

# 智能语音控制系统的设计与实现

王磊, 何勇, 张宇

(贵州大学 计算机科学与技术学院, 贵阳 550000)

**摘要:** 在未来的发展中, 语音识别与自动化控制的结合也将是未来发展的一种必然趋势; 语音交互方式是更为人性化的人机交互界面, 针对语音识别与智能控制的运用, 提出了一种基于非特定人语音识别技术的智能控制设计方法, 并使用 LD3320 芯片将语音信息采集和识别, 采用 STC10L08XE 芯片作为语音控制系统的主控制器, 并结合外部存储器和控制电路实现人机语音交互和语音控制, 设计一款区别于传统控制的语音交互设备; 通过设计一款上位机软件, 可以很好地解决外部存储器数据修改困难的问题, 并同时实时进行语音数据写入和参数配置, 结合自定义的协议帧, 使数据的传输过程更加快速、稳定和可靠, 经过实验验证, 该语音控制系统工作稳定、识别率高, 具有很好的易用性和市场应用价值。

**关键词:** 语音识别; 非特定人; LD3320; 语音控制; 上位机

## Design and Realization of Intelligent Voice—Controlled System

Wang Lei, He Yong, Zhang Yu

(Computer Science and Technology College, Guizhou University, Guizhou 550000, China)

**Abstract:** In the future, it is an inevitable trend to combine the speech recognition with the intelligent control. Man-machine voice interaction provide a more user-friendly interface. In view of the speech recognition and the application of intelligent control, this paper puts forward a intelligent control design method based on speaker-independent speech recognition technology. It uses LD3320 chip to acquire and recognize speech information and STC10L08XE chip as the main controller to control the voice system, combined the external storage and the control circuitit realizes man-machine voice interaction and voice control. Design a voice interaction is different from the traditional control devices. It is easy to resolve a issue about how to modify the datas in external storage by designing an upper computer software. At the same time, it can write the voice data into the upper computer and configure parameters synchronously. Combined with the custom protocol frame, the data transmission process is more rapid, stable and reliable. According to repeated experimental verification, this speech control system works stably and owns excellent recognition rate. It is of great usability and market application value.

**Keywords:** automatic speech recognition; speaker-independent; LD3320 chip; voice control; upper computer

## 0 引言

语言是人类沟通的桥梁, 可以很轻松地达到信息的交互。如今语音识别技术已经是世界上最热门和最具发展前景的技术之一, 通过这项技术可以让语言成为人类和机器沟通的桥梁, 可以使人类的日常生活操作变得更加简单和方便。智能语音控制系统, 采用 LD3320<sup>[1]</sup> 芯片, 主要通过语音识别技术实现在家庭中对电灯、风扇、窗帘等电器的开关控制, 并且伴随识别结果有语音播放, 达到真正的智能人机交互。

## 1 系统总体设计

基于 LD3320 的智能语音控制系统框架如图 1 所示。系统主要由电源模块、主控制器、识别模块、控制模块、外部存储器、语音输入和输出等部分组成。语音输入模块将采集到的语音信息送入 LD3320 处理芯片进行识别, LD3320 将识别结果

返回给主控模块, 同时从外部存储器中取出语音数据进行语音输出播放, 此时主控模块分析识别结果后, 利用控制模块达到最终控制家电的效果。外部存储器中的数据, 可在设备运行过程中实时更改, 利用编写的上位机通过串口将数据发往主控模块, 主控模块将数据转存至外部存储器进行存放。

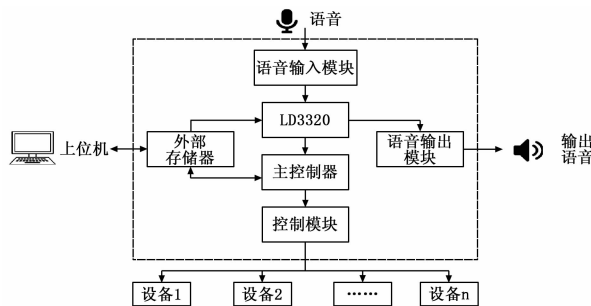


图1 智能语音控制系统框架图

收稿日期:2017-06-26; 修回日期:2017-07-24。

基金项目:黔科合 L·H 字[2014]7638; 研究生创新基金(研理工 2017078)。

作者简介:王磊(1993-),男,辽宁朝阳人,硕士研究生,主要从事物联网技术、嵌入式方向的研究。

何勇(1975-),男,贵州贵阳人,副教授,硕士研究生导师,主要从事计算机控制技术、嵌入式方向的研究。

## 2 硬件设计

智能语音控制系统中的硬件电路主要包括主控制器和语音识别两个核心模块, 语音采集成功后进入语音识别模块, LD3320 进行识别后将结果返回到主控模块, 外部存储器模块的设计主要是将返回的结果进行输出播放。由于控制模块与其

它模块供电电压不同，本文设计的电源模块就分成两路进行供电。语音输入和输出模块的设计，优化了语音采集的性能，实现了语音播放的功能。最终将控制请求命令发送给控制模块对设备进行控制操作。

### 2.1 电源模块

由于整个语音控制系统是嵌入墙体内，类似普通的插座，所以它的供电采用 HLK-PM01<sup>[2]</sup> 电源模块将 220 V 的交流电转化为 5 V 直流电，一路用来驱动 SRD-05VDC-SL-C<sup>[3]</sup> 继电器，另一路经过 ASM-1117<sup>[4]</sup> 稳压芯片将 5 V 转为 3.3 V 电压供主控模块、识别模块、和外部存储器使用。

### 2.2 主控制器

用 STC10L08XE 芯片作为智能语音控制系统的 MCU，它是宏晶科技公司生产的机器周期为 1 T 的单片机，SRAM 为 512K，片内 RAM 为 8 K；指令代码兼容传统 8051 单片机，属于增强型 8051 MCU，速度比传统的快 8~12 倍，有 EEPROM<sup>[5]</sup> 功能，2 个 16 位定时器，MCU 操作电压范围：2.4~3.6 V。

主控制器是控制设备开关的核心电路，主要处理经过 LD3320 识别后的语音数据及与外部存储器模块实现语音播放。最终主控制器分析识别结果后将控制指令发送给控制模块。

### 2.3 识别模块

语音识别模块采用 ICRout 公司设计生产的 LD3320 芯片，该芯片集成语音识别处理器、A/D 和 D/A 转换器、语音输入接口、语音输出接口等外部电路，具有快速而稳定的优化算法，完成语音识别，不需要事先训练和录音，只需要把识别的关键词语以字符串的形式传进芯片，即可以在下次识别中立即生效，识别准确率高达 95%。芯片及外围电路如图 2 所示。

寄存器控制语音输出的音量，也可以用外部的电路控制，引脚为 EQ1、EQ2、EQ3，利用图 2 中 R8 和 R5 的阻值分配可使得语音输出被放大了约 2 倍；MBS 引脚为语音输入的偏置，需要接一个相移电路，保证能输出一个浮动的电压给麦克风。

该模块采用语音识别 ASR<sup>[6]</sup> 技术，该技术是非特定人语音识别技术，不支持不同人的语音区分识别。这种识别技术的原理是通过某种数学模型，将操作者说出的语音信息，通过频谱转换为语音特征提取出来，然后与设定好的关键词条目进行一一匹配，这种匹配也只是用概率性来衡量，在 LD3320 上最多能匹配出 4 种结果，这 4 种结果分别是最优匹配结果、候选结果 1、候选结果 2、候选结果 3。通常开发只取最优结果，其他答案做参考。但是考虑实际开发和应用效果，这里将这 4 种结果加以分析，选出结果相同最多的那个结果。

简单来说，这种识别技术不关心关键词的内容，只关心语音的特征，所以不能分辨不同人的话语。最多可识别的关键词个数为 50 个，且每个关键词的长度不超过 10 个汉字或 79 个字节的标准普通话的汉语拼音（小写），每两个字（汉语拼音）之前用一个空格符相隔，如“kaideng”。

### 2.4 控制模块

控制模块采用 SRD-05VDC-SL-C 继电器来控制家电的开关，它可以控制 10 A 250 V AC、10 A 30 V DC 的负载，线圈工作电压为 5 V 直流。

采用单片机引脚控制 NPN 型三极管 S8050 的导通，从而起到控制继电器线圈两端电压的作用，并且继电器线圈端需要加一个二极管，防止产生反向感应电动势击坏电路零件，起到保护的功能。

### 2.5 外部存储器

外部存储器采用 W25Q32DW 外部 Flash<sup>[7]</sup> 存储芯片，存储空间为 4 MB，工作电压为 3.3 V，通过 SPI<sup>[8]</sup> 通信方式与单片机交互。

使用此芯片的主要目的是配合整个智能语音识别系统的语音输出功能，即播放语音文件，如 MP3<sup>[9]</sup>。同时也保存与主控制器和语音识别模块的一些配置信息，实现掉电数据不丢失。

使用上位机通过串口与主控制器交互，将需要播放的一些语音数据和配置信息逐一发送给主控制器，主控制器再将这些数据通过 SPI 的方式转存至 W25Q32DW 存储芯片，通信过程中用到自定义的协议，保证数据传输的完整性。

在系统需要播放语音文件时，由主控制器将存储器中的语音数据逐一取出，然后不断写入 LD3320 语音识别模块的寄存器中，识别模块经过自动处理，以模拟信号的方式输出至语音输出接口。

### 2.6 语音输入和输出模块

语音输入模块采用驻极体麦克风<sup>[10]</sup>，这种麦克风体积小，结构简单，成本低，电声性能好，得到很广泛的应用，在设计时加一些简单的外围电路即可。

语音输出模块采用 0.5 W 内磁式扬声器，该扬声器体积小，漏磁少，失真比较小，成本低，适合该系统应用。

将该扬声器与 LD3320 的 SPON 和 SPOP 两个引脚相连，语音识别模块在需要播放语音时就将信号发送至这两个引脚，

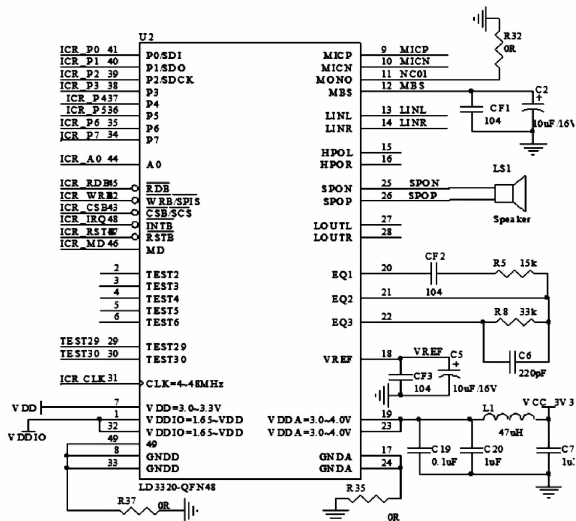


图 2 识别模块电路图

LD3320 芯片必须用外部时钟，可以接受的频率范围是 4~48MHz，操作电压为 3.0~3.3 V；可通过并行方式与主控制器连接，8 根数据线（P0-P7），4 个控制信号（ $\overline{\text{WRB}}$ 、 $\overline{\text{RDB}}$ 、 $\overline{\text{CSB}}$ ），以及一个中断返回信号（ $\overline{\text{INTB}}$ ）；可使用内部

从而经过扬声器达到最终语音播放的功能。

### 3 软件设计

具有语音识别、语音播放、智能控制、与上位机实时交互等功能的智能语音控制系统,其写数据过程由上位机通过串口将语音数据和配置信息打包至自定义协议帧中,与主控制器进行数据交互,主控制器将数据转存至外部存储器 W25Q32DW 中,在需要时,主控制器再将数据取出;其语音识别与处理输出过程由驻极体麦克风采集到的语音信息传送到 LD3320 语音识别芯片加以识别处理,返回主控制器识别结果,根据识别结果操作外部存储器 W25Q32DW,将对应的语音数据从存储器中取出,并放入 LD3320 的寄存器中,LD3320 会自动播放该语音。与此同时,主控制器操作相应引脚电平,配合控制模块电路实现对家电的控制,达到控制与语音反馈的目的。图 3 为智能语音控制系统的软件框图。

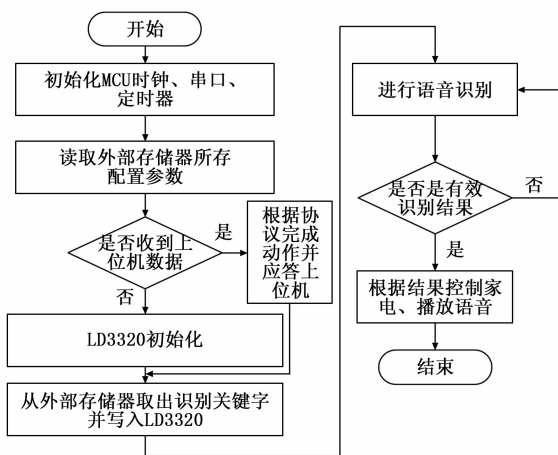


图 3 软件框图

#### 3.1 LD3320 芯片的配置

麦克风增益调整: 调整 LD3320 芯片 0x35 寄存器的数值大小。

麦克风灵敏度调整: 调整 LD3320 芯片 0xb3 寄存器的数值大小。

扬声器音量调整: 调整 LD3320 芯片 0x8e 寄存器的数值大小。

这 3 个参数的调整,直接影响 LD3320 芯片的识别准确度,所以需要根据实际情况做相应的调整。

由于这些数据需要保证在操作者修改后,不论掉电与否,都不能变,则每次主控制器与上位机交互时若修改了这些数据,立即将修改后的数据存入 W25Q32DW 中保存,在每次重启系统时,先将这些数据读出来再进行语音识别操作。

#### 3.2 语音识别

LD3320 最多只支持 50 个识别关键词,那么也就是说最多也只能支持 50 条不同语音的播放,每个语音文件在 W25Q32DW 中都有字符开始位置和结束位置,占一定长度,此时将 W25Q32DW 芯片的前 400 个字节用来存放每个语音文件的开始和结束位置,从第 401 个字节开始至第 2 900 个字节,存放最多 50 个固定字符长度为 50 个字节的关键词。从第 4097 个字节开始存放实际需要播放的语音数据。

在初始化 LD3320 完成后,需要对语音识别芯片写入识别关键词列表,此时将之前存入的关键词逐一取出,并写入 LD3320 中。

在 LD3320 工作过程中,不管是否识别成功都会产生一个中断,在播放语音文件时也会有 3 种情况使 LD3320 产生中断,第一种:声音数据已经全部播放完毕;第二种:声音数据已经全部发送至 LD3320 完毕;第三种:声音数据暂时要用完,需要向 LD3320 放入新的数据。而 LD3320 产生的这些中断,都只由一个引脚输出,则 MCU 需要分析和处理这 7 种中断情况。

如果识别正确,从 LD3320 可以得到 1 至 4 种结果,在此比较得到的几种结果,选取唯一的、二者较靠前的、多者相同最多的,作为最终 MCU 选择的结果,这个最终结果,是之前存放识别关键词的顺序序号,例如,第一个存放的关键词是“kaideng”,第一个存放的语音文件对应是播放“灯已打开”的人声,则有可能的最终识别结果是 1,代表识别到“kaideng”这个关键词。

#### 3.3 语音播放与控制家电

若最终识别结果是第一个关键词,则从 W25Q32DW 中取出第一个语音文件在存储器中的起始位置和结束位置,采用 W25Q32DW 的快速读取方式,逐一将这个范围内的数据取出并写入 LD3320,LD3320 将数据输出至扬声器。

提前设置好 MCU 与对应的控制模块的控制引脚,则根据最终识别结果的号码,控制相应引脚的电平高低,再配合控制模块达到控制家电的效果。

#### 3.4 与上位机交互协议

与上位机交互协议帧格式如表 1 所示。

表 1 协议帧格式

Header	Function	DataLength	Data	CRC	Footer
1 Byte	1 Byte	2 Byte	可变	2 Byte	1 Byte

帧头: 0xe8 (固定), 1 个字节;

功能码: 0x01—0x05, 1 个字节;

数据长度域: 由数据域字节数决定, 2 个字节;

数据域: 存放要发送的数据, 字节数不定;

CRC 校验域: 存放帧头至数据域最后一个字节的 16 位 CRC 循环冗余校验码, 2 个字节;

帧尾: 0xe6 (固定), 1 个字节。

上位机与主控制器交互总共分 5 种功能:

1) 0x01: 发送当前语音文件在 W25Q32DW 中的起始、结束位置和识别关键词。起始位置和结束位置分别占 4 个字节, 关键词所占字节根据操作者输入的关键词个数来定。

2) 0x02: 发送当前语音文件的数据, 最多一次放入 64 个字节。

3) 0x03: 发送配置信息(麦克风增益、灵敏度以及扬声器音量)。这 3 种配置信息各占 1 个字节。

4) 0x04: 发送写入初始化信息和语音文件个数, 清除 W25Q32DW 的前 2 个扇区。语音文件个数占 1 个字节。

5) 0x05: 发送写入完成信息, 告知主控制器写入完成。

主控制器根据上位机发来的帧不同, 做出不同相应, 并应

答上位机。在交互过程中，为了保护程序的鲁棒性和数据传输的完整性，是不允许系统进行语音识别的，在交互完毕后，主控制器收到写入完成的命令帧时，才开启语音识别模式。

### 3.5 上位机软件

上位机软件将操作者需要设置的数据信息经过协议帧打包并通过串口发送至主控制器，主控制器根据操作情况和过程，实时响应上位机，上位机对所接收的数据做分析、处理、显示、以及将写入过程实时反映在界面上。软件的人机交互界面友好，具有智能语音控制系统的所需功能，有很好的易用性。图 4 为上位机软件的主界面。



图 4 上位机软件界面

## 4 实验结果与分析

直接触发模式（一直处于识别状态）不适合在喧闹的环境运用，包括周围有几个人在对话。口令触发模式（第一步操作者说出口令，待正确识别口令后，第二步操作者说出操作指令，识别成功后直接执行具体操作），整体实现效果比直接触发模式好很多，误识别率很低，稳定性好。

最终麦克风增益、麦克风灵敏度调整为 21 和 18，识别距离（上接第 108 页）

步加快抑制其它元素的干扰，结合实验结果的真实性，能够实现调节量小、速度快、鲁棒性好的控制品质。

### 4 结束语

采用模糊调节的方法与自校正准则对控制器结构进行分析，并根据实际情况设置了仿真实验。实验条件设置都与现实生活息息相关。传统的自校正模糊控制虽然也建立了模糊调节机制，但是与基于动态机理的模糊规则自校正控制相比，还是存在较大误差，不能够满足实时性的要求，对于社会不断发展的今天，传统的自校正模糊控制系统已经无法抵抗外界的一些干扰因子，整个控制系统不能有机结合在一起，其设计成本与运行性能成反比关系。

通过模糊规则及动态机理结合，提出的自校正控制方法更能接近人类思维方式，有效寻找合理配置方案，使操作过程更加贴近生活，也便于将理论知识与实践模式相互结合，便于操作人员工作，为下一步发展提供支持，也为一些与自校正控制相关的企业降低能源消耗，帮助其寻找到新途径。

#### 参考文献:

[1] 孟 磊, 胡亚洲, 陈 晓. 基于自适应模糊神经网络控制器的网

离在 0~5 m, 语音正确识别率很高。

## 5 结束语

设计并实现了智能语音控制系统，功能较完善，可实现语音识别，人机语音交互，智能控制家电开关。上位机软件的设计，很好的解决了外部存储器数据修改困难的问题，并配合主控制器，使数据传输更加快捷、稳定和可靠，方便了操作者调试、修改和使用，提高了操作者的体验感，从长期测试和使用情况上看，该系统运行稳定，适合应用在智能家居<sup>[1]</sup>行业中，并有很广泛的应用价值和前景。

#### 参考文献:

[1] 洪家平. LD3320 的嵌入式语音识别系统的应用 [J]. 单片机与嵌入式系统应用, 2012, 2: 48-49.  
 [2] 马晓莉. 红外热感应智能节能插座的设计与制作 [J]. 价值工程, 2015, 28: 156-157  
 [3] 张 珺, 张俊钰. 基于 51 单片机的压力蒸汽灭菌器控制器的研制 [J]. 医疗卫生装备, 2013, 34 (5): 3-4.  
 [4] 陆二磊, 杜永文. 基于 LPC1114 的迷宫机器人的设计与实现 [J]. 广西民族大学学报 (自然科学版), 2012, 18 (4): 53-54.  
 [5] 杨 峰, 胡 彬, 罗青青. STC 单片机 AD 转换及 EEPROM 使用 [J]. 数字技术与应用, 2011, 9: 169-170.  
 [6] 刘耀文, 宁 媛. 基于 LD3320 芯片的声电密码门设计 [J]. 现代机械, 2013, 4: 55-56.  
 [7] 郑文静, 李明强, 舒继武. Flash 存储技术 [J]. 半导体技术, 2010, 47 (4): 717-718.  
 [8] 田 茂, 鲜于李可, 潘永才. SPI 模式下 SD 卡驱动的设计与实现 [J]. 现代电子技术, 2009, 14: 195-196.  
 [9] 汪永好, 周延森. 基于嵌入式 Linux 的 MP3 播放器的设计与实现 [J]. 计算机工程与设计, 2009, 30 (17): 3948-3949.  
 [10] 刘 岩, 华斯亮, 王东辉, 侯朝焕. 一种用于驻极体麦克风的 CT<sub>SC</sub>Σ-Δ 调制器 [J]. 微电子学, 2009, 39 (2): 150-151.  
 [11] 刘芮辰, 李树江, 刘 帅. 基于 IIC 和云服务器的智能家居控制系统设计 [J]. 计算机测量与控制, 2017, 24 (3): 98-100.  
 [1] 络控制系统 [J]. 计算机工程与应用, 2015, 51 (18): 90-93.  
 [2] 杨宏斌. 基于最优模糊控制的悬臂梁振动控制研究 [J]. 电子设计工程, 2016, 24 (18): 104-105.  
 [3] 刘章文, 李正东, 周志强, 等. 基于模糊控制的自适应光学校正技术 [J]. 物理学报, 2016, 65 (1): 131-138.  
 [4] 蔡 滨, 俞 岭, 彭 鑫, 等. 基于模糊控制理论的 Web 服务器多参数自优化 [J]. 小型微型计算机系统, 2015, 36 (10): 2197-2202.  
 [5] 刘 磊, 王渝红, 李兴源, 等. 基于模糊控制的变斜率 VDCOL 控制器设计 [J]. 电网技术, 2015, 39 (7): 1814-1818.  
 [6] 刘文胜, 白 璐, 马运柱, 等. 基于反函数自校正法的飞机刹车伺服系统研究 [J]. 计算机仿真, 2016, 33 (10): 91-95.  
 [7] 雷 杨, 李 军, 李绍军, 等. 高精度稳定平台伺服控制系统仿真实现 [J]. 电光与控制, 2015 (5): 72-76.  
 [8] 张彦会, 孟祥虎, 肖 婷, 等. 模糊 PID 自调整控制的锂电池均衡研究 [J]. 电子技术应用, 2015, 41 (10): 123-125.  
 [9] 刘章文, 周志强, 李正东. 基于模糊控制的波前校正技术 [J]. 激光与光电子学进展, 2017 (3): 62-69.  
 [10] 陈文艺, 李龙龙, 刘禹韬, 等. 基于 SDRAM 缓存的实时视频图像几何校正系统 [J]. 电视技术, 2016, 40 (5): 105-109.