

Research on Multimodal Emotion Recognition for
Speech and Facial Expression Signal

面向语音与面部表情信号 的多模式情感识别技术研究

韩志艳 著

面向语音与面部表情信号的多模式情感 识别技术研究

Research on Multimodal Emotion Recognition for
Speech and Facial Expression Signal

韩志艳 著

东北大学出版社

· 沈 阳 ·

© 韩志艳 2017

图书在版编目 (CIP) 数据

面向语音与面部表情信号的多模式情感识别技术研究 /
韩志艳著. — 沈阳: 东北大学出版社, 2017. 1
ISBN 978-7-5517-1538-6

I. ①面… II. ①韩… III. ①情感—识别—语音识别
IV. ①TN912.34

中国版本图书馆 CIP 数据核字 (2017) 第 032669 号

内容简介

本书系统研究了面向语音与面部表情信号的多模式情感识别技术。全书共分为 7 章。第 1 章介绍了课题的国内外研究现状、意义和需要解决的难题; 第 2 章介绍了情感语音信号预处理技术; 第 3 章介绍了面部表情信号预处理技术; 第 4 章介绍了语音情感特征参数提取技术, 着重介绍了一种自适应语音情感动态特征提取方法; 第 5 章介绍了面部表情特征参数提取技术, 着重介绍了一种基于差值模板和 PCA 的面部表情特征提取方法; 第 6 章介绍了特征参数融合与识别技术, 着重介绍了一种基于 FASPC 的多模式情感信息融合方法; 第 7 章归纳全书并对今后工作提出展望。

本书的主要特点是在面向语音与面部表情信号的多模式情感识别技术方面提出了开创性的设计和分析方法, 书中的内容来源于作者近年来的创新性研究成果, 新颖实用, 研究方法先进, 尤其注重多模式情感识别的鲁棒性和实用性, 属于当前所属研究领域的前沿问题, 具有重要的理论与应用价值。

出版者: 东北大学出版社

地址: 沈阳市和平区文化路三号巷 11 号

邮编: 110819

电话: 024-83687331(市场部) 83680267(社务部)

传真: 024-83680180(市场部) 83687332(社务部)

E-mail: neuph@neupress.com

http://www.neupress.com

印刷者: 沈阳航空发动机研究所印刷厂

发行者: 东北大学出版社

幅面尺寸: 170mm × 240mm

印张: 12

字数: 260 千字

出版时间: 2017 年 1 月第 1 版

印刷时间: 2017 年 1 月第 1 次印刷

责任编辑: 孙锋 孙德海

封面设计: 潘正一

责任校对: 图 图

责任出版: 唐敏志

ISBN 978-7-5517-1538-6

定 价: 50.00 元

前 言

随着信息技术的高速发展和人类对计算机依赖性的不断增强，人机的交互能力越来越受到研究者的重视。如何实现计算机的拟人化，使其能感知周围的环境和气氛以及对象的态度、情感等内容，自适应地为对话对象提供最舒适的对话环境，尽量消除操作者和机器之间的障碍，已经成为下一代计算机发展的目标。斯坦福大学的 Reeves 和 Nass 的研究表明，在人机交互中需要解决的问题实际上与人和人交流中的重要因素是一致的，最关键的都是“情感智能”的能力。计算机要能够更加主动地适应操作者的需要，首先必须能够识别操作者的情感，而后再根据情感的判断来调整交互对话的方式。因此，情感识别的研究工作在人机交互领域中已经成为一个热点问题。

目前，国内外情感识别的研究主要有两大类：一类是单模式情感识别，另一类是多模式情感识别。所谓单模式情感识别为只从单一信息通道中获得当前对象的情感状态，如从语音信号、面部表情信号或生理信号（血压、体温、脉搏、心电、脑电、皮肤电阻等）等。虽然单一地依靠语音信号、面部表情信号和生理信号来进行情感识别的研究取得了一定的成果，但存在着很多局限性，因为人类是通过多模式的方式表达情感信息的，具有表达的复杂性和文化的相对性。比如，在噪声环境下，当某一个通道的特征受到干扰或缺失时，多模式方法能在某种程度上产生互补效应，弥补单模式的不足，所以研究多模式情感识别的方法十分必要。

本书以语音和面部表情信号为研究对象，以情感特征参数的提取和参数融合与识别算法为研究内容，实现对人类基本情感的识别。主要研究内容和创新点有以下几个方面：

① 为了提高低信噪比下语音端点检测的准确率，提出了一种端点检测算法。其核心技术是利用短时能零积与鉴别信息的互补优势，首先利用短时能零积的方法进行判决，当遇到噪声帧与语音帧的转折帧时，利用基于子带能量鉴别信息的方法来进行复检，从而避免了噪声幅度急剧变化导致的误检，并提出了一种动态更新噪声能量门限的方法，从而能更准确地跟踪噪声能量的变化。

仿真实验结果表明,提出的方法在信噪比变化比较剧烈的情况下仍能准确快速地检测出语音的起止点,对语音信号的后续研究起到了很好的铺垫作用。

② 研究用于融合的情感特征的优劣以及情感特征提取是否全面直接影响到情感融合的效果。但是目前所提的特征参数大多是反映语音情感的静态特征。语音情感的动态特性是指从连续几帧语音中提取的特征参数,它揭示了语音前后以及相邻之间存在着密切关联,因此更能体现情感。目前的动态特征主要是通过静态特征的差分参数和加速度参数来获得,但它们并不能将动态信息挖掘得很充分,所以本书引入调制谱和多信号分类法(Multiple Signal Classification, MUSIC)的谱估计技术,提出了一种自适应的语音情感动态特征提取方法,采用基于频率的自适应窗口调节技术,使之能充分地挖掘出情感的动态信息。

③ 本书以静态图像为研究对象,常用的特征提取方法主要有基于几何特征的方法、基于模型的方法以及 Gabor 小波法等。基于 Gabor 小波法能够同时检测多尺度、多方向的纹理变化,对亮度和人脸姿态的变化不敏感,但是运算量过大,所以本书针对 Gabor 小波变换特征提取方法运算量过大的问题,提出采用差值模板来模拟 Gabor 变换进行特征提取,并利用主成分分析法 PCA (Principal Component Analysis) 算法进行降维处理,在大大降低时间消耗的情况下,达到了与 Gabor 变换特征提取基本相似的识别性能。

④ 把遗传算法和小波神经网络算法相结合,即利用遗传算法具有的高度并行、随机、自适应搜索性能来选取初值进行训练,用小波神经网络来完成给定精度的学习。这样在解决复杂和非线性问题时,具有明显的优势。并将其与 BP 算法和小波神经网络算法进行了比较,实验结果表明,该模型不但能够提高情感识别的正确率,缩短系统识别时间,而且为算法的实用性奠定了基础。

⑤ 将局部保持投影算法(LPP)的思想融入到主元分析(PCA)的目标函数中,使得在原始变量空间投影到低维空间的过程中,不仅实现了整体方差的最大化,而且保持了局部邻近结构不变,有利于全局和局部特征的全面提取,克服了传统 PCA 方法只关注全局结构特征而忽略局部特征的缺陷。对比实验结果验证了该方法的可行性和有效性。

⑥ 将获取的信号进行预处理并提取情感特征参数,然后利用模糊核聚类算法对其进行聚类分析,即利用 Mercer 核,将原始空间通过非线性映射到高维特征空间,在高维特征空间中对多模式情感特征进行模糊核聚类分析。由于经过了核函数的映射,原来没有显现的特征突现出来。实验结果验证了该方法的可行性和有效性。

⑦ 针对传统的多模式情感信息融合算法效果不佳的问题,在 Bagging 算法

和 AdaBoost 算法的启发下，提出了一种串并结合的多模式信息融合方法 FASPC (Fusion Algorithm based Series - Parallel Combination)，充分发挥了决策层融合和特征层融合的优点，使整个融合过程更加接近人类情感识别，从而提高了情感识别的准确性。其中在 FASPC 算法中，选择了两种不同类型的分类器，使个体分类器之间差异变大，集成效果变好。并通过双误差差异性选择策略挑选出大于平均差异性水平的分类器用来做最终情感识别，降低了计算复杂度，提高了精度。

本书的内容为作者几年来的研究成果，内容新颖，属于当前所属研究领域的前沿问题，具有重要的理论与应用价值。

在本书的写作过程中，渤海大学工学院尹作友教授、王健博士等提出了许多有价值的建议，硕士研究生张鹏、夏桂阳、徐静、殷允飞在手稿整理、仿真实验和校对书稿等方面做了大量的工作，作者在此向他们表示由衷的感谢。

本书的出版获得了国家自然科学基金（项目编号：61503038、61403042）和辽宁省教育厅一般项目（项目编号：L2013423）的资助，在此一并表示感谢。

由于作者水平有限，不足之处在所难免，热诚欢迎读者与同行不吝赐教。

韩志艳

2017 年 1 月于渤海大学

目 录

第 1 章 绪 论	1
1.1 情感识别研究背景概述	1
1.2 情感的定义	1
1.3 情感的分类	2
1.4 情感计算	3
1.5 情感识别国内外研究现状	5
1.6 情感识别的研究意义及应用	6
1.7 情感识别研究中的难点	8
1.8 章节安排	9
第 2 章 情感语音信号预处理技术	15
2.1 概 述	15
2.2 语音生成系统和语音感知系统	15
2.2.1 语音生成系统	15
2.2.2 语音感知系统	17
2.3 语音信号的产生模型	20
2.4 语音信号的时域波形	22
2.5 音素与音节	25
2.6 基音与四声	25
2.7 语音信号的采样和量化	26
2.8 语音信号的预加重	27
2.9 语音信号的分帧和加窗处理	28
2.10 端点检测	29
2.10.1 几种常用的端点检测算法	30
2.10.2 4 种端点检测方法比较总结	34

2.10.3	基于短时能量和鉴别信息的语音端点检测算法	34
2.11	本章小结	42
第3章	面部表情信号预处理技术	46
3.1	视觉认知心理	46
3.2	人脸面部结构	48
3.3	面部肌肉结构	48
3.4	面部动作编码系统	50
3.5	面部表情图像预处理在系统中的作用	52
3.6	表情图像的光照补偿	52
3.6.1	线性变换法	53
3.6.2	对数变换法	53
3.6.3	指数变换法	53
3.6.4	直方图均衡化法	54
3.7	表情图像的尺寸归一化	56
3.7.1	眼睛定位算法	56
3.7.2	尺寸、角度归一化	60
3.8	本章小结	61
第4章	语音情感特征参数提取	64
4.1	语音情感表达概述	64
4.2	时域特征	67
4.2.1	短时能量	67
4.2.2	短时平均幅度	67
4.2.3	短时过零率	68
4.2.4	短时自相关函数	68
4.2.5	短时平均幅度差函数	69
4.3	倒谱特征	69
4.3.1	基于线性预测的倒谱参数 LPCC	69
4.3.2	基于 Mel 频率的倒谱参数 MFCC	70
4.4	感知线性预测参数	72
4.5	共振峰特征参数	73
4.6	基因特征参数	75
4.7	谐振强度特征参数	78

4.8	基于边际谱的长时特征参数	80
4.9	语音序列的相空间重构特征参数	82
4.10	自适应语音情感动态特征参数	83
4.11	本章小结	90
第5章	面部表情特征参数提取	97
5.1	概述	97
5.2	NMF 面部表情特征	98
5.3	Gabor 小波特征	100
5.4	基于光流法的动态表情提取	102
5.4.1	时空梯度法	102
5.4.2	匹配法	103
5.4.3	基于能量的方法	104
5.4.4	基于相位的方法	105
5.5	基于时序模板的动态特征提取	106
5.6	基于分数阶傅里叶变换的表情特征提取	109
5.6.1	一维分数阶傅里叶变换的定义和相关性质	109
5.6.2	一维分数阶傅里叶变换的离散算法	111
5.6.3	二维分数阶傅里叶变换	114
5.6.4	二维分数阶傅里叶变换性质	115
5.6.5	分数阶傅里叶变换提取表情信息	116
5.7	基于 HOG 的表情特征提取	120
5.8	基于主动外观模型的表情特征提取	122
5.8.1	特征点检测方法	122
5.8.2	建立形状模型	123
5.8.3	建立外观模型	125
5.8.4	联合形状和外观建立模型	126
5.8.5	面部特征点检测	126
5.8.6	面部动画参数提取	127
5.9	基于差值模板和 PCA 的面部表情特征提取	129
5.10	本章小结	130
第6章	特征参数融合与识别	137
6.1	概述	137

6.2	基于遗传小波神经网络的语音情感识别	138
6.2.1	小波神经网络的设计	138
6.2.2	优化方法	140
6.2.3	实验结果与分析	142
6.3	基于局部保持投影和主元分析的语音情感识别	144
6.3.1	PCA 和 LPP 算法分析	145
6.3.2	算法描述	145
6.3.3	仿真实验及结果分析	148
6.4	基于模糊核聚类的多模式情感识别算法	151
6.4.1	参数提取	151
6.4.2	算法描述	153
6.4.3	仿真实验及结果分析	155
6.5	基于 FASPC 的多模式情感信息融合方法	157
6.5.1	系统结构框架	157
6.5.2	常用算法介绍	157
6.5.3	融合识别算法具体步骤	165
6.5.4	仿真实验及结果分析	166
6.6	本章小结	168
第7章	总结与展望	178
7.1	主要工作及创新点	178
7.2	进一步研究的展望	180

第1章 绪论

1.1 情感识别研究背景概述

人非草木，孰能无情？在人与人之间的交际过程中每个人都会产生情感，不同的情感会对交际产生不同的影响。可见，情感在人们相互交际过程中起着极其重要的作用。借助情感表达所伴随着的外在表现信息，如情感化的语音信号或面部表情，人们可以很方便地相互沟通、相互了解。对于人类情感方面的研究，一直是生理学、神经学、心理学等领域的重要研究方向，近几年来备受工程领域研究者的关注，其目的就是设计出一种智能化、人性化的新型人机交互系统。然而，究竟什么是情感？情感是如何表示的呢？这是人类开展情感方面的研究首先要解决的问题。本章将先叙述相关情感的定义及分类等概念，然后探讨情感计算、情感识别的国内外研究现状、情感识别的研究意义及应用、情感识别研究中的难点，最后给出本书的章节安排。

1.2 情感的定义

什么是情感？这是一个从过去到现在一直为科学领域所不断争论的问题。尽管心理学家已经对情感机理的研究做了大量的工作，但到目前为止，研究者对于情感的定义还没有达成一致的观点。在不同领域的研究当中，研究者对情感的解释和理解也各不相同。

在心理学领域，Arnold^[1]把情感定义为一种认知评价，即个人能以自己的方式对某种特殊重要情景做出一种直接的、即时的感知判断。Darwin^[2]则认为情感与个人的相关行为习惯是密不可分的，而这种习惯是由个人的生存价值观所决定的与众不同的行为模式。在哲学领域，庞学铨^[3]把情感理解成客观上把握到的一种具有空间性的力量、气氛。在生理学领域，James^[4]提出一种情绪外周理论，即将情感定义为一种独立的心理过程，突出人体的神经系统及内脏器官对情感的产生所起到的重要作用。Kleinginna 等人^[5]则总结了近百名研究

学者对情感的不同定义和理解。

1.3 情感的分类

要研究语音信号的情感，首先需要根据某些特性标准对语音情感做一个有效合理的分类，然后在不同类别的基础上研究特征参数的性质。

Plutchik 等^[6]人经过多年研究，通过在激活评价空间上对情感进行分析，认为情感分布在一个圆形的结构上，结构的中心是自然原点。由于各种情感在自然原点的周围排成了一个圆形，所以这种对情感进行分类的方法叫作“情感轮 (Emotion Wheel)”，如图 1.1 所示。



图 1.1 情感轮

Fig. 1.1 Emotion wheel

不同于 Plutchik 的分类手段，Fox 提出的三级情感模型，则是按照情感中表现的主动和被动程度的不同将情感分成不同的等级，如表 1.1 所列。等级越低，分类越粗糙；等级越高，分类越精细。

表 1.1 Fox 的情感三级分类模型

Table 1.1 The emotion three level classification model for Fox

1 st Level	Approach			Withdrawal		
2 nd Level	Joy	Interest	Anger	Distress	Disgust	Fear
3 rd Level	Pride	Concern	Hostility	Misery	Contempt	Horror
	Bliss	Responsibility	Jealousy	Agony	Resentment	Anxiety

目前在情感识别的研究中常用的情感分类大多是 Plutchik “情感轮” 的八情感模型（高兴、期望、愤怒、厌恶、悲伤、惊奇、恐惧、赞同）、六情感模

型（高兴、愤怒、惊奇、悲伤、恐惧、中性）或者四情感模型（喜、怒、惊、悲）。

1.4 情感计算

1985年，人工智能创始人之一，美国的 Minsky 教授在其专著 *The Society of Mind*^[7] 中说道：“The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions”，强调了情感对机器智能的重要作用。1995年，美国神经生理学家 Damasio^[8] 在对人类大脑神经活动方面的研究中发现，人类的智能不仅仅表现为正常的逻辑推理和理性思维能力，而且也应该表现出一种正常的情感能力。实际上，中国的古话“晓之以理，动之以情”“合情合理”“尽在情理之中”也都表达了同样的意思，即人们在决策、处事、待人时，“情”和“理”常常是密不可分的。1997年，美国麻省理工大学的 Picard 教授提出了“情感计算”^[9]（Affective Computing）的概念，即将“情感计算”定义为“关于情感、情感产生以及影响情感方面的计算”，其研究的最终目标就是使计算机具有人性化、自然化的人机交互能力。

目前，对于情感计算方面的研究是一门综合认知科学、生理学、心理学、语言学、计算机科学等多学科交叉的热点研究课题，正越来越受到国内外科研机构和研究人员的重视。1997年，美国麻省理工大学多媒体实验室 Picard 教授等人成立了世界上第一个情感计算方面的研究小组，专门侧重于有关情感信号的获取与识别。美国卡内基梅隆大学专门研究开发基于情感计算的可穿戴式计算机，致力于研究情感计算技术的实际应用。为推动我国在此国际前沿方向的研究，2004年，我国国家自然科学基金委将情感计算理论与方法的研究首次列入拟资助的重点项目指南中。2009年，由清华大学心理学系、中国科学院心理学研究所等部门在南昌联合主办的首届全国认知科学研讨会，首次将“情感计算”列为认知科学领域重点关注的前沿课题之一。2010年，为了促进“情感计算”这个新兴主题的研究和发展，IEEE 计算机协会新创办了以“情感计算”命名的国际学术期刊 *IEEE Transaction on Affective Computing*。可见，有关情感计算的研究是当前人机交互、计算机科学、认知科学等领域亟待开垦的沃土之一。

情感计算的实现，需要一个完整的情感交流过程。这就涉及到一系列相关的理论和技术，其核心技术可以分为情感信号的获取、情感识别、情感理解与

反馈以及情感表达四大部分^[10]。情感信号的获取是指能够感知情感信号（如皮肤电阻及电流、语音信号、人脸表情等）的传感器的研制。美国麻省理工大学多媒体实验室在传感器设备研制方面达到了世界一流水平。目前，他们已成功研发出皮肤电流传感器、脉压传感器、肌电流传感器及汗液传感器等多种传感器。

情感识别是指对采集的情感信号进行分析和建模，并识别出人类内在的情感类型。情感识别是情感计算研究当中最基础、最重要的内容之一。人类情感的表达，往往伴有某些表情特征信息。借助表情，人们才能“察言观色”，洞悉他人的真实想法。“相由心生”“观人口面可知心”也表达了同样的道理。依据表情产生方式与部位的差异，人的表情主要分成3种：言语表情、面部表情和身体姿态表情。相对于言语表情和从面部表情而言，身体姿态表情（如手势、运动姿势）变化的规律性比较难以获取，因而情感识别的研究目前主要侧重于言语表情和面部表情的识别。其中，对言语表情的识别，通常被称为“语音情感识别”；对面部表情的识别，通常被称为“人脸表情识别”。尽管从言语表情获取的音频信息和从面部表情获取的视觉信息在进行情感识别时所起的作用都很大，但这两者各有自己的优缺点，也有着某种程度的互补作用。例如，言语表情容易受到周围环境噪声的干扰，而面部表情就不会出现这种情况。为了在言语表情和面部表情之间取长补短，有必要将语音情感识别技术和人脸表情识别技术融合在一起，以便对人类情感的类别进行更有效的判定。这就是所谓的“多模态情感识别”，即同时融合多种表情（如言语表情和面部表情）的情感识别。

情感理解与反馈是指计算机通过分析用户的情感所产生的原因，对用户的情感变化做出最恰当的反应。例如，当用户厌烦时，计算机就会积极主动地提供更加新鲜有趣的内容。当用户困惑时，计算机就会及时地鼓励用户，并且提供对用户很有价值的建议。情感理解与反馈是情感计算技术研究的核心，只有计算机具备了这样的能力，才能真正实现人机交互的自然化。情感表达是指给予某一种情感状态，研究如何使这一情感状态在一种或几种行为或生理特征当中体现出来。也就是要求计算机交互设备能够向用户模拟并表达出情感。例如计算机交互设备通过动画、语音等方式来模拟和表达情感。在人脸面部图像合成领域，现代的计算机已经能够合成出具有混合表情特征的面部图像。同样，由计算机合成的语音也不再那么机械和单调，而是能够体现出生气、高兴、悲伤等不同的情感语音基调。

1.5 情感识别国内外研究现状

近年来,美国、日本、英国、德国、荷兰、法国、新加坡等经济发达国家和如印度等发展中国家都有专门的研究组进行情感识别方面的研究。国内的许多大学如清华大学、中国科学院大学、哈尔滨工业大学、东南大学、北京航空航天大学、北方交通大学、西南大学、中国科学技术大学等都开展了相关的研究工作,并取得了卓有成效的进展。

目前,国内外情感识别的研究主要有两大类:一类是单模式情感识别,另一类是多模式情感识别。所谓单模式情感识别为只从单一信息通道中获得当前对象的情感状态,如从语音信号、面部表情信号或生理信号(血压、体温、脉搏、心电、脑电、皮肤电阻等)等。对于语音情感识别,1990年麻省理工大学多媒体实验室构造了一个“情感编辑器”,对外界各种情感信号进行采样来识别各种情感,并让机器对各种情感做出适当的反应^[11]。中科院的颜永红等^[12]采用非均匀子带滤波器来挖掘对语音情感有益的信息,加大了各类情感之间的鉴别性,提高了情感识别的性能。北京航空航天大学的毛峡等^[13]通过用相关密度和分形维数作为情感特征参数来进行语音情感识别,获得了较好的性能。东南大学的邹采荣等^[14]提出了一种基于改进模糊矢量量化的语音情感识别方法,有效地改善了现有模糊矢量量化方法的情感识别率。东南大学的赵力等^[15]通过采用全局与时序结构特征并用的方法进行了语音情感识别。Attabi等^[16]将锚模型的思想应用到了语音情感识别中,改进了识别系统的性能。Zheng等^[17]通过对传统的最小二乘回归算法进行改进,提出了不完全稀疏最小二乘回归算法,能同时对标记和未标记语音数据进行情感识别。Mao等^[18]通过使用卷积神经网络来选择对情感有显著影响的特征,取得了很好的效果。对于面部表情识别,Ekman等^[19]于1978年开发了面部动作编码系统(Facial Action Coding System, FACS)来检测面部表情的细微变化。Essa等^[20,21]于1997年提出了基于视频的动态表情描述方法FACS+,解决了FACS中没有时间描述信息的问题。Chakraborty等^[22]将模糊逻辑的思想应用到面部表情识别中,获得了较好的识别结果。Rahulamathavan等^[23]利用局部Fisher判别分析对加密面部表情信号进行了识别研究。中国科学技术大学的文沁等^[24]提出一种基于三维数据的人脸情感识别方法,给出了基于三维特征的眼角和嘴角新的提取算法。Zheng等^[25]提出了基于组稀疏降秩回归的多视角面部表情识别方法,能够从多尺度子域中自动选择出对情感识别贡献最大的子域。对于生理信号情感识

别, Petrantonakis 等^[26]采用高阶过零技术 (Higher order crossing, HOC) 提取脑电波信号中的情感信息来进行情感识别。西南大学的刘光远等^[27]从呼吸信号中提取特征参数进行情感识别。天津理工大学的苟立晨^[28]利用心电信号的特征值进行了快乐和悲伤两种情感的识别。山东大学的王艳超^[29]通过提取脉搏信号不同特征来分析和辨别不同情感的状态。Zacharatos 等^[30]分析研究了身体姿势和动作对情感识别的重要性。

虽然单一地依靠语音信号、面部表情信号和生理信号来进行情感识别的研究取得了一定的成果,但是存在着很多局限性,因为人类是通过多模式的方式表达情感信息的,它具有表达的复杂性和文化的相对性^[31]。比如,在噪声环境下,当某一个通道的特征受到干扰或缺失时,多模式方法能在某种程度上产生互补的效应,弥补单模式的不足,所以研究多模式情感识别的方法十分必要。如 Wagner 等^[32]通过融合肌动电流、心电、皮肤电阻和呼吸 4 个通道的生理参数,进行了多模式情感识别,获得了较高的融合识别率。Kim 等^[33]融合了肌动电流、心电、皮肤电导和呼吸 4 个通道的生理参数,并采用听音乐的方式来诱发情感,实现了对积极和消极两大类情感的高效识别。东南大学的赵力、黄程韦等^[34]通过融合语音信号与心电信号进行了多模式情感识别,获得了较高的融合识别率。Wen 等^[35]通过采用多变量相关法将皮肤电反应、心率、皮肤电反应的一阶导数等特征融合起来,实现了对生气、悲伤、恐惧和中性 4 种情感的高效识别。但是上述方法均为与生理信号相融合,而生理信号的测量必须与身体接触,因此对于此通道的信号获取有一定的困难,所以语音和面部表情作为两种最主要的表征情感的方式,得到了广泛的研究。如 Busso 等^[36]分析了单一的语音情感识别与人脸表情识别在识别性能上的互补性。Hoch 等^[37]通过融合语音与表情信息,在车载环境下进行了正面(愉快)、负面(愤怒)与平静等 3 种情感状态的识别。Sayedelahl 等^[38]通过加权线性组合的方式在决策层对音视频信息中的情感特征进行了融合识别。中科院的陶建华、黄力行等^[39]通过融合语音与脸部肌肉,提出了一种自适应权重的双模态情感识别方法。Meng 等^[40]通过融合语音和面部表情信号特征,研究了自然情感表达的连续性,提高了情感识别的准确率。

1.6 情感识别的研究意义及应用

情感识别是情感计算领域中最基础、最重要的研究内容之一。根据情感的外部表现特征方面的“表情”,情感识别的研究主要包括 3 个方面:基于言语

表情的语音情感识别、基于面部表情的人脸表情识别,以及融合多种表情信息的多模式情感识别。情感识别的研究具有极其重要的应用前景,其研究的最终目的就是设计出更加智能化、人性化的新型人机交互系统。例如:

(1) 在人机交互方面的应用

语音是人类最常用、最便捷的交流方式,利用语音进行自然、和谐的人机交互是人们一直以来的奋斗目标。语音情感识别应用于人机交互方面,将能有效地改变过去计算机呆板的交互服务,提高人机交互的亲切性和准确性。一个拥有情感能力的计算机能够对人类情感进行获取、分类、识别和响应,进而帮助使用者获得高效而又亲切的感觉,并能有效减轻人们使用电脑的挫败感,甚至能帮助人们理解自己和他人的情感世界。这正是计算机技术向人类社会全面渗透的重要手段。例如,可以采用此类技术探测司机精力是否集中,并做出相应的反应;还可以在汽车中用电脑测量驾驶者感受到的压力水平,以帮助解决驾驶者的所谓“道路狂暴症”问题。

在信息家电和智能仪器中增加自动感知人们情感状态的功能,可以提供更好的服务;在信息检索应用中,通过情感分析的概念解析功能,则可提高智能信息检索的精度和效率;在远程教育平台中,情感计算技术的应用能提升教学效果;利用多模式的情感交互技术,还可以构筑更贴近人们生活的智能空间或虚拟场景。此外,情感计算还能应用在机器人、智能玩具、游戏等相关产业中,以构筑更加拟人化的风格和更加逼真的场景。

(2) 在心理学方面的应用

从认知科学的角度看,情感反映了人的心理状态,是人类思维活动最生动的体现,所以对人类情感机理的研究与探索一直是科学研究的重要方向。情感计算理论通过计算机的分析和处理手段,将对情感的研究从感性认知上升为可计算模型,对在认知科学上探索大脑对信息的分析和处理机理、进一步加深对大脑中情感概念的解析和理解将具有重要的科学意义。

(3) 在医学方面的应用

现代医学研究表明,健康应该包含更多情感方面的内容,可是现代人中患有抑郁症的比例却在逐年上升,所以医学界希望能够更早地向好的方面诱导,从而有效地帮助人们对不良情绪进行排解。

(4) 在语音识别方面的应用

语音情感识别的研究可以辅助语音识别。试想,如果能够利用语音情感识别得到说话者的情感,判断出当时的语境,将会更好地进行语音识别。另外,语音识别在带有情感的语音中不能进行很好的识别,如果我们通过语音情感的研究消除或减弱情感的影响,将很好地帮助情感语音的识别。