

# 一种改进的 DHMM 语音识别算法的 DSP 实现<sup>①</sup>

王维强

(中国空空导弹研究院, 洛阳 471009)

**摘要:** 对多种隐马尔可夫模型算法进行了分析对比, 在此基础上设计了一种改进的离散隐马尔可夫模型 (DHMM) 算法, 并将此算法成功运用到了 DSP 嵌入式语音识别系统中. 该系统采用上述算法完成了对非特定人的孤立词语音识别. 该系统以 ADSP-BF531 语音处理专用 DSP 为核心, 并辅以大容量的 SDRAM 和 ROM 为扩展, 具有小型、高速、可靠、鲁棒性好、扩展性强等多个优点; 可应用于许多特定场合, 有很好的市场前景. 试验结果表明, 该系统对非特定人的孤立词的综合识别率在 94% 以上. 对该系统应用的改进的 DHMM 算法, 硬件的实现过程以及其实际应用效果等进行了详细阐述.

**关键词:** 语音识别; DHMM; 特征图案; 矢量量化; 非特定人; 孤立词

## DSP Implementation of an Improved DHMM Algorithm

WANG Wei-Qiang

(AVIC China Airborne Missile Academy, Luoyang 471009, China)

**Abstract:** This paper analysed and contrasted multi-algorithms about DHMM (Discrete Hidden Markov Models), based on which it designed an improved DHMM algorithm, and then, applied it to the embedded speech recognition system. This system was realization of speaker-independent's isolated word recognition. The platform of this system was centered ADSP-BF531. And it was expanded by big capacity SDRAM & ROM. This system has many merits, e.g. small volume, high speed and high reliability, high robustness and high expansibility, etc. It is very convenient for some special situations and has much market potential. The experiment confirms that its speech recognition accuracy reaches 94 percent for speaker-independent and small vocabulary. This paper introduced the improved HMM algorithm, hardware design and its application effect.

**Key words:** speech recognition; DHMM; characteristic pattern; vector quantization; speaker-independent; isolated word

目前, HMM(隐马尔可夫模型)算法是语音识别领域内使用最为广泛的技术, HMM 有 DHMM(离散 HMM)、CHMM(连续 HMM)和 SCHMM(半连续 HMM) 等几种. 以上三种算法各有其特点: DHMM 的存储和计算量较小, 但其矢量量化的过程会造成性能的损失; CHMM 采用连续概率密度函数来描述观测矢量, 具有精度高的特点, 但其计算量很大; SCHMM 介于以上两者之间, 可视为 CHMM 中概率密度函数的聚类或 DHMM 中矢量量化的细化.

笔者在以上基础上借鉴并利用方绍武等人的算法思想<sup>[1]</sup>, 设计了一种改进的 DHMM 算法, 并将其应用在了嵌入式语音识别系统上. 该系统以 ADSP-BF531 为核心构建, 实现了对非特定人、孤立词的语音识别, 改进的 DHMM 算法具有计算量小、计算精度高等优点.

## 1 语音识别算法设计

### 1.1 传统 DHMM 算法

传统的 DHMM 语音识别的原理框图如图 1 所示.

<sup>①</sup> 收稿时间:2012-02-03;收到修改稿时间:2012-02-25

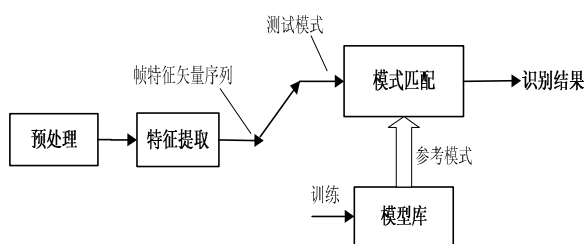


图 1 语音识别原理框图

①预处理: 对声源信号进行抗混叠滤波、A/D 转换、预加重及端点检测等内容,以获得比较理想的处理信息对象。

②特征提取: 从语音波形中提取出随时间变化的能够反映原始语音特征的矢量序列。

③语音训练: 建立声学模型,将获取的语音特征通过必要学习算法产生。

④模式匹配: 在识别时将输入的语音特征同声学模型进行比较,得到识别结果<sup>[2]</sup>。

在训练阶段,用户将词汇表中的词依次读一遍,并且将其特征矢量序列存入模板库中。在识别阶段,将输入语音的特征矢量序列依次与模板库中的每一个模板进行形似度比较,相似度最高者作为识别结果输出。

在 HMM 算法中,语音序列被看做马尔可夫随机过程的输出。假定识别系统的词汇表共包括 V 个词条,那么在训练阶段需要请很多个说话人分别将这次词条说一遍并存入数据库中。利用这些训练数据可以为每一个词条建立一套 HMM 参数  $\lambda_v$ , ( $1 \leq v \leq V$ )。

在识别时,对于每个待识别语音,可以得到一个观察矢量序列  $Y=[y_1, y_2, \dots, y_N]$ , 其中 N 为输入语音所包含的帧数。语音识别的过程就是计算每个 HMM 模型  $\lambda_v$  产生 Y 的概率  $P(Y | \lambda_v)$ , 并使得该概率达到最大的 HMM 模型,那么该模型所对应的词条即为孤立词识别的结果<sup>[3]</sup>, 即:

$$\hat{v} = \arg \max_{v=1,2,\dots,V} [P(Y | \lambda_v)] \quad (1)$$

### 1.2 改进的 DHMM 算法

传统的 DHMM 算法中,矢量量化的作用是将连续的输入参数矢量化为码本标号,每个说话人的码本独立进行训练。如利用矢量量化对全部说话人的数据进行技术综合,生成含有各个参与训练者的语音特征

图案,并用生成的特征图案来代替原始的矢量码本。

这样,原来的“码矢量”将只是这个图案的一个子集,而对应不同说话者的子集之间所含码本标号及标号的顺序均有所不同<sup>[1]</sup>。语音识别时,经分类处理后的待识别语音的码本标号与模型库进行模式匹配,即:用 Viterbi 算法进行评分,从而得到识别结果。

改进的 DHMM 训练算法流程框图如图 2 所示。

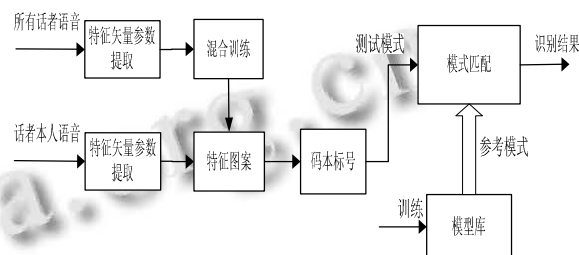


图 2 改进的 DHMM 算法流程框图

### 1.3 其他部分算法说明

对本系统中在预处理、特征提取、HMM 模型库训练以及模式匹配等模块中应用的算法分别进行说明。

#### 1.3.1 预处理和特征提取

首先采用预加重、汉明窗、双门限法等完成对语音信号的预处理;然后使用 Mel 倒谱参数(MFCC)进行特征识别, MFCC 参数提取的过程如图 3 所示,其中 Mel 滤波器组的作用是利用人耳听觉特性对语音信号的幅度平方谱进行平滑。对数操作的用途:压缩语音谱的动态范围;考虑乘性噪声,将频域中的乘性成分转换成加性成分。离散余弦变化主要是用来对不同频段的频谱成份进行解相关处理,使得各维向量之间相互独立。

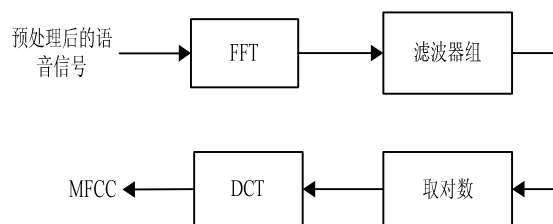


图 3 MFCC 参数求解过程

#### 1.3.2 HMM 模型训练过程

本系统选择无跨越的从左向右的 HMM 模型,状态数取 6。训练过程中利用 Baum-Welch 算法和 Viterbi 算法来计算所有观察序列的输出概率,然后对其进行累加,得到总输出概率。采用前后 2 次的输出概率的

相对变化小于一定阈值(如:  $1 \times 10^{-4}$ )或超过迭代次数作为训练结束的判据。

程序流程图如图 4 所示。

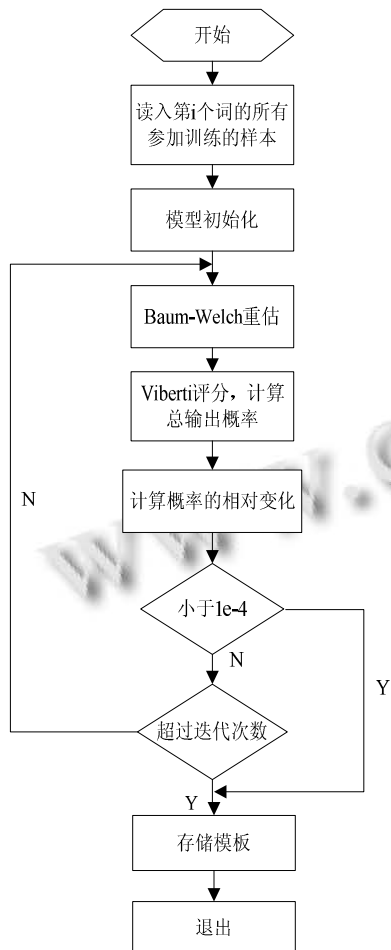


图 4 HMM 模型建立及训练流程

### 1.3.3 模式匹配

语音识别的过程即是用 Viterbi 算法将经将输入的矢量量化后的语音与模型库中的参考模板进行匹配。

Viterbi 算法是一种前向搜索算法,其可以是在给定相应的观察序列时,找出从模型  $\lambda$  中找出的最佳状态序列,即选择输出概率最大的模版作为输出结果。对数形式的 Viterbi 算法,能够避免大量的乘法运算,减少计算量,同时还可以保证有很高的动态范围,不会出现由于过多的连乘而导致溢出问题<sup>[4]</sup>。

## 2 DSP硬件设计

该系统电路主要由 DSP、音频编码器、CPLD、片外 SDRAM、FLASH 和 EEPROM 存储器以及电源、

时钟等组成。硬件接口如图 5 所示。

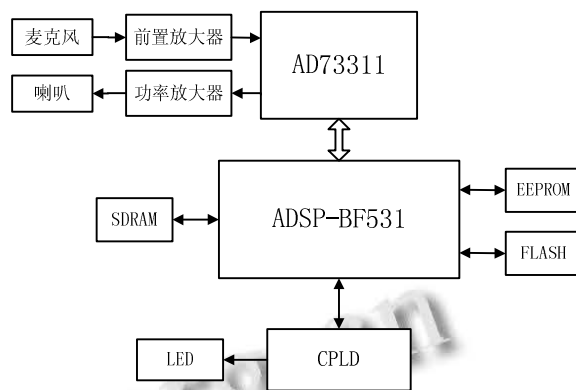


图 5 系统硬件结构图

各器件的主要功能如下:

①AD73311 将经前置放大后的麦克风音频信号经 A/D 转换后通过串行端口输入 BF531,同时完成对 BF531 输出的数字音频信号的 D/A 转换,而后输出到功放和喇叭;

②BF531 作为该系统的核心,完成对信号的特征提取和语音识别,同时对其外围的器件进行控制管理;

③CPLD 对 DSP 的外围的器件时序和数据流程进行控制,以及对 LCD 显示屏初始化检测设置;

④由于 DSP 片内的 RAM 有限,配置了一块 SDRAM 用于扩展系统的内存,以满足程序运行时数据和指令存储的要求;

⑤EEPROM 用于存放 DSP 程序代码和系统初始化所需的数据;

⑥FLASH 用于存放训练样本库。

### 2.1 AD73311 与 BF531 接口设计

本系统中设计 AD73311 的采样率为 16KHz<sup>[5]</sup>,其与 BF531 的串口 0(SPORT0)连接,通过 DMA 方式在单时钟周期内完成操作,每个 SPORT 都可以自动发送和接收多个缓冲区的数据信息。DSP 能够链接或串接 SPORT 和存储器之间的多个 DMA 序列<sup>[6]</sup>(BF531 有 12 通道的 DMA 控制器)。

BF531 支持 32bit 的串口数据传输,由于 AD73311 为 16bit 的音频器件,而且 16bit 已可满足系统精度要求,因此本系统只使用了 BF531 的主传输数据通道,即: DTOPRI 和 DROPRI,而将第二传输数据通道 DTOSEC 和 DROSEC 进行了屏蔽<sup>[7]</sup>。AD73311 与 BF531 的连接方式如图 6 所示。

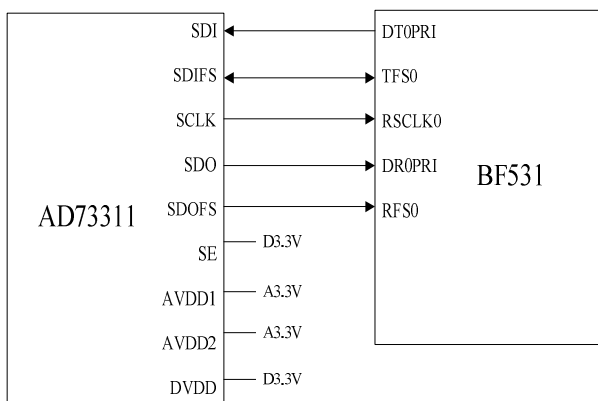


图 6 AD73311 与 BF531 连线图

## 2.2 SDRAM 接口设计

SDRAM 主要用于扩展系统内存, 这里采用了三星电子的 K4M56163PG 的移动式 SDRAM 存储器, 其容量为  $4\text{M} \times 16\text{Bit} \times 4(32\text{MB})$ <sup>[8]</sup>, 满足系统运算的要求, 其与 DSP 的 I/O 总线连接<sup>[9]</sup>, 如图 7 所示。

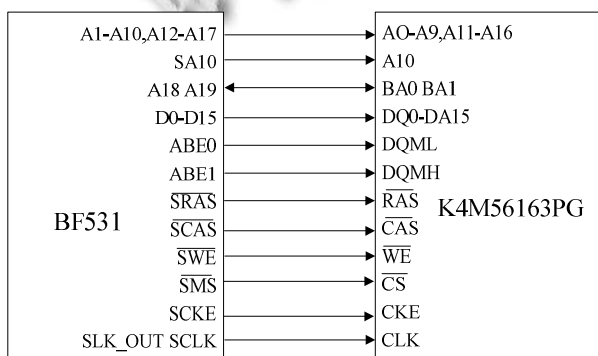


图 7 SDRAM 与 BF531 连线图

## 2.3 EEPROM 接口设计

EEPROM 主要用于存放程序代码, 选用 Atmel 公司的 AT45DB161D-TU 型的 EEPROM, 该型存储器采用 SPI 接口, 容量为  $16\text{Mb}(2\text{MB})$ <sup>[10]</sup>, 可以满足程序存储的要求, 其与 DSP 的 SPI 端口连接。

通过设置 EEPROM 存储器的 SPI 主模式启动(即设置 BMODE=11), 现实配置 BF531 为连接一个 SPI 存储器的主设备和存储器的加载。

需要注意的是, 该加载模式需要在 MISO 加上拉电阻。否则, BF531 将从 MISO 引脚读取到 0xFF(即 SPI 存储器没有写任何数据到 MISO 引脚), 不是需要的响应值<sup>[11]</sup>。

## 2.4 FLASH 接口设计

片外 FLASH 主要用于存训练样本库, 本系统采用的 NAND FLASH 为三星电子的 K9F8G08U0M-PIB0,

该 FLASH 为工业级 SLC 架构(Single Layer Cell, 单层单元)芯片, 具有速度快、可靠性高等特点, 而且容量为  $8\text{Gb}$ (即  $1\text{GB}$ )<sup>[12]</sup>, 可以满足存储大量样本数据的要求。其采用 EBIU(External Bus Interface Unit, 外部数据总线)和单个 GPIO(General Purpose Input Output, 通用输入/输出)引脚实现与 DSP 数据通讯<sup>[13]</sup>。

## 3 试验结果分析

经过试验验证, 该改进的 DHMM 算法的识别率较传统的 DHMM 高, 与 CHMM 相当。另外, 传统的 DHMM 算法对训练样本量的依赖性较强, 但这种改进的 DHMM 算法对样本量具有较强的鲁棒性, 样本量从 50 降为 10 时, 其识别率仅下降约 1%<sup>[1]</sup>。

在计算复杂度方面, 改进的算法也优于传统算法, 如果参与识别的说话人数为  $N$ , 则传统算法要将测试语音分别与  $N$  个人的矢量码本进行量化, 并分别求出其 Viterbi 评分; 而本系统中的应用的改进算法则大大节省了计算量: 只需将待识别的语音对共用的说话人特征图案量化 1 次, 再将得到的码本标号与各说话人的模型进行模式匹配。由此可见, 改进的 DHMM 算法比传统 DHMM 算法的计算量小得多。

## 4 结语

该嵌入式语音识别系统在以 ADSP-BF531 为核心的硬件基础上, 成功应用改进的 DHMM 算法完成了对非特定人孤立词的语音识别。该系统运行稳定、可靠, 对每个词条的识别时间在  $0.4\sim 0.6\text{s}$  之间, 具有很高的实时性, 对非特定人(包括参加训练和未参加训练)的综合识别率在 94% 以上, 具有较好的应用前景。

## 参考文献

- 1 方绍武, 戴蓓倩, 李霄寒. 一种具有强分类能力的离散 HMM 训练算法. 软件学报, 2001, 12(10): 1540-1543.
- 2 马斌, 郭莉莉, 韩中华, 戴敬. 基于 DSP 的语音识别系统的研究与设计. 微计算机信息, 2008, 24(8-2): 197-199.
- 3 陈志鑫, 郭华伟. 基于 TMS320C54x DSP 实时语音识别系统. 半导体技术, 2001, 26(4): 5-8.
- 4 何强, 何英. MATLAB 扩展编程. 北京: 清华大学出版社, 2002.
- 5 AD73311 Datasheet. Analog Devices, Inc., 2000.

(下转第 183 页)