基于“码袋”算法的话者自动辨认系统

陈 哲[2]，顾明亮[1,2]，王劲松[1]，冯静兰[2]，杨亦鸣[2]
(1. 徐州师范大学物理与电子工程学院，江苏徐州 221116；2. 江苏省语言科学与神经认知工程重点实验室，江苏徐州 221116)

摘要：在话者自动辨认系统中，话者数量是决定辨认时间的最主要因素。因而在大数量注册话者的辨认中如何减少辨认所需要的运算时间是一个关键问题。针对这一问题，提出了一种新的基于“码袋”的话者模型设计算法，它通过统计“码袋”中每个码字在话者语音中的概率分布来实现话者模型的设计。实验结果表明该算法在保证较高辨认率的同时，有效地降低了话者自动辨认系统的计算复杂度。

关键词：码袋；矢量量化；话者自动辨认
中图分类号：TP391 **文献标识码**：A **文章编号**：1000-3630(2010)-02-0188-04
DOI 编码：10.3969/j.issn1000-3630.2010.02.015

Automatic speaker identification based on “bag of codes”

CHEN Zhe[2], GU Ming-liang[1,2],WANG Jin-song[1], FENG Jing-lan[2], YANG Yi-ming[2]
(1. *School of Phusics and Electronic Engineering,Xuzhou Normal Universitu*, *Xuzhou* 221116, *Jiangsu*, *China*;
2. *Jiangsu Key Laboratory of Language Science and Neural Cognition Engineer*, *Xuzhou* 221116, *Jiangsu*, *China*)

Abstract: The most dominating factor of the identification time is the number of speakers. Thus, how to reduce the computational cost of evaluating large speaker database is the key problem. Due to this, a “bag of codes” algorithm is proposed. This novel algorithm can generate speaker models by estimating the probability distribution of codes in speech data. Experiments prove that the new algorithm can reduce the computational complexity in the speaker identification system with high identification accuracy.

Key words: bag of codes; vector quantization; automatic speaker identification

1 引 言

话者自动辨认始于 19 世纪 50 年代。它通过将未知话者模型与已知的注册话者模型进行比较，从而确认话者身份。话者自动辨认系统在通讯、国防及安全检查等领域均有广阔的应用前景[1]。目前在话者自动辨认系统中应用比较多的数学模型有隐马尔可夫模型、高斯混合模型、动态时间归整模型、矢量量化模型等。其中矢量量化模型，由于其较高的辨认正确率以及简单的原理，成为话者自动辨认系统中应用比较多的模型之一。话者自动辨认本质上是“一对多”的分类问题，因此辨认所需要花费的时间必然会随着注册话者数量的增加而增加。矢量量化是一种计算复杂度很高的算法，这必然会导致传统基于矢量量化的话者自动辨认系统在进行大数量话者的自动辨认时，辨认所需的时间过长。因此如何能够快速实现大数量话者的自动辨认是该领域的研究重点和难点[2]。针对这一问题，借鉴文本分类中经常使用的“词袋”算法思想，将矢量量化算法与之相结合，设计出基于“码袋”的快速话者建模算法。实验结果表明，当话者数量增加时，新系统在保证高辨认率的前提下较传统适量量化系统而言，辨认速度明显提高。从而使得话者自动辨认系统更具有实用价值。

“词袋”模型是一种典型的统计模型。该模型用一串具有代表性的单词描述一类文本，统计这组单词在未知文本中的概率分布得到该文本属于这一类别的概率[3]。在该算法中“词袋”的选择是算法的核心，它必须满足以下两个条件：(1) 清晰地描述文本的类别特征。(2) 尽量避免冗余信息对模型的影响。基于这些条件，在“码袋”的设计中，本文借用矢量量化的码本作为“码袋”，并依据“码袋”对注册话者和未知话者建模。新模型具有“词袋”模型的特点，较适宜于话者模型的快速设计。

2 基于“码袋”算法的话者模型设计及系统实现

2.1 “码袋”话者模型的设计

“码袋”话者模型设计借用矢量量化算法的思

收稿日期：2009-03-15；**修回日期**：2009-05-23
基金项目：徐州师范大学 2008 年度研究生创新计划(08YLB016)
作者简介：陈哲(1983-)，男，江苏徐州人，硕士研究生，研究方向为语音信号处理、模式识别。
通讯作者：顾明亮，E-mail: guml@xznu.edu.cn

想，首先对全体注册话者的语音特征矢量进行量化，设计出统一码本，然后依据码本对每个注册话者的语音特征进行量化，使话者特征空间压缩成一维的码字序列。最后统计每个码字的出现概率，以此生成辨认所需要的话者模型。

2.1.1 “码袋”的设计

在“词袋”模型中，“词”的选择通常是在专家知识的指导下完成的，使得所选择的“词”有极好的类代表性和可区分性。然而，通常在语音处理中所使用的语音特征并没有明确的物理意义，很难得到专家经验的指导，因此本文借鉴矢量量化的思想来实现“码袋”的设计。矢量量化是一种数据压缩方法。在话者语音处理中，它将话者的语音特征按照设计需要进行最优的聚类，在特征空间中形成不同的胞腔，求出每个胞腔的中心矢量，将其作为该特征集合的码本。它可以大致描绘出数据空间的几何分布状态。但是由于干扰信息如噪音段、静音段等的影响，矢量量化码本中的码字并不能完全地满足类代表性的要求。为了减少这些冗余信息对模型影响，本文对矢量量化后的码字进行加权，最终形成我们所需要的“码袋”。

2.1.2 权重矢量的设计

在话者特征提取的实验中，我们发现很多冗余信息如静音段以及噪声点等会对话者特征产生不同程度的干扰。鉴于此，本文设计了一套“码袋”的加权方法，其中每一码字权重的大小与相应胞腔中所含的冗余信息量成反比。该权重矢量的计算方法如下：

(1) 对静音段和噪声段等冗余信息进行采样，得到 N 个样本。

(2) 提取这些采样的声学特征，得到冗余信息采样的特征矩阵 F ($f_n \in F$ $n=1,\cdots,N$)。

(3) 根据矢量量化码本($y_l \in Y_R$)，对冗余信息采样的特征矩阵进行矢量量化。

$$CN_n = \arg\min_{l \in R}(d(f_n, y_l)),\ n=1,\cdots,N,\ l=1,\cdots,R \quad (1)$$

(4) 建立 CN 序列的概率统计直方图，得到冗余信息的统计序列 B ($b_l \in B$, $l=1,\cdots,R$) 。并据此设计出与码本相对应的权重矢量 $\boldsymbol{W}$，其元素 $w_l \in W$, $l=1,\cdots,R$ 为：

$$\begin{cases} w_l = 1/b_l & b_l \neq 0 \\ w_l = 1 & b_l = 0 \end{cases} \quad (2)$$

2.2 基于“码袋”模型的话者自动辨认系统

话者自动辨认系统由四部分组成：语音信号预处理、话者特征提取、话者模型设计与匹配以及后端分类器，如图 1 所示。

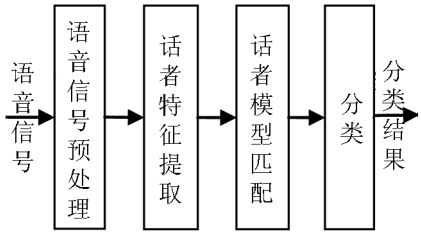


图 1　系统框图

Fig.1　System diagram

2.2.1 语音信号的预处理及话者特征的提取

语音信号的预处理过程包括语音的加窗、采样、预加重、去噪。这些工作一方面保证了语音数据能够被计算机合理地加以应用，另一方面补偿了语音在传输过程中的损失以及减少噪声、静音等干扰信息对后期处理的影响。

在话者自动辨认中，基于 Mel 频率的倒谱系数(MFCC)是一种常用的语音特征[4]。它反应了人类的听觉感知特性。本文将其运用到话者模型的设计中。

2.2.2 模型间的匹配度量及分类器设计

最近邻法是线性分类器设计中一种较为常用的方法。该方法最早由 Cover 和 Hart 于 1968 年提出，至今最近邻法仍是非参数模式识别中最重要的方法之一[5]。在最近邻法中，模式间的差异度量是该方法的核心。为了减少静音段以及噪声点对话者模型的影响，本文采用加权的欧式距离作为模型间差异度的度量函数：

设有 K 个注册话者模型 $m_k, k=1,2,\cdots,K$，未知话者模型为 m_x、权重矢量为 $\boldsymbol{W}$、欧氏距离度量为 D，则 m_k 类的判别函数为：

$$g_k(m_x) = D(W \cdot m_k, W \cdot m_x),\ k=1,2,\cdots,K \quad (3)$$

若 $g_j(m_x) = \min_k g_k(m_x),\ k=1,2,\cdots,K$

则决策 $x=j$

3　“码袋”模型与矢量量化模型的比较

模型设计与分类系统框图如图 2 所示。

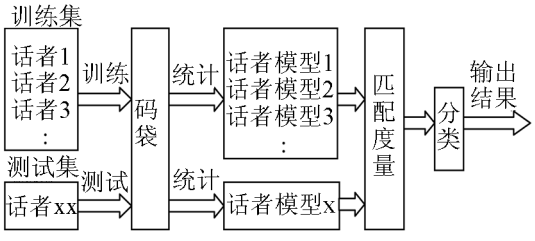


图 2　模型设计与分类系统框图

Fig.2　Diagram of model design and classifier

从图 2 可以看到：基于“码袋”的话者自动辨

认系统与传统基于矢量量化的系统相比，最大的特点是话者模型训练和分类时所使用的是统一的“码袋”，而不是针对不同话者所设计的码本。因此在基于“码袋”的话者自动辨认系统中，对未知语音的建模只需要一次矢量量化。假设 $O(XC)$ 为每个话者的语音特征序列 X 在码本 C 中寻找最近邻所需要的时间损耗，N 为话者数量，那么在基于矢量量化的辨认系统中，一次话者自动辨认所需要的时间损耗为 $O(NXC)$。而基于“码袋”算法的话者自动辨认系统的时间复杂度为 $O(XC)+O(M)$，其中 $O(M)$ 为统计每个码字出现概率的时间损耗，因为 $O(M)<<O(XC)$，所以系统的时间损耗大约为 $O(XC)$。新系统减少了时间复杂度很高的矢量量化的工作，从而使整个系统的时间复杂度显著降低。这些优点在实验中会更加直观地体现出来。同时，实验结果表明，该话者模型可以较好地体现出话者特性，即由相同话者不同语音设计得到的模型间保持较好的相似性，不同话者的模型间体现出较为明显的差异性。这一结果表明新系统可以较好地满足与文本无关的话者自动辨认的需要，如图 3 所示。

图 3 中测试模型是由时长为 3s 的与文本无关的话者语音建立而成；训练模型则是由时长为 90s 的与文本无关的话者语音建立而成。从图中我们可以看到，相同话者的模型间保持了较好的相似性，而不同话者的模型间保持了较明显的差异性。这种差异性和相似性，可以用类内差异度和类间差异度的比值 F 来描述。

$$F=\frac{\text{同话者类内差异度}}{\text{异话者类间差异度}}=\frac{D(\text{训练话者模型}\_i,\ \text{测试话者模型}\_i)}{D(\text{训练话者模型}\_i,\ \text{测试话者模型}\_j)}(i\neq j) \tag{4}$$

同时，实验结果表明话者模型以及测试模型的精度与训练及测试语音的长度有关：

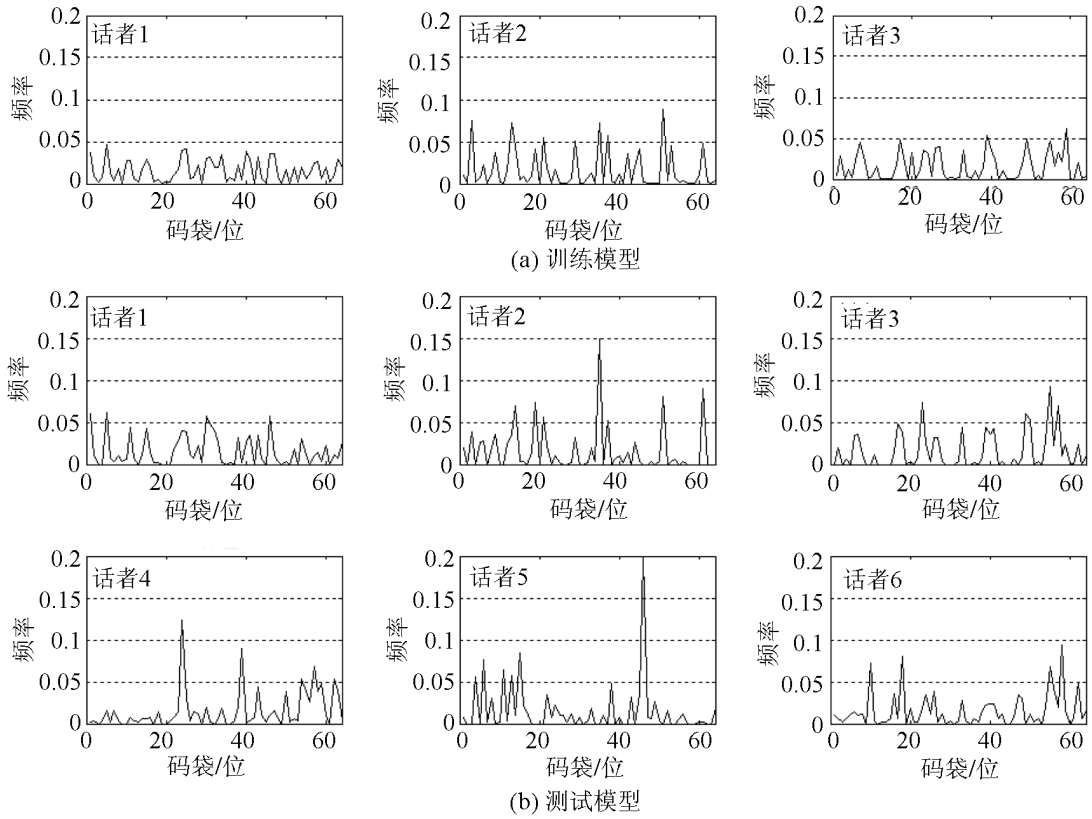
$$F\propto 1/\text{训练语音长度} \tag{5}$$

$$F\propto 1/\text{测试语音长度} \tag{6}$$

4 实验设计及结果分析

为保证对新系统的性能检测科学公正，实验所用语音是在实验室环境下录制而成，选择 30 个 19~23 岁之间的话者作为语音采集对象，性别比例为 1：1。语料的采集时间间隔为三个月，共分三次录制，每次共录制语音 150min(每人 15min)。提取的语音特征为 12 维的美尔倒谱特征。

实验结果表明，新系统可以在较短时间内实现高质量的辨认。如图 4 所示，当注册话者为 5 人、



(a) 训练模型

(b) 测试模型

图 3 话者模型设计（测试集:3s, 训练集:90s）

Fig.3 Speaker model design(test:3s, train:90s)

测试语音长度为 4s 时，新系统的辨认率可以达到95%，而系统平均时间消耗仅为 0.0844s，大约是矢量量化系统的 21.7%。当训练语音长度继续增加时，新系统的速度优势更加明显地体现出来，这点与式(6)的分析相互吻合，如表 1 所示(注册话者数量为 5 人)。同时实验结果也表明随着注册话者数量的增加，系统的时间损耗较传统系统而言也明显减小，如表 2 所示(测试语音长度为 3.2s)。此外，对表 2 中的数据进行分析可知：(1) 新系统的时间损耗约为矢量量化系统的 $1/K$，其中 K 为话者数量。(2) 在不同的注册话者数量情况下，新系统的时间复杂度较为稳定。

表 1　不同长度的测试语音下系统响应速度比较(“码袋”：BOC：矢量量化：VQ)

Table 1　Comparision of system response speed with different test speech duration(“Bag of codes”: BOC; Vector quantization: VQ)

模型	响应速度/s				
	语音长度=1.6s	2.4s	3.2s	4s	4.8s
BOC	0.0312	0.0509	0.0659	0.0844	0.1000
VQ	0.1586	0.2391	0.3079	0.3891	0.4743

表 2　不同数量的注册话者条件下系统响应速度比较(“码袋”：BOC;矢量量化：VQ)

Table 2　Comparision of system response speed with different number of enrolled speakers(“Bag of codes”: BOC; Vector quantization: VQ)

模型	响应速度/s				
	话者数量=5	10	15	20	25
BOC	0.0659	0.0656	0.0660	0.0660	0.0662
VQ	0.3079	0.6583	0.9911	1.3100	1.6511

最后，通过实验对新系统的辨认率进行了对比分析，从实验结果中我们发现，在测试语音较短或且注册话者数量较多的情况下，“码袋”系统的辨认率略低于矢量量化系统，但随着测试语音长度的增加，两者的差异逐渐减小。当测试语音达到 5s 以上时，两者的性能基本相当，如图 4 所示。说明在测试数据量足够的情况下，新系统在降低计算复杂度的同时可以实现较高正确率的话者辨识。

5　总 结

话者自动辨认是话者自动辨识研究领域中的热点，具有非常广泛的应用前景。辨认速度过慢严重制约了话者自动辨认系统的发展，本文将“词袋”算法同矢量量化算法相结合，设计出基于“码袋”算法的话者自动辨认系统。实验结果表明，新系统是一种更加适应大数量话者及高实时性要求的话者自动辨认系统。但是当新注册话者加入时，与传统的话者辨识系统相比，新系统要重新对所有注册话者进行“码袋”模型的设计，从而需要更多的训练时间。因此，在今后的工作中可以进一步研究如何改进“码袋”的设计策略，以提高系统的推广性。

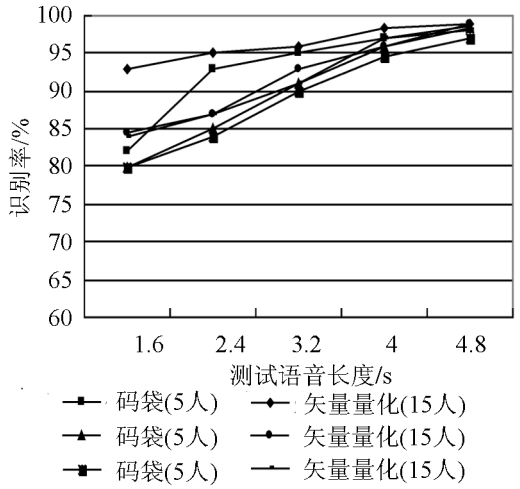


图 4　正确率比较

Fig.4　Comparision of correctness rate

参　考　文　献

[1] Atal B S. Automatic recognition of speakers from their voices[J]. Proceddings of the IEEE, 2005, **64**: 460-475.

[2] Kinnunen T, Karpov E, Fränti P. Real-time speaker identification and verification[J]. IEEE Transactions on Audio, Speech and Language Processing, 2006, **14**(1): 277-288.

[3] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, (3): 993-1022.

[4] 马静, 侯丽敏. 基于全局背景模型和竞争者模型的说话人确认系统[J]. 声学技术, 2007, **26**(1): 105-110.
MA Jing, HOU Limin. Speaker verification based on UBM and cohort model[J]. Technical Acoustics, 2007, **26**(1): 105-110.

[5] Richard O Duda, Peter E Hart, David G Stork. Pattern classification(2ndEdition)[M]. Malden: Wiley-Interscience. 2000.