

# 基于子带分析稳健的说话人识别

关海欣, 曾庆宁

(桂林电子科技大学信息与通信学院, 桂林 541004)

**摘要:** 针对说话人识别系统易受噪声干扰的问题, 考虑语音的非平稳特性, 以语音信号各个频段区间所含有信号能量大小和所携带信息量多少的不同为前提假设, 以大量的试验为基础, 研究语音信号各个子带对噪声的稳健性、所含信息的充分性。研究发现, 语音信号的低频区携带大量的说话人个性信息, 并且能量集中, 不易被噪声干扰, 再通过适当地处理, 使系统在 10dB 信噪比下识别率超过 90%。

**关键词:** 说话人识别; 子带信号; 稳健

中图分类号: TN912.3

文献标识码: A

文章编号: 1000-3630(2009)-02-0142-06

DOI 编码: 10.3969/j.issn1000-3630.2009.02.011

## Robust speaker recognition based on subband analysis

GUAN Hai-xin, ZENG Qing-ning

(Guilin university of Electronic Technology, Guilin 541004, China)

**Abstract:** Speaker recognition system is susceptible to noise. On this problem, the non-stationary of the speech signal needs to be considered based on the assumption that the energy and information are different in different sub-band signals. So, the stability of the noise and the completeness of the information for different sub-band signals are analyzed. Through a large number of experiments, it can be found that the low frequency speech signal, which can not only bring a large number of personalized information but also concentrate much energy, is not easy being interfered by noise. Some appropriate treatment has been adopted to make the system's recognition rate over 90% for the signal to noise ratio of 10dB.

**Key words:** speaker recognition; sub-band signal; robust

## 1 引言

说话人识别也称声纹识别, 可应用于系统安全接入、身份确认、电子侦听、司法鉴定及需要身份鉴别的相关领域, 具有广泛的社会需求和广阔的应用前景。目前此技术在实际应用中仍然存在困难, 主要原因在于系统易受背景噪声干扰, 稳定性欠佳。

近些年针对说话人识别系统的噪声问题依然是国际上研究的热点问题之一, 通常可采取三种途径加以解决: 一是通过前端增强处理, 提取尽可能纯净的语音信号, 如使用谱减、维纳滤波、子空间、小波等前端增强方法<sup>[1-3]</sup>; 二是对提取的特征进行处理, 如使用倒谱均值规整(CMS)<sup>[4]</sup>、RASTA 滤波<sup>[5]</sup>、特征规整(Feather Mapping)<sup>[6]</sup>、特征变换 HLDA<sup>[7]</sup>及多层特征融合<sup>[8]</sup>; 三是通过对模型和得分进行处理, 如使用统一背景模型(UBM)<sup>[9]</sup>、Z-Norm 和 T-Norm

等得分规整<sup>[10,11]</sup>。实际系统通常会级联以上多种方法提高系统在噪声下的稳健性<sup>[12]</sup>。然而以上方法多有缺陷, 主要表现在前端增强方法虽然在噪声下系统的性能有了明显提高, 然而在高信噪比下系统性能却出现下降, 说明现有的单通道语音增强算法通常都会对语音造成一定的损伤<sup>[13]</sup>, 使语音的部分信息丢失, 当然若使用多通道语音增强方法在很大程度上可以解决语音损伤问题<sup>[14, 15]</sup>, 但多通道语音往往不易获得, 且在处理上相对复杂; 而特征处理方法也存在类似问题, 如广泛使用的 CMS 方法, 有学者已经证明此方法在克服信道影响和加性噪声的同时, 也会使长时平均谱特征丢失<sup>[16]</sup>, 依然会造成高信噪比时识别率下降的后果, 而 RASTA 滤波也存在着类似问题; 模型方法中普遍使用的 UBM 模型有着广泛应用, 具有良好的噪声鲁棒性, 但在实际应用中需要大量复杂背景语音训练, 如司法鉴定中, 对比两段语音是否系同一人时, 供使用的语音仅有两组, 没有大量的说话人语音构成 UBM 模型, 此时 UBM 模型方法无法应用, 而得分规整方法作用效果有限, 不足以大幅提高识别系统的性能。此外, 还有一些学者利用长时特征<sup>[17]</sup>及研究针对噪声性质<sup>[18]</sup>的其它抗噪方法等。

收稿日期: 2008-04-15; 修回日期: 2008-06-19

基金项目: 广西自然科学基金资助项目(0639028)

作者简介: 关海欣(1982-), 男, 辽宁辽阳人, 满族, 硕士研究生, 主要研究方向为语音信号处理、模式识别、阵列信号处理。

通讯作者: 关海欣, E-mail: lantian\_guan@yahoo.com.cn

本文根据语音信号的非平稳特性, 以语音信号在各个子带中所含能量不均匀、所含信息量不均匀为前提假设, 撇开传统的噪声下稳健识别方法, 分析不同子带语音信号所携带信息的充分性和抗噪能力, 在文中信息的充分性以高信噪比时系统的识别率来衡量, 抗噪能力以低信噪比时系统识别率衡量, 从试验结果中发现本文的前提假设是正确的, 个别子带信号中含有的信息足够对说话人个体进行区分, 且抗噪性能良好, 通过适当的改进, 本文方法大幅提高了识别系统的性能。

## 2 基于 GMM 的说话人识别系统

### 2.1 GMM 模型表示

一个  $M$  阶的高斯模型的概率密度函数是由  $M$  个高斯概率密度函数加权求和得到的, 即:

$$p(\mathbf{X}|\lambda) = \sum_{i=1}^M p_i b_i(\mathbf{x}) \quad (1)$$

其中,  $\mathbf{x}$  是一个  $D$  维的特征向量;  $p_i (i=1, \dots, M)$  为混合权重, 且  $\sum_{i=1}^M p_i = 1$ ;  $b_i(\mathbf{x})$  为  $D$  维高斯函数, 即  $b_i(\mathbf{x}) = N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ 。其中,  $\boldsymbol{\mu}_i$  为均值向量,  $\boldsymbol{\Sigma}_i$  为协方差矩阵。完整的高斯混合模型由各均值向量、协方差矩阵及混合权重组成, 如式(2)所示:

$$\lambda = \{p_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\} \quad i=1, 2, \dots, M \quad (2)$$

### 2.2 GMM 模型参数估计

GMM 模型的训练就是给定一组训练数据, 依据某种准则确定模型的参数, 此处使用的参数估计方法是最大似然估计。对一组长度为  $T$  的训练矢量序列  $\mathbf{X} = (X_1, X_2, \dots, X_T)$ , GMM 似然度可表示为:

$$p(\mathbf{X}|\lambda) = \prod_{i=1}^T p(x_i|\lambda) \quad (3)$$

训练的目的就是找到一组参数  $\lambda$ , 使  $p(\mathbf{X}|\lambda)$  最大。对  $\lambda$  的估计首先使用 K-means 算法获得聚类初始值, 然后使用 EM 算法估计出新的参数  $\hat{\lambda}$ , 使新的模型参数下似然度  $P(\mathbf{X}|\hat{\lambda}) \geq P(\mathbf{X}|\lambda)$ 。新的模型参数继续训练, 这样迭代计算, 直到模型收敛, 以下为各参数估计公式。

混合权重估计公式:

$$P_i = \frac{1}{T} \sum_{i=1}^T p(i|x_i, \lambda) \quad (4)$$

均值重估公式:

$$\mu_i = \frac{\sum_{i=1}^T p(i|x_i, \lambda) x_i}{\sum_{i=1}^T p(i|x_i, \lambda)} \quad (5)$$

方差重估公式:

$$\sigma_i^2 = \frac{\sum_{i=1}^T p(i|x_i, \lambda) x_i^2}{\sum_{i=1}^T p(i|x_i, \lambda)} - \mu_i^2 \quad (6)$$

训练数据落在假定隐状态  $i$  的概率为:

$$p(i|x_i, \lambda) = \frac{p_i b_i(x)}{\sum_{k=1}^M p_k b_k(x)} \quad (7)$$

### 2.3 GMM 模型识别

给定一个语音样本, 说话人识别的目的是要确定这个语音属于  $N$  个说话人中的哪一个。在一个闭集的说话人识别集合中, 只要确认该语音属于语音库中的哪一个说话人。假定一个说话者  $i^*$  所对应的模型  $\lambda_i^*$  使得待识别语音特征矢量组  $\mathbf{X}$  具有最大后验概率  $P(\lambda_i^*|\mathbf{X})$ , 则  $i^*$  即为识别出的说话人。根据 Bayes 理论, 最大后验概率可表示为:

$$P(\lambda_i^*|\mathbf{X}) = \frac{P(\mathbf{X}|\lambda_i^*)P(\lambda_i^*)}{P(\mathbf{X})} \quad (8)$$

由于先验概率  $P(\lambda_i)$  未知, 这里假设该语音信号出自闭集的每个人的可能性相等, 即  $P(\lambda_i) = 1/N$ , 对于一个确定观察值矢量  $\mathbf{X}$ ,  $P(\mathbf{X})$  是一个确定的常数值, 对所有的说话人都相等。因此, 求取后验概率的最大值可通过求取  $P(\mathbf{X}|\lambda_i)$  获得, 这样辨认该语音属于语音库中的哪个说话人可以表示为:

$$i^* = \operatorname{argmax}_i P(\mathbf{X}|\lambda_i) \quad (9)$$

## 3 子带信号分析

主要以试验为基础, 验证语音信号各个子带对识别的贡献。

### 3.1 语音数据采集

使用一鹅颈式麦克风, 通过与计算机声卡相连进行语音采集, 采集软件使用 cooledit, 采样率为 16kHz, 采样精度为 16bit, 在一个安静的实验室内进行采集, 共采集 19 人语音, 每人约 2min, 其中男性 12 人, 女性 7 人, 语音内容为一段散文, 最后使用 cooledit 软件手工切除无声段, 形成每人一段长 30s 的语音作为训练语音和 5 段长 3s 的测试语音(共 95 段测试语音)。

### 3.2 试验说明

试验中使用第 2 节中介绍的 GMM 模型分类器, 模型阶数  $M$  取 32, 与训练语音长度相关的最佳模型阶数尚无定论, 有兴趣的读者可参考文献 [19] 所得的一些试验数据, 特征由 12 阶 MFCC 和

12 阶 MFCC 的一阶差分  $\Delta$ MFCC 组成,提取特征前对语音信号做了分帧、加窗处理,帧长 256 点,帧移 128 点,加汉明窗,注意此处并未做预加重处理。

第一个试验测试各个子带对噪声的抵抗能力,对 NOISE92 噪声库的白噪声进行了滤波处理,使用 50 阶 FIR 滤波器分别得到从 0~8000Hz 共 16 个子带噪声,每个子带噪声带宽为 500Hz,而后将各子带噪声加入测试语音送入识别系统进行测试。表 1 中的 1、2...16 代表 16 个子带。

第二个试验主要测试语音信号各个子带所含信息的充分性和抗噪性,此时对训练和测试语音都进行滤波处理,所用方法与第一个试验中对白噪声的处理相同。

第一个试验结果如表 1 和图 1 所示。

表 1 不同子带噪声、不同信噪比下系统正确识别率  
Table 1 The system's recognition rate at different sub-band noise and different SNRs

字带 序号	识别率					
	30dB	25dB	20dB	15dB	10dB	5dB
1	1	1	1	1	1	0.9684
2	1	0.9895	0.9684	0.9474	0.9158	0.7895
3	1	0.9895	0.9653	0.7790	0.6316	0.4421
4	1	0.9368	0.8642	0.8211	0.7263	0.5158
5	1	0.9895	0.8737	0.6632	0.5053	0.3895
6	0.9579	0.9158	0.7895	0.5158	0.4316	0.3368
7	0.9790	0.8105	0.6105	0.5474	0.4316	0.2632
8	0.9895	0.8737	0.6211	0.5263	0.3895	0.2737
9	0.9790	0.9790	0.4737	0.2947	0.2211	0.1368
10	0.9474	0.7263	0.4842	0.3579	0.2	0.0632
11	0.9895	0.7368	0.4947	0.3684	0.3053	0.2421
12	0.8526	0.6105	0.4211	0.3579	0.2526	0.2316
13	0.9474	0.7474	0.5895	0.4947	0.4	0.3158
14	0.9579	0.8526	0.6421	0.4421	0.2947	0.2526
15	0.9158	0.6632	0.4526	0.3474	0.3263	0.2737
16	0.9895	0.9263	0.6947	0.5158	0.4105	0.3474

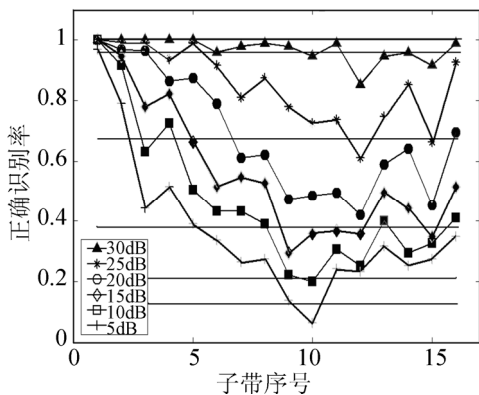


图 1 不同子带噪声、不同信噪比下系统正确识别率  
Fig.1 The system's recognition rate at different sub-noise and different SNR.

在试验一中发现,不同子带噪声对系统的影响是不同的,且差别巨大。第一个子带噪声对识别系

统的影响最小,其次是第二个子带噪声。这说明语音信号在 0~500Hz 区间所含能量丰富,不易被噪声干扰,500Hz~1000Hz 所含能量次之,若单纯使用 0~500Hz 或 500Hz~1000Hz 语音信号能够在高信噪比时得到很高的识别率,则说明此频段信号含有的说话人个性信息能够满足识别需要,再借助此频段语音能量高、不易被干扰的事实,必然能得到更好的识别性能。图 1 中 6 条横线为宽带白噪声在不同信噪比下系统的识别率,第二个试验中出现的横线与此处相同,具体识别率数据将在第 4 节给出。

第二个试验结果如表 2 和图 2 所示。

表 2 不同子带信号在不同信噪比下系统正确识别率  
Table 2 The system's recognition rate at different sub-band signal and different SNRs.

字带 序号	识别率					
	30dB	25dB	20dB	15dB	10dB	5dB
1	0.9789	0.9789	0.9579	0.8947	0.7474	0.4947
2	0.9684	0.9684	0.9158	0.7368	0.5789	0.3474
3	0.9789	0.9579	0.7789	0.4842	0.2947	0.1579
4	0.9895	0.9158	0.7053	0.4211	0.2737	0.1158
5	0.9684	0.9368	0.7158	0.4737	0.2	0.0947
6	0.9789	0.9368	0.7158	0.5368	0.3579	0.1474
7	0.9895	0.9158	0.6632	0.4211	0.2632	0.1895
8	0.9895	0.8947	0.6632	0.2947	0.1579	0.1474
9	0.9579	0.8632	0.5684	0.3368	0.1579	0.0526
10	0.8947	0.6421	0.3579	0.2	0.1053	0.0737
11	0.9053	0.6632	0.4105	0.2	0.1789	0.1053
12	0.9263	0.6	0.4421	0.1789	0.1053	0.0947
13	0.8842	0.6737	0.3263	0.0842	0.0526	0.0526
14	0.9368	0.8	0.2842	0.0526	0.0526	0.0526
15	0.9053	0.6	0.2526	0.0526	0.0526	0.0526
16	0.9368	0.7895	0.4211	0.2947	0.2	0.1053

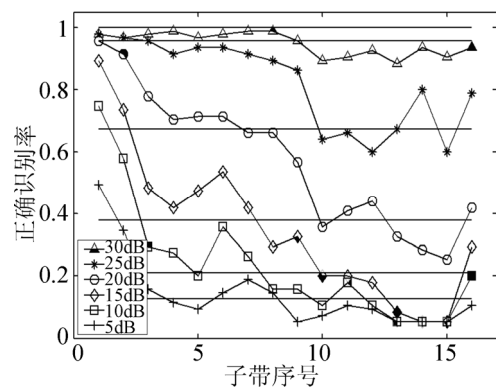


图 2 不同子带信号在不同信噪比下系统正确识别率  
Fig.2 The system's recognition rate at different sub-band signal and different SNRs.

在第二个试验中,不同子带信号对噪声的抵抗能力和携带信息的多少也是不相同的,以第一个子带信号抗噪性最佳,其次是第二子带信号,这与第一个试验是高度吻合的,并且在 30dB 高信噪比时观察发现,前 8 个子带信号所携信息相当,高于后

8 个子带信号所携带的信息。并且随着信噪比的降低, 系统的正确识别率基本呈现低频高、高频低的特点, 可以想象, 当信噪比较低时, 保留高频信息不仅不利于系统的正确辨识, 而且会对系统造成危害。

从试验一、二中发现第一个子带信号所携带的信息和抗噪性能均较佳, 其次是第二子带, 此处又对 0~800Hz 子带信号做了进一步分解, 每个子带带宽 100Hz, 子带间重叠 50Hz, 即第一个子带为 0~100Hz, 第二子带为 50Hz~150Hz, 第三子带为 100Hz~200Hz, 依次类推, 试验结果如表 3 和图 3 所示。

表 3 不同子带信号在不同信噪比下的正确识别率  
Table 3 The system's recognition rate at different sub-band signal and different SNRs.

序号	识别率					
	30dB	25dB	20dB	15dB	10dB	5dB
1	0.9789	0.9789	0.9684	0.9474	0.8947	0.7684
2	0.9579	0.9579	0.9579	0.9263	0.8211	0.6737
3	0.9579	0.9579	0.9579	0.9053	0.8	0.5789
4	0.9684	0.9684	0.9579	0.9263	0.7579	0.5158
5	0.9579	0.9579	0.9579	0.9158	0.8316	0.5053
6	0.9684	0.9684	0.9684	0.9368	0.7895	0.4421
7	0.9263	0.9263	0.9263	0.9158	0.8	0.4526
8	0.9789	0.9789	0.9579	0.9158	0.6526	0.4316
9	0.9684	0.9684	0.9368	0.8316	0.6947	0.4632
10	0.9684	0.9684	0.9368	0.8526	0.7263	0.4632
11	0.9684	0.9789	0.9474	0.8737	0.7368	0.4632
12	0.9895	0.9789	0.9474	0.8526	0.7263	0.4632
13	1	0.9789	0.9474	0.8	0.6211	0.4211
14	0.9789	0.9684	0.9158	0.7368	0.5684	0.2421
15	0.9684	0.9579	0.8947	0.7053	0.4632	0.2

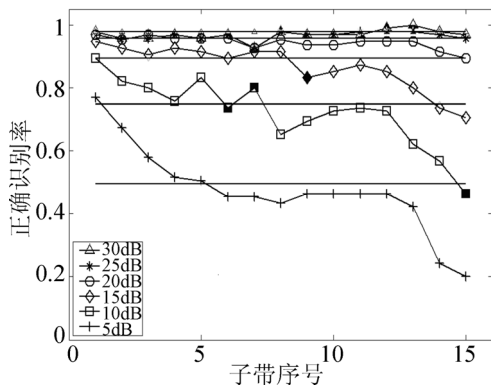


图 3 不同子带信号在不同信噪比下的正确识别率  
Fig.3 The system's recognition rate at different sub-band signal and different SNRs.

图 3 中, 横线为 0~500Hz 子带信号在不同信噪比下的正确识别率曲线。实验中发现, 前 5 个子带即 0~300Hz 信号具有比 0~500Hz 信号更好的或相当的抗噪性及所携带的个性信息, 尤以第一子带即 0~100Hz 最佳, 而随着频率的升高, 系统的识别率不断降低, 到达 650Hz 以后性能开始恶化。

### 4 子带信号进一步处理

在第 3 节的试验中发现 0~100Hz 子带信号不仅具有着较好的抗噪性能, 并且所携带的个性信息也不比 0~500Hz 子带信号所携带的信息少, 然而在高信噪比下依然不能达到 100% 的正确识别率, 此时考虑口、鼻腔的辐射特性, 声道所发语音通过辐射后按 6dB/oct 衰减, 通常采用预加重处理, 前述部分未加入预加重处理主要是考虑避免语音和噪声特性被改变, 影响试验结果, 预加重采用如下公式:

$$y(n)=x(n)-\alpha x(n-1) \tag{10}$$

$\alpha$  通常取 0.95~1 之间的值, 此处取 0.97。其幅频特性如图 4 所示:

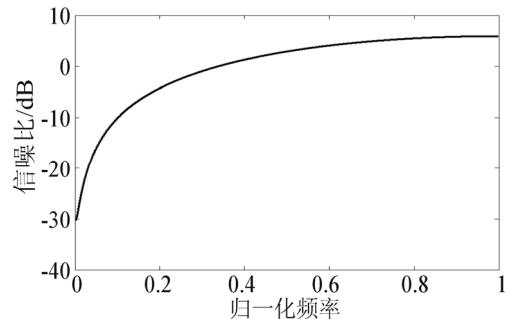


图 4 预加重滤波的幅频特性曲线  
Fig.4 The frequency response of the emphasize filter

将 0~100Hz 的子带信号再次使用预加重滤波处理, 为了进行对比, 将滤取 0~100Hz 所用的低通滤波器幅频响应图一同给出, 见图 5、6。

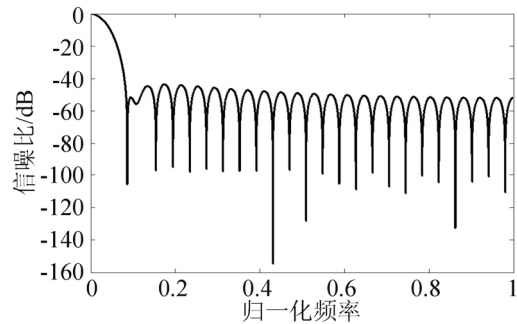


图 5 低通滤波器幅频特性曲线  
Fig.5 The frequency response of the low pass filter

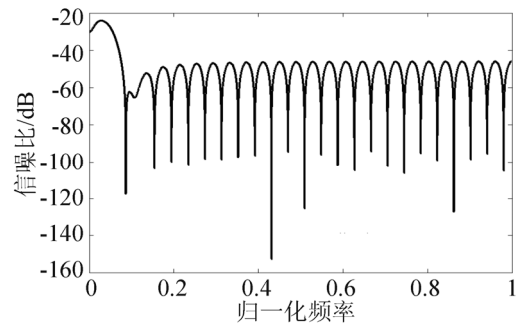


图 6 再经预加重滤波的幅频特性曲线  
Fig.6 The frequency response after emphasize filter

在前述部分图形中给出了全带信号的识别率结果，而未给出具体数值，将全带信号、全带加重信号、子带信号及子带加重信号识别率分别给出，其识别结果如表 4 所示：

表 4 全带子带的识别率对比

Table 4 The comparison between sub-band signal and origin signal

	识别率			
	全带	全带加重	子带	子带加重
30dB	0.9895	1	0.9789	1
25dB	0.9579	0.9579	0.9789	1
20dB	0.6947	0.6737	0.9684	0.9684
15dB	0.3579	0.3790	0.9474	0.7579
10dB	0.2211	0.2105	0.8947	0.4526
5dB	0.1474	0.1263	0.7684	0.1790

经过加重处理后的信号在高信噪比时有着更好的识别效果，说明加重信号的确更加符合声道特性，然而，也正是由于预加重使得能量丰富的低频段信号被削弱，信号更容易被噪声所淹没。而经过子带加重的信号无论在高信噪比时还是低信噪比时均比全带信号和全带加重信号有着更好的识别性能，说明此子带信号所含个性信息充分性较好，并且抗噪性能优于全带信号，但相比于单纯子带信号，其低信噪比时性能下降十分严重，所以还有待于进一步改进。

考虑到单纯子带信号具有很好的抗噪性，若能联合子带信号和子带加重信号，有可能进一步提高系统的识别性能。基于此，考虑子带信号所提取参数为  $C_1$ ，子带加重信号所提取参数为  $C_2$ ，组成新的参数  $C$ ：

$$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \quad (11)$$

除此之外，将特征扩展为 16 阶 MFCC 加 16 阶  $\Delta$ MFCC 共 32 阶特征参数，和 16 阶 MFCC 加 16 阶  $\Delta$ MFCC 加 16 阶  $\Delta\Delta$ MFCC 共 48 阶特征参数，作为对比，其识别性能如表 5 和图 7 所示。

表 5 子带信号的进一步改进

Table 5 The improvement of the sub-band signal

	sub+pre-ep	sub+pre-ep32	sub+pre-ep48
30dB	1	1	1
25dB	1	1	1
20dB	0.9895	1	1
15dB	0.9368	0.9789	0.9789
10dB	0.8316	0.8632	0.9053
5dB	0.5474	0.6632	0.6316

在高信噪比下，新的参数与子带加重信号相比识别率并没有下降，而低信噪比下识别率却有了大幅提升，虽然与只用于子带信号时相比在 5dB 信噪比下有所下降，但性能已经明显优于传统全带加重方

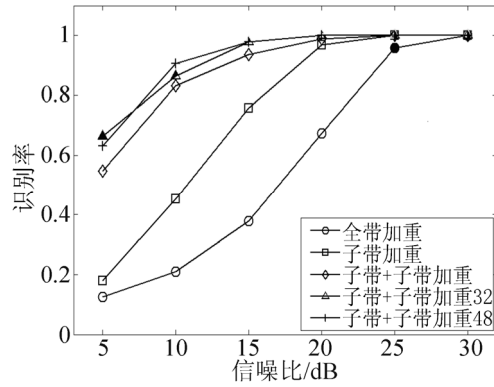


图 7 子带信号的进一步改进

Fig.7 The improvement of the sub-band signal

法，此时通过加入高阶特征参数及参数差分，使识别系统在 10dB 时的识别率超过 90%，远远高于传统的全带加重信号，为实际应用奠定了基础。

## 5 结 论

本文针对说话人识别系统易受噪声干扰问题，从子带信号分析入手，研究各个子带信号所含能量的大小和所携带信息的多少，以大量的试验为依据，最终发现语音信号在 100Hz 以下含有大量的说话人个性信息，并且不易受到噪声干扰，通过对此子带信号的进一步加重处理、特征融合，使系统的性能有了进一步提高，在信噪比为 10dB、并且不使用其它抗噪方法时系统的正确识别率就已超过 90%。就其本质来说，本文方法可归入前端处理方法，若再联合其它有效的特征、信道、模型等补偿方法，完全有理由相信可以进一步提高识别系统的性能。

## 参 考 文 献

- [1] BAI, Rong Junmei, XU Zhengbo, ZHANG Shuwu. Robust speaker recognition intergrating pitch and wiener filter[A]. IEEE ISCSLP[C]. 2004, 7: 69-73.
- [2] Jia Ching Wang, Chung Hsien Yang, Jhing Fa Wang and Hsiao Ping Lee. Robust speaker identification and verification[J]. IEEE COMPUTATIONAL INTELLIGNC- E, 2007, 5(2): 52-59.
- [3] Aliaa A. Youssif, Ebada A. Sarban. Development of automatic speaker identification system[A]. 21<sup>st</sup> National Radio Science Conference[C]. 2004, C7: 1-8.
- [4] Naik D. Pole-filtered cepstral mean subtraction[J]. IEEE ECASSP, 1995, 5(1): 157-160.
- [5] 甄斌, 吴玺宏, 刘志敏, 迟惠生. 语音信号的增强相对谱滤波[J]. 声学学报, 2001, 26(3): 252-258.
- [6] Reynolds D A. Channel robust speaker verification via feature mapping[A]. Proc. ICASSP[C]. 2003, 4(II): 53-56.
- [7] Stemmer G, Brugnara F. Integration of Heteroscedastic Linear Discriminant Analysis (HLDA) into Adaptive Training[J]. IEEE ICASSP, 2006, 5(I): 1-4.
- [8] Reynolds D A. The super SID project: Exploiting high-level infor-

- mation for high-accuracy speaker recognition[A]. In Proc. IEEE ICASSP[C]. Hong Kong, 2003.
- [9] Reynolds D A. Speaker verification using adapted Gaussian mixture models[J]. Digital Signal Processing, 2000, (10): 19-41.
- [10] Barras C Gauvain. Feature and score normalization for speaker verification of cellular data[J]. IEEE ICASSP, 2003, 4(2): 49-52.
- [11] The official website for the NIST speaker recognition evaluations [EB/OL]. <http://www.nist.gov/speech/tests/spk>.
- [12] Matějka P, Burget L. STBU System for the NIST 2006 Speaker Recognition Evaluation[J]. IEEE ICASSP, 2007, 7(IV): 221-224.
- [13] Chen J, Benesty J, Huang Y, Doclo S. New insights into the noise reduction wiener filter[J]. IEEE. Trans. Audio, Speech and Language Processing, 2006, 14(4): 1218-1234.
- [14] 关海欣, 曾庆宁. 多通道子空间算法在说话人识别中的应用[J]. 声学技术, 2008, 27(3): 396-402.
- GUAN Haixin, ZENG Qingning. Application of multi-channel subspace algorithm to speaker recognition[J]. Technical Acoustics, 2008, 27(3):296-402.
- [15] McCowan I, Pelecanos J, Sridharan S. Robust speaker recognition using microphone arrays[J]. Proceedings of 2001: A speaker odyssey, June 2001.
- [16] Douglas A. Reynolds and Richard C. Rose. Robust text-independent speaker identification using gaussian mixture models. [J] IEEE Transactions on Speech and Audio Processing, 1995, 3 (1): 72-83.
- [17] 梁振福, 李平. 环境噪声对语言识别率的影响[J]. 声学技术, 1999, 18(3): 27-31.
- [18] 王宏, 向大威. 基于长时平均频谱的“文本无关”话者识别[J]. 声学技术, 2002, 21(1-2): 59-62.
- [19] Bhattacharyya S, Srikanthan T. Ideal GMM parameters & posterior log likelihood for speaker verification[J]. IEEE Communication Magazine, 2001, 1(8): 471-480.