

高等院校计算机教育系列教材

蔡莲红 黄德智 蔡 锐 编著

现代语音技术 基础与应用

语音处理——综合性的新兴学科

语音合成——教计算机学会说话

语音识别——让计算机听懂人类语言

语音编码——去除冗余压缩数据



清华大学出版社

高等院校计算机教育系列教材

现代语音技术基础与应用

蔡莲红 黄德智 蔡 锐 编著

清华大学出版社

北 京

内 容 简 介

本书综合介绍了现代语音处理的研究与进展。内容包括语音学基础、方法及应用,是一本论述现代语音处理的专著。

全书共分8章。第1章概述语音处理研究的问题和方向;第2章介绍数字信号处理和语音信号处理基础。第3章从工程的角度出发简单介绍了语音基础知识和语料库有关内容;第4章介绍语音编码的算法和国际标准;第5章介绍语音合成的原理和文语转换系统;第6章介绍语音识别基础;第7章介绍可视语音和汉语视位;第8章介绍典型语音系统及其应用。语音信号处理是计算机应用的重要研究方向。本书主要介绍语音信号处理的基础知识以及相关的新方法和新技术。

本书内容精心挑选,叙述深入浅出、基础理论和实际应用并重。本书可供语音处理和相关学科的研究生和本科生作为教材,也可作为研究人员和工程技术人员的参考书。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

现代语音技术基础与应用/蔡莲红,黄德智,蔡锐编著.—北京:清华大学出版社,2003
(高等院校计算机教育系列教材)

ISBN 7-302-07277-9

I.现… II.①蔡…②黄…③蔡… III.语音数据处理 IV.TN912.3

中国版本图书馆 CIP 数据核字(2003)第 084072 号

出 版 者:清华大学出版社

<http://www.tup.com.cn>

社 总 机:010-62770175

地 址:北京清华大学学研大厦

邮 编:100084

客户服务:010-62776969

组稿编辑:刘建龙

文稿编辑:杨作梅

封面设计:陈刘源

印 刷 者:北京四季青印刷厂

装 订 者:三河市李旗庄少明装订厂

发 行 者:新华书店总店北京发行所

开 本:185×260 印张:23.75 字数:570千字

版 次:2003年11月第1版 2003年11月第1次印刷

书 号:ISBN 7-302-07277-9/TP·5282

印 数:1~4000

定 价:32.00元

前 言

语言是人类互动的基本手段也是人机交互最自然的方式,是通信的必要手段。专家预测,未来计算机“能听会说”,建立和谐的人机交互环境。语音处理是目前发展最为迅速的一个研究领域,并形成了一门新兴的交叉学科。

语音处理的任务是如何利用信号处理技术研究语音信号。语音的生成、接收是人类的高级行为。在计算机和网络环境中,人们还可以传输、存储语音信号。因此语音处理涉及到语音学、语言学、生理学、心理学、信号处理、计算机和数学等多个学科。

人类对语音的研究具有长远的历史,而且研究方法和技术不断进步。在语音工程中,从模拟处理变换到数字处理;在处理策略上,从专家自省、基于示例的研究方法,到基于大量数据的统计方法。总之,对语音的研究和处理,从策略到方法、从算法到模型,都在不断进步和更新。本书力图把最新的研究进展介绍给读者。

本书针对当前语音处理的技术介绍一些相关内容。语音处理主要包括语音识别、语音合成和语音编码等。数字信号处理和语音学作为语音处理的基础,本书特意安排了两章。本书还包括了汉语语音学基础知识、可视语音转换、语音处理的系统和应用。具体安排如下:

第1章 绪论。主要介绍言语的产生过程,语音处理研究的内容,语音处理的应用。

第2章 语音信号处理基础。介绍语音处理的信号处理基础,语音处理的基本算法。

第3章 语音基础知识与语料库。从工程的角度介绍语音学、汉语语音学和语料库基础知识。

第4章 语音编码。介绍了语音编码的基本原理、编码方法,以及一系列语音编码国际标准。

第5章 语音合成。介绍语音合成的基本原理和方法,特别介绍汉语文语转换的最新进展和技术。

第6章 语音识别。介绍语音识别的基本原理和方法,以及相关的数字信号处理算法,如动态规划算法(DTW)、矢量量化(VQ)、隐马尔可夫模型(HMM)等。

第7章 可视语音与汉语视位。介绍视位(Viseme)的概念,汉语可视语音转换(Text To Visual Speech, TTVS)。

第8章 系统与应用。介绍几个典型的语音技术和系统,如IBM的语音识别系统、ISIS多语言的智能语音系统、语音技术网络应用解决方案等。

本书作者所在的研究组成立于1979年,主要从事语音合成、语音识别、说话人识别等研究工作,具有扎实的基础理论知识和实际工作经验。

本书的第1章、第3章和第5章主要由蔡莲红执笔,第2章主要由蔡锐编写;第4章主要由黄德智编写;第6章主要由吴志勇编写;第7章主要由王志明编写;第8章由蒙美玲、柴海新、徐明星、任小山编写;参加本书编写工作的还有赵晟、蒋丹宁、崔丹丹、刘涛、王玮、陶建华、倪昕、燕鹏举、黄寅飞等。

现代语音处理技术涉及的知识面广、交叉性强、技术新、难度大。尽管作者作了很大

的努力，但是受到理论水平、实践经验所限，书中难免会有错误和遗漏，敬请读者给予批评和指正，我们不胜感谢。

目 录

第 1 章 绪论.....1	
1.1 言语过程.....2	
1.2 语音技术的研究范围.....3	
1.3 数字信号处理与数字语音信号 处理.....4	
第 2 章 语音信号处理基础.....5	
2.1 数字信号处理基础.....5	
2.1.1 离散时间信号与系统.....5	
2.1.2 离散傅立叶变换.....7	
2.1.3 Z 变换.....8	
2.1.4 离散余弦变换.....10	
2.1.5 卷积和滤波.....11	
2.2 语音信号产生模型.....14	
2.2.1 语音的产生机理.....14	
2.2.2 级联声管模型.....17	
2.2.3 语音生成模型.....22	
2.3 语音信号的时域处理.....23	
2.3.1 语音信号的抽样和量化.....23	
2.3.2 语音信号的短时 分析和预处理.....24	
2.3.3 短时能量、短时平均幅度 和短时平均过零率.....26	
2.3.4 语音的端点检测.....27	
2.3.5 短时自相关函数.....29	
2.3.6 短时基音周期估计.....30	
2.4 语音信号的频谱分析.....32	
2.4.1 短时傅立叶变换和语谱图.....32	
2.4.2 同态信号处理的基本原理.....34	
2.4.3 复倒谱和倒谱.....35	
2.5 语音信号的线性预测编码分析.....37	
2.5.1 线性预测的基本原理.....38	
2.5.2 线性预测方程组的解法.....40	
2.5.3 线谱对参数.....45	
2.6 语音信号的矢量量化.....47	
2.6.1 矢量量化的基本原理.....47	
2.6.2 失真测度.....49	
2.6.3 量化器和码本的设计.....50	
2.6.4 量化系统的复杂度控制.....52	
2.7 听觉特性和语音感知.....54	
2.7.1 听阈与听域.....55	
2.7.2 音调(pitch).....55	
2.7.3 响度级、响度与遮掩效应.....56	
第 3 章 语音基础知识与语料库.....58	
3.1 语音基础知识.....58	
3.1.1 国际音标.....58	
3.1.2 音素的分类和特点.....58	
3.2 汉语的特点.....62	
3.2.1 汉语拼音方案.....63	
3.2.2 汉语音素及其分类.....65	
3.3 汉语的音节.....66	
3.4 汉语的声调.....70	
3.4.1 汉语的声调的特点.....70	
3.4.2 声调的标记.....71	
3.4.3 声调的声学特性.....72	
3.4.4 动态声调.....75	
3.5 汉语的语调.....78	
3.5.1 关于语调.....78	
3.5.2 语调的模式.....78	
3.6 语音语料库.....79	
3.6.1 概述.....79	
3.6.2 语音语料库的设计与建设.....81	
3.6.3 TIMIT.....87	
3.7 基于语料库的语音学研究.....92	
3.7.1 声学参数的统计分析.....93	
3.7.2 音节聚类研究.....96	
3.7.3 音域模型初探.....100	
第 4 章 语音编码.....107	
4.1 语音编码基础.....107	

4.1.1 信息论基础	107	5.3.3 文本规范化(text normalization)	186
4.1.2 语音编码分类	108	5.3.4 语法分析	190
4.1.3 语音编码与其他语音 处理领域	109	5.3.5 韵律分析	201
4.2 语音编码的评价方法和依据	110	5.3.6 字音转换	203
4.2.1 语音编码的评价方法	110	5.3.7 小结	205
4.2.2 语音编码的评价依据	112	5.4 韵律建模	206
4.3 波形编码及其国际标准	113	5.4.1 韵律的描述	206
4.3.1 标量量化	113	5.4.2 韵律与句法	214
4.3.2 瞬时压扩	116	5.4.3 韵律建模	215
4.3.3 自适应差值脉冲编码	117	5.5 文语转换系统	222
4.3.4 G.721-32 kbps 自适应差值脉 冲编码调制(ADPCM)	119	5.5.1 汉语 TTS 系统 Sonic	223
4.4 参数编码与混合编码	124	5.5.2 基于大语料库的 TTS 系统	225
4.4.1 编码参数及其量化	125	5.5.3 基于匹配代价函数的 基元选取	227
4.4.2 LPC-10 声码器	126	5.5.4 权重的设定和训练	229
4.4.3 编码结构的改进	129	5.6 语音合成技术展望	230
4.4.4 低延时 CELP 16 kbps 语音编码标准	131	第 6 章 语音识别	232
4.4.5 G.723.1 双速率多媒体通信 传输语音编码器	137	6.1 语音识别基础	232
4.5 码本设计与生成	150	6.1.1 语音识别基本原理	233
4.5.1 G.728 激励码本	151	6.1.2 语音识别分类	233
4.5.2 G.729 激励码本	152	6.2 特征表示与提取	234
4.6 感知编码	154	6.2.1 LPC 倒谱系数(LPCC)	234
4.6.1 感知编码原理	155	6.2.2 Mel 频率倒谱系数 (MFCC)	236
4.6.2 感知编码算法	158	6.2.3 特征提取的具体问题	238
第 5 章 语音合成	166	6.3 模板匹配技术	239
5.1 语音合成研究的历史和现状	167	6.3.1 相似性度量	239
5.2 语音合成方法	168	6.3.2 动态时间规整(DTW)	240
5.2.1 共振峰合成	169	6.3.3 特征模板训练	242
5.2.2 波形拼接合成	174	6.4 隐马尔可夫模型(HMM)	244
5.3 TTS 系统的文本分析和韵律 预测	180	6.4.1 HMM 基本概念与原理	245
5.3.1 文本分析概述	180	6.4.2 前向概率与后向概率	247
5.3.2 文档结构分析	184	6.4.3 HMM 模型的三个问题	249
		6.4.4 HMM 模型的训练	252
		6.4.5 HMM 求解具体问题	258

6.4.6 HMM 的结构和类型	260	7.6 机器自动唇读	308
6.4.7 HMM 模型相似性比较	262	7.6.1 视觉特征	308
6.5 孤立词及连接词识别	262	7.6.2 实现方法	308
6.5.1 孤立词识别	262	7.7 双模态语音识别	310
6.5.2 HMM 模型参数选择	264	7.7.1 双模态信息融合的时间	311
6.5.3 HMM 模型参数训练—— 分段 K 均值算法	264	7.7.2 双模态信息的同步	311
6.5.4 连接词识别	265	7.7.3 确定双模态信息的 融合权重	312
6.6 连续语音识别	268	7.8 音视频映射	313
6.6.1 连续语音识别整体模型	269	7.8.1 语音特征选取	314
6.6.2 声学模型	269	7.8.2 基于矢量量化分类 的方法	314
6.6.3 语言模型	272	7.8.3 基于混合高斯模型的 方法	315
6.7 说话人识别	273	7.8.4 基于神经网络的方法	316
6.7.1 说话人识别的基本原理 与系统性能评价	274	7.8.5 基于隐马尔可夫模型 的方法	317
6.7.2 说话人识别的特征选择	276	7.8.6 基于支持向量回归的 方法	317
6.7.3 说话人识别的基本方法	278		
第 7 章 可视语音	282	第 8 章 系统与应用	321
7.1 概述	282	8.1 多模态三语种分布式口语 对话系统 ISIS	321
7.2 看得见的语音	283	8.1.1 概述	321
7.2.1 语音波形图	283	8.1.2 对 KQML 软件代理的 授权	325
7.2.2 语谱图	284	8.1.3 ISIS 中的多模态	327
7.2.3 可视发音器官的运动	286	8.1.4 小结	330
7.3 视位	288	8.2 人机口语对话系统与应用	330
7.3.1 视位的定义	288	8.2.1 口语对话系统	330
7.3.2 视位的参数表示方法	289	8.2.2 清华大学校园导游 系统 EasyNav	336
7.3.3 视位的非参数表示方法	290	8.2.3 电话航班订票与信息查询 系统 EasyFlight	341
7.3.4 汉语视位	294	8.3 炎黄之声语音合成服务器 SinoSonic	346
7.4 动态视位模型	296	8.3.1 语音合成服务器概述	347
7.4.1 基于权值融合的动 态视位模型	296	8.3.2 语音合成服务器体系 结构	347
7.4.2 视位数据的自动提取	298		
7.4.3 动态视位模型的参数 估计	302		
7.5 文本-可视语音转换	303		
7.5.1 基于参数控制的方法	304		
7.5.2 基于数据驱动的方法	306		

8.3.3 SinoSonic 语音合成 服务器的应用	350	8.4 IBM 语音解决方案简介	353
8.3.4 Monternet(移动梦网)统一消息 平台语音技术应用 TTS	352	8.4.1 桌面听写机系统(ViaVoice Desktop).....	354
8.3.5 TTS 在其他领域的应用和 今后的工作	353	8.4.2 电话语音识别系统(ViaVoice Telephony).....	355
		8.4.3 嵌入式 ViaVoice 技术	356
		8.4.4 WebSphere Voice Server	356

第 1 章 绪 论

随着计算机技术和信息技术的发展,语音交互已经成为人机交互的必要手段。

从第一台计算机诞生到现在,计算机技术经历了突飞猛进的发展。从科学研究到日常生活,计算机已渗透到人类生活的方方面面。在现代社会中,人们逐渐习惯借助计算机来完成各项事务。在这种形势下,如何让计算机智能化地与人通信,使人机交互更加自然方便成为现代计算机科学的一个重要研究课题。语音是人际交流的最习惯、最自然的方式,它将成为人机交互方式的理想选择。让说话替代键盘输入汉字是计算机用户的愿望,这个愿望正在变成现实,其技术基础是语音识别和理解。语音识别将人发出的声音、音节或短语转换成文字和符号,或给出响应,如执行控制、作出回答。

传统计算机与人交互是通过键盘和显示器,人们通过键盘或鼠标输入,通过视觉接收信息。然而听觉也是一个重要的信息通道,人们日常的信息交流大部分是通过语音来完成的。声音是人们最熟悉最习惯的传递信息方式,为计算机增加声音交互方式,使人机交互像人与人交流那样自然友好,这是人类的美好愿望。从第一台计算机诞生以来,专家们就为之付出了巨大的努力。设计师为计算机安上了“嘴巴”(扬声器),让计算机奏乐、讲话;还为计算机装上了“耳朵”(麦克风),让计算机听懂、理解人的讲话。网络专家还期望分布在不同地点的计算机成为“顺风耳”,实现音频实时传播。人们期望以最自然的方式与计算机交互,要求计算机不仅能处理文字和数字,还能处理声音和图像。最早进入千家万户的多媒体计算机(MPC)就是增加了音频的普通计算机。人们期待“语音打字机”,希望随时随地听到“有声信息”。

在信息处理中,音频和视频同时存在,没有声音的视频是不可接受的。可视电话、电视会议中的声音是优先级最高的信道。在身份识别系统中,声纹验证是重要的监测方式。人们预测,未来的计算机应该能够看、听、学,能用自然语言与人类进行交流,显然语音和音乐都是不可缺少的。

语音处理的研究已经有一百多年的历史。特别是在 20 世纪,语音处理技术有了长足的进步。自从 1939 年研制成功了第 1 个电子声码器,到目前为止语音合成已经得到广泛的应用。1952 年, Bell 实验室首次研制成功识别 10 个英文数字的装置, 1997 年 IBM 把语音识别软件推向市场。我国在语音处理方面的研究起步较晚。近二十年来,此领域的研究受到国家的重视和支持,特别是国家 863 计划的支持和评测,大大推动研究的进步和成果的应用。但是语音技术的进步与人们的期望还有较大的差距,而且语音处理的每一个成功,伴随而来的是人们对它更大的期望,研究者面对的是更难或许是不知答案的问题。

迫切的应用需求是研究的动力。语音技术的实用性、趣味性成为国内外诸多单位关注的热点,大到国家,小到个人,如美国的国防部、中国的国家高技术计划(863)和国家基础研究计划(973)等。研究语音技术的队伍十分庞大,包括公司、研究所、大学,如 IBM、Microsoft、Motorola、Siemens、Nokia 和 Toshiba 等公司;中国科学院和社会科学院等研究部门;还有许多国际知名的大学,如卡纳基梅隆(CMU)、麻省理工大学(MIT)、清华大学和北京大学等。据预测,带有语音功能的计算机将很快成为大众化产品,语音识别将可能取

代键盘和鼠标成为计算机的主要输入手段,使用户界面产生一次飞跃,所以语音识别所具有的商业前景是不可估量的。

本书的目的是让读者了解语音处理技术的基本理论和基本发展;熟悉语音编码、语音合成、语音识别的基本算法和应用。通过学习和练习,一方面为开展本领域的研究打下良好的基础,同时也为后续的开发和应用把握好方向。为此,在本章中,对上述问题进行一般性的讨论,使读者有一个全面概括的了解。

本章将介绍人类的言语过程和计算机的言语处理,即人类言语的产生和接收过程,计算机言语处理的分类,计算机语音研究的主要课题。

1.1 言语过程

人们能够讲话,也能听懂别人的讲话,但很少关注言语的产生、传播和接受的过程。实际上,从说话者开始想说到听话者对话音的理解是一个很复杂的过程。如图 1.1 所示,一般分为以下几个阶段:

- 想说:大脑产生说话的意向(intention),接着生成概念(concept),选择适当的词汇,按语法组织成语言。
- 说出:发音器官协调工作,发出声音(产生声波)。面部的肌肉、器官和体态与发音器官配合,送出多种信息,以便让听者更好地理解语音。与此同时讲话者的听觉系统接收到自己的声音,并随之修改。
- 传输:声波凭借质点运动而传播。
- 接收:人的听觉系统负责接收声波,这包括外耳、中耳、内耳。内耳的基低膜被声波刺激而振动,激发神经元产生脉冲,传递给大脑,从而感知到声音。
- 理解:听觉神经中枢收到脉冲信息,通过一系列复杂的处理过程,辨认出说话人,理解其信息内容。

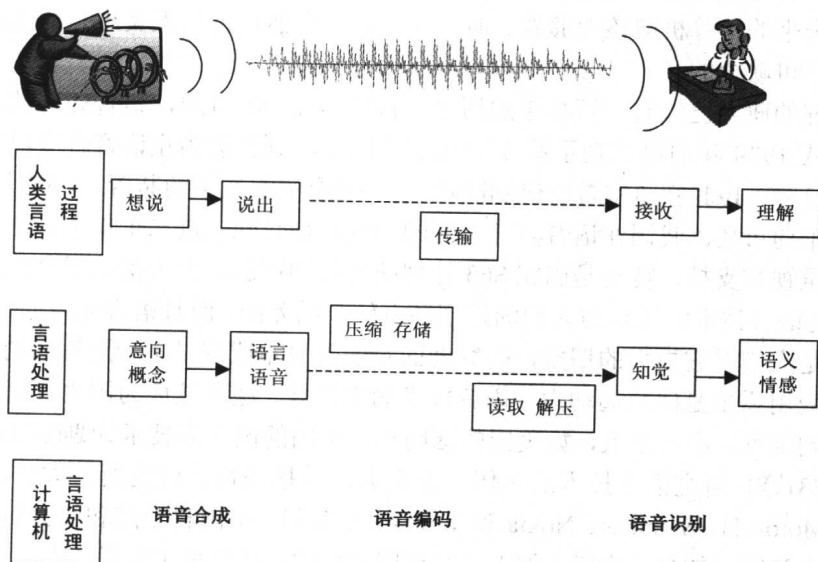


图 1.1 人类的言语过程与计算机言语处理

分析人类的言语过程，意在找出影响言语过程的因素，了解其机理；归纳出物理规律；推导出数学模型，最终研究让计算机能说、会听，能理解人的语言，并进而实现人与计算机之间的自然语言交互。

1.2 语音技术的研究范围

语音是语言的物质外壳，它涉及到生理、心理、物理、文化及社会背景等因素。因此语音处理的涉及面极广，并由此产生了一系列交叉学科分支。多个领域的专家从不同的角度来研究语音。瑞典的著名语音学家 G.Fant 把语言和听觉比喻成多边学科的花朵，如图 1.2 所示。它包括了声学、物理学、语言学、教育学、心理学、知觉、生物学、电讯技术、计算机技术和医学等。

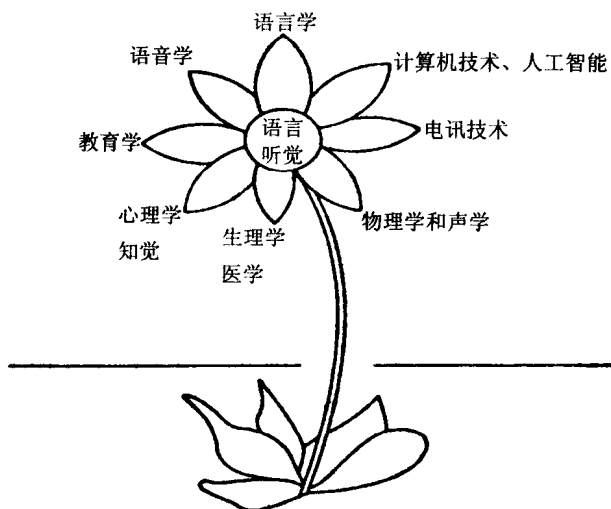


图 1.2 多边学科的花朵 (G. Fant)

这里介绍几个相关学科对语音的研究。

- 音系学：重点研究语音的差别和对立、系统和组织。这是研究语言的语音系统，对语音进行概括抽象的学科。
- 语音学：语音学研究具体客观的声音，这是语言学的基础。语音学可细分为发音语音学、声学语音学和听觉语音学。
- 心理学：研究语音的心理知觉以及与声学参数的关系。
- 信号处理：把语音看成一种信号形式，研究信号处理方法、内容描述等，包括语音的编码、识别和合成。
- 数学：研究语音信号的数学描述和模型。特别一提的是数字信号处理。20 世纪 70 年代以来，结合数字信号处理的成果，更新语音处理中策略和算法，使统计语音识别和合成有了突破性的进展。如动态规划算法(DTW)、矢量量化(VQ)、隐马尔可夫模型(HMM)、线性预测(LPC)、基音同步叠加(PSOLA)和人工神经网络(NN)等。

语音是人类交流的重要手段，也是人类期望的人与计算机交互的方式，为此语音信号受到信息科学研究领域的关注，成为智能计算机的重要研究方向。他们从数字信号处理出发，利用相关学科的研究成果，研究语音信号的处理算法和模型、系统和应用。

1.3 数字信号处理与数字语音信号处理

计算机内的信息表示都是数字(非模拟)，因此语音信号也是数字序列。数字信号处理方法和算法是本书的基础，如抽样、相关、傅立叶变换、线性预测分析和矢量量化等。

信号处理就是对信号进行分析、变换和滤波等处理。例如频谱变换、低通滤波和平均幅度计算。数字信号处理就是采用数字计算技术对信号进行处理的学科。数字语音信号处理就是数字信号处理技术对语音信号处理的学科。

语音信号处理的特点如下：

- 语音信号是时间依赖的连续媒体。因此处理的时序性要求很高，如果在时间上有 25 ms 的延迟，就会感到不连续，语音编码的延时必须很小。
- 由于语音信号不仅仅是声音的载体，同时还携带了情感的意向，故对语音信号的处理，不仅是信号处理问题，还要抽取语义等其他信息。因此可能会涉及到语言学、社会学、声学等学科。
- 一个完整的语音交互过程，应该包括语言的产生和语言的接收，也就是“说”和“听”。虽然语音处理是针对语音信号进行的，但是为了研究更好的处理算法，应该关注语音产生机理和听觉特性。

从人与计算机交互的角度来看语音信号的处理应包括：语音信号获取域表示、语音识别与理解、语音合成、语音编码、语音识别和说话人识别等。

第2章 语音信号处理基础

语音信号数字处理是研究用数字信号处理的技术对语音信号进行处理的学科。20世纪中叶以来,语音的研究与应用取得了一系列重大的进步,例如语音识别、语音合成、语音的压缩编码和语音的远距离传输等都是建立在语音信号数字处理的基础上的。语音信号数字处理是一门综合性的学科,包括计算机科学、语音学、语言学、声学、生理学、心理学和数学等诸多领域的内容。本章作为全书的基础章节,将简要介绍以下内容:语音的生成机理及其相应的产生模型;用数字信号处理技术对语音信号进行处理的基本方法;语音信号中常见的时域处理方法和频谱分析方法;线性预测分析方法的基本原理和方法;矢量量化技术的原理和基本算法;听觉系统特性和语音感知。此外,本章的第二小节简要地介绍了部分数字信号处理的基础供参考,有这方面知识的读者可以跳过这一小节。限于篇幅所限,本章介绍的内容都较为浅显,有需要的读者请自行查阅相应的参考书籍。

2.1 数字信号处理基础

随着计算机技术的飞速发展,数字信号处理的技术广泛应用于信号处理领域。本节将简要介绍数字信号处理的一些基础知识,包括:离散时间信号与系统、离散傅立叶变换、Z变换、离散余弦变换以及卷积和滤波的概念、计算方法和特性。由于篇幅所限,不详尽之处请读者查阅相关资料。

2.1.1 离散时间信号与系统

按照时间函数取值的连续性与离散性可将信号划分为连续时间信号与离散时间信号。离散信号在时间上是离散的,只在某些不连续的规定瞬间给出函数值,在其他时间没有定义。如果离散时间信号的幅值是连续的,则离散信号又叫抽样信号;如果幅值也是离散的,这样的信号又成为数字信号,如图2.1所示。

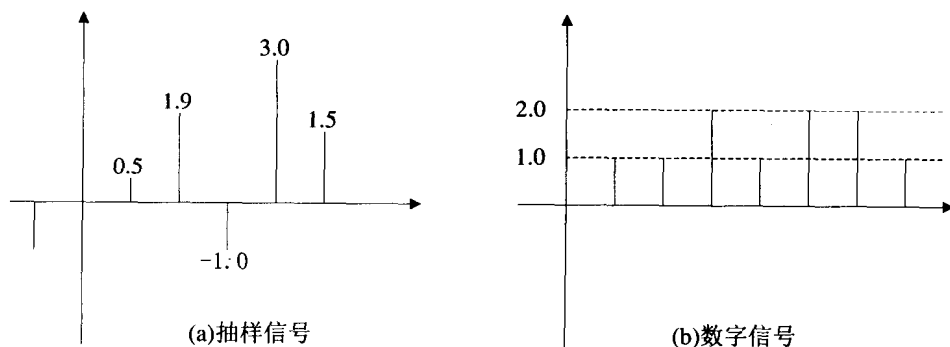


图 2.1 离散信号

通常,给出函数值的离散时刻之间的间隔是均匀的,一般以 $x(n)$ ($n=0, \pm 1, \pm 2, \dots$) 来表示

此序列, 这里 n 表示各函数值在序列中出现的序号。也可以说, 一个离散信号就是一组序列值的集合 $\{x(n)\}$ 。在离散信号的分析中, 常会遇到一些基本运算, 比如加、乘和时延。离散信号序列 $\{x(n)\}$ 和 $\{y(n)\}$ 相加是指两序列同序号的数值逐项对应相加而构成一个新序列 $\{z(n)\}$

$$z(n) = x(n) + y(n) \quad (n = 0, \pm 1, \pm 2, \dots) \quad (2-1)$$

类似的有相乘的定义

$$z(n) = x(n) \cdot y(n) \quad (n = 0, \pm 1, \pm 2, \dots) \quad (2-2)$$

序列时延是指原序列逐项依次延时 m 位而生成一个新的序列

$$z(n) = x(n-m) \quad (n = 0, \pm 1, \pm 2, \dots) \quad (2-3)$$

如果 $m > 0$ 意味着序列后移, 反之, 如果 $m < 0$ 意味着序列前移。序列的能量定义为

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2 \quad (2-4)$$

一个离散时间系统的示意图如图 2.2 所示, 其激励信号是离散信号序列 $x(n)$, 响应序列为 $y(n)$, 系统完成了两者之间的转换运算。



图 2.2 离散时间系统

按离散时间系统的性能可以划分成许多种类型, 包括线性/非线性、时变/时不变等。最常见的系统是线性时不变系统, 其最大特点就是均匀性和叠加性, 即: 若 $y_1(n)$ 和 $y_2(n)$ 分别是激励 $x_1(n)$ 和 $x_2(n)$ 的输出, 那么激励 $\alpha x_1(n) + \beta x_2(n)$ (α, β 为常数) 的响应为 $\alpha y_1(n) + \beta y_2(n)$, 如图 2.3 所示。而时不变系统指的是系统响应与激励作用于系统的时间无关, 即激励 $x(n-t)$ 将产生响应 $y(n-t)$, 如图 2.4 所示。

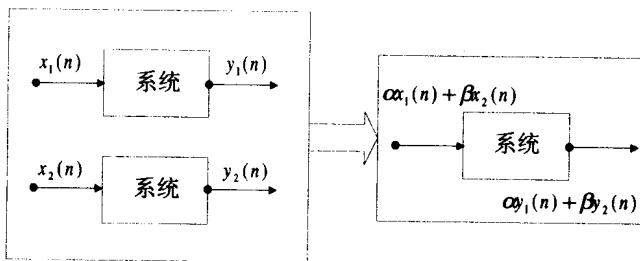


图 2.3 线性系统的均匀性和叠加性

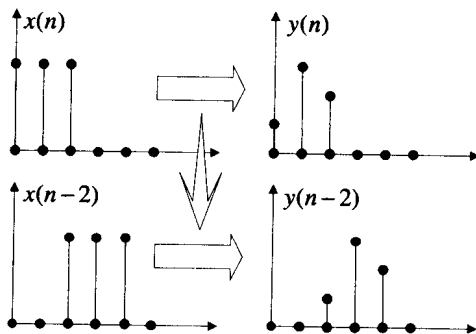


图 2.4 系统的时不变性

在连续时间系统, 信号是时间变量的连续函数, 系统可以用微积分方程来描述; 在离散时间系统, 信号的变量是离散的整型值, 因此系统的行为和性能需要用差分方程来表示。在差分方程中, 构成方程的各项包含离散变量的函数 $x(n)$, 以及此函数产生位移后得到的 $x(n+1)$ 、 $x(n-1)$ 等。求解差分方程的方法有很多, 包括迭代法、时域经典法、卷积方法和变换域法等, 限于篇幅, 请读者自行查阅相应的参考书籍。

2.1.2 离散傅立叶变换

法国数学家和物理学家 J.B.J.Fourier(1769~1830)于 1807 年提出, 任何一个连续周期信号可以分解为一族正弦曲线和余弦曲线。后来这个定理被扩充, 也适用于离散信号和非周期信号, 统称为傅立叶分析。如图 2.5 所示, 信号被分解为 9 个正弦信号和 9 个余弦信号共 18 个信号。

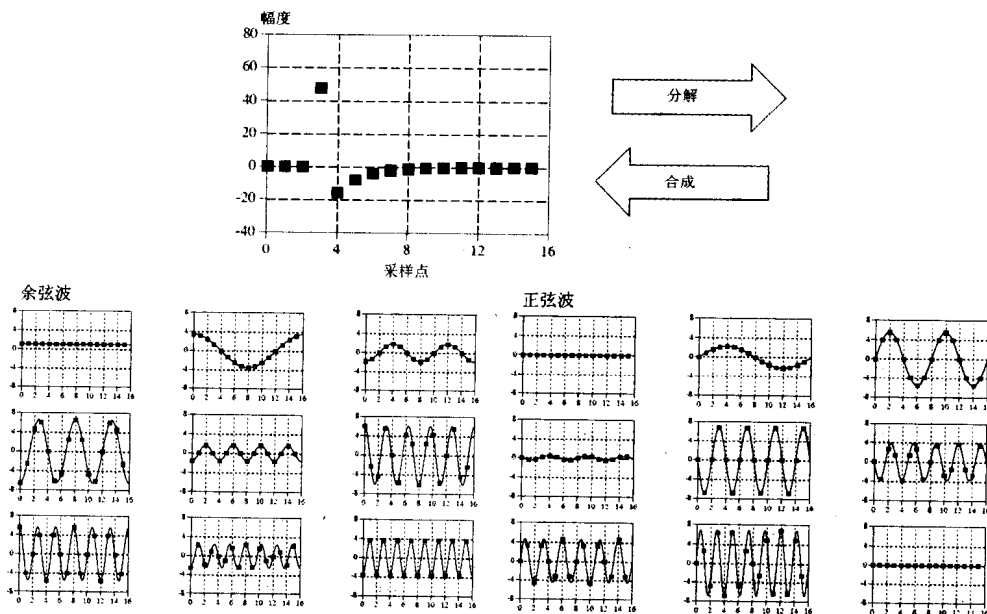


图 2.5 信号的傅立叶分解

根据被分解信号类型的不同, 相应的傅立叶分析也有着不同的定义, 如图 2.6。

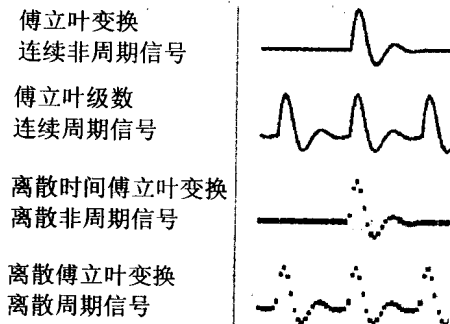


图 2.6 4 种傅立叶分析

- 傅立叶变换(Fourier Transform): 连续非周期信号。
- 傅立叶级数(Fourier Series): 连续周期信号。
- 离散时间傅立叶变换(Discrete Time Fourier Transform): 离散非周期信号。
- 离散傅立叶变换(Discrete Fourier Transform): 离散周期信号。

由于篇幅所限, 这里我们着重于介绍离散傅立叶变换(DFT), 因为它在数字信号处理领域中应用最广泛, 图 2.5 中的傅立叶分解就是 DFT。

由于 DFT 的作用对象是离散周期信号, 对于一个有限长序列 $x(n)$ ($0 \leq n \leq N-1$), 我们必须进行延拓来将其扩展成周期信号, 如图 2.7 所示。

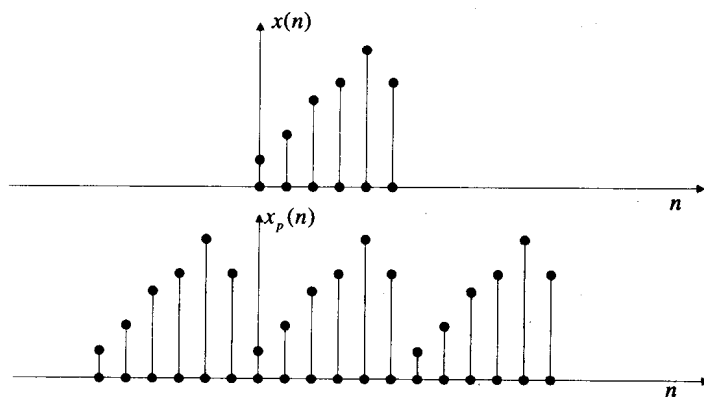


图 2.7 有限长序列 $x(n)$ 延拓成周期序列 $x_p(n)$

离散傅立叶变换 DFT[] 及其反变换 IDFT[] 的定义为

$$\begin{cases} X(k) = \text{DFT}[x(n)] = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi nk}{N}} & (0 \leq k \leq N-1) \\ x(n) = \text{IDFT}[X(k)] = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi nk}{N}} & (0 \leq n \leq N-1) \end{cases} \quad (2-5)$$

在语音信号数字处理领域中, 离散傅立叶变换扮演了重要的角色, 完成了信号序列从时域到频域的转化, 为语音的频域分析奠定了基础。为了提高 DFT 的计算速度, J.W.Cooley 和 J.W.Tukey 于 1965 年提出了著名的快速傅立叶算法(FFT, fast fourier transform), 极大地提高了傅立叶变换的速度, 使傅立叶算法得以真正地应用于实用场合, 如今, 大量的硬件已经集成了进行 FFT 计算的功能, 而且对该算法的改进也在不断进行, 如 Winograd 算法和 Mason 算法。

2.1.3 Z 变换

在离散信号与系统的理论研究中, Z 变换是一种重要的数学工具, 它把离散系统的数学模型——差分方程, 转化为简单的代数方程, 从而大大简化了求解过程。尤其是随着计算机技术的普及和使用, Z 变换得到了越来越广泛的应用。(注: 本节提及的 Z 变换为双边 Z 变换, 单边 Z 变换的定义请读者自行查阅相应书籍。)

信号序列 $x(n)$ 的 Z 变换定义为