arXiv:2602.01050v1 [cs.HC] 1 Feb 2026

嵌入式与情境式：增强现实面部训练反馈评估

AVINASH AJIT NARGUND, 美国加利福尼亚大学圣塔芭芭拉分校
ANDREA M. PARK, 美国加利福尼亚大学旧金山分校TOBIAS HÖLLERER,美国加利福尼亚大学圣塔芭芭拉分校MISHA SRA, 美国加利福尼亚大学圣塔芭芭拉分校



图 1. 一位参与者正在进行增强现实引导的面部练习，并接收旨在改善肌肉运动模式的实时反馈。我们评估了三种基于增强现实的反馈条件：(a) 嵌入式 *ARSelfie*，将反馈叠加在用户自己的面部上；(b) 代理嵌入式 人体模型，在三维虚拟形象上提供反馈；(c) 情境式 条形图，通过位于用户面部下方的条形传达反馈；以及 (d) 无反馈的 基线 条件，用户可以看到自己。我们评估了反馈的空间放置如何影响用户表现、体验和认知负荷。此图的部分内容由人工智能生成。

尽管增强现实（增强现实）研究表明，嵌入式可视化有助于大运动训练，但其在面部练习中的适用性仍未得到充分探索。鉴于面部肌肉组织的复杂性，为面部肌肉训练提供有效的实时反馈面临独特的设计挑战。我们开发了三种增强现实反馈方法，它们与用户的空间关系各不相同：情境式（屏幕固定式）、代理嵌入式（呈现在人体模型上）和完全嵌入式（叠加在用户面部上）。在一项被试内研究（N=24）中，我们测量了面部训练任务中的练习准确度、认知负荷和用户偏好。嵌入式反馈降低了认知负荷，且获得了更高的偏好评分；情境化反馈则有助于进行更精准的调整，并提高准确度。定性分析揭示了一项关键的设计张力：嵌入式反馈改善了体验，却也引发了自我意识，并造成理解困难。我们将这些发现提炼为设计考量，以应对面部训练系统面临的权衡，并探讨其对康复、表现训练和运动技能习得的启示。

作者联系信息：Avinash Ajit Nargund，anargund@ucsb.edu，电气与计算机工程，加利福尼亚大学圣塔芭芭拉分校，美国加利福尼亚州圣塔芭芭拉；Andrea M. Park，andrea.park@ucsf.edu，加利福尼亚大学旧金山分校，美国旧金山；Tobias Höllerer，holl@cs.ucsb.edu，计算机科学，加利福尼亚大学圣塔芭芭拉分校，美国加利福尼亚州圣塔芭芭拉；Misha Sra，sra@cs.ucsb.edu，计算机科学，加利福尼亚大学圣塔芭芭拉分校，美国加利福尼亚州圣塔芭芭拉。

允许为个人或课堂使用而免费制作本作品的全部或部分数字副本或纸质副本，前提是这些副本不得用于营利或商业利益，也不得为此目的进行分发，并且副本首页须保留本声明和完整引文。对于本作品中由作者以外的其他方拥有版权的部分，必须尊重其版权。注明出处后可以摘录。以其他方式复制、重新发布、发布到服务器或分发至列表，须事先获得明确许可和/或支付费用。请向 permissions@acm.org 申请许可。
© 2018 版权所有归所有者/作者所有。出版权已授权给 ACM。稿件已提交至 ACM

CCS概念：• **以人为本的计算 → 可视化中的实证研究**。

其他关键词和短语：增强现实、面部肌肉可视化、训练

ACM 引用格式：
Avinash Ajit Nargund、Andrea M. Park、Tobias Höllerer 和 Misha Sra。2018。嵌入式与情境式：增强现实面部训练反馈评估。载于请从版权确认邮件中填写正确的会议名称（会议简称'*XX*）。ACM，纽约，纽约州，美国，23 页。
https://doi.org/XXXXXXX.XXXXXXX

1 引言

面部肌肉在人类交流、情绪表达以及各种生理功能中发挥着至关重要的作用 [45]。对面部肌肉运动进行监测和训练，可在一些专业应用中带来益处。这些应用包括希望提升情绪表现力的演员和表演者 [5, 6, 18], 、旨在改善发声 [15, 21], 的公众演讲者，以及正在接受面瘫康复治疗 [4] 或面部表情减少症康复治疗 [9] 的患者。在这些场景中，精细的指导对于帮助用户分离并激活特定面部肌肉至关重要。本研究评估可视化的效果，不依赖于具体应用。我们认为，了解可视化在以下方面的影响，对于不同应用领域都很重要：1）练习准确度；2）认知与任务负荷；以及 3）用户偏好。

现有的面部肌肉训练系统提供的视觉反馈形式有限，通常显示在用户面部旁边。例如，Farapy [4] 会在用户镜像中的面部下方叠加条形图，以指示肌肉激活情况；而 GY MEDIC [22] 及相关系统则在面部旁显示原始对称性评分或文本标签 [9]。这些紧邻屏幕边缘的可视化内容与动作发生区域在空间上相互分离，可能妨碍用户理解并据此采取行动 [4]。尽管这些工具支持基本的自我监测，但无法提供与动作本身相契合的空间集成式指导。

相比之下，全身及大运动训练的交互式系统已广泛采用基于增强现实（AR）的空间嵌入式反馈 [2, 49, 52, 61]。这类系统通常会将视觉线索与用户身体或运动轨迹对齐，并借助增强现实镜、可穿戴显示器或头戴式叠加层 [48, 52, 60]呈现。这种嵌入式可视化 [55] 通过将反馈与身体动作置于同一位置，增强了用户的理解能力。尽管这些技巧已在肢体和姿势训练中得到充分研究 [47, 55], ，但对于面部细微、精细动作的应用仍缺乏充分探索。

为弥补这一空白，我们研究了在使用移动增强现实进行面部肌肉训练时，视觉反馈的空间位置如何影响用户体验和表现。我们评估了三种用于传达面部肌肉激活的可视化策略，这些策略与用户面部的空间关系各不相同（见图 1）：(a) 屏幕锚定的条形图，(b) 代理嵌入式人体模型，以及 (c) 嵌入式*ARSelfie* 视图。受 Farapy [4],所使用的可视化方式启发，条形图 通过放置在屏幕底部的比例填充条来传达肌肉激活状态。在人体模型范式中，反馈提示投射在用户面部的纯纹理三维模型上；而在*ARSelfie* 视图中，提示则直接叠加在用户的面部上。这些技巧沿着扩展现实运动引导系统 [59]视觉纠正性反馈设计空间的“WHERE”轴，代表了不同的方法。

我们开展了一项被试内研究（$N = 24$），参与者使用三种可视化条件完成了三项标准化面部练习，并设置了一个无反馈的基线条件；在该条件下，参与者可以在摄像头中看到自己的面部，但看不到任何反馈可视化。本研究围绕以下研究问题展开：提交至 ACM 的稿件

(1) 研究问题 1：与情境式条形图条件或基线条件相比，嵌入式反馈（*ARSelfie* 和人体模型条件）能否提高（a）练习准确度，即每次重复中正确激活的肌肉所占比例，以及（b）正确完成练习的重复次数？**(2) 研究问题 2：**与基线条件相比，各种反馈模态对应的认知与任务负荷如何？**(3) 研究问题 3：**用户偏好哪种条件？

我们的工作做出了以下贡献：

- 我们提出了三种可视化方法：情境式、代理嵌入式和完全嵌入式，将先前的增强现实反馈研究从大运动训练拓展到精细的面部肌肉练习。
- 通过一项被试内研究（$N = 24$），我们提供了实证证据：已被证明对全身和大运动训练有效的嵌入式反馈，同样适用于面部肌肉训练，并且与促进面部肌肉精确激活的情境化反馈相比，能够带来更好的用户体验。
- 我们综合定量与定性结果，提出设计指南，以指导未来系统如何有效平衡练习表现准确度、反馈可解释性和用户舒适度。

2 相关工作

我们的研究与增强现实的应用相关，具体而言，聚焦于面部运动学习中用于指导和反馈的情境式与嵌入式可视化。虽然已有多项研究探讨了基于增强现实的反馈在步态康复 [44], 、上肢训练 [49] 或一般物理治疗 [17], 等领域的应用，但此前研究对其在面部练习中的有效性和适用性仍鲜有探索。只有少数研究，例如 Farapy [4], 、GY MEDIC [22],和 Cai等 [9] ，探讨了基于增强现实的反馈在面部练习中的应用。本节简要概述相关研究，以说明我们研究的贡献。

2.1 运动学习中的指导与反馈

习得一项运动技能是一个复杂的过程，取决于有效的指导和反馈。运动学习包含依次进行的三个阶段：认知阶段、联结阶段和自主阶段 [19]。反馈在最初的认知阶段最为关键，此时学习者正在形成动作的心理模型；在联结阶段，学习者则不断完善自己的技巧。学习者通过自身感觉系统（例如本体感觉）获得内在反馈，而由外部来源（如人类教练或计算机）提供的增强反馈，通常有助于加快学习进程并克服表现停滞期 [42]。

对增强反馈进行分类的一种方法，是区分结果知识（*KR*）与评估鸿沟 [39], 。前者告知学习者其动作的结果（例如是否击中目标），从而缩小评估鸿沟；后者表现知识（*KP*）则提供有关动作本身是否正确及其特征的信息（例如手臂抬起的速度不够快），有助于缩小执行鸿沟 [39]。基于AR的训练系统能够实时感知用户动作，并提供详细的表现知识，而这些信息对人类导师而言很难准确传达 [47]。

我们的工作探讨如何利用嵌入式增强现实可视化，为面部练习提供精准的实时知识表现。我们将这种方法与情境式增强现实可视化进行比较，以评估它们对用户表现、体验和偏好的影响。

2.2 反馈可视化

White 提出的情境化可视化 [54], 是指将数据可视化放置在与之相关的物理实体或指涉对象附近。情境化可视化始终与指涉对象相邻，而嵌入式可视化则更进一步，将视觉元素与指涉对象的物理范围对齐 [55]。运动技能习得和物理康复领域的反馈系统研究，探索了多种可视化范式，旨在弥合用户当前状态与目标状态之间的差距。这些系统的主要区别在于反馈与用户身体之间的空间关系。空间关系会显著影响指导信息的可解读性，以及用户所承受的认知负荷 [47, 55]。

在广泛的可视化范式中，基于AR的训练系统的先前工作采用了三种主要策略：情境式、代理嵌入式和嵌入式。

情境化反馈。 情境化可视化将反馈与身体分离，并将其显示在周围环境中，例如将足部放置位置投射到跑步机上 [44], 显示专家的等身大小视觉图像以帮助用户进行攀岩 [29]或提供基于文本的运动反馈 [11, 16]。

代理嵌入式反馈。 代理嵌入式可视化将反馈映射到按用户体型和身体尺寸缩放的虚拟化身或骨骼模型上。YouMove [2] 在大型增强现实镜上显示的骨骼上可视化运动引导，而 MuscleRehab [61] 通过在虚拟三维模型上显示肌肉激活水平来提供反馈。

嵌入式反馈。 嵌入式反馈通过将提示直接与身体对齐来最小化感知—动作差距 [55]。LightGuide [48] 将二维和三维箭头直接投射到用户手上，以引导空中手势；与基于视频的引导相比，手部运动准确度提高了近85%，而 SleeveAR [49] 将颜色编码的反馈投射到定制袖套上，以支持肩部康复。一些系统使用虚拟镜子，将视觉效果叠加在用户的镜像上。这些系统将镜子的熟悉感与增强反馈相结合，以引导用户 [52] 或突出用户所犯的错误 [53]。

尽管这些范式已被证明对上肢康复和步态训练有效，但它们在面部练习中的应用仍未得到充分探索，而面部练习涉及精细运动。现有的面部锻炼系统主要采用情境化可视化，将反馈显示在用户面部旁边。例如，Farapy [4]和Cai等 [9] 在用户的自拍画面附近使用增强现实叠加层，以提供表现反馈。然而，反馈与面部相分离会造成脱节：Farapy [4] 的参与者表示，屏幕底部的条形图让他们更难解读反馈，并建议直接在面部肌肉上显示颜色变化等嵌入式提示会更合适。

在我们的研究中，我们考察了三种专为面部运动训练设计的可视化策略的有效性。借助这些可视化条件，我们首先探究已证实有益于肢体和全身训练的嵌入式可视化 [59], 是否适用于面部运动训练。随后，我们比较反馈相对于面部的空间位置如何影响面部动作的准确性、理解指导信息的难易程度，以及由此给用户带来的认知负荷。

2.3 面部可视化

先前的研究利用用户的镜像来展示个人健康信息学 [1, 13, 25, 40, 50]。WizeMirror [25]使用来自多种传感器的数据来估计心代谢风险因素，并显示一个综合的“健康指数”

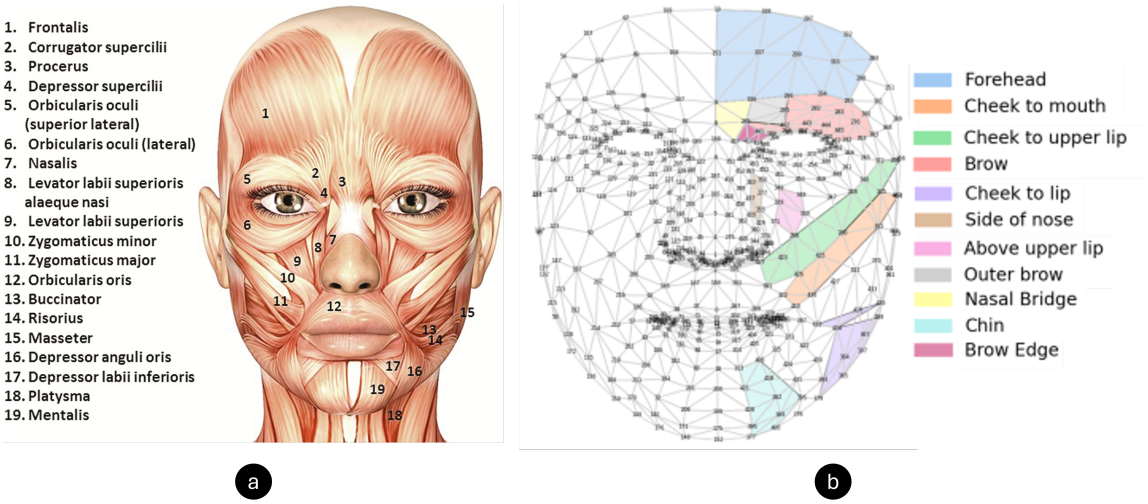


图 2. 面部肌肉解剖图 (a)（来自 [38]），用于将 Google ARCore 增强现实面部API 提供的 468 个面部关键点三维面部网格 (b) 的子集映射到本研究的目标肌肉。

在用户的镜像旁边。同样，Rigling 等人 [40] 探索了使用增强现实面部滤镜来可视化健身追踪器指标，例如步数和睡眠时间。

基于先前将健康数据可视化于身体或面部，以减少数据与其背景之间脱节的研究，我们将这一方法拓展至面部肌肉训练：在相应的面部解剖结构上实时呈现肌肉激活情况，为用户提供即时且扎根于情境的反馈。

3 研究设计与实现

为了探究我们的研究问题，我们使用 Unity[1] 游戏引擎开发了一款移动增强现实应用。该应用使用 Google ARCore 的 *AugmentedFaces* Unity API [2] 进行面部追踪。该 API 提供 468 点三维面部网格（图 2b），并利用手机前置摄像头实时追踪面部。我们利用这一密集面部网格来计算肌肉激活，并将反馈叠加层在空间上配准到用户面部。与使用原始图像像素相比，基于关键点的方法具有两大关键优势。首先，ARCore 中的预训练机器学习（ML）模型针对移动端推理进行了优化，并在大规模数据集上训练，因此对光照、头部姿态和用户外观的变化具有较强鲁棒性。其次，三维网格提供了语义组织良好且具有解剖学依据的面部表征。这便于通过测量局部关键点位移直接推导出可解释的肌肉激活估计值，而基于像素的方法往往依赖间接推断，并且需要大型标注数据集来应对外观、表情和光照的变化。

我们设计并实现了四种实验性可视化条件：*ARSelfie*（嵌入式）、人体模型（代理式嵌入）、条形图（情境式）和基线。本节将介绍每种可视化条件的设计依据、系统架构和功能特性。

[1]https://unity.com/[2]
https://developers.google.com/ar/develop/unity-arf/augmented-faces/developer-guide

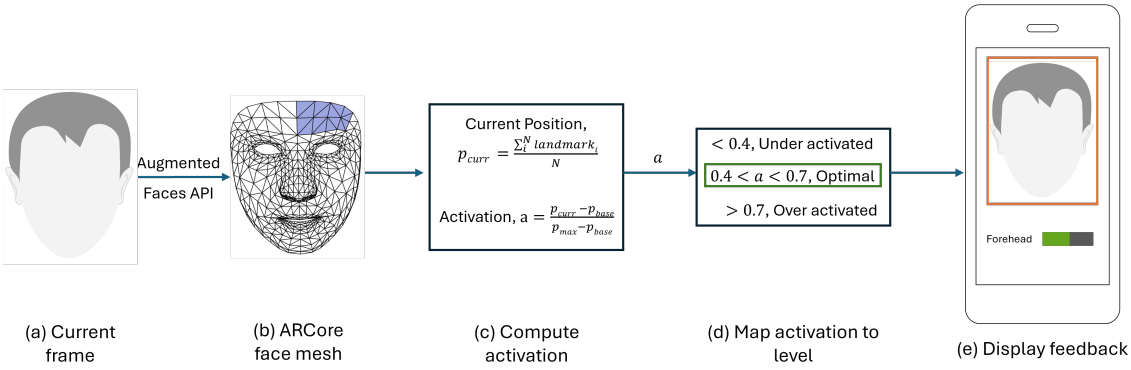


图 3. 在条形图 条件下，以左侧额肌为例展示激活估算过程。首先，手机前置*AugmentedFaces* API 摄像头捕捉用户的面部，并返回一个面部网格。利用与特定肌肉对应的特征点来估算其当前位置。随后结合该位置以及肌肉的基线位置和峰值位置，计算原始激活分数。再将该分数映射为激活标签，并利用这两者向用户提供反馈。

3.1 将特征点映射至肌肉

为了提供针对特定肌肉的反馈，我们手动将 468 个 ARCore 面部关键点中的部分关键点映射到相应的面部肌肉。我们利用这样一条生物力学原理：与大多数其他肌肉不同，面部肌肉直接附着于皮肤，而非骨骼 [12]。这种直接附着意味着肌肉收缩会引起局部皮肤形变，因此可以使用面部特征点簇作为代理来估计肌肉激活 [46, 58]。该方法类似于 Hasebe 等人 [24], 以往研究中采用的方法；该研究证实，特征点簇的运动与整合肌电图（iEMG）对肌肉激活的估计结果相关，尤其是在口部（显著相关）和眼部肌肉方面。通过将 ARCore 面部网格叠加到面部肌肉解剖图 [35], 上，我们直观地识别出位于特定肌肉上方的面部关键点簇，其位移可以指示激活情况。该映射的准确性由我们的第二作者审查并验证；该作者是一名面神经外科医生。

3.2 参与者校准

在校准阶段，系统首先提示用户保持中性面部表情，同时记录所有面部关键点的基准位置。接下来，用户尽最大努力完成每项目标练习（例如，尽可能大幅度地微笑），系统则记录峰值激活时的三维特征点位置。随后，根据每块肌肉在峰值激活和中性表情下的特征点簇质心之间的三维欧几里得距离，计算其最大活动范围。此次校准为确定激活水平建立个性化参考尺度。

3.3 肌肉激活

在练习过程中，系统通过追踪特定面部特征点簇的位移来估算肌肉激活（见图 3）。系统将每个簇的质心作为相应肌肉的位置，并持续测量其当前位置与中性表情下初始位置之间的欧几里得距离，以此计算位移。随后，将当前位移与校准步骤中确定的最大活动范围之比作为肌肉激活值。稿件已提交至 ACM

。该归一化方法根据每位用户的活动范围，为其所有肌肉建立个性化的 0 到 1 量表，从而考虑个体间差异。随后，系统将激活得分映射为三分类标签。得分低于 0.4 时归类为激活不足，高于 0.7 时归类为激活过度，介于两者之间则视为最佳。上述阈值通过先导测试并咨询面部外科医生后确定。这种三分类方法与先前研究 [4] 所采用的方法相似。

在这项研究中，我们采用基于标志点位移的启发式方法，而非机器学习（ML）模型，以确保肌肉激活估计值具有确定性和稳定性。在激活估计层面控制变异性，有助于我们分离可视化设计的影响，并避免机器学习预测不准确或噪声带来的潜在混杂因素。此外，这种启发式方法可最大限度地减少移动设备的计算负载，确保反馈界面快速响应。

3.4 可视化条件

为了探究反馈的空间位置和具身程度如何影响用户，我们设计了四种可视化条件（见图4）。

基线。我 在基线 条件下，系统显示用户的实时摄像头画面，不提供基于激活分数的反馈 .

情境式可视化：条形图。条形图条件通过放置在设备屏幕底部的水平条形图提供肌肉激活反馈，与顶部显示的用户实时摄像头画面相互分离。每个条形图对应练习中的一个目标肌肉，并以该肌肉的常用名称标注（例如，额肌标注为“额头”）。条形图填充部分的宽度与肌肉的原始激活值成正比，从而向用户持续反馈激活强度。条形图填充部分根据激活水平着色：激活不足时为蓝色，激活适当时为绿色，激活过度时为橙色。

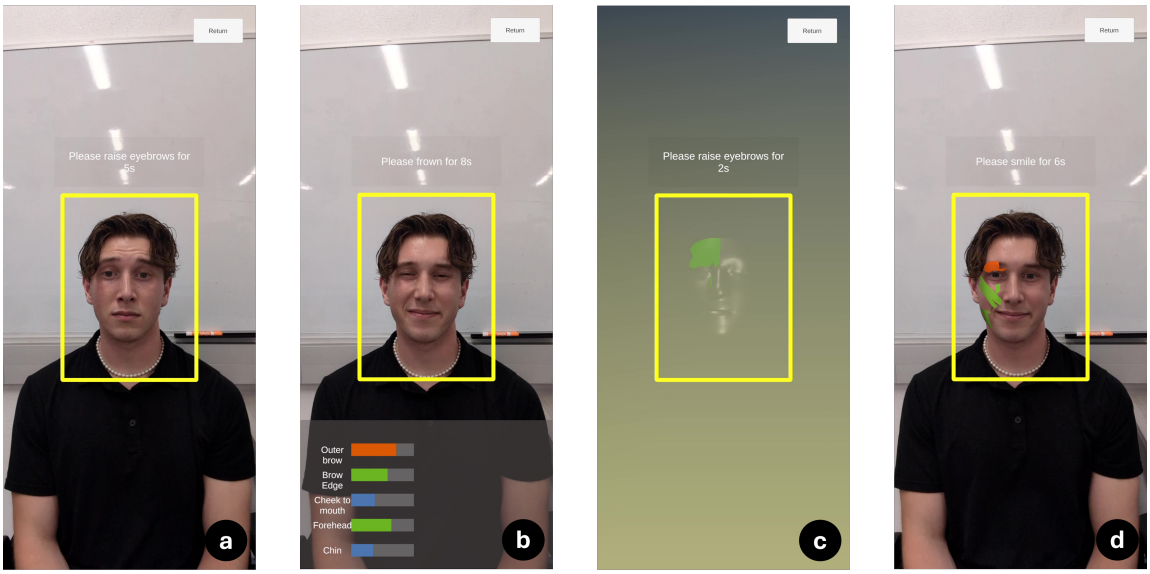


图 4. 研究中使用的四种（挑眉） ，(b)条形图 （反向皱眉），(c)人体模型（挑眉）和(d)*ARSelfie* （微笑）。 实验条件，展示一名参与者进行三种面部练习：(a)基线 *e*

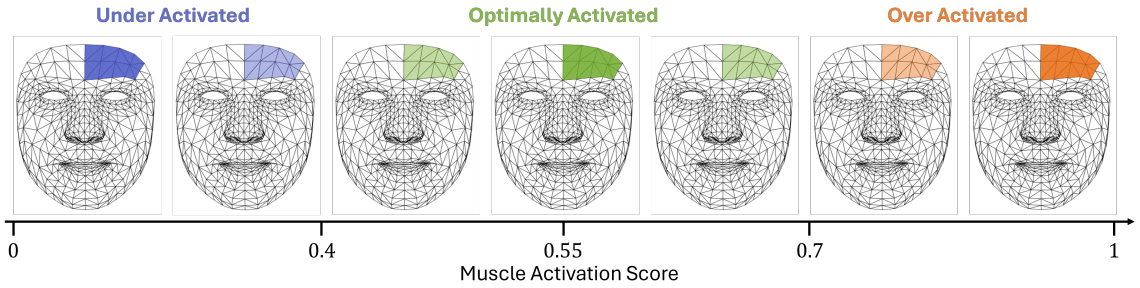


图 5. 嵌入式条件下的透明度调节反馈，以 ARCore 面部网格上的额肌为例进行说明。颜色表示激活区域：蓝色代表激活不足，绿色代表最佳范围，橙色代表激活过度。透明度编码了与最佳范围的距离：随着激活水平接近绿色区域，蓝色叠加层逐渐淡出；绿色叠加层在范围中点处最不透明，在接近下限和上限阈值时则更透明；当激活水平进一步超过最佳范围时，橙色叠加层的透明度随之增加。

代理嵌入式可视化：人体模型。人体模型条件通过一个基于 ARCore 面部网格创建的动态三维面部虚拟形象提供视觉反馈。该虚拟形象通过网格变形实时复现用户的面部动作，但不使用照片级真实纹理，因此呈现为中性的代理面部。反馈通过与特定肌肉区域对齐的彩色叠加层呈现；这些叠加层通过将网格分割成与目标肌肉对应的子区域生成。

此可视化条件同时使用颜色和透明度来传达反馈。这种方法与 Ikeda 等人 [26] 采用的方法一致，也符合 Lee 等人 [33] 所描述的情境化可视化 “Decal” 模式（见图 5 ）。叠加层的颜色表示离散的激活水平：蓝色代表激活不足，绿色代表最佳激活，橙色代表激活过度。叠加层的透明度表示当前激活分数与最佳激活分数之间的绝对差值。在激活不足范围内（得分低于 0.4），随着激活水平接近最佳阈值，蓝色叠加层的透明度会降低，从而鼓励用户进一步激活肌肉。在最佳范围内（0.4 – 0.7），绿色叠加层在范围中点（得分为 ~ 0.55）处最不透明，并朝最佳范围的下限和上限逐渐变得更透明。在激活过度范围内（得分高于 0.7），随着激活分数升高，橙色叠加层的透明度也会增加，以表示激活过度。

嵌入式可视化：*ARSelfie*。 *ARSelfie* 条件采用与人体模型相同的设计、反馈机制和颜色。主要区别在于，反馈直接叠加在实时视频流中用户的面部上，而不是代理虚拟形象面部上。

3.5 数据记录

移动应用程序会将所有目标肌肉的原始激活分数，以及当前练习的标识符，按条件分别记录到不同文件中。这些数据以应用程序的帧率采集（与 Unity 的 Update() 周期同步），用于计算性能指标，包括肌肉处于最佳激活状态的练习时间占比、达到首次最佳重复所需的时间，以及练习时间窗口内的最佳重复次数总量。

4 用户评估

我们开展了一项用户研究，以考察面部运动训练中视觉反馈的不同空间布局如何影响参与者进行面部练习时的用户表现、感知任务负荷和体验。为开展这项研究，我们开发了一款定制应用，提供三种基于增强现实的反馈：*ARSelfie*、人体模型 和 条形图，以及无反馈基线 （详见第 3 节）。研究在一间光线稳定、安静的室内房间进行。应用运行于一部 Google Pixel 8 Pro 智能手机上，手机放置在参与者前方.5 米处。研究使用下巴托固定参与者的头部，以尽量减少非预期头部移动造成的面部追踪不一致。下文将介绍研究中使用的面部练习、研究流程（见图 6）和任务。

4.1 面部练习

参与者完成了三项面部练习：微笑、挑眉和反向皱眉，每项练习都需要激活来自不同面部区域、大小各异的肌肉（见图4）。微笑依赖于口周肌肉的参与。挑眉是将眉毛向发际线抬起，主要由额肌控制。反向皱眉通常被称为“挤脸”，要求参与者将眉毛向下并拢，同时将面颊和下巴的肌肉向上移动。这一复合动作需要协调上面部与下面部的肌肉激活。

4.2 研究流程

参与者在研究开始时提供知情同意（本研究已获本机构审查委员会批准，协议编号为#anony-mous）、人口统计信息，并完成技术交互亲和度（ATI）问卷 [20],，该问卷用于评估个体在与技术交互和参与方面的差异。参与者通过教程了解了研究目的，并学习了三项面部练习。教程还解释了可视化编码，包括在各条件下使用的配色方案和不透明度调制。参与者还了解到，当该练习所涉及的所有肌肉的肌肉反馈指标均变为绿色时，即表示达到最佳重复次数。

教程结束后，参与者完成了一个校准阶段。在此阶段，他们首先被要求保持中性表情 10 秒，以记录其基线面部标志点位置。接着，他们以最大努力完成三项练习，每项练习之间休息 5 秒。此次校准也帮助参与者熟悉他们将在练习阶段进行的练习（详见第 4.3 节）。

在练习阶段，每完成一种可视化条件后，参与者都会填写 NASA-TLX 问卷 [23], 用户体验问卷（UEQ）[32]以及一份根据 [28] 改编的自定义问卷，以评估外在认知负荷（ECL）（表 1）。测量外在认知负荷有助于我们估计每种可视化条件对参与者提取与任务相关的信息并有效完成练习的要求有多高。研究结束时，参与者填写了一份研究后问卷，按偏好顺序对各条件进行排序，并在与研究人员进行的半结构化访谈中说明其偏好的理由。

4.3 研究任务

在练习阶段，参与者完成了四次试次，每种可视化条件各一次。试次顺序采用平衡拉丁方进行平衡，以尽量减少学习效应并降低条件间潜在的延续效应。每次试次包含六组，每组均包含全部三项练习，且练习顺序经过排列，以减轻顺序效应

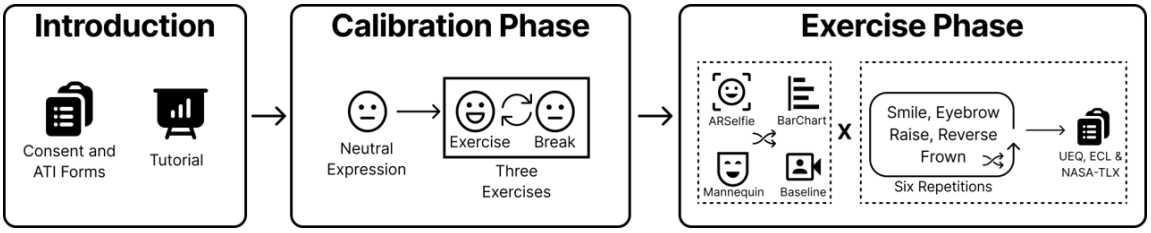


图 6. 研究流程。参与者提供知情同意并完成 ATI 问卷 [20], 后，通过演示文稿接受关于练习、研究任务和反馈条件的教程，随后切换到系统进行校准阶段和练习阶段。在校准阶段，参与者先保持中性表情（10 秒），随后以最大努力完成三项练习（每项 10 秒，间隔 5 秒）。在练习阶段，参与者完成四个平衡顺序的试次（拉丁方设计），每个条件对应一个试次。每个试次包括六次练习循环（以随机顺序进行三项练习，每项 10 秒，间隔 5 秒）。每个试次结束后，参与者完成 NASA-TLX [23], 用户体验问卷 [43], 以及一份自编的外在认知负荷问卷。

效果。每项练习中，参与者练习 10 秒，随后休息 5 秒。他们被要求在每个 10 秒时间窗口内尽可能多地完成最佳重复次数，只有当所有肌肉反馈指标均显示为绿色时，该次重复才计为最佳重复。练习和休息时长通过先导测试确定，旨在确保参与者有足够时间在每个练习时间窗口内完成多次重复，同时尽量减轻整个研究环节中的疲劳。

4.4 参与者

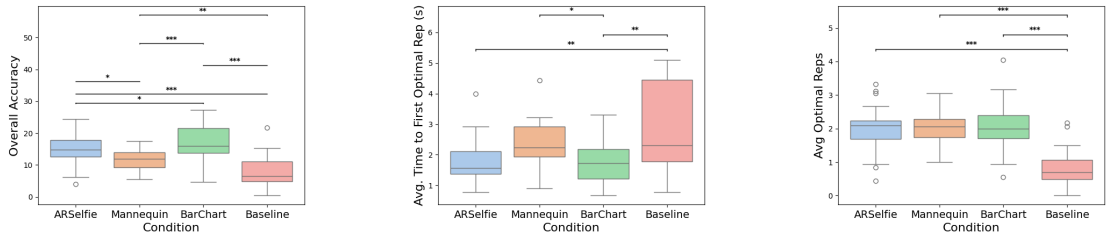
初步试测后，我们进行了功效分析，以确定所需的样本量。假设效应量较小至中等（$f = 0.2$）、显著性水平为 $\alpha = 0.05$、统计功效为 $1 - \beta = 0.8$，且重复测量之间的相关系数为 0.7，分析得出的目标样本量为 22 名参与者。随后，我们通过内部邮件列表招募了 24 名年龄在 18 至 30 岁之间的参与者（10 名女性，14 名男性；中位数 = 20 岁）。所有参与者的视力正常或矫正后正常。17 名参与者报告没有色觉缺陷。两名参与者自述存在轻微色觉缺陷，但确认能够区分研究中使用的颜色。其余五名参与者未透露色觉状况，但报告在研究开始前或过程中区分颜色没有困难。所有参与者均获得 15 美元作为报酬。

5 定量结果

我们评估了以下因变量，以考察反馈条件的影响：(a) 性能指标，包括练习准确度、首次达到最佳重复的时间和重复次数；(b) 通过用户体验问卷（UEQ）测量的用户体验评分 [43], ；(c) 通过 NASA-TLX 测量的任务负荷 [23],；以及 (d) 外在认知负荷（ECL）得分。我们对所有因变量进行了 Shapiro-Wilk 检验，以评估其正态性。满足正态性假设时，我们采用重复测量方差分析；否则采用弗里德曼检验。所有事后两两比较均采用 Bonferroni-Holm 校正。

5.1 练习表现

我们通过三项客观指标衡量练习表现：稿件已提交至 ACM



(a) 激活准确率 (b) 首次达到最佳重复的时间 (c) 重复次数

图 7. 参与者的表现通过以下指标衡量：(a) 平均准确率、(b) 首次达到最佳重复的时间，以及 (c) 最佳重复次数。使用条形图 可视化时，参与者的准确率显著更高。

(1) 准确度，肌肉处于最佳激活状态的练习时间占总练习时间的比例，并对所有练习时间窗口取平均值。**(2) 最佳重复次数**，每个 10 秒练习时间窗口内完成的最佳重复次数的平均值。**(3) 首次达到最佳重复的时间**，从练习开始到完成首次最佳重复所用的时间（以秒为单位）。

图 7 展示了各可视化条件下的平均性能指标。

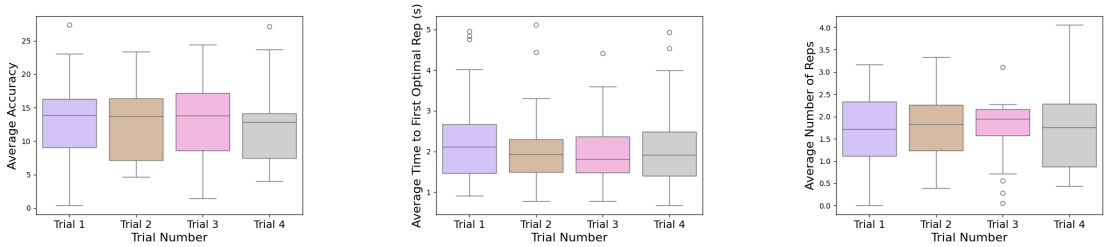
准确率。 我们发现，可视化条件对平均准确率的主效应显著，$F_{3,69} = 26.16, p < 0.001$。事后比较显示，与 *ARSelfie* （$\mu = 14.59, \sigma = 5.3, p = 0.046$）、人体模型 （$\mu = 11.49, \sigma = 3.18, p < 0.001$）和基线 （$\mu = 7.97, \sigma = 5.1, p < 0.001$）条件相比，参与者在条形图 条件下的表现显著更好（$\mu = 16.88, \sigma = 5.7$）。有趣的是，尽管人体模型 条件在概念上与*ARSelfie*相似，但人体模型 条件下的准确率（$\mu = 11.49$）显著低于*ARSelfie*条件（$\mu = 14.59, p = 0.019$）。

最佳重复次数。 重复测量方差分析显示，可视化方式对最佳重复次数的平均值有显著主效应，$F_{3, 69} = 35.9$， $p < 0.001$。事后分析表明，与基线 条件（$\mu = 0.81$；所有 $p < 0.001$）相比，参与者在*ARSelfie* （$\mu = 1.97$）、人体模型 （$\mu = 2.01$）和条形图 （$\mu = 2.08$）条件下完成的重复次数显著更多。

首次达到最佳重复的时间。 弗里德曼检验显示，可视化方式对首次达到最佳重复的平均时间有显著主效应， $Q = 13.1$， $p = 0.004$， $W = 0.18$。事后两两比较表明，与*ARSelfie*（$\mu = 1.76s$， $\sigma = 0.7s$， $p = 0.003$）和条形图 （$\mu = 1.7s$， $\sigma = 0.65$， $p = 0.005$）条件相比，参与者在基线 条件（$\mu = 2.85s$， $\sigma = 1.46s$）下首次达到最佳重复所花的时间显著更长。值得注意的是，与条形图（$p = 0.01$）相比，用户在人体模型 条件（$\mu = 2.32s$， $\sigma = 0.78s$）下的速度也显著更慢。

5.1.1 学习与迁移效应。我们分析了性能数据，以确定参与者表现是否会随着完成的试次增多而提升，或（经过平衡的）条件顺序是否影响了结果。

为检验学习效应，我们评估了四次试次中的表现（见图 8）。重复测量方差分析检验显示，试次编号对平均准确率、$F_{3, 69} = 0.19$， $p = 0.9$ 或 M



（a）激活准确率 （b）首次达到最佳重复的时间 （c）重复次数

图 8. 按试次编号（1–4）展示的三项性能指标，用于评估学习效应。图中显示，参与者在所有指标上的表现从第一次试次到最后一次试次均保持稳定，表明实验过程中未出现显著的学习效应。

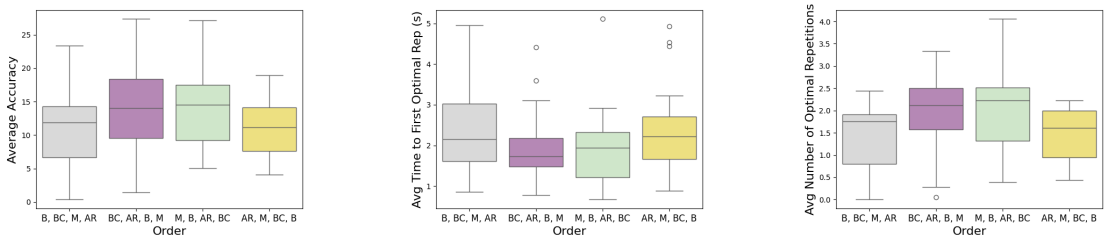
平均最佳重复次数，$F_{3,\ 69} = 0.04$，$p = 0.98$均无显著影响。同样，弗里德曼检验显示，试次编号对首次达到最佳重复的平均时间，$Q = 2.2$，$p = 0.53$，$W = 0.03$无显著影响。

为检验残留效应，我们进行了混合方差分析检验，将条件作为被试内因素，将可视化呈现顺序作为被试间因素（见图 9）。分析显示，准确率（$F = 1.55$，$p = 0.23$，$\eta_p^2 = 0.19$）和最佳重复次数（$F = 2.59$，$p = 0.08$，$\eta_p^2 = 0.28$）均无显著差异。对于非正态分布，我们使用了对齐秩变换方差分析 [56], ，结果同样显示顺序没有显著影响，$F = 2.22$，$p = 0.09$，$\eta_p^2 = 0.07$。这表明残留效应可以忽略不计。

5.1.2 ATI 和对 AR 的熟悉度的影响。 我们进行了相关性分析，以确定 ATI [20] 得分或对 AR 的熟悉度中的个体差异是否会影响表现。分析结果显示，ATI 得分或 AR 熟悉度与三项性能指标中的任何一项均不存在显著相关性。

5.2 用户体验

图 10 展示了四种条件下用户在用户体验问卷（UEQ） [43]六个量表上的评分。总体而言，*ARSelfie* 和人体模型 条件在实用性和



(a) 激活准确率 (b) 首次达到最佳重复的时间 (c) 重复次数

图 9. 用于评估残留效应的四个呈现顺序组的性能指标。在三项指标中的任何一项上，各组之间的表现均无显著差异。这表明，参与者查看可视化的顺序对其表现没有显著影响。

用户体验的享乐方面。条形图条件的评分通常高于基线，但低于两个嵌入式条件。

实用性 指与任务相关的用户体验方面。它由明晰性、效率和可靠性 子量表组成。针对明晰性，弗里德曼检验显示，可视化条件的主效应显著（$W = 0.35$，$p < 0.001$）。事后比较表明，参与者认为*ARSelfie* 和人体模型 条件比条形图 （分别为$p = 0.019$ 和 $p = 0.016$）以及基线 （两者均为$p < 0.001$ ）清晰得多。

针对效率，弗里德曼检验显示主效应显著（$W = 0.5$，$p < 0.001$）。事后比较表明，参与者认为基线 条件在引导他们方面的效率显著低于*ARSeflie*（$p < 0.001$）、人体模型 （$p < 0.001$）和条形图 （$p = 0.026$）。其他三个条件之间未发现显著差异。

对于可靠性，重复测量方差分析发现了显著的主效应（$F_{3,\ 69} = 8.41$，$p < 0.001$，$\eta_g^2 = 0.16$）。由于球形假设被违反，因此应用了 Greenhouse-Geisser 校正。事后检验显示，参与者认为*ARSelfie* 条件比条形图 （$p = 0.023$）和基线 （$p = 0.012$）都更可靠。此外，人体模型 可视化的可靠性评价也高于基线 条件（$p = 0.011$）。

享乐质量衡量交互的情感层面，由刺激性 和新颖性两个子量表构成。

对于刺激性，弗里德曼检验显示存在显著的主效应（$W = 0.65$，$p < 0.001$）。基线 条件的刺激性评分显著低于*ARSelfie* （$p < 0.001$）、人体模型 （$p < 0.001$）和条形图 （$p < 0.001$）。此外，*ARSelfie* 的刺激性评分高于条形图 （$p < 0.001$）。

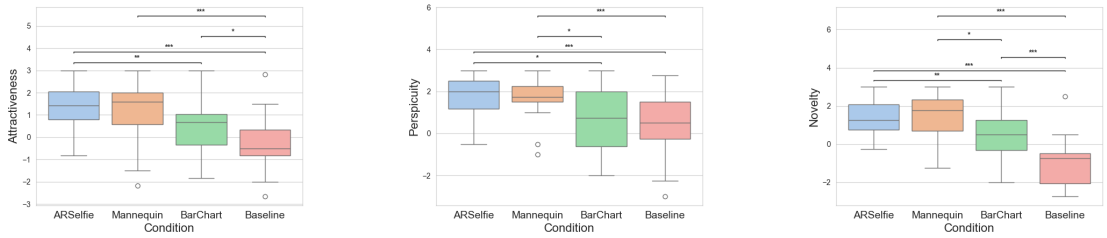
对于新颖性，弗里德曼检验显示存在显著主效应（$W = 0.62$，$p < 0.001$）。事后比较表明存在明确的排序，其中*ARSelfie* 和人体模型 被评为更具新颖性，其次是条形图 ，而基线 被评为新颖性最低。沿此排序的所有成对比较中，除*ARSelfie* 与人体模型 之间的比较外，均达到显著水平（$p < 0.05$）。

吸引力 量表提供了对每种条件整体吸引力的总体测量。重复测量方差分析发现了显著的主效应（$F_{3,\ 69} = 15.63$，$p < 0.001$，$\eta_g^2 = 0.24$）。基线 条件的吸引力评分显著低于*ARSelfie* （$p < 0.001$）、人体模型 （$p < 0.001$）和条形图 （$p = 0.014$）。此外，*ARSelfie* 的吸引力评分高于条形图 （$p = 0.008$）。

5.3 任务负荷

我们进行了一项重复测量方差分析，以评估可视化条件对加权总体 NASA-TLX 工作负荷得分的影响。由于数据违反了球形性假设，因此采用了 Greenhouse-Geisser 校正。分析显示存在显著的主效应，$F_{3,\ 69} = 4.6$，$p = 0.011$，$\eta_g^2 = 0.083$。采用 Bonferroni-Holm 校正进行事后两两比较后发现，*ARSelfie* 条件（$\mu = 42.8$，$\sigma = 18.6$）下的工作负荷显著低于条形图 条件（$\mu = 55.7$，$\sigma = 17.6$，$p = 0.009$）。基线条件（$\mu = 43.19$，$\sigma = 16.1$）下的总体工作负荷低于条形图 （$\mu = 55.7$，$\sigma = 17.6$）和人体模型 （$\mu = 46.5$，$\sigma = 17.9$）条件。然而，两两比较的差异并不显著。结果如图 11a 所示。

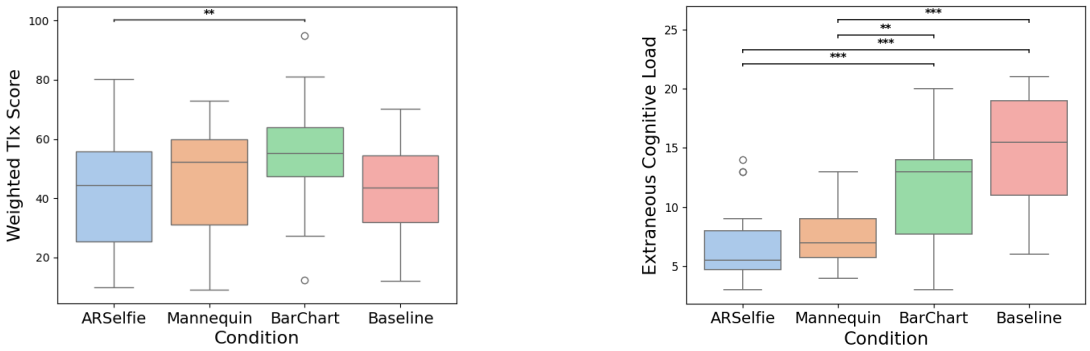
为了进一步探究任务负荷的来源，我们分析了 TLX 子量表中的六个单项子量表。结果如图 12 所示。



(a) 吸引力 (b) 明晰性 (c) 新颖性

(d) 刺激性 (e) 可靠性 (f) 效率

图 10。用户体验问卷的六个子量表 [43] 在各实验条件下的结果。须线表示显著的事后比较（* $p < 0.05$；**$p < 0.01$，*** $p < 0.001$）。得分越高越好，得分 > 0.8被视为正面，< -0.8被视为负面。

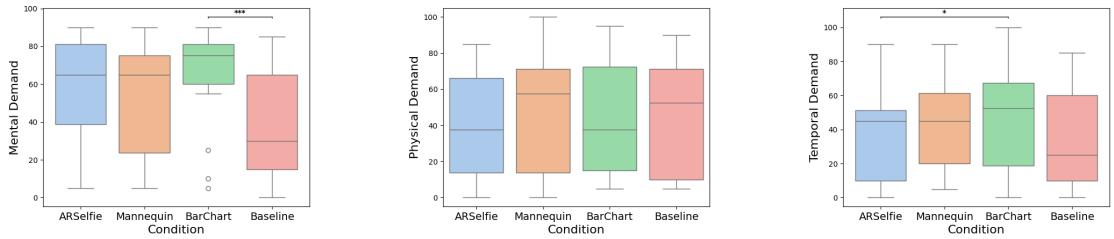


（a）加权 NASA-TLX 得分。 （b）外在认知负荷

图 11。（a）各条件下的加权 NASA-TLX 得分（越低越好）。事后比较显示，与*ARSelfie*相比，条形图 条件下的任务负荷显著更高。（b）四种可视化条件下的外在认知负荷得分（越低越好）。嵌入式条件对应的外在认知负荷显著低于条形图 和基线 条件。

5.4 外在认知负荷

外在认知负荷（ECL）由学习材料或应用程序的设计导致 [28, 51]。用户往往不得不将认知资源耗费在搜索信息、忽略无关信息等外在过程



(a) 心理需求 (b) 身体需求 (c) 时间需求

(d) 表现 (e) 努力程度 (f) 挫败感

图 12. 各可视化条件下的 NASA-TLX 子量表得分。除表现外，所有量表的得分越低越好。误差线表示显著的事后差异（* $p < 0.05$；** $p < 0.01$，*** $p < 0.001$）。

在学习或执行可能降低效率的任务时 [51]。我们使用 Klepsch等人 [28]所用问卷的修订版评估外在认知负荷。该问卷包含三个7点李克特量表问题（见表 1），Cronbach's $\alpha = 0.82$。外在认知负荷总分由三个条目得分相加得出。结果见图 11b。

表 1. 用于评估外在认知负荷的问卷。回答采用7点评分量表（1=强烈不同意，5=非常同意）。这些条目改编自 [28]，用于评估参与者从中获取所需信息的难度可视化内容，以便最佳地激活其肌肉。

1. 从该可视化内容中找到所需的肌肉激活信息令人筋疲力尽 2. 该可视化内容不便于验证肌肉是否得到最佳激活 3. 该可视化内容使人难以识别信息并将其与任务联系起来

弗里德曼检验显示，可视化条件对外在认知负荷（$Q = 21.78$，$W = 0.49$，$p < 0.001$）具有显著主效应。事后比较显示，与条形图（$\mu = 11.42$，$\sigma = 4.77$；$p = 0.001$以及 $p = 0.008$，分别）和基线（$\mu = 14.88$，$\sigma = 4.79$；$p < 0.001$ 两者均）相比，参与者在*ARSelfie*（$\mu = 6.50$，$\sigma = 3.16$）和人体模型（$\mu = 7.71$，$\sigma = 2.53$）条件下报告的外在认知负荷显著较低。此外，参与者评价人体模型 带来的外在认知负荷显著低于条形图（$p = 0.008$）。嵌入式条件之间未观察到显著差异。

6 定性结果

6.1 开放式问题

在每个可视化条件之后，参与者都以书面形式回答了以下问题：“你喜欢这个可视化的哪些方面。”完成整个研究后，他们以书面形式回答了以下问题：“总体而言，你更喜欢哪种可视化？为什么？。”最后，他们在与研究人员进行的半结构化访谈中，就不同可视化条件提供了开放式反馈。研究人员特别要求参与者详细说明他们喜欢这些可视化的哪些方面，以及为什么偏好其中一个条件而非另一个条件。研究人员使用归纳式主题分析对他们针对这些开放式问题的回答进行编码，并将其整理为不同主题 [7]。

主题 ***1***：认知与任务负荷。 参与者经常指出，反馈的位置及其视觉编码都会影响将其纳入任务所需的精力。

可视化位置。 条形图 条件要求参与者频繁在条形图和面部之间转移视线，许多人觉得这很费力。P6 提到：“上下看让评估状态变得更困难”，P23 则表示“不得不抬头低头看让我有点恼火”。其他人认为这些视线转移会分散注意力，例如 P10 表示“无法专注于面部，总是看反馈”，P11 则“最不喜欢这种（条形图 条件），因为我得看两个不同的地方”。另一方面，嵌入式条件中的同位反馈减少了视线转移。*ARSelfie* 条件下的 P24 提到，他们“不必把 [他们的] 注意力转移到屏幕的不同区域”，而人体模型 条件下的 P7 则强调，他们只需“关注一个区域”。

视觉编码。 参与者认为，条形图 条件下通过颜色变化呈现的比例条形填充提供了清晰且可量化的反馈。24 位参与者中有 3 位（P14、P22、P23）指出，条形图 让他们知道“ [他们] 需要移动多少”。嵌入式条件中的基于不透明度的编码则被认为不太直观。虽然 P5 认为这种编码有助于追踪激活情况，但 24 位参与者中有 5 位（P8、P14、P18、P22、P26）表示，这种编码更难解读，需要更加集中注意力。P18 解释说，“集中注意力观察不透明度 [要] 更费力，把注意力放在条形图的左侧/右侧更容易”，而 P14 总结道：“透明度比条形填充难理解得多。”

主题 ***2***：感知到的 表现与学习。 参与者指出了不同可视化各自的优势 .

学习。 24 位参与者中有 7 位（P7、P14、P17、P18、P19、P24、P26）认为，嵌入式条件更有利于学习如何控制面部肌肉。P7 认为，*ARSelfie* 条件“让人更容易逐渐学会需要施加多大的力”。P19 和 P24 都提到，在*ARSelfie*条件下看到自己的面部形态，有助于记住正确的动作。尽管 P18 总体上更喜欢另一种可视化方式（条形图），但他们表示，人体模型 条件“让人很容易学会哪块肌肉应该以什么方式运动”，而 P26 认为人体模型 条件“展示肌肉运动的方式最清晰，我可以据此学习并进行调整”。

相比之下，参与者表示在 条形图 条件下的学习程度有限，因为他们要么经常忽略肌肉标签，要么无法将其映射到自己面部的相应区域。P17 承认，他们“甚至都没读标签，只是看颜色”，而且“[完全] 没有学习”。同样，P24 提到他们“没有读标签”，因此“无法弄清楚 [反馈与面部肌肉之间的对应关系]”。一些参与者虽然阅读了肌肉标签，但仍难以进行映射。P3 提到自己“无法弄清楚 [需要改变] 哪个具体部位”，P26 也类似地表示，“很难将肌肉名称与《提交至 ACM 的稿件》中的特定面部区域对应起来

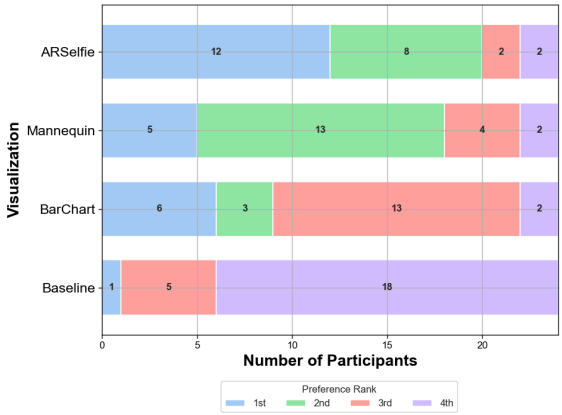


图 13. 四种可视化条件的用户偏好排名。*ARSelfie* 条件获得的第一名排名最多，而人体模型 最常被排在第二名。条形图 最常被排在第三名，基线条件则最不受青睐，获得的第四名排名最多。

面部区域”。即使理解了映射关系，注意力转移所需的额外努力也降低了反馈的实用性，并阻碍了学习。P6 回忆道，“将反馈映射到动作上需要时间和脑力”，这可能导致一些参与者（例如 P21）不再观察自己的面部，而只关注反馈。

表现。 许多人认为条形图 条件有助于即时表现。P18 指出，比例条形填充消除了“颜色／不透明度系统带来的模糊性”，让他们“更容易进行调整”。P14 表示，他们“喜欢能够看出自己离最佳激活有多近”，并观察到“自己的肌肉运动对得分产生了多大影响”[] 。另一方面，嵌入式条件中的反馈叠加层似乎让一些参与者更难进行调整。P21 提到，“滤镜让人稍微更难判断”自己的肌肉运动是否正确；P22 则表示，他们“发现由于叠加层的存在，很难将其与自己的面部状态联系起来”。

主题 ***3***：用户体验与偏好。 许多参与者更喜欢*ARSelfie* 条件，因为他们发现更容易将自己的动作与面部形态联系起来（P3、P5、P19、P20），或者只是因为他们更喜欢看自己的脸（P23、P24），而不是虚拟形象。不过，也有几位参与者（P4、P6、P21、P26）表示感到不适，原因可能是自我意识，也可能是受到干扰。P21提到，他们“ [不喜欢看]自己的脸……结果把注意力放在了五官上”，P4则表示，“看着这张脸会分散注意力，而且不想长时间看着这张脸”。

对一些参与者来说，人体模型 条件是*ARSelfie*的一种不会造成干扰的替代方案。P4提到，“它去掉了背景，让人更容易专注于面部表情”。P6、P9和P21都表示，这种方式“没有干扰”，让他们更容易专注于练习。不过，也有少数参与者认为人体模型 条件“显得不自然”（P11、P19）、“互动性较弱”（P23），甚至“令人毛骨悚然”（P20）。

6.2 偏好

汇总排名（见图13）从最受偏好到最不受偏好依次为：*ARSelfie*、人体模型、条形图和基线。*ARSelfie* 条件最受偏好，24 位参与者中有 20 位将其排在前两位；其次是人体模型 条件（24 位中有 18 位）。相比之下，条形图 （24 位中有 9 位）和基线 条件

被排在前两位的频率低得多。弗里德曼检验证实，可视化条件对偏好排序的影响具有统计显著性， $Q = 29.75$, $p < 0.001$ 且效应量中等至较大（Kendall's $W = 0.41$）。采用邦费罗尼校正进行的事后两两比较显示，基线条件的排名显著低于*ARSelfie*、人体模型 和条形图 条件。其他成对比较均未达到统计显著性。

7 讨论与设计启示

我们的研究结果表明，反馈的空间放置会显著影响面部练习中的工作负荷、用户体验和表现。我们据此解读这些结果，并提出设计未来面部锻炼系统的策略。

7.1 面部的嵌入式反馈?

我们的定量和定性结果表明，此前已在全身动作中得到验证的嵌入式反馈优势，同样适用于面部动作。NASA-TLX 和外在认知负荷结果表明，在嵌入式条件（*ARSelfie* 和 人体模型）下，视觉线索与动作区域在空间上的对齐，使用户能够直接将反馈与目标面部肌肉关联起来。这可能会减少反馈与目标之间的空间转换需求 [31]，从而相比条形图 和 基线 条件降低任务负荷和认知负荷（**RQ2**）。相比之下，条形图 条件对认知要求较高，因为用户需要在脑中将呈现在周边位置的反馈转换到自己的面部上，才能做出调整。这种空间分离可能增加了反馈与其指涉对象之间的认知距离 [27]，使分散注意力效应更容易出现 [14]。与 *ARSelfie* 条件相比，条形图条件下 NASA-TLX 时间需求子量表的得分显著更高，这进一步表明，在面部与反馈之间转移注意力，加上空间转换过程，让参与者感到更加仓促。尽管参与者在使用条形图 条件时激活肌肉的准确度更高，但这种表现提升很可能是以牺牲学习效果并增加认知努力为代价的。嵌入式可视化在明晰性和效率方面得分更高，这进一步表明，空间上相契合的反馈更容易整合。这与空间连续性原则 [36]相符。

面部练习反馈系统的设计者应采用与目标肌肉直接对齐的空间一致可视化，因为这类可视化能够降低施加给用户的额外认知与任务负荷（**研究问题 2**）。虽然情境化反馈方式有助于用户做出准确动作（**研究问题 1**），但设计者必须留意注意力转移和空间转换所带来的认知要求。总体而言，这些发现表明，为全身或上肢训练建立的嵌入式反馈设计原则，也可推广至面部练习。

7.2 自拍的权衡

总体而言，参与者更倾向于嵌入式条件，而非条形图 和 基线 条件（**研究问题 3**）；在用户体验方面，*ARSelfie*条件的评分最高，也是整体上最受偏好的条件，这表明空间对齐反馈对面部动作任务是有效的。然而，定性反馈揭示了设计嵌入式反馈时需要权衡的重要因素。尽管参与者认可嵌入式条件的空间优势，但他们对自拍视图的看法差异很大。有些人觉得在面部显示视觉叠加层具有互动性且直观，但另一些人则觉得看到自己的面部会分散注意力，令人不适。

这些截然不同的反应与自我聚焦注意相关研究的结论一致：人们天生会关注自己的面部，而这可能凌驾于自上而下注意控制之上，进而导致认知耗竭 [8,57]。自我评价、镜像焦虑和“Zoom疲劳”可能会加剧这一问题；持续关注自己的

形象可能引发负面的自我认知，并耗尽注意力资源 [3, 10]。在这种状态下，用户的注意力可能会被自己感知到的缺陷吸引，从而实际上对增强现实反馈本身视而不见 [30]。这些顾虑可能在一定程度上抵消了空间对齐对部分参与者带来的益处。

人体模型 条件通过提供一个个人色彩较弱的替代对象缓解了这一问题。虽然这种方法确实减轻了部分人的自我意识，但也有人将虚拟形象描述为“不自然”或“令人毛骨悚然”，表明其对恐怖谷效应较为敏感 [37]。

我们的研究结果揭示了一个核心的设计权衡。虽然将反馈叠加在面部上（无论是用户自己的面部还是虚拟形象的面部）可能提供更优的空间引导，但通过兼顾自我表征舒适度方面的个体差异，可以进一步提升其效果。

我们建议嵌入式面部反馈系统纳入灵活的自我表征选项，以最大化空间优势，同时尽量减少心理障碍。这可以包括渐进式的抽象层级（从风格化虚拟化身到照片级真实）、可定制的虚拟形象特征以减少恐怖谷效应、使用喜爱的虚构角色作为替代形象，或将抽象叠加层与最少面部关键点相结合的混合方法。这样可以让用户充分受益于嵌入式反馈的优势。

7.3 感知清晰度

虽然我们的结果表明，嵌入式反馈的益处也适用于面部，但其有效性高度依赖于线索的感知清晰度。

首先，参与者难以解读通过不透明度调制传达的纠正调整幅度。这种编码限制造成了偏好与表现之间的权衡：尽管用户更偏好嵌入式条件的空间优势，但借助条形图 条件中明确的幅度表示，他们能够更准确地激活面部肌肉。其次，AR叠加层本身引入了一些视觉杂乱。部分参与者发现，半透明的面部滤镜遮挡了他们试图控制的面部肌肉视野。这种视觉干扰可能不利于学习，因为清晰的视觉反馈对于培养准确的运动控制至关重要，尤其是在技能习得阶段 [41]。

因此，我们建议设计师将空间嵌入式反馈与明确的幅度指示器相结合。虽然空间线索能够有效指示用户应集中注意的位置，但还应辅以具有良好感知效果的线索，例如对肌肉区域进行比例填充、使用轮廓线，或在文本中显示数值百分比，从而清楚地传达需要进行多大幅度的调整才能达到目标。这些滤镜必须经过谨慎设计，以便提供反馈，同时不遮挡用户对自身面部的视野。

7.4 反馈几何形态

条形图 条件使用形状和尺寸一致的二维条形来表示所有肌肉，而嵌入式条件则使用贴合各块面部肌肉自然三维轮廓的叠加层。这一差异可能影响了参与者对反馈的感知和解读。条形填充可能因几何形态一致而更具优势，因为各块肌肉的变化在视觉上保持一致，因此更容易比较。但这需要频繁转移注意力。

相比之下，叠加层提供了空间对齐反馈，使参与者更容易将其与正在锻炼的肌肉联系起来，但这些叠加层的形状和范围各不相同。较大肌肉（如额头肌肉）的变化可能更为显眼，而小肌肉或薄肌肉叠加层的变化则较为细微，难以察觉。不同肌肉的反馈显著性不均，这给面部嵌入式反馈带来了独有的挑战，

虽然空间对齐能够减少操作负担并提升用户体验，但叠加层的几何变化可能造成感知失衡。

在面部锻炼系统中使用嵌入式反馈时，设计者应通过调整亮度或对比度等视觉参数来增强较小或不太显眼的肌肉所获得的反馈，从而确保感知显著性均衡。他们也可以使用细微的轮廓线或字形符号等辅助指示器，使各块肌肉的显著性保持一致。

8 局限性与未来工作

我们的研究表明，与情境化可视化相比，空间对齐的嵌入式可视化能够有效支持面部运动训练，降低认知负荷并提升用户体验。不过，若干局限性使我们的研究发现需要谨慎解读，也指明了未来的研究方向。

首先，研究在受控的实验室环境中进行，历时 45–60 分钟，重点考察了涉及 11 块不同面部肌肉的三项练习。虽然这一研究范围有助于对不同可视化条件进行有针对性的比较，但嵌入式反馈在支持面部动作学习与再训练方面的长期效果仍有待探究。未来研究应通过被试间研究评估各可视化条件带来的学习增益，并探索纵向使用、更广泛的锻炼项目组合以及生态效度较高的环境，以评估持续效果。

其次，我们的评估仅关注视觉反馈。即使在较短的训练中，一些参与者也表示，在*ARSelfie*条件下反复观看自己的面部会感到不适。在临床或康复环境中，这种不适可能会加剧，从而降低参与意愿。未来的系统应更深入地研究基于虚拟形象或代理反馈策略，并探索听觉或触觉提示等非视觉模态，尤其是在无法或不适合反复观看面部的情境下。

第三，我们的反馈流程采用 ARCore 面部网格，通过手动映射将网格顶点与深层肌肉组织对应起来，并使用启发式方法，根据检测到的面部动作推断肌肉激活状态；这两项均经面部解剖专家验证。虽然这种映射具有解剖学依据，但它仍是肌肉激活的间接代理，无法捕捉募集强度或神经肌肉协调等生理信号。在将这一基于 ARCore 的肌肉激活检测系统及其底层启发式方法用于现实环境中的面部肌肉训练应用之前，应以肌电图（EMG）等真实生理测量值对其进行验证，以确保准确性和可靠性。未来的工作还可以采用更先进的算法（例如 [34] ），直接根据图像数据预测肌肉激活；整合互补的传感模态（如 EMG）；或结合面部解剖结构的生物力学模型，从而提供更精确且可能具有诊断性的反馈。

最后，我们的参与者样本规模相对较小，年龄分布也较为单一，这限制了研究结果的可推广性。未来的研究应招募规模更大、背景更多元的群体，包括临床群体，以及表达规范或文化背景各异的个体，以便更好地评估我们设计见解的适用性与包容性。

9 结论

在本研究中，我们探讨了视觉反馈的空间位置如何影响面部运动训练中的表现、认知负荷和用户体验。我们在一项被试内研究（N=24）中比较了三种可视化技术：情境式（条形图）、代理嵌入式（人体模型）和嵌入式（ARSelfie），以及一种基线条件：参与者可以通过移动设备的前置摄像头看到自己的面部，但不会收到关于面部肌肉激活的额外反馈。结果表明，空间对齐反馈能够降低认知负担、提升用户体验，并且通常比情境式替代方案更受欢迎。稿件已提交至 ACM

情境化反馈虽然有助于提高准确性，但由于注意力转移以及需要进行空间转换，会带来更高的认知负荷。基于这些发现，我们为未来基于增强现实的面部反馈系统提出了设计启示，这些启示对康复、表现训练和技能习得均具有参考意义。

参考文献

[1] Mohammed F. Alhamid、Mohamad Eid 和 Abdulmotaleb El Saddik。2012。用于生物反馈交互的多模态智能系统。载于《2012年IEEE医学测量与应用国际研讨会论文集》。1–5。https://doi.org/10.1109/MeMeA.2012.6226653[2] Fraser Anderson、Tovi Grossman、Justin Matejka 和 George Fitzmaurice。2013。YouMove：利用增强现实镜提升动作训练效果。载于第26届ACM用户界面软件与技术研讨会论文集。311–320。[3] Jeremy N Bailenson。2021。非语言超载：关于Zoom疲劳成因的理论论述。（2021）。[4] Giuliana Barrios Dell'Olio 和 Misha Sra。2021。FaraPy：一种利用动作单元强度估计的面瘫增强现实反馈系统。载于第34届ACM用户界面软件与技术年度研讨会。ACM，美国虚拟会议，1027–1038。https://doi.org/10. 1145/3472749.3474803[5] Matthew Berry 和 Steven Brown。2022。动态面具：专业演员角色塑造的面部相关特征。《实验心理学季刊》75，5（2022年5月），936–953。https://doi.org/10.1177/17470218211047935 出版方：SAGE Publications。[6] Susana Bloch。1993。Alba Emoting：帮助演员创造和控制真实情绪的心理生理技术。《戏剧专题》3，2（1993），121–138。[7] Virginia Braun 和 Victoria Clarke。2006。在心理学中使用主题分析。《心理学定性研究》3，2（2006），77–101。[8] Serge Brédart、Marie Delchambre 和 Steven Laureys。2006。短文：自己的脸难以忽视。《实验心理学季刊》59，1（2006），46–52。[9] Xueyan Cai、Yingjing Xu、Zihong Zhou、Mengru Xue、Zhengke Li、Chentian Weng、Wei Luo、Cheng Yao、Bo Lin 和 Jianwei Yin。2025。"打破面具屏障"：面向帕金森病面部表情减少症的基于动作单元的康复训练系统。载于计算机系统中的人为因素会议扩展摘要集（CHI EA '25）。美国计算机协会，纽约，NY，美国，1–9。https: // doi.org/10.1145/3706599.3719706[10] Alecia Carolli。2023。Zoom使用与认知负荷。博士论文。约克大学。[11] Polona Caserman、Shule Liu 和 Stefan Göbel。2021。沉浸式虚拟现实健身游戏中的全身动作识别。《IEEE游戏汇刊》14，2（2021），243–252。[12] Luigi Cattaneo 和 Giovanni Pavesi。2014。面部运动系统。《神经科学与生物行为评论》38（2014年1月），135–159。https: // doi.org/10.1016/j.neubiorev.2013.11.002[13] Luigi Ceccaroni 和 Xavier Verdaguer。2004。Magical Mirror：家庭自动化中的多媒体交互式服务。载于个性化信息访问环境研讨会论文集。10–21。[14] Paul Chandler 和 John Sweller。1992。分散注意效应在教学设计中的作用。《英国教育心理学杂志》62，2（1992），233–246。[15] Lei Chen、Chee Wee Leong、Gary Feng、Chong Min Lee 和 Swapna Somasundaran。2015。利用多模态线索自动评估公开演讲表现。载于2015年情感计算与智能交互国际会议（ACII）。394–400。https://doi.org/10. 1109/ACII.2015.7344601 ISSN：2156-8111。[16] Alana Elza Fontes Da Gama、Thiago Menezes Chaves、Lucas Silva Figueiredo、Adriana Baltar、Ma Meng、Nassir Navab、Veronica Teichrieb 和 Pascal Fallavollita。2016。MirrARbilitation：一种用于增强现实康复系统、与临床相关的手势识别交互工具。《生物医学计算方法与程序》135（2016），105–114。[17] Florian Diller、Nico Henkel、Gerik Scheuermann 和 Alexander Wiebel。2025。SkillAR：利用增强现实进行运动技能训练的无处不在的现场反馈。《虚拟现实》29，1（2025年2月），33。https://doi.org/10.1007/s10055-025-01108-1[18] Paul Ekman 和 Erika L Rosenberg。1997。《面部揭示的内容：运用面部动作编码系统（FACS）对自发性表情的基础与应用研究》。牛津大学出版社，美国。[19] Paul M Fitts 和 Michael I Posner。1967。人类表现。（1967）。[20] Thomas Franke、Christiane Attig 和 Daniel Wessel。2019。技术交互的个人资源：技术交互亲和度（ATI）量表的开发与验证。《国际人机交互杂志》35，6（2019），456–467。[21] Miharu Fuyuno、Rinko Komiya 和 Takeshi Saitoh。2018。英语作为外语学习者公开演讲表现的多模态分析：应用深度学习理解成功演讲者如何运用面部动作。《亚洲应用语言学杂志》5，1（2018年4月），117–129。https://caes.hku.hk/ajal/index.php/ajal/article/view/508 编号：1。[22] Gissela M. Guanoluisa、Jimmy A. Pilatasig 和 Víctor H. Andaluz。2019。GY MEDIC：面瘫患者分析与康复系统。载于《知识建模与决策中的综合不确定性》，Hirosato Seki、Canh Hao Nguyen、Van-Nam Huynh 和 Masahiro Inuiguchi（编）。施普林格国际出版公司，查姆，63–75。https://doi.org/10.1007/978-3-030-14815-7_6

[23] Sandra G. Hart 和 Lowell E. Staveland。1988。NASA-TLX（任务负荷指数）的开发：实证与理论研究结果。载于《心理学进展》，Peter A. Hancock 和 Najmedin Meshkati（编）。《人的心理工作负荷》，第52卷。North-Holland，139–183。https: // doi.org/10.1016/S0166-4115(08)62386-9[24] Koki Hasebe、Tsuyoshi Kojima、Yusuke Okanoue、Ryohei Yuki、Hirotaka Yamamoto、Shuya Otsuki、Shintaro Fujimura 和 Ryusuke Hori。2024。利用iPhone上的三维人脸识别系统评估面神经麻痹的新方法。《Auris Nasus Larynx》51，3（2024年6月），460–464。https://doi.org/10.1016/j.anl.2024.02.003[25] Pedro Henriquez、Bogdan J. Matuszewski、Yasmina Andreu-Cabedo、Luca Bastiani、Sara Colantonio、Giuseppe Coppini、Mario D'Acunto、Riccardo Favilla、Danila Germanese、Daniela Giorgi、Paolo Marraccini、Massimo Martinelli、Maria-Aurora Morales、Maria Antonietta Pascali、Marco Righi、Ovidio Salvetti、Marcus Larsson、Tomas Strömberg、Lise Randeberg、Asgeir Bjorgan、Giorgos Giannakakis、Matthew Pediaditis、Franco Chiarugi、Eirini Christinaki、Kostas Marias 和 Manolis Tsiknakis。2017。镜子，墙上的镜子……一种用于健康状态自我评估与可视化的非侵入式智能多感官镜。《IEEE多媒体汇刊》19，7（2017年7月），1467–1481。https://doi.org/10.1109/TMM.2017.2666545[26] Atsuki Ikeda、Dong-Hyun Hwang、Hideki Koike、Gerd Bruder、Shunsuke Yoshimoto 和 Sue Cobb。2018。采用衰减动态时间规整算法的增强现实自主运动学习系统。。载于ICAT-EGVE。171–174。[27] SeungJun Kim 和 Anind K Dey。2009。作为老年驾驶员导航认知地图辅助工具的模拟增强现实挡风玻璃显示。载于SIGCHI计算系统中的人为因素会议论文集。133–142。[28] Melina Klepsch、Florian Schmitz 和 Tina Seufert。2017。用于测量内在、外在和相关认知负荷的两种工具的开发与验证。《心理学前沿》8（2017年11月）。https://doi.org/10.3389/fpsyg.2017.01997 出版方：Frontiers。[29] Felix Kosmalla、Florian Daiber、Frederik Wiehr 和 Antonio Krüger。2017。Climbvis：通过示范研究用于理解攀爬动作的现场可视化。载于2017年ACM交互式表面与空间国际会议论文集。270–279。[30] Carina Kreitz、Philip Furley、Daniel Memmert 和 Daniel J Si mons。2015。认知能力中的非注意盲视与个体差异。《PLOS ONE》10，8（2015），e0134675。[31] Axel Larsen 和 Claus Bundesen。1998。视觉模式匹配中空间分隔的影响：关于心理转换作用的证据。《实验心理学杂志：人类感知与表现》24，3（1998），719–731。https://doi.org/10.1037/0096-1523.24.3.719 地点：美国 出版方：美国心理学会。[32] Bettina Laugwitz、Theo Held 和 Martin Schrepp。2008。用户体验问卷的构建与评估。载于《教育与工作中的人机交互及可用性》，Andreas Holzinger（编）。施普林格，柏林，海德堡，63–76。https://doi.org/10.1007/978-3-540-89350-9_6[33] Benjamin Lee、Michael Sedlmair 和 Dieter Schmalstieg。2023。增强现实中情境化可视化的设计模式。《IEEE可视化与计算机图形学汇刊》（2023），1–12。https://doi.org/10.1109/TVCG.2023.3327398[34] RUOFAN LIU、YICHEN PENG、Takanori Oku、Chen-Chieh Liao、Erwin Wu、Shinichi Furuya 和 Hideki Koike。2025。从姿势到肌肉：钢琴手部肌电图的多模态学习。载于神经信息处理系统第三十九届年会。https://openreview.net/forum?id=ftZEltGArK[35] Tania Marur、Yakup Tuna 和 Selman Demirci。2014。面部解剖结构。《皮肤病学临床》32，1（2014年1月），14–23。https://doi.org/10.1016/j. clindermatol.2013.05.022[36] Richard E Mayer 和 Logan Fiorella。2014。减少多媒体学习中外在加工的12项原则：连贯性、信号提示、冗余、空间邻近和时间邻近原则。《剑桥多媒体学习手册》279（2014），279–315。[37] Masahiro Mori、Karl F MacDorman 和 Norri Kageki。2012。恐怖谷 [来自实地]。《IEEE机器人与自动化杂志》19，2（2012），98–100。[38] Mark Nestor、Glynis Ablon 和 Andy Pickett。2017。美容领域使用A型肉毒杆菌毒素的关键参数：起效时间与持续时间。《美容外科杂志》37，增刊_1（2017年5月），S20–S31。https://doi.org/10.1093/asj/sjw282[39] Donald A Norman。1986。认知工程。载于《以用户为中心的系统设计》。CRC出版社，31–62。[40] Sebastian Rigling、Xingyao Yu 和 Michael Sedlmair。2023。"就在你脸上！"：在增强现实中可视化健身追踪器数据。载于2023年计算机系统中的人为因素会议扩展摘要集。ACM，德国汉堡，1–7。https://doi.org/10.1145/3544549.3585912[41] Christelle Robin、Lucette Toussaint、Yannick Blandin 和 Annie Vinter。2004。感觉统合在学习瞄准"自我定义"目标中的作用。《运动与锻炼研究季刊》75，4（2004），381–387。[42] Richard A Schmidt 和 Craig A Wrisberg。2008。《运动学习与表现：基于情境的学习方法》。Human Kinetics。[43] Martin Schrepp、Jörg Thomaschewski 和 Andreas Hinderks。2017。用户体验问卷（UEQ）基准的构建。（2017）。[44] Yoones A Sekhavat 和 Mohammad S Namani。2018。基于投影的增强现实：步态康复中的有效视觉反馈。《IEEE人机系统汇刊》48，6（2018），626–636。[45] Todd K Shackelford 和 Randy J Larsen。1997。面部不对称是心理、情绪和生理困扰的指标。《人格与社会心理学杂志》72，2（1997），456。[46] Lun Shu、Victor R. Barradas、Zixuan Qin 和 Yasuharu Koike。2025。通过肌肉协同实现面部表情识别，并利用面部表面肌电信号和皮肤-肌肉骨骼模型估算面部关键点位移。《生物工程与生物技术前沿》13（2025年2月）。https://doi.org/10.3389/fbioe.2025.1490919 出版方：Frontiers。[47] Roland Sigrist、Georg Rauter、Robert Riener 和 Peter Wolf。2013。增强视觉、听觉、触觉及多模态反馈在运动学习中的应用：综述。《心理学通报与评论》20，1（2013），21–53。已提交给ACM的稿件

[48] Rajinder Sodhi、Hrvoje Benko 和 Andrew Wilson。2012。LightGuide：用于手部运动引导的投影视觉化。载于*SIGCHI*计算系统中的人为因素会议论文集（*CHI '12*）。美国计算机协会，纽约，纽约州，美国，179–188。https://doi.org/10.1145/2207676.2207702[49] Maurício Sousa、João Vieira、Daniel Medeiros、Artur Arsenio 和 Joaquim Jorge。2016。SleeveAR：利用实时反馈实现的康复增强现实。载于第*21*届智能用户界面国际会议论文集。ACM，美国加利福尼亚州索诺玛，175–185。https://doi.org/10.1145/2856767.2856773[50] Hariharan Subramonyam。2015。SIGCHI：Magic Mirror——面向量化自我的具身交互。载于第*33*届*ACM*计算系统中的人为因素年会扩展摘要集。ACM，韩国首尔，1699–1704。https://doi.org/10.1145/2702613. 2732884[51] John Sweller。1994。认知负荷理论、学习难度与教学设计。学习与教学 4，4（1994），295–312。[52] Richard Tang、Xing-Dong Yang、Scott Bateman、Joaquim Jorge 和 Anthony Tang。2015。Physio@ Home：探索物理治疗练习的视觉引导与反馈技术。载于第*33*届*ACM*计算系统中的人为因素年会论文集。4123–4132。[53] Milka Trajkova 和 Francesco Cafaro。2018。Takes Tutu to ballet：为增强镜设计视觉与语言反馈。*ACM*交互、移动、可穿戴与泛在技术论文集 2，1（2018），1–30。[54] Sean Michael White。2009。情境化可视化的交互与呈现技术。哥伦比亚大学。[55] Wesley Willett、Yvonne Jansen 和 Pierre Dragicevic。2016。嵌入式数据表示。*IEEE*可视化与计算机图形学汇刊23，1（2016），461–470。[56] Jacob O Wobbrock、Leah Findlater、Darren Gergle 和 James J Higgins。2011。仅使用方差分析程序进行非参数析因分析的对齐秩变换。载于*SIGCHI*计算系统中的人为因素会议论文集。143–146。[57] Michał J Wójcik、Maria M Nowicka、Ilona Kotlewska 和 Anna Nowicka。2018。自我面孔会捕获、保持并偏向注意力。心理学前沿8（2018），2371。[58] Fan Xu、Xian-wei Zou、Li-qiong Yang、Shi-cong Mo、Quan-hao Guo、Jing Zhang、Xiechuan Weng 和 Guo-gang Xing。2022。帕金森病患者在发声测试中的面部肌肉运动。神经病学前沿 13（2022年10月）。https://doi.org/10.3389/fneur.2022.1018362 出版方：Frontiers。[59] Xingyao Yu、Benjamin Lee 和 Michael Sedlmair。2024。基于XR的运动引导系统中的视觉前馈与纠正性反馈设计空间。载于*CHI*计算系统中的人为因素会议论文集。ACM，美国夏威夷州檀香山，1–15。https://doi.org/10.1145/ 3613904.3642143[60] Xingyao Yu、David Rosin、Johannes Kässinger、Benjamin Lee、Frank Dürr、Christian Becker、Oliver Röhrle 和 Michael Sedlmair。2024。PerSiVal：生物力学手臂仿真的身体穿戴式增强现实可视化。*IEEE*计算机图形学与应用 44，6（2024年11月），24–38。https: //doi.org/10.1109/MCG.2024.3494598[61] Junyi Zhu、Yuxuan Lei、Aashini Shah、Gila Schein、Hamid Ghaednia、Joseph Schwab、Casper Harteveld 和 Stefanie Mueller。2022。MuscleRehab：通过监测和可视化肌肉参与度，改进无监督物理康复。载于第*35*届*ACM*用户界面软件与技术研讨会论文集。1–14。

收稿日期：2007年2月20日；修订日期：2009年3月12日；接受日期：2009年6月5日