

文章编号:1006-8309(2009)02-0064-03

语音情绪识别的应用和基础研究

林茵^{1,2},张侃¹

(1. 中国科学院心理研究所脑与认知科学国家重点实验室,北京 100101;

2. 中国科学院研究生院,北京 100049)

摘要:文章介绍了语音情绪识别的主要应用领域,着重分析了其中基础研究的关键问题和研究现状,并探讨了未来研究的方向。

关键词:语音;情绪识别;情感计算

中图分类号:B849;TP391.42

文献标识码:A

1 引言

作为表达个人信息的最明显和最直接的方式之一,语音是传递情绪的最精确和最具有多样性的媒介^[1]。从本质上说,从语音中识别出说话人的情绪状态是一个模式识别问题,需要情绪的分类,情绪语音的特征,以及这两者之间的对应关系。这是技术实现的基础。因此,本文在关注语音情绪识别技术应用价值的同时,主要总结了基础研究的关键问题以及未来研究的趋势。

2 语音情绪识别的应用

2.1 电话服务

由于只能依赖语音获取信息,电话服务是最需要语音情绪识别技术的领域之一。一方面,电话服务需要处理顾客的抱怨,而语音系统能自动地探测到这种情绪,及时提醒客服人员注意服务态度,提供相应的关怀和帮助。另一方面,电话自助服务在识别客户请求的同时,如果能够根据客户的情绪做出快速和适当的响应,则能使人机之间的交互更加自然,提高亲和力。例如,当机器检测到用户受挫或生气的时候,可以自动给予提示^[2]。

2.2 远程交流

在语音通讯工具上赋予情绪识别的功能,可以帮助通话双方及时了解彼此的情绪、促进交流。一款手机语音智能助手能结合语音情绪识别,分析出通话对象的喜爱程度、真诚度、疲倦度、欺骗程度和友好度^[3],在增加手机通话的娱

乐性的同时,突破了听觉信息单一性的限制。

类似的语音的情绪识别技术能追踪远程教学中学习者的情绪,当学习者因遇到难题或无法理解的材料而表现出焦虑和不满时,老师或系统可以调整教学方式和进度,或给予情绪调节指导^[4]。

2.3 驾驶系统

随着语音交互设备在车载系统的普及,语音情绪识别功能也开始用于监控情绪、保障驾驶安全。研究发现,积极的情绪可以促使驾驶者更好地驾驶,减少危险驾驶的行为。当驾驶者处于情绪不稳定的状态时,系统可以给予提醒,或者自动调节驾驶参数以防止事故的发生;系统也可以根据驾驶者的情绪提供关怀式的语音,提高驾驶体验^[5,6]。一项近期研发的系统能较准确地识别厌烦、愉悦、惊讶、受挫或愤怒、悲伤或痛苦这5类情绪,并且当车载系统发出的语音与驾驶者的情绪匹配时,驾驶的安全性会得到提高^[6]。

2.4 测谎与安检

人类有目的的行为始终包含着情绪的因素^[7]。因此,在犯罪测谎和安全监测活动中,利用罪犯特有的情绪特征可以更有效地识别出犯罪者。特别是当人在欺骗他人的时候往往处在应激状态,发音会发生改变^[8]。语音自动识别系统能辅助测谎仪根据话语的节律、音调、紧张度等检测出说话者由于隐瞒事实而产生的紧张、焦虑或者负罪感。

与此同时,安检系统则可以利用犯罪分子普

作者简介:林茵(1984 -),女,四川乐山人,硕士研究生,研究方向:工程心理学,(电话)010 - 64836047(电子邮箱)linh@psych.ac.cn。

遍拥有的另一情绪特征——害怕被捕。让受检者回答设定好的问题,计算机通过录音自动检测出这种情绪。这项技术虽不能取代现有的安检措施,但如果提高识别率,将能有效地提高安检的效率和检测率。

3 语音情绪识别的基础研究

虽然语音情绪识别有着如此重要的作用,但目前其实际应用范围还比较窄。若要发展语音情绪识别技术、提高其识别率和应用价值,还需要从基础研究问题入手。

3.1 情绪的分类

关于情绪的分类,研究者们并没有统一的标准。多数研究者认同人类具有6种基本情绪:快乐、兴趣、厌恶、恐惧、痛苦(悲伤)和愤怒^[7]。但语音的情绪识别研究大多采用的是中性、快乐、悲伤和愤怒这四种区分度大的情绪。也有研究者采用五种、六种或者八种情绪^[9,10]。

另一类情绪的分类方法是层级式的,即首先划分情绪的维度,再根据情绪在各维度上的情况确定出高层次的情绪。研究广泛采用PAD情绪三维度模型^[10,11]:愉悦度(Pleasure)表示情感状态的正、负特性(高兴的-不高兴的),激活度(Arousal)表示情感的神经生理激活水平(激活的-非激活的),优势度(Dominance)表示个体对情景和他人的控制状态(主控的-受控的)^[12]。

一项比较研究发现,不管采用哪种分类方法,最终的情绪识别率相近。其中对愤怒和悲伤这两种情绪的识别率最高,均在90%以上,而其他情绪的识别率则在50%左右^[2]。

3.2 情绪分析方法

从话语中识别出说话人的情绪可以从两方面进行分析:语言学分析和副语言学分析^[5,13]。前者指根据用词来提取说话者的情绪特征。如词语“惊讶”表达高激活度,“内疚”表达低愉悦度等。而副语言分析则针对话语内容以外的信息,如音高、强度等等。对两个分析过程的整合是一个难题。一般认为,副语言学分析能提供更可靠和有效的信息,而语言学分析存在混淆的结论。例如,同一个词语用不同的语调说出来则代表不同的情绪。因此,大多数系统只采用了副语言分析过程。

3.3 情绪语音的特征

心理学和语言学的研究发现,语音的韵律特

征受到情绪的影响最明显^[14]。目前,语音情绪识别常采用以下韵律特征:基频F0(均值、范围和最大最小值等)^[2,13,15]、能量(均值、范围和方差等)^[2,11,13]、语速(一个时间段内说出的字数)^[2,11]、和流利程度(韵律词之间的停顿时间)^[11]等。研究发现情绪的愉悦度与F0的均值、最大值、范围有显著的正相关,而激活度还与F0的最小值呈正相关^[15]。另有研究表明,愉悦和愤怒的F0均值高于其他情绪;愉悦的短时能量增强,语速变快;愤怒状态下的基频提高最多;悲哀的短时能量均值接近平静状态^[16]。除此以外还有音质特征也用于区分不同的情绪^[11,13]。

虽然特定的情绪下会表现出某些特定的声学特征,然而语音的情绪识别是一个逆命题,即从声学特征反推出情绪的类别。由于这些特征并不是和情绪一一对应的,并且特征是多样和多变的,选取合适的特征集是识别的关键^[11]。

3.4 情绪语音库

情绪语音库为识别提供训练集。样本越大,最后计算的结果越不易受个体和背景的影响。一般说来,情绪语音库中的语音片段主要通过两种方式获取。一种是由专业或业余演员在模拟真实情绪状态下朗读设计好的文字^[3,11]。这种方法容易控制,操作简单,但跟现实的情绪表达有差距。另一种是从广播、电视、电话、网络等媒体收集现实生活中人们的自然语言^[2]。但这种方式收集到的语音背景音复杂,处理起来难度大。比较发现:基于自然语音库的识别率偏低,但基于模拟语音库的识别率被夸大,对强烈的情绪识别率高于对中性情绪的识别率^[2]。

4 未来研究方向

4.1 发展情绪理论研究

目前的研究多数还停留在基本情绪的层面,较少涉及复合情绪,比如焦虑、爱等^[7]。如何定义和区分这些情绪应当是未来研究的重点。另外,不同领域针对的情绪也有所不同。因此,应根据系统的应用场景,有针对性地研究。这可以从两方面进行。一是收集真实场景下人群的声音数据,进行情绪的评定和归类,二是利用模拟场景诱发相应的情绪,比如用户在软件使用过程中产生的愤怒、挫败感等。对情绪的归类可以采用多种方法,并结合特征分析结果选取最有效的方式。

4.2 提高情绪语音的特征分析

关于情绪的声学特征尚没有明确的界定。在众多特征中,哪些才是与情绪的表现最相关并且是最不易受其他因素干扰的呢?除了改进特征的计算方法以外,单纯对韵律特征的重视对于语音情绪识别是远远不够的,还应当考虑到语音的其他文化、社会环境等因素。

目前多数研究还停留在对低层次声学属性的研究,而对高层次的语音结构和内容的语言分析还很少。作为一种交流方式,声音是手段、语法是过程、内容是目的。对后两方面的分析应当是未来语音情绪识别的努力方向之一。目前的技术已经可以做到对话语的单词进行识别,但是还需要从这些识别出的字句提取说话者的情绪。比如,人在愤怒的情绪下可能会使用更多的带有攻击性和侮辱性的词语;在紧张状态下,可能出现更多语法错误。不过,对于整合这几方面的特征还没有比较好的方法。一些研究根据语言学分析的结果对副语言学分析过程进行修正,已取得了一定成效^[13]。

4.3 建立全面的情绪语音库

不少语音库是为情绪的语音合成而建立的,但合成只需要找到一种情绪表现的特例就足够了,识别却需要尽可能考虑到一种情绪的不同表达方式。目前各种情绪语音库一般只包括典型情绪,没有考虑到各种变化,所包含的语音片段数量从几十到几千不一,因此不足以获得可靠的情绪识别结果^[17]。建立全面的情绪语音库可以提高识别率,但也为情绪的分类研究提出了更高要求。

4.4 结合多通道信息研究

越来越多的人机交互系统开始利用多通道的方式,通过麦克风、摄像头、传感器等对来自语音、面部表情和姿势的情绪信息进行整合^[5],可以提高情绪识别灵敏度和准确性^[18]。各通道的分析过程可以各自独立得出结果,再进行整合;也可以对各通道的特征同时进行计算分析,最后得出统一的结论。Busso 对声音和面部特征的双通道研究发现,这两种分析方式的整体识别率基本相同,但对不同情绪的敏感度不同^[18]。由于多通道的方式需要各种设备,如何扩大这种多通道系统的应用范围也是值得研究的课题。

5 结束语

情绪的语音识别可以提高人机交互的效用和友好性、促进身份识别和检测。作为情感计算的热门课题,这项技术还需要各项理论研究的支撑。结合心理学、语言学和声学等各领域的研究,可以促进该技术发展和应用。

参考文献

- [1] Harrigan J A, Wilson K, Rosenthal R. Detecting State and Trait Anxiety from Auditory and Visual Cues: A Meta - Analysis [J]. *Personality and Social Psychology Bulletin*, 2004, 30(1):56 - 66.
- [2] Morrison D, Wang R, De Silva L C. Ensemble Methods for Spoken Emotion Recognition in Call - Centers [J]. *Speech Communication*, 2007, 49(2):98 - 112.
- [3] Yoon WJ, Cho YH, Park KS. A Study of Speech Emotion Recognition and Its Application to Mobile Services [C] // *Ubiquitous Intelligence and Computing*, 4th International Conference (UIC'07). Hong Kong, 2007.
- [4] Chen K, Yue G, Yu F, et al. Research on Speech Emotion Recognition System in E - Learning [J]. *Lecture Notes in Computer Science*, 2007, 4489(1): 555 - 558.
- [5] Balomenos T, Raouzaoui A, Karpouzis K, et al. An Introduction to Emotionally Rich Man - Machine Intelligent Systems [C] // *Third European Symposium on Intelligent Technologies, Hybrid Systems and Their Implementation on Smart Adaptive Systems (EU-NITE'03)*. Finland, 2003.
- [6] Jones C, Jonsson IM. Performance Analysis of Acoustic Emotion Recognition for In - Car Conversational Interfaces [J]. *Lecture Notes in Computer Science*. 2007, 4555(6): 411 - 420.
- [7] 孟昭兰. 情绪心理学 [M]. 北京大学出版社: 2005:12 - 16.
- [8] Hansen J H L, Patil S. Speech Under Stress: Analysis, Modeling and Recognition [M] // Muller C. *Speaker Classification*. Berlin: Springer - Verlag. 2007:108 - 137.
- [9] Tao J, Kang Y, Li A. Prosody Conversion From Neutral Speech to Emotional Speech [J]. *Audio, Speech and Language Processing*, 2006, 14(4): 1145 - 1154.
- [10] Yang H, Meng H M, Cai L. Modeling the Acoustic Correlates of Expressive Elements in Text Genres for Expressive Text - to - Speech Synthesis [C] // *Ninth International Conference on Spoken Language Processing*. USA: ISCA, 2006.

(下转封三)

(上接第 72 页)

- [5] Neuman J H, Baron R A. Workplace Violence and Workplace Aggression: Evidence Concerning Specific Forms, Potential Causes and Preferred Targets [J]. *Journal of Management*, 1998, 24(3): 391 - 419.
- [6] Chen P Y, Spector P E. Relationships of Work Stressors with Aggression, Withdrawal, Theft and Substance Use: an Exploratory Study [J]. *Journal of Occupational and Organizational Psychology*, 1992, 65(3): 177 - 184.
- [7] Fox S, Spector P E. A Model of Work Frustration - aggression [J]. *Journal of Organizational Behavior*, 1999, 20(6): 915 - 931.
- [8] Vivian C S Lau, Wing Tung Au, Jane M C Ho. A Qualitative and Quantitative Review of Antecedents of Counterproductive Behavior in Organizations [J]. *Journal of Business and Psychology*; 2003, 18(1): 73 - 94.
- [9] Judge TA, Scott B A, Ilies R. Hostility, Job Attitudes, and Workplace Deviance: Test of a Multi - level Model [J]. *Journal of Applied Psychology*, 2006, 91(1): 126 - 138.
- [10] Reeshad S Dalal. A Meta - analysis of the Relationship between Organizational Citizenship Behavior and Counterproductive Work Behavior [J]. *Journal of Applied Psychology*, 2005, 90(6): 1241 - 1255.
- [11] Skarlicki, Folger R. Retaliation in the Workplace: The Roles of Distributive, Procedural and Interactional Justice [J]. *Journal of Applied Psychology*, 1997, 82(3): 434 - 443.
- [12] Skarlicki D P, Folger R, Tesluk P. Personality as a Moderator in the Relationship between Fairness and Retaliation [J]. *Academy of Management Journal*, 1999, 42(1): 100 - 108.
- [13] Fox S, Spector P E, Don M. Counterproductive Work Behavior (CWB) in Response to Job Stressors and Organizational Justice: Some Mediator and Moderator Tests for Autonomy and Emotions [J]. *Journal of Vocational Behavior*, 2001, 59(3): 291 - 309.
- [14] Marcus B, Schuler H. Antecedents of Counterproductive Behavior at Work: A General Perspective [J]. *Journal of Applied Psychology*, 2002, 89(4): 647 - 660.
- [15] Spector P E. Organizational Frustration: a Model and Review of the Literature [J]. *Personnel Psychology*, 1978, 31(4): 815 - 829.
- [16] Rotundo M, Sackett P R. The Relative Importance of Task, Citizenship, and Counterproductive Performance to Global Ratings of Job Performance: A Policy - capturing Approach [J]. *Journal of Applied Psychology*, 2002, 87(1): 66 - 80.

[收稿日期]2008 - 06 - 03

[修回日期]2008 - 08 - 10

(上接第 66 页)

- [11] Pao T L, Chen Y T, Yeh J H, et al. Detecting Emotions in Mandarin Speech [J]. *Computational Linguistics and Chinese Language Processing*, 2005, 10(3): 347 - 361.
- [12] Osgood C E, Suci G J, Tannenbaum P H. The measurement of meaning [M]. Urbana: University of Illinois Press, 1957: 31 - 76.
- [13] Fotinea SE, Bakamidis S, Athanaselis T, et al. Emotion in Speech: Towards an Integration of Linguistic, Paralinguistic, and Psychological Analysis [C] // *Artificial Neural Networks and Neural Information Processing - ICANN 2003*. Istanbul, 2003: 1125 - 1132.
- [14] 林奕琳, 韦岗, 杨康才. 语音情感识别的研究进展 [J]. *电路与系统学报*, 2007, 12(1): 90 - 99.
- [15] Jin X, Wang Z. An Emotion Space Model for Recognition of Emotions in Spoken Chinese [C] // Tao J, Tan T N, Picard R W. *Affective Computing and Intelligent Interaction*. Beijing: Yan Shan Press, 2005. 397 - 402.
- [16] 高慧, 苏广川, 陈善广. 不同情绪状态下汉语语音的声学特征分析 [J]. *航天医学与医学工程*, 2005, 18(5): 350 - 354 .
- [17] Ververidis D, Kotropoulos C. Emotional Speech Recognition: resources, features, and methods [J]. *Speech Communication*, 2006, 48(9): 1162 - 1181.
- [18] Busso C, Deng Z, Yildirim S, et al. Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information [C] // *Proceedings of the 6th International Conference on Multimodal Interfaces*. NY: ACM Press, 2004: 205 - 211.

[收稿日期]2008 - 04 - 19

[修回日期]2008 - 06 - 13