

## 英语口语机考中英语塞音语音识别的改进型方法研究\*

刘健刚

赵力

(东南大学 外国语学院, 江苏 南京 210096) (东南大学 信息科学与工程学院, 江苏 南京 210096)

**摘要:** 文章通过论证英语口语机考中英语塞音语音识别实现的可能性和实现方法, 试图将一种新型的语音识别方法运用在英语口语机考实践中。文章探讨了“Bayes 判别准则的 2 次判别函数法”改进型语音识别方法的最佳方案并通过实验求出了它的最佳参数, 从而提出了适合于用“Bayes 判别准则的 2 次判别函数法”改进型语音识别方法进行英语塞音语音识别的语音特征参数变换方法。文章通过对英语塞音语音的识别实验, 求证了“Bayes 判别准则的 2 次判别函数法”改进型语音识别方法在英语塞音语音识别过程中具有良好的识别效果。

**关键词:** 英语口语机考, 塞音语音识别, 2 次判别函数法改进型

**中图分类号:** H017

**文献标识码:** A

## 1 研究背景

### 1.1. 大学英语口语机考

我国教育部高等教育司于 2007 年 5 月启动基于计算机和网络的大学英语四、六级考试(以下简称四、六级网考)项目, 并于 2008 年 12 月 20 日在全国 53 所高校实施四级网考试点考试<sup>[1]</sup>。四、六级网考加入了英语口语测试部分, 主要测试重点放在语音模仿方面(单句跟读), 占分比例为 10%。四、六级网考成功地将英语口语考试引入我国无纸质考试的语言考试的实践之中。

从严格的意义上来说, 机考和网考是有本质区别的<sup>[2]</sup>。机考的理念是“Computer Assisted”, 而网考则是“Internet-based”。机考可以是封闭的, 而网考则是开放的。在技术实现上, 网考则需要将网络稳定、网络噪音、网络安全等诸因素考虑进去以确保考试效度的可靠性。

尽管四、六级网考试项目启动已达 4 年之久, 但实际发展情况却令人失望。许多高校对此项目不理解, 存在疑惑和误解。语音模仿能力的测试理论根据是什么? 语调判别的依据是什么? 单句跟读中是否体现语法现象? 四、六级网考的语音识别系统使用的是什么识别系统? 基于何种语音识别方法(计算方法)? 许多大学英语老师和计算机应用工程语音识别的研究人员对此都有不解的疑问, 也无法从四、六级网考网站上获取任何关于四、六级网考评分标准的信息([www.ccets.org](http://www.ccets.org))。

中国语言学界研究人员大多从语言学应用的角度来从事研究, 具体地说就是“拿来主义”地使用国外的研究成果进行对中国英语教学和考试做一些分析和研究。对于如何从语言学角度设计出自主知识产权的计算机运作和分析系统的研究却很少。因此, 语言研究人员联手计算机研究人员共同对语言学运作和分析系统的研究在当今计算机快速发展的时代显得尤为重要。本研究基于单片机运行状态下英语口语机考的口语语音识别建模和具体算法, 不考虑网络运

\* 本文系教育部人文社会科学研究一般项目“英语口语机考评分系统的建模与应用”(10YJA740061)的理论研究阶段性成果、东南大学 2010 年度教学改革项目“大学英语四、六级网考模拟网站的建模与分析应用”(2010-54)研究成果。

行的压力、噪音、语音失真变形、网络环境下智能云计算建模等诸因素。

### 1.2. Bayes 语音识别

常用的语音识别方法一般有模板匹配法、随机模型法和概率语法分析法<sup>[3]193</sup>。模板匹配法就是在语音识别系统中按照简单的模板匹配的原理构造出特定人、小词汇量、孤立词识别系统。在语音识别过程中，将输入语音的特征矢量序列依次与模板库中的每个模板进行相似度比较，将相似度最高者作为识别结果输出，得出识别结果。随机模型法则是使用 HMM 的概率参数来对似然函数进行估计与判决，从而得到识别结果，它是目前语音识别主流的研究途径，被 IBM 和 CMU 的科学家引入语音识别研究，并取得了在语音识别技术上的巨大的成功。概率语法分析法则根据不同人的语音语谱，即语音的“区别性特征”，结合词法、语法、语义诸等约束因素，来实现大长度范围的连续语音识别。

“英语口语机考评分系统的建模与应用”课题组研究人员则采用基于概率语法分析法中 Bayes 判别准则的语音识别方法，利用“2 次判别函数法”（Quadratic Discriminant Function 简称 QDF）<sup>[4]1977-1983 [5] 199-235</sup>来进行识别。所谓符合 Bayes 判别准则的“2 次判别函数法”可表示为识别概率等于待识别类别的似然度概率与先验概率的对数和。即如果  $\vec{X}$

$(x_1, x_2, \dots, x_p)$  是维数为 P 的输入特征向量，k 是参考样本类别， $\vec{\mu}_k$  是 k 类参考样本的均值向量， $\Sigma_k$  是 k 类参考样本协方差矩阵。识别概率  $P(\vec{X} | k)$  等于待识别类别的似然度概率  $\left[ (\vec{X} - \vec{\mu}_k)^T \Sigma_k^{-1} (\vec{X} - \vec{\mu}_k) \right]$  与先验概率  $\ln p(k)$  的对数和。

“2 次判别函数法”较好地解决了处理时间顺序信息问题，但是，在用这种方法，在计算过程中，随着语音特征向量维数的增加，协方差矩阵的推定误差将增大，从而导致识别性能下降。四、六级网考是我国超大规模的标准化考试，2008 年参考人数就已达 1600 多万<sup>[6]4</sup>。在这样海量参考人员中，其语音特征形形色色，导致识别性能明显下降。目前大多院校进行基于计算机英语口语教学与测试的人数亦达上万人次，在计算机运行的语音评测目标实现中也是海量识别。对于这个问题，许多语音识别研究人员大多是从降低特征向量维数的方面着眼，通过 KL 变换等方法尽可能地降低维数以适应“2 次判别函数法”的使用要求，而对“2 次判别函数法”本身的改进等方面的研究和探讨甚少。

“英语口语机考评分系统的建模与应用”课题组研究人员从四、六级网考口语评分建模和学校英语口语计算机测试的角度出发（局域网络传输中的语音失真和变形等因素不在本课题研究范围内），试图提出一种“2 次判别函数法”改进型语音识别方法，修正“2 次判别函数法”（Modified QDF 简称 MQDF）进行英语语音的识别，求出了它的最佳参数，提出适合于用“2 次判别函数法”改进型语音识别方法进行英语语音识别的一种语音特征参数变换方法。通过对英语塞音语音的识别实验，来求证“2 次判别函数法”改进型语音识别方法的有效性。

### 1.3. 英语塞音习得

塞音是所有语言都共有的<sup>[7]99</sup>。二语塞音习得的研究，特别是二语韵首和韵尾辅音群的习得是一个热门的话题，涵盖汉语、英语、朝鲜语、汉语地方语的语音对比。为了说明和方便计算机试验，本文特意取值英汉都有的辅音，且略有细微不同的塞音作为语音识别的实验对象（图 1：英汉辅音音位构音简略比较），从而找出改进“2 次判别函数法”型语音识别的最佳方法。

英汉辅音音位构音简略比较图（图 1）

英汉辅音比较				
汉语	汉语辅音	英语	英语辅音	
双唇音	b, p, m	双唇音 (bilabial)	b, p, m, w	
唇齿音	f	唇齿音 (labio-dental)	f, v	
舌尖音	d, t, n, l	齿槽音 (alveolar)	d, t, n, l, z, c, s, θ	
舌尖前音	z, c, s	后齿槽音 (post-alveolar)	ʃ, ʒ, tʃ, dʒ	
舌尖后音	zh, ch, sh, r			
舌面音	j, q, x	硬腭音 (palatal)	j	
舌根音	g, k, h(?*), ng	软腭音 (velar)	k, g, ŋ	
喉音 (无)		喉音 (glottal)	h(?*)	
英汉塞音比较				
发音方法		汉语 (国际音标)	发音方法	英语 (R. .P)
塞音	不送气清音	b, d, g	stop	浊音
	送气清音	p, t, k	(plosive)	清音
				b, d, g
				p, t, k

从上表可以看出, "b"、"p"、"d"、"t"、"g"、"k" 这六个塞音在构音部位差别不大, 送气和不送气也可以用双源谐波模型的语音型号进行清浊分离<sup>[8]385-386</sup>。因此, 以塞音为切入点, 便于学生和课题组研究人员快速进入实质性的实验。

## 2. 修正 2 次判别函数 (改进型方法)

基于 Bayes 判别准则的“2 次判别函数法”存在着一些令人遗憾的问题。我们可以看到在上面介绍的“2 次判别函数法”中, 待识别类别的似然度概率计算协方差矩阵  $\Sigma_k$  所必要的乘法次数是  $p^2+p$ , 所以当  $X$  的维数增加时, 计算量和内存所占容量会变得很大。更为麻烦的是, 随着  $X$  的维数的增大, 协方差矩阵的推定误差将增大, 从而降低判别性能。为了避免直接计算  $\Sigma_k$  所引起的计算量和计算误差增大的问题, 我们根据协方差矩阵中本征值和本征向量的关系对“2 次判别函数法”进行修正<sup>[9] 2495-2503</sup>, 得到一种改进型“2 次判别函数法”即“修正 2 次判别函数法”。和“2 次判别函数法”相比, 改进后的“修正 2 次判别函数法”能较大地降低计算量和提高判别性能。然而, 为了寻找最佳的“2 次判别函数法”改进型语音识别方法, 课题组研究人员对“修正 2 次判别函数法”又做了进一步的修正。由于高次的本征值和本征向量的推定误差比低次的大的多, 所以课题组研究人员对高于某一阈值  $m$  的误差较大的本征值  $\lambda_{m+1}, \lambda_{m+2}, \dots, \lambda_p$  舍去不用, 而用一固定的等值  $N_0 \sigma^2 / (N+N_0)$  取代。这里  $N$  是学习样本的总数,  $N_0$  是一个信赖度常数,  $\sigma^2$  是一个和本征值有关的量。我们令  $N_0 / (N+N_0) = \alpha$  ( $0 < \alpha < 1$ ) 以及  $\alpha \sigma^2 = V^2$ , 进一步对“2 次判别函数法”进行了改进。另外我们还通过实验选择  $\sigma^2$  为  $\lambda_{m+1}, \lambda_{m+2}, \dots, \lambda_p$  的平均值, 并且找到了  $\alpha$  以及分割点  $m$  的最佳值。

## 3. 用“2 次判别函数法”改进型语音识别方法的英语塞音语音识别实验

“英语口语机考评分系统的建模与应用”课题组研究人员利用上述的改进型方法进行了不特定话者英语六个塞音音族 "b"、"p"、"d"、"t"、"g"、"k" 的识别实验。课题组研究人员选择了《大学一年级语音练习手册》<sup>[10]29-24</sup> 中 IV. Exercises for the English consonants 的练习题, 由 5 名男性在读研究生对每个音节发音三遍, 其中二遍做学习用 (共 1800 个), 一遍作识别用音节 (共 900 个)。这些音节通过 12K Hz, 16bit 的

\* 汉语中的舌根音与英语中的软腭音和喉音类似。但似乎汉语普通话的划分略微粗糙, 将喉音 (glottal)/h/ 也划入舌根音当中。

A/D 变换以及窗长为 20ms、窗移为 10ms 的汉明窗以后，提取 14 阶 LPC me1 倒谱和 1 阶的短时能量组成一个 15 维的特征向量。然后用文献<sup>[13]15-18[12]933-940</sup>所介绍的方法，进行破裂点的检测，再从破裂点前半帧开始取八帧语音信号作为学习和识别用数据。这样每个音节就变成了共有  $8 \times 15 = 120$  维的语音特征向量。

实验中为了求取  $\alpha$  和  $m$  的最佳值，课题组研究人员就不同的  $\alpha$  和分割点  $m$  进行了比较。其结果如图 2 和图 3 所示。图 2 反映了对于某一固定的  $m$  值， $\alpha$  变化时对识别率的影响，从图中可以看出，对于固定值  $m=35$ ， $\alpha$  在 0.3 时出现峰值，达到最高的识别率。图 3 反映了  $m$  和  $\alpha$  均变化时识别率的变化情况，图中的  $\alpha$  是每一个  $m$  的最佳值。实验表明各个分割点  $m$  和它所对应的最佳  $\alpha$  值基本成线性关系。从图 3 可知，当分割点  $m$  在 35 左右时，识别率最佳，最高识别率可达到 96%。

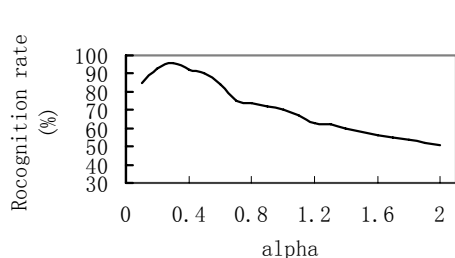


图 2 识别率随  $\alpha$  的变化曲线 ( $m=35$ )

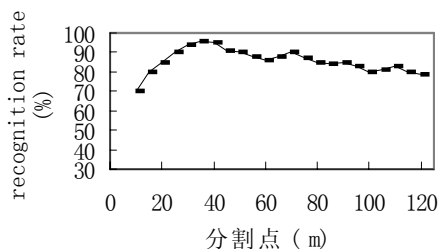


图 3 识别率随  $m$  的变化曲线 ( $\alpha$  为各  $m$  的最佳值)

一般来讲，在利用“2 次判别函数法”的 Bayes 判别准则中，都是假定参数分布是按正态分布的，对于非正态分布的特征来讲，其性能将会下降<sup>[11]106-114</sup>。为此，课题组研究人员假定特征参数只符合近似的正态分布，通过对数变换  $y = \ln(1+x)$ ，使其更加接近于正态分布（如图 4 所示），并且用变换前后的数据进行了比较实验，实验结果如图 5 所示。从图中

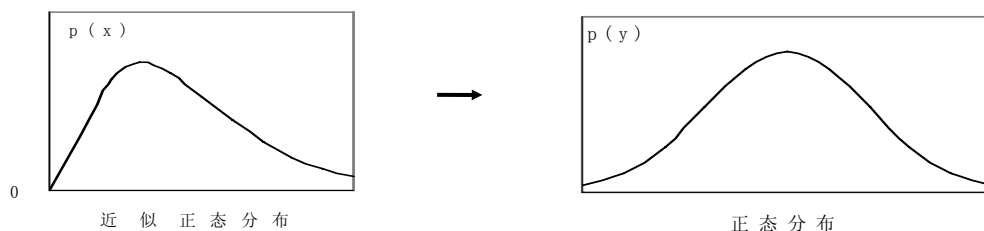


图 4 参数的正态分布变换

可以看出，变换后的最高识别率比变换前提高了近两个百分点，因此证明了这种变换在利用“2 次判别函数法”改进型语音识别方法进行识别时是有效的。

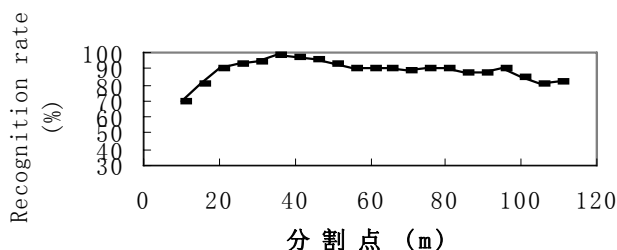


图 5 参数变换后的识别率 ( $\alpha$  为  $m$  的最佳值)

#### 4. 结论

鉴于基于 Bayes 判别准则的“2 次判别函数法”作为一种统计的识别方法今后仍将广泛地应用于语音识别等领域的这一现实情况和事实,“英语口语机考评分系统的建模与应用”课题组研究人员认为对于“2 次判别函数法”的改良在英语口语机考评分系统建模就更显得十分重要。本文探讨的“2 次判别函数法”改进型语音识别方法作为“2 次判别函数法”的一种改进方法,有其理论基础和实用价值,不仅在英语口语机语音识别上有重要意义,而且在基于计算机英语口语教学的实施以及在四、六级网考英语口语评分系统建模中也具有广泛的推广价值。

### 参考文献

- [1] 金艳,吴江. 大学英语四级网考效度初探——影响考生评价和考试成绩的因素分析[J]. 外语电化教学 2010.3.
- [2] 贾国栋. 计算机辅助语言教学——理论与实践[M]. 北京: 高等教育出版社, 2007.3.
- [3] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2009(第二版).
- [4] S.NAKAGAWA and Y.ENOMOTO. Speaker-Independent Phoneme and Word Recognition by Statistical Classification Methods for Time-Sequential Patterns[J]. IEICE 1988.
- [5] 新美康永. 语音识别[M]. 日本东京: 共立出版株式会社, 1978.
- [6] 蔡基刚. 大学英语四、六级网考: 意义和反拨作用[J]. 外语电化教学 2009.9.
- [7] 郑鲜日,李英浩. 英语、朝鲜语塞音的浊音起始时间(VOT)对比[J]. 延边大学学报(社会科学版) 2005.12.
- [8] 梁满贵,肖彦,阮秋琦. 双源谐波模型的语音信号清浊音分离[A]. 《中国声学学会》2006 全国声学学术会议论文集[C] 2006.
- [9] T. WAKABAYASHI and S TSURUOKA et al. On the Size and Variable Transformation of Feature Vector for Handwritten Character[J]. IEICE VOL.J76-D- II .NO.12, 1993.
- [10] 张冠林. A Handbook of English Pronunciation Exercises for the First-Year College Students[M]. 北京: 外语教学与研究出版社, 1999 (13 次印刷).
- [11] H.IWAMIDA and Y.NARA. A Discrimination Method for Stops With Detection of Burst Position[J]. The Preprint of the Acoustical Society of Japan VOL 2, 1990.
- [12] Zhao.Li, Y.Kobayashi,Y.Niimi. Tone recognition of Chinese continuous speech using continuous HMMs[J] 日本音响学会论文志, VOL. 53(12), 1997.
- [13] L.Lei and H.Fujisaki. Prosodic Features of Chinese Continuous Speech[J]. IEICE Technical Report, SP90-105 1990.

## A Study on Realization of English Stop Consonants Recognition

## in Computer-assisted English Oral Test Based on Revised “Quadratic Discriminant Function” Model

Liu Jian-gang

( School of Foreign Languages, Southeast University) (School of Information Science and Engineering, )  
( Nanjing, Jiangsu Province 210096, China)

Li Zhao

(Southeast University, Nanjing,)  
(Jiangsu Province 210096, China)

**Abstract:** Through discussion on modified “Quadratic Discriminant Function” method , this paper presents an approach for recognition of English stop consonants in computer-assisted English oral tests based on modified “Quadratic Discriminant Function”. In this paper, we discussed the best way to work out the “Modified Quadratic Discriminant Function” coefficients and presented a kind of phonetic feature coefficients transferring way fit for the use of “Modified Quadratic Discriminant Function”. Via experiments on the recognition of English stop consonants with the method presented in this paper, it turns out reaching the desirable recognition effect.

**Keywords:** Recognition of English speech in computer-assisted tests; Stop consonants; modified quadratic discriminant function

**作者简介:** 第一作者刘健刚(1959--), 男, 东南大学外国语学院副教授; 主要研究方向为实验语言学(740.3530)、语料库语言学、CALL和iBLL。电邮地址: [jhonliunj@163.com](mailto:jhonliunj@163.com)  
第二作者赵力(1958--), 男, 东南大学信息科学与工程学院教授, 博士生导师; 主要研究方向为语言学模型、说话人和环境自适应方法。电邮地址: [zhaoli@seu.edu.cn](mailto:zhaoli@seu.edu.cn)