

第 12 章 VR 系统和体验评估

哪个头戴式设备更好？哪些 VR 体验在长时间内更舒适？多少体验的视野是足够大的？什么是最合适的交互机制？工程师和开发人员想知道这些问题的答案；然而，需要明确的是，由于人类生理学和感知在工程系统中操作和交互的方式，这些问题还难以回答。相比之下，纯粹的工程问题，比如“电池寿命估计值是多少？”或“车辆在平面上的最高速度是多少？”，这些问题更容易理解和解答。

回想第 1.1 节涉及生物体的 VR 定义。当科学家应用 VR 来研究老鼠的神经结构和感知时，老鼠和科学家之间是有明确界限的。但是，对于 VR 来说，开发人员会经常尝试自己的想法。在这种情况下，开发人员应在科学家和老鼠的角色之间切换。这就带来了很多问题，尤其是当开发人员对感知问题还不够熟悉的情况下。

类似的适应问题进一步复杂化，会发生在各个方面。例如，几个星期内多次评估 VR 体验的人可能在最初觉得不舒服，但后来会习惯。当然，这并不意味着使新用户患病的可能性变低。此外，人与人之间的差异也很大。任何一个人，包括开发人员，都只提供一个或一组数据。对晕眩免疫的人在使用这种系统时则很有可能没有任何阻碍。

另一个因素是，大多数搭建系统的人都更偏爱他们自己创造的东西。此外，正如 8.4 节所讨论的那样，只要知道体验的具体内容就会影响到最终的观点。这些几十年来一直在研究的问题统称为人为因素。一个与之密切相关的领域是人机交互 (HCI)，它也使用本节讨论的方法。然而，由于 VR 通过破坏我们日常所信赖的感官系统的低级操作来工作，所以由 VR 产生的最低级别副作用到最高级别认知效果的并发症似乎是前所未有的。

从硬件，低级软件到内容创建引擎的各个层面，都存在结果不佳的可能性。随着硬件和低级软件的迅速发展，引擎和 VR 体验的开发人员的负担也越来越重。本章将介绍几个有助于工程师和开发人员寻求建立更好的 VR 系统和体验的主题。第 12.1 节介绍了指导他们提高区分能力的方法。开发者可以培养自己对问题的敏感性，而不是因为适应性而忽略一些重要问题。第 12.2 节应用本书的基础知识为 VR 开发人员提供简单的建议。第 12.3 节介绍了 VR 晕眩，包括主要症状和原因，以便改善 VR 系统和体验。第 12.4 节介绍了设计关于人类受试者实验的一般方法，并包括一些来自心理物理学的具体方法。本章中的所有概念都是为了取得实质性的反馈，并避免在 VR 开发过程中掉入陷阱。

12.1 感知训练

大部分第一次尝试 VR 的人都不会知道经验丰富的工程师和开发人员会有很明显的技术上的缺陷。虚拟现实体验的最佳情况应该如此：用户被主导的视觉刺激所淹没，并感觉自己仿佛置身于虚拟世界。可以存在一些微小的缺陷，因为注意力主要集中在目标体验上（如 1.1 节中 VR 的定义所考虑的那样）。有些部分可能没有按照设计运行，或者一些感知问题可能被忽略。但这样的问题也可以被改善，并产生更好的体验。更糟的情况是，这些缺陷可能会导致用户疲劳或生病。最后，这些用户也无法意识到产生了什么问题。他们可能会追究到任何一点，比如特定的视觉刺激，特殊的体验，耳机硬件，甚至 VR 的整体概念。

通过培训特定的用户和开发人员来关注常见的缺陷类型，可以一定程度上缓解此问题。通过开发感知训练计划，可以要求用户寻找特定的缺点，或者反复练习来执行一些任务。在整本书中，我们已经看到适应在人类感知过程中的重要性。例如，如果长时间产生持续刺激，则感知强度会减弱。

通过对特定的 VR 系统和体验进行反复和引导性的训练，用户可以调整他们的感知系统。这是感知学习的一种形式，它是感知心理学的一个分支，它研究生物体因环境而引起的感知系统的长期变化。由于虚拟现实成为一种新环境，学习的机会和局限性在很大程度上仍然未被探索透彻。通过主动训练，可以控制用户适应的方式，从而增加他们的感知能力和

辨别能力。这也可以用于在开发过程中提供频繁反馈的评估人员。另一种方法是开发一个可以在没有人为干预的情况下自动检测缺陷的系统。人为和自动评估的结合在未来几年中很可能非常重要。

感知学习的例子

在日常生活中，我们会遇到许多关于感官学习的例子。关于视力，医生和医疗技术人员接受培训，从图像中提取相关信息，这些信息对未经训练的人来说非常难以捕捉。癌症专家可以在 CT 和 MRI 扫描中发现肿瘤。产科医生可以毫不费力地从手持式超声扫描仪中确定胎儿结构是否正常发育。关于听力，音乐家在广泛的练习之后学会区分和分类各种音符。由于录音，压缩，扬声器和室内声学问题，发烧音乐者会注意到音乐再现中的特殊缺陷。关于味道和嗅觉，侍酒师学习辨别葡萄酒之间的微妙差异。关于触摸，盲人学会阅读盲文，这是用指尖感觉到的凸起点的微小图案完成的。所有这些例子对于新手来说似乎都是不可能的，因为看起来我们甚至没有完成它的神经条件。尽管如此，通过建立感知训练计划和反复练习，人们可以获得惊人的感知能力。所以，为什么不可以对 VR 进行评估呢？

感知学习因素和机制

当这些形式的学习发生时，人类的观察系统会发生什么？一个重要的因素是神经可塑性，它使人类的大脑能够开发专门的神经结构作为对环境刺激的适应。虽然这对小孩子来说要强得多，正如学习母语所展示的那样，但神经可塑性仍然在成年人中存在；人与人之间最终的神经可塑性结果可能不同。

另一个因素是学习的方式。随意观察或专注于刺激的针对性策略均可能会导致适应。学习发生的时间和重复性可能会有很大不同，具体取决于任务，性能要求，刺激和人员。此外，还有可能会给他/她提供有监督的培训，在他/她试图提高自身表现时直接提供反馈。又或者，可能会出现无监督的训练，训练员已经在学习环境中放置了足够的刺激，但不会干扰学习过程。

四种基本机制已经被开发出来来解释感知学习。

1. 注意力权衡：与任务相关的功能的注意力增加，同时减少对其他功能的注意力。
2. 刺激印迹：开发专门的受体来识别部分或全部相关刺激。这些可能是神经结构的功能。
3. 差异化：曾经融合在一起的不同刺激共同激活了知觉。因而细微的差异需要被放大。
4. 单元化：这个过程将许多不同的刺激结合或压缩成一个响应。这与差异化形成鲜明对比，且因与单元内部差异关联不大，对分类很有帮助。



图 12.1：蝴蝶出现在左眼的图像中，但相应右眼图像中没有蝴蝶。

本节的其余部分将在 VR 的背景下提供示例和有用的建议。该领域远离标准的感知训练课程，类似于医学影像或音乐训练。相反，我们提供关于如何移动以及在何处关注焦点的建议，同时尝试发现 VR 体验中的错误认知。这就要求人们继续意识到人为刺激造成的干扰，这与第 1.1 节所述的 VR 的定义并不完全相符。

立体声问题

图 12.1 显示了一个简单的错误，其中蝴蝶出现在场景中，而另一个中并没有。虚拟世

界的其余部分被正确渲染。对于未经训练的人眼,这可能完全不会被注意到。解决方案则是:闭左眼,同时保持右眼张开;之后,切换到左眼张开,右眼闭合。通过在打开一只眼睛之间来回切换,不匹配现象应该变得清晰。这种方法被称为闭眼法。

另一个常见错误是将右眼和左眼图像颠倒过来。(3.50)中出现符号错误后,很容易出现这个问题,或者误解了视点需要为每只眼睛转换的方式。这种现象被称为视觉幻觉,其中物体的感知凹面可能被颠倒过来。但是,在很多情况下,这种现象很难直观地检测到。解决方法:接近物体的边缘,使其一侧仅对一只眼睛可见。这可以通过使用闭眼法来验证。根据物体的几何形状,确保侧面被正确可见。

最后,立体视觉可能在虚拟瞳孔之间具有不正确的距离((3.50)中的 t 参数)。如果 $t=0$,则可以使用闭眼技巧来检测两个图像看起来是相同的。如果 t 太大或太小,则深度和尺度感知(第6.1节)会受到影响。更大的分离 t 会导致整个环境变得更小;较小的 t 会导致相反的结果。

典型的头部运动

现在考虑涉及到移动的错误,这可能是由于头部跟踪错误,渲染角度或某种组合引起的。做出谨慎的,可重复的运动是有益的,这被称为典型的头部运动。如果单独跟踪旋转,则有三个旋转自由度。要发现各种运动或视点错误,应对评估者进行培训,以仔细执行各自的基本旋转。纯偏航可以通过“否”来执行。纯俯仰表现为“是”。纯翻滚状态则更难以完成,这涉及到来回转动头部以使一只眼睛在极端情况下高于另一只眼睛。在任何这些运动中,尽可能少地转换环视点(双眼中心点)或采用9.1节头部模型引起的转换是有帮助的。

对于每一种基本旋转,评估人员都应该练习以各种恒定的角速度和幅度来执行它们。例如,评估人应该试着以恒定的速度慢慢地摇头,每个方向45°。或者,她应该尝试以很快的速度旋转,每个方向可以达到10°,频率为2 Hz。使用规范的头部运动,可以确定图9.7中给出的常见错误。其他问题,例如跟踪中的不连续性,倾斜误差,等待时间以及视点的不正确深度也可以通过这种方式被更容易地检测到。

如果对位置进行跟踪,那么三种甚至更多典型的头部运动会变得重要,每个位置都会有一个DOF。因此,可以执行水平,垂直和深度变化运动来辨别存在的问题。例如,对于水平的左右运动,可以确定运动视差是否提供了正确的提示。

VOR vs 平滑移动

回顾5.3节,5.4节和6.2节,眼动在视觉感知中起着重要作用。评估人员在评估虚拟世界中的物体是否应该是静止时,应考虑特定的眼球运动模式。如果在眼睛注视对象时进行典型的偏航运动,则利用到前庭眼球反射(VOR)进行跟踪。在这种情况下,评估者可以确定物体图像在视网膜上固定时是否出现移动或扭曲的现象。同样,如果一个物体缓慢移动并且头部固定,则评估者执行平滑追踪以将物体保持在视网膜上。如第5.4节所述,物体逐行扫描显示出扭曲的样式取决于运动是由于VOR还是由于平滑追踪。如果物体移动得非常快并且眼睛不能将其固定在视网膜上,那么可能会感觉到拉链效应。

外设问题

目前的一代VR头设具有明显的光学像差问题;根据4.3节,随着离光轴的距离增加,这会变得更加糟糕。区分以下两种情况很重要:1)在检测周边变形的同时通过镜头中心观看,以及2)旋转眼睛直接观看镜头边缘。在第一种情况下,由于外围光感受器密度较低,失真可能不太明显;然而,不匹配可能会对舒适度和晕眩产生影响。光流信号在外围较强,而不匹配值可能被视为不正确的运动。

在第二种情况下,直接通过镜片观察可能会发现由球面像差引起的边缘缺乏焦点的现象。此外,色差可能变得可见,特别是对于黑色背景下的清晰的白线。此外,随着直线看起来变弯曲,枕形失真校正中的误差可能变得更明显。这些问题不能通过单个失真校正函数来修复

（如 7.3 节所述），因为当眼睛旋转时，瞳孔会转换离开光轴。每个眼睛方向都需要一个不同的不对称校正函数，这就需要眼部跟踪来确定在各个时间点使用哪个校正函数。

为了观察枕形或桶形变形，评估人员应当在眼睛注视在物体上的同时，尽可能大幅度地施加典型的偏航运动。在这种情况下，VOR 将导致眼睛在大范围内旋转，同时从一侧到另一侧地扫过镜头，如图 12.2 所示。如果虚拟世界包含具有显着纹理或空间频率的大而平坦的墙壁，那么当运动期间墙壁看起来像在“呼吸”时，变形便清晰可见。如果墙壁上有规则的网格图案，效果可能会更明显。

最后，许多用户甚至没有注意到镜头的有限视场。回想一下第 5.4 节，放在眼前的任何平面屏幕只覆盖眼睛的一部分视野。因此，外围的感光器将不会接收到来自显示器的任何光线。在大多数情况下，它在 HMD 内部很暗，导致在显示器的可见部分周围出现黑带。一旦向用户看向这部分，就很难忽视掉这种现象。

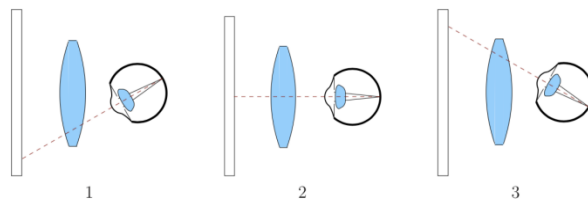


图 12.2：自上而下的视图，显示了当虚拟世界中的一个静止物体被眼睛注视时眼睛如何旋转，同时头部逆时针方向偏转（面向右侧）。周边的镜头扭曲干扰了对平稳性的看法。

延迟感知

人们对延迟的直接感受各不相同。即使不能察觉，它也是 VR 晕眩的主要原因之一[171]。适应性在其中产生了巨大的难题，因为人们可以通过长时间训练来适应固定的潜伏期；但在这种情况下回归到现实世界可能会很困难。一段时间内，现实世界中的大部分物体可能都不是静止的！

在我自己的 Oculus VR 工作中，当我在 2012 年开始使用 Oculus Rift 原型时，我可以检测到大约 40 毫秒的延迟。到 2014 年，通过以下过程，我可以把检测延迟降低至 2 毫秒。第一步是面对虚拟世界中的垂直边缘，例如门框。评估人员应保持舒适的距离，如两米。当视点固定在边缘上时，典型的偏航运动应该以非常低的幅度（例如几度）和大约 2Hz 的频率来执行。运动的幅度和频率很重要。如果幅度太大，则光学失真可能会产生干扰。如果速度太高，那么 HMD 可能会开始相对于头部翻转。如果速度太低，延迟可能不容易察觉。执行此动作时，如果存在严重的延迟，则边缘会与头部移动不同步。

回想一下，现在许多 VR 系统可以实现零有效延迟，正如第 7.4 节所述；尽管如此，由于硬件，软件和 VR 内容的特定组合，在许多系统上仍可能会出现可感知的延迟。通过使用预测，甚至可以获得负的有效延迟。配合增加或减少预测间隔的按键，我可以通过应用上述方法将有效延迟调整到 2 毫秒。该方法与心理物理学调整方法密切相关，将后面的 12.4 节中介绍。之后我能够立即将延迟时间降至 10 毫秒，而无需进行任何其他调整或比较。虽然这不是一个科学的结论（参见第 12.4 节），但看起来在花费近两年时间调试 Oculus VR 的跟踪和渲染系统以将有效延迟降至零后，我经历了一次感知学习过程。

结论

本节提供了一些培训人员发现 VR 系统问题的建议。预计未来还会出现更多。例如，为了评估虚拟世界中的听觉定位，评估人员应该闭上眼睛，以标准动作移动头部。为了在使用菲涅耳透镜的系统中检测透镜眩光，他们应该寻找在黑暗背景下由亮光形成的图案。为了检测显示闪烁（从 6.2 节回忆），特别是低至 60Hz 时，评估人员应进入一个明亮的虚拟世界，最好是白色的，并放松眼睛，直至感到明显外围振动。为了发生视觉辐辏调节冲突（回忆 5.4

节)的相关问题,可以将虚拟物体放置得非常靠近眼睛。

此外,还需要有正式的培训机制或课程,工程师和开发人员可以使用这些机制或课程来提高他们的洞察力。在这种情况下,评估人员可以通过反复练习来提高他们的技能。想象一下把 VR 体验看作一种竞争性游戏,旨在增强发现虚拟现实缺陷的洞察力。

12.2 针对开发人员的建议

随着虚拟现实耳机的广泛应用和可承受性,近年来开发虚拟现实体验的人数急剧增长。迄今为止,大多数开发人员都来自视频游戏行业,他们在开发游戏和游戏引擎方面的技能和经验被“移植”到 VR。在某些情况下,简单的改编就足够,但是游戏开发者一直对一个非常成功和流行的游戏无法直接转换为舒适,甚至是有趣的 VR 体验感到意外和困惑。大部分困惑都是由于缺乏对人类生理和感知的理解。随着该领域的发展,开发人员来自于各种不同的背景,包括电影,广播,通信,社交网络,可视化工程等。艺术家和爱好者也加入并创造了一些最有创意的经历。

本节将提供一些有用的建议,这些建议基于本书所涵盖原则的组合以及其他开发人员指南(特别是[359])的建议。这份建议当然还不完全,随着新的硬件和经验的发展,未来几年会有所增长。到目前为止,绝大多数 VR 体验都是基于成功的 3D 视频游戏。下面的大部分建议都与本书的前几部分有关,这些部分提供了科学的目的或进一步的解释。

虚拟世界

- 在虚拟世界中设置与现实世界相匹配的单位,以便更简单地匹配两者比例。例如,一个单位等于虚拟世界中的一米 这有助于深度和场景感知(第 6.1 节)。
- 确保对象完全建模,以便在用户查看屏幕上图形出现意想不到的情况时,缺少现象不再明显。
- 由于视点不同,非常薄的物体(如树上的树叶)可能在 VR 中看起来不真实。
- 设计好环境,减少运动量;例如,虚拟电梯比虚拟楼梯更舒适(8.4 节和 10.2 节)。
- 考虑视觉和听觉渲染性能问题,根据需要简化几何模型,以便在目标硬件上保持适当的帧速率(7.4 节和 11.4 节)。

视觉渲染

- 左视图和右视图之间的唯一区别是视点,而不是模型,纹理,颜色等(3.5 节和 12.1 节)。
- 切勿将文字,对象或图像固定到屏幕的一部分;所有内容应该嵌入在虚拟世界中。回想一下 2.1 节,屏幕上的静止不同于在虚拟世界中被认为是静止。
- 调整渲染视场或任何影响镜头失真的参数时要谨慎,以免造成进一步的失配(7.3 节和 12.1 节)。
- 重新评估常见的图形学技巧,如纹理映射和法线贴图,以确保它们在 VR 中有效,因为用户可以立体观看,并且能够快速更改观点(见第 7.2 节)。
- 由于视角和立体视角不同,抗锯齿技术对 VR 来说更为关键(7.2 节)。
- 渲染系统应该进行优化,以便所需的虚拟世界可以以至少与硬件要求一样高的帧率进行更新(例如, Oculus Rift 和 HTC Vive 的 90 FPS);否则,帧率可能会下降并变化,导致不适(7.4 节)。
- 避免导致大视点变化的物体移动;否则,用户可能会觉得自己正在移动(8.4 节)。
- 确定如何挑选太靠近用户脸部的几何图形的方法;否则,会出现视觉辐辏调节冲突(第 5.4 节)。
- 与游戏和摄影不同的是,视点不应以与头部追踪不匹配的方式改变,除非使用者感觉自己好像正在虚拟世界中移动。
- 为了获得适当的深度和尺度感知,用户在现实世界中的瞳孔间距离应与虚拟世界中眼

睛之间的相应视点距离相匹配（6.1 节）。

- 与屏幕上的图形相比，降低模型的亮度和对比度可提高 VR 舒适度。

追踪和匹配区域

- 切勿让头部追踪停止或延迟；否则，用户可能会立即感受到自我运动（8.4 节）。
- 考虑立体声偏移（3.5 节），头部模型（9.1 节）和运动（10.2 节），确保眼睛视点正确定位。
- 注意现实世界中存在而虚拟世界中不存在的障碍；当用户靠近障碍物时可能需要警告（见第 8.3.1 节）。
- 同样，要小心虚拟世界中存在而现实世界中不存在的障碍。例如，如果用户决定戳穿墙壁（见第 8.3.1 节），可能会产生不可预料的后果。
- 随着到达跟踪区域的边缘，逐渐降低对比度和亮度比仅保持位置固定更舒适（8.4 节）。

交互性

- 通过赋予人类“超能力”而不是应用通用模拟原则的交互机制，来达到比现实更好的效果（第 10 章）。
- 至于运动性，请遵循第 10.2 节中的建议以减少运动副作用。
- 对于在虚拟世界中的操纵，系统应要求用户在物理世界中尽可能少移动（10.3 节）。
- 关于与社会的互动性，由于恐怖谷理论，高现实度并不一定会更好（第 10.4 节）。

用户界面

- 如果出现浮动菜单，如网络浏览器或其他类型的虚拟显示器，则应将其呈现在距离用户视角至少两米远的地方，以尽量减少视觉辐辏调节冲突现象的发生（第 5.4 节）。
- 这种虚拟显示器应居中，并具有相对较窄的视野，大约为总观看区域的三分之一，以最大限度地减少眼睛和头部的移动。（第 5.3 节）。
- 如果将菜单、选项、游戏状态和其他信息嵌入虚拟世界，则应以人们熟悉的方式嵌入其中；这遵循通用模拟原理（第 10 章）。

音频

- 注意外部扬声器与耳机之间的区别；如果没有跟踪，声源本地化将无法正确实现（第 2.1 节）。
- 运动的位置和方向跟踪都应该考虑到听觉化（第 11.4 节）。
- 多普勒效应提供了强大的运动信息提示（第 11.1 节）。
- 与视觉渲染相比，音频实现过程的几何模型可以大大简化；0.5 米的空间分辨率通常来讲就已经足够（第 11.4 节）。

用户自我外观

- 如果用户能够看到他/她的身体，则其在虚拟世界中的感觉以及判断能力会得到增强。
- 简单的虚拟身体比没有虚拟身体要好得多。
- 虚拟身体和真实身体之间的差异可能令人担忧。他们可能有不同的性别，体型或物种。这可能会导致用户得到前所未有的新鲜体验，但反之可能会发生不专心的现象。
- 如果只执行头部追踪，那么虚拟身体应该满足一些基本的运动学约束，否则会让用户在虚拟世界中感到绝望（第 9.4 节）。
- 用户的自我表现会影响他们的社交行为，以及他们身边的人对他们的反应（10.4 节）。

12.3 舒适度和虚拟现实病

作为使用 VR 系统的一个副作用，体验不适是过去几十年来该技术广泛采用的一个最大的威胁。它被认为是其未能在 20 世纪 90 年代初期达到预期的主要原因。很少有人需要此技术，因为 VR 被使用时往往会导致人们的身体受到损害，且这种现象会在很多情况下并且长时间的发生。由于许多因素，例如用户之间的差异，重复使用的适应性，测量症状的困难程

度，快速变化的技术以及依赖于内容的敏感度等因素，研究人员对此也感到沮丧。显示器，传感器和计算机技术的进步已经导致硬件造成的副作用减少。随着硬件产生的副作用减少，相应的负担一直在向软件工程师和内容开发人员转移。发生这种情况是因为 VR 体验本身有可能让人感到恶心，即使硬件被认为是非常舒适的。事实上，最好的 HMD 可能比以往任何时候都更让人感到恶心！出于这些原因，VR 系统的工程师和开发人员必须了解这些副作用产生的原因，以便他们确定如何为绝大多数用户减少或消除这些副作用。

晕眩或综合症

在本书中，我们讨论到 VR 晕眩中任何意想不到的副作用。这包括很多通常与晕眩无关的症状，例如疲劳。因此，更准确的短语可能是虚拟现实适应不良综合症，其中适应不良是有害而不是有益的，综合征是指与活动相关的一组症状。

晕动症变种

了解与 VR 晕眩密切相关的术语是有帮助的，因为它们与类似的活动，症状集和潜在原因相关。这有助于进行相关的研究。最广泛的相关领域是晕动症，与真实世界中的运动症状有关。其通常涉及前庭器官（8.2 节），这意味着它们涉及感觉输入或加速冲突；事实上，没有功能性前庭器官的人不会遇到晕动的现象[145]。由于真实运动引起的运动症是由于经历了异常的力量而发生的。这可能是由于身体的运动，导致头晕和恶心。同样，这些症状是由可能产生极端力量的车辆运输而发生的。此外，各种游乐设施可以产生更极端的体验和副作用。

不幸的是，晕车远远超出娱乐范围，因为许多人在乘坐专为运输而设计的车辆时仍患有晕车病。人们分别从汽车，船只和飞机中经历晕车，晕船和晕机症状。据估计，只有约 10% 的人在运输过程中从未经历过明显的恶心[171]。由于士兵们花费了长时间进行海上巡逻和高速战斗机飞行，军队进行了最大的晕动病研究。大约 70% 的海军人员会遇到晕船问题，其中约 80% 的士兵会降低其工作效率或动力[249]。最后，一个非常规的例子是太空旅行，其中遇到微重力的宇航员也会有恶心或其他症状，这就是所谓的太空病。

视觉诱发晕动症

迄今为止的晕动症都涉及到真实的动作。相反，运动病可能因为刺激反应过激而发生，这些刺激会说服大脑发生加速，尽管事实并非如此。这就是所谓的视觉运动。最常见的研究案例是视觉诱发的视觉运动，也称为视觉运动，见 8.4 和 10.2 节。与此相关的症状是视觉诱发晕动症的一部分。

媒介过程可以通过多种方式产生。回想一下 2.3 节的图 2.20，极端情况是由于人们在摇摆的房间引起。科学家们通过运动视觉刺激来围绕受试者进行运动视觉和感知运动病的对照实验。在涉及特定移动视野的各种研究中，只有少数受试者不受副作用的影响。大约 50% 至 100% 的人会出现头晕，约 20% 至 60% 的人会出现胃部症状；更确切的情况取决于具体的实验结果[171]。

此外，可以使用显示器来产生动感。回想一下第 6.2 节，由于帧的快速连续性，在这种情况下感知的光流是频闪表现运动。然而，使用视动力鼓的晕眩研究仍然有重要的参照意义。它们揭示了视觉诱发晕动症的严重程度，即使在没有数字伪像（如显示分辨率和帧率）的限制下也是如此。

模拟器晕眩和网络病

一旦使用显示器，2.1 节中设计的选择问题则又在此重现：定屏幕（如 CAVE VR 系统）还是需要跟踪的头戴式显示器。汽车模拟器也许是 VR 的第一个重要应用，最常见的例子是驾驶汽车，驾驶飞机或直升机。用户可以坐在固定的基座上，也可以坐在一个机动的基座上对应控制。后一种情况提供了前庭刺激，运动和视觉信息的时间同步，这对于减轻晕眩至关重要。通常，整个驾驶舱都在现实世界中重建，视觉刺激主要出现在窗户外。头部可以被迫

踪以提供立体视觉和不同的视点，但是大多数情况下并不会这样做，主要是为了舒适度最大化并且技术副作用被最小化。这种活动导致的视觉诱发晕动的分支症状地称为模拟病，美国军方已经对此进行了充分研究。

术语 cybersickness [206]被提议用于涵盖与 VR（或虚拟环境）相关的任何晕眩，其中就包括模拟器晕眩。不幸的是，这个词的含义近来已经扩大到包括花费太多时间与智能手机或电脑进行交互导致的晕眩。此外，网络这个术语在过去的几十年中已经积累了许多奇怪的内涵。因此，我们将视觉诱发晕动病以及由 VR 引起的任何其他形式的不舒适症状共称为 VR 晕眩。

VR 病的常见症状

目前而言，已有各种术语用于指向 VR 运动和 VR 晕眩研究中的症状。最常见的是（基于[143,145,169,171,305]）：

- 恶心：症状开始时，使用者可能有与胃，上腹部，食道或喉咙有关的不适感。随着强度的增加，它逐渐导致呕吐感。这是 VR 晕眩最消极和最吓人的症状。
- 头晕：即使在刺激消除后，用户可能会感觉到运动的感觉，如旋转或摇摆。这也可能包括眩晕，它与前庭器官功能障碍相似且经常发生。
- 嗜睡：用户可能会变得不那么警觉，打哈欠，并最终开始入睡。
- 流涎增加：口中唾液量增加，导致通常吞咽量增加。
- 冷汗：用户开始出汗或增加汗水，但不会因环境温度升高而出汗。
- 苍白：用户可能在脸上、耳朵、脖子和胸部都会出现皮肤苍白甚至脱落的现象。
- 脸色潮红：这与感觉到的温暖突然增加相似，类似于发烧。
- 头痛：用户头痛，可能会逐渐增加强度，并在使用后很长时间持续头痛。
- 疲劳：长时间使用后，用户可能会感到疲倦，筋疲力尽。
- 眼睛疲劳：用户可能会觉得他们的眼睛疲劳甚至疼痛。
- 聚焦问题：用户可能视力模糊或难以聚焦。

以上是部分体验者为某些 VR 测试所报告的症状。研究这些症状的目标是让 VR 系统和体验尽可能地消除这些症状。此外，通过重复训练和适应，大多数症状发生的可能性会被大大降低。

其他副作用

除了刚刚列出的直接症状外，其他一些现象也与运动和 VR 晕眩密切相关，并且会持续很长时间。其中一个是 Sopite 综合征[102]，与嗜睡密切相关，但可能包括其他症状，如懒惰，缺乏社会参与，情绪变化，冷漠和睡眠障碍。即使在适应上述系统后，这些症状仍可能持续存在。另一个现象是姿势不平衡，对平衡和协调产生不利影响[171]。最后，另一种现象是在头部或身体运动过程中丧失视力[171]，这似乎是 VOR（5.3 节）适应虚拟现实系统缺陷的自然结果。这是因为尽管存在分辨率，等待时间，帧速率，光学畸变等问题，但仍然会对平稳性的感知造成影响。

体验后效应

VR 晕眩最令人担忧的一个方面是症状在使用 VR 后可能持续数小时甚至数天[304]。大多数在从 VR 体验中退出后立即出现症状的用户仍然会在后续的一段时间内表现出一些晕眩。有极少数的用户可能会持续数小时或数天出现症状。同样，一些遇过海难的人在返回地面后长期患有土地晕眩症。这对应于视觉世界的姿势不稳定性和感知不稳定性；对具有这种症状的人来讲，世界似乎在摇摆[171]。

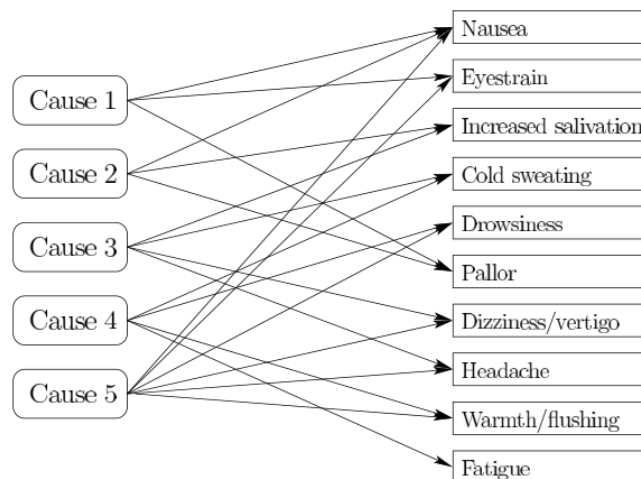


图 12.3: 症状可以观察到, 但原因并不能直接被测量到。研究人员面临逆向问题, 根据观察到的症状推测原因。但目前的问题在于每种症状可能有很多可能的原因, 其中一些可能与 VR 体验无关。

从症状到原因

症状是最终的结果, 但它们的原因是什么? 见图 12.3。VR 系统的科学家或评估者目前面临的问题在于只有症状可观察到。任何症状都有任何可能的原因。其中一些原因是已知的, 而其他可能无法确定。例如, 假设用户在 VR 体验 5 分钟后出现了轻微的恶心。除了因为他骑着自行车参加考试并忘了吃早餐, 他还会有什么产生恶心的可能性呢? 如果他前一天晚上喝醉怎么办? 也许像这样的一些用户可能被作为离群值抛弃, 但是如果前一天晚上举行了一个大型节日, 在实验之前疲劳了的测试者应该怎么对待呢? 其中一些问题可以通过将一些用户分进预计变异性较小的组来处理; 见第 12.4 节。如果他们感到恶心, 至少应该事先询问他们; 然而, 这样的行为可能导致他们更加关注虚拟内容而恶心。

通过在接受 VR 体验之前, 期间和之后获得的数据中找到的统计相关性, 可以回答有关 VR 病原因的问题。原因的分析过程并不是通过直接见证其原因及其效果, 而是通过观察明显的效果而确定的。消除不相关的原因是实验设计的重要组成部分, 其中涉及仔细选择用户的过程并提前收集适当的数据。确定更具体的原因需要更多的试验性实验。这种情况很复杂, 因为不同的试验不能轻易应用于同一用户。一旦人们生病了, 他们将无法参与, 或者至少会给出难以弥补的偏差结果。

用户之间的差异

另一个复杂因素是人们对 VR 晕眩易感性的变化性。在实验设计中必须考虑到群体间的个体差异; 参见第 12.4 节。大多数研究人员认为女性比男性更容易患晕动病[143,241]; 然而, 这个结论在[171]中有争议。关于年龄, 12 岁以下儿童的易感性似乎最高, 然后随着年龄增长而迅速下降, 同时在其一生中逐渐下降[262]。一项研究甚至认为中国人比其他一些组别更容易受到影响[309]。预测个体易晕车的最佳预测因素是确定他/她是否曾患有晕车症。最后, 要注意的是, 各组之间也可能存在差异, 如症状的严重程度, 发病的速度, 实验后的持续时间以及用户适应 VR 的速度。

感知冲突理论

除了根据违规刺激来确定因果关系之外, 我们还应该尝试理解人体为什么因 VR 产生负面作用。该过程涉及哪些生理和心理机制? 为什么一个人不能快速适应某些刺激, 而其他的人很好? 什么是特别不好的刺激? 虽然目前没有广泛包含和公认的解释晕动病或 VR 晕眩的理论, 但存在一些有用和可接受的理论。

理解 VR 晕眩必须具备的相关理论和强大理论之一是感知冲突理论[132,145]。回想一下

第 2.1 节图 2.1 中 VR 系统的高级描述，VR 中存在两种不匹配：

1.与自然刺激相比，工程刺激与中枢神经系统和大脑预期的刺激程度并不相近。示例包括由于显示分辨率，帧率，光学失真，有限色彩，合成照明模型和延迟而引起的伪影。

2.一些感官系统不接受工程刺激。它们继续以自然的方式感知周围的物理世界，并相应地发送神经信号。例子包括来自前庭和本体感受系统的信号。前庭器官继续感知到真实世界的加速度，并且可以从运动信号粗略地估计身体部位的姿势。

毫不奇怪的是，VR 中最重要的冲突涉及到加速。人类视觉系统提供与运动一致的光流读数，但来自前庭器官的信号与之不匹配。请注意，这是一种常见的晕动病现象，它在现实中往往出现在行驶中的车辆中行驶而人们不向外看的情况。例如，想象一下当一名乘客在汽车里读书时的情况。在这种情况下，前庭系统报告汽车的加速度，但没有相应的光流量。

强制融合和疲劳

回想一下第 6.4 节，我们的感知系统整合了来自不同来源，不同感知模式的线索，以获得连贯的感知解释。在线索之间存在微小差异的情况下，由此产生的解释可以被认为是强制融合[120]，尽管其中会有偏差，感知系统似乎难以形成匹配。执行传感器融合或视觉场景解释的工程系统下的情况与之类似；在存在更多噪音或不完整信息的情况下，优化或寻找合理解释的可能性会更大。强迫融合似乎直接导致疲劳和眼睛疲劳。因此，尽管存在工程上的缺陷，但人类在学习解释虚拟世界时被迫加大机体运行力度。幸运的是，反复的虚拟现实体验会导致感知系统的学习或适应，最终可能会减少疲劳。

毒药假说

感觉冲突似乎足以解释为何会产生额外负担，但似乎并不意味着会导致恶心。科学家们正在调查什么是造成这一现象和相关症状的原因。请注意，人类有能力自然地使自己摆脱不涉及技术的旋转运动。间接毒物假说认为与晕动病相关的恶心是一种在人体内进化机制带来的副产物，导致人们会呕吐意外摄入的毒素[323]。这种毒素频率的症状包括视觉和前庭线索之间的冲突。另一种解释是直接毒物假说，该假说认为恶心与毒素有关，因为它们在整个进化过程中与涉及增加或延长加速度的活动有关。[171]的第 23.9 节给出了这些替代假说及其不完整的详细评估。

虚拟现实晕眩的级别

为了改善虚拟现实系统和体验，我们必须对副作用进行适当的比较。因此，所产生的症状需要量化。对于每种症状，并不是简单的是/否回应，而是获得与相对严重程度对应的数值。对于一个特定的症状，几个重要的参量包括：

- 症状的强度。
- 刺激呈现时，症状发作或强度增加的速度。
- 刺激消除后，症状衰减或强度下降的速度。
- 在固定级别或更高级别遇到症状的用户的百分比。

前三个可以被视为随时间变化的强度图。最后一个统计属性；许多其他统计数据可以从原始数据中计算出来。

问卷调查

收集定量数据最常用的方法是让用户填写问卷。研究人员多年来设计了许多问卷[170]；最广为人知的是模拟病人问卷（SSQ）[144]。它被设计用于美国军方的模拟病症研究，已被广泛使用。其要求用户以四个等级评分 16 个标准症状中的每一个：0 无，1 轻微，2 中等和 3 严重。通常通过总计分数来汇总结果。为了确定起病率或衰退率，必须多次给予 SSQ，例如实验之前，实验后，10 分钟后，30 分钟后，甚至 60 分钟后。

但问卷调查主要有四个缺点。首先是答案是主观的。没有明确的方法来校准在用户感觉在等级“1”与等级“2”之间恶心的含义。单个用户甚至可能根据情绪甚至其他症状的发

作给出不同的评分。第二个缺点是要求用户注意他们的症状,这可能会影响他们的发病倾向。第三个缺点是用户必须中断,以便他们可以在会话期间提供分数。最后的缺点是,随着时间的推移,强度必须粗略地采样,因为每次感兴趣的时刻都必须填写新的问卷。

生理测量

另一种方法是将传感器连接到用户,以便在 VR 体验之前,之中和之后自动获取生理测量结果。这些数据可以在不中断用户或要求他们注意症状的情况下持续获得。然而,在身体上放置传感器可能会有一些不适或恐惧。研究人员通常会购买标准传感系统,如 Biopac MP150,其中包含一包传感器,记录数据并将其传输到计算机进行分析。

已经用于研究 VR 晕眩的一些生理学措施是:

- 心电图(ECG):该传感器通过将电极放置在皮肤上来记录心脏的电活动。VR 体验期间心率通常会增加。

- 胃电图(EGG):这与心电图相似,但是电极放置在胃部附近,以便估计肠胃的不适情况。

- 眼电图(EOG):将电极放在眼睛周围,以便估计眼球运动。或者,可以使用基于相机的眼动追踪系统(第 9.4 节)以确定眼旋转和眨眼率。

- 光电容积描记图(PPG):这提供了关于心脏运动的附加数据,并通过使用脉搏血氧仪获得。通常这种设备夹在指尖上来监测血液的氧饱和度。

- 电流皮肤反应(GSR):该传感器测量整个皮肤表面的电阻。当人流汗时,皮肤表面的水分会增加导电性。这提供了一种测量冷汗的方法。

- 呼吸运作:呼吸频率和振幅是从胸腔上的一个贴片测量的,它对压差或膨胀产生反应。VR 体验期间呼吸速率可能会增加。

- 皮肤苍白:可以使用相机和图像处理来测量。在最简单的情况下,IR LED 和光电二极管可用作测量皮肤反射率的发射器-检测器对。

- 头部运动:头部跟踪系统是运动数据的丰富来源,它可以帮助估计疲劳或姿势不稳定性,而无需额外的成本或分散用户的注意力。

近期,[59]对生理测量和问卷调查的方法进行了比较,甚至得出了结论:可以从独立的生理数据中确定一个人是否正在经历 VR。

减少晕眩不适的策略

通过实验研究来确定用户的 VR 晕眩频率和强度,工程师和开发人员可以迭代并生成更舒适的 VR 体验。回顾之前讲过的“平稳感”的内容,我们发现在各个环节都需要改进。大多数现实世界都被认为是静止的,它应该和虚拟世界一样。视觉显示,渲染和跟踪方面的改进应该通过提高对平稳性的认识来帮助减轻晕眩不适,还应该减少或消除光学失真,混叠,延迟和其他伪影。当它们不能被消除时,应该找到适当的折衷方案。此外,采用新的显示技术,以降低视觉辐辏调节冲突的问题,当传统屏幕和镜头组合的头戴式设备上出现紧密物体时会引起严重不适(请参阅第 5.4 节)。

即使对于理想的头戴式设备,运动也会因为视力而导致晕眩。遵循 10.2 节中提出的策略可以减少晕眩症状。一个更好的想法是设计需要很少或不需要移动的 VR 体验。

作为最后的措施,还有两种策略可能有助于缓解 VR 晕眩[145]。首先是定期练习,这会导致适应。随着身体对不寻常的刺激组合的调整,预计强制融合导致的疲劳量会减少。当然,如果 VR 体验让大多数人感到不舒服,那么让他们“通过”十几次甚至更多的练习不是一个好的主意。最后,用户可以服用降低易感性的药物,就像有些人在登机前服用空气药丸一样。这些药丸通常是抗组胺药或抗胆碱药,它们具有一些副作用,例如疲劳,嗜睡,认知能力受损,以及在某些情况下成瘾的可能性。

12.4 人类受试者的实验

想象一下，你已经开发出一种新的运动方法，希望能减少 VR 晕眩。你和几个朋友可能会尝试它，并相信它比默认方法更好。你如何说服那些怀疑它的人，尤其是那些不太可能偏向于新方法的人？你可以争辩说它更好，因为它尊重来自人体生理学和知觉的已知问题，这将是一个合理的开始。这将为首先尝试该方法提供良好的动力；然而，它本身也还不够，因为人体与技术的交互方式存在很多不确定性。解决方案是设计一个实验，科学地确定你的方法是否更好。这导致了许多挑战，比如确定有多少人应该尝试，应该做什么，应该做多长时间，谁应该分配给哪种方法，以及之后如何测量他们的晕眩程度。其中，一些困难讲述于 12.3 节。如果实验设计得很好，那么科学家会支持你的结果。如果有些人还不相信，那么至少你会得到那些相信科学方法的人的支持！幸运的是，这包括心理学家和神经科学家，甚至包括人机交互相关领域的研究人员[37,40]。

科学的方法

科学方法自古以来就已存在，并在特定的科学领域不断完善和改进。图 12.4 描述了它如何在 VR 开发中出现。想象一下，试图爬上梯子，第一步是通过研究适当的文献或获得必要的背景来设计一种可能有所改进的新方法。这将减少从梯子上坠落的可能性。第二步是设计和实施新方法。这一步可能包括对少数用户进行一些简单的评估，以确保值得进一步处理。

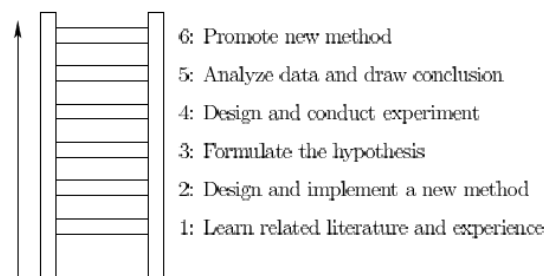


图 12.4：科学过程就像爬梯子一样。每次失败时都要小心，不要跌得太低！

第三步是精确地制定假设，关于它是如何改进的。例如：1）减少不良症状，2）提高舒适度，3）提高解决任务的效率，4）更强烈地相信虚拟世界是真实的，5）更好地享受这项活动。评估多个标准通常是有意义的，但结果可能是新方法在某些方面更好，而其他方面更糟。这是一个普遍的结果，但这总比不以任何方式改善要好！这个假设甚至可能涉及改进未来的实验程序，在[59]中举了一个例子，在这个例子中，研究人员确定了哪些生理学指标比问卷更好地指示了 VR 晕眩。最后，该假设应简化实验的第四步，同时保持结果的可靠性。

对于第四步，应该设计和实施实验来测试假设。第五步是分析数据并得出结论。如果结果是关于感兴趣的标准的“更好”的方法，那么就到达了六步，此时新方法应该呈现给世界。

任何步骤都可能产生偏差。例如，在进行实验后，可能会意识到，项目池过于有偏向性。这时就需要退一步，重新设计或重新实施实验。如果第五步的结论是该方法不是明显的改进，或者更糟，这可能需要返回到第二步甚至第一步。所以设计关键是每个步骤都要小心谨慎，避免每次失败都会导致实验被重置得太多！

人类—实验研究对象

做关于人的实验很困难，特别是如果他们是科学实验的研究对象。他们之前的虚拟现实经验，对晕动病的敏感性，对技术的怀疑，喜怒无常以及倾向性可能会有很大差异。由于好奇心，经济补偿，感兴趣程度或学位要求（心理学学生常常被迫参加实验），他们可能会同意成为实验对象。一位科学家或许能够根据诸如性别，年龄或职业等因素猜测一些人在实验中的表现。应用科学方法来制定和评估关于人群（或动物）的假设的主题被称为行为科学[152]。

最大的挑战之一是他们是否在不经意间被观察到（甚至不知道他们是否是实验的一部分），或者实验是否呈现了在现实世界中永远不会遇到的刺激或情况。人为设置有时会导致

科学家反对实验的生态效度。幸运的是，虚拟现实是我们想要评估的一个特别设计的设置。因此，关于 VR 使用的结论更可能在生态学上是有效的，特别是即使没有用户意识到实验也可以获得实验数据。头部追踪数据可以在服务器上收集，有数百万人尝试 VR 体验。

道德标准

这导致了下一个挑战，那就是人类的权利，人们可能拥有比动物更多的权利。必须避免影响其隐私或健康的实验。涉及人类受试者的科学实验必须坚持高标准的伦理道德，这是从纳粹医学实验和 20 世纪中叶的 Tuskegee 梅毒实验中学到的一课。纳粹战争罪法庭的结果产生了纽伦堡法典，其中规定了一套关于人体实验的道德原则。今天，世界各地正在认真对待人类研究的伦理标准，正在进行的辩论或细节上的差异[236]。在美国，涉及人类受试者的实验需要由法律机构审查委员会（IRB）批准。通常情况下，IRB 这个术语也用于指由审核委员会批准的实验或一组实验的提案，如声明“需要 IRB”。涉及虚拟现实的实验通常不存在争议，并且与数十年来广泛认可的模拟病症实验类似。

变量

行为科学家总是关心变量。每个变量都有一个集合中的值，可以是数字，也可以是实数，也可以是符号，如颜色，标签或名称。从他们的角度来看，最重要的三类变量是：

- 依赖性：这些是假设所关注的主要对象。
- 独立：这些值是由科学家直接改变或操纵的值。
- 滋扰：由于这些因素各不相同，它们的价值观可能会影响因变量的价值，但科学家对它们的控制力较弱，而且它们不是主要研究的对象。

高级任务是制定一个假设，可以根据独立变量和因变量之间的关系进行评估，然后设计一个可以控制滋扰变量的实验，并且可以在有限的时间，资源内完成。

用于制定变量如何行为并预测其行为的模型的基础数学是 6.4 节介绍的概率论。不幸的是，我们正面临着一个相反的问题，如图 12.3 所示。大多数行为都不是直接可观察的，这意味着我们必须收集数据并对基础模型做出推论，并尽可能获得更多的信服。因此，解决这一假设是应用统计学中的一个问题，它是概率论的自然补充或逆转。

提出假设

在最简单的情况下，科学家想要确定一个关于假设的二元结果：真或假。在更复杂的情况下，可能有许多相互排斥的假设，科学家们想确定哪一个是真的。例如，17 种不同的运动方法中哪一种最舒适？从简单的案例出发，假设已经根据 VR 晕眩确定了一种潜在更好的运动方法。令 x_1 表示使用原始方法，并让 x_2 表示使用新方法。

集合 $x = \{x_1, x_2\}$ 是自变量。每个 x_i 被称为一次实验。接受原始方法的受试者被认为是对照组。如果正在评估一种药物，对照组不使用药物的话，那么他们将接受安慰剂。

回顾第 12.3 节，可以通过各种方式评估 VR 晕眩的水平。为了举例，假设选择一段时间内平均的 EGG 电压测量值作为因变量 y ，对应于 x_1 或 x_2 条件下的胃肠不适程度。

该假设是将 x 与 y 联系起来的逻辑真/假陈述。例如，它可能是

$$H_0 : \mu_1 - \mu_2 = 0, \quad (12.1)$$

其中，通过对世界上所有人施加实验 x_i 可以得到每个 μ_i ，即在实验中同一点上 y 的“真实”平均值。假设 H_0 意味着新方法对 y 没有影响，通常称为零假设。 H_0 的负面被称为替代假设。在上述情况下，替代假设为

$$H_1 : \mu_1 - \mu_2 \neq 0, \quad (12.2)$$

这意味着新方法对胃肠不适有影响，然而，它可能会更好或更糟。

测试假设

不幸的是，科学家无法同时对所有人进行同样的实验。他们必须从人群中抽出一小部分人，并确定该假设是否属实。让指数 j 指向一个特定的选择主题，并让 $y[j]$ 作为受试者对实验的回应；每个受试者的反应是一个因变量。两个统计数据对于结合因变量的信息很重要：

平均数：

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y[j], \quad (12.3)$$

这只是 $y[j]$ 在主体上的平均值，以及方差：

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (y[j] - \hat{\mu})^2. \quad (12.4)$$

方差估计 (12.4) 被认为是“真实”方差的有偏估计量，因此，有时会应用贝塞尔的修正，即将 $n - 1$ 放入分母而不是 n ，从而得到无偏估计量。

为了检验这个假设，Student 的 t 分布（“学生”是威廉·西利戈斯特）被广泛使用，这是一个概率分布，它捕获了任意 n 个随机样本被选择及其响应 $y[j]$ 被平均后均值的分布情况，见图 12.5。假定每个个体 j 的响应 $y[j]$ 是一个正态分布（工程中称为高斯分布），它是最基本和最常见的概率分布。它的平均 μ 和标准偏差 σ 完全给出。这些分布的确切表达式在这里没有给出，采用通用形式，具体请参阅[124]以及其他有关这些以及更多数学统计的书籍。

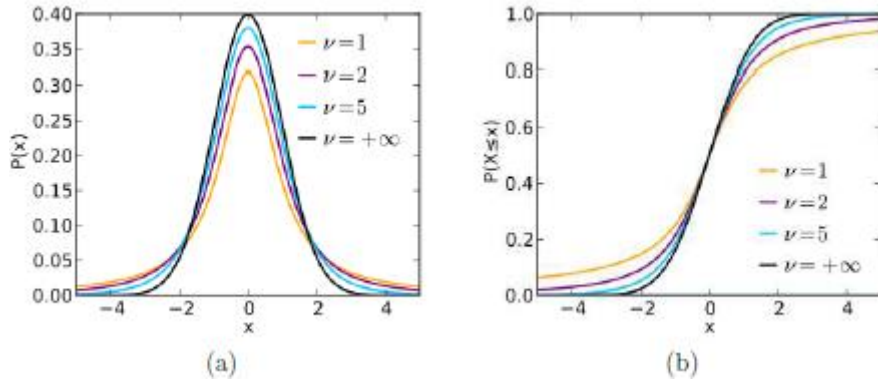


图 12.5: Student 的 t 分布: (a) 概率密度函数 (pdf) ; (b) 累积分配函数 (cdf)。在这些图中, ν 被称为自由度, 并且对于被摄体数 n , $\nu = n - 1$ 。当 ν 很小时, pdf 的尾部比正态分布更大; 然而, 当 ν 接近 ∞ 时, Student 的 t 分布收敛于正态分布。(维基百科用户 skbkekass 提供)

Student 的 t 检验[314]涉及计算以下内容:

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (12.5)$$

其中:

$$\hat{\sigma}_p = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}} \quad (12.6)$$

其中, n_i 是接受治疗 x_i 的受试者数量。表达式中减 1 和 2 是由于贝塞尔的修正。基于 t 的值, 通过查看 Student's t cdf 的表格来确定零假设 H_0 中的置信度 α (图 12 (b))。通常, $\alpha = 0.05$ 或更低就足以说明 H_1 是真实的 (相当于 95% 的置信度)。这样的表格通常是这样排

列的：对于给定的 ν 和 α ，提供确认 H_1 以置信度 $1-\alpha$ 所需的最小 t 值。请注意，如果 t 是负数，那么 x 在 y 上的效果将朝相反的方向运行，并且用 $-t$ 替换查表。

二元结果可能不够令人满意，但这不是问题，因为平均值 $\mu_1 - \mu_2$ 中的差异是应用 x_2 与 x_1 相比的变化量的估计值。这被称为平均实验效果。因此，除了通过 t 检验确定 H_1 是否为真，我们还可以估计它对结果的影响程度。

学生的 t 检验假定每个组内的方差是相同的。如果不是，那么使用 Welch 的 t 检验[343]。请注意，在这两种情况下均未预先给出差异。它们是根据实验数据“实时”估算的。如果方差恰好相同，则 Welch 的 t 检验给出与 Student's t 检验相同的结果，因此，如果存疑，最好应用 Welch 的 t 检验。科学家们则可以使用许多其他测试并在特定情况下进行辩论；见[124]。

相关系数

在许多情况下，自变量 x 和因变量 y 都是连续的（取实值）。这便引出了另一个重要的参数称为皮尔逊相关系数（或皮尔逊的 r ）。这估计了两个变量之间的线性相关量。对于每个受试者，处理条件 $x[i]$ 被应用，并且响应是 $y[i]$ 。请注意，在这种情况下，没有组的概念（或每个主题都是唯一组）。此外，任何治疗都可能适用于任何人；索引 i 仅表示特定的受试者。

当作为随机变量处理时， r 值被计算为 x 和 y 之间的估计协方差：

$$r = \frac{\sum_{i=1}^n (x[i] - \hat{\mu}_x)(y[i] - \hat{\mu}_y)}{\sqrt{\sum_{i=1}^n (x[i] - \hat{\mu}_x)^2} \sqrt{\sum_{i=1}^n (y[i] - \hat{\mu}_y)^2}}, \quad (12.7)$$

其中 μ_x 和 μ_y 分别是所有组的 $x[i]$ 和 $y[i]$ 的平均值。分母只是估计的标准偏差的乘积： $\sigma_x \sigma_y$ 。

r 值介于 -1 和 1 之间。可能发生三种定性不同的结果：

- $r > 0$ ：这意味着 x 和 y 正相关。随着 x 的增加， y 趋于增加。 r 值越大意味着效果越强。
- $r = 0$ ：这意味着 x 和 y 是不相关的，这在理论上相当于零假设。
- $r < 0$ ：这意味着 x 和 y 是负相关的。随着 x 增加， y 趋于减小。 r 值越小意味着效果越强。

在实践中，从实验数据中很难获得 $r = 0$ ；因此，绝对值 $|r|$ 给出了 y 取决于 x 的可能性的重要指示。与零假设（ $r = 0$ ）的理论等价性只会在受试者的数量趋于无穷大时才会发生。

处理滋扰变量

我们已经考虑了依赖性和自变量，但忽略了这些滋扰变量。这是实验设计中最具挑战性的部分。这里只给出了总体思路，请参阅[152,195]了解详尽的介绍。假设在查看数据时因变量 y 在很大程度上取决于对象的可识别属性，如性别。这个属性将成为一个令人麻烦的变量 z 。我们可以想象设计一个实验来确定 z 是否影响 y 以及 z 对于 y 有多大影响。

对 z 的依赖性会导致主体间的差异很大；然而，如果它们被分成每个组内具有相同 z 值的组，则方差可能会相当低。例如，如果性别是滋扰变量，那么我们会将受试者分成男性和女性小组，并发现每个小组的方差都较小。这种技术称为分块，每个组称为一个块。在块内部，如果自变量 x 保持不变， y 的方差应该很小。

接下来的问题是要确定实验适用的对象。继续这个例子，对女性给予 x_1 实验，对男性给予 x_2 实验会是一个可怕的想法。这完全混淆了因变量 y 的滋扰变量 z 和自变量 x 的依赖关系。与此相反，将 x_1 应用于一半的女性和男性， x_2 应用于另一半，这显然更好。一个简单的选择是使用随机设计，其中受试者随机分配 x_1 或 x_2 。这可以安全地消除意外偏差，并

且实验者可以轻松开展实验。

如果有多于一个滋扰变量,那么分配过程变得更加复杂,这往往导致对随机化的依赖性。如果受试者参加了不同时间应用不同疗法的多阶段实验,则必须认真分配疗法。处理它的一种方法是根据拉丁方阵分配治疗方法,拉丁方阵是 $m \times m$ 矩阵,其中每行和每列都是 m 个标签(在这种情况下为实验方案)的置换。

方差分析

剩下的主要挑战是找出会对方差产生重大影响的滋扰变量。这称为方差分析(或 ANOVA,发音为“ay nova”),考虑到这一点的方法称为 ANOVA 设计。性别是一个容易想象的因素,但其他因素可能更为微妙,例如受试者玩过的 FPS 游戏的数量,或受试者已玩的时间。这个设计过于复杂,不能完全覆盖所有情况(见[152]),但重要的点是必须在主体之间发现低方差的群集,作为将它们分块的基础。这与无监督聚类问题(或无监督学习)密切相关。方差分析也被认为是对三个或更多变量的 t 检验的推广。

更多变量

除了独立,依赖和滋扰之外的变量有时在实验中也会起重要作用。控制的变量本质上也是一个变量,通过选择主体或实验性试验来保持固定不变。例如,通过控制对象选择,可以使方差保持较低,如在实验中仅使用年龄在 18 岁至 21 岁之间的男性。该方法有助于提高对实验结论的可信度,但可能会阻止将其发现推广到控制之外的情况。

混杂变量是导致独立变量和因变量相关的无关变量,但是一旦给出了混杂变量的值,它们就变得不相关。例如,拥有较大的鞋号可能与更好的语言能力相关。在这种情况下,混杂变量就是人的年龄。一旦知道了年龄,我们就会意识到老年人有比小孩子更大的鞋码,表达能力也更好。这说明了从统计相关性推断因果关系的冒险性。

心理物理学方法

回想第 2.3 节,心理物理学将感知现象与原始刺激联系起来,这对理解 VR 来说至关重要。史蒂文斯的幂律(2.1)将感知的刺激幅度与实际幅度相关联。JND 涉及确定差异阈值,这是可检测的最小刺激变化量。这种情况的一个特例是绝对阈值,它是可检测的最小幅度刺激(与零相比)。

心理物理学的规律或关系是通过对人体的特定实验获得的。术语心理物理学和研究领域由 Gustav Fechner [76]介绍,他提出了三种基本的实验方法,下面将对此进行介绍。假设 x 代表刺激幅度。其任务是确定 x 可以变得多小,这样受试者就会感受到差异。

传统的方法是:

- 不断刺激的方法:在这种情况下,连同参考刺激一起呈现不同大小的刺激。该方法要求对于每个刺激,受试者可以感知到刺激之间的差异。幅度通常以随机顺序呈现以抑制适应。基于对许多试验的反应,计算出最佳拟合心理测量函数,如图 2.21 所示。

- 限制方法:实验者以较小或较高的限制开始以较小的增量改变刺激幅度。在每种情况下询问受试者新的刺激是否比参考刺激小,相同或更大。

- 调整方法:允许对象在短时间内上下调整刺激强度,同时还可以与参考刺激进行比较。当他报告调整后的刺激有与参考刺激相同的幅度时,实验停止。

尽管这些方法是有效的并且被广泛使用,但存在一些问题。所有受试人在实验时都可能产生偏见。对于倒数两种方法,适应性可能会干扰结果。对于最后一个方法,没有办法控制主体如何做出决定。另一个问题是效率,因为考虑到远离参考刺激的刺激,多次刺激的更新会导致时间的浪费。

自适应方法

由于这些缺点,研究人员在过去几十年中找到了许多改进的实验方法。其中很大一部分在[324]中进行了调查和比较,属于适应性心理物理学方法的领域。大多数改进的方法执行阶

梯程序，其中刺激量级从测试对象容易察觉的情况开始，并且逐渐减小梯度直至受试者出错[90]。到了这一点，再进行相反的过程，直到产生另一个错误。出现错误和改变方向的过程一直持续到在很短的多次实验中出现许多错误为止。步长必须仔细选择，甚至可以在实验过程中逐渐减小。当每次迭代收集数据时，方向（增加或减少）可以使用贝叶斯或最大似然算法来决定，该方法可以提供阈值的估计值[113,154,342]。这些方法通常属于随机逼近方法的范畴[266]。

刺激幅度估计

回想一下，史蒂文斯的幂律不是关于检测阈值，而是关于刺激的感知幅度。例如，一个盘子可能会感到两倍热。在这种情况下，可以要求受试者估计刺激之间幅度的相对差异。在足够数量的试验中，史蒂文斯幂指数（2.1）的指数可以通过选择使最小平方误差最小的值（指数）来估计（从 9.1 节回忆）。

进一步阅读

有关感知学习的调查，请参见[94,98,109,253]。感知学习的过度行为在[101,253]中进行了研究。在[283]中已经确定，感知学习可以在没有集中注意力的情况下发生。

[62,67,200,357]分析了人类对 VR 和计算机接口延迟的敏感性。在[289]中研究了立体显示器的舒适问题。关于姿势摇摆和晕眩之间的联系，请参见[294,313]。

有关与 VR 晕眩有关的一些重要研究，请参阅[13,146,147,153,223,263]。[143,169,305]给出了 VR 疾病的一般概述。[262]中调查了晕动病。参见[120,138,55,255]可以了解强制融合的更多信息。

有关人类受试者实验的数学方法和统计数据介绍，请参见[152]。[195]因其在心理学背景下对假设检验的报道而受到欢迎。有关心理物理学方法的治疗，请参阅[176,324,349]和[93]的第 3 章。