

语音信号处理

Speech Signal Processing

易克初 田斌 付强 编著



国防工业出版社

53.58
310
4

语音信号处理

Speech Signal Processing

易克初 田 斌 付 强 编著

国防工业出版社

·北京·

图书在版编目(CIP)数据

语音信号处理/易克初等编著. —北京:国防工业出版社,2000.6

ISBN 7-118-02198-9

I. 语… II. 易… III. 语言信号处理 IV. TN912.3

中国版本图书馆 CIP 数据核字(1999)第 50568 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

涿中印刷厂印刷

新华书店经售

*

开本 850×1168 1/32 印张 11 $\frac{1}{4}$ 305 千字

2000 年 6 月第 1 版 2000 年 6 月北京第 1 次印刷

印数:1—2000 册 定价:24.00 元

(本书如有印装错误,我社负责调换)

致 读 者

本书由国防科技图书出版基金资助出版。

国防科技图书出版工作是国防科技事业的一个重要方面。优秀的国防科技图书既是国防科技成果的一部分,又是国防科技水平的重要标志。为了促进国防科技事业的发展,加强社会主义物质文明和精神文明建设,培养优秀科技人才,确保国防科技优秀图书的出版,国防科工委于1988年初决定每年拨出专款,设立国防科技图书出版基金,成立评审委员会,扶持、审定出版国防科技优秀图书。

国防科技图书出版基金资助的对象是:

1. 学术水平高,内容有创见,在学科上居领先地位的基础科学理论图书;在工程技术理论方面有突破的应用科学专著。

2. 学术思想新颖,内容具体、实用,对国防科技发展具有较大推动作用的专著;密切结合科技现代化和国防现代化需要的高新技术内容的专著。

3. 有重要发展前景和有重大开拓使用价值,密切结合科技现代化和国防现代化需要的新工艺、新材料内容的科技图书。

4. 填补目前我国科技领域空白的薄弱学科和边缘学科的科技图书。

5. 特别有价值的科技论文集、译著等。

国防科技图书出版基金评审委员会在国防科工委的领导下开展工作,负责掌握出版基金的使用方向,评审受理的图书选题,决定资助的图书选题和资助金额,以及决定中断或取消资助等。经评审给予资助的图书,由国防工业出版社列出版。

国防科技事业已经取得了举世瞩目的成就。国防科技图书承

担着记载和弘扬这些成就,积累和传播科技知识的使命。在改革开放的新形势下,国防科工委率先设立出版基金,扶持出版科技图书,这是一项具有深远意义的创举。此举势必促使国防科技图书的出版随着国防科技事业的发展更加兴旺。

设立出版基金是一件新生事物,是对出版工作的一项改革。因而,评审工作需要不断地摸索、认真地总结和及时地改进,这样,才能使有限的基金发挥出巨大的效能。评审工作更需要国防科技工业战线广大科技工作者、专家、教授,以及社会各界朋友的热情支持。

让我们携起手来,为祖国昌盛、科技腾飞、出版繁荣而共同奋斗!

国防科技图书出版基金
评审委员会

国防科技图书出版基金 第三届评审委员会组成人员

名誉主任委员	怀国模
主任委员	黄宁
副主任委员	殷鹤龄 高景德 陈芳允 曾铎
秘书长	崔士义
委员	(以姓氏笔划为序)
	于景元 王小谟 尤子平 冯允成
	刘仁 朱森元 朵英贤 宋家树
	杨星豪 吴有生 何庆芝 何国伟
	何新贵 张立同 张汝果 张均武
	张涵信 陈火旺 范学虹 柯有安
	侯正明 莫梧生 崔尔杰

前 言

语音是人类交流信息的基本手段。随着信息科学技术的飞速发展,语音处理技术在最近 20 多年中取得了重大进展。语音信号处理技术的四大分支——语音识别、语音合成、语音编码和说话人识别都在蓬勃发展,学术思想丰富多彩,研究开发成果累累,并且陆续地推向实际应用。

语音信号处理,由于其学术思想的综合性、趣味性和可推广性,一直深深地吸引着众多学者,同时由于其研究成果的实用性和与人们新生活的密切关系,又受到了广大公众的密切关注。它在信息化社会中的地位,就像百花园中一颗神奇的花朵,刚一绽开花蕾就是那么鲜艳夺目。

语音信号处理技术还在蒸蒸日上地发展着。本书的目的不仅在于让读者对其基本理论和主要技术有一个总的了解,还希望使读者在学术思想上受到一些启发,因此注意收集了本学科的一些最新发展。同时还注意了所介绍方法的可操作性,并在附录中提供了几个应用程序。

本书第一章绪论介绍语音信号处理及其发展过程的基本情况;第二章集中介绍与语音信号处理技术关系十分密切的语音学知识;第三章讲语音信号时域处理方法;第四章讨论语音信号频谱分析方法,包括倒谱分析法;第五章介绍语音信号线性预测原理;第六章讨论语音信号的时频表示、参数表示和矢量表示,以及与参数矢量表示密切相关的谱失真测度。以上各章尽管涉及了时频分布、小波理论、矢量量化等新兴领域。但都是语音信号处理技术中共用的基础知识。第七章至第十章分别介绍语音处理的四大分支的专门技术。最后的附录中提供了四个应用程序。

本书的内容涉及作者承担的两个国家自然科学基金项目的部分研究成果,其中“时频分布和小波理论在信息传输中的应用”(1999—2001)正在进行之中。

本书主要考虑为从事语音处理和信息科学中的某些相关学科专家提供参考,但若增加一些习题和实验,也可作为研究生教材。

本书虽然是作者从事十余年语音信号处理教学与科研的结晶。但因作者水平有限、时间较仓促,缺点错误在所难免,敬请读者多多批评指教。

作者

于西安电子科技大学

目 录

第一章 绪论

- § 1.1 语音处理技术的发展概况 3
- § 1.2 语音信号处理的应用 8

第二章 语音学概要

- § 2.1 语音声学基础 11
- § 2.2 语音的感知 16
- § 2.3 语音的产生 24
- § 2.4 元音 35
- § 2.5 辅音 38
- § 2.6 汉语音节 44

第三章 语音信号的时域处理方法

- § 3.1 语音信号及其数字化 51
- § 3.2 语音信号的时间依赖处理 51
- § 3.3 短时能量和短时平均幅度 54
- § 3.4 短时平均过零率 54
- § 3.5 语音端点检测 56
- § 3.6 短时自相关函数 58
- § 3.7 短时平均幅度差函数 59
- § 3.8 基音周期估计方法 60

第四章 语音信号频谱分析

- § 4.1 短时傅里叶变换的定义和物理意义 67
- § 4.2 基于短时傅里叶变换的语谱图及其时频分辨率 68
- § 4.3 基于短时傅里叶变换的信号重构 71
- § 4.4 基音同步分析法 79

§ 4.5	复倒谱和倒谱	83
§ 4.6	语音信号的倒谱分析与同态解卷积	88
第五章 语音信号线性预测原理		
§ 5.1	线性预测的基本原理	97
§ 5.2	线性预测方程组的解法	100
§ 5.3	线性预测的频域解释	107
§ 5.4	线性预测的几种推演参数	108
§ 5.5	线谱对分析法	112
§ 5.6	基于线性预测参数的声道滤波器的实现	115
第六章 语音信号的表示与谱失真测度		
§ 6.1	时频分布理论和语音信号的时频表示	120
§ 6.2	小波理论及其在语音处理中的应用	132
§ 6.3	语音信号的参数表示及其失真测度	135
§ 6.4	矢量量化技术及其在语音处理中的应用	148
第七章 语音识别		
§ 7.1	语音识别系统	154
§ 7.2	语音识别中的特征提取及谱失真测度	160
§ 7.3	矢量量化技术在语音识别中的应用	161
§ 7.4	模板匹配法	164
§ 7.5	隐马尔可夫模型技术	172
§ 7.6	连接词语音识别系统	191
§ 7.7	大词汇量连续语音识别系统	197
§ 7.8	说话人自适应技术	212
§ 7.9	关键词检出	223
§ 7.10	语音理解	228
§ 7.11	人工神经网络技术在语音识别中的应用	233
第八章 说话人识别和语种辨识		
§ 8.1	说话人识别的基本原理和应用	249
§ 8.2	说话人识别的特征选择	254
§ 8.3	说话人识别的主要方法	256

§ 8.4 语种辨识的原理和应用	261
第九章 语音合成	
§ 9.1 语音合成方法	266
§ 9.2 汉语语音的按规则合成	278
第十章 语音编码	
§ 10.1 不依赖模型假定的语音编码方法	284
§ 10.2 基于模型假定的语音编码方法	297
§ 10.3 极低速率语音编码技术	323
§ 10.4 语音编码器的性能指标和评测方法	328
§ 10.5 语音质量评价	331
附 录	
1. 语音基音检测子程序	340
2. 语音信号线性预测分析子程序	345
3. 基于线性预测的倒谱系数计算子程序	347
4. Viterbi 算法子程序	348

Contents

Chapter 1 Introduction

- 1.1 Review of speech signal processing 3
- 1.2 Applications of speech signal processing 8

Chapter 2 Elementary Phonetics

- 2.1 Fundamentals of speech acoustics 11
- 2.2 Perception of speech 16
- 2.3 Generation of speech 24
- 2.4 Vowels 35
- 2.5 Consonants 38
- 2.6 Syllables in the Chinese language 44

Chapter 3 Time Domain Processing of Speech Signals

- 3.1 Speech signal and its digitization 51
- 3.2 Time-dependent processing of speech signals 51
- 3.3 Short-time energy and average amplitude 54
- 3.4 Short-time zero-crossing rate 54
- 3.5 Endpoint detection of speech signals 56
- 3.6 Short-time auto-correlation function 58
- 3.7 Short-time average amplitude difference function ... 59
- 3.8 Pitch estimating 60

Chapter 4 Spectrum Analysis of Speech Signals

- 4.1 Definition and concept of short-time fourier transform 67
- 4.2 Spectrogram and its time-frequency resolution 68
- 4.3 Reconstruction based on short-time fourier

transform	71
4.4 Pitch synchronous analysis	79
4.5 Complex cepstrum and cepstrum	83
4.6 Cepstrum analysis and homomorphic de-convolution	88
Chapter 5 Linear Prediction of Speech Signals	
5.1 Principle of linear prediction	97
5.2 Solution of linear predictive equations	100
5.3 Explanation in frequency domain for linear prediction	107
5.4 Some parameters deduced from linear prediction	108
5.5 Line spectrum pair	112
5.6 Implementation of vocal track filters	115
Chapter 6 Representation of Speech Signals and Distortion Measures	
6.1 Time-frequency representation of speech signals	120
6.2 Wavelet theory and its applications to speech processing	132
6.3 Speech signal representation and spectrum- distortion measures	135
6.4 Vector quantization and its application to speech processing	148
Chapter 7 Speech Recognition	
7.1 Speech recognition systems	154
7.2 Feature extraction and distortion measures in speech recognition	160
7.3 Vector quantization applied to speech recognition	161

7.4	Dynamic time warping	164
7.5	Hidden markov modeling	172
7.6	Speech recognition based on connected word models	191
7.7	Large vocabulary continuous speech recognition ...	197
7.8	Speaker adaptation	212
7.9	Keyword spotting	223
7.10	Speech understanding	228
7.11	Neural networks applied to speech recognition ...	233
Chapter 8 Speaker Recognition and Language Identification		
8.1	Principle and applications of speaker recognition	249
8.2	Feature selection in speaker recognition	254
8.3	Elementary methods of speaker recognition	256
8.4	Principle and applications of language identification	261
Chapter 9 Speech Synthesis		
9.1	Speech synthesizing methods	266
9.2	Chinese speech synthesizing by rules	278
Chapter 10 Speech Coding		
10.1	Speech coding without model assumption	284
10.2	Speech coding based on model assumption	297
10.3	Very low bit rate speech coding	323
10.4	Specifications and assessment of speech coders ...	328
10.5	Speech quality evaluation	331
Appendix		
1.	A C-Program of pitch detection	340
2.	A C-Program of linear prediction	345
3.	A C-Program of calculating cepstrum coefficients from LP coefficients	347
4.	A C-Program of Viterbi algorithm	348

第一章 绪 论

劳动创造了人类,创造了人类文明。而语言是人类创造和记载几千年的人类文明史的根本手段,没有语言就没有今天的人类文明。

语音是语言的声学表现,是人类交流信息最自然、最有效、最方便的手段,也是人类进行思维的一种依托。人类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、存储和获取语言信息,这对于促进社会的发展具有十分重要的意义。例如,我们可以应用自动语音识别技术,使手写文稿和手工打印文本变成自动听写机操作;把人工查阅各种书面文字资料的操作,变为口呼自动查阅各种各样的数据库;可以通过语音压缩编码技术把一些有价值的对话、演讲和广播等语音信息,进行高效率的压缩后再存储和传输;可以采用语音合成技术,将存储的语音或文字资料转化为语音高质量地回放,甚至自动翻译成另一种语言的语音回放或进行文字显示。这就是说,如果人们的听、说、读、写、查阅信息和语音通信等,都能采用先进的手段高效率地进行,这将使人们的才能得到更加充分的发挥和有效的扩展。特别是移动通信和个人通信的发展,将使地球上任意地区的人都可以随时随地通电话,人们的相互联系将更加紧密,好像大家同属“地球村”的村民。这一理想的实现已为期不远,其中语音压缩编码技术的应用当然是必不可少的。总之,语音信号处理的研究对于信息化社会的发展具有极其重要的意义。

语音信号处理与语音学存在十分密切的关系,因此本书专门安排了一章介绍语音学的一些基本知识,同时介绍作者关于汉语语音系统结构特点的新见解。语音学是研究言语过程的一门科学,

它包括发音语音学、声学语音学和听觉语音学三大分支。发音语音学研究发音器官在发音过程中的运动和语音的音位特性；声学语音学研究语音的物理属性，如语音声波的频率、振幅和频谱特性等；听觉语音学研究听觉和语音感知。

语音学是一门古老的学科，但它在近几十年中显得十分活跃，得到了很大的发展。由于它的研究手段迅速改进，由过去单纯地靠人的观察、听辨和推断，到使用仪器设备测试乃至使用计算机分析，形成了一门新兴学科——实验语音学，取得了许多重要研究成果，加速了语音处理技术的发展。

语音信号处理也可以简称为语音处理，它是以语音学和数字信号处理为基础而形成的一个综合性的学科，包括语音识别、语音合成、语音编码和说话人识别等四大分支。自动语种辨识可能是一个新出现的分支，但也可以归并到说话人识别之中。

语音处理的研究目标多种多样，处理方法丰富多彩，一直是数字信号处理技术发展的重要推动力量。无论是谱分析方法，还是数字滤波技术或压缩编码算法等，许多新方法的提出，首先是在语音处理中获得成功，然后再推广到其他领域的。高速信号处理器的诞生和发展也是与语音处理的发展分不开的，语音识别和语音编码算法的复杂性和实时处理的需要，就是促使人们去设计这样的处理器的重要推动力量之一。这种产品问世之后，又首先在语音处理应用中得到最有效的推广应用。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。

语音处理之所以能够那样长期地、深深地吸引广大学者，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。比如说，模糊集理论、神经网络理论、小波理论和时频分布理论是当前非常热门的研究领域，这些研究领域的研究常常不可避免地要把语音处理任务作为一个应用实例，而语音处理研究者也常常从这些领域的研究进展中找到突破口，使语音处理技术研究取得突破性进展。

§ 1.1 语音处理技术的发展概况

语音信号处理,简称语音处理,主要包括语音识别、语音合成、语音编码和说话人识别等四大分支,语音学则是与语音处理密切相关的一个学科。

语音处理研究的历史可以追溯到 1876 年贝尔(Bell)发明电话,那是首次采用声电、电声转换技术实现远距离语音通信。在此之前还有一件趣事值得提及,那就是奥地利皇帝命令他的科学家 Von Kempelen 研制一台能说话的机器,并且终于在 1780 年获得了一定成功。这种纯机械结构的机器可以发出一些像元音那样最简单的语音。1939 年 H. Dudley 提出了一种全新的语音信号传输方法,研制成功第一个声码器,这种声码器打破语音信号的内部结构,使之解体,提取负载信息的参量加以传输,在接收端重新合成语音,Vocoder 这个词就是他创造的。这一发明奠定了语音产生模型的基本思想,在语音信号处理领域具有划时代的意义。1947 年 Bell 实验室发明了语谱图仪,这一发明意味着自动语音识别研究的开始,不过起初只是通过人工分析语谱图进行语音识别。1948 年美国 Haskins 实验室研制成功“语图回放机”,它把手工绘制在薄膜片的语谱图自动转换为语音,可以进行语音合成。虽然这样得到的合成语音显得单调、呆板,但其学术思想是有深远意义的。共振峰合成法就是源于这一思想的,它至今还被认为是一种较好的语音合成方法。

1952 年 Bell 实验室的 Davis 等人首次研制成功能识别十个英语数字的实验装置,其识别方法是根据语音第一、二共振峰位置提取若干特征,用模拟电路实现未知语音模式与参考语音模式之间的互相关运算。1958 年, Duddley 和 Balashek 改进了这一装置,他们将语音分割为元音、辅音等语音单位。1956 年 Olson 和 Belar 等人采用 8 个带通滤波器提取频谱参量作为语音的特征,研制成功一台简单的声控打字机。