Teaching Interactively to Learn Emotions in Natural Language

RIT

AWARE-AI NSF Research Traineeship Program

Rajesh Titung[1] and Cecilia O. Alm[2]
[1]Department of Computing and Information Sciences Ph.D, [2]Department of Psychology
Rochester Institute of Technology

Abstract

We conduct a study with an underexplored interactive machine learning method, Machine Teaching (MT), for the text-based emotion prediction task, and compare against a well-studied technique, Active Learning (AL). Results show the strengths of both approaches over more resource-intensive offline supervised learning. Additionally, applying AL and MT to fine-tune a pre-trained model offers further efficiency gain.

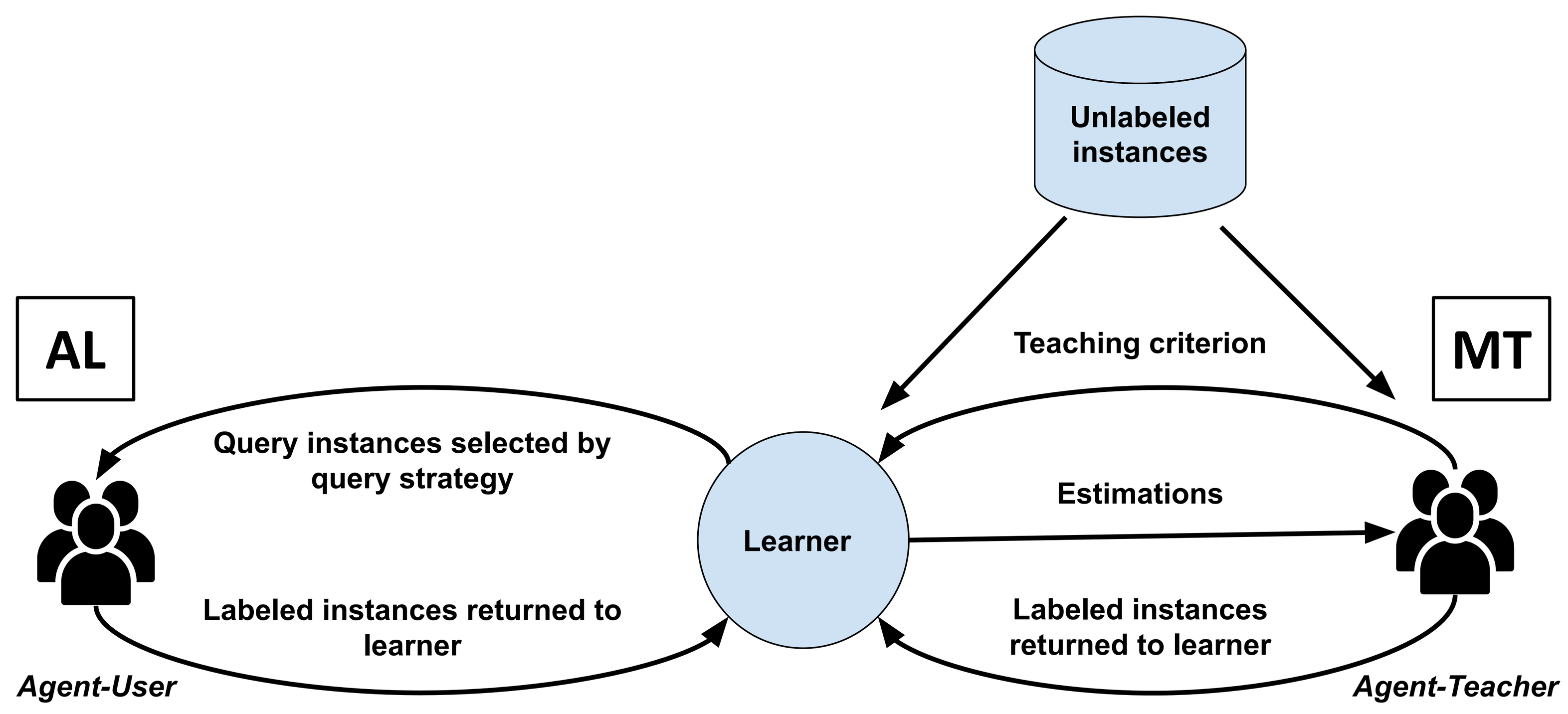


Figure 1. Comparing interactive Active Learning (left) with Machine Teaching (right). Training instances are labeled by the Agent-User (in AL) or the Agent-Teacher (in MT).

Introduction

This controlled interactive machine learning study considers a user or teacher who labels instances, simulating the human decision via dataset lookup.

RQ 1: How do interactive ML (AL and MT) perform in terms of resource-efficiency and compared with offline performance?

RQ 2: Can fine-tuning a pre-trained transformer model in combination with interactive Machine Learning further improve performance, while retaining resource-efficiency?

Method

- Active Learning: Query strategies
  - Least confidence: $x^*_{LC} = \underset{A\subset U,|A|=k}{\operatorname{argmin}} \sum_{x\in A} P_\theta(\hat{y}|x)$
  - Random, entropy
  - Margin sampling: $x^*_{MS} = \underset{A\subset U,|A|=k}{\operatorname{argmin}} \sum_{x\in A} (P_\theta(\hat{y}_1|x) - P_\theta(\hat{y}_2|x))$
- Machine Teaching: Teaching criteria
  - Error-based: $x^*_{error} = \{x|x \in A, A \subset U, \psi(x) \neq y_i\}$
  - Error-based w count*
  - State-change based: $x^*_{state\_change} = \{x|x \in A, A \subset U, y_i \neq y_{i-1}\}$

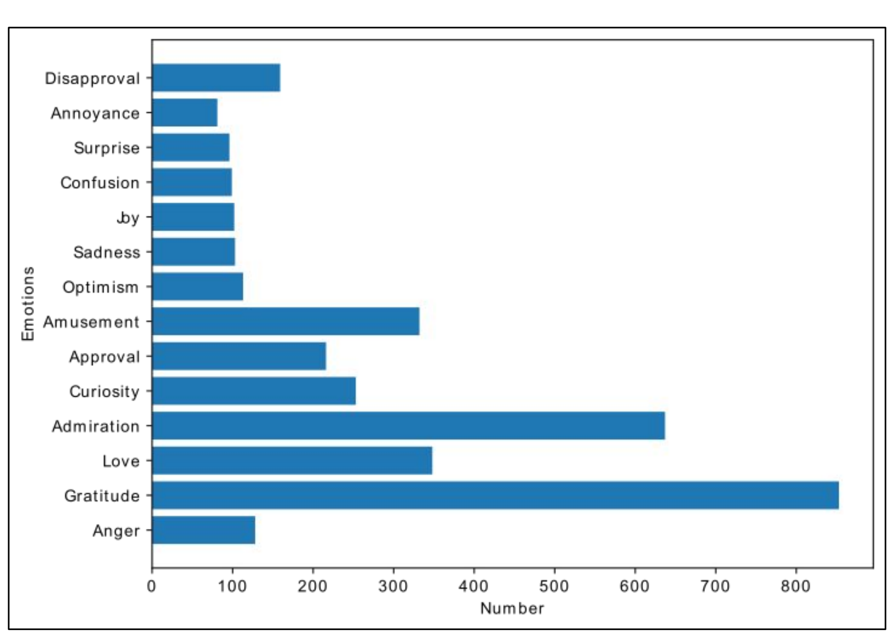
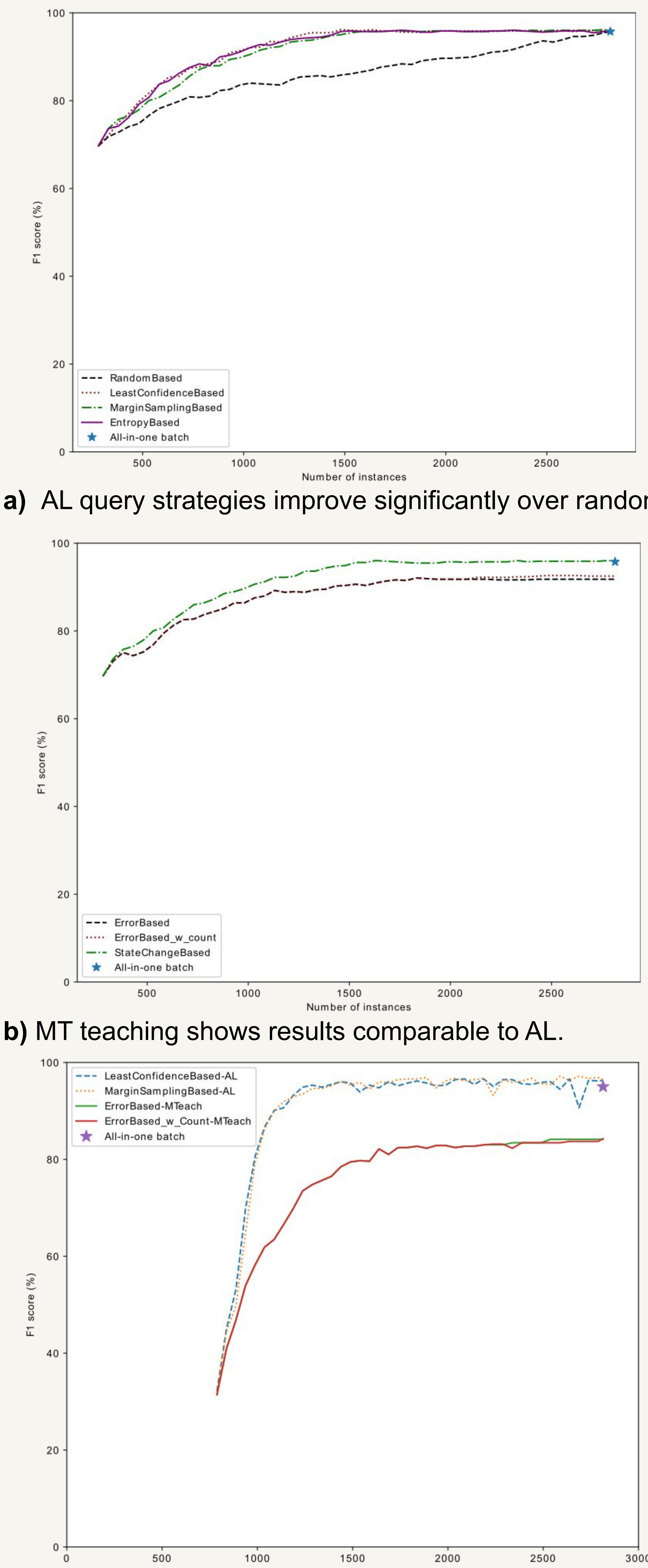


Figure 2. Class distribution for 14-class emotion prediction, using GoEmotions data.

Results



a) AL query strategies improve significantly over random.

b) MT teaching shows results comparable to AL.

c) Fine-tuning combined with iML accelerates convergence.

Figure 3. Text-based emotion prediction with (a) AL query strategies or (b) MT teaching criteria. The all-in-one batch option (green star) signifies resource-inefficient offline batch. (c) AL/MT with fine-tuning a pre-trained model.

Conclusion

- Interactive ML achieves resource-efficient outcomes.
- Combining iML with fine-tuning leverages strengths of both.
- Next directions: teacher variations, teaching criteria, and an integrated mobile framework for user elicitation with iML.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

HUMAN-AI COLLABORATION FOR UX EVALUATION: VISUALIZATIONS AND CONVERSATIONAL ASSISTANTS

CAIR Center for Accessibility and Inclusion Research

EMILY KUANG

PhD Student

Golisano College of Computing and Information Sciences

RIT AWARE-AI NSF Research Traineeship Program

NSF

INTRODUCTION

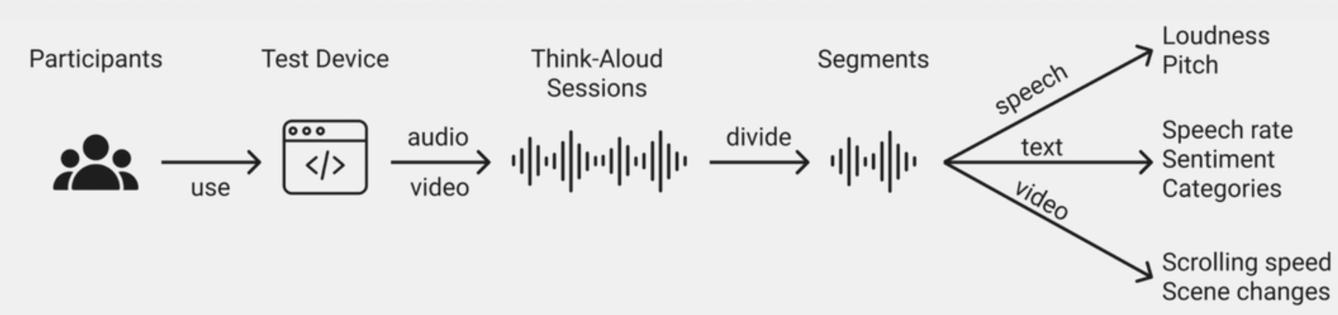
User experience (UX) is central to human-technology interdependence. UX practitioners harness **artificial intelligence (AI)** to conduct extensive usability testing. However, most systems apply an automatic approach.

APPROACH

We view **AI as an assistant** instead of a replacement:

1. **CoUX**: a collaborative visual analytics tool that provides problem-indicators.
2. A web-based tool with a conversational **AI assistant** that provides information on demand.

FEATURE EXTRACTION

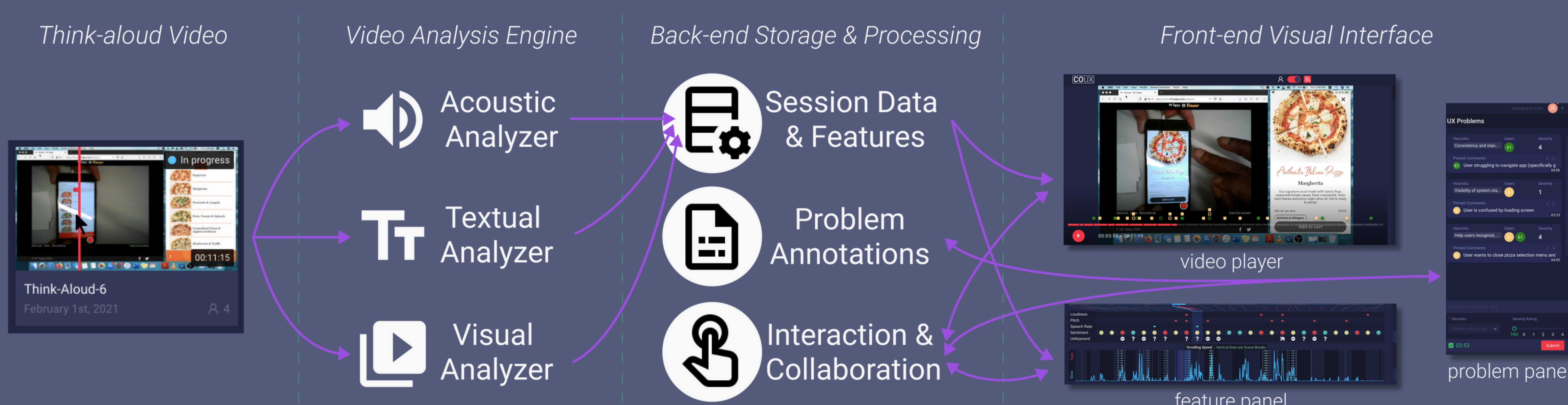


DESIGN PROCESS

1. For CoUX, we conducted **semi-structured interviews** to gather design considerations.
2. For AI assistants, we conducted a **design probe** to collect questions from UX evaluators.

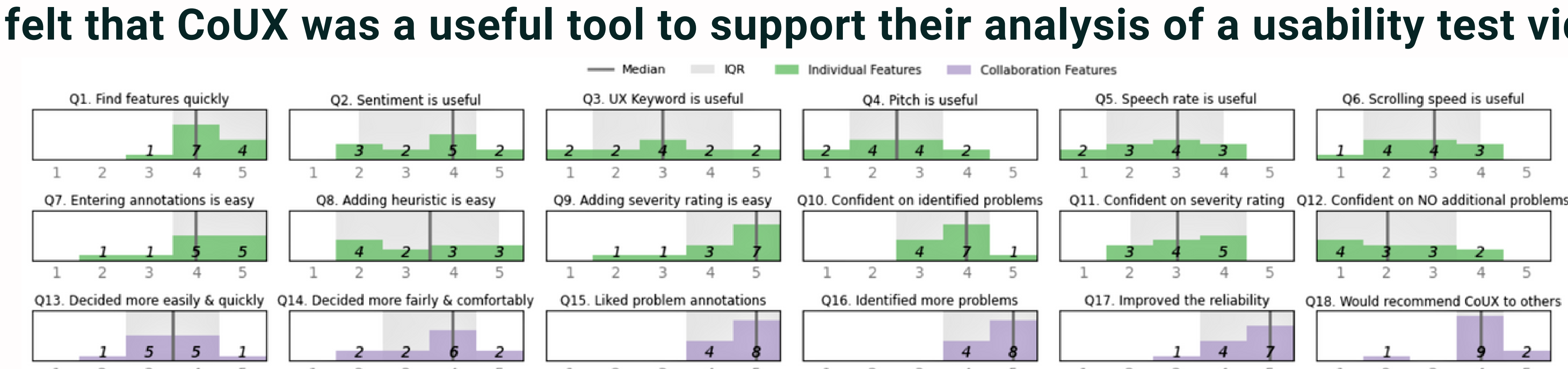
Leverage features from the video to enhance the robustness of problem identification	Provide an integrated environment for both video review and problem logging
Support collaboration with both individual and collaboration modes	Allow for both synchronous and asynchronous communication

SYSTEM OVERVIEW



RESULTS OF USER EVALUATION

Participants felt that CoUX was a useful tool to support their analysis of a usability test video recording:



Feature Panel was used in three ways:

- Hints or warnings
- Anticipations of problems
- Checks or anchors

This research is supported by the National Science Foundation (NSF) Award No. 2125362: AWARE-AI: AWAREness for Sensing Humans Responsibly with AI. Any views presented here are those of the authors alone and do not necessarily reflect those of the NSF.

Temporal Awareness and Anticipatory Performance

Margaret Gray, Abhijan Wasti, and Zhuorui Yong

Advisers: Dr. **Esa Rantanen** (CLA-PSYC) and Dr. **Jamison Heard** (COE-EME)

A Research Framework

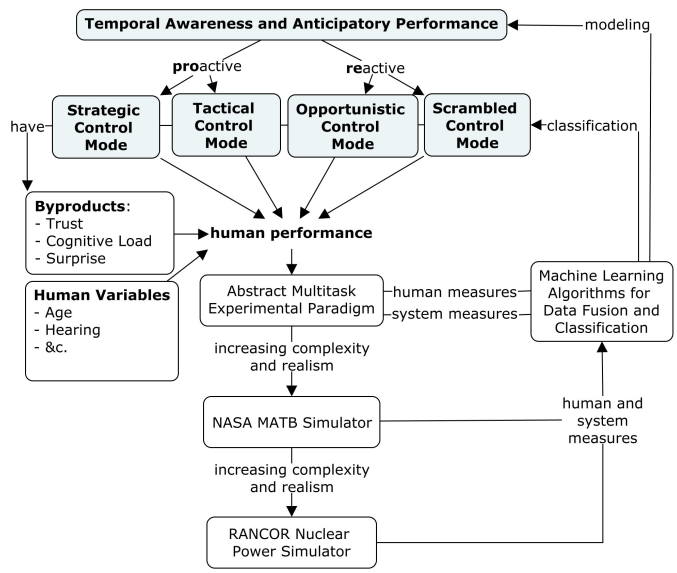
Primary theoretical (cognitive) construct to be modeled:
Temporal Awareness/Anticipatory Performance

Relevance to AI:

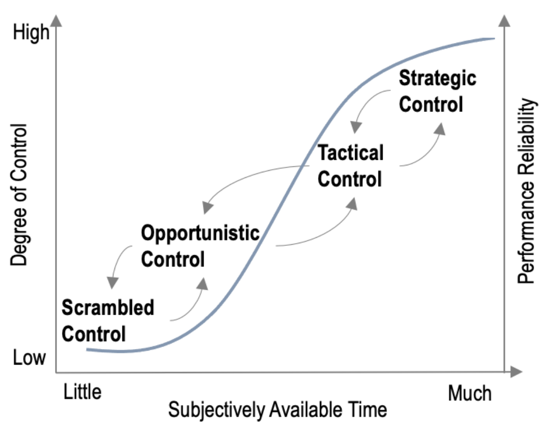
- A window, or pathway, to human mental models of (automated) systems and their dynamics
- Critical to human-automation interactions

Relevance to AWARE-AI NRT RT4:

- Multiple (human subjects) experiments (building blocks to cognitive models)
- Use of machine learning in data fusion (system and human measures)
- Research on transparency of ML algorithms



Hypothesis [1]

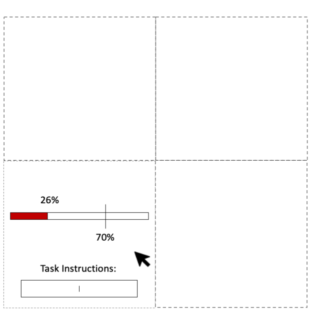


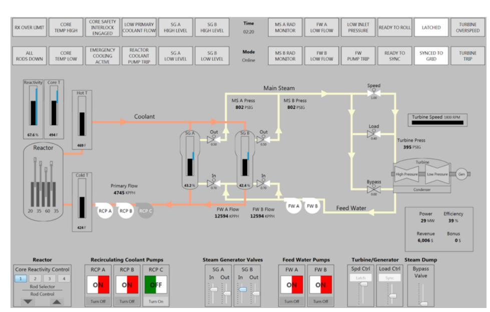
Independent Variables

- Number of subtasks
- Temporal sequencing/ predictability of subtasks
- Secondary tasks (difficulty)
- Reliability of automation (alerts, cueing)
- Subject variables (age, hearing status)

Dependent Variables

- System measures
- Behavioral measures (based on the TA vs. TR paradigm)
- Subjective measures (NASA-TLX, STS)
- Psychophysiological measures





References

[1] E. Hollnagel: *Human Reliability Analysis: Context and Control.* Academic Press, 1993.

Acknowledgments:

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

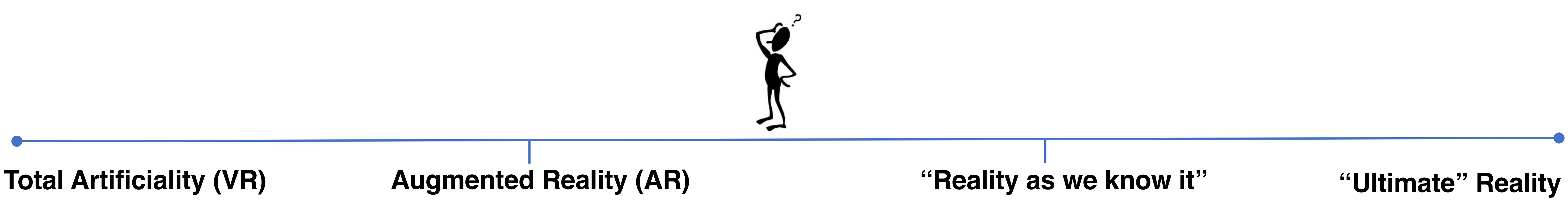
Whither Human Factors in the Age of AI?

Esa M. Rantanen, Ph.D., CPE; Department of Psychology

"For the wise men of old the cardinal problem had been how to conform the soul to reality, and the solution had been knowledge, self-discipline, and virtue. For magic and applied science alike, the problem is how to subdue reality to the wishes of men: the solution is a technique" [1]

Three Problems:

1. Completely opaque AI-based automation makes its actions inscrutable by humans.
   - No training material to the people operating the system.
   - Systems based on ML and AI evolve on their own in time, making past experience with them moot for training new operators.
2. Unintended consequences of widespread applications of AI will be felt on a larger, societal, scale.
   - Past approaches to cognitive systems engineering must be correspondingly "scaled up".
   - Increasing human variability across heterogeneous user groups presents additional challenges.
3. Push away from reality towards artificial/virtual "reality".
   - "Built-in" bias in ML, rise of false positives due to new and "better" machines and diagnostic aids in healthcare [2]
   - "Filter bubbles", and "echo-chambers" created by algorithms [3]



Total Artificiality (VR): Pure fantasy, without any correspondence to "the world as we know it".

Augmented reality (AR): Artificially generated information superimposed on reality as perceived through senses.

Perceived Reality: Unaided perception or perception aided by instrumentation that amplify human sensation.

"Ultimate" reality: Reality beyond human sensing and perception, including metaphysics.

(Traditional) Human Factors

- Human- or user-centered design
- Fitting the task to human capabilities and limitations
- Small, specialized populations: well-trained and highly experienced operators of complex technological systems
- Cognitive work accomplished by humans, with the support of technology
- Technological solutions to human limitations

New Human Factors: Key Questions

- What is a human?
- What is the ultimate purpose of a human?
- Are we using technology for ***human flourishing*** or ***dulling of the human mind***?
- What does "human flourishing" mean?
- Why, or for what purpose, to imitate human intelligence?
- Humanoid robots of ***Intelligent Infrastructure*** (II)
- Do ML- and AI-driven automation applications work?

References

[1] C. S. Lewis, *The Abolition of Man*. Macmillan, 1965.
[2] M. I. Jordan. Artificial intelligence–the revolution hasn't happened yet. *Harvard Data Science Review*, (1.1), 2019.
[3] B. Garfield. The revolution will not be monetized. *IEEE Spectrum*, 48(6):34–39, 2011.

Acknowledgments:

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

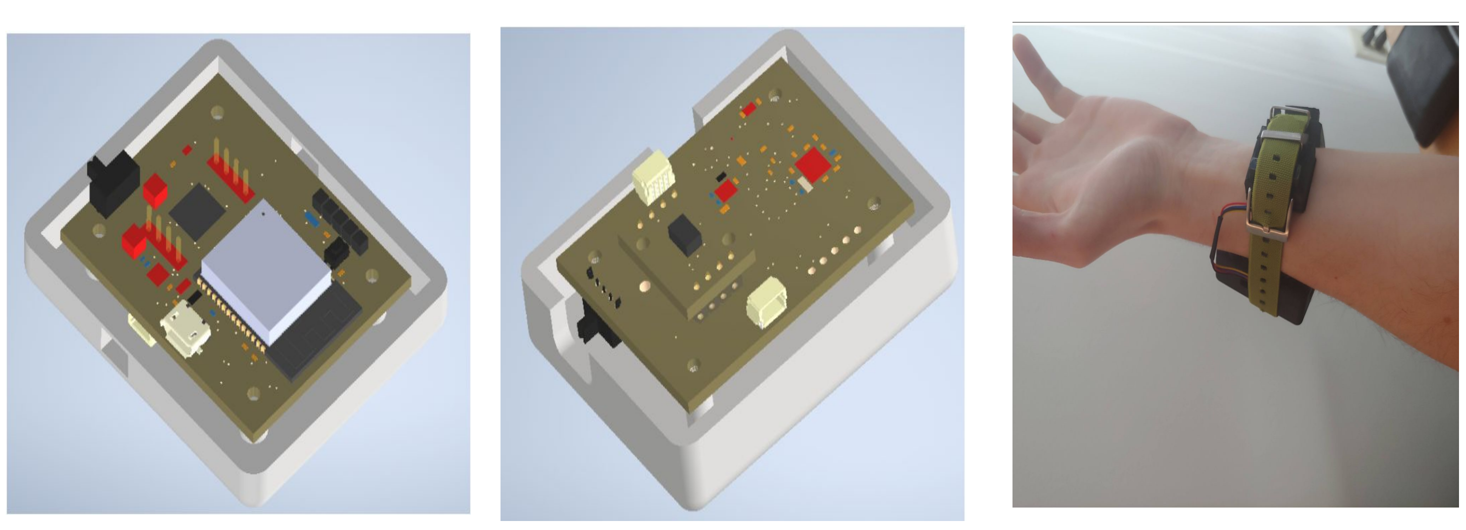
Embedded PPG Device Development for Hand Gesture Recognition?

Espen Peterson, MS. Dept of Electrical and Microelectronic Engineering
Karthik Subramanian, Ph.D Student, Dept of Electrical and Microelectronic Engineering
Ferat Sahin, Ph.D, Dept of Electrical Engineering and Microelectronic Engineering

Embedded PPG Device Development for Hand Gesture Recognition

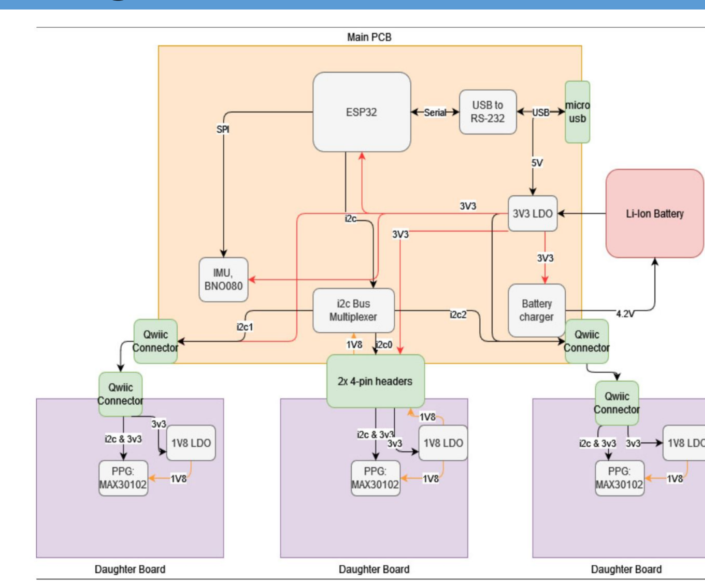
What does it do?

Hand gesture recognition (HGR) has many applications in Human Computer Interaction. Traditionally this has been achieved with the use of electromyography. The creation of the Myo-Armband, which measured the conductance of the user's skin, was one of the first ventures into a modality that did not restrict the motion of the hand itself, and instead rested on the user's forearm. In recent years, many novel approaches for hand gesture recognition have emerged. One such approach utilizes photoplethysmography (PPG) sensors for the purpose of hand gesture recognition (HGR). These sensors are typically used for heart rate estimation and detection of cardiovascular diseases. Heart rate estimates obtained from these sensors are disregarded when the arm is in motion on account of artifacts. Some research studies suggest that these artifacts are repeatable in nature based on the gestures performed. A new wearable device platform is created which contains 3 PPG sensors and a 9 degree of freedom inertial measurement unit to be able to extract these artifacts and hand movements. With use of Machine learning it becomes possible to predict the hand gestures made by users. This device also allows more freedom, as it is built on custom made printed circuit boards (PCBs). This allows researchers to access every piece of information related to how the device is designed and built, and to make incremental improvements if necessary.



Embedded PPG Device Models and realized prototype worn by a user

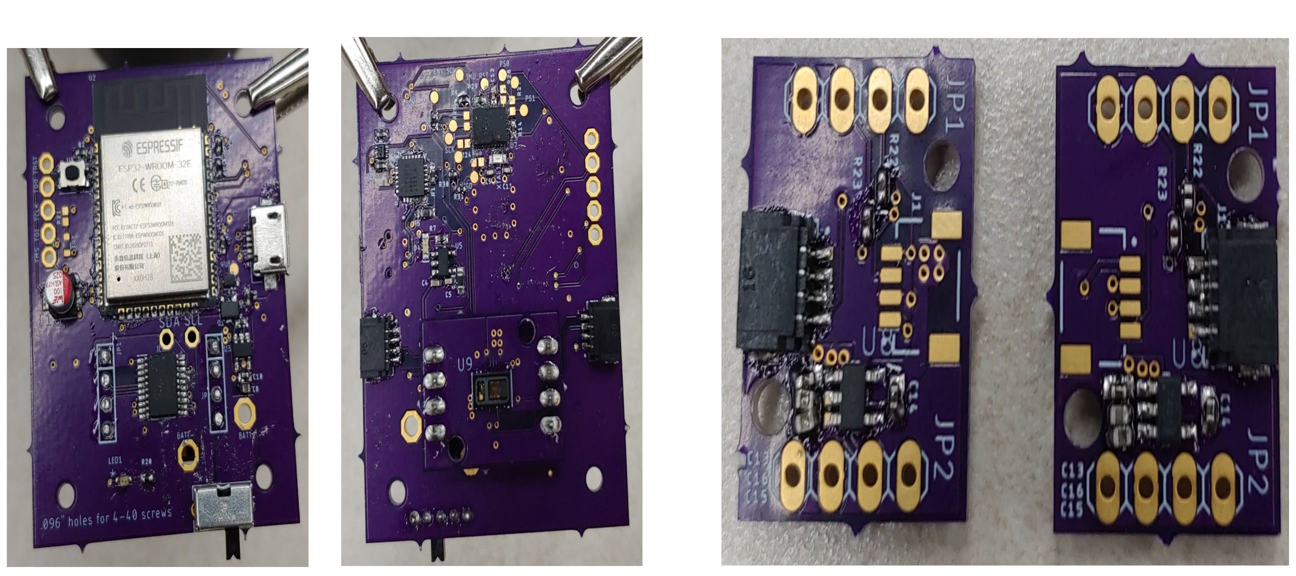
Design: Distributed Architecture



System Block Diagram for embedded Device

The system is comprised of 2 different PCB designs: a main board and a daughter board. Both PCBs are 2 layers. There is the main node, which has the micro controller that communicates with the IMU and the three PPG sensors. The PPG sensors each have a daughter board designed to break them out to either go to a QWIIC connector or the two 4-pin headers. The QWIIC connector is a special connector sold by 'SparkFun' and used on almost all their I2C breakout boards. The PPG sensor used in this project was the MAX30102. In addition to the PPG sensor on the daughter board is a 1.8V linear voltage regulator, which is necessary because the MAX30102 operates off a 1.8 Volt I2C bus. An obstacle presented by the MAX30102 is that it does not have a configurable I2C address. The solution to this was to use an I2C bus multiplexer that also could shift down to 1.8 Volts from 3.3 Volts. The IMU is placed on the main PCB, directly underneath the microcontroller. Both SPI and I2C configurations are available on this design. I2C was ultimately used over SPI, as multiple issues were encountered when trying to use SPI with the BNO080 and its C++ library. An important feature of this design is that when the boards are put together, nothing on the same face as the PPG sensor will be taller than it. This will minimize the reflections as the light travels from the sensor and hits the skin. The black matte solder-mask of the daughter board will also serve a similar purpose of preventing the light from reflecting after it has reflected off the skin.

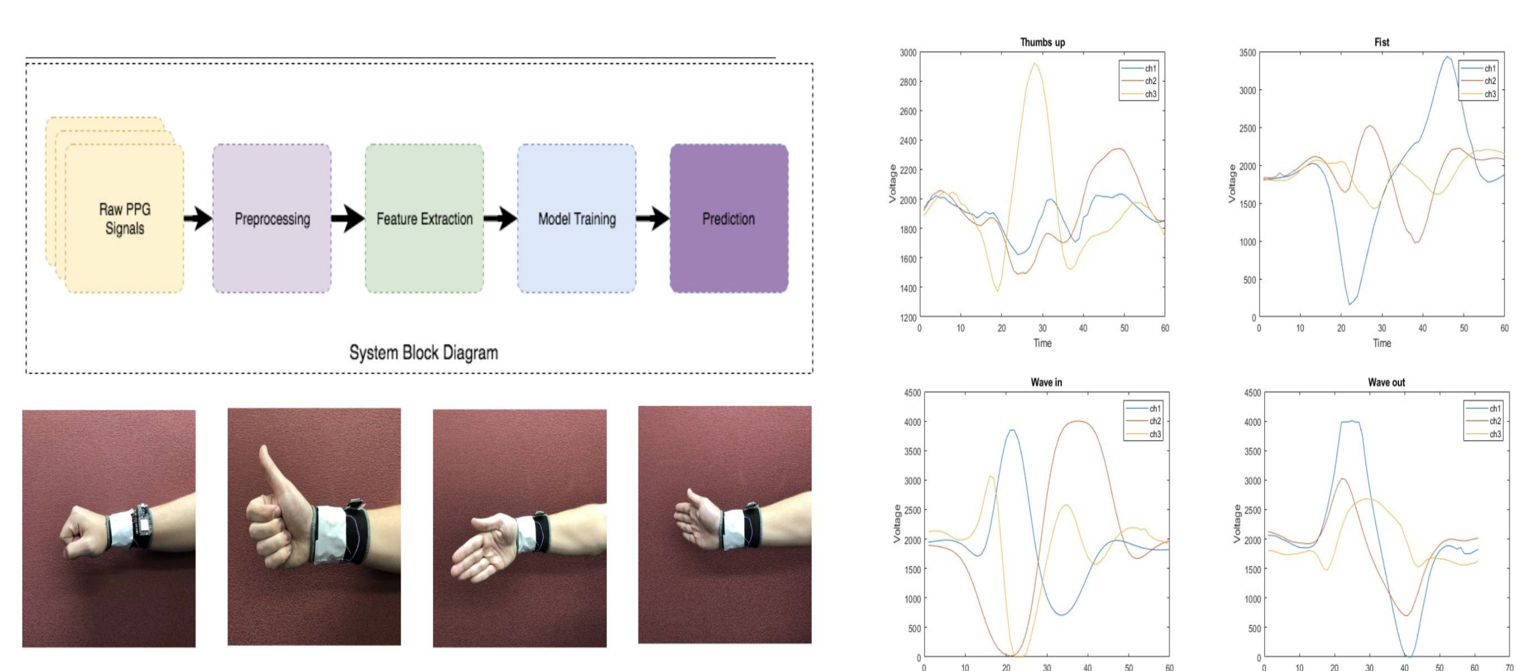
Performance and Limitations



Populated PCB main boards and Daughter boards

It was found that sampling the PPG devices could be done at a maximum rate of about 400 HZ, . It was found to be able to sample at around 500 HZ for the Accelerometer. The maximum data rate achieved with the Bluetooth link was about 170 Hz. With the current setup, 9.8% of the RAM and 73.7% of the Flash memory on the chip are used. This leaves about 295 kilobytes of RAM, and 344 kilobytes of flash memory. In its current state, it could deploy small trained models to predict a limited number of hand gestures as an Edge device. More complicated models will require streaming and evaluating the predictions using a distributed system, where the device is only responsible for the data stream.

Learning and Evaluation Methodology



1.Close Fist 2.Thumbs Up 3.Wave in 4.Wave out

Recorded PPG Signals from Hand Gestures performed

Raw data is obtained from the device. It is followed by standard pre-processing technique. Relevant features are obtained by applying feature extraction methods. Supervised learning classifiers ca be trained to generate a model with these obtained features. The then trained models will be used to make predictions on unseen test data not utilized in the training process. Using these methods, it is possible to detect hand gestures made by relaying PPG and IMU streams. Once hardware limitations have been lifted, it could be possible to evaluate the predictions of the trained models on device itself.

Acknowledgments: *This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.*

Using Deep Learning to Increase Eye-Tracking Accuracy and Precision in Virtual Reality

RIT

AWARE-AI NSF Research Traineeship Program

Kevin Barkevich, Dr. Gabriel Diaz (Imaging Science), Dr. Reynold Bailey (Comp. Science)
Golisano College of Computing and Information Sciences (GCCIS)
Rochester Institute of Technology

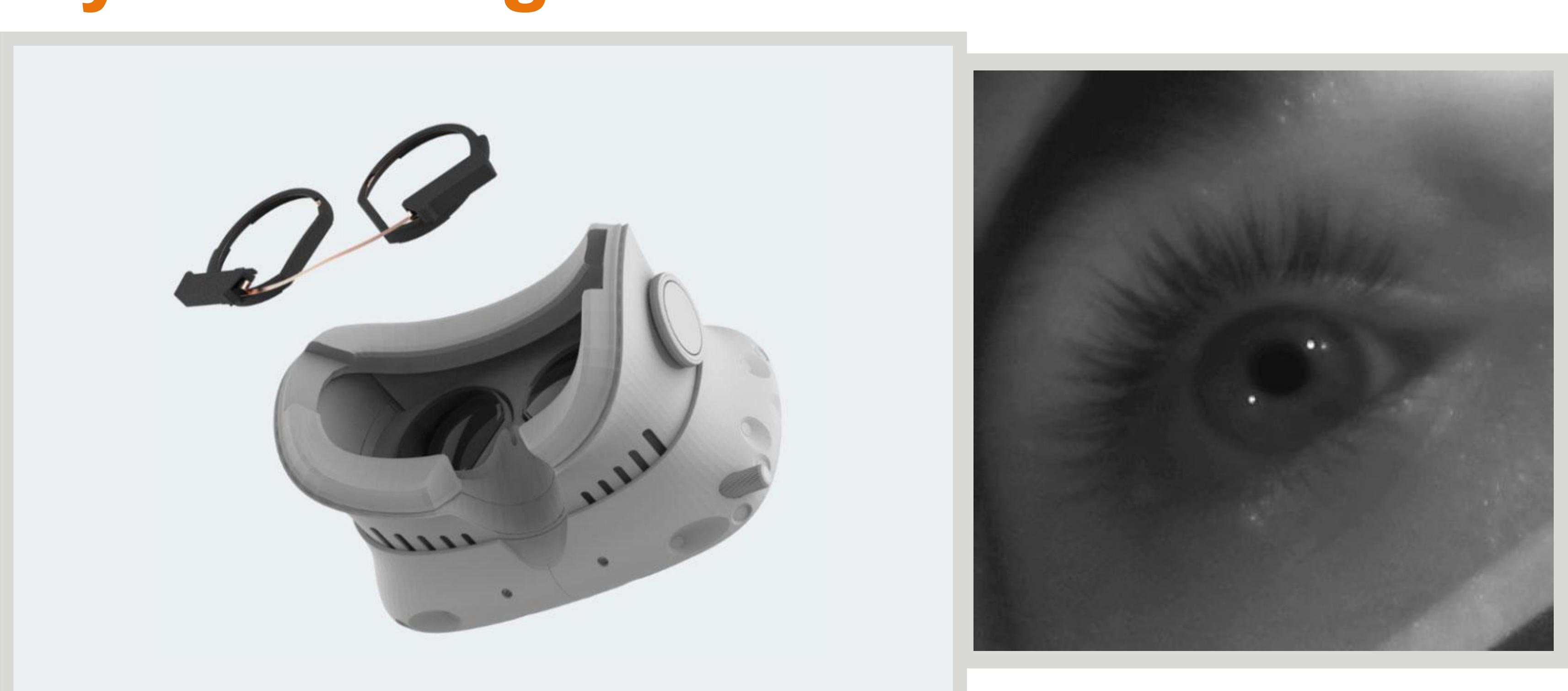
Motivation

Eye-tracking in a virtual reality environment is an ongoing topic in research, both as a subject to be researched and a tool to be used for researching other topics. Eye-tracking solutions based on traditional computer vision techniques, especially when used inside a virtual reality headset, can perform poorly due to the difficult and sometimes inconsistent environment around the wearer's eyes. Problems such as extreme camera angles, inconsistent lighting caused by changing screen content, and corneal reflections caused by infrared lighting all contribute to the difficulty of traditional computer vision-based solutions.

Approach

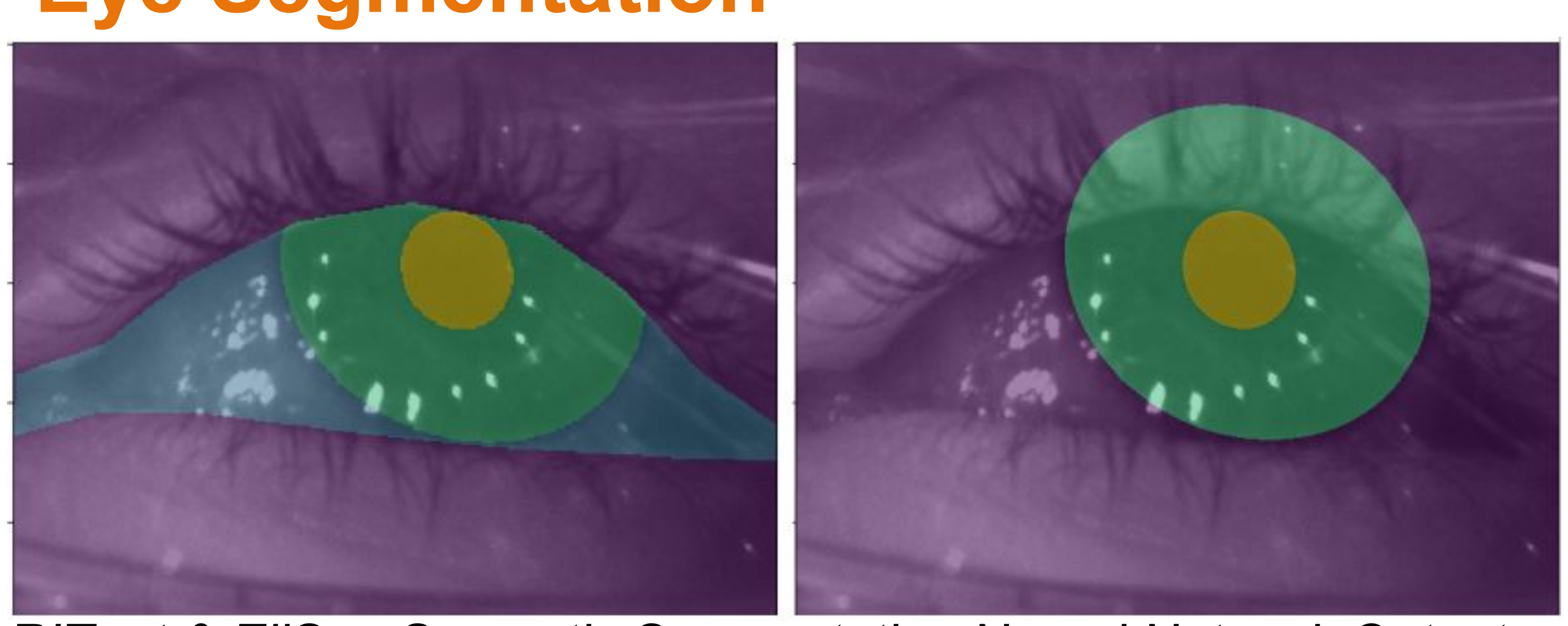
We choose to evaluate the influence of semantic segmentation on the gaze estimate using the Pupil Labs HTC Vive Pro integration rather than the Pupil Labs Core tracker (mobile with glasses-like form-factor) because the use of virtual reality (VR) allows for the controlled presentation of gaze targets. Additionally, the environment inside of a VR headset presents unique difficulties regarding lighting and camera position that we hope to alleviate using our segmentation approach.

Eye-Tracking Hardware



Pupil Labs HTC Vive Pro Eye Camera Insert & Example Eye Frame

Eye Segmentation

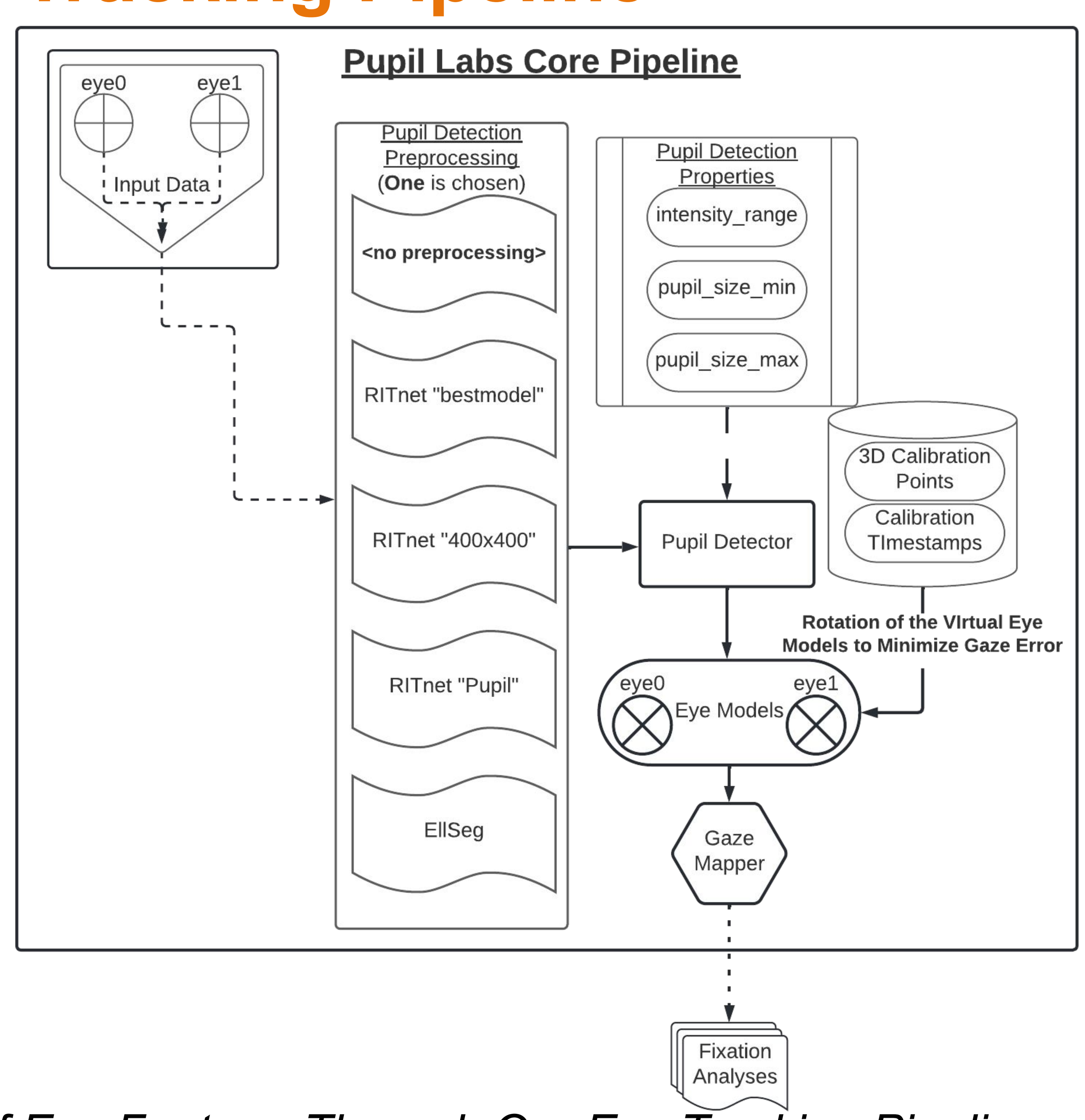


RITnet & EllSeg Semantic Segmentation Neural Network Outputs

Objectives

We are exploring the use of deep neural networks as a pre-processing step to improve the accuracy and precision of the gaze output of commercially available eye-tracking pipelines. We are focusing on the pupil as a trackable eye feature. By comparing multiple different semantic segmentation-based pupil detection methods, we hope to demonstrate the real-world effectiveness of neural networks for use in eye-tracking.

Eye-Tracking Pipeline

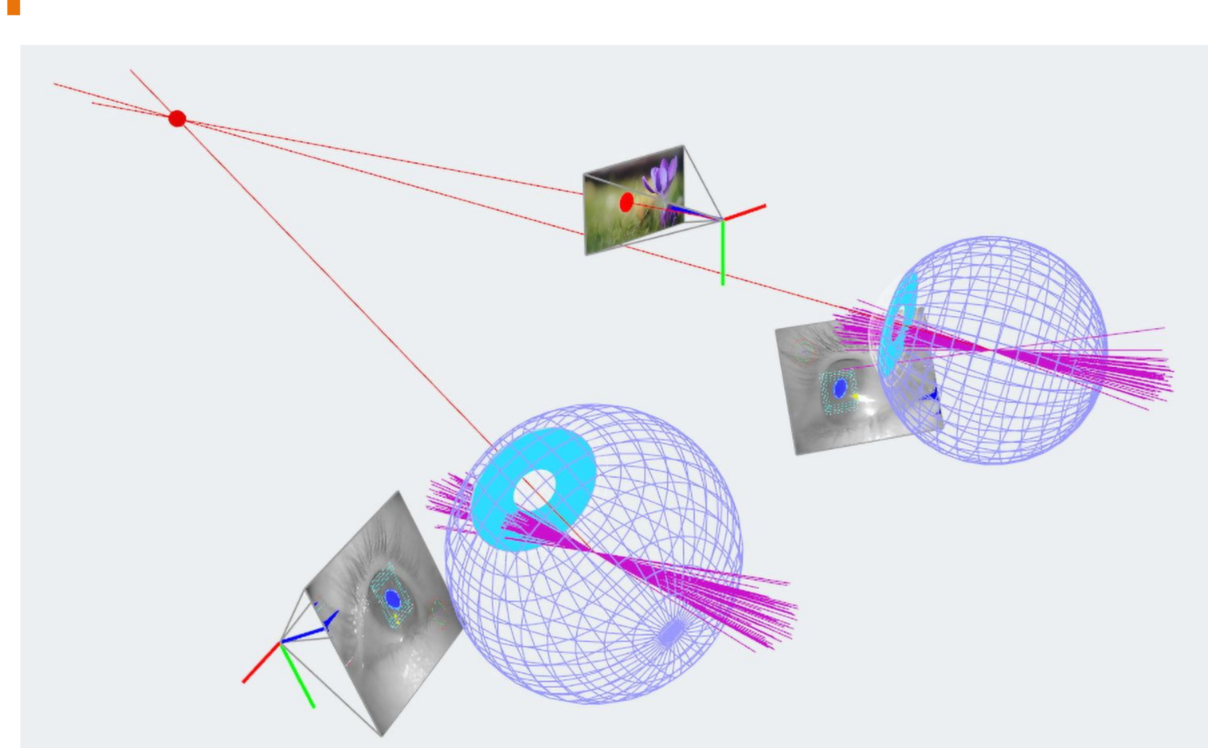


Flow of Eye Footage Through Our Eye Tracking Pipeline

Evaluation

The performance of each pre-processing step will be evaluated post-hoc using measures of *accuracy* and *precision* during periods where subjects were instructed to fixate their gaze on specific points in front of them.

Pupil Labs Gaze Estimation



Pupil Labs' Gaze Estimation Algorithm Using 3D Eye Models

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Deaf or Hard of Hearing Individuals' Collaboration With Manufacturing Robots

Margaret Gray (mag5244@rit.edu), Abhijan Wasti (abhijan@mail.rit.edu), Zhuorui Yong (zy6237@rit.edu)
Advisors: Dr. Jamison Heard, Dr. Shannon Connell

Introduction

Robots are becoming increasingly prevalent in our lives, particularly in our homes and in the workplace. However, **most robots that interact with humans are not intentionally designed to include individuals with disabilities.** Specifically in the manufacturing industry, 18% of all workers are made up of deaf/hard of hearing (D/HOH) individuals. About 17% of all D/HOH people are employed in the manufacturing industry.

RQ: Are there differences between hearing and D/HOH individuals when interacting with collaborative manufacturing robots ?

Experimental Design & Data Collected

Two tasks, each with a simple and enhanced version. These tasks simulate manufacturing assembly tasks, where humans and robots (Fig. 1) must work together to complete a task. The first task involves assembling PVC pipe parts, and the second involves assembling a Lego set (Fig. 2). Data collected includes:

- **Physiological**: eye fixations, blink data (Fig. 3), heart rate & variability, respiration rate (Fig. 4)
- **Subjective**: NASA Task Load Index, post-trial surveys
- **Behavioral**: time to complete tasks

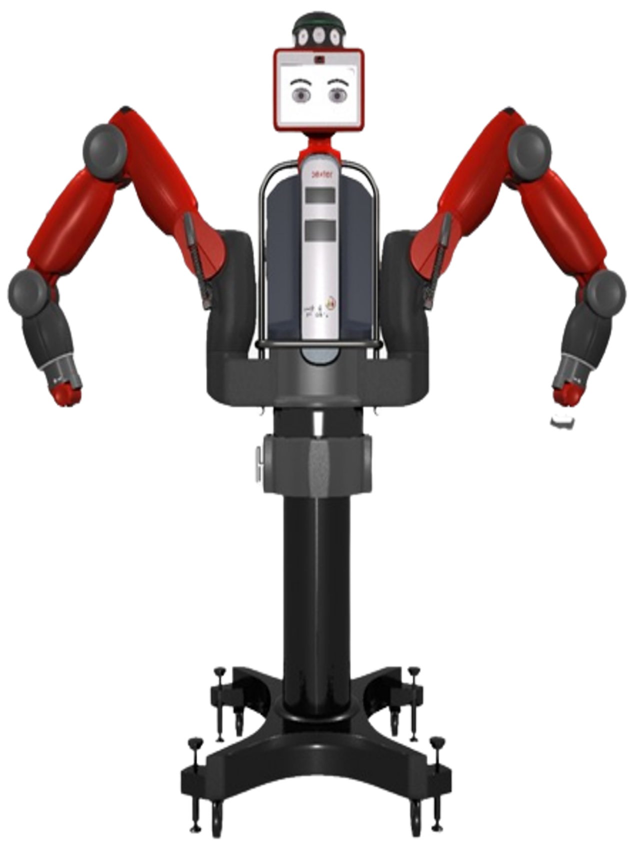


Fig. 1: Baxter Robot

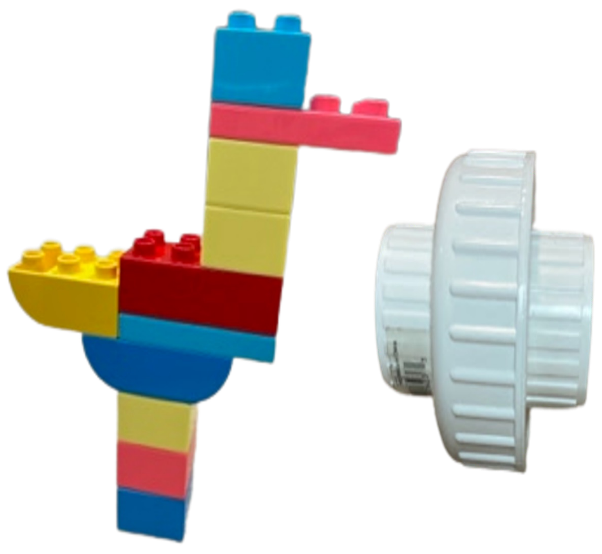


Fig. 2: Builds from Task 1 (right) and Task 2 (left)

Fig. 3: Pupil Core Glasses

Fig. 4: Zephyr Bioharness

Initial Results

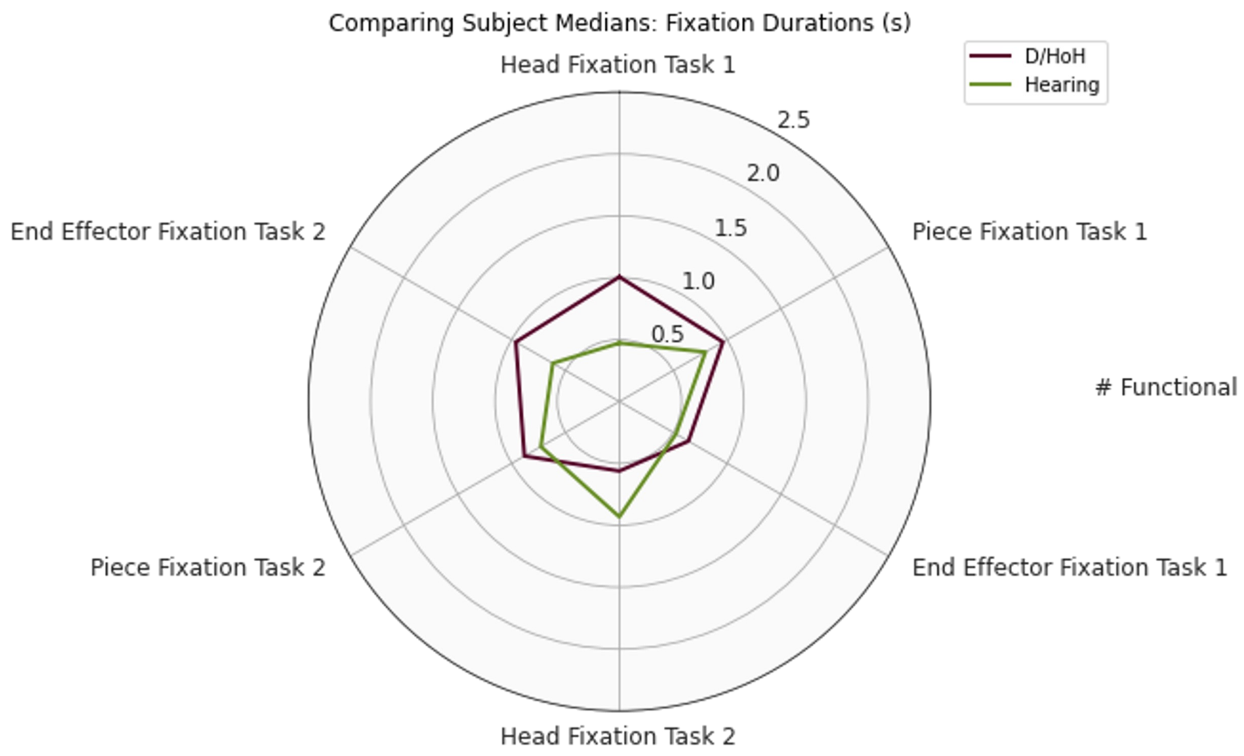


Fig. 5: Fixation Durations Spider Chart

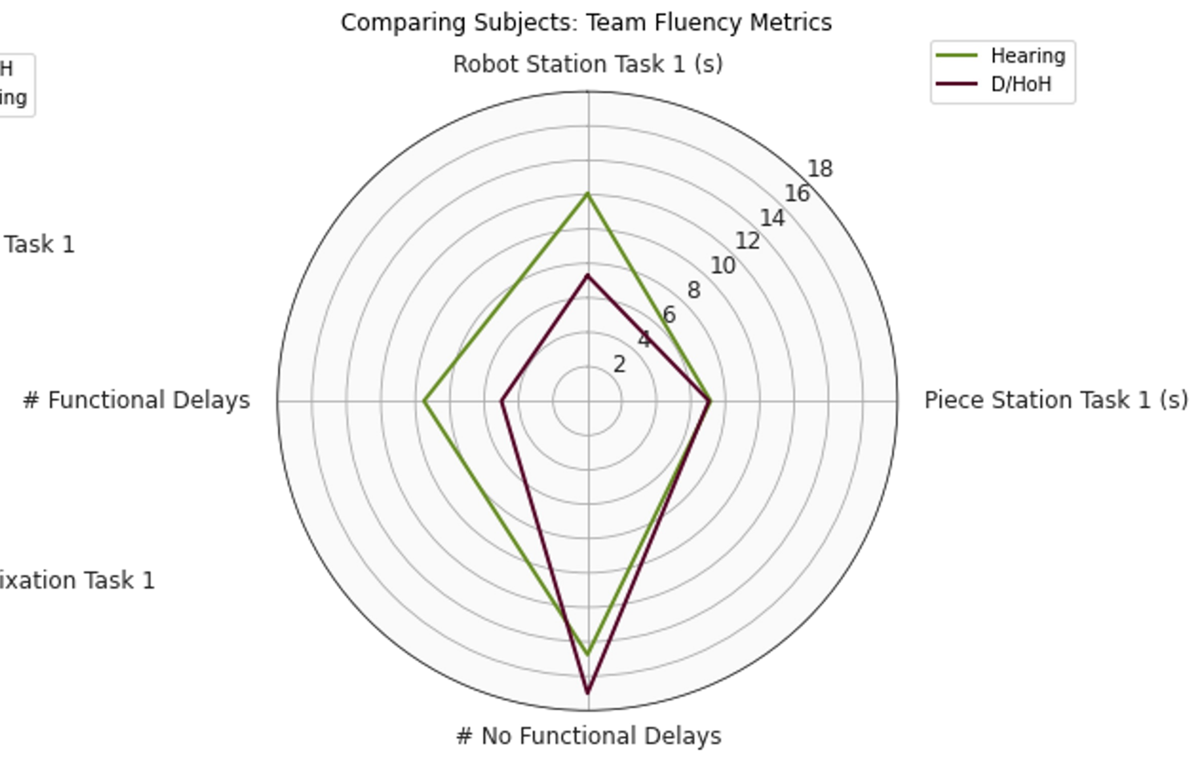


Fig. 6: Team Fluency Spider Chart

Initial results show that **D/HOH individuals generally were more fixated on their surroundings** than hearing individuals, which indicates possible distrust of the robot throughout the course of the trial (Fig. 5). Additionally, **D/HOH individuals showed more fluency with the robot** than hearing individuals, with less functional delays and less time spent at each station (Fig. 6).

Conclusions & Future Work

- **No significant subjective results** were found, but there appears to be possible physiological & behavioral differences between D/HOH and hearing individuals
  - Implies less trust and comfort from D/HOH individuals
- Aiming to recruit a total of 10 D/HOH participants to establish a more representative sample

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Special thank you to PhD student Saurav Singh and 2022 REU students A'di Dust and Carola González Lebrón

RIT | AWARE-AI NSF Research Traineeship Program

Human-Aware AI for Adaptive Human Robot Teaming

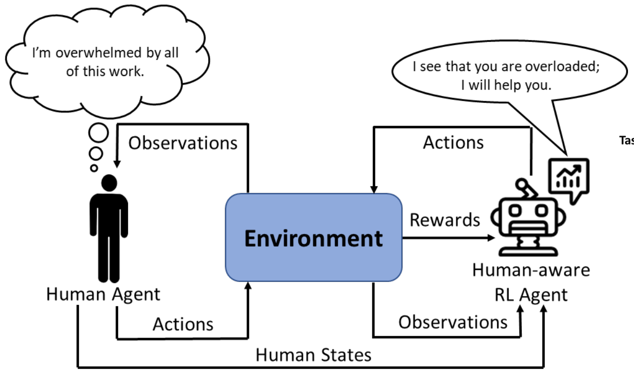
Saurav Singh & Jamison Heard, PhD. | Adaptive Human Robot Teaming Lab | Kate Gleason College of Engineering

Problem

This work presents an AI-based human-aware decision making system that incorporates internal and external human information to elicit more natural robot interactions.

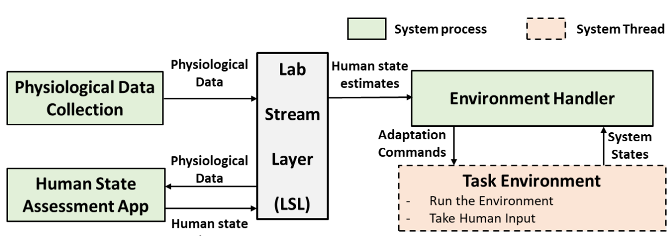
Motivation

- Society expects robots to interact naturally with humans, in a similar manner to how humans interact with other humans.
- However, robots typically only determine interactions based on external information about a human, such as the human's location or what the human is saying.
- This differs from how humans interact, where internal information (e.g., emotions, stress) drives interaction decisions along with external information.

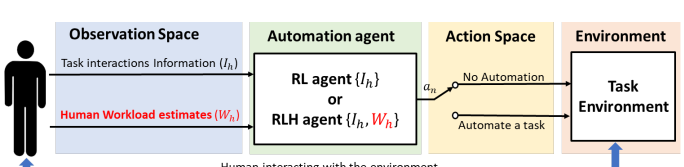


Methodology

- Human internal states are estimated objectively by measuring the human **physiological data** such as heart rate, respiration rate and speech intensity.
- **Human State Assessment App** uses Neural Networks to estimate the human internal states.
- Environment Handler is responsible for determining what tasks to automate based on human and task state information.



- Soft Actor Critic (SAC) reinforcement learning algorithm was used for the human-aware AI agent.
- Two state-space encapsulations are explored.
  - ➢ **RL** encompasses task and interaction information
  - ➢ **RLH** augments RL with human workload information,



Experimental Variables

The developed system was validated in a remotely piloted aircraft simulator (the NASA MATB-II), where the robot provided varying levels of support to the human teammate based on current task information and the human's workload level.

Workload was chosen to represent the human's internal information, due to the relationship between workload and team performance.

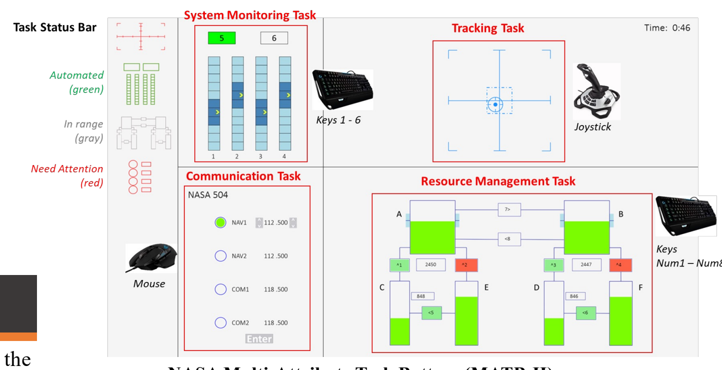
Number of Participants Recruited: 9

Independent variables:

- Workload Condition *(within-subjects variable)*
  Underload (UL) | Normal Load (NL) | Overload (OL)
- Automation strategy type *(between-subjects variable)*
  Rule Based (RB) | Reinforcement Learning agent (RL) | Reinforcement Learning agent with human states (RLH)

Dependent Variables:

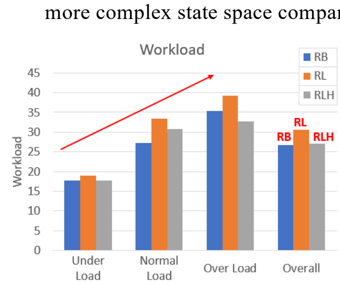
- Estimated Workload
- Rewards
- Individual Task Performance
- Automation Time

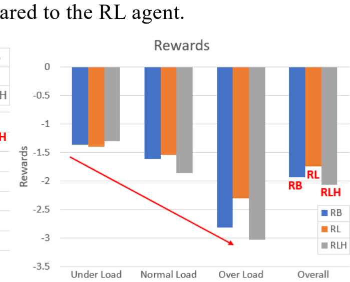


NASA Multi-Attribute Task Battery (MATB-II)

Results

- The addition of **human states** resulted in a **lower overall workload** but with **worse rewards**.
- RLH agent may have **picked up this trend** to reduce human workload, however it **could not achieve rewards** due to a much more complex state space compared to the RL agent.





Conclusion

- The current results demonstrate that an AI that uses external and internal human information can achieve similar or better team performance than traditional methods.
- This result provides the necessary foundation for promoting natural human-robot interactions in high-stress environments, such as search and rescue scenarios.

Acknowledgments: This research is supported by the National Science Foundation (NSF) Award No. 2125362: AWARE-AI: AWAREness for Sensing Humans Responsibly with AI. Any views presented here are those of the authors alone and do not necessarily reflect those of the NSF.

Co-Designing Features for AI-Supported Communication Applications using Automatic Captioning with Deaf and Hearing Pairs

RIT AWARE-AI NSF Research Traineeship Program

CAIR Center for Accessibility and Inclusion Research

RIT

Matthew Seita, Ph.D. Student
Golisano College of Computing and Information Sciences (GCCIS)
Rochester Institute of Technology

Introduction

Due to difficulty conversing using spoken English, many Deaf or Hard of Hearing (DHH) individuals face barriers communicating in meetings with hearing peers, e.g. impromptu small group conversations in the workplace, leading to miscommunication, isolation, & reduced productivity or professional outcomes. We investigate addressing this challenge via investigating AI-supported mobile applications that provide live captions, based on automatic speech recognition (ASR).

Approach

While co-designing in-person was our original approach, this was interrupted by the COVID-19 Pandemic, so we shifted to an online modality and report on methodological findings. We also present some prototype designs our participants created as part of our process.

Research Questions

RQ1: How well can pairs of DHH and hearing individuals participate in an entirely virtual co-design workshop session in which participants collaborate and create sketches of prototype designs of AI-supported technologies?

RQ2: What exploratory design solutions emerge from pairs of DHH and hearing individuals remotely co-designing features for an ASR-supported communication application, along two design dimensions: (1) How errors in ASR output should be indicated and fixed, and (2) How the system should notify users to influence speaking behavior?

Methodology

We conducted 18 co-design sessions each with a DHH-hearing pair. To address RQ1, we performed qualitative analysis of our video recordings of our workshops. Upon workshop completion, o address RQ2, we considered the design options suggested by participants as well as drawings produced during the sessions, which revealed several design options along each of the two dimensions investigated.

Communication Modalities	**Collaborative Drawing Tool**
Conversation Flow & Strategies	**Reconciliation of Ideas**

Figure 2: A snapshot of a Zoom recording capturing the DHH and hearing participants communicating via text chat and working on creating icons for notifications

Our qualitative analysis revealed four major themes related to our co-design process, shown in the figure above. They relate to utilization of interpreter and text-chat, use of the collaborative drawing tool, strategies for communication, and how differences in ideas were reconciled.

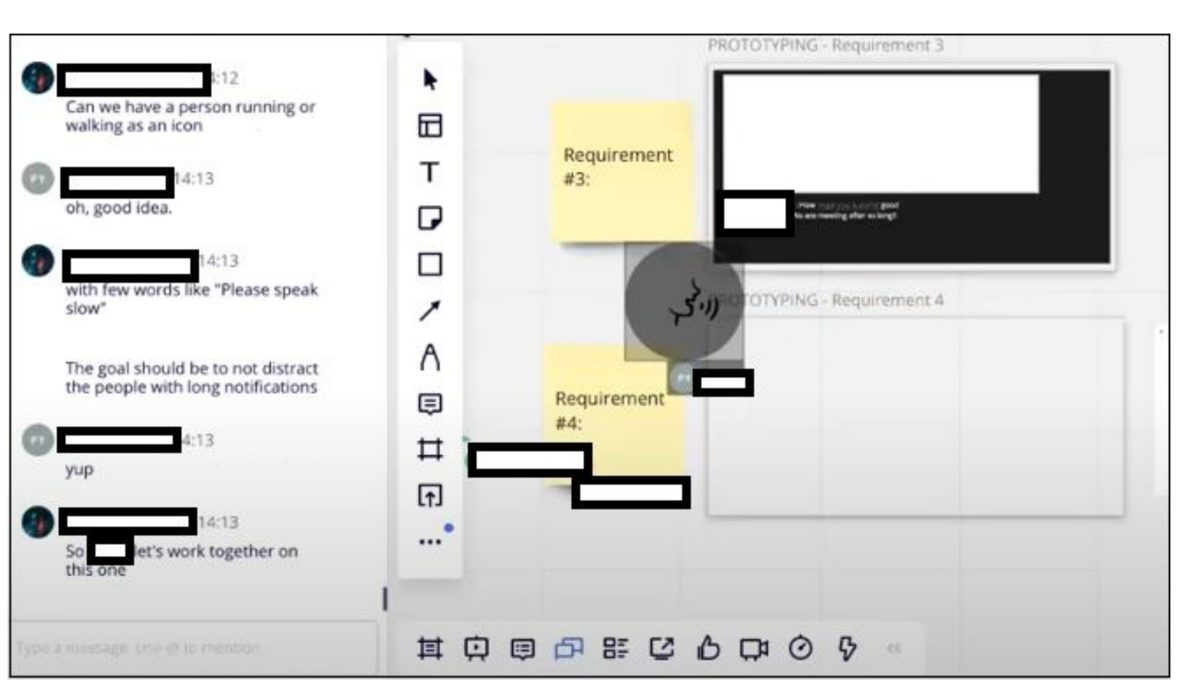


Figure 2: A snapshot of a Zoom recording capturing the DHH and hearing participants communicating via text chat and working on creating icons for notifications

Resulting Prototypes

A major point of disagreement among groups (and between the two participants within some groups) in their prototype designs was how visually prominent features should be. Each has benefits and drawbacks – for example, while notifications with lower visual salience would be less distracting it would also be harder to notice. Some examples of prototype designs are shown below.

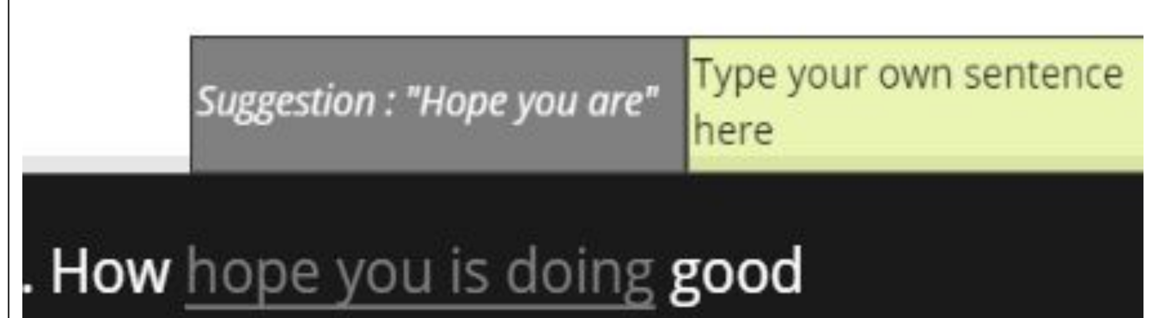


Figure 3: Prototype example for an error correction feature that participants created.

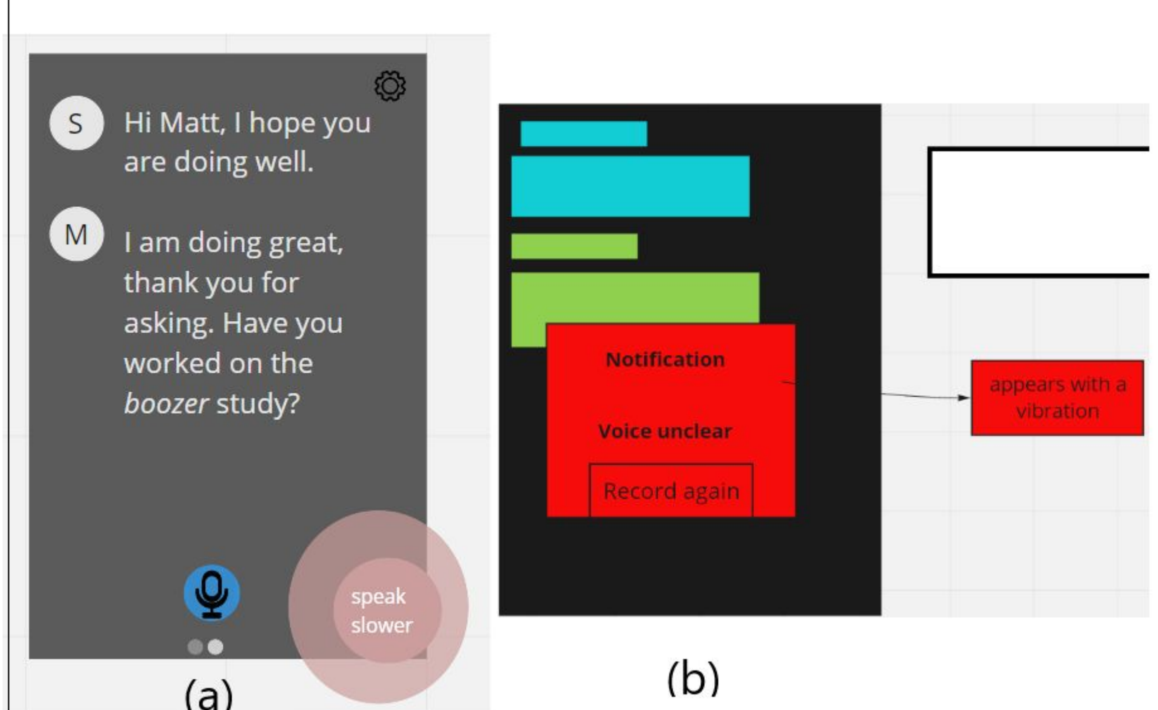


Figure 4: Prototype examples for notification systems to influence speaker behaviors. (a) shows a small notification at the corner, while (B) shows a larger pop-up.

Future Work

Some possible avenues for future work include a longer co-design session with a larger group of participants, and possibly a different topic, to see whether ideas, collaboration, and prototyping changes in this context. Future work is also needed to investigate more deeply whether the preferences for designs created by this study generalize to the preferences of the DHH & hearing communities as a whole, and in practice in real-world situations.

Acknowledgements

This material is based upon work which is generously supported by the National Science Foundation (NSF) under award No. 1954284 and the NSF Graduate Research Fellowship under award No. 1746056, as well as Award No. 2125362: AWARE-AI: AWAREness for Sensing Humans Responsibly with AI. Any views presented here are those of the authors alone and do not necessarily reflect those of the NSF.

Deep Domain Adaptation for Eye Segmentation

RIT

AWARE-AI NSF Research Traineeship Program

Viet Dung Nguyen, *Advisors:* Alexander Ororbia, Reynold Bailey
Golisano College of Computing and Information Sciences (GCCIS)
Rochester Institute of Technology

Motivation

- Current segmentation methodologies have excelled in working with domain-specific datasets
- **Problem:** when inferring on different domains, performance degrades:

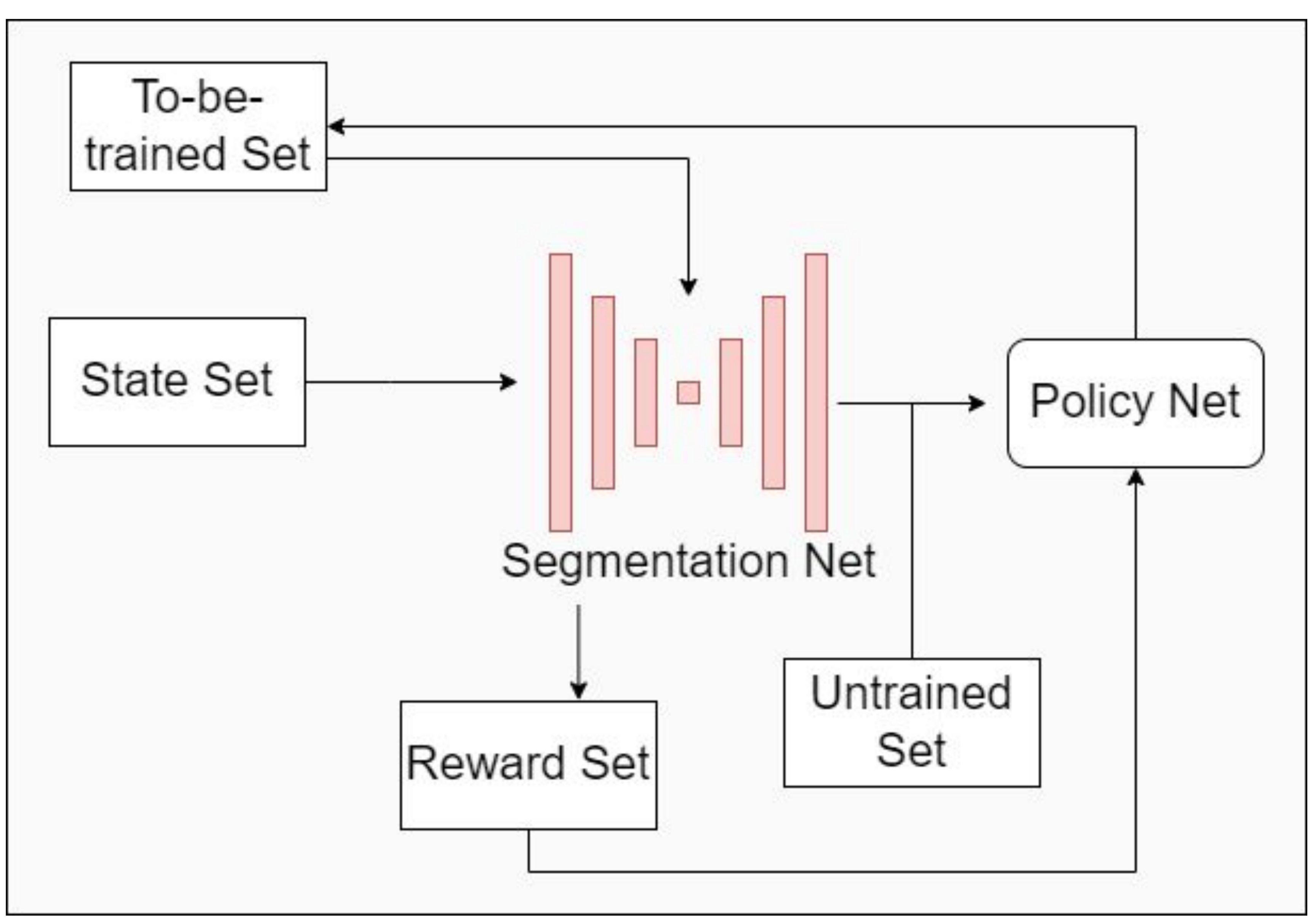
Training a segmentation model on synthetic data yields low mean intersection over union (IoU) on real data

- Issue occurs because images from both datasets come from two different distributions

Method

We develop a reinforcement learning (RL) system that selects whether or not to pick a current eye image to be further used in training the eye segmentation network

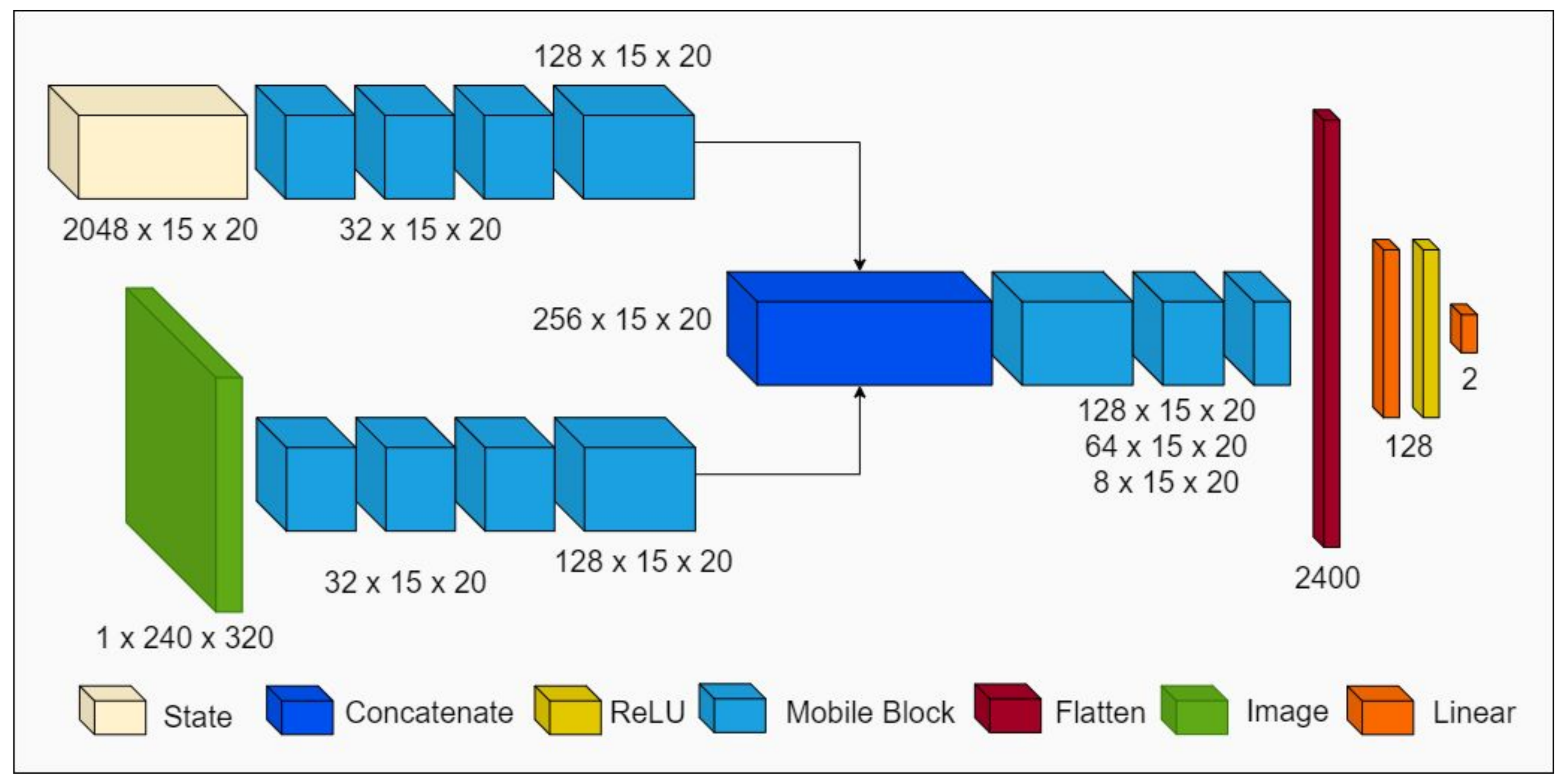
Reinforced Active Learning for Eye Segmentation



Epsilon Greedy

The agent will randomly choose/determine action using a policy network (or via random choice w/ probability epsilon)
- Compute epsilon w/ starting value of 1.0 and ending epsilon at 0.01, decayed over 200 steps

Policy Network Design

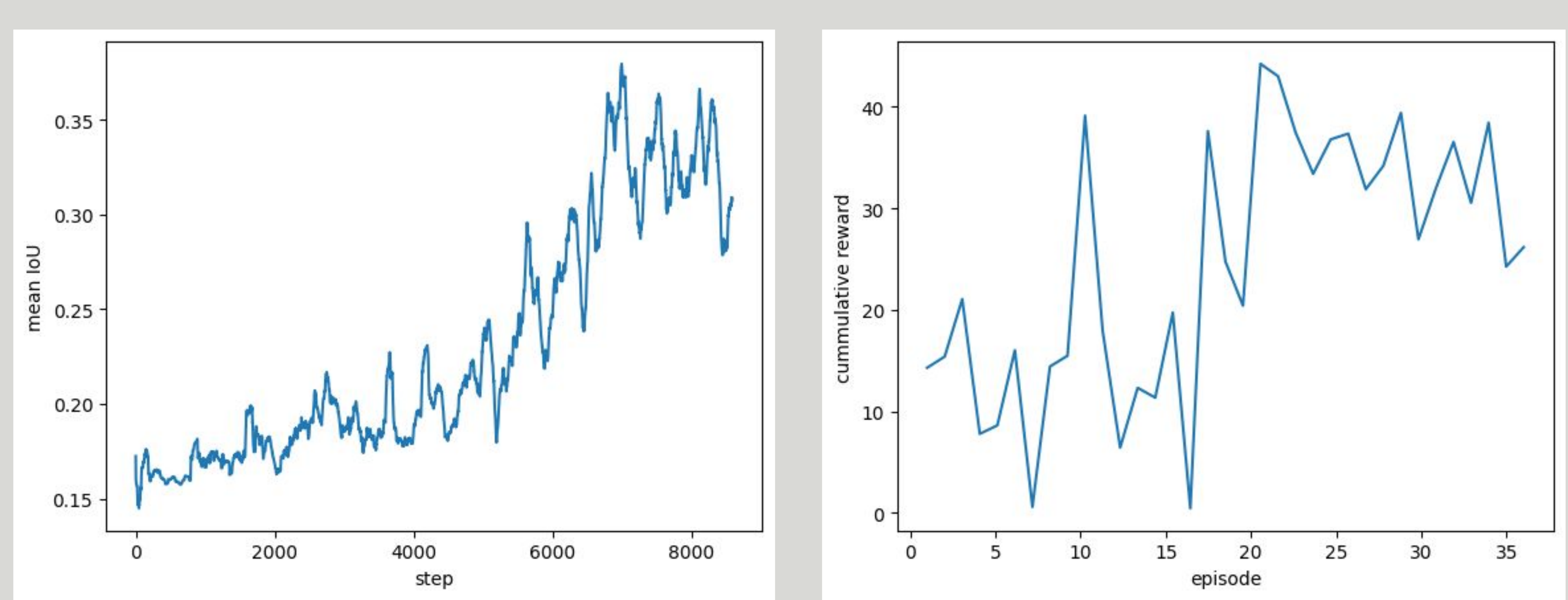


Reward Function Formulation

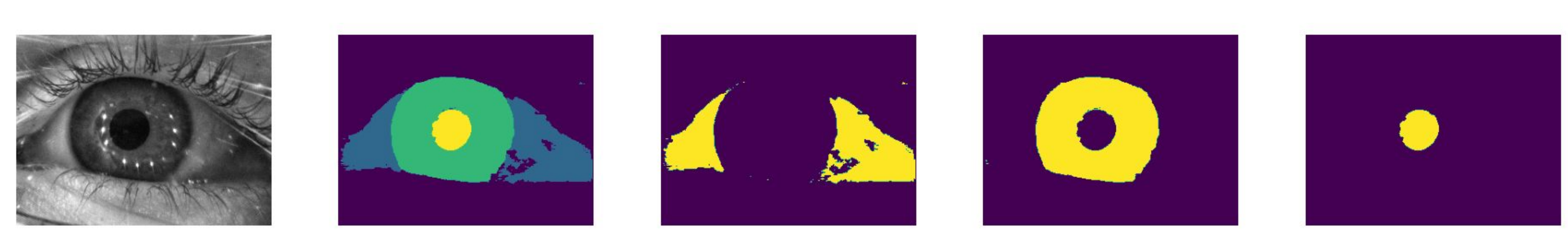
$$r_t = \begin{cases} 10(m_t^{iou} - M^{iou} + 0.05) & \text{if } m_t^{iou} \geq M^{iou} \\ 20(m_t^{iou} - M^{iou}) & \text{otherwise} \end{cases} + \begin{cases} 30(m_t^{iou} - m_{t-1}^{iou}) & \text{if } m_t^{iou} > m_{t-1}^{iou} \\ 0 & \text{otherwise} \end{cases}$$

- We use develop a reward that is a function of the mean IoU (computed using inference under current segmentation network over reward set)
- Mean IoU is appended at each (time) step to the global mean IoU queue (which has a capacity of 5)
- Reward is -0.1 if agent/system does not choose an action

Evaluation and Results



(Left): Mean Intersection over union (IoU) across simulation steps when training with synthetic data
(Right): Cumulative reward across episodes when training the system with the synthetic dataset



Sample predicted segmentation map using segmentation network trained from dataset containing both real and synthetic eye imagery. From left to right: original eye, combined predicted segmentation map, predicted sclera , predicted iris, predicted pupil.

Future Work

- Examine improved reward heuristics for the policy model to be trained in the most effective setting
- Extend framework to continual RL setting, where reward function will be learned using abstract goal states

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Artificial Intelligence in Virtual Reality-based Manufacturing Training

Zhuorui Yong, Ph.D. Student; Department of Industrial & Systems Engineering
Esa Rantanen, Ph.D., CPE; Department of Psychology,
Yunbo "WILL" Zhang, Ph.D.; Department of Industrial & Systems Engineering

Abstract:

Facing the labor shortage in the manufacturing industry, we proposed a VR training system to improve training effectiveness. The VR system combines artificial intelligence algorithms to achieve automatic performance assignment, adaptive training content generation based on trainees' performance, and task assistance.

Background

Manufacturing and Industry5.0

Industry Challenges:
- Labor shortages
- Aging workforces

Industry 5.0:
- Human-first
- Social
- Environmental

Human operators in manufacturing are essential.
While nowaday's training:
- Still rely on the mentor-apprentice system
- On-site teaching

The **high training cost** and **complex working environment** make training less efficient.

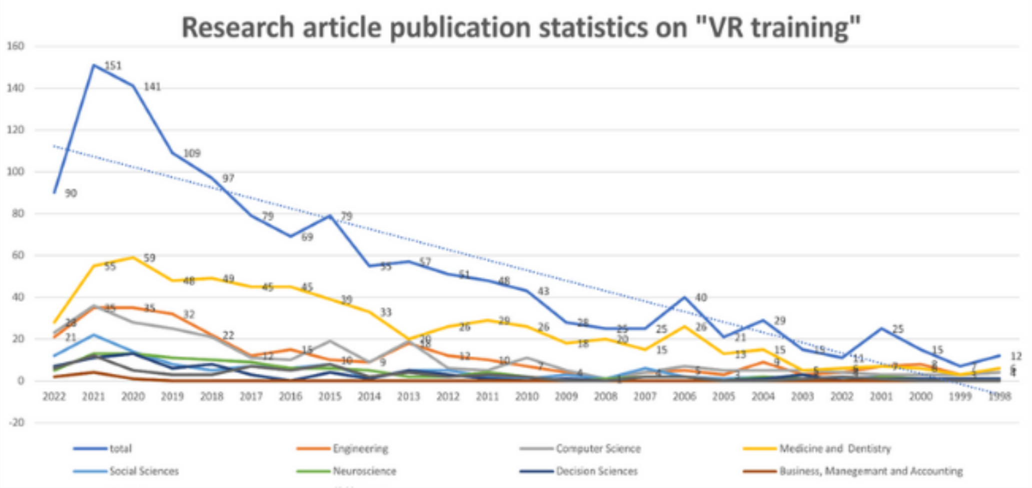
VR Training

Virtual Reality(VR) is an **immersive simulated** **experience**

Stimulus

Training: Produce the correct mapping relationship between **Stimulus** and **Response**

VR training has been a research hotspot. Engineering is an essential component of VR training study in academics.



Benefits of VR Training

- **Reusable**
  - Low cost
  - No waste of physical material
  - Software
- **Virtual Environment**
  - A safe place for novices
  - Protect Trainees
  - Protect manufacturing device
- **Digital**
  - Training log
  - Quantitative data
  - Interactable

Experiment Design

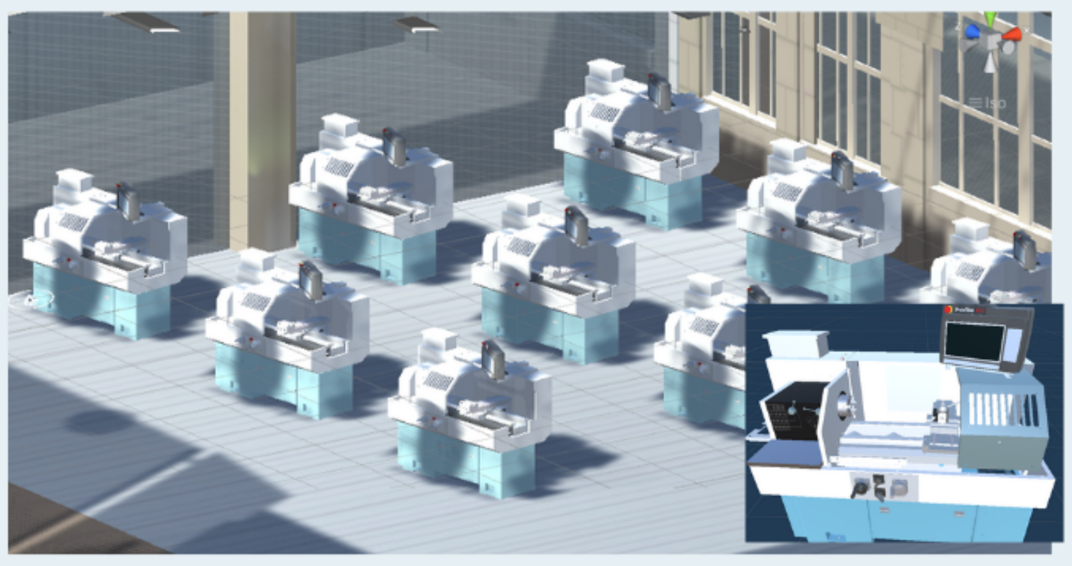
VR Training System Design

Objective: A training environment can provide realistic stimulus and measure the trainees' response.

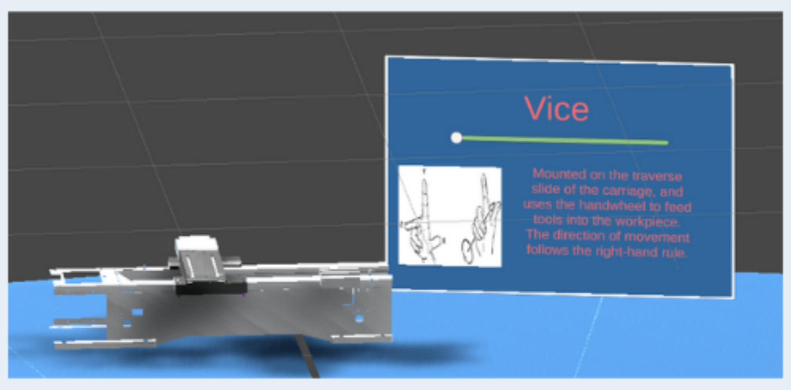
Realistic Criteria:
- Physical Fidelity
- Psychological Fidelity

Solution:
- **Physical Fidelity:**
  - Interactable digital twin models
  - Environmental Simulation

1:1 simulation model of CNC machine in RIT's Lab

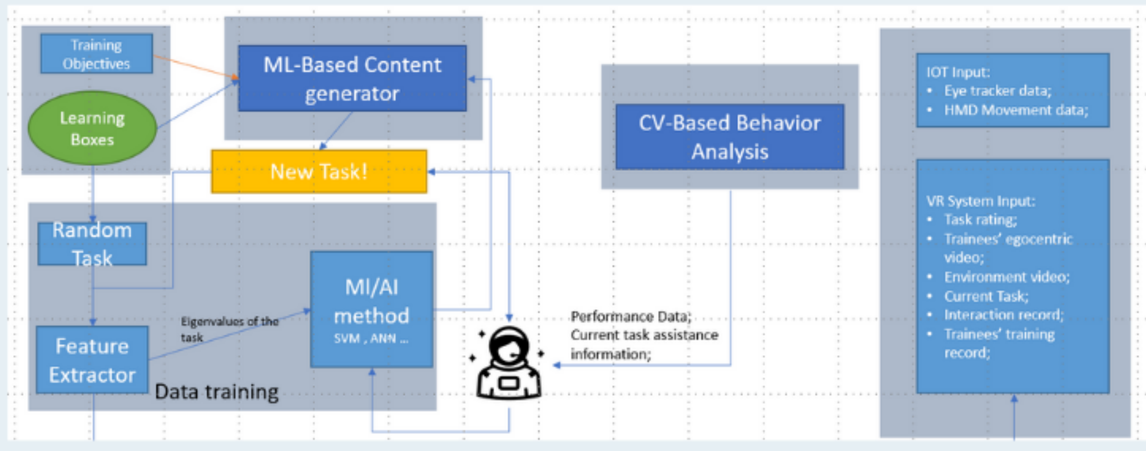


Solution:
- **Psychological Fidelity**
  - Cognitive task analysis of training task
  - Provide visual, auditory, tactile (vibration) multi-sensory stimulation based on cognitive cues



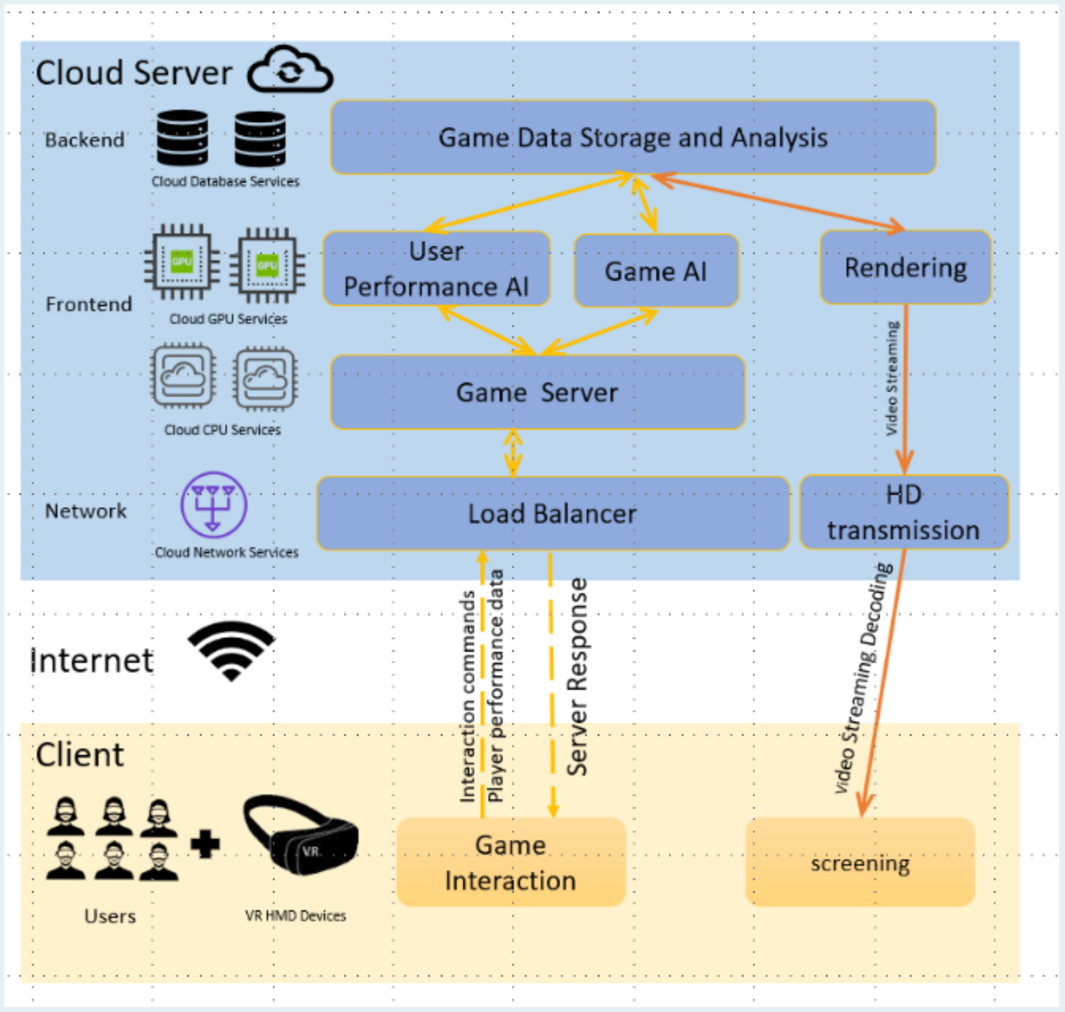
Trainees' Performance Data collection:
- **Performance Model Data Source**
  - Eye-tracker
  - Virtual cameral(Third-person)
  - Virtual egocentric camera (First-person)
  - Interaction record

AI Model:
- **Behavior Analysis and Assistance:**
  - Use Computer Vision Algothism to understand trainees' behavior
  - provide corresponding feedback to the trainees
- **Behavior-Based Adaptations:**
  - Use the Machine Learning model to generate the next training task adaptively based on trainees' skill level and task performance



Application

Cloud-Based Training system:



System Aim :
- Low cost and safe training system
- Equivalent or better training performance
- Independent training
- Personalized training plan generation
- Interesting and engaging training system

System Application:
- Cloud-based deployment
- More Beneficiaries

Future Work:
- Cognitive Modeling of Manufacturing training
- Human-centered AI model for Training evaluation

Reference

1. Maddikunta, P. K. R., Pham, Q. V., Prabadevi, B., Deepa, N., Dev, K., Gadekallu, T. R., ... & Liyanage, M. (2022). Industry 5.0: A survey on enabling technologies and potential applications. Journal of Industrial Information Integration, 26, 100257.
2. Hooshyar, D., Yousefi, M., & Lim, H. (2019). A systematic review of data-driven approaches in player modeling of educational games. Artificial Intelligence Review, 52(3), 1997-2017. Review, vol. 52, no. 3, pp. 1997-2017, 2019, doi: 10.1007/s10462-017-9609-8.
3. Broo, D. G., Kaynak, O., & Sait, S. M. (2022). Rethinking engineering education at the age of industry 5.0. Journal of Industrial Information Integration, 25, 100311.
4. Ragusa, F., Furnari, A., Battiato, S., Signorello, G., & Farinella, G. M. (2020). EGO-CH: dataset and fundamental tasks for visitors behavioral understanding using egocentric vision. Pattern Recognition Letters, 131, 150-157.

Acknowledgments: This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.