第 38 卷 第 2 期 计 算 机 工 程 2012 年 1 月
Vol.38 No.2 Computer Engineering January 2012

·人工智能及识别技术· 文章编号：1000—3428(2012)02—0169—03 文献标识码：A 中图分类号：TP37

语音情感中基于 ZCPA 的 VAP 模型

秦宇强[1,2]，张雪英[1]

(1. 太原理工大学信息工程学院，太原 030024；2. 太原科技大学经济与管理学院，太原 030024)

摘 要：分析一个基于心理学的情感空间模型原理。研究语音情感识别中 7 种情感(中性、喜悦、愤怒、惊讶、恐惧、悲伤和厌恶)的效价-激励-能量(VAP)维分布状况，根据过零峰值幅度(ZCPA)的最大值、最小值、均值和绝对值方差和，在 VAP 三维空间中分析维数水平和 ZCPA 韵律特征之间的关系。实验结果表明，该情感空间模型原理有助于描述和区分各种语音情感。

关键词：语音情感识别；效价维；激励维；能量维；过零峰值幅度

ZCPA-based VAP Model in Speech Emotion

QIN Yu-qiang[1,2], ZHANG Xue-ying[1]

(1. College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China;
2. College of Economics and Management, Taiyuan University of Science and Technology, Taiyuan 030024, China)

【Abstract】This paper presents a conception of emotion space modeling using psychological research for reference. Based on this conception, this paper studies the Valence-Arousal-Power(VAP) distribution of the seven emotions for speech emotional recognition, including joy, anger, surprise, fear, disgust, sadness and neutral, in the three dimensional space of VAP, and analyses the relationship between the dimensional ratings and the Zero Crossings with Peak Amplitudes(ZCPA) prosodic characteristics in terms of maximum, minimum, mean and absolute square difference sum of ZCPA. Experimental results show that the conception of emotion modeling is helpful to describe and distinguish speech emotions.

【Key words】speech emotional recognition; valence dimension; arousal dimension; power dimension; Zero Crossings with Peak Amplitudes (ZCPA)

DOI: 10.3969/j.issn.1000-3428.2012.02.055

1 概述

近年来在人-机交流中，为了使机器更好地理解人类，从人类语言中识别情感得到了广泛的关注。一些学者提出了大量的语音情感识别的方法，但大多只关心将语音情感划分为一些情感状态[1]。这些情感状态的定义至今没有一个统一的标准。其实精确地把语音情感划分为某种特定的情感状态是一件非常困难的事。如果这些情感能够根据类似于颜色水平位置指示器(HSI)进行建模，并且使用一些数学的方法进行描述的话，那么语音情感识别的问题就迎刃而解了。在情感语音中，把人类听觉系统过零峰值幅度(Zero Crossings with Peak Amplitudes, ZCPA)模型作为情感特征进行情感识别，是一个新颖且热点的研究课题，但这些研究仅停留在对单一的一维情感进行识别，而且其识别率有时并不理想[2]。

本文基于心理学研究的结果提出了情感结构模型的原理，并在实验中将语音信号中的 7 种情感(包括中性、喜悦、愤怒、惊讶、恐惧、悲伤和厌恶)分布在效价-激励-能量(Valence-Arousal-Power, VAP)三维情感空间中。此外，还研究 ZCPA 韵律特征与 VAP 三维空间中的维数水平之间的关系。

2 基于心理学的情感空间模型

根据相关心理学研究[3-4]，情感计算可以定义为一个三维空间，即效价维(V)、激励维(A)和能量维(P)。效价维用来判断情感是激动(心潮澎湃)的还是冷漠(无动于衷)的；激励维用来判断情感是正面的还是负面的；而能量维是用来判断情感的力度、控制程度和控制感的。根据 Plutchik 情感空间模型和 HSI 颜色水平位置指示器模型，可以建立起一个有效的三维情感空间模型，如图 1 所示。

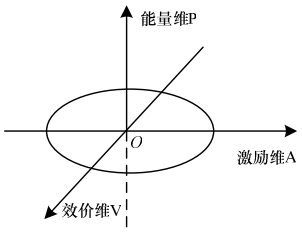


图 1 三维情感空间

3 基于人类听觉系统的 ZCPA 模型基本原理

3.1 人类听觉系统简介

人类听觉系统包括听觉外周和听觉中枢。听觉外周由外耳、中耳和耳蜗组成。声波经外耳、中耳和听小骨传至耳蜗，耳蜗将声波的机械能转换为神经编码信号，继而通过听神经将该信号传至大脑皮层的听觉中枢，最终产生听觉。

在上述过程中，耳蜗负责声音信号分析的绝大部分工作，它主要包括卵圆窗、基底膜、柯蒂氏器官和听神经。其中，卵圆窗接收听小骨的振动信号并使基底膜产生行波振动；基底膜是一片浸浴在淋巴液内的薄膜，其宽度从底部至顶部逐渐增大，弹性系数和阻尼也随之改变，从而使得不同频率的声波在基底膜不同位置处产生最大波峰，因此，基底膜具有带通滤波能力；基底膜的振动会以速度激励方式刺激柯蒂

基金项目：山西省自然科学基金资助项目(2010011020-1)；山西省国际科技合作基金资助项目(2011081047)

作者简介：秦宇强(1976－)，男，讲师、博士，主研方向：情绪语音识别；张雪英，教授、博士生导师

收稿日期：2011-07-12 **E-mail：**qinyuqiang@126.com

氏器官中的内毛细胞，由其完成振动刺激到电刺激的能量转换，并将电刺激信号导入与之相连的感音神经元；在感音神经元和听觉中枢之间，由侧抑制神经网络完成对电刺激信号的特征提取；内毛细胞对振动刺激的响应有半波整流特性(即内毛细胞只对刺激信号波形中“正”的部分产生响应)和非线性饱和特性(即当刺激强度达到一定水平时电位发生饱和)[5-6]。

3.2 基于听觉 ZCPA 模型的基本原理

ZCPA 模型是一种计算方法较为简洁的听觉模型，已有研究表明，该模型在语音识别中具有很高的识别准确率和抗噪声干扰能力。ZCPA 模型分析和处理信号的思路和方法与常规信号分析方法也存在明显区别，其实现过程如图 2 所示，大体包括基底膜带通滤波、内毛细胞与听神经特征提取和特征信息综合等 3 个主要步骤[7-8]。

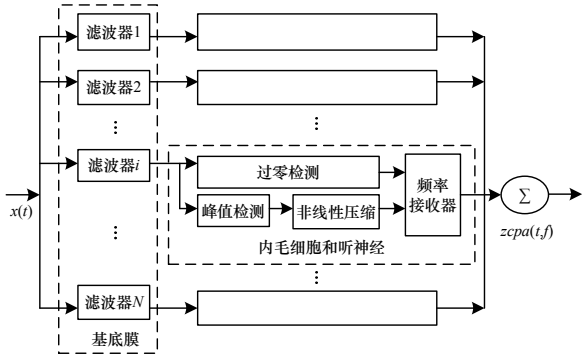


图 2 基于人类听觉系统的 ZCPA 模型原理

3.2.1 基底膜带通滤波

ZCPA 模型使用带通滤波器组模拟基底膜的频率分解特性。设共有 N 个滤波器，第 i 个滤波器为 $h(t,i)$，$h(t,i)$的中心频率由 i 确定，则基底膜对信号 $x(t)$的响应为：

$$y_1(t,i)=x(t)*_t h(t,i) \tag{1}$$

其中，$y_1(t,i)$为基底膜输出；$*_t$表示时域卷积。

对于滤波器 $h(t,i)$的设计，均以实际基底膜的滤波特性为依据，以两者的频率特性曲线尽量吻合为目标。

3.2.2 内毛细胞和听神经特征提取

信号 $x(t)$经带通滤波后得到 N 路滤波信号 $y_1(t,i)$，对每一路信号均进行过零与峰值检测、峰值非线性压缩和频率接收。其中，过零检测是指检测出以给定时间 t 为起点的一段时间内信号 $y_1(t,i)$的所有上升过零点，并计算出各相邻过零点之间的时间间隔，令第 i 个信号 $y_1(t,i)$的第 l 个与第 l+1 个过零点之间的时间间隔为 ΔT_{il}。峰值检测是指检测出两过零点之间的信号最大峰值，令第 i 个信号 $y_1(t,i)$的第 l 个与第 l+1 个过零点之间的峰值为 p_{il}。ZCPA 模型利用下式进行非线性压缩：

$$g(p_{il})=\lg(1+20p_{il}) \tag{2}$$

经过以上操作，各滤波信号的过零点时间间隔 ΔT_{il} 和经非线性压缩后的峰值 $g(p_{il})$被传给频率接收器。其中，ΔT_{il} 反映的是频率信息，由于各基底膜滤波器之间存在较大的重合度，因此，在频率轴上划分出 M 频率区间，称为频率箱，则与第 i 个滤波器对应的频率接收器的输出为：

$$y_2(t,m,i)=\sum_{l=1}^{Z_i-1}\delta_{mil}g(p_{il}),m=1,2,\cdots,M \tag{3}$$

其中，Z_i 为 $y_1(t,i)$的过零点总数；m 为频率箱的序号，每一个 m 值均对应一个频率范围；δ_{mil} 为 Kronecker 算子，若 $f_{il}(f_{il}=1/\Delta T_{il})$落入第 m 个频率箱，则 $\delta_{mil}=1$，否则 $\delta_{mil}=0$。

3.2.3 特征信息综合

如图 1 所示，将各频率接收器的输出进行累加便得到 ZCPA 模型的输出，即：

$$zcpa(t,f_m)=\sum_{i=1}^{N}y_2(t,m,i),m=1,2,\cdots,M \tag{4}$$

其中，f_m 为第 m 个频率箱的中心频率，称 $zcpa(t,f_m)$为听觉谱。

4 基于听觉韵律特征的 VAP 语音情感识别实验

4.1 情感语料库的建立

实验选择 6 句汉语作为语料，它们可以充分表达 7 种基本情感(包括高兴、愤怒、惊讶、悲伤、恐惧、厌恶和中立)；同时使用了 4 名专业演员(2 男 2 女)，每个演员每句话按照 7 种情感表达 2 次，这样可以获得 336 句语料；再随机抽取 10 名硕士研究生对这些语料进行一次主观辨听实验，消除语料中情感不明显的语句。在严格的辨听测试中，情感通过说话人表达出来，2 个辨听人判断错误情感的语料将被淘汰掉。这样一系列严格主观辨听后，剩下 214 句能够表达情感的语料脱颖而出。

4.2 语音情感的效价-激励(VA)维空间分布

实验研究了 7 种情感的 VA 的二维空间分布。空间每一维用 7 个水平划分，如−3 表示非常负面，−2 表示负面，−1 表示轻微负面，0 表示非正非负或中性，1 表示轻微正面，2 表示正面，3 表示非常正面。

然后请 20 名正常听力的听众对先前选出的 214 句情感语料进行二次主观辨听实验，并且鉴定每个语料的 7 个档次在 VA 二维空间的分布。在辨听测试的同时，听众可以比较情感语料与中性语料的区别，每种情感在 VA 二维空间的统计分布结果如图 3 所示。

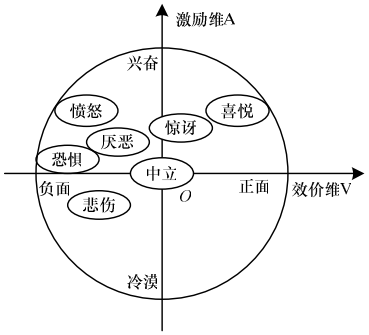


图 3 7 种情感在 VA 维空间的分布

在激励维上，喜悦和愤怒基本上是等高的，向下依次是惊讶、厌恶、恐惧、中立和悲伤；在效价维上，喜悦仍旧是最高的，向下依次是惊讶、中立、厌恶、悲伤、愤怒和恐惧。如图 4 所示，这 7 种情感的变化曲线，它们的激励维曲线要低于效价维曲线。

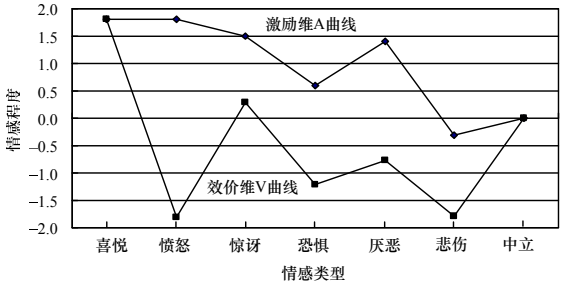
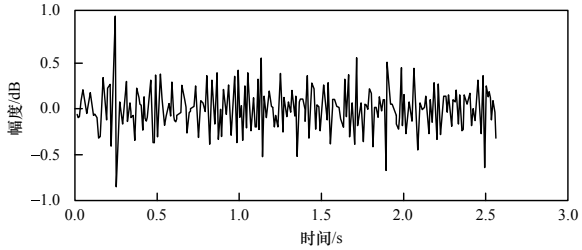


图 4 7 种情感的 VA 维数水平

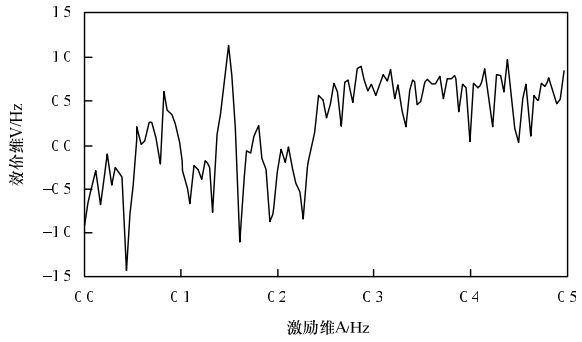
4.3 语音情感韵律特征与 VAP 三维空间的相关性

前文的论述表明了使用 ZCPA 特征作为情感识别的主要

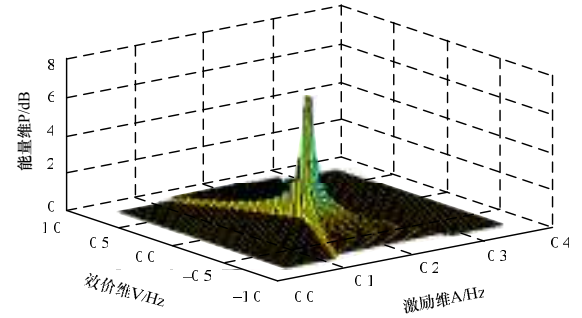
韵律学特征[9]。用 ZCPA 特征研究韵律学特征与 VAP 三维空间的关系[10]。基于 Matlab 仿真实验平台，使用 colea 语音信号处理软件，在隐马尔科夫(HMM)计算环境中，以喜悦情感语音信号“爸爸给我买了一辆车”为例，对韵律 ZCPA 进行特征提取，同时得到其在 VA 和 VAP 空间的分布，如图 5 所示。



(a)喜悦情感语音信号



(b)喜悦语音信号 VA 分布



(c)喜悦语音信号 VAP 分布

图 5　喜悦语音信号特征与 VAP 三维空间的相关性图谱

图 6 为每种情感的 HMM-ZCPA 特征统计。其中，悲伤的 ZCPA 均值和最大值是最低的，而且最大值和最小值的差异也是最小的；喜悦和愤怒的 ZCPA 均值仅次于惊讶的，而明显高于其他情感；喜悦和惊讶的 ZCPA 绝对值方差和最高，而恐惧的最低。

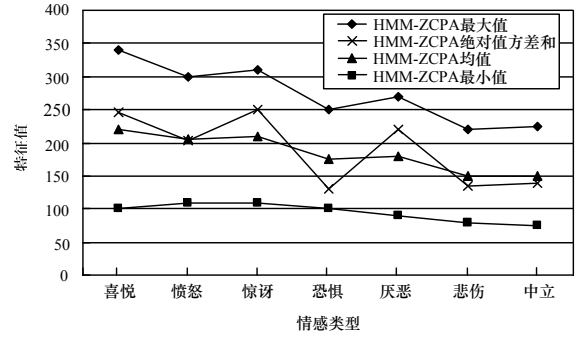


图 6　7 种情感基于 HMM 的 ZCPA 韵律特征统计

为了研究 VAP 与 ZCPA 韵律特征的相关性，计算 VAP 模型与 HMM-ZCPA 的最大值、最小值、均值和绝对值方差和的相关系数，如表 1 所示。

表 1　情感 VAP 维数水平与 ZCPA 韵律特征的相关系数

相关系数	HMM-ZCPA 最小值	HMM-ZCPA 最大值	HMM-ZCPA 均值	HMM-ZCPA 绝对值方差和
效价维 V	−0.064 3	0.478 3	0.493 3	0.465 4
激励维 A	0.593 8	0.767 4	0.806 9	0.759 1
能量维 P	0.374 6	0.885 9	0.597 7	0.246 3

从表 1 可以看出，效价维与 ZCPA 的最大值、均值、绝对值方差和有正面和显著的关系，而与 ZCPA 的最小值的负面关系并不显著；激励维和能量维则与 ZCPA 的均值、最大值、最小值和绝对值方差和都有着正面和显著的关系。

5　结束语

本文基于心理学研究的结果提出了情感结构模型的原理，将语音信号中的 7 种情感分布 VAP 三维情感空间中。实验结果表明，在 VAP 维上所有情感彼此之间都不尽相同，同时维数信息有助于区分情感，而且情感空间模型原理描述情感也是合理的。相同激励维和能量维的情感彼此之间容易混淆。同时，维数水平与 ZCPA 特征之间的相关性分析表明，根据 ZCPA 最大值、最小值、均值和绝对值方差和，相同激励维和效价维的情感共享相同的韵律学特征，而且不同水平激励维、效价维和能量维的情感有着明显差别，所以 ZCPA 作为区分情感的特征是非常有效的。

在今后的研究中，将探究控制维的情感分布，同时为了选择一种有效的韵律学特征区分语音信号中的情感，需要分析维数水平与其他韵律学特征之间的关系，使用合理的数学模型描述情感空间模型，并进行语音信号中的情感识别。

参考文献

[1] 梁青青, 杨鸿武, 郭威彤, 等. 基于语音识别和语速修改的语音复读系统[J]. 计算机工程, 2011, 37(5): 288-290.

[2] 梁五洲, 张雪英. 基于加权组合过零峰值幅度特征的抗噪语音识别[J]. 太原理工大学学报, 2006, 37(1): 77-79.

[3] Cowie R, Douglas-Cowie E, Tsapatsoulis N, et al. Emotion Recognition in Human-computer Interaction[J]. IEEE Signal Processing Magazine, 2007, 18(1): 32-80.

[4] Dellaert F, Polzin T, Waible A. Recognizing Emotion in Speech[C]// Proc. of the International Conference on Spoken Language Processing. Philadelphia, USA: [s. n.], 2006: 1970-1973.

[5] Quatieri T F. 离散时间语音信号处理——原理与应用[M]. 赵胜辉, 译. 北京: 电子工业出版社, 2004.

[6] Wang K, Shamma S. Self-normalization and Noise-robustness in Early Auditory Representations[J]. IEEE Trans. on Speech and Audio Processing, 2004, 2(3): 421-435.

[7] Oded G. Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition[J]. IEEE Trans. on Speech and Audio Processing, 2004, 2(1): 113-131.

[8] Kim Doh-Suk, Lee Soo-Young, Rhee M Kil. Auditory Processing of Speech Signal for Robust Speech Recognition in Real-word Noisy Environments[J]. IEEE Trans. on Speech and Audio Processing, 2008, 7(1): 55-68.

[9] Schuller B, Rigoll G, Lang M. Hidden Markov Model-based Speech Emotion Recognition[C]//Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE Press, 2008: 401-405.

[10] 赵　晖, 顾亚强, 唐朝京. 基于乘积 HMM 的双模态语音识别方法[J]. 计算机工程, 2010, 36(8): 7-9.

编辑　顾逸斐