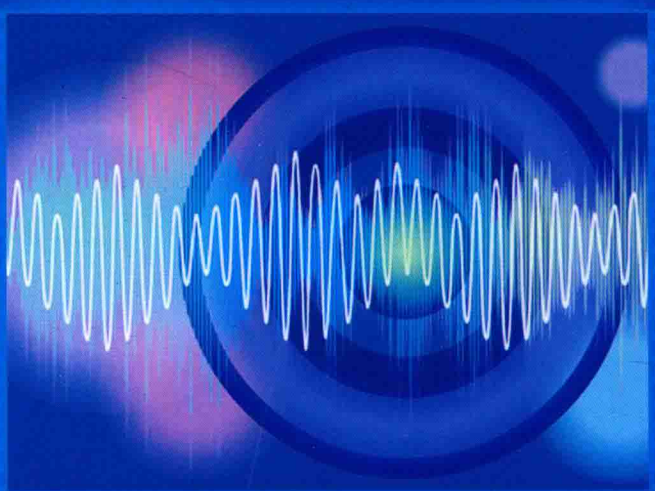
“十三五”普通高等教育规划教材

语音信号处理

（C++版）

梁瑞宇 赵 力 王青云 等编著



含电子教案和源码

http://www.cmpedu.com

机械工业出版社
CHINA MACHINE PRESS

“十三五”普通高等教育规划教材

语音信号处理

（C++版）

梁瑞宇　赵　力　王青云　唐闺臣　等编著

机械工业出版社

本书介绍了语音信号处理的基础、原理、方法和应用，并且给出一些语音信号处理关键算法的C++函数。全书共分12章。第1章介绍了语音信号处理的发展历程和相关研究方向；第2~4章介绍了语音信号处理的一些基础理论、方法和参数；第5~12章按语音信号处理的研究方向，分别介绍了语音增强、说话人识别、语音识别、语音信号情感处理、语音合成与转换、声源定位、语音隐藏和语音编码的基础理论和算法原理。在附录中，介绍了本书涉及的C++类库及引入的函数库，并且以基于Visual Studio的语音录放程序为例，详细介绍了基于MFC的语音处理框架及程序实现。

本书可作为计算机和通信与信息系统等学科相关专业的高年级本、专科学生和研究生的教材用书或教学参考用书，也可作为从事语音信号处理的科研工程技术人员的辅助读物和参考用书。

本书配有电子教案和程序代码，读者可登录机械工业出版社网站（www.cmpedu.com）免费注册，审核通过后下载，也可联系编辑索取（电话010-88379753，QQ6142415）。

图书在版编目（CIP）数据

语音信号处理：C++版/梁瑞宇等编著.—北京：机械工业出版社，2018.1
“十三五”普通高等教育规划教材
ISBN 978-7-111-58755-2

Ⅰ.①语… Ⅱ.①梁… Ⅲ.①语声信号处理-C语言-程序设计-高等学校-教材 Ⅳ.①TN912.3 ②TP312.8

中国版本图书馆CIP数据核字（2017）第312905号

机械工业出版社（北京市百万庄大街22号 邮政编码 100037）
策划编辑：李馨馨 责任编辑：李馨馨
责任校对：张艳霞
责任印制：孙 炜

北京玥实印刷有限公司印刷

2018年1月第1版·第1次印刷
184mm×260mm·23印张·558千字
0001-3000册
标准书号：ISBN 978-7-111-58755-2
定价：59.80元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务
服务咨询热线：(010)88379833
读者购书热线：(010)88379649

网络服务
机工官网：www.cmpbook.com
机工官博：weibo.com/cmp1952
教育服务网：www.cmpedu.com
金书网：www.golden-book.com

前　言

语音信号处理是以语音语言学和数字信号处理为基础而形成的一门涉及面很广的综合性学科，与心理学、生理学、计算机科学、通信与信息科学以及模式识别和人工智能等学科都有着非常密切的关系。该学科始终与信息科学中最活跃的前沿学科保持密切的联系，并且一直是数字信号处理技术发展的重要推动力量，从而能够长期地、深深地吸引广大科研工作者不断地进行研究和探讨。

本书较全面地反映了现代语音信号处理的主要内容和发展方向，主要面向信号与信息处理、电路与系统、通信与电子工程、模式识别与人工智能、计算机信息处理等学科有关专业的高年级本科生和研究生，也可以作为从事语音信号处理这一领域科研工作的技术人员参考书。因此，本书在内容上强调基本概念和基本理论方法的掌握，并突出各部分的相互联系。此外，考虑到语音信号处理的实用性很强，本书在介绍基本理论和基本算法的基础上，给出部分 C ++ 程序实现，使学习人员可以边学习理论边实践，有助于知识的理解和记忆。

本书的参考学时为本科生 32 学时、研究生 40 学时，可以根据不同的教学要求对内容进行适当取舍，灵活安排授课学时数。全书共分为 12 章，具体内容如下：

第 1 章简要介绍了语音信号处理的发展历程和当前的主要研究方法，以及本书的章节安排情况。

第 2 章介绍了语音信号处理的基础知识，包括语音的发音和感知机理、语音信号的数学模型、语音信号的基本参数以及语音的基本表征方法等。

第 3 章介绍了语音信号的预处理方法（包括分帧与加窗、趋势项和直流量的消除、预加重和去加重）以及 4 种语音信号的基本分析方法，包括时域分析、频域分析、倒谱分析和线性预测分析。

第 4 章介绍 3 种语音信号的特征提取技术，包括端点检测、基音周期估计和共振峰估计。其中，端点检测算法包括双门限法、自相关法、谱熵法、比例法和谱距离法；基音周期估计算法包括信号预处理、自相关法、平均幅度差函数法、倒谱法、简化逆滤波法以及后处理法；共振峰估计算法包括倒谱法和线性预测法。

第 5 章介绍了语音增强的基本原理和典型算法。首先介绍了语音和噪声特性、人耳的声音感知特性和语音质量的评价标准，然后依次介绍 4 种语音增强算法：谱减法、维纳滤波法、自适应滤波器法和基于听觉掩蔽效应的语音增强方法。

第 6 章介绍了说话人识别算法。首先介绍了说话人识别的原理及系统结构，然后介绍了两种典型的说话人识别系统，分别是基于 VQ 的说话人识别系统和基于 GMM 的说话人识别系统。最后介绍了说话人识别的研究难点。

第 7 章介绍了语音识别算法。首先介绍了语音识别基本原理与系统构成，然后介绍了基于动态时间规整的语音识别系统和基于隐马尔可夫模型的语音识别系统，最后介绍了算法的评测方法。

第 8 章介绍了语音信号中的情感信息处理的基本原理。首先介绍了情感理论和语音数据

库的建立方法，然后介绍了一些常用的语音情感特征及其提取算法，最后介绍了3种语音情感识别算法，包括K近邻分类器、支持向量机和人工神经网络。

第9章介绍了语音合成与转换的基本原理。首先介绍了帧合成技术，然后介绍了3种语音合成算法，包括线性预测合成法、共振峰合成法和基音同步叠加技术，接着介绍了语音信号的变速和变调的原理和实现方法，最后介绍了语音转换的基本原理和研究方向。

第10章介绍了声源定位的基本原理。依次介绍了双耳听觉定位原理及方法和3种基于传声器阵列的声源定位方法，即基于最大输出功率的可控波束形成算法、基于到达时间差的定位算法和基于高分辨率谱估计的定位算法。此外，还介绍了传声器阵列模型以及可用于声源定位研究的房间回响模型。

第11章介绍了语音隐藏的基本原理。首先介绍了信息隐藏基础理论，然后主要介绍了两种语音隐藏算法：低比特位编码法和回声隐藏算法，最后介绍了算法的常用评价指标以及未来的研究方向。

第12章介绍了语音编码的基本原理。首先介绍了语音编码的理论基础，然后介绍语音编码的主要性能指标，接着依次介绍了3种语音编码算法的基本原理和典型代表，最后对未来研究进行了展望。

在附录中，给出了书中涉及的C++类库及引入的函数库和基于Visual Studio的语音采集程序框架及实现。

需要说明的是，书中加“[C]”的章节包含关键算法的C++函数及说明。

本书主要由梁瑞宇、赵力、王青云和唐闺臣编著，并由梁瑞宇统稿。参加本书编写和校对整理工作的还有包永强、谢跃和赵立业。本书的出版得到了江苏高校品牌专业建设工程项目（项目编号：PPZY2015A035）和江苏省2016年度教育科学规划重点资助课题（项目编号：B-a/2016/01/44）的资助。作者参考和引用了一些学者的研究成果，具体见参考文献。在此，作者向这些文献的著作者表示敬意和感谢，同时诚势感谢给予此书指导和帮助的老师和同学们。

本书还可以配套《语音信号处理实验教程》（ISBN 978-7-111-53071-8）使用，以方便教师根据不同的学生层次和要求来组织实验教学，加深学生对知识的理解和掌握。

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快速发展之中，尽管作者在编写过程中始终注重理论紧密联系实际，力求以尽可能简明、通俗的语言，深入浅出、通俗易懂地将这门学科介绍给读者，但因作者水平有限、时间较仓促，缺点错误在所难免，敬请广大读者批评指正。

编　者

目　录

第1章 绪 论

1.1 语音信号的发展历程

通过语音传递信息是人类最重要、最有效、最常用和最方便的交换信息的形式。语言是人类特有的功能，声音是人类常用的工具，是相互传递信息的最主要的手段。因此，语音信号是人们进行思想沟通和情感交流的最主要的途径。并且，由于语言和语音与人的智力活动密切相关，与社会文化和进步紧密相连，所以它具有最大的信息容量和最高的智能水平。现在，人类已开始进入了信息化时代，用现代手段研究语音处理技术，使人们能更加有效地产生、传输、存储、获取和应用语音信息，这对于促进社会的发展具有十分重要的意义。

让计算机能听懂人类的语言，是人类自计算机诞生以来梦寐以求的想法。随着计算机越来越向便携化方向发展，以及计算环境的日趋复杂化，人们越来越迫切要求摆脱键盘的束缚而代之以语音输入这样便于使用的、自然的、人性化的输入方式。尤其是汉语，它的汉字输入一直是计算机应用普及的障碍，因此，利用汉语语音进行人机交互是一个极其重要的研究课题。作为高科技应用领域的研究热点，语音信号处理技术从理论的研究到产品的开发已经走过了几十个春秋并且取得了长足的进步。它正在直接与办公、交通、金融、公安、商业、旅游等行业的语音咨询与管理，工业生产部门的语声控制，电话和电信系统的自动拨号、辅助控制与查询以及医疗卫生和福利事业的生活支援系统等各种实际应用领域相接轨，并且有望成为下一代操作系统和应用程序的用户界面。可见，语音信号处理技术的研究将是一项极具市场价值和挑战性的工作。我们今天进行这一领域的研究与开拓就是要让语音信号处理技术走入人们的日常生活当中，并不断朝向更高目标而努力。

语音信号处理作为一个重要的研究领域，已经有很长的研究历史。但是它的快速发展可以说是从1940年前后Dudley的声码器和Potter等人的可见语音开始的。20世纪60年代初期，由于Faut和Stevens的努力，奠定了语音生成理论的基础，在此基础上语音合成的研究得到了扎实的进展。60年代中期形成的一系列数字信号处理方法和技术，如数字滤波器、快速傅里叶变换（FFT）等成为语音信号数字处理的理论和技术基础。在方法上，随着电子计算机的发展，以往的以硬件为中心的研究逐渐转移到以软件为主的处理研究。然而，语音识别难度使得该技术在20世纪70年代的发展几乎停滞不前。但是，在整个70年代期间还是有几项研究成果对语音信号处理技术的进步和发展产生了重大的影响：70年代初由板仓提出的动态时间规整技术，使语音识别研究在匹配算法方面开辟了新思路；70年代中期线性预测技术被用于语音信号处理，此后隐马尔可夫模型法也获得初步成功，该技术后来在语音信号处理的多个方面获得巨大成功；70年代末，Linda、Buzo、Gray和Markel等人首次解决了矢量量化码书生成的方法，并首先将矢量量化技术用于语音编码获得成功。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作用，而且很快推广到其他许多领域。因此，20世纪80年代开始出现的语音信号处理技术产品化的热潮，与上述语

音信号处理新技术的推动作用是分不开的。

20世纪80年代，由于矢量量化、隐马尔可夫模型和人工神经网络等相继被应用于语音信号处理，并经过不断改进与完善，使得语音信号处理技术产生了突破性的进展。其中，隐马尔可夫模型作为语音信号的一种统计模型，在语音信号处理的各个领域中获得了广泛的应用。其理论基础是1970年前后，由Baum等人建立起来的，随后，由美国卡内基梅隆大学的Baker和美国IBM公司的Jelinek等人将其应用到语音识别中。由于美国贝尔实验室的Rabiner等人在80年代中期，对隐马尔可夫模型深入浅出的介绍，才使世界各国从事语音信号处理的研究人员有所了解和熟悉，进而成为一个公认的研究热点，也是目前语音识别等的主流研究途径。

进入20世纪90年代以来，语音信号处理在实用化方面取得了许多实质性的研究进展。其中，语音识别逐渐从实验室走向实用化。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，说话人自适应、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等课题倍受关注。

语音信号处理这门学科之所以能够长期地、深深地吸引广大科学工作者不断地对其进行研究和探讨，除了它的实用性之外，另一个重要原因是，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。语音信号处理是以语音语言学和数字信号处理为基础而形成的一门涉及面很广的综合性的学科，与心理学、生理学、计算机科学、通信与信息科学以及模式识别和人工智能等学科都有着非常密切的关系。对语音信号处理的研究一直是数字信号处理技术发展的重要推动力量。因为许多处理的新方法的提出，都是先在语音处理领域中获得成功，然后再推广到其他领域的。例如，许多高速信号处理器的诞生和发展是与语音信号处理的研究发展分不开的，语音信号处理算法的复杂性和实时处理的要求，促使人们去设计许多先进的高速信号处理器。这种产品问世之后，又首先在语音信号处理应用中得到最有效的推广应用。语音信号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。

1.2 语音信号处理的研究方向

语音信号处理是目前发展最为迅速的信息科学技术之一，其研究涉及一系列前沿课题，且处于迅速发展之中。概括来讲，当前的研究方向主要包括九大类。

（1）语音增强

语音增强是指当语音信号被各种各样的噪声干扰、甚至淹没后，从噪声背景中提取有用的语音信号，抑制、降低噪声干扰的技术。然而，由于干扰通常都是随机的，从带噪语音中提取完全纯净的语音几乎不可能。语音增强不但与语音信号数字处理理论有关，而且涉及人的听觉感知和语音学范畴。再者，噪声的来源众多，因应用场合而异，它们的特性也各不相同。所以必须针对不同噪声，采用不同的语音增强对策。

（2）说话人识别

说话人识别通过对说话人语音信号的分析处理，自动确认识别人是否在所记录的话者集合中，以及进一步确认说话人是谁。和语音识别技术很相似，都是在提取原始语音信号中某些特征参数的基础上，建立相应的参考模板或模型，然后按照一定的判决规则进行识别。语音识别

中，尽可能将不同人说话的差异归一化；说话人识别中，力求通过将语音信号中的语义信息平均化，挖掘出包含在语音信号中的说话人的个性因素，强调不同人之间的特征差异。说话人识别是交叉运用心理学、生理学、数字信号处理、模式识别、人工智能等知识的一门综合性研究课题。根据识别对象的不同，说话人识别可分为三类，文本有关、文本无关和文本提示型。

（3）语音识别

语音识别主要指让机器听懂人说的话，即在各种情况下，准确地识别出语音的内容，从而根据其信息，执行人的各种意图。近20年来，语音识别技术取得显著进步，开始从实验室走向市场。随着云计算技术的发展，目前语音识别作为信息技术领域重要的科技发展技术，已经广泛用于工业、家电、通信、汽车电子、医疗、家庭服务、消费电子产品等各个领域。未来语音识别的关键研究在于如何进一步提高算法的鲁棒性。语音识别技术所涉及的领域包括：信号处理、模式识别、概率论和信息论、发声机理、听觉机理和人工智能等。目前，语音识别方法一般有模板匹配法、随机模型法和概率语法分析法三种。

（4）语音情感识别

计算机对从传感器采集来的信号进行分析和处理，从而得出对方（人）正处在的情感状态，这种行为叫作情感识别。从生理·心理学的观点来看，情绪是有机体的一种复合状态，既涉及体验又涉及生理反应，还包含行为。目前对于情感识别有两种方式，一种是检测生理信号如呼吸、心律和体温等，另一种是检测情感行为如面部特征表情识别、语音情感识别和姿态识别。目前，关于情感信息处理的研究正处在不断的深入之中，而其中语音信号中的情感信息处理的研究正越来越受到人们的重视。

（5）语音合成与转换

语音合成，又称文语转换技术，是将任意文字信息实时转换为标准流畅的语音朗读出来。该技术涉及声学、语言学、数字信号处理、计算机科学等多个学科，是中文信息处理领域的一项前沿技术。文语转换过程是先将文字序列转换成音韵序列，再由系统根据音韵序列生成语音波形。

和语音合成原理相似的一种语音处理应用是语音转换，和语音合成不同的是，语音合成是根据参数特征合成语音，而语音转换是将某种特征的语音转换为另一种特征语音。语音合成的研究已有多年的历史，从技术方式上讲可分为波形合成法、参数合成法和规则合成方法；从合成策略上讲可分为频谱逼近和波形逼近。

（6）声源定位

声源定位技术的研究目标是方向估计和距离估计，即主要研究系统接收到的语音信号相对于接收传感器是来自什么方向和什么距离的。声源定位是一个有广泛应用背景的研究课题，其在军用、民用、工业上都有广泛应用。声源定位技术的内容涉及了信号处理、语言科学、模式识别、计算机视觉技术、生理学、心理学、神经网络以及人工智能技术等多种学科。传统的声源定位技术分为基于最大输出功率的可控波束形成法、高分辨率谱估计法和到达时间差的声源定位法。

（7）语音隐藏

语音隐藏技术是指将特定的信息嵌入到数字化的语音中。在某些场合，信息隐藏比加密更安全，因为信息加密是隐藏信息的内容，而信息隐藏是隐藏信息的存在性。信息隐藏的目的不在于限制正常的信息存取和访问，而在于保证隐藏的信息不引起监控者的注意和重视，从而减

少被攻击的可能性。典型的数字语音信息隐藏技术主要有回声隐藏算法、相位编码算法、扩频算法、Patchwork 算法以及标量量化算法。尽管不同的使用场合对语音信息隐藏的要求不同，也没有确定的评判标准及评估系统来判断一种信息隐藏方法的优劣，但从比较广泛的应用范围来考虑，安全性、隐蔽性、鲁棒性和隐藏容量或速率是隐藏技术的主要性能指标。

（8）语音编码

编码、传输、存储和译码是语音数字传输和数字存储的必要过程。随着语音通信技术的发展，压缩语音信号的传输带宽，增加信道的传输速率，成为人们追求的目标。语音编码在实现这一目标的过程中担当了重要的角色。语音编码就是对模拟的语音信号进行编码，将模拟信号转换成数字信号，从而降低传输码率并进行数字传输。语音编码的基本方法可分为波形编码、参量编码（音源编码）和混合编码。

（9）声反馈抑制

声学回声是指扬声器播出的声音在被受话方听到的同时，也通过多种路径被送话器拾取到。在很多情况下都会产生回波，如会议电视系统、免提电话、可视电话终端及移动通信等。回波会严重影响语音的清晰度，更为致命的是，当反馈严重时会产生自激啸叫，使整个系统无法工作。常用的声反馈抑制算法主要包含三类：增益衰减、陷波器和自适应滤波器。增益衰减法的主要思路是降低回声出现通道的增益；而陷波器法最初主要针对静态因素产生的回声而设计。目前，应用较多的是自适应滤波器法。

1.3 本书结构

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。语音信号处理的理论和研究包括紧密结合的两个方面：一方面是从语音的产生和感知来对其进行研究，这一研究与语音语言学、认知科学、心理学和生理学等学科密不可分；另一方面是将语音作为一种信号来处理，包括传统的数字信号处理技术以及一些新的应用于语音信号的处理方法和技术。

本书将系统介绍语音信号处理的基础、原理、方法和应用。全书共分 12 章，其中第 2 章介绍了语音信号处理的基础知识，包括语音的产生与感知、语音产生的数学模型、语音的常用参数语音信号的数字化，以及语音信号的表征等；第 3 章介绍了语音信号的预处理以及四种基本分析方法，包括时域分析、频域分析、倒谱分析和线性预测分析；第 4 章介绍三种语音信号的特征提取技术，包括端点检测、基音周期估计和共振峰估计。第 5 ~ 12 章分别介绍了语音信号处理的各种典型应用，包括语音增强、说话人识别、语音识别、语音信号中的情感信息处理、语音合成与转换、声源定位、语音隐藏和语音编码。需要说明的是，书中加“［C］”的章节包含关键算法的 C ++ 函数及说明，所包含的基本类库可参见附录 A。

语音信号处理是目前发展最为迅速的信息科学技术之一，其研究涉及一系列前沿课题，且处于迅速发展之中。因此本书的宗旨是在系统地介绍语音信号处理的基础、原理、方法和应用的同时，向读者介绍该学科领域一些基本算法、核心理论和基础应用。数字语音信号处理属于应用科学，因此本书不同于以往教材的关键在于，不仅提供了理论知识，而且在关键章节给出了基于 C ++ 的函数功能实现代码。此外，本书附录中给出了基于 Visual Studio 2013 的语音录放程序实现案例，供语音信号处理初学者学习。本书在每一章后面都附有课外思考题，建议学习者进行选做，并进行计算机上机实验以获得实际经验，帮助自己尽快掌握所学的知识。

第 2 章　语音信号处理的基础知识

2.1　语音的产生与感知

2.1.1　人类发音系统

语音是从肺部呼出的气流通过在喉头至嘴唇的器官的各种作用而发出的。作用的方式有三种：①把从肺部呼出的直气流变为音源，即交流的断续流或者乱流；②对音源进行共振和反共振，使其带有音色；③从嘴唇或鼻孔向空间辐射。

与发出语言声音有关的各器官叫作发音器官。人的发音器官包括：肺、气管、喉（包括声带）、咽、鼻和口，如图 2-1 所示。这些器官共同形成一条形状复杂的管道。喉的部分称为声门。从声门到嘴唇的呼气通道叫作声道。声道的形状主要由嘴唇、腭和舌头的位置来决定。声道形状的不断改变会发出不同的语音。

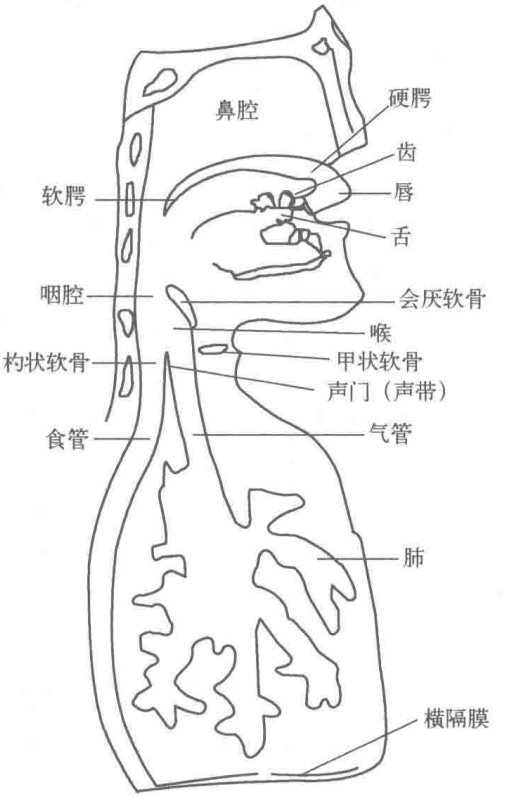


图 2-1　发音器官的部位和名称

声道是自声门（声带）之后对发音起决定性作用的器官。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。由于发不同音时，声道的形状不同，所以能够听到不同的语音。声道的形状主要由嘴唇、腭和舌头的位置来决定。声道中各器官对语音的作用称为调音。口腔是声道最重要的部分，它的大小和形状可以通过调整舌、唇、齿和腭来改变。舌最活跃，它的尖部、边缘部、中央部都能分别自由活动，整个舌体也能上下前后活动；双唇位于口腔的末端，也可活动成展开的（扁平的）或圆形的形状；齿的作用是发齿化音的关键；腭中的软腭是发鼻音与否的阀门，而硬腭及齿龈则是声道管壁的构成部分，同样参与了发音过程。

产生语音的能量，来源于正常呼吸时肺部呼出的稳定气流。气管是由一些环状软骨组成的，讲话时它将来自肺部的空气送到喉部。“喉”是由许多软骨组成的，对发音影响最大的是从喉结至杓状软骨之间的韧带褶，称为声带。呼吸时左右两声带打开，讲话时则合拢起来。而声带之间的部位称为声门。声门的开启和关闭是由两个杓状软骨控制的，它使声门呈 Λ 形状开启或关闭。讲话时声带受声门下气流的冲击而张开；但由声带韧性迅速地闭合，随

后又张开与闭合，这样不断重复。不断地张开与闭合的结果，使声门向上送出一连串喷流而形成一系列脉冲。声带每开启和闭合一次的时间即声带的振动周期就是音调周期或基音周期，它的倒数称为基音频率。基音频率范围随发音人的性别、年龄而定。老年男性偏低，小孩和青年女性偏高。基音频率决定了声音频率的高低，频率快则音调高，频率慢则音调低。

2.1.2 人类听觉系统

人的听觉系统是一个十分巧妙的音频信号处理器。听觉系统对声音信号的处理能力来自于它巧妙的生理结构。从听觉生理学角度来说，人耳的听觉系统可认为是从低到高的一个序列表示，一般分为听觉外周和听觉中枢两个部分，如图 2-2 所示。听觉外周包括位于脑及脑干以外的结构，即外耳、中耳、内耳和蜗神经，主要完成声音采集、频率分解以及声能转换等功能；听觉中枢包含位于听神经以上的所有听觉结构，对声音有加工和分析的作用，主要包括感觉声音的音色、音调、音强、判断方位等功能，还承担与语言中枢联系和实现听觉反射的功能。

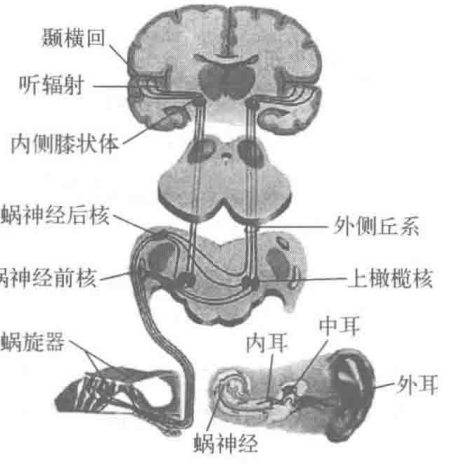


图 2-2 人耳听觉神经系统

外耳是指能从人体外部看见的耳朵部分，即耳郭和外耳道。耳郭对称地位于头两侧，主要结构为软骨。耳郭具有两种主要功能，它即能排御外来物体以保护外耳道和鼓膜，还能起到从自然环境中收集声音并导入外耳道的作用。当声音向鼓膜传送时，由于外耳道的共振效应，会使声音得到 10 dB 左右的放大。此外，外耳道具有保护鼓膜的作用，耳道的弯曲形状使异物很难直入鼓膜，耳毛和耳道分泌的耵聍也能阻止进入耳道的小物体触及鼓膜。外耳道的平均长度为 2.5 cm，可控制鼓膜及中耳的环境，保持耳道温暖湿润，使外部环境不影响中耳和鼓膜。从声音的感知角度来说，外耳主要起着声源定位和声音放大的作用。

中耳由鼓膜、中耳腔和听骨链组成。听骨链包括锤骨、砧骨和镫骨，悬于中耳腔。中耳的基本功能是把声波传送到内耳。声音以声波方式经外耳道振动鼓膜，鼓膜斜位于外耳道的末端呈凹型，正常为珍珠白色，振动的空气粒子产生的压力变化使鼓膜振动，从而使声能通过中耳结构转换成机械能。由于鼓膜前后振动使听骨链做活塞状移动，鼓膜表面积比镫骨足板大好几倍，声能在此处放大并传输到中耳。由于表面积的差异，鼓膜接收到的声波就集中到较小的空间，声波在从鼓膜传到前庭窗的能量转换过程中，听小骨使得声音的强度增加了 30 dB。同时，在一定声强范围内，听小骨对声音进行线性传递，而在特强声时，听小骨进行非线性传递，从而对内耳起到保护的作用。

内耳是位于颞骨岩部内的一系列管道腔，通常可看成三个独立的结构：半规管、前庭和耳蜗。前庭是卵圆窗内微小的、不规则开关的空腔，是半规管、镫骨足板和耳蜗的汇合处。半规管可以感知各个方向的运动，起到调节身体平衡的作用。耳蜗是被颅骨所包围的像蜗牛一样的结构，内耳在此将中耳传来的机械能转换成神经电冲动传送到大脑。耳蜗长约 3.5 cm，呈螺旋状盘旋 2.5 ~ 2.75 圈。它是一根密闭的管子，内部充满淋巴液。耳蜗由三个分隔的部分组成：鼓阶、中阶和前庭阶。其中，中阶的底膜称为基底膜，基底膜之上是柯蒂