高等院校信息与通信工程系列教材

语音信号处理

韩纪庆　张磊　郑铁然　编著

清华大学出版社

高等院校信息与通信工程系列教材

语音信号处理

韩纪庆　张磊　郑铁然　编著

清华大学出版社
北京

内 容 简 介

本书系统地介绍了语音信号处理的概念、原理、方法与应用以及该领域取得的新进展，主要内容包括语音信号处理的发展过程、语音信号的产生与人类听觉的机理、线性语音产生模型、非线性语音产生模型、语音信号的特征分析、语音信号的线性预测方法、语音信号的编码与合成技术、语音识别技术等，最后还介绍了近年来兴起的一些基于语音识别的应用技术。

本书可作为高等院校计算机应用、信号与信息处理、通信与电子系统等专业及学科的高年级本科生、研究生教材，也可供该领域的科研及工程技术人员参考。

图书在版编目(CIP)数据

语音信号处理 / 韩纪庆，张磊，郑铁然编著. —北京：清华大学出版社，2004.9
(高等院校信息与通信工程系列教材)
ISBN 7-302-08831-4

Ⅰ. 语… Ⅱ. ①韩… ②张… ③郑… Ⅲ. 语音信号处理—高等学校—教材 Ⅳ. TN912.3

中国版本图书馆 CIP 数据核字(2004)第 055787 号

出 版 者：清华大学出版社　　**地　　址**：北京清华大学学研大厦
http://www.tup.com.cn　　**邮　　编**：100084
社 总 机：010-62770175　　**客户服务**：010-62776969
组稿编辑：陈国新
文稿编辑：马幸兆
印 装 者：北京国马印刷厂
发 行 者：新华书店总店北京发行所
开　　本：185×260　**印张**：21.25　**字数**：486 千字
版　　次：2004 年 9 月第 1 版　2004 年 9 月第 1 次印刷
书　　号：ISBN 7-302-08831-4/TN・190
印　　数：1～3000
定　　价：35.00 元

前言

语音信号处理以语音为研究对象，涉及心理学、生理学、语言学、人工智能和模式识别等多项研究领域，甚至还涉及到说话时的表情、手势等人的体态语言信息。由于语音是人们在日常生活中的主要交流手段，因此语音信号处理在现代信息社会中占有重要地位。语音信号处理的研究工作最早可以追溯到19世纪70年代，而在20世纪得到了长足的发展，到了20世纪90年代，IBM、Apple、AT&T、NTT等著名公司为语音识别的实用化开发投以巨资，致使语音信号处理技术的应用掀起了热潮。

近年来，随着语音信号处理技术的日益成熟，出现了新的基于语音识别的应用方向，如语音拨号、呼叫中心、移动设备中的嵌入式命令控制、发音学习以及基于关键词检出的口语会话系统等。随着语音信号处理技术在实际生活中的应用的不断发展，语音信号处理技术已经被广泛地接受和使用。由于语音比其他形式的交互方式具有更多的优势，因此这项技术已经越来越贴近人们的生活。目前，语音信号处理技术处于蓬勃发展时期，不断有新的产品被研制开发，市场需求逐渐增加，具有良好的应用前景。

本书不仅让读者对语音信号处理技术的基本理论、方法和典型应用有一个系统的了解，并且将该学科中的一些新的发展动态介绍给读者，使读者在学术思想上受到一些启发。

书中内容涉及作者承担的国家自然科学基金项目的部分研究成果。衷心感谢国家自然科学基金项目(项目号60085001)、教育部跨世纪优秀人才培养计划基金项目的资助，书中包含了这些项目的部分研究成果。

本书从语音信号处理的共性技术出发，首先介绍了预处理、信号数字化以及特征提取技术，然后根据不同的研究方向，分别介绍了语音编码和语音识别等相关技术。内容由浅入深，知识全面、系统。并将一些较新的技术呈现给读者，如目前较受关注的语音信号的非线性产生模型、从时频分析角度考虑的短时分析技术以及语音识别中的关键词检出技术、自适应技术和一些顽健语音识别方法等。

本书的第1、7、9章由哈尔滨工业大学韩纪庆编写，第2、3、6章由黑龙江科技学院张磊编写，第4、5、8章由哈尔滨工业大学郑铁然编写，韩纪庆负责

全书的总体安排和审定，郑贵滨为书稿中的插图做了大量的工作，在此表示感谢。

虽然作者从事语音信号处理工作已 10 余年，积累了大量的语音信号处理的实践经验，但因作者水平有限，时间仓促，文中所述内容难免有不足之处，敬请读者批评指正。

作　者

2004 年 2 月于哈尔滨工业大学

目 录

第1章 绪 论

CHAPTER 1

语言是人类最重要的交流工具，它自然方便、准确高效。随着社会的不断发展，各种各样的机器参与了人类的生产活动和社会活动，因此改善人和机器之间的关系，使人对机器的操纵更加便利就显得越来越重要。随着电子计算机和人工智能机器的广泛应用，人们发现，人和机器之间最好的通信方式是语言通信，而语音是语言的声学表现形式。要使机器听懂人讲话，并能说出话来，需要做很多工作，这就是科学工作者研究了几十年的语音识别和语音合成技术。随着移动通信的迅猛发展，人们可以随时随地通过电话进行交流，其中语音压缩编码技术发挥着重要的作用。上述这些应用领域构成了语音信号处理技术的主要研究内容。

语音信号处理是语音学与数字信号处理技术相结合的交叉学科，它和认知科学、心理学、语言学、计算机科学、模式识别和人工智能等学科联系紧密。语音信号处理技术的发展依赖于这些学科的发展，而语音信号处理技术的进步也会促进这些学科的进步。

1.1 语音信号处理的发展

语音信号处理的研究工作最早可以追溯到1876年贝尔发明的电话，该电话首次用声电、电声转换技术实现了远距离的语音传输。1939年Dudley研制成功第一个声码器，从此奠定了语音产生模型的基础，这一工作在语音信号处理领域具有划时代的意义。1947年贝尔实验室发明了语谱图仪，将语音信号的时变频谱用图形表示出来，为语音信号分析提供了一个有力的工具。1948年美国Haskins实验室研制成功“语图回放机”，该回放机把手工绘制在薄膜片上的语谱图自动转换为语音，并进行语音合成。共振峰合成方法就是源于这一思想。

对语音识别的研究相对较晚，始于20世纪50年代。语音识别技术的根本目的是研究出一种具有听觉功能的机器，使机器能接受人类的语音，理解人的意图。由于语音识别本身所固有的难度，人们提出了各种限制条件下的研究任务，并由此产生了不同的研究领域。这些领域包括：针对说话人，可分为特定说话人语音识别和非特定说话人语音识

别；针对词汇量，可划分为小词汇量、中词汇量和大词汇量的识别；按说话方式，可分为孤立词识别和连续语音识别等。最简单的研究领域是特定说话人、小词汇量、孤立词的识别，而最难的研究领域是非特定说话人、大词汇量、连续语音的识别。

1952 年贝尔实验室的 Davis 等人研制了特定说话人孤立数字识别系统。该系统利用每个数字元音部分的频谱特征进行识别。1956 年 RCA 实验室的 Olson 等人也独立地研制出 10 个单音节词的识别系统，系统采用从带通滤波器组获得的频谱参数作为语音的特征。1959 年 Fry 和 Denes 等人尝试构建音素识别器来识别 4 个元音和 9 个辅音，并采用频谱分析和模式匹配来进行识别决策。其突出的贡献是，用英语音素序列中可以利用的统计信息来改进包含多个音素的词中音素的精度。与此同时，MIT 林肯实验室的 Forgie 等人，研究了 10 个元音的识别，并采用了声道的时变估计技术。

20 世纪 60 年代初期，日本的很多研究者开发了相关的特殊硬件来进行语音识别，例如东京无线电研究实验室 Suzuki 等人研制的通过硬件来进行元音识别的系统。在此期间，有很多研究工作开始进行，这些研究工作对随后近 20 年的语音识别研究产生了很大的影响。RCA 实验室的 Martin 等人在 20 世纪 60 年代末开始研究语音信号时间尺度不统一的解决办法，开发了一系列的时间归正方法，明显地改变了识别性能。与此同时，苏联的 Vintsyuk 提出了采用动态规划方法解决两个语音的时间对准问题。尽管这是动态时间弯折算法 DTW(dynamic time warping)的基础，也是其连接词识别算法的初级版，但 Vintsyuk 的研究并不为学术界的广大研究者所知道，直到 20 世纪 80 年代大家才知道 Vintsyuk 当初的研究，而这时 DTW 方法已广为人知。

值得一提的是，20 世纪 60 年代中期，美国斯坦福大学的 Reddy 就开始尝试用动态跟踪音素的方法来进行连续语音的识别。后来，Reddy 加盟到卡内基·梅隆大学，多年来在连续语音识别上开展了卓有成效的工作，直至现在仍然在连续语音识别方面居于领先地位。

20 世纪 70 年代之前，语音识别的研究特点是以孤立词的识别为主。到 20 世纪 70 年代，语音识别研究取得了诸多的成就，首先在孤立词识别方面，由日本学者 Sakoe 给出了使用动态规划方法进行语音识别的途径——DTW 算法，DTW 算法是把时间归正和距离测度计算结合起来的一种非线性归正技术。这是语音识别中一种非常成功的匹配算法，当时在小词汇量的研究中获得了成功，从而掀起了语音识别的研究热潮。Itakura 基于语音编码中广泛使用的线性预测编码(linear predictive coding，LPC)技术，通过定义基于 LPC 频谱参数的合适的距离测度，成功地将其扩展应用到语音识别中。以 IBM 为首的一些研究单位还着手开展了连续语音识别的研究，AT&T 的贝尔实验室也开展了一系列非特定说话人语音识别方面的研究工作。

应该指出的是，从 20 世纪 70 年代起人工智能技术开始被引入到语音识别中来。美国国防部的高级研究规划局 ARPA(advanced research projects agency)组织了有卡内基·梅隆大学等 5 个单位参加的一项大规模语音识别和理解的研究计划，当时专家们认为：要使语音识别研究获得突破性进展，必须让计算机像人那样具有理解语言的智能，而不必过多地在孤立词识别上下功夫。在这个历时 5 年的庞大的研究计划中，最终在语言理解、语言的统计模型等方面积累了经验，其中卡内基·梅隆大学完成的 Hearsay-Ⅱ和 Harpy 两个系统效果最好。在这两个系统中，引用了“黑板模型”来完成底层和顶层之

间不同层次的信息交换和规则调用，成为以后其他专家系统研究工作中的一种规范。但从整体上看，这个计划并没有取得突破性的进展。

20 世纪 70 年代末 80 年代初，Linda、Buzo、Gray 等人解决了矢量量化（vector quantization）码本生成的方法，并将矢量量化技术成功地应用到语音编码中。从此矢量量化技术不仅在语音识别、语音编码和说话人识别等方面发挥了重要的作用，而且很快被推广应用到其他领域。这一时代，语音识别的研究重点之一是连接词识别，典型的工作是进行数字串的识别。研究者提出了各种连接词语音识别算法，大多数工作是基于对独立的词模板进行拼接来进行匹配的方法，如两级动态规划识别算法、分层构筑（level building）、帧同步（frame synchronous）分层构筑方法等。这些方法都有各自的特点，广泛用于连接词识别当中。

20 世纪 80 年代开始，语音识别研究的一个重要进展，就是识别算法从模式匹配技术转向基于统计模型的技术，更多地追求从整体统计的角度来建立最佳的语音识别系统。隐马尔可夫模型（hidden Markov model，HMM）技术就是其中的一个典型技术。尽管开始的时候只有较少的单位采用这种模型，但由于该模型能很好地描述语音信号的时变性和平稳性，具有把从声学-语言学到句法等统计知识全部集成在一个统一框架中的优点，因此从 20 世纪 80 年代起，它被广泛地应用到语音识别研究中。直到目前为止，HMM 方法仍然是语音识别研究中的主流方法。HMM 的研究使大词汇量连续语音识别系统的开发成为可能。20 世纪 80 年代末，美国卡内基·梅隆大学用 VQ/HMM 实现了 997 个词的非特定人连续语音识别系统 SPHINX，这是世界上第一个高性能的非特定人、大词汇量、连续语音识别系统。此外，BBN 的 BYBLOS 系统，林肯实验室的识别系统等也都具有很好的性能。这些研究工作开创了语音识别的新时代。

从 20 世纪 80 年代后期和 90 年代初期开始，人工神经网络（artificial neural network，ANN）的研究异常活跃，并且也被应用到语音识别的研究中。进入 90 年代后，相应的研究工作在模型设计的细化、参数的提取和优化以及系统的自适应技术等方面取得了一些关键性的进展，这使语音识别技术进一步成熟，并且出现一些很好的产品。许多发达国家，如美国、日本、韩国，以及 IBM、Microsoft、Apple、AT&T、NTT 等著名公司都为语音识别系统的实用化开发研究投以巨资。

当今基于 HMM 和 ANN 相结合的方法得到了广泛的重视。而一些模式识别、机器学习方面的新技术也被应用到语音处理过程中，如支持矢量机（support vector machine）技术、进化计算（evolutionary computation）技术等。

支持矢量机是一种通用的机器学习方法，它的最大优点就是在小样本情况下依然可以保持很好的推广泛化能力，这是传统的机器学习方法所不具备的。传统的学习方法都是通过最小化经验风险来使期望风险最小化，因此分类性能受训练样本数目的影响很大，在小样本情况下的表现很不尽如人意。由于支持矢量机所表现出的优良性能，近年来它已广泛地应用于各个研究领域。在语音信号处理领域，已有学者尝试采用这种方法来进行端点检测、语音分类和说话人识别等方面的研究。

进化计算是当前人工智能、机器学习等领域广为关注的方法，它把自然界中进化的机制，比如变异、自然选择等引入到搜索、优化或机器学习中来，从而改进了算法的全局搜索

能力。在一些研究中已经采用这种方法来改进 HMM 技术中 B-W 算法的全局搜索能力。由于 B-W 算法是一种基于梯度的算法，其训练结果依赖于初始值的选择，因而往往只能得到初始值附近的一个局部最优解。通过引入进化计算可以改进训练模型的性能，最终提高系统的识别率，这方面的工作已成功地应用于孤立词识别和说话人识别中。

从语音识别研究的进展来看，国际上孤立词识别系统的词汇量已经扩大到几万，特定说话人或非特定说话人的连续语音识别系统已经达到了很高的识别率。从研究领域上看，在连续语音流中识别出关键词的研究，以及利用语音处理技术进行多种语言之间的自动翻译系统的研究(如著名的 CSTAR 项目)是当前较热门的课题。以电话语音识别、自动声讯信息查询、语音信息检索、自动誊写、语音自动文摘等为代表的应用研究异常活跃；以语音产生的非线性模型、实际环境下的顽健(robust)语音识别等为代表的基础研究是当前的研究热点。随着网络技术和无处不在的移动计算技术的迅速发展，出现了网络环境下的语音识别技术、嵌入式和计算资源有限时的语音识别技术、语种识别技术、基于语音的情感处理技术等一些新的研究方向。

在国内，20 世纪 50 年代末就有人尝试用电子管电路进行元音识别，而到了 70 年代才由中国科学院声学所开始进行计算机语音识别的研究。在此之后，有关专家也开始撰文介绍这方面的工作。从 80 年代开始，很多单位陆续参加到这一行列中来，它们纷纷采用不同的方法，开展了从最初的特定说话人、小词汇量孤立词识别，到非特定说话人、大词汇量连续语音识别的研究工作。80 年代末，以汉语全音节识别作为主攻方向的研究已经取得了相当大的进展，一些汉语语音输入系统已向实用化迈进。四达技术开发中心、星河公司等相继推出了相应的实际产品。清华大学、中国科学院声学所在无限词汇的汉语听写机的研制上获得成功。90 年代初，四达技术开发中心又与哈尔滨工业大学合作推出了具有自然语言理解能力的新产品。在国家“863”计划的支持下，清华大学和中国科学院自动化所等单位在汉语听写机原理样机的研制方面开展了卓有成效的研究。北京大学在说话人识别方面也做了很好的研究。

近年来，随着改革开放的不断进行，我国的国际地位与日俱增，汉语语音识别越来越受到重视，国外很多著名的公司，如 Microsoft、IBM、Motorala、Intel 等都在国内设立了研发机构，并且都将汉语语音识别作为主攻方向之一。IBM 公司于 1997 年推出了汉语连续语音识别系统 Via Voice，输入速度平均每分钟可达 150 字，平均最高识别率达到 95%，并具有“自我”学习的功能。在 2000 年发布的 Via Voice 千禧版中，用户可以通过语音导航到计算机桌面及浏览网页。1998 年微软投资 8000 万美元在中国筹建微软中国研究院(2000 年更名为微软亚洲研究院)，其开发的重点方向之一就是语音识别。1998 年 Intel 公司提出了基于 Intel 构架发展语音技术的构想，向软件开发厂商提供包括信号处理库、识别库、图像处理库在内的高性能语音函数库支持，1999 年 Intel 公司又和 L&H 公司合作，推出语音识别软件开发包 Spark3.0，其中包括 Spark 语音识别引擎和软件开发工具箱。微软也推出了基于 .net 的语音识别引擎。

尽管语音识别技术研究已经取得了很大的成绩，但到目前为止，离广泛的应用尚存在距离。很多因素影响着语音识别系统的性能，例如实际环境中的背景噪声、传输通道的频率特性、说话人生理或心理情况的变化以及应用领域的变化等都会导致语音识别系统性

能的下降，甚至使系统不能工作。研究语音识别系统顽健性(robustness)问题受到了研究者的广泛重视，国内外很多单位都开展了很好的研究。但到目前为止，所做的工作大都是针对某一种或两种影响因素进行补偿的研究，综合考虑各种影响因素补偿方法的研究还很少。

语音识别通常是指能识别出相应的语音内容，除此之外，它还有一种特殊的形式——说话人识别。说话人识别不必识别出语音信号的具体内容，而只要鉴别出该语音是哪个说话人发出的即可。从实现的技术手段上看，说话人识别和语音识别一样，都是通过提取语音信号的特征，并建立相应的参考模板来进行分类判断。说话人识别问题，最初是在第二次世界大战期间，美国国防部向贝尔实验室提出的课题。目的是根据窃听到的电话语音来判断说话人是哪一位德军高级将领，这对分析当时的德军战略部署具有重要的意义。该项目持续进行了 3 年，但并未达到预期的效果。

说话人识别研究的早期工作，主要集中在人耳听辨实验和探讨听音识别的可能性方面。随着语音识别研究的不断深入，说话人识别研究也获得了突飞猛进的发展。语音识别中很多成功的技术都可以应用到说话人识别中。目前国外已经有了一些成熟的产品。例如 AT&T 公司应用说话人识别技术研制出了智慧卡，并已应用于自动提款机。欧洲电信联盟在电信与金融结合领域应用说话人识别技术，于 1998 年完成了 CAVE 计划，在电信网上进行说话人识别。近年来，说话人识别技术应用最为成功的例子是在伊拉克战争期间，萨达姆在电视上发表讲话后，美国 FBI 宣称讲话者不是萨达姆本人，而德国的科学家应用说话人识别技术证实讲话者确实是萨达姆。从后来的情况看，德国科学家的判断是正确的。随着 Internet 的发展，网络环境下的说话人识别技术日益受到重视，已成为当今的一个研究热点。

就语音合成技术而言，最早的语音合成器是 1835 年由 W. von Kempelen 发明，经 Weston 改进的机械式会讲话的机器。该机器完全模拟人的发音生理过程，用风箱模拟来自肺部的空气动力。气流通过特别设计的哨时会产生语音中的辅音；气流通过形状可以变化的模拟口腔的软管时会产生元音。风箱、哨和软管 3 部分机械配合起来就可以产生一些音节和词。这是一个相当完善的机械式语音合成器。最早的电子式语音合成器是前面提到的 1939 年 Dudley 发明的声码器，它不是机械地模仿人发音的生理过程，而是通过电子线路来实现基于语音产生的源/滤波器理论。其中声源包括产生清音的噪声源和产生浊音的周期脉冲声源，它们分别用噪声发生器和张弛振荡器来实现，而声道的滤波作用是通过电子通道滤波器来实现的，滤波器的中心频率是用键盘上的 10 个琴键来控制。

现代的语音合成器都是利用计算机来实现的。从 20 世纪 70 年代末开始，出现了对文语转换(text to speech)系统的研究，其特点是用最基本的语音单元，如音素、双音素、半音节或音节作为合成单元，建立语音库，通过合成单元拼接而达到无限词汇的合成。为了保证合成声音具有良好的音质，在这种系统中除语音库外，还有一个相当庞大的规则库，以实现对合成语音的音段特征和超音段特征的控制。20 世纪 80 年代，由 D. Klatt 设计的串并联混合型共振峰合成器是 20 世纪最有代表性的工作。该合成器可以设置和控制多达 8 个共振峰，可模拟发音过程中的声道共振，而且还设有单独的滤波器，以模拟鼻腔和气管的共振。其中元音和浊辅音的产生用串联通道来实现，清辅音的产生用并联通道

来实现。此外，这种合成器还可以对声源做各种选择和调整，以模拟不同的嗓音。它共产生 7 种不同音色的语音，包括模拟不同年龄、性别和个性的说话人的语音。瑞典皇家理工学院 Fant 实验室在多语种文语转换系统研究方面也做出了突出的成绩，完成了英语、法语、瑞典语、西班牙语和芬兰语的文语转换系统。

20 世纪 90 年代末，日本的研究者提出了一种多样本、不等长语音拼接合成技术 PSOLA。PSOLA 在语音库中存放了大量的真人语音样本，通过选择合适的拼接语音片段来实现高质量的合成语音。在这项技术中，语音合成问题被简化为如何建立一个在语音学上充分覆盖的语音库，如何从语音库中选择合适的语音片段来拼接，以及如何对语音片段之间的拼接部分做适当的调整。

目前，有限词汇的语音合成器已经在自动报时、报警、报站、电话查询服务、智能玩具等方面得到了广泛的应用。从语音合成技术采用的方法上看，大体上可以分为：基于发音模型的合成、波形编码合成和基于语音库的合成。从研究进展上看，很多语音合成系统都具有很高的可懂度，但自然度还不尽如人意。提高语音合成的自然度是当今研究的热点。

我国的语音合成研究是从 20 世纪 80 年代开始的，声学所、自动化所、社科院语言所较早地开展了这方面的工作。早期的工作主要是参数合成，尤其是共振峰合成及线性预测合成。20 世纪 90 年代初开始，真实语音的波形拼接技术最早由清华大学应用到汉语合成中来，合成的语音清晰度明显好于参数合成。之后声学所将可以调节韵律参数的波形拼接合成技术 PSOLA 引入汉语合成，并提出了一套韵律控制方法，使合成语音的质量有突破性的提高。当前的汉语语音合成系统中，音质比较好的基本上都是采用波形拼接合成技术，如清华大学、中国科技大学、微软亚洲研究院、IBM 中国研究中心、摩托罗拉中国研究中心等完成的系统，尤其是中国科技大学在国家“863”计划支持下开发出的文语转换系统已投放市场。

就语音编码技术而言，它的研究也是始于 1939 年 Dudley 发明的声码器，但是直到 70 年代中期，除了脉冲编码调制(pulse coding modulation，PCM)和自适应差分脉冲编码调制(ADPCM)取得较好的进展之外，中、低比特率语音编码一直没有大的突破。自 70 年代起，国外就开始研究计算机网络上的语音通信，当时主要是基于 Arpanet 网络平台进行的研究和实验。1974 年首次分组语音实验是在美国西海岸南加州大学的信息科学研究所和东海岸的林肯实验室之间进行，语音编码为 9.6 Kbps 的连续可变斜率增量调制。1974 年 12 月线性预测编码(LPC)声码器首次用于分组语音通信实验，数码率为 3.5 Kbps。1975 年 1 月又首次在美国实现使用 LPC 声码器的分组语音电话会议。1977 年 Internet 工程任务组(internet engineering task force，IETF)颁发了关于分组话音通信协议的讨论文件 RFC741。20 世纪 80 年代的研究主要集中在局域网上的语音通信，因为 70 年代后期已推出带宽可达 Mbps 量级的价格较为低廉的以太网。最早的实验是由英国剑桥大学于 1982 年在 10 Mbps 的剑桥环形网上进行的。其后，意大利、美国、英国等许多国家的研究者在总线型局域网、令牌环网、3Com 以太网上进行实验，深入研究了分组时延的原因、分组语音通信协议、链路利用率和语音分组同步等问题，并试制了电话网和局域网的接口模块。1980 年美国政府公布了一种 2.4 Kbps 的线性预测编码标准算法 LPC-10，这使得在普通电话带宽信道中传输数字电话成为可能。1988 年美国又公布了一

个 4.8 Kbps 的码激励线性预测编码(CELP)语音编码标准算法，欧洲推出了一个 16 Kbps 的规则脉冲激励(RELP)线性预测编码算法，这些算法的音质都能达到很高的质量，而不像单脉冲 LPC 声码器的输出语音那样不为人们所接受。进入 20 世纪 90 年代，随着 Internet 在全球范围内的兴起和语音编码技术的发展，IP 分组语音通信技术获得了突破性的进展和实际应用。最初的应用只是在网络游戏等软件包中传送和存储语音信息，对语音质量要求低，相当于机器人的声音效果。其后计算机厂商纷纷推出对等方式或客户服务器方式语音通信免费软件，这些软件利用计算机中的声卡对语音进行打包传送，对语音一般不进行压缩。20 世纪 90 年代中期开始，有关厂商开始开发用于局域网语音通信的网关产品，实现局域网内 PC 间的语音通信以及经 PBX 和外界电话的通信，但这些产品都采用内部协议规范。20 世纪 90 年代中期还出现了很多被广泛使用的语音编码国际标准，如数码率为 5.3/6.4 Kbps 的 G.723.1、数码率为 8 Kbps 的 G.729 等。此外，也存在着各种未形成国际标准，但数码率更低的成熟的编码算法，有的算法数码率甚至可以达到 1.2 Kbps 以下，但仍能提供可懂的语音。

由于语音编码产品化的过程相对比语音识别容易些，因此其研究成果能很快转向实际应用，对通信事业的发展起了重要的推动作用。

1.2　语音信号处理的应用及新方向

语音信号处理技术是计算机智能接口与人机交互的重要手段之一。就语音识别技术而言，其基本任务是将输入语音转化为相应的文本或命令。语音识别的市场前景广泛，在一些应用领域中正迅速成为一个关键的具有竞争力的技术。例如在声控应用中，计算机识别输入的语音内容，并根据内容来执行相应的动作，这包括声控电话转换、声控语音拨号系统、声控智能玩具、信息网络查询、家庭服务、宾馆服务、旅行社服务系统、医疗服务、银行服务、股票查询服务和工业控制等。语音识别也可用于将文字以口授的方式输入到计算机中，即广泛开展的听写机研究，如声控打字机等。语音识别技术还可以用于自动口语翻译，即通过将口语识别技术、机器翻译技术、语音合成技术等相结合，可将一种语言输入的语音翻译为另一种语言的语音输出，实现跨语言的交流，例如目前美国、日本、欧洲，包括中国科学院自动化所参加的 CSTAR 计划，正在重点开展多语种口语自动翻译研究。随着计算技术的发展，各种移动计算设备、可穿戴计算设备日益增多，这些设备，其体积越来越小，并且要求在行走或驾驶时进行信息的输入，而传统的键盘输入方式已不能满足其方便、自然及随时随地、有效地输入信息的需要，采用语音识别技术可以解放用户的手眼，有效地改变人机交互手段。如目前在一些手持计算机、手机等嵌入式电子产品上已经使用语音识别技术来进行控制。

对说话人识别技术，近年来已经在安全加密、银行信息电话查询服务等方面得到了很好的应用。此外，在公安机关破案和法庭取证方面也发挥着重要的作用。

就语音合成而言，它已经在许多方面得到了实际应用，发挥了很好的社会效益，如公共交通中的自动报站、各种场合的自动报时、自动告警、电话自动查询服务和文本校对中的语音提示等。在电信声讯服务领域的智能电话查询系统中，采用语音合成技术可以弥