



“十二五”江苏省高等学校重点教材（编号2014-1-102）
高等院校通信与信息专业规划教材

语音信号处理

第3版

SPEECH SIGNAL PROCESSING



赵力 等编著



机械工业出版社
CHINA MACHINE PRESS



本书赠送电子课件



“十二五”江苏省高等学校重点教材 (编号 2014-1-102)
高等院校通信与信息专业规划教材

语音信号处理

第3版

赵力等编著



机械工业出版社

本书介绍了语音信号处理的基础、原理、方法和应用,以及该学科领域近年来取得的一些新的研究成果和技术。全书共分13章,内容包括:绪论、语音信号处理的基础知识、语音信号处理的常用算法、语音信号分析、语音信号特征提取技术、语音增强、语音识别、说话人识别、语音编码、语音合成与转换、语音信号情感处理、语音隐藏、声源定位。

本书可作为高等院校的教材或教学参考书,同时也可供语音信号处理等领域的工程技术人员参考。

本书提供电子课件,需要的教师可登录 www.cmpedu.com 免费注册,审核通过后下载,或联系编辑索取(QQ: 241151483, 电话: 010-88379753)。

图书在版编目(CIP)数据

语音信号处理 / 赵力等编著. —3版. —北京: 机械工业出版社, 2016.3
高等院校通信与信息专业规划教材
ISBN 978-7-111-53460-0

I. ①语… II. ①赵… III. ①语声信号处理—高等学校—教材
IV. ①TN912.3

中国版本图书馆CIP数据核字(2016)第070273号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

责任编辑: 李馨馨 责任校对: 张艳霞

责任印制: 李洋

北京振兴源印务有限公司印刷

2016年5月第3版·第1次印刷

184mm×260mm·18.5印张·454千字

0001—3000册

标准书号: ISBN 978-7-111-53460-0

定价: 45.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线: (010) 88379833

机工官网: www.cmpbook.com

读者购书热线: (010) 88379649

机工官博: weibo.com/cmp1952

教育服务网: www.cmpedu.com

封面无防伪标均为盗版

金书网: www.golden-book.com

高等院校通信与信息专业教材编委会名单

(按姓氏笔画排序)

编委员会主任	乐光新	北京邮电大学
编委会副主任	王金龙	解放军理工大学
	刘 陈	南京邮电大学
	张文军	上海交通大学
	张思东	北京交通大学
	杨海平	解放军理工大学
	徐澄圻	南京邮电大学
	彭启琮	电子科技大学
	曾孝平	重庆大学
编委会委员	王成华	南京航空航天大学
	王建新	南京理工大学
	冯正和	清华大学
	卢官明	南京邮电大学
	刘富强	同济大学
	刘增基	西安电子科技大学
	余 翔	重庆邮电大学
	张邦宁	解放军理工大学
	张玲华	南京邮电大学
	李少洪	北京航空航天大学
	邹家禄	东南大学
	南利平	北京信息科技大学
	赵 力	南京邮电大学
	赵尔沅	北京邮电大学
	徐惠民	北京邮电大学
	舒 勤	四川大学
秘书长	胡毓坚	机械工业出版社
副秘书长	许晔峰	解放军理工大学

前 言

本书是根据机械工业出版社高等院校通信与信息专业规划教材出版规划,由通信与信息专业规划教材编审委员会编审、推荐出版的。自从2009年3月第2版出版以来,时间已过去了近6个年头。6年来,本学科领域的理论与实践研究迅速发展,分析方法不断更新,技术应用范围日益扩展,对本教材的内容的更新和结构体系的进一步完善提出了更高的要求。面对这一情况,我们结合教学实践,在广泛听取从事本课程教学和研究的教师与学术专家意见的基础上,逐步明确了编写本书第3版的追求目标:即在相对稳定中力求变革,处理好经典理论的论述与最新技术的相互融合。正是在这样的指导思想下,编者对第2版教材进行了修订、补充和更新。

新版教材力求系统地反映语音信号处理的基本原理与方法,以及近年来该领域的新进展和新技术;突出基本概念、原理、方法、应用、研究现状及学科发展趋势。在结构上,按照基础、分析、处理、应用的顺序组织材料,从最经典的技术与应用逐步过渡到最新、最热门的技术与应用中,使之既能满足教学需要,又反映出本学科领域近年来发展的新成果。

第3版教材与第2版相比,除了增减了部分章节以外,基本保持了原作风貌,认真修订了第2版中的部分错误和疏漏。根据作者多年来给本科生和硕士研究生讲授“语音信号处理”课程的体会,做了两个主要的改变。首先,在结构上,将第2版教材中的矢量量化、隐马尔可夫模型和神经网络等章节中较为零散、繁杂的内容进行了适当删减,最终合并成一章,这样可以将后续语音信号处理技术与应用中用到的相关理论在这一章中进行集中介绍,从而更加突出后续内容。其次,增加了一些现在较流行的内容,如语音信号的情感处理、语音隐藏、声源定位等方向中出现的新方法和新应用场景等。

本书主要面向信号与信息处理、电路与系统、通信与电子工程、模式识别与人工智能、计算机信息处理等学科有关专业的本科高年级学生和研究生,也可以作为从事语音信号处理这一领域科研工作的技术人员的参考书。

本书的参考学时为本科生32学时、研究生40学时,可以根据不同的教学要求对其内容进行适当取舍,灵活安排讲课时数。

本书由赵力、梁瑞宇、魏昕、奚吉编写。本书的出版得到了江苏高校品牌专业建设工程项目PPZY2015A035的资助。作者参考和引用了一些学者的研究成果、著作和论文,具体出处见参考文献。在此,作者向这些文献的著作者表示敬意和感谢。

语音信号处理是一门理论性强、实用面广、内容新、难度大的交叉学科,同时这门学科又处于快速发展之中,尽管作者在编写过程中始终注重理论紧密联系实际,力求以尽可能简明、通俗的语言,深入浅出地将这门学科介绍给读者,但因作者水平有限,缺点错误在所难免,敬请广大读者批评指正。

为配合本书教学,本书配套了《语音信号处理实验教程》(书号:978-7-111-53071-8),以方便教师根据主教材来合理地组织实验教学,加深学生对所学知识的理解。

2015年5月

作者

目 录

前言	
第1章 绪论	1
第2章 语音信号处理的基础知识	5
2.1 语音发音及感知系统	5
2.1.1 语音发音系统	5
2.1.2 语音听觉系统	6
2.1.3 人耳听觉特性	7
2.2 语音信号生成的数学模型	10
2.2.1 激励模型	11
2.2.2 声道模型	12
2.2.3 辐射模型	15
2.2.4 语音信号的数字模型	15
2.3 语音基本概念与参数	16
2.3.1 声压与声强	16
2.3.2 响度	17
2.3.3 频率与音高	17
2.4 语音信号的数字化和预处理	17
2.4.1 预滤波、采样、A/D转换	18
2.4.2 预加重与去加重	18
2.5 语音信号处理的应用	19
2.6 思考与复习题	22
第3章 语音信号处理的常用算法	23
3.1 矢量量化	23
3.1.1 矢量量化的基本原理	23
3.1.2 矢量量化的失真测度	26
3.1.3 线性预测失真测度	27
3.1.4 识别失真测度	28
3.1.5 矢量量化器的最佳码本设计	29
3.2 隐马尔可夫模型	31
3.2.1 概述	31
3.2.2 隐马尔可夫模型的定义	34
3.2.3 隐马尔可夫模型的基本算法	36

3.2.4 隐马尔可夫模型的结构类型	41
3.3 神经网络	43
3.3.1 神经网络概述	43
3.3.2 人工神经网络的构成	45
3.3.3 神经网络模型及算法	47
3.3.4 基于神经网络的模式识别应用	52
3.4 思考与复习题	53
第4章 语音信号分析	55
4.1 概述	55
4.2 语音分帧	55
4.3 语音信号的时域分析	57
4.3.1 短时能量及短时平均幅度分析	58
4.3.2 短时过零率分析	59
4.3.3 短时相关分析	60
4.3.4 短时平均幅度差函数	63
4.4 语音信号的频域分析	64
4.4.1 利用短时傅里叶变换求语音的短时谱	65
4.4.2 语音的短时谱的临界带特征矢量	66
4.5 语音信号的倒谱分析	67
4.5.1 同态信号处理的基本原理	67
4.5.2 复倒谱和倒谱	68
4.5.3 Mel 频率倒谱系数	70
4.6 语音信号的线性预测分析	72
4.6.1 线性预测分析的基本原理	72
4.6.2 线性预测方程组的求解	74
4.6.3 LPC 谱估计和 LPC 复倒谱	76
4.6.4 线谱对 (LSP) 分析	78
4.6.5 LPC 与 LSP 参数的转换	79
4.7 语音信号的小波分析	80
4.7.1 短时傅里叶变换	81
4.7.2 连续小波变换	81
4.7.3 离散小波变换	82
4.8 思考与复习题	84
第5章 语音信号特征提取技术	85
5.1 概述	85
5.2 端点检测	85
5.2.1 双门限法	85
5.2.2 自相关法	87

5.2.3	谱熵法	88
5.2.4	比例法	89
5.2.5	对数频谱距离法	90
5.3	基音周期估计	92
5.3.1	自相关法	92
5.3.2	平均幅度差函数法	95
5.3.3	倒谱法	96
5.3.4	简化逆滤波法	98
5.3.5	基音检测的后处理	99
5.4	共振峰估计	100
5.4.1	带通滤波器组法	101
5.4.2	倒谱法	102
5.4.3	线性预测法	102
5.5	思考与复习题	104
第6章	语音增强	105
6.1	概述	105
6.2	语音特性、人耳感知特性及噪声特性	105
6.2.1	语音特性	105
6.2.2	人耳感知特性	106
6.2.3	噪声特性	106
6.3	滤波器法	107
6.3.1	陷波器法	107
6.3.2	自适应滤波器	108
6.4	相关特征法	110
6.4.1	自相关处理抗噪法语音增强技术	110
6.4.2	利用复数帧段主分量特征的降噪方法	111
6.5	非线性处理法	112
6.5.1	小波降噪法	112
6.5.2	同态滤波法	113
6.6	减谱法	115
6.6.1	基本原理	115
6.6.2	基本减谱法的改进	116
6.7	Weiner 滤波法	117
6.7.1	基本原理	117
6.7.2	Weiner 滤波的改进形式	118
6.8	思考与复习题	118
第7章	语音识别	120
7.1	概述	120

7.2	语音识别原理和识别系统的组成	123
7.2.1	预处理和参数分析	124
7.2.2	语音识别系统构成	126
7.3	孤立字(词)识别系统	129
7.3.1	基于改进的二次分类函数的汉语塞音语音识别系统	130
7.3.2	基于动态时间规整(DTW)的孤立字(词)识别系统	132
7.3.3	基于概率尺度DP识别方法的孤立字(词)识别系统	134
7.3.4	基于隐马尔可夫模型(HMM)的孤立字(词)识别系统	135
7.4	连续语音识别系统	136
7.5	连续语音识别系统的性能评测	139
7.5.1	评测方法及系统复杂性和识别能力的测度	139
7.5.2	综合评估需要考虑的其他因素	142
7.6	思考与复习题	143
第8章	说话人识别	144
8.1	概述	144
8.2	说话人识别方法和系统结构	145
8.2.1	预处理	146
8.2.2	说话人识别特征的选取	146
8.2.3	特征参量评价方法	148
8.2.4	模式匹配方法	149
8.2.5	说话人识别中判别方法和阈值的选择	150
8.2.6	说话人识别系统的评价	151
8.3	应用VQ的说话人识别系统	151
8.4	应用DTW的说话人确认系统	153
8.5	应用HMM的说话人识别系统	154
8.5.1	基于HMM的与文本有关的说话人识别	154
8.5.2	基于HMM的与文本无关的说话人识别	155
8.5.3	基于HMM的指定文本型说话人识别	155
8.5.4	说话人识别HMM的学习方法	156
8.5.5	鲁棒的HMM说话人识别技术	156
8.6	应用GMM的说话人识别系统	157
8.6.1	GMM的基本概念	157
8.6.2	GMM的参数估计	157
8.6.3	训练数据不充分的问题	159
8.6.4	GMM模型的识别问题	159
8.7	尚需进一步探索的研究课题	160
8.8	思考与复习题	161
第9章	语音编码	163

9.1 概述	163
9.2 语音编码的评价指标	165
9.2.1 语音编码的依据	165
9.2.2 语音压缩系统的性能指标和评测方法	167
9.3 语音信号的波形编码	172
9.3.1 脉冲编码调制 (PCM)	172
9.3.2 自适应预测编码 (APC)	176
9.3.3 自适应增量调制和自适应差分脉冲编码调制	177
9.3.4 子带编码 (SBC)	180
9.4 语音信号参数编码	186
9.4.1 线性预测声码器	186
9.4.2 LPC-10 编码器	188
9.5 语音信号的混合编码	191
9.6 思考与复习题	194
第 10 章 语音合成与转换	196
10.1 概述	196
10.2 语音合成算法	197
10.2.1 共振峰合成法	199
10.2.2 线性预测合成法	201
10.2.3 PSOLA 算法合成语音	202
10.3 文语转换系统 (TTS)	205
10.4 常用语音转换的方法	207
10.4.1 频谱特征参数转换	207
10.4.2 基音周期转换	209
10.4.3 韵律信息转换	210
10.5 语音转换的研究方向	211
10.6 思考与复习题	211
第 11 章 语音信号情感处理	213
11.1 概述	213
11.2 情感理论与情感诱发实验	214
11.2.1 情感的心理理论	214
11.2.2 实用语音情感数据库的建立	215
11.2.3 情感语料的诱发方法	217
11.2.4 情感语料的主观评价方法	219
11.3 情感的声学特征分析	221
11.3.1 情感特征提取	221
11.3.2 特征降维算法	226
11.4 实用语音情感的识别算法研究	228

11.4.1	支持向量机	228
11.4.2	K 近邻分类器	229
11.4.3	高斯混合模型的基本原理	230
11.5	跨语言的语音情感识别	232
11.6	应用与展望	233
11.6.1	载人航天中的应用的设想	233
11.6.2	情感多媒体搜索	234
11.6.3	个人机器人	235
11.6.4	总结与展望	235
11.7	思考与复习题	236
第 12 章	语音隐藏	237
12.1	概述	237
12.2	国内外研究现状及存在的问题	238
12.2.1	信息隐藏学科发展	238
12.2.2	语音信息隐藏国内外研究现状	239
12.2.3	存在的主要问题	240
12.3	语音信息隐藏基础	241
12.3.1	语音信息隐藏基本概念及理论	241
12.3.2	信息隐藏算法分类	242
12.3.3	信息隐藏技术应用	243
12.4	常用评价指标	244
12.4.1	感知透明性	244
12.4.2	鲁棒性	245
12.4.3	信息容量	246
12.5	基本语音信息隐藏算法	246
12.5.1	回声隐藏算法	246
12.5.2	相位编码算法	247
12.5.3	扩频算法	247
12.5.4	Patchwork 算法	248
12.5.5	量化算法	249
12.6	总结与展望	249
12.7	思考与复习题	251
第 13 章	声源定位	252
13.1	概述	252
13.2	双耳听觉定位原理及方法	253
13.2.1	人耳听觉定位原理	253
13.2.2	人耳声源定位线索	255
13.2.3	声源估计方法	256



13.3 传声器阵列模型.....	258
13.3.1 均匀线阵.....	259
13.3.2 均匀圆阵.....	260
13.4 基于传声器阵列的声源定位算法分类.....	260
13.4.1 基于最大输出功率的可控波束形成算法.....	261
13.4.2 基于到达时间差的定位算法.....	262
13.4.3 基于高分辨率谱估计的定位算法.....	265
13.5 总结与展望.....	270
13.6 思考与复习题.....	271
附录 汉英名词术语对照.....	272
参考文献.....	280

第1章 绪 论

通过语音传递信息是人类最重要、最有效、最常用和最方便的交换信息的形式。语言是人类特有的功能，声音是人类常用的工具，是相互传递信息的最主要的手段。因此，语音信号是人们构成思想疏通和感情交流的最主要的途径。并且，由于语言和语音与人的智力活动密切相关，与社会文化和进步紧密相连，所以它具有最大的信息容量和最高的智能水平。现在，人类已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、获取和应用语音信息，这对于促进社会的发展具有十分重要的意义。

让计算机能听懂人类的语言，是人类自计算机诞生以来梦寐以求的想法。随着计算机越来越向便携化方向发展，以及计算环境的日趋复杂化，人们越来越迫切地要求摆脱键盘的束缚而代之以语音输入这样便于使用的、自然的、人性化的输入方式。尤其是汉语，它的汉字输入一直是计算机应用普及的障碍，因此，利用汉语语音进行人机交互是一个极其重要的研究课题。作为高科技应用领域的研究热点，语音信号处理技术从理论研究到产品开发已经走过了几十个春秋并且取得了长足的进步。它正在直接与办公、交通、金融、公安、商业、旅游等行业的语音咨询与管理，工业生产部门的语音控制，电话·电信系统的自动拨号、辅助控制与查询，以及医疗卫生和福利事业的生活支援系统等各种实际应用领域相接轨，并且有望成为下一代操作系统和应用程序的用户界面。可见，语音信号处理技术的研究将是一项极具市场价值和挑战性的工作。我们今天进行这一领域的研究与开拓就是要让语音信号处理技术走入人们的日常生活当中，并不断朝向更高目标而努力。

语音信号处理这门学科之所以能够长期地、深深地吸引广大科学工作者不断地对其进行研究和探讨，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。语音信号处理是以语音语言学和数字信号处理为基础而形成的一门涉及面很广的综合性的学科，与心理·生理学、计算机科学、通信与信息科学以及模式识别和人工智能等学科都有着非常密切的关系。对语音信号处理的研究一直是数字信号处理技术发展的重要推动力量。因为许多处理的新方法的提出，都是先在语音处理领域中获得成功，然后再推广到其他领域的。例如，许多高速信号处理器的诞生和发展是与语音信号处理的研究发展分不开的，语音信号处理算法的复杂性和实时处理的要求，促使人们去设计许多先进的高速信号处理器。这种产品问世之后，又首先在语音信号处理应用中得到最有效的推广应用。语音信号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。

语音信号处理作为一个重要的研究领域，已经有很长的研究历史。但是它的快速发展可以说是从1940年前后Dudley的声码器(Vocoder)和Potter等人的可见语音(Visible Speech)开始的。1952年，贝尔(Bell)实验室的Davis等人首次研制成功能识别10个英语数字的实验装置；1956年，Olson和Belar等人采用8个带通滤波器组提取频谱参数作为语音的特征，研制成功一台简单的语音打字机。20世纪60年代前期，由于Faut和Stevens的努力，奠定了语音生成理论的基础，在此基础上语音合成的研究得到了扎实的进展。60年

代中期形成的一系列数字信号处理方法和技術,如数字滤波器、快速傅里叶变换(FFT)等成为语音信号数字处理的理論和技術基础。在方法上,随着电子计算机的发展,以往的以硬件为中心的研究逐渐转化为以软件为主的處理研究。然而,在语音识别领域内,初期有几种语音打字机的研究也很活跃,但后来已全部停了下来,这说明了当时人们对语音识别难度的认识得到了加深。所以1969年美国贝尔研究所的Pierce感叹地说“语音识别向何处去?”。

到了1970年,好似反驳Pierce的批评,单词识别装置开始了实用化阶段,其后实用化的进程进一步高涨,实用机的生产销售也上了轨道。此外社会上所宣传的声纹(Voice Print)识别,即说话人识别的研究也扎扎实实地开展起来,并很快达到了实用化的阶段。到了1971年,美国ARPA(American Research Projects Agency)为主导的“语音理解系统”的研究计划也开始起步。这个研究计划不仅在美国国内,而且对世界各国都产生了很大的影响。它促进了连续语音识别研究的兴起。历时5年的庞大的ARPA研究计划,虽然在语音理解、语言统计模型等方面的研究积累了一些经验,取得了许多成果,但没能达到巨大投资应得的成果,在1976年停了下来,进入了深刻的反省阶段。但是,在整个20世纪70年代还是有几项研究成果对语音信号处理技术的进步和发展产生了重大的影响。这就是70年代初由板仓(Itakura)提出的动态时间规整(DTW)技术,使语音识别研究在匹配算法方面开辟了新思路;70年代中期线性预测技术(LPC)被用于语音信号处理,此后隐马尔可夫模型法(HMM)也获得初步成功,该技术后来在语音信号处理的多个方面获得巨大成功;70年代末,Linda、Buzo、Gray和Markel等人首次解决了矢量量化(VQ)码书生成的方法,并首先将矢量量化技术用于语音编码获得成功。从此矢量量化技术不仅在语音识别、语音编码和说话人识别等方面发挥了重要作用,而且很快推广到其他许多领域。因此,80年代开始出现的语音信号处理技术产品化的热潮,与上述语音信号处理新技术的推动作用是分不开的。

20世纪80年代期间,由于矢量量化、隐马尔可夫模型和人工神经网络(ANN)等相继被应用于语音信号处理,并经过不断改进与完善,使得语音信号处理技术产生了突破性的进展。其中,隐马尔可夫模型作为语音信号的一种统计模型,在语音信号处理的各个领域获得了广泛的应用。其理论基础是1970年前后,由Baum等人建立起来的,随后,由美国卡内基·梅隆大学(CMU)的Baker和美国IBM公司的Jelinek等人将其应用到语音识别中。由于美国贝尔实验室的Rabiner等人在80年代中期,对隐马尔可夫模型深入浅出的介绍,才使世界各国从事语音信号处理的研究人员所了解和熟悉,进而成为一个公认的研究热点,也是目前语音识别等的主流研究途径。

20世纪90年代以来,语音信号处理在实用化方面取得了许多实质性的研究进展。其中,语音识别逐渐由实验室走向实用化。一方面,对声学语音学统计模型的研究逐渐深入,鲁棒的语音识别、基于语音段的建模方法及隐马尔可夫模型与人工神经网络的结合成为研究的热点。另一方面,为了语音识别实用化的需要,讲者自适应、听觉模型、快速搜索识别算法以及进一步的语言模型的研究等课题倍受关注。

在语音合成方面,有限词汇的语音合成群已在自动报时、报警、报站、电话查询服务、发音玩具等方面得到了广泛的应用。关于文本—语音自动转换系统(TTS)的研究,许多国家、多个语种都已在20世纪90年代初达到了商品化程度,其语音质量能为广大公众接受。从研究技术上可分为发音器官参数合成、声道模型参数合成和波形编辑合成;从合成策略上

讲可分为频谱逼近合成和波形逼近合成。其中采用波形拼接来合成语音的方法,越来越被广泛地应用。其中最具代表性的是基音同步叠加法(PSOLA),这种方法既能保持所发语音的主要音段特征,又能在拼接时灵活调整其基频、时长和强度等超音段特征,在语音合成中影响较大。

在50多年的时间里,语音编码已得到了迅速的发展。最早的标准化语音编码系统是速率为64 kbit/s的PCM波形编码器;到20世纪90年代中期,速率为4~8 kbit/s的波形与参数混合编码器,在语音质量上已接近前者的水平,且已达到实用化阶段。当前的研究主要集中在4 kbit/s码率以下的高音质、低延迟的声码器,提高在噪声信道中低码率编码器的性能,并能传输多种信号,包括音频信号。为此在寻找更为有效的参数量化技术、非线性预测技术(Non-Linear Prediction)、多分辨率时频分析技术(如Wavelets)和高阶统计量的使用、对人耳感知特性的进一步研究和探索等方面有较多的研究工作。

说话人识别和语种辨识是语音识别的两种特殊形式。它们和语音识别一样,都是通过提取语音信号的特征和建立相应的模型进行分类判断的。说话人识别力求找出包含在语音信号中的说话人的个性因素,强调不同人之间的特征差异;而语种辨别则要从一个语音片段中判别它是哪个语种,所以就要尽可能找出不同语种的差别特征。目前,这方面的研究重点转向对各种声学参数的线性或非线性处理以及新的模式匹配方法上,如DTW、主分量(成分)分析(PCA)、隐马尔可夫模型与神经网络组合等技术上。

包含在语音信号中的情感信息是一种很重要的信息资源,它是人们感知事物的必不可少的部分信息。例如,同样的一句话,由于说话人表现的情感不同,在给听者的感知上就可能会有较大的差别。所谓“听话听音”就是这个道理。然而传统的语音信号处理技术把这部分信息,作为模式的变动和差异,通过规则化噪声处理给去掉了。实际上,人们是同时接受各种形式的信息的,怎样有效地利用各种形式的信息以达到最佳的信息传递和交流效果,是今后信息处理研究的发展方向。所以包含在语音信号中的情感信息的计算机处理研究,分析和处理语音信号中的情感特征、判断和模拟说话人的喜怒哀乐等是一个意义重大的研究课题,也是20世纪90年代以来兴起的一个新的语音信号处理研究领域。

有关抗噪声技术的研究以及实际环境下的语音信号处理系统的开发,在国内外作为语音信号处理的非常重要的研究课题,已经做了大量的研究工作,取得了丰富的研究成果。目前国内外的研究成果大体分为三类解决方法:第一类是采用语音增强算法等;第二类方法是寻找稳健的语音特征;第三类方法是基于模型参数适应化的噪声补偿算法。然而,解决噪声问题的根本方法是实现噪声和语音的自动分离,尽管人们很早就有这种愿望,但由于技术的难度,这方面的研究进展很小。近年来,随着声场景分析技术和盲分离技术的研究发展,利用在这些领域的研究成果进行语音和噪声分离的研究取得了一些进展。

语音信号处理是研究用数字信号处理技术对语音信号进行处理的一门学科。语音信号处理的理论和研究包括紧密结合的两个方面:一方面是从语音的产生和感知来对其进行研究,这一研究与语音·语言学、认知科学、心理·生理学等学科密不可分;另一方面是将语音作为一种信号来进行处理,包括传统的数字信号处理技术以及一些新的应用于语音信号的处理方法和技术。

本书将系统介绍语音信号处理的基础、原理、方法和应用。全书共分13章,其中第2章介绍了语音信号处理的基础知识,如发音与听觉器官、语音信号的数学模型、语音信号的

预处理等;第3章介绍了当前语音信号处理应用的三个主流技术,即矢量量化技术、隐马尔可夫模型技术和人工神经网络;第4章介绍了语音信号的分析方法,包括时域分析、频域分析,同态分析、线性预测分析等;第5章介绍了语音信号的特征提取技术,包括端点检测、基音检测和共振峰检测等。从第6章开始介绍了语音信号处理的各种典型应用,包括语音增强、语音编码、语音合成与转换、语音识别、说话人识别、语音信号中的情感信息处理、语音隐藏以及声源定位等。

语音信号处理是目前发展最为迅速的信息科学技术之一,其研究涉及一系列前沿课题,且处于迅速发展之中。因此本书的宗旨是在系统地介绍语音信号处理的基础、原理、方法和应用的同时,向读者介绍该学科领域近年来取得的一些新成果、新方法及新技术。数字语音信号处理属于应用科学,要学好这门课程,关键在于理论必须联系实际应用,才能很好地掌握数字语音处理的理论和技术方法。因此,本书在每一章后面都附有课外思考题,并且在全书的最后附有两个语音处理实用程序。建议学习者仔细选做书中的习题,并进行计算机上机实验以获得实际经验,以尽快掌握所学的知识。

第2章 语音信号处理的基础知识

语音信号处理是研究用数字信号处理技术对语音信号进行处理的一门学科。它的目的有两个：一个是要通过处理得到一些反映语音信号重要特征的语音参数，以便高效地传输或存储语音信号信息；另一个是要通过处理的某种运算以达到某种用途的要求，例如人工合成出语音、辨识出讲话者、识别出讲话的内容等。因此，在研究各种语音信号数字处理技术应用之前，首先需要了解语音信号的一些重要特性，在此基础上才可以建立既实用又便于分析的语音信号产生模型和语音信号感知模型等，它们是贯穿整个语音信号数字处理的基础。

2.1 语音发音及感知系统

2.1.1 语音发音系统

人的发音器官包括：肺、气管、喉（包括声带）、咽、鼻和口。这些器官共同形成一条形状复杂的管道。喉的部分称为声门。从声门到嘴唇的呼气通道叫作声道（Vocal Tract）。声道的形状主要由嘴唇、颚和舌头的位置来决定。由声道形状的不断改变，而发出不同的语音。

语音是从肺部呼出的气流通过在喉头至嘴唇的器官的各种作用而发出的。作用的方式有三种：第一是把从肺部呼出的直气流变为音源，即变为交流的断续流或者乱流；第二是对音源起共振和反共振的作用，使它带有音色；第三是从嘴唇或鼻孔向空间辐射的作用。因此，与发出语言声音有关的各器官叫作发音器官。图2-1示出了发音器官的部位和名称。

产生语音的能量来源于正常呼吸时肺部呼出的稳定气流。气管壁是由一些环状软骨组成的，讲话时它将来自肺部的空气送到喉部。“喉”是由许多软骨组成的，对发音影响最大的是从喉结至杓状的软骨之间的韧带褶，称为声带。呼吸时左右两声带打开，讲话时则合拢起来。而声带之间的部位称为声门。声门的开启和关闭是由两个杓状软骨控制的，它使声门呈 Λ 形状开启或关闭。讲话时声带合拢因而受声门下气流的冲击而张开；但由声带韧性迅速地闭合，随后又张开与闭合，这样不断重复。不断地张开与闭合的结果，使声门向上送出一连串喷流而形成一系列脉冲。声带每开启和闭合一次的时间即声带的振动周期就是音调周期或基音周

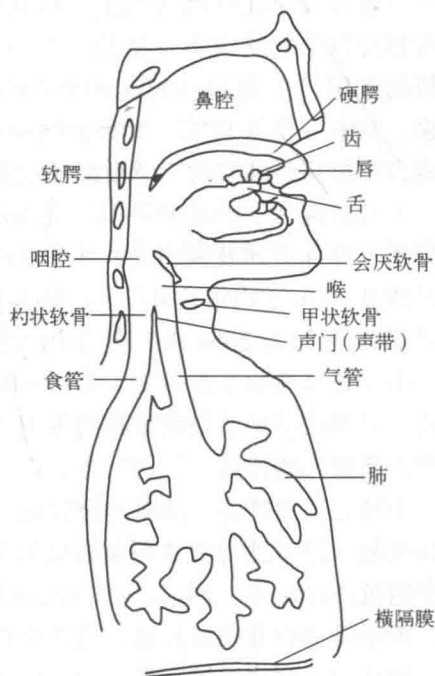


图2-1 发音器官的部位和名称