

基于 STM32 的嵌入式语音识别模块设计^{※*}

苏鹏, 周凤余, 陈磊

(山东大学 控制科学与工程学院, 济南 250061)

摘要: 介绍了一种以 ARM 为核心的嵌入式语音识别模块的设计与实现。模块的核心处理单元选用 ST 公司的基于 ARM Cortex-M3 内核的 32 位处理器 STM32F103C8T6。本模块以对话管理单元为中心, 通过以 LD3320 芯片为核心的硬件单元实现语音识别功能, 采用嵌入式操作系统 $\mu\text{C}/\text{OS-II}$ 来实现统一的任务调度和外围设备管理。经过大量的实验数据验证, 本文设计的语音识别模块具有高实时性、高识别率、高稳定性的优点。

关键词: ARM; 语音识别; 对话管理; LD3320; $\mu\text{C}/\text{OS-II}$

中图分类号: TB24

文献标识码: A

Embedded Speech Recognition Module Based on STM32^{※*}

Su Peng, Zhou Fengyu, Chen Lei

(School of Control Science and Engineering, Shandong University, Ji'nan 250061, China)

Abstract: This paper provides the design of an embedded speech recognition module based on ARM. The system core is STM32F103C8T6, a 32-bit processor based on ARM Cortex-M3 core produced by STMicroelectronics. The center of this module is the dialogue management unit. The module contains a hardware unit based on LD3320 which is used for speech recognition. The real time operating system $\mu\text{C}/\text{OS-II}$ is used for multitask scheduling and the operation of peripheral devices. A number of validations based on experimental data show this module has advantages of real-time, high recognition rate and high stability.

Key words: ARM; speech recognition; dialogue management; LD3320; $\mu\text{C}/\text{OS-II}$

引言

服务机器人以服务为目的, 因此人们需要一种更方便、更自然、更加人性化的方式与机器人交互, 而不再满足于复杂的键盘和按钮操作。基于听觉的人机交互是该领域的一个重要发展方向^[1]。目前主流的语音识别技术是基于统计模式。然而, 由于统计模型训练算法复杂, 运算量大, 一般由工控机、PC 机或笔记本来完成, 这无疑限制了它的运用。嵌入式语音交互已成为目前研究的热门课题^[2-3]。

嵌入式语音识别系统和 PC 机的语音识别系统相比, 虽然其运算速度和内存容量有一定限制, 但它具有体积小、功耗低、可靠性高、投入小、安装灵活等优点, 特别适用于智能家居、机器人及消费电子等领域。

1 模块整体方案及架构

语音识别的基本原理^[4]如图 1 所示。语音识别包括

两个阶段: 训练和识别。不管是训练还是识别, 都必须对输入语音预处理和特征提取。训练阶段所做的具体工作是通过用户输入若干次训练语音, 经过预处理和特征提取后得到特征矢量参数, 最后通过特征建模达到建立训练语音的参考模型库的目的。而识别阶段所做的主要工作是将输入语音的特征矢量参数和参考模型库中的参考模型进行相似性度量比较, 然后把相似性最高的输入特征矢量作为识别结果输出。这样, 最终就达到了语音识别的目的。

现有的语音识别技术按照识别对象可以分为特定人识别和非特定人识别。特定人识别是指识别对象为专门的人, 非特定人识别是指识别对象是针对大多数用户, 一般需要采集多个人的语音进行录音和训练, 经过学习, 达到较高的识别率。

基于现有技术开发嵌入式语音交互系统, 目前主要有

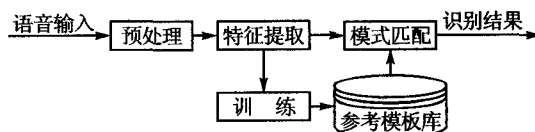


图 1 语音识别的基本原理

* 基金项目: 国家 863 计划资助项目(2009AA04Z220), 中国博士后科学基金资助项目(20100471522)。

两种方式:一种是直接在嵌入式处理器中调用语音开发包;另一种是嵌入式处理器外围扩展语音芯片^[5]。第一种方法程序量大,计算复杂,需要占用大量的处理器资源,开发周期长;第二种方法相对简单,只需要关注语音芯片的接口部分与微处理器相连,结构简单,搭建方便,微处理器的计算负担大大降低,增强了可靠性,缩短了开发周期。

语音识别技术在国内外的发展十分迅速。目前国内在PC应用领域,具有代表性的有:科大讯飞的 InterReco2.0、中科模式识别的 Pattek ASR3.0、捷通华声的 jASRv5.5;在嵌入式应用领域,具有代表性的有:凌阳的 SPCE061A、ICRoute 的 LD3320X、上海华镇电子的 WS-117^[5]。

本文的语音识别方案是以嵌入式微处理器为核心,外围加非特定人语音识别芯片及相关电路构成。语音识别芯片选用 ICRoute 公司的 LD3320 芯片。

2 硬件电路设计

如图2所示,硬件电路主要包括主控核心部分和语音识别部分。语音进入语音识别部分后,将处理过的数据并行传输到主控制器,主控制器经过处理后,发送命令数据到 USART,USART 可用于扩展外围串行设备,如语音合成模块等。

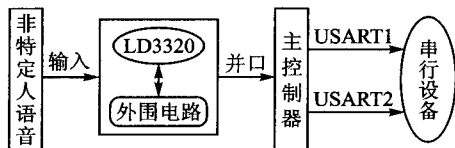


图2 硬件电路整体架构

2.1 主控制器电路

本文的主控制器选用的是 ST 公司的 STM32F103C8T6 芯片。该芯片基于 ARM Cortex-M3 32 位的 RISC 内核,工作频率最高可达 72 MHz,内置高速存储器(64 KB 的闪存和 20 KB 的 SRAM),丰富的增强 I/O 端口和联接到两条 APB 总线的外设。STM32 系列提供了全新的 32 位产品选项,结合了高性能、实时、低功耗、低电压等特性,同时保持了高集成度和易于开发的优势,将 32 位 MCU 世界的性能和功效引向一个新的级别^[6]。

编者注:STM32F103C8T6 主控制器的设计原理图^[7]见本刊网站 www.mesnet.com.cn。

2.2 语音识别电路

图3为语音识别部分原理图,参照了 ICRoute 发布的 LD3320 数据手册^[9]进行设计。LD3320 的内部集成了快速稳定的优化算法,不需外接 Flash、RAM,不需要用户事先训练和录音而完成非特定人语音识别,识别准确率高。

图中,LD3320 采用并行方式直接与 STM32F103C8T6 相接,均采用 1 k Ω 电阻上拉,A0 用于判断是数据段还是地址段;控制信号 \overline{RDB} 、 \overline{WRB} 、 \overline{CSB} ,复位信号 \overline{RSTB} 以及中断返回信号 INTB 与 STM32F103C8T6 直接相连,采用 10 k Ω 电阻上拉,辅助系统稳定工作;和 STM32F103C8T6 采用同一个外部 8 MHz 时钟;发光二极管 D1、D2 用于复位后的上电指示;MBS(引脚 12)作为麦克风偏置,接了一个 RC 电路,保证能输出一个浮动电压给麦克风。

3 软件系统设计

软件系统的设计主要包括 3 部分:主控单元的嵌入式操作系统 $\mu C/OS-II$ 移植、LD3320 的语音识别程序设计、对话管理单元的设计。

3.1 嵌入式操作系统 $\mu C/OS-II$ 移植

$\mu C/OS-II$ 是一个源码公开、可移植、可固化、可裁剪、占先式的实时多任务操作系统^[10]。它是专门为计算机的嵌入式应用设计的,绝大部分代码采用 C 语言编写,具有执行效率高、占用空间小、实时性能优良和可扩展性强等特点,最小内核可至 2 KB。在 $\mu C/OS-II$ 中,任务的概念尤为重要,它是可剥夺型的内核,因此任务优先级的划分至关重要。基于分层化和模块化的设计理念,整个系

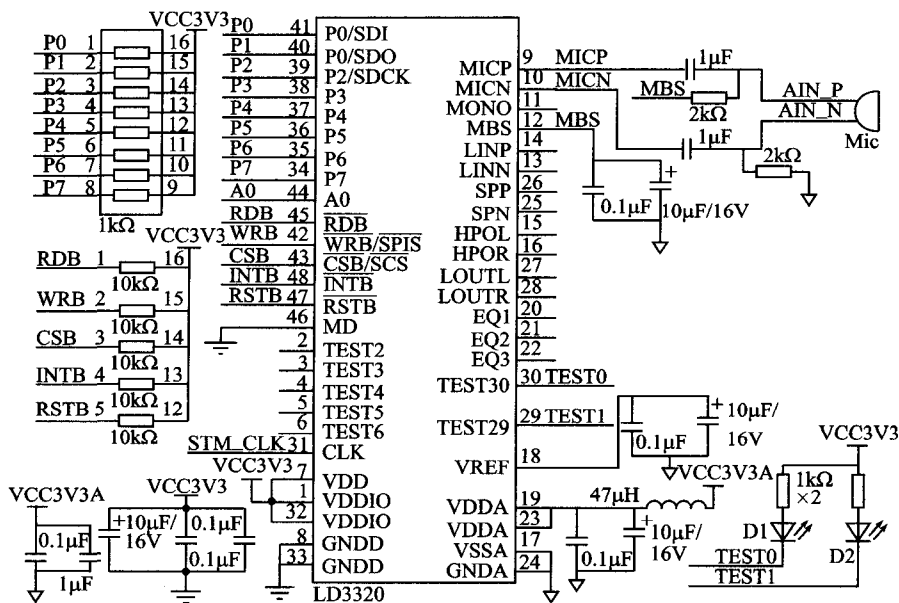


图3 LD3320 语音识别单元设计原理图

统任务的划分如表 1 所列。

表 1 主控系统任务优先级划分

任务名	任务优先级	堆栈字节 (RAM)/字	实现功能
App_TaskStart	5	64	系统初始化
App_TaskSR	6	128	语音识别
App_TaskCmd	8	128	命令解析与执行
App_TaskCom	9	64	串口命令发送
App_TaskUpdate	10	128	更新对话集
App_TaskPB	11	64	按键检测与处理
App_TaskLed	13	64	LED 指示
OSTaskStat	30	64	统计任务
OSTaskIdle	31	128	空闲任务

表 1 中除了 OSTaskStat 和 OSTaskIdle 任务为系统自带,其他 7 个任务均为用户创建。App_TaskStart 是系统的第一个任务,对系统时钟和底层设备进行初始化,创建所有事件和其他各项用户任务,并对系统状态进行监测;App_TaskSR 完成语音识别;App_TaskCmd 完成对话集中命令的解析和执行,并通过 USART1 向外发送;App_TaskCom 作为外围扩展任务,通过 USART2 向外发送指令或数据,负责控制外围扩展设备,如语音合成设备等;App_TaskUpdate 通过解析 USART1 接收的命令和数据进行对话集的更新;App_TaskPB 是按键扫描任务,负责检测 3 个独立按键,分为短按和长按检测;App_TaskLed 驱动 4 个 LED 指示灯,指示当前工作状态。

3.2 语音识别程序设计

语音识别程序的设计,参考了 LD332X 开发手册^[1],本文中采用中断方式工作,其工作流程分为通用初始化—语音识别用初始化—写入识别列表—开始识别—响应中断。

① 通用初始化和语音识别用初始化。在初始化程序里,主要完成软复位、模式设定、时钟频率设定、FIFO 设定。

② 写入识别列表。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),编号可以相同,可以不连续,但是数值要小于 256(00H~FFH)。本芯片最多支持 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)之间用一个空格间隔。本文中采取了连续不同编号的识别条目,表 2 是简单的示例。

③ 开始识别。设置几个相关的寄存器,即可开始语

表 2 识别列表示例

编号	字符串
1	ni hao
2	ren shi ni hen gao xing
3	da kai dian shi
4	ni neng zuo shen me shi qing

音的识别。图 4 是相关的流程。ADC 通道即为麦克风输入通道,ADC 增益也就是麦克风音量,可设定值 00H~7FH,建议设置值为 40H~6FH,值越大代表 MIC 音量越大,识别启动越敏感,但可能带来更多误识别;值越小代表 MIC 音量越小,需要近距离说话才能启动识别功能,好处是对远处的干扰语音没有反应。本文中设定值为 43H。

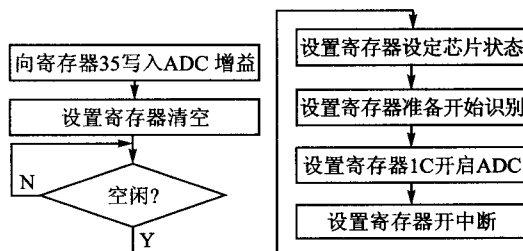


图 4 开始识别流程

④ 响应中断。如果麦克风采集到声音,不管是否识别出正常结果,都会产生一个中断信号。而中断程序要根据寄存器的值分析结果。读取 BA 寄存器的值,可以知道有几个候选答案,而 C5 寄存器里的答案是得分最高、最可能正确的答案。

3.3 对话管理单元设计

为了方便进行对话的管理,本文中设计了一个对话管理单元,用于对等待识别的语句和等待执行的命令进行存储,在主控制器中通过定义二维数组来实现。LD3320 每次识别最多可以设置 50 项候选识别句,每个识别句可以是单字、词组或短句,长度为不超过 10 个汉字或者 79 个字节的拼音串。基于上述原因,本文设计的对话管理数组如表 3 所列。

表 3 对话管理单元数组设计

名称	存储内容	数组定义
识别数组	待识别的语句	INT8U Reco[50][79]
行为数组	对应的行为编号	INT8U Act[50][6]

行为数组中存储要执行的行为编号,对应于 50 条语音识别语句,共有 50 组指令,每组指令中可以最多包含 6 个行为,并行的行为可以归为一步,通过多个行为的组合,就可以完成更复杂的任务。

4 性能测试与应用

为了保证设计的语音识别模块的语音识别率、稳定性和响应时间,本文对所描述的语音识别模块做了相应的测试,测试环境分别为安静的家庭环境和嘈杂的医院环境,共 8 条语音指令,对每条语音指令分别进行 10 次测试,每个环境下对每个特定人的总实验次数为 80 次,记录成功识别的次数。测试结果如表 4 所列。

表4 语音模块实际测试效果统计

成功识别 的次数 语句	环境	家庭(安静环境)			医院(嘈杂环境)		
		非特定 人1	非特定 人2	非特定 人3	非特定 人1	非特定 人2	非特定 人3
你好		10	9	10	10	9	10
前进		9	9	8	9	9	8
后退		9	8	9	8	8	8
左转		9	10	10	9	9	9
右转		9	9	9	9	9	9
停止		10	10	10	8	9	8
开始动作		9	9	8	7	8	6
结束动作		9	8	9	8	8	8
识别率/(%)		92.5	90	91.25	85	86.25	82.5

测试中的3个非特定人中,非特定人1为女性,非特定人2和非特定人3为男性。由表中数据可以看出,家庭环境下对非特定人的语音识别率可达90%以上,嘈杂的医院环境下的语音识别率也可达82.5%以上。识别率方面,在嘈杂环境下比在安静环境下的语音识别率有所降低;稳定性方面,在安静环境下系统的稳定性较好,语音说1遍,最多说2遍模块就可以做出正确的响应;在噪声环境下,系统的稳定性有所下降,个别语音命令需要说3遍甚至3遍以上才能被模块准确识别;实时性方面,在安静环境下的语音能保证系统响应的实时性,响应时间一般不超过1s,在噪声环境下的响应时间相对长一些。

结 语

本文讨论了基于STM32的嵌入式语音识别模块的设计和实现,对模块各个组成单元的硬件电路及软件实现进行了详细的介绍。大量实验及实际应用表明,本文设计的语音识别模块具有稳定性好、语音识别率高、抗噪声干扰能力强、结构简单和使用方便等特点。该模块实用性强,可广泛应用于服务机器人智能空间、智能家居和消费

电子产品等多个领域。

参考文献

- [1] 徐方,张希伟,杜振军.我国家庭服务机器人产业发展现状调研报告[J].机器人技术与应用,2009(2):14-19.
- [2] 孙艺红,刘媛.服务机器人语音交互解决方案[J].控制工程,2004,11(2):184-186.
- [3] 邝志刚,刘媛,崔世钢.机器人开发平台中语音交互系统的实现[J].天津职业技术学院学报,2004,14(2):1-7.
- [4] 柳春.语音识别技术研究进展[J].甘肃科技,2008,24(5):41-45.
- [5] 陈景帅.智能空间下语音交互系统的研究与实现[D].济南:山东大学,2010.
- [6] STM32F103x8/B增强型系列中容量产品数据手册[EB/OL].(2009-04)[2010-09].http://www.stmicroelectronics.com.cn/stonline/mcu/MCU_Pages.htm.
- [7] STMicroelectronics. STM32F10xxx 硬件开发使用入门[EB/OL].(2008-05)[2010-09].http://www.stmicroelectronics.com.cn/stonline/mcu/STM32_AN.htm.
- [8] DALLAS Semiconductor. DS1233A 3.3V EconoReset[EB/OL]. [2009-10-26].http://china.maxim-ic.com/quick_view2.cfm/qv_pk/2653.
- [9] ICRout. LD332X 数据手册[EB/OL]. [2010-03-10].http://www.icroute.com/web_cn/Download.html.
- [10] 任哲.嵌入式实时操作系统μC/OS-II原理及应用[M].2版.北京:北京航空航天大学出版社,2009.
- [11] ICRout. LD332X 开发手册[EB/OL]. [2010-03-10].http://www.icroute.com/web_cn/Download.html.

苏鹏、陈磊(硕士研究生),主要从事智能服务机器人的研究;周凤余(教授),研究方向为服务机器人技术、智能空间和计算机控制系统。

(收稿日期:2010-09-21)

加电磁干扰的条件下,总线只有在极少数开机时出现错误,主要原因是开机时出现的浪涌电流。

结 语

本文设计的模块化检测系统,利用TWI总线作为各个器件通信的媒介,并以此为基础构建总线式拓扑网络,简化了硬件和软件设计,缩短了系统的开发周期。在TWI总线驱动程序设计上,增加容错处理程序,使总线运行更加稳定和可靠,提高了系统的抗干扰能力。同时在从机TWI总线驱动程序设计时使用TWI中断,合理安排各个功能程序的执行时间,有效地提高了程序的运行效率。

参考文献

- [1] 肖晶晶,李天剑,吴启焱.基于TWI接口的AVR单片机通

信系统设计[J].北京信息科技大学学报,2009,24(4):66-68,86.

- [2] 马惠锋,韩震宇.AVR单片机TWI通信接口在广播机微控制器中的应用[J].中国测试技术,2005,31(2):98-100.
- [3] Atmel. ATmega128 Datasheet.
- [4] 楚天鹏,郑华耀.TWI接口多机通信的实现[J].装备制造技术,2008(6):96-98.
- [5] 郭祥军.AVR单片机的TWI总线的原理及应用[J].单片机与嵌入式系统应用,2006(8):37-39.
- [6] 江海波,王卓然,耿德根.深入浅出AVR单片机[M].北京:中国电力出版社,2008:270-280.

刘千杰(硕士研究生),主要研究领域为电力电子技术和电气测试技术。

(收稿日期:2010-09-13)