电子信息与电气学科规划教材·电子信息科学与工程类专业

数字语音处理

及MATLAB仿真

张雪英　编著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY　http://www.phei.com.cn

电子信息与电气学科规划教材·电子信息科学与工程类专业

数字语音处理及MATLAB仿真

张雪英　编著

電子工業出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书系统地阐述了语音信号处理的原理、方法、技术和应用，同时给出了部分内容对应的 MATLAB 仿真源程序。全书共 12 章，第 1 章至第 7 章是基本理论部分，包括语音信号的数字模型、语音信号的短时时域分析和频域分析、语音信号的同态处理、语音信号线性预测分析和矢量量化；第 8 章至第 12 章是应用部分，包括语音编码、语音合成、语音识别、语音增强和语音处理的实时实现。本书内容全面，重点突出，原理阐述深入浅出，注重理论与实际应用的结合，可读性强。

本书可作为高等院校通信工程、电子信息工程、自动控制、计算机技术与应用等专业高年级本科生相关课程的教材，也可供从事语音信号处理研究的研究生和科研人员参考。

图书在版编目(CIP)数据

数字语音处理及 MATLAB 仿真 / 张雪英编著. —北京：电子工业出版社，2010.7
电子信息与电气学科规划教材·电子信息科学与工程类专业
ISBN 978-7-121-11323-9

Ⅰ. ①数… Ⅱ. ①张… Ⅲ. ①语音数据处理－计算机仿真－软件包，MATLAB－高等学校－教材 Ⅳ. ①TN912.3

中国版本图书馆 CIP 数据核字(2010)第 131775 号

策划编辑：凌　毅
责任编辑：李秦华
印　　刷：涿州市京南印刷厂
装　　订：涿州市桃园装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：787×1092　1/16　印张：16　字数：410 千字
印　　次：2010 年 7 月第 1 次印刷
印　　数：4000 册　定价：28.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系。联系及邮购电话：(010)88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：(010)88258888。

前　言

语言是人类交换信息最方便、最快捷的一种方式，在高度发达的信息社会中，用数字化的方法进行语音的传送、存储、识别、合成和增强等是整个数字化通信网中最重要、最基本的组成部分之一。随着人类步入信息社会步伐的加快，越来越多的地方需要用到语音信号处理知识。语音信号处理作为一门涉及面很广的交叉学科，已经在越来越多高校中的通信工程、电子信息工程、自动控制、计算机技术与应用等专业开设这门课程，语音信号处理的教材也逐渐增多。但目前已有的教材中，基本上以理论阐述为主，内容较深，更适合硕士、博士研究生层次的学生使用，不太适合本科生。

根据教育部关于加强本科生动手实践能力培养的要求，编写本书的目的就是要让本科生通过本课程的学习，了解这门课程的基本理论，同时学会用 MATLAB 语言来处理实际的语音，提高学习本课程的兴趣，培养解决实际问题的能力。

本书最大的特色就是把理论与实际相结合，其中融入了编著者多年从事语音信号处理的科研成果。在阐述基本理论的同时，辅以 MATLAB 源程序，加上详细注释，并配有程序运行结果图。学生们在课后，可以自己动手用工具软件录制语音，照着编写程序，按自己的意图修改程序，进行语音处理的实践。克服以前学生分别学完 MATLAB、语音信号处理课程后，当要用 MATLAB 具体处理一句实际语音时，却不知从何入手的缺陷。

本书主要以高年级本科生和初次学习语音信号处理知识的研究生为读者对象，注重语音信号处理基础知识及主要应用的描述，同时对本领域的最新成果也有简单介绍。全书共 12 章，第 1 章是绪论；第 2 章是语音信号的数字模型；第 3 章是语音信号的短时时域分析；第 4 章是语音信号短时频域分析；第 5 章是语音信号的同态处理；第 6 章是语音信号线性预测分析；第 7 章是矢量量化；第 8 章是语音编码；第 9 章是语音合成；第 10 章是语音识别；第 11 章是语音增强；第 12 章是语音处理的实时实现。附录部分是本书中出现过的专业名词缩写及中英文对照，供大家学习时参考。本书第 1 章至第 7 章属于基本理论部分，所附的 MATLAB 程序较多，第 8 章至第 12 章是语音信号处理技术的应用，这方面的程序一般都比较长，且有一定难度，所以附带的程序较少，且都是相对简单的。可以说，本书是一本关于语音信号处理的入门实践教材，在学习和掌握本书内容的基础上，再进行本专业更深层次的学习是合适的。

本书前 7 章内容可以用做工科高等院校相关专业 32～40 学时课程的教程，后 5 章内容可作为选学内容。

本书提供免费的电子课件、MATLAB 仿真程序，读者可登录华信教育资源网 www.hxedu.com.cn，注册后免费下载。

本书由张雪英教授编著，马建芬副教授和李凤莲博士参编，具体分工是：第 1 章、第 2 章、第 3 章、第 4 章、第 5 章、第 7 章、第 8 章、第 9 章、第 10 章、第 12 章由张雪英编写，第 6 章和附录 A 由李凤莲编写，第 11 章由马建芬编写。在本书编写过程中，特别是 MATLAB 程序的调试过程中，得到了太原理工大学信息工程学院电路与系统专业一些硕士生和博士生的帮助，在此表示衷心感谢。

由于编著者水平有限，书中难免存在错误之处，敬请读者批评指正。

编著者

2010 年 3 月

目　录

第1章 绪 论

1.1 概 述

语言是人类交换信息最方便、最快捷的一种方式，在高度发达的信息社会中，用数字化的方法进行语音的传送、存储、识别、合成和增强等是整个数字化通信网中最重要、最基本的组成部分之一。数字电话通信、高音质的窄带语音通信系统、语言学习机、声控打字机、自动翻译机、智能机器人、新一代计算机语音智能终端及许多军事上的应用等，都要用到语音信号处理技术，随着集成电路和微电子技术的飞速发展，语音信号处理系统逐步走向实用化。

语音信号处理是一门新兴的边缘学科，它是语音学与数字信号处理两个学科相结合的产物。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有着紧密的联系。语音信号处理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领域的进步。

语音信号处理的目的是要得到某些语音特征参数以便高效地传输或存储；或者是通过某种处理运算以达到某种用途的要求，例如人工合成语音、辨识出讲话者、识别出讲话的内容等。

随着现代科学和计算机技术的发展，除了人与人之间的自然语言的通信方式之外，人机对话及智能机器等领域也开始使用语言。这些人工语言同样有词汇、语法、语法结构和语义内容等。控制论创始人维纳在1950年就曾指出过："通常，我们把语言仅仅看做人与人之间的通信手段，但是，要使人向机器、机器向人及机器向机器讲话，那也是完全办得到的"。通常认为，语音信息的交换大致上可以分为三大类：

① 人与人之间的语言通信：包括语音压缩与编码、语音增强等。

② 第一类人机语言通信问题，指的是机器讲话、人听话的研究，即语音合成。

③ 第二类人机语言通信问题，指的是人讲话、机器听话的情况，即语音识别和理解。

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。

1.2 语音信号处理的发展

早在一两千年以前，人们便对语言进行了研究。由于没有适当的仪器设备，长期以来，一直是由耳倾听和用口模仿来进行研究。因此，这种语言研究常被称为"口耳之学"，所以对语音只是停留在定性的描写上。

语音信号处理真正意义上的研究可以追溯到1876年贝尔电话的发明，该技术首次用声电、电声转换技术实现了远距离的语音传输。1939年 Homer Dudley 提出并研制成功的第一个声码器，从此奠定了语音产生模型的基础。这一发明在语音信号处理领域具有划时代的意义。19世纪60年代，亥姆霍兹应用声学方法对元音和歌唱进行了研究，从而奠定了语言的声学基础。20世纪40年代，一种语言声学的专用仪器——语谱图仪问世了。它可以把语音的

时变频谱用语图表示出来，从而得出了“可见语言”。1948 年美国 Haskins 实验室研制成功“语音回放机”，该仪器可以把手工绘制在薄膜片上的语谱图自动转换成语音，并进行语音合成。20 世纪 50 年代对语言产生的声学理论开始有了系统的论述。随着计算机的出现，语音信号处理的研究工作得到了计算机技术的帮助，使得过去受人力、时间限制的大量的语音统计分析工作，得以在电子计算机上进行。在此基础上，语音信号处理不论在基础研究方面，还是在技术应用方面，都取得了突破性的进展。下面分别论述语音信号处理的三个主要分支（语音合成技术、语音编码和语音识别技术）的发展和现状。

1.2.1 语音合成

就语音合成技术而言，最早的合成器是 1835 年由 W. von Kempelen 发明，经 Weston 改进的机械式会讲话的机器。该机器完全模仿人的发音生理过程，分别用风箱、特别设计的哨和软管来模拟肺部的空气动力、模拟口腔。而最早的电子式语音合成器是 1939 年 Homer Dudley 发明的声码器，它不是简单地模拟人的生理过程，而是通过电子线路来实现基于语音产生的源－滤波器理论。

但真正具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，主要是采用计算机产生高清晰度、高自然度的连续语音。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，1973 年 Holmes 发明的并联共振峰合成器和 1980 年 Klatt 发明的串/并联共振峰合成器，只要精心调整参数，这两个合成器都能合成出比较自然的语音。最具代表性的文语转换系统是美国 DEC 公司 1987 年开发的 DECtalk。但是，由于准确提取共振峰参数比较困难，虽然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难以达到文语转换（TTS）系统的实用要求。

自 20 世纪 80 年代末期至今，语音合成技术又有了新的进展，特别是 1990 年提出的基音同步叠加（PSOLA）方法，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 LPC 方法或共振峰合成器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，有很大的商用前景。

我国的汉语语音合成研究起步较晚，但从 20 世纪 80 年代初就基本上与国际研究同步发展。大致也经历了共振峰合成、LPC 合成到应用 PSOLA 技术的过程。在国家 863 计划、国家自然科学基金委员会、国家攻关计划、中国科学院有关项目等支持下，汉语文语转换系统研究近年来取得了令人瞩目的进展，其中不乏成功的例子，如 1993 年中国科学院声学所研制的 KX-PSOLA，1995 年研制的联想佳音；清华大学在 1993 年研制的 TH_SPEECH；1995 年中国科技大学研制的 KDTALK 等系统。这些系统基本上都采用基于 PSOLA 方法的时域波形拼接技术，其合成汉语普通话的可懂度、清晰度达到了很高的水平。然而同国外其他语种的文语转换系统一样，这些系统合成的句子及篇章语音机器味较浓，其自然度还不能达到用户可广泛接受的程度，从而制约了这项技术大规模进入市场。

现阶段语音合成的最大进展是已经能够实时地将任意文本转换成连续可懂的自然语句输出。文语转换使得数据通信和语音通信在终端一级实现交融，人们将有望在获取 Internet 信息时，使短消息服务、电子邮件等多数以文本方式提供的信息也能用语音的方式输出。语音合成技术经历了从参数合成到拼接合成，再到两者的逐步结合，其不断发展的动力是人们认知水平和需求的提高。

1.2.2 语音编码

语音编码的目的就是在保证一定语音质量的前提下，尽可能降低编码比特率，以节省频率资源。语音编码技术的研究开始于1939年军事保密通信的需要，贝尔电话实验室的Homer Dudley提出并实现了在低带宽电话电报电缆上传输语音信号的通道声码器，成为语音编码技术的鼻祖。直到20世纪70年代，国际电联(ITU-T，原CCITT)于1972年发布了64kbit/s脉冲编码调制(PCM)语音编码算法的G.711建议，它被广泛应用于数字通信、数字交换机等领域，从而占据统治地位。1980年美国政府公布了一种2.4kbit/s的线性预测编码标准算法LPC-10，这使得在普通电话带宽中传输数字电话成为可能。ITU-T也于20世纪80年代初着手研究低于64kbit/s的非PCM编码算法，并于1984年通过了32kbit/s ADPCM语音编码G.721建议，它不仅可以达到与PCM相同的语音质量，而且具有更优良的抗误码性能。1988年美国又公布了一个4.8kbit/s的码激励线性预测(CELP)编码算法。与此同时，欧洲也推出了一个16kbit/s的规则脉冲激励线性预测(RPE-LPC)编码算法。这些算法的语音质量都能达到较高的水平，大大超过LPC声码器的质量。进入20世纪90年代，随着因特网在全球范围的兴起，人们对能在网络上传输语音的VoIP技术兴趣大增，由此，IP分组语音通信技术获得了突破性进展和实际应用。ITU-T于1992年公布了16kbit/s低延迟码激励线性预测编码(LD-CELP)的G.728建议。它以其较小的延迟、较低的速率、较高的性能在实际中得到广泛的应用，也成为分组化语音通信的可选算法之一。1996年ITU-T发布了码率为5.3/6.4kbit/s的G.723.1标准。在1995年11月ITU-T SG15全会上通过了共轭代数码激励线性预测(CS-ACELP)的8kbit/s语音编码G.729建议，并于1996年6月ITU-T SG15会议上通过G.729附件A：减少复杂度的8kbit/s CS-ACELP语音编解码器，正式成为国际标准。这几种语音编码算法也成为分组化语音通信的可选算法。

语音编码技术主要有两个努力方向：一是中低速率的语音编码的实用化及如何在实用化过程中进一步提高其抗干扰、抗噪声能力；另一个是如何进一步降低其编码速率。目前已能在5～6kbit/s的速率上获得高质量的重建语音，下一个目标则是要在4kbit/s的速率上获得短延时、高质量的重建语音。特别是对中长延时编码，人们正在研究其更低速率(如400～1200bit/s)的编码算法。当编码速率降至2.4kbit/s以下时，CELP算法即使应用更高效的量化技术也无法达到预期的指标，需要其他一些更符合低速率编码要求的算法，目前比较好的算法有正弦变换编码(STC)、混合激励线性预测编码(MELPC)、时频域插值(TFI)编码、基音同步激励线性预测(PSELP)编码等，同时还要求引入新的分析技术，如非线性预测、多精度时频分析技术(包括子波变换技术)、高阶统计分析技术等，这些技术更能挖掘人耳听觉掩蔽等感知机理，更能以类似人耳的特性作为语音的分析与合成，使语音编码系统更接近于人类听觉器官的处理方式工作，从而在低速率语音编码的研究上取得突破。

20世纪90年代中期到现在，第三代移动通信技术逐渐成熟并走向商用，变速率语音编码和宽带语音编码得到了迅速的发展，不断有新的国际标准和地区标准公布。应用于第三代移动通信的变速率语音编码主要有可变速率码激励线性预测(QCELP)、增强型变速率编码器(EVRC)、自适应多速率(AMR)编码器、自适应多速率宽带(AMR-WB)编码器、可选模式声码器(SMV)和变速率多模式宽带(VMR-WB)编码器等。宽带语音的发展也经历了一个过程，1988年国际电联通过了第一个宽带语音编码器标准G.722，基于子带自适应差分脉码调制(SB-ADPCM)编码原理，速率为64kbit/s、56kbit/s和48kbit/s。宽带语音编码器的合成语音

更自然，非常适合应用到电视电话会议中。早期的宽带语音编码器的缺点就是编码效率不高，64kbit/s 的速率不利于在系统中实现。1999 年 ITU-T 公布了新的宽带语音编码国际标准 G.722.1，降低了编码速率(24kbit/s 和 32kbit/s)。2002 年 ITU-T 在对以往宽带语音编码算法改进的基础上提出 G.722.2 标准，由 9 种速率的语音模式组成，编码速率较低，而且可以根据无线环境和本地容量需求动态选择。变速率语音编码理论上仍属于 CELP，但在“变”上有了新的研究，由此引入了相关技术的研究，包括：用来检测语音通信时是否有语音存在的语音激活检测(VAD)技术、为突出“变”字而进行速率判决(RDA)的自适应技术、为避免语音帧丢失后带来负面效应的差错隐藏(ECU)技术、为克服背景噪声不连续的舒适背景噪声生成(CNG)技术等。这些相关技术的应用使变速率语音编码之后的语音合成效果几乎没有降低。随着移动通信的飞速发展，用变速率语音编码来提高频带的有效利用率，将是未来数字蜂窝和微蜂窝网的必然发展趋势。

1.2.3 语音识别

与机器进行语音交流，让机器明白你说什么，这是人们长期以来梦寐以求的事情。而语音识别技术就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高技术。由于语音本身所固有的难度，让机器识别语音的困难在某种程度上就像一个外语不好的人听外国人讲话一样，它和不同的说话人、不同的说话速度、不同的说话内容及不同的环境条件有关。语音信号本身的特点造成了语音识别的困难，这些特点包括多变性、动态性、瞬时性和连续性等。根据在不同限制条件下的研究任务，产生了不同的研究领域。这些领域包括：①根据对说话人说话方式的要求，可以分为孤立字语音识别系统、连接字语音识别系统及连续语音识别系统；②根据对说话人的依赖程度可以分为特定人和非特定人语音识别系统；③根据词汇量大小，可以分为小词汇量、中等词汇量、大词汇量及无限词汇量语音识别系统。

语音识别的研究工作真正开始于 20 世纪 50 年代 AT&T 贝尔实验室的 Audry 系统，它是第一个可以识别 10 个英文数字的语音识别系统。1956 年 RAC 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 等人采用频谱分析和模式匹配来进行识别决策构建音素识别器来辨别 4 个元音和 9 个辅音。同年，MIT 林肯实验室采用声道的时变估计技术研究 10 个元音的识别。

但语音识别的研究真正取得实质性进展，并将其作为一个重要的课题开展则是在 20 世纪 60 年代末。这一方面是因为计算机的计算能力有了迅速的提高，能够提供实现复杂算法的软件、硬件环境；另一方面，数字信号处理理论和算法在当时有了蓬勃发展，从而自 20 世纪 60 年代末开始引起了语音识别的研究热潮。这时期的重要成果是提出了动态规划(DP)和线性预测编码(LPC)分析技术，其中后者较好地解决了语音信号产生模型的问题，对整个语音识别、语音合成、语音分析、语音编码的研究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性进展。在理论上，LPC 技术得到进一步发展，动态时间弯折(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，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(DP)进行语音识别的途径——DP 算法。DP 算法是把时间规整和距离测度计算结合起来的一种非线性规整技术，这是语音识别中一种非常成功的匹配算法，并在小词汇量中获得了成功，从而掀起了语音识别的研究热潮。另外，就是学者 Itakura 基于语音编码中广泛使用的

LPC 技术，通过定义基于 LPC 频谱参数的合适的距离测度，成功地将其应用到语音识别中。同时，以 IBM 为首的一些语音研究单位还着手开展了连续语音识别的研究。

在 20 世纪 70 年代末和 80 年代初，Linda、Buzo、Gray 等人解决了矢量量化码本生成的方法，并将矢量量化成功地应用到语音编码中，从此矢量量化技术很快被推广应用到其他领域。

从 20 世纪 80 年代开始，语音识别研究进一步走向深入，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。HMM 技术就是其中的一个典型技术。最早将 HMM 用于语音识别是 20 世纪 70 年代中期，但对 HMM 的全面研究和大规模应用是 20 世纪 80 年代以后的事情。它受到广泛重视的原因是：马尔可夫链可以用来描述蕴藏于观察数据中的时变特性，这使得它能处理语音信号中常常出现的非平稳特性（即时变特性）。它不仅能用于描述各种不同层次的语音单元，甚至可以描述 VQ 中的任一码字或由声学特征定义的任一种声学单元，并且由小单元模型组成大单元模型［音节（或音素）→单词→句子］。由 Viterbi 解码可得到与语音序列相对应的最佳状态序列，从而得到语音单元的最佳分割，使子词单元的使用非常方便，大大避免了训练和识别时的分割困难，使连续语音识别问题得到解决。随着对 HMM 的深入研究和在语音识别中的需要，许多新的算法产生，如估计、平滑、外插、建立时间模型、话者自适应等，使得这一技术在语音识别中有了更深入的应用。到目前为止，HMM 方法仍然是语音识别研究中的主流方法，并使得大词汇量连续语音识别系统的开发成为可能。在 20 世纪 80 年代末，由美国卡内基梅隆大学用 VQ/HMM 实现 997 个词的非特定人连续语音识别系统 SPHINX 成为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。这些研究开创了语音识别的新时代。

20 世纪 80 年代中期重新开始的人工神经网络（ANN）研究，也给语音识别带来一片新的生机。由于 ANN 具有自组织和自动学习各种复杂分类边界的能力，以及很强的区分能力，使它特别适用于语音识别这一特殊的分类问题。人们将 ANN 和 HMM 在同一语音识别系统中结合使用，即由 ANN 完成静态的模式分类问题，而用 HMM 甚至传统的 DP 来完成时间对准问题。从实验结果来看，这种思想可行而且有效，并能使 ANN 比较容易地用于连续语音识别问题。语音识别常用的 ANN 有：时间延迟神经网络 TDNN、递归神经网络 RNN、自组织神经网络 SONN、学习矢量量化 LVQ 及混合语音识别系统。

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研究投以巨资。在 20 世纪 90 年代初期，开始出现孤立语音的英文听写机系统，在 1997 年开始出现基于说话人自适应的连续语音听写系统，并达到一定的实用化程度。从语音识别的进展来看，国际上孤立词识别系统已经扩大到数万个，特定说话人或非特定说话人的连续语音识别系统已达到了很高的识别率。从研究领域来看，在连续语音中识别关键词的研究以及多种语言之间的自动翻译、语音检索等已成为比较热门的课题。随着网络技术和语音研究工作的迅速发展，出现了语种识别技术、基于语音的情感技术、嵌入式语音识别技术等一些新的研究方向。

在国内，语音识别的研究工作起步于 20 世纪 50 年代，但是除中科院声学所外，大多数单位是 20 世纪 70 年代末及 80 年代初才开始的。到 20 世纪 80 年代末，以汉语全音节识别为主攻方向的研究已经取得相当大的进展，一些汉语输入系统已向实用化迈进。20 世纪 90 年代初，在国家“863 计划”支持下，国家 863 智能计算机专家组为语音识别技术研究专门立项。清华大学与中科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。近些年，在我国科研人员长期艰苦努力下，我国在语音技术研究水平和原型系统开发方面达到了世界级的水平，做出了当之无愧的成果。在中国科学院自动化研究所模式识别国家重点实验室，汉语非特定人、连续语音听写机系统的普通话系统，其错误率可以控制在10%以内的的水平，并具有非常好的自适应功能。尤其是在国内外首创研究开发了汉语自然口语的人机对话系统和汉语到日语、英语的直接语音翻译系统，为在未来发展民族化的语音产业打下了非常坚实的技术基础。清华大学王作英教授提出的非齐次基于段长分布的隐马尔可夫模型(DDBHMM)可以说是对语音识别模型算法的一次重大革新。以此理论为指导所设计的语音识别听写机系统在1994-1998年的全国语音识别系统评测中取得三连冠，从而显示了这一新模型的生命力和在这一研究领域内的领先水平。目前，我国语音识别技术的研究已取得令人瞩目的成绩，其基础研究涉及汉语语音学、听觉模型、人工神经网络、小波变换、分形维数和支持向量机等理论，其研究成果必将推动我国语音识别技术研究迈上新台阶。

1.3 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。从目前和整个信息社会发展趋势看，语音技术有很多的应用。语音技术包括语音识别、说话人的鉴别和确认、语种的鉴别和确认、关键词检测和确认、语音合成、语音编码等，但其中最具有挑战性和最富有应用前景的为语音识别技术。

首先对于说话人识别技术，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，说话人识别技术也在公安机关破案和法庭取证方面发挥着重要的作用。其次对于语音识别技术而言，在一些应用领域中正成为一个关键的具有竞争力的技术。例如，在声控应用中，计算机可识别输入的语音内容，并根据内容来执行相应的动作，这包括了声控电话转换、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、股票查询服务和工业控制等。在电话与通信系统中，智能语音接口正在把电话机从一个单纯的服务工具变成为一个服务的“提供者”和生活“伙伴”；使用电话与通信网络，人们可以通过语音命令方便地从远端的数据库系统中查询与提取有关的信息；随着计算机的小型化，键盘已经成为移动平台的一个很大障碍，想象一下如果手机仅仅只有一个手表那么大，再用键盘进行拨号操作已经是不可能的。再者，语音信号处理还可用于自动口语分析，如声控打字机等。随着计算机和大规模集成电路技术的发展，这些复杂的语音识别系统也已经完全可以制成专用芯片，大量生产。在西方经济发达国家，大量的语音识别产品已经进入市场和服务领域。一些用户交换机、电话机、手机已经包含了语音识别拨号功能，还有语音记事本、语音智能玩具等产品也包含了语音识别与语音合成功能。人们可以通过电话网络用语音识别口语对话系统查询有关的机票、旅游、银行信息，并且取得很好的结果。

就语音合成而言，它已经在许多方面得到了实际的应用并发挥了很大的社会作用。例如，公交汽车上的自动报站、各种场合的自动报时、自动报警、手机查询服务和各种文本校对中的语音提示等。在电信声讯服务中的智能电话查询系统中，采用语音合成技术可以弥补以往通过电话进行静态查询的不足，满足海量数据和动态查询的需求，如股票、售后服务、车站查询等信息；也可用于基于微型机的办公、教学、娱乐等智能多媒体软件，例如语言学习、教学软件、语音玩具、语音书籍等；也可与语音合成技术与机器翻译技术结合，实现语音翻译等。

对于语音编码而言，随着人类社会信息化进程的加快，语音编码技术也正在迅速发展，在移动通信、卫星通信、军事保密通信、信息高速公路和IP电话通信中得到了广泛的应用。例如低速率语音编码技术解决了信道容量问题。光纤通信技术使有线通信的信道容量得到了缓解，但对于信道价格昂贵的卫星通信及线路铺设艰难的边远山区通信，仍希望能在现有信道上得到更大的通信容量。再者由于数字加密技术具有高度可靠性，一般在军事保密通信中采用低速率语音编码器，以便对经过压缩编码后的语音数据进行加密处理，然后在窄带信道上进行传输。个人移动通信、语音存储、多媒体通信、数字数据网(DDN)中也用到语音通信技术。目前语音编码的算法发展较快，它可应用的范围也相当广泛，除了上述应用外，未来的ISDN、卫星通信、移动通信、微波接力通信和信息高速公路以及保密电话等无一例外地都会采用低速率语音编码技术。

随着信息技术的不断发展，尤其是网络技术的日益普及和完善，语音信号处理技术正发挥着越来越重要的作用，并且出现了一些新的方向。

① 基于语音的信息检索。随着网络技术及数字图书馆技术的发展，针对于传统的基于文本信息的检索技术，基于语音识别的信息检索技术正成为当今的研究热点。

② 基于语音识别的广播新闻的自动文摘技术的研究。由于广播、电视中的发音较为标准规范，在识别中避免了说话人发音上的不规范，有利于语音识别系统性能的提高。

③ VoIP技术。它是通过TCP/IP网络，而不是传统的电话网络来传输语音的新的通信方式，通常称为IP电话技术。它是在网络上对压缩的语音数据以数据包的形式进行传输和识别。随着手机、PDA等移动电子设备的发展，嵌入式语音识别算法的研究已逐渐成为研究的热点。

④ 语音训练与校正技术也是近年来语音信号处理的一个重要方向。现在越来越多的人希望掌握其他非母语语言，以便方便地进行交流。因此语言学习机已成为当今外语学习者的有利工具。

⑤ 语种识别。语种识别是近年来新出现的研究方向，它是通过分析处理一个语音片断来判别其所属语音的种类，本质上属于语音识别的研究范畴。

⑥ 基于语音的情感处理研究。在人与人的交流中，除了语音信息外，非语言信息也起着重要的作用。为了使人机交流更自然、更人性化，基于语音的情感处理研究也是非常必要的。

1.4 语音信号处理过程的总体结构

信息加工和处理的一般流程如图1.1所示。

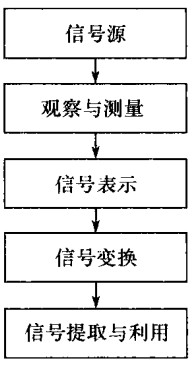

图1.1 信号加工和处理的一般流程

在语音信号的具体情况下，信息源就是说话的人，通过观察和测量得到的就是语音的波形。信号处理包括以下几个内容，首先根据一个给定的模型得到这一信号的表示；然后再用某种高级的变换把这一信号变成一种更加方便的形式；最后一步是信息的提取和使用，这一步可由听者来完成，也可由机器自动完成。

所以，语音信号处理一般有两个任务：第一，它是一种工具，利用它可以得到语音信号的一般表示，这种表示可以用波形表示也可用参数形式表示；第二，把信号从一种形式变换到另一种形式，变换后的表示形式虽然从性质上讲它的普遍性可能小一些，但对某一特殊应用却是更加合适。由此从总体上来看，语音信号处理过程可以用统一的框架来表示，其基本的结构框图如图1.2所示。

从图 1.2 可以看出：无论是语音识别还是语音编码与合成，对于输入的语音信号首先要进行预处理，对信号进行适当的放大和增益控制，并进行反混叠滤波来消除工频信号的干扰；然后进行数字化，将模拟信号转换为便于计算机处理的数字信号；随后对数字语音信号进行分析，提取一定的反映语音信息的参数；最后根据语音信号处理任务的不同，采用不同的处理方法。语音识别技术分为两个阶段：语音识别和训练阶段。在训练阶段，对用特定的参数形式表示的语音信号进行相应的处理，获得表示识别基本单元共性特点的标准数据，以此构成参考模板，并将所有能识别的基本单元的参考模板结合在一起，形成参考模式库；在识别阶段，将待识别的语音信号经特征提取后逐一与参考模型库中的各个模板按某种原则进行比较，找出最相似的参考模板所对应的发音，即为识别结果。对于语音编码技术来说，为了对语音信号进行有效的传输，需要对语音信号以某种算法进行编码，并在接受端进行解压缩。对于语音信号的合成，则是对编码后的信号进行储存。

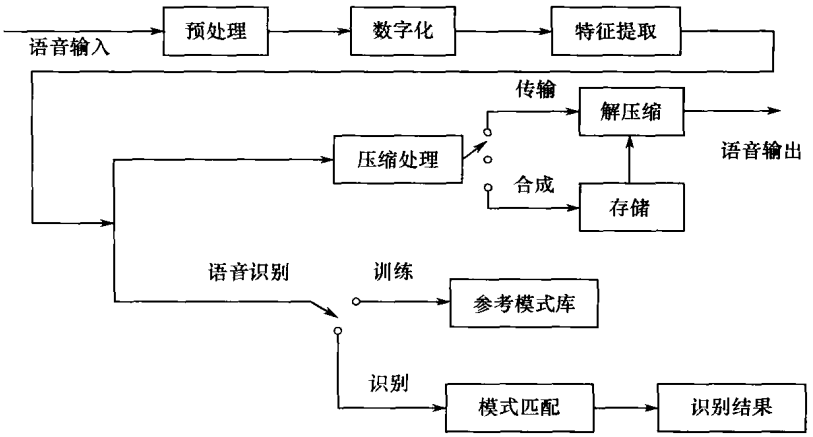

图 1.2　语音处理过程的结构框图

1.5　MATLAB 在数字语音信号处理中的应用

数字语音信号处理是将数字信号处理与语音学相结合，解决现代通信领域中人与人、人与机器之间的信息交流的学科。近几年来语音信号处理学科在世界范围内已取得了飞速的发展，又因为 MATLAB 是一种功能强大、效率高、交互性好的数值计算和可视化计算机高级语言，它将数值分析、信号处理和图形显示有机地融合一体，形成了一个极其方便、用户界面友好的操作环境。随着 MATLAB 的不断发展，其功能越来越强大，广泛应用于数字语音信号处理、数值图像处理、仿真、自动控制、小波分析和神经网络等领域。同时又由于 MATLAB 具有大量的信号处理工具箱并能利用非线性动态系统分析工具 Simulink 等优点，所以近年来 MATLAB 已成为数字信号处理的有利工具，因此也成为学习语音信号处理和进行研究工作的仿真软件工具。

下面简要介绍 MATLAB 在数字语音信号中的几方面应用。

① 通过 MATLAB 可以对数字化的语音信号进行时频域分析。通过 MATLAB 可以方便地展现语音信号的时域及频域曲线，并且根据语音的特性对语音进行分析。例如，清浊音的幅度差别、语音信号的端点、信号在频域中的共振峰频率、加不同窗和不同窗长对信号的影响、