

# 基于语音识别技术的门禁系统的研究

黎育红<sup>1</sup>, 张 熹<sup>1</sup>, 袁 荃<sup>2</sup>

- (1. 华中科技大学 数字化工程与仿真中心, 湖北 武汉 430074;
- 2. 武汉大学 经济与管理学院, 湖北 武汉 430072)

**摘要:** 通过对语音识别原理的系统分析,结合特定人语音识别的具体情况,研究了基于线性预测编码(LPC)的递推来推求倒谱编码(CC)基于滤波理论的基音周期确定算法,并将其结合起来,应用凌阳 SPCE061A 单片机,从硬件、软件、算法优化等方面实现了基于语音识别门禁系统的总体设计,在凌阳  $\mu'nSP$  IDE 平台上完成了原型系统的开发,并从应用的角度阐明了本系统理论上的合理性与鲁棒性。

**关键词:** 语音识别 特征提取 模式匹配 LPC DTW

综合集成计算机、通讯、自动识别、机械工程和安全管理等相关技术的门禁系统有效地解决了重要部门和场所的安全访问控制问题,已得到广泛的应用,成为日常工作和生活中的电子门卫。目前,门禁系统采用较多的是非接触式 RF 卡、生物识别技术、IC 卡、密码输入等方式。然而随着科学技术的发展,综合应用语音识别、指纹识别、虹膜识别、红外(热)感应等最新生物识别技术,结合电磁锁等技术的门禁系统已广泛吸引了人们的注意,并将逐步成为门禁系统发展的主流与最终目标之一。利用语音识别技术来实现门禁系统不用像其他方式一样需要触摸,具有方便、安全、准确、信息完整、独立性强、反应速度快等优点<sup>[1]</sup>。因此,基于语音识别技术的门禁系统有着非常独特的优势和发展前景。

## 1 语音识别技术及其原理

语音识别技术的关键在于准确地分辨出不同人的语音特征及其信息内容,并以此控制其他设备来满足人们的各种需要<sup>[2]</sup>。语音识别根据应用场合、使用对象、语音词汇量、算法模型等不同的分类依据,可分为不同的类型,如表 1<sup>[1]</sup>所示。针对办公室、家庭等私人小型场所,门禁系统涉及到的主要是特定人语音识别的特殊要求。

本文拟采用对用户的依赖性分类的形式,将相关语音识别分为两类:对用户声音特征的辨识和对用户发出的命令的确认。

基于对声音特征辨识的特定人语音识别在对安全要求很高的部门门禁系统中得到了较为广泛的应用。其基本原理:当系统接收到外界语音信息后,从事先训练好的语音库中找出惟一匹配的声音特征模型进行辨识,如果匹配成功则执行下一步操作;反之,如果无法匹配将会拒绝执行任何操作。

特定人语音识别应用较简单,不需要预先采集过多的样本,对硬件资源要求也较低,因此降低了系统运行中的前期成本;此外,其训练过程可以根据用户习惯,由用户任意定义控制项目的具体命令语句,因而适合大多数中小型以下企业或部门的各类应用。然而,特定人语音识别存在稳健性不理想,对有些人的语音识别率高,有些人的识别率却不高;系统刚训练完时识别率较高,但随着时间的延长会导致识别率慢慢降低等缺点<sup>[2]</sup>。为克服这些缺陷,本系统结合门禁系统语音识别的实际情况,通过从改进语音特征提取算法和模式匹配算法两方面对传统方法作了相应的改进,以提高门禁系统的语音识别性能和稳健性。

特定人语音识别分为语音训练和语音识别两个阶段。在训练阶段,通过 MIC 输入语音命令,然后对模拟语音信号进行预处理,对处理后得到的数字语音信号进行语音特征提取,为不同用户的不同语音特征参数建立一个相应

表 1 语音识别系统的分类

分类依据	用户说话的方式	对用户的依赖	模型	词汇量的大小	应用场合
类别	“简短词”语音识别系统	特定人语音识别系统	动态时间归整技术模型(DTW)	10~50 字(小)	语音命令识别
	“长句子”语音识别系统	非特定人语音识别系统 1.固定参考模型 2.自适应参考模型	隐马尔可夫模型及其扩展模型(DHMM、SDHMM、CDHMM)	50~200 字(中)	将语音转换为文字记录
	“连续发音”语音识别系统		人工神经网络模型(ANN)	200 字以上(大)	对语音的理解

# 集成电路应用

的语音特征模型库。训练完成后,进入语音识别阶段,对MIC输入事先训练好的语音命令,然后对模拟语音信号进行预处理,对处理后得到的数字语音信号提取语音特征参数,紧接着调出语音特征模型库进行匹配检测。如果在模型库中找到先前已经训练好的与之匹配的语音特征模型,就会产生识别结果;反之,则无法识别。特定人语音识别原理框图如图1所示。

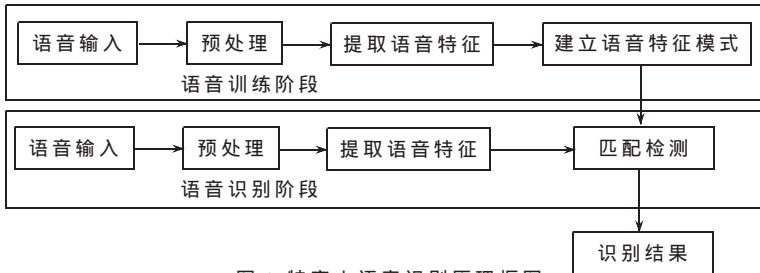


图1 特定人语音识别原理框图

## 1.1 预处理

预处理包括噪音去除以及端点检测。

### 1.1.1 噪音去除

将通过麦克风输入的一段模拟语音信号进行量化和采样,转换成数字语音信号;将这段含噪的数字语音信号去噪,得到干净的数字语音信号,再通过预加重技术滤除低频干扰(尤其是50Hz或60Hz的工频干扰)提升语音信号的高频部分,这样可以起到提升清音部分能量、抑制随机噪声和消除直流漂移的作用。

### 1.1.2 端点检测

端点检测也即是从一段语音信号中确定出语音的起点及结束点的过程。有效的端点检测不仅减少了系统的处理时间(帧数最少),而且能排除无声段的噪声干扰,从而使处理质量得到保证<sup>[3]</sup>。端点检测的困难在于无声段或者发音前后人为呼吸等产生的杂音,使得语音的端点比较模糊。

目前对端点检测采用较多的方法是单纯用过零率和能量值来检测,但是由于这种方法没有考虑到背景噪声,因此存在不可避免的缺点<sup>[4]</sup>。这里,采用当前帧与前一帧的过零率和能量值的差作为检验端点的标准就能很好地解决背景噪声的问题。

## 1.2 语音特征的提取

一般而言,都是通过对语音信号的分析处理,去除无关的冗余信息,获得影响语音识别的重要信息,来达到提取语音信号中表征不同人的不同声纹特征的目的。与非特定人语音识别不同,对特定人语音识别来讲,特征参数的选取更侧重于说话人的个人信息与特征的获取,而对语义信息获取的要求较低。

### 1.3 模型训练和模式匹配

模型训练是指按照一定的数据挖掘或知识发现规则,从大量训练样本中提取表征该样本类本质特征参数的过程。模式匹配则是根据一定准则,使待识别样本与

通过提取出来的训练样本的本质特征而建立的模型进行对比分析,以期获得最佳匹配。

语音识别所应用的模式匹配和模型训练技术主要有动态时间归整技术(DTW)、隐马尔可夫模型(HMM)和人工神经网络(ANN)。DTW是较早的一种模式匹配和模型训练技术,它应用动态规划方法成功解决了语音信号特征参数序列比较时时长不等的难题,在简短词语音识别中获得了良好性能,但它不适合连续语音大词汇量语音识别系统。相反,其对短时小词汇量的语音(有效语音长度小于3秒)的识别却既简单又有效,且识别率非常高。因此,本系统选取了DTW算法,并在具体实践中加以改进。

## 2 系统总体设计

本系统以改进的特定人语音识别技术为基础,通过对实际应用及现有文献中各类门禁系统进行了系统分析,改进了对语音信号的获取和处理的现有算法,从硬件、软件、以及算法等方面完成了对基于特定人语音识别技术的门禁系统的设计。系统的原理简图如图2所示。

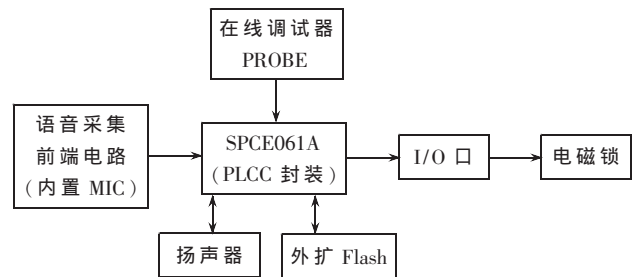


图2 系统原理简图

## 2.1 硬件组成

基于语音识别技术的门禁系统硬件组成如下。

### 2.1.1 SPCE061A 单片机

具有强大的语音处理与识别功能的凌阳SPCE061A单片机操作简单,支持基于汇编语言、C语言等高级程序设计语言的二次开发及凌阳 $\mu'nSP$  IDE 2.0.0软件开发平台的在线调试等功能。

SPCE061A单片机的工作电压范围为2.6~3.6V,工作频率范围为0.32M~9.152MHz,其指令系统 $\mu'nSPTM$ 提供具有较高运算速度、16位 $\times$ 16位的乘法运算指令和内积运算指令;具有DSP功能,因此有较高的处理速度,为复杂的数字信号处理提供了便利;支持多达10个中断向量以及14个中断源,适合实时应用的需求;内嵌2K字的SRAM和32K字的Flash,具有32位可编程的多功能I/O端口;包含有7通道10位通用A/D转换器和内置麦克风放大器与自动增益控制AGC功能的单通道声音A/D转换器,以及具有音频输出功能的双通道10位D/A转换器。

### 2.1.2 电磁锁

在门禁系统中,电磁锁是必须的配件,它可以在接

# 集成电路应用

通 12V 或者 24V 直流电源后,使锁舌缩进锁体,门被允许打开;反之,门被关上。门禁系统所用电锁一般有三种类型:电阴锁、电磁锁和电插锁,视门的具体情况选择。

另外,考虑到供电部分,应该在系统中加入直流升压器给电磁锁供电,并且还可以添加一个可以充电的蓄电池,以保证系统的持续供电。

## 2.2 软件设计

所选用的凌阳  $\mu'nSP$  IDE 集成开发平台支持汇编语言和 C 语言混合编程进行程序开发,集程序的编辑、编译、链接、调试以及仿真等功能为一体。

### 2.2.1 提示语音的生成

录好提示语音(.WAV 文件),采用 SACM\_S480 编码算法(音频编码算法),将 8K、16bit 的音频文件压缩为二进制文件,存储在程序存储器中。当应用程序运行时,再将压缩后的提示语音解压缩,以 8kHz 的频率送 D/A 还原出声音,并播放。

### 2.2.2 语音训练

在调用语音训练程序之前,首先确保识别器能正确的初始化。为了增强语音识别的可靠性,避免识别的命令倾向于噪音的问题,这里选择了两次语音训练取平均的方式。系统在发出开始训练的语音提示、延时 2 秒后开始训练,每条命令固定 1.3 秒的时长,当训练命令超出 1.3 秒时,只有前 1.3 秒命令有效。如果训练成功,则返回 0;未收到训练样本语音返回 1;需更多的训练语音样本数据时,返回 2;当训练样本语音背景太吵时,返回 3;由于单片机硬件限制,当语音特征模型库满时,返回 4;当两次输入训练样本不同时,返回 5。

“训练”函数 TrainWord():

```
int TrainWord(int WordID,int SndID)
{
    int Result;
    PlaySnd(SndID,3); //PlaySnd()为语音播放函数
    Result=BSR_Train(WordID,BSR_TRAIN_TWICE);
    //训练两次,获得训练结果

    switch(Result)
    {
        case 0:
            Break; //训练成功
        case 1:
            Break; //未收到训练样本语音
        .....
        default:
            break;
    }
    return result;
}
```

### 2.2.3 语音识别

语音识别程序包括识别程序和中断服务程序。识别程序选取词库、初始化 A/D 和定时器 TimerA,进行识别

运算及识别结果处理。中断服务程序定时读取 A/D 转换结果,并存入缓冲区。语音识别和放音分时复用 TimerA FIQ 中断,由标志位判断是语音识别处理还是放音处理。语音识别流程图如图 3 所示。

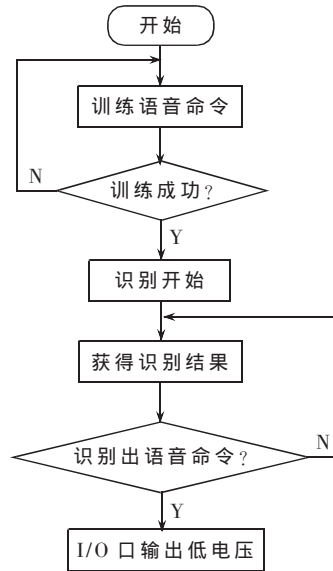


图 3 语音识别流程图

FIQ\_TimerA 中断服务程序:

```
L_FIQ_TimerA:
[P_INT_Clear]=R1;
call _BSR_FIQ_ROUTINE; //识别中断服务
call F_FIQ_Service_SACM_S480; //播放中断
pop R1,R5 from [sp];
reti;
```

识别程序如下:

```
BSR_DeleteSDGroup(0); //初始化 RAM
BSR_InitRecognizer(BSR_MIC); //初始化识别器
while(1)
{
    Result = BSR_GetResult(); //获得识别结果
    Switch(Result)
    { .....
        //识别出的命令,语音命令的序号号
        default:
            Break;
    }
}
..... (识别结果处理程序)
```

## 2.3 关键技术与算法的优化

由于特征参数选取的好坏会直接影响到识别精度,而合适的模式匹配算法可以很好地提高识别率,语音的特征提取以及模式匹配构成了基于语音识别门禁系统设计的关键。

### 2.3.1 特征参数的提取

本文设计的门禁系统综合采用了基音周期和线性

# 集成电路应用

预测倒谱编码(LPCC)共同作为特定人识别的特征参数。

(1) 基于线性预测编码(LPC)的倒谱编码(CC)推求算法。这种方法与直接计算倒谱系数相比,计算量要小。

其公式:  $\hat{x}(n) = \sum_{l=1}^p \alpha_l X(n-l)$ , 其中,  $p$  为 LPC 模型的阶数、 $X(n)$  为取样值、 $\hat{x}(n)$  为预测值、 $\alpha_l$  为预测系数。LPC 模型的阶数  $p$  的选择, 应该从频谱估计精度、计算量、存储量等多方面综合进行考虑, 而与线性预测分析的求解方法无关,  $p$  值取得太大会增加计算量和存储量。在实际语音识别应用中, 本文把阶数  $p$  取为 8~12, 如 10kHz 采样的语音信号, 取  $p=10$ , 但为了弥补鼻音中存在零点以及感冒等其他因素所引起的偏差, 通常在上述阶数的基础上再增加两个极点, 即  $p=12$ 。为了计算出倒谱系数  $C_n$ , 这里选用了 Durbin 递推算法, 由  $\alpha_l$  可递推得到, 公式为:

$$\begin{cases} C_1 = \alpha_1 \\ C_n = \sum_{l=1}^{n-1} (1 - \frac{1}{n}) \alpha_l C_{n-l} + \alpha_n, \quad 1 < n < p \end{cases}$$

(2) 基音估计的方法很多, 主要有基于短时自相关函数和基于短时平均幅度差函数(AMDF)等方法。本门禁系统采用的基音估计方法是: 对带通滤波后短时间的小词汇量语音信号进行线性预测, 求取预测残差; 再对残差信号求取自相关函数, 找到除零点外第一个峰值点位置, 便可估计出基音周期。为了得到较好的基音周期估计速度与效果, 在此考虑了一种反馈式窗长预估策略, 使得所取窗长初步满足至少大于 2 倍基音周期的要求。为克服声道共振峰特性的干扰, 可以先对语音信号进行非线性变换后再求取自相关函数。

### 2.3.2 模式匹配算法——DTW

DTW 算法不适合连续大词汇量的语音识别系统, 但对短时的小词汇量语音(有效语音长度低于 3 秒)的识别既简单又有效, 且识别率很高。DTW 算法采用动态规划最优化来找到一条从起点  $(n_0, m_0)$  到终点  $(n_N, m_M)$  的最佳搜寻路径。同时为了避免  $n$  轴的过分倾斜, 对通过路径中各点的平均斜率加以限制, 通常最大斜率定为 2, 最小斜率定为 1/2。

常规 DTW 算法要求从  $(1,1)$  出发到  $(N,M)$  结束, 这两点称为固定的起点和终点, 常规算法的缺陷是对端点的选取存在敏感依赖性。这里利用松弛的 DTW 算法, 可以使语音识别系统的识别性能得到改善, 且匹配效果较之常规方法更好, 即从起始点  $(1,1)$ 、 $(1,2)$ 、 $(2,1)$ 、 $(1,3)$ 、 $(3,1)$  以及终止点  $(N,M)$ 、 $(N-1,M)$ 、 $(N,M-1)$ 、 $(N-2,M)$ 、 $(N,M-2)$  中各选择一个最小值, 两语音样本之间的相互距离在相应的点放松后选择一个最小距离。一般情况下, 采取正确的端点检测之后, 起点和终点在纵轴和横轴方向各放松二帧, 即在两方向各放松 60ms 左右就足够了。松弛起点、终点的优点是可以克服由于端点检测不精确引起的误差。

### 2.4 系统工作原理

训练过程: 系统给出语音信号输入提示, 用户的语音通过麦克风进入语音信号采集前端电路, 第一次发出语音命令时, 由 SPCE061A 对采集的语音信号进行 A/D 转换和预处理, 然后提取用户的语音特征参数, 并存储到外扩的 Flash 内, 形成用户的特征参数模板。共进行两次训练, 第二次发出语音命令时, 提取的用户特征参数与由第一次语音输入形成的特征参数模板进行匹配, 在匹配距离小于模板更新阈值时, 将用户特征参数模板更新为两次特征参数的平均值, 形成最后的该用户的特征参数模板; 训练过程完成后再通过麦克风录入用户的语音, 再由 SPCE061A 对采集的语音信号进行处理, 将提取的用户特征参数与存储在外扩 Flash 内的特征参数模板进行匹配, 匹配距离小于认证阈值时, 通过认证; 然后再判断匹配距离是否小于认证模式下的模板更新阈值, 决定是否对模板进行更新。以上步骤完成后, 系统进入识别状态, 用户发出语音命令, 系统接受然后对模拟语音信号进行转换和处理, 在事先训练好的语音库中进行匹配检验, 如果匹配成功, 单片机在 I/O 口输出一个 3V 左右的直流电压, 经过直流升压器后, 电压升到 12V 或者 24V, 当电磁锁接通 12V 或者 24V 直流电源后, 锁舌缩进锁体, 门被打开。系统工作流程如图 4 所示。

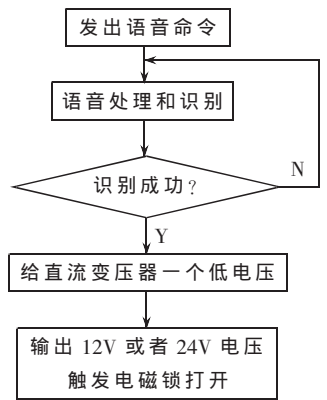


图 4 系统工作流程图

考虑到当用户感冒或声音发生暂时改变时, 系统无法辨识会将用户拒之门外, 本系统采用长密码方式进行 ID 认证。另外, 用户无论是在训练模式还是认证模式下都要输入一个可以自行修改的短密码, 以形成或找到与该用户相对应的特征参数模板, 而且系统还设置了超级管理用户, 通过超级管理用户可以对单片机内存储的所有用户进行删除或者添加。

本文介绍了门禁系统采用特定人语音识别的原因、原理。作为一种新型的现代化安全工具, 语音识别门禁系统无疑将会在未来成为从事门禁系统开发的企业以及各安全部门争相开发研究的重点。基于以上理论, 本系统在采用 SPCE061A 单片机和直流升压器以及电磁锁、在微软 WINXP 操作系统平台与凌阳  $\mu'nSP$  IDE 集成开发环境下, 应用 C 语言实现了系统原型的开发。本系统适合应用于家庭、办公室、宾馆等小型且较为安静的场所。整个系统成本低, 较为容易实现, 使用方便, 稳定性好, 识别效果好, 值得推广。但在大型场合以及很嘈杂的环境下, 由于无法对背景噪声的有效控制, 该系统在以后的研究中需要进一步改进。

(接上页)

### 参考文献

- 1 Rabiner L, Fundamentals of Speech Recognition[M]. Prentice Hall Press, 1993
- 2 刘敬伟,徐美芝,郑忠国等.基于 DTW 的语音识别和说话人识别的特征选择[J].模式识别与人工智能,2005;18(1):50~54
- 3 胡光锐.语音处理与识别[M].上海:上海科学技术文献出版社,1994
- 4 胡航.语音信号处理[M].哈尔滨:哈尔滨工业大学出版

社,2000

- 5 李晶皎.嵌入式语音技术及凌阳 16 位单片机应用[M],北京:北京航空航天大学出版社,2003
- 6 张培仁,张志坚,高修峰等,十六位单片微处理器原理及应用(凌阳 SPCE061A)[M].北京:清华大学出版社,2005
- 7 刘福才,王冬云.语音识别技术及其在控制领域中的应用研究(1~2).可编程控制器与工厂自动化,2005;(11~12)
- 8 王和顺,黄惟公,邓成中.基于 SPCE061A 的语音识别系统的研究[J],计算机工程与设计,2004;(12)

(收稿日期:2006-07-12)