# Augmenting Human Capabilities for the Mixed Reality Future

Hrvoje Benko
Director, Research Science @ Reality Labs Research

@ MIT HCI Seminar
November 8, 2022

REALITY LABS RESEARCH

Meta

Facebook F8 2017

# A Vision of All-day MR

- Sensory and social superpowers
- Communicate and collaborate at a distance
- Next computing platform

Facebook F8 2017

# Metaverse

Meta Connect 2021

Hollerer, T., Bell, B., Feiner, S., et al.
*Mobile Augmented Reality System, ISAR 2001*

Hollerer, T., Bell, B., Feiner, S., et al.
*Mobile Augmented Reality System, ISAR 2001*


Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine - ACM ISAR 2001*


Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine - ACM ISAR 2001*


Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine - ACM ISAR 2001*


Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine - ACM ISAR 2001*

# What is taking so long?


Networking
Compute
Display
Optics
Audio
Tracking
Battery


Networking
Display
Compute
Optics
Audio
Interactions & Interfaces
Battery
Tracking

**Command Line Interfaces**
(mainframes, keyboard)

**1960s**

**Graphical User Interfaces**
(personal computers, keyboard & mouse)

**1980s**

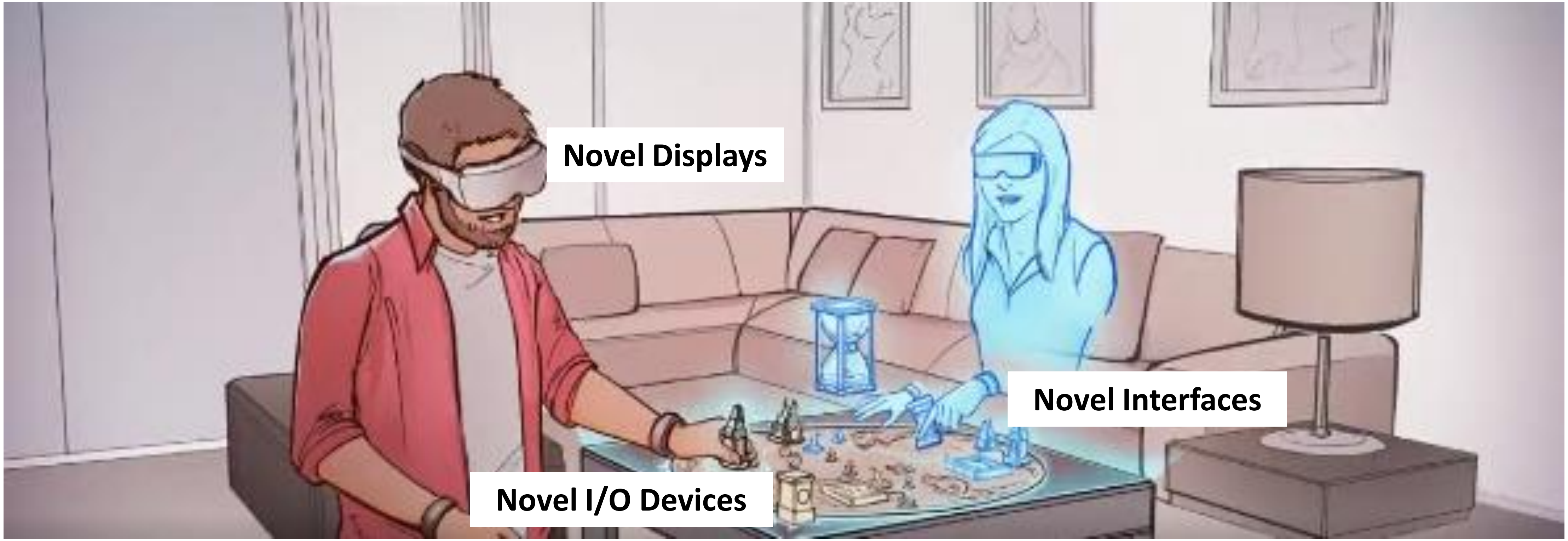
**Natural User Interfaces**
(tablets, smartphones, touch/gestures)

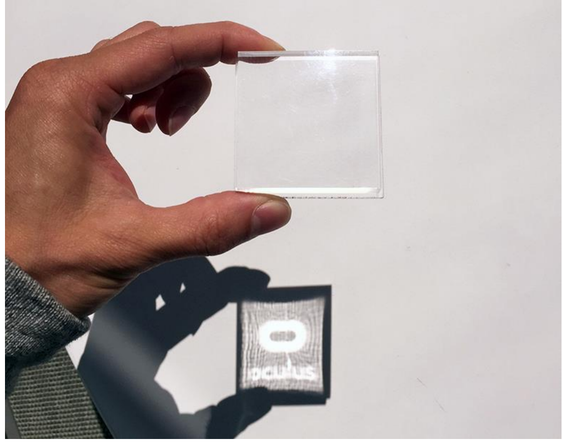
**2000s**

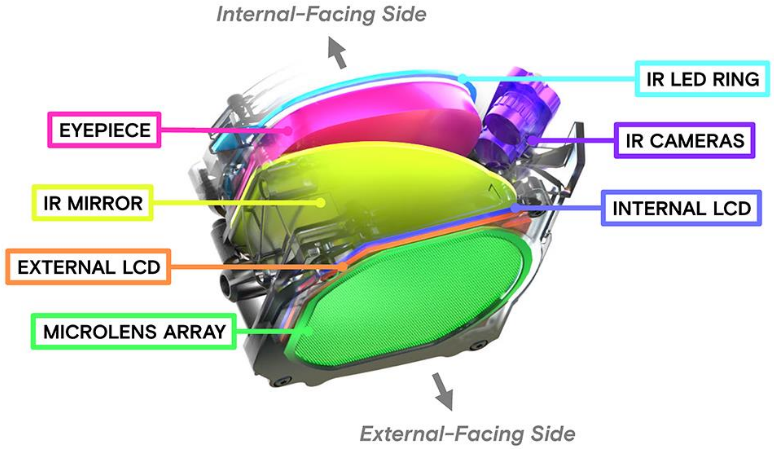
**Mixed Reality Interfaces**
(MR glasses, wristbands, ???)

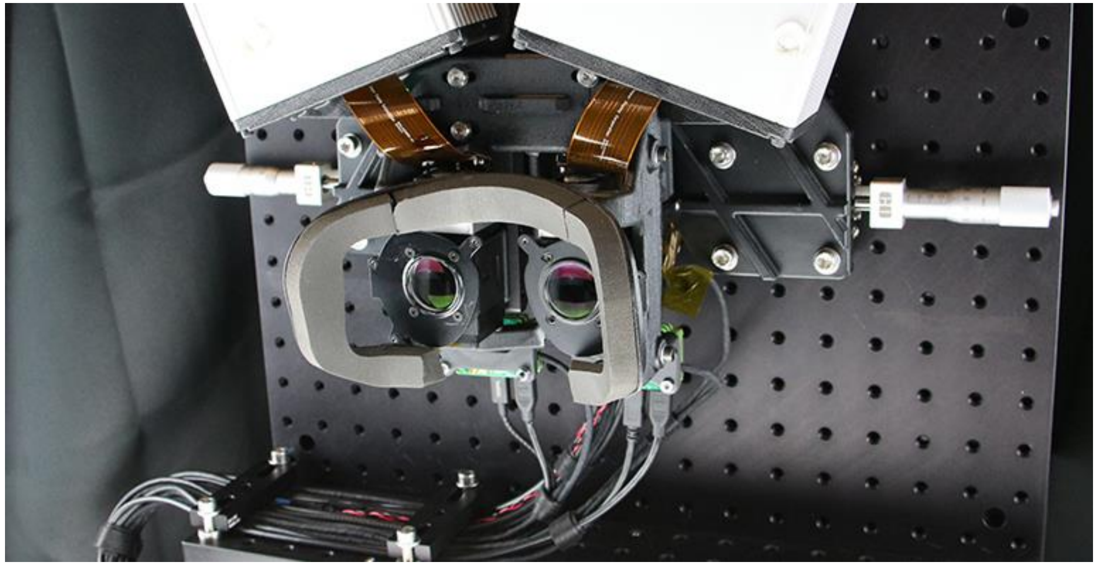
**2020s**

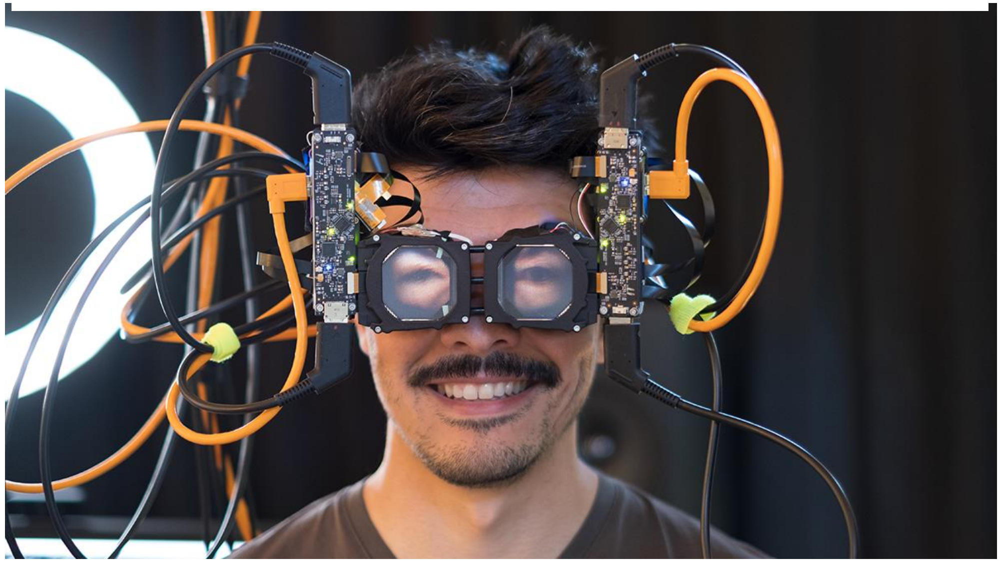
**New Computing Era =**
New Display Form Factor
+ New Input Method
+ New Interface


Novel Displays
Novel Interfaces
Novel I/O Devices



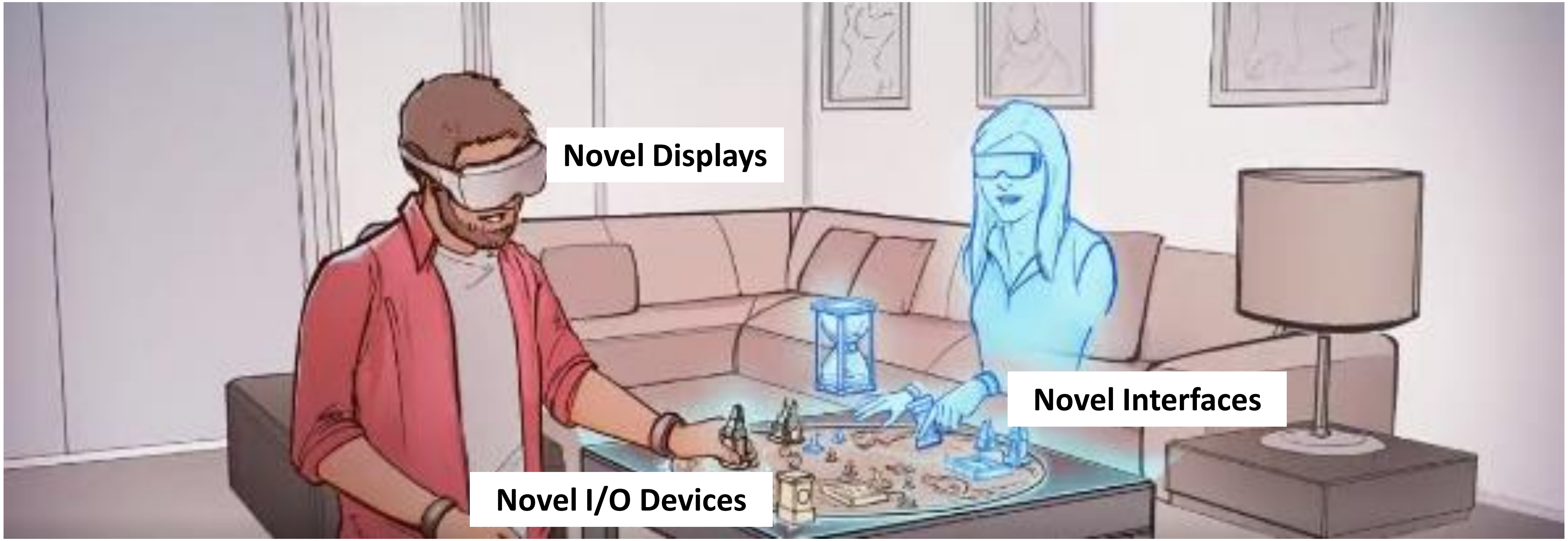


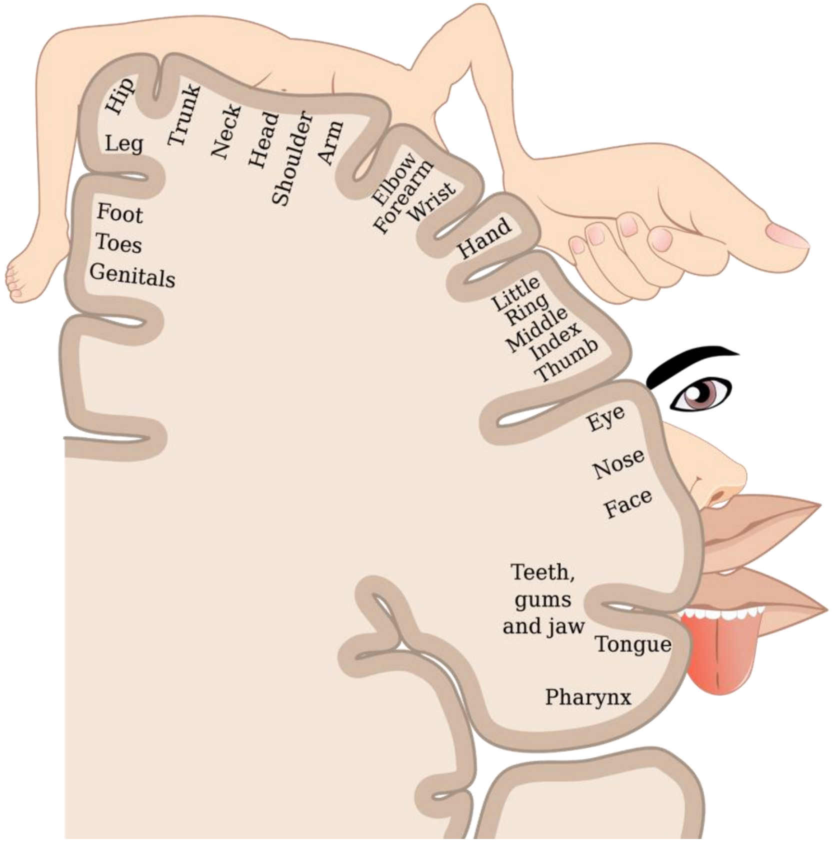


https://research.fb.com/blog/2017/05/oculus-research-spotlight-meet-the-team-behind-focal-surface-displays/

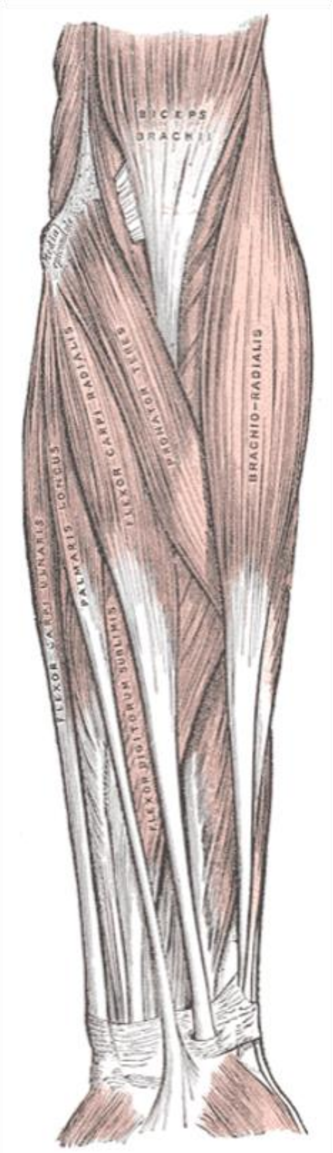
https://research.fb.com/blog/2021/08/display-systems-research-reverse-passthrough-vr/


Novel Displays
Novel Interfaces
Novel I/O Devices

# Novel XR Wristbands

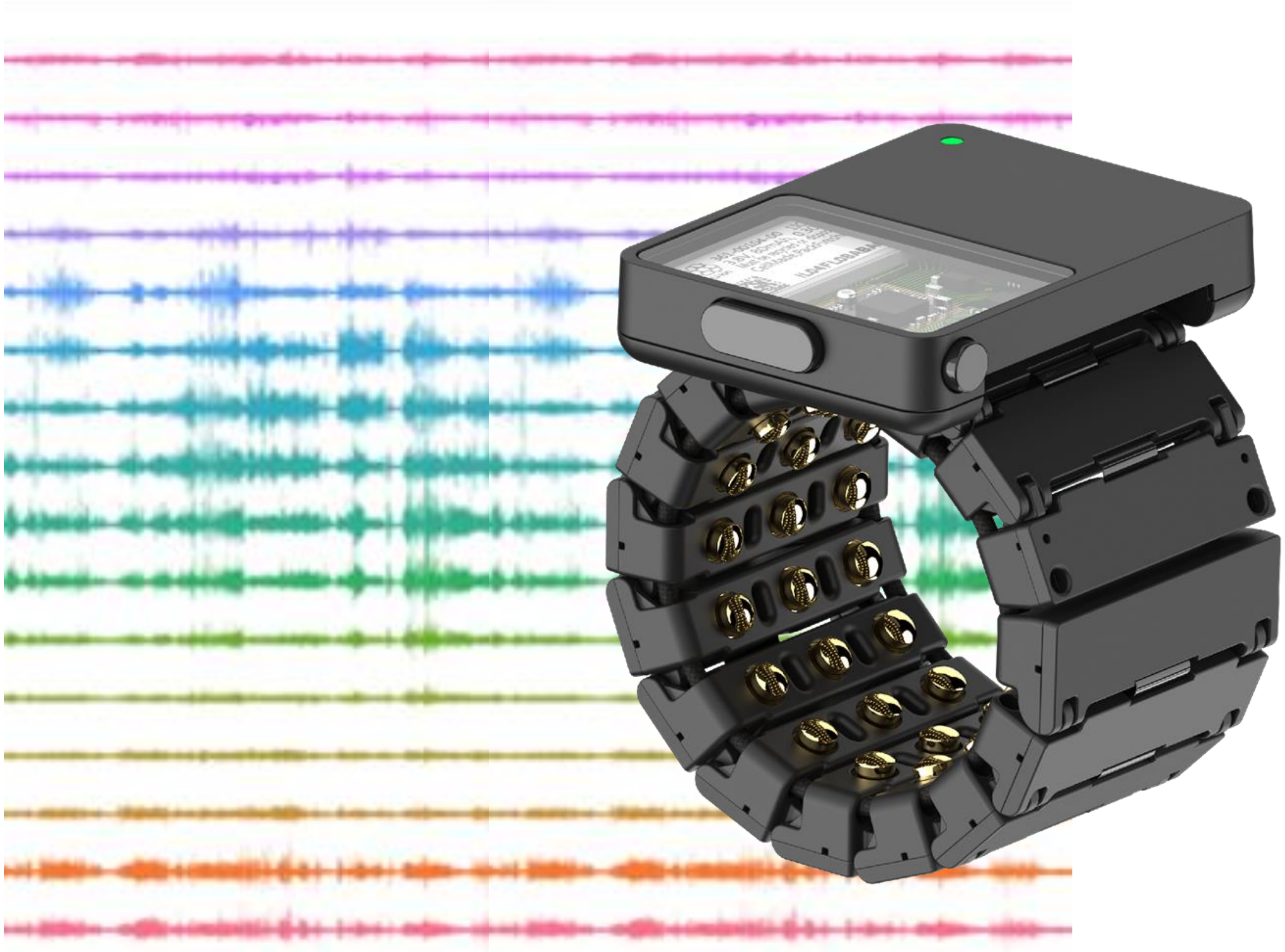


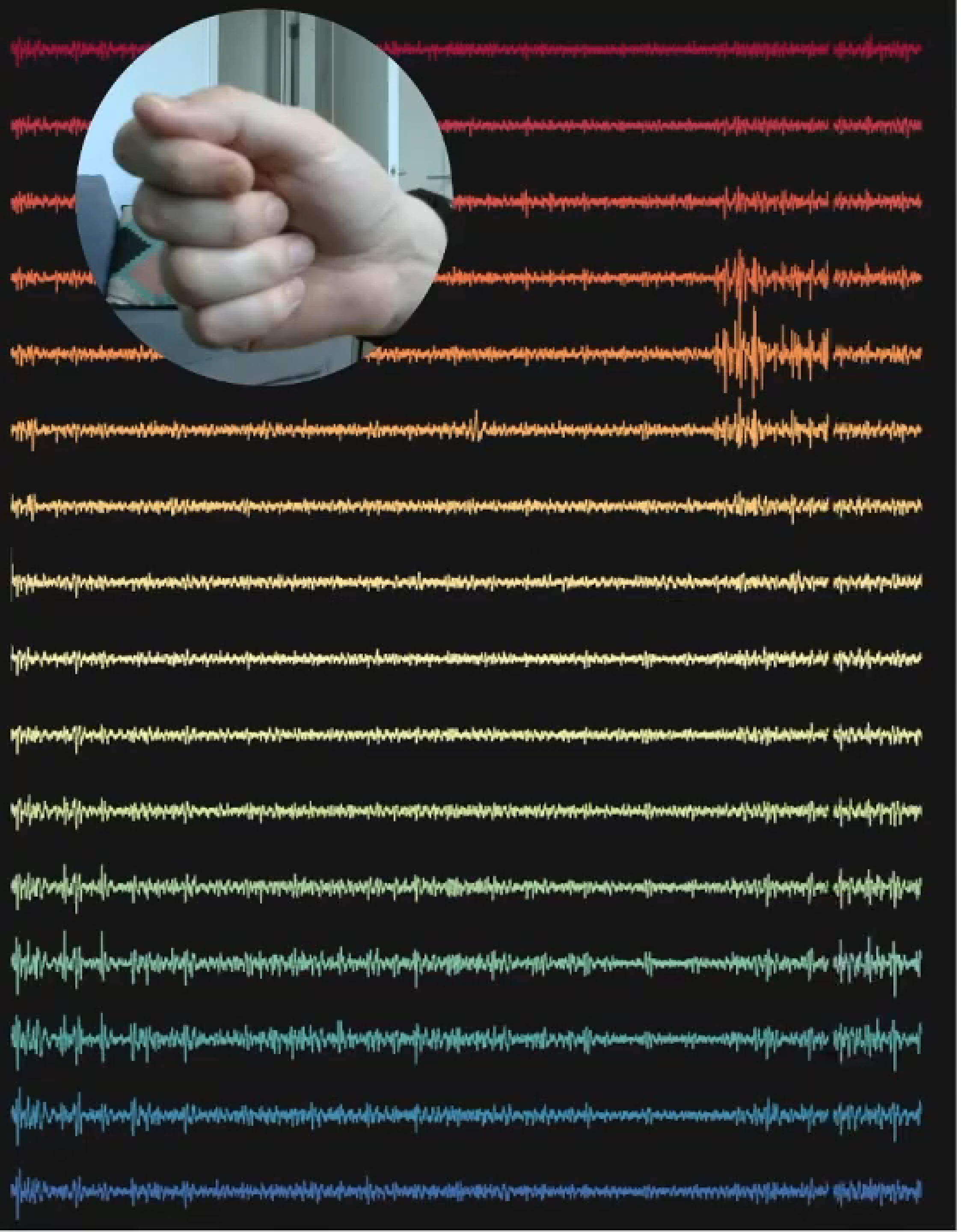
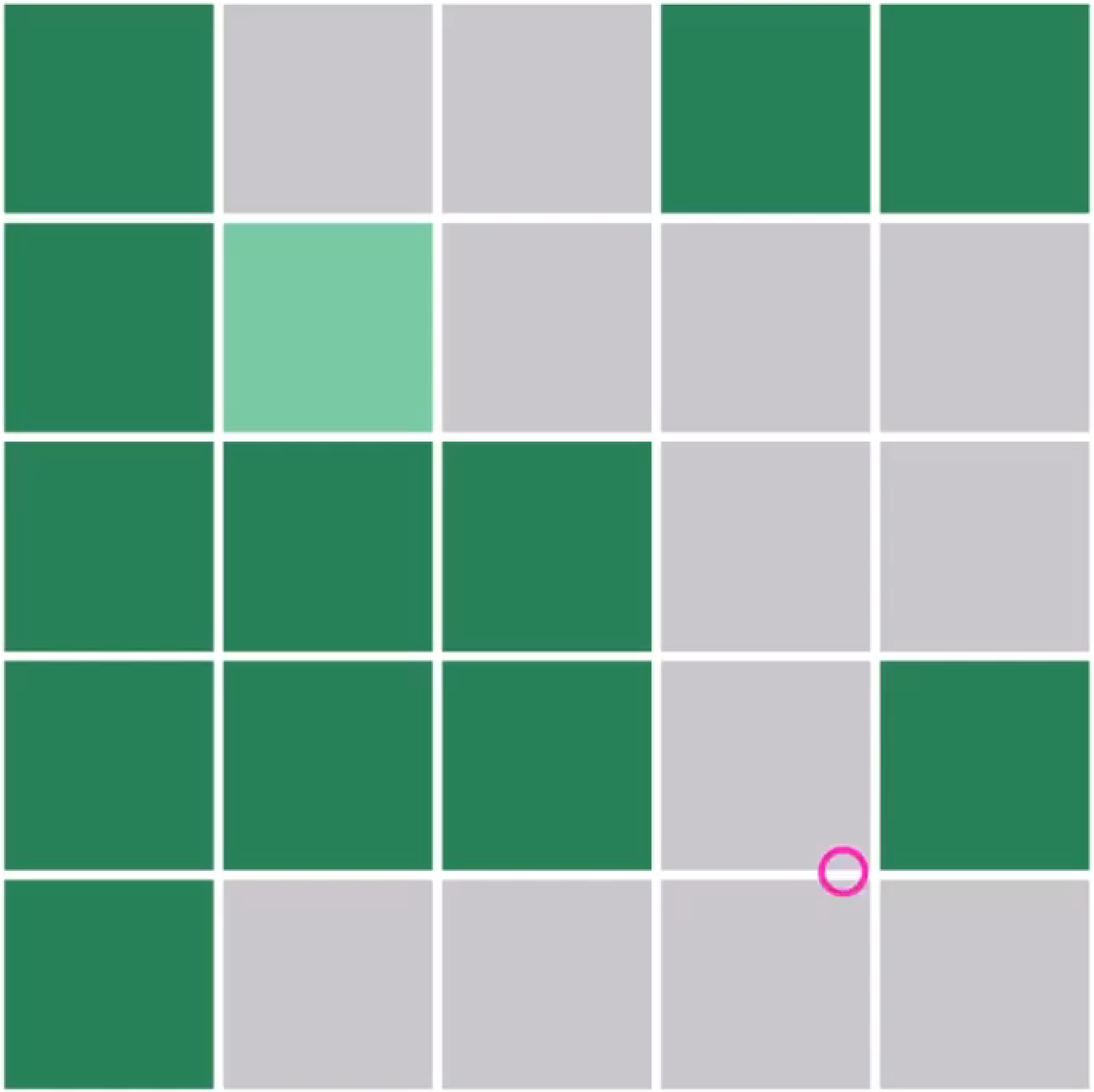

Penfield & Rasmussen. 1950



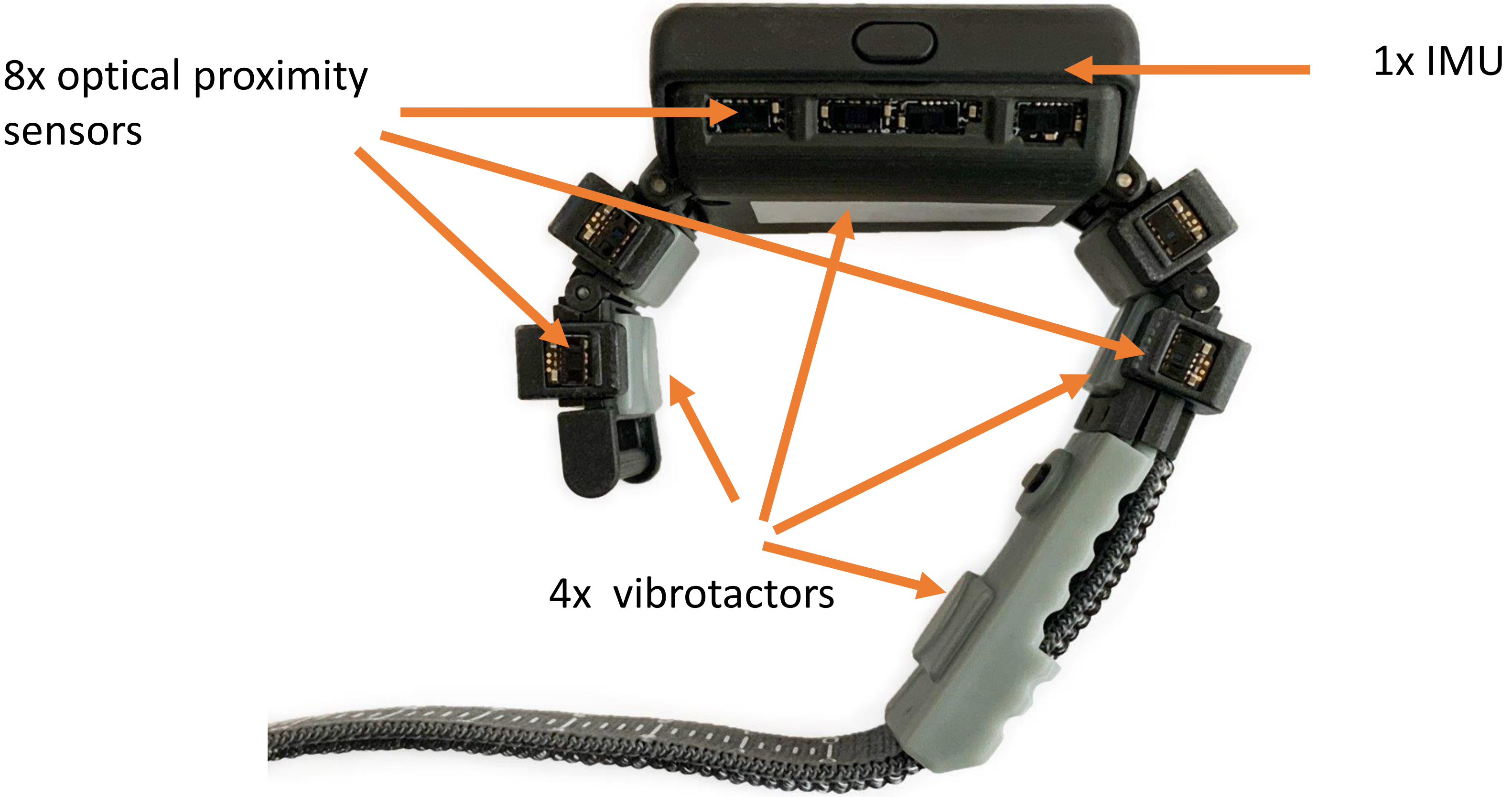

Gray’s Anatomy Plates. 1918

# Electromyography Wristbands



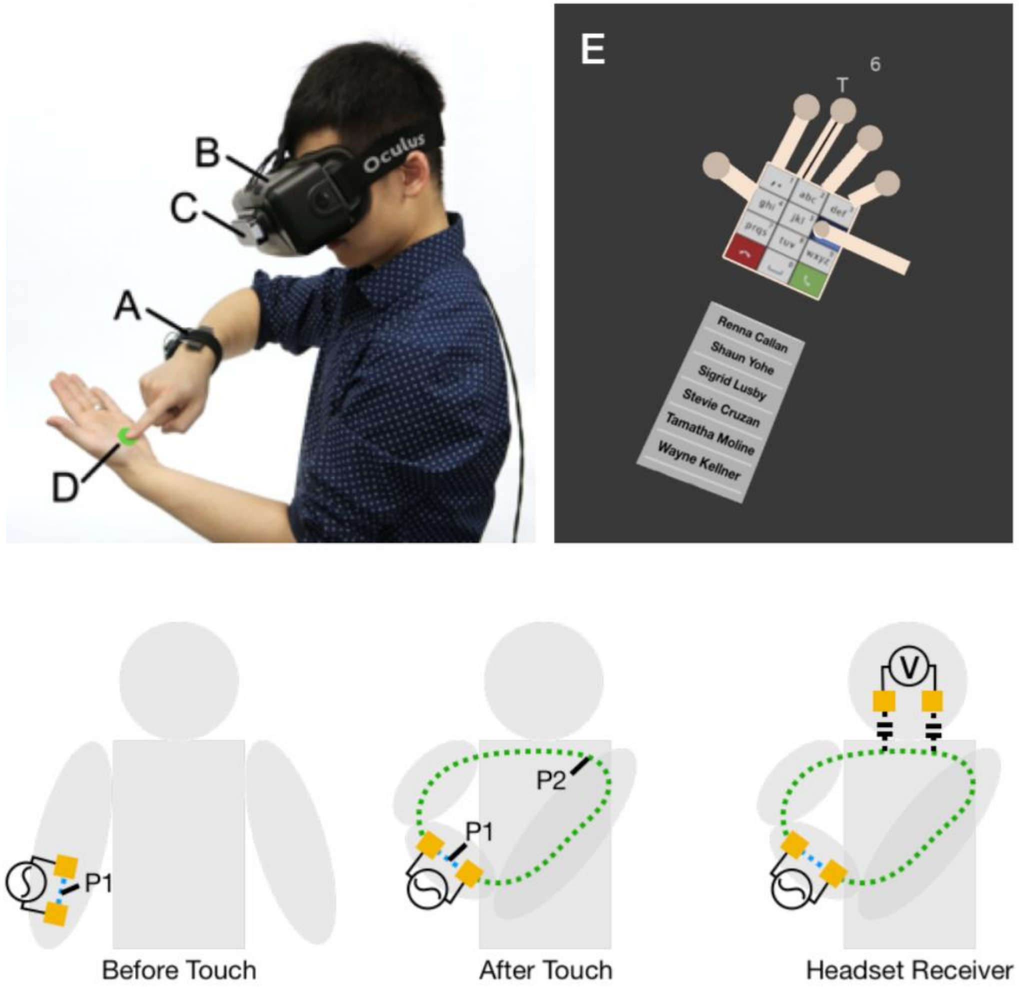



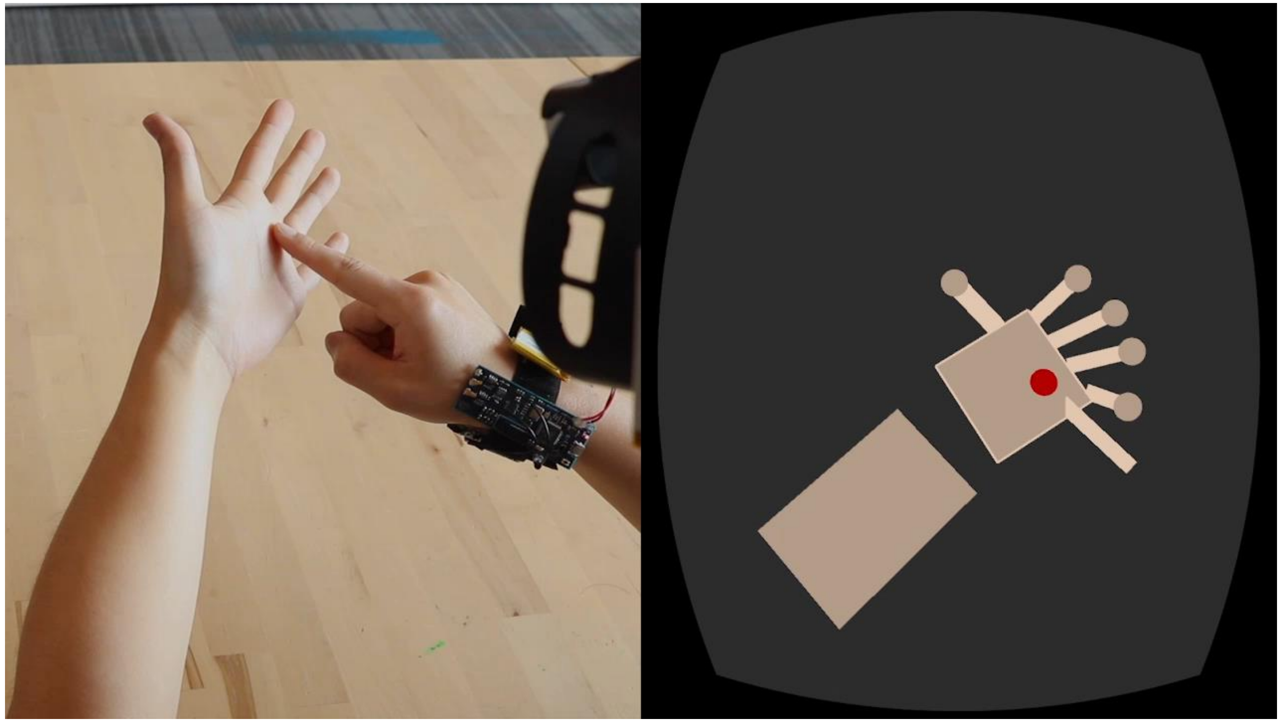
# Fusion of Optical and Inertial Sensing




Parizi, F. S., Kienzle, W., Whitmire, E., Gupta, A., and Benko, H. (2021) RotoWrist: Continuous Infrared Wrist Angle Tracking using a Wristband. In *Proceedings of ACM VRST '21.*

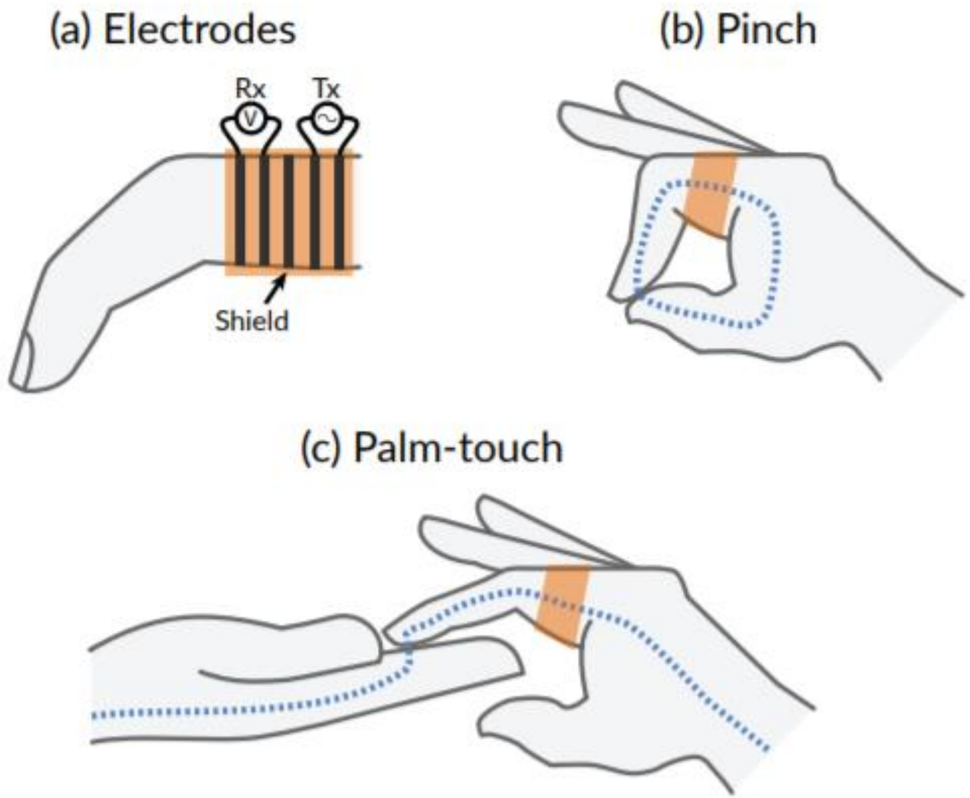
# Active Electrical Sensing of Touch and Contact

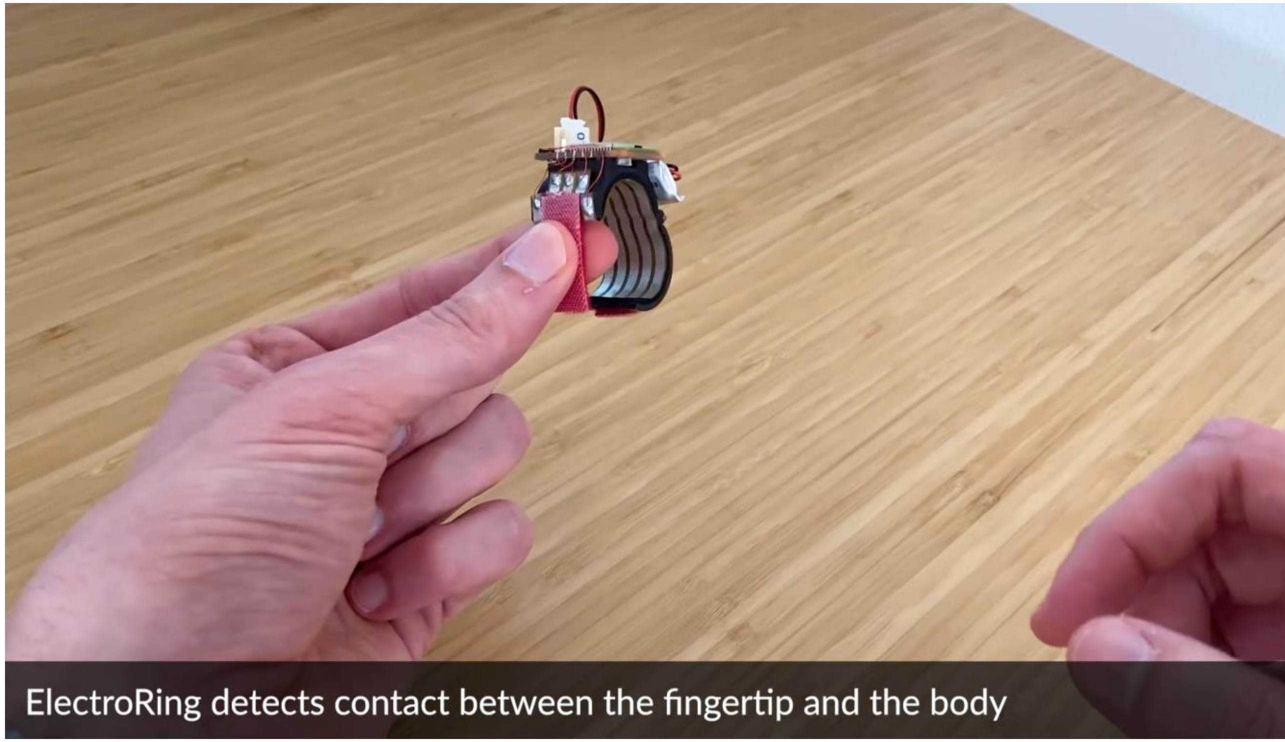


Zhang, Y., Kienzle, W., Ma, Y., Ng, S. S., Harrison, C., and Benko, H. (2019). ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. *In Proc. of ACM UIST 2019.*

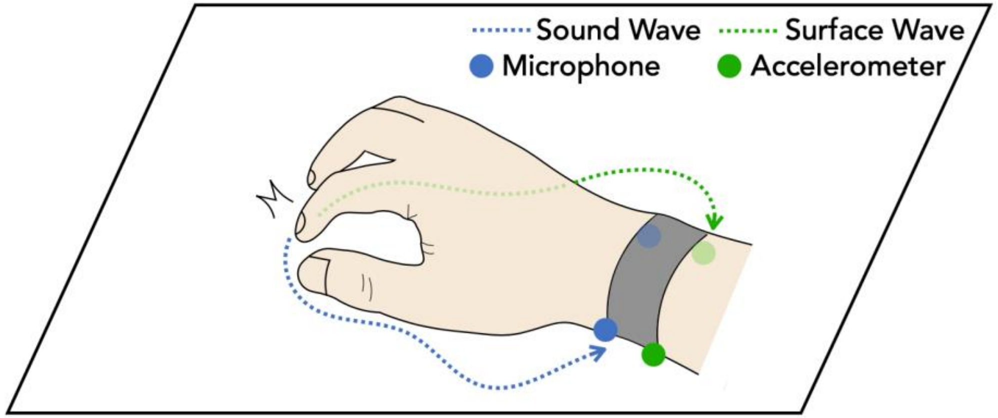
# Active Electrical Sensing of Touch and Contact

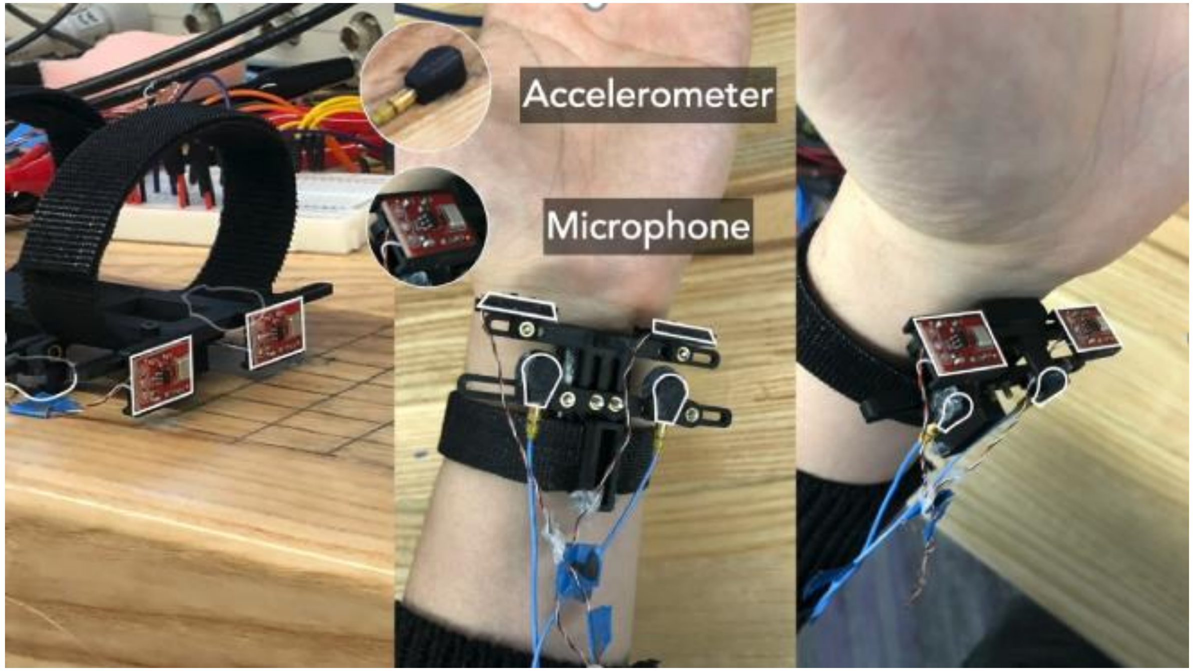


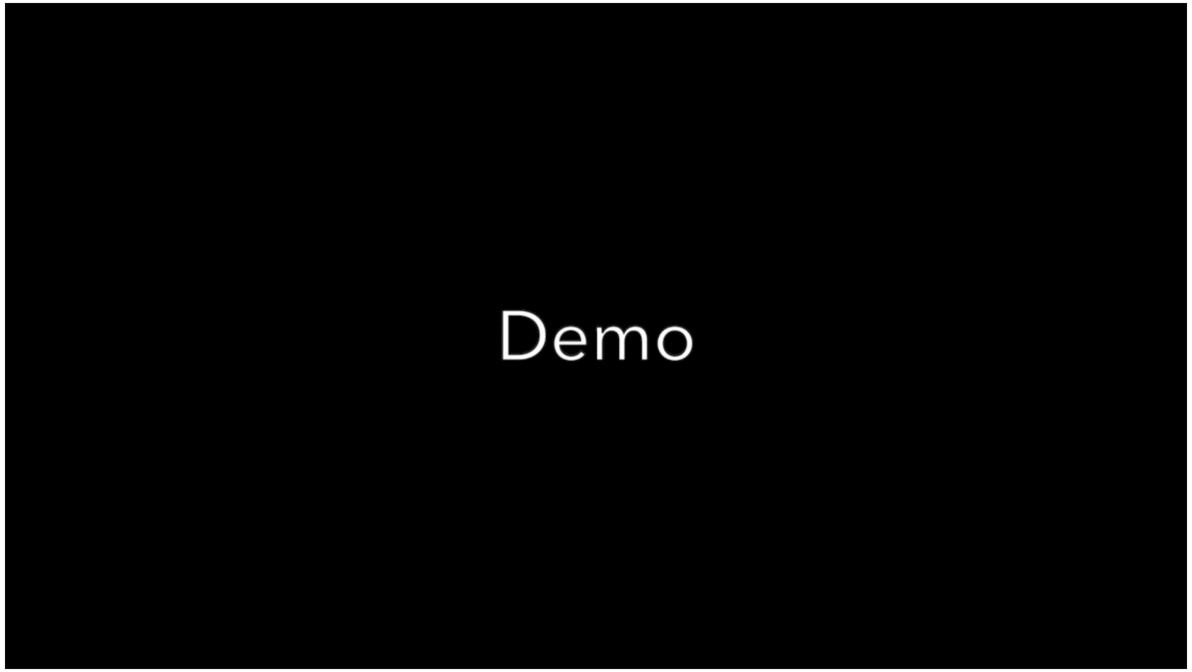




Kienzle, W., Whitmire, E., Rittaler, C., and Benko, H. (2021) ElectroRing: Subtle Pinch and Touch Detection with a Ring. In *Proceedings of ACM CHI '21.*

# Acoustic Touch Sensing on Any Surface

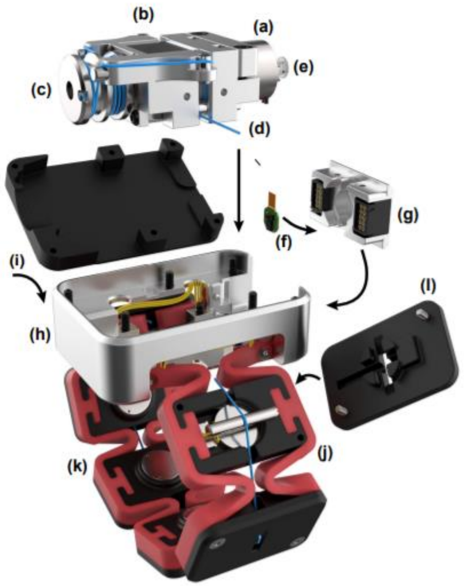


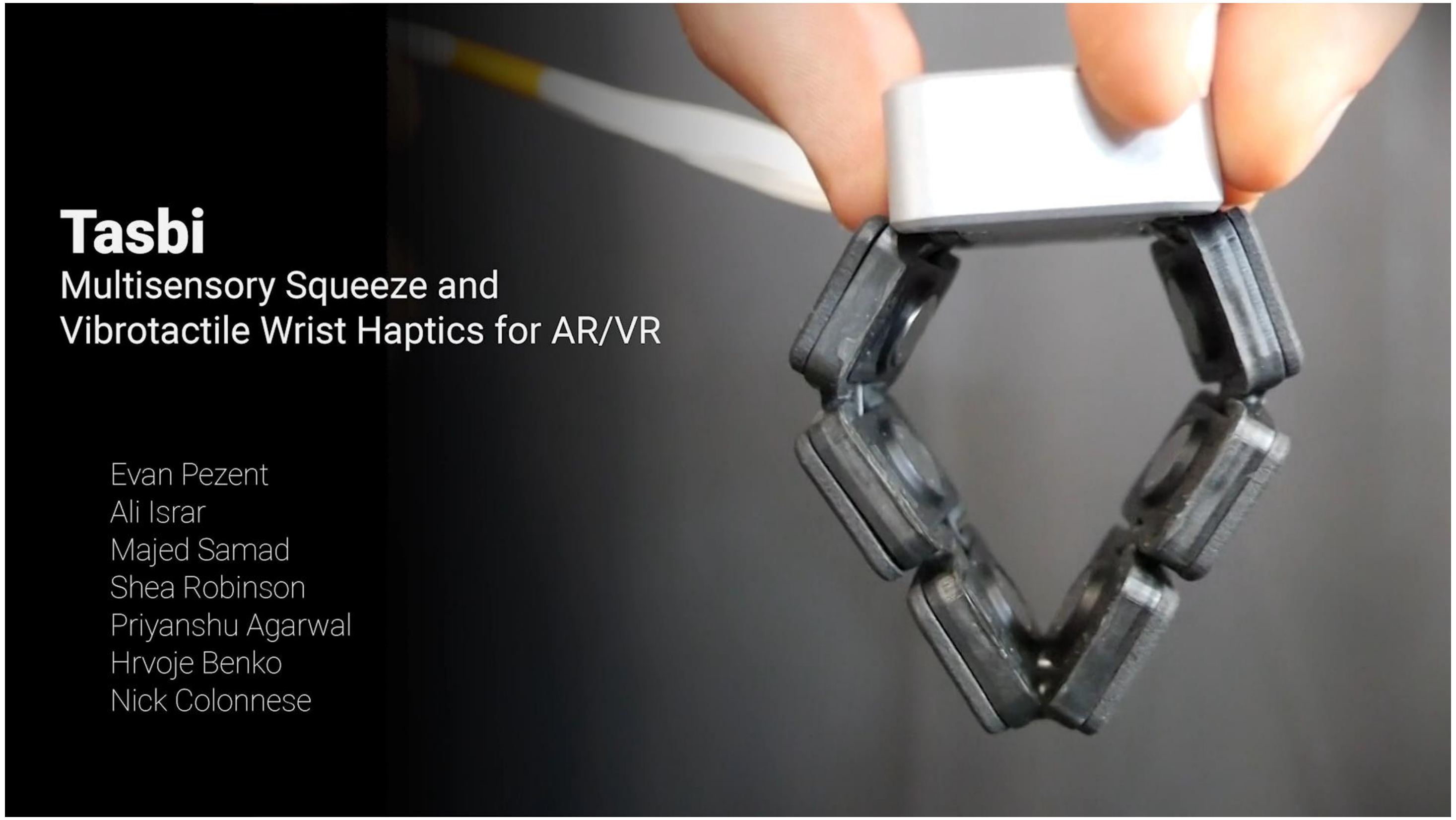




Gong, J., Gupta, A. and Benko, H. (2020). Acustico: Surface Tap Detection and Localization using Wrist-based Acoustic TDOA Sensing. *In Proceedings of ACM UIST '20.*

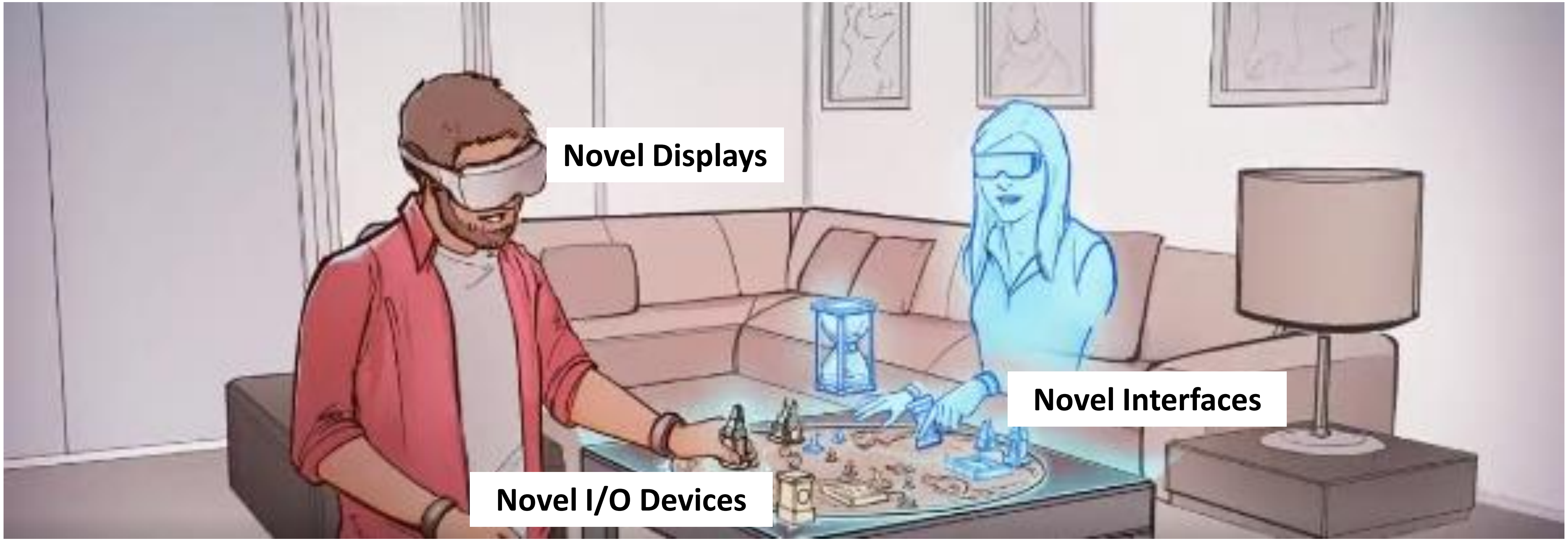
# Wrist Haptics




Penzent, E., Israr, A., Samad, M., Robinson, S., Agrawal, P., Benko, H., and Colonnese, N. (2019). Tasbi: Multisensory Squeeze and Vibrotactile Wrist Haptics for Augmented and Virtual Reality. *In Proc. of World Haptics Conference (WHC 2019).*


Haptic Gloves


Novel Displays
Novel Interfaces
Novel I/O Devices

Magic of MR interactions happens when they are tightly coupled to the user’s ~~environment.~~

context

# Context

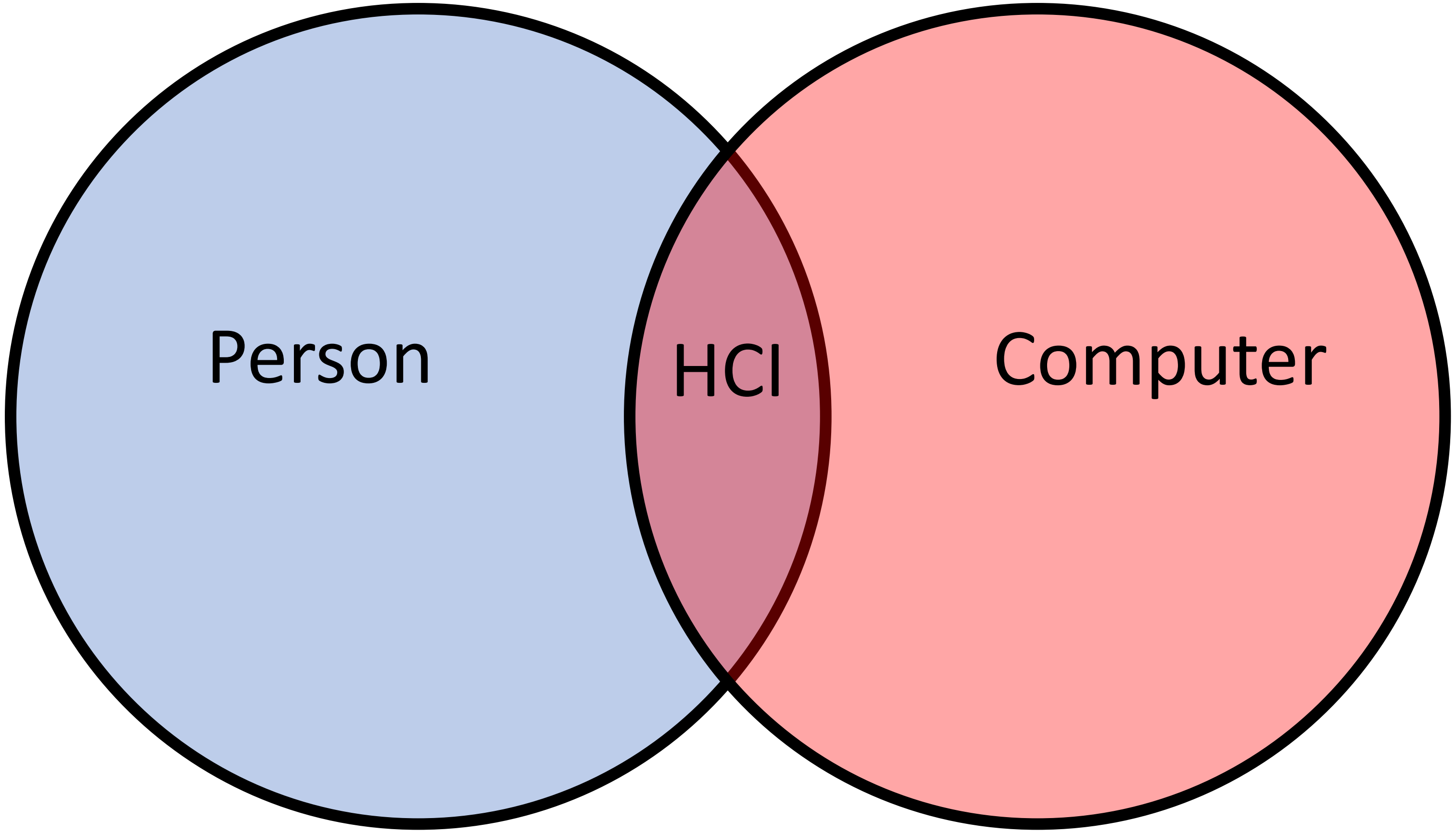
**environment**
(space geometry, object semantics, people around,…)

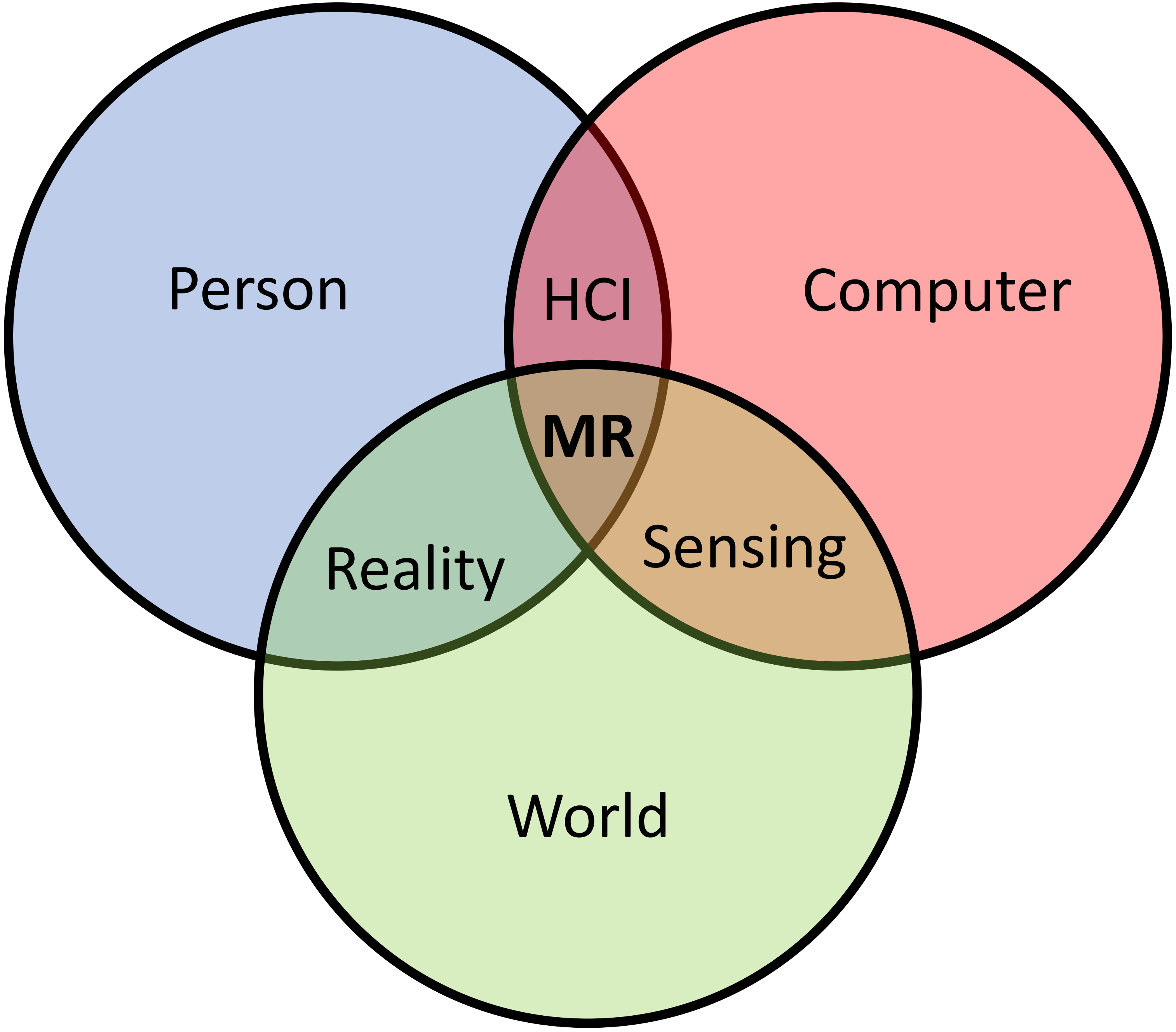
**task**
(communication, navigation, calendar,…)

**user actions**
(gestures, body pose, bio-signals,…)

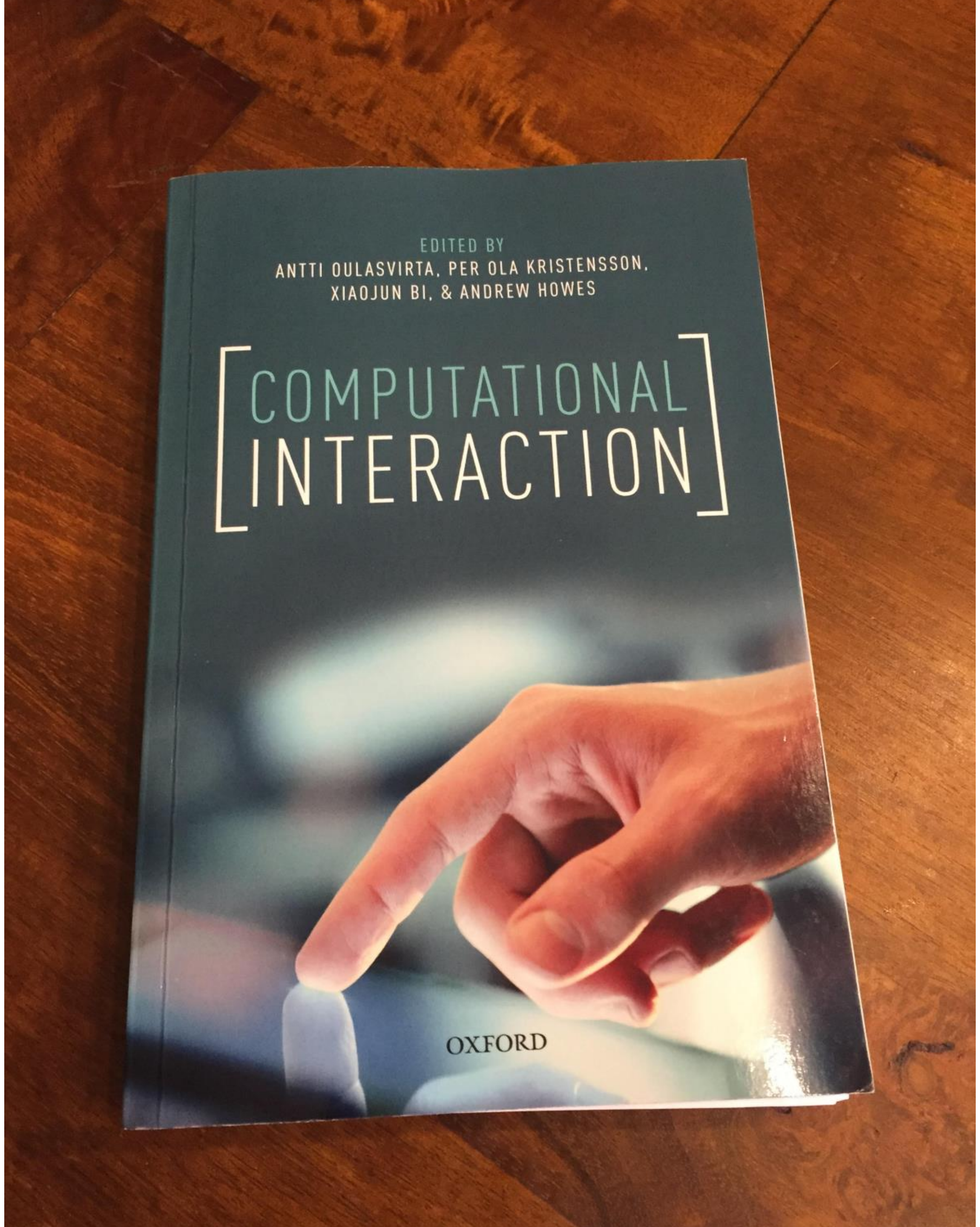
**user’s mental state**
(emotional, mental load, cognitive focus,…)

Context not known at design time.


Person
HCI
Computer

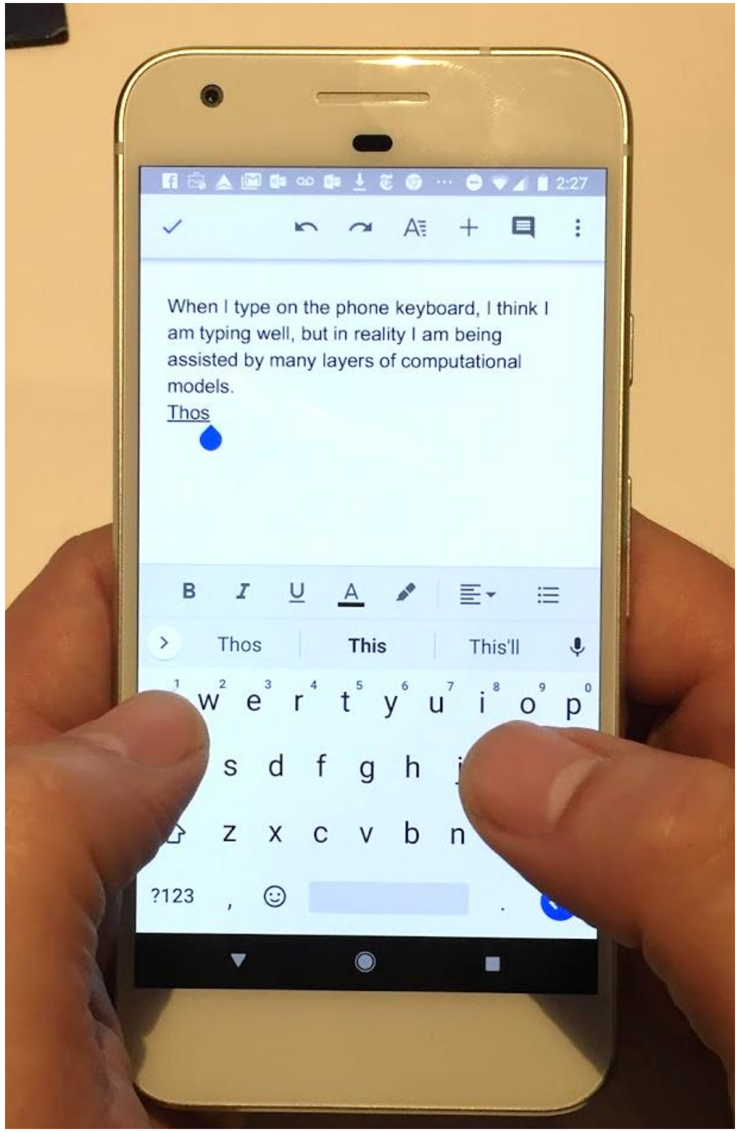

Person
HCI
Computer
MR
Reality
Sensing
World

How to deal with imprecise, noisy, but sensing-rich inputs?


EDITED BY
ANTTI OULASVIRTA, PER OLA KRISTENSSON,
XIAOJUN BI, & ANDREW HOWES
COMPUTATIONAL
INTERACTION
OXFORD

# Can you type on a phone keyboard?

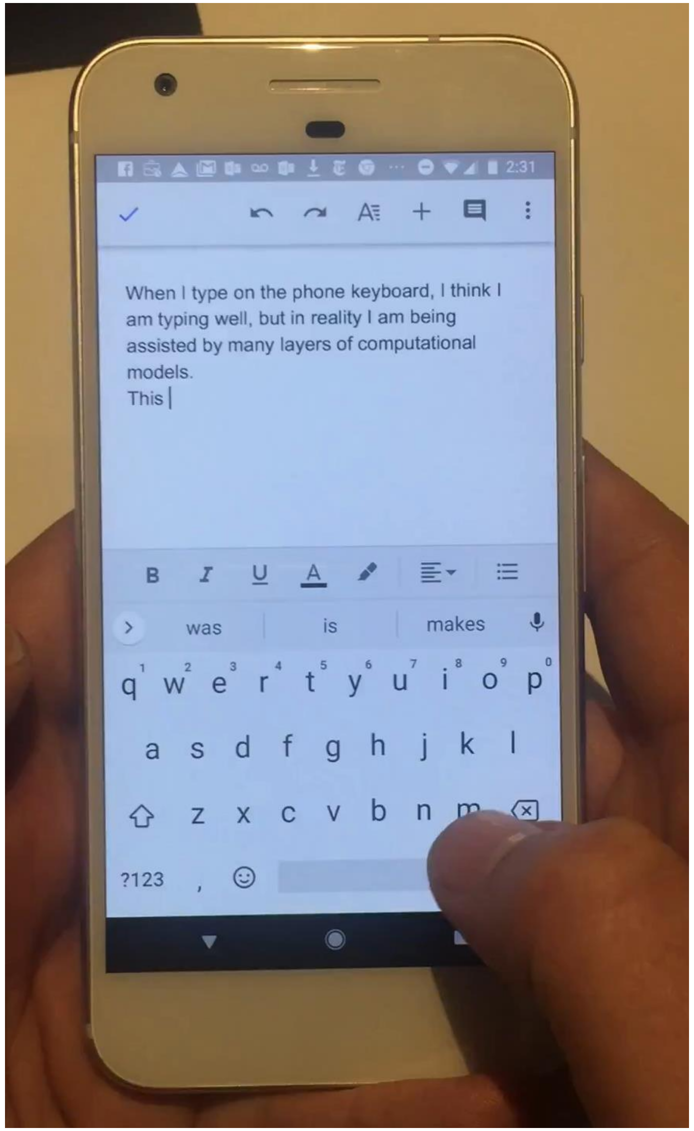
# Probabilistic Phone Touch Keyboard



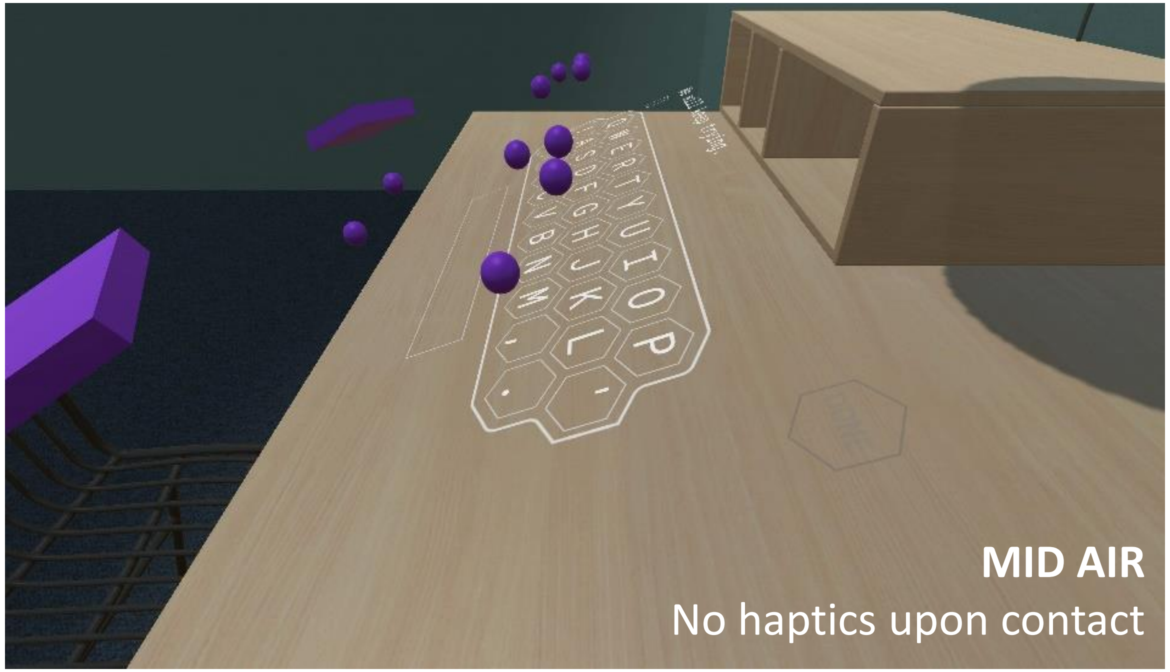

Keyboard geometry model
+ Touch precision model
+ Dictionary model
+ Language model

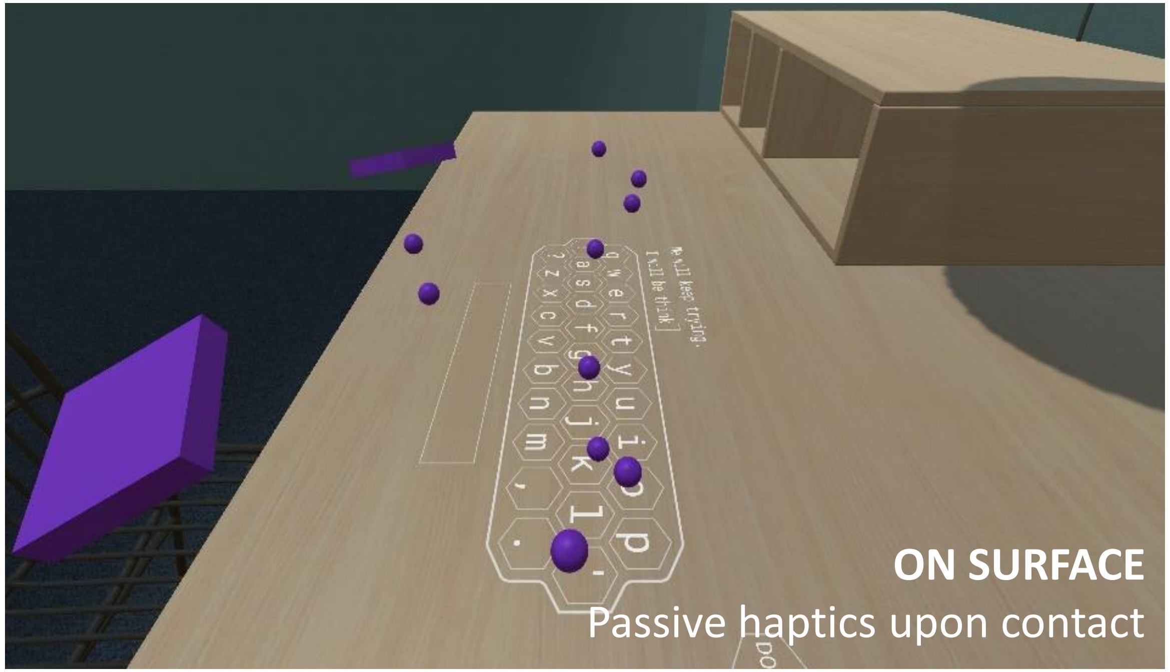
+ N-best list UI for error correction

+ Gesture model



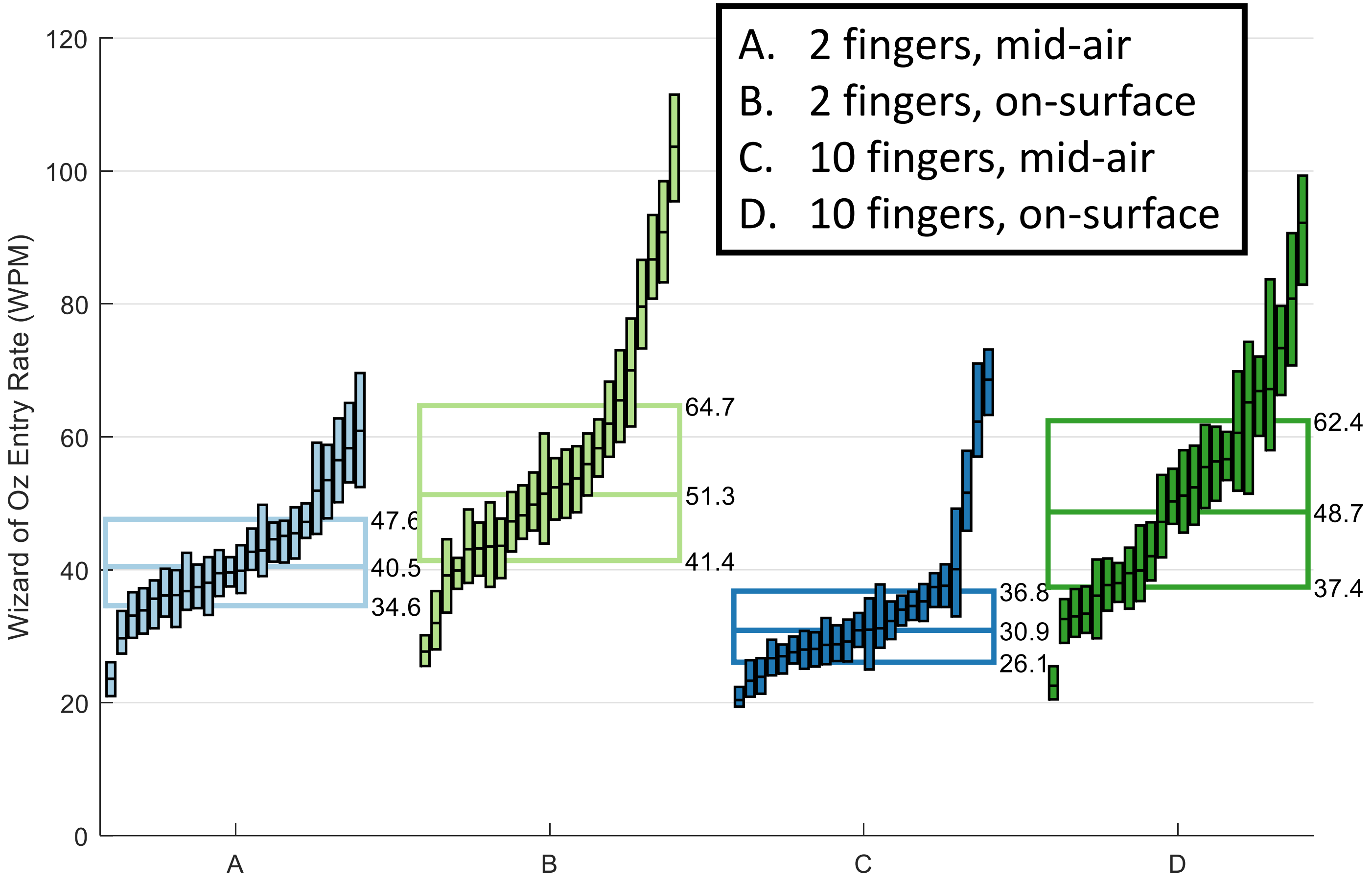
# Smart virtual keyboard can be better than a physical keyboard



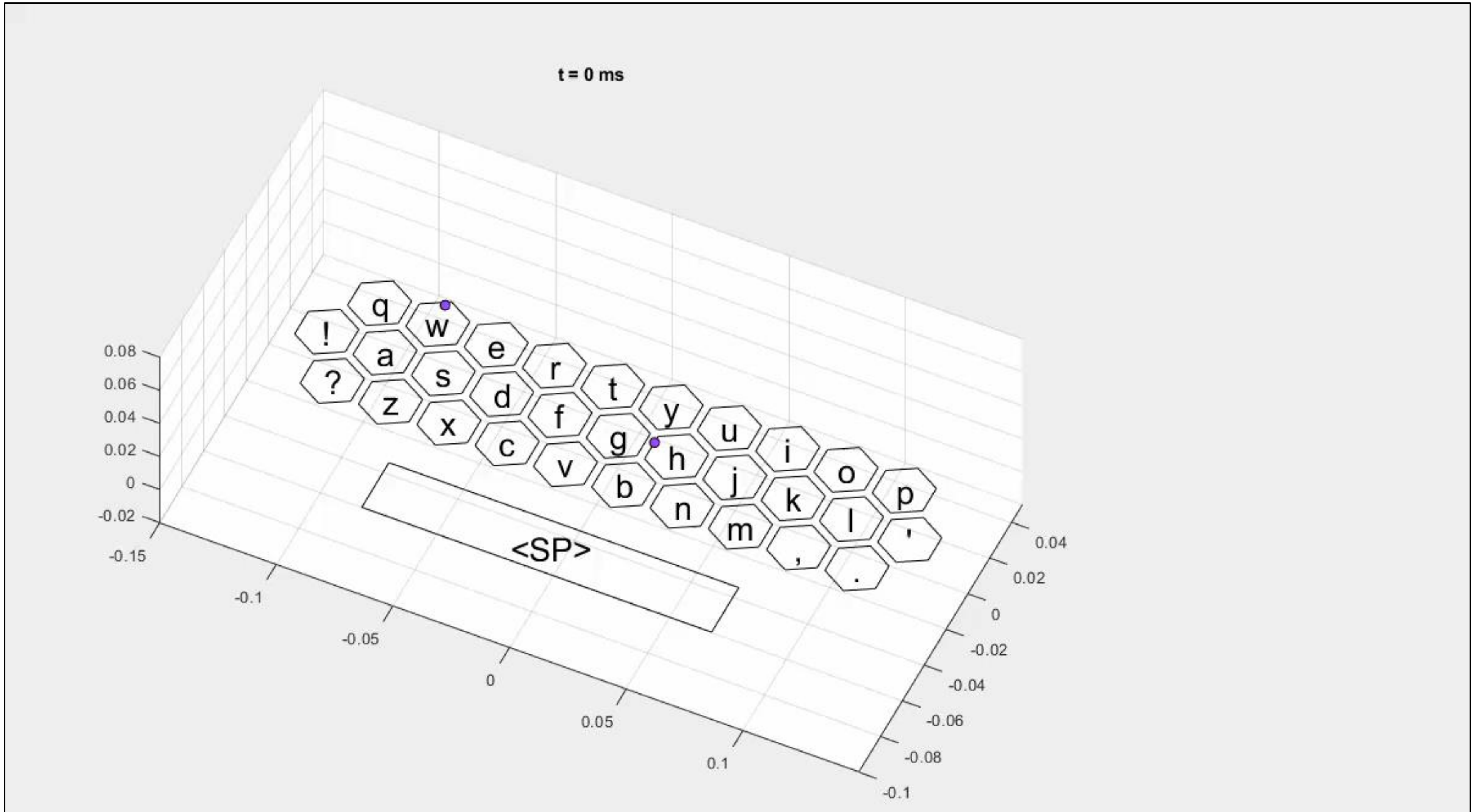

Dudley, J., Benko, H., Wigdor, D., and Kristensson, P.O. (2019). Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. *In Proc. of IEEE ISMAR 2019.*

# Entry Rate Results




*Plot shows participant $q_1$, median and $q_3$ (sorted by median) entry rates as well as lumped condition $q_1$, median and $q_3$ entry rates. Only entries where error rate < 10%.*

# 2 Finger VR Typing at >100 WPM




**Stimulus:** *How are things with you?*

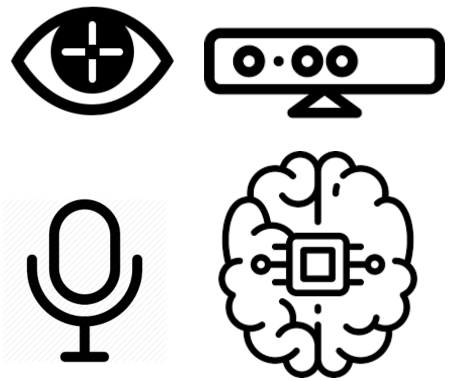
# Computational Approaches Needed

- Text entry
- Hand, body, touch input
- Object selection
- Multimodal fusion
- Layout optimizations
- Action recommendations
- Error mitigations
- Personalization


00:39
Tuesday
3
FaceTime
Calendar
Photos
Camera
Mail
Clock
Maps
Weather
Reminders
Notes
Stocks
Books
App Store
Podcasts
tv
TV
Health
Home
Wallet
Settings

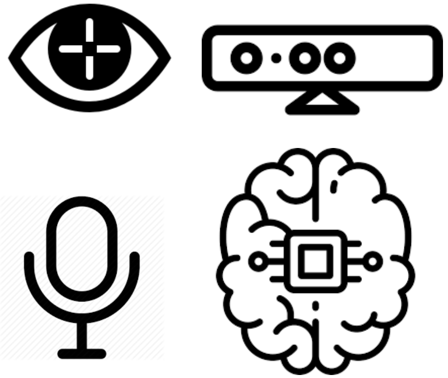
# MR Interaction Pipeline

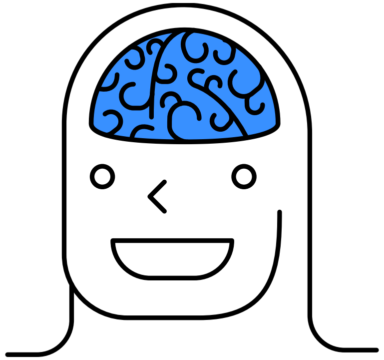
Understand Context



# MR Interaction Pipeline

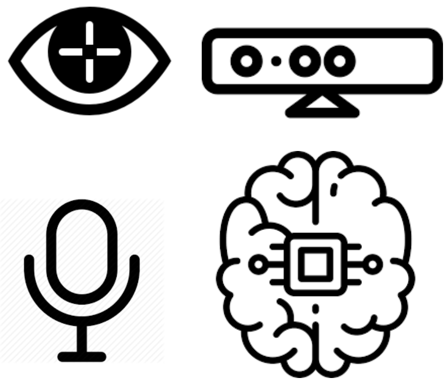
Understand Context → Infer Goals

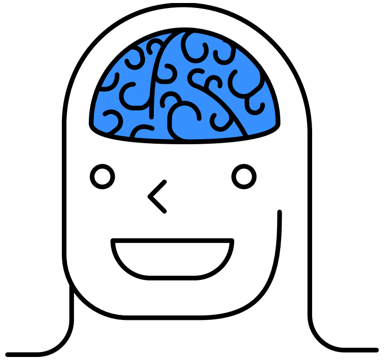


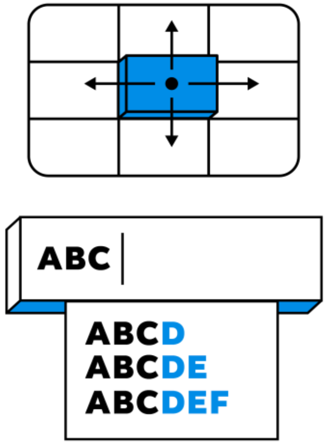


# MR Interaction Pipeline

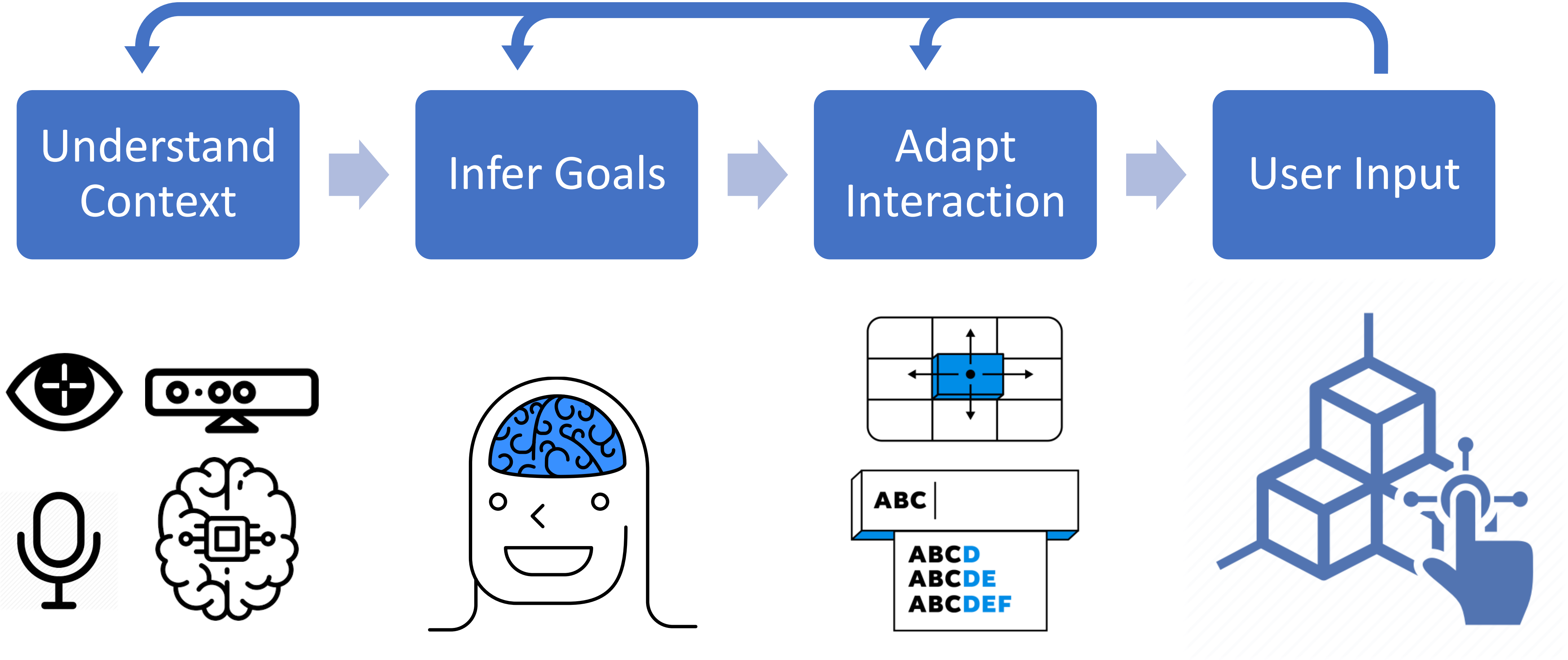
Understand Context → Infer Goals → Adapt Interaction

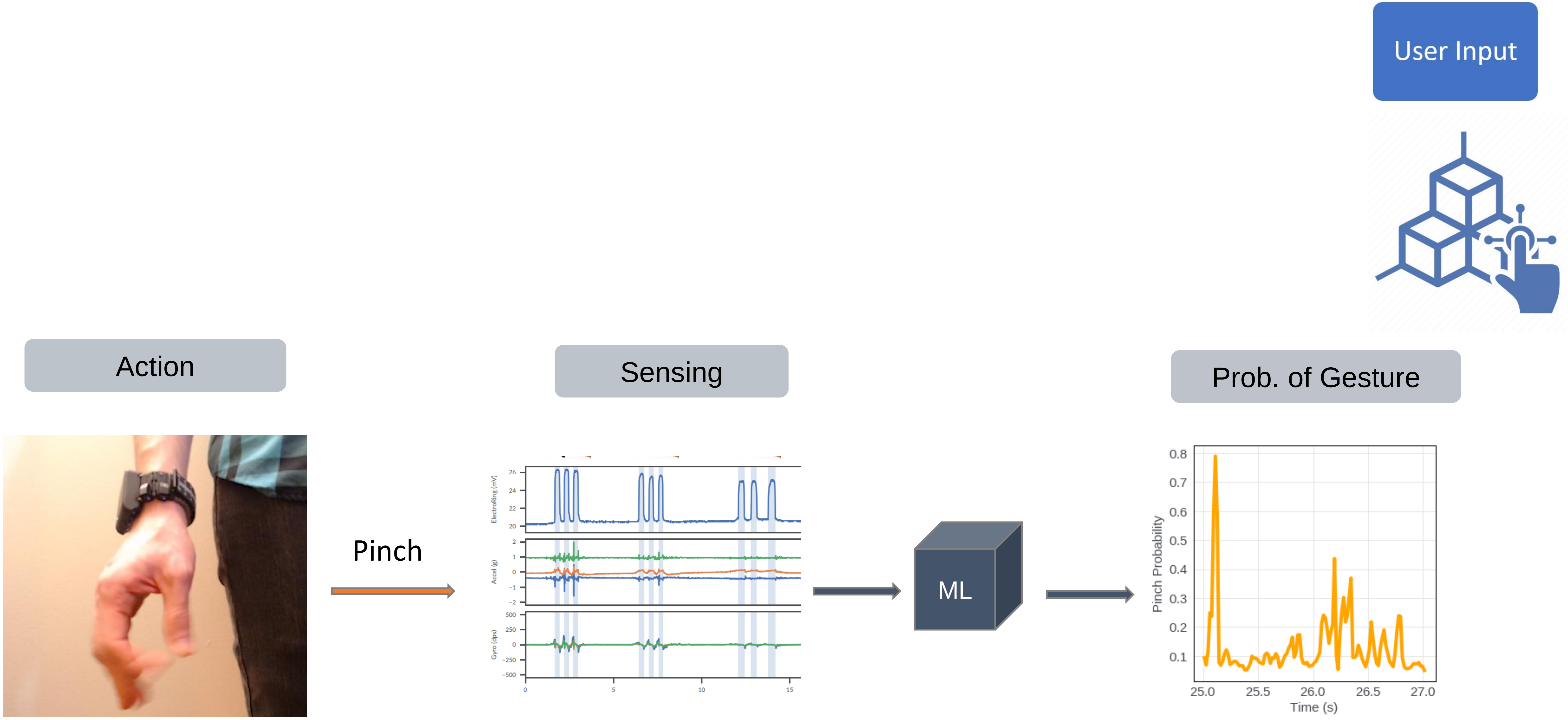






# MR Interaction Pipeline




User Input
Action
Pinch
Sensing
ML
Prob. of Gesture
Pinch Probability
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
25.0
25.5
26.0
26.5
27.0
Time (s)

# Understand the Environment

Understand Context

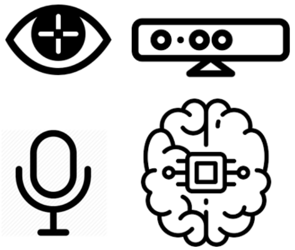
***Where am I? What is around me?***

*Project Aria* - Research glasses device to help build the 3D map of the world together with all the objects, people and their relationships

https://about.facebook.com/realitylabs/projectaria/

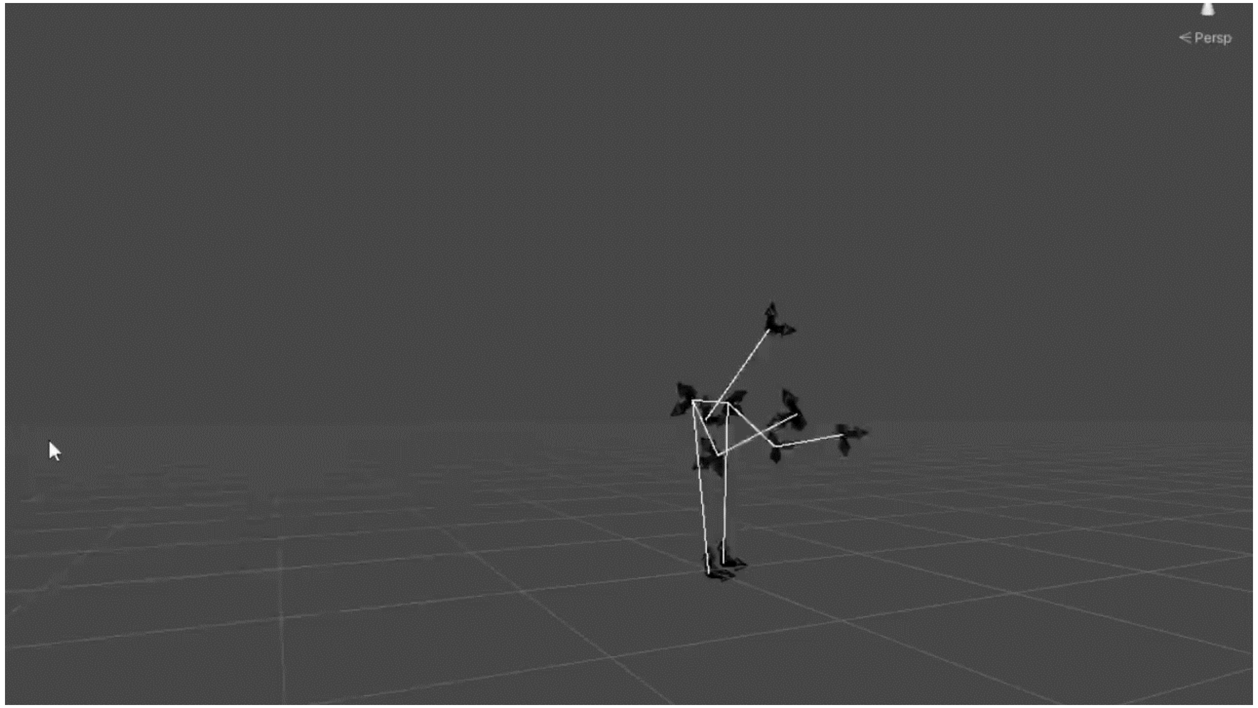
# Inferring user actions from sparse sensors

Understand Context



***What am I doing?***

Aria headset is doing SLAM + 2 wristbands with IMUs only are providing full upper-body pose and helping with action recognition



# Intent to Interact using gaze dynamics

Infer Goals

***What am I trying to accomplish?***

Predict user's intent to interact with a virtual object using eye-tracking and pupillometry features alone. (AUC-ROC = 0.77, chance 0.5)

These features are consistent across individuals.

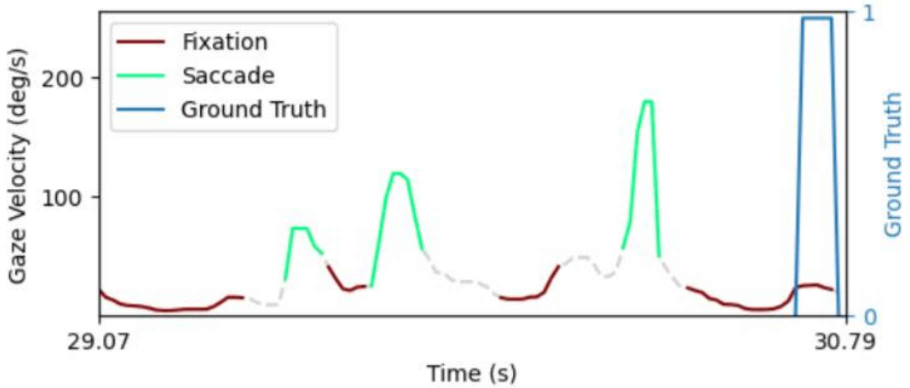


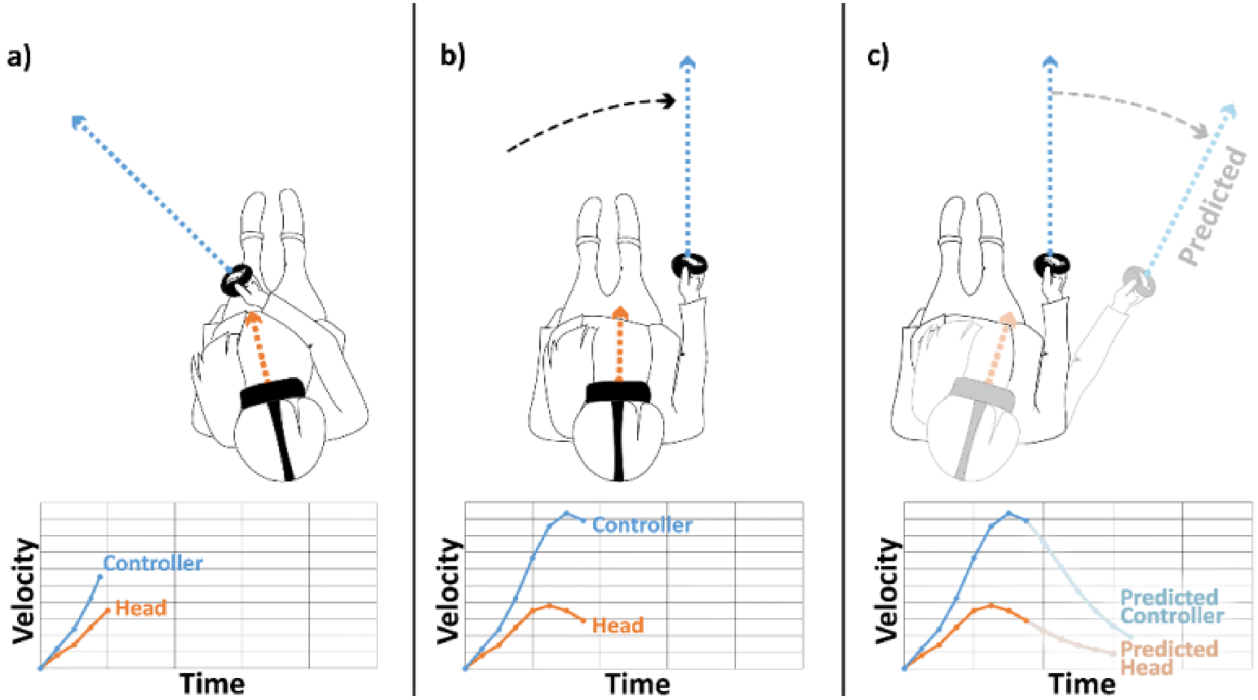

**Table 1: Features selected for model evaluation and the number of participants in which they were retained.**

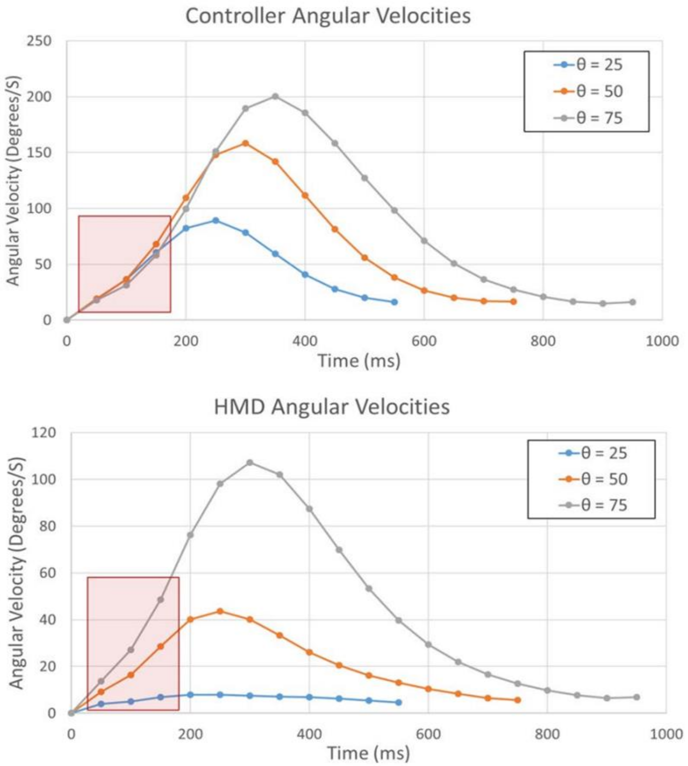
| Feature | Count | Feature | Count | Feature | Count |
|---|---|---|---|---|---|
| Fixation Detection | 13 (87%) | Std. Dev. of Vert. Gaze during Saccade | 9 (60%) | Saccade Duration | 8 (53%) |
| Gaze Vel. | 12 (80%) | Kurtosis of Vel. during Saccade | 9 (60%) | K Coefficient | 8 (53%) |
| Average Vel. during Fixation | 10 (67%) | Skew of Vel. during Saccade | 9 (60%) | Std. Dev. of Vel. during Saccade | 8 (53%) |
| Skew of Horiz. Accel. during Saccade | 10 (67%) | Skew of Horiz. Vel. during Saccade | 9 (60%) | Ang. Distance from Prev. Saccade | 8 (53%) |

David-John, B., Peacock, C., Murdison, T. S., Benko, H., Jonker, T. Towards gaze-based prediction of the intent to interact in virtual reality. ETRA 2021

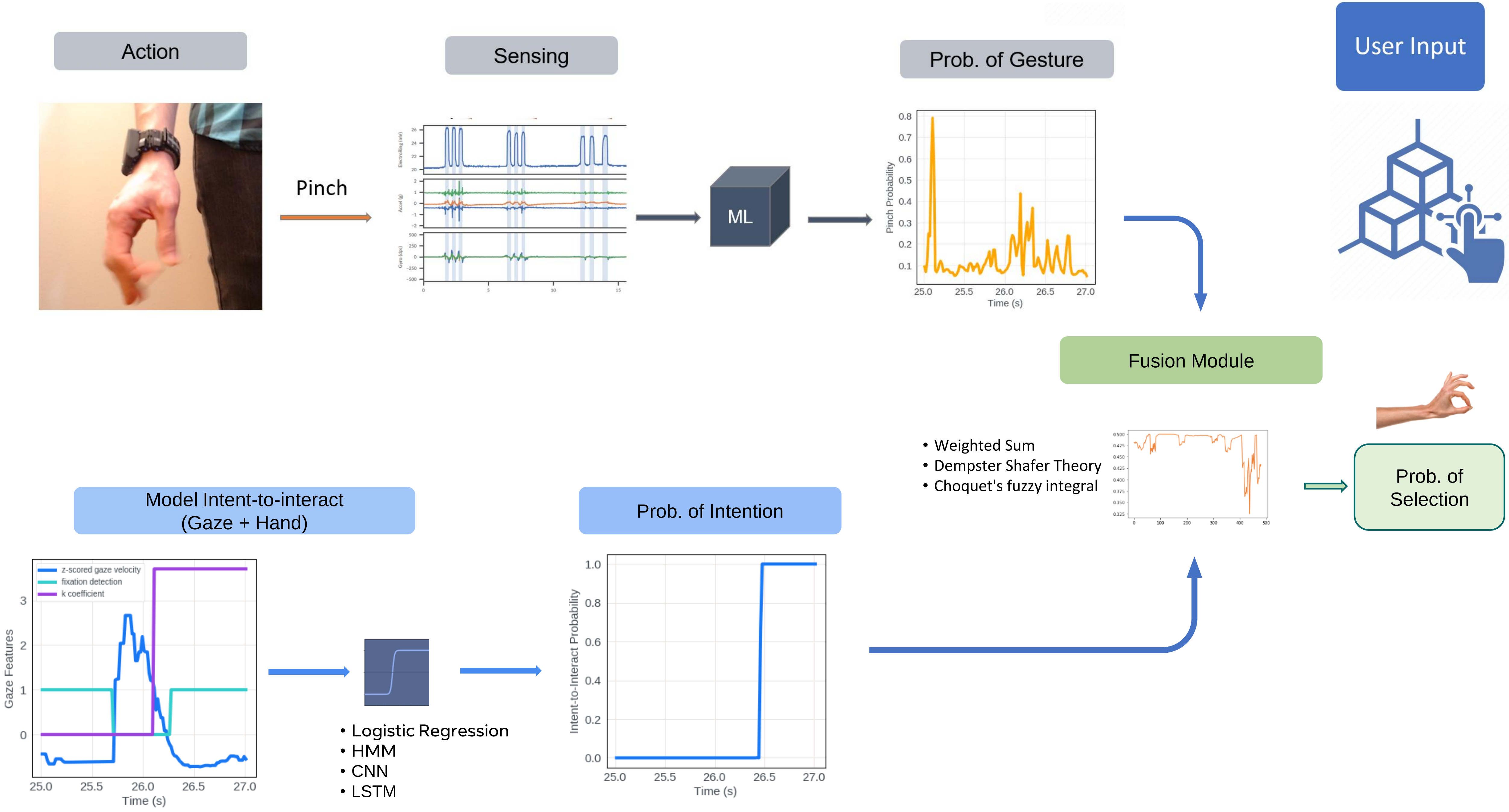
# Predictive Pointing

Infer Goals



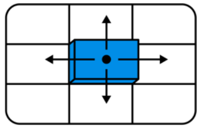

Henrikson, R., Grossman, T., Trowbridge, S., Wigdor, D., and Benko, H. (2020). Head-Coupled Kinematic Template Matching: A Prediction Model for Ray Pointing in VR. *In Proceedings of ACM CHI '20.*

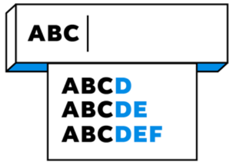


Action
Pinch
Sensing
ML
Prob. of Gesture
Pinch Probability
Time (s)
User Input
Fusion Module
• Weighted Sum
• Dempster Shafer Theory
• Choquet's fuzzy integral
Prob. of Selection
Model Intent-to-interact
(Gaze + Hand)
z-scored gaze velocity
fixation detection
k coefficient
Gaze Features
Time (s)
• Logistic Regression
• HMM
• CNN
• LSTM
Prob. of Intention
Intent-to-Interact Probability
Time (s)

# Interface Adaptation to Minimize Noise

Adapt Interaction






No adaptation – Raw gaze highlighting

Adaptation based on I2I model


David-John, B., Peacock, C., Murdison, T. S., Benko, H., Jonker, T.
Towards gaze-based prediction of the intent to interact in virtual reality. ETRA 2021

# Optimizing the Timing of Intelligent Suggestion in Virtual Reality


Difeng Yu
University of Melbourne
Melbourne, VIC, Australia

Ruta Desai
Reality Labs Research
Seattle, WA, USA

Ting Zhang
Reality Labs Research
Seattle, WA, USA

Hrvoje Benko
Reality Labs Research
Seattle, WA, USA

Tanya R. Jonker
Reality Labs Research
Seattle, WA, USA

Aakar Gupta
Reality Labs Research
Seattle, WA, USA

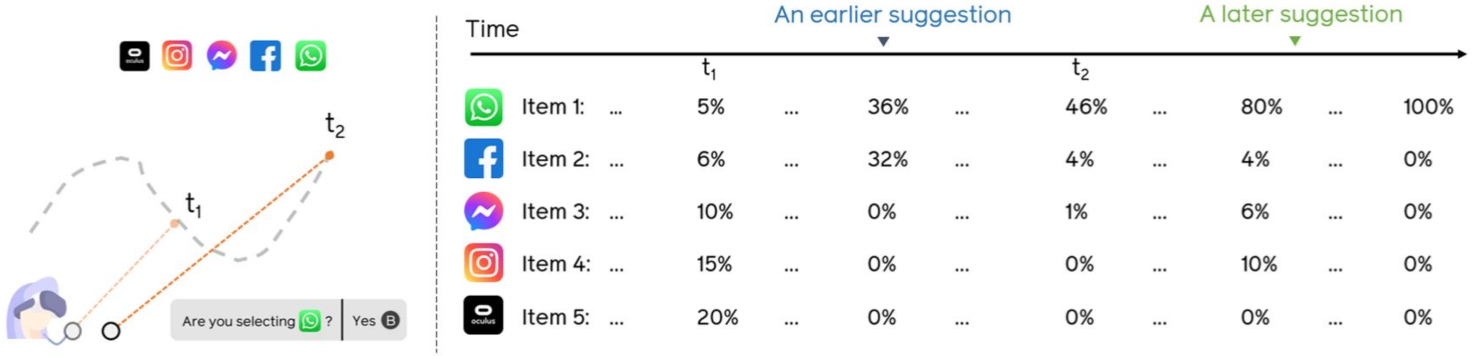




**Figure 1: An overview of the intelligent suggestion timing problem. While a user is attempting to select an icon in virtual reality, a target prediction model could be continuously estimating the likelihood that the user will select each icon (e.g., at timestamp $t_x$ and $t_y$). Depending on the results of these estimations, a system could then display an intelligent suggestion to the user that highlights the most probable icon for them to select. This suggestion, for example, could enable them to select an icon using a simple click, so that the user does not need to manually point towards the icon. While such suggestions could improve the usability of intelligent user interfaces, it is currently unknown whether early suggestions, which could save the user time and effort but may be less accurate, or later suggestions, which could save less time and effort but may be more accurate, are more beneficial for users.**

## ABSTRACT


Intelligent suggestion techniques can enable low-friction selection-based input within virtual or augmented reality (VR/AR) systems. Such techniques leverage probability estimates from a target prediction model to provide users with an easy-to-use method to select the most probable target in an environment. For example, a system and showed that it was both theoretically and empirically effective at determining the optimal timing for intelligent suggestions.

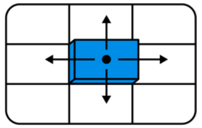

## CCS CONCEPTS

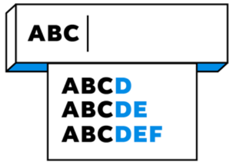
• **Human-centered computing → HCI theory, concepts and models**; *Mixed / augmented reality*; *Virtual reality*.


Yu, D., Desai, R., Zhang, T., Benko, H., Jonker, T.R., and Gupta, A. (2022). **Optimizing the Timing of Intelligent Suggestion in Virtual Reality.** In *Proceedings of ACM User Interface Systems and Technology (ACM UIST '22).*

# Many other adaptations possible

Adapt Interaction

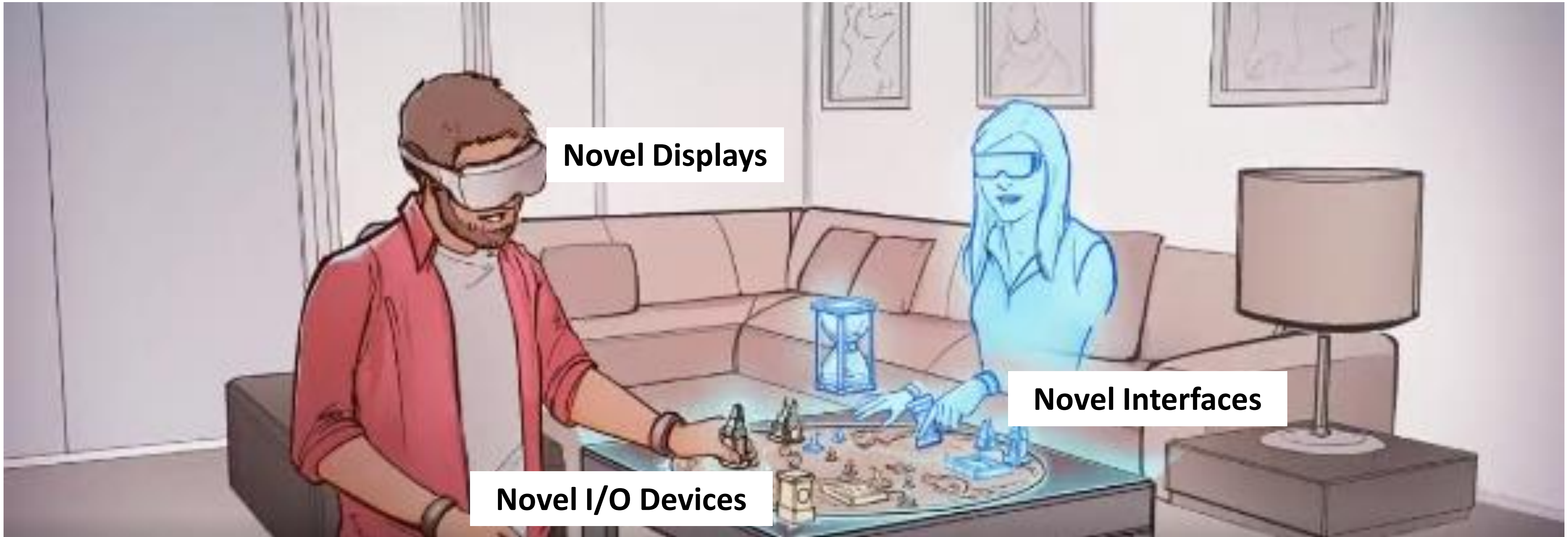





- Move content around
- Filter information content
- Correct user errors (auto-correct)
- Make it easier to complete an action (auto-complete)
- Suggest the next most likely action
- Provide optimal guidance for a task
- Change the level of detail presented


The Intelligent Click

| | | | |
|---|---|---|---|
| **Command Line Interfaces** (keyboard) | **Graphical User Interfaces** (mouse) | **Natural User Interfaces** (touch/gestures, tablets, smartphones) | **Mixed Reality Interfaces** |
| **1960s** | **1980s** | **2000s** | **2020s** |



# Thanks to all my collaborators!

**My teams are looking for interns for 2023!**

**Hrvoje Benko**

Director, Research Science
Reality Labs Research
benko@meta.com

REALITY LABS RESEARCH

Meta

# Summary

**Command Line Interfaces**
(keyboard)

**1960s**

**Graphical User Interfaces**
(mouse)

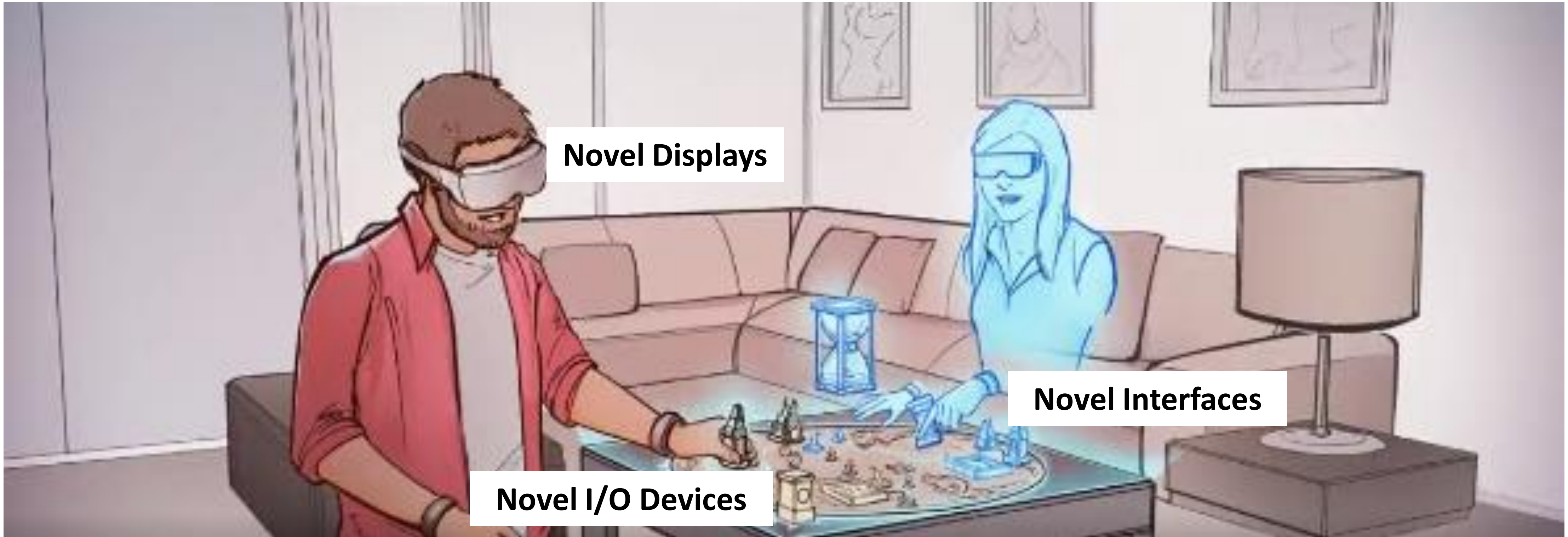
**1980s**

**Natural User Interfaces**
(touch/gestures, tablets, smartphones)

**2000s**

**Mixed Reality Interfaces**

**2020s**



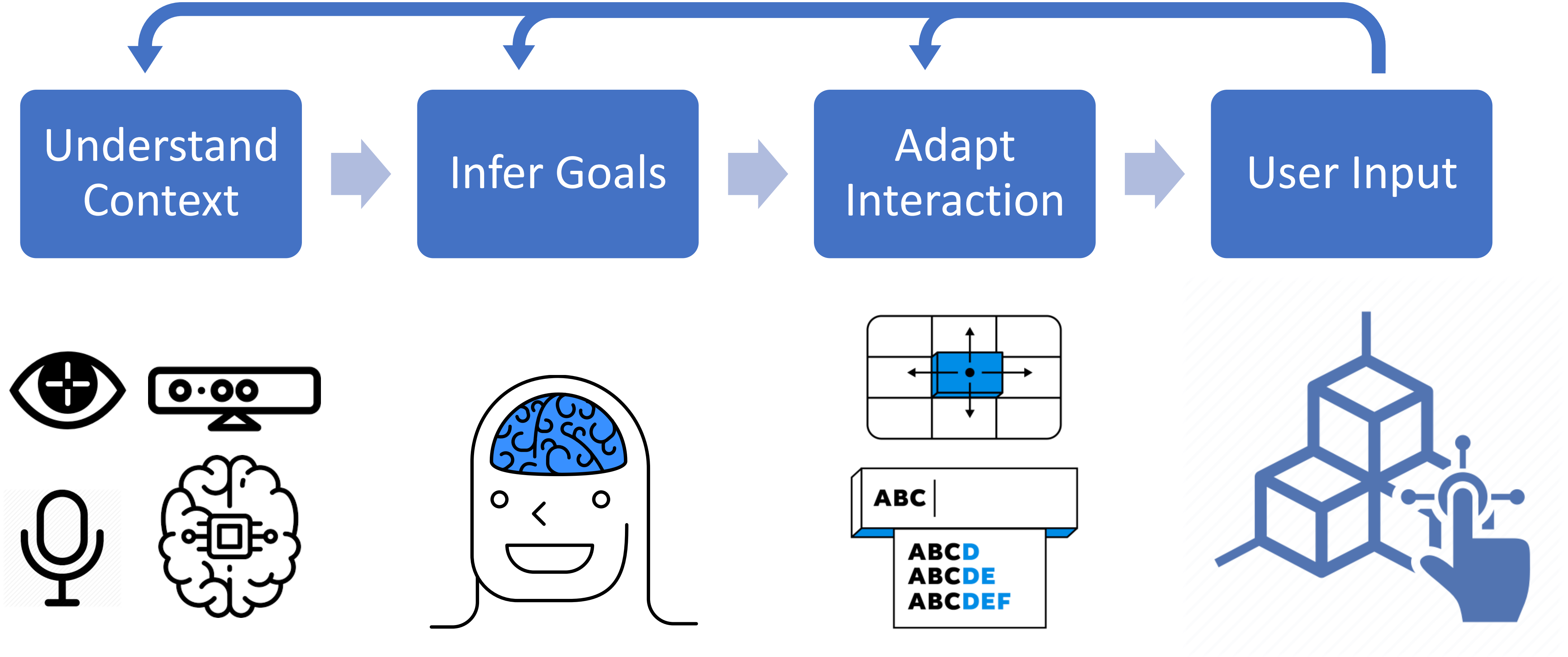
Compelling MR interactions are *adaptive* and *computational*.

# Wrist and hands are the key to subtle XR interactions

Design interactions that adapt to the user’s actions, the world around them, and the context of use.

Harness the computational methods to overcome uncertainty, scale, noise, and enable personalization.

# MR Interaction Pipeline



# Thanks to all my collaborators!

**My teams are looking for interns for 2023!**

**Hrvoje Benko**

Director, Research Science
Reality Labs Research
benko@meta.com

REALITY LABS RESEARCH

Meta

# Computational Interactions for the XR Future

REALITY LABS
RESEARCH
Meta

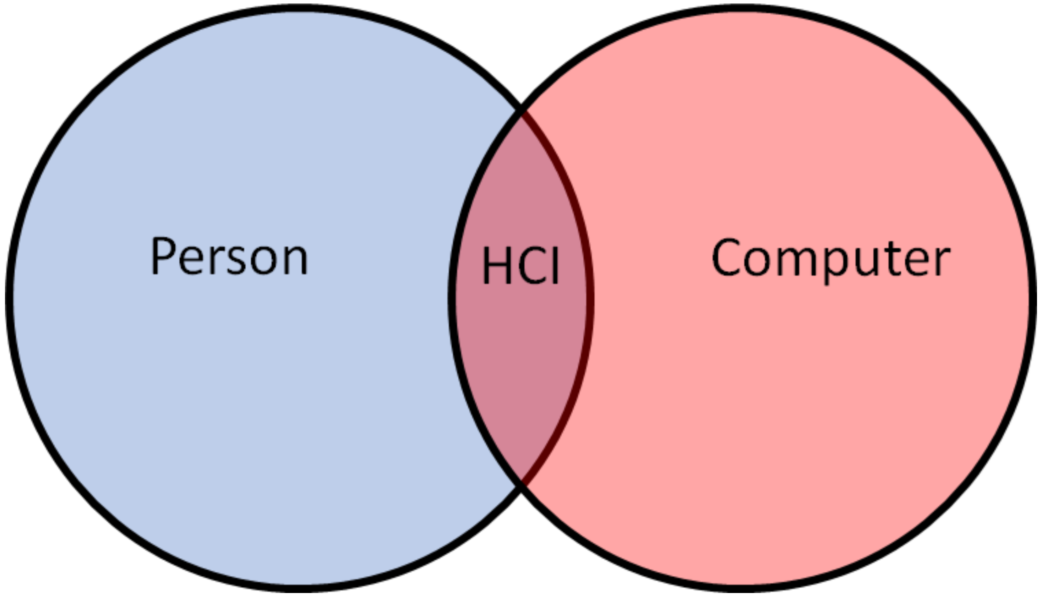
# Computational Interactions for the XR Future

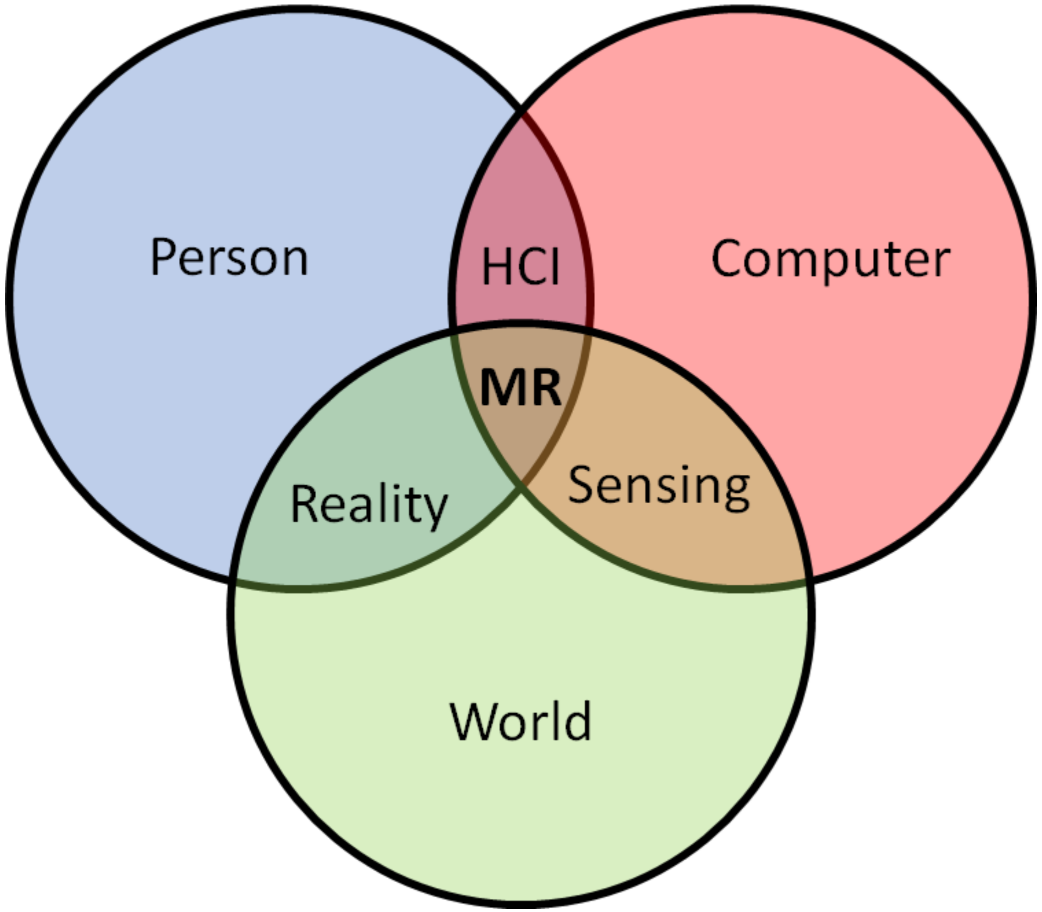
REALITY LABS RESEARCH

Meta

# Mission

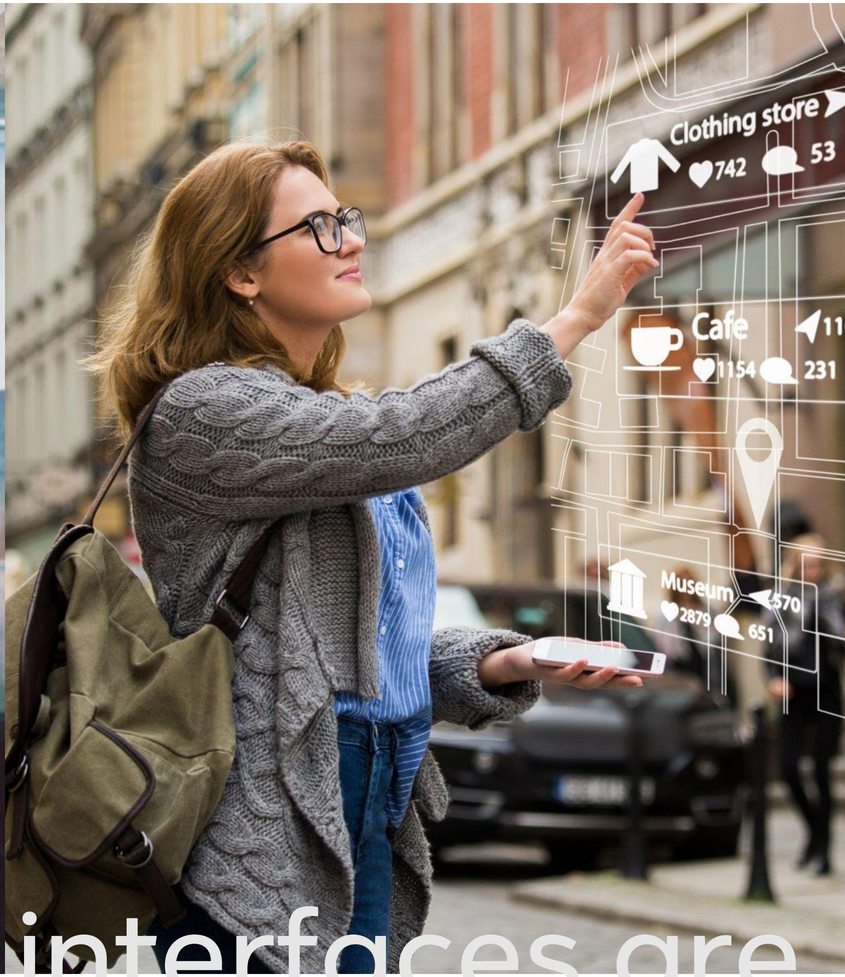
Solve the interaction problem for
future all-day wearable virtual and augmented
reality


Person
HCI
Computer


Person
HCI
Computer
MR
Reality
Sensing
World

Mixed reality is a new era of computing...



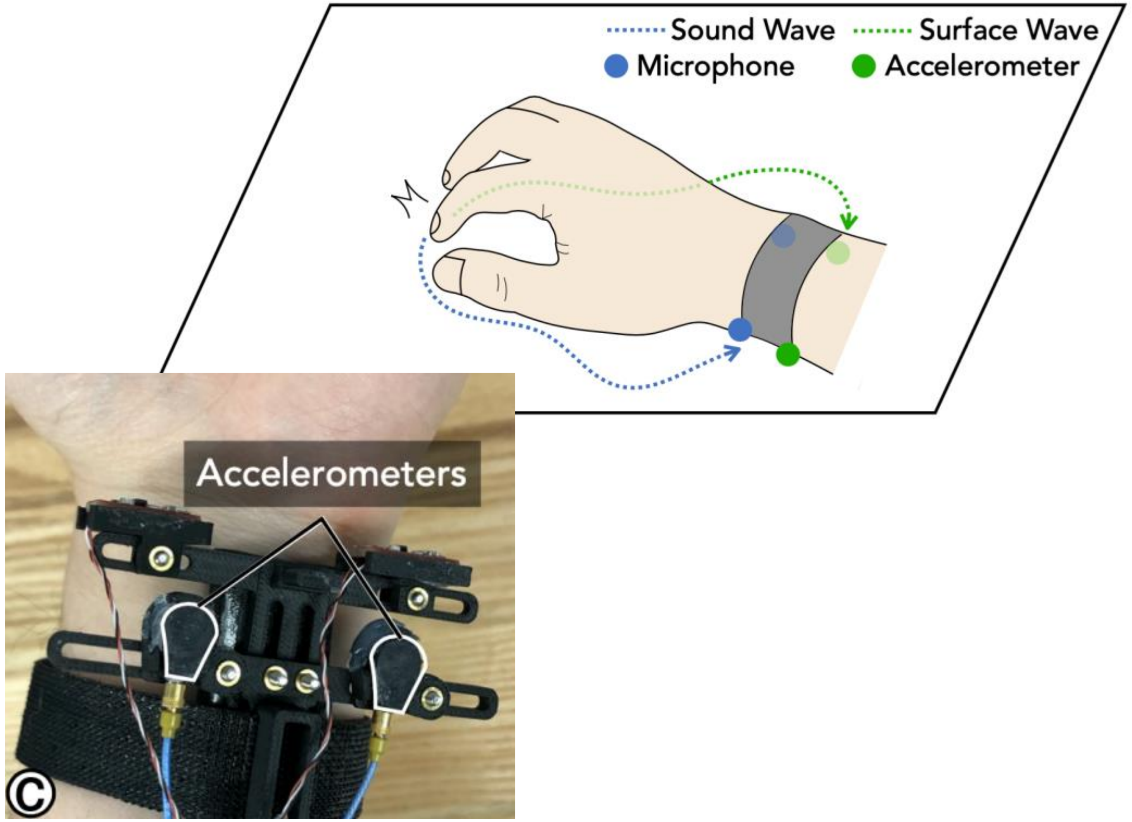



...contextually adaptive interfaces are essential

# Agenda

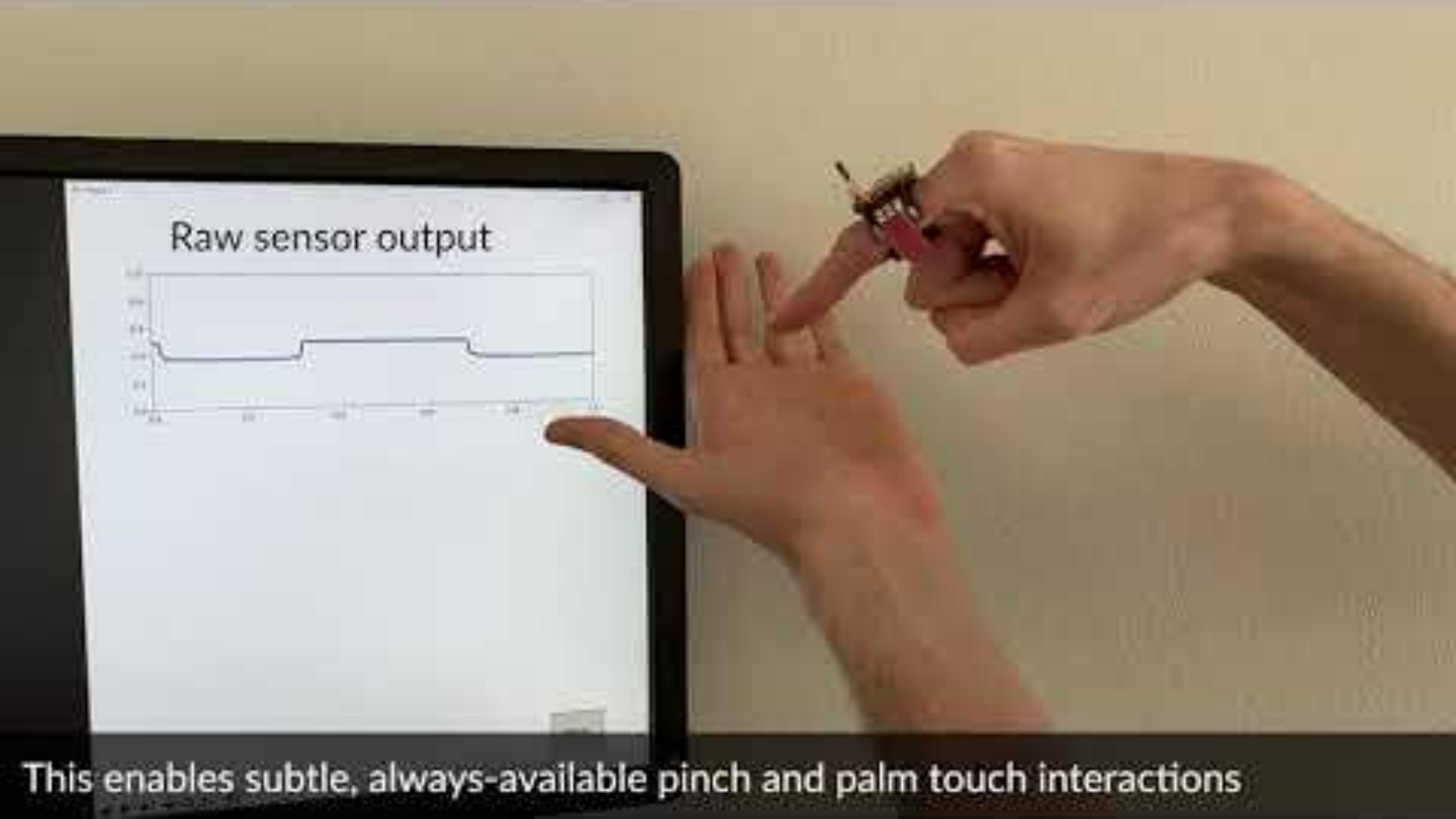
# Computational Input

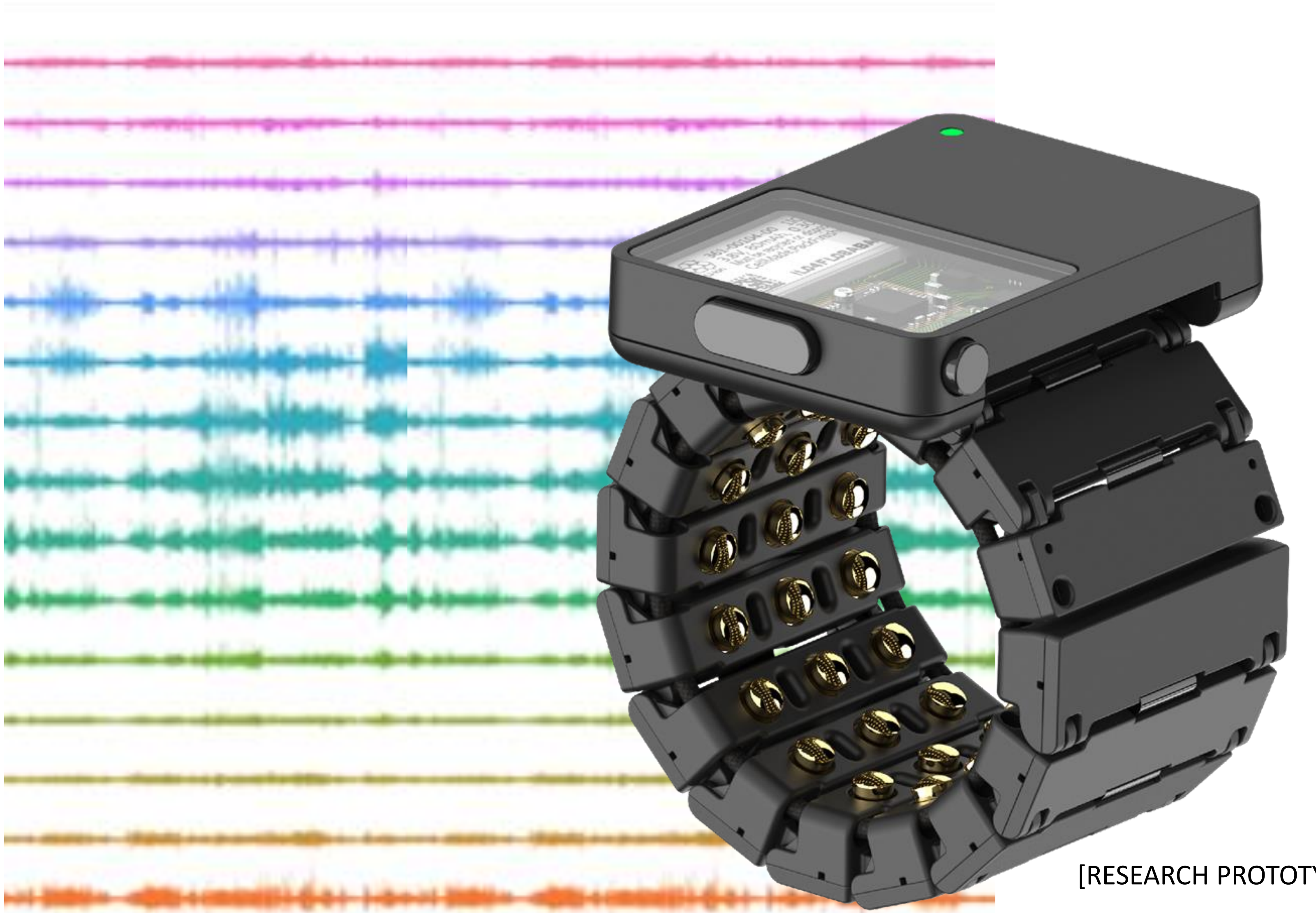
# Wearable Sensing



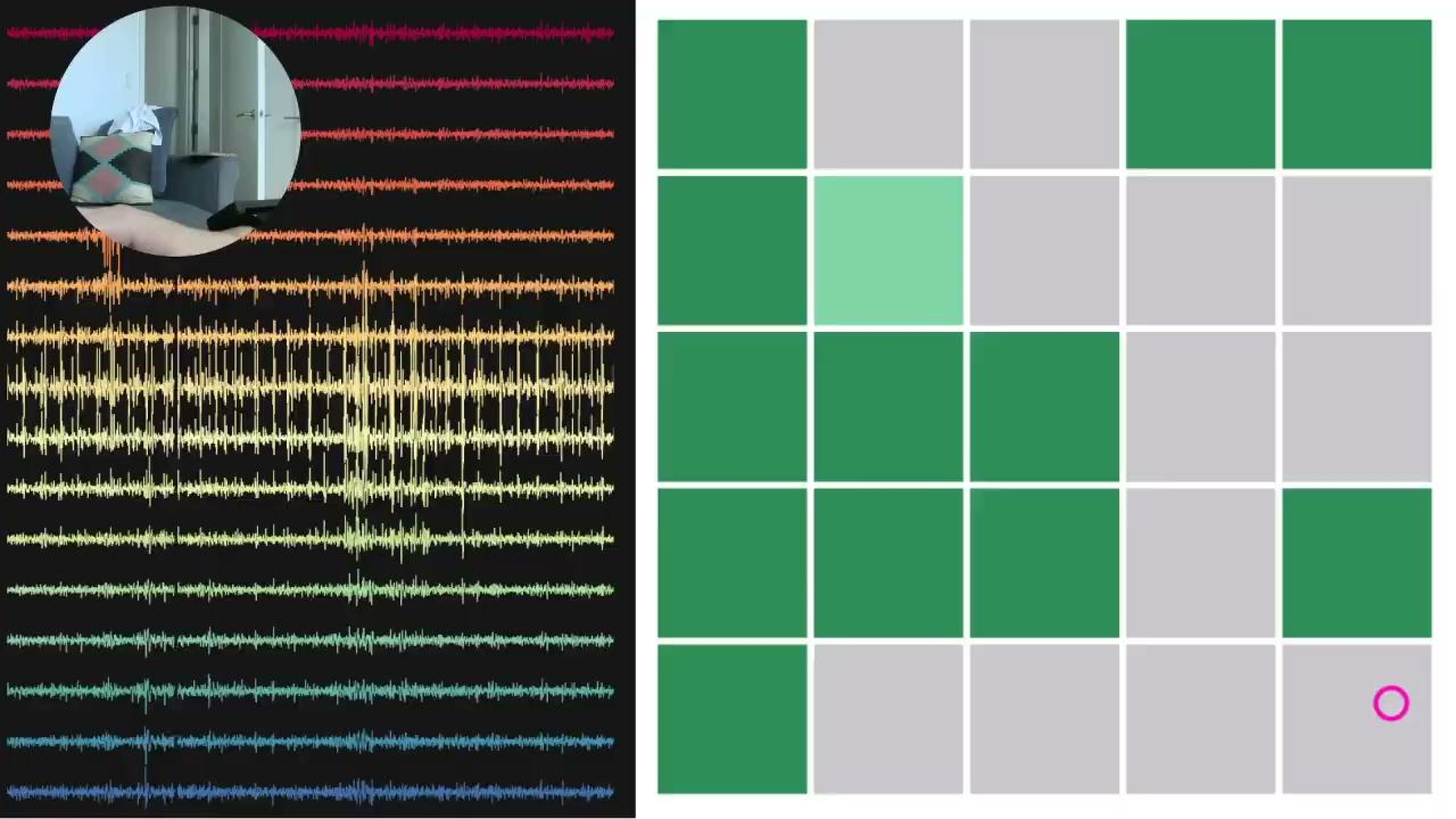


Acustico
Gong et al. *UIST 2020*

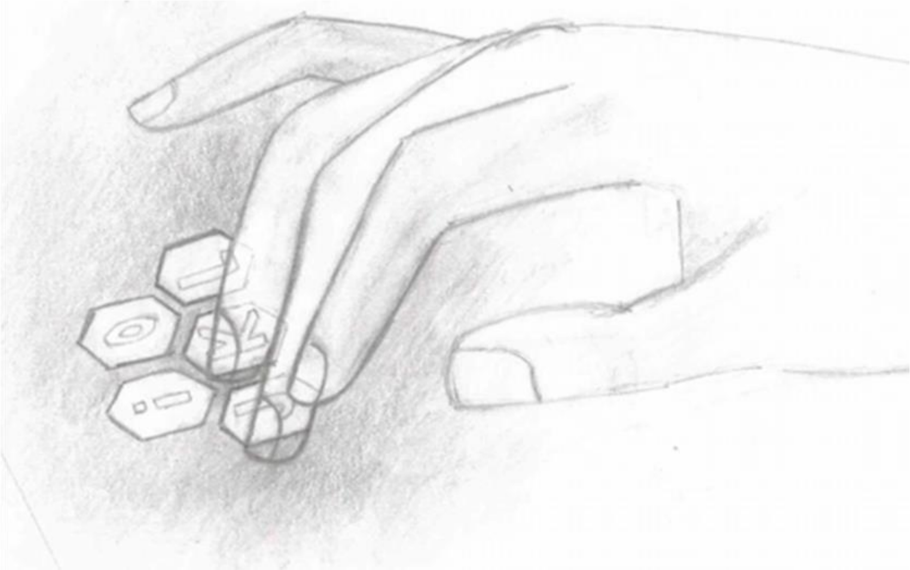

Raw sensor output
This enables subtle, always-available pinch and palm touch interactions



[RESEARCH PROTOTYPE]



# Mid-air Typing

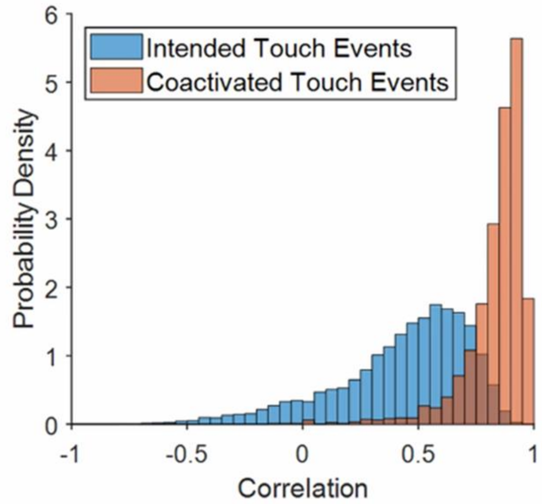



The Coactivation Problem

Understanding, Detecting and Mitigating the Effects of Coactivations in Ten-Finger Mid-Air Typing in Virtual Reality. Foy et al. *CHI 2021*

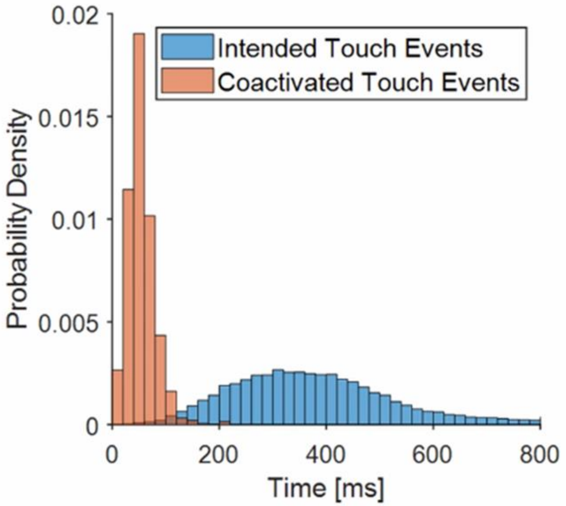
# Mid-air Typing

**Finger velocity correlation**

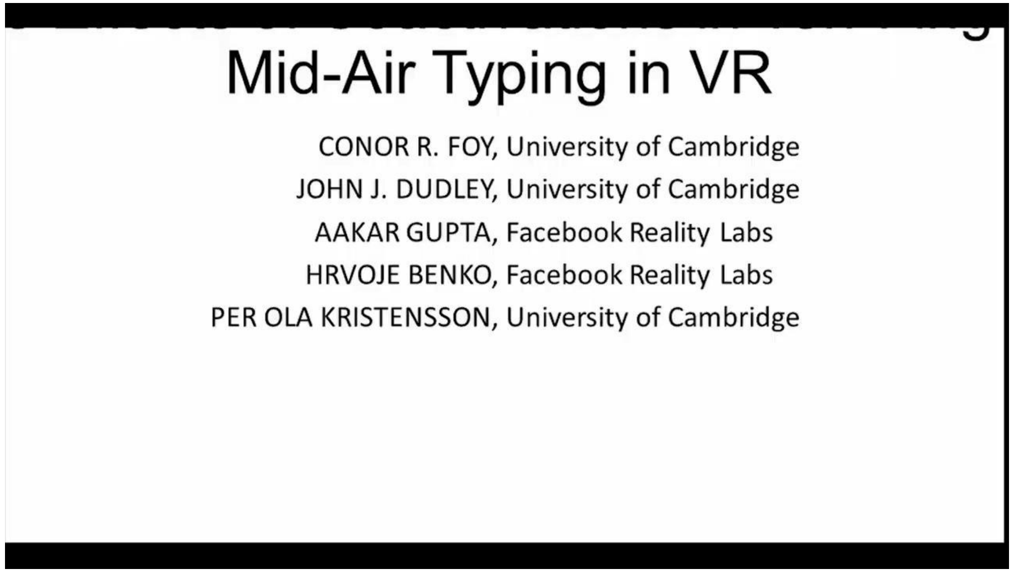



Correlation between the intended and coactivating finger velocities.

**Inter-key Interval**



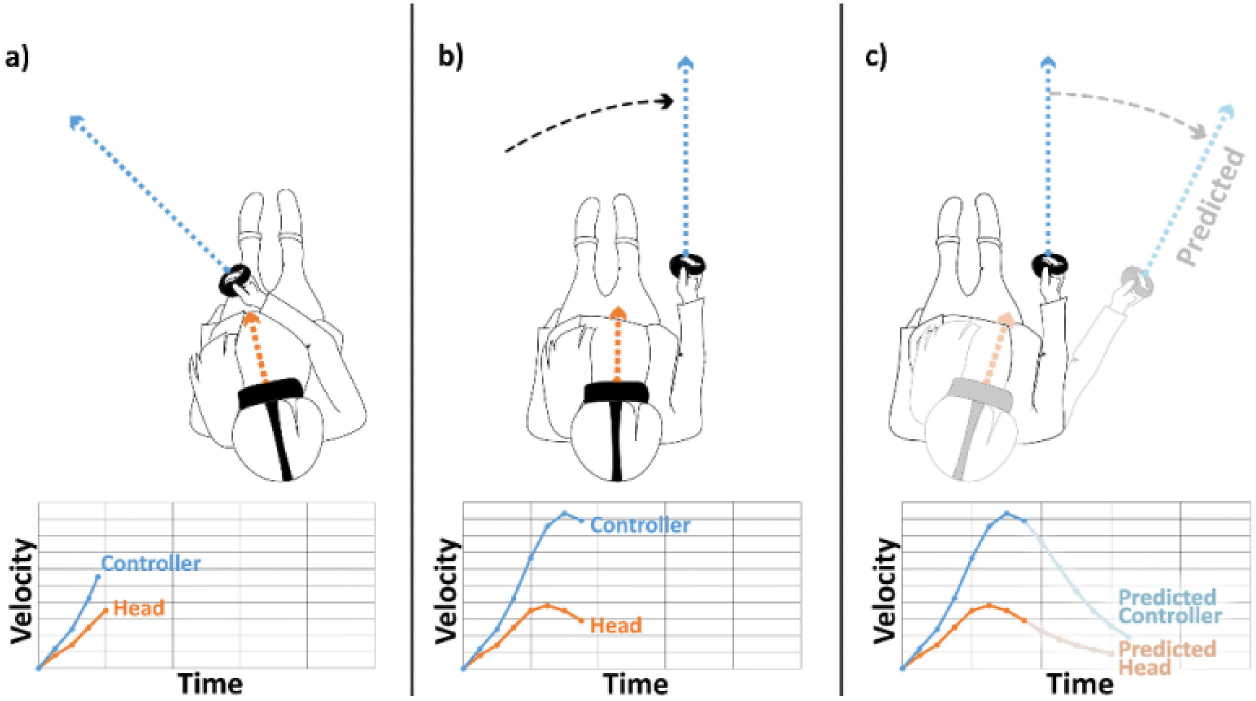

Interval between intended and coactivating touch events.

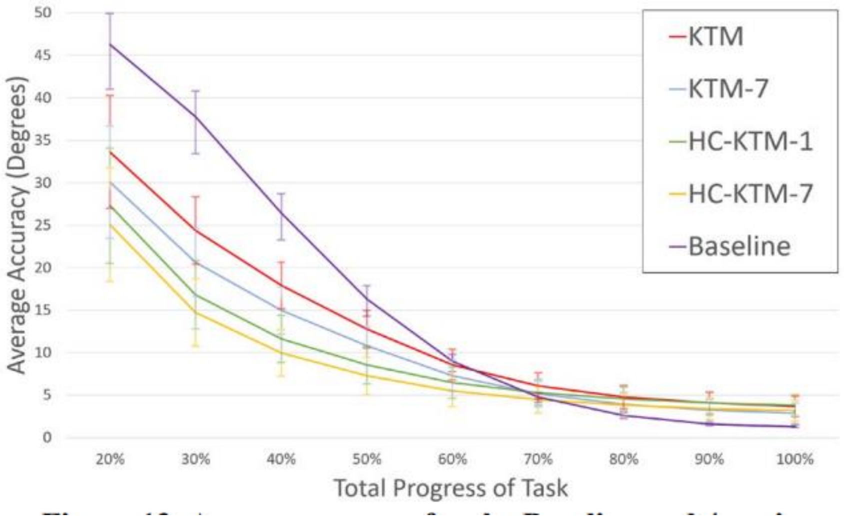



| Model | Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|---|
| Neural Net | $0.978 \pm 0.011$ | $0.882 \pm 0.072$ | $0.968 \pm 0.009$ | $0.921 \pm 0.044$ | $0.990 \pm 0.006$ |
| SVM | $0.973 \pm 0.011$ | $0.854 \pm 0.065$ | $0.971 \pm 0.010$ | $0.907 \pm 0.042$ | $0.985 \pm 0.006$ |
| Naïve Bayes | $0.976 \pm 0.005$ | $0.922 \pm 0.028$ | $0.902 \pm 0.021$ | $0.912 \pm 0.023$ | $0.984 \pm 0.008$ |

Understanding, Detecting and Mitigating the Effects of Coactivations in Ten-Finger Mid-Air Typing in Virtual Reality. Foy et al. *CHI 2021*

# Predictive Pointing

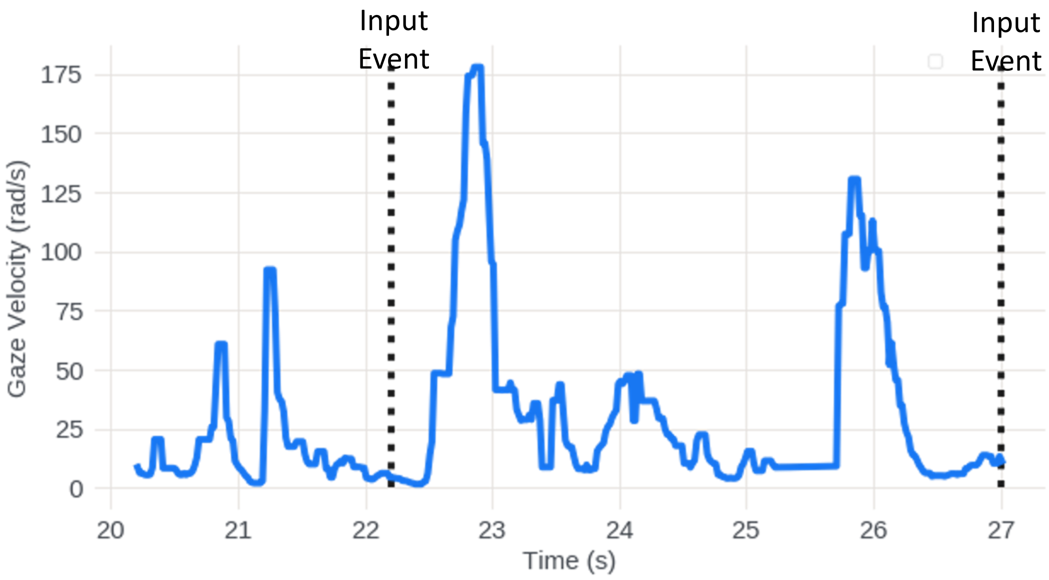


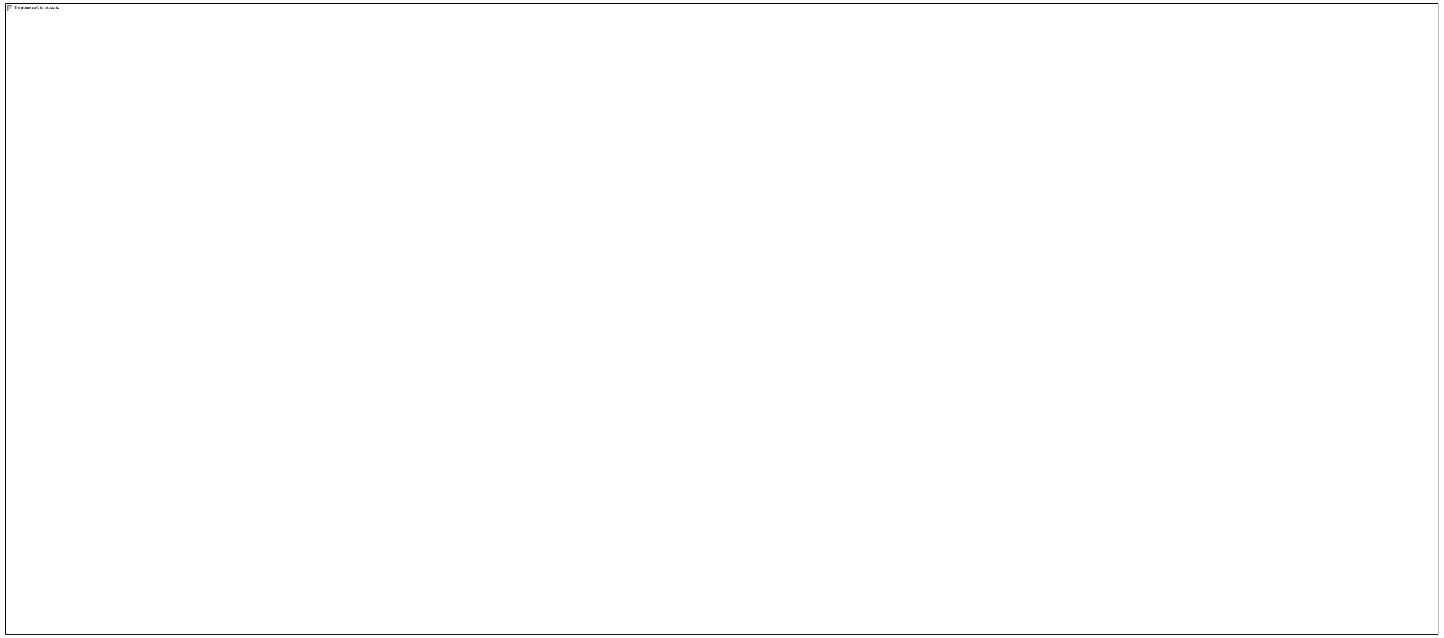




Head-Coupled Kinematic Template Matching
Henrikson et al. *CHI 2020*

# Computational Interfaces

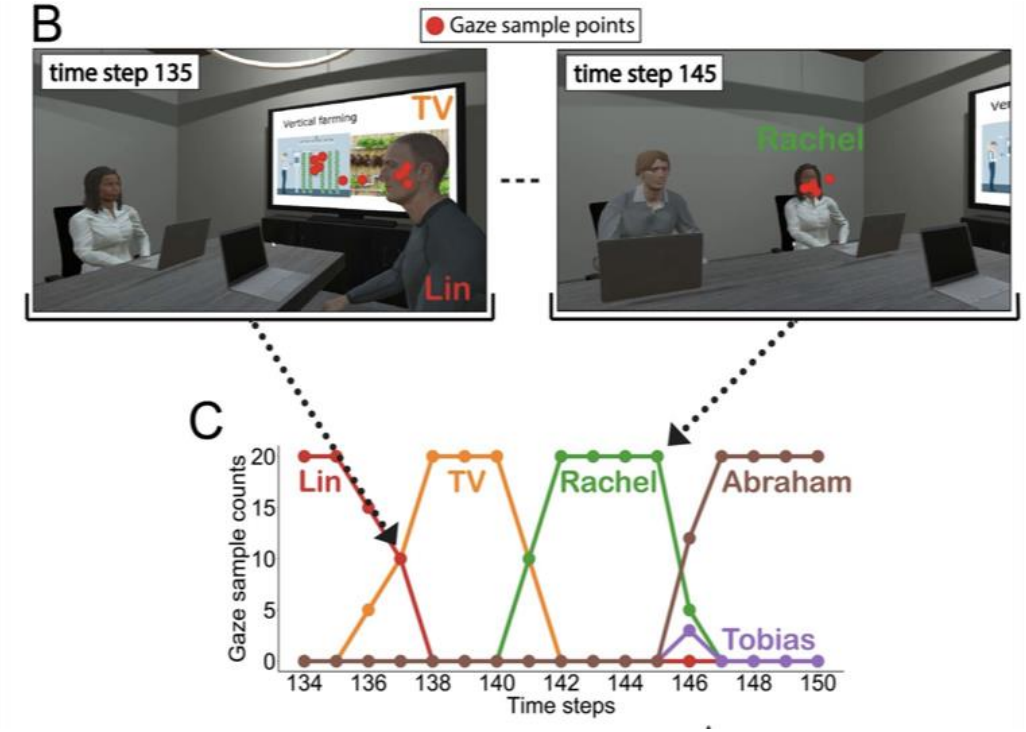
# Intent to interact using gaze dynamics



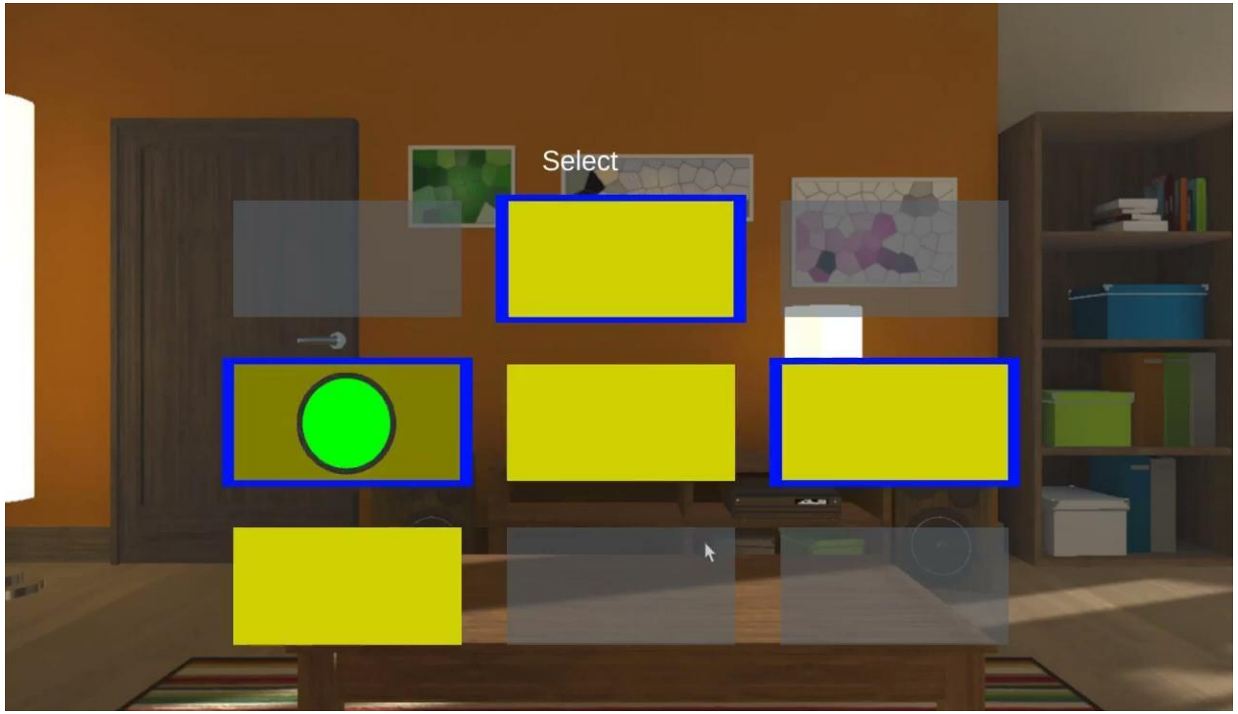
Towards gaze-based prediction of the intent to interact in virtual reality
David-John et al. *ETRA 2021*

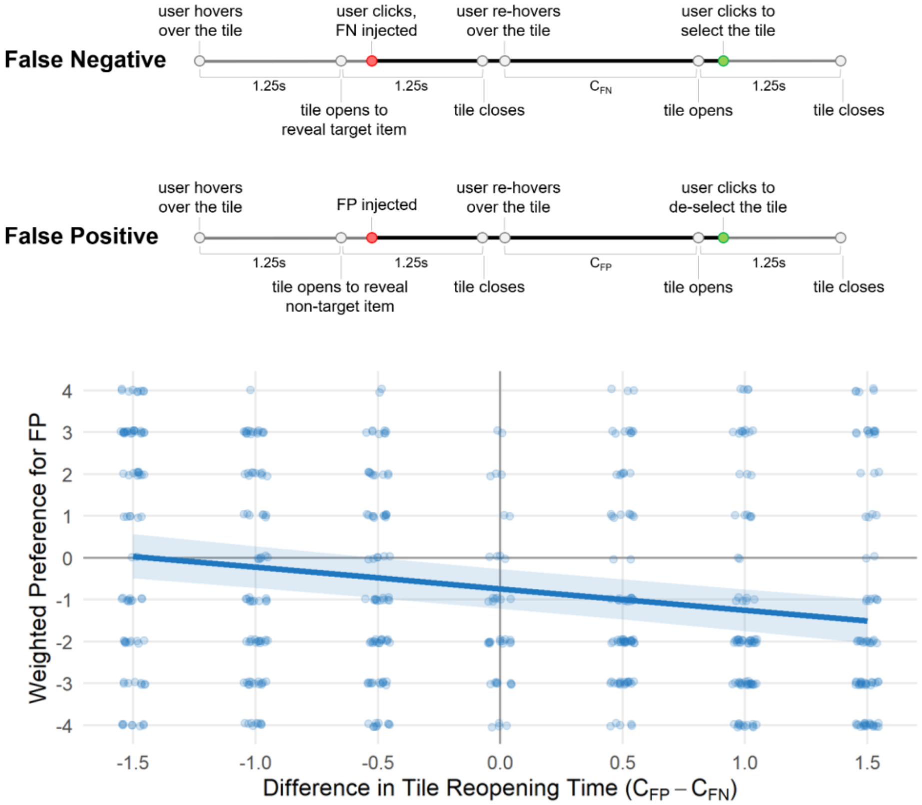
# Predicting Focus of Visual Attention




Predicting visual attention using the hidden structure in eye-gaze dynamics
Lengyal et al. *CHI 2021 Workshop*

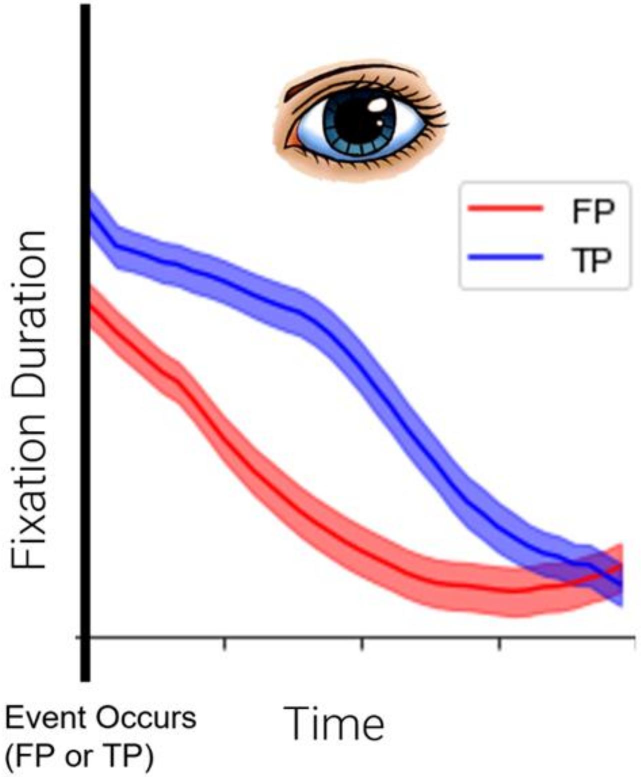
# Investigating the Costs of Input Errors



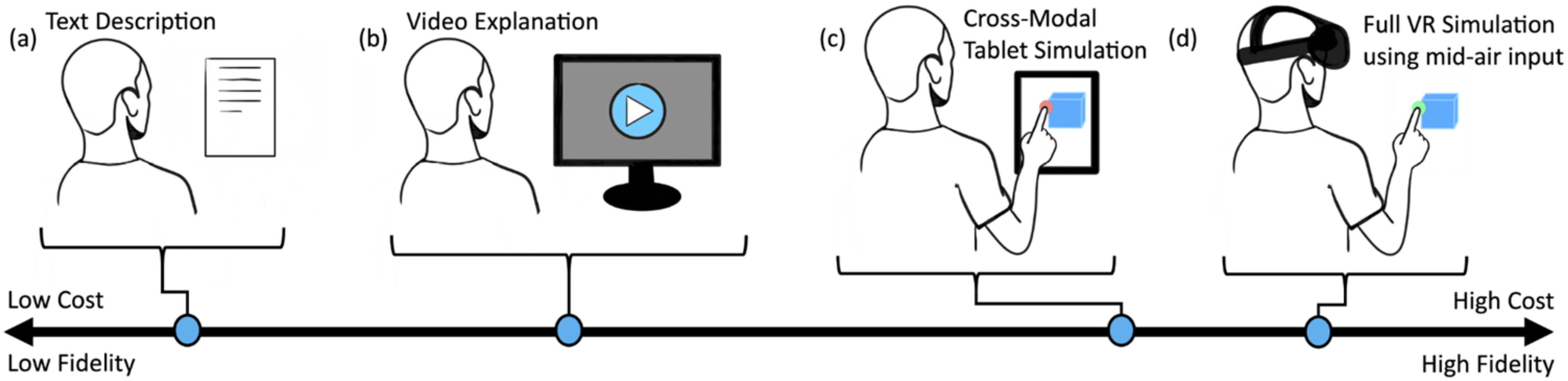

False Positives vs. False Negatives: The Effects of Recovery Time and Cognitive Costs on Input Error Preference
Lafreniere et al. *UIST 2021*

# Intelligent Error Mediation

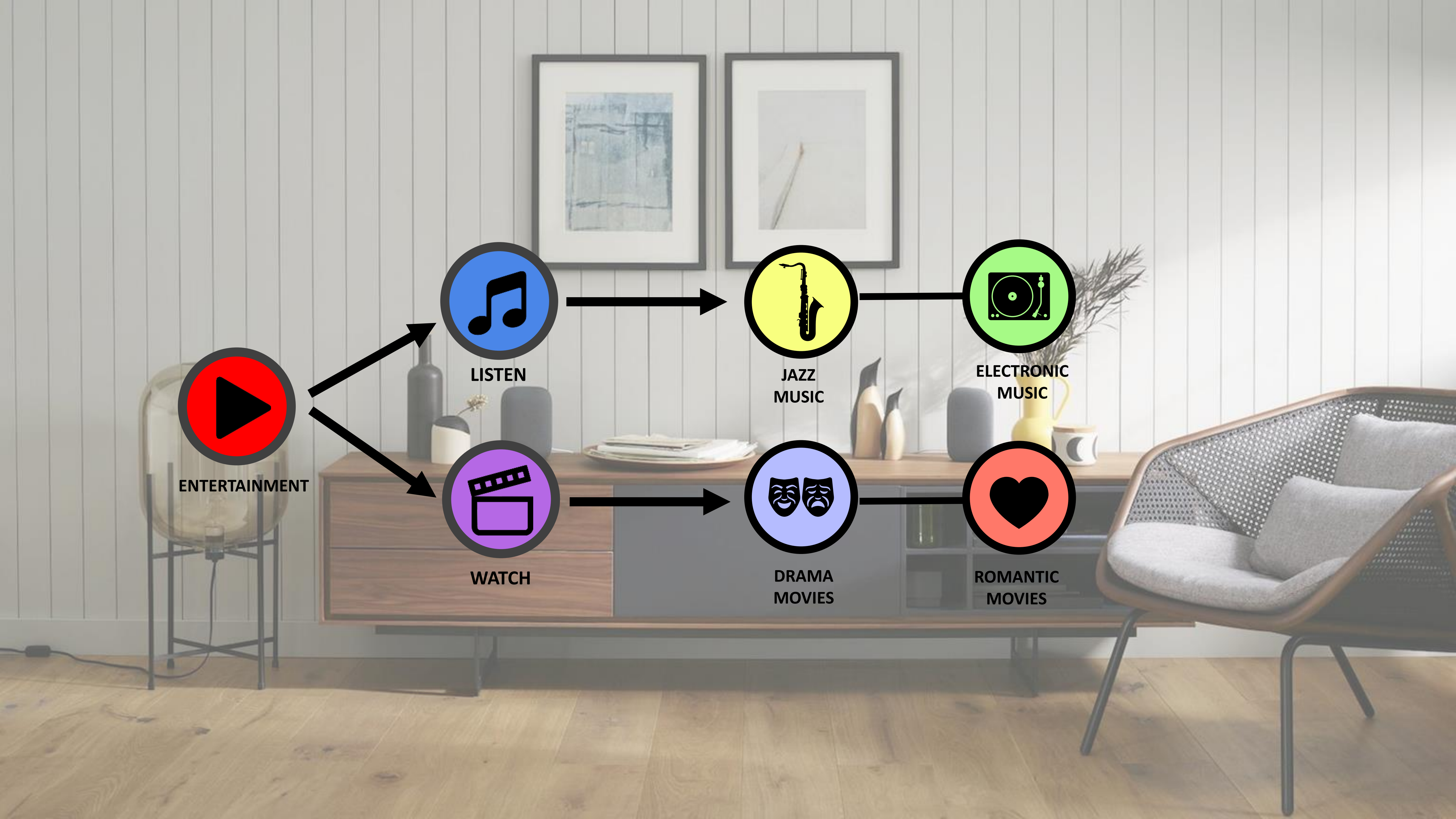

Gaze as an Indicator of Input Recognition Errors
Peacock et al. *ACM ETRA 2022*

# Novel Approaches to Evaluating Error Acceptability

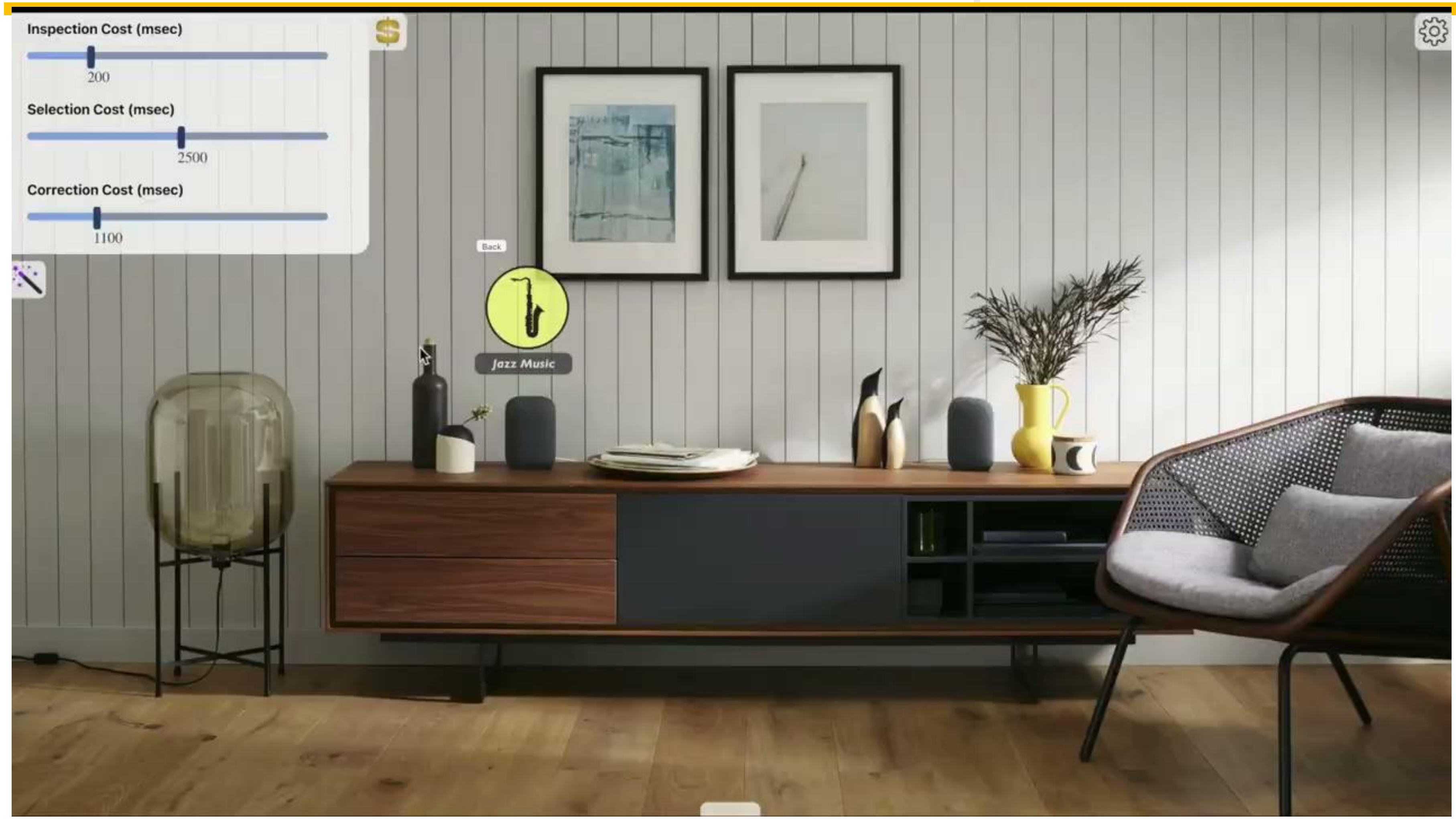



Investigating Cross-Modal Approaches for Evaluating Error Acceptability of a Recognition-Based Input Technique
Henderson et al. *ACM IMWUT 2022*

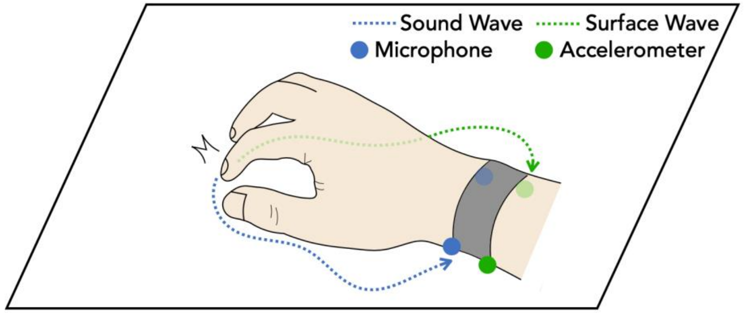

JAZZ
MUSIC
Intelligent Click
to Accept

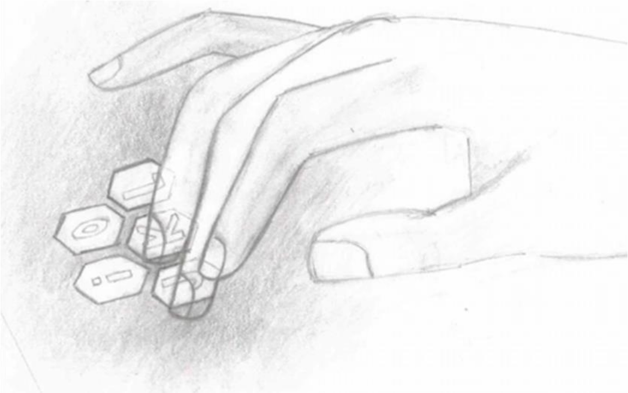

ENTERTAINMENT
LISTEN
WATCH
JAZZ
MUSIC
ELECTRONIC
MUSIC
DRAMA
MOVIES
ROMANTIC
MOVIES

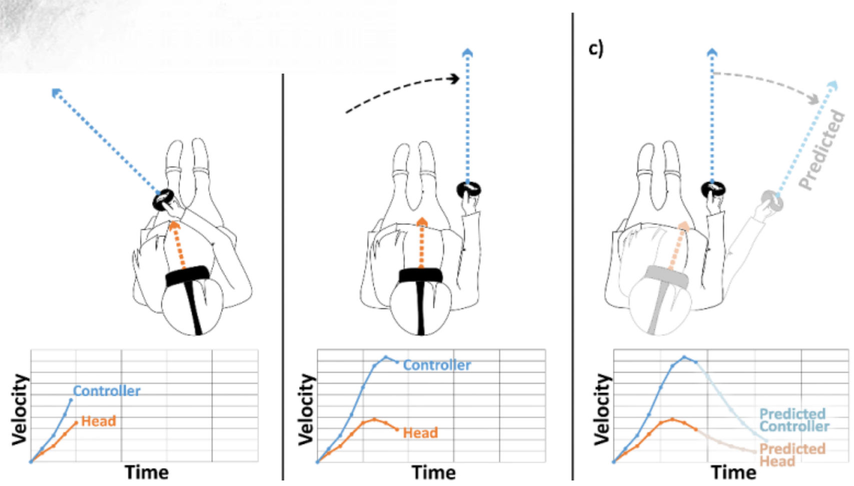
# Simulation-based Interface Adaptations

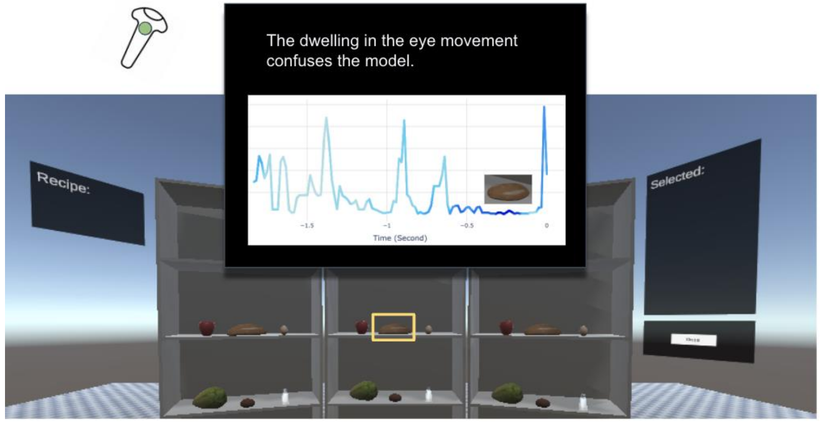


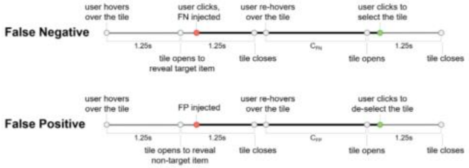

Computational Adaptation of Extended Reality Interfaces Through Interaction Simulation
Todi et al. *CHI 2022 Workshop*

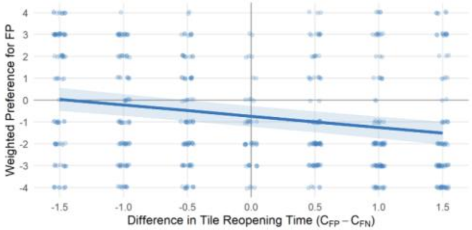

Sound Wave
Surface Wave
Microphone
Accelerometer

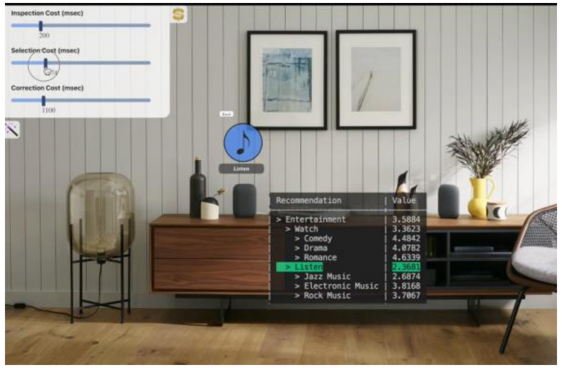



c)
Predicted
Velocity
Controller
Head
Time
Velocity
Controller
Head
Time
Velocity
Predicted Controller
Predicted Head
Time

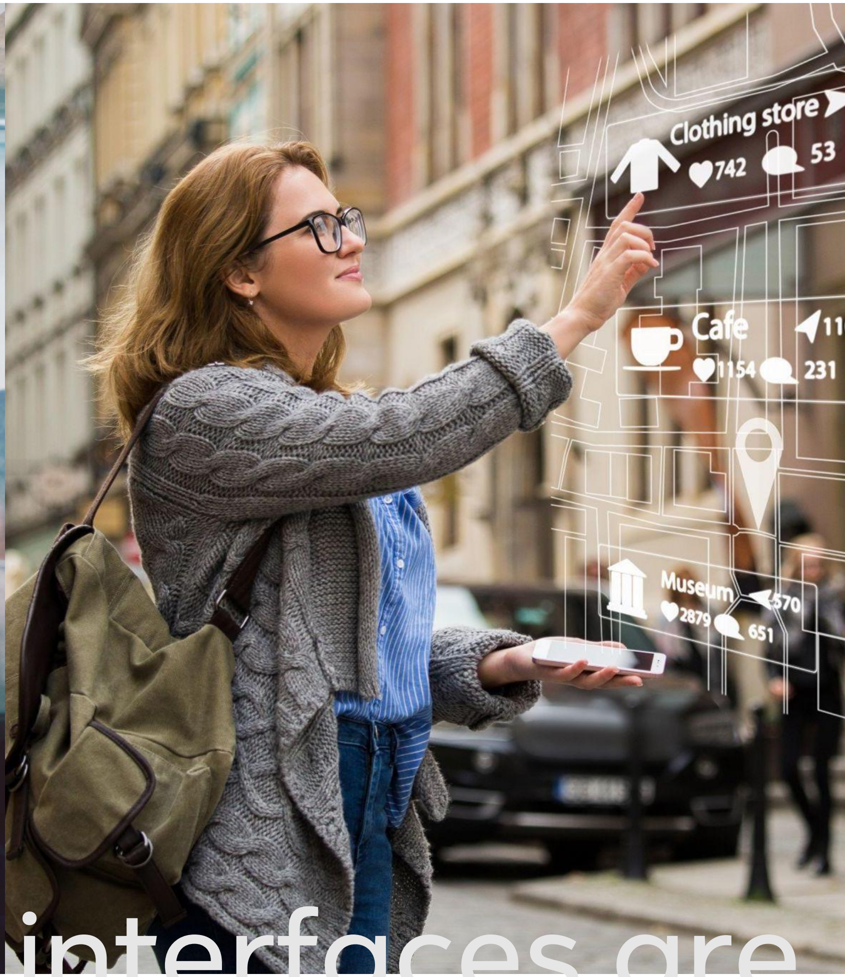

The dwelling in the eye movement confuses the model.
Time (Second)
Recipe:
Selected:


False Negative
user hovers over the tile
user clicks, FN injected
user re-hovers over the tile
user clicks to select the tile
tile opens to reveal target item
tile closes
tile opens
tile closes
False Positive
user hovers over the tile
FP injected
user re-hovers over the tile
user clicks to de-select the tile
tile opens to reveal non-target item
tile closes
tile opens
tile closes


Weighted Preference for FP
Difference in Tile Reopening Time (CFP − CFN)


Inspection Cost (msec)
Selection Cost (msec)
Correction Cost (msec)
Recommendation
Value
> Entertainment
> Watch
> Comedy
> Drama
> Romance
> Listen
> Jazz Music
> Electronic Music
> Rock Music

Mixed reality is a new era of computing…




…contextually adaptive interfaces are essential

# Building the future XR interface together

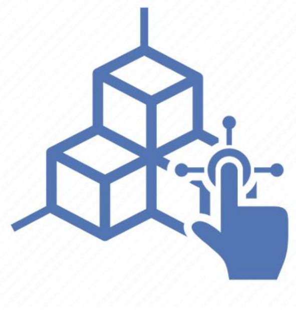
Full-time, Post-Doc, Interns

Research Scientists – HCI, Haptics, ML, AI
Research Engineers – SW, ML, HW
Design Technologists and Prototypers

www.meta.com/careers

REALITY LABS
RESEARCH

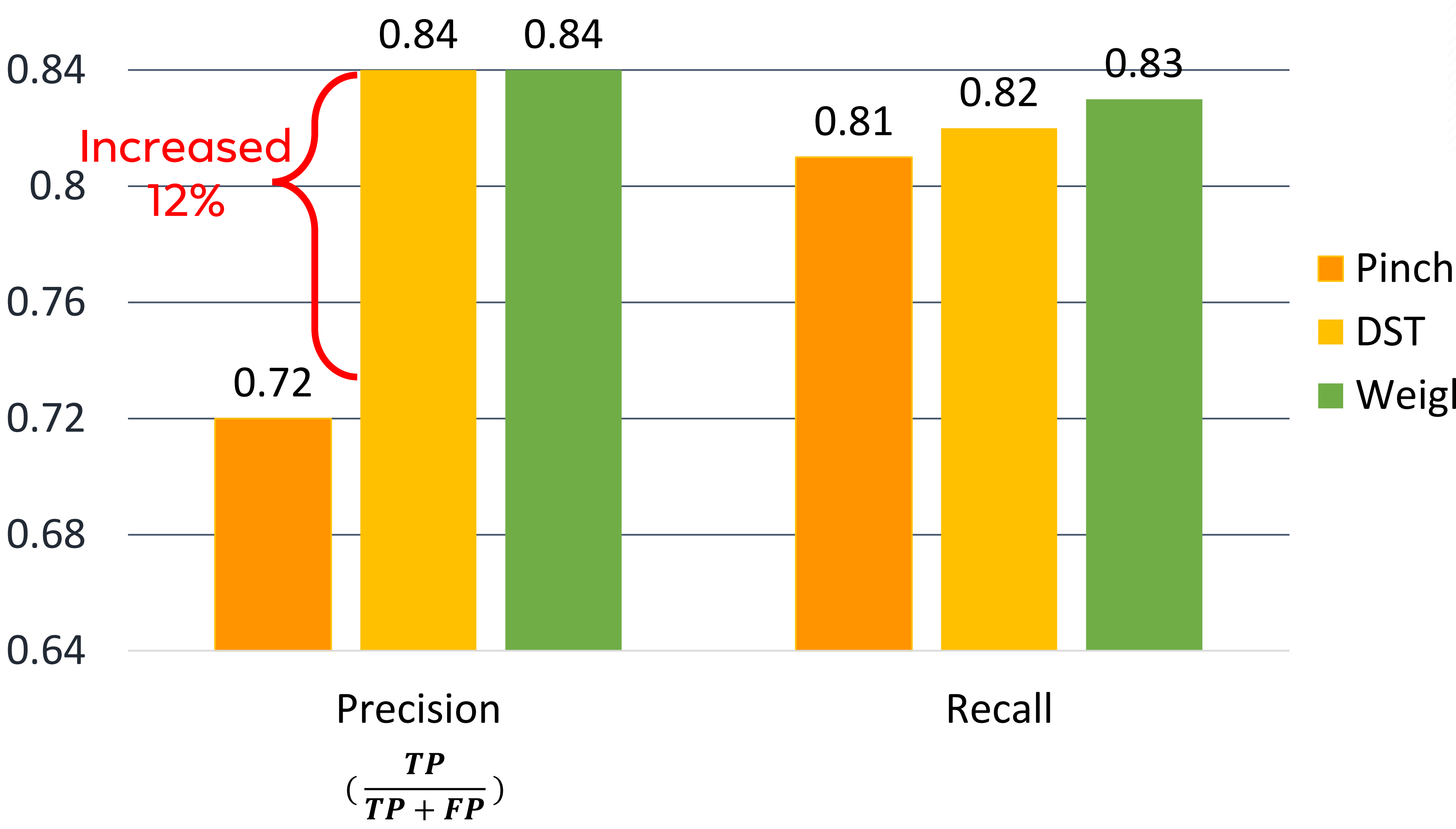
User Input



I2I

+

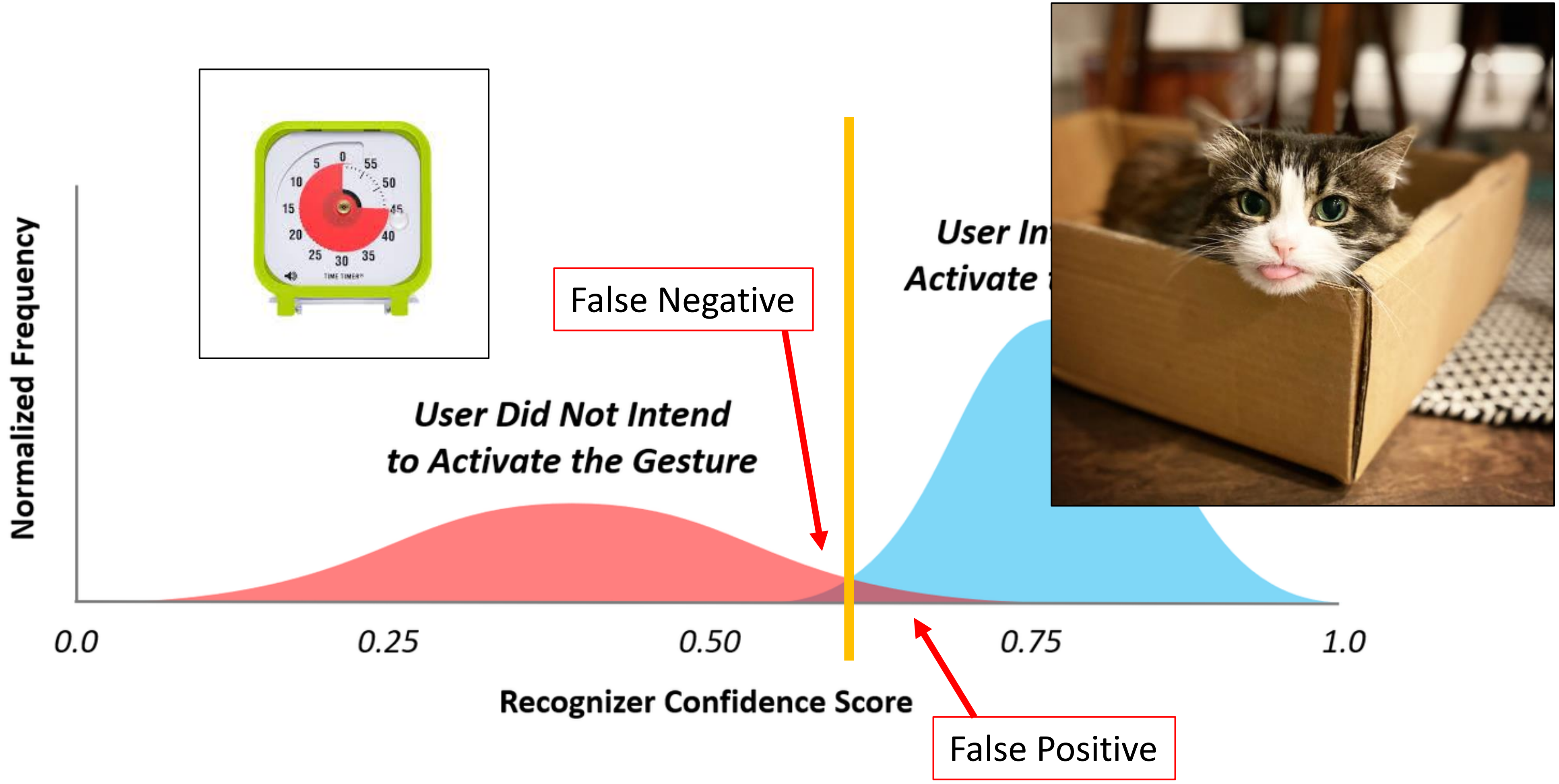
Input detection (Pinch Recognizer)

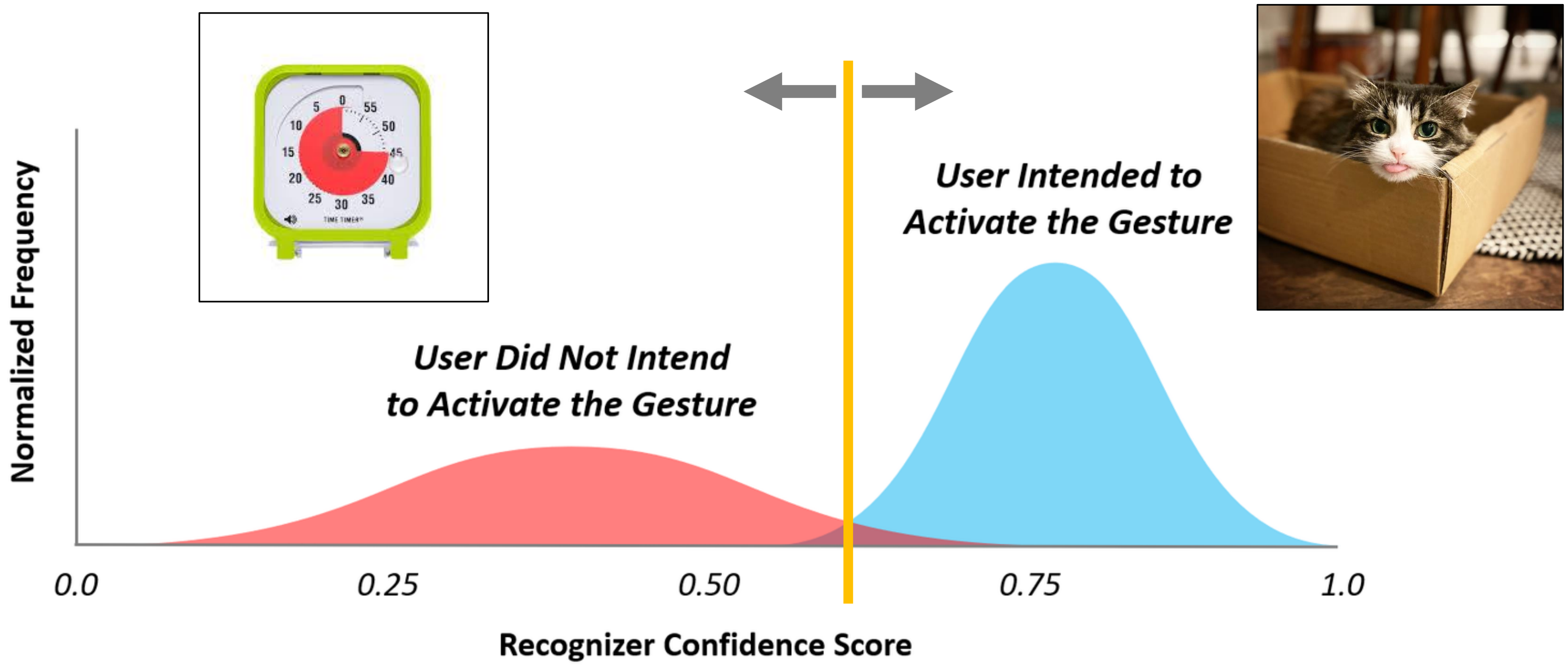


# When and how to adapt the UI?

Lafreniere, B., Jonker, T. R., Santosa, S., Parent, M., Glueck, M., Grossman, T., Benko, H., Wigdor, D. (2021) False Positives vs. False Negatives: The Effects of Recovery Time and Cognitive Costs on Input Error Preference. In *Proceedings of ACM UIST '21.*

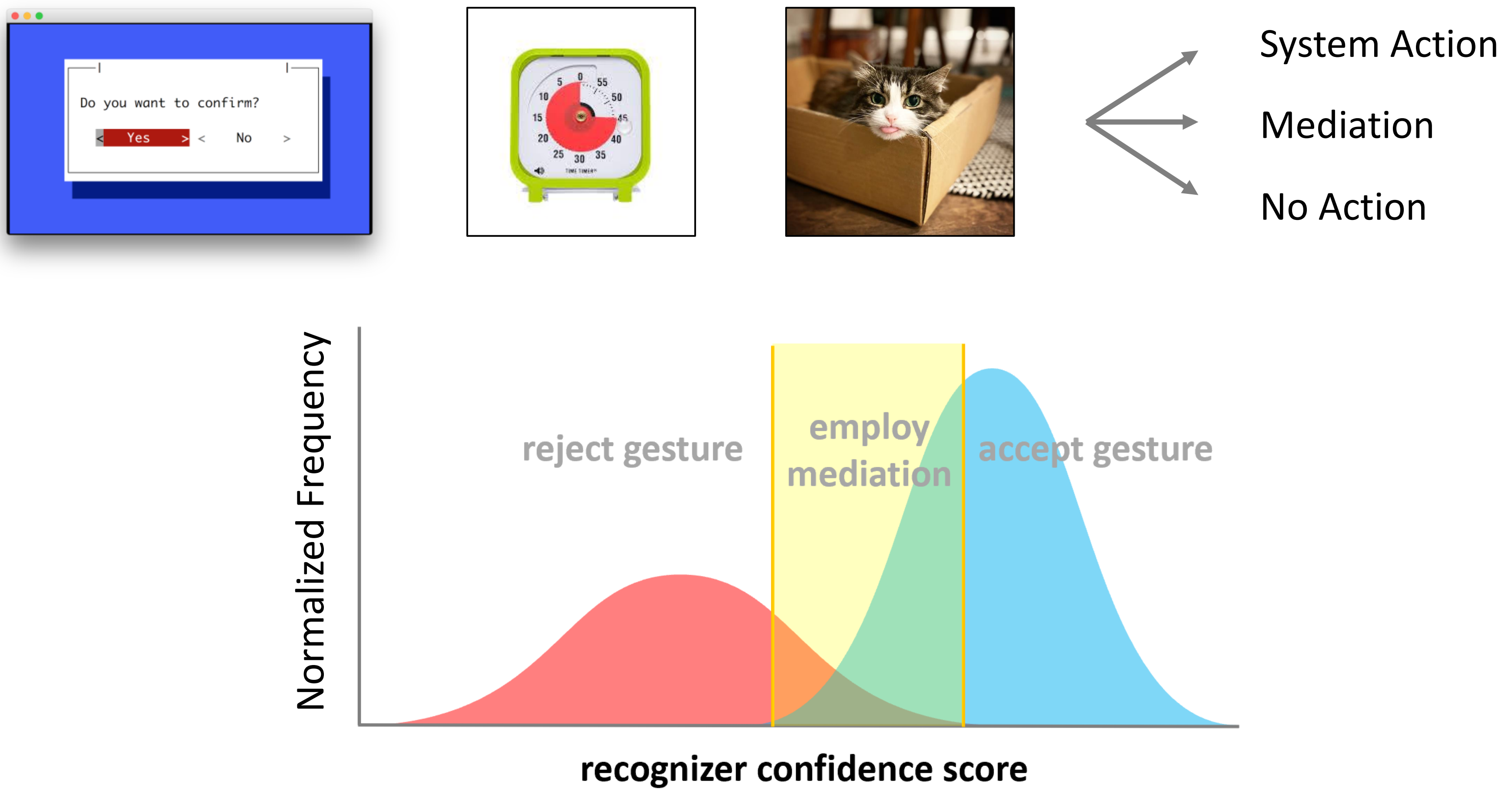
**Research Question**: How can a system assess and adapt to the costs of errors?




User Intended to
Activate the Gesture
User Did Not Intend
to Activate the Gesture
Normalized Frequency
0.0
0.25
0.50
0.75
1.0
Recognizer Confidence Score


Select


Select


Do you want to confirm?
Yes
No
System Action
Mediation
No Action
Normalized Frequency
reject gesture
employ mediation
accept gesture
recognizer confidence score

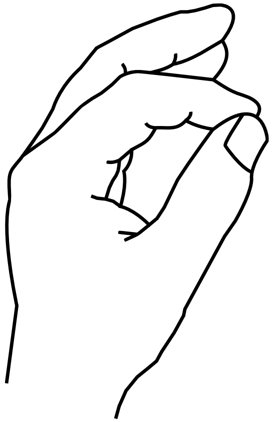
# False Positives vs. False Negatives

*The  Effects of Recovery Time and Cognitive Costs on Input Error Preference*

Ben Lafreniere, Tanya R. Jonker, Stephanie Santosa, Mark Parent,
Michael Glueck, Tovi Grossman, Hrvoje Benko, and Daniel Wigdor

FACEBOOK REALITY LABS
RESEARCH

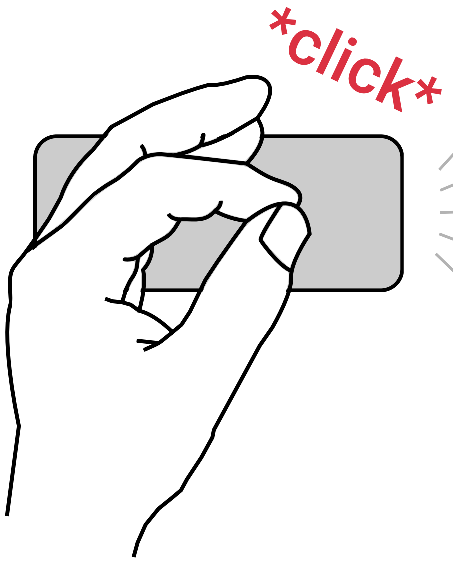
## False Negative Errors



User intentionally performs a gesture → System fails to recognize the gesture; No action is performed

## False Positive Errors

User is not intentionally performing a gesture → System recognizes a gesture anyway; An unwanted action is performed




*Compose Email*

*Attach File*

*Delete Selected Items*

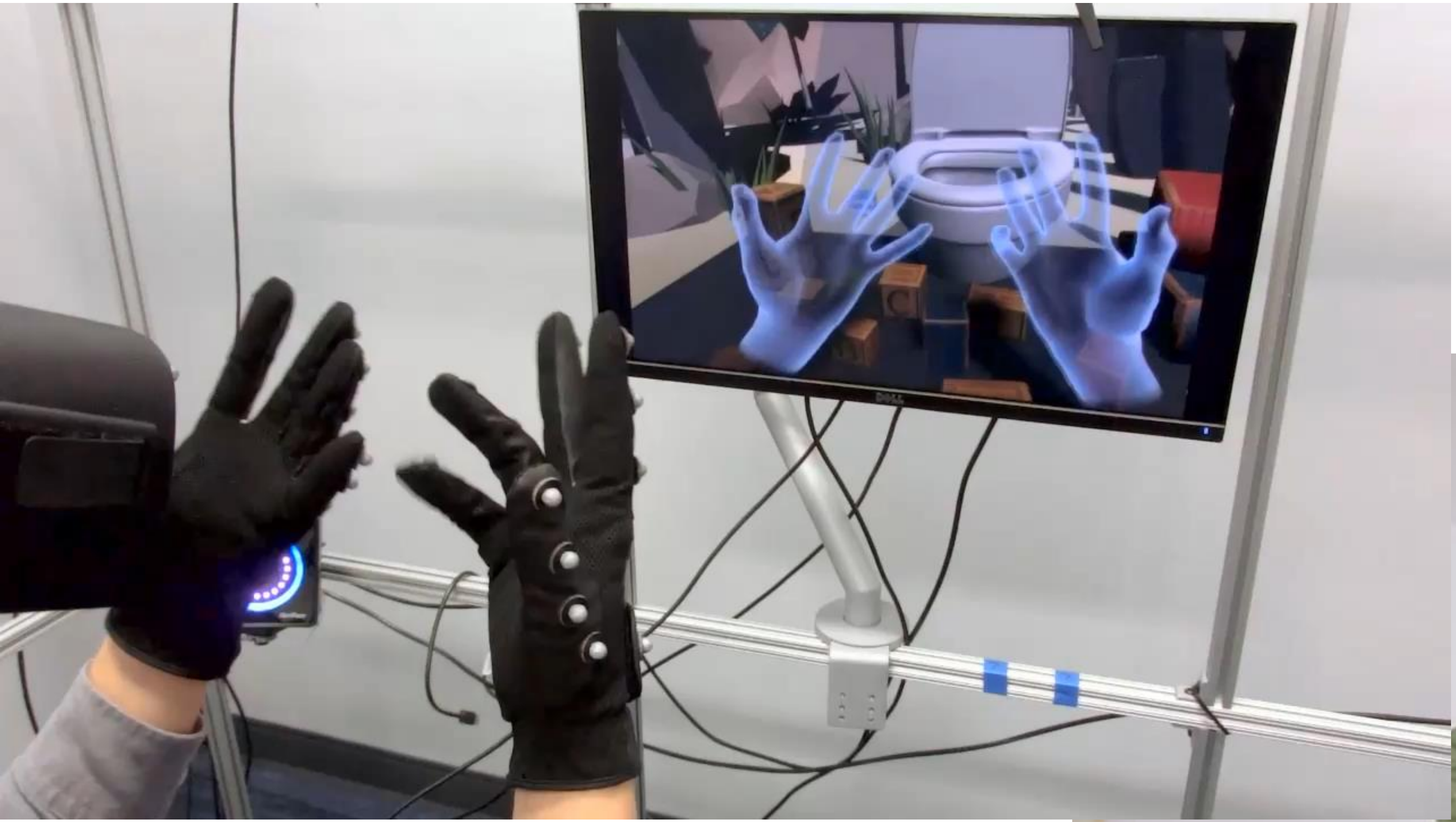
*Reply All*

*Send Message*

# Key Takeaways

- Error-type preference can be driven by differences in the temporal cost of FP and FN errors
- Users exhibit a bias against FP errors, which can be equivalent to 1.5 seconds or more of added recovery time
- FP errors impose greater attentional demands on users as compared to FN errors, which may partially explain this bias

# Hand tracking



Shangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D. Twigg, and Kenrick Kin. 2018. Online optical marker-based hand tracking with deep labels. ACM Trans. Graph. 37, 4, Article 166 (July 2018)

Hands in Oculus Quest 2

**Voice Assistant Interface**

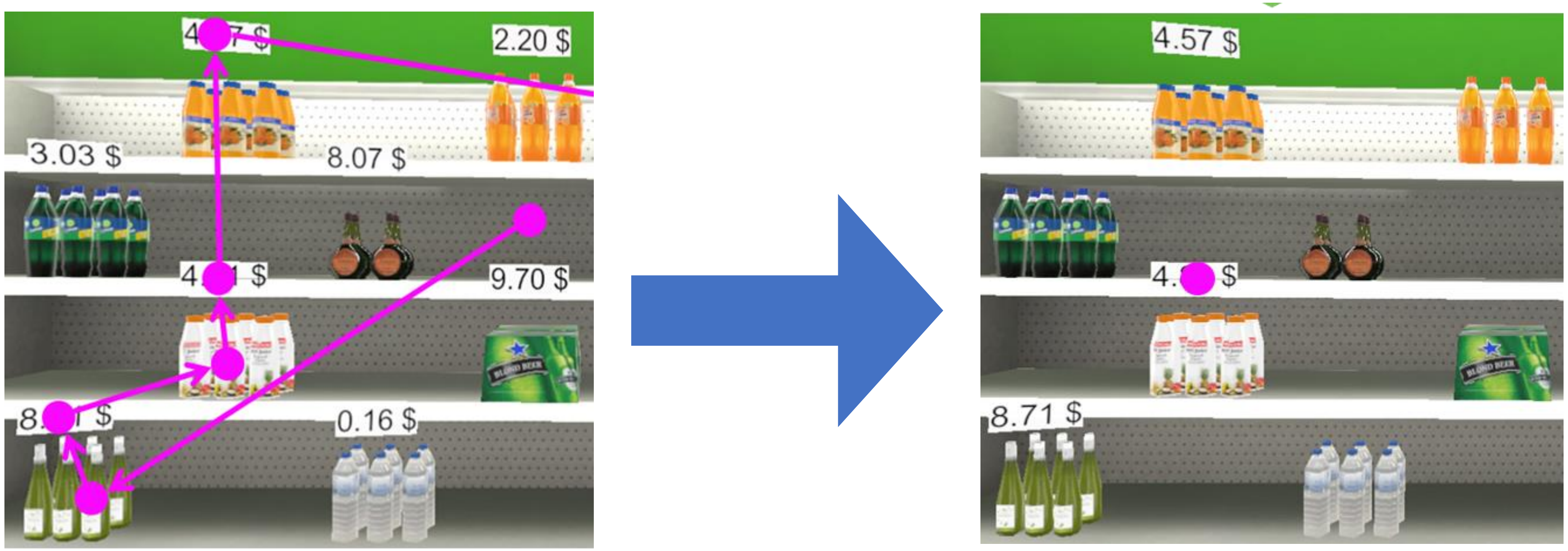
- User initiates interaction
- Limited contextual understanding
- Turn taking dialogue for disambiguation

**MR Interface**

- Proactive
- Understands context and user goals
- Engages continuous multimodal feedback to reduce errors and enable disambiguation.

# Learning MR UI Policies from Gaze Data

Trained RL agents to predict when an MR label is meaningful to the user.




**Context:** User's gaze behavior + task + environment

**Output: Inferring task-specific goals +** reduced clutter of MR labels

https://ait.ethz.ch/label-agent/
Gebhardt et al. **"Learning Cooperative Personalized Policies from Gaze Data"** ACM UIST 2019