

# LD3320 的嵌入式语音识别系统的应用

洪家平

(湖北师范学院 计算机科学与技术学院, 黄石 435002)

**摘要:** 语音交互系统是比较人性化的人机操作界面,它需要语音识别系统的支持。LD3320 就是这样一款语音识别芯片。介绍了该芯片的工作原理及应用,给出了 LD3320 与微处理器的硬件接口电路及软件程序。随着高档 MCU 的不断出现,以 MCU 为核心的嵌入式语音交互系统会有非常好的应用前景。

**关键词:** 嵌入式芯片;语音识别;语音交互系统

**中图分类号:** TP393.098

**文献标识码:** A

## Application of Embedded Speech Recognition System Based on LD3320

Hong Jiaping

(College of Computer Science and Technology, Hubei Normal University, Huangshi 435002, China)

**Abstract:** Voice interaction system is the Human-Machine Interface, and it requires the involvement of the voice recognition system. LD3320 is one such voice recognition chip. Its principle and typical application are introduced in this article, and hardware and program flowchart of the interface of LD3320 and microcontroller are provided. The Voice interaction system with the MCU will have the good application prospect with the development of the top grade MCU.

**Key words:** embedded chip; speech recognition; voice interaction system

### 1 概述

在现代社会,“懒人科技”大行其道。当面临众多繁琐的按键操作和菜单选择的时候,简单地说出指令,是最具有人情味的人机操作界面。让身边的各种电子设备可“听从”人类的语音,是从电影“星球大战”就开始的科技发展目标。虽然目前的科技还不能做到让计算机完全理解人类的所有自然语音,但是可以在一定程度上实现这个梦想。

### 2 特定人语音识别技术及原理

特定人语音识别(ASR, Auto Speech Recognition)技术是基于“关键词语列表”的识别技术,它是对大量的语音数据(相当于对数千人采集的数万小时的有效声音数据)经语言学家语音模型分析,建立数学模型,并经过反复训练提取基元语音的细节特征,以及提取各基元间的特征差异,得到在统计概率最优化意义上的各个基元语音特征,最后才由资深工程师将算法以及语音模型转换成硬件芯片并应用在嵌入式系统中。

ASR 技术每次识别的过程就是把用户说出的语音内容,通过频谱转换为语音特征,再将这个转换后的语音特

征和“关键词语列表”中的条目一一进行匹配,最优匹配的一条即作为识别结果。比如 ASR 技术在语音控制的手机应用中,这个“关键词语列表”的内容就是电话本中的人名、手机的菜单命令或手机存储卡中的歌曲名字。不论这个列表的条目内容是什么,只需要用户设置相关的寄存器,就可以把相应的待识别条目内容以字符形式传递给识别引擎。

由此可见,语音识别芯片完成的工作就是:把 MIC(麦克风)输入的声音进行频谱分析后提取语音特征,再和关键词语列表中的关键词语进行对比匹配,最后找出得分最高的关键词语作为识别结果输出。

通常基于 ASR 技术的语音识别芯片能在两种情况下给出识别结果:

① 外部送入预定时间的语音数据(比如 5 s 的语音数据),芯片对这些语音数据运算分析后,给出识别结果。

② 外部送入语音数据流,语音识别芯片通过端点检测(VAD, Voice Activity Detection)技术检测出用户停止说话,把用户开始说话到停止说话之间的语音数据进行运算分析后,给出识别结果。

对于第一种情况,可以理解为设定了一个定时录音(如 5 s 的语音数据),芯片在 5 s 后会停止把声音送入识

别引擎,并且根据已送入引擎的语音数据计算出识别结果。

对于第二种情况,需要了解 VAD 的工作原理:VAD 技术是在一段语音数据流中,判断出哪个时间点是人声音的开始,哪个时间点是人声音的结束。判断的依据是,在背景声音的基础上有了语音发音,则视为声音的开始。而后,检测到一段持续时间的背景音(比如 600 ms),则视为人声说话结束。通过 VAD 判断出人声说话的区域后,语音识别芯片会把这期间的声音数据进行识别处理,计算出识别结果。

除了以上两种情况外,语音识别算法无法“主动”地判断出是否识别出了一个结果。这是因为,在计算过程中的任何时刻,语音识别器都会对已送入识别芯片的声音数据进行分析,并根据匹配程度为识别列表中的关键词语进行打分,最匹配的打分最高。但是,由于识别算法不知道用户后面是否还继续说话,所以无法主动地判断已经识别出的结果。

### 3 语音识别芯片 LD3320 的工作原理

#### 3.1 语音识别系统原理结构

LD3320 语音识别芯片采用的就是 ASR 技术,图 1 就是由 LD3320 和单片机(或嵌入式系统)组成的语音识别系统原理框图。图中给出了 LD3320 的内部原理结构,本文中选用的 MCU 是 STC10L08XE 单片机。

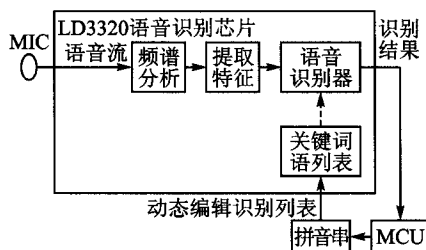


图 1 语音识别系统原理框图

语音识别芯片 LD3320 是 ICRout 公司的产品,它采用 ASR 技术,提供了一种脱离按键、键盘、鼠标、触摸屏等 GUI 操作方式且基于语音的用户界面 VUI(Voice User Interface),使得用户对该系统的操作更简单、快速和自然。

用户只需要把识别的关键词语以字符串的形式传送给芯片,即可以在下次识别中立即生效。比如,用户在 51 等主控 MCU 的编程中,简单地通过设置芯片的寄存器,把诸如“你好”这样的识别关键词语的内容动态地传入芯片中,芯片就可以识别所设定的关键词语了。每个关键词语可以是单字、词组、短句或者任何的中文发音的组合。基于 LD3320 的语音识别系统可以随着使用流程,在运行

时动态地更改关键词语列表的内容,这样可以用一个系统支持多种不同的场景,同时也不需要用户作任何的录音训练。

#### 3.2 LD3320 的用户使用模式

LD3320 有两种用户使用模式,即“触发识别模式”和“循环识别模式”。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:系统的主控 MCU 在接收到外界一个触发后(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 s),要求用户在这个定时过程中说出要识别的语音关键词语。这个过程结束后,需要用户再次触发才能再次启动一个识别过程。

循环识别模式:系统的主控 MCU 反复启动识别过程。如果没有人说话就没有识别结果,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。

### 4 语音识别系统软硬件设计

#### 4.1 硬件系统设计

由图 1 可知,由 LD3320 组成的语音识别系统硬件有单片机(或嵌入式系统)及 LD3320。图 2 和图 3 分别是由单片机 STC10L08XE 构成的主控芯片和由 LD3320A 构成的语音识别主系统。

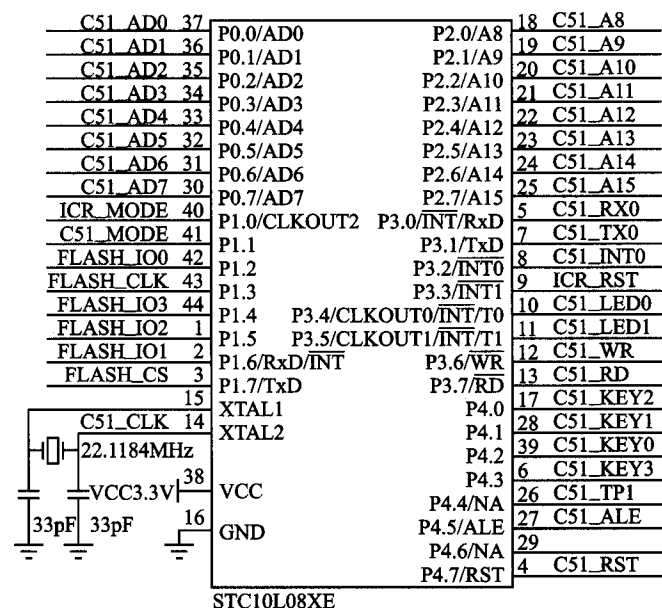


图 2 STC10L08XE 构成的主控芯片

#### 4.2 软件系统设计

语音识别的操作顺序是:先进行语音识别的初始

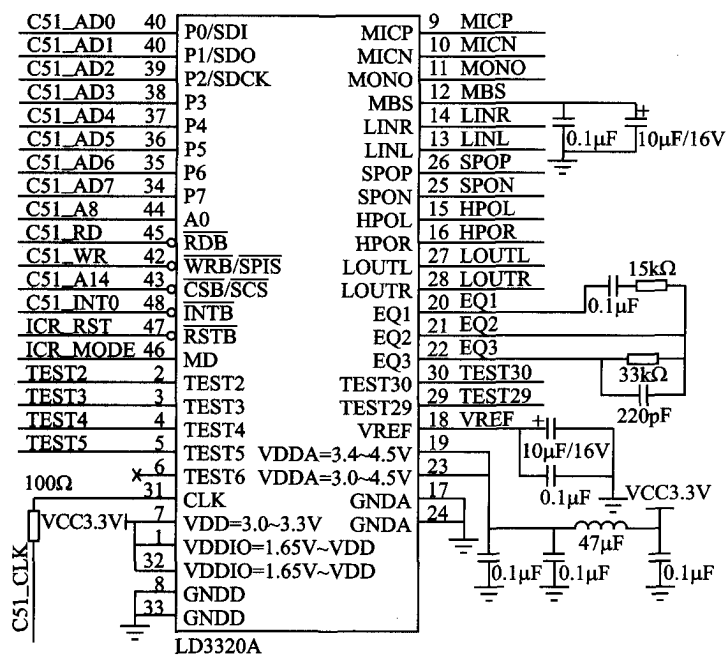


图3 LD3320A 构成的语音识别主系统

化,然后写入识别列表,系统即开始进行语音识别,并准备好中断响应函数,打开中断允许位。这里如果不用中断方式,也可以通过查询方式工作。在“开始识别”后,读取寄存器 B2H 的值,如果为 21H 就表示有识别结果产生。

下面是语音识别的初始化程序段,按照以下序列来设置寄存器:

```
void LD_Init_ASIR() {
    nLD_Mode=LD_MODE_ASIR_RUN;
    LD_Init_Common();
    LD_WriteReg(0xBD,0x00);
    LD_WriteReg(0x17,0x48);
    delay( 10 );
    LD_WriteReg(0x3C,0x80);
    LD_WriteReg(0x3E,0x07);
    LD_WriteReg(0x38,0xff);
    LD_WriteReg(0x3A,0x07);
    LD_WriteReg(0x40,0x00);
    LD_WriteReg(0x42,0x08);
    LD_WriteReg(0x44,0x00);
    LD_WriteReg(0x46,0x08);
    delay( 1 );
}
```

初始化后是写入识别列表。识别列表的规则是:每个识别条目对应一个特定的编号(1 个字节),不同识别条目的编号可以相同,而且不用连续。LD3320 芯片最多支持 50 个识别条目,每个识别条目是标准普通话的汉语拼音

(小写),每 2 个字(汉语拼音)之间用 1 个空格间隔。表 1 是一个简单的例子。

表 1 识别列表

编号	字符串
1	bei jing
1	shou du
3	shang hai
7	tian jin
8	chong qing

图 4 是由 LD3320 组成的语音识别系统主程序流程图,图 5 是语音识别系统中断服务程序流程图。

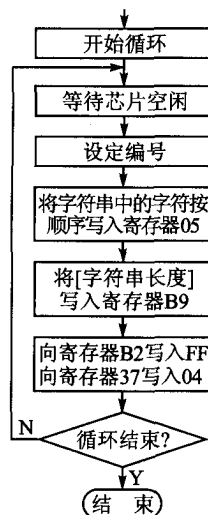


图4 语音识别系统主程序流程图

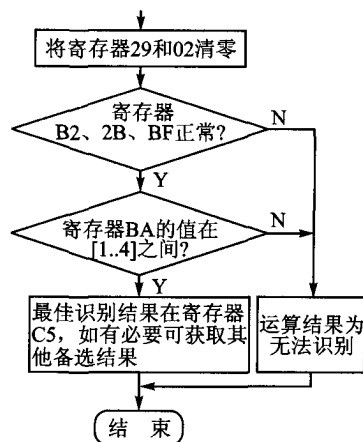


图5 语音识别系统中断服务程序流程图

## 5 声控电视遥控器

声控电视遥控器最有用的地方是,在更换频道时可以直接说出频道名称,而不是去记忆频道名称和频道数字的联系。同时,在使用电视遥控器时,遥控器是用电池工作,不能让识别芯片一直处于工作状态。因此在设计时,可以在遥控器上设置一个大一点的按键,用户在使用时,按一下这个按键,启动 LD3320 语音识别芯片,此时可以播放一声“嘀”的提示音,然后在限定的时间内(如 5 s),接收用户的语音命令,并给出识别结果。比如用户说“体育台”,识别芯片把识别结果提供给遥控器的主控 MCU。随后遥控器的 MCU 就根据事先设定好的对应关系,发出对应频道的红外编码,实现换台。可以不加确认过程直接换台。图 6 为声控电视遥控器工作流程。

## 结 语

这种语音识别系统也容易引起误识别,如当用户

```

reset 彩信的各参数,以免发生彩信尺寸已达最大等问题
OK
AT $ MMSW=1,1 //设置目的号码或邮件地址
13970975217
OK
AT $ MMSW=4,1 //彩信主题
Test0
OK
AT $ MMSW=5,1,1 //设置彩信内容
My picture
OK
AT $ MMSW=6,1,"mms.jpg" //设置要发送的附件
OK
AT $ MMSEND
OK
$ MMSTRACE: 1 //返回 1,表明发送成功

```

```

AT $ FDEL="*" //删除存储器里的文件
OK

```

## 结 语

本文根据报警系统的即时性需要,设计了一款性价比高的控制系统。该系统经过测试和应用,稳定性好,系统的远程报警功能及时,且彩信图像清晰。

## 参考文献

- [1] 江海波,王卓然,耿得根.深入浅出 AVR 单片机[M].北京:中国电力出版社,2008.
- [2] 铃木雅臣.晶体管电路设计(上)[M].周南生,译.北京:科学出版社,2004.
- [3] 戴卫平,胡耀辉,朱朝华,等.单片机系统开发实例经典[M].北京:冶金工业出版社,2006.

(责任编辑:高珍 收稿日期:2011-09-30)

49 说的内容不在识别列表内时,必然会引起误识别。为

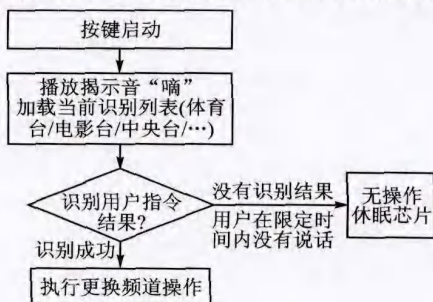


图6 声控电视遥控器工作流程

了克服这些缺点,降低误识别率,可在设定好要识别的关键词语后,再添加一些与识别列表内的单词有联系的任意其他词汇,用来吸收错误识别,从而达到降低误识别率的目的。

由 LD3320 组成的语音识别系统有很广泛的应用,如语音控制的点歌系统、语音控制的手机、音控智能导航仪、音控智能家电产品等。

## 参考文献

- [1] LD3320 数据手册[EB/OL]. (2010-06-08)[2011-09]. <http://www.icroute.com/index.html>.
- [2] 肖来胜,冯建兰,夏术泉.单片机技术实用教程[M].武汉:华中科技大学出版社,2004.
- [3] 宗光华,李大寨.多单片机系统应用技术[M].北京:国防工业出版社,2003.

洪家平(副教授),主要研究方向是嵌入式系统应用、计算机控制技术。

(责任编辑:高珍 收稿日期:2011-09-23)

## Xilinx 交付首批 Zynq-7000 可扩展处理平台

赛灵思公司(Xilinx, Inc.)已向客户交付首批 Zynq-7000 可扩展处理平台(EPP),这是其完整嵌入式处理平台发展战略的一个重大里程碑,率先为开发人员提供堪比 ASIC 的性能与功耗、FPGA 的灵活性以及微处理器的可编程性。那些先期已经采用 Zynq-7000 EPP 仿真平台、赛灵思早期试用硬件工具以及 ARM Connected Community 社区支持的标准软件工具进行系统开发的客户,现在就可以将他们的应用移植到这些器件上,并开始下一阶段的产品开发工作。

针对那些需要支持高性能及实时运算应用的系统而言,Zynq-7000 EPP 提供了传统处理解决方案所无法实现的性能水平。仿真平台、硬件开发工具、开源 Linux 支持,以及与 Cadence 设计系统公司联合开发的可扩展虚拟平台,均有助于推进 Zynq-7000 EPP 系统的开发与实现。随着可支持的操作系统越来越多,嵌入式工具和软件开发解决方案生态系统也将不断扩展。

Zynq-7000 系列将 ARM 双核 Cortex-A9 MPCore 处理系统与赛灵思可扩展的 28 nm 可编程逻辑架构完美整合在一起,可支持双核 Cortex-A9 处理器系统以及可编程逻辑中定制加速器和外设的并行开发。软件开发人员可充分利用 Eclipse 环境、Xilinx Platform Studio 软件开发套件(SDK)、ARM Development Studio 5(DS-5)和 ARM RealView Development Suite(RVDS)或编译器、调试器,以及 ARM Connected Community 社区和赛灵思联盟计划生态系统(Xilinx Alliance Program)的优秀厂商(诸如 Enea Services Linköping AB, Express Logic, Lauterbach Datentechnik GmbH, MathWorks, PetaLogix, Mentor Graphics, Micrium 和 Wind River Systems 等)提供的应用。