# Workload estimation using brain- and bio- signals for adaptive training system

Ivan Tashev, PhD

Microsoft Research Labs – Redmond, USA

# Agenda

- Hardware Experimental Setup
- Adaptive Training System
- Brain- and bio-signal scoring
- Conclusions and future work

# Collaborators

Dr. R. Michael Winters
Microsoft Research

Dr. Yu-Te Wang
Microsoft Research

David Johnston
Microsoft Research

Dr. Ivan J. Tashev
Microsoft Research

Dr. Christine Beauchene
MIT Lincoln Lab

Dr. Nathaniel Bridges
Air Force Research Lab

Justin R. Estep
Air Force Research Lab

# Adaptive Training System

Hardware experimental system

EEG

Eye Tracking

EEG Feature Extraction

Eye-Tracking Feature Extraction

Cognitive Score

Score (0 – 100)

Adaptive Training System

Learning Process Model

Trial Difficulty Level

Trials Library

**Adaptive Trial Difficulty Level**


# Adaptive Training System

Hardware experimental system

EEG

Eye Tracking

EEG Feature Extraction

Eye-Tracking Feature Extraction

Cognitive Score

Score (0 – 100)

Adaptive Training System

Learning Process Model

Trial Difficulty Level

Trials Library

**Adaptive Trial Difficulty Level**


# Hardware experimental system

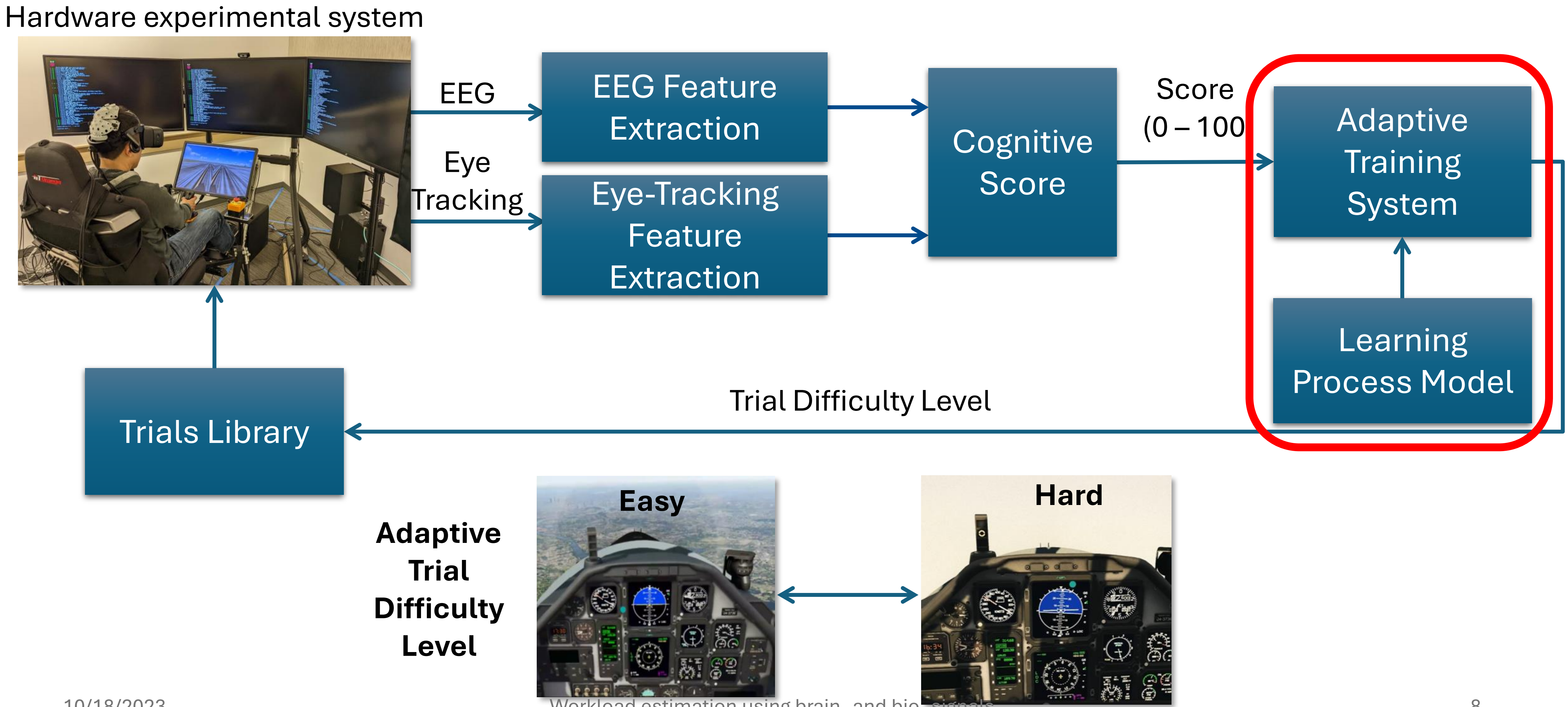
- Flight simulator
  - Prepar 3D software with T6A Texan II plugin
  - 6 DoF motion platform
  - Works in both 3+1+1 monitors setup and in VR
  - 5.1 surround sound
- VR glasses Varjo VR3 retrofitted with EEG electrodes
  - 32 EEG channels plus ground and reference
    - Custom electrodes with Ag/AgCl tips and spring
    - Three 3D printed plates for flexible shape
  - BrainVision LiveAmp pre-amplifier (24 bits, up to 1,000 Hz)
  - Gaze tracking integrated into Varjo VR3
- Auxiliary bio-signal collection
  - Electrocardiographic
  - Galvanic Skin Resistance
  - Breathing

# Trials library

- General scenario is straight line flight with two flavors:
  - Straight-and-level: maintain course, altitude, and speed
  - Glideslope: maintain speed and course, approach the runway
- Difficulty variability
  - Visibility: clouds, fog
  - Disruptions: wind, gusts, thermals
- In the final set are 11 scenarios
  - One straight-and-level with full visibility (easiest)
  - 10 glideslope scenarios
    - Starting with full visibility
    - Ending with practically zero visibility, wind gusts, and thermals (most difficult)
- Data collection with N=16 subjects
  - Experience vary from novice to licensed pilots
  - Aim is to estimate the scenario difficulties (see the training process model later)

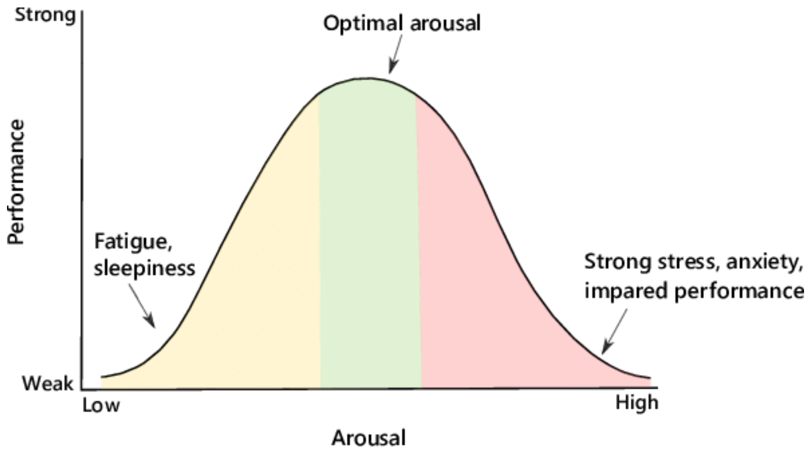
| Name | Difficulty | Limitation |
|---|---|---|
| Straight-and-Level Sc Diff 0 | 1.00 | 95.00 |
| Glideslope Sc Diff 1 | 2.86 | 77.14 |
| Glideslope Th2 Diff 2 | 4.52 | 82.83 |
| Glideslope Th2 Diff 3 | 6.59 | 92.20 |
| Glideslope Th2 Diff 4 | 8.53 | 93.34 |
| Glideslope Th2 Diff 5 | 10.36 | 100.00 |
| Glideslope Th2 Diff 6 | 12.36 | 100.00 |
| Glideslope Th2 Diff 7 | 14.24 | 99.75 |
| Glideslope Th2 Diff 8 | 16.23 | 100.00 |
| Glideslope Th2 Diff 9 | 18.16 | 82.71 |
| Glideslope Th2 Diff 10 | 20.07 | 100.00 |

# Adaptive Training System

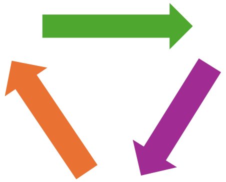


# Theories of Learning Optimization

- Hypotheses
  - Yerkes-Dodson Law valid in pilot training
  - Keeping optimal arousal increases learning speed (cognitive load theory)
- Adaptive Simulation Training
  - Keep the trainee in optimal cognitive and performance state during training




**1. Performance Measurement**
- RMS Deviation
- Kinematics of Controls
- GSR, HRV, RSA
- Gaze & Pupillometry
- EEG / MEG
- Learning-Styles, Self-Report



**2. Adaptive Logic**
- Rule-based Heuristics
- Fuzzy Logic
- Decision Trees
- KNN, SVM (Supervised Learning)
- Reinforcement Learning
- State-Control Regulators

**3. Adaptive Variable**
- Wind Speed
- Wind Direction
- Visibility
- Control of Aircraft
- Controller Sensitivity
- Task Difficulty

RM Yerkes, JD Dodson, "The relation of strength of stimulus to rapidity of habit-formation". Journal of Comparative Neurology and Psychology, 1908/18 (5): 459–482.
M. Zahabi, AM Abudl Razak, “Adaptive virtual reality-based training: a systematic literature review and framework.” Virtual Reality, 2020/24 (4), 725–752.

# Model of the scenarios and trainees

- Each scenario $l$ is characterized by: a) scenario difficulty $d_l$; b) maximum achievable score $M_l$.
- Each trainee $k$ is characterized by: a) initial absolute skill $S^0{}_k$; b) learning rate $\mu_k$.
- At each training step $n$ with scenario with difficulty $d_l$, every trainee $k$ is modelled as:

$$S_k^{(n-1)} = \exp\left(-\frac{\left(t_k^{(n)} - t_k^{(n-1)}\right)}{\tau}\right) S_k^{(n-1)}$$

Account for the absolute skill deterioration

$$\hat{Q}_k^{(n)} = M_l \left(1 - \exp\left(-\frac{\left(S_k^{(n-1)}\right)^2}{2\left(d_l^{(n)}\right)^2}\right)\right) + \mathrm{N}\left(0, \left(c_1 \exp\left(\left(-\frac{d_l^{(n)}}{c_2}\right)^{c_4} + c_3\right)\right)^2\right)$$

Compute the score, second part is random performance variation, interpolated with the difficulty of the scenario

$$S_k^{(n)} = S_k^{(n-1)} + \mu_k \frac{S_k^{(n-1)}}{\left(d_l^{(n)}\right)^2} \exp\left(-\frac{\left(S_k^{(n-1)}\right)^2}{2\left(d_l^{(n)}\right)^2}\right)$$

Compute the increase in the absolute skill

Ivan J. Tashev, R. Michael Winters, Yu-Te Wang, David Johnston, Nathaniel Bridges. “Adaptive Training System", IEEE RAPiD 2023, September 2023

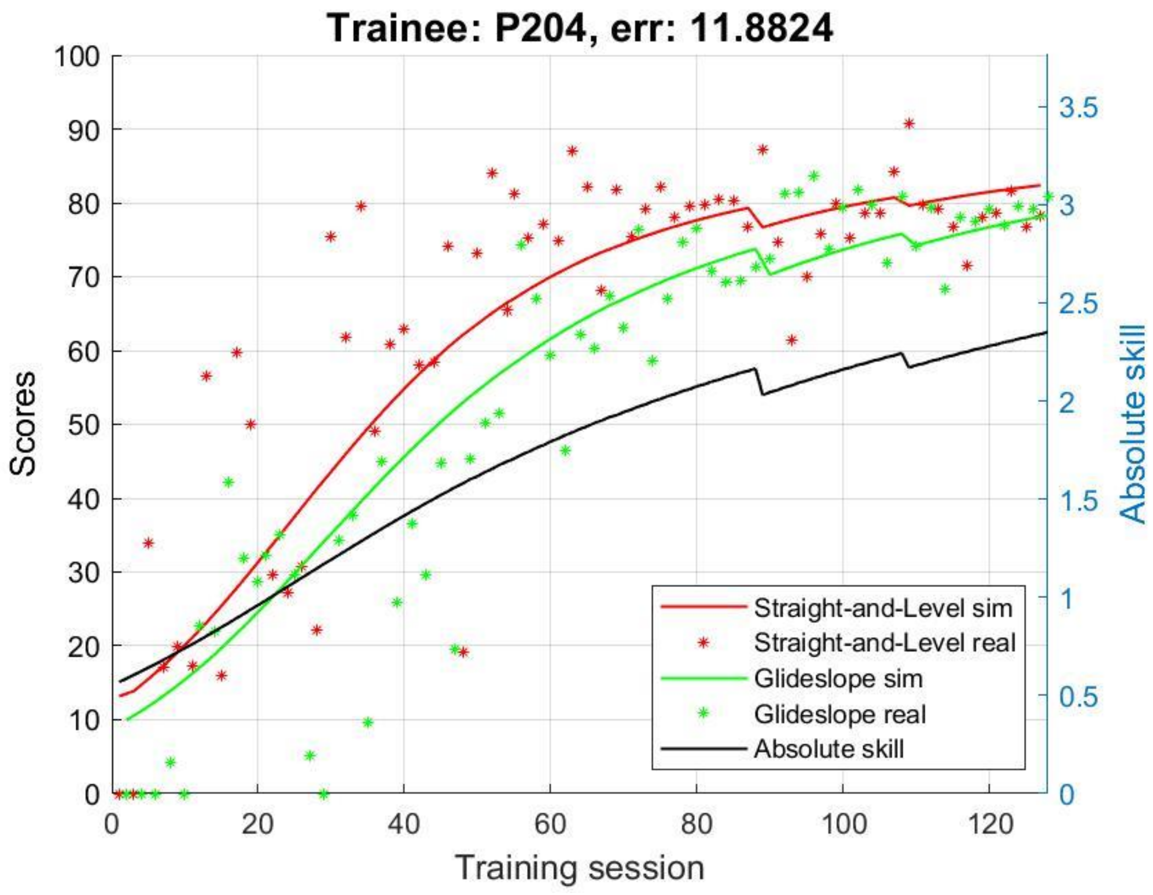
# Model of the scenarios

- Each scenario $l$ is characterized by: a) scena
- Each trainee $k$ is characterized by: a) initial a
- At each training step $n$ with scenario with dif

$$S_k^{(n-1)} = \exp\left(-\frac{\left(t_k^{(n)} - t_k^{(n-1)}\right)}{\tau}\right) S_k^{(n-1)}$$

$$\hat{Q}_k^{(n)} = M_l \left(1 - \exp\left(-\frac{\left(S_k^{(n-1)}\right)^2}{2\left(d_l^{(n)}\right)^2}\right)\right) + \mathrm{N}\left(0, \left(c_1 \exp\left(\left(-\frac{d_l^{(n)}}{c_2}\right)^{c_4} + c_3\right.\right.\right.$$

$$S_k^{(n)} = S_k^{(n-1)} + \mu_k \frac{S_k^{(n-1)}}{\left(d_l^{(n)}\right)^2} \exp\left(-\frac{\left(S_k^{(n-1)}\right)^2}{2\left(d_l^{(n)}\right)^2}\right)$$




ılty

Compute the increase in the absolute skill

Ivan J. Tashev, R. Michael Winters, Yu-Te Wang, David Johnston, Nathaniel Bridges. “Adaptive Training System", IEEE RAPiD 2023, September 2023

# Adaptive Training System – the math

- Absolute skill increase at $n$-th run of scenario with difficulty $d_l$:

$$S_k^{(n+1)} = S_k^{(n)} + \mu_k \frac{S_k^{(n)}}{\left(d_l^{(n)}\right)^2} \exp\left( -\frac{\left(S_k^{(n)}\right)^2}{2\left(d_l^{(n)}\right)^2} \right)$$

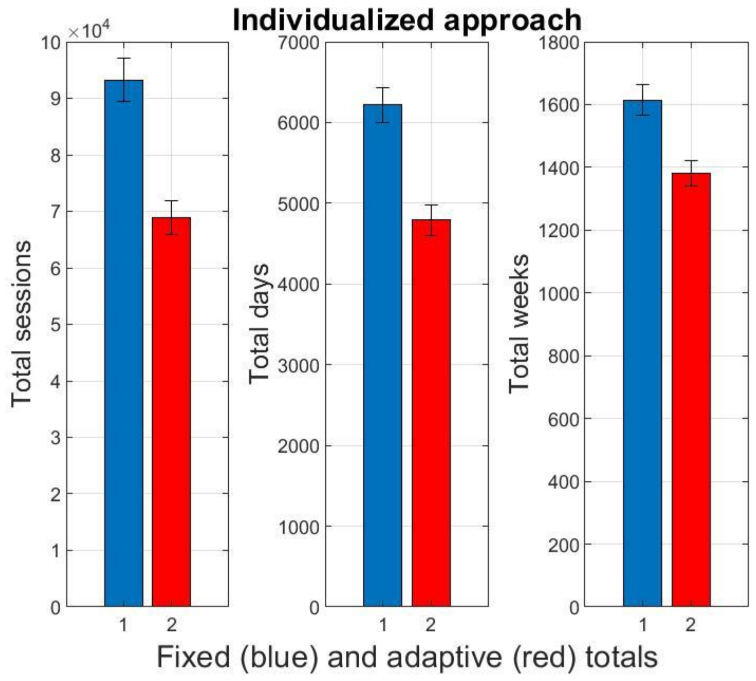
- The absolute skill increase depends on:

$$\frac{S_k^{(n)}}{\left(d_l^{(n)}\right)^2} \exp\left( -\frac{\left(S_k^{(n)}\right)^2}{2\left(d_l^{(n)}\right)^2} \right)$$

- After taking the first derivative, assigning to zero, and solving for $d_l$ the highest increase of the absolute skill is at: $d_l^{(n)} = S_k^{(n)}$
- Given the training history (scores, dates, difficulties) we can estimate the initial absolute skill and the learning rate – lead to the current skill, hence the recommended difficulty

Ivan J. Tashev, R. Michael Winters, Yu-Te Wang, David Johnston, Nathaniel Bridges. “Adaptive Training System", IEEE RAPiD 2023, September 2023

# Adaptive Training System – the




- Absolute skill increase at $n$-th run of scenario with c

$$S_k^{(n+1)} = S_k^{(n)} + \mu_k \frac{S_k^{(n)}}{\left(d_l^{(n)}\right)^2} \exp\left(-\frac{\left(S_k^{(n)}\right)^2}{2\left(d_l^{(n)}\right)^2}\right)$$

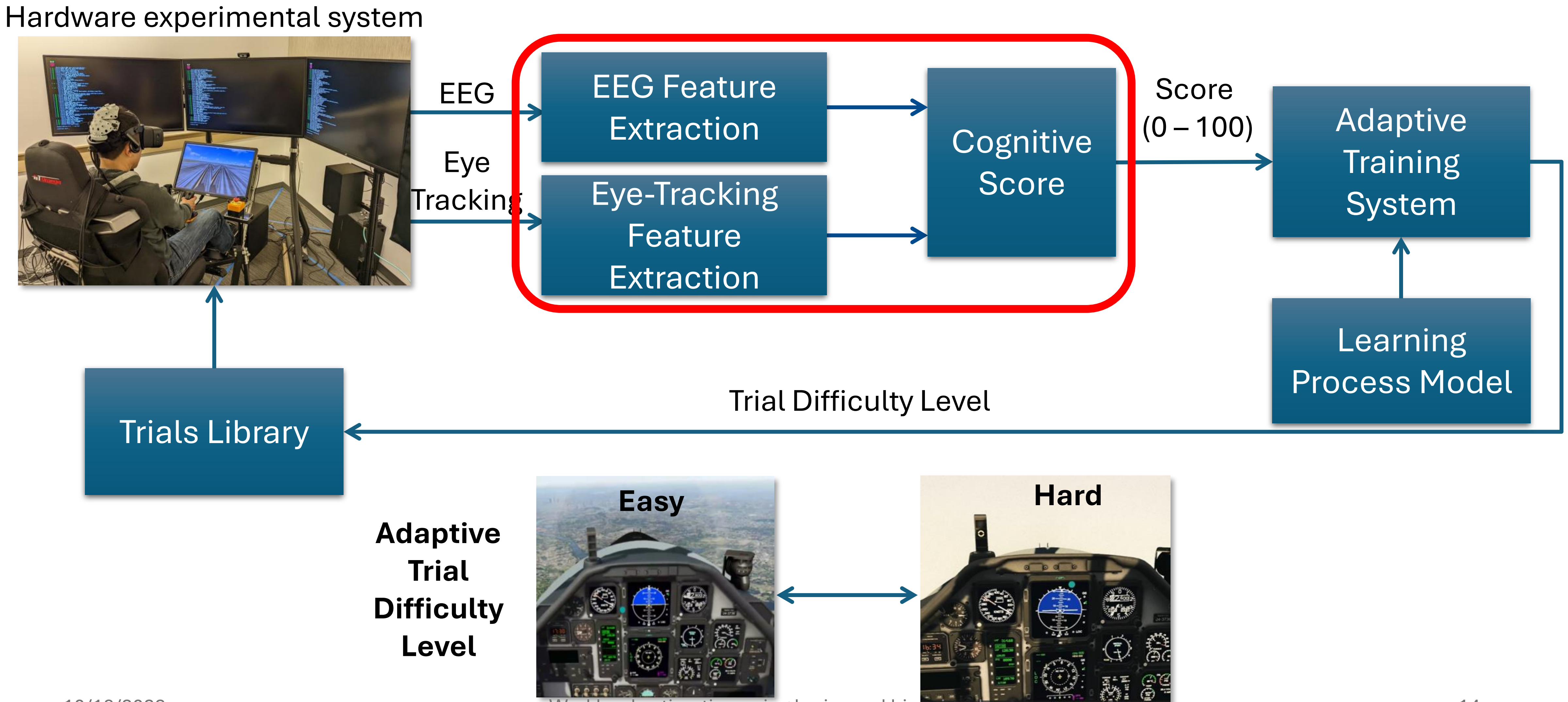
- The absolute skill increase depends on:

$$\frac{S_k^{(n)}}{\left(d_l^{(n)}\right)^2} \exp\left(-\frac{\left(S_k^{(n)}\right)^2}{2\left(d_l^{(n)}\right)^2}\right)$$

- After taking the first derivative, assigning to zero, and solving for $d_l$ the highest increase of the absolute skill is at: $d_l^{(n)} = S_k^{(n)}$
- Given the training history (scores, dates, difficulties) we can estimate the initial absolute skill and the learning rate – lead to the current skill, hence the recommended difficulty

Ivan J. Tashev, R. Michael Winters, Yu-Te Wang, David Johnston, Nathaniel Bridges. “Adaptive Training System", IEEE RAPiD 2023, September 2023

# Adaptive Training System



# EEG and Eye-tracking Preprocessing and Feature Extraction Pipeline

- 32 channel dry EEG
  - Only F5, C3, P5, F6, C4, P6, Pz channels used (frontal, central, parietal; left and right)
  - Process in 5 seconds frames
  - Data cleaning, artifact removal, frames rejection
  - Feature calculation – spectral power in delta, theta, alpha, beta, and gamma bands*
  - Variable length vector for the session

- Eye tracking
  - Convert raw gaze data to PyTrack format
  - Oculomotor feature extraction
  - Statistical measures calculation
  - Variable length vector for the session

- Dealing with the variable length vectors for each session
  - Statistics across the timeline: mean, deviation, min, max
  - One fixed set of features for each session
    - EEG: 5 bands x 4 stats x 6 channel = 120
    - EYE: 39 eye features x 4 stats = 156

* Delta: 1-4 Hz, Theta: 4-8 Hz, Alpha: 8-12 Hz, Beta: 12-30 Hz, Gamma: 30-60 Hz

# Workload vs. scoring

| Workload (WL, 0-1) | Situation | Expected score 100*(1-WL) | Skill (S) and Difficulty (d) | Simulated Score mean and deviation |
|---|---|---|---|---|
| High | Skill below difficulty | 25 | $S/d < 0.9$ | $11.57 \pm 9.31$ |
| Medium | Skill adequate to difficulty | 50 | $0.9 < S/d < 1.1$ | $37.56 \pm 3.48$ |
| Low | Skill above difficulty | 75 | $S/d > 1.1$ | $70.51 \pm 18.75$ |
| None | Calibration | 100 | N/A | 100 (assigned) |

Conclusions: 1. Simulated score is highly correlated to the workload.
2. Correlation with simulated score is a good evaluation parameter.

# Using the simulation scores as labels

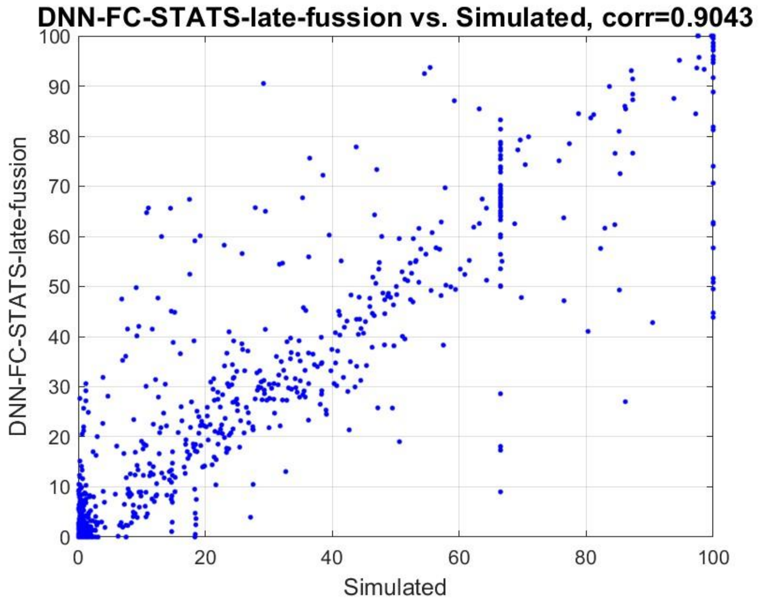
- Treat the problem as a machine learning problem:
  - Set of feature vectors with variable length as input
  - The simulated score as label – one per session
- Evaluation criterion: correlation with the simulated scores
- Try to create one scoring model for all persons and sessions
  - Take one subject for test, one for validation, train on the rest
  - Exercise all possible combinations for subjects with scores > 90, average the results
- Features: all bands (alpha, delta, theta, beta) and timeline statistics (mean, deviation, min, max)
- Fusion of the two feature sets:
  - Early – all features in one classifier
  - Late – two classifiers (for EEG and EYE), third for final score estimation
- Dataset: 20 subjects, 1223 points
- Classifiers: linear, SVM, ELM, DNN, LSTM

Ivan Tashev, R. Michael Winters, Yu-Te Wang, David Johnston, Justin Estepp, Nathaniel Bridges. “Towards Better Scoring", IEEE RAPiD 2023, September 2023

# Results

| Features | | Validation | | | | | Test | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | LIN | SVM | ELM | DNN | LSTM | LIN | SVM | ELM | DNN | LSTM |
| mean of the feature set | EEG | 0.1471 | 0.1204 | 0.1469 | 0.1375 | | **0.1471** | 0.1204 | 0.1381 | 0.1332 | |
| | EYE | 0.4592 | 0.1615 | 0.2509 | 0.4816 | | 0.4592 | 0.1615 | 0.2437 | **0.4693** | |
| | Fusion | 0.7207 | 0.724 | 0.7359 | 0.7560 | | 0.7207 | 0.724 | 0.7331 | **0.7403** | |
| | Early fusion | 0.4391 | 0.1596 | 0.1488 | 0.4814 | | 0.4391 | 0.1596 | 0.1414 | **0.4511** | |
| mean, max, min, std of the feature set | EEG | 0.1143 | 0.1250 | 0.1182 | 0.1290 | | 0.1143 | **0.125** | 0.1058 | 0.1186 | |
| | EYE | 0.4233 | 0.3262 | 0.2909 | 0.5449 | | 0.4233 | 0.3262 | 0.279 | **0.5417** | |
| | Fusion | 0.8093 | 0.8073 | 0.8394 | **0.8402** | | 0.8093 | 0.8073 | **0.8397** | 0.8376 | |
| | Early fusion | 0.2816 | 0.3417 | 0.3399 | 0.4688 | | 0.3786 | 0.3253 | 0.2743 | **0.5087** | |
| Sequence | EEG | | | | | 0.3173 | | | | | 0.2986 |
| | EYE | | | | | 0.5613 | | | | | 0.5499 |
| | Fusion | 0.839 | 0.8464 | 0.8384 | 0.8489 | | 0.839 | **0.8464** | 0.8384 | 0.8371 | |

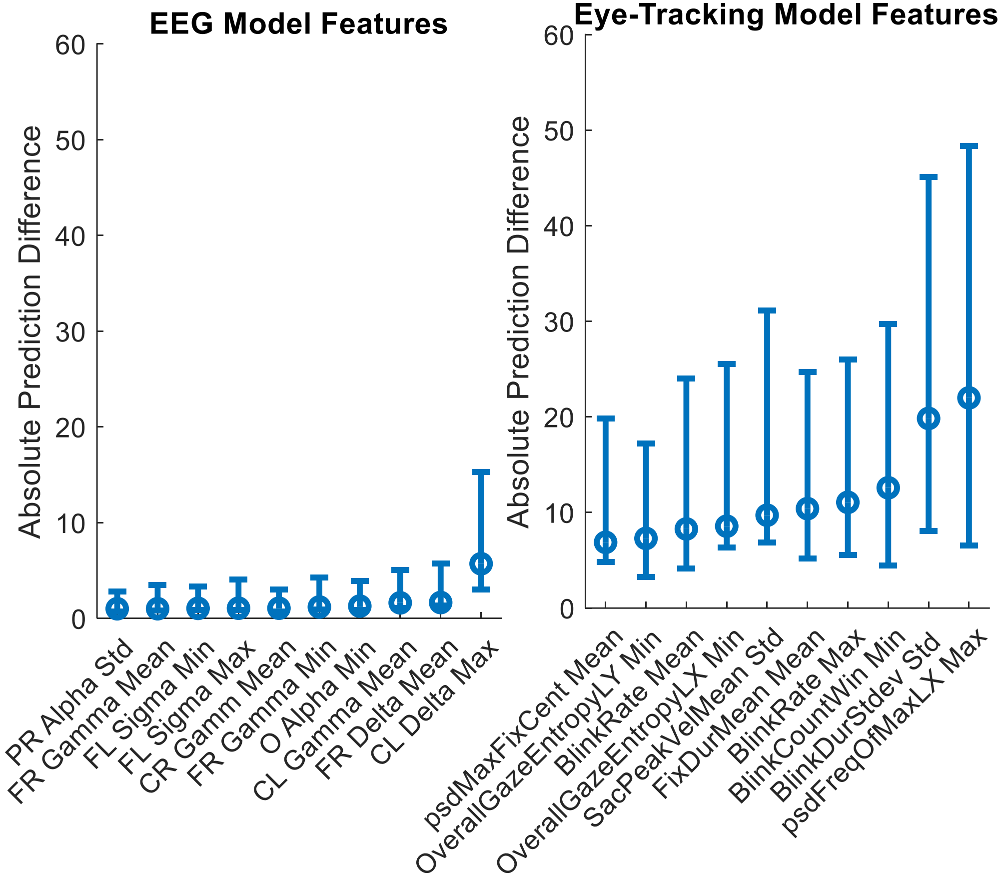
- Recommended:
  - Late fusion structure
  - DNN for EEG features + stats
  - DNN for EYE features + stats
  - DNN for late fusion




**Note:** correlation in the figure is for all data (test + validation + train)

# Ablation Study

FR: Frontal Right
FL: Frontal Left
CR: Central Right
CL: Central Left
PR: Parietal Right
O: Occipital




- Systematically "removed" each feature to identify the primary drivers of the model.
- For EEG, the primary feature is in the Delta band, which may be associated with attention [1].
- For Eye-Tracking, the primary features are related to the frequency of eye-movements and blinking, which could be associated with attention and fatigue [2]

**For predicting performance, the eye-tracking measures have a larger effect on the prediction compared to the EEG.**


[1] Liang (2022)
[2] Huette S (2016)

# EEG ablation study per feature groups

- Bands group:
  - Most useful: Alpha and Theta
  - Mostly noise: Beta, Delta
- Electrodes group:
  - Most useful: CL, FR, FL
  - Mostly noise: PR, CR, O
- Processing group:
  - Most useful: Std, Max
  - Mostly noise: Min, Mean

| **Algorithm** | **Baseline** | **Bands** | | | | | **Electrodes groups** | | | | | | | **Feature groups** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Delta | Theta | Alpha | Beta | Gamma | FL_ | CL_ | PL_ | FR_ | CR_ | PR_ | O_ | Mean | Std | Min | Max |
| LIN | 0.0544 | 0.0285 | 0.0321 | 0.0216 | 0.0607 | 0.0644 | 0.0628 | 0.0288 | 0.0494 | 0.0448 | 0.0999 | 0.1142 | 0.0733 | 0.0747 | 0.0559 | 0.0737 | 0.0304 |
| SVM | 0.1078 | 0.1480 | 0.0754 | 0.0995 | 0.1204 | 0.1155 | 0.1023 | 0.1111 | 0.1174 | 0.1024 | 0.1083 | 0.1214 | 0.1245 | 0.1274 | 0.1149 | 0.1128 | 0.0896 |
| ELM, 64 | 0.0724 | 0.0603 | 0.0610 | 0.0659 | 0.0983 | 0.0558 | 0.0787 | 0.0348 | 0.0963 | 0.0749 | 0.1234 | 0.1391 | 0.0978 | 0.1000 | 0.0896 | 0.0576 | 0.0212 |
| DNN-FC | 0.0920 | 0.1094 | 0.0848 | 0.0708 | 0.1112 | 0.1087 | 0.0567 | 0.0726 | 0.0720 | 0.0556 | 0.0832 | 0.1264 | 0.0414 | 0.0929 | 0.0953 | 0.0709 | 0.0583 |

# Conclusions and future work

- Conclusions:
  - Created a model for predicting simulated scoring based on brain- and gaze-signals
  - The model is person independent and session independent
  - The correlation with the simulated scores is above 0.84

- Future work:
  - Continue to refine the model based on feature set analysis to identify key features across participants
  - Exploring more sophisticated classifiers: CNN, LSTM, for example (while being careful of limited number of labels)
  - Adding other bio-signals collected: ECG, GSR, respiration
  - Evaluation how the model will work outside of the specific scenario

# Finally…

# Questions?

## Thank you for your attention!

BCI Project in MSR: https://www.microsoft.com/en-us/research/project/brain-computer-interfaces/

E-mail for contact: ivantash@microsoft.com