



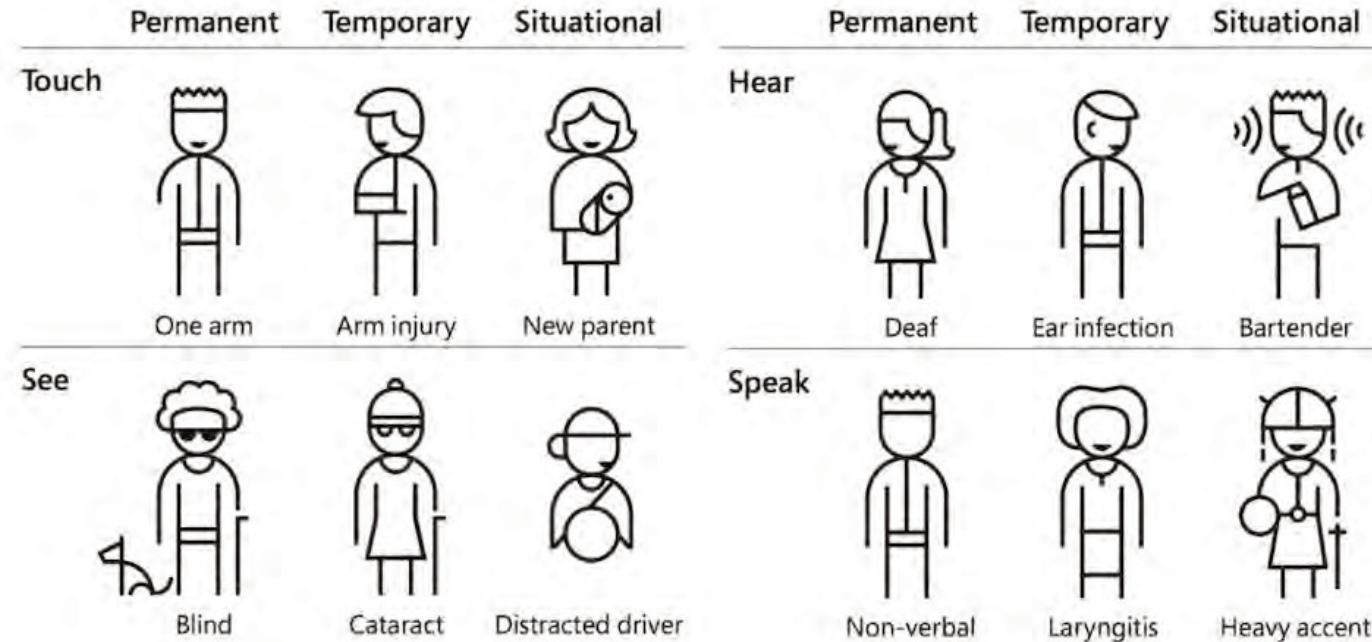
CHI 2024
Surfing the World

Human I/O: Towards a Unified Approach to Detecting Situational Impairments

Xingyu Bruce Liu, Jiahao Nick Li, David Kim,
Xiang 'Anthony' Chen, Ruofei Du

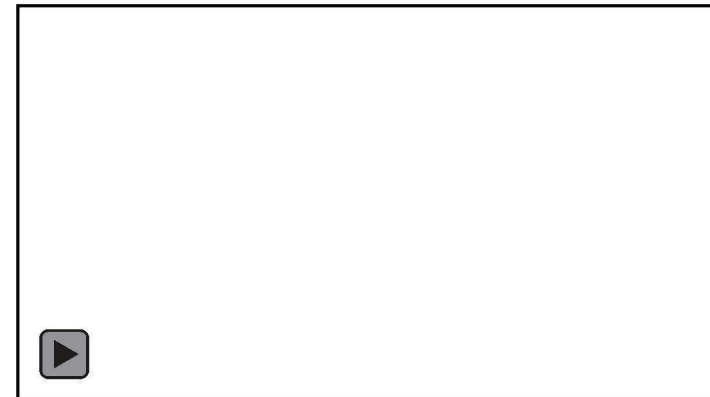
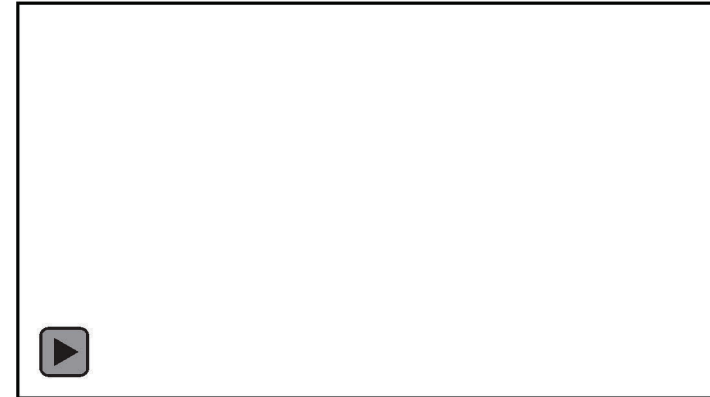
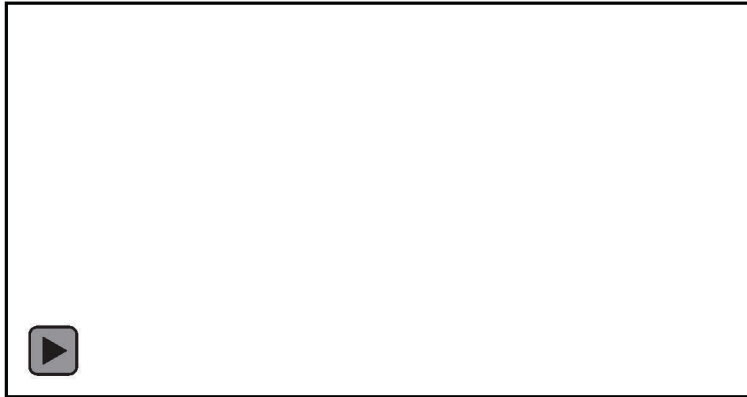
xingyuliu@ucla.edu
liubruce.me

Situational Impairments (SIIDs)



@abbott567

Situational Impairments in Daily Lives



Prior Research: Task-specific SIIDs



SwitchBack (CHI '15)





WalkType (CHI '12)





Driving Mode (Android Auto, Apple Carplay)


However, SIDs are **Dynamic** and **Pervasive**


 Can't see things when getting up at night


 Can't move freely in a crowded subway

 Can't smell when having a cold

 Can't hear things in a loud restaurant

 Can't respond to home assistants when brushing teeth






 Can't use touchscreens when hands are wet

 Can't use Face ID when wearing a mask






SIIDs as Human I/O Channel Availability

Human Input

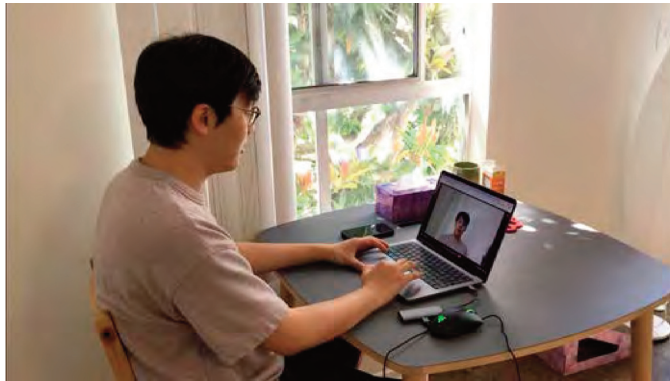
Example use cases

-  Vision *Read text message, watch video, etc.*
-  Hearing *Hear notification, phone call, etc.*
-  Tactile *Haptic feedback, feel temperature, etc.*
-  Taste *Drink, eat, etc.*
-  Smell *Smell, etc.*

Human Output

-  Eyes / Gaze *Face ID, gaze-based interaction, etc.*
-  Vocal System *Conversation, voice assistants, etc.*
-  Hands / Fingers *Touch screen, gesture control, etc.*
-  Limbs / Movement *Walk, reach, etc.*
-  Head / Face *Nodding, facial expression, etc.*

SIIDs as Human I/O Channel Availability



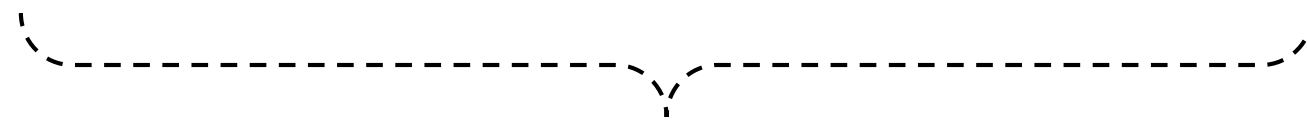
👁️ Vision: *need to make eye contact*



✋ Hands / Fingers: *hands are wet*



👂 Hearing: *noisy hairdryer*



Human I/O Channel Availability

Formative Study



Remote whiteboard session

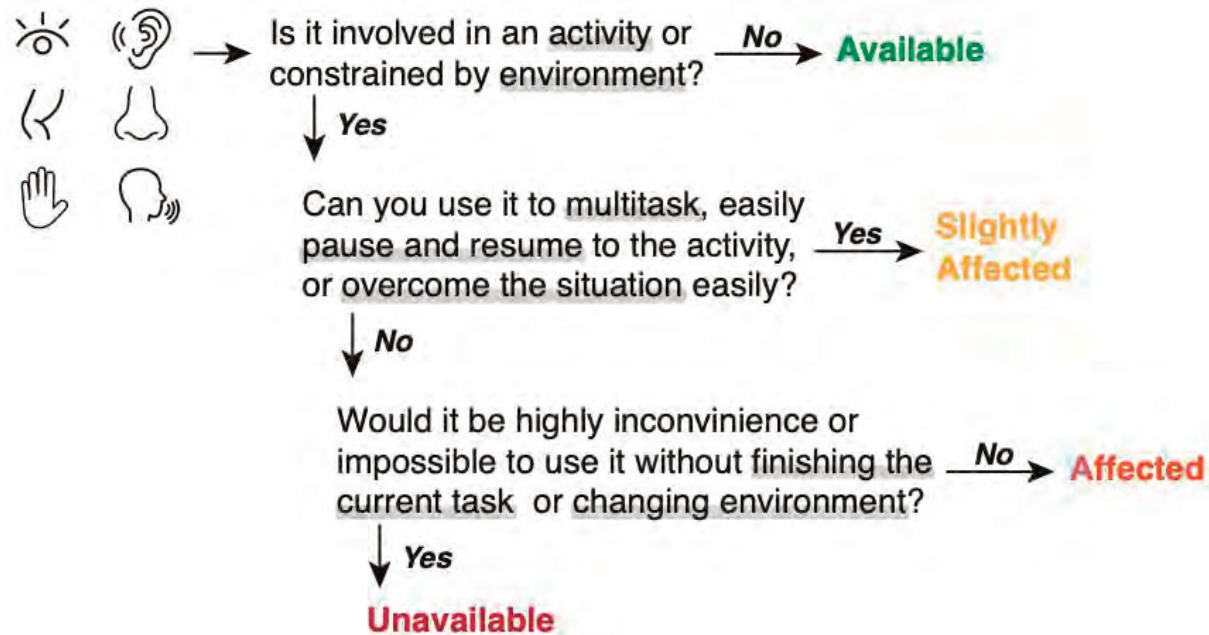


10 participants

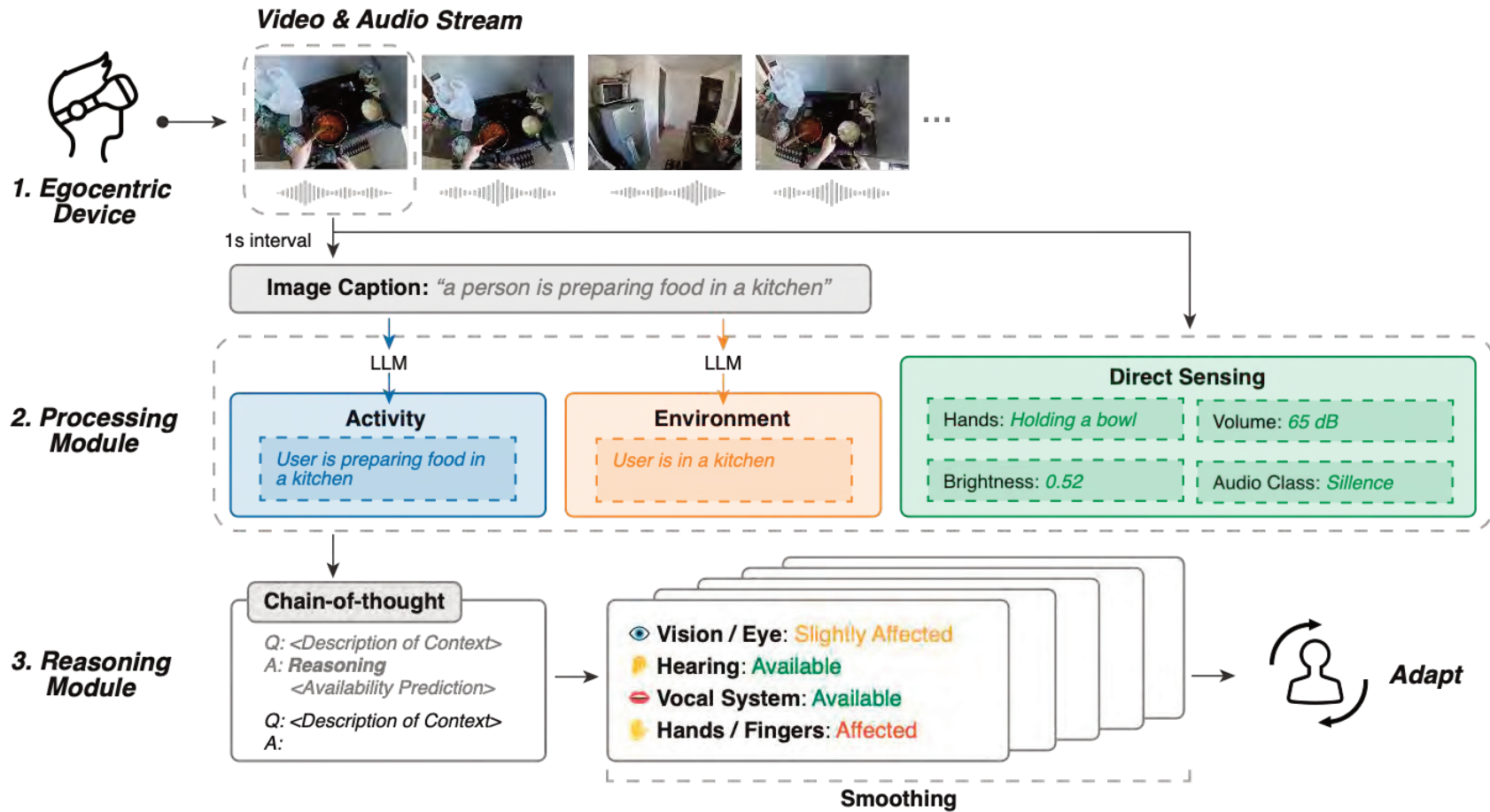
Boot Up Note Taking:
Copy this slide for your brainstorming

Input	<p>Vision</p> <ul style="list-style-type: none"> - cannot read text messages → automatically enable text-to-speech - cannot watch a movie - your eye contact disappear, not polite in a conversation → automatically turn on gaze correction? Or notify the message in a low-res way 	<p>Hearing</p> <ul style="list-style-type: none"> - cannot hear news → turn on subtitles - cannot hear announcement - fire alarm → cannot hear at the regular volume → increase volume / turn on captions or sound event detection - in a noisy restaurant → run live transcription 	<p>Touch</p> <ul style="list-style-type: none"> - cannot feel temperature → measure and display temperature of objects - cannot feel the vibrations of phones → use sound - lose haptic feedback (thus decrease selection accuracy) → switch the feedback to another modality
	<p>Eyes</p> <ul style="list-style-type: none"> - cannot use eye gaze in ar devices → use handtracking instead if it's available - cannot use faceid → automatically switch to password without having to swipe up multiple times 	<p>Vocal System</p> <ul style="list-style-type: none"> - cannot use voice assistants → - cannot talk to other people → automatically generate a message to let people know - cannot make a phone call → auto generate / reply to messages 	<p>Hands / Fingers</p> <ul style="list-style-type: none"> - cannot type → make typing easier - cannot use hand gestures → use body movements - cannot hold objects - cannot reply to text messages with fingers → automatically enable speech-to-text
Output			

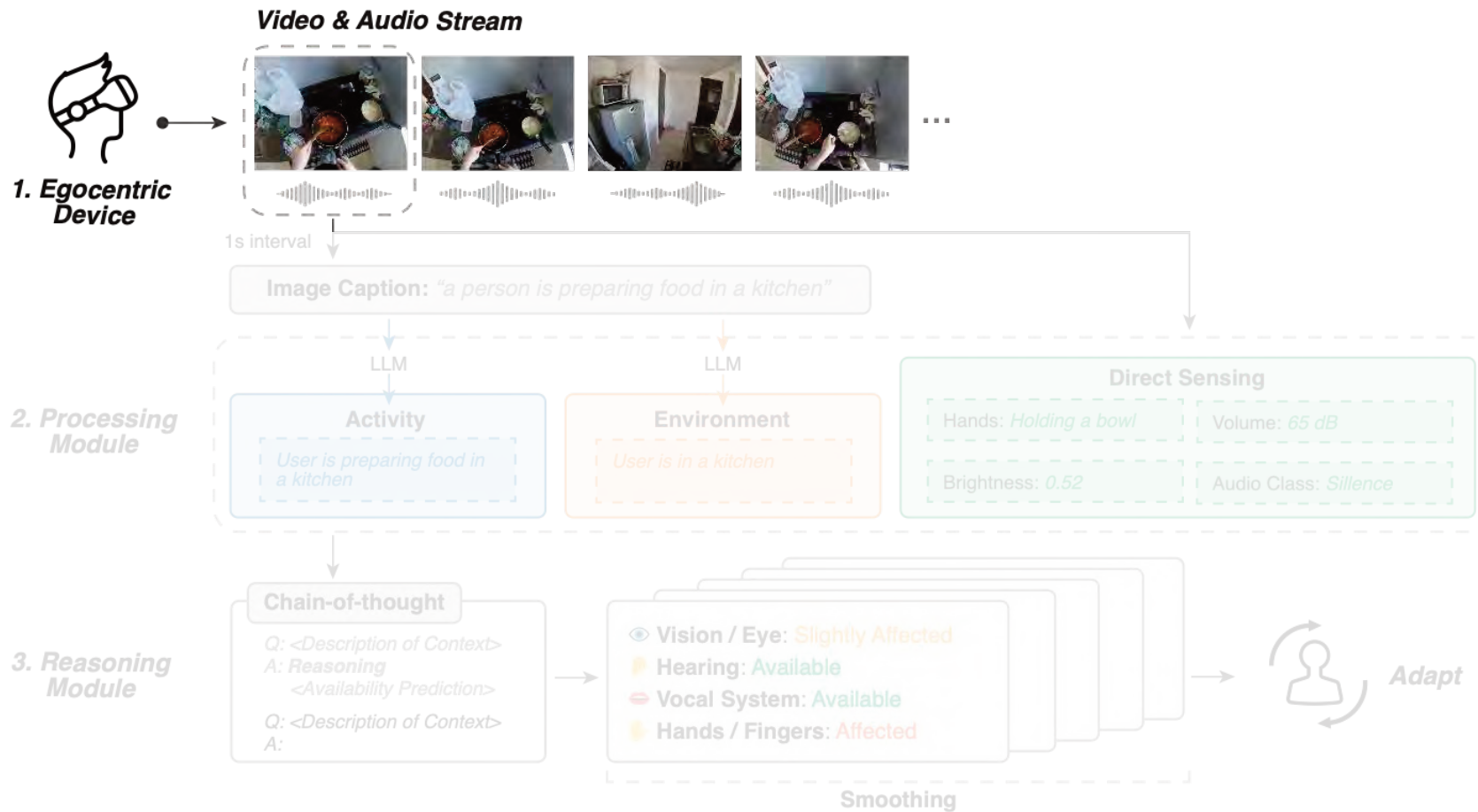
Formative Study



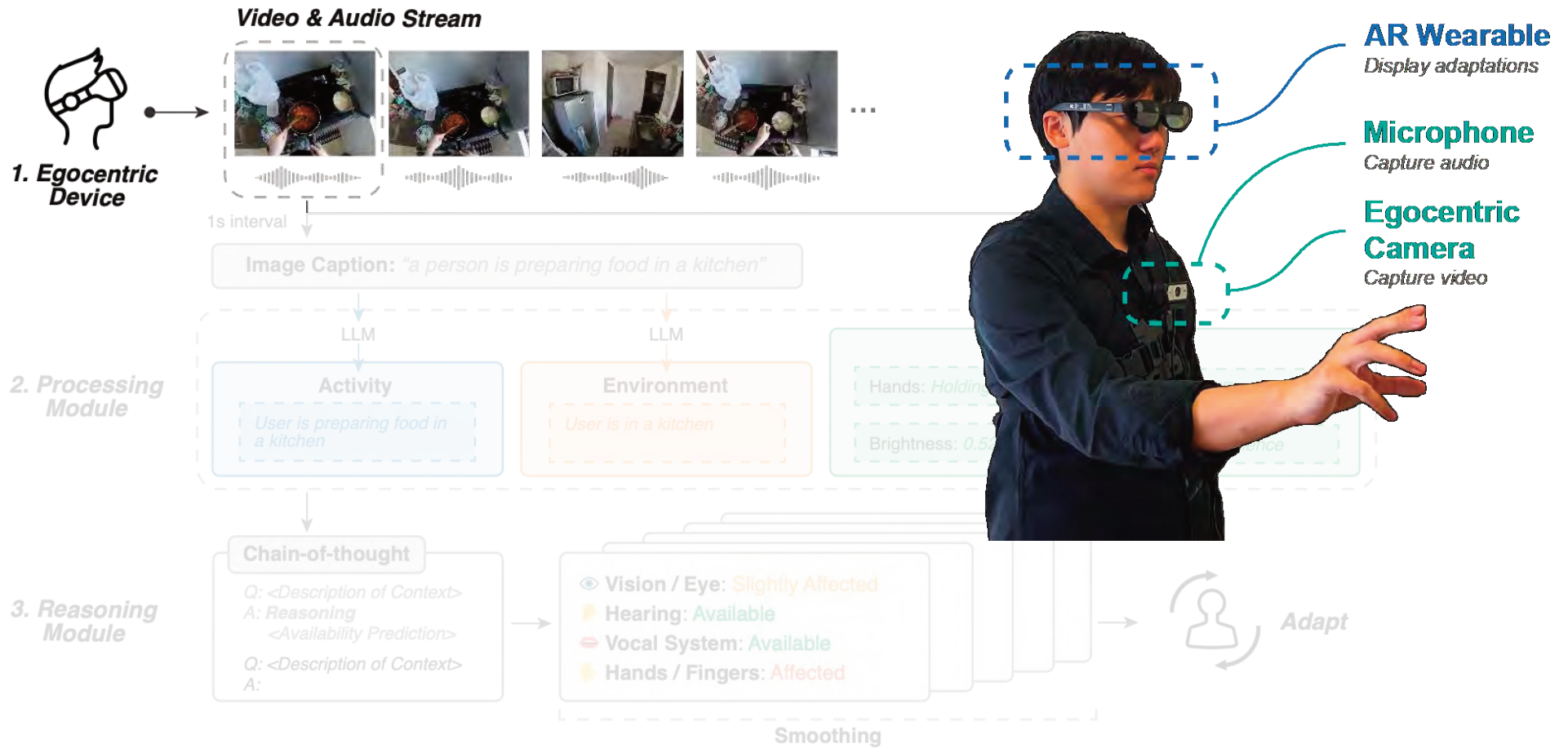
Human I/O System



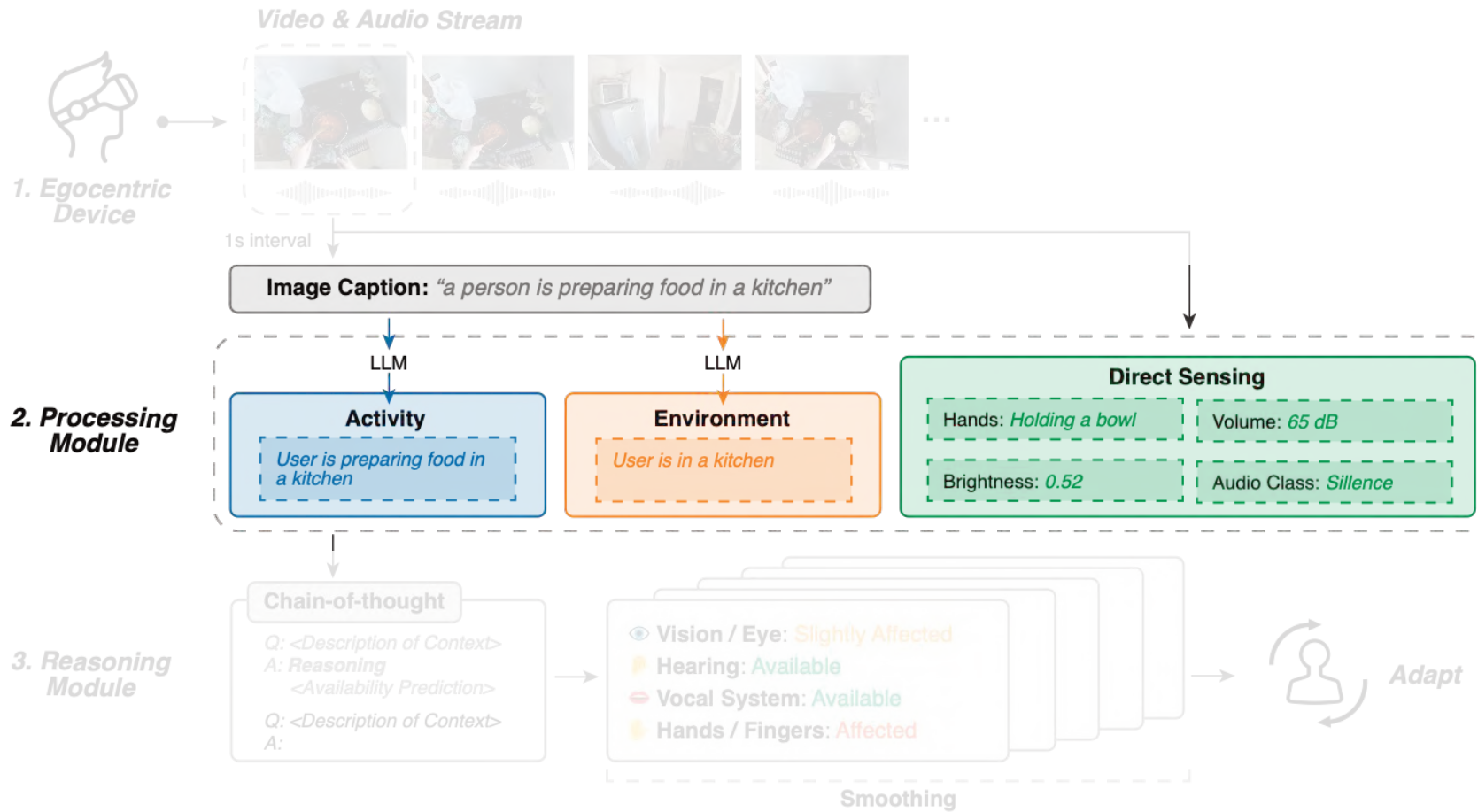
Human I/O: Data Capture



Human I/O: Data Capture



Human I/O: Processing



Human I/O: Processing



Activity

User is preparing food in a kitchen

Environment

User is in a kitchen

Direct Sensing

Hands: *Holding a bowl*

Volume: *65 dB*

Brightness: *0.52*

Audio Class: *Sillence*

Human I/O: Processing

Volume: 61.585

Silence (0.801)
Sound effect (0.008)
Speech (0.004)

Direct - Visual

- Object_Detection
- Hand_Landmar...
- Context_Results

Direct - Audio

- Sound_Level
- Audio_Classifi...

Control

- Interval 5
- BlackCanvas

Simulation

- SimulationTest

Close Controls

bowl 57%

person 59%

Holding: bowl 57%

Caption: a person is preparing food in a kitchen

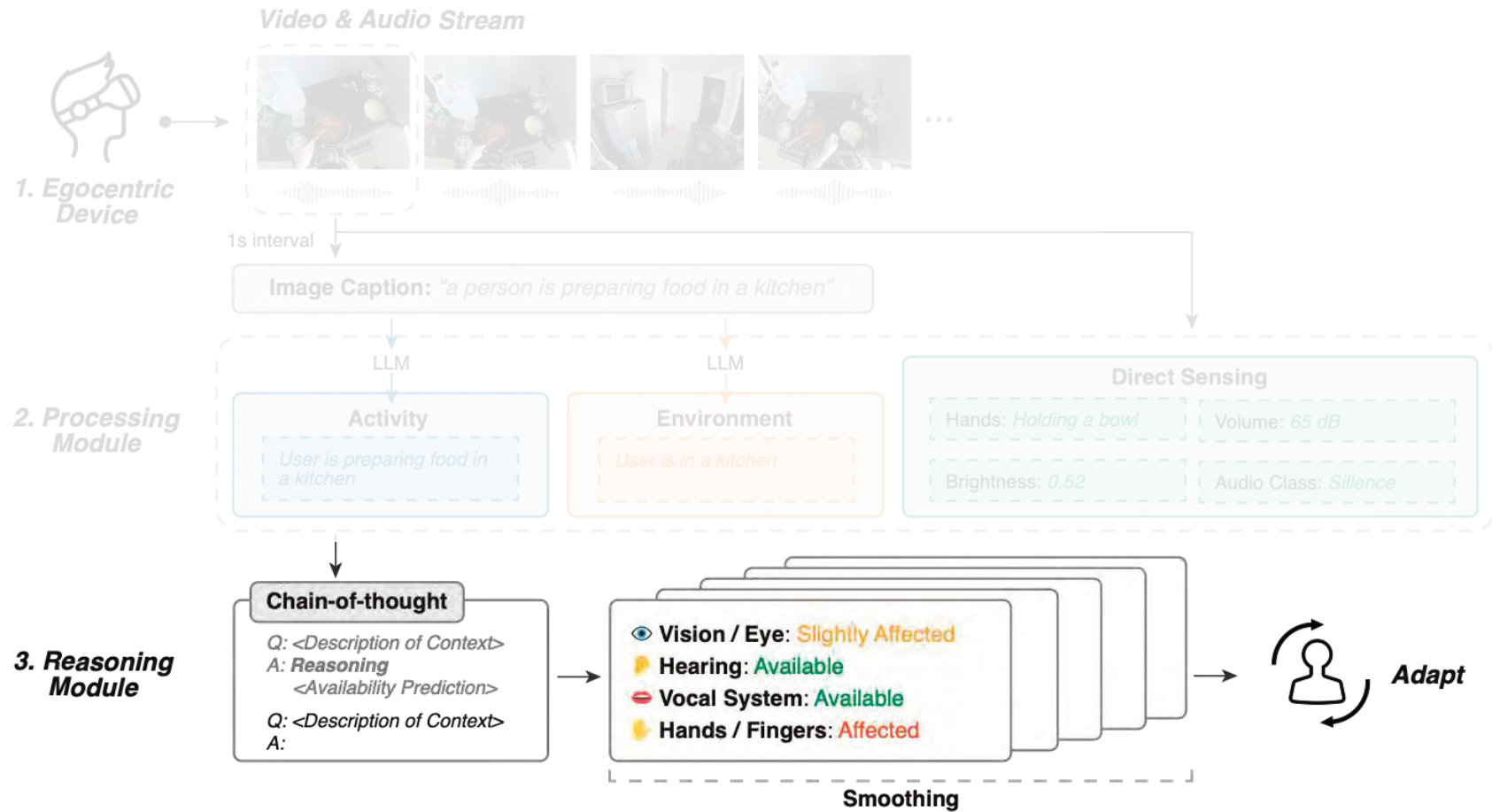
Caption Hand: preparing food

Activity: C is preparing food in a kitchen.

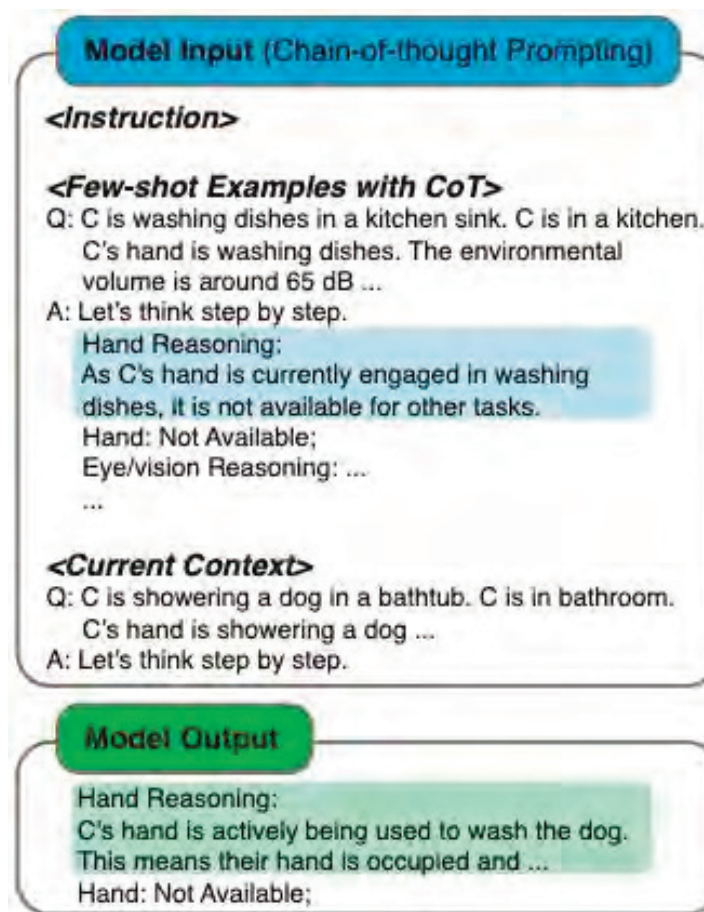
Environment: C is in a kitchen.

- Vision / Eye: *Affected*
- Hearing: *Available*
- Vocal System: *Available*
- Hands / Fingers: *Not Available*

Human I/O: Reasoning



Human I/O: Reasoning



Human I/O: Reasoning



Activity

User is preparing food in a kitchen

Environment

User is in a kitchen

Direct Sensing

Hands: *Holding a bowl*

Volume: *65 dB*

Brightness: *0.52*

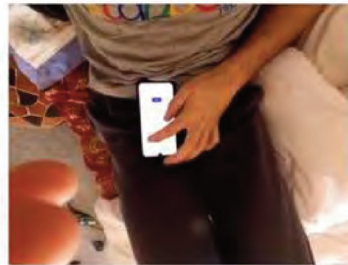
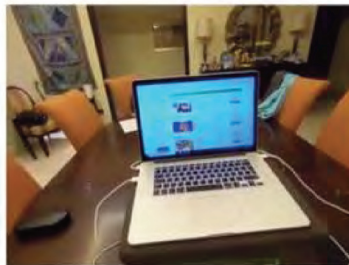
Audio Class: *Sillence*



- 👁️ **Vision / Eye:** Slightly Affected
- 👂 **Hearing:** Available
- 👄 **Vocal System:** Available
- 👋 **Hands / Fingers:** Affected

Technical Evaluation

▶ 300 egocentric video clips (32 scenarios) from Ego4D [1]



Technical Evaluation

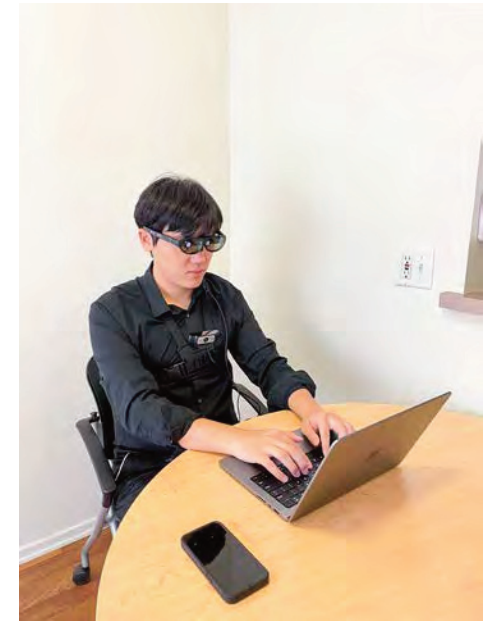
MAE = 0.22

96% of the predictions have a discrepancy of 1 or less

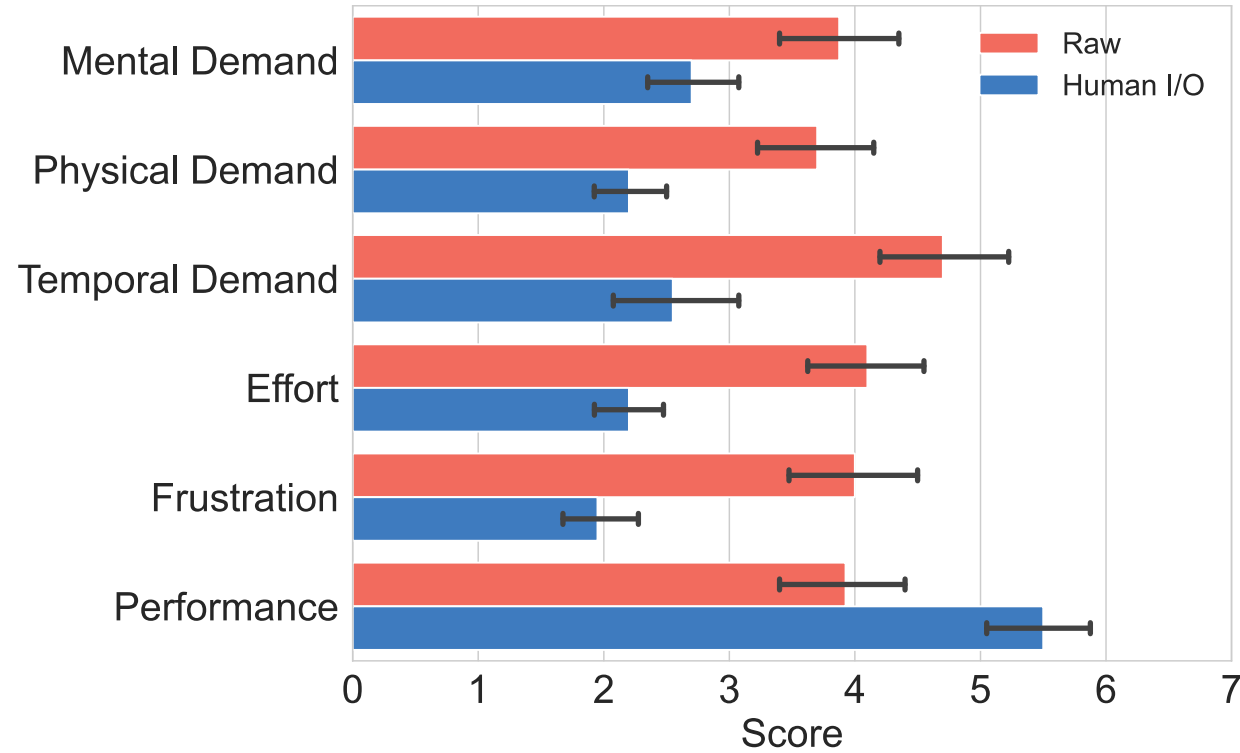
Channels	Human I/O			Human I/O Lite		
	MAE	ACC	VAR	MAE	ACC	VAR
Vision/Eyes	0.25	0.76	0.17	0.45	0.62	0.16
Hearing	0.23	0.87	0.05	0.37	0.63	0.11
Vocal System	0.08	0.92	0.01	0.23	0.87	0.20
Hands/Fingers	0.33	0.73	0.18	0.70	0.47	0.30
Total	0.22	0.82	0.10	0.44	0.65	0.19

User Study

 10 participants



User Study





CHI 2024
Surfing the World

Human I/O: Towards a Unified Approach to Detecting Situational Impairments

Xingyu Bruce Liu, Jiahao Nick Li, David Kim,
Xiang 'Anthony' Chen, Ruofei Du

xingyuliu@ucla.edu
liubruce.me