

----- 经济转型与创新发 展论丛

鲁棒性说话人识别技术

——在移动商务中的应用研究

徐丽敏 著



南京大学出版社

----- 经济转型与创新发 展论丛 -----

鲁棒性说话人识别技术

——在移动商务中的应用研究

徐丽敏 著



南京大学出版社

图书在版编目(CIP)数据

鲁棒性说话人识别技术:在移动商务中的应用研究 / 徐丽敏著. —南京:南京大学出版社, 2011. 9

ISBN 978-7-305-08890-2

I. ①鲁… II. ①徐… III. ①语音识别—鲁棒设计—研究 IV. ①TP391.4

中国版本图书馆 CIP 数据核字(2011)第 199353 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

网 址 <http://www.NjupCo.com>

出版人 左 健

丛 书 名 经济转型与创新发 展论丛

书 名 鲁棒性说话人识别技术——在移动商务中的应用研究

著 者 徐利敏

责任编辑 王日俊 梁 颖 编辑热线 025-83592193

照 排 南京紫藤制版印务中心

印 刷 盐城市华光印刷厂

开 本 880×1230 1/32 印张 5.25 字数 180 千

版 次 2011 年 12 月第 1 版 2011 年 12 月第 1 次印刷

ISBN 978-7-305-08890-2

定 价 22.00 元

发行热线 025-83685951

电子邮箱 Press@NjupCo.com

Sales@NjupCo.com(市场部)

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

本书为江苏省教育厅高校自然项目(2011年)“鲁棒性说话人识别技术在移动商务中的应用研究”(10KJB520007)、江苏高校优势学科建设工程资助项目和江苏现代服务业研究院专项科研基金项目的研究成果。

本书出版得到江苏高校优势学科建设工程资助项目支持。

前 言

移动商务是指对通过移动通讯网络进行数据传输,并且利用移动终端开展各种商业经营活动的一种新电子商务模式。移动商务是与商务活动参与主体最贴近的一类电子商务模式,其商务活动中以应用移动通讯技术使用移动终端为特性。由于用户与移动终端的对应关系,通过与移动终端的通讯可以在第一时间准确地与对象进行沟通,使用户更多脱离设备网络环境的束缚最大限度地驰骋于自由的商务空间。

随着电子商务的逐渐发展,安全隐患也有新增。例如信用卡信息被盗用,敏感信息泄密等时有发生并呈上升趋势。在开放的网络上尤其是无线网络上数据的存储、处理、传输等越来越容易,而如何保证传输的交易数据的安全成为电子商务更是移动商务能否展开并普及的最重要因素之一。因此,移动商务环境下的鲁棒性说话人识别问题也显

得特别重要。

为了提高说话人辨认系统的性能和在实际应用中的鲁棒性,作者在高斯混合模型特征变换、特征加权补偿变换和自适应直方图均衡化等进行了研究,研究成果包括:

1. 提出了基于嵌入变换的对角方差矩阵高斯混合模型的多步聚类算法。为了简便计算,高斯混合模型中的方差矩阵通常直接用对角方差矩阵代替,因而会对相似度的计算产生损失。为了弥补由于采用对角方差矩阵而引起的相似度损失,提出了基于嵌入变换的对角方差矩阵高斯混合模型的多步聚类算法。该方法采用嵌入变换的对角方差矩阵来建立模型;同时将多步聚类算法融入其中,使高斯混合模型能找到其最适合的模型混合数。与普通聚类期望最大(EM)算法相比,多步聚类算法所需的EM估计次数明显减少;与聚类EM估计的GMM方法相比,在同一语音库下平均计算时间降低了约50%,错误识别率平均减少1.4%;在自制和公开的两个语音库下,与嵌入变换的GMM估计方法相比,新方法都可以直接达到说话人辨认错误识别率的最佳点,达到了识别效果和识别时间的统一。

2. 提出了基于高斯混合模型的加权特征补偿

变换的抗噪声算法。针对特征加权算法的局限性和归一化补偿变换方法的特性,提出了基于高斯混合模型的加权特征补偿变换的抗噪声算法。一方面根据帧信噪比对特征值的贡献大小进行加权;另一方面根据说话人识别的声学特性对模型输出的似然得分进行变换,补偿了加权因子在某些环境下的局限性。对于不同信噪比的平稳和非平稳噪声环境,在自制语音库下,与特征加权算法相比,该算法平均识别率提高了2.74%和2.82%;与归一化补偿变换方法相比,平均识别率提高了3.56%和1.34%。在另一公开语音数据集下,与特征加权算法相比,该算法平均识别率提高了3.02%和2.56%;与归一化补偿变换方法相比,平均识别率提高了3.9%和1.14%。

3. 提出了基于统计模型的自适应直方图均衡化方法。针对说话人特征的统计特性和直方图均衡化在说话人识别中应用的不足之处,提出了应用于说话人辨认中的自适应直方图均衡化方法。该方法首先用较大的区间长度来构造直方图的累积函数,然后根据各区间内特征值频率增量的大小来自适应确定该区间是否需要再划分以及划分的程度。采用这种方法不仅使计算量降低,而且得到的

变换特征值的分布更符合实际特征空间,从而进一步提高了噪声环境下说话人辨认系统的识别率和鲁棒性。在同一测试集下,研究两种常用经典噪声(White 和 Babble),与普通直方图均衡化方法相比,自适应直方图均衡化方法的平均识别率分别提高了 3% 和 2.9%。

目 录

第一章 绪 论	1
第一节 说话人识别概述	2
一、背景及意义	2
二、说话人识别基本原理	3
第二节 说话人识别系统的研究与发展	6
第三节 说话人识别中的特征提取与分析	8
第四节 说话人识别中的识别方法	10
一、模板匹配法	10
二、统计概率模型法	11
三、人工神经网络	13
四、支持向量机法	13
第五节 性能评价	14
第六节 研究现状和难点	16
一、特征变换	17
二、鲁棒性说话人识别	18

第七节	研究工作概述与安排	21
一、	工作内容	21
二、	全书的组织结构	24
第二章	基于嵌入变换的多步聚类变换算法的说话人辨认	27
第一节	引言	27
第二节	基于高斯混合模型的说话人辨认	29
一、	说话人辨认系统	29
二、	高斯混合说话人模型	31
三、	前端处理	33
四、	MFCC 参数	34
第三节	基于嵌入变换的对角方差高斯混合模型	35
一、	嵌入变换的对角方差高斯混合模型参数	35
二、	最大似然估计的 EM 算法	37
第四节	基于对角方差矩阵的多步聚类 EM 算法	39
一、	基于对角方差矩阵的聚类 EM 算法	40

二、多步聚类算法及其步骤	43
第五节 实验语音库	46
第六节 实验操作	47
一、最小描述长度(MDL)比较实验	48
二、多步聚类说话人辨认实验	52
第七节 结论	57

第三章 基于加权特征补偿变换的鲁棒性说话人 辨认

第一节 引言	59
第二节 前端处理	63
一、语音检测	64
二、噪声谱估计	66
三、谱减	66
四、帧信噪比与加权因子	67
第三节 加权特征补偿变换算法	70
一、加权算法	71
二、对帧似然概率的归一化补偿变换	72
三、加权特征补偿变换的提出	75
第四节 鲁棒说话人识别仿真系统	79

一、预处理和特征提取	79
二、高斯混合模型的建立与识别	80
第五节 仿真实验与结果分析	81
一、仿真系统语音库	81
二、加权因子仿真实验	82
三、基于加权特征补偿变换的说话人识别 方法的仿真研究	84
第六节 结论	89
第四章 基于自适应直方图均衡化的鲁棒性说话 人辨认研究	91
第一节 引言	91
第二节 直方图均衡化	96
一、直方图均衡化原理	96
二、分位差直方图均衡化法	101
三、数据拟合直方图均衡化	103
第三节 改进的自适应直方图均衡化	106
一、问题的提出	106
二、自适应直方图均衡化方法	109
第四节 鲁棒说话人辨认实验的建立	111
第五节 实验与结果分析	113
一、变换性能实验	114

二、效率比较实验	119
三、识别性能比较实验	121
第六节 结论	124
第五章 总结与展望	126
第一节 本研究工作总结	126
第二节 研究展望	130
参考文献	133

第一章

绪论

众所周知,语音是人类获取信息的主要来源之一,也是人与外界交流中使用最方便、最有效、最自然的工具。最初人们是通过人耳来辨别语音的说话人身份,即“闻声识人”。随着计算机的出现和电子信息技术的发展,出现了用计算机自动识别语音的话者身份的技术,即说话人识别(Speaker Recognition)技术。说话人识别技术有着非常广阔的应用前景:在司法领域,它可以用来协助确认犯罪嫌疑人;在军事领域,它可以用于战场侦听,以辨认敌方指挥员;在银行等处的安全系统中,它可以作为身份核查或安全检查的一种手段;在日常生活中,它可以用作个人身份认证的手段,如声控门、声控命令,等等。

说话人识别根据应用的范畴可分为两类:第一类是说话人辨认(Speaker Identification),把待识别的语句判定为属于N个参考说话人中的某一位,是

一个多选一的问题；另一类是说话人确认(Speaker Verification)：确定一段说话人的语句是否与所声明的参考说话人相符，这只有两种选择：或是肯定(即接收)，或是否定(即拒绝)。由于时间和精力有限，本著作中主要基于特征变换和鲁棒性技术这两个方面对说话人辨认进行了研究，并提出了自己的一些见解。

本章的内容安排如下：首先，对说话人识别技术的组成与发展做简要介绍，并且着重介绍了说话人辨认系统的组成结构和原理；然后，综述了说话人识别中特征变换技术和鲁棒性技术的国内外研究现状，并指出了其研究的重点和难点；最后，给出本书工作的安排和具体内容。

第一节 说话人识别概述

一、背景及意义

以人类生物特征(指纹、声纹、人脸、虹膜等)进行身份验证的生物识别技术，正逐渐成为信息产业中极为重要的前沿技术。生物识别技术具有无需记忆、不会遗忘、适用方便等优点。其中说话人识

别还具有以下特点：

(1) 用户接受程度高。由于不涉及隐私问题，用户无任何心理障碍。

(2) 用户使用方便、经济。利用语音进行身份识别是最自然和最经济的方法之一；声音输入设备制造低廉，而其他生物识别技术的输入设备往往价格昂贵。

(3) 适合远程身份确认。在进行远程身份确认无法获取其他生物特征时，说话人识别只需要一个麦克风或电话、手机就可以通过网络（通讯网络或互联网络）进行基于说话人语音的身份认证。

(4) 算法复杂度低，配合一些其他措施，如通过语音识别进行内容鉴别等，可以提高准确率和安全性。

说话人识别可以应用的范围很宽，几乎遍及人们日常生活的各个角落。因此，研究高性能、高鲁棒性的说话人识别系统有重要的实际意义。

二、说话人识别基本原理

说话人识别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来识别待测语音话者身份的技术^[1]。说话人识别系统可以简单地定义

为：以说话人的语音作为输入，用训练得到的特定人模型来识别待测语音的说话人身份。图 1.1 是典型的说话人识别系统的模块示意图，从图中可以看到说话人识别系统的两个组成阶段：训练阶段和识别阶段。

在训练阶段，说话人的语音经过特征提取后得到各自的声学特征，然后系统为每个目标说话人建立相应的模型并组成说话人模型库；在识别阶段，用测试语音提取出的声学特征与说话人模型库中的模型进行比较，根据一定的相似性准则来判断测试语音发出者的身份。

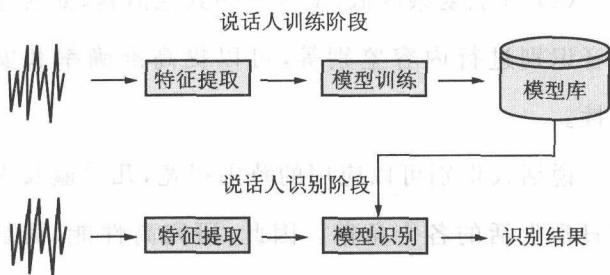


图 1.1 典型的说话人识别系统示意图

说话人识别根据应用的范畴可分为两类：说话人辨认和说话人确认。在实际应用中，说话人确认的灵活性和应用性一般要好于说话人辨认，这是因为说话人确认系统允许待测说话人不属于已知的