



中国计算机学会学术著作丛书

说话人识别 模型与方法

吴朝晖 杨莹春 著

清华大学出版社



中国计算机学会学术著作丛书

说话人识别 模型与方法

Models and Methods for
Speaker Recognition

吴朝晖 杨莹春 著

清华大学出版社
北京

内 容 简 介

说话人识别是根据语音波形中反映说话人生理和行为的特征的语音参数,自动识别说话人身份的技术。本书作者结合多年的科研工作,分5个部分介绍了说话人识别的基本概念、方法以及最新研究进展。第1部分概括介绍说话人识别的主要概念、基本原理、研究历史与现状,以及测试语料库的构建;第2部分介绍作者对特征提取提出的不同改进方法,包括特征组合与特征变换;第3部分是作者提出的新的说话人识别模型,包括支持向量机、动态贝叶斯网络、主成分分析;第4部分介绍作者在基于信息融合的说话人识别上的创新工作;第5部分介绍作者开发的一个软件平台及其基础上的两个具体应用系统,最后是全书总结并展望发展趋势。

本书可供信息工程、电子工程、计算机科学与技术、公安、军事侦察等领域的科技工作者参考,也可以作为高等院校信号与信息处理、通信与电子系统、模式识别、生物医学等学科专业的研究生或高年级本科生的教学参考书。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

说话人识别模型与方法/吴朝晖,杨莹春著. —北京:清华大学出版社,2009.3

(中国计算机学会学术著作丛书)

ISBN 978-7-302-18968-8

I. 说… II. ①吴… ②杨… III. 言语识别—计算机应用—研究 IV. TP391.4

中国版本图书馆 CIP 数据核字(2008)第 185589 号

责任编辑:薛 慧

责任校对:赵丽敏

责任印制:杨 艳

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

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

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京密云胶印厂

装 订 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:175×245 印 张:21.5 字 数:443 千字

版 次:2009 年 3 月第 1 版 印 次:2009 年 3 月第 1 次印刷

印 数:1~3000

定 价:53.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:019620-01

前言

Foreword

说

话人识别属于生物特征识别技术的一种,是一项根据语音波形中反映说话人生理和行为的特征的语音参数,自动识别说话人身份的技术。与语音识别不同的是,说话人识别利用的是语音信号中的说话人信息,而不考虑语音中的字词意思,它强调说话人的个性;而语音识别的目的是识别出语音信号中的言语内容,并不考虑说话人是谁,它强调共性。说话人识别技术的崛起得益于信号检测与处理、模式识别、人工智能、机器学习等理论与技术的发展,这是一个涉及生理学、心理学、声学、语音学等多学科的研究领域。

本书结合我们对说话人识别进行的研究和工作,在对说话人识别的基本概念和方法进行简要介绍的基础上详细介绍了我们在测试语料库、特征组合、特征变换、识别模型以及应用系统开发的最新重要研究成果。我们从事说话人识别研究至今已8年有余,简单回顾一下作者与本书写作内容有关的几个重要时间标记。

1999年,在杭州中正生物认证有限公司的资助下,我们的生物认证实验室宣告成立。

2001年3月一天的午后,中国科学院自动化所模式识别国家重点实验室徐波研究员应邀风尘仆仆来到曹光彪主楼208,将数十年语音识别技术纵横史娓娓道来,对语音研究的挚爱溢于言表,坚定了我们从事说话人识别研究的信心和

方向。

2001年10月与中国科学院自动化所联合成功组织、举办了中国第二届生物特征识别研讨会,获得国家“863”计划资助。

2002年提出了基于SVM-HMM的说话人识别模型,该成果发表于国际语音处理会议 ICSLP 2002 (International Conference on Spoken Language Processing)。获得浙江省自然科学基金青年科技人才培养专项基金和博士点专项基金资助。

2003年提出了基于主元分析(PCA)的说话人识别模型,该成果发表于机器学习会议 IJCNN 2003 (IEEE International Joint Conference on Neural Network)。又提出了基于动态贝叶斯网络(DBN)的说话人识别模型,该成果发表于国际语音处理会议 ICASSP 2003 (IEEE International Conference on Acoustics, Speech and Signal Processing)和2003年国际著名杂志 IEE Electronic Lettters。完成国内外第一个面向移动互联环境的多通道说话人识别语料库 SRMC 的采集。首次提出基于情感补偿的活体声纹识别模型研究问题,获得国家自然科学基金资助。

2004年提出了基于声门特征的倒谱补偿模型,该成果发表于机器学习会议 IJCNN 2004 (IEEE International Joint Conference on Neural Network); 提出了基于声门特征的倒谱平均减算法,该成果发表于美国声学学会年会 ASA (Proceedings of 148th Meeting of the Acoustic Society of America)。两项发明专利均获准授权。获得浙江省自然科学基金资助。“支持说话人识别研究与开发的开放式平台 SONAR”通过省级鉴定。

2005年提出了基于VQ核的SVM模型和基于声门信息的并行GMM模型,该成果发表于国际生物认证会议 AVBPA 2005 (Audio- and Video-based Biometric Person Authentication 2005)。又提出了基于混合支持向量机的说话人识别和基于得分差加权融合的多模态说话人识别方法,该成果发表于国际语音处理会议 Interspeech 2005 (9th European Conference on Speech Communication and Technology)。

本书的撰写,既参考了他人的有关文献,又结合了作者近年在该领域的研究工作,基本上遵循我们对相关问题的研究思路展开,使理论性、实用性、系统性相结合,不仅有较系统全面的原理介绍,还结合科研成果给出了许多实例与结果。希望能为说话人识别研究人员提供有益借鉴。

本书分为5部分。第1部分概括介绍说话人识别的主要概念、基本原理、研究历史与现状,以及测试语料库的构建;第2部分介绍我们对特征提取提出的不同改进方法,包括特征组合与特征变换;第3部分介绍我们提出的新的说话人识别模型,包括支持向量机、动态贝叶斯网络、主成分分析;第4部分介绍我们在基于信息融合的说话人识别上的创新工作;第5部分介绍我们开发的一个软件平

台及其基础上的两个具体应用系统,最后是全书总结并展望发展趋势。

本书是作者和学生们共同研究成果的总结。多年来,先后有忻栋、陈大为、马志友、章万锋、桑立锋、俞成功、郑海树、杨璞、吕刚、雷震春、李冬冬、单振宇、徐卢传、任舒彬、黄挺、杨旻、吴甜、周森、刘漪琰、余奇、魏春明、陈文翔、陈力等直接参与了有关研究工作,本书也使用了他们学位论文和发表文章的一些内容。衷心祝愿青年才俊们前程似锦。

1999年在我们刚迈入生物认证研究领域之时,承蒙杭州中正生物认证有限公司的梁樵女士、孙黎先生、邱柏云先生和郝云龙先生的热情关心与帮助。中国科学院自动化所模式识别国家重点实验室的谭铁牛研究员、徐波研究员,北京航空航天大学王蕴红教授对我们的研究工作给予了极大的关注和支持。作者曾与中国科学院声学所的俞铁城研究员,北京大学视觉与听觉信息处理国家重点实验室的吴玺宏教授,封举富教授,北京交通大学袁保宗教授,中国社会科学院民族所鲍怀翘研究员等进行深入、有益的探讨,在此一并向他们致以衷心的感谢。

本书的撰写先后得到了国家杰出青年基金 60525202、国家自然科学基金 60533040/60273059、教育部新世纪优秀人才计划 NCET-04-0545、国家高技术研究发展计划 2001AA4180/2006AA01Z136、浙江省自然科学基金 M603229/Y106705、浙江省自然科学基金青年科技人才培养专项基金 RC01058、博士点专项基金 20020335025 等多项资助。

作者

2008年10月于浙江大学求是园

目 录

Contents

第一篇 绪 论

第 1 章 背景与概述	3
1.1 研究背景及意义	3
1.1.1 说话人识别介绍	3
1.1.2 说话人识别的优势与应用前景	5
1.2 研究进展与趋势	6
1.2.1 研究历史	6
1.2.2 研究现状	8
1.2.3 发展趋势	9
1.2.4 存在的问题	12
1.3 本书结构	13
参考文献	14
第 2 章 技术基础与理论	16
2.1 背景知识	16
2.2 说话人识别系统结构	17
2.3 特征提取	18
2.3.1 预处理	19
2.3.2 美尔倒谱特征	22
2.3.3 线性预测系数	23
2.3.4 Delta 特征和 Delta_Delta 特征的计算	24

2.3.5 声门特征	24
2.4 说话人识别模型	26
2.4.1 高斯混合模型	27
2.4.2 隐马尔可夫模型	31
2.4.3 动态时间规整模型	36
2.4.4 向量量化模型	36
2.5 得分规整	37
2.6 系统性能评价	38
2.6.1 评价指标	38
2.6.2 性能与用户规模的关系	39
2.6.3 实际使用要求	40
2.7 小结	42
参考文献	42
第3章 说话人识别语料库	44
3.1 常用语料库	44
3.2 面向移动互联环境的说话人识别语料库(SRMC)	48
3.2.1 SRMC 的设计思路	49
3.2.2 SRMC 录音方案	49
3.2.3 SRMC 录音内容	52
3.2.4 SRMC 存储与标注	54
3.3 电话语音库(PHONE)	55
3.4 多模态说话人识别库	55
3.5 NOISEX-92 数据库	58
3.6 小结	58
参考文献	59

第二篇 特征提取

第4章 说话人特征分析与优化	63
4.1 特征性能分析	63
4.1.1 阶数的影响	63
4.1.2 帧长的影响	67
4.1.3 结论	72
4.2 特征参数优化	72
4.2.1 语音包络检测	72

4.2.2	包络最小长度限制	73
4.2.3	预加重参数选取	74
4.2.4	语音起始点的去除	74
4.2.5	Delta 特征的引入	75
4.2.6	训练音长度的影响	75
4.2.7	结论	76
4.3	特征组合	76
4.3.1	单一特征组合	77
4.3.2	不同特征组合(小规模用户)	81
4.3.3	不同特征组合(中等规模用户)	84
4.4	二次特征提取	87
4.5	小结	90
	参考文献	91
第 5 章	基于主成分分析(PCA)的说话人特征变换	92
5.1	高维说话人特征的缺陷	92
5.2	说话人特征与 PCA 变换	93
5.2.1	说话人特征	93
5.2.2	PCA 变换的流程与效果	94
5.2.3	说话人特征的 PCA 变换	95
5.3	PCA 特征变换应用于说话人鉴别	96
5.3.1	传统的说话人鉴别系统	96
5.3.2	基于 PCA 特征变换的可行性	97
5.4	局部 PCA 特征变换	97
5.4.1	基于局部 PCA 特征变换的说话人鉴别系统	97
5.4.2	实验结果分析	98
5.4.3	结论	105
5.5	全局 PCA 特征变换	106
5.5.1	基于全局 PCA 特征变换的说话人鉴别系统	106
5.5.2	实验结果分析	107
5.5.3	结论	112
5.6	基准系统、局部 PCA 变换与全局 PCA 变换的比较	112
5.6.1	可扩充性比较	112
5.6.2	识别性能比较	113
5.7	小结	117
	参考文献	118

第 6 章 基于线性判别分析(LDA)的说话人特征变换	119
6.1 LDA 变换与 PCA 变换的联系与区别	119
6.1.1 LDA 转换公式与 PCA 转换公式	119
6.1.2 LDA 变换和 PCA 变换的原理的比较	120
6.1.3 用 LDA 对说话人特征进行变换	120
6.2 LDA 特征变换	121
6.2.1 基于 LDA 特征变换的说话人鉴别系统	121
6.2.2 实验结果分析	122
6.2.3 结论	128
6.3 基准系统、全局 PCA 变换与 LDA 变换的比较	128
6.3.1 可扩充性比较	128
6.3.2 识别性能比较	129
6.4 小结	133
参考文献	134
第 7 章 基于轨线模型的说话人特征时序性发掘	135
7.1 基于段模型的说话人特征时序性发掘	135
7.1.1 段模型	135
7.1.2 段模型在语音识别中的应用	137
7.1.3 说话人特征时序性发掘方法	138
7.1.4 时序性发掘实验	141
7.2 基于 Trended HMM 的文本相关说话人识别	145
7.2.1 Trended HMM	145
7.2.2 Trended HMM 在语音识别中的应用	149
7.2.3 文本相关的说话人识别	150
7.2.4 Trended HMM 与 VIV	153
7.2.5 Trended HMM 优缺点	154
7.3 小结	155
参考文献	155

第三篇 识别模型

第 8 章 基于支持向量机的识别模型	159
8.1 研究意义	159
8.2 支持向量的区域描述	160

8.2.1 闭集与开集	160
8.2.2 支持向量的区域描述	164
8.2.3 说话人辨认	166
8.3 支持向量机的概率输出	167
8.3.1 概率	167
8.3.2 支持向量机的概率输出	170
8.3.3 内嵌支持向量机(SVM)的隐马尔可夫模型(HMM)	174
8.3.4 支持向量机(SVM)与高斯混合模型(GMM)的混合模型 ...	176
8.4 基于向量量化(VQ)模型的核方法	178
8.5 基于 GMM 模型的核方法	180
8.6 多 SVM 混合模型	182
8.7 小结	184
参考文献	184
第 9 章 基于动态贝叶斯网络的识别模型	186
9.1 动态贝叶斯网络	186
9.1.1 表达	186
9.1.2 推导	188
9.1.3 学习	195
9.1.4 结论	198
9.2 基于动态贝叶斯网络(DBN)的说话人识别	198
9.2.1 基于动态贝叶斯网络的识别框架	199
9.2.2 实验和讨论	204
9.3 小结	208
参考文献	208
第 10 章 基于主成分分析分类器的说话人识别	210
10.1 说话人分类常用算法的局限性	210
10.2 主成分分析分类原理	211
10.2.1 主成分分析的递归定义	211
10.2.2 主成分分析的分类依据	212
10.3 两种主成分分析分类器及其决策融合	213
10.3.1 基于主成分分子空间的分类器	213
10.3.2 基于截断误差子空间的分类器	214
10.3.3 两种主成分分析分类器的决策融合	215
10.4 主成分分析分类器应用于说话人鉴别	216

10.4.1	模型训练	216
10.4.2	模型测试	217
10.5	实验结果分析	217
10.5.1	无噪语料库	217
10.5.2	有噪语料库	219
10.5.3	与高斯混合模型(GMM)方法和向量量化(VQ)方法的 比较	221
10.5.4	结论	223
10.6	复杂度分析	223
10.6.1	P&T 分类器的计算复杂度	224
10.6.2	高斯混合模型的计算复杂度	225
10.6.3	两者计算复杂度的比较	226
10.7	小结	226
	参考文献	227

第四篇 信息融合

第 11 章	声门信息融合	231
11.1	基于声门特征的说话人识别研究现状	231
11.1.1	声门特征应用于说话人识别	231
11.1.2	基音周期的提取	235
11.2	基于声门特征的倒谱补偿算法	238
11.2.1	声门特征对于倒谱特征的影响	238
11.2.2	基于声门特征的倒谱补偿模型	242
11.2.3	多通道环境下的倒谱补偿	248
11.3	基于声门特征的并行高斯混合模型	255
11.3.1	并行高斯混合模型的理论框架	255
11.3.2	并行高斯混合模型的子空间划分	256
11.3.3	子空间模型的融合	257
11.3.4	实验结果及分析	258
11.4	基于声门特征的倒谱平均减	261
11.4.1	倒谱平均减技术	262
11.4.2	基于声门特征的倒谱平均减算法	263
11.4.3	实验结果及分析	265
11.5	小结	268
	参考文献	269

第 12 章 人脸信息融合	271
12.1 多模态说话人识别研究	271
12.1.1 融合框架	272
12.1.2 声纹识别模型	272
12.1.3 人脸识别模型	273
12.1.4 融合方法	275
12.1.5 融合效果分析	279
12.2 基于得分差加权和融合的双模态说话人识别	281
12.2.1 表达	282
12.2.2 实验	282
12.3 动态贝叶斯网络在多模态说话人鉴别上的应用	283
12.3.1 说话人鉴别融合框架	283
12.3.2 基于动态贝叶斯网络的特征级融合	284
12.3.3 说话人识别的实验和讨论	286
12.4 小结	288
参考文献	288

第五篇 应用展望

第 13 章 支持说话人识别研究与开发的开放式平台 SONAR	295
13.1 SONAR 平台架构	296
13.1.1 简介	296
13.1.2 SONAR 测试平台界面	297
13.1.3 SONAR 核心模块	298
13.2 特征模块	299
13.2.1 预处理算法	299
13.2.2 特征提取	300
13.3 模型模块	300
13.3.1 模型集合	300
13.3.2 模型融合判决	301
13.4 SONAR 平台可扩展性	301
13.4.1 SONAR 平台特点	301
13.4.2 可扩展性	302
13.5 小结	303
参考文献	303

第 14 章 应用系统	304
14.1 声纹打卡系统	304
14.1.1 开发背景	304
14.1.2 系统体系结构	305
14.1.3 说话人识别	308
14.1.4 性能评估	309
14.2 移动互联环境下的说话人识别系统	311
14.2.1 应用背景	311
14.2.2 系统结构	312
14.2.3 使用说明	312
14.3 小结	316
参考文献	316
第 15 章 总结与展望	318
15.1 全书总结	318
15.2 工作展望	323
15.2.1 基于声门信息的说话人识别	323
15.2.2 引入高层信息的说话人识别	323
15.2.3 基于情感补偿的活体声纹识别	325
15.3 结语	328
参考文献	328

第一篇

绪 论

1.1 研究背景及意义

1.1.1 说话人识别介绍

语音是实现人们之间沟通交流的最直接与方便的手段,而实现计算机与人之间畅通无阻的语音交流,一直是人类不懈追求的一个梦想,语音识别则是实现这一梦想的关键性技术。语音识别是指计算机对人类语音进行正确响应的技术^[1]。广义的语音识别(speech recognition)技术具体包括:语音识别(识别说话内容)、说话人识别(识别说话人是谁)、语种识别(识别说话语言种类)、语音评分(评价发音的标准程度)。

说话人识别(speaker recognition, SR)技术(也称声纹识别技术)属于生物认证技术的一种,是一项根据语音波形中反映说话人生理和行为特征的语音参数,自动识别说话人身份的技术。说话人识别技术的核心是通过预先录入说话人的声音样本,提取说话人独一无二的语音特征并保存在数据库中,应用时将待验证的声音与数据库中的特征进行匹配,从而决定说话人的身份。说话人识别技术以其独特的方便性、经济性和准确性受到世人瞩目。

语音中既包含说话人的生理特征,即先天发音器官差异,又包含说话人的行为特征,即后天的发音与言语习惯的特殊征象。说话人识别与语音识别之间有很大的区别,前者从语音中提取