

Hyperion: A Wearable Augmented Reality System for Text Extraction and Manipulation in the Air

Dimitris Chatzopoulos^{*}, Carlos Bermejo^{*}, Zhanpeng Huang^{*}, Arailym Butabayeva^{*}, Rui Zheng^{*}, Morteza Golkarifard[^], and Pan Hui^{*}

^{*}Systems and Media Lab,
The Hong Kong University of Science and Technology

[^]Sharif University of Technology

Mobile Augmented Reality

- supplements the reality of mobile users with virtual objects
- The amount of virtual objects can be huge and the selection of the most suitable ones is not trivial
 - user's context
- device limitations
- QoE guarantees

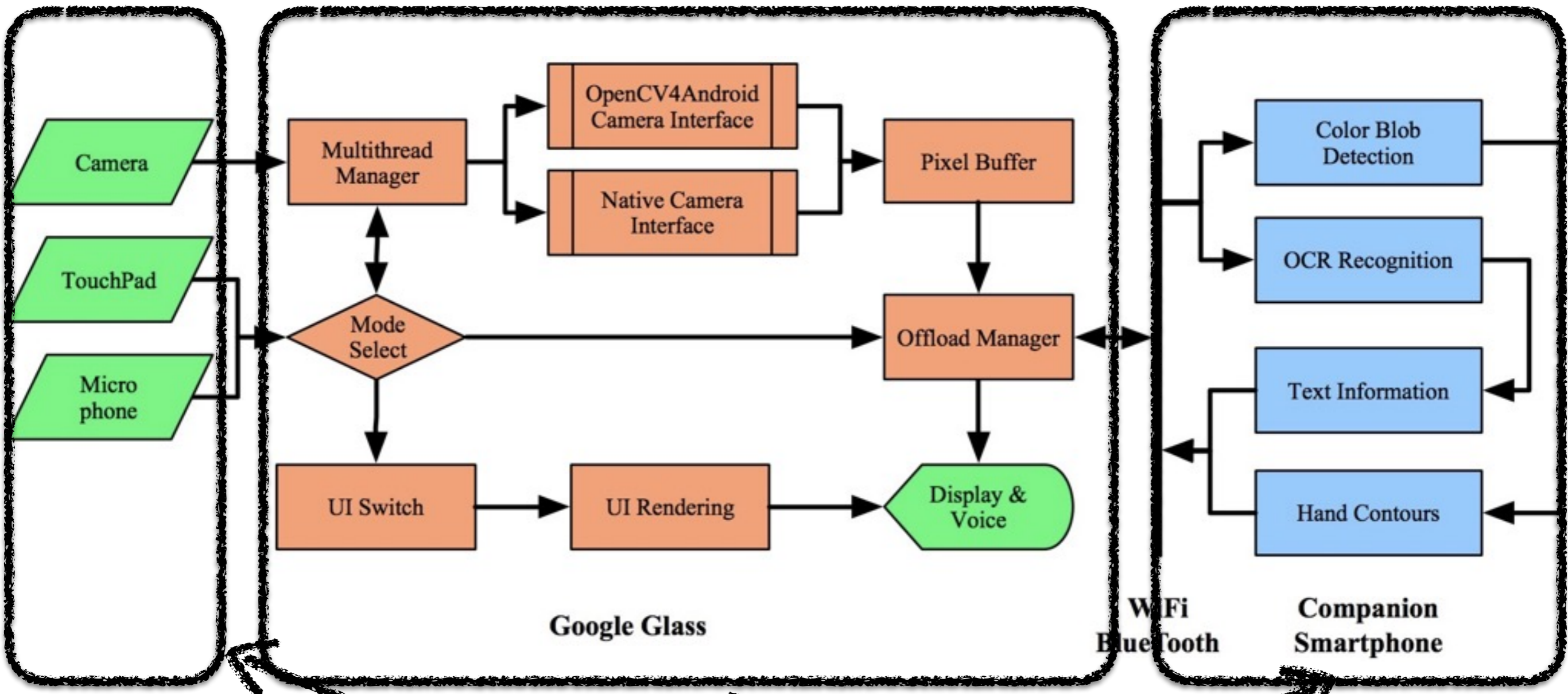
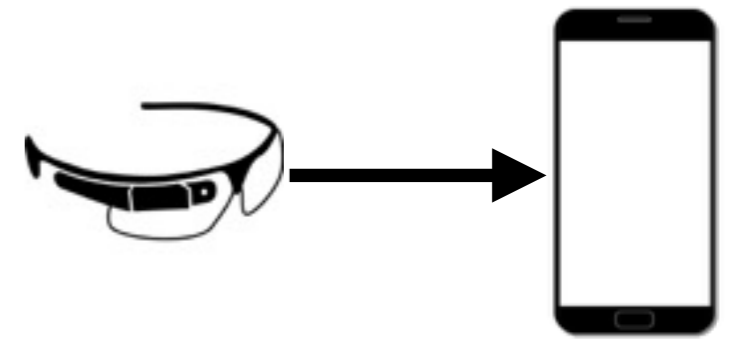
Optical Character Recognition

- extracts text from the current view of a mobile user
- can be used as an input to MAR applications
- is computationally expensive
 - computation offloading techniques are required.

Computation Offloading

	Cloud	Cloudlets	Smartphones
Pros	<ul style="list-style-type: none">• Resourceful Offloadees	<ul style="list-style-type: none">• Stable connection• Small latency• Management• No need for Internet Connection	<ul style="list-style-type: none">• Nearby Devices• No need for Internet Connection
Cons	<ul style="list-style-type: none">• Internet Connection• High latency	<ul style="list-style-type: none">• Poor availability	<ul style="list-style-type: none">• Unstable connectivity

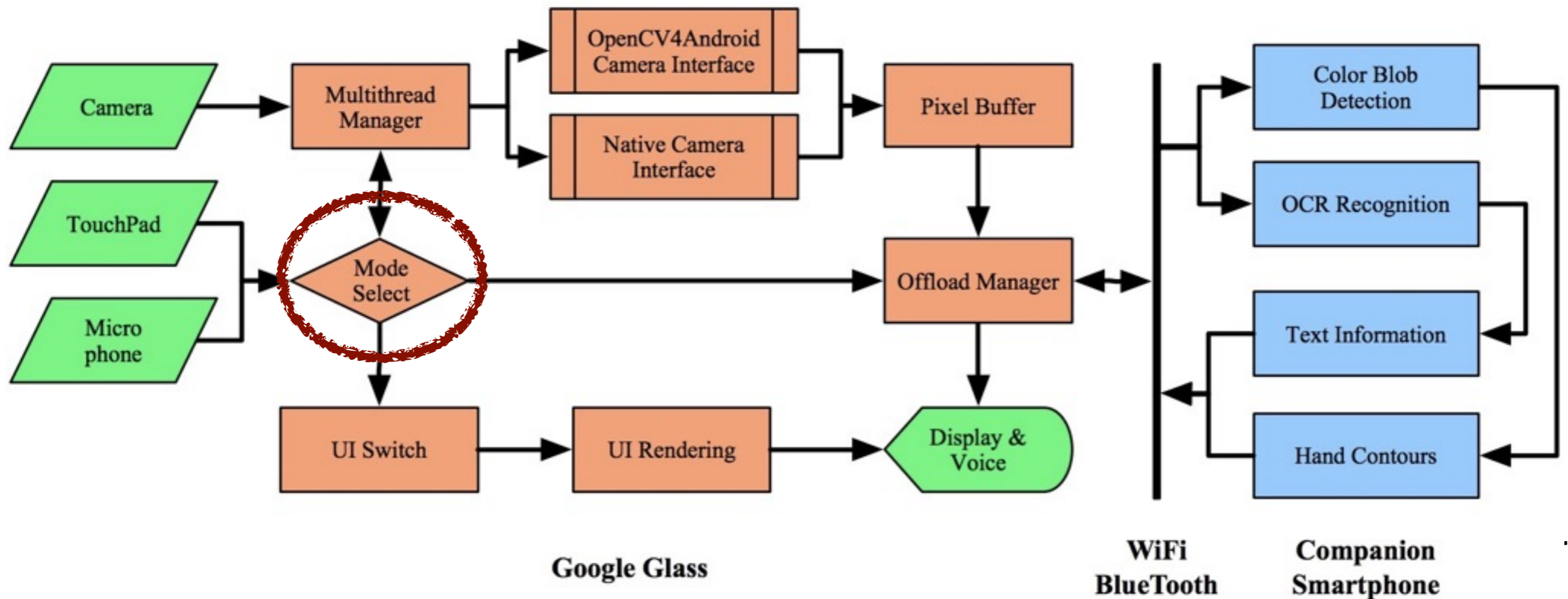
Hyperion



Offload parts

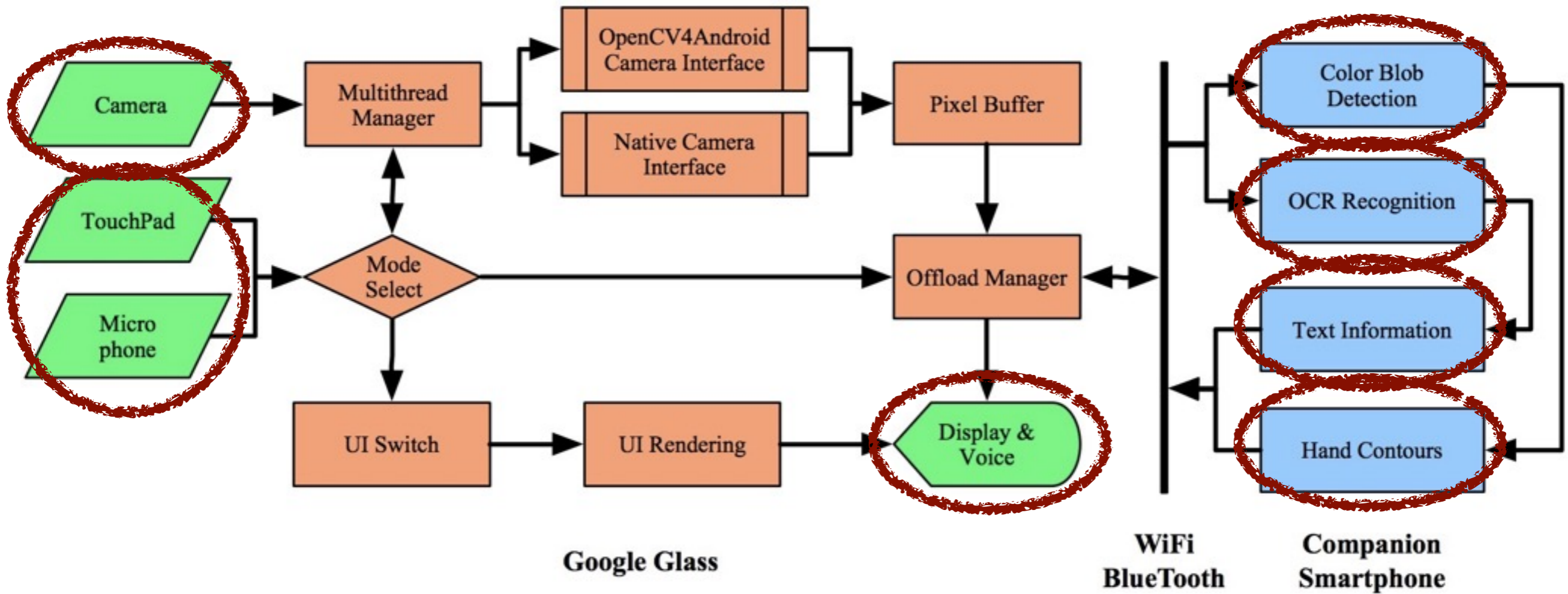
Design principles

- Simplicity
 - a simple system for a daily use
- Relevance
 - focused to the needs of the user
- Adaptivity
 - adapted to the needs of the scenario



is pointing. Translate selected text. Clip a text area.

- Meeting mode: Extract text from presentations
- Driving mode: Extract text from traffic signs
- Each modality has different inputs and outputs



- Hand gestures (colour-based)
- Speech Synthesis (Text-to-Speech)
- Visual Display (OpenCV4Android)

Hand gesture detection



- Hyperion uses a lightweight colour-based detection algorithm
- The method uses reference color for detection and requires to sample the color of the hand skin and then utilizes the skin color to extract pixel sections with similar color in current view.

Explore



Inputs

- Touchpad
- Voice commands
- Hand gestures

Output

- Visual Display
- Speech Synthesis

Reading



**Magnifying
Glass**



Inputs	Output
<ul style="list-style-type: none">• Touchpad• Hand gestures	<ul style="list-style-type: none">• Visual Display

Reading

**Abbreviation
Resolution**

**English to
Chinese
translation**

侵入的；打扰的

Graphics P

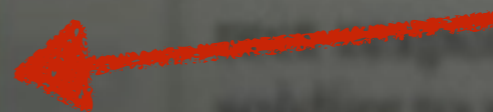
" ok glass "

Reading

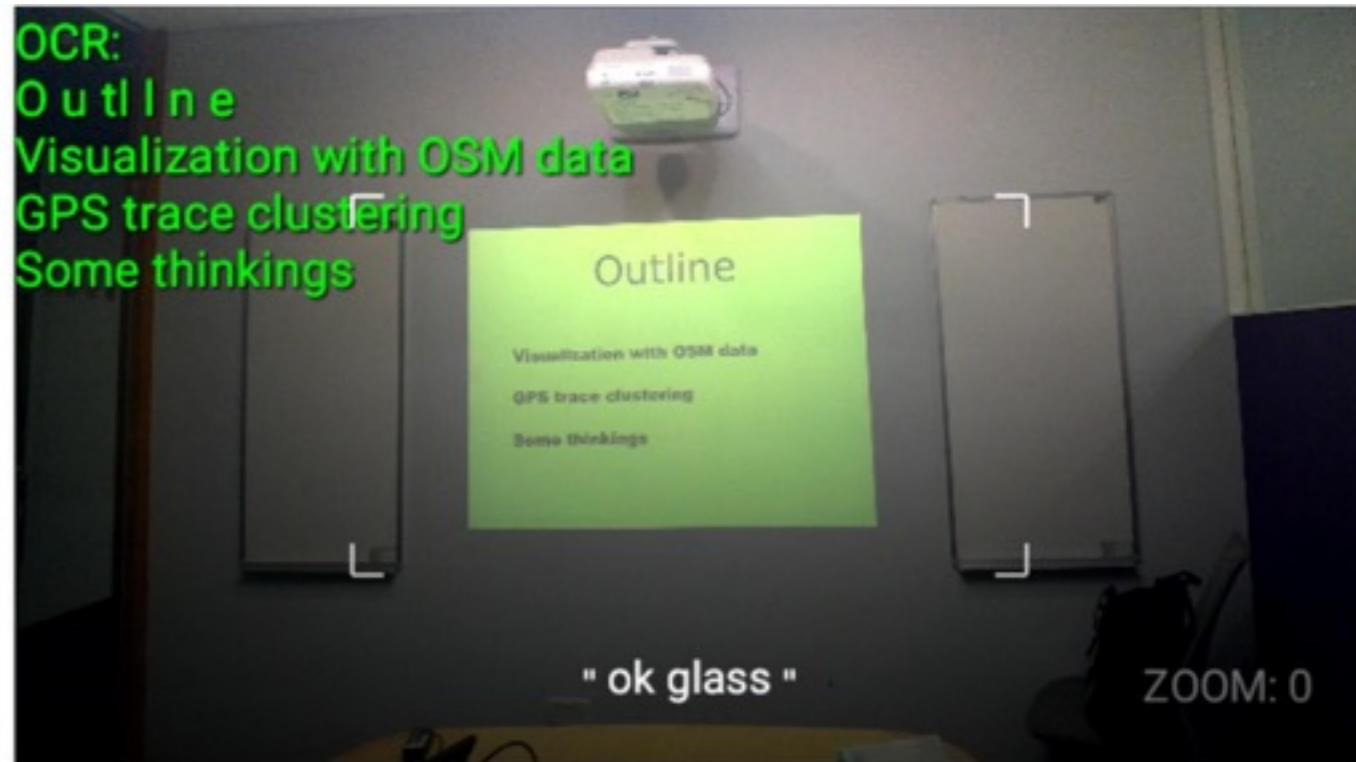
A model that approximates natural human behavior might be very different from one that drives a synthetic movie actor.

" ok glass "

Textarea Clipping



Meeting



Inputs

- Touchpad

Output

- Visual Display

Driving



Inputs

- Voice commands

Output

- Visual Display
- Speech Synthesis

Evaluation

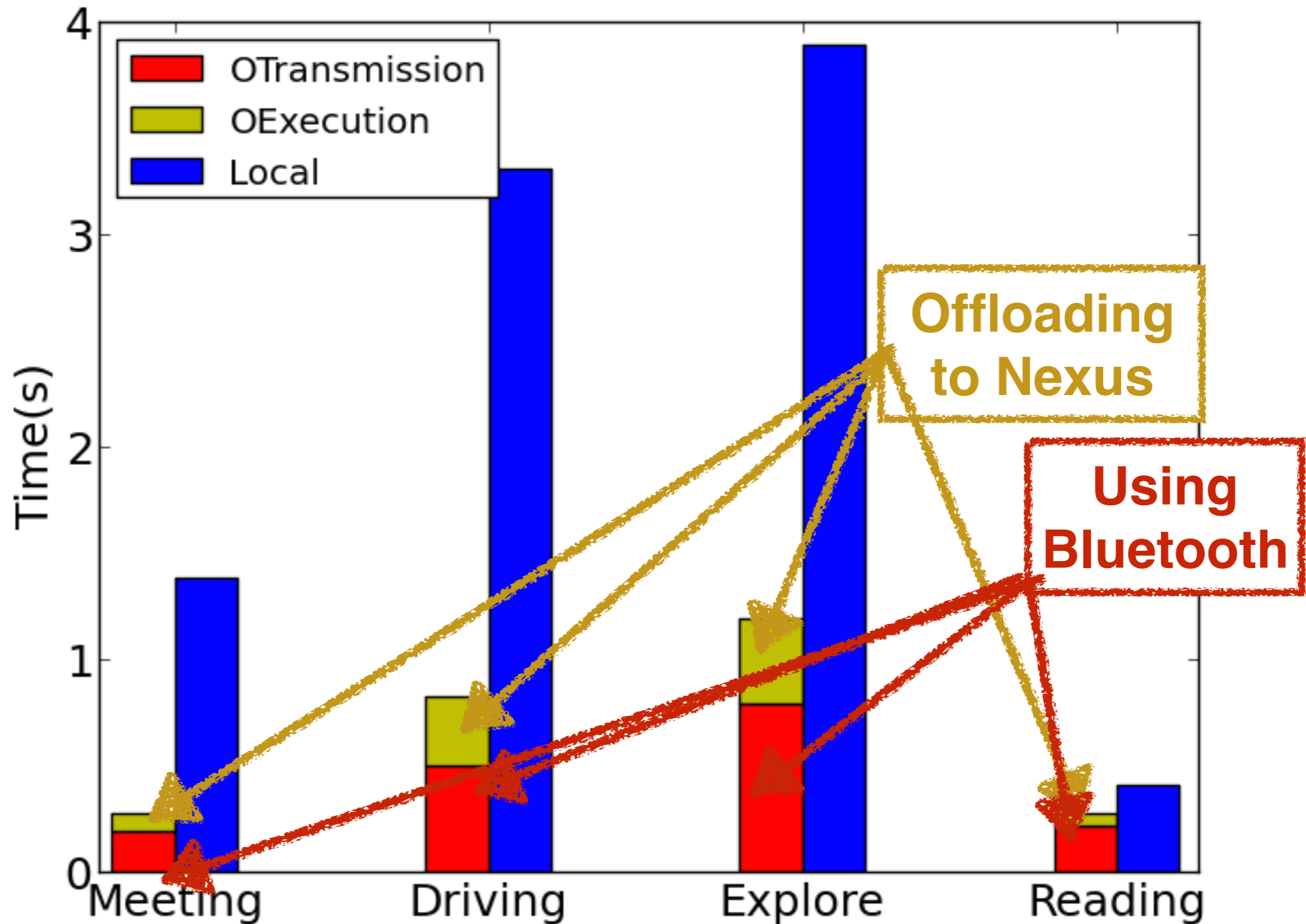
- System performance evaluation
- Quality of experience evaluation
- 4 cases, one for each modality: walking around campus, reading a book, having a meeting, and walking along a roadway.

System performance evaluation

- Google Glass
- Google Nexus 5 (2.26GHz CPU and 32 GB memory)
- Amazon EC2 instance
- Nearby cloudlet in 2hop distance

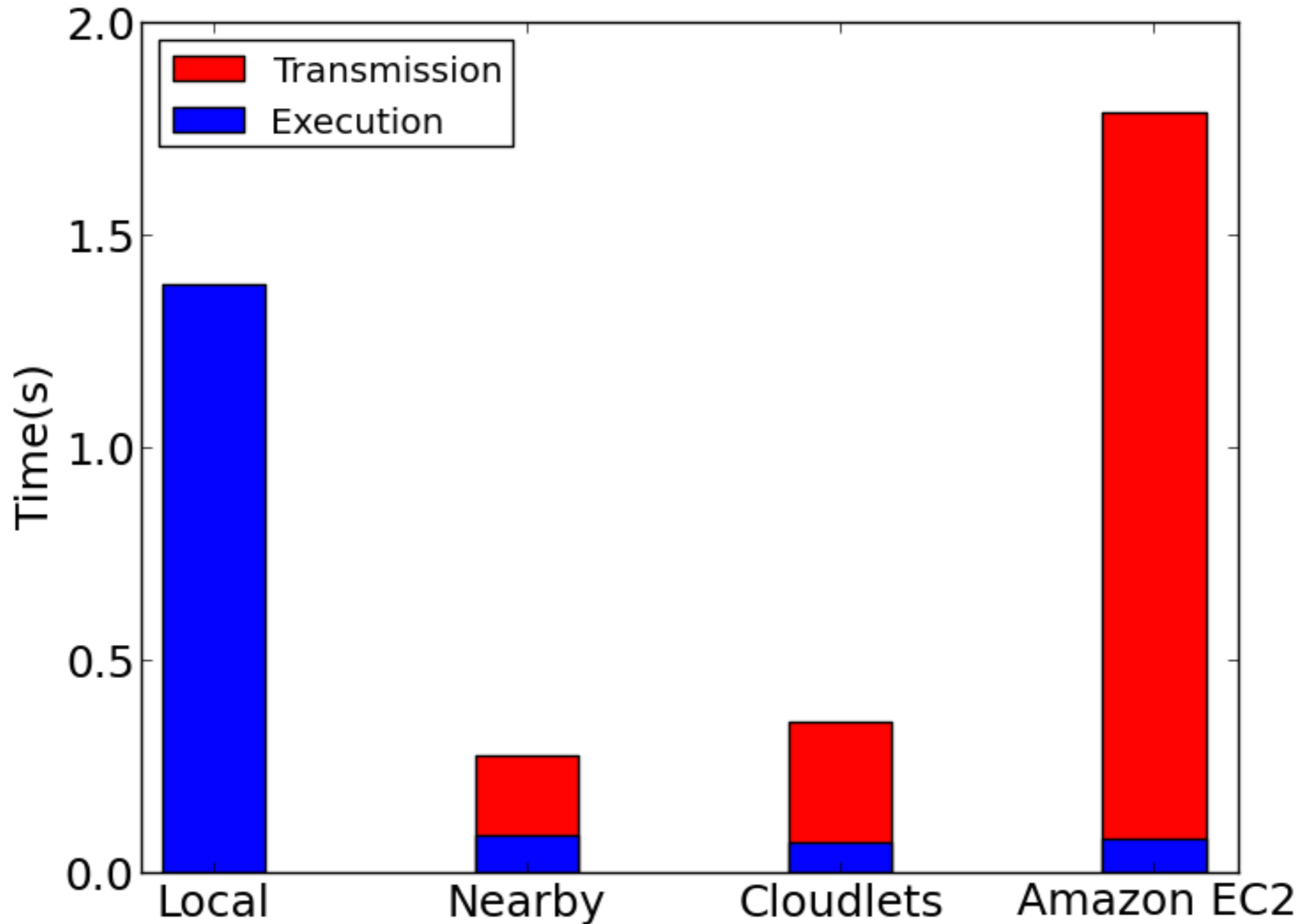
Running Time

All modes



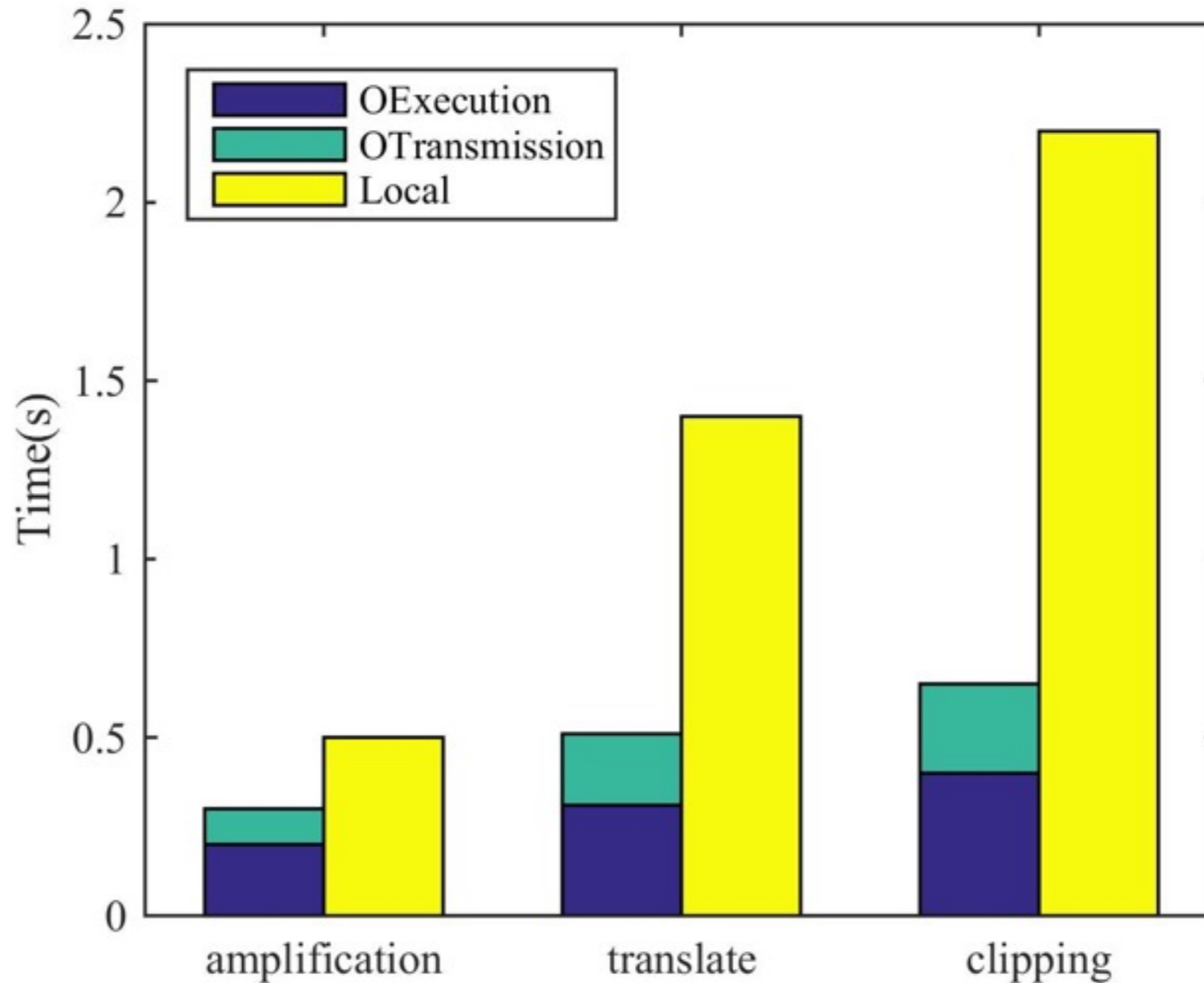
Running Time

Meeting mode



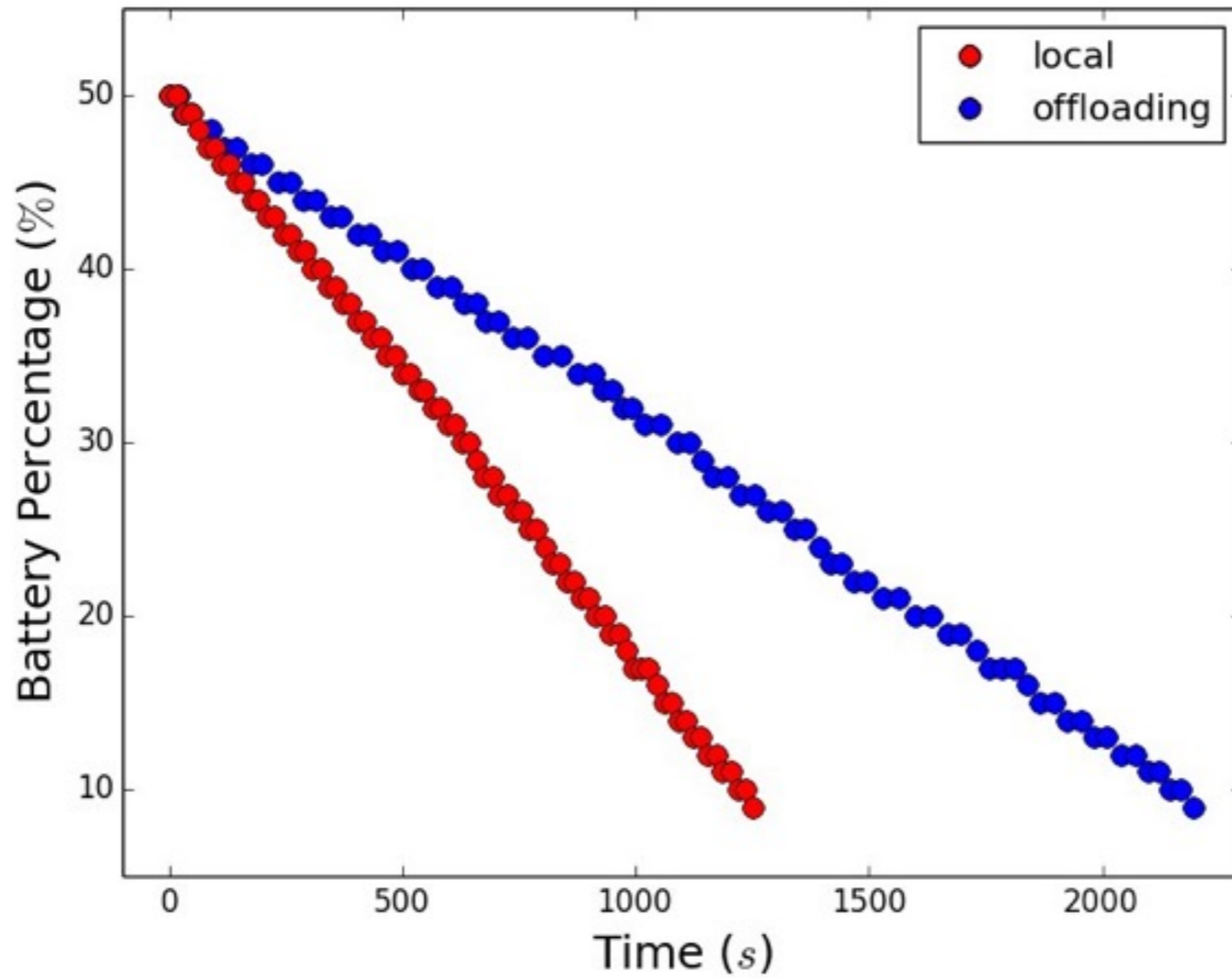
Running Time

Reading mode

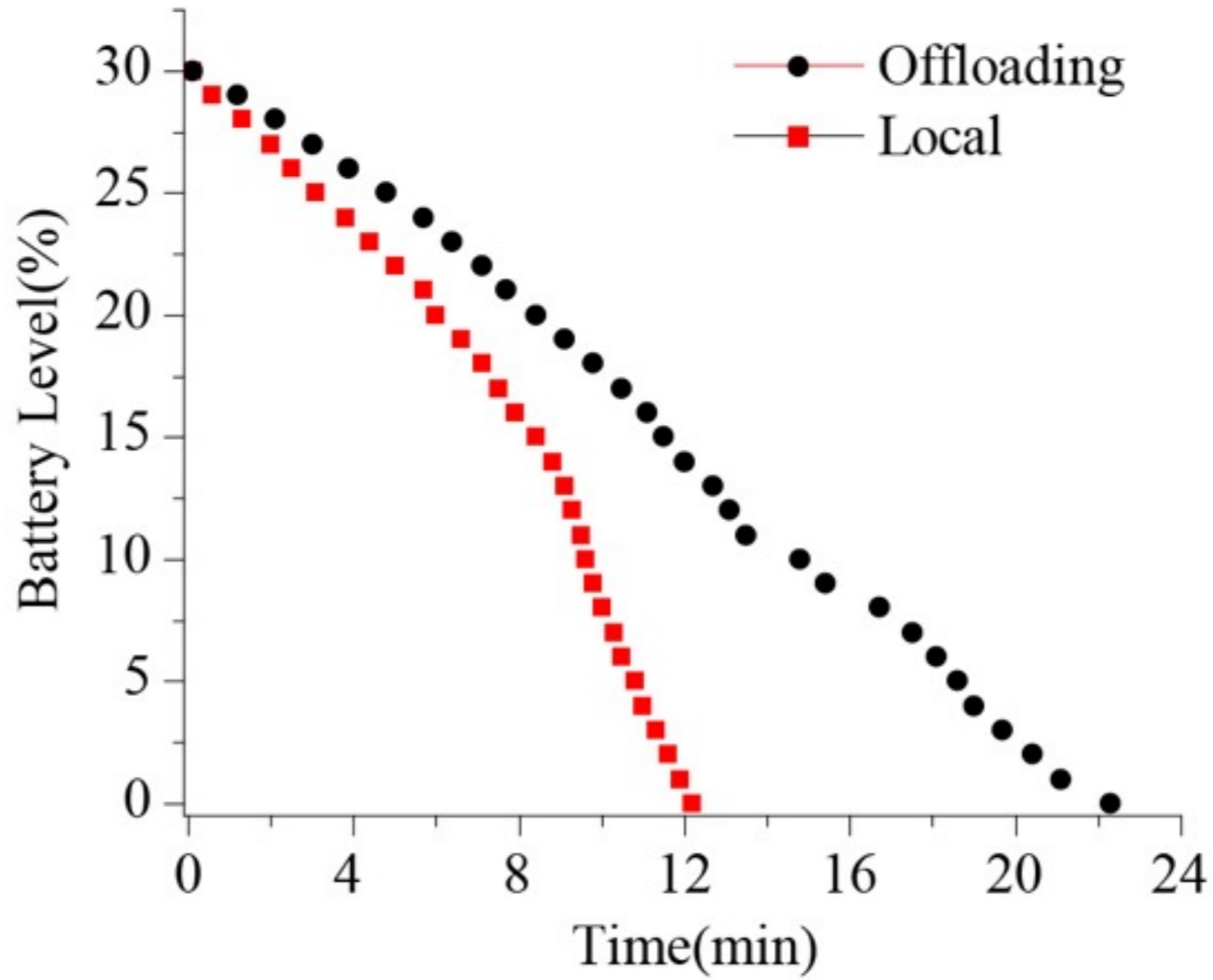


Battery gain

Explore mode



Battery gain



Rethinking visual impairment

- Affecting more than 285 million people around the world
- Blurry vision
- Cannot see objects at a distance
 - Hyperion detects text and
 - magnifies it or reads it to the user.

E

H N

D F N

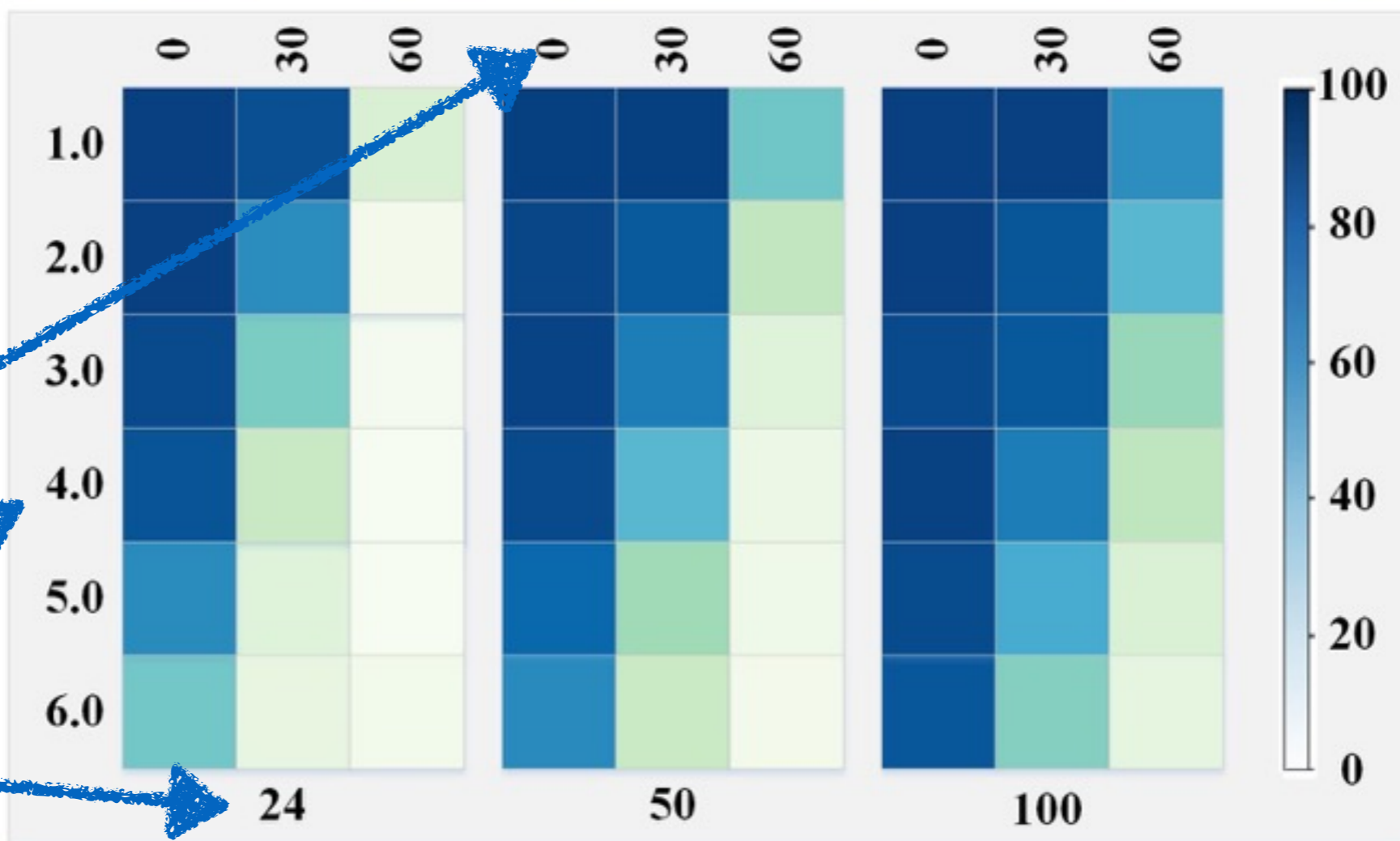
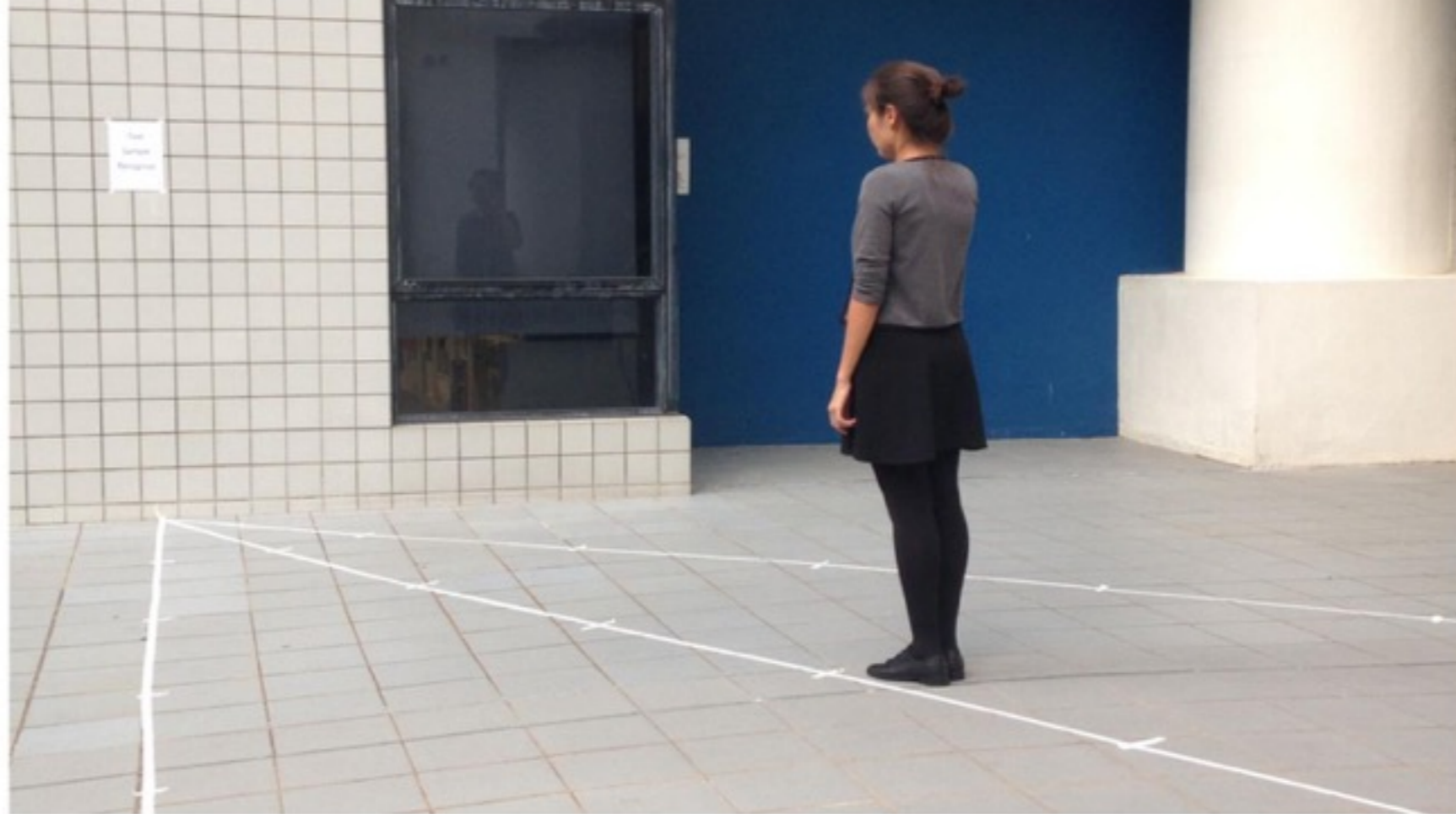
P T X Z

U Z D T F

D F N P T H

P H U N T D Z

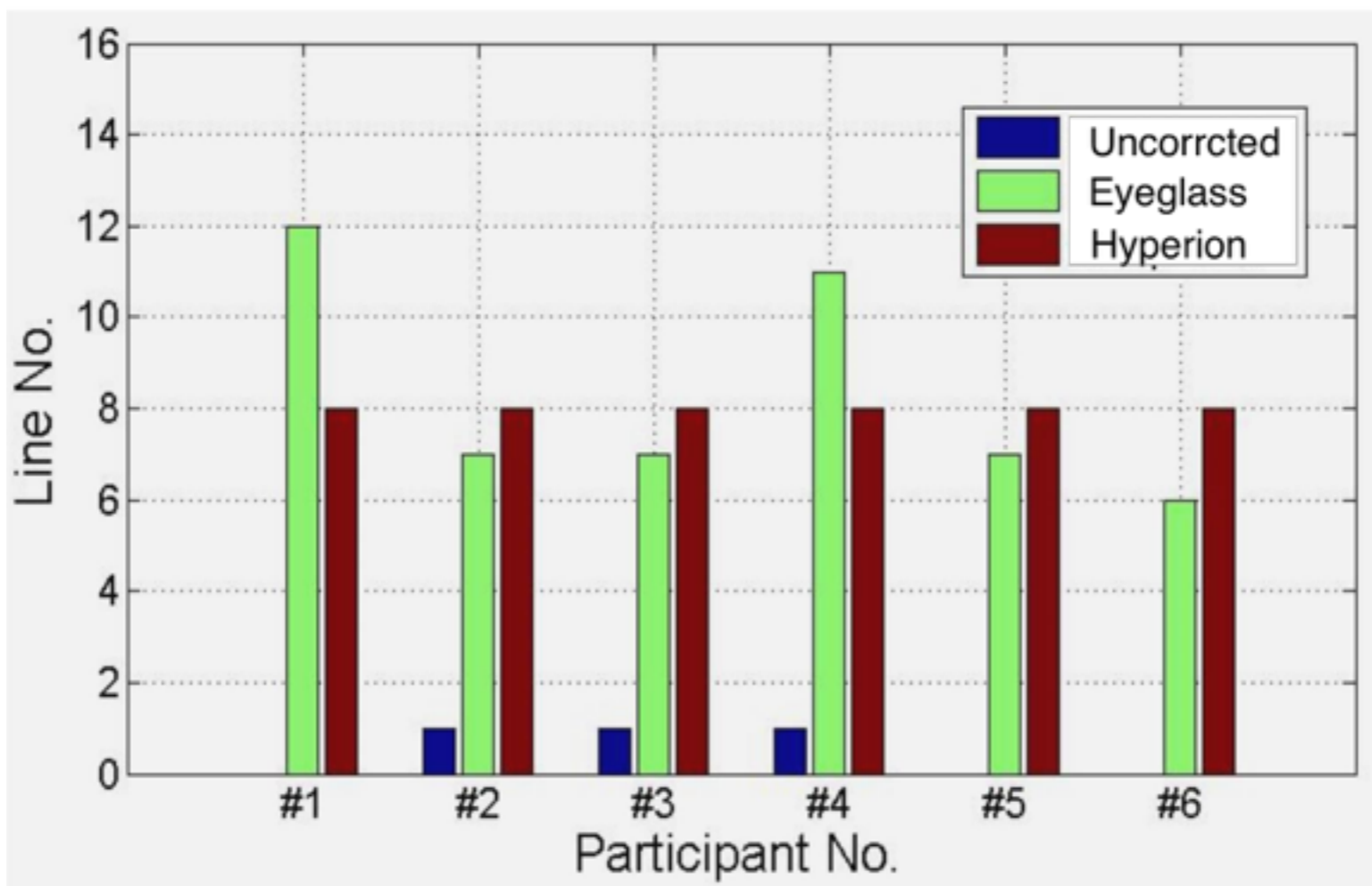
N P X T Z F H



angles

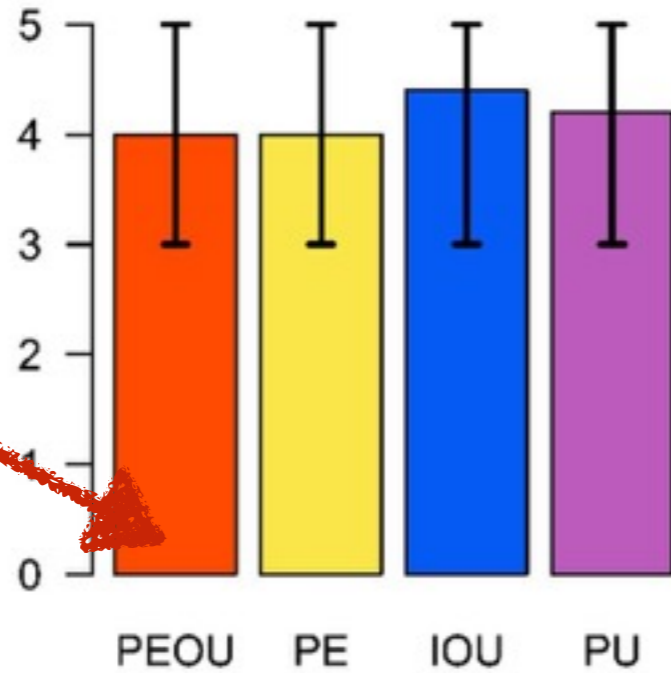
meters

font size

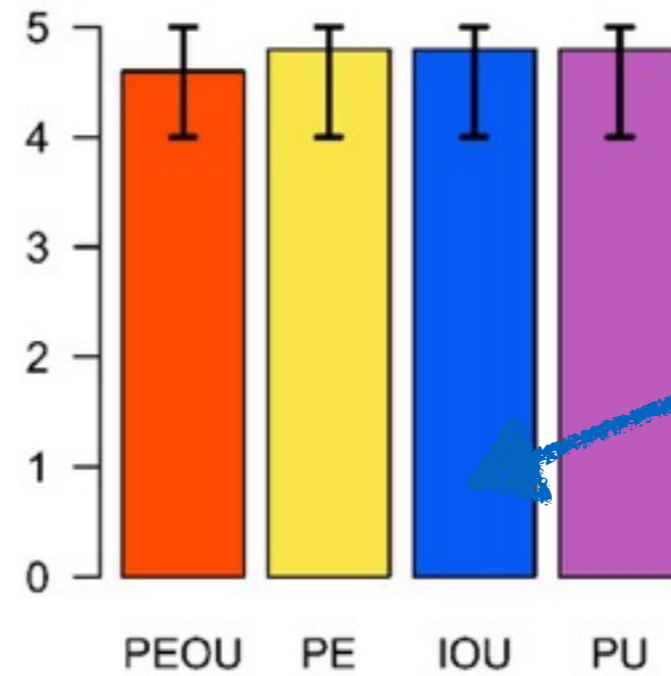


Likert scale

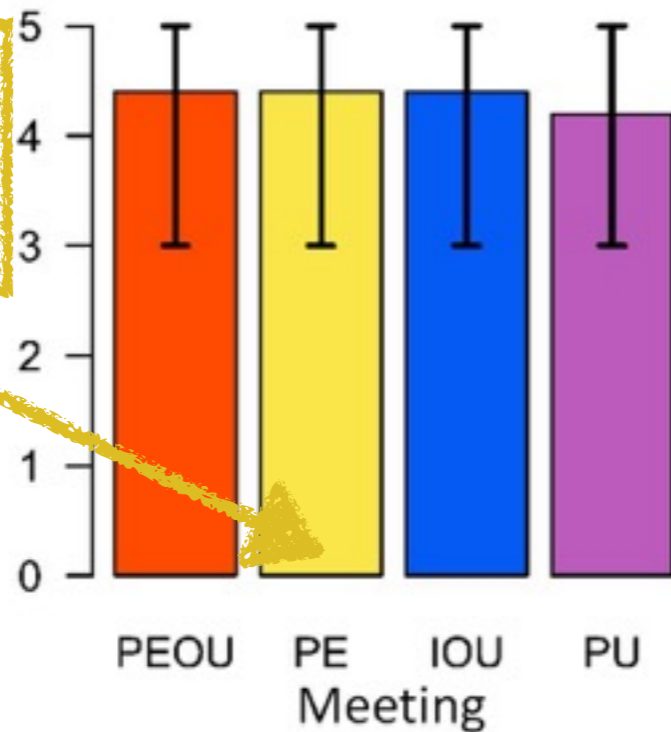
Perceived Ease of Use



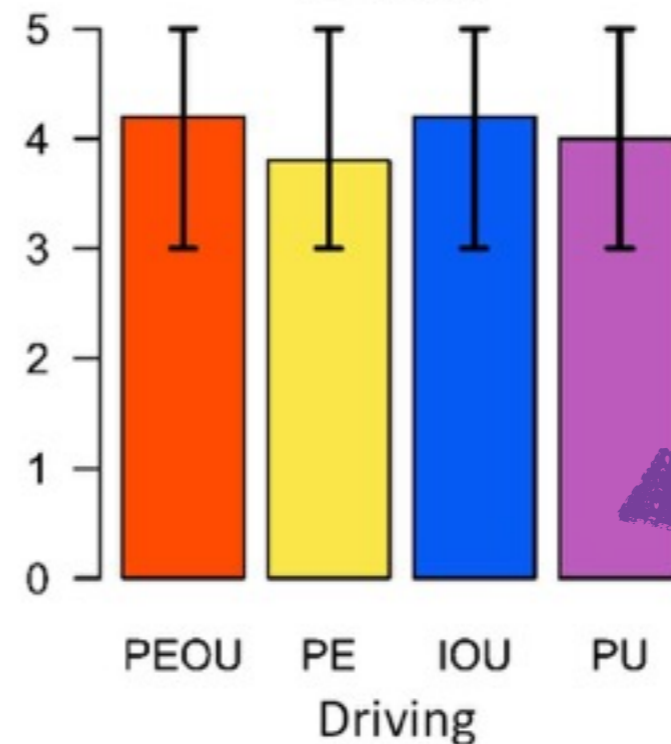
Intention of Use



Perceived Enjoyment



Perceived Usefulness



Thank you :)

Any questions ?