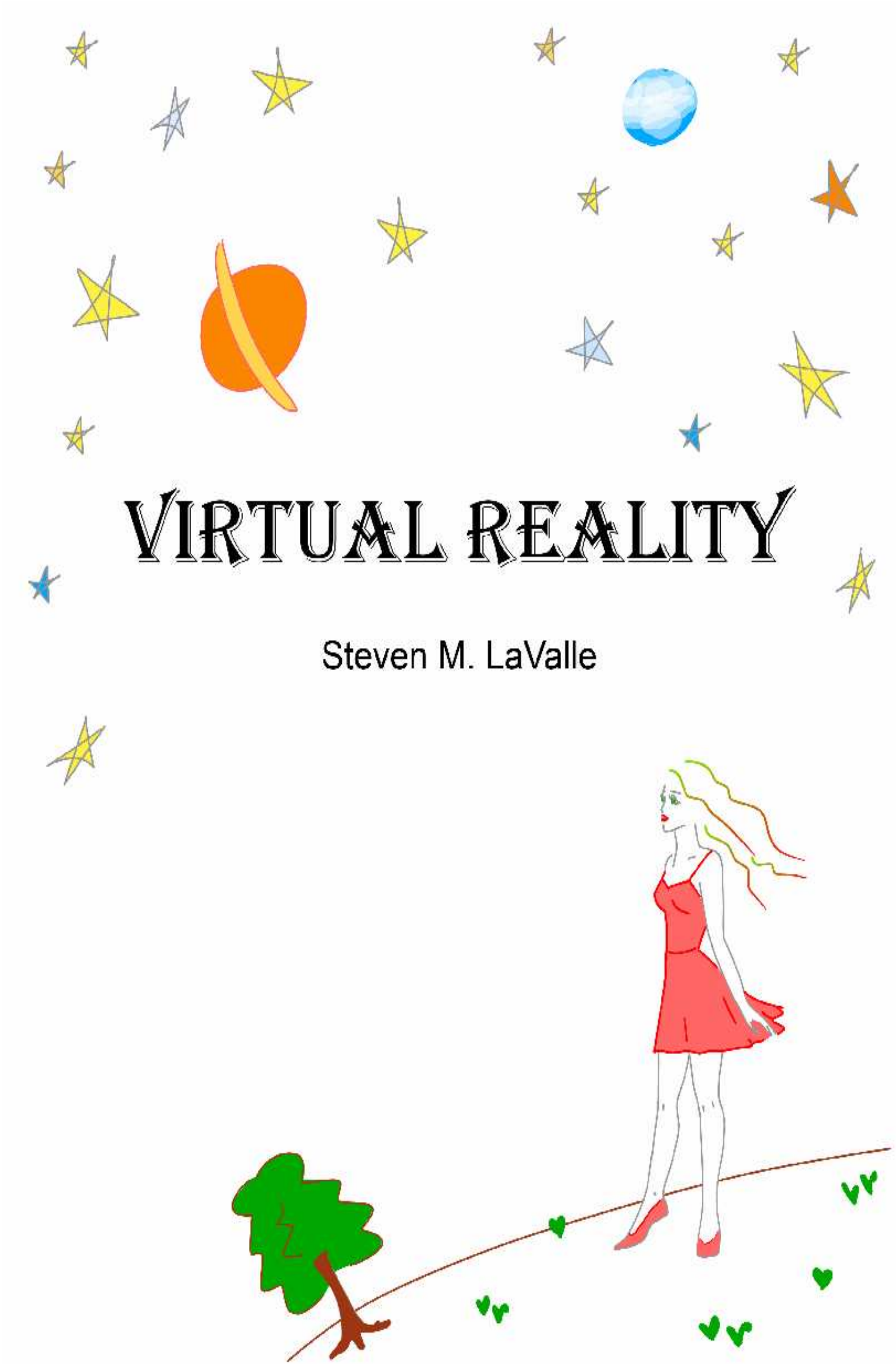


Chapter 2

Bird's-Eye View

Steven M. LaValle

University of Oulu

Available for downloading at **http://lavalle.pl/vr//**

Chapter 2

Bird's-Eye View

This chapter presents an overview of VR systems from hardware (Section 2.1) to software (Section 2.2) to human perception (Section 2.3). The purpose is to quickly provide a sweeping perspective so that the detailed subjects in the remaining chapters will be understood within the larger context. Further perspective can be gained by quickly jumping ahead to Section 12.2, which provides recommendations to VR developers. The fundamental concepts from the chapters leading up to that will provide the engineering and scientific background to understand why the recommendations are made. Furthermore, readers of this book should be able to develop new techniques and derive their own recommendations to others so that the VR systems and experiences are effective and comfortable.

2.1 Hardware

The first step to understanding how VR works is to consider what constitutes the entire *VR system*. It is tempting to think of it as being merely the hardware components, such as computers, headsets, and controllers. This would be woefully incomplete. As shown in Figure 2.1, it is equally important to account for the organism, which in this chapter will exclusively refer to a human *user*. The hardware produces stimuli that override the senses of the user. In the head-mounted display from Section 1.3 (Figure 1.30(b)), recall that tracking was needed to adjust the stimulus based on human motions. The VR hardware accomplishes this by using its own sensors, thereby *tracking* motions of the user. Head tracking is the most important, but tracking also may include button presses, controller movements, eye movements, or the movements of any other body parts. Finally, it is also important to consider the surrounding physical world as part of the VR system. In spite of stimulation provided by the VR hardware, the user will always have other senses that respond to stimuli from the real world. She also has the ability to change her environment through body motions. The VR hardware might also track objects other than the user, especially if interaction with them is part of the VR experience. Through a robotic interface, the VR hardware might also change the real world. One example is teleoperation of a robot through a VR interface.

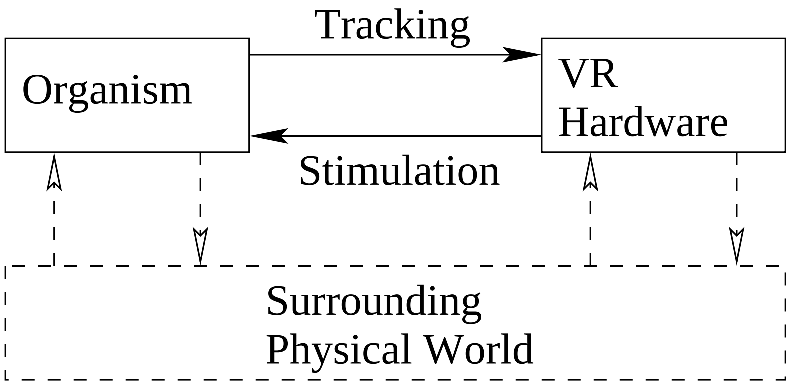


Figure 2.1: A third-person perspective of a VR system. It is wrong to assume that the engineered hardware and software are the complete VR system: The organism and its interaction with the hardware are equally important. Furthermore, interactions with the surrounding physical world continue to occur during a VR experience.

Sensors and sense organs How is information extracted from the physical world? Clearly this is crucial to a VR system. In engineering, a *transducer* refers to a device that converts energy from one form to another. A *sensor* is a special transducer that converts the energy it receives into a signal for an electrical circuit. This may be an analog or digital signal, depending on the circuit type. A sensor typically has a *receptor* that collects the energy for conversion. Organisms work in a similar way. The “sensor” is called a *sense organ*, with common examples being eyes and ears. Because our “circuits” are formed from interconnected neurons, the sense organs convert energy into *neural impulses*. As you progress through this book, keep in mind the similarities between engineered sensors and natural sense organs. They are measuring the same things and sometimes even function in a similar manner. This should not be surprising because we and our engineered devices share the same physical world: The laws of physics and chemistry remain the same.

Configuration space of sense organs As the user moves through the physical world, his sense organs move along with him. Furthermore, some sense organs move relative to the body skeleton, such as our eyes rotating within their sockets. Each sense organ has a *configuration space*, which corresponds to all possible ways it can be transformed or configured. The most important aspect of this is the number of *degrees of freedom* or *DOFs* of the sense organ. Chapter 3 will cover this thoroughly, but for now note that a rigid object that moves through

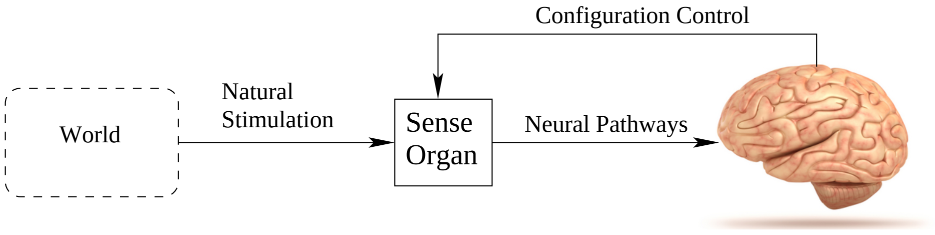


Figure 2.2: Under normal conditions, the brain (and body parts) control the configuration of sense organs (eyes, ears, fingertips) as they receive natural stimulation from the surrounding, physical world.

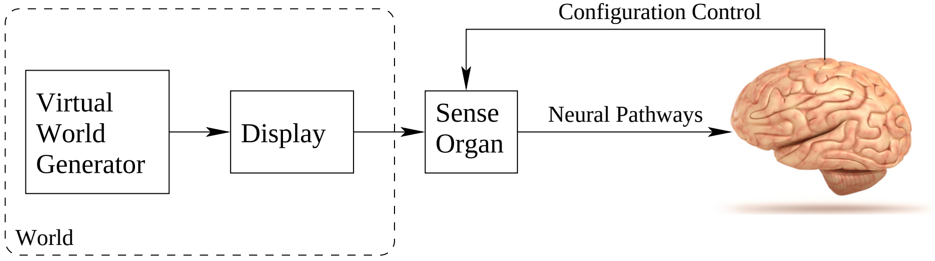


Figure 2.3: In comparison to Figure 2.2, a VR system "hijacks" each sense by replacing the natural stimulation with artificial stimulation that is provided by hardware called a display. Using a computer, a virtual world generator maintains a coherent, virtual world. Appropriate "views" of this virtual world are rendered to the display.

ordinary space has six DOFs. Three DOFs correspond to its changing position in space: 1) side-to-side motion, 2) vertical motion, and 3) closer-further motion. The other three DOFs correspond to possible ways the object could be rotated; in other words, exactly three independent parameters are needed to specify how the object is oriented. These are called yaw, pitch, and roll, and are covered in Section 3.2.

As an example, consider your left ear. As you rotate your head or move your body through space, the position of the ear changes, as well as its orientation. This yields six DOFs. The same is true for your right eye, but it also capable of rotating independently of the head. Keep in mind that our bodies have many more degrees of freedom, which affect the configuration of our sense organs. A tracking system may be necessary to determine the position and orientation of each sense organ that receives artificial stimuli, which will be explained shortly.

An abstract view Figure 2.2 illustrates the normal operation of one of our sense organs without interference from VR hardware. The brain controls its con-

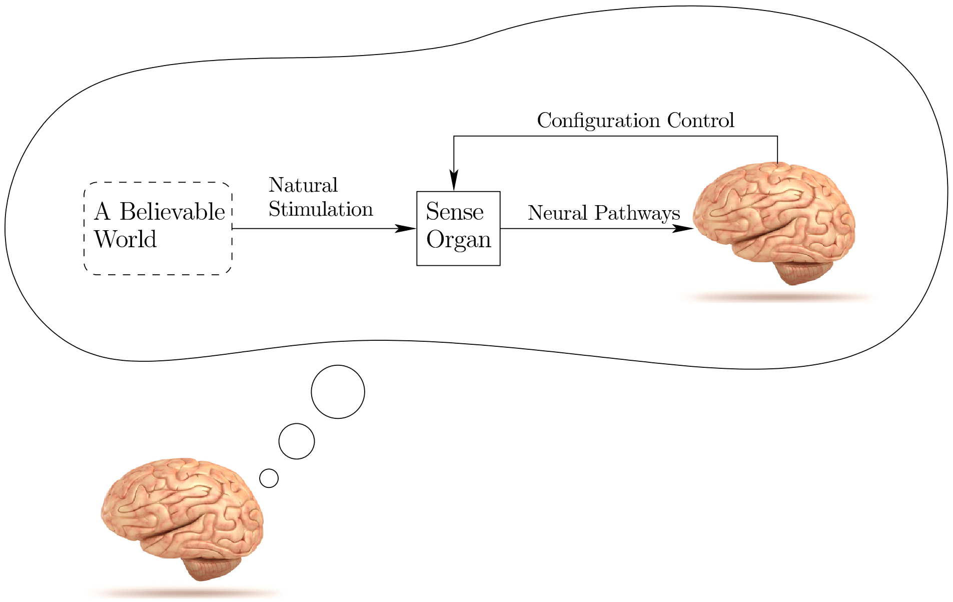


Figure 2.4: If done well, the brain is "fooled" into believing that the virtual world is in fact the surrounding physical world and natural stimulation is resulting from it.

figuration, while the sense organ converts natural stimulation from the environment into neural impulses that are sent to the brain. Figure 2.3 shows how it appears in a VR system. The VR hardware contains several components that will be discussed shortly. A *Virtual World Generator (VWG)* runs on a computer and produces "another world", which could be many possibilities, such as a pure simulation of a synthetic world, a recording of the real world, or a live connection to another part of the real world. The human perceives the virtual world through each targeted sense organ using a *display*, which emits energy that is specifically designed to mimic the type of stimulus that would appear without VR. The process of converting information from the VWG into output for the display is called *rendering*. In the case of human eyes, the display might be a smartphone screen or the screen of a video projector. In the case of ears, the display is referred to as a *speaker*. (A display need not be visual, even though this is the common usage in everyday life.) If the VR system is effective, then the brain is hopefully "fooled" in the sense shown in Figure 2.4. The user should believe that the stimulation of the senses is natural and comes from a plausible world, being consistent with at least some past experiences.

Aural: world-fixed vs. user-fixed Recall from Section 1.3 the trend of having to go somewhere for an experience, to having it in the home, and then finally to having it be completely portable. To understand these choices for VR systems and their implications on technology, it will be helpful to compare a simpler case: Audio or *aural* systems.

Figure 2.5 shows the speaker setup and listener location for a Dolby 7.1 Surround Sound theater system, which could be installed at a theater or a home family room. Seven speakers distributed around the room periphery generate most of the sound, while a subwoofer (the "1" of the "7.1") delivers the lowest frequency components. The aural displays are therefore *world-fixed*. Compare this to a listener wearing headphones, as shown in Figure 2.6. In this case, the aural displays are *user-fixed*. Hopefully, you have already experienced settings similar to these many times.

What are the key differences? In addition to the obvious portability of headphones, the following quickly come to mind:

- In the surround-sound system, the generated sound (or stimulus) is far away from the ears, whereas it is quite close for the headphones.

- One implication of the difference in distance is that much less power is needed for the headphones to generate an equivalent perceived loudness level compared with distant speakers.

- Another implication based on distance is the degree of privacy allowed by the wearer of headphones. A surround-sound system at high volume levels could generate a visit by angry neighbors.

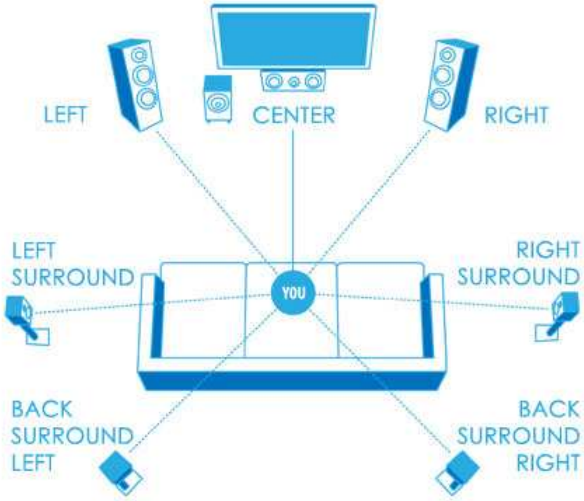


Figure 2.5: In a surround-sound system, the aural displays (speakers) are world-fixed while the user listens from the center.

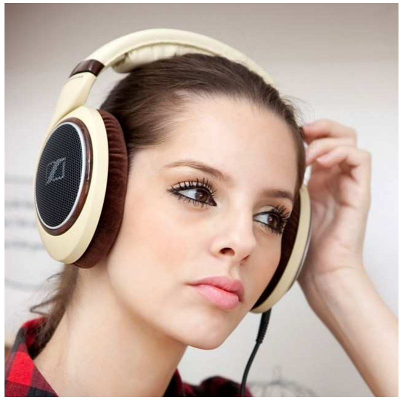


Figure 2.6: Using headphones, the displays are user-fixed, unlike the case of a surround-sound system.

- Wearing electronics on your head could be uncomfortable over long periods of time, causing a preference for surround sound over headphones.

- Several people can enjoy the same experience in a surround-sound system (although they cannot all sit in the optimal location). Using headphones, they would need to split the audio source across their individual headphones simultaneously.

- They are likely to have different costs, depending on the manufacturing difficulty and available component technology. At present, headphones are favored by costing much less than a set of surround-sound speakers (although one can spend a large amount of money on either).

All of these differences carry over to VR systems. This should not be too surprising because we could easily consider a pure audio experience to be a special kind of VR experience based on our definition from Section 1.1.

While listening to music, close your eyes and imagine you are at a live performance with the artists surrounding you. Where do you perceive the artists and their instruments to be located? Are they surrounding you, or do they seem to be in the middle of your head? Using headphones, it is most likely that they seem to be inside your head. In a surround-sound system, if recorded and displayed properly, the sounds should seem to be coming from their original locations well outside of your head. They probably seem constrained, however, into the horizontal plane that you are sitting in.

This shortcoming of headphones is not widely recognized at present, but nevertheless represents a problem that becomes much larger for VR systems that include visual displays. If you want to preserve your perception of where sounds are coming from, then headphones would need to take into account the configurations of your ears in space to adjust the output accordingly. For example, if you nod your head back and forth in a "no" gesture, then the sound being presented to each ear needs to be adjusted so that the simulated sound source is rotated in the opposite direction. In the surround-sound system, the speaker does not follow your head and therefore does not need to rotate. If the speaker rotates with your head, then a counter-rotation is needed to "undo" your head rotation so that the sound source location is perceived to be stationary.

Visual: world-fixed vs. user-fixed Now consider adding a visual display. You might not worry much about the perceived location of artists and instruments while listening to music, but you will quickly notice if their locations do not appear correct to your eyes. Our vision sense is much more powerful and complex than our sense of hearing. Figure 2.7(a) shows a CAVE system, which parallels the surround-sound system in many ways. The user again sits in the center while displays around the periphery present visual stimuli to his eyes. The speakers are replaced by video screens. Figure 2.7(b) shows a user wearing a VR headset, which parallels the headphones.

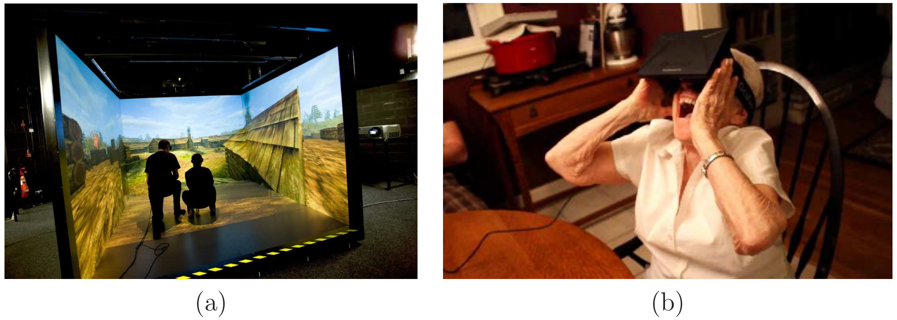
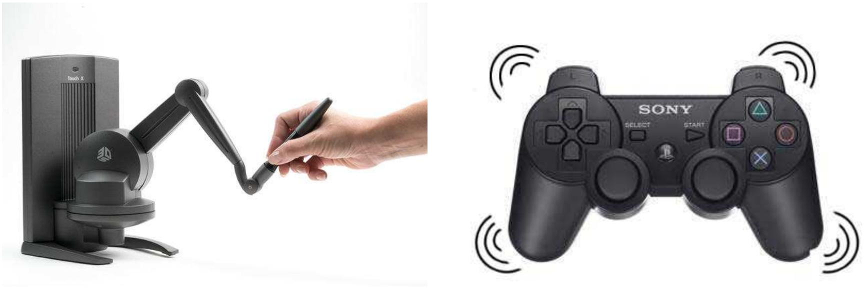


Figure 2.7: (a) A CAVE VR system developed at Teesside University, UK. (b) A 90-year-old woman (Rachel Mahassel) wearing the Oculus Rift DK1 headset in 2013.

Suppose the screen in front of the user's eyes shows a fixed image in the headset. If the user rotates his head, then the image will be perceived as being attached to the head. This would occur, for example, if you rotate your head while using the Viewmaster (recall Figure 1.29(b)). If you would like to instead perceive the image as part of a fixed world around you, then the image inside the headset must change to compensate as you rotate your head. The surrounding virtual world should be counter-rotated, the meaning of which will be made more precise in Section 3.4. Once we agree that such transformations are necessary, it becomes a significant engineering challenge to estimate the amount of head and eye movement that has occurred and apply the appropriate transformation in a timely and accurate manner. If this is not handled well, then users could have poor or unconvincing experiences. Worse yet, they could fall prey to VR sickness. This is one of the main reasons why the popularity of VR headsets waned in the 1990s. The component technology was not good enough yet. Fortunately, the situation is much improved at present. For audio, few seemed to bother with this transformation, but for the visual counterpart, it is absolutely critical. One final note is that tracking and applying transformations also becomes necessary in CAVE systems if we want the images on the screens to be altered according to changes in the eye *positions* inside of the room.

Now that you have a high-level understanding of the common hardware arrangements, we will take a closer look at hardware components that are widely available for constructing VR systems. These are expected to change quickly, with costs decreasing and performance improving. We also expect many new devices to appear in the marketplace in the coming years. In spite of this, the fundamentals in this book remain unchanged. Knowledge of the current technology provides concrete examples to make the fundamental VR concepts clearer.

The hardware components of VR systems are conveniently classified as:



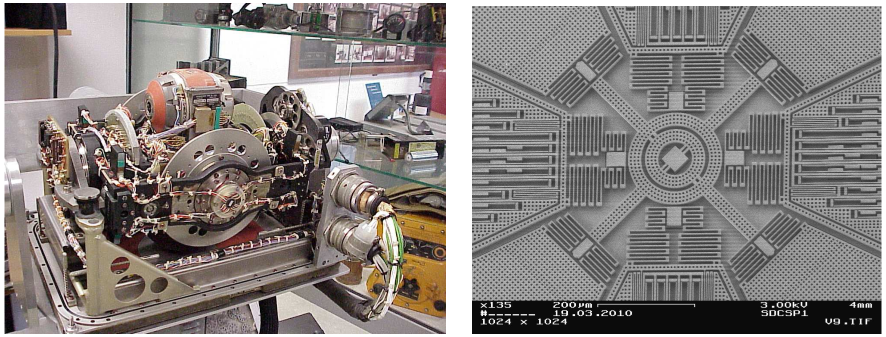
(a) (b)

Figure 2.8: Two examples of haptic feedback devices. (a) The Touch X system by 3D Systems allows the user to feel strong resistance when poking into a virtual object with a real stylus. A robot arm provides the appropriate forces. (b) Some game controllers occasionally vibrate.

- **Displays (output):** Devices that each stimulate a sense organ.
- **Sensors (input):** Devices that extract information from the real world.
- **Computers:** Devices that process inputs and outputs sequentially.

Displays A display generates stimuli for a targeted sense organ. Vision is our dominant sense, and any display constructed for the eye must cause the desired image to be formed on the retina. Because of this importance, Chapters 4 and 5 will explain displays and their connection to the human vision system. For CAVE systems, some combination of digital projectors and mirrors is used. Due to the plummeting costs, an array of large-panel displays may alternatively be employed. For headsets, a smartphone display can be placed close to the eyes and brought into focus using one magnifying lens for each eye. Screen manufacturers are currently making custom displays for VR headsets by leveraging the latest LED display technology from the smartphone industry. Some are targeting one display per eye with frame rates above 90Hz and over two megapixels per eye. Reasons for this are explained in Chapter 5.

Now imagine displays for other sense organs. Sound is displayed to the ears using classic speaker technology. Bone conduction methods may also be used, which vibrate the skull and propagate the waves to the inner ear; this method appeared Google Glass. Chapter 11 covers the auditory part of VR in detail. For the sense of touch, there are *haptic displays*. Two examples are pictured in Figure 2.8. Haptic feedback can be given in the form of vibration, pressure, or temperature. More details on displays for touch, and even taste and smell, appear in Chapter 13.

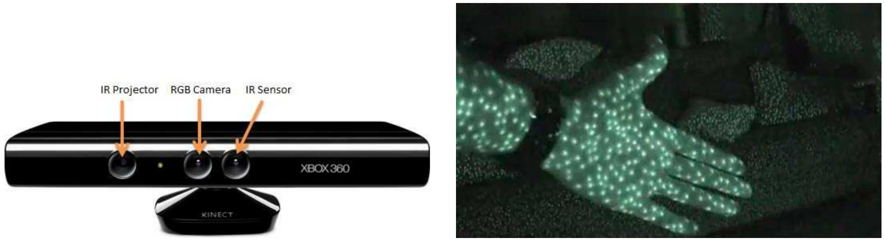


(a) (b)

Figure 2.9: Inertial measurement units (IMUs) have gone from large, heavy mechanical systems to cheap, microscopic MEMS circuits. (a) The LN-3 Inertial Navigation System, developed in the 1960s by Litton Industries. (b) The internal structures of a MEMS gyroscope, for which the total width is less than 1mm.

Sensors Consider the input side of the VR hardware. A brief overview is given here, until Chapter 9 covers sensors and tracking systems in detail. For visual and auditory body-mounted displays, the position and orientation of the sense organ must be tracked by sensors to appropriately adapt the stimulus. The orientation part is usually accomplished by an *inertial measurement unit* or *IMU*. The main component is a *gyroscope*, which measures its own rate of rotation; the rate is referred to as *angular velocity* and has three components. Measurements from the gyroscope are integrated over time to obtain an estimate of the cumulative change in orientation. The resulting error, called *drift error*, would gradually grow unless other sensors are used. To reduce drift error, IMUs also contain an *accelerometer* and possibly a *magnetometer*. Over the years, IMUs have gone from existing only as large mechanical systems in aircraft and missiles to being tiny devices inside of smartphones; see Figure 2.9. Due to their small size, weight, and cost, IMUs can be easily embedded in wearable devices. They are one of the most important enabling technologies for the current generation of VR headsets and are mainly used for tracking the user's head orientation.

Digital cameras provide another critical source of information for tracking systems. Like IMUs, they have become increasingly cheap and portable due to the smartphone industry, while at the same time improving in image quality. Cameras enable tracking approaches that exploit line-of-sight *visibility*. The idea is to identify features or markers in the image that serve as reference points for an moving object or a stationary background. Such *visibility constraints* severely limit the possible object positions and orientations. Standard cameras passively form an image by focusing the light through an optical system, much like the human eye. Once the camera calibration parameters are known, an observed



(a) (b)

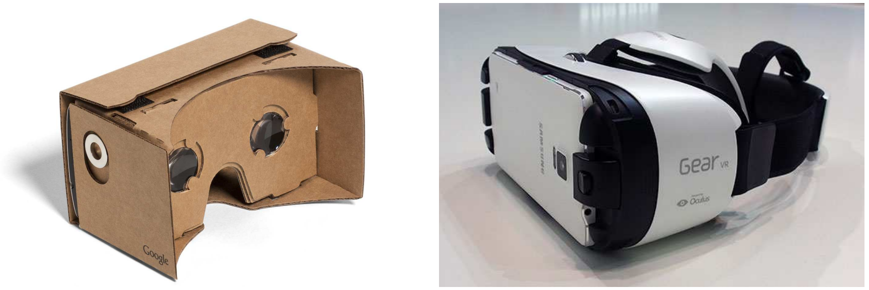
Figure 2.10: (a) The Microsoft Kinect sensor gathers both an ordinary RGB image and a depth map (the distance away from the sensor for each pixel). (b) The depth is determined by observing the locations of projected IR dots in an image obtained from an IR camera.

marker is known to lie along a ray in space. Cameras are commonly used to track eyes, heads, hands, entire human bodies, and any other objects in the physical world. One of the main challenges at present is to obtain reliable and accurate performance without placing special markers on the user or objects around the scene.

As opposed to standard cameras, *depth cameras* work actively by projecting light into the scene and then observing its reflection in the image. This is typically done in the infrared (IR) spectrum so that humans do not notice; see Figure 2.10.

In addition to these sensors, we rely heavily on good-old mechanical switches and potientiometers to create keyboards and game controllers. An optical mouse is also commonly used. One advantage of these familiar devices is that users can rapidly input data or control their characters by leveraging their existing training. A disadvantage is that they might be hard to find or interact with if their faces are covered by a headset.

Computers A computer executes the virtual world generator (VWG). Where should this computer be? Although unimportant for world-fixed displays, the location is crucial for body-fixed displays. If a separate PC is needed to power the system, then fast, reliable communication must be provided between the headset and the PC. This connection is currently made by wires, leading to an awkward tether; current wireless speeds are not sufficient. As you have noticed, most of the needed sensors exist on a smartphone, as well as a moderately powerful computer. Therefore, a smartphone can be dropped into a case with lenses to provide a VR experience with little added costs (Figure 2.11). The limitation, though, is that the VWG must be simpler than in the case of a separate PC so that it runs on less-powerful computing hardware. In the near future, we expect to see wireless, all-in-one headsets that contain all of the essential parts of smartphones for delivering VR experiences. These will eliminate unnecessary components of smartphones (such as the additional case), and will instead have customized optics, microchips, and sensors for VR.



(a) (b)

Figure 2.11: Two headsets that create a VR experience by dropping a smartphone into a case. (a) Google Cardboard works with a wide variety of smartphones. (b) Samsung Gear VR is optimized for one particular smartphone (in this case, the Samsung S6).

In addition to the main computing systems, specialized computing hardware may be utilized. Graphical processing units (GPUs) have been optimized for quickly rendering graphics to a screen and they are currently being adapted to handle the specific performance demands of VR. Also, a display interface chip converts an input video into display commands. Finally, microcontrollers are frequently used to gather information from sensing devices and send them to the main computer using standard protocols, such as USB.

To conclude with hardware, Figure 2.12 shows the hardware components for the Oculus Rift DK2, which became available in late 2014. In the lower left corner, you can see a smartphone screen that serves as the display. Above that is a circuit board that contains the IMU, display interface chip, a USB driver chip, a set of chips for driving LEDs on the headset for tracking, and a programmable microcontroller. The lenses, shown in the lower right, are placed so that the smartphone screen appears to be "infinitely far" away, but nevertheless fills most of the field of view of the user. The upper right shows flexible circuits that deliver power to IR LEDs embedded in the headset (they are hidden behind IR-transparent plastic). A camera is used for tracking, and its parts are shown in the center.

2.2 Software

From a developer's standpoint, it would be ideal to program the VR system by providing high-level descriptions and having the software determine automatically all of the low-level details. In a perfect world, there would be a *VR engine*, which

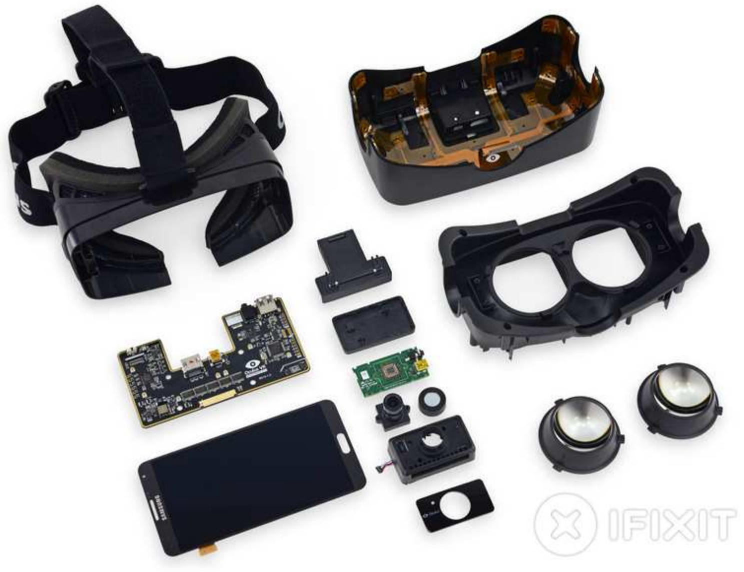


Figure 2.12: Disassembly of the Oculus Rift DK2 headset (figure from www.ifixit.com).

serves a purpose similar to the game engines available today for creating video games. If the developer follows patterns that many before her have implemented already, then many complicated details can be avoided by simply calling functions from a well-designed software library. However, if the developer wants to try something original, then she would have to design the functions from scratch. This requires a deeper understanding of the VR fundamentals, while also being familiar with lower-level system operations.

Unfortunately, we are currently a long way from having fully functional, general-purpose VR engines. As applications of VR broaden, specialized VR engines are also likely to emerge. For example, one might be targeted for immersive cinematography while another is geared toward engineering design. Which components will become more like part of a VR "operating system" and which will become higher level "engine" components? Given the current situation, developers will likely be implementing much of the functionality of their VR systems from scratch. This may involve utilizing a *software development kit (SDK)* for particular headsets that handles the lowest level operations, such as device drivers, head tracking, and display output. Alternatively, they might find themselves using a game engine that has been recently adapted for VR, even though it was fundamentally designed for video games on a screen. This can avoid substantial effort at first, but then may be cumbersome when someone wants to implement ideas that are not part of standard video games.

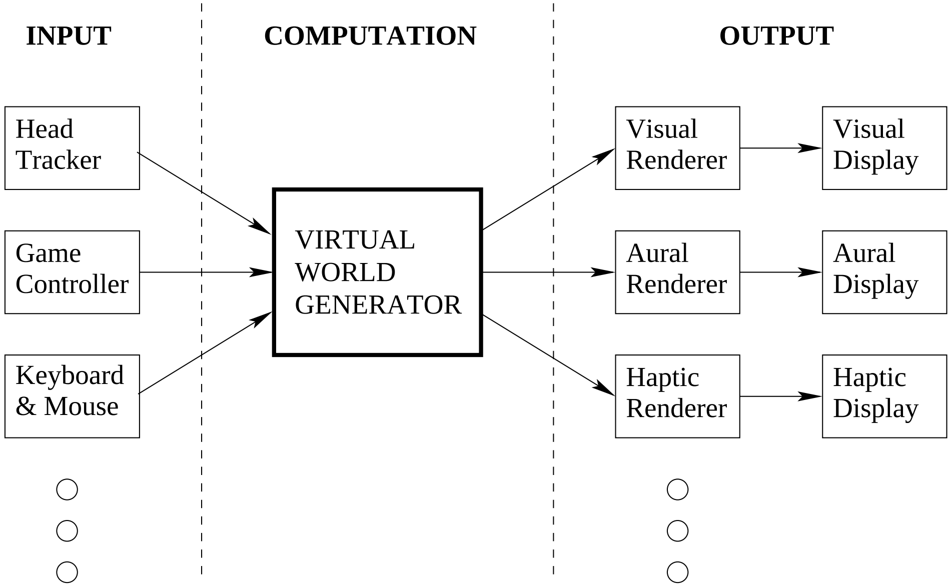


Figure 2.13: The Virtual World Generator (VWG) maintains another world, which could be synthetic, real, or some combination. From a computational perspective, the inputs are received from the user and his surroundings, and appropriate views of the world are rendered to displays.

What software components are needed to produce a VR experience? Figure 2.13 presents a high-level view that highlights the central role of the Virtual World Generator (VWG). The VWG receives inputs from low-level systems that indicate what the user is doing in the real world. A head tracker provides timely estimates of the user's head position and orientation. Keyboard, mouse, and game controller events arrive in a queue that are ready to be processed. The key role of the VWG is to maintain enough of an internal "reality" so that renderers can extract the information they need to calculate outputs for their displays.

Virtual world: real vs. synthetic At one extreme, the virtual world could be completely synthetic. In this case, numerous triangles are defined in a 3D space, along with material properties that indicate how they interact with light, sound, forces, and so on. The field of *computer graphics* addresses computer-generated images from synthetic models, and it remains important for VR; see Chapter 7. At the other extreme, the virtual world might be a recorded physical world that was captured using modern cameras, computer vision, and Simultaneous Localization and Mapping (SLAM) techniques; Figure 2.14. Many possibilities exist between

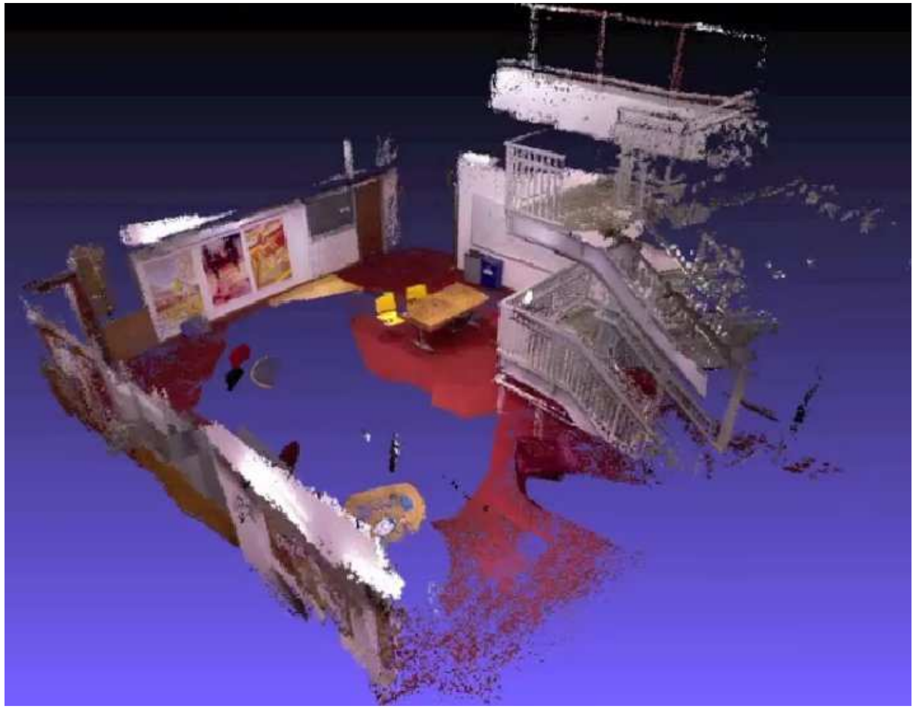


Figure 2.14: Using both color and depth information from cameras, a 3D model of the world can be extracted automatically using Simultaneous Localization and Mapping (SLAM) techniques. Figure from [2].

the extremes. For example, camera images may be taken of a real object, and then mapped onto a synthetic object in the virtual world. This is called *texture mapping*, a common operation in computer graphics; see Section 7.2.

Matched motion The most basic operation of the VWG is to maintain a correspondence between user motions in the real world and the virtual world; see Figure 2.15. In the real world, the user's motions are confined to a safe region, which we will call the *matched zone*. Imagine the matched zone as a place where the real and virtual worlds perfectly align. One of the greatest challenges is the mismatch of obstacles: What if the user is blocked in the virtual world but not in the real world? The reverse is also possible. In a seated experience, the user sits in a chair while wearing a headset. The matched zone in this case is a small region, such as one cubic meter, in which users can move their heads. Head motions should be matched between the two worlds. If the user is not constrained to a seat, then the matched zone could be an entire room or an outdoor field. Note that safety becomes an issue because the user might spill a drink, hit walls, or fall into pits that exist only in the real world, but are not visible in the virtual world. Larger matched zones tend to lead to greater safety issues. Users must make sure that the matched zone is cleared of dangers in the real world, or the developer should make them visible in the virtual world.

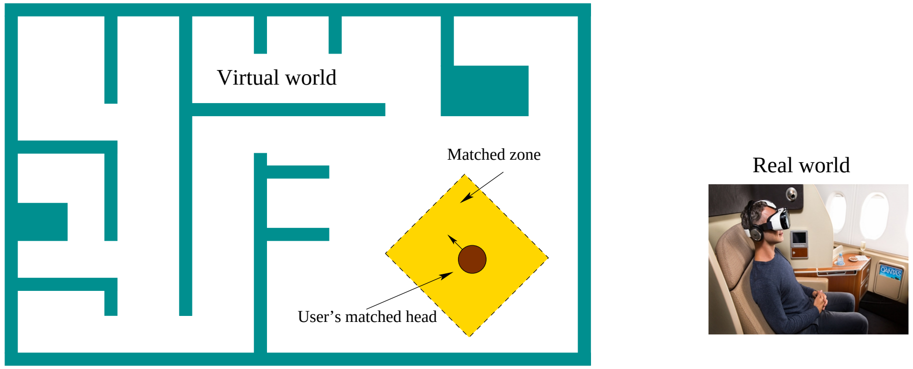


Figure 2.15: A matched zone is maintained between the user in their real world and his representation in the virtual world. The matched zone could be moved in the virtual world by using an interface, such as a game controller, while the user does not correspondingly move in the real world.

Which motions from the real world should be reflected in the virtual world? This varies among VR experiences. In a VR headset that displays images to the eyes, head motions must be matched so that the visual renderer uses the correct viewpoint in the virtual world. Other parts of the body are less critical, but may become important if the user needs to perform hand-eye coordination or looks at other parts of her body and expects them to move naturally.

User Locomotion In many VR experiences, users want to move well outside of the matched zone. This motivates *locomotion*, which means moving oneself in the virtual world, while this motion is not matched in the real world. Imagine you want to explore a virtual city while remaining seated in the real world. How should this be achieved? You could pull up a map and point to where you want to go, with a quick teleportation operation sending you to the destination. A popular option is to move oneself in the virtual world by operating a game controller, mouse, or keyboard. By pressing buttons or moving knobs, your self in the virtual world could be walking, running, jumping, swimming, flying, and so on. You could also climb aboard a vehicle in the virtual world and operate its controls to move yourself. These operations are certainly convenient, but often lead to sickness because of a mismatch between your balance and visual senses. See Sections 2.3, 10.2, and 12.3.

Physics The VWG handles the *geometric* aspects of motion by applying the appropriate mathematical transformations. In addition, the VWG usually implements some *physics* so that as time progresses, the virtual world behaves like the real world. In most cases, the basic laws of mechanics should govern how objects move in the virtual world. For example, if you drop an object, then it should accelerate to the ground due to gravitational force acting on it. One important component is a *collision detection* algorithm, which determines whether two or more bodies are intersecting in the virtual world. If a new collision occurs, then an appropriate response is needed. For example, suppose the user pokes his head through a wall in the virtual world. Should the head in the virtual world be stopped, even though it continues to move in the real world? To make it more complex, what should happen if you unload a dump truck full of basketballs into a busy street in the virtual world? Simulated physics can become quite challenging, and is a discipline in itself. There is no limit to the complexity. See Section 8.3 for more about virtual-world physics.

In addition to handling the motions of moving objects, the physics must also take into account how potential stimuli for the displays are created and propagate through the virtual world. How does light propagate through the environment? How does light interact with the surfaces in the virtual world? What are the sources of light? How do sound and smells propagate? These correspond to rendering problems, which are covered in Chapters 7 and 11 for visual and audio cases, respectively.

Networked experiences In the case of a networked VR experience, a shared virtual world is maintained by a server. Each user has a distinct matched zone. Their matched zones might overlap in a real world, but one must then be careful so that they avoid unwanted collisions. Most often, these zones are disjoint and distributed around the Earth. Within the virtual world, user interactions, including collisions, must be managed by the VWG. If multiple users are interacting in a social setting, then the burdens of matched motions may increase. As users meet each other, they could expect to see eye motions, facial expressions, and body language; see Section 10.4.

Developer choices for VWGs To summarize, a developer could start with a basic Software Development Kit (SDK) from a VR headset vendor and then build her own VWG from scratch. The SDK should provide the basic drivers and an interface to access tracking data and make calls to the graphical rendering libraries. In this case, the developer must build the physics of the virtual world from scratch, handling problems such as avatar movement, collision detection, lighting models, and audio. This gives the developer the greatest amount of control and ability to optimize performance; however, it may come in exchange for a difficult implementation burden. In some special cases, it might not be too difficult. For example, in the case of the Google Street viewer (recall Figure 1.10), the "physics" is simple: The viewing location needs to jump between panoramic images in a comfortable way while maintaining a sense of location on the Earth. In the case of telepresence using a robot, the VWG would have to take into account movements in the physical world. Failure to handle collision detection could result in a broken robot (or human!).

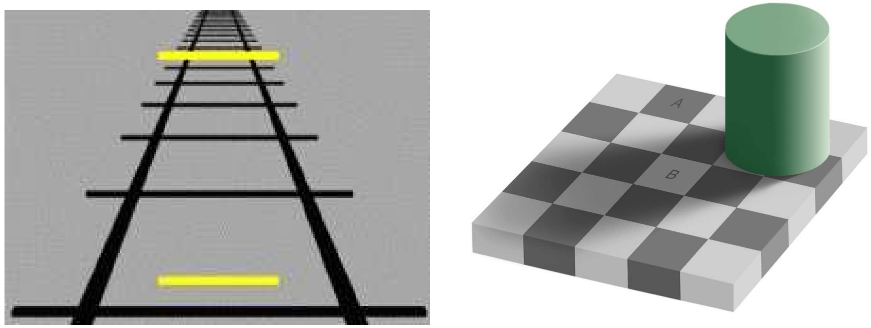
At the other extreme, a developer may use a ready-made VWG that is customized to make a particular VR experience by choosing menu options and writing high-level scripts. Examples available today are OpenSimulator, Vizard by WorldViz, Unity 3D, and Unreal Engine by Epic Games. The latter two are game engines that were adapted to work for VR, and are by far the most popular among current VR developers. The first one, OpenSimulator, was designed as an open-source alternative to Second Life for building a virtual society of avatars. As already stated, using such higher-level engines make it easy for developers to make a VR experience in little time; however, the drawback is that it is harder to make highly original experiences that were not imagined by the engine builders.

2.3 Human Physiology and Perception

Our bodies were not designed for VR. By applying artificial stimulation to the senses, we are disrupting the operation of biological mechanisms that have taken hundreds of millions of years to evolve in a natural environment. We are also providing input to the brain that is not exactly consistent with all of our other life experiences. In some instances, our bodies may adapt to the new stimuli. This could cause us to become unaware of flaws in the VR system. In other cases, we might develop heightened awareness or the ability to interpret 3D scenes that were once difficult or ambiguous. Unfortunately, there are also many cases where our bodies react by increased fatigue or headaches, partly because the brain is working harder than usual to interpret the stimuli. Finally, the worst case is the onset of VR sickness, which typically involves symptoms of dizziness and nausea.

Perceptual psychology is the science of understanding how the brain converts sensory stimulation into perceived phenomena. Here are some typical questions that arise in VR and fall under this umbrella:

- How far away does that object appear to be?
- How much video resolution is needed to avoid seeing pixels?
- How many frames per second are enough to perceive motion as continuous?
- Is the user's head appearing at the proper height in the virtual world?
- Where is that virtual sound coming from?
- Why am I feeling nauseated?



(a) (b)

Figure 2.16: Optical illusions present an unusual stimulus that highlights limitations of our vision system. (a) The *Ponzo illusion* causes the upper line segment to appear larger than the lower one, even though they are the same length. (b) The *checker shadow illusion* causes the B tile to appear lighter than the A tile, even though they are the exactly the same shade of gray (figure by Adrian Pingstone).

- Why is one experience more tiring than another?
- What is presence?

To answer these questions and more, we must understand several things: 1) basic physiology of the human body, including sense organs and neural pathways, 2) the key theories and insights of experimental perceptual psychology, and 3) the interference of the engineered VR system with our common perceptual processes and the resulting implications or side effects.

The perceptual side of VR often attracts far too little attention among developers. In the real world, perceptual processes are mostly invisible to us. Think about how much effort it requires to recognize a family member. When you see someone you know well, the process starts automatically, finishes immediately, and seems to require no effort. Scientists have conducted experiments that reveal how much work actually occurs in this and other perceptual processes. Through brain lesion studies, they are able to see the effects when a small part of the brain is not functioning correctly. Some people suffer from *prosopagnosia*, which makes them unable to recognize the faces of familiar people, including themselves in a mirror, even though nearly everything else functions normally. Scientists are also able to perform *single-unit recordings*, mostly on animals, which reveal the firings of a single neuron in response to sensory stimuli. Imagine, for example, a single neuron that fires whenever you see a sphere.

Optical illusions One of the most popular ways to appreciate the complexity of our perceptual processing is to view optical illusions. These yield surprising results and are completely unobtrusive. Each one is designed to reveal some shortcoming of our visual system by providing a stimulus that is not quite consistent with ordinary stimuli in our everyday lives. Figure 2.16 shows two. These should motivate you to appreciate the amount of work that our sense organs and neural structures are doing to fill in missing details and make interpretations based on the context of our life experiences and existing biological structures. Interfering with these without understanding them is not wise!

Sense	Stimulus	Receptor	Sense Organ
Vision	Electromagnetic energy	Photoreceptors	Eye
Auditory	Air pressure waves	Mechanoreceptors	Ear
Touch	Tissue distortion	Mechanoreceptors	Skin, muscles
		Thermoreceptors	Skin
Balance	Gravity, acceleration	Mechanoreceptors	Vestibular organs
Taste/smell	Chemical composition	Chemoreceptors	Mouth, nose

Figure 2.17: A classification of the human body senses.

Classification of senses Perception and illusions are not limited to our eyes. Figure 2.17 shows a classification of our basic senses. Recall that a sensor converts an energy source into signals in a circuit. In the case of our bodies, this means that a stimulus is converted into neural impulses. For each sense, Figure 2.17 indicates the type of energy for the stimulus and the *receptor* that converts the stimulus into neural impulses. Think of each receptor as a sensor that targets a particular kind of stimulus. This is referred to as *sensory system selectivity*. In each eye, over 100 million photoreceptors target electromagnetic energy precisely in the frequency range of visible light. Different kinds even target various colors and light levels; see Section 5.1. The auditory, touch, and balance senses involve motion, vibration, or gravitational force; these are sensed by mechanoreceptors. The physiology and perception of hearing are covered in Sections 11.2 and 11.3, respectively. The sense of touch additionally involves thermoreceptors to detect change in temperature. Touch is covered in Section 13.1. Our *balance sense* helps us to know which way our head is oriented, including sensing the direction of “up”; this is covered in Section 8.2. Finally, our sense of taste and smell is grouped into one category, called the *chemical senses*, that relies on chemoreceptors; these provide signals based on chemical composition of matter appearing on our tongue or in our nasal passages; see Section 13.2.

Note that senses have engineering equivalents, most of which appear in VR systems. Imagine you a designing a humanoid telepresence robot, which you expect to interface with through a VR headset. You could then experience life through your surrogate robotic self. Digital cameras would serve as its eyes, and microphones would be the ears. Pressure sensors and thermometers could be installed to give a sense of touch. For balance, we can install an IMU. In fact, the

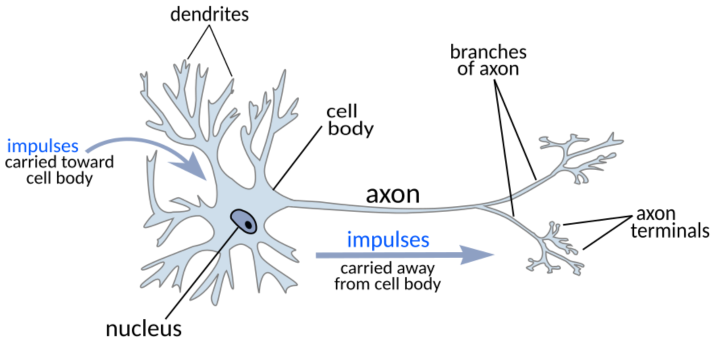


Figure 2.18: A typical neuron receives signals through dendrites, which interface to other neurons. It outputs a signal to other neurons through axons.

human *vestibular organs* and modern IMUs bear a striking resemblance in terms of the signals they produce; see Section 8.2. We could even install chemical sensors, such as a pH meter, to measure aspects of chemical composition to provide taste and smell.

Big brains Perception happens after the sense organs convert the stimuli into neural impulses. According to latest estimates [1], human bodies contain around 86 billion neurons. Around 20 billion are devoted to the part of the brain called the *cerebral cortex*, which handles perception and many other high-level functions such as attention, memory, language, and consciousness. It is a large sheet of neurons around three millimeters thick and is heavily folded so that it fits into our skulls. In case you are wondering where we lie among other animals, a roundworm, fruit fly, and rat have 302, 100 thousand, and 200 million neurons, respectively. An elephant has over 250 billion neurons, which is more than us!

Only mammals have a cerebral cortex. The cerebral cortex of a rat has around 20 million neurons. Cats and dogs are at 300 and 160 million, respectively. A gorilla has around 4 billion. A type of dolphin called the long-finned pilot whale has an estimated 37 billion neurons in its cerebral cortex, making it roughly twice as many as in the human cerebral cortex; however, scientists claim this does not imply superior cognitive abilities [5, 6].

Another important factor in perception and overall cognitive ability is the interconnection between neurons. Imagine an enormous directed graph, with the usual nodes and directed edges. The nucleus or cell body of each neuron is a node that does some kind of "processing". Figure 2.18 shows a neuron. The *dendrites* are essentially input edges to the neuron, whereas the *axons* are output edges. Through a network of dendrites, the neuron can aggregate information from numerous other neurons, which themselves may have aggregated information from others. The result is sent to one or more neurons through the axon. For a connected axon-dendrite pair, communication occurs in a gap called the *synapse*, where electrical or chemical signals are passed along. Each neuron in the human brain has on average about 7000 synaptic connections to other neurons, which results in about 10^{15} edges in our enormous brain graph!

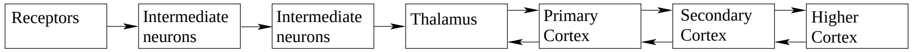


Figure 2.19: The stimulus captured by receptors works its way through a hierarchical network of neurons. In the early stages, signals are combined from multiple receptors and propagated upward. At later stages, information flows bidirectionally.

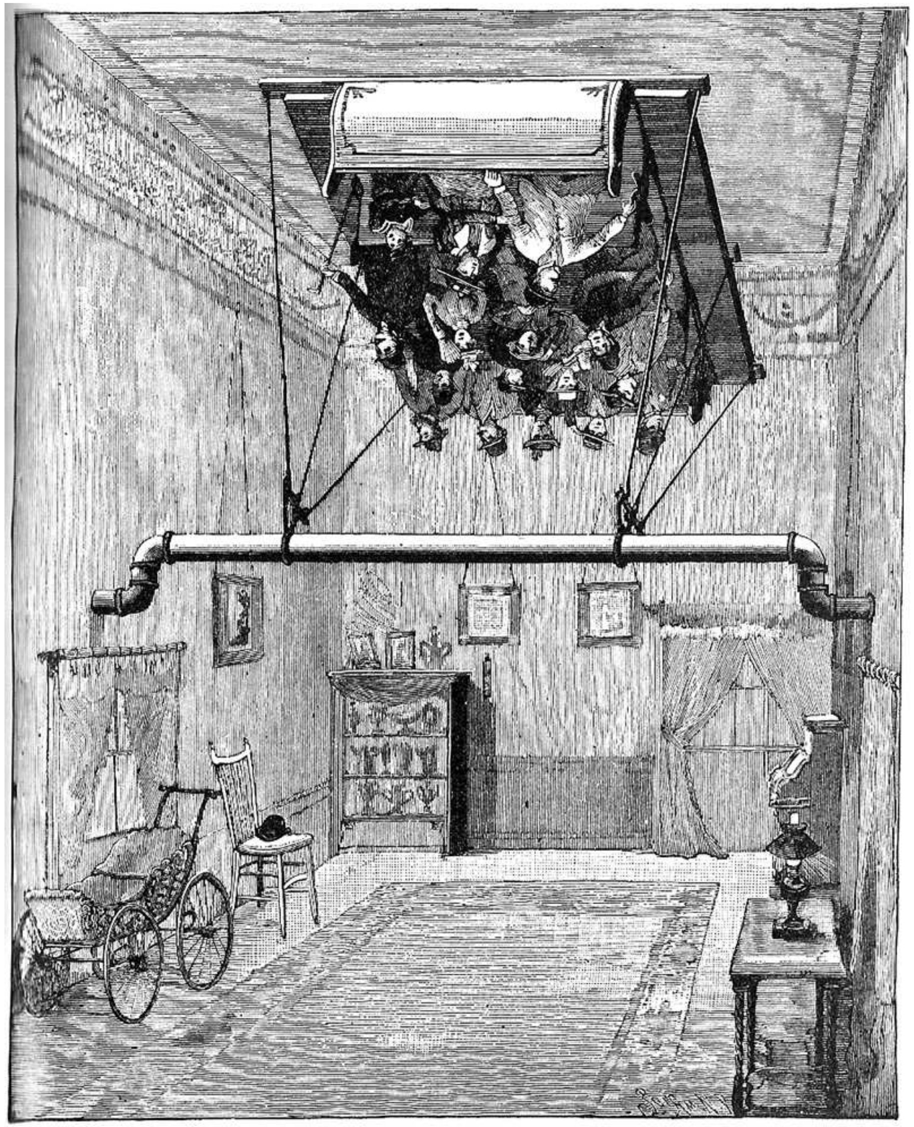
Hierarchical processing Upon leaving the sense-organ receptors, signals propagate among the neurons to eventually reach the cerebral cortex. Along the way, *hierarchical processing* is performed; see Figure 2.19. Through selectivity, each receptor responds to a narrow range of stimuli, across time, space, frequency, and so on. After passing through several neurons, signals from numerous receptors are simultaneously taken into account. This allows increasingly complex patterns to be detected in the stimulus. In the case of vision, feature detectors appear in the early hierarchical stages, enabling us to detect features such as edges, corners, and motion. Once in the cerebral cortex, the signals from sensors are combined with anything else from our life experiences that may become relevant for making an interpretation of the stimuli. Various *perceptual phenomena* occur, such as recognizing a face or identifying a song. Information or concepts that appear in the cerebral cortex tend to represent a global picture of the world around us. Surprisingly, *topographic mapping* methods reveal that spatial relationships among receptors are maintained in some cases among the distribution of neurons. Also, recall from Section 1.1 that place cells and grid cells encode spatial maps of familiar environments.

Proprioception In addition to information from senses and memory, we also use *proprioception*, which is the ability to sense the relative positions of parts of our bodies and the amount of muscular effort being involved in moving them. Close your eyes and move your arms around in an open area. You should have an idea of where your arms are located, although you might not be able to precisely reach out and touch your fingertips together without using your eyes. This information is so important to our brains that the *motor cortex*, which controls body motion, sends signals called *efference copies* to other parts of the brain to communicate what motions have been executed. Proprioception is effectively another kind of sense. Continuing our comparison with robots, it corresponds to having *encoders* on joints or wheels, to indicate how far they have moved. One interesting implication of proprioception is that you cannot tickle yourself because you know where your fingers are moving; however, if someone else tickles you, then you do not have access to their efference copies. The lack of this information is crucial to the tickling sensation.

Fusion of senses Signals from multiple senses and proprioception are being processed and combined with our experiences by our neural structures throughout our lives. In ordinary life, without VR or drugs, our brains interpret these combinations of inputs in coherent, consistent, and familiar ways. Any attempt to interfere with these operations is likely to cause a mismatch among the data from our senses. The brain may react in a variety of ways. It could be the case that we are not consciously aware of the conflict, but we may become fatigued or develop a headache. Even worse, we could develop symptoms of dizziness or nausea. In other cases, the brain might react by making us so consciously aware of the conflict that we immediately understand that the experience is artificial. This would correspond to a case in which the VR experience is failing to convince people that they are present in a virtual world. To make an effective and comfortable VR experience, trials with human subjects are essential to understand how the brain reacts. It is practically impossible to predict what would happen in an unknown scenario, unless it is almost identical to other well-studied scenarios.

One of the most important examples of bad sensory conflict in the context of VR is *vection*, which is the illusion of self motion. The conflict arises when your vision sense reports to your brain that you are accelerating, but your balance sense reports that you are motionless. As people walk down the street, their balance and vision senses are in harmony. You might have experienced vection before, even without VR. If you are stuck in traffic or stopped at a train station, you might have felt as if you are moving backwards while seeing a vehicle in your periphery that is moving forward. In the 1890s, Amariah Lake constructed an amusement park ride that consisted of a swing that remains at rest while the entire room surrounding the swing rocks back-and-forth (Figure 2.20). In VR, vection is caused by the locomotion operation described in Section 2.2. For example, if you accelerate yourself forward using a controller, rather than moving forward in the real world, then you perceive acceleration with your eyes, but not your vestibular organ. For strategies to alleviate this problem, see Section 10.2.

Adaptation A universal feature of our sensory systems is *adaptation*, which means that the perceived effect of stimuli changes over time. This may happen with any of our senses and over a wide spectrum of time intervals. For example, the perceived loudness of motor noise in an aircraft or car decreases within minutes. In the case of vision, the optical system of our eyes and the photoreceptor sensitivities adapt to change perceived brightness. Over long periods of time, *perceptual training* can lead to adaptation; see Section 12.1. In military training simulations, sickness experienced by soldiers appears to be less than expected, perhaps due to regular exposure [3]. Anecdotally, the same seems to be true of



ILLUSION PRODUCED BY A RIDE IN THE SWING.

Figure 2.20: A virtual swinging experience was made by spinning the surrounding room instead of the swing. This is known as the *haunted swing illusion*. People who tried it were entertained, but they became nauseated from an extreme version of vection. (Compiled and edited by Albert A. Hopkins, Munn & Co., Publishers, scanned by Alistair Gentry from "Magic Stage Illusions and Scientific Diversions, Including Trick Photography", 1898.)

experienced video game players. Those who have spent many hours and days in front of large screens playing first-person shooter games apparently experience less vection when locomoting themselves in VR.

Adaptation therefore becomes a crucial factor for VR. Through repeated exposure, developers may become comfortable with an experience that is nauseating to a newcomer. This gives them a terrible bias while developing an experience; recall from Section 1.1 the problem of confusing the scientist with the lab subject in the VR experiment. On the other hand, through repeated, targeted training developers may be able to improve their debugging skills by noticing flaws in the system that an "untrained eye" would easily miss. Common examples include:

- A large amount of tracking latency has appeared, which interferes with the *perception of stationarity.*
- The left and right eye views are swapped.
- Objects appear to one eye but not the other.
- One eye view has significantly more latency than the other.
- Straight lines are slightly curved due to uncorrected warping in the optical system.

This disconnect between the actual stimulus and one's perception of the stimulus leads to the next topic.

Psychophysics *Psychophysics* is the scientific study of perceptual phenomena that are produced by physical stimuli. For example, under what conditions would someone call an object "red"? The stimulus corresponds to light entering the eye, and the perceptual phenomenon is the concept of "red" forming in the brain. Other examples of perceptual phenomena are "straight", "larger", "louder", "tickles", and "sour". Figure 2.21 shows a typical scenario in a psychophysical experiment. As one parameter is varied, such as the frequency of a light, there is usually a range of values for which subjects cannot reliably classify the phenomenon. For example, there may be a region where they are not sure whether the light is red. At one extreme, they may consistently classify it as "red" and at the other extreme, they consistently classify it as "not red". For the region in between, the *probability of detection* is recorded, which corresponds to the frequency with which it is classified as "red". Section 12.4 will discuss how such experiments are designed and conducted.

Stevens' power law One of the most known results from psychophysics is *Steven's power law*, which characterizes the relationship between the magnitude of a physical stimulus and its *perceived* magnitude [7]. The hypothesis is that an exponential relationship occurs over a wide range of sensory systems and stimuli:

$$p = cm^x \tag{2.1}$$

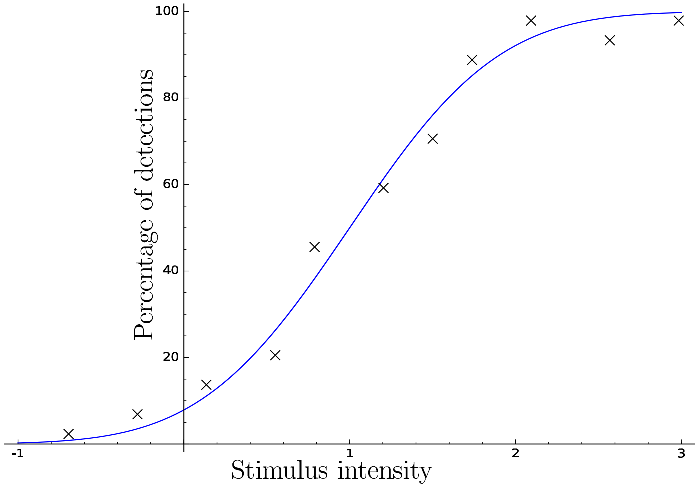


Figure 2.21: The most basic *psychometric function*. For this example, as the stimulus intensity is increased, the percentage of people detecting the phenomenon increases. The point along the curve that corresponds to 50 percent indicates a critical threshold or boundary in the stimulus intensity. The curve above corresponds to the cumulative distribution function of the error model (often assumed to be Gaussian).

in which

- m is the magnitude or intensity of the stimulus,
- p is the perceived magnitude,
- x relates the actual magnitude to the perceived magnitude, and is the most important part of the equation, and
- c is an uninteresting constant that depends on units.

Note that for $x = 1$, (2.1) is a linear relationship, $p = cm$; see Figure 2.22. An example of this is our perception of the length of an isolated line segment directly in front of our eyes. The length we perceive is proportional to its actual length. The more interesting cases are when $x \neq 1$. For the case of perceiving the brightness of a target in the dark, $x = 0.33$, which implies that a large increase in brightness is perceived as a smaller increase. In the other direction, our perception of electric shock as current through the fingers yields $x = 3.5$. A little more shock is a lot more uncomfortable!

Just noticeable difference Another key psychophysical concept is the *just noticeable difference* (*JND*). This is the amount that the stimulus needs to be

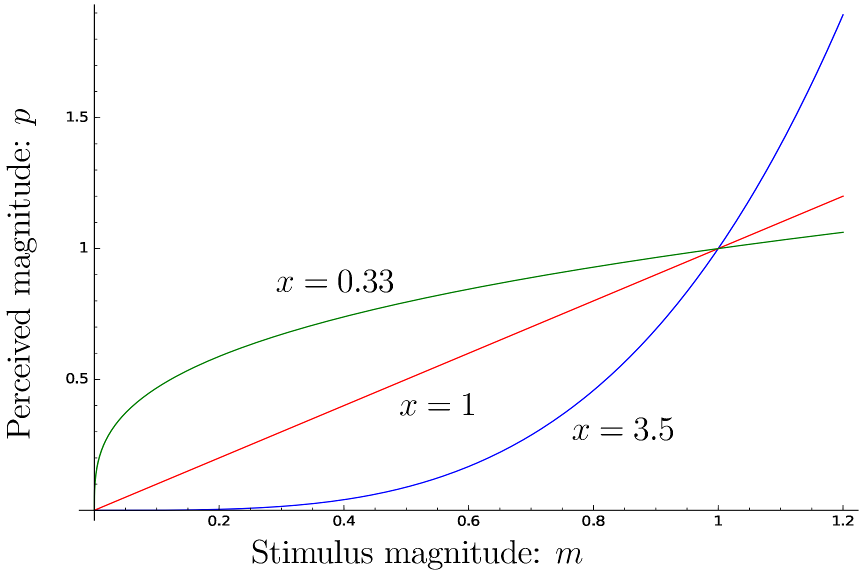


Figure 2.22: Steven's power law (2.1) captures the relationship between the magnitude of a stimulus and its perceived magnitude. The model is an exponential curve, and the exponent depends on the stimulus type.

changed so that subjects would perceive it to have changed in at least 50 percent of trials. For a large change, all or nearly all subjects would report a change. If the change is too small, then none or nearly none of the subjects would notice. The experimental challenge is to vary the amount of change until the chance of someone reporting a change is 50 percent.

Consider the JND for a stimulus with varying magnitude, such as brightness. How does the JND itself vary as the magnitude varies? This relationship is captured by *Weber's law*:

$$\frac{\Delta m}{m} = c, \tag{2.2}$$

in which Δm is the JND, m is the magnitude of the stimulus, and c is a constant.

Design of experiments VR disrupts the ordinary perceptual processes of its users. It should be clear from this section that proposed VR systems and experiences need to be evaluated on users to understand whether they are yielding the desired effect while also avoiding unwanted side effects. This amounts to applying the scientific method to make observations, formulate hypotheses, and design experiments that determine their validity. When human subjects are involved, this becomes extremely challenging. How many subjects are enough? What happens if they adapt to the experiment? How does their prior world experience affect the experiment? What if they are slightly sick the day that they try the experiment? What did they eat for breakfast? The answers to these questions could dramatically affect the outcome.

It gets worse. Suppose they already know your hypothesis going into the experiment. This will most likely bias their responses. Also, what will the data from the experiment look like? Will you ask them to fill out a questionnaire, or will you make inferences about their experience from measured data such as head motions, heart rate, and skin conductance? These choices are also critical. See Section 12.4 for more on this topic.

Further Reading

The particular software and hardware technologies described in this chapter are rapidly evolving. A quick search of the Internet at any give time should reveal the latest headsets and associated tools for developers. The core concepts, however, remain largely unchanged and are covered in the coming chapters. For broader coverage of human physiology and perception, see [4] and numerous other books with “Sensation and Perception” in the title.

Bibliography

[1] F. A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Computational Neurology*, 513:532–541, 2009.

[2] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. Visual odometry and mapping for autonomous flight using an RGB-D camera. In *Proceedings International Symposium on Robotics Research*, 2011.

[3] B. D. Lawson. Motion sickness symptomatology and origins. In K. S. Hale and K. M. Stanney, editors, *Handbook of Virtual Environments, 2nd Edition*, pages 531–600. CRC Press, Boca Raton, FL, 2015.

[4] G. Mather. *Foundations of Sensation and Perception*. Psychology Press, Hove, UK, 2008.

[5] H. S. Mortensen, B. Pakkenberg, M. Dam, R. Dietz, C. Sonne, B. Mikkelsen, and N. Eriksen. Quantitative relationships in delphinid neocortex. *Frontiers in Neuroanatomy*, 8, 2014.

[6] G. Roth and U. Dicke. Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, 9:250–257, 2005.

[7] S. S. Stevenson. On the psychophysical law. *Psychological Review*, 64(3):153–181, 1957.