

基于 PAD 情绪模型的情感语音识别

宋 静, 张雪英, 孙 颖, 张 卫

(太原理工大学 信息与工程学院, 山西 太原 030024)

摘 要: 简述梅尔频率倒谱系数、线性预测系数、韵律学特征、共振峰频率和过零峰值幅度特征, 并将这五种语音特征应用于情感语音识别. 根据识别结果从 PAD 情绪模型的三个维度进行相关性分析得到特征的权重系数, 并将识别结果融合映射到 PAD 三维情绪空间, 最终获得情感语音的 PAD 值. 利用情感语音的 PAD 值可以从连续情感理论对情感语音进行描述分析, 采用量化的方法揭示情感空间中各种情绪范畴的定位和关系.

关键词: 语音特征; 情感语音识别; PAD 情绪模型; 相关性分析

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2016)09-0128-04

Emotional Speech Recognition Based on PAD Emotion Model

SONG Jing, ZHANG Xue-ying, SUN Ying, ZHANG Wei

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: Five approaches of feature extraction: the MEL-frequency Cepstral Coefficient (MFCC), the Linear Predictor Coefficient (LPC), prosodic features, formant frequency and the Zero Crossings with Peak Amplitudes (ZCPA) are described in this paper. These features are applied to emotional speech recognition. According to the recognition results, the weight coefficients of features are obtained by correlation analysis in the three dimensions of PAD emotion model. Simultaneously the recognition results are fused to the PAD emotional space, and the PAD values of the emotional speech are obtained. The PAD values of the emotional speech can be analyzed from the theory of continuous emotion. And the quantitative analysis of emotional speech can reveal the position and relationship of emotional category in emotional space.

Key words: speech feature; emotional speech recognition; PAD emotion model; correlation analysis

1 引言

语言是人类信息交换最方便快捷的一种方式, 现代科技的飞速发展为实现人机对话、智能机器的情感交流提供可能^[1]. 语音信号中不仅仅包含着人类所要表达的文本信息, 还饱含着丰富的情感内容. 情感是人类对外界事物刺激而产生的一种心理反应. 目前, 心理学研究者已经对情感进行了大量的研究, 主要是从两个方面对情感进行描述: 一个是经典的离散情感标签, 另外一个连续维度理论^[2]. 目前情感语音识别结果多为离散情感, 离散情感标签描

述方法虽然能够清晰、简洁地定义情感状态, 但是不符合人类复杂的心理活动状态, 无法展现出情感间的相对关系以及变化, 也难以表示出混合情感的状况. 因此, 本文利用连续维度理论进行情感语音识别的研究, 对情感语音的 PAD 三个维度值进行准确的预测, 探索情感状态间的变化与联系.

2 情感语音数据库

目前, 在国内外语音研究领域, 由于研究需求、获取途径、情感分类、语种等存在差异, 所以情感语音数据库种类繁多. 本文选用的情感语音数据库为

收稿日期: 2015-12-10; **修回日期:** 2016-01-28

基金项目: 国家自然科学基金项目(61371193); 山西省青年科技研究基金项目(2013021016-2); 山西省回国留学人员科研资助项目(2013-034); 山西省研究生创新项目(2015BY24)

摘引型情感语音数据库 TYUT2. 0. 其具体建库流程如图 1 所示.

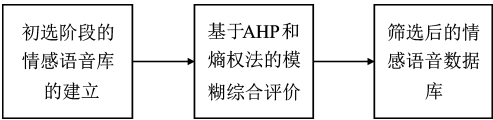


图 1 情感语音数据库流程图

根据本实验室的研究需求,通过截取广播剧的方式获得了包含高兴、惊奇、悲伤、愤怒 4 种基本情感类型的初选阶段的情感语音数据库.

通过模糊综合评价结合层次分析法(AHP)、熵权法^[3]建立模糊综合评价模型,辨听筛选初选阶段的情感语音数据库,利用该综合评价模型对初选阶段的数据库从情感准确度、背景噪声影响、自然度等角度进行评价,将定性分析转化为定量分析,根据模糊综合评价模型得到语音的综合得分来判断情感语音的优劣,筛选出真实、自然、可靠的情感语音库.该语音库为情感语音特征、情感语音识别以及合成奠定了基础,也为实验室的脑电研究提供了实验材料.

3 情感语音特征

与情感语音识别相关的声学特征大致可分为基于谱的相关特征、韵律学特征和音质特征、基于人耳听觉模型的特征、融合特征^[3].

3.1 基于谱的相关特征

随着对基于谱的相关特征研究的深入,研究者发现语音中的情感信息在各个频谱区间分布有着明显的变化.因此越来越多的研究者将基于谱的相关特征应用到情感语音识别中来.在语音情感识别系统中使用的线性谱特征一般有:LPC(Linear Predictor Coefficient),OSALPC(One Sided Autocorrelation Linear Predictor Coefficient)等;倒谱特征一般有:LPCC(Linear Predictor Cepstral Coefficient),MFCC(Mel-Frequency Cepstral Coefficient)等.

3.1.1 MFCC

梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficient,MFCC)是依据人耳听觉特性提取到的特征参数.MFCC是从对频率高低的非线性心理感觉这个角度来反映语音短时幅度谱的特征^[4].MFCC是将信号频谱的频率轴转变成梅尔刻度,然后转换到倒谱域得到.提取MFCC时,首先利用短时傅里叶变换得到语音信号的频谱,然后计算其能量谱,将能量谱通过梅尔滤波器组,即一组三角带通滤波器.该滤波器的中心频率是按照梅尔频率刻度均

匀分布的.接着,对每个滤波器的输出取对数,最后进行傅里叶逆变换即可得到MFCC.

提取MFCC特征时用到的三角带通滤波器在中高频区域的分布比较稀疏,而低频区域比较密集,所以MFCC在中高频的频率分辨率较低,可以屏蔽高频噪声干扰;在低频区内的计算精度良好,其频率分辨率也较高.由于语音信号主要集中在低频部分,因此MFCC特征可以较好地描述语音信号的特征^[5].

3.1.2 线性预测(LPC)系数

线性预测分析基于人的发声机理,通过对声道的短管级联模型的研究,可以认为系统的传递函数符合全极点数字滤波器的形式.因此可以利用前若干时刻的信号的线性组合对某个时刻的信号进行估计.通过使实际语音的采样值和线性预测采样值之间达到最小均方误差,即可得到线性预测系数LPC^[9].

3.2 韵律学特征

韵律特征描述了语音在音高、声调、快慢和重音等方面的变化,又被称为超音段特征.

本文提取了语音的常用韵律学特征包括能量、基音频率、共振峰、过零率、语速的统计参数共 14 维作为情感语音识别的基本特征参数.如表 1 所示.

表 1 语音的韵律学特征及其统计参数

韵律学特征	统计参数
短时能量	最大值,最小值,平均值,变化率的最大值、最小值、平均值
基音频率	最大值,最小值,平均值,变化率的最大值、最小值、平均值
过零率	平均过零率
语速	平均语速

3.3 音质特征

声音质量主要用于衡量语音的纯净度、清晰度以及辨识度等.喘息、颤音、哽咽等声学表现对声音质量会产生影响,而且往往出现在说话人情绪激动、心情难以平复的情形下.在情感语音识别中常用的音质特征有:共振峰频率及其带宽、频率微扰和振幅微扰、声门参数等.

本文提取的声音质量特征为共振峰频率特征,包括第 1~3 共振峰的最大值、最小值、平均值、方差、中值,变化率的最大值、最小值、平均值共 24 维.

3.4 基于人耳听觉模型的特征—ZCPA

本文提取的语音韵律学特征为过零峰值幅度特征(Zero Crossings with Peak Amplitudes,ZCPA).

ZCPA 是 1999 年由 Doh-Suk Kim 根据人耳听觉模型提出的^[7]. ZCPA 特征的提取框图如图 2 所示, 整个提取过程由带通滤波器组、过零检测器、非线性压缩和频率接收器组成^[8]. 具体过程如下:

- (1)语音信号经过 16 个 FIR 带通滤波器, 用来模拟耳蜗基底膜;
- (2)利用过零检测器、峰值检测器、非线性压缩部分仿真听觉神经纤维;
- (3)将过零检测器得到的频率信息和峰值检测器得到的强度信息经过非线性压缩后, 利用频率接收器进行时间幅度规整, 得到最终的语音信号的特征.

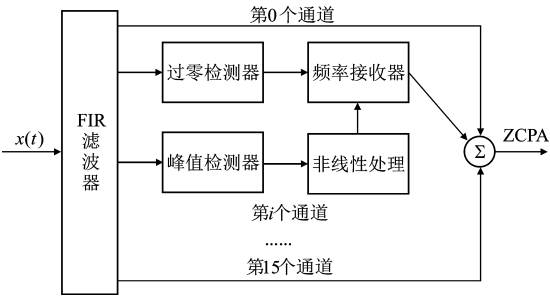


图 2 ZCPA 特征的提取框图

4 基于 PAD 三维情绪模型的情感语音识别

对情感语音进行特征提取, 得到五种特征: MF-CC、LPC、韵律特征、共振峰频率、ZCPA; 分别将五种特征输入 SVM 分类器, 得到该语音的五种判别结果, 之后分别将判别结果利用 PAD 量表映射到三维情感空间中, 根据不同特征的重要性确定权值, 最终融合得到该语音的加权后在三维情感空间中的 PAD 值. 基于 PAD 三维情绪模型的多特征融合模型如图 3 所示.

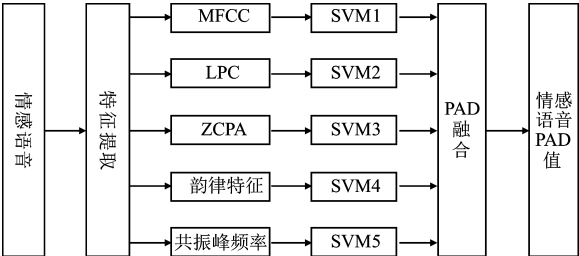


图 3 基于 PAD 三维情绪模型的多特征融合模型

4.1 PAD 三维情绪模型

PAD 三维情绪模型属于连续维度理论的一种. Mehrabian 和 Russell 根据语义差异的评价方法^[9], 定义情绪的三个维度: 愉悦度 (pleasure-displeas-

ure)、激活度 (arousal-nonarousal)、优势度 (dominance-submissiveness). 其中, 愉悦度代表个体情感状态的正负, 激活度代表个体的神经生理激活程度, 优势度体现个体对情境和他人的控制状态. PAD 三维情绪模型是对情感空间的理论描述, 建立了情感空间中不同情绪范畴的定位和关系, 提供了情感之间定量分析的途径^[10]. 目前, PAD 三维情绪模型已经被广泛的应用.

李晓明等人利用编订的中文版 PAD 情绪量表对 14 个基本情绪范畴的 PAD 值进行评定^[11]. 其中包括不同研究人员提出的基本情绪, 如高兴、悲伤、轻松、恐惧等. 本文采用语音库 TYUT2.0 中的高兴、惊奇、生气、悲伤 4 种基本情感所对应的 PAD 值进行识别结果的相关性分析, 4 种情感所对应的 PAD 值如表 2 所示.

表 2 选取的 4 种情感的 PAD 值

情感类型	平均值		
	P	A	D
高兴	2.77	1.21	1.42
愤怒	-1.98	1.10	0.60
悲伤	-0.89	0.17	-0.70
惊奇	1.72	1.71	0.22

4.2 特征权值计算

为了确定不同特征在 SVM 中识别结果的可信度, 提出一种根据识别结果进行特征加权的信息融合计算. 本文将机器识别的情感与其真实情感各自所对应的三个 PAD 值做相关性分析, 再用其结果去确定不同特征权重, 进而得到融合后的 PAD 三个维度值.

4.2.1 Pearson 相关系数

本文选择 Pearson 相关方法进行相关性分析. 总体相关系数用 ρ 表示, 一般 ρ 是未知的, 要根据样本相关系数 r 作为 ρ 的近似估计值. 由于样本取值不同, r 取值不同, 是一个随机变量, 故需要考察样本相关系数的可靠性, 进行显著性检验^[12].

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

上式为 Pearson 相关系数 r 的计算公式. 式中, $\bar{x}, \bar{y}$ 分别是变量 x, y 的均值; x_i, y_i 分别是变量 x, y

的第*i*个观测值;*r*为样本相关系数($-1 \leq r \leq 1$),表示两变量*x,y*线性相关程度和方向.*r*绝对值越接近 1,则两变量线性相关程度越大, $r > 0$ 表示正相关,*x,y*变化的方向一致; $r < 0$ 表示负相关,*x,y*变化的方向相反.

表 3 情感语音特征的相关系数结果

声学特征		PAD 值					
		愉悦度	归一化	优势度	归一化	激活度	归一化
基于谱的相关特征	MFCC	0.612	0.319	0.704	0.316	0.588	0.271
	LPC	0.317	0.165	0.302	0.136	0.534	0.247
韵律学特征	韵律特征	0.191	0.100	0.334	0.150	0.182	0.084
音质特征	共振峰频率	0.308	0.161	0.411	0.184	0.455	0.210
基于人耳听觉模型特征	ZCPA	0.490	0.255	0.477	0.214	0.407	0.188

对基于谱的特征(MFCC、LPC)、韵律学特征、音质特征(共振峰频率)、基于人耳听觉模型特征(ZCPA)四类情感特征的识别结果进行相关性分析,从 PAD 情绪模型的三个维度上分别得出了它们对应的 PAD 相关系数.分别从 PAD 三个维度(愉悦度、优势度、激活度)对比语音特征的相关系数由大到小进行排列. P(愉悦度):MFCC、ZCPA、LPC、共振峰频率、韵律特征; A(优势度):MFCC、ZCPA、共振峰频率、韵律特征、LPC; D(激活度):MFCC、LPC、共振峰频率、ZCPA、韵律特征.

4.3 加权融合预测 PAD 值

利用 4.2.2 节中得到的五种特征归一化后的相关系数实现 PAD 值的重新预测计算,将 PAD 相关系数与各自识别结果对应的 PAD 值加权融合得到该语音的 PAD 三个维度值.以某条语音的愉悦度为例,SVM 可能的决策为高兴、生气、悲伤、惊奇,用五种语音特征分别对该语音进行情感判别,采用 PAD 量表对判别出的情感独立进行三维空间的映射,将映射结果与对应的相关系数进行加权融合,

$$P = m_{MFCC}P_{MFCC} + m_{LPC}P_{LPC} + m_{ZCPA}P_{ZCPA} + m_{韵律特征}P_{韵律特征} + m_{共振峰频率}P_{共振峰频率}$$

得到该语音在三维情感空间的愉悦度值.

5 结束语

本文利用基于谱的特征(MFCC、LPC、)、韵律学特征、音质特征(共振峰频率)、基于人耳听觉模型的特征(ZCPA)四大类语音特征的识别结果进行情感语音在 PAD 三维情感空间的映射,同时分别将

4.2.2 识别结果的相关性分析

选择 Pearson 相关方法将机器识别出的情感所对应的 PAD 值与语音本身的情感对应的 PAD 值进行相关性分析,得出的五种特征的相关系数结果对比如表 3 所示.

五种特征的机器识别结果所对应的 PAD 值与语音真实情感所对应的 PAD 值进行相关性分析,根据相关系数确定不同特征的权值,最后加权融合得到语音在连续维度角度的三维数值.将难以量化的情感语音映射到三维情感空间,可以直观精确地展现出语音所包含的情感内容.

在计算 TYUT2.0 数据库语料的 PAD 值时,本文发现绝大多数语料的 PAD 值并不与基本情感的 PAD 值相对应,而是在基本情感点附近离散分布,因此还需要利用情感语音在愉悦度、激活度、优势度上的空间分布,进一步分析语音的情感状态由哪些情感状态构成,也可为进一步分析情感识别的误判原因提供基础.

参考文献:

[1] 张雪英. 数字语音处理及 Matlab 仿真[M]. 北京:电子工业出版社,2010:1-9.

[2] 赵力,黄程韦. 实用语音情感识别中的若干关键技术[J]. 数据采集与处理,2014,29(2):157-170.

[3] 韩文静,李海峰,阮华斌,等. 语音情感识别研究进展综述[J]. 软件学报,2014,25(1):37-50.

[4] Deshmukh A A, Pramod K M. Combined LPC and MFCC features based technique for isolated speech recognition[C]//Proc of ISCAS 2013. Beijing: IEEE, 2013:437-441.

[5] 孙颖,张雪英. 情感语音特征对语料库依赖性的统计分析[J]. 噪声与振动控制,2011,31(4):132-136.

[6] 吴炜烨. 基于神经网络语音识别算法的研究[D]. 长沙:中南大学,2009.

定动态路由中各标准消息的转换模型,构建相应的消息转换目标函数,通过粒子群算法和蚁群算法对消息转换模型的参数值进行修正,使误差达到最小值,保证转换前后消息的一致性.实验结果表明,所提方法具有很高的转换精度,且转换前后消息的无关性较低,转换效率高.

参考文献:

[1] 苑迎,王翠荣,王聪,等.基于DPSO负载可控的虚拟网络映射算法[J].东北大学学报(自然科学版),2014,35(1):10-14.

[2] 蒋建峰.PSN网络中基于社会属性的消息路由[J].计算机系统应用,2014,23(11):165-168.

[3] 吴高峰,丁君辉,徐远兵.基于内容的ESB消息路由机制[J].计算机系统应用,2015,24(1):139-142.

[4] 常磊,顾华玺,张之义,等.一种粒子群优化的用户优先级虚拟网络映射算法[J].西安电子科技大学学报(自然科学版),2015,42(1):16-22.

[5] 王文钊,汪斌强,王志明,等.基于混合群智能优化的虚拟网络映射算法[J].计算机应用,2014,34(4):930-934.

(上接第131页)

[7] 孙慧.粒子群优化模糊神经网络用于语音识别的研究[D].太原:太原理工大学,2010.

[8] Ying Sun, Xueying Zhang. A study of zero-crossings with peak-amplitudes in speech emotion classification [C]//The First International Conference on Pervasive Computing, Signal Processing and Applications. Harbin, China:IEEE,2010:328-331.

[9] Mehrabian A. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in Temperament[J]. Current Psychology, 1996,14(4):261-292.

[10] Backs R W, da Silva S P, Han K. A comparison of younger and older adults' self-assessment manikin ratings of affective pictures[J]. Experimental Aging Research,2005,31(4):421-440.

[11] Li X,Zhou H,Song S,et al. The reliability and validity of the chinese version of abbreviated PAD emotion

[6] 周军海,林亚平,周四望.一种低功耗的社区机会网络消息路由算法[J].计算机科学,2014,41(1):178-182.

[7] 叶晨,支海邦.物联网中一种基于内容的消息路由算法[J].小型微型计算机系统,2015,36(9):2026-2030.

[8] 夏汉铸,崔晓燕.无线 Mesh 网络中基于路由信息的拥塞控制策略研究[J].测控技术,2015,34(3):101-104.

[9] 龚丁海,覃少华.消息副本动态感知的机会网络路由算法[J].科学技术与工程,2014,14(27):256-259.

[10] 孔飞,吴学文,宋亚男.一种移动 Ad Hoc 网络路由和数据安全方案设计[J].电子设计工程,2014,22(9):142-144.

作者简介:

董振华 女,(1980-),硕士,讲师.研究方向为数据库、计算机网络技术. E-mail:zhenhuadong1230@163.com.

张开便 女,(1983-),硕士,讲师.研究方向为软件工程、计算机网络技术.

李喜艳 女,(1983-),硕士,讲师.研究方向为信息安全、图像处理.

刘佳 女,(1970-),工学博士,副教授.研究方向为计算机应用.

scales[C]// Proceedings of the International Conference on Affective Computing and Intelligent Interaction. Beijing:IEEE,2005:513-518.

[12] Caglar Oflazoglu,Serdar Yildirim. Recognizing emotion from turkish speech using acoustic features[J]. EUR-ASIP Journal on Audio, Speech and Music Processin, 2013,20013(26):1-11.

作者简介:

宋静 女,(1989-),硕士研究生.研究方向为情感语音识别. E-mail:tysongj@163.com.

张雪英 女,(1964-),教授.研究方向为语音信号处理.

孙颖 女,(1981-),讲师.研究方向为语音信号特征参数提取.

张卫 女,(1986-),博士.研究方向为语音信号处理、情感语音识别.