国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY
第 19 卷第 3 期　1997 年 6 月　Vol. 19 No. 3

一种基于改进的 SEVQ 匹配算法的汉语全音节语音识别系统*

唐朝京　吴自强　王跃科　张　南　周代英　王成友

(国防科技大学电子技术系　长沙　410073)

摘　要　本文全面介绍了一种采用改进的 SEVQ 匹配算法的特定人汉语语音识别系统，具体描述了系统总体方案、系统参数的选取策略、各种识别预处理所用方法及语音模式匹配原理和方法。系统的实时识别率超过 93%。

关键词　语音识别，模式匹配，信号处理

分类号　TN 912.34

A Whole- Syllable Chinese Speech Recognition System Based on Improved SEVQ Matching Algorithm

Tang Chaojing　Wu Ziqiang　Wang Yaoke
Zhang Nan　Zhou Daiying　Wang Chengyou
(Department of Electronic Technology, NUDT, Changsha, 410073)

Abstract　A speaker- dependent Chinese speech recognition system based on improved SEVQ matching algorithm is presented. The overall plan of the system, the choosing method of the system's parameters、methods of recognition preprocessing、principles and methods of speech mode matching are described particularly. The system's real-time recognition rate is over 93%.

Key words　speech recognition, mode matching, signal processing

语音识别技术是一项涉及到计算机、数字信号处理、模式识别、人工智能、语音学、语法学等多门学科的前沿技术，是近 10 年来国内外非常引人注目的研究领域，具有重要的学术研究价值和良好的应用前景。国防科技大学研制的华声智能语音识别系统是由以 DSP32C 信号处理器为核心的微机插卡与实时识别软件构成的特定人全音节实时汉语语

* 1996 年 12 月 25 日修订

音识别系统, 它以 LPC 倒谱参数作为特征参数, 以加权倒谱距离作为距离测度, 综合采用了多种有效的方法, 完成语音起末点检测、声/韵母分界、基音提取、四声判决、样本子集分类等识别预处理工作, 然后采用改进的有序码矢量化法(SEVQM) 先后对韵母和全音节进行模式匹配, 完成待测语音的分类分层匹配, 最后加上音字转换软件, 输出有意义的汉语文本。该系统实现了较高的识别率, 根据测试, 在 1200 多个含调音节中, 首选正确识别率超过 93%, 达到国内先进水平。本文对该系统的主要技术作一介绍。

2 系统参数的选取

2.1 特征参数

本文以语音倒谱参数作为识别的特征参数。原始语音 $S(n)$ 的倒谱 $\hat{S}(n)$ 可以从 LPC 系数中递推得到。取 LPC 分析阶数 $p=16$, p 个 LPC 系数$(\alpha_1, \cdots, \alpha_p)$ 利用协方差法求得。实验表明, 采用 16 阶分析阶数对于声韵母的匹配效果是令人满意的。

LPC 分析帧长取为 16ms, 而语音采样速率取 $f_s=10$kHz. 倒谱 $\hat{S}(n)$ 的系数 C_i 与 LPC 系数$(\alpha_1, \cdots, \alpha_p)$ 的递推关系为

$$C_n=\begin{cases}-\alpha_1 & n=1\\ -\alpha_n-\sum_{k=1}^{n-1}\frac{k}{n}\alpha_{n-k}C_k & 1<n\le p\\ -\sum_{k=1}^{p}\frac{k}{n}\alpha_{n-k}C_k & p<n\le Q\end{cases}$$

倒谱分析阶数 Q 通常选 $Q\ge\frac{3}{2}p$, 但通过实验发现: 当取 $Q=14$ 时, 已能较好地表征语音特征, 故本文选取倒谱阶数 $Q=14$, 以节约存贮资源, 并减少运算时间。

2.2 距离测度

本文选取加权倒谱距离 d_{WCEP}作为距离测度。

$$d_{WCEP}(r,t)=\sum_{i=1}^{14}w(i)[c_r(i)-c_t(i)]^2$$

其中 $c_r(i)$ 为倒谱系数向量 c_r 中的各元, $\omega(i)$ 为加权值。我们选定的加权系数为平滑群延迟加权系数:

$$\omega(i)=i\quad\exp(-i^2/2\tau^2),\ \tau\text{取为 }16$$

实验结果表明, 选取上述加权倒谱距离可以得到良好的匹配结果。

3 语音识别预处理

为了进行待测语音与训练模板的匹配运算, 首先必须进行语音起末点检测、声/韵母分界、基音检测、四声判别等一系列预处理。

3.1 起末点检测

本文主要利用多门限过零率 MZCR 对语音信号进行起点判断, 将声母从背景噪声中分割出来。首先对背景噪声进行训练, 求出噪声的统计均值 S_A 和标准差 D_{SA}:

$$S_A=\frac{1}{NL}\sum_{i=1}^{NL}Si,\quad D_{SA}=\sqrt{\frac{1}{NL}\sum_{i=1}^{NL}(Si-S_A)^2}$$

据此计算 4 对过零率门限{TH_i, TL_i}:

$$TH_i = S_A + 2^i \quad D_{SA}, TL_i = S_A - 2^i \quad D_{SA} \quad (i = 1, 2, 3, 4)$$

选取权系数 $W = \{1, 4, 8, 64\}$, 综合求出多门限过零率 MZCR:

$$MZCR = \sum_{i=1}^{4} W_i \quad Z_i, \qquad (Z_i \text{ 为穿越第 } i \text{ 对门限}\{TH_i. TL_i\} \text{ 的次数})$$

当检测输入语音的起点时, 首先求出前 50 帧 MZCR 的平均值 MZAV, 取定高门限 T_{ZH}与低门限 T_{ZL}分别为

$$T_{ZH} = 150 \quad MZVA, \quad T_{ZL} = 2 \quad MZVA$$

然后从前向后逐帧计算 MZCR, 找到首次超过 T_{ZH}的位置, 再从后向前回索, 寻找 MZCR 首次小于 T_{ZL}的位置, 此处即确定为语音起点。

末点检测利用语音的全通能量进行。从语音起点开始, 求取信号的全通能量 $EP(n)$, 找出最大值 EP_{MAX}, 同时将 $EP(n)$ 与 EP_{MAX}比较。当满足 $EP(n) < \frac{1}{50}EP_{max}$时, 即判为语音末点; 当末点与起点帧号之差小于 20 时, 可认为是干扰噪声而忽略。

3. 2 声/韵母分界点检测

本文利用语音能量、LPC 倒谱系数、归一化 LPC 残差能量及基音周期相结合实现声韵分界。在实际分割中, 首先用能量确定出粗略的声韵分界点 S, 然后在包含 S 的一个区域[DOWN, UP]内根据倒谱系数、归一化 LPC 残差能量以及能量的相对变化率确定出声/韵母分界点 D_S.

求出声/韵母分界点后, 再对基音的平稳特性逐帧判断。如果基音参数是从某帧起连续 3 帧平稳, 则该处即可认为是韵母起点。

3. 3 基音提取与四声判别

本文采用中心削波自相关法(AUTOC)提取基音[5]。为确保提取基音的稳定性和适应性, 选取分析帧长为 200 点(20ms), 帧移 40 点(4ms)。

为了加强信号中的低频信息, 进一步提高基音估值精度, 在中心削波之前加入了平方幅度变换。

实验结果表明, 四声判别正确率稳定在 98% 以上。

3. 4 声母分类

本文在对语音匹配前进行声母分类。声母分类的目的在于减小声母类间错误, 同时提高系统识别速度。经过大量实验研究, 最后确定采用过零率参数将语音的声母分成擦与非擦两大类:

擦类: f, h, s, sh, x, c, ch, q, z, zh, j

非擦类: b, d, g, p, t, k, m, n, l, r, 零声母

声母分类所用的过零率门限定为 60. 当过零率超过 60 时判为擦类声母, 否则判为非擦类。经测试对于 1200 多个汉语含调音节, 上述分类正确率可达 99%以上。

4 基于改进的 SEVQ 的分类分层匹配方法

4. 1 SEVQ 匹配法及其改进

由于语音信号在时域上具有相当大的随机性, 因此待测语音模板与参考模板的距离

必须采用时间规正技术。针对 DTW 匹配算法匹配速度慢及要求时间严格对准的缺点,我们采用 SEVQ(有序码矢量化)为核心匹配算法[4,7],并在实验的基础上对 SEVQ 进行改进,得到 SEVQM. 使用 SEVQM 作为实际的模式匹配算法,使系统识别率得到了明显的提高。

设待测语音模板 $T=\{t_i\}, i=1,2,\cdots,I$

参考语音模板 $R=\{r_j\}, j=1,2,\cdots,J$

一般 $I\neq J$. 如图 1 所示,在 TOR 平面上,直线 OO' 的斜率为 $k=I/J$,即在直线 OO' 上有 $i=kj$.

由于发音速度的随机性,待测模板与参考模板的对应关系并不总能保证落在 OO' 线上,但在一个音节的持续时间内,局部语流速度与平均速度相差不会太远,因此存在一个常数 ω,使得 r_i 的下标 i 落入如下范围:

$$k\ \ j-W_I\leq i\leq k\ \ j+W_I$$

即 i 落在图 1 中线段[A,B]内。令

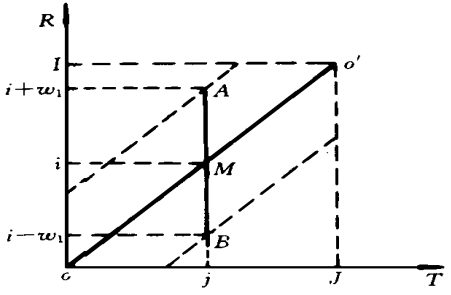

图 1 SEVQ 法匹配路径

$$\mathrm{IL}(j)=\max(k\ \ j-w_1, I),\quad \mathrm{IH}(j)=\min(k\ \ j+w_1, I)$$

则 i 的取值范围为

$$\mathrm{IL}(j)\leq i\leq \mathrm{IH}(j)$$

因此对于一特定的 w_1 值,对待测模板的每一点 j,可相应地确定一组参考矢量:

$$\{r[\mathrm{IL}(j)], r[\mathrm{IL}(j)+1], \cdots, r[\mathrm{IH}(j)]\}$$

分别计算各点与 t_j 的匹配距离,再求出其中的最小值 $d_p(j)$,全部 J 个 $d_p(j)$ 的平均值即为 SEVQ 的匹配结果,即:

$$D_{\mathrm{SEVQ}}(I,J)=\frac{1}{J}\sum_{j=1}^{j} d_p(j)$$

其中

$$d_p(j)=\min\{d(i,j)\ \ \mathrm{IL}(j)\leq i\leq \mathrm{IH}(j)\}$$

上述SEVQ 法对待测模板的每一个匹配点 j,都在线段 AB 上求得一个距离最小点。如果待测音的数据不稳定,存在跳变点,则由SEVQ 得到的匹配结果就会产生误差。鉴于此,我们对 SEVQ 匹配路径进行改进,引入了 SEVQM 法,其基本原理如图 2 所示。对应每一个匹配点 j,SEVQM 要在以M 为中心的四边形区域CDEF 中求一个距离最小点,匹配距离为:

$$D_{\mathrm{SEVQM}}(I,J)=\frac{1}{J}\sum_{j=1}^{J} d_m(j),$$

其中

$$d_m(j)=\min\{d_p(j+w)\ \ -w2\leq w\leq w2$$

$d_p(j+w)$ 定义与SEVQ 法中 $d_p(j)$ 定义相同。

SEVQM 的最大优点是:待识别音特征能抗数据点双边不稳定的跳变。本文所有的模式匹配方法均采用 SEVQM 匹配,收到了理想的效果。

4.2 韵母匹配与全音节匹配

本文的模式匹配过程利用 SEVQM 先对韵母进行匹配, 再作全音节匹配, 完成待测语音的识别。

韵母匹配的目的是筛选出最有可能的前 5 选韵母, 缩小后续全音节匹配的范围, 提高识别速度与识别正确率。实际测试结果表明: 韵母前 5 选的识别正确率超过 99%。

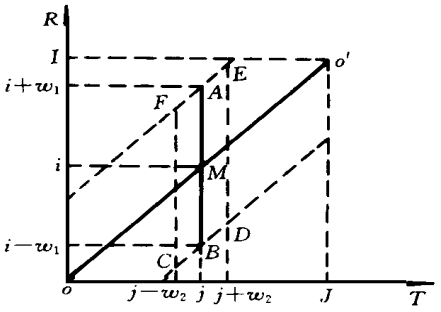

图 2 SEVQM 法匹配路径

韵母匹配过程完成以后, 接着进行全音节匹配。全音节匹配过程中, 我们根据实验语音学知识, 对声母段和韵母段的距离信息进行了分段加权处理。

一个完整的语音音节 SPEECH 可表示成声母段 C_1、过渡段 C_2、韵母段 C_3 之迭加:

$$\mathrm{SPEECH} = C_1 + C_2 + C_3$$

其中声母段 C_1 可明确区分出来, 而过渡段 C_2 一般不易区分。本文的分段帧间加权对 C_1+C_2 段各帧距离乘以权系数 v_1, 而对 C_3 段各帧距离乘以权系数 v_2, 得总的匹配距离为:

$$D = v_1 \sum_{C_1+C_2} d(i) + v_2 \sum_{C_3} d(i)$$

其中过渡段长度 C_2 及两个权值 v_1 与 v_2, 根据声母时长的不同而不同。本文通过大量实验, 确定出 V_1、V_2 和 C_2 的最佳值如表 1.

4.3 多模板匹配策略

表 1

C_1 长度	V_1	V_2	C_3
$C_1<3$	6	1	4
$3\leq C_1<9$	4	1	2
$C_1\geq 9$	1.5	1	2

为了提高系统识别率的稳定性, 匹配过程中采用了多模板策略, 在训练建库时除了对以 m, n, l, r 等声母开头的音节只建一声库外, 对其它所有音节可能存在的声调都需建库。匹配时, 将待测音节与参考音节的所有声调模板都进行匹配, 以确定正确的音节, 而待测音节的声调并不靠多模板匹配过程得出, 而是单独通过四声判别得到。实验结果表明: 采用上述多模板匹配策略以后, 识别率约提高了 8%, 而系统识别率的稳定性也大为提高, 任何人重复测试的识别结果波动不会超过 2%, 而未采用多模板匹配时系统识别率的波动范围可高达 15% 以上。

5 结束语

我们采取上述各种有效的识别技术后实现了系统的高识别率。经测试, 在 1200 多个含调音节中, 首选正确识别率超过 93%, 达到了国内外同类系统的先进水平。

本文是在我们研制成功的华声智能语音识别系统的基础上总结提炼而成的。为该系统的研制成功付出过辛勤劳动的还有朱庶、马建华、王耀勋、苗普选、刘桐、陈星耀、宁军、王敏等同志。

(下转第 60 页)

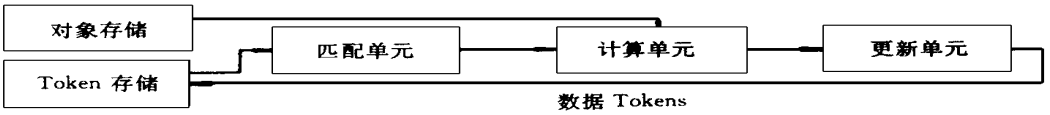

图 1　虚拟的面向对象大粒度数据流模型

象技术设计了一个面向对象高层次、高性能、易于扩展的文件 I/O 对象, 以及易于操作的网络文件 I/O 对象机制。最后, 提出了 OOLGDFM 抽象机。在体系结构方面支持 OOCPCS 的程序执行。有关系统的测试结果请见[8]。

参 考 文 献

1　Wegner P. Dimensions of Object __Based Languages Design. In: ACM OOPSLA´87 Proceeding, USA. Oct, 1987: 21～35

2　Davis A L, Keller R M. Data Flow Program Graphs. IEEE Computer, 1982, 15(2): 31～41

3　Arvind, Gostelow K P. The U-Interpreter. IEEE Computer, 1982, 15(2): 42～50

4　Robert G. Parallel Processing with Large Grain Data Flow Techniques. IEEE Computer, 1984, 17(7): 55～61

5　Grimshaw A S, Liu J W S. MENTAT: an Object-Oriented Macro Data flow System. In: OOPSLA´87 Proceedings. 1987

6　Stroustrup B. The C++ Programming Language. Addison. Westey, 1986

7　肖侬, 胡守仁. 网络的面向对象 I/O 系统开放式分布系统, 南京, 1995

8　肖侬. 基于网络环境的面向对象语言 C++ 的并行化技术的研究与实现: [博士论文]. 长沙: 国防科学技术大学计算机系, 1996

(上接第 43 页)

参 考 文 献

1　张世平. 普通话全音节识别系统: [博士论文]. 清华大学, 1988

2　易克初. 普通话全音节识别系统及语音识别与合成若干新问题的研究: [博士论文]. 清华大学, 1989. 3

3　王跃科. 机器人语音通信的理论与技术研究: [博士论文]. 华中理工大学, 1989

4　易克初, 王跃科. 有序码书矢量量化及其在语音识别中的应用, 语音图像通讯信号处理学术会议, 1989

5　陈尚勤, 罗承烈, 杨雪. 近代语音识别. 电子科技出版社, 1991

6　Tokura Y A. Weighted Cepstrum Distance Measure for Speech Recognition. IEEE Trans. 1987, ASSP-35(10)

7　Yi K H, Wang Y K. Sequential VQ and its Application to Chinese Dictation Recognizer. ICASSP-90

(责任编辑　潘　生)