基于最大似然线性回归矩阵的说话人识别算法研究

钟 山[1] 何 亮[1] 邓 妍[1] 刘 加[1]

摘 要 研究了将自适应领域的最大似然线性回归 (Maximum likelihood linear regression, MLLR) 变换矩阵作为特征进行文本无关的说话人识别算法. 本文引入了基于统一背景模型的 MLLRSV-SVM 说话人识别算法, 并在此基础上进行高层音素聚类以进一步提高识别性能. 在采用多种信道补偿技术后, 在 NIST SRE 2006 年 1 训练语段-1 测试语段同信道和跨信道数据库上, 基于 MLLR 特征的系统与其他最好的系统性能接近并有很强的互补性, 经过简单线性融合可以极大提高识别性能.

关键词 说话人识别, 最大似然线性回归, 支持向量机, 信道补偿
中图分类号 TP24

Research on MLLR Based Speaker Recognition Algorithm

ZHONG Shan[1] HE Liang[1] DENG Yan[1] LIU Jia[1]

Abstract This paper uses the maximum likelihood linear regression (MLLR) as feature for text-independent speaker recognition algorithm. We introduce a universal background model (UBM) based MLLRSV-SVM algorithm first, and then extend the algorithm to multi-class for improvement. After channel compensation, in terms of the NIST 2006 SRE 1conv4w-1conv4w/mic corpus, the MLLR based system is comparable with and complementary of the state of the art systems. The performance is greatly improved by simply linear fusion.

Key words Speaker recognition, maximum likelihood linear regression (MLLR), support vector machine (SVM), channel compensation

虽然近年来说话人识别技术的发展突飞猛进, 新的识别系统不断涌现, 但是仍然没有任何一套系统能够具有压倒性优势. 在近几次由美国国家技术和标准署 (National Institute of Standards and Technology, NIST) 举办的说话人评测中, 各参评单位都不约而同采用了多系统融合的解决方案, 只有博采众长, 才能得到最佳性能. 高斯混合模型–通用背景模型 (Gaussian mixture model – universal background model, GMM-UBM)[1] 和以高斯超矢量 (Gaussian supervector, GSV) 为特征输入的支持向量机 (Support vector machine, SVM)[2] 是当前主流的文本无关说话人识别系统, 在历年评测中都表现了相当的优越性. 但是这两种基于短时倒谱特征建模的主流算法仍然存在三点不足: 1) 它忽略了同样包含特定说话人信息的长时特征和高层信息; 2) 由于不能明确分割说话人、信道以及通话的不同, 因此针对不同的测试条件, 系统的鲁棒性不够; 3) 由于目标说话人的数据不够, 一般由 UBM 映射 (通常为最大后验概率自适应 (Maximum a posterior, MAP)) 到说话人的 GMM 上, 而 SVM 系统只是把 MAP 后的 GMM 作为分类器的输入特征. MAP 过程中考虑到了与 UBM 模型的耦合性, 而这种耦合性事实上也降低了每个说话人模型的区分性.

为了弥补这两种主流算法在以上三方面的不足, 研究人员提出了很多改进方法. 比如, 引入高层音素信息的 Phontic GMM (PGMM)[3], 分别针对 GMM 和 SVM 进行信道补偿的隐藏因子分析 (Latent factor analysis, LFA)[4] 和无用分量投影 (Nuisance attribute projection, NAP)[5], 以及对 UBM 更新量建模以消除特定说话人模型间的耦合性[6] 等.

本文从另外一个角度来解决以上问题, 研究将最大似然线性回归 (Maximum likelihood linear regression, MLLR) 变换矩阵形成超矢量作为特征的说话人识别算法. 在说话人识别系统中引入 MLLR 变换矩阵作为特征, 最早是由 Stolcke 在 2005 年提出来的[7]. 因为原本用于说话人模型自适应的 MLLR 变换矩阵本身就高度提炼了特定说话人信息, 而且不存在和 UBM 的耦合问题, 因此形成的超矢量 (MLLR supervector, MLLRSV) 可以直接用来作为说话人分类的特征输入. 同时, MLLR 可以通过大词汇连续语音识别系统 (Large vocabu-

收稿日期 2008-05-23 收修改稿日期 2008-09-16
Received May 23, 2008; in revised form September 16, 2008
国家高技术研究发展计划 (863 计划) (2006AA010101, 2007AA04Z223), 国家自然科学基金委员会与微软亚洲研究院联合资助项目 (60776800) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA010101, 2007AA04Z223), National Natural Science Foundation of China and Microsoft Research Asia (60776800)
1. 清华大学电子工程系清华信息科学与技术国家实验室 (筹) 北京 100084
1. Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Electronic Engineering, Tsinghua University, Beijing 100084
DOI: 10.3724/SP.J.1004.2009.00546

lary speech recognition system, LVCSR) 进行聚类, 从而引入高层信息.

与 Stolcke 不同, Karam[8] 等在随后的研究中认为, MLLR 不再由 LVCSR 系统自适应得到, 而是直接对 UBM 进行变换. 为了降低系统复杂度, 本文同样采用这种思想, 在 GMM-UBM 平台基础上开发 MLLRSV-SVM 系统. 同时, 为了引入高层信息, 本文用 BUT (Brno University of Technology) 提供的识别引擎将数据按音素切割聚类[9], 训练多个 UBM, 分别进行 MLLR 自适应, 并在此基础上构建多类 MLLRSV-SVM (Multi-class MLLRSV-SVM, MMLLRSV-SVM) 系统. 针对系统特点, 综合运用了 LFA、NAP 以及分数域的 T-norm (Test normalization) 技术对信道进行补偿, 以进一步提高系统性能. 本文最后将 MLLRSV-SVM 系统和 GMM-UBM、GSV-SVM 进行了线性融合.

本文的安排如下: 第 1 节简单介绍 MLLR 变换矩阵和 SVM 分类器; 第 2 节详细介绍 MMLLRSV-SVM 系统; 第 3 节在 NIST 2006 年 1conv4w-1conv4w/mic 数据库上进行各种实验; 第 4 节给出分析和总结.

1　MLLR 和 SVM 简介

1.1　MLLR

MLLR 的基本原理是用仿射变换来表示说话人无关的语音模型空间与被适应人语音空间之间的变换关系, 即

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b} = W\boldsymbol{\xi} \tag{1}$$

其中, $\boldsymbol{x}$ 为自适应前的参数矢量, $\boldsymbol{y}$ 为自适应后的参数矢量, A、$\boldsymbol{b}$ 分别为根据自适应训练语音, 用最大似然准则统计出的变换参数. $\boldsymbol{\xi}$ 为拓展均值矢量 $[1, \boldsymbol{x}^{\mathrm{T}}]^{\mathrm{T}}$, W 为拓展变换矩阵 $[\boldsymbol{b} \quad A]$, 估计变换参数即对 W 进行估计. 因为 MLLR 算法的前提假设是相近的语音共享相同的变换, 因此可以根据一定的准则对语音空间进行划分, 然后对每一类空间估计其相应的变化, 并且在划分的过程中需要综合考虑划分的类数、准则和方法等因素.

1.2　SVM

SVM 作为统计学习理论中最新的部分, 目前仍处在不断发展阶段[10]. 它将解决方案建立在训练数据的子集, 即支持向量机来解决模式识别和回归问题. SVM 方法是从线性可分情况下的最优分类面 (Optimal hyperplane) 提出来的, 将输入空间的向量映射到高维 SVM 拓展空间, 然后在高维的拓展空间中采用分类方法构造最优超平面分界面, 来解决模式识别任务. 一个 SVM 可以看成是由多个内积核函数 $K(\boldsymbol{x}_1, \boldsymbol{x}_2)$ 求和构成的一个两类分类器, 对于任意的一个矢量 $\boldsymbol{x}$, 支持向量机的输出 $f(\boldsymbol{x})$ 为

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i t_i K(\boldsymbol{x}, \boldsymbol{x}_i) + d \tag{2}$$

其中, t_i 为理想的输出, $\sum_{i=1}^{N} \alpha_i t_i = 0$, $\alpha > 0$, $\boldsymbol{x}_i$ 即支持向量. 当支持向量属于第一类时, 理想输出 t_i 为 -1, 否则输出为 $+1$. 常用的核函数包括多项式核函数、径向基核函数和 Sigmoid 核函数等.

2　MMLLR-SVM 系统

2.1　MLLRSV 核函数

对 SVM 分类器选用不同的核函数会对性能造成较大影响. 在 GSV 作为输入特征的 SVM 系统中, 美国麻省理工学院林肯实验室的 Campbell 等引入了 K-L 核函数[2]

$$K_{GSV}(\boldsymbol{m}_1, \boldsymbol{m}_2) = \sum_{n=1}^{N} \gamma_n (m_{1,n})^{\mathrm{T}} \Sigma_n^{-1} m_{2,n} = \sum_{n=1}^{N} (\sqrt{\gamma_n}\Sigma_n^{-\frac{1}{2}} m_{1,n})^{\mathrm{T}} (\sqrt{\gamma_n}\Sigma_n^{-\frac{1}{2}} m_{2,n}) = d(\boldsymbol{m}_1)^{\mathrm{T}} d(\boldsymbol{m}_2) \tag{3}$$

其中, $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_2$ 为 GSV 特征, γ_n 为第 n 个高斯模型的权重, N 为混合数. $d(\cdot)$ 为对应的映射操作. 在 K-L 核基础上, 针对 MLLRSV, 由式 (1) 和 (3), 可以替换均值超矢量得到 MLLRSV 核函数如下

$$K_{MLLRSV}(W_1, W_2) = \sum_{n=1}^{N} \gamma_n (W_1\boldsymbol{\xi})^{\mathrm{T}} \Sigma_n^{-1} W_2\boldsymbol{\xi} = \sum_{n=1}^{N} (\sqrt{\gamma_n}\Sigma_n^{-\frac{1}{2}} W_1\boldsymbol{\xi})^{\mathrm{T}} (\sqrt{\gamma_n}\Sigma_n^{-\frac{1}{2}} W_2\boldsymbol{\xi}) = f(\boldsymbol{W}_1)^{\mathrm{T}} f(\boldsymbol{W}_2) \tag{4}$$

实验表明, 针对 MLLRSV 特征提出的核函数对输入特征向量空间的刻划比其他核函数更好[8], 因此本文的算法中采用 MLLRSV 核函数.

2.2　MLLR-SVM 系统构建

在 MLLRSV-SVM 系统的构建过程中, 基本步骤可以概括为以下几步:

1) 对于男女性, 分别通过足够多的数据训练两个高阶的 UBM: U_{female} 和 U_{male}, 以描述性别相关说话人无关的特征分布;

2) 对于每段训练和测试语音, 在相应性别的 UBM 上用期望最大化 (Expectation maximiza-

tion, EM) 算法计算 MLLR 变换矩阵 W, 将矩阵参数拼接起来, 形成超矢量 $\boldsymbol{W}$;

3) 将 $\boldsymbol{W}$ 进行均值方差归一化

$$\boldsymbol{W}_{\rm norm} = \frac{\boldsymbol{W} - \boldsymbol{W}_{\rm mean}}{\boldsymbol{\sigma}_W} \tag{5}$$

$\boldsymbol{W}_{\rm mean}$ 和 $\boldsymbol{\sigma}_W$ 分别为超矢量 $\boldsymbol{W}$ 的均值和标准差;

4) 将 $\boldsymbol{W}_{\rm norm}$ 特征输入 SVM 进行训练和测试.

2.3 基于音素聚类的 MMLLR-SVM 系统

语音空间可以依据一定的准则, 如欧氏距离、似然度等进行划分, 因此在本文中, 为了研究引入音素相关的高层信息, 挖掘更多说话人相关的信息, 进一步提高 MLLRSV-SVM 系统性能, 将按照音素聚类原则对语音空间进行划分. 只有可靠的音素标注才能进行正确的聚类, 因此对音素识别前端的选择非常重要. 在目前已报道的识别引擎中, BUT 提供的基于神经网络和维特比解码的识别系统识别率高, 性能最好[9]. 虽然 NIST 评测中说话人以英语为主, 但是语种识别相关研究的结论[11] 表明, 用不同语种的识别引擎进行解码, 也能提供足够的高层音素结构化信息. 因此本文选用 BUT 提供的捷克语识别引擎, 并对解码出的音素串进行聚类, 如图 1 所示.

对于所有语音, 可以按照元音和辅音分为两类. 而元音部分, 又可以分为单元音和多元音等两类. 辅音部分, 则可以分为爆破音、塞擦音、摩擦音、流音和鼻音等五类. 类数划分越多, 对语音空间的描述越具体, 但同时也会造成训练不足的问题; 类数划分越少, 高层信息利用就越少, 但同时训练出来的模型更加鲁棒. 本文的实验部分, 对划分为一类、两类以及七类都进行了相关实验, 对类数过多造成的训练数据不足, 则用回退 (Back-off) 算法进行解决.

在多类情况下, 对确定类数 M, 将训练和测试数据通过音素解码器后得到的标注依照类别聚类. 对每类训练数据, 训练 UBM, 计算 MLLRSV 并训练支持向量机. 因此对每个特定说话人, 将有 M 个实值核函数输出 K^i_{MLLRSV}, $i \in M$, 将这 M 维分数作为第二阶段特征再进行一次 SVM 分类以得到最后判决结果. 对于第二阶段的 SVM, 本文采用线性核函数.

2.4 信道补偿

要实现真实环境下的说话人识别, 就必须解决训练集和测试集不匹配的问题. 信道类型、麦克风种类、环境噪音以及通话内容等因素的不同都会造成测试条件和训练条件的失配. 在各种失配条件中, 通常情况下信道的不匹配是系统性能下降的最主要原因, 因此各种失配问题的解决方案也称之为信道补偿技术. 为了使系统更加鲁棒, 本文采用了多种信道补偿技术. 除了常规的倒谱均值减 (Cepstral mean subtraction, CMS) 和特征弯曲 (Feature warping, FW)[12], 为进一步提高系统性能, 还综合应用近些年比较流行的 LFA 和 NAP 的方法. LFA 和 NAP 的相同点都是首先将语音特征映射到高维空间, 再在高维空间估计说话人空间和信道空间的两个子空间; 二者不同点在于, LFA 通过估计测试语音的信道因子来对说话人模型进行补偿, 用于 GMM-UBM 系统; 而 NAP 算法则消除训练语音中的信道因素来突出模型中的说话人特性, 用于 SVM 系统. LFA 和 NAP 这两种算法的差异性让我们自然想到它们之间是否能互补. 因此, 为了综合这两种信道补偿算法的特点, 同时结合本文的系统, 与 LFA 在 GMM-BUM 系统中的应用一样, 在特征域采用 LFA 算法进行补偿. 而对得到的拓展变换矩阵用 NAP 算法进一步补偿, 再进行均值方差归一化送入 SVM 分类器. 最后, 对识别分数用 T-Norm 技术进行规整.

3 实验设置和结果分析

3.1 数据库

本文的实验结果建立在 2006 年 NIST 说话人评测的 1 训练语段-1 测试语段同信道以及跨信道 (1conv4w-1conv4w/mic) 的测试条件上. UBM 采用 SRE 04 数据集中挑选的男性 248 段和女性 368 段. Mixer 5 和 SRE 05 中的部分数据用于训练 LFA 和 NAP, 其中, 男性 45 人共 1 401 段, 女性 52 人共

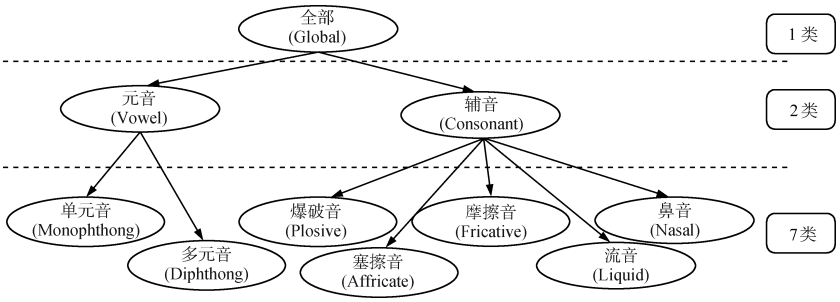


图 1 素聚类决策树结构

Fig. 1 The decision tree for phoneme clustering

1 690 段, 数据覆盖电话信道和麦克风信道. 在 SRE 02-03 以及 Mixer 1 中, 挑选男、女性各 250 人的数据作为冒充者用于 T-Norm 训练.

3.2　特征参数及系统描述

在前端, 语音信号通过预加重 (加重因子 0.97) 去除直流分量, 并经过帧宽 20 ms, 帧移 10 ms 的汉明窗. 同时采用 G.723.1 进行静音检测 (Voice activity detection, VAD), 用 12 维梅尔频率倒谱系数 (Mel frequency cepstral coefficients, MFCC) 特征加上 C0, 以及一阶、二阶、三阶差分, 形成总共 52 维特征. 之后再用线性鉴别分析 (Linear discriminant analysis, LDA) 去相关, 将特征维数从 52 维降到 39 维, 并在最后对特征进行高斯化[13]. 对于划分后的类数据, 每个 UBM 选取的高斯数为 1 024 个, MLLRSV 特征维数为 $M \times (39 \times 40) = M \times 1\,560$ 维.

3.3　实验结果

本文采用 NIST 的两个标准指标作为系统性能的评价指标: 等错误率 (Equal error rate, EER) 和最小检测代价函数 (Minimum detection cost function, MinDCF). 各实验详细设计和结果如下.

3.3.1　语音聚类结果比较

为了引入高层信息, 实验中将语音空间分别聚类为 1 类、2 类和 7 类. 为了避免类数增加造成训练数据不足, 对于缺乏训练数据的节点, 采用 Back-off 算法, 用上层节点的数据来代替. 为了单纯比较聚类性能, 除了在所有实验中都用到的倒谱均值减法和特征弯曲技术, 暂不考虑对信道进行其他特殊补偿, 结果如表 1 所示.

表 1　不同聚类的实验结果

Table 1　The results of different clusterings

EER (%) **MinDCF**	同信道男	同信道女	跨信道男	跨信道女
1 类	8.65 **0.3912**	9.33 **0.4211**	7.43 **0.3289**	9.67 **0.4338**
2 类	7.43 **0.3065**	8.72 **0.3738**	6.35 **0.3091**	9.17 **0.4179**
7 类	7.83 **0.3236**	9.08 **0.4104**	7.23 **0.3120**	4 **0.4321**

实验结果显示, 在引入高层信息之后, 识别率得到了提高. 在本文的实验中, 把语音空间聚为元音和辅音这两类能够得到最佳识别性能. 在进一步增加类数之后, 识别性能反而降低了, 这是因为类数的显著增加造成了训练不足. 尽管如此, 分 7 类的系统性能仍然优于不分类的情况. 由于分两类能得到最佳性能, 因此以下实验均建立在分两类的基础上.

3.3.2　信道补偿实验结果

在本文的系统中, 用到的信道补偿技术较多. 下面这个实验, 特意对各种补偿技术的性能进行对比, 并在最后综合运用所有补偿技术.

如表 2 所示, 每种信道补偿技术对系统性能提升都是有益的, NAP 和 LFA 这两种技术相当. 由于本文把 LFA、NAP 和 T-Norm 分别运用在特征域、模型域和分数域, 使得各种技术之间能够有效互补, 因此在综合运用各种补偿技术之后, 系统得到了最佳性能, 同时也使得系统性能更加鲁棒.

表 2　信道补偿技术实验结果

Table 2　The results of channel compensation technology

EER (%) **MinDCF**	同信道男	同信道女	跨信道男	跨信道女
基线系统	7.43 **0.3065**	8.72 **0.3738**	6.35 **0.3091**	9.17 **0.4179**
基线系统 + LFA	7.13 **0.3063**	8.35 **0.3251**	5.78 **0.2843**	8.82 **0.3979**
基线系统 + NAP	7.16 **0.3060**	8.42 **0.3282**	5.62 **0.2785**	8.75 **0.3662**
基线系统 + LFA+ NAP + T-Norm	6.83 **0.3035**	8.12 **0.3238**	4.75 **0.2591**	8.57 **0.3429**

3.3.3　系统融合性能

这部分实验将 MMLLRSV-SVM 系统和 GSV-SVM、GMM-UBM 进行比较. 这三套系统的性能以及经简单线性融合后的结果如表 3 所示, 其中两系统融合为 GSV-SVM 和 GMM-BUM 系统.

表 3　系统融合后实验结果

Table 3　The results of fusion system

EER (%) **MinDCF**	同信道男	同信道女	跨信道男	跨信道女
MMLLRSV-SVM	6.83 **0.3035**	8.12 **0.3238**	4.75 **0.2591**	8.57 **0.3429**
GSV-SVM	5.17 **0.2656**	7.01 **0.3212**	3.55 **0.1726**	7.57 **0.2916**
GMM-UBM	6.17 **0.3310**	7.69 **0.3698**	4.12 **0.2376**	5.83 **0.2701**
两系统融合	4.85 **0.2618**	6.31 **0.3176**	2.75 **0.1512**	4.94 **0.2356**
三系统融合	4.77 **0.2524**	6.14 **0.2808**	2.65 **0.1364**	4.54 **0.2143**

根据实验结果, 本文的 MMLLRSV-SVM 系统和另外两套主流系统的性能相当, 但仍有一定差距. 若将其加入融合系统中, 能进一步提高性能. 这说明 MMLLRSV-SVM 和另外两套系统确实存在互补. 男性跨信道性能最优, 远好于同信道性能, 这是因为

虽然训练和测试数据同为电话信道, 但是通话中的语种情况比较复杂. 而麦克风信道条件中的通话基本都是英语, 这说明语种不同对识别存在干扰.

4 结论

通过以上实验可知, 在划分为两类同时综合采用多种补偿技术时, MMLLRSV-SVM 系统能够得到最佳性能. MMLLRSV-SVM 系统性能和当前主流的 GSV-SVM 和 GMM-UBM 系统接近, 同时因为它采用了不同于倒谱特征的自适应变换矩阵作为系统输入, 具有很强的系统间互补性. 因此在加入融合系统后, 能够进一步提高系统性能. 在下一阶段的工作中, 我们将研究如何进一步提高 MMLLRSV-SVM 系统性能, 以达到甚至超过其他主流系统. 同时, 语种情况较复杂的同信道识别性能不如纯英语的跨信道性能, 这说明本文系统对多语种情况不够鲁棒, 严重偏向英语. 因此, 如何利用语种等边信息, 更好地进行融合和校准, 也需要进一步的研究.

References

1 Reynolds D A, Quatieri T F, Dunn R B. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 2000, **10**(1–3): 19−41

2 Campbell W M, Sturim D E, Reynolds D A. Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters*, 2006, **13**(5): 308−311

3 Castaldo F, Colibro D, Dalmasso E, Laface P, Vair C. Compensation of nuisance factors for speaker and language recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007, **15**(7): 1969−1978

4 Vair C, Colibro D, Castaldo F, Daimasso F, Laface P. Channel factors compensation in model and feature domain for speaker recognition. In: Proceedings of IEEE Odyssey: The Speaker and Language Recognition Workshop. San Juan, Puerto Rico: IEEE, 2006. 1−6

5 Campbell W M, Sturim D E, Reynolds D A, Solomonoff A. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Toulouse, France: IEEE, 2006. 97−100

6 Guo Wu, Dai Li-Rong, Wang Ren-Hua. Speaker verification based on improved updates to the SVM. *Journal of Tsinghua University (Science and Technology)*, 2008, **48**(z1): 704−707
(郭武, 戴礼荣, 王仁华. 采用 UBM 更新量作为支持向量机特征的说话人确认. 清华大学学报 (自然科学版), 2008, **48**(z1): 704−707)

7 Stolcke A, Ferrer L, Kajarekar S, Shriberg E, Venkataraman A. MLLR transforms as features in speaker recognition. In: Proceedings of the 9th European Conference on Speech Communication and Technology. Lisbon, Portugal: International Speech and Communication Association, 2005. 2425−2428

8 Karam Z N, Campbell W M. A new kernel for SVM MLLR based speaker recognition. In: Proceedings of the 8th Conference in the Annual Series of Interspeech Events and the 10th Biennial Eurospeech Conference. Antwerp, Belgium: International Speech and Communication Association, 2007. 290−293

9 Pavel M, Petr S, Jan C, Pavel C. Phonotactic language identification using high quality phoneme recognition. In: Proceedings of the 9th European Conference on Speech Communication and Technology. Lisbon, Portugal: International Speech and Communication Association, 2005. 2237−2240

10 Bian Zhao-Qi, Zhang Xue-Gong. *Pattern Recognition*. Beijing: Tsinghua University Press, 1999
(边肇祺, 张学工. 模式识别, 北京: 清华大学出版社, 1999)

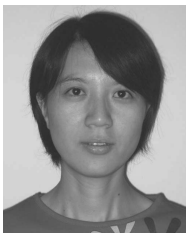
11 Muthusamy Y K, Barnard E, Cole R A. Reviewing automatic language identification. *IEEE Signal Processing Magazine*, 1994, **11**(4): 33−41

12 Pelecanos J, Sridharan S. Feature warping for robust speaker verification. In: Proceedings of the International Conference on a Speaker Odyssey: The Speaker Recognition Workshop. Crete, Greece: ISCA, 2001. 213−218

13 Xiang B, Chaudhari U V, Navratil J, Ramaswamy G N, Gopinath R A. Short-time Gaussianization for robust speaker verification. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, USA: IEEE, 2002. 681−684

钟 山 清华大学电子工程系博士研究生. 主要研究方向为说话人识别和语种识别. 本文通信作者.
E-mail: zhongshan00@mails.thu.edu.cn
(**ZHONG Shan** Ph. D. candidate in the Department of Electronic Engineering, Tsinghua University. His research interest covers speaker recognition and language recognition. Corresponding author of this paper.)

何 亮 清华大学电子工程系博士研究生. 主要研究方向为说话人识别和语种识别.
E-mail: heliang06@mails.thu.edu.cn
(**HE Liang** Ph. D. candidate in the Department of Electronic Engineering, Tsinghua University. His research interest covers speaker recognition and language recognition.)



邓 妍 清华大学电子工程系博士研究生. 主要研究方向为基于音位特征的语种识别.
E-mail: y-deng05@mails.thu.edu.cn
(**DENG Yan** Ph. D. candidate in the Department of Electronic Engineering, Tsinghua University. Her main research interest is language recognition based on phonology.)

刘 加 清华大学电子工程系教授. 主要研究方向为语音识别和信号处理.
E-mail: liuj@tsinghua.edu.cn
(**LIU Jia** Professor in the Department of Electronic Engineering, Tsinghua University. His research interest covers speech recognition and signal processing.)