

# 关于某博彩运营商【理性博彩】传播策略的 情绪反应生物计量研究

Guillermo Mejías Martínez<sup>[1]</sup>, Ubaldo Cuesta<sup>[2]</sup>

本文源于马德里康普顿斯大学与西班牙国家盲人组织合作设立的【理性博彩】及其传播研究特别教席框架。

## 1. 研究背景

赌博成瘾被定义为「持续反复的赌博行为，对个人生活造成严重负面影响」。典型症状包括：判断力失常、试图追回损失、整天想着赌博、无法自控（美国精神病学协会，1980）。研究发现，赌客希望获得理性博彩指引工具，特别是能控制预算的功能，比如查看他人评价和自身消费记录（Griffiths等，2009）。

目前理性博彩措施的普及度仍然较低。数据显示，只有5%的潜在问题赌徒会使用自我禁赌服务（Cohen等，2011）。在线上平台，仅1.2%用户设置消费限额，13%会经常使用理性博彩控制功能（Forsström等，2016）。这说明虽然大多数赌客认同理性博彩理念，但需要更有效的引导方式。

为评估相关宣传效果，本研究采用NeurolynQ设备进行神经学定向群体测试。团队曾用该方法分析过受众对反疫苗观点的生理反应，发现年轻人在讨论该话题时情绪波动最明显（Cuesta等，2022）。

## 2. 研究目标

彩票运营商需遵循欧洲彩票协会的理性博彩标准。本研究通过神经学测量，检验现有宣传材料（包括宣传册、广告、社交媒体内容等）是否清晰有效地传递了理性博彩信息。

---

<sup>1</sup>. Guillermo Mejías Martínez. Universidad Complutense de Madrid,

<sup>2</sup>. Ubaldo Cuesta Cambra. Universidad Complutense de Madrid.

### 3. 研究方法

为开展研究，我们进行了一个神经学定向群体研究，对象为 9 名年轻大学生。定向群体的讨论，流程如下：

#### 一、自由讨论环节

1. 电视广告
2. 报纸广告
3. 广播广告
4. 理性博彩概念探讨

#### 二、定向讨论环节

1. 社交媒体内容
2. 理性博彩手册
3. ONCE理性博彩政策核心问题：
  - 是否告知赌博风险？
  - 是否提供戒赌支持信息？
  - 是否明确禁止未成年人参赌？
  - 是否避免诱导过度消费？

**主持人：乌巴尔多·奎斯塔**

**参与者：19-21岁女大学生（全部女性）**

**地点：马德里康普顿斯大学信息科学学院**

情绪反应的测量是通过放置在参与者食指和无名指上的两个电极进行的，这些电极测量由参与者出汗情况决定的变化。面对焦虑、愉悦、恐惧或排斥等情况，机体的反应是改变出汗水平。这种出汗水平的变化与情绪强度的变化相关。这一数据通过测量个体的心率得到丰富，提高了皮肤电反应测量的精度，并可以通过忽略内部稳态过程引起的假阳性来验证它。

这些测量由 NeuroLynQ 系统 (Shimmersensing, 2019) 提供，该系统使用心搏间期变异性，而不仅仅是个体的心率。这一特性使我们能够利用其心率的变化（而非绝对值），并且该变化独立于参与者的静息心率阈值。这样，每个参与者都是其自身的对照。

由于在定向群体中，随着小组会议中每个主题的讨论，会接连发生一系列社会动态，因此我们着手分析 NeuroLynQ 在最相关的时间窗口内提供的结果。在本研究中，指每个讨论主题以及仅观看或收听待分析的视听作品的时间段。

在安装好图 1 中展示的所有个人和小组仪器设备后，即可开始会议。NeuroLynQ 的算法利用从皮肤电反应和心搏间期变异性（源自心电图信号提供的数据）获得的数据，并将响应分为 3 类：无响应（参与者对刺激没有情绪反应）、中等响应（参与者对刺激有中等程度的反应）和高度响应（参与者对刺激有非常情绪化的反应）。通过这种分类，我们以参与者百分比的形式，获得了个体层面和集体层面在会议每个时刻所具有的情绪反应程度的呈现。

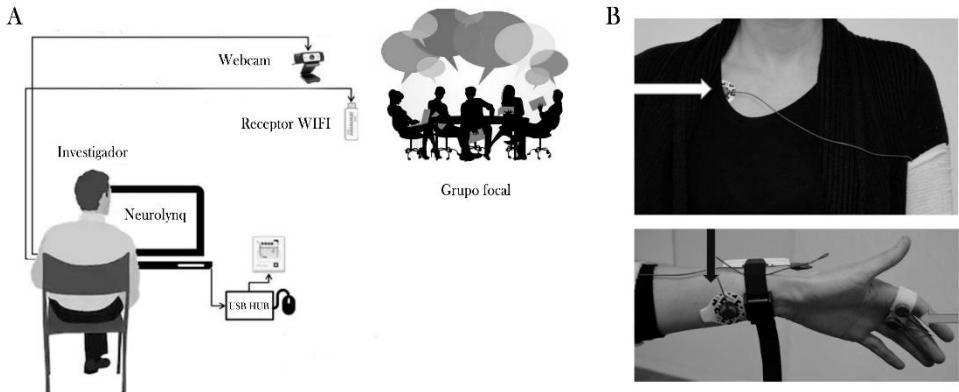


图 1. NeuroLynQ 小组及个体安装示意图

A 小组安装示意图，包括装有 NeuroLynQ 的电脑、WIFI 接收器和用于记录会议的网络摄像头；B 每个参与者的个体安装示意图，包括用于心率的电极（白色箭头）和用于皮肤电反应的电极（黑色箭头）。

在定向群体开始之前，为参与者配备了 NeuroLynQ 的个体传感器，除了告知他们在整个小组会议期间保持安装传感器的手臂放松外，没有给予任何其他额外指示。

## 4. 研究开展

### 4.1. 内容分析

当年轻人想到 ONCE 时，他们会想到帮助、募集资金和支持残疾人。他们给其品牌形象打 9 分，因此在他们中间评价很高。

他们声称 ONCE 比其竞争对手更具同理心、更团结、更无私且无害。事实上，用一个丛林动物做比喻，他们将所有博彩公司视为掠食者，而将 ONCE 视为一只乌龟——「ONCE 是团结的，帮助人们」。

然而，他们完全不认为 ONCE 在致力于宣传“负责任博彩”，尽管他们隐含地理解它确实如此。但这并非因为他们看到过 ONCE 在这方面的工作。

他们认为 ONCE 没有开展关于博彩成瘾的意识提升活动，也没有在销售点包含相关信息。尽管他们也说 ONCE 不鼓励过度购买产品，并且当其员工看到用户购买很多时会发出警告——「他们不会坚持让我奶奶多买一张彩票」。

他们也不认为禁止向未成年人销售，并将购买时刻与超市或家庭时刻联系起来。

关于创意作品：

- 电视广告活动：他们喜欢，但没有人看到负责任博彩的信息——「我没注意到它在那里」。
- 广播广告：能听到负责任博彩的信息，但出现在广告正文之外。不连贯，这使其被视为类似药品广告中的附加内容——「他们只是为了合规而加上去」。
- 报纸宣传册：看不到，有必要将其整合到广告内部。起初也不理解"Bon jugado"的概念，虽然经过简短讨论后最终理解了。凭直觉，他们不将"Bon Jugado"理解为一个帮助他人的游戏。他们觉得颜色过于饱和。
- 社交媒体活动：只有"设定一个限度"这条信息有效，认为这个口号很聪明。其他作品他们认为非常单薄，有些缺乏意义。他们认为展示的创意与音乐不协调。音乐非常欢快，而信息旨在用负责任博彩教育客户，因此两者之间没有联系——「听起来像 YouTube 教程的音乐」。
- 负责任博彩宣传册：无效，非常冗长繁琐。此外，橙色不与 ONCE 关联，因此他们不理解。字体也不被认为是品牌字体，觉得幼稚过时。此外，他们不认识 FEJAR 的标识，这引起了不信任——「看起来像橙汁的字体」「我不了解那个协会，他们应该解释一下是谁」。

关于 ONCE 负责任博彩策略的关键点：

- 他们不认为 ONCE 的职能是告知博彩者病态博彩的风险——「关于限度的信息我联想到的是生病的人，不是在说我」。
- 除一名参与者外，其余人不认为 ONCE 在其销售点包含了告知博彩成瘾支持服务的信息。当一名参与者告诉他们，根据她的经验，确实有销售员给出建议时，他们感到惊讶，认为这个因素非常积极。
- 他们将 ONCE 与家庭活动联系起来，与祖父母或父母一起购买彩票，

因此，他们没有注意到 ONCE 禁止向未成年人销售其产品——「ONCE 是一个非常家庭化的组织」。

- 他们认为 ONCE 不鼓励过度购买其产品或试图挽回损失，除了圣诞彩票的情况。

## 4.2. 情绪反应分析

接下来展示的是 NeuroLynQ 算法根据皮肤电反应和心搏间期变异性提供的结果，该结果以小组中处于高度、中等或某种（高度 + 中等）响应状态的参与者百分比形式编码，对应不同的时间窗口。

整个定向群体的响应百分比（中等响应 + 高度响应）高于认为参与者对活动开展产生积极响应的阈值（Shu 等人，2018）。这向我们证实，参与者不仅在个体层面上被情绪激活，在小组会议期间也在群体层面上被激活。得益于 NeurolynQ 提供的数据，我们可以使用像定向群体这样重要的定性技术，特别是在分析像内隐情绪反应这样重要但又“难以把握”的东西时，因为迄今为止，对这些反应的分析完全依赖于主持人的经验和直觉，通过观察和解读小组成员的非语言行为（情绪手势、微笑等）来“捕捉”它们。

数据显示，在整个小组会议期间，小组成员在话语的特定时刻产生情绪波动，有时共同达到 100% 的某种情绪响应峰值（图 3 中的蓝线）和一个 87.50% 的高度情绪响应的瞬时点。平均而言，某种响应为 50.35%，高度响应为 15.40%，这表明参与者保持了持续的注意力并投入到他们的回应中。

表 1 显示了定向群体每个讨论主题获得的不同数值。

主题	平均高度响应 %	最大高度响应 %	平均某种响应 %	最大某种响应 %
电视广告讨论	16,22	50	54,48	100
广播广告讨论	15,11	62,50	51,87	100
报纸宣传册讨论	16,45	62,50	57,39	87,50
社交媒体讨论	14,28	87,50	45,66	100

理性博彩	16,42	62,50	47,18	87,50
ONCE理性博彩策略	13,50	50	47,01	87,50
观看电视广告	13,72	50	43,03	87,50
收听广播广告	7,20	25	34,92	75
观看社交媒体作品	9,91	37,50	39,94	75

表 1. Neurolynq 测量的不同定向群体阶段及观看创意作品期间的高度响应%、某种响应%及最大响应%结果。来源：自行绘制。

同时，对观看/收听内容的统计分析显示，无论是在高度响应百分比 (ANOVA, 创意类型:  $F(3,005)=25.68; p<0.001$ )，所有创意类型之间的交互作用均显著 (Tukey, HSD,  $p<0.001$ ) 方面，还是在某种响应百分比 (ANOVA, 创意类型:  $F(3,005)=33.37; p<0.001$ )，电视广告-广播广告和电视广告-社交媒体之间的交互作用显著 (Tukey, HSD,  $p<0.001$ ) 方面，所发现的差异均是显著的。情感影响最大的创意类型是电视广告。结合口头回应，社交媒体创意作品引起的兴趣尽管同属视听作品，但低于电视广告。不同讨论主题之间的响应没有显著差异。

因此，使用像神经学定向群体这样的技术似乎证实，可以使像定向群体这样重要的定性技术的运用更加“严谨化”，特别是在分析像内隐情绪反应这样重要但又“难以把握”的东西时。我们必须记住，迄今为止，对这些反应的分析完全依赖于主持人的经验和直觉，通过观察和解读小组成员的非语言行为（情绪手势、微笑等）来“捕捉”它们 (Dodaro, 2016; Canales, 2006)。借助 NeuroLynQ 仪器技术，我们可以测量在像定向群体这样的社会动态中产生的内隐情绪，通过将这种分析与外显回应（参与者的陈述）相结合，获得对所研究目标群体的完整洞察。

## 5. 结论

所进行的定向群体内容分析显示，ONCE 实体在年轻人中评价很高，他们将 ONCE 与金钱联系起来，但目的是回馈所有人（不仅是盲人，而是帮助整个社会）。然而，绝大多数参与者在任何其广告作品中都未感知到负责任博彩。尽管如此，参与者将所有博彩公司视为掠食者，但 ONCE 除外。

关于创意作品，小组对信息的有效性提出了合理的质疑，并支持使信息更可信、更真实、创意更丰富、并优先考虑负责任博彩信息的必要性。他们将展示的广告作品归类为适合 40 岁以上人群，并认为如果改变信息和创意，他们可以开展一个更有趣的社交媒体活动。

情绪反应分析表明，无论是外显还是内隐层面，效果最好的创意是电视广告，但负责任博彩的信息被忽视了，因此有必要重新设计该创意，以最大化其在负责任博彩方面的影响力。

## 6. 参考文献

(参考文献部分保持不变，因主要为原文列表)

Canales, M. (2006). El grupo de discusión y el grupo focal (pp. 265-287). *Metodologías de investigación social. Introducción a los oficios*. LOM Ediciones

Cuesta, U., Mejías, G., Roberts, M., & Bengochea-González, C. (2022). Respuesta psicofisiológica de una audiencia en relación a líderes de opinión anti-vacunas: prueba de concepto. *Estrategias de comunicación: Género, Persuasion y Redes sociales*. ISBN 978-84-18914-55-3. Gedisa.

Cohen, I. M., McCormick, A. V., Corrado, R. R., Corporation, B. C. L., Responsibility, B. C. C. for S., & Valley, U. of the F. (2011). *BCLC's Voluntary Self-Exclusion Program: perceptions and experiences of a sample of program participants / Irwin M. Cohen, Amanda V. McCormick, and Raymond R. Corrado*. <https://doi.org/http://dx.doi.org/10.11575/PRISM/9433>

Dodaro , C . (2016). "Decime quien consume a quien" Estudio sobre la creación de disposiciones al consumo de bebidas con alcohol generadas por la publicidad en jóvenes. *Revista de Comunicación y Salud*, 61, 55-69. [https://doi.org/10.35669/revistadecomunicacionysalud.2016.6\(1\).55-69](https://doi.org/10.35669/revistadecomunicacionysalud.2016.6(1).55-69)

Forsström, D., Hesser, H., & Carlbring, P. (2016). Usage of a Responsible Gambling Tool: A Descriptive Analysis and Latent Class Analysis of User Behavior. *Journal of Gambling Studies*, 32(3), 889–904. <https://doi.org/10.1007/s10899-015-9590-6>

Griffiths, M. D., Wood, R. T. A., & Parke, J. (2009). Social responsibility tools in online gambling: a survey of attitudes and behavior among Internet gamblers. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 12(4), 413–421. <https://doi.org/10.1089/cpb.2009.0062>

Nelson, S. E., LaPlante, D. A., Peller, A. J., Schumann, A., LaBrie, R. A., & Shaffer, H. J. (2008). Real limits in the virtual world: self-limiting behavior of Internet gamblers. *Journal of Gambling Studies*, 24(4), 463–477. <https://doi.org/10.1007/S10899-008-9106-8>

Shimmersensing. (2019). NeuroLynQ v2.0 User Guide.

[http://shimmersensing.com/images/uploads/docs/NeuroLynQ\\_User\\_Guide\\_rev2.0a.pdf](http://shimmersensing.com/images/uploads/docs/NeuroLynQ_User_Guide_rev2.0a.pdf) Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., & Yang, X. (2018). A Review of Emotion Recognition Using Physiological Signals. *Sensors* (Basel, Switzerland), 18(7). <https://doi.org/10.3390/s18072074>

# 基于生理信号量化观众的情感反应

Sriram Raju Dandu  
弗吉尼亚大学 电气与计算机工程  
系, 美国夏洛茨维尔市  
sd9aj@virginia.edu

Geoff Gill  
Shimmer Americas, 美国  
波士顿  
ggill@shimmersensing.com

Caleb Siefert  
密歇根大学迪尔伯恩分校 行为  
科学系, 美国  
csiefert@umich.edu

**摘要**—本文提出了一种新颖的度量方法，从心电图和皮肤电活动信号中提取特征，以近乎实时地将每位观众的反应量化为无、中等和高。该度量方法直观，并且与生物特征数据的手动编码结果吻合良好。该度量方法有效性和可靠性的初步证据来自一系列使用不同刺激（即体育赛事；广告；情景喜剧）和不同参与者群体（N=231）的研究。本文还讨论了这种方法在多种用途上的实用性。

**关键词**—媒体研究, 生理信号, 心率, 皮肤电导, 拐点, 信号处理, 自适应阈值。

## I. 引言

情感反应可以广义地定义为对特定内心感受的反应，伴随着可能或可能不外显的生理变化。简而言之，情感涉及生理唤醒和认知归因（即标签化）。情感唤醒涉及自主神经系统活动的变化。自主神经系统是一个通用生理系统，负责调节外周功能，如心率和呼吸节奏[1]。该系统由交感神经和副交感神经分支组成，通常分别与激活和放松相关。由于ANS的通用性，其活动不仅仅是情感反应的功能，还涵盖了与消化、稳态、努力、注意力等多种其他功能。

ANS激活的常用评估指标基于皮肤电（即汗腺）或心血管（即血液循环系统）反应。皮肤电活动是人体导致皮肤电特性持续变化的属性。EDA的测量，通常称为皮肤电导，由交感神经活动自主调节，交感神经活动在潜意识层面驱动人类行为、认知和情感状态。SC为自主情绪调节提供了直接的洞察。Najström和Jansson [2] 研究了在压力性生活事件后，对掩蔽威胁图片的皮肤电导反应性对情感反应的预测价值。EDA可用于索引几种不同的过程，如激活、注意力以及任务重要性或刺激的情感强度[3]。Simons等人[4]表明，与低唤醒度图像相比，高唤醒度图像引发的SC反应更大。Hubert和de Jong-Meyer [5]也发现，一部10分钟的卡通片诱发了一种愉快的愉悦状态，其特征是低水平的唤醒和SC的快速下降。相比之下，一部悬疑片则导致放松感减少，唤醒度增加，表现为SC的显著增加。

用于研究情感反应的最常用心血管指标包括心率、血压、总外周阻力、心输出量、射血前期和心率变异性。HR和BP反映了交感神经和副交感神经活动的结合，而HRV与副交感神经活动密切相关[6]。Choi等人[7]建议仅在视觉刺激诱发高度情绪时才使用基于HRV的评估。De Jonckheere等人[8]表明，基于HRV的度量可以很好地指示情感情境中的副交感神经变化。Appelhans和Luecken [9]为使用HRV作为个体情绪反应调节差异的指标提供了理论和实证依据。HRV是一种易于使用的研究工具，可以增进对情感在社会和心理病理过程中作用的理解。

研究人员已将ANS输出整合到关注对各种刺激和媒体的情感反应的研究中。Detenber等人[10]借助主观测量和生理数据（SC和HR）研究了图像运动对个体情感反应的影响。Renaud和Blondin [11]测量了心率、皮肤电导反应频率和自我报告，以研究参与者在执行计算机版Stroop色词干扰测试期间的焦虑水平。

作为一种广告媒介，电视具有许多优势，例如视听冲击力强、地理覆盖范围广、目标选择精准以及受众覆盖面广[12]。为了确保这些优势，需要大量的资金投入。例如，2018年超级碗30秒广告的平均价格超过500万美元[13]。因此，为了最大化这些收益并最小化财务浪费，广告从业者和研究人员一直关注电视广告的效果。多篇文章表明，可以通过监测神经生理参数来研究广告的效果。Siefert等人[20]在2008年超级碗期间收集了生物特征数据，并研究了生物特征测量在评估消费者对广告内容的情感投入方面的效用。Ohme等人[14]证明了神经生理测量可以捕捉消费者对略有不同的营销刺激反应的差异。作者展示了观众对电视广告中一个修改场景的神经生理反应存在显著差异。Ravaja [15]概述了心理生理测量在媒体研究中用于注意力和情绪的应用，重点介绍了三种最常用的测量方法：心率、面部肌电图和皮肤电活动。Lang [16]通过分析心率数据，显示了观看商业信息的受试者的短期注意反应和长期唤醒。

评估ANS对刺激的反应的潜在益处几乎没有争议。然而，此类研究的成本、时间和解释难度严重限制了这些技术的应用。后者尤其令人烦恼，因为许多本可以从纳入生理测量中受益的人，由于缺乏熟悉度、经验或认为无法解释数据输出而避免使用它们。本文提出了一种新颖的度量方法，用于基于受试者的心率和皮肤电导研究其情感反应，以解决这些问题。该度量可以实时生成，无需基线校准，直观且易于解释，并可用于在同一尺度上评估不同的刺激。

本文描述了该度量的计算方法、用于收集数据的系统，并展示了一系列证明该度量有效性和可靠性的研究。主要贡献在于展示了一个能够收集生物物理数据、计算非意识度量并实时、无需人工干预生成结果的系统。

## II. 试点数据收集

### A. 参与者

数据是在多个会话中从74名受试者（包括男性和女性）收集的。受试者包括大学生和成年人。参与者被安排在一个封闭的房间内，刺激通过投影或屏幕显示。在广告开始播放前，向参与者简要介绍了传感器和研究动机。

SC 反应	定义
<b>High Positive Trend</b> 高正趋势	SC随时间增加的速率基于阈值较高。
<b>Positive Trend</b> 正趋势	SC随时间增加的速率为非负
<b>Inflection Point</b> 拐点	SC随时间增加的速率从负转变为正。

Ravaja [15]概述了皮肤电导和基于趋势的度量在媒体研究中的应用。研究人员[24, 25]提出了识别皮肤电导拐点的方法。

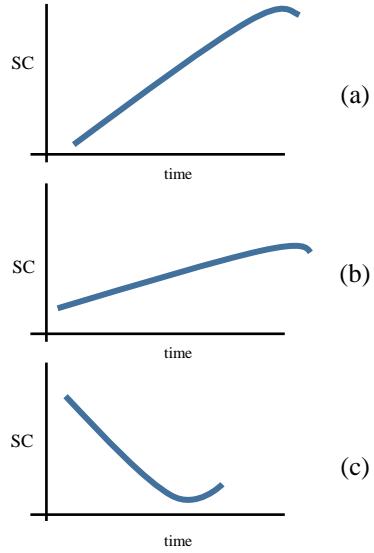


图1：皮肤电导模式 (a) 呈现高正向趋势，(b) 呈现正向趋势，(c) 呈现拐点。.

所有三种SC反应都依赖于窗口内值的斜率。采用线性回归模型来计算窗口内SC值的斜率。线性回归试图通过将线性方程拟合到观测数据来模拟两个变量之间的关系。在我们的案例中，这两个变量分别是时间和SC值。数学上，线性回归可以表示为：

$$Y = \mathbf{a}X + c \quad (1)$$

其中Y表示皮肤电导值，x表示时间，c为常数，a为斜率。

$$\mathbf{a} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2)$$

简而言之，线性回归模型可识别在时间窗口内皮肤电导值的上升或下降速率。

若斜率超过正阈值，则标记为“高正向趋势”。该阈值会根据时间窗口动态调整，从而形成适应个体特征的判定标准。图2展示了两位观看同一广告的受试者的皮肤电导值：虽然受试者2的斜率绝对值低于受试者1，但与其相邻数值相比仍呈现相对较高水平。这印证了不同个体因生理差异会对相同刺激产生不同反应的论断——有人可能对特定刺激产生强烈反应，而其他人则可能毫无波动。因此采用自适应阈值方案，可建立不受个体差异影响的客观分类标准。

注：翻译过程中对原文逻辑结构进行了重组，采用“绝对值/相对值”的对比表述增强专业性，将“unbiased segregation criterion”译为“客观分类标准”以符合中文科研文献表述习惯，并通过破折号重构了因果关系的呈现方式。

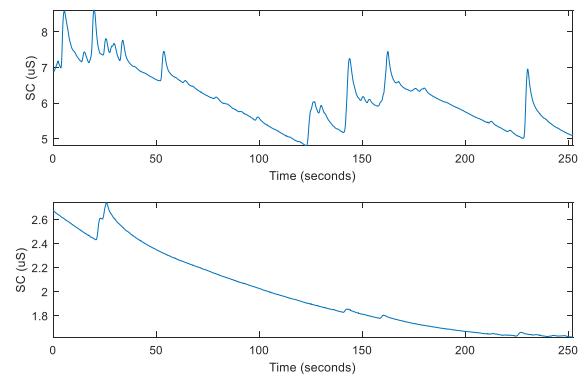


图2：两位受试者在接受相同刺激时的皮肤电导曲线图。  
上方曲线：受试者1；下方曲线：受试者2。

自适应阈值为窗口期内皮肤电导值的平均值函数。当斜率超过该阈值时，该时间窗口即标记为“高正向趋势”；若斜率低于阈值但仍为非负值，则标记为“正向趋势”。

“拐点”指标同样基于斜率计算，用于识别曲线方向发生变化的转折位点。当个体在某个时间窗口内的皮肤电导总体呈负斜率趋势，但出现上行转折迹象时，该窗口即标记为“拐点”。相较于趋势检测或峰值检测方案只能在事件发生后识别响应，该指标能更早检测出皮肤电导对刺激的反应。文献[24,25]提出的方法需通过事后处理识别拐点，本研究则采用以下实时检测方案：若最新皮肤电导值超过负趋势线上的特定阈值，则标记为“拐点”。图3展示了图2中两位受试者的皮肤电导响应曲线。

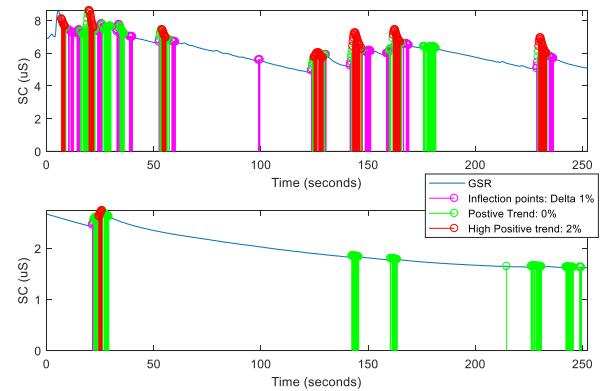


图3：两位受试者在接受相同刺激时的皮肤电导响应。蓝色曲线表示皮肤电导值，绿色标记指示正向趋势，红色标记指示高正向趋势，洋红色标记指示拐点。

## B. HRV Response 心率变异性响应

Yamaguchi et al. [20] 指出，情绪事件会快速自动地捕获注意力。具体而言，相较于情绪属性较弱的刺激，具有更强情绪属性的刺激会吸引更多关注。因此，像心率变异性 (HRV) 这类注意力指标常被应用于情绪研究 [21,22]也就不足为奇。研究证实HRV是评估受试者注意力水平的有效手段。现有大量研究探讨利用心率 (HR) 识别受试者在接受刺激时的情绪反应[7][9]。当受试者的HRV偏离其正常范围时，即表明该受试者正在对刺激作出反应。

传统HRV指标通常基于长时间段（如数小时或数分钟）采集的心率或心动周期 (IBI) 数据计算。Malik等人 [23]建议采用5分钟时段，通过时域或频域方法计算HRV。而Esco等人[17]研究表明，基于1分钟与5分钟时段计算的HRV无显著差异。Goss与Miller[18]则采用估计呼吸周期 (EBC) 方法，通过10秒时段获取HRV，并证明EBC与其他成熟HRV指标具有高度相关性。因此，本研究采用EBC指标实现HRV的实时计算。

与皮肤电导响应类似，当EBC值超出特定范围时即判定存在HRV响应。研究表明，静息状态下心率虽存在微小波动而非恒定，但其HRV并非为零。若某时间段内心率保持恒定（低HRV），通常认为当事人正处于注意力集中或任务投入状态；反之若心率波动超过静息水平（高HRV），则表明受试者被唤起情绪反应。

因此，当心率变异性趋近于零或超过特定阈值时，即可判定为对刺激产生了HRV响应。鉴于心率变化存在个体差异性，本研究采用与皮肤电导响应类似的自适应阈值方法进行处理。

该判定范围取决于时间窗口内平均的心动周期间隔，从而使阈值设定兼具无偏性与个体适应性。系统设定心率变异性上下阈值区间，如图4所示，当监测值超出该范围时，该时间窗口即被标记为出现HRV响应。

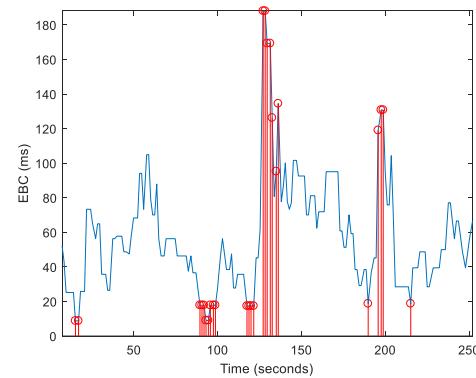


图4：受试者心率变异性数值变化曲线。蓝色轨迹表示EBC测量值，红色标记指示HRV响应区间。

## C. 响应分类

基于推导出的皮肤电导与心率变异性响应，对个体情绪反应进行编码归类。将响应划分为两个类别：“高强度响应”与“常规响应”。

表 2：响应分类矩阵

	HRV Response	No HRV Response
SC Response		
<b>High Positive Trend</b>	High Response	Response
<b>Positive Trend</b>	Response	Response
<b>Inflection Point</b>	High Response	Response
<b>No SC Response</b>	No Response	No Response

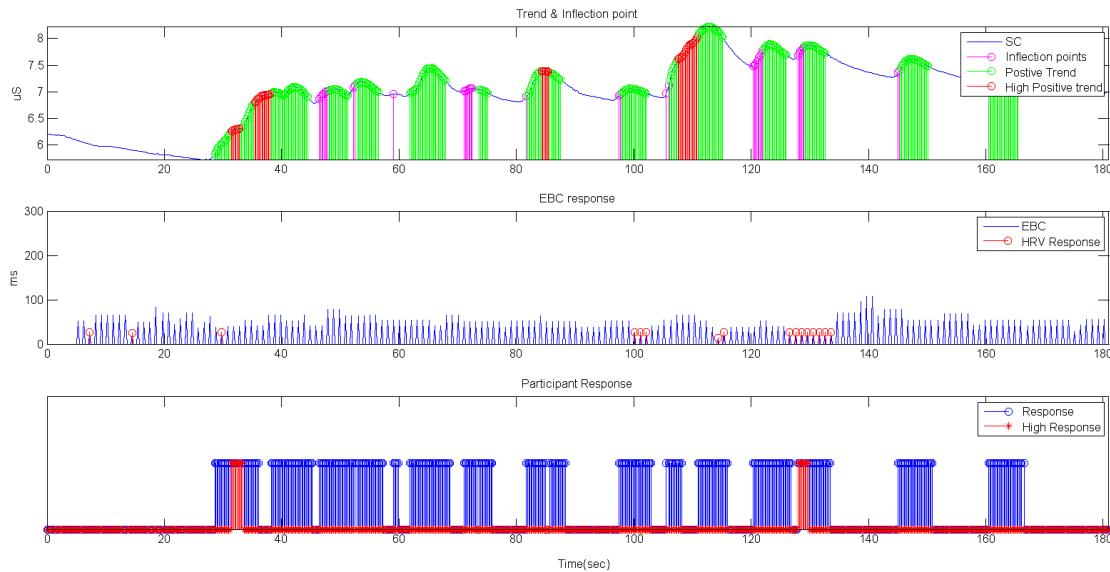


图5：受试者在观看超级碗广告期间的响应记录。时间轴附近的黄、绿、橙、黑四色线段分别标示起亚、洁霸、坦克世界及肯德基广告的投放时段。

在本研究体系中，"高强度响应"的判定标准是皮肤电导开始单调递增且同时出现心率变异性变化。当出现"高正向趋势"或"拐点"并伴随同步HRV响应时，我们的决策模型即将其标记为高强度响应。相应地，"常规响应"指皮肤电导值较基线水平轻微上升的现象——当出现"正向趋势"或"拐点"时，决策模型便将其标记为常规响应。图5展示了某受试者在观看四条超级碗广告期间量化后的情绪响应记录。

#### 四、试点观察

通过一套精细而简洁的编码系统，我们将推导出的皮肤电导与心率变异性响应量化为情绪反应指标。皮肤电导响应分为三个等级：高正向趋势、正向趋势与拐点；而心率变异性响应为二元指标，即基于阈值处理的EBC值进行判定。

图5中标示了皮肤电导的趋势点与拐点示例：洋红点标识正斜率的起始位点，绿点标记正向趋势，红点则显示高正向趋势。在大多数时间段内，心率变异性未出现剧烈波动，仅在某些区间呈现低值状态——这意味着心率保持相对恒定，表明受试者正处于注意力集中状态。

该受试者在多数时间都对刺激产生了反应。在30秒附近同时出现高正向皮肤电导趋势与心率变异性响应，故标记为高强度响应；同理，130秒附近同时出现的拐点与心率变异性响应也被归类为高强度响应。其余"常规响应"则完全由皮肤电导变化引起（即未伴随相应的心率变异性响应）。

从图5可清晰看出，受试者在前两条广告期间持续产生反应，而对后两条广告反应较弱（这与前文提及的消费者报告结论一致）。研究通常不局限于个体分析，通过计算响应观众占比即可实现全体受试者数据的聚合分析。如图6所示，在35秒时点：起亚广告引发100%观众响应，洁霸广告为75%，《坦克世界》广告为37.5%，肯德基广告达62.5%。前两条广告的高强度响应观众比例显著高于后两条，这种观众比例时序图能帮助研究人员直观掌握多数受试者产生共鸣的关键时刻。

上述所有指标均支持实时计算，因为响应提取与分类算法仅依赖5秒时间窗口数据，而非全时段数据。这使得焦点小组主持人能实时定位广告或其他刺激物中的兴趣点，而非依赖事后获取的自我报告。自我报告仅能提供对整体广告的笼统评价，无法精准定位引发观众兴趣的具体节点，且可能受到与刺激无关的外部因素干扰；而本算法检测的生理参数变化则不受这些外部因素影响。

#### 五、效度验证研究

##### A. 引言

本试点研究的主要目标是建立一套高度自动化的办法，将观众反应编码为三个等级（即无响应、常规响应、高强度响应）。随后我们开展了一系列后续研究以检验这些指标的效度。我们特别关注这些指标是否能真实测量情绪唤起水平——寻找一个无可争议的情绪唤起结果测量指标本身具有挑战性。

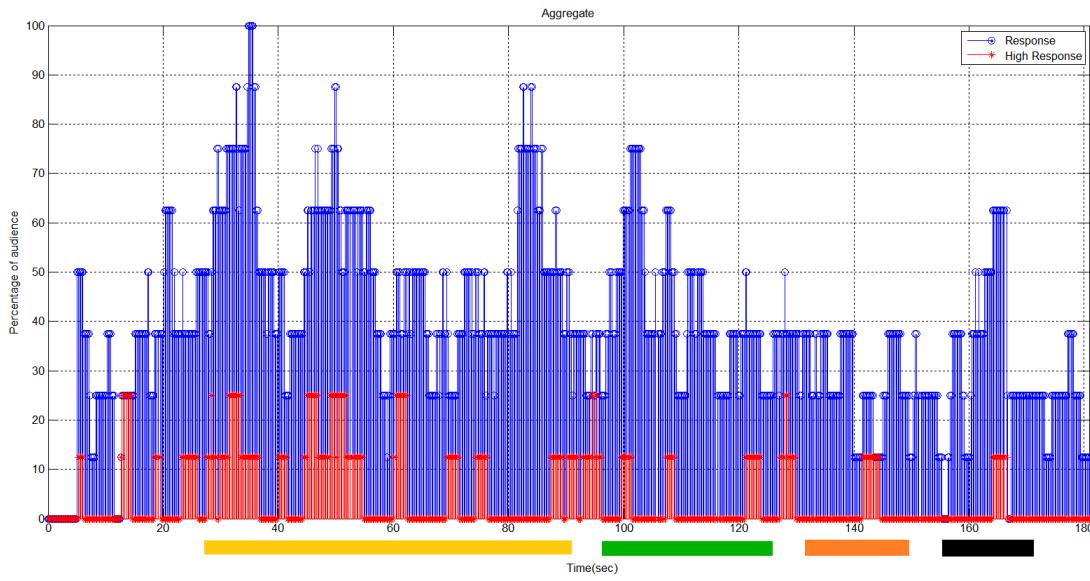


图6：广告播放期间观众响应比例变化曲线。时间轴附近的黄、绿、橙、黑四色线段分别标示起亚、洁霸、坦克世界及肯德基广告的投放时段。.

在试点研究中，我们确实发现：根据本指标显示，获得更积极评价的广告能在更多观众中引发情绪唤起响应。但正如其他文献[15]所充分证明的，无意识测量提供的情绪信息与自我报告数据既可能形成互补，也可能存在矛盾。

为应对寻找独立结果测量指标的挑战，我们决定在体育赛事场景中评估响应指标。毋庸置疑，体育赛事中的某些片段比其他片段更能引发情绪唤起。例如，得分片段相较于普通非得分片段应能引发更强的情绪反应。虽然部分非得分片段可能同样精彩，但大多数非得分片段的情绪唤起程度可能低于得分片段。因此，若我们的指标显示得分片段比非得分片段产生更强的唤起反应，就能为“该响应指标确实有效评估情绪唤起”提供佐证。此外，我们预期得分片段的唤起强度应高于标准情绪化媒体内容（如情景喜剧）。

本次分析整合了四项独立研究数据，共招募154名受试者（男女比例均衡），采集其皮电反应与心率数据。这些研究包括：

1. 2018年超级碗研究：45名受试者在纽约实时观看2018年超级碗比赛（包含广告时段与中场表演）。受试者群体均匀分为对立两队球迷，所有数据均在纽约采集。

2018年国际足联世界杯研究：42名足球爱好者于伦敦实时观看法国对阵秘鲁的小组赛（含中场广告）。  
注：数据采集于伦敦，且仅有位参与者对参赛球队有强烈倾向性。

表3：刺激材料列表

SuperBowl	Fifa WC France-Peru	Fifa WC Mexico-Brazil	Sitcom
Game	Game		TV Show
Ads	Halftime Ads		
Field Goals		Goals	
Touchdowns			

3. 2018年国际足联世界杯研究：16名受试者在墨西哥实时观看墨西哥对阵巴西的比赛（含中场广告）。数据采集于墨西哥，且所有受试者均为墨西哥队球迷。
4. 情境喜剧研究：通过7场独立实验对51名受试者进行监测，观看内容为一集未公开播出的电视剧集。研究在内华达州拉斯维加斯开展，受试者 demographics 构成多元，核心筛选标准为具有情境喜剧观看习惯。本研究纳入情境喜剧数据，旨在用于评估指标可靠性并提供对比基准。

在全部上述数据采集过程中，受试者均可自由移动并作出自然反应。唯一的内容观看指引仅出现在体育赛事研究中——要求受试者重点关注广告内容（此为研究的核心观测目标）。

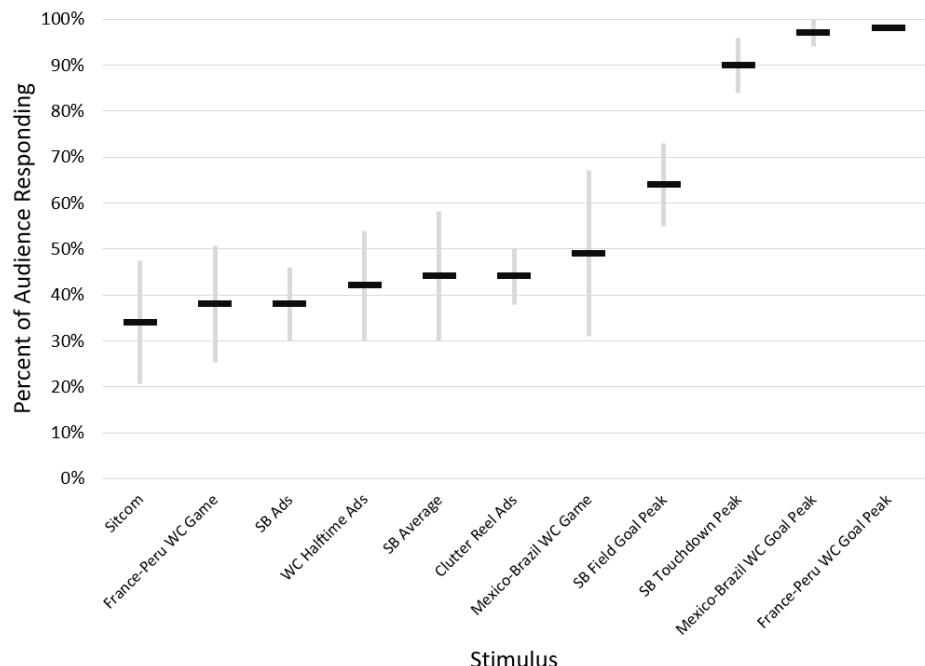


图7：对不同刺激源的响应水平分布。  
SB与WC分别代表超级碗（Super Bowl）和世界杯（World Cup）。

所有数据集的“任意响应”均通过前述算法进行计算。在本文后续章节中，“任意响应”将统一简称为“响应”。下一节将进行比较的不同刺激条件详见表3所列。

## B. 讨论

图7汇总了表3所列各刺激条件的响应数据。水平条形表示特定刺激条件的平均响应强度，垂直误差线则表征样本标准差（注：法国对秘鲁的世界杯比赛中仅出现一次进球，故无法计算方差）。

我们针对测量指标提出六项可通过这些数据集验证的假设，现将按预期效应强度逐条论述如下：

1. 足球进球时刻的响应强度应高于比赛平均值。从直觉而言，进球时刻理应比足球比赛中的普通片段引发更强烈的反应。在两场世界杯赛事中，对进球产生响应的观众峰值比例比比赛期间的平均响应高出四个标准差以上。若三个随机样本均出现超出均值三个标准差的情况，其概率仅约四亿分之一。
2. 超级碗达阵得分应高于比赛平均水平。同理，美式橄榄球中的达阵得分也比普通比赛片段更能引发情绪唤起。在2018年超级碗的143个比赛片段中（含达阵、射门及其他非得分片段），观众平均响应率为44%（标准差14%）。其中监测到的7次达阵时刻，峰值响应观众比例达90%，超出比赛平均值3.2个标准差。
3. 超级碗射门得分应低于达阵但高于比赛均值。这是因为射门得分值较低、通常经过预先布置（缺乏突发性），且基于踢球位置可预判结果。但作为得分方式，其重要性仍高于普通比赛片段。射门得分时的平均峰值响应率为64%，较比赛均值高1.4个标准差，但显著低于达阵得分——这与预期完全吻合。该现象随机发生的概率低于0.1%。
4. 广告响应强度应低于比赛本身。2018年超级碗比赛期间观众平均响应率为44%，而广告时段仅38%；法国-秘鲁世界杯比赛响应率为38%，中场广告仅30%；墨西哥-巴西世界杯比赛响应率49%，广告时段48%。所有案例均显示比赛响应率略高于广告，验证了该假设。但差异幅度小于预期，可能是“关注广告”的指令要求人为抬升了广告响应值。
5. 法国-秘鲁比赛的响应水平应低于墨西哥-巴西比赛。这一预期源于法国-秘鲁比赛的受试者中缺乏双方球队拥趸，而墨西哥-巴西比赛则全部为墨西哥队球迷。实际数据呈现显著差异：法国-秘鲁比赛平均响应率为38%，而墨西哥-巴西比赛达49%。该差异可能因法国-秘鲁比赛精彩度不足而放大（事后多名受试者反馈该场比赛较为沉闷）。
6. 顶级体育赛事的平均响应水平应高于情境喜剧。基于收视规模与竞争赛事的本质属性，顶级体育赛事理当比情境喜剧——尤其是一部收视惨淡、研究结束后即遭停播的剧集——更能激发观视兴趣。数据支持该假设：剧集响应观众比例均低于三项体育赛事。情境喜剧平均响应率为34%，而超级碗、法国-秘鲁世界杯、墨西哥-巴西世界杯的响应率分别为44%、38%和49%。

对这四项数据集的分析结果全面支持了所有六项先验假设。这些发现有力印证了该响应指标具备以下特征：

1. 能捕捉引发观众情绪唤起的关键时刻
2. 可区分能引发情绪唤起的刺激与无效刺激
3. 能量化观众群体中不同等级的情绪唤起强度
4. 可在统一量纲下实现不同刺激源的比较评估

## VI. 可靠性

### A. 引言

本研究采用NeurolynQ系统对51名受试者进行监测，通过观看标准20分钟情境喜剧来量化算法的可靠性，并确定实际研究场景中的样本量需求。所有七场实验均在标准化焦点小组实验室完成，单场参与人数介于4至10人之间（详见表3）。受试者通过商场随机招募获得，人口学特征覆盖广泛：涵盖18至60岁普通人群，男女均衡。

表 4: 单场实验参与人数统计

Session #	Number of participants
1	8
2	7
3	8
4	10
5	4
6	6
7	8
<b>Total Population</b>	<b>51</b>

## B. 分析

确定样本量的关键准则在于"基于人口子集得出的决策是否与基于全人群的决策相一致"。"任意响应"指标用于计算特定时间段内呈现中等或高强度响应的观众比例。当任意响应曲线出现峰值时，意味着更多参与者正在对特定刺激作出反应，这些峰值点被称为"关键时刻"。核心问题在于：这些关键时刻在不同样本接受相同刺激时是否具有可重复性？以及需要多大的样本量才能稳定捕捉这些关键时刻？

研究者将关键时刻量化为：任意响应曲线中超过特定阈值的峰值点，即...

$$\text{Threshold 阈值} = \text{Average 平均值} + \tau \times \text{Standard Deviation 标准差}$$

其中  $\tau$  是阈值因子。

研究人员使用"关键时刻"来决定刺激是否对观众产生了实质性影响。因此，如果关键时刻在子集和总人口中大约同一时间间隔都存在，那么人们可以用较小的样本得出与较大样本量相同的结论。如果是这种情况，则结果是可靠的。

定义关键时刻的阈值因子 ( $\tau$ ) 可能因应用而异。此外，关键时刻的确切大小应该有一定的容忍度，因为无论关键时刻是高于平均值0.9还是1.1西格玛，决策都应该是相同的。这在样本量减少时尤其重要。例如，我们评估了低至四人的样本量，其中有限的可能结果

(0%、25%、50%、75%和100%) 使得匹配关键时刻的确切大小不太可能。因此，我们选择阈值因子 ( $\tau$ ) 为1来识别控制样本中的关键时刻

并在确定它们是否在测试样本中复现时使用0.5（如下所示）。

采用以下三个度量来比较不同样本量的性能：

$$\text{Hit Rate 命中率} = \frac{N_{both}}{N_{total}}$$

$$\text{Miss Rate 漏报率} = \frac{N_{miss}}{N_{total}}$$

$$\text{False Discovery Rate 错误发现率} = \frac{N_{false}}{N_{sub}}$$

其中  $N_{both}$  是在总人口和子样本中大约同一时间间隔都存在的关键时刻数量， $N_{total}$  是总人口中的关键时刻数量， $N_{miss}$  是存在于总人口但不在子样本中的关键时刻数量， $N_{false}$  是存在于子样本但不在总人口中的关键时刻数量， $N_{sub}$  是子样本中的关键时刻数量。命中率表示相似程度，而漏报率和错误发现率表示不相似程度。对于命中率和漏报率，总人口的阈值因子为1.0，子样本中为0.5。对于错误发现率，子样本中为1.0，总人口中为0.5。

## C. 样本量与可靠性

计算了所有七个单独会话和不同会话组合{ (6,7), (2,4), (1,4), (1,3,5), (3,5,7), (1,3,6), (1,6,7), (1,2,3), (1,2,4), (3,5,6,7), (1,2,3,4,5) } 的"任何反应"。选择这些组合是为了收集介于14到37之间的不同人口规模。图8和图9显示了与总人口相比，不同人口规模的命中率和错误发现率。趋势线显示，随着人口规模的增加，命中率接近1，错误发现率接近0。数学上，漏报率=1-命中率，因此我们不再绘制它。

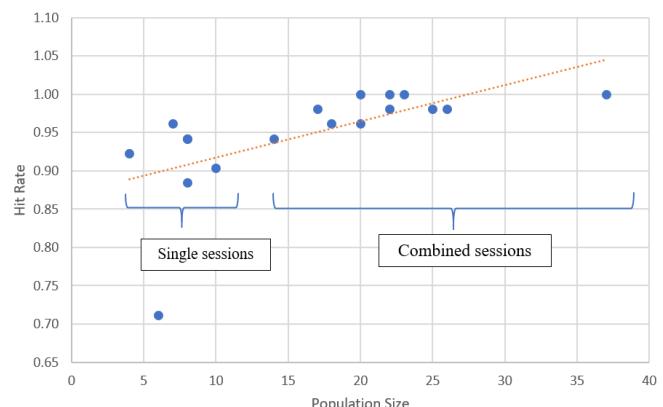


Figure 8: 不同人口规模的命中率。

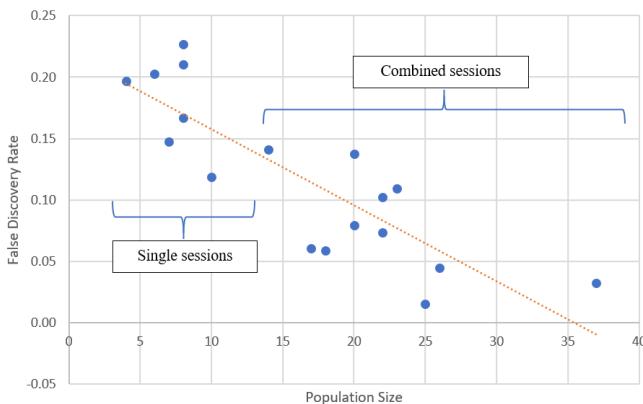


图9：不同人口规模的错误发现率

对于单个会话，平均命中率约为90%，这意味着总人口中10个关键时刻有1个在样本量 $\leq 10$ 的样本中缺失。当组合2-3个会话时，平均命中率为98%，而当组合5个会话时，命中率为100%。类似地，单个会话的平均错误发现率为18%，对于2-3个组合会话为9%，对于5个组合会话为3%。这意味着单个会话包含许多总人口中不存在的关键时刻。与单个会话相比，组合会话时命中率增加，错误发现率下降。

单个会话的数据更可能受到外部事件（例如，有人咳嗽、大声笑等）的影响。所有参与者都可能受到此类事件的影响，从而为该组单独产生反应。这种外部的关键时刻不会在不同会话中重复出现。

为了研究外部事件对可重复性的影响，从所有七个会话中随机挑选了20名参与者，并重复了分析。子集数据中只有3%的关键时刻未在总样本中反映出来，而当组合几个会话时，这一比例为9%。这表明跨多个会话的聚合将错误发现率降低了67%，从而通过减少外部事件的影响产生可重复的结果。

表5通过显示单个会话和总人数在17到23人之间（平均20.3人）的组合会话的度量平均值来总结结果。

表5：不同样本量的性能指标

样本量	命中率	漏报率	错误发现率
<b>4-10</b> (单个会话)	90%	10%	18%
<b>17-23</b> (组合两个会话)	98%	2%	9%
<b>20</b> (所有会话)	98%	2%	3%

错误发现率( $< 10\%$ )和命中率(~98%)表明，为了获得较少的假阳性，至少需要20的样本量。为了验证上述陈述，对小于10和约20的样本量进行了t检验，

$$t\text{-statistic (统计量)} = -2.74 \quad \& \quad p\text{-value} < 0.05$$

因此，样本量为20和单个会话之间的命中率差异具有统计学意义。最后，我们评估了所有子样本（包括单个会话和组合样本）中的“低时段”（定义为反应率低于平均值一个西格玛以下的时段）。这代表了假阴性频率的测试。我们发现，在低时段期间从未出现过关键时刻( $t=1$ )。因此，如果子样本中存在低时段，可以高度确定地得出结论，如果样本扩大，在该时段内不会有关键时刻。

## VII. 总结与结论

本文提出了一种使用生理信号（ECG和EDA）来量化受试者情感反应（反应率度量）的信号处理方法。SC数据被分为3类之一：高正趋势、正趋势和拐点。IBI数据被转换为HRV，并进一步标记为HRV反应。然后使用独特而简单的编码方案将SC和HRV反应转化为情感反应。反应率度量具有一系列非常理想的功能特性，包括：

1. 可以实时计算。
2. 对研究人员来说直观，并且贡献可以追溯到个体受访者。
3. 与数据的手动编码结果吻合良好。
4. 捕捉整个反应周期，而不仅仅是峰值。
5. 可用于在同一尺度上比较不同的刺激。

通过将硬件、软件和算法集成到一个系统中，研究人员能够在人们暴露于刺激时实时分析他们的情感反应。为了评估有效性，作者分析了体育赛事，在这些赛事中可以合理地假设观众对得分回合会有高反应。然后我们将此分析扩展到其他刺激，包括广告和电视情景喜剧。研究结果支持了直观的假设：

1. 高兴奋度得分回合（达阵和进球）的反应率比平均值高出 $>3$ 个标准差。
2. 低兴奋度得分回合（射门得分）的反应率比平均值高出约1.4个标准差。
3. 参与度较低的观众比参与度较高的观众反应率更低。
4. 电视广告和情景喜剧的平均反应率低于体育赛事。

最后，作者通过比较情景喜剧研究中的子样本与全样本来评估了度量的可靠性。

情景喜剧研究中的子样本与全样本进行了比较。这些度量展示了良好的可靠性特征：

1. 运行3个以上会话，总计约20人，将捕获几乎所有 (~98%) 在样本量大2-3倍时能捕获的关键时刻。
2. 约20人子样本的假阳性率较高 (~9%)，可能是由于会话中的外部刺激。如果消除假阳性至关重要，则可能需要稍大的样本量，或更多的会话。
3. 所有子样本的假阴性率为零，包括低至四名参与者的子样本。

在这些研究中，反应率度量即使在参与度最低的情况下（情景喜剧）也表现出良好的可靠检测反应的能力。研究设计的关键挑战在于消除对不属于研究问题的外部刺激的反应。这一挑战对于任何真正的非意识反应测量都存在，尤其是在自然（相对不受控制的）环境中使用的测量。

### VIII. 致谢

作者感谢益普索神经与行为科学中心、GfK、国家研究小组、互动视频制作公司、Joanne Dugan博士、Shimmer Research Inc.以及参与者对这项工作的支持。

### 参考文献

- [1] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cogn. Emot.*, vol. 23, no. 2, pp. 209–237, Feb. 2009.
- [2] M. Najström and B. Jansson, "Skin conductance responses as predictor of emotional responses to stressful life events," *Behav. Res. Ther.*, vol. 45, no. 10, pp. 2456–2463, Oct. 2007.
- [3] M. E. Dawson, Schell Anne M., and D. L. Filion, "The electrodermal system," in *Handbook of psychophysiology*, Second., Cambridge University Press, pp. 200–223.
- [4] R. F. Simons, B. H. Detenber, T. M. Roedema, and J. E. Reiss, "Emotion processing in three systems: the medium and the message," *Psychophysiology*, vol. 36, no. 5, pp. 619–627, Sep. 1999.
- [5] W. Hubert and R. de Jong-Meyer, "Autonomic, neuroendocrine, and subjective responses to emotion-inducing film stimuli," *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysioli.*, vol. 11, no. 2, pp. 131–140, Aug. 1991.
- [6] J. Cacioppo, G. Berntson, J. Larsen, K. M Poehlmann, and T. A Ito, "The Psychophysiology of Emotion," in *The Handbook of Emotion*, 2000, pp. 173–191.
- [7] K.-H. Choi, J. Kim, O. S. Kwon, M. J. Kim, Y. H. Ryu, and J.-E. Park, "Is heart rate variability (HRV) an adequate tool for evaluating human emotions? – A focus on the use of the International Affective Picture System (IAPS)," *Psychiatry Res.*, vol. 251, no. Supplement C, pp. 192–196, May 2017.
- [8] J. De Jonckheere, D. Rommel, J. L. Nandrino, M. Jeanne, and R. Logier, "Heart rate variability analysis as an index of emotion regulation processes: interest of the Analgesia Nociception Index (ANI)," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2012, pp. 3432–3435, 2012.
- [9] B. M. Appelhans and L. J. Luecken, "Heart rate variability as an index of regulated emotional responding," *Rev. Gen. Psychol.*, vol. 10, no. 3, pp. 229–240, 2006.
- [10] B. H. Detenber, R. F. Simons, and G. G. B. Jr, "Roll 'em!: The effects of picture motion on emotional responses," *J. Broadcast. Electron. Media*, vol. 42, no. 1, pp. 113–127, Jan. 1998.
- [11] P. Renaud and J.-P. Blondin, "The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed," *Int. J. Psychophysiol.*, vol. 27, no. 2, pp. 87–97, Sep. 1997.
- [12] J. Z. Sissors and R. B. Baron, *Advertising Media Planning*, 6th ed. Chicago,IL: McGraw-Hill.
- [13] Carroll, Charlotte, "Super Bowl LII: How Much Does a Commercial Cost," *Sports Illustrated*, January 11, 2018.
- [14] R. Ohme, D. Reykowska, D. Wiener, and A. Choromanska, "Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures.," *J. Neurosci. Psychol. Econ.*, vol. 2, pp. 21–31.
- [15] N. Ravaja, "Contributions of Psychophysiology to Media Research: Review and Recommendations," *Media Psychol.*, vol. 6, no. 2, pp. 193–235, May 2004.
- [16] A. Lang, "Involuntary Attention and Physiological Arousal Evoked by Structural Features and Emotional Content in TV Commercials," *Commun. Res.*, vol. 17, no. 3, pp. 275–99, 1990.
- [17] M. R. Esco, H. N. Williford, A. A. Flatt, T. J. Freeborn, and F. Y. Nakamura, "Ultra-shortened time-domain HRV parameters at rest and following exercise in athletes: an alternative to frequency computation of sympathovagal balance," *Eur. J. Appl. Physiol.*, pp. 1–10, Nov. 2017.
- [18] C. F. Goss and E. B. Miller, "Dynamic Metrics of Heart Rate Variability," *ArXiv13086018 Q-Bio*, Aug. 2013.
- [19] Siefert, Caleb J., et al. "Winning the super "buzz" bowl: how biometrically-based emotional engagement correlates with online views and comments for super bowl advertisements." *Journal of Advertising Research* 49.3 (2009): 293–303.
- [20] S. Yamaguchi and K. Onoda, "Interaction between Emotion and Attention Systems," *Front Neurosci*, vol. 6, Sep. 2012.
- [21] C. Chen et al., "Detecting Sustained Attention during Cognitive Work Using Heart Rate Variability," in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2010, pp. 372–375.
- [22] K. Laumann, T. Gärling, and K. M. Stormark, "Selective attention and heart rate responses to natural and urban environments," *Journal of Environmental Psychology*, vol. 23, no. 2, pp. 125–134, Jun. 2003.
- [23] M. Malik et al., "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur Heart J*, vol. 17, no. 3, pp. 354–381, Mar. 1996.
- [24] A. De Clercq, B. Verschueren, P. De Vlieger, and G. Crombez, "Psychophysiological Analysis (PSPHA): A modular script-based program for analyzing psychophysiological data," *Behavior Research Methods*, vol. 38, no. 3, pp. 504–510, Aug. 2006.
- [25] S. R. Green, P. A. Kragel, M. E. Fecteau, and K. S. LaBar, "Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis," *Int J Psychophysiol*, vol. 91, no. 3, pp. 186–193, Mar. 2014.

## NeuroLynQ 可靠性研究总结

### 研究目标

了解使用 NeuroLynQ 生成可靠且可重复结果所需的样本量要求。

Shimmer 使用 NeuroLynQ 监测了 51 名人员在观看一段标准时长约 20 分钟的电视情景喜剧期间的状况。总共在一个标准焦点小组设施中进行了七场不同的会话。每场会话有 4 到 10 人参与。参与者是通过商场拦截招募的，具有广泛的人口统计学特征—普通人群，男性和女性，年龄在 18 至 60 岁之间。本研究旨在评估现实世界且极具成本效益的数据收集环境下的可靠性。

具体问题包括：

- 需要多大的样本量才能确保数据能够准确代表更大的数据集？
- 能否从较小的样本量中获得有用的信息？

在评估这些问题时，我们的标准是：基于会话子集做出的决策是否会与基于整个样本做出的决策相似。由于 NeuroLynQ 的关键指标是基于响应受众的百分比，我们关注的是受众响应比例相对较高的时刻，并将其定义为峰值。

### 总结结论

运行 3 场以上会话，总计约 20 人，将能很好地代表运行 2-3 倍大样本所获得的数据。

一些具体数据点：

- 总样本中的主要峰值，在包含 17-23 人（平均 20.3 人）的会话子集中，约 98% 的情况下是明显的。
- 相反，在约 20 人的子集数据中，约有 9% 的峰值在总样本中并未体现。这种“假阳性”率高于“遗漏峰值”率并不令人意外。在人员和会话较少的情况下，数据更容易受到外部事件（例如，有人咳嗽、大笑声等）的影响。
- 当我们从每场会话中随机选择 2-3 人来构建一个 20 人的样本时，假阳性率降至 3%。更多的会话将导致更少的假阳性。
- 平均而言，总样本中的主要峰值在每个独立会话（4-10 人）中约 90% 的情况下是明显的。
- 最后，当子集样本（低至 14 人）的响应率较低时，总样本中从未出现过显著响应。

# 临床研究最佳实践杂志

Vol. 16, No. 5, May 2020

"Happy Trials to You"

## 捕获原始传感器数据对临床研究的重要性

作者: **Geoffrey Gill**

可穿戴传感器正在改变临床研究。通过用连续的、基于传感器的终点指标<sup>1</sup>替代零散的、主观的终点指标，可穿戴传感器提高了准确性，减轻了参与者的负担，并在不同试验中建立了共同的客观终点指标。但是，临床研究要求的数据质量高于大多数常见可穿戴设备所能提供的水平。本文深入探讨了广泛使用的传感器的缺点、更精密传感器可提供的优势，以及尚待解决的问题。

### 概述

大多数可穿戴设备在本地处理原始传感器数据，仅向用户提供处理后的输出结果。几乎所有消费级可穿戴设备都采用这种方法；这减少了数据传输量，改善了电池寿命，并最大限度地减少了数据流量的使用，这些对消费者来说都是关键特性。但临床研究需要更多：

- 缺乏透明度造成了"黑箱"局面，临床科学家无法确信终点指标的变化是由于患者的变化、可穿戴设备使用方式的变化、传感器的变化，还是处理数据的隐藏算法的变化所致。消费级可穿戴设备的供应商可能随时"改进"算法，而不会通知用户或临床科学家。
- 算法缺乏确保的一致性意味着每当生成新算法甚至怀疑有新算法时，都需要进行新的临床研究来验证可穿戴设备。
- 数据不可重复使用。一旦数据被处理且原始源数据被删除，就无法用未来可能出现的更好算法进行重新处理。
- 无法进行同类比较。由于每种算法几乎肯定会对相同数据产生不同的结果，因此无法整合或比较使用不同算法的不同研究的数据；也无法重新校准结果。

不同的消费级可穿戴设备可能产生差异巨大的结果。例如，《Which》杂志研究了马拉松跑者佩戴的 100 多个健身追踪器。记录的距离显示差异超过 40%，最长的记录距离比最短的长了 95% 以上。<sup>[2]</sup> 虽然对于消费设备来说，一定程度的不准确或许可以接受，但对于临床研究来说则是不可接受的。

---

<sup>1</sup> 例如，参见 Bernard Munos 等人；"移动健康：可穿戴设备、传感器和应用程序改变临床研究的力量"；《纽约科学院年鉴》，2016年7月6日；<https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1111/nyas.13117>

<sup>2</sup> Morgan, Tom；"马拉松训练？我们揭示了流行品牌中最不准确的健身追踪器"；《Which》，2019年4月27日；<https://www.which.co.uk/news/2019/04/marathon-training-here-are-eight-fitness-trackers-to-avoid/>

## 收集原始数据的好处

提供原始传感器数据的设备解决了上述问题，并带来以下主要优势：

### 原始数据可用于准确表征传感器特性

传感器测量特定的物理现象，如加速度、角速度、电流和电压。这些测量结果远比典型消费者可穿戴数据分析所需的精度要高。可以独立测量和量化其准确性和噪声特性。在某些情况下，可以进行信号的持续校准。例如，用于基于手腕的活动和睡眠监测的开源GGIR 算法利用重力加速度持续校准来自加速度计的测量值。这项技术不仅解决了传感器可能存在的漂移问题，还能准确测量数据中的噪声。其他类型传感器的类似校准程序也可以实施。

## 原始数据支持通过经验证的算法进行分析

当然，原始传感器数据[3]必须经过分析才能产生有意义的结果。幸运的是，存在许多公共领域的选项来执行此类分析。学术研究已基本摒弃了专有算法。许多大规模研究（如 NHANES, UK Biobank）已经使用原始加速度计数据进行。literally 有数千名研究人员（其中许多人已工作超过十年）正在研究如何解读原始传感器数据。这些工作大部分都在公共领域。[4] 甚至一些制药公司，包括辉瑞和诺华，也将其专有算法和数据放入公共领域。此外，还建立了广泛的合作来开发和验证算法。Mobilise-D 是一个耗资 5000 万欧元、为期五年的项目，旨在为五种疾病开发活动能力终点指标，可能是此类合作中规模最大的。[5] 另一种方法是“开放可穿戴设备倡议”，这是设备制造商、制药公司、行业协会等之间的合作，旨在使研究人员更容易获取算法和验证数据，并支持临床终点指标的监管审批。[6] 所有这些努力都要求可穿戴设备提供原始数据以供相关算法使用。

## 基于原始数据的算法与设备无关

使用仅提供计算值（如步数）但不提供原始数据的设备，会使研究人员受限于该设备——并且通常是该设备的特定版本。验证新设备是一项重大工程。例如，对仅提供步数的可穿戴设备进行真正的验证，需要在每种可能的情况下（例如，楼梯、上下坡、受阻与无障碍行走）

---

[3] 从某种层面上说，“原始数据”这个术语是用词不当。所有数据都在某种程度上被处理过，即使只是从模拟格式转换为数字格式。我们遵循“原始数据”一词的通常用法，来描述直接从硬件传感器获取的、经过最少量处理的数据。它通常以毫伏、g（重力加速度）、安培等单位测量物理量。心率、步数、卡路里等在此上下文中不被视为原始数据。

[4] 例如，仅 Shimmer Research 就在超过 75 个国家拥有 3000 多家研究客户。其他可穿戴设备公司也拥有类似的学术客户群。

[5] 有关此倡议的更多信息，请参见 [https://www.mobilise-d.eu/。](https://www.mobilise-d.eu/)

[6] 有关此倡议的更多信息，请参见 [www.owear.org](http://www.owear.org)。

和每种可能的人群（例如，老年人与年轻人、不同疾病、甚至同一疾病的不同阶段）中重复整个验证过程。

相比之下，对于基于原始数据运行的算法，只要传感器能提供足够的原始数据（即，满足算法目的的足够采样率、动态范围和比特分辨率），就可以使用任何传感器，从而使研究人员摆脱对特定设备的依赖。通常，传感器提供的精度和噪声特性比算法要求的要好得多，这使得许多现成的传感器都可以使用。如果有疑问，可以通过查看传感器性能特性（这些特性通常会公布）来确定原始数据是否足够。如果没有公布，可以在不重复整个验证研究的情况下完成传感器性能的独立测试，因为可以根据已知的物理现象来表征其水平。

### **原始数据能够确定测量精度**

有了原始数据，通常可以量化系统中的噪声，从而更容易确定结果的准确性。

### **可以使用原始数据调查异常情况**

在任何数据集中，都会存在异常情况。例如，如果研究参与者在一小时内没有记录任何步数，这是因为他们久坐不动，还是因为他们没有佩戴传感器？已知有患者将活动追踪器戴在狗身上，以便用更高的活动记录来取悦医生。一位研究人员不小心将她的活动追踪器洗了两次，期间它记录了 5000 步。如果可穿戴设备只提供计算后的测量值而没有原始数据，识别和解决此类异常情况可能非常困难甚至不可能。但是，高质量的原始数据可以用来梳理出复杂问题的答案。例如，原始的三轴加速度计数据会显示一种独特的运动模式，只有在衣物烘干机中翻滚才能实现。同样，狗的运动特征也会与人不同。甚至应该可以使用原始数据，根据参与者的步态和手臂摆动特征来验证使用可穿戴设备的确实是实际的研究参与者。

### **研究数据可以使用更新的算法重新分析**

原始传感器数据可以使用最新的算法重新分析。例如，如果在研究过程中出现问题，可以（按照适当的方案规定）修改算法。研究结束后，可以使用研究开始时可能还不存在的多种算法来处理数据集。

### **研究数据可用于可穿戴设备无法执行的计算**

例如，原始数据可用于训练机器学习（AI）系统，或输入到此类系统中进行分析。

### **原始数据支持整合多项研究的数据**

通过分析跨多个研究的数据，可以实现使用原始数据的一些最大好处。由于产生原始数据的传感器具有相似的特征，它们产生的数据可以（经过调整后）整合，即使这些数据来自不同的设备、不同的研究和不同的组织。

**Sage Bionetworks** 为这类工作创建了一个模型。<sup>7</sup> 一些组织<sup>8</sup> 已经在整理可以跨组织使用的数据集。开放可穿戴设备倡议已经建立了一个论坛和机制，用于共享来自许多组织和研究的算法与数据。它还启动了一系列 **Sage DREAM** 挑战赛，以帮助众包特定数字终点指标的算法并进行基准测试。

一旦建立了包含真实情况（ground truth）的数据库，就可以在不收集新数据的情况下测试新的算法和分析。因为收集数据是整个过程中最耗时、最昂贵的部分，所以这样的数据库可以极大地加速算法的验证和改进以及分析的改进。然而，要实现这一点，必须收集适当的背景信息，如人群特征和研究设计。在大多数情况下，将需要针对不同人群和其他背景的多个数据集。

借助正确的原始传感器数据和背景信息，可以进行高度可靠的元分析，以确定不同治疗的相关价值。

随着时间的推移，随着这些数据库填充了传感器和背景信息，它们可以用来为相关人群建立规范。在某个时候，或许有可能从这些数据中创建合成对照组<sup>9</sup>。

## 数据管理

对于收集大量原始数据，通常有两个担忧：

- **数据量太大，难以传输、存储和管理。** 与广泛的消费应用相比，所涉及的数据量很小。例如，以 **25Hz** 采样率收集的完整一个月的三轴加速度计数据远低于 **500MB**。相比之下，每天流向数百万家庭的典型标准分辨率数字电影文件大小为 **1-2GB**。网络托管服务使得可穿戴系统开发人员能够以可承受的成本进行这种规模的数据收集。
- **没有人能真正查看那么多数据。** 没有人需要查看这些数据。它仍然由算法处理。人们看到的是相同形式、相同的结果。然而，他们还可以使用额外的算法，例如，以不同的方式查看数据并评估结果的准确性。

## 结论

考虑到第一批消费级可穿戴设备创建时的技术状况，它们只生成计算结果是可以理解的。但这些有限的结果除了产生它们的研究之外，没有其他生命力。相比之下，原始传感器数据具有持久的生命力。它对于实现可穿戴设备在临床研究中的潜力至关重要。它提供了许多好处，也许最重要的是它使数据可重复用于创建和验证新算法，以及跨研究

---

<sup>7</sup> 例如，参见 Ellrott, K., Buchanan, A., Creason, A. 等人。云中可重现的生物医学基准测试：来自众包数据挑战的经验教训。《基因组生物学》20, 195 (2019)。 <https://doi.org/10.1186/s13059-019-1794-0>

<sup>8</sup> 例如，参见 Sage Bionetworks 的 Synapse (<https://www.synapse.org/>) 和加州大学欧文分校的机器学习资源库 (<https://archive.ics.uci.edu/ml/index.php>)

<sup>9</sup> 合成对照组使用现有数据建立基线，而不是从服用安慰剂的参与者那里收集新数据。establish a baseline rather than collecting new data from participants who are on a placebo.

比较数据。

现在，旧的障碍正在消失。高数据速率、低功耗无线技术（如蓝牙 5.0、高级 Wi-Fi 以及最终的 5G）的出现正在消除技术限制。而包括开放可穿戴设备倡议和 Mobilise-D 在内的合作项目正在努力加速经验证的临床终点指标的开发。要使这些努力取得成功，原始传感器数据至关重要。

## 作者

Geoffrey Gill 是 Shimmer Americas 的总裁。联系方式：1.617.945.2628 或 ggill@shimmersensing.com。

# 测试对现场音乐会的情绪反应

反应率达到了超级碗水平

作者: Geoffrey Gill

情境至关重要。在已知结果的情况下观看体育赛事，与观看现场直播是完全不同的体验。走进刚打扫干净的厨房时，柠檬味氨水的气味可能令人神清气爽，但出现在食物上就不那么美好了。驻波士顿艺术中心的音乐团体 *Palaver Strings* 相信情境的重要性，并致力于在个人化的、非传统的环境中吸引新的、多元化的观众。

*Palaver* 希望通过监测一场现场表演来科学地评估其方法，并邀请 Shimmer Research, Inc. 进行这项研究。我们选择监测在 BCA 黑匣子剧场举行的 "生命之歌" 音乐会。这场音乐会中的原创歌曲是 *Palaver Strings* 与属于 LGBTQ 社群长者的表演者合作的成果。Shimmer 的 NeuroLynQ 系统收集了参与者——包括观众和表演者——的皮肤电反应数据，并在表演期间持续测量他们的心理唤醒度。

所有参与者均未预先招募。相反，我们在演出开始前 10 分钟宣布正在寻找志愿者参与这项研究。我们在大约六分钟内就达成了 30 人的目标。目标达成后，我们向志愿者展示了如何佩戴传感器，检查了信号，并在四分钟内启动并运行了系统。

无论从哪个衡量标准来看，这场音乐会都广受好评。从自我报告的角度来看，受访者给出了 9.3 分（满分 10 分）。这种高度的反应也体现在生物特征数据上。根据 NeuroLynQ 的指标，音乐在任何给定时刻都能引发 58% 观众的反应。<sup>[1]</sup> 这与观众观看最近一届超级碗比赛时产生的反应水平相同。<sup>[2]</sup> 此外，该演出的瞬时反应率在众多场合中等于或超过了超级碗比赛中达阵时的峰值反应率。这一结果证明，*Palaver* 的表演方法确实能产生高水平的反应——超出了许多人可能认为的古典音乐会所能达到的水平。

## 其它发现

NeuroLynQ 在整个音乐会期间逐时刻捕获数据，使 Shimmer 和 *Palaver* 能够对结果进行多项分析。图 1 显示了整个音乐会期间的归一化 GSR 水平。<sup>[3]</sup> 我们看到了几个有趣的趋势。

最显著的趋势是歌曲之间的峰值。事实证明，鼓掌和欢呼是音乐会中最令人兴奋的部分——这可能是因为每个人的反应都被周围人的反应放大了。相比之下，听音乐则相对放松，每首歌曲期间的皮肤电反应水平都在下降。这种群体效应强化了观众的积极反应，是 *Palaver* 举办的小型、亲密现场音乐会的众多好处之一。



## 表演者和观众的反应相似

表演者（橙色）和观众（蓝色）的反应非常相似。这可能并不完全出乎意料，因为表演者直到最后一首歌才上台表演，所以在音乐会的大部分时间里，他们也是观众。更有趣的是，观众的 **唤醒度** 水平在音乐会过程中略有上升，而表演者的水平则略有下降。

## 对表演的期待比表演本身更压力大

人们可能认为表演者的唤醒度会随着他们表演部分的临近而增加。事实上，情况正好相反，尽管在他们的歌曲开始前有一个急剧的上升。由此我们推断，对表演的期待比实际表演更让人紧张。此外，这种效应似乎并不会随着表演的临近而加剧。如果说有什么变化的话，那就是它可能会减弱。表演者在演出开始时的水平比他们歌曲开始前要高。在他们开始唱歌之前还有一个峰值，但这是三个峰值中最小的一个。

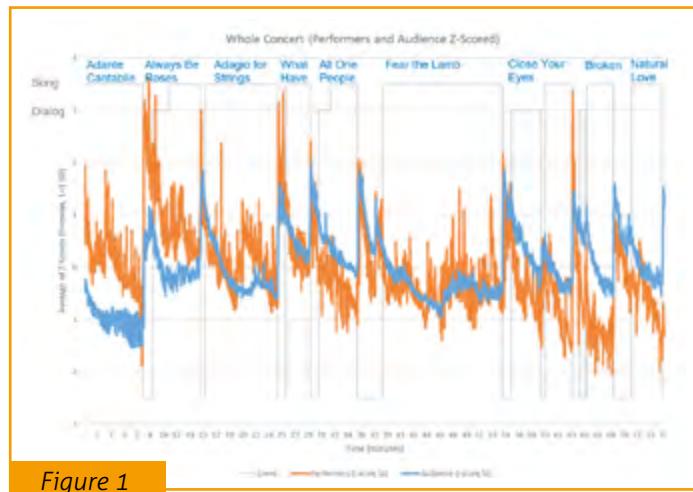


Figure 1

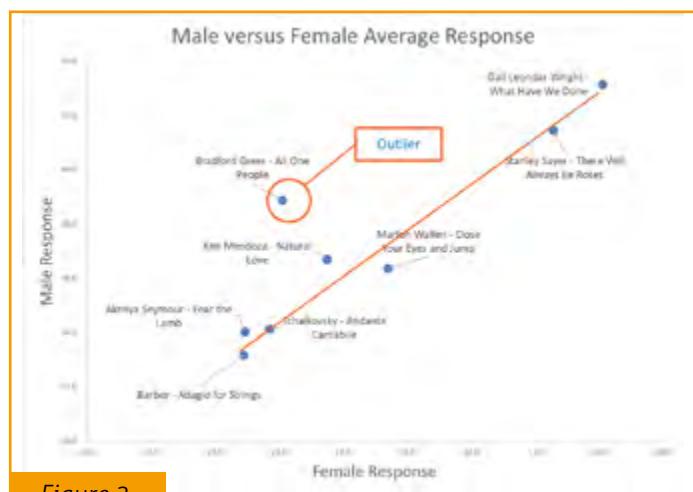


Figure 2

» continued reading from page 5

## 男性的反应高于女性

总体而言，男性的反应率远高于女性，男性平均反应时间为 40%，而女性为 33%。然而，除一个异常值外，他们的相对反应率非常接近（图 2）。这个异常值（歌曲 "All One People"）的主要差异在于，当表演者唱到他与母亲的对话时，男性的反应率非常高。相比之下，女性的反应率很低，表明歌曲的这一部分引起了男性更强烈的共鸣。

## 年龄差异比性别差异更显著

我们曾认为这些歌曲可能会在年长观众（与作曲者同时代的人）中引起比年轻观众更强烈的共鸣。但总体而言，情况恰恰相反，年轻成员的反应时间为 36%，年长成员为 31%。平均差异可能反映了年轻人平均而言情绪更丰富。然而，当我们查看单首歌曲的结果时（图 3），结果远不如性别结果那样一致。虽然我们注意到年轻人对音乐音量的反应更大，而年长者对 upbeat 节奏的反应更大，但这些因素并不能解释为什么有三首歌曲在年长观众中脱颖而出。很可能是歌词中的主题，而非音乐本身，导致了年长观众较高的反应率。

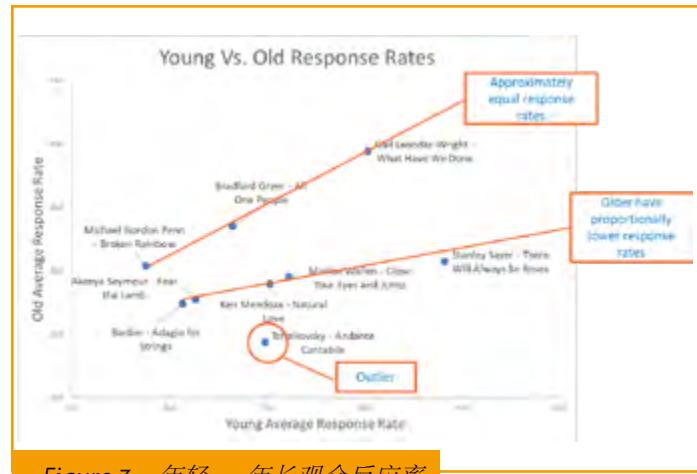


Figure 3: 年轻 vs. 年长观众反应率

## 结果

通过使用 Shimmer 的 NeuroLynQ 系统，Palaver Strings 能够证明其方法所产生的高水平反应。年轻人和男性——他们并非古典音乐的典型消费者——的反应率甚至高于平均水平。Palaver 正在将这一信息用于资助申请和其他营销活动中。详细的分析还提供了宝贵的见解，Palaver 可以将其纳入未来的音乐会和社区合作中。例如，该研究强调了讲故事对于与观众建立联系的重要性。还可以选择音乐，以针对不同观众优化节奏和音量。Palaver 正在将这些发现纳入其未来的节目开发中。

## 关于作者：

Geoffrey 陈 是 Shimmer Americas 的总裁。  
Shimmer 是一家可穿戴技术服务与传感器制造公司，总部位于爱尔兰都柏林，客户遍布 75 多个国家。

## 参考文献：

- 1: 关于这些指标的描述，参见 Dandu, Gill, and Siefert, "Quantifying Emotional Responses of Viewers based on Physiological Signals"，可在此处查看。
- 2: 关于 Shimmer 与 Ipsos 合作的超级碗项目描述，请参见此处。
- 3: 标准化 GSR 的计算方法：将每个个体在每个时间点的实际 GSR 水平减去其在整个体验过程中的平均 GSR 水平，然后除以其 GSR 水平的标准差。然后将所有个体在每个时间点的结果进行平均，得到平均 z 分数。

# 解读现场橄榄球联赛的体育场体验及其对广告的影响

Vodafone Warriors（沃达丰勇士队）是新西兰最令人振奋的职业运动队之一，参加澳大拉西亚地区的NRL（国家橄榄球联赛）。该俱乐部拥有遍布新西兰的忠实粉丝群，并与沃达丰、Canterbury of New Zealand、马自达和ANZsahi Beverages（Woodstock）等品牌达成了众多长期赞助协议——这些企业希望与球队同甘共苦。

但是，在沃达丰勇士队的主场——斯玛特山体育场——作为一名比赛日的支持者，体验究竟如何？这对赞助商又意味着什么？为了回答这些问题，有必要在整个体验过程中对体育场内的观众进行监测。沃达丰勇士队与行为洞察机构NeuroSpot合作，利用Shimmer Research的NeuroLynQ系统，以科学的视角探究球迷在观看第24轮对阵南悉尼兔子队比赛时的感受。

## 研究方法

在本研究中，NeuroSpot利用Shimmer Research的NeuroLynQ生物特征测量系统，在整个比赛过程中测量皮肤电反应的变化——这是由情绪反应变化引起的皮肤反应的细微变化。通过在手指上佩戴两个小型传感器，可以记录球迷群体的兴奋度生理变化，而无需他们告诉我们感受如何。（参见图1。）



图1:

作为这项研究的一部分，20名沃达丰勇士队球迷被邀请在斯玛特山体育场的一个媒体包厢内观看比赛。在整个比赛过程中，NeuroSpot能够测量球迷情绪的逐秒波动——并识别出导致波动的原因。图2显示了媒体包厢的视野。比赛开始前，观众观看了一档谈论该球队和本场比赛的体育脱口秀电视节目，以建立情绪基线。他们还观看了一系列来自非体育场广告赞助商和赞助商的广告，以建立广告反应的基线。

## 研究结果

人们常谈论观看现场体育比赛的兴奋感，这清晰地体现在人们的情绪反应中，高涨情绪相比基线水平提升了38%。比赛中最激动人心的时刻出现在第37分钟，肯·毛马洛为沃达丰勇士队达阵得分（第二次达阵）——此时高涨情绪水平相比基线飙升了44%。这紧随主队两分钟前由亚当·庞培完成的第一次达阵之后，那次达阵使得高涨情绪相比基线提升了20%。

当兔子队达阵得分时，兴奋度水平继续达到峰值——在戴恩·加盖完成一次单刀达阵后，兴奋度跃升了25%。这意味着什么？现场体育赛事带来的兴奋感并不仅仅来自于你支持的球队得分——人们会做出反馈。



图2：图注：媒体包厢的视野

竞争双方球队的对抗以及周围观众的反应所带来的兴奋感。对于像沃达丰和马自达这样的赞助商来，一个关键问题是：这种合作关系如何影响他们的品牌？沃达丰勇士队的球迷是否会将这种情感转移到赞助商品牌上，并可能改变长期的品牌联想？为了找出答案，在比赛开始前和中场休息时，向球迷展示了一系列赞助商广告，以观察反应有何不同。

关键发现是，对于相同广告的情绪反应，在中场休息时相比赛前增加了21%——这种情绪提升可以与赞助商品牌关联起来。有趣的是，我们在赞助商广告中观察到了这种提升，而且由于赞助商标识在整个体育场和球员队服上高度可见，这为嵌入这些品牌联想提供了多重机会。（参见下面的图3）。

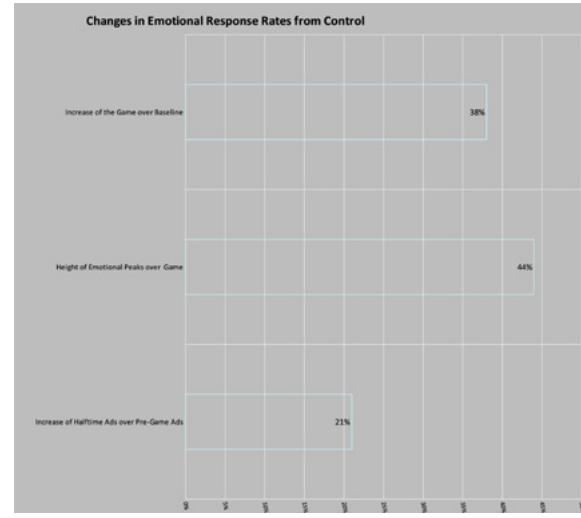


图3：关键数据

中场广告情绪反应相比赛前广告的提升：21%  
比赛中情绪峰值最高相比基线的提升：44%  
整场比赛平均情绪相比基线的提升：38%

## 最终思考

我们相信，越来越多的神经营销工作将需要在现实环境中进行，并采用被动数据收集方式，以捕捉参与者的真实反应。生物特征测量技术，尤其是皮肤电反应和心率，非常适合在现实世界中进行数据收集。并非说这些测量方法没有挑战，但与其他技术相比，挑战更少。事实上，如今已有数亿人通过消费级健身追踪器和智能手表监测这些指标。本研究是众多证明这些测量方法能够在现实环境中捕获有趣且有意义的数据的案例之一。

## 结论

在新西兰奥克兰的斯玛特山体育场主场观看沃达丰勇士队的现场比赛，会在球迷中引发高水平的情绪反应——平均比基线水平高出38%。情绪峰值在比赛过程中多次出现，既包括主队得分时（相比基线最高提升44%），也包括客队得分时。

对球队赞助商而言重要的是，这种情绪提升会延续到赞助商广告上，中场休息时播放的广告相比赛前播放的相同广告，引发的情绪反应高出21%。这说明了与此类赛事关联对赞助商的价值。

## CONTACT INFORMATION

### NeuroSpot

[www.neurospot.co.nz](http://www.neurospot.co.nz)  
Cole Armstrong  
[cole@neurospot.co.nz](mailto:cole@neurospot.co.nz)

### Vodafone Warriors

[www.warriors.kiwi](http://www.warriors.kiwi)  
Aaron Lawton  
[aaron@warriors.kiwi](mailto:aaron@warriors.kiwi)

### Shimmer Research

[www.shimmersensing.com](http://www.shimmersensing.com)  
Geoffrey Gill  
[gill@shimmersensing.com](mailto:gill@shimmersensing.com)

# 观看山水画能减轻压力吗？

疫情普遍加剧了每个人的情绪健康问题，从一线工作者和必要行业员工所处的压力环境，到居家的老年人、在线教学的教师以及在家学习的学生，无一幸免。接触自然已被证实有助于减压，但作为自然替代品的山水画——能否产生同样的效果？

本研究比较了在真实环境中观赏自然（一个由弗雷德里克·奥姆斯特德设计的公园——见图1照片）与观赏由历史上重要的哈德逊河画派艺术家（包括弗雷德里克·丘奇等）创作的山水画（位于康涅狄格州中部的美国新不列颠艺术博物馆，示例见图2）的减压效果。博物馆永久收藏中的此类山水画，以其对自然空间奇妙、崇高且坦率地说令人放松的唤起作用，持续吸引着仰慕者。本研究让参与者在博物馆画廊观赏展出的哈德逊河画派山水画，并在博物馆户外露台观赏真实自然景观，以帮助理解人们是否能从这种代理性的自然表征中，获得与亲身体验真实自然相似的减压和注意力恢复效益。

## 研究方法

博物馆的参观者被告知正在进行一项参与式的“艺术即科学”讲座研究。

共有33名研究参与者自行选择并同意在整个讲座实验过程中佩戴NeurolynQ生物特征传感器，以监测他们的皮肤电反应水平。参与者被随机分配到讲座室的两侧之一，这决定了他们接受“处理”的顺序。参与者首先观看一个五分钟的纽约市交通蒙太奇视频，旨在模拟压力体验。在接触交通蒙太奇视频后，一组参与者转移到博物馆附近的一个画廊，坐着观赏山水画七分钟；而另一组则转移到讲座室外，在真实的户外景观中坐七分钟。观看期结束后，参与者返回讲座室进行五分钟的休息重置。重置后，参与者再次观看五分钟的交通蒙太奇视频，然后两组交换景观观察设置：第一组到室外体验户外景观七分钟，第二组在相同条件下到画廊观赏山水画七分钟。

NeurolynQ系统允许进行非常快速的分析。每组由独立的Shimmer NeurolynQ系统监控，该系统提供小组反应率的实时分析。



图1：弗雷德里克·奥姆斯特德设计的公园



图2：弗雷德里克·埃德温·丘奇，《纽黑文的西岩》，1849年，美国新不列颠艺术博物馆，约翰·巴特勒·塔尔科特基金购买，1950.10

反应率通过分析每位参与者的皮肤电反应并将其分类为三个级别来计算：无反应、有些反应、高反应。然后通过计算每个时刻每个类别中观众所占的百分比来汇总个体结果。

## 结果：

本研究的数据分析仍在进行中，结果计划在同行评审期刊上发表。然而，所使用的技术提供了实时分析，使我们能够在此分享一些定性观察结果。实时分析基于在任何给定时间产生可测量反应的观众百分比（反应率）。图1显示了两种体验（公园与画廊）的对比，数据采用了一分钟的追踪平均反应率。

我们的定性观察包括：

- 无顺序效应。两组的体验似乎非常相似，表明接触顺序的影响可以忽略不计。
- 交通视频不令人放松。毫不意外，交通视频期间的反应率高于公园或画廊体验。
- 公园和画廊的放松效果相似。在公园和画廊环境中，参与者的反应率最初都因安顿下来而有所上升，随后下降，这表明了相似的压力减轻模式。
- 在画廊中放松所需时间更长。参与者在画廊中似乎需要更长的时间才能放松。也许画廊是一个相对不那么熟悉的环境，即使对于常去博物馆的观众也是如此。
- 画廊在初始期后较低的反应率表明更深度的放松。这可能是因为画廊环境是真正静态的，而公园中则有动静，包括风声和远处走动的人。
- 随着时间的推移，两种体验的压力减轻效果都回归基线。这很可能是预期效应，因为人们知道每次体验都有特定的时间限制。

对艺术的认知处理分阶段进行，在对作品进行全面消化的阶段之后，会转向对构成整体的个别细节进行细微处理。最初的放松阶段可能被活跃的认知吸收阶段所覆盖，达到一个积极的刺激-反应水平。这种类型的反应被称为情感环状模型，在艺术观赏参与者中有充分记载。

## 结论：

这些观察结果提示了以下几点结论：

- 自然景观和山水画似乎具有相似的压力减轻效果。
- 随着时间的推移，这些效果可能会减弱。
- 增加一些动静/变化可能会延长放松效果。

这些分析基于一项使用NeuroLynQ实时指标进行的探索性研究。因此，它们需要通过对原始数据进行额外的统计分析和进一步的研究来确认。然而，至少拥有快速的结果对于方向性分析和生成假设非常有用，以便通过更严谨的统计工作来进行更全面的探索。

## 最终思考：

大多数关于亲近自然对健康益处的研究都在城市绿地中进行，或者试验了未来主义的自然模拟版本，如虚拟现实或增强现实。这项研究则回溯过去，测试一种比照



Figure 3: Comparison of park and gallery experiences

片更早的媒介是否能提供减压效果。对于那些因身体原因、距离问题或单纯不喜欢而不愿出门的人来说，皮肤电反应数据证明，弗雷德里克·丘奇（的画作）与弗雷德里克·奥姆斯特德（的公园）有着同样的impact（效果/影响力）！未来的研究方向可能会在更广泛的受众中测试这种“合成自然”，而不仅仅是年长的博物馆常客，或者测试在机构环境中工作或生活的人，以及医院和学校的休息区。

## CONTACT INFORMATION

### Shimmer Research, Inc.

United States

[www.shimmersensing.com](http://www.shimmersensing.com) [www.neurollynq.com](http://www.neurollynq.com)

Geoff Gill

[info@shimmersensing.com](mailto:info@shimmersensing.com)

### Harvard University, School of Public Health

#### United States

Linda Powers Tomasso

[tomasso@hsph.harvard.edu](mailto:tomasso@hsph.harvard.edu)

[www.hsph.harvard.edu](http://www.hsph.harvard.edu)