将目标语音信号分解为目标语音区段以及目标无声区段。高斯对准模块 92 对已补偿的信号 97 进行运算并产生输出 99，后者被送往卷积偏差模块 84。

（当不具备高斯对准模块 92 时，已补偿的信号 97，取代输出 99，被送往卷积偏差模块 84。）尽管典型地将分解功能纳入到语音识别器之中，但为了讨论的目的，在这里它被描述为卷积噪声模块 82 的一部分。事实上，在不离开本发明的特性和范围的前提下，本文所描述的许多部件的特定位置都是可以改变的。因此，卷积偏差模块 84 求出目标语音信号的平均值以及对应于参考模型 70 的高斯数据的平均值。卷积偏差模块 84 还计算介于高斯数据的平均值以及目标语音区段的平均值之间的差值。其结果得到一个可以在由信道适应矩阵进行转换之后，添加到处于倒频谱域之中的参考模型 70 中的卷积偏差 $C(H_{tar}) - C(H_{ref})$ 。

II. 第二方案

一个用以估计卷积偏差的优选的增量式解决方案允许我们考虑

目标噪声 N_{tar} 。这个方法利用卷积偏差的先前估计（例如，对前一个句子计算出来的估计），并通过使用当前句子的信号来改进这个估计。因此，从一句到下一句，以增量方式对卷积偏差作出估计，只要讲话者或环境不发生改变，就有可能获得改进。

对当前句子的语音区段的信号求平均值让我们得到：

$$\begin{aligned}\hat{C}(S_{tar}) &= \frac{1}{T} \sum_{i=1}^T (C(H_{tar} \cdot S_i + N_{tar})) \\ &= \frac{1}{T} \sum_{i=1}^T \left(C(H_{tar}) + C\left(S_i + \frac{N_{tar}}{H_{tar}}\right) \right) \\ &= C(H_{tar}) + \frac{1}{T} \sum_{i=1}^T C\left(S_i + \frac{N_{tar}}{H_{tar}}\right)\end{aligned}$$

在先前的句子中，已经对适应模型 $C(H_{ref}S + \frac{H_{ref}}{H_{tar}} N_{tar})$ 作出估计。

因此，就有可能使用这一组模型来计算参考帧的平均值：

$$\begin{aligned}\hat{C}(S_{ref}) &= \frac{1}{T} \sum_{i=1}^T \left(C\left(H_{ref} \cdot S_i + \frac{H_{ref}}{H_{tar}} N_{tar}\right) \right) \\ &= \frac{1}{T} \sum_{i=1}^T \left(C(H_{ref}) + C\left(S_i + \frac{N_{tar}}{H_{tar}}\right) \right) \\ &= C(H_{ref}) + \frac{1}{T} \sum_{i=1}^T C\left(S_i + \frac{N_{tar}}{H_{tar}}\right)\end{aligned}$$

将前面两个等式相减，就给出由等式（16）表示的信道偏差。用这个方法，我们必须假设第一个句子的目标与参考信道是相同的。

重要的是要注意到，上述的联合补偿方案享受到与仅针对加性噪声补偿而描述的 α -Jacobian 适应有关的许多好处。正如已经讨论过的那样， α -Jacobian 适应源于 Jacobian 算法。 α -Jacobian 适应不同于 Jacobian 适应之处在于实现近似的方法：后者利用线性函数，它能最好地近似在训练条件下的 PMC，而前者以实验方法来计算一个能针对一组选定的可能测试条件集合较好地近似 PMC 的线性函数。其结果是，在真实的测试条件下， α -Jacobian 适应比古典的 Jacobian 适应表现得更好。

通过使用主成分分析来减少需要的转换矩阵的数目，在不使识别结果变坏的前提下，我们已经将这个算法的时间和存储器复杂性降低

了一半以上。

由于在倒频谱域中对观测信号所进行的转换在数学上等效于在时域中的线性滤波，所以普通的专业人士的第一印象可能是，如图 7 和图 8 所示的本发明的配置在使均方误差最小化方面是次最佳的。然而，在实际应用中发现本发明的配置并不是这样的情况。

从本质上来说，本发明的说明仅仅是示例性的，因此，作者打算将那些不背离本发明的要点的改动都纳入本发明的范围之内。这样的改动不应被认为是对本发明的精神实质和范围的背离。

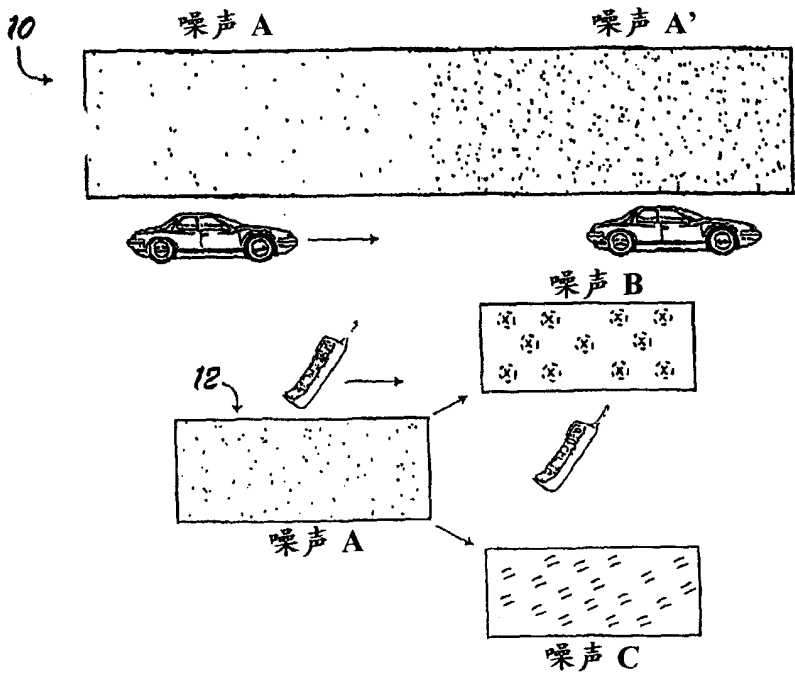


图1

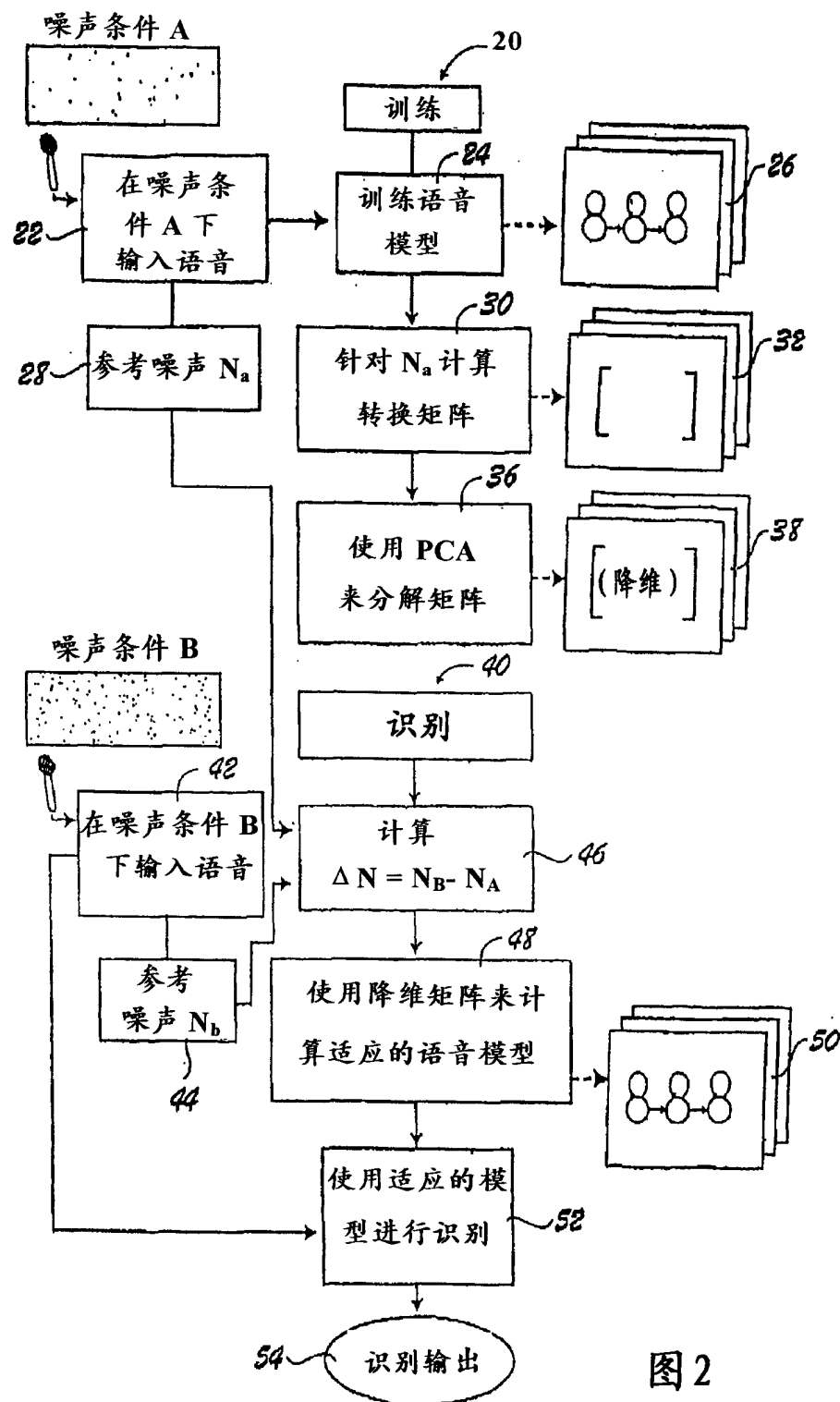


图 2

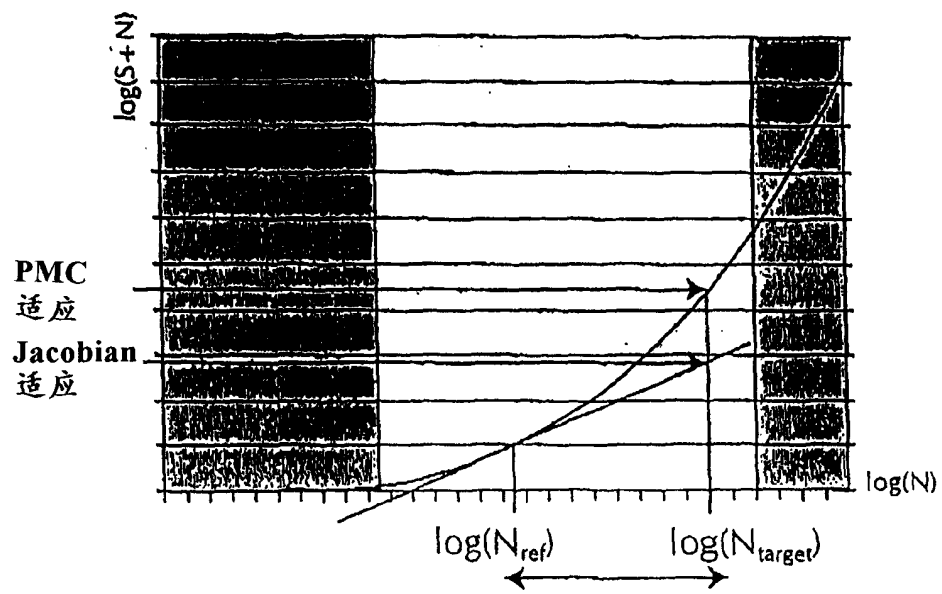


图 3

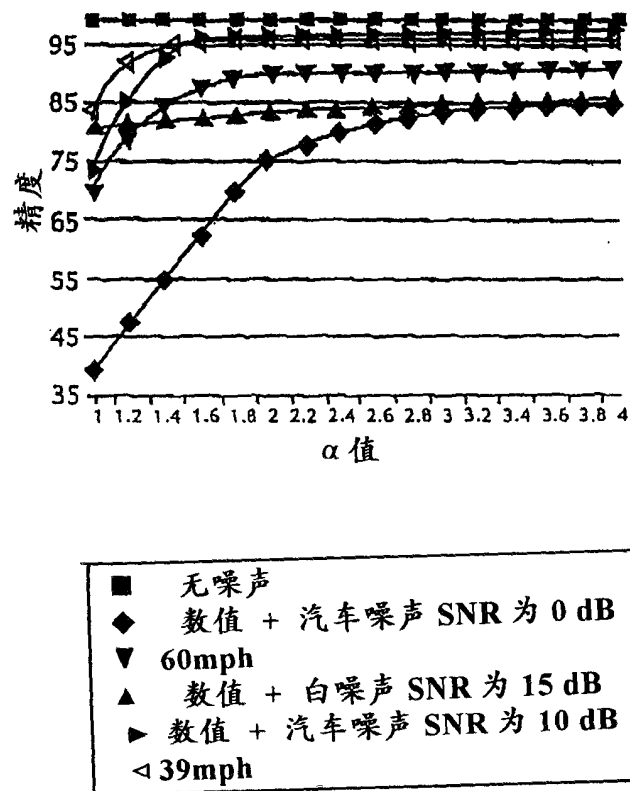


图 4

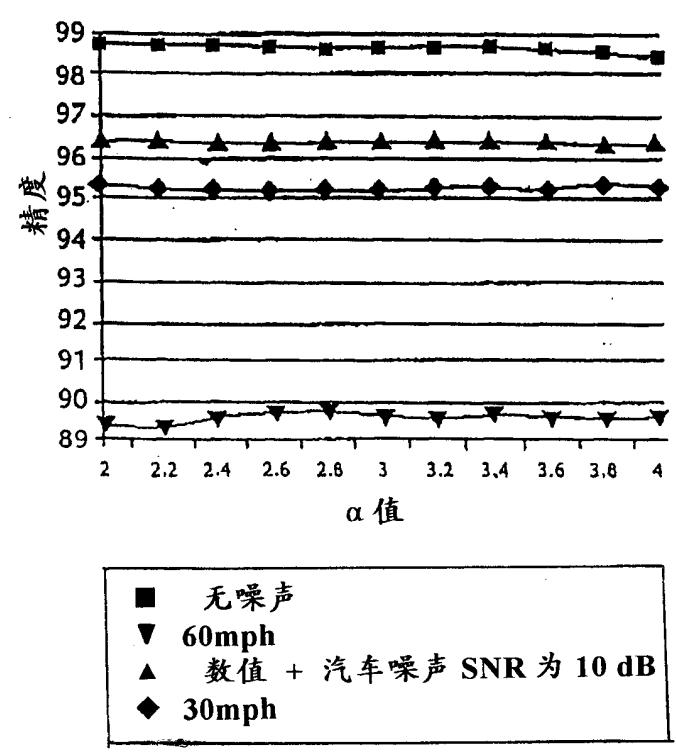


图5

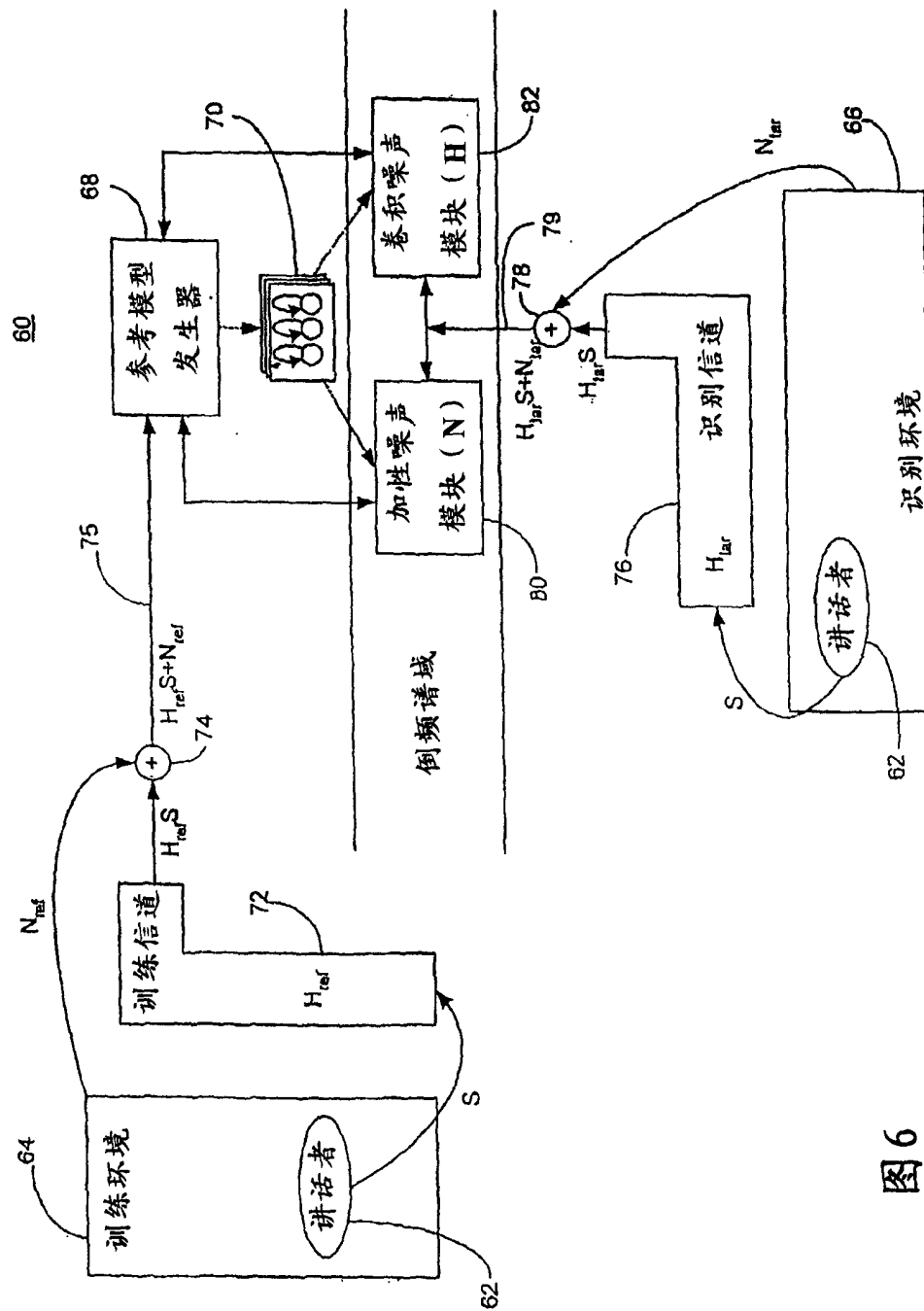


图6

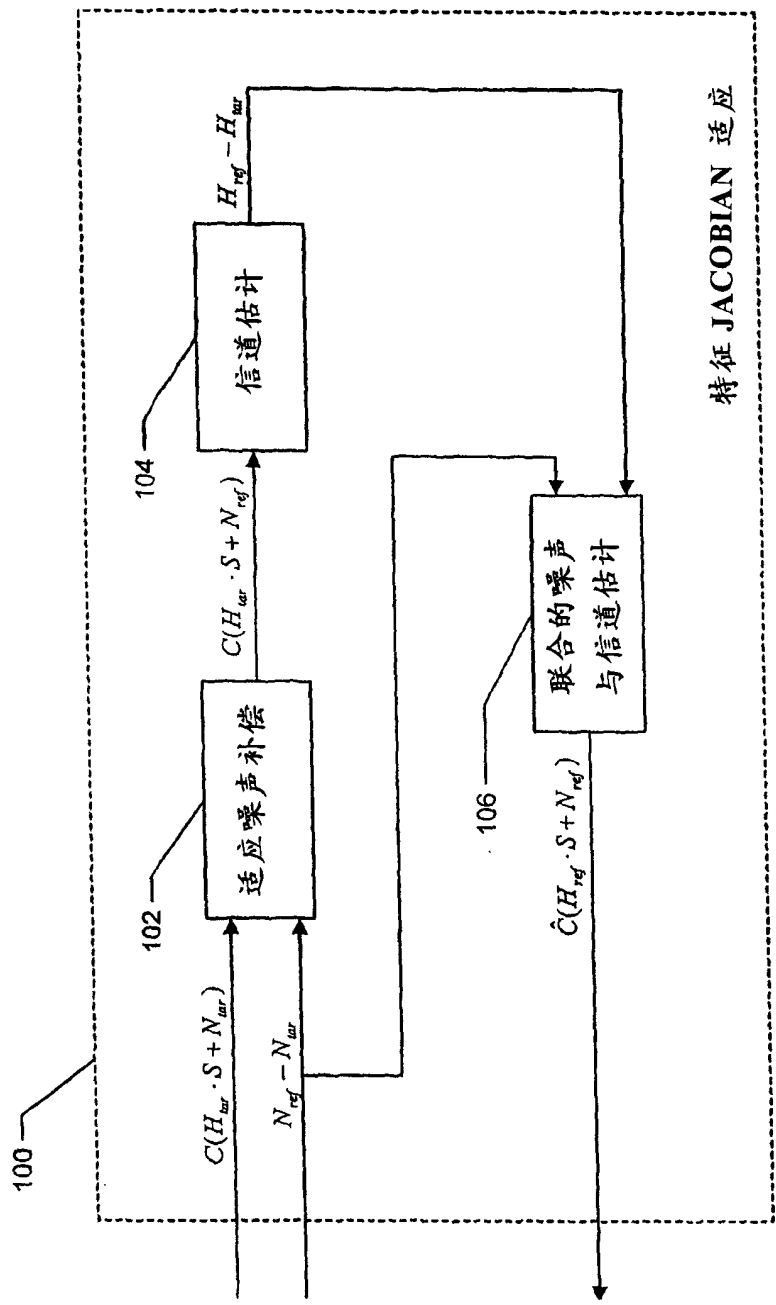


图7

