

「人工智能」

智能音乐创作系统的疗愈价值及神经心理的机制研究 ——立足于青少年音乐教育与音乐疗愈课堂

曲达志

首都师范大学音乐学院
北京 100048

摘要：随着全球青少年心理健康问题的日益凸显，如何把音乐疗愈融入日常中小学课堂的音乐教育中已成为跨学科研究的重要方向。本研究立足于青少年音乐教育与音乐疗愈课堂的交叉场域，构建并验证了一套名为“育愈和鸣”的人工智能音乐创作系统。本文的讨论和研究首先从神经音乐学、多模态情绪识别理论发展心理学的视角，生动系统地阐述了音乐疗愈课堂的运行理论基础与作用实践机制；其次，本研究设计并实现了一类融合视觉情感计算与声学情感分析的多模态情绪感知引擎，另基于 Transformer 架构的可控音乐生成引擎，形成了“情绪识别—特征映射—自适应音乐生成”的闭环干预框架；在此基础上，提出了“双源并行、动态干预”的课堂设计理念，架构了“核心曲库—扩展曲库—AI 生成语料库”三位一体的系统引导性曲库，构建了音乐特征与 PAD 情绪间维度的量化控制、映射；本研究的最后，通过包含了自我报告、生理测量与行为观察的多模态情绪测量实验，验证了此方案在捕捉青少年情绪状态方面的一致性和有效性。

关键词：音乐疗愈课堂；人工智能音乐创作；多模态情绪识别；神经音乐学；青少年心理健康；音乐教育

中图分类号：J6

一、引言

1. 音乐疗愈与音乐教育相结合的概述与现状

从 2020 年伊始，全球青少年的心理健康问题已逐渐演变为一项重大的公共卫生挑战。世界卫生组织（WHO）2022 年的年度报告中表示，焦虑和抑郁是现阶段较为主要的两大心理障碍，呈现出普遍化和低龄化的特征^[1]。立足于青少年音乐教育中出

现的部分心理、生理问题，本文通过 AI 赋能的音乐疗愈手段加以干预，旨在构建一个“人工智能音乐创作系统的疗愈及神经心理机制”的理论模型与“音乐疗愈课堂”实践框架。

2. 音乐疗愈与教育交互作用的必然性：从多层面讨论

在技术理论层面，本文将力求探索人工智能赋能下音乐创作系统在青少年的音乐疗愈与教育中新型的应用范式，在深入研究神经心理机制后，

为音乐教育和疗愈结合的新发展提供一定的科学理论依据。“音乐疗愈课堂”这一实践框架概念新颖，可操作性强，同时融合了音乐教育学、发展心理学、神经教育学、音乐艺术感知等多个学科领域的交叉，具有较大的研究空白。

在方法创新层面，笔者将成为结合音乐疗愈、教育以及神经科学评估结合的先行者，使用神经科学研究法引入传统的音乐教育评估体系，并通

通讯作者：曲达志 **邮箱：**2241102078@cun.edu.cn

收稿日期：2025-12-09 **录用日期：**2026-04-21

DOI: <https://doi.org/10.58244/sha.263969>

过人工智能赋能下的音乐疗愈进行相关干预与治疗。

在实践与创新层面,尝试开发与教育一线深度融合的AI音乐创作系统,使得文中一切的探讨和设想得以落地生根,在教育场景中扎扎实实地进行实施。

在内在机制层面,笔者通过多学科交叉的研究方法深入探索人工智能音乐创作系统如何影响青少年心理状态的内在机制,调节情绪的有效性以及时效性。为构建科学的音乐教育评价体系 and 个体学生的音乐疗愈方案提供理论支撑。

二、学理支撑与理论构建

1. 音乐疗愈的范式演进与理论源头

音乐作为一种疗愈力量的历史源远流长,其理论建构经历了从“现象学描述”到“实证科学验证”的模式转变。

现代音乐疗愈的理论体系通过多方研究的讨论与证实,确立了其建立在“心理动力学理论”、“行为主义理论”、“人本主义理论”和“神经科学理论”四大支柱之上。整体学科以及行业发展轨迹也呈现出从“辅助性手段”到“独立性干预”的学科化进程。值得一提的是,本文中探讨并提出的“音乐疗愈课堂”概念与整个的学科发展潮流相适应,并且交叉融入了音乐教育、神经心理学等学科。

2. 神经音乐学视角下的机制阐释

神经音乐学是研究音乐疗愈与音

乐教育绕不开的学科,这一特殊神经学科的兴起为音乐疗愈提供了坚实的生物学基础。经脑电研究发现,音乐刺激可以通过多感觉整合通道作用于中枢神经系统,产生复杂的“神经可塑性”变化^[2]。

本文的研究中心及重点紧紧围绕“音乐疗愈课堂”这一概念,这是一种基于教育神经科学原理的创新型教育范式,能够通过结构化音乐干预将音乐元素系统地融入教育领域,实现“认识——情感”的整合发展。这一概念超越了传统的审美教育范畴,构建了一个多维度的心理发展促进系统。

课程内容架构遵循发展适应性原则,确保音乐干预与学生认知发展阶段和心理需求相匹配。同时考虑到课堂属性,在未来,发展完善的理想化“音乐疗愈课堂”应承担一定的教育任务,通过对课本内音乐的节奏、调式、和声等进行教学,同样的,这些音乐要素可以作为人工智能音乐创作系统下的可控音乐参数,实现特定的心理调节目标。其作用机制模型在本文中

主要介绍两种,一种是针对课堂的框架设计,另一种是人工智能赋能下音乐创作系统大模型以及语料库的训练原理。基于生物心理社会模型,笔者构建了“音乐疗愈课堂”的多重通路作用模型。

基于深度学习的方法是一种特殊的机器学习方法,可以通过多层神经网络来进行情绪识别,例如卷积神经网络(Convolution Neural Networks,CNN),亦或是循环神经网络(Recurrent Neural Network,RNN)^[3]。上述方法需要大量的数据和计算资源进行大模型训练,但是有部分内容可以自动提取特征和模式。

卷积神经网络(Convolution Neural Networks,CNN)是深度学习的一种重要应用和手段,由于深度学习的本质是一种仿生思想,也就是通过构建深层的神经网络来模拟现存生物的神经系统的结构与功能。

人工神经网络是上文中提到的深度学习模型下的核心计算架构,其基础结构如图1所示,本文探讨和研究

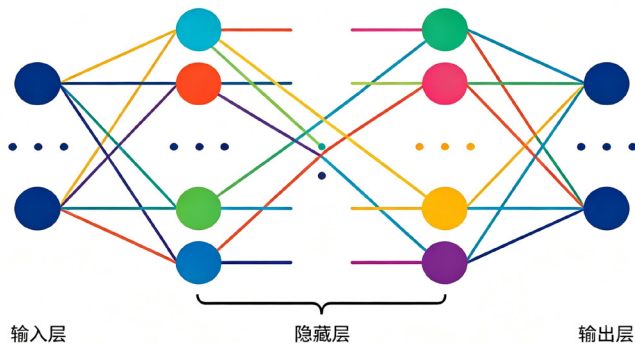


图1.

中所涉及的各种算法均由此计算架构演化而来。依据这一理论的提出人霍普菲尔德(Hopfield)的观点,一个典型的神经网络包含输入层、隐藏层和输出层三点。

在最新研究中,情绪数据采集与采集分析技术已取得了显著的阶段性成果。尤其是在实时监测领域,动态情绪识别与监测技术的准确性与效能均得到了长足的进步。并且值得注意的是,Scheibe Susanne等人通过实验研究了不同个体佩戴口罩对心理监测数据的影响后,发现面部信息的部分缺失对情绪识别的结果干扰程度是比较有限的,这一较为颠覆以往认知的发现为在自然日常环境下获取相对有效的生态学术数据提供了至关重要依据。

在动态情绪识别理论方法与分析层面,相关领域的学者们提出了纷繁复杂且有效的分析结构性框架。Johnston开创性地采用正负半轴区分,分别表征愉悦与悲伤的情绪,并通过相应的强度指标实现了情绪量化处理^[4]。在这一理论的基础上,Russell和Mehrabian进行了一定程度的优化,提出了包含“愉悦度-唤醒度-控制度”的三维评估体系^[5],目前已成为各分支情绪研究的主流范式之一,这一评估体系也是本文探讨与研究的评价体系中最为核心的一环^[6]。除此之外,Ortony提出的OCC模型和Picard引入的隐马尔可夫模型法,也从不同角度丰富了情绪识别的研究维度,隐约呈现了一种百家争鸣的态势。

3. 多模态情绪识别的基础理论

音乐疗愈课堂的情绪识别建立在情感计算与心理生理学的交叉学科基础之上,现代情绪科学认知体系中认为,情绪是包含了“主观体验”、“生理唤醒”、“表情行为”、“认知评价”的多成分过程^[7]。基于此,笔者在本文的探讨之中构建了“多模态情绪识别框架”的雏形,通过“数据融合”技术实现对情绪状态的全面捕捉与量化分析。

4. 情绪识别技术分类

基于介入性考虑,情绪识别技术主要分为非介入性情绪识别和介入性情绪识别两种。介入性的情绪识别系统是指进行情绪识别时需要直接接触人体,例如通过监测人体的生理反应(例如心电图、皮电反应等)来判断其情绪状态。这种技术通常需要使用一些生物传感器或生理监测设备来收集数据,在一些团体的认知当中,这都会对人体造成一定的影响和干扰,导致排斥的负面情绪和行为。

非介入性的情绪识别指的是,在不需要直接接触任何人体部位的情况下,通过分析肢体语言、面部肌肉运动、个体声音、全局动作姿态、部分生理反应等多种信号来识别人类动态情绪状态。

(1) 生理信号的情绪识别路径(介入式)

笔者在结合研究国内外矢量研究后发现,自主神经系统活动监测(介入式)是情绪识别的重要技术路径,可以通过几项身体结构的数值得以分析:

心血管信号分析:通过心电图和

脉搏波进行数据采集,用脉搏传导时间的变化来反映交感神经激活程度,以心率加速能力作为情绪唤醒度的客观指标。

皮电反应监测信号分析:通过导电的水平与导电的状态、数值来反映情绪的波动,以SCL基线值反映长期的情绪波动以及状态倾向^[8],SCR恢复时间反映情绪调节的能力^[9]。

(2) 生理信号的情绪识别路径(非介入式)

笔者在结合研究国内外矢量研究后发现,自主神经系统活动监测(非介入式)是情绪识别的重要技术路径,可以通过几项身体结构的数值进行分析:

面部表情分析:基于正态面部动作编码系统,这种系统与技术广泛出现在游戏角色影像(NBA2K、FIFA等)以及电影特效中的动态面部捕捉中,例如《指环王》中的咕噜,主要依托于计算机视觉技术实现。动作单元的识别与组合分析,可以分配到类似token的单元节,如AU4(皱眉)AU5(嘴角上扬)与积极或消极的情绪相关^[10]。

面部肌肉运动速度分析:包含肌肉移动的速度与幅度能够量化并反映情绪强度。

微表情捕捉分析:在心理学中,微表情一直被广泛关注,尤其在一些特殊的机构面谈中,更能体现被观察者的心理活动,对细节的观察和分析更能揭示个体企图隐藏的真实情绪。值得一提的是,此类技术的出现,提高了非介入式情绪识别的效率,人的

表情是复杂的运动和活动，表面的内容不容易完全的、或完全不会展示内心的动态，微表情捕捉分析技术或将对非介入式的技术产生重大的提高。

体态与手势分析：旨在关注非言语行为的情绪信息，这与微表情的分析有类似的部分，但并非完全重合。身体姿态的开放性与情绪效价高度相关，动作的流畅度反应情绪状态的一致性，受试者手势的空间范围大小与情绪唤醒度呈正相关。

语音特征分析：主要侧重于提取韵律特征作为情绪指标，人声频率中的基频均衡值和范围与情绪唤醒度相关，在音乐疗愈课堂中，在校青少年（受试者）在表达时的语速、停顿模式、音质特征（共振峰、抖动度）可以反映情绪的稳定性与相关信息。

三、“育愈和鸣”人工智能创作引擎

本文中所讨论的音乐疗愈属性的人工智能音乐创作系统，是采用微服务构架以实现高内聚、低耦合与弹性扩展的原理，是一套人工智能自适应性的靶向音乐创作系统，考虑到本系统最终面向的是青少年音乐课堂，所以笔者将其命名为“育愈和鸣”人工智能创作引擎（后文称“和鸣引擎”）。

1. 总体架构与设计哲学

和鸣引擎人工智能自适应靶向音乐创作系统在笔者的设计中并不是单一的工具软件，而是深度融合“音乐疗愈课堂”理念生态的“智能干预框架”，

本文中的“疗愈”一词与传统的医学系统中的“治疗”、“治愈”均不同，更注重精神上的干预，而不是希腊语中的“φυσικός (physical)”，尤其在青少年群体中，更多的只是简单的情绪波动，若不加以干预则会对成年后的自我认知、情绪调节、额叶发育造成影响，但实际上尚未有恶劣的现状^[11]。

此架构设计的核心要素可以具体表现为三大核心原则：普惠范式教育、神经元素与 AI 协同、闭环感知与自适应驱动。

超低门槛的教育普惠：和鸣引擎的重要思想是拆除音乐创作技术壁垒，借助人工智能赋能下的音乐创作系统，让基于任何学习背景的青少年儿童都可以创作音乐，并且比以往的音乐创作生成系统更适配偏日常化的口述表达描述，生成音乐的歌词、调式调性、编配更倾向于疗愈类音乐。将学生的非结构化输入（日常用语表达）精准

解析并映射为音乐生成的控制信号，力求确保教育的公平性，使音乐表达成为所有学生均可触及的情感出口。

以精神元素协同 AI 确保干预程度：为实现真正有效的疗愈干预，生成的音乐需要具备足够的艺术性通识且兼具科学上的靶向性，基于深度学习的生成模型可以在采集到学生的心理动态变化情况并进行创作，结合传统音乐的调式调性、和声走向等专业内容，输出合乎乐理、契合教学、又足以精准服务于青少年宣泄、镇静、愉悦等特定疗愈目标，实现了创作的自由度与干预精准性的统一。（图 2）

2. 技术支持与实现路径

“和鸣引擎”的核心功能由两大关键模块协同实现：多模态情绪感知引擎（系统的“感官”）与可控音乐生成引擎（系统的“创造核心”）。其协同工作流程如下方的系统数据流程图所示：（图 3）

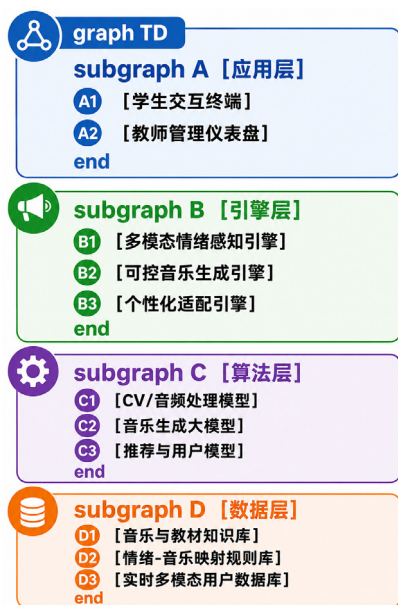


图 2.



图 3.

下面，我们将对图中各个模块进行详细阐述。

(1) 多模态情感感知引擎

该引擎负责将课堂环境中非结构化的用户信号，转化为量化的、可计算的情绪状态向量 $E=(valence,arousal)$ ，为音乐生成提供精准的靶向目标。

(2) 视觉情感计算通道

输入：课堂摄像头实时视频流。

预处理：使用 MediaPipe Face Mesh 进行实时人脸检测与面部几何归一化，以消除头部姿态偏移带来的影响。

核心模型：采用在大型情感数据集 AffectNet 上预训练的 DeepAffectNet 或 Vision Transformer(ViT) 模型。考虑到课堂部署的实时性要求，可对模型进行知识蒸馏，得到一个轻量化的学生模型，在保证精度的同时大幅提升推理速度。

输出：从每帧图像中提取的面部表情特征，并基于时序卷积网络或 Transformer 编码器对视频片段的特征序列进行建模，最终输出一个平滑的、代表短期情绪状态的效应唤醒度向量。

(3) 声学情感分析通道

输入：学生语音（如在课堂分享环节）。

特征工程：提取 egEMAPS 标准声学特征集，包括基频、能量、频谱、频率微扰等。

输出：对语音片段的整体情绪进行分类与维度回归。

(4) 数据融合与决策层

策略：采用基于注意力机制的晚期融合。分别为视觉和声学模态的特

流程图 T0

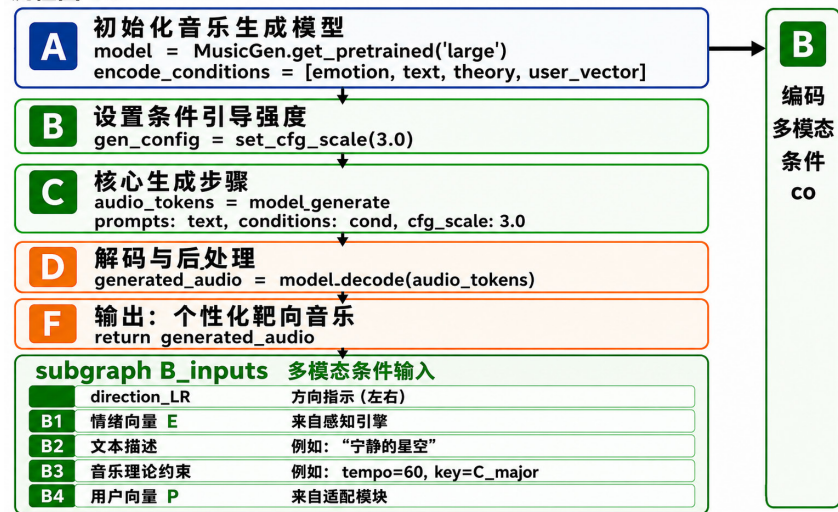


图 4.

征向量训练一个注意力网络，该网络能动态计算在特定时刻、针对特定用户，各模态信息的可信度权重。

公式简化示意： $E_{final}=\alpha * E_{visual}+\beta * E_{acoustic}$ （其中 α 和 β 由注意力网络动态生成）。

优势：此类融合方式下的系统得以在面对某单个模态出现信号质量差、灯光昏暗等问题时，可以自动降低其权重，以此来保证整体动态情绪估计的鲁棒性。（图 4）

3. 可控音乐生成引擎

这是“和鸣引擎”的灵魂，其任务是在多重条件约束下，生成具有特定疗愈与教育目标的音乐。

基础模型：选择 Meta 的 MusicGen 或 Google 的 MusicLM 这类基于 Transformer 架构的自回归语言模型。它们将音频离散化为声学 token，将音乐生成问题转化为序列预测问题，成熟度高，生态完善。

领域自适应：为实现“育愈”目

标，必须对基础模型进行监督微调。

微调数据集：构建一个专用的“教育-疗愈”音乐语料库。该库不仅包含大量按情绪标签（愉悦、悲伤、平静、兴奋）分类的音乐，更重要的是整合了人教版中小学音乐教材的全部曲目及分析（如标注出《春江花月夜》的宁静、《金蛇狂舞》的欢腾）。

“育愈合鸣”系统从实验室走向教室，笔者认为其核心挑战在于如何与现行的国家课程体系深度融合，实现人工智能赋能下“疗愈”与“教育”的共融共生、共荣共升。

4. 课堂设计理念与课程整合框架

“音乐疗愈课堂”的开展宗旨将秉承着“双源并行、动态干预”这一设计思路与理念。“源”字取自“问渠那得清如许，唯有源头活水来”的诗句，该理念旨在统一教学时空，达成两个层面目标的协同发展。

其一为教育源，课堂将严格遵循并完成《义务教育艺术课程标准

(2022 年版)》所规定的课程目标, 包括核心素养中的“审美感知”、“艺术表现”、“创意实践”、“文化理解”。在此源中, 确保系统的介入具备合法的教育身份与意图, 是日常教学活动开展的“本体”和“载体”。

另一为疗愈源, 即利用“和鸣引擎”的多模态情绪感知能力, 在学生进行创作、欣赏、表达音乐的过程中, 无感化地实时 (real time) 捕捉受试者的情绪状态, 进而通过 AI 生成个性化的音乐内容, 进行无痕的情绪疏导、压力缓解与心理韧性培养。此环节主要注重前文中所提到的、系统带来的“赋能”与“增值”。

5. 多模态情绪识别在课堂中的无感化部署

若要将动态识别与监测技术引入我国的基础教育课堂内, 必须在有效性、合规性与伦理之间取得微妙平衡, 中学生尤其是高中生的学习生活是紧张且有一定压力的, 所以也会有一定阻力。

轻量化与无感化的技术路径

视觉模态: 群体情绪基调分析

硬件部署: 在教室后方或侧后方部署单个广角高清摄像头, 其视野覆盖大部分学生区域。设备固定安装, 不进行人脸追踪或特写拍摄。

算法策略: 采用在 FERplus 和 AffectNet 数据集上预训练的、经过剪枝的轻量化卷积神经网络 (如 MobileNetV3-Small)。模型不进行个体身份识别, 而是将整个画面或划分为若干区域的画面作为输入, 直接输出群体性

的情绪概率分布。

声学模态: 环境音韵律分析

硬件部署: 利用教室现有的多媒体系统麦克风或部署定向麦克风, 用于采集课堂环境音, 如集体朗读、小组讨论、学生发言等声音。

算法策略: 从音频流中实时提取梅尔频率倒谱系数 (MFCCs)、基频 (F0) 和能量 (Energy) 等声学特征。使用一维卷积神经网络 (1D-CNN) 分析这些特征的时序变化, 将其映射为课堂氛围的唤醒度指标。例如, 持续高能量、高基频的环境音可能对应较高的集体唤醒度。

输出结果: 声学模态的分析结果主要用于交叉验证和丰富视觉模态得出的 'A_class' (唤醒度) 数据, 增强系统判断的鲁棒性。

上述多模态技术的部署架构与数据流如下所示:

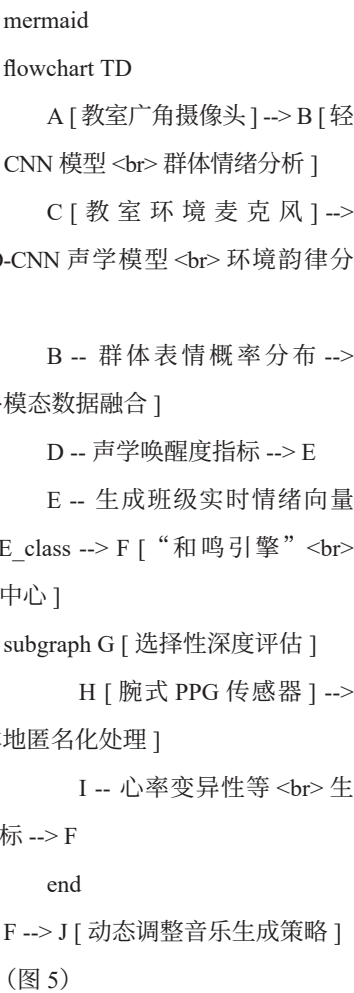


图 5

四、系统引导曲库的构建与音乐特征的分析

“和鸣引擎”的疗愈效力与效度不仅依赖于精准的情绪识别，更根植于其核心，即经过科学设计与严密构建的音乐知识库（曲库）在不同的教学场景中能及时调用适配的音乐素材。传统的曲库模式运用的多是人工标签进行划分。本系统摒弃了这种分类模式，转而采用“音乐特征 - 情绪映射 - 教育适配”三位一体的量化分类建模法，使人工智能赋能下音乐创作系统兼具艺术性与科学性。

1. “三位一体”的曲库构建模型

本研究构建的曲库由三个层级递进、支撑，架构起三维的构建场域：

核心曲库（教育锚定层级）：这一层级采用教育部审定的、人教版中小学音乐教育教材中的全部曲目为基石。在使用过程中，通过对这些经典曲目进行音乐信息检索与分析（Music In-



图 6. 情绪唤醒度标轴示意图

formation Retrieval, 下文中统称 MIR) 提取其客观声学特征，逐步建立起符合新课标要求的“标准特征曲库”，确保该系统音乐的文化正统性与教育适切性，确保适合面向青少年学生进行音乐教育和疗愈活动。

扩展曲库（疗愈深入层级）：这一层级广泛地纳入已经过心理学实验或临床实践证实的、具有情绪调节效用

的音乐以及符合疗愈活动的音乐，涵盖特定的音乐风格，包括但不限于古典乐、声景学作品、噪音音乐、电子音乐、现代音乐（现代派、Pop、爵士乐等）以及中国民族器乐作品等。这些作品同样经过 MIR 分析，与核心曲库一同构成了更为广泛的“疗愈音乐特征空间”。

AI 生成语料库（动态 real time 创造层级）：这一层级是“和鸣引擎”创造性的核心。将前两层曲目的声学特征与通过大量文献研究、实验建立的 PAD 情绪标签进行关联与映射，形成一个庞大的“情绪 - 音乐”参数对应规则集合。这个语料库不存储具体的音乐文件，只存储用于指导 AI 生成的方法参数。

2. 音乐特征与情绪的量化映射

为了实现精准可控的生成，笔者在本研究定义了以下核心音乐特征维度，并且建立了其与 PAD 情绪维度的典型映射关系（如下表），这一系列的

表 1

速度 (Tempo)	单位时间拍数 (BPM)	唤醒度 (A)	低 (<80)：诱导放松、冥想 中 (80-120)：轻度愉悦 高 (>120)：提升活力、兴奋
调性 (Tonality)	大调 / 小调 / 五声调式	愉悦度 (P)	大调类：关联积极明亮的情绪 小调类：关联忧伤深沉的情绪
和声 (Harmony)	和声进行的协和度与复杂度	愉悦度 (P) / 控制度 (D)	协和进行：带来稳定感 不协和进行：制造张力与悬念 可用于宣泄调节后的解决
音高与旋律 (Pitch、Melody)	音高范围 旋律线走向	愉悦度 (P) / 唤醒度 (A)	上行旋律：常伴随情绪上扬 下行旋律：常伴随情绪释放

特征参数即为“育愈合鸣”引擎进行条件话音乐生成时的控制旋钮。(表 1)

3. 唤醒度衡量探讨

唤醒度这一概念是反应情绪激活水平的一个重要的概念，描述了受试者从沉睡到兴奋的连续状态。(图 6)

五、课堂效果评价流程设计

为科学验证本文探讨的研究是否准确且科学，在“育愈合鸣”引擎下的“音乐疗愈课堂”是否具有较高的有效性，笔者构思形成了“设计 - 实施 - 评估 - 优化”的闭环库式课堂评价流程，利用混合方法评估模型，融合量化数据与质性洞察，兼顾及时的反馈和成效，从多个维度对系统的疗愈效果与教育价值进行综合评判。

1. 混合方法评估模型框架

本文的研究和探讨的创新之一，在于提出了一整套融合了主观报告与客观数据的多模态情绪测量方案，这套方案的有效性是“育愈合鸣”系统实现精准干预的逻辑前提。(图 7)



图 7.

(1) 实验设计

本实验的核心主旨是量化地回答下列问题：

基于生理信号（心率、血氧）和行为数据（面部表情、微动作）的情绪识别结果，是否与个体自我报告的主观情绪感受相一致？

三种测量方式各自有何种优势和局限？是否能够有效互补并共同构成一个更具鲁棒性的情绪判定系统？

参与本实验的受试者定位为北京市景山中学大兴分校中学部初二和高二的在校生共 60 名（其中男女各 30 名），平均年龄 15.5 岁（SD=1.5），体现出样本年龄分布相对集中，是一个比较理想的、同质性较高的实验群体。所有受试者视力均正常（含矫正后），听力正常、无严重身心疾病史。实验之前均获得了本人、班主任以及家长的书面知情同意书。

由于课堂实践时间相对有限，8 名参与受试者的数据因采集原因以及设备故障导致不完整而被剔除，最终有效数据为 52 份。

测量工具中，生理测量板块选用华为 Band7 智能手环，考虑到中学生普遍无法在校内使用智能手机，所以老师携带平板电脑与手环通过蓝牙连接，以每分钟一次的频率采集静息心率和经过创作后的心率数据进行采集、对比。

每次生理测量都有相对规范的变量控制，被测试学生应静息三分钟，左手佩戴手环，手腕朝上，身体倚靠桌椅靠背，心脏与手臂横截面呈就



8.

九十度夹角，以得到最精准的心率数据采集结果。

避免测量前的过度运动、情绪波动、激烈言语等影响静息检测结果的一系列行为，从而得到更具有科学研究意义的数

据。实验流程如图 8：

实验教室选用北京市景山学校中学部校区中三楼的音乐教室，实验在 2025 年 11 月 27 日下午第三节课正式开展。

准备阶段（4 分钟）：由引导人（老师为主）向课堂内受试同学简单介绍实验的流程，准备 10 个左右手环，在整个实验开始前，受试同学佩戴好智能手环，并确保摄像头所处的角度足够清晰地捕捉其面部。首先在教室内播放相对舒缓的音乐，参与受试者

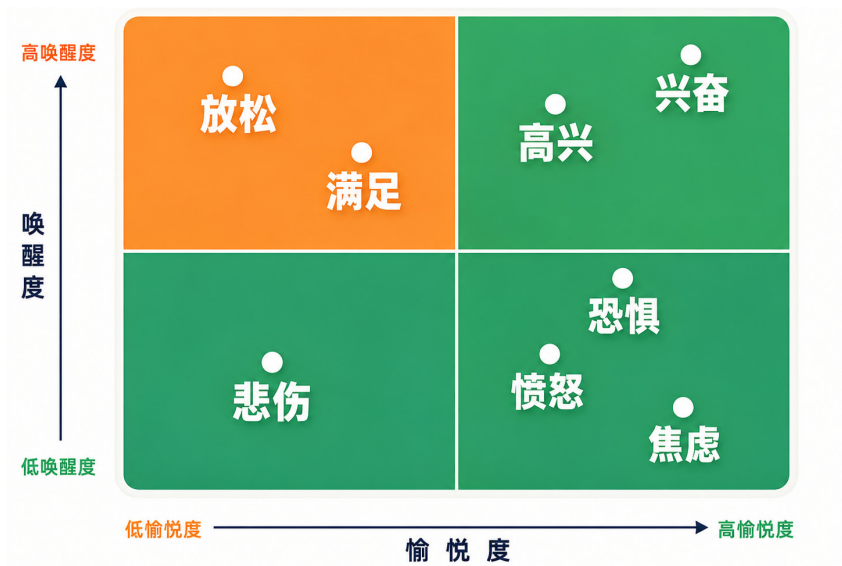


图9. PAD 情绪模型：愉悦度 - 唤醒度二维平面图

需静坐并采集3分钟内的静息心率、血氧作为基线。

基线情绪报告（2分钟内）：完成第一次PAD情绪自我报告。

情绪诱发与数据采集阶段（12分钟内）：播放本节课的赏析乐曲，观察在播放期间受试者的面部表情和同步心率。

最终报告（2分钟内）：在整个视频播放结束后参与者针对整个流程的感受完成一次PAD情绪自我报告。

(2) PAD 简述

上文中多次提到的PAD可能会给读者带来一定的困惑，这里特别做一定的解释和介绍。PAD是一个经典的情绪模型，在这个模型中所有情绪都可以通过三个核心维度来精确地进行描述和量化，这三个维度分别为：

P= 愉悦度 (Pleasure-Displeasure)

A= 唤醒度 (Arousal-Nonarousal)

D= 控制度 (Dominance-Submissiveness)

如果用北斗卫星导航系统来举例，PAD可以理解为北斗卫星的自有绘制图，任何一种情绪状态的动态分级都可以在这个三维空间中找到属于自己的坐标。

针对整个的“育愈合鸣”引擎，坐标系的建立可以附带PAD属性，例

如PAD分别对应了0.8、0.6、0.7三个数值，那么引擎就可以学习在这一区域坐标附近的情绪可以使用的相关音乐，这使得音乐生成有了明确的、可量化的目标。（图9）

2. 数据处理与分析

整体的生理和数据依据本研究制定的PAD规则映射至坐标系空间，下表中展示了愉悦度和唤醒度在三中测量方式中的差值，可以更直观的了解整体的测量相关参数与系数，并通过坐标系具体的表现出来。（图10）

3. 结果分析与讨论

实验证明，“高兴”、“焦虑”在青少年受试者中是最常出现的两种情绪。本讨论将着重对比这两者的结果。

一致性分析：生理测量和行为观察的结果与自我报告基准在愉悦度（P）和唤醒度（A）两个核心维度上均呈现出高度显著的正相关性，即所

情绪唤醒度测量方式与评分标准



图10. 情绪唤醒度测量方式与评分标准

有 $p < 0.01$ 。因此,这一结果有力地证明了通过客观数据进行反向推演的情绪状态与个体的主观感受具有良好的 consistency。也从实证角度验证了本研究所涉及的综合性情绪测量方案的高度有效性质。综上,本文中提出的方案能够较为准确的捕捉和量化个体的动态情绪状态。

互补性分析:行为观察(非介入式)中在愉悦度(P)维度上与自我报告的相关性是最高的,这有力体现了人的表情作为情绪直观的外显通道,在识别情绪效应(积极情绪或消极情绪)方面具有极高的灵敏度和辨识度。

六、结论与展望

本研究通过系统的理论探索、技术实现与实践验证,取得了一定的创新性研究成果,主要体现于理论、技术与实践三个层面。

1. 研究结论:构建了“育愈合鸣”的交叉研究范式

本文中的探讨与研究率先将音乐疗愈、心理治疗、音乐教育与人工智能进行了深度融合,并且首次提出了“音乐疗愈课堂”这一独到理念,基于教学的属性,整个音乐疗愈过程并非仅仅是现存的中小学心理健康咨询这一性质,更多聚焦于教育中的疗愈作用,让学生在学习正常的课标教材课程的同时得到疗愈效果。

这一体系突破了传统音乐教育以技能与美育为主的限制,更深层地触及到学生的内心世界,针对在青春期、

社交困境、学习压力等环节中可能出现的问题。同样也超脱于单一的心理咨询模式,借助于人工智能的强大效能与算力,实现疗愈、教育、人工智能技术三者有机统一的理论框架。

2. 现存问题:提出了多模态情绪测量的综合性解决办法

人的情绪极具主观性与复杂性,故而本文中研究设计了一套以自我报告为基础的、融合生理测量与行为观察的多模态情绪测量方案,结合上文提到的PAD情绪空间映射规则,将不同来源的数据进行有效的整合分析与交叉验证,对青少年课堂中的情绪进行更为全面科学的量化评估,为精准的情绪干预提供可靠依据。

3. 技术与方法层面的创新范式

本研究设计并实现了基于多模态情绪感知的可控音乐生成系统,将情绪向量映射到音乐参数,实现了科技与艺术的对接,使得情绪识别到音乐生成的端口成为智能化流程。采用了条件化生成与符号相约束的结合技术路径,确保了生成的音乐具有一定的音乐性、艺术创造性,并符合音乐理论与疗愈目标,即敞开心灵与艺术大门。

4. 自述与展望

由于笔者独特的学习经历,以钢琴演奏伊始,横跨艺术与科技、音乐录音以及制作、音乐科技、音乐教育、声乐歌剧艺术指导以及人工智能音乐疗愈,因此对整个研究的发展有着独到的理解和要求。相信,当下不断发展的科技水平和音乐艺术的进步可以让更多的传统项目和科技进行深度融

合,未来的音乐疗愈行业一定会更深的融入人工智能领域,科学技术的发展与传统学科的支撑深刻交融,助力青少年的健康成长,造福普罗大众的日常生活。

参考文献:

- [1] 易敏, (2024), 全球儿童心理健康教育的问题、特点及启示,《基础教育参考》, (02), 71-80.
- [2] 林佳川, 吴佳倩, 霍岩, 张剑宁, (2025), 脑源性神经营养因子在耳鸣发病中的中枢调控作用,《中华耳科学杂志》, 23(09), 1073-1078.
- [3] 张一心, 蒋林, 李远成, 纪辰, 面向可重构结构的CNN剪枝与量化压缩方法,《计算机应用》, 1-9.
- [4] Johnston V S., (1999). Why we feel: The science of human emotions. New York: Perseus publishing.
- [5] Russell J A., (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.
- [6] Mehrabian A., (1980). Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies. Cambridge: Oelgeschlager Gunn & Hain Cambridge, MA.
- [7] 李新, 郑奕珂, 杨现民, (2024),

- 心理生理学数据的教育应用: 科学释义、实践探索与发展趋向,《现代远距离教育》, (05), 44-58.
- [8] 孙晓乐, 袁方, 陈环, (2025), 多元化认知行为干预在腹腔镜前列腺癌根治术患者围术期的应用及对FACT-P评分、SCL-90评分的影响,《全科医学临床与教育》23(11), 1051-1053.
- [9] 刘静, 刘旭薇, 付小星, 一种基于SiGe/ Si异质结势垒调控的内嵌HBT触发型SCR ESD防护机理及特性分析,《电力电子技术》, 1-13.
- [10] 路绳方, 刘莉莉, 唐震, (2025), 基于Gabor小波变换的面部表情识别及应用分析研究,《南京工程学院学报(自然科学版)》, 23(02), 60-66.
- [11] 王秋琳, 方长乐, 熊柳林, (2025), 法洛四联症胎儿前额叶皮层神经发育异常及代偿机制的解析,《第11届亚太解剖学大会暨中国解剖学会第39届(2025年)学术年会摘要汇编》(pp.433), 遵义医科大学.

作者简介:

曲达志 中央民族大学英才班传承与创新艺术指导, 首都师范大学硕士。

Interdisciplinary Approach: Theoretical Construction and Preliminary Research on the Integration of Music Therapy and Music Education Empowered by a New Type of Artificial Intelligence

Qu Zhida

(School of Music, Capital Normal University, Beijing 100048, China.)

Abstract: With the increasing prominence of mental health issues among adolescents globally, integrating music therapy into daily music education in primary and secondary classrooms has become an important direction in interdisciplinary research. This study is grounded in the intersection of adolescent music education and music therapy classrooms, constructing and validating an artificial intelligence music creation system named “Yuyu Harmony”. The discussion and research in this paper firstly elaborates on the theoretical basis and practical mechanisms of music therapy classrooms from the perspectives of neuromusicology, multimodal emotion recognition theory, and developmental psychology. Secondly, this study designs and implements a multimodal emotion perception engine that integrates visual affective computing and acoustic emotion analysis, and a controllable music generation engine based on the Transformer architecture, forming a closed-loop intervention framework of “emotion recognition - feature mapping - adaptive music generation”. On this basis, it proposes a classroom design concept of “dual-source parallel and dynamic intervention”, constructs a three-in-one systematic guidance music library of “core music library - extended music library - AI-generated corpus”, and establishes quantitative control and mapping between music features and PAD emotions. Finally, through multimodal emotion measurement experiments involving self-reporting, physiological measurements, and behavioral observations, this study verifies the consistency and effectiveness of this approach in capturing the emotional states of adolescents.

Keywords: Music therapy classroom; Artificial intelligence music creation; Multimodal emotion recognition; Neuromusicology; Adolescent mental health; Music education