

Khhalasi-I/O

MAKING MULTI-TASKING MORE EFFECTIVE

Situationally Induced Impairments and Disabilities (SIIDs)

SIIDs refer to temporary impairments or disabilities that arise due to specific circumstances or situations, rather than being permanent or inherent conditions.



Activity: User is preparing food in a kitchen.

Environment: User is in a kitchen.

Hand: Holding a bowl

Volume: 65 dB



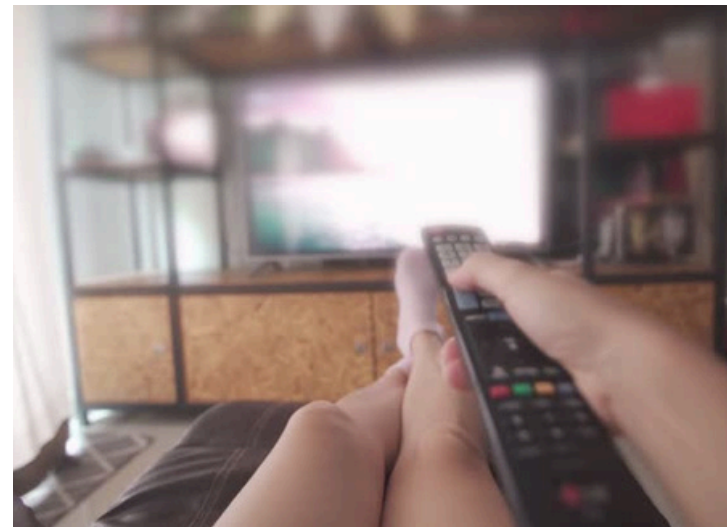
Activity: User is preparing food in a kitchen.
Environment: User is in a kitchen.
Hand: Holding a bowl
Volume: 65 dB



Incoming
Call



User can't directly pick up the call, user needs to multi-task and divert attention.



Activity: user watching television
Environment: User in a closed interior space
Hand: Holding Remote
Volume: 70DB



Doorbell
Rings



User may not be able to hear in the current environment situations.



These interruptions in modern age are generally mitigated using a smart device.

They require user to divert attention and multi-task.

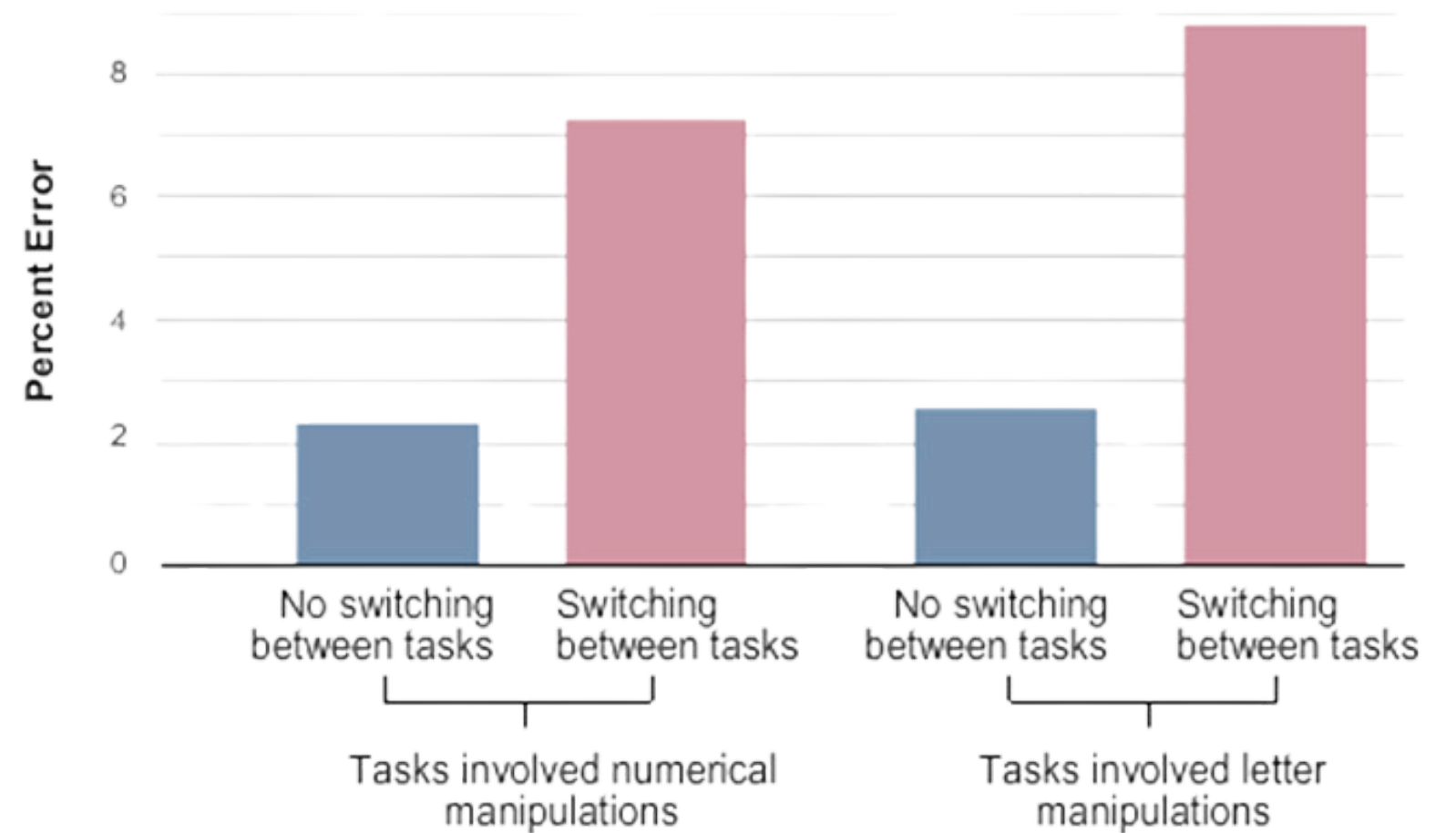
Types of multitasking

1. Digital - Physical
2. Digital - Digital

prone to errors

Multi-tasking is prone to errors and lack of attention in a task may cause an accident.

error percentages for
no-switching and switching activities



attention residue

Attention residue reflects the persistence of cognitive activity about a Task A even though one stopped working on Task A and currently performs a Task B.

It has been proven that attention residue, decreases efficient multi-tasking.

problem statement

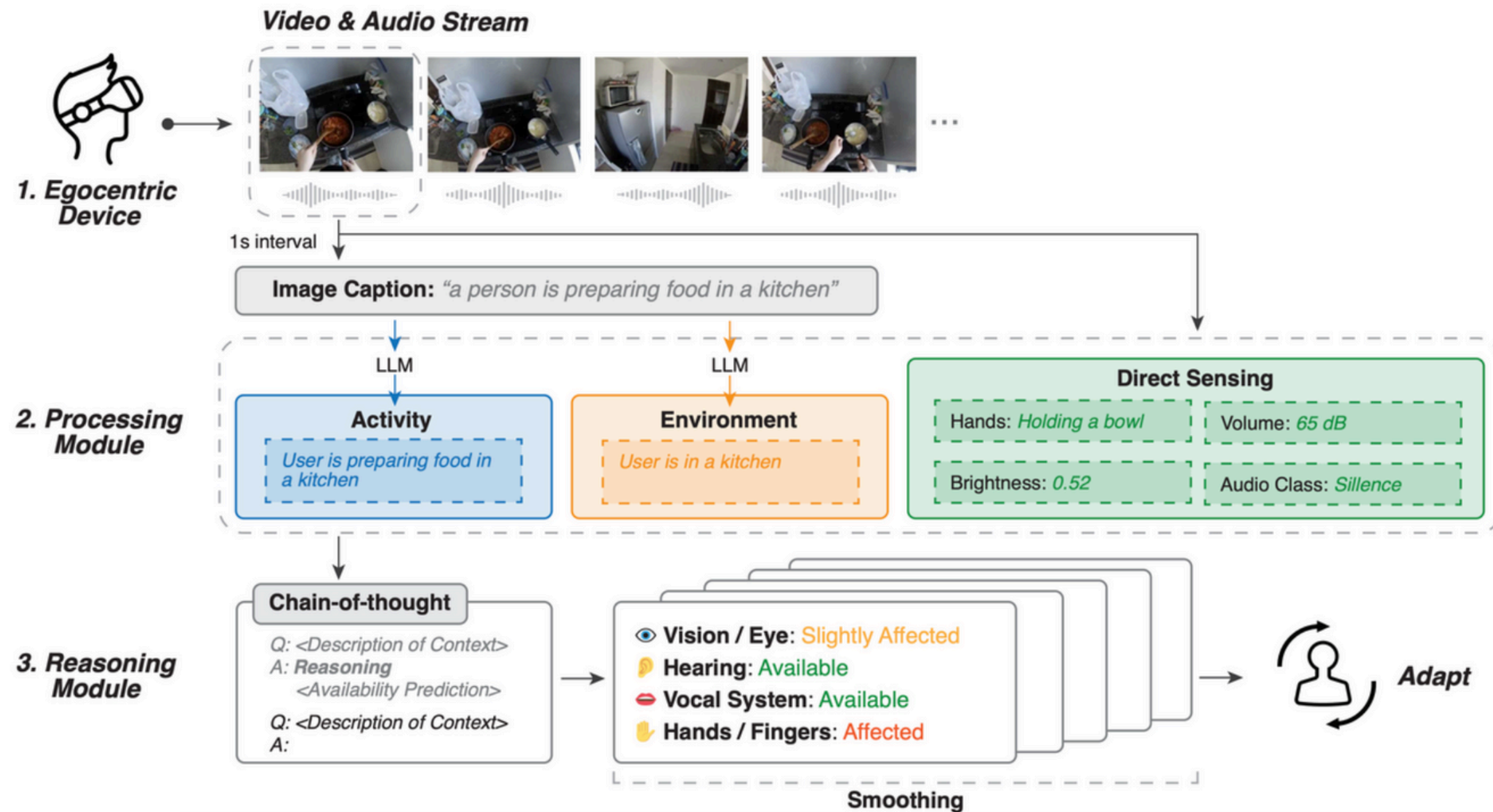
Effective multitasking has become standard as users frequently juggle between tasks across diverse devices, daily activities, and dynamic environments. This leads to SIIDs which can severely hinder a user's ability to multitask, reducing their capacity to engage with technology or respond appropriately to dynamic challenges present in the environment.

diliverable

Developing an open-source middleware for real-time inference of human sensory channels for AR/VR applications to improve user experience by eliminating SIIDs.

Literature Review

HumanIO [CHI24, Honourable Mention]



Inference

Lite Model w/o CoT: **2.07 sec**

Model w CoT: **19.95 sec**

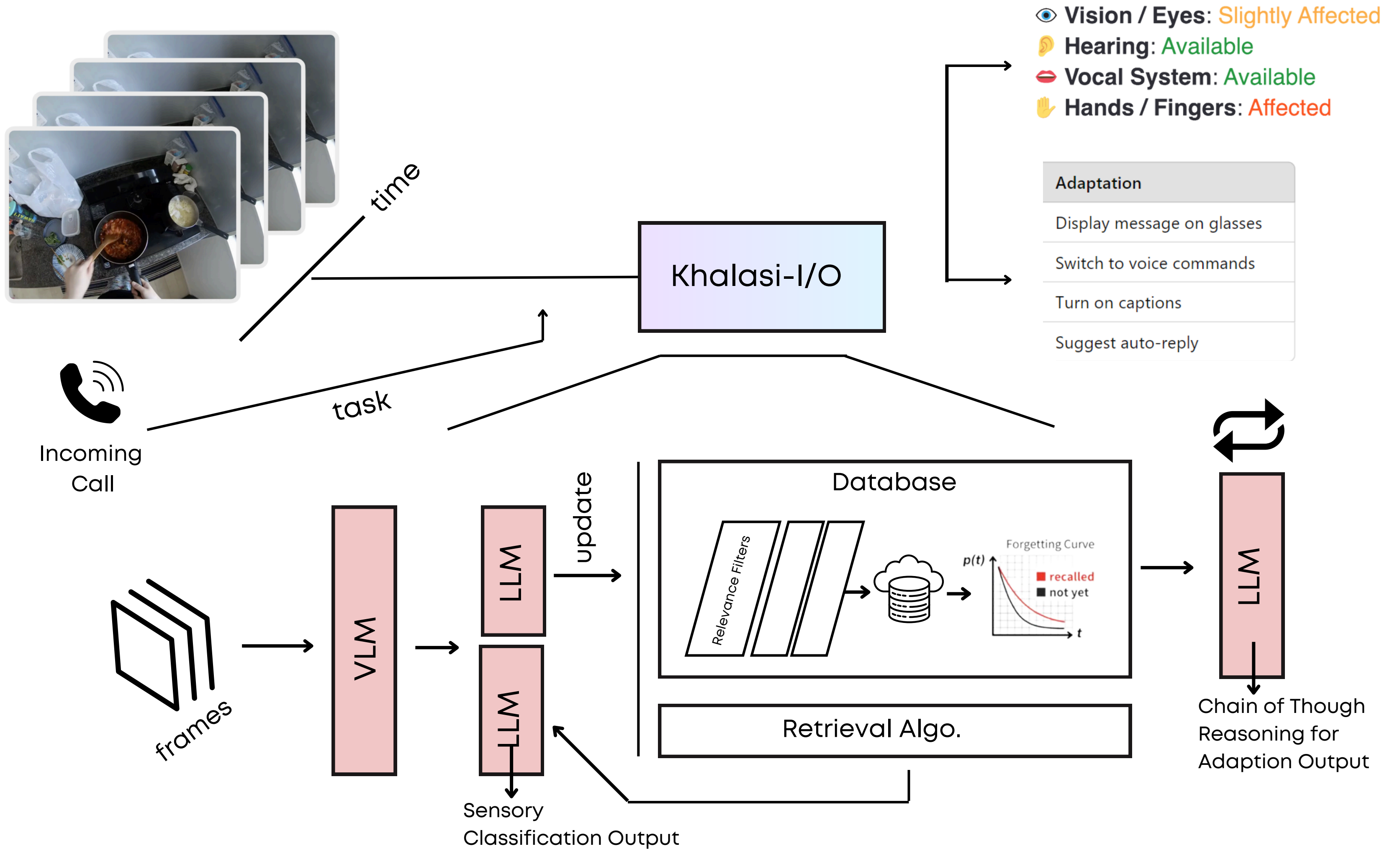
Results

Achieves a 0.22 mean absolute error and a 82% accuracy in availability prediction across 60 in-the-wild.

Dataset - our own



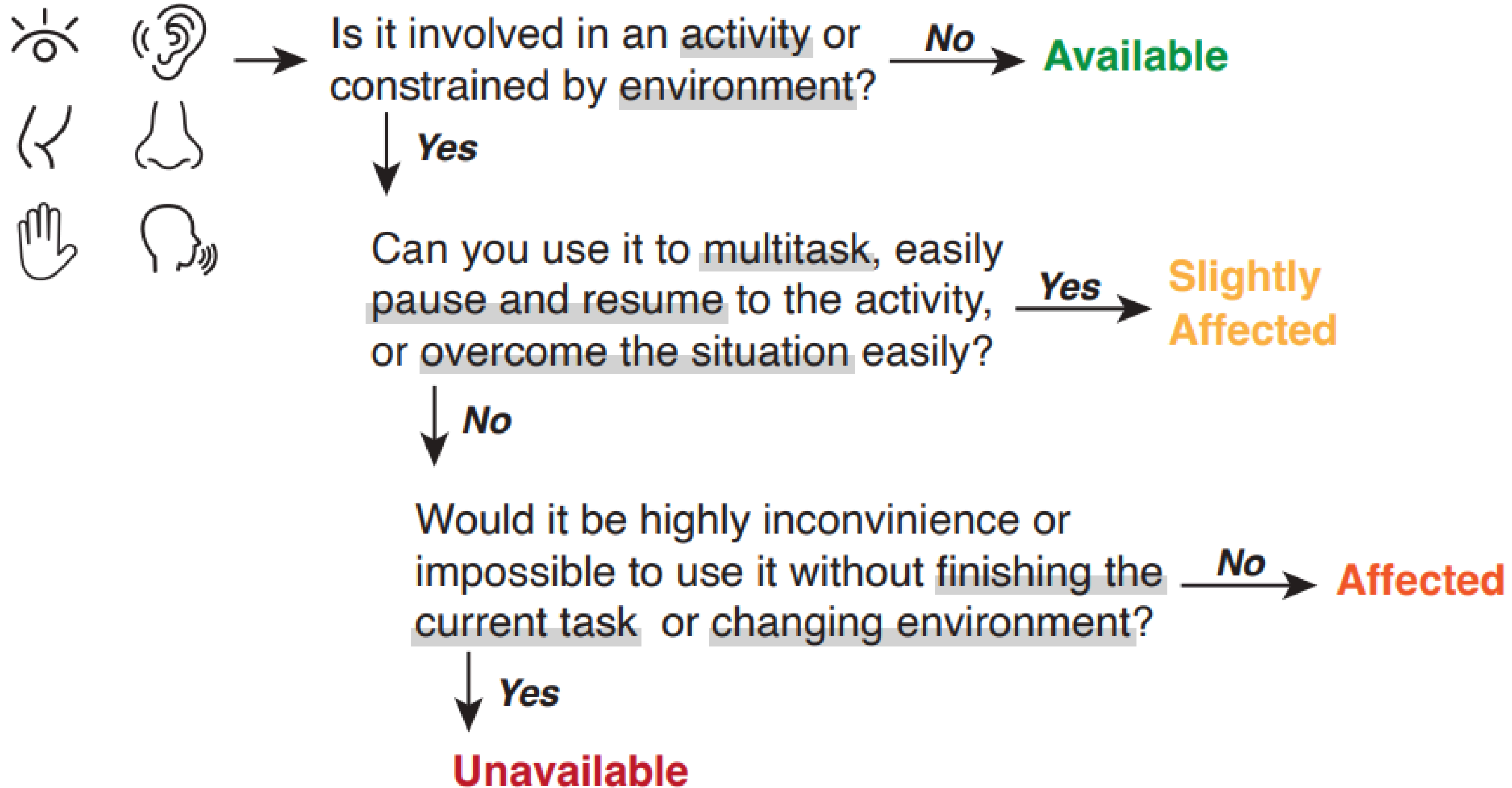
Our Approach





	hand_left_state	hand_right_state	vision	intent	critical_focus_level
0	holding sink	holding black object	bright, daytime lighting	handwashing	Low
1	wet	holding faucet handle	well-illuminated, natural light	washing hands	Low
2	submerged in water	holding soap	bright, daytime lighting	washing hands	Low
3	wet, holding hands together	wet, holding hands together	well-illuminated, neutral lighting	drying hands	Low
4	holding paper towel	holding metal dispenser	well-illuminated, natural lighting	drying hands	Low
5	reaching for towel	holding towel dispenser	bright, daytime lighting	drying hands	Low
6	placed on counter	near faucet	well-illuminated, daylight	drying hands	Low
7	not-visible	visible, in motion	bright, warm lighting, moderate angle of incid...	moving/walking	Low
8	holding green object	holding red object	well-illuminated, daytime lighting	moving towards destination	Mid
9	bent at the elbow, near the chest	holding [object] (not specified what object)	brightly lit, daytime, with visible sunlight i...	not specified	Low
10	not-visible	holding door frame	well-illuminated, neutral lighting	entering a room	Low
11	not-visible	not-visible	well-illuminated, natural lighting	exploration	Low
12	not-visible	holding keyboard	well-illuminated, neutral lighting	using laptop	Mid
13	hovering over mouse	resting on keyboard	bright, neutral lighting	gaming	Mid
14	typing on keyboard	typing on keyboard	dimly-illuminated, focused on laptop screen	working on coding project	Mid

Chain of Thought



Evaluation Metrics

A. Evaluating Model Performance

B. Evaluating Effect on Human
Experience using Khalasi-I/O

Mean Absolute Error (MAE):

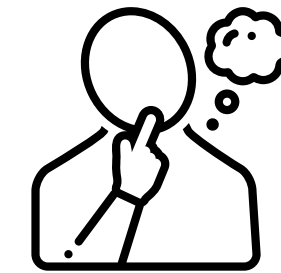
metric captures the average discrepancy between predicted availability levels and the actual ground truth for each of the four channels within every video clip. Given the nature of our data, MAE offers a precise error depiction. This is because misclassifying a level as “Slightly Affected” instead of Affected is less erroneous than confusing “Unavailable” with “Available”. Availability levels are numerically translated to values ranging from **1 to 4**.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

10 mins video - 60fps - 60x60x10 - 3600 frames -> 360 frames (Manual Ground truth)

Experiment Design

16 participants (age range: **20 – 23**)



vs

Khalasi-I/O

given 3 chosen tasks \longrightarrow given incoming interruption

w/o Intervention vs w Khalasi-IO Intervention } **S1**

Human Adaptation vs Khalasi-IO Adaptation } **S2**

8 students

Scenario	Incoming Task
Working (Typing)	Receive a text message
Hand Washing	Receive a phone call
Hair Drying	Watch a video



Control Group

NASA - TLX

- *Mental Demand*
- *Physical Demand*
- *Temporal Demand*
- *Overall Performance*
- *Effort*
- *Frustration*

Adaptations

3 chosen tasks



Khalasi-I/O

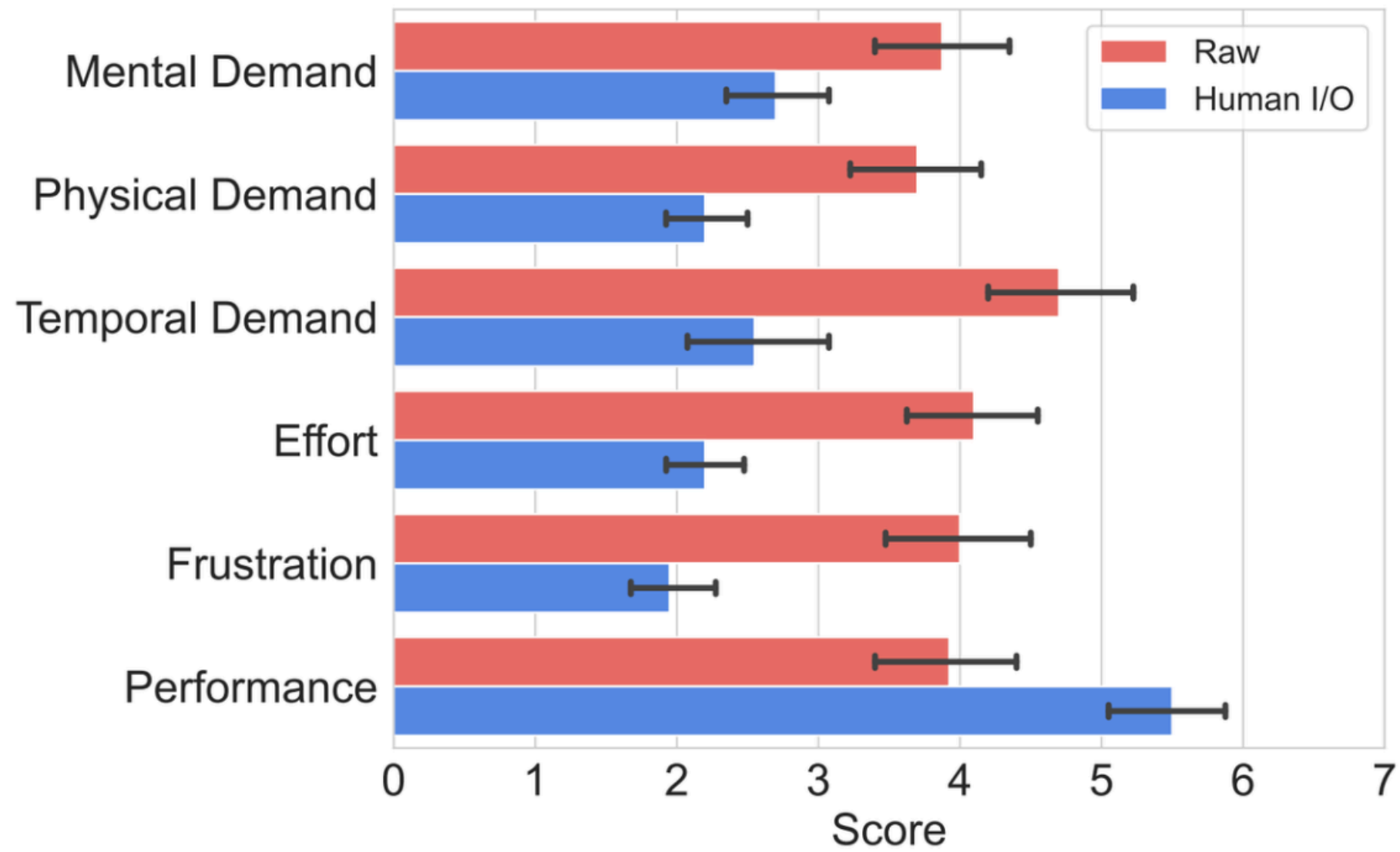


Adaptation
Display message on glasses
Switch to voice commands
Turn on captions



S1
Treatment
8 students

NASA - TLX

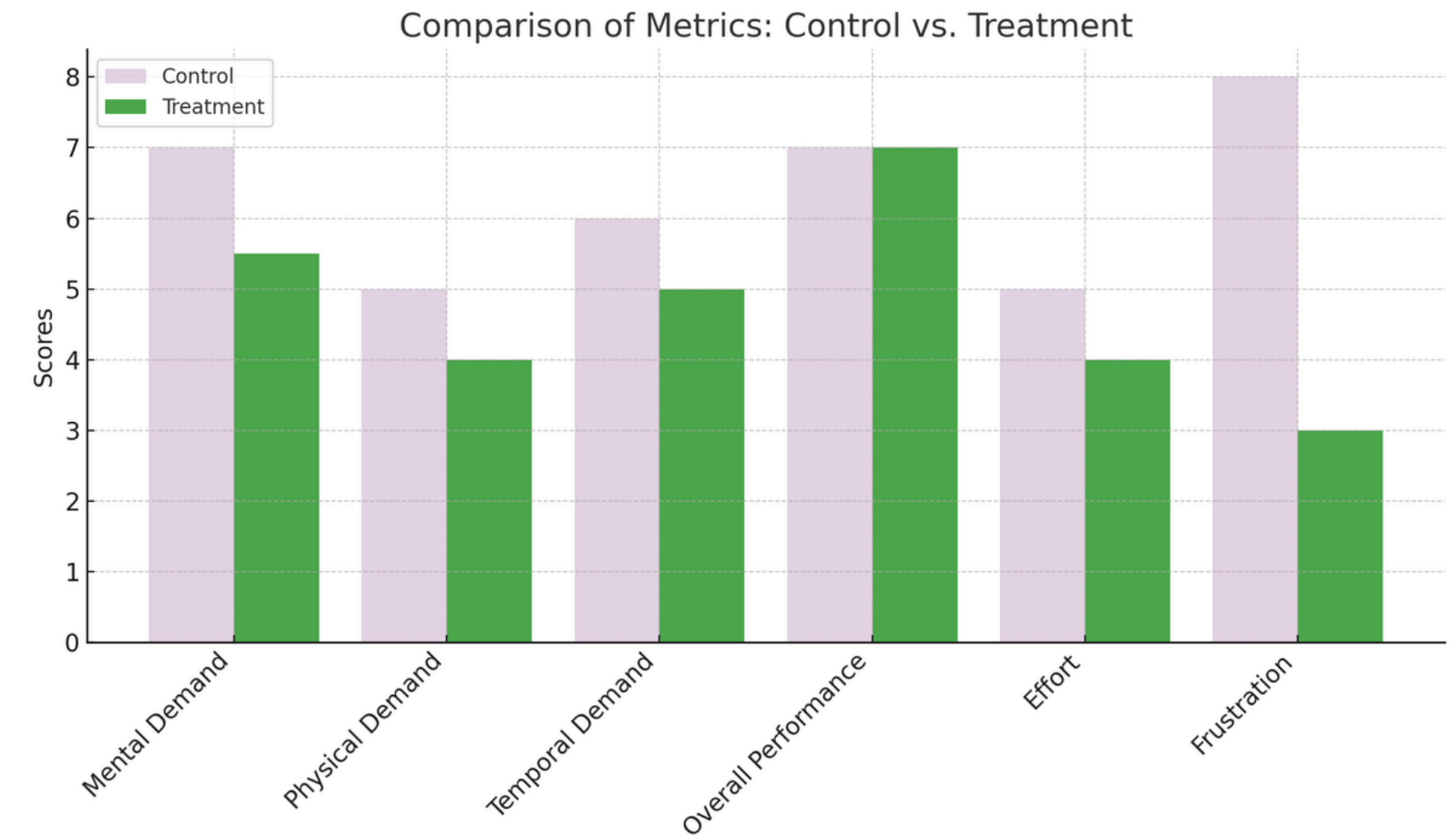


HumanIO [CHI24, Honourable Mention]

Adaptations

I would not mind to have an auto reply, but if it is important call like interview then something better or more safeguard thing I would prefer

6:58 PM



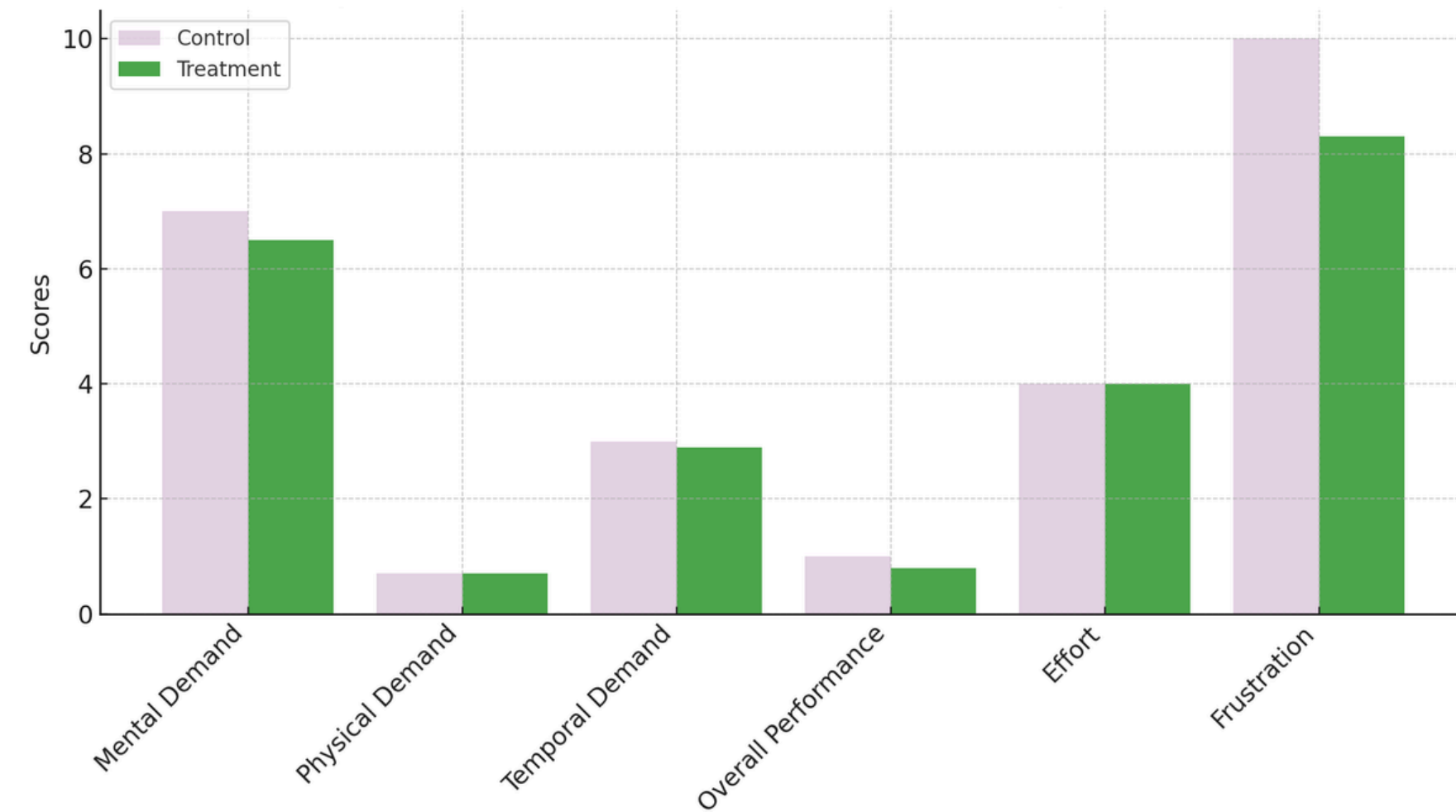
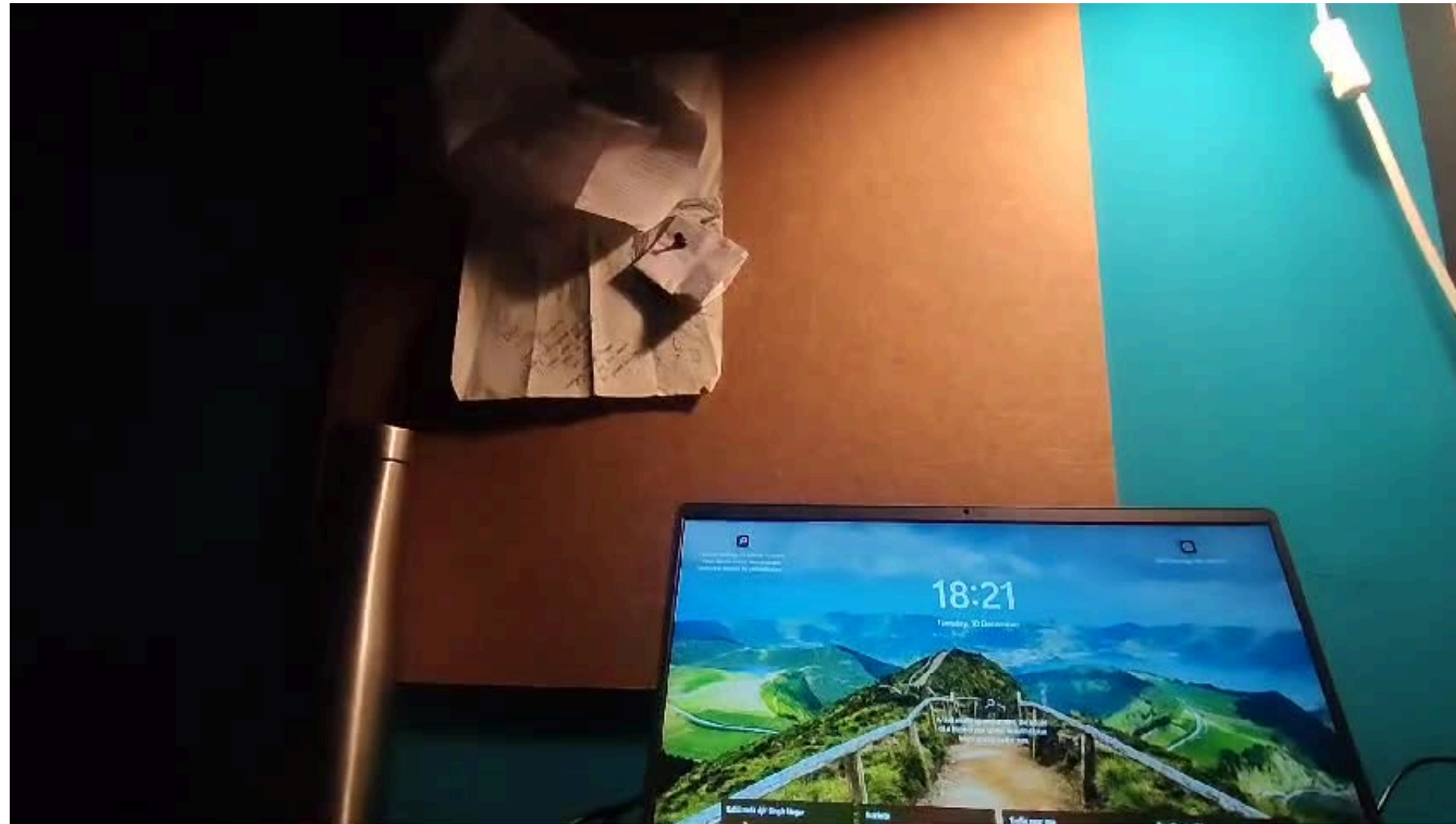
Scenario	Incoming Task	Impaired Channel
Hand Washing	Receive a phone call	Hands/Fingers

Our Adaptations

Speaker voice assist

Adaptations

if im am ina space where I cant pickup calls for example an interview... I would want my phone to ring for a second or two and then move to vibration (if the phone is near me/ on me) if the phone is away then it should ring as normal so that others can picku[

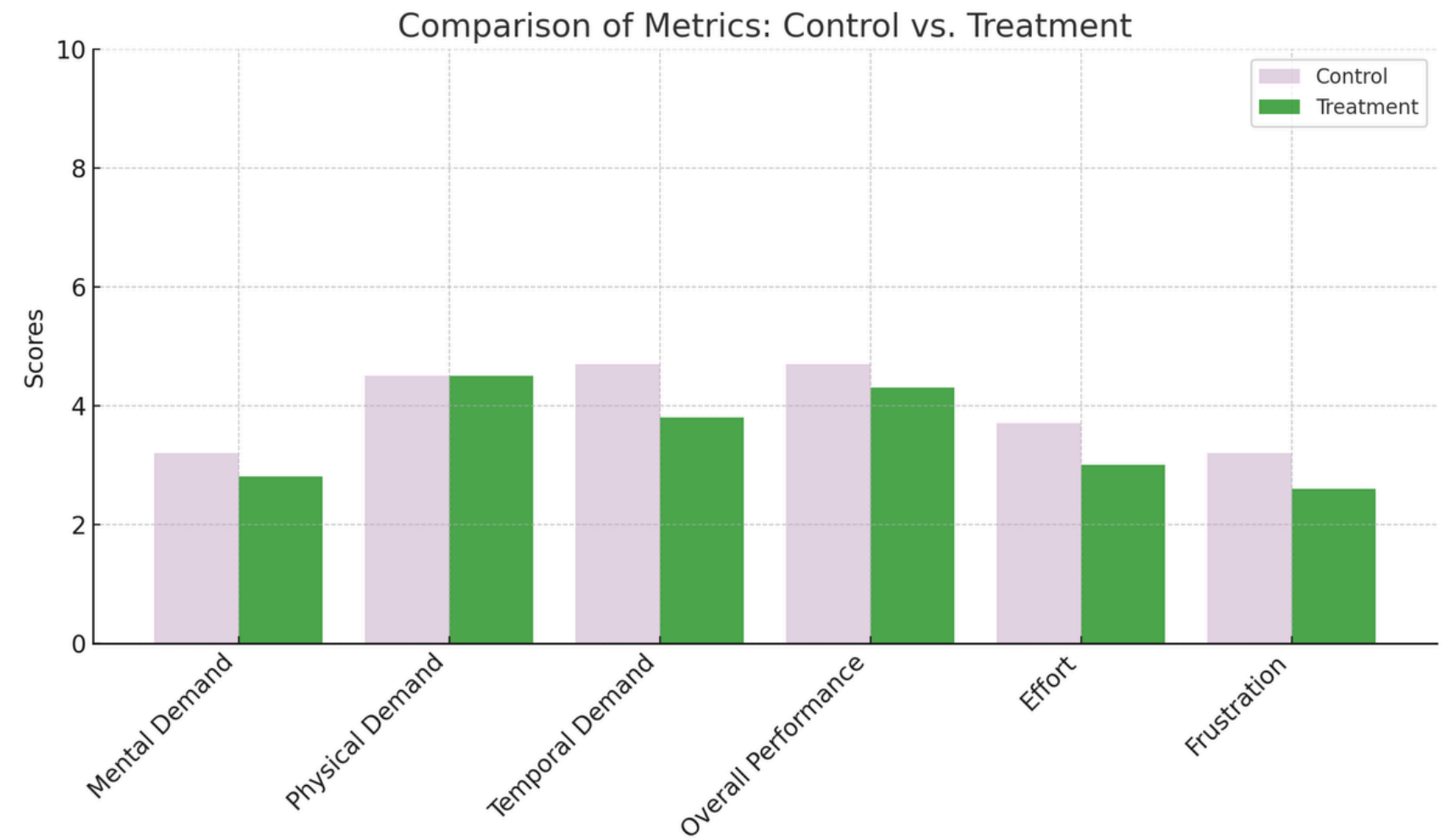


Scenario	Incoming Task	Impaired Channel
Working (Typing)	Receive a text message	Vision/Eyes

Our Adaptations

Ring once and vibrate then

user- I would prefer silent dryer, but voice increase works



Scenario	Incoming Task	Impaired Channel
Hair Drying	Watch a video	Hearing

Our Adaptations

increase volume and start captions

Objective:

Compare the raw human interaction with the environment against our model's intervention in real-time scenarios involving SIIDs.

Evaluation Method:

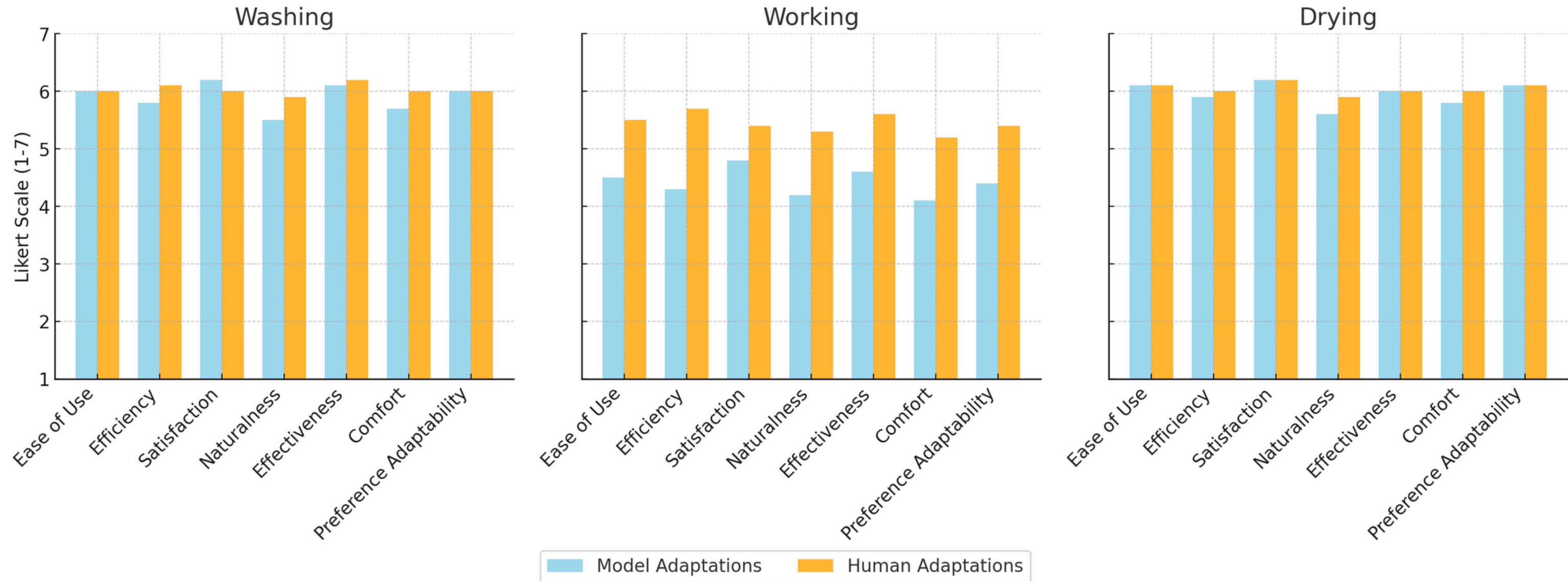
Use the **Wilcoxon signed-rank** test to assess differences between the two conditions.

Significance Threshold:

All rating differences will be tested for statistical significance with $p < 0.05$.

This will let us **confidently determine** if the **model intervention** has a **statistically significant positive or negative effect** on **user experience** compared to **raw human interaction and preferences**

Comparison of Model vs Human Adaptations Across Tasks



```
Washing: WilcoxonResult(statistic=2.0, pvalue=0.13621686984456766),  
Working: WilcoxonResult(statistic=0.0, pvalue=0.015625),  
Drying: WilcoxonResult(statistic=0.0, pvalue=0.10880943004054568)
```

Metrics for Assessment: *w/o Intervention* **V/S** *w Navigator I/o Intervention*

We will adopt the **NASA Task Load Index (NASA-TLX)**, assessing the following on a 7-point scale:

- Mental Demand
- Physical Demand
- Temporal Demand
- Overall Performance
- Effort
- Frustration

NASA-TLX Add languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

The **NASA Task Load Index (NASA-TLX)** is a widely used,^[1] subjective, multidimensional assessment tool that rates perceived workload in order to assess a task, system, or team's effectiveness or other aspects of performance (task loading). It was developed by the Human Performance Group at NASA's Ames Research Center over a three-year development cycle that included more than 40 laboratory simulations.^{[2][3]} It has been cited in over 4,400 studies,^[4] highlighting the influence the NASA-TLX has had in human factors research. It has been used in a variety of domains, including aviation, healthcare and other complex socio-technical domains.^[1] It is a subjective self-reporting set of scores, and is not an objective measure of the Task Load that should be measured using objective metrics that examine the product of the speed and accuracy of users performing a task.

Anticipated Outcomes:

- Like the Human I/O study, we expect our system to significantly reduce effort and improve user experience in real-time interactions.
- We will also evaluate whether our system raises awareness of SIIDs and fosters new interaction possibilities.

Metrics for Assessment: *Human Adaptation V/S Navigator I/O Adaptation*

We will adopt the **Likert Scale Index**, assessing the following on a 7-point scale:

- Ease of Use Efficiency
- Satisfaction
- Naturalness
- Effectiveness
- Comfort
- Preference Adaptability

Anticipated Outcomes:

- We expect our system to show adaptations which are better or at least on the same scale with human preferred adaptations
- We will also evaluate whether our system raises awareness of SIIDs and fosters new interaction possibilities.

Likert scale

[Article](#) [Talk](#)

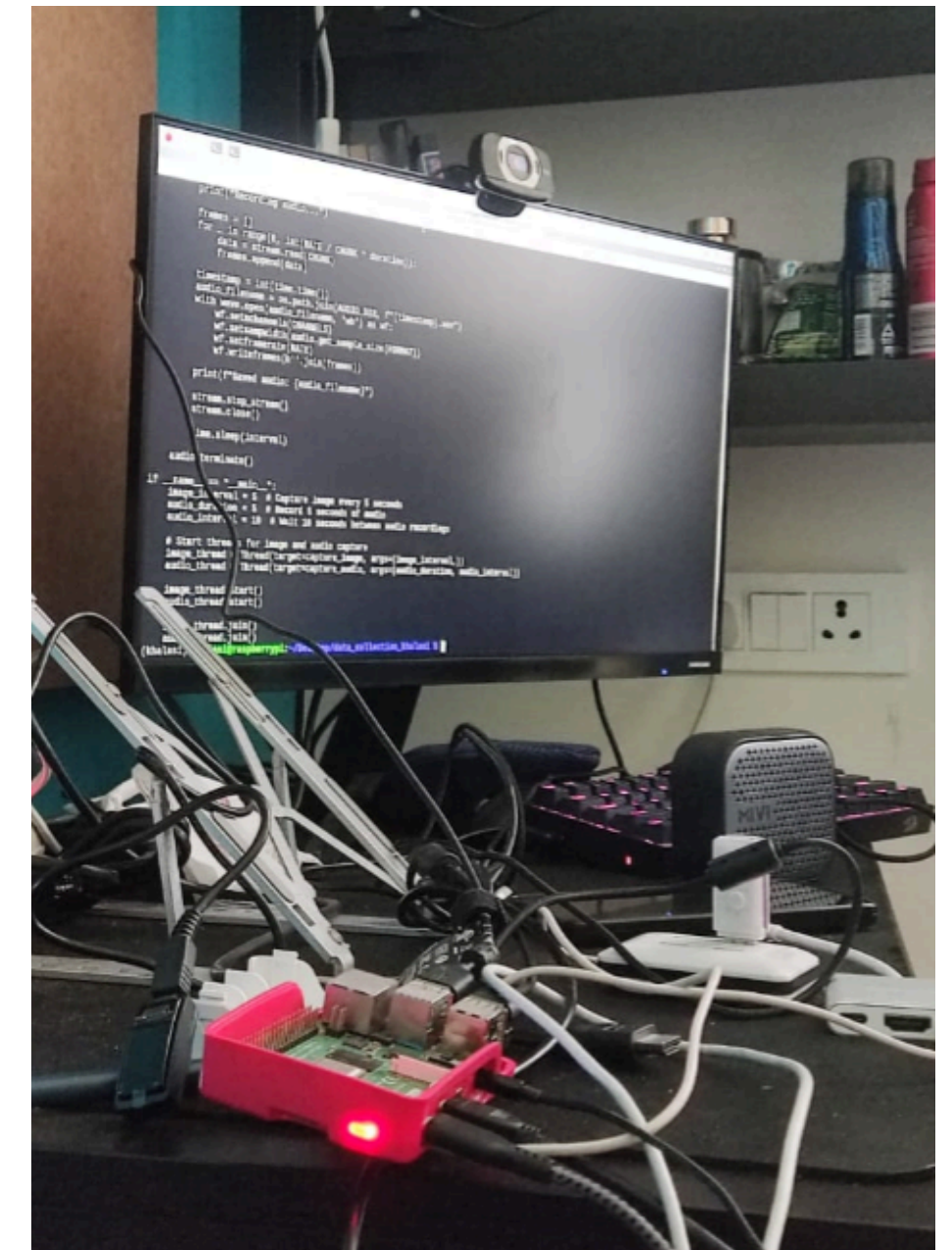
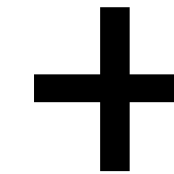
From Wikipedia, the free encyclopedia

A **Likert scale** (/ˈlɪkərt/ *LIK-ərt*,^{[1][note 1]}) is a **psychometric** scale named after its inventor, American social psychologist **Rensis Likert**,^[2] which is commonly used in research **questionnaires**. It is the most widely used approach to scaling responses in survey research, such that the term (or more fully the **Likert-type scale**) is often used interchangeably with *rating scale*, although there are other types of rating scales.

Personality and recommendation diversity. User Modeling and User-Adapted Interaction, Chen, L., Wu, W., & He, L. (2016). 26(2–3), 259–289. Springer International Publishing.

Impact

	hand_left_state	hand_right_state	vision	intent	critical_focus_level
0	holding sink	holding black object	bright, daytime lighting	handwashing	Low
1	wet	holding faucet handle	well-illuminated, natural light	washing hands	Low
2	submerged in water	holding soap	bright, daytime lighting	washing hands	Low
3	wet, holding hands together	wet, holding hands together	well-illuminated, neutral lighting	drying hands	Low
4	holding paper towel	holding metal dispenser	well-illuminated, natural lighting	drying hands	Low
5	reaching for towel	holding towel dispenser	bright, daytime lighting	drying hands	Low
6	placed on counter	near faucet	well-illuminated, daylight	drying hands	Low
7	not-visible	visible, in motion	bright, warm lighting, moderate angle of incid...	moving/walking	Low
8	holding green object	holding red object	well-illuminated, daytime lighting	moving towards destination	Mid
9	bent at the elbow, near the chest	holding [object] (not specified what object)	brightly lit, daytime, with visible sunlight i...	not specified	Low
10	not-visible	holding door frame	well-illuminated, neutral lighting	entering a room	Low
11	not-visible	not-visible	well-illuminated, natural lighting	exploration	Low
12	not-visible	holding keyboard	well-illuminated, neutral lighting	using laptop	Mid
13	hovering over mouse	resting on keyboard	bright, neutral lighting	gaming	Mid
14	typing on keyboard	typing on keyboard	dimly-illuminated, focused on laptop screen	working on coding project	Mid



VR/AR

multimodal detection

personalization

intent

middle-ware

Deployability

- **Improving Latency**

- To use Khalasi-IO real time we need to make sure the system is fast enough.

- **Cost Optimization**

- We are utilizing multiple LLM and VLM instances, hence doing token optimization is crucial.

- **Dealing with Hallucination**

- We are well aware of the hallucination with LLMs and VLMs. Eg. “walking in the kitchen” and “working in the kitchen”.



	hand_left_state	hand_right_state	vision	intent	critical_focus_level
0	not-visible	open	well-illuminated, warm lighting	standing	Low
1	not-visible	holding a metal pole or railing	well-illuminated, natural light	waiting or anticipated unknown	Low
2	not-visible	resting on railing	well-illuminated, daytime lighting	waiting	Low
3	not-visible	holding door handle (silver)	well-illuminated, bright overhead lighting	waiting for someone or something	Mid
4	relaxed by side	gesturing	well-illuminated, natural lighting	presenting or discussing	Mid
5	not-visible	raised, palm facing outward	well-illuminated, neutral lighting	giving presentation in a meeting or event	Mid

LIVE DEMO

thanks !