中山大学学报（自然科学版）
第31卷 第2期 ACTA SCIENTIARUM NATURALIUM Vol.31 №.2
1992年 UNIVERSITATIS SUNYATSENI 1992

发音人语音特征参量的选择和提取

郭 铭 陈云凤
（无线电电子学系）

摘 要 本文研究发音人识别特征参量的选择和提取，探讨参量选择的原则，提出一种混合特征矢量，并以时间域规正法规正语音长短。

关键词 发音人识别，特征参量，选择，混合

说话人识别系统识别率的高低，在很大程度上依赖于语音的特征参量的正确选择和准确提取[1]，一般公认的选择特征参量的原则有：①能有效地描述与发音人有关的信息。②稳定、不随环境影响而显著变化。③易于度量、提取和存储。④不易被模仿。⑤在混合参量的识别中，所选各种特征参量之间的相关性应尽量小。

大部分特征参量，都可以归为声道参量和波形参量两大类[2]，前者是通过建立发音人的时变声道模型，进行线性预测分析而得到的参量，综合反映发音人说话时声道的特征、形状和位置等。例如：LPC系数，LPCC系数，共振峰频率等。后者是对语音波形直接进行不同的数学变换和分析而得到的参量，综合反映说话人习惯、语气、音色等特征。例如：基音周期、短时能量、短时谱等。

我们用基音周期、短时能量、LPC的PARCOR系数、LPCC系数以及它们的动态参数构成混合参数矢量，进行混合参数识别，取得了较好的效果。

1 特征参量

1.1 基音周期

基音周期反映说话人发音时声带的振动频率，它与说话人的声带长短，发音时的情绪以及发音强度有关。基音周期的检测是先对语音信号用自关函数法进行估值，然后进行非线性滤波，最后再用一个32Hz低通、线性相位、FIR数字滤波器进行平滑。

1.2 短时能量

短时能量给出反映语音信号幅度变化的合适描述方法。为简化计算，每一帧语音的短时平均幅度加以归一化，得到短时平均能量这一识别参量。

1.3 线性预测系数（LPC）

线性预测系数综合反映发音人的声道特性。因为每个人的声道具有生理上的差异和发音习惯上的动作差异，因此该参数包含了用以区分不同发音人的特征。

本文1991年6月12日收到

LPC分析采用格型法求PARCOR系数，即K参数。帧长256个采样点（32ms），帧间重叠128个采样点，阶数选8和12两种。

1.4 倒谱系数（LPCC）

倒谱系数从时域上反映了语音总的频谱形状信息。典型的倒谱计算需进行两次DFT，运算量大。我们利用LPC分析的K参数，很方便地推得倒谱系数，方法如下：

用下式递推p次得到p阶LPC预测系数

$$\begin{cases} a_i^{(i)} = K_i \\ a_j^{(i)} = a_j^{(i-1)} - K_i\, a_{i-j}^{(i-1)} & 1 \leqslant j \leqslant i-1 \end{cases}$$

然后把$a_j^{(p)}$（$p=12$或8，$j=1,2,\cdots,12$或$j=1,2,\cdots,8$）代入下式中（简写为a_i），推出倒谱系数C_n（$n=1,2,\cdots,12$或$n=1,2,\cdots,8$）

$$C_1 = a_1$$

$$C_n = \sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)a_i C_{n-i} + a_n \qquad 1 < n \leqslant p$$

1.5 语音的动态特征

语音的动态特征在语音感知中起着十分重要的作用[3]。我们尝试用特征参数曲线的斜率特征反映语音的动态变化。

由于在语音感知中，50～60ms是保持语言动态特征的最佳时间长度[2]，在我们的系统中，一帧为32ms，所以只需求取相邻帧间差值即可。

设p阶LPC或LPCC参数系列是$f(i,j)$，i是帧号，j是阶号，则得到p阶动态参数：

$$D(i,j) = f(i,j) - f(i-1,j) \qquad j = 1,2,\cdots,p$$

这组动态参数可以与其它参量一起构成多维识别矢量。

2 系统实现

2.1 硬软件构成

本文研究的硬件基础是一个以IBM-PC/XT为主机，配以高速数字处理系统ATD-20A及语言输入、输出接口卡所构成的一个语音处理实验系统[4]。

说话人确认系统的信号处理方式见图1。

端点检测在语音处理中是一个很重要的问题，这里采用能量和过零率联合检测的方法[2]。

2.2 时间域规正

声母和韵母相连相拼的部分

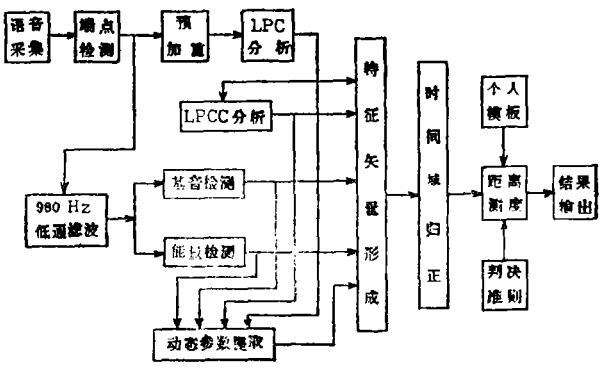

图1 发音人确认系统的信号处理方式

Fig. 1 The speaker verification system and the flow chart for date processing

叫过渡音，虽然它只占发音中的很小一部分时间，但往往带有强烈的个人特征，人们的说话特点、习惯、语气、音调等很多特征都在过渡音中反映出来，而韵母段数据量大，信息量小，可以加以裁减。

参考俞铁城和王迎庆的工作[5,6]，我们提出一种在时间域上归一化语音的方法。这种方法具有两个优点：①归一化了不同长度的语音，使模板匹配得以进行。②压缩了共性较多的稳态部分，相对强调了个性较多的过渡音部分，从而使得参数矢量空间中的类间重叠减少，有利于提高识别率。

设混合特征参数矢量为$K(n,i)$,其中n是帧序号，i是矢量维号。

定义第n帧和第$n-1$帧之间距离为：

$$D(n)=\sum_{i=1}^{m}[K(n,i)-K(n-1,i)]$$

时间域规正法就是不断寻找并压缩帧间距离最小的帧，直到帧数等于测试文本的标准帧数。

2.3 测试文本

为全面试验不同特征参量的优劣，尽量选择用不同发音方法和发音部位的汉字作测试文本，仔细研究汉语发音特点后[7]，我们选3个不同发音方法的声母：塞音g，鼻音m和边音 l，同时选5个不同发音部位的韵母：单韵母a,i,u,鼻韵母 ing,带介音的韵母uan，这些声母和韵母相拼，照顾4个声调，选出5个字作为测试文本集：秘(Mi)，拉(La)，古(Gu)，灵(Ling)，馆(Guan)。

3　实验及分析

以确认错误率EP (Error Probability) 作为对发音人确认系统性能的评价尺度，定义：

$$EP=\sqrt{FA\cdot FR}$$

式中FA是对冒认者的错误接受率，FR是对系统注册者的错误拒绝率，当$FA=FR$时，EP最小，下面实验所称的确认错误率即是此时的EP值。

F比是一种类别间可分性的测度，定义为：

$$F=\frac{\text{均值的方差（全部类别之间）}}{\text{方差的均值（同一类别之内）}}$$

模板匹配采用绝对值距离测度，传输媒介是话筒-录音机-确认系统，注册者模板是由在不同时间的10次训练取平均得到的。

对每个测试文本，每个注册者作100次确认测试，冒认者集合作100次确认测试，由这200次测试结果统计得到该文本的确认错误率曲线。

3.1 LPCC及LPC分析阶数选择和特征参量选择

3.1.1　条件　注册者：男1人，年龄，28岁。冒认者：男，10人，年龄，24～28岁。测试文本：秘(Mi),拉(La),古(Gu),灵(Ling)。

2.1.2　结果　综合4个测试文本的4次实验结果，得出不同分析阶数和特征矢量的确认错误率示于表1。

表1 特征矢量和分析阶数选择

Tab.1 Feature and analysis order selections

特征	LPCC+					LPC+				
	P	P+E	P+PDC	LDC	AVE	P	P+E	P+PDC	LDC	AVE
8阶(%)	4.0	4.7	6.0	9.8	6.1	5.2	9.6	8.7	11.1	8.7
12阶(%)	3.4	3.8	5.2	9.4	5.5	4.7	9.3	9.1	10.3	8.4

注：P为基音周期，E为短时能量，xDC为x的动态参数，AVE为平均

3.1.3 分析 ① LPCC参数显示出比LPC参数更大的优越性，含LPCC参数的混合矢量比含LPC参数的混合矢量，计算量增加约15%，但性能提高约29%（8阶）和31%（12阶）。② 含12阶LPCC的混合矢量，平均$EP=5.45\%$，含8阶LPCC的混合矢量，平均$EP=6.13\%$，可见，12阶分析比3阶分析，错误率下降0.68%，性能提高11%，计算量增加约12%。③ LPCC,LPC，基音周期是较好的识别参量，而它们的动态特征参数和短时能量不好，混合特征矢量并非包含的参量越多越好，某些参量的加入反而会降低识别效果。

3.2 Fisher判据与确认错误率

3.2.1 条件 注册者：男，1人，年龄，28岁，女，1人，年龄22岁。冒认者：男，10人，女，10人，年龄，24～28岁。测试文本：秘(Mi),灵(Ling),馆(Guan)。

3.2.2 结果 我们考虑基音周期，短时能量，12阶LPCC，它们的动态参量以及混合矢量的F值与确认错误率EP，结果分别示于表2和表3。

3.2.3 分析 ①在单一特征参量中，以12阶LPCC参数的F值最高，基音周期第二，

表2 特征参量的F比

Tab.2 The F of the features

特征	P	PDC	E	EDC	L	LDC	P+L	P+E+L	P+L+LDC	P+E+L+LDC	P+E+L+DC
男	2.70	0.30	0.38	0.28	5.39	0.96	8.02	7.62	6.36	7.00	4.90
女	0.80	0.60	0.59	0.15	5.52	0.03	4.70	3.72	3.56	3.58	2.60
平均	1.76	0.45	0.49	0.22	5.46	0.50	6.36	5.67	4.96	5.29	3.75

注：P为基音周期，E为短时能量，L为LPCC系数，xDC为x的动态参数

表3 特征参量的错误率

Tab.3 The EP of the features

特征	P	PDC	E	EDC	L	LDC	P+L	P+E+L	P+L+LDC	P+E+L+LDC	P+E+L+DC
男(%)	13.7	36.0	35.7	37.7	5.80	28.3	2.47	2.10	3.07	2.73	4.80
女(%)	23.0	29.0	34.0	41.7	7.03	53.0	5.87	6.13	7.87	7.60	13.0
平均(%)	13.4	32.5	34.9	39.7	6.42	40.7	4.17	4.12	5.47	5.17	8.90

注：P为基音周期，E为短时能量，L为LPCC系数，xDC为x的动态参数

短时能量和LPCC动态参数次之。② 混合矢量的F值与其中每一种参量的F值有很大关系，大体上看，$F>1$的参量对总F起正作用，而$F<1$的参量对总F起负作用。③ F与EP的关系见图2，图中各点代表各种特征矢量。④ 依据②和③，在混合参量识别中，并非参量越多越好，要避免采用$F<1$的参量。⑤ 实验中发现，男女在声道参数上差别不明显，较稳定，但在波形参数上差别明显，不稳定。所以，在同性冒认者中，LPCC参数比基音周期效果好，而在异性冒认者中，则相反，故两种参数是互补的，有很好的混合效果。

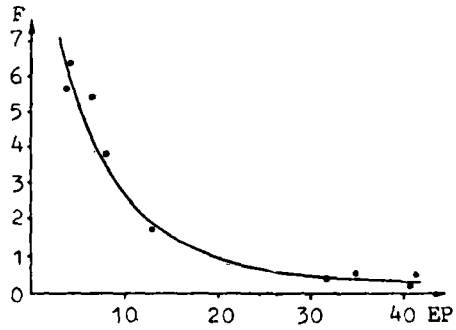

图2　F比与错误率的关系

Fig. 2　The relation between F and EP

4　结论和讨论

(1)　特征参量的研究是要继续进行的重要工作，互补性是构造混合矢量的一个重要条件，特征参量评价的一般准则和选择的一般规律还有待深入探讨。

(2)对语音进行时间对准时，实验证明计算简单的时间域规正法对于短时间语音(如汉语单字)是很有效的。

(3)　距离测度的精心选择对识别效果非常重要，但有两点须注意：① 时间域规正法中采用的帧间距离测度应与模板匹配时的距离测度一致，否则会导致效果下降，不能发挥好的距离测度的优势。② 由于距离测度与特征参量有关，故对采用混合参量的识别，应用不同参量取不同距离测度的方法，以达到最佳效果。

参　考　文　献

1　Naik J M. IEEE Communication, 1990,28(1): 42~48

2　Rabiner L R, Schafter R W(朱雪龙等译). 语音信号数字处理. 北京：科学出版社，1983

3　Furui S. IEEE trans ASSP, 1981,29(3): 342~350

4　陈志成，陈云凤. 中山大学学报(自然科学版)，1991,30(1): 45~51

5　俞铁城. 物理学报，1978, 26(5):508~515

6　王迎庆. 电声技术，1988(2): 73~76

7　张志公. 现代汉语. 北京：人民教育出版社，1982

Feature Selection and Extraction on Speaker Recognization

Guo Ming　Chen Yunfeng*

Abstract　This paper studies the feature selection and extraction on a speaker verification system based on IBM-PC/XT microcomputer and highspeed DSP system ATD320-A. Some standards for the feature selection are discussed. The method of the time region uniform is used to solve the problem of speech dynamic variation.

Keywords　speaker recognition, feature parameters, selection, mix

* Department of Radio and Electronics