

## 采用模型和得分非监督自适应的说话人识别

王尔玉<sup>1</sup> 郭武<sup>1</sup> 李轶杰<sup>1</sup> 戴礼荣<sup>1</sup> 王仁华<sup>1</sup>

**摘要** 在说话人识别的研究中,使用以前的测试语句信息对模型参数或者测试得分进行动态更新,使模型可以更精确地反映测试语句和说话人模型之间的关系,这种更新策略称为非监督模式,这方面的研究对实际的说话人识别系统具有非常重要的意义.本文除了采用非监督的说话人模型自适应更新方法之外,还提出了非监督的得分域自适应算法:首先采用双高斯函数对得分建立一个先验的得分模型,利用最大后验概率准则对得分规整的模型进行调整.在测试过程中,采用得分域和模型域的非监督算法可以互相补充,提高识别率,在 NIST SRE 2006 年 1 训练语段 -1 测试语段数据库上,使用模型域和得分域非监督自适应的系统能够取得等错误率 4.3% 和检测代价函数 0.021 的结果.

**关键词** 说话人确认,混合高斯模型,非监督模式,得分规整  
**中图分类号** TP391

### Speaker Verification with Model-based and Score-based Unsupervised Adaptation Method

WANG Er-Yu<sup>1</sup> GUO Wu<sup>1</sup> LI Yi-Jie<sup>1</sup> DAI Li-Rong<sup>1</sup> WANG Ren-Hua<sup>1</sup>

**Abstract** In the text-independent speaker verification research, the information of previous trials can be adopted to update the speaker models or the test scores dynamically. This process is defined as the unsupervised mode, which can make a coupling between the trials and the speaker models. The unsupervised mode is very useful for real speaker recognition application. In this paper, a score-based unsupervised adaptation is proposed as well as model-based unsupervised adaptation. In the score-based unsupervised adaptation mode, a bi-Gaussian model is introduced as a prior score distribution. Then the MAP (maximum a posteriori) method is adopted to adjust the parameters of the score normalization. In the test process, the unsupervised score adaptation and unsupervised model adaptation can both improve the performance. In the case of NIST SRE 2006 1conv4w-1conv4w corpus, the equal error rate (EER) of the proposed system is 4.3% and the minimum detection cost function (minDCF) is 0.021.

**Key words** Speaker verification, Gaussian mixture model (GMM), unsupervised mode, score normalization

在真实的说话人识别研究中,训练集和测试集一般不会取得一致.这种不一致来源于多个方面,比如环境噪声、信道类型以及不同的麦克风种类等因素.这些因素之中,训练和测试信道的不匹配是系统性能衰减的最大原因.解决这种不匹配问题常用的方法是模型域的因子分析 (Factor analysis)<sup>[1]</sup> 和扰动属性投影 (Nuisance attribute projection, NAP)<sup>[2]</sup>,这两种算法都可以消除信道的影响,使得说话人模型和测试语句在信道层面上更加匹配.另一方面,如果采用适当的策略来把测试语句的信息应用到说话人模型处理上,即在测试的过程中使用测试语句来调整说话人的模型,使说话人模型可以更好地反映测试环境下说话人的特征,那么不仅可

以减少测试和训练中不匹配的问题,而且可以增加训练的数据,从而使模型更加稳定,这种方式称为非监督模式 (Unsupervised mode).在由美国国家技术和标准署 (National Institute of Standards and Technology, NIST) 举办的说话人测试大赛上,从 2006 年开始,就允许采用非监督模式<sup>[3]</sup>.

几乎所有的研究机构对于非监督的模式都是采用模型参数更新的方法,即根据测试中的得分范围确定一个门限<sup>[4-5]</sup>,如果一个测试语句的得分大于门限,那么就用这个语句的数据更新说话人的模型.也有研究者采用最大后验概率 (Maximum a posteriori, MAP)<sup>[6]</sup> 准则的方法来更新模型.这些方法称为模型域的非监督自适应.

本文将非监督的思想引入到得分域的得分规整之中,提出了得分域的非监督自适应策略.首先建立目标说话人和冒认者得分分布的概率分布模型,在得到了每一次测试得分的情况下,检测这次得分对于两个先验得分模型的占有率.然后采用最大后验概率准则的方式,利用得到的占有率信息对先验的得分模型进行更新,将更新后的先验得分模型用于下一次得分的规整.这种方法改善了得分规整的性

收稿日期 2007-12-03 收修改稿日期 2008-09-06  
Received December 3, 2007; in revised form September 6, 2008  
国家高技术研究发展计划 (863 计划) (2006AA010104) 资助  
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA010104)  
1. 中国科学技术大学电子工程与信息科学系科大讯飞语音实验室 合肥 230027  
1. iFly Speech Laboratory, Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027  
DOI: 10.3724/SP.J.1004.2009.00267

能,同时具有和模型域非监督自适应方法相似的性能.通过实验可知,非监督模式的这两种方法具有一定的互补性,融合起来可得到较好的性能.

另一方面,混合高斯模型和通用背景模型(Gaussian mixture model-universal background model, GMM-UBM)<sup>[7]</sup>仍是目前最广泛的说话人识别算法,目前性能较好的主流单系统仍然是 GMM-UBM 模型结合因子分析技术的改进系统<sup>[5]</sup>,本文的所有实验都是基于这个框架.

本文结构安排如下:第 1 节介绍 GMM-UBM 系统和因子分析;第 2 节详细介绍得分域的非监督自适应算法;第 3 节介绍得分域自适应算法在 NIST 2006 上的相关实验;第 4 节是对得分域自适应算法性能的分析 and 结论.

## 1 混合高斯模型-通用背景模型和联合因子分析

按照混合高斯模型的框架,一个说话人模型  $\lambda$  可以用  $N$  个高斯函数  $p_i(\mathbf{x}; \mathbf{m}_i, \Sigma_i)$  的加权和表示如下

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^N \omega_i p_i(\mathbf{x}; \mathbf{m}_i, \Sigma_i) \quad (1)$$

其中,  $\mathbf{m}_i, \Sigma_i$  分别代表均值和方差参数,  $\omega_i$  代表权重.

由于单独采用每个人的数据无法训练一个稳定的高斯模型, Reynolds<sup>[7]</sup> 提出先用大量数据训练一个 UBM 模型,然后通过最大后验概率准则的方法自适应到每个说话人的 GMM 上去.

如果把说话人的所有 GMM 的均值排列起来<sup>[8]</sup>,那么就构成一个均值超矢量,联合因子分析就是在这个均值超矢量空间进行的一种处理.按照联合因子分析的思路,可以认为每句话形成的超矢量中,不仅包含说话人的信息,也包含信道的信息,因此这个超矢量可以写成如下形式

$$\mathbf{m}_h(s) = \mathbf{m}_{ubm} + \Delta \mathbf{y}(s) + U \mathbf{x}_h(s) \quad (2)$$

这里认为  $\Delta \mathbf{y}(s)$  是一个与信道无关的特定说话人在 UBM 模型均值超矢量  $\mathbf{m}_{ubm}$  上的偏移,  $U$  代表信道空间,  $\mathbf{x}_h(s)$  代表具体某句话中的信道因子.因此式(2)的意义很明确,即对每句话,希望尽量把  $U \mathbf{x}_h(s)$  信道信息去除,只保留与特定说话人相关的信息.在估计出信道空间  $U$  和每句话的信道因子  $\mathbf{x}_h(s)$  之后,采用特征映射的方法在前端声学参数上滤除信道部分的信息<sup>[9]</sup>.本文后面的实验均采用了这种算法处理声学特征.

## 2 得分域的非监督自适应算法

得分规整的公式为

$$\hat{S} = \frac{S - \mu}{\sigma} \quad (3)$$

其中,  $\mu$  和  $\sigma$  为大量冒认数据得分的均值和方差,  $S$  和  $\hat{S}$  分别为规整前后的得分.

最典型的得分规整有两种算法, Znorm<sup>[10]</sup> 和 Tnorm<sup>[10]</sup>. Znorm 是通过大量的冒认语音来对特定的说话人模型进行测试,从而估计出对于这个特定说话人模型得分的均值和方差; Tnorm 是通过测试语音对大量冒认说话人进行测试,从而估计出这个特定的测试语音对于不同模型得分的均值和方差.

Znorm 和 Tnorm 的参数估计中都需要使用大量的冒认数据(冒认语句或冒认说话人模型).得分规整带来的性能提高与这些数据的挑选有很大的关系,即需要选择合适的开发集.由于测试集在测试之前是未知的,因此选择的开发集几乎不可能与测试集完全匹配,这就影响了得分规整的效果.而得分域的非监督自适应算法,根据每次测试的得分来动态地调整由开发集得到的规整参数,使得调整之后的规整参数能够更好地反映出当前测试集的得分分布情况.这时再使用这些规整参数来进行得分域的规整,就可在一定程度上消除开发集与测试集的不匹配现象.下面详细介绍这种方法的流程.

在测试过程中,根据已经得出的评分来修改原来的规整参数.但是,这里有一个很关键的问题,即在式(3)中规整参数是冒认语音的得分分布,而在实际的测试中无法确定当前的测试得分是否为冒认语音数据,这就需要确定当前得分是否属于冒认得分,所以需要衡量每一个得分对于改变冒认者得分分布所能做出的贡献.这里,使用置信度这个概念来衡量这个贡献.

为了确定这个置信度,首先需要建立体现目标语音和冒认语音两部分得分分布关系的双高斯先验分布.对具体的每个目标说话人而言,由于目标测试语音不可能足够多来得到目标说话人的得分的分布,因此采用一个替代的策略,即用一个开发集估计出目标语音和冒认语音总体上的得分分布.这个分布只是大致反映目标语音和冒认语音在分布上的关系,而对于每个具体说话人来说,正确与冒认得分分布需要做一定的修正.

比如要做 NIST 2006 年的测试,那么 NIST 2005 年测试集的数据就可以用来估计出目标说话人和冒认者得分构成的双高斯分布,两部分表示为  $\Gamma_{imp.all}$  和  $\Gamma_{tar.all}$ , 其对应的参数均值和标准差分别为  $\{\mu_{imp.all}, \sigma_{imp.all}\}$  和  $\{\mu_{tar.all}, \sigma_{tar.all}\}$ . 得到的这两个分布大致上反映了目标语音和冒认语音的分

布关系.

在具体每个人的测试过程中, 首先可以采用 Znorm 的方法估计出该说话人的冒认语句的分布  $\lambda_{imp}$ , 其所对应的均值和标准差为  $\{\mu_{imp}, \sigma_{imp}\}$ . 需要估计其所对应的目标说话人的得分分布  $\lambda_{tar}$  以及其参数  $\{\mu_{tar}, \sigma_{tar}\}$ , 采用式 (4) 和 (5) 来估计, 即

$$\mu_{tar} = \mu_{imp} + \frac{\sigma_{imp}}{\sigma_{imp-all}}(\mu_{tar-all} - \mu_{imp-all}) \quad (4)$$

$$\sigma_{tar} = \frac{\sigma_{tar-all}}{\sigma_{imp-all}}\sigma_{imp} \quad (5)$$

式 (4) 和 (5) 估计出来的目标说话人的分布是很不准确的, 但是这个分布仅仅是作为一种先验分布, 粗略地反映了对于某个具体说话者模型的正确和错误的得分分布关系, 之后还需要用非监督的方法来做修正, 所以这里允许稍微有些偏差. 在这些参数的初始估计中, 由于没有任何先验性质的知识, 因此引入了一个假设, 即所有目标说话人的得分分布与这个说话人冒认语句的得分分布之间的相对位置关系, 和开发集中目标说话人得分与冒认者得分分布之间的相对位置关系 (即前文中提到的两个双高斯分布所反映的目标说话人和冒认者得分分布的关系) 是一致的.

对每一次得分  $S$ , 可以得到相对  $\lambda_{imp}$  和  $\lambda_{tar}$  的条件概率  $p(S|\lambda_{tar})$  和  $p(S|\lambda_{imp})$ , 因此当前的测试得分是冒认者得分的置信度  $C$  可以按式 (6) 计算. 其中冒认得分和目标得分的点分布概率  $p(\lambda_{tar})$ ,  $p(\lambda_{imp})$  的取值对计算影响不大, 在实验中取 0.5.

$$C = \frac{p(\lambda_{imp}) \times p(S|\lambda_{imp})}{p(\lambda_{tar}) \times p(S|\lambda_{tar}) + p(\lambda_{imp}) \times p(S|\lambda_{imp})} \quad (6)$$

根据经典的 MAP 算法, 使用实时得到的测试得分, 结合式 (6) 得到的置信度来更新规整参数. 对于得分规整参数, 我们只更新均值, 不更新方差. 设更新之后的均值为  $\hat{\mu}_{imp}$ , 原来的均值为  $\mu_{imp}$ , 更新的自适应因子为  $r$ , 实时测试得分为  $S$ , 更新公式如式 (7) 所示

$$\hat{\mu}_{imp} = \frac{C}{C+r} \times S + \left(1 - \frac{C}{C+r}\right) \times \mu_{imp} \quad (7)$$

目标说话人的得分分布  $\lambda_{tar}$  以及其参数  $\{\mu_{tar}, \sigma_{tar}\}$  同样使用式 (7) 予以更新. 在得到了更新之后的冒认语音分布的参数后, 对于下一个测试得分  $S$ , 就可以采用式 (3) 进行规整了.

式 (7) 中的自适应因子  $r$  表示原来的得分模型参数对于新模型的贡献大小.  $r$  越大, 原来得分模型参数对于更新贡献越大. 在 SRE 06 的实验中,  $r$  的取值在 16~120 之间性能都有一定的提高, 在 32~

64 之间时得到的性能提升更大一些; 在 SRE 05 的实验中,  $r$  的取值在 48~80 左右时性能提升较大. 可见系统性能对于  $r$  的取值并不是十分敏感, 在一个很大的取值范围之内系统都可以获得性能的提升. 可以根据经验, 选取  $r = 48$  作为实验中的参数. 当然在新的任务中,  $r$  的取值也可以由与任务匹配的开发集来确定.

### 3 实验设置及结果

#### 3.1 数据库

本文采用 SRE 06 测试任务中的核心测试: 1 训练语段 -1 测试语段 (1conv4w-1conv4w) 的数据集进行实验, 一共有 810 个目标说话人, 总计 51448 次测试. 测试语音和训练语音均为从 5 分钟左右的双声道对话语音数据中抽出的一个声道的声音, 实际其中有语音部分大约为 3 分钟左右.

NIST SRE 04 和 SRE 05 的训练集和测试集中有多段同一说话人在不同信道下采集的语音, 我们选择这些语句作为信道子空间估计的训练数据. 经过筛选后, 总共有 724 人及 9897 段语音, 平均每人 13 段.

选择 SRE 04 的 1conv4w-1conv4w 训练集来作为 Tnorm 参数估计的冒认数据集; 选择 SRE 05 的 1conv4w-1conv4w 训练集来作为 Znorm 参数估计的冒认数据集. 实验中使用性别相关的得分规整策略.

#### 3.2 特征参数提取

本文采用的是梅尔刻度式倒谱 (Mel frequency cepstral coefficients, MFCC) 参数, 对于 MFCC 参数提取, 语音信号先去直流, 预加重 (因子为 0.97), 经过帧宽 25 ms, 帧移 10 ms 的汉明窗. 在抽取 MFCC 特征参数的同时, 采用基于能量的寂静帧检测算法去除寂静音. 抽取 0~12 维 MFCCs, 总计为 13 维, 特征参数通过 CMN (Cepstral mean normalization)<sup>[7]</sup> 和 RASTA (Relative spectral)<sup>[7]</sup> 进行倒谱域滤波去除信道卷积噪声, 通过一阶差分、二阶差分总计构成 39 维, 特征再通过高斯化模块以提高识别率.

#### 3.3 系统描述

为了比较得分域的非监督自适应算法与经典的得分规整和模型域的非监督自适应算法的性能, 这里选择采用信道补偿和特征映射的 512 维的 GMM-UBM 系统做对比系统, 即表 1 (见下页) 中的系统 1. 系统 2 和系统 3 为采用经典的得分规整的系统, 系统 4 为采用模型域的非监督自适应算法的系统.

表 1 NIST SRE 2006 1conv4w-1conv4w 测试任务上不同系统的实验结果

Table 1 Performance of different systems on NIST SRE 2006 1conv4w-1conv4w task

序号	系统描述	minDCF/EER
1	因子分析之后 GMM-UBM, 512 GMM	0.031/5.8 %
2	因子分析之后 GMM-UBM, 512 GMM, Tnorm	0.027/5.6 %
3	因子分析之后 GMM-UBM, 512 GMM, Tnorm, Znorm	0.025/5.1 %
4	因子分析之后 GMM-UBM, 512 GMM, Tnorm, 非监督的模型训练	0.022/4.5 %
5	因子分析之后 GMM-UBM, 512 GMM, Tnorm, 非监督的得分规整	0.023/4.7 %
6	因子分析之后 GMM-UBM, 512 GMM, Tnorm, 非监督的模型训练, 非监督的得分规整	0.021/4.3 %

对于使用因子分析技术的系统均分为两部分. 首先是前端因子分析, 采用特征映射的方式去除信道影响<sup>[8]</sup>, 这部分的高斯数与训练数据的大小关系紧密, 实验中选取 512 高斯, 信道因子的个数取为 30, 男声、女声分别单独处理.

对于后端的真正的说话人识别系统, 采用性别相关的 GMM-UBM 系统, 其模型高斯数均为 512 个. 采用经典的相关 MAP 自适应方式由 UBM 模型得到具体的每个人的 GMM 模型, 相关因子为 16.

### 3.4 实验结果

本文中衡量系统性能的指标采用 NIST 标准中的两个评测指标: 等错误率 (Equal error rate, EER) 和最小检测代价函数 (Minimum detection cost function, minDCF)<sup>[8]</sup>, 结果见表 1.

对比采用因子分析技术的 GMM-UBM 基线系统, 使用了非监督算法的系统性能都有比较大的提升, 其中采用模型域和得分域融合的非监督算法的系统 6 性能最佳, 优于文献 [4-5] 中描述的性能. 实验中采用的所有系统的检测错误代价 (Detection error trade-off, DET) 性能曲线如图 1 所示.

## 4 分析及相关结论

得分域的非监督自适应方法是将非监督模式与得分规整结合到一起的一种新的方法. 这个新方法的性能好于经典的得分规整, 与传统的模型域自适应算法性能相当.

由表 1 及图 1 可见, 系统 2 和系统 3 采用的是经典的得分规整方法, 使得系统性能较基线系统有了明显提升. 系统 5 采用了得分域非监督自适应算法, 较系统 3 仍有 7.84% 的系统提升. 由此可知这

种方法可以很大程度地改进得分规整的效果. 即使对于进行过较好的规整的系统, 仍然可以在此基础上继续提升系统的性能. 这就说明, 非监督的得分规整有效地利用了观测到的数据分布情况, 实时地调整了规整的参数, 使其与测试集合分布更加匹配. 这种方法的好处在于, 得分规整参数的设定不再仅仅依靠经验性的初始设定, 还可以根据实时信息灵活有效地调整, 最终提高系统的性能.

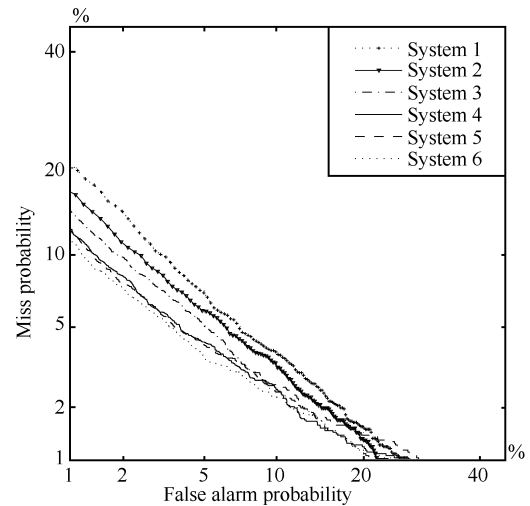


图 1 NIST SRE 2006 1conv4w-1conv4w 测试任务上各系统的 DET 性能曲线图

Fig. 1 The DET of different systems on NIST SRE 2006 1conv4w-1conv4w task

对于采用模型域的非监督自适应方法的系统 4, 获得了比得分规整更好的性能. 对于模型域自适应方法, 只要合理设置更新规则, 等同于增加系统的训练集数据. 同时在测试过程中获得之前未出现过的模型信息, 并且可以以此来完善说话人的模型, 显著地提高系统性能. 而系统 5, 应用得分域的非监督算法, 得到的 DET 曲线和系统 4 几乎重合在一起, 说明它与模型域的方法性能相似. 但是这种方法的数据计算量小, 训练复杂度低. 在实时处理的需求中, 得分域的方法可以代替模型域的方法, 并且可以达到与模型域处理相近的结果.

非监督模式的得分域算法和模型域算法性能接近, 且各具优点. 将得分域和模型域的非监督方法加以融合, 就可以充分利用两种方法的优点, 在一定程度上解决了训练数据有限的问题以及测试与训练的数据不匹配问题. 应用在相关的系统之中, 非监督自适应的方法可以获得很大性能提升, 系统 6 是实验中得到的最佳系统.

## References

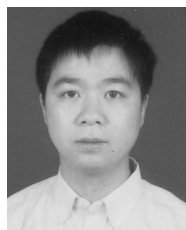
- 1 Vogt R, Sridharan S. Experiments in session variability modeling for speaker verification. In: Proceedings of IEEE Inter-

national Conference on Acoustics, Speech, and Signal Processing, Washington D. C., USA: IEEE, 2006. 897–900

- 2 Campbell W M, Sturim D E, Reynolds D 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. Washington D. C., USA: IEEE, 2006. 97–100
- 3 The NIST. Speaker Recognition Evaluation Plan [Online], available: <http://www.nist.gov/speech/tests/sre>, September 1, 2008
- 4 Vair C, Colibro D, Castaldo F, Dalmaso E, Laface P. Loquendo-politecnico di Torino's 2006 NIST speaker recognition evaluation system. In: Proceedings of the 8th Conference in the Annual Series of INTERSPEECH Events and the 10th Biennial EUROSPEECH Conference. Antwerp, Belgium: ISCA, 2007. 1238–1241
- 5 Matejka P, Burget L, Schwarz P, Glembek O, Karafiat M, Grezl F. STBU system for the NIST 2006 speaker recognition evaluation. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington D. C., USA: IEEE, 2007. 221–224
- 6 Pretil A, Bonastrel J F, Matrouf D. Confidence measure based unsupervised target model adaptation for speaker verification. In: Proceedings of the 8th Conference in the Annual Series of INTERSPEECH Events and the 10th Biennial EUROSPEECH Conference. Antwerp, Belgium: ISCA, 2007. 754–757
- 7 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
- 8 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
- 9 Castaldo F, Colibro D, Dalmaso 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
- 10 Bimbot F, Bonastre J F, Fredouille C, Gravier G, Magrin-Chagnolleau I, Meignier S. A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing*, 2004, **2004**(4): 430–451



**王尔玉** 中国科学技术大学硕士研究生. 主要研究方向为说话人识别和语种识别. E-mail: [eryuwang@mail.ustc.edu.cn](mailto:eryuwang@mail.ustc.edu.cn)  
(**WANG Er-Yu** Master student at University of Science and Technology of China. His research interest covers speaker recognition and language identification.)



**郭武** 中国科学技术大学电子工程与信息科学系讲师. 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: [guowu@ustc.edu.cn](mailto:guowu@ustc.edu.cn)

(**GUO Wu** Lecturer in the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His research

interest covers speaker recognition and language identification. Corresponding author of this paper.)



**李轶杰** 中国科学技术大学硕士研究生. 主要研究方向为说话人识别和语种识别.

E-mail: [andylyj@mail.ustc.edu.cn](mailto:andylyj@mail.ustc.edu.cn)

(**LI Yi-Jie** Master student at University of Science and Technology of China. His research interest covers speaker recognition and language identification.)



**戴礼荣** 中国科学技术大学电子工程与信息科学系教授. 主要研究方向为语音识别和信号处理.

E-mail: [lrdai@ustc.edu.cn](mailto:lrdai@ustc.edu.cn)

(**DAI Li-Rong** Professor in the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His re-

search interest covers speech recognition and signal processing.)



**王仁华** 中国科学技术大学电子工程与信息科学系教授. 主要研究方向为文语合成和语音识别.

E-mail: [rhw@ustc.edu.cn](mailto:rhw@ustc.edu.cn)

(**WANG Ren-Hua** Professor in the Department of Electronic Engineering and Information Science, University of Science and Technology of China. His

research interest covers text to speech and speech recognition.)