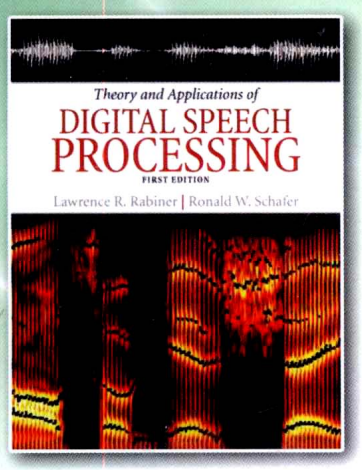
国外电子与通信教材系列

英文版

PEARSON

数字语音处理理论与应用



Theory and Applications of Digital Speech Processing

[美] Lawrence R. Rabiner
Ronald W. Schafer 著

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

数字语音处理理论与应用

（英文版）

Theory and Applications of Digital Speech Processing

[美] Lawrence R. Rabiner
Ronald W. Schafer 著

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

内容简介

本书是作者继1978年出版的经典教材 *Digital Processing of Speech Signals* 之后的又一著作，全书除有简练精辟的基础知识介绍外，系统讲解了近30年来语音信号处理的新理论、新方法和在应用上的新进展。全书共14章，分四部分：第一部分介绍语音信号处理基础知识，主要包括数字信号处理基础、语音产生机理、(人的)听觉和听感知机理，以及声道中的声传播原理；第二部分介绍语音信号的时、频域表示和分析；第三部分介绍语音参数估计算法；第四部分介绍语音信号处理的应用，主要包括语音编码、语音和音频信号的频域编码、语音合成、语音识别及自然语言理解。

本书可供高等院校通信、电子、信息、计算机等专业作为研究生和本科生的教材，也可以供有关科研和工程技术人员参考，是一本既有系统的基础理论讲解、又有最新研究前沿介绍并密切结合应用发展的教材。

版权贸易合同登记号 图字：01-2010-7880

图书在版编目（CIP）数据

数字语音处理理论与应用：英文/（美）拉比纳（Rabiner, L. R.），（美）谢弗（Schafer, R. W.）著.

北京：电子工业出版社，2011.1

（国外电子与通信教材系列）

书名原文：Theory and Applications of Digital Speech Processing

ISBN 978-7-121-12409-9

Ⅰ.①数… Ⅱ.①拉… ②谢… Ⅲ.①语音数据处理－高等学校－教材－英文 Ⅳ.TN912.3

中国版本图书馆 CIP 数据核字（2011）第231370号

策划编辑：马 岚

责任编辑：冯小贝

印 刷：北京市顺义兴华印刷厂

装 订：三河市双峰印刷装订有限公司

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 × 980 1/16 印张：66.25 字数：1484千字 彩插：2

印 次：2011年1月第1次印刷

定 价：118.00元

凡所购买电子工业出版社的图书有缺损问题，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。

序

2001年7月间，电子工业出版社的领导同志邀请各高校十几位通信领域方面的老师，商量引进国外教材问题。与会同志对出版社提出的计划十分赞同，大家认为，这对我国通信事业、特别是对高等院校通信学科的教学工作会很有好处。

教材建设是高校教学建设的主要内容之一。编写、出版一本好的教材，意味着开设了一门好的课程，甚至可能预示着一个崭新学科的诞生。20世纪40年代MIT林肯实验室出版的一套28本雷达丛书，对近代电子学科、特别是对雷达技术的推动作用，就是一个很好的例子。

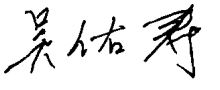
我国领导部门对教材建设一直非常重视。20世纪80年代，在原教委教材编审委员会的领导下，汇集了高等院校几百位富有教学经验的专家，编写、出版了一大批教材；很多院校还根据学校的特点和需要，陆续编写了大量的讲义和参考书。这些教材对高校的教学工作发挥了极好的作用。近年来，随着教学改革不断深入和科学技术的飞速进步，有的教材内容已比较陈旧、落后，难以适应教学的要求，特别是在电子学和通信技术发展神速、可以讲是日新月异的今天，如何适应这种情况，更是一个必须认真考虑的问题。解决这个问题，除了依靠高校的老师和专家撰写新的符合要求的教科书外，引进和出版一些国外优秀电子与通信教材，尤其是有选择地引进一批英文原版教材，是会有好处的。

一年多来，电子工业出版社为此做了很多工作。他们成立了一个“国外电子与通信教材系列”项目组，选派了富有经验的业务骨干负责有关工作，收集了230余种通信教材和参考书的详细资料，调来了100余种原版教材样书，依靠由20余位专家组成的出版委员会，从中精选了40多种，内容丰富，覆盖了电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等方面，既可作为通信专业本科生和研究生的教学用书，也可作为有关专业人员的参考材料。此外，这批教材，有的翻译为中文，还有部分教材直接影印出版，以供教师用英语直接授课。希望这些教材的引进和出版对高校通信教学和教材改革能起一定作用。

在这里，我还要感谢参加工作的各位教授、专家、老师与参加翻译、编辑和出版的同志们。各位专家认真负责、严谨细致、不辞辛劳、不怕琐碎和精益求精的态度，充分体现了中国教育工作者和出版工作者的良好美德。

随着我国经济建设的发展和科学技术的不断进步，对高校教学工作会不断提出新的要求和希望。我想，无论如何，要做好引进国外教材的工作，一定要联系我国的实际。教材和学术专著不同，既要注意科学性、学术性，也要重视可读性，要深入浅出，便于读者自学；引进的教材要适应高校教学改革的需要，针对目前一些教材内容较为陈旧的问题，有目的地引进一些先进的和正在发展中的交叉学科的参考书；要与国内出版的教材相配套，安排好出版英文原版教材和翻译教材的比例。我们努力使这套教材能尽量满足上述要求，希望它们能放在学生们的课桌上，发挥一定的作用。

最后，预祝“国外电子与通信教材系列”项目取得成功，为我国电子与通信教学和通信产业的发展培土施肥。也恳切希望读者能对这些书籍的不足之处、特别是翻译中存在的问题，提出意见和建议，以便再版时更正。



中国工程院院士、清华大学教授

“国外电子与通信教材系列”出版委员会主任

出 版 说 明

进入21世纪以来，我国信息产业在生产和科研方面都大大加快了发展速度，并已成为国民经济发展的支柱产业之一。但是，与世界上其他信息产业发达的国家相比，我国在技术开发、教育培训等方面都还存在着较大的差距。特别是在加入WTO后的今天，我国信息产业面临着国外竞争对手的严峻挑战。

作为我国信息产业的专业科技出版社，我们始终关注着全球电子信息技术的发展方向，始终把引进国外优秀电子与通信信息技术教材和专业书籍放在我们工作的重要位置上。在2000年至2001年间，我社先后从世界著名出版公司引进出版了40余种教材，形成了一套“国外计算机科学教材系列”，在全国高校以及科研部门中受到了欢迎和好评，得到了计算机领域的广大教师与科研工作者的充分肯定。

引进和出版一些国外优秀电子与通信教材，尤其是有选择地引进一批英文原版教材，将有助于我国信息产业培养具有国际竞争能力的技术人才，也将有助于我国国内在电子与通信教学工作中掌握和跟踪国际发展水平。根据国内信息产业的现状、教育部《关于“十五”期间普通高等教育教材建设与改革的意见》的指示精神以及高等院校老师们反映的各种意见，我们决定引进“国外电子与通信教材系列”，并随后开展了大量准备工作。此次引进的国外电子与通信教材均来自国际著名出版商，其中影印教材约占一半。教材内容涉及的学科方向包括电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等，其中既有本科专业课程教材，也有研究生课程教材，以适应不同院系、不同专业、不同层次的师生对教材的需求，广大师生可自由选择和自由组合使用。我们还将与国外出版商一起，陆续推出一些教材的教学支持资料，为授课教师提供帮助。

此外，“国外电子与通信教材系列”的引进和出版工作得到了教育部高等教育司的大力支持和帮助，其中的部分引进教材已通过“教育部高等学校电子信息科学与工程类专业教学指导委员会”的审核，并得到教育部高等教育司的批准，纳入了“教育部高等教育司推荐——国外优秀信息科学与技术系列教学用书”。

为做好该系列教材的翻译工作，我们聘请了清华大学、北京大学、北京邮电大学、南京邮电大学、东南大学、西安交通大学、天津大学、西安电子科技大学、电子科技大学、中山大学、哈尔滨工业大学、西南交通大学等著名高校的教授和骨干教师参与教材的翻译和审校工作。许多教授在国内电子与通信专业领域享有较高的声望，具有丰富的教学经验，他们的渊博学识从根本上保证了教材的翻译质量和专业学术方面的严格与准确。我们在此对他们的辛勤工作与贡献表示衷心的感谢。此外，对于编辑的选择，我们达到了专业对口；对于从英文原书中发现的错误，我们通过与作者联络、从网上下载勘误表等方式，逐一进行了修订；同时，我们对审校、排版、印制质量进行了严格把关。

今后，我们将进一步加强同各高校教师的密切关系，努力引进更多的国外优秀教材和教学参考书，为我国电子与通信教材达到世界先进水平而努力。由于我们对国内外电子与通信教育的发展仍存在一些认识上的不足，在选题、翻译、出版等方面的工作中还有许多需要改进的地方，恳请广大师生和读者提出批评及建议。

电子工业出版社

教材出版委员会

主　任	吴佑寿	中国工程院院士、清华大学教授
副主任	林金桐	北京邮电大学校长、教授、博士生导师
	杨千里	总参通信部副部长，中国电子学会会士、副理事长 中国通信学会常务理事、博士生导师
委　员	林孝康	清华大学教授、博士生导师、电子工程系副主任、通信与微波研究所所长 教育部电子信息科学与工程类专业教学指导分委员会委员 清华大学深圳研究生院副院长
	徐安士	北京大学教授、博士生导师、电子学系主任
	樊昌信	西安电子科技大学教授、博士生导师 中国通信学会理事、IEEE 会士
	程时昕	东南大学教授、博士生导师
	郁道银	天津大学副校长、教授、博士生导师 教育部电子信息科学与工程类专业教学指导分委员会委员
	阮秋琦	北京交通大学教授、博士生导师 计算机与信息技术学院院长、信息科学研究所所长 国务院学位委员会学科评议组成员
	张晓林	北京航空航天大学教授、博士生导师、电子信息工程学院院长 教育部电子信息科学与电气信息类基础课程教学指导分委员会副主任委员 中国电子学会常务理事
	郑宝玉	南京邮电大学副校长、教授、博士生导师 教育部电子信息科学与工程类专业教学指导分委员会副主任委员
	朱世华	西安交通大学副校长、教授、博士生导师 教育部电子信息科学与工程类专业教学指导分委员会副主任委员
	彭启琮	电子科技大学教授、博士生导师
	毛军发	上海交通大学教授、博士生导师、电子信息与电气工程学院副院长 教育部电子信息与电气学科教学指导委员会委员
	赵尔沅	北京邮电大学教授、《中国邮电高校学报（英文版）》编委会主任
	钟允若	原邮电科学研究院副院长、总工程师
	刘　彩	中国通信学会副理事长兼秘书长，教授级高工 信息产业部通信科技委副主任
	杜振民	电子工业出版社原副社长
	王志功	东南大学教授、博士生导师、射频与光电集成电路研究所所长 教育部高等学校电子电气基础课程教学指导分委员会主任委员
	张中兆	哈尔滨工业大学教授、博士生导师、电子与信息技术研究院院长
	范平志	西南交通大学教授、博士生导师、信息科学与技术学院院长

前　言

70多年以来，语音信号处理一直是一个活跃的并不断发展的领域。最早的语音处理系统是模拟系统，如20世纪30年代由Homer Dudley及其同事们在Bell实验室开发的Voder（语音演示记录器）系统，该系统可通过手工操作合成出语音，并于1939年在纽约世博会上展出；同一年代同样在Bell实验室，Homer Dudley还开发出了通道声码器或称为声音编码器；20世纪40年代在Bell实验室，Koenig及其同事们开发出了声音语谱图系统，该系统可以在时域和频域展示语音的时变特征；另外，20世纪50年代在全世界很多研究实验室都开发出了早期语音单词识别系统。

数字信号处理（DSP）起源于20世纪60年代，在DSP应用的广泛领域中，语音处理是其早期发展的驱动力。在此期间，先驱研究者们如MIT Lincoln实验室的Ben Gold和Charlie Rader，Bell实验室的Jim Flanagan、Roger Golden和Jim Kaiser，他们开始研究数字滤波器的设计和应用方法，并用于语音处理系统的模拟仿真。随着1965年Jim Cooley和John Tukey的快速傅里叶变换（FFT）技术的面世，以及之后FFT在快速卷积和谱分析方面的广泛应用，模拟技术的束缚和局限逐渐被打破，数字语音处理随之产生并展现出一种清晰的面貌。

本书的作者（LRR和RWS）从1968年至1974年在Bell实验室一起密切地工作，期间DSP领域发生了很多基础性的进展。当RWS 1975年离开Bell实验室并在Georgia Tech任学术职位时，数字语音处理领域已经蓬勃发展，于是我们觉得写一本关于语音信号数字处理方法和系统的教材的时机到了。到1976年，我们相信数字语音处理的理论发展得已经足够完备，精心撰写一本教材不但可以作为传授数字语音处理基础知识的教材，还可以作为将来语音处理实际应用系统设计的参考书。1978年Prentice Hall出版了这本教材《*Digital Processing of Speech Signals*》（语音信号数字处理）。采用这本教材，RWS在其新职位上开设了第一门数字语音处理的研究生课程，期间LRR仍在Bell实验室从事数字语音处理的基础研究工作。（LRR在AT&T Bell实验室和AT&T实验室工作了40年，2002年他也加入了学术界，在Rutgers大学和California大学Santa Barbara分校任教。RWS在Georgia Tech工作了30年，于2004年加入了Hewlett Packard实验室。）

1978年出版的教材的目标是呈现语音的基础科学和一系列数字语音处理方法，用以构建强大的语音信号处理系统。从大的方面来讲，我们达到了最初的目标。这本教材按我们的预想服务了30多年，令人高兴的是，直到今天它仍然广泛应用于本科生和研究生的语音信号处理课的教学中。然而，根据过去20年来教授语音处理课程的经验来看，原书的基础尚可，但其中的很多素材已经和当代语音信号处理系统脱钩，对当前很多研究热点方向也没有涉及。这本新书正是改正这些弱点的一个尝试。

在着手处理统一数字语音处理现行理论和实践这样艰巨的任务时，我们发现原书中的很多内容还是正确和相关的，所以有一个很好的起点来开始这本新书。进一步，从语音处理的科研和教学实际经验中了解到，1978年的教材的组织材料虽然基本不错，但它已经不适合用来理解当代的语音处理系统。针对这些弱点，我们采用了新的框架来组织新书的材料，对比原书有两大框架上的改变。首先，新书

中包含了已有的数字语音处理知识体系结构的概念。这种体系的第一层是关于语音基础科学和工程方面的基础知识，第二层是集中在语音信号的各种表示。原书主要侧重了这两层，但是一些关键的主题有所缺失。第三层是操作、处理和抽取语音信号中信息的各种算法，这些算法是基于前面两个底层的科学和技术知识。顶层（也就是第四层）是语音处理算法的各种应用，以及处理语音通信系统中问题的技术。

我们努力沿着这种体系结构［第1章中称为语音金字塔（语音堆）］来展现新书的材料。为了达到这样目的，在第2~5章中，集中在金字塔底层构建一个坚实的基础，包括语音产生和感知基础知识、DSP基础知识回顾，以及声学、语音学、语言学、语音感知、声道中声音传播的讨论等。在第6~9章，了解如何通过基本信号处理原理对数字语音信号进行不同的（短时）表示（构成了语音金字塔的第二层）。在第10章，展示了如何设计可靠和稳健的语音算法来估计感兴趣的语音参数（构成了语音金字塔第三层的基础）。最后，在第11~14章，展示了如何利用语音金字塔前面几层的知识设计和实现各种语音应用（构成了语音金字塔的第四层）。

新书在结构和行文上的另外一个主要变化是为了尽可能地方便教学，我们在呈现材料的同时侧重学习新思想的三个方面，即理论、概念和实现。于是对于本书介绍的每一个基本概念，都用很容易理解的DSP概念进行理论阐释；类似地，为了加深理解，每一个新概念都提供了简单的数学解释和精心准备的例子及插图；最后，基于教学中对基础知识的理解，针对每一个新概念的实现都提供了可实现特定的语音处理操作的MATLAB参考代码（通常包含在每一章中），每章习题中配备了带有详尽文档作为课外练习的MATLAB习题。我们还在教学网站上提供了解决所有MATLAB习题所需的材料，包括一些特定的MATLAB代码、访问简单数据库、访问一系列的语音文件等。最后我们提供了几种语音处理系统结果的音频演示。通过这种方式，读者可以获得各种语音信号操作处理后语音质量方面的直观感觉。

更具体地讲，这本新书的组织结构如下。第1章大体介绍语音处理的领域，并对贯穿本书主题相关的应用领域进行了简要的讨论。第2章简要回顾了DSP中概念，侧重于与语音处理系统中密切相关的几个关键概念：

1. 从时域到频域的转换（通过离散时间傅里叶变换方法）；
2. 理解采样在频域的影响（即时域的混叠）；
3. 理解采样在时域的影响（包括降采样和升采样），以及在频域的混叠和镜像。

在回顾DSP技术的基础之后，第3章和第4章转到了对语音产生和感知基础的讨论。这两章与第2章和第5章一起，构成了语音金字塔的底层。从这里，开始讨论语音产生的声学理论，对不同的语音发音，我们导出了一系列声学语音模型，并展示了语言学和语用学如何与语音发声学一起相互作用生成语音信号及其在语言上的解释。从讨论语音在人耳中如何处理开始，到声音转换为通往大脑的听感知神经通路中的神经信号结束，通过分析语音感知过程，完成了潜藏在语音通信背后的基础过程的讨论。我们简要地讨论了几种在一些语音处理应用中可能嵌入语音感知知识到听感知模型的方法。接着，在第5章，讨论了关于人类声音在声道中传播问题的基础知识。我们展示了和声道相似的均匀无损声管具有共振结构，以此阐明了语音中的共振（共振峰）频率。我们展示了如何通过适当的“终端模拟”数字系统表示一系列级联声管的传播特性。该“终端模拟”数字系统具备了特定的激励函数、对应不同长度和面积的声管的特定系统响应，以及对应声音在口唇端传输的特定辐射特征。

本书接下来的四章介绍主要的四种数字语音信号的表示（语音金字塔的第二层），每章介绍一种。首先在第6章，从语音产生的时域模型开始，逐步展示如何通过简单的、基于时域测量的方法估计模型中基本的时变属性。在第7章，展示了短时傅里叶分析概念如何以一种简单而一致的方式应用于语音信号，以至于可以实现一种完全透明（无失真）的分析/合成系统。取决于要进一步处理信息的性质，我们展示了两种短时傅里叶分析/合成系统的解释，两者都有着广泛的应用。在第8章，描述了语音的同态（倒谱）表示，其中用到了卷积信号（如语音）可以转换为一系列加性分量这一性质。基于语音信号可以表示为激励信号和声道系统的卷积认识，很容易明白语音信号非常适合这种分析。最后在第9章，涉及了线性预测分析的理论和实践，线性预测是语音信号的一种模型表示，当前的语音采样可以通过先前p个语音采样的线性组合建模表示，通过寻找最优线性预测器（最小均方误差）的系数，实现在给定一段时间内最优的匹配语音信号。

第10章，代表语音金字塔的第三层，涉及到使用前面章节介绍的信号处理的表示和语音信号的基础知识，作为测量或估计语音信号性质和属性的基础。这里展示了短时（对数）能量、短时过零率、短时自相关函数这些测量值如何用来估计基本的语音属性，例如分析的信号段是语音还是静音（背景信号），语音段是浊音还是清音，浊音语音段的基音周期（基音频率），语音段的共振峰（声道共振），等等。对于许多语音属性，展示了四种语音表示的每一种都可以作为估计语音属性的高效算法的基础使用。与此相似，还展示了如何基于四种语音表示中的两种测量法来估计共振峰。

第11~14章代表语音金字塔的顶层（语音应用），涉及到几种主要的语音和音频信号处理技术应用。这些应用是深入理解语音和音频技术的成果，它们代表了人们几十年来研究如何最好地综合各种语音表示和测量方法，使每一种语音应用都能给出最好性能。我们讨论语音应用的目标是给读者如何构建这些应用提供一种感觉，它们在不同比特率和不同应用场景的性能如何。具体来讲，第11章涉及语音编码系统（包括开环和闭环系统）；第12章涉及基于人们熟知的感知掩蔽准则最小编码感知误差的音频编码系统；第13章涉及构建用于口语对话系统中的文语转换的语音合成系统；第14章处理语音识别和自然语言处理系统，以及它们在一系列面向任务的场景中的应用。我们在这些章的目标是提供最新的例子，但不求全面覆盖。关于这些应用每一个领域已经有很多教材出版。

在学生具有基本DSP基础的前提下，这本书的材料可以作为一个学期的语音处理课程来讲授。在我们自己的教学实践中，重点强调第3~11章，并选取其他章节的部分材料进行授课，使学生对音频编码、语音合成和语音识别系统的也有一定的认识。为了辅助教学过程，每章都配有一套有代表性的课后习题，用于强化每章所讨论的思想。如前所述，成功完成一定比例的课后习题对理解语音处理的数学和理论概念非常重要。然而，也正如读者看到的一样，很多的语音处理是经验性的，这一点是由其本质决定的。于是我们包含了一系列MATLAB习题（或者作为正文，或者作为习题的一部分）来强化学生对语音处理基本概念的理解。我们也提供了教学网站（http://www.pearsonhighered/Rabiner.com）并随时更新材料，包括所需的语音文件、数据库和解决MATLAB习题的MATLAB代码，以及一系列语音处理概念的演示。

致　谢

在语音处理的整个职业生涯中，我们非常幸运在拥有杰出的研究和学术机构的单位工作，这些单位提供了充满激情的研究环境并且鼓励分享知识。对于LRR，这些单位包括Bell实验室、AT&T实验室、Rutgers大学和California大学Santa Barbara分校；对于RWS，这些单位包括Bell实验室、Georgia Tech ECE和Hewlett-Packard实验室。没有这些单位的同事和领导的支持与鼓励，这本书不会存在。

很多人对这本书展现的内容有直接或间接的重大影响，但最应感谢的是James L. Flanagan博士，他是我们两人职业生涯中很多关键点的导师和益友。Jim为我们如何从事科研、如何清晰合理地呈现研究结果提供了鼓舞人心的楷模。无论是对这本书还是我们各自的职业，他的影响都是非常深远的。

我们有幸合作和互相学习的其他人，包括我们的导师MIT的Alan Oppenheim教授和Kenneth Stevens教授，以及我们的同事Georgia Tech的Tom Barnwell教授、Mark Clements教授、Chin Lee教授、Fred Juang教授、Jim McClellan教授和Russ Mersereau教授。这些人既是我们的同事，又是我们的老师，我们感激他们的睿智和多年来的指导。

直接参与本书准备工作的同事包括Bishnu Atal博士、Victor Zue教授、Jim Glass教授和Peter Noll教授，他们每人都提供了具有深刻见解的和高度技术含量的成果，这些成果对展现在本书中的很多内容都产生了巨大的影响。其他人允许我们使用其发表物中的图表，他们包括Alex Acero、Joe Campbell、Raymond Chen、Eric Cosatto、Rich Cox、Ron Crochiere、Thierry Dutoit、Oded Ghitza、Al Gorin、Hynek Hermansky、Nelson Kiang、Rich Lippman、Dick Lyon、Marion Macchi、John Makhoul、Mehryar Mohri、Joern Ostermann、David Pallett、Roberto Pieraccini、Tom Quatieri、Juergen Schroeter、Stephanie Seneff、Malcolm Slaney、Peter Vary和Vishu Viswanathan。

我们感谢Lucent-Alcatel、IEEE、美国声学学会及House-Ear Institute提供的支持，允许使用已发表或存档的图表。

同时，我们感谢Pearson Prentice Hall那些帮助本书出版的人们，包括策划编辑Andrew Gilfillan、生产编辑Clare Romeo和助理编辑William Opaluch。我们也感谢TexTech International负责文字编辑和校对工作的Maheswari PonSaravanan。

最后，我们感谢赞助商Suzanne和Dorothy，感谢他在我们看似无休止的写书期间的关爱、耐心和支持。

Lawrence R. Rabiner
Ronald W. Schafer

Contents

CHAPTER 1

Introduction to Digital Speech Processing

This book is about subjects that are as old as the study of human language and as new as the latest computer chip. Since before the time of Alexander Graham Bell's revolutionary invention of the telephone, engineers and scientists have studied the processes of speech communication with the goal of creating more efficient and effective systems for human-to-human and (more recently) human-to-machine communication. In the 1960s, digital signal processing (DSP) began to assume a central role in speech communication studies, and today DSP technology enables a myriad of applications of the knowledge that has been gained over decades of research. In the interim, concomitant advances in integrated circuit technology, DSP algorithms, and computer architecture have aligned to create a technological environment with virtually limitless opportunities for innovation in speech processing as well as other fields such as image and video processing, radar and sonar, medical diagnosis systems, and many areas of consumer electronics. It is important to note that DSP and speech processing have evolved hand-in-hand over the past 50 years or more, with speech applications stimulating many advances in DSP theory and algorithm research and those advances finding ready applications in speech communication research and technology. It is reasonable to expect that this symbiotic relationship will continue into the indefinite future.

In order to fully appreciate a technology such as digital speech processing, three levels of understanding must be reached: the theoretical level (theory), the conceptual level (concepts), and the working level (practice). This technology pyramid is depicted in Figure 1.1.[1] In the case of speech technology, the theoretical level is comprised of the acoustic theory of speech production, the basic mathematics of speech signal representations, the derivations of various properties of speech associated with each representation, and the basic signal processing mathematics that relates the speech signal to the real world via sampling, aliasing, filtering, etc. The conceptual level is concerned with how speech processing theory is applied in order to make various speech measurements and to estimate and quantify various attributes of the speech signal. Finally, for a technology to realize its full potential, it is essential to be able to

[1] We use the term "technology pyramid" (rather than "technology triangle") to emphasize the fact that each layer has breadth and depth and supports all higher layers.

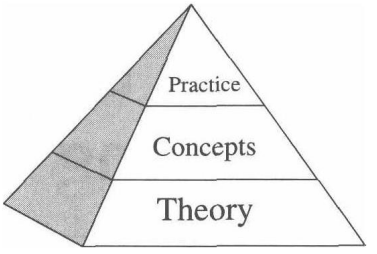


FIGURE 1.1
The technology pyramid—theory, concepts, and practice.

convert theory and conceptual understanding to practice; that is to be able to implement speech processing systems that solve specific application problems. This process involves knowledge of the constraints and goals of the application, engineering tradeoffs and judgments, and the ability to produce implementations in working computer code (most often as a program written in MATLAB®, C, or C++) or as specialized code running on real-time signal processing chips [e.g., application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs), or DSP chips].

The continuous improvement of digital implementation technology capabilities, in turn, opens up new application areas that once were considered impossible or impractical, and this is certainly true in digital speech processing. Therefore, our approach in this book is to emphasize the first two levels, but always to keep in mind the ultimate technology payoff at the third (implementation) level of the technology pyramid. The fundamentals and basic concepts of the field of digital speech processing will certainly continue to evolve and expand, but what has been learned in the past 50 years will continue to be the basis for the applications that we will see in the next decades. Therefore, for every topic in speech processing that is covered in this book, we will endeavor to provide as much understanding as possible at the theory and concepts level, and we will provide a set of exercises that enable the reader to gain expertise at the practice level (usually via MATLAB exercises included within the problems at the end of each chapter).

In the remainder of this introductory chapter we begin with an introduction to the speech communication process and the speech signal and conclude with a survey of the important application areas for digital speech processing techniques. The remainder of the text is designed to provide a solid grounding in the fundamentals and to highlight the central role of DSP techniques in modern speech communication research and applications. Our goal is to present a comprehensive overview of digital speech processing that ranges from the basic nature of the speech signal, through a variety of methods of representing speech in digital form, to overviews of applications in voice communication and automatic synthesis and recognition of speech. In the process we hope to provide answers to such questions as:

- What is the nature of the speech signal?
- How do DSP techniques play a role in learning about the speech signal?

- What are the basic digital representations of speech signals, and how are they used in algorithms for speech processing?
- What are the important applications that are enabled by digital speech processing methods?

We begin our study by taking a look at the speech signal and getting a feel of its nature and properties.

1.1 THE SPEECH SIGNAL

The fundamental purpose of speech is human communication; i.e., the transmission of messages between a speaker and a listener. According to Shannon's information theory [364], a message represented as a sequence of discrete symbols can be quantified by its *information content* in bits, where the rate of transmission of information is measured in bits per second (bps). In speech production, as well as in many human-engineered electronic communication systems, the information to be transmitted is encoded in the form of a continuously varying (analog) waveform that can be transmitted, recorded (stored), manipulated, and ultimately decoded by a human listener. The fundamental analog form of the message is an acoustic waveform that we call the *speech signal*. Speech signals, such as the one illustrated in Figure 1.2, can be converted to an electrical waveform by a microphone, further manipulated by both analog and digital signal processing methods, and then converted back to acoustic form by a loudspeaker, a telephone handset, or headphone, as desired. This form of speech processing is, of course, the basis for Bell's telephone invention as well as today's multitude of devices for recording, transmitting, and manipulating speech and audio signals. In Bell's own words [47],

> Watson, if I can get a mechanism which will make a current of electricity vary its intensity as the air varies in density when sound is passing through it, I can telegraph any sound, even the sound of speech.

Although Bell made his great invention without knowing about information theory, the principles of information theory have assumed great importance in the design of sophisticated modern digital communications systems. Therefore, even though our main focus will be mostly on the speech waveform and its representation in the form of parametric models, it is nevertheless useful to begin with a discussion of the information that is encoded in the speech waveform.

Figure 1.3 shows a pictorial representation of the complete process of producing and perceiving speech—from the formulation of a message in the brain of a speaker, to the creation of the speech signal, and finally to the understanding of the message by a listener. In their classic introduction to speech science, Denes and Pinson appropriately referred to this process as the "speech chain" [88]. A more refined block diagram representation of the speech chain is shown in Figure 1.4. The process starts in the upper left as a message represented somehow in the brain of the speaker. The message information can be thought of as having a number of different representations during the process of speech production (the upper path in Figure 1.4). For example