文章编号:1671 - 7147(2003)02 - 0142 - 03

孤立词语音识别系统的一种实用精简算法

李 挺
(江南大学 机械工程学院,江苏 无锡 214063)

摘 要: 提出了一种以降低识别计算代价为目标的孤立词语音识别系统的设计算法. 语音识别系统要求具有很强的实时性,同时应保证较好的识别率. 该设计对语音信号采用了处理速度较快的时间轴和幅值上规正化的数据压缩算法,并采用加权离散度法进行识别,算法精确、简便、可靠,适合作为小型语音识别产品的主要算法.

关键词: 语音识别; 孤立词; 聚类分析
中图分类号: TP 391. 42 **文献标识码**: A

A Practical and Efficient Arithmetic for Isolated Speech Recognition System

L I Ting
(School of Mechanical Engineering, Southern Yangtze University, Wuxi 214063, China)

Abstract: The paper provides a method for designing a set of isolated speech recognition system for the purpose of reducing expense in recognition calculation. Speech recognition has strict real time requirement and high recognition accuracy. More rapid data compress arithmetic is used to process speech signal data by normalizing them in time and amplitude. A more efficient, easier and higher dependability recognition method-the method of discrete degree added by power is provided in the paper. The arithmetic is suited for designing a small speech recognition production.

Key words: speech recognition; isolated word; clustering analysis

语言是人类最重要的交流工具,也应是人机之间最有效的通信手段. 目前,各国对机器语音识别及合成的研究已达到相当的水平,语音合成也早已商品化. 语音识别较合成难度大,商品化困难,但对于孤立语音识别,一些识别方法已相当有效. 作者提出了一种以降低识别计算代价为目标的孤立语音识别方案,按该方案设计的语音识别程序甚至能在 IBM PC/ XT 一级计算机上达到较好的识别效果,当然其也可移植到单片机系统上形成实用化的语音识别产品.

1 语音识别系统结构方案

语音识别系统采用模拟带通滤波器组进行抽取,经模拟量数值量化后,由软件进行幅值规整和非时间规整,形成语音特征参数矩阵. 硬件及软件系统结构如图 1 所示[1].

语音识别硬件配置为:1) IBM PC 兼容机;2) 语音输入接口卡;3) A/ D 转换卡;4) 麦克风.

收稿日期:2002 - 05 - 23; **修订日期**:2003 - 03 - 10.
作者简介: 李挺(1968 -),男,江苏无锡人,工学硕士,讲师.

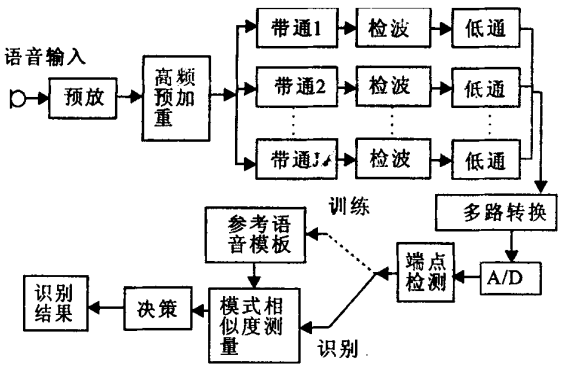

图 1　语音识别软硬件结构原理

Fig. 1　The principle diagram of software and hardware for speech recognition

2　特征抽取

一般认为语音在 10～30 ms 短时段内平稳[2],根据采样定理,采样率设置为 200 Hz,以对应 5 ms 采样周期. 语音信号经 J 组(本系统 $J = 16$)带通滤波后,进行检波平滑. 系统能自动判断语音的开始和结束. 一旦判定语音开始,就将取得的 16 个通道的语谱值,依次存放到内存,作为矩阵第一行 S_1,等二次采样得第 2 组 16 个数据,作为矩阵第 2 行 S_2,……语音结束前的 N 行数据,构成一个 $I \times J$($J = 16$)的原始语谱矩阵 S.

$$S = [S_1, S_2, \cdots, S_i, \cdots, S_I]^{\mathrm{T}} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1J} \\ S_{21} & S_{22} & \cdots & S_{2J} \\ \cdots & & & \\ S_{I1} & S_{I2} & \cdots & S_{IJ} \end{bmatrix} = \{S_{ij}\}_{I \times J} \tag{1}$$

3　数据压缩

3.1　语音原始谱矩阵的时间规整

语音原始谱矩阵数据的压缩方法可按以下两个步骤进行.

3.1.1　计算语音原始谱矩阵总长度　采用 Chebyshev 范数距离度量,则对应于原始谱矩阵 S 的特征轨迹总长度 L 为

$$L = \sum_{i=1}^{I} \sum_{j=1}^{J} |S_{(i-1)j} - S_{ij}| \tag{2}$$

式中,设 $S_{0j} = 0$,　$j = 1, 2, 3, \cdots J$.

3.1.2　原始谱矩阵在时间上的段长计算　将特征轨迹分为 M 段,则对应于特征轨迹各分段段长 ΔL 为

$$\Delta L = L / M \tag{3}$$

3.1.3　原始谱矩阵在时间上重抽样

重抽样按下列原则进行:

对第 i 段　　$1 \leq i' \leq M$,

i 段在 S 中对应的末行行号为 $l_{i'}$,则当

$$\sum_{i=l_{i'-1}}^{l_{i'-1}} \sum_{j=1}^{J} |S_{(i-1)j} - S_{ij}| < i' \cdot \Delta L < \sum_{i=l_{i'}}^{l_{i'}} \sum_{j=1}^{J} |S_{(i-1)j} - S_{ij}| \tag{4}$$

有　$\overline{S}_{i'j} = \dfrac{1}{l_{i'} - l_{i'-1}} \cdot \sum_{K=l_{i'-1}+1}^{l_{i'}} S_{kj}$　　(5)

这样,由 $\overline{S}_{i'j}$ 可得到经时间上重分段的语音. 特征矩阵 $\overline{S}$

$$\overline{S} = \begin{bmatrix} \overline{S}_{11} & \overline{S}_{12} & \cdots & \overline{S}_{1j} & \cdots & \overline{S}_{1J} \\ \overline{S}_{21} & \overline{S}_{22} & \cdots & \overline{S}_{2j} & \cdots & \overline{S}_{2J} \\ \cdots & & & & & \\ \overline{S}_{M1} & \overline{S}_{M2} & \cdots & \overline{S}_{Mj} & \cdots & \overline{S}_{MJ} \end{bmatrix} = \{\overline{S}_{i'j}\}_{M \times J} \tag{6}$$

经重新分段后的新的特征矩阵 $\overline{S}$ 较 S,数据得到了压缩,此外由(4),(5)式可见,对应于语音平稳段,数据得到了更大压缩. 因此对原始谱的时间分段法,可更有效地压缩语音平稳段的冗余信息.

3.2　语音数据的幅值规整

对应于同一语音,两次发音程度上的差异会导致同一单词间语音样本之间差距增大. 为了克服这一缺点,对 $\overline{S}$ 矩阵作幅度归一化处理.

$$\Delta = \overline{S}_{\max} - \overline{S}_{\min}$$
$$\Delta_{i'j} = \overline{S}_{i'j} - \overline{S}_{\min} \tag{7}$$

式(7)中,$\overline{S}_{\max}$, $\overline{S}_{\min}$ 分别为矩阵 $\overline{S}$ 中最大和最小元素,归一化按下式进行

$$\alpha_{i'j} = [(2^n - 1)\frac{\Delta_{i'j}}{\Delta}]_{\mathrm{INT}} \tag{8}$$

式(8)中,算符 $[\cdot]$ 表示取整运算,n 为 A/D 量化后二进制的位数,构成的新矩阵 A 即为幅值归一化后新特征矩阵

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1J} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2J} \\ \cdots & & & \\ \alpha_{M1} & \alpha_{M2} & \cdots & \alpha_{MJ} \end{bmatrix} = \{\alpha_{i'j}\}_{M \times J} \tag{9}$$

4　参考模式的建立及其聚类分析

语音是一种由生物学特点起因的模式,这种模式所含有的语义信息往往和讲话人的各种生理和心理的状况有关. 从发声的过程中,设法收集由上述因素引起的每一个附加特征足够的统计信息较为困难. 由于上述因素,使得语音信号带有某种程度的随机性和模糊性. 在建立语音参考模式时,若仅使用某一次的语音输入作为参考模式,就不可能反映出语音的模糊性,而必须多次重复输入(训练)同

一语音,才能总结出该语音的统计特性[4],见图2.

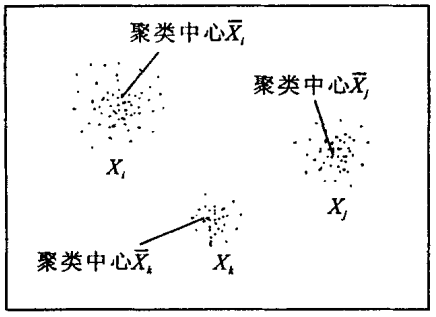

图2 语音聚类分析图

Fig. 2 The clustering analysis diagram for speech recognition

由于语音的随机性和模糊性,对于同一单词发音 n 次,n 个发音得到 n 个语音样本,形成一个区域的集合 X_i,不同词的发音可以得到不同区域的集合 X_j, X_k 等. 分别找出各个区域的聚类中心 $\overline{X_i}$, $\overline{X_j}$, $\overline{X_k}$ …,将待识别的语音样本 X 与各区域的聚类中心 $\overline{X_i}$, $\overline{X_j}$, $\overline{X_k}$ …相比较,从而确定其属于哪一类.

整个语音识别系统的参考模式可以表示为语音参考模式类的集合.

$$S = \{ X_1, X_2, \cdots, X_r \} \tag{10}$$

$X_i = \{ X_{i1}, X_{i2}, \cdots, X_{in} \}$, X_{ij} 为第 i 个词第 j 次训练所得的参考模式

$$X_{ij} = A_{ij} = \begin{bmatrix} \alpha_{11}^{ij} & \alpha_{12}^{ij} & \cdots & \alpha_{1J}^{ij} \\ \alpha_{21}^{ij} & \alpha_{22}^{ij} & \cdots & \alpha_{2J}^{ij} \\ \cdots & & & \\ \alpha_{M1}^{ij} & \alpha_{M2}^{ij} & \cdots & \alpha_{MJ}^{ij} \end{bmatrix} = \{ \alpha_{pq}^{ij} \}_{M \times J} \tag{11}$$

词语数 $i = 1, 2, \cdots, r$; 训练次数 $j = 1, 2, \cdots n$.

X_i 的聚类中心 $\overline{X_i}$ 为

$$\overline{X_i} = \overline{A_i} = \frac{1}{n} \sum_{j=1}^{n} A_{ij} = \{ \bar{\alpha}_{pq}^{j} \}_{M \times J} \tag{12}$$

式中

$$\bar{\alpha}_{pq}^{j} = \frac{1}{n} \sum_{j=1}^{n} \alpha_{pq}^{ij} \tag{13}$$

此外,本方法还应用了离散度概念,应用离散度主要为了解决以下问题:假定语音样本 X 与参考模式类 X_i, X_j, X_k …的聚类中心 $\overline{X_i}$, $\overline{X_j}$, $\overline{X_k}$ …的距离相等,在利用距离进行判别时,不可能判别 X 是属于 X_i 还是 X_j,这时就必须利用离散度解决这个问题. 由图可知, X_i 类的语音的参考模式的离散度较大,因此,可以认为 X 属于 X_i. 图2中,定义 X_i 的离散度 λ_j 为

$$\lambda_i = \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{M \times J} \sum_{p=1}^{M} \sum_{q=1}^{J} | \alpha_{pq}^{ij} - \bar{\alpha}_{pq}^{j} | \right) \tag{14}$$

5 语音识别

由式(12),(13),(14)可以求得 X_i 类($i = 1, 2, \cdots, r$).

语音参考模式的聚类中心分别为

$\overline{A_1}, \overline{A_2}, \overline{A_3}, \cdots, \overline{A_r}$

现有新的识别语音模式为

$$A = \{ \alpha_{pq} \}_{M \times J}$$

定义距离 $d_i = \sum_{p=1}^{M} \sum_{q=1}^{J} | \alpha_{pq} - \bar{\alpha}_{pq}^{j} |$ (15)

由(15)式可以求得距离向量为

$$D = [d_1, d_2, d_3, \cdots, d_r]^{\mathrm{T}}$$

将向量 D 中元素按从大到小排列得

$$D' = [d_1', d_2', d_3', \cdots d_r']^{\mathrm{T}}$$

则 d_1 所对应参考模式 $\overline{A_i}$ 即为 A 识别后的结果归类模式,其中 $i = 1, 2, \cdots, r$.

若 $d_1' = d_2$ 所对应参考模式为 $\overline{A_i}$ 和 $\overline{A_j}$, $j = 1, 2, \cdots, r$; $j \neq i$ 且 $\lambda_i > \lambda_j$. 则 $\overline{A_i}$ 为 A 识别后所得结果的归类模式.

6 结 语

文中介绍的语音识别方法,已由作者在 PC/XT 上用汇编程序成功实现,并已对机器人实现一系列动作的控制. 比如:步行机器人[5]的“前进”、“后退”、“左转”、“右转”,“前进 x 步”命令等;以及控制工业机械手“抓取 A 物体”,“A 上放上 B”,“A 上卸下 B”命令等. 其中 x, A, B 分别是数字和物体名. 使用结果表明该算法实时性强,识别率较高.

参考文献:

[1] 李挺. 孤立语音识别系统在自动编程中的应用研究[D]. 南京:南京航空航天大学,1993.

[2] RABINER L R, SAMBUR M R. An algorithm for determining the endpdints of isolated utterances[J]. Tech J, 1975, 54(2): 297 - 315.

[3] 许才刚. 数控机床的语音交互式控制系统研究与实现[D]. 南京:南京航空学院,1990.

[4] 王松年. Fuzzy 聚类分析的语音识别[J]. 模糊数学,1984, 127: 47 - 54.

[5] 尉忠信、竺钦尧、李挺,等. 智能双足步行机器人技术报告[Z]. 南京:南京航空航天大学机电工程学院,1995.

(责任编辑:彭守敏)