

低频噪声方面,比传统的消声器效果好得多.计算机模拟结果表明,一个矩形截面(40 cm × 40cm)管道,当只计及平面波时,从20Hz到450Hz噪声功率谱分量下降20—35dB.传统消声器在这一频段一般只能降噪10dB左右.

为了使LMS算法在该系统正确收敛,必须采用一辅助滤波器对输入信号加以处理.声反馈效应可通过一补偿滤波器来消除.

滤波器的阶数选128阶左右适宜.太低或太高,总降噪量均有所下降.与管壁有吸收的

情形相比,刚性管壁自适应抵消系统的抵消效果更好.气流速度或温度上升,总降噪量相应有所下降,在较大变化范围内下降量不大于2dB.

参 考 文 献

- [1] 李毅民,声学学报,13(1988),336—342.
- [2] 李毅民,声学学报,14(1989),370—376.
- [3] Bernard Widrow et al.,*Proceedings of the IEEE*, 63-12(1975), 1692—1716.

说话人辨认中有效参数的研究

王炳锡 张白愚

(中国人民解放军信息工程学院)

1991年1月23日收到

说话人辨认技术是语音识别中的一项重要应用.在我们研究的系统中,LPC参数并不是都很有效.我们用统计分析方法对12个预测系数、12个部分相关系数、12个对数面积比系数、12个倒谱系数、12个相关系数、短时能量、短时平均过零率及基音共63个参数,进行方差比检验,用10名男青年的三个元音的发音,在半年内采集97组语音作为试验材料,选出15个方差比较大的作为识别参数,识别率为89.19%,采用样本刷新技术后,识别率达到97.3%.

一、引 言

自动说话人辨认是说话人识别中的一个重要课题.它可以用于警卫辅助装置以控制要害部门的出入,控制计算机网中信息资源的访问,银行辨认提款人,辅助司法公安部门对罪犯的辨认,以及在其它公用事业,军事部门中的特殊应用.但是由于声纹不同于指纹,指纹是静态的、固定不变的物理特性,而声纹是动态的、时变的,它和说话人所处的环境、情绪、健康状况都有密切的关系.如果被辨认人不合作,他(她)可以伪装发音欺骗机器.因而说话人辨认的研究具有很大的困难.文献[1]介绍了这方面的研究,认为对说话人识别有价值的个人特

征可分成两类:一类是说话人生理决定的固有特征,反映在基音和低频共振峰,这类特征不易模仿,却容易受健康状况影响;另一类是声道运动的动态特征,也就是发音方式、习惯等,这类特征相对比较容易模仿.因此在机器自动辨认说话人时,提取表征说话人固有特征的参数是非常重要的.究竟什么特征参数最有效,用什么方法选取较有效的参数,我们通过实验,进行了统计分析的研究.

二、实验设计

系统硬件构成如图1,采样率10kHz, A/D字长12比特,低通滤波器截止频率4kHz.用常规的方法完成端点检测,采用基音同步方式

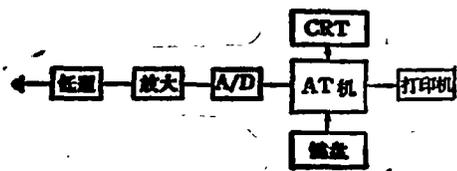


图1 说话人辨认硬件系统图

提取基音及其它参数,以欧氏距离为距离测度,用DTW方法时间配准^[4],生成参考样本,然后辨认判决。

我们选10名讲普通话的男青年,发汉语拼音“a”、“e”、“i”三个单元音,系统将采集的数据存入数据库中,进行文本有关的说话人辨认的研究。间隔1—3周作一次采集,共采集语音数据97组,291个音的语音数据,总时延半年。

文献[3]给出了普通话单元音的舌位图,发“i”时的舌位最高,最前,发“e”时舌位居中,发“a”时舌位最低,较后。但是个人之间是有差别的,正象音位存在变体一样,舌高点也存在“个人”的变体。从“普通话单元音五位发音人的生理舌位图”可以看出,这三个单元音不同程度地反映性别差异和个人声腔差别。另外从唇形的突度和开度曲线,也可以看出个人的发音特点不同。我们考虑到对这三个单元音在一般情况下发音人的发音比较稳定,因此我们采用这三个单元音作为试验样本。

三、各种参数的对比试验

理论分析表明,基音是表征说话人固有特征的参数。另外,发音时声道的某些特征也应是表征说话人固有特征的参数。但是,因分析方法不同,得到的参数不同,所表征的物理意义不同。以全极点模型为基础的线性预测参数可以较为精确地反映语音信号的频谱幅度。把声道模拟为一系列长度相等、截面积不同的无损声管级联,部分相关系数可以表征声波在各声管边界处的反射,对数面积比参数可以表征声管截面积的差别;把语音信号每一短段看成准周期脉冲或随机噪声信号激励一个线性时不变

应用声学

系统的输出,解卷积的方法将激励信号和系统冲激响应分开,那么低时倒谱系数就反映声道的共振性能。

我们分别以12个预测系数(α),12个部分相关系数(k),12个对数面积比系数(g),12个倒谱系数(c),及12个相关系数(r)各为一组,分别就“a”、“e”、“i”,以及这三个单元音组合作为识别样本。取每个人发每个单元音的最初6次的语音数据进行训练,生成参考样本,然后进行发话人辨认。结果如表1。

表1 各种参数对不同样本辨认率(百分数)

样 本	参 数				
	倒谱 c	预测系 数 α	部分相 关 k	对数面 积比 g	相关系 数 r
a	43.24	29.73	45.95	45.95	48.65
e	67.57	35.24	68.86	59.46	48.65
i	72.97	51.35	48.65	43.24	48.15
a、e、i	70.27	48.65	64.86	75.68	51.35

从表1的结果看,对单元音“i”倒谱系数比较有效,辨认率较高,可达72.97%;对发“e”音部分相关系数比较有效,辨认率可达68.86%,而对“a”音各种参数的辨认率都不高。但是将“a”、“e”、“i”三个单元音结合作为识别样本时,对数面积比系数的作用更有效,比任何一个单元音用任何一种参数辨认的辨认率都高,达到75.68%。

四、用方差比测度选择参数的试验

从以上试验结果分析,对于同一个音,在说话人辨认中不同的参数的有效性是不同的;对不同的音,同一种参数的有效性也是不同的。这给说话人辨认带来了复杂性。我们用方差比测度进行参数选择。

设有 N 个发话人被辨认,以 n 给予编号,第 n 个人用作训练样本的发音共有 M_n 次,以 i 表示发音次数排序,每次发音分为 H_{ni} 帧,以 h 作为帧序号,则对于任何一种参数 x ,以 i 作为参数序号,有 x_i, h_{ni} 表示第 n 个人在第 i 次发音

时第 h 帧的第 i 个 x 参数。

定义：个人某参数的均值 $\mu_{x_{in}}$ ：

$$\mu_{x_{in}} = \frac{\sum_{i=1}^{M_n} \sum_{h=1}^{H_{in}} x_{ikhin}}{\sum_{i=1}^{M_n} H_{in}} \dots\dots (1)$$

所有人某参数的总均值 $\bar{\mu}_{xi}$

$$\bar{\mu}_{xi} = \frac{\sum_{n=1}^N \mu_{xin}}{N} \dots\dots (2)$$

发话人之间某参数的均值方差 D_{xi}

$$D_{xi} = \frac{\sum_{n=1}^N (\mu_{xin} - \bar{\mu}_{xi})^2}{N} \dots\dots (3)$$

第 n 个发话人某参数本身方差的均值 d_{xin}

$$d_{xin} = \frac{\sum_{i=1}^{M_n} \sum_{h=1}^{H_{in}} (x_{ikhin} - \mu_{xin})^2}{\sum_{i=1}^{M_n} H_{in}} \dots\dots (4)$$

所有发话人某参数方差的总均值 \bar{d}_{xi}

$$\bar{d}_{xi} = \frac{\sum_{n=1}^N d_{xin}}{N} \dots\dots (5)$$

方差比测度 F_{xi}

$$F_{xi} = \frac{D_{xi}}{\bar{d}_{xi}} \dots\dots (6)$$

我们用基音 p , 短时平均过零率 z , 短时平均能量 E , 12 个预测系数 (α), 12 个部分相关系数 (k), 12 个对数面积比系数 (g), 12 个短时相关系数 (r), 12 个倒谱系数 (c) 作方差比检验, 以方差比的大小排列如表 2、3、4。

我们从表中可以看出

1. 对于不同的音方差比值的排列顺序不同。

2. 没有一种参数的方差比都很大, 全排在前边, 也没有一种参数的方差比都很少, 全排在后边。这就是很多文献用不同类的参数都可以对说话人辨认的原因。

表 2 发“a”音各参数的方差比值表

序号	方差比值	参数	序号	方差比值	参数	序号	方差比值	参数
1	2.948	P	22	.955	c2	43	.698	c6
2	1.663	k4	23	.955	a2	44	.662	a10
3	1.647	g4	24	.952	c5	45	.624	a3
4	1.594	g1	25	.924	a5	46	.621	a9
5	1.464	c4	26	.892	r5	47	.620	k10
6	1.443	c3	27	.891	r12	48	.620	g10
7	1.328	k1	28	.885	a4	49	.618	k3
8	1.317	g2	29	.866	r11	50	.611	g3
9	1.285	g7	30	.853	r4	51	.567	c11
10	1.259	k7	31	.837	r2	52	.565	c12
11	1.241	c10	32	.830	r1	53	.552	a6
12	1.215	Z	33	.824	k9	54	.515	r3
13	1.208	r8	34	.819	k8	55	.461	a11
14	1.172	k6	35	.815	g9	56	.428	a7
15	1.172	k2	36	.802	r9	57	.326	k12
16	1.164	c8	37	.799	g8	58	.324	g12
17	1.155	g6	38	.799	c9	59	.269	a12
18	1.134	c7	39	.791	r6	60	.210	E
19	1.111	r7	40	.751	a8	61	.187	g5
20	1.060	c1	41	.723	k11	62	.184	k5
21	1.056	r10	42	.712	g11			

表 3 发“i”音各参数的方差比值表

序号	方差比值	参数	序号	方差比值	参数	序号	方差比值	参数
1	3.297	c5	22	.849	k10	43	.458	g7
2	2.694	P	23	.848	g10	44	.440	c7
3	2.294	a6	24	.826	k4	45	.418	k1
4	1.565	a5	25	.777	k8	46	.395	r4
5	1.416	a12	26	.772	c1	47	.381	c3
6	1.412	k6	27	.768	g8	48	.377	k3
7	1.392	g6	28	.759	g4	49	.369	r6
8	1.324	c9	29	.722	g1	50	.362	g3
9	1.247	c2	30	.693	c4	51	.348	r2
10	1.247	a2	31	.692	a10	52	.323	r1
11	1.189	c11	32	.662	c6	53	.318	r7
12	1.169	k11	33	.627	a4	54	.314	c8
13	1.168	g11	34	.625	k9	55	.302	r5
14	1.159	g5	35	.623	g9	56	.282	a11
15	1.156	a3	36	.568	a9	57	.268	r8
16	1.134	k5	37	.556	r3	58	.229	r10
17	.991	c10	38	.522	E	59	.228	r9
18	.980	k2	39	.504	c12	60	.203	r11
19	.978	g2	40	.491	Z	61	.164	r12
20	.883	k12	41	.470	a7	62	.102	a8
21	.881	g12	42	.458	k7			

表4 发“e”音各参数的方差比值表

序号	方差比值	参数	序号	方差比值	参数	序号	方差比值	参数
1	4.309	c5	22	1.212	Z	43	.824	α_{11}
2	3.541	α_6	23	1.098	c6	44	.756	c3
3	3.240	P	24	1.098	r4	45	.738	c4
4	2.867	g6	25	1.070	k11	46	.670	r12
5	2.788	k6	26	1.068	g11	47	.659	α_7
6	2.406	c7	27	1.060	α_3	48	.643	g3
7	1.653	c11	28	1.060	k1	49	.636	k3
8	1.500	α_5	29	1.033	r1	50	.627	r11
9	1.454	c2	30	1.031	r2	51	.610	k9
10	1.454	α_2	31	1.003	r8	52	.598	g9
11	1.412	k8	32	1.002	r7	53	.598	c10
12	1.407	g8	33	.986	r5	54	.590	k10
13	1.324	k7	34	.973	α_4	55	.585	g10
14	1.311	α_8	35	.965	r3	56	.544	α_9
15	1.292	k5	36	.960	r6	57	.399	c9
16	1.286	c1	37	.891	g4	58	.369	c12
17	1.284	g7	38	.880	r9	59	.268	α_{10}
18	1.282	k2	39	.879	k4	60	.251	E
19	1.281	g5	40	.879	r10	61	.223	g12
20	1.274	g2	41	.830	α_{12}	62	.223	k12
21	1.260	g1	42	.827	c8			

3. 基音 p 在每个单元音中, 方差比都很大, 是说话人辨认的重要参数。

4. 部分相关系数和对数面积比系数在前 20 个参数排序中, 方差比值很接近, 而且参数序号几乎相同, 这就说明在说话人辨认时, 这两种参数的作用相同, 用一种就可以了。

5. 短时平均过零率 z , 短时平均能量 E , 短

(上接第 45 页)

器的换能器是一个矩形厚板三维耦合振动换能器, 它比圆形换能器有更多的自由度, 而每一个自由度各自有相对应的振动频率, 因此矩形厚板换能器除了主频率 40kHz 之外, 在受到冲击激励时, 还会因谐振和畸变同时复合共振产生到 90kHz 的多种频率, 在清洗槽内产生多频率的复合超声场, 克服了单一频率固有的极为明显的驻波声场的缺陷, 因此 CQF-1 超声复频清洗器具有多频率超声波清洗的特点, 适合精密制品和复杂工件的清洗。

CQF-1 超声复频清洗器具有超声功率调节功能, 功率强弱可连续调节, 这特别适用于脆弱零件的清洗。该仪器设有全波和半波工作状态的选择, 将适应于不同的清洗液体, 从而使清洗效果更佳; 清洗槽设有自动温控装置, 槽液温度可通过清洗槽面板上的“温度控

应用声学

时相关系数(r)对说话人辨认贡献不大。

6. 从前 20 个方差比比值看, 随着发音舌位的升高, 有效参数的序号有变大的趋势, 而且预测系数的作用也越来越显著。

我们以方差比比值大小排列, 选择前 20 个参数中的 15 个 (扣除对数面积比参数), 对样本进行训练, 得到参考样本, 然后进行辨认。发“a”音的辨认率 56.76%, 发“e”音的辨认率 86.49%, 发“i”音的辨认率 75.68%。设置一定的置信度, 综合发“a”、“e”、“i”三个音的辨认率 89.19%。我们采用样本刷新技术后, 最后辨认率可达 97.3%。

五、结束语

方差比选择的参数在自动说话人辨认中是有效的, 可以大大提高辨认率。但是用方差比选择的参数是怎样表征个人的生理固有特征, 说话习惯, 这是一个相当复杂的问题, 有待进一步研究。

参 考 文 献

- [1] 陈永彬, 王仁华, 语音信号处理, 中国科技大学出版社, 1990.
- [2] 文成义等译, 语音处理, 国防工业出版社, 1990, 219—224.
- [3] 吴宗济, 林茂灿主编, 实验语音学概要, 高等教育出版社, 1989, 78—81.

制²调节旋钮来设定并自动保温, 选择合适的槽液温度可使清洗效果达到最佳的理想状态。

CQF-1 超声复频清洗器噪音低结构新颖, 采用特种换能器清洗槽及全不锈钢结构, 耐腐蚀, 清洗工作稳定可靠。

主要技术指标

1. 超声波主频 40kHz; 2. 输出功率连续可调 全波时最大功率不小于 300W, 半波时最大功率减半; 3. 电源 220±22V 50Hz; 4. 发生器电源功耗: ≤700W
5. 清洗槽加热电源功耗 900W; 6. 连续工作时间 8h; 7. 外形尺寸 发生器 350mm×300mm×210mm, 清洗槽 460mm×360mm×340mm

(高 玉)