EYE TRACKING IN 3D ENVIRONMENTS

Gabriel J. Diaz
Kamran Binaee
Rakshit Kothari
Cayla Fromm

PERFORM LAB
(Perception For Movement)

ROCHESTER INSTITUTE OF TECHNOLOGY
R·I·T
1829

OUTLINE

- Active research
- Hardware
  - HMD
  - Motion capture
  - Eye trackers
- Vision science
  - Vergence accommodation
  - Correct perspective
- Software
  - Game engines
  - What you get for “free”
- Analysis
  - EIH to GIW
  - Angular distance
  - Reducing spatial inaccuracies
  - Spatio-temporal inaccuracies
  - Determining the object of fixation
  - Gaze event classification
- Jupyter notebooks
  - Examine calibration
  - Eye-in-head to gaze-in-world
  - Find saccades

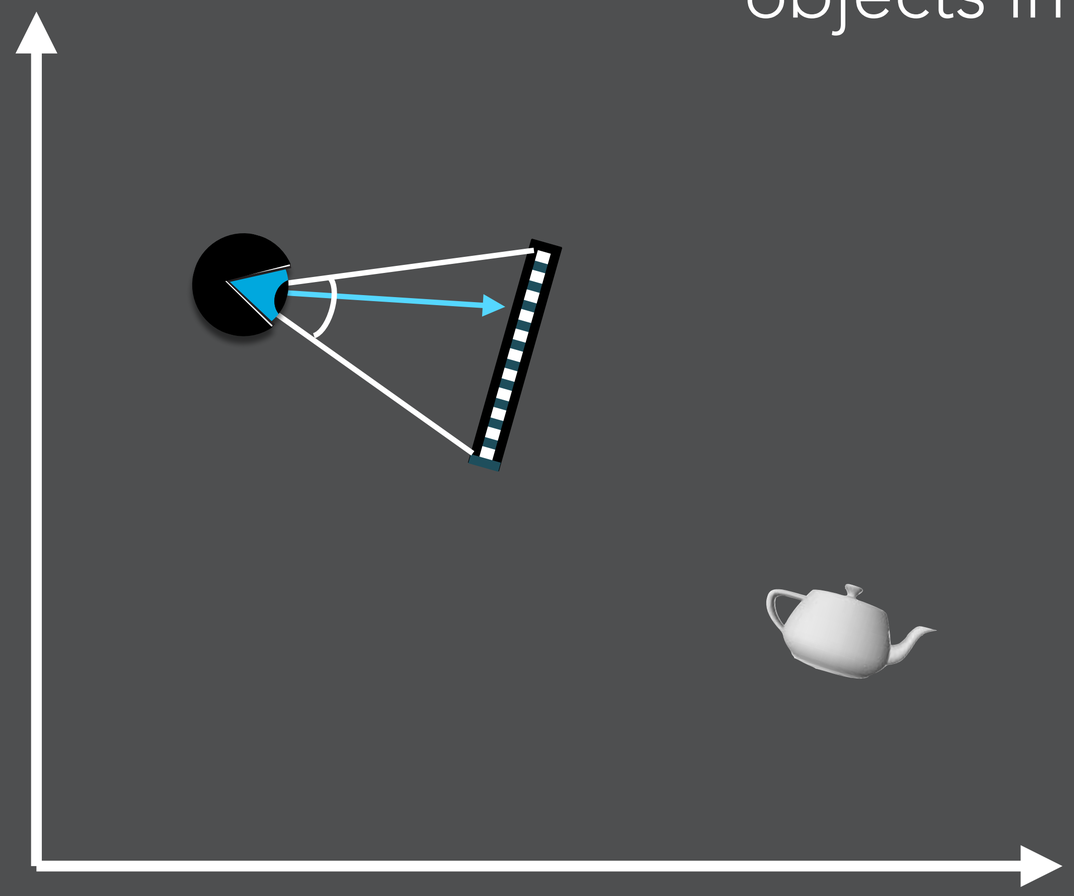
WHAT DOES VR BUY YOU?

EyeLink 2000

EYE MOVEMENTS WHILE HANDING BANKNOTES - PELZ & DIAZ

VR aligns both the head and objects within a world coordinate system.

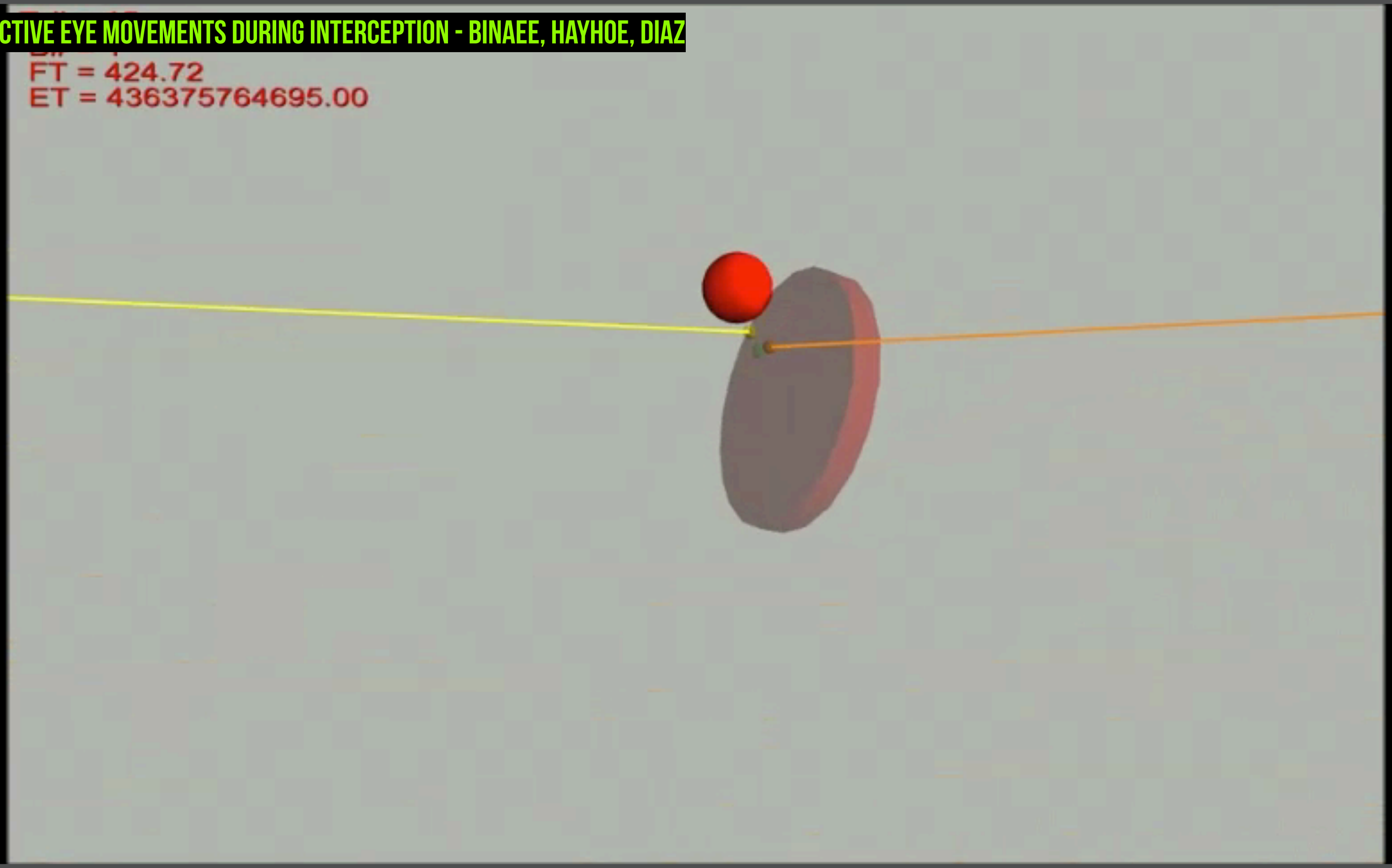
This facilitates computational analysis of gaze behavior with respect to objects in known locations.



ACTIVE RESEARCH

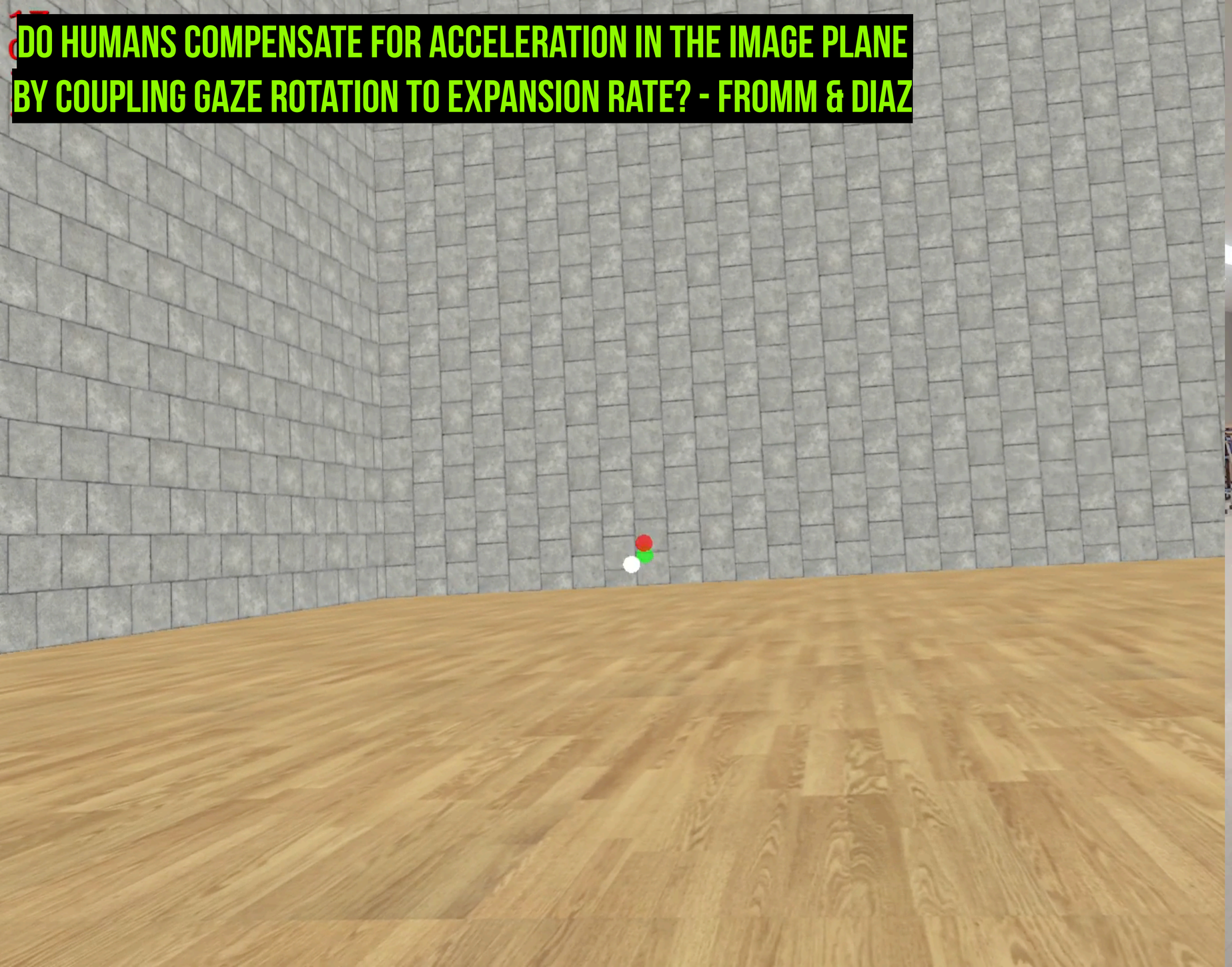
A small, biased sample.

PREDICTIVE EYE MOVEMENTS DURING INTERCEPTION - BINAEE, HAYHOE, DIAZ



VISUALLY GUIDED WALKING - DIAZ

DO HUMANS COMPENSATE FOR ACCELERATION IN THE IMAGE PLANE BY COUPLING GAZE ROTATION TO EXPANSION RATE? - FROMM & DIAZ



VISUAL REHABILITATION - HUXLIN, FROMM, DIAZ

[A work in progress!]

EQUIPMENT - HEAD MOUNTED DISPLAYS

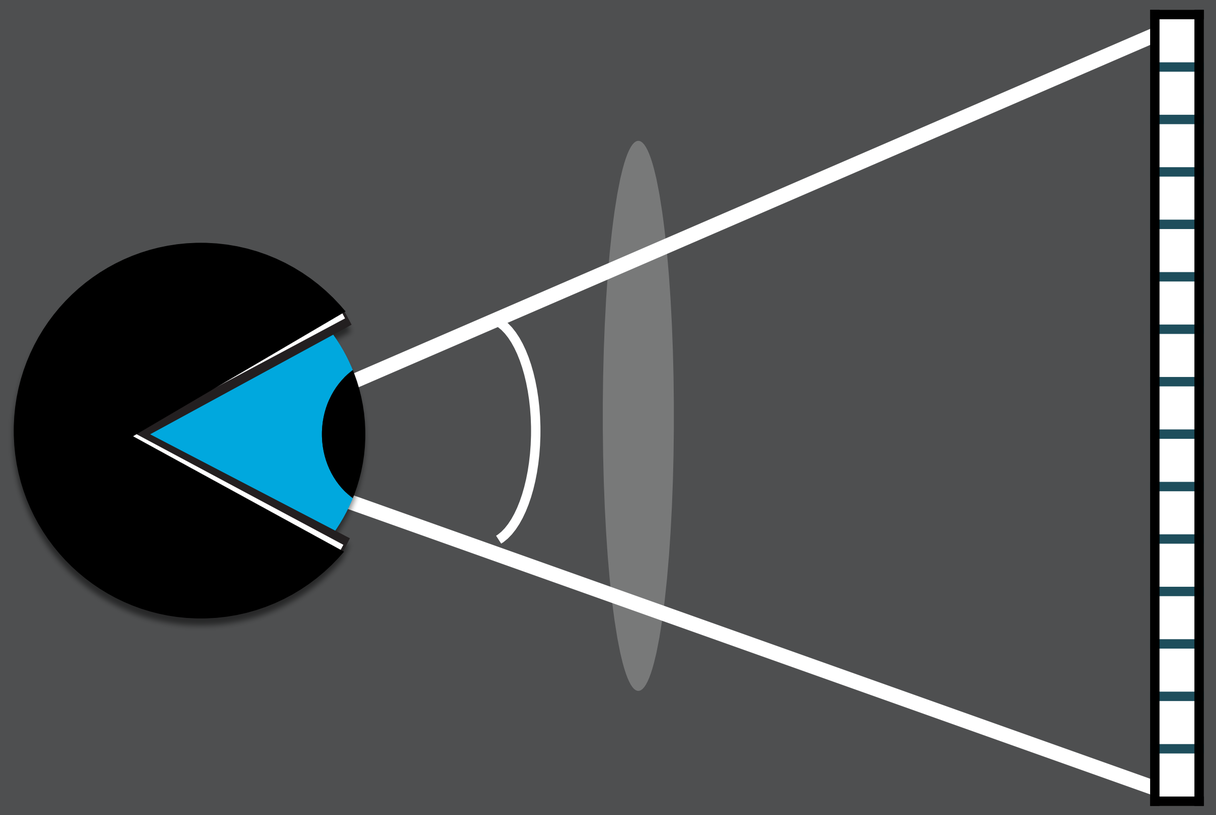
OCULUS RIFT

- Resolution: 2160 x 1200 pixels
  - (two 1080x1200 AMOLED displays)
  - After optics: **13.8 pixels/degree**
- 90 Hz refresh rate
- ~90 degrees FOV (binocular)
  - approx. 70 degrees monocular

HTC VIVE

- Resolution: 2160 x 1200 pixels
  - (two 1080x1200 AMOLED displays)
  - After optics: **10.8 pixels/degree**
- 90 Hz refresh rate
- ~110 degrees FOV (binocular)
  - approx. 90 degrees monocular

WHAT ANGULAR RESOLUTION?
(PIXELS PER DEGREE)



REAL RESOLUTION DEPENDS ON THE OPTICS

Optics in VR...

Need to be: (pick two!)

Small diameter

Lightweight

Short focal length

Large field of view

Inexpensive

Aberration free

We *would like* them to be:

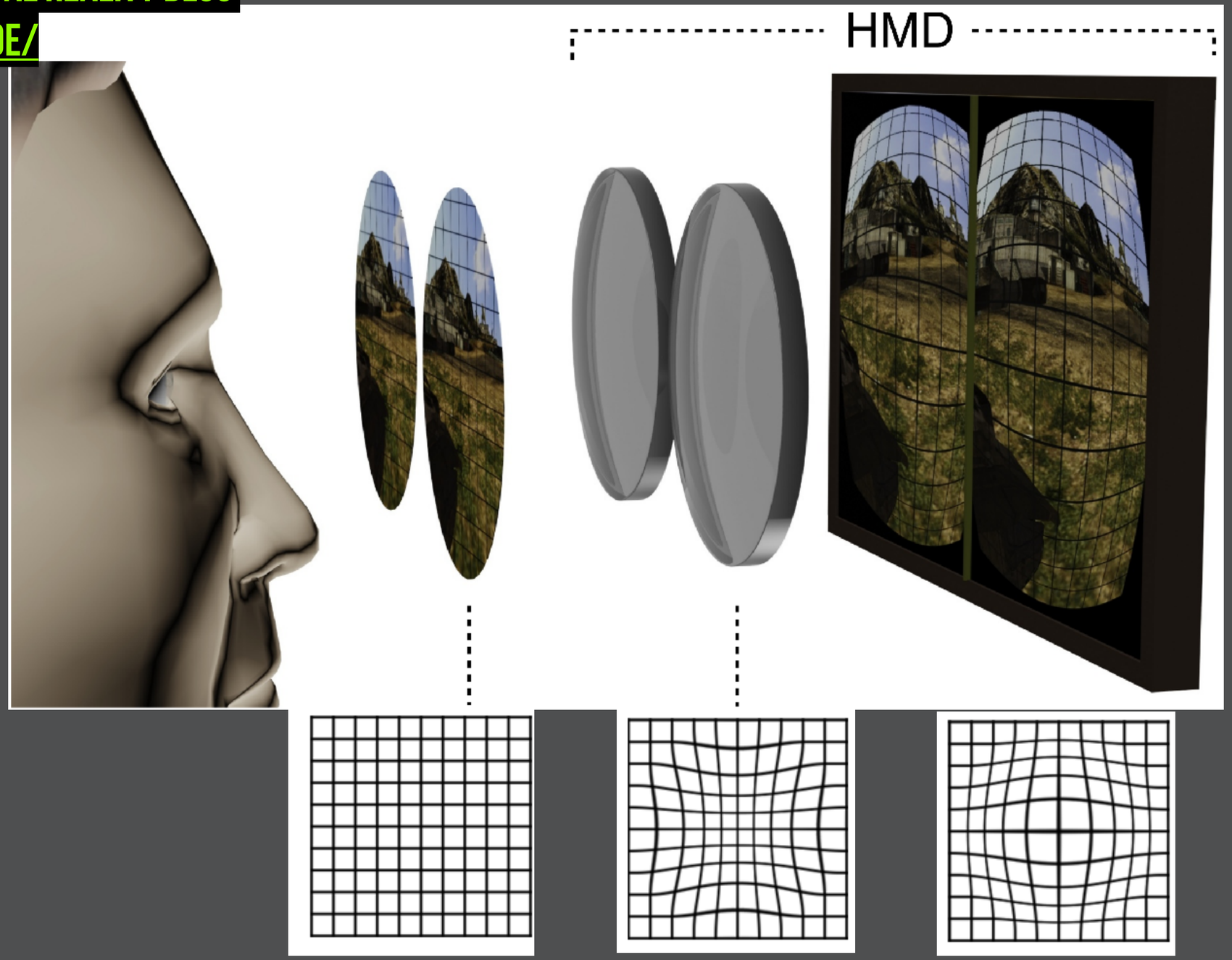
Dynamic focal length

Dynamic optical axis

Fresnel lens used in HTC Vive

Image from https://www.ifixit.com/Guide/HTC+Vive+Lens+Replacement/98127

DANIEL POHL'S VIRTUAL REALITY BLOG
HTTP://BLOG.QWRT.DE/



should be higher resolution at the EDGE to have uniform all o=on the display, see doc ok

An example of the HMD image before lensing.

There is a higher pixel density in the center of the visual field.

WWW.DOC-OK.ORG

Pincushion rendering in the Vive

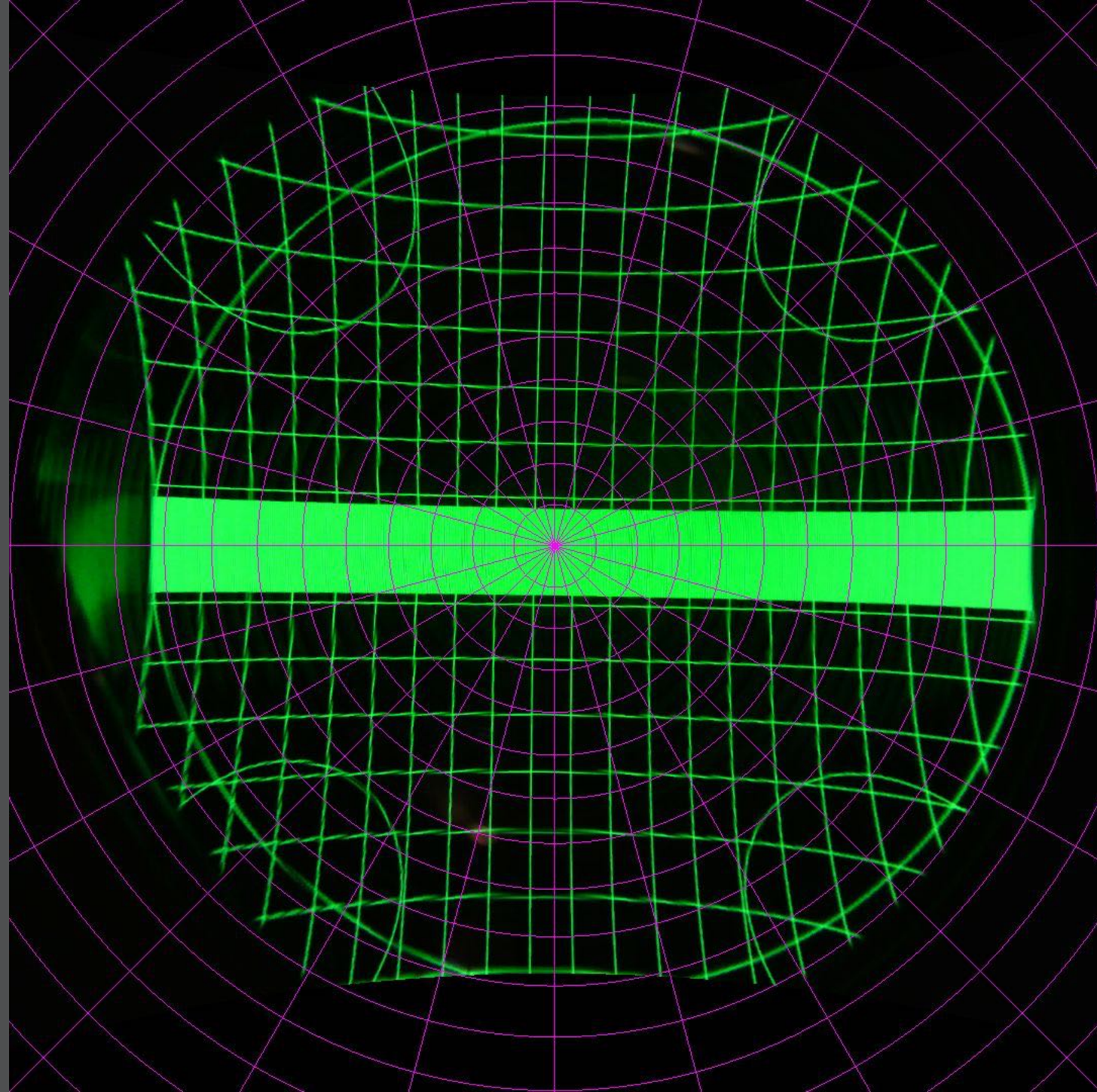
Camera 10 mm from lens

Test pattern rendered in green

- Green squares 53 x 75 pixels

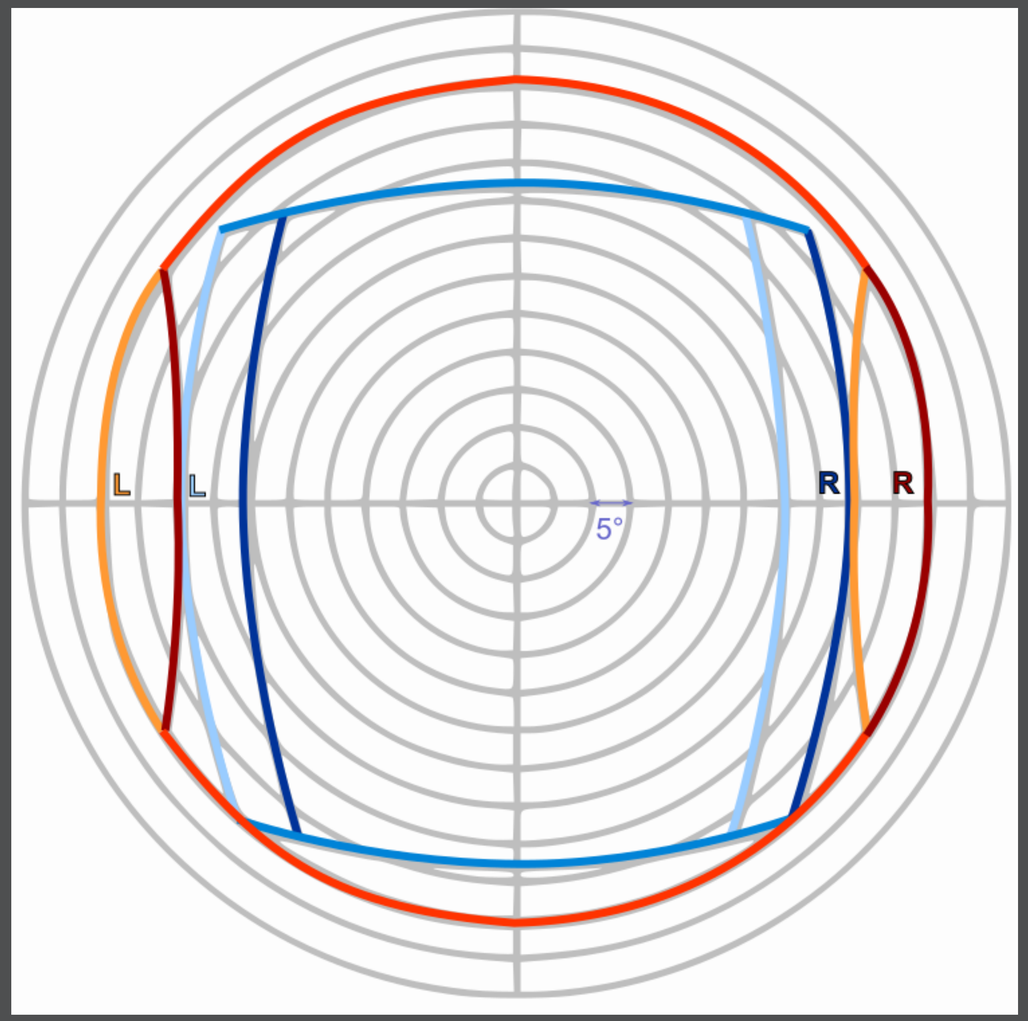
Purple grid for reference

- Purple circles 5° apart



TRADE-OFF BETWEEN RESOLUTION AND FIELD OF VIEW

- Where to put the most pixels and which optics to use to:
  - Maximize FoV?
  - Maximize useful resolution ?



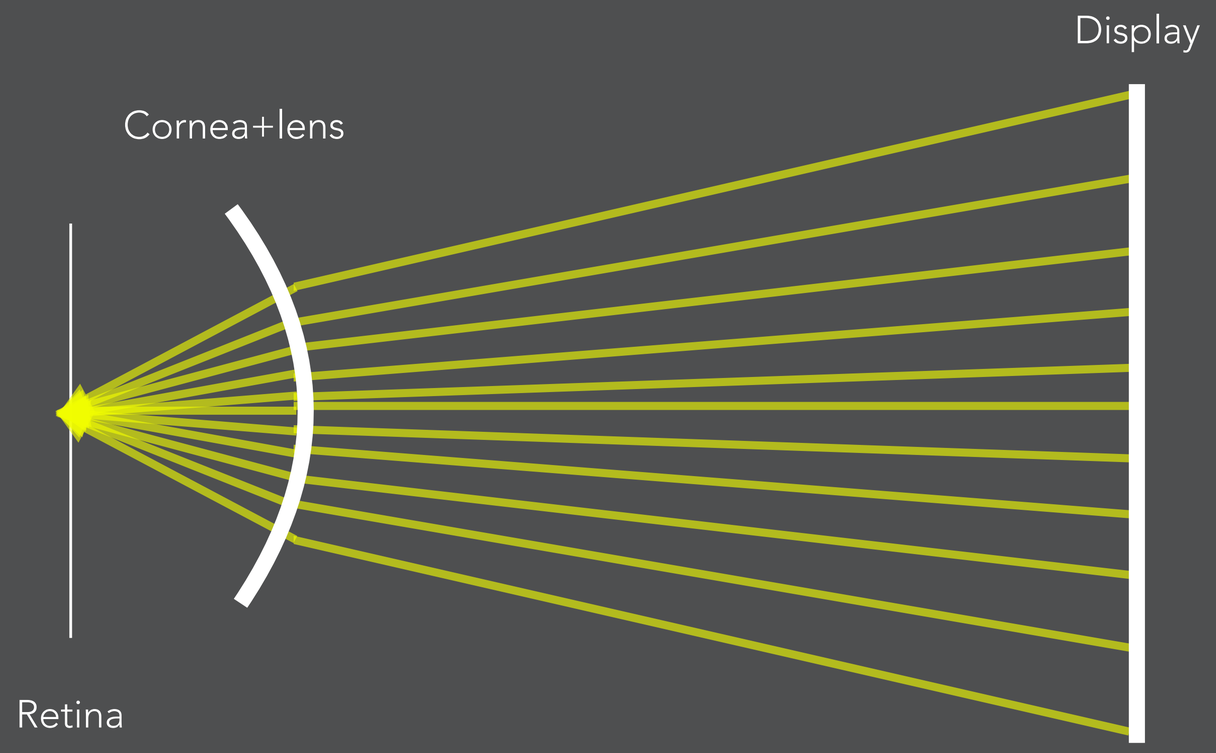
Rift FoV in blue, Vive FoV in red, 5 degree rings

Figure adapted from: “The Display Resolution of Head-mounted Displays, Revisited” by Oliver Kreylos, www.doc-ok.org

VR LACKS DEPTH FROM FOCAL CUES

Blur and accommodation.

VR is a single focal-plane display.
Although VR helmets present a different image to each eye, the images come from common depth.



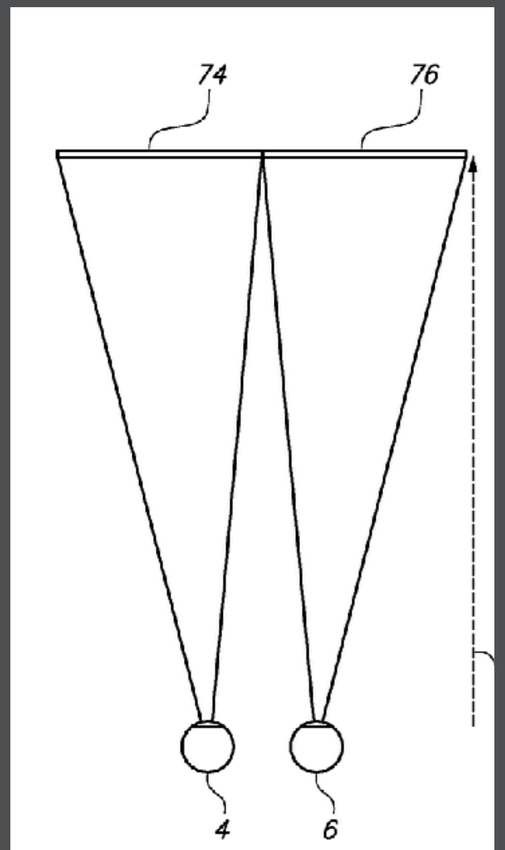
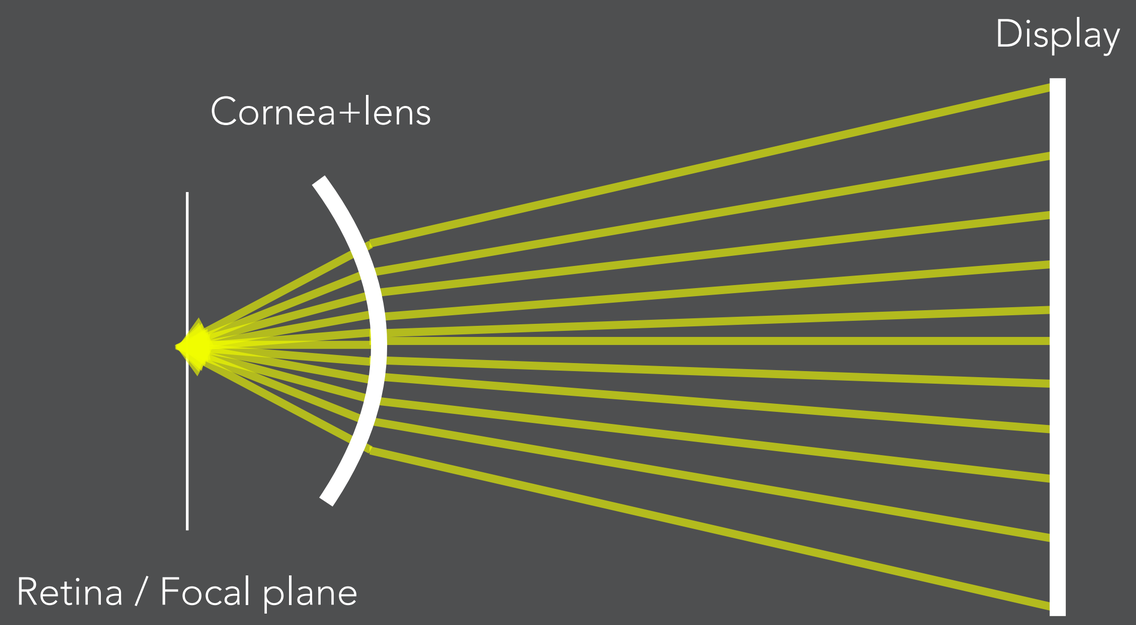
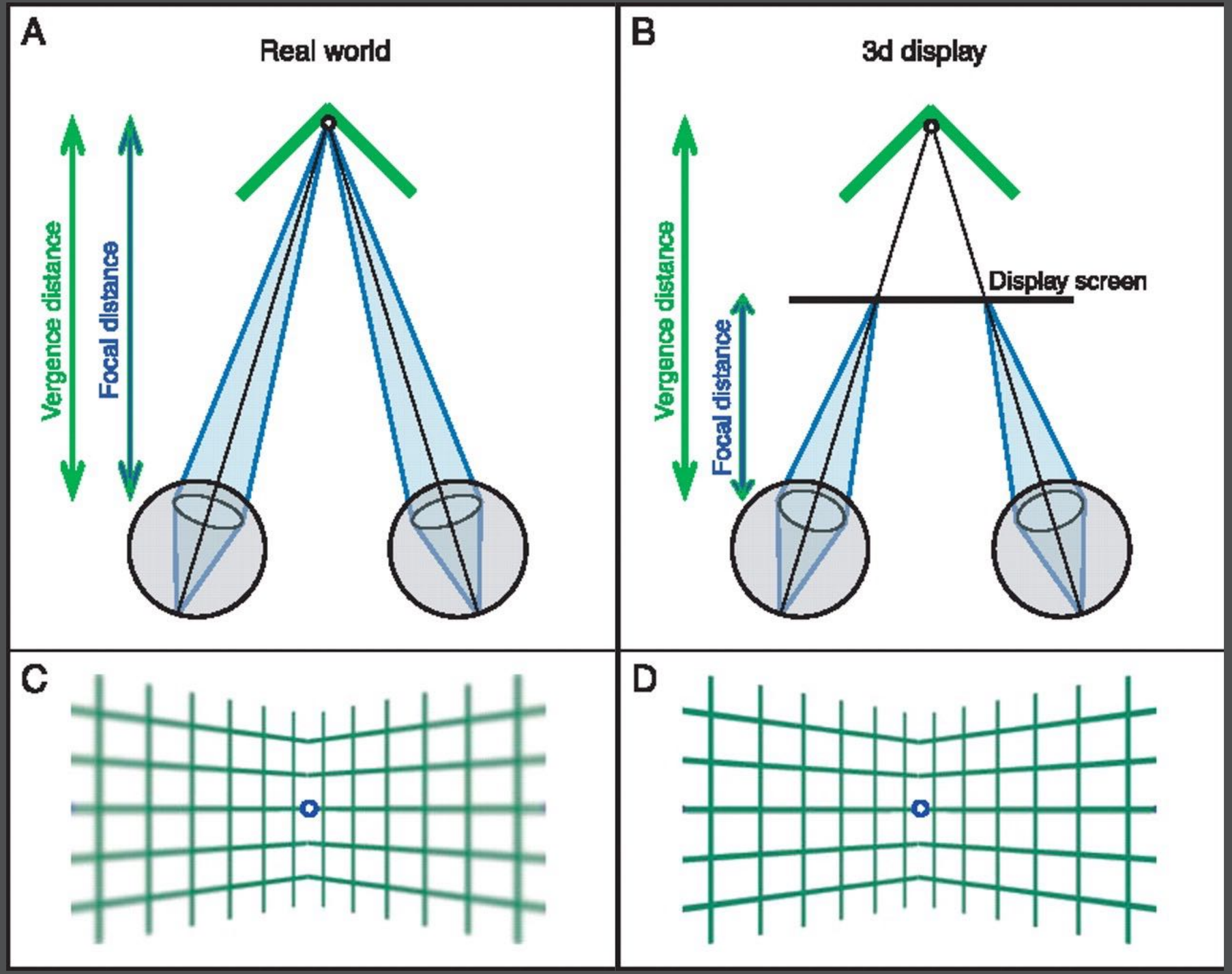


FIG. 2

BLUR IS NOT PRESENT IN SINGLE FOCAL PLANE DISPLAYS



In VR, one only needs to accommodate to the depth the display.
Because the display is at a single depth, everything will appear uniformly sharp and in focus.



Single focal plane displays don't typically replicate blur.

* Figure from Marty Banks

ROBERT T. HELD, EMILY A. COOPER, JAMES F. O'BRIEN, AND MARTIN S. BANKS. "USING BLUR TO AFFECT PERCEIVED DISTANCE AND SIZE". ACM TRANSACTIONS ON GRAPHICS, 29(2):19:1–16, MARCH 2010.

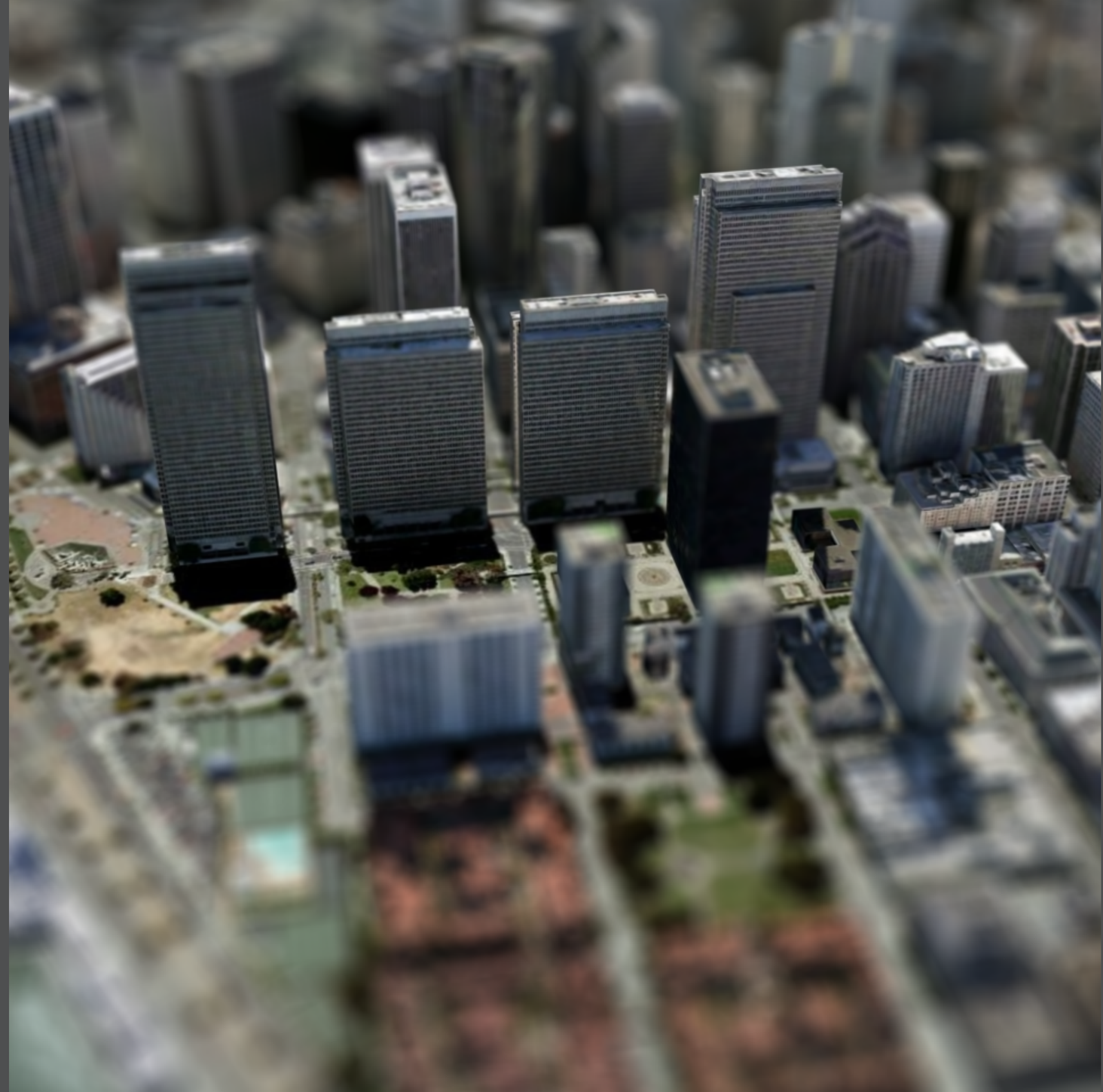
Because depth of field normally increases with distance, the blur in the above image suggests that we are focusing upon a small object that is nearby.

Pinhole image (no blur)

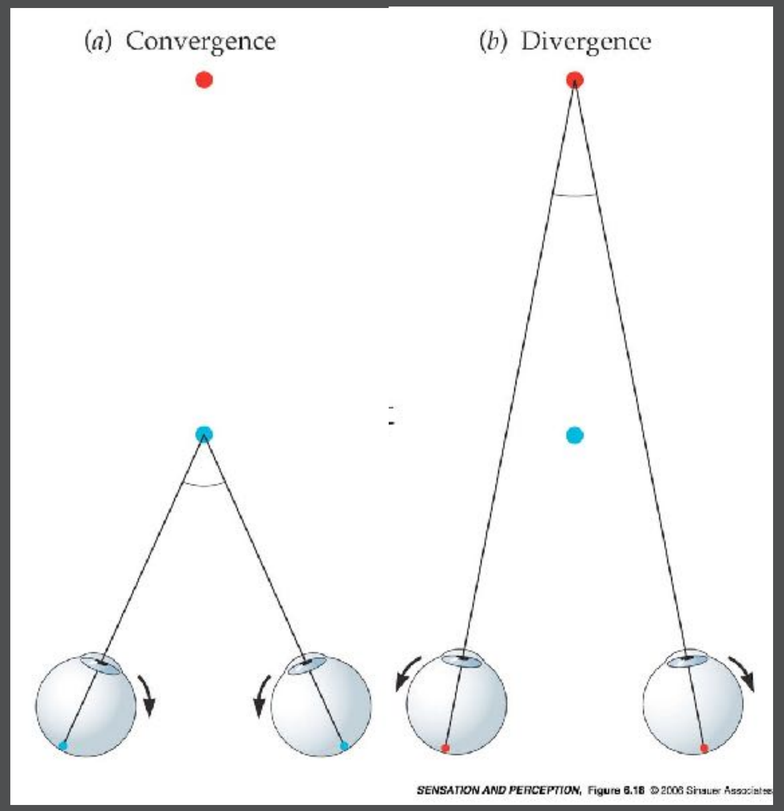
0.36m Simulated Focal Distance

0.15m Simulated Focal Distance

0.6m Simulated Focal Distance

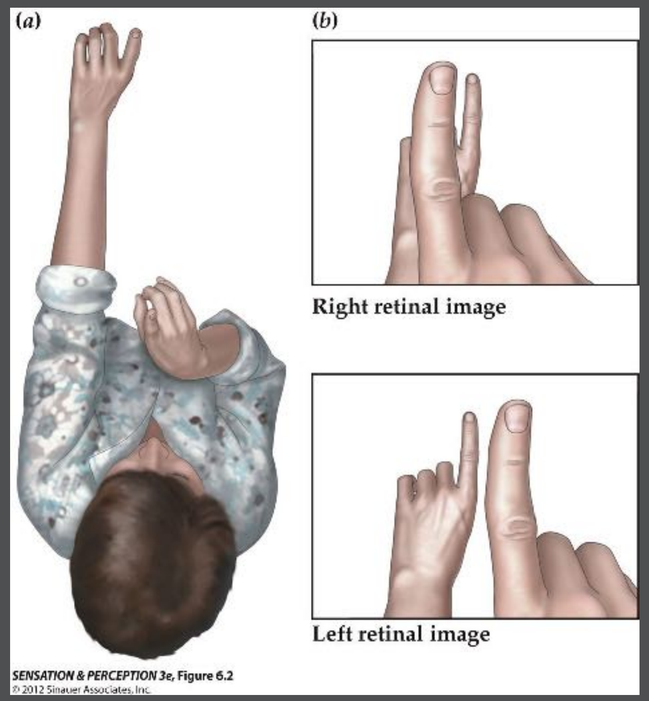


Disparity and blur act as inputs into the system that drives vergence eye movements.



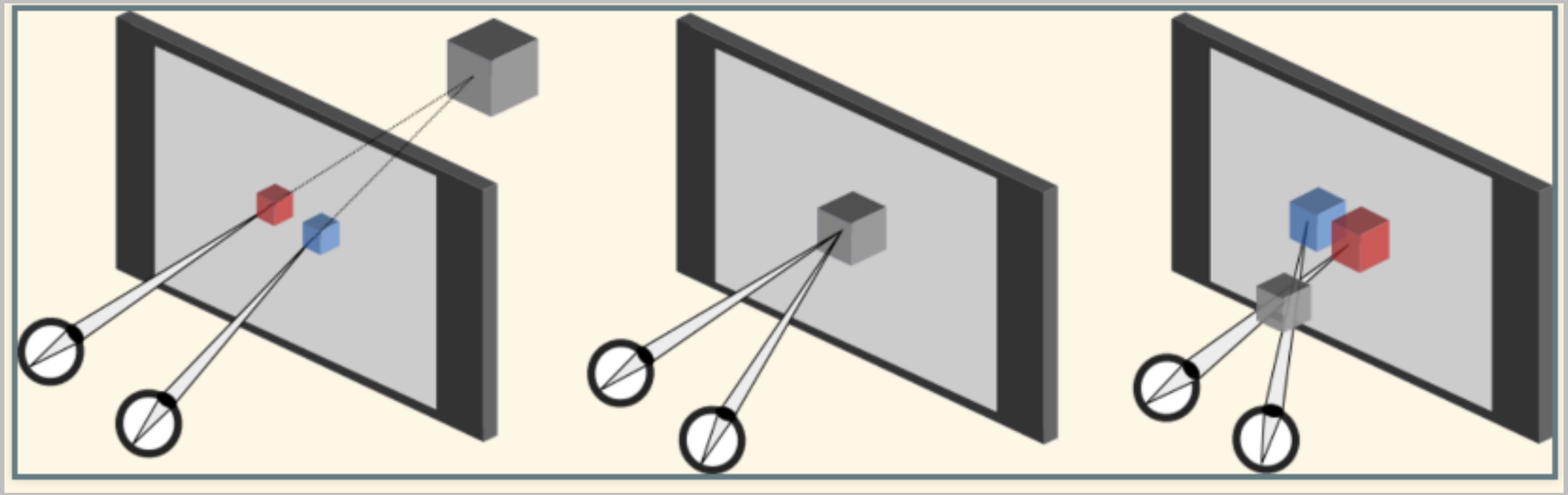
Blur.

Disparity.



To "converge" upon an object in depth, rotate your eyes until disparity and blur are minimized.

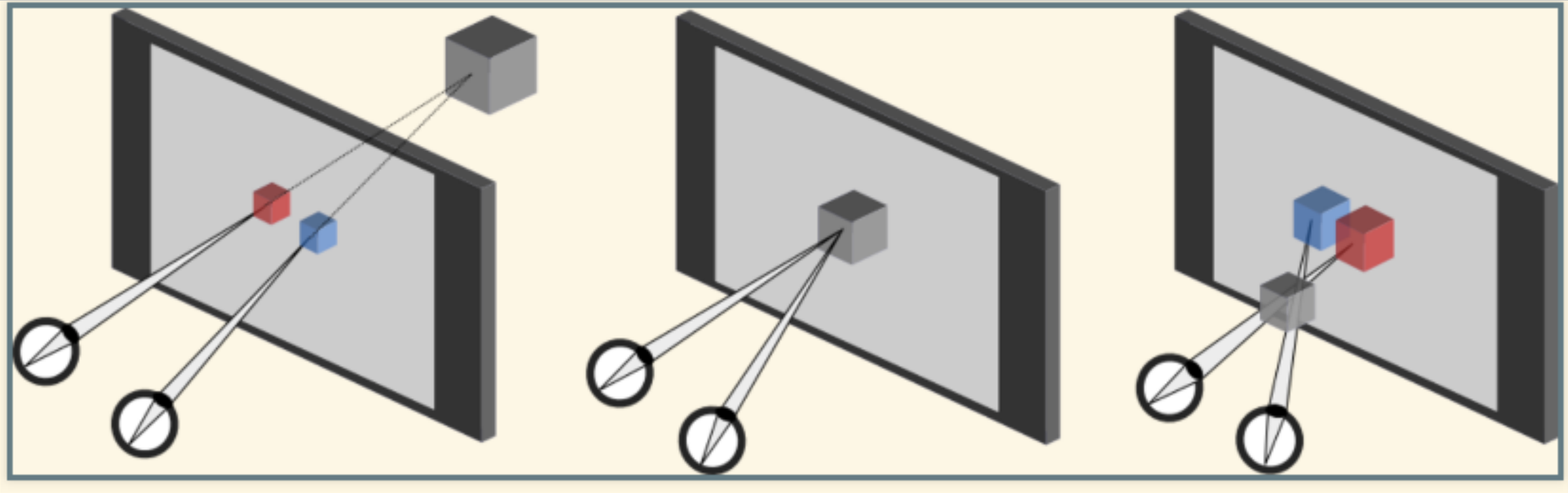
VERGENCE/ACCOMMODATION CONFLICT



Frey, Appriou, Lotte, Hachet, "Classifying EEG signals during stereoscopic visualization to estimate visual comfort" Computational intelligence and neuroscience (2017)

Accommodation is appropriate for a single depth - that of the screen.
Disparity signals suggest that objects exist at different depths.

VERGENCE/ACCOMMODATION CONFLICT



Frey, Appriou, Lotte, Hachet, "Classifying EEG signals during stereoscopic visualization to estimate visual comfort" Computational intelligence and neuroscience (2017)

Little is known of how V/A conflict affects eye movements.

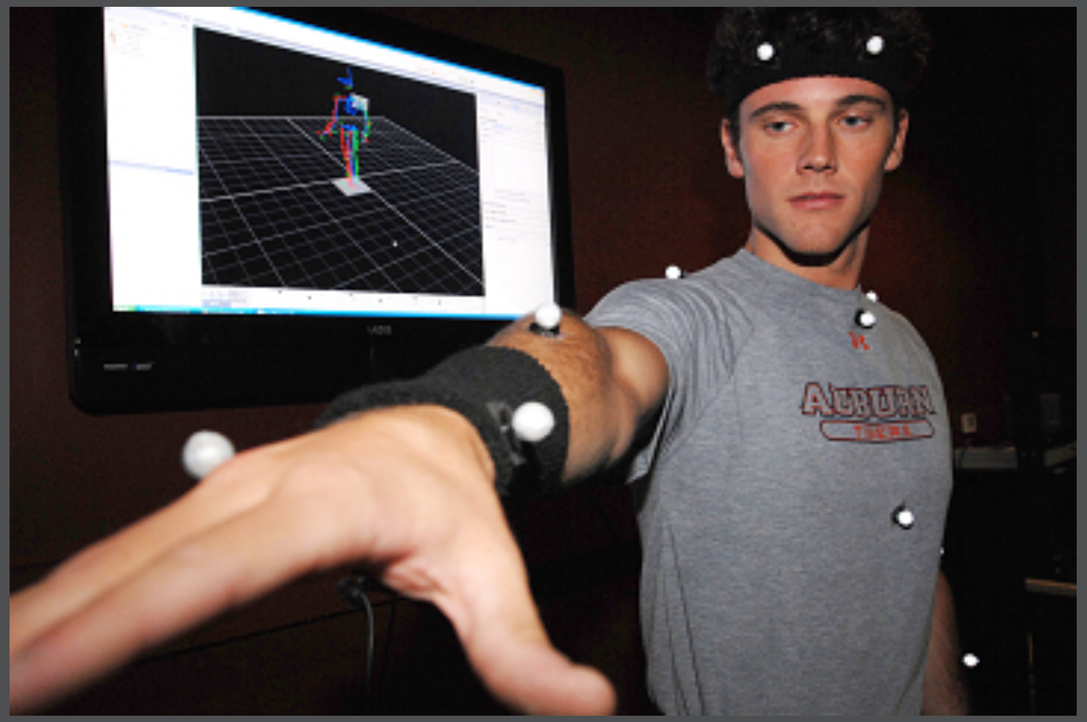
EQUIPMENT - MOTION CAPTURE

OPTICAL - PASSIVE MARKER SYSTEMS

qualisys

vicon

motion analysis



OPTICAL ACTIVE MARKER SYSTEMS

Phasespace

Northern digital “optotrak”

MOBILE SYSTEMS (SUIT BASED)

MotionStar

magnetic

Motion Shadow

imu's

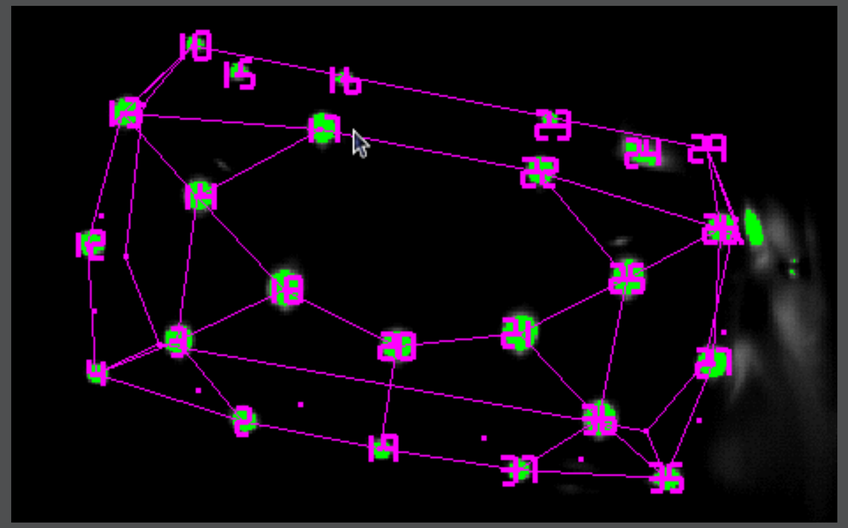
Xsens

imu's

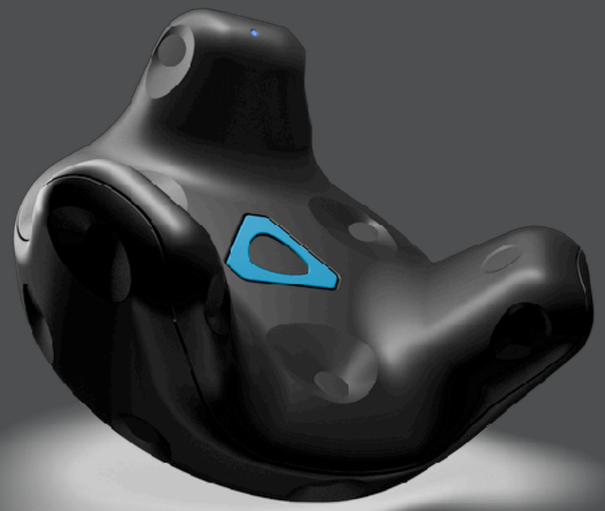
HEAD + HAND TRACKING FOR VR

HTC Vive

HTC Vive



HTC POSITION TRACKING



HYPERKIN
DUCK
SEASON

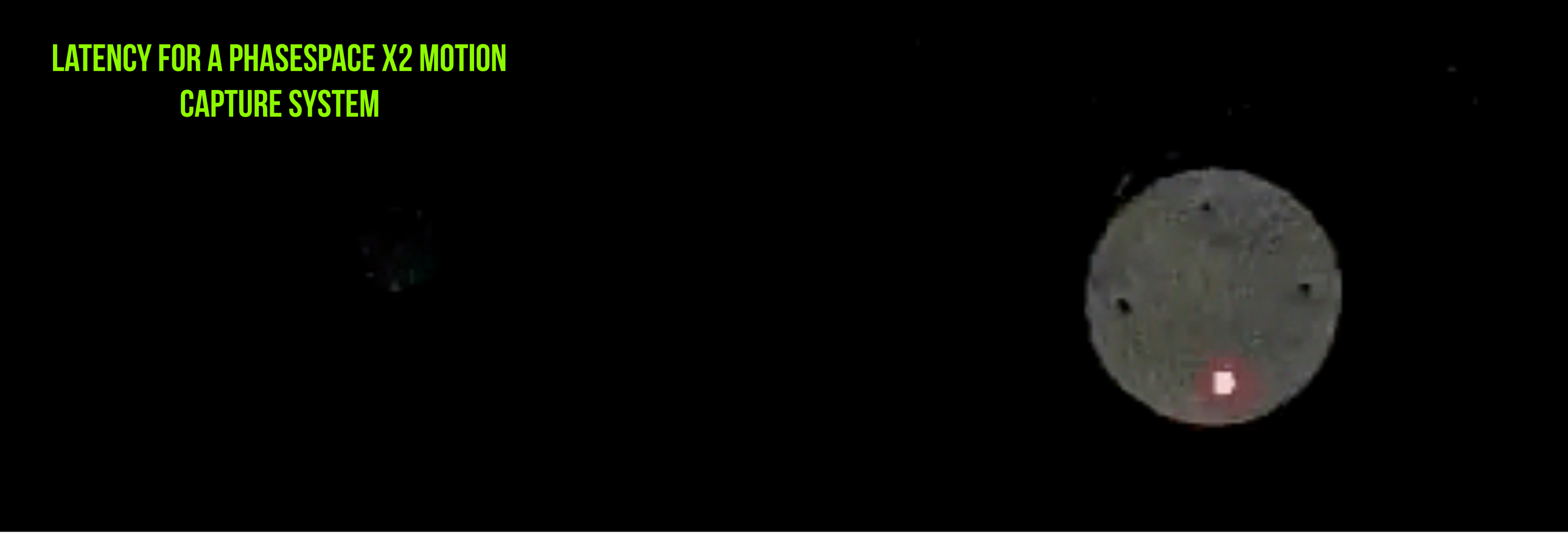
HTC LIGHTHOUSE TRACKING - TIME BASED

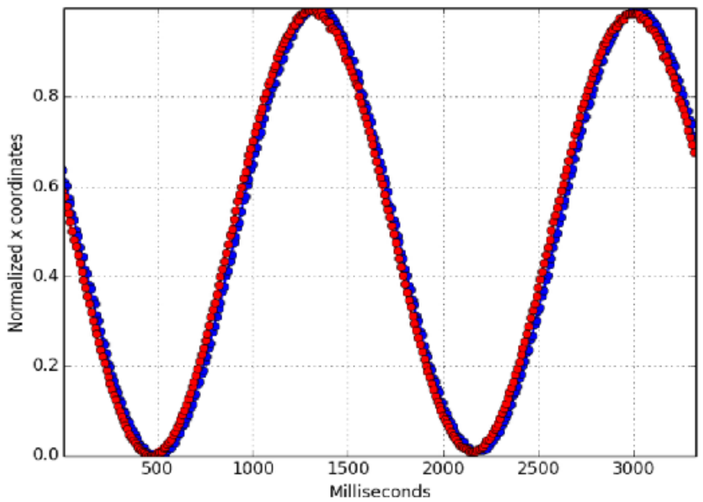
VIDEO BY RVDM88 (YOUTUBE)

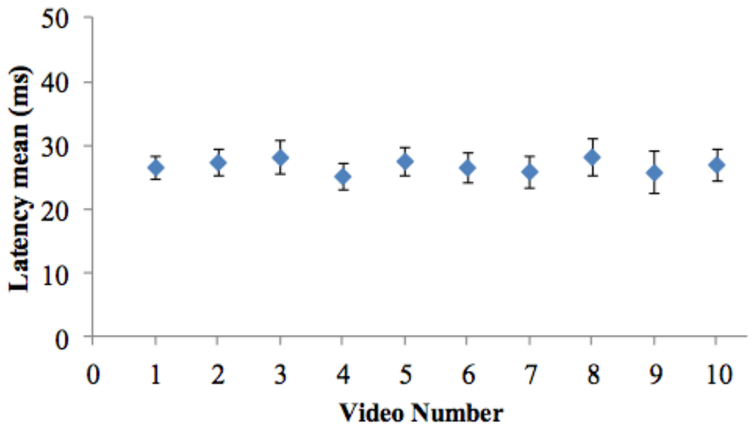
END-TO-END LATENCY (MOTION TO PHOTON)

- Vive
  - ~22 ms (Niehorster, Li, Lappe 2017),
- Oculus Rift:
  - ~22 ms
- Phasespace: ~27 ms (From my own lab, unpublished)

LATENCY FOR A PHASESPACE X2 MOTION CAPTURE SYSTEM







NIEHORSTER, LI, LAPPE 2017

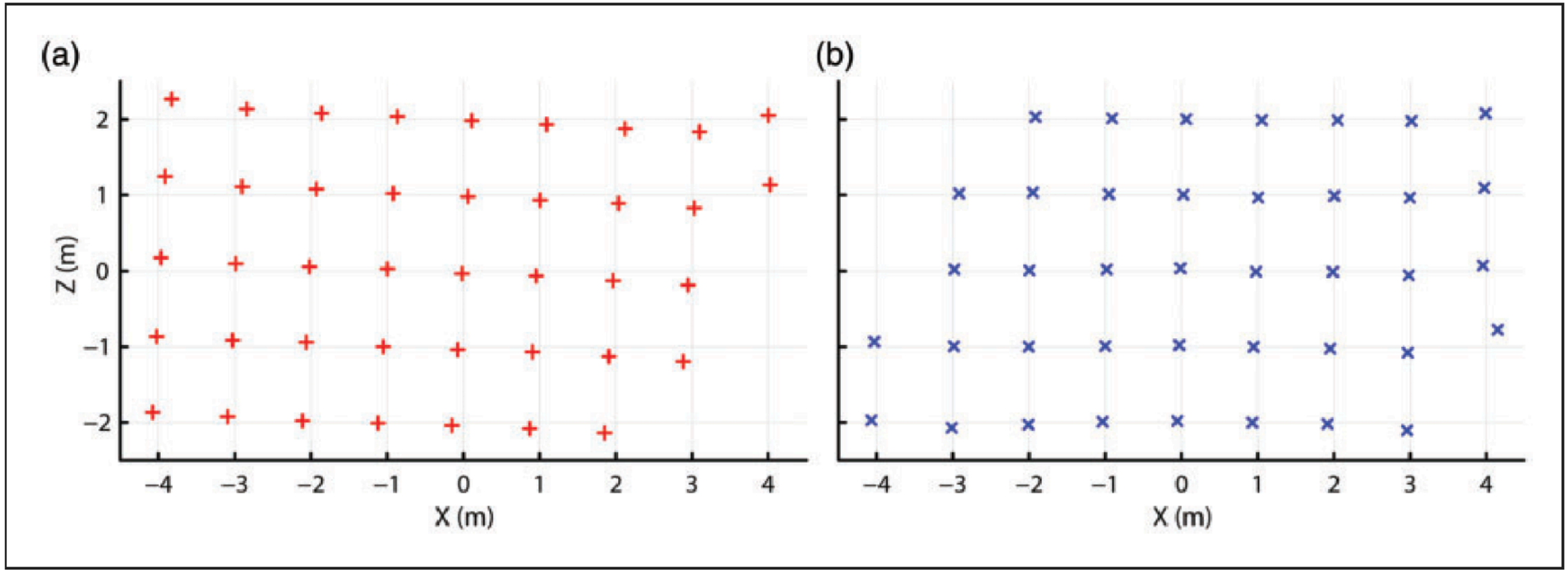
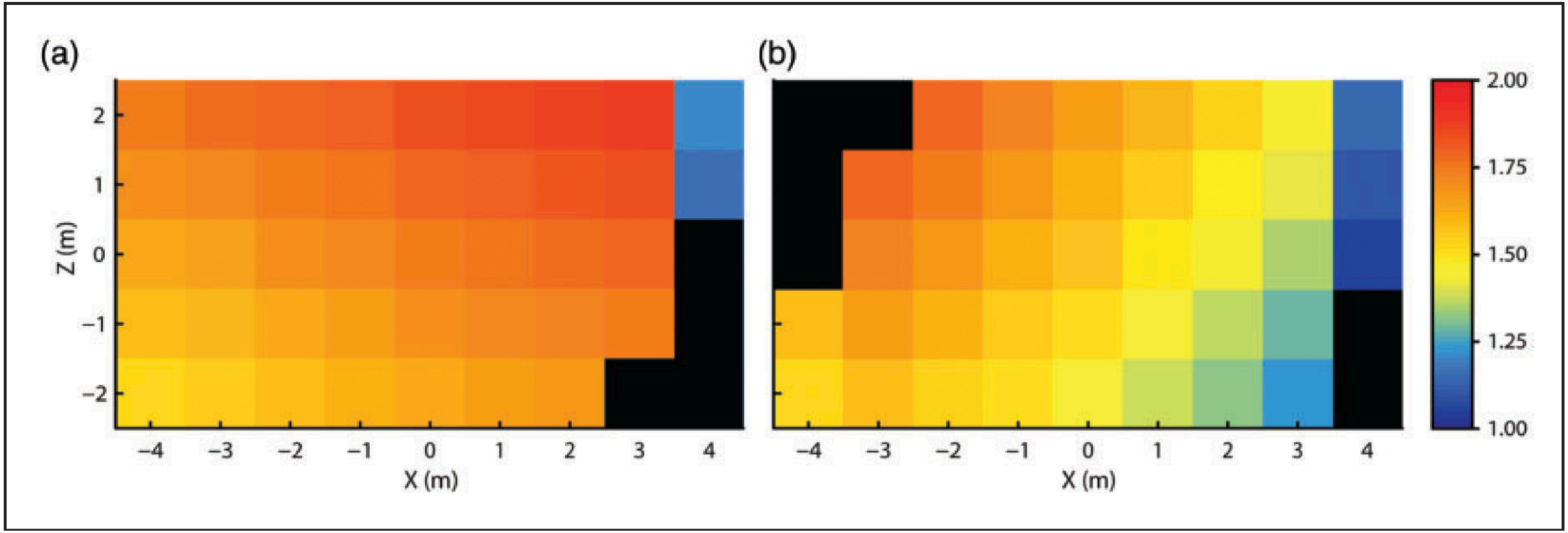


Figure 2. Recorded *X*-*Z* position (a) for when the headset faced along the positive *X* axis and (b) for when the headset faced along the negative *X* axis.

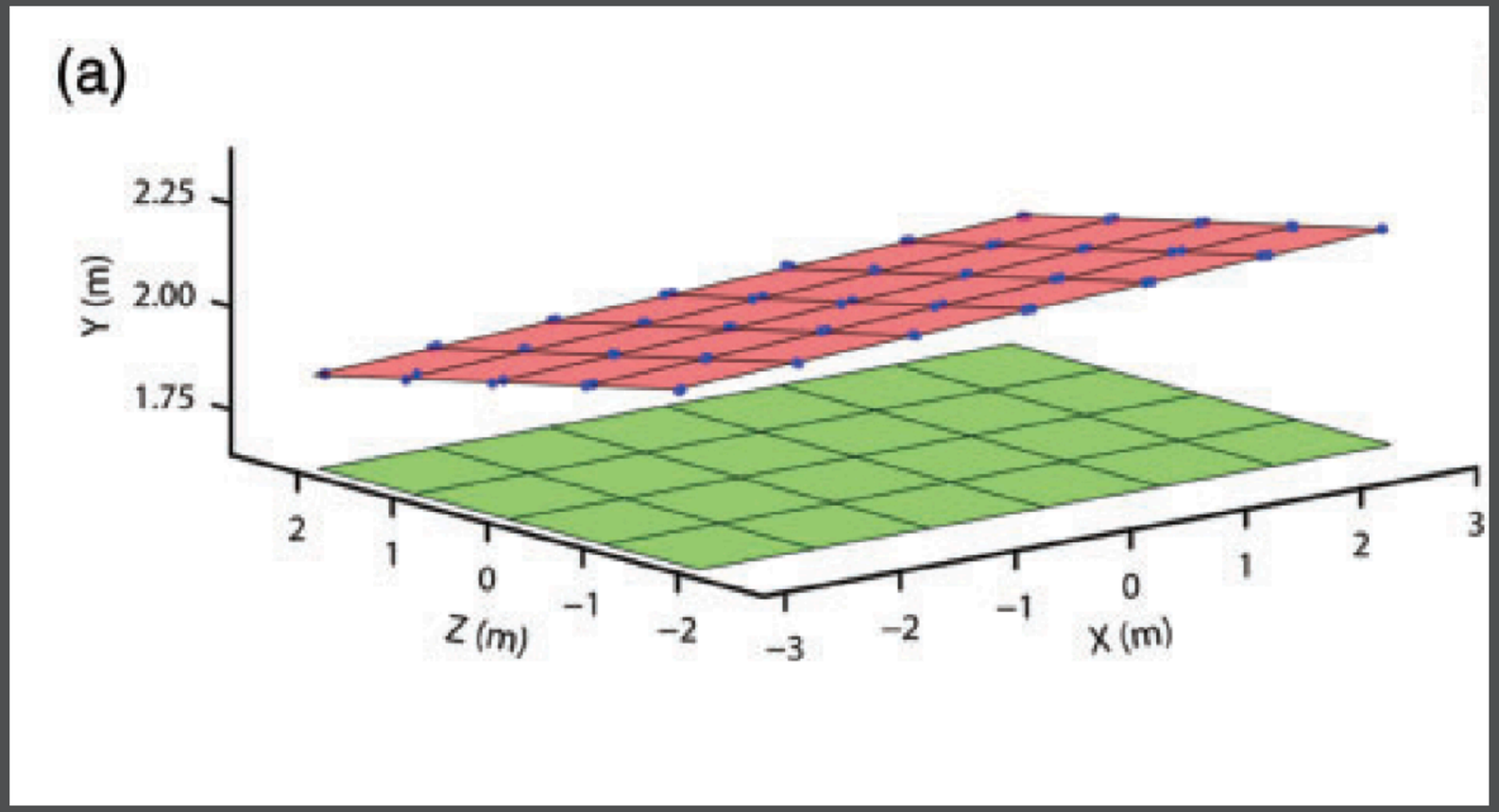
NIEHORSTER, LI, LAPPE 2017

A systematic miscalculation of height in the Vive tracking system.



True height was approximately 1.65m.

NIEHORSTER, LI, LAPPE 2017

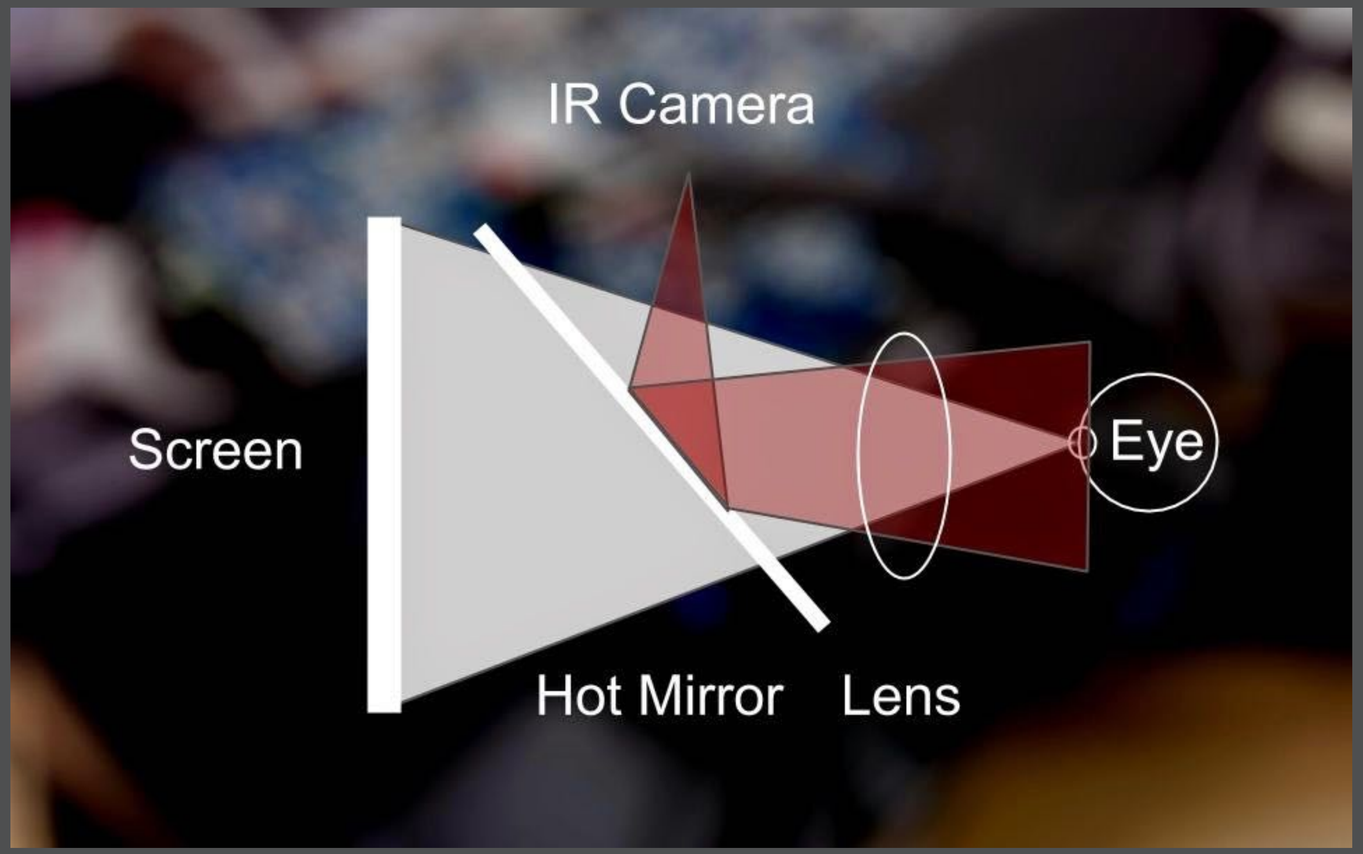


Spatially dependent miscalculations of height, yaw, pitch, and roll are consistent with a tilted reference plane.

EQUIPMENT - EYE TRACKERS

THIS IS THE SECTION WHERE I WAS GOING TO SING PRAISES ABOUT SMI EQUIPMENT.

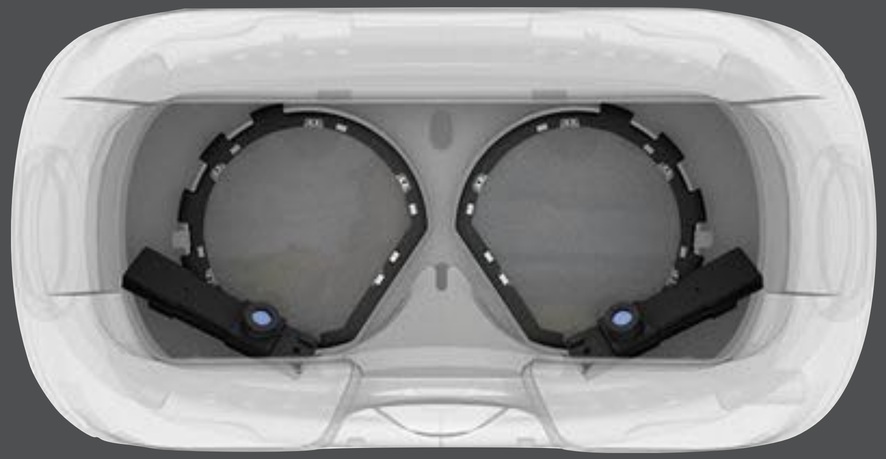
TOBII



eye image *not* from a Tobii

TOBII AND PUPIL LABS

- Refresh rate: 120 Hz
- Latency: Approximately 10ms from mid exposure to when data available on client interface.

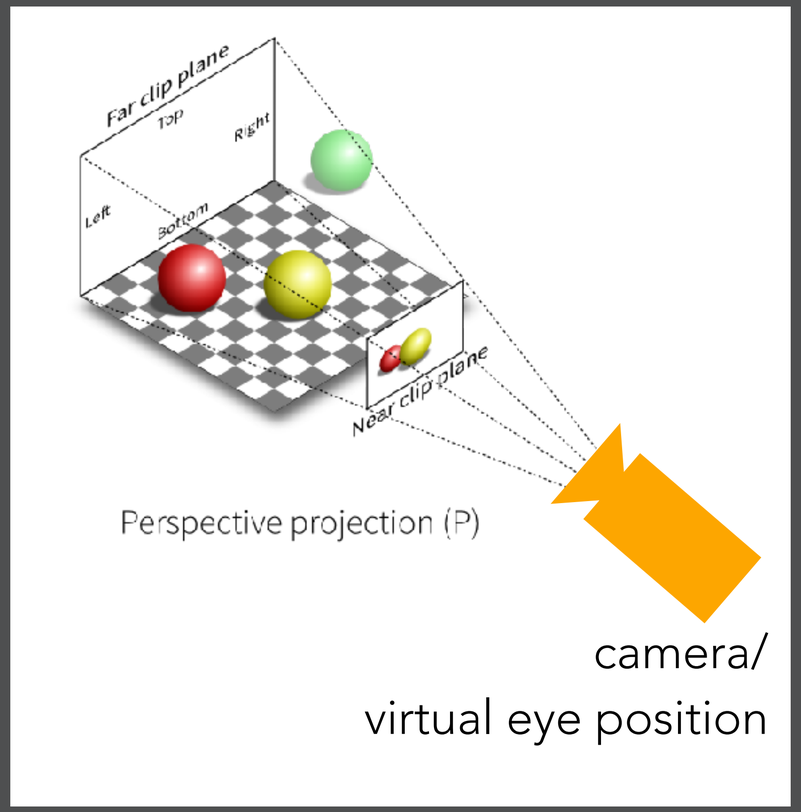


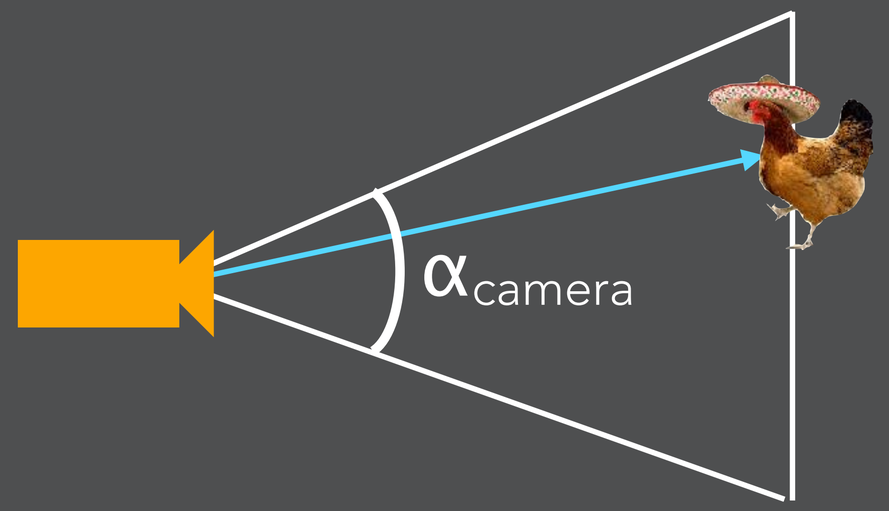
THE IMPORTANCE OF PROPER PERSPECTIVE

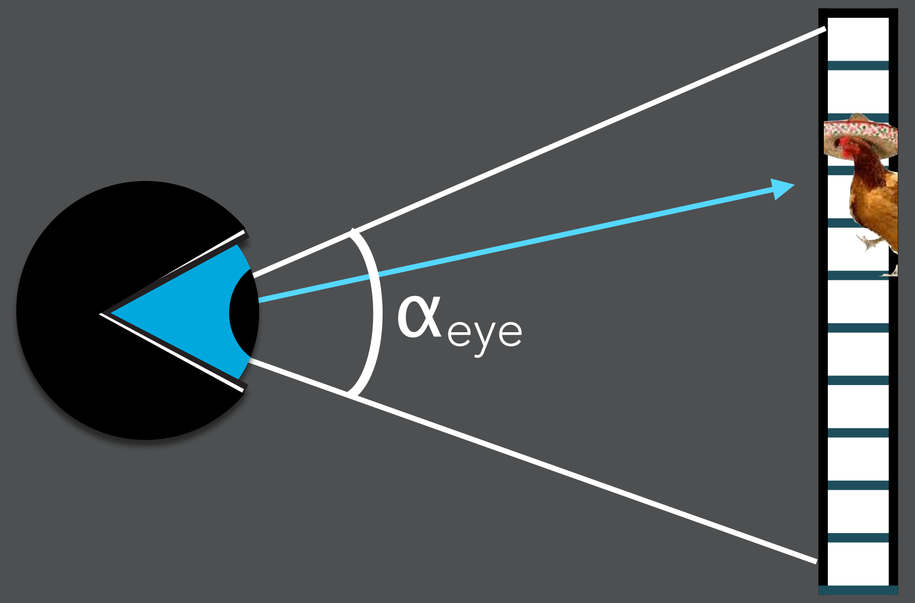
PERSPECTIVE RENDERING IN COMPUTER GRAPHICS:

- Camera position
- Camera orientation
- Field of view
- Aspect ratio

The projected image is then displayed on the HMD screen.



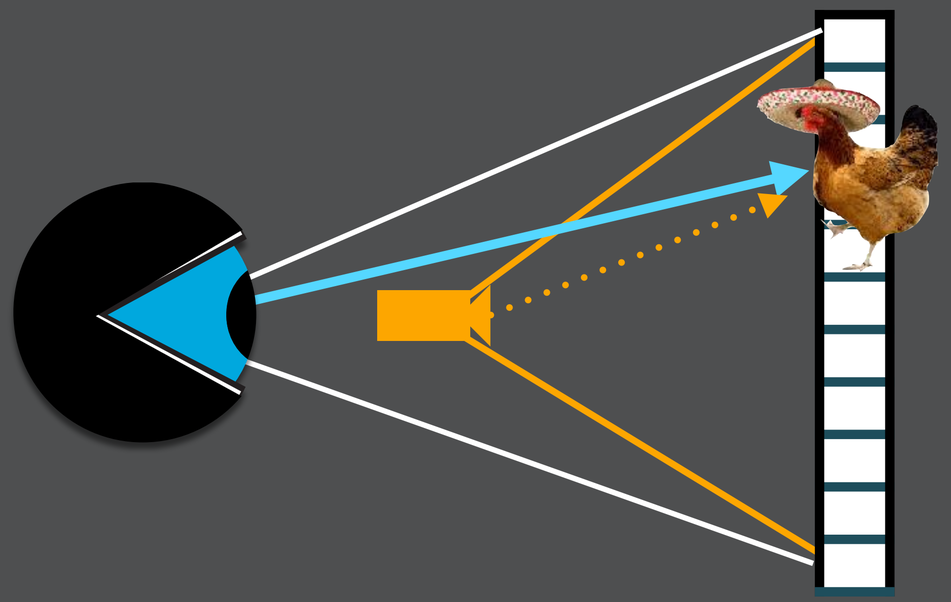




virtual gaze angle
from the virtual camera to an object in the virtual world

physiological gaze angle
from the eye to an object on the display

If geometries are set properly, then when the subject is looking at a virtual object, $\alpha_{eye}=\alpha_{camera}$

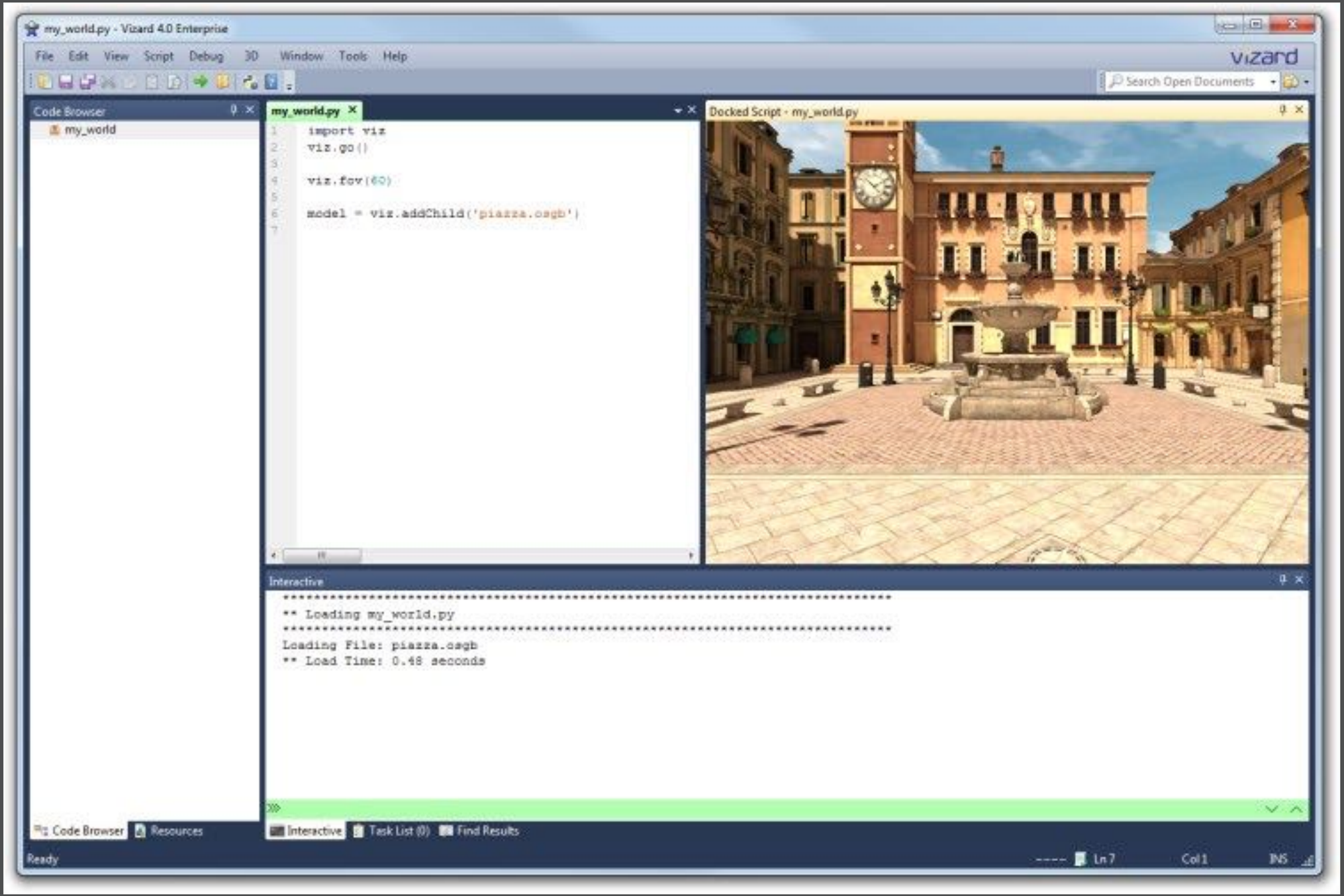


If geometries are **not set** properly, $\alpha_{eye} \neq \alpha_{camera}$.

The geometry above would lead to an angular compression of the visual field.

TOOLS FOR EXPERIMENTAL DESIGN

Vizard, by Worldviz



unity

Unity - Scene01.unity - Project - PC, Mac & Linux Standalone

File Edit Assets GameObject Component Window Help

Pivot Global

Layers Layout

Scene Animator

Textured RGB 2D Effects Gizmos All

Persp

Game

Free Aspect Maximize on Play Stats Gizmos

Hierarchy

Create All

- Character
  - Armature
  - Eye_L
  - Eye_R
  - Hands
  - Head
  - Head_001
  - Pants
  - Shirt
  - Shoes
- Directional light
- Main Camera
- Terrain

Project

Create

- Materials
- Standard Assets
- Character
  - Eye_L
  - Head
  - Head_001
  - Shirt
  - Hands
  - Pants
  - Shoes
  - Eye_R
  - Idle
  - TurnLeft
  - TurnRight
  - Walk
  - CharacterAvatar
  - Armature
  - Eye_L
  - Eye_R
  - Hands
  - Head
  - Head_001
  - Pants
  - Shirt
  - Shoes
  - CharacterController
  - Eye
  - Head
  - IdleWalk
- New Terrain
- Scene01

Inspector

IdleWalk Import Settings

Open... Execution Order...

Imported Object

```
using UnityEngine;
using System.Collections;

public class IdleWalk : MonoBehaviour
{
    protected Animator animator;
    public float DirectionDampTime = 0.25f;

    // Use this for initialization
    void Start ()
    {
        animator = GetComponent<Animator>();
    }

    // Update is called once per frame
    void Update ()
    {
        if(animator)
        {
            float h = Input.GetAxis("Horizontal");
//left/right arrow keys
            float v = Input.GetAxis("Vertical");
//up/down arrow keys

            animator.SetFloat("Speed", h*h + v*v);
            animator.SetFloat("Direction", h,
DirectionDampTime, Time.deltaTime);
        }
    }
}
```

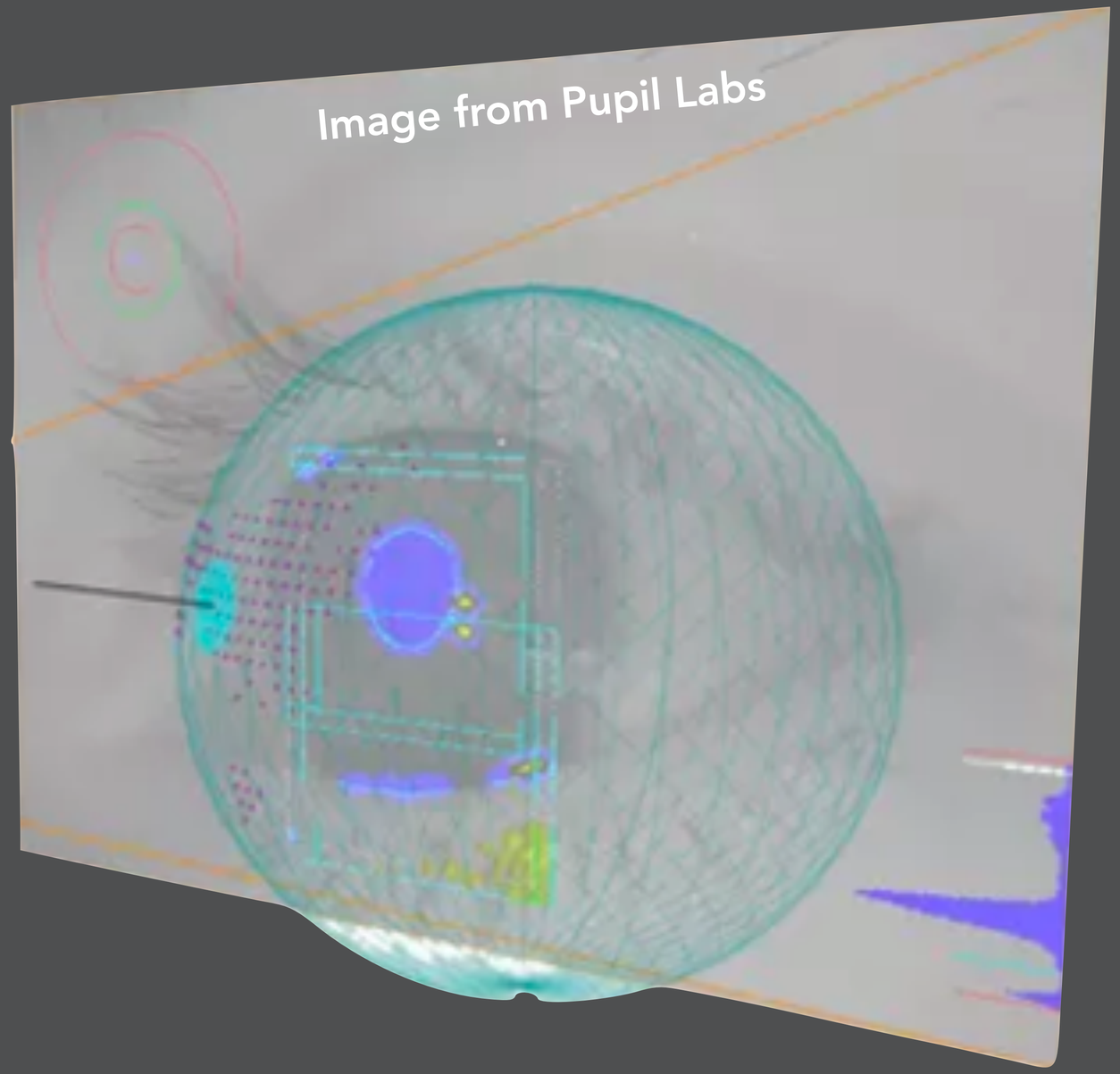
Asset Labels

The Unreal Engine

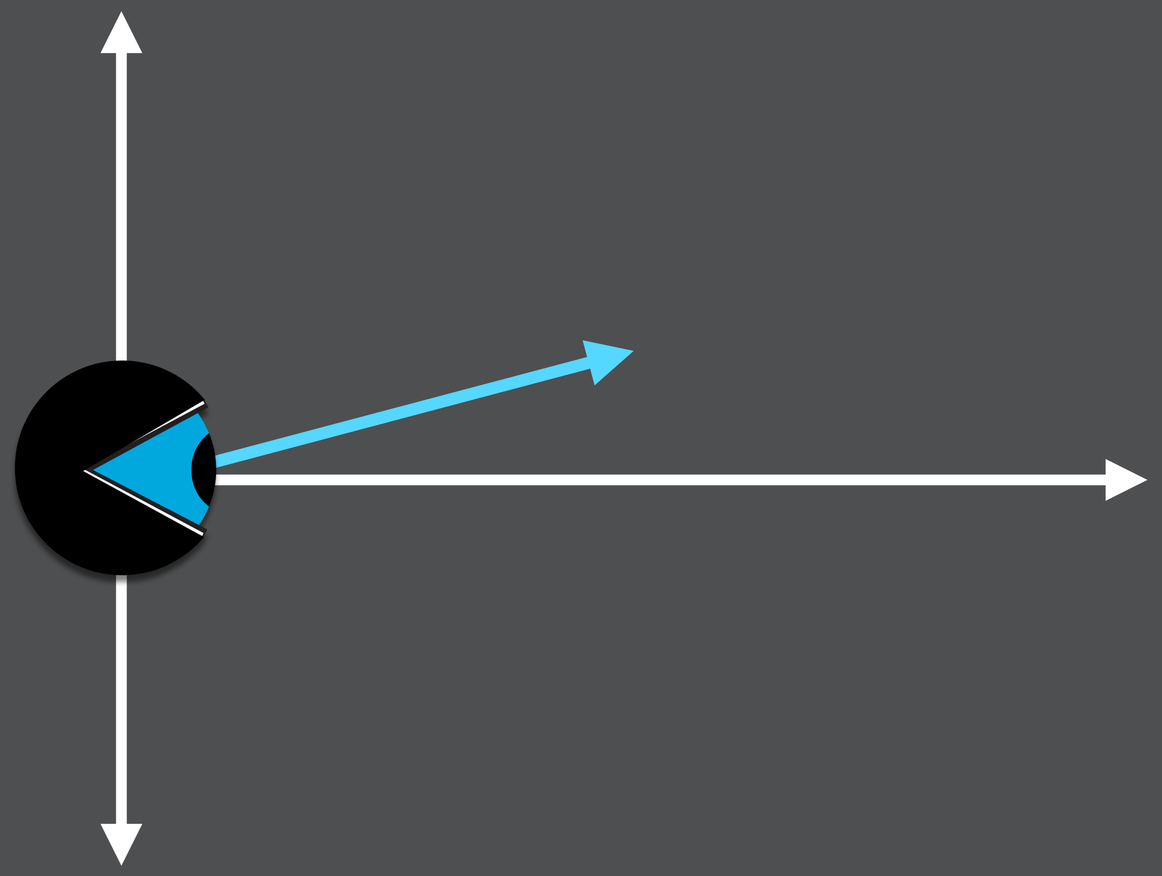
WHAT YOUR SOFTWARE DOES FOR YOU

A SINGLE SAMPLE OF GAZE DATA CONTAINS:

- Timestamp
- Index
- Gaze origin
- Pupil position
- Pupil size
- Eye center
- Eye-in-head direction [XYZ]



EYE-IN-HEAD DIRECTION [XYZ]

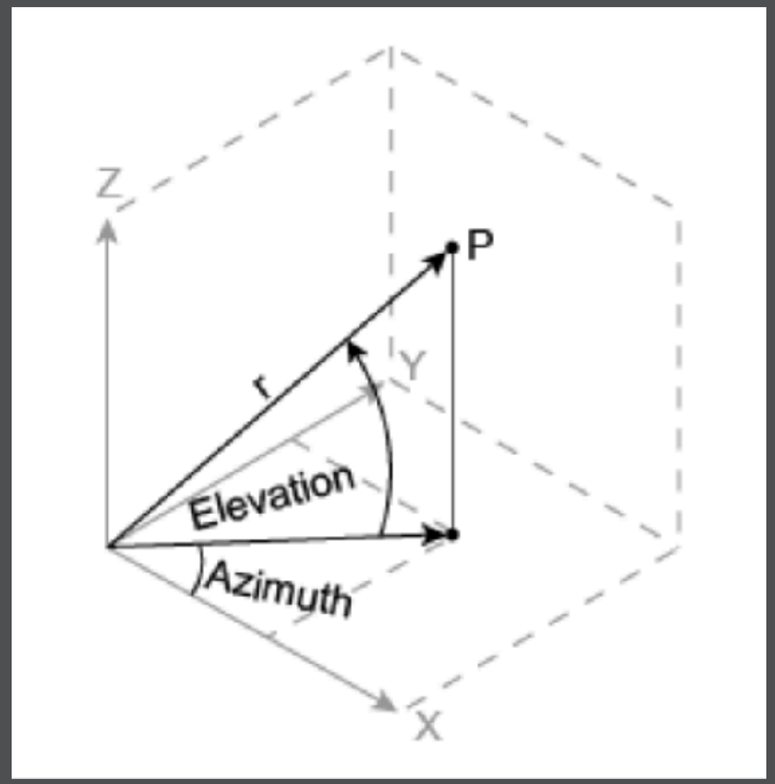


It is a unit-length vector (i.e. 1 meter long), indicating a point along the gaze vector.
It is defined within the head's local coordinate system.

SPHERICAL COORDINATES

(i.e. azimuth and elevation)

An intuitive alternative to a vector representation.



NOTEBOOK 1: INSPECT DATA

ANACONDA/JUPYTER NOTEBOOK:

MODULES:

- PLOTLY
- PANDAS
- NUMPY
- SCIPY

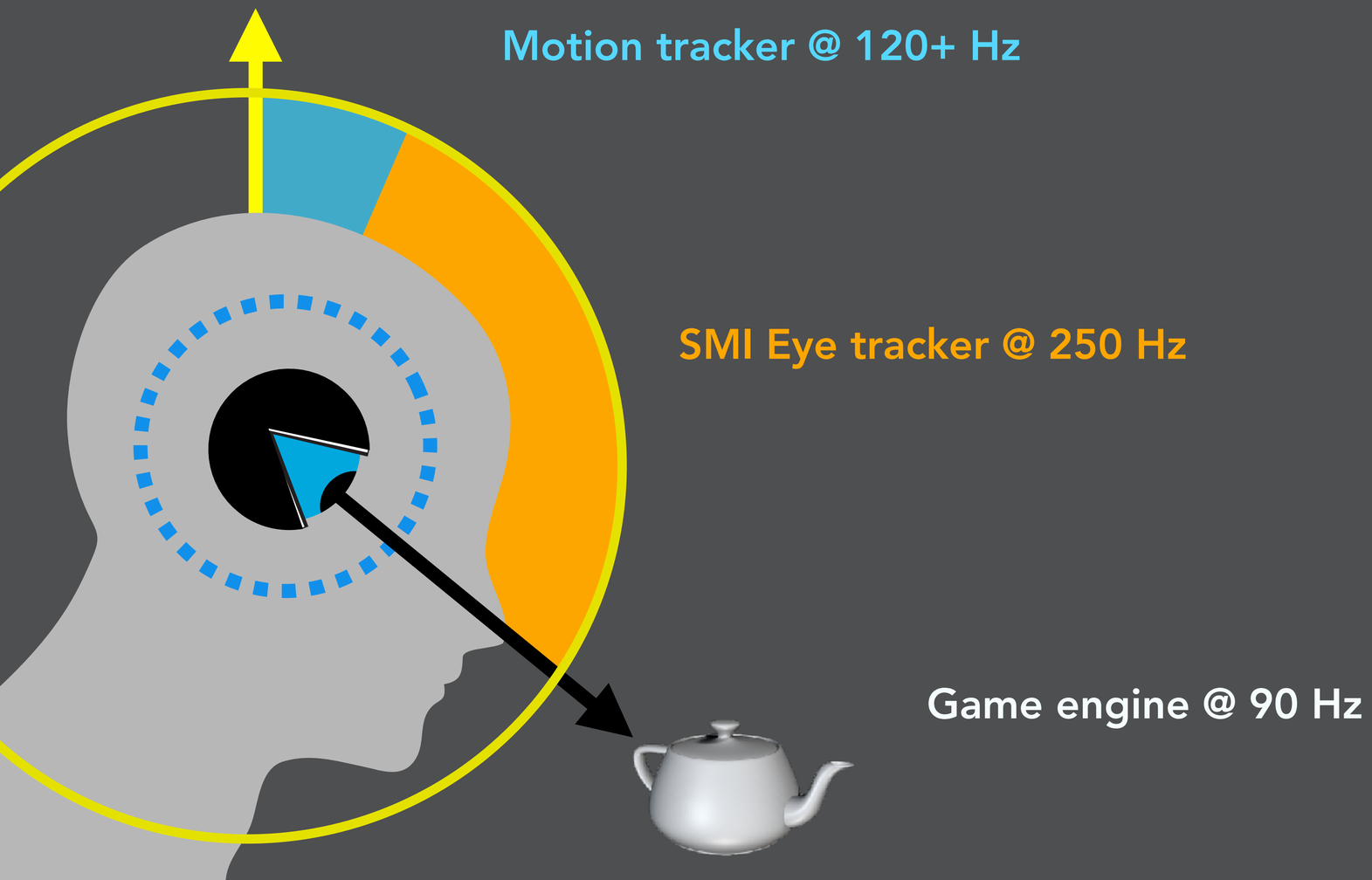
NOW, LETS SEE SOME HIGHER-QUALITY DATA FROM A NEWER EYE TRACKER

WHAT YOU REALLY WANT TO KNOW:

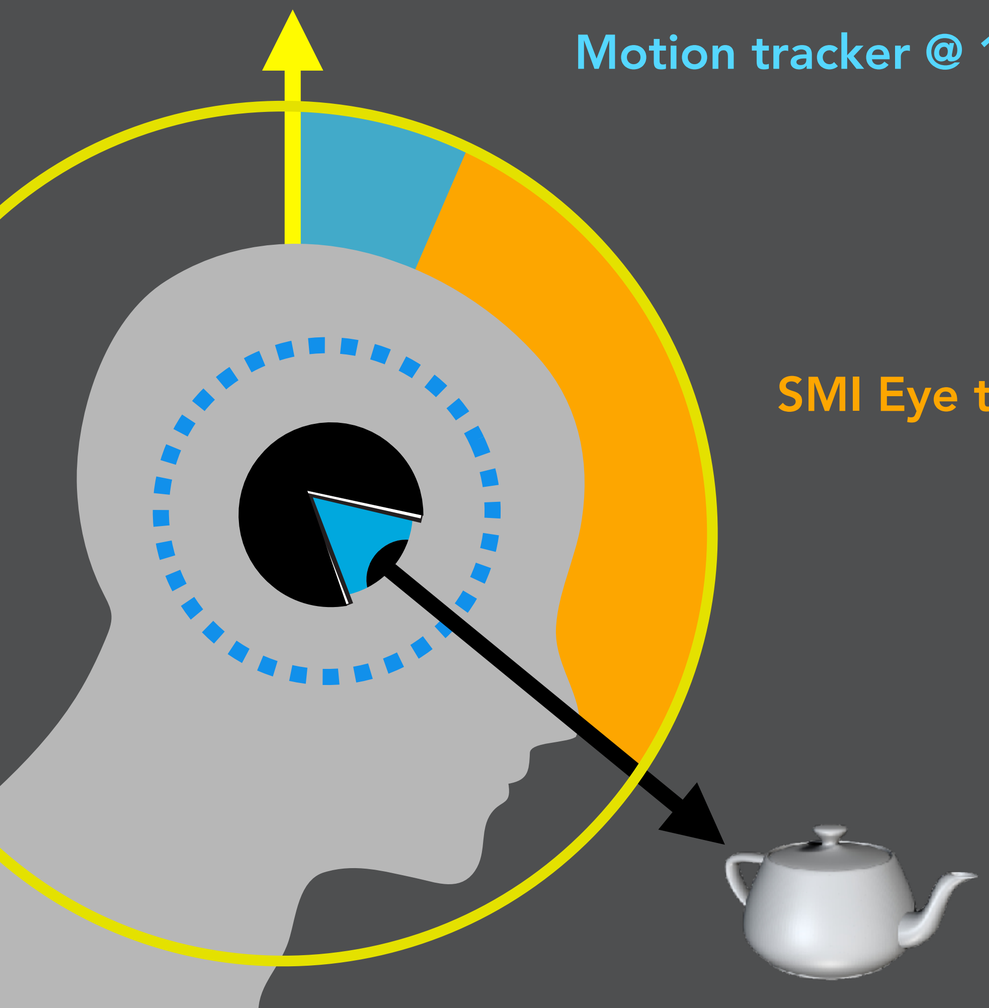
- **What** is the person looking at?
- **Where** on the object or region of interest (ROI) is the person looking?
- How close to an object/ROI is a person looking?
- What kind of eye/head movements have occurred?
- What are the kinematics?

TODAY, WE WILL TAKE A STEP IN THIS DIRECTION.

TEMPORAL ALIGNMENT


Motion tracker @ 120+ Hz
SMI Eye tracker @ 250 Hz
Game engine @ 90 Hz

May differ in….



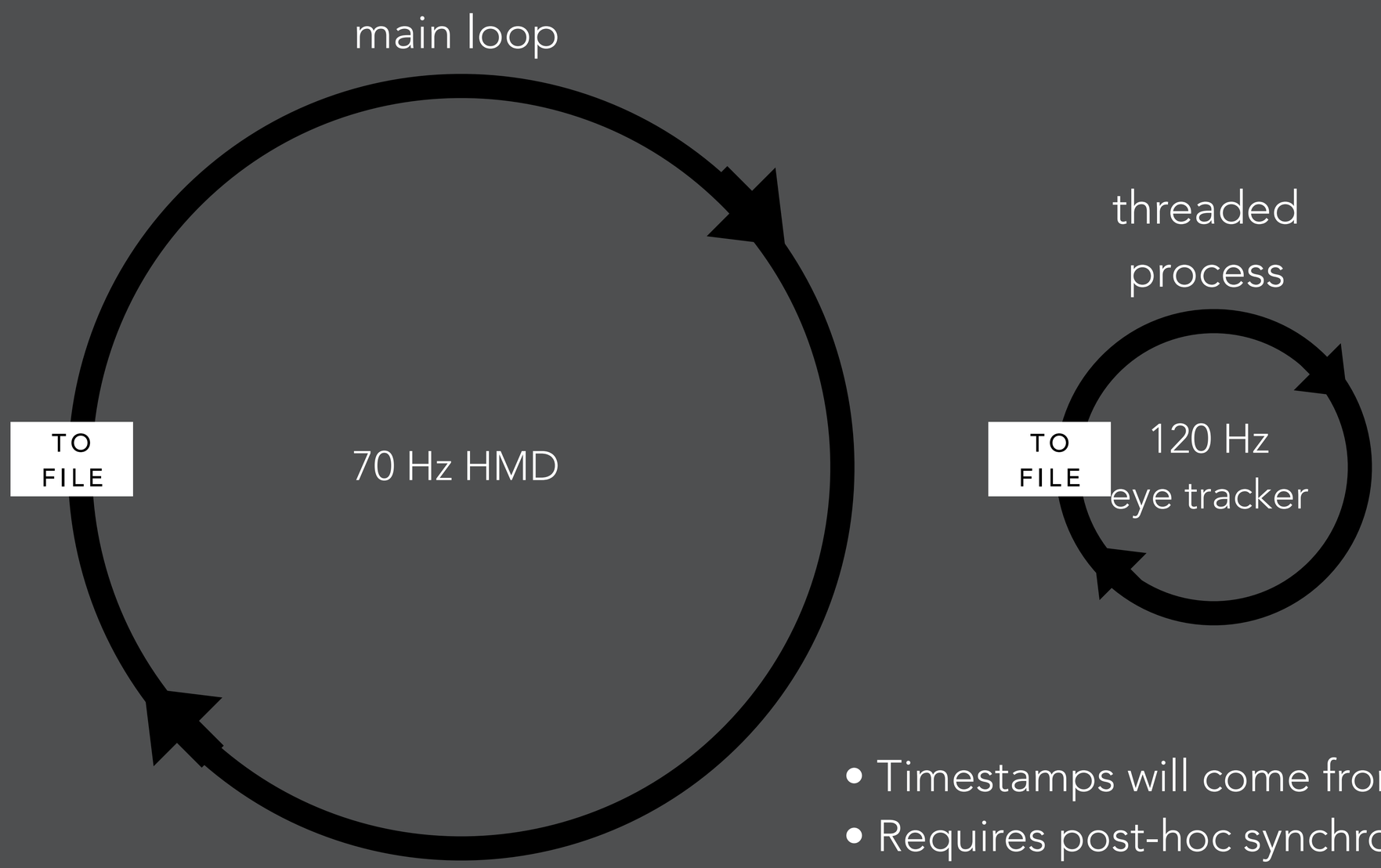
Sampling rate

Coordinate systems

Latency

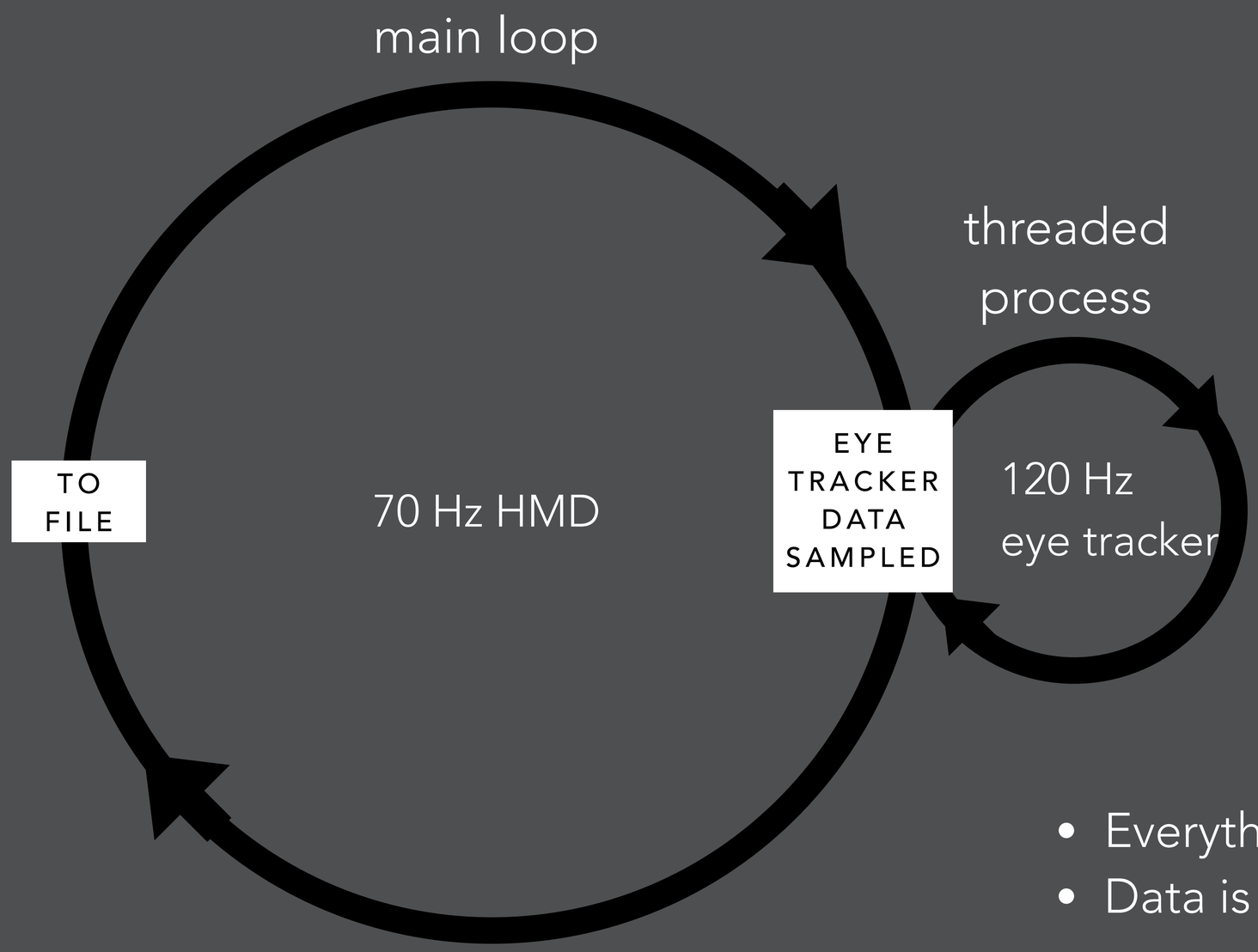
Phasing

INDEPENDENT DATA OUTPUT



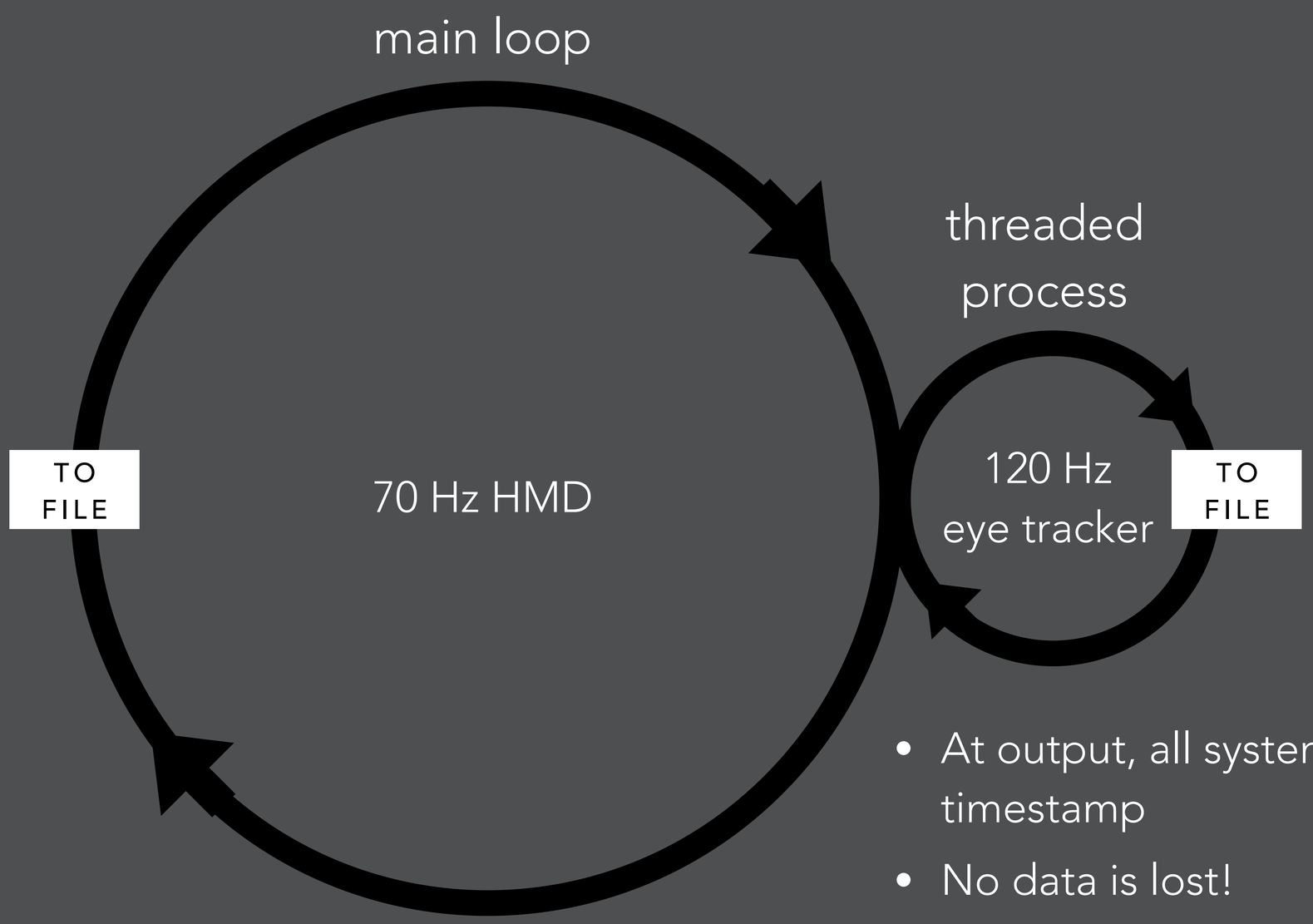
- Timestamps will come from separate clocks
- Requires post-hoc synchronization

PRE-ALIGNED AT LOWEST REFRESH



- Everything sampled at lowest refresh rate
- Data is lost!

PRE-ALIGNED

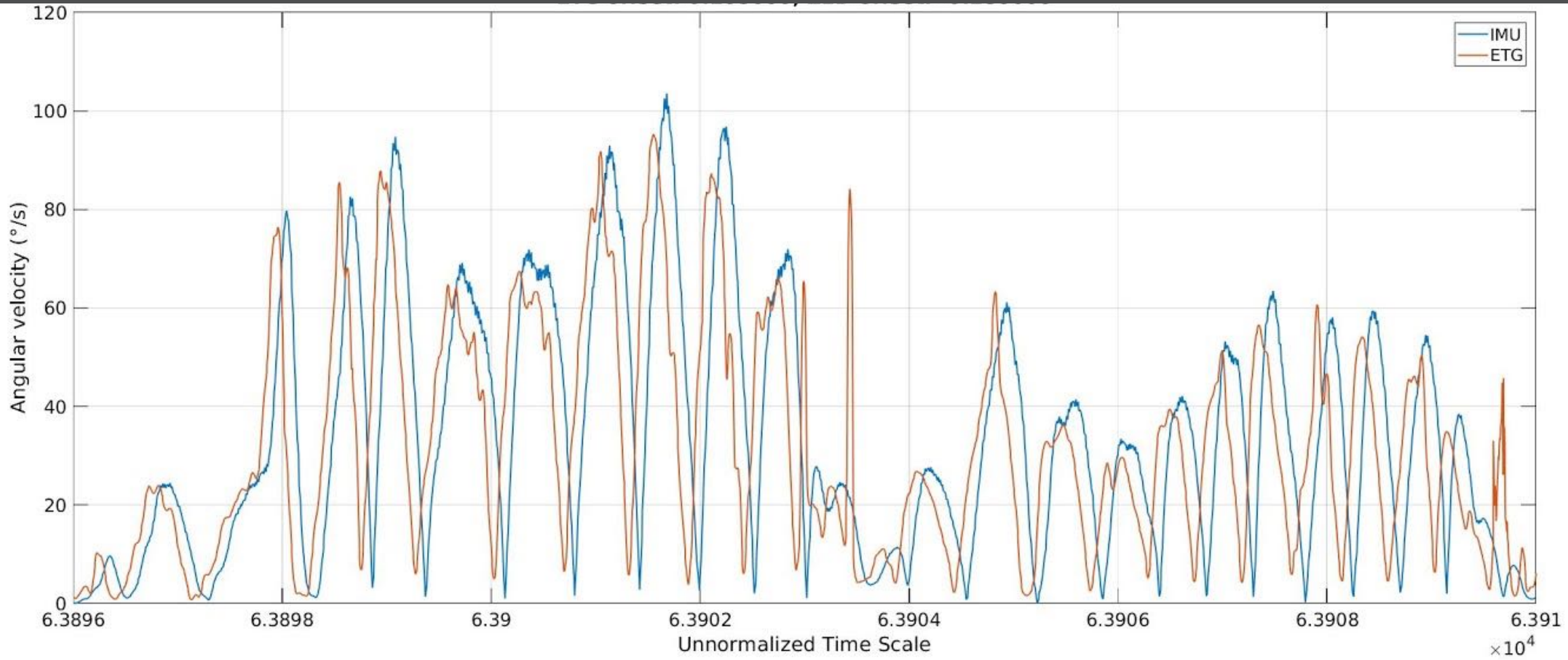


- At output, all systems have access to a common clock/ timestamp
- No data is lost!
- More difficult to implement

3 SOLUTIONS FOR TEMPORAL

- **Pre-aligned/synchronized:** Record at Unity refresh.
- **Pre-aligned:** Create subprocess to record at fastest rate - (1 output file: data for slower subprocess is repeated)
- **Independent:** Create subprocess. Each subprocess records independently and is aligned post-hoc (2 output files)

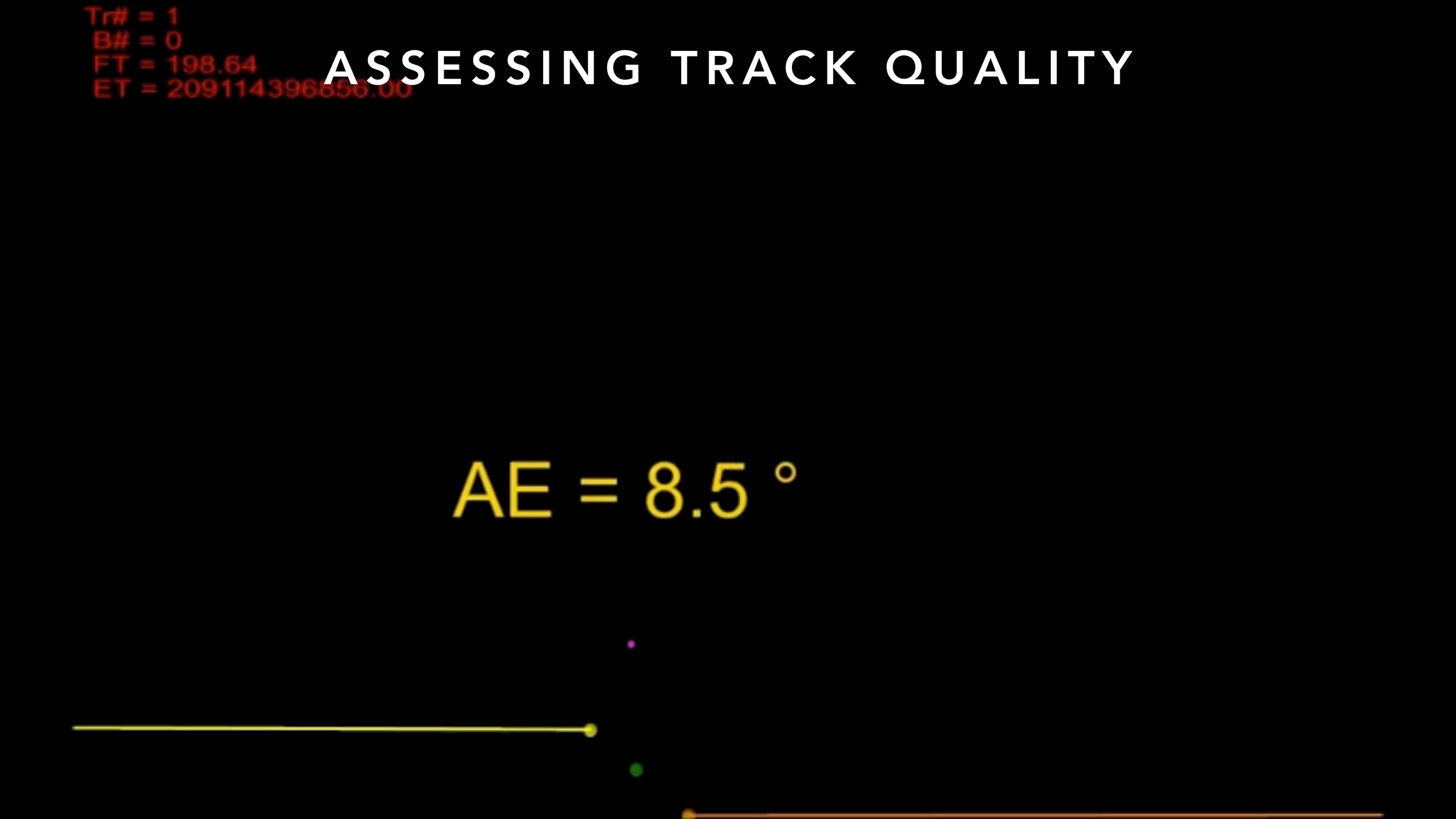
TEMPORAL OFFSET = 10.8 MS



TIME (MS)

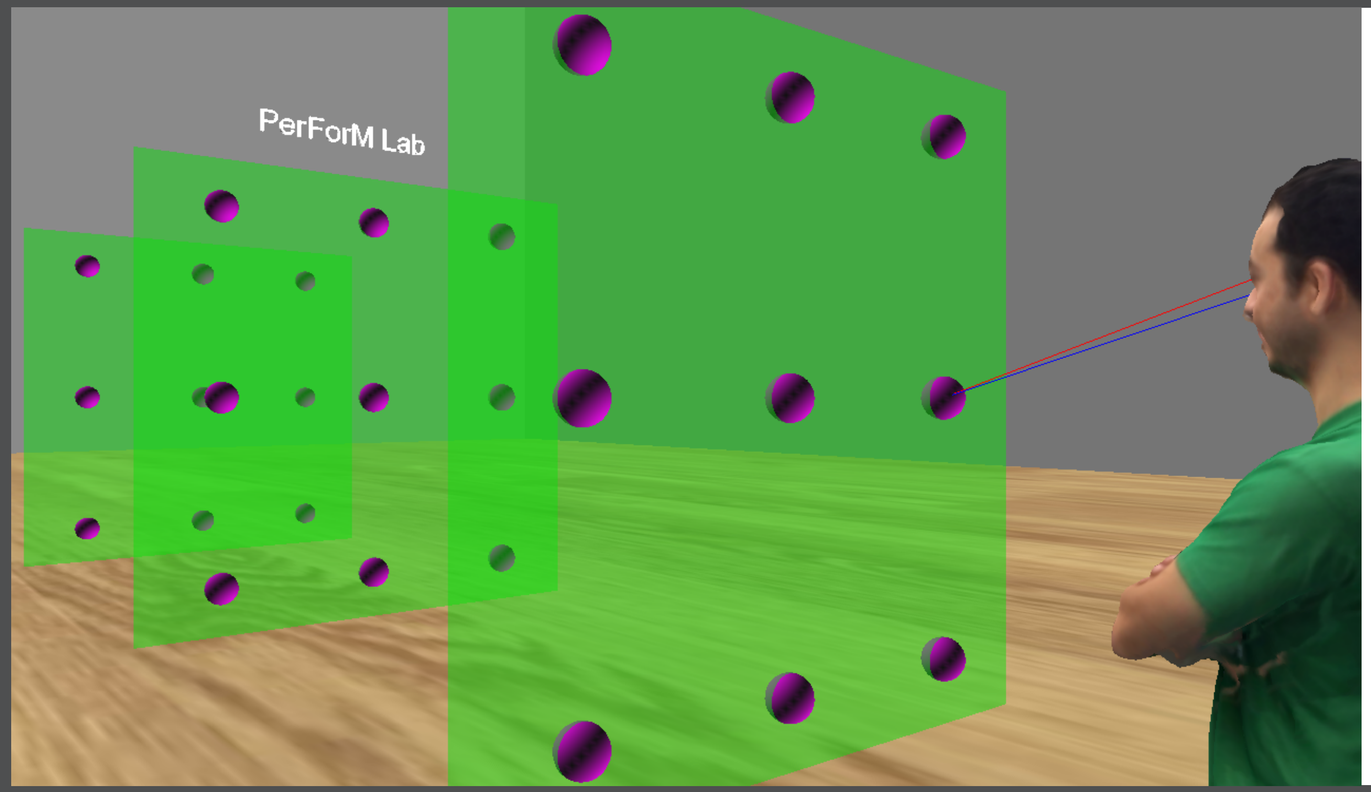
SPATIAL ACCURACY

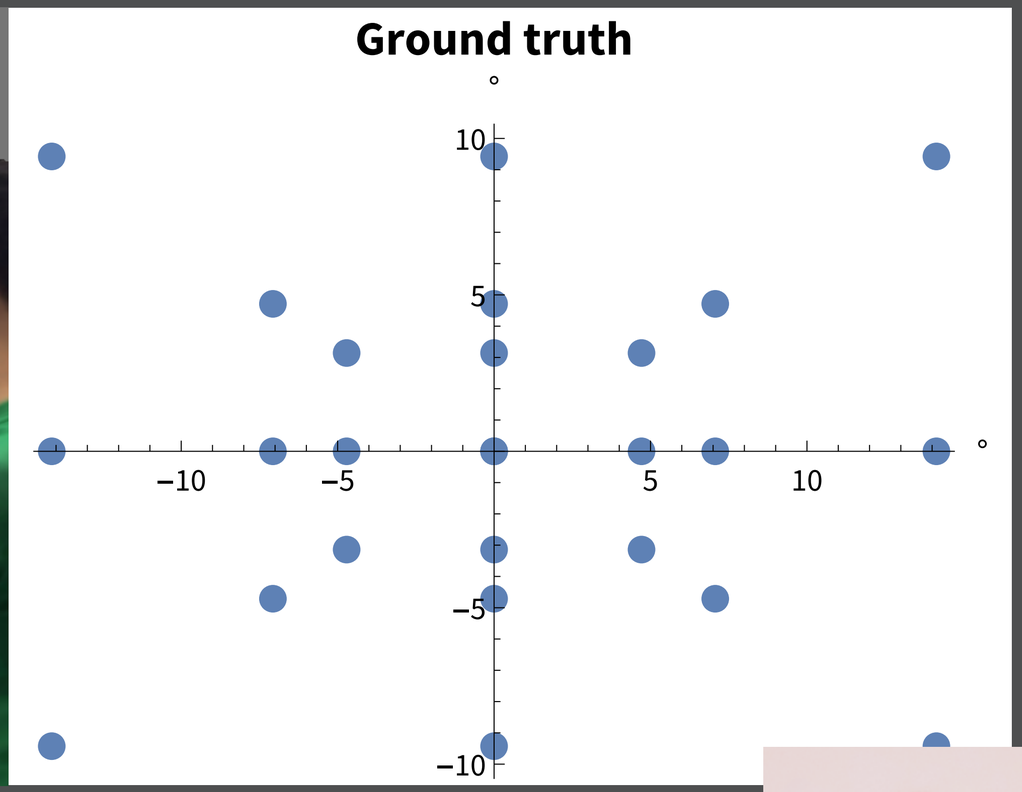
ASSESSING TRACK QUALITY

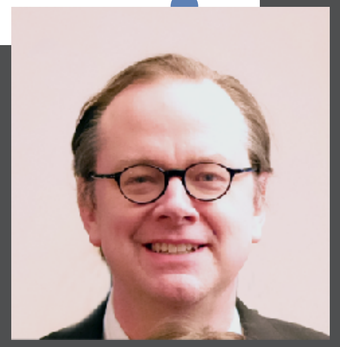


NOTEBOOK 2: MEASURE TRACK QUALITY

DEALING WITH HELMET SLIPPAGE AND CALIBRATION WOES



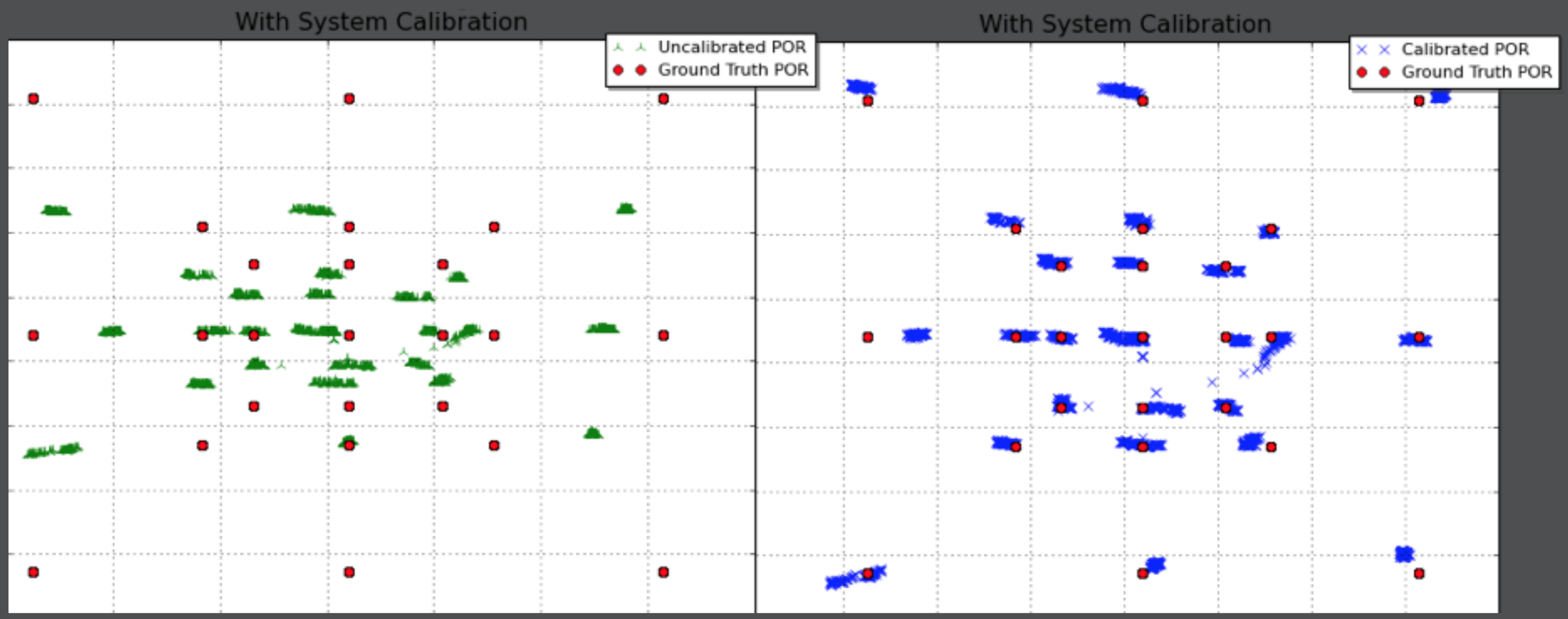




Work in collaboration with Flip Phillips @ Skidmore College.

CALIBRATION PROCEDURE

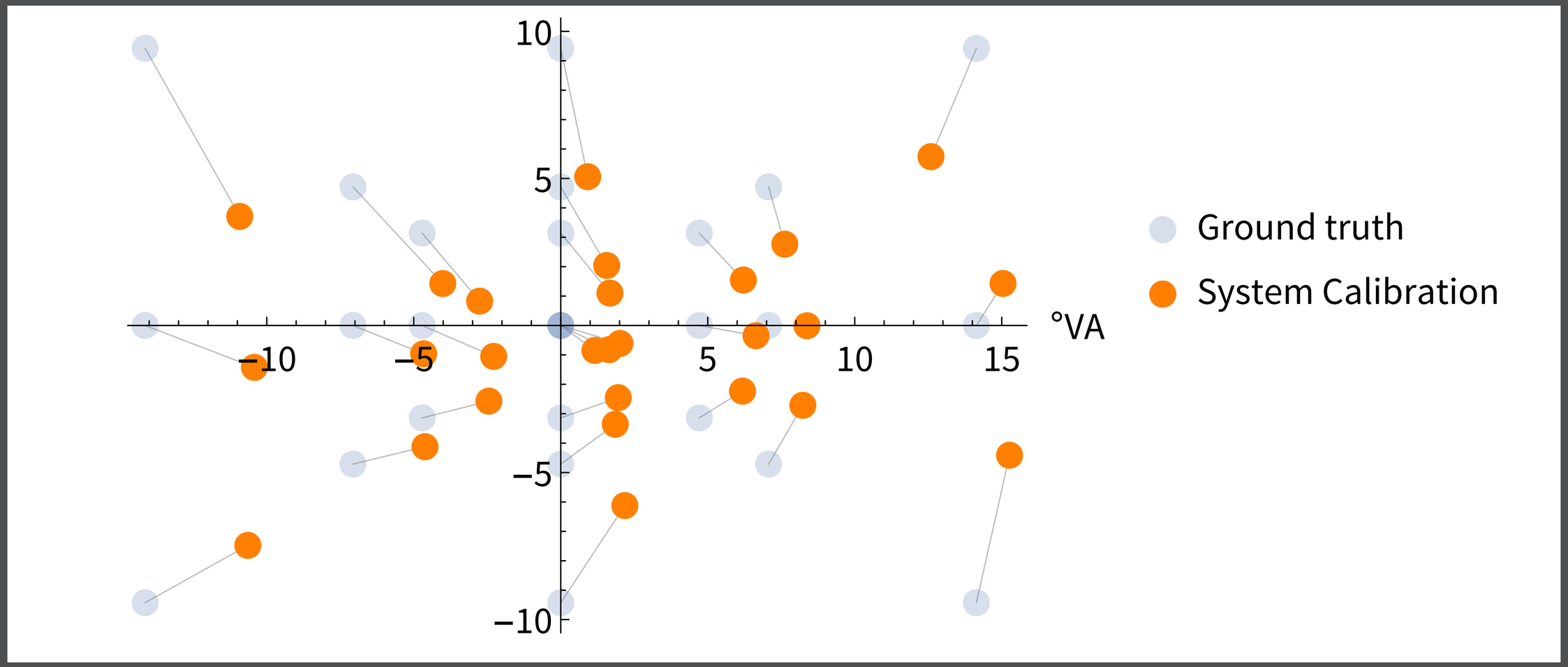
- Record 100 frames per calibration point
- Calculate the 3*3 homography via least squared error

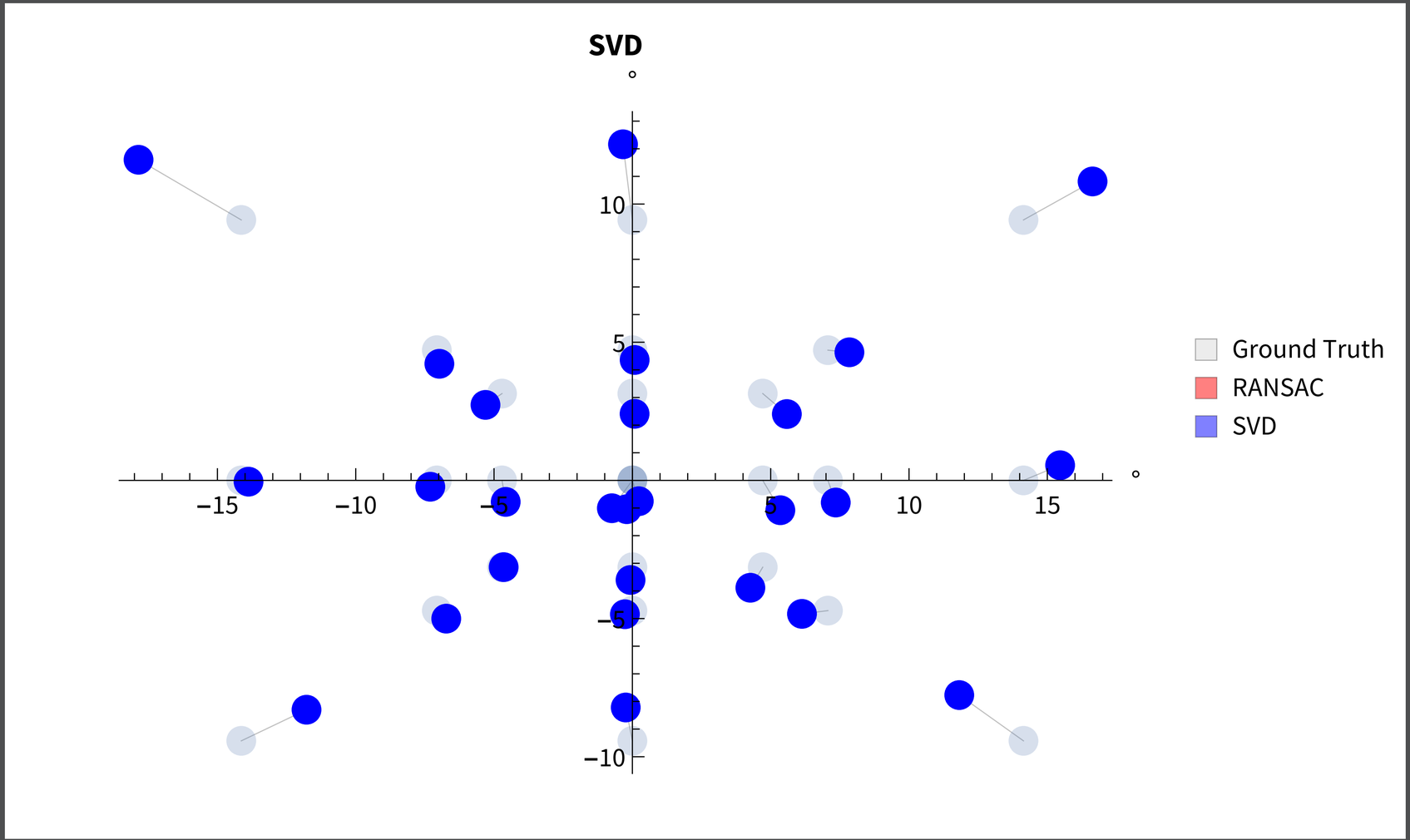


Before Calibration

After Calibration

TRACK DEGRADATION DUE TO HELMET SLIPPAGE OVER TIME

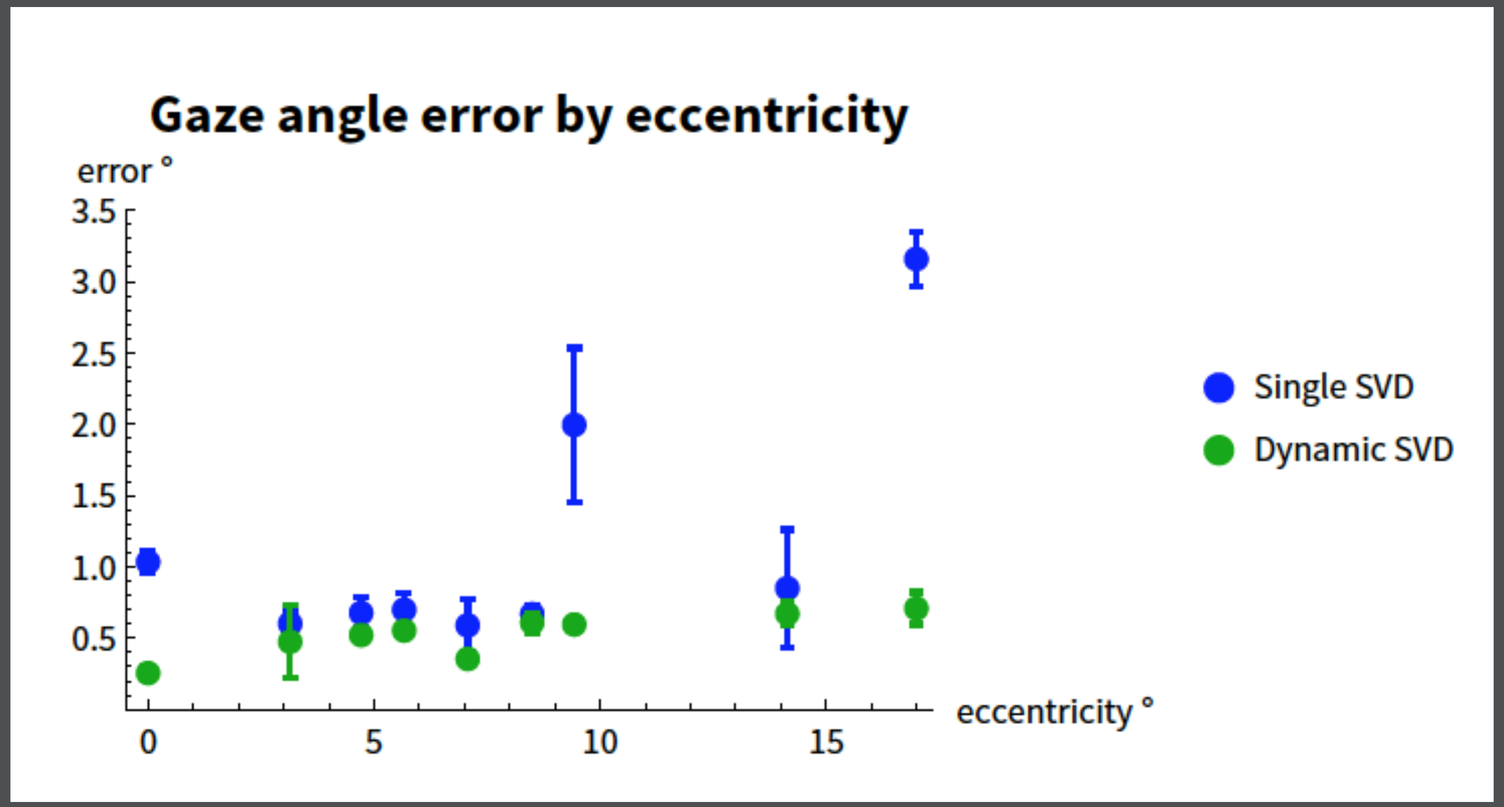



SVD
10
5
−5
−10
−15
−10
−5
5
10
15
Ground Truth
RANSAC
SVD

OFFLINE DYNAMIC RECALIBRATION

Typically, calibrations events cause discrete updates to the eye/screen mapping.

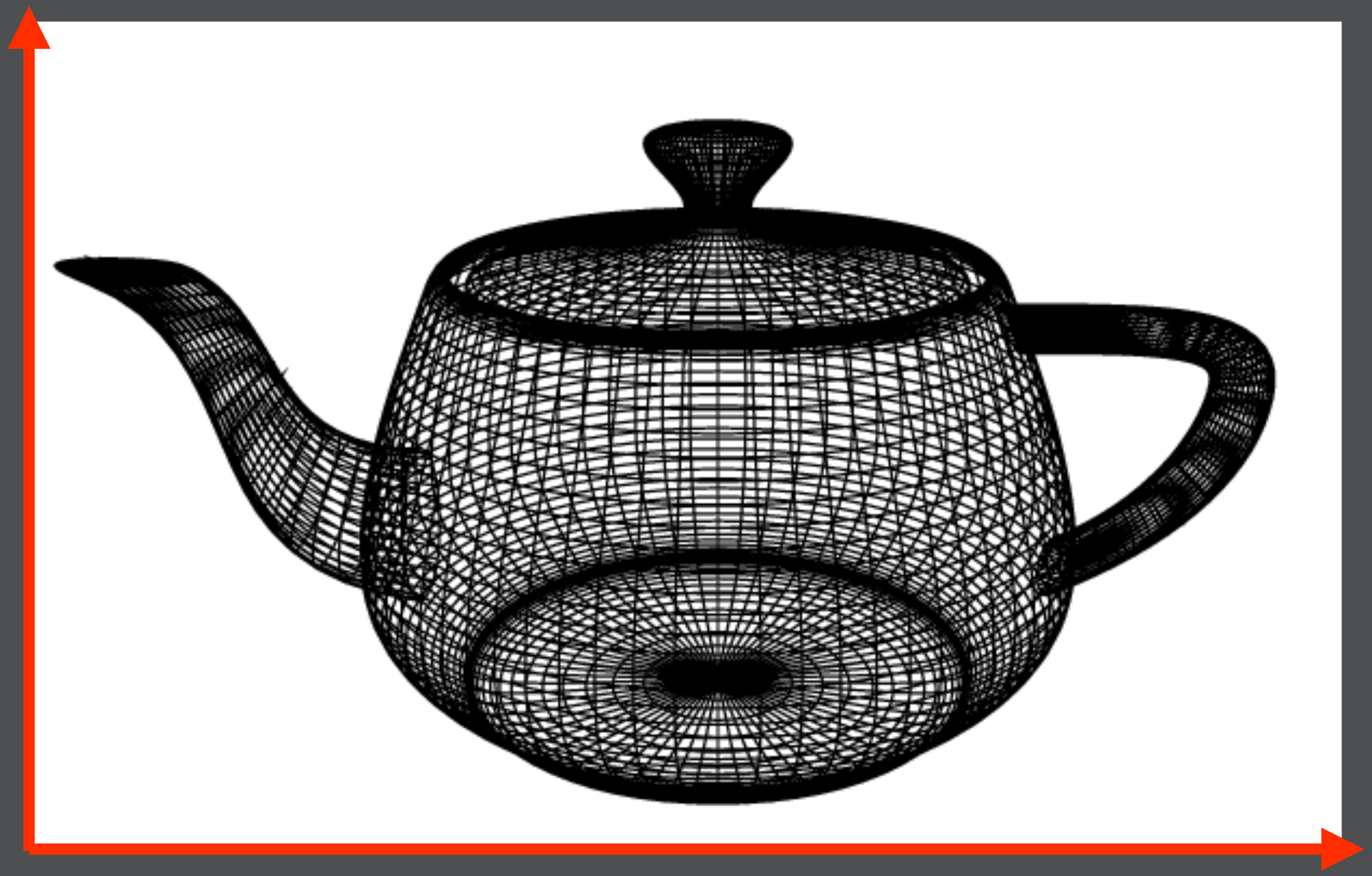
In the case of slippage, a continuous interpolations, or a "dynamic recalibration" may be more appropriate.



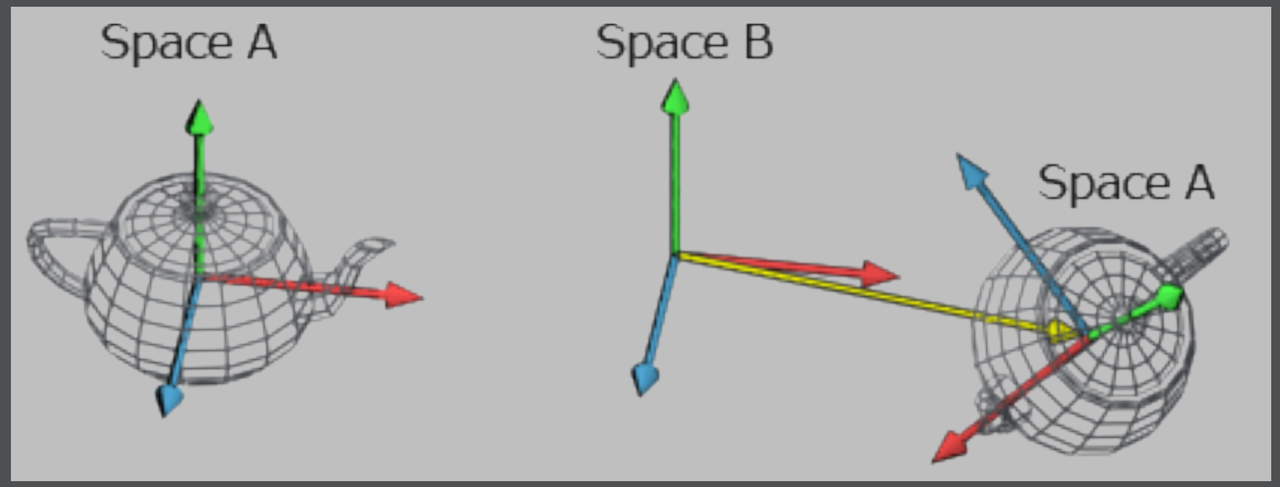
Method	$\overline{x}_{err}$	s_{err}
System	2.90	1.60
RANSAC	2.39	1.11
SVD	1.18	1.03
Dynamic SVD	0.538	0.0668

THE SCENE GRAPH, AND TRANSFORMATION MATRICES

A 3D Object is a collection of vertices within a local coordinate system.



Translate, rotate, the vertices quickly and efficiently using the transformation matrix.



$$\underbrace{\begin{bmatrix} .96 & .07 & -0.28 & -0.98 \\ -.01 & 0.98 & 0.19 & 1.5 \\ 0.29 & -0.18 & 0.94 & 0.14 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{transformation matrix}} \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{\text{vertex position}} = \underbrace{\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}}_{\text{new vertex position}}$$

The position, scale, and orientation of the virtual camera and objects in the world are encoded using these matrices.

THE LINEAR/AFFINE TRANSFORMATION MATRICES COLUMN VECTOR.

Linear transformations
(scale/rotate/shear)

$$\begin{bmatrix} 0 & 4 & 8 & 12 \\ 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \end{bmatrix}$$

Translation

Projection

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

identity

$$\begin{pmatrix} 1 & 0 & 0 & tx \\ 0 & 1 & 0 & ty \\ 0 & 0 & 1 & tz \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

translate

$$\begin{pmatrix} sx & 0 & 0 & 0 \\ 0 & sy & 0 & 0 \\ 0 & 0 & sz & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

scale

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(d) & -\sin(d) & 0 \\ 0 & \sin(d) & \cos(d) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

rotate around x

$$\begin{pmatrix} \cos(d) & 0 & \sin(d) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(d) & 0 & \cos(d) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

rotate around y

$$\begin{pmatrix} \cos(d) & -\sin(d) & 0 & 0 \\ \sin(d) & \cos(d) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

rotate around z

Operations of scaling, rotation, translation, can be composed into a single transformation matrix

$$\underbrace{\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix}}_{\text{transformation}}$$

$$M = R_y \cdot T \cdot S \cdot R_x$$

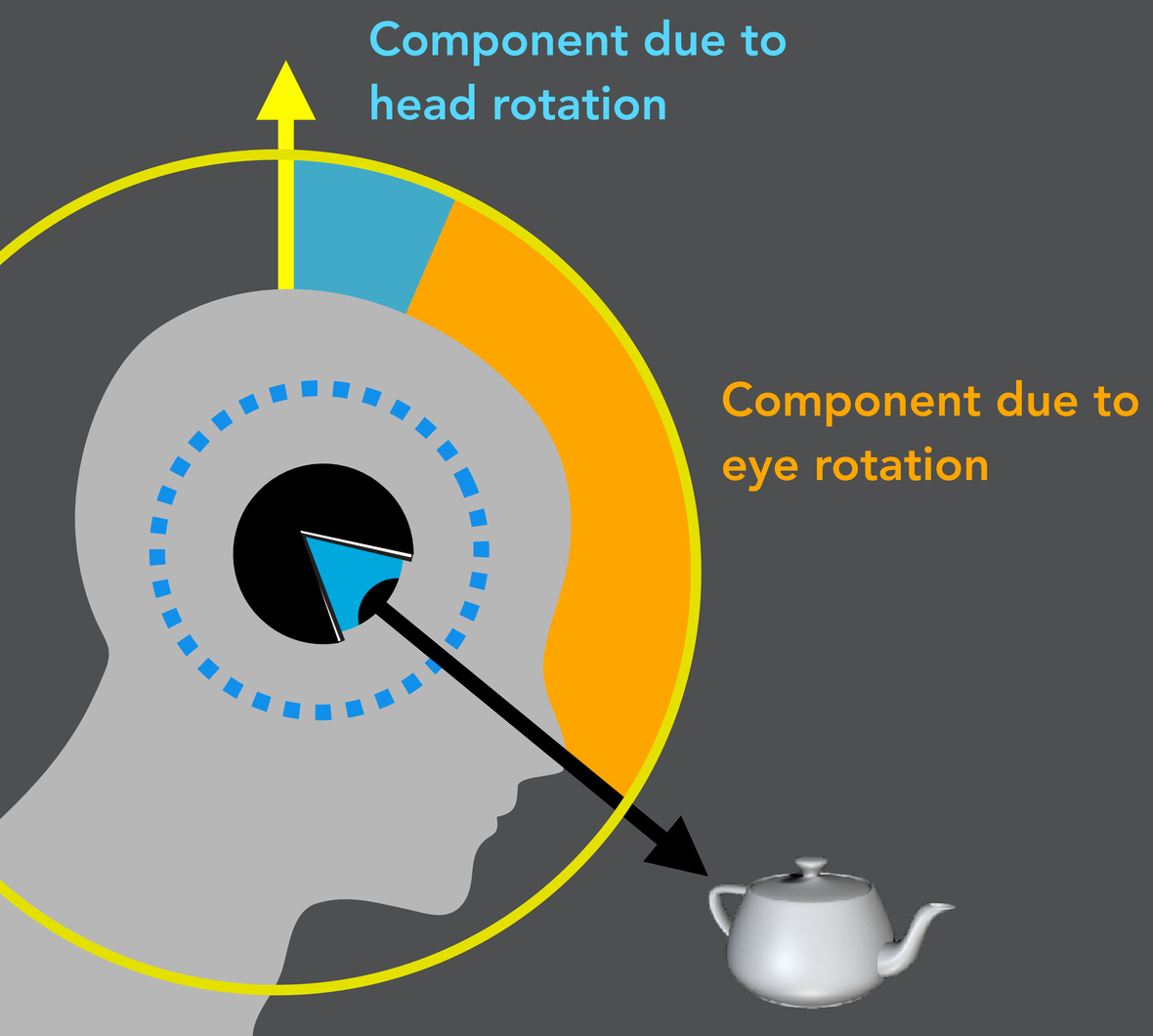
The position, scale, and orientation of the virtual camera and objects in the world are encoded using these matrices.

NEED TO BRUSH UP ON MATRICES?

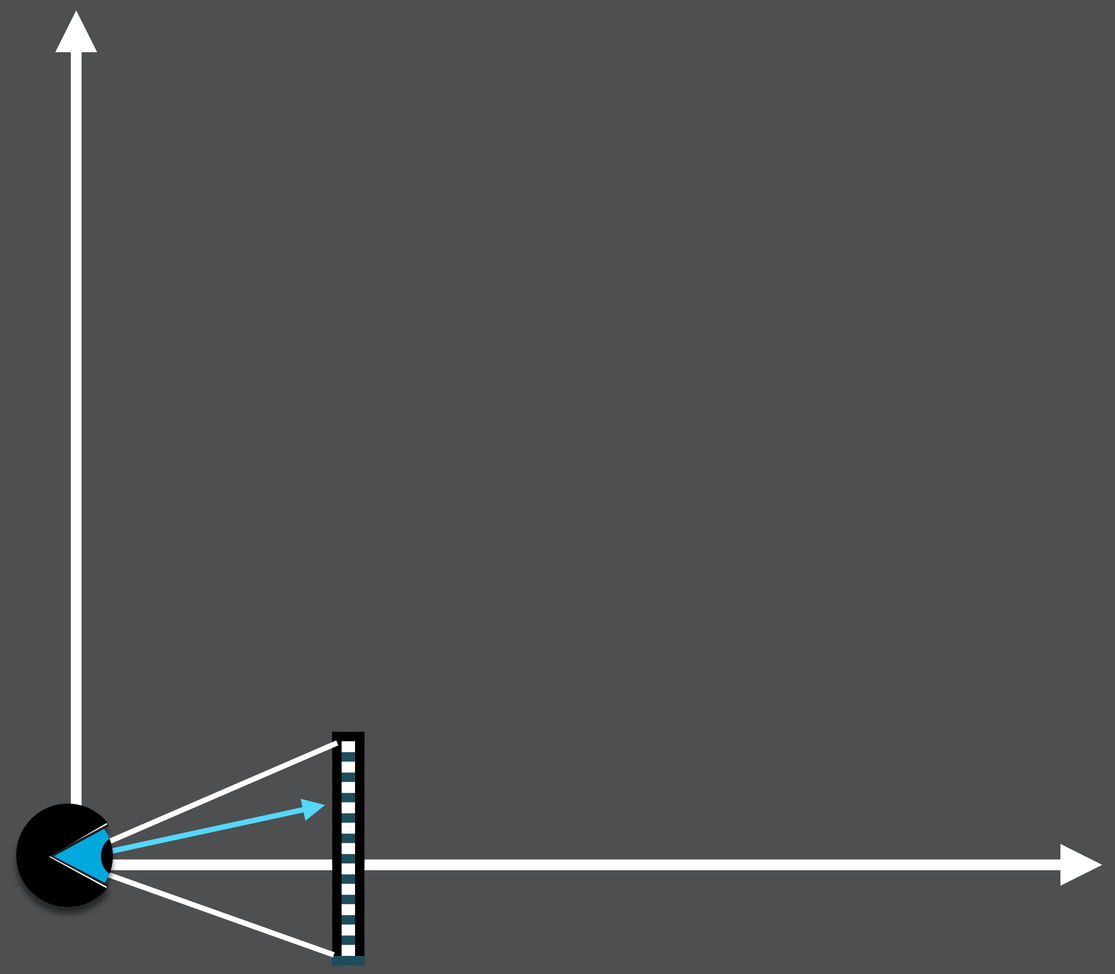
http://www.scratchapixel.com/lessons/mathematics-physics-for-computer-graphics/geometry/matrices

ANALYSIS - CALCULATING GAZE IN WORLD (USING TRANSFORMATION MATRICES)

FROM EYE IN HEAD TO GAZE-IN-WORLD



EYE-IN-HEAD TO GAZE IN WORLD

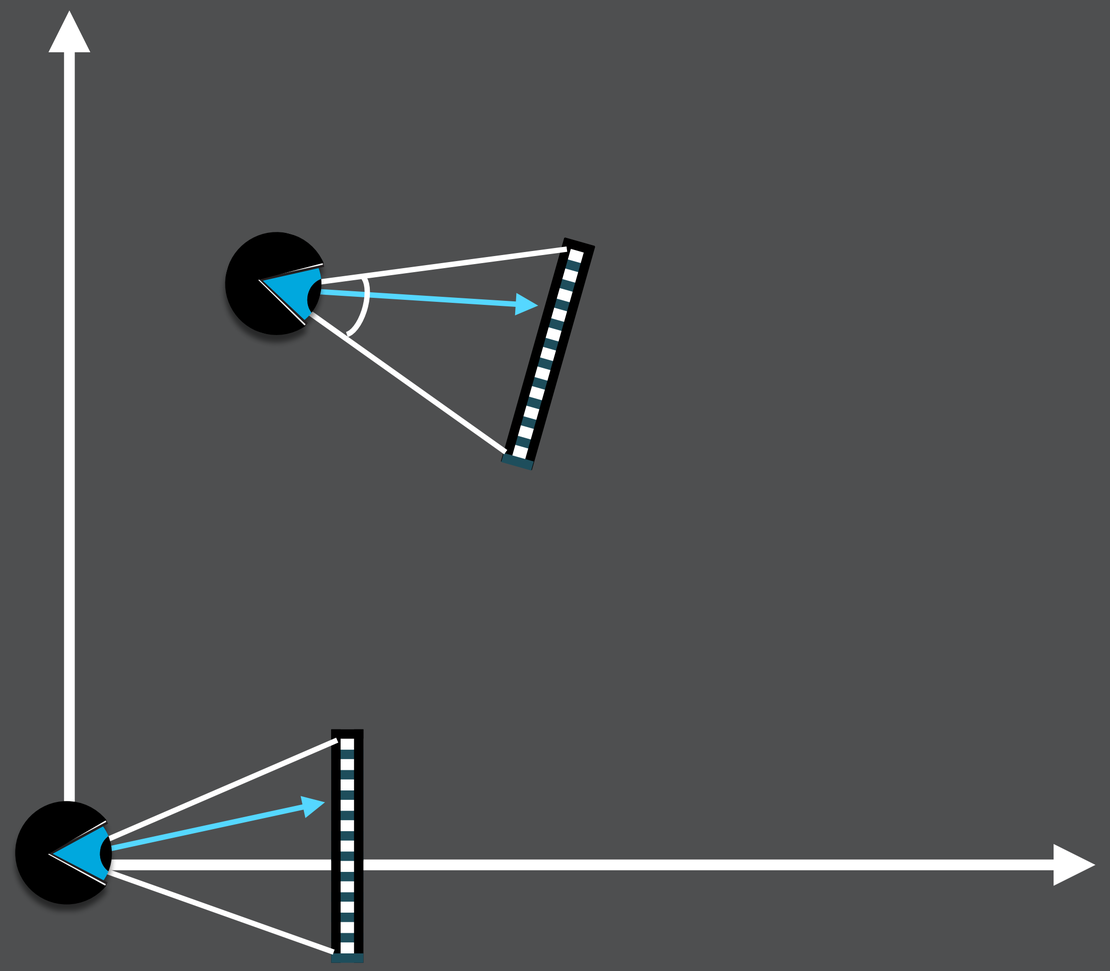


The eye-tracker provides the direction of the eye-in-head vector **(e)** to a point of reguard on the HMD screen [x,y,z].

It is a unit vector.

$\mathbf{e} = [x,y,z]$

$|\mathbf{e}| = \mathrm{sqrt}(x^2 + y^2 + z^2) = 1$



Step 1: Transform eye in head into gaze in world

THE TRANSFORMATION MATRIX CAN BE COMPOSED FROM A COMBINATION OF LINEAR OPERATIONS

(E.G. TRANSLATIONS, ROTATIONS, SCALING)

translation/rotation of the head

eye in head

e

gaze in world

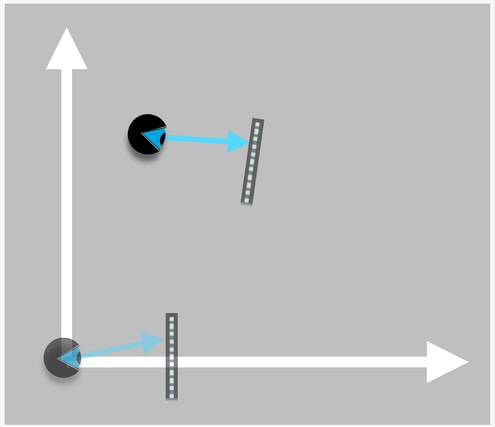
g

$$\begin{bmatrix} .96 & .07 & -0.28 & -0.98 \\ -.01 & 0.98 & 0.19 & 1.5 \\ 0.29 & -0.18 & 0.94 & 0.14 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}$$

THE STARTING POINT: EYE-IN-HEAD DATA

in python…

```python
import numpy as np

# Lets imagine the subject is looking down the Z axis
eihDir_xyz = [0,0,1]

# add a 1 prior to multiplication!
eihDir_xyzw = np.hstack([eihDir_xyz,1]) # now, it's [0,1,0,1]

# Retrieve the 4x4 matrix from my recorded data
transf_4x4 = sessionDict['raw']['viewMat'].values[fr].reshape([4,4]).T

# Multiply the transformation matrix by the EIH vector
# This converts EIH into GIW
xyzwOut = np.dot(transf_4x4,eihDir_xyzw)
```



A unit vector representing the direction of the gaze in world vector!

```python
# xyzwOut
```

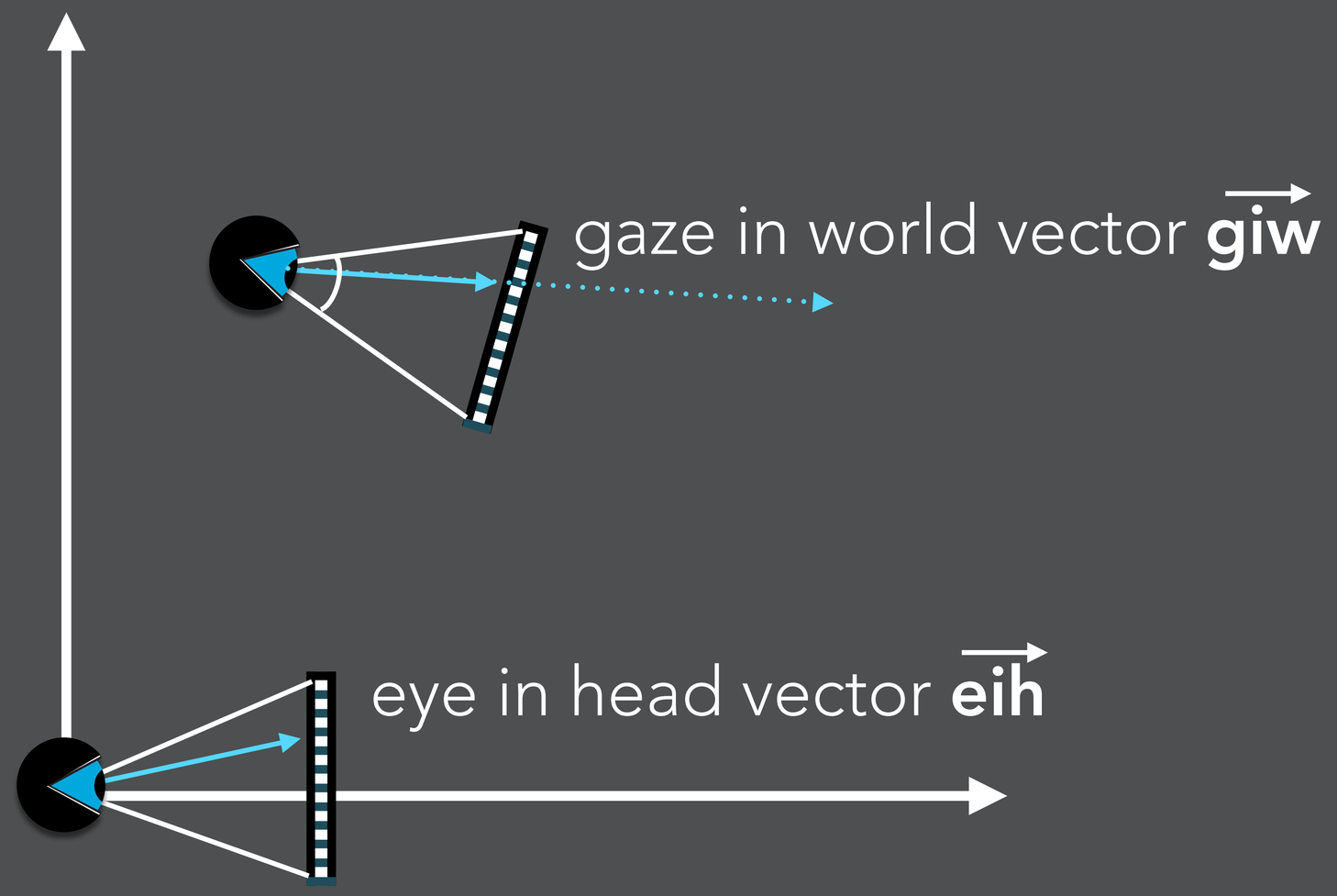
```
array([-1.1077543497085571, 1.6018794104456902, 1.6959261298179626, 1.0], dtype=object)
```

Step 1: Move the head and EIH direction into its 3d position and orientation.

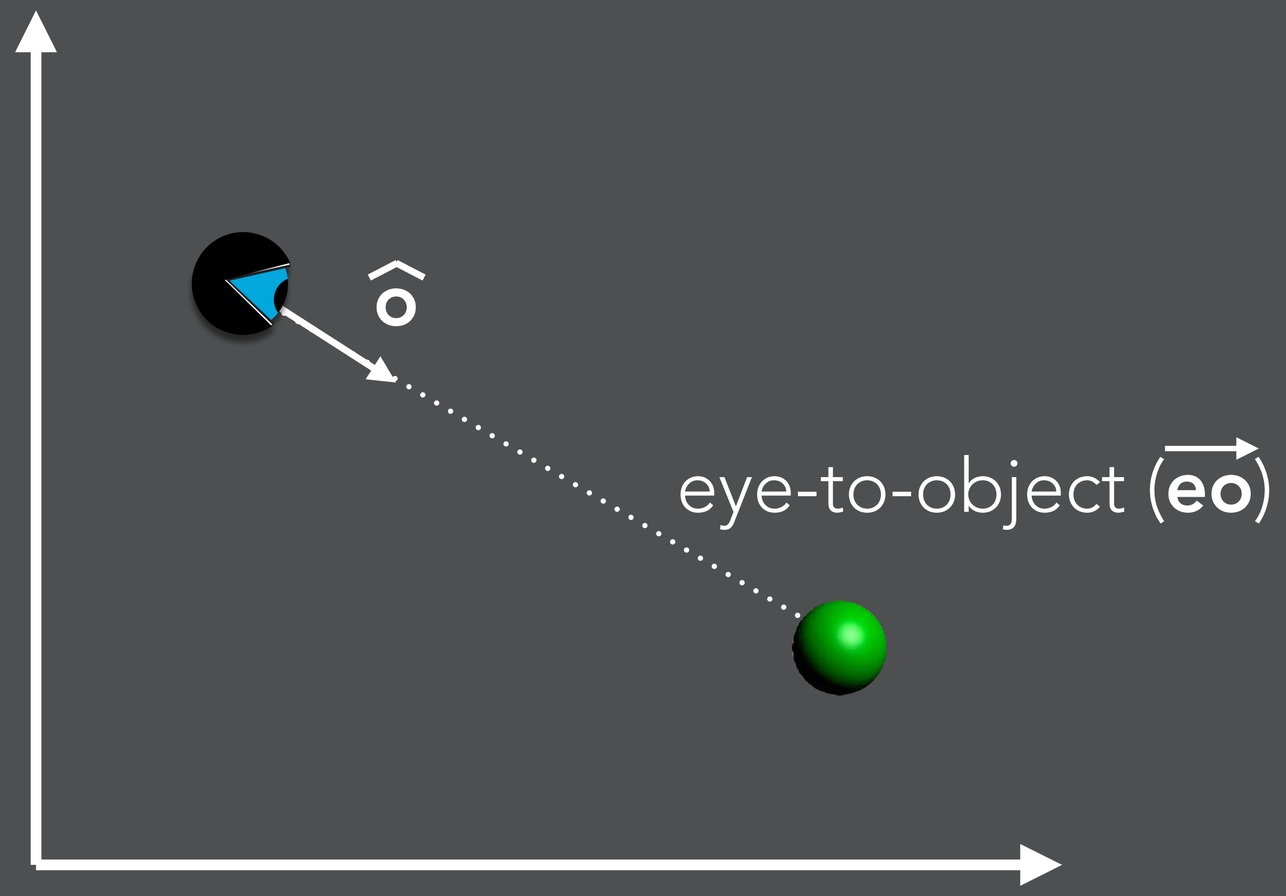
NOTEBOOK 2 : EIH TO GIW

ANGULAR DISTANCE WITHIN THE WORLD COORDINATE SYSTEM

Step 1: Transform eye in head into gaze in world

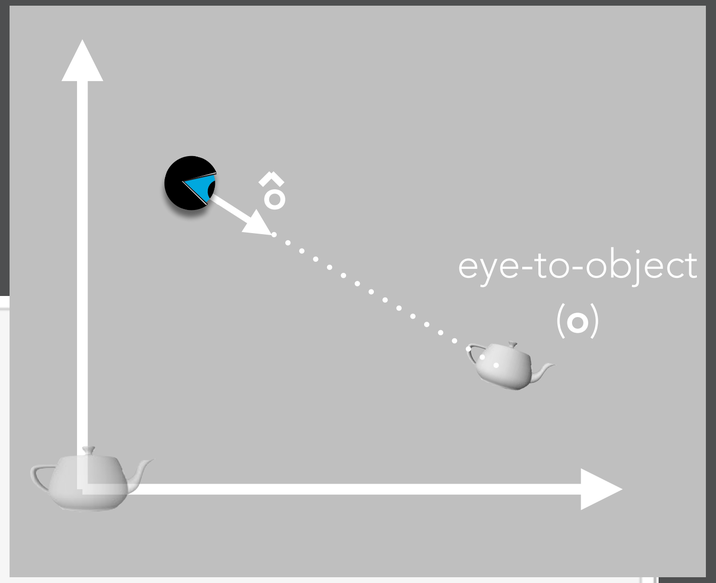


Step 2: Normalize eye-to-ball vector (**o**) so that its length is 1.



$$\hat{\mathbf{o}} = \frac{[o_x, o_y, o_z]}{\mathrm{sqrt}(o_x^2 + o_y^2 + o_z^2)}$$

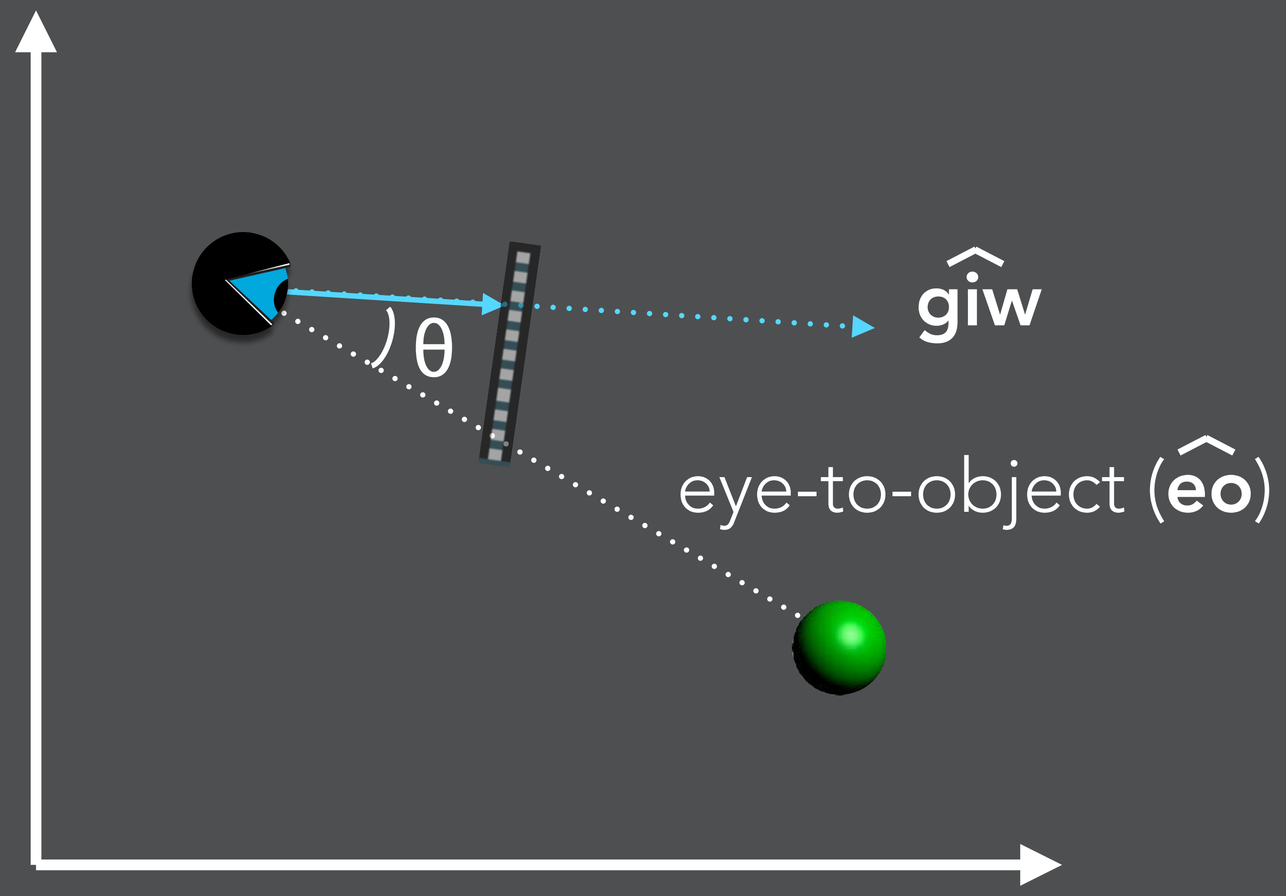
VECTOR NORMALIZATION




```python
import numpy as np

objDist_xyz = [1,2,3]

# Calculate the magnitude
mag = np.linalg.norm(objDist_xyz)

print 'Original values: %1.2f , %1.2f, %1.2f' % (objDist_xyz[0], objDist_xyz[1], objDist_xyz[2])
print 'Unnormalized mag: %1.2f\n' % np.linalg.norm(objDist_xyz)


# Normalize the distance, so that it's a unit vector
# indicating direction
objDir_xyz = objDist_xyz / np.linalg.norm(objDist_xyz)
normMag = np.linalg.norm(objDir_xyz)

print 'New values: %1.2f , %1.2f, %1.2f' % (objDir_xyz[0], objDir_xyz[1], objDir_xyz[2])
print '  Normalized mag: %1.2f' % normMag
```

```
Original values: 1.00 , 2.00, 3.00
Unnormalized mag: 3.74

New values: 0.27 , 0.53, 0.80
  Normalized mag: 1.00
```

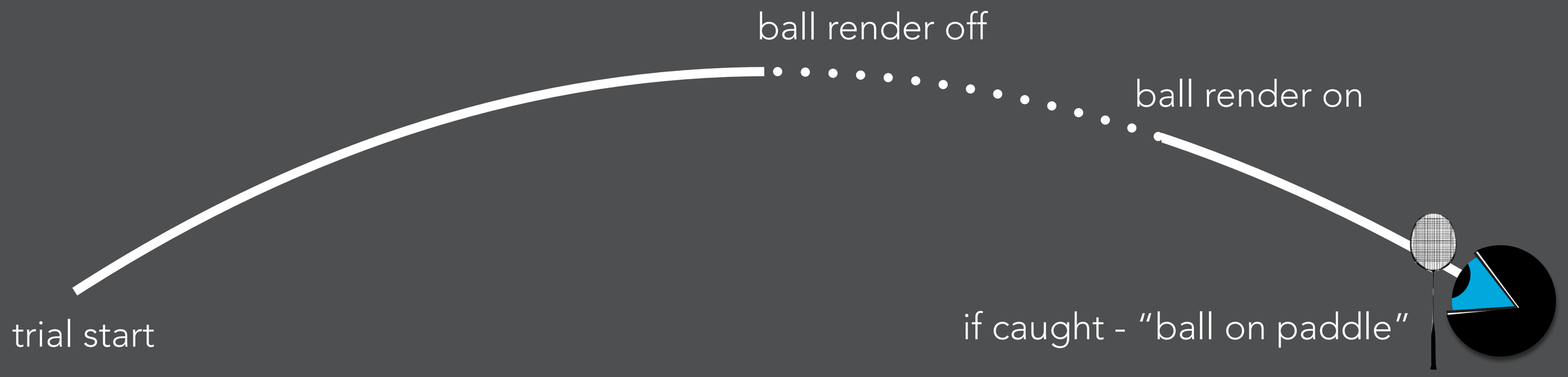
Step 3: Calculate gaze-to-object distance (degrees).

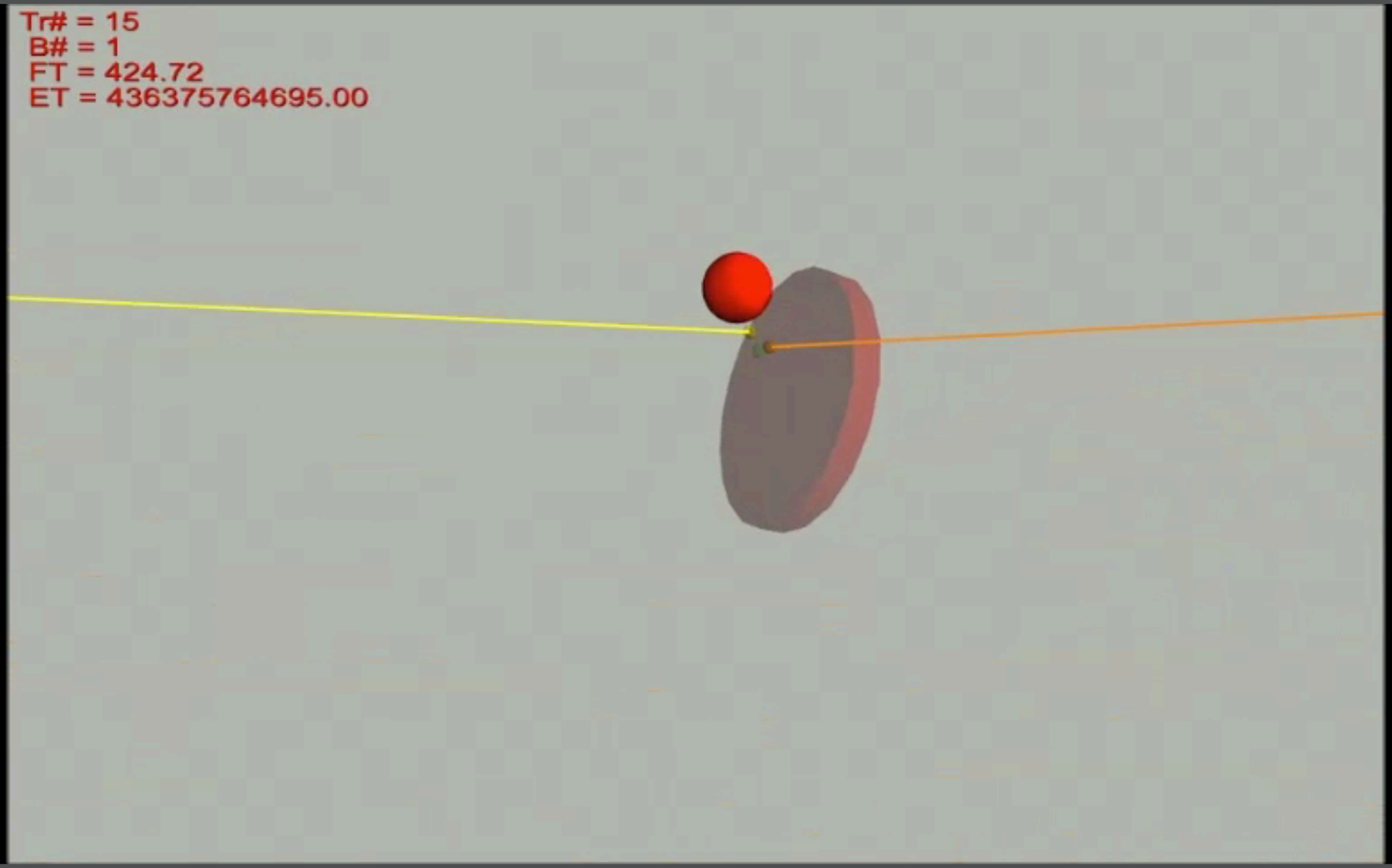


$$\theta = \arccos\left(\frac{\widehat{giw} \bullet \widehat{eo}}{|\widehat{giw}||\widehat{eo}|}\right)$$

The data is from a single frame of a ball catching task.

Event labels

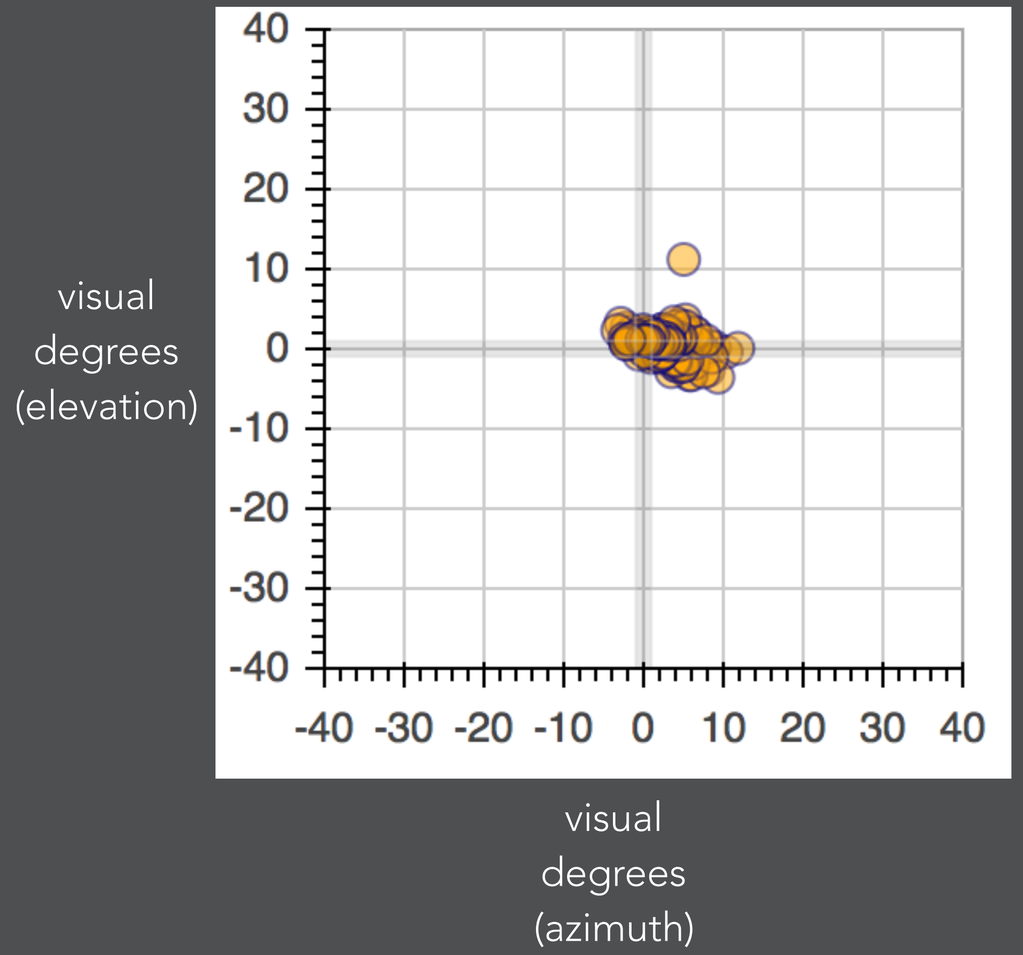


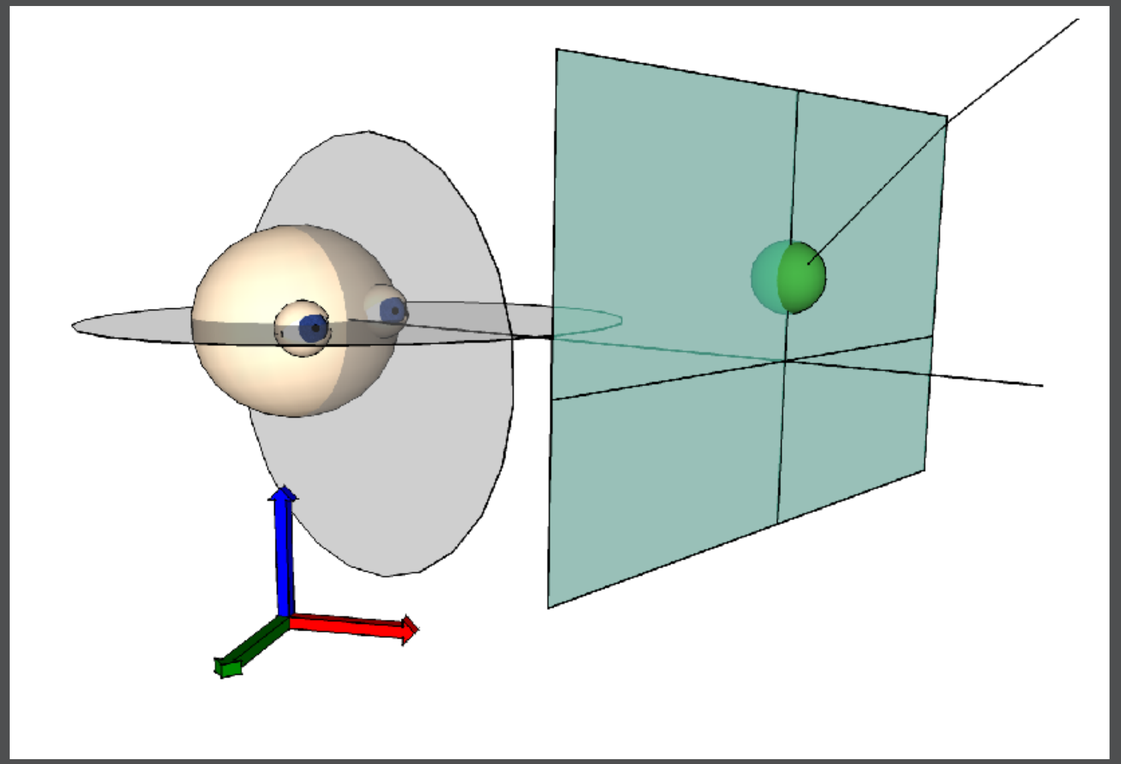

Tr# = 15
B# = 1
FT = 424.72
ET = 436375764695.00

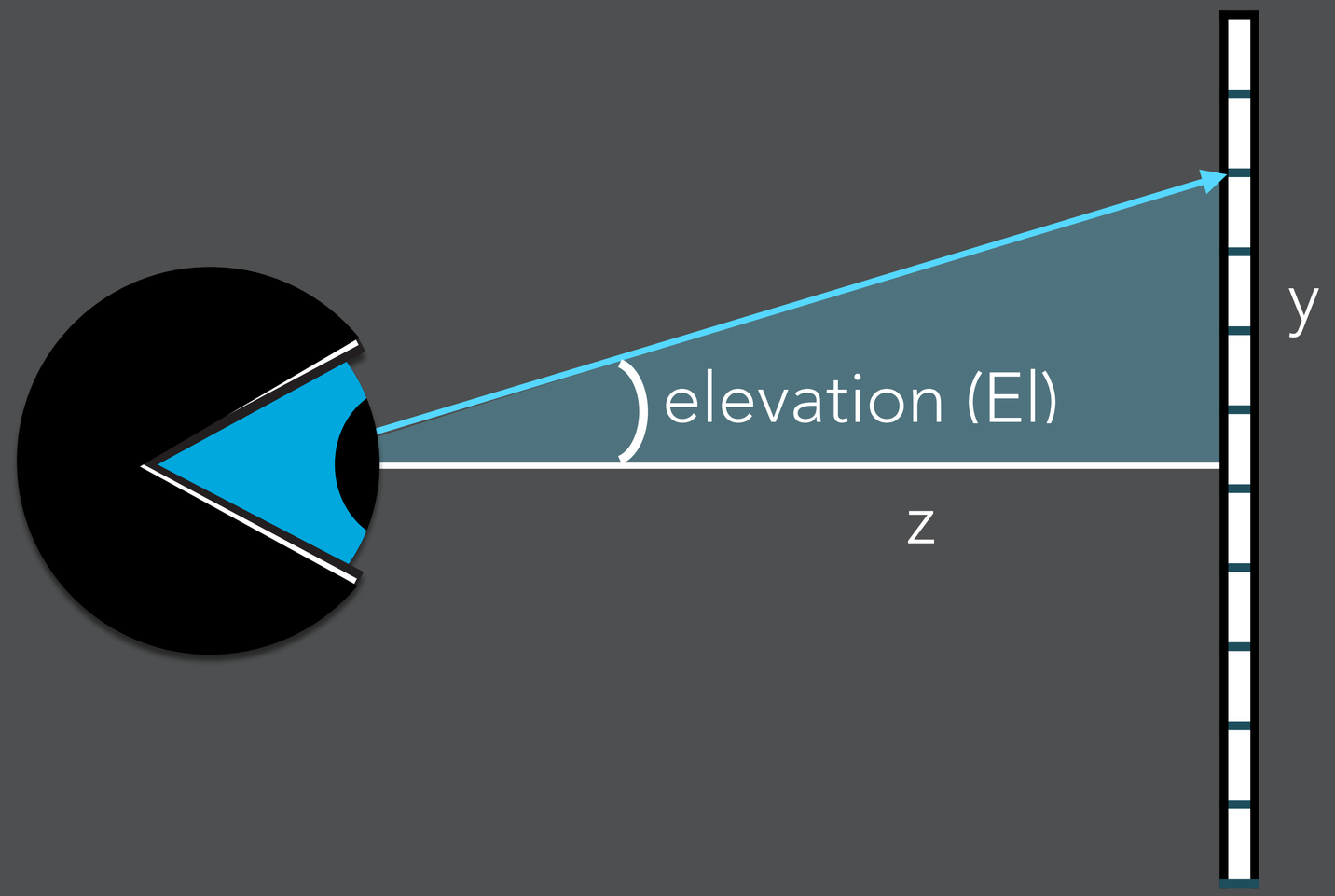
Where is the person looking relative to the object?

the ball's angular position relative to the cyclopean gaze vector

data points (orange circles) are not scaled to angular ball size.







$$Az = atan(x/z) *(180/pi)$$

$$El = atan(y/z) *(180/pi)$$

NOTEBOOK 3 :ANGULAR DISTANCE

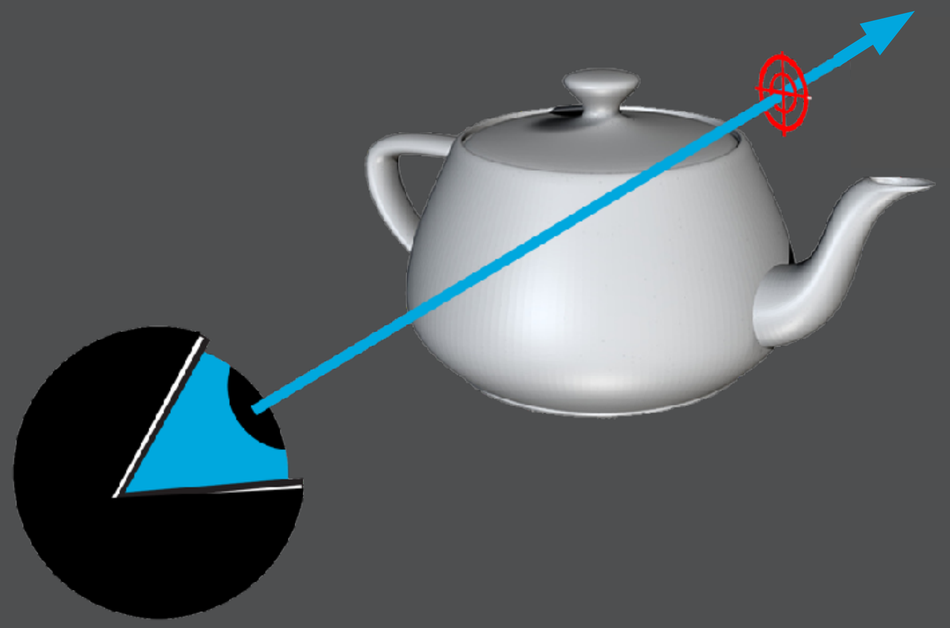
DETERMINING THE OBJECT OF FIXATION

The Analysis Pipeline

- **What** is the person looking at?
- **Where** is the person looking relative to the object?
- **How** are movements of the eyes and head coordinated to bring about shifts in gaze?

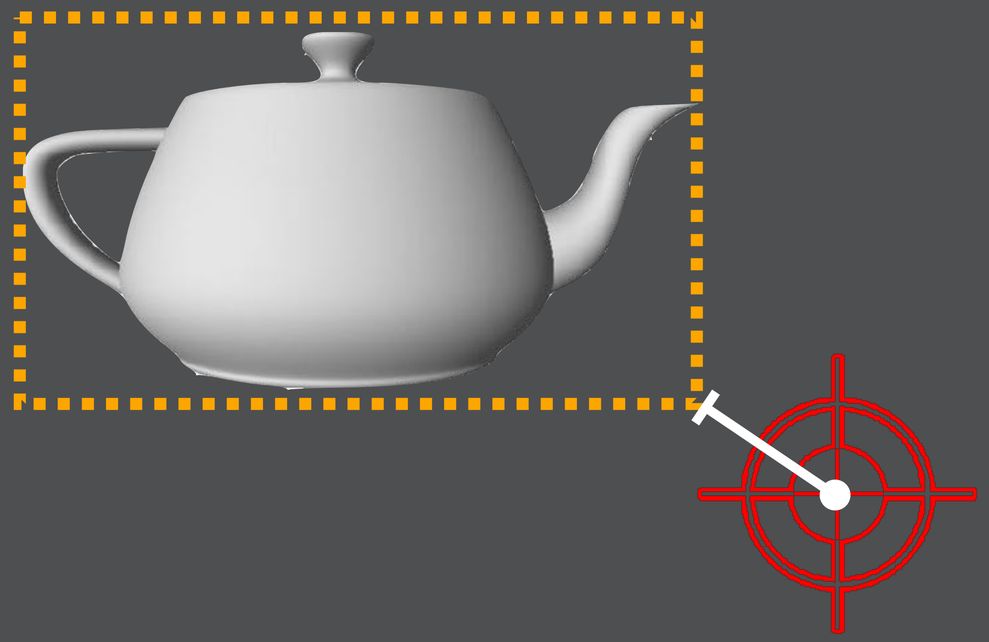
WHAT IS THE PERSON LOOKING AT?

Due to biological and tracker noise, fixations will often fall near to, but not on the target object. This makes detecting fixations via ray-object collision ineffective.



WHAT IS THE PERSON LOOKING AT?

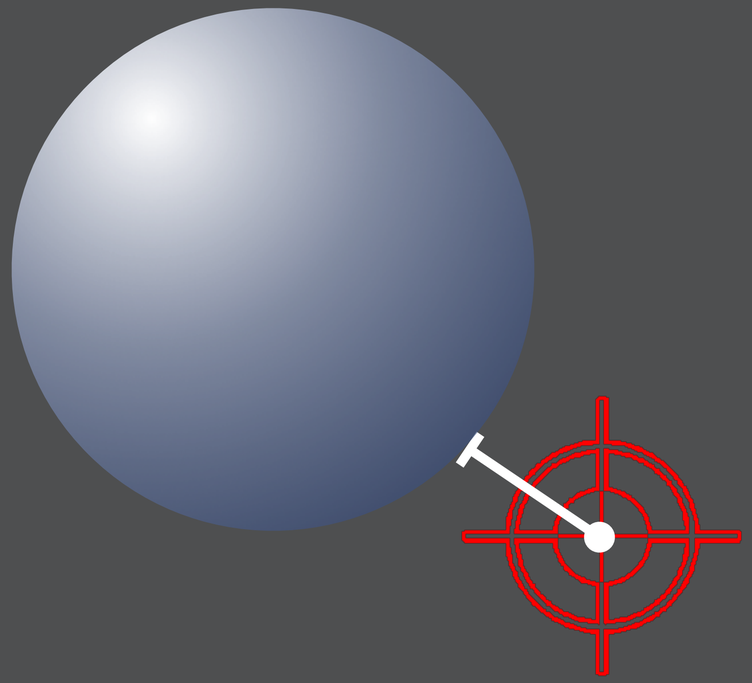
Bounding boxes or hull-based operations are poor approximations.



We are still in need of solutions for minimum angle-to-hull distances.

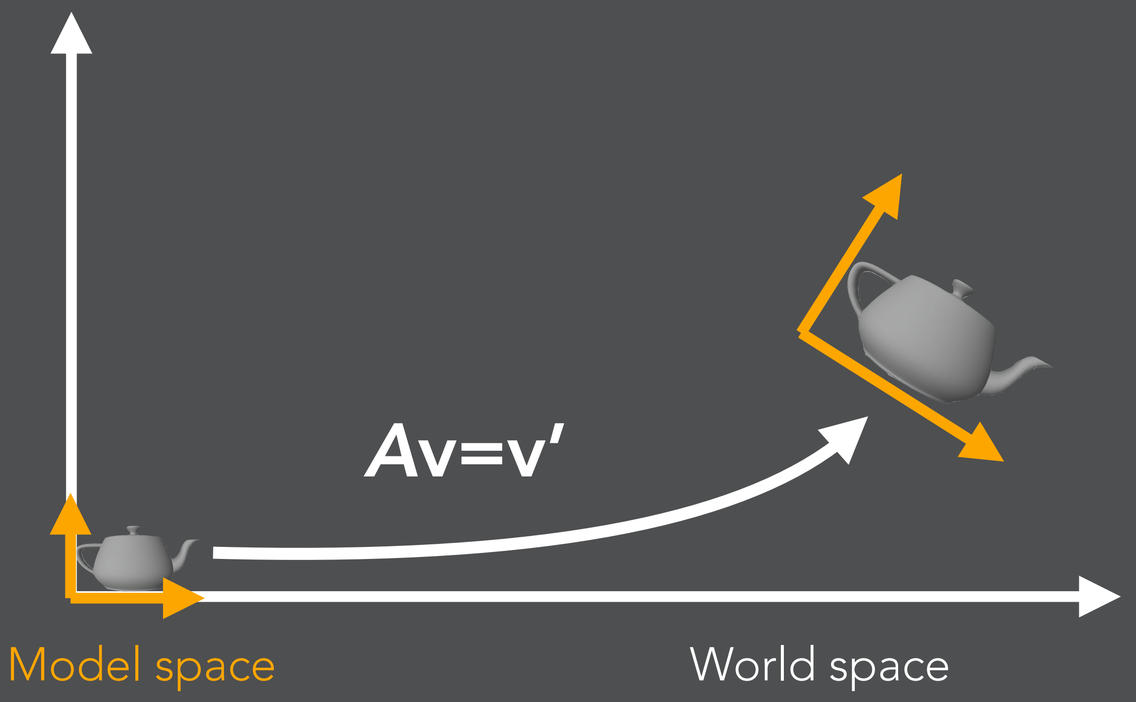
NOTEBOOK: ANGULAR DISTANCE

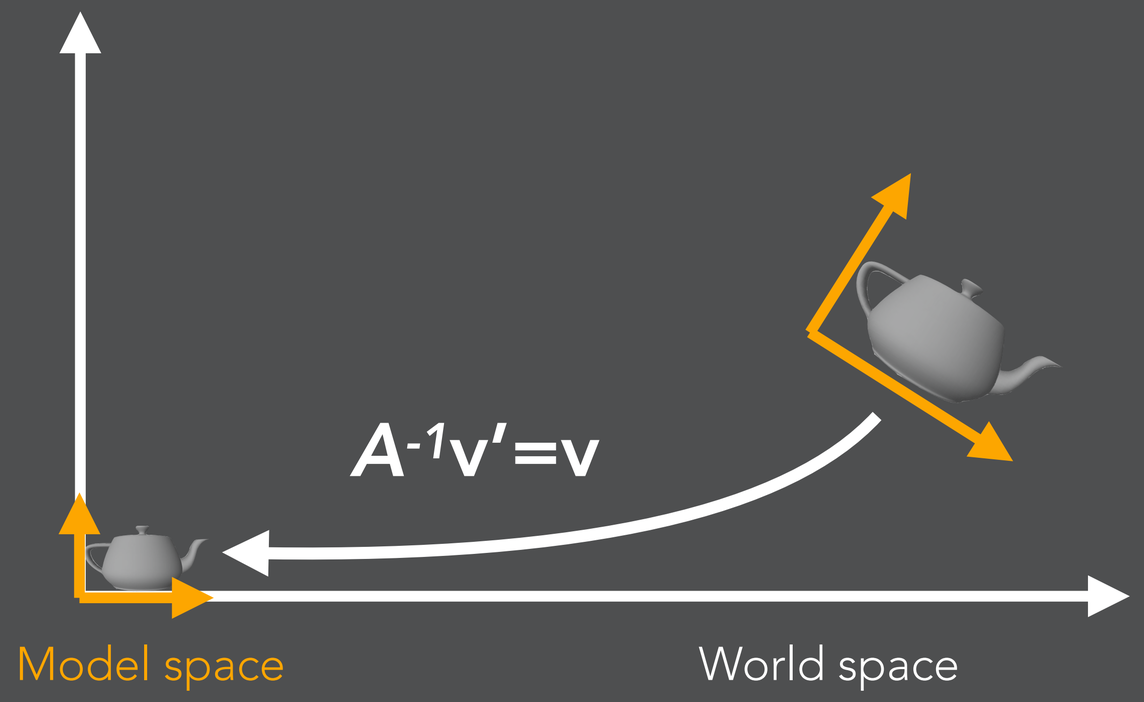
Things are much easier if your world is full of spheres!



CHANGING COORDINATE SYSTEMS

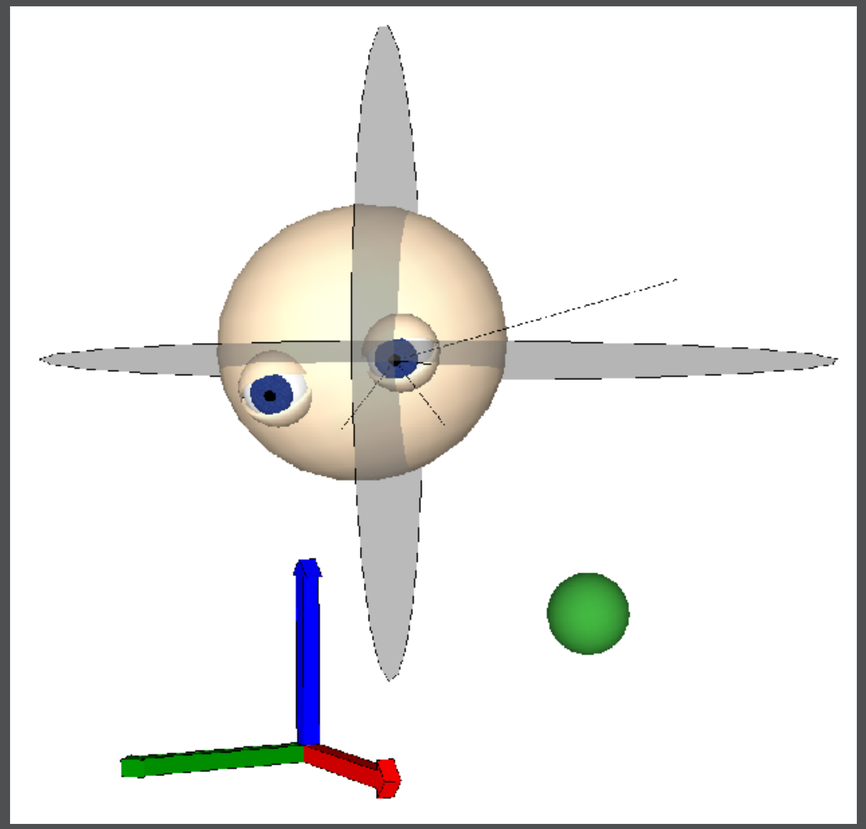
Just as the model matrix ***A*** is used to move all vertices **v** from model space to world space….



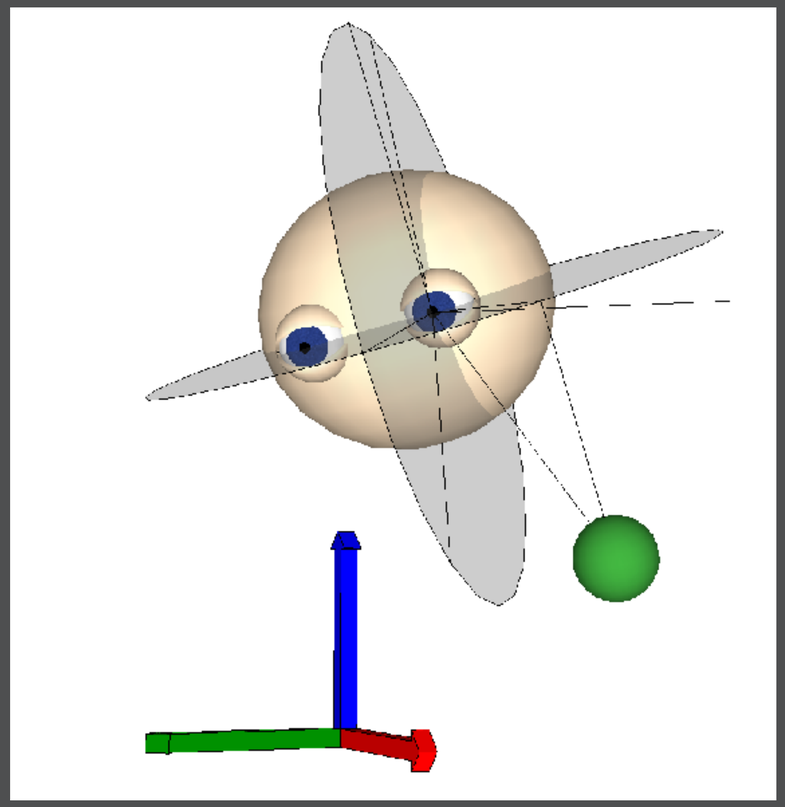


One can also use the inverse of ***A*** to move in the other direction (e.g. from world space to model-space).

ANALYSIS IN DIFFERENT COORDINATE FRAMES



Axis aligned with gravity.

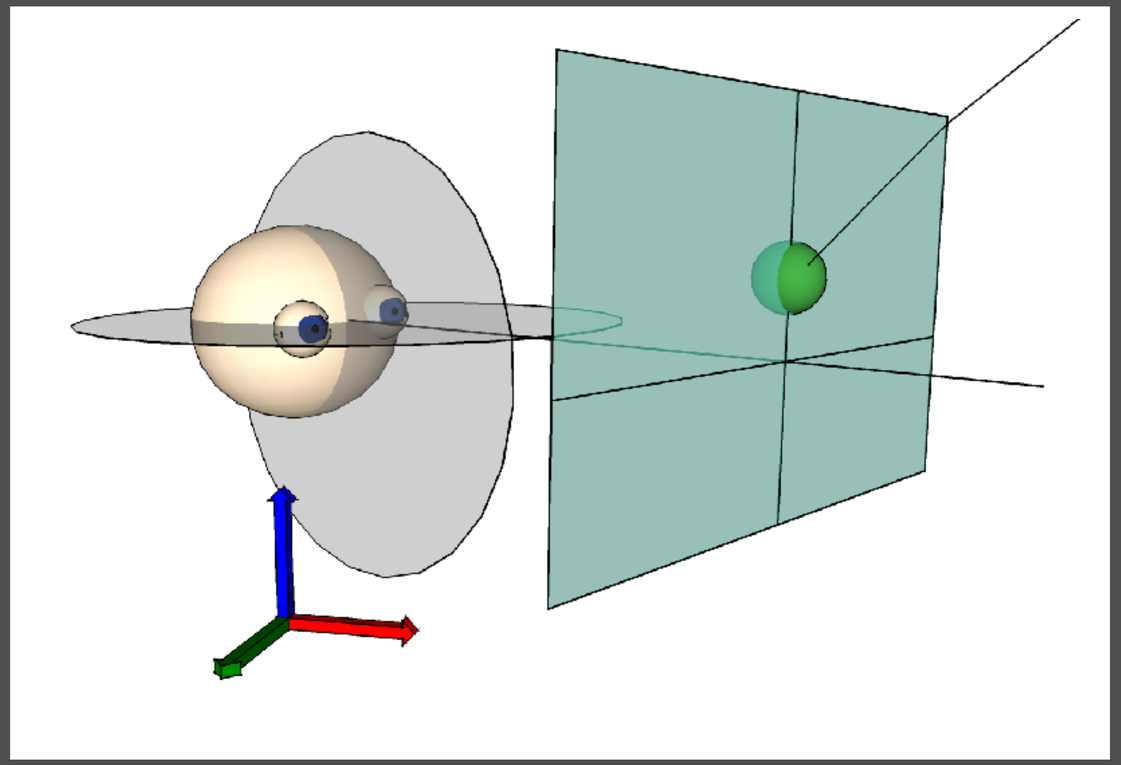


Axis aligned with head.

Where is the person looking relative to the object?

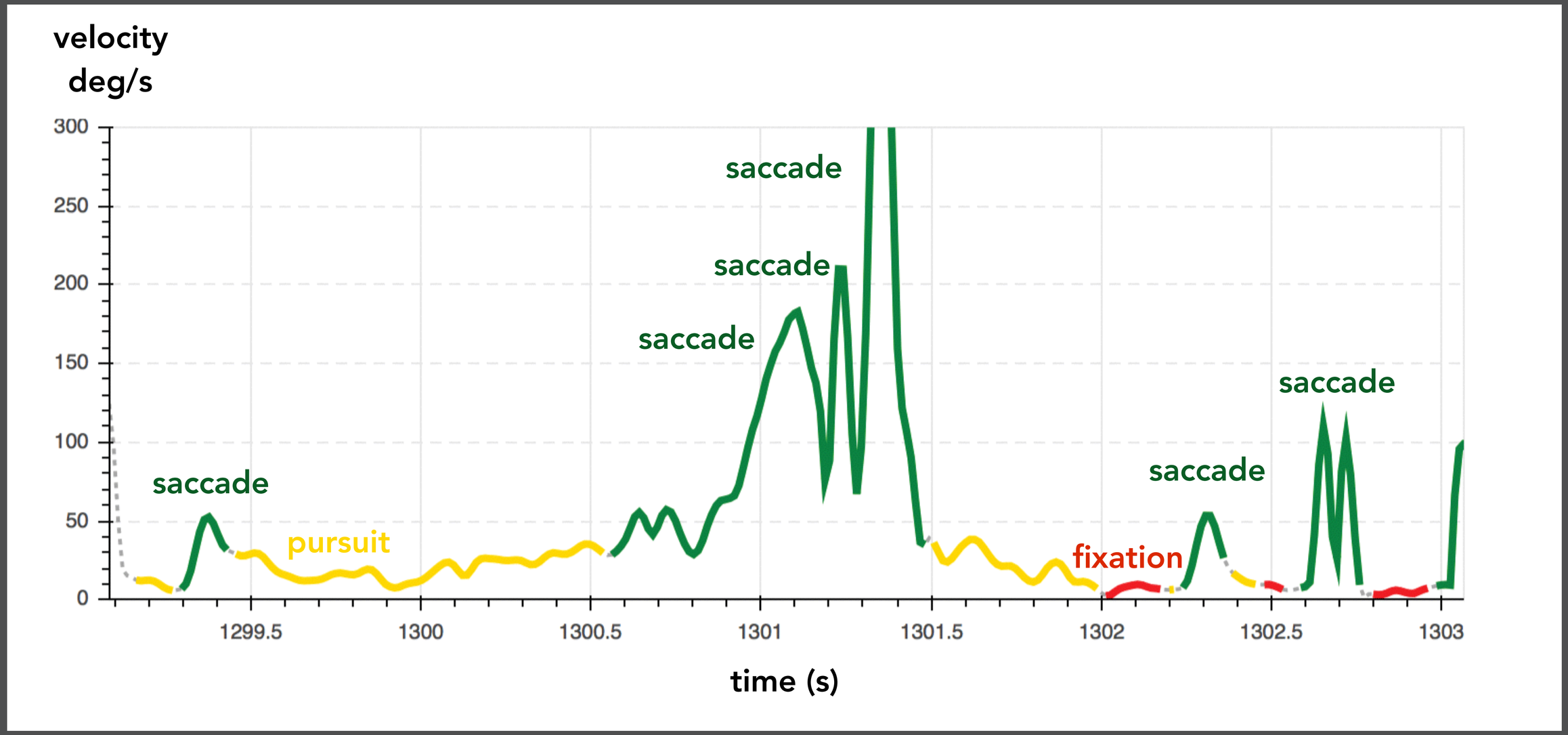
the ball's angular position relative to the cyclopean gaze vector

data points (orange circles) are not scaled to angular ball size.

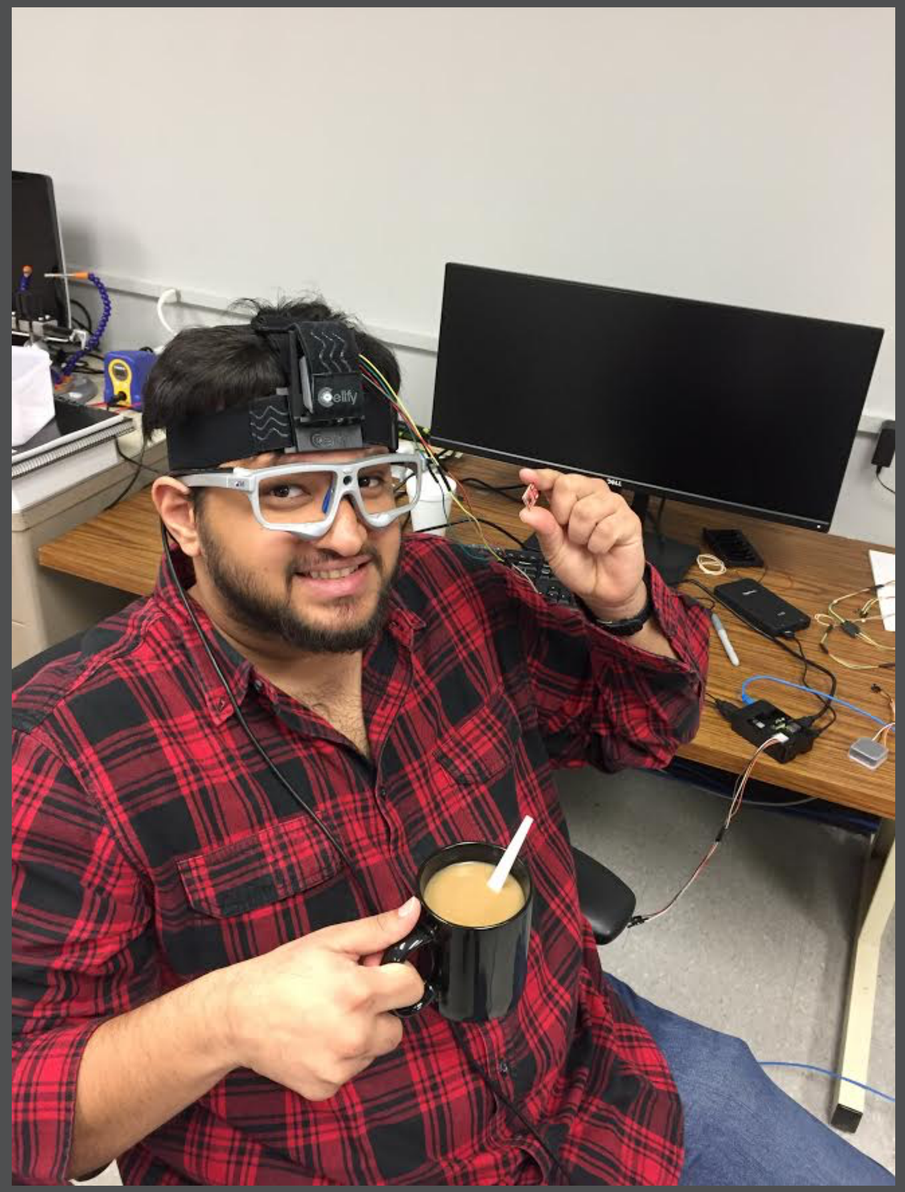


FILTERING AND
GAZE EVENT CLASSIFICATION

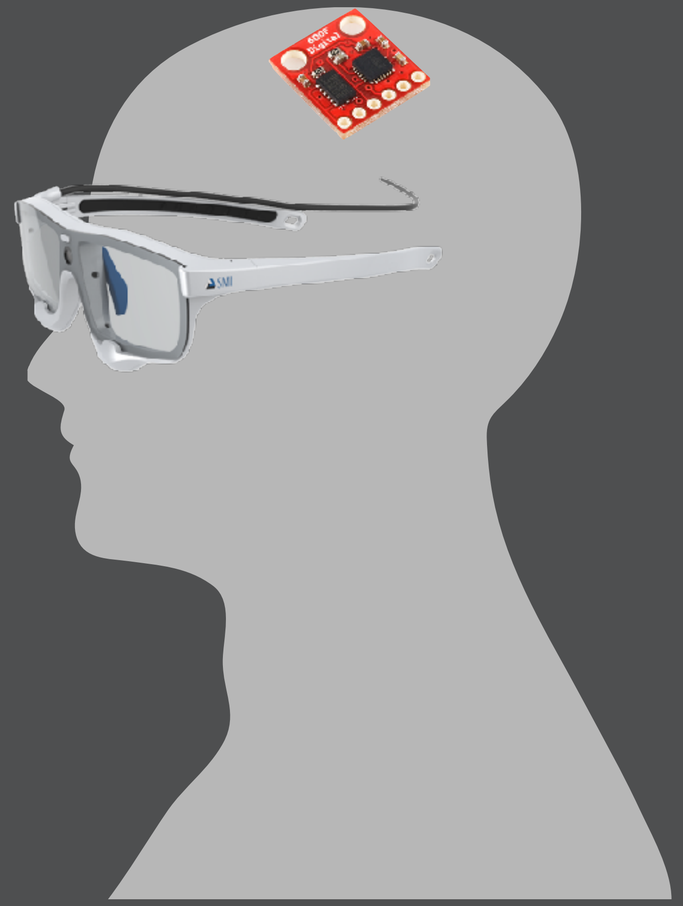
HOW ARE MOVEMENTS OF THE EYES AND HEAD COORDINATED TO BRING ABOUT SHIFTS IN GAZE?



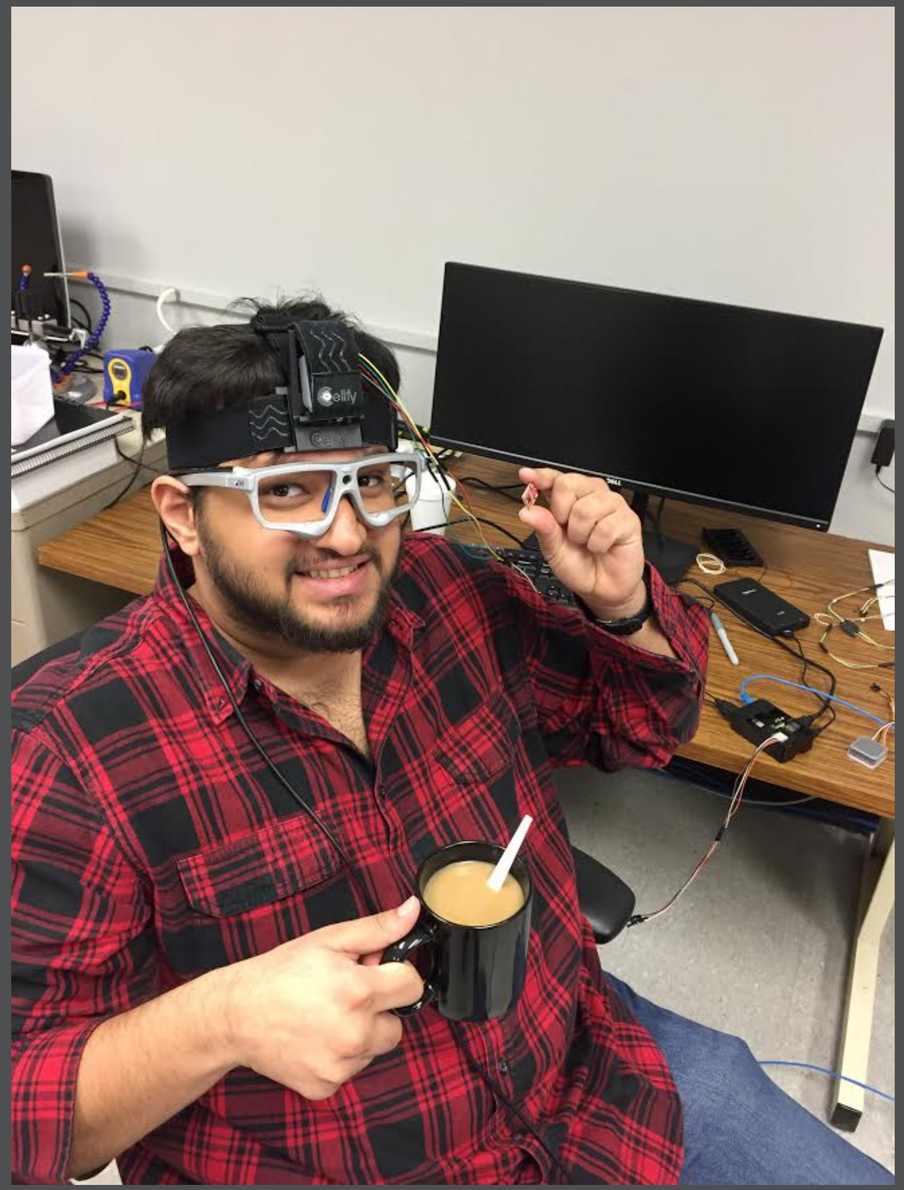
Data from an SMI 60 Hz tracker integrated into a Oculus DK2



Rakshit Kothari



Eye tracker (eye-in-head vector)
IMU (rotational velocity of the head)

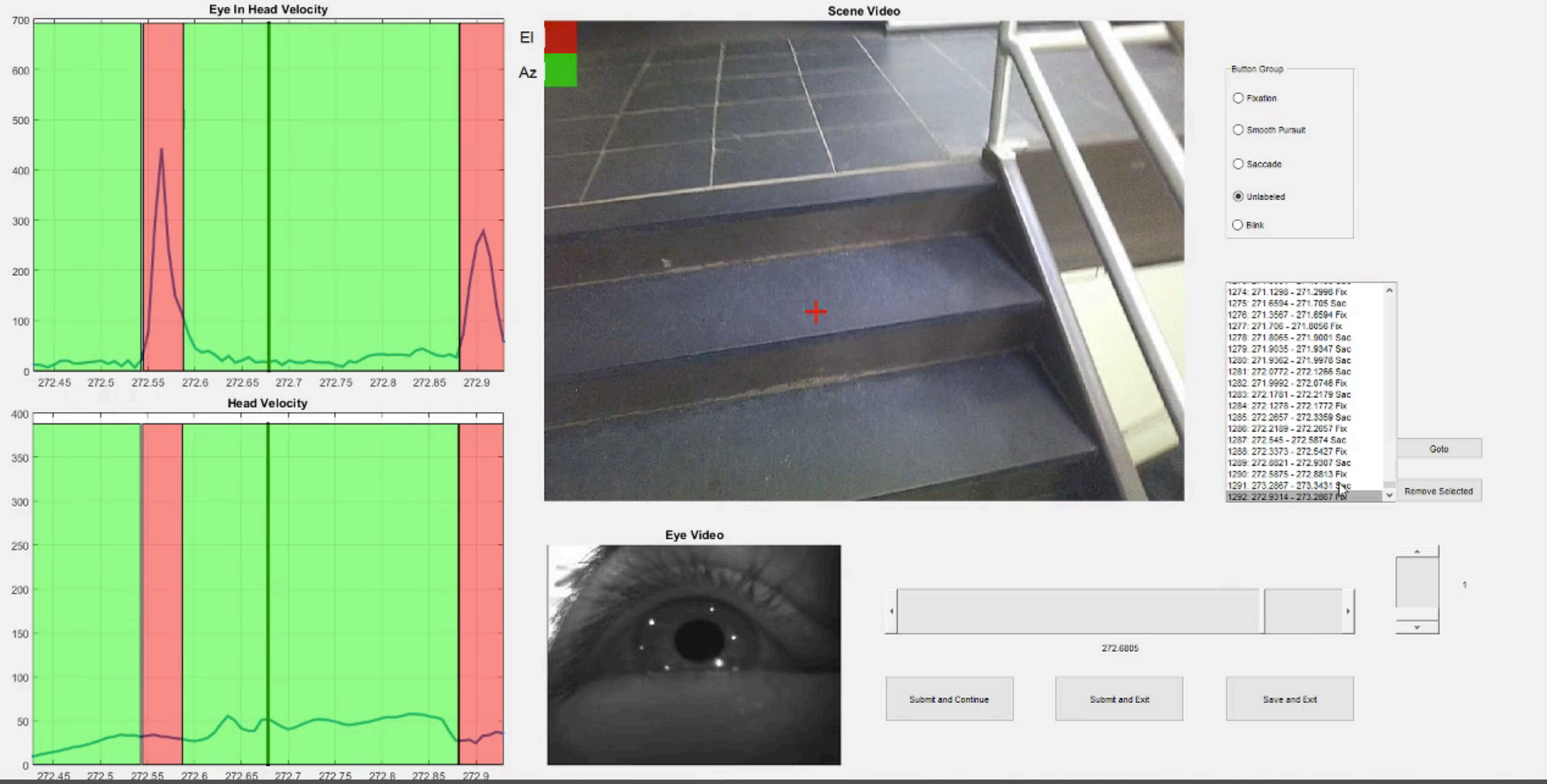


Rakshit Kothari

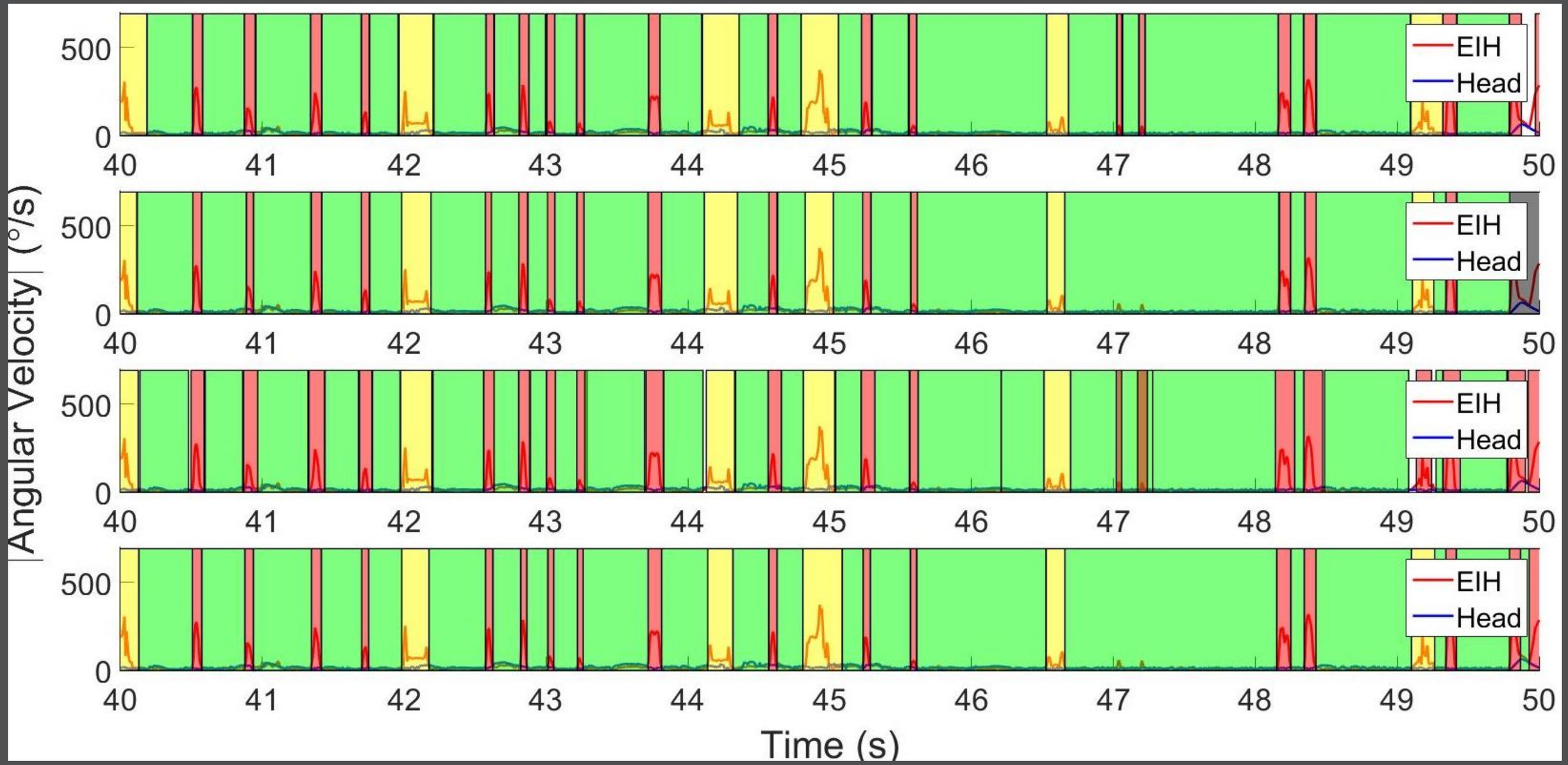
Supervised machine learning for gaze event classification

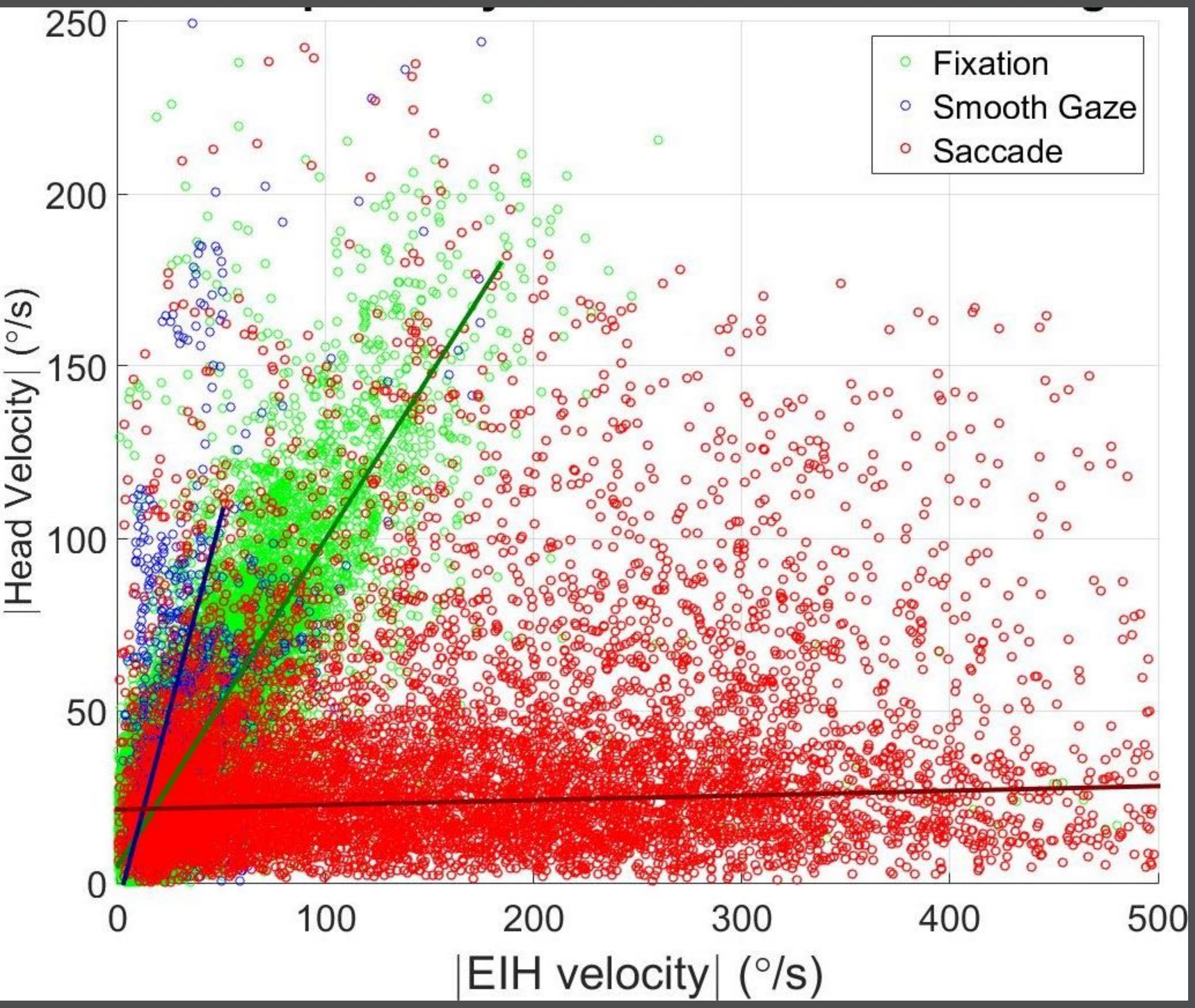
Participant tasks:

- Walk while talking to a friend
- Catch a ball
- Make a sandwich


Eye In Head Velocity
700
600
500
400
300
200
100
0
272.45
272.5
272.55
272.6
272.65
272.7
272.75
272.8
272.85
272.9
Head Velocity
400
350
300
250
200
150
100
50
0
Scene Video
El
Az
Button Group
Fixation
Smooth Pursuit
Saccade
Unlabeled
Blink
1274: 271.1298 - 271.2998 Fix
1275: 271.6594 - 271.705 Sac
1276: 271.3567 - 271.6594 Fix
1277: 271.706 - 271.8056 Fix
1278: 271.8065 - 271.9001 Sac
1279: 271.9035 - 271.9347 Sac
1280: 271.9362 - 271.9978 Sac
1281: 272.0772 - 272.1266 Sac
1282: 271.9992 - 272.0748 Fix
1283: 272.1781 - 272.2179 Sac
1284: 272.1278 - 272.1772 Fix
1285: 272.2657 - 272.3359 Sac
1286: 272.2189 - 272.2657 Fix
1287: 272.545 - 272.5874 Sac
1288: 272.3373 - 272.5427 Fix
1289: 272.8821 - 272.9307 Sac
1290: 272.5875 - 272.8813 Fix
1291: 273.2867 - 273.3431 Sac
1292: 272.9314 - 273.2867 Fix
Goto
Remove Selected
Eye Video
272.6805
Submit and Continue
Submit and Exit
Save and Exit

Training data weighted according to inter-rater consensus/reliability.



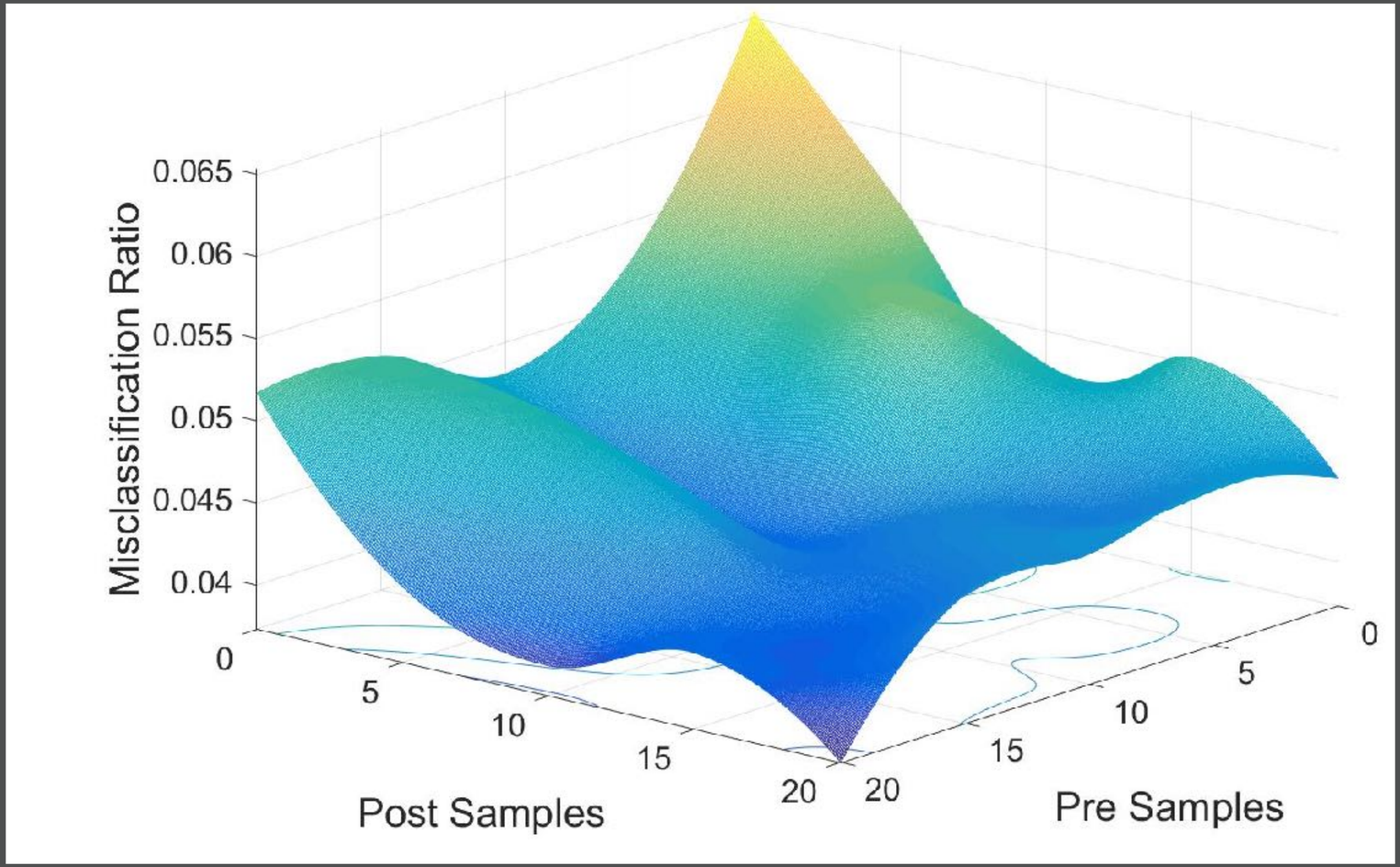


Separability of individual samples within velocity space.

Additional feature spaces will improve separability.

	Pre	Post	Misclass (%)	Fix acc (%)	SG acc (%)	Sacc acc (%)
NN*	15	15	4.29	98.16	60.30	81.54

Smooth gaze movements are underrepresented in the *skewed* dataset.



WHAT TO INCLUDE IN YOUR METHODS SECTION / WHAT WRITE IN YOUR GRANT PROPOSALS

- **HMD/Display**
  - Spatial density (pixels/degree)
  - Field of view (X/Y)
  - Stereo overlap
  - Weight

- **Motion capture**
  - Sampling rate
  - Latency
  - Measures of spatial accuracy (less common)

- **Eye tracker**
  - Calibration accuracy (in central/peripheral visual field)
  - Sampling rate
  - Latency

THE FUTURE OF EYE TRACKING IN 3D

Blur
can be used to convey

THANK YOU

- Gabriel Diaz

PERFORM LAB

(Perception For Movement)