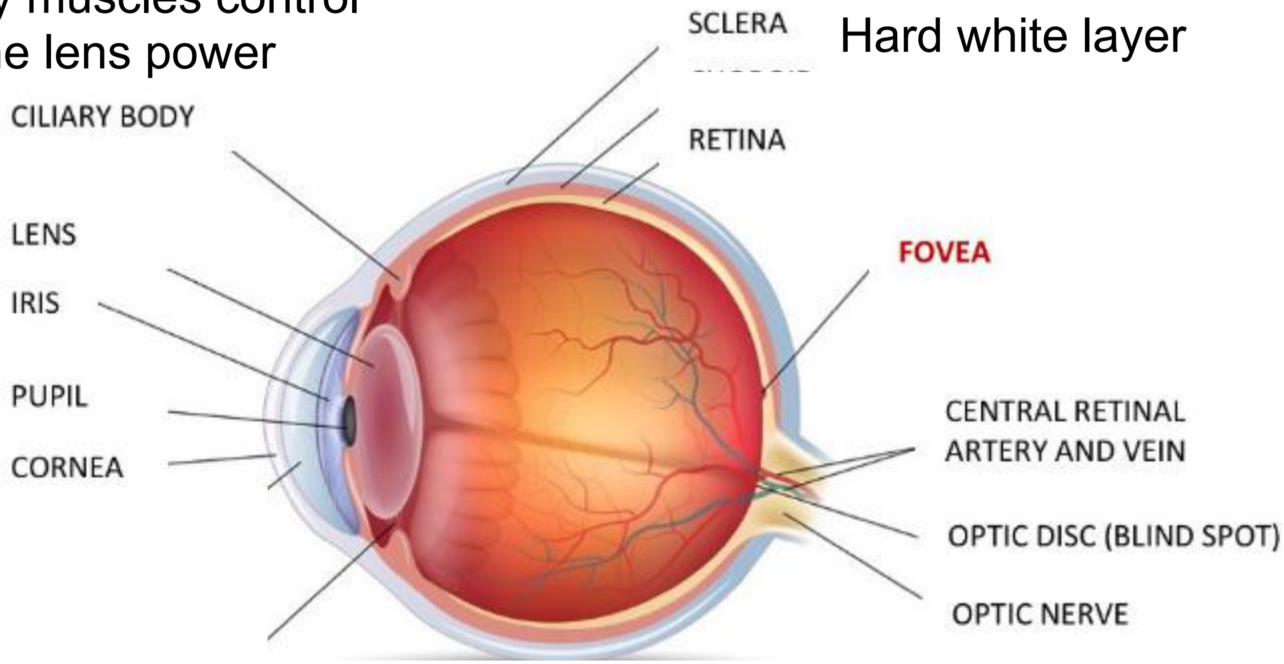
LECTURE 7: VISUAL PERCEPTION

Ehsan Aryafar
earyafar@pdx.edu
http://web.cecs.pdx.edu/~aryafare/VR.html

Recall: Anatomy of Human Eye

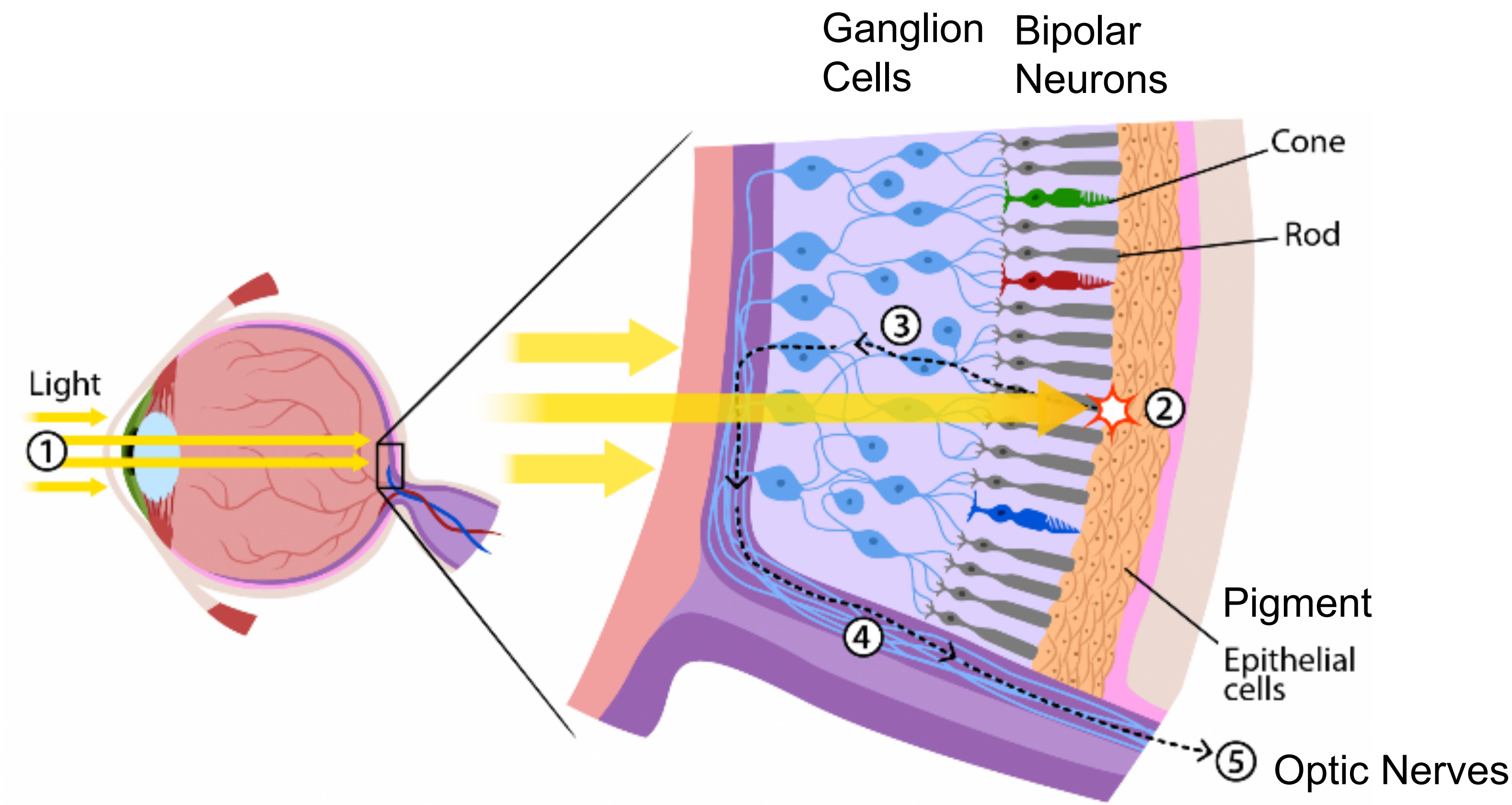
Ciliary muscles control the lens power

Hard white layer



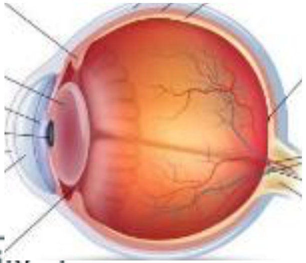
Eye: spherical, 24 mm diameter; Interior is a gelatinous mass, allows light to penetrate
Cornea is a hard, transparent surface through which light enters (high optical power)
Light enters the lens by passing through pupil, the size of which controlled by Iris
Light eventually hits the retina, which is disc shaped (180 degrees)
Retina is covered by photoreceptors, which **behave like input pixels**
Fovea is the most important part of retina with highest visual activity
Optical disc is a hole in retina through which neural pulses are sent outside the eye

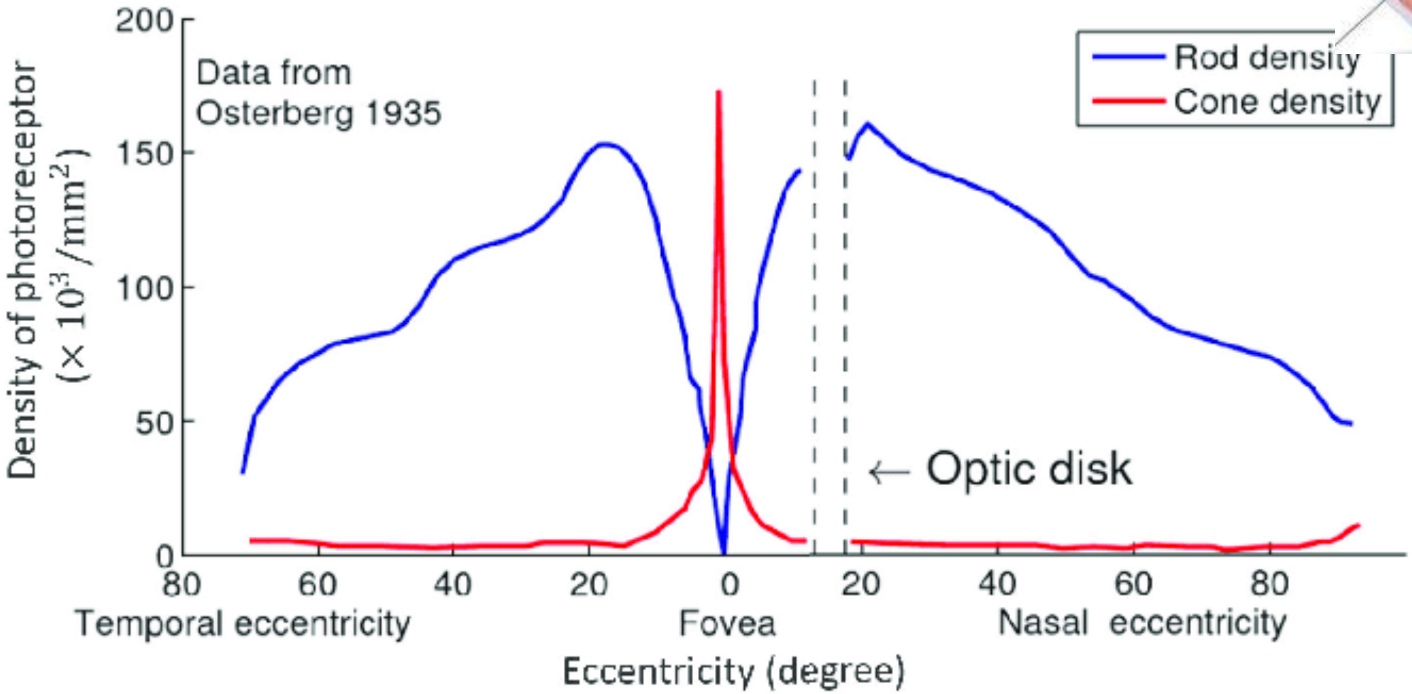
Recall: Neuron Layers in the Eyes



- Light appears to be moving in the wrong directions (passing through neurons before hitting photoreceptors)
- A result of inside-out retina: a hole punctured into retinas to pass optical nerves to brain
- Some bipolar neurons connect only to cones, other to both rids and cones
- An on bipolar activates when the rate of photons hitting its photoreceptors increases
- An off bipolar neuron activates with a decreasing photon absorption
- Human vision even at the lowest layers is not totally understood
- The well understood parts greatly contribute to the design of VR systems

Recall: Photoreceptor Density





- Cone density is highest at the fovea
  - Fovea has a diameter of only 0.5 mm (+/-0.85 degree angular range)
  - Eye must be pointed straight a target to see a sharp colored image
  - Corners have low cone density
    - We are good at detecting motion on the peripheries but not colors
- Blind spot is due to the canal through which eyes route neural signals to brain

Recall: Illusion of a Sharp Colorful Image

- If there is a blind spot and cones with narrow coverage, then why then do we perceive the world as a sharp colorful image with a wide angular range?
  - Frequent eye movements
  - Having two eyes
  - Perceptual processes created by our brain fills the missing details using contextual information and memory/history

Objective

- Transition from human eye physiology to visual perception
  - Analogy: from low level hardware to high-layer software and algorithms
- The goal of VR is to fool our visual stimuli by a display
  - Important to understand how we perceive the world around us
  - We have already seen some optical illusions
  - VR can be thought of as the grand illusion
    - When do we succeed or fail?

Outline

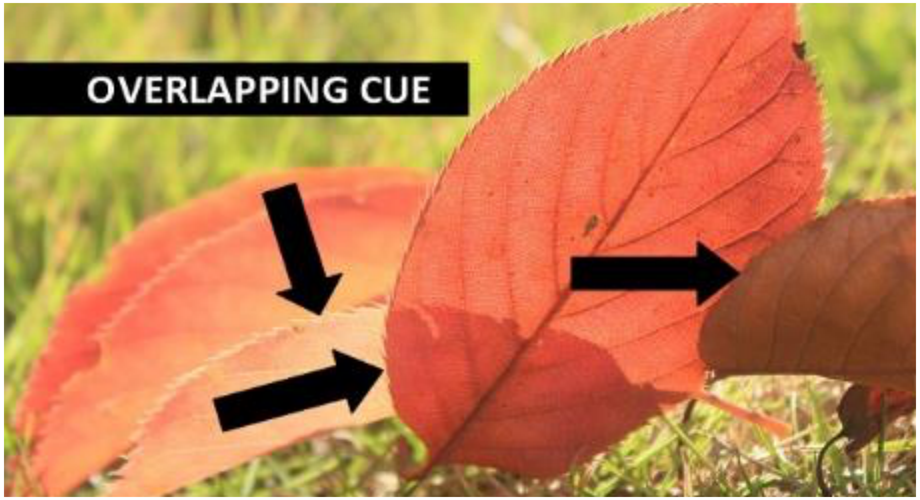
- Perception of distance of objects and object scale
- Motion perception
  - Perception of motion in videos
    - Already an illusion as video is just a sequence of frames

Outline

- Perception of colors
  - Why do we use only three colors to create spectrum of colors
- How brain combines information from multiple resources to produce a perceptual experience

Perception of Depth

- How do humans judge distance
  - In meters or in order (one object close than other)
- Depth/distance cues used by brain
  - Cue: a piece of information derived from sensory stimulation and used for perception
  - A depth cue by only a single eye movement is called monocular depth cue

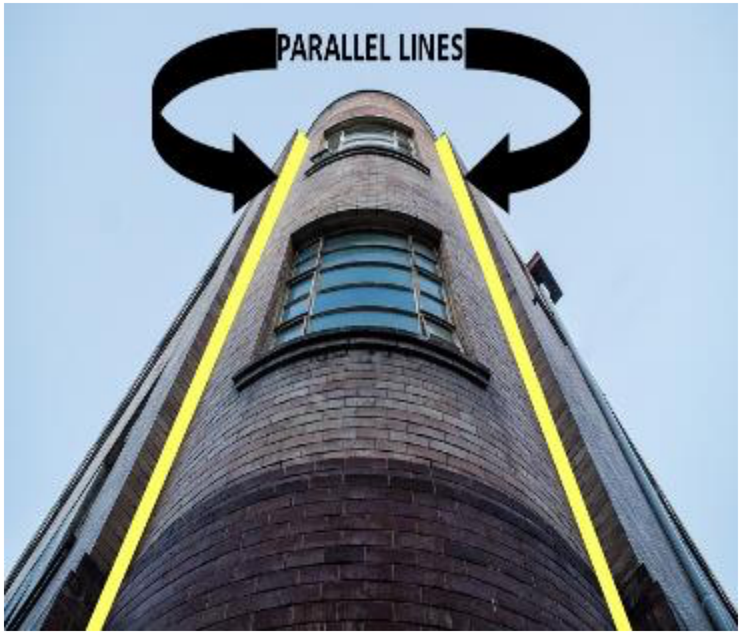


Stereo Depth Cues

- Stereo depth cue: both eyes are required
- Many more monocular depth cues than stereo depth cues
  - That is why we can infer so much information from a single image
  - The cues used by humans to extract depth information are also used by computer vision algorithms

Monocular Depth Cues: Linear Perspective Depth Cue

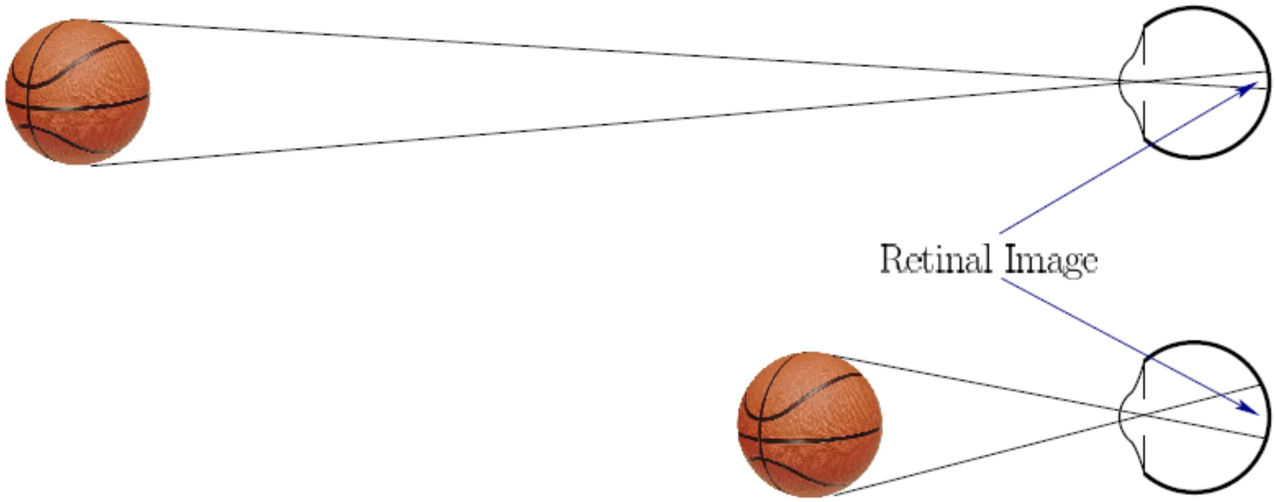
- While observing objects, you may see parallel lines. By evaluating how parallel lines converge in the distance w.r.t objects between lines, we can develop a perspective of objects' relative depth to one another



Ponzo illusion: both lines are the same length

Monocular Depth Cue: Retinal Image Size

- The further an object’s distance, the smaller the size of the image on the retina
  - Object appears naturally with no conflict with other depth cues
  - Viewer is familiar with an object size: car
    - Our brain assumes the object size is the same (called **size consistency**)
      - Therefore, we can judge the distance of the object



Moon Illusion

- **The moon illusion is an optical illusion**, which causes the moon to appear larger near the horizon than it does higher up in the sky
  - You can take photos at fixed location, which show that what your brain shows you is just an illusion

Reason not fully known, but perceived (vs real) retinal image size (or visual angle) could play a role!

Monocular Depth Cue: Height in the Visual Field

- The distance of objects from the horizon gives us another cue
  - The closer the object is to the horizon, the further the perceived distance

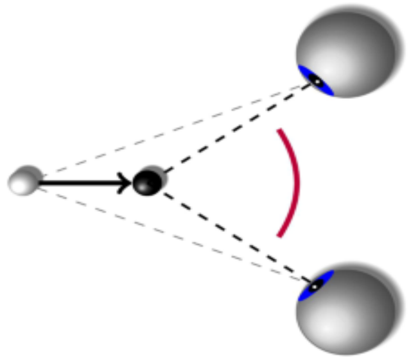
Other Monocular Cues

- There are many more monocular cues
  - Shadows, blur, contrast, …, all provide cues and help us perceive distance, depth, and other visual metrics
  - There are also cues that are **not given by the photoreceptors**
    - For example, our brain controls the focus of the lens in our eyes. The required tension level is fed to our brain, which helps with depth perception, among others.

Stereo Depth Cues

- Recall vergence: eye muscles are controlled to focus on an object (convergence vs divergence)

- Motor control of the eye muscles for vergence motions provides information to the brain about the amount of convergence, thereby providing a direct estimate of distance!
  - Note that each eye provides a different retina image
    - Called binocular disparity
  - However eyes rotate to fixate on an object, which reduces disparity



Implications for VR: Incorrect Scale Perception

- Need to get the scaling right for VR world generation. But there are many challenges
  - The VR world may not have many familiar objects to help with depth and size perception
  - Also, the VR world needs to take the interpupillary distance right
    - If the users' pupils are 64mm apart in the real world and only 50mm apart in the VR world, then the VR world may appear much larger
      - Recall chaining transformations for each eye!
- You can also play with different size and depth experiments, e.g., create a VR world where the user would look huge

Implications for VR: Mismatches

- In the real world, all depth (and other) cues work in harmony
- In a VR world it is easy to create mismatches
- If mismatch is undesired, then we should do our best to eliminate it
  - One source of mismatch could be delay in e.g., head tracking
    - Can cause the visual stimuli to not appear at the right place at the right time

Implications for VR: Monocular cues may be enough!

- We are bombarded by stereo and 3D displays for better depth perception
  - Stereo displays show images with offsets to left and right eye, so that our brain can create perception of depth
  - 3D movies give you glasses that essentially feed a different picture to each eye
- Drawback of feeding different left/right images: rendering cost increases
- You do not need stereo images to perceive the world as 3D

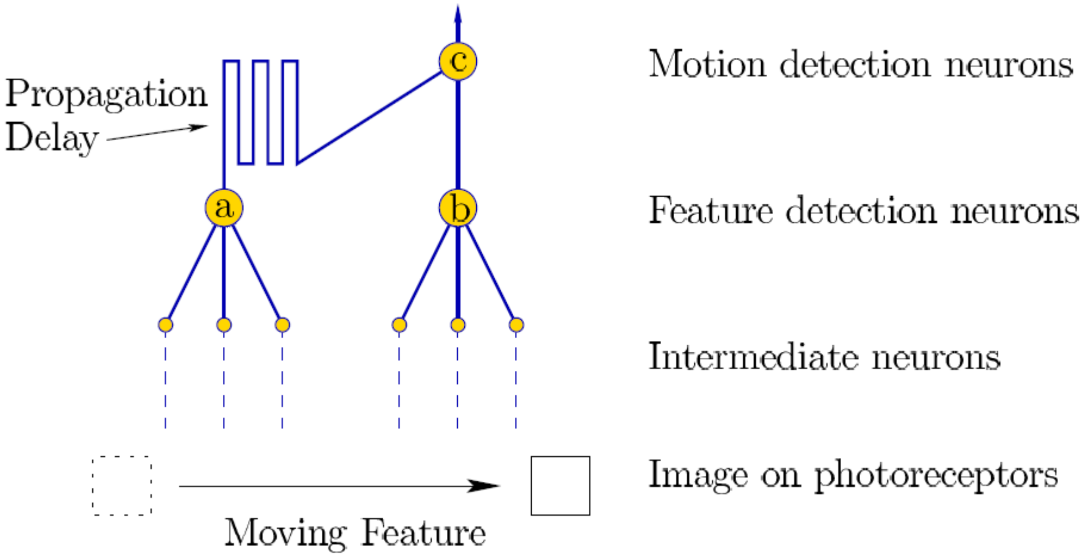
Depth Perception for VR

- Feed each eye a different image
  - Easy to do when the VR world in synthetic
    - The drawback is the higher rendering cost
- It is still possible to create immersive 3D experience , even when the same image is fed to both eyes
  - Capturing panoramic images and movies is fast and cheap
  - Smartphone apps can create panoramic images or videos or take shots which are panoramic
  - The images and videos are already 3D enough due to wide filed of view and sufficient monocular depth cues
  - Billions of images/video on Google and YouTube, which give you immersive 3D experience by using Google Cardboard and other VR headsets

Motion Perception

- We rely on our vision for motion perception, with many uses
  - Separate mobile object from stationary: animal in the forest
  - Assess 3D objects: rotate a fruit in market
- In the VR world, slight error in time or position of objects can give perception of motion

Reichardt Directional Motion Detector



Could be in Ganglion cells or Thalamus

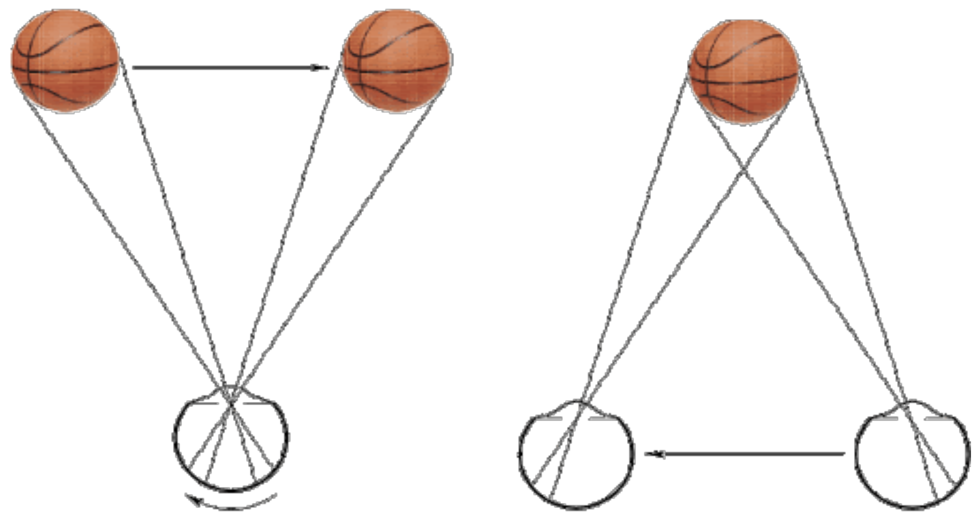
As the image features move across the retina, nearby feature detection neurons (labeled a and b) activate in succession. Their outputs connect to motion detection neurons (labeled c). A and b have different lengths/delays to c. Thus, c activates when the feature was detected by a slightly before b

The detector can be easily fooled! Based on the spacing between features and their speed, the detector may inadvertently fire (e.g., motion can be perceived in the opposite direction)!

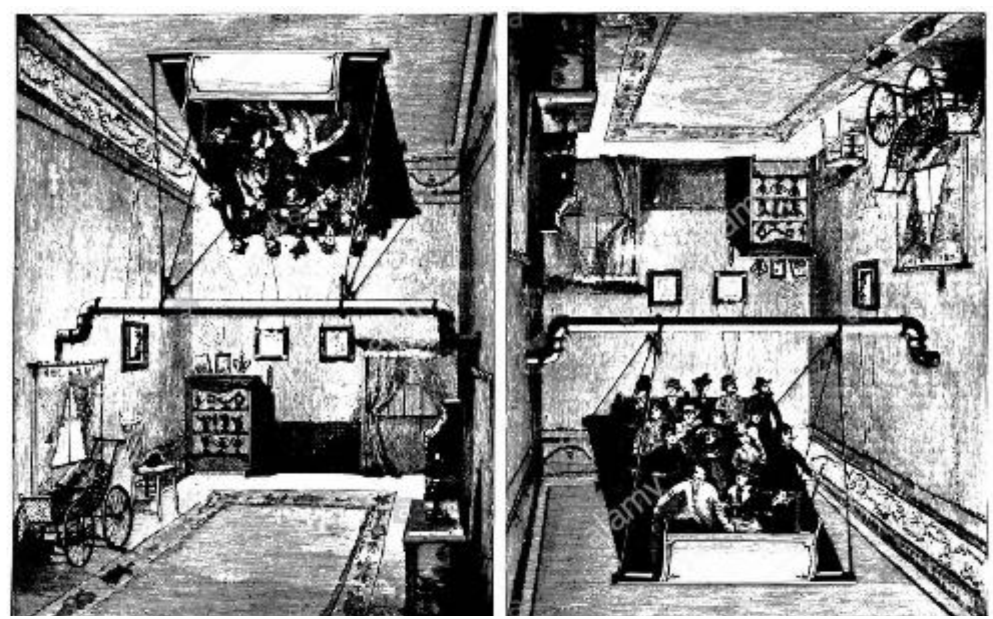
Wagon-Wheel Effect: A Wheel May Appear to be Rotating in the Opposite Direction

Distinguishing Object Motion from Observer Motion

- Brain uses several cues to distinguish between the two
  - Eye movement
  - Body motor signals
  - Large scale motion
- If the entire scene is moving, brain assumes the observer motion
  - Explains why haunted swing illusion is so effective

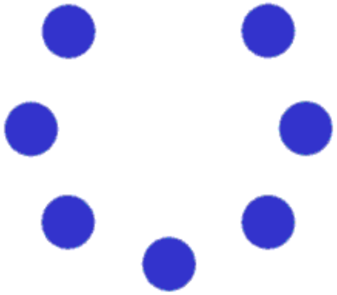


(a) (b)



Stroboscopic Apparent Motion

- Perceived motion through a sequence of frames
  - So it is an illusion: Zoetrope developed around 1834
  - The idea behind motion picture, TV, smartphone
- Why is the illusion so effective?
  - Earlier believed to be since visual cortex latches on
    - Somewhat true: images persist for 100msec
  - However, stroboscopic motions perceived even at 2 frame/sec
    - Example, sequence of blinking lights
  - Believed to be because it triggers the neural motion detector



Perception of Motion

24: Sound and frame needed to be synchronized

FPS	Occurrence
2	Stroboscopic apparent motion starts
10	Ability to distinguish individual frames is lost
16	Old home movies; early silent films
24	Hollywood classic standard
25	PAL television before interlacing
30	NTSC television before interlacing
48	Two-blade shutter; proposed new Hollywood standard
50	Interlaced PAL television
60	Interlaced NTSC television; perceived flicker in some displays
72	Three-blade shutter; minimum CRT refresh rate for comfort
90	Modern VR headsets; no more discomfort from flicker

Black in between frames, caused flicker

72: People sat closer to monitors

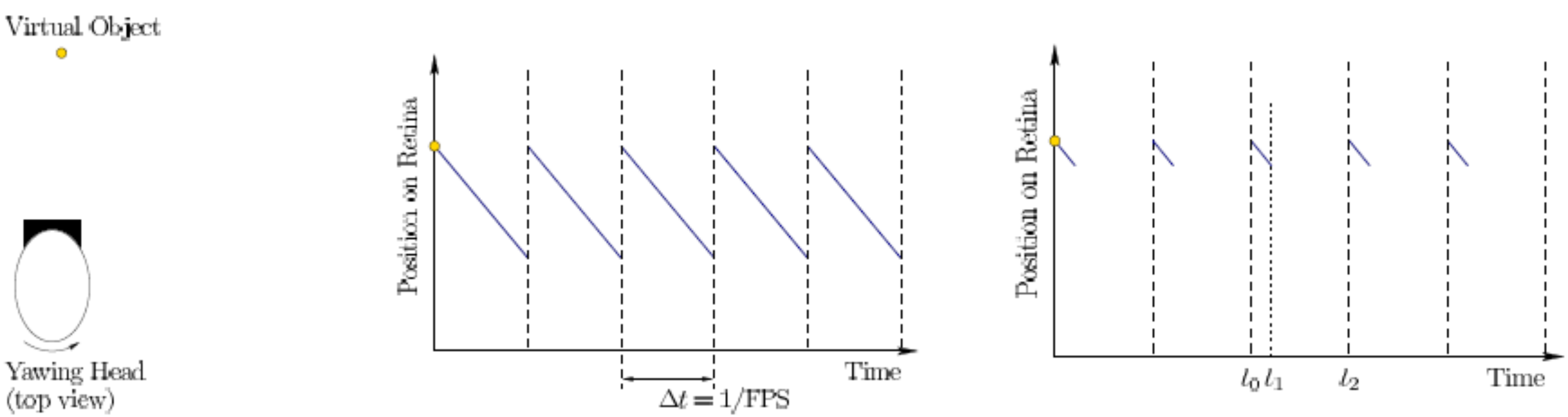
Modern LCD and LED displays in smartphones, TVs, monitors support 60, 120, and even 240 FPS

Figure 6.17: Various frame rates and comments on the corresponding stroboscopic apparent motion. Units are in Frames Per Second (FPS).

2 or 3 blade: show a single picture across 2 or 3 frames
Interlacing: show half of the image in one frame, and the other half in another

Implications for VR

- Higher FPS needed compared to movie standards of 24 FPS or even 60 FPS
- Image fixating on a virtual object and moving your head
  - The image on VR screen needs to shift as you rotate your head
  - The problem is that the position of object slips on retina
    - At 60 FPS, object position is fixed for 16.67 ms
  - Engineering hack at 60 FPS:
    - Keep the image for 2 ms and then turn off
    - Photoreceptors can catch the image and brain can perceive
    - This is called low-persistence
      - 60 FPS low-persistence causes flicker; goes away at 90 FPS



Display Selection for VR

- VR requires fast pixel switching
- LCD Display
  - As high as 20ms delay to reach desired intensity level
    - Can blur the virtual object
    - Problem observed in Oculus Rift DK1
- Modern OLED display
  - Pixel reaches its maximum intensity in 0.1ms

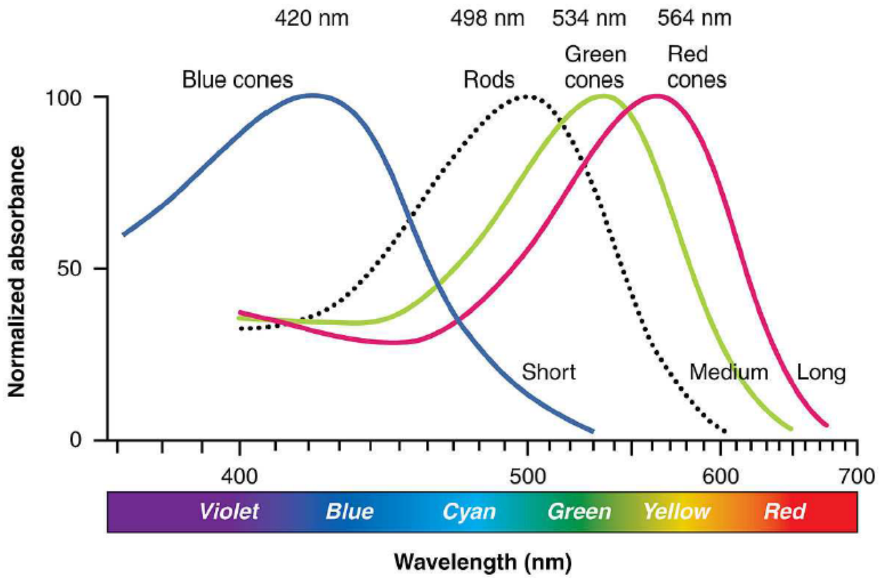
Perception of Color

- Color perception is purely result of our visual physiology and neural structures
  - In other words, it's all in your head
- This contrasts other perception topics such as motion, depth, and scale
  - These can be measured through instruments

Color illusion: based on the precise combination of colors and lighting conditions, its appearance fell on the boundary of what human color perceptual system can handle

Color Dimensionality Reduction

- Light energy is a jumble of infinite wavelengths
  - Spectrum of wavelengths with different magnitude values
- The color we see, is based on the wavelengths the object reflects
- The rod and cones have a bias towards certain wavelengths
  - Three essentially: red, green, and blue
  - Dimensionality reduction from infinite to three
  - Our RGB displays target the same colors

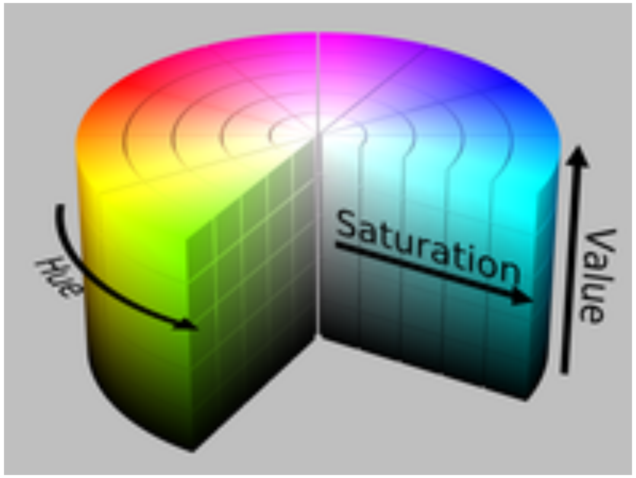


So What is Yellow Color?

- Suppose we use laser to create the precise light wavelength of yellow
  - Our eyes do not have a yellow photoreceptor
  - The wavelength causes some reception by rods and cones
  - Same effect can be achieved by an appropriate selection of red and green wavelengths
    - You can create the same photoreceptor response

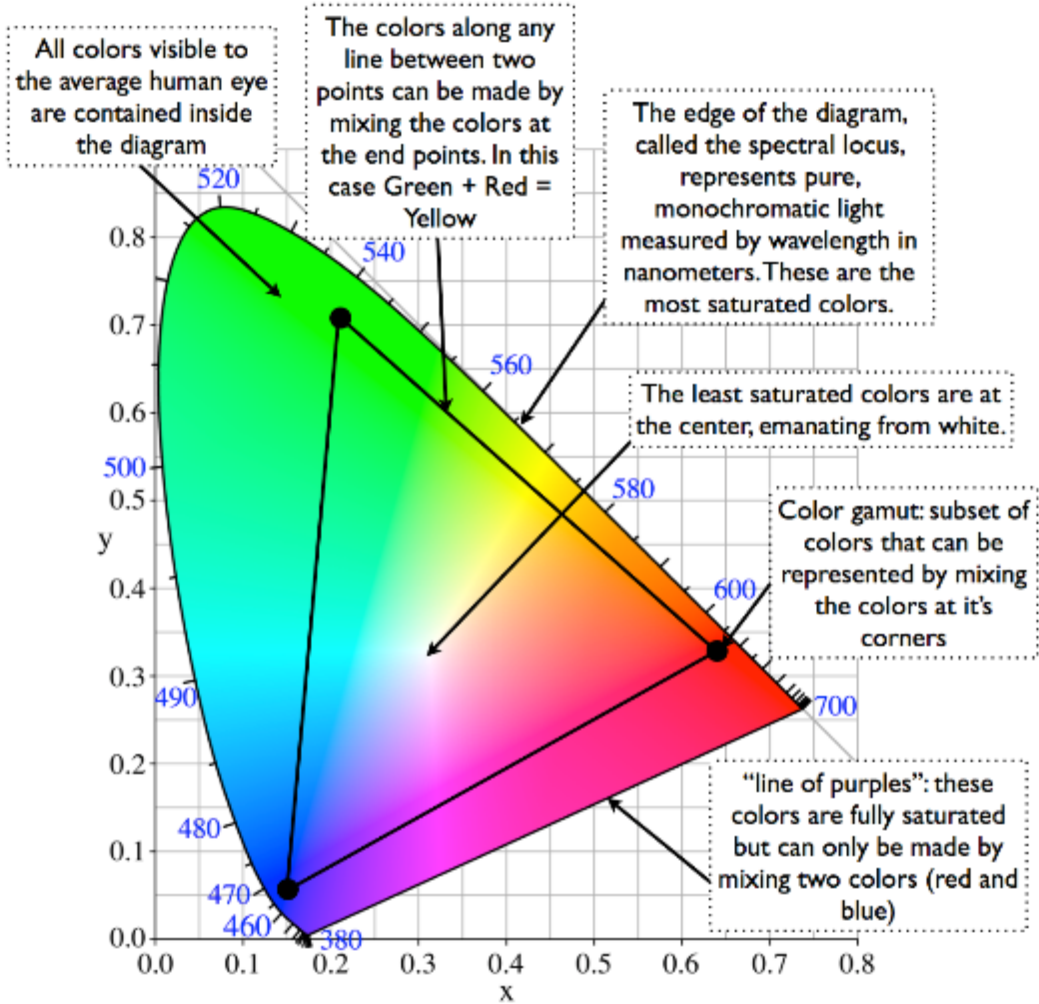
Color Representation in Computer Graphics

- HSV is commonly used in CG
  - A parameterized color space
  - Hue: corresponds directly to the perceived color
  - Saturation: purity of the color. How much energy is coming from wavelengths other than the hue wavelength
  - Value corresponds to brightness



CIE: Color Standard Used by Vision Scientists

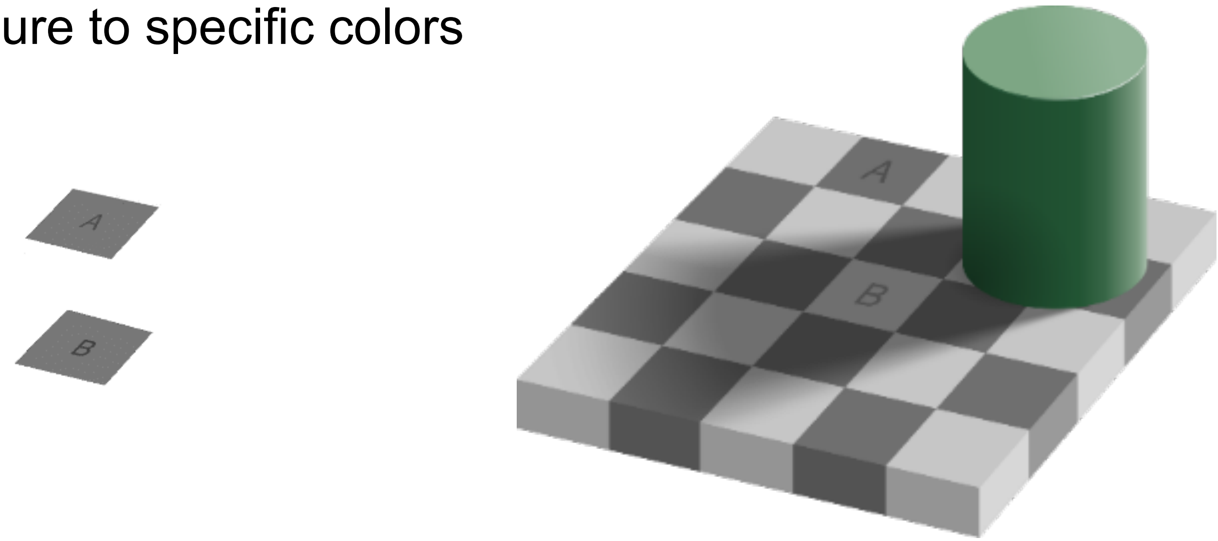
- The distance between two points corresponds directly to the amount of distinguishability
  - Vision scientists designed CIE standard to achieve this
- Is 2D: Does not have brightness
  - Which is independent of color perception



Anatomy of a CIE Chromaticity Diagram

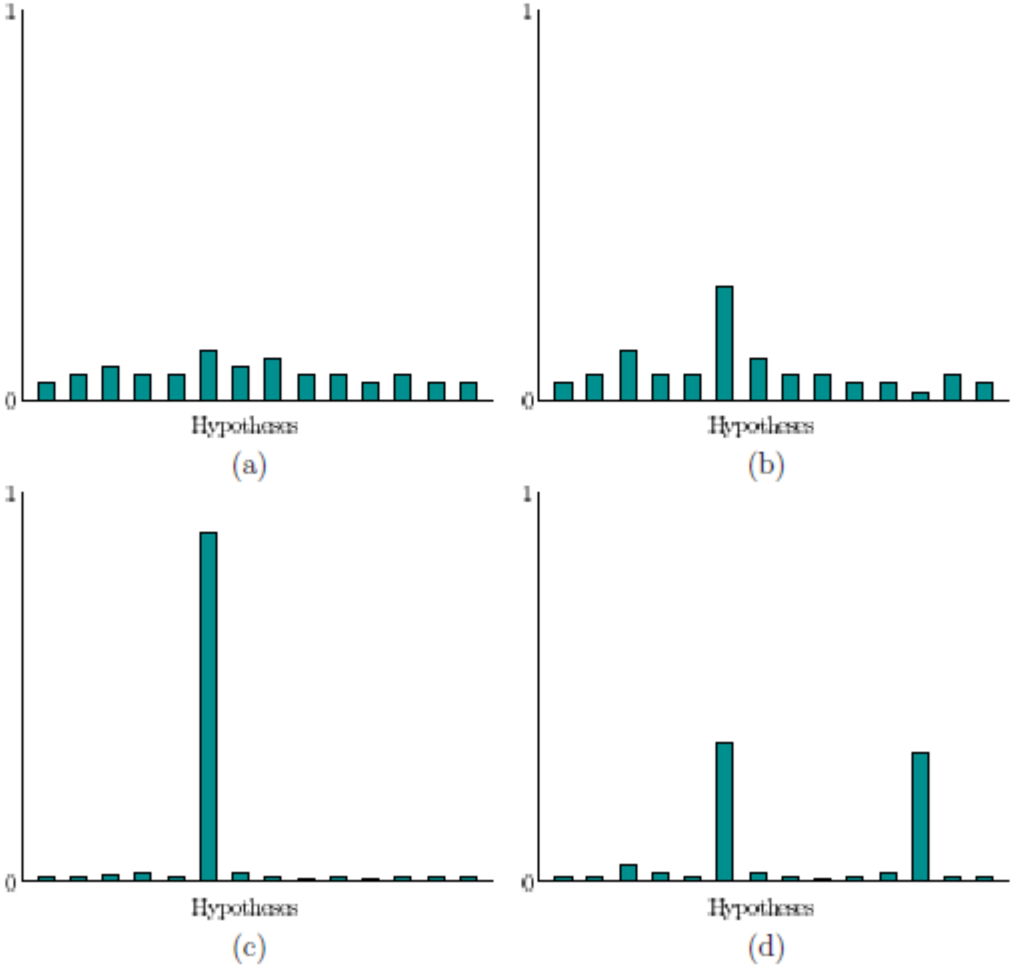
Constancy

- Ordinarily, human color perception is robust to the source of color
  - A red shirt appears red whether in sunlight or in a room at night
  - Correspond to vastly different lighting conditions
- **Our ability to perceive an object as having the same color under vastly different lighting conditions is called color constancy**
- Many issues affect our perception of object lightness and color
  - Memory about how objects are colored in the environment
  - Expectations from colors of surrounding objects
  - Prolonged exposure to specific colors



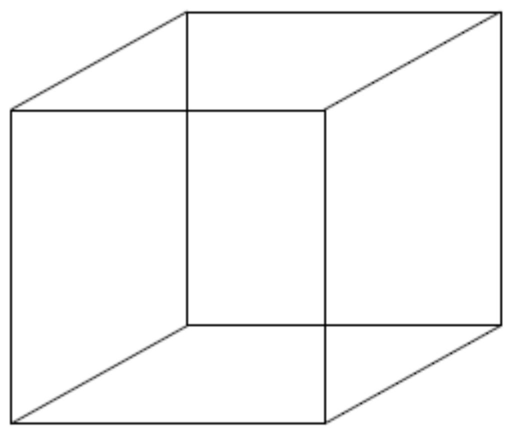
Combining Multiple Sources of Information

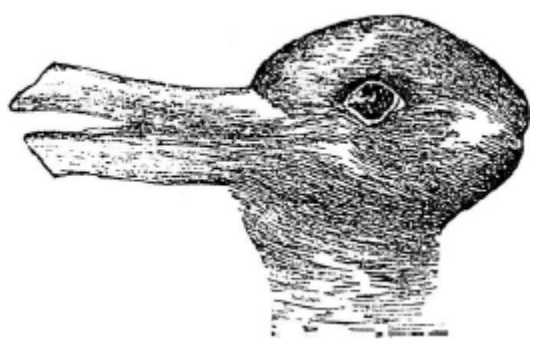
- Our perception system works by combining multiple cues
  - We also take into account prior cues
- Let Hypothesis be for example face of a person as we see the image
  - Cues could be for example, color of the eye, etc.
  - Ambiguity arises if two or more hypotheses have the same probabilities



Multistable Perception

- Our perception system may alternate between two or more conclusions
  - Which side of the cube is at the front
  - Picture of a dock or rabbit
  - What we see first is highly impacted by our prior biases





Implications for VR

- Not all senses are taken over by VR, which naturally creates mismatch between the real and virtual worlds
  - Vection: sickness causing conflict between visual and vestibular cues rising from apparent self-motion in VR while stationary
  - In addition to mismatch among senses, imperfections in VR hardware, software, and content cause inconsistencies
- Can also cause many multi-stable perception scenarios
  - Example, a menu popping up in VR and rushing towards you
    - The menu coming to user or vice versa
    - The visional stimuli is very overpowering
      - Can cause extreme fatigue and sickness