arXiv:2304.01293v1 [cs.CY] 3 Apr 2023

基于可穿戴传感器的社交焦虑个体在不同社交情境中的多模态生理反应

预印本

Emma R.Toner*, MarkRucker*,Zhiyuan Wang, Maria A. Larrazabal, Lihua Cai, Debajyoti Datta, Elizabeth Thompson, Haroon Lone, Mehdi Boukhechva, Bethany A. Teachman 和 Laura E. Barnes

摘要

利用被动佩戴的传感器准确识别个体所处的社交情境，有望为社交焦虑障碍提供即时自适应干预（JITAIs）。本研究介绍了一项被试内实验的结果，该实验通过被动采集的数据评估了不同社交情境（即独处与和他人在一起）、社交阶段（即互动前后与互动期间）、社交互动规模（即双人互动与群体互动）以及社交威胁程度（即隐式与显式社会评价）下的生理反应。研究参与者（$N = 46$）在社交互动焦虑量表（Social Interaction Anxiety Scale）上的得分为（$\geq$34 / 80），表明其社交焦虑症状达到中度至重度。研究采用单变量配对差异检验、多变量随机森林模型以及后续聚类分析，探究不同社交与非社交情境下的生理反应模式。结果表明，相较于社交阶段、群体规模或社交威胁程度，社交情境更容易被可靠地区分；但即使在这些可区分的情境中，生理反应模式仍存在显著差异。本文讨论了这些发现对现实世界情境检测及即时自适应干预部署的启示。

1 引言

社交焦虑障碍（SAD）是美国最常见的精神健康障碍之一：据估计，13% 的成年人一生中会被诊断为 SAD [1]。社交焦虑障碍的特征包括对社交情境感到强烈焦虑、害怕受到负面评价，以及回避社交情境 [2], 。SAD 与工作及社交领域的显著功能损害相关 [3]。尽管 SAD 带来了沉重负担，大多数确诊者并未寻求治疗或无法获得治疗 [4] ，许多人甚至要等待数十年才能接受治疗 [5]。既然已有针对 SAD 的有效心理治疗 [6], , 为什么仍有这么多人无法获得治疗？一个尤其有说服力的解释是，主流心理治疗服务模式（即由受过训练的心理健康专业人员面对面提供的一对一治疗）难以获得，无法满足巨大的需求 [7]。对社交焦虑个体而言，寻求治疗还常常受到后勤、经济和情感方面的障碍（例如羞耻感、与服务提供者交谈时的焦虑）影响 [8]。

数字心理健康干预（DMHIs）通常易于获取、成本低且具有隐私性，因此为标准的一对一治疗提供模式提供了潜在替代方案。尤其是即时自适应干预（JITAIs； [9]），旨在通过在最需要的时刻提供与情境相关的干预，使数字心理健康干预更具个性化并得到优化。

*Emma R. Toner 和 Mark Rucker 对本文的贡献相同，并列为共同第一作者。本出版物的研究部分得到了 3Cavaliers 种子基金以及美国国立卫生研究院下属美国国家精神卫生研究所的资助，资助编号为 R01MH132138。本文内容完全由作者负责，不一定代表美国国立卫生研究院的官方观点。Emma R. Toner、Maria A. Larrazabal 和 Bethany A. Teachman 任职于弗吉尼亚州夏洛茨维尔的弗吉尼亚大学心理学系。Mark Rucker、Zhiyuan Wang、Debajyoti Datta、Elizabeth Thompson 和 Laura Barnes 任职于弗吉尼亚州夏洛茨维尔的弗吉尼亚大学系统与信息工程系。Lihua Cai 任职于中国广州的华南师范大学阿伯丁数据科学与人工智能研究院。Haroon Lone 任职于印度博帕尔的印度博帕尔科学教育与研究学院电气工程与计算机科学系。Mehdi Boukhechba 任职于美国新泽西州蒂特斯维尔的强生旗下杨森制药公司，邮编 08560。

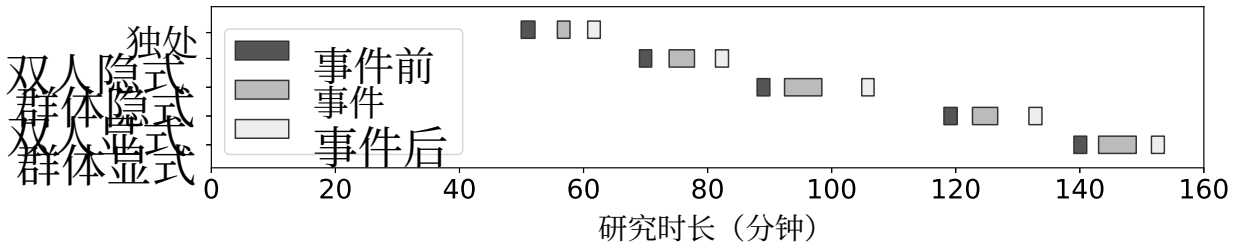


图1：我们展示了某位参与者的一次完整实验环节。每次实验中，事件的顺序（纵轴）均为随机安排，只有独处环节始终排在最前。最初45分钟的延迟是因研究准备工作所需。

这或许特别有助于弥合社交焦虑治疗缺口。为了最有效地实施即时自适应干预，有必要确定人们何时最需要干预，以及根据其当前情境（例如所处位置），哪种或哪些干预最有用、最合适。一种收集这些信息的方法是请参与者报告其日常生活中的情绪体验和情境背景（例如通过生态瞬时评估），但这可能会增加负担 [10]，并要求参与者频繁且准确地回应提示。此外，当参与者正处于关注的情境中（例如社交互动）时，收集自我报告数据可能颇具挑战，因为回应提示意味着必须中断他们正在做的事情。直接询问参与者体验的潜在替代方法，是通过非侵入式设备和传感器（例如腕带、手机）被动收集数据（例如生理反应、GPS位置），以帮助识别其当前的情绪状态和/或情境背景。

识别社交情境对于社交焦虑干预至关重要，因为与其他情境相比，社交焦虑个体更可能在某些情境中感到焦虑（例如，社交与非社交情境，或需要接受评价与无需接受评价的情境； [11]）。此外，根据情境不同，某些干预策略在实施上的可行性也有所不同（例如，如果一个人正积极参与社交互动，他们可能无法详细写下并评估自己的所有想法）。因此，了解某人身处何处或正在做什么，有助于判断何时采用干预措施，以及选择可能适当且有效的干预措施。

移动技术和被动感知技术的进步，使研究人员能够在实验研究和日常生活中，基于生物行为数据检测社交焦虑和社交情境。例如，可穿戴设备采集的生理数据（如皮肤温度、皮肤电导）已被用于准确检测演讲任务情境下的社交焦虑严重程度 [12],，而手机采集的其他被动感知数据（如全球定位系统（GPS）、通话与短信数据）则已被用于检测日常生活中的社交焦虑症状 [13,14]。此外，心率、皮肤电导、皮肤温度、基频和手部动作等生理生物标志物，可用于准确区分社交焦虑的基线、预期和同期阶段 [12,15]，以及演讲任务中不同程度的评价威胁 [16]。在日常生活研究中，音频数据已被用于识别参与者是否处于人类语音附近（如对话； [17,18]），而位置数据则有助于检测人们何时在群组中互动，并区分群体规模 [19]。

综上，这些发现表明，被动传感数据是一种重要工具，可用于判断个体何时可能受益于靶向干预。然而，仍有许多问题尚待解答。迄今为止，大多数检测社交焦虑或社交情境的研究都强调结合多种生物行为特征，以最大限度地提高预测效果，但这会降低此类方法的可扩展性（例如，如果要求参与者佩戴多种设备，以收集准确预测所需的全部数据），也会使人难以确定哪些特征最为必要（因此在规划数据收集时应优先考虑），以及哪些特征可能是冗余的。
本研究利用了一项实验研究的数据。该实验通过 Zoom 在线开展，高特质性社交焦虑个体完成了一项非社交任务，以及四种社交互动；这些互动涉及不同数量的互动对象，并设置了经实验操纵的不同程度的社交评价威胁。研究通过 Empatica E4 腕带被动评估焦虑的生理指标，并依据既有文献提取了13项与焦虑和压力检测相关的特征（见表3）。这项探索性研究旨在从三个主要方面加深我们对社交焦虑个体情境特异性生理反应的理解：

1. 描述参与者在社交与非社交情境及各阶段中的生理反应。2. 确定社交与非社交情境及各阶段中存在显著差异的生理反应。3. 通过聚类分析探讨能否用少数几种反应来表征生理反应，或者这些反应在社交与非社交情境及各阶段中是否相对异质。

表 1：本研究使用的四种传感器。

Name	Rate	描述	位置	分辨率
PPG	64Hz	光体积描记法	手腕外侧	0.9 纳瓦
ACC	32Hz	三轴加速度计	手腕外侧	0.016 克
EDA	4Hz	皮肤电活动	手腕内侧	∼900 pS
TMP	4Hz	皮肤温度	手腕外侧	0.02 °C

2 方法

2.1 参与者

参与者为美国东南部一所大型公立大学的本科生。$N = 55$ 参与者同意参加研究以获得心理学课程学分。所有参与者均完成了社交互动焦虑量表（SIAS； [20]）；只有得分达到34分或以上、表明存在中度至重度社交焦虑症状的参与者才符合参加资格。由于SIAS得分较低，9名参与者被排除，最终样本为 $N = 46$。入组前SIAS平均得分为45.57（标准差 = 8.87），得分范围为34至69分（满分为80分）。此外，若参与者表示经常使用某些药物（苯二氮䓬类、兴奋剂、抗精神病药、情绪稳定剂、β受体阻滞剂、单胺氧化酶抑制剂、可穿过血脑屏障的药物或止痛药），或被诊断患有心血管疾病，则不符合参与资格，因为这些因素均可能影响心理生理反应性。因此，研究还要求参与者在研究时段开始前至少两小时避免使用上述药物以及咖啡因、尼古丁和大麻，并避免进行剧烈运动；同时，在研究时段开始前至少24小时避免饮酒及使用非法药物。

2.2 程序

所有研究程序均经大学机构审查委员会批准（UVA-SBS IRB #3004）。参与者以四至六人为一组，参加为期三小时的研究环节。由于 COVID-19 疫情，研究通过 Zoom 在线进行。研究环节由接受过培训的心理学本科生和/或研究生主持。在研究环节中，参与者最多完成四项社交任务和一项非社交任务。首先完成的任务始终是非社交基线：参与者独自观看三段中性视频之一，视频内容为彩色移动图形，时长两分钟（这些中性视频的效价依据先前的 mTurk 样本确定； [21]）。随后，参与者以随机顺序完成两次双人对话（每次四分钟）和两次小组对话（每次六分钟）。每次对话中，参与者会从四个预选话题中选取一个（例如：“如果你赢得一百万美元，你会如何使用这笔钱？为什么？”）。在两次双人对话和两次小组对话中，各有一次被指定为显式 社交评价情境：参与者在对话开始前会被告知，对话伙伴将在交谈后根据他们的讨人喜欢程度和对话技巧进行评价。另外两次对话不包含这一指示，因此属于隐式 社交评价情境；之所以如此认定，是因为社交焦虑个体可能会认为，任何陌生的社交互动都存在潜在的社交评价威胁，即使他们并未被明确告知这种威胁 [11]。全部五项任务也都分为三个独立的时间阶段：（1）预期阶段或事件前 阶段：向参与者介绍即将进行的任务，并要求他们用两分钟生动地想象自己完成该任务（例如，想象自己正在与一位陌生人交谈，而对方随后会对自己进行评价）；（2）同时进行 阶段：参与者完成任务；（3）事件后 阶段：要求参与者用两分钟生动地回想刚完成的任务，并思考自己觉得任务完成得如何。在整个三小时研究环节中，研究人员通过佩戴在参与者左腕上的 Empatica E4 腕带（https://www.empatica.com/research/e4/）持续监测其生理反应。研究程序示意图见图1。

2.3 数据

本研究聚焦于从 Empatica E4 的四种传感器采集的数据：光电容积脉搏波（PPG）、加速度计、皮肤电活动（EDA）和皮肤温度。这四种传感器是整个研究数据集的一部分；该数据集还包含来自其他传感器和设备的数据。选择这四种传感器，是因为它们与被动式焦虑检测领域的现有研究相符（ [22–29]）。本研究所用四种传感器的详细技术说明见表1。

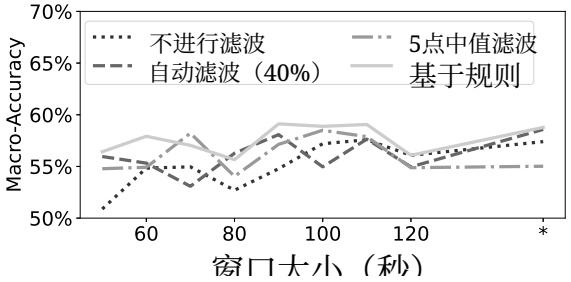


图2：就第3节所述的五项任务的随机森林模型性能而言，并不存在“最佳”的神经网络间期滤波器。此处分析仅纳入神经网络特征，并采用 NLOPOCV 流程评估模型性能。星号表示使用全部数据（3分钟）。自动滤波器见 [32]。规则滤波器会剔除超出阈值范围的间期。

2.4 数据清理

筛选 我们首先剔除了在实验阶段及体验之外采集的所有传感器读数，以此开始清理数据。这样一来，每位参与者留下了 9 至 15 段不同的传感器读数（具体数量取决于参与者完成的任务数），各段之间间隔数分钟（见图1）。

光电容积脉搏波 我们对原始光电容积脉搏波读数应用了一个 3^{rd} 阶巴特沃斯滤波器，低截止频率为 0.5Hz，高截止频率为 8Hz。这去除了正常心率频率范围之外的伪迹。随后，我们使用 [30] 中的峰值检测算法提取收缩峰，并计算正常间期（NN间期）。滤波器和峰值检测算法均通过 Neurokit2 [31]应用。

NN间期 我们评估了多种方法，以去除原始 NN间期中的噪声。我们采用嵌套留一参与者交叉验证（NLOPOCV），训练随机森林（RF）模型，根据 NN特征预测第 3 节中描述的各项任务（参见表3）。如果某种 NN间期清理方法所对应的下游随机森林模型在所有预测任务上的平均宏平均准确率（宏平均准确率）更高，我们就认为该方法更优。我们没有发现一种普遍“最佳”的方法（见图2）。对于较短的时间窗口，中值滤波器似乎效果最好；而较长的时间窗口则似乎更适合自动滤波器 [32]（采用 40% 的截止规则）。

皮肤电活动 我们对原始皮肤电活动信号应用了一个低通 4^{th} 阶巴特沃斯滤波器，高截止频率为 3Hz。与光电容积脉搏波的处理相同，我们使用 Neurokit2 [31] 对皮肤电活动信号进行了清理。

加速度计和皮肤温度 未对加速度计和皮肤温度读数进行滤波或清理。假设这些传感器上的噪声是加性的且均值为零，这对于从这些传感器导出的特征（即均值和标准差）是适当的。

2.5 特征

为回答我们的问题，我们从文献中选取了十三个特征；研究表明，这些特征能够准确预测焦虑和/或压力，且可根据我们的传感器数据流计算得出。我们使用清理后的数据，针对每位参与者的每个事件阶段间隔计算这些特征。计算后得到一个数据集，其中包含 594 个样本，每个样本包含 13 个特征。特征完整列表见表3。

其中十个特征属于时域特征，可直接根据清理后的数据计算得出。另有两个特征是频域统计量，需要将时域数据转换为频域数据。为此，我们对时域样本应用 Lomb-Scargle方法，使用在 0.01 至 0.5Hz 之间均匀间隔的基函数。

2.6 特征相关性

图3展示了13个特征的相关矩阵。大多数特征之间的相关性较弱，这表明每个特征都提供了不同的信息。相关性最高的特征均源自光电容积脉搏波传感器。

2.7 特征条件化

在进行分析之前，我们考虑了是否要在参与者层面对数据进行中心化或缩放。中心化时，逐项特征减去该参与者相应特征的中位数。缩放时，将参与者的各项特征除以该参与者的特征四分位距。不同条件下，第 3 节所有任务的随机森林模型宏平均准确率均值见表 2。根据结果，我们在分析前于参与者层面进行中心化，但不进行缩放

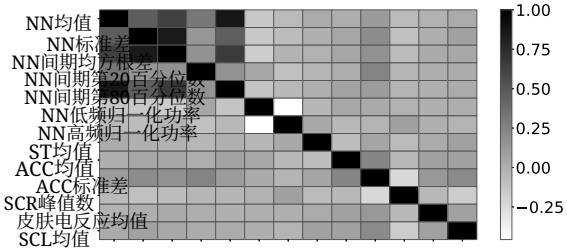


图3：完整数据集的特征相关矩阵。大多数特征之间的相关性较低，这表明每个特征都提供了独特的信息。

表 2：在第 3 节考察的所有任务中，预处理时在参与者层面对特征进行中心化但不进行缩放，可获得最高的宏平均准确率均值。报告的性能由通过 NLOPOCV 训练的随机森林模型实现。

参与者		随机森林宏平均准确率	
中心化	缩放	Mean	SEM
✓	✓	61。3%	2.3%
✓	X	**65.7**%	2.4%
X	✓	61。7%	2.1%
X	X	61。3%	2.1%

表3：本研究中使用的十三个特征。

特征	传感器	单位	引文	描述
NN均值	PPG	秒	[22, 29]	NN间期均值
NN标准差	PPG	秒	[29]	NN间期标准差
NN间期均方根差	PPG	秒	[24, 26, 29]	相邻差值的均方根
NN间期第20百分位数	PPG	秒	[31]	NN间期的第20百分位数
NN间期第80百分位数	PPG	秒	[33]	NN间期的第80百分位数
NN低频归一化功率	PPG	% 功率	[24, 25, 29]	0.04-0.15 Hz 频段功率 / 总功率
NN高频归一化功率	PPG	% 功率	[24, 25, 29]	0.15-0.40 Hz 功率 / 总功率
ST均值	TMP	◦摄氏度	[29]	温度样本均值
ACC均值	ACC	重力加速度	[23]	$\|\|[x, y, z]\|\|$的均值
ACC标准差	ACC	重力加速度	[23]	$\|\|[x, y, z]\|\|$的标准差
SCL均值	EDA	微西门子	[26, 28]	慢性振幅均值
皮肤电反应均值	EDA	微西门子	[26, 28]	相位性振幅均值
SCR峰值数	EDA	峰值/时间	[22, 27]	SCR 峰值数 / 总时间。

结果中的随机森林分析 。值得注意的是，虽然随机森林模型通常不需要缩放，但在本例中，这样做是合适的是因为应用于每位参与者的。

3 结果

我们围绕研究中的参与者回答四个问题：

1. 独处与处于社交情境时，生物标志物有何差异?
2. 社交情境的事件前、事件阶段和事件后阶段，生物标志物有何差异?
3. 双人互动与群组互动中的生物标志物有何差异?
4. 隐式与显式评价性互动中的生物标志物有何差异?

在后续分析中，我们在任务层级通过威尔科克森符号秩检验和Benjamini-Hochberg程序确定统计显著性，且 $\alpha = .05$。特征图中的点估计表示中位数及其95%自助法置信区间（这两者与非参数威尔科克森符号秩检验的契合度更高）。比较模型性能的图表展示了NLOPOCV各折性能的平均值。性能指标宏平均准确率是经宏平均处理的准确率，用于控制测试数据中的类别不平衡。在我们的所有分析中，宏平均的基线性能均为50%。模型性能图中的误差线表示标准误。我们在内部交叉验证循环中进行了超参数调优，以确定随机森林复杂度剪枝以及特征选择方法（即互信息标准或方差分析F值）。

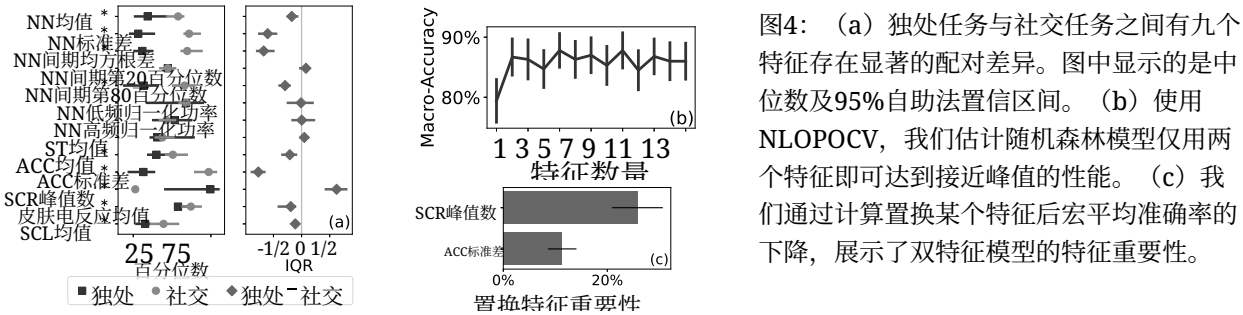
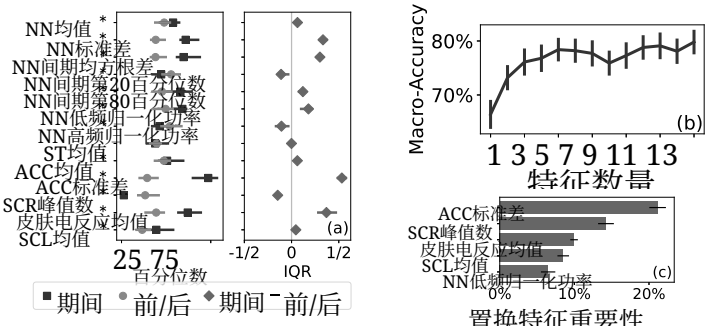


图4：（a）独处任务与社交任务之间有九个特征存在显著的配对差异。图中显示的是中位数及95%自助法置信区间。（b）使用NLOPOCV，我们估计随机森林模型仅用两个特征即可达到接近峰值的性能。（c）我们通过计算置换某个特征后宏平均准确率的下降，展示了双特征模型的特征重要性。

图 5：(a) 社交事件与其前/后阶段之间有 12 个特征存在显著的配对差异。图中展示的是中位数及其 95% 自助法置信区间。(b) 使用 NLOPOCV，我们估计随机森林模型使用 13 个特征时可达到最佳性能。(c) 我们通过置换某个特征后宏观准确率的下降，展示仅使用 5 个特征的模型中的特征重要性。



3.1 社交情境

为比较不同社交情境下的被动特征（即独处与社交任务），我们将事件观测分为独处组和社交组：独处组仅包含独处任务同期阶段的数据，社交组则包含所有双人及群体对话同期阶段的数据。

独处与社交数据集之间有 9 个特征存在显著差异（见图 4(a)）。差异最大的 3 个特征是 NN间期均方根差、ACC标准差和 SCR峰值数。我们进行多变量分析，以评估各特征的独立性。使用 NLOPOCV，我们发现，仅用 2 个特征，随机森林模型就能以最佳性能预测某个测试样本属于独处数据集还是社交数据集（见图 4(b)）。

使用完整数据集选出两个最佳特征后，我们再拟合一个随机森林，以通过置换来考察各特征的重要性（见图 4(c)）。按照这一流程得到的模型仅使用皮肤电活动和 ACC 传感器的特征。我们可以从图 4(b) 推断，来自光电容积脉搏波和 TMP 传感器的额外特征并未为模型提供更多信息。尽管如此，若干由光电容积脉搏波导出的特征在两组之间的中位数确实存在显著差异。

3.2 社交时序

为了解特征如何随时间变化，我们进行了两项分析：（1）比较社交事件期间记录的特征与合并后的前后阶段特征；（2）比较所有社交任务中社交任务前阶段与社交任务后阶段记录的特征。比较事件前后社交阶段的依据是相关文献指出，这些阶段可能具有相关但不同的过程特征。例如，预期性焦虑（即事件前焦虑）可能与当下焦虑更为相似，并涉及生理唤醒，而事件后加工可能主要是一种认知过程 [11]。

3.2.1 事件与前后阶段

社交事件与其前后阶段之间有 12 个特征存在显著的配对差异。此分析排除了独处体验的所有数据。差异最大的三项显著特征是 ACC 标准差、皮肤电反应均值和 NN 标准差。完整的特征分析见图 5（a）。

与之前一样，我们通过使用一小部分特征拟合随机森林模型来进行多变量分析。使用 NLOPOCV，我们确定只需五个特征，模型性能即可达到距峰值性能一个标准误以内。我们使用完整数据集选出最佳的五个特征，然后通过随机森林模型进行特征置换，以确定特征重要性。重要性排序依次为 ACC标准差、SCR峰值数、皮肤电反应均值、SCL均值和 NN低频归一化功率（见图 5）。

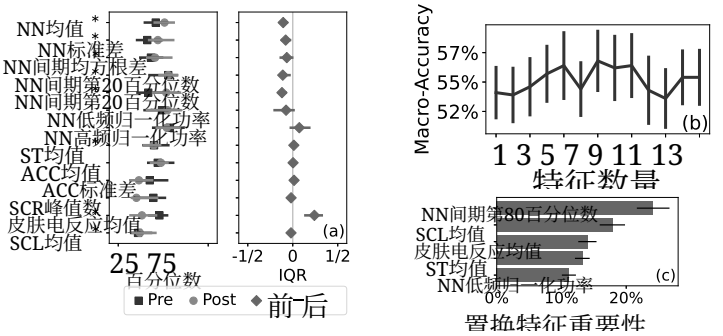


图 6：(a) 前阶段与后阶段之间有八个特征存在显著的配对差异。图中显示的是中位数及其 95% 自助法置信区间。(b) 使用 NLOPOCV，我们估计随机森林模型在使用五个特征时达到峰值性能。(c) 我们通过计算特征置换后宏平均准确率的下降幅度，展示五特征模型的特征重要性。

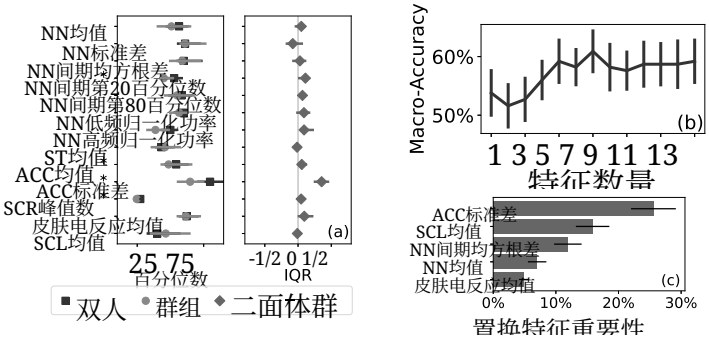


图 7：（a）四个特征在双人事件与群体事件之间存在显著的配对差异。图中展示的是中位数及其 95% 自助法置信区间。（b）我们使用 NLOPOCV 估计随机森林模型在使用七个特征时可达到最佳性能。（c）我们通过计算置换某个特征后宏平均准确率的下降幅度，展示五个特征模型中的特征重要性。

有趣的是，尽管我们的分析将特征筛选为五个，但模型似乎一直在提升，直至使用全部十三个特征。

3.2.2 事件前与事件后

我们进行了第二次时间分析，比较社交任务事件前阶段与事件后阶段的生物标志物特征。结果显示有八项显著的配对差异：五项显著的 NN 特征、ST均值、皮肤电反应均值和 SCL均值。完整结果见图 6(a)。通过多变量分析并使用 NLOPOCV 确定最小模型，我们未能有效剔除任何特征（即使纳入所有特征，整体表现也不理想）。这表明，根据我们所采用的特征，无法有效区分社交情境的事件前阶段与事件后阶段。

3.3 社交互动规模

我们使用二元事件特征和群体事件特征重复分析，以确定被动传感器能否区分社交互动中的人数。我们仅观察到四项显著的特征差异：ACC标准差、ACC均值、NN间期第20百分位数和 SCR峰值数。拟合随机森林模型后，预测准确率为 58%，仅略高于随机猜测水平（即 50%）；五特征模型的特征重要性主要取决于 ACC标准差。

3.4 社交威胁

比较隐性社交威胁情境与显性社交威胁情境下的生物标志物差异后，我们发现只有一个特征存在显著差异，即 SCL均值。同样，采用 NLOPOCV 训练的随机森林模型难以达到高于随机猜测（即 50%）的表现。我们的模型持续低于随机猜测水平，这表明参与者之间存在系统性差异（因为 NLOPOCV 每次只在一名参与者的数据上进行测试）。

3.5 事后分析

已有研究 [34,35] 表明，不同个体的生理数据可能存在显著异质性。基于这一点，我们对自身数据进行事后聚类分析，以评估其异质性。在此分析中，我们聚焦于第 3 节介绍的两项任务：(1) 独处与社交，以及 (2) 期间／同时与之前／之后。之所以选择这两项任务，是因为只有这两项任务的预测模型显著优于随机猜测

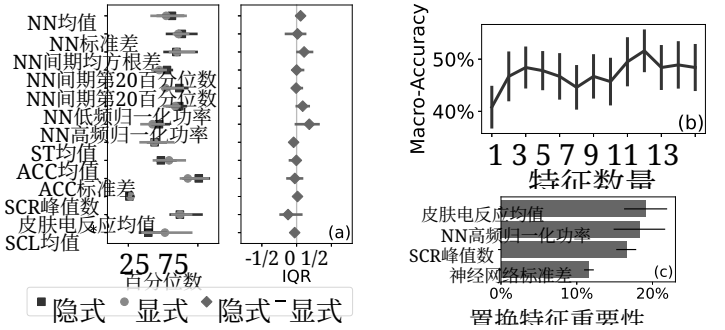


图8：（a）在隐式与显式评估之间，有一项特征的配对差异具有统计显著性。我们绘制了中位数及其95%自助法置信区间。（b）使用NLOPOCV，我们估计随机森林模型在采用10项特征时达到峰值性能。（c）对于包含4项特征的模型，我们通过计算置换某项特征后宏平均准确率的下降幅度，展示该特征的重要性。

表4：在前两项任务中，我们识别出28个潜在聚类。这表明数据具有高度异质性。唯一的例外是“独处与社交”中的独处数据集。几乎所有独处样本点要么属于同一个簇，要么被视为离群值。下表中的纯度以簇内多数类标签所占的百分比表示。

群组		簇			离群值
Task	标签	计数	E[大小]	E[纯度]	计数
独处与社交	独处	1	33	84.8%	16
独处与社交	社交	5	11.8	98.8%	95
进行中 vs. 前/后	进行中	7	10	74%	73
进行中 vs. 前/后	前/后	15	11.8	85%	129

的 50% 基线。在聚类之前，我们针对每项任务筛选出对预测最重要的特征（见图 4 (c) 和图 5 (c)）。随后，我们对数据应用基于层次密度的含噪声空间聚类（HDBSCAN） [36] 算法。该算法既能确定合适的聚类数目，也能过滤离群数据点。

对于独处与社交，HDBSCAN识别出6个簇，平均纯度超过90%（纯度是指簇中属于主导类别的点所占的百分比）。此外，54%的点被标记为离群值，因此未被分配到任何簇。对于同期与前/后期，HDBSCAN识别出22个簇。主要包含任务“进行中”事件期间所收集数据的簇，平均纯度为74%；而主要包含前/后期数据的簇，平均纯度为85%。这一分析的完整结果见表4，独处与社交聚类的可视化结果见图9。

4 讨论

本研究旨在描述与不同社交和非社交任务相关的生理模式，确定这些模式是否以及如何随情境、阶段和评价程度而变化，并探讨社交焦虑个体在不同任务中是否具有共同的生理反应模式。为探讨这些问题，我们：（1）考察与焦虑和压力检测相关的被动感测生物标志物，是否会因社交任务与非社交任务、社交互动的不同阶段、社交互动的群体规模以及评价威胁程度而异；（2）使用随机森林模型，确定达到最佳模型性能所需的特征数量，以及用于情境检测的最重要个体特征；（3）使用事后聚类分析，探讨与非社交任务相比，参与者在社交任务期间的生理反应是否呈现共同模式。

我们的结果表明，本研究使用的被动感测生物标志物能够可靠地区分个体参与社交互动与未参与社交互动的时段（即社交互动与非社交任务；社交互动期间与社交互动前后）。然而，这些生理特征无法区分社交事件发生前、紧接事件之前的预期性焦虑与事件后加工，也无法很好地区分具有细微差异的不同类型社交互动（例如，双人互动与群体互动；隐式与显性社会评价）。这些发现与此前的研究一致：该研究表明，可以使用被动感测生物标志物来区分基线、预期性焦虑和同时发生的焦虑阶段 [12, 15]。然而，这些发现与成功利用被动

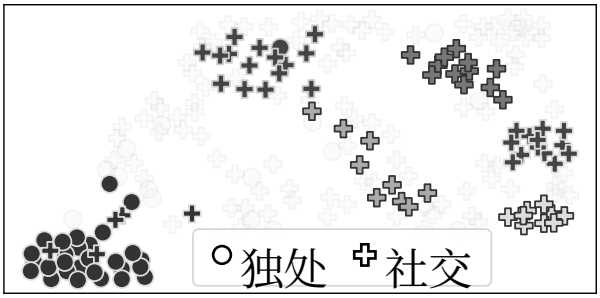


图9：来自社交体验的样本形成了六个不同的簇，而非社交样本大多归属于同一个簇。这个二维图是对真实底层数据的 t-SNE 表示。因此，有些簇看起来可能比预期更加分散。

感测来区分社交互动不同情境特征的研究并不一致。例如，一项研究在 Virtual Reality 演讲任务中发现，心率、皮肤电导和语音特征会随社交评价程度而变化 [16]；另一项研究则利用位置数据识别社交互动中的不同群体规模 [19],，不过，后一项研究依赖 WiFi 和实时定位服务来识别彼此距离较近且佩戴相同设备的人。在本研究中，实验操纵（例如，告知参与者他们将接受评价或不会接受评价）可能没有引起足够大的焦虑变化，因此无法通过生理指标检测出来。事实上，焦虑反应常表现出“不一致性” [37]，也就是说，焦虑的各个成分之间缺乏强关联（例如，威胁感知增强，却没有相应的生理唤醒）。与此一致，先前一项使用该数据集的研究未发现证据表明，在社交互动期间，社交焦虑的情感、认知、行为和生理成分之间存在一致性 [21]。

有趣的是，在某些情况下，我们观察到某项特征的中位数在不同情境之间存在显著的配对差异，但预测模型却无法可靠地区分这些情境。这种结果模式表明，这些特征在集中趋势方面可能因情境而异，但在整个时间窗口内，数据存在大量重叠和变异。换言之，有时可以利用多名参与者的数据来提高检验功效，从而观察到特征之间的显著差异。然而，这并不一定能转化为有效的模型，因为模型在进行预测时只能获取单个样本。

就具体的特征层面发现而言，加速度计标准差始终是区分社交与非社交情境最有效的特征。这可能反映了这样一个事实：参与者只有在积极参与某项社交互动任务时才会说话，而在其他时候则被要求安静地坐在电脑前。鉴于人们在说话时往往会做手势 [38],，这一特征特别适用于检测涉及对话的情境背景，也就不足为奇了。这些发现也与既往研究一致，后者表明，加速度计数据可用于准确检测人们在对话中何时说话 [39]。此外，群体互动期间的加速度计标准差显著低于双人互动期间的数值，并且在社交群组规模模型中成为信息量最大的特征；这或许可以用对话轮替模式来解释 [40]。也就是说，与小组对话相比，双人对话中可能有更多发言（进而做手势）的机会，也有更强的发言预期，这与对话中的人数有关。综合来看，我们的发现与既往研究表明，旨在实时检测社交互动的研究若将加速度计传感器纳入被动传感数据流，将会有所裨益。

我们的结果还表明，皮肤电活动特征，尤其是皮肤电导中的相位性峰值数量，对于检测一个人何时积极参与社交互动十分重要。这些发现与已有研究一致，相关研究发现皮肤电活动增强与压力和焦虑有关（例如， [29, 41]）。然而，当前研究结果不同于此前的研究；此前研究指出，从腕戴式传感器提取的皮肤电活动特征在对话场景中可能不可靠 [42]。事实上，有研究认为，Empatica E4 的 4Hz 采样率可能过低，无法准确检测数据中的皮肤电反应 [43]。虽然腕戴式设备测得的皮肤电活动指标在获得客观准确的读数方面可能存在局限，尤其是与实验室级心理生理设备相比时 [44],，但在区分不同情境方面，它们仍可能具有一定的现实场景实用性。无论如何，还需要使用其他设备重复验证我们的发现。

为更深入地探究这些发现，我们针对模型能够可靠区分的两组任务进行了聚类分析：独处与社交，以及任务期间与任务前后。这些分析旨在考察参与者在这些任务中表现最突出的特征时，是否呈现出共同或异质的生理反应。在这两组任务中，我们的结果表明，同一情境下不同个体的生理反应模式既存在显著差异，也具有重要共性。例如，在所有社交互动中，共有五个簇代表了不同的生理反应模式，

这表明，参与者在社交互动中的 ACC 标准差和 SCR峰值数可能存在个体差异，但这些模式并非完全具有个体特异性。因此，接下来一个重要步骤是明确这些簇具体代表什么。例如，这些簇可能代表状态焦虑水平；社交焦虑个体会根据当下的焦虑程度，呈现不同的动作和皮肤电导模式。了解这些共同的生理反应模式，以及它们对应的感受或行为，有助于在最需要的时间和地点部署即时自适应干预。

4.1 局限性与未来方向

解读本研究时必须考虑其局限性，并应承认，研究设计的某些特征可能会限制这些发现对现实情境的可推广性。首先，本研究通过 Zoom 在线进行，实验者在被试内操纵的对话条件（例如，双人对话与群组对话）均涉及与陌生人讨论无敏感性的主题，因此互动类型有限。目前尚不清楚这些发现对不同条件的适用程度；例如，与朋友而非陌生人交谈，或与他人进行面对面互动时，参与者可能较少依赖手势表达这一沟通工具，而 ACC 标准差在区分社交与非社交情境方面的参考价值也可能较低。其次，我们的样本在人口特征方面相对同质，包括年龄（M = 19.28，标准差 = 1.91）、性别（女性占 76.1%）、种族（白人占 67.4%）和族裔（非拉丁裔/西班牙裔占 89.1%）。鉴于研究发现，诸如皮肤电活动等公认的“金标准”心理生理测量指标是在以白人参与者为主的样本中开发并测试的，而典型的数据清理程序可能会系统性地排除反应性较低的非白人参与者 [45]，因此开展涉及更多元样本的后续研究至关重要。第三，使用腕戴式设备采集某些心理生理数据存在固有局限。例如，Empatica E4 等腕带使用干电极（而非涂有凝胶、可增强皮肤电导率的湿电极），且采样率较低，这使得准确测量皮肤电导极具挑战性 [44]。未来，使用不同设备、规模更大且更多元的样本，并在现实环境中开展重复验证，将是关键。

4.2 结论

尽管存在局限性，本研究仍在几个关键方面推进了我们对通过被动感测进行社交情境检测的理解。首先，使用本研究中的传感器，我们能够可靠地区分社交与非社交情境，但无法区分具有更细微情境差异的社交互动。加速度计标准差和皮肤电反应峰值数量是区分社交与非社交情境时最重要的两个个体特征。在这些特征上，生理反应模式存在相当大的差异，但也出现了可能具有意义的簇，为参与者之间既存在个体差异又具有共性提供了一些证据。综合来看，这些发现表明，加速度计和皮肤电导传感器对于识别何时有人正在进行社交互动可能尤其重要，但无法识别该互动的具体特征。重要的是，即使仅了解某人是否与他人在一起，也有助于适当地定制和部署即时自适应干预。然而，未来还需开展研究，以更好地检测社交互动的特征，并准确阐明社交互动中的生理反应模式簇所代表的含义，从而进一步改进即时自适应干预。

参考文献

[1] R. C. Kessler、M. Petukhova、N. A. Sampson、A. M. Zaslavsky 和 H.-U. Wittchen，“美国焦虑障碍和心境障碍的12个月患病率、终生患病率及终生发病风险”，《国际精神病学研究方法杂志》，第21卷第3期，第169–184页，2012年。[2] A. P. Association，《精神障碍诊断与统计手册》，第5版。华盛顿哥伦比亚特区：美国精神病学协会，华盛顿哥伦比亚特区，2022年。[3] I. M. Aderka、S. G. Hofmann、A. Nickerson、H. Hermesh、E. Gilboa-Schechtman 和 S. Marom，“社交焦虑障碍导致的功能损害”，《焦虑障碍杂志》，第26卷第3期，第393–400页，2012年4月。[4] B. F. Grant、D. S. Hasin、C. Blanco、F. S. Stinson、S. P.Chou、R. B. Goldstein、D. A. Dawson、S. Smith、T. D. Saha 和 B. Huang，“美国社交焦虑障碍的流行病学：全国酒精及相关疾病流行病学调查结果”，《临床精神病学杂志》，第66卷第11期，第6546页，2005年11月，出版商：Physicians Postgraduate Press, Inc.

[5] P. S. Wang、P. Berglund、M. Olfson、H. A. Pincus、K. B. Wells 和 R. C. Kessler，“全国共病调查复现研究中，精神障碍首次发作后初次治疗接触的失败与延迟”，《普通精神病学档案》，第62卷第6期，第603–613页，2005年6月。[6] C. Acarturk、P. Cuijpers、A. van Straten 和 R. de Graaf，“社交焦虑障碍的心理治疗：一项荟萃分析”，《心理医学》，第39卷第2期，第241–254页，2009年2月。[7] A. E. Kazdin，“弥合治疗缺口：扩大循证心理社会干预的关键挑战”，《行为研究与治疗》，第88卷，第7–18页，2017年1月。[8] E. M. Goetter、M. R. Frumkin、S. A. Palitz、M. B. Swee、A. W. Baker、E. Bui 和 N. M. Simon，“社交焦虑障碍和广泛性焦虑障碍患者接受心理健康治疗的障碍”，《心理服务》，第17卷第1期，第5–12页，2020年2月。[9] I. Nahum-Shani、S. N. Smith、B. J. Spring、L. M. Collins、K. Witkiewitz、A. Tewari 和 S. A. Murphy，“移动健康中的即时自适应干预（JITAIs）：持续健康行为支持的关键组成部分与设计原则”，《行为医学年鉴》，第52卷第6期，第446–462页，2018年5月。[10] C. R. van Genugten、J. Schuurmans、F. Lamers、H. Riese、B. W. Penninx、R. A. Schoevers、H. M. Riper 和 J. H. Smit，“情感障碍患者使用基于智能手机的生态瞬时评估所感受到的负担及其依从性”，《临床医学杂志》，第9卷第2期，第322页，2020年1月。[11] R. G. Heimberg、F. A. Brozovich 和 R. M. Rapee，“社交焦虑障碍的认知行为模型”，载于《社交焦虑：临床、发展与社会视角》，第三版。学术出版社，2014年，第705–728页。[12] R. Shaukat-Jali、N. v. Zalk 和 D. E. Boyle，“利用腕戴式可穿戴设备采集的生理数据检测亚临床社交焦虑：小规模可行性研究”，《*JMIR*形成性研究》，第5卷第10期，文章编号e32656，2021年10月。[13] H. Rashid、S. Mendu、K. E. Daniel、M. L. Beltzer、B. A.Teachman、M. Boukhechba 和 L. E. Barnes，“利用稀疏采集的移动传感器数据预测社交焦虑的主观测量”，《*ACM*交互、移动、可穿戴与普适技术会议论文集》，第4卷第3期，第1–24页，2020年9月。[14] M. Boukhechba、P. Chow、K. Fua、B. A. Teachman 和 L. E. Barnes，“基于大学生全球定位系统轨迹预测社交焦虑：可行性研究”，《*JMIR*精神健康》，第5卷第3期，文章编号e10101，2018年7月。[15] K. Pisanski、A. Kobylarek、L. Jakubowska、J. Nowak、A. Walter、K. Błaszczy´nski、M. Kasprzyk、K. Łysenko、I. Sukiennik、K. Pi.atek、T. Frackowiak 和 P. Sorokowski，“多模态压力检测：检验特里尔社会压力测试中语音、激素及生理反应的协变关系”，《激素与行为》，第106卷，第52–61页，2018年11月。[16] M. Barreda-Ángeles、S. Aleix-Guillaume 和 A. Pereda-Baños，“用户对立体360°视频中虚拟观众表面态度的心理生理、语音及自我报告反应”，*Virtual Reality*，第24卷第2期，第289–302页，2020年6月。[17] R. Wang、F. Chen、Z. Chen、T. Li、G. Harari、S. Tignor、X. Zhou、D. Ben-Zeev 和 A. T. Campbell，“StudentLife：利用智能手机评估大学生的心理健康、学业表现和行为趋势”，载于《*2014*年*ACM*普适与泛在计算国际联合会议论文集》，系列：UbiComp ’14。美国纽约州纽约：计算机协会，2014年9月，第3–14页。[18] D. Ben-Zeev、E. A. Scherer、R. Wang、H. Xie 和 A. T. Campbell，“新一代精神科评估：利用智能手机传感器监测行为与心理健康”，《精神康复杂志》，第38卷第3期，第218–226页，2015年9月。[19] N. C. B. ZAKARIA、R. BALAN 和 Y. LEE，“StressMon：通过被动感测工作惯例和群体互动的变化，实现可扩展的感知压力与抑郁检测”，《*ACM*人机交互会议论文集》，第3卷，第37:1–29页，2019年11月。[20] R. P. Mattick 和 J. C. Clarke，“社交恐惧症受审视恐惧和社交互动焦虑测量工具的开发与验证”，《行为研究与治疗》，第36卷第4期，第455–470页，1998年4月。[21] E. R. Toner、M. A. Larrazabal、L. Cai、T. R. Henry、J. MacCormack、M. Boukhechba、L. Barnes 和 B. Teachman，“不同评价威胁水平下社交焦虑与情绪反应的一致性”，2022年10月。[22] F.-T. Sun、C. Kuo、H.-T. Cheng、S. Buthpitiya、P. Collins 和 M. Griss，“利用生理传感器进行活动感知型心理压力检测”，载于《移动计算、应用与服务：第二届*ICST*国际会议，*MobiCASE 2010*，美国加利福尼亚州圣克拉拉，*2010*年*10*月*25*日至*28*日，修订版精选论文*2*》。施普林格，2012年，第282–301页。

[23] M. Boukhechba、A. R. Daros、K. Fua、P. I. Chow、B. A. Teachman 和 L. E. Barnes，“Demonicsalmon：利用智能手机监测大学生的心理健康与社交互动”，*Smart Health*，第 9 卷，页码 192–203，2018年。[24] D. Hernando、S. Roca、J. Sancho、Á. Alesanco 和 R. Bailón，“Apple Watch 在健康受试者放松及心理压力期间测量心率变异性的验证”，传感器，第 18 卷，第 8 期，文章编号 2619，2018年。[25] J. W. Ahn、Y. Ku 和 H. C. Kim，“一种用于压力评估的新型可穿戴式脑电与心电记录系统”，传感器，第 19 卷，第 9 期，文章编号 1991，2019年。[26] M. Barreda-Ángeles、S. Aleix-Guillaume 和 A. Pereda-Baños，“ 用户对立体 360 度视频中虚拟观众表面态度的心理生理、声音及自我报告反应 ”，*Virtual Reality*，第 24 卷，第 2 期，页码 289–302，2020年。[27] F. R. Ihmig、F. Neurohr-Parakenings、S. K. Schäfer、J. Lass-Hennemann 和 T. Michael，“基于生物信号的在线焦虑水平检测：基于一项针对惧蛛个体的随机对照试验的机器学习”，*PLOS ONE*，第 15 卷，第 6 期，文章编号 e0231517，2020年。[28] L. Petrescu、C. Petrescu、O. Mitruț, G. Moise、A. Moldoveanu、F. Moldoveanu 和 M. Leordeanu，“将生物信号测量融入虚拟现实环境以进行焦虑检测”，传感器，第 20 卷，第 24 期，文章编号 7088，2020年。[29] B. A. Hickey、T. Chalmers、P. Newton、C.-T. Lin、D. Sibbritt、C. S. McLachlan、R. Clifton-Bligh、J. Morley 和 S. Lal，“用于检测和监测心理健康状况与压力的智能设备和可穿戴技术：系统综述”，传感器，第 21 卷，第 10 期，文章编号 3461，2021年。[30] M. Elgendi、I. Norton、M. Brearley、D. Abbott 和 D. Schuurmans，“热带环境下测得的应急救援人员加速度光电容积脉搏波中的收缩峰检测”，*PLOS ONE*，第 8 卷，第 10 期，文章编号 e76585，2013年。[31] D. Makowski、T. Pham、Z. J. Lau、J. C. Brammer、F. Lespinasse、H. Pham、C. Schölzel 和 S. H. A. Chen，“NeuroKit2：用于神经生理信号处理的 Python 工具箱”，*BehaviorResearch Methods*，第 53 卷，第 4 期，页码 1689–1696，2021年2月。[32] M. Karlsson、R. Hörnsten、A. Rydberg 和 U. Wiklund，“心率变异性分析前对动态心电图记录中 RR 间期离群值进行自动滤波：与仔细编辑的数据进行比较”，*Biomedical Engineering Online*，第 11 卷，页码 1–12，2012年。[33] K. Plarre、A. Raij、S. M. Hossain、A. A. Ali、M. Nakajima、M. Al’Absi、E. Ertin、T. Kamarck、S. Kumar、M. Scott等，“基于在自然环境中收集的感官测量数据，对心理压力进行连续推断”，载于第十届 *ACM/IEEE* 传感器网络信息处理国际会议论文集。IEEE，2011年，页码 97–108。[34] E. H. Siegel、M. K. Sands、W. Van den Noortgate、P. Condon、Y. Chang、J. Dy、K. S. Quigley 和 L. F. Barrett，“情绪指纹还是情绪群体？情绪类别自主神经特征的荟萃分析研究。”*Psychological Bulletin*，第 144 卷，第 4 期，页码 343，2018年。[35] K. Hoemann、Z. Khan、M. J. Feldman、C. Nielson、M. Devlin、J. Dy、L. F. Barrett、J. B. Wormwood 和 K. S. Quigley，“情境感知体验抽样揭示情感体验的变异规模”，*Scientific Reports*，第 10 卷，第 1 期，页码 1–16，2020年。[36] L. McInnes、J. Healy 和 S. Astels，“HDBSCAN：基于层次密度的聚类”，*TheJournalofOpenSourceSoftware*，第 2 卷，第 11 期，2017年3月。[37] T. Hollenstein 和 D. Lanteigne，“情绪一致性的模型与方法”，*Biological Psychology*，第 98 卷，页码 1–5，2014年4月。[38] S. Goldin-Meadow，“手势在沟通与思维中的作用”，*Trends inCognitive Sciences*，第 3 卷，第 11 期，页码 419–429，1999 年 11 月。[39] H. Hung 、 G. Englebienne 和 L. Cabrera Quiros，“ 利用单个佩戴式加速度计检测交谈群体”，载于第十六届多模态交互国际会议论文集。土耳其伊斯坦布尔：ACM，2014年11月，页码 84–91。[40] T. Stivers、N. J. Enfield、P. Brown、C. Englert、M. Hayashi、T. Heinemann、G. Hoymann、F. Rossano、J. P. de Ruiter、K.-E. Yoon 和 S. C. Levinson，“对话中轮流发言的普遍性与文化差异”，美国国家科学院院刊，第 106 卷，第 26 期，页码 10 587–10 592，2009年6月，出版方：美国国家科学院院刊。

[41] J. Globisch、A. O. Hamm、F. Esteves 和 A. Öhman，“恐惧迅速显现：动物恐惧受试者惊跳反射增强的时间进程”，心理生理学，第 36 卷，第 1 期，页码 66–75，1999年。[42] N. Milstein 和 I. Gordon，“验证在涉及互动二元状态的研究环境中使用 Empatica E4 得到的皮肤电活动和心率变异性测量指标”，*Frontiers in Behavioral Neuroscience*，第 14 卷，文章编号 148，2020年8月。[43] J. J. Braithwaite、D. G. Watson、R. Jones 和 M. Rowe，“心理学实验中皮肤电活动（EDA）与皮肤电反应（SCRs）的分析指南。”心理生理学，第 49 卷，第 1 期，页码 1017–1034，2013年。[44] I. R. Kleckner、M. J. Feldman、M. S. Goodwin 和 K. S. Quigley，“心理生理学研究中移动设备的选择与基准评估框架”，*Behavior ResearchMethods*，第 53 卷，第 2 期，页码 518–535，2021年4月。[45] D. E. Bradford、A. DeFalco、E. R. Perkins、I. Carbajal、J. Kwasa、F. R. Goodman、F. Jackson、L. N. S. Richardson、N. Woodley、L. Neuberger、J. A. Sandoval、H. J. Huang 和 K. J. Joyner，“谁的信号正在被放大？迈向更公平的临床心理生理学”，《临床心理科学》，文章编号 21677026221112117，2022年12月，出版方：SAGE Publications Inc。