

文章编号:1001-9081(2010)04-1135-06

## 普通话声母的客观评测

汤霖<sup>1</sup>, 黄建中<sup>1</sup>, 尹俊勋<sup>2</sup>

(1. 江门职业技术学院 现代教育技术中心, 广东 江门 529090; 2. 华南理工大学 电子与信息学院, 广州 510641)

(china-tl@163.com)

**摘要:**通过罗列分析声母读音错误的主要表现形式,提出了利用语音知识引导的两层两级声母客观评测算法。根据普通话声母的特点,总结出98种普通话声韵组合作为声母的评测基元。实验结果表明,所提出的算法比单独用隐马尔可夫模型(HMM)算法评测的主客观符合率高2.56%,比单独用BP神经网络算法评测的主客观符合率高3.65%,比只用单层算法评测的主客观符合率高1.42%,证明该算法不但能减少计算量,还能提高评测的精度。

**关键词:**语音客观评测;声母;语音信号处理;普通话水平测试;计算机辅助语言学习

**中图分类号:** TP391.6    **文献标志码:**A

## Objective evaluation of Mandarin initials

TANG Lin<sup>1</sup>, HUANG Jian-zhong<sup>1</sup>, YIN Jun-xun<sup>2</sup>

(1. Modern Educational Technology Centre, Jiangmen Polytechnic, Jiangmen Guangdong 529090, China;

2. School of Electronic and Information Engineering, South China University of Technology, Guangzhou Guangdong 510641, China)

**Abstract:** After searching and analyzing the error forms of spoken initials, a two-time-two-level objective evaluation algorithm of initials was advanced on the basis of the phonetic knowledge. Ninety-eight combinations of initial and final were summarized as the basic elements of the initial objective evaluation. The experiment has manifested that the accuracy of the two-time-two-level algorithm is 2.56% higher than that of the Hidden Markov Model (HMM) algorithm, and 3.65% higher than that of BP neural network algorithm, and also 1.42% higher than that of single-time-two-level algorithm. The results prove that the calculation amount of decreases, and the accuracy increases.

**Key words:** speech objective evaluation; initial; speech signal processing; Mandarin proficiency test; computer assisted language learning

### 0 引言

普通话是我国法定的通用语言,是各民族交流的公共语言。我国是一个方言众多的国度,学习普通话是解决沟通困难的唯一途径。同时,随着我国的改革开放,同世界各国的交往与日俱增,越来越多的外国人也学习普通话。由于各种需要,许多人参加普通话水平测试,但目前的普通话水平测试一年只进行两次,而且采用人工评测的方式。这种方式存在测试次数满足不了需求、评测方法主观性强、人为因素多和测试员劳动强度大等弊端。因此,开展计算机辅助普通话水平测试的研究具有十分重要的现实意义和广泛的应用前景。

声母的客观评测是普通话水平客观测试系统中的一个重要组成部分。事实上,声母发音的准确程度是衡量普通话水平高低的指标之一,如从平翘舌声母的发音准确程度就可以略窥其普通话水平。

语音客观评测有两个分支:一是通信系统中基于输入输出的语音可懂度与清晰度的客观评测,另一种是基于计算机辅助语言学习(Computer Assisted Language Learning, CALL)系统的语音客观评测。普通话水平客观测试属于后一种。

早在1990年,文献[1]作者就已经描述了基于计算机学习英语的言语交互语言学习系统的研制工作,该系统是为那些英语作为第二语言的人学习英语所开发的。1996年,美国斯坦福研究中心(SRI)语音技术研究组开发出了VILTS系

统<sup>[2-3]</sup>,该系统利用基于后验概率的评价策略对语言学习中的发音人进行总体发音水平评价。英国剑桥大学人工智能实验室语音组和麻省理工学院人工智能实验室也联合研制出了SCILL语言学习系统<sup>[4-5]</sup>,它主要侧重于语言学习中的发音错误检测和小尺度的发音质量评测。荷兰奈梅亨大学研制的VICK系统<sup>[6]</sup>,主要侧重于研究人工打分的合理性,以及人工打分受韵律、流畅度和音段质量等的影响程度。

文献[7-8]作者针对外国人学习普通话的特点,探讨和研究了计算机辅助汉语教学系统中语音评价体系的组成与实现方法。文献[9]作者也提出了应用语音识别技术进行计算机辅助语言教学的一些设想。文献[10]作者进行了普通话水平测试客观评价研究。文献[11]作者利用语言学专家知识,引入语料选择的自适应算法改进了传统的语音评测算法。文献[12]作者在隐马尔可夫模型的对数后验概率算法的基础上,引入普通话发音的语言学知识,降低了运算量,取得了良好效果。

以上这些研究基本上是以音节或词组为单位进行评测,无法针对音素提供具体的评价得分,应用有一定的局限性。本文所开展的以汉语音素为单位的客观评测,能针对音素提供具体的评价得分,为普通话学习者提供精确的指导。

### 1 普通话声母特点

普通话声母有22个,按发音方法可分为不送气塞音3

收稿日期:2009-08-18;修回日期:2009-11-07。

作者简介:汤霖(1962-),男,湖南醴陵人,副教授,博士,主要研究方向:语音处理、计算机网络安全; 黄建中(1965-),男,广东新会人,实验师,主要研究方向:网络管理、语音处理; 尹俊勋(1942-),男,广东东莞人,教授,博士生导师,主要研究方向:通信、音视频信号处理。

个:b,d,g,送气塞音3个:p,t,k,鼻音2个:m,n,擦音6个:f,h,x,sh,s,r,边音1个:l,送气塞擦音3个:q,ch,c,不送气塞擦音3个:j,zh,z。

普通话声母还有一个零声母,零声母也是一种声母。实验语音学证明,零声母往往也有特定的、具有某些辅音特性的起始方式。普通话零声母可以分为两类,一类是开口呼零声母,一类是非开口呼零声母。

非开口呼零声母,即除开口呼以外的齐齿呼、合口呼、撮口呼三种零声母的起始方式。

齐齿呼零声母音节汉语拼音用隔音字母y开头,由于起始部分没有辅音字母,实际发音带有轻微摩擦,是半元音[j],半元音仍属辅音类。合口呼零声母音节汉语拼音用隔音字母w开头,实际发音带有轻微摩擦,是半元音[w]或齿唇通音[u]。撮口呼零声母音节汉语拼音用隔音字母y(yu)开头,实际发音带有轻微的摩擦,是半元音[q]。

开口呼零声母汉语拼音字母不表示。不经过专门的语音训练,人们一般感觉不到以a,o,e开头的音节还有微弱的辅

音(喉塞音[?]或舌面后擦音[y])存在,因为这些音节开头的辅音成分没有辨义作用,可以忽略不计。

## 2 声母评测方法

### 2.1 声母的分类

声母存在着变体和音征互载现象。变体是指一个声母和不同的韵母结合时,这个声母本身的音段特性(即频谱特性)会有所不同,如擦音h对每个后接元音都有一个变体。音征互载是指一个语音单元的信息由相邻的其他单元携带的现象,如不送气塞音,其区别特征几乎全部由过渡段携带。

因此,在声母评测单元的选择上,采用声母与韵母相结合的组合。从对韵母的声学分析看,韵母与声母的结合音素有12种,即[A][a][ə][ɛ][e][i][ɪ][ʊ][ɔ][u][y],为减少模型总数,将其中的[A][a][ə]合并为[a],[y][ə]合并为[ə],这样处理后,剩下9类,与声母的组合为98个,见表1。 $\varnothing$ 代表零声母。

表1 声韵结合表

序号	声母	结合点的韵母音素类	个数	序号	声母	结合点的韵母音素类	个数
1	b	[a][ə][ɛ][e][i][ɪ][ʊ][ɔ][u]	6	15	zh	[a][ə][ɪ][ʊ]	4
2	p	[a][ə][ɛ][e][i][ɪ][ʊ][ɔ][u]	6	16	ch	[a][ə][ɪ][ʊ]	4
3	m	[a][ə][ɛ][e][i][ɪ][ʊ][ɔ][u]	6	17	sh	[a][ə][e][ɪ][ʊ]	5
4	f	[a][ə][ɛ][e][o][u]	5	18	r	[a][ə][ɪ][ʊ]	4
5	d	[a][ə][ɛ][e][i][u]	5	19	g	[a][ə][e][u]	4
6	t	[a][ə][ɛ][i][u]	4	20	k	[a][ə][e][u]	4
7	n	[a][ə][ɛ][e][i][u][y]	6	21	h	[a][ə][e][u]	4
8	l	[a][ə][ɛ][e][i][u][y]	6	22	$\varnothing$	[a]	1
9	j	[i][y]	2	23	$\varnothing$	[ə]	1
10	q	[i][y]	2	24	$\varnothing$	[o]	1
11	x	[i][y]	2	25	$\varnothing$	[i]	1
12	z	[a][ə][e][ɪ][ʊ]	5	26	$\varnothing$	[u]	1
13	c	[a][ə][ɪ][u]	4	27	$\varnothing$	[y]	1
14	s	[a][?][ɪ][u]	4				
合计			63				35

### 2.2 声母客观评测策略

#### 2.2.1 普通话水平测试中声母读音评测准则

在普通话水平测试中,对声母的测试分散在各测试项中。声母读音误差有两类:一类是语音错误,另一类是语音缺陷。语音错误就是将一个声母读成另一个声母,如将舌尖前音读成舌尖后音。语音缺陷是指声母的发音部位不够准确,但还不是把普通话里的某一类声母读成另一类声母,比如舌面前音j,q,x,读得接近z,c,s;或者把普通话里的某一类声母的正确发音部位用比较接近的部位代替,比如把舌面前音读得接近舌叶音;或者读翘舌音声母时舌尖接触或接近上腭的位置过于靠后或靠前等。

#### 2.2.2 声母读音客观评测策略

声母读音客观评测就是根据已有文本通过各种算法自动地判断发音是正确的、错误的、还是存在语音缺陷的。

设提取到的语音特征为 $X,X_t$ 为 $X$ 经切分后的第 $t$ 个音素的语音特征,语音对应的字符串为 $W_s$ ,对应的第 $t$ 个音素的音素标准模型为 $q_{st}$ ,则该语音音素的得分 $G(X_t)$ 由式(1)给出:

$$G(X_t) = \log \left\{ \frac{P(q_{st} | X_t)}{\max_{q \neq q_{st}} [P(q | X_t)]} \right\} \quad (1)$$

其中 $q$ 为标准音素模型集 $Q$ 中的任一个。利用贝叶斯准则,

$G(X_t)$ 可以表示为:

$$\begin{aligned} G(X_t) &= \log \left\{ \frac{P(q_{st} | X_t)}{\max_{q \neq q_{st}} [P(q | X_t)]} \right\} = \\ &\log \left[ \frac{P(X_t | q_{st}) P(q_{st})}{\sum_{f \in Q} P(X_t | f) P(f)} \right] - \\ &\log \left\{ \max_{q \neq q_{st}} \left[ \frac{P(X_t | q) P(q)}{\sum_{f \in Q} P(X_t | f) P(f)} \right] \right\} = \\ &\log [P(X_t | q_{st}) P(q_{st})] - \\ &\log [\max_{q \neq q_{st}} (P(X_t | q) P(q))] \end{aligned} \quad (2)$$

设备音素的概率相等,则式(2)可再简化为:

$$\begin{aligned} G(X_t) &= \log [P(X_t | q_{st})] - \log \left[ \max_{q \neq q_{st}} P(X_t | q) \right] = \\ &\log \left[ \frac{P(X_t | q_{st})}{\max_{q \neq q_{st}} P(X_t | q)} \right] \end{aligned} \quad (3)$$

事实上,从 $P(q_{st} | X_t)$ 出发,也可以得到相似的结论。利用贝叶斯准则:

$$P(q_{st} | X_t) = \frac{P(X_t | q_{st}) P(q_{st})}{P(X)} =$$

$$\begin{aligned} \frac{P(X_t | q_{st}) P(q_{st})}{\sum_{i=1}^M P(X_t | q_i) P(q_i)} &\approx \\ \frac{P(X_t | q_{st}) P(q_{st})}{\sum_{q_i \in U} P(X_t | q_i) P(q_i)} &\approx \\ \frac{P(X_t | q_{st}) P(q_{st})}{\max_{q_i \in U} (P(X_t | q_i) P(q_i))} & \quad (4) \end{aligned}$$

其中,  $M$  为音素模型集合的模型总数,  $U$  为  $X_t$  所对应的易读错音素的模型集合。

这样,该分值还不能直接对应于普通话水平测试中的 3 个评价:正确、错误、缺陷,通常还要通过一个非线性映射函数来实现。

声母客观评测就是根据测试语音与该语音对应的标准语音的  $P(q_s | X)$  值同测试语音与其他语音的  $P(q | X)$  值的差别来确定对语音质量的评价。从式(3)可知,通过计算  $G(X)$  可得语音的客观评测得分,也可以由式(4)通过计算  $P(q_s | X)$  来计算语音的客观评测得分。实际上,这两种计算方法扩展后对应两种评测算法。

第一种方法就是把声母同所有声母标准模板比较,如果其隶属度最高的模板就是该声母对应的模板,而且该隶属度与紧随其后的对应其他声母的最高隶属度之比大于设定的域值,则认为该声母发音正确;如果其隶属度最高的模板不是该声母对应的模板,而且该隶属度与紧随其后的与该声母对应的最高隶属度之比大于设定的域值,则认为该声母发音错误,

否则认为该声母发音存在语音缺陷。该方法无需语音先验知识,实现起来比较容易。

第二种方法是先收集所有可能的声母发音错误和发音缺陷的情况,形成各声母发音错误和发音缺陷的对照表。把声母同该声母所对应的标准语音模板进行比较,并且与该声母所对应的发音错误和发音缺陷的对照表中的所有模板比较,通过式(4)求出该声母对应的评分,进行非线性映射得到该声母的等级评定。该方法具有非常强的针对性,计算次数少,但是由于中国方言繁多,要收集全声母发音错误和发音缺陷的样板语音十分困难,而且,实际的发音例外情况也不少,因此,该方法难以完全实现。

本系统采用两种办法相结合的方法,也就是语音知识指导下的全模板比较法,具体方法如下:

- 1) 形成声母常见错误读音对照表(包括多音字表)。
- 2) 如果该声母在声母错误读音表中为空,直接进行下一步,否则按第二种方法将该声母同该声母对应的标准模板和错误模板比较,通过非线性映射得到该声母的等级评定,如果其隶属度大于设定的域值,就认为该声母的等级评定可以接受,结束本声母的评测,否则进行下一步。
- 3) 按第一种方法将该声母与所有声母标准模板比较,通过非线性映射得到该声母的等级评定,结束本声母的评测。

## 2.3 声母的客观评测算法

### 2.3.1 声母错误读音对照表

表 2 声母常见错误读音对照表(含多音字读错)

序号	声母	误读成的声母	序号	声母	误读成的声母
1	b	p,m	14	x	h,sh,q,s,j,ch,φ[u],φ[y]
2	p	b	15	z	zh,j,c
3	m	b	16	c	ch,z,s
4	f	h	17	s	sh,c,x
5	d	t,ch,sh	18	zh	z,n,sh,ch,φ[i]
6	t	d	19	ch	c,sh,zh,j,x,d
7	n	l,zh,φ[a]	20	sh	s,x,zh,d,ch
8	l	n,r	21	r	l,φ[i]
9	g	j,h	22	φ[a]	n,φ[ɔ],φ[i]
10	k	h,q	23	φ[ə]	φ[a]
11	h	x,k,f,g	24	φ[i]	φ[a],zh
12	j	q,x,g,ch,z	25	φ[u]	x,φ[y]
13	q	x,j,k	26	φ[y]	x,φ[u]

### 2.3.2 语音评测参数

在 21 个声母中,除 m,n,l,r 4 个为浊声母外,全部为清声母,在语音波形的表现中,呈现出明显的噪声特性,频域参数对这些清声母的区分能力有限,稳定性也比较差。因此,在本系统中,采用时域特征与频域特征相结合的办法。爆破音的音长还特别短,它们的主要特征依附于随后的韵母。因此,在特征上,还加上了随后 64 ms 韵母段特征。

时域特征采用四个自适应门限过零率、声母音长、一阶差分归一化能量和二阶差分归一化能量,其他参数采用 12 阶 MFCC、12 阶一阶差分 MFCC 和 12 阶二阶差分 MFCC。这样,除声母音长外,第  $k$  帧的语音参数  $T(k, n)$  有 42 维。

### 2.3.3 声母客观评测算法

隐马尔可夫模型(Hidden Markov Models, HMM)<sup>[13-14]</sup> 是一种用参数表示的,用于描述随机过程统计特性的概率模型,

它是由马尔可夫链演变而来,并广泛应用于语音处理的各个领域。目前,HMM 模型成为语音处理的基本模型。

神经网络中的多层感知器(Multi-Layer Perceptron, MLP)模型同样也是语音处理的常用模型之一,特别是基于 BP(Back-Propagation)算法的 MLP 更是被广泛应用<sup>[15]</sup>。

本文将 HMM 模型与 BP 模型作为声母客观评测的基本模型。

用标准语音库训练 98 个 HMM 模型分别代表 98 个声韵结合体,接着采用语音知识引导的两层两级模型进行声母客观评测。

第一层采用声母客观评测策略中的第二种方法。设各声母的先验概率  $P(q)$  相同,用 HMM 模型得到  $P(X | \lambda_s)$  和  $\max_{\lambda \in U} (P(X | \lambda))$ ,其中  $\lambda_s$  代表同  $X$  对应的正确的声母 HMM 模型,  $U$  代表该正确声母的常见错误表中的声母集。将得到的

两个值连同其比值输入1个只有3个输入节点和3个输出节点的BP网络,其中输出节点代表“正确”、“错误”和“不确定”,如果得到“正确”和“错误”的结果,并且其输出值大于设定的域值,则接受评测结果,结束该声母的评测;否则,继续下一层的评测。

第二层采用声母客观评测策略中的第一种方法。计算98个HMM模型的 $P(X|\lambda)$ 值,与评价策略中不同的是,算法中将这98模型的输出连同声母音长共99个参数作为第二级BP模型的输入,第二级的输出层只有2个,分别对应“正确”和“错误”,其中的BP模型也有98个。如果输出值大于设定的域值,则接受评测结果,否则认定该声母为“缺陷”。

HMM模型采用连续型,状态数统一为4个,为从左之右

无跳转型,高斯混合数取6个。

第一层评测中BP模型的隐层取为6个节点,第二层评测中BP模型的隐层取为20个节点。

### 3 实验结果

#### 3.1 实验数据

取语音数据库中的标准普通话语音库中的所有单音节字词语和多音节字词语的语音数据,193 140个音节作为训练集1。测试语音数据库中的60人的单音节字词和多音节字词,12 000个音节作为训练集2。测试语音数据库中的另外的60个测试者的单音节字词和多音节字词项的语音数据,12 000个音节作为系统测试集。

表3 实验数据统计

声韵结合体	训练数据	测试数据	声韵结合体	训练数据	测试数据	声韵结合体	训练数据	测试数据
b[a]	3 192	132	n[e]	270	18	zh[u]	4 158	306
b[e]	924	24	n[ə]	432	18	ch[a]	2 442	48
b[ə]	720	18	n[i]	1 776	228	ch[ə]	2 418	78
b[i]	2 892	192	n[u]	858	84	ch[ʌ]	774	72
b[o]	924	60	n[y]	390	66	ch[u]	3 246	186
b[u]	1 548	36	l[a]	1 914	78	sh[a]	1 860	150
p[a]	1 506	66	l[e]	582	6	sh[e]	18	0
p[e]	522	18	l[ə]	816	42	sh[ə]	4 290	168
p[ə]	426	30	l[i]	4 938	384	sh[ʌ]	2 058	168
p[i]	2 334	120	l[u]	2 406	48	sh[u]	2 886	258
p[o]	714	66	l[y]	906	96	r[a]	1 062	36
p[u]	474	6	j[i]	10 440	522	r[ə]	2 148	150
m[a]	1 962	108	j[y]	2 754	144	r[ʌ]	432	36
m[e]	786	30	q[i]	5 418	342	r[u]	1 536	60
m[ə]	882	36	q[y]	2 394	216	g[a]	2 172	102
m[i]	2 634	186	x[i]	8 976	552	g[e]	18	0
m[o]	1 086	60	x[y]	2 886	186	g[ə]	1 734	96
m[u]	816	6	z[a]	1 452	102	g[u]	6 276	336
f[a]	3 330	108	z[e]	60	6	k[a]	1 320	78
f[e]	1 050	60	z[ə]	948	102	k[e]	18	0
f[ə]	1 704	48	z[i]	1 224	126	k[ə]	1 854	144
f[o]	90	0	z[u]	2 052	144	k[u]	2 940	222
f[u]	1 110	12	c[a]	2 046	48	h[a]	1 902	84
d[a]	4 230	216	c[ə]	852	24	h[e]	138	12
d[e]	42	6	c[i]	672	60	h[ə]	2 328	78
d[ə]	1 554	60	c[u]	1 638	54	h[u]	6 084	342
d[i]	2 628	198	s[a]	912	12	φ[a]	1 782	252
d[u]	3 396	120	s[ə]	696	12	φ[ə]	1 890	252
t[a]	2 064	120	s[i]	918	54	φ[o]	72	0
t[ə]	1 218	66	s[u]	2 340	180	φ[i]	9 828	684
t[i]	2 442	156	zh[a]	1 566	72	φ[u]	7 458	510
t[u]	2 994	186	zh[ə]	2 016	108	φ[y]	5 214	246

#### 3.2 实验结果

从表3可以看到,f[o]、sh[e]、g[e]、k[e]和φ[i]这5种声韵组合的测试数据为零,还有5种组合只有6个数据。另外,在普通话水平测试中,第1、2项测试项中语音出现声母“错误”和“缺陷”的平均次数为8次左右,出现“错误”和“缺陷”的声韵组合还比较集中,大部分声韵组合基本不出现“错误”和“缺陷”。因此,在训练时,无法只用标记好的训练数据来确定“错误”和“缺陷”的判决域值。此外,即使有比较充实的训练数据,也因标记为“错误”和“缺陷”的数据比标记为“正确”的

数据少得多,因此,在训练各个声韵综合的声学模型时,还要取与“正确”的数据等量的,来自其他声母训练数据中与该声母的后接韵母相同的数据作为“错误”训练数据,同时,只确定“正确”与“错误”的评测确定域值,如果在判决中既不满足“正确”,也不满足“错误”的条件时,则评定为“缺陷”。

##### 3.2.1 实验1

对采用HMM模型、BP神经网络模型和使用语音知识引导的两层两级模型进行对照实验。

采用HMM模型时,先用训练集1训练,得到98个HMM

模型,用训练集1和训练集2确定98个正确评测域值 $\alpha_i$ 和98个错误评测域值 $\beta_i$ ,客观评测规则采用2.2.2节“声母读音客观评价策略”中的第一种方法,具体步骤如下。

- 1) 计算 $P(q_i | X); i = 1, \dots, 98$ 。
- 2) 如果 $j = \operatorname{argmax}(P(q_i | X))$ 就是待评测声母对应的正确声母的标号,而且 $P(q_j | X)$ 的值与其他声母 $P(q_i | X)$ 的最大值的比值大于该声母对应的正确声母的评测域值 $\alpha_j$ ,就评价该声母发音“正确”,退出比较,否则进行下一步。
- 3) 如果 $j$ 不是待评测声母对应的正确声母的标号,而且 $P(q_j | X)$ 的值与正确声母 $P(q_i | X)$ 的最大值的比值大于该声母对应的错误声母的评测域值 $\beta_j$ ,就评价该声母发音为“错误”,退出比较,否则进行下一步。
- 4) 该声母评价为“缺陷”。
- 5) 算法结束。

采用BP神经网络模型时,用训练集1和训练集2中的标注为“正确”和“错误”的语音数据训练出分别与98个声韵组合对应的98个BP神经网络模型,其中的输入层节点数由训练集1中各声母集的平均帧数乘42确定,输出只有2个节

点,即“正确”和“错误”;同时,确定评测确信域值,如果在判决中既不满足“正确”的确信域值,也不满足“错误”的确信域值,则评定为“缺陷”。其中,隐层统一采用20个节点。

语音知识引导的两层两级模型用训练集1训练HMM模型、训练集1和训练集2训练BP模型来表示。3种模型得到的比较结果见表4。

从表4可以看到,语音知识引导的两层两级模型的总体平均主客观符合率分别比单独采用HMM模型与采用BP神经网络模型高2.56%和3.65%,说明该模型能减少模式之间的混淆,因此能提高主客观符合率。

从整体看,3种模型的零声母的评测主客观符合率都是最高的,这说明3种模型对元音的建模效果都很好,另一方面,零声母的第一元音基本上没有朗读“错误”和“缺陷”。所有模型中,声母h的评测主客观符合率最低,这也反映了声母h的声学多变性严重影响了语音评测的准确度。

HMM模型比BP神经网络模型高1.09%,主要原因是BP模型要输入归一化,对于爆破音,因其音长短,语音特征主要由随后的韵母表达,而归一化对特征的表达有较大影响。

表4 HMM模型、BP神经网络模型和使用语音知识引导的两层两级联模型的主客观符合率 %

声母名称	HMM模型	BP神经网络模型	两层两级模型	声母名称	HMM模型	BP神经网络模型	两层两级模型
b	78.14	76.19	79.87	zh	85.28	82.98	87.94
p	77.78	76.80	80.39	ch	86.98	85.68	88.80
m	85.21	84.51	87.32	sh	86.16	86.29	89.25
f	84.21	84.21	85.96	r	82.62	82.98	84.75
d	86.50	82.17	87.17	g	85.96	85.96	88.95
t	84.85	83.14	87.31	k	85.14	83.78	89.86
n	85.74	83.94	88.96	h	72.67	71.12	74.22
l	84.40	85.32	88.38	φ[a]	100.00	100.00	100.00
j	87.84	85.89	89.49	φ[ə]	95.63	94.05	98.02
q	85.30	84.41	89.78	φ[o]	0.00	0.00	0.00
x	87.13	86.31	89.84	φ[i]	95.32	94.74	97.51
z	80.21	79.17	82.92	φ[u]	94.90	94.51	97.06
c	82.80	81.72	85.48	φ[y]	95.53	94.72	97.15
s	85.27	83.72	90.31	总体平均	86.07	84.98	88.63

### 3.2.2 实验2

对采用语音知识引导的两层两级模型与直接采用单层两级模型的两种方法进行对照实验。这两种模型的差别就是单层两级模型没有两层两级模型的第一层,也就是没有根据声母常见错误读音对照表直接对声母进行判别“正确”和“错误”这一层。实验结果如表5所示。

从表5可以看出,基于语音知识引导的两层两级模型比直接采用单层两级模型的语音评测主客观符合率高1.42%,这说明采用语音知识引导的两层两级模型不仅减少了评测系统的计算量,还提高了评测精度。

## 4 结语

本文统计了声母读音错误的主要表现形式。根据声母的特点,总结出98种声韵合作为声母的评测基元。利用声母错误对照表,提出了利用语音知识的两层两级的声母评测算法。实验结果表明,该算法比单独用HMM算法评测的主客观符合率高2.56%,比单独用BP神经网络算法评测的主客观符合率高3.65%,比只用单层算法算法评测的主客观符合

率高1.42%。说明该算法不但能减少计算量,还能提高评测的精度,是一种有效的普通话水平的声母客观评测算法。

表5 单层两级模型与两层两级模型的主客观符合率 %

声母名称	单层两级模型	两层两级模型	声母名称	单层两级模型	两层两级模型
b	78.79	79.87	zh	85.64	87.94
p	77.78	80.39	ch	87.50	88.80
m	86.62	87.32	sh	87.23	89.25
f	84.21	85.96	r	84.04	84.75
d	87.17	87.17	g	86.70	88.95
t	86.74	87.31	k	87.16	89.86
n	87.35	88.96	h	73.06	74.22
l	86.54	88.38	φ[a]	100.00	100.00
j	89.49	89.49	φ[ə]	97.22	98.02
q	86.56	89.78	φ[o]	0.00	0.00
x	88.35	89.84	φ[i]	96.35	97.51
z	81.88	82.92	φ[u]	95.49	97.06
c	84.95	85.48	φ[y]	95.93	97.15
s	87.98	90.31	总体平均	87.21	88.63

## 参考文献:

- [1] PETERSEN M J. An evaluation of voxbox, a computer-based voice-interactive language learning system for teaching English as a second language[ D]. Nairobi, Kenya: United States International University, 1990.
- [2] FRANCO H, NEUMEYER L, KIMAND Y, et al. Automatic pronunciation scoring for language instruction[ C]// Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing. Washington, DC: IEEE Computer Society, 1997: 1471 – 1474.
- [3] NEUMEYER L, FRANCO H, WEINTRAUB M, et al. Automatic text-independent pronunciation scoring of foreign language student speech [ EB/OL]. [2009 – 06 – 20]. <http://www.asel.udel.edu/icslp/cdrom/vol3/670/a670.pdf>.
- [4] WITT S M. Use of speech recognitionin computer-assisted language learning[ D]. Cambridge: University of Cambridge, 1999.
- [5] WITT S M, YOUNG S J. Phone-level pronunciation scoring and assessment for interactive language learning[ J]. Speech Communication, 2000, 30(2): 95 – 108.
- [6] CUCCIAHINI C, STRIK H, BOVES L. Automatic evaluation of dutch pronunciation by using speech recognition technology[ C]// 1997 IEEE Workshop on Automatic Speech Recognition and Understanding. New York: IEEE, 1997: 622 – 629.
- [7] 郭巧, 陆际联. 具有语音评价功能的计算机辅助汉语教学实验系统[ J]. 微型机与应用, 1999(8): 49 – 51.
- [8] 郭巧, 陆际联. 计算机辅助汉语教学系统中语音评价体系初探[ J]. 中文信息学报, 1999, 13(3): 48 – 53.
- [9] 岳东剑, 柴佩琪. 面向汉语的计算机辅助语音学习系统特征的研究[ J]. 小型微型计算机系统, 2001, 22(7): 848 – 850.
- [10] 魏思, 刘庆升, 胡郁, 等. 带方言口音普通话自动水平测试[ EB/OL]. [2009 – 07 – 01]. <http://cpfd.enki.com.cn/Area/CPFD-CONFArticleList-ZGZR200510001.htm>.
- [11] 魏思, 刘庆升, 胡郁, 等. 普通话水平测试电子化系统[ J]. 中文信息学报, 2006, 20(6): 89 – 96.
- [12] 刘庆升, 魏思, 胡郁, 等. 基于语言学知识的发音质量评价算法改进[ J]. 中文信息学报, 2007, 21(4): 92 – 96.
- [13] RABINER L R, LEVINSON S E, SONDHI M M. On the application of vector quantization and hidden Markov models to speaker independent, isolated word recognition [ J]. Bell System Technical Journal, 1983, 62(4): 1075 – 1105.
- [14] RABINER L R, LEVINSON S E, SONDHI M M. On the use hidden Markov models for speaker independent recognition of isolated word recognition from a medium-size vocabulary[ J]. Bell System Technical Journal, 1984, 63(4): 627 – 642.
- [15] 沈清, 汤霖. 模式识别导论[ M]. 长沙: 国防科技大学出版社, 1991.

(上接第 1131 页)

从表 3 可以看出, 训练好的支持向量机对含有训练样本的测试视频识别效果较好, 识别准确率均在 96% 以上, 这说明支持向量机对于小样本、非线性的分类问题分类效果显著; 同时说明对于新的应用环境, 需要加入新的训练样本进行训练, 以保证更高的识别准确率。对于误识别的情况, 经分析可知, 出现误判的原因跟训练样本的选取有关。因为支持向量机识别的准确性和可靠性在很大程度上取决于样本的选取, 所以样本的选取应尽量具有代表性, 应尽可能多地涵盖不同场景不同材料的火灾图片, 并且有火训练样本和无火训练样本的数目要尽量平衡, 以免使得最优分类面偏移。

本文最后还对利用 BP 神经网络、RBF 神经网络和支持向量机识别火灾的 3 种算法做了比较, 3 种算法对视频 1 的识别准确率比较结果如表 4 所示。

表 4 3 种算法识别准确率比较

算法名称	测试样本数目	误判数目	识别准确率/%
BP 网络	56	6	89.20
RBF 网络	64	4	93.70
径向基 SVM	30	1	96.70

从表 4 可以看出, 利用支持向量机识别火灾的准确率高于利用神经网络识别的准确率, 这是因为支持向量机克服了神经网络容易产生过学习和陷入局部最小点的缺点。相比较而言, 支持向量机识别效果最佳。

## 4 结语

火灾探测问题, 实质上是一个对火焰和疑似火焰物体识

别和分类的问题。本文首先对视频分帧处理, 形成图像序列, 并利用颜色信息在 HSI 空间进行图像分割, 提取出疑似火焰区域; 然后对图像序列进行特征提取, 提取出待识别图像的面积变化率、圆形度和尖角数目 3 个特征量; 接着选取合适的核函数和惩罚因子; 最后建立 SVM 模型并利用支持向量机分类器对特征数据进行分类识别, 并与 BP 网络、RBF 网络的识别结果做了比较。从实验结果来看, 此算法的识别准确率较高; 在各个特征量的提取中, 不用通过多次实验人为设定阈值来区分火焰和干扰物体; 在探测环境以及燃烧材料改变的情况下, 通过训练支持向量机将会自动的产生分类超平面, 将各类样本准确分开, 具有自适应性。

## 参考文献:

- [1] CRISTAINI N, SHWAE-TAYLOR J. 支持向量机导论[ M]. 李国正, 王猛, 曾华军, 译. 北京: 电子工业出版社, 2004: 82 – 108.
- [2] 郝祖龙, 刘吉臻, 田亮. 基于支持向量机的电站锅炉燃烧定性判断[ J]. 华北电力大学学报, 2007, 34(4): 51 – 55.
- [3] 安丰凌, 孙劲光, 张新君. 一种火灾报警系统中火焰区域的检测方法[ J]. 现代计算机, 2007(10): 44 – 46.
- [4] 王俊明, 杨永跃, 付贵权. 多判据图像型火灾探测系统的研究[ J]. 工业控制计算机, 2008, 21(2): 50 – 51.
- [5] 唐发明. 基于统计学习的支持向量机算法研究[ D]. 武汉: 华中科技大学, 2005: 16 – 27.
- [6] 肖靓. 基于支持向量机的图像分类研究[ D]. 上海: 同济大学, 2006: 49 – 54.