

语音识别 SoC UniLite 的系统设计

杨之佐¹, 董明¹, 刘加¹, 刘润生¹, 孙旭东²

(1. 清华大学电子工程系, 北京 100084; 2. 中航一集团沈阳飞机设计研究所, 沈阳 110035)

摘要: UniLite 系统以 DSP 为核心, 集成了片内存储器、AD/DA 通道和多种的输入输出端口, 有效控制了系统的成本和功耗, 在有限的 RAM 空间上实现了高性能非特定人语音识别以及语音编解码的功能。系统采用连续隐含马尔科夫 (CHMM) 算法, 运用了分阶段提取特征、束搜索等策略, 在保证系统识别性能的同时大大降低了内存消耗, 提高了识别速度, 识别率在 98.5% 以上, 识别时间在 0.5 倍实时以下。

关键词: 嵌入式系统; 语音识别; DSP; SoC 设计

System Design of Speech Recognition SoC UniLite

YANG Zhizuo¹, DONG Ming¹, LIU Jia¹, LIU Runsheng¹, SUN Xudong²

(1. Department of Electronic Engineering, Tsinghua University, Beijing 100084;

2. China Aviation Shenyang Aircraft Design & Research Institute, Shenyang 110035)

【Abstract】 The UniLite chip is realized around an architecture which is composed of a 16-bit DSP core, on-chip memory, an ADC/DAC channel as well as other common I/O interfaces. High performance speaker independent speech recognition and speech coding are implemented on limited RAM memory. This results in low power consumption and reduction in cost. The system is based on continuous hidden Markov model. Two-pass feature extraction and beam search strategy are also employed to reduce the consumption of the RAM and enhance the recognition rate. Test shows that the UniLite system provides recognition accurate as high as 98.5% and requires less than 0.5 real time.

【Key words】 Embedded system; Speech recognition; DSP; SoC design

随着半导体工业和语音识别、编解码技术的发展, 市场对于嵌入式语音处理系统的需求也前所未有的迫切。但是大部分语音识别芯片较高的成本和功耗限制了在智能玩具、语音遥控器等方面的应用。为此, 清华大学设计开发了新一代的语音识别专用芯片 UniLite。

根据系统的设计要求, UniLite 语音处理系统以一个 DSP 核为核心集成了片内 RAM 和 ROM 存储器及 Codec、UART、SPI 等外围接口, 采用了基于子词的连续隐含马尔科夫模型的语音识别算法及 G.723.1 低速率语音编解码算法, 在单一芯片上实现了高性能语音综合处理系统。

1 系统设计特点

1.1 系统的设计目标

根据应用需求, 市场对 UniLite 语音识别系统提出以下要求: (1) 高性能中小词汇量文本输入的非特定人语音识别; (2) 低速率语音压缩, 语音压缩率在 8Kbps 以下; (3) 玩具、遥控器等应用要求芯片的总体成本要控制在 2 美元以下; (4) 应用产品的可移动性对系统的功耗有严格限制, 系统平均功耗需控制在 100mW 以内; (5) 高清晰度录音; (6) 对外控制的功能; (7) 实时性的要求, 要求系统能够实现实时的语音识别和语音压缩。

1.2 系统核心处理器的设计特点

根据系统算法的复杂度以及对于实时性的要求, 选用 16 位高速定点 DSP 核 OAK 作为核心运算处理器件, 其最高速度可达 104MHz。另外 OAK DSP 核还承担了核心控制的任务。单 DSP 核的结构简化了系统结构, 使得系统资源得到有

效调配, 降低了系统的整体功耗和成本。系统工作频率为 60MHz 时, 整体功耗仅为 90mW。

1.3 系统存储器的设计特点

为了提高系统运算速度, UniLite 芯片采用了哈佛总线结构, 根据系统语音识别及语音压缩算法的要求, 片内存储设备包括: 16KB 的 SRAM 片内程序存储器, 存储应用程序和驱动程序; 12KB 的 SRAM 片内数据存储器, 存储系统运行时所需要的数据; 32KB 的 ROM 片内程序存储器, 存储语音识别以及语音压缩的算法; 20KB 的 ROM 片内数据存储器, 存储算法中所用的常数表。RAM 存储器与 ROM 存储器相结合的结构特点, 在保证一定应用灵活性的同时大大降低了系统的整体成本。

1.4 系统模拟通道的设计特点

为了降低系统整体成本, 提高处理语音信号的质量, UniLite 系统集成了一路片内 AD/DA 通道, 采用 $\Sigma\Delta$ 调制方式, 能够在 8kHz 或者 16kHz 的采样频率下得到 12 位线性量化的数据, 可以根据具体应用的要求得到不同音质的语音或音乐。

1.5 系统的硬件结构

除 DSP 核、存储设备、AD/DA 通道之外, 系统还在片内

基金项目: 国家自然科学基金资助项目(60572083)

作者简介: 杨之佐(1981 -), 男, 硕士生, 主研方向: 嵌入式语音识别系统开发; 董明, 博士生; 刘加, 教授; 刘润生, 博士; 孙旭东, 高工

收稿日期: 2005-12-01 **E-mail:** yangzz03@mails.tsinghua.edu.cn

集成了 16 个GPIO接口、UART、SPI等其他外围通信设备接口^[1]。实现与外部存储器、PC等外部设备的数据交换，提供与用户间的辅助交互接口。系统的硬件结构如图 1 所示。

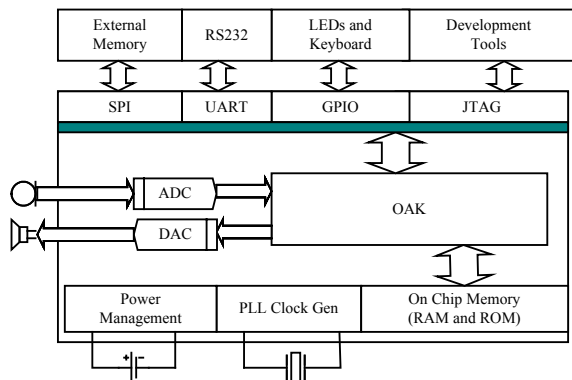


图 1 UniLite 系统的硬件结构

1.6 系统的风险控制

由于 UniLite 采用 0.18 μ m CMOS 工艺制造，ASIC 的整体成本与系统设计的风险有很大的关系，芯片的再次生产将使得系统的整体成本大大提高。因此保证一次性成功率成为系统设计中的重要因素。生产前需要做大量的测试以降低硬件系统整体风险。就软件系统来说，启动程序的风险性是最高的。系统的启动是系统加载程序的前提，因此启动程序中的问题将使得整体系统不能工作。

为了降低 SoC 设计的风险，减少生产成本，在 UniLite 语音识别系统中设计了一种特别的“双启动模式”。根据外部设置，用户可选择硬件启动或软件启动模式。两种启动模式均可将外部存储设备中的特定地址段的程序读入内存后启动系统工作。这样任何一种启动模式都可保障系统正常工作，大大降低了系统设计的风险，提高了系统的可靠性。

系统的硬件驱动程序的风险性也相当高，驱动程序的问题将导致某个硬件模块不能使用。为了降低驱动程序的风险性，将驱动程序存储在外部存储设备中，由启动程序读入 RAM 后使用。这种设计使得驱动程序在芯片生产后随时可进行修正，大大降低了驱动程序的风险性。

系统的核心算法程序的风险性相对于启动和驱动程序要小得多。由于 UniLite 系统采用 OAK DSP 核作为核心运算处理器件，软件仿真器可对系统核心算法程序进行仿真。因此大部分核心算法程序用 ROM 存储器存储，减少了 RAM 空间的开销，降低了系统成本，增加对核心算法的保密性。

1.7 系统的功耗控制

UniLite 系统面向语音遥控器、智能玩具等应用产品。系统对功耗控制提出了很高的要求，而系统的整体功耗与系统时钟有密切联系。在 UniLite 系统中，SPI、UART、GPIO 和 CODEC 等模块的时钟与 DSP 核的时钟是独立的，系统可根据应用的需要切断暂时不用的模块的时钟。

UniLite 系统作为综合的语音处理系统，要实现语音识别、提示音播放、对外控制及键盘扫描等多种子任务。根据系统实时性的要求，处理不同的子任务所需要的最低时钟频率有所不同，单一时钟频率的 DSP 核很难兼顾系统性能和功耗。基于这样的考虑，UniLite 的电源管理系统可根据系统要求调整核心运算控制器件 OAK 的时钟频率，在保证系统实时性的基础上大大降低了系统功耗。系统的不同算法所需的最低时钟频率见表 1。

表 1 算法所需的时钟频率

算法	工作频率
G.723 编码	60 MHz
G.723 解码	20 MHz
特征提取 (13 维)	10 MHz
识别	104 MHz

由于各种算法的复杂度不同，可根据当前算法适当地降低系统时钟频率从而降低系统的整体功耗。另外，在等待外部输入的情况下，可将系统的时钟降到 1MHz，只处理简单的输入信号。如果系统长时间不工作，可彻底关闭 DSP 的时钟，使得系统进入休眠状态，仅响应特定的外部中断信号，将系统再次唤醒。在休眠状态下，系统的整体功耗将降至在 2 μ W 以下。

1.8 系统的加密保护

UniLite 系统对程序代码作了有效的加密保护措施。任何试图访问程序存储器中代码的操作必须得到相应的密钥授权认证。系统对程序存储器设置了 4 层密钥保护，对程序存储设备的未授权访问将使得系统重新启动。为了满足某些用户访问程序 RAM 存储设备的要求，UniLite 系统对程序 RAM 存储器设置了 4 个级别的密钥保护，可对用户不同层次地开放 RAM 存储器。在提高保密性的前提下，满足了特定用户的需求。

2 系统软件体系结构

2.1 软件的体系结构

UniLite 软件系统分为 3 层：驱动层，算法层和应用层。驱动层程序对应一个实际的硬件模块，把这部分硬件模块的功能抽象出来，完成虚拟硬件系统的功能。大部分算法层程序存储在 ROM 存储器中，仅保留对外的程序接口。应用层是整个系统中最灵活的部分，可根据用户实际需要实现某个产品的具体功能。在这样的系统框架结构下，可以根据实际需要迅速建立起新的综合语音处理系统，大大提高了系统的灵活性。

由于 UniLite 系统中的片内数据存储器仅为 12KB，在有限的内存空间上实现高性能的语音识别需要对系统软件进行新的改进。

2.2 两阶段识别算法

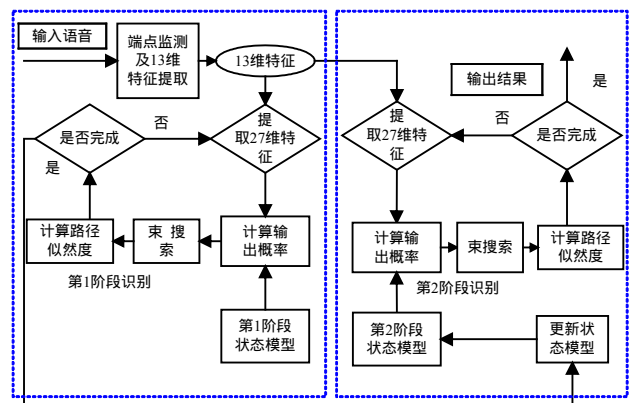


图 2 识别算法的基本框架

在基于HMM识别算法中，系统的识别效果与模型的复杂度有很大的关系，精细的模型能够提高系统的识别性能，但是会大大增加系统的计算量和内存占用量。根据系统的这一特点，识别算法采用了两阶段的识别算法^[2,3]，先使用粗糙声

学模型得到多候选识别的结果,然后采用复杂的声学模型作第2次识别,就可以得到和原来的系统相似的识别性能。在进行第2次识别时,识别词表规模可以从原来的数十甚至上百词条减少为数个词条,其搜索时间远小于同样性能的一阶段Viterbi解码算法。另外第2阶段识别过程中可以直接利用第1阶段输出概率分数占用的内存空间,提高了系统内存的使用效率。系统识别算法的基本框架如图2所示。

2.3 MFCC 特征的分阶段提取

在以往的语音识别系统中,为了提高系统的识别性能,通常直接提取27维或者39维MFCC特征。以27维MFCC特征为例,UniLite系统中语音的帧移为16ms,这样当待识别的语音长度为1.6s时,存储语音特征的内存空间就为5400B,这样大的内存消耗是系统不能承受的。根据MFCC特征的特点,高维特征是与低维特征相关的。仅低13维的MFCC特征中就包含了语音识别所需要的所有信息。为了减少UniLite芯片片内RAM存储空间的消耗,特征提取时仅提取13维的MFCC特征,即12维MFCC参数以及短时归一化对数能量E。为了进一步提高系统的识别性能,在系统进行Viterbi解码之前逐帧提取高14维特征,即12维 Δ MFCC(MFCC的一阶差分)短时归一化对数能量E的一阶差分 ΔE 和二阶差分 $\Delta^2 E$ 。由于系统工作的最高时钟频率为104MHz,而提取高维特征的计算量并不大。这样虽然使得识别时间略有增长,但是用于存储100帧语音特征的内存空间消耗量就减少为2600B,兼顾了系统的识别性能,大大节省了系统RAM空间的消耗。

2.4 基于滑动窗的束搜索策略

为了进一步节省硬件的RAM存储空间,UniLite系统采用了基于滑动窗的束搜索策略^[4]。根据t时刻的路径似然度设置了一定的束宽,如果路径似然度落在束宽以外,则该路径会被剪枝,不能继续向后延伸。在识别率下降不大的情况下,基于滑动窗的束搜索降低了系统的内存消耗,节省了Viterbi解码的时间。例如当搜索窗宽度为10,候选词条为6个字的时候,可以将搜索速度提高至原Viterbi解码算法的3倍以上,内存占用量减少约250B。

(上接第187页)

法的缺点,但是建立和管理一个索引结构需要较多的时间和较大的空间开销,它需要根据匹配的情况不断地进行修改,以使得经常被匹配上的案例总能位于案例库的前面,同时最坏情况的可能仍然存在。

以上改进的算法还只是实验数据样品空间下进行的比较和验证,对于系统实际应用时的行业实际数据是否适用,相关的一些参数是否需要调整,算法的合理性和效率的验证等问题,由于时间有限,仍需在将来系统的推广应用中得到校验和改进。

3 结论

基于案例的推理是伴随着认知心理学的研究而发展起来的一种推理方法,是一种类比推理。它的出现丰富了构造专家系统的方法,同时解决了基于规则推理所面临的困难和弊端。本系统采用了2级推理的模式,即基于案例推理和基于规则推理相结合的方法,一定程度上解决了案例不全、效率不高、对推理结果不满意等问题。但是本系统仍存在一些问

3 实验结果

系统的训练集采用经过信道转录后的863大词汇量连续语音数据库;系统的识别集为本实验室自录的孤立词库,10名男性,10名女性,录音和测试环境为办公室环境,8kHz采样,12位线性量化,束搜索的滑动窗宽度为10,候选词条为2~6个字。测试结果如表2所示。

表2 UniLite 系统测试结果

测试候选词条数	平均识别率	平均识别时间
50个	98.5%	0.47倍实时
20个	99.2%	0.24倍实时
10个	99.6%	0.13倍实时

4 小结

本文介绍了一个新的片上语音识别系统的设计目标及软硬件的设计特点。系统在以一个DSP核为核心的芯片上集成了一路AD/DA通道、片内存储设备及其它外围设备接口。在28KB的SRAM上实现了语音识别、提示音、语音回放等功能。当识别词条为50时,识别率达到98.5%,识别时间在0.5倍实时以下,满足了实际应用的要求。由于UniLite是清华大学为语音识别“量身定做”的芯片,充分考虑了语音识别算法的特点和市场对于嵌入式系统的低功耗、低成本的要求。该芯片将广泛应用于智能玩具、语音遥控器、录音机等语音信号综合处理的应用领域。

致谢:感谢英飞凌公司在UniLite芯片测试和生产方面给予的大力支持。

参考文献

- 1 Chan Eric Infineon Technologies. M7005-UniSpeech-Lite Device Specification[Z]. 2005.
- 2 Zhu Xuan, Chen Yining, Liu Jia, et al. A Novel Efficient Decoding Algorithm for Cdhmm-based Speech Recognizer on Chip[C]. Proc. of ICASSP'03, 2003.
- 3 朱璇,陈一宁,刘加等.语音识别片上系统的多级搜索算法[J].电子学报,2004,32(1):150-153.
- 4 朱璇,李虎生,刘加等.高性能汉语数码串快速识别算法研究[J].计算机研究与发展,2001,38(7):845-850.

题,例如知识的表示和知识获取、案例库的尺寸和维护、提高系统智能性等,这些对本系统的改进和CBR的探究,都是很好的研究方向。

参考文献

- 1 杨善林,倪志伟.机器学习与智能决策支持系统[M].北京:科学出版社,2004:79-112.
- 2 郭艳红,邓贵仕.基于事例的推理(CBR)研究综述[J].计算机工程与应用,2004,40(21):1-5.
- 3 张良,田盛丰.中国人工智能进展[M].北京:北京邮电大学出版社,2003:777-780.
- 4 Leake D. A Case Study of Case-based CBR[C]. Proceedings of the 2nd International Conference on Case-based Reasoning, 1997:371-382.
- 5 Reategui E B, Campbell J A. A Case-based Model that Integrates Specific and General Knowledge in Reasoning[J]. Applied Intelligence, 1997,7(1):79-90.