第 23 卷第 6 期　　声 学 学 报　　Vol. 23, No. 6
1998 年 11 月　　ACTA ACUSTICA　　Nov., 1998

几种高鲁棒性通道及说话人自适应语音识别算法研究

陈景东　姚　磊　黄泰翼

(中国科学院自动化研究所模式识别国家实验室　北京 100080)

1997 年 5 月 9 日收到
1997 年 9 月 5 日定稿

摘要　鲁棒性问题是决定语音识别技术能否在实际中得以应用和推广的关键问题之一。概括起来说，导致语音识别系统性能变坏的原因大体上来自三个方面，即噪声 (加性噪声、卷积噪声) 、信道变化和不同的讲话者 (不同的声道形状、不同的发音方式等) 。本文对三种高鲁棒性自适应语音识别方法进行了研究和改进，并对它们的性能进行了比较，这三种方法分别是 VQ 码本自适应法、 HMM 参数自适应法和基于正则相关分析的谱变换补偿方法。实验结果表明，这三种方法都能提高非特定人语音识别系统对信道以及说话人的鲁棒性，而且基于正则相关分析的谱变换补偿方法具有最好的性能，它能够补偿由三种失真源同时引起的训练条件与测试条件之间的不匹配，因此适合作为一种通用的自适应方法。

PACS 数　43.70

Channel and speaker adaptation techniques for robust speech recognition

CHEN Jingdong　YAO Lei　HUANG Taiyi

(*National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences* Beijing 100080)

Received May 9, 1997
Revised September 5, 1997

Abstract Acoustical mismatches between training and testing environments of HMM-based speech recognizer often cause sever degradations in recognition performance. The mismatch is mainly caused by noise, changes of the channel through which the speech signal is transmitted and differences of speakers. This paper addresses the problem of changes of the recording channel and variations of speakers. Three adaptation methods are presented to deal with the problem, i.e., the adaptation via VQ prototype modification, the adaptation via HMM parameters modification and the canonical correlation based compensation method (short for CCBC). Experimental results have shown that all the three techniques can make our speaker-independent recognition system robust to channels and speakers. Among the three techniques, the CCBC has

the best performance and it can be used as a unified approach to cope with mismatch caused both by noise, by differences of channels and by variations of speakers.

引言

当训练与测试条件不匹配时，语音识别系统的性能就会严重下降。例如，一个大词汇量非特定人语音识别系统，其对训练集外讲话者的识别率往往要比对训练集内讲话者的识别率低得多。同样，对于一个已训练好的语音识别器，如果输入通道改变 (如换一个话筒) 时，其性能也会下降。

概括起来讲，训练条件与测试条件之间的不匹配可以分为三大类：即不同的讲话者、信道畸变 (信道干扰、采用不同的输入设备) 、以及噪声的影响。噪声、信道和讲话者三者以不同的方式改变着语音信号。本文中我们主要研究如何消除信道畸变和不同的讲话者对系统性能产生的影响。

信道是信号赖以传输的通路。我们所接收到的语音信号实际上是原始语音信号同信道传输函数的卷积。在训练和识别过程中，如果信道发生了变化，识别系统的性能将会受到严重影响。例如：电话信道可以看作是下限频率为 200 Hz，上限频率为 3200 Hz 的带通滤波器，而且在不同频率段对语音信号的衰减不同，如果在训练和识别过程中，信道的频率特性不一致时，系统的识别率就会急剧下降。

采用不同的输入设备，同样也会导致训练与测试条件的不匹配。例如，Acero[1] 等人在 Sphinx 语音识别系统中的实验结果表明，同样一个识别系统，当使用 Sennheiser close-talking 话筒进行测试时，其识别率为 85%，而用 desk-top 话筒进行测试时，其识别率不到 20%。又如，我们在程控交换机上设计实现了一个电话语音自动拨号系统，当训练和测试均使用话机所带的话筒时，其识别率为 95%，但如果测试时使用免提方式，则其识别率不到 65%.

不同的说话人，其声道形状、声道长度、发音方式均不相同，甚至说话人的身体状况 (如感冒等) 、说话人的情绪 (当说话人意识到面对的是机器而不是人) 等因素都会使语音信号的结构发生很大变化，从而影响识别系统的性能。

提高识别系统对信道和说话人的鲁棒性的最佳方法是在测试条件下采集语音数据，重新训练模型，但这对于大词汇量语音识别系统来说显然是不现实的。本文中，我们介绍了三种信道及说话人自适应方法，它们都是利用少量的自适应语料，修改已训练好的模型参数或对测试语音的声学特征进行某种变换，使系统达到适应新环境的目的。这三种方法分别是 VQ 码本自适应法、HMM 参数自适应法和基于正则相关分析的谱变换补偿方法。实验结果表明，这三种方法都能够大大提高我们所设计的基于 VQ-DHMM 识别系统对信道和说话人的鲁棒性，而且基于正则相关分析的谱变换补偿方法具有最好的性能，它能够补偿由噪声、信道和讲话者三种失真源同时引起的训练条件与测试条件之间的不匹配。

本文的其余部分安排如下：第二部分阐述了 VQ 码本自适应方法，第三部分介绍了 HMM 参数自适应法 [2]，第四部分介绍了基于正则相关分析的谱变换补偿方法 [3]，第五部分给出了三种方法的实验结果，并对结果作了分析，最后，在第六部分中给出了本文的重要结论。

1 VQ 码本自适应法

在文献 [8] 中，我们曾介绍了一种用于说话人自适应的 VQ 码本自适应方法，其基本思想是利用新说话人的少量自适应语料，通过 K-means 聚类算法修改原来的码本，以减少新说话人特征矢量量化时的总体误差，而不影响已训练好的 HMM 参数。显然，这种自适应是在量化过程中进行的，所得的自适应 VQ 码本是基于自适应语料的一种最小失真估计。下面我们介绍一种新的用

于通道和说话人自适应的 VQ 码本自适应方法：首先为新条件下的自适应语料建立一个 SCHMM 模型，然后利用少量的自适语料和 Forward–backward 算法对模型进行训练，训练结束后，即可得到新环境下的自适应 VQ 码本。利用这种方法所得的自适应 VQ 码本实际上是基于自适应语料的一种最大似然估计。需要特别指出的是，这种 VQ 码本自适应也是在矢量量化过程中进行的，所建立的 SCHMM 只是为估计新的 VQ 码本所用，识别过程中所用的 HMM 模型及参数保持不变。这种 VQ 码本自适应算法的具体步骤如下：

1 为自适应语料建立一个 SCHMM 模型

设 $\{C_1, C_2, \cdots, C_M\}$ 为 VQ 码本中码字的标号，a_{ij} 代表从状态 S_i 向状态 S_j 转移。在时刻 t, 我们可以将从状态 S_i 转移到状态 S_j 时，语音特征矢量 x_i 出现的概率表示为

$$P(x_t|a_{ij}) = \sum_{k=1}^{M} P(x_t|C_k, a_{ij})P(C_k|a_{ij}), \tag{1}$$

为了简化模型和减少自由参数，假设 [9]

$$P(x_t|C_k, a_{ij}) = P(x_t|C_k), \tag{2}$$

这样 (1) 式简化为

$$P(x_t|a_{ij}) = \sum_{k=1}^{M} P(x_t|C_k)P(C_k|a_{ij}). \tag{3}$$

上式即为自适应语料的 SCHMM 模型。式中 $P(x_t|C_k)$ 是具有连续分布的输出概率，它描述了声学矢量与码字之间相关性的大小。实际上，(3) 式也可以看作是由离散概率 $P(C_k|a_{ij})$ 对 $P(x_t|a_{ij})$ 进行加权的 M 维混合密度 CDHMM 。

2 对 SCHMM 进行训练以获得自适应 VQ 码本

假定 $P(x_t|C_k)$ 服从高斯分布，且其均值为 μ_k, 对角协方差矩阵为 Σ_k ，则我们可以利用少量的自适应语料和 forward–backward 训练算法进行 VQ 码本的自适应。其具体步骤为： 1). 利用识别系统码本中原有的码矢作为均值矢量来初始化概率 $P(x_t|C_k)$ ，其方差可以初始化为单位矩阵。2). 利用识别系统中原有 HMM 参数的转移概率来初始化 SCHMM 模型的转移概率。3). 进行多次 forward–backward 迭代运算，在每次迭代结束后，更新输出概率的均值和方差，但转移概率保持不变。4). 整个迭代结束后，将训练所得的概率 $P(x_t|C_k)$ 的均值矢量作为新条件下的 VQ 码本。

3 利用新的 VQ 码本和原有的 HMM 参数构成新环境下的 VQ/DHMM 识别器。

2 HMM 参数自适应法

Nishmura[2] 的研究表明，通过线性映射变换可以将已训练好的 HMM 参数适应于新的说话人，在本文中，我们对这种方法进行了改进，用于通道自适应和说话人自适应。

设 S_k 是 HMM 的一个状态，C_i 是 HMM 的原始码矢，$1 \le i \le M$ ，C_j 是对应自适应后 HMM 的码矢，$1 \le j \le M$ 。自适应后 HMM 的状态输出概率为：

$$P(C_j|S_k) = \sum_{C_i=1}^{M} P(C_j|C_i, S_k)P(C_i|S_k), \tag{4}$$

设 $P(C_j|C_i, S_k)$ 与状态 S_k 无关，则上式可改写为

$$P(C_j|S_k) = \sum_{C_i=1}^{M} P(C_j|C_i)P(C_i|S_k), \tag{5}$$

下面的问题是如何估算 $P(C_j|C_i)$。利用 Viterbi 算法将自适应语料的每一词 $\omega(1 \le \omega \le W)$，每一帧 $t(1 \le t \le T)$，对应于 HMM 的某一状态 $S = V[C(\omega,t)]$，对 C_j，C_i 的关系进行统计

$$R(C_j, C_i) = \sum_{C(\omega,i)=C_j} P(C_i|V[C(\omega,t)]), \tag{6}$$

式中 $R(C_j,C_i)$ 为在输出 C_j 的状态输出 C_i 的累加概率。归一化得

$$P(C_j|C_i) = R(C_j,C_i) / \sum_{C_j=1} R(C_j,C_i). \tag{7}$$

同样，上述过程也适用于状态转移概率 $P(T_j|S_k)$, 其中 T_j 为 HMM 的状态转移，j 为状态转移数，$1 \le j \le N$。

对应于自适应语料中未出现的 C_j，T_j，其对应的概率不作修改。并对 $P(C_j|S_k)$ 与 $P(C_i|S_k)$, $P(T_j|S_k)$ 与 $P(T_i|S_k)$ 进行适当的加权，以获得较高的识别率。

3 基于正则相关分析的谱变换补偿自适应算法

语音信号可以表示成为一个特征矢量序列，每个特征矢量分别可以看作是特征空间中的点。测试集和训练集之间的特征差异可以通过正则相关技术，利用少量的参考语料和自适应语料来进行补偿 (以下简称此方法为 CCBC(Canonical Correlation Based Compensation) 方法)。不过，CCBC 方法并不是将测试语音特征矢量直接映射到训练空间，也不是将训练语音特征矢量直接映射到测试空间，而是分别将训练语音矢量和测试语音矢量同时映射到第三个谱空间 (我们称之为参考空间), 并使它们在该空间中的相关性达到最大。在文献 [3] 中，我们曾利用 CCBC 方法提高了识别系统对噪声的鲁棒性，文中对此方法进行了改进和推广，用以提高系统对信道、说话人和噪声的鲁棒性。

记 $\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$ 分别为测试集和训练集中的倒谱特征矢量，设倒谱矢量的长时均值为 0，即

$$\varepsilon[\boldsymbol{X}^{(1)}] = \varepsilon[\boldsymbol{X}^{(2)}] = 0, \tag{8}$$

如果倒谱矢量的长时均值不为 0, 则采用倒谱归一化技术，用 $\boldsymbol{X}^{(1)} - \varepsilon[\boldsymbol{X}^{(1)}]$ 和 $\boldsymbol{X}^{(2)} - \varepsilon[\boldsymbol{X}^{(2)}]$ 来代替 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$。

假设可以利用线性变换将 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 映射到参考空间中，则有

$$\begin{aligned} U &= \boldsymbol{A}'\boldsymbol{X}^{(1)}, \\ V &= \boldsymbol{B}'\boldsymbol{X}^{(2)}, \end{aligned} \tag{9}$$

式中 $\boldsymbol{A}'$ 和 $\boldsymbol{B}'$ 分别是对应于 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的系数变换矩阵，“$'$”表示转置，U, V 分别是对应于 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 在参考谱空间中的映射。U 和 V 之间的均方差为

$$\xi = \{|U - V|^2\}. \tag{10}$$

利用导数理论和多维随机变量的统计分析理论 [4], 使 ξ 达到最小，即可使 U 和 V 在参考空间中的相关性达到最大。具体的求解过程如下：

令

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{bmatrix} \tag{11}$$

$\boldsymbol{X}$ 的方差矩阵可以分解为

$$\boldsymbol{X}=\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}, \tag{12}$$

设 U 和 V 的方差为 1，则有

$$1=\varepsilon[U^2]=\boldsymbol{A}'\Sigma_{11}\boldsymbol{A}, \tag{13}$$

$$1=\varepsilon[V^2]=\boldsymbol{B}'\Sigma_{22}\boldsymbol{B}, \tag{14}$$

$$\varepsilon[UV']=\varepsilon[\boldsymbol{A}'\boldsymbol{X}^{(1)}\boldsymbol{X}^{(2)\prime}\boldsymbol{B}]=\boldsymbol{A}'\Sigma_{12}\boldsymbol{B}. \tag{15}$$

令

$$\psi=\boldsymbol{A}'\Sigma_{12}\boldsymbol{B}-\frac{1}{2}\lambda(\boldsymbol{A}'\Sigma_{11}\boldsymbol{A}-1)-\frac{1}{2}\mu(\boldsymbol{B}'\Sigma_{22}\boldsymbol{B}-1). \tag{16}$$

上式中 λ 和 μ 为 Lagrange 算子。不难证明，使 (10) 式达到最小等价于使 (16) 式达到最大。将 ψ 分别对 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 求偏导，并令其导数为 0，不难推得 $\boldsymbol{A}$, $\boldsymbol{B}$ 满足：

$$[\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}-\lambda^2 I]\boldsymbol{A}=0, \tag{17}$$

和

$$\Sigma_{21}\boldsymbol{A}-\lambda\Sigma_{22}\boldsymbol{B}=0, \tag{18}$$

显然，要使 (17) 式有解，其左边的矩阵必须为奇异阵，如果令 $\eta=\lambda^2$, 则有

$$|\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}-\eta I|=0. \tag{19}$$

可以证明 (19) 式有 P 个正根 [10], 分别记为 $\boldsymbol{\eta}_1\geq\boldsymbol{\eta}_2\geq\cdots\geq\boldsymbol{\eta}_p$, 求解方程 (19), 则正则相关问题就转化成了普通的特征值问题。相应于 $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_p$ 的特征向量 $\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\cdots,\boldsymbol{\alpha}^{(p)}$ 即为矩阵 $\boldsymbol{A}$ 的列向量。再根据 (18) 式可得

$$\boldsymbol{\beta}^{(i)}=\lambda_i^{-1}\Sigma_{22}^{-1}\Sigma_{21}\boldsymbol{\alpha}^{(i)}\qquad(i=1,2,\cdots,P). \tag{20}$$

式中 $\boldsymbol{\beta}^{(i)}$ 为矩阵 $\boldsymbol{B}$ 的第 i 列列向量。

求得矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 后，计算 $[\boldsymbol{A}']^{-1}\boldsymbol{B}'\boldsymbol{X}^{(2)}$，即可将测试倒谱矢量映射到训练谱空间。更一般地，如果语音倒谱的长时均值不为 0，则测试倒谱矢量在训练空间中的映射为 $[\boldsymbol{A}']^{-1}\boldsymbol{B}'(\boldsymbol{X}^{(2)}-\varepsilon[\boldsymbol{X}^{(2)}])+\varepsilon[\boldsymbol{X}^{(1)}]$ 。

4 实验及其结果

为了验证各自适应方法的有效性，我们做了大量的实验。实验是在基于 VQ/DHMM 的非特定人语音识别系统上进行的。

实验中使用的语音库包括 10 个男声和 10 个女声共 20 个人的发音。每人发 1—7 字词共 646 词和 1—7 字长数字串共 70 个。实验时，用 18 人 (男女各 9 人) 作为训练集，其余的 2 人作为测试集。语料选用 500 词，包括 1—7 音节。

语音库建立时，是由 DAT 数字录音机、放大器、滤波器和 TMS320C30 开发板构成的系统所采集，采样频率为 16 kHz，有效位为 14 bit 。

我们利用 Hamming 窗对语音信号进行加窗处理，窗长为 24 ms, 采用 12 阶 MFC 倒谱和 12 阶一阶差分 MFC 倒谱为特征矢量，码本大小为 $M=256$ 。提取 MFC 倒谱时所使用的三角滤波器个数为 25 个。

实验中采用的声学模型为细化的声韵母模型，其中包括 100 个细化的声母模型和 37 个非细化的韵母模型，共 137 个模型单元。自适应词表是从 500 词中选出的 108 词，该自适应词表能够覆盖细化的声韵母模型。

我们首先考查了在由不同说话人引起的不匹配情况下，各种方法的性能。表 1 给出了系统对训练集外两位测试者的识别结果。

表 1 说话人自适应实验结果

自适应方法	识别率	
	对测试者 ZBR 的识别率 (%)	对测试者 ZXJ 的识别率 (%)
无自适应	93.0	92.0
方法 1	94.6	93.4
方法 2	94.4	93.2
方法 3	95.2	93.8
方法 4	95.6	94.6
方法 5	96.0	95.1
方法 6	96.1	95.8

表中，各方法分别为：方法 1: VQ 码本自适应法；方法 2: HMM 参数自适应法；方法 3: VQ 码本自适应法与 HMM 参数自适应法相结合；方法 4: CCBC 方法；方法 5: VQ 码本自适应法与 CCBC 方法相结合；方法 6: VQ 码本自适应法、 HMM 参数自适应法和 CCBC 方法三者相结合。

从表中可以看出，各自适应方法都能够提高系统对新的说话人的识别率，同 VQ 码本自适应法、 HMM 参数自适应法相比， CCBC 方法具有更好的性能，而且如果将不同自适应方法结合起来，能够进一步改善系统的性能。

为了考查在由不同信道和不同说话人同时引起的不匹配情况下，各自适应方法的性能，我们采用 Sound Blaster 声卡采集了三个男声的语音，词表与上述试验中所选的 500 词相同，然后利用各自适应方法对这三位男声语音进行自适应处理和识别，各方法相应的误识率如图 1 所示。为了将本文的自适应方法与语音识别中最常用的通道自适应技术相比较，图中还给出了利用倒谱归一化技术 [5] 的识别结果。

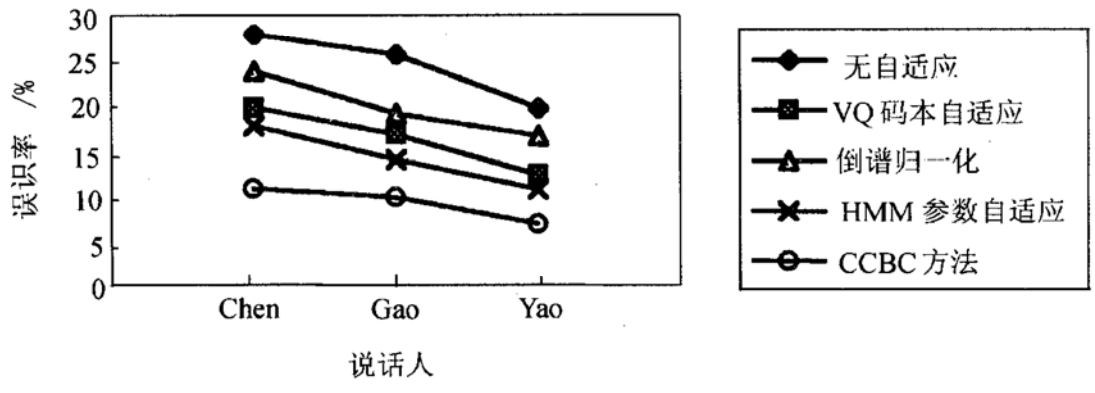

图 1 信道及说话人自适应实验结果

由图 1 不难看出，在由不同信道和不同说话人同时引起的不匹配情况下， VQ 码本自适应法、 HMM 参数自适应法以及 CCBC 方法都能够大大减小系统的误识率，而且它们的性能都超过了用于通道自适应的倒谱归一化技术。在所研究的方法中， CCBC 方法的性能最好。

在实验中，我们发现， CCBC 方法可以同时补偿由噪声、不同信道以及不同讲话者三者引起的不匹配。为了证实这一点，我们在安静办公室环境下，利用 Sound Blaster 声卡录制了一个女声，词表仍为 500 词。然后在语音样本中加入了仿真噪声。图 2 给出了各种自适应方法的实验结果。为了同其它的噪声自适应技术相比较，图中还给出了利用 Lin–Log RASTA 技术 [6] 的识别结果。

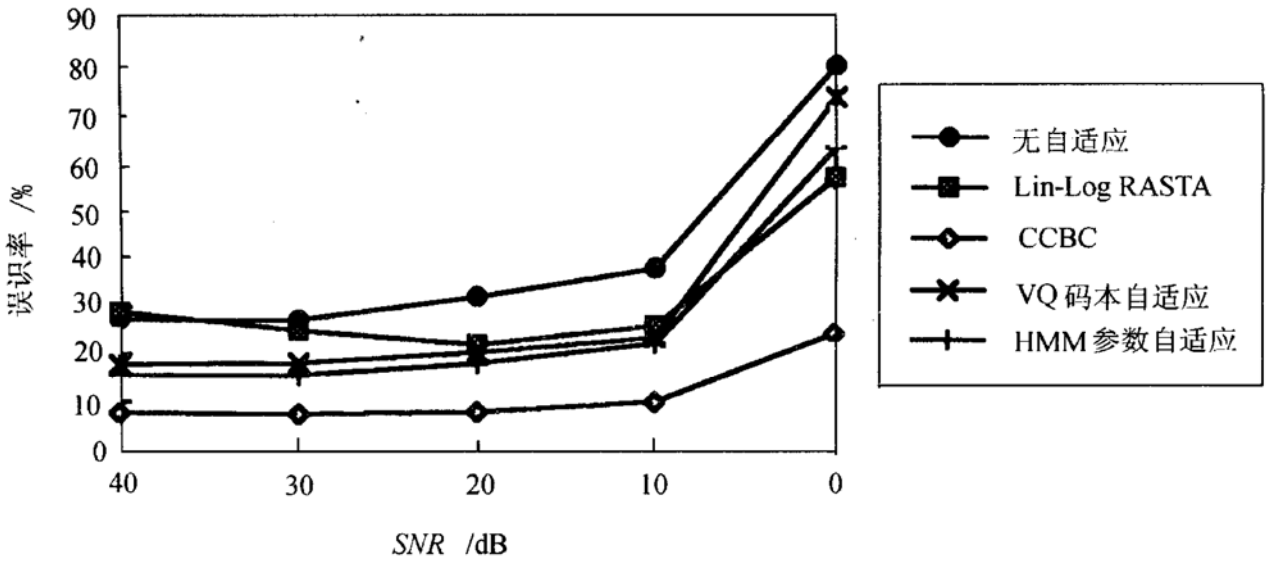

图 2　噪声、信道及说话人自适应实验的结果

从图 2 可直观地看出，在由噪声、不同信道以及不同讲话者同时引起的不匹配情况下，图中所列出的各种自适应方法都在不同程度上改善了系统的性能。但当信噪比较低时 (SNR<10 dB)，即使是使用了 Lin–Log RASTA 、 VQ 码本自适应或 HMM 参数自适应等自适应技术，识别系统的性能仍急聚下降。这是因为 Lin–Log RASTA 技术实际上相当于一种噪声屏蔽技术 [7] ，它在屏蔽噪声的同时，也屏蔽了语音信号中的一些有用成份，噪声电平越高，语音信号中被屏蔽的有用成份越多，系统的性能也就越差。 Lin–Log RASTA 技术的另一个缺点是使系统对纯净语音的识别率降低，这一点在图 2 中也能看出来。 VQ 码本自适应法、 HMM 参数自适应法都是利用少量的自适应语料来估计新环境下的 VQ 码本或 HMM 参数，显然，噪声越大，这种估计就越不准确，系统的性能也就越差。相比之下， CCBC 方法的性能比较稳定，无论是在高信噪比，还是在低信噪比情况下，都能够大大改善系统的性能。这是因为 CCBC 方法通过使测试语音倒谱矢量和训练语音倒谱矢量在参考空间中的相关性达到最大，从而比较好地补偿了由噪声、信道和讲话者三者同时引起的训练集和测试集之间的不匹配，同时， CCBC 方法不修改已训练好的 VQ 码本和 HMM 参数，从而避免了由噪声带来的估计误差。

5　结论

本文对 VQ 码本自适应法、 HMM 参数自适应法和 CCBC 方法进行了研究和改进，用以提高基于 VQ/HMM 的语音识别系统对信道以及说话人的鲁棒性。实验结果表明，这三种方法都能

够大大降低在由信道和不同的讲话者同时引起的不匹配情况下系统的误识率。在这三种方法中，CCBC 方法的自适应效果最好。

如果将这三种方法进一步扩展，用以提高系统对噪声、信道和说话人三者的鲁棒性，实验结果表明，当信噪比较低时 (SNR<10 dB), VQ 码本自适应方法和 HMM 参数自适应方法的自适应效果不明显，但 CCBC 方法仍能够大大改善系统的性能，这说明 CCBC 方法能够补尝由噪声、信道和说话人三者同时引起的不匹配，适合作为一种通用的自适应方法。

参 考 文 献

1 Alejandro Acero, Richard M Stern. Environmental Robustness in Automatic Speech Recognition. in *Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing*, 1990, 849—852

2 Nismura Masafumi, Sugawara Kazuhide. Speaker Adaptation Method for HMM-based Speech Recognition. in *Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing*, 1988, 207—210

3 Dong Yu, Huang Taiyi. Canonical Correlation Based Compensation Approach for Robust Speech Recognition in Noisy Environment. in *Proc. EUROSPEECH'95*, 477—480

4 Anderson T W. An Introduction to Multivariate Statistical Analysis. the second Edition, John Wiley & Sons, Inc., 1984

5 Liu F H, Stern R M, Acero A, Moreno P J. Environment normalization for robust speech recognition using direct cepstral comparison. in *Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing*, April 1994, **I**: 61—64

6 Koehler J, Morgan N, Hermanskey H, Hirsch H G, Tong G. Integrating RASTA-PLP into speech recognition. in *Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing*, April 1994, **I**: 412—424

7 Openshaw J P, Manson J S. On the Limitations of Cepstral features in Noise. in *Proc., IEEE Int., Conf., Acoust., Speech, Signal Processing*, April 1994, **I**: 49—53

8 张希军，徐 波，黄泰翼. 基于 VQ/HMM 的非特定人语音识别与说话人自适应. 软件学报，863 专刊，1996 年 10 月

9 Dimitry Rtischev, David Nahamoo, Michael Picheny. Speaker adaptation via protype modification. *IEEE Transactions on Speech and Audio Processing*, 1994, **2**(1): Part I

10 程云鹏，张凯院，徐仲等著. 矩阵论，西安：西北工业大学出版社，1989. 6 月