

基于鲁棒听觉特征的说话人识别

林 琳, 陈 虹, 陈 建

(吉林大学通信工程学院, 吉林长春 130022)

摘 要: 为了提高噪声环境中说话人识别系统的性能, 本文提出了一种鲁棒听觉特征提取的算法, 并将其应用到说话人识别系统中. 运用自适应压缩 Gammachirp 滤波器组模拟人耳耳蜗的听觉特性, 对输入的语音信号进行频域子带滤波, 将得到的对数子带能量作为听觉特征参数. 分别运用离散余弦变换和核主成分分析方法, 对提取的特征参数进行特征变换, 降低特征参数的维数, 提高特征参数的噪声鲁棒性和个性表现力. 实验结果表明, 将提取的新听觉特征参数应用到说话人识别系统中, 新特征参数在鲁棒性和识别性能上均优于梅尔倒谱系数和基于 Gammatone 的听觉特征参数.

关键词: 说话人识别; 自适应压缩 Gammachirp 滤波器; 核主成分分析; 特征提取

中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2013)03-0619-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2013.03.034

Speaker Recognition Based on Robust Auditory Feature

LIN Lin, CHEN Hong, CHEN Jian

(College of Communication Engineering, Jilin University, Changchun, Jilin 130022, China)

Abstract: In order to improve the performance of speaker recognition system in noisy environment, this paper presents an auditory feature extraction algorithm. It used adaptive compression Gammachirp filter banks to simulate the auditory characteristics of human cochlea, and the input speech signal was sub-band filtered in the frequency-domain. After logarithmic transformation, it can get the logarithmic sub-band energy as the auditory feature parameter. It respectively used discrete cosine transform and kernel principal component analysis method to transform the auditory feature and get the two new auditory features, which not only can reduce the dimension of the feature parameters, but also can improve the robustness and personality expression of feature parameters. The experimental results show that speaker recognition system with the new auditory feature parameters can get the better results in the robustness and recognition performance than Mel cepstral coefficients and auditory feature parameters based on Gammatone filter banks.

Key words: speaker recognition; adaptive compression Gammachirp filter; kernel principal component analysis; feature extraction

1 引言

鲁棒特征参数的提取是提高实际应用环境中说话人识别系统性能的关键. 人耳具有较强的噪声鲁棒性, 能够在极低的信噪比条件下具有较好的识别能力, 因此, 利用人耳听觉特性提取的 Mel 频率倒谱系数 (Mel-Frequency Cepstrum Coefficient, MFCC) 及衍生特征是目前说话人识别中最为普遍的鲁棒性特征. 尽管 MFCC 已经部分考虑了人的听觉感知特性, 但在低信噪比条件下性能不理想.

为了充分利用人耳的听觉鲁棒性, Yang^[1]等利用模

拟耳蜗基底膜特性的 Gammatone 听觉滤波器提取语音特征参数. Qi Li 等^[2]在 Yang 的基础上研究了 Gammatone 滤波器组的可逆变换, 提取的听觉特征进一步提高了低信噪比下系统的性能. 王 ■^[3]等对基于 Gammatone 滤波器组的特征参数提取进行了研究, 提高了系统鲁棒性. 但 Gammatone 滤波器的幅频响应是关于中心频率对称的^[4,5], 无法体现基底膜曲线的非线性, 同时 Gammatone 函数中没有强度相关的参数, 其幅频响应曲线是与强度无关的, 不符合基底膜的强度相关特性. 因此, Irino 和 Patterson^[6]提出了一种能够描述耳蜗基底膜对于频率响应非线性特性的 Gammachirp 听觉滤波器. 然而, 研究表

明,听觉滤波器是高度的非线性和动态的,尤其是听觉滤波器的频率响应展现出了等级相关的非对称性^[7]和输入/输出的压缩特性^[8],而这些性质归根结底是由滤波器的动态特性决定的.听觉滤波器的非线性和动态特性会提高人耳在噪声环境下的鲁棒性,因此,本文采用一种基于自适应压缩的 Gammachirp 滤波器组模拟耳蜗的听觉滤波器,对输入的语音信号进行频域子带滤波,将得到的对数子带能量作为听觉特征参数.自适应压缩 Gammachirp 滤波器组能够根据信号的幅度对滤波器的参数进行自适应调整,使滤波器能够更好地体现动态和非线性特性,提高听觉参数的鲁棒性.分别运用离散余弦变换和核主成分分析方法,对提取的特征参数进行特征变换,降低特征参数的维数,进一步提高特征参数的噪声鲁棒性和个性表现力.

2 自适应压缩 Gammachirp 滤波器

2.1 压缩 Gammachirp 滤波器

Gammachirp 滤波器的冲激响应公式为

$$g_c(t) = at^{n_1-1} \exp(-2\pi b_1 \text{ERB}(f_{r1})t) \times \exp(j2\pi \cdot f_{r1}t + j\varphi_1 \ln t + \varphi_1) \quad (1)$$

其中,时间 $t > 0$, a 为振幅,阶次 n_1 与参数 b_1 是定义 Gamma 函数分布情况的参数, c_1 为啁啾因子,会随着声压位阶而变化, $\text{ERB}(f_{r1})$ 是中心频率为 f_{r1} 的听觉滤波器等效矩形带宽的值, φ_1 是初始相位.当啁啾因子 $c_1 = 0$ 时,式(1)退化为 Gammatone 函数的冲激响应.

心理生理学实验表明,耳蜗滤波器具有非线性压缩特性,因此, Irino 与 Patterson^[9] 利用一个表示被动基底膜的被动 Gammachirp (passive Gammachirp, pGC) 滤波器和反映耳蜗主动机制的强度自适应、高通非对称函数 (High-Pass Asymmetric Function, HP-AF) 构成压缩 Gammachirp 滤波器,定义其幅度谱为 $|G_{cc}(f)|$

$$|G_{cc}(f)| = |G_{cp}(f)| \cdot \exp(c_2 \theta_2(f)) \quad (2)$$

$$|G_{cp}(f)| = a_{\Gamma} \cdot |G_{\Gamma}(f)| \cdot \exp(c_1 \theta_1(f)) \quad (3)$$

其中, $|G_{cp}(f)|$ 和 $\exp(c_2 \theta_2(f))$ 分别表示被动 Gammachirp 滤波器 pGC 和高通滤波器 HP-AF 的频域表达式, $|G_{cp}(f)|$ 可由式(3)得到. c_1 和 c_2 为定值, a_{Γ} 是常数, $|G_{\Gamma}(f)|$ 是 Gammatone 滤波器的傅立叶幅度谱.若假设 pGC 和 HP-AF 滤波器的中心频率分别为 f_{r1} 和 f_{r2} , 则

$$\theta_1(f) = \arctan\left(\frac{f - f_{r1}}{b_1 \text{ERB}(f_{r1})}\right) \quad (4)$$

$$\theta_2(f) = \arctan\left(\frac{f - f_{r2}}{b_2 \text{ERB}(f_{r2})}\right) \quad (5)$$

HP-AF 滤波器的中心频率 f_{r2} 由频率比率 f_{rat} 和被动 Gammachirp 滤波器的峰值频率 f_{p1} 的乘积计算得到.

$$f_{r2} = f_{\text{rat}} \times f_{p1} \quad (6)$$

其中

$$f_{\text{rat}} = f_{\text{rat}}^{(0)} + f_{\text{rat}}^{(1)} \times P_{\text{gcp}} \quad (7)$$

$$f_{p1} = f_{r1} + c_1 b_1 \text{ERB}(f_{r1}) / n_1 \quad (8)$$

式(7)中, $f_{\text{rat}}^{(0)}$ 和 $f_{\text{rat}}^{(1)}$ 为常数,文中分别取 0.446 和 0.0109. P_{gcp} 是被动 Gammachirp 滤波器输出的总强度,被用来控制高通非对称函数 HP-AF 的位置.在压缩 Gammachirp 滤波器中,被动 Gammachirp 滤波器的时域表达式可以用式(1)来表示,而高通非对称函数对应的高通滤波器没有相应的时域表达,文中采用 Irino 和 Unoki^[9] 提出的方法得到高通滤波器的时域实现.

2.2 自适应压缩 Gammachirp 滤波器

自适应压缩 Gammachirp 滤波器包括两个过程:强度估计过程和信号滤波过程,其结构如图 1 所示.图中上行路径是强度估计过程,它由一个被动 Gammachirp 滤波器 pGC 和一个高通滤波器 HP-AF 串联组成.输入信号首先通过一个参数为 b_1 、 c_1 、 f_{r1} 、 f_{p1} 的被动 Gammachirp 滤波器 pGC, f_{p1} 表示 pGC 滤波器的峰值频率,满足

$$\text{ERB}_{\text{rate}}(f_{p1L}) = \text{ERB}_{\text{rate}}(f_{p1}) + r_{\text{EL}} \quad (9)$$

其中, $\text{ERB}_{\text{rate}}(f)$ 是在频率 f 处的等效矩形带宽比率, r_{EL} 表示在 ERB 比率轴上两个 pGC 滤波器之间频率的间隔.

然后,将 pGC 的输出进行 HP-AF 滤波器滤波.为了更好地考虑耳蜗上层抑制和下层抑制区域增长的不同比率,利用 pGC 和 HP-AF 滤波器在强度估计路径中的输出,对控制强度 P_c 进行估计,从而使得 HP-AF 的参数能够根据控制强度 P_c 的不同进行自适应调节,达到自适应滤波的目的.

假设 pGC 和 HP-AF 在强度估计路径中的输出分别是 $s_1(t)$ 和 $s_2(t)$, 则估计的线性等级 $\bar{s}_1(t)$ 和 $\bar{s}_2(t)$ 定义为

$$\bar{s}_1(t) = \max\{\bar{s}_1(t - \Delta t) \cdot e^{-\ln 2 \cdot (\Delta t / \tau_L)}, \max(s_1(t)), 0\} \quad (10)$$

$$\bar{s}_2(t) = \max\{\bar{s}_2(t - \Delta t) \cdot e^{-\ln 2 \cdot (\Delta t / \tau_L)}, \max(s_2(t)), 0\} \quad (11)$$

其中 Δt 是采样时间, τ_L 是指数式衰减的半衰期.这是一种“快速产生慢衰减”的强度估计形式,当信号强度增加时,估计的强度跟踪滤波器的正输出,但在达到峰值之后,估计值不再跟踪滤波器的输出,而是以半衰期进行衰减.控制强度 $P_c(t)$ 定义为线性强度的加权和

$$P_c(t) = 20 \log_{10} \left\{ w_L \cdot a_{\text{RL}} \left(\frac{\bar{s}_1(t)}{a_{\text{RL}}} \right)^{v_{\text{L}}} + (1 - w_L) \cdot a_{\text{RL}} \left(\frac{\bar{s}_2(t)}{a_{\text{RL}}} \right)^{v_{\text{H}}} \right\} \quad (12)$$

$$a_{\text{RL}} = 10^{P_{\text{RL}}/20} \quad (13)$$

其中 w_{1L}, v_{1L}, v_{2L} 是加权参数, P_{RL} 是参数 a 的参考强度, 单位是 dB.

图 1 中的下行路径是信号滤波过程, 信号滤波过程与强度估计过程的组成基本相同, 也是由一个被动 Gammachirp 滤波器 pGC 和一个高通非对称滤波器 HP-AF 串联组成, 这里 pGC 的参数有 b_1, c_1, f_{r1}, f_{p1} , HP-AF 的参数有 $b_2, c_2, f_{r2} (f_{r2} = f_{rat} * f_{p1})$, 两个滤波器参数的取值与前面的设置基本相同. 两条路径唯一的区别就是两个过程中的频率比率不同, 强度估计过程中, 频率比率 f_{ratL} 是强度独立的, 取 $f_{ratL} = 1.08$; 而对于信号滤波过程中的频率比率 f_{rat} , 它是主要的强度自适应参数, 需要利用控制强度 P_c 代替式(7)中的 P_{gcp} 计算得到. 由于滤波器组在信号滤波过程中, HP-AF 参数能够根据控制强度 P_c 进行自适应调节, 从而使滤波器组的输出峰值频率 f_{p2} 也是强度自适应的, 因此, 将上述滤波器组称为自适应压缩 Gammachirp 滤波器组 AcGC (Adaptive Compression Gammachirp).

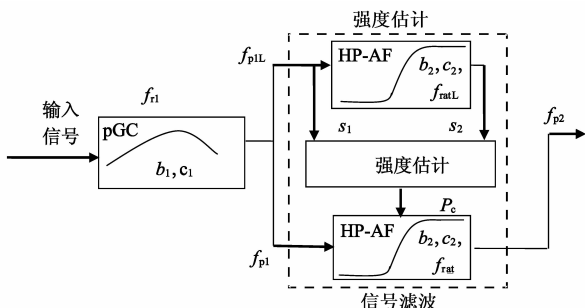


图1 自适应压缩Gammachirp滤波器结构图

3 核主成分分析

核主成分分析 (Kernel Principal Component Analysis, KPCA) 是一种简洁高效的非线性特征抽取方法, 其算法执行过程如下:

(1) 给定数据集 $S = \{x_1, x_2, \dots, x_N\}$, 计算 $N \times N$ 核矩阵

$$K = (K_{ij})_{N \times N} = K(x_i, x_j); \quad (14)$$

(2) 计算特征空间对映射数据进行中心化处理后的核矩阵 \tilde{K}

$$\tilde{K} = K - \mathbf{1}_N K - K \mathbf{1}_N - \mathbf{1}_N K_N \quad (15)$$

其中 $\mathbf{1}_N$ 是元素全部为 $\frac{1}{N}$ 的 $N \times N$ 维矩阵.

(3) 求解矩阵 \tilde{K} 的特征值和特征向量 $\tilde{K}A = A\Lambda$, 这里 Λ 是核矩阵 \tilde{K} 的特征值矩阵, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$, 其中 λ_k 是矩阵最小非零特征值; $A = [\alpha_1, \alpha_2, \dots, \alpha_N]$ 是核矩阵 \tilde{K} 特征值矩阵所对应的特征向量矩阵.

(4) 对任意输入样本 x_i 提取主成分, 计算样本 x_i 变换后的数据

$$y_i^l = \frac{1}{\sqrt{\lambda_l}} \sum_{j=1}^N \alpha_j^l K(x_j, x), l = 1, 2, \dots, k \quad (16)$$

4 基于鲁棒听觉特征参数的说话人识别

4.1 鲁棒听觉特征参数的提取

鲁棒听觉特征其算法框图如图 2 所示. 图中, 输入语音经过预处理, 进行离散傅里叶变换 DFT, 利用自适应压缩 Gammachirp 滤波器组在频域上对信号进行子带滤波, 然后对数变换, 得到对数子带能量, 至此得到基于自适应压缩 Gammachirp 滤波器的听觉特征参数 AcGCF (Adaptive Compression Gammachirp Feature). 在 AcGCF 特征参数提取的过程中, 为了使特征参数具有较好的鲁棒性, 通常滤波器组的个数较大, 如 128 个或 64 个, 提取的特征参数维数大大超过了用于说话人识别的传统特征. 因此, 分别运用离散余弦变换和核主成分分析 KPCA 变换对 AcGCF 特征参数进行特征变换, 得到特征参数 AcGCF-DCT 和 AcGCF-KPCA. 特征变换不仅能够降低特征参数的维数, 而且能够去除噪声和特征分量之间的冗余信息, 降低计算的复杂度.

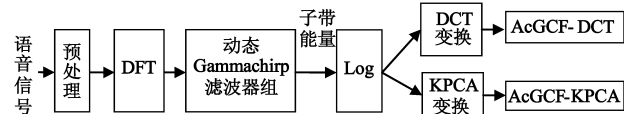


图2 鲁棒听觉特征参数提取算法框图

4.2 说话人识别系统的构成

基于 AcGCF-DCT 特征参数的说话人识别系统与传统说话人识别系统结构一样, 但基于自适应压缩 Gammachirp 滤波器组和 KPCA 的说话人识别系统比传统说话人识别系统多了两个模块: “计算映射空间主成分” 模块和 “KPCA 变换” 模块, 其系统框图如图 3 所示.

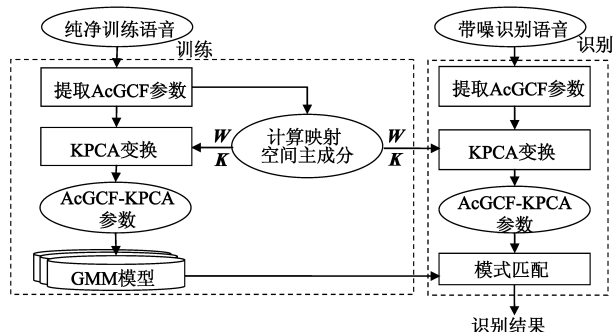


图3 基于KPCA变换的鲁棒听觉特征参数说话人识别系统框图

训练阶段, 对输入的纯净训练语音提取听觉特征参数 AcGCF, 然后利用 “计算映射空间主成分” 模块计算核矩阵 K 和核空间上的变换矩阵 W . 将核矩阵 K 和变换矩阵 W 输入到 “KPCA 变换” 模块, 计算训练特征参数 AcGCF-KPCA. 利用 AcGCF-KPCA 参数训练说话人模型, 文中使用高斯混合模型作为说话人模型. 最后, 将说话

人的高斯混合模型、核矩阵 \mathbf{K} 、变换矩阵 \mathbf{W} 、训练得到的 AcGCF 参数存储到说话人数据库中。

在识别阶段,将得到的识别语音特征参数 AcGCF 输入到“KPCA 变换”模块,利用训练得到的 AcGCF 参数、核矩阵 \mathbf{K} 、变换矩阵 \mathbf{W} 计算新的特征矢量,将新的特征矢量输入模式匹配与对应的说话人模型进行得分运算,根据决策规则得到识别结果。

5 仿真实验

实验中采用的语音数据为 PKU-SRSC 语音数据库^[10],进行与文本无关的说话人辨认实验.从数据库中任选 66 个说话人(24 个男生和 42 女生),使用每个说话人的 10 次录音数据,其中说话人的每次录音间隔为一周,每个说话人的每次录音数据包括 20 个语音文件.选择其中第一次录音的部分语音作为训练语音,其余的所有语音作为识别语音,每个识别语音约为 1s 左右.实验中采用高斯混合模型作为说话人的语音模型,模型混合度取 16。

实验 1 听觉滤波器组对识别性能的影响

为了验证自适应压缩 Gammachirp 听觉滤波器对听觉特征参数提取的有效性,分别使用 Mel 滤波器组、Gammatone 滤波器组、Gammachirp 滤波器组、压缩 Gammachirp 滤波器组、自适应压缩 Gammachirp 听觉滤波器对图 2 中 DFT 变换后的信号进行滤波,采用 DCT 变换,提取听觉参数 MFCC、Gammatone 特征参数 (Gammatone Feature, GTF)、Gammachirp 特征参数 (Gammachirp Feature, GCF)、压缩 Gammachirp 特征参数 (Compression Gammachirp Feature, cGCF)、自适应压缩 Gammachirp 特征参数 (Adaptive Compression Gammachirp Feature, AcGCF),在纯净语音和噪声语音下,分别考察不同听觉滤波器对听觉特征参数提取的影响.其中, MFCC 的维数是 20 维, GTF 的维数是 23 维, GCF、cGCF、AcGCF 的维数是 24 维.提取上述参数的一阶动态参数,去掉动静组合参数的第一维,将剩余的组合作为说话人的特征参数.从数据库中选取 20 个说话人,使用 NOISEX-92 噪声库中的白噪声信号,在信噪比分别为 20dB、15dB、10dB、5dB、0dB 条件下,进行说话人识别系统性能的比较,结果如图 4 所示。

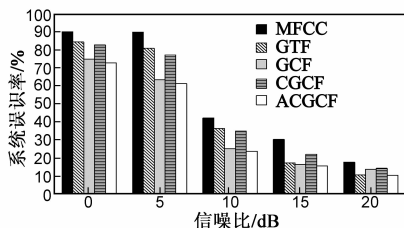


图4 不同信噪比下不同特征参数的系统误识率

从图 4 中可以看出,无论是在纯净语音还是在不同信噪比的噪声条件下,采用基于 Gammachirp 滤波器的听觉特征参数 GCF、CGCF、AcGCF 误识率均低于基于 Mel 滤波器、Gammatone 滤波器的特征参数,可见 Gammachirp 滤波器组更适合用于描述耳蜗基底膜对于频率响应的非线性特性.随着压缩特性和幅度自适应的加入,本文的自适应压缩 Gammachirp 滤波器能够更好地展现听觉滤波器的非线性和动态特性,因此,提取的 AcGCF 特征参数在纯净语音和不同信噪比条件下均能够得到最好的识别性能,误识率最低。

实验 2 核参数对识别结果的影响

采用 KPCA 对听觉特征参数 AcGCF 进行变换时,不同的核参数意味着将特征映射的高维空间不同, KPCA 变换后得到的特征参数的鲁棒性和个性表现力也不同,因此,文中选用高斯核函数,在纯净语音下,考察不同高斯核宽度 σ 对系统性能的影响.实验数据选取与实验 1 相同,自适应压缩 Gammachirp 滤波器的个数设置为 128,取前 24 维特征参数,计算一阶动态参数,去掉第一维,将剩余的动静组合参数作为说话人的特征参数进行识别.图 5 给出了系统错误率随高斯核宽度 σ 的变化曲线。

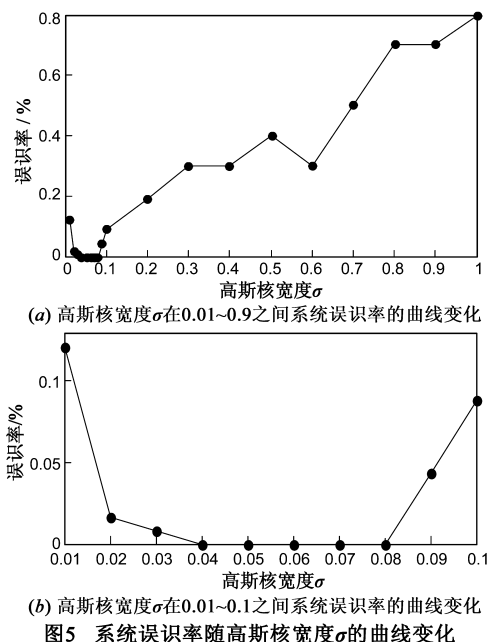


图5 系统误识率随高斯核宽度 σ 的曲线变化

在图 5 中,图 5(a)和图 5(b)分别给出了高斯核宽度 σ 在 0.01 ~ 0.9 和 0.01 ~ 0.1 之间误识率的变化曲线.从图 5(b)中可以看出,当高斯核宽度 σ 值小于 0.04 时,随着 σ 值的增加系统误识率逐渐降低,当 σ 在 0.04 ~ 0.08 之间系统误识率为 0,当 σ 大于 0.08 时,随着 σ 的增加,系统误识率逐渐增大.因此,选高斯核 $\sigma = 0.05$ 进行后续实验。

实验 3 KPCA 对识别结果的影响

为了验证 KPCA 对听觉参数变换的有效性,将图 2 中提取的 AcGCF 参数分别进行 DCT 变换和主成分分析 PCA 变换,得到的特征参数分别为 AcGCF-DCT、AcGCF-PCA,将上述的两种特征参数与本文提出的 AcGCF-KPCA 参数在纯净语音和噪声语音下进行比较.使用 66 个说话人的语音数据,选取 NOISEX-92 噪声库中的白噪声,在信噪比分别为 20dB、15dB、10dB、5dB、0dB 条件下,进行说话人识别系统性能的比较,结果如图 6 所示.

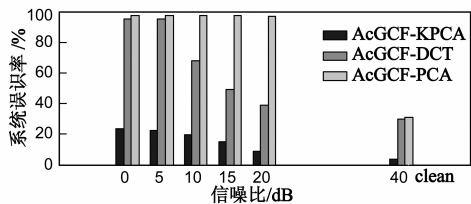


图6 三种特征变换对应的系统误识率曲线

从图 6 可以看出,无论在纯净语音还是在不同信噪比的噪声环境下,AcGCF-KPCA 和 AcGCF-DCT 特征参数性能均优于 AcGCF-PCA 参数,尤其在噪声环境下,基于 PCA 变换的特征参数 AcGCF-PCA 性能急剧下降,受噪声影响较大. AcGCF-KPCA 在不同条件下均能够得到较 AcGCF-DCT 更好的识别性能,尤其在 0dB 条件下,其识别率提高了 75.02%. 这主要是由于 KPCA 将特征参数映射到一个高维特征空间,增加了特征之间的可区分性,在噪声环境下,能够更有效地减少特征参数的噪声分量.可见,基于 KPCA 的特征参数较基于 DCT 和 PCA 的特征参数更具有噪声鲁棒性和可区分性.

实验 4 不同特征参数的系统性能比较

为了验证本文提取的特征参数的鲁棒性,将本文提取的特征参数 AcGCF-KPCA、AcGCF-DCT 与 MFCC、文献[2]提取的特征参数,在纯净语音下和噪声条件下进行比较,提取上述四种参数的动态参数,将各参数的组合参数作为说话人的语音特征参数.使用 66 个说话的语音数据,选取 NOISEX-92 噪声库中三种典型的噪声,白噪声、Babble 噪声、Pink 噪声,在信噪比分别为 20dB、15dB、10dB、5dB、0dB 条件下,进行说话人识别系统性能的比较,结果如图 7、图 8、图 9 所示.

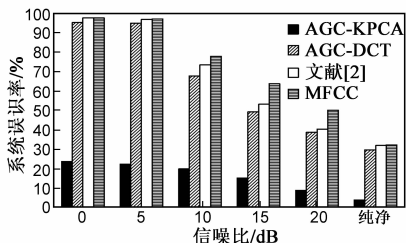


图7 纯净语音和白噪声条件下,不同特征参数的系统误识率曲线

从图 7~9 中可以看出,随着信噪比的降低,各种参数的识别性能逐渐下降.在三种噪声条件下,使用不同的信噪比,本文提出的 AcGCF-KPCA 和 AcGCF-DCT 特征参数都能得到较好的识别性能.这主要是由于采用的自适应压缩 Gammachirp 滤波器能够更好地展现听觉滤波器的非线性和动态特性,提高噪声环境下参数的鲁棒性.基于 KPCA 变换 AcGCF-KPCA 较基于 DCT 变换的 AcGCF-DCT 特征参数能够在不同信噪比条件下得到更好地识别性能.可见,KPCA 对降低特征参数的噪声分量,提高特征参数之间的可区分性具有更好地效果.

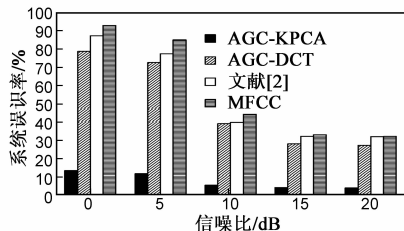


图8 Babble噪声条件下,不同特征参数的系统误识率曲线

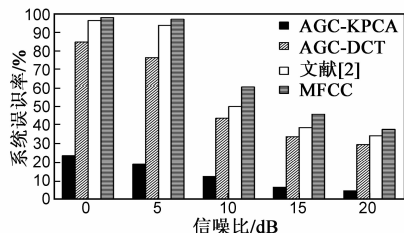


图9 Pink噪声条件下,不同特征参数的系统误识率曲线

6 结论

复杂多变的应用环境,使得说话人识别系统的性能迅速下降,鲁棒语音特征参数的提取是提高实际环境中说话人识别系统性能的关键.本文提出了一种鲁棒的听觉特征参数提取的新算法并将其应用到说话人识别系统中.运用自适应压缩 Gammachirp 滤波器组模拟人耳耳蜗的听觉特性,分别运用离散余弦变换和核主成分分析方法,对提取的特征参数进行特征变换,降低特征参数的维数,提高特征参数的噪声鲁棒性和个性表现力.实验结果表明,将提取的新听觉特征参数应用到说话人识别系统中,新特征参数在鲁棒性和识别性能上均优于梅尔倒谱系数和基于 Gammatone 的听觉特征参数.

参考文献

- [1] Yang Shao, Zhao zhang Jin, DeLiang Wang. An auditory-based feature for robust speech recognition [A]. International Conference on Acoustics, Speech, and Signal Processing [C]. Taipei, Taiwan: IEEE Signal Processing Society, 2009. 4625 - 4628.
- [2] Qi (Peter) Li, Yan Huang. Robust speaker identification using

- an auditory-based feature [A]. International Conference on Acoustics, Speech, and Signal Processing [C]. Dallas, Texas, USA: IEEE Signal Processing Society, 2010. 4514 – 4517.
- [3] 王 ■, 钱志鸿, 王雪, 程光明. 基于伽马通滤波器组的听觉特征提取算法研究[J]. 电子学报, 2010, 38(3): 525 – 528.
Wang Yue, Qian Zhi-hong, Wang Xue, Cheng Guang-ming. An auditory feature extraction algorithm based on Γ -tone filterbanks [J]. Acta Electronica Sinica, 2010, 38(3): 525 – 528. (in Chinese)
- [4] Yang Shao, Soundararajan Srinivasan, Zhaozhang Jin, Deliang Wang. A computational auditory scene analysis system for speech segregation and robust speech recognition [J]. Computer Speech and Language, 2010, 24(1): 77 – 93.
- [5] Peng Li, Yong Guan, Shijin Wang, Bo Xua, Wenju Liu. Monaural speech separation based on MAXVQ and CASA for robust speech recognition [J]. Computer Speech and Language, 2010, 24(1): 30 – 44.
- [6] T Irino, R D Patterson. A time-domain level-dependent auditory filter: The gammachirp [J]. The Journal of the Acoustical Society of America, 1997, 101(1): 412 – 419.
- [7] B C J Moore. An Introduction of the Psychology of Hearing [M]. 5th Edition. Oxford, UK: Academic, 2003.
- [8] C J Plack. The Sense of Hearing [M]. London, UK: Lawrence Erlbaum Associates, 2005.
- [9] T Irino, R D Patterson. A compressive gammachirp auditory

filter for both physiological and psychophysical data [J]. J Acoust Soc Amer, 2001, 109(5): 2008 – 2022.

- [10] 吴玺宏. 一个面向说话人识别的汉语语音数据库[OL].
<http://nlpr-web.ia.ac.cn/english/irds/chinese/sinobiometricspdf/wuxihong.pdf>, 2002.

作者简介



林琳女, 1979年7月出生于辽宁省大连市. 2007年获得吉林大学通信工程学院博士学位. 现为吉林大学通信工程学院讲师, 博士后. 主要研究方向为说话人识别, 语音信号处理, DSP技术. E-mail: lin-lin@jlu.edu.cn.



陈虹女, 1963年出生于浙江. 吉林大学通信工程学院教授, 博士生导师. 研究方向为系统仿真与先进控制.

陈建(通信作者)男, 1977年12月出生于山东单县. 吉林大学通信工程学院讲师, 博士. 研究方向: 数字信号处理, 阵列信号处理.