## 面向普通话辅音检测的区别特征参数测量

冯晓亮,孟子厚

(中国传媒大学传播声学研究所,北京100024)

摘要: 以建立基于特征参数的解析化的普通话辅音发音检测方法为应用目的。根据普通话辅音按发音方式和发音部 位分类的特点并结合区别特征理论的二元对立思想,首先提取分析了普通话 21 个辅音的美尔倒谱系数 MFCC 和美 尔滤波器能量这两类特征,并进一步得到能区别发音方式或发音部位的区别性特征参数 k1~k11。在此基础上构建了 面向普通话辅音检测的二元分类决策树。与基于 HTK 的分类结果比较表明:使用基于区别性特征参数的决策树判 决的方法对辅音进行分类检测和识别的结果比较稳定,准确率大多在 80%以上且有更好的鲁棒性。

关键词: 普通话; 辅音; 区别特征参数; 语音检测; MFCC;

中图分类号: TN912.34 文献标识码: A

DOI 编码: 10.3969/j.issn1000-3630.2010.03.013

## Distinctive parameter survey of mandarin consonants for speech evaluation

#### FENG Xiao-liang, MENG Zi-hou

(Communication Acoustics Laboratory, Communication University of China, Beijing 100024, China)

Abstract: This paper is aimed at build a mandarin consonants recognition system based on distinctive parameters. According to the classification method of consonants based on articulatory position and manner in combination with binary opposition of distinctive features theory, the two types of features: Mel frequency cepstrum coefficients (MFCC) and Mel-filter energy are analyzed, and then distinctive parameters k1~k11corresponding to articulatory position and manner are further derived. A decision tree based on these parameters is built for classification of consonants. The consonants can be recognized by a binary search process from the decision tree. The comparison with HTK results shows that the recognition accuracy of most mandarin consonants based on distinctive parameters is over 80% and has a better robustness.

Key words: mandarin; consonant; distinctive parameter; speech evaluation; MFCC;

1 引言

区别特征的原始概念起源于 20 世纪 30 年代的 "布拉格学派",其代表人物 Trubetzkoy 创立了音 位对立的二分概念。Jakobson 在 Trubetzkoy 的基础 上把语音的本质特点归纳为十二对区别特征<sup>[1,2]</sup>,此 后 Chomsky, Ladefoged 等学者在区别特征数目和 多元特征对立方面对区别特征理论进行了扩充<sup>[3,4]</sup>。 在国内,吴宗济<sup>[5,6]</sup>建立了普通话区别特征矩阵,并 探讨了它们之间的相互关系模型。

区别特征理论的出现使人们认识到音位并不是 语音的最小结构单位,在音位内部还可以进一步细 化分析。此外,区别特征的思想也影响到了许多应 用领域。张家騄根据知觉混淆结果,构建了普通话 声韵调区别特征体系<sup>[7,8]</sup>,与传统区别特征理论不同,这一区别特征系统是面向实际应用的。在言语 工程中,语音错误或是语音缺陷往往发生在最小对 立体之间,利用区别特征体系,将辅音的检测过程 转化为一系列二元判别。通过上述方法不但可以对 辅音进行识别,还可以找出对应的语音特征,为评 价发音质量提供依据。目前对于区别特征的声学参 数研究还较少,研究区别特征参数有利于区别特征 理论在言语工程中的应用,对于研究语音的不变特 征也有参考价值。本文从区别特征二元对立的思想 出发,根据普通话辅音的不同特征提取了对应的参 数,在此基础上建立了以特征参数为基础的决策 树,实现了辅音分类模型的构建。

文章编号: 1000-3630(2010)-03-0297-09

## 2 辅音区别特征参数的线索

寻找普通话特征参数的线索,一种方式是根据 已有的辅音区别特征寻找对应的区别特征参数。一

收稿日期: 2009-03-05; 修回日期: 2009-06-23

作者简介:冯晓亮(1982-),男,上海人,博士研究生,研究方向为语音 评测。

通讯作者:孟子厚, E-mail: mzh@cuc.edu.cn

般认为区别特征是语音功能的聚合特征<sup>19</sup>,即每一 个区别特征对辅音的判别标准是一致的,但在参数 的层面却难以实现。例如: b 与 g 这对辅音,它们 的不同体现在"聚集的/发散的"这对区别特征 上, b 为发散的,而 g 为聚集的。类似的, z 与 zh 也在"聚集的/发散的"这对区别特征上产生差 异。尽管这两对辅音的区别特征是相同的,但在发 音部位上, b 与 g 的差异在于双唇音与舌根音,而 z 与 zh 的差异在于舌尖音与卷舌音,所以适用于区 分 b 与 g 的参数可能在区分 z 与 zh 时失效。

另一种方式是根据辅音的语音特征,结合二元 对立的特点寻找对应的参数,普通话辅音的语音特 征主要表现在发音部位和发音方式上,按照这一方 法对普通话全部辅音进行分类,其结果如表 1 所 示。由于辅音的语音特征作用于音素这一层次<sup>[9]</sup>, 对辅音的分类更为具体和明确。尽管采用这一方式 在参数的数量上不够精简,会在参数之间产生冗 余,但这种冗余可以保证系统的可靠性<sup>[8]</sup>。因此, 本文根据普通话辅音在发音上的特征,结合二元对 立的特点,寻找对应的参数。

表 1 普通话辅音的分类 Table 1 Classification of mandarin consonants

		双唇	唇齿	舌尖	舌根	舌面	卷舌
केंग्र केंग्र	非送气	b[p]		d[t]	g[k]		
基百	送气	p[p <sup>h</sup> ]		t[t <sup>h</sup> ]	k[k <sup>h</sup> ]		
宝坛立	非送气			z[ts]		j[tc]	zh[tş]
基摖首	送气			c[ts <sup>h</sup> ]		q[tc <sup>h</sup> ]	ch[ts <sup>h</sup> ]
協立	清音		f[f]	s[s]	h[x]	x[¢]	sh[ş]
综百	浊音						r[z]
鼻音	浊音	m[m]		n[n]			
边音	浊音			1[1]			

## 3 基于发音方式分类的区别特征参数

本文中使用的语料库包括 212 位女生和 126 位 男生,年龄范围是 18-22 岁。本文使用的声母语料 是从发音较为清晰的单音节中,通过人工切分的方 式得到的。其中 40%的语料用于寻找与特征对应的 参数,剩余 60%用于测试这些参数的有效性。

## 3.1 清擦音、塞擦音与其他辅音的区别特征参数

美尔频率倒谱系数 MFCC 是语音分析中最常用 的声学线索之一,包含一维对数能量和 12 维符合 人耳听觉特性的参数,由于对数能量受环境影响较 大,结果不稳定,因此将其舍去,保留其余 12 维。之前的实验表明,由于辅音持续时间较短,辅 音在被分帧后,各帧之间的差异不大<sup>[10]</sup>,因此使用 各帧的 MFCC 能量平均值并进行归一化后的结果来 表示该辅音的总体特征。这里用 *M<sub>i</sub>* 表示第*i* 维 MFCC 值(*i*=1,2,…,12)。图 1 为全部辅音的 MFCC 曲线。



从图 1 中可以发现, 普通话辅音可以被分为两 个大类, 一类辅音的特征是 $M_1 < M_2$ , 这类辅音包 括两种不同发音方式: 塞擦音和清擦音(h[x]除 外)。另外一类辅音的特征是 $M_1 > M_2$ , 这类辅音包 括塞音、鼻音、边音、浊擦音和 h[x], 通过该特征 可以将辅音划分为两类。设特征参数为 $k_1$ :

 $k_1 = M_1 - M_2$ 

(1)

对全部的辅音计算其 k<sub>1</sub>值,发现这两类辅音的 概率密度分布符合正态分布,如图 2 所示,将判决 阈限设为两类辅音概率密度分布的交点-0.1。表 2 为全部辅音在 k<sub>1</sub>下分类准确率。



# 3.2 鼻音、边音、非送气塞音、浊擦音与送气辅音和舌根音的区别特征参数

通过参数 k<sub>1</sub>, 全部辅音被分成两组, 其中 k<sub>1</sub>> -0.1 的一组有 11 个辅音(b[p], p[p<sup>h</sup>], d[t], t[t<sup>h</sup>], m[m], n[n], l[l], r[z], g[k], k[k<sup>h</sup>], h[x]), 这些辅音属于五个类 别: 塞音、鼻音、边音、浊擦音和舌根音。这些辅 音的 MFCC 曲线如图 3 所示。

从图 3 中可以发现鼻音、边音、浊擦音以及送

(2)

表 2 特征参数  $k_1$ 的分类结果及其准确率 Table 2 The accuracy of classification using parameter  $k_1$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
k <sub>1</sub> >-0.1	塞音       鼻音       边音       浊擦音       舌根音	b[p], p[p <sup>h</sup> ], d[t], t[t <sup>h</sup> ] m[m], n[n] l[1] r[z] g[k], k[k <sup>h</sup> ], h[x]	98.7%	98.5%
k₁≤−0.1	<sup>塞擦音</sup> 清擦音 (不包括 h[x])	z[ts], j <b>[tɕ]</b> , zh[tʂ] c[tsʰ], q <b>[tɕʰ],</b> ch[tʂʰ] s[s], x <b>[ɕ]</b> , sh[ʂ], f[f]	99.0%	99.9%



图 3 寒音、鼻音、边音、浊擦音和舌根音 MFCC 曲线 Fig.3 MFCC plots of stops, nasals, laterals, voiced fricatives and velars

气塞音的特点是 $M_3 > M_4$ ,而这一组中其余辅音的 特点则是 $M_3 < M_4$ 。这一特征将鼻音(m[m], n[n])、 边音(l[1])、浊擦音(r[z])以及非送气塞音(b[p], d[t])与 送气辅音(p[p<sup>h</sup>], t[t<sup>h</sup>])、舌根音(g[k], k[k<sup>h</sup>], h[x])区分 开来,设特征参数为 $k_2$ :

 $k_2 = M_3 - M_4$ 

计算这 11 个辅音的 k<sub>2</sub> 值,图 4 为这 11 个相关 辅音的 k<sub>2</sub> 概率密度分布,设判决阈限为两曲线的交 点-0.03。表 3 为相关辅音在 k,下的分类准确率。



图 4 塞音、鼻音、边音、浊擦音、舌根音的 $k_2$ 概率密度分布 Fig.4 The probability density distribution of  $k_2$ , of stops, nasals, laterals, voiced fricatives and velars

表 3 特征参数  $k_2$  的分类结果及其准确率 Table 3 The accuracy of classification using parameter  $k_2$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
k <sub>2</sub> >-0.03	鼻音 边音 浊擦音 非送气塞音	m[m],n[n] l[l] r[z] b[p],d[t]	94.2%	91.2%
<i>k</i> ₂≤−0.03	送气塞音 舌根音	p[p <sup>h</sup> ],t[t <sup>h</sup> ] g[k],k[k <sup>h</sup> ], h[x]	98.4%	92.7%

## 3.3 非送气塞音与鼻音、边音、浊擦音的区别特征 参数

m[m], n[n], l[l], r[z], b[p], d[t]这六个辅音属于四 类不同的发音方式,使用 MFCC 参数并没有观察到 这六个辅音之间明显的区别,因此考虑使用分辨能 力更佳的美尔滤波器能量进行区分。美尔滤波器能 量共有 40 维,本文中使用 *fb*(*i*)表示第 *i* 维美尔滤 波器能量。同时对这一参数也进行归一化处理。这 六个辅音的美尔滤波器能量如图 5 所示。



图 5 非送气塞音、鼻音、边音和浊擦音的美尔滤波器能量曲线 Fig.5 Mel filter energies of unaspirated stops, nasals, laterals and voiced fricatives

从图 5 中可发现辅音 b[p]和 d[t]在美尔滤波器 能量的第 5 维至第 10 维,大于其余四个辅音,取 *fb*(5)、*fb*(6)、*fb*(7)这三个参数的平均值作为区分 b[p]和 d[t]鼻音、边音与浊擦音的特征参数,设其 为*k*<sub>a</sub>:

$$k_3 = (fb(5) + fb(6) + fb(7))/3 \tag{3}$$

计算这 6 个辅音的 k<sub>3</sub>值,其概率密度分布如图 6 所示,设判决阈限为两曲线的交点 0.59。表 4 为 这 6 个辅音在 k<sub>3</sub>下的分类准确率。

### 3.4 浊擦音与鼻音、边音的区别特征参数

**r**[**z**], **m**[**m**], **n**[**n**], **l**[**l**]这四个辅音归属于浊擦音、 鼻音和边音三类辅音, 他们的 MFCC 曲线如图 7 所 示, 从图 7 中可以看出浊擦音 **r**[**z**]的 *M*<sub>1</sub>小于 *M*<sub>3</sub>, 而鼻音(**m**[**m**],**n**[**n**])和边音(**l**[**l**])的 *M*<sub>1</sub> 值则大于 *M*<sub>3</sub>。



图 6 非送气塞音、鼻音、边音和浊擦音 k<sub>3</sub> 概率密度分布 Fig.6 The probability density distribution of k<sub>3</sub>, of unaspirated stops, nasals, laterals and voiced fricatives

表 4 特征参数  $k_3$  的分类结果及其准确率 Table 4 The accuracy of classification using parameter  $k_3$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
k <sub>3</sub> >0.59	非送气塞音	b[p], d[t]	91.1%	88.3%
	鼻音	m[m], n[n]		
$k_3 \le 0.59$	边音	1[1]	94.8%	92.1%
	浊擦音	r[z]		



设特征参数为k<sub>4</sub>:

$$k_4 = M_1 - M_3$$

计算这 4 个辅音的 $k_4$ 值,其概率密度分布如图 8 所示,设判决阈限为两曲线的交点 0.05。表 5 为这 4 个辅音在 $k_4$ 下的分类准确率。

#### 3.5 鼻音与边音的区别特征参数

对于 m[m], n[n], l[l]三个辅音, 使用 MFCC 无法 提取有效的参数, 通过美尔滤波器能量进行分析, 如图 9 所示。其中边音 l[l]的 fb(20)大于 fb(27), 而 m[m]和 n[n]则都是 fb(20)小于 fb(27)。设特征 参数为  $k_s$ :

$$k_5 = fb(20) - fb(27) \tag{5}$$

计算这 3 个辅音的 $k_5$ 值,其概率密度分布如图 10 所示,设判决阈限为两曲线的交点—0.02。表 6



图 8 r[z]、m[m]、n[n] 和 l[1]的 k<sub>4</sub> 概率密度分布

Fig.8 The probability density distribution of  $k_4$ , of r[z], m[m], n[n] and l[l]

表 5 特征参数  $k_4$  的分类结果及其准确率 Table 5 The accuracy of classification using parameter  $k_4$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
k <sub>4</sub> >0.05	鼻音 边音	m[m], n[n] l[l]	100%	85.3%
$k_4 \le 0.05$	浊擦音	r[z]	98.1%	95.3%



表 6 特征参数  $k_s$  的分类结果及其准确率 Table 6 The accuracy of classification using parameter k

10010 0 1		Chapping attended to	Paralite	<b>101</b> <i>N</i> 5
会粉	斌立米刊	枯立	准确率	准确率
少奴	<b></b> 相日天空	相日	(女生)	(男生)
k <sub>5</sub> >-0.02	边音	1[1]	91.2%	87.2%
$k_5 \leq -0.02$	鼻音	m[m], n[n]	91.1%	88.3%

为这3个辅音在k,下的分类准确率。

## 4 基于发音部位分类的区别特征参数

上面根据不同辅音的发音方式提取了 5 个特征 参数对辅音进行初步的分类,其中需要根据发音部 位进一步分类的有以下几组:第一组根据 $k_1 \le -0.1$ 而得的塞擦音与清擦音共 10 个辅音,第二组 5 个 辅音( $p[p^h]$ ,  $t[t^h]$ , g[k],  $k[k^h]$ , h[x]),第三组(b[p], d[t]) 和第四组(m[m], n[n])均只有 2 个辅音。以下分别进



Fig.10 The probability density distribution of  $k_5$ , of m[m], n[n] and l[l]

#### 行讨论。

## 4.1 唇齿音与舌尖音、舌面音、卷舌音的区别特征 参数

第一组共有 10 个辅音(j[tɕ], q[tɕ<sup>h</sup>], x[c], z[ts], c[ts<sup>h</sup>], s[s], zh[tş], ch[tş<sup>h</sup>], sh[ş], f[f]), 按照发音部位 可以分为 4 类: 唇齿音、舌尖音、舌面音和卷舌 音。图 11 所示为这 10 个辅音的 MFCC 曲线。

在这 10 个辅音中, f[f]是唯一的唇齿音, 其他 9 个辅音均和舌部相关。从图 11 中可以发现, f[f]和 其他辅音不同的特点在于它的 $M_2$ 小于 $M_3$ 并且 $M_3$ 小于 $M_4$ , 而其他与舌部相关的辅音并没有这样的 特点。设布尔型特征参数 $k_6$ 为:

$$k_6 = (M_2 < M_3 \&\&M_3 < M_4) \tag{6}$$

其中**&&** 表示"与"运算,表 7 为相关辅音在 *k*<sub>6</sub>下的分类准确率。





表 7 特征参数  $k_6$  的分类结果及其准确率 Table 7 The accuracy of classification using parameter  $k_6$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
$k_6 = 1$	唇齿音	f[f]	88.8%	86.3%
	舌面音	[¢]x ,[ <sup>h</sup> at]p ,[عt]j		
$k_6 = 0$	舌尖音	z[ts], c[ts <sup>h</sup> ], s[s]	92.1%	90.3%
_	卷舌音	zh[tş], ch[tş <sup>h</sup> ], sh[ş]		

## 4.2 舌面音与舌尖音、卷舌音的区别特征参数

由特征参数 $k_6$  划分出了 3 类与舌部相关的辅音,它们是舌面音、舌尖音和卷舌音。其 MFCC 曲 线如图 12 所示。从图 12 可以发现,三个舌面音的特征为 $M_2$ 大于 $M_4$ ,而其余两类辅音则是 $M_2$ 小于 $M_4$ ,设特征参数为 $k_7$ :





图 12 舌面音、舌尖音与卷舌音的 MFCC 曲线 Fig.12 MFCC plots of front palatals, blade-alveolars and retroflexes

计算这 9 个辅音的 k<sub>7</sub> 值,其概率密度曲线如图 13 所示,设判决阈限为两曲线的交点-0.02,表 8 为这 9 个辅音在 k<sub>7</sub> 下分类准确率。



图 13 舌面音、舌尖音与卷舌音的 k<sub>7</sub> 概率密度分布 Fig.13 The probability density distribution of k<sub>7</sub>, of front palatals, blade-alveolars and retroflexes

表 8 特征参数  $k_7$  的分类结果及其准确率 Table 8 The accuracy of classification using parameter  $k_7$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
k <sub>7</sub> >0.02	舌面音	[ə]x ,[ <sup>h</sup> at]p ,[at]į	98.4%	97.2%
k7≤0.02	舌尖音 卷舌音	z[ts],c[ts <sup>h</sup> ], s[s] zh[tş], ch[tş <sup>h</sup> ], sh[ş]	96.1%	94.3%

#### 4.3 舌尖音与卷舌音的区别特征参数

舌尖音与卷舌音包括 6 个辅音,这两类辅音十 分容易混淆,在很多方言中经常要根据上下文环境 才能分辨。它们的 MFCC 曲线如图 14 所示。从图 14 发现卷舌音的 $M_3$ 大于 $M_2$ 与 $M_4$ 的平均值,而舌 (8)



尖音则恰好相反,设特征参数为 $k_8$ :  $k_8 = M_3 - (M_2 + M_4)/2$ 

计算舌尖音与卷舌音的 k<sub>s</sub>值,它们的概率密度 分布如图 15 所示,设判决阈限为两曲线的交点 0。 表 9 为这 6 个辅音在 k<sub>s</sub>下的分类准确率。



Table 9 The accuracy of classification using parameter  $k_s$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
$k_8 > 0$	卷舌音	zh[tş], ch[tşʰ], sh[ş]	96.9%	98.2%
$k_8 \leq 0$	舌尖音	z[ts], c[ts <sup>h</sup> ], s[s]	98.8%	93.1%

## 4.4 送气塞音与舌根音的区别特征参数

前 3 个小节对第一组 10 个辅音进行了分类。第 二组中包含 5 个辅音(p[p<sup>h</sup>], t[t<sup>h</sup>], g[k], k[k<sup>h</sup>], h[x]), 其中 p[p<sup>h</sup>], t[t<sup>h</sup>]按发音部位划分为双唇音和舌尖 音,而 g[k], k[k<sup>h</sup>], h[x]则均为舌根音,它们的美尔 滤波器能量如图 16 所示。从图 16 可发现,p[p<sup>h</sup>]、 t[t<sup>h</sup>]在第 3 至第 7 维参数中有两个局部的最大值, 而舌根音 g[k], k[k<sup>h</sup>], h[x]则不具有上述特点,设特 征参数为*k*。:

$$k_9 = (fb(3) + fb(4))/2 + (fb(6) + fb(7))/2$$
(9)

计算这 5 个辅音的 k, 值, 它们的概率密度分布 如图 17 所示, 设判决阈限为两曲线的交点 0.8。表 10 为 k, 的分类结果。



图 17 舌根音与  $p[p^h]$ 、 $t[t^h]$ 的  $k_9$  概率密度分布 Fig.17 The probability density distribution of  $k_9$ , of velar,  $p[p^h]$  and  $t[t^h]$ 

表 10 特征参数  $k_9$  的分类结果及其准确率 Table 10 The accuracy of classification using parameter  $k_9$ 

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
k <sub>9</sub> >0.8	送气塞音	p[p <sup>h</sup> ], t[t <sup>h</sup> ]	87.9%	89.0%
$k_9 \leq 0.8$	舌根音	g[k], k[k <sup>h</sup> ], h[x]	94.4%	91.9%

## 4.5 双唇塞音与舌尖塞音的区别特征参数

由于特征参数 $k_9$ 将 p[p<sup>h</sup>], t[t<sup>h</sup>]分离出来,可以将 b[p], p[p<sup>h</sup>], d[t], t[t<sup>h</sup>]这四个辅音一起分析,按照发音 部位划分, b[p], p[p<sup>h</sup>]为双唇塞音而 d[t], t[t<sup>h</sup>]则是舌 尖塞音。它们的 MFCC 参数如图 18 所示。从图 18 可以发现 b[p], p[p<sup>h</sup>]的 $M_2$ 大于 $M_3$ 而 d[t], t[t<sup>h</sup>]则是  $M_2$ 小于 $M_3$ 。设特征参数为 $k_{10}$ :

$$k_{10} = M_2 - M_3 \tag{10}$$

计算这 4 个辅音的 k<sub>10</sub> 值,它们的概率密度分布 如图 19 所示,设判决阈限为两曲线的交点-0.02。 表 11 为 k<sub>10</sub> 的分类结果。

## 4.6 双唇鼻音与舌尖鼻音的区别特征参数

最后分析 m[m]和 n[n]这两个鼻音, 其中 m[m]



图 19 b[p]、p[ph]、d[t]、t[th]的 k<sub>10</sub> 概率密度分布图 Fig.19 The probability density distribution of k<sub>10</sub>, of b[p], p[p<sup>h</sup>], d[t], t[t<sup>h</sup>]

表 11	特征参数 k <sub>10</sub>	的分类结果及其	准确率	
Table 11 The	accuracy of cl	assification using	g parameter	k,

参数	辅音类型	辅音	准确率 (女生)	准确率 (男生)
$k_{10} > -0.02$	双唇塞音	b[p], p[p <sup>h</sup> ]	87.9%	86.2%
$k_{10} \le -0.02$	舌尖塞音	$d[t], t[t^h]$	82.8%	81.1%

是双唇鼻音,而 n[n]是舌尖鼻音。使用美尔滤波器 进行能量分析,如图 20 所示。从图 20 可以发现, m[m]的 *fb*(9) 至 *fb*(1) 区间的平均值大于 *fb*(18) 至 *fb*(20) 区间的平均值,而 n[n]在这两区间内的能量 则较为接近。设特征参数 *k*<sub>1</sub>:

$$k_{11} = i_1 - i_2$$

$$i_1 = (fb(9) + fb(10) + fb(11))/3$$

$$i_2 = (fb(18) + fb(19) + fb(20))/3$$
 (11)

计算 m[m]和 n[n]这两个辅音的 k<sub>11</sub> 值,其概率 密度分布如图 21 所示,设判决阈限为两曲线的交 点 0.15。表 12 为 k<sub>11</sub> 对这两个辅音的分类准确率。

5 辅音区别特征参数树状图

## 5.1 辅音区别特征参数树状图的构建

根据以上各个层次的特征参数,可以构建一个 普通话辅音区别特征参数树状图,如图 22 所示。可



表 12 特征参数  $k_{11}$  的分类结果及其准确率 Table 12 The accuracy of classification using parameter  $k_{11}$ 

	•		0.	11
参数	辅音类型	辅音	准确率	准确率
			(女生)	(男生)
$k_{11} > 0.15$	双唇鼻音	m[m]	77.9%	80.4%
$k_{11} \le 0.15$	舌尖鼻音	n[n]	76.8%	83.1%

以发现在树状图的上层,主要依靠发音部位的不同 进行区分,而在树状图的底层则是根据辅音的发音 方式进行区别。说明文中使用的 MFCC 和美尔滤波 器能量参数对于不同发音部位的区别效果要强于不 同的发音方式。





#### 5.2 普通话辅音的分类结果

根据上面的特征参数树状图,每一个辅音对应 树状图中的一条路径,通过路径中的节点所对应的 参数进行二元判决就可以对语音进行分类。例如: 使用特征参数 k<sub>1</sub>、k<sub>2</sub>、k<sub>3</sub>和 k<sub>10</sub>就可以对 b[p]进行 判别。利用剩余 60%的测试数据进行验证实验,提 取这些辅音相关的 MFCC 参数以及美尔滤波器能量 参数,根据每个特征参数的要求进行计算。表 13 给出了决策树对测试数据的判别结果。

表 13 普通话辅音特征参数体系判别输出结果 Table 13 The accuracy of consonants recognition using the decision tree

辅音	准确率 (女生)	准确率 (男生)	辅音	准确率 (女生)	准确率 (男生)
f[f]	85.6%	83.6%	r[z]	92.3%	84.3%
[at]j	96.3%	97.3%	1[1]	75.4%	82.5%
[ <sup>م</sup> يt]p	92.3%	90.1%	m[m]	75.3%	76.3%
x[¢]	88.1%	86.3%	n[n]	72.3%	80.2%
z[ts]	96.1%	94.3%	b[p]	86.5%	81.2%
c[ts <sup>h</sup> ]	89.4%	92.6%	p[p <sup>h</sup> ]	83.8%	80.2%
s[s]	84.6%	92.6%	d[t]	79.3%	81.6%
zh[tş]	87.1%	86.2%	t[t <sup>h</sup> ]	78.4%	79.3%
ch[tş <sup>h</sup> ]	86.5%	85.4%	g[k]	85.6%	86.9%
sh[ş]	85.4%	85.1%	k[k <sup>h</sup> ]	84.5%	86.7%
			h[x]	82.6%	83.8%

从普通话辅音判别的输出结果表明,对于大多数辅音,使用 MFCC 和美尔能量参数的正确率集中在 80%以上,而一些在知觉相似度较为接近的辅音,如 m[m],n[n]<sup>[7]</sup>,结果也低于平均水平,需要进一步从其他角度提取特征参数区分它们。

在文献[7]中,作者根据噪声条件知觉实验,给 出了普通话辅音相似度的树状结构图。从分层的结 构上,这一结果优先使用发音方式对辅音进行区 别。而本文这一参数体系建立在语音特征与声学参 数关联的基础上,对于发音部位的区别效果较好。 其原因可能是噪声传递条件对人们的认知线索产生 了影响。虽然两者出发点不同,但根据主观知觉的 实验结果,为实际应用中利用特征参数给出了目标 和方向。为了能与主观评测结果有着更好的一致 性,本文提取的参数在判别过程中的权重可能还需 要进一步调整。

#### 5.3 基于统计识别方法

表 13 给出了基于规则的识别结果,而统计识别 方法仍是较为常用的技术手段。为了评价区别特征 参数体系的判别效果,就需要建立一个用于对比的 基线系统,基线系统采用比较成熟的隐含马尔科夫 模型 HMM,利用 HTK 工具包搭建识别系统。 在训练的过程中,将帧长定为 16ms,特征参数同样选取 12 维的 MFCC 参数。由于该 HMM 系统只是针对辅音搭建的,因此只需定义 4 个状态,就可完全满足要求。在本基线系统中,每一个状态设为 4 个高斯混合度时,系统的效果最佳。系统在训练时使用了 40%的数据,其余 60%的数据用于测试。系统按照 HTK 标准的 15 次训练过程对系统模型参数加以修正。集外测试的结果如表 14 所示。

表 14 基于 HTK 的普通话辅音特征参数体系判别输出结果 Table 14 The accuracy of consonants recognition using HTK

辅音	准确率	准确率	辅音	准确率	准确率
	(女生)	(男生)		(女生)	(男生)
f[f]	94.3%	97.6%	r[z]	100.0%	98.8%
j[tç]	86.6%	81.4%	1[1]	88.7%	40.7%
q[t¢ <sup>h</sup> ]	76.0%	74.4%	m[m]	87.3%	93.0%
x[¢]	71.8%	62.7%	n[n]	83.1%	36.0%
z[ts]	66.9%	51.1%	b[p]	61.9%	55.8%
c[ts <sup>h</sup> ]	69.7%	40.7%	p[p <sup>h</sup> ]	81.6%	76.7%
s[s]	66.9%	77.9%	d[t]	80.2%	37.2%
zh[tş]	88.0%	79.0%	t[t <sup>h</sup> ]	88.7%	86.0%
ch[tş <sup>h</sup> ]	88.7%	81.4%	g[k]	100.0%	100.0%
sh[ş]	97.1%	96.5%	k[k <sup>h</sup> ]	91.5%	96.5%
			h[x]	97.1%	97.6%

通过表 14 可以发现,采用 HTK 搭建的统计识 别方法对于数据较为敏感,不同语音数据的识别效 果差异较大。而采用区别特征决策树进行分类的结 果比较稳定,说明采用区别特征参数决策树判别的 方法有更好的鲁棒性。从识别效果上,基于区别特 征参数的识别方法对于大多数辅音的识别效果大多 集中在 70%~80%, 而基于统计模型的识别方法, 虽然有的低于 40%, 但对一部分辅音的识别效果高 于特征参数判别方法。如果从提高识别效果的角度 出发,可以考虑结合两种方法对于不同辅音的判别 优势,建立一个融合的系统对辅音加以识别。从数 理基础上看,以 HTK 为代表的 HMM 语音识别算 法以统计数学为基础,统计数学方法往往习惯于寻 找外在特征,这在工程上可能较为实用,但却不能 很好的揭示语音的内在规律。而区别特征理论是一 种解析化的分析方法,对于探讨语音的本质特性是 有启发意义的。

## 6 总结

特征参数对于实现基于非统计模式识别的解析 化的语音识别、语音评测有着重要意义。本文从应 用的角度出发,以区别特征理论的二元对立思想为 基础,根据普通话辅音的具体特征提取了对应的参 数。虽然在特征的数目上不如直接根据已有的区别 特征寻找更为精简,但参数的冗余保证了对辅音区 别的可靠性。通过这些参数构建了普通话辅音特征 决策树。在实际应用中,待测语音经过一系列规则 判别可以转化为在决策树中的一条判决路径,这条 判决路径可以提供丰富的语音特征信息,这些语音 特征信息为语音评测中判别发音缺陷和错误类型提 供了线索。寻找其他特征参数来完善辅音特征参数 体系,并结合人工评测结果调整不同参数的权重, 使得结果与人工评测有更好的一致性,是下一步的 工作重点。

#### 参考文献

- Jakobson R, Fant G, Halle M. Preliminaries to speech analysis, the distinctive features and their correlates[C]. Cambridge, MA, USA: MIT Acoustic Laboratory, Tech Report no.13, 1952. 3-8.
- [2] Jakobson R, Halle M. Fundamentals of language[M]. s-Gravenhage: Mouton, 1956. 20-30.
- [3] Chomsky, Halle, The sound pattern of English[M]. Cambridge, MA, USA: MIT Press, 1991. 64-68.
- [4] Peter Ladefoged. Preliminaries to linguistic phonetics[M]. Chicago

University of Chicago Press, 1971. 91-100.

- [5] 吴宗济. 试论普通话语音的"区别特征"及其相互关系[J]. 中国语 文 1980, 5(3): 321-327.
   WU Zongji. The distinctive features of mandarin and their relationships[J]. Chinese, 1980, 5(3): 321-327.
- [6] 吴宗济,林茂灿. 实验语音学概要[M]. 北京:高等教育出版社, 1989. 264-268.

WU Zongji, LIN Maocan. The foundations of experimental phonetics[M]. Beijing: Higher Education Press, 1989. 264-268.

[7] 张家騄. 汉语普通话区别特征系统[J]. 声学学报, 2005, 30(6): 506-514.

ZHANG Jialu. The distinctive features system of standard Chinese[J]. ACAT Acoustica, 2005, **30**(6): 506-514.

[8] 张家騄. 汉语普通话区别特征系统树状图[J]. 声学学报, 2006, 31(3): 193-198.

ZHANG Jialu. The distinctive features tree of standard Chinese[J]. ACAT Acoustica, 2006, **31**(3): 193-198.

- [9] 马学良. 普通语言学[M]. 北京: 中央民族大学出版社, 1997. 88.
   MA Xueliang. General linguistics[M]. Beijing: Minzu University of China Press, 1997. 88.
- [10] 冯晓亮, 孟子厚. 舌尖辅音和舌面辅音的区别特征参数[A]. 第八届中 国语音学学术会议[C]. 北京, 2008. 149. FENG Xiaoliang, MENG Zihou. The distinctive parameters for blade of tongue/tip of tongue consonants in mandarin[A]. The 8th Phonetic Conference of China[C]. Beijing, 2008. 149.