中国电子教育学会高教分会推荐
普通高等教育电子信息类“十三五”课改规划教材

数字音视频处理

主　编　韩　冰
副主编　杨　曦　张建龙

西安电子科技大学出版社
http://www.xduph.com

中国电子教育学会高教分会推荐
普通高等教育电子信息类“十三五”课改规划教材

数字音视频处理

主　编　韩　冰
副主编　杨　曦　张建龙

西安电子科技大学出版社

内容简介

本书从人类听觉、视觉的处理机制出发，系统地介绍了听觉和视觉感知模型、数字音频技术、音视频(图像)压缩编码技术、音视频(图像)处理技术和基于内容的音视频(图像)检索技术等内容。在介绍各部分内容的同时，给出了相关知识的应用实例，具有较高的参考学习和实用价值。本书覆盖的学科领域十分广泛，包括人工智能、信号处理、图像处理、语音处理、视频处理和模式识别等一系列学科。读者可以通过本书，学习到很多具有普遍价值的知识和具体的应用方法。

本书可作为高等学校电子信息工程、通信工程和计算科学与技术等相关专业的本科、研究生教材，同时也可作为高职、高专音视频应用相关课程的参考书。

图书在版编目(CIP)数据

数字音视频处理/韩冰主编．—西安：西安电子科技大学出版社，2018.10
ISBN 978-7-5606-4765-4

Ⅰ.① 数… Ⅱ.① 韩… Ⅲ.① 数字技术—应用—音频设备 ② 数字技术—应用—视频信号 Ⅳ.①TN912.27 ②TN941.3

中国版本图书馆 CIP 数据核字(2017)第 305442 号

策　　划　毛红兵
责任编辑　王　斌　毛红兵
出版发行　西安电子科技大学出版社(西安市太白南路 2 号)
电　　话　(029)88242885　88201467　　邮　　编　710071
网　　址　www.xduph.com　　电子邮箱　xdupfxb001@163.com
经　　销　新华书店
印刷单位　陕西利达印务有限责任公司
版　　次　2018 年 10 月第 1 版　2018 年 10 月第 1 次印刷
开　　本　787 毫米×1092 毫米　1/16　印张　16.5
字　　数　389 千字
印　　数　1～1000 册
定　　价　36.00 元
ISBN 978-7-5606-4765-4/TN
XDUP 5067001-1

前　言

随着多媒体、计算机等技术的飞速发展，人们获取信息的渠道日益增多，音频、图像以及视频的应用也越来越广泛。这主要体现在两个方面：一是人们对音频、图像以及视频信息获取的需求越来越强烈；二是通信、计算机、广播电视技术的发展，特别是微电子技术的进步，为音频、图像以及视频信息的处理和通信提供了实现的可能。

如今，越来越多的研究和应用领域都已离不开音频、图像以及视频的处理，特别是在高等院校的教学和科研等工作中，对音频、图像以及视频知识的需求更加迫切。然而，一些相关音频、图像以及视频处理书籍基本都会将音频、图像或者视频分开讲解，针对这三者统一说明和阐述的书并不多见。针对现状，本书结合了音频、图像以及视频三方面的知识，首先从人类脑科学的听觉、视觉出发，介绍了听觉、视觉感知模型，随后给出了如何应用软件获取和处理音频、图像以及视频数据。在简单回顾了数据信号处理，尤其是语音信号处理的基本理论后，本书系统地介绍了音频、图像以及视频处理的基本理论及其相关压缩标准。最后，本书针对音频、图像以及视频的应用问题，分别给出了基于内容的音频、图像以及视频的检索实例。

本书由韩冰担任主编，杨曦、张建龙担任副主编。杨曦编写了第7章和第8章，张建龙编写了第9章，韩冰编写了其余章节。西安电子科技大学的博士研究生王平做了大量的组织协调和整理工作，仇文亮和赵晓静等同学参与了收集资料、绘制图表等工作。可以说本书的出版与他们的努力工作是分不开的。本书的编写还得到西安电子科技大学高新波教授、焦李成教授、石光明教授、苏涛教授和史林教授等老师的鼓励和支持，在此向他们表示深深的谢意。本书在编写过程中参考了大量的文献，在此向所有参考文献的作者表示衷心的感谢。最后，要特别感谢西安电子科技大学出版社社长胡方明的耐心指导和责任编辑王斌的辛勤工作，有了他们的付出，本书才能够顺利出版。

感谢国家自然科学基金委员会一直以来对我们课题组在音频、图像以及视

频方面研究的大力支持，使我们先后得到三个国家自然科学基金(61572384、60902082 和 41031064)的资助。同时，我们还得到了陕西省自然科学基础研究计划资助项目(2011JQ8019)、海洋公益性行业科研专项(201005017)以及教育部留学回国人员科研启动基金的支持与资助。正是在这些基金和项目的资助下，我们的编写工作才能正常有序地进行。

由于编者水平有限，书中难免存在疏漏和不足之处，敬请广大读者批评指正。

韩　冰

2018 年 6 月于西安

目　录

第1章 绪 论

近年来，多媒体数据的表示与通信技术取得了惊人的进展，作为引领全球消费电子行业的国际视频编解码标准也正处于更新换代的时期。目前我国在数字音视频编解码、传输与服务方面的核心关键技术仍然受制于人。开展新一代数字音视频媒体综合服务标准的研究，对形成我国自主知识产权、提高我国信息技术自主创新能力，为我国信息产业可持续发展提供核心技术，带动信息产业突破性发展具有重要意义。

1.1 数字音视频基础

随着现代科技的不断发展，以信息技术产业为代表的高新技术产业，得到了迅猛的发展，推动了全球产业结构转型和优化升级，带来了人类生产和生活的深刻变化。数字音视频技术已经成为当前最流行、使用最频繁、应用范围最广泛的新技术。它与人们的生活、工作和娱乐紧密联系在一起，正日益改变着人们的生活方式，掀起了一场声势浩大的信息革命。

数字音视频技术是信息领域的基础技术之一，随着大规模集成电路、计算机数字技术的发展，传统的影视传媒、消费类电子以及通信行业几乎全部实现了数字化。数字化促进了这些行业的迅速发展，同时也将原来不同的行业——计算机、通信、影视传媒和消费类电子等汇聚在一起。所谓数字化，是指信息的采集、传输、交换和处理过程全面采用数字化技术。

数字音视频技术是对音视频信息(如文本、图形、图像、声音、动画和视频等)进行采集、获取、压缩、解压缩、编辑、存储、传输及再现等环节全部数字化的技术。数字音视频技术的发展推动了音视频产品的发展，音视频产品的数字化进一步提高了产品的技术含量。与传统模拟技术相比，数字音视频技术有以下特点：

(1) 传输效率较高。音视频数字信号被压缩后，可以在6～8 MHz的传输信道内传送2～4套标准清晰度电视(SDTV)节目或一套高清晰度电视(HDTV)节目。

(2) 信息传输存储灵活方便。数字信号便于存储、控制、修改，存储时间与信号特点无关。存储媒体的存储容量大，存储媒体可以是CMOS(互补金属氧化物半导体)型存储器，也可以是计算机硬盘、高密度激光盘等。

(3) 信息传输存储的可靠性高。数字信号的检错、容错、纠错能力强，在数字信号传输放大过程中若出现误码，则很容易实现检错与纠错。

(4) 抗干扰能力强。数字信号不会产生噪声和失真的累积。

(5) 有效保护信息和进行版权管理。数字信号便于实现加密/解密技术和加扰/解扰技术，便于专业应用(军用、商用、民用)或条件接收、视频点播、双向互动传送等。

(6) 具有可扩展性、可分级性和可操作性。数字音视频技术易于与其他系统配合使

用，在各类通信信道和网络中传输，构成一个灵活、通用、多功能的综合业务信息传输网。

(7) 便于与其他数字设备融合。因为数字设备的信号语言是相同的，只要有一套数字信号传输编码、调制协议，就可以做到互联、互通。以音视频数字化为代表的消费类电子，正逐渐与电子计算机、通信技术相融合。

(8) 易于集成化和大规模生产，其性能一致性好且成本低。

1.2 数字音视频技术的发展趋势

数字音视频技术的主要关键技术为音频和视频的获取、信源编码技术和信道编码技术、音频处理、视频处理。信源编码技术包括视频编码技术和音频编码技术。视频编码技术的主要目的是在保证一定重构质量的前提下，以尽可能少的比特数来表征视频信息。目前国内外通用视频压缩标准都是基于混合编码框架的，但随着计算机网络的发展和应用需求的多样化，视频编码技术的研究开始向可伸缩编码、多视点编码等分支方向发展。而音频编码主要是完成声音信息的压缩，其主要分为两类：一类为基于线性预测技术的混合编码；另一类是基于变换的感知音频编码。近几年来，音频编码技术也开始向无损编码、可伸缩编码等分支方向发展。音频处理主要是音频的合成、音频的检测与分类等技术。视频处理主要包括视频处理中的关键技术研究。最后就是基于这些处理技术的音视频检索研究。

1. 国外发展趋势

目前，国外音视频技术领域正在发展的主要技术包括：

(1) 压缩码率更高和算法更先进的音视频数字信号压缩编解码技术。

(2) 传输效率更高和传输质量更优的数字信号调制解调技术。

(3) 加快已成熟的数字音视频技术产品的商品化，推广、普及高清电视技术，通过卫星电视直播接收、电缆电视传输系统和地面广播等三个途径实现模拟电视向数字电视的过渡。

(4) 发展存储容量更大的存储媒体，如高集成度的 CMOS 半导体存储器、固体存储器和采用蓝光技术的高密度光盘等。

(5) 发展新型显示器件，提高显示器件的清晰度、对比度、亮度，降低成本，提高重显彩色色域，寻求新型显示方式和新型发光材料。目前除比较成熟的平面型阴极射线管显示器之外，还有液晶显示屏(LCD)、等离子显示屏(PDP)、有机发光二极管(OLED)型显示器等。

(6) 发展新型电声显示屏和数字音频技术，包括微传声器、基于传声器阵列的语言增强和说话定位技术、多声道回声抵消技术等。

(7) 数字音视频与科研领域的新应用，包括音视频新技术的发展、新算法的研究等及其在人们生活中的应用研究。

2. 我国的现状与差距

我国的音视频技术通过引进、消化、吸收、创新、国产化，走出了一条发展快、技术新的成功道路，不仅缩小了与国外先进国家的差距，提高了广大人民群众的生活质量，满足

了人们日益增长的物质文明和精神文明的需要，而且带动了国民经济持续、稳定和健康发展。2014年我国彩色电视机的产量已经达到15 541.94万台，在国际市场上，我国已经成为以彩色电视机、彩色显像管等为代表的音视频产品的重要生产基地。伴随着互联网的普及，更多的音视频资源被发布在网站上，越来越多的消费者选择在网络上进行音视频点播操作。音视频产品对电子信息产业的生产增长贡献率达到45%以上。此外，音视频技术领域的飞速发展，带动了模具制造、精密机械制造、微电子、光机电、冶金及化工等相关产业的发展。

目前，我国的音视频行业基本掌握了产品的设计技术和生产制造技术，能自行设计、制造出具有先进水平的音视频产品。其产品价廉物美，具有一定的国际竞争能力，成为名副其实的生产、制造和出口大国。但与先进国家相比，我国的音视频技术仍有一定的距离。首先，健全的科技创新体系还未成熟。我国在音视频技术领域的专利技术很少，关键技术大多掌握在国外大公司手中，制约了我国产品进入国际市场的利润空间。音视频产品的某些关键器件仍然依靠进口，尤其是专用超大规模集成电路、关键的显示器件(如PDP显示屏、LCD显示屏等)，我们仍不能自主开发、生产，其中，音视频产品中的专用集成电路95%以上需要依靠进口，PDP和LCD显示屏等仍被国外几个大公司垄断。

另外，我国建立了多个音视频研究中心和实验室，为音视频技术的发展提供有力的技术保障和支持。西安交通大学获准建设数字音频编解码技术与处理芯片国家工程实验室，近年来，在郑南宁教授的带领下，在多媒体数据的表示与通信技术方面取得了惊人的进展。哈尔滨工业大学的NELVT实验室的主要研究领域包括数字音视频编解码技术、基于内容的海量多媒体信息检索、人体生物特征检测与识别、智能人机交互技术和应用算法学等，并承担国家973计划、国家自然科学基金和国家863计划等多项重大研究课题。国内的高校和研究所的研究为数字音视频的理论发展提供了基础。

3. 我国未来发展趋势

音视频产业作为我国信息产业的重要领域，在行业政策的落实与新技术的创新下，不断发展壮大。经过多年的发展，我国数字音视频产业已经成为我国重要的支柱产业和基础产业。

在数字化、网络化、无线化和融合化的发展趋势引导下，中国音视频产业不断发展与完善，数字音视频产品已经被大众消费者广泛应用到生活、工作中。随着新型显示、移动互联网、云计算等新技术的应用，音视频产业向全数字、大屏幕、超高清、网络化和智能化方向发展。

另外，数字音视频领域从音视频算法应用方面转向偏重于相关理论方面的研究，也就是从工程师到研究者的华丽转变，即要从算法等根本上解决音视频的信息化等，这也是我们未来最为重要的研究方向。

1.3 数字音视频系统的组成

数字音视频信息系统模型如图1-1所示。其中，信源编码和信源解码统称为信源编码，信源编码主要解决有效性问题，只有通过对信源的压缩、扰乱和加密等一系列处理，才能用最少的码数去传递最大的信息量，使信号更适宜传输存储；信道编码和信道解码统称为信道编码，主要解决可靠性问题，旨在尽可能使处理的信号在传输/存储过程中不出

错或少出错，即使出错了也要能自动检错和自动纠错，通常包括调制编码和纠错编码，前者主要解决码间干扰产生的错误，后者解决“噪声”引起的突发性错误(如光盘划伤、污迹等)；格式编码和格式解码统称为格式编码，主要解决高效性问题，旨在通过对所存储/传输数据的组织达到提高数据存取速度的目的。传输信道或介质统称为信道，实际上信道可以是由光缆或电缆构成的有线信道，也可以是由高频无线线路、微波线路或卫星中继等构成的无线信道。存储介质可以是磁带、磁盘和光盘等。无论是何种介质，都将受到不同性质的噪声干扰。信源和信宿指的是音视频的采集和重放等终端设备。

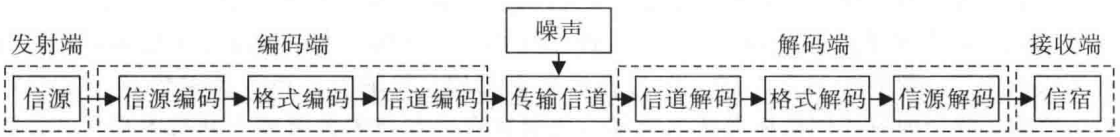

图 1-1 数字音视频信息系统模型

1. 数字音视频信息处理

1) 信息处理

信息处理包括信息的获取、交换、存储，信息特征的提取与选择，信息的分类与识别、传递、处理分析以及信息安全标准化技术等方面的内容。

信息获取是信息处理的基础，主要包括界面接口技术和提取技术两个主要方面。提取技术是指从已经获取的信号中提取感兴趣的信息，它是信号处理技术的一种应用。信息获取的一般过程如图 1-2 所示。其主要流程是：首先为信息需求，即对所需信息进行精确定位；其次对信息来源进行选择；随后确定获取信息所用的方法；最后对获取的信息进行评价。

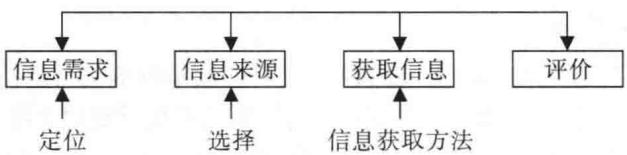

图 1-2 信息获取的一般过程

各种音频信号、图像信号都是音视频信源信息的载体，从这些信号中提取音视频信息的主要特征，利用计算机进行自动分类与识别是信息处理的基本方式，也是信息处理的主要内容。

音视频信息的主要特征包括数字化特征、结构特征、几何特征和空时特性等。而特征提取与选择的主要任务是根据既定的识别任务，按照预先给定的判别准则，选择合适的特征以便更好地完成分类与识别任务，因而它主要是一个统计优化问题。信息特征选择的优劣会最直接地影响到识别性能。由于自然界中的音视频信息的复杂多样性以及海量性，特征选择作为解决这些问题的基本和重要途径，自然而然也就成为了音视频信息处理中最重要和最复杂的问题之一。

对于音视频信息处理中的重要部分，视觉和听觉信息的识别，即语音识别、图像识别或者文字识别等，也是模式识别的主要内容。

信息交换也称为存储和转发交换，包括通过网络从节点到节点的信息传送。信息存储

是指将获得的或加工后的信息保存起来，以备未来应用。信息存储不是一个孤立的环节，它始终贯穿于信息处理的全过程。信息安全也是信息处理的重要内容。信息安全主要是指信息系统的信息不被泄露给非授权用户、实体或供其使用。

2）信息的数字化处理

计算机系统能够处理通过键盘接收到的字符信息，通过扫描仪、视频接收器等接收到的图像信息以及通过话筒或其他语音设备接收到的音频信息等。但计算机并不能直接处理这些不同形态的信息，而是必须通过信息的数字化处理。信息的数字化是指通过计算机中的编码转换器把各种不同形态的信息转换成机器能识别与运算的二进制数字形式来进行加工处理。数字化是计算机处理信息的基础。而数字化的重要手段就是利用数字信号处理技术对各种信号进行数字化处理。

数字信号处理技术通常是对数字信号进行采集、检测、变换、调制、压缩和降噪等方面的处理。

2. 数字音频信息处理系统

数字音频信息处理系统是指对音频信号进行采集、获取、编码、解码、存储、变换、合成、识别、理解、传输和编辑等。数字音频是一个关键且重要的概念，可以用来表示声音强弱的数据序列，并由模拟声音经抽样(即每隔一个时间间隔在模拟声音波形上取一个幅度值)、量化、编码(即把声音数据写成计算机的数据格式)后而得到。模拟/数字转换器把模拟声音变成数字声音；数字/模拟转换器可以恢复出模拟的声音。

语音信息处理以心理声学、物理声学、生理声学、语言学和语音学为基础，涉及电子技术、信号处理技术、微电子技术和计算机控制技术等多个学科领域。

目前，语音处理研究的重点是完善语言产生模型，加强语言感知模型的研究，建立一个完整的语言产生模型；利用听觉的心理特性，如声音的掩蔽效应，实现大压缩比高效编码等应用。

3. 数字图像/视频处理系统

一般的图像/数字视频处理系统模型包括：图像/视频输入设备、存储设备、控制设备、用户存取/通信设备、输出设备以及专用图像/视频处理硬件设备。不同的应用环境，所需要的硬件设备、软件环境也不同。

(1) 图像/视频输入设备：主要用于将待处理的图像/视频信号输入系统装置或者计算机等。图像视频的装置一般有摄像头、数字照相机、扫描仪、数字摄像机、磁盘和视频采集卡等多种静态或动态图像生成、存储设备或装置。

(2) 图像/视频存储设备：主要用于在处理图像/视频过程中对视频/图像信息本身和其他相关信息进行暂时或永久性的保存，如U盘、RAM、ROM、硬盘和磁带等。

(3) 图像/视频控制设备：处理图像/视频过程中的相关控制设备，如鼠标、键盘、操纵杆和开关等。

(4) 用户存取/通信设备：主要用于将图像/视频信号提取或存入视频处理模块。

(5) 图像/视频输出设备：主要用于将经过系统或计算机处理后的图像/视频信号以用户能感知的形式显示出来，常见的有显示器、打印机、绘图仪和影像输出系统等。

(6) 专用图像/视频处理设备：主要用于对待处理的图像/视频信号进行给定任务的处

理。视频处理设备一般可分为两类：一类是软件型视频处理系统，即将视频处理卡插入计算机，视频处理卡中有专用硬件，而相应的处理工作则由计算机软件来完成；另一类是专用型计算机图像/视频处理系统，由专用硬件对图像/视频进行处理。

4. 数字音视频系统的应用模型

当前，信息技术的发展日新月异，信息技术的普及和应用对于经济、政治、社会、文化和军事等方面，都有深远的影响。以数字音视频为代表的信息技术已衍生出很多应用，包括音频/图像/视频检索、数字图书馆、视频监控、视频点播、视频会议、远程教育、游戏互动、远程医疗、军事监控、远程监控和远程交易等。

数字音视频技术是电子信息数字化的核心内容，目前被广泛应用在广播电视、计算机、通信、网络和家用电器等领域，对经济发展和社会进步都具有重要的意义。

1.4 本章小结

本章介绍了数字音频和视频的主要技术以及国内外的发展趋势，为后续分析数字音视频技术提供了基础。

第 2 章　听视觉处理的脑机制

在众多的生物系统中，人脑是最有效的生物智能系统，它具有感知、识别、学习、联想、记忆和推理等功能。人类感知外部世界主要是通过视觉、触觉、听觉和嗅觉等感觉器官，其中最主要的是通过视觉感知外界信息。据统计，人类感知的信息有 80%来自视觉。为此，研究生物体的视觉功能，解析其内在的机理，并用机器来实现，成为科学研究领域的一个重要方面。进一步来讲，研究其过程有助于我们从根本上研究音视频处理的脑机制，模拟这些机制与规律，研究音视频处理的各种方法与模型，同时为开发智能化信息处理模式开拓新的途径。

2.1　听觉的生理基础

随着信息化社会的发展，生命科学是信息科学领域最值得期待的学科。脑和神经系统的信息加工和信息处理方式已成为信息科学家们着力研究的对象。而语音信息处理作为信息科学的重要组成部分，这一领域的研究人员努力使计算机语音识别逼近听觉感知过程，而对听觉感知模型的研究就是实现这一目标的途径。

听觉是一个接收、理解声音信息的过程，是听者对说话人所传来的声音信息进行编码的过程。感知是指作用于我们的听觉感受器官的声音的各种属性在我们大脑中的反应。听觉感知模型研究是用数学表达式对听觉系统的特征和信息处理方式做出抽象和描述，从而构成具有人类听觉系统特性的语音信号处理系统的研究。听觉感知模型研究是一种跨学科的研究，它涉及生理声学(研究听觉器官和生理特征的科学)、心理声学(研究声音的主观感知与客观参数间关系的科学)、数理科学和信息科学等。听觉感知模型主要是对生理声学和心理声学的基本理论和实验资料进行分析综合，找出听觉信息加工的本质要素，用适当的数学表达式来描述这些本质要素，并利用这些数学表达式来构成语音信号处理系统。

医用人工耳蜗的研制与听觉感知模型有关，但听觉感知模型研究的最重要的意义在于它将对信息科学和计算机科学提供新的线索和新的思路。

能否有效地将人的听觉处理机制融合到语音信号处理系统中，取得人们所期望的效果，取决于很多条件。首先，需要对听觉系统的处理机制有足够地理解；其次，对于听觉系统的处理机制要能够进行有效的建模，并与相应的语音处理系统有机地结合。

经过多年的研究，人们对于听觉系统已经获得了比较详尽的知识，有些甚至到了细胞层次，例如，耳蜗中内外毛细胞的纤毛排列与听神经突触的触点等细节。人们在生理学实验、理论建模与应用方面取得了可喜的成果。

听觉心理学实验，从宏观角度研究听觉行为与现象，研究人对声信号和语言的主观感受能力，包括频率选择性、声音响度、基音、声信号在时间和空间的处理、听觉模式的感知与语音处理。其主要研究方法是将人看成是黑箱系统，由输入(声音刺激)和输出(人的反

应)考察听觉系统的感知特性。某些听觉感知特性虽然可由听觉生理机制解释，但通常不能确定到底是由听觉系统中的哪些部分完成的。比较有代表性的工作是罗宾逊(Robinson)等人对等响度曲线的测量、弗莱彻(Fletcher)通过遮蔽实验提出临界带(Critical Bandwidth，CBW)的概念、摩尔(Moore)等人通过系列遮蔽实验推出听觉滤波器的形状及听觉的频率选择性并通过调制阈值确定听觉的时间分辨率以及布莱格曼(Bregman)的专著《Auditory Scene Analysis》(《听觉场景分析》)的出版等。听觉心理学实验获得的一些数据，也可以直接用于听觉建模。例如，根据临界带划分的滤波器组及其派生的 Mel 频率倒谱系数(Mel Frequency Cepstrum Coefficients，MFCC)，现在被广泛应用于语音识别的前端处理及特征提取。听觉心理学与听觉生理学研究相互支撑、相互印证，共同推动听觉研究的发展。

2.1.1 听觉感知模型的国内外研究现状

计算机语音识别系统需要听觉感知模型研究解决的问题有：① 找到更简洁有效的参数(语音信息在听觉通路的各个阶段特别是在大脑皮层的反应模式能提供启发)；② 更好地利用过渡音的信息来提高识别率(可以从听觉系统信息处理的非线性和动态特性中得到启发)；③ 找到去除个人信息而只保留音韵信息的参数，来解决非特定说话人的问题；④ 提高抗噪声能力(如语音识别装置在嘈杂的场合几乎丧失了语音识别能力)。

自从 1961 年贝克西(Bekesy)揭示了内耳基底膜机制以来，随着听觉心理和听觉生理科学的发展，对于听觉模型的研究出现了几个高潮：① 20 世纪 60 年代的物理模型，即对外耳、中耳和内耳基底膜的物理特性的模型化，如对耳蜗管这种一端封闭短管的声学特性进行模块化；② 20 世纪 70 年代的神经生理模型，即对内毛细胞将声波震动转化为电脉冲发放的机理和特性的模型化及对听觉神经纤维电脉冲发放模式的模型化；③ 20 世纪 80 年代的表征模型，即对于声信号在听觉系统中表征(Representation)模式的研究和模型化；④ 20世纪 90 年代著名的听觉模型，即美国麻省理工学院的 Seneff 模型；⑤现在主要以注意选择为主的听觉模型。

注意的选择理论有以下四个。

1. 过滤器理论

1958 年，英国心理学家布罗德本特(Broadbent，1958)根据双耳分听的一系列实验结果，提出了解释注意选择作用的一种理论：过滤器理论(Filter Theory)。布罗德本特认为：神经系统在加工信息的容量方面是有限度的，不可能对所有的感觉刺激都进行加工。当信息通过各种感觉通道进入神经系统时，首先要经过一个过滤装置。只有一部分信息可以通过这个机制，并接受进一步的加工；而其他的信息就被阻断在它的外面。布罗德本特把这种过滤机制比喻为一个狭长的瓶口，当人们往瓶内灌水时，一部分水通过瓶颈进入瓶内，而另一部分水由于瓶颈狭小，通道容量有限，而被留在了瓶外。这种理论也称为瓶颈理论或单通道理论。

2. 衰减理论

过滤器理论得到了某些实验事实的支持，但进一步研究发现，这种理论并不完善。例如，格雷(Gray，1960)在双耳分听的研究中，研究发现来自非追随耳的信息仍然受到了加工。基于日常生活观察和实验研究的结果，特瑞斯曼(Treisman，1964)提出了衰减理论。衰减理论主张：当信息通过过滤装置时，不被注意的信息只是在强度上减弱了，但不会完

全消失。特瑞斯曼指出，不同刺激的激活阈限是不同的。有些刺激对人有重要意义，如自己的名字、火警信号等，它们的激活阈限低，容易激活。当它们出现在非追随的通道时，容易被人们所接受。特瑞斯曼的理论与布罗德本特的理论对过滤装置的具体作用有不同的看法，但两种理论又有共同的地方：① 两种理论有相同的出发点，即主张人的信息加工系统的容量有限，所以，对外来的信息需要经过过滤或衰减装置加以筛选；② 两种理论都假定信息的选择过程发生在对信息的充分加工之前，只有经过选择以后的信息，才能进一步加工和处理。

3. 后期选择理论

多伊奇(Deutsch，1963)等人提出了选择性注意的一种观点——后期选择理论，后由诺尔曼(Norman，1968)加以完善。后期选择理论认为，所有进入过滤或衰减装置的信息是经过充分分析的，因此对信息的选择发生在加工后期的反应阶段。后期选择理论也称为完善加工理论、反应选择理论或记忆选择理论。

4. 多阶段选择理论

过滤器理论、衰减理论及后期选择理论都假设，注意的选择过程发生在信息加工的某个特定阶段上。约翰斯顿(Johnston，1978)等人提出了一个较灵活的模型，其认为选择过程在不同的加工阶段都有可能发生，这就是多阶段选择理论。这一理论的两个主要假设是：① 在进行选择之前的加工阶段越多，所需要的认知加工资源就越多；② 选择发生的阶段依赖于当前的任务要求。多阶段选择理论看起来更有弹性，由于强调任务要求对选择阶段的影响，因而避免了过于绝对化的假设所带来的问题。

多阶段选择理论在很大程度上带有假设的性质，原来的很多在注意方面的研究也能够为多阶段加工理论提供一定的实验依据，如有的研究支持早期选择理论，而有的研究则支持后期选择加工理论。

上述理论试图解释注意对信息进行选择的机制，而认知资源理论是关于注意分配的理论，它从另一个角度来解释注意，即注意是如何协调不同的认知任务或认知活动。

认知资源理论就是从注意是如何协调不同的认知任务或认知活动的角度来理解注意的，不同的认知活动对注意提出的要求是不相同的。注意的认知资源理论有以下两个。

1) 认知资源分配理论

认知资源分配理论是由心理学家卡里曼(Kahneman)提出的，他认为注意资源和容量是有限的。这些资源可以灵活地分配去完成各种各样的任务，甚至可以同时做多件事情，但完成任务的前提是所要求的资源和容量不能超过所能提供的资源和容量。认知资源理论从另外一个角度来理解注意，即注意是如何协调不同的认知活动或认知任务的。不同的认知活动对注意所提出的要求也不相同。认知资源理论认为，与其把注意看成一个有限容量的加工通道，不如将其看成是一组对刺激进行归类和识别的认知资源或认知能力。这些认知资源是有限的，对刺激的识别需要占用认知资源，当刺激越复杂或加工任务越复杂时，占用的认知资源就越多。例如，在没有人的高速公路上，熟练的汽车司机可以一边开车，一边和车内的人说话。他之所以能够同时进行两种或两种以上的活动，是因为这些活动所要求的注意容量在他所提供的容量范围之内。而若是在行人拥挤的街道上开车，大量的视觉和听觉刺激占用了他的注意容量，他也就不能再与同伴聊天了。