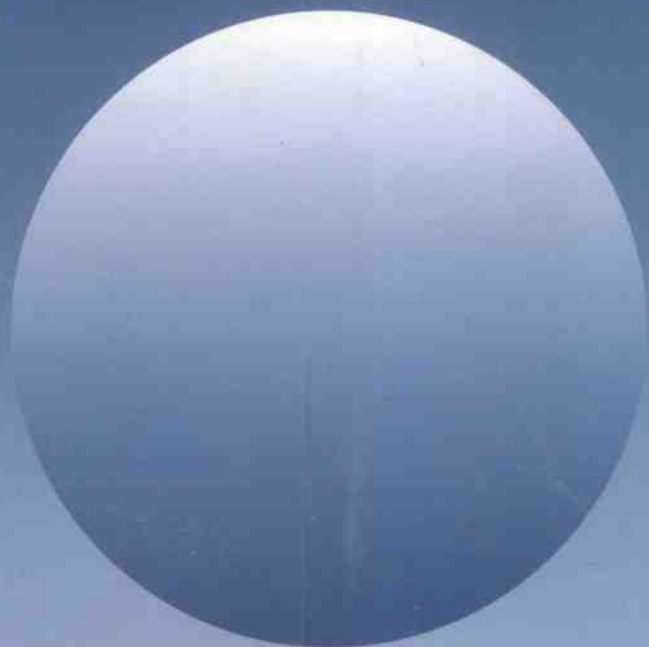


● 研 究 生 用 书 ●

DIGITAL SPEECH
PROCESSING

华中科技大学出版社



姚天任

室
导航、制导与控制

数字语音处理

数字语音处理

姚天任

华中科技大学出版社
(中国·武汉)

中国版本图书馆 CIP 数据核字(2007)第 125027 号

数字语音处理

姚天任

责任编辑 李凤英

责任校对 卢金锋

*

华中科技大学出版社出版发行

(武昌喻家山)

湖北省京山德兴印刷有限公司印刷

*

开本:850mm×1168mm 1/32 印张:13 插页:2 字数:316 000

1992 年 4 月第 1 版 2007 年 8 月第 7 次印刷

ISBN 978-7-5609-0618-8/TN·20

定价:20.00 元

(本书若有印装质量问题,请向出版社发行部调换)

内 容 简 介

本书系统全面地讨论了语音信号数字处理的理论基础、各种方法和某些重要应用领域。

全书十章,分别论及语音信号的基本性质和数字模型,短时时域处理技术,短时傅里叶分析,语音波形数字编码,同态语音处理和倒谱分析,线性预测编码,矢量量化,隐马尔柯夫模型,以及语音压缩、语音合成、语音识别和语音增强等典型应用领域。

本书主要是为研究生和本科高年级学生写的,但也可作为工程技术人员和科学研究工作者的一本有用的参考书。

ABSTRACT

This book discusses systematically and comprehensively the theoretical basis, the various methods, and some important application fields of the digital processing of speech signals.

The book is divided into ten chapters dealing, respectively, with the fundamental features and the digital model of speech signal, the short-time time-domain processing techniques, the short-time Fourier analysis, the digital coding of speech waveforms, the homomorphic speech processing and cepstral analysis, the linear predictive coding, the vector quantization, the hidden Markov models, as well some typical application areas such as speech compression, speech synthesis, speech recognition, and speech enhancement.

This book is primarily written for graduate or senior students, but it may also be used as a useful reference by practicing engineering professionals and scientific researchers.

写在“研究生用书”出版 10 周年

在今天,面对科技的迅速发展,知识经济的已见端倪,国际竞争也日趋激烈,显然,国家之间的竞争是国家综合实力的竞争,国家综合实力的竞争关键是经济实力的竞争,而经济实力的竞争关键又在于科技(特别是高科技)的竞争,科技(特别是高科技)的竞争归根结底是人才(特别是高层次人才)的竞争,而人才(特别是高层次人才)的竞争基础又在于教育。“百年大计,教育为本;国家兴亡,人才为基。”十六个字、四句话,确是极其深刻的论断。目前,国际形势清楚表明:我们国家的强大与民族的繁荣,主要立足于自己,以“自力更生”为主;把希望寄托于他人,只是一种不切实际的幻想。这里,我们决不是要再搞“闭关锁国”,搞“自我封闭”,因为那是没有出路的;我们强调的是要“自信,自尊,自立,自强”,要“自力更生”为主,走自己发展的道路。

显然,知识经济最关键的是人才,是高层次人才的培养,而作为高层次人才培养的研究生教育就在一个国家的方方面面的工作中,占有十分重要的战略地位。可以说,没有研究生教育,就没有威伟雄壮的科技局面,就没有国家的强大实力,就没有国家在国际上的位置,就会挨打,就会受压,就会被淘汰,还说什么知识经济与国家强大?!

“工欲善其事,必先利其器。”教学用书是教学的重要

基本工具与条件。这是所有从事教育的专家所熟知的事实。所以,正如许多专家所知,也正是原来的《“研究生用书”总序》中所指出,研究生教材建设是保证与提高研究生教学质量的重要环节,是一项具有战略性的基本建设。没有研究生的质量,就没有研究生教育的一切。

我校从1978年招收研究生以来,即着力从事于研究生教材与教学用书的建设。积十多年建设与实践的经验,我校从1989年起,正式分批出版“研究生用书”。第一任研究生院院长陈珏教授就为之写了《“研究生用书”总序》,表达了我校编写这套用书的指导思想与具体要求,“要力求‘研究生用书’具备科学性、系统性、先进性”。后三任研究生院院长,也就是各任校长黄树槐教授、我本人和周济教授完全赞同这一指导思想与具体要求,从多方面对这套用书加以关心与支持。

我是十分支持出版“研究生用书”的。早在1988年我在为列入这套书中的第一本,即《机械工程测试·信息·信号分析》写“代序”时就提出:一个研究生应该博览群书,博采百家,思路开阔,有所创见。但这不等于他在一切方面均能如此,有所不为才能有所为。如果一个研究生的主要兴趣与工作不在“这一特定方面”,他也可以选择一本有关的书作为了解与学习这方面专业知识的参考;如果一个研究生的主要兴趣在“这一特定方面”,他更应选择一本有关的书作为主要学习用书,寻觅主要学习线索,并缘此展开,博览群书。这就是我赞成为研究生编写系列教学用书的原因。

目前,这套书自第一本于1990年问世以来,已经渡

过了10个春秋,出版了8批共49种,初步形成规模,逐渐为更多读者所认可。在已出版的书中,有15种分获国家级、省部级图书奖,有16种一再重印,久销不衰。采用此套书的一些兄弟院校教师纷纷来信,赞誉此书为研究生培养与学科建设作出了贡献,解决了他们的“燃眉之急”。我们感谢这些赞誉与鼓励,并将这些作为对我们的鞭策与鼓励,“衷心藏之,何日忘之?!”

现在,正是江南春天,“最是一年春好处”。华工园内,红梅怒放,迎春盛开,柳枝油绿,梧叶含苞,松柏青翠,樟桂换新,如同我们的国家正在迅猛发展、欣欣向荣一样,一派盎然生机。尽管春天还有乍寒时候,我们国家在前进中还有种种困难与险阻,来自国内与来自国外的阻挠与干扰,有的还很严峻;但是,潮流是不可阻挡的,春意会越来越浓,国家发展会越来越好。我们教师所编的、所著的、所编著的这套教学用书,也会在解决前进中的种种问题中继续发展。然而,我们十分明白,这套书尽管饱含了我们教师的辛勤的长期的教学与科研工作的劳动结晶,作为教学用书百花园中的一丛鲜花正在怒放,然而总会有这种或那种的不妥、错误与不足,我衷心希望在这美好的春日,广大的专家与读者,不吝拔冗相助,对这套教学用书提出批评建议,予以指教启迪,为这丛鲜花除害灭病,抗风防寒,以进一步提高质量,提高水平,更上一层楼,我们不胜感激。我们深知,“一个篱笆三个桩”,没有专家的指导与支持,没有读者的关心与帮助,也就没有这套教学用书的今天。我衷心祝愿在我们学校第三次大发展的今天,在百年之交与千年之交的时候,这套教学用书会以更

雄健的步伐，走向更美好的未来。

诗云：“嘒其鸣矣，求其友声。”这是我们的心声。

中国科学院院士

华中理工大学学术委员会主任

杨叔子

于华工园内

1999年5月15日

前 言

未来的历史学家或许会把20世纪70和80年代称为计算机革命的时代。这是计算机技术迅猛发展并得到普及应用的时期。但是,为了使用计算机,人们必须首先学会计算机所使用的语言。这意味着,人必须去适应计算机,因此,只有部分专家或受过专门训练的人才能够使用计算机。人们盼望着改变这种局面,盼望着计算机具有“听”、“说”人类语言的能力。这样,就可以要求计算机来适应人类。如果计算机能够用人类语言和人进行通话,至少会带来以下好处:第一,使用计算机的人不必具有计算机方面的专门知识和技能,任何具有听说能力的人都能得心应手地使用计算机;第二,不需要专门的终端设备,只要通过现有的电话线路,甚至坐在行驶着的汽车里,也可以随心所欲地让计算机为你服务;第三,人们在使用计算机的同时,能够把自己的手和眼腾出来集中到主要工作上,而与计算机打交道的事就让嘴和耳去对付就够了。这正是数字语音处理的一个目的。

曾几何时,人们还认为制造“口授打字机”只是一种梦想。可是今天,人类智慧创造性的劳动终于开始结出一些硕果。人们欣喜地看到,数字语音处理的研究成果正从实验室逐步走向商品市场,走向非常广阔的应用领域。近些年来,各种语音识别器和语音合成器陆续投放市场,甚至在儿童的学习工具和玩具上也得到了体现。当然,这些成果远不是语音处理研究的最终目标。为了能够达到更加光辉的顶点,科学家们正在那崎岖而又漫长的道路上不畏艰辛地前赴后继地勇敢攀登。

数字语音处理是一个跨学科、综合性的研究领域,涉及到声学、语音学、语言学、生理学、心理学、数字信号处理、信息工程、通

信理论、电子技术、计算机科学、模式识别、人工智能等众多的学科。关于语音处理的文献浩如烟海。本书只是关于该学科基础理论、基本方法、典型应用的概论。

本书是以作者近年来在华中理工大学为研究生讲授“数字语音处理”课程的讲稿为基础,参考所搜集的大量文献资料编写而成。作者关于汉语语音识别和语音信息压缩的科研成果以及所指导的历届研究生的研究成果,使得本书的内容得以丰富。编写本书时力图体现出这样的特色,既注意到内容的全面和广泛,又照顾到理论和实际的结合,并尽可能反映该领域中的新成果。希望这种特点能引起从事研究生教学和科学研究的同行们的兴趣,也能引起各领域中有可能涉及语音处理的工程技术人员的兴趣。

华中理工大学研究生院的有关领导对于作者编写本书给予了鼓励、关心和支持;华中理工大学出版社的同志对于本书的出版付出了大量辛勤的劳动;华中理工大学电子与信息工程系近几年的研究生们,在学习“数字语音处理”课程的过程中,曾经和作者进行过许多有益的讨论。对于以上各方面的同志,作者在此谨致衷心的感谢。

作者自感才疏学浅,书中必有不少缺点甚至不当之处,诚恳欢迎读者指出。

作者

1991年1月

“研究生用书”书目(第一批~第四批)

书 名	作 者
机械工程测试·信息·信号分析	卢文祥等
粘弹性力学	杨挺青
网络分析与综合原理	曾凡刊
现代数字设计	陈耀奎
随机过程	李鼎煊
应用泛函简明教程	申李华
协同学原理和应用	李大大
现代中央银行导论	黄芳泉
高级英语阅读系列文选(1)	朱月珍
高级英语阅读系列文选(2)	樊长荣等
时间序列分析与工程应用(上)(下) (获中国图书奖)	杨叔子等
实用偏微分方程数值解法	徐长发等
遥感图象数字处理	万发贵等
线性多变量系统	庞富胜
内燃机燃烧学	金国栋
控制系统CAD基础	罗宗虔
企业管理专家模拟系统	黎志成等
数字语音处理	姚天任
数据库设计与分析	刘云生
辩证法史论稿	阳作华等
大型复杂结构优化设计	宋天霞等
自适应光学	叶嘉雄等
机械振动系统——分析、测试、建模与对策(上)(下)	师汉民等
薄膜生长理论	王敬义
高等弹性力学	王钟伟
高等内燃机教程	刘永长
液压故障分析与状态监测	李壮云
塑性成型模拟理论	肖景容
微波磁学和光磁学	何华辉
有机化学中的单电子转移反应及其应用	赵成学等
生命科学中的微量元素硒	徐辉碧等
计算机几何	熊有伦

目 录

第一章 绪论	(1)
§ 1.1 数字语音处理研究的内容	(1)
§ 1.2 语音处理的发展历史	(3)
§ 1.3 本书的内容	(4)
第二章 语音信号产生的数字模型	(6)
§ 2.1 人类的语言器官	(6)
§ 2.2 语音产生过程	(8)
§ 2.3 语音信号产生的数字模型	(11)
§ 2.4 语音信号的特性	(14)
2.4.1 语音的声学特性	(14)
2.4.2 语音的时间波形和频谱特性	(15)
2.4.3 语音信号的统计特性	(20)
§ 2.5 人类的听觉功能	(24)
第三章 语音波形的数字编码	(27)
§ 3.1 脉冲编码调制(PCM)	(28)
3.1.1 语音信号的取样	(29)
3.1.2 取样语音信号的量化	(31)
3.1.3 减小量化噪声影响的方法	(37)
§ 3.2 差分脉冲编码调制(DPCM)	(41)
3.2.1 DPCM 原理	(41)
3.2.2 DPCM 的信噪比	(44)
3.2.3 预测系数对差分增益的影响	(47)
§ 3.3 增量调制(DM)	(50)
3.3.1 DM 原理	(50)
3.3.2 DM 的斜率过载失真和颗粒噪声	(52)
3.3.3 DM 的信噪比	(53)

§ 3.4	自适应技术在语音波形编码中的应用	(55)
3.4.1	一般原理	(55)
3.4.2	自适应量化	(56)
3.4.3	自适应增量调制(ADM)	(67)
3.4.4	自适应线性预测	(72)
§ 3.5	压缩比特率的其它方法	(81)
3.5.1	残差信号压缩	(81)
3.5.2	噪声整形	(83)
3.5.3	多脉冲技术	(85)
第四章	短时域处理技术	(87)
§ 4.1	语音信号的短时处理方法	(88)
§ 4.2	短时能量和短时平均幅度	(89)
4.2.1	短时能量	(89)
4.2.2	短时平均幅度	(93)
§ 4.3	短时平均过零率	(97)
§ 4.4	短时自相关函数	(100)
4.4.1	短时自相关函数的定义	(100)
4.4.2	减少短时自相关函数计算量的方法	(101)
4.4.3	语音信号的短时自相关函数的实例	(103)
4.4.4	短时自相关函数的另一种计算方法	(104)
4.4.5	短时平均幅度差函数	(108)
§ 4.5	短时域处理技术应用举例	(109)
4.5.1	语音段起止端点判别	(109)
4.5.2	基音周期的估计	(111)
§ 4.6	中值滤波在语音短时域处理中的应用	(114)
第五章	短时傅里叶分析	(120)
§ 5.1	短时傅里叶变换的定义	(120)
5.1.1	定义	(120)
5.1.2	移动窗形状对短时傅里叶变换的影响	(121)
5.1.3	窗宽对短时频谱的影响	(124)
5.1.4	结论	(125)
§ 5.2	短时傅里叶变换的某些性质	(126)

§ 5.3	短时傅里叶变换的线性滤波实现	(127)
§ 5.4	短时傅里叶谱的取样	(130)
5.4.1	短时傅里叶变换的时域取样	(130)
5.4.2	短时傅里叶变换的频域取样	(131)
5.4.3	短时傅里叶变换时域和频域总取样率	(131)
§ 5.5	语音的短时合成技术	(132)
5.5.1	语音短时合成的滤波器组相加法	(133)
5.5.2	短时傅里叶变换的欠速率取样	(137)
5.5.3	语音短时合成的叠接相加法	(139)
5.5.4	短时频谱变化对合成结果的影响	(143)
§ 5.6	短时分析-合成数字滤波器组的设计	(148)
5.6.1	设计中需要考虑的实际问题	(149)
5.6.2	IIR 滤波器组的设计	(157)
5.6.3	FIR 带通滤波器组的设计	(159)
§ 5.7	用快速傅里叶变换进行短时傅里叶分析	(167)
5.7.1	基本原理	(167)
5.7.2	减少计算量的其它考虑	(171)
5.7.3	用 FFT 节省短时谱合成语音信号的计算量	(172)

第六章 语音信号的线性预测

§ 6.1	线性预测基本原理	(177)
§ 6.2	线性预测和信号模型之间的关系	(179)
§ 6.3	Levinson-Durbin 算法	(184)
§ 6.4	格型滤波器	(191)
6.4.1	前向预测和反向预测	(191)
6.4.2	格型滤波器的结构	(194)
6.4.3	格型滤波器的性质	(196)
§ 6.5	由已知数据计算预测系数的方法	(199)
6.5.1	自相关法或 Yule-Walker 法	(199)
6.5.2	协方差法	(202)
6.5.3	Burg 法	(203)
§ 6.6	线性预测的频域解释	(206)
6.6.1	最小预测误差	(207)

6.6.2	线性预测谱匹配性质	(208)
§ 6.7	线性预测模型的局限性	(211)
6.7.1	基音频率对预测系数的影响	(212)
6.7.2	高频损失问题	(213)
§ 6.8	线性预测分析应用举例	(214)
6.8.1	基音检测	(214)
6.8.2	共振峰估计	(217)
§ 6.9	对数面积比(LAR)参数	(219)
第七章	语音信号的同态滤波和倒谱分析	(220)
§ 7.1	广义叠加原理	(220)
§ 7.2	卷积同态系统	(223)
§ 7.3	复倒谱的定义	(225)
7.3.1	复对数的多值性问题	(225)
7.3.2	$\tilde{X}(z)$ 的解析性问题	(226)
§ 7.4	复倒谱的性质和计算方法	(227)
7.4.1	复倒谱的性质	(227)
7.4.2	复倒谱的计算方法	(228)
§ 7.5	语音的倒谱分析	(231)
第八章	矢量量化	(236)
§ 8.1	矢量量化基本原理	(237)
8.1.1	矢量量化过程	(237)
8.1.2	失真的测度	(240)
8.1.3	码本的设计	(244)
8.1.4	计算量和存储量的估计	(247)
8.1.5	几个例子	(248)
§ 8.2	矢量量化器的理论性能	(256)
8.2.1	率-失真理论	(256)
8.2.2	率-失真理论的一些结果	(259)
8.2.3	标量量化	(265)
8.2.4	矢量量化的一些理论结果	(269)
§ 8.3	矢量源的标量量化和矢量量化	(273)
8.3.1	比特分配	(274)

8.3.2	相关源的矢量旋转	(277)
8.3.3	标量量化和矢量量化的比较	(281)
§ 8.4	减少矢量量化中计算量和存贮量的方法	(285)
8.4.1	二叉树搜索	(285)
8.4.2	多级矢量量化	(289)
8.4.3	乘积码	(293)
§ 8.5	码本的训练和测试	(296)
§ 8.6	码本的鲁棒性	(299)
§ 8.7	时间依赖矢量量化	(302)
8.7.1	选帧传送	(302)
8.7.2	分段量化	(302)
8.7.3	自适应矢量量化器	(303)
§ 8.8	语音波形矢量量化	(305)
8.8.1	波形标量量化	(305)
8.8.2	波形矢量量化	(307)
第九章	隐马尔柯夫模型(HMM)	(312)
§ 9.1	隐马尔柯夫模型的定义	(312)
9.1.1	信号模型	(312)
9.1.2	离散马尔柯夫过程	(313)
9.1.3	隐马尔柯夫模型的概念	(316)
9.1.4	隐马尔柯夫模型的参数	(319)
§ 9.2	隐马尔柯夫模型的三个基本问题	(320)
9.2.1	三个基本问题的提出	(320)
9.2.2	第1个问题的求解	(322)
9.2.3	第2个问题的求解	(327)
9.2.4	第3个问题的求解	(329)
§ 9.3	隐马尔柯夫模型的类型	(332)
9.3.1	隐马尔柯夫模型中的连续观测密度	(334)
9.3.2	自回归隐马尔柯夫模型	(336)
9.3.3	隐马尔柯夫模型的变型	(338)
9.3.4	含有状态持续时间密度显函数的隐马尔柯夫模型	(339)
9.3.5	最佳判据	(344)

9.3.6 隐马尔柯夫模型的比较	(345)
§ 9.4 隐马尔柯夫模型的实现问题	(347)
9.4.1 定标	(347)
9.4.2 多个观测序列	(351)
9.4.3 隐马尔柯夫模型参数的初始估计	(352)
9.4.4 训练数据不够的影响	(352)
9.4.5 模型选择	(354)
第十章 数字语音处理的应用	(356)
§ 10.1 语音压缩	(356)
10.1.1 声码器的基本结构	(357)
10.1.2 通道声码器	(358)
10.1.3 共振峰声码器	(361)
10.1.4 线性预测编码声码器	(362)
10.1.5 矢量量化在语音压缩中的应用	(363)
§ 10.2 语音合成	(365)
10.2.1 以单词为基础的合成方法	(365)
10.2.2 以音节为基础的合成方法	(366)
10.2.3 以音素为基础的合成方法	(367)
10.2.4 语音合成器芯片和语音合成系统	(369)
§ 10.3 语音识别	(371)
10.3.1 孤立单词语音识别	(374)
10.3.2 线性时间归一化	(376)
10.3.3 非线性时间归一化	(378)
10.3.4 采用 DTW 的孤立单词语音识别	(382)
10.3.5 隐马尔柯夫模型在孤立单词语音识别中的应用 ...	(384)
10.3.6 连续语音识别	(388)
§ 10.4 语音增强	(394)
参考文献	(397)