

文章编号 1004-924X(2013)06-1598-07

基于自适应高斯混合模型与静动态 听觉特征融合的说话人识别

吴迪^{1*}, 曹洁^{1,2}, 王进花¹

(1. 兰州理工大学 电气工程与信息工程学院, 甘肃 兰州 730050;

2. 兰州理工大学 计算机与通信学院, 甘肃 兰州 730050)

摘要:对特征参数和高斯混合模型进行改进,提出了一种特征域和模型域混合补偿的方法用于解决说话人识别特征受噪声影响较大以及高斯混合模型随训练样本长度减小而性能下降的问题。通过模拟人耳听觉,给出了基于伽马通滤波器的伽马通滤波倒谱系数;考虑其只反映了语音的静态特征,提取了能够反映语音动态特征的伽马通滑动差分倒谱系数。基于因子分析技术,利用移动因子表示高斯混合模型的自适应过程,通过训练语料较充分的说话人模型中的均值向量补偿受训练语料长度影响较大的分量的均值向量。仿真实验表明:在纯净背景下,本文方法的识别率达到了 98.46%;在不同噪声环境下,本文提出的混合补偿方法能有效提高说话人识别系统的性能。

关键词:高斯混合模型;伽马通滤波器;滑动差分倒谱;因子分析;听觉特征

中图分类号:TP391.41 **文献标识码:**A **doi:**10.3788/OPE.20132106.1598

Speaker recognition based on adapted Gaussian mixture model and static and dynamic auditory feature fusion

WU Di^{1*}, CAO Jie^{1,2}, WANG Jin-hua¹

(1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China; 2. College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China)

* Corresponding author, E-mail: wudi6152007@163.com

Abstract: By optimizing the feature vectors and Gaussian Mixture Models(GMMs), a hybrid compensation method in model and feature domains is proposed. With the method, the speaker recognition features effected by the noise and the declined performance of GMM with reducing length of the training data under different unexpected noise environments are improved. By emulating human auditory, Gammatone Filter Cepstral Coefficients(GFCC) is given out based on Gammatone Filter bank models. As the GFCC only reflects the static properties, the Gammatone Filter Shifted Delta Cepstral Coefficients(GFSDCC) is extracted based on Shifted Delta Cepstral. Then, the adaptive process for each GMM model with sufficient training data is transformed to the shift factor based on factor analysis.

收稿日期:2013-02-07;修订日期:2013-02-26.

基金项目:国家自然科学基金资助项目(No. 61263031);甘肃省自然科学基金资助项目(No. 1010RJZA046)

Furthermore, when the training data are insufficient, the coordinate of the shift factor is learned from the GMM mixtures of insensitive to the training data and then it is adapted to compensate other GMM mixtures. The experiment result shows that the recognition rate of the method proposed is 98.46%. The conclusion is that the performance of speaker recognition system is improved under several kinds of noise environments.

Key words: Gaussian mixture model; Gammatone filter; shifted delta cepstra; factor analysis; auditory feature

1 引 言

说话人识别隶属于语音识别,长期以来一直是语音处理领域的研究热点之一,国内外研究人员已在这一领域取得了相当多的研究成果^[1]。在低噪声环境下,说话人识别目前已经达到较高识别率,但在实际环境下,不同类型的噪声会在一定程度上降低说话人识别的性能,因此如何提高复杂噪声环境下说话人识别算法的性能显得至关重要。目前用于提高说话人识别算法鲁棒性的方法可分为 3 类:(1)基于特征的方法^[2],即提取对噪声敏感度较小的稳健参数,常用的方法有谱减法、倒谱均值归一化(CNN)、RAS-TA 滤波处理、特征映射等;(2)基于模型补偿的方法^[3],即对语音模型参数进行修正处理,从而补偿训练环境和实际环境之间的差异,常用的方法有说话人模型合成(SMS)、因子分析(FA)、扰动属性投影(NAP);(3)基于得分规整的方法^[4],主要是利用各种归一化方法对得分进行规整,以达到降低信道影响的目的,常用的方法有 HNorm、TNorm、ZNorm 等。

目前常用的说话人特征为时变特征性参数,如梅尔倒谱系数(MFCC)、感知线性预测系数(PLP)或韵律特征,这些参数对噪声敏感,在噪声环境下其性能会大幅度降低。虽然目前常用的基于特征的方法能够一定程度上消除噪声的影响,但是需要预知噪声形式,不能处理未知的干扰类型。因此,提取具有鲁棒稳健的语音特征在说话人识别中显得尤为重要。

高斯混合模型(GMM)是目前解决说话人识别的主流算法。实验结果表明,GMM 系统纯净的长语音数据库上具有优异的性能,但在实际应用中的性能远达不到要求。尤其是目标说话者语料长度较短,对 GMM 的识别性能有很大影响。

在很多实际应用环境,如电子监听和刑侦侦破中,目标说话者训练语料时常较短,此时系统识别性能会因为训练语料长度减小而下降,原因在由参数训练阶段模型参数训练不充分。

本文综合考虑基于特征的方法和基于模型补偿的方法,提出了一种基于自适应高斯混合模型与静动态听觉特征融合的说话人识别方法。在特征提取阶段,基于伽马通滤波器^[5],提取了基于人耳听觉特性的伽马通滤波倒谱系数(Gammatone Filter Cepstral Coefficients, GFCC),由于其只反映了语音的静态特征,没有考虑语音频谱结构在时间上的变化,故本文利用滑动差分倒谱^[6],在伽马通滤波器倒谱系数的基础上,提取了一种语音的动态特征-伽马通滤波滑动差分倒谱(Gammatone Filter Shifted Delta Cepstral Coefficients, GFSDCC)。通过使用由若干块跨多帧语音的差分倒谱构成的滑动差分倒谱特征向量,使一帧特征能够包含其前后多帧语音的声学信息。在 GMM 模型训练阶段,基于训练语音长度减少而造成的模型参数训练不充分问题,提出了一种自适应高斯混合模型。将各高斯分量均值按照顺序排列起来形成一个高维超向量,在不调整权重和方差的情况下,每个话者模型均可以用均值超向量唯一表示。基于因子分析技术^[7],将均值仿射为移动因子,利用均值超向量的移动规律对均值参数进行补偿,从而提高系统的性能。

2 基于伽马通滤波器组的静动态听觉特征提取

2.1 Gammatone 滤波器简介

Gammatone 函数最早由 Johannesma 提出,是一个标准的耳蜗听觉滤波器^[8],被用来描述由猫的听觉神经响应测量值经反向相关得到的数

据,该滤波器的时域脉冲响应为:

$$h(t) = A t^{N-1} \exp(-2\pi ERB(f_c)t) \cdot \cos(2\pi f_c t + \varphi), \quad (1)$$

其中: A 为输出增益, N 为滤波器阶数, f_c 是中心频率, φ 是相位, $ERB(f_c)$ 为等效矩形带宽(Equivalent Rectangular Bandwidth, ERB),它决定了脉冲响应的衰减速度,可定义为中心频率 f_c 的函数:

$$ERB(f_c) = 24.7 + 0.108 f_c, \quad (2)$$

本文中的 Gammatone 滤波器由 64 个滤波器组叠加成 64 通道的 Gammatone 滤波器组来实现耳蜗滤波器模型,中心频率 f_c 由式(3)给出:

$$f_c = (f_H + 228.7) \exp\left(-\frac{v_i}{9.26}\right) - 228.7, \quad (3)$$

其中: f_H 为滤波器的截止频率, v_i 是滤波器重叠因子。

2.2 伽马滤波倒谱系数提取

MFCC 是按帧计算的,其中 Mel 滤波器的作用是利用三角滤波器组对语音信号的幅度平方谱进行平滑。本文利用伽马通滤波器组代替 Mel 滤波器组,图 1 显示了伽马通滤波倒谱系数的提取流程图,首先对语音信号进行相关的预处理,接下来对语音帧进行傅里叶变换,获得语音信号频谱,通过伽马通滤波器将线性频率能量转换为伽马通频率能量,最后计算其倒谱,即先计算能量的对数,再进行离散余弦变换,得到伽马通滤波倒谱系数(Gammatone Filter Cepstral Coefficients, GFCC)。

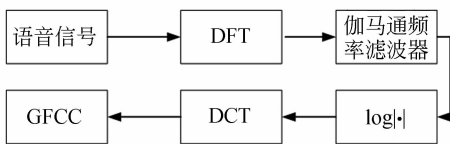


图 1 伽马通滤波倒谱系数提取过程

Fig. 1 Extraction process of Gammatone filter cepstral coefficients

2.3 伽马通滤波滑动差分倒谱系数提取

伽马通滤波滑动差分倒谱系数(GFSDCC)是在 GFCC 特征的基础上利用滑动差分倒谱提取的^[9],通常由多个连续语音帧的一阶差分谱扩展连接而成,通过串联 k 块差分倒谱在一帧内对差分倒谱进行了扩展,其中每块差分倒谱向后滑动

了 p 帧。

设第 t 帧语音的 GFCC 系数 $C(t) = [c_0(t), c_1(t), c_2(t), \dots, c_{N-1}(t)]$, $c_j(t)$ 是第 t 帧语音的第 j 个倒谱系数,每帧语音包含 N 个倒谱系数,那么第 t 帧语音的一阶差分倒谱可表示为^[10]:

$$\Delta C(t) = C(t+d) - C(t-d) = [\Delta c_0(t), \Delta c_1(t), \Delta c_2(t), \dots, \Delta c_{N-1}(t)], \quad (4)$$

同理可求:

$$\Delta C(t+p) = C(t+p+d) - C(t+p-d), \quad (5)$$

$$\Delta C[t+(k-1)p] = C[t+(k-1)p+d] - C[t+(k-1)p-d], \quad (6)$$

把 t 时刻, $t+p$ 时刻一直到 $t+(k-1)p$ 时刻的一阶差分倒谱系数连接起来,就可以求出当前帧的 GFSDCC 特征向量:

$$GFSDCC(t) = [\Delta C(t), \Delta C(t+p), \dots, \Delta C(t+(k-1)p)]. \quad (7)$$

3 自适应高斯混合模型

3.1 GMM 模型

GMM 模型是目前说话人识别中比较常用的识别模型^[11],一个 M 阶 GMM 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:

$$P(\mathbf{x}_t | \lambda) = \sum_{i=1}^M \omega_i b_i(\mathbf{x}_t), \quad (8)$$

这里 \mathbf{x}_t 代表的是说话人的特征向量, $b_i(\mathbf{x}_t)$ 是成员密度函数, ω_i 是混合权值,均值矢量为 $\boldsymbol{\mu}_i$,协方差矩阵为 $\boldsymbol{\Sigma}_i$ 的高斯函数,形式如下:

$$b_i(\mathbf{x}_t) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}_i|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_i)\right\}, \quad (9)$$

其中混合权值满足条件: $\sum_{i=1}^M \omega_i = 1$ 。完整的高斯混合模型由均值向量、协方差矩阵和混合权值参数组成,即:

$$\lambda = \{(\boldsymbol{\omega}_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), i = 1, 2, \dots, M\}. \quad (10)$$

3.2 GMM 模型中基于权重的均值分类

GMM 模型中各分量均值向量附近的特征向量在训练数据中所占有的百分比即权重 ω ,将说话人 GMM 模型各分量按照权重大小进行排序,排序在前 1/3 分量构成高权重分量集合 H ,排序在最后 1/3 分量构成低权重向量集合 L ,中间的

分量构成向量集合 M 。

GMM-UBM 模型是目前最常用的 GMM 模型,假设其 UBM 模型按照权重降序排列的均值超向量为 μ_{ubm} ,说话 s 人与 UBM 模型高斯分量排序一致的均值超向量为 μ_s ,那么两者之间的差值 $\mu_{s,u}$ 就代表了该说话人在自适应过程中其均值超向量在高维空间的移动。按照高斯分量的权重大小排列将其划分为 $\{\mu_{s,u,H}, \mu_{s,u,M}, \mu_{s,u,L}\}$,由于高权重集合 H 受训练样本长度影响较小,故 $\mu_{s,u,H}$ 信任度比较高,而集合 L 受训练样本长度影响较大,故 $\mu_{s,u,L}$ 信任度较小,有时会产生失真。

3.3 基于自适应均值的 GMM 模型

本文分析 $\mu_{s,u,H}$ 和 $\mu_{s,u,L}$ 之间的关联规律,试图通过信任度较高的来补偿失真条件下的 $\mu_{s,u,L}$,以达到说话人模型在训练样本长度较小的情况下仍能获得良好的模型的目的。

因子分析(Factor Analysis)^[12]是一种建立在均值超向量上的算法,近年来被成功应用于说话人识别,主要解决训练样本与测试样本信道不匹配问题。其优势在于与 GMM-UBM 算法和最大后验概率算法集合紧密,理论完备。基于因子分析的核心思想,本文提出了一种基于移动因子补偿的 GMM 自适应模型。

基于因子分析技术,本文将 $\mu_{s,u}$ 分解成式(11)^[13]:

$$\mu_{s,u} = \mu_s - \mu_u = \mathbf{V}\mathbf{y}(s) + \mathbf{D}\mathbf{z}(s), \quad (11)$$

则 $\mu_{s,u}$ 的第 k 维元素可表示为:

$$\mu_{s,u,k} = \sum_{i=1}^h v_{ki}\mathbf{y}(s)_i + \eta_k, \quad (12)$$

其中: η_k 为 \mathbf{D} 中的第 k 行行向量与 $\mathbf{z}(s)$ 的内积。

设特征向量的维数为 F ,GMM 模型中高斯混合数为 M ,则 $\mu_{s,u}, \mu_s, \mu_u$ 的维数为 $MF \times 1, \mathbf{V}$ 为均值超向量移动空间载荷矩阵,维数为 $MF \times h, h$ 为移动空间因子数,本文中为 100, $\mathbf{y}(s)$ 为移动因子,服从 $N(0, I)$ 分布,维数为 $h \times 1, \mathbf{D}$ 为移动残荷矩阵,代表的是 \mathbf{V} 无法描述的说话人模型均值超向量空间移动特性,维数为 $MF \times MF, \mathbf{z}(s)$ 为移动残差因子,服从均值为 0 的正态分布,维数为 $MF \times 1$ 。

从 $\mathbf{y}(s)$ 的定义和式(12)可以得知,移动因子 $\mathbf{y}(s)$ 不仅反映了 $\mu_{s,u}$ 在空间 \mathbf{V} 中整体的相关性,同时还体现出 $\mu_{s,u}$ 中的各元素通过 $\mathbf{y}(s)$ 在空间 \mathbf{V} 中建立联系。前面按照权重大小将 GMM 模型

的均值超向量划分为 $\{\mu_{s,u,H}, \mu_{s,u,M}, \mu_{s,u,L}\}$,结合式(11),可以得到下式:

$$\mu_{s,u} = \mu_s - \mu_u = \begin{Bmatrix} \mu_{s,u,H} \\ \mu_{s,u,M} \\ \mu_{s,u,L} \end{Bmatrix} = \begin{bmatrix} \mathbf{V}_H \\ \mathbf{V}_M \\ \mathbf{V}_L \end{bmatrix} \mathbf{y}(s) + \mathbf{D}\mathbf{z}(s), \quad (13)$$

式(13)将矩阵 \mathbf{V} 划分为 3 个部分: $\{\mathbf{V}_H, \mathbf{V}_M, \mathbf{V}_L\}$,由于在计算时 $\mathbf{D}\mathbf{z}(s)$ 的数值非常小,本文将其忽略不计, $\mu_{s,u,M}$ 为零向量也可忽略不计,故式(13)可以变为:

$$\mu_{s,u,H} = \mathbf{V}_H \mathbf{y}(s), \quad (14)$$

$$\mu_{s,u,L} = \mathbf{V}_L \mathbf{y}(s). \quad (15)$$

由于 $\mu_{s,u,H}$ 受训练样本长度影响较小,故可以通过 $\mu_{s,u,H}$ 和 \mathbf{V}_H 估计出移动因子 $\mathbf{y}(s)$,然后再利用 $\mathbf{y}(s)$ 和 \mathbf{V}_L 求取失真较大的 $\mu_{s,u,L}$,最终实现对集合 L 中各高斯分量均值参数的有效补偿。

4 仿真实验设置及结果分析

4.1 实验设置

本文的实验数据库来源于 CLEAR 评测,包括 4 个部分:(1). UBM 训练数据库,包含 60 段语音,每段语音长约 4 min,总计约 240 min;(2). GMM 模型参数训练数据库,同样包含 60 段语音,每段语音长约 3 min,总计约 180 min;(3). 因子分析参数训练数据库,包含 40 段语音,每段约 2 min,总计约 80 min;(4). 测试数据库,包含 300 段语音,每段约 10 s,总计约 50 min。

本文采用的特征是伽马通倒谱系数(GFCC)和伽马通滑动差分倒谱系数(GFSDCC),语音信号先进行预处理,首先使用了基于能量和过零率的方法进行静音检测,再通过因子为 0.97 的滤波器进行预加重,接着进行长度为 20 ms,帧移为 10 ms 的分帧,最后进行汉明窗加窗处理。然后提取 0~12 维 GFCC,总计为 13 维,提取 GFSDCC 特征时,参数 $N-d-P-k$ 的选取依据文献^[15]给出的最优参数组合组合,其中 N 为每帧语音中包含的倒谱系数个数, d 为计算差分倒谱的时移, p 为差分倒谱块的滑动帧数, k 为一个 SDC 特征向量中包含的差分倒谱块的个数。则 GFSDCC 总计为 39 维,在特征层融合两种特征时,总计 52 维特征向量。

4.2 不同特征组合的说话人识别结果

首先在 GMM 识别模型中,在不同语音环境下,利用 MFCC、本文提出的 GFCC 语音特征、GFSDCC 语音特征及其两者的融合特征进行说话人识别实验,实验结果如表 1 所示。

表 1 各种语音环境下采用不同声学特征的识别率

Tab. 1 Recognition rates based on different voice features under different voice environments (%)

语音环境 声学特征	SNR	MFCC	GFCC	GFSDCC	GFCC+ GFSDCC
纯净背景	>40 dB	88.72	90.02	92.38	96.54
	0 dB	34.63	34.17	42.01	49.76
	5 dB	47.77	48.17	54.19	58.72
White 噪声	10 dB	55.25	59.13	62.48	66.33
	15 dB	64.58	67.66	70.24	74.24
	20 dB	76.56	77.32	79.22	81.30
	0 dB	32.72	33.71	34.64	36.95
Babble 噪声	5 dB	47.03	48.4	50.99	56.84
	10 dB	66.13	68.04	71.00	76.62
	15 dB	75.44	77.27	78.71	83.17
	20 dB	79.79	85.64	87.24	90.09
平均值		60.78	62.68	65.73	70.05

从表 1 中的数据中可以看出,基于 GFCC 语音特征的识别率无论是在纯净背景下,还是在各种噪声环境下,识别效果都要好于基于 MFCC 语音特征的识别率,平均高出 1.9%,这是因为基于 Gammatone 滤波器组的 GFCC 语音特征比基于 Mel 滤波器组的 MFCC 特征具有更好的区分性,同时这种听觉模型更好地体现了人耳听觉系统的抗噪能力。

而基于滑动差分倒谱的 GFSDCC 特征不仅利用了 Gammatone 滤波器组的抗造能力,同时可以在一个特征向量内融入比较长的时序信息,有力地刻画了语音特征的动态信息,在纯净背景下,基于 GFSDCC 特征的识别率达到了 92.38%,比基于 GFCC 特征的识别效果高出 2.36%。在 White 噪声和 babble 噪声背景下,基于 GFSDCC 特征的识别效果也有明显改善。GFCC 和 GFSDCC 都只反映了语音特征的一个方面,本文在特征层融合二者,综合考虑语音特征的静动态特性。在纯净背景下,基于融合特征的识别率达到了 96.54%,比基于 MFCC、GFCC、GFSDCC 的识别

效果分别高出了 7.82%、6.52%和 4.16%。在不同噪声环境下,基于融合特征的识别效果也明显好于基于单特征的识别效果,在 babble 噪声环境下,当 SNR 为 20 时,识别率达到了 90.09%。

4.3 GMM 模型与自适应 GMM 模型的说话人识别结果

在纯净背景下,对 GMM 模型和本文提出的自适应 GMM 模型的识别性能进行比较,采用的识别特征为 GFCC 和 GFSDCC,实验结果如表 2 所示。

表 2 GMM 与自适应 GMM 的识别率比较

Tab. 2 Recognition rate comparison between GMM and adaptive GMM

序号	识别模型—特征向量	识别率/(%)
1	GMM-GFCC	90.02
2	GMM-GFSDCC	92.38
3	自适应 GMM-GFCC	92.69
4	自适应 GMM-GFSDCC	94.52

从表 2 可以看出,利用因子分析对 GMM 模型进行自适应均值补偿后,说话人识别系统的性能有良好提升,识别率在两种特征下分别为 92.69%和 94.52%,能够相对提高约 2%。得到的结果说明本文所提出的方法是有效的,原因在于自适应 GMM 模型对分量集合中各分量的均值参数进行了补偿,使得 GMM 模型能够更好地描述说话人的发音特性,从而使得目标话者的似然高斯分布与冒名话者的似然高斯分布之间的空间距离加大,故系统在训练语音长度较小的时候仍能取得较好的识别效果。

4.4 不同噪声环境下基于自适应高斯混合模型与静动态听觉特征融合的说话人识别

表 3 为在不同噪声背景下,利用本文提出的融合识别特征和自适应 GMM 模型进行说话人识别时的识别效果。

从表 3 的数据中可以看出,利用 GFCC 和 GFSDCC 特征组合与自适应 GMM 模型的说话人识别系统,在纯净背景下,识别率达到了 98.46%,比表 2 中利用自适应 GMM 模型单特征的识别系统的识别率分别高出了 5.77%和 3.94%,从而体现出了本文所提出方法的优越性。

表 3 不同噪声背景下本文方法的识别率比较

Tab. 3 Recognition rates based on the proposed method under different noise backgrounds

语音环境	SNR/dB	识别率/(%)
纯净背景	>40	98.46
	0	53.41
	5	62.72
	10	70.33
White 噪声	15	78.58
	20	84.93
	0	43.46
	5	66.13
Babble 噪声	10	79.46
	15	88.43
	20	94.24

在 White 噪声和 Babble 噪声背景下,在不同 SNR 情况下,本文提出的方法比表 1 中相同条件时的识别率同样有着一定幅度的提高,当 SNR 为 20 时,在 White 噪声环境下,识别率为 84.93%,而在 Babble 环境下,识别率为 94.23%,比表 1 中

参考文献:

- [1] KINNUNEN T, LI H Z. An overview of text-independent speaker recognition: from features to supervectors [J]. *Speech Communication*, 2010, 52: 12-40.
- [2] HAMID R, SEYYED A, HOSSEIN B, et al.. A new representation for speech frame recognition based on redundant wavelet filter banks [J]. *Speech Communication*, 2012, 54: 256-271.
- [3] TYLER K P, STEPHANIE N, JOHN D, et al.. Human voice recognition depends on language ability [J]. *Science*, 2011, 333: 595.
- [4] PARVIN Z, SEYYED A. Robust speech recognition by extracting invariant features [J]. *Procedia - Social and Behavioral Sciences*, 2012, 32 (3): 230-237.
- [5] SHAO Y, JIN ZH ZH, WANG D L. An auditory based feature for robust speech recognition [C]. *ICASSP*, 2009: 4625-4628.
- [6] MAK B K W, LAI T C, TSANG I W, et al.. Maximum penalized likelihood kernel regression for fast adaptation [J]. *IEEE Transactions on Audio, Speech and Language Processing*, 2009, 17 (7):

利用 GMM 模型,融合特征时的识别率分别提高了 3.63% 和 4.86%。

5 结 论

针对 MFCC 特征受噪声影响较大且只反映语音的静态特征, GMM 模型受训练样本长度减小而噪声的说话人识别系统性能下降的问题,本文提出了基于伽马通滤波器的伽马通滤波倒谱系数和伽马通滑动差分倒谱系数,使得语音特征更好地反映了人耳的听觉特性;基于因子分析,对 GMM 模型进行参数补偿,使得说话人识别系统在训练语料较小的情况下仍能取得较好的性能。

进一步的工作将针对围绕下面 2 个方面展开: (1) 如何进一步提高语音特征的抗噪能力及如何最优确定滑动差分倒谱系数提取时的参数组合, (2) 如何更好地解决说话人识别时的信道失真问题及利用因子分析进行参数补偿时, 载荷矩阵和残荷矩阵的最优估计。

1372-1381.

- [7] 翟优, 曾奕, 熊伟. 基于不变特征描述符实现星点匹配[J]. *光学精密工程*, 2012, 20(11): 2531-2539. ZHAI Y, ZENG L, XIONG W. Star matching based on invariant feature descriptor [J]. *Opt. Precision Eng.*, 2012, 20(11): 2531-2539. (in Chinese)
- [8] DU J, HUO Q. A feature compensation approach using high-order vector Taylor series approximation of an explicit distortion model for noisy speech recognition [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, 19(8): 2285-2293.
- [9] JEONG Y. Speaker adaptation based on the multi-linear decomposition of training speaker models [C]. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Dallas, USA: IEEE, 2010: 4870-4873.
- [10] HE Y J, HAN J. Gaussian specific compensation for channel distortion in speech recognition [J]. *IEEE SIGNAL PROCESSING LETTERS*, 2011, 18(10): 599-602.
- [11] OMID D, BIN M, ENG S, et al.. Discriminative feature extraction for speech recognition using continuous output codes [J]. *Pattern Recognition Letters*, 2012, 33: 1703-1709.

- [12] 史思琦,石光明,李甫. 基于轮廓特征多层描述和评价的部分遮挡目标匹配[J]. 光学 精密工程, 2012, 20(12):2804-2811.
SHI S Q, SHI G M, LI F. Partially occluded object matching via multi-level description and evaluation of contour feature [J]. *Opt. Precision Eng.*, 2012, 20(12):2804-2811. (in Chinese)
- [13] BALWANT A. SONKAMBLE, DOYE D D. A novel linear-polynomial kernel to construct support vector machines for speech recognition[J]. *Journal of Computer Science*, 2011, 7 (7): 991-996.
- [14] TOMAS P, PETER R. Real-time recognition of affective states from nonverbal features of speech and its application for public speaking skill analysis [J]. *IEEE Transactions on Affective Computing*, 2011, 2(2):66-78.
- [15] SANTHOSH K C, MOHANDAS V P. Robust features for multilingual acoustic modeling[J]. *Int J Speech Technol*, 2011, 14:147-155.

作者简介:



吴 迪(1985—),男,湖南湘潭人,博士研究生,2007年于九江学院获得学士学位,2010年于兰州理工大学获得硕士学位,主要从事多源信息融合及智能信息处理方面的研究。E-mail: wudi6152007@163.com



王进花(1978—),女,甘肃天水人,讲师,2001年于西安科技大学获得学士学位,2010年于兰州理工大学获得硕士学位,主要从事多源信息融合、视频目标跟踪等方面的研究。E-mail: 770274257@qq.com



曹 洁(1966—),女,安徽宿州人,教授,博士生导师,1987年于兰州理工大学获得学士学位,1994年于西安交通大学获得硕士学位,主要从事为智能交通、信息融合理论与应用、信息检测与估计等方面的研究。E-mail: caoj@lut.cn

(版权所有 未经许可 不得转载)