

基于单片机的智能语音识别系统设计 (硬件部分)

系 别:

专 业 班:

姓 名:

学 号:

指导教师:

基于单片机的智能语音识别系统设计 (硬件部分)

The Design of Intelligent Speech Recognition System Based on Single-chip Computer (HardWare)

摘 要

本文设计一个让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高技术的语音识别系统。本语音识别系统以 LD3320 语音识别芯片为核心部件，主控 MCU 选用 STC10L08XE 主控 MCU 通过控制 LD3320 内部寄存器以及 SPI flash 实现语音识别和对话。通过麦克风将声音信息输入 LD3320 进行频谱分析，分析后将提取到的语音特征和关键词语列表中的关键词语进行对比匹配，找出得分最高的关键词语作为识别结果输出给 MCU。MCU 针对不同的语音输入情况通过继电器对语音命令所对应的电器实现控制。同时也可以通过对寄存器中语音片段的调用，实现人机对话。

设计中，电源模块采用 3.3V 供电，主要控制及识别部分采用 LM1117-3.3 稳压芯片，语音播放及继电器部分采用 7812 为其提供稳定的电流电压。寄存器采用一片华邦 SPI flash 芯片 W25Q40AVSNIG 大小为 512Kbyte。系统声音接收模块采用的传感器为一小型麦克风——驻极体话筒，在它接收到声音信号后会产生微弱的电压信号并送给 MCU。另外系统还采用单片机产生不同的频率信号驱动蜂鸣器来完成声音提示，此方案能完成声音提示功能，给人以提示的可懂性不高，但在一定程度上能满足要求，而且易于实现，成本也不高。

关键词：语音识别 LD3320 STC10L08XE 单片机 频谱分析

Abstract

This paper designs a hi-tech speech recognition system which enables machines to transfer speech signals into corresponding texts or orders by recognizing and comprehending. The centerpiece of the speech recognition system is LD3320 voice recognition chip, its master MCU is STC10L08XE. Master MCU achieve voice conversation by controlling the internal registers and SPI flash LD3320. The sound information is inputted into LD3320 by microphone to do spectrum analysis. After analyzing the voice characteristics extracted are compared and matched with the key words in the list of key words. Then the highest scores of key words found would be output to MCU as recognition results. MCU can control the corresponding electrical real of speech recognition for different voice input through the relays and can also achieve voice conversation through a call to voice clips in register.

In the design, power module uses 3.3V. The main control and identification part adopt LM1117-3.3 voltage regulator chip, and 7812 is used to provide stable current and voltage for the part of voice broadcast and relay. Register uses chip SPI flash W25Q40AVSNIG which is 512Kbyte. The sensor used in the speech reception module of the design is microphone, namely electret microphone. After receiving the sound signal, it can produce a weak voltage signal which will be sent to MCU. In addition, the system also adopts a different frequency signals generated by microcontroller to drive the buzzer to complete the voice prompt, and this program can complete the voice prompt. The program gives a relatively poor intelligibility. However, to some extent, it can meet the requirements and is easy to implement and the cost is not high.

Key words : Speech Recognition LD3320 STC10L08XE Single-chip computer
Spectrum Analysis

目 录

摘要 I

Abstract II

绪论 1

1 设计方案 5

1.1 系统设计要求 5

1.2 总体方案设计 5

2 系统硬件电路设计 6

2.1 电源模块 6

2.2 寄存器模块 6

2.3 控制单元模块 7

2.3.1 STC10L08XE 单片机简介 8

2.3.2 STC11/10xx 系列单片机的内部结构 10

2.4 声音接收器模块 10

2.5 声光指示模块 11

2.6 语音识别模块 11

2.6.1 LD3320 芯片简介 11

2.6.2 功能介绍 12

2.6.3 应用场景 13

2.6.4 芯片模式选择 15

2.6.5 吸收错误识别 16

2.6.6 口令触发模式 17

2.6.7 关键词的 ID 及其设置 18

2.6.8 反应时间 18

3 系统软件设计 20

3.1 系统程序流程图 20

3.2 系统各模块程序设计 20

3.2.1 主程序 20

3.2.2 芯片复位程序 27

3.2.3 语音识别程序 28

3.2.4 声音播放程序 37

4 系统调试 44

4.1 软件调试 44

4.1.1 上电调试 44

4.1.2 读写寄存器调试 44

4.1.3 检查寄存器初始值 44

4.2 硬件电路调试 45

4.2.1 硬件检查 45

4.2.2 硬件功能检查 45

4.3 综合调试 46

结论 47

致谢 48

参考文献 49

附录 1 实物图片 50

附录 2 系统电路图 51

绪 论

课题背景及意义

让机器听懂人类的语音，这是人们长期以来梦寐以求的事情。语音识别是一门交叉学科，关系到多学科的研究领域，不同领域上的研究成果都对语音识别的发展作了贡献。语音识别技术就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高技术。

计算机语音识别过程与人对语音识别处理过程基本上是一致的。目前主流的语音识别技术是基于统计模式识别的基本理论。一个完整的语音识别系统可大致分为三部分：

- (1) 语音特征提取：目的是从语音波形中提取随时间变化的语音特征序列。
- (2) 声学模型与模式匹配（识别算法）：声学模型是识别系统的底层模型，并且是语音识别系统中最关键的一部分。声学模型通常由获取的语音特征通过训练产生，目的是为每个发音建立发音模板。在识别时将未知的语音特征同声学模型（模式）进行匹配与比较，计算未知语音的特征矢量序列和每个发音模板之间的距离。声学模型的设计和语言发音特点密切相关。声学模型单元大小（字发音模型、半音节模型或音素模型）对语音训练数据量大小、系统识别率，以及灵活性有较大影响。
- (3) 语义理解：计算机对识别结果进行语法、语义分析。明白语言的意义以便作出相应的反应。通常是通过语言模型来实现。

语音识别过程实际上是一种认识过程。就像人们听语音时，并不把语音和语言的语法结构、语义结构分开来，因为当语音发音模糊时人们可以用这些知识来指导对语言的理解过程。对机器来说，识别系统也要利用这些方面的知识，只是鉴于语音信号的多变性、动态性、瞬时性，语音识别技术与人类的大脑还有一定的差距。因此在语言识别技术选择，其应用模式和系统开发上，都需要专业的设计和调试，以实现最好的应用效果^[1]。

1996年9月，Charles Schwab 开通了首个大规模商用语音识别应用系统：股票报价系统。该系统有效地提高了服务质量和客户满意度，并减少了呼叫中心费用。不久，Schwab又开通了语音股票交易系统。

美国主要电信运营商 **Sprint** 的 **PCS**部门以卓越和有创新的客户服务著称。自2000年来为客户开通了语音驱动系统，提供客户服务、语音拨号、查号和更改地址等业务。2001年9月开通的可以自然方式对话的咨询系统，更实现了以自然、开放的询问方式实时获得所需要的信息。

Bell Canada，加拿大最大的电信运营商，也拥有多个语音驱动系统，提供从客户服务，增值业务和资讯服务多种功能。这些系统不但减少了用户的投诉，也为无线网络服务增值，增强了客户的忠诚度并开辟了新的收入来源。

仔细分析这些案例不难发现，一个成功的语音应用首先是基于详细准确的用户需求调查。只有详细了解客户需求，才能根据需要选用不同类型的产品。目前市场上的语音识别产品基本可分为：

特定人/非特定人：是按照声学模型建立的方式来划分。特定人识别的声学模型是针对某一特定用户训练的。一般来说用户需要先训练系统，然后才能识别该用户的发音。而非特定人识别的声学模型是针对某一种特定的语言来训练的，发音人不需要训练既可使用。虽然在建立模型时需要大量的语料，对用户来说却提供了更大方便，他们甚至不需要了解该过程。

嵌入式/服务器模式：嵌入式是将语言识别软件及模型，写在设备(如手机)的存储器里，识别过程在终端完成。在服务器模式，终端只负责收集和传导语音信号，由服务器负责完成识别。因此，对于大规模，多用户和有大量识别需求的系统，服务器模式提供了有效的方式。

目前运营商和企业用户的需求多数适合采用非特定人服务器模式的服务。例如上面提到的 **Charles Schwab**，**SprintPCS** 和 **Bell Canada** 都选用了 **Nuance Communications** 提供的语音识别软件。加之专业的系统集成方案及 **Nuance** 技术服务所提供的良好的语音用户界面设计和后期优化，使得应用系统获得了很高的客户满意度。在很多方面超过了接线员。

国内外发展现状

60年代，计算机的应用推动了语音识别的发展。这时期的重要成果是提出了动态规划(DP)和线性预测分析技术(LP)，其中后者较好地解决了语音信号产生模型的问题，对语音识别的发展产生了深远影响。

70 年代，语音识别领域取得了突破。在理论上，LP 技术得到进一步发展，动态时间归正技术(DTW)基本成熟，特别是提出了矢量量化(VQ)和隐马尔可夫模型(HMM)理论。在实践上，实现了基于线性预测倒谱和 DTW 技术的特定人孤立语音识别系统。

80 年代，语音识别研究进一步走向深入，其显著特征是 HMM 模型和人工神经网络(ANN)在语音识别中的成功应用。HMM 模型的广泛应用应归功于 AT&T Bell 实验室 Rabiner 等科学家的努力，他们把原本艰涩的 HMM 数学模型工程化，从而为更多研究者了解和认识。ANN 和 HMM 模型建立的语音识别系统，性能相当。

进入 90 年代，随着多媒体时代的来临，迫切要求语音识别系统从实验室走向实用。许多发达国家如美国、日本、韩国以及 IBM、Apple、AT&T、NTT 等著名公司都为语音识别系统的实用化开发研究投以巨资。

我国对语音识别的研究起步相比于发达国家来说比较晚，但是经过科研工作者们的不懈努力，现在已经取得了巨大的突破。1988 年，在清华大学和中科院声学所的努力下，无限词汇的汉语听写机被研制成功。此后不少公司陆续推出自己的汉语听写机产品，大大促进了汉语语音识别的发展。之后，在国家的大力支持下，很多国内单位在语音识别上进行了大量的探索研究，为汉语的语音识别做出了杰出的贡献。

语音识别的应用场合

目前，语音识别技术已经在生活中的诸多方面有了广泛的应用，而这方面的例子已经屡见不鲜，其主要的应用场合如下

(1) 声控应用

在很多场合与特定环境中，使用声控系统可以大大提高生活质量和便利程度。比如司机通过语音进行电话拨号，可以减少交通事故的发生率，而很多行动不便的人通过声控系统更是可以将生活质量提高。

(2) 数据库检索

在不少行业部门当中都需要对庞大的数据进行检索和查询，这是一项非常复杂的工作。语音识别的引入可以大大的减少劳动量，既方便又节约人力成本。

(3) 语音听写机

用口述来代替人工输入文档，提高了输入速度，而且更加方便，可以为办公

出品的 Via Voicc 系列产品就是这个方面典型的应用。

(4) 自动口译

将语音识别应用于自动口语翻译是未来的一大趋势，不仅可以节省大量的人力。同时也让翻译工作变得更加轻松简单。一些发达国家在自动口译方面的研究已经取得了一定的进展。

目前，语音识别技术仍在高速发展之中。随着时代的进步，越来越多的领域开始引入这项技术，人类的日常生活会级此而带来更大改变。

虽然在过去的几十年语音识别技术已经取得了长足的发展，但是现在的技术还不足以让电脑像人一样随意交谈。说来，主要有以下几个方面：

首先，语音识别系统的鲁棒性不够强，对环境的依赖程度过高。在某一种环境下训练的语音识别系统换了一种环境之后性能就会下降。

其次，语音识别对于外部噪声特别敏感。这不仅是因为外部噪声会导致语音信号发生变化，而且由于嘈杂的环境下人的音调，语速以及音量都会改变，因此识别难度也更大。

再次，语音的随机性很强。就算是同一个人在不同的时刻，由于身心状态的差异，导致语音的特征也会不一样。

最后，由于目前对人类的听觉理解、知识积累和神经系统的机理等方面的研究水平不足，限制了语音识别的发展。

为了解决上述问题，研究者们想出了各种方法，比如自适应训练、神经网络等。这些做法虽然都取得了一定的成绩，然而，如果要想让语音识别系统的性能得到大幅的提高，还有大量的工作要做。

目前，市场上大词汇量的语音识别系统多采样 PC 机作为硬件平台，而基于嵌入式的小词汇量的语音识别系统，其硬件设计常采用 DSP 或者 AUM 这样的高性能芯片，这样硬件成本较高。对于单片机来说，虽然成本低，但由于单片机本身计算能力有限，而语音识别的计算量过大，这对系统在单片机上的实现带来了很大的困难。因此，如何改进算法以减少计算量成为了语音识别能否在单片机上运行的一大难点。

1.1 系统设计要求

(2) 语音点播：

- ①唱儿歌
- ②朗诵一古诗
- ③讲笑话

本课题旨在开发一个将语音识别技术，传感器技术等技术应用其中，并集教育性、娱乐性和互动性于一体的智能语音识别系统。

1.2 总体方案设计

本语音识别系统以 LD3320语音识别芯片核心部件，制作了一个语音识别控制系统。本系统主控 MCU选用 STC10L08XE 主控 MCU通过控制 LD3320内部寄存器以及 SPI flash 实现语音识别语音对话。通过 MIC将声音信息输入 LD3320进行频谱分析，分析后将提取到的语音特征和关键词语列表中的关键词语进行对比匹配，找出得分最高的关键词语作为识别结果输出给MCUMCU针对不同的语音输入情况同通过继电器对语音命令所对应的电器实现控制。同时也可以通过对寄存器语音片段的调用，实现人机对话。该系统框架图如下图1-1 所示：

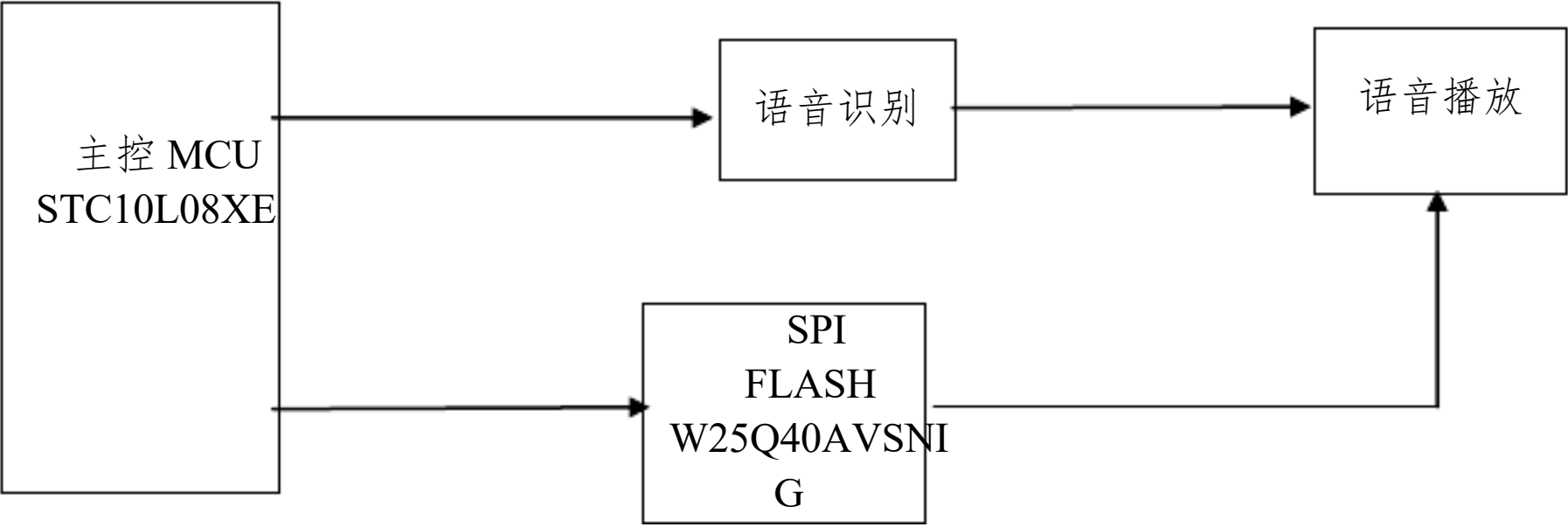


图 1-1 系统框架图

2.1 电源模块

供电，主要控制及识别部分采用 LM1117-3.3 稳压芯片，语音播放及继电器部分采用 7812 为其提供稳定的电流电压。电源模块电路图如图 2-1 所示：

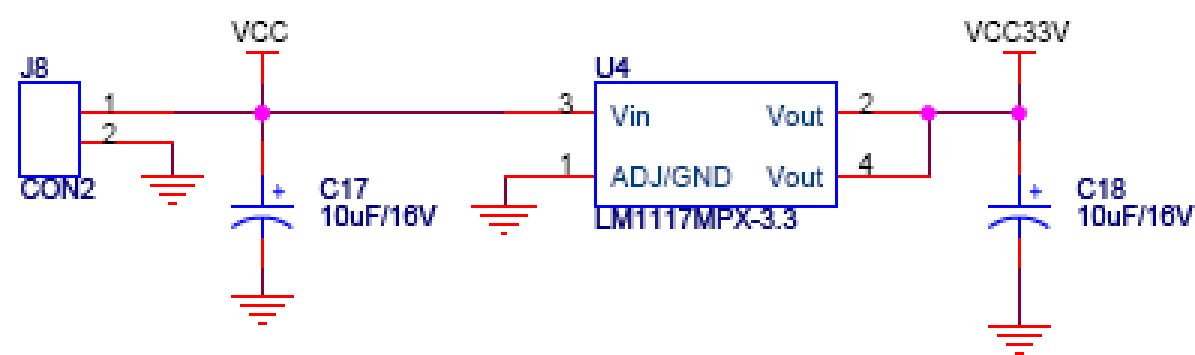


图 2-1 电源模块电路图

(1) 参数特性

LM1117-3.3 参数特性：

输入电压 (V) 4.75~12

输出电流 (A) 1

输出电压 (V) 3.3

压差 (V) 1.3

(2) 封装

类型：SOT-223

引脚：3

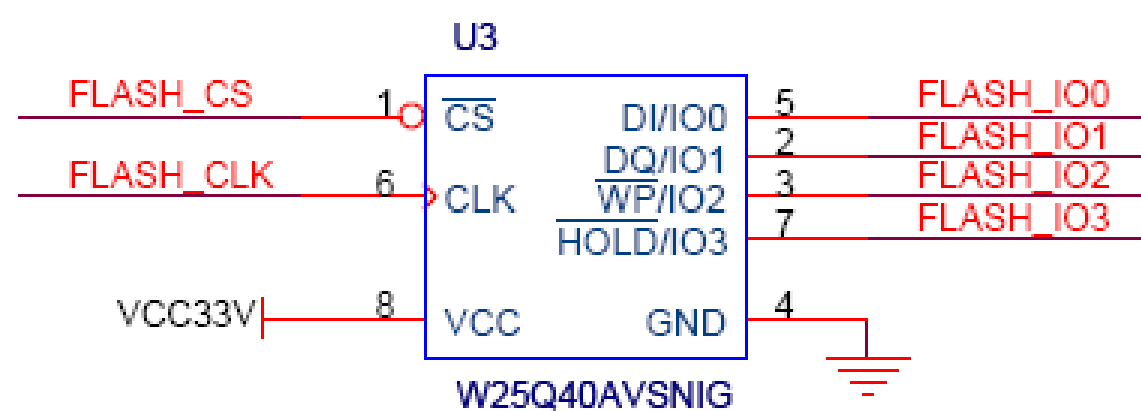
体宽：6.30-6.71 mm

长度：3.30-3.71 mm

引脚间距：2.29 mm

2.2 寄存器模块

本系统采用一片华邦 SPI flash 芯片 W25Q40AVSNIG 大小为 512Kbyte。用于存储人机对话时的声音素材。寄存器模块电路图如图 2-2 所示：



寄存器模块电路图

- (1) 能够减少微处理器的封装管脚数目；
- (2) 线路板 (PCB) 可以更小，更简单；
- (3) 能够减少系统电路切换噪声；
- (4) 能够减低系统功能及制造成本。

2.3 控制单元模块

采用 STC10L08XE 单片机作为系统的主控制单元。STC10L08XE 系列单片机具有强大的数据处理功能，且 I/O 口控制灵活。控制单元模块电路图如图 2-3 所示：

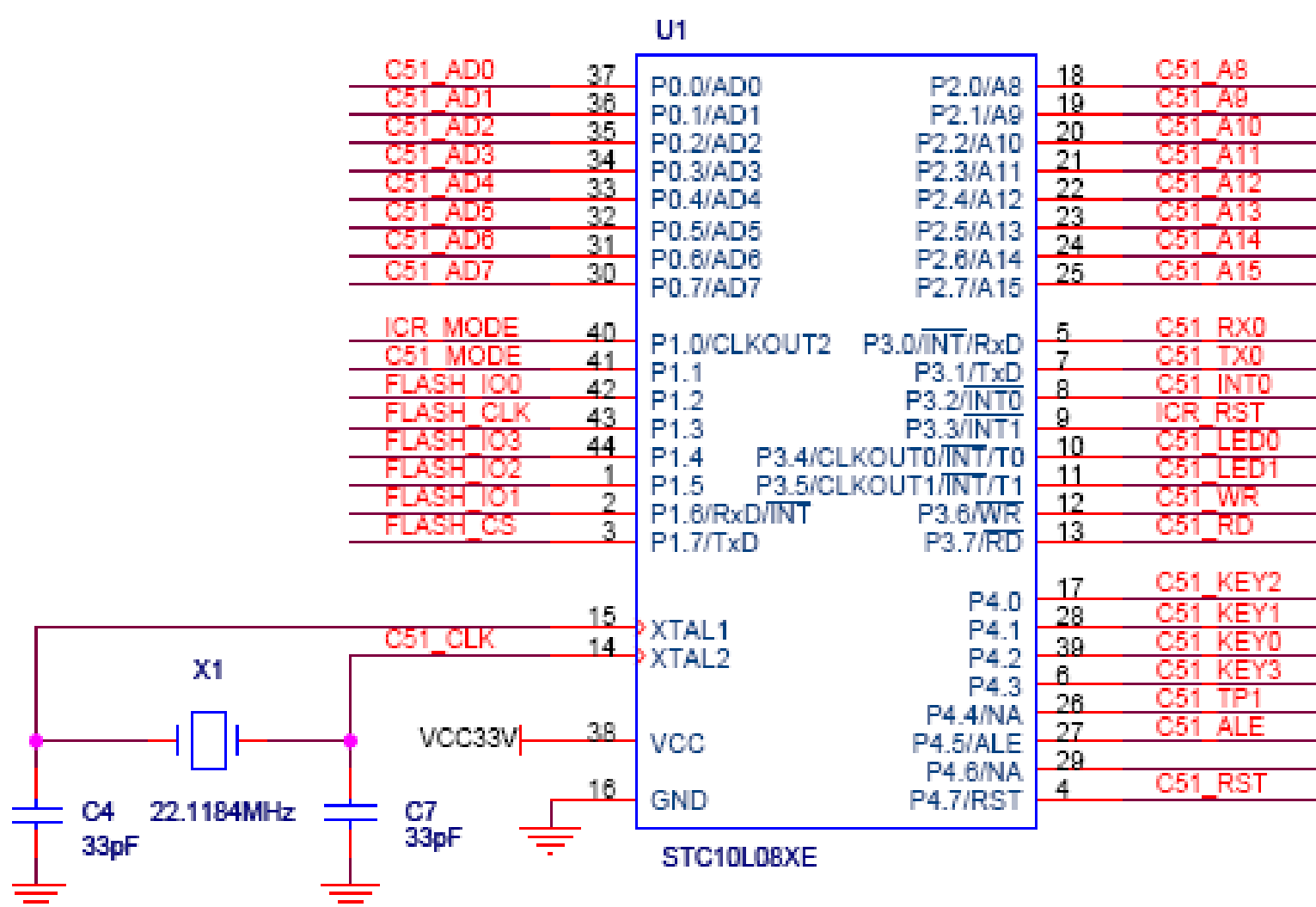


图 2-3 控制单元模块电路图

/机器期(IT) 的单片机, 是高速/低功耗/超强抗干扰的新一代 8051 单片机, 指令代码完全兼容传统 8051, 但速度快 8-12 倍。内部集成高可靠复位电路, 针对高速通信, 智能控制, 强干扰场合^[1]。

11/10xx 系列单片机的定时器 0/定时器 1 串行口与传统 8051 兼容, 增加了独立波特率发生器, 省去了定时器 2。传统 8051 的 111 条指令执行速度全面提速, 最快的指令快 24 倍, 最慢的指令快 3 倍。

增强型 8051 CPU, IT, 单时钟/机器周期, 指令代码完全兼容传统 8051。

工作电压:

STC11F 系列工作电压:5.5V - 4.1V / 3.7V (5V 单片机);

STC11F 系列工作电压:3.6V - 2.4V / 2.1V (3V 单片机);

STC11Fxx系列工作电压:5.5V - 3.8V / 3.3V (5V 单片机);

STC11Lxx系列工作电压:3.6V - 2.4V / 2.1V (3V 单片机)。

工作频率范围:0 - 35MHz, 相当于普通 8051 的 0~420MHz

STC11F/Lxx 系列单片机用户应用程序空间:

1/2/3/4/5/6/8/16/20/32/40/48/52/56/60/62K 字节。

STC11F/Lxx系列单片机用户应用程序空间: 4K/6K/8K/10K/12K/14K 字节。

STC11xx系列单片机:片上集成 1280 字节或 256 字节 RAM;

STC11XX系列单片机:片上集成 51 字节或 256 字节 RAM

通用 I/O 口 (36/40/12/14/16 个) 每个 I/O 口驱动能力均可达到 20mA复位后为:

(1) 准双向口/弱上拉(普通 8051 传统 I/O 口);

(2) 可设置成四种模式:

①准双向口/弱上拉;

②强推挽/强上拉;

③仅为输入/高阻;

④开漏。

ISP(在系统可编程)/IAP(在应用可编程), 无需专用编程器, 无需专用仿真器。

可通过串口(RxDIP3.0, TxD/P3.1) 直接下载用户程序, 数秒即可完成一片^[2]。

功能。

看门狗。

内部集成 MAX810 专用复位电路(晶体频率在 24MHz 以下时, 要选择高的复位门槛电压, 如 4.1V 以下复位, 晶体频率在 12MHz 以下时, 可选择低的复位门槛电压, 如 3.7V 以下复位, 复位脚接 1K 电阻到地)。

内置一个对内部 Vcc。进行掉电检测的掉电检测电路, 可设置为中断或复位。5V 单片机掉电检测门槛电压为 4.1V~3.7V 附近, 3.3V 单片机掉电检测门槛电压为 2.4V 附近。

时钟源: 外部高精度晶体/时钟, 内部 R/C 振荡器。用户在下载用户程序时, 可选择是使用内部 R/C 振荡器还是外部晶体/时钟;

常温下内部侧 R/C 振荡器频率为: 4MHz~8MHz 精度要求不高时, 可选择使用内部时钟, 但因为有制造误差和温漂, 以实际测试为准。共 2 个 16 位定时器(与传统 8051 兼容的定时器/计数器, 16 位定时器 T0 和 T1) STC10xx 全系列都有 1 个独立波特率发生器。3 个时钟输出口, 可由 T0 的溢出在 P3.4/T0 输出时钟, 可由 T1 的溢出在 P3.5/T1 输出时钟, 独立波特率发生器可以在 P1.0 口输出时钟。外部中断 I/O 口 5 路, 传统的下降沿中断或低电平触发中断, Power Down 模式可由外部中断唤醒, INTO/P3.2, INTI/P3.3, INT/T0/P3.4, INT/T1/P3.5, INT/INT/RxD/P3.0 或 INT/RxD/P1.6

Power Down(掉电)模式可由内部掉电唤醒, 可用定时器唤醒(STC11xx 系列有此功能, STC10xx 无此功能), 也可由上面提到的外部中断口中断唤醒, 由于 INT/RxD 支持下降沿中断, 故也可支持远程通信唤醒。

一个独立的通用全双工异步串行口(UART, 做主机时可以当 2 个串口使用 [RxD/P3.0, TxD/P3.1] 可以切换到 [RxD/P1.6, TxD/P1.7], 通过将串口在 P3 口和 P1 口之间来回切换, 将 1 个串口作为 2 个主串口分时复用, 可低成本实现 2 个串口, 当然有其局限性。

工作温度范围: -4°C - +85°C (工业级) / 0°C - 75°C (商业级)。

封装: SOP16/DIP

16/DIP18/SOP20/DIP20/LSSOP20/PDIP-40/LQFP-44/PLCC-44(暂时不要选 PLCC

-44)。SOP16有 12 个 I/O 口，DIP18 有 14 个 I/O 口，SOP20/PDIP20/LSSOP20有 16 个 I/O 口，LQFP44有 40 个 I/O 口。

2.3.2 STC11/10xx 系列单片机的内部结构

STC11/10xx系列单片机的内部结构框图如图 2-1 所示。STC11/10xx 单片机中含中央处理器(CPU)、程序存储器(Flash) 、数据存储器(SRA)、定时/计数器、UART串口、I/O 接口、看门狗及片内振荡器和外部晶体振荡电路等模块。STC11/10xx 系列单片机几乎包含了数据采集和控制中所需的所有单元模块，可称得上一个片上系统。图 2-4 STC11/10xx 的功能与结构图如图 2-4 所示：

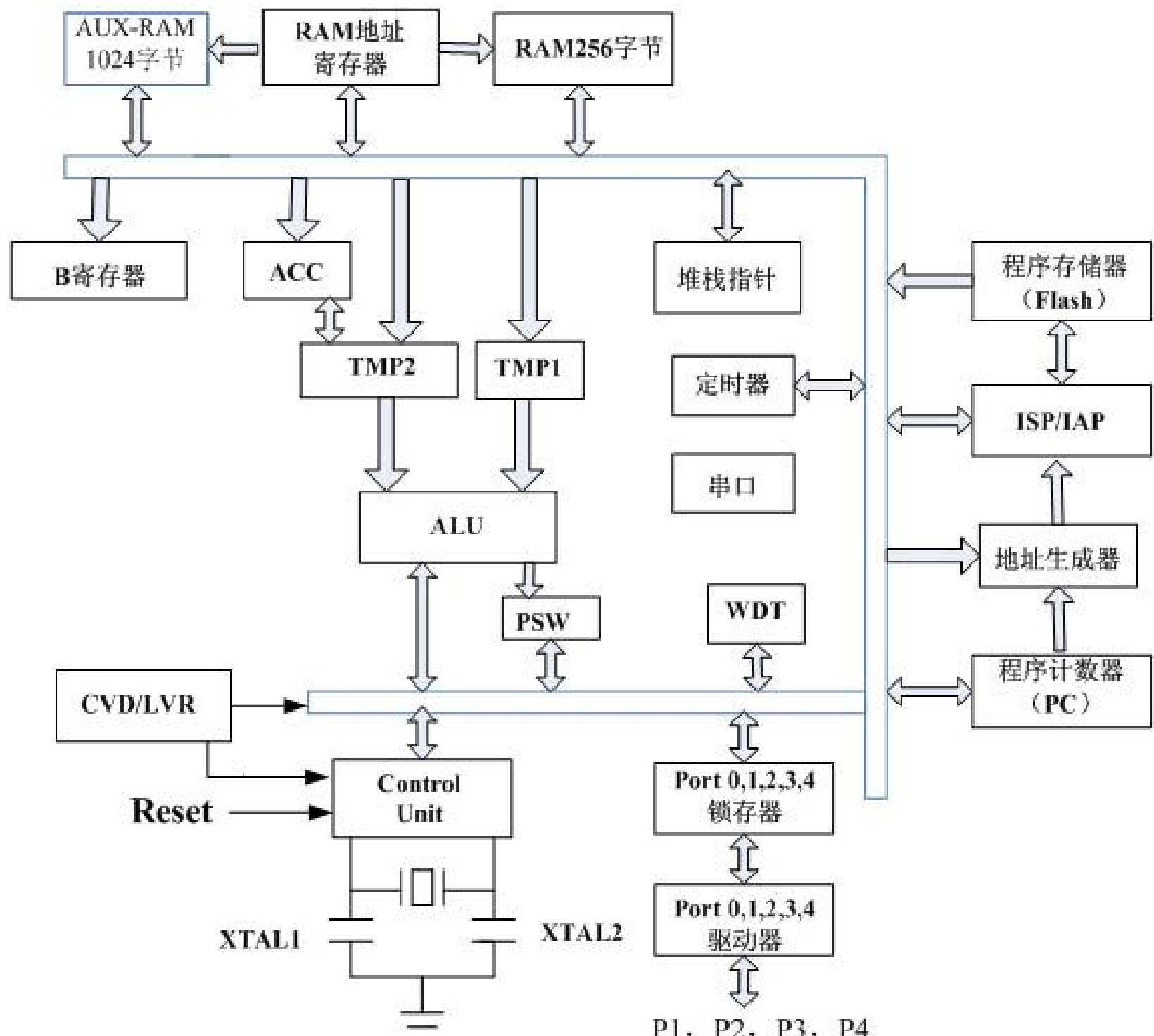


图 2-4 STC11/10xx 系列单片机的内部结构框

2.4 声音接收器模块

系统声音接收模块采用的传感器为一小型麦克风—驻极体话筒。在它接收到声音信号后会产生微弱的电压信号。声音接收器模块如图2-4所示

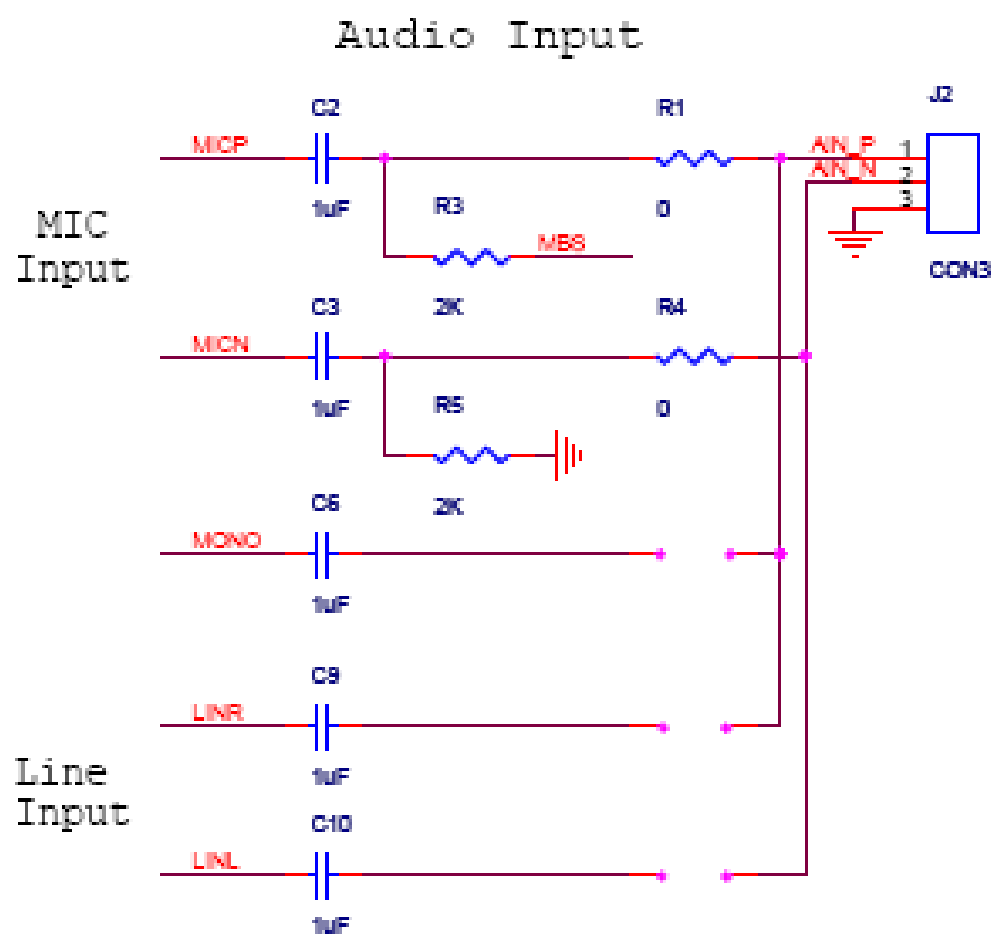


图 2-4 声音接收器模块电路图

2.5 声光指示模块

采用单片机产生不同的频率信号驱动蜂鸣器来完成声音提示，此方案能完成声音提示功能，给人以提示的可懂性不高，但在一定程度上能满足要求，而且易于实现，成本也不高。

2.6 语音识别模块

采用芯片 LD3320，一颗基于非特定人语音识别(SI — ASR: Speaker — Independent Automatic Speech Recognition) 技术的语音识别/声控芯片。提供了真正的单芯片语音识别解决方案。语音识别模块电路图如图2-6所示。

2.6.1 LD3320 芯片简介

LD3320 是一颗基于非特定人语音识别(SI-ASR: Speaker-Independent Automatic Speech Recognition) 技术的语音识别/声控芯片。提供了真正的单芯片语音识别解决方案。

LD3320 芯片上集成了高精度的A/D和D/A接口，不再需要外接辅助的Flash 和RAM即可以实现语音识别/声控/人机对话功能。并且，识别的关键词语列表是可以动态编辑的。

基于 LD3320 可以在任何的电子产品中，甚至包括最简单的 51 作为主控芯片的系统中，轻松实现语音识别/声控/人机对话功能。为所有的电子产品增加 VUI(Voice

User Interface) 语音用户操作界面。

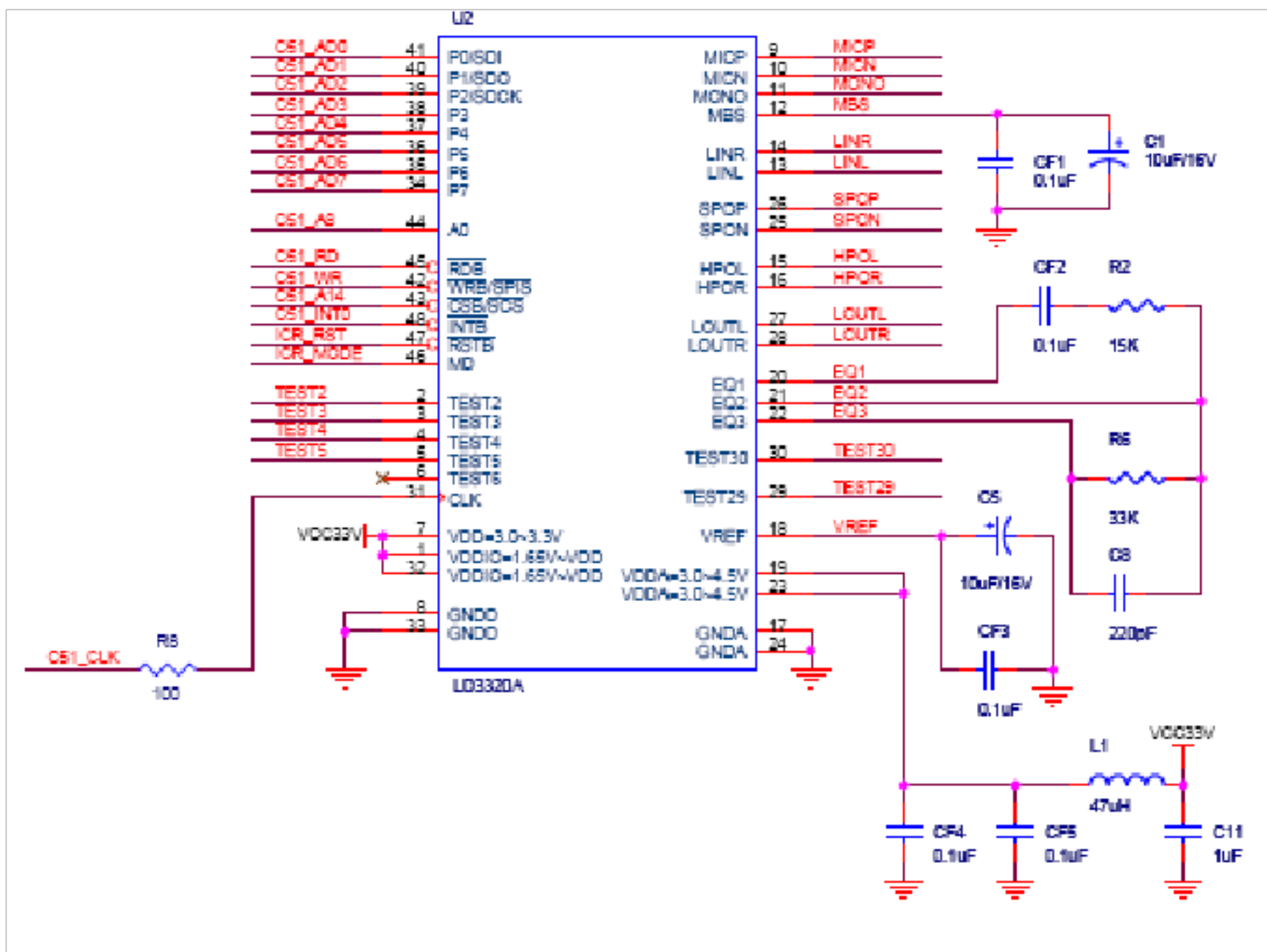


图 2-6 语音识别模块电路图

2.6.2 功能介绍

尺寸:2*6.2cm

排针:2*20 标准 DIP40 排针。

LD3320芯片的音频模拟管脚连接相应的电容/电阻后通过排针引出。M-LD3320模块上设计有 2 个音频插座，直接引出 **MIC**输入和 **Speak:** 输出信号。用户可以用一个带麦克风的耳机验证语音识别和声音播放，十分方便。M-LD3320模块上没有电源芯片，相应的电源管脚由排针引出，由开发者连接入 3.3v 电源输入。M-LD3320模块上的 **CLK**输入可以选择如下任意一种：

- (1) 直接将晶振信号通过排针输入到 LD3320 的相应管脚。

- (2) 或者用户可以自行焊接晶振, 在模块上预留晶振的空间和连接点 [3]。M-LD3320 模块上有两个 LED 灯, 连接到 LD3320 芯片的 29, 30 管脚上, 在 LD3320 上电重启复

位(RSTB*)并稳定工作后, 29, 30 管脚会稳定输出低电平, 因此这两个 LED 灯可以作为芯片上电指示。

2.6.3 应用场景

由于 LD3320 可以动态编辑的识别关键词语列表, 因此其可以应用的范围大大超过了那些不可以改变识别列表的芯片。又由于 LD3320 是一颗真正的单芯片解决方案, 不需要辅助的外围 Flash/RAM/AD 从而降低了系统的成本。

其潜在的应用场景主要包括:

(1) 电磁炉/微波炉/智能家电操作

在现在的家电中, 各种各样的设置越来越繁复。用户在使用过程中, 还要不断地对家电进行功能切换。在引入 LD3320 芯片后, 可以用语音直接控制这些家电。比如用语音来控制电磁炉把火力调整到“煎炸”或者是“慢炖”。

(2) 导航仪

在导航上, 语音识别应用最主要的限制因素包括:

①POI 地名数量过多, (一个城市经常有数万个 POI 地名), 使得在这么多的地名中进行语音识别搜索很难达到满意的地步。

②录音效果的制约。现在一般的导航机器都是安置在车子的前端, 距离用户的距离很遥远, 录音的质量无法得到保障, 也就无法达到满意的识别效果。

因此在导航上比较可行的语音操作方案是在语音命令/控制/查询导航系统上。也就是用语音在开车过程中去对导航进行一些操作, 比如重新规划路径, 查询前方加油站, 播报方位, 查询附近的餐饮, 查询车速等等这样的命令。同时, 对于录音距离的问题, 可以在麦克风集成 LD3320 后, 安装在保险带上(司机的胸前位置)。这样可以最好效果地接受司机的语音指令, 并把相应的识别结果发送到导航仪上就可以(比如用最简单的红外就可以把识别结果通知到导航仪)。这样的应用, 在导航车载环境中, 是最具可行性和实用性的语音操作功能。

(3) MP3/MP4

目前, 在手机中, 语音点歌是一个很受市场欢迎的亮点。在独立的 MP3/MP4 设备中, 加入 LD3320 芯片后, 也可以方便地实现语音点歌的功能。随着用户把歌曲拷贝进 MP3/MP4 同时把歌曲名字传入进 LD3320 芯片, 这样用户再说出歌曲名字的时候

候，就可以得到识别，并播放该歌曲。

(4) 数码像框

数码像框中存放了许多的照片和视频，同时又具有多种播放的方式。一般是通过按键或者遥控器的方式来对其进行操作，但是这样的操作并不方便。在引入了 LD3320 提供的语音识别功能后，用户可以最自然地用语音去点播想要显示的照片，或者改变数码像框的显示方式。使得这样的数码产品更加具有人性化的操作界面。

(5) 机顶盒/彩电遥控器

随着数字电视的普及，家庭中可以收看到的电视节目也越来越丰富。大家也就苦于在众多的频道中迅速选择到自己想要看的频道。在把 LD3320 语音识别芯片集成进机顶盒/彩电遥控器后，用户只需要对着遥控器说出想要看的电视频道的名字，就可以快速地选择。比如用户可以说出“奥运体育”，就可以转到体育频道来观看精彩的体育比赛了。

(6) 智能玩具/对话玩具

在电视购物中曾经出现过可以人机对话的玩具，比如金福猪，金福狗等。

这些玩具采用的语音识别只能支持固定的 10 条左右的语音命令，比如“你好”，“我想听歌”等等。

采用 LD3320 芯片，可以利用其动态编辑识别关键词语列表的性能，让玩具实现及其复杂的对话脚本。避免了玩具的严重同质化。

(7) PMP/ 游戏机

目前的游戏机操作都是基于键盘的操作。在引入了语音界面后，可以用语音对游戏进行操作，增加游戏的趣味性。比如在游戏中，可以通过语音来吟诵一段魔法咒语。或者用语音与游戏对着猜拳。又或者可以通过语音来快速地切换游戏的视角，组队。

(8) 自动售货机

在有了 LD3320 语音界面后，人们可以对着售货机说出要买的商品就可以，比如“可口可乐”或者“面巾纸”。

(9) 地铁自动售票机

在大都市中，无数不熟悉道路的人在买地铁票时，总是无法知道自己要去的站

是多少票价。有了语音界面后，只需要对着售票机说出要去的地方，就可以方便地买到车票。

(10) 导游机

在引入 LD3320 的语音操作界面后，游客可以随时用语音向导游机发出指令，查询某个景点的介绍和特征。避免了使用按键在导游机上麻烦地输入景点的名称，也避免了导游机没有人机互动 Ifu 成为单纯的播放声音的录音带。

(11) 楼宇电视的广告点播

目前分众传媒等公司的楼宇广告设施，遍布几乎所有的写字楼，也在广告投放上取得了良好的效果。但是目前用户在接受广告时，都是被动地去接受信息。对于其中感兴趣的广告，只能是等待下次再看到时进行仔细地了解，没有办法进行主动式地广告查询。引入 LD3320 语音识别芯片后，用户可以用语音去方便地查询想要了解的广告信息。比如操作楼宇广告“重新播放”“上一条”来重新观看一条广告。或者“汽车”来点播想要看的汽车广告。这样的语音操作，不需要增加额外的键盘输入和触摸屏输入，又可以让用户与广告充分互动，取得更佳的效果。

(12) 公共照明系统/卫生系统/智能家居的声控

可以用声音来对照明，卫生，或者其他一些家居系统作命令控制。躺着就可以说话控制家里的灯光开关，在公共场合帮助残疾人士非接触地去控制卫生系统的运行，可以听懂人说话的智能家居，这些都是人 I 智能科技力图达到的效果。

使用 LD3320 芯片可以在一定程度上完成这样的语音控制系统，给人们的生活带来更便利的语音交互界面。VUI(Voice UserInterface) 作为一种新的人机交互界面，会逐步地走入人们的口常生活，在适合用语音控制的地方给人们提供更多的便利。

2.6.4 芯片模式选择

用户可以通过编程，设置两种不同的用户使用模式：“触发识别模式”和“循环识别模式”：

(1) 触发识别模式：

系统的主控 MCU 在接受到外界一个触发后(比如用户按动某个按键)，启动 LD3320 芯片的一个定时识别过程([比如 5 秒钟)，要求用户在这个定时过程中说出要识别的语音关键词语。过了这个过程后，需要用户再次触发才能再次启动一个识别过程。

(2) 循环识别模式:

系统的主控 MCU 反复启动识别过程。如果没有人说话没有识别结果,则每次识别过程的定时到时后再启动一个识别过程;如果有识别结果,则根据识别作相应处理后(比如播放某个声音作为回答)再启动一个识别过程。一般来说,触发识别适合于识别精度要求比较高的场合。外界触发后,产品可以播放提示音或者其他方式来提示用户在接下来的几秒钟内说出要识别的内容,这样来引导用户在规定的时间内只说出要识别的内容,从而保证比较高的识别率。而循环识别比较适合需要始终进行语音监控的场合,或者没有按键等其他设备控制识别开始的场合。而这种状态,识别准确度会有一定下降,在循环识别的过程中,用户的其他说话声音,或者外界的其他声音,都有可能被识别引擎误识别出错误的结果,需要产品的控制逻辑都作相应的处理。

在识别精度要求高的场景中,应该采用“触发识别”模式。原因是

(1) 用户在每次按热键后,精神处于最集中的状态,此时用户说的语音命令会比较认真,清晰。避免了用户过于随意的发音导致的识别误差。

(2) 每次按热键后,产品应该给以一个明显的开始信号,比如发出“当”的一声或者其他提示信号,可以给用户一个明确开始的提示,方便用户掌握说语音命令的时间。

(3) 由于按键触发后,用户就会贴近麦克风并说出语音命令,避免了其他环境声音被录入 LD3320 芯片导致的误识别。

另:这种方式还是一种省电的方式,在不识别时,彻底不让芯片工作以省电。

2.6.5 吸收错误识别

在设定好要识别的关键词语后,为了进一步降低误识别率,可以再添加一些其他的任意词汇进识别列表,用来吸收错误识别,从而达到降低误识别率的目的。可以把这些关键词语称之为“垃圾关键词语”。

比如,某个应用场景中,需要识别的关键词语是 4 条,“前进”,“后退”,“开门”,“关门”。在把这 4 个关键词语设置进 LD3320 后,可以再另外设置 1030 个词语进 LD3320 比如“前门”,“后门”,“阿阿阿”,“呜呜”等等。

只有识别结果是 4 个关键词语之内的,才认为识别有效。如果识别结果是“垃圾

关键词语”，则说明是其他的声音导致的误识别，产品应该重新开始一次识别过程。这样，可以非常非常有效地降低误识别率。极大地提高终端用户的主观使用体验。“垃圾关键词语”的选取，最好可以选择一些字数和关键词语一样的词语，用来吸收可能发生的错误识别。

需要说明的是：这一方法，即可以应用在“触发识别”模式中，也可以应用在“循环识别”模式中。这样作的原理是非特定人语音识别技术 **ASR** 是一个基于关键词语列表的匹配识别技术，算法本质是在提取输入声音的特征后，在关键词语列表中寻找一个相似度最高的词语作为识别结果。因此，任何的声音输入进语音识别芯片，都会去和关键词语列表中的词语进行匹配对比，并且_也都会依次打分。这样，其他人在随意聊天，或者任意说一个不在关键词语列表中的命令，或者是其他毫无联系的说话声音，都可能会匹配到某一个关键词语并作为结果输出。从而造成误识别。虽然算法设计中有一定的算法来避免出现这样的误识别，但还是不可避免。产品开发者可以在芯片外部针对性的处理以降低误识别率。本节提供的方法，是非常有效的一种方法，在实际应用中具有非常重要的地位。

2.6.6 口令触发模式

在一些应用场合，希望识别精度高，但是又无法要求用户每次都用手按键来“触发识别”。此时，可以采用“口令触发模式”。

产品定义一句短语，作为触发口令。比如，可以定义“芝麻开门”作为触发口令。产品在等待用户触发时，启动一个“循环识别”模式，把触发口令“芝麻开门”和其他几十个用来吸收错误的词汇设置进 **LD3320**。只有当检测到识别出的结果是触发口令时，才认为是终端用户叫了这个口令。此时，给出提示音，并启动一个“触发识别模式”，并把相应的识别列表设置进 **LD3320**，提示用户在提示音后几秒钟内说出要执行的操作。

在等待用户的过程时，如果识别的结果是那些用来吸收错误的词汇，则认为是误识别，或者其他的声音干扰，而不进行任何的处理，直接再次进入“循环识别”模式。

这种口令触发模式，融合了其他两种模式的优点，并且_结合第二节提到的“垃圾关键词语”的方法，可以为产品提供更加方便实用的语音操作特性。

2.6.7 关键词的 ID 及其设置

在把关键词语设置进 LD3320 时，是把关键词语的拼音串传入 LD3320 并同时传入一个 ID，用来代表这个关键词语。LD3320 的识别结果，也是把识别出的关键词语的 ID 作为结果输出。

在 LD3320 芯片，不同的关键词语是可以对应同一个 ID。而且 ID 不需要是连续的。这样就为产品开发者提供了很方便的编程手段。

例如：“北京”，“首都”，可以设置为同一个 ID，进行后续处理。

例如：在使用第二节提到的“垃圾关键词语”时，可以把添加的这些用来吸收错误的关键词语的 ID 都标记成一个值，或者把它们标注为比较特殊的 ID 值，如大于 200。在程序中就比较简单，很容易处理误识别，避免了添加进很多关键词语后，写程序中需要为这些关键词语的处理增加过多的程序分支。

终端用户在说语音指令时，可能对同一个词汇有不同的发音习惯。

例如，“打开电灯”，用户可能会说“开灯”，“打开灯”，“打开电灯”，“把灯打开”等等。

充分利用 LD3320 的 50 条可动态编辑的关键识别条目的特性，开发者可以把这些习惯发音都设置进芯片，这样无论用户怎么说，都会被正确识别出来，进一步增加终端用户的良好体验。

同时，可以结合第四条秘籍“巧妙利用关键词语的”，在编程中可以很方便地处理这些多个习惯发音。值得注意的是，如果用来进行控制工作，需要加入一些垃圾关键词吸收错误以降低误识率。

2.6.8 反应时间

LD3320 芯片内部是通过 VAD(端点检测) 机制来判断人是否说完语音，并给出识别结果根据 VAD 机制，语音识别芯片监测出有一段连续的背景噪音后，认为用户已经说完了语音识别命令，然后再给出识别结果。

默认设置是监测到在人声开始后有连续的 600 毫秒的不说话时，才会给出识别结果。也就是说，根据默认设置，从人说话结束，到语音识别芯片主动送出结果中断，至少要有 600 毫秒的间隔，如果用户希望调节这个反应间隔，可以从以下几方面入手：

(1) 改变使用方式

采用类似于步话机的方式，每次人按键后，按下不放，开始说命令，说完命令后，松开按键，每次检测到松开按键时，主控的单片机通过设置 BC寄存器来立即获得识别结果。

(2) 修改 VAD判断的寄存器

ASR: Vad Silence End 在语音检测到语音数据段以后，又检测到背景噪音段，连续检测到多长时间的背景噪音段才可以确认为是真正的语音结束。每 1 单位，10 毫秒。**Default:** 60, 相当十 600 毫秒数值范围:20200(相当于 2002000 毫秒)但是这个修改会导致，如果这个时间过短，导致用户在说话时的说话停顿也会造成 VAD检测认为说话结束，从而降低某些用户的识别率。

(3) 修改麦克风的音量，35 寄存器，(建议调整范围在 40H~58H之间)，看哪个录音增益适合使用的麦克风，以及使用的环境。

(4) 修改 B8 寄存器

比如修改为 2，那么这意味着，无论如何，在每次识别开始后 2 秒钟的时间内，必然会停止识别给出一个识别结果。（这个设置不影响 VAD检测）。

如果 b8 值特别小，比如设置:1, 2, 3, 就需要在开始识别前，给用户一个很明确的提示，要开始识别了。免得用户还没有准备就识别时间过去了。但这个间隔设置的过短，也必然会引起一些可能存在的误识别，比如语音命令比较长，那么这个时间设置的太小，就会造成比较长的语音命令无法在特定时间内完整念完引起误识别。所以当这个数值设置比较小的时候，一般建议使用“触发识别”的用户界面，避免使用“循环识别”的用户界面。

(5) 改变使用环境，或许在某些环境中的噪声或者回声会影响到判断说话结束。以及说话人自己的音量，如果声音很低，也会导致判断人说话是否结束比较困难。改变命令词语内容，比较好念，开口音响亮等，方便使用者连续清晰念出语音命令。

3.1 系统程序流程图

所示：

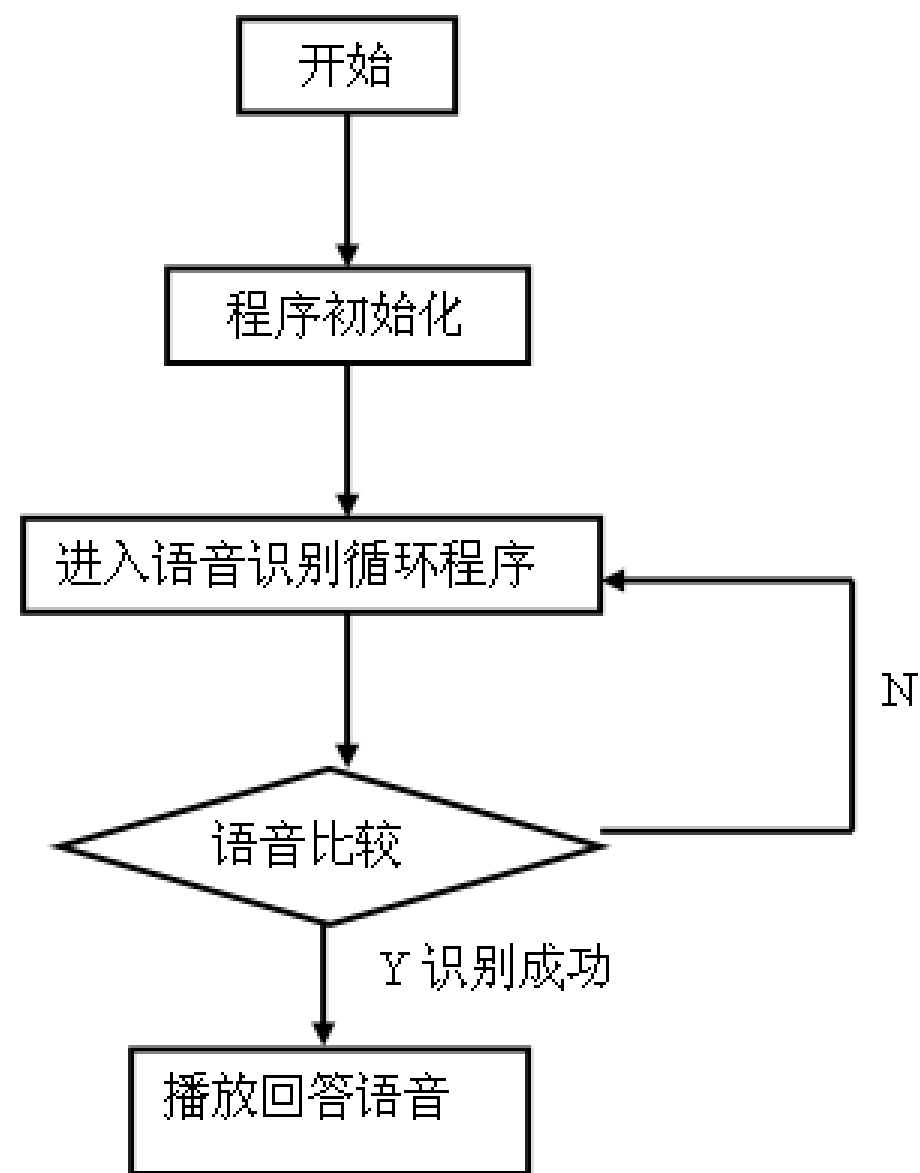


图 3-1 系统程序流程图

3.2 系统各模块程序设计

3.2.1 主程序

主程序如下所示：

```
void main()
{
    uint8 idata nAsrRes;
    FlashLED(3);
    MCU_init();
    LD_reset();
```

```

nAsrStatus = LD_ASR_NONE;          //          ASR
PlaySound(CODE_DEFAULT); // 播放提示音
while(1)
{
    if (bMp3Play!=0)                // 如果还在播放 MP3中, 则等待, 直到 MP3
播放完毕 即 bMp3Play==0
        continue;                  // bMp3Play 是定义的一个全局变量用来
记录 MP3播放的状态, 不是LD3320芯片内部的寄存器
    switch(nAsrStatus)
    {
        case LD_ASR_RUNING:
        case LD_ASR_ERROR:
            break;
        case LD_ASR_NONE:
        {
            FlashLED(4);
            nAsrStatus=LD_ASR_RUNING;
            if (RunASR()==0) // 启动一次 ASR识别流程: ASR初始化,
ASR添加关键词, 启动 ASR运算
            {
                nAsrStatus = LD_ASR_ERROR;
                LED1=0;
                LED2=0;
            }
            break;
        }
        case LD_ASR_FOUNDOK:
        {

```

```

FlashLED(2);

nAsrRes = LD_GetResult();          //      ASR识别流程结束，去
取 ASR识别结果

PlaySound(nAsrRes);
nAsrStatus = LD_ASR_NONE;
break;
}

case LD_ASR_FOUNDZERO:
default:
{

FlashLED(1);

PlaySound(CODE_DEFAULT);
nAsrStatus = LD_ASR_NONE;
break;

}

} // switch

} // while
}

void MCU_init()
{

P1M0 |= 1;
P3M0 |= 8;
LED1=0;
LED2=0;
P0 = 0xff;
P1 = 0xff;
P2 = 0xff;
P3 = 0xf7;

```

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/718134142120006126>