





The Research on speaker recognition of wireless access control  
system

by  
HU Jin ping

B.E.(Lanzhou University of Technology)2008

A thesis submitted in partial satisfaction of the

Requirements for the degree of

Master of engineering

In

Circuits and Systems

in the

Graduate School

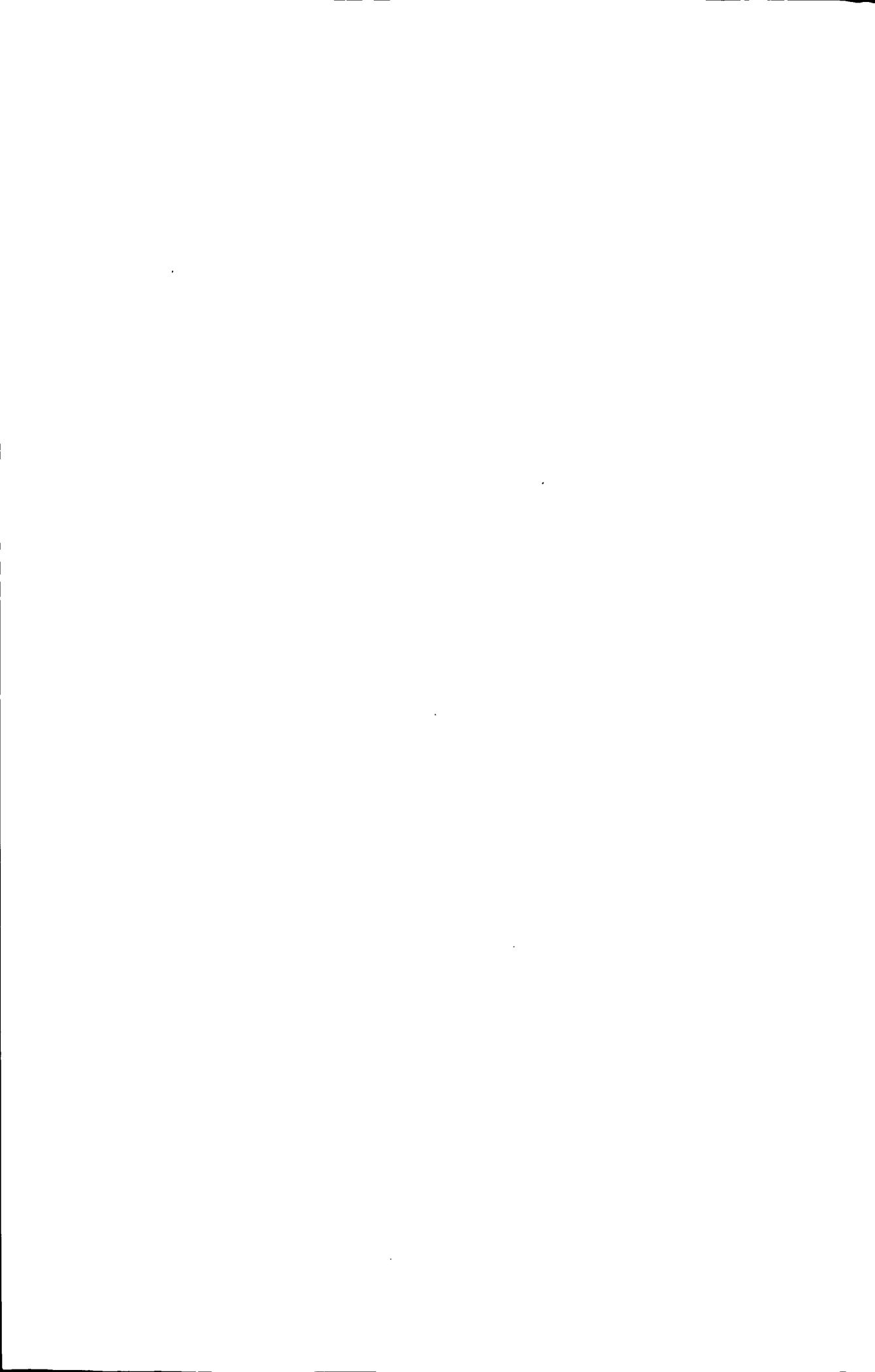
of

Lanzhou University of Technology

Supervisor

Engineer Chen Ruozhu

May,2011



# 兰州理工大学学位论文原创性声明和使用授权说明

## 原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：胡金平

日期：2011年6月7日

## 学位论文版权使用授权书

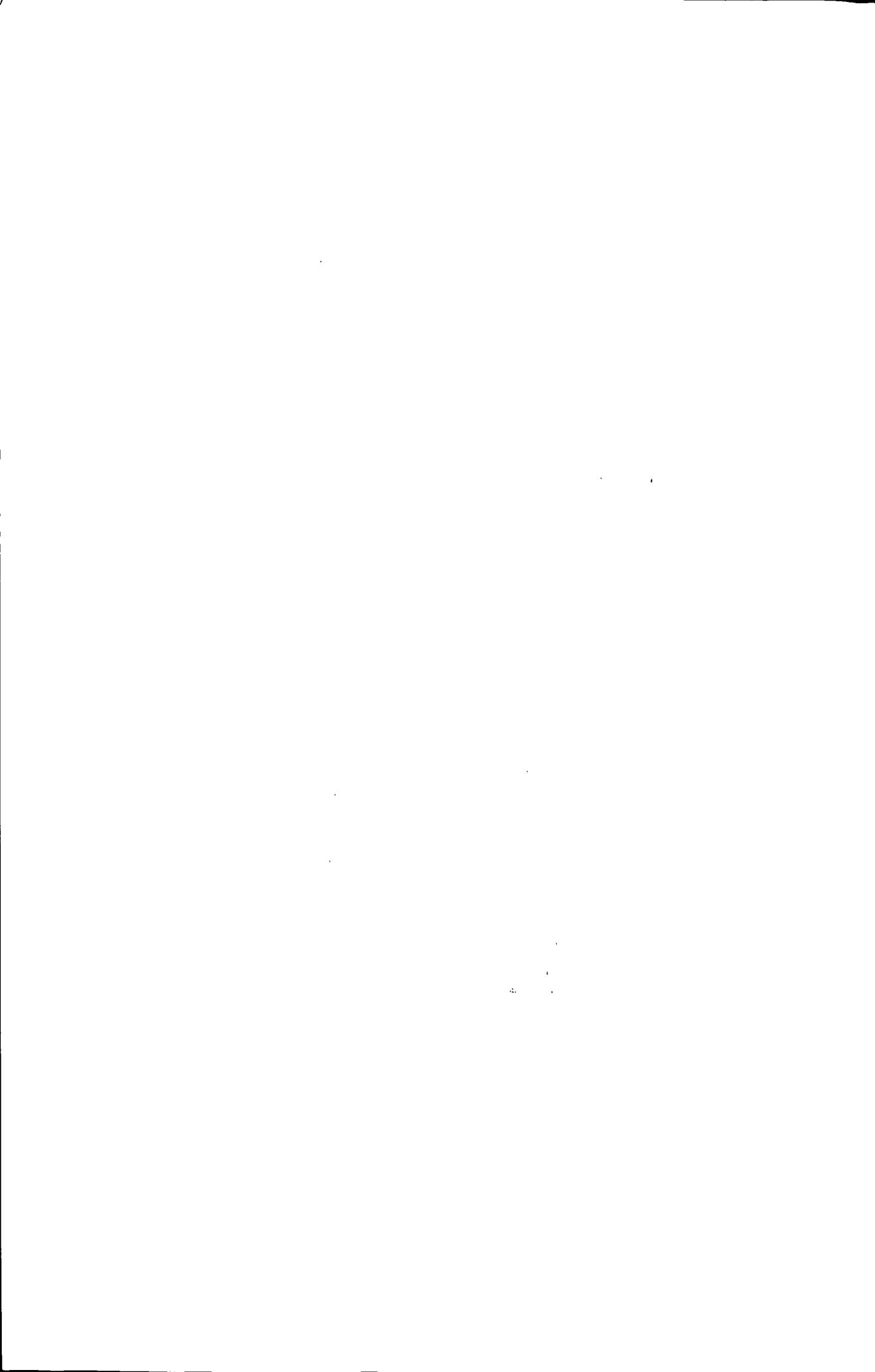
本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权兰州理工大学可以将本学位论文的全部或部分内 容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》，并通过网络向社会公众提供信息服务。

作者签名：胡金平

日期：2011年6月7日

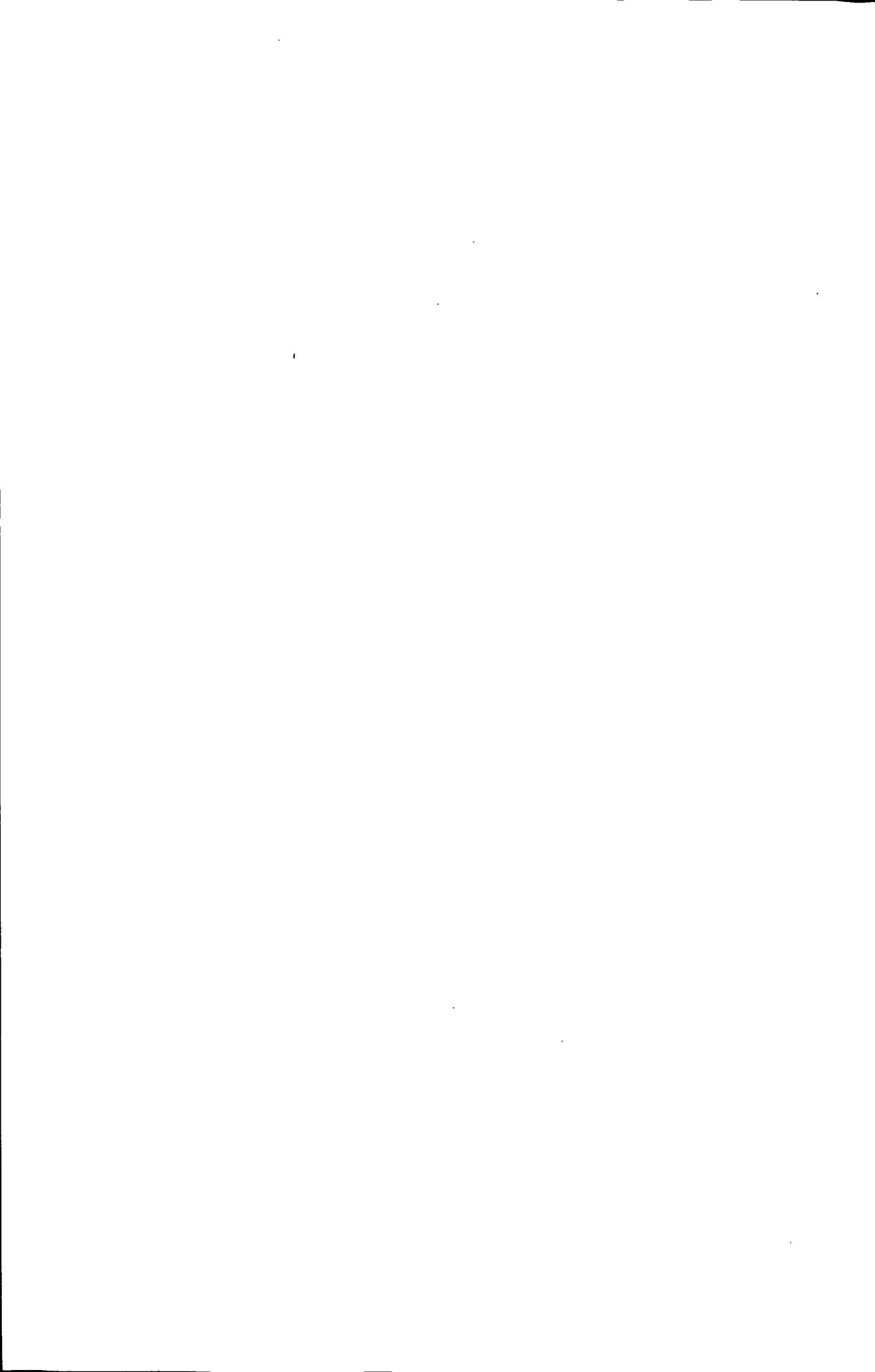
导师签名：陈君平

日期：2011年6月7日

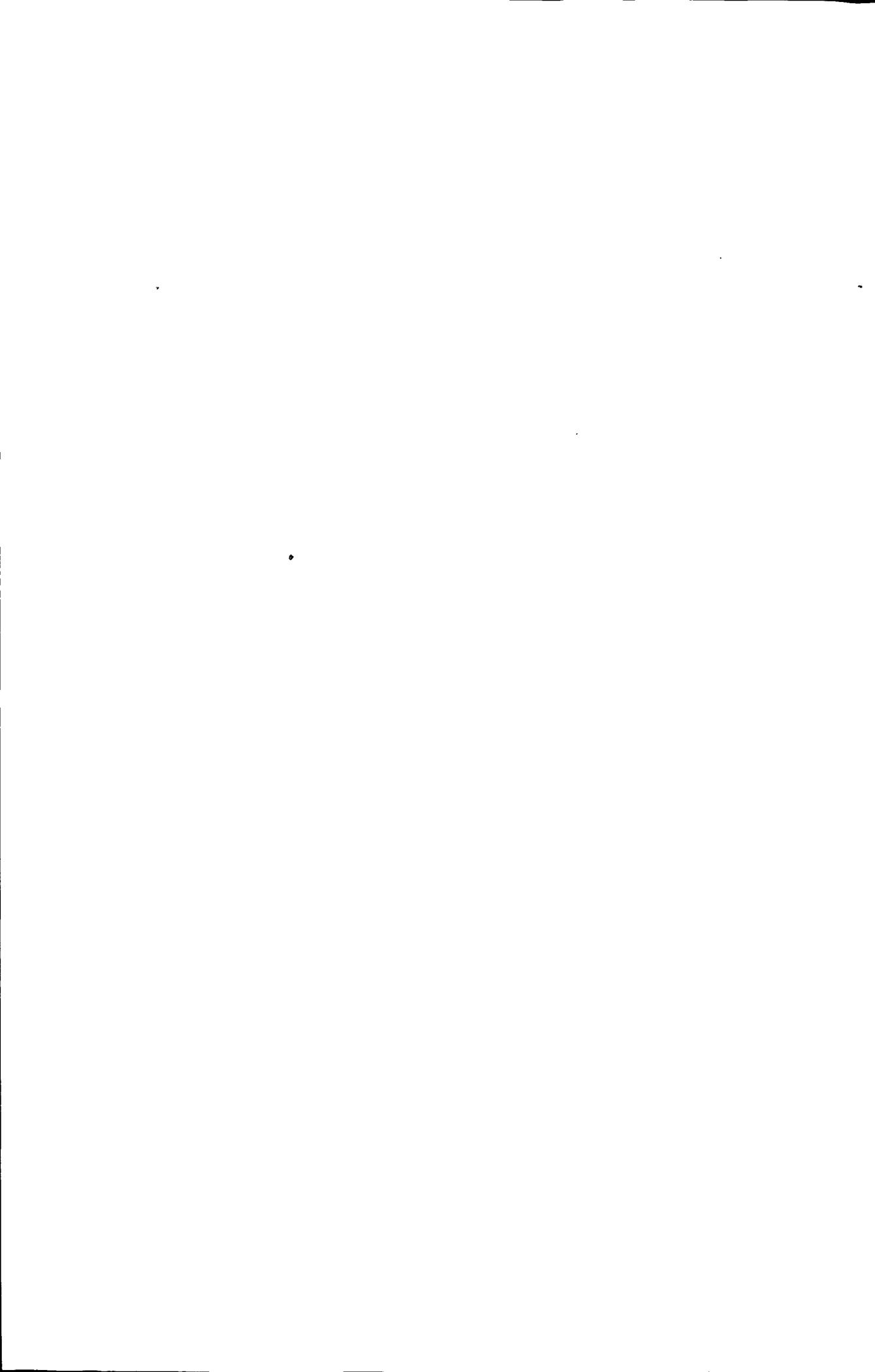


## 目 录

|                                |     |
|--------------------------------|-----|
| 摘 要.....                       | III |
| Abstract.....                  | IV  |
| 插图索引 .....                     | V   |
| 附表索引.....                      | VII |
| 第 1 章 绪 论.....                 | 1   |
| 1.1 选题背景及意义.....               | 1   |
| 1.2 国内外研究现状.....               | 1   |
| 1.3 研究的主要内容及论文结构.....          | 6   |
| 1.4 本章小结.....                  | 6   |
| 第 2 章 门禁系统的设计方案与相关理论研究.....    | 7   |
| 2.1 说话人识别原理.....               | 7   |
| 2.2 语音信号的预处理.....              | 8   |
| 2.2.1 预加重.....                 | 8   |
| 2.2.2 分帧与加窗.....               | 8   |
| 2.3 语音信号的端点检测.....             | 9   |
| 2.4 语音信号的特征提取.....             | 11  |
| 2.4.1 线性预测系数.....              | 11  |
| 2.4.2 美尔倒谱系数 MFCC.....         | 12  |
| 2.5 语音信号的模式匹配.....             | 14  |
| 2.6 本章小结.....                  | 17  |
| 第 3 章 基于说话人识别的门禁系统的硬件设计.....   | 18  |
| 3.1 系统的硬件结构.....               | 18  |
| 3.2 TMS320VC55x DSP 的基本结构..... | 18  |
| 3.2.1 C55x 的 CPU 结构.....       | 20  |
| 3.2.2 存储器映射.....               | 21  |
| 3.2.3 DSP 的片内外设.....           | 22  |
| 3.2.4 C55x 的指令流水线和并行性.....     | 22  |
| 3.3 各主要单元电路.....               | 23  |
| 3.3.1 电源电路.....                | 23  |
| 3.3.2 时钟电路.....                | 24  |
| 3.3.3 音频编解码模块.....             | 26  |
| 3.3.4 外扩存储电路.....              | 29  |
| 3.3.5 上电自举模式与语音处理模式的设定.....    | 33  |



|                                    |           |
|------------------------------------|-----------|
| 3.3.6 JTAG 接口电路.....               | 35        |
| 3.3.7 无线传输电路.....                  | 35        |
| 3.4 硬件电路板的抗干扰设计.....               | 38        |
| 3.5 本章小结.....                      | 38        |
| <b>第 4 章 基于说话人识别门禁系统的软件设计.....</b> | <b>39</b> |
| 4.1 说话人识别的结构框图.....                | 39        |
| 4.2 说话人识别的软件设计.....                | 40        |
| 4.2.1 系统的初始化.....                  | 40        |
| 4.2.2 AIC23 语音信号采集.....            | 43        |
| 4.2.3 数据处理.....                    | 43        |
| 4.2.4 BOOT LOADER 程序设计.....        | 44        |
| 4.2.5 非易失存储器 FLASH 在线编程.....       | 45        |
| 4.3 无线通信的软件设计.....                 | 48        |
| 4.4 本章小结.....                      | 51        |
| <b>第 5 章 系统联调测试与结果分析.....</b>      | <b>52</b> |
| 5.1 系统的硬件调试.....                   | 52        |
| 5.2 系统的软件调试.....                   | 52        |
| 5.3 系统的测试设备.....                   | 54        |
| 5.4 测试结果与分析.....                   | 55        |
| 5.5 本章小结.....                      | 57        |
| <b>结论与展望.....</b>                  | <b>58</b> |
| <b>参考文献.....</b>                   | <b>60</b> |
| <b>致 谢.....</b>                    | <b>64</b> |
| <b>附录 攻读学位期间所发表的学术论文目录.....</b>    | <b>65</b> |



## 摘要

说话人识别是生物识别技术的一种，其独特的方便性、经济性和准确性等优势越来越得到人们的关注。但目前国内外的绝大多数说话人识别系统都是基于PC机来完成的，从而成本过高，并且现有的语音门禁系统产品也仅限于单用户使用，这些在一定程度上阻碍着说话人识别门禁系统的发展。

本文回顾了说话人识别技术发展的历史，根据说话人识别系统的构成模型，介绍了说话人识别的基本原理，包括语音信号的预处理、端点检测。在研究传统端点检测的基础上提出双门限的比较方法。特征提取，重点分析了LPC倒谱和MFCC倒谱；模式匹配采用改进的动态时间规整算法并进行详细的分析。对算法的各个部分进行了MATLAB仿真验证。

本文以 TMS320VC5509A DSP 芯片为核心，以高保真语音编解码芯片 TLV320AIC23，同步动态随机存储器 SDRAM 和异步存储器 FLASH 为主要外设完成了系统的硬件设计，具体硬件设计电路包括：电源模块、时钟模块、外扩存储器模块、音频编解码器模块、JTAG 接口、Si4432 模块。根据系统实现的功能要求，软件基于 CCS 开发环境，采用并行 EMIF 引导方式，实现了 TMS320VC5509A DSP 扩展 FLASH 的在线烧写；采用了模块化的程序设计方法，程序整体采用 C 语言和汇编语言混合方式编程，完成了语音识别工程及无线数据收发工程。

本文基于实验室环境，实现了说话人识别的无线门禁系统的软硬件设计，基本实现了预期的目标，充分体现系统的可行性，具有较强的实用价值。

**关键词：**说话人识别；DSP；动态时间规整；Si4432；门禁系统



## Abstract

Speaker recognition is a kind of biological certification technology. In the biological certification area, speaker recognition widely draws the attention because of its convenience, efficiency and accuracy. At present, the speaker recognition system are work depend on computer mostly, so the system is expensive and serve just for a single user. To a certain extent, it hindered the speaker recognition access control system development

This article review the courses of speech recognition technology. On the basis of the above, the basic theory of speaker recognition is introduced according to the model of speaker recognition system, which includes the voice signals' pretreatment and the detection of endpoint. Comparing the traditional method of detection, this paper proposes a double threshold comparative method. The feature is extracted and it is focus on the inverted pedigree of LPC and Mel scale frequency. The speaker recognition algorithm DTW is introduced and analyzed in the paper. Each part of the algorithm is verified on the MATLAB simulation.

In the hardware design, the system uses the digital signal processor of TMS320VC5509A as the core, high-fidelity voice codec chip TLV320AIC23, synchronous dynamic random access memory SDRAM and asynchronous memory FLASH as the main peripherals. The part of hardware circuit design specifically includes power modules, clock modules, external expansion memory modules, audio codec modules, JTAG interface and Si4432 mode . Based on CCS development environment, this piece of paper has parallel EMIF-boot mode to achieve the on-chip bootloader of the digital signal processor of TMS320VC5509A after online simulation. Using modular programming methods, procedures using C language and assembly language hybrids programming. It finished the speech recognition engineering and wireless data transceiver engineering. It is proved that the design scheme of the system is effective.

**Key words:** speaker identification; DSP; DTW; Si4432; wireless access control system



## 插图索引

|  |    |
|--|----|
| 图 2.1 门禁系统框图 .....                       | 7  |
| 图 2.2 说话人识别原理图 .....                     | 8  |
| 图 2.3 改进前的端点检测.....                      | 11 |
| 图 2.4 改进后的端点检测.....                      | 11 |
| 图 2.5 Mel 滤波器组.....                      | 13 |
| 图 2.6 MFCC 计算流程图.....                    | 14 |
| 图 2.7 DTW 算法搜索路径.....                    | 15 |
| 图 2.8 匹配路径约束示意图.....                     | 16 |
| 图 3.1 系统硬件结构图.....                       | 18 |
| 图 3.2 C55x 的 CPU 结构图.....                | 20 |
| 图 3.3 TMS320VC5509A 封装存储器映射图 .....       | 21 |
| 图 3.4 C55x 流水线示意图 .....                  | 23 |
| 图 3.5 电源电路.....                          | 24 |
| 图 3.6 时钟电路.....                          | 25 |
| 图 3.7 TLV320AIC23 芯片引脚图 .....            | 26 |
| 图 3.8 TMS320VC5509A 与 AIC23 接口示意图 .....  | 28 |
| 图 3.9 I2C 模块数据传输图.....                   | 29 |
| 图 3.10 AIC23 在 DSP 模式下的串口时序图 .....       | 29 |
| 图 3.11 TMS320VC5509A 与 SDRAM 接口示意图.....  | 31 |
| 图 3.12 TMS320VC5509A 与 FLASH 接口示意图 ..... | 32 |
| 图 3.13 LVC16245 连接图 .....                | 34 |
| 图 3.14 上电自举模式示意图.....                    | 34 |
| 图 3.15 JTAG 接口电路 .....                   | 35 |
| 图 3.16 Si4432 硬件连接图.....                 | 37 |
| 图 4.1 说话人识别的软件流程图.....                   | 39 |
| 图 4.2 擦除和编程命令示意图.....                    | 46 |
| 图 4.3 触发位检测算法流程图.....                    | 47 |
| 图 4.4 FLASH 擦除图 .....                    | 48 |
| 图 4.5 FLASH 烧写图 .....                    | 48 |
| 图 4.6 DSP 与 Si4432 连接图 .....             | 49 |
| 图 4.7 SPCR1 控制寄存器图 .....                 | 49 |
| 图 4.8 SPCR2 控制寄存器图 .....                 | 49 |
| 图 4.9 SPI 操作模式 3 时序图.....                | 50 |
| 图 4.10 Si4432 发送软件流程图.....               | 50 |
| 图 5.1 语音工程 yuyinshibie.pjt 界面 .....      | 53 |

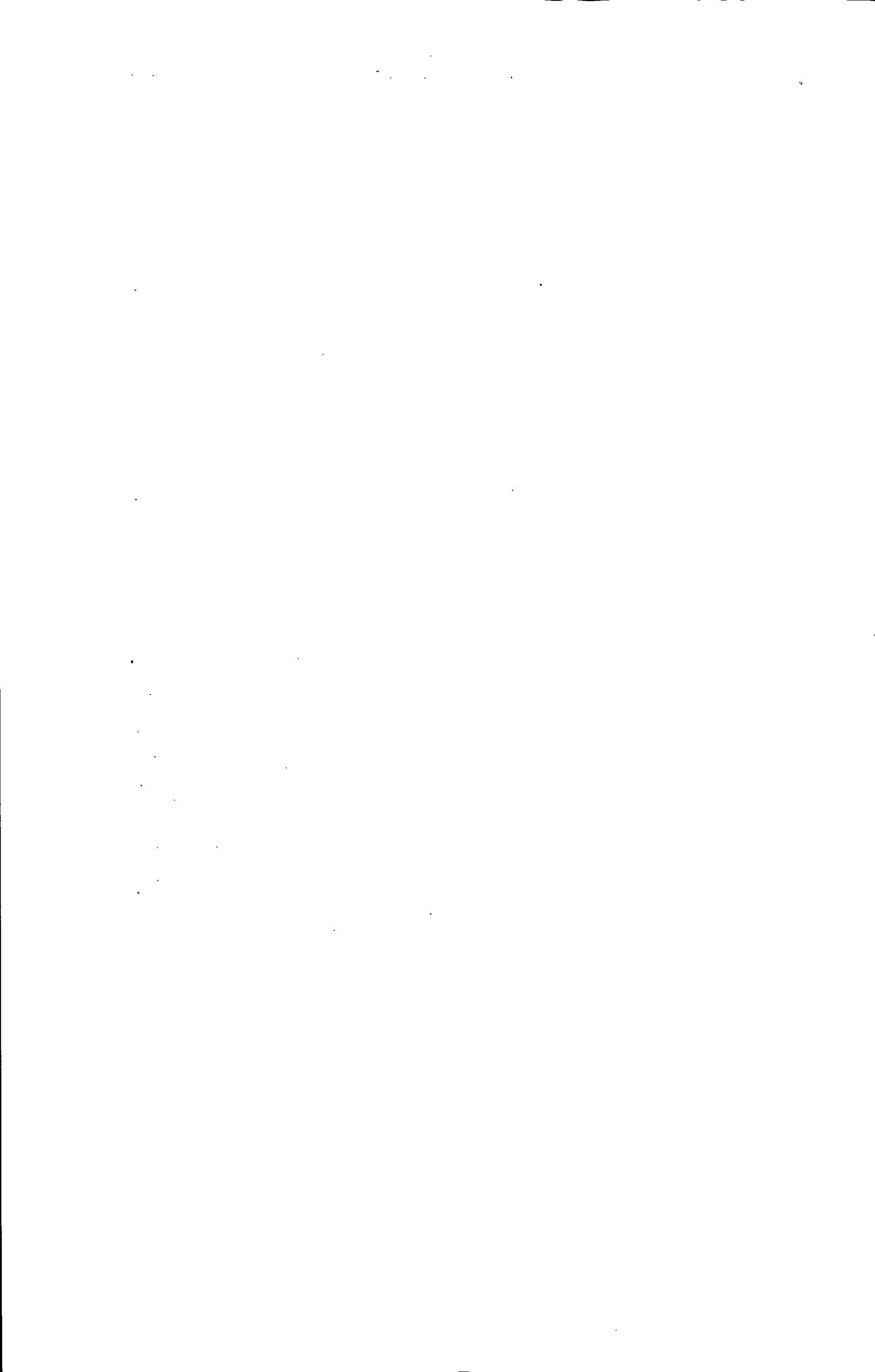


|                               |    |
|-------------------------------|----|
| 图 5.2 烧写工程 flash.pjt 界面 ..... | 53 |
| 图 5.3 DSP 硬件系统 .....          | 54 |
| 图 5.4 Si4432 硬件系统 .....       | 54 |
| 图 5.5 系统的测试平台 .....           | 54 |
| 图 5.6 说话人识别实验结果 .....         | 55 |
| 图 5.7 端点检测前波形图 .....          | 56 |
| 图 5.8 端点检测后波形图 .....          | 56 |
| 图 5.9 Profile 观察窗口 .....      | 56 |



## 附表索引

|                               |    |
|-------------------------------|----|
| 表 2.1 矩形窗与汉明窗的比较.....         | 9  |
| 表 3.1 C54x 与 C55x 的比较.....    | 19 |
| 表 3.2 C55x 的特征和优点.....        | 19 |
| 表 3.3 时钟模式寄存器 CLKMD.....      | 25 |
| 表 3.4 TLV320AIC23 内部寄存器.....  | 27 |
| 表 3.5 HY57V641620 引脚说明.....   | 30 |
| 表 3.6 全局控制寄存器 EGCR.....       | 31 |
| 表 3.7 TMS320VC5509A 引导方式..... | 33 |
| 表 3.8 LVC16245 功能表.....       | 34 |
| 表 3.9 Si4432 的引脚定义与描述.....    | 37 |
| 表 4.1 引导构.....                | 44 |
| 表 4.2 AM29LV800B 操作命令说明.....  | 45 |
| 表 4.3 写操作各时间参数.....           | 47 |
| 表 5.1 说话人识别实验结果.....          | 55 |



# 第1章 绪论

## 1.1 选题背景及意义

伴随着全球数字化、信息化时代的来临,我们对更便捷更安全的生物门禁系统的需求也不断增长。传统的门锁或密码锁为特征的身份认证技术不方便、容易被盗等特点暴露出很大的缺点,很难满足现代社会更高安全性和快速便捷性的要求,而在生物技术和信息科学迅速发展的今天,生物认证技术是一种方便、先进的信息安全技术在现实生活中已经得到广泛的应用。人的语音信号是自然界生物信号的一种,每个人的声音都与环境、情绪、声道模型、健康状况等有关,都有自己独特的特性,所以可以通过分析语音信号来识别说话人成为可能<sup>[1-2]</sup>。特别是进入到21世纪以来,科研人员对说话人识别进行了很广泛的研究,在生物识别领域中,利用说话人的语音信号来进行说话人识别有许多独特的优点,如语音信号相对于其它的生物信号,成本比较低廉,采集相对简单,没有严格的方向性,可以在黑暗中传播等,说话人识别技术被广泛的应用于医学应用、声控电子锁、搜索罪犯等领域。

而综合集成了大型计算机、无线通讯、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合的安全控制问题,已经被应用的越来越广泛,并产生了深远的影响。目前市场上的门禁系统大多是采用有线的通信方式,小区单元通过有线的方式与各个用户端相连;与这类有线网络通信的楼宇控制产品相比,无线门禁的解决方案安装布置更加灵活、费用低廉、楼宇间的设备重组更加方便、可移动性强等方面的优势。因此,无线门禁系统受到相关公司、研究团队的重视,其将在门禁市场中占据越来越重要的地位。

本课题针对人们对智能安防系统日益增长的需求,提出了基于说话人识别的无线门禁系统的应用研究。该系统可应用于小区单元及用户控制锁的工程实践中,充分体现了生物信号中说话人识别的便利性,无线通信的实时性,为现代化的门禁系统提供一种可靠的方案。

## 1.2 国内外研究现状

对说话人识别的研究<sup>[3-4]</sup>开始于上世纪六十年代,主要研究内容是人工的说话人识别。1966年,美国法院第一次采用语谱图进行了取证<sup>[5-6]</sup>。70年代中期BSA<sup>tal</sup>研究了线性预测LPC(Linear Predictive Coding)系数、倒谱系数等不同的特征参数,并论证了倒谱系数在进行说话人识别中是比较理想的特征参数。许多学者和研究人将对矢量化模型VQ(Vector Quantization)、隐马尔科夫模型HMM(Hidden Markov Model)<sup>[7]</sup>和人工神经网络ANN(Artificial Neural Network)<sup>[8-10]</sup>引入到说话人识别研究中,大大地推动了说话人识别技术的发展。

我国说话人识别的研究工作开始于五十年代,伴随着计算机技术的快速发展,根据我国的具体实际情况,特别是在汉语识别技术上,创造出适合汉语识别的技术。说话人识别要求从语音信号中提取个体差异,提取出能反映说话人是谁的信息,从而进行说话人识别。其基本原理是为每一个说话人建立一个能够描述这一说话人个性特征的模型,作为此说话人个性特征的描述。说话人识别可以分为两个范畴,即说话人辨认和说话人确认。前者是把要检测的语句判为N个训练说话人之一所说,是一个多择一的问题;后者则是把带检测人的语句与其参考说话人的相比较,相符的即得到确认,不相符的则得到否定,是二择一的问题。说话人辨识时取与测试语音匹配距离最小的说话人模型所对应的说话人作为说话人辨识的结果,说话人确认时与所声称的说话人模型进行匹配,若匹配距离小于规定的阈值,则该说话人得到证实,否则该说话人不是他所声称的说话人。要实现说话人识别系统需要解决三个问题:1)首先对说话人的语音信号进行预处理,端点检测和特征提取;2)说话人模板的建立和模板参数的训练;3)测试的语音模板与说话人模板的匹配距离计算。

针对当前的说话人识别,大致可归为三类:

第一类,模板匹配法。以前的说话人识别是根据简单的模板匹配的原理构造的特定人、小词汇量、孤立词识别系统<sup>[11-12]</sup>。在训练和识别的时候,必须保证两个模板采用相同的帧数,相同的预处理与窗函数。但是在实际的应用中,说话人的语音具有很大的似然性,所以要把两帧延长到相同的长度。于是,日本学者板仓(Itakura)将动态规划算法(Dynamic Programming, DP)的概念用于解决孤立词识别时的说话速度不均匀的难题,提出了著名的动态时间规整算法(Dynamic Time Warping, 简称DTW)。它是最近被广泛研究和应用的一种模式识别的方法,在说话人识别中获得了比较广泛的应用<sup>[14-15]</sup>。DTW是一个典型的求解最优化问题,将说话人的语音信号在时间轴上进行不均匀的弯曲,使其测试模板与参考模板对齐,并计算最小的帧匹配距离,从而获得累计距离最小的规整函数,这就保证了参考模板的特征和测试模板特征之间最大的声学相似性和最小的误差失真。但是对于非特定人、大词汇量、连续语音识别系统来讲,这种模板匹配就显得力不从心,就必须寻求模板匹配以外的其它识别方法—随机模型法及概率语法分析法。

第二类,概率模型法。突出的代表是(1)隐马尔可夫模型HMM方法<sup>[16-18]</sup>。可以使用HMM的概率参数来对似然函数进行估计与判决,从而得到识别结果。在用于语音建模时,它是双重的随机过程。2)高斯混合模型(Gaussian Mixture Model, GMM)<sup>[19-21]</sup>,因为每一个说话人的个性特征都可由形式一致的高斯混合概率密度函数中的参数来表示,故为每一个说话人建立一个GMM混合模型来训练,从而获得模型中的参数。

第三类,概率语法分析法。对于大词汇量、连续非特定人的语音识别具有较

大的优势。专家学者通过采集分析语音信号，并建立不同的约束规则来描述不同人的语音信号，从而达到识别说话人的目的。而说话人的语音具有较大的随机性和时变性，此方法的重点是如何选取这些约束规则以及它们之间相互转化的条件等，虽然说话人识别的正确率大大提高，但是算法计算的复杂度要远远的高于前两种方法，并且占用较大的存储空间，移植起来具有很大的难度。

在介绍了常用的说话人识别的方法外，还有很多像神经网络方法<sup>[22-23]</sup>算法的识别方法。人工神经网络可采用物理可实现的器件或利用现有的计算机来模拟生物神经网络的某些结构与功能，包括神经元、网络拓扑和网络的训练算法三个部分。神经网络对复杂不确定性的问题具有自适应和自学习能力，快速优化能力，对大量定性或定量信息的分布存储能力，并行处理与合成能力。神经网络还具有较强的鲁棒性，若局部或部分神经元损坏后，不会影响全局的活动。

最近采用不同方法的融合技术<sup>[24-26]</sup>有了广泛的应用。1) 维数约简与识别方式相结合：局部PCA (Principal Component Analysis)<sup>[27]</sup>和GMM结合的说话人识别系统。2) 聚类与识别方法的结合；3) 识别方式的结合：主要是采用了SVM (support vector machine)<sup>[28-29]</sup>分类器，充分利用了SVM适合不同样本之间的分类，对于异类样本具有很强的分类能力。可以与其它很多模式识别的算法相结合，利用不同分类器的特点，实现识别结果的最优化。

近年来虽然说话人识别的研究已经取得很大的进展，但应用于实际情况还需要解决很多问题<sup>[30]</sup>。

问题一：高维说话人特征的缺陷。为了提高说话人识别系统的识别率，往往采用更高维的特征，然而，高维的特征矢量会增加计算量与存储空间，不利于实时的对语音信号进行处理。

问题二：说话人分类方法的局限性。目前虽然有很多说话人识别的方法，在说话人识别的领域中也取得了很好的识别效果，如矢量量化模型、隐马尔科夫模型、神经网络模型、支持向量机模型等。但是选用这些高维向量，需要不断的迭代计算，使得说话人识别的时间相对较长。因此，高复杂度的精确的模式匹配方法并不适合于实时的说话人识别应用，这需要强大的计算处理能力与数据存储设备。现有的低复杂度低计算量的模式识别方法，虽然所需的训练时间和测试时间都比较短，但对说话人识别的准确率有所下降，只适用于识别人数相对较少的情况。一旦说话人的数量增多，识别性能就会急剧下降。

随着说话人识别中对于语音信号处理技术和计算机硬件、DSP的高速发展，语音信号处理的硬件方式发生着质的飞跃。随着C55x系列DSP的推出和应用开发的逐步深入，在实时性和低功耗方面体现出其优良的性能，C55x正成为通信和个人消费领域的主流DSP产品<sup>[31]</sup>。作为TI公司低功耗系列的DSP芯片，2000系列主要用于工程电机等的自动控制中；5000系列主要用于语音信号处理，算法的实现；6000系列主要用于图像的压缩、增强等。对于市场上的C8051单片机、ARM及FPGA

等，它在功耗、信号算法的处理等方面具有显著的优势<sup>[32]</sup>。DSP主要发展现状如下：

(1)制造工艺与存储容量。目前的 DSP 芯片都采用 CMOS 工艺。芯片的引脚越来越多，集成的功能也越来越强，需要设计的外围电路越来越少，TI 公司在 2005 年 12 月发布的达芬奇系统已经把音视频部件集成在 DSP 内部，内部集成的数据和程序存储器也不断增大，成本、体积和功耗不断下降。

(2)内存结构和运算速度。目前的 DSP 模块基本上都采用数据总线和程序总线分离的哈佛结构或改进的哈佛结构，多级流水线指令，内部集成了锁相环，可以相应的进行倍频或者分频，在一定范围内实现任意的时钟频率，使 CUP 在一个指令周期内可多次对程序空间和数据空间进行访问，大大提高了 DSP 的运行速度。

(3)开发工具。具有完善的软件仿真工具和硬件开发工具，如：软件开发环境 CCS，软件仿真器 Simulator，在线仿真器 Emulator、C 编译器，在线分析器 Profile 等，给开发带来了极大的便利。其中 CCS 软件开发环境是 TI 公司针对本公司的 DSP 产品开发的集成开发环境，它集成了代码的编译、链接和调试等诸多功能，而且支持 C/C++和汇编的混合编程，开放式的结构允许用户外扩自身的模块。

随着芯片工艺的不断发展，DSP 性能的进一步增强，为说话人识别技术走向实践应用奠定了基础<sup>[33-34]</sup>。基于说话人识别的无线门禁系统，作为新型的现代化智能化安全管理系统，它包括电子信息技术、计算机技术、无线通讯技术、生物识别技术等诸多新技术。如何设置参考模板和测试模板，何时打开门锁，遇到特殊情况是否要发出警报等，以达到安全快捷的目的，是门禁系统最基本的功能。

在六十年代，计算机的应用推动了说话人识别的发展。日本的东京无线电研究实验室、京都大学和 NEC 实验室都制作了能够进行说话人识别的专用硬件，对说话人识别领域进行了开拓性的研究工作。近年来，高性能数字信号处理芯片 DSP(Digital Signal Processor)技术的迅速发展，为说话人识别的实时实现提供了可能。

说话人识别的研究人员也一直致力于语音识别专用芯片和系统的研究，但是，大多数的语音识别专用芯片识别性能差，不具备实用的要求。直到近十年以来，随着说话人识别算法的深入研究和集成电路技术的发展，才出现了一些具有实用价值和市场前景的说话人识别专用芯片。其中，较为成功的两个芯片详细介绍如下：

#### 1. RSC-364

由美国的公司开发，2000 年开始生产，是一颗为消费类电子产品应用的低价位的语音识别专用芯片，即 RS-364，它使用预先学习好的人工神经网络进行

非特定人语音识别,不需要经过训练就可以识别" Yes''、" 0"、" 0k'"等简单语句,其说明书上介绍其识别率为 97%。此外,RSC-364 可以识别特定人、孤立词命令语句,约 60 条左右,其说明书上称其识别率为 99%以上。RSC-364 还具有 5-15kb/s 的语音合成,其语音合成由 Snesory 专门设计,其音质较好。它还具有改进的 ADPCM(自适应差分脉冲调制)语音编解码功能,用作语音回放。

## 2. UniSpeech-SDA80D51

德国 Infineon 公司 2000 年开始生产的产品,它是一颗高性能的语音专用芯片。这样的设计能够满足立体声处理或者消除外界干扰等功能要求。UniSpeech-SDA80D51 的语音处理软件包括:利用 DTW 算法的特定人语音识别,能够识别 100 条语句;利用 HMM 算法的非特定人语音识别,词汇量可以达到 100 条语句。高质量、低码率(2.4-13kb/s)的语音编解码,用作语音提示和语音降低外界的噪声干扰;说话人识别功能等等。

国内在语音识别专用芯片的开发与研究方面起步较晚。比如清华大学“语音识别专用芯片研究开发”课题,研究开发了国内第一个具有自主知识产权的语音识别专用芯片的软件部分。该实验室是国内最早专门从事语音技术与专用芯片设计研究单位之一。芯片能够识别约 20-30 条特定人语音命令,同时具有语音合成(提示)与语音编解码(回放)功能。语音识别率达到 98%以上。

随着现代社会的发展,人们对门禁系统的要求越来越严格和现代化、智能化,门禁系统的应用范围也越来越广泛。近年来,国内的门禁市场增长比较快速,门禁管理系统应用于现代化智能大厦的出入控制、公司考勤、安防报警等许多场合,也出现了许多生产门禁管理系统的公司,也研发出许多新型的门禁控制系统。国内外研发和使用的门禁控制系统主要都使用 RFID 门禁系统和生物识别门禁系统。从 1994 年 RFID(射频卡)进入中国,就带动了我国 RFID 卡的技术快速革新。在国内 RFID 的应用领域越来越广,它的应用也促进了门禁管理系统的发展。近年来我国经济总体水平明显提高,各行各业对于现代化、智能化安防设备的需求也会越来越多,安防产品也随之迅速发展。

目前在欧美的发达国家,它们的门禁系统市场已经进入到细化的阶段,其产业的分工越来越细化,生产卡的公司单独生产卡;生产控制器的公司独立生产控制器和配套的软件,都集中生产某一种产品,而不是生产所有的设备。最近几年,欧美各大公司也开始投入到无线门禁系统的研究设计、生产中来。正是由于 DSP、计算机数字化的大量普及,高性能的模拟数字电路的不断完善,各种保护电路的不断成熟改进,各种芯片的大幅降价,使得全世界的工程师才有机会设计出性能良好、结构完善、性价比极高的优秀产品,表现出门禁系统对安防工作以及对日常管理工作的真实贡献。

由于磁卡、接触式 IC 卡对操作环境的要求相对较高,磁卡及 IC 卡容易磨损,本身的工作时间也是一定的,再加上识别过程中使用者的动作和所需时间明显的

要多于非接触式卡，所以，IC卡和生物识别的门禁系统相比，有很多缺点。从科学技术长远的发展角度上看，基于说话人识别的无线门禁系统或与以上技术相结合的新型门禁产品将成为未来门禁系统发展的重要方向。

本文针对说话人识别门禁系统的现状，从研究与使用的角度出发，提出了基于说话人识别的无线门禁系统并进行相关的研究设计。主要研究设计是针对小区的单元门禁，包括每个小区单元的主机与用户端的分机，整个系统由一台主机和多个用户端共同完成。本文针对用户端进行了设计研究，音频编解码模块实现对语音信号的采集，FLASH 模块实现说话人模板的存储，主控模块实现说话人识别，无线模块实现用户端与主机之间的通信，用 LED 的状态来实现说话人识别的结果与电磁锁的状态。

### 1.3 研究的主要内容及论文结构

论文共分五章。

第一章为绪论，介绍选题的目的和意义，国内外研究说话人识别的主要方法、DSP 的发展和主要特点、门禁系统的研究现状以及论文的主要内容。

第二章详细阐述了门禁系统的设计方案，说话人识别的基本原理，说话人识别系统的构成、对说话人语音信号处理的流程。详细研究了动态时间规整及改进算法的原理。采用双门限检测法，通过对过零率和短时能量参数的检测来判断起始点和结束点，提取出语音信号数据。采用 Mel 倒谱 MFCC 提取语音信号的特征参数，这些特征参数组成特征矢量，从而构成参考的语音模板。在模式匹配中，采用了改进的动态时间归整 DTW 方法，将测试模板与参考模板进行匹配计算，比较两者之间的失真，最后计算出最小的匹配距离作为规整函数的路径，从而得出说话人识别判决的结果。

第三章为基于说话人识别门禁系统的硬件电路设计。以单元中各个用户为对象，详细设计了整个系统的硬件结构与元器件的选型标准，各个主要模块电路的构成与 PCB 制图，电路的抗干扰设计等。

第四章为系统的软件设计。基于 CCS 开发环境，采用并行 EMIF 引导方式，实现了 TMS320VC5509A DSP 扩展 FLASH 的在线烧写；完成各个主要模块的初始化过程，采用 C 语言完成语音识别工程及无线数据收发工程。

第五章介绍了系统软硬件调试过程，matlab 仿真结果与移植的实现结果。

最后为结论与展望，论述了论文的主要贡献以及在研究中所遇到的主要不足之处，指出接下来研究的主要方向。

### 1.4 本章小结

本章介绍了论文选题的目的和意义，课题的国内外研究现状以及论文研究的主要内容、论文结构等等。

## 第 2 章 门禁系统的设计方案与相关理论研究

本文根据实际应用,提出了基于说话人识别的无线门禁系统。主要目标是基于各个小区单元,以各个用户为单位,实现用户与楼宇单元门之间的无线通信并且实现用户端的说话人识别,以开启用户端自家的门锁,达到便捷性,降低成本等的目的。门禁系统的设计框图如图 2.1 所示,在用户端,通过采集说话人的语音信号,经过预处理,端点检测,特征提取与参考模板相比较,输出比较结果。如果匹配成功,则打开用户端的门锁。当在小区单元门口来访人员按下用户门铃时,主机会通过无线收发模块寻找用户端的 ID 号,进而建立主机与用户端的无线连接。如果用户允许来访人员进入,则用户端给主机发个开锁的信号,主机接收到这个信号后,打开单元门。要实现这些要求,要解决的主要理论问题是说话人识别的原理。

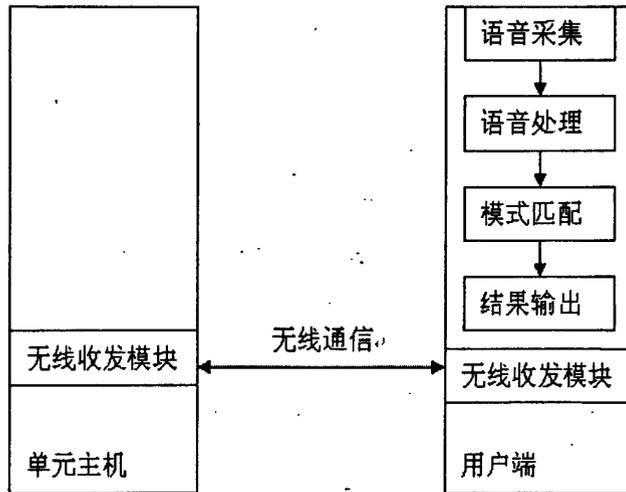


图 2.1 门禁系统框图

### 2.1 说话人识别原理

无论是小词汇、特定人、文本有关的说话人识别,还是连续词汇、非特定人、文本无关的说话人识别,其基本原理都是先要建立说话人语音信号特征参数的参考模板。在进行说话人识别时,先计算参考模板与测试模板两帧之间的帧匹配距离矩阵,再计算出最小的累积距离,作为规整函数的路径。因此,若要实现对说话人的识别,需解决以下几个基本问题:

- (1) 对语音信号的加重、分帧、加窗、端点检测和特征提取;
- (2) 建立说话人的参考模板和测试模板;
- (3) 计算测试模板与参考模板的最小累积距离。

其中关键问题是建立说话人识别的参考模板,为了保证测试模板与参考模

板的特征参数一致，常将每一说话人语音信号进行预加重、分帧、加窗等，它们的模型结构取得相同，不同的只是模型中的参数<sup>[1]</sup>。伴随着计算机和语音信号处理、高速的 DSP 芯片等外围电路的不断发展，说话人模型已经从单一的 DTW、VQ、HMM、ANN 等模型向着多模型方向发展，同时也从实验室无噪声的背景下向着复杂噪声的应用场合发展，说话人识别从实验环境走向实用有了坚实的理论基础。

从本质上讲，说话人识别是一个语音信号的模式识别问题，它由训练和识别两个过程来完成。说话人识别的基本原理框图如图 2.2 示：

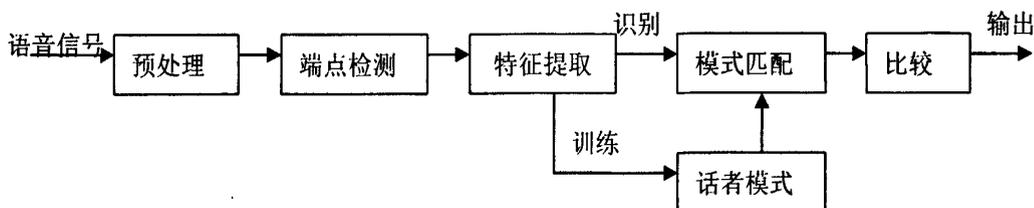


图 2.2 说话人识别原理图

## 2.2 语音信号的预处理

在对语音信号进行时域或频域分析前，首先要对语音信号进行一定的预处理，预处理主要包括对语音信号的预加重、分帧和加窗。

### 2.2.1 预加重

说话人的语音信号经过 A/D 将模拟信号变为数字信号后，依次存入一个数据区，在语音信号处理中一般用循环队列的方式来存储这些数据。根据语音信号产生的机制，高频端大约在 800Hz 以上按 6dB/倍频程跌落，这就需要进行预加重来提升高频部分，使信号的频谱变得平坦<sup>[1]</sup>。预加重一般是一阶的数字滤波器：

$$H(Z)=1-\mu z^{-1} \quad u \text{ 值接近为 } 1 \quad (2.1)$$

$u$  决定截止频率，此滤波器为单零点滤波器，呈高通特性， $u=0.95$ 。

### 2.2.2 分帧与加窗

在完成预加重之后，要对语音信号要进行分帧与加窗处理。根据语音信号的短时平稳性，语音信号以帧为单位进行处理。一般每秒的帧数为 33~100 帧，要根据实际情况而定。分帧应该采用部分互相交错的形式，目的是为了使语音信息损失最小，帧与帧之间平滑的过渡，保持其连续性。前一帧与后一帧的重合部分称为帧移。帧移与帧长的比值一般取为 0~1/2，帧与帧之间重合太多，会增大计算量，重合太少会漏掉部分信息。分帧是通过一个有限长度窗口进行加权的方法来实现，用窗函数  $w(n)$  乘语音信号  $s(n)$ ，得到  $sw(n)=s(n)*w(n)$ 。本文采样频率为 8KHz，每帧长 240 点 (30ms)，帧移 80 点，采集 3 秒的数据，约为 120 帧。

窗函数  $w(n)$  长度与形状的选择，影响到短时参数的分析。因此在对话话人

的语音信号进行预处理时要选择合适的窗函数,使其短时参数更好的反映语音信号的特性变化。常用的窗函数主要是矩形窗和汉明窗,它们的表达式如下(其中N为帧长):

$$\text{矩形窗: } \omega(n) = \begin{cases} 1, & 0 \leq n \leq (N-1) \\ 0, & n = \textit{else} \end{cases} \quad (2.2)$$

$$\text{汉明窗: } \omega(n) = \begin{cases} 0.54 - 0.46 \cos[2\pi n / N - 1], & 0 \leq n \leq N-1 \\ 0, & n = \textit{else} \end{cases} \quad (2.3)$$

矩形窗和汉明窗的比较数据如表 2.1 所示:

表 2.1 矩形窗与汉明窗的比较<sup>[1]</sup>

| 窗类型 | 旁瓣峰值 | 主瓣宽度     | 最小阻带素衰减 |
|-----|------|----------|---------|
| 矩形窗 | -13  | $4\pi/N$ | -21     |
| 汉明窗 | -41  | $8\pi/N$ | -53     |

从表 2.1 可以看出,汉明窗的带宽增加一倍,能减小时间窗两端的坡度,使得语音信号能够很平滑的过渡,语音信号变化的幅度不是很大,减小语音帧的截断效应。所以本文选择汉明窗。

### 2.3 语音信号的端点检测

从背景的噪声环境中提取出说话人语音信号的特征信息不是一件容易的事情,一个比较好的端点检测算法可以精确的检测到语音信号的开始点与结束点,在一定程度上能够有效提高说话人识别系统的识别率。信号的幅度用短时能量来表示,过零率定义为一帧语音信号穿越零电平的次数,两者的结合可以有效的检测到语音信号的起始点和终止点,故本文采用短时能量和过零率的双门限检测法来进行端点检测,并把可容忍的静音区间延长为 15 帧。

设语音波形的时域信号为  $x(l)$ , 加窗分帧处理后得得的第  $n$  帧语音信号为  $x_n(m)$ , 则  $x_n(m)$  满足:  $x_n(m) = w(m)x(n+m) \quad 0 \leq m \leq N-1$  (2.4)

其中,  $w(m)$  为汉明窗,  $n=0,1T,2T,\dots$ , 并且  $N$  为帧移长度。

设第  $n$  帧语音信号  $x_n(m)$  的短时能量用  $E_n$  表示, 则

$$E_n = \sum_{m=1}^N |x_n(m)| \quad (2.5)$$

$$E_n = \sum_{m=1}^N x_n^2(m) \quad (2.6)$$

$$E_n = \sum_{m=1}^N \log x_n^2(m) \quad (2.7)$$

表达式分别为绝对值的累加、平方累加和平方的对数的累加,只是表达方式的不同,实质上是一致的,选择其中的一种就可以。根据人的发音机理和声道模型,人的语音信号分清音和浊音两种。浊音的幅度相对高,相对的能量也比较

大，而清音的幅度比较小，所以可以用短时能量来区分清音段与浊音段。根据实验及前人的工作总结，短时能量高门限的值为低门限的 5 倍。

通过观察比较能够发现，在静音段语音信号的短时能量低门限很容易被超过，能量比较低，波形的变化相对平缓；而在语音段，语音信号的能量已经超过了双门限中的高门限，幅度变化比较剧烈，通常用一帧信号中波形穿越零电平的次数来表达这种变化的程度，称为过零率。它的公式是：

$$\begin{aligned} Zn &= \sum_{m=-\infty}^{\infty} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]w(n-m)| \\ &= \sum_{m=n}^{n+N-1} |\operatorname{sgn}[x(m)] - \operatorname{sgn}[x(m-1)]w(n-m)| \end{aligned} \quad (2.8)$$

式中， $\operatorname{sgn}[]$ 是符号函数，即

$$\operatorname{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \quad (2.9)$$

而  $w(n)$  为窗口序列，设

$$w(n) = \begin{cases} \frac{1}{2N}, & 0 \leq n \leq N-1 \\ 0, & \text{其它} \end{cases} \quad (2.10)$$

在实际的应用中，静音段的随机噪声有可能会产生过高的过零率，往往先设定一个低门限，在静音段，如果短时能量或过零率超过低门限，就认为是起始点，进入到过渡段。在语音段，一些突发性的噪声也可能引起短时能量或过零率的数值较高，但是往往不能维持很长时间。如果两个参数的数值降低到低门限以下，且一直持续 15 帧，则认为是一段噪音而放弃<sup>[34]</sup>，语音进入停止。如果说话人的语音信号中包含 2 个字的时间间隔很长，双门限端点检测法有可能只检测到第一个字，从而导致语音信号端点检测的错误，漏掉需要检测的语音信息，使得语音信号的识别率下降。如图 2.3 所示，检测到语音信号中“你好”的第一个字，为了避免这种情况，这需要把双门限端点检测可容忍的静音区间适当的延长，本文把静音区间设置为 15 帧。假设 15 帧之内语音信号没有短时能量或过零率超过低门限，就认为语音信号结束；反之，则把短时能量和过零率计算在内语音段内。图 2.4 是以语音“你好”为例，是改进后的双门限端点检测，可以精确的检测到语音的起始点和终止点。

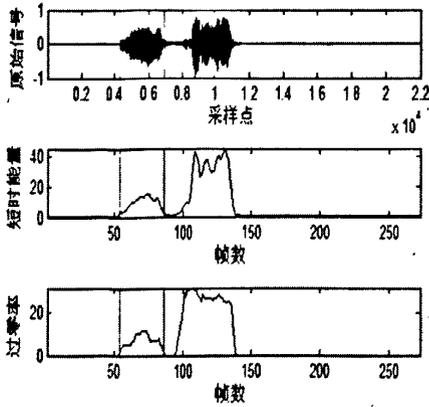


图 2.3 改进前的端点检测

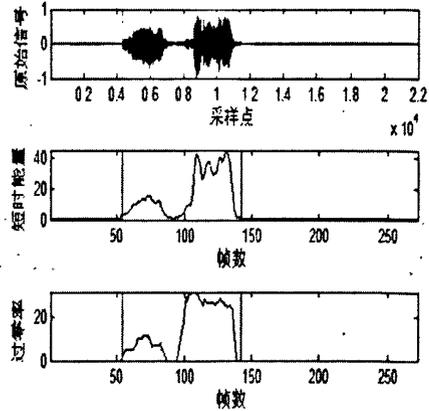


图 2.4 改进后的端点检测

根据前面的理论分析，短时过零率可以检测到语音信号的清音段，短时能量可以检测到浊音段，所以采用延长帧的双门限端点检测法可以准确的判断语音信号的起始点和结束点，使得端点检测更精确，丢弃一些噪声信号帧，减少系统的计算量，降低噪声对系统性能的影响。采用改进后的端点检测算法，对单个汉字或多个汉字的语音命令均识别正常。

## 2.4 语音信号的特征提取

在说话人识别时，特征参数的选取是至关重要的，能够表征说话人个性的特征有：语音帧能量、基音频率、语音信号的短时频谱、线性预测系数、倒谱特征、基因轮廓特征等。目前进行语音信号特征分析最有效、最重要的方法有线性预测系数和美尔倒谱系数。

### 2.4.1 线性预测系数

说话人的语音信号是一种典型的时变信号，如果语音信号的观察时间尽可能短，则语音信号可以看做是近似平稳的信号。全极点线性预测模型 LPC(Linear predictive coding, 简称 LPC)可以对声管模型进行很好的描述，它的基本思想是利用语音信号的相关性，用过去的取样值来预测现在或未来的取样值，即利用过去若干个语音信号的取样值与线性组合来逼近语音信号的取样值，LPC 是基于发音模型建立的。对于一个 LPC 系统，采样点  $n$  的输出  $s(n)$  用  $p$  个样本的线性组合来表示：

$$s(n) = a_1s(n-1) + a_2s(n-2) + \dots + a_p s(n-p), a_1, a_2, \dots, a_p \text{ 为常数} \quad (2.11)$$

$$\text{上式可简写为 } s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2.12)$$

其中  $Gu(n)$  是一个归一化冲击响应及其增益系数的乘积，该式的 Z 域表达式为：

$$S(z) = \sum_{k=1}^p a_k z^{-k} S(z) + GU(z) \quad (2.13)$$

系统的传递函数为:

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{1}{A(z)} \quad (2.14)$$

系统的估计误差为:

$$e_n = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (2.15)$$

相应的误差函数为:

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p a_k z^{-k} \quad (2.16)$$

误差平方和为:

$$E_n = \sum_m e_n^2(m) = \sum_m \left[ s_n(m) - \sum_{k=1}^p a_k s_n(m-k) \right]^2 \quad (2.17)$$

上式对各阶 LPC 系数求导数, 令其分别为 0, 可以得到:

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^p \hat{a}_k \sum_m s_n(m-i)s_n(m-k) \quad (2.18)$$

根据相关函数的定义, 可以得到自相关函数为:

$$\varnothing_n(i, 0) = \sum_{k=1}^p \hat{a}_k \varnothing_n(i, k), k = 1, 2, \dots, p \quad (2.19)$$

该式表示 p 个方程构成的方程组, 未知数为 p 个, 求该方程组就可以得到系统的线性预测系数。

#### 2.4.2 美尔倒谱系数 MFCC

LPC模型与线性预测倒谱系数LPCC(Linear Predictive Cepstral Coding)是基于发音模型建立的, 没有充分利用人耳的听觉特性。美尔倒谱系数(Mel-Frequency Cepstral Coefficients, 简称MFCC)着眼于人耳的听觉特性, 依据听觉实验的结果来分析语音信号的频谱, 有很好的识别率和噪声鲁棒性。Mel频率尺度的值和实际频率的值成对数分布关系。Mel频率与实际频率的具体关系可用(2.20)来表示:

$$MEL(f) = 2595 \lg(1 + f/700) \quad (2.20)$$

公式中 f 单位为 Hz。

M为滤波器的个数, 取M=24, N为一帧语音信号的点数, 本文取N=240点。滤波器在频域上为简单的三角形, 中心频率为  $f_m$ , 它们在Mel频率轴上是分布均

匀的。根据Zwicker的工作，临界频率带宽随着频率的变化而改变，并与Mel频率的增长一致。根据临界频带的划分，可将语音信号的频率划分成一系列三角形的滤波器，即Mel滤波器组，如图2.5所示。在Matlab中，调用了melbank.m函数，用来计算M个滤波器的系数，语法为：

```
h=melbankm(16,240,8000);
```

```
plot(h');
```

显示的结果就是图2.5。

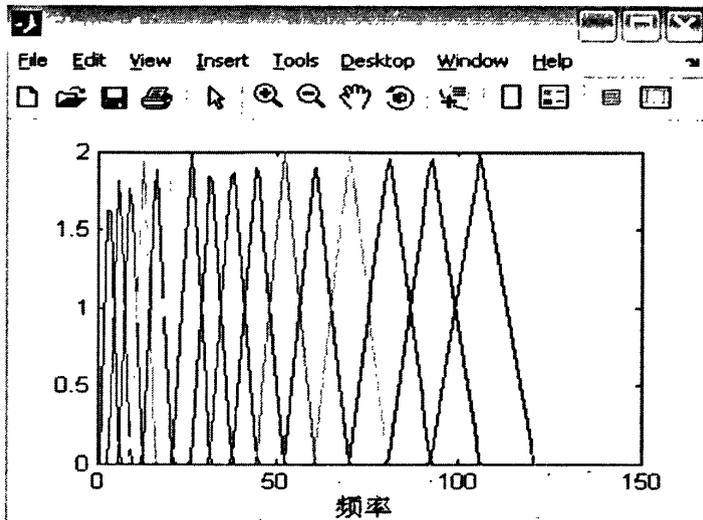


图 2.5 Mel 滤波器组

Mel滤波器组的中心频率分别为  $f(m), m=1,2,\dots,M$ ，Mel滤波器有两个作用，一是把实际的频率映射到Mel频域上去；二是对能量谱进行滤波，它每一个滤波器覆盖的范围都近似于人耳的临界带宽，以此来模拟人耳的掩蔽效应。则Mel滤波器组的频率响应公式是：

$$H_m(k) = \begin{cases} 0 & k \leq f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad 0 \leq m \leq M \quad (2.21)$$

其中  $\sum_{m=0}^{M-1} H_m(k) = 1$ ， $f(m)$  是每个滤波器的中心频率，其计算公式为：

$$f(m) = \left(\frac{N}{f_s}\right) B^{-1} \left( B(f_l) + m \frac{B(f_h) - B(f_l)}{M+1} \right) \quad (2.22)$$

$f_h$  和  $f_l$  是滤波器组应用的最高频率和最低频率,  $f_s$  是采样速率。  $B^{-1}$  是函数  $B$  的反函数。  $B$  的计算公式如下

$$B(f) = 1125 \times \ln(f/700 + 1) \quad (2.23)$$

MFCC计算过程如图2.6所示, 具体计算过程如下:

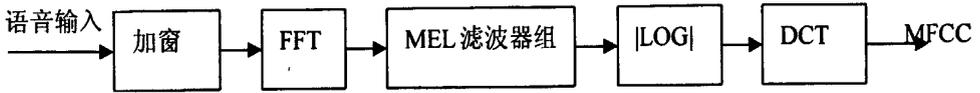


图 2.6 MFCC 计算流程图

(1) 首先要计算一帧语音信号的点数, 本系统取  $N=240$  点。对语音信号每帧序列  $s(n)$  进行预加重处理后再经过离散FFT变换, 把时域的语音信号变到频域, 取模的平方得到离散功率谱  $S(n)$ 。

(2) 计算  $S(n)$  通过  $M$  个  $H_m(n)$  后所得的功率值, 即计算  $S(n)$  和  $H_m(n)$  在各离散频点上乘积之和, 得到  $M$  个参数并取自然对数得到  $S(m)$

$$S(m) = \ln \left( \sum_{k=0}^{N-1} |X(k)|^2 H_m(k) \right) \quad 0 \leq m \leq M \quad (2.24)$$

(3) 对得到的对数频谱进行离散余弦变换, 得到  $n$  阶 Mel 频率倒谱系数  $c(n)$ :

$$c(n) = \sum_{m=0}^{M-1} S(m) \cos\left(\frac{\pi m(m + \frac{1}{2})}{M}\right) \quad 1 \leq m \leq M, \quad n = 1, 2, 3 \dots N \quad (2.25)$$

MFCC利用了人耳听觉系统的临界带效应, 描述了人耳对频率高低的非线性心理感觉, 识别性能和抗噪性能均明显优于传统的线性预测倒谱参数LPCC, 是目前国内外最重要、最实用的用于说话人识别的特征参数。

## 2.5 语音信号的模式匹配

完成特征提取之后要进行模式识别, 对于小词汇, 特定人识别系统, DTW显示了独特的优越性, 算法相对简单, 计算量较少, 故本文采用DTW算法。假设测试模板用  $T$ , 参考模板用  $R$  表示, 它们之间的相似度用它们之间的距离  $D[T, R]$  来度量, 距离越小相似度就越高<sup>[33]</sup>。首先要计算两帧之间的帧匹配距离, 设  $n$  和  $m$  分别是  $T$  和  $R$  中任意选择的帧号,  $D[T(n), R(m)]$  表示这两帧特征矢量之间的距离。

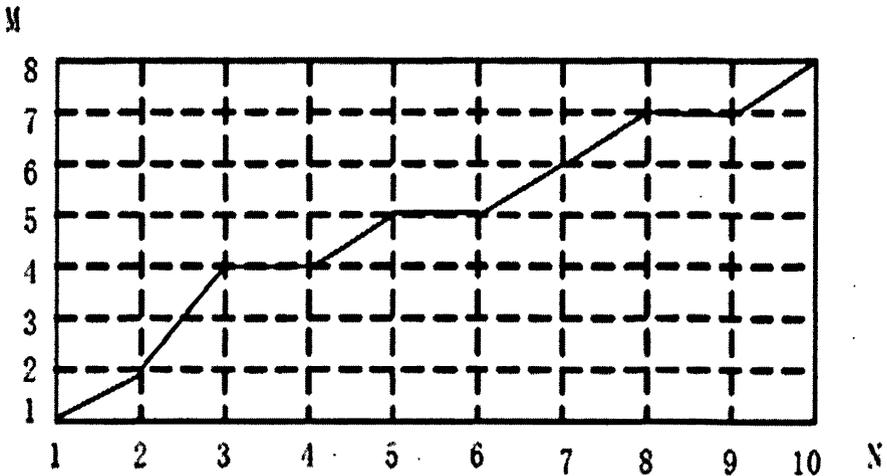


图 2.7 DTW 算法搜索路径

如图2.7所示，可以把测试模板和参考模板映射到二维坐标系中，横轴代表测试模板的帧号，纵轴代表是参考模板的帧号。DTW算法的实质就是寻找一条通过此各点的路径，使得累计距离为最小。其中规整函数必须满足三个约束条件：

- (1) 边界条件，路径的起始点是(1,1)，结束点是(N,M)；
- (2) 单调性限制，因为说话人的发音是有先后顺序的，所以规整函数必须是从左下角到右上角，不能违背时间顺序的原则；
- (3) 连续性约束，在横轴上的测试模板，每一帧都可能代表着说话人的特征信息，所以为了保证信息损失最小，规整函数不能跳过任何一帧。

根据规整函数的约束条件以及前人的经验总结，为了使路径不至于过分倾斜，约束斜率在0.5~2的范围内。求规整函数最佳路径的问题即为：

$$\sum_{\substack{n=1 \\ m=\Phi(n) \leq \eta}}^N D[n, m_i] = \min_{\Phi(\bullet)} \sum_{\substack{n=1 \\ m=\Phi(n) \leq \eta}}^N D[n, m_i], \text{ 式中 } m_i = \hat{\Phi}(n) \text{ 为最佳路径函数}$$

$$D[(n_i, m_i)] = d[T(n_i), R(m_i)] + D[(n_i - 1, m_i - 1)] \quad (2.26)$$

其中的  $D[(n_i - 1, m_i - 1)]$  由下式决定；

$$D[(n_i - 1, m_i - 1)] = \min\{D[(n_i - 1, m_i)], D[(n_i - 1, m_i - 1)], D[(n_i - 1, m_i - 2)]\} \quad (2.27)$$

计算DTW的过程，首先要分配2个 $N \times M$ 的矩阵，一个存放两个模板中各帧之间的距离，及求出帧匹配距离矩阵，另一个存放各帧之间的累积距离矩阵，搜索路径从左下角到右上角，最后求出一条最佳路径，这保证了它们之间存在最大的声学相似特性<sup>[1]</sup>。

DTW算法虽然简单有效，但是需要两个 $N \times M$ 的矩阵，占用的存储空间也比较大，计算量相应增大。但是在实际的应用中，有很多格点实际上是达不到的，通常，规整函数  $\Phi(n_i)$  被限制在一个平行四边形内，斜率为1/2~2，规整函数的起点从左下角(1,1)点开始，终止点是右上角点(N,M)。如图2.8所示。

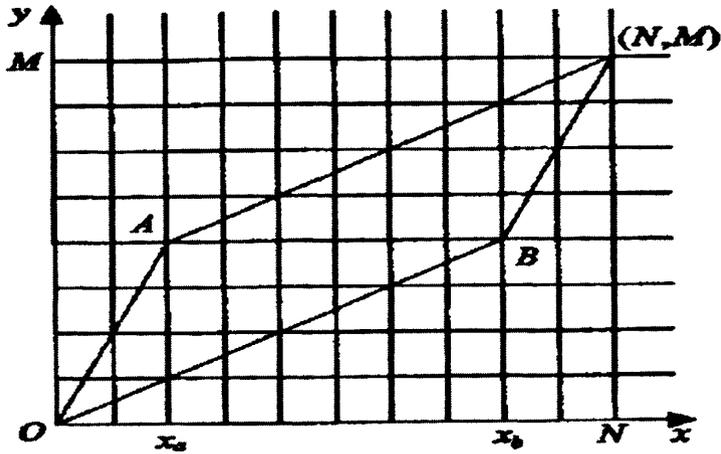


图 2.8 匹配路径约束示意图

根据图2.8所示，实际的动态弯折为三段， $(1, X_a)$ ,  $(X_a+1, X_b)$ 和 $(X_b+1, N)$ ，其中：

$$\begin{cases} X_a = \frac{1}{3}(2M - N) \\ X_b = \frac{2}{3}(2N - M) \end{cases} \quad (2.28)$$

由于 $X_a$ 和 $X_b$ 代表的是帧数，所以取整数。根据约束条件有：

$$\begin{cases} 2M - N \geq 3 \\ 2N - M \geq 2 \end{cases} \quad (2.29)$$

当不满足以上条件时，认为两者差别实在太大，无法进行动态弯曲匹配。

在X轴上每前进一帧，只需用到前一列的累积距离，而不需要与Y轴上的每一帧进行比较，而只是与Y轴上 $[y_{min}, y_{max}]$ 间的帧进行比较。 $y_{min}$ 和 $y_{max}$ 的计算如下：

$$y_{min} = \begin{cases} \frac{1}{2}x, 0 \leq x \leq X_b \\ 2x + (M - 2N), X_b < x \leq N \end{cases} \quad (2.30)$$

$$y_{max} = \begin{cases} 2x, 0 \leq x \leq X_a \\ \frac{1}{2}x + (M - \frac{1}{2}N), X_a < x \leq N \end{cases} \quad (2.31)$$

如果出现 $X_a > X_b$ 的情况，此时弯折匹配的三段为 $(1, X_b)$ ， $(X_b+1, X_a)$ 和 $(X_a+1, N)$ 。累积距离的计算可以通过式2.32不断更新来实现。

$$D(x, y) = d(x, y) + \min[D(x-1, y), D(x-1, y-1), D(x-1, y-2)] \quad (2.32)$$

由于X轴上每前进一帧，只需要用到前一列的累积距离，整个DTW算法的计算过程递推循环进行，然后重复使用，而不用保存所有的帧匹配距离和累积距离，这样可以大大减少存储空间，降低计算量，从而提高说话人的识别率。

## 2.7 本章小结

本文针对的目标是小词汇系统的说话人识别,所以本章对语音信号各个组成部分进行设计研究,包括对语音信号的分帧、加窗、预加重等。为了有效提高说话人识别系统的识别率,采用延长帧的双门限检测法来判断语音信号的起始点,分析提取MFCC倒谱系数的具体方法,着重讨论了DTW算法及改进的高效算法,对后续的硬件设计提供理论依据。

## 第 3 章 系统的硬件设计

根据说话人识别算法的步骤,说话人识别门禁系统包括用户端与小区单元主机端两个主要部分。在用户端,通过采集说话人的语音信号,经过预处理,端点检测,特征提取与参考模板相比较,输出比较结果。如果匹配成功,则打开用户端的门锁。当在小区单元门口来访人员按下用户门铃时,主机会寻找用户端的 ID 号,进而建立主机与用户端的无线连接。如果用户允许来访人员进入,则用户端给主机发个开锁的信号,主机接收到这个信号后,打开单元门。本文主要对用户端的门禁系统进行设计,系统由 TLV320AIC23 语音编解码芯片对语音信号进行采集与量化, TMS320VC5509A DSP 为核心电路对语音信号进行处理, SDRAM 和 FLASH 分别用于数据存储和程序存储,键盘部分进行模式选择,控制部分由 4 个 LED 灯来实现。

### 3.1 系统的硬件结构

各用户端的系统硬件结构框图如图 3.1 所示,各主要模块为:语音输入模块,主要是采用麦克风实现;音频编解码模块 TLV320AIC23 实现语音信号的采集与传输; TMS320VC5509A DSP 处理器主要负责对语音信号的处理; SDRAM 是暂存程序运行中的数据; FLASH 保存说话人识别的参考模板;无线模块负责用户端与主机端之间的无线通信等。

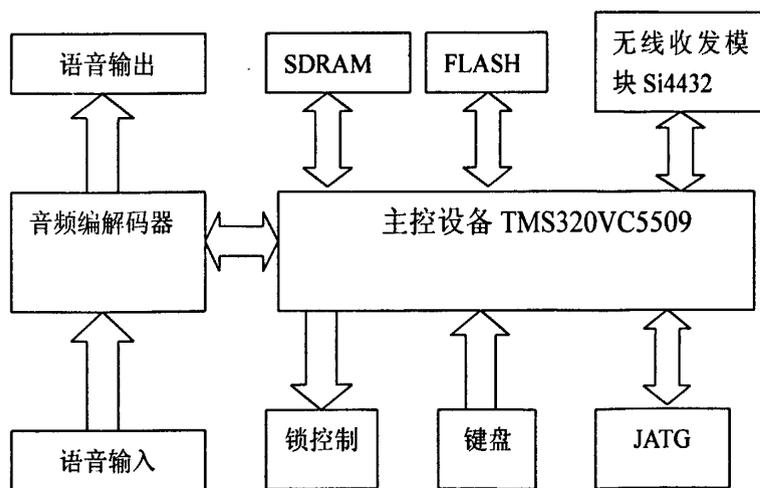


图 3.1 系统硬件结构框图

### 3.2 TMS320VC55x DSP 的基本结构

TMS320C55x 数字信号处理器是在 C54x 的基础上开发的新型低功耗、高性能的 DSP 芯片,其软件兼容 C54x 代码集,乘法器变成两个,极大地节省了 C54x 向 C55x 的转化时间。而采用最新的芯片制造技术大幅度提升了 DSP 的主频,从

而提高了 C55x 系列处理器的处理能力。C55x 系列的 DSP 功耗为 VC54xx 的 1/6, 性能为 VC54xx 的 5 倍。C55x 主要有三个部分组成: CPU、存储空间、片内外设。不同的芯片体系结构相同, 它们有相同的 CUP 内核, 但片上存储器和外围电路配置有所不同。与 C54x 相比, C55x 在硬件方面的扩展及 C55x 的特征和优点分别如表 3.1, 3.2 所示,

表 3.1 C54x 与 C55x 的比较

| 比较内容      | C54x       | C55x           |
|-----------|------------|----------------|
| 乘法单元(MAC) | 1          | 2              |
| 累加器(ACC)  | 2          | 4              |
| 读总线       | 2          | 3              |
| 写总线       | 1          | 2              |
| 地址总线      | 4          | 6              |
| 指令字长      | 16位        | 8/16/24/32/48位 |
| 数据字长      | 16位        | 16位            |
| 算术逻辑单元ALU | 1(40位)     | 1(40位)、1(16位)  |
| 辅助寄存器字长   | 2字节(每个16位) | 3字节(每个24位)     |
| 辅助寄存器     | 8          | 8              |
| 存储器空间     | 独立程序数据空间   | 统一程序数据空间       |
| 数据存储器     | 0          | 4              |

表 3.2 C55x 的特征和优点

| 特征   | 优点                                     |
|--|--|
| 一个32位x16位的指令缓冲队列   | 缓冲变长指令并完成有效的块重复操作                      |
| 两个17位x17位的乘法累加器  | 在一个单周期执行双累加操作                          |
| 一个40位桶形移位寄存器   | 实现高精度算数和逻辑操作                           |
| 一个16位算术逻辑单元  | 能够将一个40位的计算结果最高向左移31位或右移32位            |
| 4个40位的累加器  | 对主ALU并行完成简单的算数操作                       |
| 12条独立总线, 包括<br>3条读数据总线<br>2条写数据总线<br>5条数据地址线<br>1条读程序总线<br>1条程序地址线 | 为各种计算单元并行地提供将要处理的指令和操作数-利用C55x的并行机制的优点 |
| 用户可配置IDLE域   | 改进了低功耗电源管理的灵活性                         |

### 3.2.1 C55x 的 CPU 结构

C55x 系列具有统一的 CPU 内核，5 个功能单元：

(1) 存储器接口单元(M 单元)。M 单元是一个内部数据流、指令流接口，管理所有来自 CPU、数据空间或 I/O 空间的数据和指令，负责 CPU 和数据空间或 CPU 和 I/O 空间的数据传输。

(2) 指令缓存单元(I 单元)。此单元将指令流从存储器送到 CPU。

(3) 程序流程单元(P 单元)。此单元由程序地址产生电路和寄存器组成。程序流程单元产生所有程序空间的地址，并控制指令的读取顺序。在 P 单元中使用的寄存器分为 5 种类型：程序流寄存器、块重复寄存器、单重复寄存器、中断寄存器和状态寄存器。

(4) 地址数据流单元(A 单元)。A 单元为访问数据空间读和写产生地址。该单元包含 3 组数据读地址总线和两组数据写地址总线产生地址所必需的所有逻辑与寄存器，还包括带有移位能力的通用 16-bit 算数逻辑单元，提高了指令的并行性。

(5) 数据运算单元(D 单元)。这是 DSP 的心脏，对所处理的数据执行算数运算。它包括 MAC、主 ALU 以及累加寄存器。另外还包括一个桶形移位器、舍入和饱和控制以及用于算法的专门硬件。

图 3.2 是 C55x 的 CPU 结构图，采用改进的哈佛结构，共有 12 条总线，实现 CPU 对外部存储器的访问，使得 C55x 具有强大的处理功能<sup>[35]</sup>。

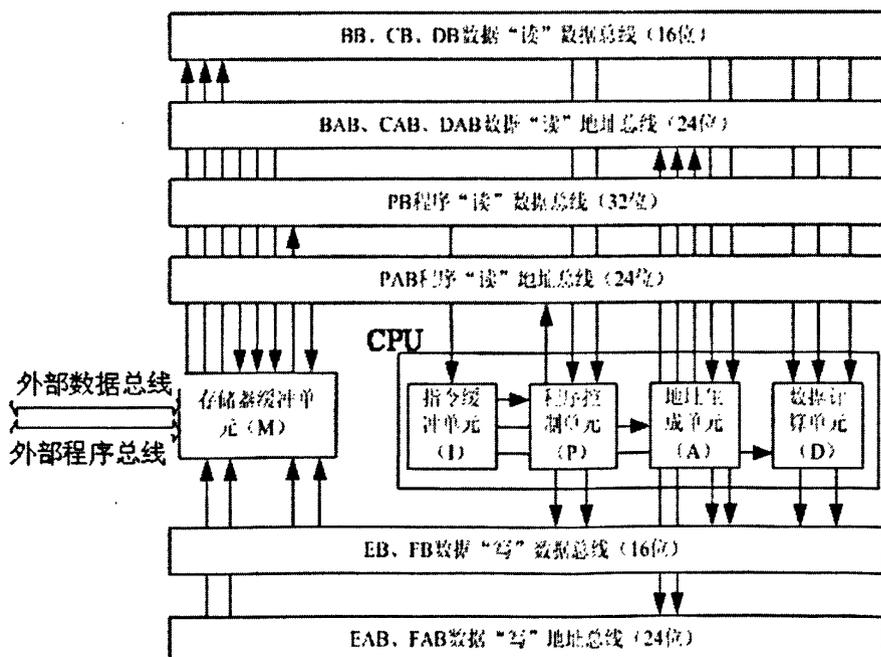


图 3.2 C55x 的 CPU 结构图

### 3.2.2 存储器映射

TMS320VC5509A DSP 存储器映射图如图 3.3 所示, 它的地址寻址有比较特殊的地方, TMS320VC5509A DSP 采用采用统一的编址方式, 在图 3.3 中, 左边的是按照程序存储器空间寻址, 为 16M\*8bit; 右边的是按照数据空间寻址, 为 8M\*16bit。当存储器按照程序存储空间寻址时, 地址编码采用字节寻址方式, 即每8位存储器占有一个地址编号, 此时A0有效, 而按照数据存储器空间使用时, 地址编码采用字寻址方式, 即 16 位存储器占有一个地址编号, 此时 A0 无效<sup>[42-43]</sup>。

| Byte Address (Hex) <sup>†</sup>   | Memory Blocks   | Block Size  |  |           |
|-----------------------------------|---|---|--|-----------|
| 000000                            | MMR (Reserved)  |   |  |           |
| 0000C0                            | -----   |   |  |           |
|                                   | DARAM / HPI Access  | (32K - 192) Bytes   |  |           |
| 008000                            | DARAM <sup>‡</sup>  | 32K Bytes   |  |           |
| 010000                            | SARAM <sup>§</sup>  | 192K Bytes  |  |           |
| 040000                            | External <sup>¶</sup> - $\overline{CE0}$  | 16K Bytes - Asynchronous<br>4M Bytes - 256K Bytes SDRAM <sup>#</sup>  |  |           |
| 400000                            | External <sup>¶</sup> - $\overline{CE1}$  | 16K Bytes - Asynchronous<br>4M Bytes - SDRAM  |  |           |
| 800000                            | External <sup>¶</sup> - $\overline{CE2}$  | 16K Bytes - Asynchronous<br>4M Bytes - SDRAM  |  |           |
| C00000                            | External <sup>¶</sup> - $\overline{CE3}$  | 16K Bytes - Asynchronous<br>4M Bytes - SDRAM (MPNMC = 1)<br>4M Bytes - 64K Bytes if internal ROM selected (MPNMC = 0) |  |           |
| FF0000                            | <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>ROM<sup>  </sup><br/>(if MPNMC=0)</td> <td>External<sup>¶</sup> - <math>\overline{CE3}</math><br/>(if MPNMC=1)</td> </tr> </table> | ROM <sup>  </sup><br>(if MPNMC=0)   | External <sup>¶</sup> - $\overline{CE3}$<br>(if MPNMC=1) | 32K Bytes |
| ROM <sup>  </sup><br>(if MPNMC=0) | External <sup>¶</sup> - $\overline{CE3}$<br>(if MPNMC=1)  |   |  |           |
| FF8000                            | <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>ROM<sup>  </sup><br/>(if MPNMC=0)</td> <td>External<sup>¶</sup> - <math>\overline{CE3}</math><br/>(if MPNMC=1)</td> </tr> </table> | ROM <sup>  </sup><br>(if MPNMC=0)   | External <sup>¶</sup> - $\overline{CE3}$<br>(if MPNMC=1) | 16K Bytes |
| ROM <sup>  </sup><br>(if MPNMC=0) | External <sup>¶</sup> - $\overline{CE3}$<br>(if MPNMC=1)  |   |  |           |
| FFC000                            | External <sup>¶</sup> - $\overline{CE3}$<br>(if MPNMC=1)  | 16K Bytes   |  |           |
| FFFFFF                            |   |   |  |           |

图 3.3 TMS320VC5509A 封装存储器映射图

### 3.2.3 DSP 的片内外设

DSP 的 CPU 主要完成运算与控制, 而片内外设完成的功能包括原始数据的采集、处理结果的输出、控制外部其他设备等。TMS320VC5509A 的主要片内外设有:

(1) 时钟与定时器, 包括时钟产生器、通用定时器、实时时钟以及看门狗定时器等。时钟发生器<sup>[36]</sup>通过 CLKIN 引脚获得输入时钟, 对 DSP 提供时钟信号。内部集成了一个数字锁相环 PLL, 通过配置可以得到期望的时钟信号。通用定时器、实时时钟及看门狗定时器的功能是通过计数器为系统提供时钟信号。

(2) 信号采集, 包括采集模拟信号的 ADC 和提供数字信号输入、输出功能的通用 I/O 接口。5509A 的 ADC 在一个时刻可以对最多 4 路模拟输入 (AIN0~AIN3) 的采样, 并用 10bit 的数字形式表示, 最大的采样率为 21.5kHz。因此, 很适合对慢速的模拟信号进行采样。5509A 有通用 I/O 引脚 GPIO 8 个并且可以配置。

(3) 外部设备连接接口, 包括外部存储器接口 (EMIF), 控制 DSP 与外部存储器之间的所有数据传输, EMIF 为异步存储器、同步突发 SRAM、同步 DRAM 提供了无缝接口。增强型主机接口 (EHPI), 是一个 16-bit 的并口, 外部主机通过它直接访问 DSP 内部的 DARAM。

(4) 通信接口, 包括多通道缓冲串口 (Multichannel Buffered Serial Ports, 简称 McBSPs)、I2C、异步串口、USB 接口以及多媒体卡 SD 卡接口。5509A 有 3 组 McBSP, 它可以连接串行存储器、A/D、D/A, 并可以通过该接口实现与其他处理器的高速串行连接。

(5) 其他外设, 包括 DMA 控制器、指令流水线等, 主要用来辅助 CPU 工作, 提高 DSP 的工作效率。

为了方便实现 5509A 对外设的控制, TI 公司为用户提供了片上外设支持库及控制片上外设的函数、宏等工具, 用户可以通过程序或 DSP/BIOS 完成这些函数和宏的调用。

### 3.2.4 C55x 的指令流水线和并行性

C55x CPU 采用指令流水线工作方式, C55x 的指令流水线包括两个阶段:

第一阶段是取指令阶段, 即从内存中取出 32 位的指令包, 放入指令缓冲队中, 然后为流水线的第二阶段提供 48 位的指令包。

第二阶段是执行指令阶段。这部分的功能是对指令进行解码, 完成数据的存取和计算。

图 3.4 为 C55x 流水线示意图, 采用指令流水线的执行方式, 可以大大提高系统的执行效率, 使得系统可以“低延迟”或“无延迟”执行较复杂的指令<sup>[36]</sup>。

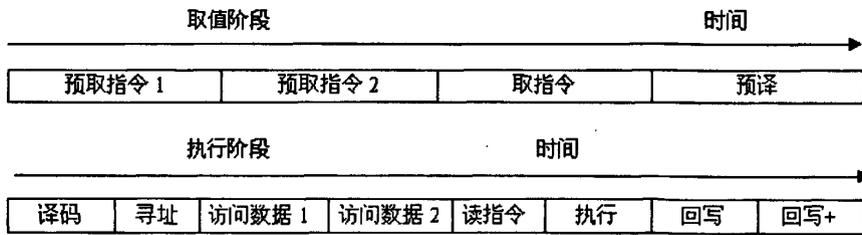


图 3.4 C55x 流水线示意图

### 3.3 各主要单元电路

#### 3.3.1 电源电路

电源是电路工作的能量来源和基础，而系统核心控制板正常工作需要 3.3V、1.6V 两组电压，VC5509A DSP 采用低电压设计，双电源供电，需要两组电源电压，即内核电源 CVDD 和 I/O 电源 DVDD。每种电源又可分为数字电源和模拟电源，相对应的也有模拟地和数字地。而整个系统芯片所需要的电压共三种为 +5V、+3.3V 及+1.6V。

由于TMS320VC5509的内核电压为1.6V，I/O电压为3.3V，因此该系统采用了TI公司的双路低压差电压调整器TPS767D301电源方案。它有两路电压输出，一路输出电压固定为3.3V，另外一路输出电压可调节，范围为1.5V~5.5V，电流输出范围是0mA~1.0A<sup>[37]</sup>。此外，电源芯片具有上电复位和监控功能。上电后在输出电压达到正常输出电压的95%时， $\overline{RST2}$  引脚就保持200ms的低电平信号。在系统运行正常时，如果电源电压VCC一旦降到该门限值以下，同样会产生一个200ms的低电平脉冲信号<sup>[38]</sup>。图3.5为系统电源电路图。

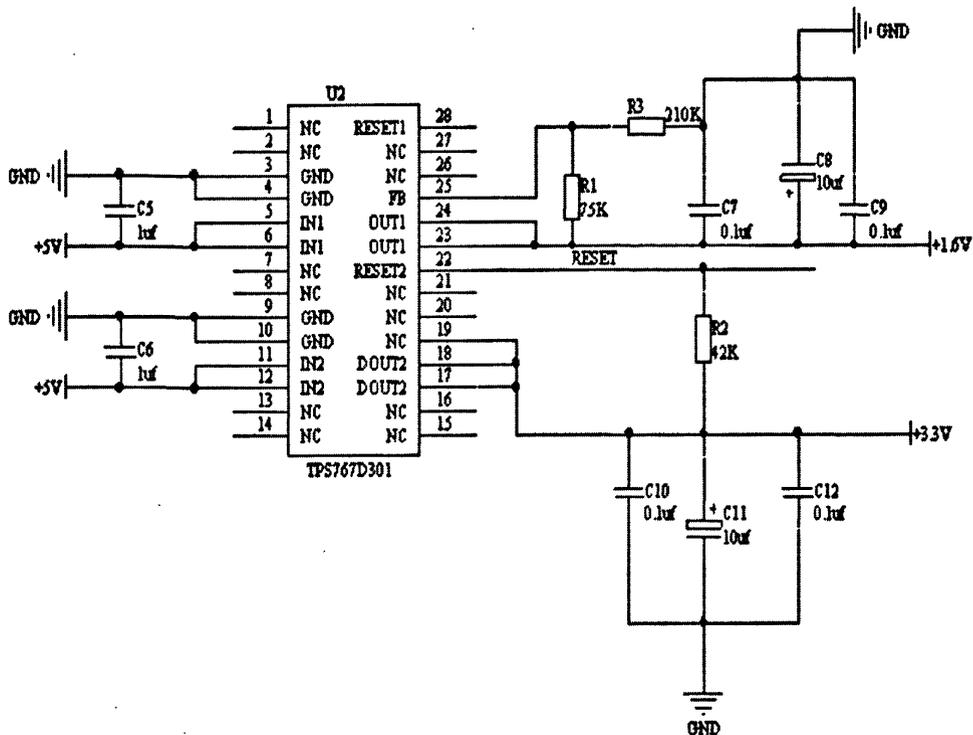


图 3.5 电源电路

图 3.5 中对于第二路可调输出电压 1.5V~5.5V，想要输出+1.6V 的电压，要通过一个分压电路来实现，根据图 3.5 可知：

$$\frac{V_o - V_{ref}}{R1} = \frac{V_{ref}}{R3} \quad (3-1)$$

$$V_o = V_{ref} \times \left(1 + \frac{R1}{R3}\right) \quad (3-2)$$

$$V_{ref} = 1.1834V \quad (3-3)$$

$$R1 = \left(\frac{V_o}{V_{ref}} - 1\right) \times R3 \quad (3-4)$$

要输出电压  $V_o = +1.6V$ ，由“式(3-4)”可计算出  $R1/R3 = 1/3$ ，故取  $R1 = 75K$ ， $R3 = 210K$  即可。

### 3.3.2 时钟电路

为 DSP 芯片提供时钟一般有两种方式：一种是使用外部时钟源，将外部时钟信号直接加到 VC5509 芯片的 X2/CLKIN 引脚，且 X1 引脚悬空；另外一种是用 DSP 芯片内部的振荡器构成时钟电路，在芯片的 X1 和 X2/CLKIN 引脚之间接入一个晶体，用于启动内部振荡器。由于 DSP 是核心的数据处理单元，对时钟的稳定性要求更高，所以本文采用有源晶振作为时钟源，它的实现通常是采用

一个电容和电感构成的 PI 型滤波网络。根据有源晶振的参考文档，有源晶体的接法是：一脚悬空，二脚接地，三脚输出，四脚接电压。时钟电路如图 3.6 所示：

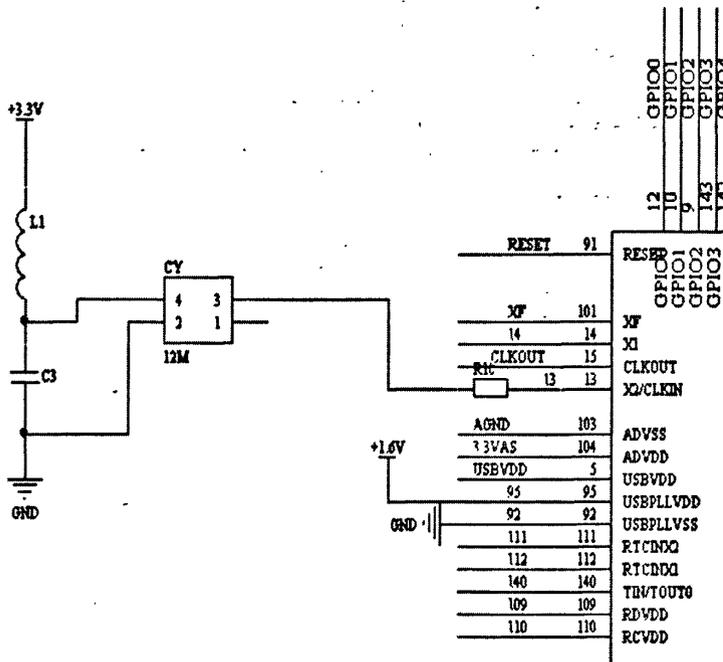


图 3.6 时钟电路

TMS320VC5509A DSP 内部集成了一个可以配置的数字锁相环(PLL)，可以通过配置时钟发生器得到所期望频率的时钟信号。它有一个时钟模式寄存器 CLKMD，用来控制和监视时钟发生器，有两种工作模式：旁路模式和锁定模式。

本设计中采用锁定模式，输出频率= $(PLL \text{ MULT}/(PLL \text{ DIV}+1)) \times$  输入时钟频率。时钟模式寄存器 (CLKMD) 中各个位符号如表 3.3 所示：

表 3.3 时钟模式寄存器 CLKMD

| 15   | 14  | 13  | 12   | 11-7     | 6-5     | 4          | 3-2        | 1        | 0    |
|------|-----|-----|------|----------|---------|------------|------------|----------|------|
| Rsvd | IAI | IOB | TEST | PLL MULT | PLL DIV | PLL ENABLE | BYPASS DIV | BREAK LN | LOCK |

输入时钟为 12MHz 分频，在锁定模式和旁路模式下，Rsvd 与 TEST 位都不适用，PLL MULT 取值范围是 2~31，PLL DIV 取值范围是 0~3，所以设置 CLKMD 为 6613h。

### 3.3.3 音频编解码模块

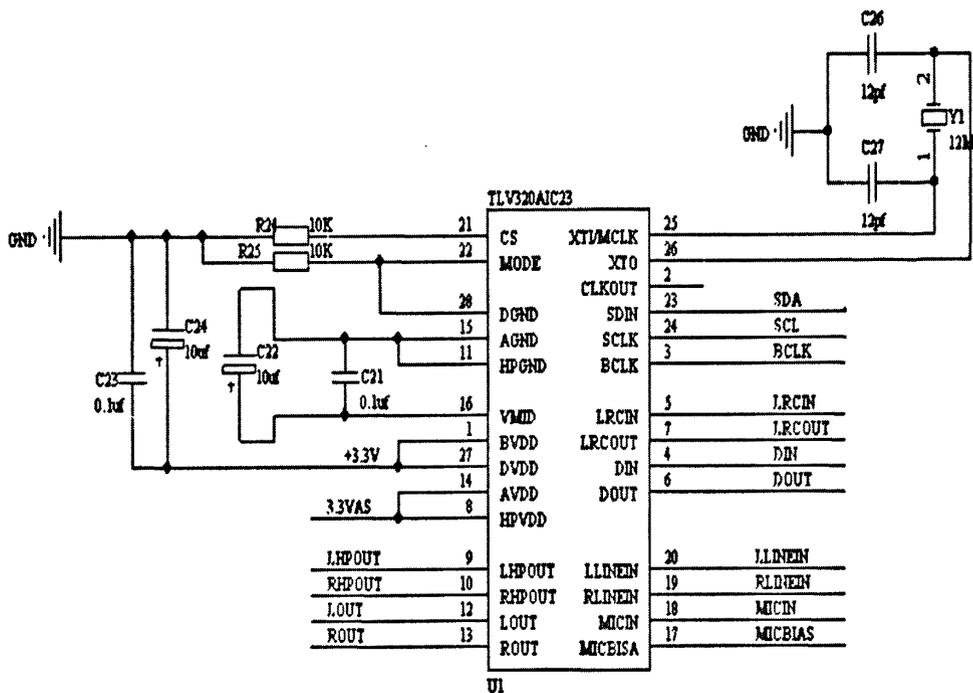


图 3.7 TLV320AIC23 芯片引脚图

TLV320AIC23 是 TI 公司生产的一种高性能立体音频编解码芯片<sup>[39-41]</sup>。它内部集成了 ADC 和 DAC，有两种输入方式 MIC 和 LINE IN，可以在 8~96kHz 的频率范围内提供了 16 位、20 位、24 位和 32 位的采样<sup>[42-44]</sup>。AIC23 还具有低功耗的特点，回放模式下功耗仅为 23mW，省电模式下更是小于 15uW。说话人的语音信号频率在 300Hz-3400Hz，根据乃奎斯特定理，采样频率应该为最高频率的 2 倍，在本文中通过设置 AIC23 的寄存器，以 8KHz，16 位量化对语音信号进行采集，并且把采集到的语音数据传送给 DSP 进行处理。AIC23 的引脚连接如图 3.7 所示。

AIC23 是 28 引脚的 TSSOP 封装，主要引脚的功能如表 3.4 所示：

图 3-4 AIC23 引脚功能

| 引脚      | 优点                           |
|---------|------------------------------|
| BCLK    | AIC23主模式，时钟由AIC23产生，反之由DSP产生 |
| LRCIN   | I2S DAC字时钟信号                 |
| LRCOUT  | I2S ADC字时钟信号                 |
| DIN     | 输入到DAC的I2S格式串行数据             |
| DOUT    | 从ADC输出的I2S格式串行数据             |
| MICBIAS | 为麦克风输入提供偏置电压                 |
| MICIN   | 麦克风输入端                       |
| LLINEIN | 左声道LINE IN输入，增益可调，默认值是0dB    |
| RLLNEIN | 右声道LINE IN输入，增益可调，默认值是0dB    |
| LHPOUT  | 左声道耳机放大输出                    |
| RHPOUT  | 右声道耳机放大输出                    |
| SDIN    | 控制口串行数据输入                    |
| SCLK    | 控制口串行数据时钟信号                  |

在对 TLV320AIC23 进行编程时，要了解内部寄存器的地址和主要功能，进而才能把 AIC23 配置为需要的模式。内部寄存器的含义如表 3.4 所示：

表 3.4 TLV320AIC23 内部寄存器

| ADDRESS | REGISTER |
|---------|----------|
| 0000000 | 左声道输入控制  |
| 0000001 | 右声道输入控制  |
| 0000010 | 左耳机通道控制  |
| 0000011 | 右耳机通道控制  |
| 0000100 | 模拟音频通道控制 |
| 0000101 | 数字音频通道控制 |
| 0000110 | 启动控制     |
| 0000111 | 数字音频格式   |
| 0001000 | 样本速度控制   |
| 0001001 | 数字界面激活   |
| 0001111 | 初始化寄存器   |

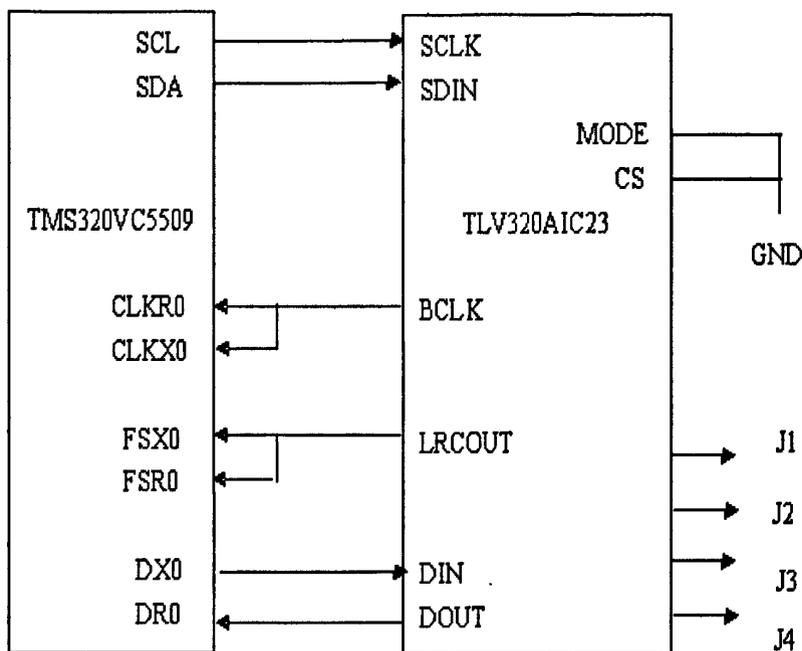


图 3.8 TMS320VC5509A 与 AIC23 接口示意图

TMS320VC5509A 与 TLV320AIC23 的连接如图 3.8 所示。把 AIC23 配置为主模式，时钟为 12MHz，8KHz 采样，16 位量化，USB 模式、耳机音量增益为 6dB，打开数字界面。工作步骤：

(1)初始化设置：AIC23 控制接口有两种工作模式， $I^2C$  模式与 SPI 模式。当 MODE=0 时，选择  $I^2C$  模式。DSP 通过  $I^2C$  总线配置 AIC23，把 AIC23 配置为 8KHz 采样，16 位量化；

(2)语音信号的输入：通过设置 AIC23 芯片的 11 个寄存器，通过 A/D 转换把数据发送到 DSP 的 McBSP0 接口上，数据保存在 SDRAM 中，配置到 DSP 的 CE0 空间，地址从 20000 开始；

(3)语音信号的输出：DSP 可以从 CE0 空间中将语音数据通过多通道缓冲串口 McBSP0 传给 AIC23，AIC23 的 DA 器件将它们变成模拟信号通过耳机输出<sup>[43]</sup>。

AIC23 通过两个独立的通道进行通信。一路控制 AIC23 的端口配置寄存器；另一路发送和接收数字音频信号。利用 5509A 的  $I^2C$  总线作为单向控制通道，控制通道只在配置 AIC23 时才用，当传输语言信号时一般是空闲的。McBSP 被用作双向数据通道，所有的语音数据都是通过数据通道传送的。

图 3.9 为采用 7-bit 地址传送 8-bit 数据的数据传输时序图。SDI 用来传输串行数据，SCLK 作为串行时钟。对数据的操作，必须严格按照  $I^2C$  的传输协议进行操作。当 SCLK 处于高电平，SDI 的下降沿启动数据发送；开始启动后首先传送 7 位的地址，由它决定哪个设备接收数据。R/W 决定数据发送的方向，识别到地址的器件在第 9 个时钟期间把 SDI 拉低，通知要进行数据传送，接下来时钟传送 2 个 8 位的数据块；数据传送完毕，SCLK 为高电平，SDI 的上升沿停止

传输<sup>[43]</sup>。配置为：DSP 为主发送模式，时钟分频寄存器 I2CCLKL 和 I2CCLKH 的值均设为 10。I<sup>2</sup>C 主要功能是为了配置 AIC23，配置完成后就处于空闲状态。

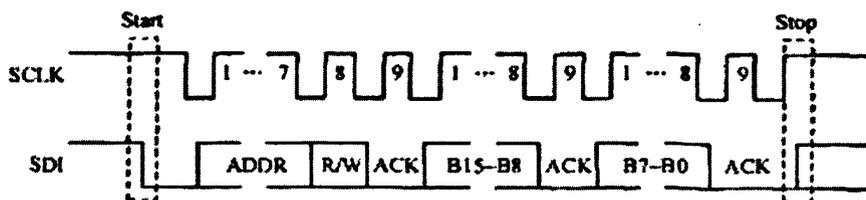


图 3.9 I<sup>2</sup>C 模块数据传输图

DSP 的多通道缓冲串口 McBSP0 作为双向数据通道，首先要进行初始化，才能发送和接收数字音频信号。AIC与DSP的接口主要有4种：右判断模式、左判断模式、I<sup>2</sup>S 模式和 DSP 模式。由于本文所用 TMS320VC5509A 的缓冲串口与 AIC23 相兼容，所以 McBSP0 被配置为 DSP 模式：AIC23 的时钟信号 LRCIN 和 LRCOUT 必须与 McBSP0 的帧同步信号相连，根据它的数据手册，在 LRCOUT/LRCIN 的下降沿开始串口数据传送；先传送左通道的数据字，再传送右通道的数据字<sup>[43]</sup>。每传送完一个字内部会产生一个中断，通知 DSP 要保存数据，为下一步数据处理做准备。要传送的字长由 IWL 寄存器来决定，McBSP0 与 AIC23 之间的数据传输，采用 32-bit，单帧无延时，下降沿发送数据。图 3.10 为 AIC23 在 DSP 模式下的串口时序图。

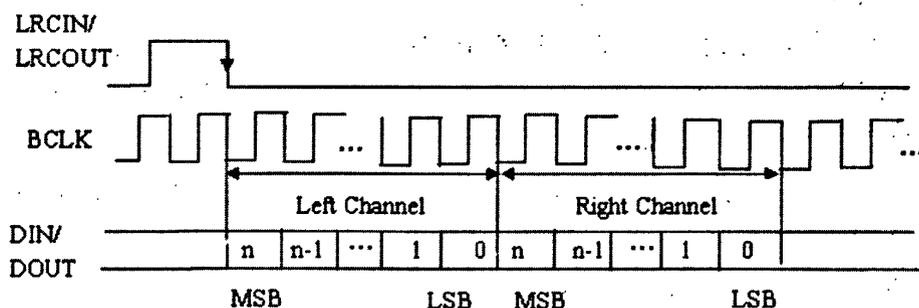


图 3.10 AIC23 在 DSP 模式下的串口时序图

### 3.3.4 外扩存储电路

TMS320VC5509 访问片外存储器时必须通过外部存储器接口 EMIF(External MEMORY Interface)。它具有很强的接口能力，不仅具有很高的数据吞吐率，而且通过 EMIF 接口可以灵活地和各种同步或异步存储器件进行无缝连接<sup>[45]</sup>。EMIF 除了对异步存储器的支持以外，还提供了对同步突发静态存储器(SBSRAM)和同步动态存储器(SDRAM)的支持。异步存储器可以是静态随机存储器(SRAM)、只读存储器(ROM)和闪存存储器(Flash)等。

#### 1、SDRAM

在设计中，选用了hynix公司的 1M\*4Banks\*16bit 的同步突发动态存储器——HY57V641620FTP<sup>[46]</sup>，工作电压为 3.3V，兼容 LVTTL 接口，常见封装是 54

脚 TSOP 封装, 时钟频率最大为143MHz, 读写周期最小为7ns, 64ms 的刷新周期。引脚说明如表 3.5 所示。

表 3.5 HY57V641620 引脚说明

| 引脚         | 引脚名         | 说明                                      |
|------------|-------------|---|
| CLK        | 时钟          | 系统时钟输入, 在CLK信号的上升沿输入其它信号                |
| CKE        | 时钟使能        | 控制内部时钟信息                                |
| CS         | 芯片选择        | 使能和禁止除了CLK、CKE、DQM信号的输入                 |
| BA0、BA1    | 块地址         | 当信号RAS使能时, 选择块使能<br>当CAS为低电平时, 选择块使能    |
| A0-A11     | 地址线         | 行地址线, RA0-RA11; 列地址线CA0-CA7, 自动切换标志为A10 |
| RAS、CAS、WE | 行列地址选通, 写使能 | 低电平有效                                   |
| LDQM、UDQM  | 数据输入/输出掩码   | 控制输出缓冲在读模式下, 掩码输入数据在写模式下                |
| DQ0-DQ15   | 数据输入/输出     | 数据线                                     |
| VDD/VSS    | 电源和地        | 内部电路和缓冲的电源和地                            |
| VDDQ/VSSQ  | 数字电源和地      | 为输出缓冲器提高电                               |
| NC         | 无连接         | 无连接                                     |

EMIF 与 SDRAM 的连接可以通过编程寄存器来配置, 把 SDRAM 扩展在 CE0 空间, 地址是从 020000 开始, 把 5509A 的 CE0 引脚与 SDRAM 的片选  $\overline{CS}$  相连, 数据线 DQ0~DQ15 连接 5509A 的数据线 D0~D15, 地址线 A0~A9(因为所用 SDRAM 每块大小为 1M) 连接 5509A 的 A1~A10(因为使用字寻址, 所以 5509A 的 A0 无效),  $\overline{RAS}$ 、 $\overline{CAS}$ 、 $\overline{WE}$  分别连接 SDRAM 接口专用信号。连接如图 3.11 所示。

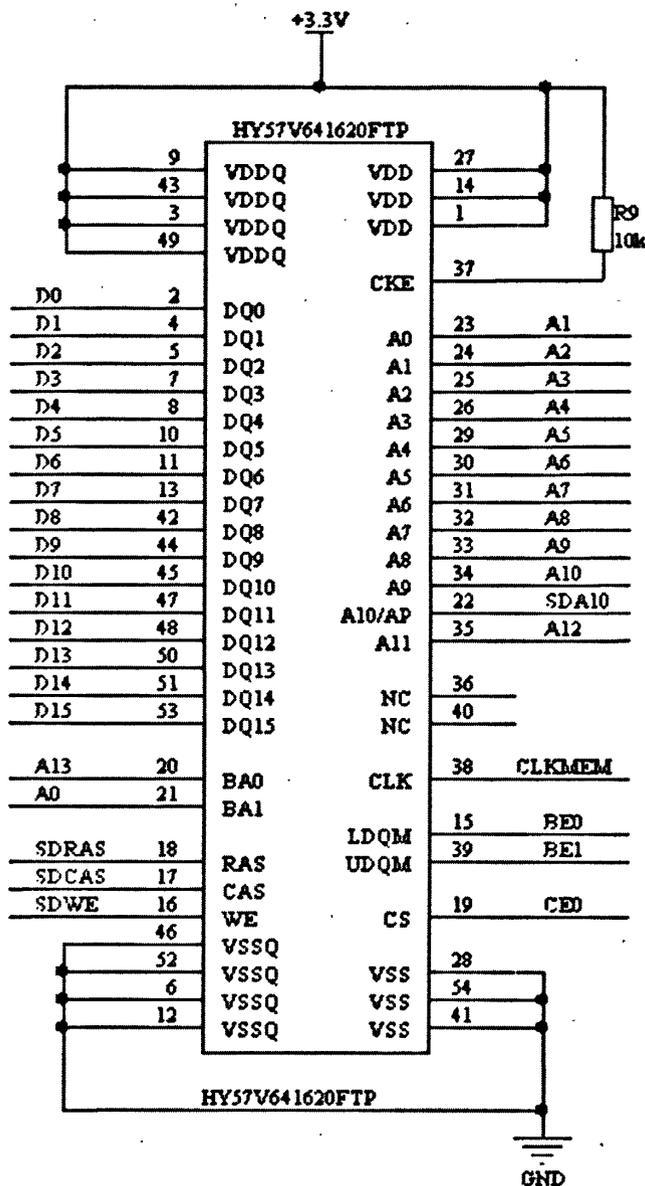


图 3.11 TMS320VC5509A 与 SDRAM 接口示意图

设置全局控制寄存器(EMIF Global Control Register)EGCR 和 CE 空间控制寄存器(CE Space Control Register)把 SDRAM 映射到 CE0 空间。CE0 空间具体的软件配置过程为:

(1)设置 EBSR(External Bus Selection Register)为 1, 选择完全 EMIF 模式;

(2)EMIF 全局控制寄存器 EGCR 如表 3.6 所示: 设置 MEMCEN=1, 说明由 CLKMEM 引脚提供存储器时钟; 设置 MEMFREQ=01b,CLKMEM 频率为 CPU 时钟频率的一半;

表 3.6 全局控制寄存器 EGCR

|       |         |    |     |    |        |        |      |     |
|-------|---------|----|-----|----|--------|--------|------|-----|
| 15~12 | 11~9    | 8  | 7   | 6  | 5      | 4      | 3    | 2~0 |
| 保留    | MEMFREQ | 保留 | WPE | 保留 | MEMCEN | ARDYQF | ARDY | 保留  |

(3)设置 CE0 空间控制寄存器，CE0\_1 的 MTYPE=011b，配置为 32-bit 的 SDRAM；并设置 SDRAM 的控制寄存器 SDC1 和 SDC2，配置 SDRAM 的时序和刷新操作；

(4)配置 SDRAM 初始化寄存器。把新值写入 EMIF 配置寄存器后，需要消耗 6 个 CPU 时钟周期等待新值传过 EMIF 逻辑，随后 EMIF 开始 SDRAM 的初始化进程。

## 2、FLASH

在语音识别过程中，要保存语音模板，由于 5509A 内部没有 FLASH，其程序加载和系统掉电之后的参数需要保存到外部存储器中，采用 Flash 存储器进行系统扩展是合适的选择。本设计选用了 AM29LV800B 芯片作为外部 ROM 固化程序使用<sup>[47]</sup>。芯片使用 3.3V 单电源供电，48 引脚的 TSOP 封装，DSP 与 Flash 的硬件连接如图 3.12 所示。

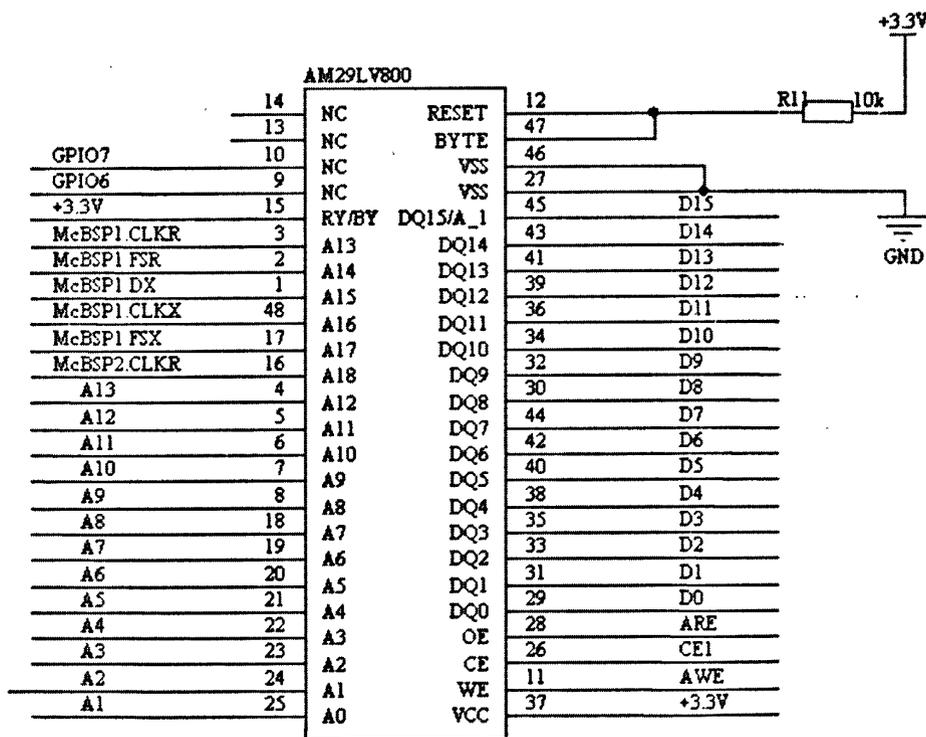


图 3.12 TMS320VC5509A 与 FLASH 接口示意图

芯片主要引脚功能说明如下：

- A0-A18 是 19 根地址线
- DQ0-DQ14: 15 根数据输入/输出线
- DQ15/A-1: 字模式，选用 DQ15 为数据输入/输出线；字节模式，为 A-1(LSB 地址输入)
- $\overline{BYTE}$ : 选择 8-bit 或 16-bit 模式
- $\overline{CE}$ : 芯片使能端
- $\overline{OE}$ : 输出使能端

- $\overline{WE}$  : 写使能端
- $\overline{RESET}$  : 硬件复位引脚, 低电平有效
- $\overline{RY}/\overline{BY}$  : 准备好/忙输出信号
- $V_{cc}$  : 标准电压范围: 3.0~3.6V(-90R)、扩展电压范围: 2.7~3.6V(-100,-120,-150)
- $V_{ss}$ : 地
- NC: 内部未连接引脚

DSP 在访问程序存储器时, 采用字节寻址方式; 访问数据存储器时, 采用字寻址方式。5509A 只有 14 根地址线 A0~A13, 而 FLASH 作为数据存储空间使用时的地址编码采用字寻址方式, 因此 DSP 的 A0 信号线无效, 由图 3-12 可知把 FLASH 映到在 CE1 空间, 起始地址为字地址 200000h, 而不是字节地址 400000h<sup>[43]</sup>。Flash 的低 13 位地址线 A0~A12 连接到 C5509A 的地址线 A1~A13, 高 6 位地址线 A13~A18 由 DSP 多缓冲串口 McBSP 来扩展。

对于程序加载部分, 加载方式由上电时系统读取 GPIO0~GPIO3 的状态决定, 本论文选用 16-bit 并行 EMIF 加载, 表 3.7 为 TMS320VC5509A 的加载方式。

表 3.7 TMS320VC5509A 引导方式

| BOOT[3:0]/[GPIO0 GPIO3<br>GPIO2 GPIO1] | 引导方式                       |
|--|----------------------------|
| 0001                                   | Mcbsp0的串行EEPROM引导(24bit地址) |
| 0010                                   | USB接口引导                    |
| 0101                                   | EHPI(多元引导)方式               |
| 0110                                   | EHPI(非多元引导)方式              |
| 1000                                   | 来自外部16bit异步内存引导            |
| 1001                                   | Mcbsp0的串行EEPROM引导(16bit地址) |
| 1011                                   | 并行EMIF引导方式(16bit异步内存)      |
| 1110                                   | Mcbsp0的同步串行引导(16bit数据)     |
| 1111                                   | Mcbsp0的同步串行引导(8bit数据)      |

### 3.3.5 上电自举模式与语音处理模式的设定

由于 DSP 的工作电压为 3.3V, 其它芯片的工作电压为 5V, 它们不能直接接口。故设计中加入了 LVC16245, 它是带三态输出的 16bit 总线收发器, 起双向

缓冲与电平转换的作用，使 DSP 芯片与外部器件之间的数据不发生冲突<sup>[48]</sup>。它有三种工作状态，如表 3.8 所示：

表 3.8 LVC16245 功能表

| INPUT           |     | OPERATION       |
|-----------------|-----|-----------------|
| $\overline{OE}$ | DIR |                 |
| L               | L   | B data to A bus |
| L               | H   | A data to B bus |
| H               | X   | Isolation       |

LVC16245 的连接图如图 3.13 所示，输出使能端  $\overline{OE} = 0$ ，方向控制端 DIR 为高电平时，数据从 A 总线到 B 总线，即数据从 DSP 到外部扩展的 8 个串口；当 DIR 为低电平时，数据从 B 总线到 A 总线，即读取 XGPIO 口的状态送到 DSP，来设定上电自举模式，如图 3.14 所示。

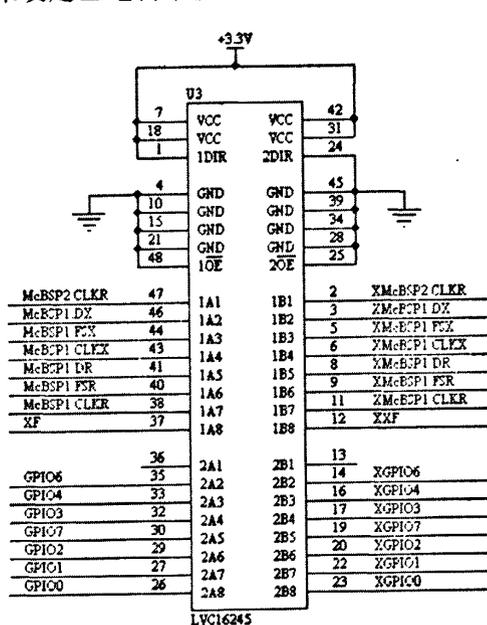


图 3.13 LVC16245 连接图

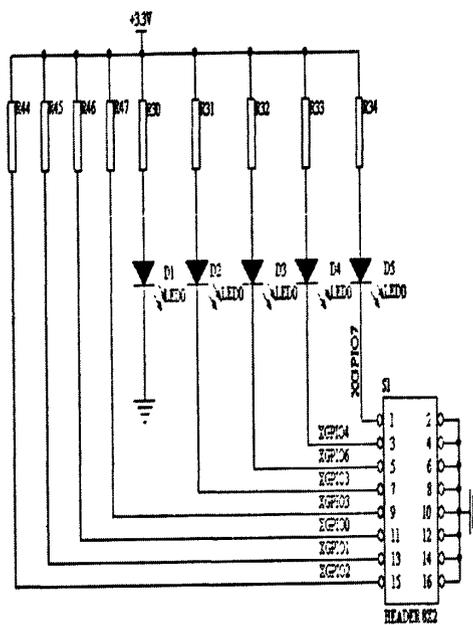


图 3.14 上电自举模式示意图

在图 3.14 中，把 DSP 自举模式设置为 16 位异步并行模式，即用短接端子把 7,8 脚短接，即设置 GPIO0, GPIO3, GPIO2, GPIO1 为 1011。可以通过发光二极管 D2 到 D5 的不同组合，如 D5~D2=1110，可以认为识别结果是第一个人，这样最多可以显示 16 组说话人识别的结果。

### 3.3.6 JTAG 接口电路

JTAG(Joint Test Action Group)接口电路与 IEEE 1149.1 标准给出的扫描逻辑电路一致,是DSP的调试接口,用户可以利用JTAG接口完成程序的下载,调试和信息的输出,通过该接口可以查看DSP的存储器、寄存器等的內容。本设计选用的是瑞泰 ICETEK-5100USB 仿真器,图 3.15 为 TMS320VC5509A 与 JTAG 的连接图。

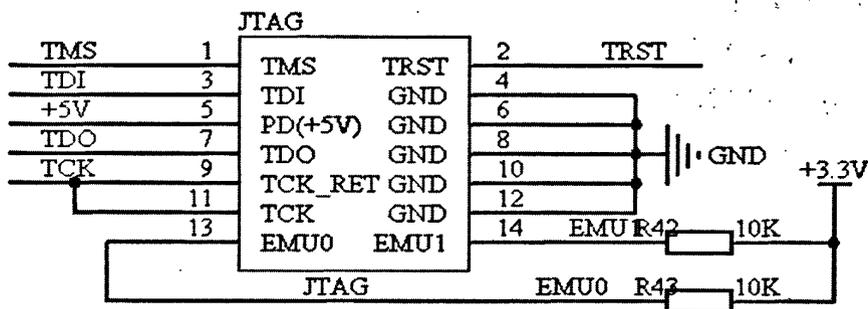


图 3.15 JTAG 接口电路

主要的 JTAG 引脚说明:

- (1)TCK: IEEE 标准 1149.1 测试时钟输入引脚;
- (2)TDI: IEEE 标准 1149.1 测试数据输入引脚;
- (3)TDO: IEEE 标准 1149.1 测试数据输出引脚;
- (4)TMS: IEEE 标准 1149.1 测试方式选择信号;
- (5) $\overline{TRST}$ : IEEE 标准 1149.1 测试复位信号引脚,低电平有效;
- (6)EMU0,EMU1: 仿真器中断引脚。

### 3.3.7 无线传输电路

本文实现用户端与小区单元之间的通信主要是通过 Silicon Labs 公司的 Si4432 这个芯片来实现的。Si4432 模块是一块完整的、体积小巧的、低功耗的无线收发模块。模块集成了芯片所需的外围器件,可以通过数字 I/O 口,控制芯片内部的寄存器,实现对无线数据发送、接收、RF 参数设置等功能。

Si4432 是 Silicon Labs 公司推出的射频收发一体芯片, Si4432 保留着以前产品的所有优越的特点,比如:高集成度、低成本、灵活性,外围电路精简和很容易设计等特点。综合起来使得它适合更复杂的应用和提供一些强大的功能和特性。包括连续频率覆盖范围: 240Mhz-930Mhz, 输出功率达到+20dbm (100mW), 接收灵敏度达到-116dbm.还包括一些内置功能,比如天线的分集算法、唤醒定时器、低电压检测、温度传感器、常用的 A/D 转换、TX/RX 先进先出缓冲寄存器 (FIFOs)、上电复位和通用 I/O 口。这些特点使得系统设计非常简单和允许使

用低端的单片机，内部集成一个 RF 输出+20dbm 的高效率的功放，从而不必要增加外置的 PA。并且在任何标准的 ISM 频段 Si4432 完全符合 FCC 和 ETSI 的要求。在大量的产品应用中当体积和成本严格要求时仅仅一个 30Mhz 的晶体和需要几个外围被动匹配元件和滤波元件就能轻松完成设计。

它的主要特征为：

- 完整的 FSK, GFSK, OOK 收发器
- 工作频率范围是 433MHz-470MHz，发射功率典型值是 17dBm，接收灵敏度是-115dBm@9.6Kbps

- 传输速率最大为 128Kbps
- FSK 频偏可编程 1-160KHz
- 接收带宽可编程 10.6KHz-230.4KHz
- 只需要使用 5 个 I/O 口 SDI、SDO、SCK、NSEL、NIRQ 即可工作
- SPI 兼容的控制接口
- 低功耗、自带唤醒定时器、低关断电流 10nA
- 内置温度传感器与 8 位的 ADC
- 集成电压调节器
- 针对窄带系统，内置晶体负载电容微调电路，方便生产调试。

Si4432 具有如下优势：

- 1)采用 SPI 接口协议与外部微处理器通信，易于编程、操作；
- 2)内部集成一个专用的晶体振荡器，能够提供准确的定时；
- 3)支持快速跳频机制；
- 4)支持 FIFO 与非 FIFO 两种模式进行数据的读、写操作，能够满足多种实际需要；

5)Si4432 支持较大的数据传输速率，数字信号的传输速率可达 128kbps；

6)外围器件少，应用电路设计简单，方便 PCB 版面设计和制作。此外，由于 Si4432 包含自动频率控制功能，降低了对外部晶体的要求，应用中仅需一支普通的 10MHz 外部晶振便能够正常工作，降低了对辅助器件的要求；

8)Si4432 采用 20 引脚的封装，能够在极少使用外围元器件的条件下正常工作。

要实现用户端与主机端之间的无线通信，本文采用低功耗的 Si4432 无线收发芯片。它与单片机、晶振和少许的元件一起工作，从而构成了一个非常低成本的系统。如图 3.16 所示，片上集成了电压调整器，允许从 1.8V 到 3.6V 的宽范围的工作电压，一个标准的 4 线 SPI 总线用于和 DSP 的多通道缓冲串口通讯。

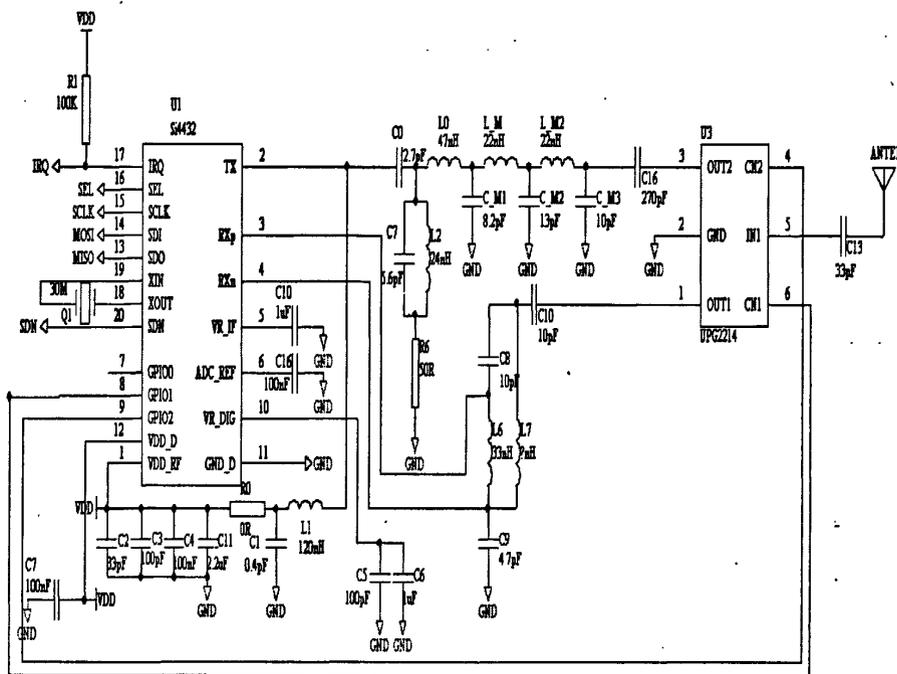


图 3.16 Si4432 硬件连接图

Si4432 的引脚定义<sup>[49]</sup>和描述如表 3.9 所示:

表 3.9 Si4432 的引脚定义与描述

| 引脚  | 引脚名        | I/O | 描述   |
|-----|------------|-----|--|
| 1   | VDD_RF     | VDD | +1.8V 到+3.6V 提供电压输入  |
| 2   | TX         | O   | 发射输出引脚   |
| 3   | RFp        | I   | LNA 的差分 RF 输入引脚  |
| 4   | RXn        | I   |  |
| 5   | VR_IF      | O   | 1.7V 调整器输出电压, 需要 1uF 退耦电容。   |
| 6   | ADC_REF    | O   | ADC 参考电压退耦, 需要 1uF 退耦电容。   |
| 7   | GPIO_0     | I/O | 通用数字 I/O, 通过 SPI 寄存器配置完成许多功能   |
| 8   | GPIO_1     | I/O |  |
| 9   | GPIO_2     | I/O |  |
| 10  | VR_DIG     | O   | 1.7V 调整器输出电压, 需要 1uF 退耦电容。   |
| 11  | GND_DIG    | O   | 数字地引脚  |
| 12  | VDD_DIG    | VDD | +1.8V 到+3.6V 供电电压输入  |
| 13  | SDO        | O   | 0-VDDv 数字输出提供内部控制寄存器串行读回功能。  |
| 14  | SDI        | I   | 串行数据输入引脚。  |
| 15  | SCLK       | I   | 串行数据输入引脚   |
| 16  | nSEL       | I   | 串行接口选择输入引脚   |
| 17  | nIRQ       | O   | 中断状态输出引脚。  |
| 18  | XOUT       | O   | 晶体振荡器输出  |
| 19  | XIN        | I   | 晶体振荡器输入  |
| 20  | SDN        | I   | 关闭输入引脚, 0-VDDv, 数字输入, 在所有模式除关闭模式外 SDN=0, 当 SDN=1, 芯片将完全关闭且寄存器的内容将丢失。 |
| PKG | PADDLE_GND | GND | RFIC 芯片底部导线架金属裸露提供整个芯片 RF 和电路接地                                      |

### 3.4 硬件电路板的抗干扰设计

为了做出一个稳定可靠的 DSP 系统, 必须从各个方面来消除干扰, 即使不能完全消除, 也要尽量减少到最小。对于 DSP 系统而言, 主要干扰来自于以下几个方面:

- (1) 输入输出通道的干扰, 模拟电路与数字电路之间的干扰;
- (2) 电源干扰和时钟干扰;
- (3) 空间辐射耦合干扰。

针对 DSP 系统板的主要干扰, 可以从以下三个方面消除信号干扰: 一是设计双层板; 二是把数字地模拟地和数字地分开; 三是使用电容滤波。具体在画 PCB 板时, 体现在以下方面:

(1) 电源芯片是系统工作的基本保证, 但同时它的散热量比较大, 因此应该把电源芯片放在 PCB 板的边缘并且要周边敷铜, 电源滤波要尽量靠近电源输入或输出;

(2) 时钟电路往往是产生电磁辐射的主要来源, 也是进行 PCB 设计时需要特别处理的地方。一方面要使时钟源尽量靠近 DSP 芯片, 使时钟走线尽量短; 另一方面, 对时钟线要优先考虑布内层, 并进行必要的匹配, 屏蔽处理;

(3) 数字信号对模拟信号的干扰主要是串扰, 在进行 PCB 放置元件和布线时, 要把模拟器件与高速数字器件分开, 模拟地与数字地也要分开处理;

(4) 在整个 DSP 系统中, DSP 与 Flash、SRAM 之间是主要的高速数字信号线, 所以高速器件之间的距离要尽量靠近, 其连线尽可能短, 并且直接连接;

(5) 在布线时, 导线的拐角处一般不要走成  $90^\circ$  折线, 以减小高频信号对外的发射耦合;

(6) 输入与输出的边线应避免相邻平行, 以避免产生反射干扰; 必要时加地线隔离。两相邻层的布线要互相垂直, 平行容易产生耦合。

(7) 电源线与地线一般要比信号线粗, 在允许的范围内尽量加宽;

(8) 在进行 PCB 设计时, 可依据信号的流向, 对整个电路进行模块划分, 从而保证整个系统布局的合理性, 达到整体走线最短, 各个模块互不交错, 减少模块间相互干扰的可能行<sup>[50]</sup>。

### 3.5 本章小结

第三章为系统硬件设计及实现部分介绍了系统硬件总的方案设计及系统总体原理图、PCB 图; 根据说话人识别无线门禁系统的要求, 设计了各个主要单元电路以及电路参数的选择与芯片的选取; 画出了每个单元电路的原理图并进行了硬件系统的抗干扰设计。在完成了说话人识别的算法设计和系统的硬件设计, 就可以进行说话人识别无线门禁系统的软件设计。

## 第 4 章系统的软件设计

### 4.1 说话人识别的结构框图

说话人识别的软件实现流程如图 4.1 所示,系统主要分为训练和识别两部分。训练的目的在于提取说话人模型参数并将其保存在 flash 中,为识别提供参考模型,训练过程与识别过程前半部分基本一致。识别的目的在于读取待识别者语音信息参数并与参考模型参数进行比较,从而获得识别结果,实现语音锁的开启与闭合。

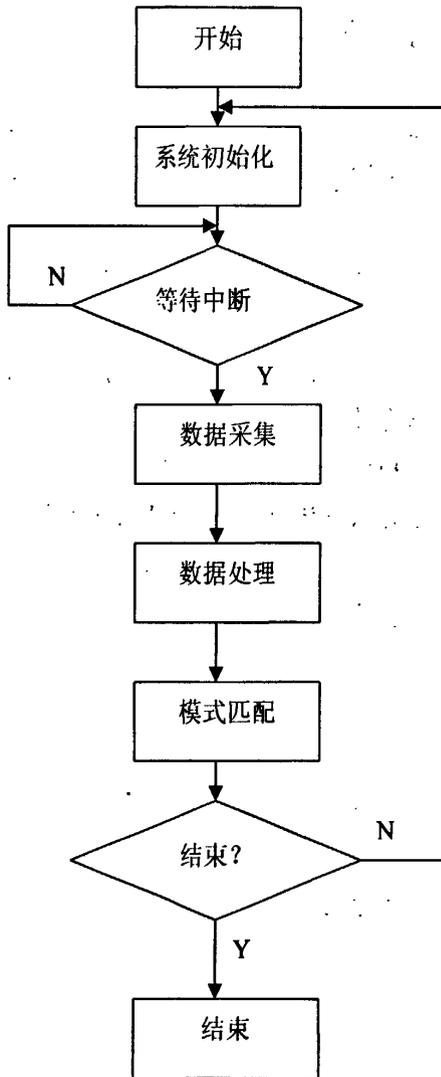


图 4.1 说话人识别的软件流程图

## 4.2 说话人识别的软件设计

### 4.2.1 系统的初始化

根据说话人识别的软件实现流程图可知,在系统进行语音数据采集之前,系统的初始化,是保证芯片的正确运行的先决条件。初始化程序主要包括:系统时钟初始化、SDRAM初始化、中断初始化、音频编解码芯片 AIC23 初始化、I<sup>2</sup>C 总线寄存器初始化和多通道缓冲口 McBSP0 初始化。初始化完成后,可以通过 McBSP0 进行实时的语音数据的采集与录放音,才能进一步的对说话人的语音信息进行处理。语音工程 yuyinshibie.pjt 中的部分程序为:

```
main()
{
    PLL_Init();      //初始化时钟
    SDRAM_init();   // SDRAM 初始化
    INTR_init();    //中断初始化
    AIC23_Init();   //初始化语音芯片
    InitMcBSP();    //McBSP0 初始化
}
```

初始化程序具体配置如下:

(1)时钟发生器初始化:通过配置时钟模式寄存器 CLKMD 来控制 DSP 的时钟发生器<sup>[40]</sup>;也可以通过 DSP 片级支持库 CSL 来配置系统时钟,输出频率= $(PLL\_MULT/(PLL\_DIV+1))*$ 输入时钟频率,配置为 144MHz。

```
PLL_Config myConfig = {
    0,    //IAI, 休眠后重新锁相
    1,    //IOB, 失锁后进入旁路模式并重新锁相
    24,   //pllmult
    1     //plldiv
};
```

(2)SDRAM 初始化包括:

a、配置外部总线选择寄存器 EBSR 为完全 EMIF 模式;

b、EMIF 的初始化:要实现 5509A 与 SDRAM 的无缝连接,需要设置 EMIF 全局控制寄存器 EGCR,由 CLKMEM 引脚提供存储器时钟且为 CPU 时钟的一半;设置 EMIF 全局复位寄存器 EMI\_RST;配置 CE 空间控制寄存器,使 SDRAM 映射在 CE0 空间;最后设置 SDRAM 控制寄存器 SDC1, SDC2、SDRAM 初值寄存器 init。

(3)中断初始化:只有初始化中断使能寄存器 IER(Interrupt Enable Register)和中断标志寄存器 IFR(Interrupt Flag Register),才能进入中断服务子程序 ISRs。

(4)音频编解码器芯片 AIC23 初始化:主要是完成对 AIC23 内部 11 个映射寄存器的配置,设置 AIC23 为主模式、8KHz 采样率、16 位量化、帧同步、基本速度比率、12MHz;启动 McBSP0 初始化。AIC23 所有的配置字都是通过 DSP 的 I<sup>2</sup>C 总线发送到 AIC23 寄存器的,所以在配置 AIC23 之前,要先对 I<sup>2</sup>C 协议进行配置。

(5)I<sup>2</sup>C 总线寄存器初始化:通过配置 I<sup>2</sup>C 模式寄存器 ICMDR,设置 I<sup>2</sup>C 工作方式为主模式,通过时钟分频寄存器 I2CCLKL 和 I2CCLKH 配置时钟控制寄存器为 100MHz 输出,具体配置如下:

```

I2C_Init()                //I2C 初始化函数
{
    PC55XX_I2C pI2C = (PC55XX_I2C)C55XX_I2C_ADDR;
                                //I2C 复位, I2CSTR 所有状态恢复到默认值
    ClearMask(pI2C -> icmdr, ICMDR_IRS); //重置 I2C 寄存器 ICMDR,
        目的是发送一个信号激活 I2C 总线, 预分频寄存器 12MHz
    pI2C -> icpsc = dspclk.pllmult; //配置 I2C 总线的时钟信号, 由于 AIC23
        使用的是晶振输入, 所以 I2C 时钟寄存器 ICPSK 直接使用时钟信号
        dspclk.pllmult; 设置时钟控制寄存器(100KHz 时钟输出)
    Write(pI2C -> icclk, 10);
    Write(pI2C -> icclkh, 10); //配置 I2C 信号的时钟信号 (SCL) 频率, 设
        置自身地址寄存器
    WriteMask(pI2C->icoar, ICOAR_OADDR, ICOAR_MASK_7); //使能
        I2C 模块 并工作在主模式
    SetMask(pI2C->icmdr,ICMDR_IRS|ICMDR_MST); //第二次对 ICMDR 赋
        值, 表示 I2C 寄存器配置完成, I2C 总线准备工作
}

```

(6)I2C 配置 AIC23: 根据第三章图 3-9 可知, SCLK 和 SDI 信号同时发送高电平“1”, 启动 I<sup>2</sup>C 总线, SDI 可以传输数据。注意 SDI 传输数据的流程:

首先, 发送 ADDR (从地址)+R/W (读写), 通过这个信号来确定 DSP 到底是与什么芯片进行通信, 这个由从地址决定, 而且同时通知从芯片是读还是写。

然后, 发送 8 位数据, ACK 信号是 I<sup>2</sup>C 传输自带的信号, 用来划分数数据流; 最后 SDI 和 SCLK 同时为“1”, I<sup>2</sup>C 总线数据传输完成。

```

void I2C_Write(unsigned short int device, int count, unsigned char *bytedata)
{
    PC55XX_I2C pI2C = (PC55XX_I2C)C55XX_I2C_ADDR;
    int i;

```

```

Write(pI2C -> iccnt, count); //设置需要通过 I2C 传输的数据的大小
WriteMask(pI2C -> icsar, device, ICSAR_MASK_7); // 主要是设置 DSP 想要
读写的从地址
WriteMask(pI2C -> icmdr,
          ICMDR_STT | ICMDR_STP | ICMDR_TRX,
          ICMDR_STT | ICMDR_STP | ICMDR_TRX); // 激活 I2C 寄存器中的相
应的寄存器
for (i = 0; i < count ; i++) {
    Write(pI2C -> icdxr, bytedata[i]); // DSP 通过 ICDXR 这个寄存器发送数
据 bytedata[i]
    while(!(pI2C -> icstr & ICSTR_ICXRDY)); // 检测 DSP 发送寄存器
ICDRX 是否发送完毕
}
}

```

(6)McBSP0 寄存器初始化: 将 McBSP0 设置为 32 位, 单帧无延时, 在时钟的下降沿发送数据。初始化步骤为:

```

void McBSP0_InitSlave()
{
    PC55XX_MCSP pMCBSP0 = (PC55XX_MCSP)C55XX_MSP0_ADDR;
    //MCBSP0 复位
    Write(pMCBSP0 -> spcr1, 0);
    Write(pMCBSP0 -> spcr2, 0); //初始化 SPCR1 和 SPCR2 这两个控制寄存
器,
    Write(pMCBSP0 -> xcr1, XWDLEN1_32); // 设置输出寄存器 xcr1 的传
输字节长度, 配置帧参数(32-bit,单帧无延时)
    Write(pMCBSP0 -> xcr2, XPHASE_SINGLE | XDATDLY_0); // 设置输出
寄存器 xcr1 在传输数据时, 是否需要延迟一个时钟输出
    Write(pMCBSP0 -> rcr1, RWDLEN1_32);
    Write(pMCBSP0 -> rcr2, RPHASE_SINGLE | RDATDLY_0);

    Write(pMCBSP0 -> pcr, PCR_CLKXP); //设置 PCR 寄存器 CLKXP 位,
主要功能是确定时钟信号的极性, 当 CLKXP 为 0 的时候为正极性, 时钟以上升
沿开始; 反之, 时钟以下降沿开始, 配置为下降沿发送数据
    SetMask(pMCBSP0 -> spcr2, SPCR2_XRST);
    SetMask(pMCBSP0 -> spcr1, SPCR1_RRST); // 分别初始化 SPCR2 中的
XRST 寄存器和 SPCR1 中 RRST 寄存器, 发送串口使能

```

}

#### 4.2.2 AIC23 语音信号采集

语音信号的采集主要通过 TLV320AIC23 来实现的, 它为我们提供了 11 个映像寄存器, 通过对这些寄存器合理的设置, 实现音频编解码器的正常工作, 并产生预期的效果。设置语音采集为单声道 8KHz 采样, 16 位量化。

```
void AIC23_Init()//语音芯片初始化
{
    // Reset the AIC23 and turn on all power
    AIC23_Write(AIC23_RESET_REG, 0);//初始化 AIC23 寄存器
    AIC23_Write(AIC23_POWER_DOWN_CTL, 0); //初始化节电模式控制
    AIC23_Write(AIC23_ANALOG_AUDIO_CTL, 0x0014);//模拟音频通道控制
    AIC23_Write(AIC23_DIGITAL_AUDIO_CTL, 0);//数字音频通道控制
    // Turn on volume for line inputs 为音频输入打开设备
    AIC23_Write(AIC23_LT_LINE_CTL, 0x000);//左声道输入音量控制
    AIC23_Write(AIC23_RT_LINE_CTL, 0x000);//右声道输入音量控制
    // Configure the AIC23 for master mode, 44.1KHz stereo, 16 bit samples 数字
    音频接口格式
    AIC23_Write(AIC23_DIGITAL_IF_FORMAT, 0x0043);
    AIC23_Write(AIC23_SAMPLE_RATE_CTL, SRC_SR_44 | SRC_BOSR |
SRC_MO);//采样率控制
    // Turn on headphone volume and digital interface
    AIC23_Write(AIC23_LT_HP_CTL, 0x07f); //左耳机通道控制
    AIC23_Write(AIC23_RT_HP_CTL, 0x07f); //右耳机通道控制
    AIC23_Write(AIC23_DIG_IF_ACTIVATE, 0x0001);//数字界面激活
}
```

#### 4.2.3 数据处理

数据处理主要包括说话人信息的分帧, 预加重, 加窗, 断点检测, 特征提取。

(1) 分帧: 语音信号的采集采用单声道 8KHz 采样, 16 位量化精度, 采集 3S, 帧长为 240, 帧移 80, 则总的帧数为 100 帧左右, 考虑到帧移, 可以把帧数设置为 120 帧。

```
for(count=0;count<120;count++) //共120帧
{
    for(i=0;i<240;i++) //每帧240个点
        {voice_data_f[i]=0;}
}
```

(2) 预加重与加窗：预加重一般是在语音信号数字化之后，实质就是一阶高通滤波器  $H(z)=1-0.95z^{-1}$ ，pre\_emphrase()子程序实现对输入语音数据进行预加重，window()子程序为汉明窗函数，都以帧数为单位。

(3) 端点检测：采用的是双门限检测法，包括短时能量和过零率，分别检测语音信号的清音和浊音。Power子程序与zero\_crossing子程序分别为短时能量函数与过零率函数。

(4) 特征提取：mfcc()子程序计算24阶的MFCC加权倒谱系数。

#### 4.2.4 BOOT LOADER 程序设计

由于C5509片内不具有FLASH也就是说DSP的程序掉电会丢失，这当然需要外接非易失性存储器来完成DSP程序上电状态的引导。DSP芯片的Bootloader程序用于上电时将用户程序从外部非易失性、慢速存储器或外部控制器中装载到片内高速RAM中，保证用户程序在DSP内部高速运行，这个过程就是自举加载。

TMS320VC5509A 为用户提供了多种加载方式，由表 3.3.3 可见，其加载方式有多种：无引导模式、并行 EMIF 引导模式、SPI 引导模式和标准串口引导模式<sup>[51-52]</sup>。选择何种引导模式，主要是通过设置 GPIO0~3 的状态来实现。在系统上电时，读取 GPIO 口的状态，GPIO0~3=1110，即本文采用并行 EMIF 引导模式。应用程序的代码是以引导表的形式存放在外部 flash 中的，引导表携带了代码段和数据段的信息，DSP 下载程序的入口地址、寄存器配置的信息、段的起始地址和可编程延时信息，应用这些信息来配置 DSP 完成下载过程<sup>[52]</sup>。引导表的结构如表 4.1 所示：

表 4.1 引导表结构

| 字节地址<br>+0  | 字节地址<br>+1 | 字节地址<br>+2 | 字节地址<br>+3 |
|-------------|------------|------------|------------|
| 入口点地址(32位)  |            |            |            |
| 寄存器地址(16位)  |            | 寄存器地址16位   |            |
| 延迟指针(16位)   |            | 延迟计数(16位)  |            |
| 段字节数(32位)   |            |            |            |
| 段起始地址(32位)  |            |            |            |
| 数据(字节)      | 数据(字节)     | 数据(字节)     | 数据(字节)     |
| 数据(字节)      | 数据(字节)     | 数据(字节)     | 数据(字节)     |
| 32位0(引导表结束) |            |            |            |

其中程序入口地址是指bootloader完成代码的搬移后，用户程序开始执行的

地址,也是用户程序生成的map文件中显示的入口地址;配置的寄存器数目决定了外围设备需要配置寄存器的个数;外围寄存器的配置需要一定的时间延时,因此通过设置延时标志0xFFFF确定延时的时钟周期数;段长度、段起始地址和数据则为用户程序中定义的各个段的内容,该部分根据具体程序的初始化段个数重复;引导表以32个0为结束标志<sup>[53]</sup>。

要生成引导表,可以将 CCS 最终编译生成的.OUT 文件通过 CCS 自带的 hex55.exe 转换程序得到。在进行转换操作之前,先把用户所有的程序——语音工程(yuyinshibie.pjt)生成的.out 输出文件、CMD 文件 hex5509.cmd 和转化工具 hex55.exe 放在相同的文件夹里,通过 DOS 命令格式调用 hex55.hex,即可完成。

在编写程序时, hex5509.cmd 为引导表提供相关的配置信息,文中用到的 hex5509.cmd 的内容是:

```

Debug\yuyinshibie.out    /*输入文件*/
-o flash.hex             /*输出文件*/
-boot                    /*创建一个引导表*/
-v5510:2                 /*选择生成 55X boot 文件格式*/
-a                       /*输出格式为 straight ASCII*/
-parallel16              /*16 位并行异步加载模式*/
-memwidth 16             /*DSP 数据宽度*/
-romwidth 16             /*Flash 数据宽度*/
-map yuyinshibie.map     /*输出 map 文件*/
-delay 0xffff            /*延时*/
    
```

#### 4.2.5 非易失存储器 FLASH 在线编程

FLASH 是掉电后信息不易丢失的存储器,可在线进行电擦写,它具有很低的功耗、存储容量非常大、擦写迅速等特点。AM29LV800B 是 AMD 公司生产的 FLASH 存储器,3V 单电源供电,可使内部产生高电压进行编程和擦除操作。

根据 FLASH 存储器的顺序进行操作,先复位,再擦除,最后烧写,不能违背操作顺序,否则可能导致 FLASH 存储器复位。表 4.2 为 FLASH 的操作步骤说明(对芯片的操作都是按照字进行的),先写地址,再写数据,就可以实现对 FLASH 的编程。表中所有的数据都是十六进制数,PA 为编程地址,PD 为编程数据。

表 4.2 AM29LV800B 操作命令说明

| 命令序列 | 周期 | 第1步<br>地址 数据 | 第2步<br>地址 数据 | 第3步<br>地址 数据 | 第4步<br>地址 数据 | 第5步<br>地址 数据 | 第6步<br>地址 数据 |
|------|----|--------------|--------------|--------------|--------------|--------------|--------------|
| 复位   | 1  | XXX F0       |              |              |              |              |              |
| 芯片擦除 | 6  | 555 AA       | 2AA 55       | 555 80       | 555 AA       | 2AA 55       | 555 10       |
| 编程   | 4  | 555 AA       | 2AA 55       | 555 A0       | PA PD        |              |              |

在 DSP 对 AM29LV800 进行读写访问之前, 要对 DSP 内相关寄存器进行正确设置, 把 FLASH 配置扩展在 CE1 空间, 起始地址为字地址 200000, 所以操作时所有地址要加上 200000, 例如在编写擦除部分命令时为:

```

deminaddr = (int *)CESECT1;
addbias = 0x0555;
*(deminaddr+addbias) = 0x00aa;
addbias = 0x02aa;
*(deminaddr+addbias) = 0x0055;
addbias = 0x0555;
*(deminaddr+addbias) = 0x0080;
addbias = 0x0555;
*(deminaddr+addbias) = 0x00aa;
addbias = 0x02aa;
*(deminaddr+addbias) = 0x0055;
addbias = 0x0555;
*(deminaddr+addbias) = 0x0010;
    
```

由表4.2知FLASH芯片擦除总共需要六个时钟周期, 编程需要四个时钟周期, 只要在特定的寄存器地址中写入特定的值就可以实现FLASH的擦除与编程。图4.2为flash擦除与编程的波形图, 说明了芯片进行擦除和编程的过程。

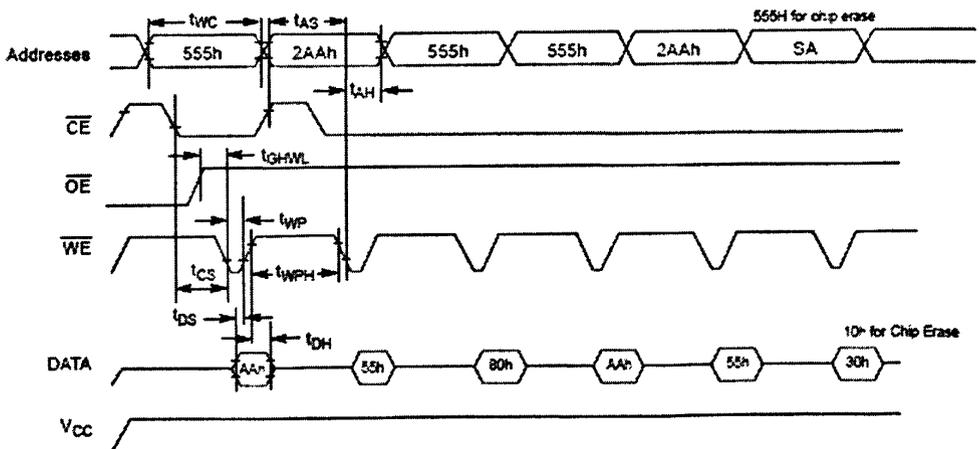


图 4.2 擦除和编程命令示意图

图4.2中显示的AM29LV800B操作需要的最小时间如表4.3所示。

表 4.3 写操作的时间参数

| 参数实体 | $t_{WC}$ | $t_{AS}$ | $t_{AH}$ | $t_{GHWL}$ | $t_{WP}$ | $t_{WPH}$ | $t_{CS}$             | $t_{DS}$ | $t_{DH}$ |
|------|----------|----------|----------|------------|----------|-----------|----------------------|----------|----------|
| 描述   | 写周期      | 地址建立时间   | 地址保持时间   | 写操作前恢复时间   | 写脉冲宽度    | 写脉冲高电平宽度  | $\overline{CE}$ 建立时间 | 数据建立时间   | 数据保持时间   |
| Min  | 90ns     | 0ns      | 50ns     | 0ns        | 50ns     | 30ns      | 0ns                  | 50ns     | 0ns      |

要判断FLASH是否编程或擦出成功，检测的方法有三种：

- (1) 检测 flash 中的 RY/BY 引脚的状态；
- (2) 检测数据之间进行切换位 DQ6 (Toggle Bit) 的状态；
- (3) 数据轮询校验。

本文采用的是第二种数据切换位检测方法，即检测跳变位 DQ6。图 4.3 为触发位校验的流程图，编程和擦除的内部算法必须在 DQ5 状态为“0”时才完成。

$DQ6 = Toggle?$  意思是判断 DQ6 的值是否在切换。

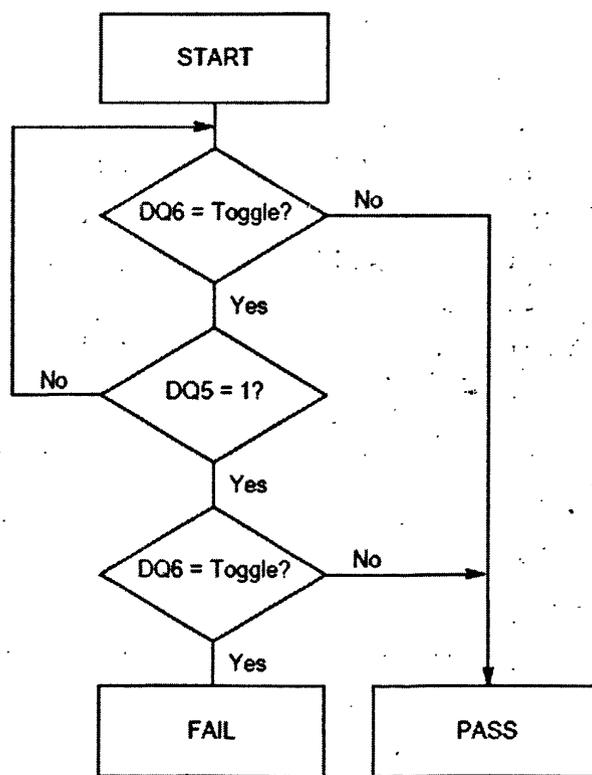


图 4.3 触发位检测算法流程图

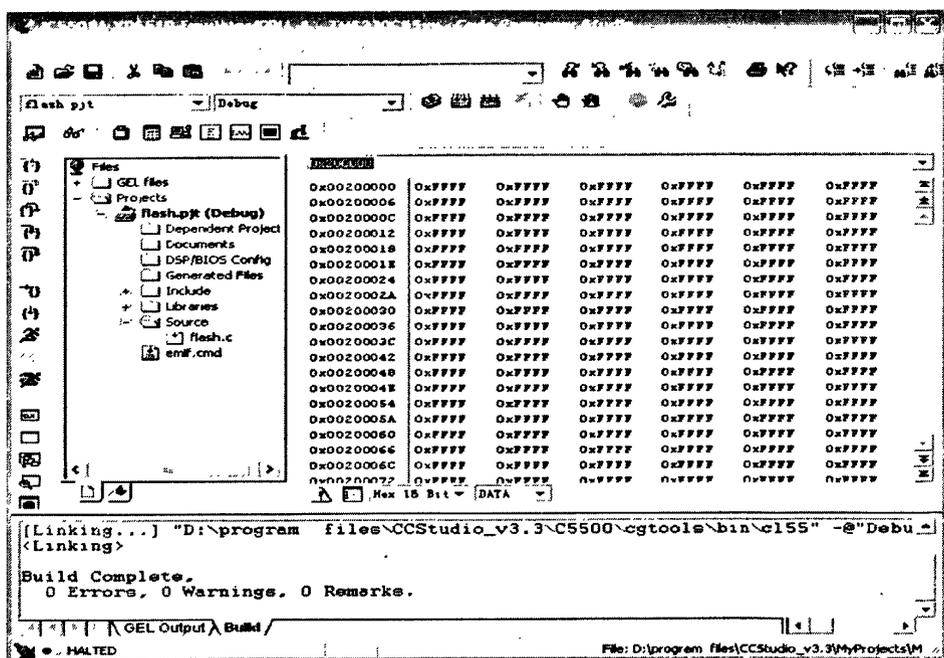


图 4.4 FLASH 擦除图

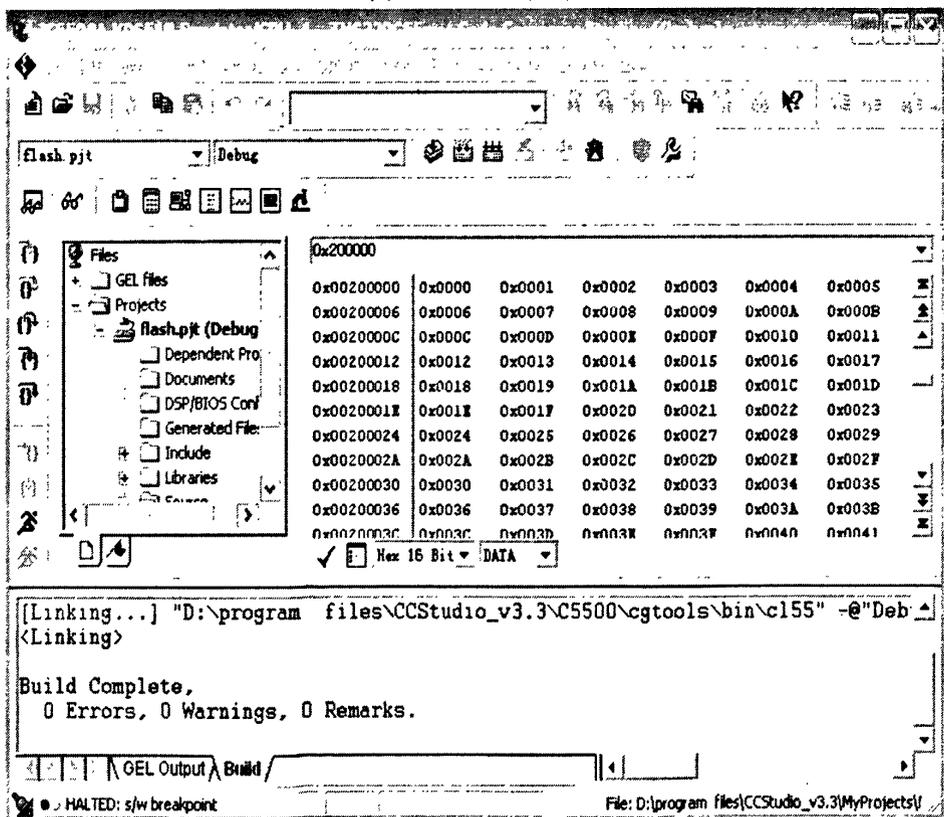


图 4.5 FLASH 烧写图

### 4.3 无线通信的软件设计

因为 Si432 提供的是 SPI 接口, 所以把 TMS320VC5509<sup>[56-58]</sup>的 McBSP2 配置成 SPI 接口, 与 Si4432 进行通信。SPI 接口共有四种操作模式, 操作模式决定

了时钟信号的上升沿还是下降沿进行数据的发送或接收。本文将 TMS320VC5509 的 McBSP2 设置为 SPI 协议的主设备。具体连接如图 4.6 所示：

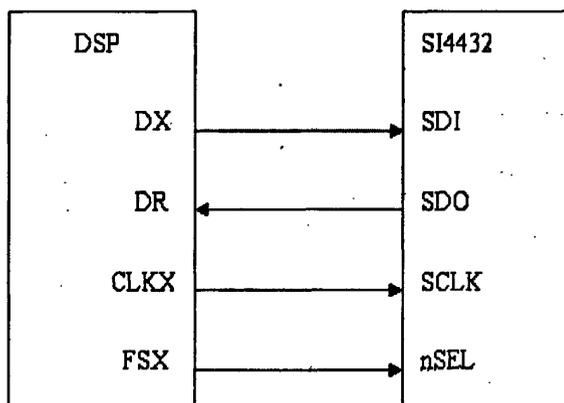


图 4.6 DSP 和 Si4432 连接图

本文设置为时钟停止模式3，即将多通道缓冲串口McBSP的Serial Port Control Register1 (SPCR1)中的CLKSTP位配置为11b，将Pin Control Register (PCR) 中的CLKXP位配置成1b的模式，为上升沿开始，有延时。SPCR1, SPCR2 控制寄存器分别如图4.7、4.8所示：

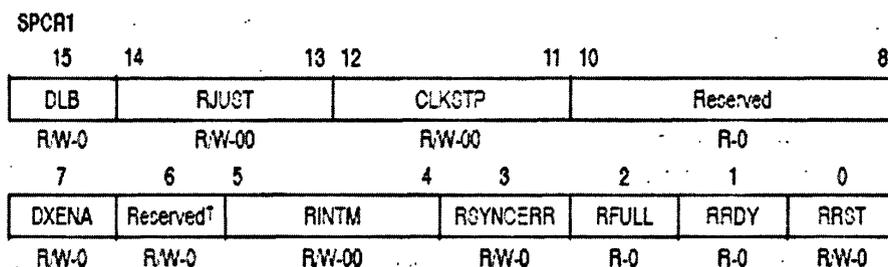


图 4.7 SPCR1 控制寄存器图

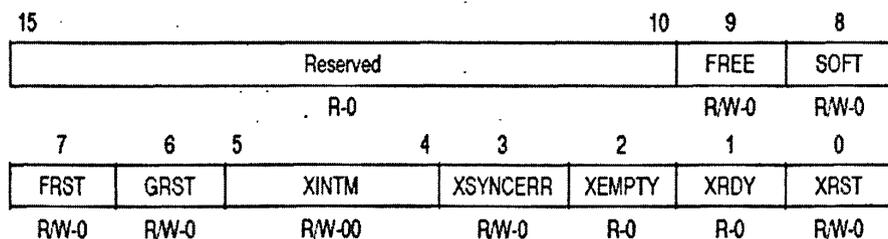


图 4.8 SPCR2 控制寄存器图

当作为SPI主设备时，McBSP内部的采样率发生器驱动时钟 CLKX 输出。FSX 输出作为从设备使能信号。在McBSP端，对收/发的内部操作而言，时钟信号是连续的，需要设置SRGR寄存器的CLKSM位选择采样率发生器的时钟源，由CLKGDV位设置需要的SPI数据传输速率。McBSP在CLKX下降沿的半个时钟周期之前发送数据，在CLKR的下降沿读取数据。如图4.9所示：

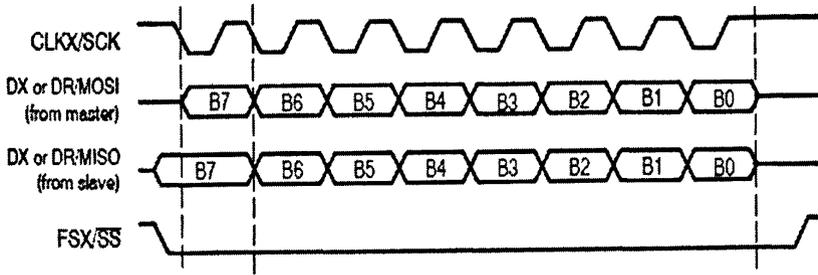


图 4.9 SPI 操作模式 3 时序图

SI4432 的发送软件流程图如图 4.10 所示，因为接收是发送的逆过程，这里就不重复叙述了。

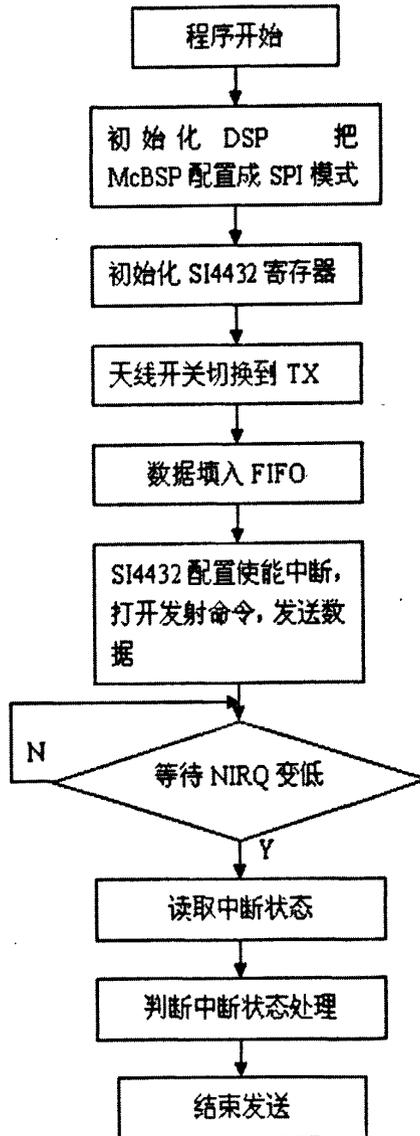


图 4.10 Si4432 发送软件流程图

## 4.4 本章小结

本章介绍了基于 CCS 开发平台的说话人识别软件开发流程。详细设计了说话人识别的框架及整个系统初始化程序的编写；针对 DSP 与 AIC23 的通信，设计语音采集传输的过程，研究了 AM29LV800B 其在线编程的具体实现方法；设计了 5509A 与 SI4432 SPI 通信的软件编写。为第五章系统联调及结果分析提高软件支持。

## 第 5 章 系统联调测试与结果分析

### 5.1 系统的硬件调试

在进行电路板焊接之前，首先原理图要和 PCB 进行对照，检查印制电路板的连线是否有短路或连接错误的走线。同时，焊接以各单元电路为单位，焊接器件的步骤及硬件调试时采取的步骤为：

(1) 电源是各个模块的基础，首先要保证电源模块工作正常。主要包括 7805 电源模块和 TPS767D301 电源芯片部分，TPS767D301 输出电压为 3.3V 和 1.6V，7805 输出 5V。

(2) 主控模块 TMS320VC5509A，因为是 QFP-144 引脚封装，要采用拖焊的方法焊接，完成后，用酒精擦除多余的松香，最后用刷子擦除干净。

(3) 因为是采用 12MHZ 有源晶振，用示波器检查 DSP 的 X2/CLKIN 引脚的信号是否正常，若正常则表明可以为 DSP 提供时钟，DSP 可以正常工作。

电源电路、晶振电路和复位电路是整个系统正常工作的基础，应首先保证它们的正常工作。

(4) 在系统的两类存储器中，SDRAM 与 FLASH 存储器控制信号较多，对于他们的操作要严格按照顺序，否则会引起烧写失败。

(5) 音频编解码模块的调试，同样先要调试系统的电源与时钟。测试程序根据瑞泰创新的 ICETEK-VC5509-A 评估板提供的语音采集与放送程序，通过耳机可以听到麦克风里输出的语音。

(6) 外围电路测试，AD 采样电路及上电自举模式等设定电路。

### 5.2 系统的软件调试

所有的软件代码调试是基于 CCS3.3(C5000)环境下进行的。整个系统软件共包含三个工程文件：语音工程 yuyinshibie.pjt，烧写工程 flash.pjt，无线收发工程 si4432.pjt。调试步骤为：首先设置软件的启动环境为 C55xx XDS510 Emulator 仿真器，芯片选择为 5509A，打开语音工程、烧写工程界面分别如图 5.1、5.2 所示。通过编译链接生成 OUT 文件，运行 file-load program 加载运行，把程序下载到 DSP 中，查看寄存器或内存单元的值，完成软件的调试过程。



### 5.3 系统的测试设备

麦克风 1 个、耳机 1 个；PC 机 1 台，电路板一套。硬件系统和测试平台分别图所示：

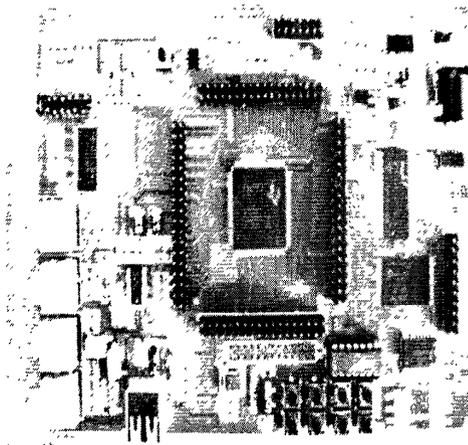


图 5.3 DSP 硬件系统

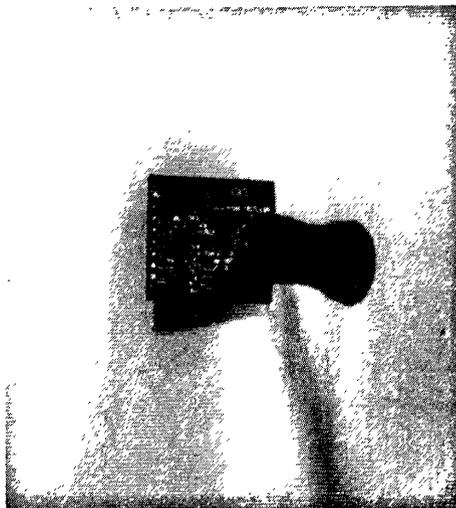


图 5.4 Si4432 硬件系统

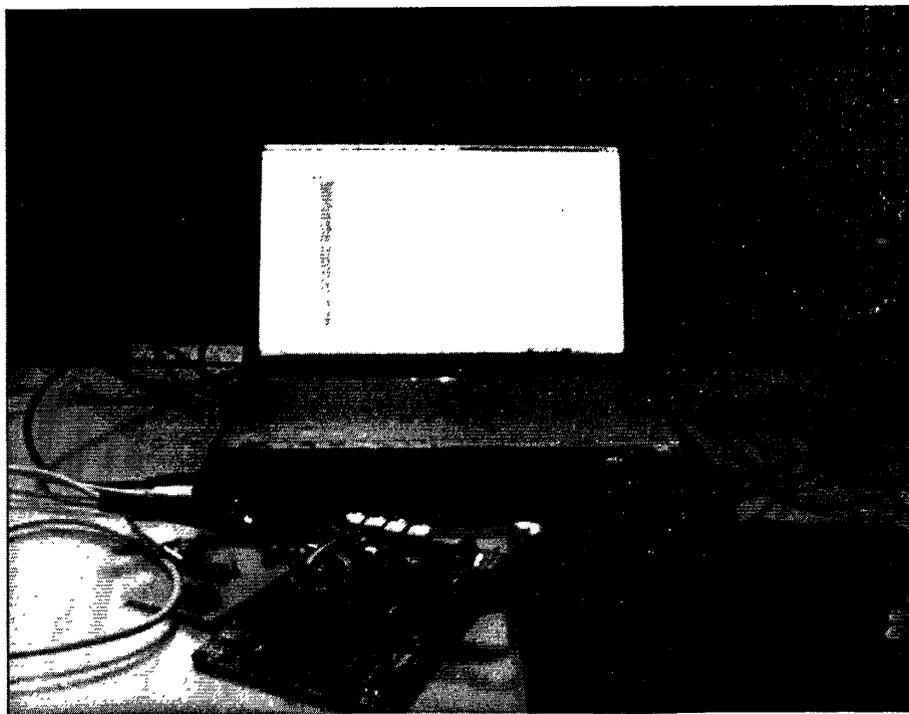


图 5.5 系统的测试平台

针对本系统，做了如下的优化：

(1)将数据缓存至 SDRAM 中。把 SDRAM 扩展在 DSP 的 CE0 空间，方便数据的保存与读取，提高计算速率。

(2)用指针方式访问语音数据的数组。改用指针的方式访问数组改善了程序的执行效率，运行速度有一定提升。

(3)用读表法获取汉明窗函数。最初的程序是通过运算公式的方式得到窗函数的各个样点值的,后改用读表法,减少了运算时间,提高运算速率。

(4)用中断方式采集语音。可实现每采集完一个人的语音数据便进行处理,极大地提升了处理速度。

### 5.4 测试结果与分析

本系统采用改进的端点检测方法, MFCC 特征提取和 DTW 改进算法来实现说话人识别。在实验室环境下,通过麦克风与 PC 机连接,把录音机配置为单声道,语音采样频率为 8KHZ, 16 位量化精度,预加重系数  $a=0.95$ ,语音每帧为 30ms, 240 点为一帧,帧移为 80,窗函数采用 Hamming 窗。采集 5 个女生, 10 个男生的数据。共分为两组,第一组是对 0~9 十个数字的说话人识别;第二组是对汉语的说话人识别。每一组的参考模板为 150 个,在相同的实验室条件下,建立测试模板 150 组。试验数据如表 5.1 所示:

表 5.1 说话人识别实验结果

| 算法        | 参考模板个数 | 测试模板个数 | 正确识别个数 | 错误识别个数 | 识别率   |
|-----------|--------|--------|--------|--------|-------|
| 传统算法数字识别  | 150    | 150    | 137    | 13     | 91.3% |
| 改进算法数字识别  | 150    | 150    | 145    | 5      | 96.7% |
| 传统算法说话人识别 | 150    | 150    | 127    | 23     | 85.7% |
| 改进算法说话人识别 | 150    | 150    | 136    | 14     | 90.7% |

在 MATLAB 中仿真结果如图 5.6 所示:

```

正在匹配模板参数.....
正在计算模板的参数....
正在进行模板匹配...
正在计算模板匹配结果....
测试模板1的识别结果为:第1个说话人
测试模板2的识别结果为:第2个说话人
测试模板3的识别结果为:第3个说话人
测试模板4的识别结果为:第4个说话人
测试模板5的识别结果为:第5个说话人
测试模板6的识别结果为:第6个说话人
测试模板7的识别结果为:第7个说话人
测试模板8的识别结果为:第8个说话人
测试模板9的识别结果为:第9个说话人
测试模板10的识别结果为:第10个说话人
    
```

图 5.6 说话人识别实验结果

本系统硬件实现的是 5 个人的说话人识别,用 DSP 板上的 5 个独立按键来

训练并保存在 FLASH 中。用 DSP 板上 4 个 LED 灯的组合来显示识别结果，最多可以显示 16 个人的识别结果。程序来自麦克风的外部模拟语音信号经 AIC23 转换为数字信号，通过多通道缓冲串口送至 TMS320VC5509A，DSP 芯片对输入的数字信号进行端点检测处理，端点检测“one”的前后波形图如图 5.7、5.8 所示：

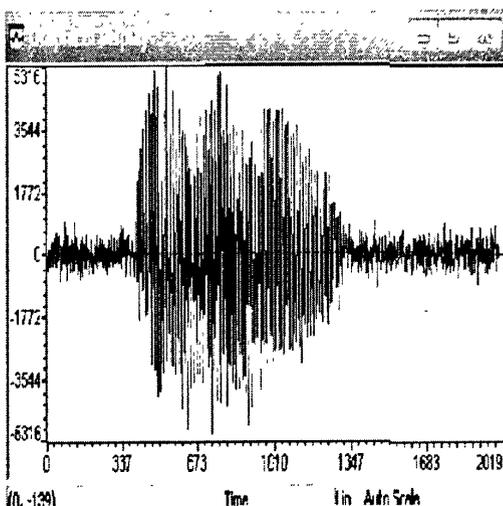


图 5.7 端点检测前波形图

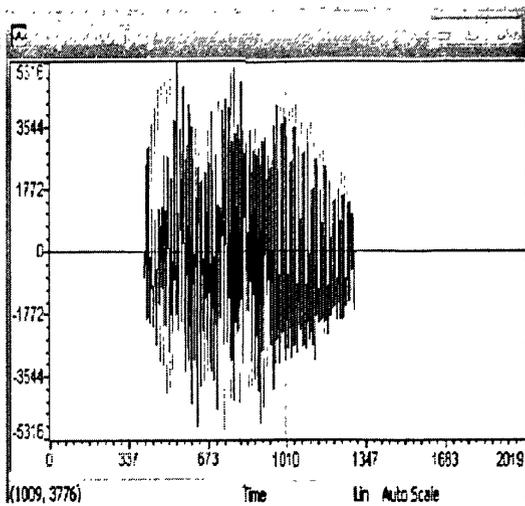


图 5.8 端点检测后波形图

根据说话人识别特征提取知，Mel滤波器组取24，每一个Mel滤波器覆盖的范围都近似于人耳的临界带宽，以此来模拟人耳的听觉效应。在Mel域上用滤波器组等分频谱，24个基本能够覆盖，但是特征系数取得太大，相应的计算量就增大，在CCS软件中，通过计算MFCC共消耗453684272个时钟周期，再加上其它函数的运算、延时及干扰等，时间相对较长。如果取MFCC为16，特征提取的共消耗254773047个时钟周期，计算量减小一半，但是识别率却下降2%。具体的可以通过CCS自带的Profile实时分析工具来获得，如图5.9所示：

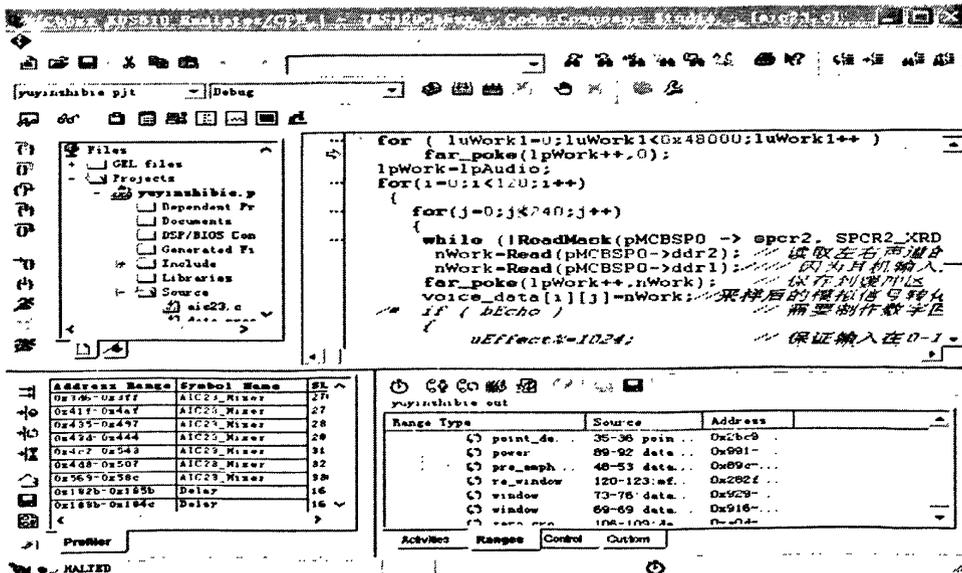


图 5.9 Profile 观察窗口

通过对语音信号的预加重,端点检测,特征提取,与语音库的模式匹配几个环节进行了详细的研究。采用双门限检测法,通过对过零率和短时能量参数的检测来判断起始点和结束点,提取出语音信号数据。利用Mel倒谱系数MFCC提取出语音信号的特征参数,这些特征参数组成特征矢量,从而构成语音模板。在模式匹配中,采用了改进的动态时间归整方法,将规整函数搜索的路径限制在一个平行四边形内,并将测试模板与参考模板进行匹配,比较两者之间的失真,求出规整函数最小的累积距离作为输出。并对系统的各个部分进行了仿真与硬件调试,能够完成小词汇量的说话人识别,效果良好。

数据的无线收发,通过把DSP的多通道缓冲串口配置成SPI模式,与Si4432相兼容。配置Si4432的载波频率是433MHz,前导码长度是2byte,FSK调制,发送数据的波特率是9600,通过判断Si4432引脚nIRQ的状态来确定数据是否发送完毕,发送完成后,进入中断,点亮LED灯,功能在系统上实现,效果良好。

## 5.5 本章小结

本章从硬件和软件两个方面叙述了系统调试方法及过程,分析并解决了调试过程中出现的问题;通过硬件采集了一段基于CCS的语音数据;并对语音数据进行分析得到了匹配模板,给出了基于MATLAB的说话人识别结果及硬件移植的结果,并对实验的结果进行了相关的分析。

## 结论与展望

本文阐述了说话人识别的背景、研究现状及基本理论,介绍了双门限的端点检测,预处理等说话人识别的过程,详细讨论了美尔倒谱系数,动态时间规整及改进算法的识别原理,设计了一个基于DSP的说话人识别系统及无线通信系统,算法在MATLAB中实现,最优方案在TMS320VC5509A及SI4432系统开发板上实现,论文针对系统的实际应用问题,合理设计流程,优化算法,提高识别率,改善可操作性。经过测试,本系统运行正确,性能良好,达到了预期的效果。

本论文的主要工作为:

1.研究了说话人识别的基本理论,采用延长帧的双门限端点检测法,MFCC特征提取,改进的动态时间规整算法,并用MATLAB进行了仿真研究,有效的提高了系统的识别率。

2.设计了整个系统的硬件原理图,PCB制板,焊接与调试,完成说话人识别系统板与无线通信板的研究设计。

3.采用了C语言与汇编语言实现整个系统的算法软件,设计并实现了包括语音信号采集、语音模板生成、说话人识别训练子系统,无线收发子系统。

4.完成了DSP芯片的上电复位和上电加载电路的设计并调试成功,使系统可以在脱机状态下独立运行。设计完成语音芯片TLV320AIC23与TMS320VC5509的电路连接,并在CCS开发环境下完成了芯片之间进行数据传输的I2C和McBSP串口的设置,完成TLV320AIC23芯片的寄存器配置和语音信号的读取。

结合本文的工作情况,本系统还存在着有待改进的地方:

1.本文只研究了小词汇量的说话人识别,如何开发出识别率更高,词汇量更大,对环境噪声更鲁棒的识别系统,需要更进一步的研究。在目前工作的基础上,研究更准确的建模方法,如利用隐马尔科夫模型,神经网络模型等,以及它们的组合来改善系统。

2.整个硬件系统板上采用FLASH来存储程序和模板,由于FLASH只有整片擦除和块擦除方式,这给模板更新过程带来不便,可以在系统上扩充一片掉电不丢失的SRAM来代替,这样就可以对模板进行选择性的训练与更新。

3.目前识别时间相对较长,以后工作可围绕优化系统方面展开,对算法中关键部分更多使用汇编语言,并且寻找更多方法优化系统,以获得更高的算法执行效率。

4.在系统实用性方面多做工作。首先,本系统是在实验室的环境下完成的,以后可研究在高噪声条件下提高系统的稳健性和识别性能。其次,本系识别范围为5个人,可以用系统板上自带的LED灯的组合来显示识别结果,当识别人数较多时难以实现,以后工作中可以考虑用语音提示的方式来控制操作和显示识别结果。

5. 虽然本文对训练和识别的速度问题进行了相关研究,但是相比较其他说话人识别方法,仍然达不到理想的效果。

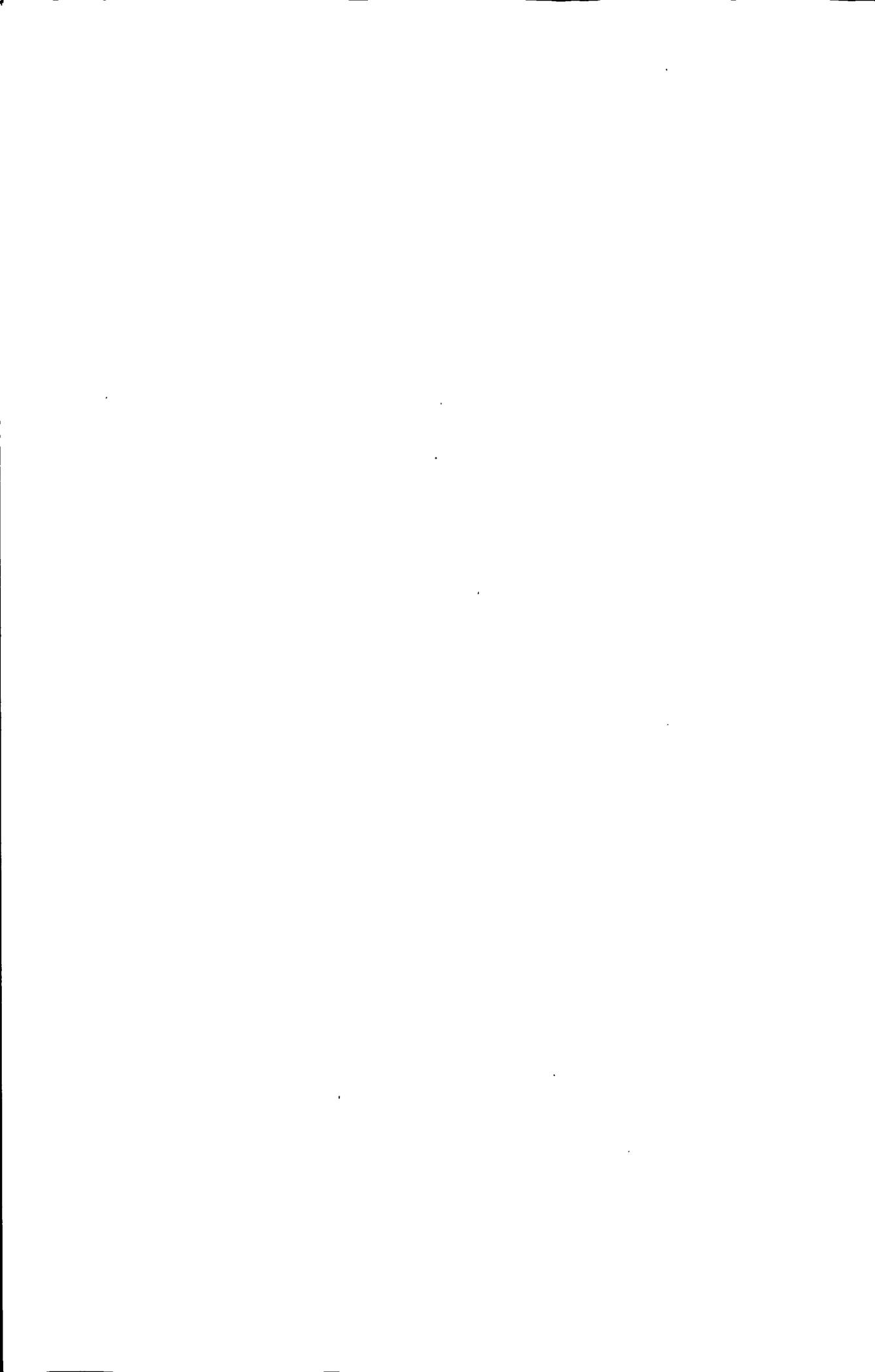
## 参考文献

- [1] 赵力.语音信号处理.第2版.北京:机械工业出版社,2003,236-253
- [2] Campbell J P.Speaker Recognition;A Tutorial [J].Proceedings of the IEEE.1977, Vol, 85, NO 9:1437-1462
- [3] Dat Tran and Michael Wagner, Fuzzy Hidden Markov Models for Speech and Speaker Recognition, [J].Lecture Notes in Computer Science,2002, 133-139
- [4] Brian Mak, James T. Kwok, and Simon Ho, USING KERNEL PCA TO IMPROVE EIGENVOICE SPEAKER ADAPTATION, Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004
- [5] V.Vapnik,The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995
- [6] D.Petrovska-Delacr'etaz,J.Cernocky,J.Hennebert,and G.Chollet, "Segmental approaches for automatic speaker verification," Digital Signal Processing, vol. 10, no.1-3, pp.198-212,2000
- [7] 姜天辰. 基于VQ/HMM的自学习语音识别系统.计算机工程与应用.[J].1998, 34(5),24-25
- [8] Chin-Ta Chen, Ching-Tai ChiangZ, Yuan-Hwang Chen, Efficient Based on Overlapped Subframes for Speaker Identification, 0-7803-6720-0/O1, 2001, IEEE
- [9] A.Harrag,T.Mohamadi, J.F.Serignat, LDA Combination of Pitch and MFCC Features in Speaker Recognition, IEEE Indicon 2005 Conference, Chennai, India, 11-13 Dec.2005
- [10]Shai Fine, Jiwi Navratil, and Ramesh A. Gopinath, A hybrid GMM/SVM approach to speaker recognition, in Proceeding of the International Conference on Acoustics. Speech, and Signal Processing, 2001, pp. I-17-I-420
- [11]徐爽. 小波分析理论在说话人识别中的应用研究.[D].燕山大学硕士论文, 2004:5-6
- [12]邱政权, 尹俊勋.针对说话人识别进行的加权小波去噪方法.[J].电声技术, 2005(9):53-56
- [13]芮贤义, 俞一彪.基于小波变换的鲁棒型特征提取及说话人识别.[J].电路与系统学报, 2005,10(5):129-132
- [14]王娜, 刘政连.基于DTW的孤立词语音识别系统的研究与实现.[J]九江学院学报(自然科学版), 2010, (3):31-33

- [15] Skowronski M D, Harris J G. Increased MFCC filter bandwidth for noise-robust phoneme recognition [J]. ICASSP2002. 2002, 801-804
- [16] 赵力. 一种引入帧间相关信息的HMM语音识别方法.[J]. 电子与信息学报, 2001,23(4): 327-331
- [17] 李灿伟, 杨震. 基于HMM的说话人确认系统的研究. [J]. 南京邮电学院学报(自然科学版), 2001,21(2):52-55
- [18] 吴金南, 宫宁生. 基于HMM和LVQ网络混合的语音识别方法.[J]. 微电子学与计算机, 2009, 26(3):49-51
- [19] 张永亮, 张先庭, 鲁宇明. FMFCC和HMM的说话人识别.[J]. 计算机仿真, 2010, 27(5): 352-358
- [20] 丁爱明, 基于MFCC和GMM的说话人识别系统研究, 河海大学, 硕士论文, 2006
- [21] 刘明辉, 戴蓓倩, 黄中伟. 自适应GMM聚类的SVM话者确认.[J]. 小型微型计算机系统, 211,(6):21-23
- [22] 连翰, 汪峥, 王建军等. 基于自适应RBFN阵列的说话人识别.[J]. 信息与电子工程, 2004,2(03):171-175
- [23] 李战明, 陈迪俊. 一种基于小波神经网络混合模型的说话人识别方法[J]. 兰州理工大学学报, 2007,33(2):77-80
- [24] 钱博, 唐振民, 李燕萍等. 基于背景噪声估计的说话人识别算法.[J]. 计算机工程. 2008,34(14):14-16
- [25] 刘雅琴, 裘雪红. 应用小波包变换提取说话人识别的特征参数[J]. 计算机工程与应用, 2006(9): 67-69
- [26] 鲍焕军, 郑方. GMM-UBM和SVM说话人辨认系统及融合的分析.[J]. 清华大学学报(自然科学版), 2008,48(z1):693-698
- [27] 陈永彬, 王仁华. 语音信号处理. 第2版. 合肥: 中国科学技术大学出版社, 1990, 25-30
- [28] S. Fine, J. Navratil, and R. A. Gopinath, "Enhancing GMM scores using SVM "hints", in Proc. 7th European Conference on Speech Communication and Technology (Eurospeech '01), Aalborg, Denmark, September 2001
- [29] Brian Mak, James T. Kwok, and Simon Ho, USING KERNEL PCA TO IMPROVE EIGENVOICE SPEAKER ADAPTATION, Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004
- [30] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, no. 1, pp. 43-49, 1978
- [31] Texas Instruments Incorporated. TMS320VC55x系列DSP的CPU与外设[M]. 北

- 京:清华大学出版社.2005, 151~388
- [32]张雄伟, 陈亮. DSP芯片的原理与开发应用.第2版.北京: 电子工业出版社, 2003, 156~367
- [33]何强,何英. MATLAB扩展编程.第1版.北京: 清华大学出版社, 2002, 291-293
- [34]刘金伟, 黄樟钦, 侯义斌.基于片上系统的孤立词语音识别算法设计.[J]计算机工程.2007,33(13), 25-27
- [35]彭启琮,武乐琴,张舰.TMS320VC55x 系列 DSP 的 CPU 与外设.第 1 版.北京: 清华大学出版社,2005
- [36]胡庆钟,李小刚,吴钰淳.TMS320C55x DSP 原理、应用和设计.第 2 版.北京: 机械工业出版社,2006
- [37]Texas Instruments, Incorporated.TPS767D301,TPS767D318,TPS767D325. 1999
- [38]张云安,冯志华.双路输出低压差电压调整器 TPS767D301 及其应用.[J]电子元器件应用, 2006
- [39]计丹,黄光明,雷环利.TLV320AIC23 CODEC 与 DSP 的接口应用.电声技术,2002(10):39~42
- [40]Texas Instruments, Incorporated.TLV320AIC23 Stereo Audio CODEC,8-to96-kHz, With Integrated Headphone Amplifer.2001
- [41]丁祥,余小清,万旺根.音频编解码器 TLV320AIC23 及其与 DSP 接口设计. [J]单片机与嵌入式系统应用.2002,06(12).35-37
- [42]胡瑛, 胡安峰.一种基于 TMS320VC5509 的语音处理系统的设计.[J]电子科技.2009,22(5).34-37
- [43]瑞泰创新.ICETEK-VC5509-A 评估板使用指导.V3.10 版
- [44]李双勋, 欧建平.TLV320AIC23 在音频处理中的应用.国外电子元器件,2003(10):16~19
- [45]汪春梅,孙洪波.TMS320C55x DSP 原理及应用.第二版.北京: 电子工业出版社,2008
- [46]hynix.4Bank x 1M x 16bits Synchronous DRAM Revision History 1.0.Apr. 2007
- [47]Am29lv800B Data Sheet.AMD. November 1997
- [48]Texas Instruments, Incorporated.SN74LVC16245 Data Sheet . NOVEMBER 1993
- [49]Si4432 Datasheet.pdf
- [50]姜雪松, 王鹰. 电磁兼容与 PCB 设计.第 1 版.北京: 机械工业出版社.2008.160-162
- [51]Texas Instruments, Incorporated.Using the TMS320C5509/C5509A Bootloader. 2003(9).
- [52]陈明义,王冠星.TMS320VC5509A 系列 DSP 的程序自举引导和加密方法.上海

- 电器技术,2008(1):30-32
- [53]薛金辉,王忠勇.TMS320VC5501 烧写 FLASH 并自举启动方法研究.[J]通信技术.2009,9(42).193-195
- [54]罗剑,胡荣强.DSP 并行引导多种方法实现的研究.[J]仪表技术.2003.25-27
- [55]郭炜,邵诗逸.DSP 通过外部 FLASH 实现自引导启动.[J]微处理机.2007.24-26
- [56]Texas Instruments, TMS320VC5509 Digital Signal Processor Silicon Errata.SPRZ006E, 2004
- [57]Texas Instruments, Incorporated.TMS320C55x DSP Programmer's Reference Guide, Literature no.SPRU376, 2000
- [58]Texas Instruments, Incorporated.TMS320VC5503/5507/5509 DSP External Memory Interface(EMIF) Reference Guide.SPRU670A,SPRU670, 2004



## 致 谢

首先感谢培育我三年的母校提供给我们良好的生活学习环境,让我在浩瀚的知识海洋中学习知识。

感谢我的导师陈若珠正高工和李战明教授的精心指导和悉心关怀,他们他们交给我的不只是科研的方法与技术,更重要的是交给我做人的道理。在攻读硕士研究生的三年中,陈老师求真务实的治学精神、渊博的知识、勇于开拓的科学精神和平易近人的态度令我终身难忘,并将深深的影响我以后的学习与工作。在此,谨向我的导师致以最崇高的敬意和最诚挚的谢意!

感谢我宿舍的舍友,生活上给我无数的关怀;感谢班级所有的同学,他们的无私的奉献让我的研究生生活充满了大家庭温暖和快乐!同时还要感谢实验室的同门师兄师姐以及师弟师妹们,认识你们使我受益良多。

最后,谨向在百忙之中抽出时间来参加我的论文答辩的各位专家和老师表示衷心的感谢!



## 附录 攻读学位期间所发表的学术论文目录

- [1] 胡金平, 陈若珠, 李战明. 基于语音识别中 DTW 改进算法的研究. 微型机与应用. 已录用, 刊登在 2011 年第 2 月。
- [2] 陈若珠, 胡金平, 李战明. TMS320VC5509 在线烧写 FLASH 并自举启动方法研究. 微型机与应用. 已录用, 刊登在 2011 年第 2 月。

