基于脑电、心电、肌电等神经信号的音乐情感表示模型

吴澹宁
(浙江音乐学院 图书馆,浙江 杭州 310024)

摘 要 现有音乐情感研究主要集中在情感识别领域,情感表示方法的研究偏少,且几乎没有本土化的音乐情感模型.于是,文章设计了基于脑电、心电、肌电等神经信号的音乐情感模型,探索音乐情感的表示方法.其主要思路:从被试者脑电、心电、肌电图中提取平均能量、幅值差、样本熵等特征值,借助相关性分析和随机森林筛选关键特征;基于关键特征与主观评分间映射方式的适用性评估,以LASSO回归法筛选回归项,综合比较特征保留度及性能指标,确定预测愉悦度(P)、唤醒度(A)和支配度(D)的三个表达式.最后,根据自建的40套测试样本的测试结果,模型的单一预测偏差多数在10%以内,如P组88%、A组93%、D组90%测试样本的预测偏差小于±5%;根据自建的295套训练样本的三维凸包,音乐情感表示模型的总体预测结果波动明显小于主观评分,故基于神经信号的情感表示方法可信度更高.

关键词 音乐情感;情感计算;脑电图;心电图;肌电图
中图分类号 Q42;J639 **文献标志码** A **文章编号** 1672-4321(2026)05-0648-10
doi:10.20056/j.cnki.ZNMDZK.20260741

A music mood representation model based on the neural signals of electroencephalograph, electrocardiograph and electromyograph

Wu Danning
(Library, Zhejiang Conservatory of Music, Hangzhou 310024, Zhejiang China)

Abstract Current research on music mood has primarily focused on the field of emotion recognition, with relatively few studies dedicated to the methods of mood representation and absence of localized music mood models. Accordingly, this paper designs a music mood model based on neural signals of electroencephalogram (EEG), electrocardiogram (ECG) and electromyogram (EMG), and explores the representation method of music mood. Its core methodologies are as follows: feature values such as average energy, amplitude difference and sample entropy are extracted from the EEG, ECG and EMG of subjects, while key features are selected via correlation analysis and random forest. Based on an applicability evaluation of mapping modes between key features and subjective ratings, regression terms are filtered using LASSO regression, and three formulas to predict Pleasure (P), Arousal (A) and Dominance (D) values are determined by comprehensive comparison of feature retention and performance metrics. Finally, test results using a self-constructed dataset of 40 testing samples indicate that the single prediction bias are mostly within 10%, for instance, the prediction bias of 88% of P samples, 93% of A samples and 90% of D samples are less than ±5%. The 3-D convex hull constructed from 295 training samples indicate that the overall predictions of music mood model have significantly smaller fluctuation than subjective ratings. Thus, the method of mood representation based on neural signals has higher credibility.

Keywords music mood; affective computing; EEG; ECG; EMG

收稿日期 2025-12-02
作者简介 吴澹宁(1987-),男,高级工程师,研究方向:数字音乐、情感计算、大数据分析与挖掘 E-mail:wdn@zjcm.edu.cn
基金项目 浙江省哲学社会科学规划课题资助项目(25NDJC200YBMS)

音乐是由声音通过媒介按一定规律组织而成的,它被人类听觉器官接收、在大脑皮层诱发生物电刺激,继而产生诸多感性的心理变化.《礼记·乐记》中记载:“凡音之起,由人心生也”,故音乐被认为是表达情感、塑造意象的艺术形式和文化活动.20世纪50年代后,一门将音乐与科技紧密结合的交叉学科(即音乐科技/计算机音乐)悄然兴起.作为音乐科技的重要方向,“音乐信息检索”(Music Information Retrieval, MIR)专注于使用数字技术理解和分析音频内容(尤指音乐).[1]对音乐情感的分析、表示、识别是MIR子课题“音乐高层语义分析”的研究内容之一,在音乐制作、标注、推荐、理疗等场景有着广泛的应用[2].

据《中国数字音乐产业报告(2023)》统计,2023年我国数字音乐市场总规模达1907.5亿元(同比增长22.7%),平台歌曲累计2.13亿首(比上一年度新增8100万首).[3]随着数字音乐市场规模、作品数量的爆发式增长,音乐信息检索的用户需求、产业形态日益多元化和专业化.如何发挥数字技术的赋能提质作用与融合创新优势,解决数字音乐领域的痛点难点问题,培育和发展文化产业的新质生产力,已成为当下的热点话题.2025年政府工作报告指出:“持续推进‘人工智能+’行动,将数字技术与制造优势、市场优势更好结合起来,支持大模型广泛应用”.[4]本文通过音乐听众脑电、心电、肌电等神经信号的综合分析,探索模拟真实情感的三维空间表示方法,用于后续音乐资源的标注与编目,进而推动“人工智能+音乐”跨学科、多领域的深度融合,为打造具有国际竞争力的数字音乐集群提供智力支撑.结构上,文章第二部分回顾了情感理论、模型、计算等方面的研究成果,分析其优势与不足;第三部分简述了神经信号采集实验的设计细节,包括被试材料、对象、数据采集和预处理;第四部分论述了脑电、心电、肌电特征的提取方法和关键特征的筛选过程;第五部分评估了关键特征与主观评分间的映射方式,基于训练集构建音乐情感表示模型,并验证了模型的预测效能及此类表示方法的可行性;第六部分总结了研究的主要内容,提出了未来的工作方向.

1　研究现状

1.1　情感理论

心理学上,“情感”(Mood)是人类独有的、稳定的、具有社会性的高级情绪(Emotion)反应[5].情感的主体是人,关于音乐传递出的情感,学界存在两种观点.其一认为音乐情感是由作者谱曲编创或演奏者演绎赋予的,即“表达理论”,另一认为是听众聆听音乐时产生的体验感受,即“唤起理论”[6].“表达理论”聚焦音乐的声学特征解析,如能量、频域、时域等[7];“唤起理论”则强调对听众反馈信息的捕捉分析,包括语言评价和生理表现.出于数据采集、分析难度的考量,多数研究采纳“表达理论”.但就应用而言,“唤起理论”更符合个性化推荐的需求[8].笔者以为,音乐的初始情感源自作曲者的谱曲编创,可通过作品的调性、旋律的音高、节拍、和弦等特征析出;乐曲经演绎后呈现的各项声学特征,融入了演奏者对乐谱的个人理解与即兴发挥,属于音乐情感的二次加工(即一阶情感);而听众的体验感受(或引起共鸣、或触发联想),是对作品、演奏所表达情感的又一次加工,可视为二阶情感.

1.2　情感模型

构建模型是从情感理论走向情感计算的关键步骤.心理学界主要有两种方式——分类概念化和维度概念化[9].分类概念化从范畴观的角度对情感进行划分,国内称为离散模型.音乐情感方向的离散模型始于1936年Kate Hevner归纳的8类别、66个情感形容词[10].维度概念化是利用向量值表示情感在二维或三维空间中的特定位置,国内称为维度模型.1974年,Albert Mehrabian和James A. Russell解释了基础情感的三个维度——愉悦度(Pleasure,以下简称P)、唤醒度(Arousal,以下简称A)和支配度(Dominance,以下简称D)[11].1980年,James A. Russell将愉悦度改为效价度(Valence,以下简称V),构建了二维象限的情感模型,其横轴表示V、纵轴表示A.离散与维度模型分别对应了计算科学中的分类法和回归法.前者易于理解、可解释性强,但碍于语言、文化间差异和文字表述的模糊性,通用性和量化能力较弱.后者所描述连续过渡、动态转化的情感状态,更贴近人类情感的实际情况,在情感计算领域更具可行性.

1.3　基于神经科学的情感计算

神经科学是研究神经系统结构、功能,探索人类感知、记忆、思维等方面原理和规律的科学[12].过去,社会科学研究往往依靠口头报告的量表问卷得出结论.但自我评价带有很强的主观性,被试者可能出于某些特殊动机,隐藏、回避或高估对某事物的真实想法、能力[13].运用神经科学方法获取的生理数据受意识控制程度低,能更准确地分析和预测个

体对特定音乐的情感表达或唤起[14].

脑电图(Electroencephalogram, EEG)是通过头皮表面或内置的电极探针,记录锥体神经元的突触后电位,绘制成包含频域、时域信息的波形图[15].EEG经变换、整形、平滑、降维处理后,被重构为5种节律波(γ、β、α、θ和δ),再提取能量、幅值、微分熵等关键特征,用于大脑的情感研究[16].例如,李洪伟等运用互信息计算建立了不同波段的动态脑网络,观察到α和γ波段上不同音乐情感间的差异性最为明显[17].心电图(Electrocardiogram, ECG)是心脏周期性电活动的测量结果,常用指标有P、QRS、T波的时频域信息和心率变异性(Heart Rate Variability, HRV).例如,喻一梵等提取了连续R波间隔(RR-I)、P波峰值(P-H)、心率等9个心电特征,有效识别了8种正负性情绪[18].肌电图(Electromyogram, EMG)采集的是面部、颈部等处肌肉运动后产生的电信号.例如,Lisa P. Chan等探究了针对歌曲演唱的面部反馈情况,处于欢快情境时,听众的颧大肌活动增强,悲伤情境时则皱眉肌更为活跃[19].除了单模信号分析,研究者还尝试了多模信号的融合分析.例如,李路宝等使用图神经网络和注意力机制对EEG、ECG进行分类,再通过证据理论融合分类结果,在视频诱导实验中,双模态的情感识别效率略优于单模态EEG、远胜于单模态ECG[20].

通过文献检索可知,国内外在情感模型、基于神经信号的情感计算方面成果丰富,但仍有下列不足:①以文本、图片或视频诱发的情感研究较多,源自听觉刺激的音乐情感相对匮乏;②涉及情感形容词、VA模型的较多,本土化的三维模型建构较少,如表1所示.而且,过往研究更关注情感的识别问题,情感表示方法的研究偏少.情感识别常遵循分类概念化的思路,预设若干情感形容词或主要类别,聚焦不同算法的精度表现与适用场景,其优势在于对情感大类(如正负性、积极消极)的识别精度极高,但较难细化不同曲目的情感烈度.而情感表示属于更基础的研究范畴,深入挖掘高精度的情感表示方法能够有效反哺情感识别研究.关于情感表示方法,国内外多采用自报告形式的量表数据,受个体认知、文化环境等因素的影响较大.因此,本文利用纯听觉材料诱发听众的音乐情感,创建基于神经信号的三维表示模型,以期填补上述不足、拓展此方面的研究.

表1 国内外情感模型、基于神经信号的情感计算的研究统计

Tab. 1 Statistics on research of domestic and foreign emotion models and neural signal-based affective computing

检索词	总文献数量	音乐情感比例	三维模型比例
情感/情绪/感情模型(Mood/Emotion/Affect Model),	国内1574	国内9.9%	国内8.4%
包括但不限于情感形容词、VA模型、PAD模型	国外5297	国外1.7%	国外13.2%
脑电+情感/情绪/感情	国内4553	国内16.8%	国内10.6%
(EEG+Mood/Emotion/Affect)	国外2553	国外3.0%	国外11.1%
心电/心率变异性+情感/情绪/感情	国内1318	国内14.0%	国内5.1%
(ECG/HRV+Mood/Emotion/Affect)	国外483	国外2.7%	国外0%
肌电+情感/情绪/感情	国内492	国内14.2%	国内10.3%
(EMG+Mood/Emotion/Affect)	国外115	国外3.5%	国外0%

注:①国内文献以中国知网为准,国外以谷歌学术为准;②检索时间为2025年7月31日,检索方式为主题词,使用NoteExpress和DeepSeek R1初筛检索结果,去除关联度较小的文献;③“音乐情感比例”=“音乐情感”文献/总文献数量,“三维模型比例”=“三维模型”文献/“音乐情感”文献.

2 实验设计

2.1 被试信息

本实验的诱发材料选自Alexander Street、中国音网、网易云音乐等在线平台,涵盖西方古典音乐、世界民族音乐、中国传统音乐和华语流行音乐(每种类型各20首).随后,邀请志愿者对上述曲目开展预测评,选取情感诱发强烈的20首(每种类型各5首),分别截取60秒高潮选段作为正式的被试材料[21].另召集20位在校师生(音乐院校与综合院校各半、年龄19~53岁、无知觉障碍和神经系统疾病),借助巴塞罗那音乐奖励问卷(Barcelona Musical Reward Questionnaire)[22]筛选出音乐高敏感10人、中敏感8人参加正式实验.实验前24小时,被试者无兴奋或抑制性物质摄入,并确保8小时以上睡眠.

2.2 采集流程

正式实验流程:①告知被试者实验目的、流程

及注意事项,根据《世界医学会赫尔辛基宣言》和学校学术委员会审查要求,请被试者签署情况告知书;②被试者聆听第一首曲目,完毕后60秒内填写中科院PAD量表[23];③播放240秒水滴白噪音,以便被试者平复状态;④更换第二首曲目并重复②、③步骤,直至所有曲目播放完毕.由于大脑前额叶皮层在情感调节和意识经验中起到重要作用[24-25],本实验使用ErgoLAB Dry干电极脑电仪采集前额区的FP1、FP2、AF7、AF8四通道EEG信号[26].同时,使用ErgoLAB Hub穿戴式多模态生理记录仪采集ECG和EMG信号[27].其中,EMG主要监测颧肌(ZYG)、皱眉肌(CORR)和胸锁乳突肌(SCM)区域,分别指向微笑、皱眉、点头等行为[28-29].信号采集区间为音乐播放前30秒静息状态和60秒播放状态,共计形成360套EEG、ECG、EMG样本数据.

2.3　数据预处理

信号采集过程可能受外部因素干扰(如电极移动),本实验通过数据清洗操作,去除存在数据缺失、异常波动的(整套)样本数据.关于信号的噪声问题,实验场所选在经特殊声场处理的视听体验馆(由作者所在单位建设),从源头削弱设备产生的工频噪声.通常,EEG频率范围0.1~50 Hz[16],ECG频率0.05~100 Hz[30],EMG频率50~150 Hz[31].考虑到心电电极位置距离头部较远,ECG对EEG、EMG的伪迹影响较小.EEG与EMG则借助监测设备的带通滤波器,以0.002~50 Hz和50~200 Hz为截止频率分别滤除无用的频率成分,并采用独立成分分析法去除眼动的噪声影响[32].经筛选,最终得到335套有效数据(每首曲目14—18套)用于特征提取.其中,每首曲目随机抽取2套(共40套)作为测试集,其余用于模型训练,模型的整体构建思路如图1所示.

3　神经信号处理

3.1　特征提取

在计算机科学领域,EEG、ECG、EMG等波形信号具有非线性、非平稳、非高斯分布等特性[33],传统的时频域分析很难克服其不利影响.1998年,Norden E. Huang等根据信号本身的尺度特征,自适应地将其分解成若干连续的、具有物理意义的固有模态函数(Intrinsic Mode Function, IMF)分量,有效解决了信号的非线性、非平稳问题[34].本研究对335套EEG、ECG、EMG有效数据的特征提取采用一种完全非递归算法——变分模态分解[35],按通道、区域分解为一系列围绕各自中心频率的IMF,以减少分量间的模态混叠现象.EEG特征的提取过程:①使用磷虾群算法确定最优分解层数K和二次罚项α;②根据公式(1)初始化子信号μ_1及其中心频率ω_1;③引入增广拉格朗日函数,并运用交替方向乘子法更新μ_k和ω_k,直至收敛条件(2).以03号被试、05号曲目为例,首先,将60秒播放状态下FP1通道EEG分解为5个IMF分量图和频谱图;其次,截取前3阶分量计算相应的平均能量、幅值差和样本熵,如公式(3)-(5);最后,计算前30秒静息状态的EEG、ECG、EMG特征值作为基准,播放状态与静息状态的差值即为03号被试、05号曲目的实际生理反馈.

$$\min_{\{\mu_k\},\{\omega_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*\mu_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\},(1)$$
$$s.t.\sum_k\mu_k=f$$

其中,f为节律波信号,μ_k为第k层子信号,ω_k为第k层子信号的中心频率,δ为狄拉克分布,j^2=-1.

$$\sum_k\frac{\left\|\mu_k^{n+1}-\mu_k^n\right\|_2^2}{\left\|\mu_k^n\right\|_2^2}<e,(2)$$

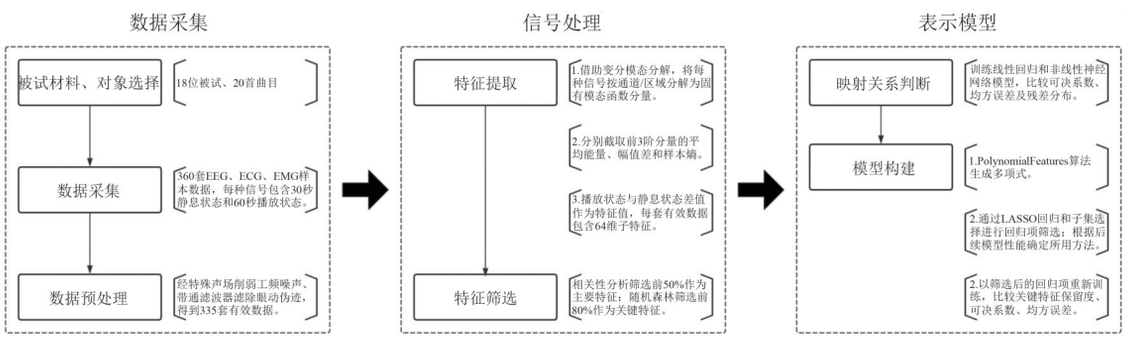


图1　音乐情感表示模型的整体构建思路

Fig. 1 Overall construction idea of music emotion representation model

$$E(i)=\frac{1}{n}\sum_{t=1}^{n}\left|e(t)\right|^2 \quad ,(3)$$

$$A(i,j)=\frac{1}{n}\sum_{t=1}^{n}\left|a_i(t)-a_j(t)\right| \quad ,(4)$$

$$\mathrm{SPE}(i)=\ln\left(\frac{1}{N-m+1}\sum_{l=1}^{N-m+1}C_l^m\right)-\ln\left(\frac{1}{N-m}\sum_{l=1}^{N-m}C_l^{m+1}\right)$$
$$C^m=\frac{1}{N-m+1}B(r) \quad ,(5)$$

其中,i, j为IMF阶数,n为IMF中采样数,$e(t)$为采样点能量,$a(t)$为采样点幅值.

其中,N为样本点数,嵌入维数m=2,B为两序列间最大差值小于容限r的数量,r=0.2个标准差.

3.2 特征筛选

特征筛选的常用方法有适用线性关系的相关性分析、Lasso回归等,适用非线性、含交互的树模型、互信息等.本研究选用线性与非线性相结合方法,筛选对PAD评分贡献度较高的子特征,即先用Pearson相关分析作先验筛选,过滤明显无关的特征,再用随机森林算法作进阶筛选,兼顾非线性与鲁棒性.具体过程:①对295套训练特征值进行标准化处理,与P、A、D评分组合为P组、A组、D组数据,依次输入SPSS 27.0作信效度检测.结果显示,三组数据的Cronbach's α系数为0.902、0.811、0.881,均高于0.8,说明具有较好的内部一致性;KMO测度为0.938、0.733、0.828,均高于0.7,Bartlett球形度检验为0.000,适合进行相关性分析.②对三组数据分别做Pearson相关分析,保留相关系数(绝对值)前50%作为主要特征,如表2所示.根据近似冗余特征的定义(见公式(6)),P组中SPE ($FP1_3$)与E (ZYG_2)间信息重叠高、存在近似冗余,故舍去E (ZYG_2),剩余31项主要特征.同理,A组剩余31项、D组剩余30项.③使用随机森林算法对主要特征进行二次筛选,按累积重要性大于80%获得P组14项关键特征、A组和D组各13项关键特征,如表3-4和表5所示.

$$\text{当}\rho_{j,y}\geqslant\rho_{i,y},\rho_{\max}-\rho_{i,y}>\alpha,$$
$$\text{若}\rho_{i,j}>\left(\bar{\rho}+\rho_{i,y}\right)/2,\text{则特征}i\text{是近似冗余的} \quad ,(6)$$

其中p为相关系数,i, j为子特征,y为目标值,$p_{\max}$为所有子特征与目标值相关系数中最大值,α一般为0.03~0.1.

表2 64维子特征与P值的相关性分析

Tab. 2 Correlation analysis of 64-dimensional sub-features and P-values

子特征	相关系数	子特征	相关系数	子特征	相关系数	子特征	相关系数
E ($FP1_1$)	0.388	E ($AF7_1$)	-0.278	E (ECG_1)	0.348	E (COR_1)	0.347
E ($FP1_2$)	0.091	E ($AF7_2$)	-0.148	E (ECG_2)	-0.458	E (COR_2)	-0.209
E ($FP1_3$)	-0.153	E ($AF7_3$)	0.038	E (ECG_3)	-0.371	E (COR_3)	0.318
A ($FP1_{1,2}$)	0.526	A ($AF7_{1,2}$)	0.283	A ($ECG_{1,2}$)	-0.311	A ($COR_{1,2}$)	-0.330
A ($FP1_{2,3}$)	-0.194	A ($AF7_{2,3}$)	-0.218	A ($ECG_{2,3}$)	-0.486	A ($COR_{2,3}$)	-0.464
SPE ($FP1_1$)	0.173	SPE ($AF7_1$)	0.084	SPE (ECG_1)	-0.268	SPE (COR_1)	0.035
SPE ($FP1_2$)	0.031	SPE ($AF7_2$)	0.170	SPE (ECG_2)	-0.092	SPE (COR_2)	0.010
SPE ($FP1_3$)	-0.371	SPE ($AF7_3$)	-0.015	SPE (ECG_3)	-0.421	SPE (COR_3)	-0.313
E ($FP2_1$)	-0.054	E ($AF8_1$)	-0.216	E (ZYG_1)	-0.521	E (SCM_1)	0.102
E ($FP2_2$)	0.182	E ($AF8_2$)	-0.178	E (ZYG_2)	-0.365	E (SCM_2)	0.081
E ($FP2_3$)	-0.007	E ($AF8_3$)	0.261	E (ZYG_3)	0.234	E (SCM_3)	-0.327
A ($FP2_{1,2}$)	0.127	A ($AF8_{1,2}$)	-0.048	A ($ZYG_{1,2}$)	0.019	A ($SCM_{1,2}$)	0.007
A ($FP2_{2,3}$)	-0.298	A ($AF8_{2,3}$)	-0.087	A ($ZYG_{2,3}$)	0.069	A ($SCM_{2,3}$)	-0.091
SPE ($FP2_1$)	-0.456	SPE ($AF8_1$)	0.345	SPE (ZYG_1)	0.657	SPE (SCM_1)	0.489
SPE ($FP2_2$)	-0.102	SPE ($AF8_2$)	0.143	SPE (ZYG_2)	-0.008	SPE (SCM_2)	-0.305
SPE ($FP2_3$)	-0.479	SPE ($AF8_3$)	0.171	SPE (ZYG_3)	0.032	SPE (SCM_3)	-0.265

注:红色字体表示相关性前50%,黄色背景表示存在近似冗余.

4 音乐情感表示

4.1 映射方式

数据间映射方式的常见类型有函数映射(线性)、概率映射(非线性)等.本研究使用Python语言构建并训练线性回归模型和非线性神经网络模型(相关参数见表6),对比二者的可决系数R^2、均方误差MSE及残差分布,判断关键特征与主观评分间映

表3　P组主要特征的随机森林筛选

Tab. 3 Random forest screening of main features in group P

主要特征	重要性	累积重要性	主要特征	重要性	累积重要性
A（$ECG_{1,2}$）	12.2%	12.2%	E（ECG_1）	4.4%	58.5%
E（$AF8_3$）	8.6%	20.8%	SPE（COR_3）	4.3%	62.8%
SPE（$FP2_1$）	7.6%	28.4%	A（$COR_{2,3}$）	4.2%	67.0%
A（$FP1_{1,2}$）	7.0%	35.4%	E（ZYG_1）	4.2%	71.2%
A（$ECG_{2,3}$）	6.9%	42.3%	E（$FP1_1$）	4.2%	75.3%
E（SCM_3）	5.9%	48.3%	SPE（ECG_3）	4.1%	79.5%
A（$COR_{1,2}$）	5.9%	54.1%	E（ECG_2）	4.1%	83.5%

注:仅罗列累积重要性大于80%的关键特征,下同.

表4　A组主要特征的随机森林筛选

Tab. 4 Random forest screening of main features in group A

主要特征	重要性	累积重要性	主要特征	重要性	累积重要性
SPE（$FP1_2$）	8.8%	8.8%	E（$AF7_1$）	5.6%	57.7%
E（ECG_2）	8.6%	17.4%	SPE（$AF7_2$）	5.6%	63.3%
SPE（$FP2_2$）	7.9%	25.2%	SPE（$FP2_3$）	5.5%	68.7%
A（$AF8_{1,2}$）	7.4%	32.6%	E（$FP2_2$）	4.4%	73.2%
E（ZYG_1）	7.2%	39.8%	SPE（$FP1_1$）	4.4%	77.6%
E（COR_3）	6.2%	46.0%	SPE（$AF7_1$）	4.3%	81.9%
A（$SCM_{1,2}$）	6.1%	52.1%	...	...	...

表5　D组主要特征的随机森林筛选

Tab. 5 Random forest screening of main features in group D

主要特征	重要性	累积重要性	主要特征	重要性	累积重要性
A（$SCM_{1,2}$）	10.3%	10.3%	E（SCM_3）	5.3%	60.3%
E（$AF7_1$）	9.7%	20.0%	A（$ECG_{2,3}$）	5.2%	65.5%
SPE（$FP1_2$）	8.4%	28.5%	SPE（$AF7_1$）	4.6%	70.1%
SPE（$FP2_2$）	7.5%	36.0%	A（$FP1_{1,2}$）	4.5%	74.5%
SPE（ECG_3）	7.0%	42.9%	A（$ZYG_{1,2}$）	4.0%	78.5%
E（$AF8_3$）	6.0%	49.0%	E（ECG_1）	3.7%	82.2%
SPE（$FP2_1$）	6.0%	55.0%	...	...	...

射方式的适用性.以P组为例,线性回归模型的R^2大于非线性神经网络模型,说明其解释能力更强;MSE小于非线性神经网络模型,说明其预测偏差更小;且多数预测的残差位于[-1, 1]区间内,如图2所示.

表6　非线性神经网络模型的关键参数

Tab. 6 Key parameters of nonlinear neural network model

层数	各层节点数	各层激活函数	训练轮数	优化器
3	[64, 32, 1]	['ReLU', 'ReLU', 'Linear']	100	Adam

4.2　模型构建

本研究使用机器学习库Sklearn中PolynomialFeatures算法生成二次多项式.例如,P组14维关键特征共生成119个回归项,包括14个一次项(x_1, x_2...)、14个二次项(x_1^2, x_2^2...)和91个交互项($x_1 \times x_2$, $x_1 \times x_3$...).然后,尝试LASSO回归(L1正则化)和子集选择(基于信息准则)两种方法进行初步筛选:前者通过施加正则化参数α,自动压缩冗余项系数至零;后者基于AIC和BIC准则平衡拟合度与复杂度,寻找最优回归项子集.同时,设置了get_feature_origin函数,追溯回归项的原始特征,如P组交互项A（$FP1_{1,2}$）×E（ZYG_1）表征A（$FP1_{1,2}$）和E（ZYG_1）两个关键特征.以P组为例,LASSO回归筛选得到9个二次项和28个交互项;子集选择筛选得到6个一次项、7个二次项和27个交互项.两种方法的筛选结果各自按系数绝对值排列,假设前若干项为实际回归项,重新进行模型训练并计算R^2和MSE,如表7和表8所示.结果表明,使用LASSO法筛选回归项的模型性能普遍优于子集选择法.在综合比较关键特征保留度和性能指标后,认为P组基于前9个回归项构建的模型性能最优,回归方程如公式(7)所示.同理,得出A组、D组的最优方程如公式(8)、(9)所示.

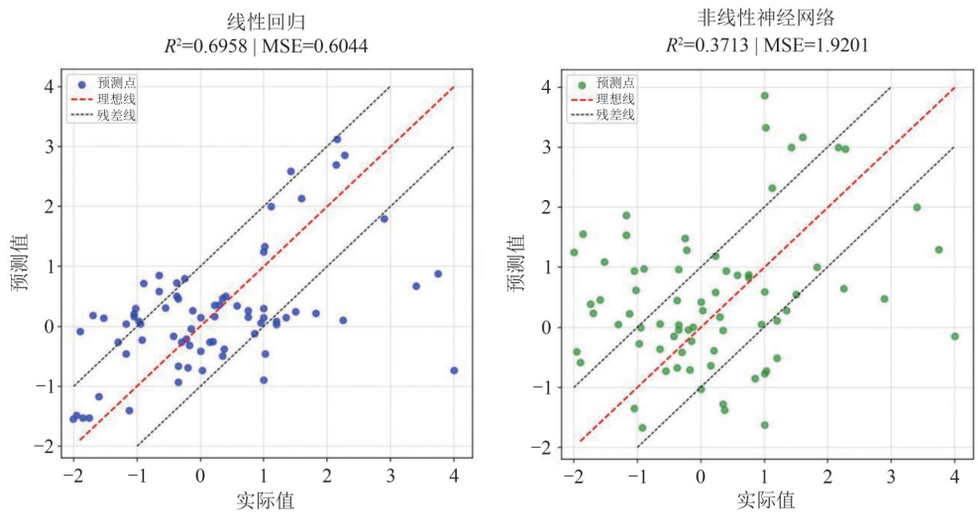


图2 线性回归与非线性神经网络模型的性能比较(P组)

Fig. 2 Performance comparison between linear regression and nonlinear neural network model (group P)

表7 LASSO回归筛选后的模型性能比较(P组)

Tab. 7 Performance comparison of models screened by LASSO regression (group P)

回归项	关键特征	R^2	MSE
前14个:$A(ECG_{1,2})^2$, $E(ECG_2)$*$SPE(ECG_3)$, $E(ECG_2)$*$A(ECG_{1,2})$, $E(AF8_3)^2$, $E(ECG_1)$*$A(ECG_{1,2})$, $SPE(ECG_3)$*$E(SCM_3)$, $SPE(FP2_1)$*$E(ECG_2)$, $A(FP1_{1,2})$*$E(ECG_2)$, $SPE(FP2_1)$*$E(ZYG_1)$, $A(ECG_{2,3})$*$E(ZYG_1)$, $SPE(COR_3)$*$E(SCM_3)$, $E(AF8_3)$*$E(ECG_1)$, $SPE(FP2_1)^2$, $E(ECG_2)$*$A(COR_{1,2})$	12	0.6376	0.1348
前13个	11	0.6288	0.2468
前12个	11	0.7200	0.1587
前11个	11	0.7090	0.1037
前10个	10	0.6017	0.1836
前9个	9	0.7451	0.1061
前8个	8	0.5705	0.1259
...	...	...	...

表8 子集选择筛选后的模型性能比较(P组)

Tab. 8 Performance comparison of models screened by subset selection (group P)

回归项	关键特征	R^2	MSE
前16个:$A(ECG_{1,2})$, $E(ECG_2)$, $A(ECG_{2,3})$, $E(FP1_1)$, $E(SCM_3)$, $A(FP1_{1,2})$, $E(AF8_3)$*$E(ECG_1)$, $SPE(ECG_3)$*$E(SCM_3)$, $A(ECG_{1,2})^2$, $A(ECG_{1,2})$*$A(COR_{1,2})$, $A(ECG_{2,3})$*$A(COR_{1,2})$, $E(ECG_2)$*$A(ECG_{1,2})$, $E(ZYG_1)$*$A(COR_{1,2})$, $E(ECG_1)$*$A(ECG_{1,2})$, $E(ZYG_1)$*$SPE(COR_3)$, $SPE(FP2_1)$*$A(ECG_{2,3})$	13	0.3043	0.6100
前15个	12	0.2958	0.6224
前14个	11	0.2866	0.6058
前13个	11	0.3428	0.5560
前12个	10	0.3727	0.4562
前11个	10	0.3675	0.4638
前10个	10	0.4672	0.4342
前9个	9	0.4657	0.4065
...	...	...	...

$$P_p = -0.3008 + 0.0404\mathrm{E}(\mathrm{AF8}_3)^2 - 0.0385\mathrm{A}(\mathrm{ECG}_{1,2})^2 + 0.0629\mathrm{A}(\mathrm{FP1}_{1,2})\mathrm{E}(\mathrm{ECG}_2) + 0.0551\mathrm{SPE}(\mathrm{ECG}_3)\mathrm{E}(\mathrm{SCM}_3) + 0.043\mathrm{SPE}(\mathrm{FP2}_1)\mathrm{E}(\mathrm{ECG}_2) - 0.0392\mathrm{E}(\mathrm{ECG}_2)\mathrm{SPE}(\mathrm{ECG}_3) + 0.028\mathrm{E}(\mathrm{ECG}_2)\mathrm{A}(\mathrm{ECG}_{1,2}) + 0.0254\mathrm{SPE}(\mathrm{FP2}_1)\mathrm{E}(\mathrm{ZYG}_1) + 0.024\mathrm{E}(\mathrm{ECG}_1)\mathrm{A}(\mathrm{ECG}_{1,2}) \quad ,(7)$$

$$A_p = -0.0451 + 0.0228\mathrm{SPE}(\mathrm{FP1}_2)^2 + 0.0334\mathrm{SPE}(\mathrm{FP1}_1)\mathrm{E}(\mathrm{ZYG}_1) - 0.0315\mathrm{SPE}(\mathrm{FP1}_1)\mathrm{SPE}(\mathrm{FP2}_3) + 0.0302\mathrm{SPE}(\mathrm{FP1}_1)\mathrm{A}(\mathrm{SCM}_{1,2}) - 0.0284\mathrm{E}(\mathrm{AF7}_1)\mathrm{A}(\mathrm{AF8}_{1,2}) + 0.0253\mathrm{SPE}(\mathrm{FP2}_2)\mathrm{SPE}(\mathrm{AF7}_1) + 0.0212\mathrm{E}(\mathrm{FP2}_2)\mathrm{E}(\mathrm{COR}_3) - 0.0189\mathrm{SPE}(\mathrm{AF7}_2)\mathrm{E}(\mathrm{ECG}_2) \quad ,(8)$$

$$D_p = 0.7688 - 0.0477\mathrm{SPE}(\mathrm{FP1}_2)^2 - 0.025\mathrm{A}(\mathrm{ECG}_{2,3})^2 - 0.0399\mathrm{E}(\mathrm{ECG}_1)\mathrm{A}(\mathrm{SCM}_{1,2}) - 0.0332\mathrm{SPE}(\mathrm{FP2}_1)\mathrm{A}(\mathrm{SCM}_{1,2}) + 0.0307\mathrm{SPE}(\mathrm{FP2}_2)\mathrm{E}(\mathrm{AF8}_3) - 0.0242\mathrm{SPE}(\mathrm{AF7}_1)\mathrm{E}(\mathrm{SCM}_3) + 0.0239\mathrm{SPE}(\mathrm{ECG}_3)\mathrm{A}(\mathrm{SCM}_{1,2}) + 0.0188\mathrm{E}(\mathrm{ECG}_1)\mathrm{A}(\mathrm{ECG}_{2,3}) \quad .(9)$$

4.3　测试分析

将40套测试特征值输入回归方程(7)-(9),计算残差、R^2和MSE,如图3、图4和图5所示.从单一偏差视角,所有预测的残差均位于[-0.62, 0.90]区间内,范围占总区间([-4,4])的19%.其中,P组90%残差位于[-0.56, 0.22](总区间的9.6%)、50%位于[-0.14, 0.04](2.3%),A组90%残差位于[-0.40, 0.22](7.8%)、50%位于[-0.21, 0.04](3.1%),D组90%残差位于[-0.41, 0.23](8.0%)、50%位于[-0.08, 0.09](2.1%),说明测试集的单一偏差控制在20%以内(多数在10%以内、近半在3%左右),如表9所示.其次,将295套训练特征值输入回归方程,生成295个预测情感的三维坐标,按曲目依次构建主观评分和预测情感的三维凸包,并计算其空间体积,如图6和表10所示.从空间分布视角,所有曲目的预测凸包与评分凸包均有重合,且多在50%以上(即单个坐标偏差±20%),说明训练集的单一偏差与测试集的结果相近.另外,多数预测凸包的体积显著小于主观评分(减小80%以上),说明相较于主观评分,回归方程的预测结果波动更小、收敛性较高.根据最新文献,基于EEG的卷积神经网络及其变体的识别精度为81.77 %~99.42 %[36].

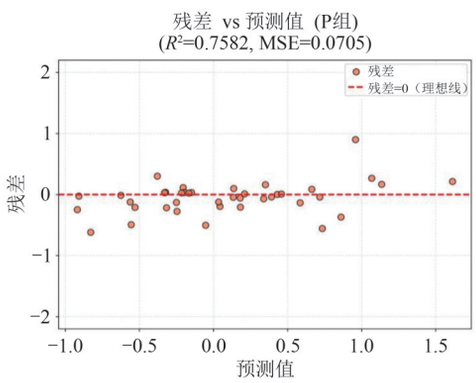


图3　基于40套训练特征值的预测性能分析(P组)

Fig. 3 Prediction performance analysis based on 40 sets of training eigenvalues (group P)

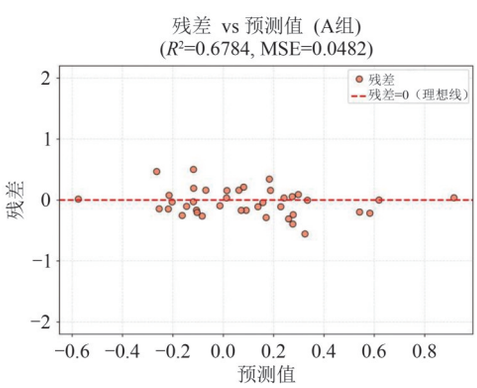


图4　基于40套训练特征值的预测性能分析(A组)

Fig. 4 Prediction performance analysis based on 40 sets of training eigenvalues (group A)

5　结语

本文旨在探究听众视角的音乐情感表示方法:首先,基于20首曲目的反馈实验,收集到335套EEG、ECG、EMG数据,使用变分模态分解法绘制固有模态函数分量图和频谱图,提取平均能量、幅值差、样本熵等64维特征值,又借助相关性分析和随机森林筛选出13-14项关键特征.经过回归模型和神经网络的模拟分析,判断关键特征值与PAD主观评分间线性的映射关系更具适用性.其后,以LASSO回归筛选二次多项式的119个回归项,综合比较关键特征保留度和R^2、MSE等性能指标,确定预测P、A、D值的三个表达式——音乐情感的三维表示模型.最后,将40套测试特征值输入表示模型,比较单一P、A、D预测值的偏差,得出多数偏差在10%以内、近半偏差在3%左右;基于295套训练特征值计算并构建三维凸包,与相应的主观评分凸包比较后发现,表示模型的预测结果波动更小,说明这种神经信号辅助的情感表示方法可信度更高.

表9 基于40套训练特征值的单一偏差分析

Tab. 9 Single deviation analysis based on 40 sets of training eigenvalues

残差及偏差	P	A	D
整体残差区间	[-0.62, 0.90]	[-0.56, 0.51]	[-0.46, 0.58]
偏差20%(残差±0.8)的数量及比例	39, 98%	40, 100%	40, 100%
偏差10%(残差±0.4)的数量及比例	35, 88%	37, 93%	36, 90%
偏差5%(残差±0.2)的数量及比例	26, 65%	27, 68%	29, 73%
偏差3%(残差±0.12)的数量及比例	19, 48%	16, 40%	22, 55%
偏差1%(残差±0.04)的数量及比例	12, 30%	8, 20%	8, 20%

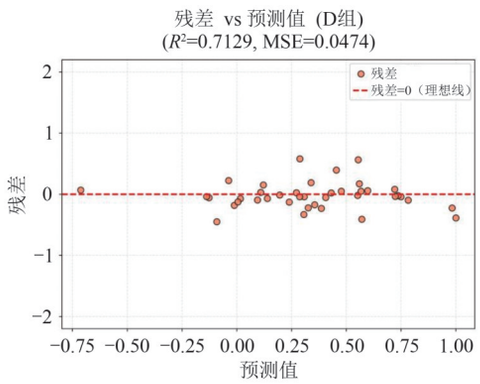


图5 基于40套训练特征值的预测性能分析(D组)

Fig. 5 Prediction performance analysis based on 40 sets of training eigenvalues (group D)

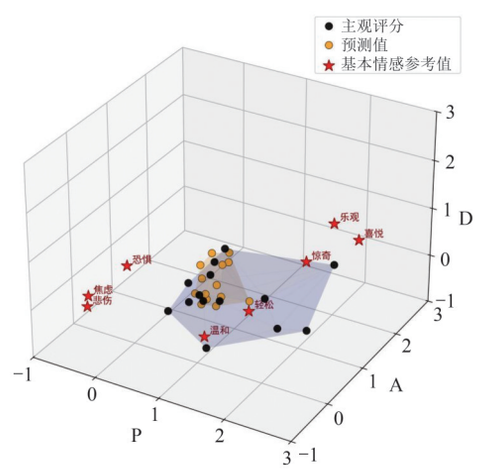


图6 主观评分和预测情感的三维凸包(14号曲目)

Fig. 6 Three-dimensional convex hull of subjective scores and predicted emotions (track 14)

虽然表示模型的预测表现较为理想,但其预测值集中于[-1.3, 1.6]区间,对极端情感(数值大于2或小于-2)的预测能力尚有待验证,后续将继续深入和细化此方面研究.随着神经信号监测设备及脑机接口技术的不断升级迭代,音乐情感表示模型能够充分运用至各类资源的情感标注中,进一步修正和丰富编目信息,为精准的用户需求分析和个性化推荐奠定坚实基础.

表10 各曲目的三维凸包分布比较

Tab. 10 Comparison of three-dimensional convex hull distribution of each track

曲目	凸包体积			重合体积	重合比例
	主观评分	预测情感	变化比例		
01	2.356	0.311	-86.8%	0.272	87.5%
02	5.396	0.360	-93.3%	0.228	63.3%
03	3.873	0.172	-95.6%	0.140	81.4%
04	1.542	0.213	-86.2%	0.119	55.9%
05	3.517	0.291	-91.7%	0.195	67.0%
06	2.145	0.171	-92.0%	0.066	38.6%
07	6.627	0.182	-97.3%	0.172	94.5%
08	8.836	0.816	-90.8%	0.612	75.0%
09	3.004	0.370	-87.7%	0.265	71.6%
10	0.724	0.108	-85.1%	0.052	48.1%
11	4.478	0.368	-91.8%	0.225	61.1%
12	3.194	0.367	-88.5%	0.291	79.3%
13	8.531	0.131	-98.5%	0.056	42.7%
14	1.738	0.144	-91.7%	0.099	68.8%
...	...	...	...	...	...

注:重合比例=重合体积/预测情感的凸包体积;变化比例=预测情感与主观评分的凸包体积差/主观评分的凸包体积.

参 考 文 献

[1] 李伟, 李子晋, 高永伟. 理解数字音乐——音乐信息检索技术综述[J]. 复旦学报(自然科学版), 2018, 57(3): 271-313.

[2] 李伟, 高智辉. 音乐信息检索技术: 音乐与人工智能的融合[J]. 艺术探索, 2018, 32(5): 112-116.

[3] 崔恒勇, 王炜茹, 王羽萱. 2023年中国数字音乐产业发展情况解析——基于《中国数字音乐产业报告(2023)》数据[J]. 中国数字出版, 2024, 2(6): 72-79.

[4] 新华社. 政府工作报告[EB/OL]. [2025-10-31]. https://www.gov.cn/yaowen/liebiao/202503/content_7013163.htm.

[5] 杨治良, 郝兴昌. 心理学辞典[M]. 上海: 上海辞书出版社, 2016.

[6] 斯蒂芬·戴维斯. 音乐的意义与表现[M]. 长沙: 湖南文艺出版社, 2007.

[7] 康健, 王海龙, 苏贵斌, 等. 音乐情感识别研究

综述[J]. 计算机工程与应用, 2022, 58(4): 64-72.
[8] 陈晓鸥, 杨德顺. 音乐情感识别研究进展[J]. 复旦学报(自然科学版), 2017, 56(2): 136-148.
[9] Yang Y H, Chen H H. Machine recognition of music emotion: A review[J]. ACM Transactions on Intelligent Systems and Technology, 2012, 3(3): 1-30.
[10] Hevner K. Experimental studies of the elements of expression in music [J]. The American Journal of Psychology, 1936, 48(2): 246.
[11] Mehrabian A, Russell J A. An Approach to Environmental Psychology[M]. Cambridge, Mass.: M.I. T. Press, 1974.
[12] 张洪亮, 王立平, 张旭. 我国神经科学发展现状和未来战略思考[J]. 中国科学基金, 2021, 35(2): 328-338.
[13] 吴澹宁. 我国大学生数字素养框架研究[J]. 中国广播电视学刊, 2024(2): 25-30.
[14] Ahmad Z, Khan N. A survey on physiological signal-based emotion recognition [J]. Bioengineering, 2022, 9(11): 688.
[15] 黄潇婷, 鞠东辰, 吕兴洋. 神经科学方法在旅游研究中的应用: 一项系统综述[J/OL]. 旅游科学, 2024: 1-21. (2024-10-08). https://doi.org/10.16323/j.cnki.lykx.20240930.001.
[16] 权学良, 曾志刚, 蒋建华, 等. 基于生理信号的情感计算研究综述[J]. 自动化学报, 2021, 47(8): 1769-1784.
[17] 李洪伟, 李海峰, 马琳. 基于动态脑网络的长时音乐情感研究[J]. 复旦学报(自然科学版), 2020, 59(3): 330-337.
[18] 喻一梵, 乔晓艳. 基于深度学习算法的正负性情绪识别研究[J]. 测试技术学报, 2017, 31(5): 398-403.
[19] Chan L P, Livingstone S R, Russo F A. Facial mimicry in response to song [J]. Music Perception, 2013, 30(4): 361-367.
[20] 李路宝, 陈田, 任福继, 等. 基于图神经网络和注意力的双模态情感识别方法[J]. 计算机应用, 2023, 43(3): 700-705.
[21] 音乐情感主观评价问卷[EB/OL]. [2025-10-31]. https://jsj.top/f/ry8Xbv.
[22] Mas-Herrero E, Marco-Pallares J, Lorenzo-Seva U, et al. Individual differences in music reward experiences [J]. Music Perception, 2013, 31(2): 118-138.
[23] 邓丽, 陈波, 庞茜月, 等. 脑电技术在文化创意产品情感化设计中的应用[J]. 图学学报, 2018, 39(2): 327-332.
[24] Dennis T A, Solomon B. Frontal EEG and emotion regulation: Electrocortical activity in response to emotional film clips is associated with reduced mood induction and attention interference effects [J]. Biological Psychology, 2010, 85(3): 456-464.
[25] Wagh K P, Vasanth K. Performance evaluation of multi-channel electroencephalogram signal (EEG) based time frequency analysis for human emotion recognition [J]. Biomedical Signal Processing and Control, 2022, 78: 103966.
[26] ErgoLAB Dry-EEG 干电极脑电仪-津发科技[EB/OL]. [2025-10-31]. https://www.kingfar.net/products/ergolab-dry-eeg-device.
[27] ErgoLAB HUB 穿戴式多模态生理记录仪-津发科技[EB/OL]. [2025-10-31]. https://www.kingfar.net/products/ergolab-hub.
[28] Bradley M M, Lang P J. Affective reactions to acoustic stimuli[J]. Psychophysiology, 2000, 37(2): 204-215.
[29] 杨钟亮, 陈育苗. 基于 GGA-Elman 网络的头部体态语言 sEMG 识别[J]. 智能系统学报, 2014, 9(4): 385-391.
[30] 季虎, 孙即祥, 林成龙. 基于离散平稳小波变换的心电信号去噪方法[J]. 计算机应用, 2005, 25(6): 1318-1320.
[31] Wang J, Tang L, Bronlund JE. Surface EMG signal amplification and filtering [J]. International Journal of Computer Applications, 2013, 82(1): 15-22.
[32] 陈立伟, 黎秀红. 基于独立成分分析的脑电伪迹消除方法[J]. 西南科技大学学报(自然科学版), 2004, 19(2): 103-106.
[33] Popivanov D, Mineva A. Testing procedures for non-stationarity and non-linearity in physiological signals [J]. Mathematical Biosciences, 1999, 157(1-2): 303-320.
[34] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995.
[35] Dragomiretskiy K, Zosso D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
[36] Li X, Yan Z, Gong P, et al. Research progress on emotion recognition based on electroencephalogram signals [J]. Biomedical Signal Processing and Control, 2026, 113: 109188.

(责编&校对　雷建云)