

# 特定人语音识别技术在汽车控制上的应用

作者：张志刚 陈昌巨

**关键词：** 特定人语音识别 预处理 LPCC MPCC

## 1 引言

从 20 世纪 50 年代开始对语音识别的研究开始，经过几十年的发展已经达到一定的高度，有的已经从实验室走向市场，如一些玩具、某些部门密码语音输入等，随着 DSP 和专用集成电路技术的发展，快速傅立叶变换以及近来嵌入式操作系统的研究，使得特定人识别尤其是计算量小的特定人识别成为可能。因此，对特定人语音识别技术在汽车控制上的应用的研究是很有前途的。

## 2 特定人语音识别的方法

目前，常用的说话人识别方法有模板匹配法、统计建模法、联接主义法(即人工神经网络实现)。考虑到数据量、实时性以及识别率的问题，笔者采用基于矢量量化和隐马尔可夫模型(HMM)相结合的方法。

说话人识别的系统主要由语音特征矢量提取单元(前端处理)、训练单元、识别单元和后处理单元组成，其系统构成如图 1 所示。

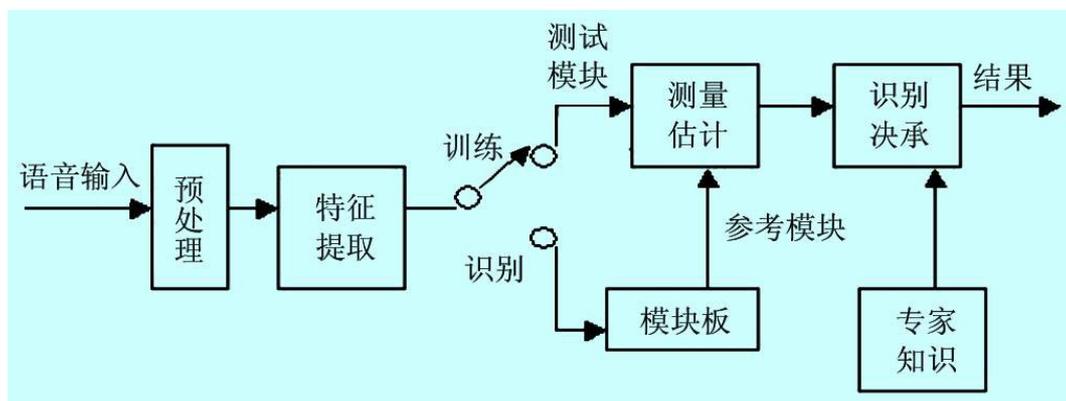


图 1 系统构成

由上图也可以看出，每个司机在购买车后必须将自己的语音输入系统，也就是训练过程，当然最好是在安静、次数达到一定的数目。从此在以后驾驶过程中就可以利用这个系统了。

所谓预处理是指对语音信号的特殊处理:预加重，分帧处理。预加重的目的是提升高频部分，使信号的频谱变得平坦，以便于进行频谱分析或声道参数分析。用具有 6d

B/倍频程的提升高频特性的预加重数字滤波器实现。虽然语音信号是非平稳时变的，但是可以认为是局部短时平稳。故语音信号分析常分段或分帧来处理。

## 2.1 语音特征矢量提取单元

说话人识别系统设计中的根本问题是如何从语音信号中提取表征人的基本特征。即语音特征矢量的提取是整个说话人识别系统的基础，对说话人识别的错误拒绝率和错误接受率有着极其重要的影响。同语音识别不同，说话人识别利用的是语音信号中的说话人信息，而不考虑语音中的字词意思，它强调说话人的个性。因此，单一的语音特征矢量很难提高识别率。该系统在说话人的识别中采用倒谱系数加基音周期参数，而在对控制命令的语音识别中仅采用倒谱系数。其中，常用的倒谱系数有 2 种，即 LPC(线性预测系数)和倒谱参数(LPCC)，一种是基于 Mel 刻度的 MFLL(频率倒谱系数)参数(Mel 频率谱系数)。

对于 LPCC 参数的提取，可先采用 Durbin 递推算法、格型算法或者 Schur 递推算法来求 LPC 系数，然后求 LPC 参数。设第 l 帧语音的 LPC 系数为  $\alpha_n$ ，则 LPCC 的参数

$$c_n = \alpha_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \alpha_k c_{n-k}$$

$$1 < n \leq p$$

其中  $p$  为 LPCC 系数的阶数， $k$  为 LPCC 系数的递推次数。

进一步的研究发现，引入一阶和二阶差分倒谱可以提高识别率。

对于 MPCC 参数的提取，若根据 Mel 曲线将语音信号频谱分为  $K$  个频带，每个频带的能量为  $\theta(M_k)$ ，则 MFCC 参数为

$$C_{m,n}(x) = \sum_{k=1}^K \theta(M_k) \cos\left(x(k-0.5)\frac{\pi}{K}\right)$$

$$1 < n \leq p$$

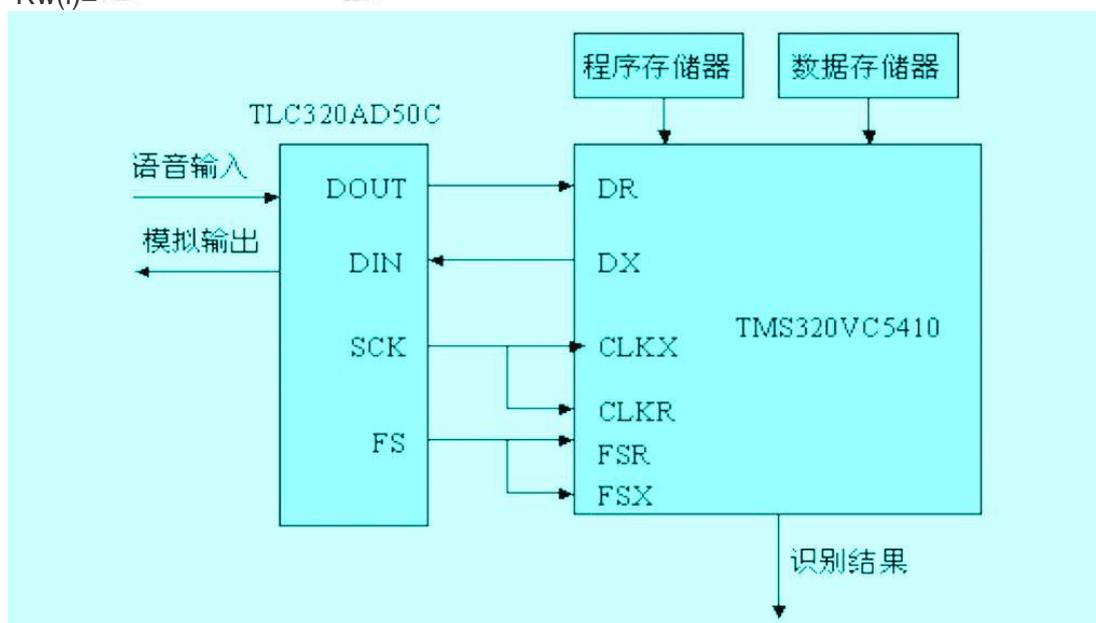
通过对 LPCC 和 MFCC 参数对识别率影响的实验比较，笔者选取 LPCC 参数及其一阶和二阶差分倒谱稀疏作为特征参数。

基音周期估计的方法很多，主要有基于求短时自相关函数的算法、基于求短时平均幅度差函数(AMDF)的算法、基于同态信号处理和线性预测编码的算法。笔者仅介绍基于求短时自相关函数的算法。

设  $Sw(n)$  是一段加窗语音信号，它的非零区间为  $0 < n \leq n-1$ 。 $Sw(n)$  的自相关函数称

为语音信号的  $S(n)$  的短时自相关函数，用  $Rw(l)$  表示，即

$$Rw(l) = \sum_{n=0}^{N-l} S_w(n) S_w(n+l) = \sum_{n=0}^{N-l} S_w(n) S_w(n+l)$$



可知短时自相关函数在  $Rw(0)$  处最大，且在基音周期的各个整数倍点上有很大的峰值，选择合适的窗函数(窗长为 40ms 的 Hamming 窗)与滤波器(带宽为 60~900Hz 的带通滤波器)后，只要找到自相关函数的第一最大峰值点的位置并计算它与零点的距离，便能估计出基音周期。

## 2.2 训练单元

训练单元的功能是把事先收集到的语音利用一定的算法为每一个待识别的说话人训练出与之相匹配的参数。针对说话人识别在汽车应用中的不同的要求，训练单元也分为 2 部分:对说话人识别的训练和对待识别词的训练。

对于说话人识别部分的训练，针对说话人的特征进行训练，为每个合法用户建立一套或多套 HMM 模型，同时采用基于矢量量化(VQ)的方法，为每个合法用户建立 VQ 码本。VQ 码本的设计采用 LBG 算法，初始码本的设置采用分裂法初始码本。

第 2 部分针对控制命令中用到的每个孤立的词条建立多个训练样本，或称为词条样本，估计出该词条的 HMM 参数(一套或多套)。对一个 HMM 过程的完整的描述包括: 2 个模型参数  $N$  和  $M$ ，3 组概率度量  $A$ ， $B$  和  $\pi$ 。为了方便起见，通常采用如下方式

表示一个完整的模型:  $\lambda = (N, M, \pi, A, B)$ , 或者简写为:  $\lambda = (\pi, A, B)$ 。而对于每一个词条  $V$  的模型参数  $\lambda = \{A, B\}$ ,  $V=1 \sim V$ , 可以用 Baum-Welch 重估算法。

### 2.3 识别单元

识别单元的功能是利用经训练已经获得的 HMM 模型参数 和测得的说话人的基音周期在一定的判决条件下辨认出待识别的说话人并估计出待识别的控制命令词串。针对 HMM 模型参数通常采用的判决条件是最大后验概率, 用 Viterbi 算法实现。

### 2.4 后处理单元

充分利用每个说话人的声道参数和词条中各状态持续时间的概率分布来改进系统的识别率。

## 3 系统的实现

由于汽车的控制命令是有限的词条和数字串的组合, 对这些语音命令的识别属于特定人小词汇量的连接词的识别以及与文本有关的说话人确认, 不论是从目前的 DSP 运算速度还是存储空间来说, 实时实现这些语音命令的识别都是完全可能的。

识别系统组成框图如图所示:在此系统中, 对运算能力和存储单元要求非常高的语音识别部分完全由 DSP 完成。

框图中识别系统的功能是完成语音的输入、A/D 转换及识别, 系统中核心部分采用 TMS320VC5410。其原因是它的运算速度和存储空间都能满足要求, 同时它的一些并行运算硬件结构也非常适合语音识别的各种算法, 程序和已经脱机训练好的 HMM 参数表及相应的词典存放在程序存储器中, 数据存储器存放识别过程中的中间计算数据。A/D 芯片采用 TLC320AD50C, 里面含有 A/D、D/A 以及低通滤波器和采样保持电路。模拟语音信号的输入主要是通过传声器, 保证语音门禁的安全性, 转换后的数字语音数据以同步串行通信方式传送给 DSP。如图 2。

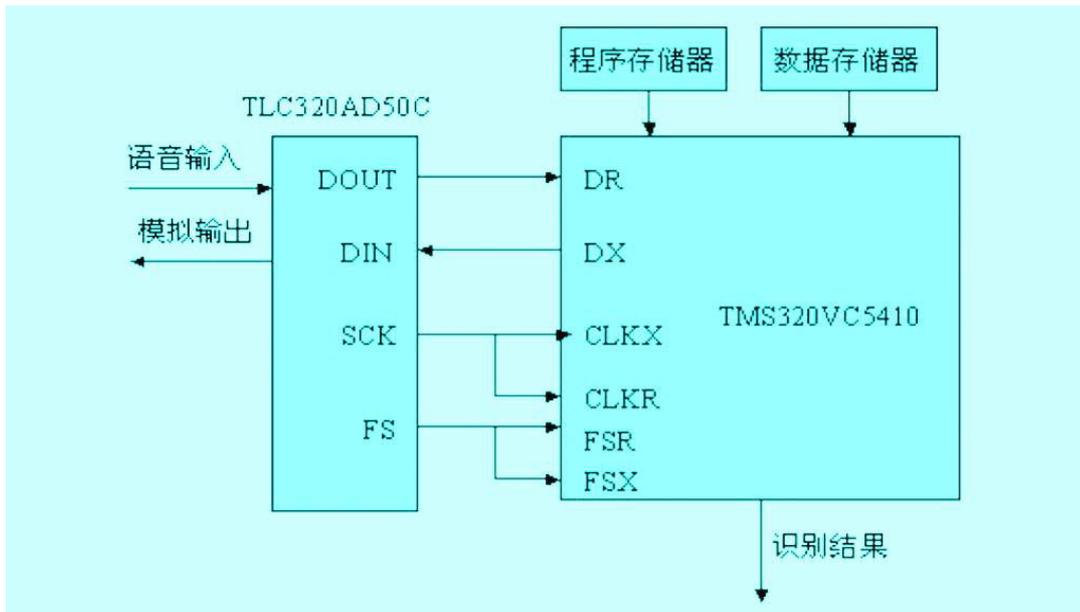


图 2 识别系统的组成框图

#### 4 结束语

语音控制汽车是未来的一种趋势。目前，将语音技术应用于汽车的产品只有在一些玩具中用到，由此可想利用语音技术进行汽车控制这一领域蕴涵着相当大的潜在市场。

而且，说话人识别技术已经发展到可以应用到实际的阶段了，但目前对说话人识别的应用并不是很多。笔者尝试提出一种比较容易实现的方案，将说话人识别技术应用到实际中。但在实际应用中，说话人识别系统都面临一个共同的问题，即无法区分一个发音是现场发音还是录音回放。针对该现象，笔者提出的说话人识别系统可以有效地防止这种情况发生。具体实现说话人识别系统时，可采用随机或其它方法来生成提示文本。如随机的数字串，以使假冒者无法事先录音，增加驾驶的安全性。