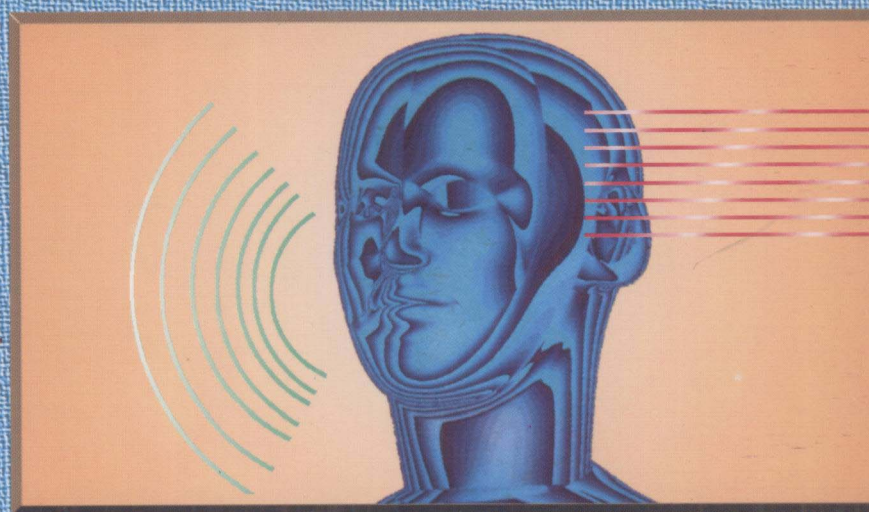
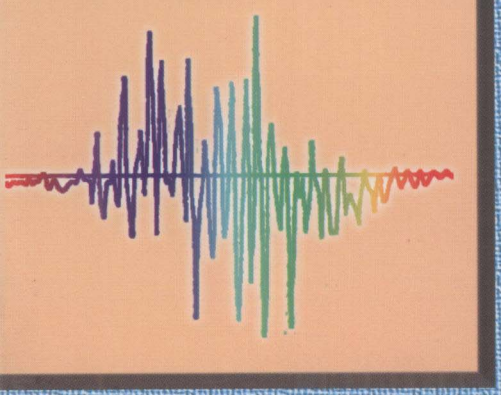
高等学校“十一五”规划教材

SPEECH SIGNAL PROCESSING





语音信号处理

（第4版）

胡 航 编著

哈尔滨工业大学出版社

高等学校“十一五”规划教材

语音信号处理

（第4版）

胡 航 编著

哈尔滨工业大学出版社

内 容 简 介

本书系统地介绍了语音信号处理的基础、概念、原理、方法与应用，以及该学科领域取得的新进展，同时介绍了本门学科的背景知识、发展概况、研究现状、应用前景和发展趋势与方向。既着重基本理论、方法的阐述，又着重新方法和新技术。全书分3篇共17章，其中第1篇语音信号处理基础，包括第1章绪论，第2章语音信号处理的基础知识；第2篇语音信号分析，包括第3章至第9章，介绍语音信号的各种分析方法和技术，包括时域分析、短时傅里叶分析、同态滤波及倒谱分析、线性预测分析、矢量量化技术、隐马尔可夫模型技术以及语音检测分析；第3篇语音信号处理技术与应用，包括第10章至第17章，分别介绍语音编码(1)——波形编码、语音编码(2)——声码器技术及混合编码、语音合成、语音识别、说话人识别、语音增强、神经网络在语音信号处理中的应用及语音信号处理中的一些新兴与前沿技术。

本书物理概念清晰、分析透彻，原理阐述深入浅出、简洁明了，取材广泛、选编得当，内容丰富而新颖，并介绍了本学科领域的一些最新的研究进展；语言通俗易懂、简洁流畅；全书层次分明、条理清晰、结构严谨，并注意各部分内容的有机结合；既有较强的理论系统性，又体现一定应用的观点。

本书可作为高等院校信号与信息处理、通信与电子系统、计算机应用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/胡航编著.—4版.—哈尔滨：哈尔滨工业大学出版社，2009.7

ISBN 978-7-5603-1489-1

Ⅰ.语… Ⅱ.胡… 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(2009)第107427号

责任编辑 张秀华
封面设计 卞秉利
出版发行 哈尔滨工业大学出版社
社　　址 哈尔滨市南岗区复华四道街10号　邮编150006
传　　真 0451-86414749
网　　址 http://hitpress.hit.edu.cn
印　　刷 肇东粮食印刷厂
开　　本 787mm×1092mm　1/16　印张18.5　字数433千字
版　　次 2009年7月第4版　2009年7月第6次印刷
书　　号 ISBN 978-7-5603-1489-1
定　　价 30.00元

前　言

语音信号处理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，是一门新兴的交叉学科，是在多门学科基础上发展起来的综合性技术。它涉及到数字信号处理、模式识别、语言学、语音学、生理学、心理学及认知科学和人工智能等许多学科领域。

语音信号处理是目前发展最为迅速的信息科学研究领域中的一个，其研究涉及一系列前沿课题，且处于迅速发展之中。其研究成果具有重要的学术及应用价值。

从技术角度讲，语音信号处理是信息高速公路、多媒体技术、办公自动化、现代通信及智能系统等新兴领域应用的核心技术之一。用数字化的方法进行语音的传送、储存、识别、合成、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。同时，自然语言作为一种理想的人机通信方式，可为计算机、自动化系统等建立良好的人机交互环境，提高社会的信息化和自动化程度。目前，语音技术处于蓬勃发展时期，有大量产品投放市场，并且不断有新产品被开发研制，具有广阔的市场需要和应用前景。

本书介绍了语音信号处理的基础、原理、方法和应用，以及该学科领域取得的一些新成果、新进展及新技术。全书共分 15 章，其中第 2 章介绍的是语音信号处理的基础知识。第 3 章到第 8 章介绍的是语音信号的各种分析和处理技术，包括传统方法，如时域、频域处理等，还包括新方法和新技术，如同态处理、线性预测分析、矢量量化、隐马尔可夫模型技术等。第 2 章至第 6 章是各种具体应用领域的共同基础部分。第 9 章至第 15 章介绍的是语音信号的各种处理及应用，包括基音提取与共振峰估值、波形编码、声码器、语音合成、语音识别、说话人识别及语音增强。

本书的特点如下：

1. 力求系统地反映语音信号处理的基本原理与方法，以及该领域的新进展和新技术；在篇幅上，按照基础—分析—处理与应用的顺序组织材料；在选材上，使之既能满足教学需要，又反映出国内外某些具有代表性的新成果。对具体技术在整个语音处理体系中的地位、作用以及与其他技术之间的关系，也给予介绍。

2. 以尽可能简明、通俗的语言，以尽可能少的篇幅，以深入浅出、通俗易懂的方式将这门学科介绍给读者，突出基本概念、原理、方法、应用、研究现状及学科发展趋势，而不是去过多追求数学推导和证明的严谨性。为便于理解，书中配备了较多的曲线、图表及实例。

3. 为了突出重点和节省篇幅，书中对语音信号处理的基础知识部分，包括语音学、语音生成、语音感知等内容进行了大幅度压缩，只对其最基本和必要的部分进行了介绍。有关这

一部分的详细内容请参阅有关文献。同时，书中将主要篇幅放在语音信号处理的原理与方法的阐述上，尽量避免涉及语音学及语言学等内容。

4. 除每章末附有参考文献外，书末将常用的专业名词以中英文对照的方式列出。

本书是在作者为哈尔滨工业大学电子与通信工程系信息工程专业本科生讲授“语音信号处理”课及为信号与信息处理学科硕士研究生开设“语音信号处理”课所编写的校内教材基础上，重新进行编写的。

本书可作为高等院校信号与信息处理、通信、模式识别与人工智能等专业及学科高年级本科生、硕士研究生教材，也可供该领域的科研及工程技术人员参考。

本书在编写过程中参考了大量国内外文献资料，在此向著译者们致谢！

语音信号处理是一门理论性强、实用面广、内容新、难度大的交叉学科，同时这门学科又处于迅速发展之中，尽管作者在各方面作了很大努力，但受水平、学识和经验所限，编写时间又很仓促，书中难免会有缺点及疏漏之处，敬请给予批评指正。

作　者

1999 年 9 月

目　　录

第 1 篇　语音信号处理基础

第 2 篇　语音信号分析

第3篇 语音信号处理技术与应用

第1篇 语音信号处理基础

第1章 绪 论

1.1 语音信号处理概述

通过语言相互传递信息是人类最重要的基本功能之一。语言是从千百万人的言语中概括总结出来的规律性的符号系统，是人们进行思维、交际的形式。语言是人类特有的功能，它是创造和记载几千年人类文明史的根本手段，没有语言就没有今天的人类文明。语音是语言的声学表现，是声音和意义的结合体，是相互传递信息的最重要的手段，是人类最重要、最有效、最常用和最方便的交换信息的形式。语音中除包含实际发音内容的语言信息外，还包括发音者是谁及喜怒哀乐等各种信息。在人类已构成的通信系统中，语音通信方式(比如日常的电话通信)早已成为主要的信息传递途径之一，具有最方便和最快捷的特点。语言和语音也是人类进行思维的一种依托，它与人的智力活动密切相关，与文化和社会的进步紧密相连，具有最大的信息容量和最高的智能水平。

语音信号处理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，它是一门新兴的学科，同时又是综合性的多学科领域和涉及面很广的交叉学科。虽然从事这一领域研究的人员主要来自信号与信息处理及计算机应用等学科，但是它与语音学、语言学、声学、认知科学、生理学、心理学等许多学科也有非常密切的联系。

语音信号处理是许多信息领域应用的核心技术之一，是目前发展最为迅速的信息科学研究领域中的一个。语音信号处理是目前极为活跃和热门的研究领域，其研究涉及一系列前沿科研课题，且处于迅速发展之中；其研究成果具有重要的学术及应用价值。

20世纪60年代中期形成的一系列数字信号处理的理论和算法，如数字滤波器、快速傅里叶变换(FFT)等是语音信号数字处理的理论和技术基础。随着信息科学技术的飞速发展，语音信号处理在最近20多年中取得了重大进展：进入70年代之后，提出了用于语音信号的信息压缩和特征提取的线性预测技术(LPC)，已成为语音信号处理最强有力的工具，广泛应用于语音信号的分析、合成及各个应用领域；以及用于输入语音与参考样本之间时间匹配的动态规划方法。80年代初一种新的基于聚类分析的高效数据压缩技术—矢量量化(VQ)应用于语音信号处理中；而用隐式马尔可夫模型(HMM)描述语音信号过程的产生是80年代语音信号处理技术的重大进展，目前HMM已构成了现代语音识别研究的重要基石。

近年来人工神经网络的研究取得了迅速发展，语音信号处理的各项课题是促使其发展的重要动力之一；同时，它的许多成果也体现在有关语音信号处理的各项应用之中，尤其语音识别是神经网络的一个重要应用领域。

从技术角度讲，语音信号处理是信息高速公路、多媒体技术、办公自动化、现代通信及智能系统等新兴领域应用的核心技术之一。在高度发达的信息社会用数字化的方法进行语音的传送、存储、识别、合成、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。同时，语言不仅是人类相互间进行沟通的最自然和最方便的形式，也是人与机器之间进行通信的重要工具，它是一种理想的人机通信方式，因而可为计算机、自动化系统等建立良好的人机交互环境，进一步推动计算机和其他智能机器的应用，提高社会的信息化和自动化程度。

语音处理技术的应用极其广泛，包括工业、军事、交通、医学、民用等各个领域。目前，语音处理技术处于蓬勃发展时期，已有大量产品投放市场，并且不断有新产品被开发研制，具有极其广阔的市场需要和应用前景。

目前对语音信号均采用数字处理，这是因为数字处理与模拟处理相比具有许多优点。其表现为：① 数字技术能够完成许多很复杂的信号处理工作；② 通过语音进行交换的信息本质上具有离散的性质，因为语音可以看做是音素的组合，这就特别适合于数字处理；③ 数字系统具有高可靠性、廉价、快速等特点，很容易完成实时处理任务；④ 数字语音适于在强干扰信道中传输，也易于进行加密传输。因此，数字语音信号处理是语音信息处理的主要方法。

语音信号处理是一门边缘学科，它主要是数字信号处理和语音学等学科相结合的产物，所以它必然受这些学科的影响，同时也随着这些学科的发展而发展。语音信号处理又简称为语音处理，它的研究目的和处理方法多种多样，一直是数字信号处理技术发展的重要推动力量，而数字信号处理的很大部分内容也涉及语音信号处理。数字信号处理技术的发展，其中的一部分就是由数字语音处理的研究中得到的。无论是谱分析方法，还是数字滤波技术或压缩编码方法等，许多新方法的提出，首先是在语音处理中获得成功，然后再推广到其他领域的。同时，它始终与当时信息科学中最活跃的前沿学科保持密切的联系，并且一起发展。比如说，神经网络、模糊集理论、子波分析和时频分析等研究领域常将语音处理作为一个应用实例，而语音处理也常常从这些领域的研究进展中取得突破。

高速数字信号处理器的诞生和发展也是与语音处理的发展分不开的，语音识别和语音编码算法的复杂性和实时处理的需要，就是促使人们去设计这样的处理器的重要推动力量之一。这种产品问世之后，又首先在语音处理的应用中得到最有效的推广应用。语音处理产品的商品化对这样的处理器有着巨大的需求，因此它反过来又进一步推动了微电子技术的发展。

语音信号处理需要有两方面的知识作为基础，除了数字信号处理外，还有语音学。语音信号处理与语音学存在十分密切的关系。语音学是研究言语过程的一门科学，它包括三个研究内容：发音器官在发音过程中的运动和语音的音位特性；语音的物理属性；以及听觉和语音感知。

1.2 语音信号处理的发展概况

1874年电话的发明可以认为是现代语音通信的开端。电话的理论基础是尽可能不失真地传送语音波形，这种“波形原则”几乎统治了整整一百年。1939年产生了一种概念全新的语音通信技术，这就是通道声码器技术。这种声码器打破语音信号的内部结构，使之解体，提取其参数加以传输，在接收端重新合成语音。这一技术包含了其后出现的语音参数模型的基本思想，在语音信号处理领域具有划时代的意义。40年代后期，研制成功了将语音信号的时变谱用图形表示出来的仪器——语谱仪，为语音信号分析提供了一个有力的工具。语谱仪的研制成功对声学语音学的发展曾经起过很大的推动作用。在语音信号分析研究的基础上，电话通信技术得到了很大发展，同时也开展了人机自然语音通信的研究。这样，便在50年代初出现了第一台口授打字机和第一台英语单词语音识别器。但由于语音信号分析的理论尚未取得决定性成熟，工艺技术水平尚未达到一定高度，这些研究工作都未取得决定性成功。进入60年代，语音信号处理的研究工作取得了新的进展，其主要标志是1960年瑞典科学家Fant的著名论文《语音产生的声学理论》的发表，它为建立语音信号数字模型奠定了基础。另一方面，数字计算机的应用得到了推广。特别重要的是60年代中期数字信号处理的技术和方法取得了突破性进展，其主要标志是快速傅里叶变换算法的成功应用。这样，出现了第一台以数字计算机为基础的孤立词语音识别器，继而又研制出第一台有限连续语音识别器。70年代初，Flanagan出版的重要著作《语音分析、合成和感知》，奠定了数字语音处理的系统的理论基础。与此同时，倒谱分析技术和线性预测技术在语音处理中的成功应用，微电子学和集成电路技术取得的进展，价格低廉的微处理器芯片及专用信号处理芯片的不断问世，再次给数字语音处理技术的发展和推广应用以巨大的推动力。发展到今天，虽然语音信号处理领域中还有许多关键问题尚未很好解决，但已经在很多研究中取得了巨大进展。可以相信，经过长期不断的艰苦努力，必将取得更大的成果。

语音信号处理有着广泛的应用领域，其中最重要的包括语音编码、语音合成、语音识别、说话人识别及语音增强。

语音编码技术是伴随着语音的数字化而产生的，目前主要应用在数字语音通信领域。语音信号的数字化传输，一直是通信的发展方向之一。采用低速率语音编码技术进行语音传输比语音信号的模拟传输有诸多优点。由于简单地将连续语音信号抽样量化得到的数字语音信号，在传输时要占用较多的信道资源，因此，为在尽量减少失真的情况下，使得同样的信道容量能够传输更多路的信号，就必须对模拟语音信号进行高效率的数字表示，即进行压缩编码，就成为语音编码技术的主要内容；如何在中低速率上获得高质量的语音，一直是其研究的主要目标。低数码率编码在无线通讯、网络安全、数字电话及存储系统等方面有广泛的应用前景。语音编码技术的研究开始于20世纪30年代末发明的声码器，但是直至70年代中期，中低比特率语音编码一直没有大的突破。而在最近20多年中整个语音编码技术产生了一个大的飞跃。1980年世界上公布了一种2.4 kbit/s的标准编码算法，使人们所希望的在普通电话带宽信道中传输数字电话的愿望终于变成事实，而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。然而，上述的线性预测编码的音质并不令人满意。80年代以后，提出了众多新型编码算法，可以在16 kbit/s、4.8 kbit/s以至2.4 kbit/s

上提供高质量的语音，而且这些算法都可用单片数字信号处理器实时实现。目前，实用系统的最低压缩速率已经达到2.4 kbit/s甚至更低，在大大节省信道带宽的同时还保证了语音质量。目前的研究是努力减小编码解码过程所产生的时延，以使其在移动通信中得到广泛应用。

近年来，高质量的语音编码技术已经开始大规模地走向实用化，各种国际标准的制定集中反映了这种技术发展的水平和趋势。语音编码的研究和通信技术的发展密切相关。现代通信的重要标志是实现数字化，语音编码技术的根本作用是使语音通信数字化，而语音通信的数字化将使通信技术的水平提高一大步。对于目前的蓬勃兴起的移动通信和个人通信，语音编码技术是非常重要的支撑技术。语音编码技术的进展对通信新业务的发展有着极为明显的影响。同时，语音编码产品化的过程比语音识别容易，其研究成果能很快推向实用，对通信事业的发展将起重要的推动作用。

目前计算机已经得到了广泛的应用，但计算机使用起来还不够方便，因为人与计算机的通信通常是采用键盘和显示器，这种方式在很多场合效率低下，操作也不方便。因而人们期望着计算机具有智能的接口。其目的是使人们能够更加方便、更加自然与计算机打交道，即使计算机象人一样能接收、识别并理解声、文、图信息，能够看懂文字、听懂语言、朗读文章，甚至能够进行不同语言之间的翻译。智能接口技术的研究既有巨大的应用价值，又有基础的理论意义，多年来一直是最活跃的研究领域，成果也最为显著。这里，人们特别期望的是智能的语音接口，最理想的是能用自然语言与机器进行对话。语音是人与人之间以及人与计算机之间的最方便的一种信息交换方式。因此，使计算机具有类似于人的听觉功能和发音功能，是人们长期追求的目标。

语音识别与语音合成为人机交流开辟了一条新的途径。语音识别和语音合成的研究是智能接口技术中的标志性成果。语音合成和语音识别是人工智能的重要课题。语音合成的目的是使计算机说话。它是一种人机语音通信技术，其应用领域十分广泛，这些应用已经发挥了很好的社会效益。对语音合成应用的社会需求是广泛和迫切的，因而语音合成技术的研究和产品开发具有很好的发展前景。目前，有限词汇的语音合成技术比较成熟，在自动报时、报警、报站、电话查询服务等方面得到了广泛应用。而无限词汇语音合成的音质的改善存在较大困难，仍未达到完美的程度。这是当前语音合成研究的主要方向，从社会需求来看也是迫切需要解决的问题。

语音识别是使计算机判断出所说的话的内容。语音识别和语音合成一样，也是一种人机语音通信技术。语音识别的研究具有重要意义，特别是对于汉语来讲，由于汉字的书写和录入比较困难，通过语音输入汉字信息就是显得特别重要。计算机终端的微型化也使键盘操作不方便，使语音输入代替键盘输入的必要性变得更加突出。在计算机智能接口技术及多媒体技术的研究中，语音识别技术具有很大的应用潜力。同时，为了实现人机语音通信，必须具备语音识别和语音理解两种功能。

语音识别的研究比语音合成困难得多，其起步也较晚。它的研究始于20世纪50年代，已有近半个世纪的历史，到目前已取得了长足的进步，而且近年来不断有语音识别器(主要是集成电路芯片)投放市场。目前，小词汇量特定人孤立词语音识别技术已经成熟，而大词汇量连续语音识别系统的性能有待进一步改善。自90年代以来，语音识别的研究重点便集中在大词汇量非特定人连续语音识别上，目前比较有代表性的是1997年IBM公司推出的

Via Voice 大词汇量连续语音识别系统。

20 世纪 90 年代以来，语音识别的研究逐渐由实验室走向实用化。一方面，对声学语音学统计模型的研究逐渐深入，鲁棒的语音识别、基于语音段的建模方法及隐马尔可夫模型与人工神经网络的结合成为研究的热点。另一方面，为了语音识别实用化的需要，听觉模型、快速搜索识别算法，以及进一步的语言模型的研究课题受到很大的关注。在语音识别方面，很多专业人员对其理论和应用进行了广泛的研究，有关这方面的文献浩如烟海。然而，语音识别是一项综合性的、难度很大的高科技项目，从语音中提取满意的信息的过程是一项艰巨复杂的任务。语音识别研究中一直面临着许多难以解决的问题，可以说存在着无穷无尽的困难。目前是语音识别研究的黄金时期，该领域的研究得到了前所未有的重视，国内外均投入了大量人力物力，语音识别因而成为科学与技术研究的热点。

计算机和集成电路技术的发展，推动了语音信号处理的实用化。目前有很多专用语音处理芯片，这些芯片与微处理机或微型计算机相结合可以组成各种复杂的语音处理系统。

1.3 本书的内容

本书系统地介绍了语音信号处理的基础、原理、方法、应用、研究现状及学科发展趋势，以及新方法与新技术。全书分为 3 篇，共 17 章。第 1 篇为语音信号处理基础，其中第 1 章为绪论；第 2 章介绍了进行语音信号处理所必需的有关语音信号的基础知识；第 2 篇为语音信号分析，其中第 3 章至第 8 章介绍了语音的各种分析和处理方法，包括时域分析、短时傅里叶分析、同态滤波和倒谱分析、线性预测分析、矢量量化技术及隐马尔可夫模型技术等。我们认为，这些内容包括了当前语音信号处理中最重要的理论和方法。而第 9 章介绍了语音特征参数的提取方法。第 3 篇为语音信号处理技术及应用，包括第 10 章至第 17 章，介绍了语音处理的理论和方法的各种实际应用，包括语音编码（分为波形编码、声码器技术及混合编码两部分）、语音合成、语音识别、说话人识别、语音增强、神经网络在语音信号处理中的应用及语音信号处理中的一些新兴与前沿技术。这一部分起着理论联系实际的作用，从应用的角度介绍了各种语音处理方法所形成的实际系统，同时也从发展的角度介绍了语音处理学科的发展动态和前景。

第2章 基础知识

2.1 概 述

在研究分析各种语音信号处理技术及其应用之前，必须了解有关语音信号的一些基本特性。为了对语音信号进行数字处理，需要建立一个能够精确描述语音产生过程和语音全部特征的数字模型，即根据语音的产生过程建立一个既实用又便于分析的语音信号模型。为了处理和实现上的简便，这个模型应尽可能简单。然而，人类语音的产生过程很复杂，语音中所包含的信息又十分丰富和多样，因而至今尚未找到一种能够细致描述语音产生过程和所有特征的理想的模型。在已经提出来的许多种模型中，Fant 于 1960 年提出的线性模型是模拟语音主要特征的较成功的模型之一。该模型以人类语音的发音生理过程和语音信号的声学特性为基础，成功地表达了语音的主要特征，在语音编码、语音识别和语音合成等领域得到了广泛应用。这是本章所要介绍的模型，也是以后各章讨论的基础。

本章还将介绍与语音处理关系密切的语音学的一些基本内容。语音学是研究言语过程的一门科学。语音就是人类说话的声音，它是语言信息的声学表现。语言交际是通过连结说话人大脑和听话人大脑的一连串心理、生理和物理的转换过程实现的，这个过程分为“发音—传递—感知”三个阶段。因此现代语音学发展为与此相应的三个主要分支：发音语音学、声学语音学、听觉语音学。

发音语音学主要研究语音产生机理，借助仪器观察发音器官，以确定发音部位和发音方法。这一学科目前已相当成熟。声学语音学研究语音传递阶段的声学特性，它与传统语音学和现代语音分析手段相结合，用声学和非平稳信号分析理论来解释各种语音现象，是近几十年中发展非常迅速的一门新学科。听觉语音学研究语音感知阶段的生理和心理特征，也就是研究耳朵是怎样收听语音的，大脑是怎样理解这些语音的，以及语言信息在大脑中存储的部位和形式。听觉语音学与心理学关系密切，是近几十年才发展起来的新兴学科，目前还处于探索阶段。语音信号处理的进一步发展在很多方面依赖于语音信息的研究，以此为目的的语音学的研究工作也非常活跃。

本章要介绍的语音的产生过程属于发音语音学的内容，语音的声学特性属于声学语音学的内容，而语音感知属于听觉语音学的内容。

本章所介绍的基础知识对于语音信号处理的任何一个研究领域都是必需的，其中贯穿全书的是语音信号产生的数字模型。

2.2 语音产生的过程

声音是一种波，能被人耳听到，它的振动频率在 20 ~ 20 000 Hz 之间。自然界中包含各

种各样的声音，如风声、雷声、雨声、机械发出的声音，乐器发出的声音等。而语音是声音的一种，它是由人的发音器官发出的、具有一定语法和意义的声音。语音的振动频率最高可达15 000 Hz左右。

人类生成语音过程的第一阶段是决定想传给对方的内容是什么，然后将内容转换为语言的形式。选择表现其内容的适当语句，将其按文法规则排列，便能构成语言的形式。由大脑对发音器官发出运动神经指令，发音器官各种肌肉运动，振动空气而形成语音波。这个过程可分为神经和肌肉的生理学阶段和产生语音波、传递语音波的物理阶段。

人类的语音是由人体发音器官在大脑控制下的生理运动产生的。人的发音器官包括肺、气管、喉(包括声带)、咽、鼻和口等，如图2-1所示。这些器官共同形成一条形状复杂的管道，其中喉以上的部分称为声道，随着发出声音的不同其形状是变化的；而喉的部分称为声门。在发音器官中，肺和气管是整个系统的能源，喉是主要的声音生成机构，而声道则对生成的声音进行调制。

产生语音的能量，来源于正常呼吸时肺部呼出的稳定气流，喉部的声带既是阀门，又是振动部件。在说话的时候，声门处气流冲击声带产生振动，然后通过声道响应变成语音。由于发不同的音时，声道的形状不同，所以听到不同的声音。

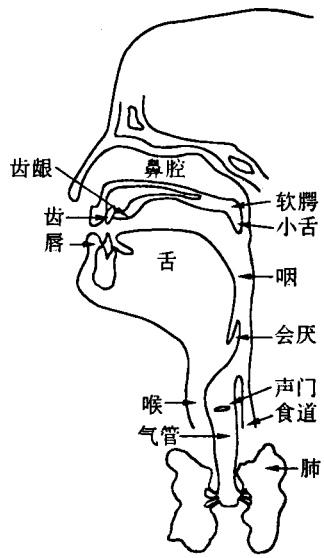


图2-1 人的发音器官简图

喉部的声带是对发音影响很大的器官。声带的声学功能是为语音提供主要的激励源：由声带振动产生声音，是形成声音的基本声源。呼吸时左右两声带打开，讲话时则合拢起来。两声带之间的部位也称为声门。讲话时声带合拢因而受声门下气流的冲击而张开；但由于声带韧性迅速地闭合，随后又张开而闭合……。声带开启和闭合使气流形成一系列脉冲。每开启和闭合一次的时间即振动周期称为音调周期或基音周期，其倒数称为基音频率，也简称为基频。基音频率取决于声带的尺寸和特性，也决定于它所受的张力。声带振动的频率即基频决定了声音频率的高低，频率快则音调高，频率慢则音调低。基音的范围约为80～500 Hz左右，它随发音人的性别、年龄及具体情况而定，老年男性偏低，小孩和青年女性偏高。

语音由声带振动或不经声带振动来产生，其中由声带振动产生的音统称为浊音，而不由声带振动产生的音统称为清音。浊音中包括所有的元音和一些辅音，而清音中包括另一部分辅音。

声道是声门至嘴唇的所有器官，由咽、口腔和鼻腔组成，它是一根从声门延伸至口唇的非均匀截面的声管，其外形变化是时间的函数，发不同音时其形状变化是非常复杂的。成年男子声道的平均长度约17 cm，而声道的截面积取决于其他发音器官的位置，它可以从零(完全闭合)变化到20cm^2。在产生声音的过程中，声道的非均匀截面又是在随着时间不断地变化。声道是气流自声门声带之后最重要的、对发音起决定性作用的器官。

下面介绍语音的产生过程。空气从肺部排出形成气流。空气通过声带时，如果声带是绷紧的，则声带将产生张驰振动，即声带周期性地启开和闭合。声带开启时，空气流从声门

喷射出来，形成一个脉冲；声带闭合时相应于脉冲序列的间歇期。因此，这种情况下在声门处产生出一个准周期性脉冲序列的空气流，该空气流经过声道后最终从嘴唇辐射出声波，这便是“浊音”语音。如果声带是完全舒展开来，则肺部发出的空气流将不受影响地通过声门。空气流通过声门后，会遇到两种不同的情况：一种情况是，如果声道的某个部位发出了收缩而形成一个狭窄的通道，当空气流到达此处时被迫以高速冲过收缩区，并在附近产生出空气的湍流，这种湍流通过声道后便形成“摩擦音”或“清音”；另一种情况是，如果声道的某个部位完全闭合在一起，当空气流到达时便在此处建立空气压力，一旦闭合点突然开启便会让气压快速释放，经过声道后便形成“爆破音”。

由此可见，语音是由空气流激励声道最后从嘴唇或鼻孔或同时从嘴唇和鼻孔辐射出来而产生的。对于浊音、清音和爆破音来说，激励源是不同的，浊音语音是位于声门处的准周期脉冲序列，清音的激励源是位于声道的某个收缩区的空气湍流（类似于噪声），而爆破音的激励源是位于声道某个闭合点处建立起来的气压及其突然释放。

当一个物体（或空腔）作受迫振动，所加驱动（或激励）频率等于振动体的固有频率，便以最大的振幅来振荡，在这个频率上其传递函数具有极大值，这种现象称之为共振。实际上，共振体的共振作用，常常不只是在一个固有频率上起作用，它可能有多个响应强度不同的共振频率。

声道是一个分布参数系统，它是一谐振腔，因而有许多谐振频率。谐振频率由每一瞬间的声道外形决定。讲话时，舌和唇连续运动，使声道常常改变外形和尺寸，随即改变谐振频率。如果声道的截面是均匀的，谐振频率将发生在

$$F_n = \frac{(2n-1)c}{4L}, \qquad n = 1,2,3\cdots \tag{2-1}$$

式中，c 为声速，在空气中为 $c = 350$ m/s；L 为声道长度，n 表示谐振频率的序号。如果 $L = 17$ cm，则谐振频率发生在 500 Hz 的奇数倍上，即 $F_1 = 500$ Hz，$F_2 = 1\,500$ Hz，$F_3 = 2\,500$ Hz，…。发元音 e[ə] 时声道截面最接近于均匀断面，所以谐振频率也最接近于上述值。而发其他音时，声道形状很少是均匀断面的，这些谐振点之间的间隔不同，但平均仍然大约为每 1 kHz 有一个谐振点。

这些谐振频率称为共振峰频率，简称为共振峰，它是声道的重要声学特性。声道对于一个激励信号的响应，可以用一个含有多对极点的线性系统来近似描述。每对极点都对应一个共振峰频率。这个线性系统的频率响应特性称为共振峰特性，它决定信号频谱的总轮廓，或称谱包络。共振峰和声道的形状与大小有关，一种形状对应着一套共振峰。当声音沿着声道传播时，其频谱形状就会随声道而改变。语音的频率特性主要是由共振峰决定的。而声道的共振峰特性决定所发声音的频谱特性，即音色。人在说话时，元音的音色和区别特征主要取决于声道的共振峰特性。共振峰特性可以从语音信号频谱分析得到的幅频特性观察到。声门脉冲序列具有丰富的谐波成分，这些频率成分与声道的共振频率之间相互作用的结果对语音的音质有很大影响。由于声道的大小随不同讲话而不同，因此共振峰频率与讲话者有密切关系。即使是音素相同，但因讲话者不同，共振峰也有相当大的变化。

共振峰用依次增加的多个频率表示，如 F_1、F_2… 等，称为第一共振峰、第二共振峰 … 等。为了得到高质量的语音，或者说为了精确描述语音，必须采用尽可能多的共振峰。但在实际应用中，只有头三个共振峰才是重要的。在声学语音学中通常考虑 F_1 和 F_2，但在语音识