

Virtual and Augmented Reality:

A Technology Tutorial

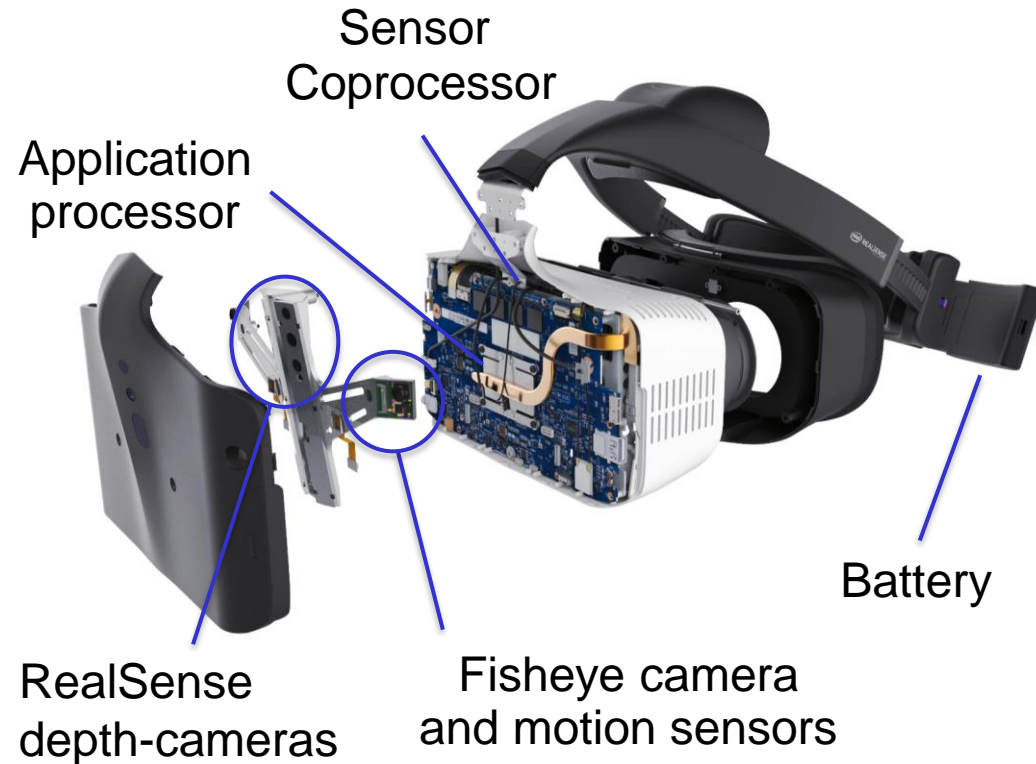
Achin Bhowmik, PhD

CTO & EVP of Engineering, Starkey Hearing Technologies

Adjunct Professor, Stanford University School of Medicine



From My Oct '16 BA-SID Presentation



Untethered, all-in-one design

- Multi-room scale mobility, no setup

Intel® RealSense™ Technology

- Inside-out 6DOF tracking
- Hand tracking & natural interactions
- Merged reality experiences

High-performance computing


- PC-class processor, graphics, computer vision & artificial intelligence accelerators

Virtual & Augmented Reality: Objectives

- Place users in simulated virtual environments, and virtual objects in the real-world.
- Generate sensory stimuli (visual, auditory, vestibular, haptic, ...) that create the illusion of “presence” and “immersion”.
- Ensure consistency of perceptual cues between the physical and digital elements, and enable natural interactions.

Ultimate Goal:

Merge the “Real” and the “Virtual” worlds,
so they are indistinguishable from one another!

A man in a dark blue suit, white shirt, and red tie stands with his hands on his hips in a server room. He has a skeptical expression. In the background, there are computer monitors displaying data on desks under blue and white lighting.

Virtual reality?



Thanks For Watching....!!

Movie Fan

Johnny English Strikes Again

The background of the image is a close-up, slightly blurred shot of Morpheus from the movie The Matrix. He is looking directly at the camera with a serious expression. His face is the central focus, with his eyes and nose clearly visible. The lighting is dramatic, with strong highlights and shadows. The text is overlaid on this image.

“What is real?”

“How do you define '**real**'?”

“If you're talking about what you can **feel**,
what you can **smell**, what you can **taste** and **see**,
then '**real**' is simply **electrical signals**
interpreted by your brain.”

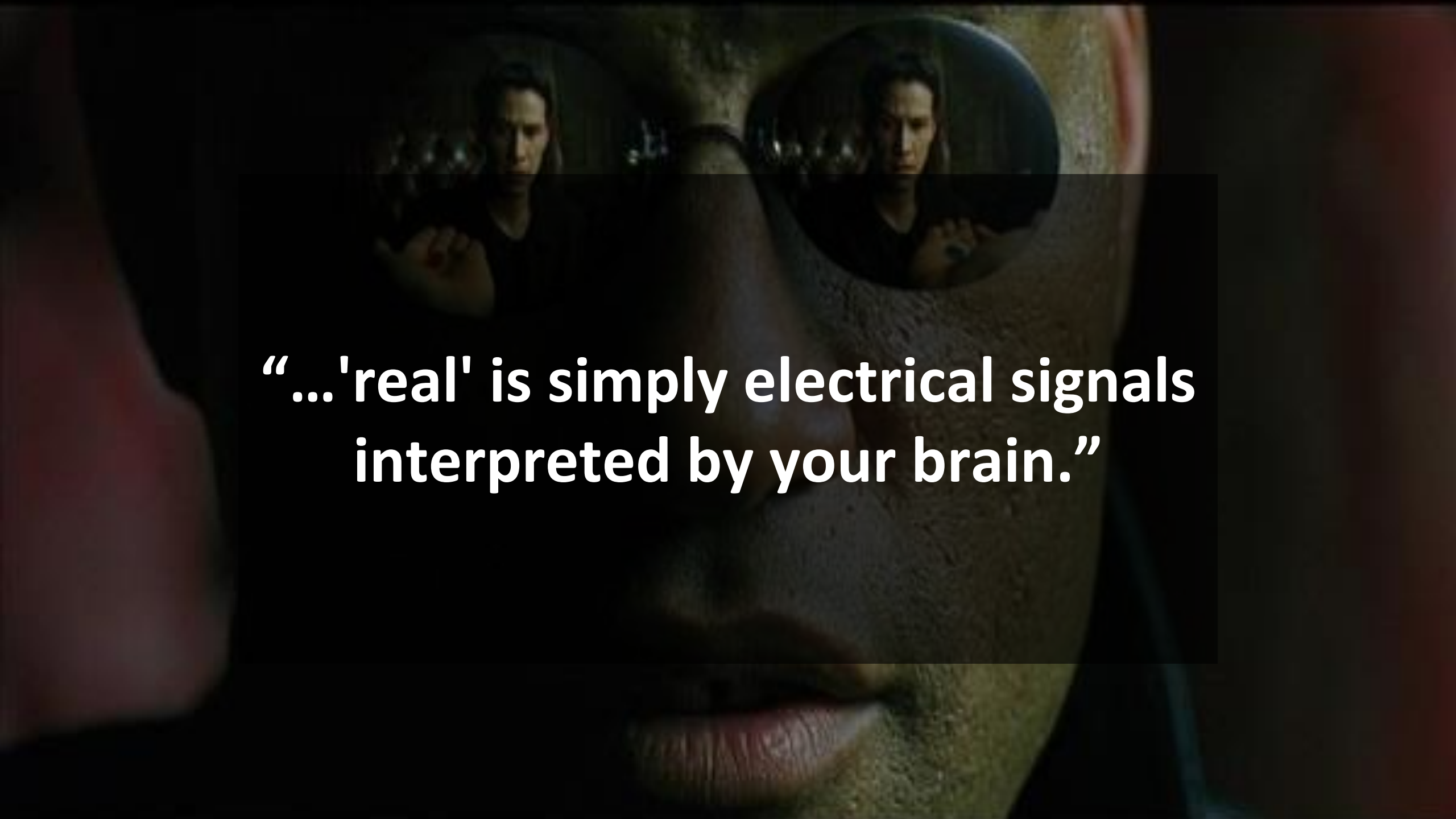
- *Morpheus in Matrix (1999)*









A close-up, artistic photograph of a person's face, focusing on the eyes. The person has dark hair and is wearing a dark top. The eyes are highly reflective, showing a clear reflection of the person's own face and upper body. The lighting is dramatic, with strong highlights and deep shadows, creating a moody and contemplative atmosphere. The skin appears slightly textured and has a warm, golden-brown tone.

**“... 'real' is simply electrical signals
interpreted by your brain.”**

Human Factors Considerations: Sensory & Perception System

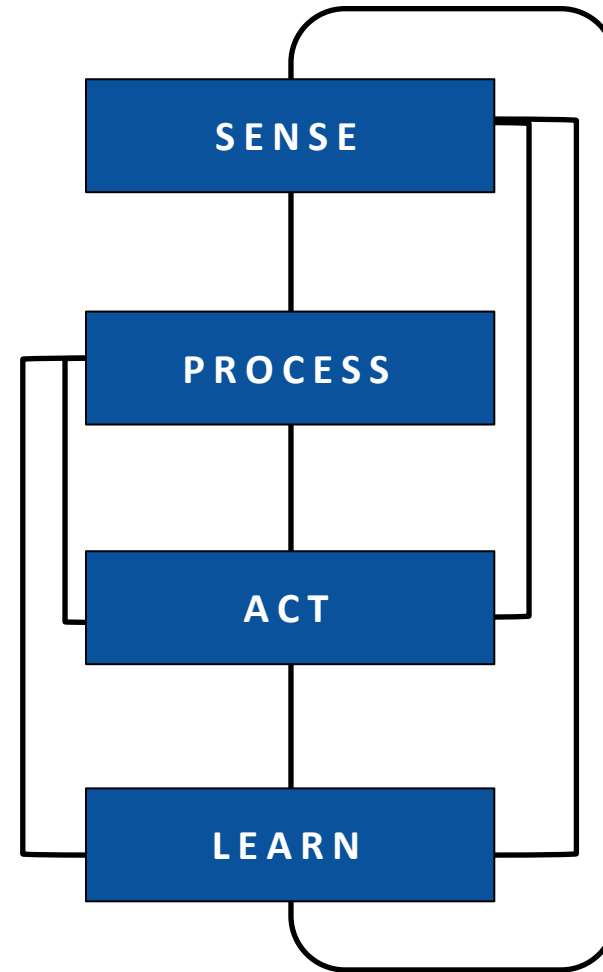
(In)famous “User Experience” Issues in VR & AR

- Inconsistent perceptual cues from the visual and vestibular sensory systems
 - Motion that is “felt” but not “seen”, or “seen” but not “felt”
 - Both systems detect and estimate motion, but they do not correspond
- Visual experiences
 - “Screen-door” effects due to “low” display resolutions
 - Field-of-view (FOV) limitations
- Inconsistent oculomotor cues
 - Eye accommodation and convergence mismatch
 - Incorrect focus and blur cues

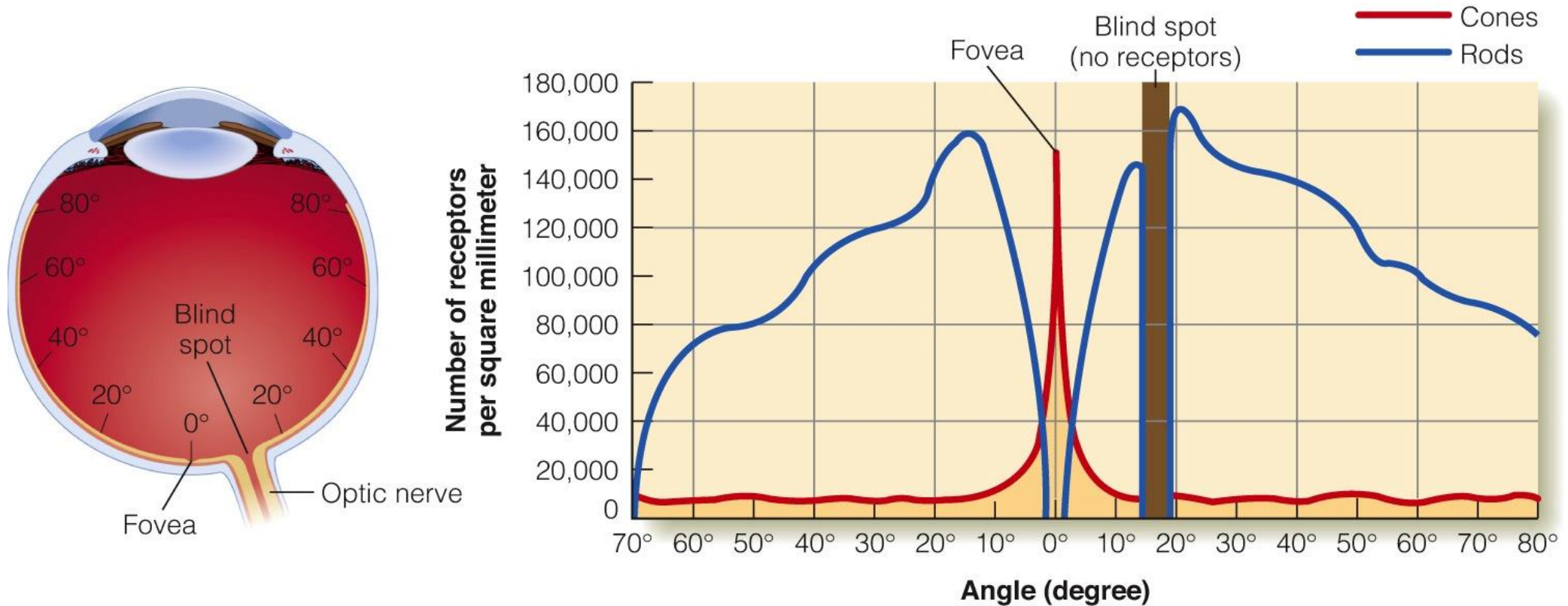
(In)famous “User Experience” Issues in VR & AR

- Systems and ergonomic issues
 - Latencies, weight, comfort, tether, eye relief, power dissipation, battery life, ...
 - Tracking setup vs. “inside-out” tracking, 3-DOF vs. 6-DOF
- Missing or inconsistent proprioception cues
- Other perceptual issues we will inevitably discover in the future...

Perceptual Computing

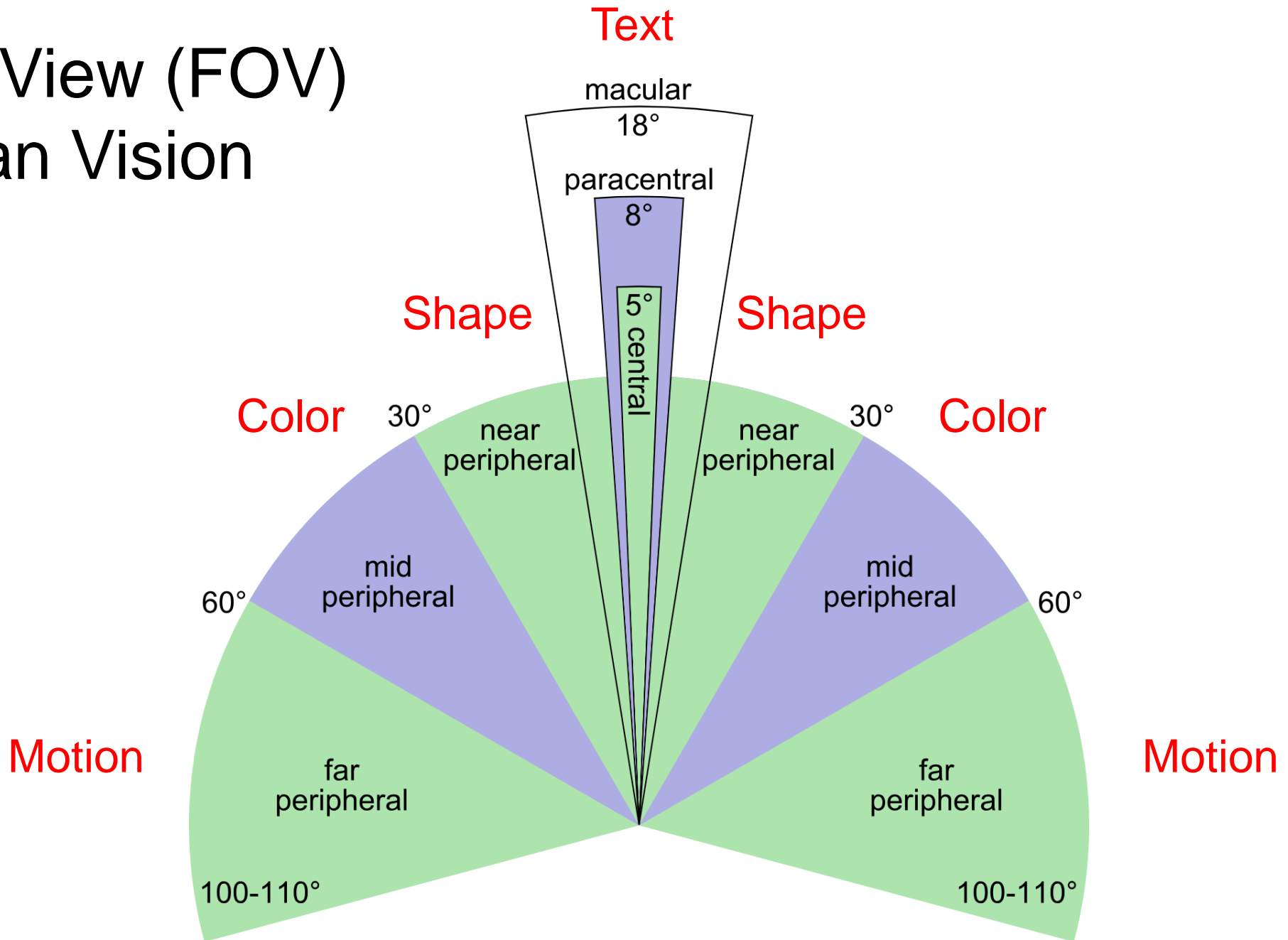


Construction & “Resolution” of the Human Eye



Visual Acuity for Foveal Vision: $\sim 1/60^{\text{th}}$ of a Degree

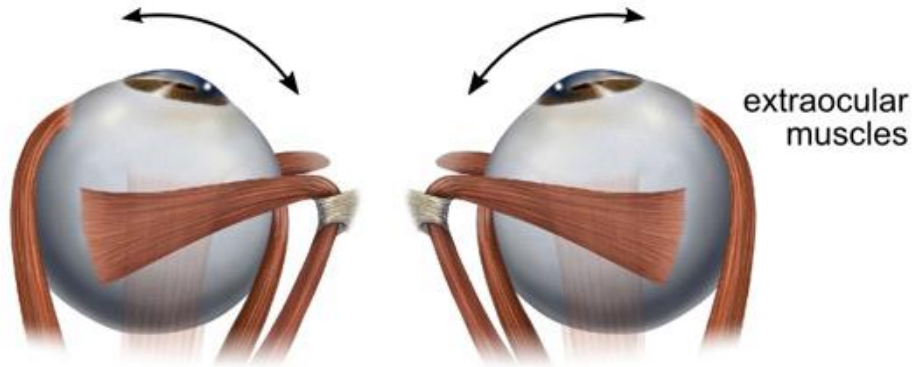
Field of View (FOV) of Human Vision





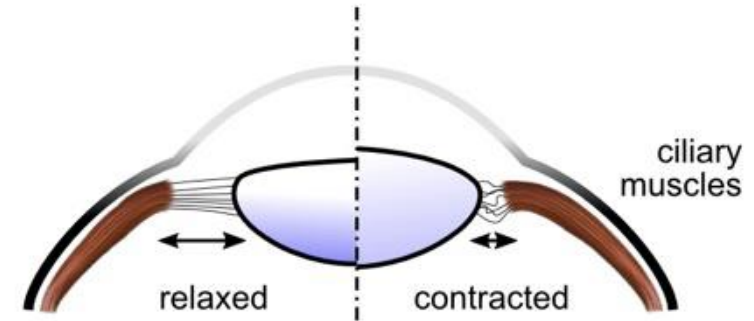
Oculomotor Cue

Stereopsis (Binocular)



Vergence

Focus Cues (Monocular)



Accommodation

Visual Cue

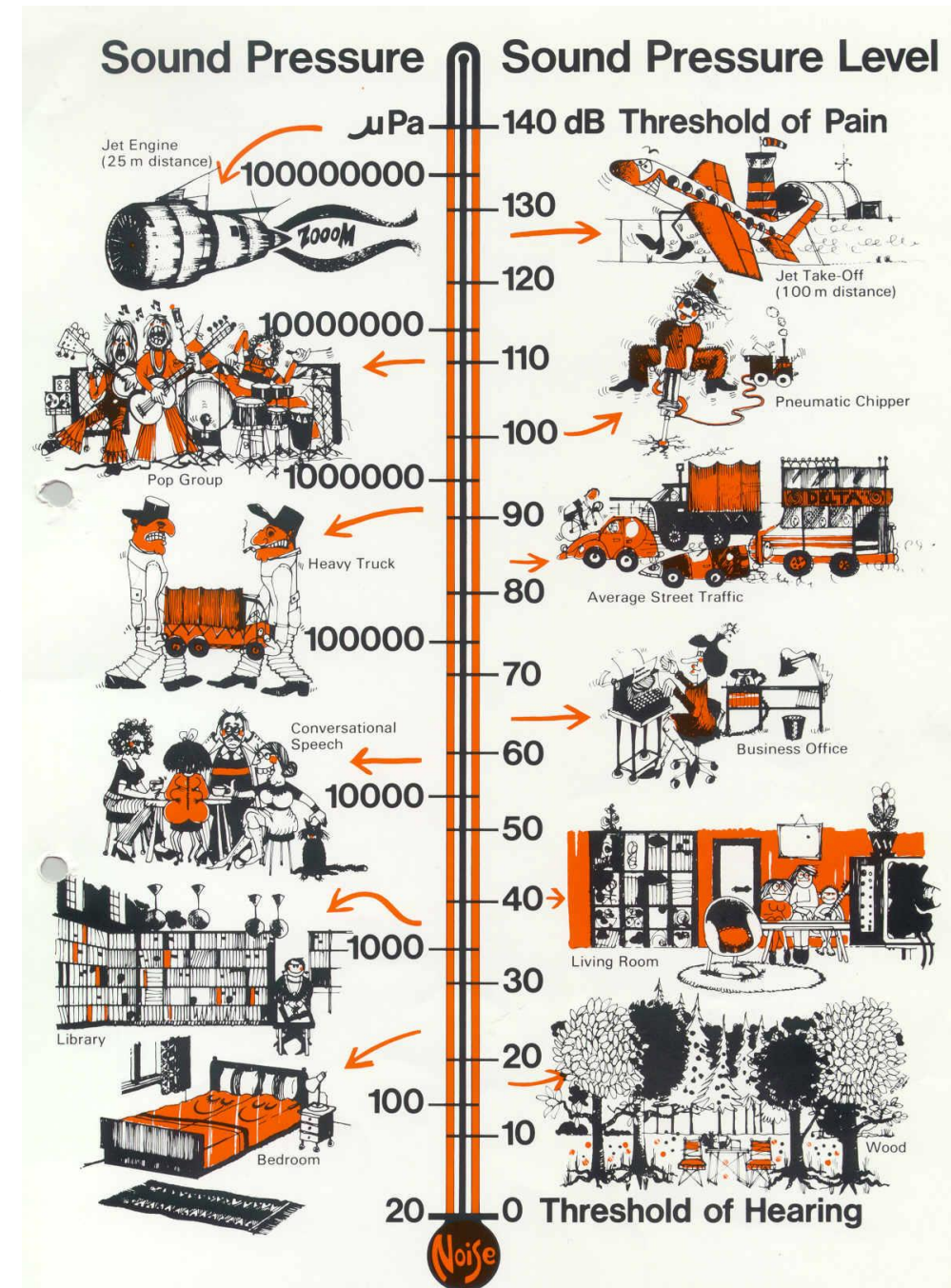
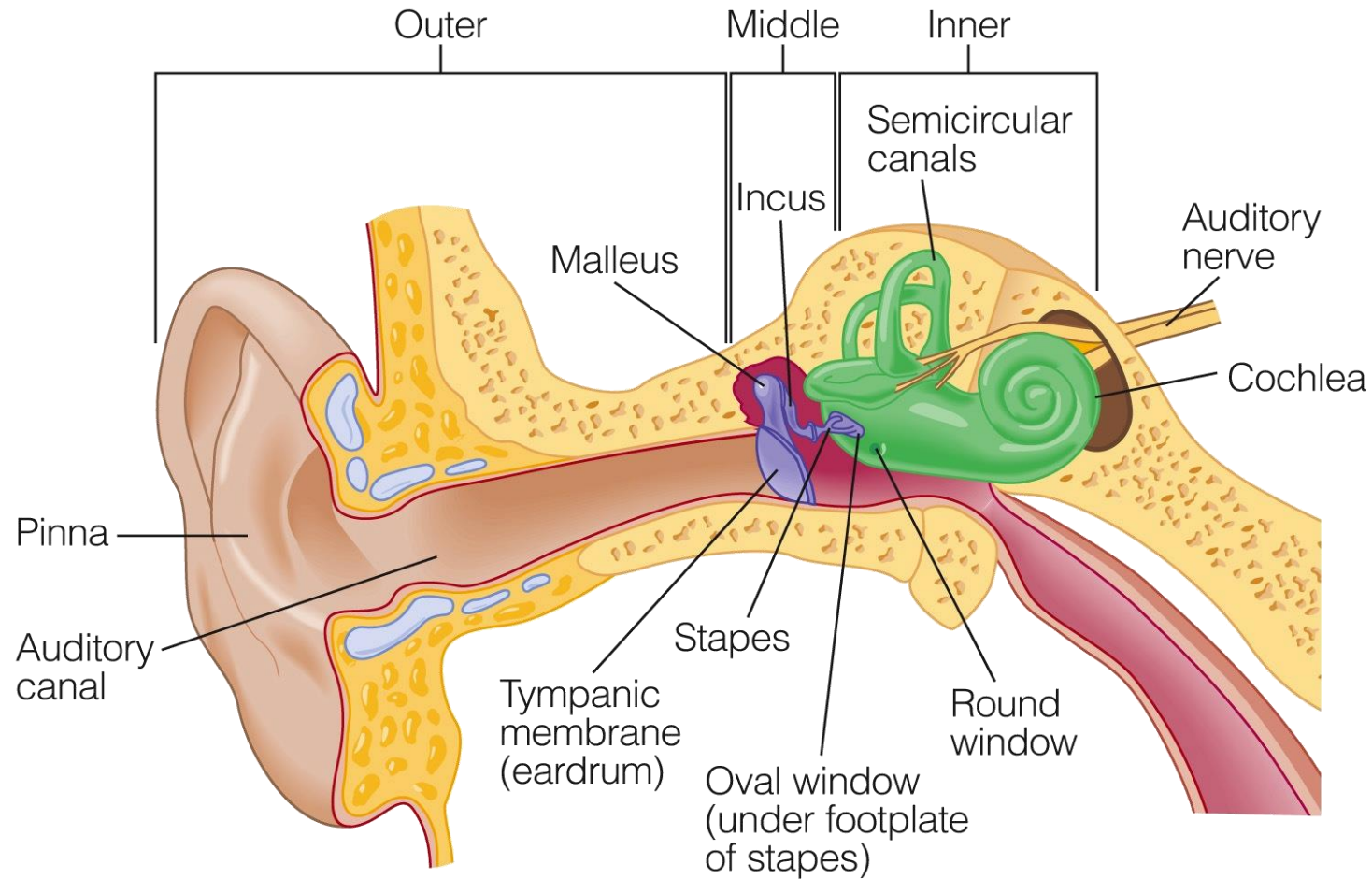


Binocular Disparity



Retinal Blur

The Ear & Sound Transduction



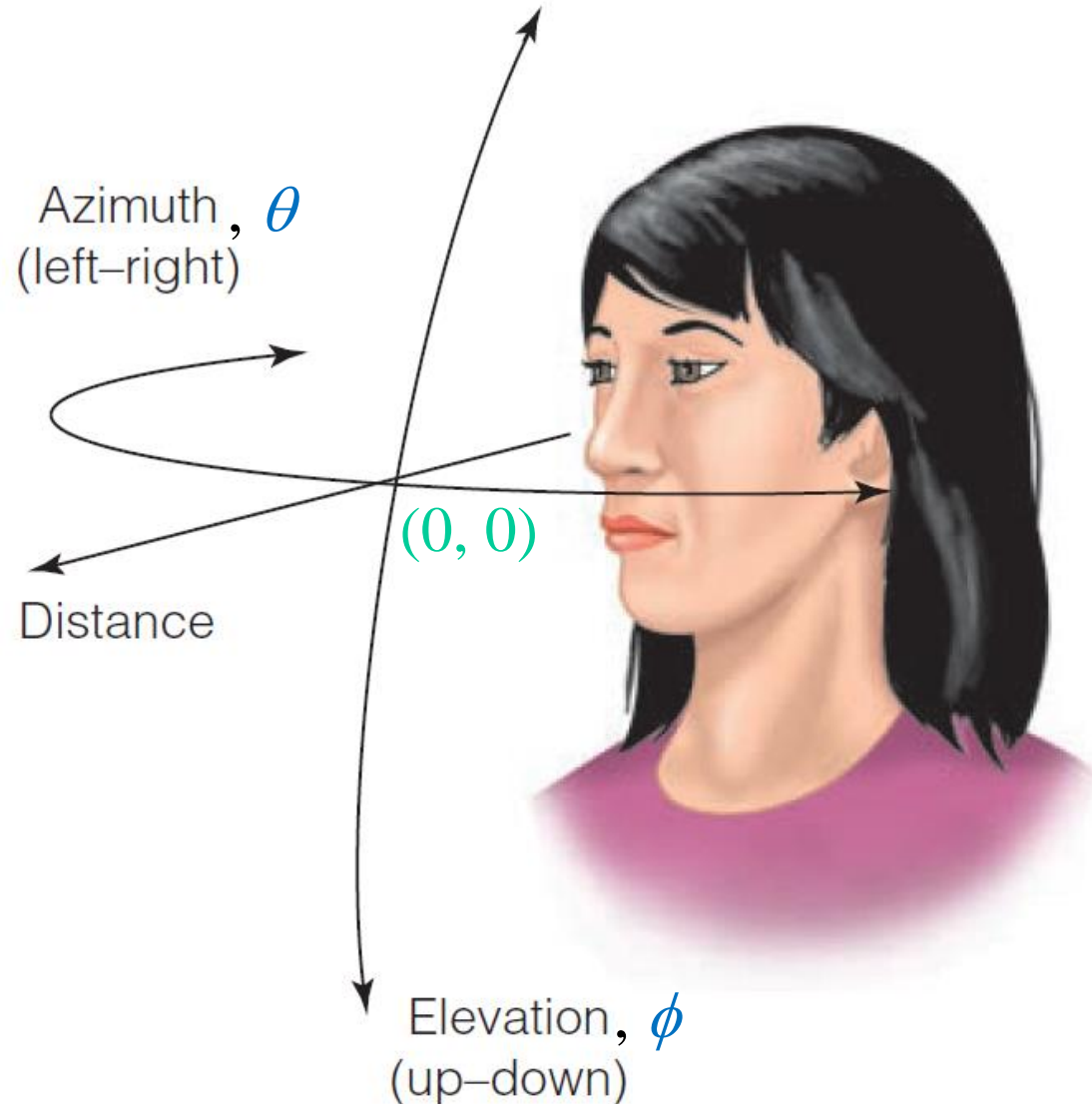
Auditory Localization

Binaural Cues:

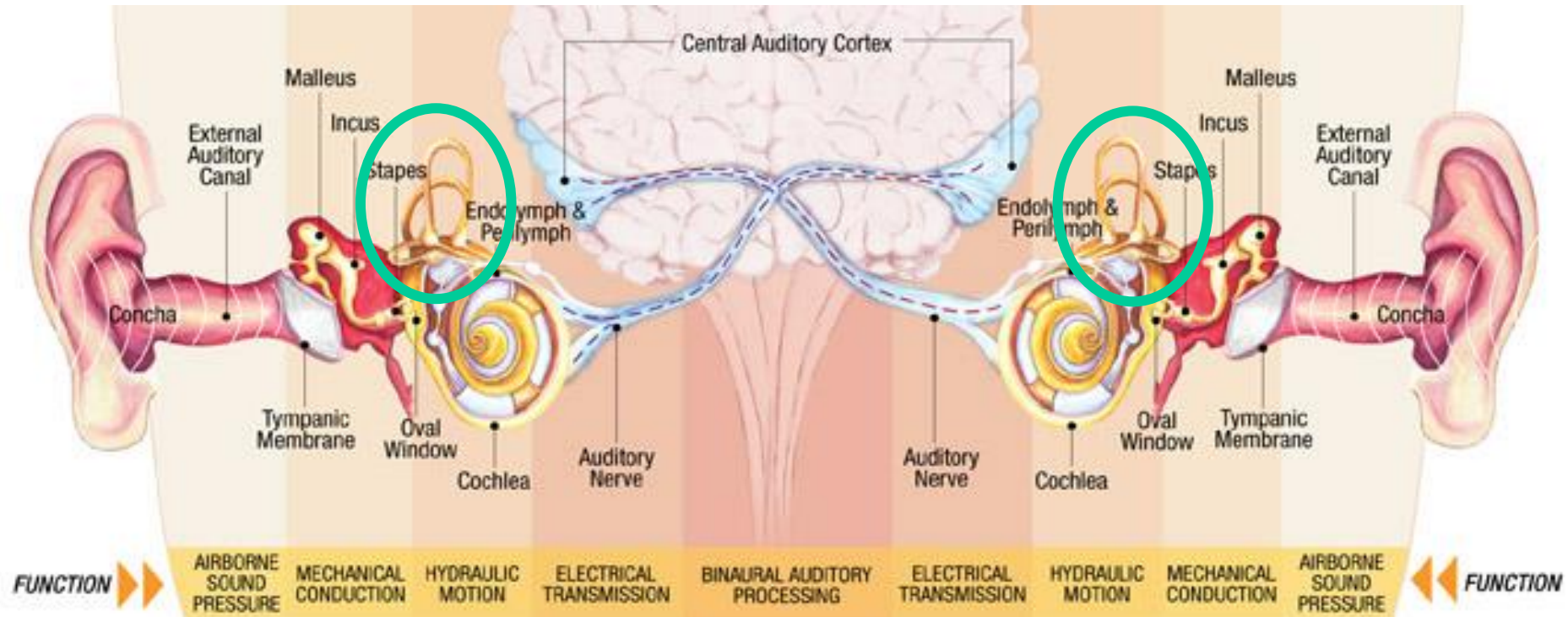
- Interaural Time Difference (ITD)
- Interaural Intensity Difference (IID)

Monoaural Cues

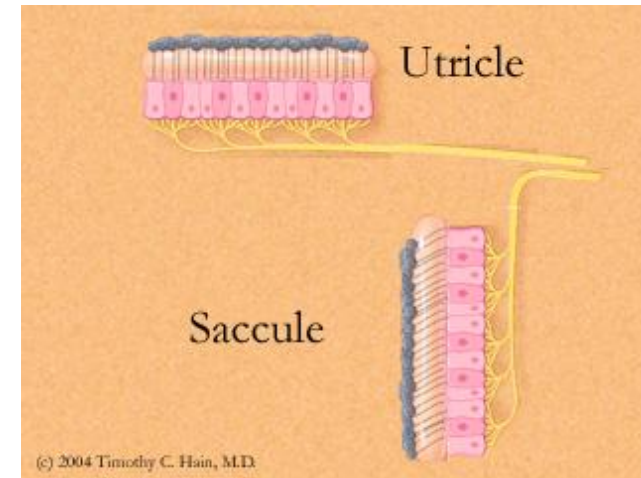
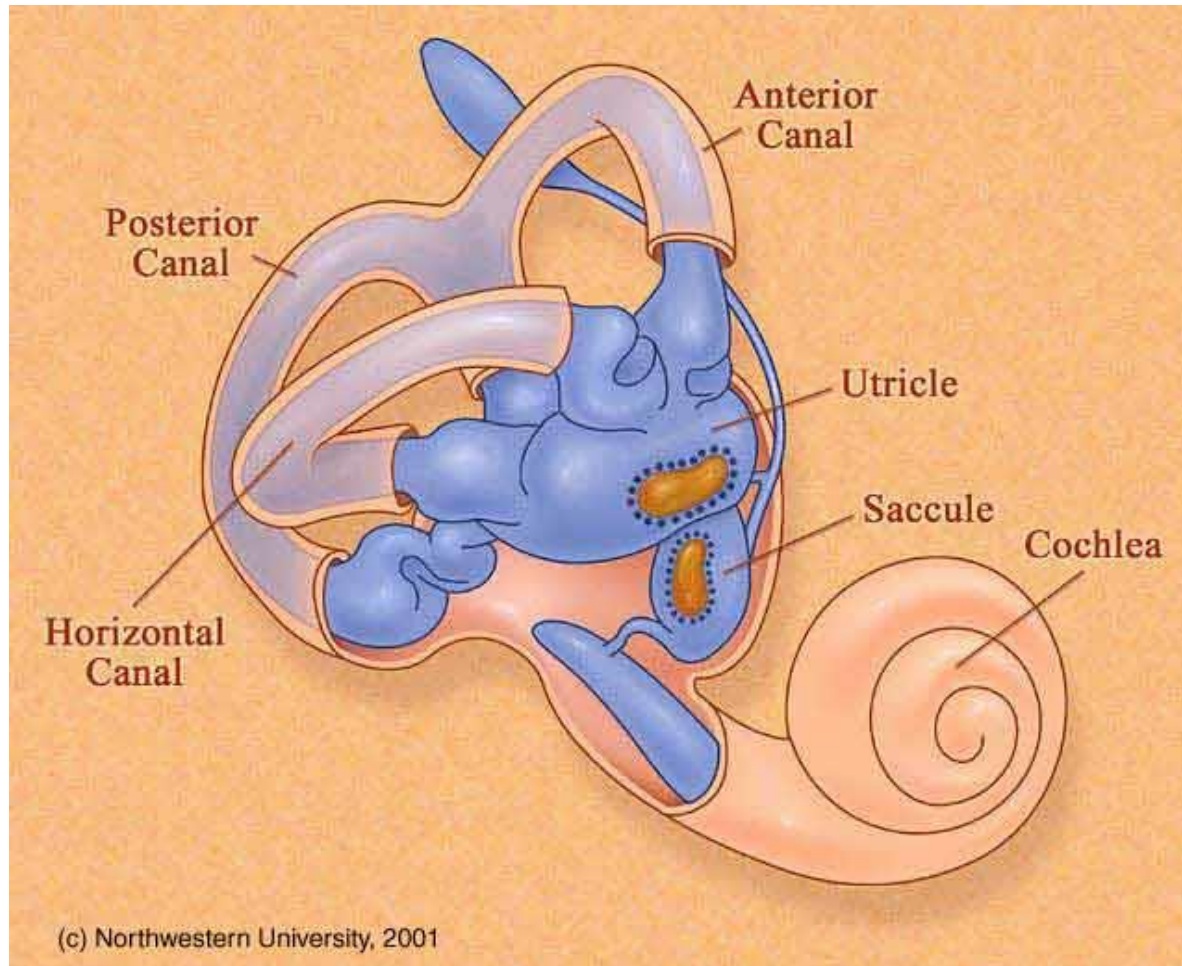
- Spatial-dependence of the sound frequency spectra (filtering by the outer ear and head)



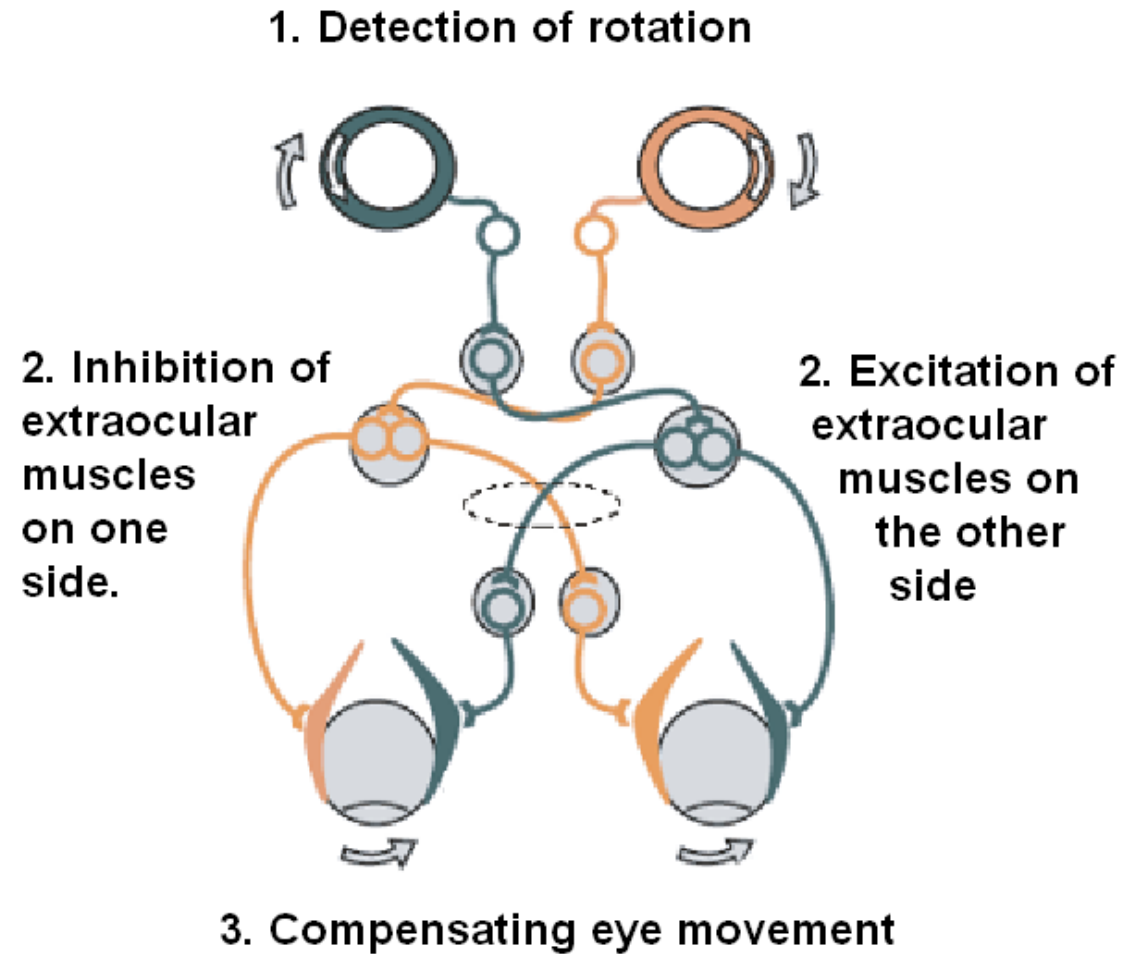
Human Vestibular System



Human Vestibular System

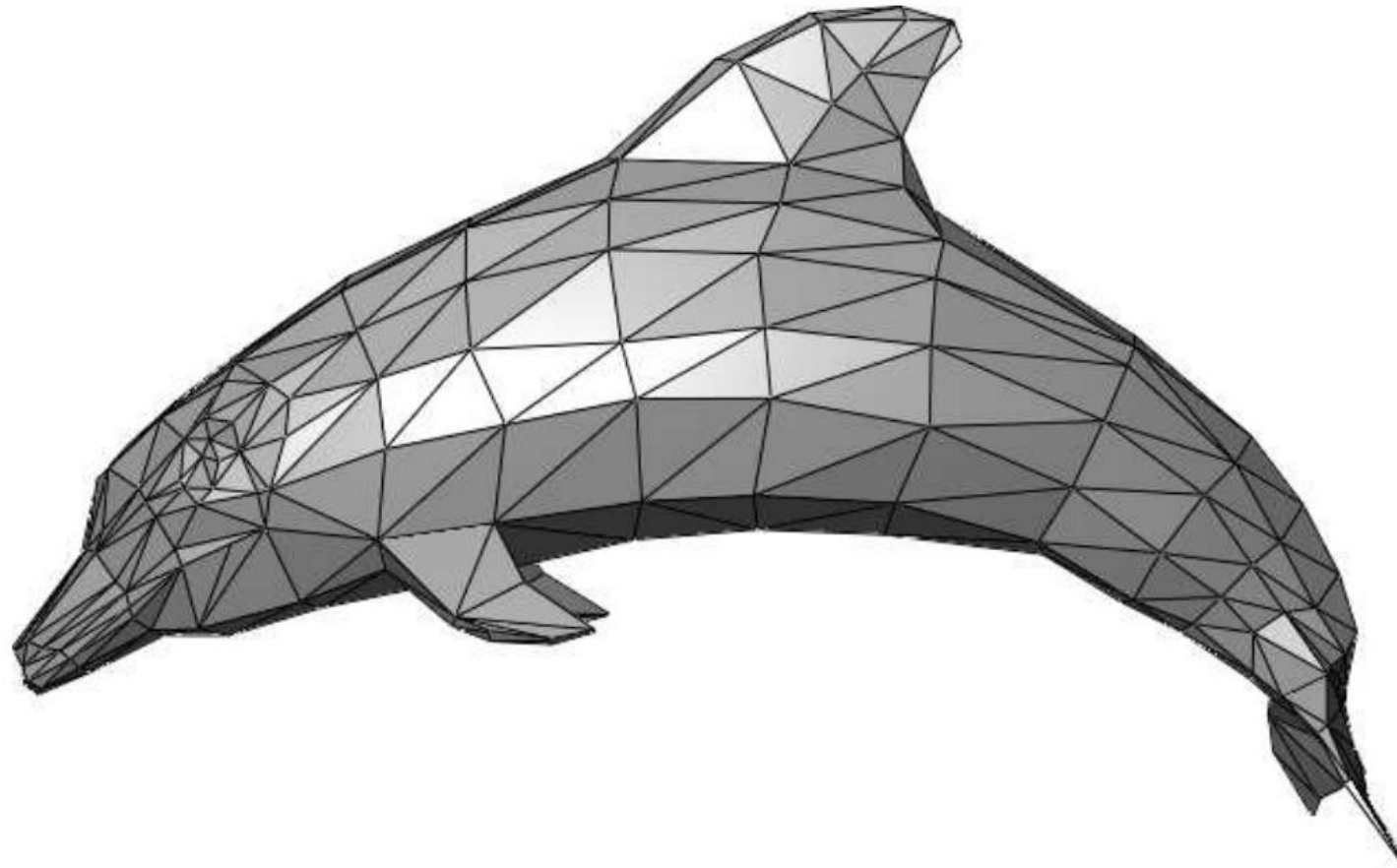


Vestibulo-Ocular Reflex

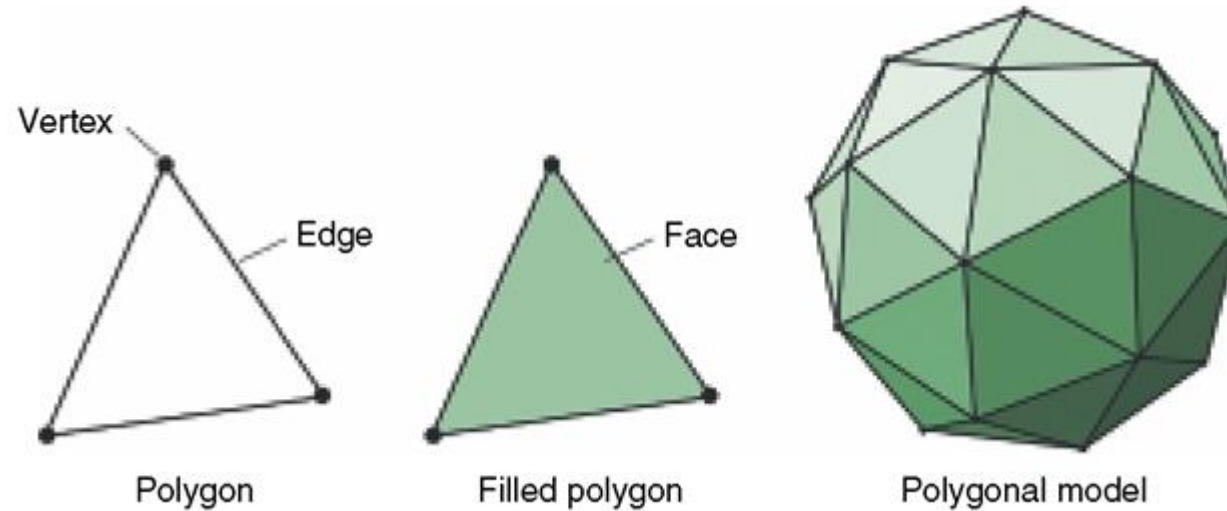


Virtual World Creation: 3D Graphics/Computation for VR/AR

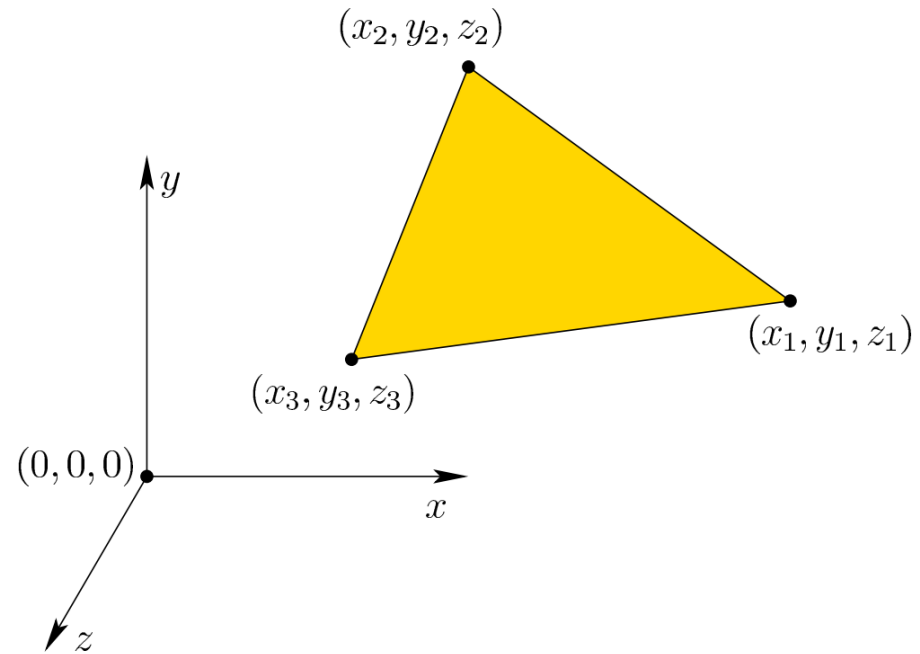
Geometric Models in VR/AR/MR: Mesh of 3D Primitives



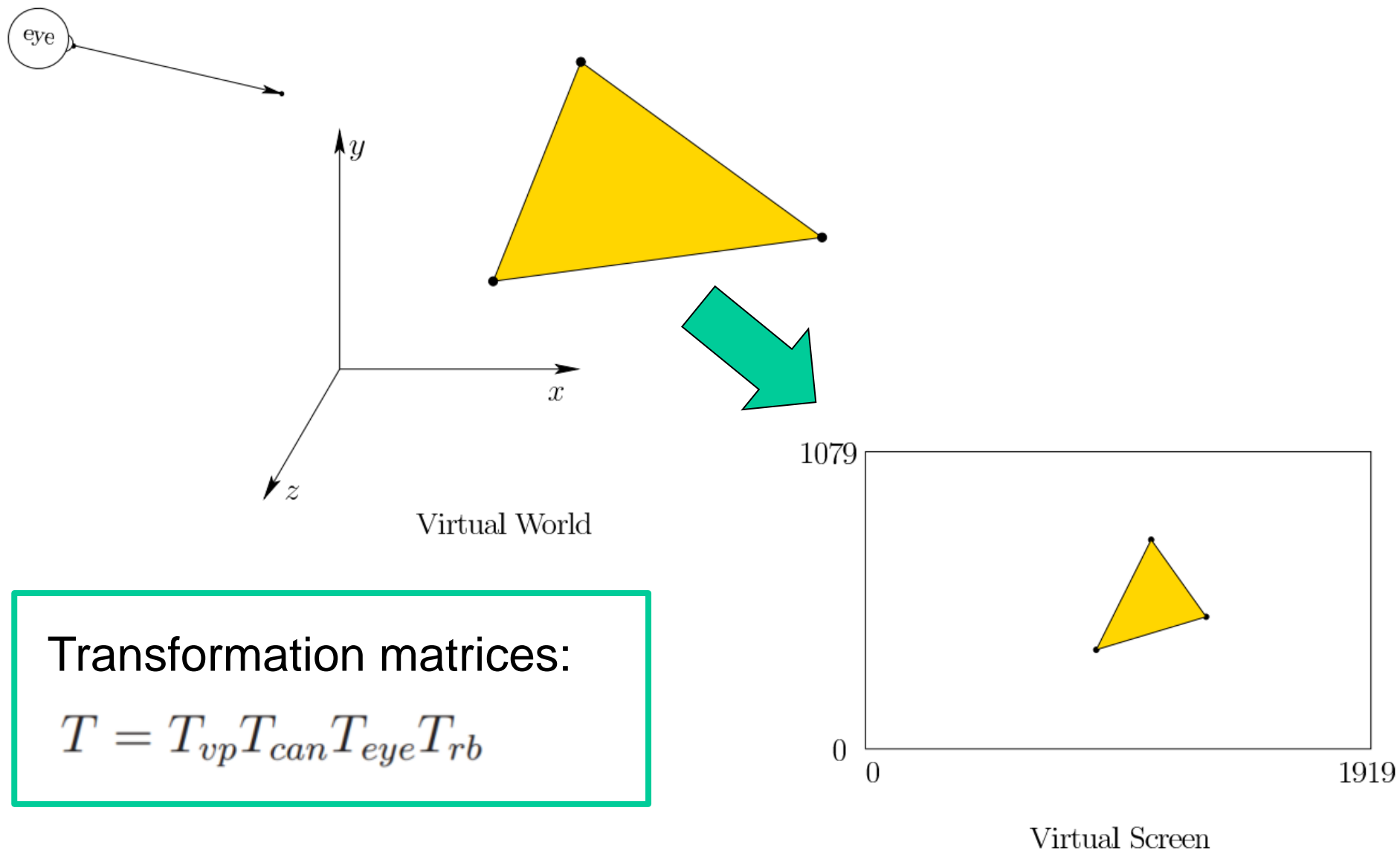
Geometric Models in VR/AR/MR: Mesh of 3D Primitives



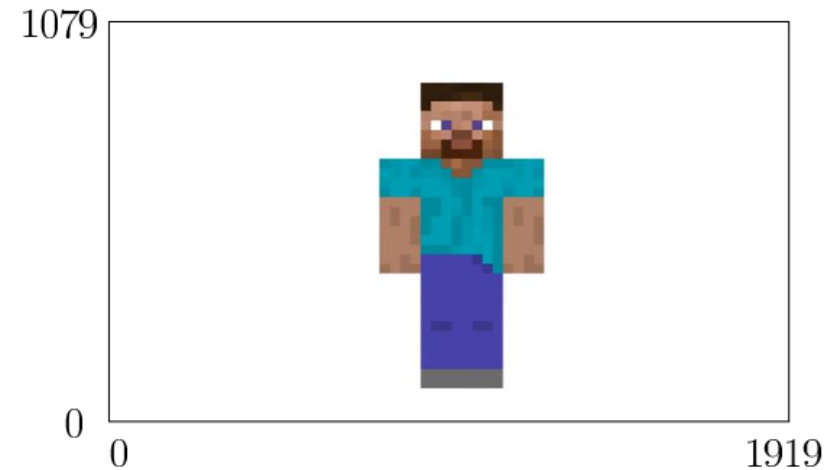
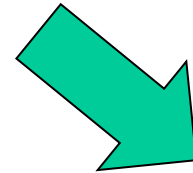
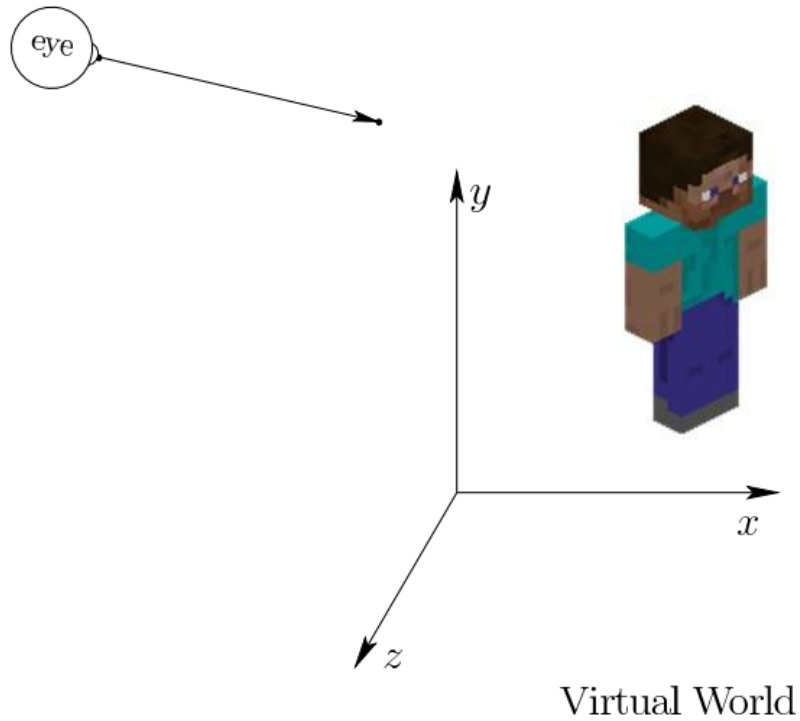
3D Primitive



Viewing Transformations



Viewing Transformations

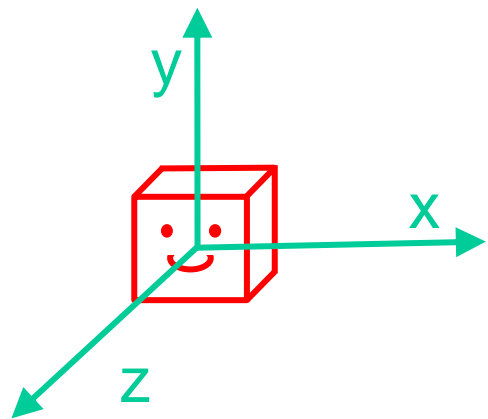


Transformation matrices:

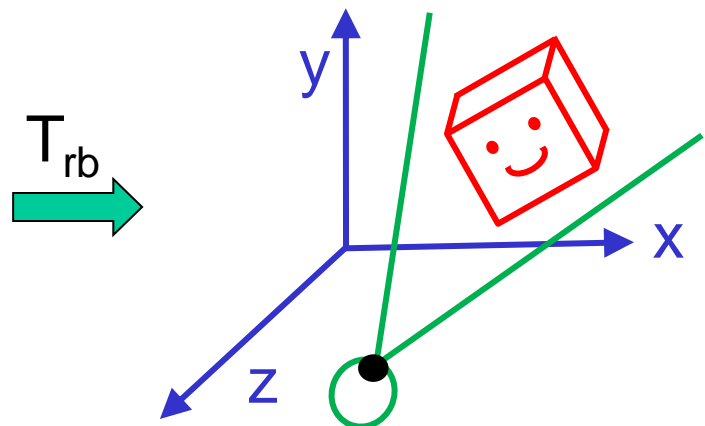
$$T = T_{vp}T_{can}T_{eye}T_{rb}$$

Virtual Screen

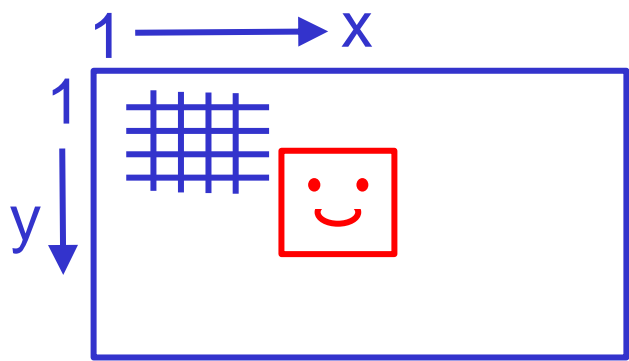
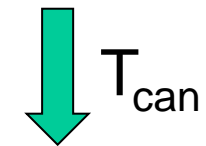
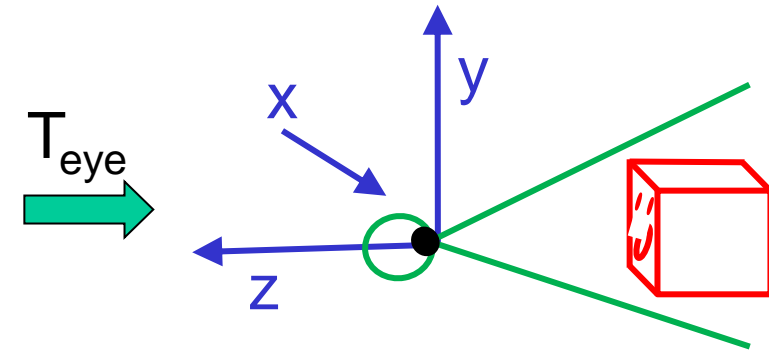
Object Frame



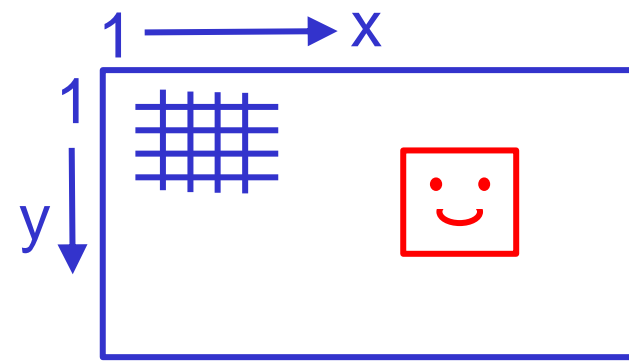
World Frame



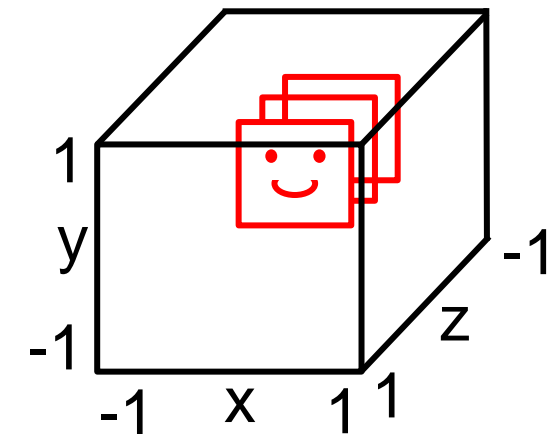
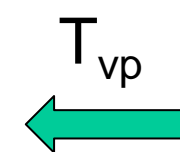
Eye Frame



Display Frame (L)

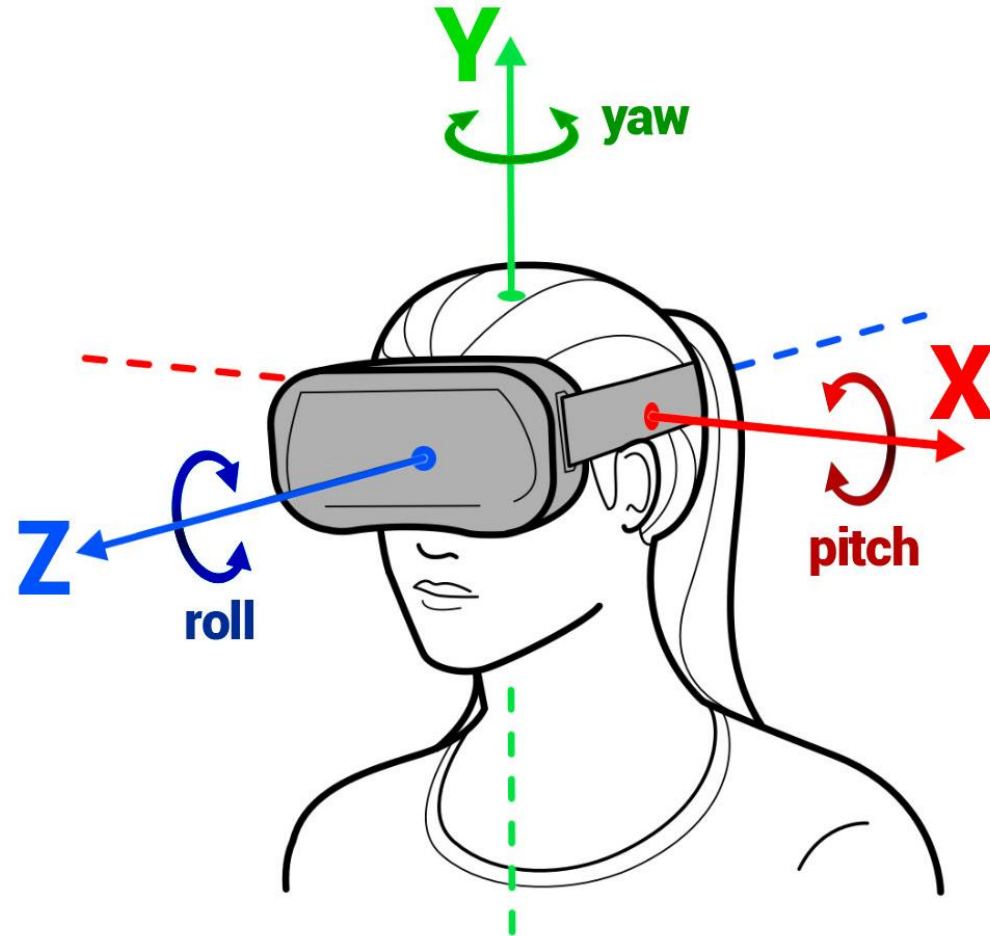


Display Frame (R)

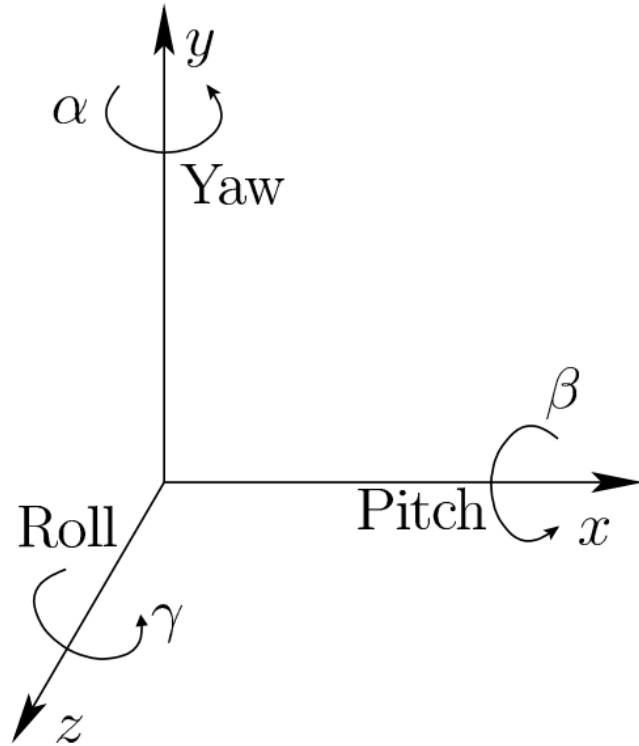


Canonical Frame

3D Rotation and 3D Translation: Six Degrees of Freedom



3D Rotation: Yaw, Pitch and Roll



$$R_y(\alpha) = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{bmatrix} \quad \text{Yaw}$$

$$R_x(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix} \quad \text{Pitch}$$

$$R_z(\gamma) = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{Roll}$$

Combining rotations: $R(\alpha, \beta, \gamma) = R_y(\alpha)R_x(\beta)R_z(\gamma)$

3D Rotation and 3D Translation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}$$

Can the rotation and translation be combined into a single matrix?

Not with a single 3X3 matrix...

Combining 3D Rotation & 3D Translation: Homogeneous Transformation Matrix

$$T_{rb} = \begin{bmatrix} \boxed{\begin{matrix} & & \\ & R & \\ & & \end{matrix}} & \begin{matrix} x_t \\ y_t \\ z_t \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{bmatrix}$$

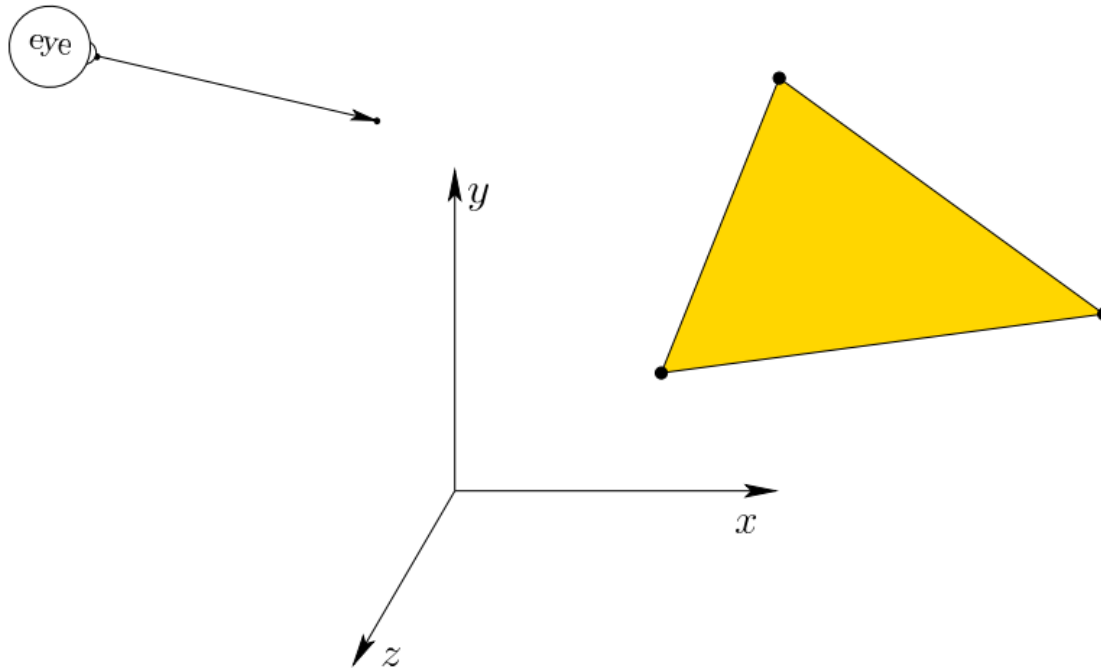
$$\begin{bmatrix} \boxed{\begin{matrix} & & \\ & R & \\ & & \end{matrix}} & \begin{matrix} x_t \\ y_t \\ z_t \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix}$$

Note:

T_{rb} = Rigid body transform

The sequence in this matrix is rotation followed by translation

Eye Transform (T_{eye})



Virtual World

Eye position & orientation:

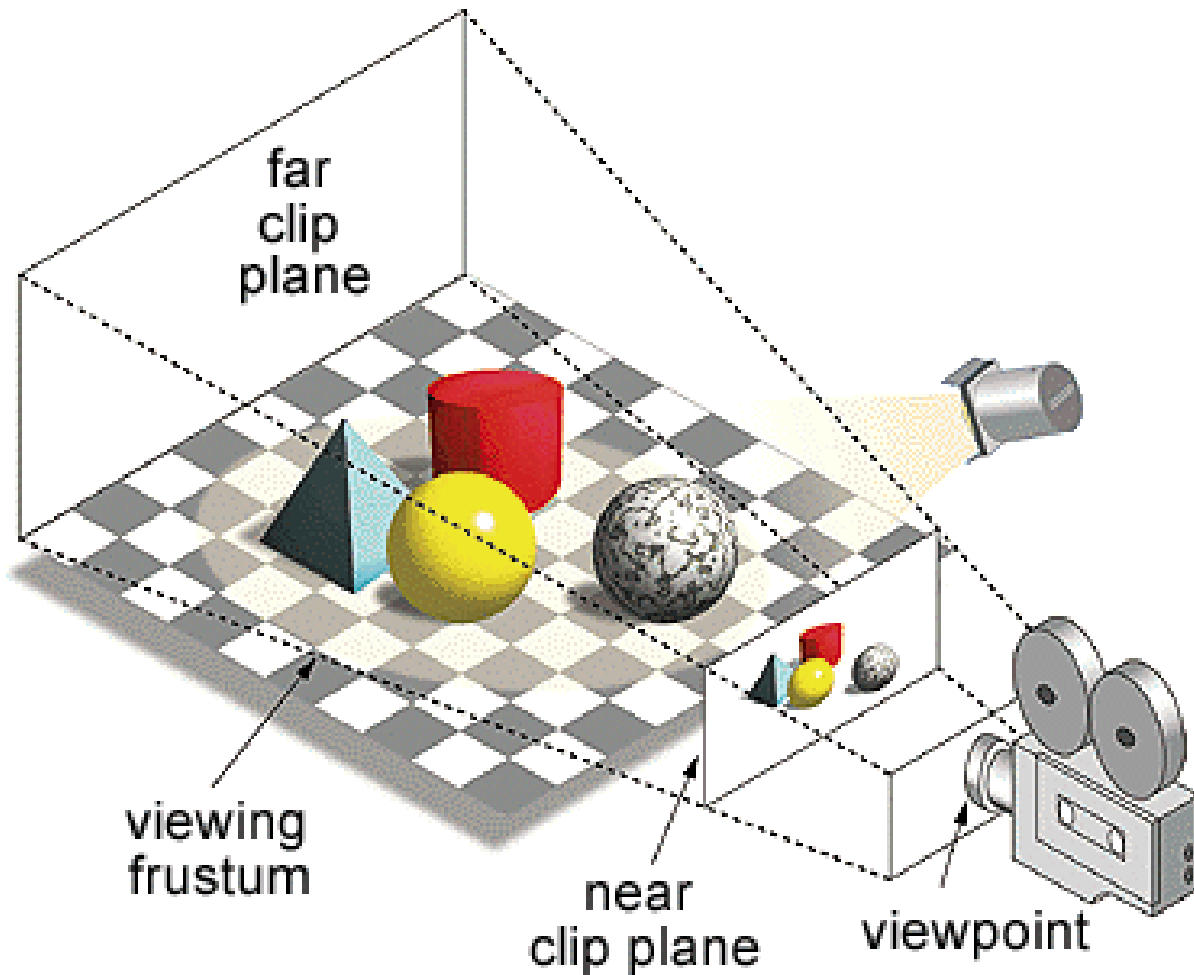
$$e = (e_1, e_2, e_3)$$

$$R_{eye} = \begin{bmatrix} \hat{x}_1 & \hat{y}_1 & \hat{z}_1 \\ \hat{x}_2 & \hat{y}_2 & \hat{z}_2 \\ \hat{x}_3 & \hat{y}_3 & \hat{z}_3 \end{bmatrix}$$

$$T_{eye} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & 0 \\ \hat{y}_1 & \hat{y}_2 & \hat{y}_3 & 0 \\ \hat{z}_1 & \hat{z}_2 & \hat{z}_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -e_1 \\ 0 & 1 & 0 & -e_2 \\ 0 & 0 & 1 & -e_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Canonical View Transform (T_{can})

Perspective Projection, Scaling & Translation



$$T_{can} = T_{st}T_p$$

$$T_p = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n+f & -fn \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$T_{st} = \begin{bmatrix} \frac{2}{r-\ell} & 0 & 0 & 0 \\ 0 & \frac{2}{t-b} & 0 & 0 \\ 0 & 0 & \frac{2}{n-f} & -\frac{n+f}{n-f} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Finally: Viewport Transform (T_{vp})

Bring the projected points to the coordinates used to index pixels on the physical display screen:

$$T_{vp} = \begin{bmatrix} \frac{m}{2} & 0 & 0 & \frac{m-1}{2} \\ 0 & \frac{n}{2} & 0 & \frac{n-1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

m : number of horizontal pixels

n : number of vertical pixels

For example, $n = 1080$ and $m = 1920$ for a 1080p display.

One More Thing: “Cyclopean” to Stereo Views

Assuming interpupillary Distance (IPD) = t

$$T_{left} = \begin{bmatrix} 1 & 0 & 0 & \frac{t}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

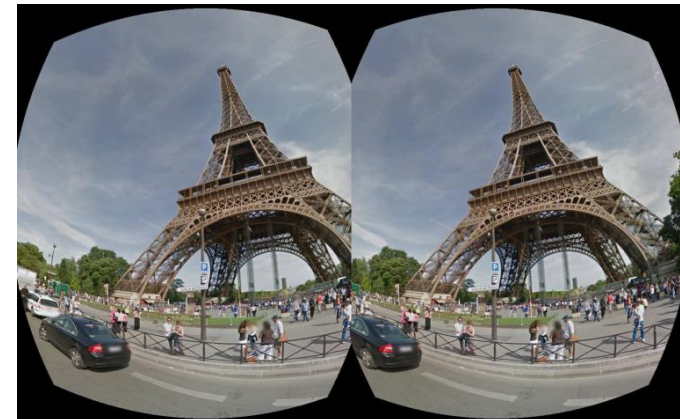
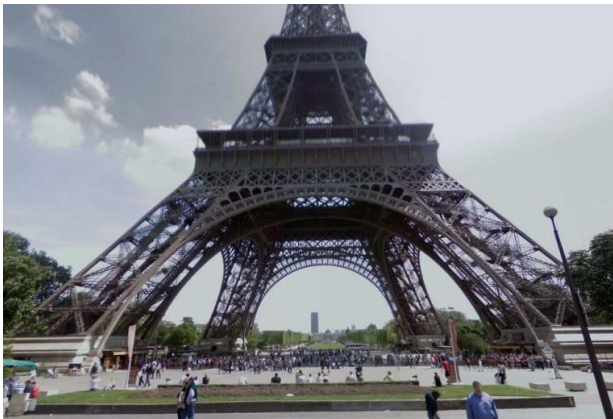
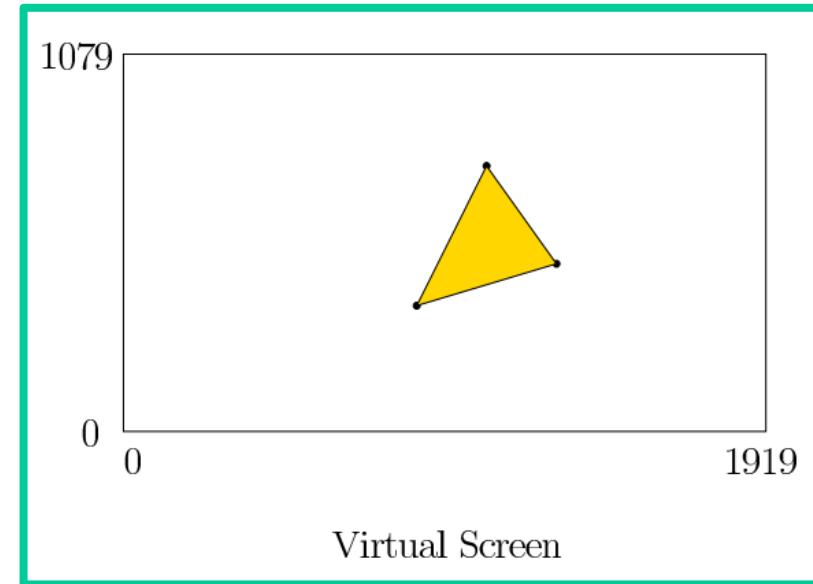
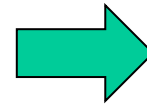
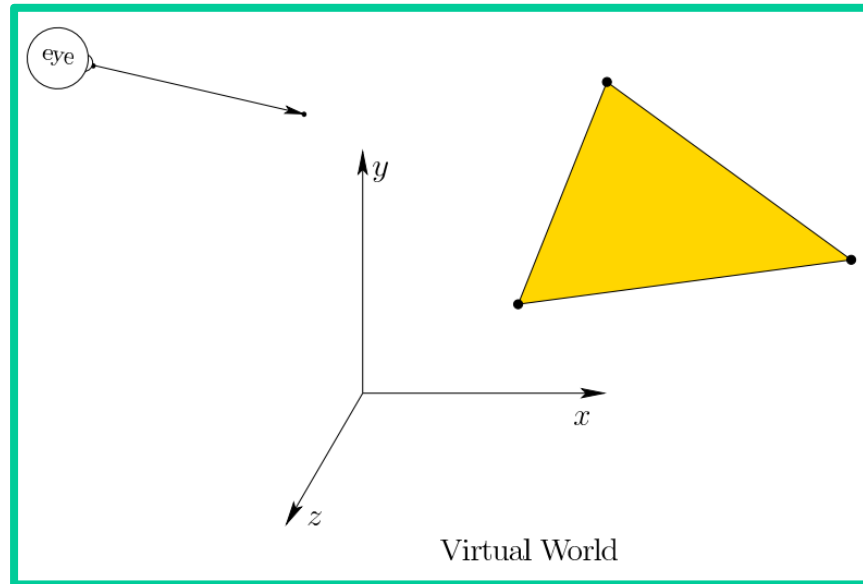
$$T = T_{vp} T_{can} T_{eye} T_{rb}$$



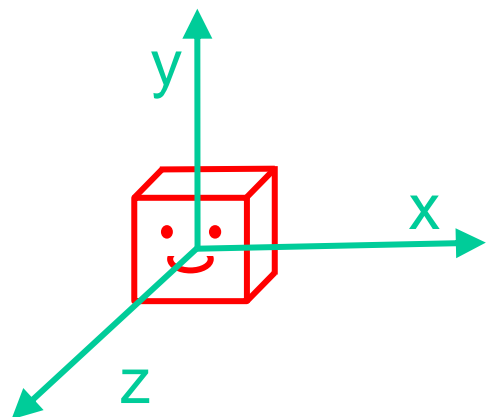
$$T = T_{vp} T_{can} T_{left} T_{eye} T_{rb}$$

Similarly, $T_{right} = \begin{bmatrix} 1 & 0 & 0 & -\frac{t}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

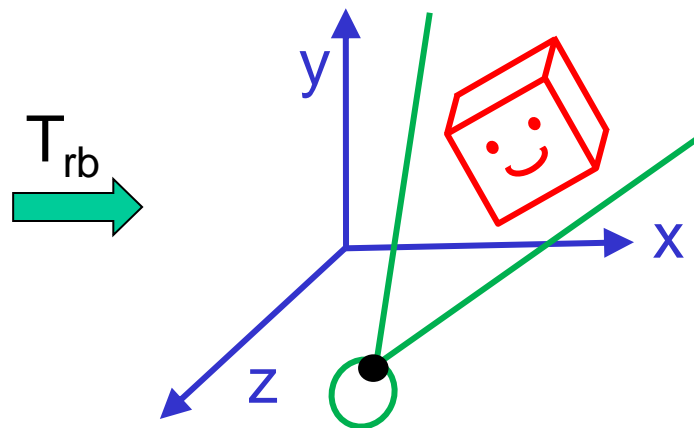
So, after applying all the viewing transformations...



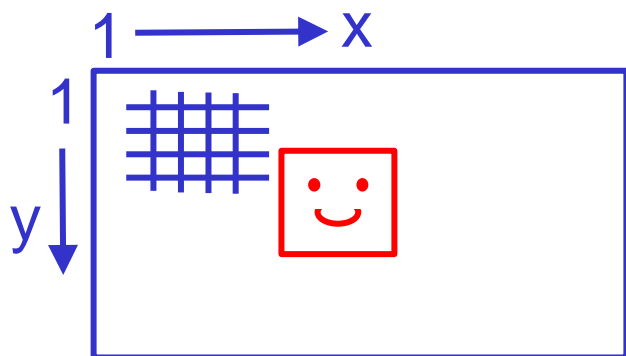
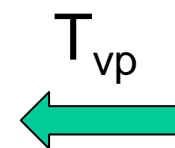
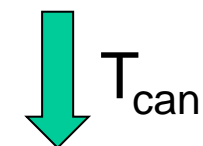
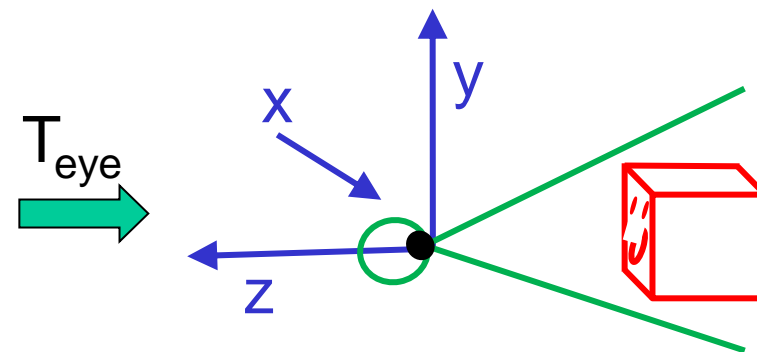
Object Frame



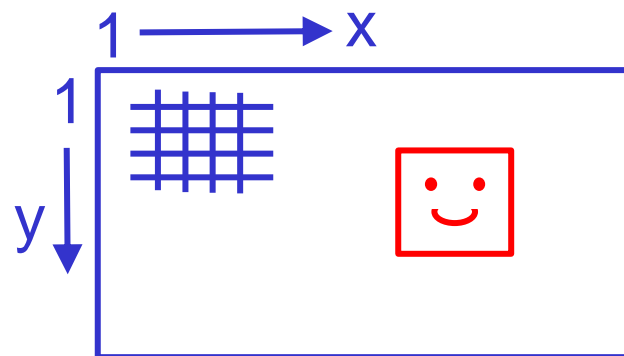
World Frame



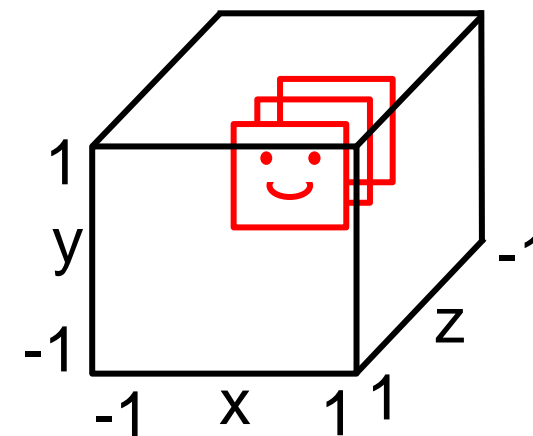
Eye Frame



Display Frame (L)

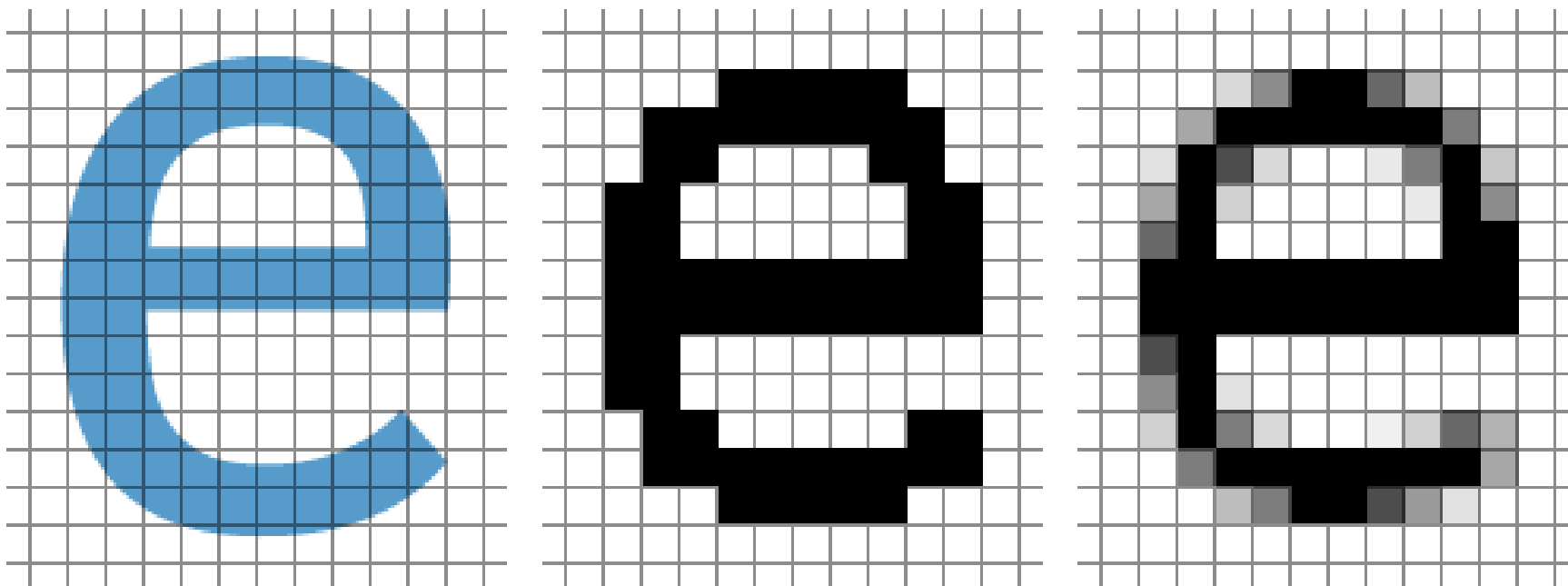


Display Frame (R)



Canonical Frame

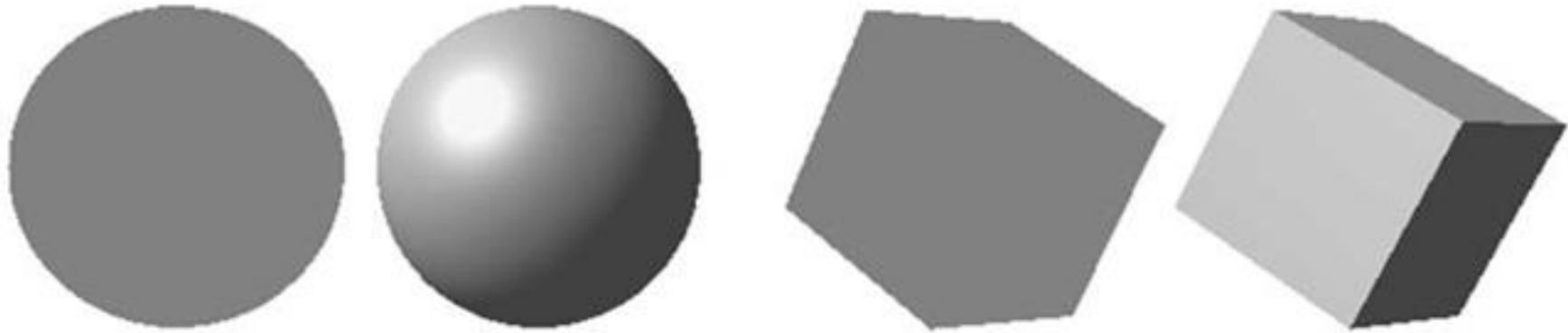
Rasterization + Antialiasing



Illumination + Shading



Illumination + Shading



Geometric Models in VR/AR/MR: How many polygons are needed?



60
triangles



600
triangles



6000
triangles



60000
triangles



Sensing & Tracking Technologies

“When it comes to VR, a system is only as good as your tracking setup.”

“A chief complaint for many current high-end VR systems is that they’re just too damn complicated.”

- TechCrunch, Jan 25, ‘17

3D Position Tracking: Artificial Features Tracked by External Cameras



Photo: iFixit

“Constellation” tracking system
used in Oculus Rift

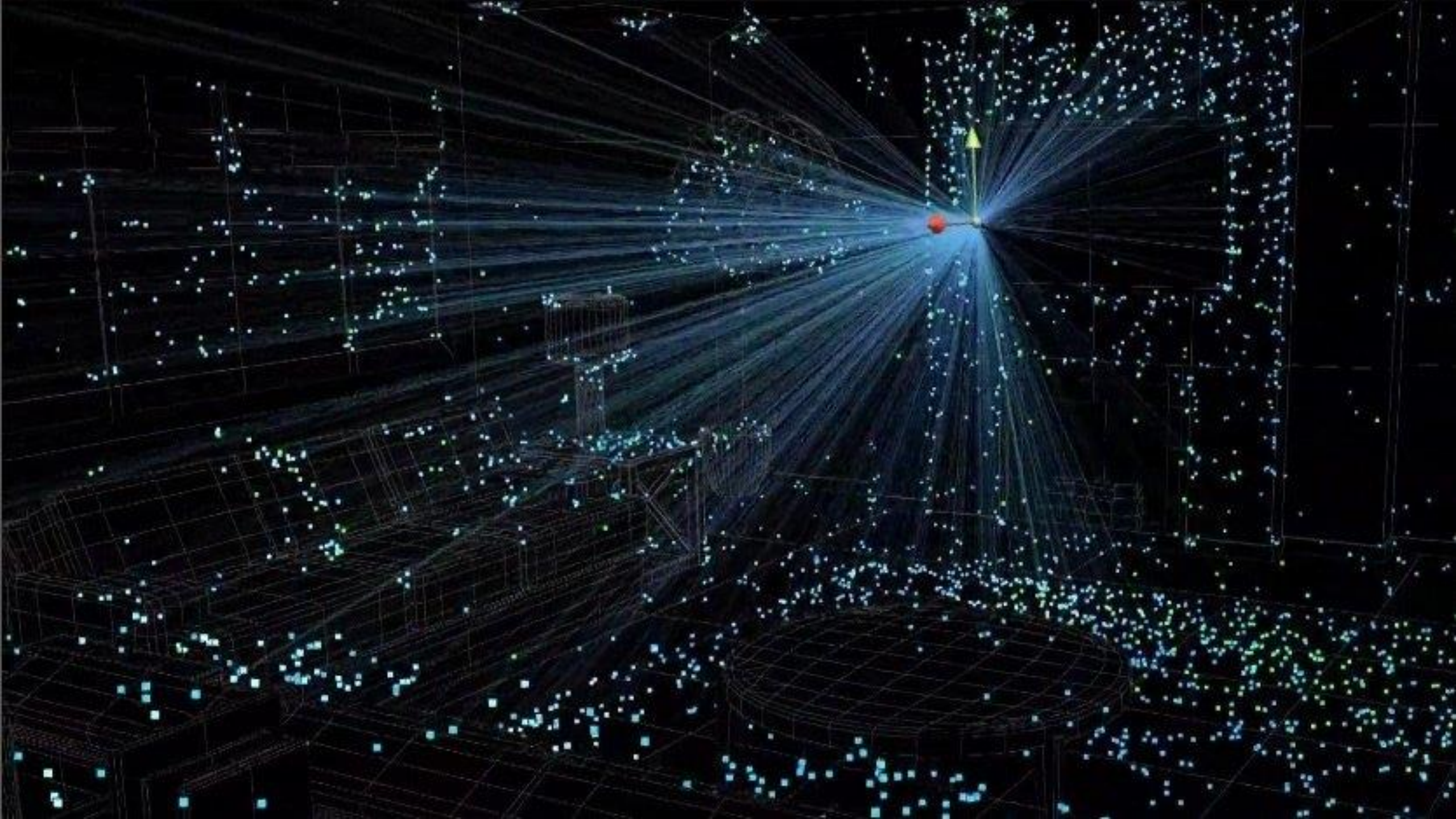


“Lighthouse” tracking system
used in HTC Vive

“Inside Out” Tracking:

Natural Visual Features Extracted and Tracked with
Computer Vision along with Inertial Motion





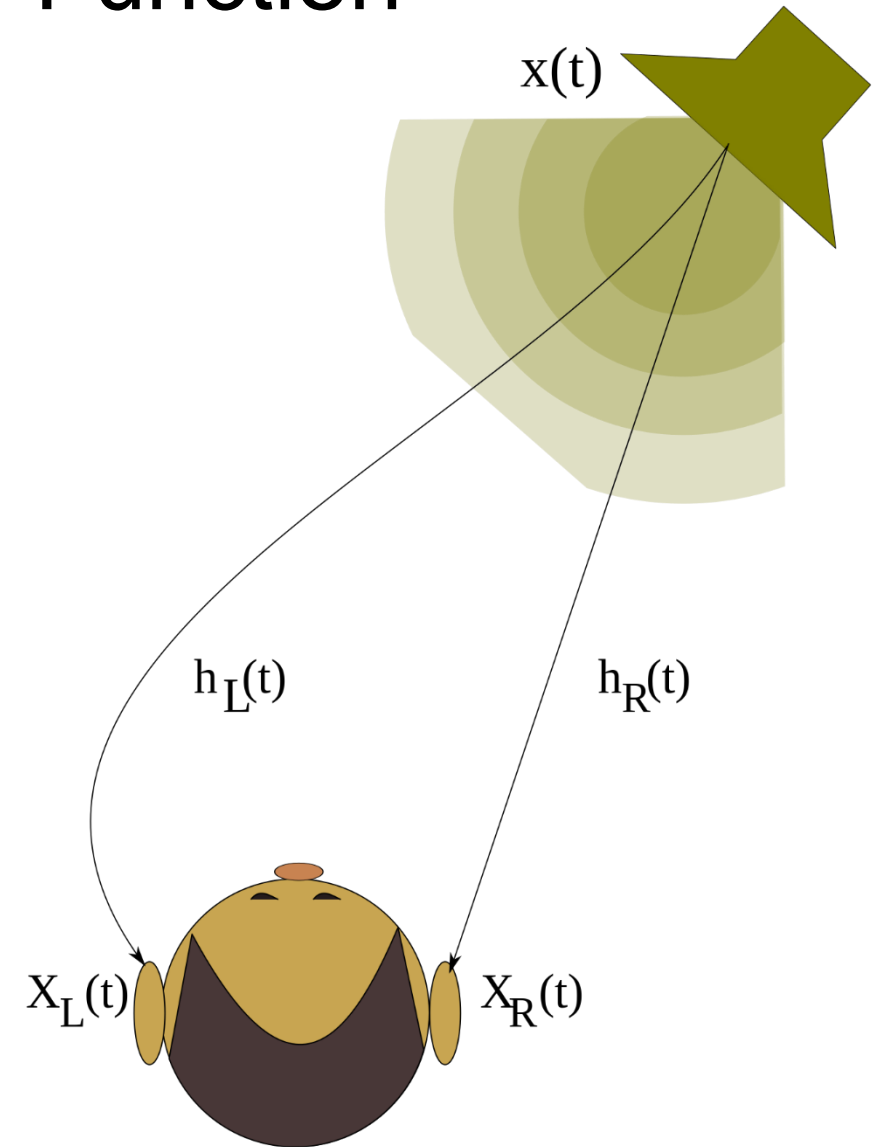


3D Sound Localization

Head-Related Transfer Function (HRTF)

HRTF characterizes how an ear receives a sound from a point in space.

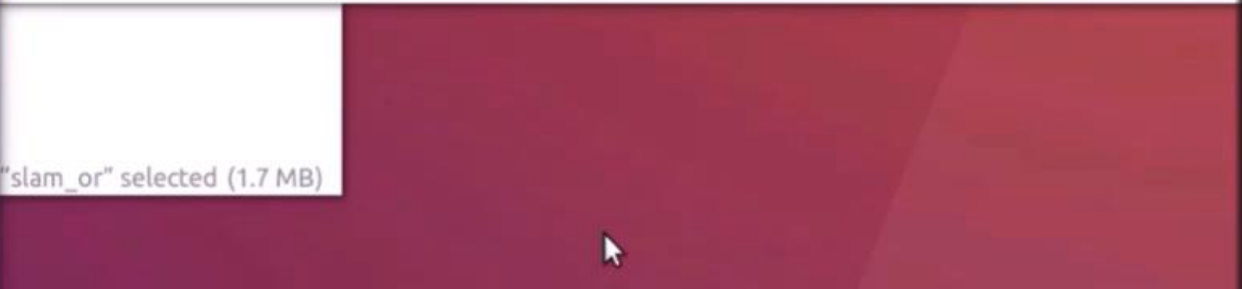
The size and shape of the head, ears, ear canal, density of the head, size and shape of nasal and oral cavities, all transform the sound and affect how it is perceived, boosting some frequencies and attenuating others.



Some Key Trends...

OR

bed: 0.980362 (541.72, 111.975, 1748.05)
pillow: 0.995586 (117.218, -386.776, 3074.43)
chair: 0.949045 (-784.102, 151.479, 1962.88)





ObjectList

- struct StaticMesh
- struct StartLocation
- struct StaticMesh
- struct StaticMesh
- struct StaticMesh
- struct GoldMeshObject

Inspector

▼ Scene Outlets

- 01_hands
- 02_music
- 03_paddle
- 04_physics
- 05_sculpting
- 06_glass
- 07_collaborative
- 08_pure
- 09_lathe
- morpheus
- painting
- robot
- sci_f1
- sci_f12



VentureBeat

AR and VR need to move from what developers like to what consumers want

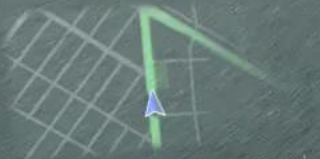


Yukon St.

PEPS SMART

Home
6.55 mi

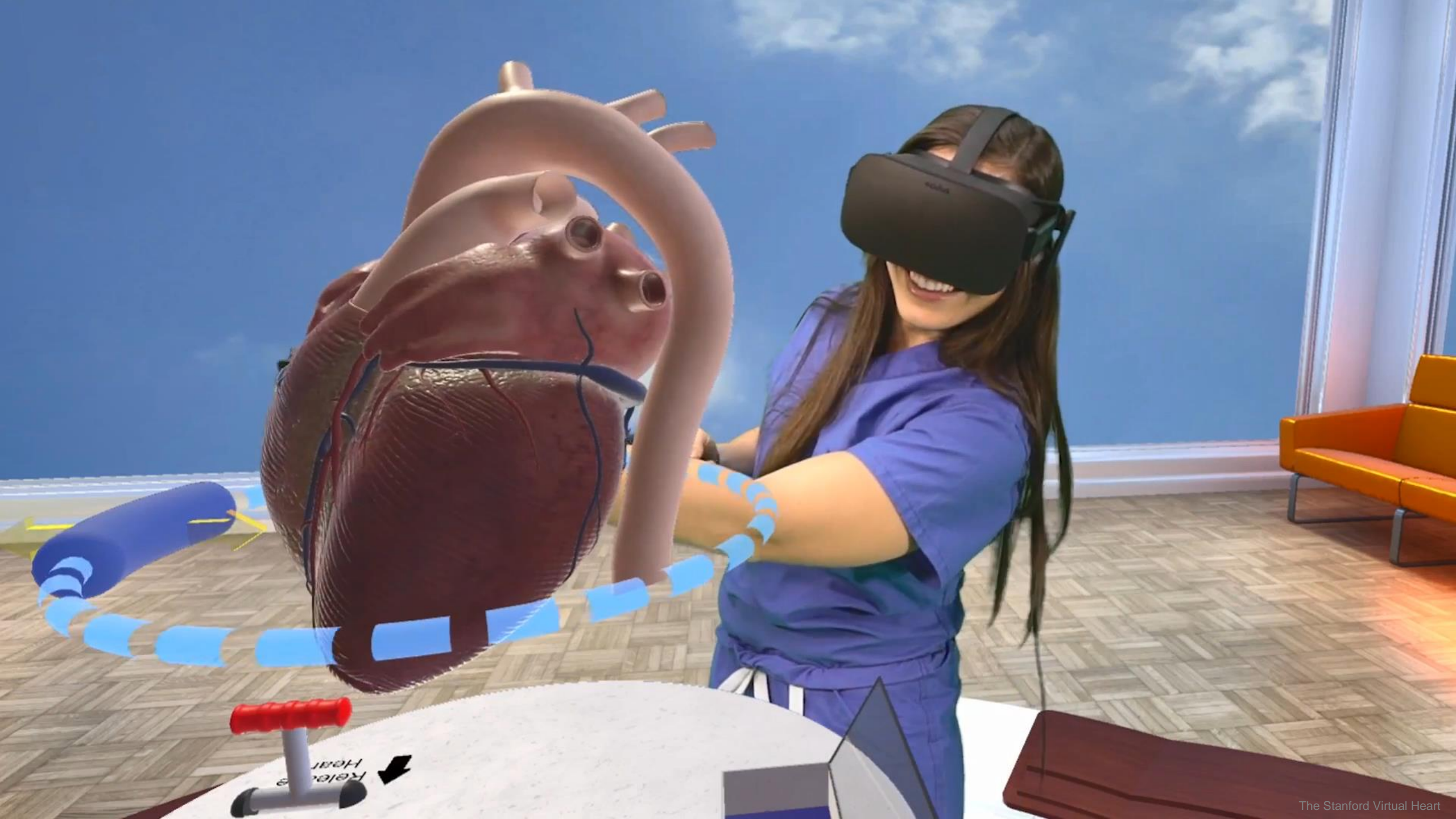
Take exit right on Cambie St towards
City Center



Home
6.55 mi
arrival in 13:24

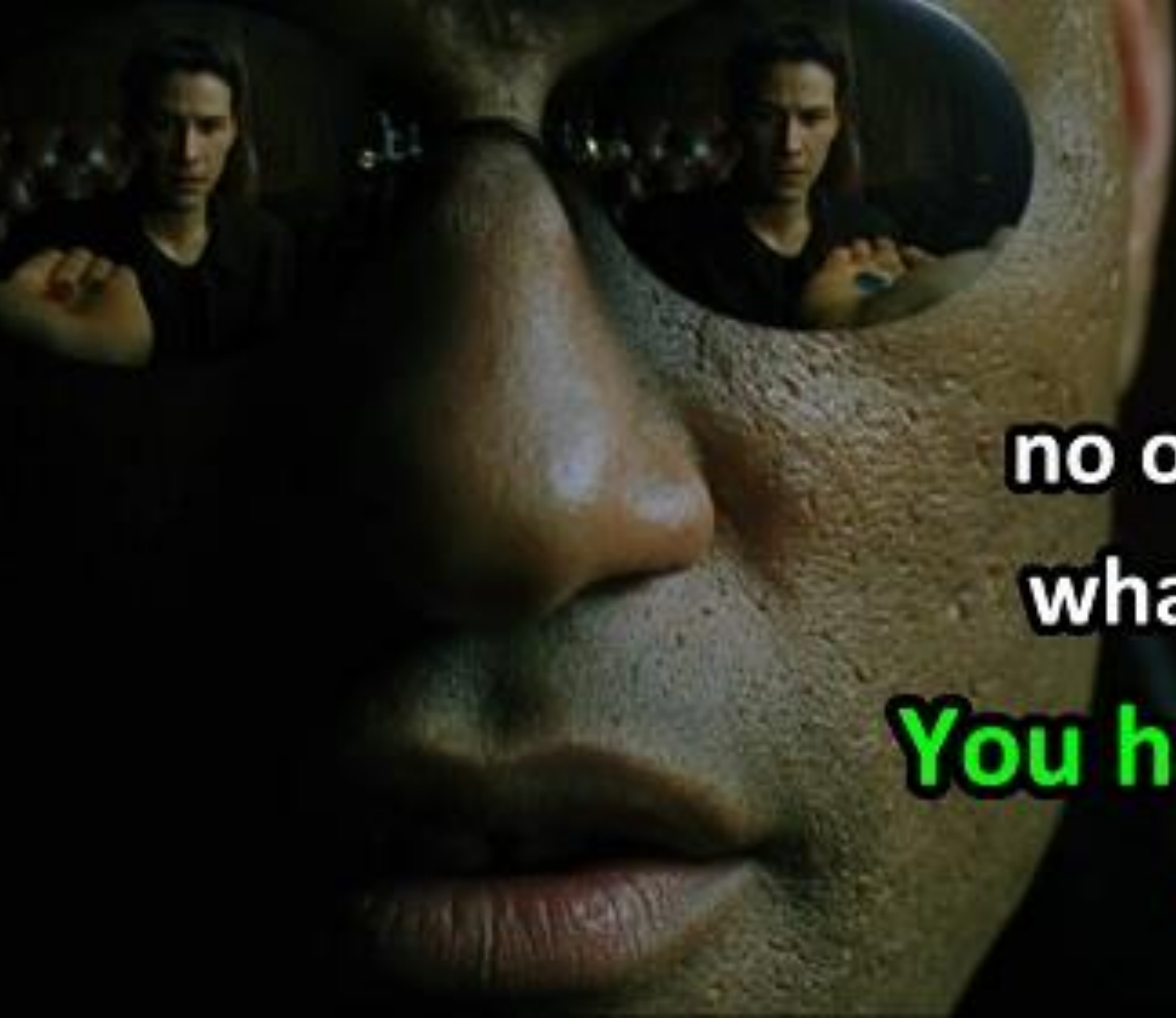
11 mph 25

4:35 74°
aug 6
Mr. Blue Sky
Electric Light Orchestra



Parting Thoughts...

- Key requirements for virtual and augmented reality devices and applications are now relatively well-understood.
- Lots of progress recently in the development of technologies and components, but some major challenges remain to be solved for mass adoption.
- Ingredients are important, but end-to-end integration (hardware, software, system, interfaces) is critical.
- Commercial success will hinge on delivering compelling applications with natural user experiences.



Unfortunately
no one can be told
what the matrix is
You have to see it
for yourself

Thank You!

achintya.k.bhowmik@gmail.com

너를 만났다

어딤어?

References & Acknowledgements

- The Matrix (1999 Sci-Fi film), wikipedia.org/wiki/The_Matrix
- African Bull Frog Ant Crusher, www.youtube.com/watch?v=WIEzvdIYRes
- E. Adelson, Checker Shadow Illusion, persci.mit.edu/gallery/checkersshadow
- R. Shepard, wikipedia.org/wiki/Roger_Shepard
- E. Lingelbach, wikipedia.org/wiki/Grid_illusion
- The McGurk Effect, www.youtube.com/watch?v=aFPtc8BVdJk
- www.project-ican.com
- E. Bruce Goldstein, Sensation and Perception, Cengage Learning
- blog.gingkoapp.com/essays/seeing-an-idea
- G. Wetzstein, EI 2017 Plenary: VR 2.0: Making Virtual Reality better than Reality
- www.levelcapped.com/ivan-sutherlands-sword-of-damocles-virtual-reality-head-mounted-viewer/
- S. Lavalley, Virtual Reality, vr.cs.uiuc.edu
- what-when-how.com
- dsky9.com/rift/vr-tech-6dof
- Virtual Reality in Architecture - The Courtyard, oneirosvr.com
- The Stanford Virtual Heart, www.stanfordchildrens.org/en/innovation/virtual-reality/stanford-virtual-heart
- clinicalarchitecture.com
- Phiar, www.youtube.com/watch?v=CQqPWEp0Zew
- www.marketplace.org, Tim P. Whitby / Getty Images
- Johnny English Strikes Again (2018 action comedy film), wikipedia.org/wiki/Johnny_English_Strikes_Again
- iFixit
- Wikipedia

A (Brief) Historical Perspective... and Why Now?



The Sensorama

Morton Heilig described his vision of a multi-sensory theater in 1955.

Built a prototype in 1962, along with five short films. Halted due to lack of financial support.

Capabilities:

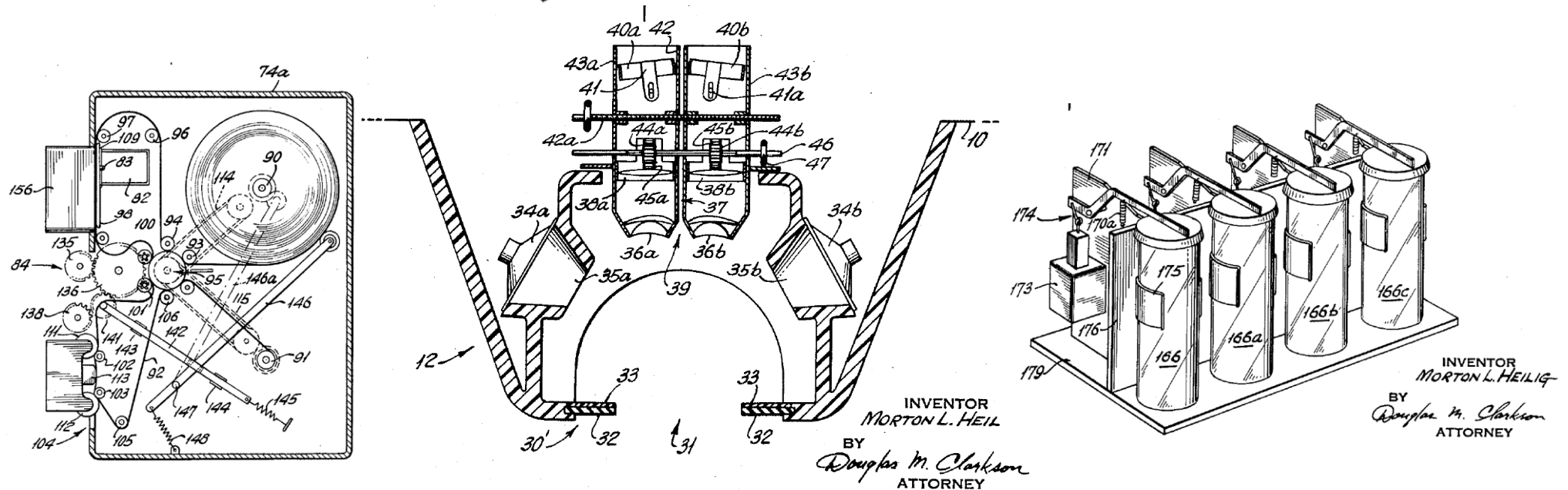
- Stereoscopic 3-D images
- Body tilting
- Stereo sound
- Wind and aromas

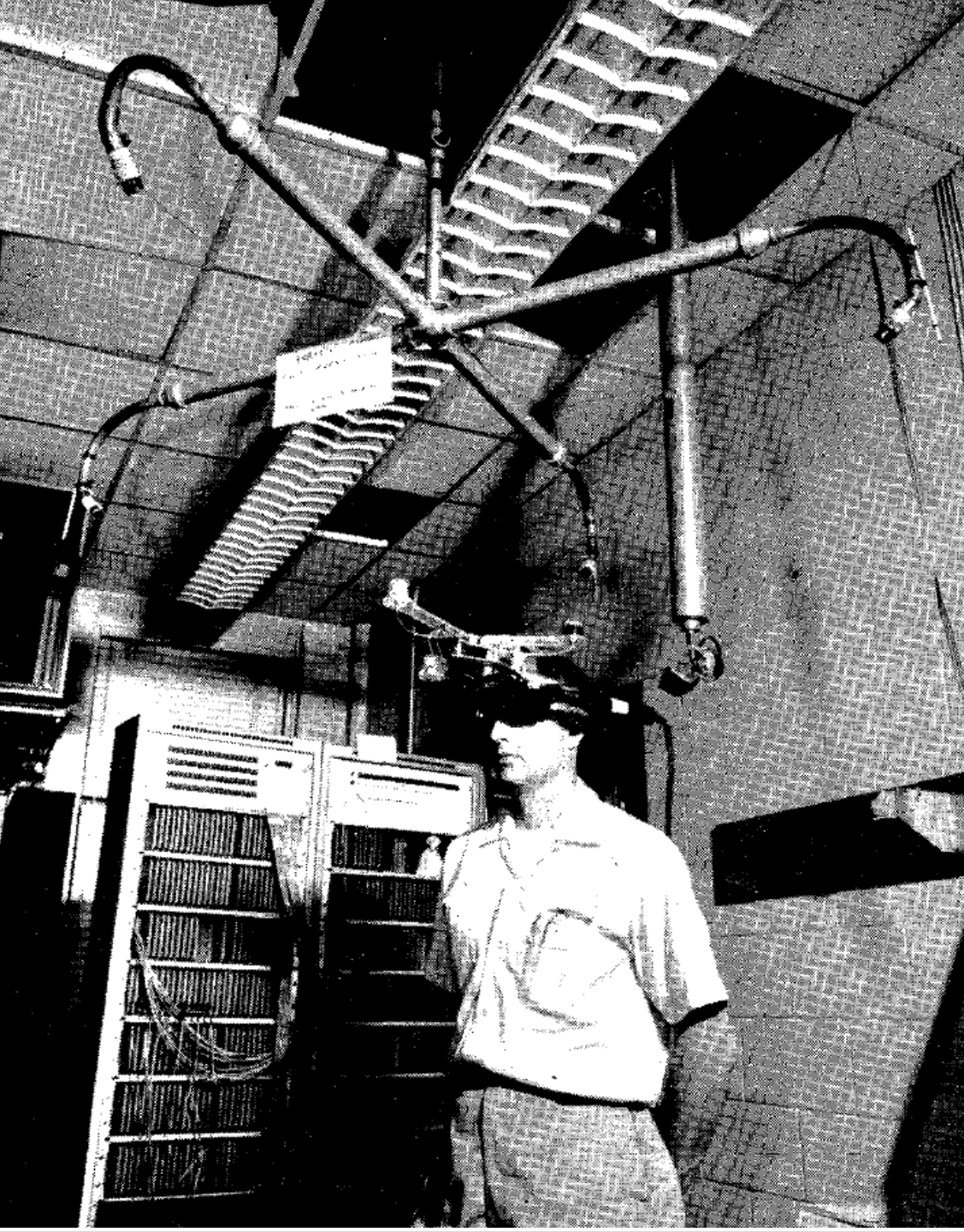
References:

- Heilig M. (1962). US Patent #3,050,870
- Rheingold, H. (1992). Virtual Reality, Simon & Schuster, New York, N.Y.
- wikipedia.org/wiki/Sensorama

3,050,870
SENSORAMA SIMULATOR
Morton L. Heilig, Long Beach, N.Y.
(10 Sheridan Square, New York 14, N.Y.)
Filed Jan. 10, 1961, Ser. No. 81,864
13 Claims. (Cl. 35—1)

The present invention, generally, relates to simulator apparatus and, more particularly, to apparatus to stimulate the senses of an individual to simulate an actual experience realistically.





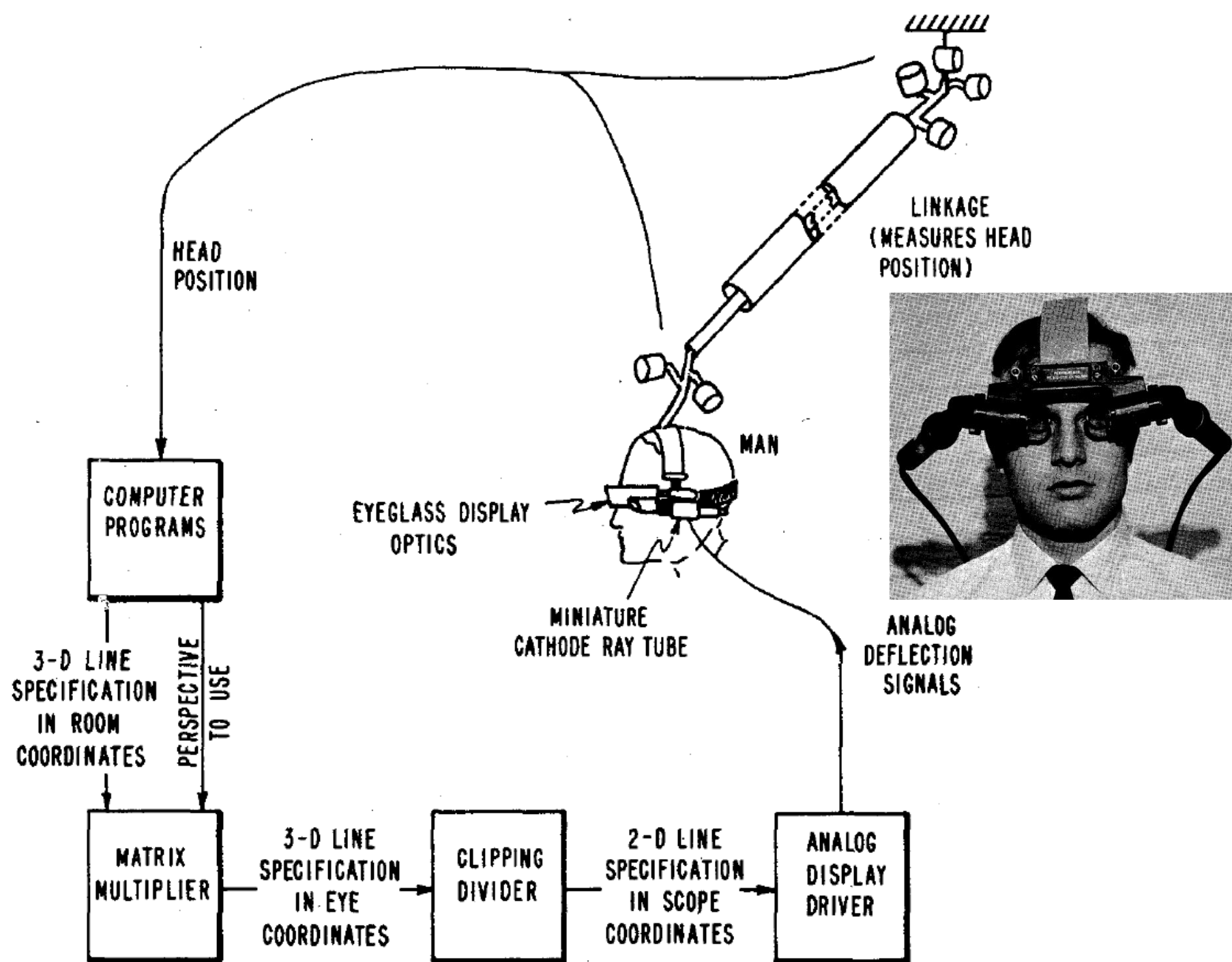
The Sword of Damocles

Ivan Sutherland and his student **Bob Sproull** created the first VR/AR system with head-mounted displays in 1968.

Displayed output from a computer program in a binocular display (2 CRTs). The perspective changed with user's head tracking.

References:

- Sutherland, I. E. (1968). "A head-mounted three dimensional display". Proceedings of AFIPS 68, pp. 757-764
- [wikipedia.org/wiki/The_Sword_of_Damocles_\(virtual_reality\)](https://wikipedia.org/wiki/The_Sword_of_Damocles_(virtual_reality))





VPL EyePhone & DataGlove

Jaron Lanier founded **VPL Research** in 1984, and was among the first to build and sell virtual reality products.

The EyePhone was an HMD intended to immerse users into a computer simulation. It could track head movements. The headset used Fresnel lenses.

Model 1: 2.4Kg

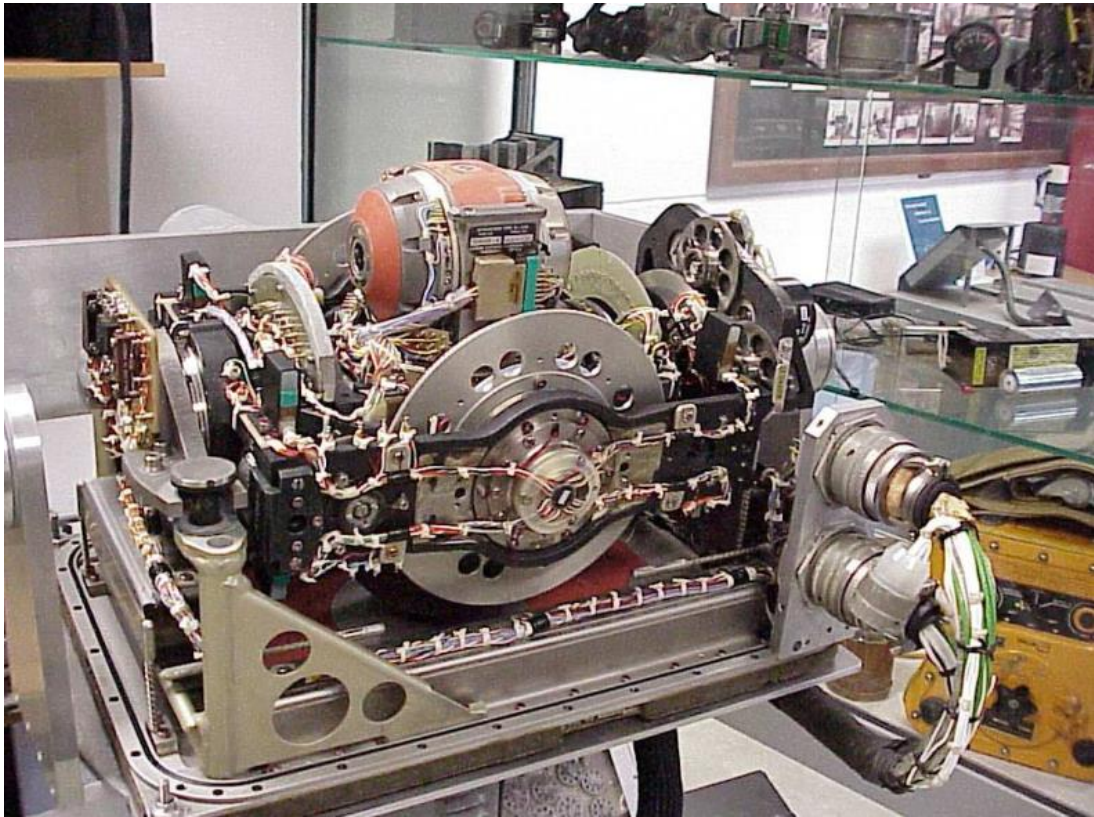
Model 2: 2.5lbs

Both >\$9K

References:

- wikipedia.org/wiki/VPL_Research
- vrwiki.wikispaces.com/VPL+EyePhone

So... Why Now? An Example...



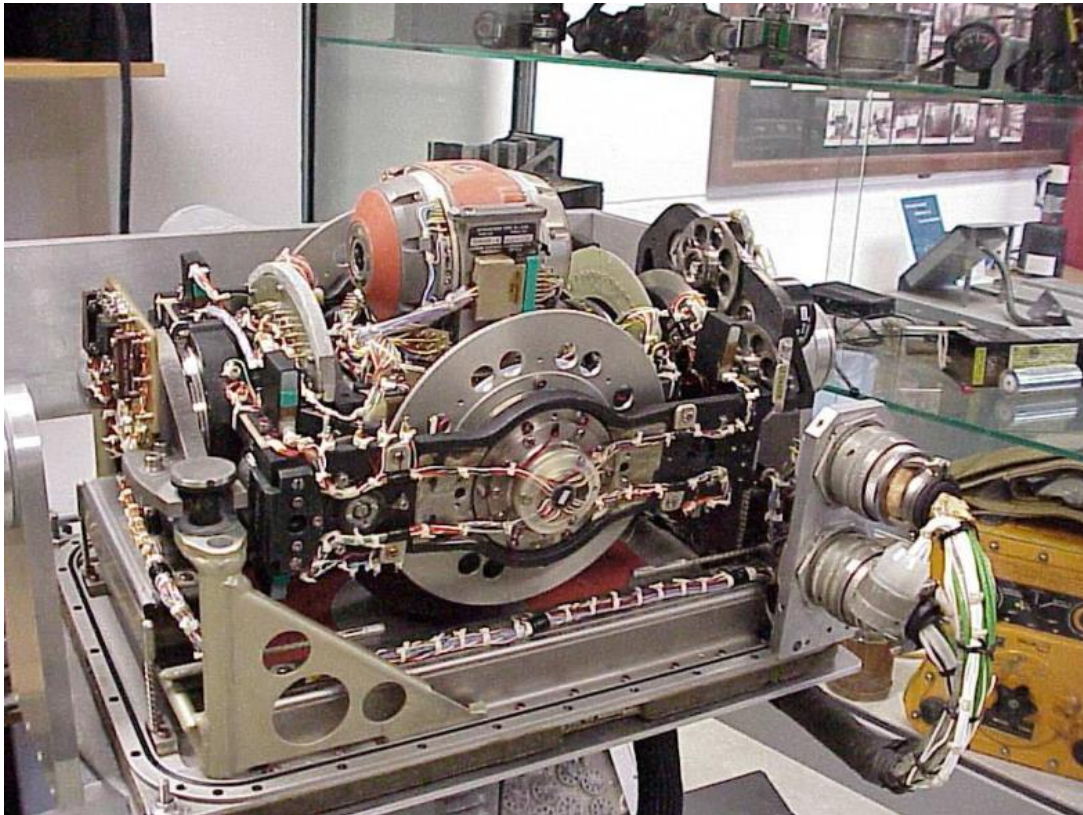
The LN-3 Inertial Navigation System, developed in the 1960s by Litton Industries

"They tell me about electric motors that are the size of the nail on your small finger... But that's nothing; that's the most primitive, halting step in the direction I intend to discuss."

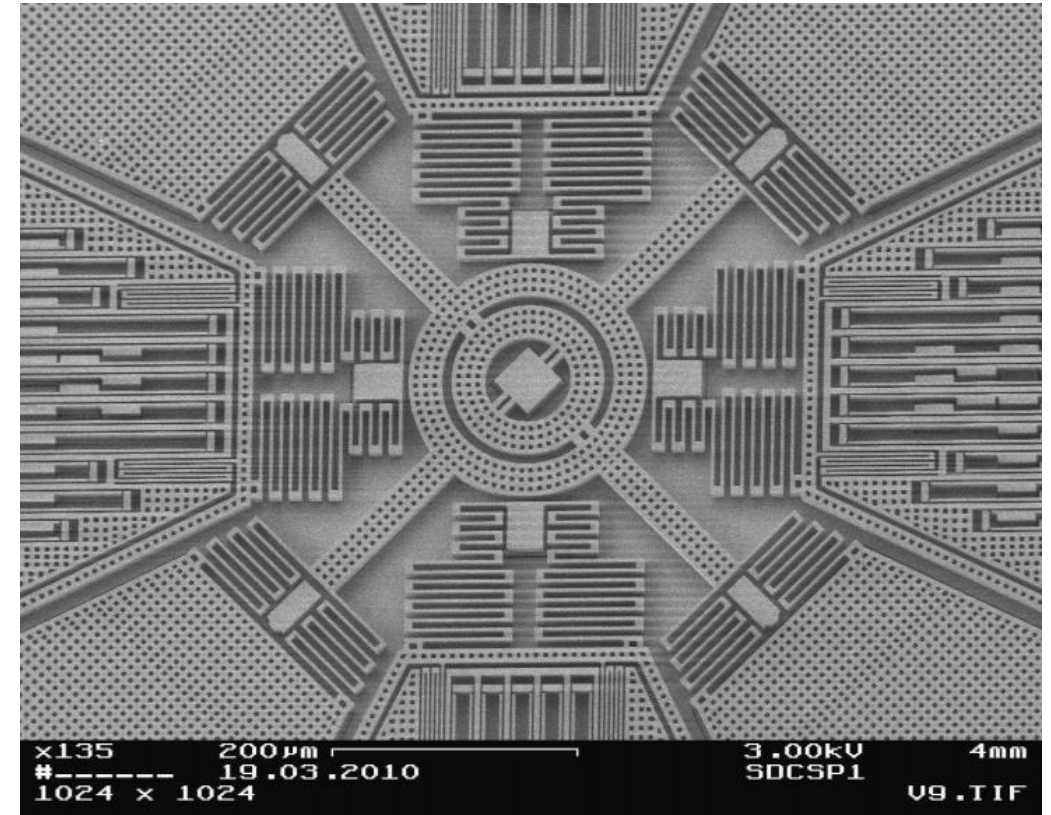
It is a staggeringly small world that is below. *In the year 2000, when they look back at this age, they will wonder why it was not until the year 1960 that anybody began seriously to move in this direction."*

– **Richard Feynman**, "There's Plenty of Room at the Bottom", at the 1959 annual meeting of American Physical Society

So... Why Now? An Example...



The LN-3 Inertial Navigation System, developed in the 1960s by Litton Industries



The internal structures of a modern MEMS gyroscope; the total width is less than 1mm

So... Why Now? Another Example...



CRT Display, Circa 2000



55" OLED TV, LG, 2012

Examples of Modern-Day VR Devices (PC and Console Based)



HTC Vive
with Valve



Facebook
Oculus Rift



Sony
PlayStation VR

Examples of Modern-Day VR Devices (Smartphone Based)



Samsung
Gear VR



Google
Daydream

Examples of Modern-Day AR/MR Devices (Standalone and Tethered)



Microsoft
HoloLens 2



Magic Leap
One



Examples of Phone/Tablet-Based AR



Apple ARKit

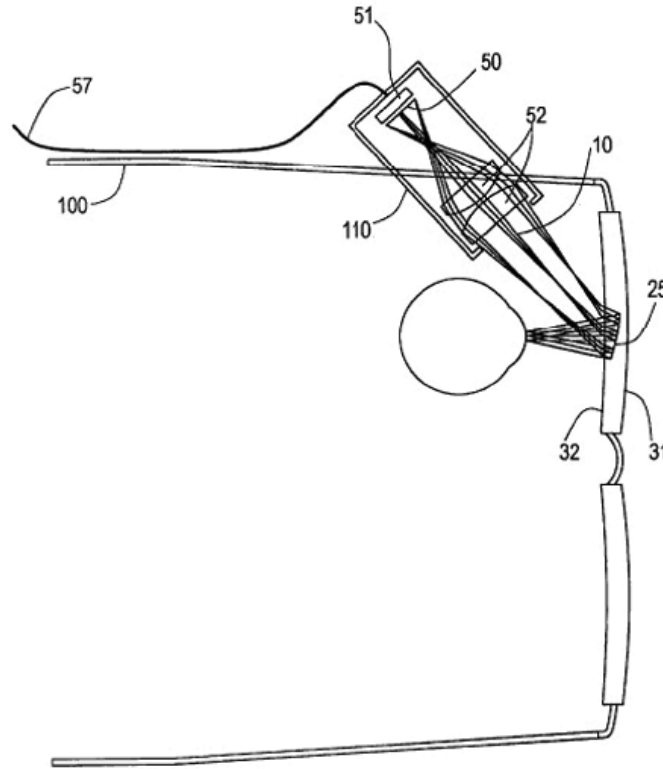


Google ARCore

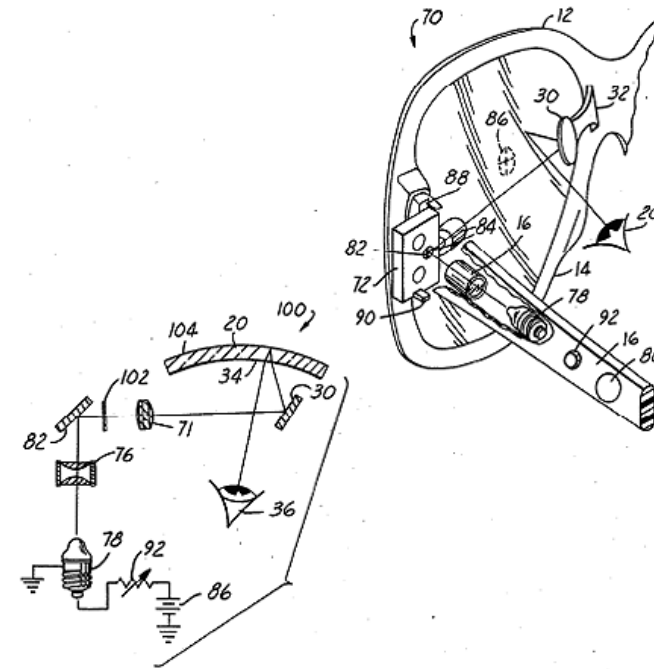
Examples of Optics for See-Through Head-Worn Displays (HWD) in Augmented/Mixed Reality Systems

HWD Examples (Slide 1 of 18)

Spitzer. Eyeglass Display Lens System Employing Off-Axis Optical Design.
US 6,353,503
Mar. 5, 2002

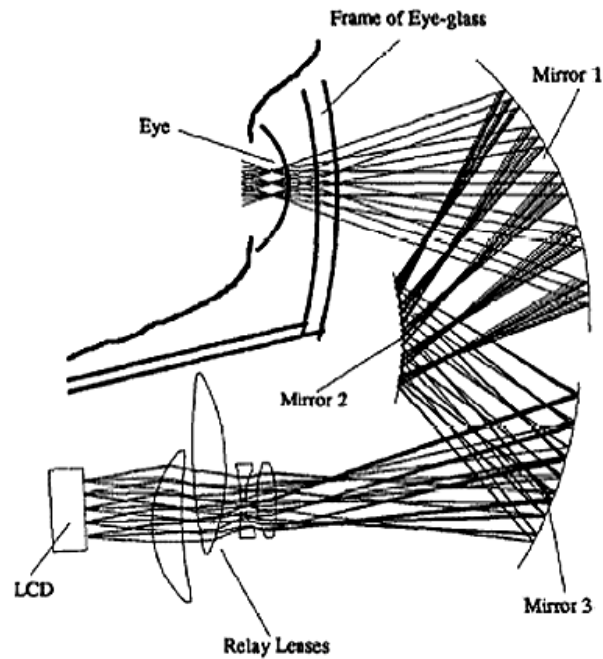


Bettinger. Spectacle-mounted ocular display apparatus.
US 4,806,011
Feb. 21, 1989

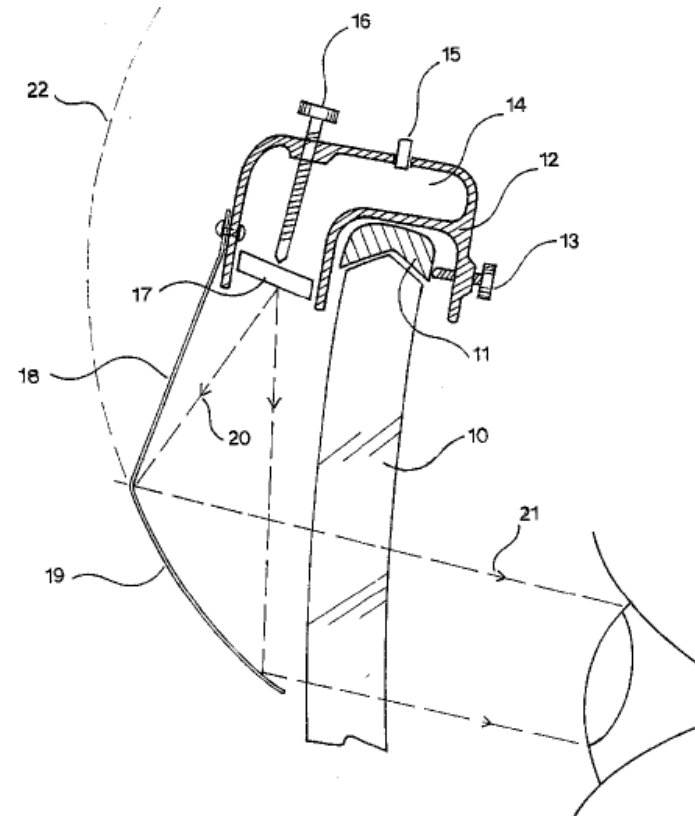


HWD Examples (Slide 2 of 18)

Hoshi *et al.* Off-axial HMD optical system consisting of aspherical surfaces without rotational symmetry.
In Proc. of SPIE Vol. 2653



Perera. Display Projection Optical System for Spectacles or Sunglasses.
US 4,867,551
Sep. 19, 1989

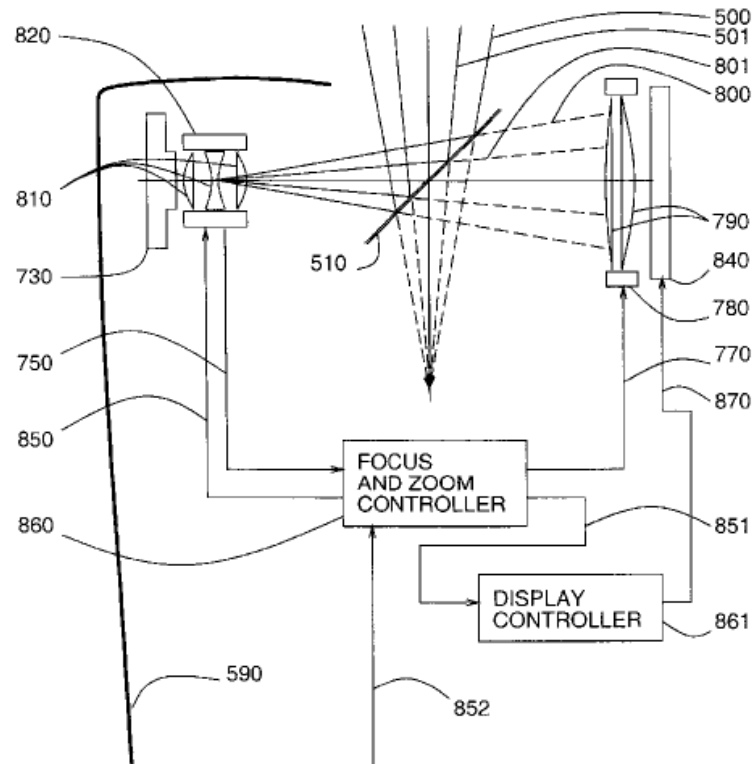


HWD Examples (Slide 3 of 18)

Mann. Wearable Camera
System With Viewfinder Means.

US 6,307,526

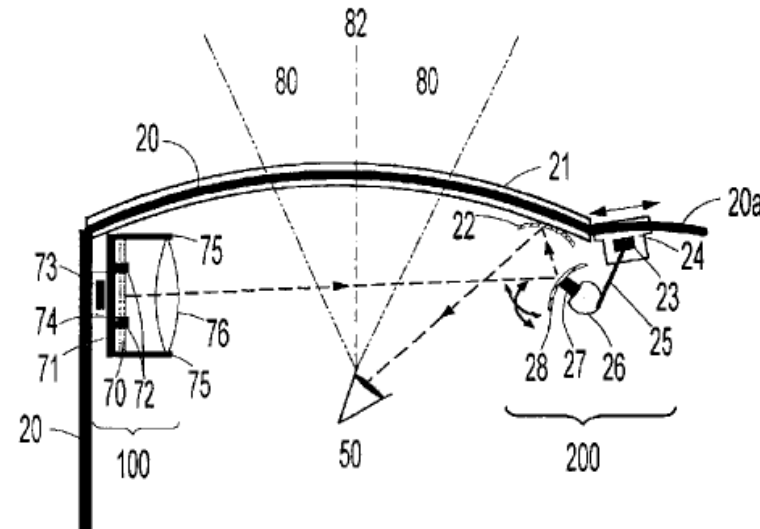
Oct. 23, 2001



Geist. Head-mounted virtual display
apparatus with near-eye deflecting element
in the peripheral field-of-view.

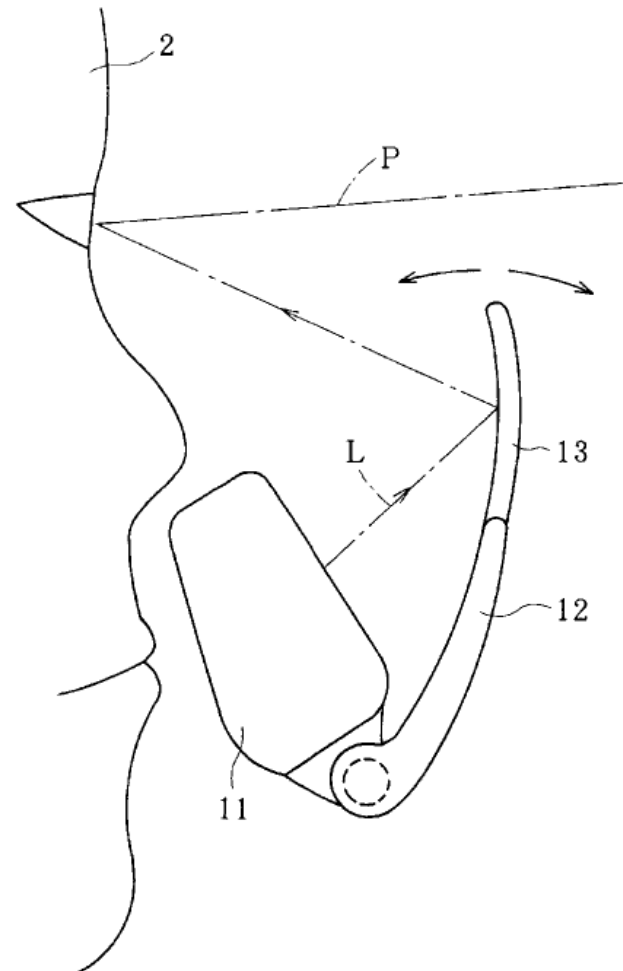
US 6, 771,423

Aug. 3, 2004

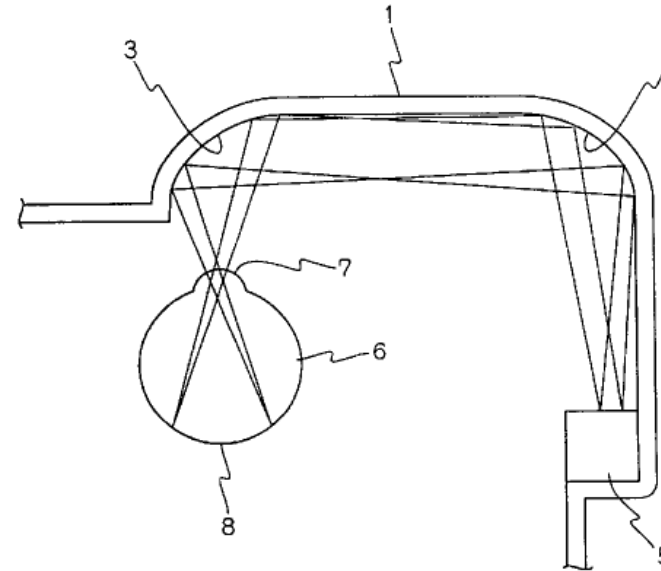


HWD Examples (Slide 4 of 18)

Amafuji. Head Mounted Display Device.
US 6,359,602
Sep. 19, 2002

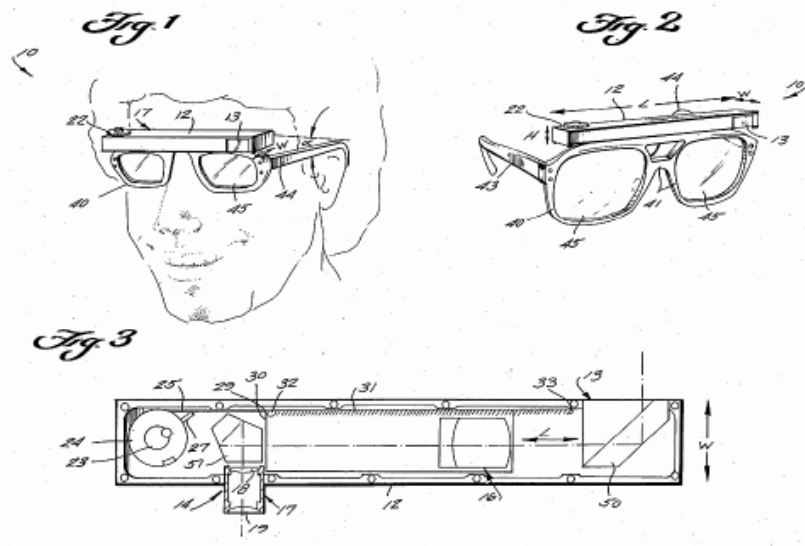


Kuriyama. Image Display Apparatus
US 6,081,304
Jun. 27, 2000

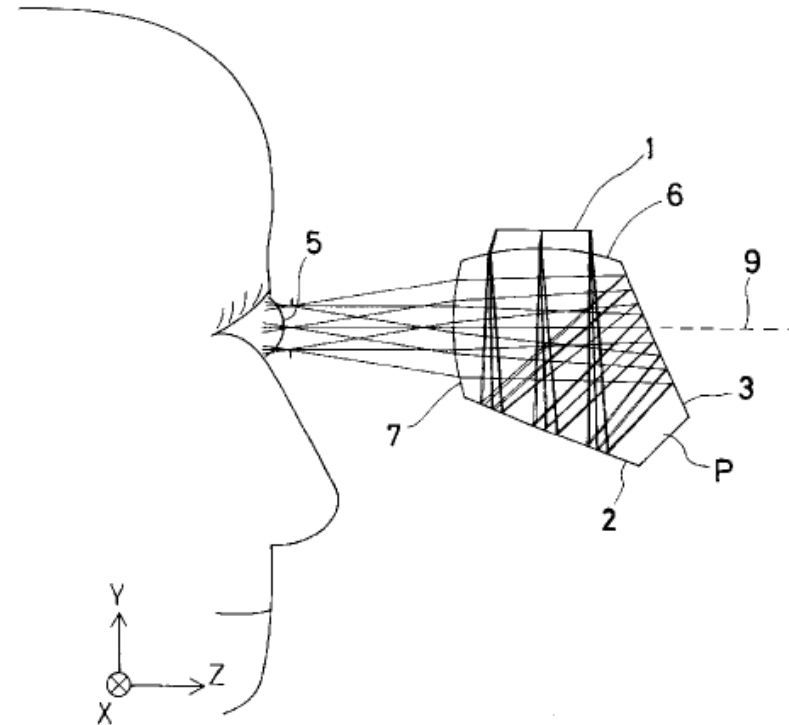


HWD Examples (Slide 5 of 18)

Pekar. Vision enhancing system.
US 4,704,000
Nov. 3, 1987



Togino. Prosm Optical System.
US 5,991,103
Nov. 23, 1999

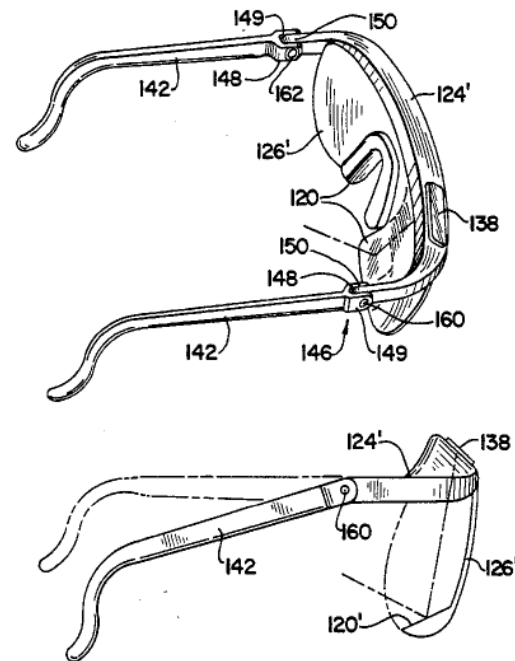


HWD Examples (Slide 6 of 18)

Furness. Display System for a Head Mounted Viewing Transparency.

US 5,162,828

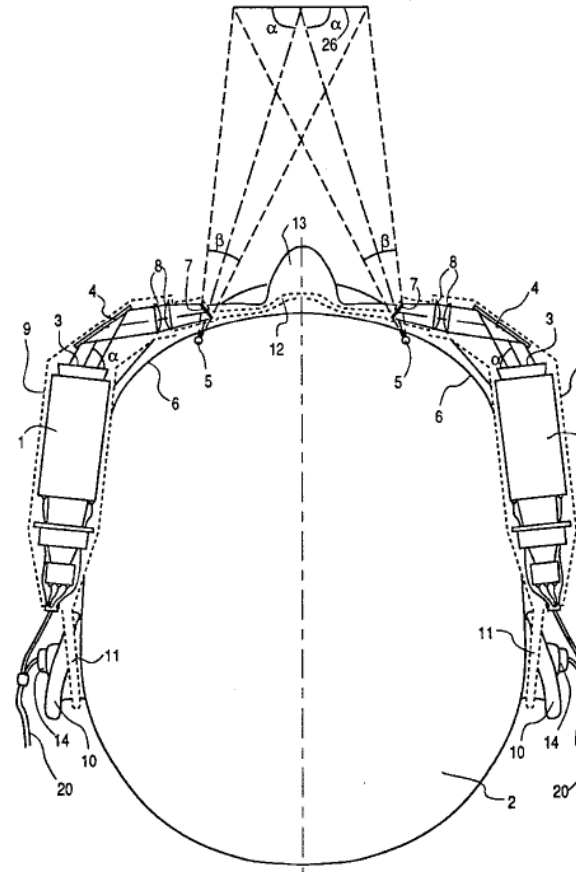
Nov. 10, 1982



Holakovszky. Stereoscopic video image display appliance wearable on head like spectacles.

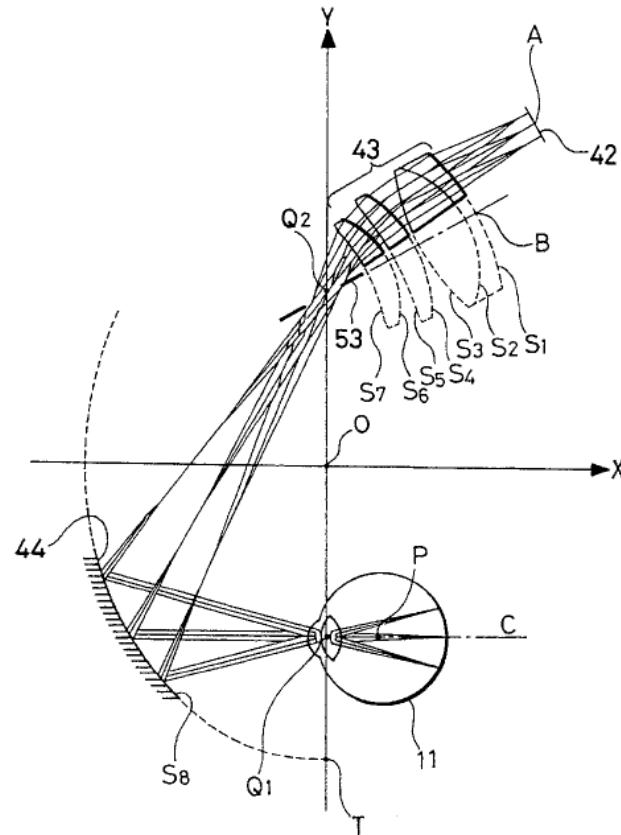
US 5,129,716

Jul. 14, 1992

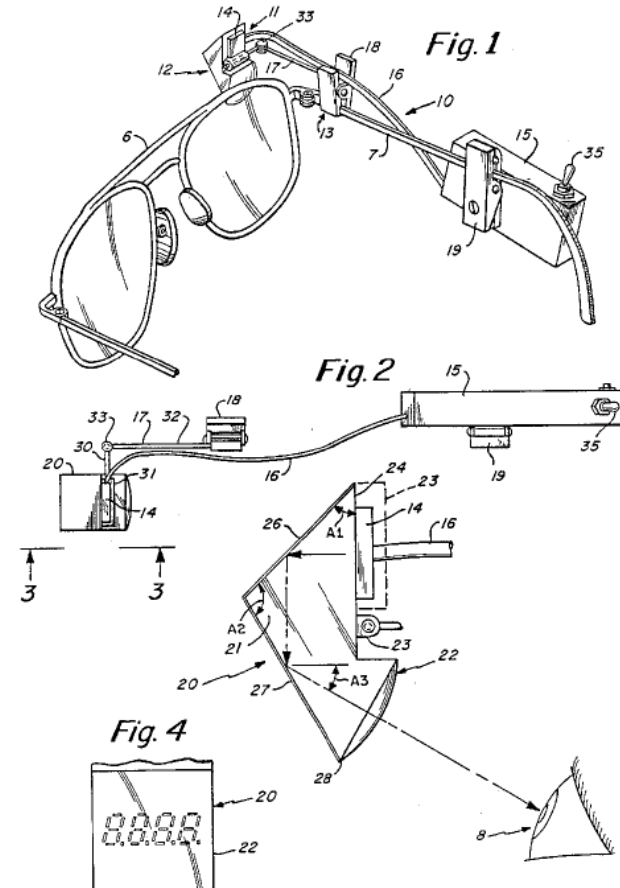


HWD Examples (Slide 7 of 18)

Iba. Image Observation Device.
US 5,384,654
Jan. 24, 1995



Kubik. Headwear-mounted
Periscopis Display Device.
US 4,753,514
Jun. 28, 1993

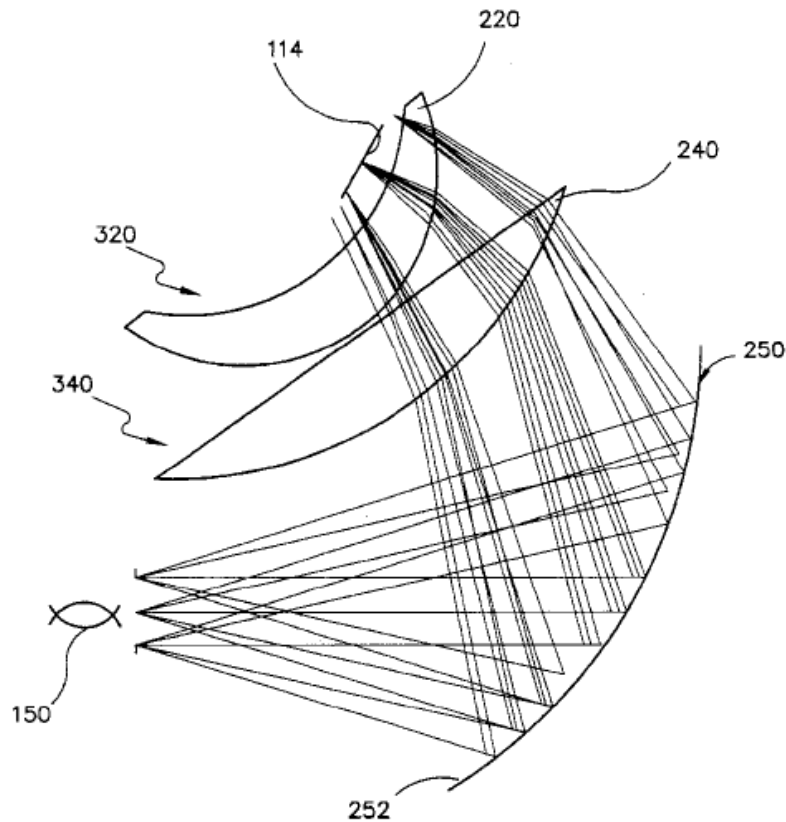


HWD Examples (Slide 8 of 18)

Ferrin. Headgear Display System
Using Off-axis Image Sources.

US 5,576,887

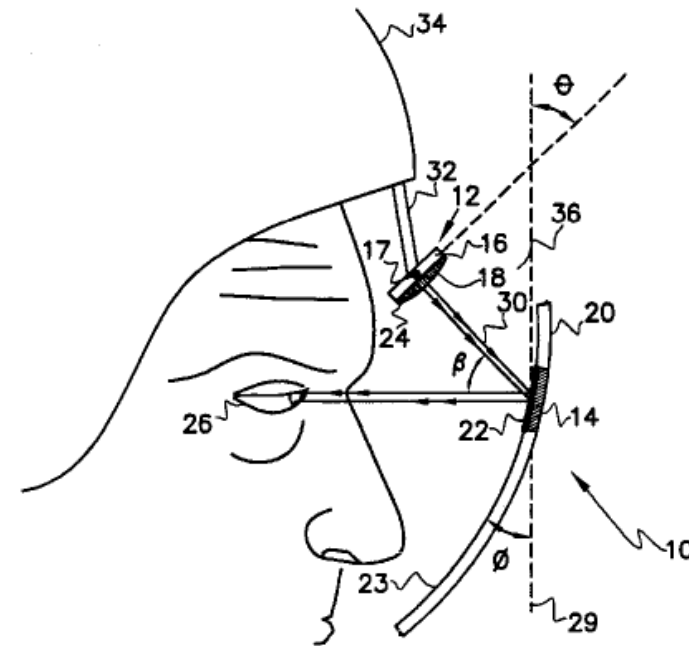
Nov. 19, 1996



Lippert. Visor Display with
Fiber Optic Faceplate Correction.

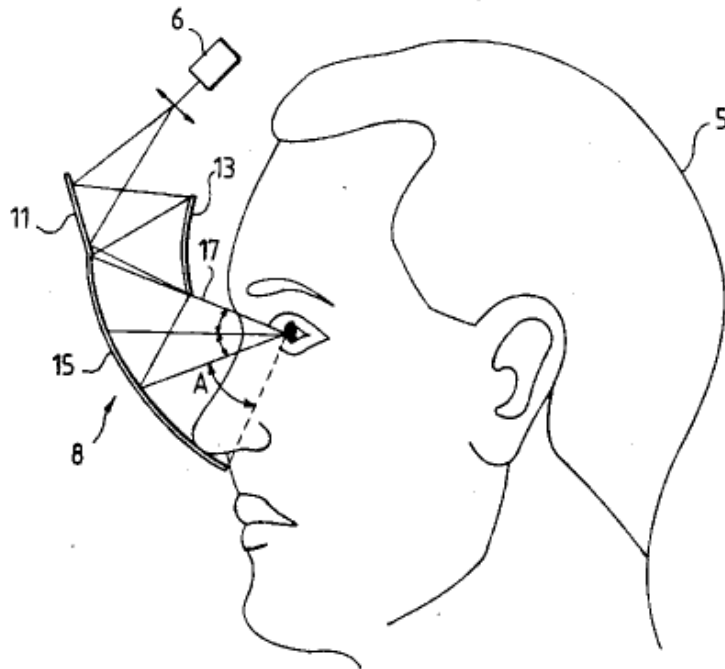
US 5,309,169

May 3, 1994

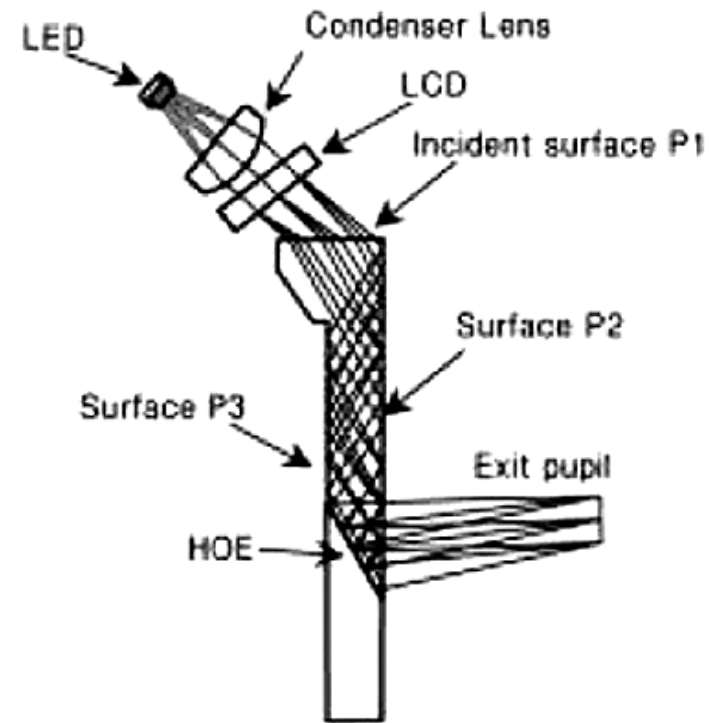


HWD Examples (Slide 9 of 18)

Lacroix. Device for the Display
of Simulated Images for Helmets.
US 5,184,250
Feb. 2, 1993



Kasai. A Forgettable Near-Eye Display.
ISWC 2000

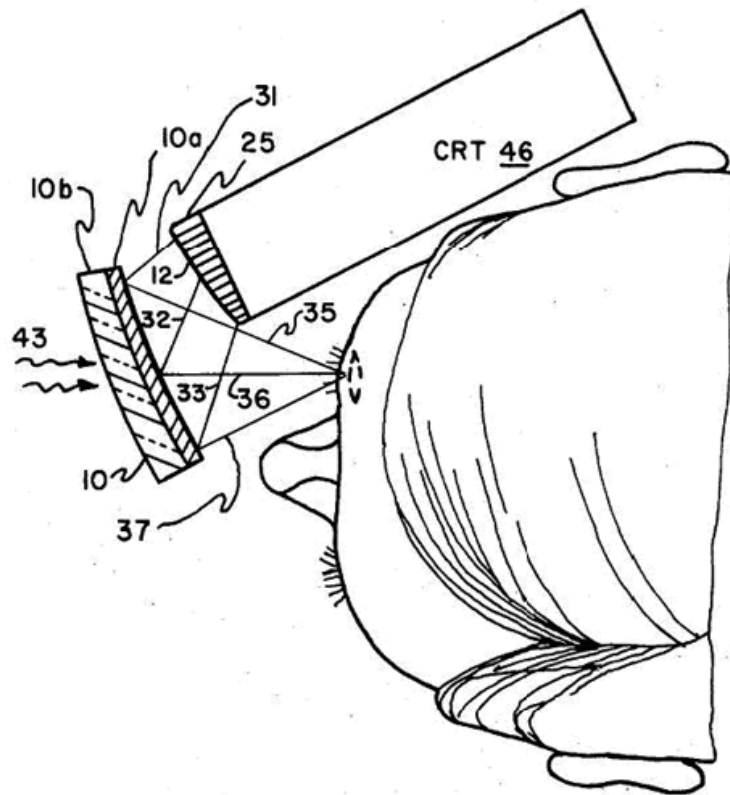


HWD Examples (Slide 10 of 18)

Bosserman. Toric reflector display.

US 4,026,641

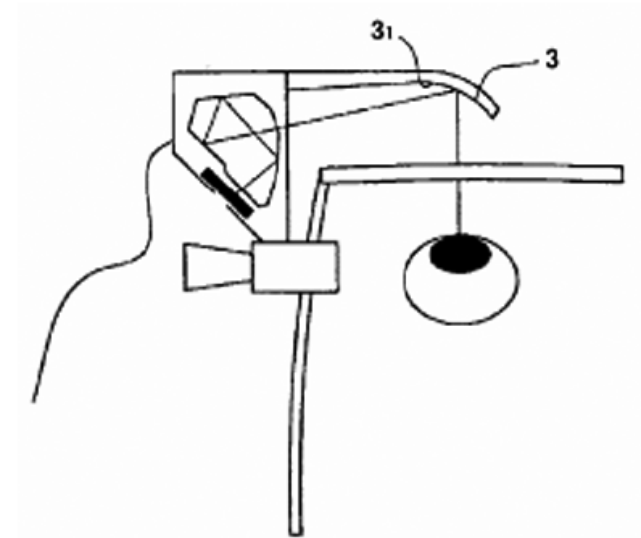
May 31, 1977



Nagaoka. Light weight head mounted image display device.

US 6,697,200

Feb. 24, 2004

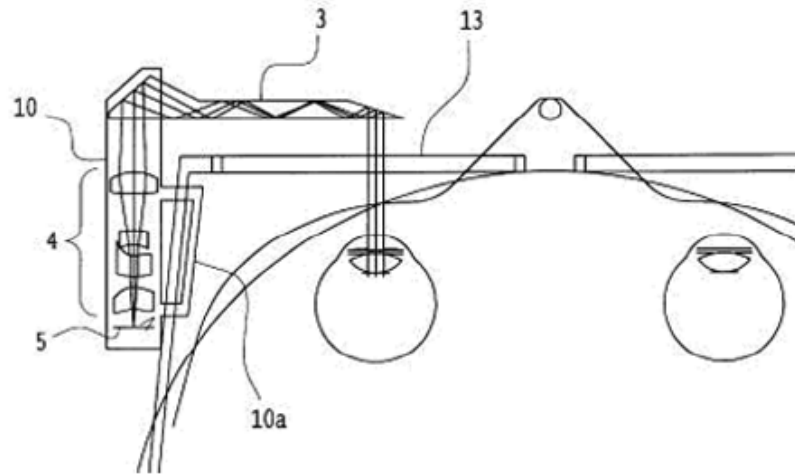


HWD Examples (Slide 11 of 18)

Takeyama. Observation optical system.

US 6,710,902

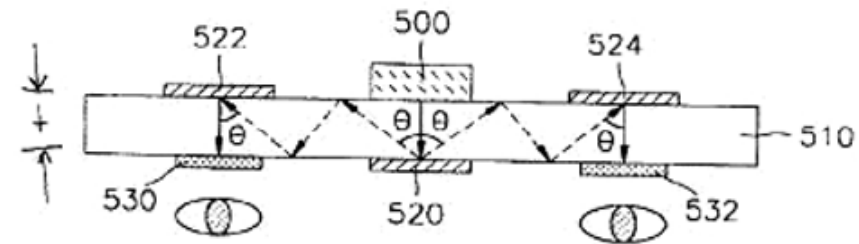
Mar. 23, 2004



Song. Wearable display system.

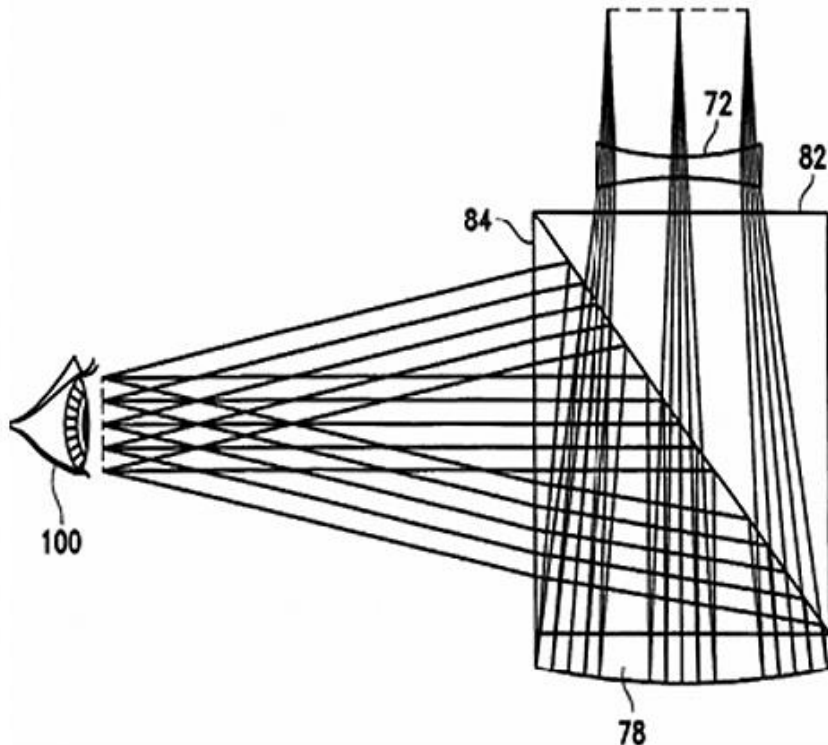
US 6,882,479

Apr. 19, 2005

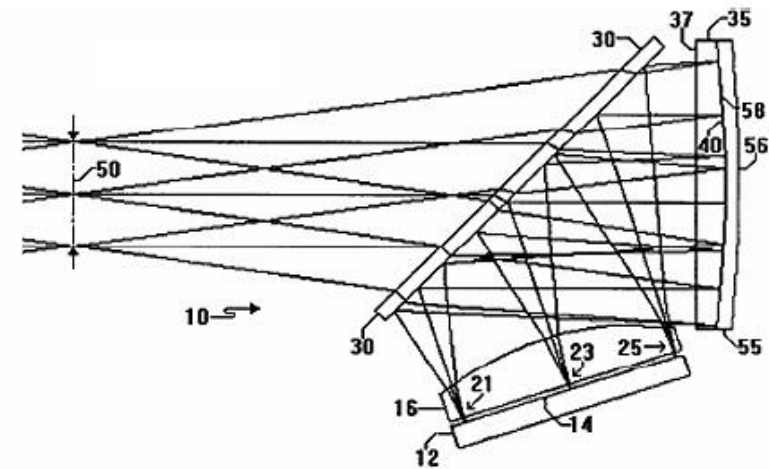


HWD Examples (Slide 12 of 18)

Robinson. Video headset.
US 5,696,521
Dec. 9, 1997

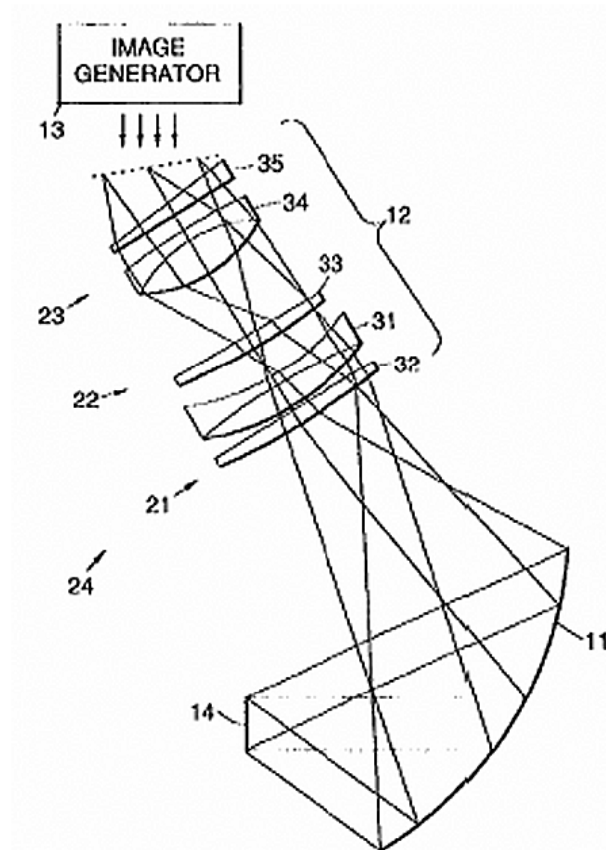


Fritz. Head mounted display
using mangin mirror combiner.
US 5,838,490

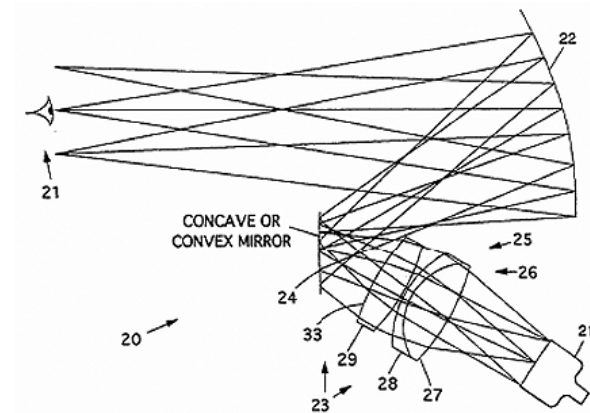


HWD Examples (Slide 13 of 18)

Chen. Helmet visor display employing reflective, refractive and diffractive optical components
US 5,526,183
Jun. 11, 1996



Chen. Wide spectral bandwidth virtual image display system.
US 5,436,763
Jul. 25, 1995

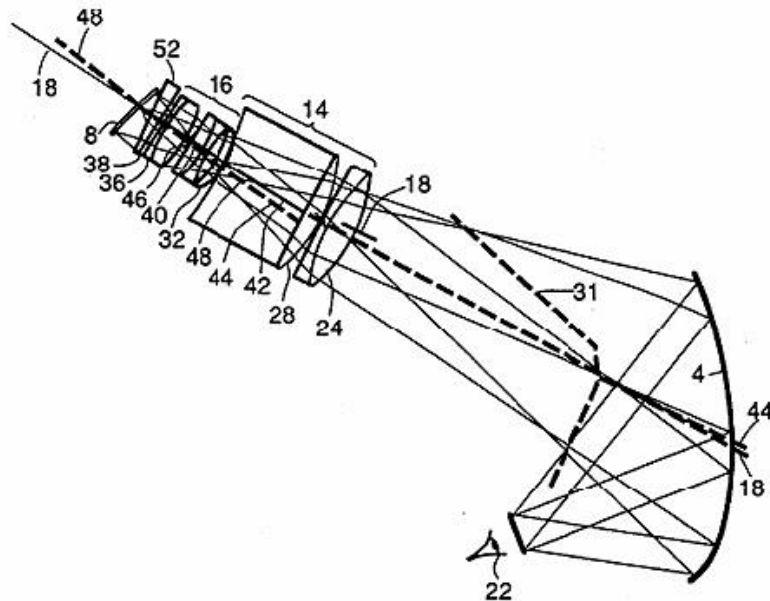


HWD Examples (Slide 14 of 18)

Chen. Ultra-wide field of view, broad spectral band visor display optical system.

US 5,499,139

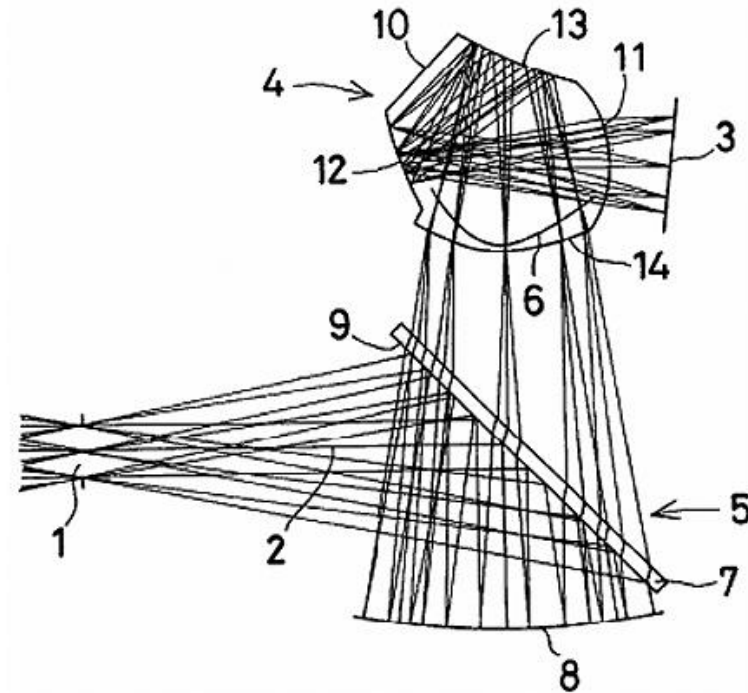
Mar. 12, 1996



Takeyama. Image display apparatus.

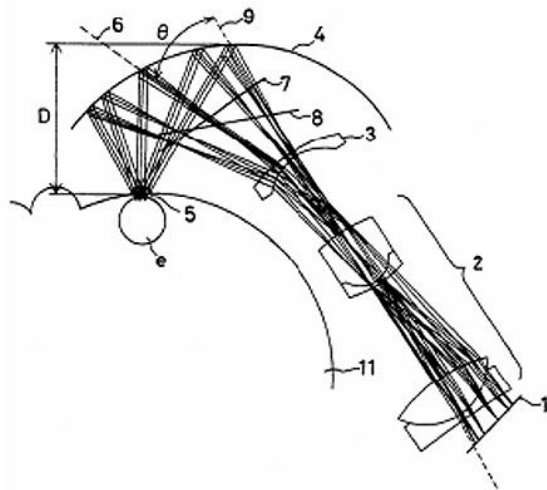
US 6,342,871

Jan. 29, 2002

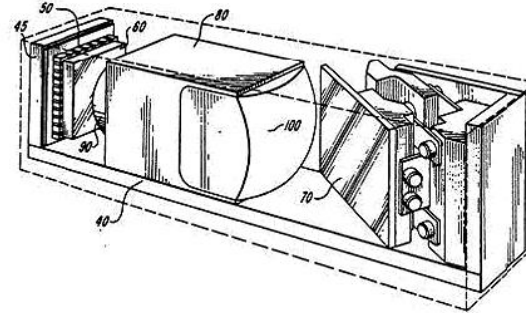


HWD Examples (Slide 15 of 18)

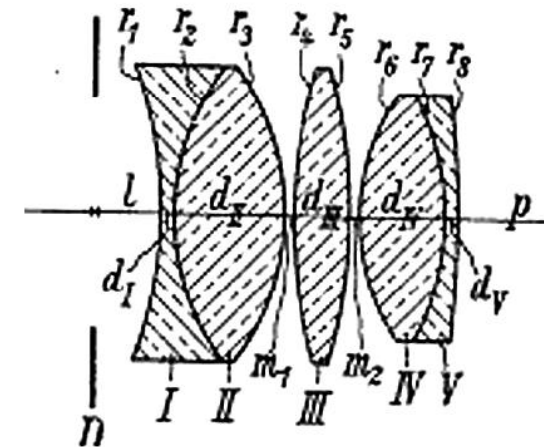
Togino. Visual display apparatus
US 5,436,765
Jul. 25, 1995



Becker. Head Mounted Display for
Minian Video Display System.
US 5,003,300
Mar. 26, 1991

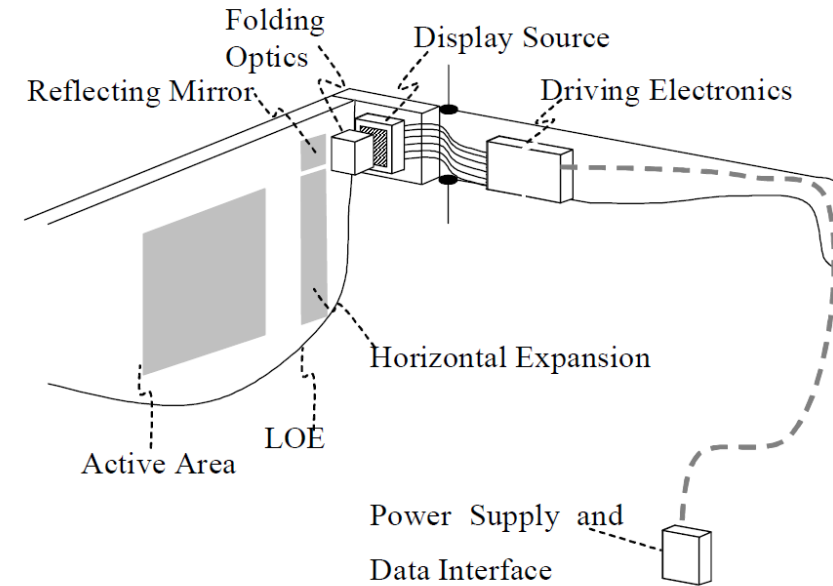
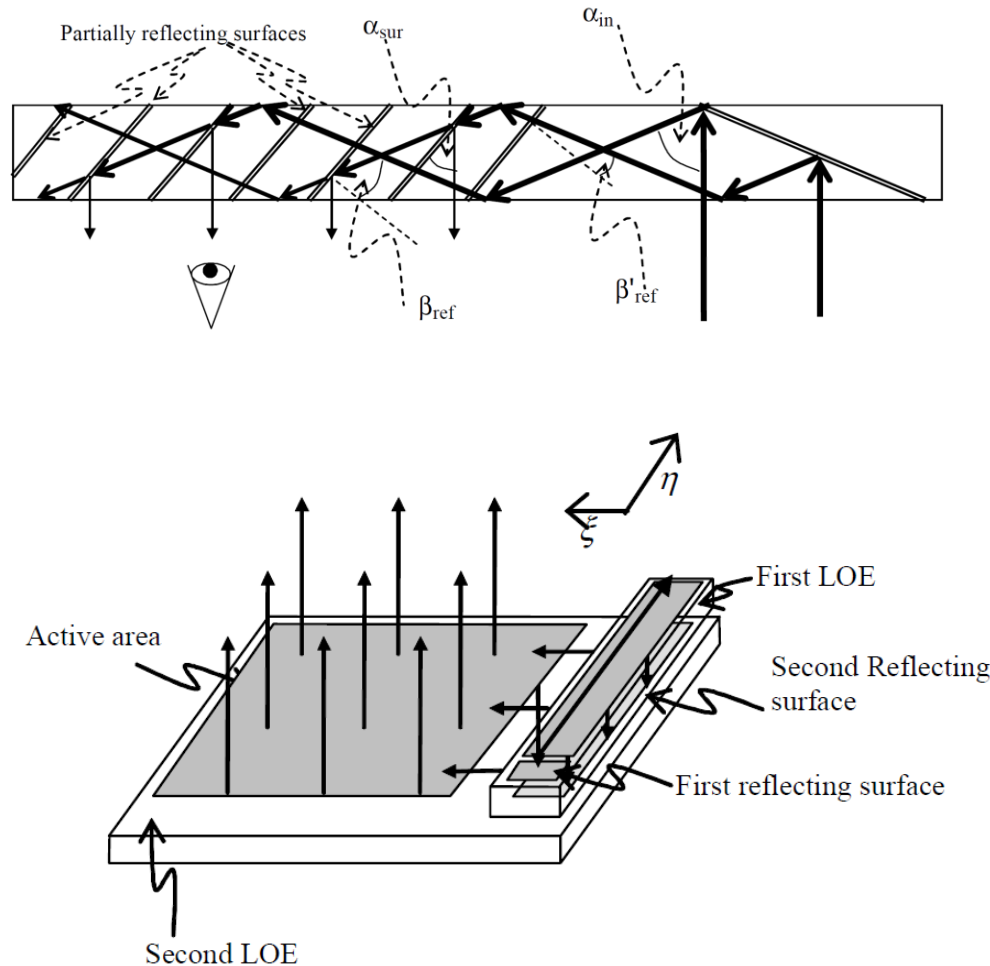


Erffle. Ocular.
US 1,478,704
Dec. 25, 1923



HWD Examples (Slide 16 of 18)

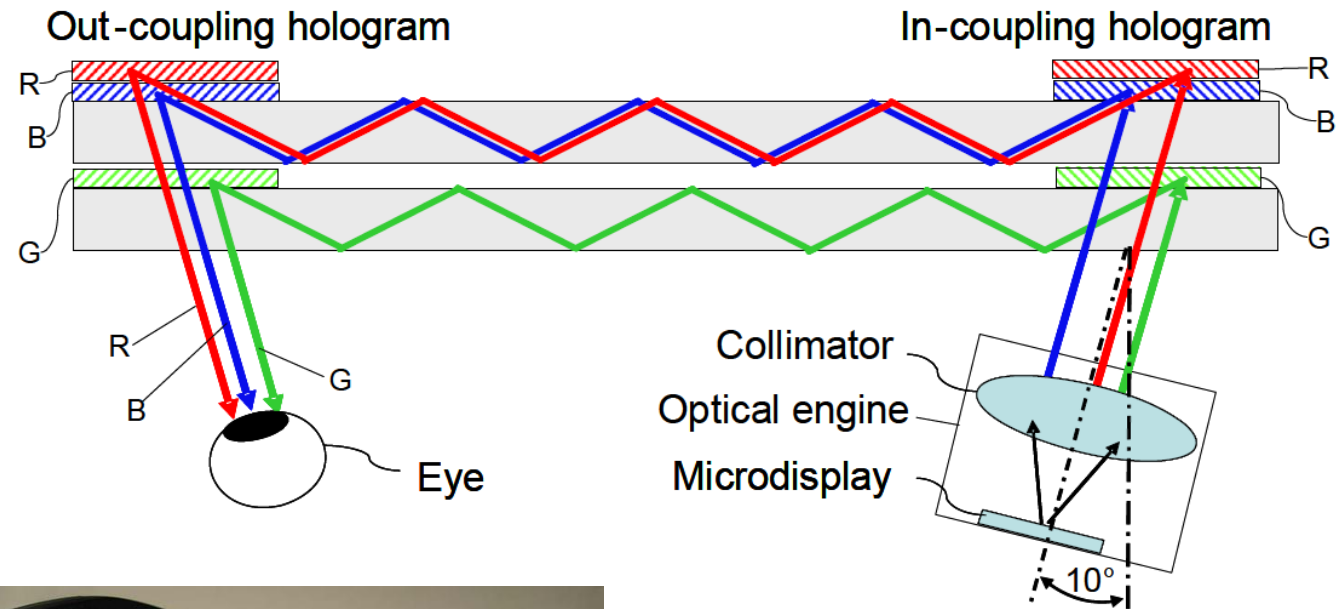
Waveguide with Cascaded Mirror-Array (Lumus)



Y. Amitai, SID Symposium Digest of
Technical Papers, 2005

HWD Examples (Slide 17 of 18)

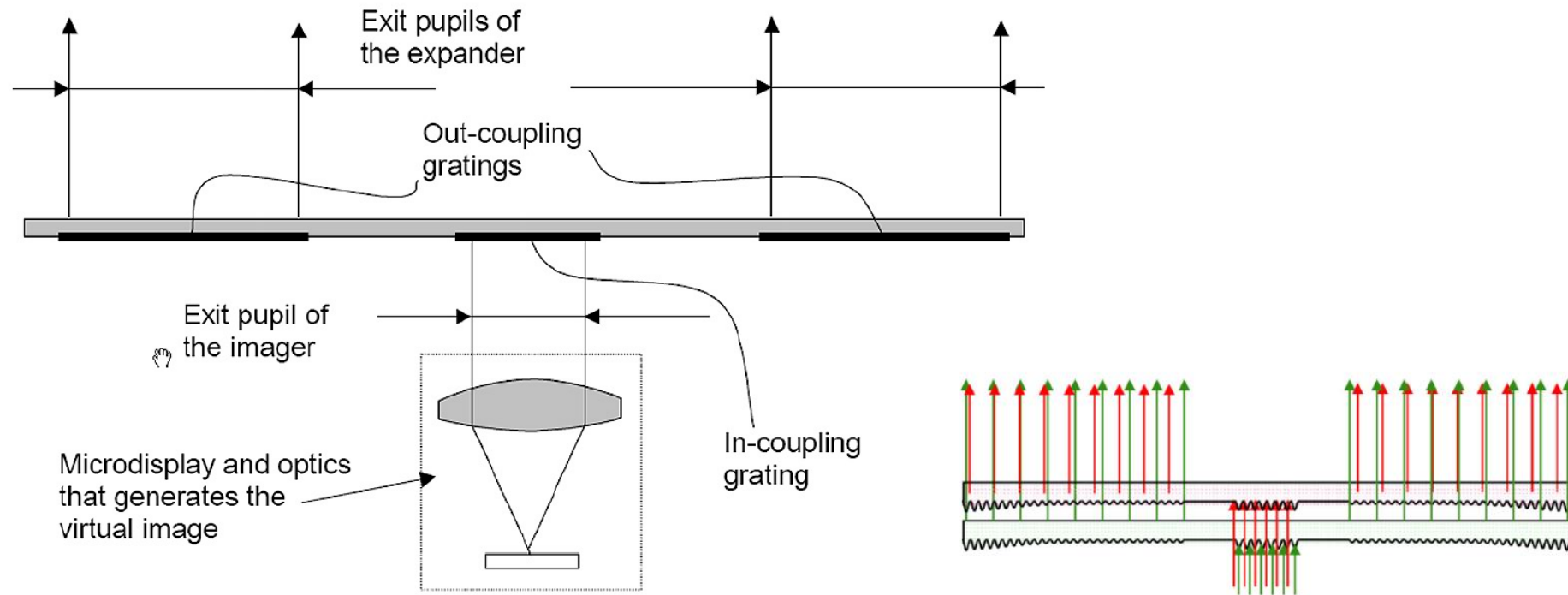
Holographic Planar Waveguide (Sony)



H. Mukawa, et al., SID Symposium
Digest of Technical Papers, 2008

HWD Examples (Slide 18 of 18)

Diffraction Waveguide (Nokia, Vuzix)



T. Levola, Journal of the SID 14/5,
2006