

# **EE3801 Data Engineering Principles**

## **Data Pipelines**

# Data Pipelines

## What you will learn

- Hands on processing of large dataset (140 GB)
- 45 days of recordings of eye movements and behavioral data
- 3 days each with 40 GB of neural data
- Staged processing of data, with subsequent analyses dependent on earlier analyses
- Optimization of pipeline, processing algorithms, and data visualization

# Data Pipelines

## Organization

Week	Monday	Thursday
7	Quiz 1 (Part 1)	Lecture
8	Lab 4	Lecture
9	Lab 5	Lecture
10	Lab 6	Lecture
11	Lab 7	Lecture
12	Lab 8	Lecture
13	Public Holiday	Quiz 2 (Part 2)

Part A of each lab will have to be completed in class so TAs can help you get started.

# Data Pipelines

## Organization

- Monday (labs) and Thursday (lectures)
- 5 lab assignments (1 per week)
- Lab assignments will be released every Thursday
- Start working on the lab assignment before Mondays
- On Mondays:
  - Join pre-assigned lab group in E2-03-08 and E2-03-09
  - TAs will help you get going with the labs
  - Complete Part A in class and submit to Canvas
- Continue working on the lab assignment after class and submit Part B by Wednesday 9 pm

# Data Pipelines

## Groups in E2-03-08

▼ Group 1

10 students

⋮

⋮ ANG HUI KAI, DAV...	⋮ CALEB ANG SI KAI	⋮ CHIN WEI HERNG
⋮ IAN ISAIAH TAN JU...	⋮ LEE EN-YI HANNAH	⋮ LIAU KAI JIE
⋮ MA XUDONG	⋮ PANG CHIN KIAT, ...	⋮ PHUA ZHUO WEI
⋮ SENTHILKUMAR S...		

▼ Group 2

11 students

⋮

⋮ ANG LIN MIN, NIC...	⋮ BELLAKKA KRISHN...	⋮ BENJAMIN CHRIST...
⋮ BUI DUC THANH	⋮ CHAI ZONG LUN	⋮ CHAN KENG JIT
⋮ SI JENG HWEE	⋮ SIA KIAN ZHONG	⋮ TAY JIUN YUAN
⋮ YEO MENG HAN	⋮ ZHUANG SHENGBIN	

▼ Group 3

12 students

⋮

⋮ ALISJA BTE ZAQY	⋮ TAN YI XUAN	⋮ THA JUN LIM
⋮ TIMOTHY ANTONI	⋮ TRAN THI PHUON...	⋮ VARUN AGARWAL
⋮ WEI WEN JIE	⋮ YANG HONG	⋮ YANG YUCHEN
⋮ YEOH JUN TING	⋮ YONG JIA FENG	⋮ YUAN XINRUI

# Data Pipelines

## Groups in E2-03-09

### ▼ Group 4

12 students

⋮

⋮ CHEW JIA REN, JO...	⋮ DESMOND ENG KI...	⋮ GUAN DINGHE
⋮ GULATI SHOBHIT	⋮ GUO HUIQI	⋮ HUANG SHANSHAN
⋮ IMPERIAL EDWAR...	⋮ JAVIER SOON QUA...	⋮ JEREMIAH ONG RAY
⋮ JOAN LIEW YU MIN	⋮ KENNETH NG YEE ...	⋮ KHOO XIN ZHEN

### ▼ Group 5

11 students

⋮

⋮ PICHANON RATTA...	⋮ POJCHARAPOL LE...	⋮ RADIT MUHTASHI...
⋮ SAMUEL TAN SZE ...	⋮ SU YATING	⋮ SYED OMAR ZORAN
⋮ TAM LI NA	⋮ TAN HSIEN RONG	⋮ TAN LING JEN
⋮ TAN WEE SIANG	⋮ TAN WEI LI	

### ▼ Group 6

11 students

⋮

⋮ LAW YONG HUI G...	⋮ LEONG JUN KAI D...	⋮ LI JIANKUN
⋮ LIM YUH CHING	⋮ LINN HTET AUNG	⋮ LOH YEE KAI
⋮ NEO JUN QIAO	⋮ NGIENG YIH HONG	⋮ NIGEL NG
⋮ ONG QING EN	⋮ PANG KAI LIN	

# Data Pipelines

## Discussion Groups

- Post questions to Canvas Discussion Group X for your TA
- Search across Discussion groups for answers to your questions before posting

▼ Discussions

Ordered by recent activity

Group	Section	Actions
Group 6 (Teng Chiong)	All sections	Checkmark, bookmark, more
Group 5 (Haoxuan)	All sections	Checkmark, bookmark, more
Group 4 (Nicholas)	All sections	Checkmark, bookmark, more
Group 3 (Sharif)	All sections	Checkmark, bookmark, more
Group 2 (Sinan)	All sections	Checkmark, bookmark, more
Group 1 (Huy)	All sections	Checkmark, bookmark, more

# Data Pipelines

## Tips

- Familiarize yourself with the bash shell
  - <https://gist.github.com/LeCoupa/122b12050f5fb267e75f>
- Familiarize yourself with the nano editor
  - <https://www.nano-editor.org/dist/latest/cheatsheet.html>
- Familiarize yourself with GitHub and VS Code:
  - <https://lab.github.com/githubtraining/introduction-to-github>
  - <https://code.visualstudio.com/docs/sourcecontrol/overview>
- Read the lab instructions carefully
  - Typos matter
  - Check that you did not miss out a space or period when things do not work

# Data Pipelines

## AWS Account

- AWS now requires your AWS account to be associated with a NUS email address in order to receive AWS credits
- Check that you used your NUS email address to sign up for your AWS account
- Enter your NUS email address linked to your AWS account into the Canvas Ungraded Survey (under Quizzes) “Information for Labs”
- If you did not use your NUS email address when you created your AWS account, you will need to sign up for another AWS account linked to your NUS email address
- Enter the new AWS account number and email address linked to your AWS account into the Canvas Ungraded Survey (under Quizzes) “Information for Labs”
- Submit by Monday (Oct 9), 2 pm

# **EE3801 Data Engineering Principles**

## **Data Acquisition and Processing**

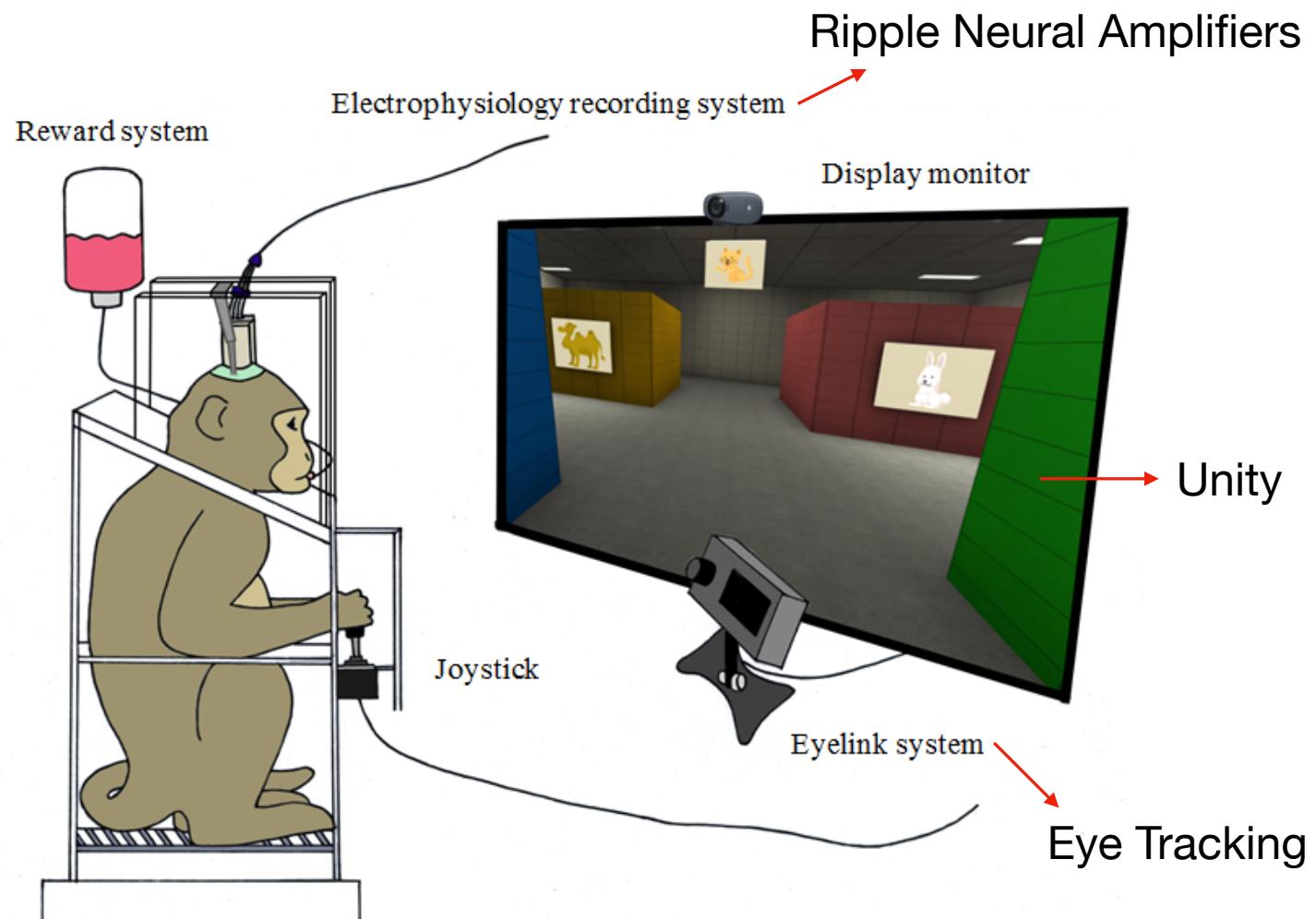
# Data Acquisition & Processing

## Experimental Data

- Experiment consisted of two types of sessions
  - Navigation
  - Data include:
    - Position (Unity)
    - Eye tracking (Eyelink)
    - Neural (Ripple)
  - Fixation (eye movements)

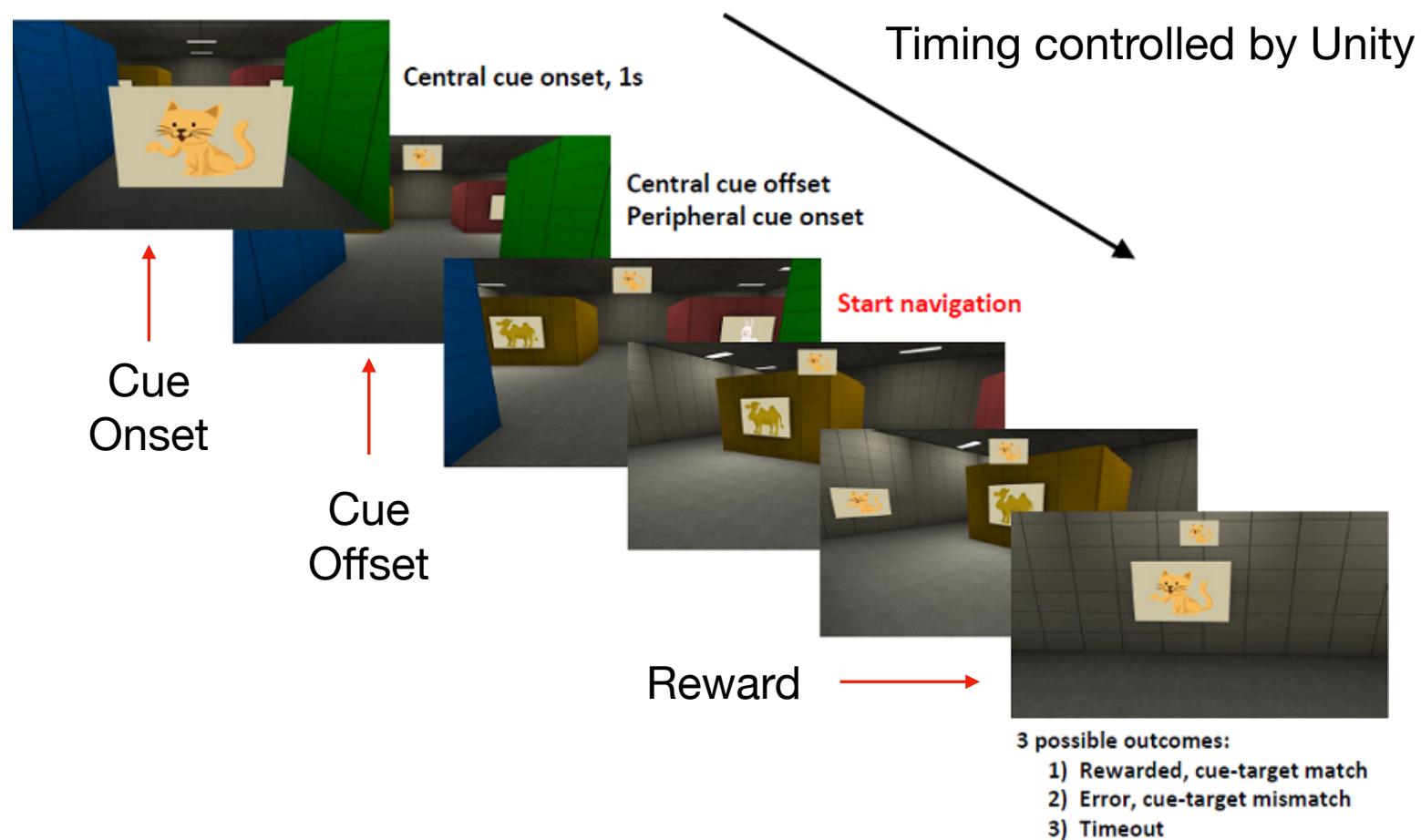
# Data Acquisition & Processing

## Experimental Setup



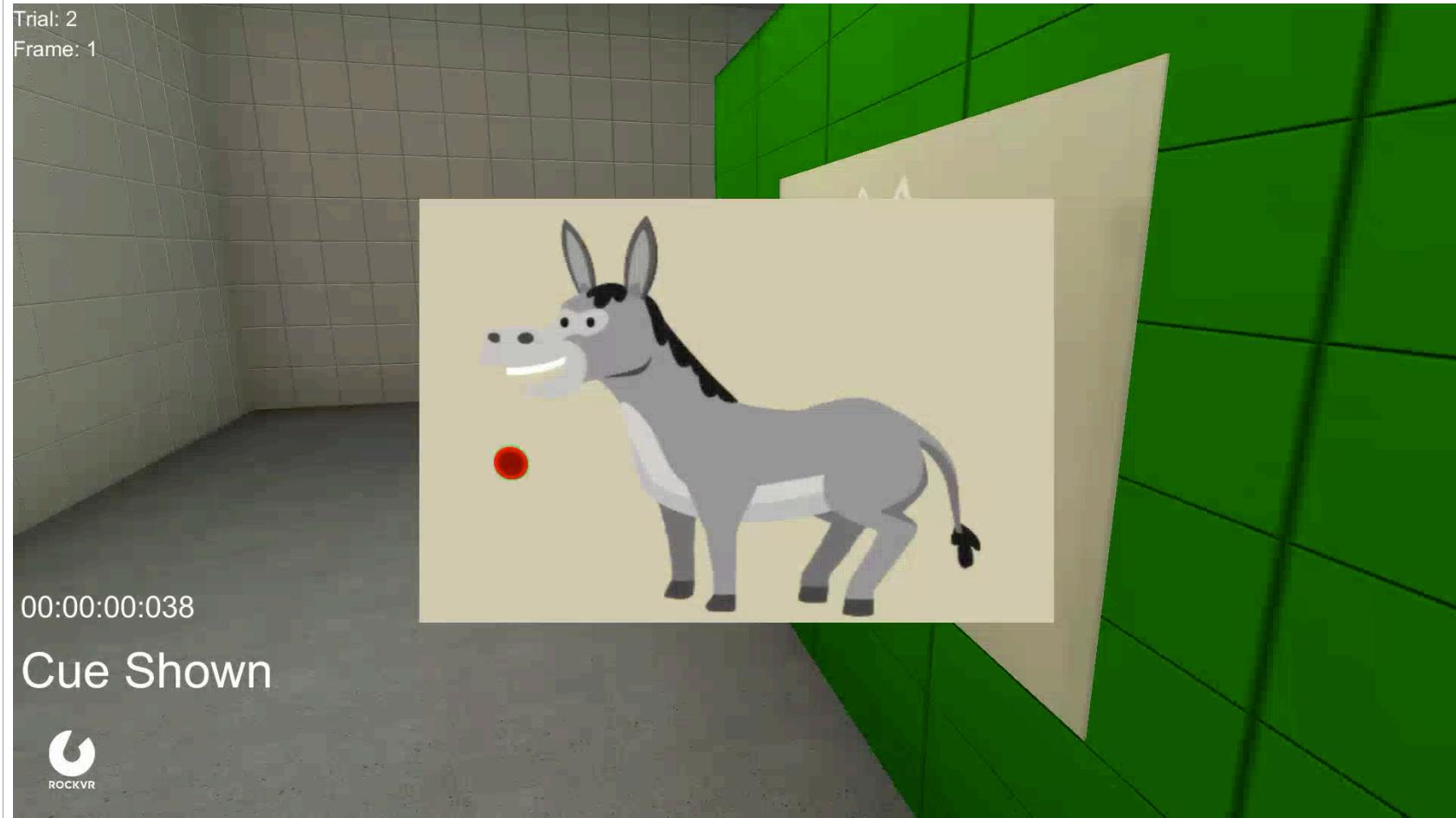
# Data Acquisition & Processing

## Experimental Task for Navigation Sessions



# Data Acquisition & Processing

## Experimental Task for Navigation Sessions



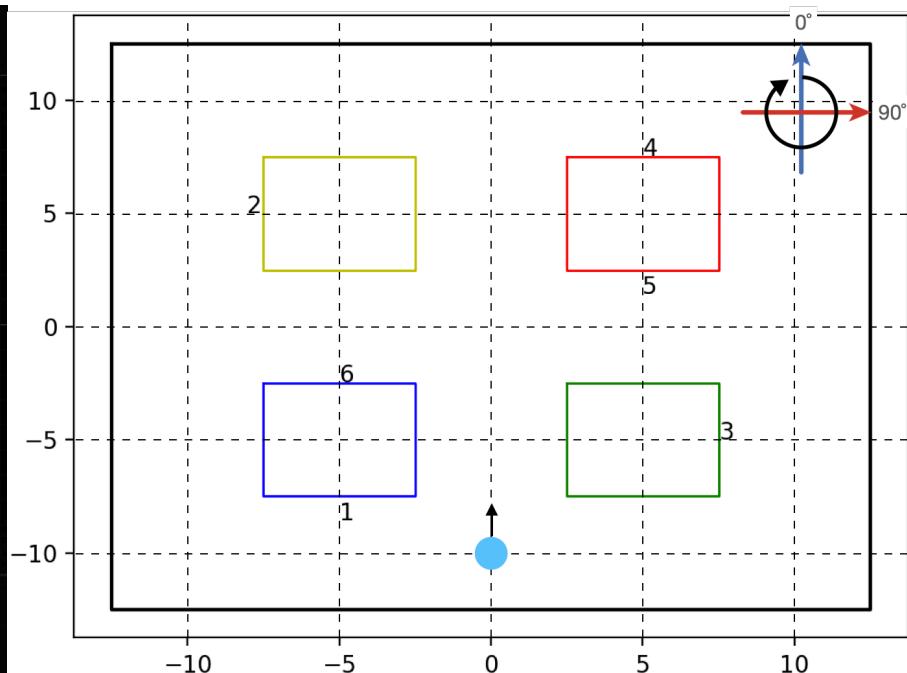
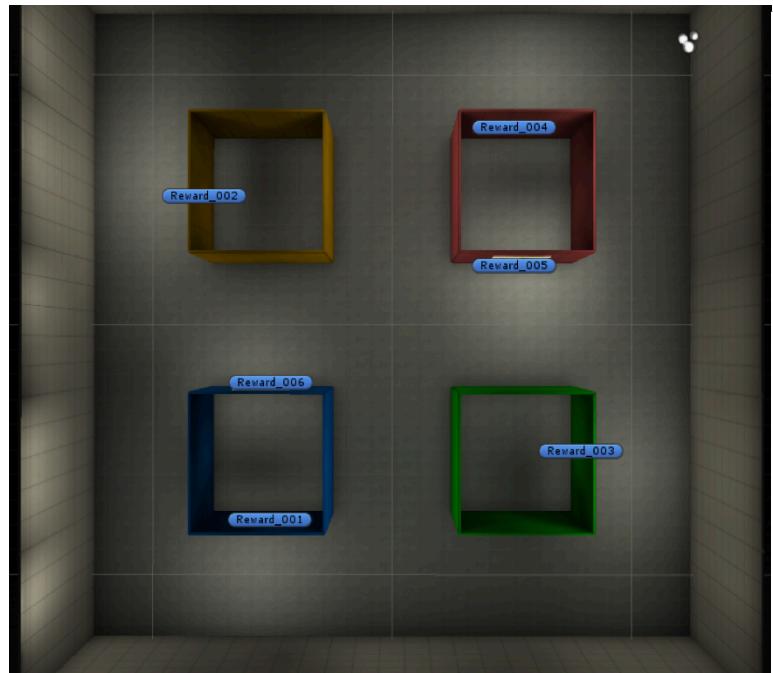
# Data Acquisition & Processing

## Experimental Data

- Experiment consisted of two types of sessions
  - Navigation
  - Data include:
    - Position (Unity)
    - Eye tracking (Eyelink)
    - Neural (Ripple)
  - Fixation (eye movements)

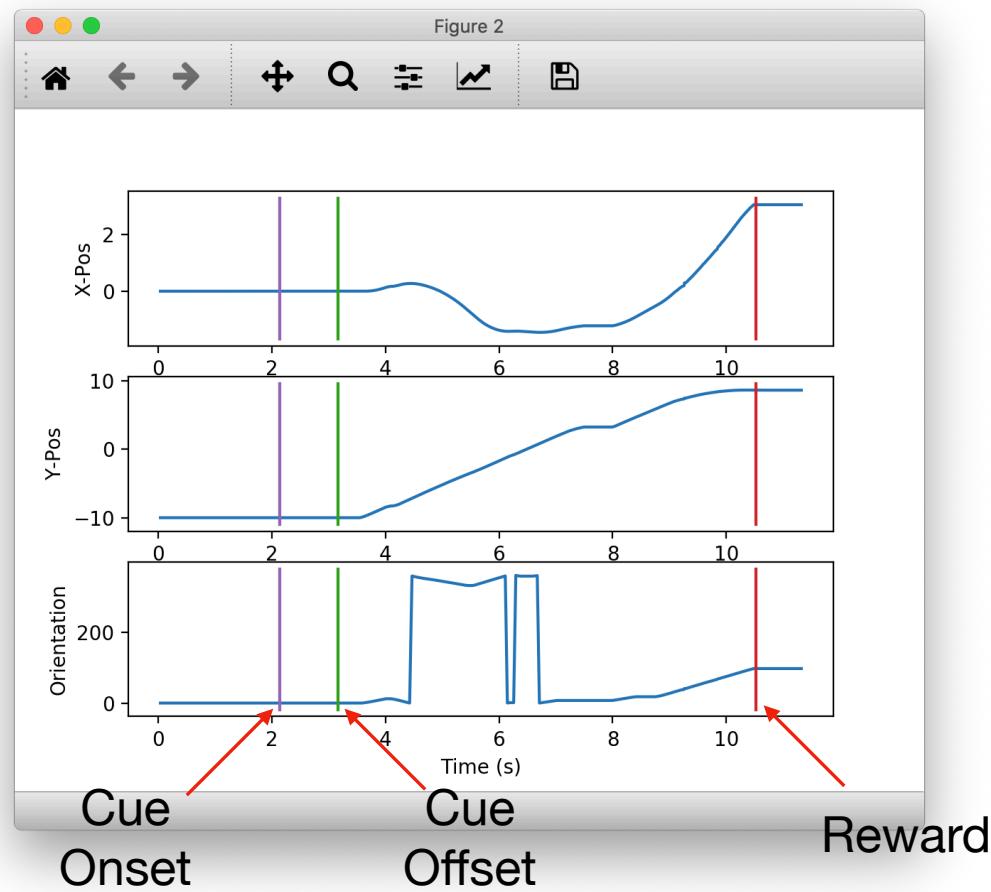
# Data Acquisition & Processing

## Position Data (Unity)



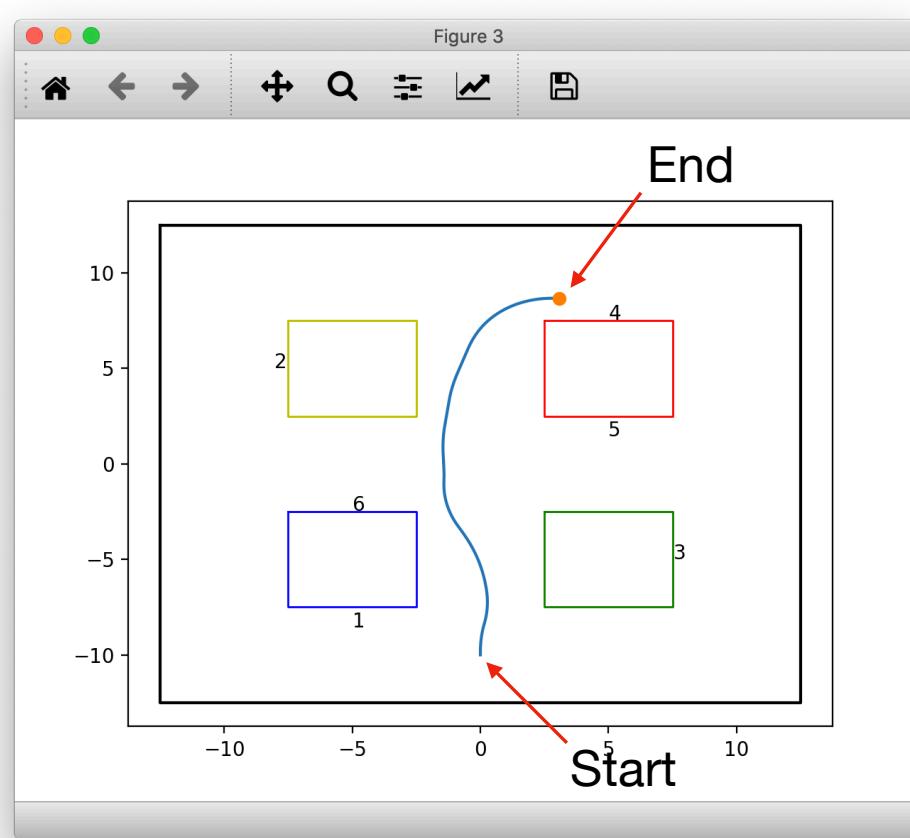
# Data Acquisition & Processing

## Position Data (Unity)



# Data Acquisition & Processing

## Position Data (Unity)



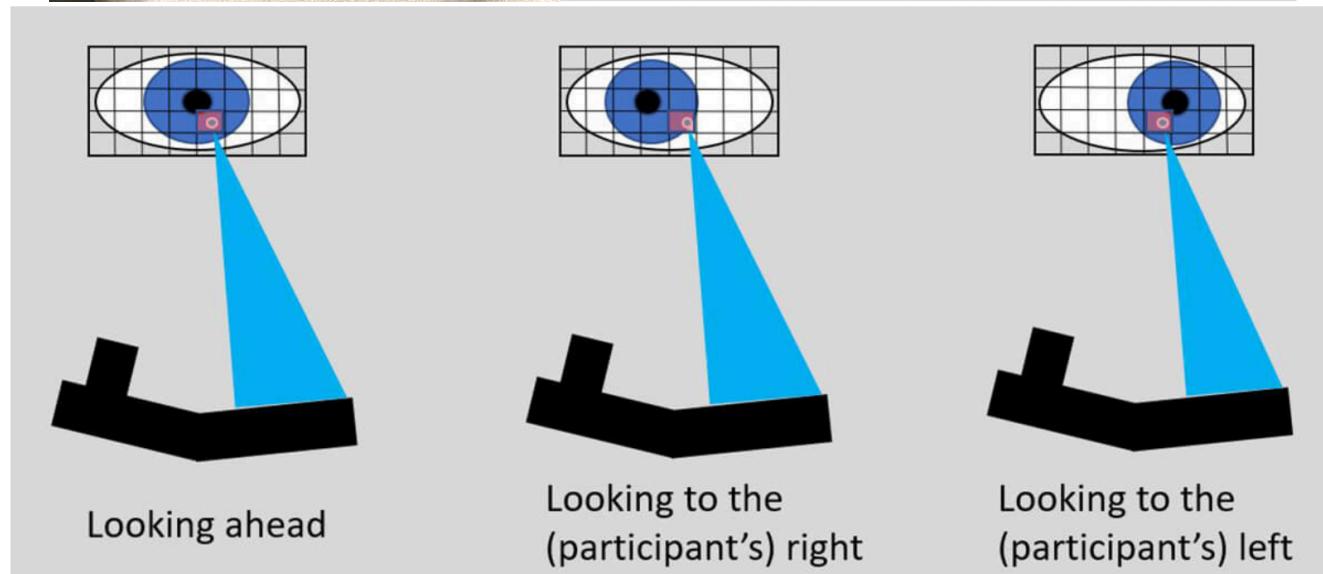
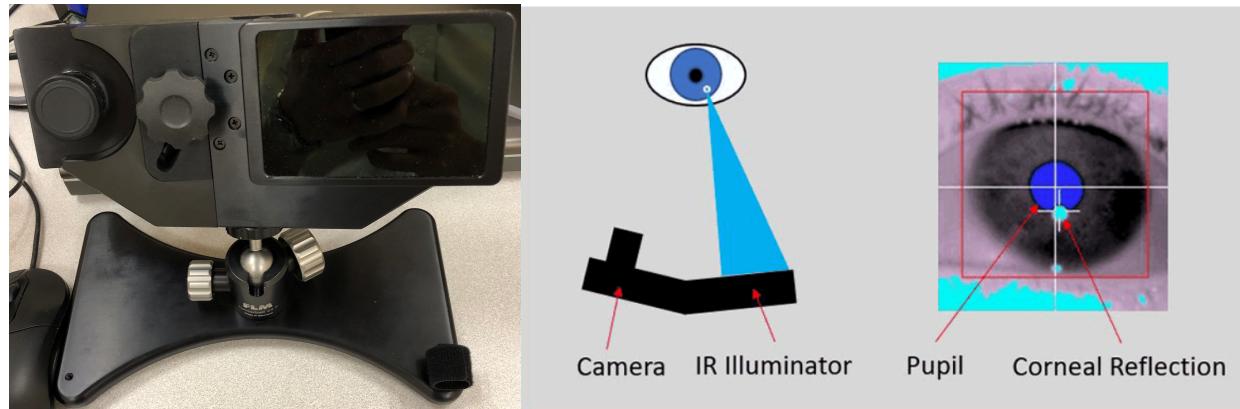
# Data Acquisition & Processing

## Experimental Data

- Experiment consisted of two types of sessions
  - Navigation
  - Data include:
    - Position (Unity)
    - Eye tracking (Eyelink)
    - Neural (Ripple)
  - Fixation (eye movements)

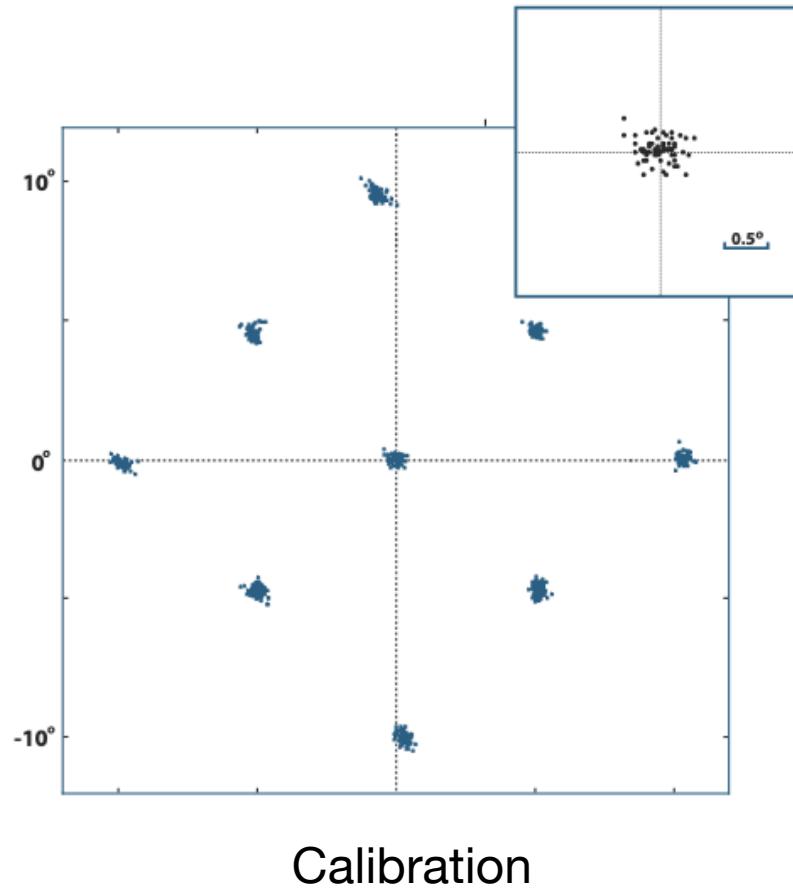
# Data Acquisition & Processing

## Eye-tracking Data (Eyelink)



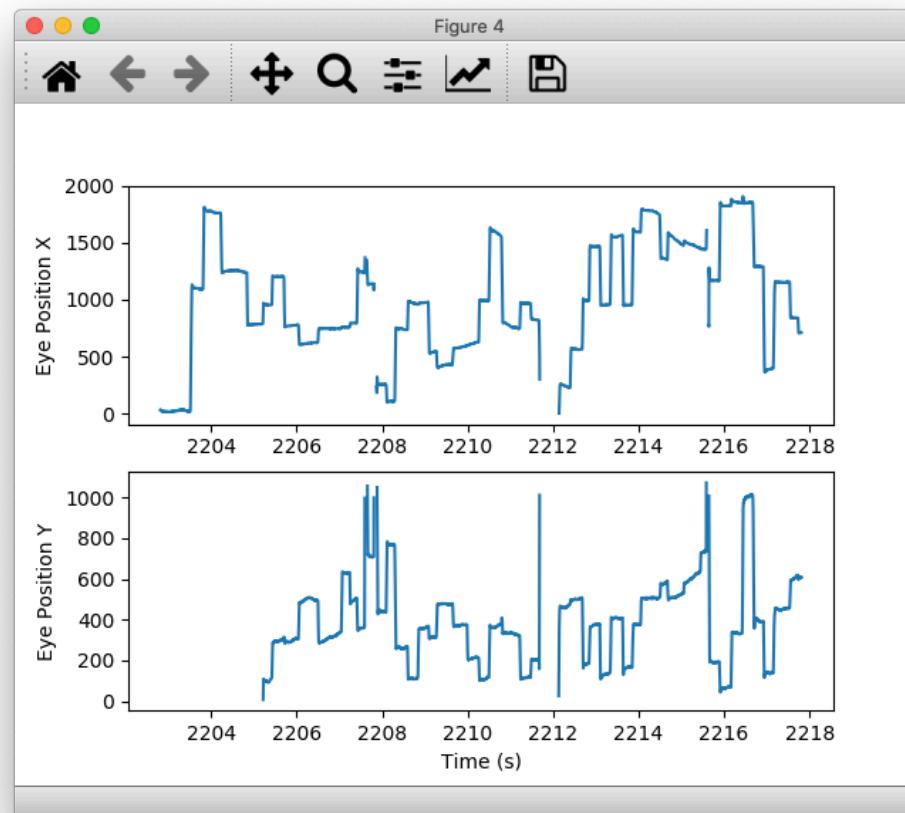
# Data Acquisition & Processing

## Eye-tracking Data (Eyelink)



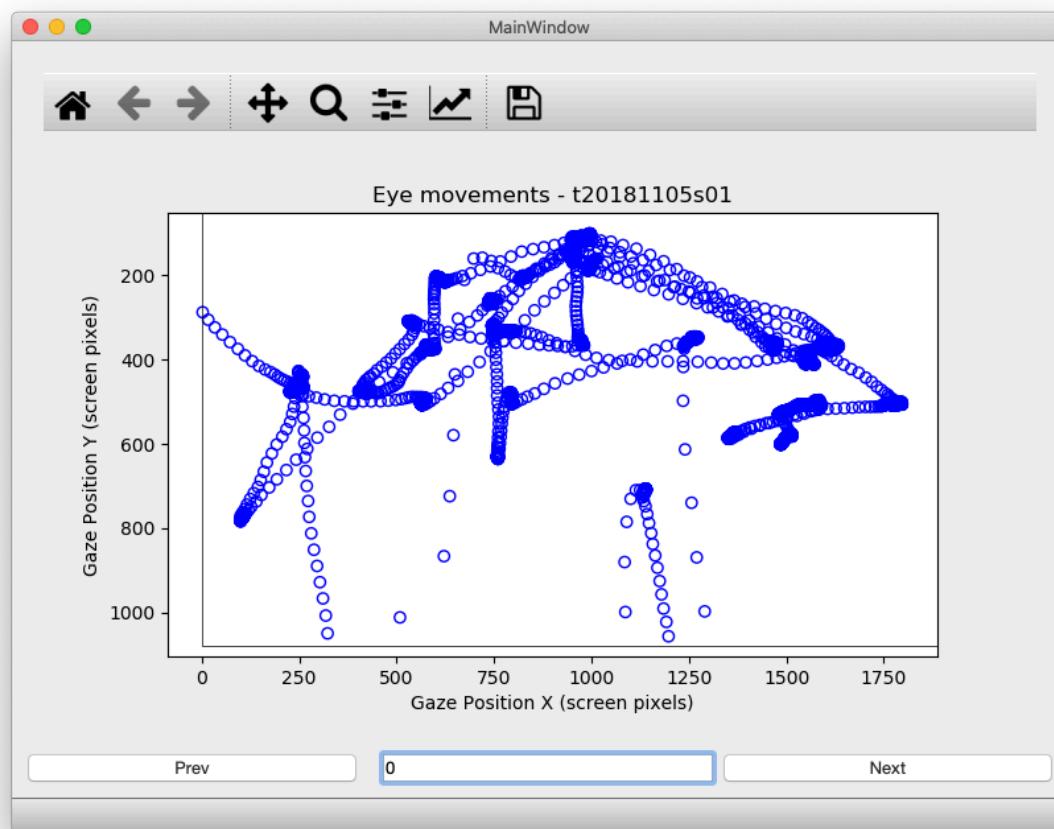
# Data Acquisition & Processing

## Eye-tracking Data (Eyelink)



# Data Acquisition & Processing

## Eye-tracking Data (Eyelink)



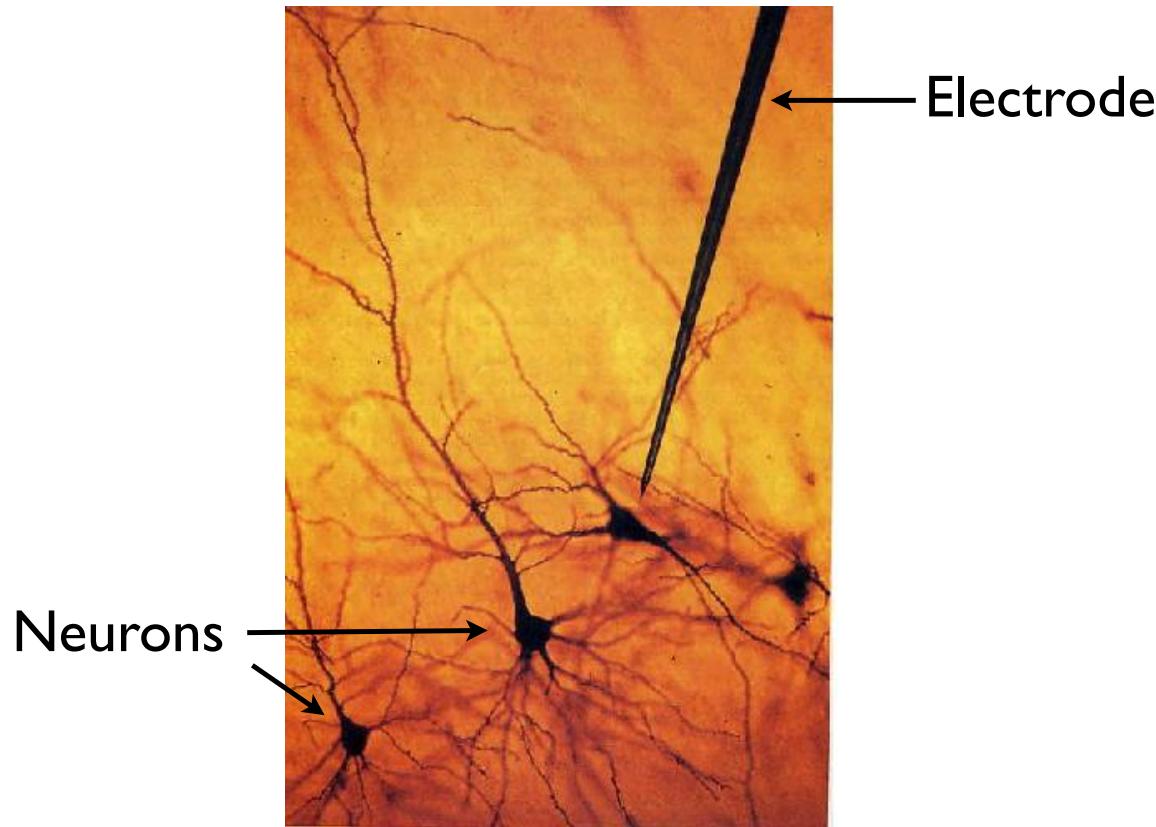
# Data Acquisition & Processing

## Experimental Data

- Experiment consisted of two types of sessions
  - Navigation
  - Data include:
    - Position (Unity)
    - Eye tracking (Eyelink)
    - Neural (Ripple)
  - Fixation (eye movements)

# Data Acquisition & Processing

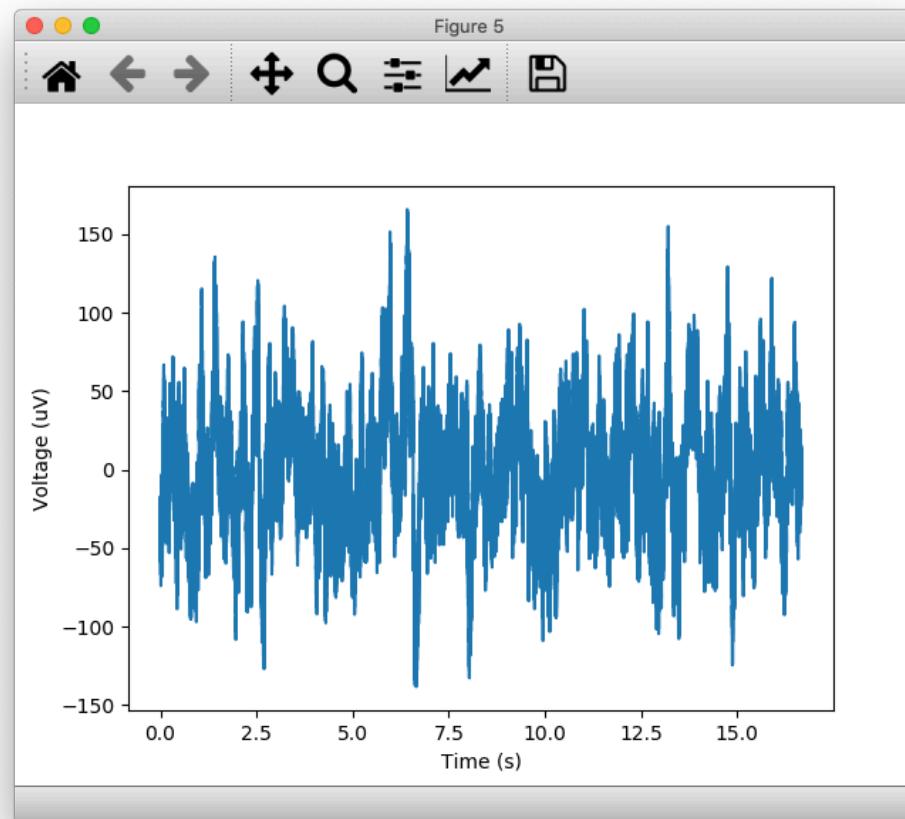
## Neural Data (Ripple)



Capable of recording from multiple neurons

# Data Acquisition & Processing

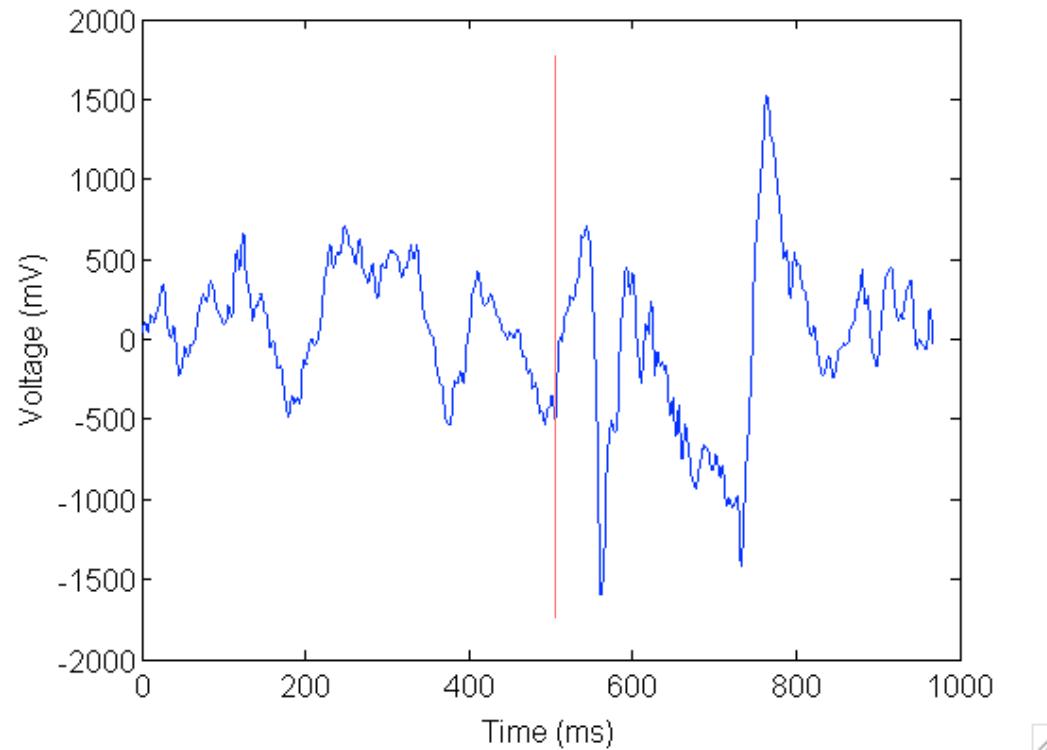
## Neural Data (Ripple)



Raw data

# Data Acquisition & Processing

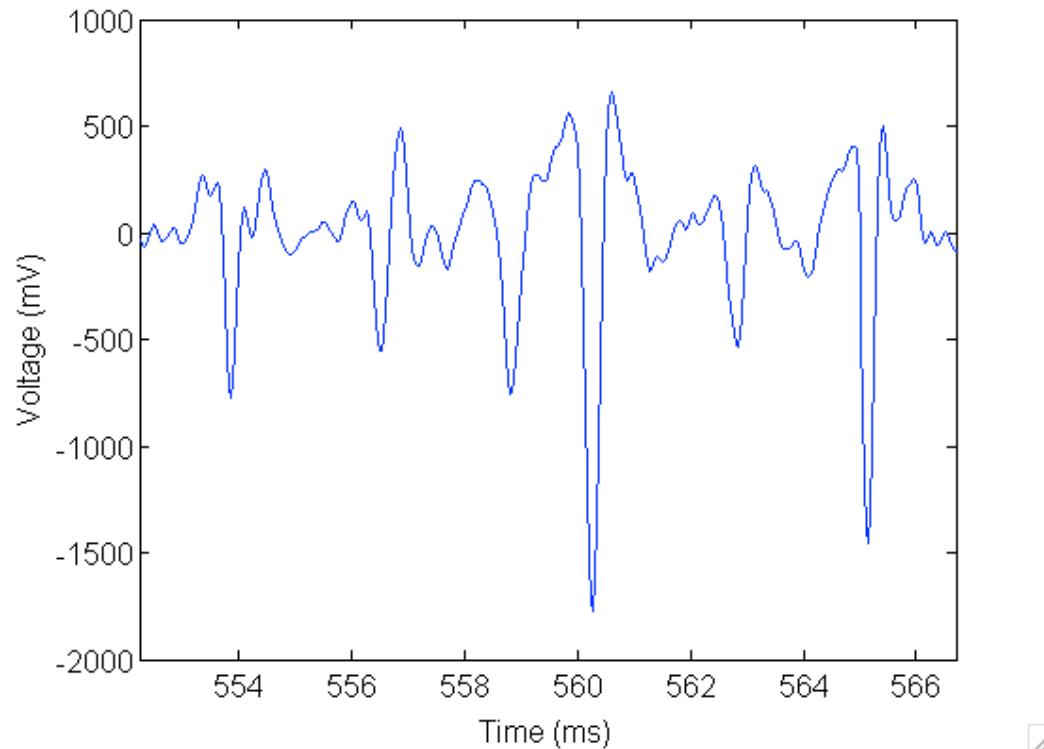
## Neural Data (Ripple)



Low-pass filtered from 1 – 150 Hz  
Local Field Potential (LFP) - population synaptic activity

# Data Acquisition & Processing

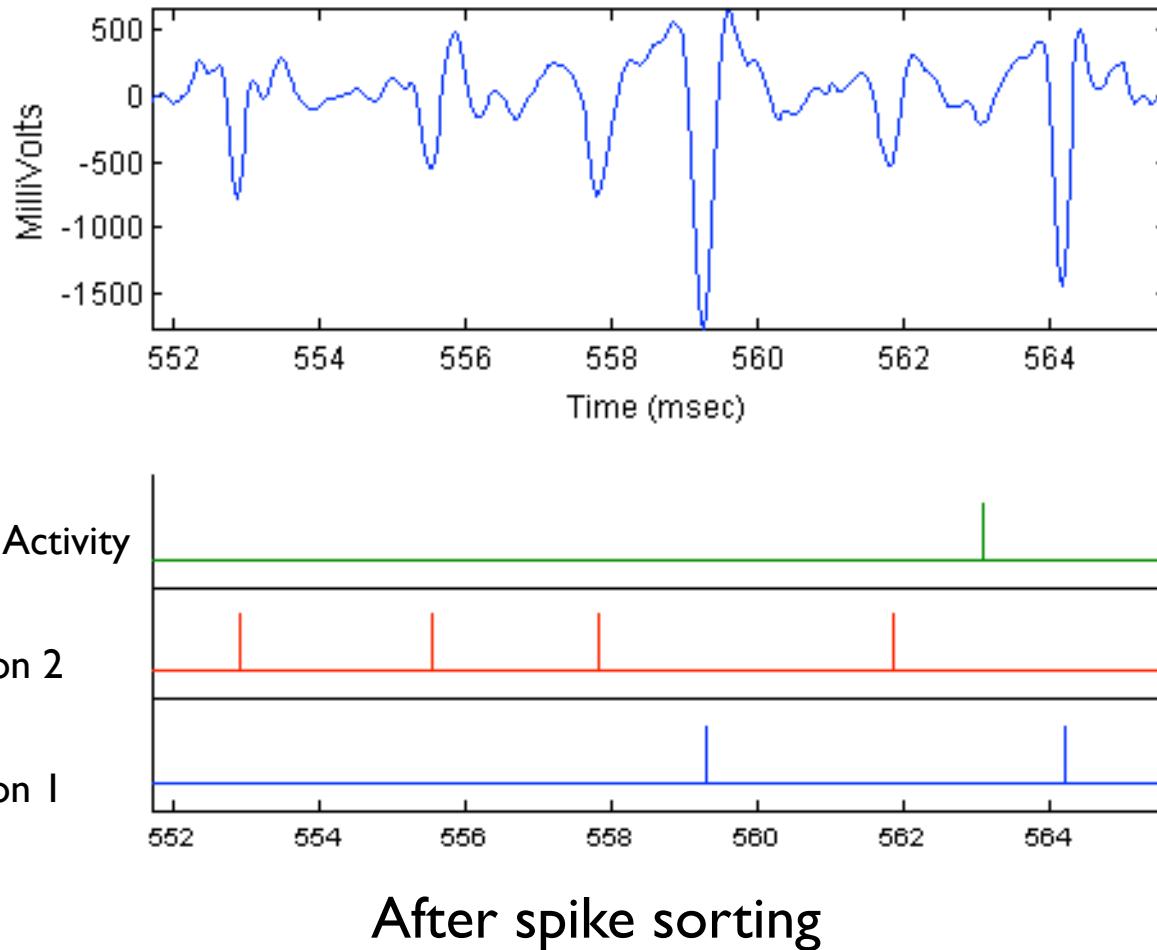
## Neural Data (Ripple)



High-pass filtered from 500 – 7,500 Hz  
Action potentials (spikes) from multiple neurons present

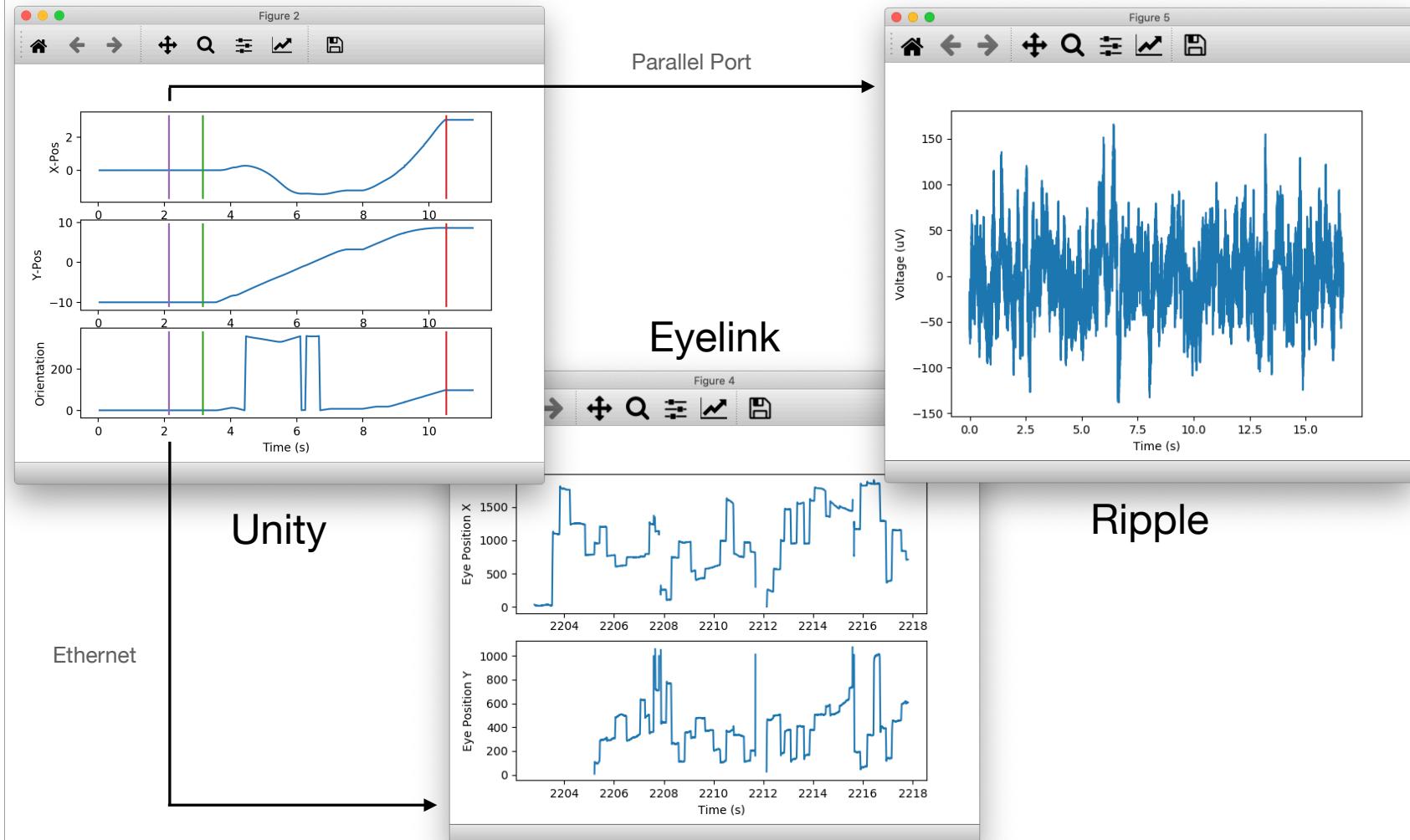
# Data Acquisition & Processing

## Neural Data (Ripple)



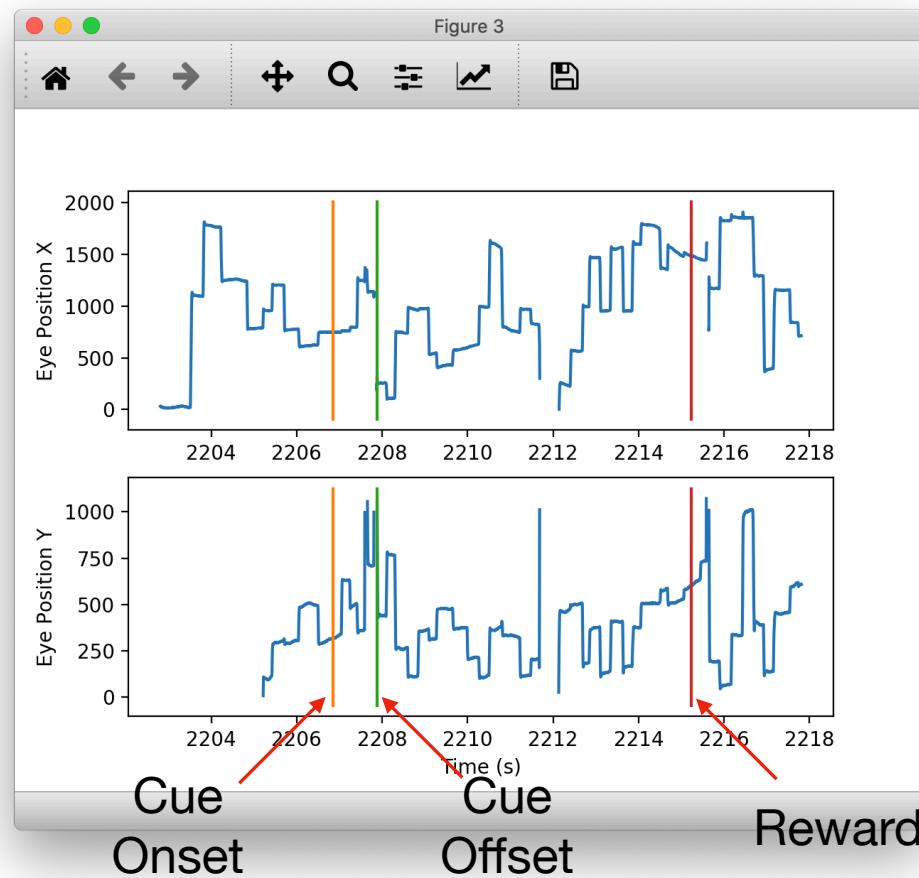
# Data Acquisition & Processing

## Data Timelines



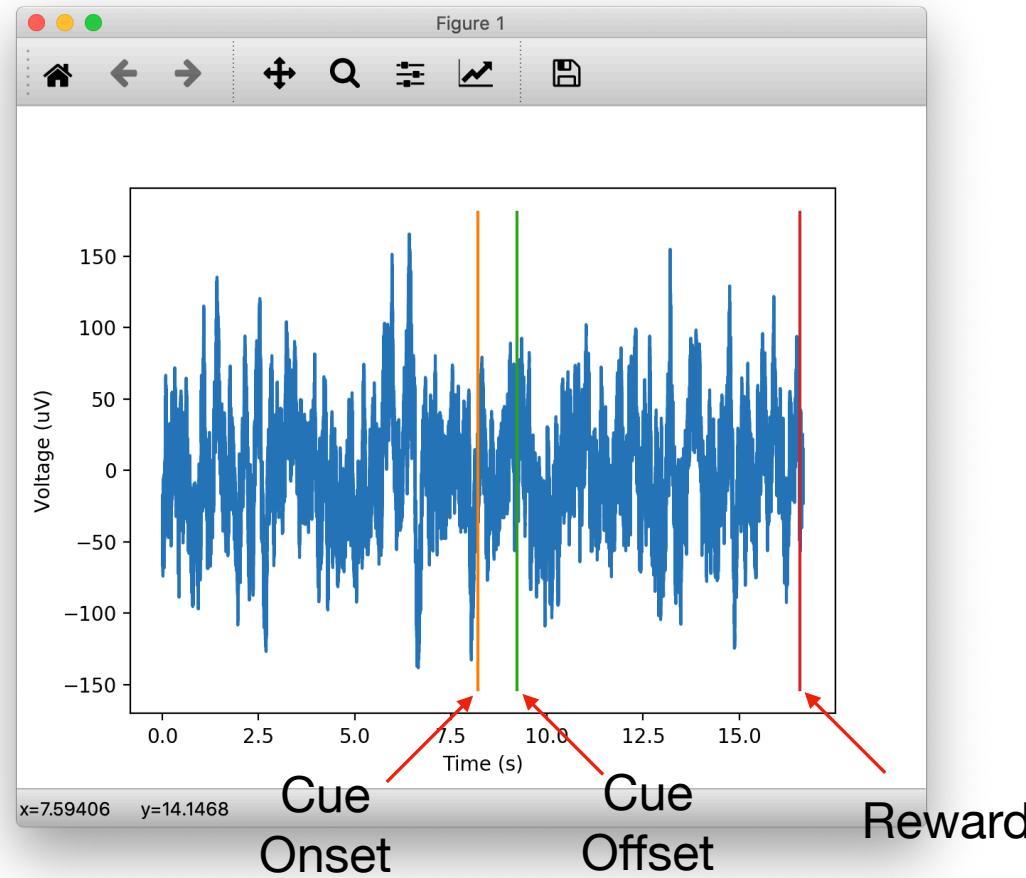
# Data Acquisition & Processing

## Eye-tracking Data (Eyelink)



# Data Acquisition & Processing

## Neural Data (Ripple)



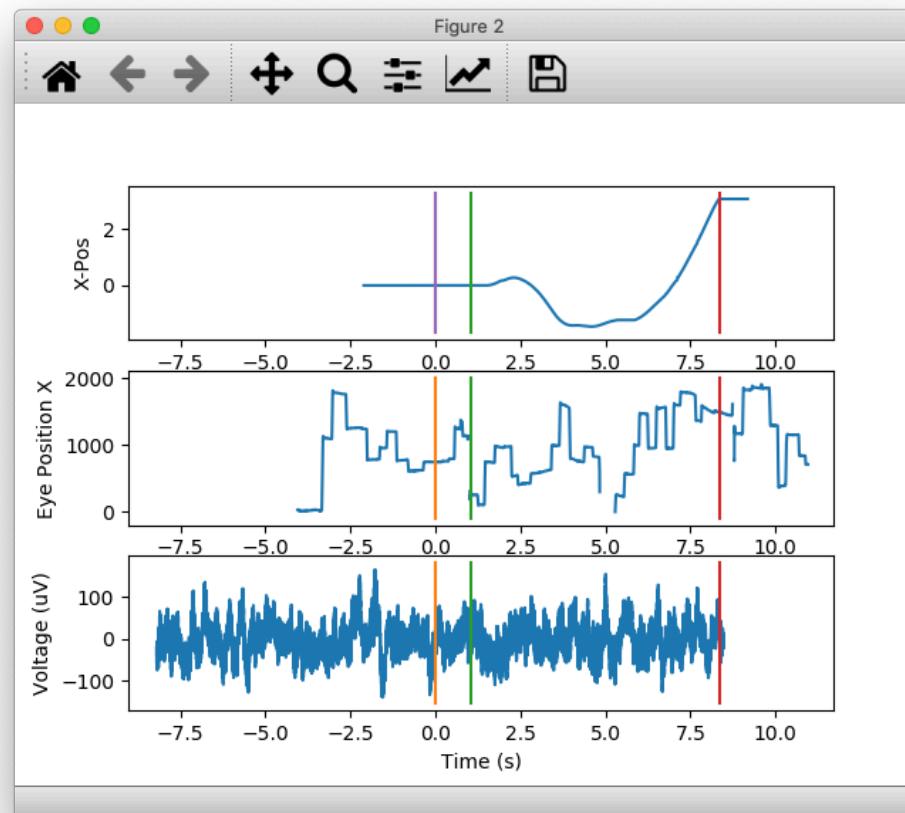
# Data Acquisition & Processing

## Aligning Data

Unity

Eyelink

Ripple



# Data Acquisition & Processing

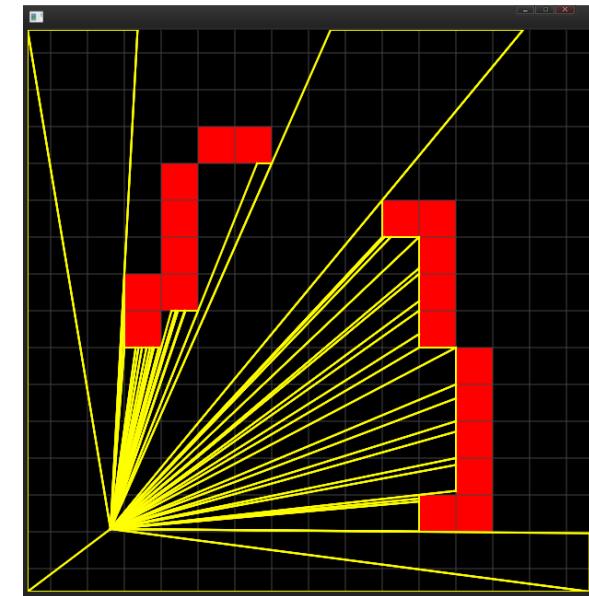
## Aligning Data



# Data Acquisition & Processing

## Raycasting

Trace rays from the camera for each eye position to find the closest object blocking the path of that ray. This helps us identify the object that the animal is looking at.



# Data Acquisition & Processing

## Raycasting

Type	Time	Object	eye_x	eye_y	pos_x	pos_y	pos_z	pos_o	fix_x	fix_y	fix_z	obj_x	
A	B	C	D	E	F	G	H	I	J	K	L	M	
1	SAMPLE_TYPE	1689040	Ground	1.725972	-0.25171	0	0	0	1.725972	1.12E-18	-0.25171	1125.6	362.5
2	SAMPLE_TYPE	1689041	Ground	1.718677	-0.25171	0	0	0	1.718677	1.12E-18	-0.25171	1124.9	362.5
3	SAMPLE_TYPE	1689042	Ground	1.723815	-0.24622	0	0	0	1.723815	1.09E-18	-0.24622	1125.3	362.6
4	SAMPLE_TYPE	1689043	Ground	1.734831	-0.20758	0	0	0	1.734831	9.22E-19	-0.20758	1125.7	363.3
5	SAMPLE_TYPE	1689044	Ground	1.740743	-0.17421	0	0	0	1.740743	7.74E-19	-0.17421	1125.7	363.9
6	SAMPLE_TYPE	1689045	Ground	1.740495	-0.15184	0	0	0	1.740495	6.74E-19	-0.15184	1125.3	364.3
7	SAMPLE_TYPE	1689046	Ground	1.734246	-0.15744	0	0	0	1.734246	6.99E-19	-0.15744	1124.8	364.2
8	SAMPLE_TYPE	1689047	Ground	1.727297	-0.19092	0	0	0	1.727297	8.48E-19	-0.19092	1124.7	363.6
9	SAMPLE_TYPE	1689048	Ground	1.72353	-0.22418	0	0	0	1.72353	9.96E-19	-0.22418	1124.9	363
10	SAMPLE_TYPE	1689049	Ground	1.720907	-0.26269	0	0	0	1.720907	1.17E-18	-0.26269	1125.3	362.3
11	SAMPLE_TYPE	1689050	Ground	1.727359	-0.27909	0	0	0	1.727359	1.24E-18	-0.27909	1126.2	362
12	SAMPLE_TYPE	1689051	Ground	1.741421	-0.23521	0	0	0	1.741421	1.04E-18	-0.23521	1126.8	362.8
13	SAMPLE_TYPE	1689052	Ground	1.750311	-0.18536	0	0	0	1.750311	8.23E-19	-0.18536	1126.8	363.7
14	SAMPLE_TYPE	1689053	Ground	1.753022	-0.14062	0	0	0	1.753022	6.24E-19	-0.14062	1126.3	364.5
15	SAMPLE_TYPE	1689054	Ground	1.739819	-0.18536	0	0	0	1.739819	8.23E-19	-0.18536	1125.8	363.7
16	SAMPLE_TYPE	1689055	Ground	1.737117	-0.22418	0	0	0	1.737117	9.96E-19	-0.22418	1126.2	363
17	SAMPLE_TYPE	1689056	Ground	1.731517	-0.27909	0	0	0	1.731517	1.24E-18	-0.27909	1126.6	362
18	SAMPLE_TYPE	1689057	Ground	1.732752	-0.29546	0	0	0	1.732752	1.31E-18	-0.29546	1127	361.7
19	SAMPLE_TYPE	1689058	Ground	1.744612	-0.24072	0	0	0	1.744612	1.07E-18	-0.24072	1127.2	362.7
20	SAMPLE_TYPE	1689059	Ground	1.761654	-0.16304	0	0	0	1.761654	7.24E-19	-0.16304	1127.5	364.1
21	SAMPLE_TYPE	1689060	Ground	1.778969	-0.08412	0	0	0	1.778969	3.74E-19	-0.08412	1127.8	365.5
22	SAMPLE_TYPE	1689061	Ground	1.778154	-0.11811	0	0	0	1.778154	5.24E-19	-0.11811	1128.3	364.9
23	SAMPLE_TYPE	1689062	Ground	1.774189	-0.15184	0	0	0	1.774189	6.74E-19	-0.15184	1128.5	364.3
24	SAMPLE_TYPE	1689063	Ground	1.765052	-0.19092	0	0	0	1.765052	8.48E-19	-0.19092	1128.3	363.6
25	SAMPLE_TYPE	1689064	Ground	1.759908	-0.20203	0	0	0	1.759908	8.97E-19	-0.20203	1128	363.4
26	SAMPLE_TYPE	1689065	Ground	1.75478	-0.21312	0	0	0	1.75478	9.46E-19	-0.21312	1127.7	363.2
27	SAMPLE_TYPE	1689066	Ground	1.763053	-0.20203	0	0	0	1.763053	8.97E-19	-0.20203	1128.3	363.4
28	SAMPLE_TYPE	1689067	Ground	1.760106	-0.16003	0	0	0	1.760106	8.48E-19	-0.16003	1128.6	363.6

### Description of Columns

- Type of data in the row (string/text).
- Timestamp of the gaze data used to identify the fixated object (unsigned int)
- Name of the object fixated by the gaze, or message received by EyeLink (string/text).

### Raw Gaze Data (Pixels)

- gx data from the .edf file used to raycast.
- gy data from the .edf file used to raycast.

### Subject Location in Worldspace (Unity Units)

- X coordinate of the subject's location in Worldspace.
- Y coordinate of the subject's location in Worldspace.
- Z coordinate of the subject's location in Worldspace.
- Orientation of the subject about Unity's Y axis (degrees).

### Gaze Location (Unity Units)

- X Worldspace coordinate where the gaze lands.
- Y Worldspace coordinate where the gaze lands.
- Z Worldspace coordinate where the gaze lands.

### Gazed Object Location (Unity Units)

- X Worldspace coordinate of the center of the fixated object.
- Y Worldspace coordinate of the center of the fixated object.
- Z Worldspace coordinate of the center of the fixated object.

### Gaze Position in Object (Unity Units)

- 2D X location of gaze with reference to the center of the fixated object.
- 2D Y location of gaze with reference to the center of the fixated object.

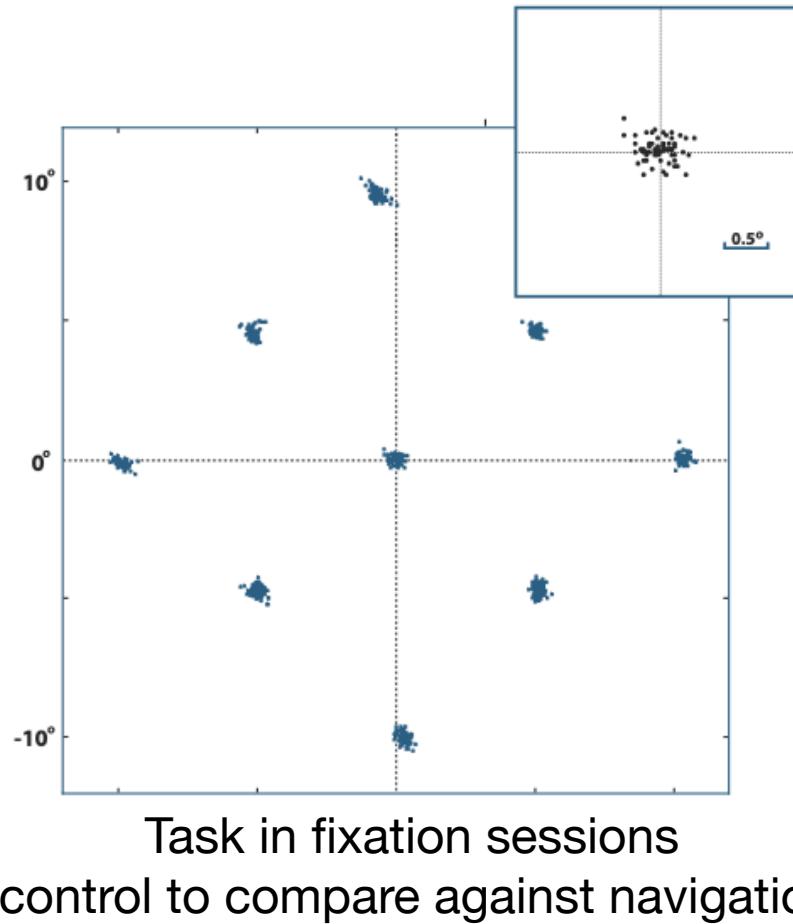
# Data Acquisition & Processing

## Experimental Data

- Experiment consisted of two types of sessions
  - Navigation
  - Data include:
    - Position (Unity)
    - Eye tracking (Eyelink)
    - Neural (Ripple)
  - Fixation (eye movements)

# Data Acquisition & Processing

## Experimental Task for Fixation Sessions



# **Data Acquisition & Processing**

## **Lab Instructions**

- Lab 4 Instructions:
  - <https://ee3801.github.io/Lab4/>
- Submit to Canvas Assignments (Lab 4->Lab 4A & Lab 4->Lab 4B)
- Submit in PDF format
- Name the files Lab4A\_YourName.pdf and Lab4B\_YourName.pdf
- Part A due on Monday (Oct 9) 2 pm
- Part B due on Wednesday (Oct 11) 9 pm

# Questions?