

# 基于语音参数的普通话韵母区别特征

王孟杰, 孟子厚

(中国传媒大学传播声学研究所, 北京 100024)

**摘要:** 现有区别特征理论缺少相应语音声学参数的支持, 尝试从声学参数的角度来建立普通话韵母的区别特征。首先根据语音参数距离, 使用层次聚类方法生成韵母决策树。在此基础上, 按开尾韵和有尾韵的韵母分类方案, 分析其聚类逻辑, 提出了基于参数的普通话韵母区别特征。在非介音韵母中, 区别特征层次首先根据主元音音位的高低区分; 而在介音韵母中首先通过介音进行区分。使用支持向量机检测区别特征, 韵母由区别特征节点组成的搜索路径所确定。通过在区别特征树上的搜索过程实现韵母的检测分类, 其中非介音韵母的平均准确率为 93.1%, 介音韵母为 92%。

**关键词:** 韵母; 区别特征; 语音参数

中图分类号: TN912

文献标识码: A

文章编号: 1000-3630(2011)-01-0088-05

DOI 编码: 10.3969/j.issn1000-3630.2011.01.016

## Research on distinctive features of mandarin finals based on speech parameters

WANG Meng-jie, MENG Zi-hou

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

**Abstract:** There is little study about acoustic cues corresponding to the existing theory of distinctive features, this paper attempts to establish distinctive features of Mandarin finals based on speech parameters. Firstly, decision trees of Mandarin finals are constructed by using multi-layer clustering methods based on distances of speech parameters. According to finals classification scheme which include rhyme and open rhyme, and by analyzing the cluster logic of Mandarin finals' decision tree, the distinctive features based on acoustic parameters are constructed. The distinctive features are arranged by the phoneme of main vowel firstly in finals without glides, and by the glides head in finals with glides. Distinctive features are checked by support vector machine and finals are identified according to a specific search path in distinctive feature trees. The accuracy of finals without glides is 93.1% and with glides is 92%.

**Key words:** finals; distinctive features; speech parameters

## 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 首先创立了音位对立的二分(区别特征)概念<sup>[1]</sup>。其后, Jakobson 把所有语言的区别特征列为十二项“最小对立体”, 作为分析语音的标尺<sup>[2]</sup>。上世纪 60 年代后, Chomsky 和 Halle 提出的区别特征体系中, 对二元特征做了修订, 允许多元特征的存在<sup>[3]</sup>。国内对普通话的区别特征研究开始于上世纪 80 年代。吴宗济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<sup>[4]</sup>。其后, 叶蜚声等列出了“北京话音位的区别特征表”<sup>[5]</sup>。陆致极进一步采用 12 对区别特征

来建立普通话音位的区别特征矩阵<sup>[6]</sup>。最近, 张家骥根据汉语语音知觉混淆的群集分析结果, 建立声母-韵母音位系统的区别特征<sup>[7,8]</sup>。此外, 王洪君对韵母组合音段的区别特征进行了研究<sup>[9]</sup>。

由于缺少相应声学参数的支持, 现有的区别特征理论很少应用于语音工程之中。虽然张家骥提出区别特征系统时是以声学参数为核心的, 但是并没有实际的语音声学参数作支撑。此外, 王洪君对韵母根据其组合位置(介音、韵腹和韵母), 使用传统区别特征进行划分。由于传统区别特征没有对应的声学参数, 因此在此基础上的组合, 也是没有声学参数支持的。区别特征是语音学中的概念, 而在语音工程中常用的语音参数是从信号处理或是结合听感模型提出的, 这两者从起源上就存在差别, 因此传统的区别特征在应用于语音识别、语音评测等语音工程实践中时可能并不合适。因此, 有必要从语音声学参数的角度考虑来建立区别特征。

收稿日期: 2010-01-30; 修回日期: 2010-04-07

作者简介: 王孟杰(1982-), 男, 安徽淮南人, 博士, 研究方向为语音信号处理。

通讯作者: 孟子厚, E-mail: mzh@cuc.edu.cn

## 2 特征计算

在本文的研究中,使用了在中国传媒大学录制的语音数据库中的数据,发音人为播音专业的学生。采用了语音库中单独录制的普通话韵母语音,每个韵母大约有 300 个左右的样本。本文主要使用了 LPCC、PLP 和基于 MEL 滤波器组的主成分参数来描述语音信号的不同特征。

LPC 的倒谱系数(LPCC),包含了信号谱的包络信息,可以看作是对原始信号短时倒谱的一种近似,对准稳态浊音部分有很好的描述。本文在计算中采用 13 阶的 LPCC 参数,以表征声道特征。同时,PLP(Perceptual Linear Prediction)参数在使用全极点模型计算线性预测系数的过程中,把临界带分析、等响曲线的预加重等人耳听觉特征应用到信号处理中,其结果更加符合人耳的听觉特点。本文根据非特定人的语音识别实验结果,PLP 选择 5 阶参数进行计算<sup>[10]</sup>。

为了反映区别特征在各个频段的特征,使用类似于耳蜗作用的 Mel 频率滤波器组<sup>[11]</sup>。滤波器组包括 40 个滤波器,包含有 13 个线性滤波器和 27 个对数滤波器。滤波器组由混迭的三角窗构成,将每个滤波器频带内的能量进行叠加作为滤波器的输出。Mel 频率滤波器组的输出参数较多,使用主成份分析(Principal Component Analysis)对参数进行降维。当 Mel 滤波器组输出降成 7 维主成份时,累计贡献率为 85%。将滤波器组的输出参数根据载荷矩阵转换为 7 维的 Mel-PCA 参数。将以上三种参数按顺序进行拼接,得到一个包含 25 维参数的特征向量,本文以此为基础构建韵母决策树。

## 3 生成决策树

### 3.1 层次聚类

由于每个韵母有多个语音样本,首先需要对于每个韵母生成标准模板。本文使用改进的 K 均值算法<sup>[12]</sup>(Modified K-Means, MKM)来获得模板的参数。在得到每个韵母的参数模板后,使用层次聚类的方法进行自动聚类的处理。考虑对  $n$  个样本聚成  $c$  类的情况。首先,将所有样本分成  $n$  类,每类正好含有一个样本。然后,将样本分为  $n-1$  类,接下来是  $n-2$  类,一直到所有的样本都被分为一类。把聚类数目  $c=n-k+1$  对应层次结构的第  $k$  层,因此第 1 层对应  $n$  个类别,而第  $n$  层对应一个类别。在层次聚

类中,令  $c=1$ ,就可以生成完整的树形层次结构聚类图。在每一层上,两个最相似的类之间的距离可以反映它们的相似性。

### 3.2 聚类结果

对于语音样本,通过 DTW 算法计算它们之间的匹配距离。使用无需先验知识的凝聚式分层聚类方法进行处理。所有韵母最初各自自成一类,相似度最高的两个韵母首先聚成一个小类,依据重心法计算这个小类和其余类之间的相似度即距离,然后继续是相似度最高的两类聚成一类,照此步骤往复,最终得到分层聚类的结果。为了得到更为简洁的音位系统,本文对于普通话韵母的处理分为介音韵母和非介音韵母两个独立的部分进行,通过分层聚类的方法生成韵母的树形图,见图 1 和图 2。

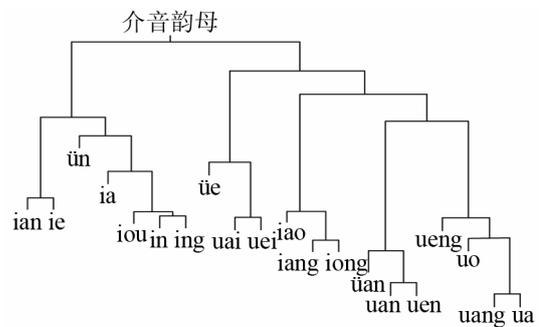


图 1 基于参数距离的介音韵母树形图  
Fig.1 The tree of finals with glides based on parameters distance

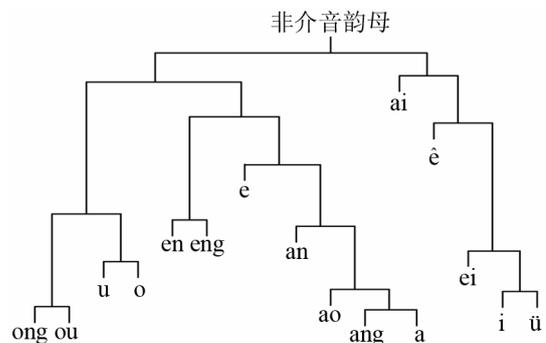


图 2 基于参数距离的非介音韵母树形图  
Fig.2 The tree of finals without glides based on parameters distance

### 3.3 聚类逻辑

韵母有两种分类方案<sup>[9]</sup>。一种是在传统的教学中常用的单韵母、复韵母和鼻韵母的分类方案,另一种是按开尾韵和有尾韵进行划分。后一方案是根据韵母的组成成分,介音和韵尾的异同来分类的,其具有更好的系统价值,有利于对于音系规则的描写。为了研究韵母聚类层次的内在联系,本文也采用这一分类规则在韵母聚类树形图的基础上分析



音韵母,对于两者分别构建区别特征节点分类器,以实现区别特征的检测。在此基础上,通过在树状图上进行搜索,检测韵母具有的区别特征来实现韵母的区分。

### 5.1 节点计算

对于区别特征节点,使用支持向量机(Support Vector Machine, SVM)构建两类分类器<sup>[13]</sup>。在节点的一侧的韵母作为一类,另一侧的韵母作为另一类。使用一半发音人的数据进行模型的训练,另外一半数据进行测试。

SVM 通过寻求结构风险最小化来实现实际风险最小化,从而在样本量较小的情况下也能获得良好的学习效果。训练集向量  $x_i$  通过函数  $\varphi$  被转换到高维空间。SVM 在这个高维空间寻找使训练样本距离最大的线性分界面。其中,

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (1)$$

被称为核函数。这里,选取径向基函数(Radial Basis Function, RBF)为内积核函数,见式(2)。

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

RBF 可以把样本非线性映射到高维空间。所以,与线性的核函数不同,RBF 可以处理样本与特征之间非线性的对应关系。

由于每个韵母语音信号的长度不一致,而对于 SVM 算法来说,语音特征必须具有统一的长度。同时考虑到韵母随时间的动态变化特征,每个语音样本按时间顺序分成 5 段<sup>[14]</sup>,每段之间的时长比例为 1: 2: 4: 2: 1。在每一段内,帧与帧之间的参数进行平均,当作该段的特征参数。然后将 5 段特征向量按时间顺序进行拼接,所得参数作为支持向量机所使用的特征向量。在表 1 和表 2 中分别给出了非介音韵母和介音韵母区别特征检测的结果。对于非介音韵母和介音韵母使用了不同的区别特征表述方式,如非介音韵母特有的“开、齐\合、撮”的区别,介音韵母中特有的“ng 尾\开尾”的区别,此外还有些是两者共有的特征,如“低\非低”等。本文对于所有区别特征的节点检测结果准确率均在 95%以上。

### 5.2 韵母的分类

根据图 5 和 6 中所确定的搜索路径,通过对于

表 1 非介音韵母区别特征检测结果(单位: %)

Table 1 Classification results of distinctive features for finals without glides(unit: %)

低	开尾	元音尾	开齐	开	合	i 尾	n 尾	开
非低	有尾	鼻尾	合撮	齐	撮	u 尾	ng 尾	合
97.3	98.2	97.7	99.2	99.7	100	99.8	98.8	97.3

表 2 介音韵母区别特征检测结果(单位: %)

Table 2 Classification results of distinctive features for finals with glides(unit: %)

齐	合	开尾	元音尾	鼻尾
合撮	撮	有尾	开尾、鼻尾	元音尾
99.1	99.8	98.7	99.0	97.7
ng 尾	ng 尾	低	n 尾	
n 尾	开尾	非低	ng 尾、开尾	
95.4	95.4	97.9	98.3	

不同区别特征节点的检测,对韵母分为无介音韵母和有介音韵母进行区分。比如,以韵母“ia”为例,通过在图 6 中所标示的区别特征节点,以“齐\合、撮”、“开尾\有尾”和“低\非低”三者所确定搜索路径,把“ia”区分出来。测试结果表明,大部分韵母的识别率在 90%以上,详细的分类结果见表 3 和 4。由于部分节点的语音相似度较高,如 ng 尾\ n 尾节点下的韵母 in 和 ing 比此区别特征节点下其它韵母对发音更相近,因此在使用统一的模型训练方式时准确率较低。整体的分类结果为:无介音韵母的平均准确率为 93.1%,介音韵母为 92%。使用区别特征的韵母分类方法在得到较好分类效果的同时,使用了更少的分类器,并且具有更为明确的语音学意义。

表 3 无介音韵母分类结果(单位: %)

Table 3 The classification results of finals without glides(unit: %)

a	o/e	eng	ong	i	u	ü	ai	ei	au	ou	an	en	ang
95	92	91	92	98	97	100	97	99	86	89	94	98	78

表 4 介音韵母分类结果(单位: %)

Table 4 The classification results of finals with glides(unit: %)

ia	ie	ing	in	iao	iou	ian	iang	iong	uo
97	97	76	80	99	91	83	87	95	93
ua	uai	uei	uan	uen	uang	ueng	üe	ün	üan
99	92	99	95	92	80	94	98	98	94

## 6 结论

本文在构建韵母决策树的基础上,通过分析韵母聚类逻辑,提出了基于参数的普通话韵母区别特征。在据此进行的韵母分类实验中,取得了平均准确率在 90%以上的效果。这里使用介音和韵尾的异同来描述区别特征,更为符合汉语的音系规则;而由语音参数决定区别特征的层次安排。区别特征体系的构建可以从语音学知识、声学参数、听感实验等多方面进行。其中,区别特征的声学参数是很重要的一个方面。本文所提出的区别特征,更侧重于音段的区分,以语音参数为基础,面向实际的语音工程应用。区别特征的研究在语音的识别、合成和

语音评测方面都有重要意义。利用区别特征的概念,可以加快搜索速度,提高语音判别的准确率,同时对区别特征的检测,为解决语音评测中语音发音错误类型和发音缺陷的判定问题提供了可能。

### 参 考 文 献

- [1] Jakobson R, Fant G, Halle M. Preliminaries to Speech Analysis, the Distinctive Features and their Correlates[R]. MIT Acoustic Laboratory, Tech Report no.13, 1952.
- [2] Jakobson R, Halle M. Fundamentals of language[M]. Mouton, 1956.
- [3] Chomsky N, Halle M. The Sound Pattern of English[M]. The MIT Press, 1991.
- [4] 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): 321-327.  
WU Zongji. A preliminary study of distinctive features and their correlations in Standard Chinese[J]. Zhongguo Yuwen, 1980, (5): 321-327.
- [5] 叶蜚声, 徐通锵. 语言学纲要[M]. 北京: 北京大学出版社, 2004.  
YE Feisheng, XU Tongqiang. An outline of linguistics[M]. Beijing: Peking University Press, 2004.
- [6] 陆致极. 试论普通话音位的区别特征[J]. 语文研究, 1987, (4): 10-21.  
LU Zhiji. The distinctive features of Mandarin phoneme[J]. Linguistic Research, 1987, (4): 10-21.
- [7] 张家驊. 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 for Standard Chinese (Putonghua)[J]. Acta Acustica, 2005, 30(6): 506-514.
- [8] 张家驊. 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 feature trees of Standard Chinese(Putonghua)[J]. Acta Acustica, 2006, 31(3): 193-198.
- [9] 王洪君. 汉语非线性音系学[M]. 北京: 北京大学出版社, 2008.  
WANG Hongjun. Non-Linear Phonology of Chinese[M]. Beijing: Peking University Press, 2008.
- [10] Hermansky H. Perceptual linear predictive (PLP) analysis of speech[J]. J. Acoust. Soc. Am., 1990, 87(4): 1738-1572.
- [11] Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition[C]. IEEE Trans. Acoust. Speech Signal Process, 1980, 28(4): 357-366.
- [12] 蔡莲红, 黄德智, 蔡锐. 现代语音技术基础与应用[M]. 北京: 清华大学出版社, 2003.  
CAI Lianhong, HUANG Dezhi, CAI Rui. The basis and application of modern speech technology[M]. Beijing: Tsinghua University Press, 2003.
- [13] Theodoridis S, Koutroumbas K. Pattern Recognition[M]. Academic Press, Third Edition, 2006.
- [14] Clarkson P, Moreno P J. On the use of support vector machines for phonetic classification[C]. 1999. ICASSP '99. Proceedings, 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999, 2: 585-588.