

## 四、结 论

从图4—图6以及表1可看出,在傅里叶域插值前进行坐标平移比在插值后进行坐标平移重建图象的质量高得多;傅里叶域覆盖区A和B边界附近点的插值精度对图象重建的质量有很大的影响。此外,从表1中图4的距离评价量与非插值算法重建图象距离评价量的比较可以看出,SADCT的傅里叶域插值算法重建图象的质量较高。另外,从本文图象重建的结果来看,距离评价指标的优劣与对重建图象进

行直观观察或用直线 $y = -0.605$ 上重建值与真实值之间比较进行重建图象质量判别的结果相一致。然而,使用距离评价量却比较简单,且易于进行定量评价。

### 参 考 文 献

- [1] Nahamoo D., Pan S. X. and Kak A. C., *IEEE Trans. on Sonics and Ultrasonics*, 31-4(1984), 218—229.
- [2] Pan S. X. and Kak A. C., *IEEE Trans. on Acoust. Speech Signal Processing*, 31(1983), 1262—1276.
- [3] Lu Jian-yu, The Proceedings of the China-Japan Joint Conference on Ultrasonics, 1987, 395—399.
- [4] Herman G. T., *Image Reconstruction from Projections*, Academic Press, 1980, 157—160.

# 基于声韵分割的汉语单音节识别方法

徐近霏 杨子云 沙宗先

(哈尔滨工业大学)

1987年7月30日收到

本文提出一种以声韵分割为基础的汉语无调单音节识别方法。这种方法由于引入声韵分割信息和声韵分段判别,显著改善了单音节的正识率。用三男一女的发音对全部403个汉语无调音节考核结果表明,第一位平均正识率为83.9%,前三位平均正识率为96.0%。

## 一、引 言

国内已推出的语音识别装置都是以词为单元、采用整体匹配方法进行识别。因而要求对每个词进行训练并建立模板。在用于汉语文本读入时,词汇量要扩大到几万,这种方法将引起内存空间、识别和训练时间等困难。利用音节作为识别单元是解决这些困难的有效途径。因为汉语音节类型有限,结构简单,而且有较稳定的声学特性。汉语音节由声母、韵母和声调三部分构成,根据基频与共振峰基本无关的结论<sup>[1]</sup>,可以分别进行声调识别和无调音节识别,从而把全部识别单元由1280个有调音节进一步压缩到403个无调音节。但是汉语单音节词表具有高度混淆性,有大量彼此相差仅一个辅

应用声学

音、元音或鼻尾音的相似音节对,如ban、dan、bang、ben等。采用整体识别时,这种微小差异将被整体匹配的随机误差所淹没,产生严重误识。

本文提出一种基于声韵分割,并对声母和韵母分别进行分段矢量量化和判别的汉语无调单音节识别方法。试验表明,这种方法可以显著提高单音节的识别精度。

## 二、识别方法

识别系统框图见图1。

### 1. 特征计算和头尾判定

语音经预放和带通滤波(100—4500 Hz)后,用A/D进行10KHz采样和12bit量化;用快速信号处理器TMS320,以帧长12.8ms、帧

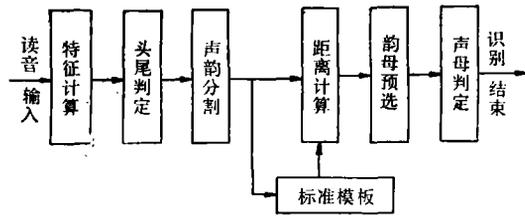


图1 识别系统框图

移 6.4ms, 在采样同时逐帧计算短时能量和短时过零率, 用二次判头判尾法获取有效语音段<sup>[2]</sup>; 用部分自相关法提取基音周期; 用 FFT 法求取加汉明窗语音帧的对数功率谱, 作 18 段 Bark 分割后, 用 DCT 变换求取语音帧的前 5 个 DCT 系数, 和短时能量、短时过零率、基音周期一起作为后期使用的特征。

## 2. 声韵分割

其基本思想是先用基音周期的开始作为分割点的第一级预测, 然后用其余各维特征变化率的综合打分求得分割点的第二级预测, 最后综合两次预测结果得到实际声韵分割点。由此得到的声母段中包含了声韵间的过渡段。方法详见文<sup>[3]</sup>。

## 3. 标准模板的建立

标准模板是将同一说话人的三次发音用上述分割方法作声韵分割后, 对声母段和韵母段分别用矢量量化方法求得的。使用的特征是除基音周期之外的七个特征参数。对声母段, 将几次训练发音的声母段各等分六段, 将每段的平均矢量作为六个初始聚类中心, 然后用 K——均值法迭代求得总方差最小的聚类中心, 将此作为声母模板的六个码字。同时三次发音的平均声母段时长也作为一个特征参数保留。韵母段采用分段码本表达, 将每个韵母等分为三段, 每段用聚类方法得到四个码字, 过程同声母段。

最终, 我们得到一个单音节的标准模板为

$$T = (LT, ST, YT)$$

其中,  $LT$  为平均声母时长(帧数);  $ST = (t_1, t_2, \dots, t_6)$ ; 声母标准模板,  $t_i$  为 7 维特征矢量;  $YT = [(t_{11}, t_{12}, t_{13}, t_{14}), (t_{21}, t_{22}, t_{23}, t_{24}), (t_{31}, t_{32}, t_{33}, t_{34})]$ , 韵母标准模板,  $t_{ij}$  为 7 维特征矢量。

## 4. 距离计算

对声母段和韵母段分别求匹配距离。设被识单音节声韵分割后可表达为

$$S = (LS, SS, YS)$$

其中,  $LS$  为声母时长;  $SS = (s_1, s_2, \dots, s_{LS})$ , 声母段帧特征矢量序列;  $YS = [(s_{21}, \dots, s_{2L}); (s_{31}, \dots, s_{3L})]$ ,

韵母段三个等分子段帧特征矢量序列。其中  $L = \left\lfloor \frac{LG - LS + 1}{3} \right\rfloor$ ;  $LG$  为音节长,  $L$  为子段长。

这样, 对于一个标准模板

$$T_m = (LT_m, ST_m, YT_m)$$

我们三个部分的匹配距离

(1) 声母时长匹配距离

$$DL_m = (LS - LT_m)^2$$

(2) 声母匹配距离

$$DS_m = \sum_{i=1}^{LS} \min_{1 \leq j \leq 6} d^2(s_i, t_j^m)$$

(3) 韵母匹配距离

$$DY_m = \sum_{i=1}^3 \sum_{j=1}^L \min_{1 \leq k \leq 4} d^2(s_{ij}, t_{ik}^m)$$

其中, 两矢量间距离采用欧氏距离

$$d(X, Y) = \left[ \sum_{i=1}^7 (x_i - y_i)^2 \right]^{1/2}$$

由此得到  $M$  个标准模板的所有匹配距离

$$(DL_m, DS_m, DY_m) \quad 1 \leq m \leq M$$

## 5. 韵母预选

用于韵母预选的匹配距离定义为

$$D_m = W_1 \cdot DS_m + (1 - W_1) \cdot DY_m$$

$$1 \leq m \leq M$$

其中, 加权值  $W_1 (0 \leq W_1 \leq 1)$  可以调整  $D_m$  中声、韵段的比重。韵母预选时, 应以韵母段为主, 适当考虑声母段影响, 所以  $W_1$  应取较小值。

然后, 将  $D_m$  从小到大排序

$$D_{m1} \leq D_{m2} \leq D_{m3} \leq D_{m4} \dots$$

发现正确的韵母类常处于前四个候选中, 并占有优势。为此, 统计前四个候选中每类韵母的打分

$$\mu_i = \sum_{j=1}^{n_i} D_{m1} / D_m(i, j)$$

其中,  $D_m(i, j)$ ,  $n_i$  是前四个候选模板中韵母为  $i$  的匹配距离和个数。

我们取打分值最大的韵母类  $I_0$  作为候选韵母

$$I_0 = \arg(\max_i \mu_i)$$

并选择下列模板集作为下级声母判定的候选

(1) 所有韵母为  $i_0$  的模板; (2) 排序为第一位的模板(如其韵母不为  $I_0$ ); (3) 前四个候选中如有  $er$ ,  $ueng$  模板, 因它们只有一个同韵母音节, 也予保留。

实验表明, 采用这种韵母预选方法, 可以提高韵母正识率, 同时减少了出现同一声母后接不同韵母情况, 避免下一级声母判定的困难。

### 6. 声母判定

用于声母判定的匹配距离定义为

$$D_{m_i} = W_2 \cdot DS_{m_i} + (1 - W_2) \cdot DY_{m_i} + W_1 \cdot DL_{m_i} \quad 1 \leq i \leq N$$

其中, 权  $W_2$  ( $0 \leq W_2 \leq 1$ ) 用于考虑韵母对声母判定影响; 权  $W_1$  用于考虑声母时长对声母判定影响;  $N$  则是韵母预选级提供的候选模板数。

将各模板按其匹配距离由小到大排序, 取距离最小的模板  $I_1$

$$I_1 = \arg(\min_i D_{m_i})$$

为识别结果。同时输出前三个候选。它们既是声母也是音节识别结果。

由上述识别算法可见, 韵母、声母匹配距离计算公式相似, 只是权系数不同。因此, 只需计算一次部分距离后, 即可方便地用相同程序实现。

## 三、实验结果

对上述无调单音节识别方法做了如下考核实验

1. 以 200 无调单音节为实验词表 (是全部 403 个无调单音节中由全部声母与  $a$ 、 $o$ 、 $e$  族及部分  $i$  族韵母构成), 由一男性说话人, 发音六次, 三次作训练, 三次作识别。测试不同加权值

应用声学

时的识别结果。

(1)  $W_1$  对韵母正识率影响见表 1。从表 1 可见, 在  $W_1 = 0.1-0.5$  范围内, 韵母正识率无显著变化。在  $W_1 > 0.5$  后, 由于匹配距离转以声母段为主, 使韵母正识率降低。

表 1  $W_1$  对韵母正识率影响

$W_1$	0	0.1	0.2	0.3	0.4	0.5	0.7
$p_{c_1}(\%)$	95.5	97.5	97	97	97.5	97	94.5

(2)  $W_2$  对前三位累计音节正识率影响见表 2。由表 2 可见, 在  $W_2 = 1.0-0.8$  范围内, 音节正识率无显著变化。在  $W_2 < 0.8$  后, 由于匹配距离中韵母段影响过大, 使声母正识率(从而音节正识率)有明显降低。

表 2  $W_2$  对前三位累计音节正识率影响

$W_2$	1.0	0.9	0.8	0.7
$p_{c_1}(\%)$	76.0	74.0	75.5	73.5
$p_{c_2}(\%)$	91.0	90.5	90.0	87.0
$p_{c_3}(\%)$	93.5	93.5	94.0	93.0

(3)  $W_1$  对前三位累计音节正识率影响见表 3 ( $W_2 = 1.0, W_1 = 0.1$  时)。可见, 声母时长参与匹配, 且  $W_1 = 17-35$  时(此时声母时长距离与声母特征距离有相近数量级), 对音节第一位正识率有明显改善。主要是改进了声学特征相近而时长相差明显的声母族(如送气/不送气塞音)间的区分性。

表 3  $W_1$  对前三位累计音节正识率影响

$W_1$	0	3	9	17	23	30	35
$p_{c_1}(\%)$	76.0	76.5	79.0	80.5	80.5	81.5	80.5
$p_{c_2}(\%)$	91.0	91.5	91.5	91.5	91.5	92.5	92.5
$p_{c_3}(\%)$	93.5	94.5	94.5	94.5	94.5	95.0	94.5

2. 作为比较, 以同样 200 无调单音节词表, 同一男性说话人发音, 同样特征参数, 但采用一般整体匹配、动态规划方法进行识别, 其音节第一位正识率只有 62.3%。

3. 进一步用全部 403 无调单音节词表, 由

: 7 :

三男性说话人和一女性说话人,各发音四次,三次作训练,一次作识别,考核结果如下:

(1) 不同说话人前三位累计音节正识率情况见表4 ( $W_1 = 0.3, W_2 = 0.9, W_l = 30$ 时)。

表4 不同说话人前三位累计音节正识率情况

说话人	男1	男2	男3	女1	平均
$p_{c_1}(\%)$	85.85	77.41	87.59	84.86	83.93
$p_{c_2}(\%)$	93.75	91.05	94.53	95.03	93.60
$p_{c_3}(\%)$	95.77	93.28	97.25	97.51	95.95

(2)  $W_1$  对前三位音节平均正识率影响见表5 ( $W_2 = 0.5, W_l = 30$ 时)。由表可见,  $W_1$  在0.1—0.5范围内变化时,对音节正识率无明显影响。这是因为 $W_1$ 只影响韵母正识率,而 $W_2$ 在此范围内韵母正识率无显著变化。

表5  $W_1$ 对前三位音节平均正识率影响情况

$W_1$	0.1	0.2	0.3	0.4	0.5
$p_{c_1}(\%)$	83.55	83.47	83.93	83.61	83.74
$p_{c_2}(\%)$	93.17	93.35	93.60	93.66	93.83
$p_{c_3}(\%)$	95.38	95.64	95.95	96.01	96.25

(3)  $W_2$ 对前三位音节平均正识率影响见表6 ( $W_1 = 0.3, W_l = 30$ 时)。由表可见,音节正识率在 $W_2 = 0.9$ 时最高。这是由于在声母匹配距离中引入少量韵母段距离可加强声韵过渡段影响,对提高第一位正识率有利。但韵母

段影响过大( $W_2 < 0.8$ )则将使声母(从而音节)正识率明显降低。

表6  $W_2$ 对前三位音节平均正识率影响情况

$W_2$	1.0	0.9	0.8	0.7	0.6
$p_{c_1}(\%)$	81.20	83.93	83.68	82.44	80.21
$p_{c_2}(\%)$	93.04	93.60	92.73	91.80	90.62
$p_{c_3}(\%)$	96.08	95.95	95.23	94.28	94.03

(4) 为分析声母、韵母混淆情况,将上述4个说话人共 $4 \times 403 = 1612$ 个发音在 $W_1 = 0.3, W_2 = 0.9, W_l = 30$ 条件下的声母和韵母第一位误识情况统计结果分别列于表7和表8中。从表中可见(a)声母误识中,误识严重的声母及其误识率为: $p(36.8\%), t(32.9\%), b(31.3\%), d(18.2\%), m(23.6\%), s(20.3\%), q(19.6\%)$ 。(b)声母混淆主要发生在同一声母族内,不同声母族其族内误识所占比例为“一, $b, d, g$ ”(66.7%),“ $p, t, k$ ”(71.7%),“ $m, n, l, r$ ”(89.5%),“ $zh, ch, sh$ ”(57%),“ $z, c, s$ ”(74.1%),“ $j, q, x$ ”(96.2%)。对全部声母而言,发生在族内的误识占71.4%。(c)韵母误识类型及占其误识总数的比例为:韵头 $i, u, u$ 的插入、丢失或代换占43.8%;韵尾 $n, ng, i$ 的插入、丢失或代换占18.8%;韵腹 $a, o, e$ 的代换占10%;全部错占27.4%。可见,韵头和韵尾的错误占较大比例,这与声韵分割和尾部判定不准确有关。

表7 声母误识统计

误识数	误识对	误识数	误识对		
—	11/136	b(1), d(1), g(3), k(4), l(2)	f	4/36	b(1), d(1), t(1), z(1)
b	20/64	—(4), d(11), g(1), p(1), t(1), k(1), l(1)	h	8/76	—(1), g(3), t(2), k(2)
d	16/88	—(2), b(10), g(3), j(1)	z	6/68	b(2), d(2), c(1), s(1)
g	7/76	—(1), d(4), k(2)	c	8/64	p(1), t(1), z(4), s(1), sh(1)
p	25/68	t(16), k(3), q(1), h(5)	s	13/64	z(6), c(7)
t	25/76	p(12), k(5), C(2), ch(6)	zh	11/80	g(1), k(1), f(1), z(1), ch(5), sh(2)
k	3/76	b(1), t(2)	ch	4/72	f(1), c(1), zh(1), sh(1)
m	17/72	n(17)	sh	6/76	z(1), s(1), zh(1), ch(2), q(1)
n	7/96	m(6), l(1)	j	7/56	p(1), q(2), x(4)
l	9/100	—(1), d(1), m(1), n(3), r(3)	q	11/56	j(2), x(9)
r	5/56	—(1), q(1), l(3)	x	8/56	j(2), q(6)

表 8 韵母误识统计

	误识数	误识对		误识数	误识对		误识数	误识对
a	3/72	ang(1), ia(1), ua(1)	ong	2/52	eng(1), uen(1)	ua	4/24	a(4)
ai	0/68		i	2/72	eng(2)	uai	3/28	ai(2), uan(1)
ao	3/72	ong(1), iao(2)	ia	0/24		uan	3/60	an(1), ang(1), uai(1)
an	2/76	ang(1), en(1)	ian	0/44		uang	4/28	ang(4)
ang	6/76	a(4), eng(2)	iang	0/24		uei	0/52	
e	2/60	iong(1), uo(1)	iao	0/44		uen	0/56	
ei	1/56	uei(1)	ie	0/44		ueng	3/4	eng(1), ong(1), uen(1)
eri	6/64	an(4), uen(1), üan(1)	in	6/36	ing(6)	uo	10/60	e(3), o(7)
eng	5/72	an(4), ang(1)	ing	1/44	ian(1)	ü	0/24	
er	0/4		iou	0/32		üe	3/24	ie(3)
o	5/20	uo(5)	iong	0/16		üan	0/16	
ou	4/72	ao(4)	u	0/76		ün	2/16	in(2)

## 四、结 论

1. 对汉语单音节识别而言,利用声、韵分割和声、韵分别匹配方法较音节整体匹配方法更为有效。

2. 声母时长对于单音节中声母识别是一个有效特征。它可以减少声母时长相差较大声母族之间的误识(如 p, t, k/b, d, g),使混淆主要限

于同一声母族内。

3. 为继续提高正识率,需要进一步探索声母族内的区别特征和提高声韵分割和头尾判定的准确性。

## 参 考 文 献

- [1] 张家骥,齐士铃,声学学报, 4-1(1979),23—29.
- [2] 徐近需,杨枫,谭志强,计算机学报, 8-3(1985), 213—222.
- [3] 杨子云,徐近需,沙宗先,模式识别与人工智能, 1-1 (1988).

# 计算机在声强测量中的应用

任明章 唐撷茹 顾崇銜

(西安交通大学机械制造教研室)

1987年12月1日收到

声强测量作为新的测试手段在噪声控制领域中应用愈来愈广。为了提高实验效率,扩大应用范围,本文实现了声强测量的计算机控制。编制了测算声功率及测绘声强三维分布图的应用软件。在此基础上,通过实验说明了声强测量在测量声功率及鉴别噪声源方面的应用及其优点。

## 一、前 言

声强测量是噪声控制领域内出现的一门新技术。由于声强反映了声能传播的大小和方向用它来识别噪声源、测量声功率及研究结构

的传声损失等比用声压具有许多优点。例如,声强可在近场进行测量<sup>[1]</sup>,因此可获得振动物体表面附近的声强分布图,直观地再现机器各部分辐射声能的大小,为采取最有效的降噪措施提供依据;利用声强法测量声功率,可以有效地抵抗稳态背景声的干扰,实现在位 (*in-situ*) 测