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。

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

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

第四篇 信息融合

第五篇 应 用 展 望

第一篇

绪　论

第1章 背景与概述

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