

车载语音识别系统设计与试验*

王登峰 曹晓琳 于吉龙

(吉林大学汽车动态模拟国家重点实验室, 长春 130025)

【摘要】 用改进的动态时间归整(DTW)算法进行语音辨识,设计了语音识别系统软件,以 SPCE061A 单片机为核心设计了车载语音识别系统的电控单元(ECU)。在车内不同噪声环境下进行了语音识别试验,结果表明语音识别率随着车内环境噪声的增加而下降,随着语音字数的增加而降低;在车辆怠速车内噪声小于 50 dB(A)的环境中 2 字语音平均识别率为 90%、4 字语音的识别率为 85%;在车辆定置油门半开、车内噪声为 60~70 dB(A)的环境中,2 字语音平均识别率为 85%、4 字语音的识别率为 80%。

关键词: 车辆 语音识别系统 动态时间归整算法 设计 试验

中图分类号: U463.6; TN912.34 文献标识码: A 文章编号: 1000-1298(2010)01-0041-05

Design and Experiment of a Vehicular Speech Recognition System

Wang Dengfeng Cao Xiaolin Yu Jilong

(State Key Laboratory of Automotive Dynamic Simulation, Jilin University, Changchun 130025, China)

Abstract

An improved dynamic time warping (DTW) speech recognition method is proposed. An electronic control unit of vehicular speech recognition system is developed based on chip microcontroller SPCE061A. The corresponding control software is programmed and the speech recognition tests are carried out under the different noise conditions within an experimental vehicle. The test results show that the speech recognition ratio reduces with the increase of the vehicle interior noise level and the speech length. The average recognition ratio is 90% for two Chinese characters speech and 85% for four Chinese characters speech when the vehicle is under the idle running condition and the interior noise is less than 50 dB(A). And the average recognition ratio is 85% for two Chinese characters speech and 80% for four Chinese Characters speech when the vehicle is under the fixation and half-throttle running condition and the interior noise level range is from 60 to 70 dB(A). The speech recognition results are satisfactory.

Key words Vehicle, Speech recognition system, Dynamic time warping method, Design, Experiment

引言

车载语音识别系统技术难点是对环境依赖性、语言模型的构建以及语音的端点检测和特征提取^[1-3]。近年来,随着人们对车辆性能要求的不断提高,语音识别技术日渐在车辆上实现一些智能操作,以满足现代车辆消费者在车内工作、休息、娱乐

以及行车安全等方面的要求。通过语音识别技术,车上装备的音频和视频播放器、空调、电动车窗、移动电话、全球卫星定位系统、卫星电台和因特网浏览器等电子系统均可通过语音识别进行操控^[4]。

本文设计开发一种车载语音识别系统,该系统利用改进的动态时间归整(DTW)语音识别方法进行语音辨识,设计出车载语音识别系统硬件和软件,

收稿日期: 2009-02-16 修回日期: 2009-05-07
* 吉林省科技发展计划资助项目(20076030、20040322)和山东省泰山学者特聘专家计划资助项目
作者简介: 王登峰,教授,博士生导师,主要从事汽车振动噪声分析与控制研究,E-mail: caewdf@jlu.edu.cn

系统经训练后能够通过驾驶员的简单语音指令,完成对车载电子电器系统的语音控制。

1 改进的动态时间归整(DTW)语音识别方法

动态时间归整(DTW)算法^[5]就是将未知语音和参考语音在时间轴上进行非线性的弯曲修正,以使2次发音能更好地匹配。为了将2次发音进行相似性度量,将特征模板库中的各个模板称为参考模板,用 R 表示;将待识别的输入语音经预处理和特征提取之后得到的特征矢量序列称为测试模板,用 T 表示。对比测试模板 T 和参考模板 R 之间的相似度,可以计算它们之间的失真,失真越小,相似度越高。测试模板和参考模板之间的总体失真可以表示为 $D[T,R]$ 。为了计算这一失真,应从测试模板和参考模板之间对应的每一帧的失真算起。假设 n,m 分别为 T,R 中任意选择的帧号,则两帧的失真用 $d[T(n),R(m)]$ 来表示。距离函数取决于实际采用的距离度量,在DTW算法中通常采用欧氏距离,其计算公式为

$$D[X,Y]=\frac{1}{K}\sqrt{\sum_{i=1}^K(x_i-y_i)^2}$$

(1)

考虑到在计算总体失真 $D[T,R]$ 时,因测试模板和参考模板之间的语音帧数往往不等,这时需要将测试模板用某种伸缩的方法映射到参考模板上,然后再分别计算新的各帧之间的失真,从而求得总失真 $D[T,R]$,具体方法如下:

采用一个二维直角坐标系标定测试模板的 N 帧和参考模板的 M 帧,构建如图1所示的网格,网格中的每一个交叉点表示测试模板中的某一帧与参考模板中的某一帧的交汇点,并且失真度为 $d[T(n),R(m)]$ 。计算各交汇点失真度

$$D[T,R]=\sum_{n=m=1}^N d[T(n),R(m)]$$

(2)

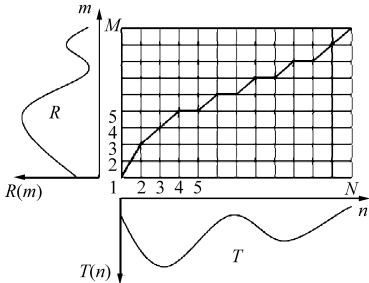


图1 动态时间归整求最小失真示意图
Fig.1 Calculating minimum distortion diagram by dynamic time warping algorithm

即可获得帧匹配距离矩阵。

DTW在测试模板和参考模板首尾端点对准的前提下,寻找一个归整函数 $m_i=\Phi(n_i)$,将测试矢量

的时间轴 N 非线性地映射到参考模板的时间轴 M 上,通过各交叉点到终止点的最优化路径,使得该路径上所有交叉点的帧失真度总和达到最小,即

$$D=\min_{\Phi(n_i)}\sum_{i=1}^N d[T(n_i),R(\Phi(n_i))]$$

(3)

语音识别时对不同的语音建立相应的参考模板,各参考模板和需要识别的测试模板进行DTW算法模式匹配,得到不同的累计距离 $D[N,M]$,累积距离最小时说明相似度最大,所对应的语音就是识别结果。

上述动态时间归整算法存在的缺点是模式匹配的运算量大,当测试和参考模板长度较大时,训练和识别算法的运算量非常大,难以满足对识别算法高实时性要求,因此需要对DTW算法进行改进。根据搜索路径的单调性约束,路径中的任何一点的斜率总是大于零。为了防止无目的搜索,可以去除那些向 x 轴或 y 轴过分倾斜的路径,在匹配过程中限定弯折的斜率如图2所示。

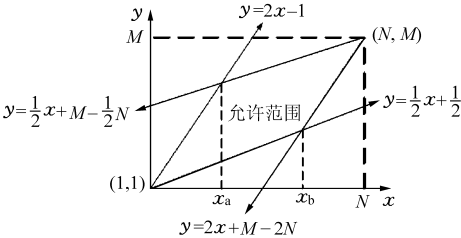


图2 匹配路径约束方法
Fig.2 Matching path restraint method

图2中平行四边形包括的允许范围之外的格点对应的帧匹配距离就不再需要计算,也不用保存所有的帧匹配距离矩阵和累计距离矩阵,充分利用这两个特点可以减少计算量和存储空间。当 $x_a < x_b$ 时,把动态弯折分成3段: $(1,x_a)$ 、 (x_a+1,x_b) 和 (x_b+1,N) ,其中 x_a 和 x_b 都取最近的整数,由此也得出对 M 和 N 的限制条件

$$\begin{cases} x_a = \frac{1}{3}(2M - N + 2) \\ x_b = \frac{2}{3}\left(2N - M + \frac{1}{2}\right) \end{cases}$$

(4)

$$\begin{cases} 2M - N \geq 3 \\ 2N - M \geq 2 \end{cases}$$

(5)

当不满足式(5)时,认为两者差距太大,不进行动态弯折匹配。此时, x 轴上的每一帧不再需要与 y 轴上的每一帧进行比较,而只是与 y 轴的 $[y_{\min}, y_{\max}]$ 区间内的帧进行比较, $y_{\min}$ 和 $y_{\max}$ 的计算公式为

$$y_{\min} = \begin{cases} \frac{1}{2}x + \frac{1}{2} & (1 \leq x \leq x_b) \\ 2x + M - 2N & (x_b < x \leq N) \end{cases}$$

(6)

$$y_{\max} = \begin{cases} 2x - 1 & (1 \leq x \leq x_a) \\ \frac{1}{2}x + M - \frac{1}{2}N & (x_a < x \leq N) \end{cases} \quad (7)$$

当 $x_a \geq x_b$ 时,此时弯折匹配的 3 段为 $(1, x_b)$ 、 $(x_b + 1, x_a)$ 和 $(x_a + 1, N)$ 。求取 $y_{\min}$ 和 $y_{\max}$ 的计算公式为

$$y_{\min} = \begin{cases} \frac{1}{2}x + \frac{1}{2} & (1 \leq x \leq x_a) \\ 2x + M - 2N & (x_a < x \leq N) \end{cases} \quad (8)$$

$$y_{\max} = \begin{cases} 2x - 1 & (1 \leq x \leq x_b) \\ \frac{1}{2}x + M - \frac{1}{2}N & (x_b < x \leq N) \end{cases} \quad (9)$$

搜索区域确定后,累积距离为

$$D[x, y] = d[x, y] + \min \begin{cases} D[x - 1, y] \\ D[x - 1, y - 1] \\ D[x - 1, y - 2] \end{cases} \quad (10)$$

对于 x 轴上每前进一帧,只需两个列矢量 D 和 d 分别保存前一列的累积距离并计算当前列的累积距离,无需保存整个距离矩阵,即按式(10)利用前一列的累积距离 D 和当前列的所有帧匹配距离 $d[x, y]$,求出当前帧的累积距离,保存于矢量 d 中,再把新的距离 d 赋值给 D ,作为新的累积距离,供下一列使用如图 3 所示,直到 x 轴上的最后一列,矢量 D 的第 M 个元素即为两个模板动态弯折的匹配距离。语音识别试验表明,本文设计的改进 DTW 算法相对于普通 DTW 算法的识别精度虽然会有 5% 左右的降低,但其运算时间可缩短至原 DTW 算法的 30% ~ 40%,能够满足对语音实时识别的要求。

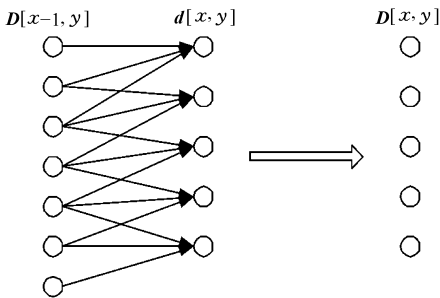


图 3 累积距离矢量的动态迭代示意图

Fig. 3 Dynamic iteration diagram of cumulative distance vector

2 硬件设计

开发的车载语音识别系统以 SPCE061A 单片机作为微控制器,外围电路包括键盘和 MIC 输入模块、语音与驱动信号输出模块、LED 显示模块、并行通讯模块和稳压电源模块,其硬件结构框图如图 4 所示,开发出的语音识别系统电控单元(ECU)如图 5 所示。车辆蓄电池 12 V 经稳压后为系统供电。操作者通过键盘输入指令使系统进入训练模式,微

控制器利用操作者的训练语句完成系统的语音识别训练后,由输出模块发出语音信号,提示操作者发出正式的语音指令,即可进入语音识别模式。

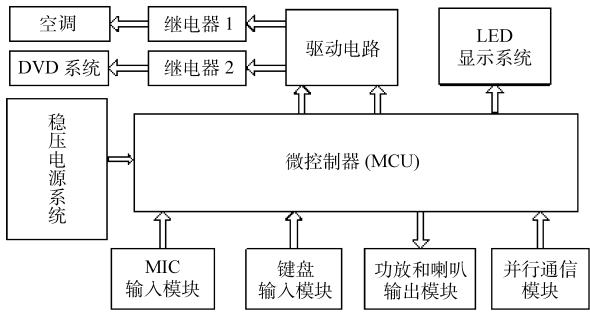


图 4 车载语音识别系统硬件结构示意图

Fig. 4 Hardware of the vehicular speech recognition system

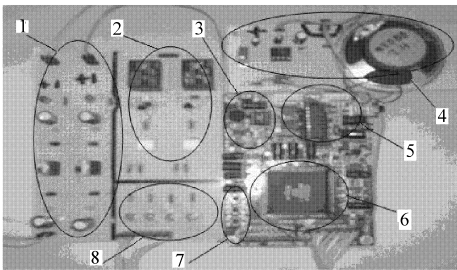


图 5 车载语音识别系统 ECU

Fig. 5 ECU for vehicular speech recognition system

- 1. 稳压电源 2. 驱动控制系统 3. MIC 音频输入模块 4. 音频输出模块 5. 并行通信模块 6. 微控制器 7. 键盘输入模块 8. LED 显示系统

在语音识别模式下,MIC 输入模块接收语音,输入微控制器,利用语音识别算法完成语音识别,并根据识别结果发出高低电平信号控制继电器的吸合与释放以控制车内空调、音频或视频等器件工作或停止,同时还可利用功放和喇叭输出模块、LED 显示模块同步输出识别结果。

3 软件设计

软件主要包括驱动程序和语音识别算法程序两部分。其中语音识别算法程序是系统的核心,由语音训练和识别程序组成。语音训练程序完成语音参考模板的生成,最终形成模板库并存储于单片机 FLASH 中,语音训练程序流程如图 6 所示。语音识别程序完成待识别语音信号与模板库中的参考模板匹配,将得到的最终识别结果输出。

(1) 车内噪声干扰滤波

根据车内噪声频谱特性和人类语音频谱特性不同的特点,对车内噪声进行抗干扰滤波。大量的车内噪声试验表明,车内怠速噪声小于 50 dB(A),城市工况车速小于 70 km/h 车内噪声小于 70 dB(A),高速行驶工况 70 ~ 120 km/h 车内噪声小于 80 dB(A)。

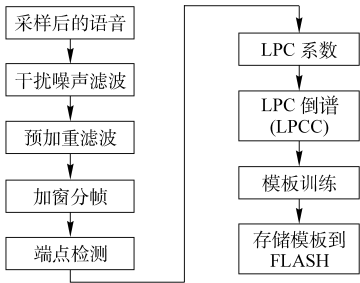


图 6 语音训练程序流程

Fig. 6 Flowchart of the speech training software

车内噪声呈窄带随机噪声特性,其频率范围为 20 Hz ~ 20 kHz,主要噪声能量集中在 20 ~ 500 Hz 的低频范围。而人类语音信号的频谱分布在 300 ~ 3 400 Hz 的范围内,且呈现共振峰结构。其中对语音感知、语意识别作用更为重要的第 2 共振峰大部分位于 1 kHz 以上,因此本文设计 1 个二阶 II R 数字滤波器,在语音识别时滤除车内 500 Hz 以内的低频噪声干扰信号。该滤波器采用的差分方程表示为

$$y(j) = \sum_{r=0}^M b_r x(j-r) + \sum_{k=1}^M a_k y(j-k) =$$
$$0.309\,85x(j) - 0.619\,7x(j-1) + 0.309\,85x(j-2) +$$
$$0.628\,85y(j-1) - 0.571\,4y(j-2) \quad (11)$$

(2) 预加重滤波

为使信号频谱变得平坦便于频谱分析或声道参数分析,采用数字滤波器

$$H(z) = 1 - az^{-1} \quad (0.9 \leq a \leq 1.0) \quad (12)$$

实现预加重。

式中 a ——预加重系数,设计中取值 0.96

(3) 加窗分帧

系统采样频率设定 8 kHz、帧长 32 ms、帧移 $\frac{1}{2}$,

采用汉明窗,以减小帧起始和结束处的信号不连续性。

(4) 端点检测

采用双级门限端点检测算法如图 7 所示。

该方法根据语音中浊音的能量较大,设定一个较高的门限 a_{mp2} ,用来确定语音已经开始(AB 段)。再取一个比 a_{mp2} 稍低的门限 a_{mp1} ,并从 A 点向左、B 点向右进行搜索,以确定与门限 a_{mp1} 相交的两个点 C 和 D。还用另一个过零率门限 z_{rc1} 来确定语音的真正起止端点。只要 z_{rc1} 取得合适,根据背景噪声的过零率值明显低于语音的过零率值的特点,沿一级判断之后的语音帧分别向前和向后搜索,就可以较精确判断出语音的端点 E 和 F。

(5) 线性预测倒谱系数和线性预测系数的提取
线性预测倒谱系数(LPCC)主要优点是彻底去掉了语音产生过程中的激励信息,主要反映声道响应,

而且只需十几个倒谱系数就能较好地描述语音的共振峰特性,在语音识别应用中有良好的效果。本文用 LPCC 作为语音特征参数,LPCC 参数选为 12 阶。
LPC 系数通过求解方程组 $\phi_n(i,0) = \sum_{k=1}^p a_k \phi_n(i,k)$, $i=1,2,\dots,p$ 获得,式中 ϕ_n 为语音抽样 $s(n)$ 的自相关函数,通过 LPC 系数即可得到 LPCC 系数^[1]。

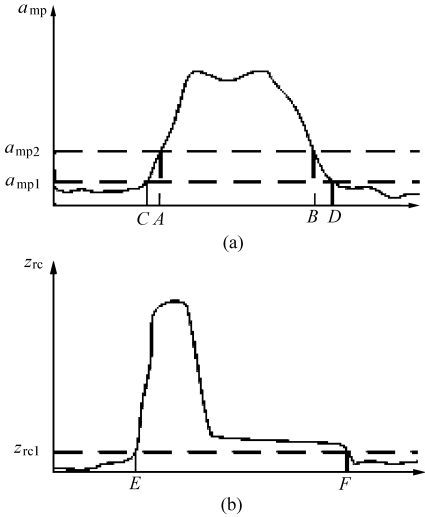


图 7 双级门限端点检测示意图

Fig. 7 Detection of the two-level threshold end point

(a) 短时能量 (b) 过零率

语音模型训练完成后,即可进入语音识别模式。以改进动态时间归整(DTW)算法为核心的语音识别流程如图 8 所示。

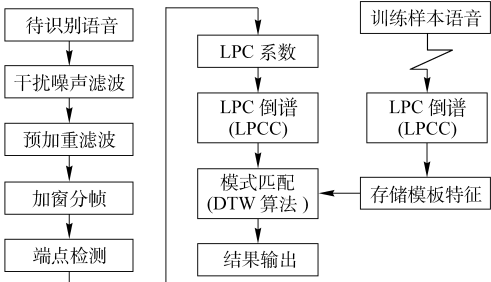


图 8 语音识别程序流程图

Fig. 8 Flowchart of the speech recognition software

语音识别时,确定一个可以衡量测试模板和参考模板相似度的测度函数,应用一种最佳的测度准则作为识别决策,对待识别的语音做出最后判决,获得识别输出结果。

4 语音识别试验与结果分析

为了得到真实的车内噪声环境,语音识别试验分别在车辆定置怠速车内噪声小于 50 dB(A) 和油门半开、车内噪声在 60 ~ 70 dB(A) 的环境下进行,以 2 字孤立词“开机”、“关机”、“打开”、“关闭”和 4 字孤立短语“打开音响”、“关闭音响”、“打开空

调”、“关闭空调”作为语音命令,分别进行 50 次识别试验,记录试验结果并计算,相应的识别率和系统平均识别率如表 1、2 所示。

表 1 车载语音识别系统 2 字语音命令识别结果

Tab.1 Speech recognition results of vehicular speech recognition system for two Chinese characters

试验环境	语音命令	识别率 /%	平均识别率/%	系统识别率/%
车辆怠速工况 车内噪声 小于 50 dB(A)	打开	90	90	87.5
	关闭	92		
	开机	90		
	关机	88		
车辆定置油门 半开、车内噪声 60 ~ 70 dB(A)	打开	90	85	85
	关闭	80		
	开机	85		
	关机	85		

结果表明,设计开发的车载语音识别系统的语音识别率随着车内环境噪声级的增加而下降,随着语音字数的增加而降低;在车辆怠速车内噪声小于 50 dB(A)的环境中 2 字语音平均识别率为 90%、4 字语音的识别率为 85%;在车辆定置油门半开、车内噪声为 60 ~ 70 dB(A)的环境中,2 字语音平均识别率为 85%、4 字语音的识别率为 80%;取得了较好的识别效果。

参 考 文 献

1 Juang B H. The past, present, and future of speech processing[J]. IEEE Signal Processing Magazine, 1998, 37(12): 24 ~ 48.

2 俞铁城. 语音识别的发展现状[J]. 通讯世界,2005(2): 56 ~ 57.
Yu Tiecheng. State and development of speech recognition[J]. Telecom World, 2005(2): 56 ~ 57. (in Chinese)

3 黄湘松,赵春晖,陈立伟. 基于 CDHMM/SOFMNN 噪声背景下的语音识别方法[J]. 应用科技, 2005, 32(9):4 ~ 6.
Huang Xiangsong, Zhao Chunhui, Chen Liwei. Speech recognition system based on CDHMM/SOFMNN in noisy environment [J]. Applied Science and Technology, 2005, 32(9): 4 ~ 6. (in Chinese)

4 Li Weifeng, Takeda K, Itakura F. Adaptive log-spectral regression for in-car speech recognition using multiple distributed microphones[J]. IEEE Signal Processing Letters, 2005, 12(4): 340 ~ 343.

5 林遂芳,张海英,潘永湘. 基于 DTW 和 LVQ 网络混合模型的语音识别方法[J]. 系统仿真学报,2005,17(8):1 959 ~ 1 961.
Lin Suifang, Zhang Haiying, Pan Yongxiang. Speech recognition method based on hybrid model of dynamic time warping and learning vector quantization[J]. Journal of System Simulation, 2005, 17(8):1 959 ~ 1 961. (in Chinese)

表 2 车载语音识别系统 4 字语音命令识别结果

Tab.2 Speech recognition results of vehicular speech recognition system for four Chinese characters

试验环境	语音命令	识别率/%	平均识别率/%	系统识别率/%
车辆怠速工况 车内噪声 小于 50 dB(A)	打开音响	90	85	82.5
	关闭音响	85		
	打开空调	85		
	关闭空调	80		
车辆定置油门 半开、车内噪声 60 ~ 70 dB(A)	打开音响	85	80	80
	关闭音响	75		
	打开空调	80		
	关闭空调	80		

5 结 论

(1) 建立了用改进的动态时间归整(DTW)算法进行语音识别的方法。开发了车载语音识别系统的电控单元(ECU)及语音识别软件。

(2) 通过在车内不同噪声环境下进行了语音识别试验,结果表明设计开发的车载语音识别系统的语音识别率随着车内环境噪声的增加而下降,随着待识别语音字数的增加而降低,2 字语音的平均识别率为 87.5%,4 字语音的平均识别率为 82.5%,取得了较满意的语音识别效果。