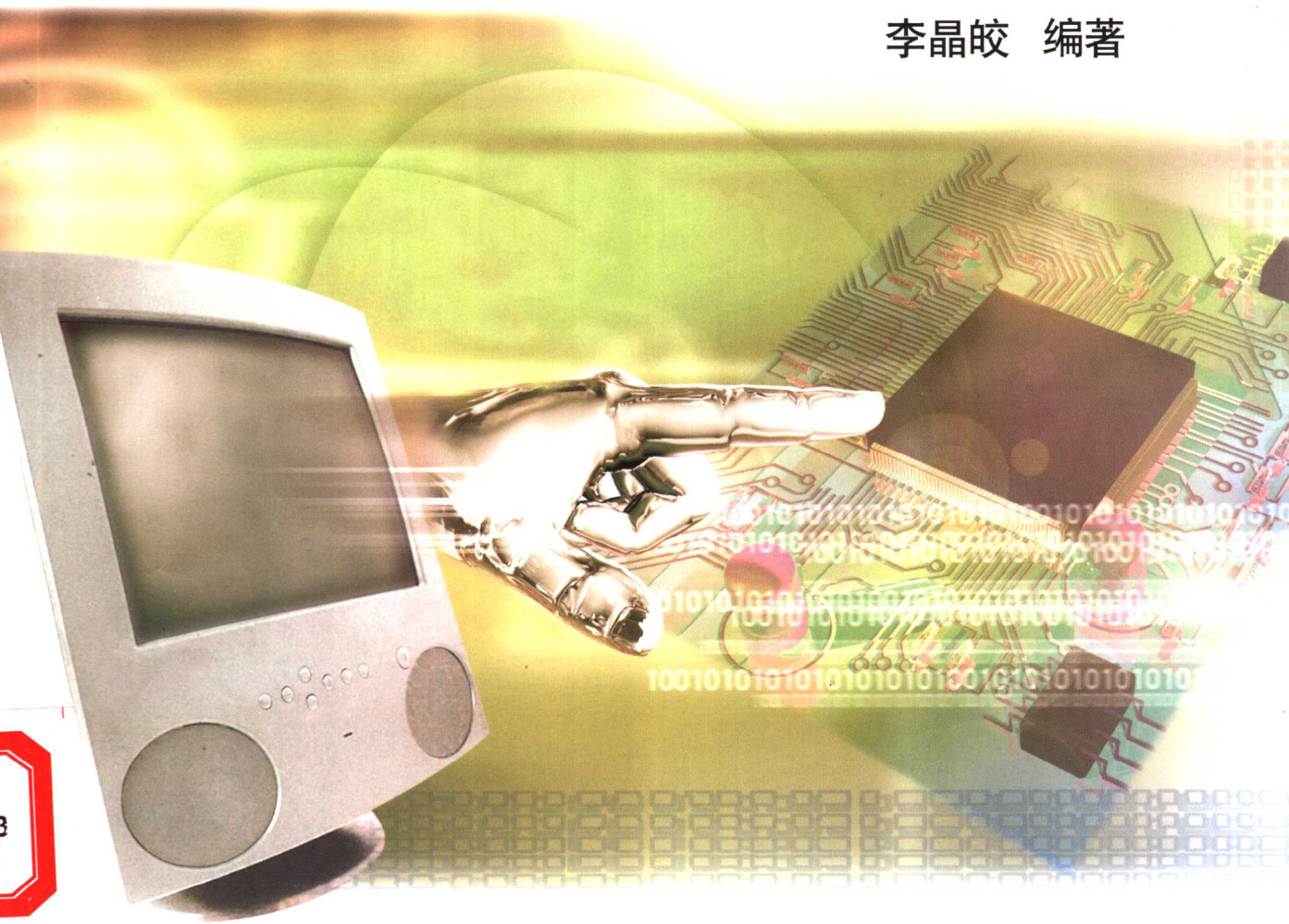
MCU

嵌入式语音技术及凌阳16位单片机应用

李晶皎 编著



北京航空航天大学出版社

嵌入式语音技术及凌阳16位单片机应用

李晶皎　编著

北京航空航天大学出版社

内 容 简 介

本书全面系统地阐述语音技术的基础、原理及方法，并结合凌阳16位单片机的应用，介绍嵌入式语音识别技术和语音合成技术的综合应用系统的设计方法。全书共10章，介绍了语音分析技术、语音存储与回放技术、语音识别技术及语音合成技术，并结合凌阳单片机给出了应用实例。另外，还介绍了凌阳公司的16位单片机，并通过实例、典型应用电路等，讲述硬件电路设计和软件编程方法，提供了解、熟悉和掌握嵌入式语音技术应用系统设计的途径。

本书配光盘一张。其内容包括实验设备图片、各种应用例程、集成开发环境IDE及相关资料文档。

本书内容充实，系统性强，具有广泛的应用性，既可作高等院校相关专业的教材，也适合于从事语音识别、人工智能、模式识别、信息与控制及计算机应用的科技人员阅读。

图书在版编目(CIP)数据

嵌入式语音技术及凌阳16位单片机应用/李晶皎编著.
北京：北京航空航天大学出版社，2003.11
ISBN 7-81077-365-8

Ⅰ.嵌… Ⅱ.李… Ⅲ.单片微型计算机－语音数据处理 Ⅳ.TP368.1

中国版本图书馆CIP数据核字(2003)第086595号

嵌入式语音技术及凌阳16位单片机应用

李晶皎 编著

责任编辑 王 实

*

北京航空航天大学出版社出版发行

北京市海淀区学院路37号(100083) 发行部电话：(010)82317024 传真：(010)82328026

http://www.buaapress.com.cn E-mail:bhpress@263.net

北京市云西华都印刷厂印装 各地书店经销

*

开本：787×1 092 1/16 印张：18.5 字数：474千字

2003年11月第1版 2003年11月第1次印刷 印数：5 000册

ISBN 7-81077-365-8 定价：32.00元

前　言

语音技术是研究用数字信号处理技术对语音信号进行处理的一门学科，是在多学科基础上发展起来的综合性技术。语音技术涉及数字信号处理、模式识别、语音学、语言学和人工智能等许多学科。

随着芯片制造水平的不断提高，芯片的功能越来越强，而价格越来越低；同时，语音处理技术日趋完善，从而使嵌入式语音技术得到发展和应用。

台湾凌阳科技有限公司近年推出了 μ'nSP 系列单片机，有 8 位和 16 位共 50 多种型号和产品。值得特别介绍的是 16 位单片机。由于在 16 位单片机内核中增加了 DSP 功能，使其特别适合语音识别、语音应答及语音编码/解码等方面的应用。

本书的主要特点是：

① 系统地讲解语音信号处理的基本原理和方法，并详细介绍适合嵌入式系统应用的方法。本书的前 5 章主要介绍语音信号处理技术的发展和基础知识，部分章节结合凌阳单片机给出了应用实例。

② 后 5 章重点介绍凌阳 16 位单片机片内资源及其与语音处理相关软件的编程方法，并给出了嵌入式语音识别和语音应答等应用实例。

③ 读者可自己动手用语音信号处理的知识编制相关软件，用凌阳 16 位单片机硬件平台构造自己的嵌入式产品，也可使用凌阳公司提供的 API 函数，制作各种语音压缩的应用产品。

④ 本书配备光盘。其内容包括实验设备图片、实用例程、集成开发环境 IDE 及相关的资料文档等。

本书由李晶皎任主编。第 2 章和第 5 章由胡峻辉编写；第 6 章、第 7 章及第 9 章由王爱侠、张广渊及胡明涵编写；第 10 章由王显巍、甄广启及赵骥编写；其余章节由李晶皎、张俐编写。

感谢台湾凌阳公司和北京北阳公司在产品资料和开发工具等方面提供的各种帮助和支持。

语音处理技术本身就是一门理论性强、实用面广及难度较大的交叉学科，而要在单片机上实现具有特定功能的嵌入式产品，更增加了难度。尽管作者在各方面作了很大努力，但受水平和经验所限，编写时间又很仓促，书中会有缺点及疏漏之处，敬请读者给予批评指正。

有关 SPCE 单片机的资料、应用信息和最新动态，请访问如下网站：

凌阳公司　http://www.sunplus.com.tw

北阳公司　http://www.unsp.com.cn

本书得到东北大学学位与研究生教育科学研究计划基金的资助。

感谢您选择了这本书，希望您喜欢！

作　者

2003 年 6 月于东北大学信息学院

目　录

第 6 章　凌阳 16 位单片机介绍

第 7 章　指令系统

第 8 章　SPCE061A 硬件结构

第1章 概 述

1.1 语音处理技术的发展

语言既是人类创造的，亦是人类区别于其他地球生命的本质特征之一。语音是语言最本质、最自然、最直接的表现形式或载体，其表现形式为声波——一种由空气分子振动而形成的机械波。人类用语言交流的过程可以看成是一个复杂的通信过程，为了获取便于分析和处理的语音信源，必须将在空气中传播的声波转变为包含语音信息并且记载着声波物理性质的模拟(或数字)电信号，即语音信号，因而语音信号就成为语音的表现形式或载体。

人们对语言的研究早已有之，其中语音学是研究人类语音的产生、传播及感知等过程机理的学科，包括发音语音学、声学语音学和听觉语音学3个分支。发音语音学研究发音器官在发音过程中的运动和语音的音位特性；声学语音学研究语音的物理属性(语音声波的振幅、频率和频谱特性等)；听觉语音学研究听觉和语音感知。数字信号处理是一门通过计算机或其他专用设备，对离散信号用数字方式进行增强、压缩、滤波、变换及识别等处理的新兴学科。语音学和数字信号处理的交叉结合便形成了语音信号处理。语音信号处理(简称语音处理技术或语音技术)是建立在语音学和数字信号处理基础之上的，对语音信号模型进行分析、存储(编码)、传输、识别和合成等方面研究的一门综合性学科。它包括语音编码、语音识别、说话人识别和语音合成四大学科分支，并由此形成了语音分析技术、语音存储(编码)技术、语音识别技术和语音合成技术四大实用技术。

自1876年Bell发明了采用声电转换技术实现远距离语音通信的电话开始，语音处理技术的发展大致经历了以下几个阶段。

(1) 萌芽阶段

在这一阶段(20世纪30年代至50年代)，人们对语音处理的研究主要是根据语音学知识，提取若干特征参数，并利用这些参数制作成模拟电路来模仿人的发音过程，实现简单的语音处理功能。

1930年，H. Dudley首次成功研制出采用传输并合成的方法从语音中提取表征语音信息特征参量的声码器，其创造性的设计思路形成了语音产生模型的基本思想；1948年，美国Haskins实验室首先研制出由语谱图自动合成语音的语图回放机，这一发明直接孕育出共振峰合成法这一至今仍被认为是较好的语音合成的方法；1952年，Bell实验室的Davis等人首次成功研制出可以识别10个英语数字的语音识别系统——Audry系统，它根据语音第一、二共振峰提取若干特征参数以形成参考语音模式，然后通过计算参考语音模式的语音与未知语音之间的互相关程度来达到识别目的；1958年，Duddley和Balashek对其进行了改进，将语音分割为元音和辅音等语音单位，开创了音素识别的先河；1956年，Olson和Belar等人采用了8个带通滤波器，提取频谱参量作为语音特征，研制成功了一台简单的声控打字机。

(2) 发展阶段

在这一阶段(20世纪60年代至80年代初),随着集成电路技术和计算机技术的发展,语音处理的理论和技术亦日趋完善和成熟。

1960年,Denes和Mathew把数字计算机引入语音识别,从而改变了采用模拟电路进行语音处理的传统做法。计算机的应用推动了语音识别的发展,形成了线性预测分析技术LP(Linear Prediction)和动态规划DP(Dynmic Programming)两项重要成果。LP较好地解决了语音信号产生模型的问题,对语音识别的发展产生了深远影响。1966年美国麻省理工学院林肯实验室的Gold等人采用16个带通滤波器、基音检测器、浊音检测器和一台计算机构成了一个语音识别系统。在整个20世纪60年代,语音识别的研究主要是根据语音产生机理和人耳对不同频率语音的感知差别,采用硬件实现的滤波器组提取频谱特征,通过计算机进行匹配和判决。

20世纪70年代至80年代初期,语音处理技术得到了迅速的发展,期间产生了一些重大的理论突破:

- 20世纪70年代,由Itakura提出的表示语音参数相似度测量的线性预测残差原理以及线性预测编码(LPC)较好地解决了语音特征的提取问题,从而使线性预测(LP)技术成功地应用于语音编码及语音识别。
- 人们把用于解决有序优化问题的动态规划(DP)技术应用到语音识别中。1972年由Sakoe提出的动态时间弯曲DTW(Dynamic Time Warping)算法有效地解决了语音两次发音之间的时间变形问题,对特定人孤立词的识别十分有效。从此,基于LPC分析及DTW算法的中、小字表孤立词特定人语音识别系统纷纷建立起来,语音识别开始走出实验室而进入实际应用阶段。
- 20世纪70年代末至80年代初,Gray和Markel等人解决了矢量量化码书生成的方法。于是,一些学者就将这项原本用于信息压缩理论中的矢量量化VQ(Vector Quantizaton)技术成功用于语音编码以及语音识别之中。矢量量化的作用就是进行数据压缩,将连续的语音特征空间离散化,降低系统在时间及空间(存储)上的开销,从而减小语音处理的复杂程度。矢量量化的另一个作用就是通过聚类分析获取一人或者多人的多次语音样本所共有的语音特征。
- 产生于20世纪六七十年代,却从20世纪80年代中期开始得到极大发展并成为语音处理研究热点的隐马尔可夫模型HMM(Hidden Markov Models),逐渐成为现代语音处理领域的重要理论基础之一,并在语音处理的各个领域中得到十分广泛的应用。

(3) 实用阶段

在这一阶段(20世纪80年代至今),随着遵循摩尔定律的超大规模集成电路技术的迅速发展,PC机的触角深入到千家万户,极大地促进了计算机技术和人工智能技术的迅猛发展,使人类社会进入到数字信息时代。在此社会背景下,人们对语音技术的实际需求愈发迫切,极大地促进了语音处理技术的不断深入和发展,使语音处理实用化产品不断出现。

从20世纪70年代末开始,由于大规模集成电路技术和语音理论与技术的成熟,语音处理技术开始步入实用化阶段。1976年,Votrax公司推出的Computalker语音合成器投入市场。它采用8080微处理器,并采用S-100总线与其他微计算机系统连接,有6 KB的存储器存储

音素表和程序。当输入机器可读的标准语音表代码后，Computalker 产生合成语音。虽然合成语音的音质很差，但是合成语音技术本身已经为人们所接受。同年，Votrax 公司推出另一款语音合成器——ML－I，它是第一个由规则合成语音的产品。ML－I 采用 80 个音节、8 级音高和4 级不同发音持续时间，并提供了一份包含 625 个单词和短语的词典。1978 年，TI 公司首次推出采用超大规模集成电路技术的单片语音合成器 TMCO280。该产品成为语音 DSP 芯片的前身，并使 TI 公司遥遥领先于同行。TI 公司用此芯片推出了一种产品——Speak'n Spell Toy，使语音技术走出实验室进入市场。其硬件结构采用 4 位微处理器 TMS1000，2 个 128 K位的 ROM 存储约 330 个单词和短语(语音持续 3～4 min)，数据传输率为 1 200 b/s；采用线性预测合成法，由格型滤波器实现，10 级格型滤波器用 10 个反射系数表示；语音合成的控制参数有 12 个：10 个反射系数、1 个能量参数和 1 个音高参数。继 1978 年 TI 公司推出会讲话的 Speak'n Spell Toy 之后，又出现了许多会讲话的产品，如会讲话的怀表、会讲话的微波灶、会讲话的弹球机、会讲话的计算器等，可以把它们看成是嵌入式语音产品的雏形。

20 世纪 80 年代初，Votrax 公司推出了采用音素合成技术的大规模集成电路芯片 SC－01。随着 PC 机技术的飞速发展和小词汇量特定人孤立词语的语音识别技术日趋成熟，特别是对 HMM 的深入研究和广泛应用，出现了语音处理技术产品化的热潮。1985 年，东京的 Matsushita 研究所研制了非特定人孤立系统，它包括：LPC 倒谱系数、辅音号分段、元音和半元音识别、辅音识别、音节序列和词的匹配。该系统对 274 个词的识别率为 95.6%；在 Yamatokoriyama 的夏普信息系统实验室，根据日语的特点(日语大约有 100 个音节，语音以音节为单位)，用音节作为识别基元，用音韵规则得到每个音节间的关系，并采用 DTW 方法，对 300 个孤立字的特定人的识别率是 94%。1985 年，Sharp 公司在超级市场放置了一个用声音操作的字处理系统，引起人们的关注。ATR 将神经元网络用于语音识别。1988 年 Waibel 用时延网络 TDNN 解决了难以区分的 B，D 和 G 的问题。网络能自学一些特征，因此，神经元的识别率是 98.5%，而 HMM 方法的识别率为 93.7%。1989 年，NEC 公司的 Sokoe 将神经元用于孤立字非特定人的语音识别，它利用动态规划技术 DTW 的优点，通过自学习改进性能，对日语数字的识别率为 99.3%。

20 世纪 90 年代初，国外许多研究机构都研制出词汇量达到几万的大词汇量识别系统，例如能识别 70 000 个词汇的 Dragon Dictate 词汇翻译系统，识别率为 80%以上。在非特定人连续词语识别方面有代表性的系统是卡内基梅隆大学(CMU)研制成功 SPHINX 系统，它能识别包括 997 个词汇的连续语句，识别率达到 95.8%。IBM 公司在语音识别领域的研究已有 30 年，其语音技术一直处于世界领先地位，并在这一领域拥有近百项专利。早在 1985 年，IBM 公司就成功研制出 5 000 个词的英语听写机 Tangora－5；20 世纪 80 年代末，研制出能识别 20 000 个词汇的 Tangora－20，识别率达到 94.6%，而且具有快速自适应于说话者的特性；1997 年，又推出了汉语听写机产品 ViaVoice，为语音识别技术在汉字输入方面的实际应用开辟了新的道路。此后，IBM 公司的 ViaVoice 抢占了中国 90%的语音识别市场。

鉴于语音识别产品的鲁棒性较差，对语音处理技术的研究将更加深入。美国 DARPA 战略计算计划局提出研究口语系统(spoken language system)。该系统要求把语音识别与自然语言理解结合起来，即让计算机像人一样具有语言理解能力，而不须过多地在孤立词识别上下功夫，从而形成了新一代语音识别系统。

我国对语音处理技术的研究起步要比先进国家晚一些，在20世纪70年代末期只有中科院声学所、清华大学计算机系等单位从事语音识别的研究工作。经过20多年的努力，我国在语音处理领域取得了长足的进步。

(1) 汉语听写机方面

与英语相比，由于汉语语音输入的迫切性，以及汉语音节种类较少，结构很规则，协同发音和音变相对不严重，便于以音节识别为基础实现无限词汇识别，因而很快获得了可与国际先进水平相比拟的成果，一些汉语语音输入系统已经实用化。1988年，清华大学、中国科学院声学所首先研制成功无限词汇的汉语听写机；20世纪80年代末，四达技术开发中心率先推出汉语输入的实用产品，并于1991年与哈尔滨工业大学合作推出具有自然语言理解处理能力的汉语听写机。这一时期的汉语听写机系统基本上都是基于特定人孤立音节识别技术的。孤立音节识别系统只能一个字一个字地读入，断断续续的，既不自然又很费力，这样的听写机产品自然不能为广大用户所接受。在国家"863"计划支持下，近几年来清华大学和中国科学院自动化所等单位研制的听写机原理样机，不仅包含一个很大的多音节词表，而且能用于非特定人连续语句输入，用起来就方便多了。目前正在考虑改进性能、开发产品。近几年来，汉语语音识别受到了前所未有的重视，国外有多个公司投资巨款，猛攻汉语听写机的研制，台湾也在听写机研究方面下了很大功夫。国家"863"计划最近几年也加大了这一研究的投资力度，并组织了定期的测试评比活动。显然，当前仍是语音识别研究的黄金时期，做出真正实用的汉语听写机已为期不远了。人们用实现登月计划来比喻研制出真正实用的汉语听写机，目前看来这个日子已经临近。

(2) 汉语语音识别方面

1984年，清华大学研制成功能够识别1 000词的汉语语音识别系统；1986年，哈尔滨工业大学研制出3 000词的汉语语音识别系统；1987年，中科院声学所研制成功通用实时语音识别系统——RTSRS，可以识别几百个汉语成语；1988年，清华大学利用矢量量化和隐马尔可夫模型首次研制成功小字表的非特定人语音识别系统。大字表语音识别系统的汉语普通话全音节识别已经取得了相当大的进展，具有代表性的系统有：汉语孤立字全音节实时识别系统(中科院声学所)，汉语大词汇量语音识别与口呼文本输入系统(中科院自动化所)，HEWTS基于汉语全音节的大字表连接词的语音识别系统(清华大学)。汉语非特定人语音识别也取得了令人鼓舞的成果：1990年，清华大学研制成功中字表和大字表的非特定人语音识别系统，可以识别3 400条汉语成语。连续语音识别和理解的研究工作从20世纪90年代开始，有北方交通大学和四川大学实现的火车订票的实验研究系统。非特定人语音识别在中、小词汇量方面已趋成熟，正在向大字表语音识别系统迈进。1992年，四川大学和西安交通大学联合研制出的连续英汉语音翻译系统，由相对独立的连续语音识别、机器翻译和语音合成3个部分组成，是一个特定人、主题受限、中等词汇的连续语音翻译系统，英语词汇量为150个，且合成汉语语音与汉语分析生成是独立的，所选定的主题是"航空订票及信息查询"。语音识别使用的是离散概率密度的隐马尔可夫模型，识别率为97%。1994年，清华大学计算机系研制的非特定人连续语音输入系统的首选正确率为82%，前五个候选正确率为93%。该系统的非特定人全音节输入系统的首选正确率为男声78.2%、女声72.2%，前四个候选的正确率为男声95.2%、女声98.4%。

(3) 汉语语音合成方面

有限词汇的语音合成器已在自动报时、报警、报站、电话查询服务、电子玩具等方面得到广泛应用。关于文本/语音自动转换系统的研究,许多国家、多个语种都已在20世纪90年代初达到了商品化程度,其语音质量能为广大公众所接受。我国汉语文本/语音转换系统的研究,虽然早有许多单位获得成功,但是到1993年底之前,合成语音质量还不能令人满意。最近几年由于国家"863"计划的大力支持,这方面的研究取得明显的进展。语句可懂度和自然度都大幅度改善,目前正在努力推广应用。

根据以上对国内外语音处理技术发展历史的概述,分类归纳一下各种语音处理技术的发展趋势:

- 语音存储技术的核心是语音编码技术。语音编码的研究始于1939年Dudley的创造性发明——声码器。从那时开始直至20世纪70年代中期,除PCM(脉冲编码调制)和ADPCM(自适应差分脉冲编码调制)已取得较大进展之外,中低比特率语音编码一直没有实质性的突破。到了1980年美国政府公布了一种2.4 Kb/s的线性预测编码标准算法LPC-10以后,整个语音编码技术领域发生了一次质的飞跃,人们梦寐以求的、在普通电话带宽信道中传输数字电话的愿望终于变成现实。众所周知,数字电话具有保密性高、容易克服噪声累计现象、便于进行程控交换等优点。但是,64 Kb/s的PCM、32 Kb/s的ADPCM要占用几十千赫信道带宽,都不便于在普通话路中传输,因此语音压缩编码技术一直是一个令人关注的课题。除PCM,ADPCM,AM(增量调制),LPC(线性预测编码),ME-LPC(多脉冲激励线性预测编码)等声码器之外,美国于1988年又公布了一个4.8 Kb/s的CELP(码激励线性预测编码)语音编码标准算法,欧洲也推出了一个16 Kb/s的规则脉冲激励(RELP)线性预测编码算法,其语音质量都能达到高音质,而不再像单脉冲LPC声码器的输出语音那样不为人们所接受。近几年又出现了更好的编码算法——多带激励声码器(MBU),它可以在2.4 Kb/s的速率下提供较高质量的语音。这些算法都可用单片数字信号处理器实时实现,目前正努力进一步减小时延,使之在移动通信中得到广泛应用。语音编码产品化的过程比语音识别来得容易,因此其研究成果能很快转向实际应用,对通信领域的发展起到了重要的推动作用。
- 语音识别的研究工作大约开始于20世纪50年代,当时AT&T Bell实验室实现了第一个可识别10个英语数字的语音识别系统——Audry系统。20世纪60年代,计算机的应用推动了语音识别的发展。这时期的重要成果是提出了线性预测分析(LP)技术和动态规划(DP)。前者较好地解决了语音信号产生模型的问题,对语音识别的发展产生了深远影响。20世纪70年代,语音识别领域取得了突破。在理论上,LP技术得到进一步发展,动态时间弯度(DTW)技术基本成熟,特别是提出了矢量量化(VQ)和隐马尔可夫模型(HMM)理论;在实践上,实现了基于线性预测倒谱和DTW技术的特定人孤立词语音识别系统。20世纪80年代,语音识别研究进一步走向深入,其显著特征是HMM模型和人工神经元网络(ANN)在语音识别中的成功应用。HMM模型的广泛应用应归功于AT&T Bell实验室Rabiner等科学家的努力,他们把原本艰涩的HMM纯数学模型工程化,从而被更多研究者了解和认识。采用HMM和ANN模型建立的语音识别系统,性能相当好。进入20世纪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等著名公司都为语音识别系统的实用化开发和研究投以巨资。

- 语音合成技术的研究已有200多年的历史，但是真正有实用意义的近代语音合成技术是随着计算机技术和数字信号处理技术的发展而发展起来的，它主要是使计算机能够产生高清晰度和高自然度的连续语音。近几十年来国际和国内的研究主要集中在按规则的文/语转换，即将书面语言转换成口头语言。在语音合成技术的发展中，早期的研究主要是采用参数合成方法。值得提及的是Holmes的并联共振峰合成器(1973)和Klatt的串/并联共振峰合成器(1980)。只要精心调整参数，这两种合成器都能合成出非常自然的语音。而最具代表性的文/语转换系统是美国DEC公司的DECtalk系统(1987)。该系统采用Klatt的串/并联共振峰合成器，可以通过标准的接口与计算机联网或单独接到电话网上提供各种语音信息服务，它的发音清晰，并可产生7种不同音色的声音，供用户选择。多年的研究与实践表明，由于准确提取共振峰参数比较困难，尽管利用共振峰合成器可以得到许多逼真的合成语音，但是整体合成语音的音质仍难以达到文/语转换系统的实用要求。自20世纪80年代末期以来，语言合成技术又有了新的进展，特别是基音同步叠加(PSOLA)方法的提出(1990)，使基于时域波形拼接方法合成的语音的音色和自然度大大提高。20世纪90年代初，基于PSOLA技术的法语、德语、英语、日语等语种的文/语转换系统都已经研制成功。这些系统的自然度比以前基于共振峰合成器或LPC方法的文/语合成系统的自然度要高，并且基于PSOLA方法的合成器结构简单易于实时实现，有广阔的商用前景。最近几年，一种新的基于数据库的语音合成方法正引起人们的注意。在这种方法中，合成语句的语音单元是从一个预先录下的、庞大的语音数据库中挑选出来的。不难想像，只要语音数据库足够大，包括了各种可能语境下的语音单元，从理论上讲就有可能拼接出任何语句。由于合成的语音基元都是来自自然的原始发音，因此合成语句的清晰度和自然度将会非常高。

1.2 嵌入式语音处理技术的发展

嵌入式语音处理技术得到广泛应用的是语音编码技术。语音编码技术促进了移动通信的发展，同时也被广泛用于语音复读机中。而语音识别的嵌入式应用一直是人们研究的热点。

20世纪六七十年代以来，语音识别的研究人员一直致力于语音识别专用芯片的研究，但是，大多数语音识别专用芯片识别性能差，不符合实用的要求。直到近10年来，随着语音识别算法的深入研究和集成电路技术的发展，才出现了一些具有实用价值和市场前景的语音识别专用芯片。其中，较为成功的芯片是由美国Sensory Integrated Circuit公司开发的RSC系列语音识别芯片。它是为消费类电子产品所应用的、低价位的语音识别专用芯片。

根据语音识别性能及识别算法的不同，语音识别专用芯片大致有以下几种类型：

- 由多带通滤波器及线性匹配电路构成的专用IC。这是20世纪80年代初期的产品，也是最早期的语音识别专用集成电路。它由一组带通滤波器组成特征提取电路，然后用线性匹配电路进行模式匹配。这种电路的语音识别性能低，现已很少应用。

- 由单片机(MCU)组成的语音识别专用 IC。它以 8 位机或 16 位机为计算核心，外加 A/D 转换、D/A 转换及存储器。由于单片机的运算能力有限，因而其识别算法不可能复杂，精度也低，故一般识别率不会太高。典型芯片是 1996 年美国 Sensory 公司生产的 RSC－146。
- 由数字信号处理器 DSP 组成的语音识别系统。它一般由定点 16 位 DSP，外加 A/D 转换和 D/A 转换，以及 ROM，RAM，Flash 等存储器组成。由于 DSP 包含用做数字信号处理运算的专用部件，因而运算能力强，精度高，适于组成较高性能的语音识别系统。最常用的 DSP 芯片有 TI 公司的 TMS320AC54XX 系列、AD 公司的 ADSP218X 系列及 DSPG 公司开发的 OAK 系列。用 DSP 组成的语音识别系统可以实现孤立词特定人和非特定人语音识别功能，识别词条可以达到中等词汇量；此外，还可以实现说话人识别以及高质量、高压缩率语音编解码功能，同时可以产生高品质的语音合成和语音回放功能。这是当前语音识别专用芯片的主流组成。
- 由人工神经网络构成的语音识别专用芯片。由于语音信号是一个在时间区间动态变化的信号，因此一般采用多层前向感知器算法。但由于人工神经网络很难达到与语音信号的最佳匹配，因此用人工神经网络实现的语音识别系统的识别性能很不理想。而如果采用时延单元神经网络，并且与其他方法配合，则可以实现较高性能的语音识别。例如 1991 年 GMResLab 利用时延单元神经网络 TDNN(Time Delay Neural Network)模拟芯片实现了特定人英语数字串的识别，8 个数字串的识别率为 98%以上。
- 语音识别系统级芯片 SOC(System on Circuit)。它将 MCU 或 DSP、A/D、D/A、RAM、ROM 以及预放、功放等电路集成在一个芯片上，只要加上极少的电源供电等单元，就可以实现语音识别、语音合成以及语音回放等功能。这是最近两年出现的最先进的语音识别芯片，其性能价格比较高，功耗省。最有代表性的是 Sensory 公司的 RSC－364 及 Infineon 公司的 UniSpeech－SDA80D51。凌阳公司的 SPCE061A 也是这类产品。

嵌入式语音识别系统框图如图 1.1 所示。

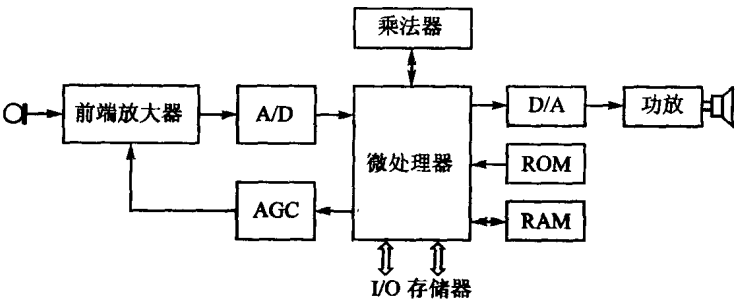


图 1.1　嵌入式语音识别系统框图

语音识别芯片的算法特点如下：

语音信号输入后首先经过滤波器，去除干扰及可能造成混淆的成分，然后由前端处理模块提取语音识别所需的特征参数。目前语音识别所用的特征参数主要有两种类型：线性预测倒

谱系数(LPCC)和MEL频标倒谱系数(MFCC)。

LPCC系数主要是模拟人的发声模型,未考虑人耳的听觉特性。它对元音有较好的描述能力,而对辅音描述能力及抗噪性能比较差。其优点为计算量小,易于实现。

MFCC系数则考虑到了人耳的听觉特性,具有较好的识别性能。由于它需要进行快速傅里叶变换,将语音信号由时域变换到频域上处理,因此其计算量大且计算精度要求高,必须在DSP上完成。

语音识别模块的作用是将输入信号的特征与模板库中已训练好的语音模板进行比较识别,找到最好的识别结果。现在应用较为广泛的语音识别的算法主要有以下几种:动态时间规整、离散隐马尔可夫模型、连续隐马尔可夫模型、人工神经网。

语音识别专用芯片的运算处理器是一颗低功耗、低价位的智能芯片。与PC机的语音识别系统相比,其运算速度、存储容量都非常有限。这些由专用芯片实现的语音识别系统有如下几个特点:

- 多为中、小词汇量的语音识别系统,即只能够识别10～100条词条。
- 一般仅限于特定人语音识别的实现,即需要让使用者对所识别的词条先进行学习或训练。这一类识别功能对语种、方言和词条没有限制。
- 由此芯片组成一个完整的语音识别系统。除语音识别功能外,为了有一个友好的人机界面和对识别正确与否的验证,该系统还必须具备语音提示(语音合成)及语音回放(语音编解码记录)功能。
- 多为实时系统,即当用户说完待识别的词条后,系统立即完成识别功能并有所回应。这就对电路的运算速度有较高的要求。
- 除了要求有尽可能好的识别性能外,还要求具有体积尽可能小、可靠性高、耗电省、价钱低等特点。

下面介绍几种典型的语音识别专用芯片:

- RSC-364是美国Sensory Integrated Circuit公司开发的,2000年开始生产的产品。它是一颗为消费类电子产品应用的、低价位的语音识别专用芯片。RSC-364使用预先学习好的人工神经网络进行非特定人语音识别,不需要经过训练就可以识别如Yes,No,Ok等简单语句,其Data Book上称其识别率为97%。此外,RSC-364可以识别特定人、孤立词命令语句,约60条,其Data Book上称其识别率为99%以上。RSC-364还具有5～15 Kb/s的语音合成速率,其语音合成由Sensory专门设计,音质较好。它还具有改进的ADPCM(自适应差分脉冲调制)语音编解码功能,用做语音回放。
- UniSpeech-SDA80D51是德国Infineon公司2000年开始生产的产品。它是一颗高性能的语音专用芯片,能够满足立体声处理或者消除外界干扰等功能要求,例如在汽车上使用时,可以消除发动机和轮胎转动产生的噪声干扰等。UniSpeech-SDA80D51的语音处理软件包括:利用DTW算法的特定人语音识别,能够识别100条语句;利用HMM算法的非特定人语音识别,词汇量可以达到100条语句;高质量、低码率(2.4～13 Kb/s)的语音编解码,用做语音提示和语音回放;回声消除技术,降低外界的噪声干扰;说话人识别功能等等。
- ISD-SR3000是一个嵌入式语音识别器件,是ISD公司开发的Simon系列芯片的第一