

## 第二章 总体架构

本章阐述了 VR 系统从硬件（第 2.1 节）到软件（第 2.2 节）再到人类感知（第 2.3 节）的概述。其目的是迅速地提供一个全面的视角，以便在更大范围内理解其余章节中的详细主题。12.2 节为 VR 开发人员提供了一些建议，有助于更深刻的理解 VR。这些章节中的基本概念将主要面向工程和科学背景的读者。此外，本书的读者读完本书后应该能够具备开发新技术并向他人提出他们自己的建议的能力，以便高效学习 VR 系统，并有好的学习体验。

### 2.1 硬件基础

理解 VR 如何工作，第一步是考虑整个 VR 系统的构成。人们很容易认为它只是硬件组件，例如电脑、头戴式设备和控制器，这是不全面的。如图 2.1 所示是 VR 系统的组织构成。硬件会产生覆盖用户感官的刺激。在第 1.3 节的达摩克利斯之剑（图 1.29（b））中，回想一下，我们需要用跟踪并根据人体的运动来调整刺激。VR 硬件通过使用自己的传感器来完成这一点，从而跟踪用户的运动。头部跟踪是最重要的，但跟踪也可能包括按钮按压，控制器移动，眼球运动或任何其他身体部位的移动。最后，将周围的物理世界视为 VR 系统的一部分也很重要。尽管 VR 硬件提供了刺激，但用户总是会有其他感官响应来自现实世界的刺激，他还有能力通过身体运动改变他的环境。VR 硬件还可能跟踪用户以外的对象，特别是如果与他们进行交互是 VR 体验的一部分。通过自动化界面，VR 硬件也可能改变现实世界，比如通过 VR 界面遥控机器人。

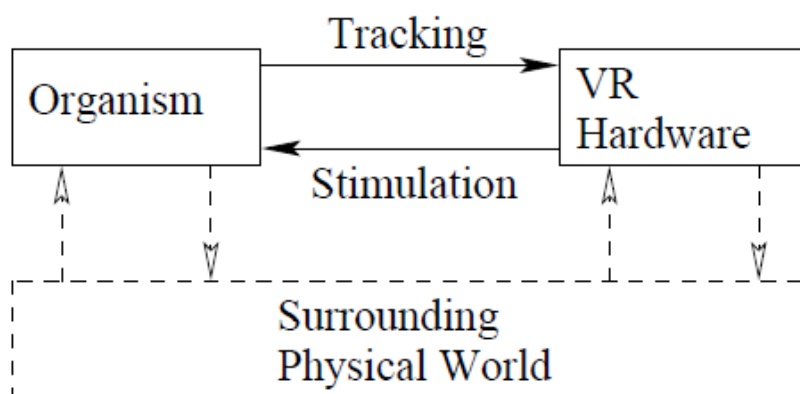


图 2.1 VR 系统的第三人称视角。设计的硬件和软件就是完整的 VR 系统，这种假设是错误的：人及其与硬件的交互同样重要。此外，在 VR 体验中，与周围物理世界的交互仍然存在。

### 传感器和感官

传感器和感觉器官如何从现实世界提取信息？显然，这对于 VR 系统至关重要。在工程中，传感器是指将能量从一种形式转换为另一种形式的装置。传感器是一种特殊的转换器件，可将接收到的能量转换为电信号，这取决于电路类型，可能是模拟或数字信号。传感器通常具有收集用于转换的能量接收器，有机体以相似的方式工作。“传感器”被称为感官，比如眼睛和耳朵。因为我们的“电路”是由相互连接的神经元形成的，所以感官将能量转化为神经冲动。在阅读本书时，请留心传感器与感官之间的相似之处，他们测量的东西相同，有时甚至以类似的方式运作。这不应感到意外，因为我们和我们的工程设备共享相同的物理世界：物理和化学定律保持不变。

### 感觉器官的空间构造

当用户在物理世界中移动时，他的感觉器官随之移动。此外，一些感觉器官相对于身体骨骼移动，例如我们的眼睛在其眼窝内旋转。每个感官器官都有一个空间构造，它对应于所有可能的方式进行转换或配置。这个最重要的方面是感官自由度的数量。第3章将全面介绍这一点，但现在要注意的是，在普通空间中移动的刚体有六个自由度。三个自由度与其在空间中的变化位置相对应：1) 侧向运动，2) 垂直运动，以及3) 远近运动。另外三个自由度对应于物体可能的旋转方式；换句话说，需要恰好三个独立的参数来指定对象如何定向。这些被称为偏航，俯仰和滚转，并在3.2节中介绍。

举个例子，考虑你的左耳。当你旋转你的头部或者在空间中移动你的身体时，耳朵的位置以及它的方向都会改变，这产生六个自由度。右眼也是如此，但它也能够独立于头部旋转。请记住，我们的身体有更多的自由度，这会影响到我们感官的构造。跟踪系统可能是必要的，以确定接收人工刺激的每个感觉器官的位置和方向，接下来会尽快对其进行解释。

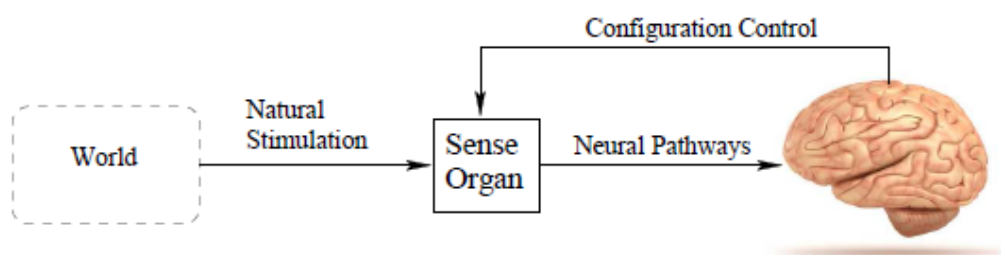


图 2.2: 在正常情况下，大脑（和身体部位）在接受来自周围物理世界的自然刺激时控制感觉器官（眼睛，耳朵，指尖）的配置。

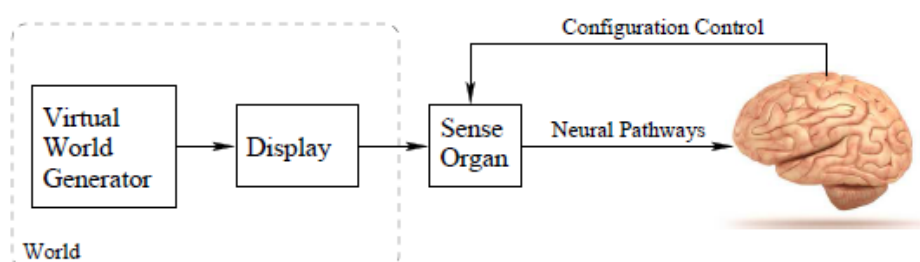


图 2.3: 与图 2.2 相比，VR 系统通过用称为显示器的硬件提供的人工刺激代替自然刺激来“劫持”每种感觉。通过使用计算机，虚拟世界生成器维持接通的虚拟世界。这个虚拟世界以适当的“视图”来呈现给显示器。

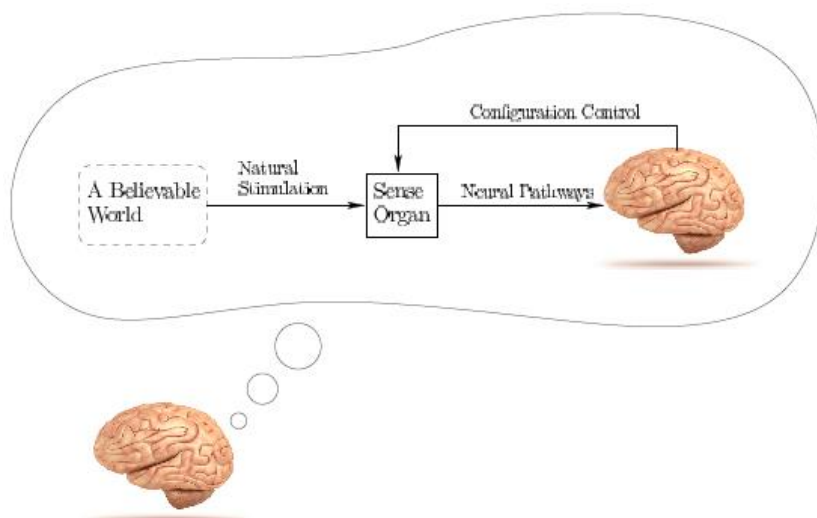


图 2.4: 如果做得好，大脑会被“愚弄”并相信虚拟世界实际上是周围的物理世界，自然的刺激源于其中。

## 抽象视图

图 2.2 说明了在没有 VR 硬件的干扰下，我们的感觉器官的正常操作。大脑控制着它的构形，而感觉器官将来自环境的自然刺激转化为发送到大脑的神经冲动。图 2.3 显示了它如何体现在 VR 系统中。VR 硬件包含了几个部分，后续我们将对其进行讨论。虚拟世界生成器（VWG）运行在计算机上并产生“另一个世界”，这可能有许多可能性，例如合成世界的完全模拟，对真实世界的记录或与另一部分真实世界的实时连接。人类通过每个使用显示器的目标感觉器官来感知虚拟世界，该感官感知到专门设计用于模拟在无 VR 的情况下将出现的刺激类型的能量。将来自 VWG 的信息转换为显示输出的过程称为渲染。对于人眼，显示器可能是智能手机屏幕或视频投影仪的屏幕。对于耳朵，这被称为扬声器。（即使这是日常生活中常见的用法，显示器也不一定是可见的。）如果 VR 系统是有效的，那么大脑有希望被“愚弄”，如图 2.4 所示。用户应该相信感官的刺激是自然的，并且来自合理的世界，与一些过去的经验一致。

## 听觉：全局固定与对象固定

从第 1.3 节中回想一下这样一种趋势：从不得不去某个地方体验一下，到可以在家中体验，再到最终让它变得完全便携。为了理解 VR 系统的这些选择及其对技术的影响，比较一个更简单的情况是有帮助的：音频或听觉系统。

图 2.5 显示了杜比 7.1 环绕声系统的扬声器设置和听众位置，可安装在剧院或家庭娱乐室。分布在房间周围的七个扬声器产生大部分声音，而低音音响（“7.1”的“1”）提供最低频率的部件。因此，听觉显示是全局固定的。将此与佩戴耳机的听众相比较，如图 2.6 所示。在这种情况下，听觉显示是对象固定的。希望你已经多次经历过类似这样的设置。

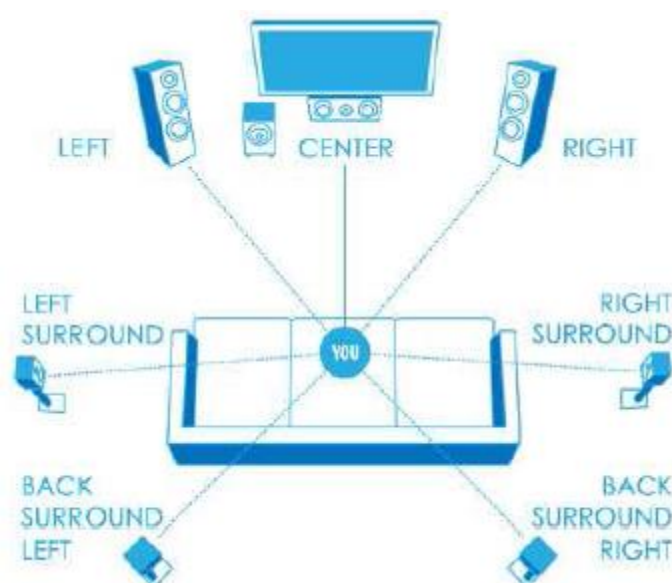


图 2.5：在环绕声系统中，当用户从中心收听时，听觉显示器（扬声器）是全局坐标中固定的。



图 2.6: 使用耳机，显示器是用户固定的，与环绕声系统不同。

它们之间关键的区别是什么？除了头戴式耳机明显的便携性之外，下面很快就会想到：在环绕声系统中，生成的声音（或刺激）远离耳朵，但耳机实则非常接近。

距离差异的含义是与远端扬声器相比，若产生相等的感知响度水平，耳机需要的功率要少得多。距离的另一个意义是耳机佩戴者可允许的隐私程度。高音量的环绕声系统可能会引起街坊邻居的愤怒。长时间戴头戴电子设备可能会感到不舒服，从而更偏爱环绕声。可同时几个人在环绕声系统中享受相同的体验（尽管他们不能全部坐在最佳位置）。如果使用耳机，那他们将需要将音频源同时分配到各自的耳机里。根据制造难度和可用的组件技术，它们可能会有不同的成本。目前，耳机比一套环绕声扬声器成本低得多（尽管可以花费大量的金钱）。所有这些差异都会传递到 VR 系统。这应该不会让人感到意外，因为根据第 1.1 节的定义，我们可以轻松地将纯音频体验视为一种特殊的 VR 体验。

在听音乐的同时，闭上你的眼睛，想象你周围环绕着艺术家，你正在现场欣赏演奏。你认为艺术家和他们的乐器在哪里？他们在你周围，还是他们似乎在你的头顶？使用耳机，他们很可能在你的脑袋里。在环绕声系统中，如果正确记录并将其显示，声音应该看起来好像来自它们原来位于头部以外的位置。它们可能感觉起来不太自然，然而却进入你的水平面。

耳机的这个缺点目前尚未得到广泛认可，但对于包含比如视频显示器的 VR 系统而言，这个问题变得尤为严重。如果你想保留你对声音来自哪里的感知，那么耳机就需要考虑你的耳朵在空间中的配置来相应地调整输出。例如，如果您以“不”手势左右摇头，则需要调整呈现给每只耳朵的声音，以便模拟声源以相反的方向旋转。在环绕声系统中，扬声器不会跟随您的头部，因此不需要旋转。如果扬声器随着您的头部旋转，则需要反向旋转以“抵消”你的头部旋转，以使声源位置被认为是静止的。

### 视觉：全局固定与对象固定

现在考虑添加一个可视化显示。在听音乐时，您可能不太担心艺术家和乐器的位置，但您很快就会注意到他们的位置是否显得不正确。我们的视觉感比我们的听觉感更强大和复杂。图 2.7 (a) 显示了一个 CAVE 系统，它在很多方面与环绕声系统相似。用户再次坐在中央，周围的显示器给他的眼睛带来视觉刺激。扬声器被视频屏幕取代。图 2.7 (b) 显示了佩戴与头戴式耳机相似的 VR 头戴式耳机的用户。

假设用户眼前的屏幕显示了在耳机中固定的图像，如果用户旋转他的头部，则图像将被视为与头部相连。例如，如果在使用 Viewmaster 时旋转头部（回忆图 1.28 (b)），就会发生这种情况。如果您想将图像视为周围固定环境的一部分，那么在您旋转头部时，头戴



式耳机中的图像必须更改以进行补偿。周围的虚拟世界应该是反向旋转的，其意义将在第 3.4 节中更加精确。



图 2.7: (a) 在英国 Teesside 大学开发的 CAVE VR 系统。(b) 2013 年,一位 90 岁女性 (Rachel Mahassel) 戴着 Oculus Rift DK1 耳机。

一旦我们认可这种转变是必要的,那么估计已发生的头部和眼球运动的量并且以及时和准确的方式应用适当的转换便成为了重大的工程挑战。如果处理得不好,那么用户可能会有不好的用户体验。更糟糕的是,他们可能会成为 VR 疾病的牺牲品。这是 VR 头戴式耳机在九十年代流行的主要原因之一,而且组件技术还不够好。幸运的是,目前情况有了很大改善。对于音频而言,似乎很少有人对这种转变感到困扰,但对于视觉对象来说,这绝对是关键。最后一点需要注意的是,如果我们希望屏幕上的图像根据房间内眼睛位置的变化而改变,那么 CAVE 系统中的跟踪和应用转换也是必要的。

现在您已经对通用硬件架构有了高层次的了解,我们将深入研究构建 VR 系统时可用的硬件组件。预计这些变化很快,成本下降,性能提高。我们也预计未来几年将有许多新设备出现在市场中。尽管如此,本书的基本面仍然没有改变。依据当前技术的了解提供具体的例子,使 VR 基本概念更加清晰。

VR 系统的硬件组件可以方便地分类为:

- 显示(输出)**: 刺激感官的装置。
- 传感器(输入)**: 从现实世界提取信息的设备。
- 计算机**: 依次处理输入和输出的设备。

## 显示器

显示器的目的是为目标感觉器官生成刺激。视觉是我们的主要意义,任何为眼睛构造的显示都必须在视网膜上形成所需的图像。由于这个重要性,第 4 章和第 5 章将分别解释光学系统和人类视觉系统。对于 CAVE 系统,则对数字投影仪和镜像进行了组合。由于成本暴跌,一系列大屏幕显示器可能会被采用。头戴设备可以将智能手机显示屏靠近眼睛放置,并为每只眼睛使用一个放大镜将其聚焦。屏幕制造商目前正在通过利用智能手机行业的最新 LED 显示技术制造 VR 头戴式定制显示器。有些厂商则瞄准每个眼睛配置一个显示器,帧速率高于 90Hz,单眼超过两百万像素。第 5 章解释了这个原因。



图 2.8: 触觉反馈设备的两个例子。(a) Geomagic Punk 允许用户用真实的触针刺入虚拟物体时感受到强大的阻力。机器人手臂提供适当的力量。(b) 有些游戏控制器偶尔产生振动。

为 VR 定制的新显示技术正在迅速兴起。通过使用微型投影技术（包括 DLP（数字光处理），LCD（液晶显示器）和 LCoS（硅上的液晶））可以直接实现视网膜刺激。使用此技术的产品包括 Google Glass, Microsoft HoloLens 和 Avegant Glyph。为了解决舒适性问题，如聚散度调节不匹配（参见第 5.4 节），更多的显示技术已经进行了原型设计，但仍然可以在像素密度，视场，帧率，可制造性，成本等方面与其他显示技术相媲美。两个主要的类群是光场显示器[74,161,198]和多焦平面显示器[4,127,187]。

现在设想展现其他感官。使用经典扬声器技术将声音传送到耳朵，也可以使用骨传导方法，其使头骨振动并将波传播至内耳；这种方法被运用在 Google Glass 上。第 11 章详细介绍了 VR 的听觉部分。对于触觉而言也会有触觉的显示，图 2.8 给出了两个例子，触觉反馈可以以振动，压力或温度的形式给出。有关触觉，味觉和嗅觉的更多细节，请参阅第 13 章。

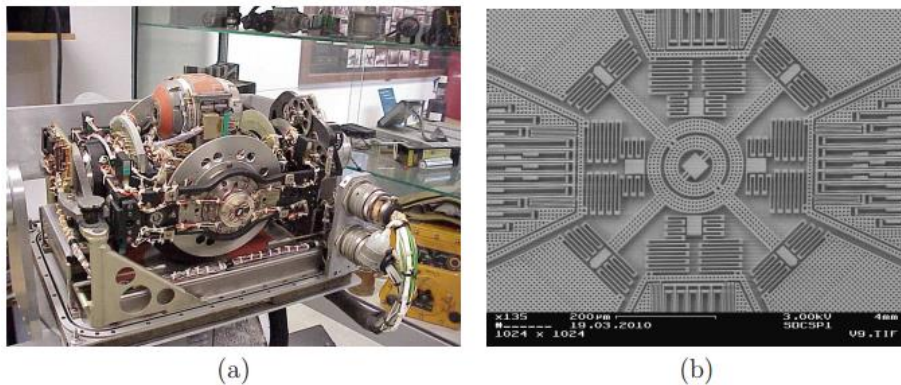


图 2.9 惯性测量单元（IMU）已经从大型重型机械系统转向廉价微型 MEMS 电路。(a) Litton Industries 在 20 世纪 60 年代开发的 LN-3 惯性导航系统。(b) MEMS 陀螺仪的内部结构，其总宽度小于 1mm。

## 传感器

让我们考虑一下 VR 硬件的输入端。这里只是给出一个简要概述，我们会在第 9 章详细的介绍传感器和跟踪系统。对于视觉和听觉的显示器，人体感官的位置和方向必须由传感器进行跟踪以适当调整刺激。定向部分通常由惯性测量单元或 IMU 完成。主要部件是一个陀螺仪，它测量自己的旋转速度；该速率被称为角速度并且具有三个分量。来自陀螺仪的测量值随时间增长以获得方位累积变化的估计值。所导致的误差称为漂移误差，如果不使用其他传感器，其值将逐渐增加。为了减少漂移误差，IMU 还包含一个加速计和一个磁力计。多年来，IMU 已经从曾经只存在于飞机和导弹等大型机械系统中到变成了手机中的小型设备，见图 2.9。由于其体积小，重量轻，成本低，IMU 可以轻松嵌入可穿戴设备中。它们是当前

VR 头戴式耳机最重要的支持技术之一，主要用于跟踪用户的头部方向。

数码相机为跟踪系统提供了另一个重要的信息来源。与 IMU 类似，由于智能手机行业的需求，它们已变得越来越便宜和便携，同时也提高了图像质量。相机运用了“视线可见性”的跟踪方法，这个方法可识别图像中用作移动物体或静止背景参考点的特征或标记。这种可见性约束严重限制了可能的对象位置和方向。标准相机通过将光线聚焦于光学系统而形成图像，非常像人眼的成像原理。一旦摄像机校准参数已知，则可以利用观察到的标记去定位空间中的光线。照相机通常用于跟踪眼睛，头部，手部，整个人体以及物理世界中的其他物体。目前面临的主要挑战之一是：当不在用户或场景周围放置特殊标记时，也可获得可靠和准确的性能。

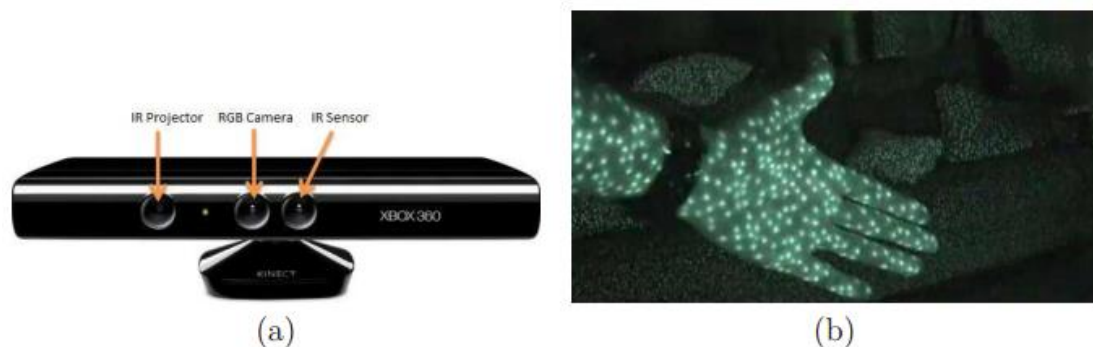


图 2.10: (a) Microsoft Kinect 传感器收集普通的 RGB 图像和深度图（每个像素距离传感器的距离）。(b) 通过观察从 IR 相机获得的图像中投影的 IR 点的位置来确定深度。

与标准相机相反，深度相机通过将光投影到场景中并观察其在影像中的反射来工作。这通常是在红外（IR）光谱中完成的，因此人眼不会注意到，见图 2.10。

除了这些传感器之外，我们还非常依赖古老的机械开关和电位计来创建键盘和游戏控制器，光学鼠标也是常用的。这些熟悉设备的一个优点是用户可以利用他们现有的知识快速输入数据或控制他们的角色。一个缺点是，如果他们的面部被头戴式耳机覆盖，他们可能很难互动。

## 计算机

虚拟世界来自于计算机。而这台电脑应该在哪里？尽管对于“world-fixed”的显示器并不重要，但对于“body-fixed”的显示器而言，该位置至关重要。如果在系统中需要一台独立的 PC，则必须在头戴式设备和 PC 之间提供快速，可靠的通信。这种连接目前通过电线进行，而有线总是带来体验的不适感，但目前的无线速度又是不够的。正如你已经注意到的，大多数传感器都存在于智能手机上，以及功能适中的计算机上。因此，可以将智能手机放入带有镜头的盒子中，以提供几乎不增加成本的 VR 体验（图 2.11）。但是，这样的做法的一个限制是模拟出的虚拟世界必须比在 PC 上运行要更简单，以便它可以在不太强大的计算硬件上运行。在不久的将来，我们希望看到一体式的无线多功能头戴式设备，以提供 VR 体验。在这种一体机中将消除智能手机中不必要的组件，而是专门为 VR 量身定制光学器件，微芯片和传感器。





图 2.11：通过将智能手机放入盒子创建 VR 体验的两款头戴式设备。（a）Google Cardboard，适用于各种智能手机。（b）三星 Gear VR，针对特定的智能手机进行了优化（本例中为三星 S6）。

在 VR 系统中，除了主要的计算系统之外，还需要使用专门的计算硬件。图形处理单元（GPU）已经经过优化可以快速将图形渲染到屏幕上，并且目前正在适应处理 VR 的特定性能需求。而且，显示接口芯片将输入视频转换为一条条的显示命令，最后，微控制器从传感设备收集信息并使用标准协议（如 USB）将它们发送到主计算机。



图 2.12：Oculus Rift DK2 耳机的拆卸（摄图为 ifixit）

图 2.12 显示了 Oculus Rift DK2 的硬件组件，该组件在 2014 年底推出。在图中的左下角可以看到一个智能手机屏幕作为显示屏，在图中上方还包含 IMU，显示接口芯片，USB 驱动芯片，用于驱动耳机上的 LED 以追踪 LED 的芯片集以及可编程微控制器的电路板，右下角是镜头，它可以使智能手机屏幕看起来“无限远”，但仍填满了用户视野的大部分区域。右上方显示的是灵活的电路，为嵌入在耳机中的 IR LED 提供电源（它们隐藏在红外透明塑料后面）。相机用于跟踪，其部件显示在上图的中心部分。

## 2.2 软件

从开发人员的角度来看，通过提供高级描述和让软件自动确定所有低级细节来编程实现 VR 系统将是一个理想的选择。在一个近乎真实的虚拟世界里，会有一个 VR 引擎，它的用途类似于当今用于制作视频游戏的游戏引擎。如果开发者在实现之前遵循许多模式，那么只需调用精心设计的软件库中的函数即可避免许多复杂的细节。但是，如果开发人员想要尝试一些原创的东西，那么将不得不从头开始设计这些功能。这需要对 VR 基础知识有更深入的了解，同时也熟悉底层级别的系统操作。

不幸的是，我们距离拥有功能齐全的通用 VR 引擎还有很长的路要走。随着虚拟现实技



术的应用越来越广泛，专门的 VR 引擎也可能出现。例如，一个可能是针对沉浸式电影摄影，而另一个则是面向工程设计。哪些组件将变得更像 VR“操作系统”的一部分，并将成为更高级别的“引擎”组件？鉴于目前的情况，开发人员可能会从零开始实施其 VR 系统的大部分功能。这可能涉及利用软件开发工具包（SDK）处理特定头戴式设备，以处理最底层级别的操作，例如设备驱动程序，磁头跟踪和显示输出。另外，他们可能会发现自己使用最近已经适用于 VR 的游戏引擎，尽管它基本上是为屏幕上的视频游戏而设计的。这可以首先避免大量的工作，但是当有人想要实现不属于标准视频游戏时，这可能会比较困难。

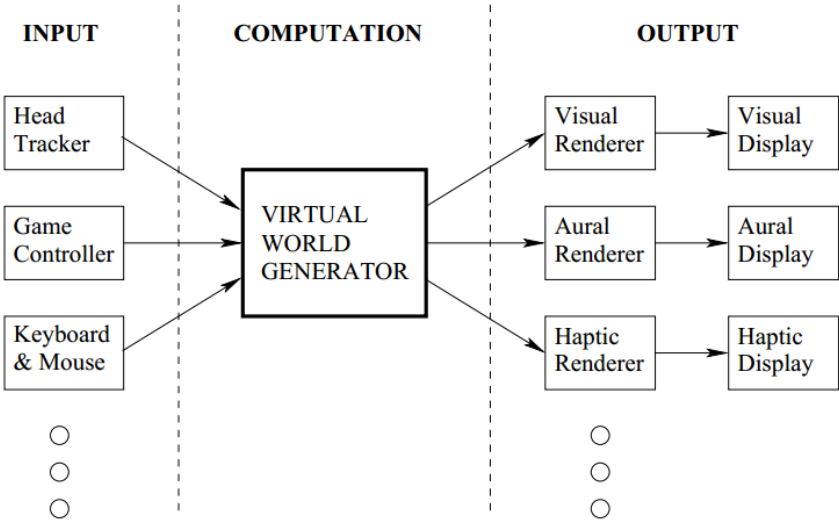


图 2.13：虚拟世界生成器（VWG）维护着另一个世界，它可能是合成的，真实的或者某种组合。从计算角度来看，输入是从用户和他的周围环境中接收到的，并且适当的场景被呈现给显示器。

需要哪些软件组件来产生 VR 体验？图 2.13 展示了一个高级视图，强调了虚拟世界生成器（VWG）的核心作用。VWG 接收来自低级系统的输入，表明用户在现实世界中正在做什么。头部跟踪器提供用户头部位置和方向的及时估计。键盘，鼠标和游戏控制器事件到达可准备处理的队列。VWG 的关键作用是保持足够的内部“现实”，以便渲染者可以提取他们需要的信息来计算其显示的输出。

### 虚拟世界：真实与合成

虚拟世界可以被完全合成。在这种情况下，许多三角形态在 3D 空间中定义，以及指示它们如何与光，声音，力等相互作用的材质属性。在计算机图形学领域较为认可的一个观点是，在合成模型中计算机生成的图像对 VR 来说非常重要，可参见第 7 章。另一方面，虚拟世界刻画了一个还原纪实的物理世界，它运用了现代相机，计算机视觉和同步定位和映射技术（SLAM），如图 2.14。极端之间存在许多可能性。例如，可以拍摄真实物体的相机图像，然后将其映射到虚拟世界中的合成物体上，这被称为纹理映射，是计算机图形学中的一种常见操作，见第 7.2 节。



图 2.14：使用摄像机的颜色和深度信息，可以使用同步定位和映射（SLAM）技术自动提取 3D 模型。

## 匹配运动

VWG 的最基本操作是保持现实世界中的用户运动与虚拟世界之间的对应关系，见图 2.15。在现实世界中，用户的运动被限制在一个安全区域，我们称之为匹配区域。想象一下，匹配区域是真实和虚拟世界完美对齐的地方。最大的挑战之一就是障碍的不匹配：如果用户在虚拟世界中碰到了障碍而在现实世界中没有的话，该怎么办？在一个有座椅设备的体验中，用户戴着头戴式设备坐在椅子上。在这种情况下，匹配区域是一个小区域，例如一立方米，用户可以在其中移动头部。头部动作在两个世界中应该是相匹配。如果用户不受限于座位，则匹配的区域可以是整个房间或室外场地。请注意这里安全性成为一个问题，因为用户可能会饮酒，撞墙，或陷入只存在于现实世界中但在虚拟世界中不可见的凹坑。较大的匹配区域往往导致更大的安全问题。用户必须确保匹配的区域在现实世界中没有危险，否则开发人员应该在虚拟世界中使其可见。

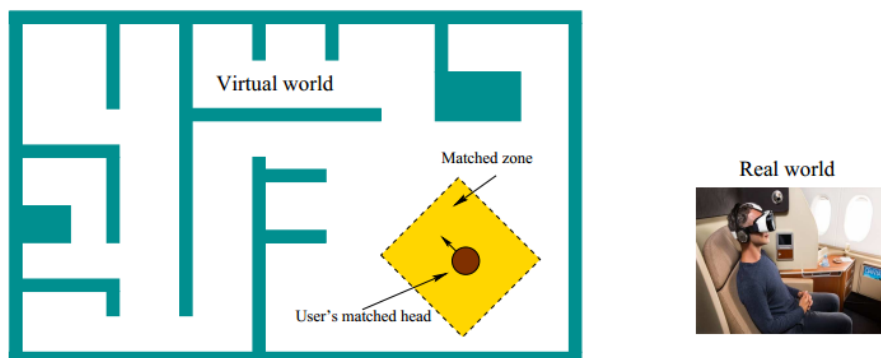


图 2.15：匹配区域是指用户在现实世界和虚拟世界中的一个显示关系。用户可以通过使用诸如游戏控制器等在虚拟世界中移动，而用户在现实世界中不会进行相应地移动。

虚拟世界中哪些动作应该反映在虚拟世界中？这在不同的 VR 体验中各不相同。在向头部显示图像的 VR 头戴式设备中，头部运动必须匹配，以便视觉渲染器在虚拟世界中使用正确的视点。身体的其他部位的匹配不那么重要，但如果用户需要进行手眼协调或查看身体的其他部位并期望实现自然移动时，则会变得十分重要。

## 用户运动

在许多 VR 体验中，用户想要在匹配区域外移动。这意味着用户在虚拟世界中自我移动，而这个移动在现实世界中是不匹配的。想象一下，你想要在真实世界中坐下来探索虚拟城市。这应该如何实现？你可以拉一张地图，并指向你想要去的地方，用一个快速的传送操作把你

送到目的地。一种流行的选择是通过操作游戏控制器，鼠标或键盘来在虚拟世界中移动自己。通过按下按钮或移动旋钮，你在虚拟世界中实现步行，跑步，跳跃，游泳，飞行等等操作。你也可以坐上虚拟世界中的一辆车，并操纵其控制来移动自己。这些操作当然很方便，但由于平衡和视觉感官不匹配，往往会导致晕眩，具体请参阅第 2.3,10.2 和 12.3 节。

## 物理学

VWG 通过应用适当的数学变换来处理运动的几何方面。另外，VWG 通常实施一些物理学，以便随着时间的推移，虚拟世界的行为就像现实世界一样。在大多数情况下，力学的基本定律控制物体在虚拟世界中的移动方式。例如，如果你放下一个物体，那么它应该由于作用在其上的重力而加速到地面。一个重要的组成部分是碰撞检测算法，它确定两个或多个物体是否在虚拟世界中相交，如果发生碰撞，则需要适当的响应。例如，假设用户在虚拟世界中穿过墙壁探出头部。虚拟世界中的头部是否应该停下来，即使它在现实世界中继续前进？为了使它更复杂，如果你将装满篮球的大货车卸载到虚拟世界的一条繁忙街道上会发生什么？模拟物理学可能变得非常具有挑战性，并且这本身就是一门学科。有关虚拟世界物理的更多信息，请参见第 8.3 节。

除了处理运动物体的运动特征之外，物理学还必须考虑如何创建不可见的介质。比如，光线如何在环境中传播？光线如何与虚拟世界中的表面相互作用？什么是光源？声音和气味如何传播？这些与渲染问题相对应，分别在视觉和音频案例的第 7 章和第 11 章中介绍。

## 网络体验

在互联网 VR 体验的情况下，共享虚拟世界由服务器维护。每个用户都有一个独特的匹配区域。它们的匹配区域可能在现实世界中重叠，但必须小心，以避免不必要的碰撞。大多数情况下，这些区域是分离的，分布在地球周围。在虚拟世界中，用户交互（包括冲突）必须由 VWG 管理。如果多个用户在社交场合中互动，则匹配动作的负担可能会增加。当用户相互见面时，他们可以期望看到眼部、面部表情和肢体语言，可参见 10.4 节。

## VWG 的开发人员选择

总而言之，开发人员可以从 VR 头戴式设备供应商的基本软件开发套件（SDK）开始，然后从头开始构建自己的 VWG。SDK 应提供基本的驱动程序和接口，以访问跟踪数据并调用图形渲染库。在这种情况下，开发人员必须从头开始构建虚拟世界的物理模型，处理诸如头像移动，碰撞检测，照明模型和音频等问题。这为开发人员提供了最大限度的控制和优化性能的权限，然而在有些情况下它可能会换来实施上的困难，但在一些特定的情形中它可能又不会变的太过复杂。例如，对于 Google Street 浏览器（如上一章的图 1.9），其物理模型很简单：观看位置需要以方便的方式在全景图像之间来回切换，同时又要保持观看者在地面的位置感。在使用机器人进行远程呈现的情况下，VWG 将不得不考虑在物理世界中的移动。如果未能处理碰撞检测，可能会导致机器人破损，甚至人受伤。

另一方面，开发人员可以使用现成的 VWG，通过选择菜单选项和编写高级脚本来定制 VWG，以制作特定的 VR 体验。目前可用的例子有 OpenSimulator，World-Viz 的 Vizard，Unity 3D 和 Epic Games 的虚幻引擎。后两者是适用于 VR 的游戏引擎，并且是目前 VR 开发者中最受欢迎的游戏引擎。OpenSimulator 是 Second Life 的替代产品，用于构建虚拟社会。如前所述，使用这种更高级别的引擎可以让开发人员在短时间内轻松制作 VR 体验；然而，其缺点是难以做出制造商无法想象的高度原创体验。

## 2.3 人体生理学与认知

人类的身体并非是为 VR 而设计的。通过对感官进行人工刺激，我们正在破坏生物机制的运作，这些机制经历了数亿年时间在自然环境中演变而来。我们也向大脑提供与现实体验不完全一致的信息。在某些情况下，我们的身体可能会适应新的刺激，这可能会使得我们无法察觉 VR 系统中的缺陷。在另一些情况下，我们可能会提高解读含义不清的 3D 场景的能力。但不幸的是，在有很多情况下，我们的身体会产生疲劳或头痛等症状，部分原因是大脑比平常更高速地运转，以理解这类刺激。而最糟糕的情况是 VR 疾病的发作，通常会出现眩晕和恶心的症状。人体感知心理学是理解大脑如何将感官刺激转化为感知现象的科学。以下是 VR 中出现的一些典型问题，属于这个范畴：

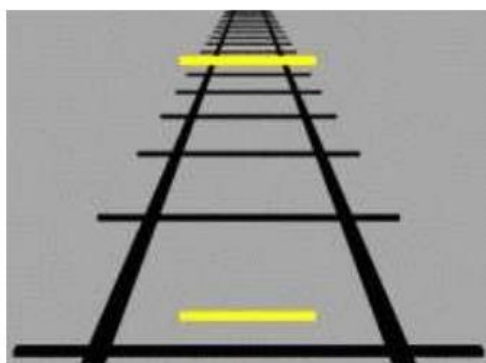
- 该对象看起来有多远？
- 需要多少视频分辨率才能避免看到像素？
- 每秒多少帧的刷新率可以将运动视为连续性的？
- 用户的头部是否出现在虚拟世界的适当高度？
- 虚拟声音来自哪里？
- 为什么会感到眩晕？
- 为什么一个人的体验比另一个人更累？
- 什么是存在？

要回答以上甚至更多的问题，我们必须了解以下几点：1) 人体的基本生理结构，主要包括感官和神经通路；2) 基于实验的感知心理学的关键理论和见解；3) VR 系统对人类通常的感知过程产生的影响或副作用。

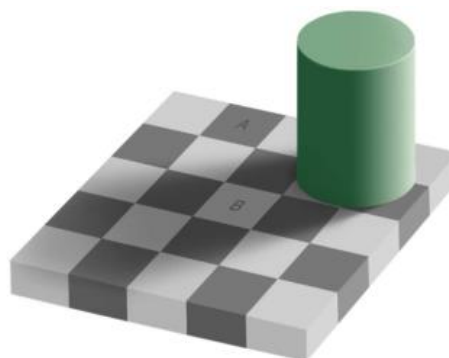
虚拟现实的感知往往无法引起开发人员的重视。在现实世界中，感知过程大多是我们看不见的。想想认出一个家庭成员需要花费多少精力，当你看到一个你熟悉的人时，这个过程会自动开始，立即结束，似乎不需要付出任何代价。科学家们进行了实验，揭示了与此类似的感知过程中究竟发生了多少工作。通过脑部神经研究，他们能够看到一小部分脑功能不正常时的影响。有些人患有面部失认症，这使得他们即使在几乎所有其他功能都正常时，也无法认出熟悉的人的面部，包括他们自己在镜子中的形象。科学家们也能够对大脑中某个神经元的状态进行记录，主要在动物身上进行实验，这可以揭示单个神经元对感官刺激的反应。

### 视觉错觉

认知人类感知复杂性的最普遍的一种方式研究视觉错觉。视觉错觉会产生人们平常毫不在意却令人惊讶的结果。该项研究通过提供与日常生活中普通刺激不太相同的刺激来揭示人类视觉系统的一些缺陷。下图展示了其中的两种错觉。这些结果或许会激励你去了解人体感官和神经结构每一次感知所要进行的工作，并根据生活经验和生物结构对此进行解释。当然，在不理解感知过程的情况下对其进行干扰是十分不明智的。



(a)



(b)



图 2.16 视觉错觉是一种非同寻常的视觉刺激，凸显出了人类视觉系统的局限性。（a）Ponzo 错觉导致上部线段比下部线段显得更大，即使它们的长度相同。（b）阴影错觉使得 B 瓦片比 A 瓦片看起来更亮，尽管它们完全相同的灰度

感知分类

感知与错觉不仅限于人眼。图 2.17 显示了基本感觉的分类。其中传感器将能量源转换为电路中的信号。就我们的身体而言，这意味着刺激会转化为神经冲动。对于每种感觉，图 2.17 说明了刺激和将刺激转化为神经冲动的受体类型。将每个受体看作是针对特定类型刺激的传感器。这被称为感知系统选择性。在单个眼睛内，有超过 1 亿个光感受器精确捕捉可见光频率范围内的电磁能量。不同种类的光感受其甚至可以针对不同的颜色和光照水平，具体见 5.1 节。听觉，触觉和平衡感涉及运动，振动和重力，这些刺激被机械感受器感知。有关于听觉的生理知识还会在 11.2 节和 11.3 节中介绍。触觉还包括温度的感知，有关触觉的内容还将在 13.1 节中介绍。平衡感有助于我们知道头部朝向哪个方向，包括“向上”的方向，将在 8.2 节继续介绍。最后，我们将味觉和嗅觉归为一类，称为化学感觉，依赖于化学感受器，这类信号的产生基于在舌头或鼻腔中物质的化学成分，详见 13.2 节。

感觉种类	刺激形式	接收单元	感官
视觉	电磁能量	感光器	眼睛
听觉	空气压力波	机械感受器	耳朵
触觉	组织扭曲	机械感受器	皮肤，肌肉
		a 热觉感受器	皮肤
平衡感	引力，加速	机械感受器	前庭器官
味道/气味	化学成分	化学感受器	嘴，鼻子

图 2.17 人体感觉分类

要注意的是，这些感官具有等价物体，其中大部分会出现在 VR 系统中。想象一下，假如你设计了一款远程呈现的人形机器人，并期望通过 VR 设备进行交互。你可以通过代理机器人体验生活。数码相机将成为它的眼睛，而手机将成为耳朵。机器人身上可以安装压力传感器和温度计以提供触感。至于平衡感，可以安装一个 IMU。事实上，人类前庭器官和现代 IMUs 在产生信号方面有惊人的相似之处，参见 8.2 节。我们甚至可以安装化学传感器，如 pH 计，测量化学成分的各个方面，以提供味道和气味。

大脑

感觉发生在感官将刺激转换为神经冲动之后。根据最新研究[15]，人体含有约 860 亿个神经元。大约有 200 亿个神经元为大脑皮层工作，处理知觉和许多其他更高级的功能，如注意力，记忆力，语言和意识。它是一个约三毫米厚的神经中枢，大量折叠以适配头骨结构。再如其他生物，蛔虫，果蝇和大鼠分别有 302,10 万和 2 亿个神经元。一头大象拥有超过 2500 亿个神经元，这比人类的还多！

只有哺乳动物有一个大脑皮层。老鼠的大脑皮层有约 2000 万个神经元。猫和狗分别为 300 万和 1.6 亿。大猩猩有大约 40 亿。一种名为长肢领航鲸的海豚，其大脑皮层中估计有 370 亿个神经元，大约是人类大脑皮质的两倍。然而，科学家声称更多的神经元并不意味着更良好的认知能力[220,273]。

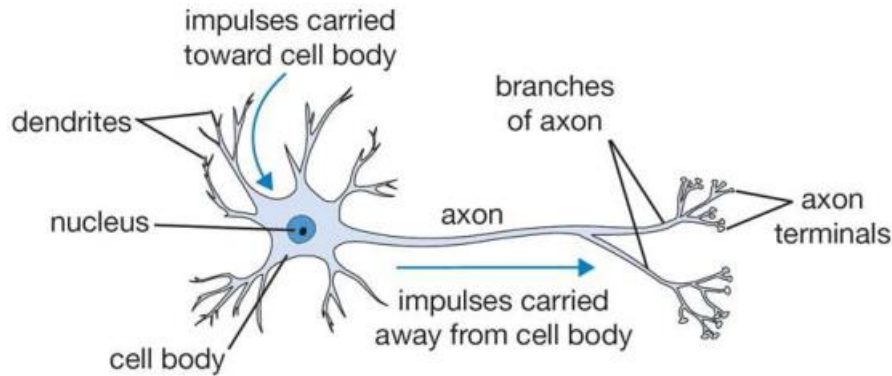


图 2.18 典型神经元，通过树突接收信号

感知和整体认知能力的另一个重要因素是神经元之间的相互联系。想象一个巨大的有向图，具有通常的节点和有向边。每个神经元的核或细胞体便是某种“处理”的节点。图 2.18 为我们展示了一种典型的神经元。树突是神经元的输入边缘，而轴突是输出边缘。通过树突网络，神经元可以汇集来自众多其他神经元的信号，这些神经元本身可能已经汇总了来自其他神经元的信号。结果通过轴突发送给一个或多个神经元。对于连接的轴突-树突对，信息交互发生在称为突触的中间结构中，其中传递电信号或化学信号。人脑中的每个神经元与其他神经元平均具有约 7000 个突触连接，这导致我们巨大的脑图中有约  $10^{15}$  个边缘！

### 分层处理

离开感官受体后，信号在神经元之间传播，最终到达大脑皮层。在传播途径上，实现分层处理，具体过程见图 2.19。通过选择性，每种受体对刺激在时间、空间、频率等各种方面的响应范围都变得相当窄。通过几个神经元后，将同时处理来自多个受体的信号，因而这种机制可以在刺激中检测到复杂的形式。在视觉感知情况下，特征检测器出现在早期的分层阶段，使人们能够检测到如边缘，角落和运动等特征。一旦进入大脑皮层，来自传感器的信号就会与我们生活中经历过的任何其他信息相结合，这些信息可能与刺激的解释有关，由此出现各种感知现象，例如识别脸部或识别歌曲。最终出现在大脑皮层的信息便映射出了我们周围甚至全景环境的情况。令人惊讶的是，有方法揭示，在神经元分布的某些情况下，受体间的空间关系可以维持不变。此外，在此前的 1.1 节中也已提到过空间识别细胞和定位细胞可以识别出熟悉环境的空间关系。



图 2.19 刺激信号的传播受体捕获的刺激通过神经元的分级网络传递。早期阶段，信号由多个受体组合而成，并向上传播。后期阶段，信息双向流动。

### 本体感觉

除了感官和记忆的信息之外，人体还具备本体感觉，这是一种能够感受到身体各部位相对位置并移动相应肌肉的行为。闭上你的眼睛，在空旷的地方移动手臂，尽管你可能无法精确地用手指进行触摸，但你应该大致知道你的手臂位于哪里。这种可以控制人体运动的信息对大脑来说非常重要，它向大脑其他部分发送一种名为“efference copies”的信号，传达身体已执行的运动。将这种感觉继续与机器人进行比较，它的等价物是在关节上安装的编码器，用来得到机器人移动的距离。因为本体感觉，从而有了以下这种有趣的现象：人们并不能把自己挠痒，因为我们知道自己的手指在哪里移动，但其他人却极有可能把你挠痒。所以这种

本体运动信息对搔痒感的产生是极为重要的。

### 感觉融合

来自多种感觉和本体感觉的信号可以被同时处理，并与我们的生活经历相结合。在日常生活中，没有虚拟现实或药物，我们的大脑对这些输入组合以一种连贯，一致和熟悉的方式进行解释。任何干扰这个处理过程的行为都可能导致感官数据的不匹配，此时大脑会以各种方式作出紧急反应，最后可能使得我们没有意识到这场冲突，但我们会感到疲惫不堪。更糟的是我们可能会出现头晕或恶心的症状。在其他一些情况下，大脑可能会让我们有意识地感受到冲突，以明白这是人为体验，这与目前 VR 体验还未能令人沉浸于虚拟世界的情况相对应。为了获得有效且舒适的 VR 体验，VR 试验最重要的是要了解大脑对于虚拟场景是如何反应的。

在虚拟现实的背景下，最糟糕的一种感官冲突就是动感，这是自我运动的幻觉。当你的视觉向大脑告知你正在加速时，冲突便会产生，此时你的平衡感会认为你并没有在动。而当人们真正走在大街上时，他们的平衡感和视觉是匹配的。即使没有 VR，我们也可能在之前就经历过这种感受。当你遇到交通拥堵看到正在向前移动的车辆的同时，你可能会觉得自己正在向后移动。在 19 世纪 90 年代，亚玛利雅湖建造了一套游乐设施，该设施包括一个始终保持静止的秋千，而围绕秋千的整个房间进行前后摇摆（图 2.20）。在 VR 中，相对运动错觉是由 2.2 节中描述的运动行为所引起的。例如我们使用控制器加速前进，那么只有我们的眼睛会感受到加速度，前庭器官并不能。有关缓解此问题的方法会在后续 10.2 节中介绍。

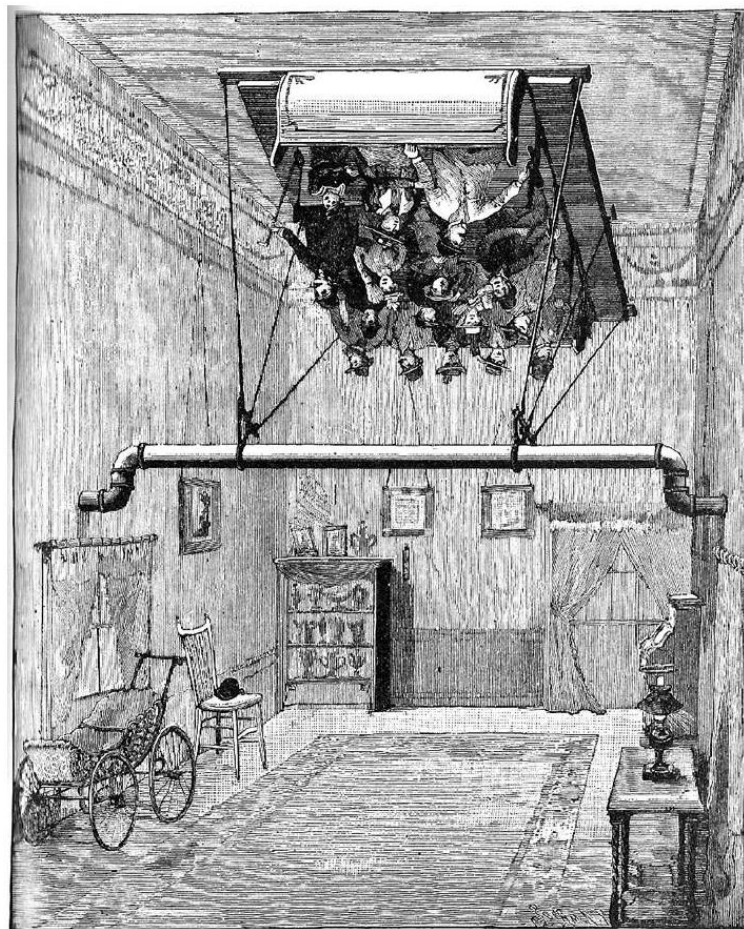


图 2.20 十九世纪九十年代，通过旋转周围的房间来创造虚拟的摇摆体验。这被称为摇摆幻觉。尝试过它的人既觉得有趣又会感到恶心。

## 适应性

人体感官系统的一个普遍特征是适应性，这意味着刺激的感知效果会随着时间而改变。例如，飞机或汽车的电机噪声感知到的响度会在几分钟内逐渐减小。对于视觉而言，人眼光学系统和感光器会调节以适应环境亮度。经过长时间的感知训练，适应能力也可以提升，参见 12.1 节。在军事训练中，士兵所经历的疾病似乎低于预期，可能是长期的风吹日晒所致 [171]。有趣的是，具有丰富经验的视频游戏玩家似乎也是如此。那些投入了大量时间和精力在第三人称射击游戏上的玩家，也可以更快地适应 VR 场景。

因此适应性也是 VR 的关键因素。通过反复的体验，开发者可能会对原本感到恶心的体验逐渐适应，但这也会对他们面向大众的开发产生不利。但从另一方面来讲，通过重复的有针对性的培训，开发人员在调试时会有更敏锐的观察能力，常见的例子包括：

- 大量跟踪延迟，这会对视野的平稳性产生影响。
- 左右眼视图已交换。
- 物体只存在于某只眼睛的视野中。
- 双眼显示图像的延迟不同。
- 直线扭曲。

## 心理物理学

心理物理学是对由物理刺激产生的感知现象的科学研究。例如，在什么情况下人们会说某个物体是“红色”的？刺激与进入眼睛的光线相对应，感知现象是大脑中形成“红色”概念的过程。感知现象的其他例子有“直线”，“更大”，“更响”，“发痒”，“酸”等等。图 2.21 展示了一个心理物理实验的典型场景。由于某个参数是可以变化的，例如光的频率，通常存在一定范围内的值，受试者不能可靠地对现象进行分类。例如，可能在某个区域内人们并不能确定光线是否为红色。在某种极端情况下，他们可能将其归类为“红色”，而在另一种极端情况下，他们则认为光线“不是红色”，对于介于两者之间的情况，记录检测的概率。在 12.4 节中还将继续讨论如何设计和实施这些实验。

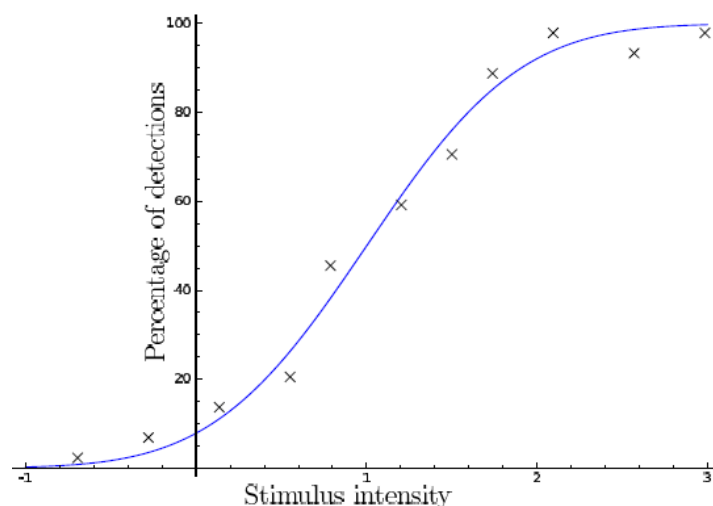


图 2.21 最基本的心理测量函数。就该函数而言，随着刺激强度的增加，被检测到实验现象的人的比例增加。沿着曲线对应于 50% 的点表示刺激强度的临界阈值。

## 史蒂文斯幂律

心理物理学中最著名的规律之一是史蒂文斯幂律，它描述了物理刺激强度与其感知强度



之间的关系[311]，认为两者存在指数关系：

$$p=cm^x \quad (2.1)$$

其中：

- $m$  是刺激强度，
- $p$  是感知强度，
- $x$  将刺激强度与感知强度相关联，是等式中最重要的一部分，
- $c$  是一个无关紧要的常数，取决于单位。

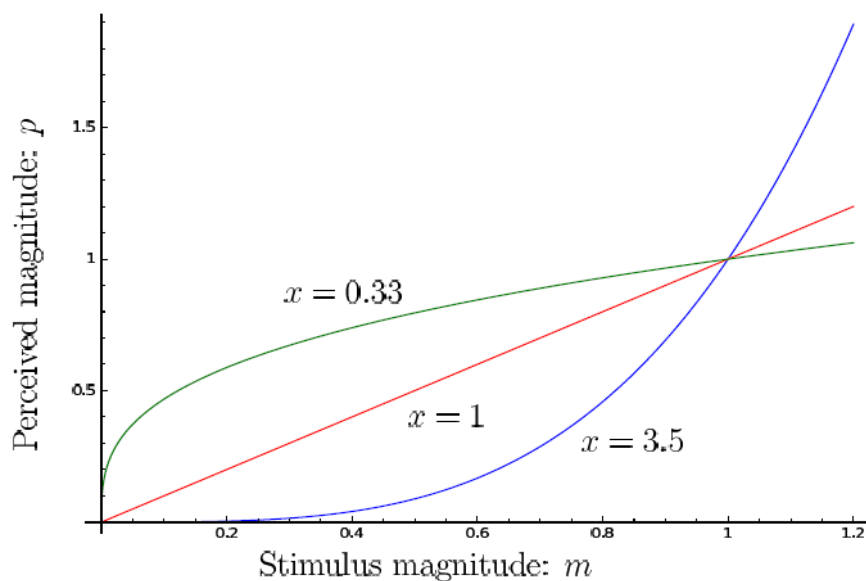


图 2.22 史蒂文斯幂律捕捉到了刺激强度与其感知强度之间的关系。该模型是指数曲线，具体参数取决于刺激类型。

结合图 2.22 来看，当  $x = 1$  时，公式 (2.1) 为线性关系， $p = cm$ 。这方面的一个例子是我们对眼前孤立线段长度的感知，我们感知的长度与其实际长度是成正比的。但更有趣的情况是当  $x \neq 1$  时，对于黑暗中感知目标亮度的情况， $x = 0.33$ ，这意味着感受到的光强度变化比实际的亮度变化更小。另一方面，人们感受通过手指的电流时的  $x = 3.5$ ，因而微弱的电击就可以带来巨大的疼痛感。

### 最小可觉差

另一个重要的心理物理学概念是最小可觉差 (JND)，表示人类或者动物对于某一特定的感官刺激所能察觉的最小改变,在具体实验中，需要至少 50% 的测试者感受到这一最小变化。对于大的刺激变化，几乎所有测试者都会察觉到变化。反之如果变化太小，几乎没有测试者会注意到。因而实验的难点就在于找到这个 50% 变化的测试点。

考虑不同强度刺激下的 JND，例如亮度。那么 JND 本身是如何随着强度的变化而变化的呢？根据韦伯定律得到了如下关系：

$$\frac{\Delta m}{m} = c \quad (2.2)$$

式中  $\Delta m$  代表 JND， $m$  为刺激强度， $c$  为常数。

## 实验设计

VR 破坏了用户通常的感知过程。从本章前述中应该明白一点，人们所提出的 VR 系统和 VR 体验需要从用户的角度进行评估，以了解 VR 技术是否产生了期望的效果，同时根据评估结果去避免不良影响。这需要应用科学方法进行观察，制定假设，并设计验证其有效性的实验。当以人为受试者时，这将变得极具挑战性。有多少测试样本才足够？当他们适应了实验环境之后会发生什么？他们的先前经验会对实验产生怎样的影响？如果他们在测试的那一天身体有恙，该怎么办？他们早餐吃了什么，会有什么影响？这些问题的答案都可能会影响最终的结果。

再深入思考下去情况将变得更为糟糕。假设测试者已经知道了实验的假设，这很可能会影响他们的反馈。另外，实验数据应是什么形式？是要求他们填写一份问卷，还是根据头部运动，心率和皮肤电导等测量数据推断他们在测试时的状态？这些选择也很关键。有关这个主题的更多信息，参见 12.4 节。

## 进一步阅读

本章介绍的软硬件技术正在迅速发展。我们可以随时在互联网上查找到 VR 开发者所设计的最新型的头戴式设备和相关工具。但是，VR 所包含的核心概念到目前为止也几乎没有改变，这也会在接下来的章节中进行介绍。为了更全面地认识人类的感知过程，请参阅[204]和其他相关书籍中“感觉与感知”的部分。