# 基于 DSP 的语音识别计算器设计

梁 俊，杨燕翔，王 娟，李海忠
（西华大学 电气信息学院 四川 成都 610039）

**摘要**：为解决特殊群体使用计算器困难的问题，设计了一种基于 TMS320VC5509 DSP 的可语音识别的计算器系统。该计算器系统的核心是采用 HMM 算法建立语音识别模型。通过对实时语音信号(数字、运算符号等语音)进行处理，将得到的参数与模板库参数进行匹配并加以识别，利用 TMS320VC5509 DSP 自带的计算模块实现语音信号整数 100 以内的加、减、乘、除等计算功能。实验结果表明，该计算器系统在低噪声场合和高噪声场合下识别率分别达到 94.73%和 76.55%。

**关键词**：语音识别；DSP；HMM；计算器；TMS320VC5509
**中图分类号**：TP391 **文献标识码**：A **文章编号**：1674-6236(2010)05-0135-04

## Design of speech recognition calculators based on DSP

LIANG Jun, YANG Yan-xiang, WANG Juan, LI Hai-zhong
(*School of Electrical and Information Engineering, Xihua University, Chengdu* 610039, *China*)

**Abstract:** In order to solve the problem of special groups to use the calculator difficultly, this paper described a calculator system of speech recognition based on TMS320VC5509 . The core of calculator system adopted the HMM algorithms to establish a model for speech recognition. Through real-time speech signals (numbers, operation symbols, etc.)processing, the parameters was matched with the template library and identified. Then the system took advantage of calculation module on the TMS320VC5509 DSP to realize 100 within the integer add, subtract, multiply, divide and other computing functions. The experimental results show that the recognition rate of the calculator system in the low-noise situations and high-noise situations reached 94.73% and 76.55% respectively, and achieved the purpose of design.

**Key words:** speech recognition; DSP; HMM; calculator; TMS320VC5509

随着电子技术的高速发展，现代普通民用计算器在保留基本的加减乘除等运算外，加入了大量如三角函数、幂函数等比较复杂的运算。但是其基本的操作没有发生变化，依然是运用手指操作，对于需要进行实时数字计算的一些特殊人群(残疾人士)或是在一些特殊场合在无法手动操作计算器的情况下，用加入了语音识别模块的计算器来进行实时数字计算就有相当的必要。

语音识别[2]技术是人机最自然、最简洁的交流方式，它就是让机器能够自动识别并理解说话人要表达的意思，将语音信号转变为正确的文本或者命令的高科技技术。根据实际的应用，语音识别可以分为：特定人与非特定人的识别、孤立词与连续词的识别、中小词汇量与无限词汇量的识别。

考虑到成本及使用范围因素，本文中应用的是基于 TMS320VC5509 DSP 的非特定人、孤立词、小词汇量的语音识别系统。通过实际测试，使用该 DSP 的语音识别系统有着较高的实时性、识别率，基于该系统的计算器对实时数字计算有较高准确性，基本能解决特殊群体和特殊地点使用计算器困难的情况。

## 1 系统硬件设计

### 1.1 语音识别系统

语音识别的基本原理框图如图 1 所示。语音识别过程主要包括语音信号前处理、特征提取、模式匹配等部分。语音信号输入之后，预处理和数字化是进行语音识别的前提条件。特征提取是进行语音信号训练和识别必不可少的步骤，本文采用的是提取每帧的 Mel 系数[3]的倒谱参数作为语音信号的特征值。模板匹配算法目前有 DTW 算法、HMM 隐马尔科夫模型、ANN 人工神经网络等。本文采用 HMM 隐马尔科夫模型的方法，提取出的特征值存入参考模式库中，用来匹配待识别语音信号的特征值。匹配计算是进行语音识别的核心部分，由待识别人的语音经过特征提取[2]后，与系统训练时产生的模板进行匹配，在说话人辨认中，取与待识别语音相似度

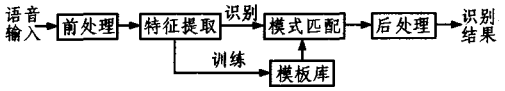


图 1 语音识别基本原理框图

**收稿日期**：2009-10-19 **稿件编号**：200910059
**作者简介**：梁 俊(1981—)，男，安徽宣城人，硕士研究生。研究方向：DSP 技术及应用。

最大的模型所对应的语音作为识别结果。

### 1.2 系统硬件结构

图 2 为系统硬件结构框图。此系统的核心器件是 TI 公司的 TMS320VC5509 定点 DSP。在本系统中，它不仅是语音识别的核心，还负责计算器的运算部分。TMS320VC5509[4]是系统的运算处理单元，具有 2 个乘法器（MAC），4 个累加器（ACC）；40 位、16 位的算术逻辑单元（ALU）各一个，这大大增强了 DSP 的运算能力；指令字长不只单一的 16 位，可扩展到最高 48 位，数据字长 16 位；可通过 USB 接口对TMS320VC5509 烧写程序而不必借助仿真器。正是基于这些优点，选择该器件可节省开发资金，减小电路板面积。DSP 与 TLV320AIC23 的接口电路如图 3 所示。

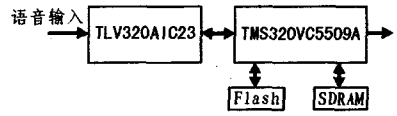


图 2 系统硬件结构框图

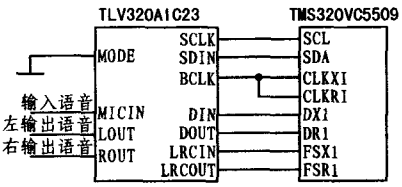


图 3 TLV320AIC23 与 TMS320VC5509 的接口电路

TLV320AIC23[5]是 TI 公司的一款低成本、低功耗的音频编解码器（CODEC），在本系统中负责采集语音信号。它与本系统相关的性能参数有：支持 8~96 kHz 可调采样率；可调 1~5 dB 的完整缓存放大系统等。图 4 是 TLV320AIC23 的电路图。

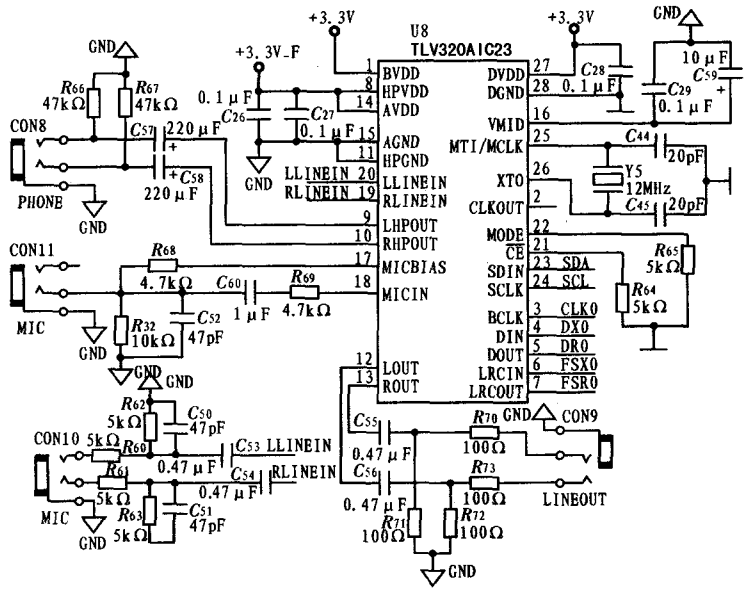


图 4 TLV320AIC23 电路

AM29LV800B 存储器又称闪存（Flash），它具有在线电擦写、低功耗、大容量等特点，其存储容量为 8 Mbit。上电后，DSP 从外部 Flash 加载并执行程序代码，使系统能够脱机运行。在本系统中，它主要用来存储程序代码、语音模型、以及压缩后的语音数据。

HY57V641620 同步动态存储器（SDRAM），容量为 4 M×16 bit。作为 RAM 的扩展，它大大增强了 DSP 的存储与运算能力。在系统初始化的时候，用来装载放在 Flash 中的声学模型。这样在语音识别的过程中可以通过片外的 SDRAM 来访问声学模型，比直接访问 Flash 来获取声学模型数据要快。LCD 显示器用来实时显示经过语音识别后的数字、运算符号，并在得到需要显示最终结果的提示后显示答案。

## 2 系统软件设计

### 2.1 系统软件流程

图 5 为系统的软件流程。整个系统开始运行后，初始化 DSP 及 TLV320AIC23，以使各个寄存器的初值符合要求。在系统通过 TLV320AIC23 采集语音信号后，首先要进行预滤波和预加重；接着将语音信号进行分帧；然后计算每帧信号的短时能量与短时平均过零率，为接下来的门限判决提供依据；利用门限判决进行端点检测后，提取每帧的 Mel 倒谱参

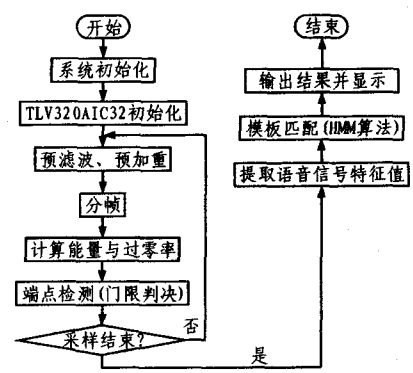


图 5 系统软件流程

数(MFCC),作为该帧信号的特征值;最后,用处理后的语音信号的特征值与模板进行匹配,这一部分是系统的重点。以相似度最大的模板锁对应的语音信号为识别结果。根据识别的结果在显示器上显示数字和运算符号,由运算规则得出结果并显示。

### 2.2 前处理

前处理是对语音信号采样、A/D转换、预滤波和预加重、分帧等。以8 kHz和16位的采样频率采集的语音模拟信号。本系统使用带通滤波器来滤波,上截频率为3.4 kHz,下截频率为60 Hz。由于语音信号具有极强的相关性,因此,分帧时要考虑帧重复的问题。本文将语音信号以256个采样点为一帧,两顿之间的重复点数为80,通过一个一阶的滤波器 $H(z)=1-a/z$ 对采集的信号进行处理。

端点检测就是从说话人的语音命令中,检测出孤立词的语音开始和结束的始点。端点检测是语音识别过程的一个重要环节,只有将孤立词从说话人的背景噪声中分割出来,才能够进一步进行语音识别工作。本文采用短时能量和过零率检测端点。语音信号的短时能量分析给出了反应其幅度变化的一个合适描述方法。

短时过零率,即指每帧内信号通过零值的次数,能够在一定程度上反映信号的频谱特性。一帧语音信号内短时平均过零率[4]定义为:

$$Z_n=\sum_{m=n}^{n+N-1}\left|\mathrm{sgn}[x_w(m)]-\mathrm{sgn}[x_w(m-1)]\right|w(n-m) \qquad (1)$$

用短时能量参数检测结束点,信号$\{x(n)\}$的短时能量定义为:

$$E_n=\sum_{m=-\infty}^{\infty}[x(m)w(n-m)]^2 \qquad (2)$$

式中,$\{x(n)\}$为输入信号序列。

在正式端点检测开始后,短时能量与短时过零率作为门限来判决说话人命令字的开始与结束:连续5帧语音信号超过门限值视为说话人命令字的开始,连续8帧语音信号低于门限值视为说话人命令字的结束。

### 2.3 特征值提取

提取每帧的Mel倒谱参数(MFCC)[6]为该帧信号的特征值。由倒谱特征是用于说话人个性特征和说话人识别的最有效的特征之一,它是基于人耳模型而提出的。其提取过程如下:

1)原始语音信号 $S(n)$ 经过预加重、加窗等处理,得到每个语音帧的时域信号 $x(n)$。然后经过离散傅里叶变换(DFT)后得到离散频谱 $X(k)$。

$$x_n(k)\sum_{n=0}^{N-1}x(n)e^{-j2\pi nk/N} \qquad 0\leqslant k\leqslant N \qquad (3)$$

式中,$N$ 表示傅里叶变换的点数。

2)将离散谱 $X(k)$ 通过 $M$ 个Mel频率滤波器组可得到Mel频谱并通过对数能量的处理,得到对数频谱 $S(n)$。计算 $S(n)$ 通过每一个滤波器的输出,得到 $M$ 个 $h(m)$ 参数。

$$h(m)=\sum_{k=f(m-1)}^{f(m+1)}W_m(k)S(k) \quad m=1,2,\cdots,M \qquad (4)$$

3)对所有滤波器输出进行对数运算,再进一步进行离散余弦变换(DCT),即可得到MFCC参数。

$$C_{mfcc}(i)=\sqrt{\frac{2}{N}}\sum_{m=1}^{M}\ln h(m)\cos\left\{(m-\frac{1}{2})\frac{i\pi}{M}\right\} \qquad (5)$$

一般在Mel滤波器的选择中,Mel滤波器组都选择三角形的滤波器,但也可以是其他形状,如正弦形的滤波器组等。

### 2.4 模板匹配(HMM算法)

本文采用隐马尔科夫模型(HMM算法)[5]进行模式匹配。它将特征矢量作为模板,在语音识别模式匹配时,对输人的语音与模板库中的模板进行比较,最后将相似度最高的作为输出结果。HMM算法解决由于说话人语速不同和连续说话的而带来的失真问题,还能大大减少运算时间,提高识别率。

隐马尔可夫模型是一个双重随机过程的统计模型,其基本随机过程是隐藏起来观测不到的,另一个随机过程则产生观测序列。对于语音识别系统,观测序列 $O$ 就是矢量量化后的结果序列,模型 $\lambda$ 就是由训练语音得到的模板。语音的训练过程就是产生模板 $\lambda$ 的过程,而语音的识别过程就是求出在模板 $\lambda$ 下,待识别语音的结果序列 $O$ 的条件概率 $P[O/\lambda]$。由 $\alpha_t(i)$ 和 $\beta_t(i)$ 的定义可直接得到:$P[O/\lambda]=\alpha_t(i)\beta_t(i)$。而语音的训练算法则较复杂,目前都采用迭代的方法得到 $a$ 和 $b$ 的近似解,其迭代公式[4]如:

$$\hat{a}_{ij}=\frac{\sum_{t=1}^{T-1}a_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N}a_t(i)a_{ij}b_j(Q_{t+1})\beta_{t+1}(j)} \qquad (6)$$

$$\hat{b}_j(k)=\frac{\sum_{t=1}^{T-1}a_t(i)\beta_t(i)}{\sum_{t=1}^{T-1}\sum_{j=1}^{N}a_t(i)a_{ij}b_j(Q_{t+1})\beta_{t+1}(j)} \qquad (7)$$

在实际应用中,仅对词条的少数次发音进行训练的语音识别系统,不可能对各种复杂语境下的不同发音都有较高的识别率。某些较陈旧的识别算法如动态时间弯曲法,只能把单词的多次训练发音形成多个模板,造成模板数量成倍增加,影响系统的实时性。而HMM能够对一个词的多个训练序列进行有效的融合而形成一个模板。当训练发音的数量增多时,只会造成训练过程的计算量增大,而不会使识别过程的计算量有丝毫增加,这对系统的实时性是相当有利的。

## 3 系统测试

针对计算器的使用特点和环境,分别在2个地点测试系统的性能。1)封闭的实验室(地点1),噪声较小,采集的信号较为良好,缺点是有回声。2)课间休息的教室(地点2),噪声及大,干扰很强,信号的采集质量很差。

因为整个系统的设计是实现计算器的计算功能,因此本次的实验是在系统识别数字和运算符号等语音后在显示器上显示数学运算公式,并在识别出"等于"或"得出"2个词组的

语音后显示出"="和最后的计算结果。

在测试前预先采集 5 男 5 女共 1 000 个语音样本（内容为数字 0 到 100，加、减、乘、除、等于和十、百、千、万等基本计算所需要的数字和运算符号读音），并且对所有样本进行训练。另外找 10 人(4 女，6 男)在各实验地点进行实时测试，每人 10 个，共 100 个未经训练的样本。用这些样本对系统进行测试，其测试结果如表 1 所示。

**表 1 测试结果**

| 内容 | 参加训练样本 | 未训练样本 | 合计 |
|---|---|---|---|
| 样本数 | 1 000 | 100 | 1 100 |
| 地点 1 正确识别率 | 95.4% | 88% | 94.73% |
| 地点 2 正确识别率 | 77.5% | 67% | 76.55% |

由表 1 所示，在相同的实验设备条件下，在噪声较小的环境下的系统识别率要远高于在嘈杂的环境下。特别是非经训练的样本在嘈杂环境下的识别率比较低，主要是因为环境中的噪声相当复杂，查看频谱图发现噪声几乎与说话人语音混杂叠加，算法难以识别。

## 4 结 论

本文设计的语音识别计算器系统，除兼有语音识别的功能，还能对识别的语音信号做进一步处理。由于采用 HMM 模型对语音信号进行端点检测，大大提高语音信号起止点判断准确性，提高了识别的准确率。由于系统运算复杂，计算量和存储量都很大，同时也需要实时处理语音信号与算法，系统所采用的 TMS320VC5509，由于其具有 0.05 MW/MIPS 的功耗，800 MIPS 的运算能力等优越的性能，完全能够满足实时识别功能。实验表明，该计算器系统处理速度快，运行稳定，达到了设计要求。

**参考文献：**

[1] Texas Instruments.TMS320VC5509A fixed-point digital signal processor[EB/OL]. 2006.http://www.ti.com.

[2] 蔡莲红，黄德智，蔡锐.现代语音技术基础与应用[M]. 北京：清华大学出版社，2003.

[3] 韩纪庆，张磊，郑铁然.语音信号处理[M].北京：清华大学出版社，2004.

[4] 程正兴.小波分析算法与应用[M].西安：西安交通大学出版社，2003.

[5] 张雄伟，陈亮，徐光辉.DSP 器件的原理与开发应用[M].北京：电子工业出版社，2004.

[6] 郭春霞，裘雪红.基于 MFCC 的说话人识别系统[J]. 电子科技，2005(11)：53- 56.