

实时语音识别系统在家庭监护机器人的实现

语音是人类最常用的交流方式，也是人类和计算机交流最渴望的方式。因此用语音同计算机交流也成为了最近研究的热点，计算机对语音的理解是计算机科学中的一个引人入胜的、富有挑战性的课题。

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 等着名公司都为语音识别系统的实用化开发研究投以巨资。IBM 公司于 1997 年开发出汉语 ViaVoice 语音识别系统，次年又开发出可以识别上海话、广东话和四川话等地方口音的语音识别系统 ViaVoice'98。目前市场上已经出现了语音识别电话、语音识别记事本等产品，如美国 VPTC 公司的 Voice Organizer 和法国的 Parrot 等。

我国语音识别研究工作开始的较晚，但近年来发展得很快，一直紧跟国际水平，国家也很重视，并把大词汇量语音识别的研究列入“863”计划，由中科院声学所、自动化所、清华大学电子工程系及北京大学等单位研究开发，取得了高水平的科研成果，如中科院自动化所研制的非特定人、连续语音听写系统和汉语语音人机对话系统，其字准确率或系统响应率可达 90% 以上。鉴于中国未来庞大的市场，国外也非常重视汉语语音识别的研究。美国、新加坡等地聚集了一批来自大陆、台湾、香港等地的学者，研究成果已达到相当高水平。

1 系统设计

文中是家庭监护机器人项目中的语音识别系统设计部分，设计目的是设计出一种可以识别语音的、协助监护家庭行动不方便人员的机器人。为实现该语音识别系统，设计了语音识别系统总体结构框图，如图 1 所示。

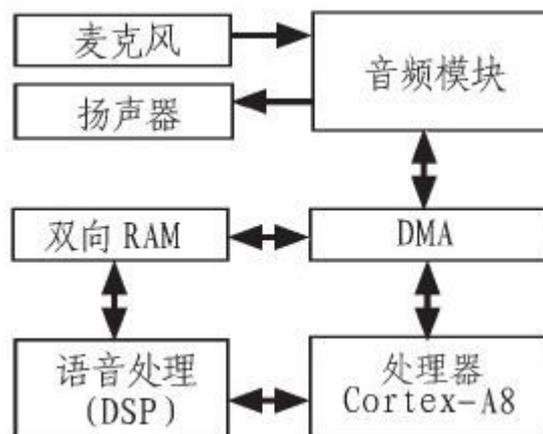


图 1 系统总体结构框图

1.1 硬件设计

文中所研究和设计的功能，都是应用在移动机器人上的。因而系统的研究设计需要考虑到体积小、省电、便于移动的特性，并需具有便于家庭用户操作的友好显示界面。对于语音识别部分，需要用到用于语音识别算法处理的处理器、语音采集电路和语音输出电路，如图

2 所示。其中语音识别算法运算的处理器主要负责算法的运算处理，相当于机器人的大脑；语音采集电路负责采集外部的声音信号，相当于机器人的耳朵；语音输出电路负责输出话语声音，相当于机器人的嘴巴。

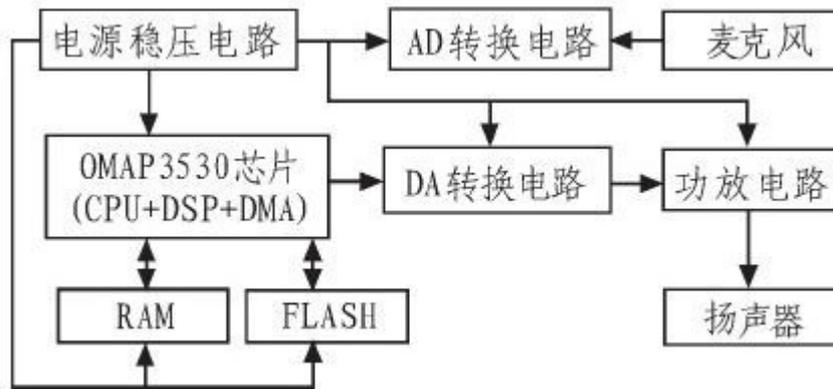


图 2 系统硬件结构图

1) 语音识别算法处理器选择

根据系统设计功能的要求，目前常用的语音识别芯片种类一般有：单片机（MCU）、DSP 和 SoC（System on Circuit）。考虑到普通单片机（MCU）资源的紧缺及运行速度较慢的缺点，因而在本系统设计将不考虑使用单片机（MCU）作为语音识别的处理器。DSP 包含用作数字信号处理的专用部件，运算能力强、精度高，但目前 DSP 的价格比较高，同时考虑到本系统的特性，需要选择一种既有较强的运算能力，适合于语音识别的功能，并且能实现较好的用户操作界面，并带有文件系统（用于识别地图）的功能，因而选择 DSP 并不是明智之举。目前 Texas Instruments 公司新推出的一款芯片 OMAP3530，它具有双内核 ARM CortexTM-A8 的内核和 TMS320C64+TM DSP 内核，属于高性能的 OMAP35x 架构系列产品，满足了系统设计的各种功能特性要求。

2) 语音编解码芯片选择

机器人选择一款合适的语音处理芯片是非常重要的。考虑到系统中用到了各种电源，并需要对电源进行管理，因此选择 TI 公司配套的 TPS 65930 芯片来作为系统语音识别部分的音频编解码处理功能的硬件平台是非常合适的。该芯片是一个集成了电源管理、ADC、嵌入式电源控制（EPC）、全功能的音频编解码器于一体的芯片，满足了系统所有电源管理和音频编解码的需要，为设计的 PCB 板节省了空间，同时减少了多电源硬件设计的负责布线烦恼。

3) 电路设计

本文的设计是用在移动机器人上的，因而需要语音的输入、识别处理及语音输出的功能。对于语音的输入采集，本文使用声音传感器麦克风及外围电路来实现。对于语音输出部分，使用功率放大器结合喇叭来使用。设计语音部分原理图如图 3 所示。

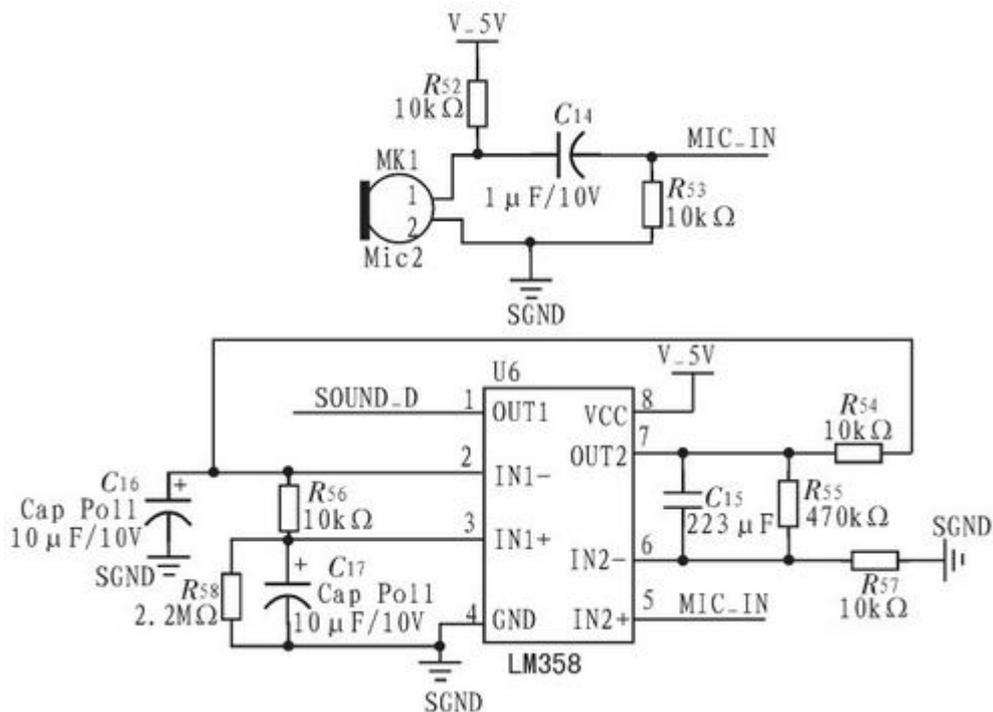


图 3 语音输入原理图

1.2 软件设计

HTK (Hidden Markov Model Toolkit) 是一套专门的建立和处理隐马可夫模型 (HMMs) 的实验工具包, 由英国剑桥大学工程系 (Cambridge University Engineering Department, CUED) 开发的, 主要应用于语音识别领域, 也可以应用于语音合成、字符识别和 DNA 排序等研究领域。HTK 经过剑桥大学、Entropic 公司及 Microsoft 公司的不断增强和改进, 使其在语音识别领域处于世界领先水平。

基于 HTK 的语言识别时, 识别结果适用只能显示在 DOS 或终端上, 而且不利于将结果保存、移植或者二次开发利用。在本语音识别系统中使用了 HTK 接口工具 ATK

(An Application Toolkit for HTK)。ATK 是由英国剑桥大学开发的开源语音识别工具, 是对 HTK 的 C++ 多线程封装, 跟 HTK 一样, 它支持 Linux 和 Windows, 它包括 HTK (HTKLib)、AHTK、AGram、ANGram、ADict、AHMMs、AResource、ARMan、ARec、ACode、ASource、ATee、AComponent、ABuffer、APacket、Asyn、FLite (SYNLib)、ALog 模块部件。

基于 ATK 的语音识别软件应用系统的由语音信号采集模块、基于 DMA 的双向高速 RAM 存取模块、ATK 语音识别模块、系统管理模块、语音输出模块等模块组成, 如图 4 所示。

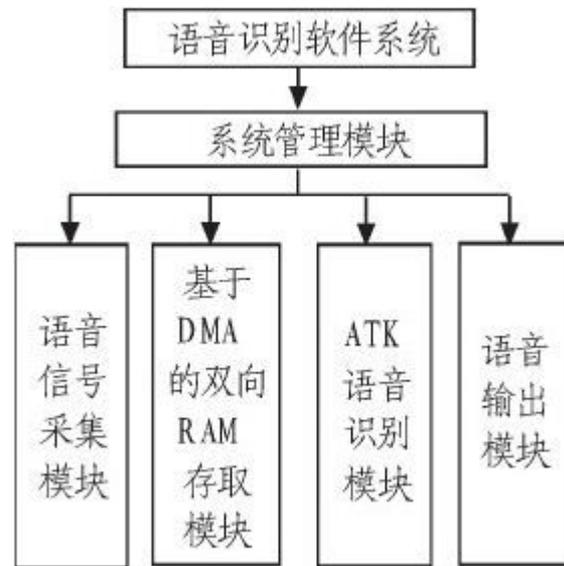


图 4 系统软件设计结构图

在软件设计中，系统管理模块主要负责系统的总体管理调度，是应用系统的调度中心；语音信号采集模块主要负责控制数据采集芯片 TPS 65930；基于 DMA 的双向 RAM 存取模块主要负责实现 DMA 驱动及双向 RAM 的读写存取，使用了通道 1 来实现高速地把语音信号采集到的数据存储到 RAM 上，并使用通道 2 实现高速地把 RAM 的数据取出来，用于语音的识别；语音输出模块主要负责把相应的音频数据送到 TPS65930，并控制 TPS6 5930 对接收到的音频解码输出到功放电路，实现语音输出的功能。软件的设计流程图如图 5 所示。

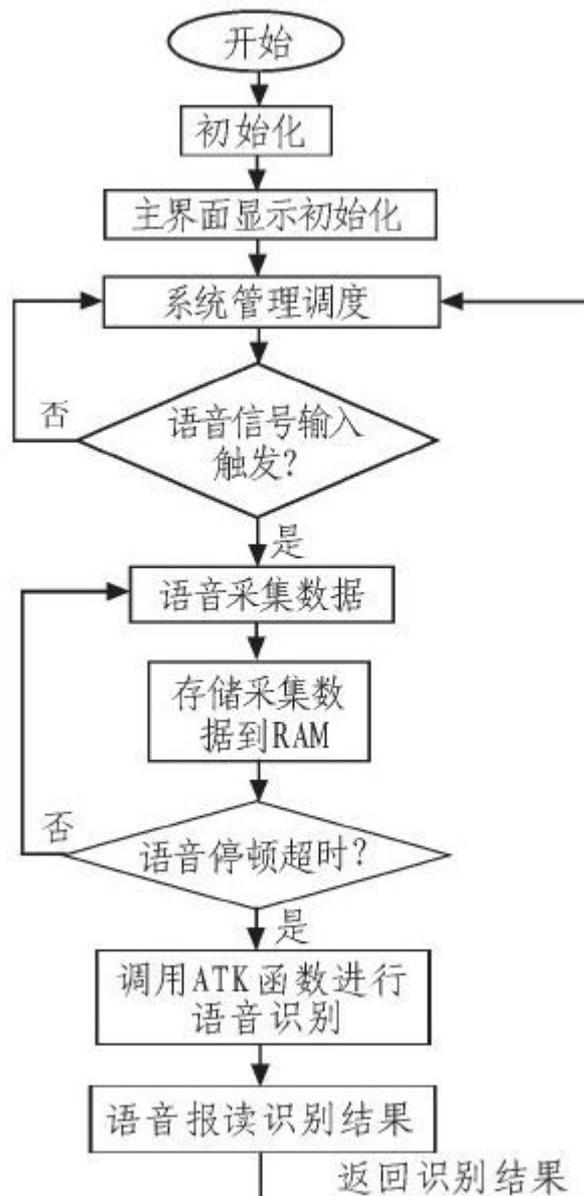


图 5 软件设计流程

2 语音识别系统实验及结果

文中设计的语音识别系统如图 6 所示。在实验中总共进行了 3 轮话语测试，每轮 300 句话语测试，其中 150 句为家庭监护机器人需要识别的话语，150 句话语为机器人不予置理的无关话语。本系统只设置 10 个需要识别的话语，由 15 名学生分别读音进行测试。同时对无关话语也是由该 15 名学生，每人 10 句分别随机读音测试。从显示结果可以看出，第 1 轮中，先测试的 150 句无关话语中，能正确识别出无关话语数为 150 句，识别出无关话语率为 100%，但对需识别的话语中，正确识别出 148 句，2 句识别出错，识别率为 98.67%；在第 2 轮中，同样先测试的 150 句无关话语中，能正确识别出无关话语数为 150 句，识别出为无关话语率为 100%，但对需识别的话语中，正确识别出 142 句，8 句识别出错，识别率为 94.67%；在第 3 轮中，同样先测试的 150 句无关话语中，能正确识别出无关话语数为

150 句，识别出无关话语率为 100%，但对需识别的话语中，正确识别出 146 句，识别率为 97.33%。在 3 轮测试中，系统都能实时响应所有语句，未出现漏句现象。并且在实验测试的过程中，系统的响应速度都非常快，感观上没有时间延迟，与真人交流速度相当。

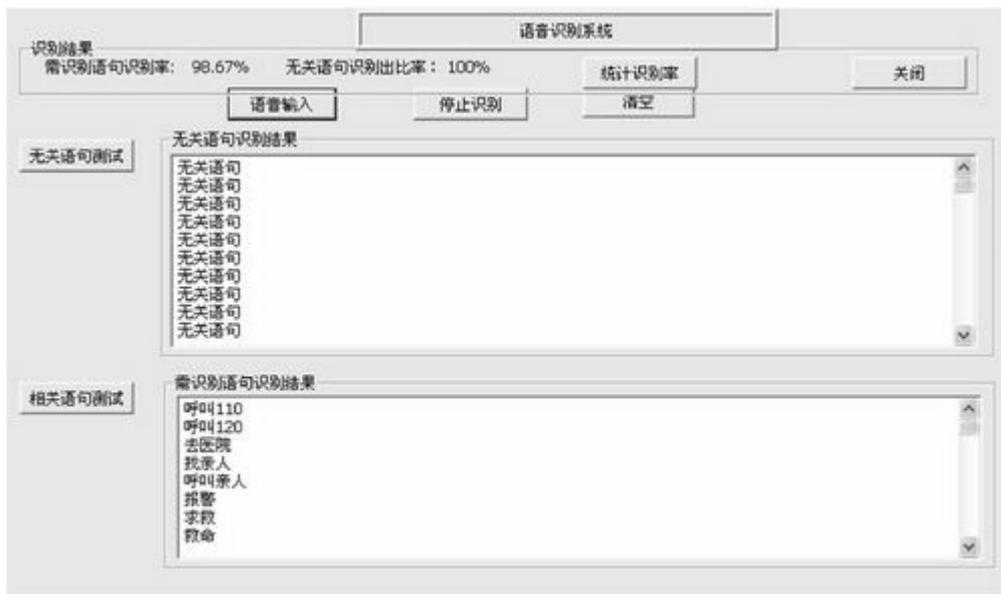


图 6 系统实验界面

从实验测试结果可看出，对于识别 10 个需要识别的话语的识别率高达 94.67% 以上，具有较高的识别率，因而该语音识别系统较好达到了家庭监护机器人使用的要求。

3 结论

文中通过 DSP、DMA 和 ARM Cortex-A8 的并行处理，利用双缓冲的方法，在嵌入式 Linux 上实现了基于 ATK 的实时语音识别系统。该系统可以实时地实现语音识别，具有较高识别率，较快的响应速度。可以应用在家庭监护机器人及其相关领域中。