智能信息处理丛书

VoIP语音处理与识别

VoIP Speech Processing and Recognition

屈丹　王波　李弼程　等编著

国防工业出版社
National Defense Industry Press

《智能信息处理》丛书

VoIP 语音处理与识别

VoIP Speech Processing and Recognition

屈 丹 王 波 李弼程
张连海 陈 琦 张文林 编著

国防工业出版社

·北京·

图书在版编目(CIP)数据

VoIP 语音处理与识别 / 屈丹等编著. —北京:国防工业出版社,2010.5
(智能信息处理丛书)
ISBN 978-7-118-06668-5

Ⅰ. ①V... Ⅱ. ①屈... Ⅲ. ①互联网络-语音数据处理 Ⅳ. ①TN912.3

中国版本图书馆 CIP 数据核字(2010)第 034604 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 100048)
北京四季青印刷厂印刷
新华书店经售
*
开本 880×1230 1/32 印张 16⅛ 字数 460 千字
2010 年 5 月第 1 版第 1 次印刷 印数 1—4000 册 定价 48.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)68428422 发行邮购:(010)68414474
发行传真:(010)68411535 发行业务:(010)68472764

序　言

随着通信和信息技术的发展，特别是网络技术的发展，海量文本、语音、图像和视频等媒体为人们提供了丰富的信息资源。对广大用户来说，主要是从海量信息环境中获取有用的信息。对于管理人员来说，除了信息获取，还需要对海量信息的内容进行监管。

无论是为了获取有价值的信息，还是对海量信息的内容进行监管，在广泛收集信息的同时，需要对获得的信息进行有效的采集、高效的内容识别、深层的检索与挖掘。在海量信息智能处理中，信息采集是基础、信息内容识别是核心、信息检索与挖掘是手段、信息获取与监管是目的。

信息工程大学信息工程学院“智能信息处理”方向长期从事文本分析与理解、语音处理与识别、图像/视频处理与识别、多源信息融合、信息检索与挖掘等前沿学科领域的教学与科研，获得了国家社科基金（重大）、国家自然科学基金、国家“863”、国防预研等多个项目的资助。“智能信息处理”方向的多名作者在总结和提升多年教学、科研成果的基础上，编写了这套《智能信息处理》丛书。国防工业出版社为该丛书的出版给予了大力支持。

《智能信息处理》丛书共8个分册，比较系统、全面地介绍了智能信息处理技术及其应用，重点阐述了文本、语音、图像及视频等媒体的内容识别、检索、挖掘和监管。该丛书可作为计算机科学与技术、电子工程、信息与通信工程、自动控制、指挥自动化、情报学、图书馆学、信息管理等相关专业方向的高年级本科生和研究生相关课程的教材和相关领域的科研、工程技术人员的参考书。

希望《智能信息处理》丛书的出版能为信息处理技术的应用与发

展添砖加瓦，做出应有的贡献。同时，由于编写组学识有限，加上时间紧张，错误缺点在所难免，敬请批评、指正。

《智能信息处理》丛书编写组

2010 年 1 月 28 日于郑州

前　言

近几年，Internet 在各个领域的应用迅速发展，使得各行各业都在注视着电信业务的 IP 网络环境这个巨大市场。目前世界各国都在竞相投入大量的人力、财力进行相关理论研究，从而使 VoIP 语音识别技术研究和应用成为一个非常活跃的领域。因此，系统、全面地掌握 VoIP 语音处理与识别领域的原理和方法，了解该领域的最新技术、最新研究成果与动态，对于语音识别领域工作的开展具有重要意义。

作者多年来一直从事语音识别领域的研究和开发工作，深深体验到语音识别技术的巨大发展潜力和广阔的应用前景。作者承担了国家 863 项目——海量语音识别综合处理系统（项目编号：2006AA01Z146），还承担了省部级项目多项，取得了多项研究成果，获省部级科技进步二等奖 3 项，这些都为本书提供了强有力的支持。本书是作者在之前编写的《实用语音识别基础》一书和“VoIP 语音识别”讲义的基础之上，结合多年的教学经验和实验室多年相关科研成果，并参考相关文献资料编写而成。

本书紧跟语音识别的发展前沿，从 PSTN 网信号处理、IP 包信号处理和后端模型算法与处理识别 3 个方面论述了 VoIP 识别技术及其应用。本书可作为电子工程、自动控制、信号与信息处理、指挥自动化等相关专业方向的高年级本科生和研究生“VoIP 语音处理和识别”课程的教材和相关领域的科研、工程技术人员的参考书。

全书共分为 14 章。第 1 章是绪论，介绍 VoIP 语音技术的基本概念与原理，传统语音识别及 VoIP 语音识别的发展历程；第 2 章介绍了 VoIP 语音通信中常用的语音编码算法；第 3 章讨论了通信网语音信号和话带数据检测技术；第 4 章讨论如何对信号是否经过 VoIP 信道进行

检测；第5章介绍语音信号的特征提取技术；第6章介绍编解码失配补偿技术，研究编解码方式失配时对语音识别系统的影响及其补偿算法；第7章介绍通话模式分析技术，对通信网获取的语音信号进行通话内容的分析，判别其音频种类、去除对语音识别产生不利影响的非语音信号的技术；第8章介绍VoIP中常用的协议标准和信息获取技术；第9章讨论了丢包处理技术，研究网络丢包对语音识别性能的影响，并进行丢包检测和补偿；第10章讨论了基于压缩码流的语音特征提取技术，包括半解码特征提取和基于帧结构映射的码流特征直接提取技术；第11章介绍特征选择和特征变换的方法；第12章介绍常用识别模型算法，包括隐马尔可夫模型、高斯混合模型、人工神经网络和动态贝叶斯网络；第13章讨论了当前语音识别领域常用的稳健性处理技术；第14章介绍语音识别的应用，包括说话人识别、语言辨识、关键词识别、连续语音识别和情感识别。

这些内容不仅有利于读者将理论与实际相结合，加深对理论方法的理解，让读者较系统地掌握VoIP语音识别的理论精髓和相关技术，同时书中给出的应用实例，为科研人员应用VoIP语音识别技术解决相关领域的实际问题，提供了具体思路和方法。

本书的特点是介绍了目前最前沿的VoIP语音处理与识别理论和技术，反映了语音识别技术的最新技术与发展趋势，具有与本学科学术水平相适应的先进性；由浅入深地安排章节，先理论后应用，知识结构合理，章节之间紧密配合、前后呼应，具有很强的科学性和系统性；以自己的研究和实验成果为主，提供了大量的实际参数、图表，与实际工作联系紧密，具有很强的可操作性与实用性。通过本书，读者可以了解当前最前沿的VoIP语音识别技术，可以掌握语音识别的基本知识和系统理论，可以获得应用实用语音识别技术的基本技能。另外，通过作者在书中描述的启发式研究过程，读者既可以提高自学能力，又可以在创新思维方面得到很大的提高。

本书由屈丹、王波、李弼程、张连海、陈琦和张文林编写。第1、5、12、14章由屈丹编写，第2章由陈琦编写，第4章由张连海编写，第7、8章由王波编写，第3章由王波、张文林共同编写，第6、9、10章由屈丹、李弼程共同编写，第11章由屈丹、陈琦共同编写，第13章由屈丹、张连

海共同编写。全书由屈丹、李弼程负责审校和统稿。

本书的编写得到了信息工程大学各级领导的关心和支持，得到了国内外同行学者的支持和帮助，和著名专家进行过有益的交流、研讨，这些使作者受益匪浅。他们的卓越见解提升了本书的理论价值和可用性，在此向他们表示深切感谢。

感谢信息工程大学信息工程学院王炳锡教授多年的教诲和指导。王教授高屋建瓴，引领语音信号处理方向经过20年的发展进入了全新发展期。王教授为我们规划了未来的发展方向，本书也正是在我们沿着他的脚步前进的一些成果的总结。

感谢信息工程大学信息工程学院语音信号处理方向毕业或在读的博士生和硕士生，他们是侯风雷、彭煊、徐望、牛铜、唐晖、高新建、戴冠男、张强、薛晓燕和张宝奇等同学。他们结合自己的研究课题展开的实用性的研究为本书打下了坚实的理论基础，他们卓有成效的工作，为语音信号处理方向的发展提供了强有力的支持，更使本书言之有理、言之有物。

最后，本书参考了国内外很多同行的论文、著作，引用了其中的相关结论和数据，这些都是原作者的辛勤工作，是他们把语音识别推向前进，向他们表示衷心感谢。

最后要特别感谢国家科技部对“海量语音识别综合处理系统”项目（批准号：2006AA01Z146）和国家自然科学基金委员会对“电话信道自然语音的语言辨识技术研究”项目（批准号：60372038）的支持。

由于作者学识有限，书中不足之处在所难免，敬请读者批评、指正。

作 者

2009年9月于郑州

目　录

第 1 章　绪论 …… 1

1.1　VoIP 的基本概念和系统组成 …… 1

1.1.1　VoIP 基本概念 …… 1

1.1.2　VoIP 基本原理 …… 3

1.1.3　VoIP 系统基本组成 …… 5

1.1.4　VoIP 主要特点 …… 7

1.1.5　VoIP 的关键技术 …… 8

1.2　语音识别的基本原理与研究内容 …… 10

1.2.1　语音识别基本原理 …… 11

1.2.2　传统语音识别 …… 14

1.2.3　VoIP 语音识别 …… 20

1.3　传统语音识别的发展历程 …… 23

1.4　VoIP 语音识别的发展历程 …… 33

参考文献 …… 36

第 2 章　VoIP 语音编码标准 …… 43

2.1　G.711 语音编码 …… 43

2.1.1　G.711 语音编码原理 …… 43

2.1.2　*A* 律压缩 …… 44

2.1.3　μ 律压缩 …… 48

2.2　G.729 语音编码 …… 51

2.2.1　G.729 编码原理 …… 51

2.2.2　编码器实现技术 …… 53

2.2.3　解码器功能说明 …… 73

2.3　G.723.1 语音编码 …… 80

2.3.1 编码器原理 …… 81
2.3.2 编码器实现技术 …… 81
2.3.3 解码器原理 …… 95
2.3.4 解码器实现技术 …… 95
参考文献 …… 99
第3章 语音和话带数据检测 …… 100
3.1 话带数据简介 …… 100
3.1.1 传真 …… 100
3.1.2 调制解调器数据 …… 101
3.1.3 单音信号和双音多频信号(DTMF) …… 103
3.2 语音和话带数据波形的特点 …… 104
3.2.1 语音波形的特点 …… 104
3.2.2 话带数据信号波形的特点 …… 105
3.3 语音和话带数据识别的特征分析 …… 107
3.3.1 能量相关参数 …… 107
3.3.2 过零率相关参数 …… 109
3.3.3 归一化自相关函数 …… 111
3.3.4 基音周期 …… 111
3.3.5 谱特征 …… 112
3.4 语音和话带数据分离方法 …… 112
3.5 语音与话带数据检测技术应用 …… 115
参考文献 …… 117
附录 …… 118
第4章 VoIP信道检测 …… 121
4.1 声道参数 …… 121
4.2 语音信号统计参数 …… 123
4.2.1 偏度与峰度 …… 123
4.2.2 LPC峰度及偏度 …… 124
4.2.3 LPCC峰度及偏度 …… 125
4.3 静态信噪比 …… 127
4.4 非自然周期性参数 …… 130

4.4.1 非自然嘟嘟声 …… 130
4.4.2 机器性参数 …… 131
4.4.3 帧重复性 …… 133
4.5 哑声参数与中断参数 …… 133
4.5.1 哑声参数 …… 133
4.5.2 中断参数 …… 134
4.6 基于支持矢量机的 VoIP 信道检测方法 …… 135
4.6.1 算法流程 …… 135
4.6.2 支持矢量机 …… 135
4.6.3 算法的具体实现 …… 137
参考文献 …… 137
第 5 章 语音信号的特征提取 …… 138
5.1 基音周期 …… 138
5.1.1 基音检测的难点及方法分类 …… 139
5.1.2 自相关法及其改进 …… 139
5.1.3 并行处理法 …… 142
5.1.4 倒谱法 …… 144
5.1.5 简化逆滤波法 …… 145
5.2 线性预测参数 …… 145
5.2.1 线性预测信号模型 …… 146
5.2.2 线性预测误差滤波 …… 147
5.2.3 语音信号的线性预测分析 …… 151
5.2.4 线性预测分析的解法 …… 152
5.2.5 斜格法(Lattice Method)及其改进 …… 153
5.3 线谱对(LSP)参数 …… 160
5.3.1 线谱对分析原理 …… 160
5.3.2 线谱对分析的求解 …… 162
5.4 倒谱系数及差分参数 …… 163
5.4.1 LPCC 参数 …… 163
5.4.2 MFCC 参数 …… 164
5.4.3 ASCC 参数 …… 166

5.4.4 差分参数 …… 167
5.5 感觉加权的线性预测(PLP)特征 …… 168
5.5.1 PLP 参数 …… 168
5.5.2 RASTA-PLP 参数 …… 169
5.6 高阶信号谱类特征 …… 170
5.6.1 WV 谱的定义及其主要性质 …… 170
5.6.2 WV 谱计算式的一些变形 …… 171
参考文献 …… 173
第 6 章 编解码失配补偿 …… 175
6.1 编解码失配影响 …… 175
6.1.1 VoIP 系统语音传输 …… 176
6.1.2 编解码失配对说话人辨认系统的影响 …… 176
6.1.3 编解码失配对说话人确认系统的影响 …… 177
6.1.4 特征参数的编码失真 …… 179
6.2 常用编解码失配补偿方法 …… 180
6.2.1 失配补偿的基本思想 …… 181
6.2.2 经验补偿技术 …… 181
6.2.3 盲补偿 …… 182
6.2.4 基于特征及模型的补偿 …… 184
6.3 基于编码失真的加权 GMM 模型算法 …… 188
6.3.1 加权 GMM 模型 …… 189
6.3.2 权重矩阵 C 的确定 …… 191
6.3.3 实验及分析 …… 191
6.4 编码自动匹配方法 …… 192
6.4.1 编码自动匹配方法的基本思想 …… 192
6.4.2 语音编码检测器 …… 193
6.4.3 实验及分析 …… 193
6.5 统计匹配特征变换失配补偿算法 …… 195
6.5.1 统计匹配的基本思想 …… 195
6.5.2 线性特征变换式 …… 195
6.5.3 非线性特征变换式 …… 196

6.5.4 M-step 迭代根的求解 …… 199
6.5.5 基于统计匹配的编解码失配补偿实验 …… 201
6.6 分数归一化补偿算法 …… 204
6.6.1 分数归一化算法的基本思想 …… 204
6.6.2 实验及分析 …… 206
参考文献 …… 207
第 7 章 通话模式分析 …… 210
7.1 通话模式分析的基本概念与研究内容 …… 210
7.1.1 通话模式分析的定义 …… 211
7.1.2 通话模式分析的关键技术 …… 211
7.1.3 通话模式分析的研究内容 …… 212
7.2 通话模式分析的基本方法 …… 215
7.2.1 基于 KL2 距离的音频分割算法 …… 215
7.2.2 基于隐马尔可夫模型的音频分割算法 …… 216
7.2.3 基于贝叶斯信息准则的音频分割算法 …… 217
7.2.4 基于熵变化趋势检测的音频分割算法 …… 219
7.2.5 基于可信度变化趋势检测的音频分割算法 …… 221
7.3 多人的说话人识别方法 …… 225
7.3.1 多人说话人识别的基本思想 …… 225
7.3.2 说话人分段 …… 226
7.3.3 说话人聚类 …… 227
7.3.4 彩铃的检测与分割算法 …… 227
7.4 电信网特有噪声检测算法 …… 230
参考文献 …… 232
第 8 章 VoIP 协议分析及数据获取 …… 234
8.1 VoIP 协议简介 …… 234
8.2 SIP 协议通信流程及识别 …… 235
8.2.1 SIP 协议的功能 …… 235
8.2.2 SIP 协议的通信方式 …… 236
8.2.3 基于 SIP 协议的 VoIP 信息识别 …… 237
8.3 H.323 协议通信流程与识别 …… 240

8.3.1 H.323 通信流程 …… 240
8.3.2 H.323 协议的动态特征 …… 244
8.3.3 H.323 的识别方法 …… 248
参考文献 …… 251
第 9 章 丢包处理 …… 252
9.1 网络丢包模型 …… 252
9.2 网络丢包对说话人识别的影响 …… 254
9.2.1 合成语音说话人识别实验 …… 254
9.2.2 解码参数说话人识别实验 …… 255
9.2.3 压缩码流说话人识别实验 …… 256
9.3 网络丢包处理技术 …… 258
9.3.1 丢包恢复技术 …… 258
9.3.2 丢包隐藏技术 …… 261
9.4 语音识别系统中的丢包补偿方法 …… 265
9.4.1 丢包检测 …… 265
9.4.2 丢包补偿 …… 266
9.4.3 有效性分析 …… 268
9.4.4 丢包补偿实验结果 …… 268
参考文献 …… 270
第 10 章 码流特征提取 …… 273
10.1 码流语音识别的原理 …… 273
10.2 G.729 码流特征提取 …… 274
10.2.1 编码原理 …… 274
10.2.2 基于解码参数的 G.729 码流特征提取 …… 275
10.2.3 基于帧结构映射的 G.729 码流特征提取 …… 279
10.3 G.723.1 码流特征提取 …… 280
10.3.1 编码原理 …… 280
10.3.2 基于解码参数的 G.723.1 码流特征提取 …… 281
10.3.3 基于帧结构映射的 G.723.1 码流特征提取 …… 282
10.4 GSM 码流特征提取 …… 283
10.4.1 编码原理 …… 283

10.4.2 基于解码参数的 GSM 码流特征提取 …… 284
10.4.3 基于帧结构映射的 GSM 码流特征提取 …… 286
10.5 码流特征提取实验 …… 287
10.5.1 解码参数实验 …… 287
10.5.2 基于帧结构映射参数实验 …… 288
参考文献 …… 289
第 11 章 特征选择与特征变换 …… 291
11.1 特征选择的基本概念 …… 291
11.1.1 特征矢量和特征空间 …… 292
11.1.2 特征的形成 …… 292
11.1.3 特征的特点 …… 292
11.1.4 特征的选择及作用 …… 293
11.2 类的可分性判据 …… 294
11.2.1 基于距离的可分性判据 …… 294
11.2.2 基于概率密度函数的可分性判据 …… 297
11.3 特征选择的方法 …… 299
11.3.1 最优搜索算法 …… 300
11.3.2 次优搜索算法 …… 301
11.3.3 遗传算法 …… 303
11.4 线性判别分析——LDA …… 306
11.4.1 线性判别分析的概念 …… 307
11.4.2 广义线性判别函数 …… 308
11.4.3 Fisher 线性判别 …… 310
11.4.4 多类问题 …… 314
11.5 主分量分析——PCA …… 315
11.5.1 基于 K－L 变换的主分量分析 …… 316
11.5.2 随机矢量的 K－L 展开 …… 316
11.5.3 基于 K－L 变换的降维 …… 318
11.6 独立分量分析 …… 319
11.6.1 线性独立分量分析 …… 319
11.6.2 线性独立分量分析算法 …… 324

11.6.3 独立分量分析的预处理 …… 329
11.6.4 非线性独立分量分析 …… 330
11.7 特征变换举例 …… 334
11.7.1 特征变换方法 …… 334
11.7.2 特征变换实验 …… 336
参考文献 …… 337
第12章 语音识别的模型 …… 339
12.1 动态时间规整 …… 339
12.1.1 动态时间规整的定义 …… 339
12.1.2 动态规划技术(DP) …… 340
12.1.3 DTW 算法的改进 …… 342
12.2 隐马尔可夫模型 …… 344
12.2.1 隐马尔可夫模型的定义 …… 344
12.2.2 HMM 中的 3 个基本问题及其解决方案 …… 346
12.2.3 隐马尔可夫模型的类型 …… 352
12.2.4 HMM 算法实现的问题 …… 352
12.3 分类模型——SVM …… 362
12.3.1 学习问题 …… 362
12.3.2 学习过程一致性的条件 …… 363
12.3.3 学习过程收敛速度的界 …… 365
12.3.4 结构风险最小归纳原理 …… 367
12.3.5 支持矢量机 …… 370
12.4 人工神经网络 …… 377
12.4.1 神经元的基本模型 …… 377
12.4.2 前向网络 …… 379
12.4.3 反馈网络 …… 382
12.5 高斯混合模型(GMM) …… 387
12.5.1 高斯混合模型的定义 …… 388
12.5.2 参数调整算法——EM 算法 …… 388
12.6 动态贝叶斯网络 …… 390
12.6.1 贝叶斯网络 …… 391

12.6.2 动态贝叶斯网络 …… 393
12.6.3 动态贝叶斯网络在语音识别中的应用 …… 395
12.6.4 基于 DBN 的语音识别软件 GMTK …… 398
参考文献 …… 400
第 13 章 稳健性识别技术 …… 402
13.1 稳健性识别技术概述 …… 402
13.2 语音增强 …… 405
13.2.1 多带谱减法
(Multi-Band Spectral Subtraction, MBSS) …… 406
13.2.2 短时谱估计(Short Time Spectral Estimator) …… 407
13.2.3 瞬时维纳滤波
(Instantaneous Wiener Filtering, IWF) …… 407
13.2.4 子空间法(Subspace) …… 408
13.3 信道补偿 …… 410
13.3.1 多重风格训练 …… 411
13.3.2 HMM 分解 …… 411
13.3.3 并行模型组合 PMC
(Parallel Model Combination) …… 413
13.3.4 矢量泰勒级数(Vector Taylor Series, VTS)方法 … 417
13.3.5 雅可比自适应(Jacobian Adaptation) …… 420
13.3.6 其他补偿方法 …… 421
13.4 说话人自适应技术 …… 421
13.4.1 最大似然度线性回归算法 …… 423
13.4.2 最大后验概率算法 …… 431
13.4.3 说话人聚类 …… 435
13.5 说话人归一化技术 …… 442
13.5.1 说话人归一化技术原理 …… 442
13.5.2 频率折叠因子的选取 …… 444
13.5.3 折叠方法的选取 …… 447
参考文献 …… 452