

人类对现实世界的感知是通过各种感官产生的感觉来实现的。VR 技术则利用人类感知的特点,根据特定需求构建具有逼真的视觉、听觉、触觉、味觉以及嗅觉等感官体验的虚拟环境,为用户提供身临其境的感觉。本章主要介绍人的视觉、听觉、触觉、体觉等感知方面的基础知识及其在 VR 系统设计中的一些应用。

3.1 感知原理

个体通过眼、耳、鼻、口及皮肤等感官进行视觉、听觉、嗅觉、味觉及触觉的感知。这些感官中含有各类感受器,例如,视网膜上的光感受器,耳廓内部的听觉刺激感受器,关节与肌肉中用于感知重力和加速度的机械刺激感受器,皮肤中感受外界压力的机械刺激感受器和感知温度的温度感受器,以及鼻腔及舌头上感知化学成分的化学感受器。不同感官中的感受器将来自外部物理世界中的电磁能量、化学成分、组织畸变、气压波动、机械动能、热量变化、重力和加速度等自然刺激转换为神经电信号传递给大脑,相应形成了视觉、听觉、嗅觉、味觉与触觉。另外,人类还能感知自身肢体的姿态,例如,闭上眼睛仍然能够感知四肢和手的位置,这称为本体感觉,是由关节与肌肉中的本体感受器形成的。

VR 则是通过计算机及各种输出设备模拟真实的物理世界,提供人工刺激代替自然刺激欺骗大脑来产生视觉、听觉、触觉、味觉、嗅觉以及体感等感觉。如果构造的虚拟环境足够理想,人类的大脑就可能认为虚拟环境是真实的。

大脑受到感官刺激产生一系列感知体验后,也会进一步根据当前情境提供反馈,并对当前感知进行调整。其中一个典型的例子就是感知适应现象。例如,在黑夜中行走一段时间后,眼睛就会逐步适应暗光环境,能够看清周围的环境。听觉也有适应现象,如声音高于人的听阈 10~15dB 时,会导致听觉不适现象,但离开噪声环境几分钟后,听觉可以完全恢复正常。人体的嗅觉、味觉等感受器也都会产生适应现象。

人的感知适应对 VR 中的用户体验有着较大影响。例如,对于经常在大屏幕前玩第一人称射击游戏的资深用户而言,在体验 VR 环境时会明显感到不够刺激;又如对于新手用户而言,VR 环境可能会导致头晕等不适感,但随着游戏体验时间的增加,其不适感会逐步降低。人体的感知适应是提高虚拟现实真实感、降低不适感的一个重要因素。

在日常生活中,大脑会依据长期的生活经验,对感官刺激形成一套相对固定的神经反应机

制。当感官受到干扰时,固定的神经反应机制就会出现不匹配现象,导致人体出现不适感。例如,戴上度数很深的眼镜时就会出现头晕目眩现象。同样,在 VR 环境中,当人工生成的虚拟刺激与真实刺激出现差异时,就会干扰神经反应机制导致不适感产生。有时用户或许在意识层面没有感知到这种冲突,但身体却会产生疲惫或头痛,甚至出现头晕或恶心的症状。例如,当头戴式显示设备的眼间距与用户实际眼间距差异较大时,虽然用户仍然能够产生立体视觉,但会容易出现头晕现象;或者大脑会清晰地感知到这种冲突,此时将无法产生良好的 VR 体验。

在众多的感觉中,视觉和听觉是最常见、最重要的人体获取外界信息的两个通道,人们通过看和听基本上可以实现对外界环境的感知。除了常见的视觉、听觉定位,其他感官也可协同辅助产生空间感、方向感。如在第 9 章介绍的虚拟飞翔影院系统中,除了基于视听技术构建的虚拟场景外,还提供一个可以模拟飞机驾驶的六自由度动感座椅(具体介绍也参见第 4 章硬件基础知识)。该平台可以给予观众除了视觉和听觉以外的本体感觉,并可通过吹风的方式刺激皮肤。这些感官的协作大大提高了玩家的空间感、方向感,增强了 VR 体验。

目前,在 VR 中比较成熟的技术应用也主要是视听技术,用于构建具有真实感、沉浸感的虚拟环境。在构建虚拟环境时,需要充分考虑人的视听感官的生理机制和 VR 体验特征。因此,本章下面重点介绍与 VR 相关的视觉和听觉知识,并简单介绍触觉和力觉、本体与内部感觉等基本的体觉知识。



3.2 视觉

视觉是人类最重要的感觉通道,人类从周围世界获取的信息约有 80% 是通过视觉得到的。满足视觉感知是构建真实感、沉浸感的虚拟环境的首要任务。在建模环节中要保证虚拟对象模型的形态与动作具有逼真的效果,同时还要保证显示设备能够实现高清实时显示。因此,本节将从视觉的生理机制、眼球运动、深度感知、立体视觉以及颜色感知等方面介绍与 VR 相关的视觉感知知识。

3.2.1 视觉的生理机制

视觉活动始于光。眼睛接收光线,转换为电信号,再传递给大脑,形成对外部世界的感知。眼睛的生理构造如图 3.1 所示。当光线穿过角膜时,它们会经过含有房水的小室(前房),之后光线经过瞳孔进入晶状体。瞳孔的大小由称为虹膜的盘状结构控制,用来调节允许通过的光量。睫状肌可以改变晶状体的光焦度。穿过晶状体后,光线通过玻璃体射向内层的视网膜。视网膜包含两种光感受器:视锥细胞和视杆细胞。视锥细胞只有在光线明亮的情况下才起作用,具有辨别光波波长的能力,因此对颜色十分敏感,特别对光谱中黄色部分最敏感,在视网膜中部最多。而视杆细胞比视锥细胞灵敏度高,在暗的光线下就能起作用,没有辨别颜色的能力。视网膜的中央是黄斑区,负责视觉和色觉的视锥细胞就分布于该区域,黄斑中央的凹陷称为中央凹,视锥细胞在此分布最为集中,是视力最敏锐的地方。

视网膜上不仅分布大量的视细胞,同时还存在一个盲点,这是视神经进入眼睛的入口。盲点中没有锥状体和杆状体,在视觉系统的自我调节下,人们无法察觉盲点。视网膜上还有一种特殊的神经细胞,即视神经中枢,可以帮助人们察觉运动和形式上的变化。

视觉感知可分为两个阶段:接收信息阶段和解释信息阶段。需要注意的是:一方面,视

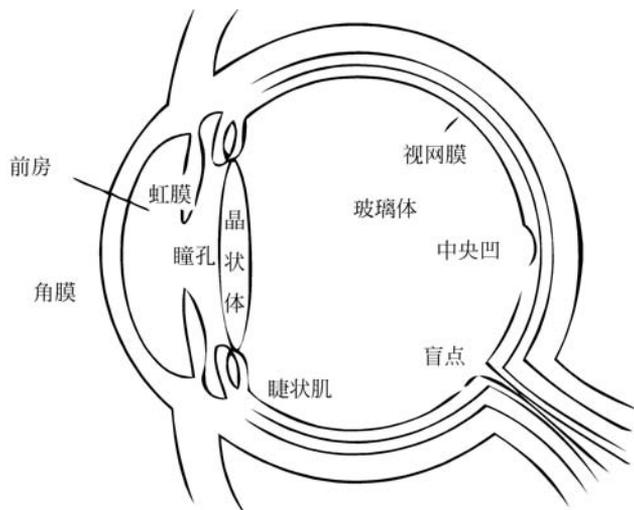


图 3.1 眼睛的结构图

觉系统的物理特性决定了人类无法看到某些事物；另一方面，视觉系统解释处理信息时可对不完全信息发挥一定的想象力。因此，进行 VR 设计时需要清楚这两个阶段及其影响，了解人类真正能够看到的信息。

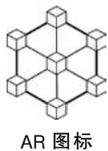
3.2.2 人眼的视觉暂留现象

现代影视主要基于视觉暂留现象。这一概念于 1824 年由英国伦敦大学教授皮特·马克·罗葛特在研究报告《移动物体的视觉暂留现象》中最先提出。在此基础上，法国卢米埃尔兄弟通过不断实验探索，于 1895 年拍摄和放映了人类第一部电影《火车进站》，开启了现代影视技术的新时代。影像内容的刷新率超过 10 帧就可以产生视觉暂留，达到 20 帧左右时就可以产生较好的动态视觉感受。在很长一段时间中，电影的图像刷新率都是以 24 帧为标准的。

因此，进行 VR 开发还需要了解各类显示设备的图像刷新率。目前常见的液晶显示器和投影仪的标准图像刷新率为 60 帧，而当进行 3D 立体显示时，图像刷新率会提高到 120 帧。显示器和投影仪需要按顺序交替显示左眼画面和右眼画面，同时配合液晶快门眼镜的左右眼镜片与画面播放次序同步开闭，以产生立体视觉感官，此时对于单眼而言画面刷新率仍是 60 帧。目前，松下等投影仪厂家已经推出 240 帧刷新率的投影仪设备，我们看到数字化显示设备的图像刷新率都是以 60 帧为基础并进行倍增的，这是欧美地区交流电 60Hz 交变频率而约定的。VR 显示设备的图像刷新率降低就会在视觉上造成画面卡顿。一般来说，当显示设备图像刷新率低于 60 帧时，用户视觉上就会感到轻微的卡顿；刷新率降至 30 帧左右时，画面的卡顿体验就比较明显。

当观看国产经典老电影时，即使刷新率只有 24 帧，画面却仍然流畅，这是由于电影画面与游戏及 VR 视频画面的生成机制不同而导致的。电影画面是通过摄像机进行现场实拍产生，在一帧画面的 $1/24s$ 时间内，胶片或 CCD 器件不断接收和记录实际场景中的光信号，这是一个光信号累加积分的过程，因此实拍电影画面中的运动物体含有“拖影”（见图 3.2(a)），这种拖影虽然细微，但在连续播放时人眼是可以捕捉和感知到的，这符合人眼观察实际环境的特性。例如，当在眼前快速摆动手指时，就会看到手指的拖影。实拍电影由于记录下了运动物体

的拖影,因此人眼对运动物体画面的感受就比较舒适和流畅。而计算机游戏和 VR 视频是通过虚拟场景模型进行实时渲染生成的,是一个瞬时采样拍摄过程,因而画面中的运动物体就没有拖影(见图 3.2(b)),导致人眼感受到不适感与卡顿感,因而需要靠较高的图像刷新率才能加以弥补。由于上述原因,在游戏和 VR 制作中,对高速移动的物体,例如赛车模型渲染时,会通过 Shader 编程产生拖影特效,但这种人为生成的运动拖影,仍无法真实和自然地再现实际环境中人眼产生的拖影,过分使用拖影特效有时会适得其反。



AR 图标



图 3.2 实拍电影中含有轻微拖影,实时渲染生成的 VR 视频中没有拖影

在观看立体电影时,有时需要佩戴主动式快门 3D 眼镜。主动快门式 3D 技术,英文为 Active Shutter 3D,配合主动式快门 3D 眼镜使用。主动快门式 3D 技术主要是通过提高画面的刷新率来实现 3D 效果的,通过把图像按帧一分为二,形成对应左眼和右眼的两组画面,连续交替显示出来,同时红外信号发射器将同步控制快门式 3D 眼镜的左右镜片开关,使左、右双眼能够在正确的时刻看到相应画面。因人的眼睛具有视觉暂留的特性,让人眼感觉左右眼不同的画面是同时播放的,从而在大脑中产生所观看画面的立体成像,以达到立体投影效果。屏幕刷新频率必须达到 120Hz 以上,也就是让左、右眼均接收到频率在 60Hz 以上的图像,才能保证用户看到连续而不闪烁的 3D 图像效果。

3.2.3 眼球运动

为了准确注视目标物体,人的眼球会发生运动来进行调节。通过快速扫视将感兴趣的特征定位在中央凹上,人眼能够准确地瞄准目标。快速扫视是眼睛的快速移动,能够快速使中心凹重新定位,以最高的视敏度来感知场景中的重要特征。人眼也可通过平稳移动来减少视网膜上的运动模糊,保持图像稳定。前庭眼动反射(Vestibulo-Ocular Reflex)也可提供图像稳定功能。人类还可通过聚散(Vergence,也称为辐辏)这种眼睛运动保持立体视觉并防止适应恒定刺激。立体视觉是指双眼被固定在相同物体上,从而产生单个感知图像。人眼可出现两种聚散运动将眼睛与物体对齐(见图 3.3):如果物体比先前的位置更近,则发生会聚运动,两个瞳孔会越来越靠近;如果物体更远,则发生发散运动,两个瞳孔会逐渐分开。聚散运动产生的眼睛方向提供了有关物体距离的重要信息。

另外,大部分时间眼球和头部是在一起移动的。眼睛可以向左或向右偏转 35° ,且这两种偏转是对称的。而眼睛的俯仰则不对称——人眼可向上倾斜 20° ,而向下则是倾斜 25° 。这表明当眼睛直视前方时,将 VR 显示器置于瞳孔下方的中心位置可能是最佳选择。在前庭眼动反射的情况下,控制眼球旋转以抵消头部运动。在平稳移动的情况下,头部和眼睛可一起移动将移动目标保持在首选观看区域中。

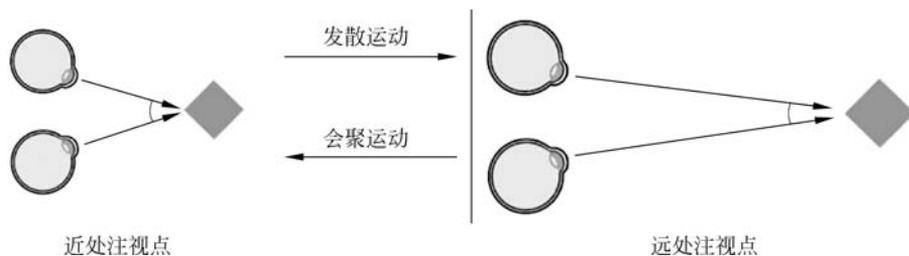


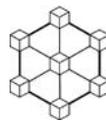
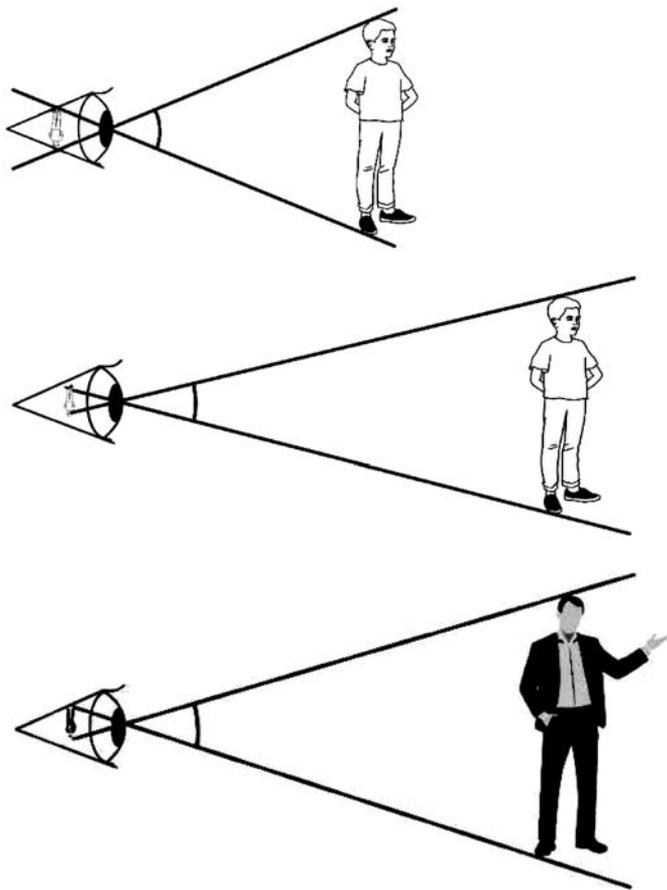
图 3.3 人眼的两种聚散运动：发散运动和会聚运动



AR 图标

3.2.4 深度感知与立体视觉

人类通过视觉可以感知物体大小、深度和相对距离,在大脑中形成空间深度感。要了解人的眼睛如何感知物体大小、深度和相对距离,首先需要了解物体是如何在眼睛的视网膜上成像的。物体反射的光线在视网膜上形成一个倒像,像的大小和视角有关。视角反映了物体占据人眼视域空间的大小,视角的大小与物体离眼睛的距离、物体的大小这两个要素有着密切的关系:两个同样大小的物体被放在离眼睛不一样远的地方,离眼睛较远者会形成较小的视角;两个与眼睛距离一样远的物体,大者会形成较大视角(见图 3.4)。

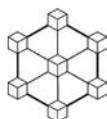


AR 图标

图 3.4 人的视角

人眼可以借用很多线索实现深度和相对距离的感知,其中包括单眼线索和双眼线索。人的视觉景象中有很多线索让人可以仅凭单眼便能感知物体的深度和相对距离。常用的单眼线索如下。

- (1) 线条透视线索: 通过地平线和平行线条的消失点可判断远近关系,如图 3.5(a)所示。
- (2) 纹理梯度线索: 可以通过纹理的变化判断深度关系,如图 3.5(b)所示。
- (3) 遮挡线索: 可通过遮挡关系确定远近。如果两个物体重叠,那么被部分覆盖的物体被看作是在背景处,自然离得比较远,如图 3.5(c)所示。
- (4) 大小线索: 人们平时所熟悉物体的大小和高度为人们判断物体的深度提供了一个重要线索。一个人如果非常熟悉一个物体,他对物体的大小在头脑中事先有一个期望和预测,就会在判断物体距离时很容易和他看到的物体的大小联系起来,如图 3.5(d)所示。
- (5) 阴影线索: 光源遇到物体时投下的阴影提供了一个重要线索,如图 3.5(e)所示,阴影解决了球中模糊的深度和阴影错觉问题。
- (6) 图像模糊线索: 根据焦点的变化的锐度来推断深度。如图 3.5(f)所示,由于图像模糊,一个对象似乎比另一个要近得多。
- (7) 大气线索: 如空气湿度导致远处的风景具有较低的对比度,从而看起来更远。如图 3.5(g)所示显示这个场景提供了一个大气信号: 一些景物被认为更远一些,因为它的对比度较低。
- (8) 运动视差线索: 通过感知的物体运动速度判断远近关系。例如,坐在火车上,远处的物体运动速度慢,近处的物体运动速度快。



AR 图标

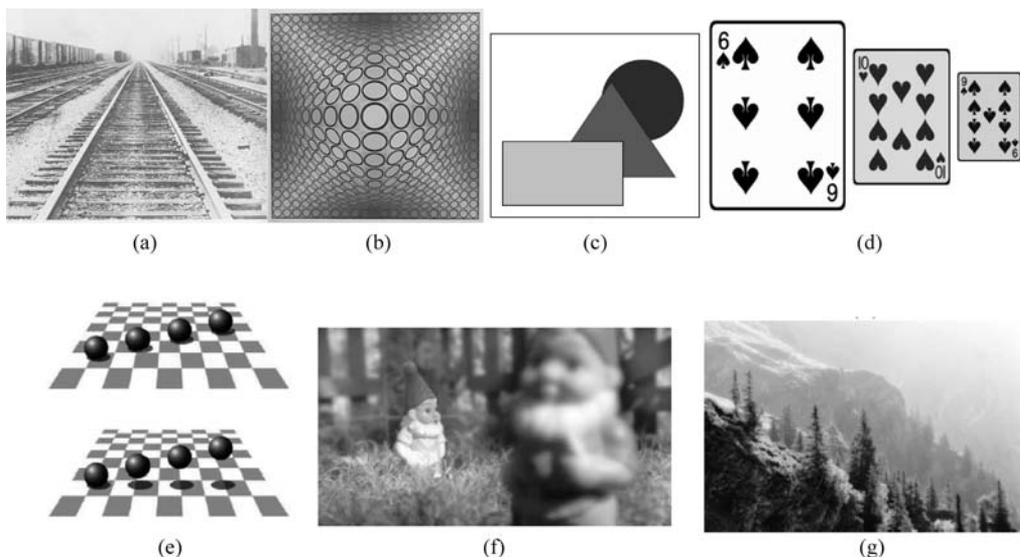


图 3.5 深度感知线索

将两只眼睛聚焦在同一个物体上会提高深度感知。由于人类是用两只眼睛同时观看,人的左、右眼之间有一定的间距(大约 6.5cm),所以在看同一物体时左眼图像与右眼图像会有细微差异,称为双眼视差(见图 3.6)。大脑会自动将两幅图像合二为一,产生对物体的立体及空间观感。双眼视差是一种双眼深度线索,大脑可以利用对这种视差的测量,估计出物体到眼睛的距离。3D 电影的制作原理便基于双眼视差原理,即通过立体摄像机获取具有差异的左、右

两组图像,然后通过相应的 3D 显示技术分别播放给人的左右眼,从而在人脑中呈现虚拟的立体场景。在 VR 系统中,则是通过设置左、右并列的两台虚拟摄像机拍摄渲染虚拟场景图像,形成左、右两路视频,对应输出到用户的左、右眼中,使用户产生真实的立体视觉感受。

VR 中不当的刺激呈现可能影响深度感知。例如,有些 VR 系统常常把视点高度设置成固定值,但是对于更高或更矮的用户来说,对于虚拟场景的感知就会出现空间感知的不适。另外,如果 VR 中瞳孔间距的设置与实际情况不匹配,还会出现额外的并发症。例如,真人用户的实际眼间距为 6.4cm,而在 VR 中扮演的虚拟角色眼间距被设置为 5cm,那么虚拟世界就会看起来大得多,这会严重影响深度感知。同样,如果虚拟角色的眼间距比真人用户大,虚拟世界就会比实际尺寸看起来小。

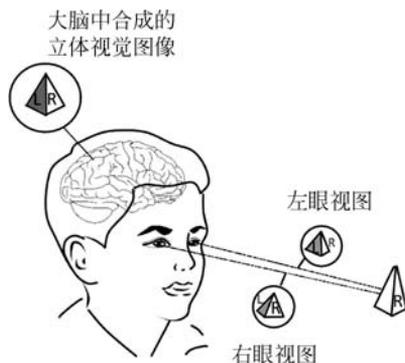


图 3.6 双目深度感知

3.2.5 颜色感知

人能感觉到不同的颜色,是眼睛接受不同波长的光的结果。颜色通常用三种属性表示:色度、饱和度、明度或亮度。色度是由光的波长决定的,正常的眼睛可感受到的光谱波长为 $400\sim 700\mu\text{m}$ 。视网膜对不同波长的光敏感度不同,同样强度的光而颜色不同,有的看起来会亮一些,有的看起来会暗一些。对于人眼来说,位于可见光谱中央位置处的绿色光($550\mu\text{m}$)最为明亮,而位于可见光谱两端的红色光($400\mu\text{m}$)和紫色光($700\mu\text{m}$)就比较暗。

在三维空间中,可以用一个圆锥体表示颜色的三种基本属性。该模型称为 HSV 颜色模型,H、S、V 分别代表色度、饱和度和明度。该模型对应于圆柱坐标系的一个圆锥形子集(见图 3.7)。圆锥的顶面对应于 $V=1$,代表的颜色最亮。色彩 H 由绕 V 轴的旋转角给定,红色对应角度为 0° ,绿色对应角度为 120° ,蓝色对应角度为 240° 。在 HSV 颜色模型中,每一种颜色和它的补色相差 180° 。饱和度 S 取值 $0\sim 1$,由圆心向圆周过渡。在圆锥的顶点处, $V=0$, H 和 S 无定义,代表黑色;圆锥顶面中心处 $S=0$, $V=1$, H 无定义,代表白色;从该点到原点代表亮度渐暗的白色,即不同灰度的白色。任何 $V=1$, $S=1$ 的颜色都是纯色。

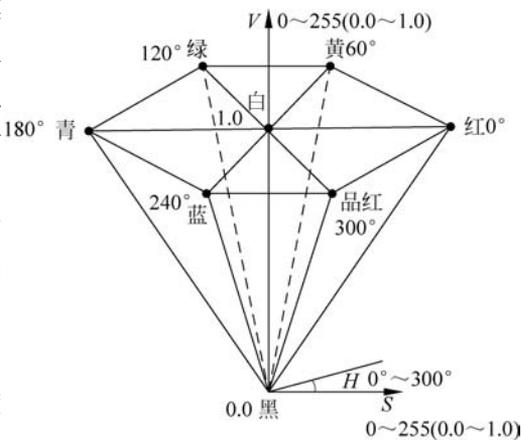
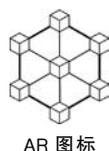


图 3.7 HSV 颜色图集

由于具有不同光谱分布的光产生的颜色有可能是一样的,所以需要采用其他颜色模型定义颜色,使光与颜色可以保持一一对应。RGB 颜色模型是 VR 中常用的一种颜色模型。该模型基于三基色学说,以红、绿、蓝为原色,各个原色混合在一起可以产生复合色,如图 3.8 所示。RGB 颜色模型通常采用如图 3.9 所示的单位立方体来表示,在正方体的主对角线上,各原色的强度相等,产生由暗到明的白色,也就是不同的灰度值。其中, $(0,0,0)$ 为黑色, $(1,1,1)$ 为白色。正方体的其他六个顶点分别为红、黄、绿、青、蓝和品红。



AR 图标

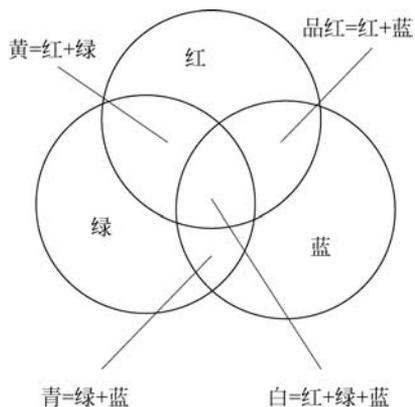


图 3.8 RGB 三原色混合效果

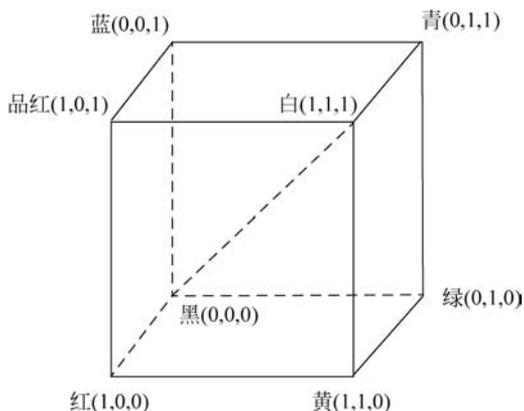


图 3.9 RGB 颜色模型

由于 RGB 颜色模型是以红、绿、蓝为原色,将各个原色混合在一起产生复合色,因此称为加色法颜色系统。该系统属于发射光原理,如头盔显示器采用的就是这种系统。

与加色法颜色系统相对应的是减色法颜色系统。如我们看到的物体的颜色,是由于物体表面上的颜料,吸收了日光中一部分的光波,反射日光其他的色光到人眼所看到的。该系统属于反射光原理,常运用在颜料的混合中,也广泛地运用在印刷技术之中。CMYK 颜色模型就是一种减色法颜色系统,主要以红、绿、蓝的补色青(Cyan)、品红(Magenta)、黄(Yellow)为原色(见图 3.10)。在实际印刷中,一般采用青(C)、品红(M)、黄(Y)、黑(BK)进行四色印刷,在印刷的中间调至暗调增加黑版。如打印机采用的就是 CMYK 颜色模型。

CMYK 颜色模型对应的直角坐标系的子空间与 RGB 颜色模型所对应的子空间几乎完全相同(见图 3.11)。差别仅在于前者的原点为白,而后者的原点为黑。前者是定义在白色中减去某种颜色来定义一种颜色,而后者是通过从黑色中加入颜色来定义一种颜色。



AR 图标

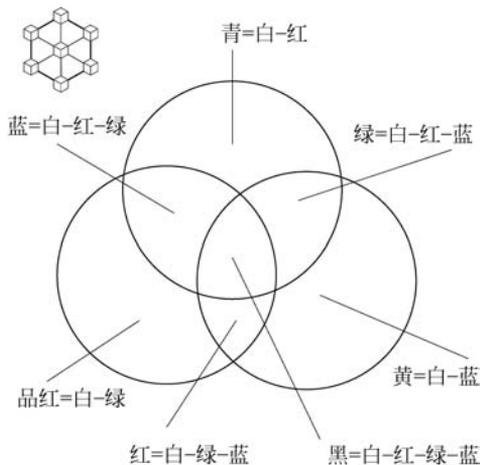


图 3.10 CMYK 原色的减色效果

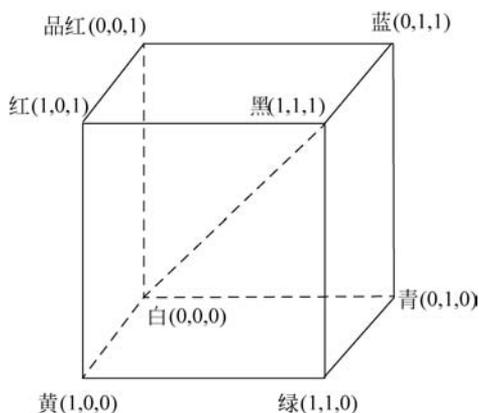


图 3.11 CMYK 颜色模型

在投影式 VR 系统中,DLP 投影仪主要通过色轮过滤的方式将彩色图像分解为红、绿、蓝三原色分别投影。色轮就是一片圆形镀膜玻璃片,分为红、绿、蓝三种色段,色轮在电机驱动下高速旋转,对白色光源进行分色。LCD 投影仪中安装有三片高温多晶硅液晶面板,而投影仪的白色光源也通过三片分色镜,被分成红、绿、蓝三束光线,分别照射对应的液晶面板,产生出

相应的红、绿、蓝单色位面图像,随后通过一体化光学棱镜,将红、绿、蓝单色位面图像合成为彩色图像,由光学镜头聚焦投影到屏幕上(见第4章)。另外,红蓝滤色眼镜等偏光式眼镜通过不同颜色的滤镜分离双眼视差图像,形成立体视觉。

3.3 听觉

人类通过听觉对客观世界产生的感知信息仅次于视觉。在 VR 系统中,听觉的作用一方面为虚拟场景伴音,通过有效的视听融合提供真实感体验,另一方面则可以提供空间方位信息。本节主要介绍听觉的生理机制和听觉定位的基本知识,及其在 VR 系统设计中的一些应用。

3.3.1 听觉的生理机制

听觉所涉及的问题和视觉类似,即耳朵接受刺激,把刺激信号转换为神经兴奋,并对信息进行加工,然后传递到大脑。耳朵是听觉的外周感觉器官,由三部分组成:外耳、中耳和内耳。外耳包括耳廓和外耳道两部分;中耳则是一个小腔,通过耳膜与外耳相连;内耳由前庭器官和耳蜗组成(见图 3.12)。

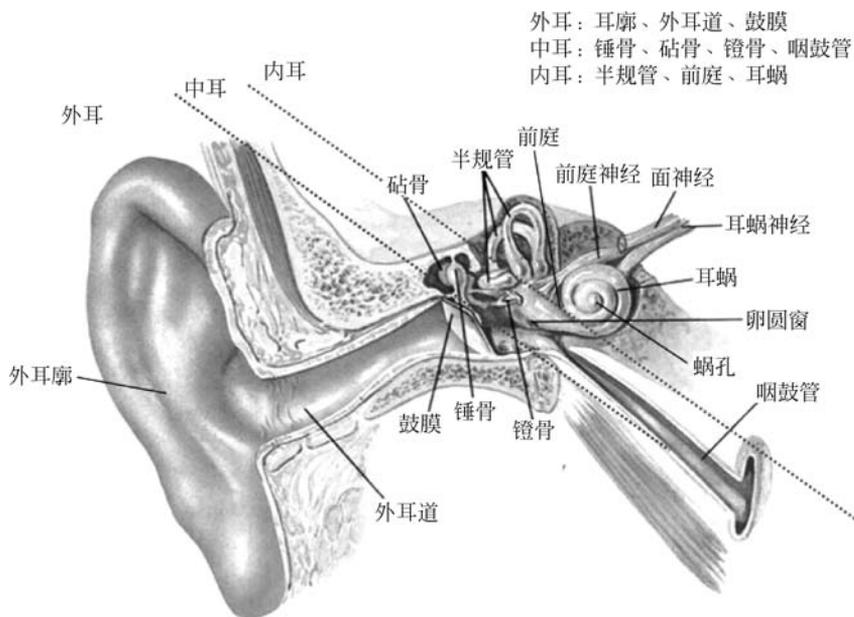


图 3.12 耳朵的生理结构图

声音是由物体振动产生的声波。当声波由空气到达外耳廓时,耳廓像喇叭一样的结构会初步将这些声波进行聚拢,以使声波信号足够强烈,然后声波就会顺着耳孔进入耳道,最终到达耳膜。耳膜是一层薄薄的弹性组织,它非常灵敏,只要稍微有点儿声波振动,就能够随之产生振动,并带动与它相连的听小骨等结构。听小骨由锤骨、砧骨及镫骨组成,可以将声波信号进行放大。最终这些放大的振动信号就会到达耳蜗。声波进入充满淋巴液的耳蜗,通过耳蜗内大量纤毛的弯曲来刺激听觉神经。神经脉冲从左耳蜗和右耳蜗一直传递到大脑中的主要听觉皮层。当信号通过神经结构组合时,发生分层处理,这使得大脑能够分析多个频率和相移。一种被称为上橄榄的早期结构接收来自双耳的信号,可以分辨出声波的振幅和相位差异。这

对于确定声源的位置是非常重要的。

人类听到的声音可由音调、响度和音色三个主要特征来描述。

音调主要是由声波频率决定的听觉特性。声波频率不同,人们听到的音调高低也不同。人的听觉的频率范围为 $16\sim 20\,000\text{Hz}$ 。其中, $1000\sim 4000\text{Hz}$ 是人耳最敏感的区域。 16Hz 是人的音调的下限, $20\,000\text{Hz}$ 是人的音调的上限。当声波的频率约为 1000Hz 、响度超过 40dB 时,人耳能分辨出 0.3% 的频率变化范围。但同时音调也是一种心理量,它和实际声波频率的变化不完全对应。在 1000Hz 以上,频率与音调的关系几乎是线性的;在 1000Hz 以下,频率与音调的关系不是线性的。

音响是由声音强度决定的一种听觉特性。强度大,听起来响度高;强度小,听起来响度低。对人来说,音响的下限为 0dB ,它的物理强度为 $2\times 10^{-9}\text{N/cm}^2$ 。上限约 130dB ,上限的物理强度约为下限物理强度的100万倍。音响还和声音频率有关,在相同的声压水平上,不同频率的声音响度是不同的,而不同的声压水平却可产生同样的音响。

音色是指不同的声音的频率表现在波形方面总是有与众不同的特性,不同的物体振动有不同的特点。不同的发声体由于其材料、结构不同,其发出声音的音色也不同。

3.3.2 虚拟环境中的听觉内容设计

在人类的各类感知中,虽然视觉大约占比 80% ,但如果没有听觉,人类就难以正确感知现实世界。听觉画面与视觉画面一起在人的脑海中形成了对自然景观及社会景观的感知。加拿大声学家与作曲家谢弗(R. Murray Schafer)于1977年出版的《声景:我们的声音环境和世界的调谐》(*Soundscape: Our Sonic Environment and the Tuning of the World*)一书中提出了声景(Soundscape)的概念,并逐步发展研究听觉、声环境与社会之间相互关系的学术领域。谢弗将声景中的声音划分为基调声(Keynote Sounds)、信号声(Signals)和标志声(Sound Marks)三类。虚拟现实环境中的听觉内容也可以按照这种分类进行相应设计。

基调声指现实世界中的背景环境声。背景环境声提供了听觉感知中的背景基本信息。例如,旷野背景中有风声,高大的厅堂环境中隐隐含有混响声,机房环境中会有机器嗡嗡声等。背景基调声不能被简单地归于背景噪声。基调声能够间接传达空间大小、建筑材质、气候气象等基本背景信息,能够使听者对所处环境产生整体性的感知。在现实环境中,人们不会刻意注意基调声,甚至会忽视基调声。但在VR环境中,如果没有合理有效的基调声,就会大大削弱用户的沉浸感与真实感,例如,在虚拟海底漫游环境中需要添加汩汩的气泡声以加强用户对环境真实感的感知。

信号声则是能够引起人们刻意倾听的前景声,是人与环境的一种重要的交互信息。例如,车辆向你驶来时的轰鸣声,使用工具时的叮当声,推门时发出的吱扭声等。在VR中,信号声有两个重要作用:一是增强交互的真实感。当用户触碰一个虚拟气球模型时,不仅要提供视觉与触觉反馈,还要配以薄膜的摩擦声音反馈才会使虚拟体验更为真实。二是引导视线作用。如果希望用户在虚拟环境中注意某个对象的出现,或吸引用户按预设路线漫游时,就可以使用信号声进行导引。虽然物体或空间处于用户视线之外,但逼真的声效同样可以吸引用户的注意。例如,在虚拟海底漫游环境中,如果希望用户注意背后游来的鱼,就可以通过鱼儿划水声的远近与方位变化引导用户视线。在这种情况下,需要开发者注意对声源移动的仿真,并对用户头部转动所引起的听觉变化进行实时响应,以再现真实感的听觉效果。

现实环境中的标志声是指“标志性的、有一定的文化含义的、独一无二的声音”。例如,听

到如诉的马头琴声,脑海中就想象出风吹草低的蒙古草原等。相应地,标志声在VR环境中的一个重要作用就是激发用户自身的想象,有效提高虚拟环境的代入感。同时,在用户进行或完成某项虚拟交互时,也可以人为放大或创造某些特殊的标志音效来增强交互体验感。例如,在虚拟海底漫游环境中,如果用户用手柄开启刚发现的一个海底宝箱时,就可以对开启宝箱的标志音进行夸张放大。虽然实际水下环境中开箱的声音几乎是听不到的,但这样可以加深用户的交互体验,激发用户继续漫游的兴趣。

在VR环境开发时,听觉内容的设计与制作要综合考虑基调声、信号声及标志声的运用与传达,三类声音的综合运用可以有效提高用户VR体验的真实感与沉浸感。反之,如果只使用单一声音(如缺乏经验的设计者常使用单一背景环境声)来搭建整个虚拟声景,不但不能辅助提升VR内容呈现质量,反而会产生负面效果。

另外,声音的传播需要空气、水和固体等介质。声音在不同的介质中传播的速度是不同的,且声音的传播速度随物质的坚韧性的增大而增加,随物质的密度的减小而减少。声音在空气中的传播速度也与压强、温度和阻力有关,还会因外界物质的阻挡而发生折射,产生回声等。因此,高度真实感的声音仿真仍需进一步研究探索。

在VR系统中,还要做到声音与视觉画面的同步融合。一个VR系统,当听觉与视觉能良好融合时,有助于提供令人沉浸、舒适的VR体验。例如,在虚拟射击影院中,逼真的射击声和射击特效画面的同步将增强射击的体验。如果VR中的场景和声音不适配,则会导致眼睛和耳朵的失调,进而引起在VR中观看时的疲劳感。

3.3.3 听觉定位与立体声

与视觉一样,人类通过听觉也能产生空间定位,即可通过听到声音来估计声源的位置。例如,如果人们正在与虚拟代理对话,那么不同代理的声音应该来自相应的位置。这对许多VR体验也是至关重要的。

人耳能够判断声音的位置和方位。人脑识别声源的位置和方向,一方面是利用了两耳听到的声音的混响时间差和混响强度差。混响时间差是两耳感受同一声源在时间先后上的不同,混响强度差则表示两耳感受同一声源在响度上的不同。另一方面,人耳听觉系统对声源的定位还与身体结构有关。声音在进入人耳之前会在听者的面部、肩部和外耳廓上发生散射,这就使得音源的声音频谱与人耳听到的声音频谱产生差异,而且两只耳朵听到的声音频谱也存在差异。这种差异可以通过测量声源的频谱和人耳鼓膜处的频谱获得。通过频谱差异的分析,就可以得出声音在进入内耳之前在人体头部区域的变化规律,即为“头部相关传递函数”(Head-Related Transfer Functions, HRTF)。利用该函数对虚拟场景中的声音进行处理后,那么即使用户使用耳机收听,也能感觉到三维空间中的声音立体感和真实性。

听觉定位线索可以分为使用单耳和双耳两种。单声道线索依赖于到达单个耳朵的声音来约束可能的声源集合。也就是说,人仅凭一只耳朵,就能初步辨出声音的方向,从而对声音进行定位。单声道线索主要包括:

(1) 耳廓形状不对称产生的声音扭曲。这取决于声音到达的方向,尤其是仰角。虽然人们没有意识到这种扭曲,但听觉系统可用它来定位。

(2) 声音的波幅与距离的关系。对于熟悉的声音,感知到的声音波幅(响度)可帮助估计距离。熟悉度越高越有助于对距离的判断。

(3) 频谱失真。声音的高频分量比低频分量衰减得更快,因此远距离声音会发生频谱失真。

(4) 声音混响。进入耳朵的混响可以提供强大的单声道线索,这种线索被称为回声定位,在室内环境中会更加明显。

如果双耳都参与声源感知定位,由于位于头部两侧的两耳之间有一定的距离,因此声源发出的声音到达两耳传播路径长短不一,导致声音的延时和衰减就不同,声音到达的时间和强度也就不同,从而可以产生听觉的位置感空间感(见图 3.13)。最简单的情况是耳间水平差,它是每只耳朵听到的声音幅度的差异。例如,一个耳朵可能面向声源,而另一个耳朵则背向声源,较近的耳朵就会比另一只耳朵受到更强烈的振动(见图 3.13(a))。另一个双耳线索是耳间时差。两只耳朵之间的距离约为 21.5cm,这导致来自声源的声音到达时间不同(见图 3.13(b))。由于声音在约 0.6ms 内传播 21.5cm,因此很小的时差就可以用于定位。另外,头部运动不仅可以提供视觉定位线索,也可以提供听觉定位线索,这是因为近的音频源比远的音频源会更快地改变其方位角和仰角。

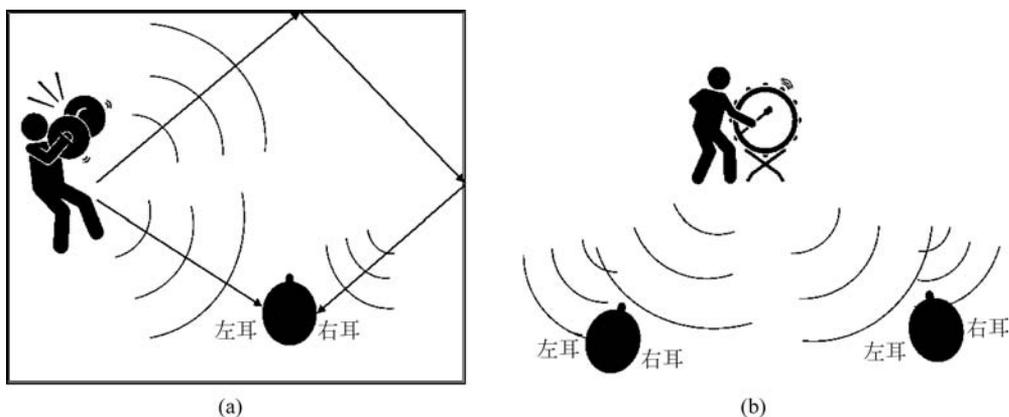


图 3.13 通过双耳对声源感知定位

为了配合人类的双耳听觉感知,音频产品常采用立体声模式,通过双通道录音并配合耳机进行双通道播放,使用户双耳听到的声音产生差异,由此产生出声音的立体感与层次感。但在 VR 环境中,单纯的立体声播放无法与视觉配合产生真实感与沉浸感。如图 3.14 所示,当用户佩戴头戴式设备观看乐团演出的 VR 内容时,无论用户如何转动头部,耳机中左声道和右声道的声音信号都不会发生改变。通过听觉感知定位,用户感知鼓手永远在前、提琴手永远在后、竖琴手永远在左、号手永远在右,即乐团成员是随着头部转动而转动的,这与头戴式设备中看到的 VR 画面效果产生矛盾,降低了沉浸感与真实感。

为在 VR 环境中产生真实的听觉感知效果,需要使用立体音频处理与播放模式,如图 3.15 所示。通过传感设备实时感知用户头部的转动方位数据,据此对左、右耳声道中不同声源的声音信号进行实时调整,实时模拟出用户当前位置所应听到的声音内容,使耳机中的听觉感知与头戴式设备产生的视觉感知相匹配,保证视听同步感知,这样能够有效提高用户 VR 体验的真实感与沉浸感。

对于大空间投影式 VR 系统,用户无须佩戴耳机,而是在场地中选用环绕声播放,可以允许多人同时共场进行 VR 体验。这类方式需要借助一定的专业知识在空间中布置若干音箱,以形成一个环绕式播放系统(见图 3.16),来模拟真实世界中包围式的声景环境。当用户在场地中自由行走时,双耳听到的声音会随着与各音箱之间的相对距离方位产生细微变化,产生出高度真实感的听觉效果。

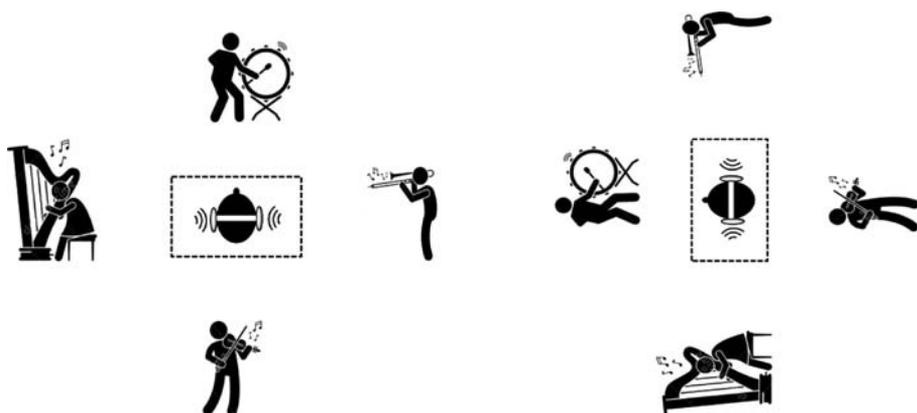


图 3.14 单纯的立体声播放无法满足虚拟现实环境要求

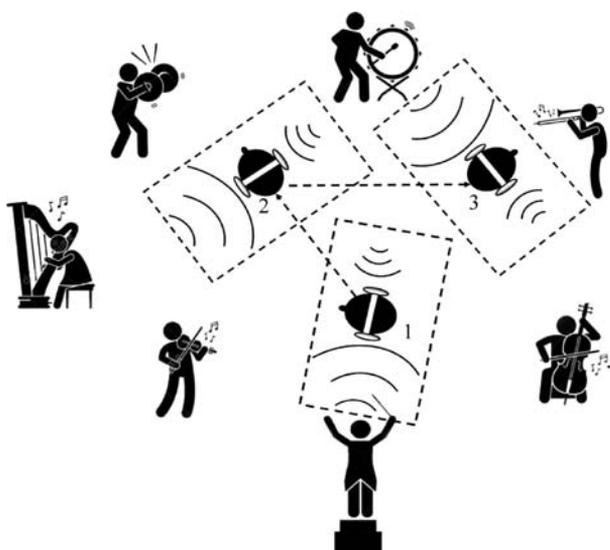


图 3.15 配合头戴式设备的立体音频播放能够产生逼真的听觉感知



图 3.16 在基于大空间投影的 VR 环境中采用多音箱构成环绕式播放系统



3.4 体觉

视觉和听觉是 VR 系统设计的重点,但触觉、力觉和本体感觉等体觉系统对提供良好的 VR 体验也有重要作用。它们也可以反馈交互环境中的许多关键信息,帮助用户产生更身临其境的体验。

3.4.1 触觉和力觉

虽然比起视觉和听觉,触觉的作用要弱些,但触觉在交互中的作用是不可低估的。如通过触摸感觉东西的冷或热可以作为进一步动作的预警信号,人们通过触觉反馈可以使动作更加精确和敏捷。

触觉的感知机理与视觉和听觉的最大不同在于它的非局部性。人的全身皮肤布满了各种触觉和力觉感受器,用以感知触觉和力觉的刺激。皮肤中包含三类感受器:温度感受器(Thermoreceptors)、伤害感受器(Nociceptors)和机械刺激感受器(Mechanoreceptors)。它们分别用来感受冷热、疼痛和压力。机械刺激感受器分为快速适应机械刺激感受器(Rapidly Adapting Mechanoreceptors)与慢速适应机械刺激感受器(Slowly Adapting Mechanoreceptors)。前者可以感受瞬间的压力,而受到持续压力时不再有反应。后者则对持续压力比较敏感,用来形成人对持续压力的感觉。

触觉感受器主要包括以下几种。

(1) 游离末梢神经(Free Nerve Endings): 主要分布在表皮层,功能是感受外界温度(热和冷),以及皮肤损伤的疼痛。

(2) 鲁菲尼小体(Ruffini's Endings): 位于真皮层,是一种慢适应感受器,也称为鲁菲尼末梢。

(3) 帕西尼小体(Pacinian Corpuscles): 广泛分布于皮下神经末梢,响应深压触动和高频振动。

(4) 默克尔小体(Merkel's Disks): 位于表皮之下,并对静态压力(随时间变化不大或没有变化)做出反应,具有缓慢的时间响应。

(5) 麦斯纳小体(Meissner's Corpuscles): 位于表皮之下,对较轻的触觉做出反应,反应速度比默克尔和鲁菲尼小体快,感受压力频率不如帕西尼小体。

(6) 毛囊感受器(Hair Follicle Receptors): 位于发根周围;对轻微的触觉做出反应,例如脱毛也会感知疼痛。

体感系统的神经通路类似于视觉途径的工作方式。信号通过丘脑传送,相关信息最终到达大脑的体感皮层,在那里进行更高级的处理。早在丘脑之前,一些信号也通过脊髓传递到控制肌肉的运动神经元。这使得运动响应十分迅速,可以快速地从疼痛刺激中退出。在主体感皮层内,神经元在空间上的排列对应于它们在身体上的位置。一些神经元也有与皮肤上的局部斑片相对应的感受域,这与接受性视觉的视觉效果十分相似。同样地,横向抑制和空间对立存在并形成检测器,允许人们估测沿着皮肤表面的尖锐压力特征。

触觉具有空间和时间分辨率。人体对触觉的感知敏感度取决于人体组织中的感受器分布密度。实验表明,人体的各个部位对触觉的敏感程度是不同的:人体在指尖和舌头处机械刺激感受器的密度最大,可以区分 2~3mm 范围的点;头部密度低一些,敏感度约为 20mm;在

背部很低,敏感度约为 60mm。对人身各部位触觉敏感度的了解有助于更好地设计 VR 触觉交互设备。VR 系统就是通过各种手段来刺激人体表面的神经末梢,从而使用户达到身临其境的接触感。

通过手指触摸可以感知物体表面的纹理。对于粗糙纹理的感知主要通过空间线索,即通过手指对表面施压来感知纹理结构。对于细小纹理则主要是通过时间线索来感知的,手指滑过表面时导致压力振动,可以由帕西尼和麦斯纳小体来感知更精细的纹理(需要保持振动频率在 250~350Hz 以下进行较慢的运动)。一般地,通过反复训练可以改善触觉感知。

力觉感知一般是指皮肤深层的肌肉、肌腱和关节运动感受到的力量感和方向感,例如用户感受到的物体重力、方向力和阻力等。图 3.17 展示了一种带后坐力的交互仿真枪,其后坐力模块可以模拟枪的射速和后坐力力度。后坐力模拟部件由驱动电路、电磁铁、归位皮筋以及活动枪托四部分组成。控制器向驱动电路发送控制信号,驱动电路驱动电磁铁工作。然后,电磁铁芯向右运动,击打活动枪托。枪托作用于玩家肩部,使玩家感受到后坐力,体验到更真实、沉浸的交互体验。

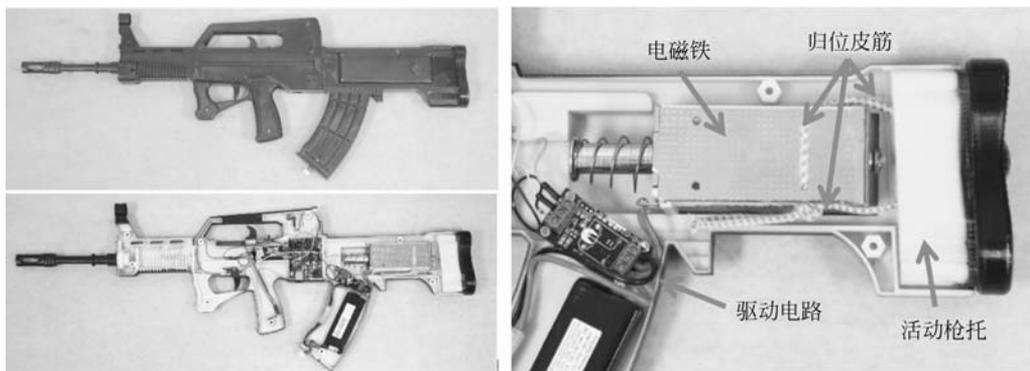


图 3.17 带后坐力触觉反馈的游戏仿真枪

图 3.18 还展示了一些常见例子。图 3.18(a)显示了美国初创公司 Haptx 推出的触觉反馈手套,支持用户在 VR 中感受到物体的存在,产生与现实环境中相似的触感体验。图 3.18(b)展示了 Go Touch VR 的一款简单有效的手指触觉设备 VR Touch,可配合 Oculus Rift 和 Leap Motion 使用。该设备用尼龙搭扣松紧带固定在指尖,通过塑料片让指尖感受到不同力量。它带有一个小电动机,通过戴在指尖的塑料片来产生反馈力,可以将触觉反馈到指尖上,让使用者体验到 VR 中更强的沉浸感和更加自然的交互体验(抓、触摸、按压等)。图 3.18(c)展示了一种 3D 打印笔的引导臂,有振动、抑制、摩擦等多种模拟触觉,可引导用户更好地完成立体模型创作。当用户在创作时超出模型范围或不知如何下笔时,机械臂会提供轻微的动作指引,让用户能顺利地进行下一步创作。用户经过多次使用后会慢慢地掌握技巧,轻松完成立体模型。图 3.18(d)展示了斯坦福大学研发的名为 Wolverine 的触觉反馈设备。这款设备固定在四只手指上,拇指处是蓝牙模块、控制模块以及电源。而在另外三只手指上的滑杆上均有低功率的制动锁定滑块,能根据用户在虚拟世界中抓取的物体提供抓取阻力,让用户感受到抓取物品的感觉。

VR 系统在触觉和力觉接口方面的研究还比较有限。虽然目前已经制造出了各种刺激用户指尖的手套和其他触觉的力反馈设备,但它们只是提供简单的高频振动、小范围的形状或压力分布以及温度特性,由此来刺激皮肤表面上的感受器。然而,这些仍然不能完全满足用户对

这方面沉浸感的需要,这是因为触觉感受器遍布在人的全身。为了在更大比例接收器上提供刺激,可能需要触觉服,它可以在套装上的不同点提供力、振动,甚至电刺激。

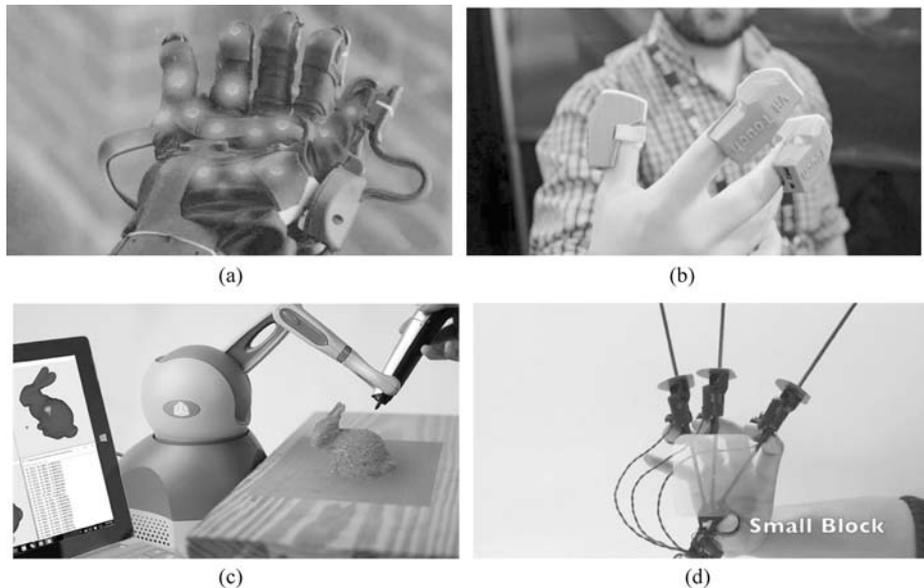


图 3.18 一些触觉反馈设备

3.4.2 本体感觉

本体感觉是一种能够感受到身体躯干和四肢各部位的位置、平衡、关节角度等姿态的感觉。当个体闭上眼睛在空旷的空间中移动手臂时,尽管无法精确地通过视觉和触觉进行定位,但仍然可以感知到手臂的空间位置。这体现了本体感觉对控制人体运动有着重要的作用。

人的体位感知器位于关节、肌肉和深层组织中,可分为以下三种类型。

- (1) 快速适应感受器(Rapidly Adapting Receptors): 用来感受四肢在某个方向的运动。
- (2) 慢速适应感受器(Slowly Adapting Receptors): 用来感受身体的移动和静态的位置。
- (3) 位置感受器(Positional Receptors): 用来感受人的一条胳膊或腿在空间的静止位置。

这些感受器的作用原理比较复杂。对关节角度的感知涉及位于皮肤、组织、关节、肌肉内的不同感受器的共同刺激。这些刺激信号组合在一起才能判断出关节信息。这些感觉不仅影响人的舒适感,而且影响人的行为表现。例如,通过手柄进行交互时,对手的相对位置的感知和手柄对手指的力反馈都是非常重要的。

大脑将多个传感模态信号组合来提供体感的感知。视觉和体感系统之间的信息不匹配会产生错觉。橡胶手错觉是最广为人知的该类现象之一(见图 3.19)。为此,科学家曾进行了一项实验,让被试坐在一张桌子前,并将双臂置于桌子上。其中,被试的左臂被遮盖且在其旁边放置了一个替代的橡胶手臂。该假手臂保持让被试可以看见,让被试有种假手臂好像是自己的左臂一样的感觉。实验者用画刷同时轻抚被试的真左手臂和假左臂,以帮助被试建立与假手臂的视觉和触觉联系。科学家利用功能磁共振成像扫描仪,发现无论轻抚被试的真手臂还是假手臂,被试大脑的同一部分都会被激活。此外,研究甚至发现,用针头刺假手臂会导致被试预期的疼痛,他们甚至会有撤回真正的左臂的趋势(尽管实际上并没有受到伤害),而且甚至可以通过联想来感知冷和热。

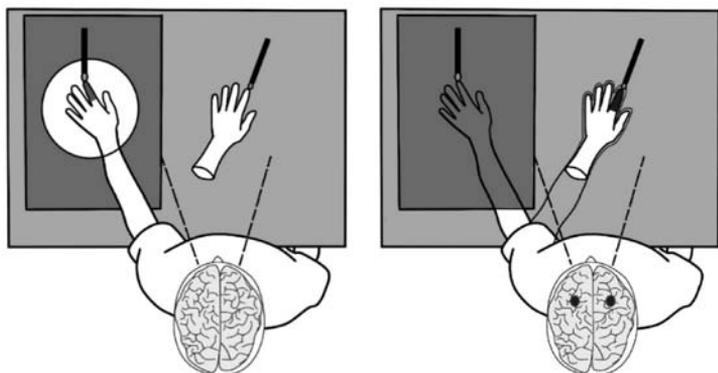
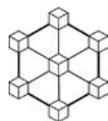


图 3.19 橡胶手错觉：被试对假手的反应就好像是自己的一样

该错觉也叫作身体转移错觉。在图 3.20 中显示了一个利用身体转移错觉来让人体验被捕食的鱼类处境的 VR 系统的例子。在此实验中，会让体验者带上 VR 头显设备并对其双脚进行捆绑，然后令其平躺在一个悬吊的模拟器上。在 VR 系统中体验者化身一条被追捕的鱼。由于体验者的身体被束缚，只能通过左右摇摆腿部来控制自己的身体。随后，VR 系统中体验者的鱼类化身会被鱼网网住并且挣扎，体验者也会感觉好像自己被网住一样。这种身体转移错觉现象也在 VR 应用中用于转移注意力或帮助截肢者克服幻肢感觉等。这种错觉还可以有助于通过控制肌肉，从 VR 中得到视觉反馈来提升沉浸感。



图 3.20 利用身体错觉让人体验被捕食的鱼类的处境



AR 图标

习题

1. 简述深度感知线索与立体视觉原理。
2. 简述听觉定位与立体声原理。
3. 简述几种主要感知现象的基本原理及在 VR 设计中的应用。
4. 试结合一个 VR 系统实例，说明该系统是如何结合用户多感知的原理和特点进行设计的。