# Towards Driver Cognitive Load Detection Based on Visual Attention Information

Liu, Cheng Chen

MASc Oral Exam, August 31st, 2017



The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

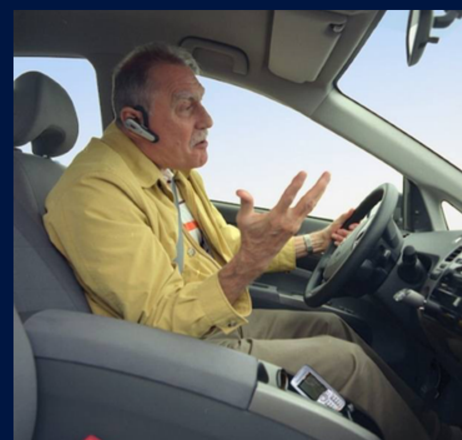
FACULTY OF APPLIED SCIENCE & ENGINEERING

# Motivation: Driver Cognitive Load

Increasing amount of technologies incorporated into vehicles

1. Engagement into secondary tasks
2. Needs for human intervention under complex situations remains

Necessity for monitoring driver **cognitive** states



This research is developing a practical system capable of estimating **driver cognitive load,** with a focus on <u>visual information</u>.

# Outline

- Contribution 1: Data Collection
  - Experiment design for modeling three cognitive load levels
  - Implementation and resulted dataset
- Contribution 2: Estimation Method Development
  - Meta-features for capturing visual attention variations
  - Training classification models with five algorithms
  - Develop towards a comprehensive evaluation framework
- Conclusion and Future Work

# Data Collection



The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

FACULTY
OF APPLIED
SCIENCE &
ENGINEERING

# Contribution 1: Data Collection

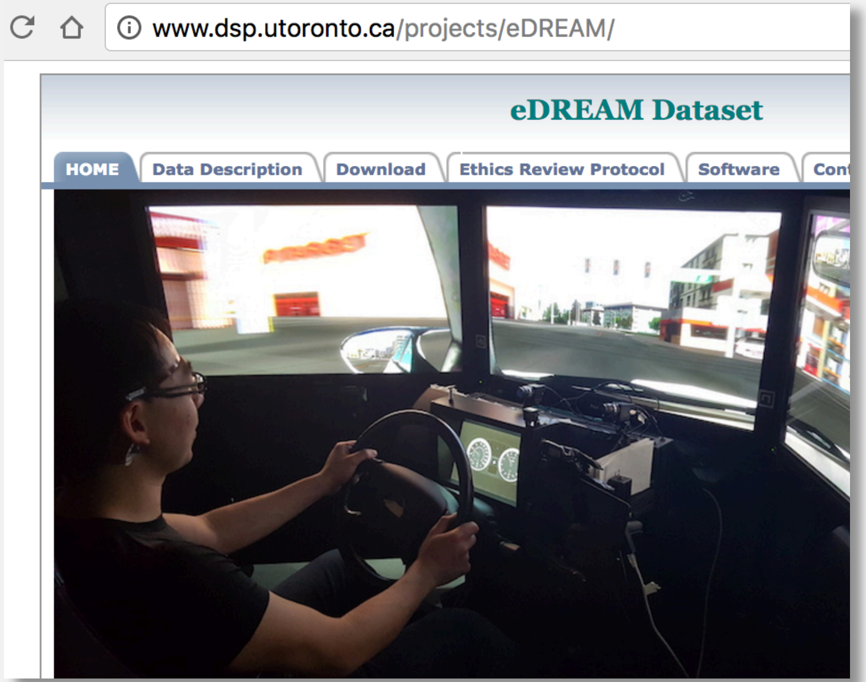
- Objective: gather drivers' responses under differing cognitive load.
  - ➢ Necessary to support studies of driver cognitive load
    - – Dataset featuring this specific problem was not publicly available
  - ➢ The collected data features visual information, as well as a comprehensive set of commonly used measurements from performance, physiological and subjective aspects
- Challenges:
  - ➢ How to effectively control participant's cognitive load under the driving context?
  - ➢ Incorporating a large number of sensors
    - – Both in terms of design considerations, and during the collecting process

# Dataset Overview

The eDREAM dataset was created to facilitate research on using advanced sensor and/or vision technologies to analyze cognitive loads of drivers.

- eDREAM = "Enhancing Driver Interaction with Digital Media through Cognitive Monitoring"

- 37 participants: experienced drivers, gender-balanced, age under 35
- 3 driving sessions: a different level of cognitive load (modeled by secondary tasks) in each drive.
- Mid-fidelity fixed-base simulator



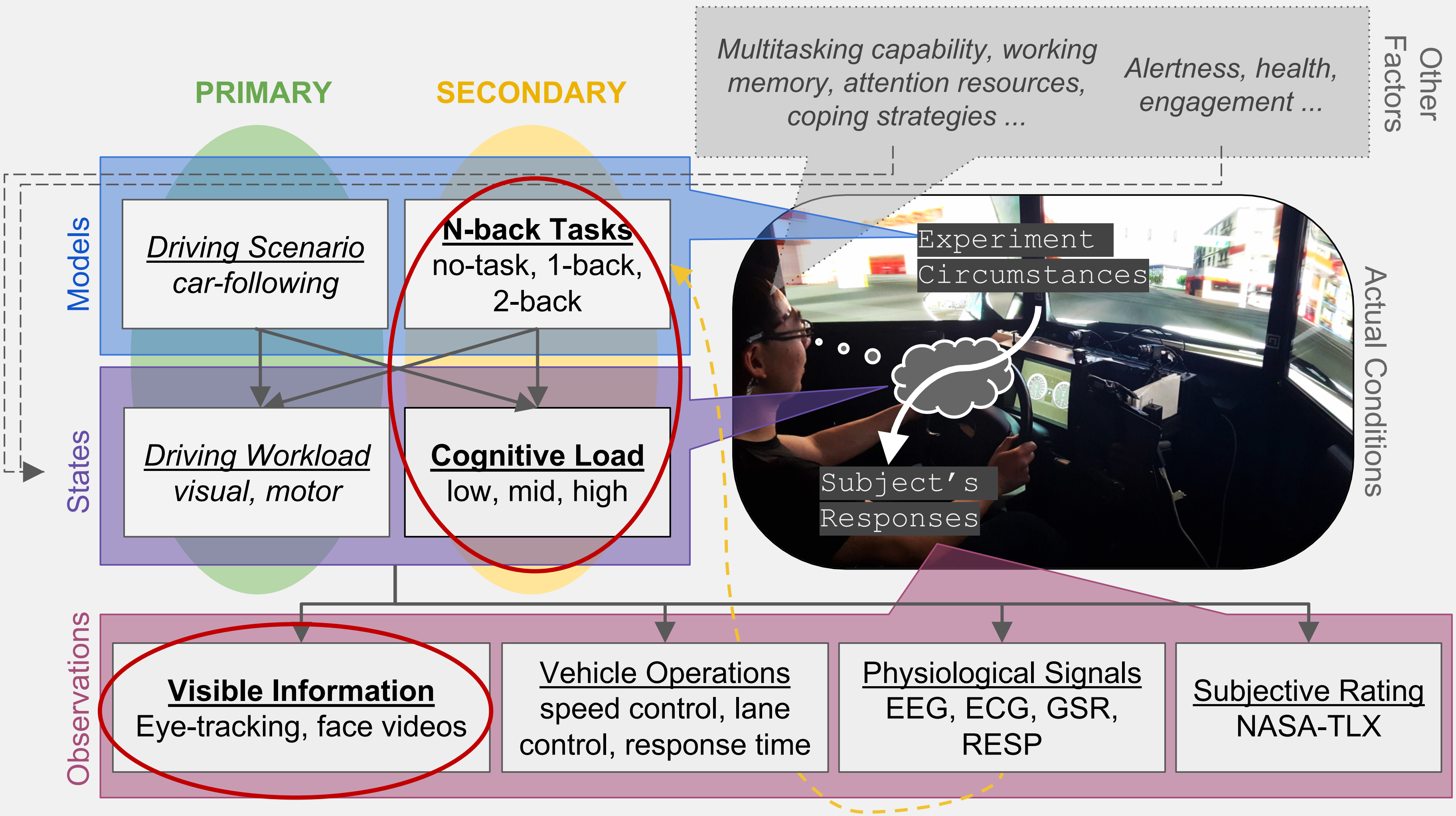

Dataset Webpage

# Modeling Cognitive Load (the *n*-back Task)

- Each task is an audio recording of 10 letters, participant need to count how many n-back patterns are presented
  - ➢ 1-back: two identical letters appeared in pairs
    - e.g. "C B H **H** C A C B F B", answer: 1
  - ➢ 2-back: two identical letters appeared in pairs with one letter in between
    - e.g. "C B H H C A **C** B F **B**", answer: 2
- The load-factor ("n") controls the number of items the participant is required to maintain and process cognitively
  - ➢ Used extensively in neuroscience and psychology [1,2], adapted for driving in [3].
  - ➢ 3 cognitive load levels: no-task → low, 1-back → medium, 2-back → high

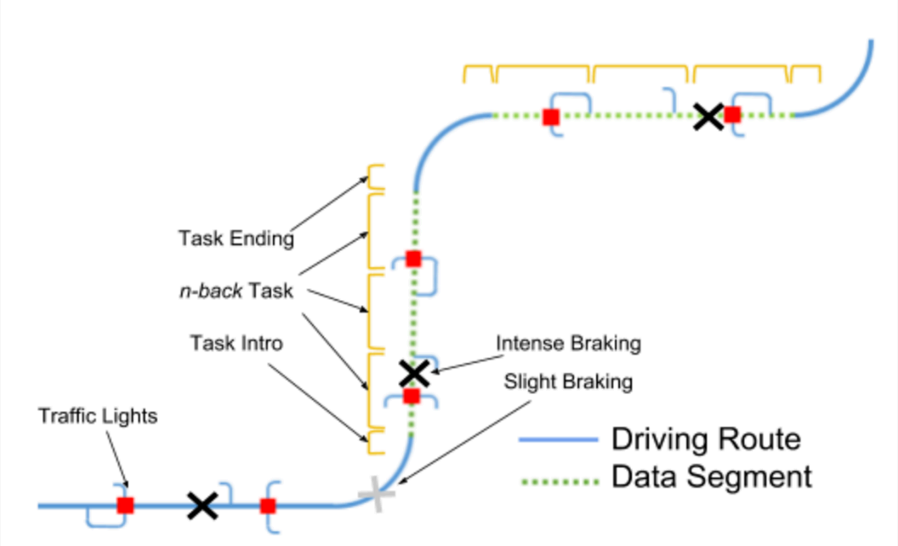
[1] C. H. Chatham, et al., "From an executive network to executive control: a computational model of the n-back task," Journal of cognitive neuroscience, 2011.
[2] S. M. Jaeggi, et al., "The concurrent validity of the n-back task as a working memory measure," Memory, 2010.
[3] B. Mehler et al., "Mit agelab delayed digit recall task (n-back)," Cambridge, Massachusetts Institute of Technology, 2011.

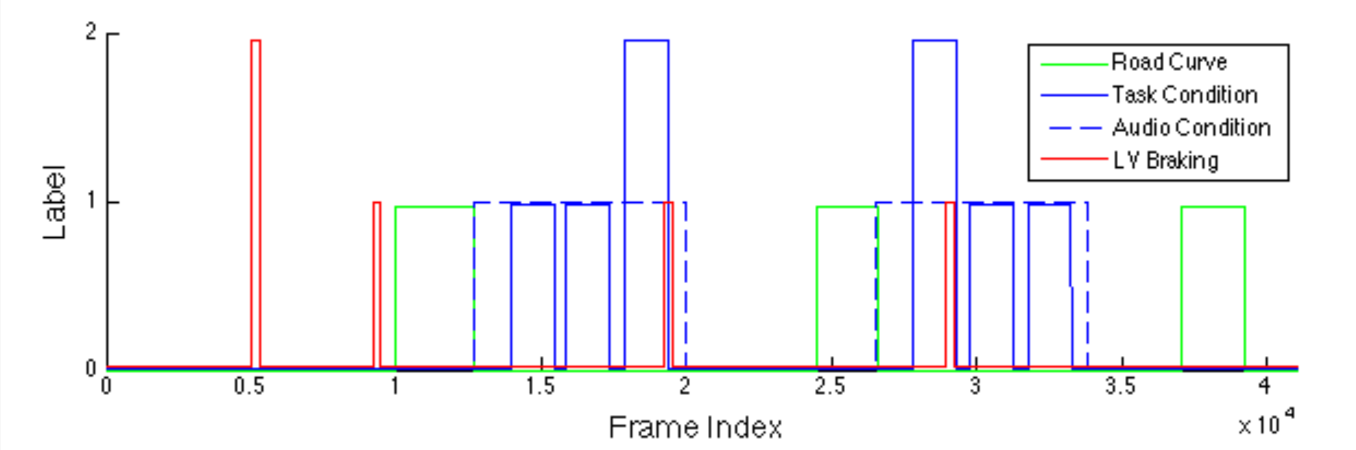
# Experimental Design



# Focus Periods

- Control of conditions are recorded in driving simulator logs:
  - ➢ Primary task: driving conditions
  - ➢ Secondary task: presence of n-back task
- Identify "focus periods" where the cognitive load is optimally controlled




Example log of conditions over one driving session (sampling frequency: 60 HZ)

# Contribution 1: Data Collection

- Objective: gather drivers' responses under differing cognitive load.
  - ➢ Necessary to support future studies focusing on driver cognitive load
  - ➢ Considers a wide range of measures concurrently (visual, performance, physiological and subjective)
- Outcomes:
  - ➢ Design and implementation of a driving experiment with three levels of cognitive load.
  - ➢ Completed a comprehensive dataset consists a total of eight measurements.

# Estimation Method Development



The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

FACULTY
OF APPLIED
SCIENCE &
ENGINEERING

# Contribution 2: Estimation Method Development

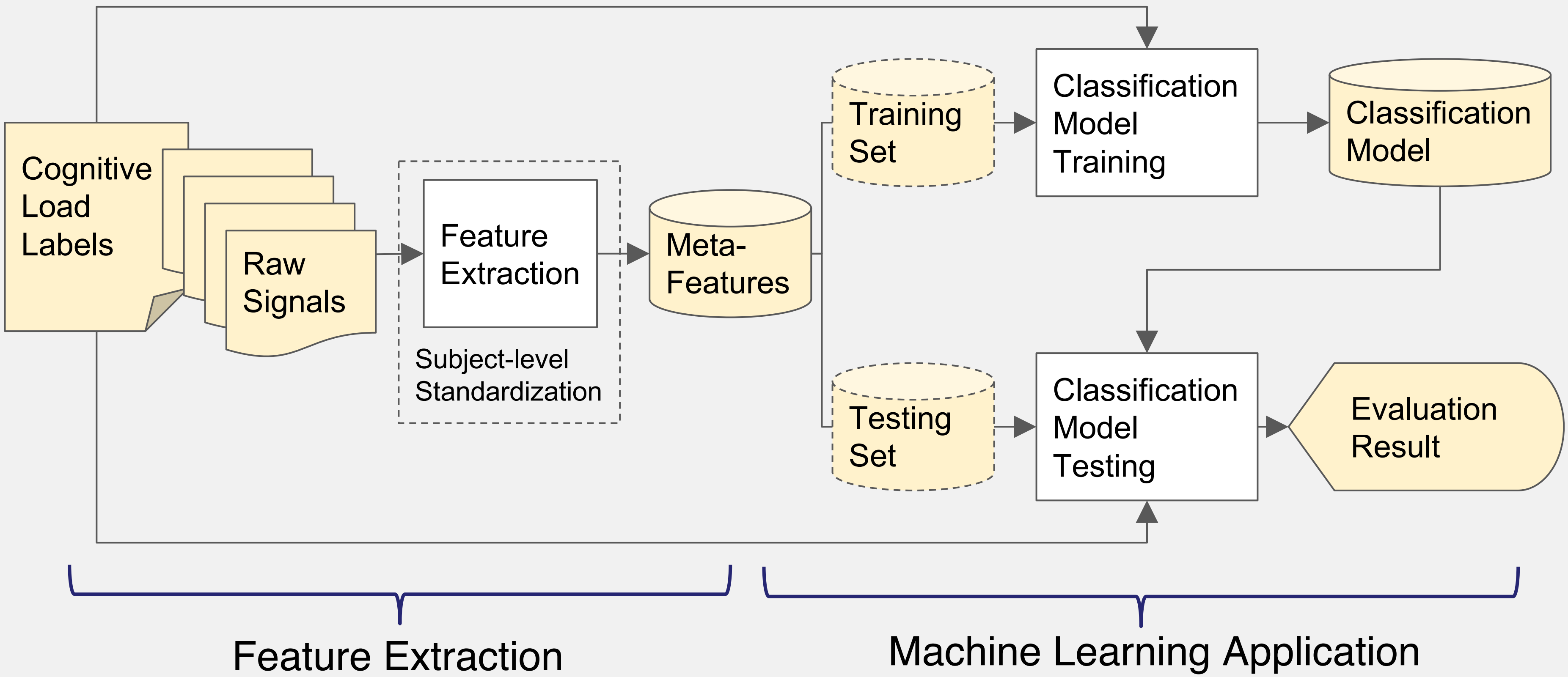
- Objective: explore the feasibility of estimating driver cognitive load based on **visual attention information**.
  - ➢ Explore features and algorithm that could extract predictive information
- Challenges:
  - ➢ Lack of established features and algorithms from previous studies
    - Some of the features might not be compatible
  - ➢ Inconsistent training/testing procedures when applying machine learning algorithms

# Prior Works

| Research Focus | Feature | Algorithm | Evaluation Result |
| --- | --- | --- | --- |
| Preliminary exploration of using machine learning [1] | Gaze fixation duration, pupil diameter, lane deviation | Decision tree trained with 20-trial boosting | Achieved accuracy (ACC) of 81.2% with 30-sec window |
| Real-time detection of cognitive distraction [2] | Eye movement pattern, vehicle measures | Support Vector Machines (SVMs) | Average ACC of 83.1%. |
| Detecting added arithmetic tasks during driving [3] | Mean and SD of gaze rotation, head rotation | AdaBoost with decision stumps, SVMs | Average ACC of 81.6% with AdaBoost, and 77.1% with SVMs. |
| Classification between higher/lower cognitive distraction (based on continuous rating) [4] | Statistics of Facial action units, visual attention, auditory responses, vehicle measures | KNN, SVMs, Linear Bayes Normal Classifier (LDC) | F-score of 0.794 with LDC, 0.681 with KNN and 0.790 with SVM (linear kernel). |

[1] Y. Zhang et al., "Driver cognitive workload estimation: A data-driven perspective", Proc. of ITSC, 2004.
[2] Y. Liang et al, "Real-time detection of driver cognitive distraction using support vector machines", IEEE Transactions on Intelligent Transportation Systems, 2007.
[3] M. Miyaji, H. Kawanaka, and K. Oguri, "Driver's cognitive distraction detection using physiological features by the AdaBoost," in ITSC'09, 2009.
[4] N. Li and C. Busso, "Predicting perceived visual and cognitive distractions of drivers with multimodal features," IEEE Transactions on Intelligent Transportation Systems, 2015

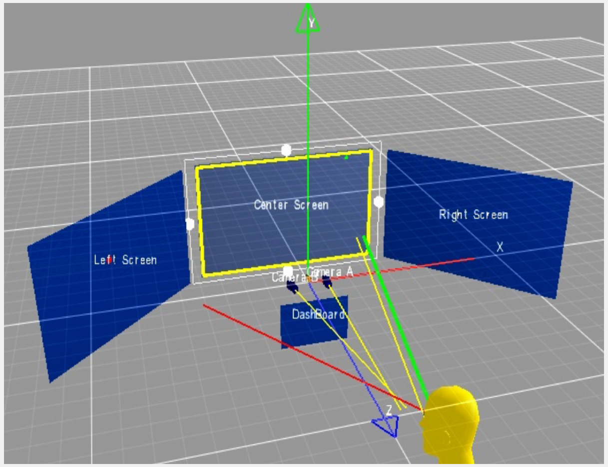
# High-level Framework

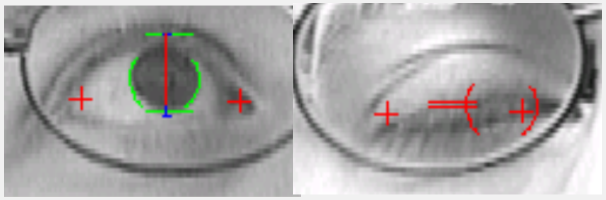



- Classification of cognitive load level is framed as a supervised learning problem:
  - Estimating the target class (no-task, 1-back, 2-back) based on meta-features extracted from raw signals (eye-tracker recordings).

# Meta-Feature Extraction

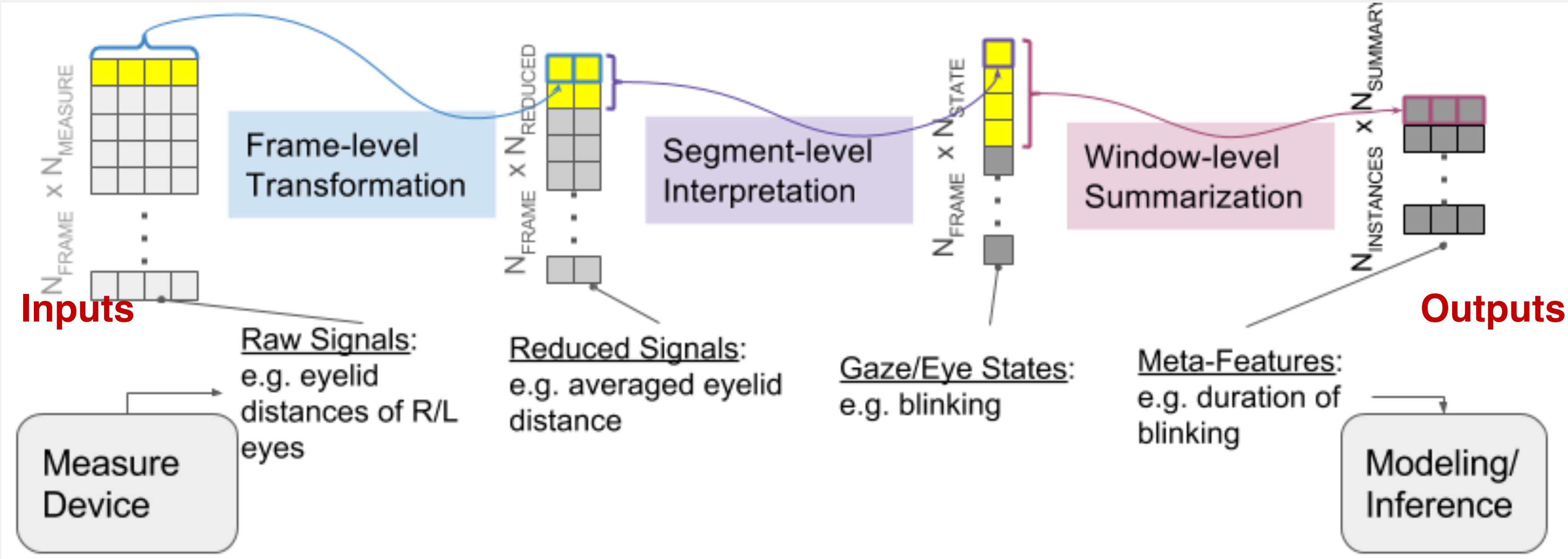
| Prior Knowledge | Proposed Meta-features | Raw signals (from eye-tracker) |
|---|---|---|
| Gaze concentration under high load[1] | Duration and count of off-center glances within 10-sec | GAZE_ROT: a pair of Euler angles for rotations in pitch and yaw (in radians). |
| Conflict of visual and cognitive attention [2] | Duration and count of large eye closures within 10-sec | EYE_CLOS: The fraction of the iris covered by eye-lids |



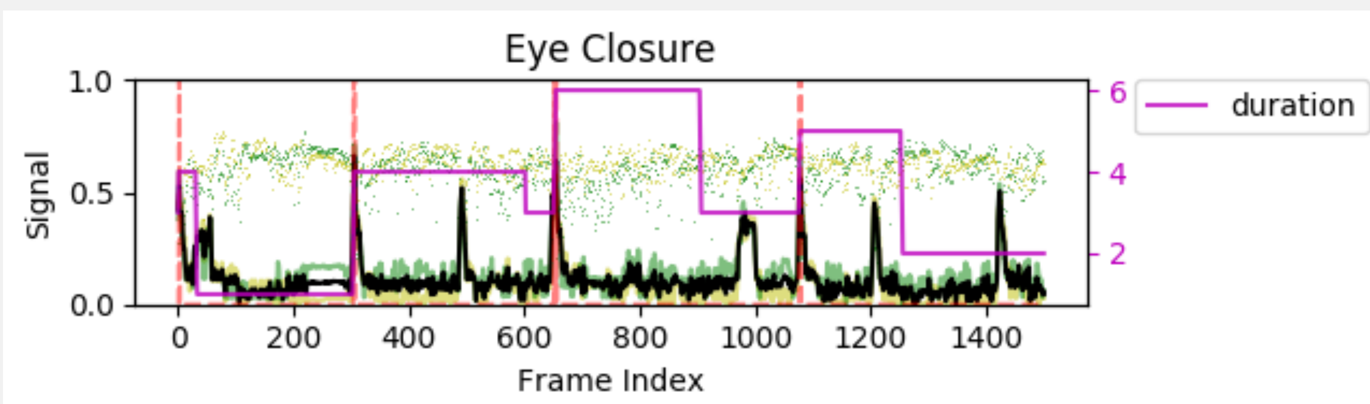



[1] J. L. Harbluk et al., "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance," Accident Analysis & Prevention, 2007.
[2] M. A . Recarte et al., "Mental workload and visual impairment: Differences between pupil, blink, and subjective rating," The Spanish journal of psychology, 2008.

# Meta-Feature Extraction Process




# Example



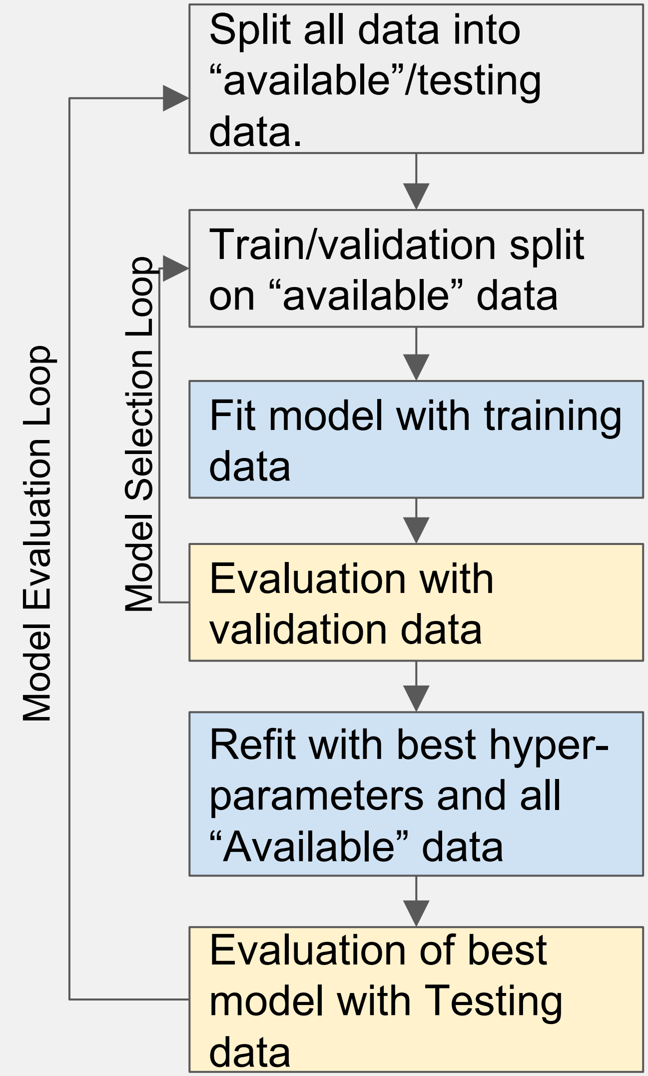

Note: The sampling frequency is 60 frames per second.

# Machine Learning Workflow

- Inputs: four meta-features
- Targets: cognitive load levels
- Algorithms: five candidates
- Evaluation:
  - ➢ Cross Validation (CV) is applied for:
    - – **Model selection** (hyper-parameter)
    - – **Model evaluation** (testing)
      - – 5-fold CV with three data grouping methods
  - ➢ Scoring Metric: accuracy (ACC)

$$Accuracy(\hat{y}, y) = \frac{1}{n_{samples}} \sum_{n_{samples}-1}^{0} [\hat{y}_i = y_i],$$

where [] represents the indicator function.



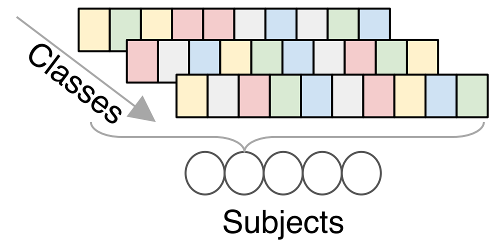
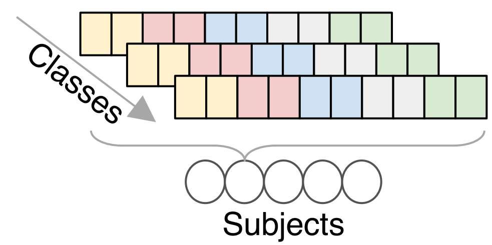
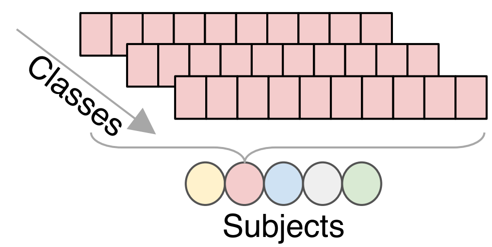
# Algorithms and Hyper-Parameters

| Algorithm Name | Implementation/ Optimization Details | Hyper-Parameters | Hyper-Parameter Range |
|---|---|---|---|
| K Nearest Neighbors (KNN) | Distance calculation: Minkowski metric. | # of neighbors | 1, 5, 10, 20, 50, 100 |
| Logistic Regression (LR) | Optimization algorithm: Newton conjugate gradient | Inverse of regularization strength | $2^{-5}, 2^{-4}, \ldots, 2^{5}$ |
| Support Vector Machines (SVM) | Soft margin is applied. Transfer kernel: Radial Basis Function (RBF) | Influence of each training sample | $2^{-5}, 2^{-4}, \ldots, 2^{5}$ [1] |
| | | Cost of misclassifying samples | $2^{-5}, 2^{-4}, \ldots, 2^{5}$ [1] |
| AdaBoost | Base classifier is CART decision tree with maximum depth of 3. | # of estimators | 5, 10, 100, 500, 1000 |
| | | Learning Rate | 0.001 |
| Random Forest | Same base classifier. The algorithm | # of estimators | 5, 10, 50 |

[1] Y. Liang et al., “Real-time detection of driver cognitive distraction using support vector machines”, IEEE Transactions on Intelligent Transportation Systems, 2007.

# Evaluation Procedure: Grouping for CV

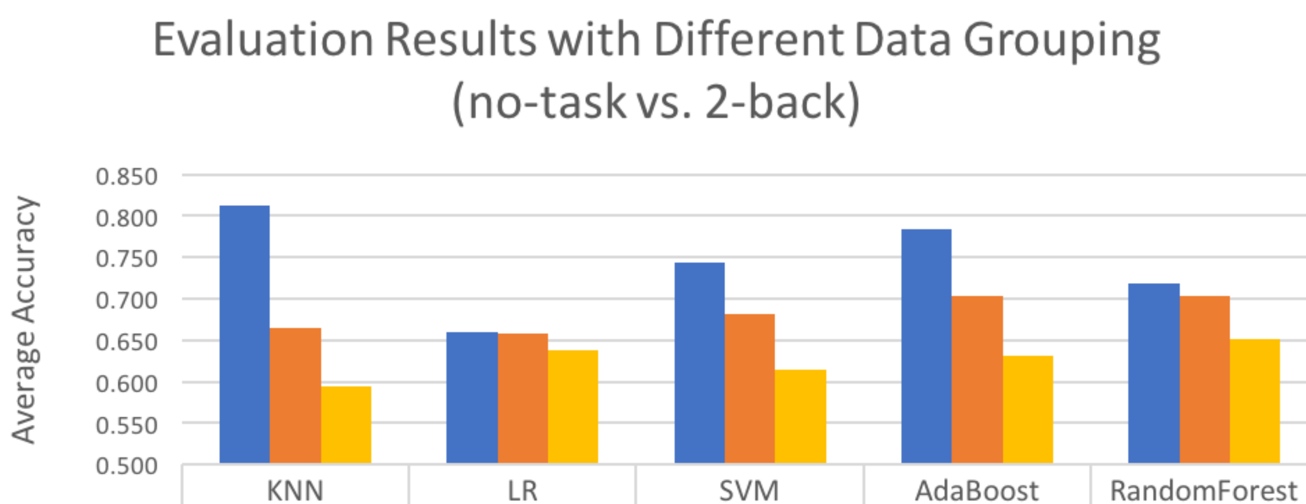
- Evaluation of the subject-independent model can be performed at three difficulty levels:

| Grouping Method | Correspondent Scenario | Implementation | Illustration Color-code for folds 0 1 2 3 4 |
|---|---|---|---|
| None | Data instances are i.i.d. | Drawing samples into training set or testing set randomly |  |
| Time-based | Train a model with data from some subjects, and apply the model to predict data of unseen periods from the same subjects | Group the data from the same run into several blocks, and cross validate at the group level |  |
| Subject-based | Train a model with data from some subjects, and apply the model to predict data from unseen subjects | Always put the data from a subject into one fold |  |

*Closer to practical application*

# Result: Binary

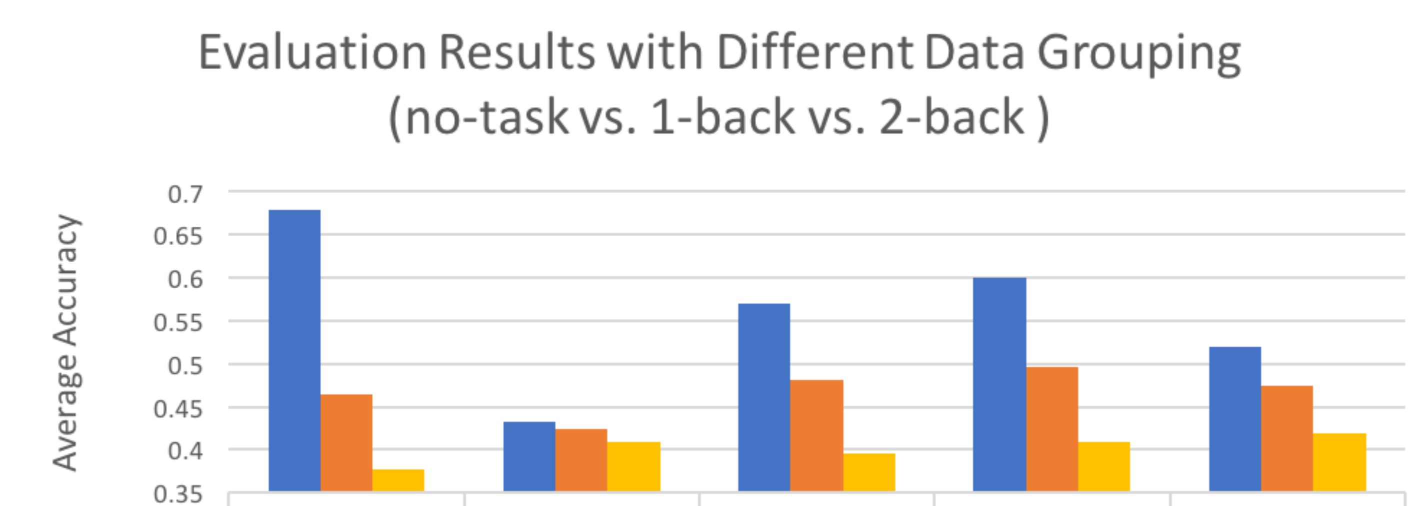
- The grouping method applied in evaluation clearly impacted the results.
  - KNN experienced most significant impact
  - Random Forest and LR are more robust against overfitting.




| | KNN | LR | SVM | AdaBoost | RandomForest |
|---|---|---|---|---|---|
| No-grouping | 0.813* | 0.660 | 0.744 | 0.784 | 0.719 |
| Time-based | 0.665 | 0.658 | 0.681 | 0.703 * | 0.703 * |
| Subject-based | 0.594 | 0.638 | 0.615 | 0.631 | 0.652 * |

# Result: Ternary

- Better-than-guess performance when the same approach for binary classification is adopted for the ternary case
  - Ensemble of decision trees methods are slightly more optimum when evaluating using grouping.




| | KNN | LR | SVM | AdaBoost | RandomForest |
|---|---|---|---|---|---|
| No-grouping | 0.679* | 0.433 | 0.570 | 0.599 | 0.520 |
| Time-based | 0.464 | 0.424 | 0.480 | 0.496* | 0.474 |
| Subject-based | 0.377 | 0.409 | 0.396 | 0.410 | 0.419* |

# Contribution 2: Estimation Method Development

- Objective: explore the feasibility of estimating driver cognitive load based on **visual attention information**.
  - ➢ Explore features and algorithm that could extract predictive information
  - ➢ Determining the appropriate procedure for evaluating models with practical meanings
- Outcomes:
  - ➢ Proposed method to estimate cognitive load levels based on visual attention information
    - Designed more flexible meta-feature based on prior knowledge
    - Applied five classification algorithms for automatic information extraction
  - ➢ Discussed the effect of different training/testing data partitioning methods

# Conclusions

- Data collection featuring driver cognitive load data:
  - ➢ Requires considering both the primary task (driving) and the secondary task.
- Proposed estimation method:
  - ➢ Visual attention information carried estimation power as all classifiers achieved better-than-guess performance.
  - ➢ Not sufficient to be relied alone if high accuracy is desired
- Evaluation framework:
  - ➢ Explicit grouping based on time or subject would be desired to carry more practical significance
  - ➢ The larger the grouping unit is, the more challenging the problem seems to become

# Future Work

- Incorporate more observations (e.g. vehicle speed) and consider strategies for combining feature values
- Employ time-series models and more intelligent method for hyper-parameter search
- Application to naturalistic dataset collected with instrumented vehicle over larger population and longer time range (e.g. SHRP2 [1])

[1] Campbell, Kenneth L. "The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety." TR News , 2012.

# Thank you very much for your attention.

# What is Cognitive Load?

Examples for comparing cognitive load with cognitive distraction and high arousal:

| High Cognitive Load | Distraction | High Arousal | Possible Driver Situation |
|---|---|---|---|
| 0 | 0 | 0 | Calm, comfortable driving. |
| 0 | 0 | 1 | Nervous due to bad weather (e.g. snow storm). |
| 0 | 1 | 0 | Operating air conditioning controls. |
| 0 | 1 | 1 | Crying baby in the backseats. |
| 1 | 0 | 0 | Daydreaming, or listening to radio. |
| 1 | 0 | 1 | Following GPS on unfamiliar routes stressfully. |
| 1 | 1 | 0 | Attempting to interact with a voice-command system. |
| 1 | 1 | 1 | Frustrated by an important phone conversation. |

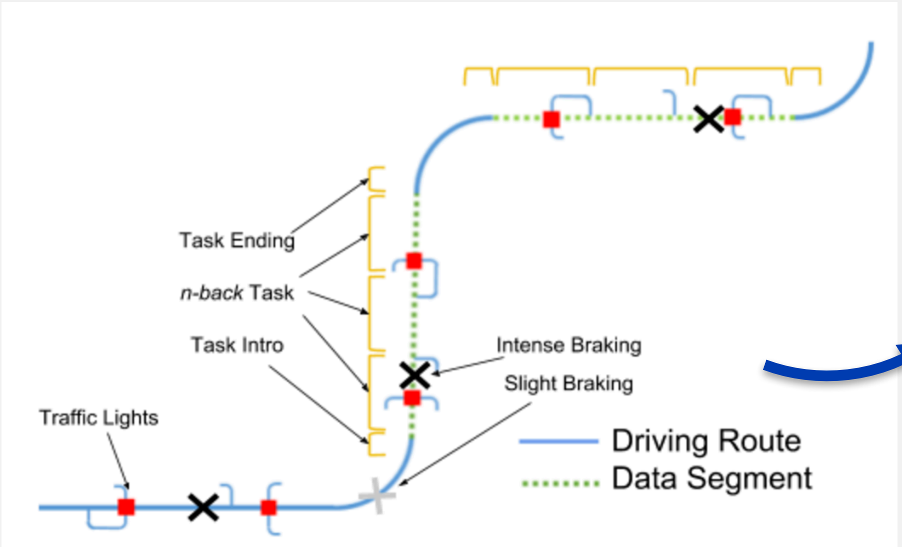
# Data Collection

- Experimental design: 2015 summer
- Implementation: 2015 fall
- Pilot testing: 2016 spring
- Data collection: 2016 summer-fall
- Data organization: 2016 winter

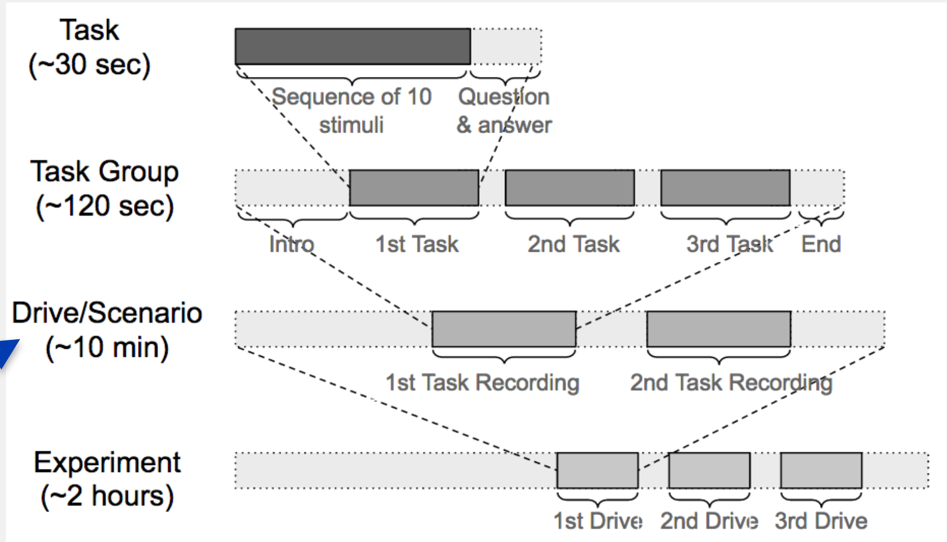
| Modality | Information | Size |
|---|---|---|
| EEG | Brain electrical activities recorded from four positions | 7 GB |
| Physiological (ECG, GSR, respiration) | Heart rate, skin conductance and breath depth/rates | 1.6 GB |
| Vehicle | Vehicle, brake pedal and steering wheel states | 1.8 GB |
| Eye-tracking | Head position, gaze position and eye closure information | 23 GB |
| Videos | Recordings from a participant-facing colour camera | 33 GB |
| Subjective Ratings | Perceived task-load level in sub-categories (e.g. mental demand) | <1MB |

# Experimental Conditions

- Primary task: following a Lead Vehicle (LV) at 40 Mph on a 4-lane urban route
  - ➢ No turning or merging, but brakes abruptly at specific moments
- Secondary task: completing 6 tasks per n-back drive (none in the no-task drive)
  - ➢ The ordering of presenting 3 cognitive load is counterbalanced
- Carefully controlling various conditions: route and road, traffic and pedestrian, presence of *n*-back tasks




*Spatial map*




*Scheduling*

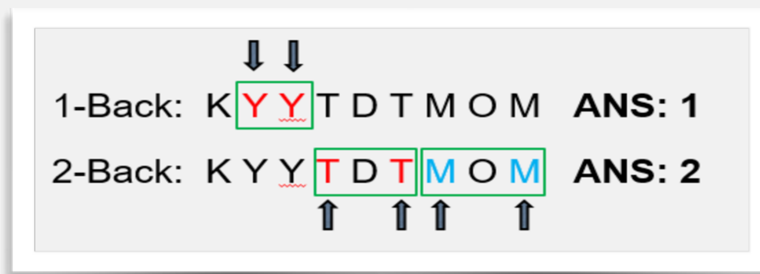
# More about *n*-back Tasks

Implementation of the *n*-back task:

- During specific periods (“focus periods”) Participants listen to a sequence of letters and count how many times the current stimuli is identical to the one presented n-steps ago.
- Used extensively in neuroscience and psychology for collecting physiological data or individual performance differences.
- Adapted for driving studies as a surrogate task [1].

n-back involves multiple cognitive processes:

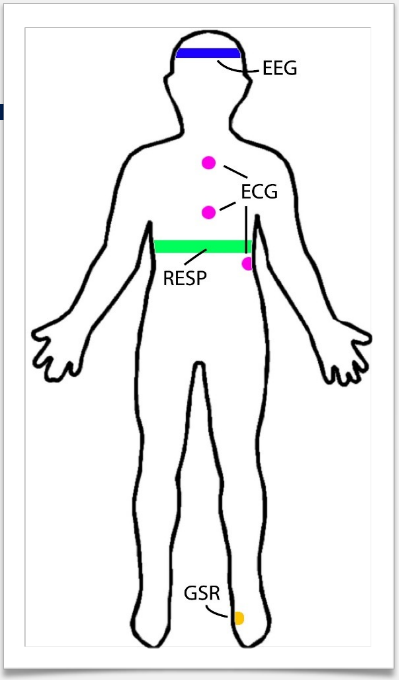
- Perceiving and encoding incoming stimuli
- Maintenance and updating of memory
- Matching/analyzing/selecting of materials




Example of *n*-back task

*[1] “MIT AgeLab Delayed Digit Recall Task (n-back) ”, B. Mehler, B. Reimer and J. A. Dusek*

# Apparatus

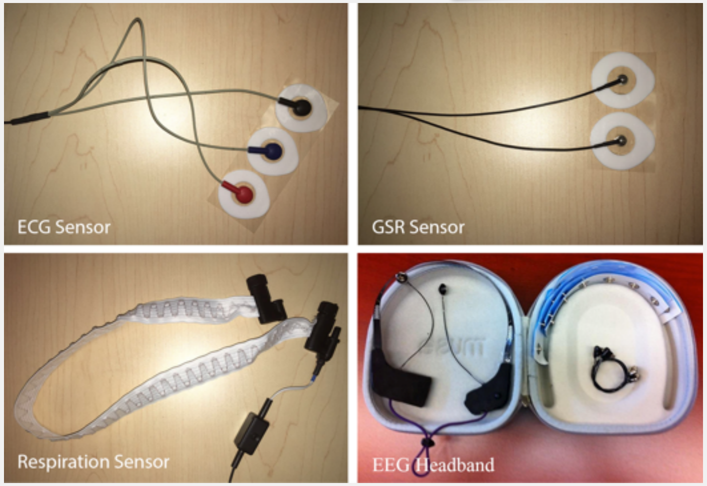



## Driving Simulator:

- miniSim by NADS
  - ➢ Also records driver operations, and various vehicle measurements (e.g. speed or lane deviation)

## Physiological Sensors:

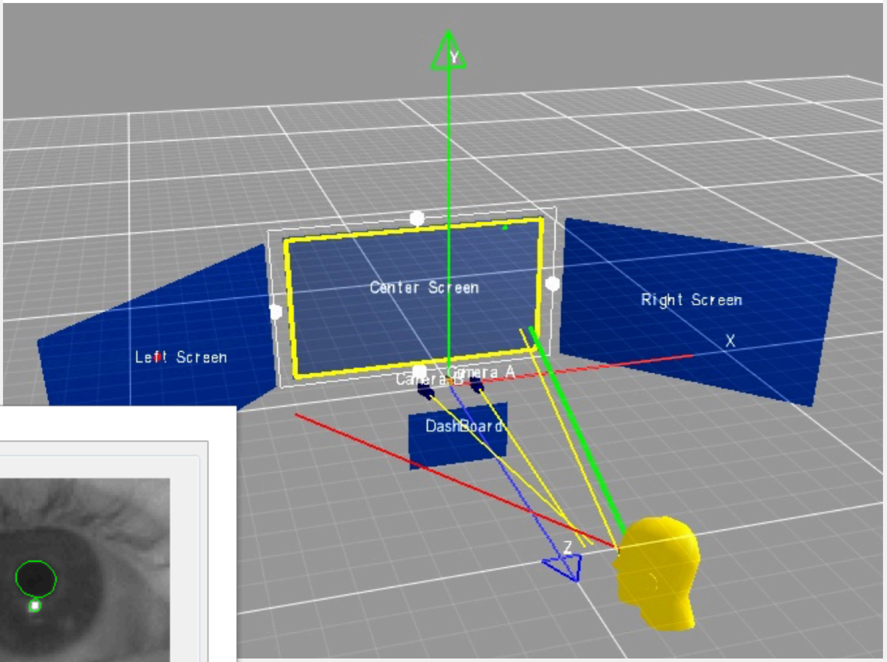
- EEG (Muse headband), ECG, GSR, Respiration

# Apparatus

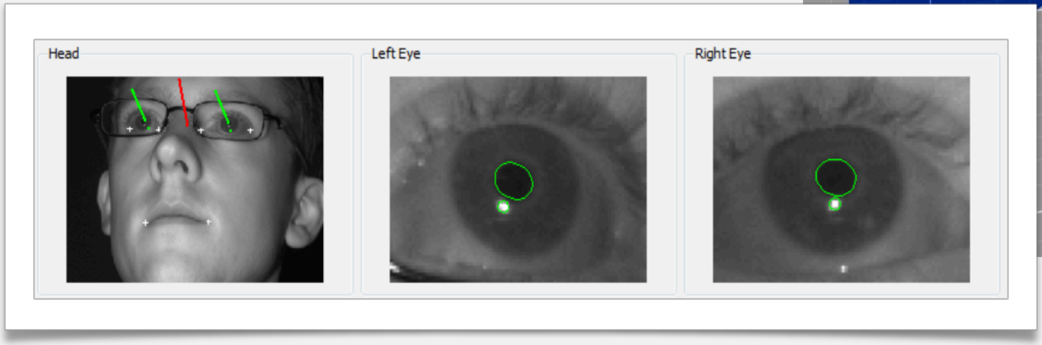
**Camera and eye-tracker:**

- Participant-facing color cameras
- A pair of Near-Infrared cameras for the faceLAB eye-tracker



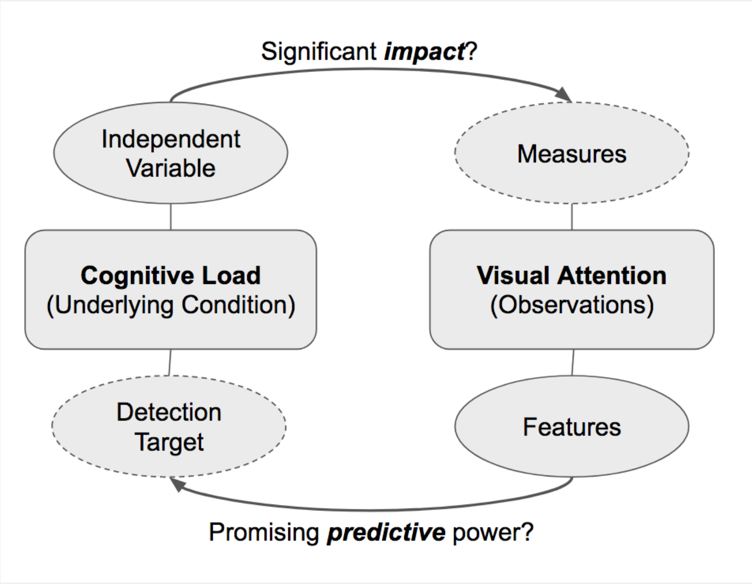

**Subjective ratings:**

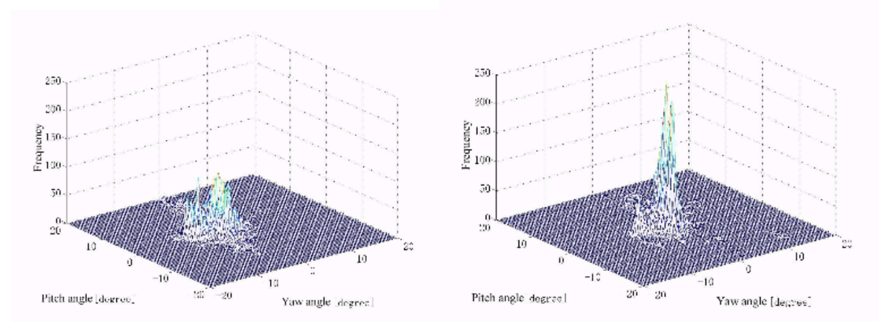
- NASA Task Load Index (NASA-TLX)



# Background:
# Cognitive Loads and Visual Attention

- This study exploits the subjects' **visual attention information** for cognitive load assessment.
  - ➢ It's been shown significantly impacted by added cognitive tasks.
    - – Increased blinking –attention allocated for visual observation is reduced, which could result in increased blinking.
      - – Consistently observed in multiple naturalistic or simulator-based studies.
    - – Concentrated gaze – reduced checking for peripheral environment or devices.
      - – More mixed results due to this measure's sensitivity to visual loads
  - ➢ It was also used as promising features for prediction models.
    - – Detection systems have been proposed



(Ordinary) (Arithmetic)

Fig. 12. Distribution of the frontal focal points

*[1] "Effect of pattern recognition features on detection for driver's cognitive distraction", Miyaji et al, 2010*

# Raw Signals from Eye-tracker

- **EYE_CLOS**: the fraction of the iris covered by eye-lids

$$x_{EC} = 1 - \frac{eyelid\_distance}{iris\_size},$$

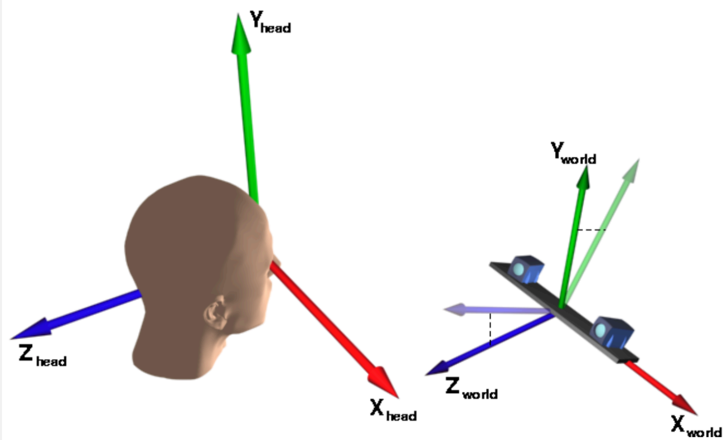
where $eyelid\_distance$ is the distance between top and bottom eyelids, and $iris\_size$ is the iris size, which is 12mm by default.

- **GAZE_ROT**: a pair of Euler angles in radians for rotations around the world x-axis (pitch) and world y-axis (yaw).

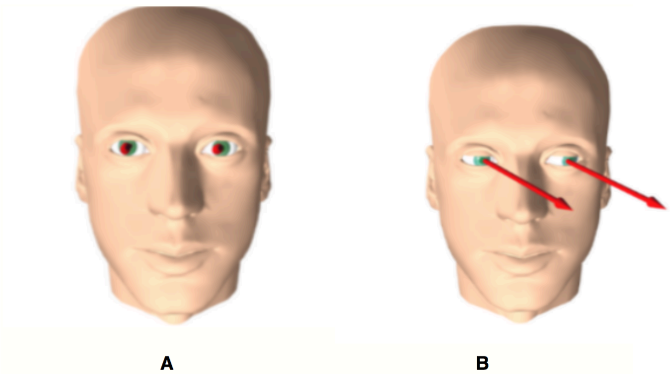
$$\begin{aligned} x_{GD} &= [X_{GD,pitch}, X_{GD,yaw}] \\ &= [arccos(\sqrt{u_x^2 + u_z^2}), arccos(\sqrt{u_y^2})], \end{aligned}$$

where $\vec{u} = [u_x, u_y, u_z]$ is the unit vector pointing from pupil center to the object being looked at in World Coordinate .

- Each raw signal is measured for right and left eyes independently.




*The Head Coordinate Frame shown with the World Coordinate Frame and the Stereo-Head Coordinate Frame*

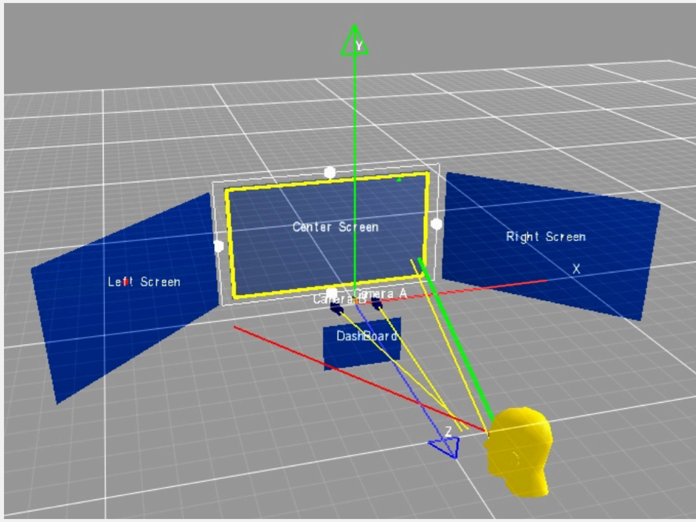



*The gaze direction vector indicates direction in the World Coordinate Frame from the origin of the gaze - the centre of the eye balls for (A) Looking straight ahead (B) looking to the left.*

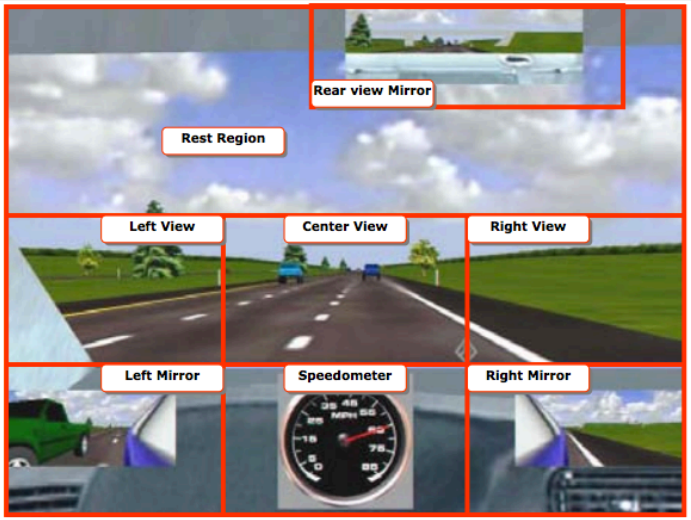
*Graphics from faceLAB manual*

# Initial Inputs: Raw Signals (+1)

- The faceLAB eye-tracker can also estimate:
  - ➢ The gaze fixation location in the world coordinate (in X, Y, Z) or the plane coordinate (in X, Y).
    - This has been used as the base signal previously [1, 2, 3, 4]
    - However, this value depends largely on the setup of the “world” in faceLAB and in the driving environment.
      - The plane in eDREAM is quite small and does not cover side or rear mirrors, which is quite different then the setup in other studies (e.g. [2]).
  - ➢ Therefore, direct analysis of the gaze rotation angle provides better generalization capability.




*The multi-screen setup in eDREAM.*




*The single-screen setup in [2]*

*[1]“Driver cognitive workload estimation: A data-driven perspective”, Y. Zhang et al.*

*[2] “Real-time detection of driver cognitive distraction using support vector machines”, Y. Liang et al.*

*[3] “Detecting Cognitive Workload Using Driving Performance and Eye Movement in a Driving Simulator”, J. Son, M. Park and H. Oh*

*[4] “Impact of Cognitive Task Complexity on Drivers’ Visual Tunneling”, B. Reimer*

# Proposed Meta-Features

Visual attention was found to be impacted by cognitive load in two ways:

- Reduced checking towards peripheral environment or mirror/speedometer [?]
- Loss of attention towards visual perception [?]

Gaze concentration is captured by gaze direction (GD) features

Loss of visual attention is reflected by eye closure (EC) features

| Notation | Description |
|---|---|
| $X_{GD\_DUR}$ | Total duration of gaze-off-center: number of frames that the gaze direction (GD) is deviated from the reference direction by more than the threshold $thres_{GD}$. |
| $X_{GD\_CNT}$ | Count of gaze-off-center times: number of times GD crossed $thres_{GD}$. |
| $X_{EC\_DUR}$ | Total duration of eye closure (EC): number of frames that EC is greater than the threshold $thres_{EC}$. |
| $X_{EC\_CNT}$ | Count of blinking times: number of times EC crossed $thres_{GD}$. |

➢ We propose to compute duration (DUR) and count (CNT) of over-threshold incidents within a 10-second sliding window to capture the interested patterns

– Better compatibility across different data collection setups

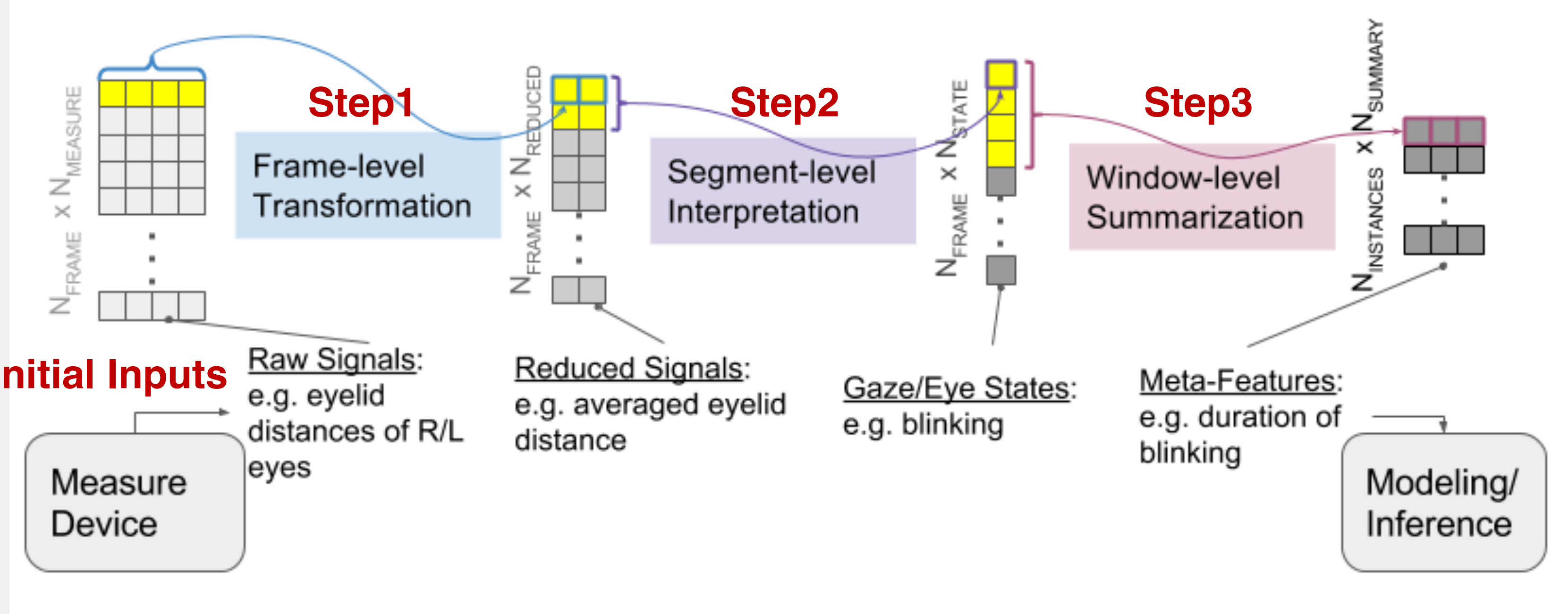
# Visual Attention Meta-Features

- Duration and count of large eye closures are for capturing the changes of blinking behaviors, which are hypothesized to indicate **the amount of visual attention demands** [7].
  - This has also led to use of the following features in previous studies:
    - PERCLOS [5], Mean blink frequency [4, 5]
- Duration and count of off-center glances are for capturing **temporal variation of visual attention direction.**
  - These are similar to the following features in previous studies:
    - Duration and count of glances to center/off-center regions [1, 2, 3]
    - Gaze fixation/pursuit duration [3, 4]
    - SD of fixation position [4, 6, 8]
  - Also commonly considered in previous studies is the spatial characteristics of visual attention (such as mean of fixation position). It is not captured with proposed features.

[1] "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance", J.L. Harbluk et al.
[2] "Sensitivity of eye-movement measures to in-vehicle task difficulty", T.W. Victor, J.L Harbluk, J.A. Engstrom.
[3]"Driver cognitive workload estimation: A data-driven perspective", Y. Zhang et al.
[4] "Real-time detection of driver cognitive distraction using support vector machines", Y. Liang et al.
[5] "Driver distraction detection using semi-supervised machine learning", T. Liu et al.
[6] "Detecting Cognitive Workload Using Driving Performance and Eye Movement in a Driving Simulator", J. Son, M. Park and H. Oh
[7] "Mental workload and visual impairment: differences between pupil, blink and subjective rating", M. Recarte, et al.
[8] "Impact of Cognitive Task Complexity on Drivers' Visual Tunneling", B. Reimer

# Extraction Process

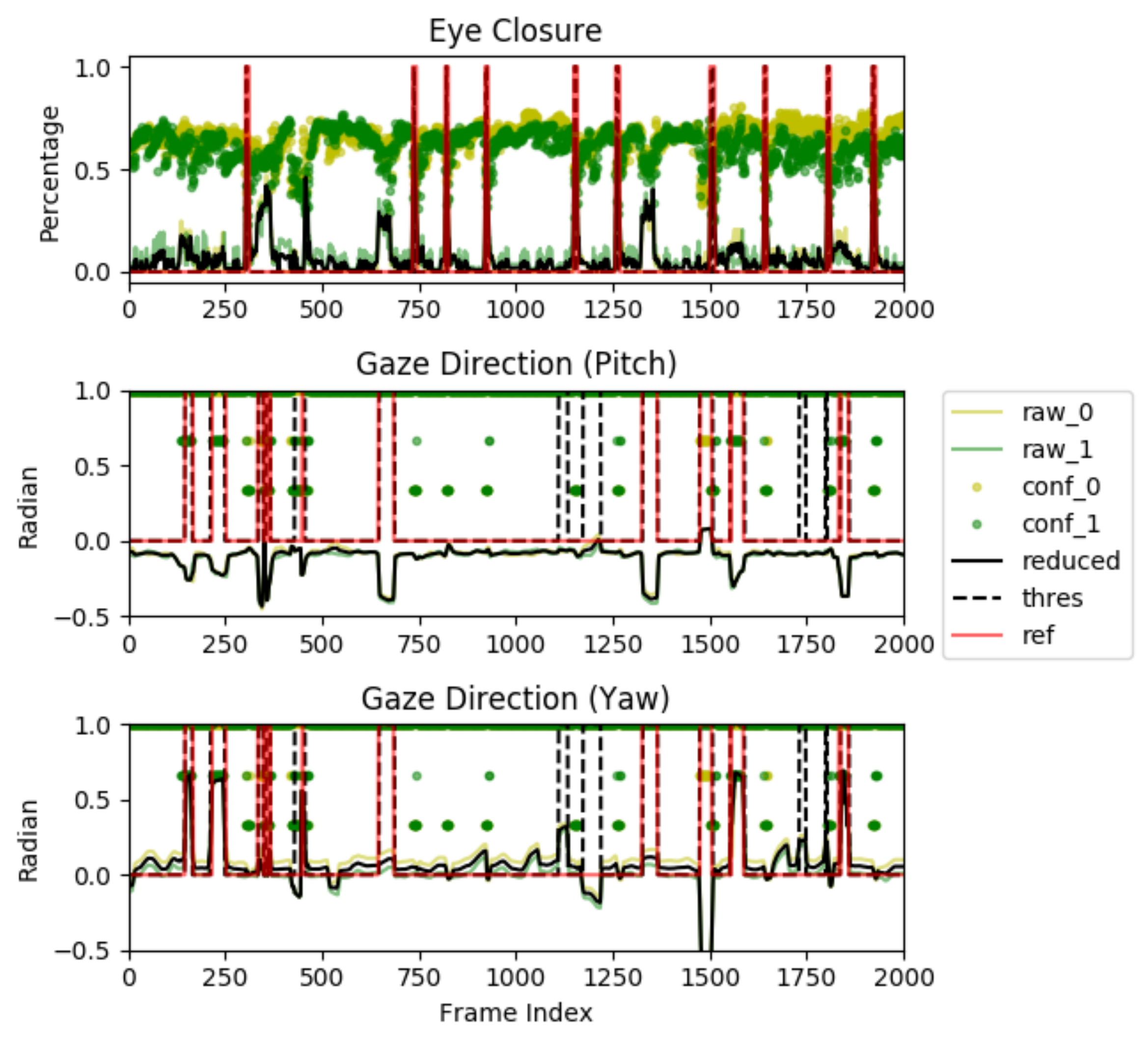
- To obtain the proposed meta-features from the complicated eye-tracking data, an extraction process of multiple levels is designed:
  - ➢Also involves preprocessing and standardize
  - ➢The process could be different for each specific meta-features.
  - ➢They are further explained in the following slides.



# Extraction: Reduction and Interpretation

Given the raw signals from a period-of-interests:

1. Combine values estimated for left/right eyes into one channel (black)
2. Detect blinks from EC, (black-dotted)
   - Mean-removal for the GD signals
3. Calculate the frequency and duration based on the thresholded signal (see next slide)



# Extraction: Window-level Summarization

- Example of extracted "duration" values (in red) on top of the thresholded signals (in dotted blue), taken from Participant 07.
    - ➢ Each row shows a segment of data for a different cognitive load. The feature values' vertical axis are on the right.
- There exists considerable individual differences in the extracted features
    - ➢ Subject-level standardization is performed using reference data from the periods near the beginning of each driving sessions.
        - – Example for duration of off-center glances shown on the right.