

## 基于粒子对协同优化的说话人辨认

薛丽萍<sup>1,2</sup>, 尹俊勋<sup>1</sup>, 周家锐<sup>2</sup>, 纪 震<sup>2</sup>

(1. 华南理工大学电子与信息学院, 广东广州 510640; 2. 深圳大学信息工程学院, 广东深圳 518060)

**摘 要:** 本文提出了一种新的说话人码本的优化设计方法—粒子对协同优化算法, 应用于矢量量化的说话人辨认. 此算法利用两个初始粒子对分别在每次迭代中执行粒子群优化算法的速度、位置更新和标准 LBG 算法实现并行搜索最优码本, 粒子对由两个粒子构成, 每隔一定的迭代次数通过交换粒子实现粒子对间的信息交流, 最后分别选出两个较优粒子组成精英粒子对进一步搜索. 此算法避免传统 LBG 算法陷入局部最优的缺点. 实验结果表明, 本算法始终稳定地取得显著优于 LBG、FCM、FRLVQ-FVQ、FEP 和 PSO 算法的说话人辨认性能, 较好地解决了初始码本影响优化结果的问题, 且在计算时间和收敛速度方面有优势.

**关键词:** 说话人辨认; 矢量量化; 与文本无关; 粒子群优化; 粒子对; 协同进化

**中图分类号:** TN912.3; TP18 **文献标识码:** A **文章编号:** 0372-2112 (2009) 01-0207-05

## Speaker Identification Based on Particle-Pair Cooperative Optimizer

XUE Li-ping<sup>1,2</sup>, YIN Jun-xun<sup>1</sup>, ZHOU Jia-rui<sup>2</sup>, JI Zhen<sup>2</sup>

(1. School of Electronics and Information Engineering, South China University of Technology, Guangzhou, Guangdong 510641, China;

2. Faculty of Information Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China)

**Abstract:** A novel particle-pair cooperative optimizer (PPCO) is proposed for speaker identification based on vector quantization. In this algorithm, two initial particle-pairs simultaneously explore for the best speaker codebook, and each particle-pair which consists of two particles performs basic operations of particle swarm optimization (velocity updating and position updating) and conventional LBG algorithm in sequence at each iteration. Information is exchanged when particle-pairs are reorganized periodically. And then two elitist particles selected from two initial particle-pairs respectively continue to move toward the global optimum. Experimental results have demonstrated that the performance of this new algorithm is much better than that of LBG, FCM, FRLVQ-FVQ, FEP and PSO consistently with lower speaker identification error rates, shorter computational time and higher convergence rate. The dependence of the final codebook on the selection of the initial codebook is also reduced effectively.

**Key words:** speaker identification; vector quantization; text-independent; particle swarm optimization; particle-pair; co-evolution

### 1 引言

说话人辨认系统的实用化发展, 使短语音说话人辨认成为一个关键问题. 对于较少的训练语音 (10 秒左右), 矢量量化比统计建模方法具有更多的优越性<sup>[1]</sup>. 码本优化设计是建立一个有效的说话人码本模型的关键, 直接影响着辨认性能. 文献[2]采用 LBG 算法, 易于实现, 常作为各种改进算法的基础, 但它对初始码本的依赖性大, 容易陷入局部最优值. 文献[3]利用模糊 C 均值 (fuzzy C-means, FCM) 聚类算法, 其码本质量有所提高, 但它对初值也十分敏感, 且计算模糊隶属度十分耗时. 文献[4]提出了 FRLVQ-FVQ 算法, 利用加强学习作为模糊矢量量化的预处理, 有效地改善了 FCM 对初值敏感

的缺陷. 文献[5]利用遗传算法, 文献[6]提出基于蚁群算法的 FCM, 均提高了辨认性能, 但这些算法对参数的设置敏感, 计算量大, 求优过程耗时, 收敛速度慢. 如何在提高码本性能的同时, 降低算法复杂度是值得研究的.

粒子群优化 (Particle Swarm Optimization, PSO) 算法<sup>[7]</sup> 作为一类基于群体智能的优化算法, 与其它进化算法相比, 是一种更高效的并行搜索算法, 设置参数少, 程序实现异常简洁, 近年来受到学术界广泛关注. 文献[8]在 PSO 算法的基础上, 提出了三粒子模糊 C 均值聚类算法, 有效地提高了辨认性能. 但模糊隶属度计算量大, 因此算法计算时间长. 文献[9]提出了一种新的粒子对算法, 用两个粒子构成群体规模较小的粒子对, 形成协同

收稿日期: 2007-10-24; 修回日期: 2008-09-22

基金项目: 国家自然科学基金 (No. 60572100); 深圳大学科研启动基金 (No. 200637); 深圳市科技计划 (No. 200706)

工作关系,在优化效率上优势明显.本文利用粒子对思想研究矢量量化的说话人辨认,提出了一种新的粒子对协同优化算法(Particle-Pair Cooperative Optimizer, PPCO).此算法初始群体规模为4个粒子,分为两个初始粒子对,在训练矢量空间并行搜索最优码本,粒子对在每次迭代中执行PSO和LBG混合更新操作,间隔一定的迭代次数,通过粒子的迁移实现信息交换,达到粒子对的协同进化.初始粒子对在协同进化竞争后,分别选出两个较优粒子组成精英粒子对,进一步进行搜索,并向最优靠拢.PPCO算法参数设置简单,计算时间短,寻优能力较强.经过实验仿真,PPCO在算法说话人辨认性能方面优于LBG、FCM、FRLVQ-FVQ、快速进化规划(Fast Evolutionary Programming, FEP)<sup>[10]</sup>和PSO算法,且在计算时间和收敛速度方面比FRLVQ-FVQ、FEP和PSO算法更具有优势.

## 2 矢量量化说话人辨认方法

在矢量量化说话人辨认中,将说话人看作一个信源,用一个码本表征.令说话人的 $L$ 维训练语音特征矢量集为 $X = \{x_1, x_2, \dots, x_T\}$ ,  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iL}\}$ ,  $T$ 为训练矢量的个数.对 $X$ 进行聚类,是用有限个( $M$ 个) $L$ 维码字的集合表示说话人码本模型,即 $Y = \{y_1, y_2, \dots, y_j, \dots, y_M\}$ ,  $y_j \in \mathbf{R}^L$ ,  $\forall j = 1, 2, \dots, M$ ,因此码本尺寸 $M$ 比 $T$ 小. $y_j$ 每一维的取值限定在训练矢量集对应维的最小值和最大值之间.码本设计的主要目标是寻找将 $T$ 个 $L$ 维的训练矢量分成 $M$ 个类别的最佳方案.

码本的设计质量通常是由训练矢量与对应的最近码字之间的均方误差(MSE)来表示,简称为 $\tilde{D}$ .

$$\tilde{D} = \frac{1}{T} \sum_{i=1}^T [d_{\min}(x_i)]^2 \quad (1)$$

其中 $d_{\min}(x_i) = \min_y d(x_i, y_j)$ ,  $d(x_i, y_j) = \|x_i - y_j\|$ 为欧氏距离.

在辨认阶段,用系统中建立的 $N$ 个说话人码本依次对测试语音特征矢量集进行矢量量化,计算各自的量化误差,选择平均量化误差最小的码本所对应的说话人作为辨认的结果.

系统的辨认性能与码本的设计质量密切相关.LBG和FCM本质上是一种局部的搜索技术,对初始码本敏感、容易陷入局部最优,一般得到的是一个局部最优码本.针对LBG等算法的缺点,本文利用PSO探索算法简单容易实现、优化效率高、码本设计质量好的优化设计方法.

## 3 粒子对协同优化算法

PPCO算法利用两个初始粒子对,在训练语音样本矢量空间中的并行搜索最优VQ码本.在搜索过程中是

作为两个独立的群体进行速度更新和位置更新和迭代次数为3的LBG算法,每隔一定的迭代次数,实施粒子对之间的信息交换:随机从一个粒子对中选取一个粒子与另一个粒子对的粒子交换,实现粒子对的协同进化.反复不断搜索并进化,直至满足初始粒子对的最大迭代次数,较优者将被选为精英粒子,组合成一个新的精英粒子对,继续进行搜索和进化,最优者将被选为最终解.

### 3.1 粒子结构和适应度函数

PPCO采用基于码本的优化方案,粒子结构的设计是基于码字的.每个粒子代表着一个码本.把说话人训练特征矢量集聚类成 $M$ 簇,每个码字 $y_j \in \mathbf{R}^L$  ( $j = 1, 2, \dots, M$ )代表一簇,第 $i$ 个粒子的位置, $z_i = (z_{i1}, z_{i2}, \dots, z_{iD})$ 由表1表示.粒子的维数为 $D = M \times L$ .

表1 粒子结构示意图

$y_{11}, y_{12}, \dots, y_{1L}$	...	$y_{j1}, y_{j2}, \dots, y_{jL}$	...	$y_{M1}, y_{M2}, \dots, y_{ML}$
---------------------------------	-----	---------------------------------	-----	---------------------------------

PPCO选取均方误差(MSE)作为适应度函数,利用式(1)计算.

### 3.2 粒子对更新策略

PPCO算法采用PSO和LBG混合更新策略.粒子完成PSO速度和位置更新之后,在进入下一次迭代之前,进行LBG局部优化,使粒子能够尽快移动到最优点上,加快算法的收敛速度.

#### (1) 粒子对的PSO更新

在每个粒子对中,粒子根据以下公式更新速度和位置:

$$v_{id}^{k+1} = wv_{id}^k + c_1 r_1 (p_{id} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k) \quad (2)$$

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \quad (3)$$

其中, $i = 1, 2; d = 1, 2, \dots, D; k$ 是迭代次数; $r_1$ 和 $r_2$ 为均匀分布在 $[0, 1]$ 之间的随机数; $w$ 为惯性权重; $c_1, c_2$ 为学习因子.每个粒子都有自我学习和向优秀粒子学习的能力,根据粒子 $i$ 自身迄今为止搜索到的最优位置 $p_i(p_{i1}, p_{i2}, \dots, p_{iD})$ 和粒子对迄今为止搜索到的最优位置 $p_g(p_{g1}, p_{g2}, \dots, p_{gD})$ 估计和调整每个码字的最佳移动方向,向自己的历史最优点以及粒子对历史最优点靠近,使码字趋于最优.

#### (2) 粒子对的LBG操作

粒子执行迭代次数为3的LBG算法.更新过程中粒子的码字数值超过最大值将会导致空胞腔的出现,在粒子执行完LBG操作后会处理空胞腔,用具有较大误差的训练矢量来代替具有越界问题的码字,兼顾训练矢量稀疏区域和密集区域.

### 3.3 PPCO说话人码本优化设计

PPCO训练最优码本,其算法主要步骤如下:

#### (1) 说话人初始码本的生成

随机初始化群体,从训练特征矢量集随机地选取  $M$  个矢量作为每个粒子的初始码字。

#### (2) 两个初始粒子对的更新和信息交换

按适应值更新确定  $p_i$  和  $p_g$ ;

粒子按式(2)、(3)更新速度和位置;

粒子执行 LBG 操作,并处理空码字;

判断如果达到交换粒子的间隔迭代次数,实施粒子对之间的粒子交换。

重复执行、、步,直至满足初始粒子对的最大迭代次数。

#### (3) 精英粒子对的更新

从初始粒子对中分别选出较优者组成精英粒子对。利用公式(2)、(3)对精英粒子对进行速度和位置更新以及 LBG 操作,直至达到精英粒子对的最大迭代次数。精英粒子对的最佳粒子作为训练结果,得到说话人码本模型。

### 3.4 PPCO 说话人码本优化策略分析

PPCO 采用群体规模小、多粒子对、协同进化、PSO 和 LBG 混合操作、精英粒子对等多种策略进行码本优化设计。

(1) PSO 一般群体规模较大,基于码本的优化方案,群体中粒子数量过多的情况下会难于协调多个粒子之间  $M$  个点的对应位置关系,且运算量较大,利用粒子对形成协作工作关系,不降低优化精度,计算时间短<sup>[9]</sup>,因此 PPCO 初始群体规模仅为 4 个粒子,每个子群体只使用一个粒子对。

(2) 标准 PSO 用一个粒子群在解空间中搜索最优解。由于搜索过程的集中化(1 个吸引子)导致粒子们陷入局部最优而无法摆脱。码本可以视为一个多极值问题,PPCO 首先利用 2 个粒子对在训练矢量空间中并行搜索,通过 2 个吸引子来避免早熟,提高搜索过程的多样化。

(3) 基于种间竞争机制的协同进化思想,通过初始粒子对之间粒子的迁移达到信息的交流,实现粒子对的协同进化。这样既保持了粒子对良好的聚集性能,能集中搜索训练矢量空间中的一个有希望的区域,又能完成粒子对之间远距离的信息传递,这就使得粒子对的信息能相互流动,更进一步加强两个初始粒子对的寻优能力。

(4) 精英粒子对充分利用初始粒子对搜索得到的较接近最优解的信息,进一步搜索最优码本,可以跳出初始粒子对可能陷入的局部最优点,找到更好的解,从而提高算法收敛的全局性。

(5) 粒子对采用 PSO 和 LBG 混合更新策略,利用粒子的 PSO 操作进行广度搜索,LBG 操作进行局部深度搜索,从而取得算法在精细搜索与扩展搜索空间之间

的协调。

## 4 实验结果

实验仿真平台为 Windows XP, Matlab 7.0, CPU 为 P4 3.0GHz, 内存为 1GB。采用误识率 ( $ERR$ )、 $MSE(\tilde{D})$  和每个人码本的平均计算时间 ( $T$ ) 对 PPCO 算法与 LBG、FCM、FRLVQ-FVQ、FEP 和 PSO 算法进行客观评价。

实验语音数据取自 TIMIT 语音数据库<sup>[11]</sup>, 随机抽取 100 个说话人作为辨认的集合, 30 个女性, 70 个男性。语音信号经过预加重系数为 0.95 的滤波, 采用汉明窗进行分帧和加窗, 提取 15 维 MFCC 作为语音特征矢量。每个人的训练和辨认语句分别为 5 句, 每句的有效语音时间约 2.5 秒。每次辨认实验共进行 500 次与文本无关的说话人辨认测试。

PPCO 初始群体规模为 4, 中止条件为: 初始粒子对最大迭代次数为 20 次, 交换粒子的间隔迭代次数为 5 次, 精英粒子对最大迭代次数为 15 次。

FRLVQ-FVQ 的参数设置见文献[4]。FEP 的参数设置见文献[10], FEP 群体规模为 20, 采用变异和迭代 3 次 LBG 的混合操作。PSO 群体规模为 10, 采用 PSO 更新和迭代 1 次 LBG 的混合操作。FEP 和 PSO 中止条件为最大迭代次数 35 次。

LBG、FCM 和 FRLVQ-FVQ 算法的中止条件为  $(\tilde{D}^{(k-1)} - \tilde{D}^{(k)}) / \tilde{D}^{(k)}$ 。参数设置如表 2, 其中  $m$  为模糊度。

表 2 LBG、FCM、PSO 和 PPCO 算法的参数设置

算法	$m$	$w$	$c_1$	$c_2$	
LBG	0.001	—	—	—	
FCM	0.001	1.1	—	—	
PSO	—	—	0.1	0.3	0.5
SPPO	—	—	0.1	0.3	0.5

码本尺寸为  $M = 8$ , 分别应用各算法对 100 个说话人进行 10 次训练和辨认实验, 其统计平均值如表 3 所示。PPCO 的平均辨认误识率比 LBG、FCM、FRLVQ-FVQ、FEP 和 PSO 算法分别降低了 2.78%、1.72%、2.58%、1.62% 和 0.82%。在计算时间  $T$  上分别比 FRLVQ-FVQ、FEP 和 PSO 算法节省了 33.3%、91.8% 和 60.7%, 计算时间优势明显。可见 PPCO 算法的优化码本的策略是有效的。

表 3 六种算法的说话人辨认实验结果比较

性能指标	算 法					
	LBG	FCM	FRLVQ-FVQ	FEP	PSO	PPCO
$ERR(\%)$	7.30	6.24	7.10	6.14	5.34	4.52
$T(s)$	0.6	7.0	19.8	160.9	33.6	13.2

在 10 次随机生成初始码本的实验中, LBG、FCM、FRLVQ-FVQ、FEP、PSO 和 PPCO 算法的辨认误识率的变化范围分别为 1.4 %、2.2 %、0.4 %、3.0 %、1.8 % 和 0.4 %, FRLVQ-FVQ 和 PPCO 受初始码本的影响最小, PSO 最好的实验结果比 PPCO 好, 但误识率的变化范围大, 存在早熟收敛. PPCO 表现稳定, 如图 1, 其说话人辨认性能受初始码本的影响小, 能有效地解决由于初值设置不当而导致系统性能下降的问题.

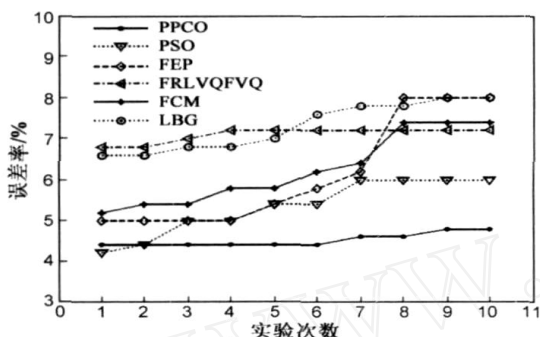


图1 六种算法的10次实验结果比较

在图 2 中, PPCO 的前 20 次迭代的失真测度 MSE 值取自于较优的初始粒子对, 之后迭代的 MSE 值取自于精英粒子对. 可见 PPCO 算法的收敛速度快, 且其 MSE 值始终稳定地保持在其它各算法之下. 其它 99 个说话人的码本训练也得到了如图 2 同样的收敛曲线.

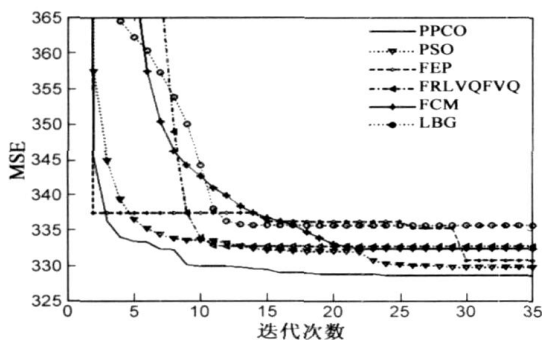


图2 六种算法的收敛特性比较

在码本尺寸分别为  $M = 8, 16, 32, 64$  时, 各算法对 100 个说话人进行 10 次训练和辨认实验, 其平均统计值如表 4 所示. 可见 PPCO 算法的码书设计质量比 LBG、FKM、FRLVQ-FVQ、FEP 和 PSO 都好, 且随着码本尺寸的增加, 计算时间优势明显. 当码本尺寸达到 32 和 64 时, PPCO 计算时间比 FCM、FRLVQ-FVQ、FEP 和 PSO 算法都短.

在不同的码本尺寸条件下, 不同算法的 100 个说话人训练和辨认实验的平均误识率如图 3 所示. PPCO 均能得到最好的辨认性能, 误识率最低. 不过随着码本尺寸的增加, 其它五种算法的辨认误识率都在降低. 在系统实现中增大码本的尺寸会增加系统的计算量, 增加系统对说话人码本模型存储器的要求. PPCO 算法的优

势在于可以用较小的码本尺寸得到较好的辨认效果.

表 4 六种算法不同码本的平均失真测度和计算时间比较

算 法	$M = 8$		$M = 16$		$M = 32$		$M = 64$	
	$\bar{D}$	$T(s)$	$\bar{D}$	$T(s)$	$\bar{D}$	$T(s)$	$\bar{D}$	$T(s)$
LBG	247.6	0.6	192.3	1	149.9	2.3	114.7	1.5
FCM	246.3	7	190.7	15.2	147.2	51.9	111.2	162.9
FRLVQ-FVQ	247.0	23.8	191.7	29.3	148.5	95.4	112.7	215.4
FEP	246.5	160.9	197.6	229.8	159.6	553.9	126.8	1272.1
PSO	243.2	33.6	188.4	49.2	145.9	79.8	109.8	149.8
PPCO	242.8	13.2	188.0	19.8	145.6	45.6	109.5	64.7

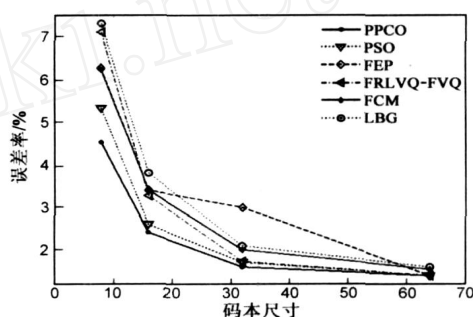


图3 不同的码本尺寸六种算法的平均辨认误识率曲线图

## 5 结论

本文提出了一种新的粒子对协同优化算法, 利用两个初始粒子对在训练矢量空间中分别通过交替执行 PSO 和 LBG 算法实现并行探索最优说话人码本, 通过粒子迁移实现粒子对协同优化, 最后选出两个精英粒子进一步搜索全局最优码本. 新算法参数设置简单, 计算时间短, 寻优能力较强. 100 个说话人辨认实验结果表明新算法在辨认性能方面显著优于 LBG、FCM、FRLVQ-FVQ、FEP 和 PSO 算法, 不依赖于说话人初始码本的选取, 且收敛速度快, 在计算时间方面优于 FRLVQ-FVQ、FEP 和 PSO 算法, 便于实际应用.

## 参考文献:

- [1] Yu K, Mason J, Oglesby J. Speaker recognition using hidden Markov models, dynamic time warping and vector quantisation [J]. IEE Proc., Vis. Image Signal Processing, 1995, 142(5): 313 - 318.
- [2] Soong F K, Rosenberg A E, Rabiner L R, et al. A vector quantization approach to speaker recognition[A]. In Proc. IEEE ICASSP '85[C]. Florida: IEEE Press, 1985. 387 - 390.
- [3] Tran D, Wagner M, Van L T. A proposed decision rule for speaker recognition based on fuzzy C-means clustering[A]. In 5th Int Conf on Spoken Language Processing[C]. Sydney: ASSTA Press, 1998. 755 - 758.

- [4] Wenhuan Xu, A K Nandi, Jihong Zhang. Novel vector quantizer design using reinforced learning as a pre-process [J]. Signal Processing, 2005, 85(7): 1315 - 1333.
- [5] 王金明, 李恩波. 基于 GA/VQ 的说话人辨认的研究与实现[J]. 解放军理工大学学报(自然科学版), 2005, 6(3): 214 - 218.  
Wang Jiming, Li Enbo. Research and implementation of speaker identification based on GA/VQ[J]. Journal of PLA University of Science and Technology(Natural Science Edition), 2005, 6(3): 214 - 218. (in Chinese)
- [6] 胡恒滔, 龙建忠. 基于蚁群算法的模糊 C-均值聚类算法在声纹辨认中的应用[J]. 四川大学学报(自然科学版), 2007, 44(3): 543 - 547.  
Hu Hengtao, Long Jianzhong. Speaker identification using fuzzy C-means clustering algorithm based on improved ant colony algorithm[J]. Journal of Sichuan University(Natural Science Edition), 2007, 44(3): 543 - 547. (in Chinese)
- [7] Kennedy J, Eberhart R. Particle swarm optimization [A]. In Proc. IEEE Int Conf on Neural Networks 1995 [C]. Piscataway: IEEE Press, 1995. 1942 - 1948.
- [8] 薛丽萍, 尹俊勋, 纪震. 基于粒子群优化-模糊聚类的说话人辨认[J]. 深圳大学学报理工版, 2008, 25(2): 178 - 183.  
Xue Liping, Yin Junxun, Ji Zhen. Speaker recognition based on particle swarm optimization and fuzzy clustering Analysis [J]. Journal of Shenzhen University Science and Engineering, 2008, 25(3): 178 - 183. (in Chinese)
- [9] 纪震, 廖惠连, 许文焕, 等. 粒子对算法在图像矢量量化中的应用[J]. 电子学报, 2007, 38(10): 1916 - 1920.  
Ji Zhen, Liao Hui-lian, Xu Wen-huan, et al. A strategy of Particle-pair for vector quantization in image coding[J]. Acta Electronics Sinica, 2007, 38(10): 1916 - 1920. (in Chinese)

- [10] Yao X, Liu Y, Lin G M. Evolutionary Programming Made Faster[J]. IEEE Transactions on Evolutionary Computation, 1999, 3(2): 82 - 102.
- [11] Garofolo J S, Lamel L F. TIMIT Acoustic-Phonetic Continuous Speech Corpus [DB/CD]. <http://www.ldc.upenn.edu/Catalog/>, 2007-12

## 作者简介:



薛丽萍 女, 1962年8月出生于陕西省铜川市, 博士生, 副教授, IEEE 会员. 主要研究方向: 语音信号处理、智能计算、DSP 及嵌入式系统. E-mail: xuelp@szu.edu.cn



尹俊勋 男, 1942年7月出生于广东省东莞市, 华南理工大学电子与信息学院教授, 博士生导师. 主要研究方向: 音频信号处理、通信新技术、DSP 及嵌入式系统. E-mail: eejyin@scut.edu.cn

周家锐 男, 1984年7月出生于广东省韶关市, 硕士生. 主要研究方向: 智能计算. E-mail: 030bug@gmail.com

纪震 男, 1973年8月出生于江苏省溧阳市, 工学博士, 教授, IEEE 会员. 主要研究方向: 智能计算、图像处理、数字水印以及数字信号处理硬件系统. E-mail: jizhen@szu.edu.cn