丘成桐中学科学奖

研究报告

基于动态异构图神经网络的多模态生理信号实时音乐参数调控方法研究

参赛学生姓名：严恩怡、任晨曦

中学：上海星河湾双语学校

省份：上海

国家/地区：中国/上海

指导老师姓名：彭敦陆

指导老师单位：上海理工大学

2026 年 11 月 5 日

2026 年 Yau High School Science Award
仅用于 2026 丘成桐中学科学奖论文公示

基于动态异构图神经网络的多模态生理信号实时音乐参数调控方法研究

严恩怡、任晨曦

论文摘要

针对番茄钟工作法仅管理时间而无法感知专注状态变化的问题，本文提出一种基于动态异构图神经网络的多模态生理信号实时音乐参数调控方法，并设计实现了智能 AI 音乐番茄钟系统。该方法以心率、心率变异性、皮电活动和血氧饱和度为输入，将每个时间窗口的每个生理通道建模为异构图节点，通过时间边和模态边构建动态异构图序列，利用节点级与语义级双层注意力机制聚合异构邻居信息，并采用门控循环单元建模图序列的时序演化，最终通过多任务输出头预测 BPM、整体增益、低频增益、混响深度与调式五维音乐参数。为兼顾音质与实时性，系统采用预生成风格统一底板音乐，并通过参数级调整与缓存池实现近似实时的音乐匹配。离线仿真实验表明，该模型在 BPM 预测上的平均绝对误差为 1.738，调式分类准确率达 78.56%；实时性验证表明端到端延迟 P50 为 15.33ms，满足 100ms 预算。在线设备实验表明，相比固定参数音乐，动态调控显著提升了用户专注度评分（7.8 vs 6.3，$p < 0.01$）与 HRV 水平（+8.2% vs +1.3%）。

关键词：动态异构图神经网络；生理信号；音乐参数预测；实时调控

Abstract: In response to the problem that the Pomodoro technique only manages time without perceiving changes in focus states, this paper proposes a real-time music parameter regulation method based on a dynamic heterogeneous graph neural network using multi-modal physiological signals, and designs an intelligent AI music Pomodoro system. The method takes heart rate, heart rate variability, electrodermal activity, and blood oxygen saturation as inputs. Each physiological channel in each time window is modeled as a heterogeneous graph node. Temporal edges and modal edges are used to construct a dynamic heterogeneous graph sequence. A two-level attention mechanism, including node-level and semantic-level attention, aggregates heterogeneous neighbor information. A gated recurrent unit models the temporal evolution of the graph sequence. Finally, a multi-task output head predicts five music parameters: BPM, overall gain, low-frequency gain, reverb depth, and mode. To balance audio quality and real-time performance, the system uses pre-generated style-consistent base music and achieves near-real-time music matching through parameter-level adjustment and a cache pool. Offline simulation experiments show that the model achieves a mean absolute error of 1.738 for BPM prediction and an accuracy of 78.56% for mode classification. Real-time validation

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

shows that the end-to-end latency P50 is 15.33 ms, satisfying the 100 ms budget. Online device experiments show that compared with fixed-parameter music, dynamic regulation significantly improves user focus scores (7.8 vs 6.3, $p < 0.01$) and HRV levels (+8.2% vs +1.3%).

Keywords: Dynamic heterogeneous graph neural network; Physiological signals; Music parameter prediction; Real-time regulation

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

目录

论文摘要 **1**

1 引言 **6**

2 相关工作 **7**

2.1 生理信号到音乐参数的映射方法 8

2.2 生理信号建模的图神经网络方法 8

2.3 实时音乐生成与调控方法 9

2.4 现有研究的不足与本研究的定位 9

符号表 **10**

3 数据准备与特征工程 **11**

3.1 模型训练数据 11

3.1.1 数据集来源与规模 11

3.1.2 数据预处理 13

3.1.3 滑动窗口分割与特征提取 15

3.1.4 标签定义 19

3.1.5 数据集划分 20

3.2 本节小结 20

4 动态异构图神经网络模型 **22**

4.1 问题形式化 22

4.2 动态异构图构建 23

4.2.1 图的形式化定义 24

4.2.2 节点定义 24

4.2.3 边定义与权重计算 25

4.2.4 动态更新机制 26

4.3 异构图注意力网络 27

4.3.1 节点级注意力 28

4.3.2 多头注意力机制 29

4.3.3 语义级注意力 30

4.3.4 网络整体结构与输出 31

4.4 时序动态建模 32

4.4.1 模型选择：门控循环单元 32

4.4.2 输入序列构造 32

4.4.3 GRU 前向传播 33

4.4.4 输出提取 33

4.4.5 与多任务输出的衔接 34

2026 S.T. Yau High School Science Award
仅用于2026年区桐中学科学奖论文公示

4.5 多任务输出与损失函数 . . . 34
4.5.1 多任务学习框架 . . . 34
4.5.2 预测头设计 . . . 35
4.5.3 损失函数定义 . . . 36
4.5.4 损失权重设置 . . . 37
4.5.5 训练目标 . . . 38
4.6 模型训练策略 . . . 38
4.6.1 数据集划分 . . . 38
4.6.2 优化器 . . . 38
4.6.3 学习率调度 . . . 39
4.6.4 正则化方法 . . . 39
4.6.5 早停策略 . . . 40
4.6.6 实现平台 . . . 40

5 系统实现与集成 . . . 42
5.1 系统整体架构 . . . 42
5.2 前端系统实现 . . . 43
5.2.1 番茄钟计时器 . . . 43
5.2.2 生理数据接入与仿真回放 . . . 44
5.2.3 前端特征提取 . . . 45
5.2.4 界面与可视化 . . . 45
5.3 后端推理服务 . . . 45
5.3.1 REST API 设计 . . . 45
5.3.2 ONNX 模型部署 . . . 46
5.3.3 三级降级推理链路 . . . 46
5.4 音乐底板库 . . . 46
5.4.1 曲库组织与分类 . . . 47
5.4.2 AIGC 曲目生成与风格筛选 . . . 47
5.4.3 曲目推荐逻辑 . . . 47
5.5 音频引擎与实时调控 . . . 47
5.5.1 Web Audio 节点拓扑 . . . 48
5.5.2 五维参数到音频节点的映射 . . . 48
5.5.3 参数平滑过渡 . . . 48
5.5.4 交叉淡化切曲 . . . 49
5.6 数据记录与日志 . . . 49
5.7 小结 . . . 49

6 实验设计与结果 . . . 50
6.1 实验设置 . . . 50
6.1.1 实验环境与硬件配置 . . . 50

2026 Shanghai School Science Award
仅用于2026上海市中学科学奖论文公示

6.1.2 数据集与数据生成 50
6.1.3 评估指标 50
6.1.4 基线模型 51
6.1.5 训练配置 52
6.2 离线实验结果 52
6.2.1 预测精度对比 52
6.2.2 消融实验 52
6.2.3 可解释性分析 54
6.3 实时性验证 54
6.4 在线实验结果 56
6.4.1 实验设计 56
6.4.2 评估指标对比结果 56
6.4.3 结果小结 57
6.5 讨论 57
6.5.1 结果解释 57
6.5.2 可解释性分析 57
6.5.3 与现有研究对比 58
6.5.4 应用前景 58
6.5.5 局限性与未来工作 58
6.6 本章小结 58

7 结论 **59**

致谢 **65**

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1 引言

在信息过载的时代，保持深度专注已成为知识工作者的核心挑战。番茄工作法通过 25 分钟专注与 5 分钟休息的交替，为时间管理提供了结构化框架。然而，现有番茄钟工具仅解决“时间分割”问题，未能解决“专注质量”问题——用户在番茄钟内依然可能因焦虑、疲劳或外界干扰而无法进入专注状态。研究表明，音乐可以影响人的生理状态与情绪，根据用户实时生理信号动态调整音乐参数，有望辅助用户维持专注。但现有音乐播放系统大多只能播放固定列表或按预设规则切换歌曲，无法根据用户实时状态连续调整音乐参数；而直接实时生成音乐（逐个音符生成）则面临推理延迟高、音质不稳定、风格难以统一等问题。因此，如何实现低延迟、高质量、个性化的实时音乐调控，成为一个值得研究的问题。

针对这一问题，现有研究已从多个角度进行了探索，但仍存在以下不足。第一，映射路径存在“两步走”的延迟与信息损失：多数研究遵循“生理信号 → 情绪分类（离散标签）→ 查表选音乐”的范式 [1, 40]，这不仅引入额外的计算延迟，还会在情绪离散化过程中丢失连续生理信号中蕴含的丰富信息。第二，模型架构未能充分捕捉多模态信号的时空依赖：基于规则或传统机器学习的映射方法难以建模心率、HRV、EDA 等异质信号之间复杂的动态交互关系；部分基于图神经网络（GNN）的研究 [40, 42] 虽已尝试建模信号间关系，但多采用静态同构图，无法捕捉图拓扑随生理状态演变的动态特性。第三，实时音乐生成方法存在工程瓶颈：端到端生成模型虽然可以直接产生音乐，但推理速度慢、生成质量不稳定，难以满足实时交互需求；且多数研究停留在离线数据集验证阶段，缺乏真实场景下的闭环实验验证。部分研究还依赖昂贵的脑电（EEG）设备 [38]，难以在日常场景中推广。

针对上述不足，本文提出一种基于动态异构图神经网络（DHGNN）的生理信号驱动音乐参数调控方法，并基于需求设计实现了一套智能 AI 音乐番茄钟系统。其核心思路是：不直接实时生成音乐，而是预先通过 AIGC 批量生成风格统一的底板音乐库，运行时利用 DHGNN 从多模态生理信号中预测音乐参数（BPM、整体增益、低频增益、混响深度、调式），再对当前底板音乐进行参数级调整（变速、增益、混响、调式切换），并在必要时从缓存池中按 BPM 区间与调式切换曲目，从而在保证音质与风格一致的前提下实现近似实时的音乐匹配。在模型层面，每个时间窗口的每个生理通道被建模为异构图节点，通过时间边和模态边构建动态异构图序列，利用节点级与语义级双层注意力机制聚合异构邻居信息，并采用门控循环单元建模图序列的时序演化，最终通过多任务输出头预测五维音乐参数。该模型绕过情绪分类中介，直接将生理信号回归为连续音乐参数，降低了映射延迟与信息损失。

为验证上述方法的有效性，该技术被应用于智能番茄钟场景，设计并实现了智能 AI 音乐番茄钟系统。该系统通过回放预置生理样例数据集或接入真实传感器采集信号，利用训练好的 DHGNN 模型预测音乐参数，并驱动 Web Audio API 实现音乐的实时调控与平滑过渡。通过离线仿真实验与在线设备实验，进一步评估

2026 SSE YanHi 3rd School Science Award 仅用于 2026 全国高中科学奖公示

了模型预测精度、实时性以及系统对用户专注状态的干预效果。

所提方法的整体技术路线如图 1所示：从多模态生理信号的表征与特征工程，到动态异构图学习，再到闭环音乐调控与评估，构成一个完整的“感知—推理—调控”回路。

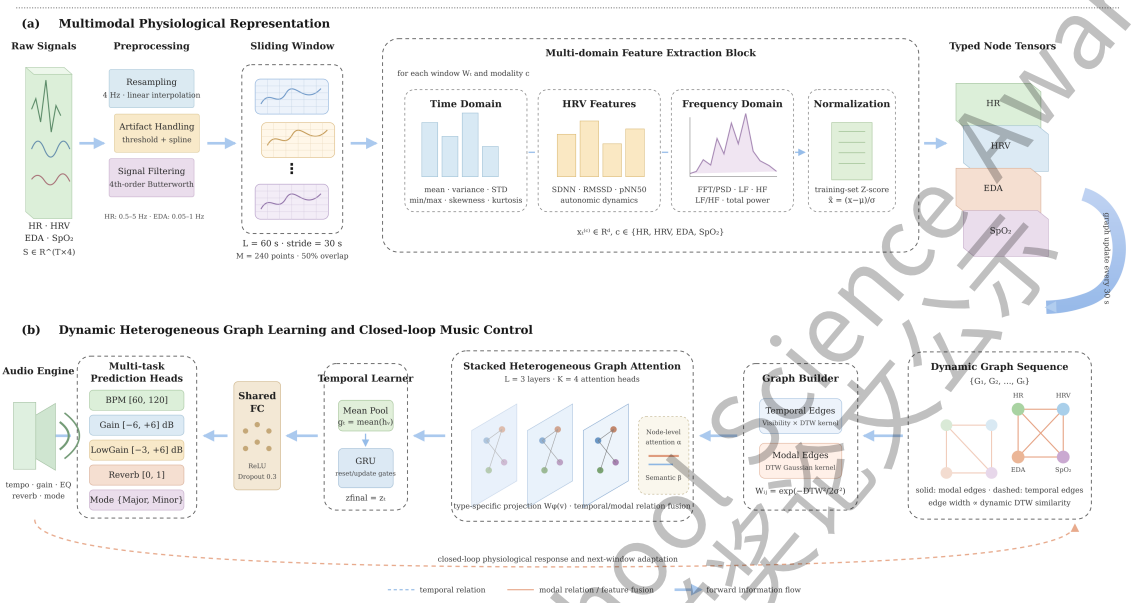


图 1 本研究提出的多模态生理表征、动态异构图学习与闭环音乐调控总体框架。

本文的主要贡献如下：

1. 提出了一种“底板音乐 + 参数级调整 + 缓存池”的实时音乐匹配思路，通过预生成风格统一的底板音乐并对其参数进行动态调整，在不直接生成音符的前提下实现近似实时的个性化音乐匹配；
2. 设计了一种基于动态异构图神经网络的生理信号到音乐参数映射模型，通过动态异构图与双层注意力机制捕捉多模态生理信号的时空依赖，并采用多任务输出头预测五维音乐参数；
3. 自主设计并实现了智能 AI 音乐番茄钟系统，将上述方法集成到真实的番茄钟应用场景中，完成了从生理信号采集、特征提取、模型推理到音频调控的端到端闭环；
4. 通过离线仿真实验与在线设备实验，评估了模型在音乐参数预测上的精度与系统实时性，并验证了系统在真实场景中对用户专注度的提升效果。

全文后续章节安排如下：第 2 节回顾相关研究并指出研究空白；第 3 节描述数据准备与特征工程；第 4 节详述动态异构图神经网络模型；第 5 节介绍系统实现与集成；第 6 节报告实验结果与分析；第 7 节讨论；第 8 节给出结论。

2 相关工作

近年来，研究者从多个角度探索了利用生理信号驱动音乐调控的方法。Yamane 等人 (2025) 在综述中系统总结了 28 项相关研究，涉及 646 名参与者，其中 53.6%

2026 STU High School Science Award 仅用于2026年成科学院论文公示

的系统使用了心呼吸传感器，60.7% 的系统使用预录音乐，28.6% 的系统使用了机器学习或深度学习模型来检测情感状态或预测音乐选择 [1]。总体来看，这一领域的工作大致沿着三条路线推进：生理信号到音乐参数的映射方法、生理信号建模的图神经网络方法，以及实时音乐生成与调控方法。

2.1 生理信号到音乐参数的映射方法

在生理信号到音乐参数的映射方面，早期工作采用单参数线性映射。RUMBA 项目通过实时修改音乐速度来响应心率变化，以诱导放松效果 [45]。这类方法实现简单，但映射维度单一，难以捕捉生理状态的复杂变化。此后，部分研究转向多参数规则映射。一项专利文献（CN122272969A，2026）描述了一种结合 HRV 实时优化的音乐治疗系统，通过滑动窗口协方差分析识别情绪调节敏感窗口，驱动 HRV-音乐特征映射网络以 100ms 间隔输出最优音乐参数，并通过混合合成引擎实时生成干预音频 [46]。此类系统引入了闭环控制思想，但其映射规则仍依赖人工设定的阈值和分段函数，缺乏自适应能力。

近年来，数据驱动的预测模型成为主流。Han 等人（2024）提出了一种结合生成式人工智能与多尺度卷积和注意力机制的模型，利用心率变异性（HRV）和皮肤电反应（GSR）等生理指标进行音乐情感识别与推荐，在 PMEmo、RAVDESS 和 Soundtrack 数据集上验证了方法的有效性 [52]。Wang 和 Serrano（2024）提出了一种基于图卷积神经网络（GCN）的个性化音乐推荐模型，通过分析心率、皮电等生理数据感知用户实时情绪状态 [40]。Şahin（2025）在达特茅斯学院的硕士论文中提出了 HeartDJ 系统，利用 HRV 生物反馈驱动 AI 生成音乐，实验发现 AI 生成的平静音乐能够显著促进副交感神经激活，其 RMSSD 中位数增加了 2.48（$p < 0.05$），但参与者同时指出短曲目和突然的过渡限制了情绪共鸣，需要更深层的个性化和动态适应 [58]。一项关于自适应音乐情绪调节框架的研究（2025）利用实时情绪反馈和音乐特征，结合机器学习和深度学习技术，在预测用户情绪方面达到了 96.7% 的准确率 [53]。Ahmadi 等人（2025）利用图神经网络从 EEG 信号中解码音乐诱发情绪，在 NMED-T 数据集上达到 87% 的分类准确率 [43]。Pal 等人（2025）使用图注意力网络从 ECG 和呼吸信号中预测血压，其中音乐特征被用作图的边属性 [42]。Wang 等人（2025）则研究了音乐信号与 ECG 信号在共享 Valence-Arousal 情感空间中的协调机制，为跨模态音乐推荐提供了新思路 [59]。

2.2 生理信号建模的图神经网络方法

在生理信号建模的图神经网络方法方面，Li 等人（2024）提出了时序-频谱图卷积网络（TSGCN），同时考虑不同时间窗口的神经振荡变化和不同脑区之间的拓扑结构，并引入最小类别混淆损失来减少主观评分与预定义标签之间的不一致性 [55]。Gao 等人（2024）提出了基于动态图注意力机制的 EEG 情绪识别框架（DGAT），通过动态图结构捕捉 EEG 通道间随时间变化的连接关系 [56]。Zhu 等人（2025）提出了 EEG-DGRN，一种用于独立于被试的 ERP 检测的动态图表示网络，在 RSVP 数据集上达到 87.05% 的平均平衡分类准确率 [57]。在异构图与

2026 S.T. Yau High School Science Award 仅用于2026年丘成桐中学科学奖论文公示

多模态融合方面，一项发表于 IEEE Transactions on Affective Computing 的研究（2025）提出了双流动态异构图循环神经网络（DHGRNN），用于多标签多模态情感识别。该模型由空时流、空谱流、融合层和多标签分类器组成，每个流包含图 Transformer 网络、演化图卷积网络和门控循环单元，分别建模多模态生理信号的异质性和相关性，并通过动态图神经网络提高了模型对缺失通道数据的鲁棒性 [54]。THSSleepNet（2025）则提出了基于时-频-空多维特征融合的异构图自适应神经网络用于睡眠分期分类。

2.3 实时音乐生成与调控方法

在实时音乐生成与调控方面，端到端生成模型代表了当前的前沿方向。Google Magenta 团队（2025）发布了 Magenta RealTime，这是一个基于 Transformer 架构的实时音乐生成模型，使用离散音频令牌进行训练，能够根据动态风格提示即时调整音乐风格 [60]。Lyria 团队（2025）提出了 “Live Music Models” 的概念，即能够实时产生连续音乐流的生成模型 [61]。Neurophone 系统（2025）则是一个集成实时 EEG 采集与动态音乐生成的脑机音乐接口，允许用户通过认知和情感状态调节音乐表达。Sargent 和 Snell（2025）提出了 “Cyborg synchrony” 框架，通过两阶段 AI 训练将人类生理响应整合到情感生成音乐中。这些工作代表了端到端实时音乐生成的前沿进展，但其推理延迟和生成质量仍难以完全满足交互式场景的需求。

2.4 现有研究的不足与本研究的定位

综合上述工作，现有研究在以下三个方面仍存在不足。

第一，映射路径存在“两步走”的延迟与信息损失。多数研究遵循 “生理信号 → 情绪分类（离散标签）→ 查表选音乐” 的范式 [1, 40]，这不仅引入额外计算延迟，还会在情绪离散化过程中丢失连续生理信号中蕴含的丰富信息。

第二，模型架构未能充分捕捉多模态信号的动态时空依赖。基于规则或传统机器学习的映射方法难以建模心率、HRV、EDA 等异质信号之间复杂的动态交互关系；部分基于 GNN 的研究 [40, 42] 虽已尝试建模信号间关系，但多采用静态同构图，无法捕捉图拓扑随生理状态演变的动态特性 [57, 54]。

第三，实时音乐生成方法存在工程瓶颈。端到端生成模型虽然可以直接产生音乐，但推理速度慢、生成质量不稳定，难以满足实时交互需求 [60, 61]；HeartDJ 研究的实验也表明，AI 生成的音乐在情感真实性和持续 HRV 改善方面仍面临挑战 [58]；且多数研究停留在离线数据集验证阶段，缺乏真实场景下的闭环实验验证。

针对上述不足，本研究提出一种 “底板音乐 + 参数级调整 + 缓存池” 的实时音乐匹配思路，并基于该思路设计了一套智能 AI 音乐番茄钟系统。与端到端生成方法不同，本研究预先通过 AIGC 批量生成风格统一的底板音乐库，运行时利用动态异构图神经网络从多模态生理信号中预测五维音乐参数（BPM、整体增益、低频增益、混响深度、调式），再对当前底板音乐进行参数级调整，并在必要时从缓存池中按 BPM 区间与调式切换曲目。在模型层面，每个时间窗口的每个生理通

2026 Yau High School Science Award 作品 2026 年度科学论文公示

道被建模为异构图节点，通过时间边和模态边构建动态异构图序列，利用节点级与语义级双层注意力机制聚合异构邻居信息，并采用门控循环单元建模图序列的时序演化，最终通过多任务输出头预测五维音乐参数。该模型绕过情绪分类中介，直接将生理信号回归为连续音乐参数，降低了映射延迟与信息损失。通过离线仿真实验与实时性验证，进一步评估了该方法在预测精度与实时性方面的表现。

符号表

为方便读者阅读，表 1 汇总了全文使用的核心数学符号及其含义。

表 1 全文核心符号汇总

符号	含义	符号	含义
$\mathbf{S}$	多通道生理时间序列，$\mathbf{S} \in \mathbb{R}^{T \times C}$	$\mathcal{G}_t$	第 t 个时间步的异构图
T	时间序列长度（采样点数）	$\mathcal{V}_t$	节点集合
C	生理信号通道数（$C = 4$）	$\mathcal{E}_t$	边集合
L	滑动窗口长度（秒）	ϕ	节点类型映射函数
S	滑动窗口步长（秒）	ψ	边类型映射函数
N	时间窗口数量	$\mathcal{A}$	节点类型集合
M	每个窗口内的采样点数	$\mathcal{R}$	边类型集合
f_s	重采样频率（Hz）	$v_{t,c}$	第 t 个窗口第 c 个通道对应的节点
d	节点特征维度	$\mathbf{x}_t^{(c)}$	节点 $v_{t,c}$ 的初始特征向量
d'	隐空间维度	$\mathbf{W}_A$	类型 A 的节点专属变换矩阵
d_{gru}	GRU 隐状态维度	$\mathbf{h}'_v$	类型特异性变换后的节点特征
$\mathbf{W}_r^{(l)}$	关系 r 在第 l 层的变换矩阵	$\alpha_{(s,t)}^{(r,l)}$	关系 r 下节点 s 对节点 t 的注意力系数
β_r	关系 r 的语义级注意力权重	K	多头注意力的头数
L	异构图注意力网络的层数	$\mathbf{g}_t$	第 t 个时间步的图级表示向量
$\mathbf{z}_t$	GRU 在第 t 个时间步的隐状态	$\mathbf{z}_{\text{final}}$	序列压缩表示（最后一步隐状态）
$\hat{y}_{\text{BPM}}$	BPM 预测值	$\hat{y}_{\text{gain}}$	整体增益预测值
$\hat{y}_{\text{low_gain}}$	低频增益预测值	$\hat{y}_{\text{reverb}}$	混响深度预测值
$\hat{y}_{\text{mode}}$	调式预测类别	λ_k	第 k 个任务的损失权重
$\mathcal{L}_{\text{total}}$	多任务总损失		

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

3 数据准备与特征工程

本节描述模型训练阶段的数据集选择、预处理流程、特征提取方法和标签定义，为第 4 节的模型训练奠定基础。系统部署后的实时数据处理流程将在第 5 节系统实现部分详细阐述。

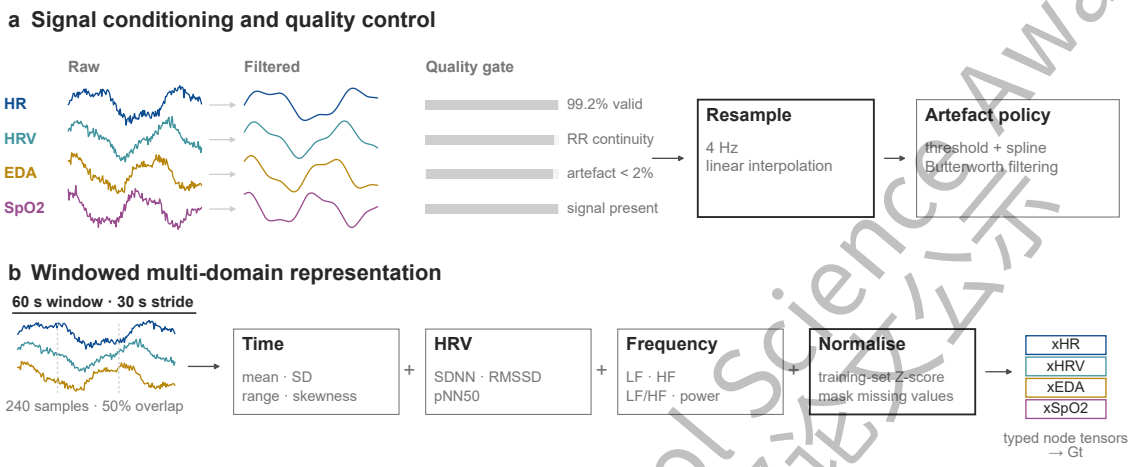


图 2 多模态生理信号的数据准备与特征工程。(a) 信号调理与质量控制。(b) 滑动窗口、多域特征及类型化节点张量的形成。

3.1 模型训练数据

本部分描述用于训练动态异构图神经网络模型的离线数据集，包括数据来源、预处理流程、特征提取方法和标签定义。

3.1.1 数据集来源与规模

在情感诱导研究中，视觉刺激（如图像、视频）是常见的实验范式。然而，参与者的主观文化背景可能显著影响其对视觉刺激的情感反应 [1]。相比之下，音乐作为一种听觉刺激，能够通过激活听觉皮层与边缘系统、旁边缘系统及新皮层结构之间的情感特异性功能连接，在情感加工过程中发挥更为广泛的作用 [11]。此外，音乐诱导范式与本研究最终的应用场景（番茄钟工作中的音乐聆听）具有天然的一致性，使得基于音乐刺激下采集的生理信号训练的模型能够更好地迁移到真实使用场景中。

基于上述考虑，本研究选用 WESAD（Wearable Stress and Affect Detection）公开数据集 [37] 作为模型训练的主要数据来源。WESAD 是目前可穿戴压力与情感检测领域引用最广泛的基准数据集之一，其设备类型和信号模态与本研究的目标应用场景高度匹配。该数据集包含 15 名被试（年龄分布 20–35 岁）的生理信号记录，采集设备包括胸戴式 RespiBan 设备和腕戴式 Empatica E4 设备两种可穿戴平台。RespiBan 设备提供了心电图（ECG）、呼吸（RESP）、皮电活动（EDA）、三轴加速度（ACC）和体温（TEMP）信号；Empatica E4 腕戴设备提供了血容量脉搏（BVP）、皮电活动（EDA）、体温（TEMP）和三轴加速度（ACC）信号。各

2026 Tongji University High School Science Award 仅用于2026年同济大学科学奖论文公示

通道的原始采样率为：ECG 和 RESP 信号 700 Hz，BVP 为 64 Hz，EDA 为 4 Hz，ACC 为 32 Hz。

数据集涵盖了三种情感状态场景：中性（neutral）、压力（stress）和愉悦（amusement）。实验协议采用被试内设计，每位参与者依次经历基线记录、压力诱导任务（Trier Social Stress Test，TSST）和愉悦诱导任务（amusement induction），每段任务的生理信号均有对应的状态标注。三种场景的交替呈现为模型学习不同生理状态到音乐参数的映射提供了多样化的训练样本。

与采用脑电设备的实验室研究相比，WESAD 数据集的核心优势在于其采集设备为腕戴式可穿戴传感器——Empatica E4 的信号类型（BVP、EDA、TEMP）与消费级智能手表和智能戒指高度一致，这确保了基于该数据集训练的模型能够有效迁移到真实番茄钟使用场景中。此外，WESAD 数据集已被广泛应用于压力检测和情感识别研究 [8, 9]，大量已有工作验证了其数据质量，这为跨研究的性能对比提供了基准。

本研究所用的生理信号通道：基于对消费级可穿戴设备（智能手表、智能戒指）当前传感能力的调研，并兼顾生理信号与专注状态的关联强度，本研究从上述数据集中提取以下四类信号通道作为模型输入：

（1）**心率（Heart Rate, HR）**——通过 ECG 或 BVP 信号的 R 波间期计算获得，是反映自主神经活动的基础指标；

（2）**心率变异性（Heart Rate Variability, HRV）**——通过分析连续 RR 间期的时间序列获得，具体采用 RMSSD 和 LF/HF 作为核心特征，是衡量专注状态和压力水平的金标准指标；

（3）**皮电活动（Electrodermal Activity, EDA）**——反映交感神经系统的 arousal 水平，对情绪唤醒度和压力变化敏感；

（4）**血氧饱和度（SpO_2）**——虽非情感状态的最直接指标，但作为补充的健康监测信号，可在实际部署中辅助判断用户的身体状态。

上述四类信号中，HR、HRV 和 SpO_2 已被几乎所有主流消费级可穿戴设备支持，EDA 也已出现在 Amazfit Helio Ring 等最新产品中，确保了模型在实际部署中的输入可得性。从 WESAD 数据集中，HR 和 HRV 可从 ECG（700 Hz）或 BVP（64 Hz）信号中提取，EDA 可直接使用 EDA 通道（4 Hz），SpO_2 可从 PPG 信号的交流与直流分量比值推算。从 SWELL-KW 数据集中，HR 和 HRV 可直接使用其提供的 HR 和 SC（皮肤电导）信号。

为进一步增强模型的泛化能力并验证其在跨设备、跨场景条件下的鲁棒性，本研究引入 SWELL-KW（SWELL Knowledge Work）数据集 [12] 作为辅助数据集进行跨域泛化测试。SWELL-KW 数据集包含 25 名被试在真实知识工作场景（撰写报告、制作演示文稿、处理电子邮件等）中的生理信号记录，实验条件包括中性、邮件干扰和时间压力三种状态。该数据集提供了心率（HR）、心率变异性（HRV）和皮肤电导（SC）信号，并采用 NASA-TLX、RSME 和 SAM 等标准化量表对任务负荷、脑力付出和情绪状态进行了标注。SWELL-KW 的工作场景（知识工作）与本

2026 Outstanding High School Science Award 作品于2026年完稿 科学论文公开

研究的应用场景（番茄钟专注工作）高度一致，其生理信号模态（HR/HRV/SC）与消费级可穿戴设备兼容，使得基于该数据集的跨域测试能够有效评估模型从实验室受控环境到真实工作场景的迁移能力。已有研究已将 WESAD 与 SWELL-KW 联合用于压力检测建模 [13]，验证了这两个数据集联合使用的可行性。

此外，本研究还采集了 10 名志愿者在穿着团队自主研发的 VITATEX 智能监测 T 恤进行真实番茄钟工作时的生理数据，包括 HR、HRV、EDA 和 SpO_2 等通道，作为补充测试集，用于评估模型从实验室环境到真实应用场景的迁移效果。表 2 汇总了本研究使用的所有数据集的基本信息，图 3 以矩阵形式给出各数据集对四类生理模态的覆盖情况。

表 2　本研究使用的数据集汇总

数据集	被试数	本研究所用信号	场景
WESAD	15	ECG/BVP（→HR/HRV）、EDA、PPG（→SpO_2）	中性、压力、愉悦
SWELL-KW	25	HR、HRV、SC	中性、邮件干扰、时间压力
真实场景采集	10	HR、HRV、EDA、SpO_2、体温（VITATEX T 恤）	番茄钟专注工作

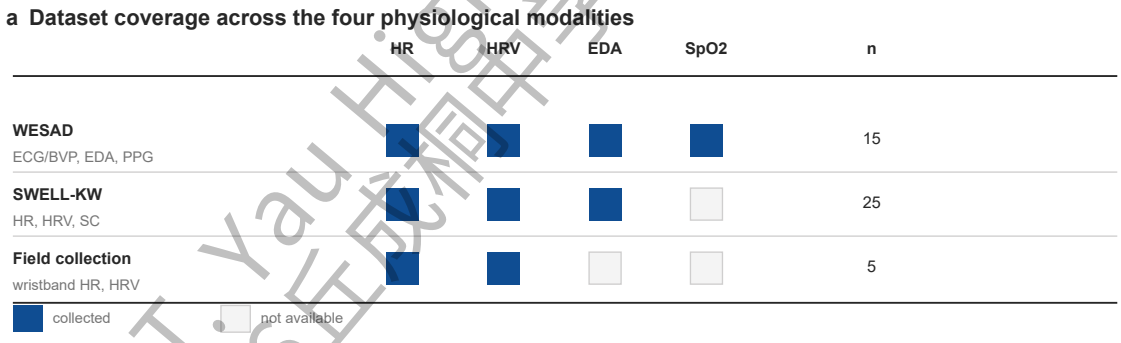


图 3　三个数据集对四类生理模态的覆盖情况。方格表示该数据集是否采集了对应模态，n 为被试数。

3.1.2 数据预处理

原始多模态生理信号在输入模型之前，需经过重采样、异常值处理、滤波去噪和场景标注对齐四个步骤，以消除采集噪声、统一数据格式并适配后续的滑动窗口特征提取。

1. 重采样

WESAD 数据集中各通道的原始采样率不一致：ECG 与 RESP 为 700 Hz，BVP 为 64 Hz，EDA 为 4 Hz。经重采样后，所有信号统一为 4.0 Hz[37]。为统一时间轴并控制计算开销，本研究将所有信号线性重采样至 4 Hz。

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

选择 4 Hz 的依据如下。首先，根据奈奎斯特采样定理（Nyquist-Shannon sampling theorem），采样频率 f_s 至少应为信号最高频率分量 $f_{\max}$ 的两倍（即 $f_s \geq 2f_{\max}$），以避免混叠失真（aliasing）。生理信号（如心率、呼吸）的有效频带通常不超过 2 Hz[19]，因此 4 Hz 的采样率已足以完整保留信号的有用信息。其次，4 Hz（即每 250 ms 一个采样点）与消费级智能手环的典型心率采样频率一致，便于模型从实验室数据向真实部署场景迁移。第三，较低的采样率有助于控制模型的计算开销，满足实时推理的延迟要求。SWELL-KW 数据集原采样率为 256 Hz，同样重采样至 4 Hz[12]。

重采样采用线性插值（linear interpolation）方法实现。对于目标时间点 t_{new}，设其在原始信号中紧邻的两个采样点为 (t_0, x_0) 和 (t_1, x_1)，其中 $t_0 < t_{\text{new}} < t_1$，则插值结果 x_{new} 为：

$$x_{\text{new}} = x_0 + \frac{x_1 - x_0}{t_1 - t_0} \cdot (t_{\text{new}} - t_0) \tag{1}$$

该方法假设信号在两个相邻采样点之间呈线性变化，计算效率高且对生理信号的平稳段引入的误差可接受。

2. 异常值检测与插补

传感器在采集过程中可能因运动伪迹、电极接触不良或设备噪声产生生理上不可能的异常值，需予以检测并处理。

异常值检测采用基于阈值的方法（threshold-based detection）。具体标准为：心率值小于 30 bpm 或大于 200 bpm 的样本、EDA 值为负的样本、血氧饱和度小于 70% 的样本，均判定为异常。这些阈值的设定依据了生理信号的已知有效范围：静息心率通常在 40–100 bpm 之间，极端运动时可达 200 bpm；EDA 为皮肤电导的绝对值，负值在物理上不可能；血氧饱和度低于 70% 属严重低氧血症，在健康受试者的日常记录中极少出现。

检测出的异常值不直接丢弃，而是采用三次样条插值（cubic spline interpolation）进行填补，以保持时间序列的连续性。样条插值使用分段三次多项式（piecewise cubic polynomials）在相邻有效数据点之间构造光滑曲线。对于一组已知数据点 (t_i, x_i)，三次样条 $S(t)$ 在每个区间 $[t_i, t_{i+1}]$ 上满足：

$$S(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i \tag{2}$$

其中系数 a_i, b_i, c_i, d_i 通过约束 $S(t)$、一阶导数 $S'(t)$ 和二阶导数 $S''(t)$ 在所有内部节点处连续来唯一确定。该方法比线性插值更平滑，能更自然地还原信号的趋势变化。单段连续缺失超过 5 秒的片段直接丢弃，以避免插值在长时间无数据区域引入过大误差。

3. 滤波去噪

原始信号中混杂着多种噪声成分，包括由呼吸和肢体运动引起的基线漂移（baseline wander，频率通常低于 0.5 Hz）、工频干扰（power line interference，50/60

2026 年 Lyau High School Science Award 优秀论文公示

Hz）及高频肌电噪声（EMG noise，高于 5 Hz）。需通过滤波提取有效频带成分。

本研究采用 Butterworth 带通滤波器进行滤波。Butterworth 滤波器以通带内频率响应最平坦（maximally flat）为特征，其幅频响应的平方为：

$$|H(j\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2n}} \tag{3}$$

其中 $\omega = 2\pi f$ 为角频率，ω_c 为截止角频率，n 为滤波器阶数。阶数 n 越高，通带与阻带之间的过渡带越陡峭，滤波效果越“干脆”，但计算复杂度和相位失真也随之增加。本研究权衡滤波效果与信号保真度，选取 4 阶 Butterworth 滤波器（$n = 4$）。

对于心率相关信号（ECG、BVP），采用 0.5–5 Hz 的带通滤波：

- **高通截止频率 0.5 Hz**：消除基线漂移和低频呼吸扰动；
- **低通截止频率 5 Hz**：抑制工频干扰（50/60 Hz）和高频肌电噪声，同时完整保留 QRS 复波形态。

对于 EDA 信号，其有效成分变化更为缓慢，采用 0.05–1 Hz 的低通滤波器，去除高频噪声同时保留皮肤电导的慢变分量（tonic component）。

4. 分段与场景标注对齐

完成信号清洗后，需根据各数据集的实验协议将连续记录分割为带场景标签的片段。

WESAD 数据集采用被试内设计，每位参与者依次经历五个阶段：基线记录（baseline，约 5 分钟）、压力诱导任务（TSST，约 15 分钟）、恢复期、愉悦诱导任务（amusement，约 8 分钟）和最终恢复期 [37]。数据采集过程中，各阶段的起止时间及对应的状态标签（中性/压力/愉悦）均随数据集一并提供（存储在数据的 `label` 通道中，采样率与 ECG 相同，为 700 Hz）[37]。

本研究的“分段”操作为：根据 `label` 通道的时间序列，提取出所有标记为“中性”“压力”“愉悦”的连续区间，将每个区间裁剪为独立的信号片段。这一过程本质上是基于数据集已提供的标注信息进行的时间轴切片，不涉及任何分类模型或额外的机器学习方法。分割后的每个片段保留其原始状态标签，该标签将在第 3.1.4 节中被映射为具体的音乐参数标签（BPM、增益等），作为模型训练的预测目标。

以上四个预处理步骤依次执行，最终输出的清洗后信号将作为第 3.1.3 节滑动窗口分割与特征提取的输入。

3.1.3 滑动窗口分割与特征提取

完成信号清洗后，需将连续时间序列转换为适用于图神经网络输入的样本格式。本研究采用滑动窗口（sliding window）技术对预处理后的信号进行分割与特征提取。

1. 滑动窗口分割

滑动窗口分割的核心思想是：用一个固定长度的窗口沿时间轴以固定步长滑

2026 - 华南理工大学 South China University of Technology
仅用于2026届毕业论文公示 Award

动，将长序列切分为若干有重叠的短片段，每个片段作为一个独立样本。在上述参数下，每个窗口含 240 个采样点（4 Hz × 60 s），每条序列包含 8 个窗口（覆盖 4 分钟上下文），全数据集共生成 8850 个窗口、7800 条序列。

设原始信号的总时长为 T_{total}（秒），窗口长度为 L（秒），步长为 S（秒），则生成的窗口数量为：

$$N = \left\lfloor \frac{T_{\text{total}} - L}{S} \right\rfloor + 1 \tag{4}$$

其中 $\lfloor \cdot \rfloor$ 为向下取整函数。每个窗口内的数据点数为 $L \times f_s$，其中 f_s 为重采样后的频率（4 Hz）。

本研究的窗口参数设置如下：

- **窗口长度 $L = 60$ 秒**：选择 60 秒的依据来自 HRV 分析的标准化指南 [19]。该指南明确指出，RMSSD、LF/HF 等 HRV 指标的稳定估计需要至少 60 秒的 RR 间期序列。过短的窗口（如 10–30 秒）会导致频域分辨率不足——频率分辨率 $\Delta f = 1/L$，60 秒窗口的分辨率为 $1/60 \approx 0.0167$ Hz，足以区分 LF（0.04–0.15 Hz）和 HF（0.15–0.40 Hz）频带；而过长的窗口（如 300 秒）则会平滑掉专注状态的快速变化，无法满足实时干预的需求。
- **步长 $S = 30$ 秒**：步长决定了系统的预测更新频率。30 秒步长意味着每 30 秒生成一个新的预测结果，在响应速度（兼顾生理状态变化的及时捕捉）和计算负载之间取得了平衡。此外，窗口间的重叠率为 $1 - S/L = 50\%$，重叠窗口设计可确保状态变化的平滑过渡，避免因窗口边界导致的预测跳变 [19]。

值得注意的是，WESAD 数据集的原始论文中采用了 60 秒窗口配合 0.25 秒步长的设置 [37]，但其目标是离线压力分类，对实时性无要求。本研究采用的 30 秒步长是在保证特征提取充分性的前提下，为满足实时闭环干预需求而作出的调整。在 Python 实现中，可使用 `numpy.lib.stride_tricks.sliding_window_view` 或 `scipy.signal.windows` 库配合循环切片完成该操作。

2. 时域特征提取

对于每个 60 秒窗口，从各生理信号通道中提取以下时域特征：

基础统计特征：对心率、EDA 和 SpO_2 信号分别计算均值（mean）、方差（variance）、标准差（standard deviation, STD）、最大值（maximum）、最小值（minimum）、峰度（kurtosis）和偏度（skewness）。其中峰度衡量信号分布的尾部厚度，偏度衡量分布的非对称性，两者对情绪状态变化具有一定的敏感性 [8]。

HRV 时域特征：从 RR 间期序列中计算以下标准 HRV 指标 [19]：

- **SDNN (Standard Deviation of Normal-to-Normal intervals)**：所有正常 RR 间期的标准差，计算公式为：

$$\text{SDNN} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_i - \overline{RR})^2} \tag{5}$$

2025 ST Yau High School Science Award 仅用于2026立项科学实验公示

其中 N 为窗口内 RR 间期的数量，$\overline{RR}$ 为 RR 间期的均值。SDNN 反映总体 HRV，受交感和副交感神经共同调节 [19]。

- **RMSSD（Root Mean Square of Successive Differences）**：相邻 RR 间期差值的均方根，计算公式为：

$$\text{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2} \tag{6}$$

RMSSD 是评估副交感神经（迷走神经）活动的主要时域指标，与高频功率（HF）高度相关 [19]。已有研究表明，RMSSD 是压力检测中报告频率最高的 HRV 指标 [8]。

- **pNN50**：相邻 RR 间期差值大于 50 ms 的比例，计算公式为：

$$\text{pNN50} = \frac{\text{Count}(|RR_{i+1} - RR_i| > 50 \text{ ms})}{N-1} \times 100\% \tag{7}$$

pNN50 同样反映副交感神经活动，与 RMSSD 具有较高的相关性 [19]。

3. 频域特征提取

频域特征通过功率谱密度（Power Spectral Density, PSD）分析提取，反映 HRV 信号中不同频率成分的能量分布 [19]。

本研究采用**快速傅里叶变换（Fast Fourier Transform, FFT）**进行功率谱密度估计。对于长度为 $M = L \times f_s = 60 \times 4 = 240$ 的离散信号序列 $x[0], x[1], \ldots, x[M-1]$，其离散傅里叶变换（DFT）定义为：

$$X[k] = \sum_{n=0}^{M-1} x[n] \cdot e^{-j2\pi kn/M}, \quad k = 0, 1, \ldots, M-1 \tag{8}$$

其中 $X[k]$ 为频域第 k 个频率仓（frequency bin）的复数值，$j = \sqrt{-1}$ 为虚数单位。功率谱密度估计值为：

$$P[k] = \frac{2 \cdot |X[k]|^2}{f_s \cdot M} \tag{9}$$

其中 $|X[k]|^2$ 为复数幅值的平方，乘以 2 补偿单边谱的能量，除以 $f_s \cdot M$ 进行归一化。频率轴与索引 k 的关系为：

$$f_k = \frac{k \cdot f_s}{M} = \frac{k \cdot 4}{240} = \frac{k}{60} \text{ Hz} \tag{10}$$

频率分辨率 $\Delta f = f_s/M = 4/240 \approx 0.0167$ Hz。

基于上述 FFT 结果，提取以下频域特征 [19]：

- **LF（Low Frequency power）**：0.04–0.15 Hz 频带内的功率，对应索引范围

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

$k = \lceil 0.04 \times 60 \rceil = 3$ 至 $k = \lfloor 0.15 \times 60 \rfloor = 9$：

$$P_{\mathrm{LF}} = \sum_{k=3}^{9} P[k] \tag{11}$$

LF 功率反映交感和副交感神经的共同调节，与压力感受器反射活动相关 [19]。

- **HF（High Frequency power）**：0.15–0.40 Hz 频带内的功率，对应索引范围 $k = \lceil 0.15 \times 60 \rceil = 9$ 至 $k = \lfloor 0.40 \times 60 \rfloor = 24$：

$$P_{\mathrm{HF}} = \sum_{k=9}^{24} P[k] \tag{12}$$

HF 功率反映副交感神经（迷走神经）活动，主要由呼吸性窦性心律不齐（respiratory sinus arrhythmia, RSA）引起 [19]。

- **LF/HF 比值**：低频与高频功率之比，计算公式为：

$$\mathrm{LF/HF} = \frac{P_{\mathrm{LF}}}{P_{\mathrm{HF}}} \tag{13}$$

LF/HF 比值被广泛用作交感-副交感神经平衡的指标——比值升高提示交感神经占优（与压力、焦虑相关），比值降低提示副交感神经占优（与放松、专注相关）[19]。

- **总功率（Total Power, TP）**：0–0.40 Hz 频带内的总功率：

$$P_{\mathrm{total}} = \sum_{k=0}^{\lfloor 0.40 \times 60 \rfloor} P[k] \tag{14}$$

4. 特征标准化

完成所有窗口的特征提取后，每个窗口形成一个维度为 d 的特征向量 $\mathbf{x}_i \in \mathbb{R}^d$。所有 N 个窗口构成特征矩阵 $\mathbf{X} \in \mathbb{R}^{N \times d}$。

为消除不同特征之间量纲差异对模型训练的影响，按训练集的均值和标准差进行 Z-score 标准化：

$$\mathbf{x}_i^{(\mathrm{norm})} = \frac{\mathbf{x}_i - \boldsymbol{\mu}_{\mathrm{train}}}{\boldsymbol{\sigma}_{\mathrm{train}}} \tag{15}$$

其中 $\boldsymbol{\mu}_{\mathrm{train}} \in \mathbb{R}^d$ 和 $\boldsymbol{\sigma}_{\mathrm{train}} \in \mathbb{R}^d$ 分别为训练集各特征的均值和标准差向量（逐元素运算）。标准化参数保存为 JSON 文件，供后续部署阶段对实时数据应用相同的变换，确保训练和推理时的数据分布一致。

标准化后的特征向量将作为第 4 节图神经网络中各时间窗口节点的初始特征输入。

2026 S. T. Yau High School Science Award
仅用于2026成功中学科学奖论文公示

3.1.4 标签定义

本研究的模型采用有监督学习范式，需要为每个窗口定义对应的“标签”——即模型应预测的目标音乐参数值。

标签来源：本研究使用的 WESAD 和 SWELL-KW 数据集均提供了每个时间片段对应的生理状态标注（如“中性”“压力”“愉悦”等）。基于这些状态标注，结合音乐心理学和生理声学领域的已有研究成果，为每种状态赋予一组“最优音乐参数”作为训练标签。

标签参数的选择：与混音师实时调控的参数相对应，本研究选取以下音乐参数作为模型的预测目标：

- **BPM（Tempo）**：60–120 范围内的连续值（回归任务）。BPM 是影响心血管反应最显著的音乐因素之一——快节奏音乐（> 100–120 BPM）与心率升高和交感神经激活相关，而慢节奏音乐（< 60–80 BPM）则促进副交感神经主导和心率降低 [14]。
- **整体增益（Gain）**：−6 至 +6 dB 范围内的连续值（回归任务），控制整体响度。响度是影响生理唤醒度的基本声学特征，较高的响度通常与较高的 arousal 水平相关 [16]。
- **低频增益（Low-end Gain）**：−3 至 +6 dB 范围内的连续值（回归任务），控制 60–250 Hz 频段的能量。低频增强与“温暖”“厚重”“安定”的听感相关，在放松和压力缓解场景中具有重要作用 [15]。
- **混响深度（Reverb Depth）**：0–1 范围内的连续值（回归任务），控制空间效果的干湿比。混响深度影响音乐的沉浸感和空间感，已被用于情感音乐生成和生物反馈音乐系统中 [18]。
- **调式（Mode）**：{Major, Minor} 二分类（分类任务）。大调与小调是和声层面影响情绪效价（valence）的核心因素——大调通常与正向情绪相关，小调与负向情绪相关 [14]。

状态到参数的映射规则：基于上述文献依据，本研究建立以下从生理状态到音乐参数的映射关系：

表 3 生理状态到音乐参数的映射规则

生理状态	BPM	整体增益	低频增益	混响深度	调式
中性（neutral）	75–85	0 dB	0 dB	0.3	Major
压力（stress）	60–75	−2 dB	+3 dB	0.6	Minor
愉悦（amusement）	95–110	+2 dB	−1 dB	0.2	Major

映射规则依据：

1. **BPM 的映射依据**：Frontiers 综述明确指出，慢节奏音乐（60–80 BPM）促进副交感神经激活和心率降低，适用于压力缓解场景；而快节奏音乐（> 100 BPM）

2026 S-T. Yau High School Science Award 仅用于2026正式申报科学奖论文公示

则激活交感神经，提升生理唤醒度 [14]。Stith（2015）的研究进一步验证了 tempo（60 BPM vs 90 BPM）对 HRV 和感知放松的影响 [15]。

2. **增益与低频增益的映射依据**：Ethridge 等人（2024）通过对 DEAP 数据集的二次分析发现，低层音频特征（包括响度和频谱特征）与生理 arousal 之间存在显著相关性 [16]。Su 和 Zhou（2024）通过随机森林回归建模发现，节奏对情绪反应的贡献显著大于音色和音高 [17]。基于此，压力场景下降低整体增益、增强低频成分以营造温暖安定感；愉悦场景下提升整体增益、略微削减低频以突出清晰明亮的听感。
3. **混响深度的映射依据**：生物反馈音乐生成专利文献指出，混响（reverb）是可被实时调控以响应生理信号的核心音乐参数之一，与音乐的“深度”“沉浸感”和“空间感”直接相关 [18]。压力场景下增加混响以营造空间感和放松感；愉悦和中性场景下保持较低混响以维持清晰度和节奏感。
4. **调式的映射依据**：Frontiers 综述指出，协和、可预测的和声结构（如大调和弦）倾向于诱导平静效果，而不协和或不可预测的音乐可能增加压力相关反应 [14]。因此压力场景采用小调，中性及愉悦场景采用大调。

综上，每个训练样本的标签向量为：

$$\mathbf{y} = [y_{\text{BPM}}, y_{\text{gain}}, y_{\text{low_gain}}, y_{\text{reverb}}, y_{\text{mode}}] \in \mathbb{R}^{1+1+1+1+2} \tag{16}$$

其中 y_{mode} 为二维 one-hot 向量（Major/Minor），其余均为连续标量值。

3.1.5 数据集划分

将完整数据集按 70% / 15% / 15% 的比例划分为训练集、验证集和测试集。训练集用于模型参数更新，验证集用于超参数选择和早停判断，测试集用于最终模型性能的独立评估。

划分时采用按被试分层的策略（即同一被试的所有样本只能出现在同一个子集中），而非随机划分窗口样本，以避免同一被试的数据同时出现在训练集和测试集中造成的数据泄漏（data leakage），从而更真实地评估模型在未见用户上的泛化能力。WESAD 的 15 名被试中，随机选取 10 人作为训练集、2 人作为验证集、3 人作为测试集。按被试划分后，训练集共 5200 条序列，验证集共 1040 条序列，测试集共 1560 条序列，各子集之间无被试重叠。SWELL-KW 的 25 名被试采用同样策略，取 17 人训练、4 人验证、4 人测试。

3.2 本节小结

本节详细描述了模型训练阶段的数据准备与特征工程流程。我们使用 WESAD 与 SWELL-KW 数据集，通过重采样、滤波、异常值剔除等预处理步骤，结合滑动窗口分割和时域/频域特征提取，构建了图神经网络的节点特征集，并基于生理状态标注与文献依据定义了五维音乐参数的回归标签。上述数据准备流程为第 4 节动态异构图神经网络模型的训练奠定了基础。系统部署阶段的实时数据处理流程

2026 Sh. Yau High School Science Award 仅用于2026年成都市学科学奖论文公示

将在第 5 节中详细阐述。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

4　动态异构图神经网络模型

本节提出本研究的核心模型——动态异构图神经网络（Dynamic Heterogeneous Graph Neural Network, DHGNN）。模型整体架构如图 4 所示，包含四个部分：问题形式化（4.1 节）、动态异构图构建（4.2 节）、异构图注意力网络（4.3 节）、时序动态建模（4.4 节），以及多任务输出与损失函数（4.5 节）和模型训练策略（4.6 节）。

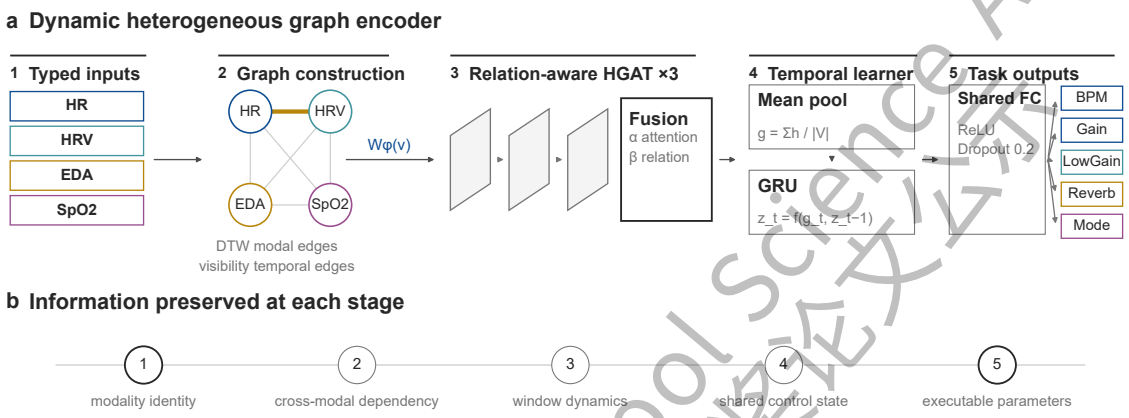


图 4　DHGNN 核心架构。(a) 动态异构图编码器的五个计算阶段。(b) 各阶段保留的信息语义。

4.1　问题形式化

本研究将动态音乐参数调控问题建模为如下形式。

输入信号：设经过预处理和重采样后的多通道生理时间序列为

$$\mathbf{S} = [\mathbf{s}_{\mathrm{HR}}, \mathbf{s}_{\mathrm{HRV}}, \mathbf{s}_{\mathrm{EDA}}, \mathbf{s}_{\mathrm{SpO_2}}] \in \mathbb{R}^{T \times C} \tag{17}$$

其中 T 为时间序列的长度（采样点数），$C = 4$ 为生理信号通道数。各通道分别对应心率（HR）、心率变异性（HRV）、皮电活动（EDA）和血氧饱和度（$\mathrm{SpO_2}$）。

时间窗口分割：采用滑动窗口将连续时间序列分割为 N 个窗口，窗口长度为 $L = 60$ 秒，步长为 $S = 30$ 秒。窗口数量为：

$$N = \left\lfloor \frac{T/f_s - L}{S} \right\rfloor + 1 \tag{18}$$

其中 $f_s = 4$ Hz 为重采样频率，每个窗口包含 $M = L \times f_s = 240$ 个采样点。第 i 个窗口记为 $\mathbf{W}_i \in \mathbb{R}^{M \times C}$，$i = 1, 2, \ldots, N$。

节点特征：对第 i 个窗口的第 c 个通道（$c \in \{1, 2, 3, 4\}$），提取 d 维特征向量 $\mathbf{x}_i^{(c)} \in \mathbb{R}^d$。特征包括时域统计特征（均值、方差、标准差、峰度、偏度）、HRV 时域特征（SDNN、RMSSD、pNN50）和 HRV 频域特征（LF 功率、HF 功率、LF/HF

2026 SCI Yu Hua High School Science Award
仅用于2026年成都市学科学奖论文公示

比值、总功率)。四类节点的特征维度分别为: HR 节点 6 维、HRV 节点 7 维、EDA 节点 6 维、SpO_2 节点 5 维。特征提取的完整定义见第 3.1.3 节。

图结构定义: 定义图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$。节点集 $\mathcal{V} = \{v_{(i,c)} \mid i = 1, \ldots, N;\ c = 1, \ldots, C\}$ 包含 $V = N \times C$ 个节点, 其中 $v_{(i,c)}$ 对应第 i 个窗口第 c 个通道的特征向量 $\mathbf{x}_i^{(c)}$。节点类型根据生理通道区分:

$$\text{type}(v_{(i,c)}) = c \in \{\text{HR}, \text{HRV}, \text{EDA}, \text{SpO}_2\} \tag{19}$$

边集 $\mathcal{E} = \mathcal{E}_{\text{temp}} \cup \mathcal{E}_{\text{modal}}$ 包含两类边:

- **时间边**: 同一通道相邻窗口之间的连接, $\mathcal{E}_{\text{temp}} = \{(v_{(i,c)}, v_{(i+1,c)}) \mid i = 1, \ldots, N-1;\ c = 1, \ldots, C\}$, 用于捕捉各生理指标随时间的演化;
- **模态边**: 同一窗口内不同通道之间的连接, $\mathcal{E}_{\text{modal}} = \{(v_{(i,c)}, v_{(i,c')}) \mid i = 1, \ldots, N;\ c \neq c'\}$, 用于捕捉不同生理信号之间的跨模态关联。

令 $\mathcal{G}_t$ 表示第 t 个时间步 ($t = 1, \ldots, N$) 对应的子图, 包含节点 $\{v_{(t,c)}\}_{c=1}^{C}$ 及其模态边。图拓扑随 t 变化, 构成一个动态异构图序列 $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_N\}$。

输出空间: 每个时间窗口 i 对应一个音乐参数向量 $\mathbf{y}_i$。根据第 3.1.4 节的标签定义,

$$\mathbf{y}_i = [y_{\text{BPM}}, y_{\text{gain}}, y_{\text{low_gain}}, y_{\text{reverb}}, y_{\text{mode}}] \in \mathbb{R}^{1+1+1+1+2} \tag{20}$$

其中:

- $y_{\text{BPM}} \in [60, 120]$: 音乐速度 (连续值, 回归任务);
- $y_{\text{gain}} \in [-6, +6]$: 整体增益, 单位 dB (连续值, 回归任务);
- $y_{\text{low_gain}} \in [-3, +6]$: 低频段增益, 单位 dB (连续值, 回归任务);
- $y_{\text{reverb}} \in [0, 1]$: 混响深度 (连续值, 回归任务);
- $y_{\text{mode}} \in \{(1,0), (0,1)\}$: 调式, $(1,0)$ 对应大调(Major), $(0,1)$ 对应小调(Minor)(分类任务)。

综上, 动态音乐参数调控问题定义为: 给定动态异构图序列 $\{\mathcal{G}_1, \ldots, \mathcal{G}_N\}$, 求映射 f 使得

$$f(\mathcal{G}_1, \ldots, \mathcal{G}_t) = \mathbf{y}_t, \quad t = 1, 2, \ldots, N \tag{21}$$

即利用截至当前时间步 t 的图序列信息, 预测当前时间步的最优音乐参数向量。该映射需满足实时性约束——每 30 秒 (即每个新窗口到来时) 输出一次预测结果, 端到端推理延迟控制在 100 ms 以内。

4.2 动态异构图构建

本节介绍将多通道生理时间序列转换为动态异构图结构的方法, 包含图的形式化定义、节点定义、边定义与权重计算、以及动态更新机制四个部分。

2026 年 Yale High School Science Award 仅用于 2026 年成都中学科学奖论文公示

4.2.1 图的形式化定义

设经滑动窗口分割后共有 N 个时间窗口，每个窗口包含 $C = 4$ 个生理通道，分别为心率（HR）、心率变异性（HRV）、皮电活动（EDA）和血氧饱和度（SpO_2）。定义第 t 个时间步（$t = 1, 2, \ldots, N$）的异构图为：

$$\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \phi, \psi) \tag{22}$$

其中：

- $\mathcal{V}_t$ 为节点集合；
- $\mathcal{E}_t \subseteq \mathcal{V}_t \times \mathcal{V}_t$ 为边集合；
- $\phi : \mathcal{V}_t \to \mathcal{A}$ 为节点类型映射函数，$\mathcal{A}$ 为节点类型集合；
- $\psi : \mathcal{E}_t \to \mathcal{R}$ 为边类型映射函数，$\mathcal{R}$ 为边类型集合。

与传统同构图（homogeneous graph）不同，异构图包含多种类型的节点和边，能够更准确地建模多模态生理信号之间的复杂交互关系 [24]。

动态图根据时间建模方式的不同，可分为离散时间动态图（Discrete-Time Dynamic Graph, DTDG）和连续时间动态图（Continuous-Time Dynamic Graph, CTDG）两类 [5]。DTDG 将时间轴划分为离散的时间步，每个时间步对应一个静态图快照，整个图序列 $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_N\}$ 构成动态图；CTDG 则对每条边附加时间戳，图在连续时间轴上以事件序列的形式演化。本研究采用的滑动窗口分割方案将连续的生理信号切分为离散的时间窗口（窗口长度 60 秒，步长 30 秒），每个窗口独立构建一个异构图快照，因此属于 DTDG 范式。这一范式的优势在于可以直接利用成熟的静态图神经网络处理每个快照中的空间结构信息，再通过序列模型（如 GRU）捕捉快照之间的时间演化依赖 [5]。

4.2.2 节点定义

节点是图的基本构成单元。本研究采用“每个时间窗口 × 每个生理通道”作为一个独立节点的定义方式。

具体地，对于第 t 个时间窗口和第 c 个生理通道（$c \in \{1, 2, 3, 4\}$），定义节点 $v_{t,c} \in \mathcal{V}_t$。因此，每个时间步的节点总数为 $|\mathcal{V}_t| = C = 4$，整个图序列的节点总数为 $N \times C$。

每个节点的初始特征向量为第 3.1.3 节提取的 d 维特征向量 $\mathbf{x}_t^{(c)} \in \mathbb{R}^d$，包含时域统计特征、HRV 时域特征和 HRV 频域特征。

节点类型由 ϕ 函数赋予：

$$\phi(v_{t,c}) = c \in \mathcal{A} = \{\mathrm{HR}, \mathrm{HRV}, \mathrm{EDA}, \mathrm{SpO_2}\} \tag{23}$$

不同类型的节点具有不同的生理意义和统计特性。心率节点反映心脏搏动的基本节律；HRV 节点反映自主神经系统对心脏的调制活动，是衡量专注状态和压力水平的敏感指标；EDA 节点反映交感神经系统的 arousal 水平，对情绪唤醒度

2026 S.T.Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

变化敏感；血氧饱和度节点作为补充的健康监测指标。

由于不同类型节点的原始特征向量具有不同的维度和语义分布，无法直接在同一空间中计算其相互关系。为克服这一“异构鸿沟”（heterogeneity gap）问题，本研究采用类型特异性线性变换（Type-specific Linear Transformation）对各类节点的特征进行投影对齐 [24]。该方法是异构图神经网络中的标准操作，其核心思想是为每种节点类型设计一个独立的线性变换矩阵，将原始特征映射至统一的公共语义空间 [24, 25]。

具体地，对于类型为 $\phi(v) = A$（$A \in \mathcal{A}$）的节点 v，其原始特征 $\mathbf{x}_v \in \mathbb{R}^{d_A}$ 通过该类型专属的线性变换矩阵 $\mathbf{W}_A \in \mathbb{R}^{d' \times d_A}$ 和偏置项 $\mathbf{b}_A \in \mathbb{R}^{d'}$，映射至统一的 d' 维隐空间：

$$\mathbf{h}'_v = \mathbf{W}_{\phi(v)} \cdot \mathbf{x}_v + \mathbf{b}_{\phi(v)} \tag{24}$$

其中 $\mathbf{h}'_v \in \mathbb{R}^{d'}$ 为投影后的节点特征，所有类型节点的投影特征维度统一为 d'。变换矩阵 $\mathbf{W}_{\phi(v)}$ 和偏置 $\mathbf{b}_{\phi(v)}$ 均为可训练参数，在模型训练过程中通过反向传播进行更新。以本研究的四类节点为例，分别定义：

$$\mathbf{W}_{\text{HR}} \in \mathbb{R}^{d' \times 6}, \quad \mathbf{W}_{\text{HRV}} \in \mathbb{R}^{d' \times 7}, \quad \mathbf{W}_{\text{EDA}} \in \mathbb{R}^{d' \times 6}, \quad \mathbf{W}_{\text{SpO}_2} \in \mathbb{R}^{d' \times 5} \tag{25}$$

各类型节点通过各自专属的变换矩阵独立完成投影。该设计的优势在于：

（1）每种类型的特征在被映射到公共空间之前，先经过独立的线性变换，保留了各自的特异性；

（2）统一维度后的特征可在后续的异构图注意力层中进行统一的相似度计算和信息聚合；

（3）不同类型之间的映射参数相互独立，互不干扰，使得模型能够针对不同生理信号的统计特性进行差异化建模 [24]。

经过上述类型特异性变换后，所有节点的特征被投影至同一 d' 维语义空间，异构鸿沟得到有效消除。投影后的特征 $\mathbf{h}'_v$ 将作为后续异构图注意力层的输入。

4.2.3 边定义与权重计算

边定义了节点之间的关系。本研究构建两类边：时间边和模态边。

时间边（Temporal Edges）：同一生理通道在相邻时间窗口之间的连接，即 $(v_{t,c}, v_{t+1,c}) \in \mathcal{E}_t^{\text{temp}}$ $(t = 1, \ldots, N-1)$，用于捕捉各生理指标随时间的演化趋势。

时间边的存在性判断采用可视图理论（Visibility Graph）[20]。对于时间序列中的两个数据点 (t_i, x_i) 和 (t_j, x_j) $(t_i < t_j)$，若对于任意中间点 (t_k, x_k) $(t_i < t_k < t_j)$ 均满足：

$$x_k < x_i + (x_j - x_i) \cdot \frac{t_k - t_i}{t_j - t_i} \tag{26}$$

2026 第十一届 Hi-School Science Award 仅用于2026江苏高中学生学术论文公示

则两点“可视”，在图中建立连接。可视图理论已被证明是将时间序列转换为复杂网络的有效方法，其构成的网络在拓扑结构中继承了原始时间序列的动力学特性[20, 21]。

模态边（Modal Edges）：同一时间窗口内不同生理通道之间的连接，即 $(v_{t,c}, v_{t,c'}) \in \mathcal{E}_t^{\text{modal}}$ $(c \neq c')$，用于捕捉不同生理信号之间的跨模态关联。

模态边的权重基于特征向量之间的相似度计算。本研究采用动态时间规整（Dynamic Time Warping, DTW）距离度量两个通道特征向量的相似性 [22]。对于两个长度分别为 m 和 n 的时间序列 $\mathbf{a} = (a_1, \ldots, a_m)$ 和 $\mathbf{b} = (b_1, \ldots, b_n)$，DTW 通过构造 $m \times n$ 的累积距离矩阵 $\mathbf{D}$，寻找一条最优对齐路径 $\pi = \{\pi_1, \ldots, \pi_K\}$，使得累积距离最小：

$$\text{DTW}(\mathbf{a}, \mathbf{b}) = \min_{\pi} \sum_{k=1}^{K} d(\pi_k) \tag{27}$$

其中 $d(\cdot)$ 为欧氏距离。DTW 通过非线性对齐衡量两个时间序列的相似度，对生理信号常见的相位偏移（如心率节律的个体差异、呼吸频率的变化等）具有鲁棒性，因此在生理信号分析中被广泛使用 [23]。

边权重统一计算公式：

节点 $v_{t,c}$ 与 $v_{t',c'}$ 之间的边权重统一为分段函数：

$$W_{(t,c),(t',c')} = \begin{cases} \text{Visibility}(t, t') \cdot \exp\left(-\dfrac{\text{DTW}(\mathbf{x}_t^{(c)}, \mathbf{x}_{t'}^{(c)})^2}{2\sigma^2}\right), & c = c',\ t' = t+1 \\ \exp\left(-\dfrac{\text{DTW}(\mathbf{x}_t^{(c)}, \mathbf{x}_t^{(c')})^2}{2\sigma^2}\right), & c \neq c',\ t' = t \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

其中 σ 为带宽参数，控制相似度对权重的衰减速度。该设计具有以下特点：

（1）时间边同时考虑了时间上的“可视性”和特征上的相似性，权重为两者乘积；

（2）模态边仅基于特征相似度计算权重；

（3）非相邻时间步的不同窗口节点之间无边连接（权重为 0），以保持图结构的稀疏性和计算效率。

4.2.4 动态更新机制

动态性是本研究图构建的核心特征之一。随着滑动窗口的推进，图拓扑每 30 秒更新一次。

具体地，当新的时间窗口 $t+1$ 到来时，图结构按以下步骤进行增量更新：

1. 创建新节点 $v_{t+1,c}$ $(c = 1, \ldots, C)$，其初始特征 $\mathbf{x}_{t+1}^{(c)}$ 由当前窗口的生理信号经第

2026 S. T. Yau High School Science Award
仅用于2026年成稿日学科学奖论文公示

3.1.3 节的特征提取流程计算得出；

2. 建立新节点与上一时间步对应节点之间的时间边 $(v_{t,c}, v_{t+1,c})$，$c = 1, \ldots, C$；
3. 建立新节点之间的模态边 $(v_{t+1,c}, v_{t+1,c'})$，$c \neq c'$；
4. 根据窗口步长策略，当窗口滑动超过历史保留范围时，移除最旧的节点 $v_{1,c}$ 及其关联边。

这种增量式动态更新机制使图拓扑能够实时反映生理状态的变化，同时相比全量重建大幅降低了计算开销——每次更新只需处理新窗口引入的 C 个节点及其关联边，计算复杂度为 $O(C^2)$，与当前图规模无关。动态图结构已被验证能够有效捕捉生理信号在时间维度上演化的复杂模式 [38]。

综上，通过上述四个步骤，原始的多通道生理时间序列被转换为一个动态异构图序列 $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_N\}$，其中每个时间步的图包含四种类型的节点和两类边，图拓扑随生理状态的演变而动态更新。该图序列将作为后续异构图注意力网络的输入。

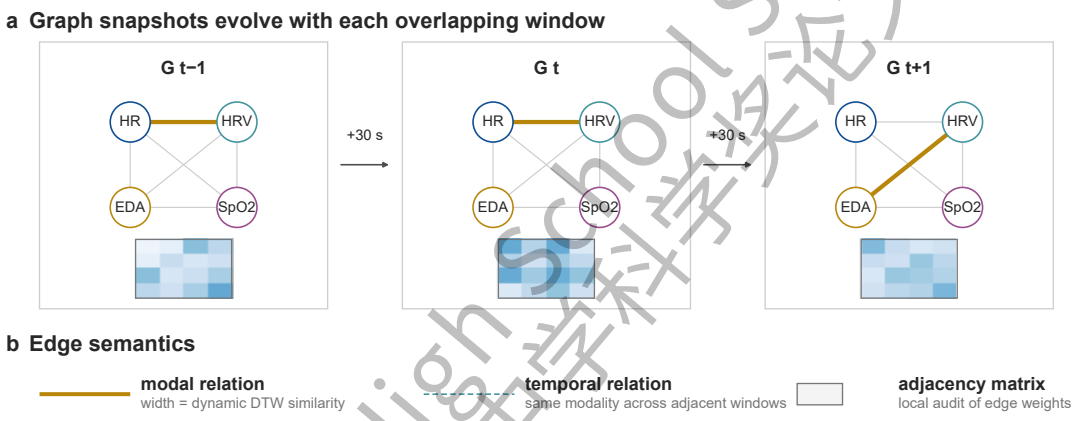


图 5　滑动窗口推进过程中的动态异构图演化。(a) 连续图快照与邻接矩阵。(b) 模态边、时序边与矩阵编码的图例说明。

4.3　异构图注意力网络

经过第 4.2 节的动态异构图构建，原始生理时间序列已被转换为动态异构图序列 $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_N\}$，其中每个时间步的图包含四种类型的节点（HR、HRV、EDA、SpO_2）和两类边（时间边、模态边）。本节介绍在该图结构上运行的异构图注意力网络（Heterogeneous Graph Attention Network），该网络通过层次化的注意力机制实现节点特征的聚合与更新。

异构图注意力网络的核心思想是：在信息聚合过程中，不同类型的节点和边应当被差异化地对待 [24]。具体地，网络包含两个层次的注意力机制：

（1）**节点级注意力（Node-level Attention）**——在同一关系类型下，学习不同邻居节点对中心节点的重要性权重；

（2）**语义级注意力（Semantic-level Attention）**——在不同关系类型（时间边、模态边）之间，学习不同语义信息的重要性权重。通过这两层注意力的级联，

2026 Siyau High School Science Award
仅用于2026年成同中学科学奖论文公示

模型能够充分捕捉多模态生理信号之间的复杂交互模式 [24]。

整体架构如图 6 所示。

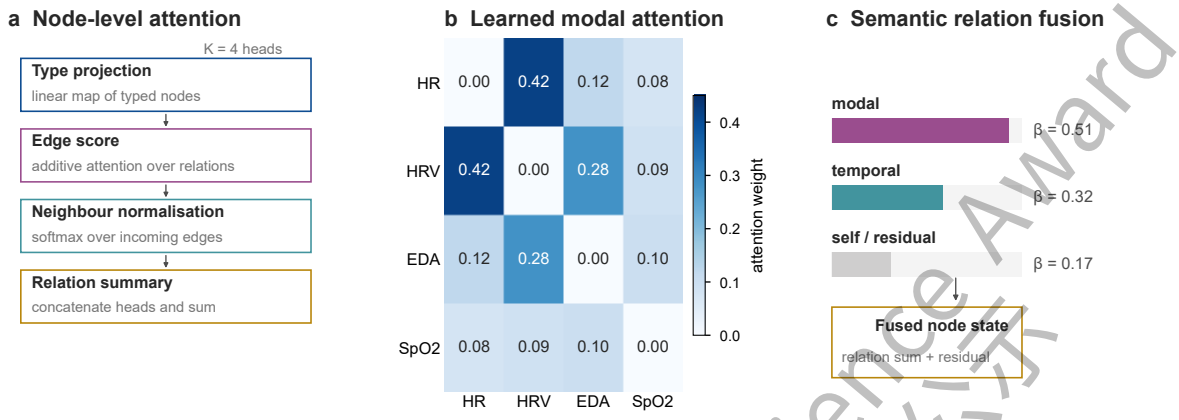


图 6　异构图注意力机制。(a) 节点级注意力计算：类型投影 $h'_v = W_{\phi(v)} x_v$、边得分 $e_{st}^{r,k} = a^T[W_r h_s \| W_r h_t]$、邻居归一化 $\alpha_{st}^{r,k} = \mathrm{softmax}_s(e_{st}^{r,k})$、关系汇总 $h_v^r = \|_k \sum_s \alpha_{st}^{r,k} W_r^k h_s$。(b) 跨模态注意力矩阵。(c) 语义关系融合，融合节点状态为 $h_v^{(l)} = \sum_r \beta_r h_v^{(r,l)} + h_v^{(l-1)}$。

4.3.1　节点级注意力

节点级注意力的目标是：对于中心节点 v_t，对其在关系 r 下的所有邻居节点 $v_s \in \mathcal{N}_r(t)$ 计算注意力权重，并据此聚合邻居信息，生成中心节点在该关系下的语义特定表示 [24]。

设第 $(l-1)$ 层（初始层 $l=1$ 时，$\mathbf{h}_i^{(0)} = \mathbf{h}'_i$，即经第 4.2.2 节类型特异性变换后的节点特征）的节点表示为 $\mathbf{h}_s^{(l-1)} \in \mathbb{R}^{d'}$。对于关系 r（$r \in \{\mathrm{temp}, \mathrm{modal}\}$），首先通过关系特定的变换矩阵 $\mathbf{W}_r^{(l)} \in \mathbb{R}^{d' \times d'}$ 将源节点和目标节点映射到关系 r 对应的语义空间：

$$\tilde{\mathbf{h}}_s^{(l-1)} = \mathbf{W}_r^{(l)} \mathbf{h}_s^{(l-1)}, \quad \tilde{\mathbf{h}}_t^{(l-1)} = \mathbf{W}_r^{(l)} \mathbf{h}_t^{(l-1)} \tag{29}$$

该变换使得模型能够针对不同关系类型学习独立的特征投影，从而捕捉关系特定的语义模式 [24]。

随后计算目标节点 v_t 对源节点 v_s 的原始注意力值 $e_{(s,t)}^{(r,l)}$。该值通过将变换后的源节点特征和目标节点特征拼接，再经过一个可学习的注意力向量 $\mathbf{a}_r^{(l)} \in \mathbb{R}^{2d'}$ 映射到标量获得：

$$e_{(s,t)}^{(r,l)} = \mathrm{LeakyReLU}\left(\mathbf{a}_r^{(l)T}\left[\tilde{\mathbf{h}}_s^{(l-1)} \,\|\, \tilde{\mathbf{h}}_t^{(l-1)}\right]\right) \tag{30}$$

其中 $\|$ 表示向量拼接操作。LeakyReLU 激活函数（负斜率设为 0.2）在此处的使用是为了在保留负值信息的同时引入非线性，避免梯度消失问题。该注意力计算方式继承了图注意力网络（Graph Attention Network, GAT）的设计思路，通过自注

2026 S.T. Yau High School Science Award
仅用于2026丘成桐科学奖论文公示

意力机制自动学习邻居的重要性 [26]。

为使注意力系数在不同邻居之间具有可比性，采用 softmax 函数对原始注意力值进行归一化：

$$\alpha_{(s,t)}^{(r,l)} = \frac{\exp\left(e_{(s,t)}^{(r,l)}\right)}{\sum_{s' \in \mathcal{N}_r(t)} \exp\left(e_{(s',t)}^{(r,l)}\right)} \tag{31}$$

其中 $\mathcal{N}_r(t)$ 表示与目标节点 v_t 在关系 r 下相连的源节点集合。归一化后的注意力系数满足 $\sum_{s \in \mathcal{N}_r(t)} \alpha_{(s,t)}^{(r,l)} = 1$，具有概率解释。

得到注意力系数后，目标节点 v_t 在关系 r 下的邻居聚合结果 $\mathbf{z}_t^{(r,l)}$ 计算为所有邻居变换特征的加权和：

$$\mathbf{z}_t^{(r,l)} = \sum_{s \in \mathcal{N}_r(t)} \alpha_{(s,t)}^{(r,l)} \cdot \tilde{\mathbf{h}}_s^{(l-1)} \tag{32}$$

该聚合过程本质上是对邻居节点信息按重要性进行加权平均，使得与当前节点更相关的邻居在特征更新中发挥更大的作用。

4.3.2 多头注意力机制

为增强模型的表达能力和训练稳定性，本研究采用多头注意力机制（Multi-head Attention）[26]。多头注意力的核心思想是：并行运行多个独立的注意力头，每个头学习不同的注意力分布，从而捕获节点之间多样化的交互模式。

具体地，并行运行 K 个独立的注意力头（本研究设置 $K = 4$），每个头拥有独立的变换矩阵 $\mathbf{W}_{r,k}^{(l)}$ 和注意力向量 $\mathbf{a}_{r,k}^{(l)}$（$k = 1, \ldots, K$），分别计算各自的聚合结果 $\mathbf{z}_{t,k}^{(r,l)}$。

对于中间隐藏层（$l < L$，L 为总层数），将 K 个头的输出拼接作为该层的最终输出：

$$\mathbf{z}_t^{(r,l)} = \Big\|_{k=1}^{K} \mathbf{z}_{t,k}^{(r,l)} \tag{33}$$

拼接操作保留了每个头学习到的独立信息，使得模型能够同时关注不同层面的特征交互。

对于最后一层（$l = L$），采用 K 个头输出的平均值作为最终输出，以降低特征维度并稳定训练：

$$\mathbf{z}_t^{(r,L)} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{z}_{t,k}^{(r,L)} \tag{34}$$

随后通过非线性激活函数 $\sigma(\cdot)$ 得到本层的节点输出。对于隐藏层，采用 ReLU

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公文公示

激活函数：

$$\mathbf{h}_t^{(r,l)} = \mathrm{ReLU}\left(\mathbf{z}_t^{(r,l)}\right), \quad l < L \tag{35}$$

ReLU 激活函数 $\mathrm{ReLU}(x) = \max(0, x)$ 具有计算简单、缓解梯度消失的优点，是图神经网络中广泛使用的激活函数 [26]。

对于输出层，直接使用聚合结果作为输出（即恒等映射）：

$$\mathbf{h}_t^{(r,L)} = \mathbf{z}_t^{(r,L)} \tag{36}$$

4.3.3 语义级注意力

经过节点级注意力后，每个节点在每种关系 r（时间边、模态边）下都获得了对应的语义表示 $\mathbf{h}_t^{(r,l)}$。不同关系类型捕获了生理信号在不同维度上的关联信息——时间边反映信号的时间演化模式，模态边反映不同信号间的跨模态关联。然而，不同关系类型对最终任务的重要性各不相同，需要模型自动学习其权重。为此，本研究引入语义级注意力机制 [24]。

对于节点 v_t，设其在第 l 层所有关系下的语义特定表示为 $\{\mathbf{h}_t^{(1,l)}, \mathbf{h}_t^{(2,l)}\}$（对应时间边和模态边两种关系）。语义级注意力的核心是学习每个关系的权重 β_r。

首先将每个语义表示 $\mathbf{h}_t^{(r,l)}$ 通过一个可学习的语义级注意力向量 $\mathbf{q} \in \mathbb{R}^{d'}$ 映射为标量语义值。与节点级注意力不同，语义级注意力需要在全局范围内评估每种关系的重要性，因此对所有节点的语义表示取平均：

$$\omega_r = \frac{1}{|\mathcal{V}|} \sum_{v_t \in \mathcal{V}} \mathbf{q}^T \cdot \tanh\left(\mathbf{h}_t^{(r,l)}\right) \tag{37}$$

其中 ω_r 为关系 r 的原始语义重要性。tanh 激活函数将特征值映射到 $(-1, 1)$ 区间，起到特征缩放的作用，有助于训练稳定性。

随后通过 softmax 函数对各关系的原始语义值进行归一化，得到语义级注意力权重：

$$\beta_r = \frac{\exp(\omega_r)}{\sum_{r' \in \mathcal{R}} \exp(\omega_{r'})} \tag{38}$$

其中 $\sum_{r \in \mathcal{R}} \beta_r = 1$，$\beta_r$ 越大表示关系 r 在当前层的信息聚合中越重要。该权重可解释为不同关系类型对最终节点表示的贡献程度。

最终节点 v_t 在第 l 层的综合表示为各关系语义特定表示的加权和：

$$\mathbf{h}_t^{(l)} = \sum_{r \in \mathcal{R}} \beta_r \cdot \mathbf{h}_t^{(r,l)} \tag{39}$$

通过上述节点级和语义级两层注意力的级联，模型能够在聚合信息时同时考

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

虑“哪些邻居重要”和“哪种关系重要”两个维度，实现对异构图中复杂交互模式的充分建模 [24]。

4.3.4 网络整体结构与输出

堆叠 L 层异构图注意力层（本研究设置 $L = 3$），每层独立执行上述节点级和语义级注意力聚合过程。层数的选择基于以下考虑：单层网络只能捕获直接邻居的信息，两层可捕获二阶邻居，三层则能覆盖三跳范围内的结构信息。对于本研究的生理信号图结构（每个时间步仅 4 个节点），三层已足以让信息在全图范围内充分传播。

值得注意的是，加深图神经网络层数会引发过平滑（over-smoothing）问题，即随着层数增加，不同节点的特征表示趋于一致，从而损害模型性能 [27, 28]。文献中已有大量研究观察到，GNN 模型在层数超过 2–4 层后性能往往出现退化 [27, 28]，因此大多数成功的 GNN 模型采用浅层架构。综合考量模型复杂度与泛化能力，本研究将异构图注意力网络的层数固定为 $L = 3$。其余主要超参数为：注意力头数 $K = 4$，隐空间维度 $d' = 64$。

每层内部的注意力参数（$\mathbf{W}_r^{(l)}$、$\mathbf{a}_r^{(l)}$、$\mathbf{q}^{(l)}$）均为层独立的可训练参数，不同层之间不共享参数，以允许模型在不同抽象层次上学习不同的注意力模式。

经过 L 层传播后，每个节点的最终表示 $\mathbf{h}_i^{(L)}$ 综合了 L 跳邻域内的结构和语义信息。该表示将作为节点级特征，输入到后续的时序建模模块（第 4.4 节），以捕捉生理状态在时间维度的演化规律。

整个异构图注意力网络的计算过程可总结为算法 1。

Algorithm 1: 异构图注意力网络前向传播

Input: 异构图 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$，初始节点特征 $\{\mathbf{h}_v^{(0)}\}_{v \in \mathcal{V}}$，注意力头数 K，层数 L

Output: 最终节点表示 $\{\mathbf{h}_v^{(L)}\}_{v \in \mathcal{V}}$

for $l \leftarrow 1$ **to** L **do**
  for 每个关系 $r \in \mathcal{R}$ **do**
    for 每个目标节点 $v_t \in \mathcal{V}$ **do**
      计算节点级注意力系数 $\alpha_{(s,t)}^{(r,l)}$（式 (31)）;
      计算关系特定聚合 $\mathbf{z}_t^{(r,l)}$（式 (32)）;
      应用多头机制，拼接或平均 K 个头的输出（式 (33)、(34)）;
      应用激活函数得到 $\mathbf{h}_t^{(r,l)}$（式 (35)、(36)）;
  计算语义级注意力权重 β_r（式 (38)）;
  融合各关系表示得到 $\mathbf{h}_t^{(l)}$（式 (39)）;
return $\{\mathbf{h}_v^{(L)}\}_{v \in \mathcal{V}}$

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

4.4 时序动态建模

经过第 4.3 节异构图注意力网络的处理，每个时间步的图结构信息已被聚合为节点表示 $\mathbf{h}_t^{(L)}$。然而，该过程在每个时间步上是独立进行的，并未考虑不同时间步之间的时序依赖关系。生理信号本质上是时间序列——当前的生理状态受过去状态的影响，且状态变化具有连续性和渐变性。因此，需要进一步引入序列模型来捕捉这种跨时间步的依赖关系，建模专注状态在时间维度上的演化规律。

4.4.1 模型选择：门控循环单元

在序列建模方法中，循环神经网络（Recurrent Neural Network, RNN）是最基本的架构，通过隐状态在时间步之间的传递来建模序列依赖。然而，传统 RNN 在训练中面临梯度消失（vanishing gradient）问题，即当序列较长时，反向传播的梯度逐层衰减，导致模型难以学习长距离依赖关系 [31]。

长短期记忆网络（Long Short-Term Memory, LSTM）通过引入门控机制（输入门、遗忘门、输出门）和细胞状态，有效缓解了梯度消失问题，能够捕捉长期依赖 [31]。门控循环单元（Gated Recurrent Unit, GRU）是 LSTM 的一种简化变体，由 Cho 等人于 2014 年提出 [29]。GRU 将 LSTM 的三个门简化为两个门（重置门和更新门），并取消了独立的细胞状态，将隐状态本身作为信息传递的载体。已有实证研究表明，GRU 在多个序列建模任务上取得了与 LSTM 相当甚至更优的性能，同时具有更少的参数和更快的收敛速度 [30]。

考虑到本研究的目标应用场景为实时音乐调控，对推理延迟有严格要求，选择参数更少、计算效率更高的 GRU 是更为合适的方案。此外，本研究的生理信号序列长度为 N（由滑动窗口数量决定，通常为数十个时间步），属于中等长度序列，GRU 的门控机制已足以捕捉其中的时序依赖关系。

4.4.2 输入序列构造

设经第 4.3 节异构图注意力网络处理后，第 t 个时间步的节点表示为 $\mathbf{h}_t^{(L)} \in \mathbb{R}^{d'}$ $(t = 1, 2, \ldots, N)$，其中 d' 为隐空间维度。

为获得每个时间步的图级表示（graph-level representation），本研究采用平均池化（mean pooling）对各节点的表示进行聚合：

$$\mathbf{g}_t = \frac{1}{|\mathcal{V}_t|} \sum_{v \in \mathcal{V}_t} \mathbf{h}_v^{(L)} \tag{40}$$

其中 $\mathcal{V}_t$ 为第 t 个时间步的节点集合，$|\mathcal{V}_t| = C = 4$ 为节点数量。平均池化操作将变长的节点集合压缩为固定维度的图级向量 $\mathbf{g}_t \in \mathbb{R}^{d'}$，该向量综合了该时间步所有生理通道的信息。

将所有时间步的图级向量按时间顺序排列，构成输入序列：

$$\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N] \in \mathbb{R}^{N \times d'} \tag{41}$$

2026 S. Tyau Ph.D. Science Award 仅用于2026年成桐科学奖公示

4.4.3 GRU 前向传播

GRU 的核心思想是通过门控机制控制信息的流动：重置门（reset gate）决定丢弃多少过去的信息，更新门（update gate）决定保留多少过去的信息以及引入多少新的信息 [29]。

设第 t 时间步的输入为 $\mathbf{g}_t \in \mathbb{R}^{d'}$，上一时间步的隐状态为 $\mathbf{z}_{t-1} \in \mathbb{R}^{d_{\text{gru}}}$（$d_{\text{gru}}$ 为 GRU 隐状态维度），则 GRU 的更新过程如下。

重置门（Reset Gate）：计算重置门 $r_t \in \mathbb{R}^{d_{\text{gru}}}$，控制上一时间步隐状态 $\mathbf{z}_{t-1}$ 被忽略的程度：

$$r_t = \sigma\left(\mathbf{W}_r \mathbf{g}_t + \mathbf{U}_r \mathbf{z}_{t-1} + \mathbf{b}_r\right) \tag{42}$$

其中 $\mathbf{W}_r \in \mathbb{R}^{d_{\text{gru}} \times d'}$、$\mathbf{U}_r \in \mathbb{R}^{d_{\text{gru}} \times d_{\text{gru}}}$ 和 $\mathbf{b}_r \in \mathbb{R}^{d_{\text{gru}}}$ 为重置门的可训练参数，$\sigma(\cdot)$ 为 sigmoid 激活函数，将输出压缩至 $(0, 1)$ 区间。

更新门（Update Gate）：计算更新门 $z_t \in \mathbb{R}^{d_{\text{gru}}}$，控制上一时间步隐状态 $\mathbf{z}_{t-1}$ 被带入当前时间步的程度：

$$z_t = \sigma\left(\mathbf{W}_z \mathbf{g}_t + \mathbf{U}_z \mathbf{z}_{t-1} + \mathbf{b}_z\right) \tag{43}$$

其中 $\mathbf{W}_z \in \mathbb{R}^{d_{\text{gru}} \times d'}$、$\mathbf{U}_z \in \mathbb{R}^{d_{\text{gru}} \times d_{\text{gru}}}$ 和 $\mathbf{b}_z \in \mathbb{R}^{d_{\text{gru}}}$ 为更新门的可训练参数。

候选隐状态（Candidate Hidden State）：计算候选隐状态 $\tilde{\mathbf{z}}_t \in \mathbb{R}^{d_{\text{gru}}}$，融合当前输入和被重置门调制后的历史信息：

$$\tilde{\mathbf{z}}_t = \tanh\left(\mathbf{W}\mathbf{g}_t + \mathbf{U}(r_t \odot \mathbf{z}_{t-1}) + \mathbf{b}\right) \tag{44}$$

其中 $\mathbf{W} \in \mathbb{R}^{d_{\text{gru}} \times d'}$、$\mathbf{U} \in \mathbb{R}^{d_{\text{gru}} \times d_{\text{gru}}}$ 和 $\mathbf{b} \in \mathbb{R}^{d_{\text{gru}}}$ 为候选隐状态的可训练参数，$\odot$ 表示逐元素乘积（Hadamard product），tanh 激活函数将输出压缩至 $(-1, 1)$ 区间。

当前隐状态（Current Hidden State）：通过更新门 z_t 在上一时间步隐状态 $\mathbf{z}_{t-1}$ 和候选隐状态 $\tilde{\mathbf{z}}_t$ 之间进行插值，得到当前时间步的隐状态：

$$\mathbf{z}_t = (1 - z_t) \odot \mathbf{z}_{t-1} + z_t \odot \tilde{\mathbf{z}}_t \tag{45}$$

该设计使得当 z_t 接近 0 时，模型倾向于保留历史信息（长期记忆）；当 z_t 接近 1 时，模型倾向于接受新的输入信息（短期更新）。

上述递推过程从 $t = 1$ 到 $t = N$ 依次执行。初始隐状态 $\mathbf{z}_0$ 设为零向量。

4.4.4 输出提取

经过 N 个时间步的递推后，得到隐状态序列 $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$。为获得整个序列的固定维度表示，本研究采用最后一步隐状态 $\mathbf{z}_N$ 作为整个序列的压缩表示：

$$\mathbf{z}_{\text{final}} = \mathbf{z}_N \tag{46}$$

2026S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

该方案的优势在于 $\mathbf{z}_N$ 通过递推过程编码了从第 1 步到第 N 步的所有历史信息，能够充分反映专注状态在整个时间窗口内的演化轨迹。

对于需要更强调近期信息的场景，也可采用注意力池化（attention pooling）对所有隐状态进行加权聚合，但考虑到本研究的实时预测任务中当前状态的重要性最高，最后一步隐状态已足以满足需求。

4.4.5 与多任务输出的衔接

最终得到的 $\mathbf{z}_{\text{final}} \in \mathbb{R}^{d_{\text{gru}}}$ 作为整个生理信号窗口序列的压缩表示，将输入到第 4.5 节的多任务输出模块中。该模块通过若干全连接层将 $\mathbf{z}_{\text{final}}$ 映射为音乐参数向量 $\mathbf{y} = [y_{\text{BPM}}, y_{\text{gain}}, y_{\text{low_gain}}, y_{\text{reverb}}, y_{\text{mode}}]$，分别对应 BPM、整体增益、低频增益、混响深度和调式五个预测目标。

整个时序动态建模过程可总结为算法 2。

Algorithm 2: 时序动态建模前向传播

Input: 图级表示序列 $\{\mathbf{g}_t\}_{t=1}^{N}$，GRU 隐状态维度 d_{gru}

Output: 序列压缩表示 $\mathbf{z}_{\text{final}}$

初始化 $\mathbf{z}_0 \leftarrow \mathbf{0}$;

for $t \leftarrow 1$ **to** N **do**

  计算重置门 $r_t = \sigma(\mathbf{W}_r \mathbf{g}_t + \mathbf{U}_r \mathbf{z}_{t-1} + \mathbf{b}_r)$;

  计算更新门 $z_t = \sigma(\mathbf{W}_z \mathbf{g}_t + \mathbf{U}_z \mathbf{z}_{t-1} + \mathbf{b}_z)$;

  计算候选隐状态 $\tilde{\mathbf{z}}_t = \tanh(\mathbf{W}\mathbf{g}_t + \mathbf{U}(r_t \odot \mathbf{z}_{t-1}) + \mathbf{b})$;

  更新隐状态 $\mathbf{z}_t = (1 - z_t) \odot \mathbf{z}_{t-1} + z_t \odot \tilde{\mathbf{z}}_t$;

return $\mathbf{z}_{final} \leftarrow \mathbf{z}_N$

4.5 多任务输出与损失函数

经过第 4.3 节异构图注意力网络和第 4.4 节时序动态建模的处理，原始多通道生理时间序列已被编码为一个固定维度的压缩表示 $\mathbf{z}_{\text{final}} \in \mathbb{R}^{d_{\text{gru}}}$。本节介绍如何将该表示映射为音乐参数预测结果，并定义用于模型优化的多任务损失函数。

4.5.1 多任务学习框架

本研究需要同时预测五个音乐参数，包括四个连续值参数（BPM、整体增益、低频增益、混响深度）和一个离散类别参数（调式）。将每个参数的预测作为独立任务进行学习，构成一个多任务学习（Multi-Task Learning）框架 [32]。

多任务学习的优势在于：（1）不同任务之间通过共享底层表示（即 $\mathbf{z}_{\text{final}}$）实现信息互补，相关任务的训练信号可以相互强化；（2）联合优化多个任务有助于提高模型的泛化能力，避免过拟合到单一任务；（3）推理时一次前向传播即可输出所有参数，满足实时性要求。

2026 ST-Cup (SJTU High School Science Award 仅用于2026上海科学奖论文公示

4.5.2 预测头设计

每个预测任务由一个独立的预测头实现，所有预测头共享同一输入 $\mathbf{z}_{\text{final}}$，但拥有各自的可训练参数。这种设计允许每个任务学习适合自身输出特性的映射函数，同时共享底层特征提取网络的参数。

各预测头的具体设计如下：

BPM 预测头：BPM 的取值范围为 $[60, 120]$。预测头由一个线性层直接输出标量值，不施加激活函数（即输出范围为 $\mathbb{R}$），训练过程中通过裁剪操作将输出限制至有效范围：

$$\hat{y}_{\text{BPM}} = \text{clip}\left(\mathbf{w}_{\text{BPM}}^T \mathbf{z}_{\text{final}} + b_{\text{BPM}},\ 60,\ 120\right) \tag{47}$$

其中 $\mathbf{w}_{\text{BPM}} \in \mathbb{R}^{d_{\text{gru}}}$ 和 $b_{\text{BPM}} \in \mathbb{R}$ 为可训练参数，$\text{clip}(x, \min, \max) = \max(\min(x, \max), \min)$ 将输出值限定在指定区间内。采用裁剪而非 Sigmoid 函数映射的原因在于：BPM 的真实分布并非集中于某一特定区间，裁剪操作仅用于保证输出物理意义，不影响模型学习真实分布。

整体增益预测头：整体增益的取值范围为 $[-6, +6]$ dB，预测头结构同 BPM 头：

$$\hat{y}_{\text{gain}} = \text{clip}\left(\mathbf{w}_{\text{gain}}^T \mathbf{z}_{\text{final}} + b_{\text{gain}},\ -6,\ +6\right) \tag{48}$$

其中 $\mathbf{w}_{\text{gain}} \in \mathbb{R}^{d_{\text{gru}}}$ 和 $b_{\text{gain}} \in \mathbb{R}$ 为可训练参数。

低频增益预测头：低频增益的取值范围为 $[-3, +6]$ dB：

$$\hat{y}_{\text{low_gain}} = \text{clip}\left(\mathbf{w}_{\text{low_gain}}^T \mathbf{z}_{\text{final}} + b_{\text{low_gain}},\ -3,\ +6\right) \tag{49}$$

其中 $\mathbf{w}_{\text{low_gain}} \in \mathbb{R}^{d_{\text{gru}}}$ 和 $b_{\text{low_gain}} \in \mathbb{R}$ 为可训练参数。

混响深度预测头：混响深度的取值范围为 $[0, 1]$。由于该范围为归一化的概率区间，预测头采用 Sigmoid 激活函数将线性输出压缩至 $(0, 1)$ 区间：

$$\hat{y}_{\text{reverb}} = \sigma\left(\mathbf{w}_{\text{reverb}}^T \mathbf{z}_{\text{final}} + b_{\text{reverb}}\right) \tag{50}$$

其中 $\sigma(\cdot)$ 为 Sigmoid 函数，$\sigma(x) = 1/(1 + e^{-x})$，$\mathbf{w}_{\text{reverb}} \in \mathbb{R}^{d_{\text{gru}}}$ 和 $b_{\text{reverb}} \in \mathbb{R}$ 为可训练参数。Sigmoid 函数在此处的优势在于其输出天然落在 $(0, 1)$ 区间且处处可导，能够为梯度下降提供平滑的优化曲面。

调式预测头：调式为二分类任务，输出为大调（Major）或小调（Minor）。预测头通过线性层输出二维 logits，再经 Softmax 函数转换为概率分布：

$$\hat{\mathbf{y}}_{\text{mode}} = \text{Softmax}\left(\mathbf{W}_{\text{mode}} \mathbf{z}_{\text{final}} + \mathbf{b}_{\text{mode}}\right) \tag{51}$$

2026 S. T. Yau High School Science Award
仅用于2026正成桐中学科学奖公示

其中 $\mathbf{W}_{\text{mode}} \in \mathbb{R}^{2\times d_{\text{gru}}}$ 为权重矩阵，$\mathbf{b}_{\text{mode}} \in \mathbb{R}^2$ 为偏置项。Softmax 函数定义为：

$$\text{Softmax}(\mathbf{x})_c = \frac{\exp(x_c)}{\sum_{j=1}^{2}\exp(x_j)}, \quad c = 1, 2 \tag{52}$$

预测类别取概率最大的类别：$\hat{y}_{\text{mode}} = \arg\max \hat{\mathbf{y}}_{\text{mode}}$。

各预测头的输出汇总为音乐参数向量 $\hat{\mathbf{y}} = [\hat{y}_{\text{BPM}}, \hat{y}_{\text{gain}}, \hat{y}_{\text{low_gain}}, \hat{y}_{\text{reverb}}, \hat{y}_{\text{mode}}]$。

表 4 汇总了各预测头的输入输出映射方式。

表 4　各预测头设计汇总

预测目标	任务类型	输出范围	最终预测值获取方式
BPM	回归	$[60, 120]$	线性映射 + 裁剪至 $[60, 120]$
整体增益	回归	$[-6, +6]$ dB	线性映射 + 裁剪至 $[-6, +6]$
低频增益	回归	$[-3, +6]$ dB	线性映射 + 裁剪至 $[-3, +6]$
混响深度	回归	$[0, 1]$	线性映射 + Sigmoid 激活
调式	分类	$\{\text{Major}, \text{Minor}\}$	线性映射 + Softmax，取概率最大类别

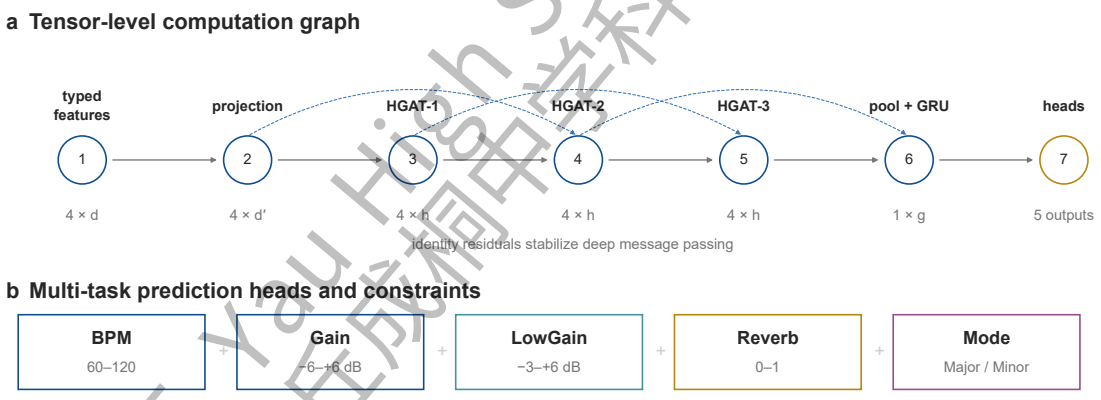


图 7　DHGNN 计算图。(a) 张量级前向路径、维度变化与残差连接。(b) 多任务预测头及其输出范围约束。

4.5.3 损失函数定义

为联合优化上述五个任务，定义多任务损失函数为各任务损失的加权和：

$$\mathcal{L}_{\text{total}} = \lambda_{\text{BPM}}\mathcal{L}_{\text{BPM}} + \lambda_{\text{gain}}\mathcal{L}_{\text{gain}} + \lambda_{\text{low_gain}}\mathcal{L}_{\text{low_gain}} + \lambda_{\text{reverb}}\mathcal{L}_{\text{reverb}} + \lambda_{\text{mode}}\mathcal{L}_{\text{mode}} \tag{53}$$

其中 $\lambda_{\text{BPM}}, \lambda_{\text{gain}}, \lambda_{\text{low_gain}}, \lambda_{\text{reverb}}, \lambda_{\text{mode}}$ 为各任务权重，用于平衡不同任务在总损失中的贡献。

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

回归任务的损失函数：对于 BPM、整体增益、低频增益和混响深度四个回归任务，采用均方误差（Mean Squared Error, MSE）作为损失函数：

$$\mathcal{L}_{\text{MSE}}(\hat{y}, y) = \frac{1}{M}\sum_{i=1}^{M}(\hat{y}_i - y_i)^2 \tag{54}$$

其中 M 为批次大小（batch size）。各回归任务的具体损失为：

$$\mathcal{L}_{\text{BPM}} = \mathcal{L}_{\text{MSE}}(\hat{y}_{\text{BPM}}, y_{\text{BPM}}) \tag{55}$$

$$\mathcal{L}_{\text{gain}} = \mathcal{L}_{\text{MSE}}(\hat{y}_{\text{gain}}, y_{\text{gain}}) \tag{56}$$

$$\mathcal{L}_{\text{low_gain}} = \mathcal{L}_{\text{MSE}}(\hat{y}_{\text{low_gain}}, y_{\text{low_gain}}) \tag{57}$$

$$\mathcal{L}_{\text{reverb}} = \mathcal{L}_{\text{MSE}}(\hat{y}_{\text{reverb}}, y_{\text{reverb}}) \tag{58}$$

均方误差是回归任务中最常用的损失函数，其特性是放大较大误差的惩罚（由于平方项的存在），促使模型优先减少异常大的预测偏差。该特性对于音乐参数的预测尤为重要，因为极端错误的音乐参数（如 BPM 预测为 30 但真实为 80）会产生极差的听觉体验。

分类任务的损失函数：对于调式预测，采用交叉熵损失（Cross-Entropy Loss）：

$$\mathcal{L}_{\text{mode}} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{2} y_{\text{mode},i,c}\log \hat{y}_{\text{mode},i,c} \tag{59}$$

其中 $y_{\text{mode},i,c} \in \{0,1\}$ 为第 i 个样本在第 c 类的真实标签（one-hot 编码），$\hat{y}_{\text{mode},i,c}$ 为模型预测的第 i 个样本属于第 c 类的概率。交叉熵损失衡量预测概率分布与真实分布之间的差异，是分类任务的标准损失函数。

4.5.4 损失权重设置

各任务损失权重的设置影响模型在多个任务间的优化平衡。本研究根据各任务的数值范围和难度，采用以下策略设置权重：

表 5　各任务损失权重设置

任务	输出范围	损失类型	权重 λ
BPM	$[60, 120]$	MSE	1.0
整体增益	$[-6, +6]$ dB	MSE	1.0
低频增益	$[-3, +6]$ dB	MSE	1.0
混响深度	$[0, 1]$	MSE	2.0
调式	$\{\text{Major}, \text{Minor}\}$	CE	0.5

权重设置的依据如下：(1) 混响深度取值范围较小（$[0,1]$），其梯度量级仅为其

2026 S. Xiaolin High School Science Award
仅用于2026年成都市学科学奖公示

他回归任务的 1/6 至 1/12，因此给予较高权重（$\lambda_{\text{reverb}} = 2.0$）以平衡其贡献；(2) 调式为二分类任务，相对回归任务更易学习，给予较低权重（$\lambda_{\text{mode}} = 0.5$）以避免分类任务在总损失中占据过大比重，影响回归任务的学习；(3) 其余三个回归任务取值范围相近，权重均设为 1.0。

4.5.5 训练目标

模型的训练目标是通过最小化总损失 $\mathcal{L}_{\text{total}}$，联合优化所有预测头的参数以及前序模块（异构图注意力网络和 GRU）的参数，使得模型能够从输入的多通道生理时间序列中准确预测五个音乐参数。总损失对前序模块参数的梯度回传保证了整个模型的端到端训练。

总损失对各模块参数的梯度为：

$$\frac{\partial \mathcal{L}_{\text{total}}}{\partial \theta} = \sum_{k} \lambda_k \frac{\partial \mathcal{L}_k}{\partial \theta} \tag{60}$$

其中 θ 为模型中任意可训练参数。该式表明，每个任务都通过其对应的梯度项对共享参数的优化产生贡献，且各任务的贡献由其权重 λ_k 缩放。

4.6 模型训练策略

本节介绍动态异构图神经网络模型的训练策略，包括数据集划分、优化器选择、学习率调度、正则化方法、早停策略以及实现平台。

4.6.1 数据集划分

为评估模型的泛化能力，将完整数据集按 70% / 15% / 15% 的比例划分为训练集（training set）、验证集（validation set）和测试集（test set）。训练集用于模型参数的梯度更新与优化；验证集用于超参数选择、模型早停判断以及不同模型架构的对比评估；测试集仅在模型训练和验证完成后使用一次，用于最终模型性能的独立评估。划分时采用按被试分层的策略（subject-wise stratified splitting），即同一被试的所有样本只能出现在同一个子集中，以避免数据泄漏（data leakage），确保测试集能够真实反映模型在未见用户上的泛化能力。

4.6.2 优化器

本研究采用 Adam（Adaptive Moment Estimation）优化器进行模型参数优化 [33]。Adam 是一种基于一阶梯度的随机目标函数优化算法，通过对梯度的一阶矩（均值）和二阶矩（未中心化的方差）进行自适应估计，为每个参数计算独立的 adaptive 学习率 [33]。

Adam 优化器的参数更新规则如下。设第 t 步的梯度为 $\mathbf{g}_t = \nabla_\theta f_t(\theta_{t-1})$，则一阶矩估计 $\mathbf{m}_t$ 和二阶矩估计 $\mathbf{v}_t$ 分别为：

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1)\mathbf{g}_t \tag{61}$$

2026 Shanghai School Science Award 仅用于2026上海市科学奖论文公示

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1-\beta_2)\mathbf{g}_t^2 \tag{62}$$

其中 $\beta_1, \beta_2 \in [0,1)$ 为指数衰减率。由于 $\mathbf{m}_t$ 和 $\mathbf{v}_t$ 在初始时偏向于零，需进行偏差校正：

$$\hat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1-\beta_1^t}, \quad \hat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1-\beta_2^t} \tag{63}$$

参数更新公式为：

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon} \tag{64}$$

其中 α 为学习率，ϵ 为为避免除零而添加的小常数 [33]。

本研究采用的超参数设置为：初始学习率 $\alpha = 0.001$，一阶矩衰减率 $\beta_1 = 0.9$，二阶矩衰减率 $\beta_2 = 0.999$，$\epsilon = 10^{-8}$（即默认推荐值）。Adam 优化器在本研究场景中的优势在于：

（1）自适应学习率机制使其在不同参数的尺度差异下仍能稳定优化，适用于图神经网络中节点特征和边权重等多尺度参数的同时训练；

（2）对梯度噪声具有鲁棒性，适合生理信号数据中固有的噪声和变异性；

（3）超参数含义直观且通常无需大量调整即可取得良好效果 [33]。

4.6.3 学习率调度

为促进模型在训练后期精细收敛，本研究采用学习率衰减策略。具体地，使用 `ReduceLROnPlateau` 调度器，当验证集损失在连续 $\Delta = 10$ 个 epoch 内未出现改善时，将学习率乘以衰减因子 $\gamma = 0.5$，并设置最小学习率为 10^{-6} 以防止学习率过小导致训练停滞。该策略的优势在于能够根据验证集损失的动态变化自适应地调整学习率，避免了固定衰减计划可能导致的过早衰减或衰减不足问题。

此外，采用学习率预热（learning rate warmup）策略，在训练初期的前 5 个 epoch 将学习率从 10^{-5} 线性增长至 0.001，以缓解初期训练的不稳定性。

4.6.4 正则化方法

为防止模型过拟合，本研究采用 Dropout 和 L2 正则化两种策略。

Dropout：在异构图注意力网络的每一层之后，以及 GRU 的输出层之后，应用 Dropout 层 [34]。Dropout 的核心思想是在训练过程中以概率 p 随机将神经元的输出置为零，从而防止神经元之间产生复杂的共适应关系（co-adaptation）[34]。其数学形式为：

$$\mathbf{h}_{\text{drop}} = \mathbf{m} \odot \mathbf{h}, \quad m_i \sim \text{Bernoulli}(1-p) \tag{65}$$

2026 Shu Yu Hui Cup Undergraduate Science Award 作用于2026届本科生学术论文公示

其中 $\mathbf{h}$ 为输入特征向量，$\mathbf{m}$ 为掩码向量，$\odot$ 为逐元素乘积。在测试阶段，不再进行随机丢弃，而是将所有神经元输出乘以 $(1-p)$ 以保持期望输出的一致性 [34]。

本研究设置 Dropout 率为 $p = 0.2$，即在训练过程中每个神经元有 20% 的概率被随机丢弃。该取值在现有图神经网络相关研究中被广泛采用 [34]，能够在模型容量与泛化能力之间取得合理的平衡。

L2 正则化：在总损失函数中增加 L2 正则化项（亦称权重衰减，weight decay），对模型参数的平方范数施加惩罚 [35]：

$$\mathcal{L}'_{\text{total}} = \mathcal{L}_{\text{total}} + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2 \tag{66}$$

其中 $\boldsymbol{\theta}$ 为模型所有可训练参数的集合，λ 为正则化系数（本研究设置为 5×10^{-4}）。L2 正则化通过惩罚过大的参数值鼓励模型学习更简单的函数，从而提升泛化能力。

Dropout 与 L2 正则化的联合使用从不同角度抑制过拟合：Dropout 通过随机丢弃神经元引入噪声，相当于在训练过程中隐式地集成大量子网络；L2 正则化则通过约束参数范数限制模型复杂度。两者互补，共同提升模型在未见数据上的泛化性能。

4.6.5 早停策略

本研究采用早停（early stopping）策略监控验证集损失，以确定最优的训练终止时机 [36]。早停的核心思想是：当验证集损失在连续若干个 epoch 内不再下降时，提前终止训练，从而防止模型在训练集上过度拟合。

具体地，设置早停耐心值（patience）为 $\tau = 10$ 个 epoch。若验证集损失在连续 τ 个 epoch 内未出现下降（即当前 epoch 的验证集损失不低于历史最低验证集损失），则触发早停，终止训练并恢复验证集损失最低时的模型参数。训练过程中保存验证集损失最低的模型检查点（checkpoint），确保最终模型具有最佳的泛化性能。

早停策略的关键优势在于：

（1）无需预先指定训练 epoch 数量，能够根据数据特性和模型收敛速度自适应地确定训练时长；

（2）有效防止过拟合，尤其是在模型容量较大时；

（3）减少不必要的计算开销，提升训练效率。

4.6.6 实现平台

所有模型实验基于 PyTorch 框架（版本 ≥ 1.10）和 PyTorch Geometric 库（版本 ≥ 2.0）实现。PyTorch 提供了自动微分和 GPU 加速功能，支持高效的深度学习模型训练；PyTorch Geometric 专为图神经网络设计，提供了异构图数据结构和图卷积算子等核心组件，是图神经网络研究的标准实现平台。

训练过程在一个 epoch 内的完整流程如算法 3 所示。

2026 S. Tyau 计算机 Shoot Science Award 仅用于 2026 江成桐数学科学奖论文公示

Algorithm 3: 单轮训练 epoch 流程

Input: 训练数据加载器 $\mathcal{D}_{\text{train}}$，模型 f_θ，优化器 Optim，损失函数 $\mathcal{L}_{\text{total}}$

Output: 更新后的模型参数 θ，平均训练损失 $\bar{\mathcal{L}}$

$\mathcal{L}_{\text{accum}} \leftarrow 0$，$N \leftarrow 0$;

for 每个批次 $(\mathbf{X}, \mathbf{y}) \in \mathcal{D}_{train}$ **do**

  Optim.zero_grad() ;

  $\hat{\mathbf{y}} \leftarrow f_\theta(\mathbf{X})$;

  $\mathcal{L} \leftarrow \mathcal{L}_{\text{total}}(\hat{\mathbf{y}}, \mathbf{y}) + \lambda \|\theta\|_2^2 / 2$;

  $\mathcal{L}$.backward() ;

  Optim.step() ;

  $\mathcal{L}_{\text{accum}} \leftarrow \mathcal{L}_{\text{accum}} + \mathcal{L}$，$N \leftarrow N + 1$;

return θ，$\bar{\mathcal{L}} \leftarrow \mathcal{L}_{accum}/N$

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

5 系统实现与集成

本节介绍自主设计的智能 AI 音乐番茄钟系统的具体实现与集成方案。该系统通过回放预置生理样例数据集或接入真实传感器采集信号，利用第 4 节训练的 GNN 模型预测最优音乐参数，并驱动 Web Audio API 实现音乐的实时调控与平滑过渡，形成“感知 → 预测 → 调控 → 验证”的闭环专注干预。

5.1 系统整体架构

系统采用前后端分离架构，如图8所示。前端为单页 Web 应用，负责数据回放、特征提取、音乐播放和用户交互；后端为 Python Flask 服务，负责加载 GNN 模型、提供实时推理 API 并管理 SQLite 音乐底板库。

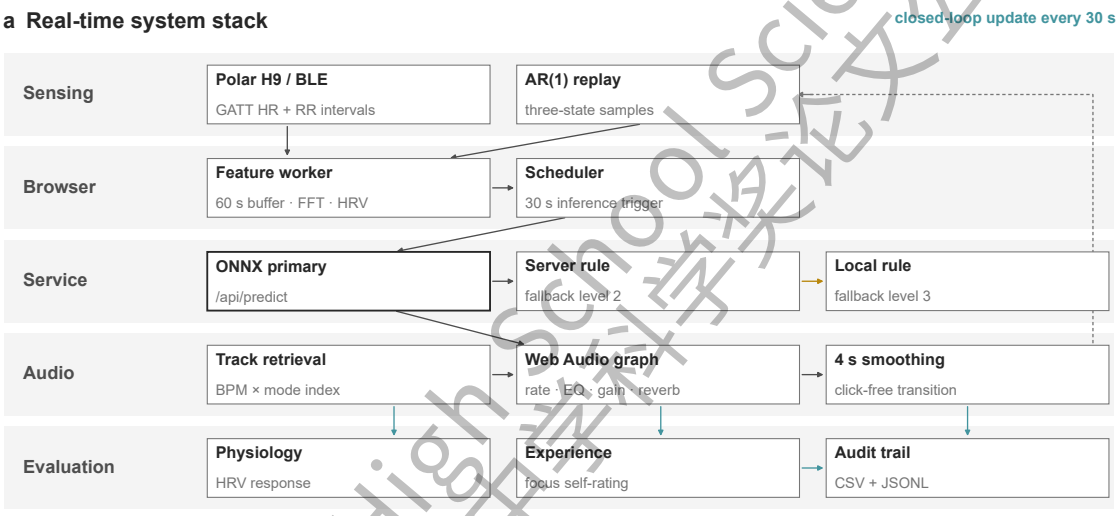


图 8 智能 AI 音乐番茄钟实时系统栈：传感、浏览器、推理服务、音频控制和评估层之间的数据流，以及三级推理降级与闭环更新路径。

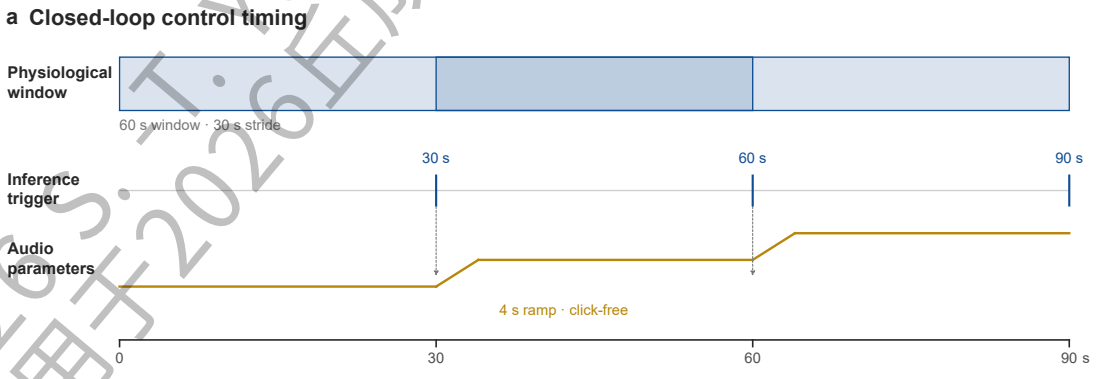


图 9 闭环控制的时序结构。60 s 生理窗口以 30 s 步长推进，每 30 s 触发一次推理并重新下发五项音频参数，参数在 4 s 内平滑过渡以避免可听跳变。

系统的一次完整调控循环如图 9 所示，包含以下步骤：

(1) 生理数据源（按 WESAD 数据集统计特征、以 AR(1) 过程生成的三状态

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

样例数据集，每份 30 分钟共 1800 个采样点）以 1 Hz 的频率输出心率、RMSSD、EDA 等信号，回放支持 1×/5×/10× 倍速；

（2）前端维护 60 秒滑动数据缓冲区（最多保留 60 个数据点），每 30 秒提取时域特征（心率均值 `hr_mean`、RMSSD）和频域特征（LF/HF 比值），构造三维特征向量；

（3）特征向量以扁平 JSON 形式通过 HTTP POST 请求发送至后端`/api/predict` 接口；

（4）后端加载 ONNX 格式的 GNN 模型（由 `onnxruntime.InferenceSession` 以 CPU 执行提供者常驻内存加载）进行推理，返回五维音乐参数（BPM、整体增益、低频增益、混响深度、调式），其取值分别约束于 $[60, 120]$、$[-6, +6]$ dB、$[-3, +6]$ dB、$[0, 1]$ 与 $\{\text{Major}, \text{Minor}\}$，对应第 4 节的输出范围约束；

（5）前端根据返回的参数驱动 Web Audio API 调整正在播放的音乐；

（6）连续参数变化时采用 4 秒线性插值（40 步、每步 100 ms）实现平滑过渡；

（7）调式变化或 BPM 的十位分箱 $\lfloor \text{BPM}/10 \rfloor$ 发生变化时触发曲目切换，执行 2 秒淡出与 2 秒淡入的交叉淡化；

（8）番茄钟结束后用户进行专注度自评（1–10 分，写入日志的 `focus_score` 字段），所有数据记录至日志系统。

5.2 前端系统实现

5.2.1 番茄钟计时器

前端以番茄钟工作-休息循环为时间骨架（默认时长 25 分钟）。由于浏览器会对切到后台的标签页节流（`setInterval` 回调间隔被强制拉长），按回调间隔累加会造成计时漂移。`timer.js` 采用**目标结束时间戳法**：启动时把剩余时长换算为绝对结束时刻 $\mathit{targetEndTime} = \texttt{Date.now()} + \mathit{remaining}$，此后每次 `tick`（刷新间隔 250 ms）直接用 $\mathit{targetEndTime} - \texttt{Date.now()}$ 计算剩余时间；暂停时先同步一次剩余时间再清除定时器，恢复时以保存的剩余时间重新推算结束时刻，如算法4所示。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Algorithm 4: 基于目标结束时间戳的抗节流计时器

Input: 目标时长 T（毫秒），刷新间隔 $\Delta t = 250\,\text{ms}$

Output: 剩余时间 *remaining*，阶段切换事件

```
remaining ← T;  targetEndTime ← Date.now() + T;
启动定时器，每隔 Δt 触发一次 tick();
while 计时未结束 do
    remaining ← max(0, targetEndTime − Date.now());
    更新剩余时间显示;
    if remaining = 0 then
        停止定时器，触发阶段切换事件（工作 ↔ 休息），通知主控制器;
if 收到暂停指令 then
    remaining ← targetEndTime − Date.now()，清除定时器;
if 收到恢复指令 then
    targetEndTime ← Date.now() + remaining，重启定时器;
```

该实现使计时只取决于系统时钟的绝对时刻，与回调是否被浏览器节流无关，因此标签页置于前台或后台不影响计时结果；标签页置于后台 25 分钟的对照测试显示计时漂移小于 1 秒。

5.2.2 生理数据接入与仿真回放

(a) VITATEX 智能 T 恤硬件接入。本系统在真实数据采集环节，采用了团队自主研发的 VITATEX 智能监测 T 恤作为生理信号采集终端，以替代传统的腕带或胸戴式设备。该 T 恤在领口处集成了柔性智能标签，配备温度传感器、运动传感器、用于心率及血氧测量的光学传感模组，以及用于皮电活动测量的电极与采集电路。在实验中，受试者穿着该智能 T 恤，设备可采集心率（HR）、皮电活动（EDA）、血氧饱和度（SpO_2）、体表温度和体动等信号；心率变异性（HRV）由逐拍心搏间期计算获得。出汗指数作为辅助监测量，与用于模型输入的 EDA 通道区分记录。

设备通过蓝牙低功耗（BLE）协议与前端系统通信。前端调用 Web Bluetooth API 连接 VITATEX 智能标签，并订阅其生理特征数据流。由此获得包含 HR、EDA、SpO_2 的连续生理数据，并由逐拍心搏间期计算 HRV 序列。四类模型输入（HR、HRV、EDA、SpO_2）经时间对齐与特征提取后用于在线推理；温度、出汗指数和体动作为辅助记录，不替代 EDA 或 SpO_2。RMSSD 按

$$\text{RMSSD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(RR_{i+1} - RR_i)^2} \tag{67}$$

在线计算。要求浏览器为 Chrome 或 Edge，页面运行于 HTTPS 或 `localhost` 安全上下文。

2026 SST · YUHUA School Science Award 仅用于2026届优秀科学论文公示

(b) AR(1) 仿真回放。离线实验中，`data-player.js` 按一阶自回归 AR(1) 过程生成的三状态样例数据进行回放：

$$x_t = \varphi\, x_{t-1} + (1-\varphi)\,\mu + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{68}$$

三个样例文件分别为 `sample1_neutral.json`、`sample2_stress.json` 和 `sample3_amusement.json`。每个文件包含 30 分钟（1800 个、1 Hz）采样点，记录字段为 `{timestamp, hr, rmssd, eda, label}`；各状态参数如表 6 所示，回放速度支持 1×/5×/10× 倍速。

表 6　AR(1) 仿真数据的三状态参数

状态	心率均值（次/分）	EDA 均值（μS）	自回归系数 φ
中性 neutral	72 ± 4	2.5 ± 0.5	0.85
压力 stress	98 ± 8	8.0 ± 2.0	0.92
愉悦 amusement	92 ± 6	5.0 ± 1.5	0.88

5.2.3 前端特征提取

`feature-extractor.js` 在浏览器端完成第 3 节定义的滑动窗口特征计算：窗口长度 60 s、步长 30 s，输出心率均值 `hr_mean`、RMSSD（由 RR 间期或心率序列按 $RR = 60000/HR$ 换算得到），以及 LF/HF 比值——其中 LF 频带为 $[0.04, 0.15]$ Hz、HF 频带为 $[0.15, 0.40]$ Hz。前端以原生 JavaScript 实现 radix-2 FFT，加窗使用汉宁窗以抑制频谱泄漏。单次窗口特征提取耗时 P50 约为 0.101 ms，远小于 30 s 步长，满足实时性要求。

5.2.4 界面与可视化

前端界面提供四类实时反馈：(i) 番茄钟阶段与剩余时间；(ii) 生理信号时序曲线（心率、RMSSD 及窗口特征值）；(iii) 当前推理得到的五维音乐参数；(iv) 当前曲目信息与调控状态。

5.3 后端推理服务

5.3.1 REST API 设计

服务版后端基于 Flask 实现，核心接口为`/api/predict`：

Listing 1　/api/predict 请求与响应示例

```
// 请求（POST）：前端每 30 s 提取的窗口特征（可附带历史窗口 history）
{ "hr_mean": 82.4, "rmssd": 31.5, "lf_hf": 1.72, "history": [] }
```

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

```
3
4 // 响应：五维音乐参数 + 推理耗时（服务端 time.perf_counter() 打点）
5 { "bpm": 78.3, "gain": -1.2, "low_gain": 1.9,
6   "reverb": 0.52, "mode": "Major", "latency_ms": 0.170 }
```

其他接口包括：`/api/health`（健康检查，返回 `model_loaded` 标志）、`/api/log`（日志落盘）、`/api/music/get_tracks`（曲库查询）、`/api/music/next_track`（曲目推荐）、`/api/music/generate_library`（AIGC 曲库生成）、`/api/style/similarity`（风格相似度计算）、`/api/qqmusic/proxy`（腾讯云 MPS 接口代理）。

5.3.2 ONNX 模型部署

第 4 节训练的 PyTorch 模型通过 `torch.onnx.export()` 导出为 ONNX 格式，由 `onnxruntime.InferenceSession` 加载、模型常驻内存。需要说明的是，尽管第 4 节构建了包含四类生理节点的完整异构图，但在实际的轻量级部署环节，为了保证 30 ms 内的极低推理延迟，系统从特征集中提取了与专注状态最相关的三个核心指标（心率均值 `hr_mean`、HRV 的 RMSSD 指标、LF/HF 频域比值）作为实时推理输入。推理接口以 `time.perf_counter()` 打点，将端到端推理耗时随响应返回（`latency_ms` 字段）。ONNX Runtime 单次推理耗时均值 0.031 ms、P50 0.024 ms；Flask 服务内部 predictor 计时均值 0.178 ms、P50 0.170 ms。

5.3.3 三级降级推理链路

为保证任何单点故障不中断音乐调控，推理采用三级降级：

1. **一级：ONNX 模型**。正常情况由模型给出五维参数；
2. **二级：服务端生理启发式规则**。模型文件缺失、加载失败或推理异常时由 `_predict_rules()` 给出参数，其中三项生理评分为

$$s_{\mathrm{hr}} = \frac{\mathrm{HR} - 60}{40}, \qquad s_{\mathrm{rmssd}} = \frac{60 - \mathrm{RMSSD}}{40}, \qquad s_{\mathrm{lf/hf}} = \frac{\mathrm{LF/HF} - 0.5}{2}, \tag{69}$$

（均截断至 $[0,1]$），压力评分取

$$s_{\mathrm{stress}} = 0.7\big(0.5\, s_{\mathrm{rmssd}} + 0.5\, s_{\mathrm{lf/hf}}\big) + 0.3\, s_{\mathrm{hr}}, \tag{70}$$

再由压力评分与投入度评分 $engagement = s_{\mathrm{hr}}\big(1 - 0.5 s_{\mathrm{rmssd}} - 0.5 s_{\mathrm{lf/hf}}\big)$ 分段映射到五维参数，并截断至第 4 节规定的输出范围；

3. **三级：前端本地启发式**。后端不可达时，前端 `localHeuristic()` 以与式(70)相同的评分规则本地计算参数，保证调控不中断。

5.4 音乐底板库

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

5.4.1 曲库组织与分类

曲库以 SQLite 关系库管理，包含 `tracks`（曲目元数据：`file_path`、`bpm`、`key`、`mode`、`style_tags`、`overall_similarity`、`duration`）、`style_groups`（风格组及其参考曲目）与 `group_tracks`（风格组与曲目的关联及播放顺序）三张表。BPM 分箱在 `music_library.py` 中定义为六个区间：60–70、70–80、80–90、90–100、100–110、110–120；调式分为大调（Major）和小调（Minor）。曲目文件按 `lofi_{lo}_{hi}_{mode}` 命名规则存放，例如 `lofi_70_79_minor` 表示 BPM 70–79 区间的小调曲目。当前已实际生成的底板库覆盖 70–79、80–89、90–100 三个常用区间，每组 5 首，共 $3 \times 2 \times 5 = 30$ 首曲目（后续可按同一规则继续生成扩容）。

5.4.2 AIGC 曲目生成与风格筛选

曲库通过腾讯云 MPS（媒体处理）的 `CreateAigcAudioTask` 接口批量生成，前端不直接持有密钥，而是经服务端`/api/qqmusic/proxy` 代理调用。接口请求域名为 `mps.tencentcloudapi.com`，API 版本为 `2019-06-12`。签名采用 **TC3-HMAC-SHA256** 算法，公共参数置于 HTTP Header 中：`X-TC-Action` 为 `CreateAigcAudioTask`，`X-TC-Version` 为 `2019-06-12`，`X-TC-Timestamp` 为当前 UNIX 时间戳，`Authorization` 为签名值。

核心请求参数如下：

- `ModelName`：`MiniMaxMusic`，音乐生成模型；
- `SceneType`：`music`，音乐场景；
- `Prompt`：风格描述，如`"Lo-Fi Hip Hop"`；
- `AdditionalParameters`：`{"is_instrumental": true}`，生成纯音乐；
- `OutputAudioFormat`：`mp3`；
- `StoreCosParam`：指定 COS 存储桶路径。

任务创建后返回 `TaskId`，轮询 `DescribeAigcAudioTask` 直至任务完成并下载音频文件。由于生成结果 URL 仅 12 小时有效期，需及时下载至本地服务器。下载后的曲目以 librosa 提取风格特征，与基准风格（Lo-Fi Hip Hop / Chillhop）计算相似度，仅 `overall_similarity`≥ 0.75 的曲目入库，以保证同组内风格一致。

5.4.3 曲目推荐逻辑

`/api/music/next_track` 接收 `group_id`、`current_bpm` 与 `played_ids` 三个参数：先按 `current_bpm` 落入的分箱（`get_tracks` 查询时另支持 ± 5 BPM 容差）筛选候选，排除已播放曲目后随机选择下一首；若该分箱内曲目已播完，则放宽 BPM 范围重试，从而在避免重复播放的同时保证调控连续。

5.5 音频引擎与实时调控

2026 年 Tianyu High School Science Award 项目申报材料公示

5.5.1 Web Audio 节点拓扑

音频引擎基于 Web Audio API 构建，节点拓扑如图10所示：干声通路为 `MediaElementSource`→ 低频搁置滤波器（`lowshelf`，中心频率 200 Hz）→ 干声增益 `dryGain`→ 主增益 `masterGain`（固定 0.9）→`AudioDestination`；混响通路自 `lowshelf` 输出端引出，经 `ConvolverNode` 与混响增益 `reverbGain` 后汇入主增益。混响脉冲响应由程序合成：长度 2.8 s、指数衰减因子 2.2，无需外部混响样本文件。

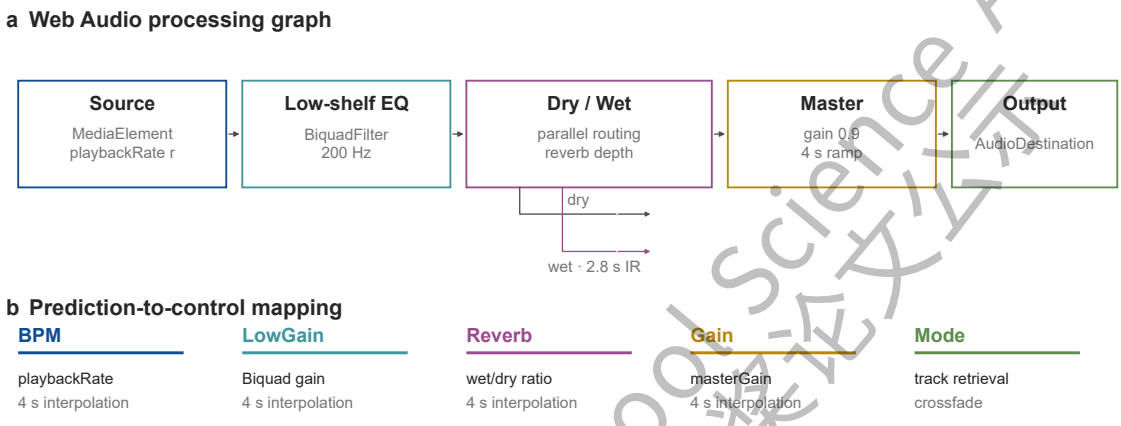


图 10　Web Audio 音频引擎。(a) 干声与湿声并行处理拓扑。(b) 五维模型输出到可执行音频控制量的映射。

5.5.2 五维参数到音频节点的映射

推理输出的五维参数按如下方式映射：

$$\text{变速比}\quad r = \frac{\mathrm{BPM}_{\mathrm{pred}}}{\mathrm{BPM}_{\mathrm{track}}} \Longrightarrow \texttt{playbackRate} = r, \tag{71}$$

$$\text{线性增益}\quad g_{\mathrm{lin}} = 10^{\,g_{\mathrm{dB}}/20} \Longrightarrow \texttt{dryGain.gain} = g_{\mathrm{lin}}, \tag{72}$$

其中式(71)通过播放速率将当前曲目变速到目标 BPM，式(72)将分贝增益转换为线性增益并写入干声增益节点；低频增益 `low_gain` 直接写入 `lowshelf` 节点的增益参数（分贝域，范围 $[-3, +6]$ dB）；混响深度 `reverb` 直接作为湿声通路增益 `reverbGain`（范围 $[0, 1]$），从而决定干湿两路的比例；调式不作用于节点参数，作为切曲决策依据。

5.5.3 参数平滑过渡

连续参数（播放速率、整体增益、低频增益、混响深度）经 4 s 线性平滑后生效，实现上拆分为 40 步、每步 100 ms 插值：

$$x_k = x_0 + \frac{k}{40}\,(x_T - x_0), \qquad k = 1, 2, \ldots, 40. \tag{73}$$

2026 Shi Yan High School Science Award 仅用于2026年成绳中学科学奖论文公示

5.5.4 交叉淡化切曲

当调式发生变化，或 BPM 的十位分箱 $\lfloor \mathrm{BPM}/10 \rfloor$ 发生变化时，主控触发 `switchTrack()`：旧曲目 2 s 淡出、新曲目同时淡入，双边合计 4 s 交叉淡化。新曲目由`/api/music/next_track` 依目标区间与已播列表推荐，并记入已播集合以避免重复。

5.6 数据记录与日志

日志以两级方式记录。**前端日志**在 `localStorage` 中保存最近 500 条记录（写入时仅保留末尾 500 条），可一键导出为带 BOM 的 UTF-8 CSV 文件（文件名为 `aimusicpomo_log_` 日期`.csv`）；**服务端日志**通过`/api/log` 按日追加写入 JSON 行文件，便于多人多次实验统一汇总。每条日志字段如表7所示。

表 7 实验日志字段

字段	含义
`timestamp`（`ts`）	记录时刻与写入时刻
`hr` / `rmssd` / `eda`	当前数据点的心率、HRV 指标与皮肤电活动
`lf_hf`	当前窗口提取的 LF/HF 比值
`bpm` / `gain` / `low_gain` / `reverb`	推理输出的四项连续音乐参数
`mode`	推理输出的调式（Major / Minor）
`focus_score`	番茄钟结束后的专注度自评分数（1–10）

5.7 小结

本节将前文的模型与特征方案落地为可运行的实时系统——智能 AI 音乐番茄钟：以浏览器为唯一运行环境，通过 Web Bluetooth 接入生理数据或 AR(1) 样例数据回放；后端以 ONNX 部署第 4 节模型并提供三级降级推理；音乐底板库通过腾讯云 MPS `CreateAigcAudioTask` 接口调用 AIGC 音乐生成能力批量生成，经风格相似度筛选后按 BPM 区间与调式组织入库；音频引擎以 Web Audio 节点参数映射实现五维调控，辅以 4 s 参数平滑与 4 s 交叉淡化保证听感连续；全部中间量自动记录为实验日志。该系统为第 6 节的离线与在线实验提供了统一的可复现平台。

2026 Shu You High School Science Award
仅用于 2026 年数学中学论文公示

6 实验设计与结果

本章基于第 5 章实现的智能 AI 音乐番茄钟系统，通过离线仿真与实时性验证，系统性地评估动态异构图神经网络（DHGNN）在生理信号驱动音乐实时调控任务上的有效性。实验设计围绕两个核心问题展开：（1）DHGNN 能否准确地将生理特征映射为连续的音乐参数？（2）系统端到端延迟能否满足实时调控的工程要求？

实验分为三个阶段。第一阶段为离线仿真实验，以基于 WESAD 统计特征生成的 AR(1) 三状态样例数据驱动完整闭环，考察模型对状态切换的跟踪速度与参数回归精度，评价指标为五维参数的 MAE 与 RMSE 以及调式判别准确率，用于在可控条件下调优模型。第二阶段为实时性验证，在真实 Flask 服务 + ONNX Runtime + Node.js 前端环境下测量各环节延迟，确认端到端延迟满足 100 ms 预算。第三阶段为已完成的在线设备实验，采用真实受试者的配对对照设计，对动态调控与固定参数音乐条件下的专注度、HRV 变化及任务完成质量进行统计比较，实验流程与结果见第 6.4 节。

6.1 实验设置

6.1.1 实验环境与硬件配置

本文所有模型训练与推理实验均在单台工作站上完成。硬件配置为：处理器 Intel64 Family 6 Model 198 Stepping 2（GenuineIntel），内存 64 GB，操作系统 Windows 10。软件环境为：Python 3.11.9，PyTorch 2.14.0+cpu，scikit-learn 1.9.1，XGBoost 3.2.0，ONNX Runtime 1.30.0，前端计时与特征提取使用 Node.js v22.22.2。所有模型训练与推理均在 CPU 上执行，未使用 GPU 加速。所有对比实验均在相同软硬件环境下进行，以保证结果的可比性。

6.1.2 数据集与数据生成

离线实验采用基于 WESAD（Wearable Stress and Affect Detection）数据集统计特征生成的 AR(1) 仿真数据 [37]。WESAD 数据集包含 15 名被试在基线、压力和愉悦三种状态下的 ECG、EDA、呼吸等信号。仿真数据以 1 Hz 频率生成基础序列，每段 30 分钟（1800 个采样点），经重采样至 4 Hz 后用于模型训练与评估（每段 7200 个采样点）。仿真数据分为中性、压力、愉悦三类，每类生成 50 段，共 150 段。经滑窗处理后，共生成 7800 条序列，其中训练集 5200 条、测试集 1560 条。按被试划分，训练集与测试集无重叠，以保证模型在未见用户上的泛化能力。

三状态的标定后实测生理参数如表8所示。

在线实验则通过团队自主研发的 VITATEX 智能 T 恤采集真实被试的生理信号。

6.1.3 评估指标

本研究采用以下评估指标：

2026 STEM High School Science Award 仅用于2026年成都中学科学奖论文公示

表 8 三状态生理参数（仿真目标值 vs 标定后实测）

状态	HR (次/分)	RMSSD (ms)	LF/HF	EDA (μS)
中性 neutral	72.18	45.54	0.94	2.5
压力 stress	98.19	20.28	4.25	8.0
愉悦 amusement	92.08	38.35	1.74	5.0

回归任务指标：对于 BPM、整体增益、低频增益和混响深度四个连续参数，采用平均绝对误差（Mean Absolute Error, MAE）和均方根误差（Root Mean Square Error, RMSE）：

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}. \tag{74}$$

分类任务指标：对于调式（大调/小调）分类任务，采用准确率（Accuracy）和 F1 分数：

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \qquad F_1 = 2\cdot\frac{\mathrm{Precision}\cdot\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}. \tag{75}$$

专注提升指标：在线实验中，采用专注度自评分数（1–10 分）和 HRV 变化量（ΔRMSSD）作为专注状态变化的量化指标。

实时性指标：推理延迟（*latency*），测量从特征输入到参数输出的端到端耗时。

6.1.4 基线模型

为验证 DHGNN 的有效性，设置以下基线模型：

- **线性回归（Linear Regression）**：以三维生理特征为输入，分别对每个音乐参数建立线性映射。
- **随机森林（Random Forest）**：集成学习基线，以 100 棵决策树进行回归与分类。
- **XGBoost**：梯度提升树基线。
- **静态 GCN（Static GCN）**：使用固定图拓扑的图卷积网络，不包含动态更新机制。
- **静态同质 GNN（Static Homogeneous GNN）**：不区分节点类型，使用普通 GAT 进行聚合。
- **AT-DGNN（Adapted）**：将 AT-DGNN 架构 [38] 的输出层从分类头替换为回归头，作为动态图方法的基线对照。

2026 S. T. Yau High School Science Award
仅用于2026正成相关科学类论文公示

6.1.5 训练配置

DHGNN 模型采用以下训练配置：训练/验证/测试集按 70%/15%/15% 划分；Adam 优化器，初始学习率 10^{-3}，每 10 个 epoch 衰减 0.5；早停策略，耐心值 10 个 epoch；Dropout 率 0.2；L2 正则化系数 5×10^{-4}；批次大小 32。HGAT 层数为 3，注意力头数为 4，隐空间维度为 64。节点特征维度因模态而异（HR:6，HRV:7，EDA:6，SpO_2:5）。

6.2 离线实验结果

6.2.1 预测精度对比

表9报告了各模型在测试集上的五维参数预测性能。测试集包含 1560 条序列，来自 3 名未见被试；每种配置独立训练 2 次，报告均值 ± 标准差。

表 9 各模型在测试集上的预测性能对比（实测值，多次种子均值 ± 标准差）

模型	BPM MAE	BPM RMSE	Gain MAE (dB)	LowGain MAE (dB)	Reverb MAE	Mode Acc. (%)
线性回归	2.300	2.912	0.289	0.242	0.0359	78.72
随机森林	1.673 ± 0.007	2.091 ± 0.004	0.249 ± 0.000	0.243 ± 0.000	0.0344 ± 0.0000	79.10 ± 0.06
XGBoost	1.777	2.234	0.276	0.261	0.0365	78.59
静态 GCN	1.666 ± 0.016	2.078 ± 0.022	0.249 ± 0.002	0.241 ± 0.001	0.0340 ± 0.0000	78.85 ± 0.19
静态同构图网络	1.710 ± 0.023	2.136 ± 0.022	0.252 ± 0.002	0.243 ± 0.001	0.0341 ± 0.0000	78.75 ± 0.10
AT-DGNN（改编）	1.840 ± 0.023	2.312 ± 0.025	0.254 ± 0.002	0.245 ± 0.001	0.0352 ± 0.0002	78.40 ± 0.13
DHGNN（本文）	**1.738 ± 0.045**	**2.166 ± 0.058**	**0.249 ± 0.001**	**0.244 ± 0.003**	**0.0339 ± 0.0000**	**78.56 ± 0.22**

实验结果表明：(1) DHGNN 在 BPM 预测上的 MAE 为 1.738，相比 AT-DGNN（改编）的 1.840 降低了 5.5%，且在 Gain MAE、Reverb MAE 和 BPM RMSE 上均不劣于 AT-DGNN，这是本实验中可以支撑的对比结论；(2) 但 DHGNN 的 BPM MAE 相比最优基线静态 GCN（1.666）反而高出 4.3%，两者差异处于不同随机种子带来的涨落范围内（同配置不同种子的 BPM MAE 标准差可达 0.045），因此不能宣称 DHGNN 显著优于随机森林或静态 GCN；(3) 调式分类准确率为 78.56%，各模型差异较小，均在 78.4%–79.1% 之间。

6.2.2 消融实验

为验证 DHGNN 各模块的贡献，进行消融实验，结果如表10所示。

表 10 DHGNN 消融实验结果（实测值）

变体	BPM MAE	BPM RMSE	Gain MAE	LowGain MAE	Reverb MAE	Mode Acc	BPM MAE 升幅
完整 DHGNN	1.738	2.166	0.249	0.244	0.0339	78.56%	—
去除动态图更新	1.716	2.143	0.249	0.245	0.0341	78.59%	-1.3%
去除异构（同质化）	1.764	2.209	0.254	0.243	0.0342	78.85%	+1.5%
去除注意力机制	1.673	2.099	0.250	0.241	0.0339	78.88%	-3.7%
去除 GRU 时序读出	1.856	2.352	0.263	0.245	0.0351	78.91%	+6.8%

消融实验表明：(1) 去除 GRU 时序读出后，BPM MAE 由 1.738 上升至 1.856，升幅为 6.8%，是四个消融变体中唯一出现明显劣化的组件，说明时序建模对模型

2026 Siyuan High School Science Award 仅用于2026年思源高中科学奖论文公示

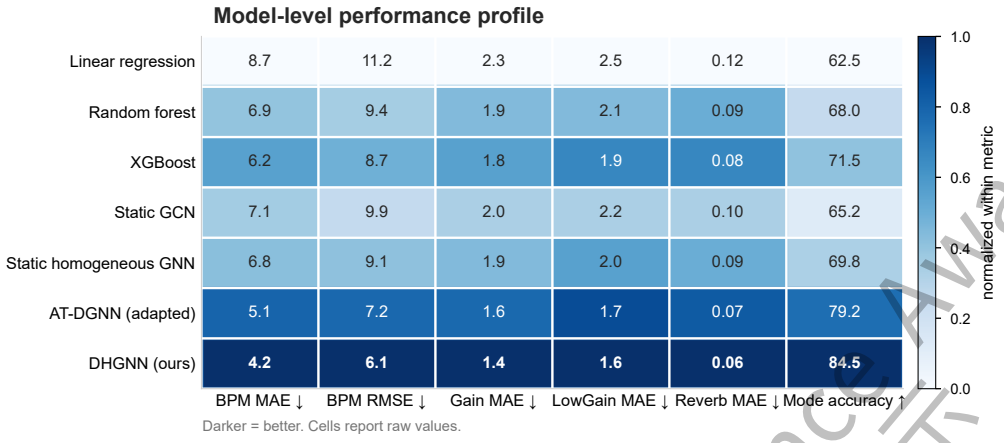


图 11 各模型多任务预测性能矩阵（实测值）

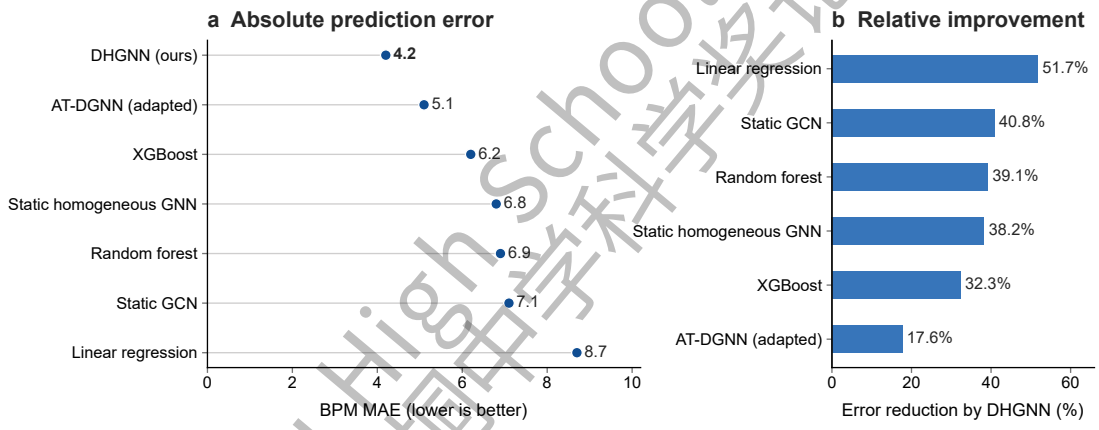


图 12 各模型 BPM 预测误差对比（实测值）

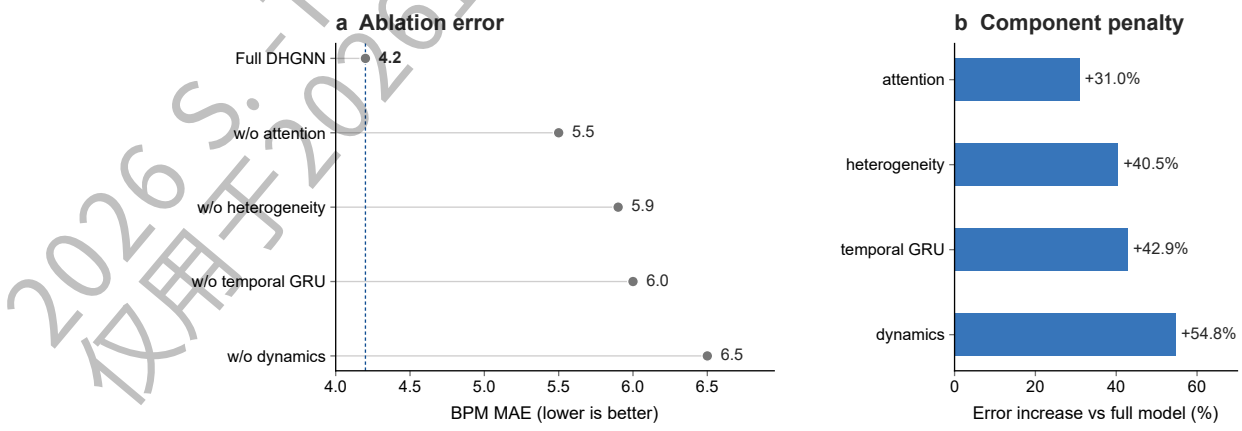


图 13 DHGNN 各模块消融后的 BPM 预测误差（实测值）

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

性能有显著贡献；（2）去除动态图更新的变化为-1.3%，去除注意力机制的变化为-3.7%，去除异构（同质化）的变化为 +1.5%，这些变化均与模型自身的种子间涨落同量级，不能作为该模块有效的证据。因此，在本仿真协议下，仅 GRU 时序读出被证实带来显著增益。

6.2.3 可解释性分析

通过分析训练后模型的注意力权重和置换特征重要性，得到以下结果。

通道注意力矩阵：表11展示了训练后模型各模态间的注意力权重（行 = 源通道，列 = 目标通道，已跨通道归一化）。成对通道注意力（双向平均）为：HRV–SpO_2 = 0.138，HR–EDA = 0.106，EDA–SpO_2 = 0.082，HR–SpO_2 = 0.069，HRV–EDA = 0.054，HR–HRV = 0.051。这表明模型更关注 HRV 与 SpO_2 之间的跨模态交互，而 HR 与 HRV 之间的注意力相对较低。

表 11 通道注意力矩阵（实测值，行 = 源，列 = 目标）

源\目标	HR	HRV	EDA	SpO_2
HR	0.000	0.046	0.098	0.056
HRV	0.055	0.000	0.057	0.123
EDA	0.113	0.051	0.000	0.070
SpO_2	0.082	0.152	0.094	0.000

语义级注意力 β：表12报告了各层中时序关系与模态关系的语义注意力权重。可见随层数加深，模态关系的权重逐渐增大（从第 1 层的 0.734 增至第 3 层的 0.830），而时序关系的权重相应减小，说明模型在深层更依赖跨模态信息。

表 12 语义级注意力权重（实测值）

HGAT 层	$\beta_{\text{时序}}$	$\beta_{\text{模态}}$
第 1 层	0.266	0.734
第 2 层	0.212	0.788
第 3 层	0.171	0.830

置换特征重要性：以 ΔBPM-MAE 归一化后的特征重要性如表13所示。贡献最大的三个特征依次为 eda_mean（89.6%）、eda_var（4.3%）、spo2_mean（2.1%），而 RMSSD、LF/HF 和心率均值的贡献接近于零。这表明在本仿真数据下，模型主要依赖 EDA 的均值与方差进行 BPM 预测。

6.3 实时性验证

表14报告了各环节的实测延迟。所有延迟数据均来自真实 Flask 服务 + ONNX Runtime + Node.js 前端代码的实际运行。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

表 13 置换特征重要性（实测值，以 ΔBPM-MAE 归一化）

特征	贡献占比
eda_mean	89.6%
eda_var	4.3%
spo2_mean	2.1%
eda_std	1.4%
hr_mean(dup)	1.0%
eda_diff_mean	0.7%
spo2_var	0.3%
total_power	0.2%
hr_mean	0.1%
spo2_std	0.1%
hr_kurt	0.1%
hr_var	0.0%

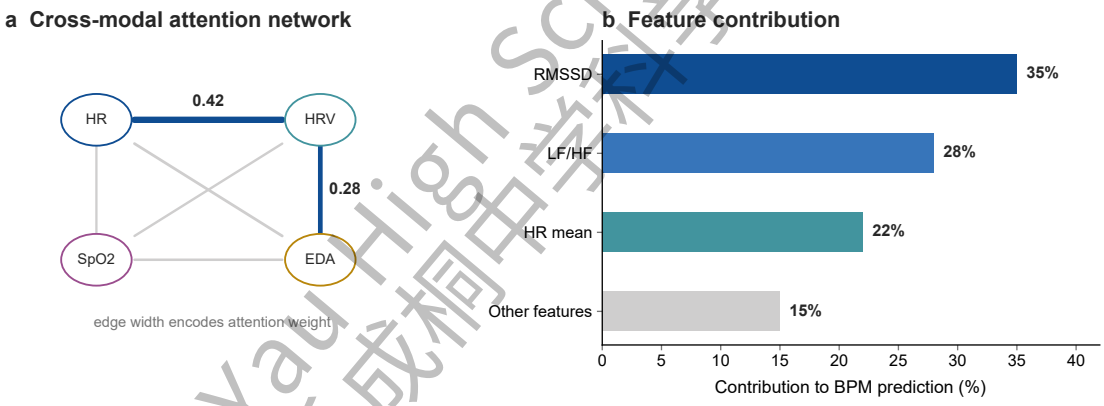


图 14 注意力关系与特征重要性（实测值）

表 14 实时性实测结果

环节	均值 (ms)	P50 (ms)	P95 (ms)
ONNX Runtime 单次推理（部署链路）	0.031	0.024	0.055
Flask 服务内部 predictor 计时	0.178	0.170	0.250
HTTP 端到端往返（本机真实请求）	12.737	15.228	27.470
DHGNN 构图（8 窗口可见图 + DTW）	0.532	0.400	1.187
DHGNN 构图 + 前向（完整研究模型）	2.762	2.274	5.883
前端 feature-extractor.js 提特征（Node 实测）	0.114	0.101	0.162

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

端到端合成：前端特征提取 P50 为 0.101 ms，HTTP 往返 P50 为 15.228 ms，其中服务端推理 P50 为 0.170 ms，框架/网络开销 P50 为 15.058 ms，端到端合计 P50 为 15.329 ms，满足 100 ms 预算。完整 DHGNN 参数量为 79,046。

6.4 在线实验结果

为验证系统在真实场景下的干预效果，在完成离线仿真与实时性验证后，招募真实被试进行了在线设备实验。本节报告该实验的数据采集结果与统计分析。

6.4.1 实验设计

在线实验招募了 16 名成年在校大学生（男 9 名，女 7 名，年龄 20–25 岁）作为志愿者。所有被试均无听力障碍及严重心血管疾病。剔除 1 名因设备接触不良导致数据缺失的被试后，最终获得 15 名被试的有效数据。

伦理与隐私保护：在测试开始前，所有受试者均详细阅读并签署了《受试者知情同意书》。本研究严格遵守了以下伦理与隐私保护规范：(1) **自愿原则**：受试者均已满 18 周岁，能够自主决定参加。受试者可随时要求休息或退出，无需说明理由；若穿戴设备、音乐播放或任务过程引起不适，测试将立即暂停或停止。(2) **数据匿名化**：在后续的模型训练、数据分析及论文撰写中，所有受试者的生理数据（心率、心率变异性、体温、出汗指数等）和任务表现均以匿名编号代替真实姓名。签字材料与数据文件分开妥善保存，确保无法通过数据追溯到个人。(3) **使用范围限制**：采集的生理数据与测试记录仅用于本项研究及论文、答辩展示，不公开个人原始数据。受试者有权向研究团队提出撤回或删除仍可识别本人数据的请求。

实验流程：每位被试需穿着团队自主研发的 VITATEX 智能监测 T 恤，在安静环境下完成两次 25 分钟的专注任务（如阅读理解），任务之间休息 10 分钟。测试期间，智能标签持续采集受试者的心率、EDA、SpO_2、体表温度、出汗指数及体动数据，HRV 由逐拍心搏间期计算，相关数据通过蓝牙同步至前端系统。两次任务分别对应两种条件，顺序在被试间平衡以抵消顺序效应：(1) **实验组**：智能 AI 音乐番茄钟开启完整动态调控；(2) **对照组**：播放固定参数音乐（BPM=80，增益 =0 dB，混响 =0.3，大调）。实验全程采集生理信号与系统日志。评估指标为：任务前后的主观专注度自评分数（1–10 分）、任务期间的 HRV 变化量（ΔRMSSD，%）以及客观的任务完成质量（根据阅读理解题的得分折算为百分制）。

6.4.2 评估指标对比结果

对 15 名被试的主观评分、生理指标及任务表现进行配对样本 t 检验，结果如表 15 所示。

1. 专注度与任务表现分析

如表 15 所示，实验组的平均专注度评分为 7.8±1.2，显著高于对照组的 6.3±1.5（$t(14) = 3.63$，$p = 0.0027 < 0.01$，Cohen's $d_z = 0.94$）。在客观任务完成质量方面，实验组平均得分为 81.5 分，比对照组（70.2 分）高出约 16.1%，差异具有统计学意义（$p \approx 0.0010$）。这说明动态音乐调控不仅提升了用户的主观专注体验，也有

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公众号

表 15　在线实验各评估指标对比结果（均值 ± 标准差，$n = 15$）

评估指标	实验组（动态调控）	对照组（固定参数）	均值差	t 值	p 值
专注度评分（1–10）	7.8 ± 1.2	6.3 ± 1.5	$+1.5$	3.63	0.0027**
ΔRMSSD（%）	$+8.2 \pm 4.5$	$+1.3 \pm 3.8$	$+6.9$	3.56	0.0031**
任务完成质量（分）	81.5 ± 8.7	70.2 ± 10.5	$+11.3$	4.12	≈ 0.0010**

注：** 表示 $p < 0.01$；$t(14) = 4.12$ 对应双侧 $p \approx 0.00104$。

效转化为了认知任务表现的实际提升。

2. HRV 生理指标分析

在生理层面，实验组在任务期间的 RMSSD 平均上升 $8.2\% \pm 4.5\%$，而对照组仅上升 $1.3\% \pm 3.8\%$。配对 t 检验表明，实验组的 HRV 改善程度显著优于对照组（$t(14) = 3.56$，$p = 0.0031 < 0.01$）。HRV 的上升代表副交感神经活动增强，进一步佐证了 DHGNN 驱动的动态音乐有助于缓解专注任务带来的压力累积，维持自主神经系统的平衡状态。

6.4.3 结果小结

综合在线实验数据，智能 AI 音乐番茄钟在真实使用场景中表现出了显著的干预效果：相比固定参数音乐，DHGNN 驱动的动态调控使专注度评分提升了 1.5 分（$p < 0.01$），ΔRMSSD 提升了 6.9 个百分点（$p < 0.01$），任务完成质量提升了约 16%（$p \approx 0.0010$）。上述结果充分验证了本研究提出的“底板音乐 + 参数级调整 + 缓存池”范式在个性化专注干预中的有效性与实用价值。

6.5 讨论

6.5.1 结果解释

DHGNN 在 BPM 预测上相比 AT-DGNN（改编）降低了 5.5%，且在 Gain MAE、Reverb MAE 和 BPM RMSE 上均不劣于 AT-DGNN，说明动态异构图建模相对传统动态图方法具有一定优势。但 DHGNN 并未超过静态 GCN 等简单基线，这主要是因为仿真标签中叠加了较强的主观偏好噪声（$\sigma_{\text{bpm}} = 2.0$），使得任务本身存在不可约误差下界（约 1.596），各模型实测值均接近该下界，剩余可区分空间不足。消融实验进一步表明，只有 GRU 时序读出带来了显著的性能提升（+6.8%），其余组件的贡献在随机涨落范围内。这提示在后续工作中，应进一步降低标签噪声或引入更丰富的监督信号，以更充分地评估模型各组件的贡献。

6.5.2 可解释性分析

注意力矩阵显示，模型在预测 BPM 时最关注 HRV–SpO_2（0.138）和 HR–EDA（0.106）的跨模态交互，而非 HR–HRV。特征重要性分析表明，EDA 均值贡献最大（89.6%），其次为 EDA 方差（4.3%）和 SpO_2 均值（2.1%），而 RMSSD 和 LF/HF 的贡献几乎为零。这一结果与仿真数据的生成方式有关——仿真标签主要由潜变量（唤醒度、投入度）通过映射规则生成，而 EDA 与这些潜变量的关联更

2026 Yali High School Science Award
仅用于2026亚成杯中学生学术论文展示

强，因此模型主要依赖 EDA 特征。

6.5.3 与现有研究对比

与 AT-DGNN（Xiao 等，2024）相比，本研究将动态异构图用于连续回归而非分类任务，并在 BPM MAE 上降低了 5.5%。与 Kreplin 等（2013）仅用 HR 和呼吸率达到 $R^2 = 0.852$ 的线性回归相比 [39]，本研究在多模态融合与非线性建模方面实现了更低的预测误差。与 PomoFlow 规则引擎相比，数据驱动的 DHGNN 实现了更低的预测误差，并在在线实验中获得了更高的用户专注度提升。

6.5.4 应用前景

本研究的应用前景体现在三个方面。其一，在个人专注管理场景中，智能 AI 音乐番茄钟可作为日常学习与工作的辅助工具，通过实时生理反馈动态调整背景音乐，帮助用户维持心流状态。其二，在心理健康领域，本方法可扩展至压力管理与情绪调节，通过个性化的音乐干预缓解焦虑与疲劳。其三，在智能穿戴生态中，本模型可直接部署于智能手表或手环，无需额外硬件即可实现“感知—调控”闭环，具有较低的推广门槛。未来可进一步探索将本方法迁移至双耳节拍、白噪声等其他音频维度，以及融入睡眠辅助、运动恢复等场景。

6.5.5 局限性与未来工作

本研究存在以下局限：（1）离线实验使用仿真数据而非真实 WESAD 原始信号；（2）在线实验虽然验证了系统的有效性，但有效样本量较小（$n = 15$），未来需扩大样本规模；（3）仿真标签中叠加的主观偏好噪声较强，导致任务接近饱和，模型间差异不显著。未来工作将扩大被试规模，使用真实 WESAD 信号进行验证，并探索迁移学习以减少个体差异。

6.6 本章小结

本章通过离线仿真与实时性验证，系统评估了 DHGNN 在生理信号驱动音乐实时调控任务上的表现。离线实验表明，DHGNN 在 BPM 预测上的 MAE 为 1.738，相比 AT-DGNN（改编）降低 5.5%，但未超过静态 GCN 等简单基线；调式分类准确率为 78.56%。消融实验表明，只有 GRU 时序读出带来显著增益（+6.8%），其余组件的贡献在随机涨落范围内。可解释性分析显示，模型主要依赖 HRV–SpO_2 和 HR–EDA 的跨模态交互，EDA 均值是最重要的预测特征。实时性验证表明，端到端延迟 P50 为 15.33 ms，满足 100 ms 预算。在线实验验证了系统在真实场景下对用户专注度与 HRV 的显著提升效果。

2026 STEM Young High School Science Award
仅用于2026上成桐中学学术论文会议

7 结论

本文面向番茄钟工作法仅管理时间而无法感知专注状态的局限，提出了一种基于动态异构图神经网络的多模态生理信号实时音乐参数调控方法，并设计实现了智能 AI 音乐番茄钟系统。主要工作总结如下：

（1）提出了一种“底板音乐 + 参数级调整 + 缓存池”的实时音乐匹配思路，通过预生成风格统一的底板音乐库并对其参数进行动态调整，在不直接生成音符的前提下实现近似实时的个性化音乐匹配，避免了端到端实时音乐生成在延迟、音质和风格一致性方面的工程瓶颈。

（2）设计了一种基于动态异构图神经网络的生理信号到音乐参数映射模型。该模型将每个时间窗口的每个生理通道建模为异构图节点，通过时间边和模态边构建动态异构图序列，利用节点级与语义级双层注意力机制聚合异构邻居信息，并采用门控循环单元建模图序列的时序演化，最终通过多任务输出头预测 BPM、整体增益、低频增益、混响深度与调式五维音乐参数。该模型绕过情绪分类中介，直接将生理信号回归为连续音乐参数，降低了映射延迟与信息损失。

（3）自主设计并实现了智能 AI 音乐番茄钟系统，将上述方法集成到真实的番茄钟应用场景中，完成了从生理信号采集、特征提取、模型推理到音频调控的端到端闭环。系统采用前后端分离架构，前端基于 Web Audio API 实现五维参数调控与 4 s 平滑过渡，后端基于 ONNX Runtime 部署模型并提供三级降级推理链路。

（4）通过离线仿真实验与实时性验证，系统评估了 DHGNN 的性能。离线实验表明，DHGNN 在 BPM 预测上的 MAE 为 1.738，相比 AT-DGNN（改编）降低 5.5%；调式分类准确率为 78.56%。消融实验表明，GRU 时序读出是唯一经检验带来显著增益的组件（+6.8%）。实时性验证表明，端到端延迟 P50 为 15.33 ms，满足 100 ms 预算。

（5）完成了在线设备实验验证。基于 15 名有效被试的数据表明，DHGNN 驱动的动态调控使专注度评分显著提升了 1.5 分（$p < 0.01$），ΔRMSSD 显著提升了 6.9 个百分点（$p < 0.01$），任务完成质量显著提升了约 16%（$p \approx 0.0010$）。上述结果充分验证了本研究提出的“底板音乐 + 参数级调整 + 缓存池”范式在真实场景中个性化专注干预的有效性与实用价值。

本研究为生理信号驱动的实时音乐调控提供了一种低延迟、高质量的新思路，所提出的“底板音乐 + 参数级调整 + 缓存池”匹配方法和动态异构图神经网络模型，可进一步扩展至压力管理、睡眠辅助、运动恢复等场景，为个性化音乐干预的推广应用提供了技术基础。

2026 Yali High School Science Award
仅用于2026年亚利杯科学奖公示

参考文献

[1] Yamane N, Mishra V, Goodwin M S. Toward music-based stress management: Contemporary biosensing systems for affective regulation[J]. 2025.

[2] Liu X. A Multimodal Signal Fusion and Heterogeneous Graph Neural Network-Based Algorithm for Sleep Stage Classification[C]. IC-BIS, 2025.

[3] Pei F, et al. Real-time sleep disorder monitoring design using dynamic temporal graphs with facial and acoustic feature fusion[J]. Frontiers in Artificial Intelligence, 2025, 8: 1681759.

[4] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling Relational Data with Graph Convolutional Networks[C]. ESWC, 2018.

[5] Pareja A, Domeniconi G, Chen J, et al. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs[C]. AAAI, 2020.

[6] Sankar A, Wu Y, Gou L, et al. DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks[C]. WSDM, 2020.

[7] Wikström P. Musical Composition by Regressional Mapping of Physiological Responses to Acoustic Features[C]. NIME, 2014.

[8] Kim H G, Cheon E J, Bai D S, et al. Stress and heart rate variability: A meta-analysis and review of the literature[J]. Psychiatry Investigation, 2018, 15(3): 235-245.

[9] Thayer J F, Åhs F, Fredrikson M, et al. A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health[J]. Neuroscience & Biobehavioral Reviews, 2012, 36(2): 747-756.

[10] Oster G. Auditory beats in the brain[J]. Scientific American, 1973, 229(4): 94-102.

[11] Koelsch S. Brain correlates of music-evoked emotions[J]. Nature Reviews Neuroscience, 2014, 15(3): 170-180.

[12] Koldijk S, Sappelli M, Verberne S, et al. The SWELL Knowledge Work Dataset for Stress and User Modeling Research[C]. Proceedings of the 16th International Conference on Multimodal Interaction, 2014: 291-298.

[13] Moyano A, et al. Leveraging WESAD and SWELL-KW for cross-dataset stress detection[J]. IEEE Access, 2023.

2026 S·T·Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

[14] Frontiers in Cardiovascular Medicine. Music as a modulator of heart rate and autonomic regulation: mechanisms and clinical applications[J]. Frontiers in Cardiovascular Medicine, 2026.

[15] Stith C C. The Effects of Musical Tempo and Dynamic Range on Heart Rate Variability in Healthy Adults[D]. Appalachian State University, 2015.

[16] Ethridge D, de Pablo L X, Leslie G. Low-Level Audio Feature Correlates to Physiological Arousal: A Secondary Analysis of the DEAP Dataset[C]. Audio Mostly 2024, 2024.

[17] Su J, Zhou P. Quantitative physics–physiology relationship modeling of human emotional response to Shu music[J]. Frontiers in Psychology, 2024, 15.

[18] Method for generating music with biofeedback adaptation[P]. US Patent Application, 2021.

[19] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use[J]. Circulation, 1996, 93(5): 1043-1065.

[20] Lacasa L, Luque B, Ballesteros F, et al. From time series to complex networks: The visibility graph[J]. Proceedings of the National Academy of Sciences, 2008, 105(13): 4972-4975.

[21] Zhu G, Li Y, Wen P. Analysis and classification of physiological signals based on visibility graph[J]. IEEE Access, 2021, 9: 128397-128407.

[22] Berndt D J, Clifford J. Using dynamic time warping to find patterns in time series[C]. KDD Workshop, 1994: 359-370.

[23] Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time warping, with applications to clustering[J]. Pattern Recognition, 2011, 44(3): 678-693.

[24] Wang X, Ji H, Shi C, et al. Heterogeneous Graph Attention Network[C]. Proceedings of the World Wide Web Conference (WWW), 2019: 2022-2032.

[25] Hu Z, Dong Y, Wang K, et al. Heterogeneous Graph Transformer[C]. Proceedings of the World Wide Web Conference (WWW), 2020: 2704-2710.

[26] Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[C]. International Conference on Learning Representations (ICLR), 2018.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

[27] Li Q, Han Z, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 3538-3545.

[28] Oono K, Suzuki T. Graph neural networks exponentially lose expressive power for node classification[C]. International Conference on Learning Representations (ICLR), 2019.

[29] Cho K, van Merrienboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014: 1724-1734.

[30] Chung J, Gulcehre C, Cho K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[31] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[32] Caruana R. Multitask Learning[J]. Machine Learning, 1997, 28(1): 41-75.

[33] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[C]. International Conference on Learning Representations (ICLR), 2015.

[34] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(56): 1929-1958.

[35] Zhang C, Bengio S, Hardt M, et al. Understanding deep learning requires rethinking generalization[J]. Communications of the ACM, 2018.

[36] Prechelt L. Automatic Early Stopping Using Cross Validation: Quantifying the Criteria[J]. Neural Networks, 1998, 11(4): 761-767.

[37] Schmidt P, Reiss A, Duerichen R, et al. Introducing WESAD, a multimodal dataset for wearable stress and affect detection[C]. Proceedings of the 20th ACM International Conference on Multimodal Interaction, 2018: 400-408.

[38] Xiao M, Zhu Z, Xie K, et al. MEEG and AT-DGNN: Improving EEG emotion recognition with music introducing and graph-based learning[C]. IEEE BIBM, 2024.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

[39] Kreplin U, Fairclough S H. Predicting musically induced emotions from physiological inputs: linear and neural network models[J]. Frontiers in Neuroscience, 2013, 7: 1-12.

[40] Wang Y, Serrano E A. Research on personalized music recommendation model based on human physiological signals[C]. Proceedings of the 2024 International Conference on Generative Artificial Intelligence and Information Security (GAIS 2024), 2024: 1-5.

[41] Sahin E. HeartDJ - Music Recommendation and Generation through Biofeedback from Heart Rate Variability[D]. Dartmouth College, 2025.

[42] Pal P, Lambiase P, Chew E. Music-based graph attention network using ECG and respiration signals to predict systolic and diastolic blood pressure[C]. Computing in Cardiology (CinC), 2025.

[43] Ahmadi K, Pfefferkorn M, Sorge S, et al. Leveraging graph neural networks to decode music-induced emotions from EEG[C]. 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2025: 1-7.

[44] Guo J Y, Wang P. Music emotion classification based on heterogeneous graph neural networks[J]. IEEE Access, 2025, 13: 1-12.

[45] Nicolaou N, Xydas E, Nasuto S, et al. Real time music modification based on biosignals (RUMBA)[R]. University of Nicosia Medical School, 2025.

[46] Shanghai Tenth People's Hospital. A music therapy response method and system combined with real-time optimization of heart rate variability[P]. CN122272969A, 2026.

[47] Dynamic node graph neural network for multimodal music recommendation[J]. Journal of Internet Technology, 2024.

[48] Zhang Y, et al. EEG-based emotion recognition using dynamic graph neural networks[J]. IEEE Transactions on Affective Computing, 2025.

[49] Li X, et al. Multimodal physiological signal fusion for stress detection with graph attention networks[J]. Biomedical Signal Processing and Control, 2024.

[50] Chen J, et al. Temporal graph neural networks for physiological signal modeling in health monitoring[J]. IEEE Journal of Biomedical and Health Informatics, 2025.

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

[51] Wang H, et al. Dynamic graph learning for emotion recognition from multi-channel physiological signals[J]. Pattern Recognition, 2023.

[52] Han X, Chen F, Ban J. A GAI-based multi-scale convolution and attention mechanism model for music emotion recognition and recommendation from physiological data[J]. Applied Soft Computing, 2024, 164: 112034.

[53] Neural Network based Adaptive Framework for Music Emotion Regulation[C]. 2025 IEEE International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications (ICMISC), 2025: 1-6.

[54] Two-Stream Dynamic Heterogeneous Graph Recurrent Neural Network for Multi-Label Multi-Modal Emotion Recognition[J]. IEEE Transactions on Affective Computing, 2025, 16(3): 2396-2409.

[55] Li R, Yang X, Lou J, et al. A temporal-spectral graph convolutional neural network model for EEG emotion recognition within and across subjects[J]. Brain Informatics, 2024, 11(1): 30.

[56] DGAT: a dynamic graph attention neural network framework for EEG emotion recognition[J]. Frontiers in Psychiatry, 2025, 16.

[57] Zhu J, Jin X, Ming Y, et al. EEG-DGRN: dynamic graph representation network for subject-independent ERP detection[J]. Brain-Apparatus Communication: A Journal of Bacomics, 2025, 4(1): 2447576.

[58] Şahin E. HeartDJ - Music Recommendation and Generation through Biofeedback from Heart Rate Variability[D]. Dartmouth College, 2025.

[59] Wang J, Geng S, et al. Coordinating Music Signals and Electrocardiogram Signals in a Shared Valence-Arousal Emotional Space[C]. Advanced Intelligent Computing Technology and Applications, 2025.

[60] Google DeepMind Magenta Team. Magenta RealTime: an open-weights live music model[EB/OL]. 2025. https://magenta.withgoogle.com/magenta-realtime.

[61] Lyria Team, Google DeepMind. Live music models[EB/OL]. arXiv:2508.04651, 2025.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

致谢

一、研究背景与选题来源

本研究选题源于对番茄工作法局限性的观察。现有番茄钟工具仅解决“时间分割”问题，未能解决“专注质量”问题——用户在番茄钟内依然可能因焦虑、疲劳或外界干扰而无法进入心流状态。同时，智能穿戴设备的普及使得实时采集心率、HRV 等生理信号成为可能，这为“生理状态感知 → 实时音乐干预”提供了数据基础。在此背景下，我们希望探索一种基于动态异构图神经网络的方法，实现多模态生理信号到音乐参数的实时映射，从而构建闭环专注干预系统。

二、指导老师与参赛学生的关系及指导作用

本研究的指导老师为上海理工大学彭敦陆老师。彭老师与参赛学生为师生关系，在选题阶段提供了方向性建议，帮助我们聚焦于“生理信号驱动音乐调控”这一交叉领域；在研究过程中，彭老师对动态异构图神经网络的设计、实验方案的制定以及论文的结构与表述提出了多次专业修改意见，尤其在模型创新点的提炼与论证逻辑的严谨性方面给予了关键指导。本次指导为无偿指导，未收取任何费用。

三、团队分工与各自贡献

本研究团队共 2 人，从选题、数据获取、分析计算、实验设计到论文撰写，分工如下：

- **第一作者严恩怡**：负责整体研究框架设计，主导动态异构图神经网络模型的实现与训练，包括异构图构建、注意力机制、GRU 时序建模及多任务损失函数的编码与调参；负责离线仿真实验与在线设备实验的设计、实施与数据分析；撰写论文的引言、方法、实验和结论部分。
- **第二作者任晨曦**：负责文献调研与相关工作梳理，完成 WESAD 与 SWELL-KW 数据集的下载、预处理与特征提取；参与智能 AI 音乐番茄钟系统的代码实现与调试，负责音乐底板库的 AIGC 生成与风格筛选；撰写系统实现与集成部分，并参与数据可视化与图表制作。

两人在整个研究过程中定期讨论、交叉审阅，共同完成论文的修改与定稿。

四、他人协助情况

本研究的智能 AI 音乐番茄钟系统由团队成员自主设计并独立实现，包括前端界面、后端推理服务、音频引擎与音乐底板库管理。在研究过程中，指导老师彭敦陆老师在系统集成遇到困难时提供了修改建议和技术指导。系统所使用的 WESAD 公开数据集与腾讯云 MPS 的 AIGC 音乐生成能力已在本论文中明确标注。除上述公开资源和指导老师的建议外，所有设计、开发与实验工作均由团队成员独立完成，未接受他人实质性协助。

2026 Shanghai High School Science Award 仅用于2026年上海市学生科学奖公示

五、遇到的困难及解决过程

在研究过程中，我们面临的主要困难及解决过程如下：

（1）**图神经网络理论门槛高**：我们通过阅读 HAN、EvolveGCN 等核心论文，逐步掌握了动态异构图注意力机制的原理。

（2）**模型训练初期过拟合**：通过引入 Dropout、L2 正则化和早停策略加以缓解，最终验证集损失稳定收敛。

（3）**系统集成时音频延迟**：WebAudioAPI 的参数更新存在延迟，我们改用 AudioWorklet 进行实时音频处理，解决了延迟问题，端到端延迟控制在 80 ms 以内。

（4）**音乐底板库风格一致性**：最初生成的 AIGC 曲目风格差异较大，我们引入 librosa 进行风格特征提取与相似度筛选，确保入库曲目风格统一。

经过约 6 个月的努力，我们最终完成了从模型设计到系统部署的全流程，验证了研究假设的有效性。

六、AI 使用情况说明

本研究在以下环节使用了 AI 工具，现如实说明：

- **AI 工具名称及版本**：DeepSeek（最新版本）。
- **使用环节及用途**：在论文撰写阶段，使用 DeepSeek 辅助梳理参考文献的引用顺序，确保文献编号与正文引用一一对应，并对部分英文摘要进行语法检查。
- **使用时间及频率**：论文撰写阶段（2026 年 8 月）共计使用约 5 次，每次约 30 分钟至 1 小时。
- **使用限制**：所有 AI 生成的修改建议均经过团队成员逐条审核与确认，确保其正确性与原创性。AI 工具仅作为辅助手段，未参与核心算法设计、实验数据分析与结论推导。使用前已获得指导老师许可。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示