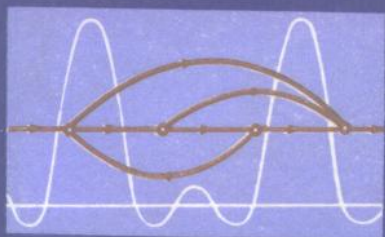


胡光锐



语音处理与识别

YUYIN CHULI YU SHIBIE



上海科学技术文献出版社

7.4.10
386

语音处理与识别

胡光锐



上海科学技术文献出版社

9510051

内 容 提 要

语音处理与识别是利用计算机对语音信号进行分析和综合,实现对人类语音的自动理解和处理的一门学科。本书共八章,即语音信号数字处理基础、语音信号时域处理方法、语音波形数字编码、短时傅里叶分析、语音信号同态处理及线性预测编码、语音识别方法与系统等。

本书可作通信工程、无线电技术和计算机应用科学等专业的本科生、研究生以及科技人员参考。

D029/2608

语 音 处 理 与 识 别

胡 光 锐

*

上海科学技术文献出版社出版发行
(上海市武康路2号 邮政编码200031)

全国新华书店经销

上海科技文献出版社昆山联营厂印刷

*

开本 850×1168 1/32 印张 10.75 字数 299,000

1994年12月第1版 1994年12月第1次印刷

印数: 1—1,000

ISBN 7-5439-0548-5/T·346

定价: 15.00元

科技新书目: 332-283

(沪)新登字301号

序

语音处理与识别是利用计算机对语音信号进行分析和综合,从而实现对人类语音的自动理解和处理的一门学科,它是信息科学中一个十分活跃的研究领域。

当1946年第一台电子计算机问世时,就有人提出利用计算机进行语言翻译的设想,在50年代中期,机器的自动翻译成了语言处理中优先探索的课题。另一方面,语音处理的研究是和通信技术的发展密切相关的。例如,为提高通信系统的效率研究语音编码方法,为满足低码率的通信要求,研究语音分析合成系统——声码器等。自70年代以来,语音信号处理成为语音通信技术中的一个重要研究课题。它的研究内容包括语音分析、语音合成、语音增强、语音编码和语音识别。语音分析之目的是提取语音的特征参数,而语音合成是让机器说话,语音合成技术上已趋成熟,至80年代,超大规模集成电路的语音合成器芯片已作为商品出售,可以合成出可懂度较高能满足实际需要的语音。语音增强是为了提高语音信号的信噪比,语音编码则需要解决传输速率和语音质量的矛盾问题,而语音识别则是让机器听懂和理解人的语音。语音识别和合成是人工智能的重要课题,是实现人机语音通信的基础。为了实现人机语音通信,必须具备语音识别和语音理解两种功能。前者提取语音的特征参数,进行语音初判,而后者根据初判结果和语法规则确定语音的内容和意义。当前,特定说话人、中小词汇量的单词识别系统已开始走向实

用，语音识别正朝着解决非特定说话人、大词汇量和连续语音三大难题的方向进行研究。

电子计算机和大规模集成电路技术的进步促进了语音信号处理的发展，反过来，又对语音处理与识别提出了新的课题，语音信号处理的发展将更加引人注目，它无疑会给某些应用部门带来新的变革。

我高兴地向读者介绍这一本书。让我们共同努力，促进语音处理与识别技术的更快发展。

中国科学院院士
上海交通大学教授



1994年2月

前 言

本书的适用对象是通信工程、无线电技术和计算机应用等有关专业的本科生,以及通信与电子系统、信号与信息处理以及模式识别与智能控制等有关专业的研究生,和从事通信、控制、计算机等工作的科技人员参考。

本书共分八章。第一章扼要介绍语音产生的过程与汉语的语音特性,并从语音信号的无损声管模型推导出语音信号的数字模型。这一章是语音处理的基础。

第二章论述语音处理的时域方法,包括语音时域处理中的一些基本定义,如短时能量函数、平均幅度函数、短时平均过零率、短时自相关函数和短时平均幅度差函数等。在这一章中,还讨论了语音信号的中值滤波,这是一种非线性处理方法,在语音处理中具有很好效果。

第三章论述语音波形的数字编码,即波形编码,从简单的增量调制(DM)到自适应脉冲编码调制(ADPCM)。

第四章介绍短时傅里叶分析以及通道声码器。第五章讨论同态语音处理以及同态声码器。第六章介绍语音的线性预测编码(LPC)及其在声码器等方面的应用。

第七章介绍语音识别的基本方法和系统组成,并讨论了语音识别的有关技术。第八章讨论数字语音处理与识别的其他方法,介绍了含噪语音的自适应滤波方法,语音的隐马尔柯夫模型(HMM)及其在语音识别中的应用,也讨论了几种主要的矢量量化和聚类方法。

在每一章之后附有习题,以便通过习题来消化全书的理论,书中附有四个语音处理的计算机程序,可供实验时参考。

本书的内容采用了作者所进行的研究课题的有关成果,该课

题得到国家攀登计划认知科学(神经网络)重大关键项目和电子科学基金的资助。在本书编写过程中,曾得到校内外有关同志的鼓励支持,谨此一并表示衷心的感谢。

由于作者水平所限,编写时间仓促,书中的错误或不足之处,欢迎读者批评指正。

胡光锐

1993年9月

于上海交大

目 录

引言	(1)
第一章 语音信号数字处理基础	(4)
1.1 发音的生理机构与过程	(4)
1.2 汉语语音基本特性	(7)
1.2.1 汉语语音基础	(8)
1.2.2 元音和辅音	(8)
1.2.3 声母和韵母	(10)
1.2.4 音调(字调)	(11)
1.2.5 音节(字)构成	(11)
1.2.6 汉语的波形特征	(12)
1.2.7 元音的频谱特性	(13)
1.2.8 辅音的频谱特性	(13)
1.2.9 汉语语音的韵律特性	(16)
1.3 语音信号的无损声管模型	(16)
1.3.1 级联无损声管声波传播关系式	(16)
1.3.2 级联无损声管与数字滤波器的关系	(21)
1.3.3 无损声管模型的传输函数	(25)
1.4 语音信号的数字模型	(30)
习题	(35)
第二章 语音信号的时域处理方法	(39)
2.1 短时平均幅度和能量	(39)
2.2 短时平均过零率	(48)
2.3 语音的端点检测	(45)
2.3.1 能量与过零率检测法	(45)

2.3.2	改进的端点检测方法	(47)
2.4	短时自关函数	(48)
2.5	短时平均幅度差函数	(55)
2.6	基音检测时域方法	(56)
2.6.1	并联处理方法	(56)
2.6.2	削波修正自相关基音检测法(AUTOC)	(59)
2.6.3	平均幅度差函数基音检测法	(63)
2.7	语音信号的中值滤波	(65)
	习题	(68)
第三章	语音波形的数字编码	(74)
3.1	取样和多取样率信号处理	(74)
3.2	瞬时量化器	(79)
3.2.1	均匀量化	(80)
3.2.2	瞬时压扩	(84)
3.3	自适应量化器	(87)
3.3.1	前馈自适应量化器	(88)
3.3.2	反馈自适应量化器	(90)
3.4	增量调制器	(93)
3.4.1	线性增量调制器	(94)
3.4.2	自适应增量调制器	(96)
3.4.3	增量调制中的高阶预测器	(99)
3.5	差值脉冲编码调制器	(100)
3.5.1	具有自适应量化的 DPCM	(101)
3.5.2	带自适应预测的 DPCM	(102)
3.6	数字编码的直接转换	(106)
3.6.1	LDM-PCM 转换	(106)
3.6.2	PCM-ADPCM 转换	(107)
	习题	(108)
第四章	短时傅里叶分析方法	(114)
4.1	时变傅里叶变换	(114)

4.1.1	从第一种定义来解释	(115)
4.1.2	从第二种定义来解释	(117)
4.1.3	$X_n(e^{j\omega})$ 的时域及频域采样率	(120)
4.2	短时综合的滤波器组相加法	(123)
4.3	短时综合的重叠相加法	(129)
4.4	短时频谱变化对综合结果的影响	(132)
4.4.1	滤波器组相加法(FBS)	(132)
4.4.2	重叠相加法(OLA)	(134)
4.4.3	加性改变	(135)
4.5	滤波器组设计考虑	(136)
4.6	利用FFT实现滤波器组相加法	(144)
4.6.1	分析技术	(144)
4.6.2	综合技术	(146)
4.7	基音检测	(148)
4.8	综合分析法	(150)
4.8.1	基音同步频谱分析	(150)
4.8.2	共振峰估计	(152)
4.8.3	声门波的基音同步估值	(155)
4.9	通道声码器	(155)
	习题	(158)
第五章	语音信号的同态处理	(164)
5.1	同态与卷积同态系统	(164)
5.1.1	广义叠加原理	(164)
5.1.2	卷积同态系统	(168)
5.2	复倒谱的性质	(173)
5.3	特征系统 D_* 的实现	(179)
5.4	语音的同态处理	(182)
5.5	基音周期估计	(188)
5.6	共振峰分析方法	(189)
5.7	同态声码器	(193)

习题	(195)
第六章 语音信号的线性预测编码	(200)
6.1 线性预测编码的基本原理.....	(201)
6.1.1 自相关法.....	(205)
6.1.2 协方差法.....	(207)
6.1.3 模型增益的计算.....	(209)
6.2 解线性预测参数方程组的算法.....	(211)
6.2.1 利用乔里斯基分解计算协方差法的方程组.....	(212)
6.2.2 利用杜宾递推算法计算自相关法的方程组.....	(216)
6.2.3 利用格型法求解线性预测系数.....	(218)
6.2.4 各种算法的比较.....	(222)
6.3 归一化预测误差.....	(224)
6.4 选择线性预测.....	(226)
6.5 各种语音参数之间的关系.....	(229)
6.5.1 预测器多项式的根.....	(229)
6.5.2 LPC 系统冲激响应的倒谱	(230)
6.5.3 全极点系统的冲激响应.....	(230)
6.5.4 冲激响应的自关函数.....	(230)
6.5.5 预测器多项式的自关函数.....	(231)
6.5.6 部分相关系数.....	(231)
6.5.7 对数面积比系数.....	(232)
6.6 线性预测编码的应用.....	(232)
6.6.1 利用 LPC 参数综合语音	(232)
6.6.2 利用 LPC 参数进行基音检测	(234)
6.6.3 利用 LPC 参数的共振峰分析	(235)
6.6.4 LPC 声码器	(236)
习题	(238)
第七章 语音识别方法与系统	(244)
7.1 语音识别的基本方法.....	(244)
7.2 语音识别系统的分类与组成.....	(248)

7.2.1	孤立词语音识别系统	(249)
7.2.2	连续语音数字识别系统	(252)
7.2.3	大词汇量单字识别系统	(255)
7.3	应用于语音识别的有关技术	(256)
7.3.1	失真测度的选择	(256)
7.3.2	矢量量化语音识别算法	(262)
第八章	语音处理与识别的其他方法	(265)
8.1	含噪语音信号的自适应滤波	(265)
8.1.1	自适应滤波的概念	(266)
8.1.2	语音信号的噪声滤除	(267)
8.1.3	噪声滤除的算法	(268)
8.1.4	单频信号的分析	(269)
8.1.5	噪声滤除的效果	(271)
8.2	语音的 HMM 模型及其在语音识别中的应用	(272)
8.2.1	语音信号 HMM 的物理含义	(272)
8.2.2	实变信号 HMM 模型的建立和参数估计	(273)
8.2.3	HMM 在语音识别中的应用	(278)
8.2.4	FHMM(Fuzzy-HMM)的相应算法	(280)
8.2.5	几种传统的识别系统	(282)
8.2.6	基于 FVQ 和 FHMM 多人通用孤立词识别系统	(285)
8.3	几种主要的 VQ 及聚类方法	(288)
8.3.1	K-Means 算法	(291)
8.3.2	Binary-Split 方法	(292)
8.3.3	MKM 算法	(293)
8.3.4	“长轴”类分裂算法	(294)
8.3.5	FVQ 方法简介	(295)
附录	关于计算机程序的说明	(298)
程序 1	线性预测分析计算机程序(AUTO-COVAR)	(299)

引 言

语音信号是人们构成思想疏通和感情交流的必要手段，是人类所特有的也是最重要的交际手段。语音具有两重属性，一方面语音具有表义功能；另一方面，语音毕竟是一种声音，它是由人的头脑中产生的意念通过一组神经信号去控制发音器官，变成空气的振动信号，然后由空气传递到人的耳朵或受话器中的信号。

语音的基本作用是进行信息交换，构成语音通信。在语音通信系统中，语音信号的传输存贮和处理的方式是各种各样的。本书的任务就是要探讨数字技术在语音处理中的应用。利用数字技术可以灵活而方便地实现极其复杂的信号处理任务，是许多模拟信号处理方法根本无法实现的，可以达到高精度、高稳定性、高机动性。事实上，数字信号处理技术的发展，其中的一部分就是由数字语音处理的研究中得来的。

大规模集成电路(VLSI)工艺和技术的发展，使得一个极其复杂的语音处理系统可以在一块芯片上实现。其工作速度已能满足实时语音处理的要求。

此外，语音信号采用数字编码就可以在强干扰的信道中可靠地进行传输，也可以在数据传输通信网中传输。同时，数字语音可以方便地实现保密通信。因此，数字技术在语音通信中得到广泛的应用。

大体上说，语音处理的研究可以分为以下几个方面：即语音分析、语音增强、语音编码、语音合成和语音识别与理解。其中，语音分析是语音处理的基本手段；语音增强是语音处理的准备过程；语音合成系统是模仿或代替人口的发音功能；语音识别系统则模仿或代替人耳的听觉功能，而语音理解系统则企图模仿人脑的思维能力，是语音处理的高级阶段。可以预料，随着数字语音处理和

VLSI 技术的发展, 具有发音、听觉和思维功能的机器人将会在国民经济的各个部门和日常生活的各个方面登上舞台, 充分发挥作用, 给生产和生活带来极大方便。

数字语音处理的最早应用是声码器 (Voice coder 的组合词 vocoder), 对语音信号波形进行数字编码 (如 LDM, PCM, DPCM, ADM 等), 其数据率 (每秒比特数) 为 20 至 200 kbit/s, 而采用参数编码, 其数据率为 0.5 至 10 kbit/s. 这为通信系统节省频带提供可能。此外, 数字传输也便于保密和加密。

计算机语音应答系统是人们通过键盘或终端向计算机提问, 而计算机用语音告诉人们所要知道的信息, 它是一种语音合成系统, 用于自动电话查号辅助系统, 自动信息服务系统, 以及股票行情系统等。说话人确认和辨认系统是一种语音分析系统, 分为合作的和非合作的两种。它的作用是在一大批人中证实或辨别出一个说话人。这种系统可以用于自动信贷交易, 法庭审讯取证或旁证等。语音增强系统的基本应用是去混响, 或去回声, 消除噪声, 或从潜水员使用的氮氧混合气体下的发音中恢复出易懂的语音等。数字语音处理的另一项应用是帮助听力障碍者和发音困难的人学习说话和矫正发音, 可以采用语音信息的视觉反馈显示装置来教聋人说话, 进行发音技巧的基本训练, 指导发音。

自然语音是人与人之间以及人与计算机之间的最方便的一种信息交换方式。因此, 使计算机具有类似于人的听觉功能和发音功能, 是人们长期追求的目标。语音理解系统是语音处理的高级阶段。对于语音识别, 其识别的对象是词汇, 而语音理解的对象是语句。语句是由词汇组成的, 但不是词汇的简单组合, 组合方式取决于语法规则、语义内容和使用环境。语音理解系统一般包括语音分析与识别、词语匹配、语义分析和语句组合等。语音识别与理解是人机口语对话的基础, 是机器人智能化的必要条件之一。在口语翻译、口授打字、口语控制、管理自动化与指挥自动化方面都期待着数字语音处理和语音识别与理解技术的发展与提高。

数字语音处理属于应用科学，要学好本课程，关键在于实践，必须联系实际应用，才能很好掌握数字语音处理的理论和技术，因此本书收集了大约 80 个课外习题和 4 个语音处理程序，建议学习者仔细选做书中的习题，并进行计算机实验，以获得实际经验。

第一章 语音信号数字处理基础

为了弄清数字语音处理的基本原理，必须了解发音的生理结构与过程，以及汉语语音的特性。从语音产生的声学理论可以得到语音信号的无损声管模型，并在此基础上可以建立语音信号的数字模型。

在这一章中所得到的语音信号数字模型是本书其余部分讨论的基础。语音分析和合成的基本方法都是在此数字模型基础上得到的。在语音分析中，我们的目的是如何从一个自然语音信号中估计这些模型参数，自然语音就是这一模型的输出。在语音合成中，我们控制这一模型的参数来产生语音信号。语音识别，发言人的辨认与证实的各种方法也都与语音信号的数字模型有关。

因此，本章讨论的内容是语音信号数字处理的基础。

在这一章中，我们先讨论发音的生理机构与过程以及汉语语音的基本特性，然后将语音信号看成是线性时变系统在随机噪声和准周期脉冲序列激励下的输出。应用这一方法就得到了语音信号的数字模型。本书其他各章讨论的方法都是在这个数字模型的基础上建立的。

1.1 发音的生理机构与过程

人的发音生理机构见图 1.1。发音时由肺部收缩送出一股直流空气，经气管流至喉头声门处（声门即声带开口处），在发声之初，声门处的声带肌肉收缩，声带并拢间隙小于 1mm，这股直流空气冲过这很小的缝隙，使声带得到横向和纵向的速度，此时声带间两边运动缝隙增大（成年男性开到最大时，截面积约为 20 mm^2 ），声门处压力下降，弹性恢复力将声带拉回平衡位置并继续趋向闭合，即声带产生振动，而且具有一定的振动周期。