工业和信息产业科技与教育专著出版资金资助出版

现代语音信号处理

Modern Speech Signal Processing

◆胡 航 编著

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

工业和信息产业科技与教育专著出版资金资助出版

现代语音信号处理

胡　航　编著

電子工業出版社
Publishing House of Electronics Industry

内 容 简 介

本书系统介绍了语音信号处理的基础、原理、方法、应用、新理论、新成果与新技术，以及该研究领域的背景知识、研究现状、应用前景和发展趋势。

全书分三篇共 17 章。第一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；第二篇语音信号分析，包括第 3 章时域分析，第 4 章短时傅里叶分析，第 5 章倒谱分析与同态滤波，第 6 章线性预测分析，第 7 章语音信号的非线性分析，第 8 章语音特征参数检测与估计，第 9 章矢量量化，第 10 章隐马尔可夫模型；第三篇语音信号处理技术与应用，包括第 11 章语音编码，第 12 章语音合成，第 13 章语音识别，第 14 章说话人识别和语种辨识，第 15 章智能信息处理技术在语音信号处理中的应用，第 16 章语音增强，第 17 章基于麦克风阵列的语音信号处理。

本书体系完整，结构严谨；系统性强，层次分明；内容深入浅出，原理阐述透彻；取材广泛，繁简适中；内容丰富而新颖；联系实际应用。

本书可作为高等院校信号与信息处理、通信与电子工程、电路与系统、模式识别与人工智能等专业及学科的高年级本科生及研究生教材，也可供该领域的科研及工程技术人员参考。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

现代语音信号处理 / 胡航编著. —北京：电子工业出版社，2014.7
ISBN 978-7-121-22625-0

Ⅰ. ①现… Ⅱ. ①胡… Ⅲ. ①语声信号处理-高等学校-教材 Ⅳ. ①TN912.3

中国版本图书馆 CIP 数据核字（2014）第 045290 号

责任编辑：韩同平　　特约编辑：李佩乾
印　　刷：北京丰源印刷厂
装　　订：三河市鹏成印业有限公司
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：26.5　　字数：750 千字
版　　次：2014 年 7 月第 1 版
印　　次：2014 年 7 月第 1 次印刷
印　　数：2 500 册　　定价：65.00 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

语音信号处理是在多学科基础上发展起来的综合性研究领域与技术，涉及数字信号处理、语音学、语言学、生理学、心理学、计算机科学、模式识别、认知科学和智能信息处理等学科。它是发展非常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。近年来，该领域取得大量成果，在理论与学术研究上取得长足发展。同时，其研究成果也在很多领域得到广泛应用；目前语音技术处于蓬勃发展时期，有大量产品投放市场，且不断有新产品被开发研制，具有广阔的市场需求和前景。

本书系统介绍了语音信号处理的基础、原理、方法、应用、新成果与新技术，以及该研究领域的背景知识、研究现状、应用前景和发展趋势。本书内容编排按基础－分析－处理与应用的顺序组织材料。

本书作者于2000年在哈尔滨工业大学出版社出版《语音信号处理》，后又多次修订。

这次的《现代语音信号处理》对原书内容、结构等进行了大幅度修订，以适应目前语音信号处理研究的不断发展及高等学校相关专业对本门课程新的教学要求。除传统的语音信号处理外，本书用大量篇幅介绍了现代语音信号处理的内容，包括以下3方面:

（1）语音信号处理领域的一些新技术与新成果，包括语音产生的非线性模型，非线性预测编码，基于HMM的参数化语音合成，可视及双模语音识别，说话人自适应，语音理解，基于子空间分解的语音增强等。

（2）智能信息处理与现代信号处理技术在语音处理中的应用。介绍了一些新兴及前沿的理论与技术，包括混沌与分形、支持向量机、神经网络、模糊理论、遗传算法（及其他智能优化算法）、以及高阶累积量、盲源分离、小波变换、信号子空间分解等在语音信号分析与处理中的应用。

语音信号处理研究已经历了几十年，特别是近30年来已取得很多重要进展；但该领域仍蕴含着很大的潜力，也面临许多理论与方法上的困难，并存在一些难以解决的问题。近年兴起并得到迅速发展的智能信息处理与现代信号处理中的一些理论与技术，是解决这些问题的工具之一；它们已在语音信号处理研究中得到广泛应用，并取得了大量成果，对该领域的发展起到了重要推动作用。

（3）语音麦克风阵列信号处理，包括基于麦克风阵列的声源定位，语音盲分离及语音增强等。基于麦克风阵列的语音信号处理是阵列信号处理与语音信号处理的交叉学科，且涉及声学信号处理的内容。应用于语音信号处理的阵列处理技术与应用于雷达、移动通信及声呐等领域的阵列处理技术有很大不同。这部分内容反映了作者从事阵列信号处理、相控阵雷达及电子侦

察与对抗等领域研究所取得的一些体会与认识。

本书体系完整、结构严谨；系统性强；内容深入浅出，原理阐述透彻；取材广泛，繁简适中；内容丰富而新颖；联系实际应用。可作为高等院校信号与信息处理、通信与电子工程、电路与系统、模式识别与人工智能等专业及学科的高年级本科生及研究生教材，也可供该领域的科研及工程技术人员参考。

感谢工业和信息产业科技与教育专著出版资金对本书出版的资助。

著名信息科学专家、北京交通大学袁保宗教授在百忙之中审阅了本书，提出了很多宝贵的指导性意见，并推荐本书出版；在此向袁先生表示深切的敬意与感谢！同时感谢鲍长春教授提出的宝贵建议。

栾学鹏老师参加了部分编写工作，金玉宝同学提供了帮助，在此一并致谢。

本书力求反映作者多年从事语音信号处理课程教学的经验与体会。鉴于该研究领域内容丰富，涉及众多学科及前沿领域，有很强的实用性，又处于迅速发展之中，受作者水平等多方面因素所限，书中难免存在一些问题与不足，敬请批评指正。

作　者

试读结束：需要全本请在线购买： www.ertongbook.com

目　录

第一篇　语音信号处理基础

第二篇　语音信号分析

第三篇 语音信号处理技术与应用

试读结束：需要全本请在线购买： www.ertongbook.com

第一篇　语音信号处理基础

第1章　绪　　论

1.1　语音信号处理的发展历史

通过语言交流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来的规律性的符号系统，是思维、交际的形式。语言是人类特有的功能，是创造和记载几千年人类文明史的根本手段。语音是语言的声学表现，是声音和意义的结合体，是人类最重要、最有效、最常用和最方便的信息传递与交换形式。语音中除包含实际发音内容的语言信息，还包括发音者是谁及其喜怒哀乐等各种信息。在人类已有的通信系统中，语音通信方式(如日常的电话通信)早已成为主要的信息传递途径之一。语言和语音也是人类思维的一种依托，其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。

语音信号处理简称为语音处理，是用数字信号处理技术对语音信号进行处理的一门学科；其是一门新兴的综合性交叉学科。从事该领域的研究人员主要来自信号与信息处理及计算机应用等领域，但其与语音学、语言学、声学、电声技术、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；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、办公自动化、现代通信及智能系统等领域的核心技术之一。在高度发达的信息社会，用数字化方式进行语音的传送、存储、识别、合成、增强等是数字化通信网中最重要、最基本的组成部分之一。同时，语言不仅是人类沟通的最自然和最方便的形式，也是人与机器通信的重要工具，是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，进一步推动计算机和其他智能机器的应用，提高社会的信息化与自动化程度。语音处理技术的应用包括工业、军事、交通、医学、民用各领域；已有大量产品投放市场，并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。

语音信号均采用数字方式进行处理，数字处理与模拟处理相比有许多优势。表现为：(1) 数字技术可完成很多很复杂的信号处理工作；(2) 通过语音进行交换的信息，有离散的性质，因为语音可看作是音素的组合，从而很适合于数字处理；(3) 数字系统有高可靠性、廉价、快速等特点，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，也易于加密传输。因此，数字语音信号处理是语音信号处理的主要方式。

语音信号的数字表示可分为两类：波形表示和参数表示。波形表示仅通过采样和量化保存模拟语音信号的波形；而参数表示将语音信号表示为某种语音产生模型的输出，是对数字化语音进行分析和处理后得到的。

1874 年 Bell 发明的电话可认为是现代语音通信的开端，其首次用声电-电声转换实现远距

离传输语音。电话的理论基础是尽可能无失真地传送语音波形，这种波形原则统治了很多年。1939 年，美国 Dudley 提出一种新概念的语音通信技术，即通道声码器。其打破了语音信号的内部结构，使之解体，提取参数进行传输，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，在语音信号处理处理领域具有划时代的意义。20 世纪 40 年代后期，美国 Bell 实验室研制出将语音信号时变谱用图形表示的仪器——语谱仪，为语音信号分析提供了有力工具，对声学语音学的发展起到过重要的推动作用。在语音信号分析研究的基础上，电话通信技术得到很大发展，同时也开展了人机自然语音通信的研究。这样，20 世纪 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。但由于此时语音信号分析理论尚未取得决定性的成熟，工艺技术水平未达到一定高度，这些研究工作未取得决定性成功。

20 世纪 60 年代后，语音信号处理的研究取得新进展，主要标志是 1960 年瑞典科学家 Fant 的论文《语音产生的声学理论》发表，为建立语音信号的数字模型奠定了基础。另一方面，数字计算机的应用得到推广。特别重要的是，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、FFT 等是语音信号数字处理的理论与技术基础。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，又研制出第一台有限连续语音识别器。

20 世纪 70 年代初，Bell 实验室的 Flanagan 的重要著作《语音分析、合成和感知》奠定了数字语音处理的理论基础。另外，有几项研究成果对语音信号处理技术的进步与发展产生重大影响。如 70 年代初，Itakura 提出用于输入语音与参考样本间时间匹配的动态时间规整(DTW)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，用于语音信息压缩及特征提取的线性预测技术(LPC)被用于语音信号处理，成为语音信号处理最有力的工具，广泛用于语音分析、合成及其他各应用领域；隐马尔可夫模型(HMM)也取得初步成功。80 年代开始出现的语音信号处理技术产品化的热潮，是与上述语音信号处理新技术的推动作用分不开的。另一方面，倒谱分析与 LPC 在语音处理中得到应用，微电子学和集成电路技术取得进展，价格较低的微处理器芯片及专用信号处理芯片不断出现，再次给数字语音处理技术的发展和推广应用以很大的推动力量。

20 世纪 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)用于语音处理，不仅在语音识别、语音编码及说话人识别等方面发挥了重要作用，且很快推广到其他许多领域。而用 HMM 描述语音信号过程是 80 年代的一项重大进展，HMM 已构成现代语音识别的重要基石。其使语音识别算法从模式匹配转向基于统计模型的技术，更多地追求从整体统计的角度建立最佳语音识别系统。作为语音信号的统计模型，HMM 的理论基础于 1970 年前后由 Baum 等人建立，后被用于语音识别。Bell 实验室的 Rabiner 等学者在 80 年代中期对 HMM 进行了深入的介绍，使其被语音处理领域的研究人员所了解；从而使 HMM 成为研究热点，并成为目前为止语音识别的主流方法。80 年代末 90 年代初，人工神经网络(ANN)的研究异常活跃，取得迅速发展，而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，其许多成果也体现在语音信号处理的各项应用中，尤其语音识别是神经网络的重要应用领域。总之，VQ、HMM 及神经网络相继用于语音信号处理，且不断改进与完善，使语音信号处理技术产生了突破性进展。

语音信号处理为交叉学科，主要是数字信号处理和语音学等学科结合的产物，因而必然受这些学科的影响，同时也随这些学科的发展而发展。语音信号处理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；数字信号处理学科与技术发展的一部分来源于数字语音处理的研究。无论是谱分析，还是数字滤波或压缩编码等，许多新方法的提出首先在语音处理中获得成功，再推广到其他领域。同时，它与信息科学中最活跃的前沿学科保持密切联系，并且一起发展。如神经网

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，而语音处理也常从这些领域的研究进展中取得发展。

语音信号处理以两方面知识为基础，除数字信号处理外还有语音学。语音信号处理与语音学有密切关系。语音学是研究言语过程的一门科学，包括三部分研究内容：发音器官在发音过程中的运动及语音音位特性；语音物理属性；听觉和语音感知。

另一方面，高速数字信号处理器(DSP)的诞生与发展也与语音处理密切相关，语音识别与语音编码算法的复杂性及实时处理的需要，是促使设计这样的处理器的重要推动力量之一。这种产品问世后又首先在语音处理应用中得到有效的推广应用。语音处理产品的商品化对这样的处理器有很大需求，因此其反过来又推动了微电子技术的发展。

1.2 语音信号处理的主要研究内容及发展概况

语音信号处理有广泛的应用领域，最重要的包括语音编码、语音合成、语音识别、说话人识别、语音增强、麦克风阵列语音信号处理等。

（1）语音编码

语音编码技术是伴随语音数字化而产生的，主要应用于数字语音通信领域。语音信号的数字化传输，一直是通信的发展方向之一。语音信号的低速率编码传输比模拟传输有很多优点。直接将连续语音信号取样量化而成为数字信号，要占用较多的信道资源。因而，应在失真尽可能小的情况下，使同样道容量能够传输更多路的信号，这需要对模拟语音信号进行高效率的数字表示，即进行压缩编码；这已成为语音编码的主要内容。

在中低速率上获得高质量的语音一直是语音编码研究的主要目标。低数码率编码在无线通信、网络安全、数字电话及存储系统等方面有广泛应用。语音编码研究始于 1939 年 Dudley 发明的声码器，但直至 20 世纪 70 年代中期，除 PCM 及 ADPCM 取得较好的进展外，中低比特率语音编码一直没有大的突破。80 年代后，语音编码技术产主大的飞跃；1980 年美国公布了一种 2.4b/s 的标准编码算法，使一直期待的在普通电话带宽信道中传输数字电话的愿望成为事实，而数字电话有保密性高、易克服噪声累计、便于程控交换等优点。但上述 LPC 编码的音质不令人满意。80 年代后，提出很多新型编码算法，在 16kb/s、4.8kb/s 以至 2.4kb/s 上提供高质量语音，且均可用单片 DSP 实时实现。目前，在 2.4kbit/以上的编码速率，合成语音质量已得到认可并广泛应用。而实用系统的最低压缩速率达 2.4kb/s 甚至更低，在大大节省信道带宽的同时保证了语音质量。目前的研究是减小编解码过程产生的时延，以广泛用于移动通信。未来研究重点是 2.4kb/s 以下极低速率的语音编码技术和算法。

近年来，高质量的语音编码技术大规模实用化，各种国际标准的制定反映了其发展水平与趋势。20 世纪 70 年代推出 64kb/s 的 PCM 语音编码国际标准，以后又有 32kb/s 的 ADPCM。1980 年美国公布了 2.4kb/s 的线性预测编码标准算法 LPC-10，使在普通电话带宽信道中传输数字电话成为可能。1988 年又公布了 4.8kb/s 的 CELP 语音编码标准算法，而欧洲推出了 16kb/s 的 RPELPC，这些算法的音质都能达到很高质量，而不像 LPC 声码器的输出语音那样不为人们所接受。此外，还有 16kb/s 的 LD-CELP、8kb/s 的 CA-ACELP 等国际标准。90 年代中期，出现了很多广泛使用的语音编码国际标准，如基于 CELP 技术的 5.3/6.4kb/s 的 G.723.1、8kb/s 的 CS-ACELP(即 G.729)等。另一方面，还有一些地区性或业务性标准，如第二代移动通信系统中的语音编码，美国国防部制定的 4.8kb/s 及 2.4kb/s 保密电话标准等。同时，还有各种未形成国际标准、但数码率更低的成熟编码算法，有的在 1.2kb/s 以下仍可提供可懂的语音。

语音编码目前的研究集中在低数码率的高音质、低延迟声码器，提高噪声信道中低数码率编码器的性能，并能传输多种信号(包括音频)。为此，应采用更有效的参数矢量量化、非线性预测、多分辨率时频分析(如小波)及高阶统计量技术，并对人耳感知特性进行进一步的研究与探索等。

语音编码与通信技术的发展密切相关：现代通信的重要标志是数字化；而语音编码的根本作用是使语音通信数字化，它将使通信技术水平提高一大步。语音编码是移动通信及个人通信非常重要的支撑技术，对通信新业务的发展有十分重要的影响。同时，语音编码的产品化比语音识别容易，研究成果可很快实用化。

(2)语音合成

目前，计算机使用还不够方便，人与计算机的通信需利用键盘和显示器，效率低下且操作也不方便。因而希望计算机有智能接口，使人能够方便自然地与计算机打交道。语音是人与人、人与计算机间最方便的信息交换方式，因而人们特别期望有智能的语音接口。最理想的是，计算机有人那样的听觉功能及发音功能；从而人可用自然语言与计算机对话，使其可接收、识别并理解声、图、文信息，看懂文字、听懂语言、朗读文章，甚至进行不同语言间的翻译。智能接口技术有重大的应用价值，又有基础的理论意义，多年来一直是最活跃的研究领域。而语音识别与语音合成为人机智能接口开辟了新途径，是智能接口技术中的标志性成果，也是人工智能的重要课题。

这里，语音合成是使计算机说话，它是一种人机语音通信技术，应用领域十分广泛，且已发挥了很好的社会效益。对语音合成的社会需求十分广泛，其研究和产品开发有很好的前景。目前，有限词汇语音合成已成熟，在自动报时、报警、报站、电话查询服务等方面得到广泛应用。

最简单的语音合成是语声响应系统，其非常简单：在计算机内建立一个语言库，将可能用到的字、词组或一些句子的声音信号，编码后存入计算机；键入所需要的字、词组或句子代码时，就可调出对应的数码信号，并转换成声音。

20 世纪 70 年代末，开始对文本-语音转换系统(TTS)进行研究；其基于规则的文字-语音合成系统，将文字转换为语言，以使计算机模仿人来朗读文本。这种系统的特点是用最基本的语音单元，如音素、音节等作为合成单元，建立语音库，通过合成单元拼接达到无限词汇的合成。输入文字信息后，将其按照语言规则转换为由基本单元组成的序列；根据说话时单元连接的规则进行控制，并发出声音。为保证合成声音的音质，系统中除语音库外还有一个很庞大的规则库，实现对合成语音的音段和超音段特征的控制。20 世纪 90 年代初，文-语转换系统在很多国家、多个语种都达到商品化程度，语音质量亦被公众接受。其中，波形拼接合成方法得到越来越广泛应用，最有代表性的为基音同步叠加法(PSOLA)，在语音合成中影响较大。PSOLA 于 20 世纪 90 年代末提出，是多样本的不等长语音拼接合成技术；其在语音库中存放大量语音样本，通过选择合适的拼接语音片段实现高质量的合成语音。它可保持所发语音的主要音段特征，又能在拼接时灵活调整一些特征及参数；其将语音合成问题简化为建立一个充分的语音库，选择合适的语音片段进行拼接，及对语音片段的拼接部分进行调整的过程。

20 世纪 90 年代中期，随语音识别中统计模型方法的日益成熟，提出可训练的语音合成方法；基本思想是基于统计模型和机器学习方法，根据一定语音数据进行训练，并快速构建合成系统。随着声学合成性能的提高，在此基础上又发展了统计参数语音合成方法，其中以 HMM 的建模与合成为代表。基于 HMM 的参数语音合成无须人工干预，可快速构建合成系统，且对不同发音人、发音风格及语种依赖很小，是近年语音合成研究的热点。而更高层次的合成是概

念或意向到语音的合成；即将想法、意向组成语言并变为语音，就如大脑形成说话内容并控制发声器官产生语音那样。

目前，很多语音合成系统有较高的可懂度，但在提高自然度方面还有很大空间，这是目前研究的重点。另一方面，无限词汇语音合成的音质改善存在一定困难，还未达到完美的程度；这是当前语音合成研究的主要方向，从社会需求上看也是要迫切解决的问题。语音合成有很好的应用前景，如与机器翻译相结合，可实现语音翻译；与图像处理结合，可输出视觉语音。

（3）语音识别

语音识别就是使计算机判断出人说话的内容。语音识别的根本目的是使计算机有人那样的听觉功能，能接受人的语音、理解人的意图。语音识别与语音合成类似，也是人机语音通信技术。语音识别的研究有重要意义，特别是对于汉语来说，汉字的书写和录入较为困难，因而通过语音来输入汉字信息就特别重要。而且，用计算机键盘进行操作也不方便，因而用语音输入代替键盘输入的必要性更为突出。在计算机智能接口及多媒体的研究中，语音识别有很大应用潜力。同时，为实现人机语音通信，需要有语音识别及语音理解等两种功能。

如上说述，语音识别可用于将文字以口授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。其可用于自动口语翻译，即通过语音识别、机器翻译及语音合成等技术的结合，可将某种语言输入的语音翻译为另一种语言的语音，实现跨语言交流。

语音识别的研究比语音合成要困难得多，其起步也较晚。语音识别的研究始于 20 世纪 50 年代，目前已取得很大进展，近年不断有语音识别器(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已成熟，而大词汇量连续语音识别系统的性能需进一步改善。20 世纪 80 年代以来，语音识别研究的重点转向大词汇量非特定人连续语音识别。80 年代末实现的 997 个词的 SPHINX，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。而有代表性的是 1997 年 IBM 公司推出的 Via Voice 大词汇量连续语音识别系统，其输入速度平均每分钟达 150 字，平均最高识别率 95%，且有自学习功能。

20 世纪 90 年代以来，语音识别已从理论研究走向实用化。一方面，声学语音学统计模型的研究日益深入，鲁棒的语音识别、基于语音段的建模方法及 HMM 与神经网络的结合成为研究热点。另一方面，为适应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研究课题受到很大关注。

目前，语音识别技术距其广泛应用还存在距离。很多因素影响语音识别系统的性能，甚至使其无法工作。如实际环境中的背景噪声、传输通道的频率特性、说话人生理或心理的变化，以及应用领域的变化等。因而，鲁棒(顽健的)的语音识别方法受到广泛重视。但目前为止，所做研究多是针对一两种因素进行的补偿，而综合考虑各种因素进行补偿的研究还很少。

随着 Internet 技术的发展，出现了 Internet 电话，即 IP 电话技术。对这种经数据压缩，并由网络以数据包形式传输的语音进行识别，与传统的语音识别技术有很大不同；这就是网络环境下的语音识别问题，其在电子商务及国防军事领域有广阔的应用前景。

迄今为止，对语音识别的理论与应用已进行了广泛研究，这方面已有相当庞大数量的文献。但语音识别是一项综合性的、难度很大的研究课题，从语音中提取满意的信息是一项艰巨复杂的任务。语音识别研究中面临很多难以解决的问题与困难。目前，国内外均投入大量人力物力来解决这些问题。

（4）说话人识别

说话人识别可看作一种特殊的语音识别，它是根据语音来辨别出说话人是谁。说话人识别与语音识别类似，通过提取语音信号的特征和建立相应模型进行分类判断。但与语音识别不同，其并不注意语音信号中的语义内容，而是从语音信号中分析和提取个人特征，以去除不含

个人特征的语音信息；即力求找出包含在语音信号中的说话人的个性因素，即不同人之间的特征差异。

对说话人识别的研究，随着语音识别研究的不断深入也得到迅速发展。语音识别中很多成功的技术，如 VQ、HMM 等均被用于说话人识别。20 世纪 90 年代，提出单状态 HMM，即后来的高斯混合模型(GMM)；其与多状态 HMM 几乎有相同的识别性能，在说话人识别研究中日益受到重视。

（5）语种辨识

语种辨识是近年出现的研究领域，也可看作一种特殊的语音识别。它是从一个语音片段中判别语音属于哪个语种。语种辨识能够实现的依据是，世界上的不同语种之间有多种区别特征，因而应找出不同语种间的特征差别。语种辨识和语音识别及说话人识别系统有很多相似之处；其可用于多语言语音识别的前端处理，在信息检索、军事领域及国家安全等领域有重要应用。

（6）语音理解

语音理解是利用知识表达和组织等人工智能技术，来进行语句识别和语意理解的。其与语音识别的不同之处在于对语法和语义知识的充分利用程度。人对语音有广泛的知识，对要说的话有一定的预见性，即对语音有感知分析能力。依靠人对语言及其内容所具有的广泛知识来提高计算机理解语言的能力，是语音理解研究的核心。语音理解可看作信号处理与知识处理的产物。

（7）语音增强

实际应用环境中，语音会不同程度地受到环境噪声的干扰。因而，语音抗噪声技术的研究及实际环境下语音处理系统的开发，是语音信号处理中非常重要的研究课题。目前这一研究大体分为三类方法，即语音增强、寻找稳健的语音特征及基于模型参数适应化的噪声补偿。然而，解决噪声问题的根本方法应是噪声和语音的自动分离，但其技术难度较大。近年来，随声场景分析及盲分离技术的发展，语音和噪声分离的研究取得一定进展。

语音增强是语音抗噪声技术的一种，即对带噪语音进行处理，以尽可能去除噪声并改善听觉效果。有些语音编码和语音识别系统在无噪声或噪声很小的环境中性能很好，但环境噪声增大时，其性能将急剧下降。因而，语音增强也是语音编码及语音识别等系统实际应用中所必须解决的问题。

（8）基于麦克风阵列的语音信号处理

麦克风阵列处理是语音信号处理中的一项新技术。与单一麦克风（通道）的语音信号处理相比，麦克风阵列在时域和频域处理的基础上增加了空域处理，可对来自空间不同方向的信号进行空-时-频联合处理，以弥补单个麦克风在噪声抑制、声源定位跟踪、语音分离等方面的不足，从而广泛用于有嘈杂背景的语音通信环境。麦克风阵列处理的研究主要包括声源定位、语音增强、声源盲分离、去混响等。

麦克风阵列信号处理是阵列信号处理领域中一个新的分支，它继承和发展了阵列信号处理的理论与算法。阵列信号处理理论的发展促进了麦克风阵列信号处理的发展，很多用于阵列信号处理的方法、技术与体系可用于麦克风阵列，其为麦克风阵列处理的发展提供了动力。

20 世纪七八十年代，开始将麦克风阵列用于语音信号处理中。1985 年，Flanagan 将麦克风阵列引入大型会议的语音增强中，后来麦克风阵列又被引入语音识别系统、移动环境的语音获取、说话人识别及混响环境下的语音捕获。90 年代后，这一技术成为研究热点。1996 年被用于声源定位，以确定和实时跟踪说话人位置。麦克风阵列处理在军事上也有重要应用，如声呐系统对水下潜艇的跟踪，无源定位直升机和其他发声设备等。在国外，IBM、Bell 实验室等

试读结束：需要全本请在线购买： www.ertongbook.com