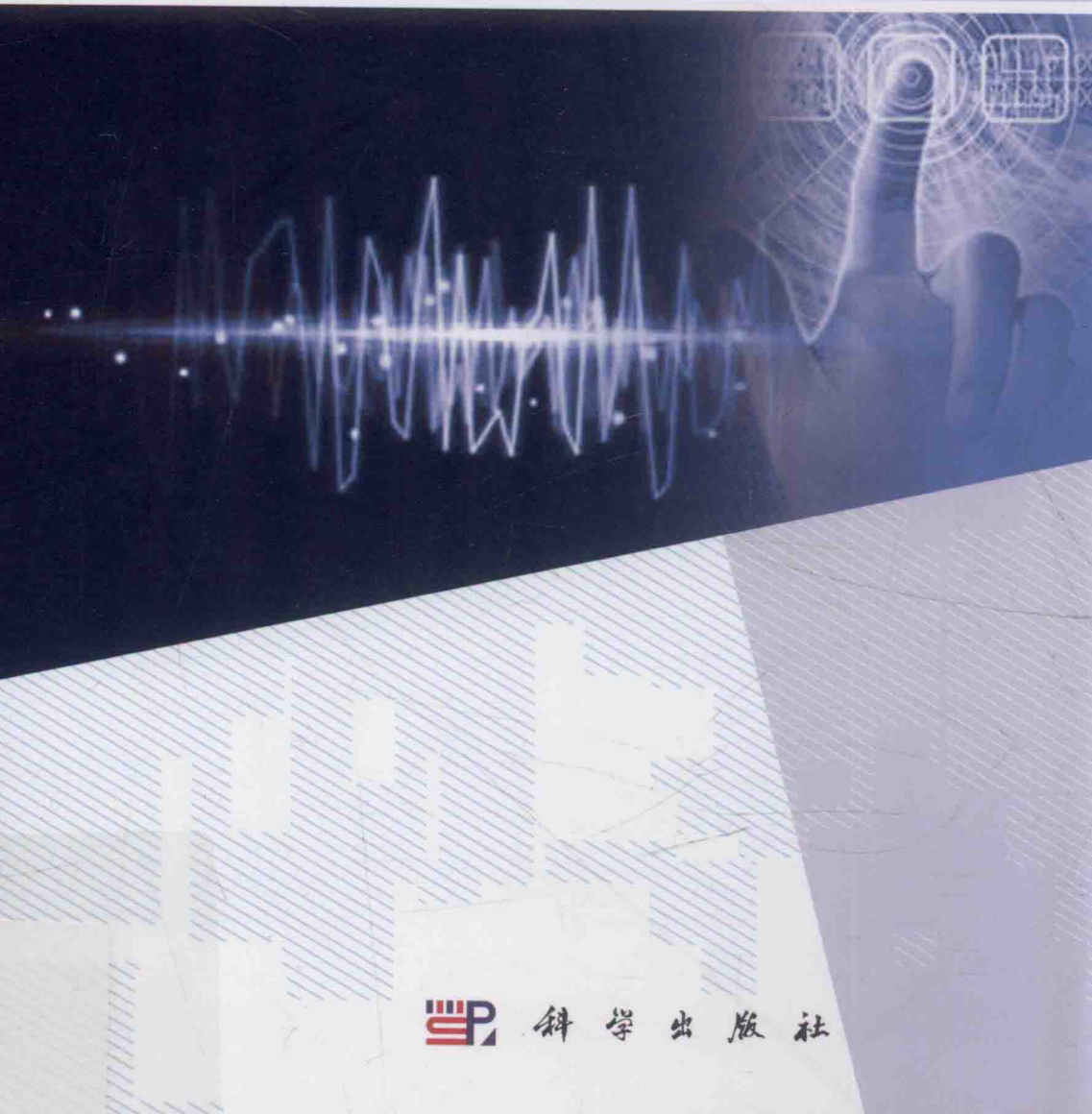




智能科学/技术/著作/丛书

语音处理及人机交互技术

张毅 刘想德 罗元等 编著



科学出版社

智能科学技术著作丛书

语音处理及人机交互技术

张毅 刘想德 罗元等 编著



科学出版社

北京

内 容 简 介

在总结目前国内语音处理和人机交互研究现状的基础上,本书着重介绍作者在语音信号处理、语音特征提取、模式识别及人机交互应用等方面取得的研究成果。内容主要包括语音信号的声学基础、语音信号的分析、语音增强、模式识别算法及应用、语音信号的识别、语音信号的合成、语音信号定位与分离技术的理论与方法等。本书内容全面,重点突出,对原理的阐述清晰易懂,可读性强。

本书可作为智能科学与技术、控制科学与工程、计算机科学与技术、机械电子工程、通信工程、生物信息技术、光电工程等与语音处理及其人机交互有关学科及专业的本科生和研究生的教材或参考书,也可供从事语音信号处理、语音人机交互、智能机器人等方向的技术人员参考。

图书在版编目(CIP)数据

语音处理及人机交互技术/张毅等编著. —北京:科学出版社,2016.3
(智能科学技术著作丛书)

ISBN 978-7-03-030324-0

I. ①语… II. ①张… III. ①自然语言处理-研究 ②人-机系统-研究
IV. ①TP391 ②TB18

中国版本图书馆 CIP 数据核字(2016)第 039549 号

责任编辑:张艳芬 纪四穗 / 责任校对:郭瑞芝
责任印制:张 倩 / 封面设计:左讯科技

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2016年3月第 一 版 开本:720×1000 1/16

2016年3月第一次印刷 印张:13 1/2

字数:252 000

定价:88.00 元

(如有印装质量问题,我社负责调换)

《智能科学技术著作丛书》编委会

名誉主编：吴文俊

主 编：涂序彦

副主编：钟义信 史忠植 何华灿 何新贵 李德毅 蔡自兴 孙增圻
谭 民 韩力群 黄河燕

秘书长：黄河燕

编 委：(按姓氏汉语拼音排序)

蔡庆生(中国科学技术大学)

杜军平(北京邮电大学)

何华灿(西北工业大学)

何新贵(北京大学)

黄心汉(华中科技大学)

李德毅(中国人民解放军总参谋部第六十一研究所)

李祖枢(重庆大学)

刘 清(南昌大学)

邱玉辉(西南师范大学)

史忠植(中国科学院计算技术研究所)

谭 民(中国科学院自动化研究所)

涂序彦(北京科技大学)

王家钦(清华大学)

吴文俊(中国科学院数学与系统科学研究院)

杨义先(北京邮电大学)

张琴珠(华东师范大学)

钟义信(北京邮电大学)

蔡自兴(中南大学)

韩力群(北京工商大学)

何 清(中国科学院计算技术研究所)

黄河燕(北京理工大学)

焦李成(西安电子科技大学)

刘 宏(北京大学)

秦世引(北京航空航天大学)

阮秋琦(北京交通大学)

孙增圻(清华大学)

谭铁牛(中国科学院自动化研究所)

王国胤(重庆邮电学院)

王万森(首都师范大学)

于洪珍(中国矿业大学)

赵沁平(北京航空航天大学)

庄越挺(浙江大学)

《智能科学技术著作丛书》序

“智能”是“信息”的精彩结晶，“智能科学技术”是“信息科学技术”的辉煌篇章，“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”(intelligence science & technology, IST)是关于“广义智能”的理论方法和应用技术的综合性科学技术领域,其研究对象包括:

- “自然智能”(natural intelligence, NI),包括“人的智能”(human intelligence, HI)及其他“生物智能”(biological intelligence, BI)。
- “人工智能”(artificial intelligence, AI),包括“机器智能”(machine intelligence, MI)与“智能机器”(intelligent machine, IM)。
- “集成智能”(integrated intelligence, II),即“人的智能”与“机器智能”人机互补的集成智能。
- “协同智能”(cooperative intelligence, CI),指“个体智能”相互协调共生的群体协同智能。
- “分布智能”(distributed intelligence, DI),如广域信息网、分散大系统的分布式智能。

“人工智能”学科自1956年诞生的,五十余年来,在起伏、曲折的科学征途上不断前进、发展,从狭义人工智能走向广义人工智能,从个体人工智能到群体人工智能,从集中式人工智能到分布式人工智能,在理论方法研究和应用技术开发方面都取得了重大进展。如果说当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合,那么可以认为,现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融。

1981年,“中国人工智能学会”(Chinese Association for Artificial Intelligence, CAAD)正式成立,25年来,从艰苦创业到成长壮大,从学习跟踪到自主研发,团结我国广大学者,在“人工智能”的研究开发及应用方面取得了显著的进展,促进了“智能科学技术”的发展。在华夏文化与东方哲学影响下,我国智能科学技术的研究、开发及应用,在学术思想与科学方法上,具有综合性、整体性、协调性的特色,在理论方法研究与应用技术开发方面,取得了具有创新性、开拓性的成果。“智能化”已成为当前新技术、新产品的发展方向和显著标志。

为了适时总结、交流、宣传我国学者在“智能科学技术”领域的研究开发及应用成果,中国人工智能学会与科学出版社合作编辑出版《智能科学技术著作丛书》。需要强调的是,这套丛书将优先出版那些有助于将科学技术转化为生产力以及对社会和国民经济建设有重大作用和应用前景的著作。

我们相信,有广大智能科学技术工作者的积极参与和大力支持,以及编委们的共同努力,《智能科学技术著作丛书》将为繁荣我国智能科学技术事业、增强自

主创新能力、建设创新型国家做出应有的贡献。

祝《智能科学技术著作丛书》出版,特赋贺诗一首:

**智能科技领域广
人机集成智能强
群体智能协同好
智能创新更辉煌**

涂序彦

中国人工智能学会荣誉理事长

2005年12月18日

前 言

语音识别技术以人类语音为研究对象,通过计算机处理将语音信号转换为命令或文字,结合语音合成实现语音人机交互的功能。语音识别在互联网、消费电子、国防、汽车工业等领域用于人与机器设备的交互,这种语音人机交互可以带来很大的便利性。作为一门交叉学科,语音识别涉及领域广泛,如语言学、计算机科学、生理学以及信号处理等。在相关领域科学家孜孜不倦的研究下,语音识别的研究和应用正不断深入和发展,已成为一门综合人体智能研究、数字信号处理、模式识别研究等在内的新兴学科。

随着语音识别技术的日益成熟,基于语音识别的人机交互技术已成为现实。目前,全球语音技术市场规模超过 30 亿美元,近年来年增长率保持在 25% 以上。未来语音识别市场被看好,其中电信行业(VoIP 等)、移动应用领域(手机、学习机、平板电脑、车载系统等移动设备)都会呈现出爆发式增长。各大搜索引擎公司也都把语音搜索作为未来重要的应用方向,投入了大量人力、物力进行研究和应用推广。

本书立足语音识别的基本理论,在其他研究人员研究成果的基础上,着重介绍作者在语音识别领域所取得的成果和理论突破。首先介绍语音处理技术的研究现状,然后详细阐述语音识别的主要步骤,包括语音信号的预处理、特征提取、模式识别等内容;接下来介绍作者在语音识别领域取得的主要研究成果,包括语音信号定位与分离的基本理论与方法、多声源环境下基于人耳听觉中枢的语音识别等;最后在语音识别理论的基础上,介绍人机交互技术的实现及作者已完成的一些应用。

本书由重庆邮电大学张毅教授组织撰写。具体分工如下:第 1、2、9、10 章由张毅撰写,第 3、4、6 章由罗元、张毅撰写,第 5、7、8 章内容由刘想德、徐晓东、张毅撰写,全书由张毅统稿。参加本书资料搜集整理、文字校对、实验仿真系统软硬件开发工作的人员还包括汪培培、谢延义、孙龙、吴承军、黎小松、许方煜、王可佳、倪雷、姚钦、颜博等。在此对他们付出的辛勤劳动表示由衷感谢。

此外,本书的出版得到了重庆市科学技术委员会项目(编号:cstc2015jcyjB0241)的资助。特别感谢参考文献中所列各位作者,包括未能在参考文献中一一列出的作者,正是他们在各自领域的独到见解和特别贡献为作者提供了宝贵的参考资料、研究视角和丰富的写作源泉,使作者能够在系统总结此领域科研工作成果的基础上,汲取各家之长,及时形成本书。

语音处理领域的内涵丰富,技术发展迅速,应用需求多样。作者从事智能机器人、人工智能、多模人机交互研究已有 10 多年,特别是在语音识别领域积累了大量经验。本书的撰写是作者在总结以往研究成果基础上的一次努力尝试,限于作者水平和学识,书中难免存在疏漏和不足之处,诚望读者不吝赐教,以利修正,让更多读者获益。

作 者

2015 年 10 月

目 录

《智能科学技术著作丛书》序

前言

第 1 章 绪论	1
1.1 语音信号技术的历史	1
1.2 语音信号技术应用及新方向	3
1.2.1 语音信号的识别	3
1.2.2 语音信号的合成	4
1.2.3 语音信号的编码	4
参考文献	5
第 2 章 语音信号的声学基础	7
2.1 语音信号的产生	7
2.1.1 语音发音器官	7
2.1.2 语音声学特征	10
2.1.3 语音信号的时域及频域表示	11
2.1.4 汉语语音分类	13
2.1.5 汉语语音韵律特性	16
2.2 语音信号的感知	16
2.2.1 人的听觉系统	17
2.2.2 人耳的听觉特性	19
2.2.3 人耳听觉的掩蔽效应	20
2.3 语音信号生成的数学模型	26
2.3.1 激励模型	26
2.3.2 声道模型	28
2.3.3 辐射模型	30
2.3.4 语音信号的数学模型	31
2.4 语音产生的非线性模型	33
2.4.1 调频-调幅模型的基本原理	33
2.4.2 Teager 能量算子	34
2.4.3 能量分离算法	34
2.4.4 调频-调幅模型的应用	36

参考文献	36
第3章 语音信号的分析	38
3.1 语音信号预处理步骤	38
3.1.1 采样与量化	38
3.1.2 预加重	39
3.1.3 分帧与加窗	40
3.1.4 端点检测	42
3.2 语音信号的时域分析	43
3.2.1 短时能量分析	44
3.2.2 短时平均过零率	44
3.2.3 短时自相关函数和短时平均幅度差函数	45
3.3 语音信号的频域分析	47
3.3.1 滤波器组方法	47
3.3.2 傅里叶频谱分析	48
3.3.3 线性预测倒谱系数	49
3.3.4 Mel 频率倒谱系数	51
3.4 语音信号的倒谱分析	53
3.4.1 同态信号处理的基本原理	54
3.4.2 复倒谱和倒谱	56
3.5 语音信号的小波分析	57
参考文献	58
第4章 语音增强	60
4.1 语音信号与语音增强	60
4.1.1 语音特性	60
4.1.2 人耳感知特性	61
4.1.3 噪声特性	61
4.1.4 语音增强的信号模型	63
4.2 谱减法语音增强	64
4.2.1 基本原理	65
4.2.2 基本谱减法的改进	66
4.3 维纳滤波法语音增强	67
4.4 卡尔曼滤波法语音增强	68
4.5 最小均方误差法的语音增强	71
4.6 语音增强的一些新发展	73
参考文献	75

第 5 章 模式识别算法及应用	77
5.1 矢量量化的基本原理	77
5.2 矢量量化的失真测度	78
5.2.1 欧氏距离测度	79
5.2.2 线性预测失真测度	80
5.2.3 识别失真测度	81
5.3 神经网络的基本原理	81
5.4 神经网络的基本构成	83
5.4.1 人工神经元模型	83
5.4.2 神经网络有向图表示	85
5.4.3 神经网络结构及工作方式	86
5.5 基本神经网络模型及算法	87
5.5.1 单层感知器	87
5.5.2 多层感知器	90
5.5.3 径向基函数神经网络模型	92
5.5.4 循环神经网络	93
5.5.5 支持向量机	95
参考文献	98
第 6 章 语音信号的识别	99
6.1 语音识别概述	99
6.2 孤立词语音识别	100
6.3 连续词语音识别	102
6.4 说话人识别	103
6.4.1 说话人识别概述	103
6.4.2 说话人识别特征选取	103
6.4.3 说话人识别系统的结构	107
6.4.4 说话人识别中的识别方法	108
6.4.5 说话人识别中需进一步研究的问题	109
6.5 语音信号预处理	110
6.6 语音信号识别特征提取	111
6.7 语音信号识别方法	113
6.8 隐马尔可夫模型基本原理及在语音识别中的应用	114
6.8.1 马尔可夫链	114
6.8.2 隐马尔可夫模型的定义	115
6.8.3 隐马尔可夫模型的三个问题	115

6.8.4 隐马尔可夫模型在语音识别中的应用	118
6.9 动态时间规整算法	120
参考文献	122
第7章 语音信号的合成	123
7.1 概述	123
7.2 语音合成的基本方法	124
7.2.1 以单词为基础的合成方法	124
7.2.2 以音节为基础的合成方法	125
7.2.3 以音素为基础的合成方法	126
7.3 共振峰合成法	127
7.4 线性预测合成法	129
7.5 语音合成专用硬件介绍	132
7.6 PSOLA 算法合成语音	134
7.7 文语转换系统	136
参考文献	138
第8章 语音信号定位与分离技术	140
8.1 声源定位技术	140
8.1.1 基于麦克风阵列的定位方法	140
8.1.2 基于人耳听觉机理的声源定位	143
8.2 语音分离技术	149
8.2.1 计算听觉场景分析	149
8.2.2 计算听觉场景分析的语音分离模型	154
8.2.3 独立分量分析	155
8.3 声源定位与语音分离实现方法	167
参考文献	169
第9章 基于听觉机制的语音识别	171
9.1 人耳的听觉机制	171
9.1.1 人耳听觉中枢的物理结构	171
9.1.2 人耳听觉中枢的听觉机理	172
9.2 人耳听觉系统的基本模型	172
9.2.1 人耳听觉系统的外周听觉模型	173
9.2.2 人耳听觉系统的内毛细胞模型	177
9.2.3 人耳听觉系统的重合神经元模型	181
9.3 多声源环境下人耳语音定位模型	183
9.3.1 双耳线索	183

9.3.2 耳廓效应	184
9.3.3 双耳时间差模型	185
9.3.4 多声源环境下声源定位系统的设计	185
9.4 多声源环境下人耳语音分离模型	188
9.4.1 语音信号分离的反演模型	188
9.4.2 语音信号分离系统的设计	189
9.5 基于听觉机制的语音识别系统	190
参考文献	191
第 10 章 基于语音的人机交互及应用	193
10.1 人机交互的研究及发展	193
10.1.1 人机交互概述	193
10.1.2 人机交互的研究内容	194
10.1.3 人机交互的发展历程	195
10.2 语音人机交互及其应用	196
10.2.1 语音人机交互概述	196
10.2.2 语音人机交互与智能轮椅	196
10.2.3 语音人机交互与智能电视	197
10.2.4 语音人机交互的其他应用	198
参考文献	199

第 1 章 绪 论

随着计算机和人工智能技术的迅速发展,语音人机交互已经成为人与机器之间最好的沟通方式之一。要使机器“听懂”人类的语言,与人类进行有效的语音信息交互,这就关系到语音信号处理技术。

1.1 语音信号技术的历史

早在一两千年以前,人们便对语音信号进行了研究。由于没有适当的仪器设备,长期以来,一直是通过由耳倾听和用口模仿来进行研究的。所以,这种语言研究常被称为“口耳之学”,而对语音只是停留在定性的描写上。

语音信号处理真正意义上的研究可以追溯到 1876 年贝尔电话的发明,该技术首次用声电、电声转换技术实现了远距离的语音传输。1939 年, Homer Dudley 提出并研制成功第一个声码器,从此奠定了语音产生模型的基础。这一发明在语音信号处理领域具有划时代的意义^[1-3]。19 世纪 60 年代,亥姆霍兹应用声学方法对元音和歌唱进行了研究,从而奠定了语音的声学基础。20 世纪 40 年代,贝尔实验室发明了一种语言声学的专用仪器——语谱图仪。它可以把语音的时变频谱用语图表示出来,从而得到“可见语言”。1948 年,美国 Haskins 实验室研制成功“语音回放机”,该仪器可以把手工绘制在薄膜片上的语谱图自动转换成语音并进行语音合成。20 世纪 50 年代,科研人员对语言产生的声学理论开始有了系统论述。语音识别技术的根本目的是研究出一种具有听觉功能的机器,使机器能接收人类的语音,理解人的意图。1952 年,贝尔实验室的 Davis 等研制了特定说话人孤立数字识别系统,该系统利用每个数字元音部分的频谱特征进行识别。1956 年, RCA 实验室独立研制出 10 个单音节词的识别系统^[4-7]。1959 年, Fry 和 Denes 等尝试建立音素识别器来识别 4 个元音和 9 个辅音,并采用模式匹配和频谱分析来进行识别决策。与此同时,麻省理工学院林肯实验室的 Forgie 等研究了 10 个元音的识别并采用了声道的时变估计技术。

20 世纪 60 年代初,日本很多研究者开发了相关的特殊硬件来进行语音识别,例如,东京无线电研究实验室的 Suzuki 等研制的通过硬件来进行元音识别的系统。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^[8-10]。60 年代中期形成的一系列数字信号方法和技术,如数字滤波器、快速傅里叶变换等成为语音信号数字处理的理

论和技术基础。在方法上,随着计算机的发展,以往的以硬件为中心的研究逐渐转化为以软件为主的处理研究。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,但后来已全部停下来,这说明当时人们对语音识别难度的认识得到了加深。

1970年,单词识别装置进入了实用化阶段,其后实用化的进程进一步高涨,实用机的生产销售业步入正轨。到1971年,以美国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。这个研究计划不仅只在美国国内,而且对世界各国都产生了很大影响,它促进语音识别研究的兴起。20世纪70年代中期,线性预测技术被用于语音信号处理,此后隐马尔可夫模型(HMM)法也获得了初步成功,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等首次解决了矢量量化(VQ)码书生成的问题,并首先将矢量量化技术用于语音编码并获得成功。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领域^[11-14]。

20世纪80年代,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,语音信号处理技术产生了突破性进展。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得到了广泛应用。

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以及隐马尔可夫模型和人工神经网络的结合成为研究的热点^[15]。另一方面,为了语音识别实用化的需要,讲者自适应模型、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备受关注。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应用。

语音信号处理学科之所以能够长期地、深深地吸引广大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联系,并且一起发展。

随着计算机的出现,语音信号处理的研究得到了飞速发展,过去受人力、时间限制的大量的语音统计分析工作,得以在电子计算机上进行。在此基础上,语音信号处理无论在基础研究方面,还是在技术应用方面,都取得了突破性进展。

语音是人类交换信息最方便、最快捷的一种方式,在高度发达的信息社会中,数字化的方法进行语音的传送、存储、识别、合成和增强等是整个数字化通信中最重要、最基本的组成部分之一^[16]。数字电话、高音质的窄带语音通信系统、语音学习机、声控打字机、自动翻译机、智能机器人、新一代计算机语音智能终端及许多

军事上的应用等,都要用到语音信号处理技术,随着集成电路和微电子技术的飞速发展,语音信号处理系统逐步走向实用化。语音信号处理是一门新兴的边缘学科,它是语音学和数字信号处理两个学科相结合的产物^[17]。它和认知学、心理学、语言学、计算机科学、模式识别和人工智能等学科有着紧密的联系。语音信号处理的发展依赖于这些学科的发展,而语音信号处理技术的进步也会促进这些领域的进步。语音信号处理的目的是要得到某些语音特征参数以便高效地传输或储存;或者是通过某种处理运算以达到某种用途的要求,如人工合成语音、辨识出讲话者、识别出讲话的内容等。随着现代科学和计算机技术的发展,除了人与人之间的自然语言的通信方式,人机对话及智能机器等领域也开始使用语言。这些人工语言同样有词汇、语法、语法结构和语义内容等。控制论创始人维纳在1950年就曾指出过:“通常,我们把语言仅仅看作人与人之间的通信手段,但是,要使人向机器、机器向人以及机器向机器讲话,那也是完全办得到的”。

1.2 语音信号技术应用及新方向

语音信号处理技术在语音识别、语音合成、语音编码方面有十分广泛的应用,同时随着市场需求的增加,该技术正在不断地改进和完善。

1.2.1 语音信号的识别

语音识别以语音为研究对象,它是语音信号处理的一个重要研究方向,是模式识别的一个分支,涉及生理学、心理学、语言学、计算机科学以及信号处理等诸多领域,甚至还涉及人的体态语言(如人在说话时的表情、手势等行为动作可帮助对方理解),其最终目标是实现人与机器进行自然语言通信。

传统方向上语音识别的基本任务是将语音转变成文本或命令。在声控应用上,计算机识别语音内容,并实施相应的动作,其中典型的系统有声控电话转换、声控语音拨号、声控智能玩具、信息网络查询、银行服务、家庭服务等。对于听写系统,是以口授方式将文字输入计算机的。而自动口语翻译是将一种语言翻译成另一种语言,如中国科学院开展的CSTAR计划。语音识别还可应用于人机交互、移动计算设备语音输入、说话人识别、安全加密、法庭取证、银行信息电话查询、公安机关破案等。

目前我国语音识别技术已取得了令人瞩目的成就,研究工作一直紧跟国际水平,国家也很重视,并把大词汇量语音识别的研究列入863计划,由中国科学院声学研究所、中国科学院自动化研究所及北京大学等单位研究开发。其基础研究涉及汉语语音学、听觉模型、人工神经网络、小波变换、分形维数和支持向量机等理论,其研究成果必将推动我国语音识别技术研究迈上新台阶^[18]。鉴于中国未来庞

大的市场,国外也非常重视汉语语音识别的研究。美国、新加坡等地聚集了一批来自中国的学者,研究成果已达到相当高的水平。因此,国内除了要加强理论研究,更要加快从实验室演示系统到商品的转化。

1.2.2 语音信号的合成

对于语音合成技术,最早的合成器是1835年由Kempelen发明,经Weston改进的机械式会讲话的机器。该机器完全模仿人的发音生理过程,分别用风箱、特别设计的哨和软管来模拟肺部的空气运动、模拟口腔。而最早的电子式语音合成器是1939年Homer Dudley发明的声码器,它不是简单地模拟人的生理过程,而是通过电子线路来实现基于语音产生的源——滤波器理论。但是真正具有实用意义的近代语音合成技术是随着计算机技术和数字信号处理技术的发展而发展起来的,主要是采用计算机产生高清晰度、高自然度的连续语音^[19]。在语音合成技术的发展中,早期的研究主要是采用参数合成方法。值得提及的是,1973年Holmes发明的并联共振峰合成器和1980年Klatt发明的串/并联共振峰合成器,只要精确调整参数,都能合成出比较自然的语音。最具代表性的文语转换(TTS)系统是美国DEC公司于1987年开发的DECtalk。由于准确提取共振峰参数比较困难,虽然利用共振峰合成器可以得到许多逼真的合成语音,但是整体合成语音的音质难以达到文语转换系统的实用要求。

语音合成已经应用于公共交通上的自动报站、各种场合的自动报时、自动告警、文本校对中的语音提示、电话查询服务等诸多方面。目前研究者也致力于新领域的开发,将语音合成与Internet结合使有声Email、网上信息的有声获取、语音聊天等相继出现在人们的生活中;将其与机器翻译技术结合从而实现语音翻译;将其与图像、视频技术结合产生视觉语音(visual speech)。语音合成的应用更加趋向高质量、高清晰度自然语音合成,以及表达人类情感的语音^[20]。

1.2.3 语音信号的编码

语音编码的目的就是在保证一定语音质量的前提下,尽可能降低编码的比特率,以节省频率资源。语音编码技术主要有两个潜在的发展方向:一是中低速率的语音编码的实用化以及如何实用化过程中进一步提高其抗干扰、抗噪声能力;另一个是如何进一步降低其编码速率。在数字通信、移动通信、卫星通信、IP电话通信、军事保密语音通信、呼叫服务(数字录音电话、语音信箱等)都需要用到语音编码。语音编码正向着低码率、高质量的音频编码方向发展,未来的综合业务数字网(ISDN)、微波接力通信和信息高速公路等正朝着这个新的方向前行。

另外,对于语音信号的处理目前正在拓展更多新的应用,如下所述。