Depth and Layout Encoded Simulated Prosthetic Vision for Obstacle Avoidance and Object Recognition

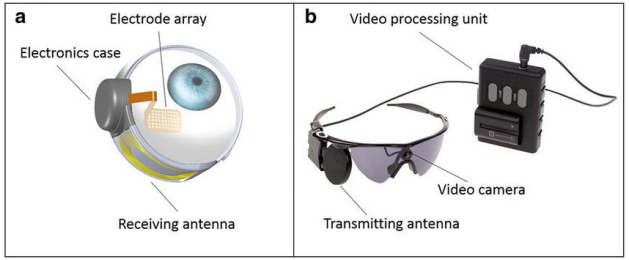
Alex Rasla

MS Project Defense

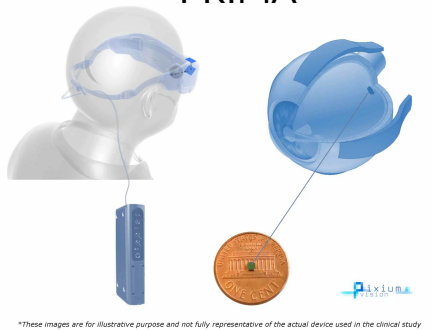
Background: Retinal Prostheses

- Retinal prostheses have the potential to treat degenerative diseases such as retinitis pigmentosa and age-related macular degeneration
- Current devices generally consist of:
  - External camera
  - Vision processing unit (VPU)
  - Electrode array
- Electrical stimulation leads people to perceiving flashes of light (“phosphenes”)

Argus II



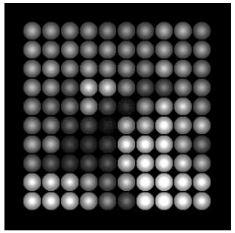
PRIMA



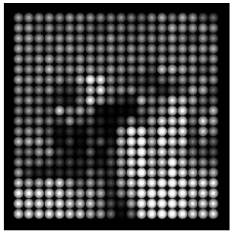
*These images are for illustrative purpose and not fully representative of the actual device used in the clinical study

IRIS II

Background: Simulated Prosthetic Vision Models



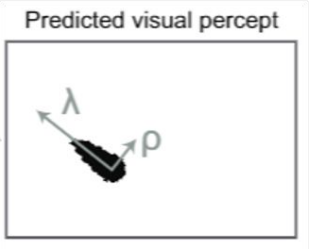
10×10

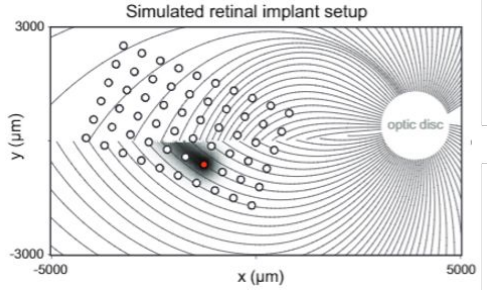


20×20

100×100

- Linear Scoreboard Model
  - assumes a linear transformation from stimulus to percept
  - cannot account for percept shape reported by actual patients: elongated, different for each electrode, but consistent and repeatable across trials
- Axon Map Model
  - takes into account inadvertent activation of axon fiber bundles in the retina
  - λ describes current spread along the axon fibers
  - ρ describes current spread perpendicular to axon fibers



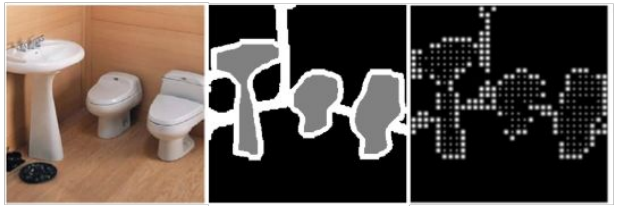


Beyeler et al. (2019)

Scene Simplification Using Computer Vision

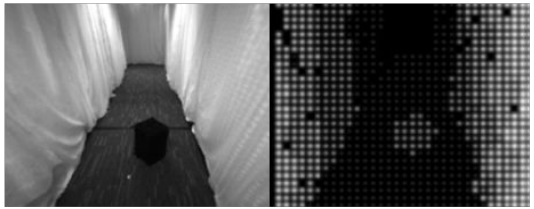
- Many of these devices do not have enough electrodes to faithfully represent the visual scene
  - There is a need to simplify the scene
- Previous works have used computer vision for scene simplification:
  - Semantic information: extracting edges, segmenting objects
  - Depth information: highlighting nearby obstacles

Semantic & structural segmentation



Sanchez-Garcia et al., 2020

Substituting depth for intensity



McCarthy et al., 2011, 2014

Common limitations:

- No previous works use a realistic phosphene model
- No previous works used a *combination* of scene simplification methods

Methods: Scene Simplification

- In this study, we explored the effectiveness of combining depth and object (i.e., object outlines, room layout) cues for obstacle avoidance and object recognition in a virtual environment
  - Depth cues might be more important for obstacle avoidance
  - Structural cues might be more important for object recognition
- 4 simulated prosthetic vision (SPV) modes viewed through a simulated 20x15 electrode array (ρ = 300, λ = 550):
  - *DepthOnly*
    - show only the depth of the walls and every object within the room
  - *LayoutOnly*
    - show only the structural layout of the walls and every object within the room
  - *DepthAndLayout*:
    - show both the depth and structural layout of the walls and the objects within the room together
  - *DepthOrLayout*
    - switch between *DepthOnly* and *LayoutOnly*
- Tutorial Room: Participants were given a chance to familiarize themselves with the different SPV modes at the start of the experiment

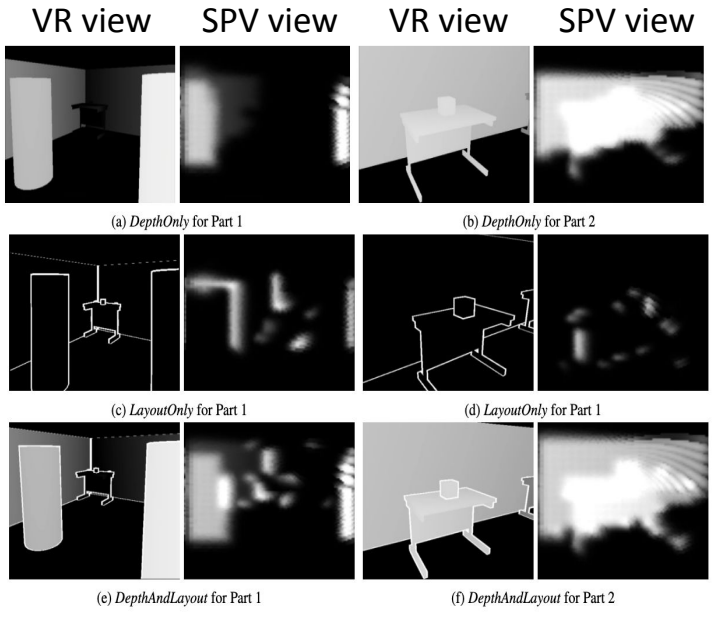
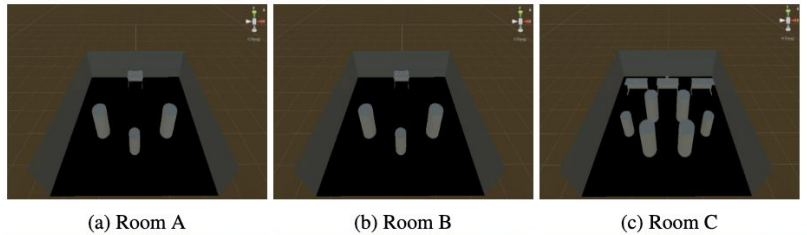


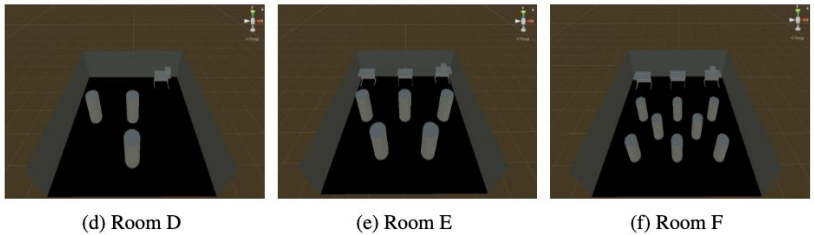
Figure 2: The different SPV modes

Methods: Task

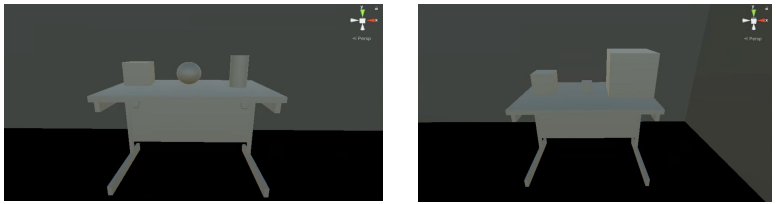
- Part 1: Obstacle avoidance
  - Goal: Walk to the other end of the room while avoiding obstacles
  - Performance metric: Number of collisions
- Part 2: Object Recognition
  - Goal: Locate and select the medium-sized cube on a table
  - Performance metric: Accuracy
  - Always 3 objects to choose from:
    - a sphere, a cylinder, and a medium cube
    - a small cube, a medium cube, and a large cube
- 6 different room layouts of increasing complexity
  - between 3 and 7 obstacles to avoid
  - either 1 or 3 tables in the room
  - Order of rooms was **randomized**



(a) Room A (b) Room B (c) Room C



(d) Room D (e) Room E (f) Room F

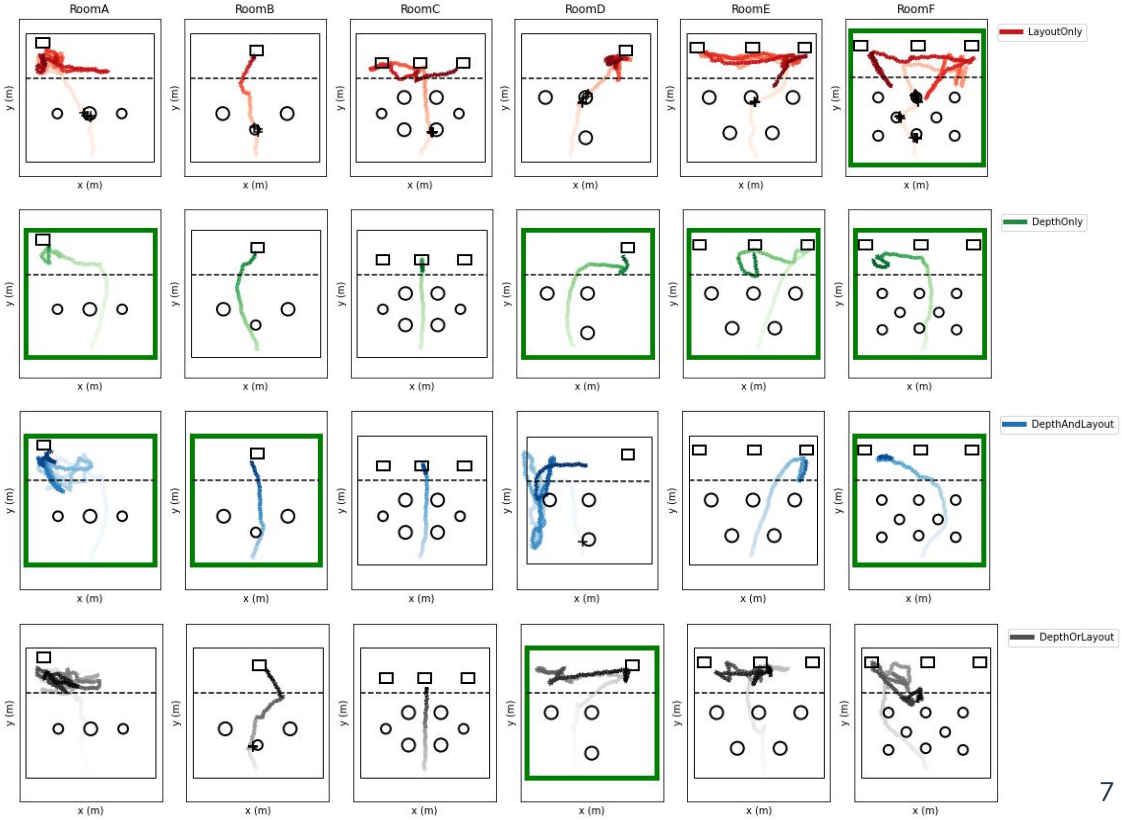


Results: Paths Taken by Participants

- Subject 13
  - Had used VR 1-5 times before
- Collision in **every room** for *LayoutOnly*
- **No collisions** using *DepthOnly*
- Some collisions with *DepthAndLayout* and *DepthOrLayout*
- Preferred *DepthAndLayout* for Part 1, Part 2, and overall

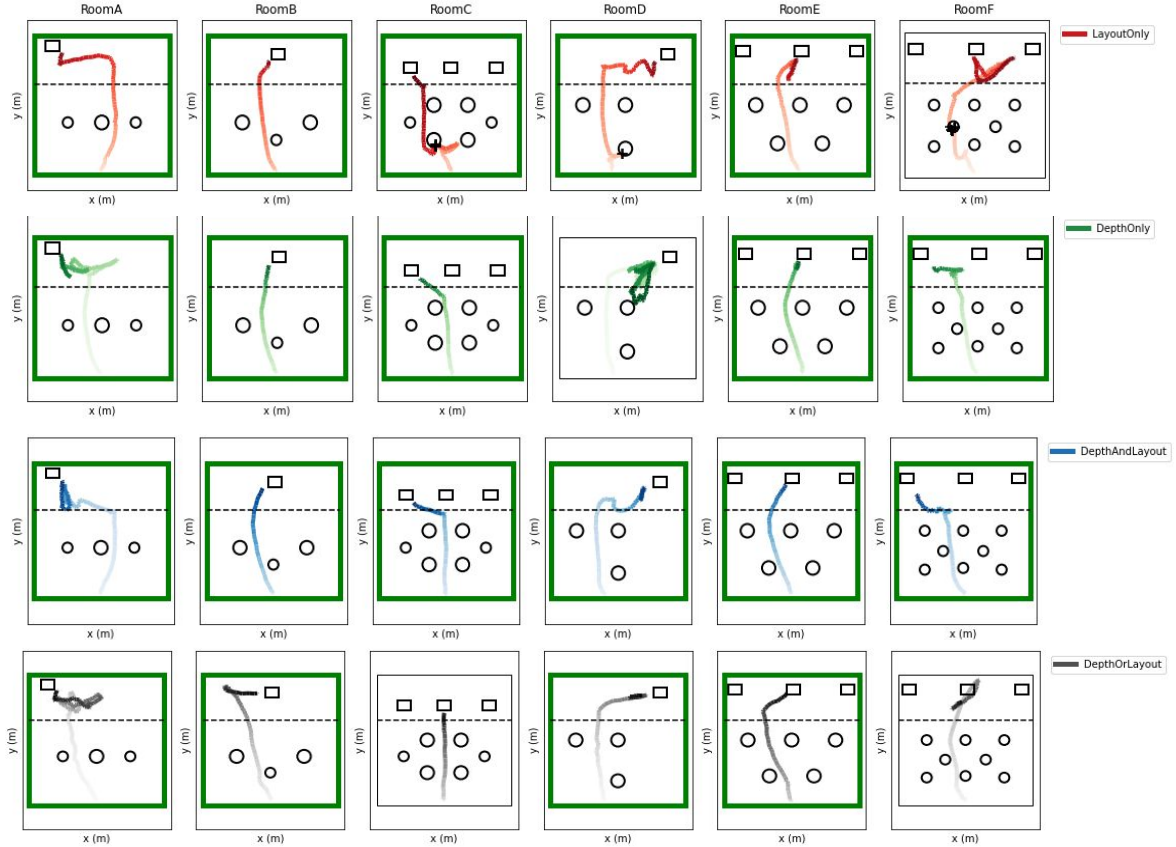
Legend:

- Collisions marked by +
- Green outline means correct object selected
- Path colors get more intense as time in room increases
- Dashed line in room represents start of part 2



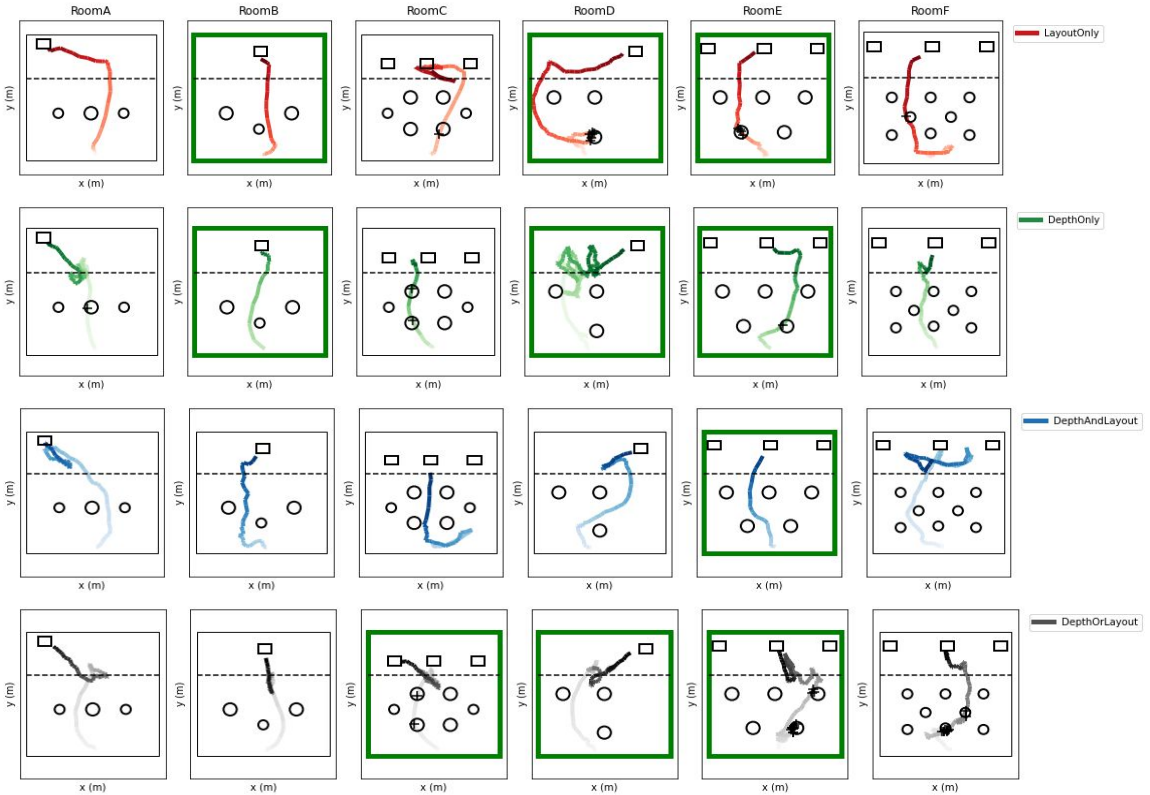
Results: Paths Taken by Participants

- Subject 6
  - No prior VR Experience
- **Some collisions** using *LayoutOnly*
- **No collisions** using *any* Depth encoded mode
- Preferred *DepthOrLayout* for obstacle avoidance, object recognition, and overall



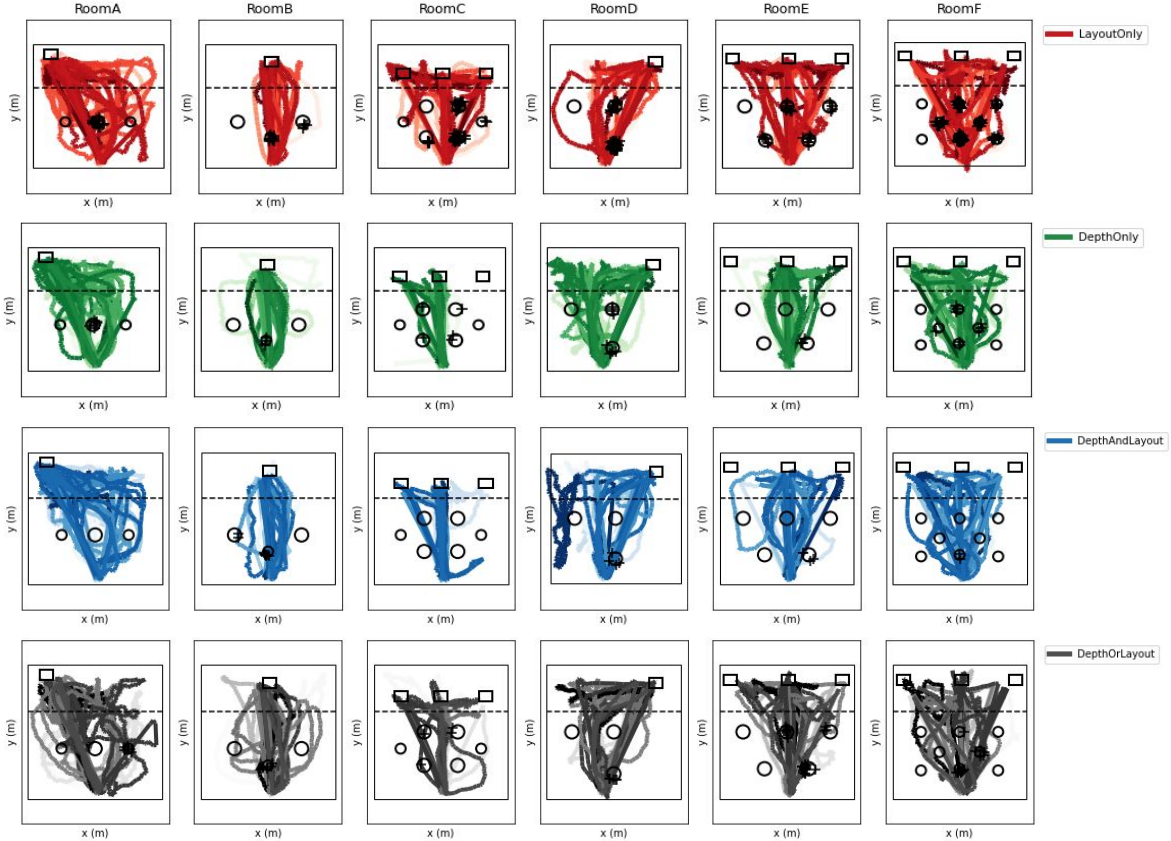
Results: Paths Taken by Participants

- Subject 10
  - 20+ times using VR
- Similar performance in *DepthOnly* and *LayoutOnly*
- **Most collisions** in *DepthOrLayoutOnly*
- Selected correct object at most half the time
- Preferred *DepthOnly* for obstacle avoidance
- Preferred *DepthAndLayout* for object recognition and overall

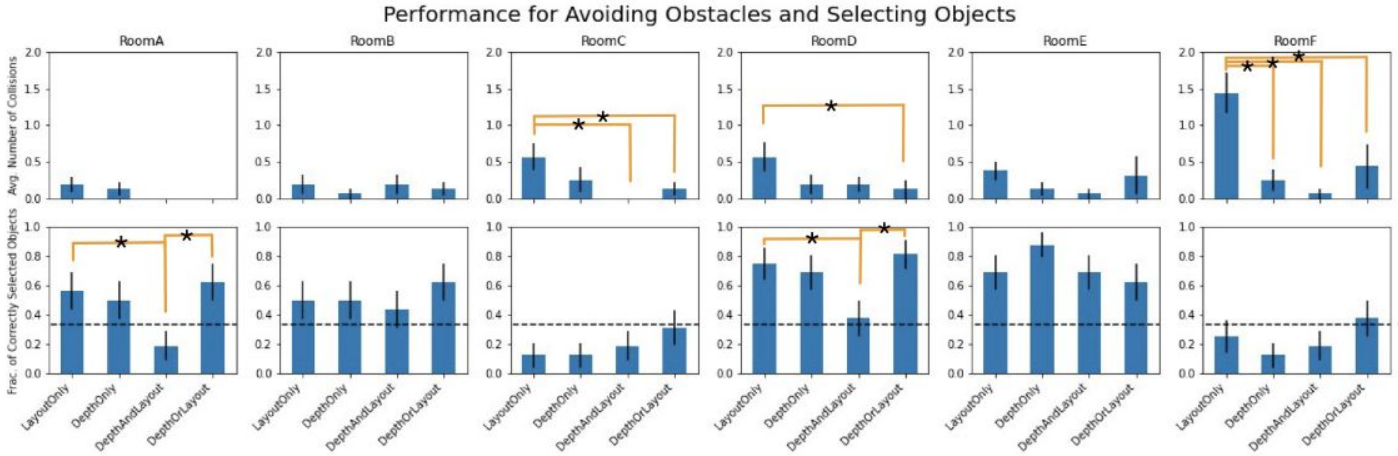


Results: Paths Taken by All Participants

- **RoomA** and **RoomF** similar in area covered, but F many more collision obstacles
- **RoomB:** most users went straight in and collided with center obstacle
- **RoomC** users went left
- **RoomD** users went right
- **LayoutOnly** much more variable paths, more collisions that other modes

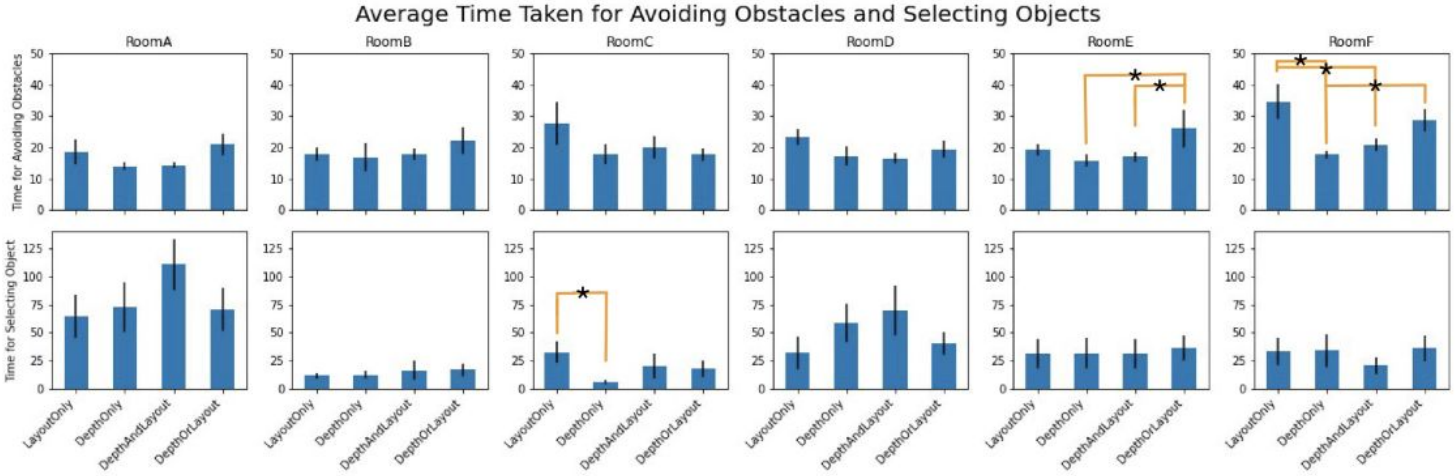


Results: Performance Within Rooms



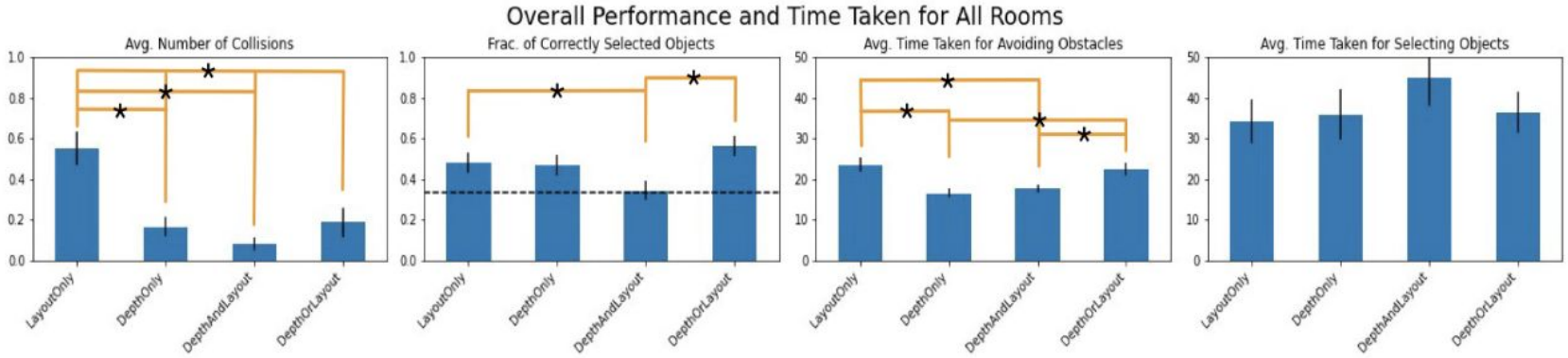
- **RoomC**: Participants collided less using *DepthAndLayout* or *DepthOrLayout* than *LayoutOnly*
- **RoomD**: Participants collided less using *DepthOrLayout* than *LayoutOnly*
- **RoomF:** Participants collided less using any mode other than *LayoutOnly*
- Participants were worse at selecting the correct object using *DepthAndLayout* compared to *DepthOrLayout* or *LayoutOnly* in RoomA, D

Results: Times Within Rooms



- **RoomE**: Took longer to avoid obstacles using D*epthOrLayout* compared to using *DepthAndLayout* and *DepthOnly*
- **RoomF**: Took longer to avoid obstacles using *DepthOrLayout* compared to *DepthOnly*, and longer for *LayoutOnly* compared to *DepthAndLayout* or *DepthOnly*
- Took less time selecting objects using *DepthOnly* than *LayoutOnly* in RoomC

Results: Across Rooms



- When avoiding obstacles, users collided with significantly more objects when using *LayoutOnly* compared to all other modes
- When selecting objects, *DepthOrLayout* and *LayoutOnly* performed better across the set of all the rooms than *DepthAndLayout*
- For time taken to avoid obstacles, *DepthOnly* was faster than both *LayoutOnly* and *DepthOrLayout*, *DepthAndLayout* was faster than *DepthOrLayout* and *LayoutOnly*

Results: User Preference

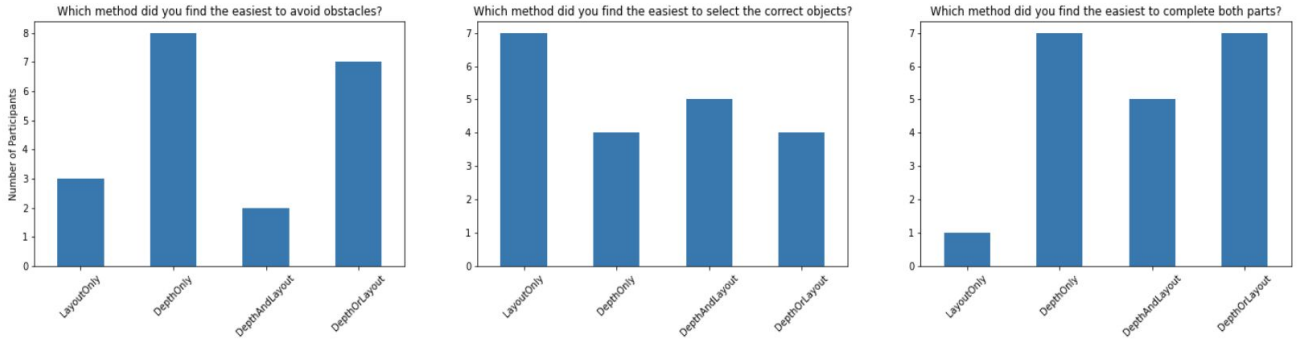
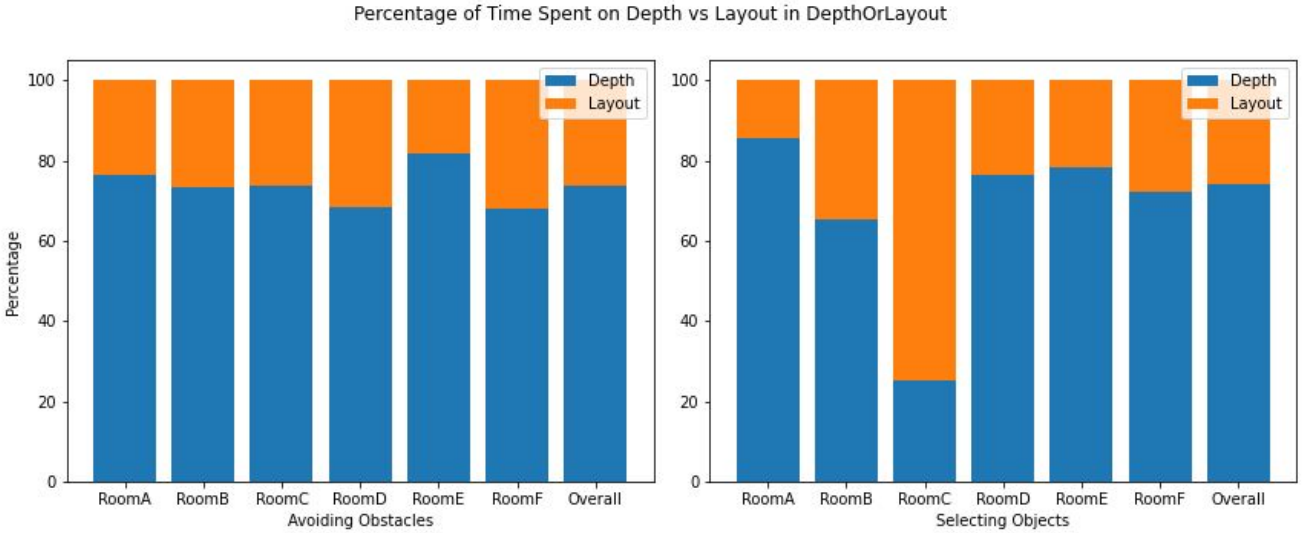


Figure 9: A histogram of each subject's preferred mode for each part of the experiment and overall

- Most preferred mode for avoiding obstacles was *DepthOnly* followed by *DepthOrLayout*
- Most preferred mode for selecting objects was *LayoutOnly*
- Overall most preferred modes were ***DepthOrLayout*** and ***DepthOnly***

Results: Depth vs. Layout Time when Switching

DepthOrLayout: Most time was spent in *DepthOnly* mode



Discussion

- Here, we studied the effectiveness of using a **combination** of computer vision based image preprocessing techniques for **obstacle avoidance** and **object recognition**
- We found that:
  - *LayoutOnly* mode led to significantly **more collisions** than modes that had any sort of depth-encoded information
  - *DepthAndLayout* led to participants having a harder time selecting the correct object, perhaps because of the over-stimulation of electrodes
  - Given the choice between depth and layout, participants chose to see the depth encoded information most of the time
  - Participants preferred the *DepthOrLayout* or *DepthOnly* overall compared to the other modes
- Overall, *some* type of depth encoded information is more useful than layout encoded information
  - More informative? More resemblance of natural vision?

Limitations & Future Work

- Limitations:
  - Most of the differences between modes were not statistically significant. Limited sample size?
  - Performance varied a lot across rooms:
    - Number and location of tables varied across rooms
    - Number and size of collider objects varied across rooms
  - Motivation of the users
    - Generally a hard task – SPV creates unnatural, distorted vision
    - Towards the end of the session, participants often got tired and simply wanted to finish the study
- Future work:
  - Testing different types of computer vision-based modes
  - Outdoor scenes as opposed to indoor scenes
  - More realistic objects to select and obstacles to avoid

Questions?