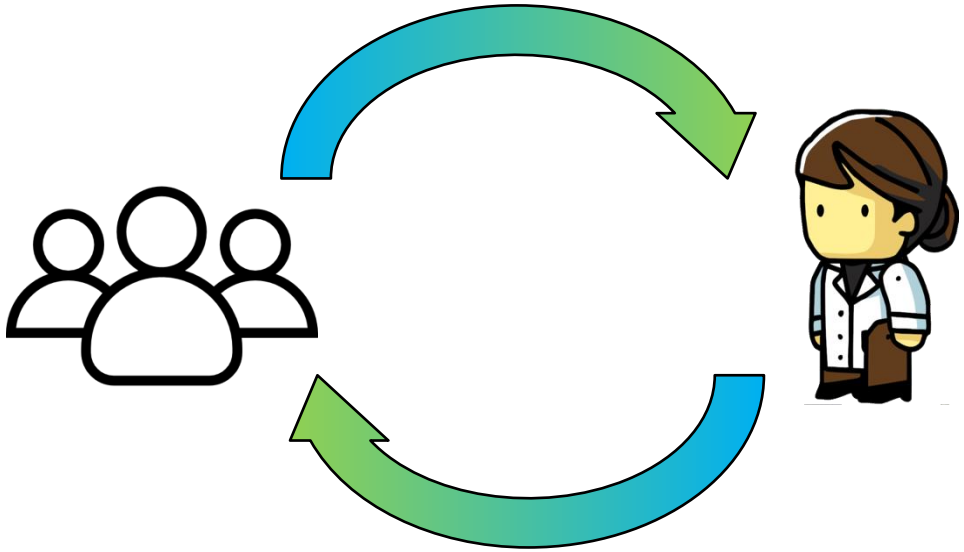
Closing the loop in healthcare: Bringing the benefits of machine learning back to the people



Hsin-Yu (Jane) Lai

NIH/NIDA P50DA054039
NIH/NIBIB and OD P41EB028242

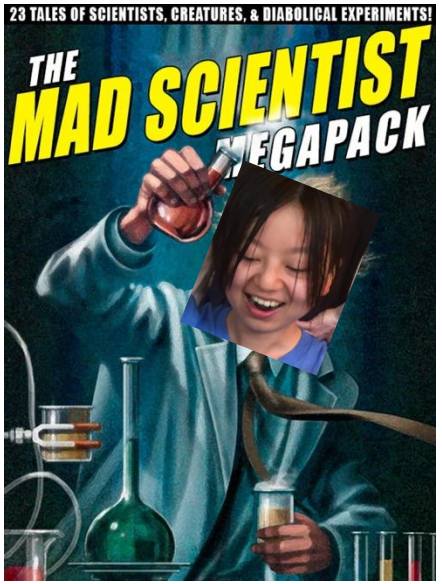
Self introduction?



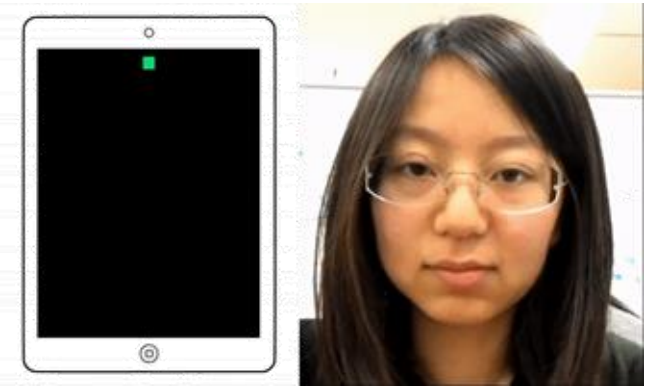
What my friends think



What my parents think



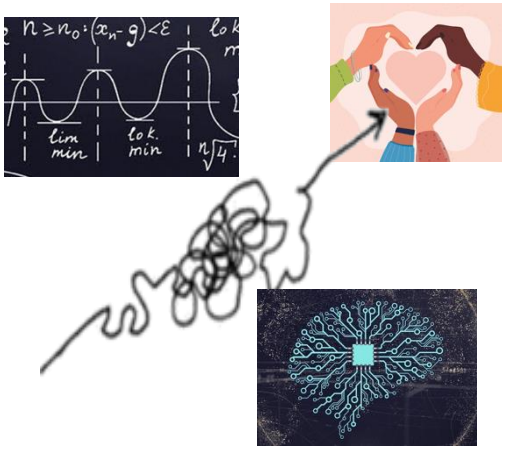
What society think



What my advisors think



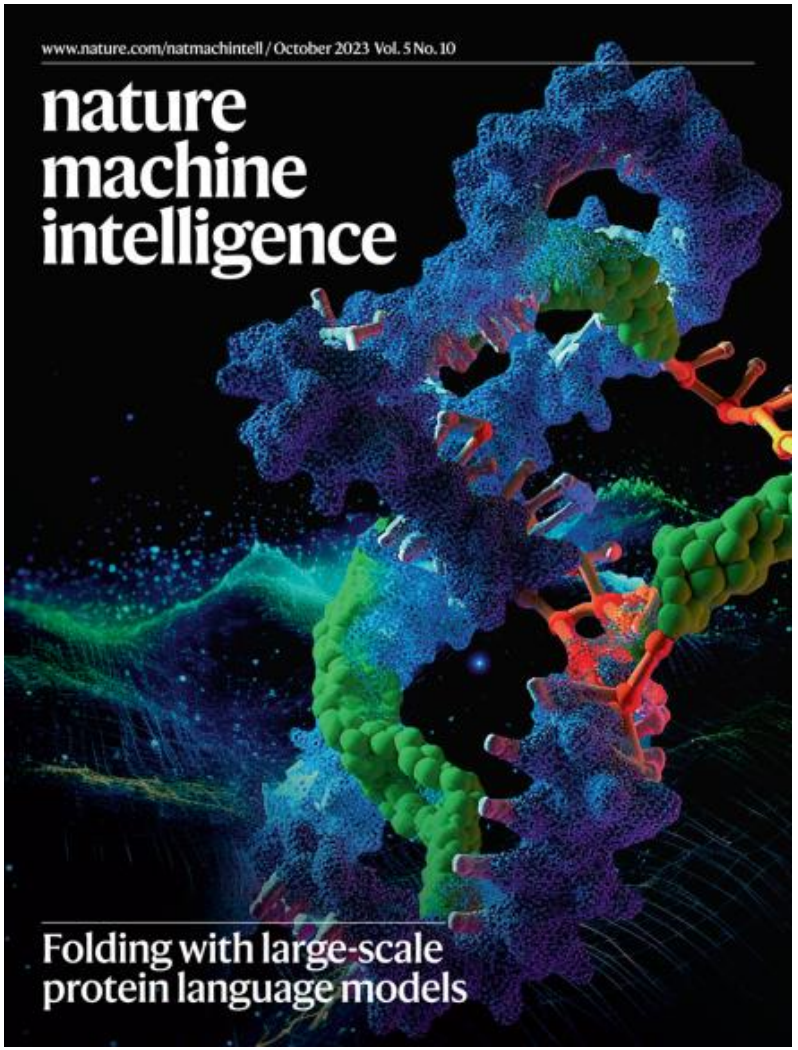
What I think



Who I actually am(?)

Machine learning (ML) has made groundbreaking progress in healthcare applications





ML helps doctors/scientists analyze data

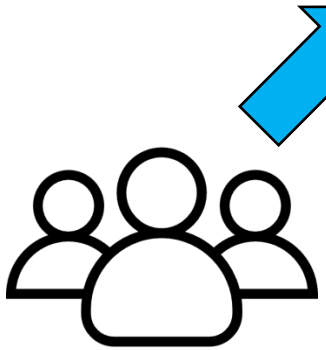
‘Omic data,
Sensor data,
Medical imaging,
Medical records

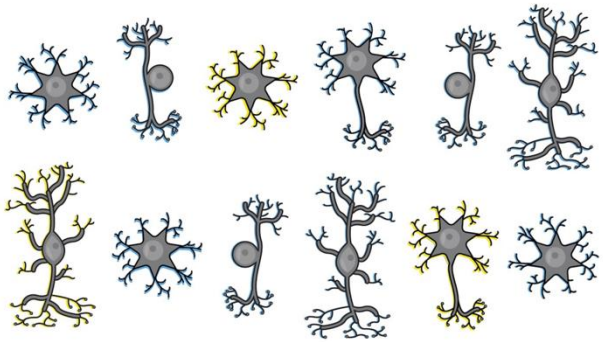
Signal processing,
Deep learning,
Statistics

Scientific discovery
Diagnosis
Treatment design
Disease progression monitoring

Data

Downstream tasks

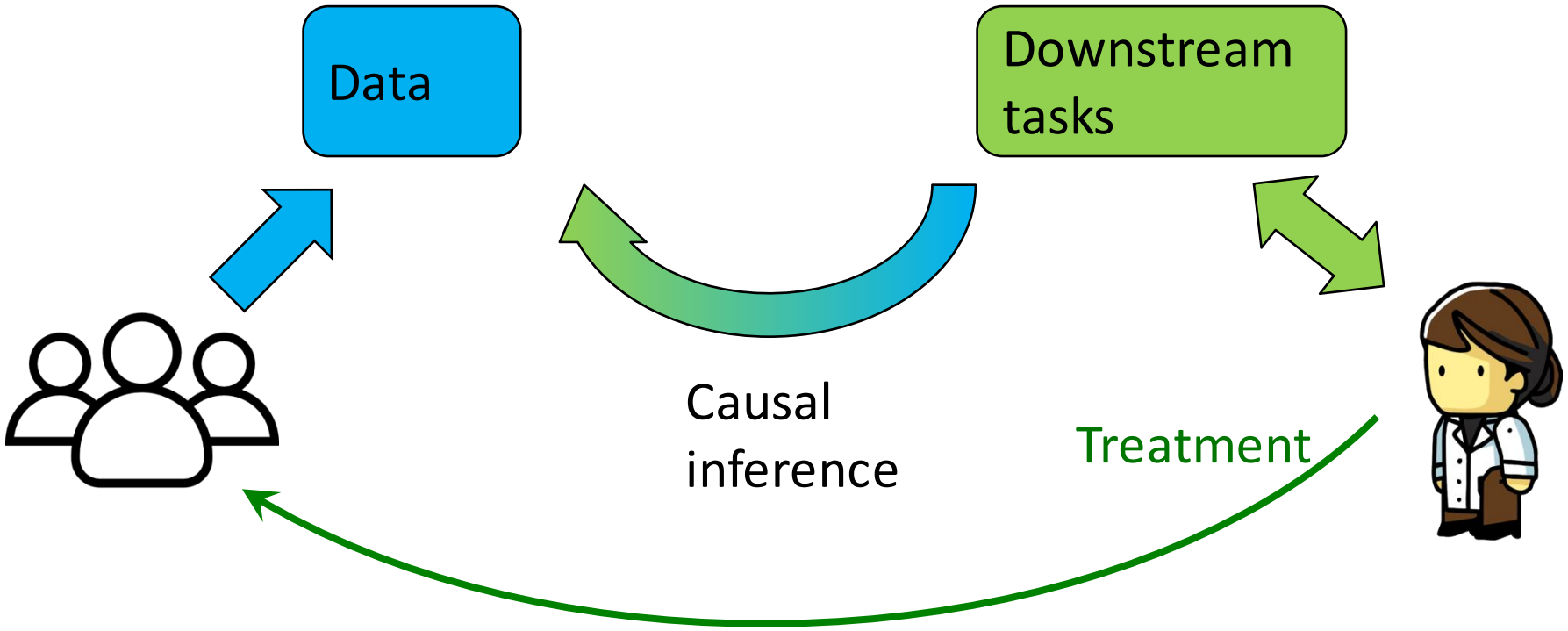




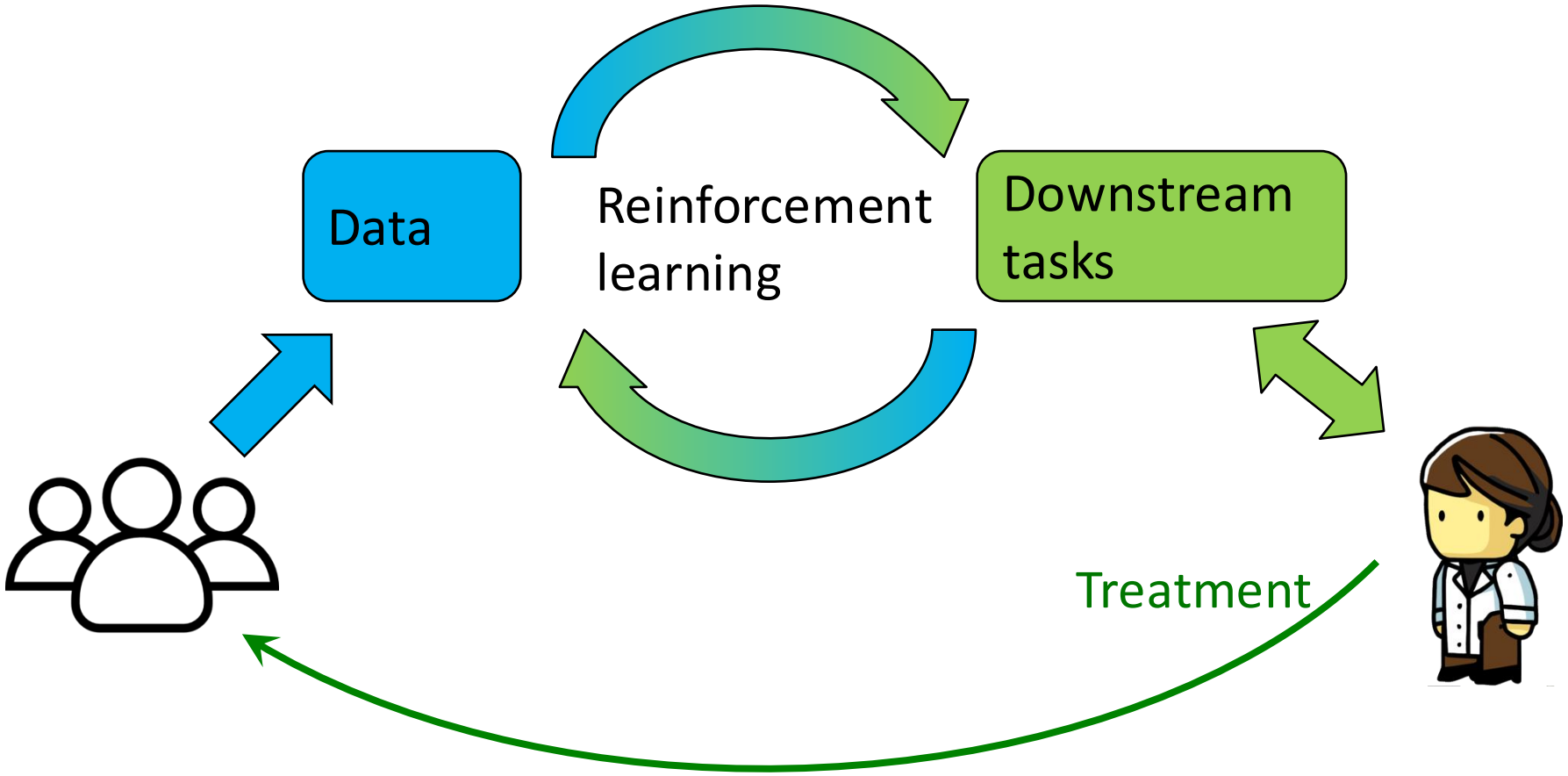
From 08.05.20. Allen Institute News



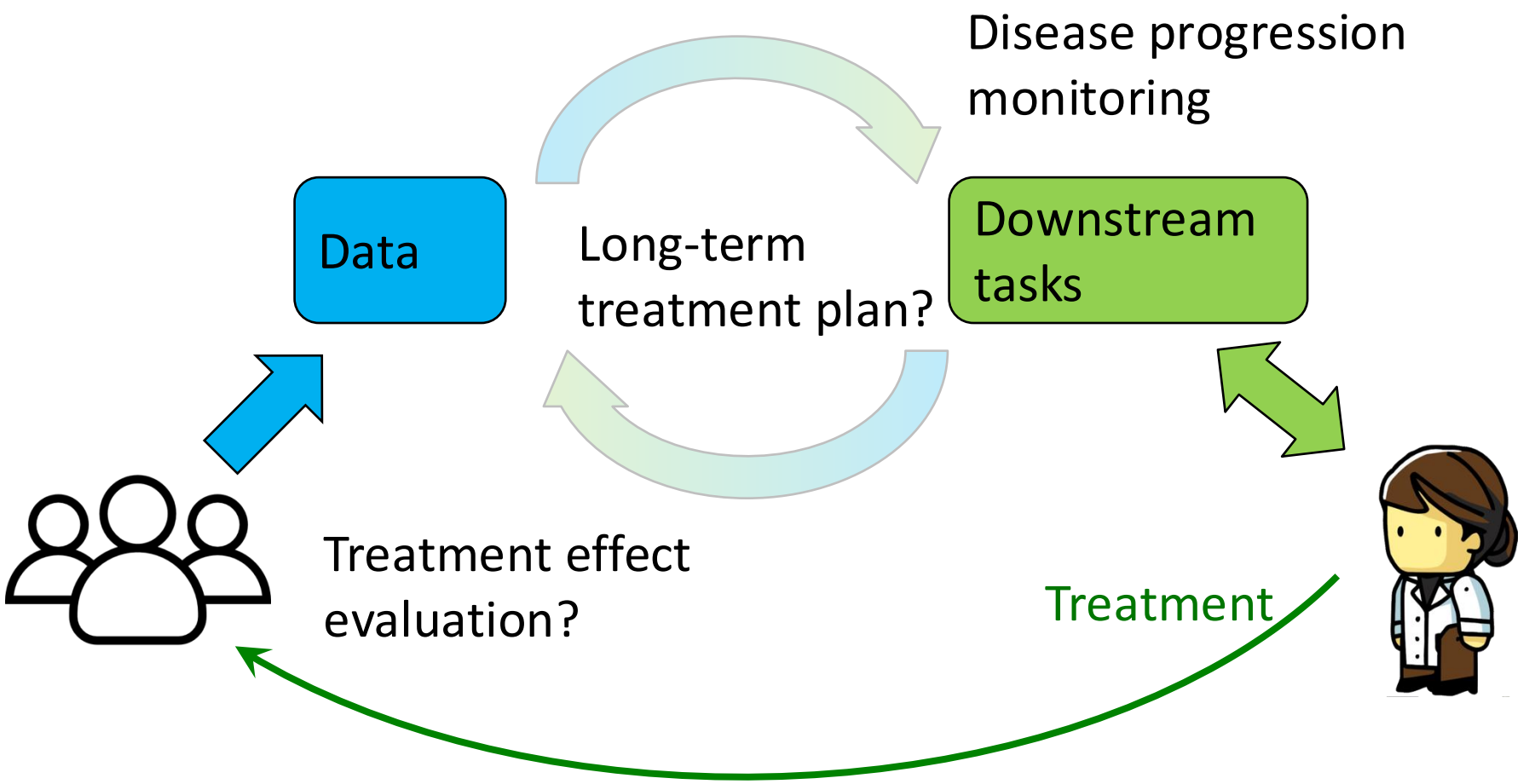
ML helps doctors/scientists evaluate causal relationships



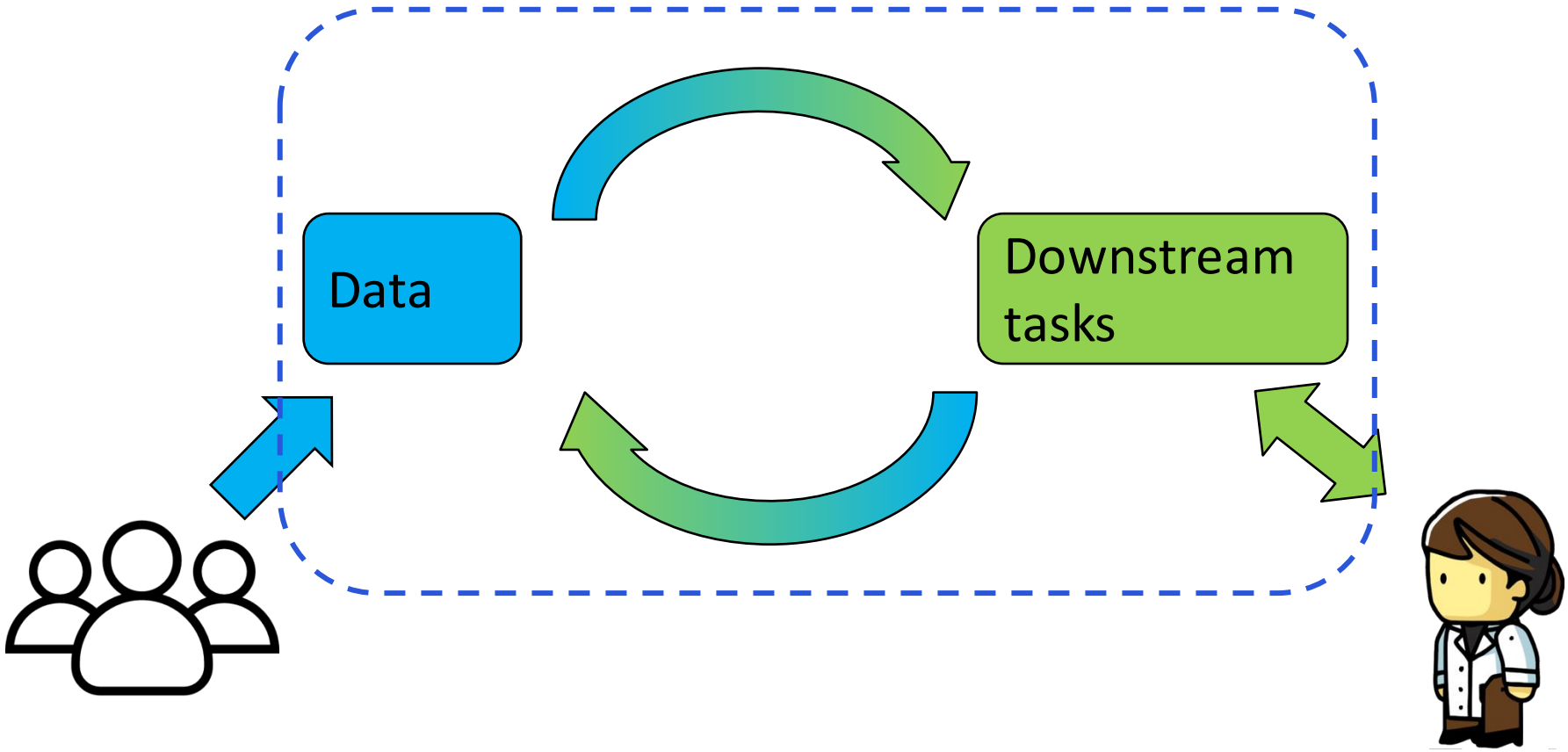
ML helps doctors/scientists make long-term plans



Could we close the loop?



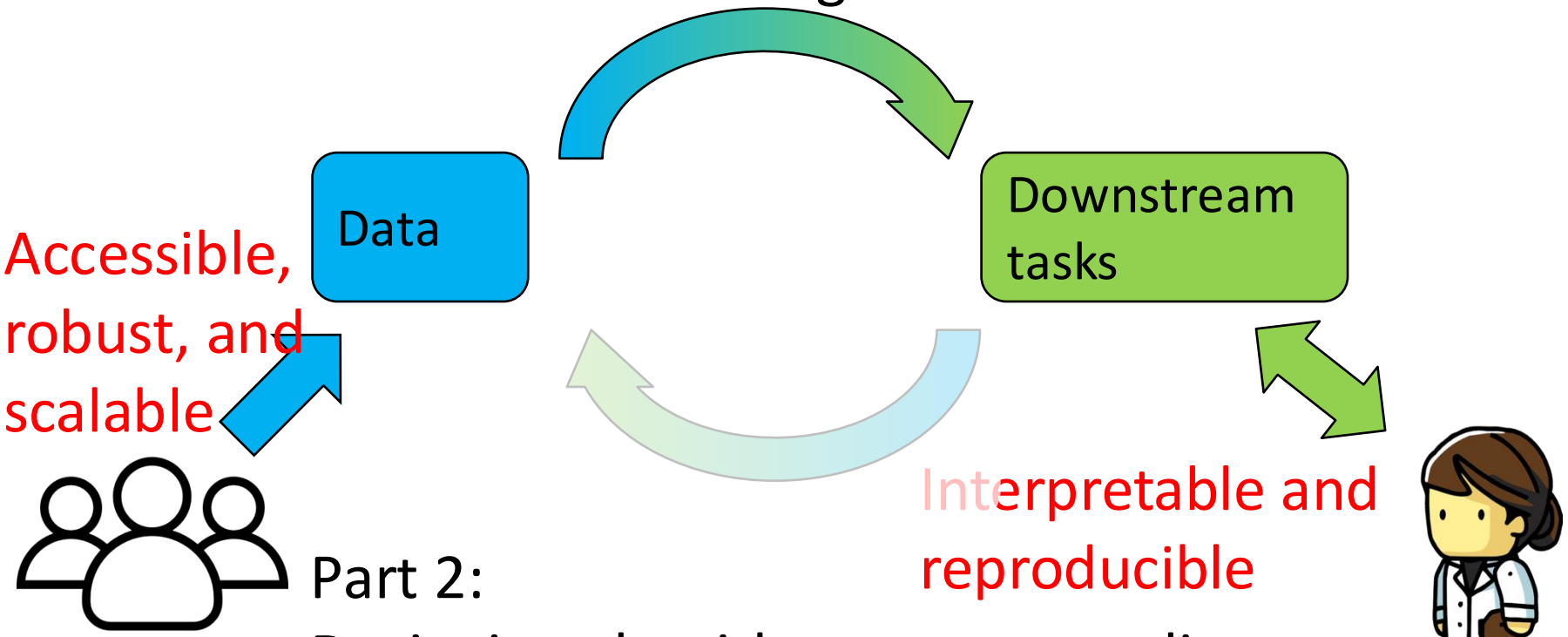
While many ML scientists focus on one of these domains,



My vision is to close the loop and bring the benefits back to the people

Part 1:

Tracking Neurodegenerative-Disease-Affected Eye-Movement Features Using Mobile Devices



Part 2:

Designing algorithms to personalize digital interventions for long-term goals

Acknowledgement



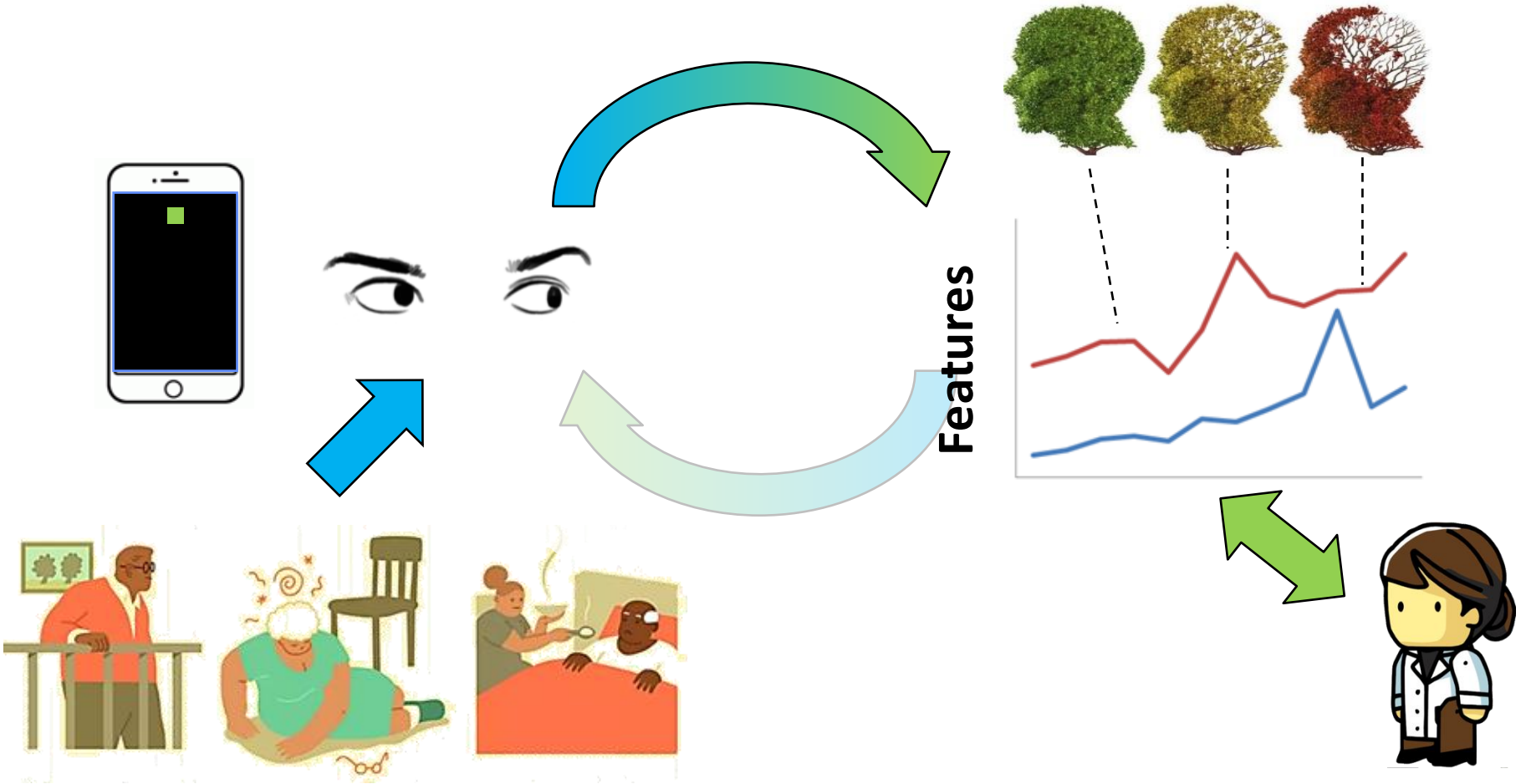








Part 1:
Tracking Neurodegenerative-Disease-Affected Eye-Movement Features Using Mobile Devices



Current measurements of neuro disease progression are qualitative and require trained specialists

What is the year? Season? Date?

Follow my finger…



These tests are sparse, as they are taken only in clinics.

The scores may be affected by individual physician measurement bias and confounding factors such as patient fatigue.

Anderson et al. *Nature Reviews Neurology*, 2013

An accessible and quantitative disease progression monitor may help with…

Detection of disease onset



Start taking measurements here

Evaluation of individual treatment effects



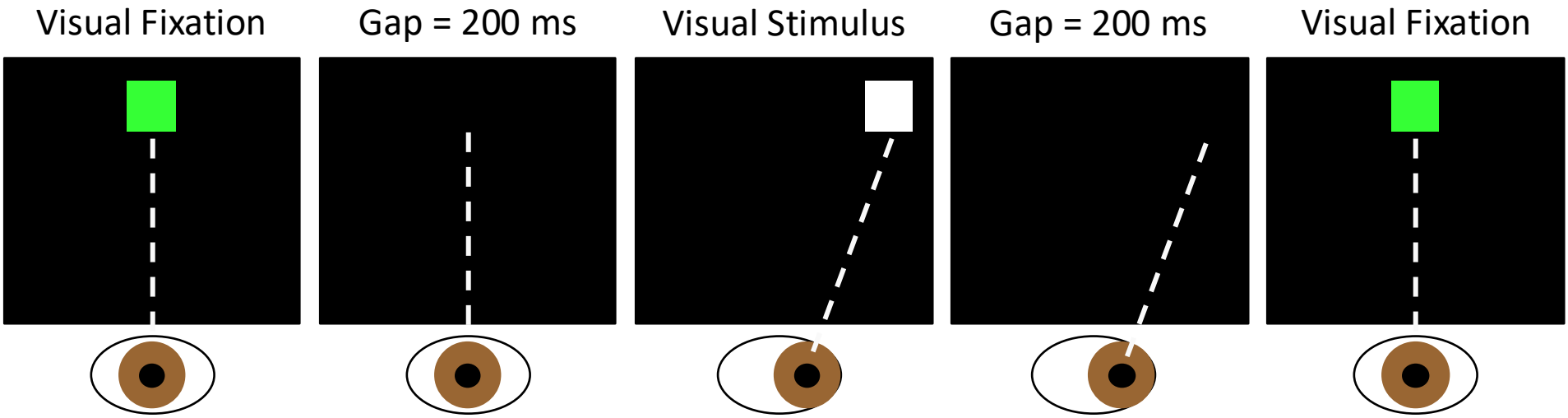
Moreover, the impact of normal variations and fatigue effects may be minimized through frequent measurements.

Can we develop a simple and quantitative metric to track neuro disease progression?

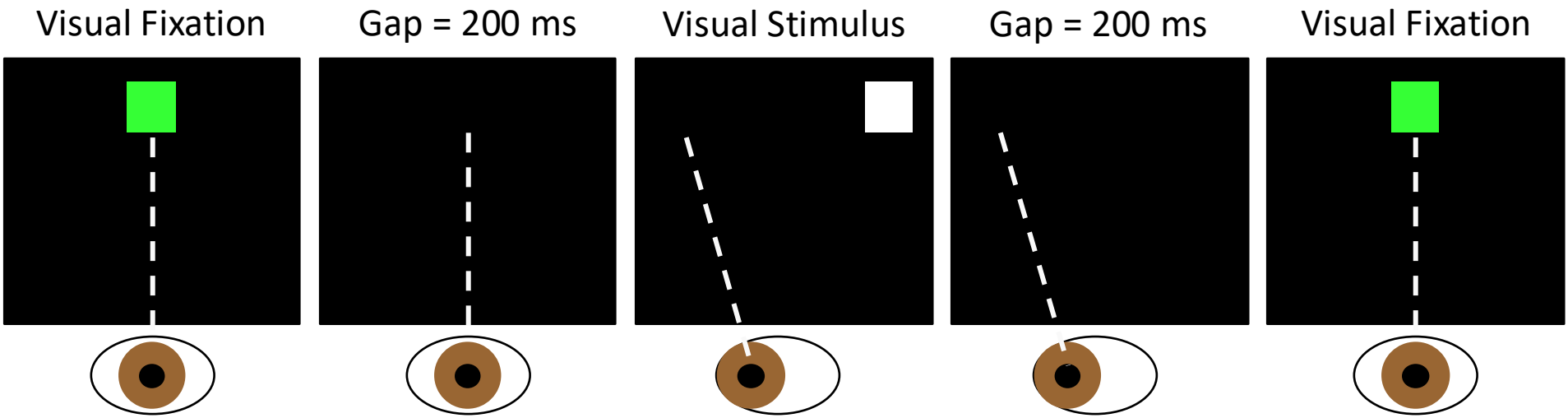


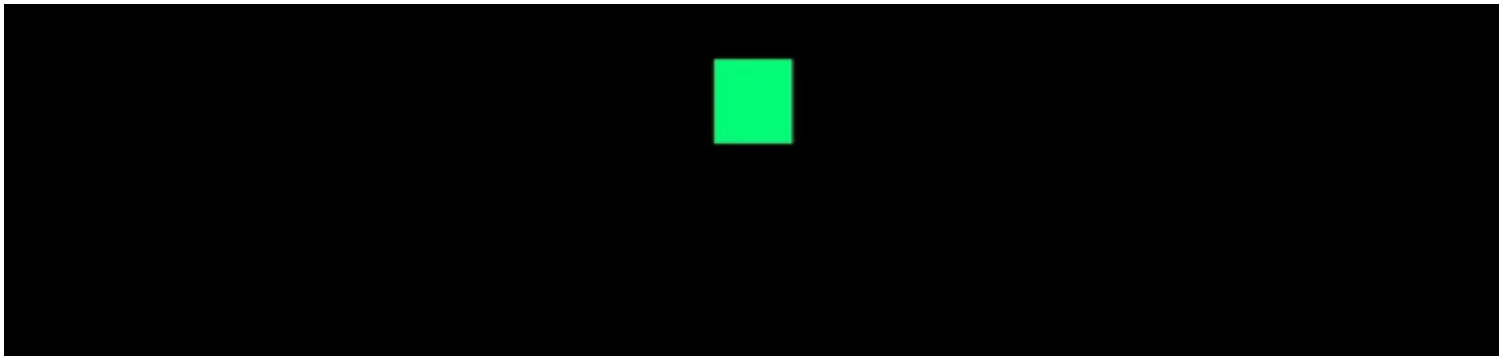
Pro-saccade task & anti-saccade task

Pro-saccade



Anti-saccade





Attention is required to not make a mistake.

Attention is affected by neurodegenerative diseases.

Brain regions involved in a saccadic eye movement are affected by neurodegenerative diseases.

Munoz et al. *Nature Reviews Neuroscience*, 2004
Anderson et al. *Nature Reviews Neurology*, 2013

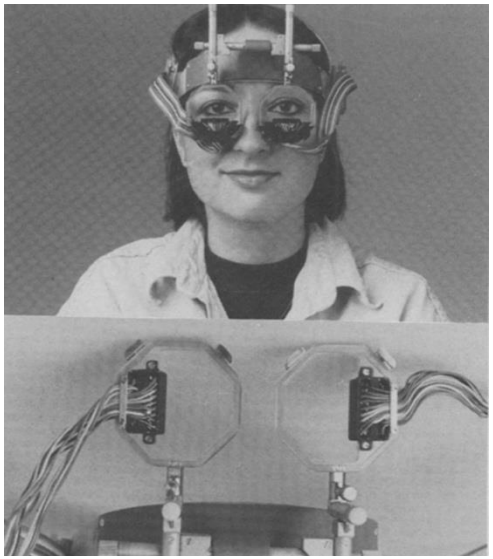
Measurements in clinical environments rely on specialized machines

High-speed camera



PHANTOM V25-11

IR illumination



[Reulen et al., Med. & Biol. Eng. & Comp, 1988.]

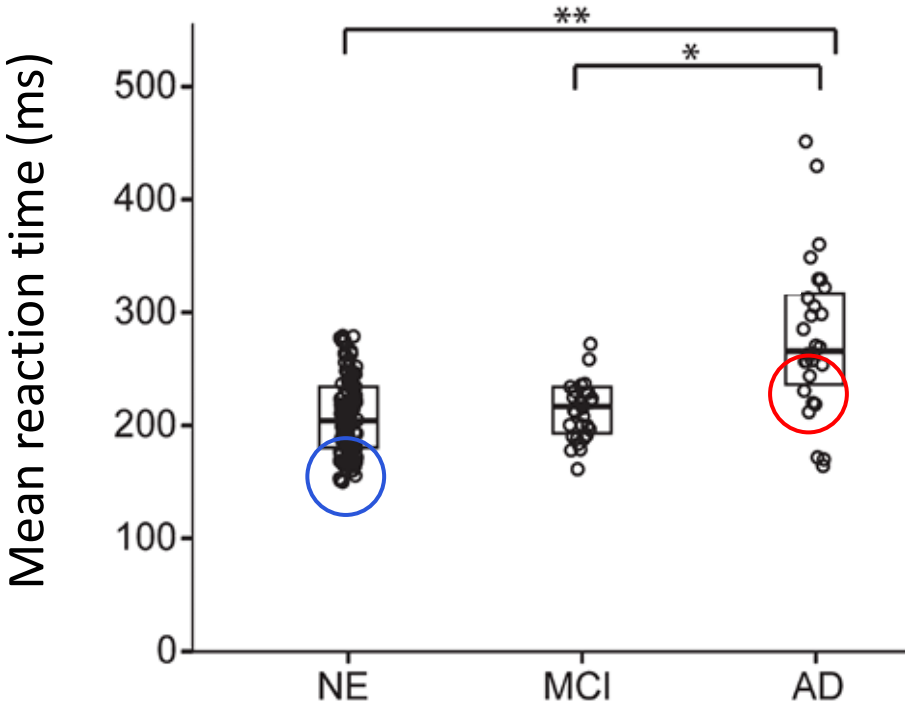
Substantial head support



SR EYELINK 1000 PLUS

These specialized machines are not widely accessible and cannot measure eye-movement features frequently.

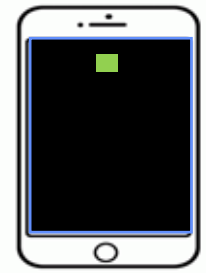
Clinical literature did not account for individual effects



NE: Normal Elderly
MCI: Mild Cognitive Impairment
AD: Alzheimer's Disease

Heuer et al. *Neuro* 81(14), 2013

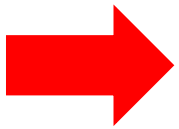
We need a way to track an individual over time…

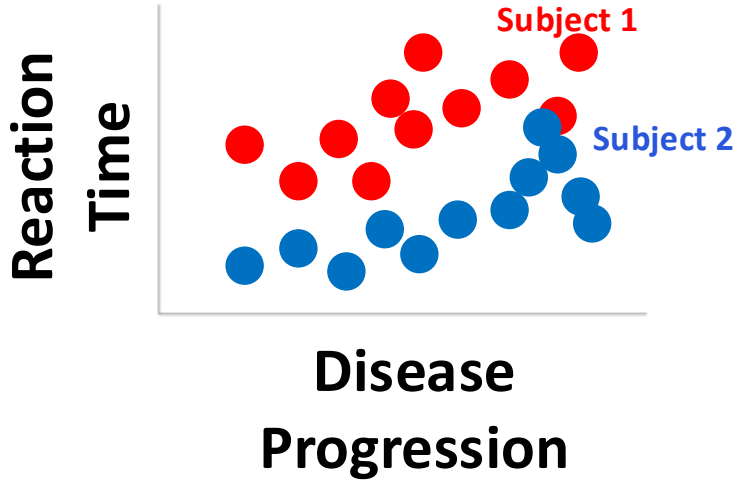


Display the task on the screen

Record the movement with the frontal camera

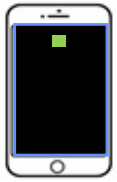






To enable individualized tracking of eye-movement features, we need …

A self-guided data-collection system



A measurement pipeline for large-scale mobile-device recordings

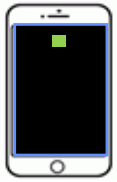


Characterization of healthy subjects



To enable individualized tracking of eye-movement features, we need …

A self-guided data-collection system



Previous work

Relied on experts and specialized equipment





This work

Designed an **instructive** and **easy-to-use** iOS app for self-recording

---

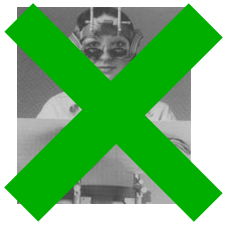
Key Contributions:

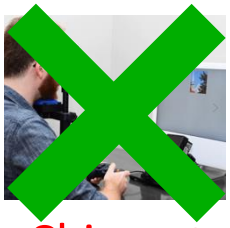
- A self-guided iPad app
- Large-scale data collection (100x recordings than in the literature)

Accessible, robust

To enable individualized tracking of eye-movement features, we need …


High-speed


IR


Chinrest

Previous work

This work

A measurement pipeline for large-scale mobile-device recordings



Required a constrained environment



Developed a **robust and automated** measurement pipeline

Key Contributions:

- Robust, automated, and accurate reaction time/error rate measurements from mobile devices

Robust, scalable, interpretable

To enable individualized tracking of eye-movement features, we need …

Are the changes in features related to disease progression?

⇒ We need to first understand how these features change over time in healthy subjects.

⇒ If we can characterize eye-movement features from healthy subjects, we may identify disease-related changes.

	Previous work	**This work**
Characterization of healthy subjects 	Lacked longitudinal data **?**	Characterized healthy subjects with an **individualized** and **interpretable** model

To enable individualized tracking of eye-movement features, we need …

Key Contributions:

- A Gaussian process (GP) model that characterizes healthy subjects with more than 25 days of recordings

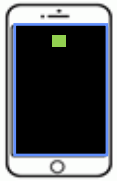
Significance:

Our work can help put into context how disease progression affects eye-movement features.

Robust, interpretable, reproducible

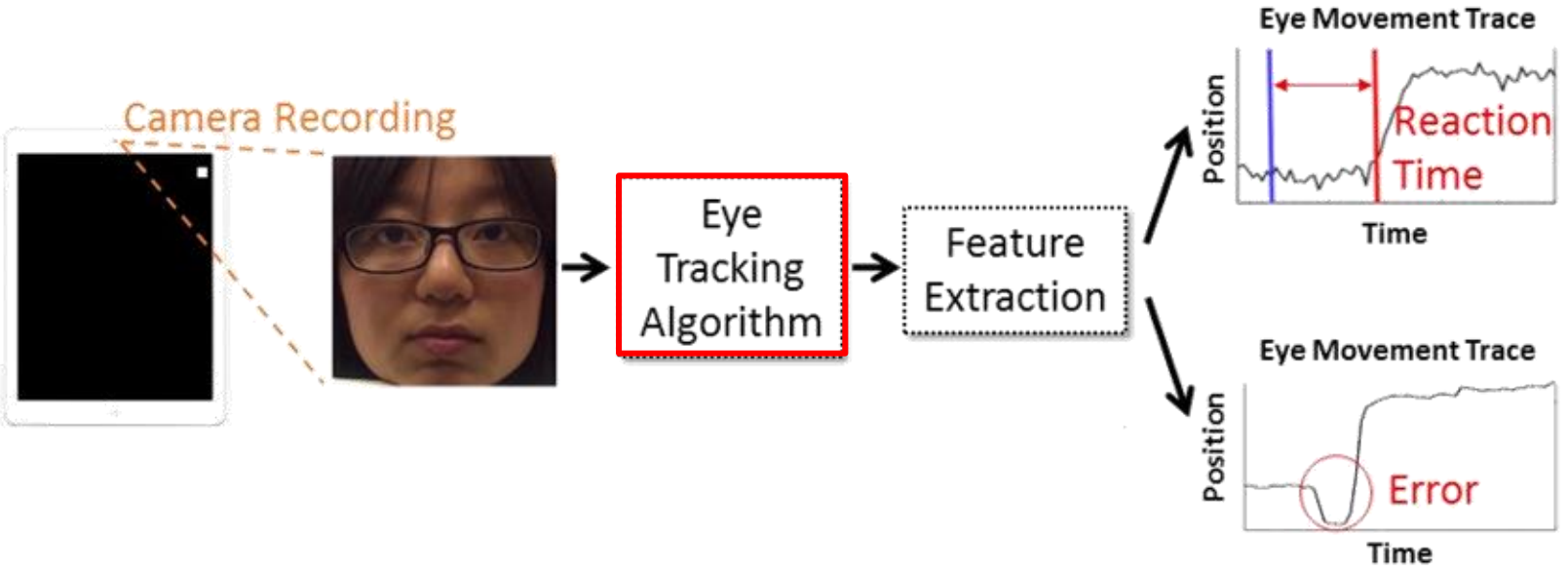
	Previous work	**This work**
Characterization of healthy subjects 	Lacked longitudinal data **?**	Characterized healthy subjects with an **individualized** and **interpretable** model

Roadmap of this part:

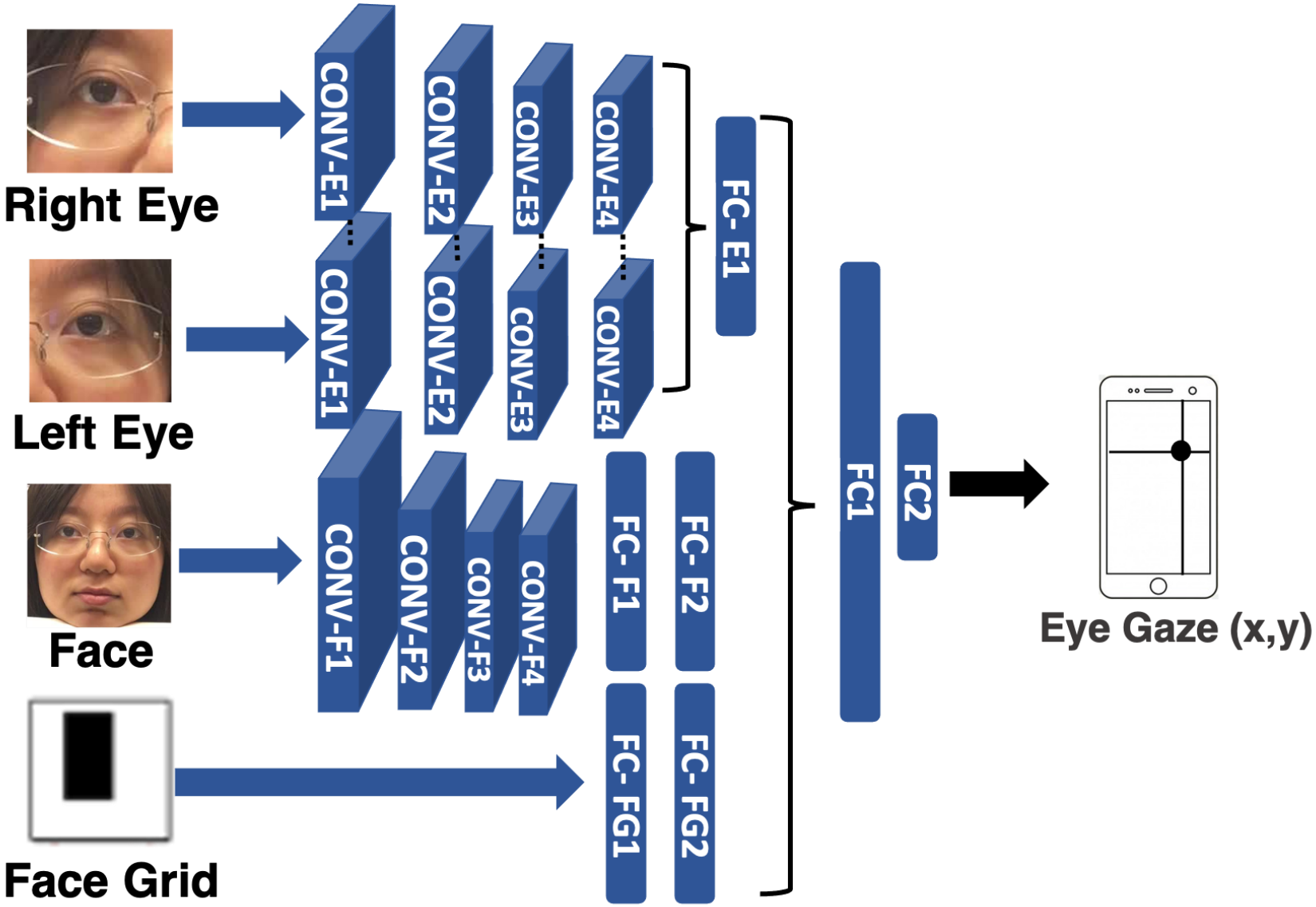
	Previous work	This work
A self-guided data-collection system 	Relied on experts and specialized equipment  	Designed an **instructive** and **easy-to-use** iOS app for self-recording
A measurement pipeline for large-scale mobile-device recordings 	Required a constrained environment 	Developed a **robust and automated** measurement pipeline **In collaboration with Gladynel Saavedra-Peña**
Characterization of healthy subjects 	Lacked longitudinal data **?**	Characterized healthy subjects with an **individualized** and **interpretable** model

The pipeline consists of an eye tracking algorithm and feature extraction algorithms

Goal: Find a **robust** eye-tracking algorithm that is able to show **clear saccade onsets**.



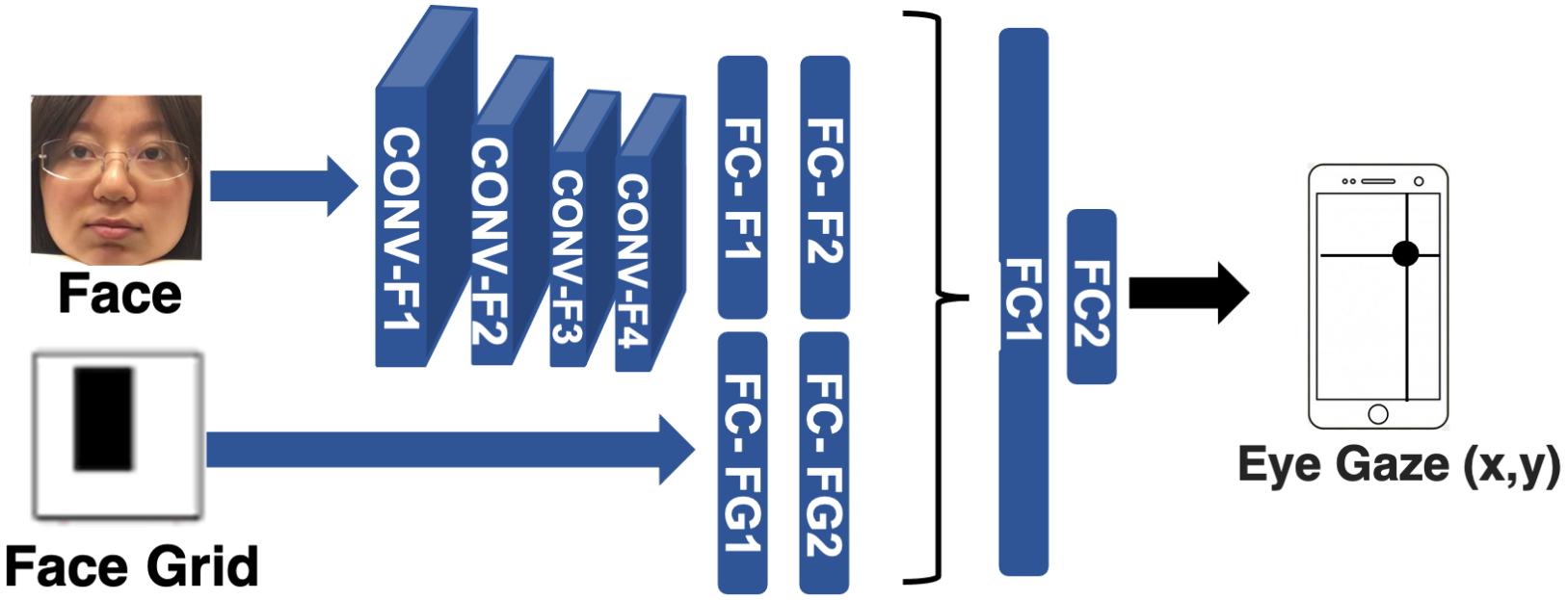
iTracker is a CNN-based gaze estimation algorithm



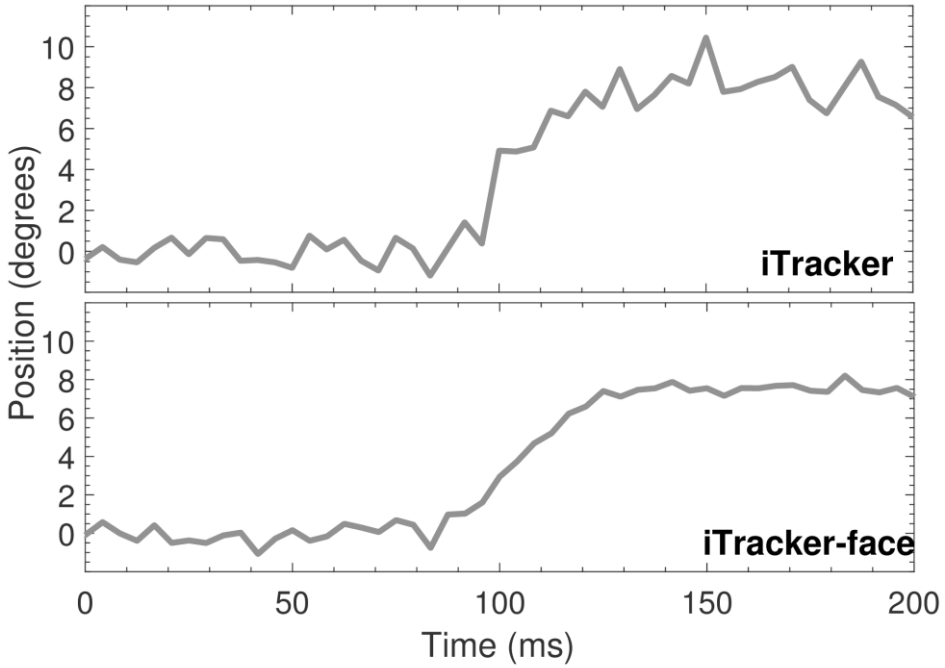
We propose iTracker-face for clear onset presentation

Our goal is to estimate **when** the gaze changes.

i.e. Gaze ~~location~~ ⇒ Clear saccade onset



iTracker-face produces traces with clearer saccade onsets compared to iTracker



We also tested and improved feature-based and phase-based algorithms.

How about robustness?

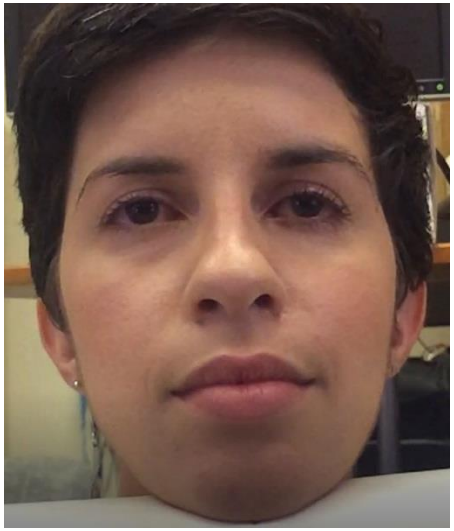
We tested

- under four lighting conditions (highest to lowest)
- on five subjects
- with and without glasses.

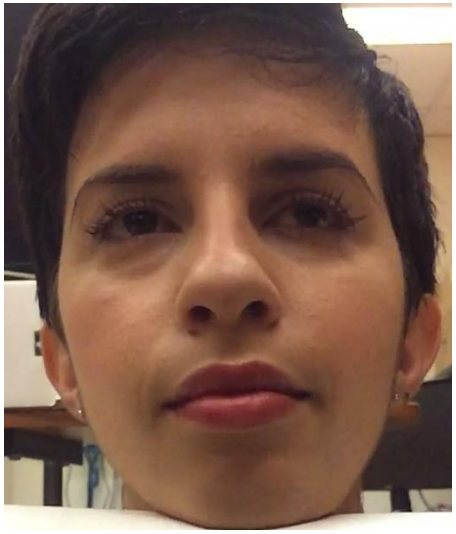
⇒ For this experiment, we acquired **9,600 saccades** in total.

Sample frames from four lighting conditions

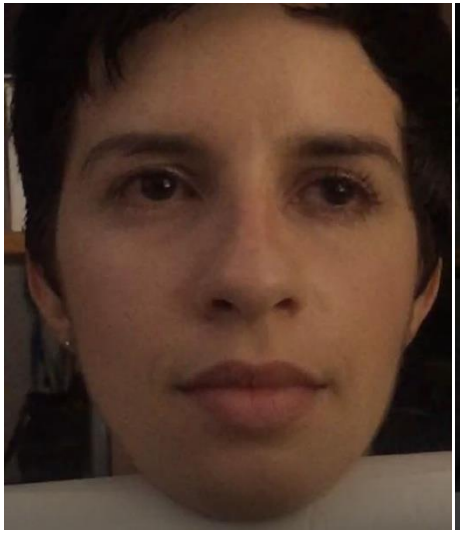
278 Lux



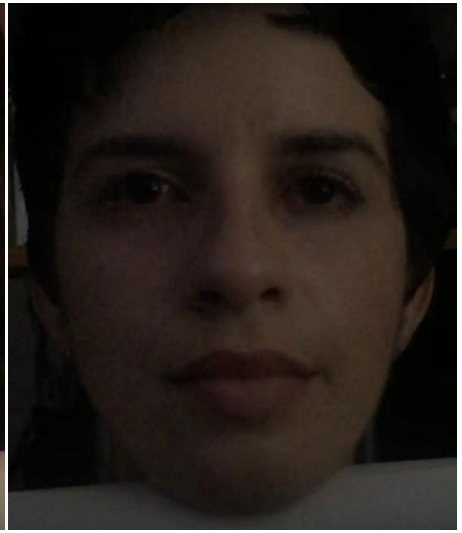
220 Lux



54 Lux



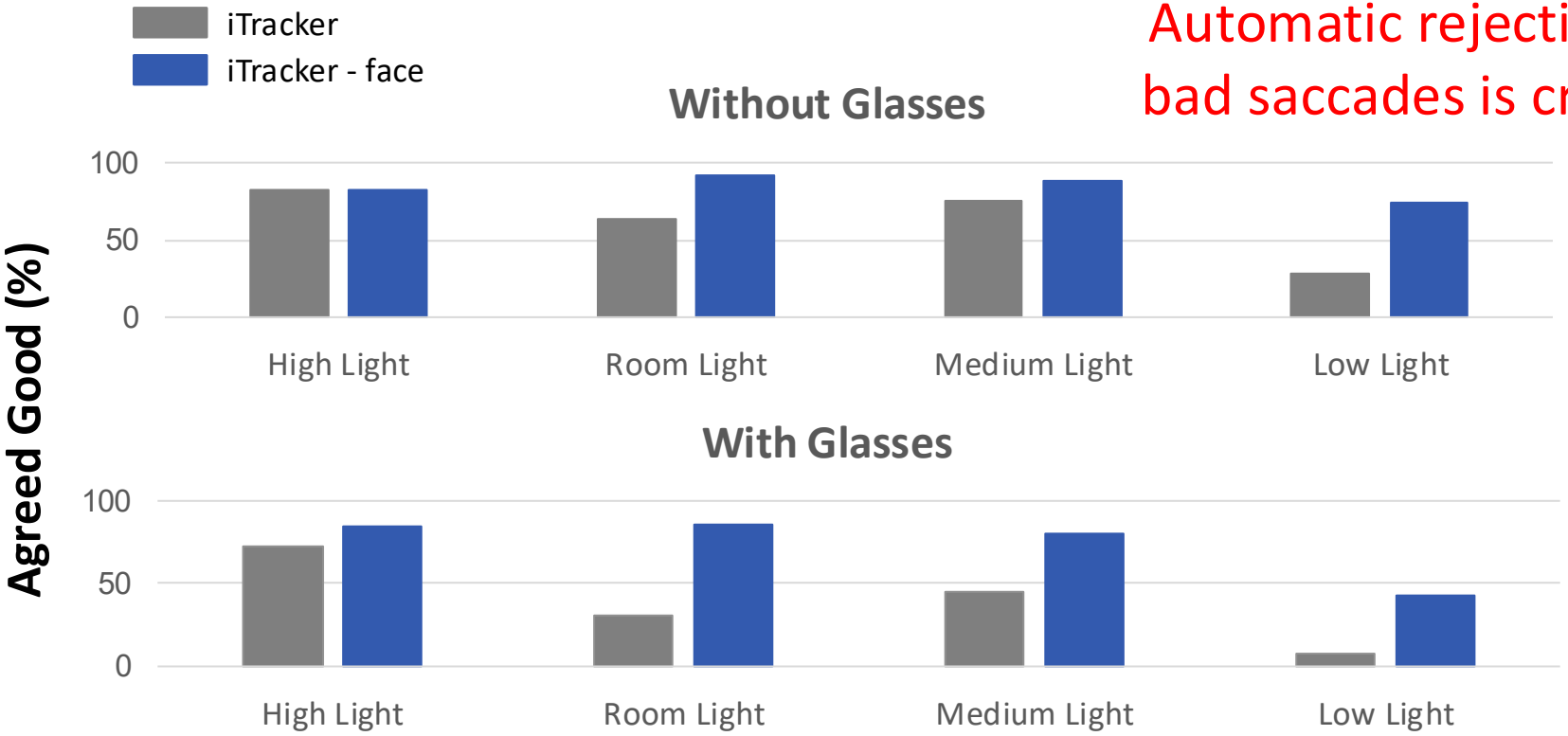
26 Lux



Robustness is compared by annotating good and bad saccades

- Two expert annotators visualized 9,600 traces and labeled them as good/bad saccades (clear/unclear onset).
- It took each annotator approximately 10-15 hours to do so.

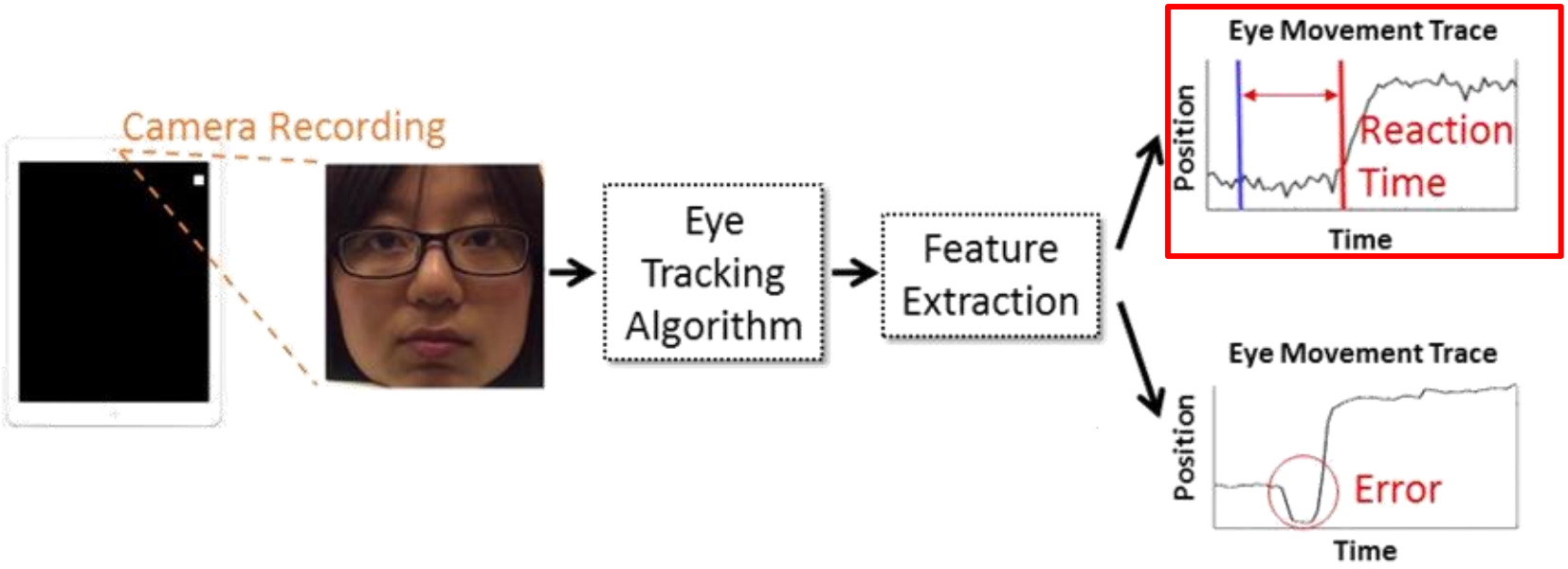
Automatic rejection of bad saccades is crucial.



iTracker-face is much more robust to lighting conditions and glasses.

We choose iTracker-face to be the eye tracking algorithm

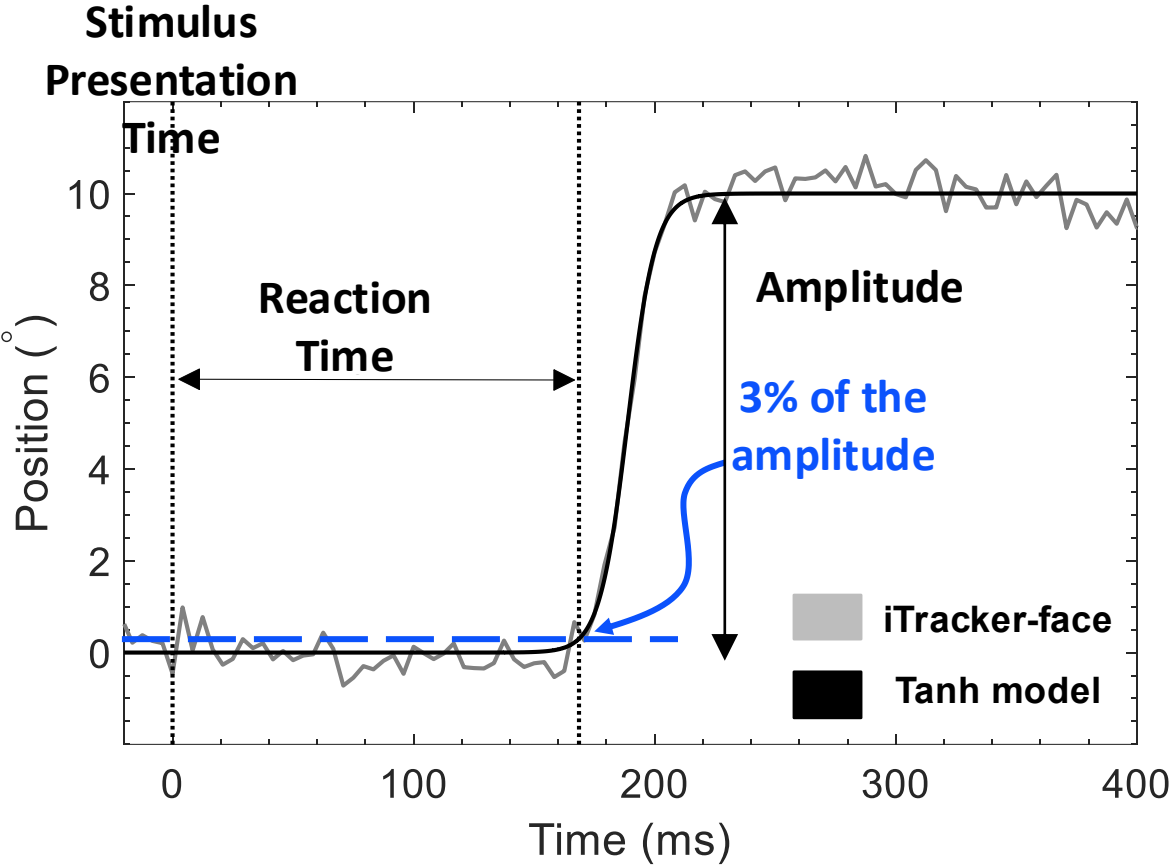
Goal: Find a **saccade onset detection** method and a way to **automatically detect bad saccades**.



In the literature, a velocity threshold is used to detect the onset.

=> May be affected by frame rate and cannot determine bad saccades.

Tanh fitting can measure saccade onset and detect “bad” saccades



- Horizontal saccades are modeled using **tanh.**
- The normalized root-mean-square error (**NRMSE**) between the model and the eye movement trace can be **used to reject bad saccades**.

Recall that we have 9,600 annotated traces from the robustness test.

Using an adequate NRMSE threshold, we can classify traces as “good” or “bad” with an **average true-positive rate of 0.87** and an **average false-positive rate of 0.18.**

How do mobile-device recordings compare to recordings made with a research-grade camera?

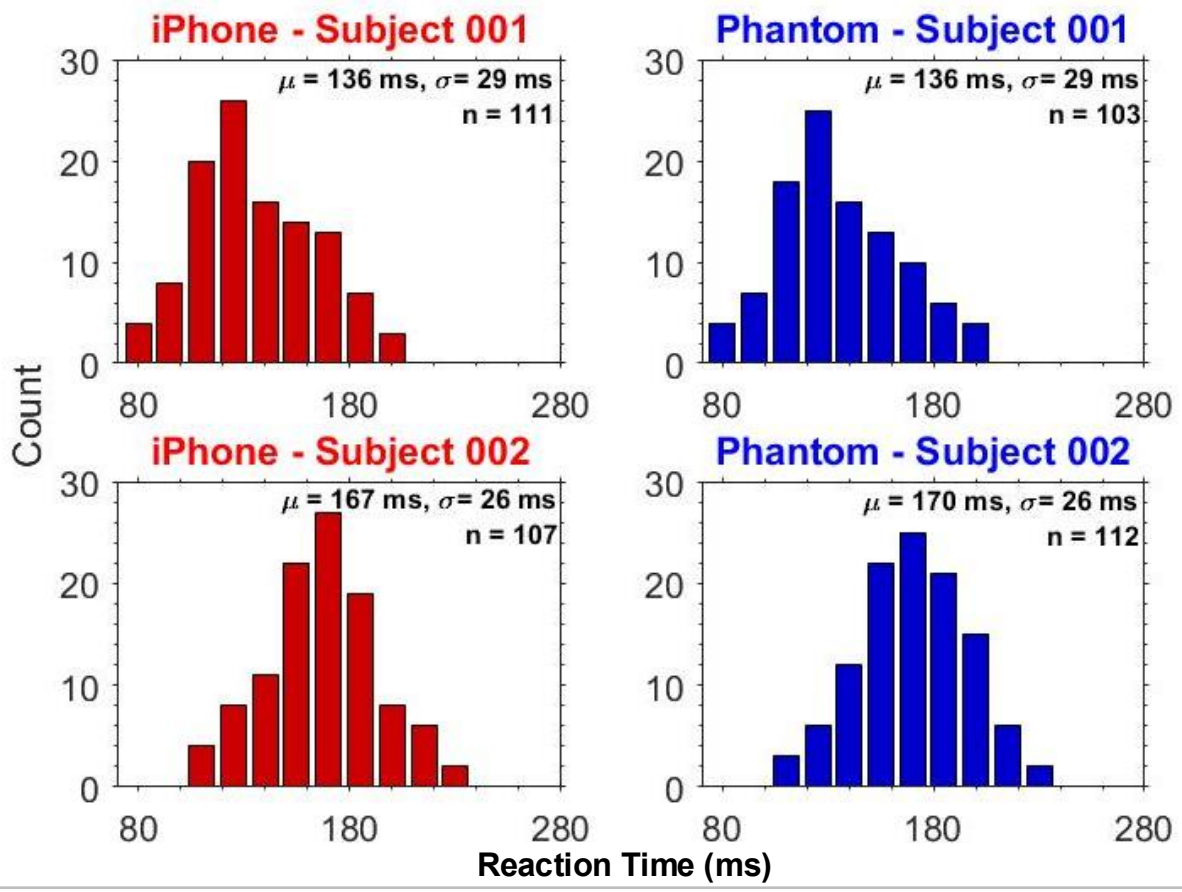
We achieved similar statistics compared with a research-grade camera

Consumer grade camera
iPhone 6, cost <$1k,
1.5 µm pixel size, rolling shutter



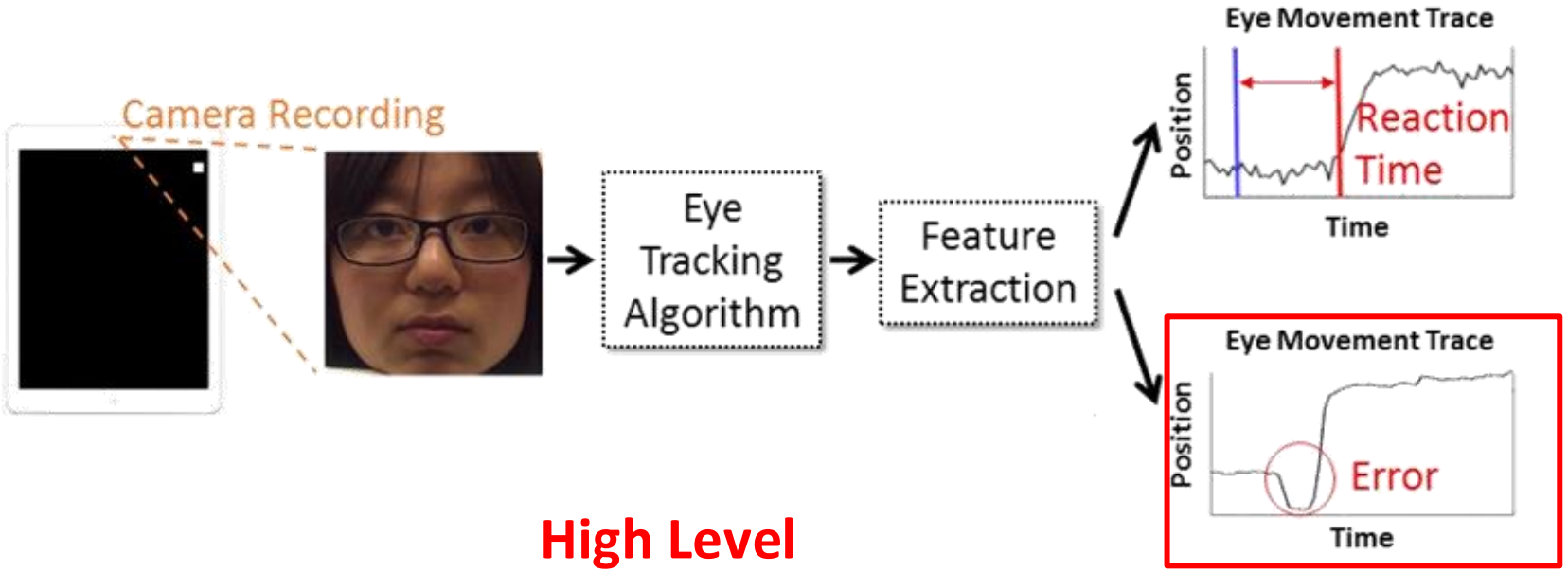
High-speed research-grade camera
Phantom v2511,cost $100k,
28 µm pixel size, global shutter





We enabled robust reaction time measurement

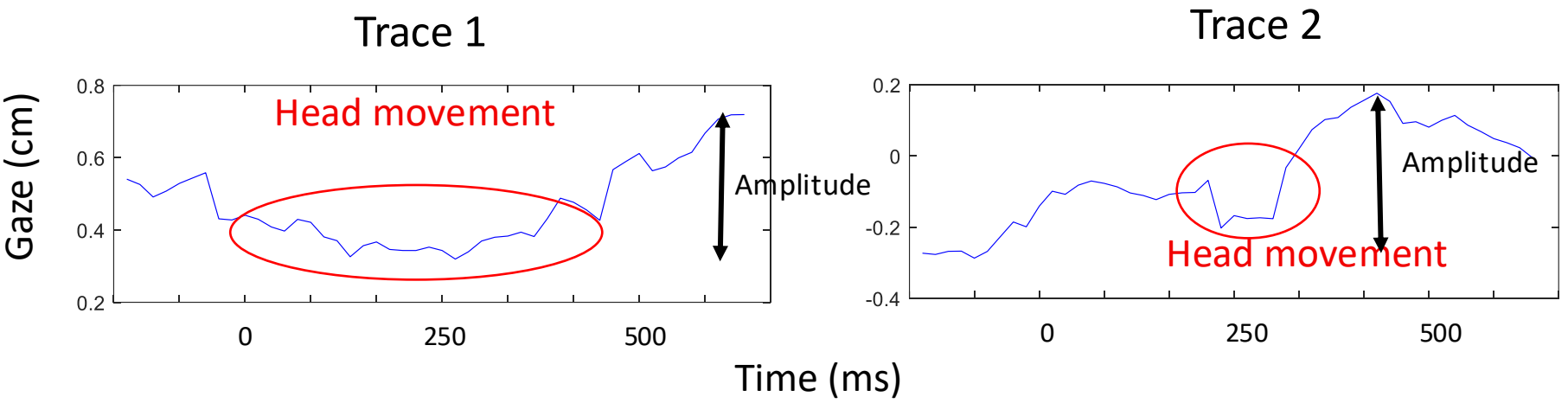
Goal: Find an **error detection** method and a **definition of error rate** accounting for noisy eye movements.



High Level

No automated and robust error detection algorithm in the literature.
=> We developed an error detection algorithm with 0.97 sensitivity and 0.97 specificity and defined an error rate that can account for low signal where head movement can look like an eye movement.

When the eye movement is too small, head movement may look like an error

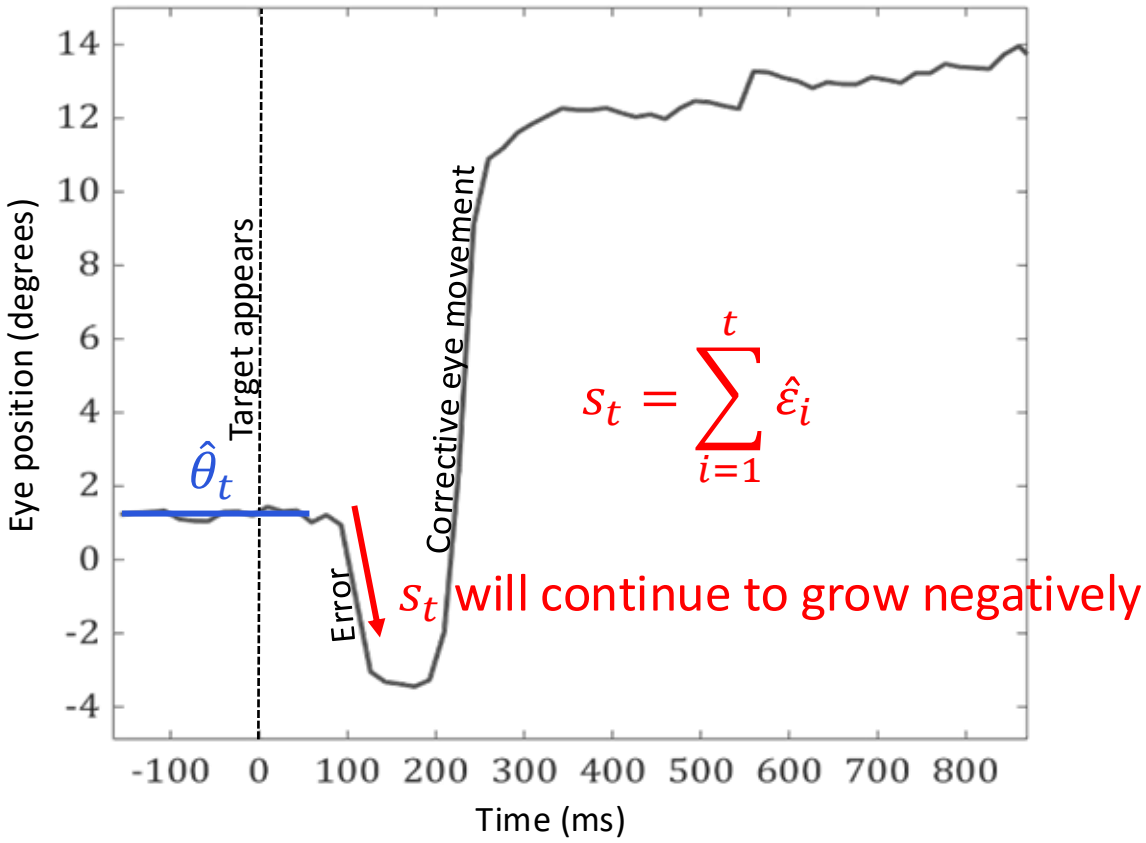


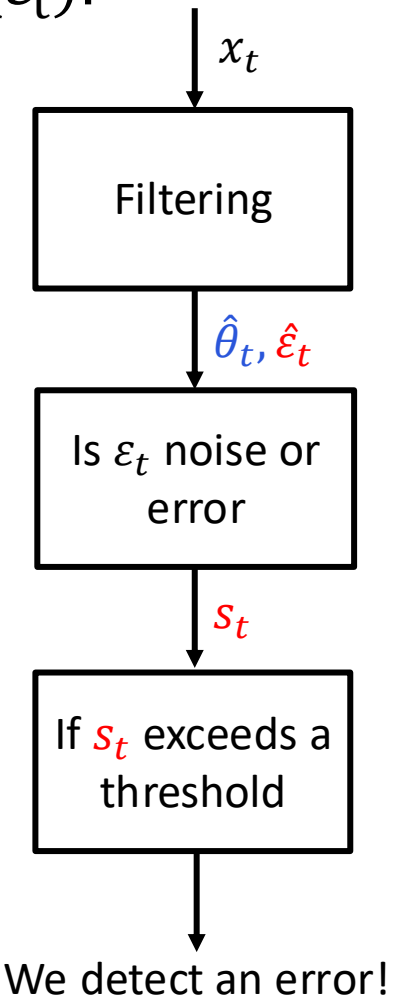
We therefore label traces as “LS (Low Signal)” if the amplitude of the signal is too small.

Change-detection algorithm is the base algorithm for error detection

Assume that the measurement x_t consisting of the fixation eye movement (θ_t) and noise/error (ε_t):

$x_t = \theta_t + \varepsilon_t$



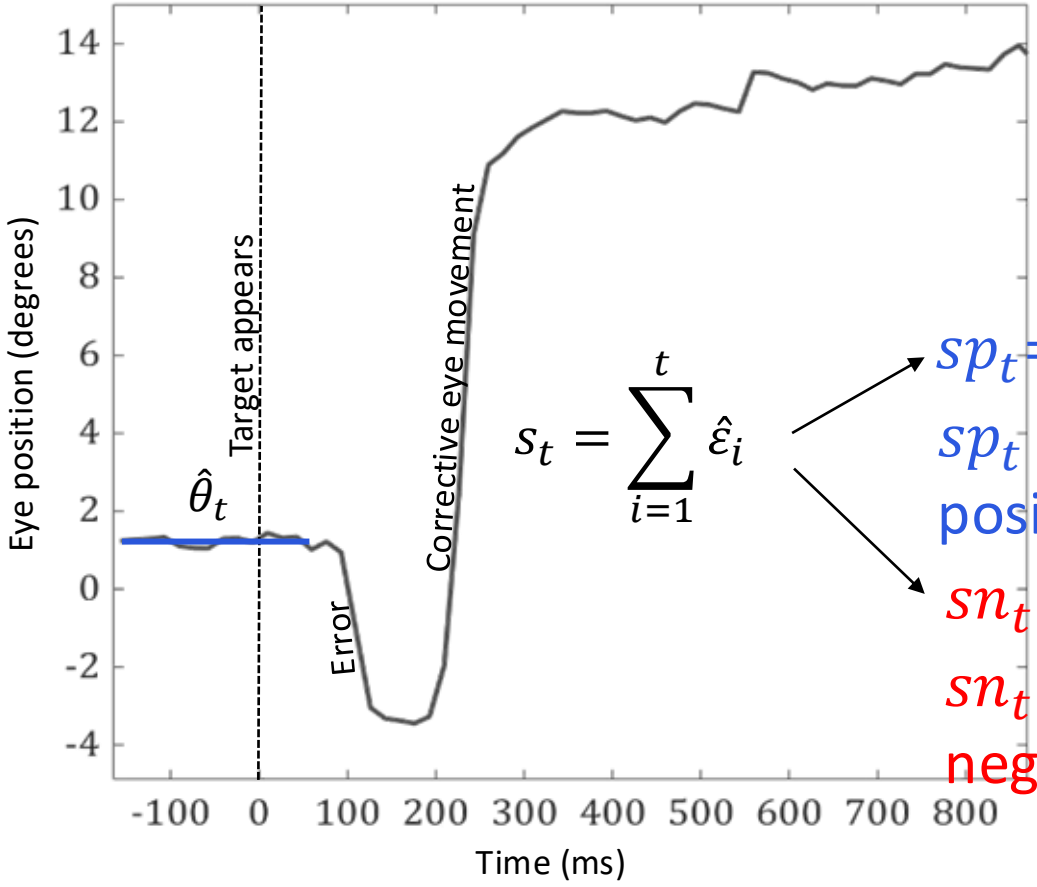


Gustaffson et al., Adaptive Filtering and Change Detection, *John Wiley & Sons*, 2000.

To detect an error, we design statistics for both positive changes and negative changes

Error definition: An initial eye movement towards the wrong direction.

=> An error happens after 0 ms **before any correct eye movement**.



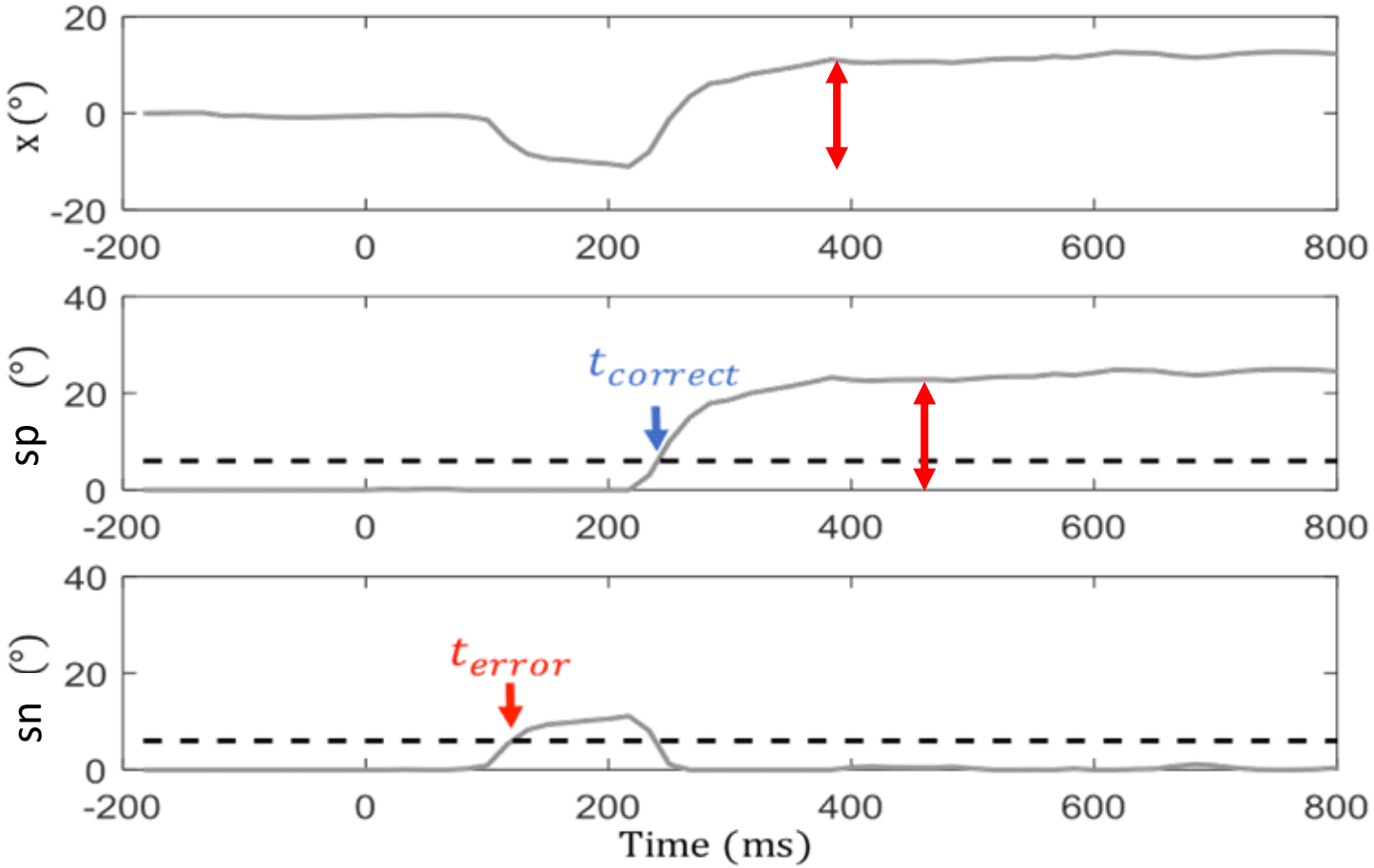
$sp_t = \max\{sp_{t-1} + \hat{\epsilon}_t, 0\}$

sp_t increases when there is a positive trend

$sn_t = \max\{sn_{t-1} - \hat{\epsilon}_t, 0\}$

sn_t increases when there is a negative trend

We modify a change-detection algorithm to measure whether a negative change occurs before a positive change

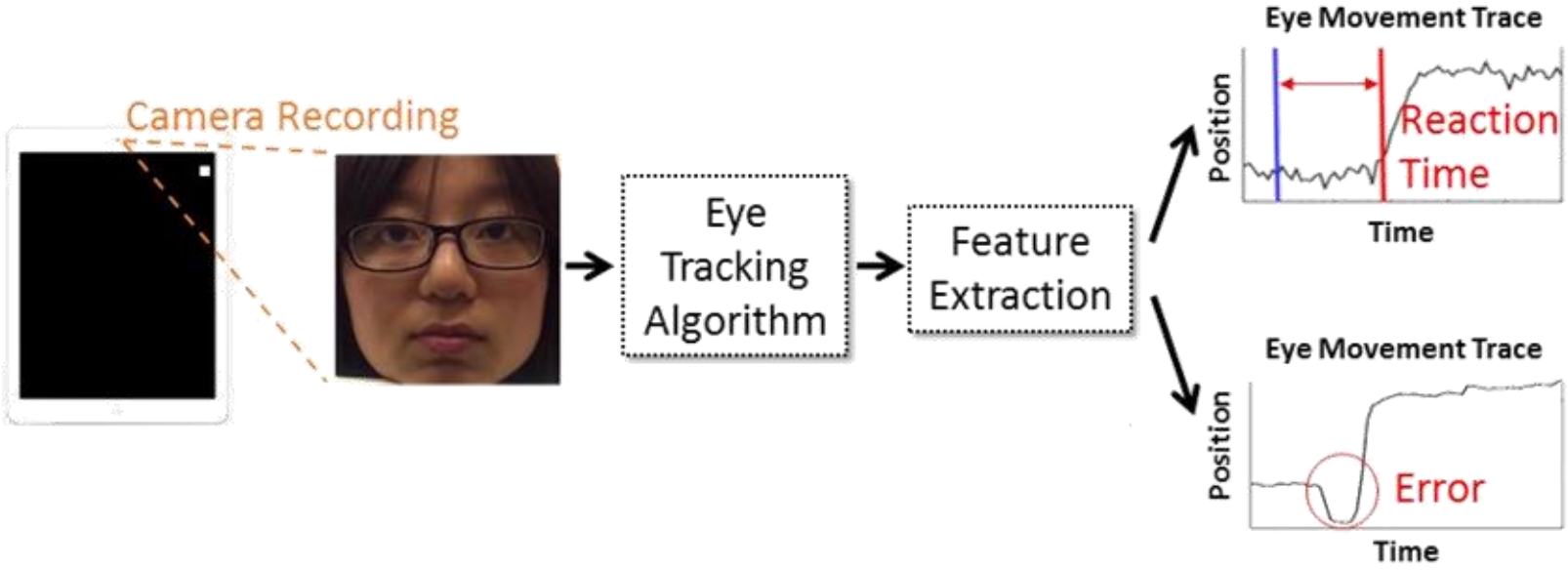


$\max\{sp, sn\}$ estimates the amplitude of the eye movement.

Consider $T \cdot \max\{sp, sn\}$ as the threshold.

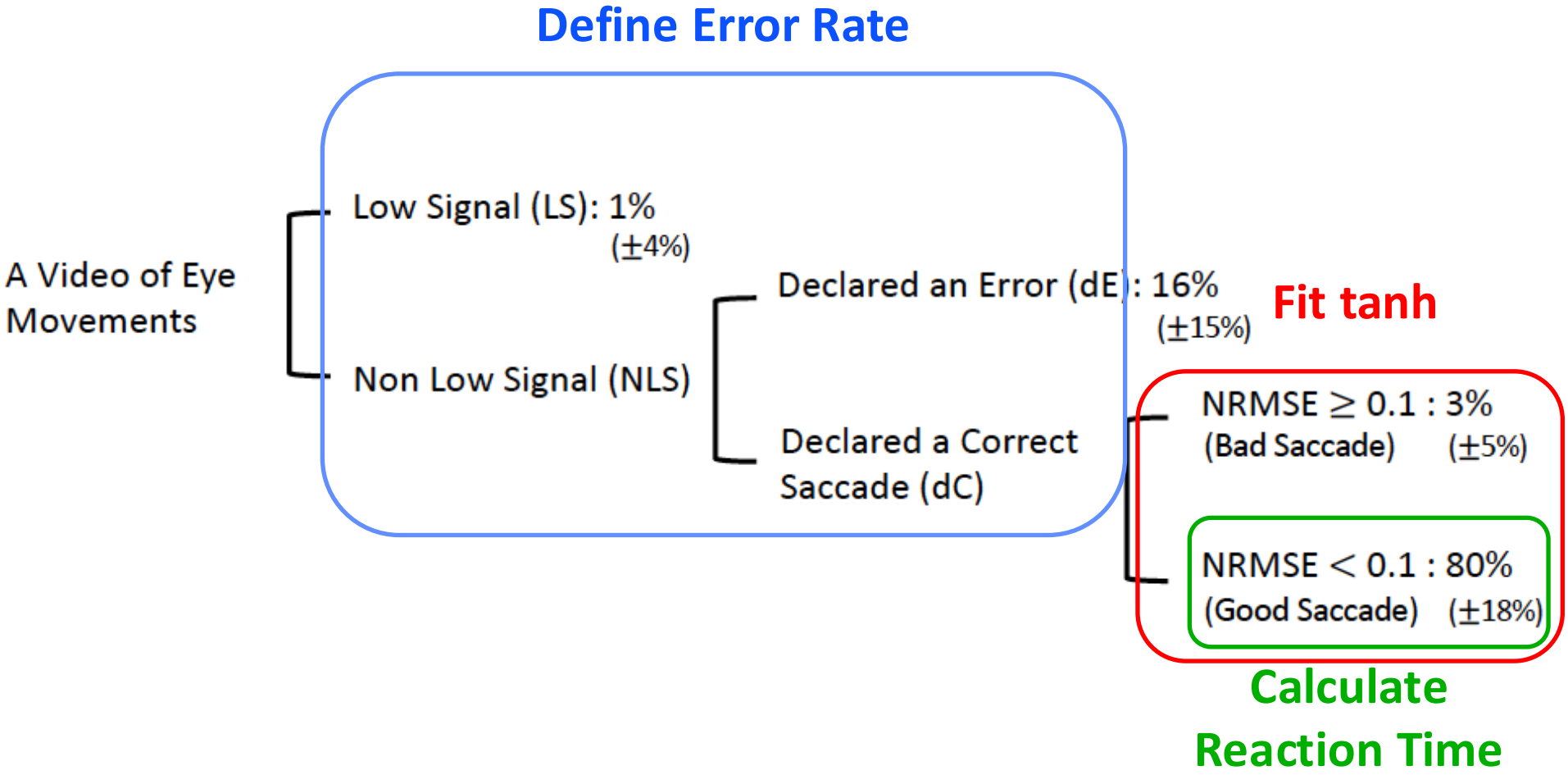
We achieve 0.97 sensitivity and 0.97 specificity.

We have an error detection algorithm



Goal: Find a **definition of error rate** accounting for “Low Signals”.

We accomplish an automated pipeline for feature extractions

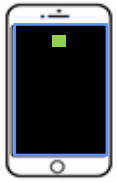
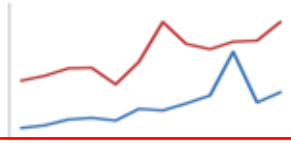


Summary

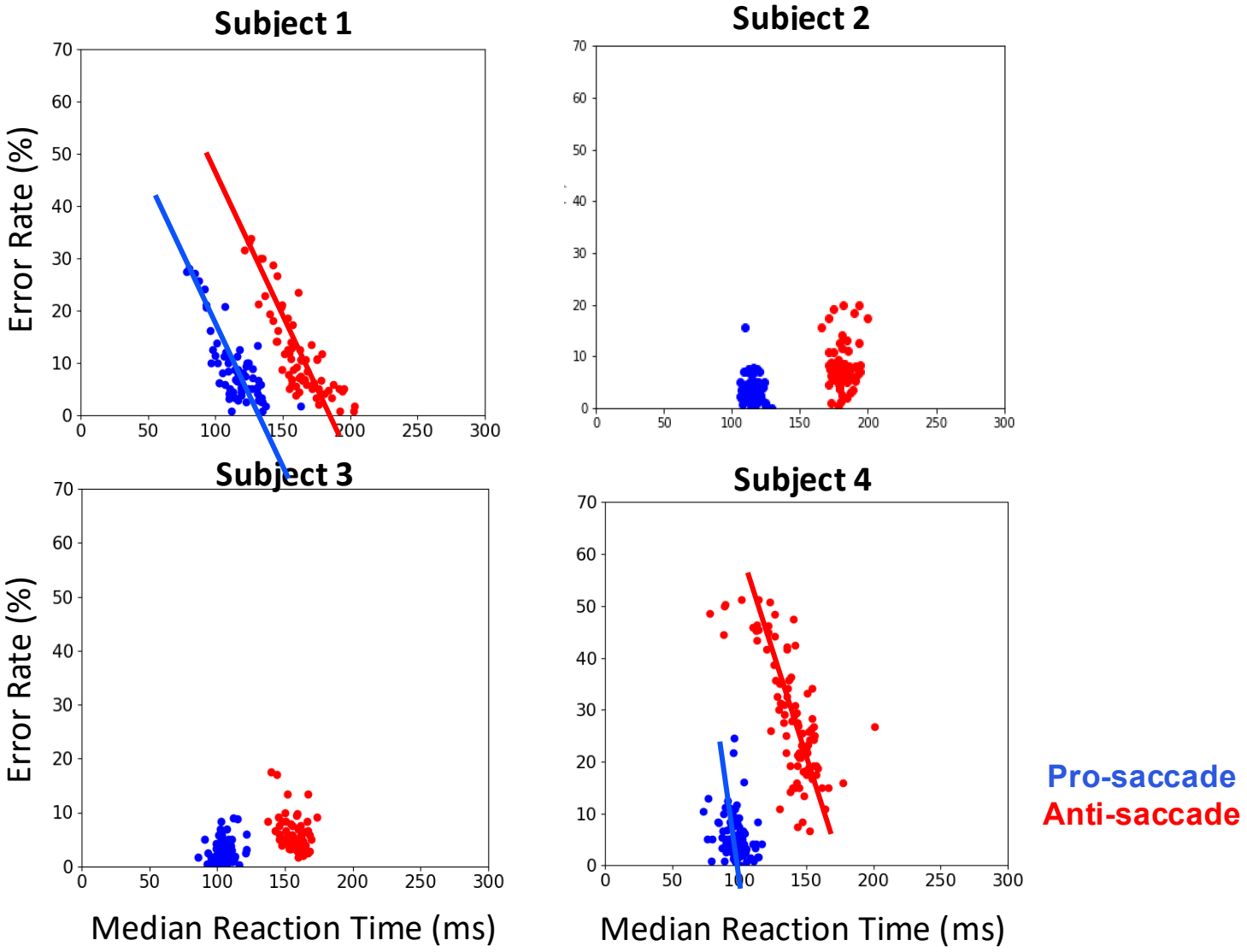
- A **robust** and **automated** pipeline is the key to measure reaction time and error rate outside of a clinical environment.
- With the pipeline, large-scale measurements of saccade features are possible.

Next goal is to characterize how eye-movement features change over time in healthy subjects.

Roadmap of part 1:

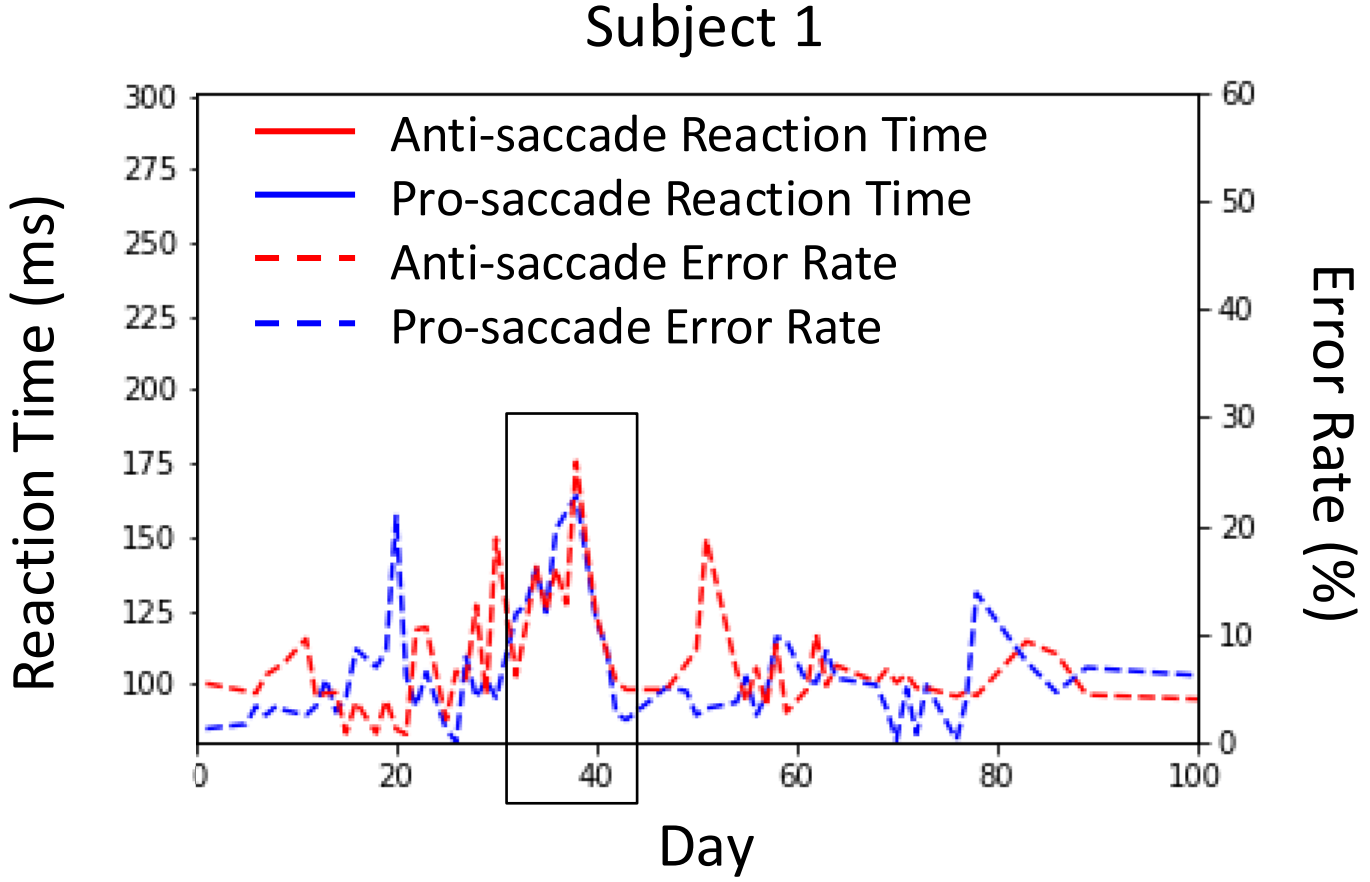
	Previous work	This work
A self-guided data-collection system 	Relied on experts and specialized equipment  	Designed an **instructive** and **easy-to-use** iOS app for self-recording
A measurement pipeline for large-scale mobile-device recordings 	Required a constrained environment and manual annotations  	Developed a **robust and automated** measurement pipeline
Characterization of healthy subjects 	Lacked longitudinal data **?**	Characterized healthy subjects with an **individualized** and **interpretable** model

There are significant intra- and inter-subject variability



We need an individualized model.

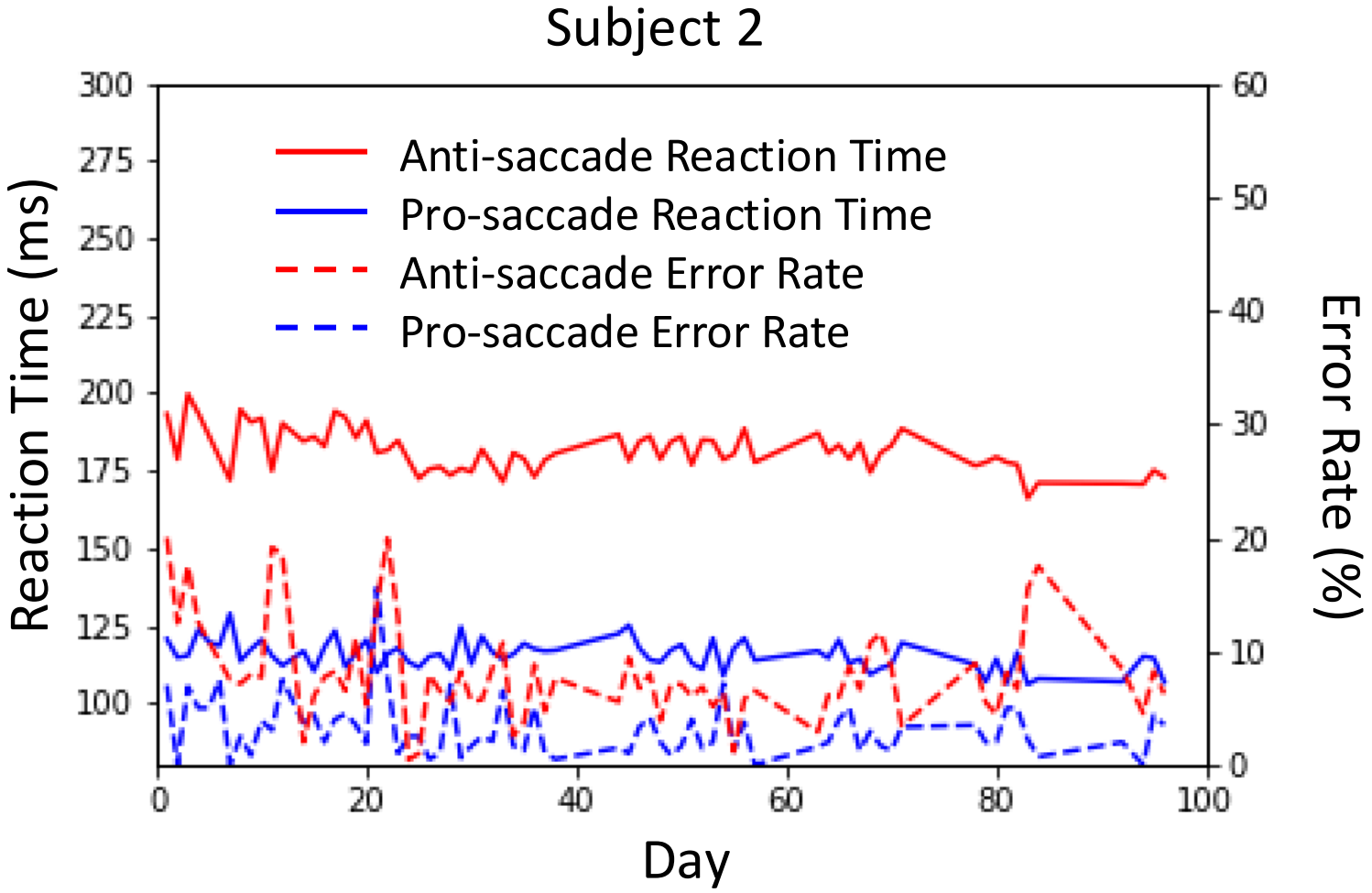
Some subjects may have a noticeable shared process across all features



The features also change smoothly over time.

The model should capture both correlation over features and correlation over time.

Some subjects may not have a noticeable shared process



Correlation over features should be individualized.

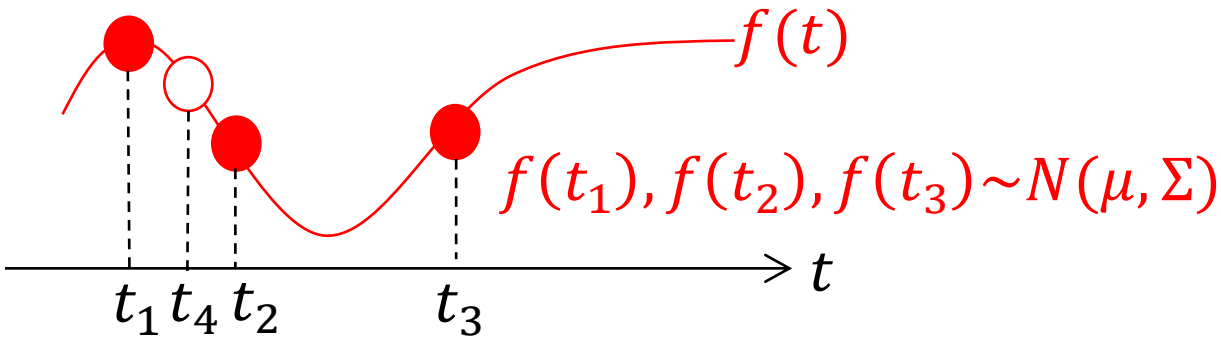
We want to develop a model that

- Unsupervisedly tracks pro/anti-saccade reaction time/error rate
- Captures correlation over time and individual correlation over features
- Can be extended to a disease progression model

We consider a Gaussian process (GP) model.

We consider a Gaussian Process (GP) model because

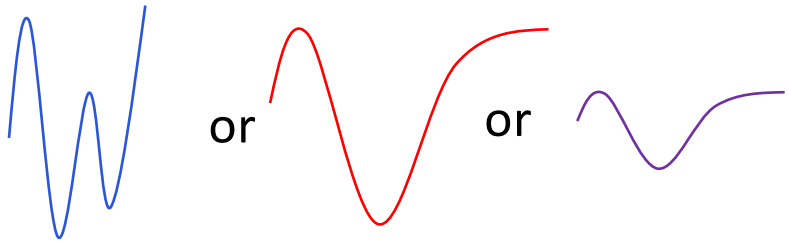
- It is computable.



- It is interpretable.

  $f(t) \sim GP(0, K(t, t'))$

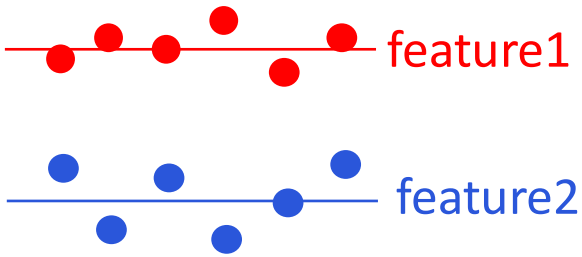
  Hyperparameters in $K(t, t')$ can determine the shape of $f(t)$.



Models to be compared

Baseline

$$y_{pi}(t) \sim N(0, \frac{1}{\beta_i})$$

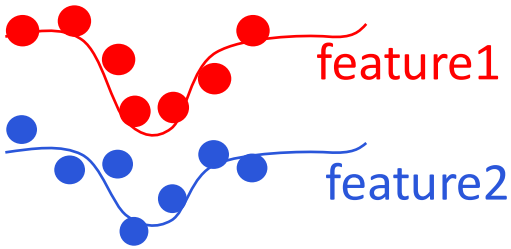


p: The p-th subject
i: The i-th feature

Multitask-GP

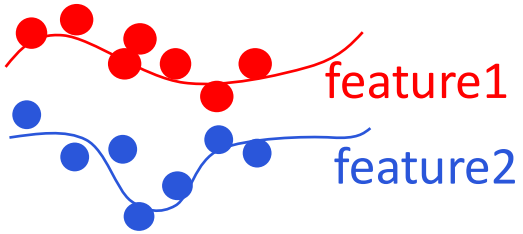
$$y_{pi}(t) \sim N\left(f_{pi}(t), \frac{1}{\beta_i}\right), \quad f_{pi}(t) = w_{pi} g_p(t)$$

$$g_p(t) \sim GP(0, K(t, t'))$$



Feature-specific GP

$$y_{pi}(t) \sim N\left(f_{pi}(t), \frac{1}{\beta_i}\right), \quad f_{pi}(t) \sim GP(0, K_i(t, t'))$$



Mixed GP

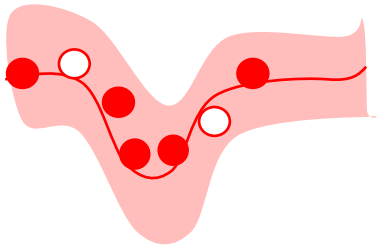
$$y_{pi}(t) \sim N\left(f_{pi}(t), \frac{1}{\beta_i}\right), \quad f_{pi}(t) = w_{pi} g_p(t) + h_{pi}(t)$$

We can evaluate how well the model fits the data by…

- Normalized L2

$$\frac{\|y-\hat{y}\|}{\|y\|}=\frac{\|y-\hat{\mu}\|}{\|y\|}$$

Prediction: $\hat{y}\sim N(\hat{\mu},\hat{\Sigma})$

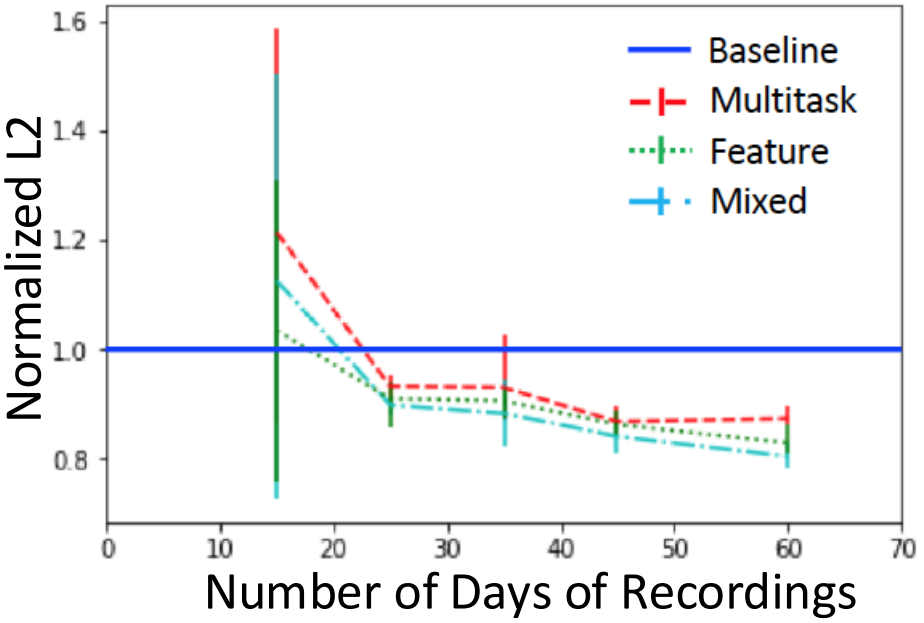


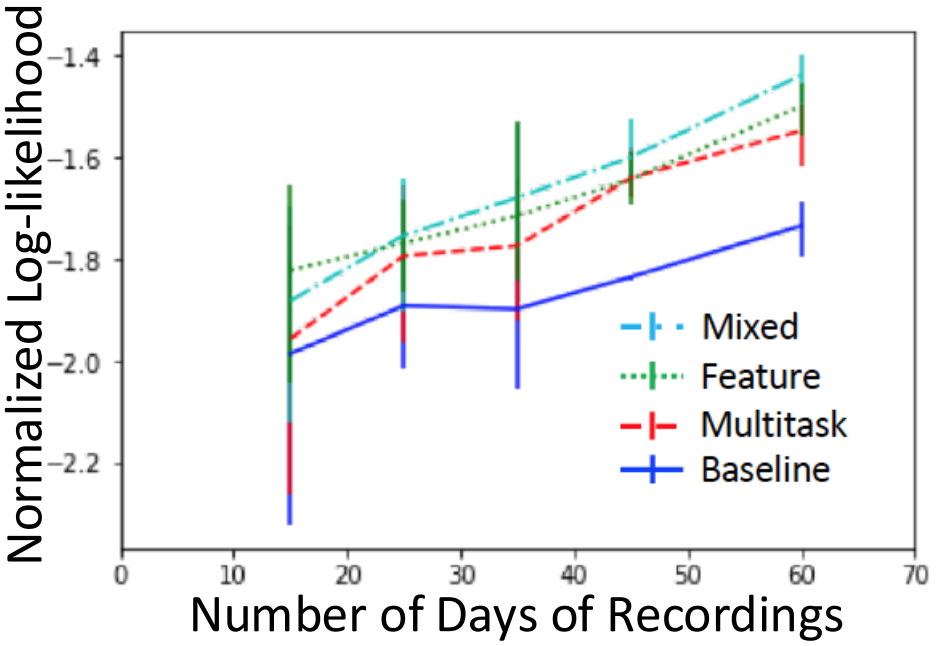
The baseline model will be setting $\hat{\mu}$ = 0. => Normalized L2 = 1.

- Normalized Log Likelihood

$$\frac{1}{N}\log P(y)=-\frac{1}{2}\log 2\pi-\frac{1}{2N}\log\left|\hat{\Sigma}\right|-\frac{1}{2N}(y-\hat{\mu})^{T}\hat{\Sigma}^{-1}(y-\hat{\mu})$$

Correlation over time and over features should be modeled





1. GP models perform better when there are more than 25 days of recordings.
2. Performance: Mixed model > Feature-specific model > Multi-task model

The mixed model can characterize healthy subjects with more than 25 days of recordings.

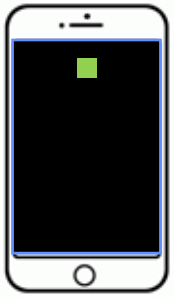
Summary

- An **individualized** and **interpretable** model is the key to characterize longitudinal eye-movement features.
- With the model, we can put into context how disease progression may affect eye-movement features.

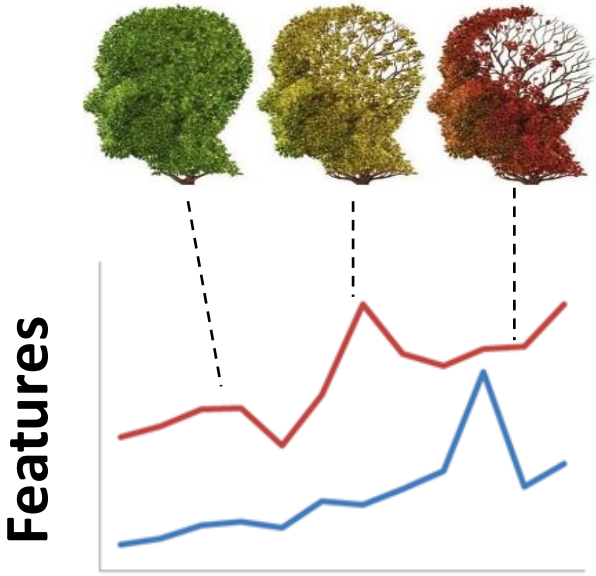
We have enabled individualized tracking of eye-movement features from healthy subjects

Interpretable and reproducible

Accessible, robust, and scalable



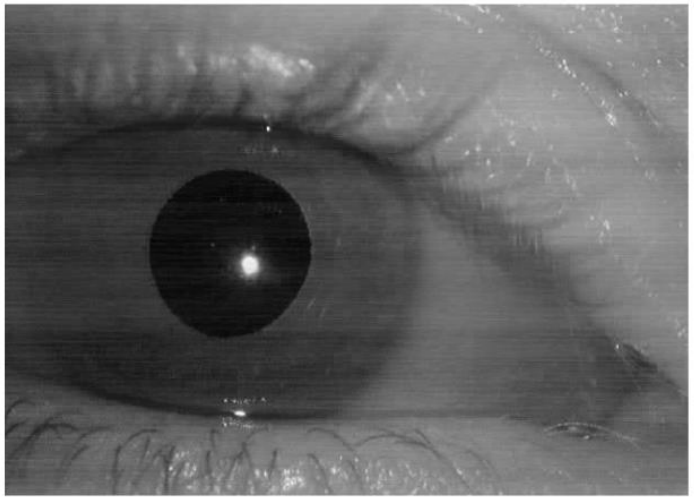




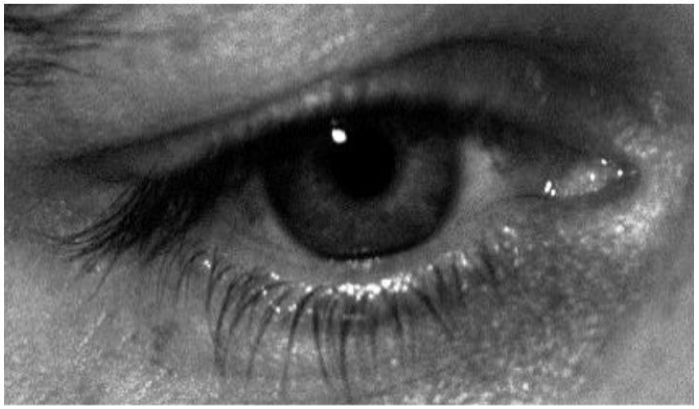
We hope by comparing the trends of the features with the clinical assessments, **we can evaluate the potential to use eye-movement features to track disease progression more frequently and widely than previously possible.**

Back-up

Starburst is applied on IR images



Starburst tracks the movement of the pupil over time.



Pupils are hard to find under natural light.

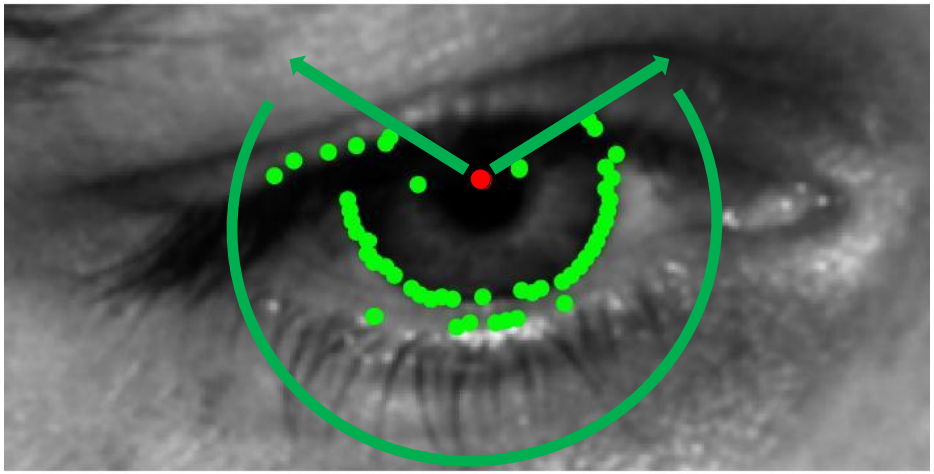
$\Rightarrow$ We track irises instead.

$\Rightarrow$ There are eyelid occlusions!!

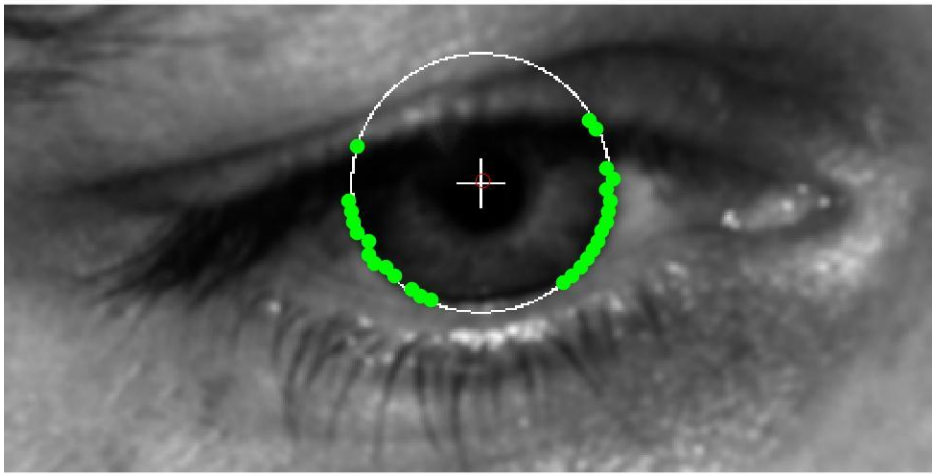
Li et al., *CVPR workshop*, 2005

We modify Starburst for mobile recordings.

Edge detection from a starting point.
(Gradient crosses a threshold => Max gradient)

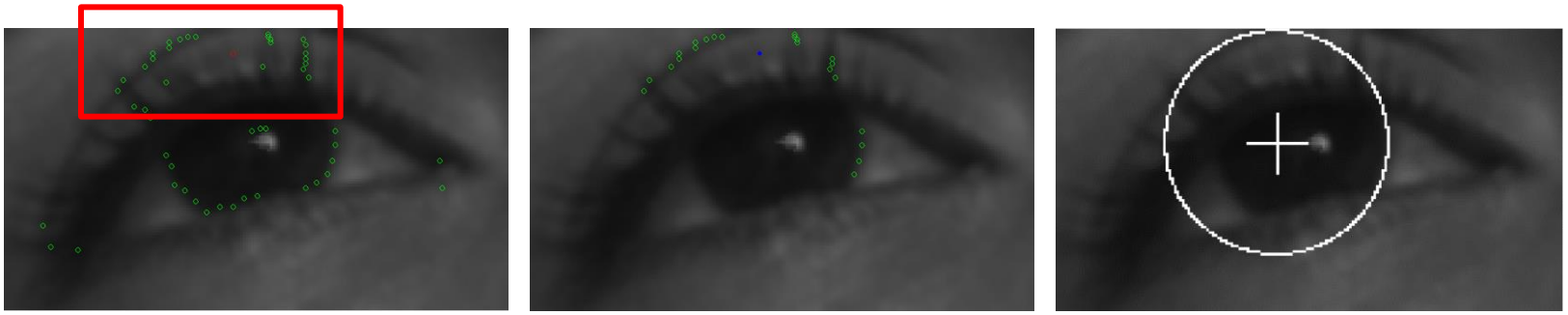


Fit a circle on the edges.
Use the center of the circle as the estimated iris center for this frame and the initial guess for the next frame.
(Ellipse=> Circle.
Pupil=> Iris)

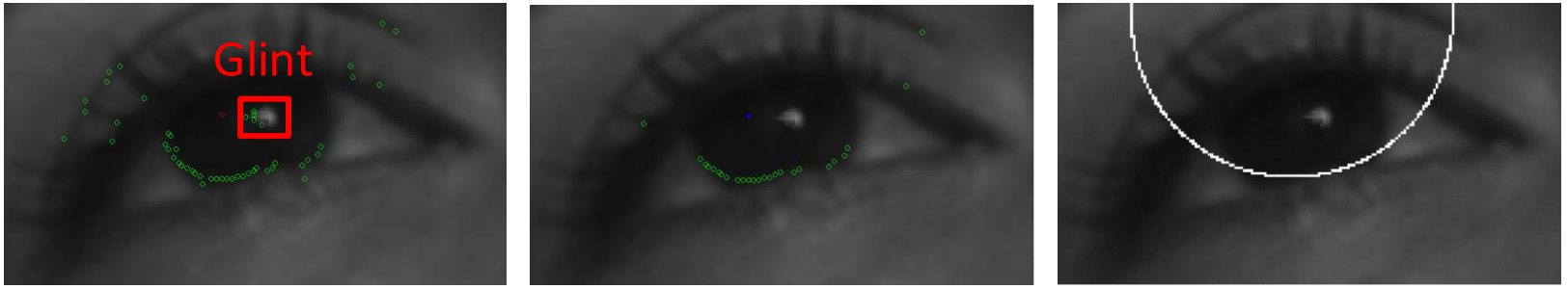


Feature-based algorithm is not robust to various eye appearances

Dark Eyelashes



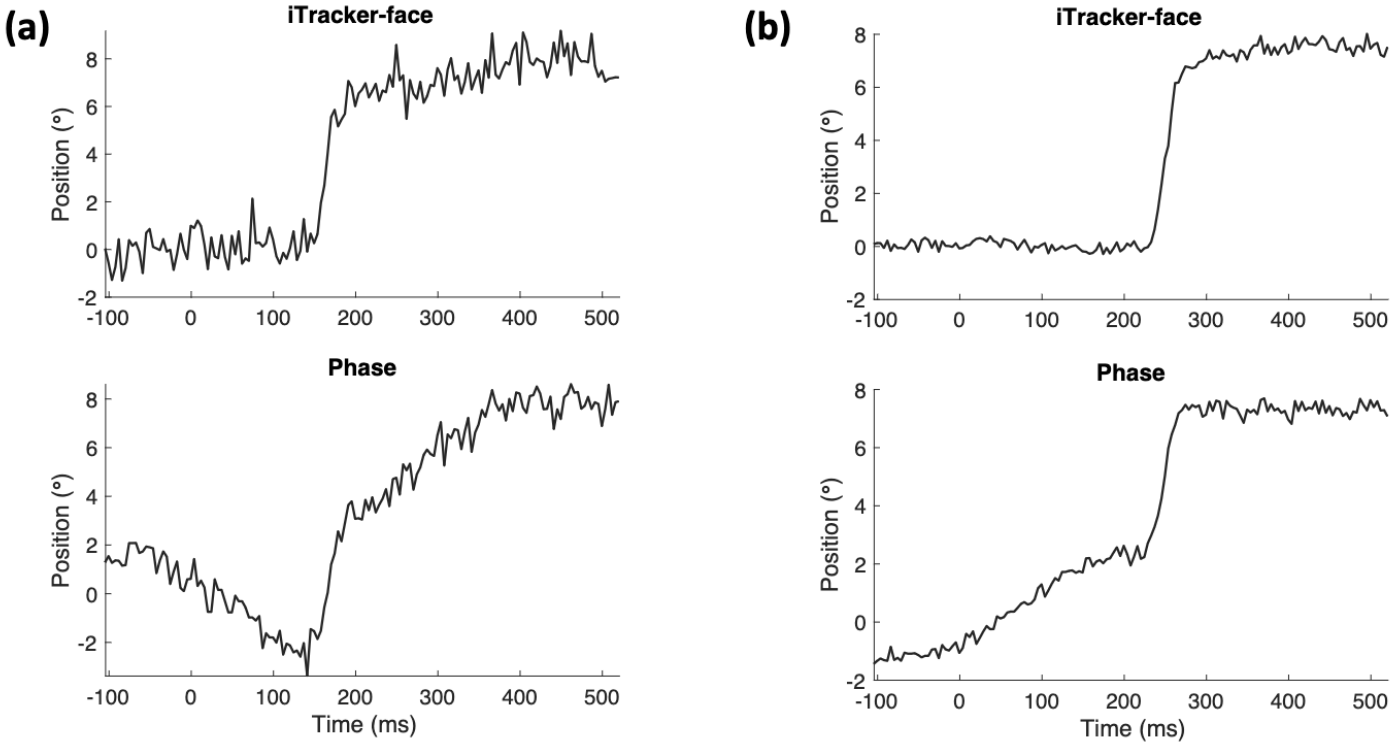
Glint



Circle Fitting Not Suitable

Phase-based algorithms use the relationship between phase shifts and translations.

We edited Visual Microphone to estimate horizontal eye movements.



Even though we can detect when the eyes move in the "Phase" traces, it is hard to find an automated method to reject bad saccades.

Davis et al., *SIGGRAPH*, 2014

Our model captures both individual strategies and population effects

$$p(y_p|g,h) = \prod_{i=1}^{P}\prod_{n=1}^{N} N(y_{pin}; \Sigma_{j=1}^{Q} w_{pij} g_{pj}(t_n) + h_{pi}(t_n), \beta_i^{-1})$$

Multi-task Gaussian Process

Want g_j to learn the "strategy". I currently set Q=1.

$$g_{pj}(t) \sim GP(0, K_{pj}(t,t'))$$

$$h_{pi}(t) \sim GP(\Phi_{sub,i}(t)^T r_{sub,c_p,i} + \Phi_{ind,i}(t)^T r_{ind,p,i}, K_i^h(t,t')),$$

Sub-population Effect Individual Effect

$$r_{ind,p} \sim N(\mu_r, \Sigma_r)$$

$$c_p|x_p \sim Multinomial(\pi_{c_p}), \pi_{c_p} \propto \exp w_c^T x_p,$$

People in the same group shared the same "sub-population" effect.
The assignment of the group is learned by the covariates.

y: Measurements
p: The p-th subject
i: The i-th feature
j: The j-th shared strategy
n: The n-th sample

Want the mean functions of h_i to learn the trend across features such as learning effects or disease progression.

x_p: The covariates (e.g. age, gender, …) of subject p
c_p: The group subject p belongs to

The screen and the camera need to be synchronized

We use the screen timestamps to know the stimulus presentation time.

We use the camera timestamps to measure the onset of the eye movement.



Reaction time = Onset of the eye movement – Stimulus presentation time

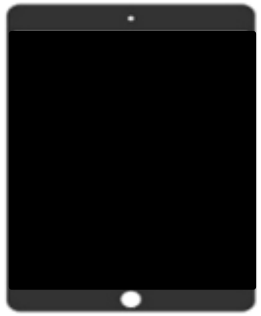
We achieved a synchronization error < STD of reaction time.

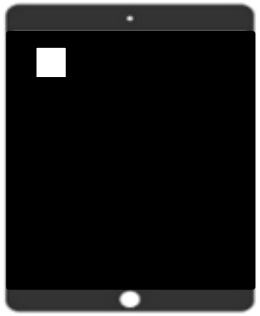
(<5 ms) (~ 40 ms)

Use a mirror to evaluate the synchronization between the screen and the camera

Experiment: Place an iPad in front of a mirror and record itself

Screen

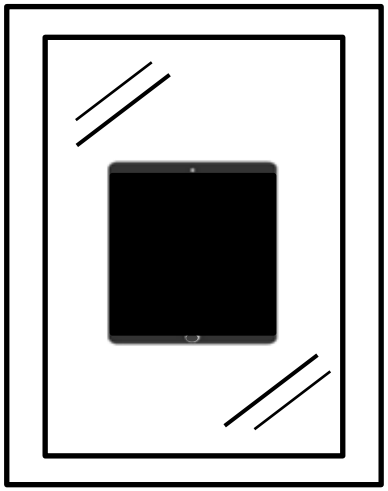


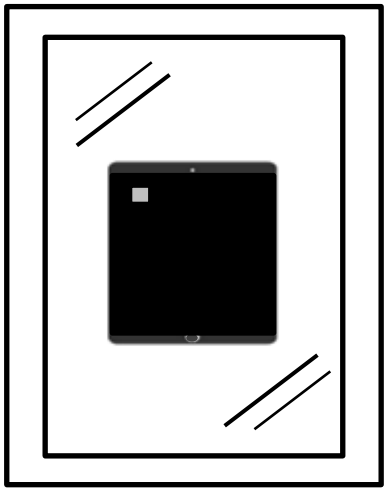


Screen Timestamp = s ms
Ground Truth = s $+D_1$ ms

$$\underbrace{r + D_2 - \frac{1000}{60}}_{\text{Last Camera Frame}} < s + D_1 < \underbrace{r + D_2}_{\text{First Frame with Stimulus}}$$

Camera





Camera Timestamp = r ms
Ground Truth = r $+D_2$ ms

Let $D_1 - D_2 = D$.

We can use this equation to estimate D, the synchronization error.

Estimate D

$$r_i - \frac{1000}{60} < s_i + D < r_i$$

We do not know D. But if we assume D, we can check all the camera timestamps to define $\tilde{r}_i(D)$ to satisfy

$$\tilde{r}_i(D) - \frac{1000}{60} < s_i + D < \tilde{r}_i(D).$$

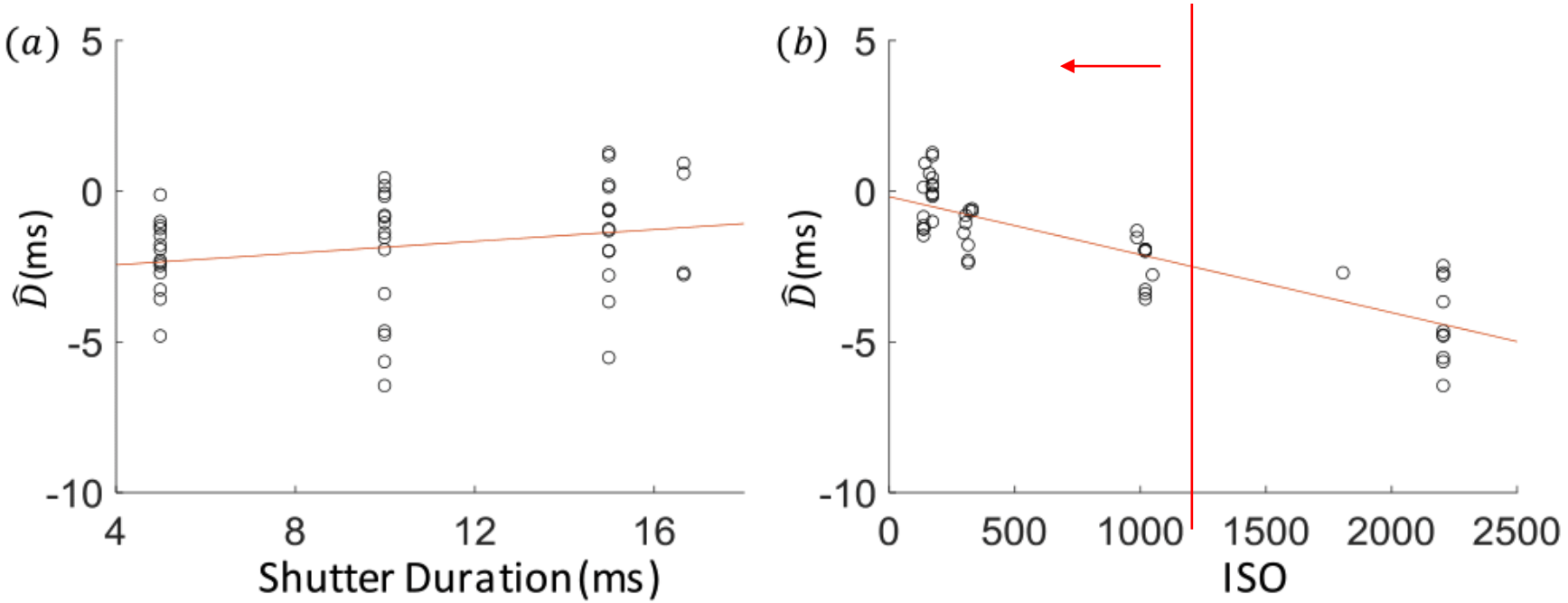
If the estimated D is correct, then $\tilde{r}_i(D)$=r_i.

$$\hat{D} = argmin_D \sum_i |\tilde{r}_i(D) - r_i|.$$

After some iterations of app development, we can bound $\widehat{D}$ to be within 5 ms.

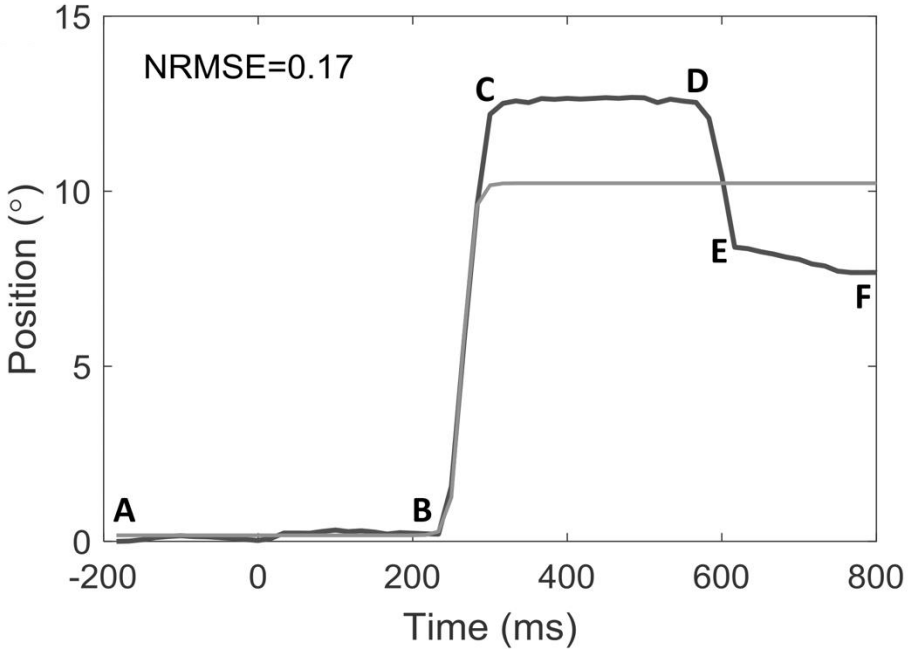
The synchronization error is within 5 ms

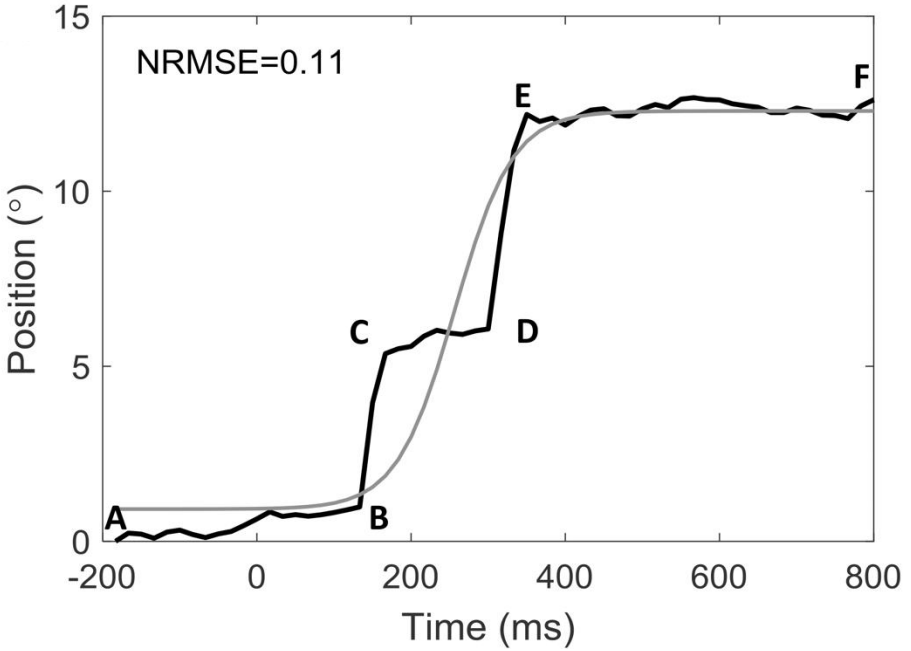
We require ISO < 1000.



The window of fit needs to be adaptive to capture different traces

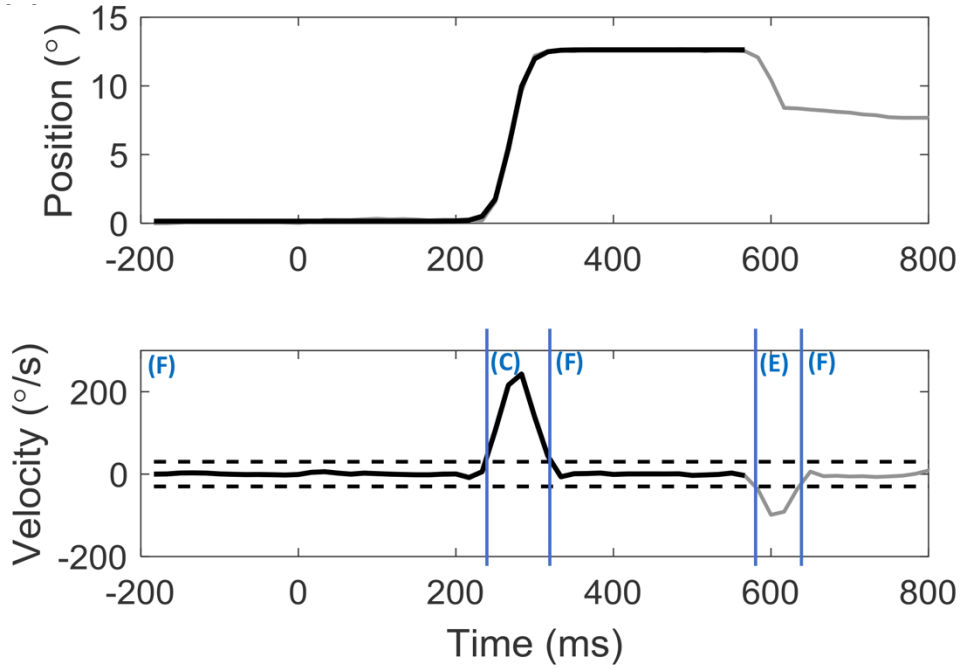
This is one of the phenotypical responses of a Parkinson's disease patient.

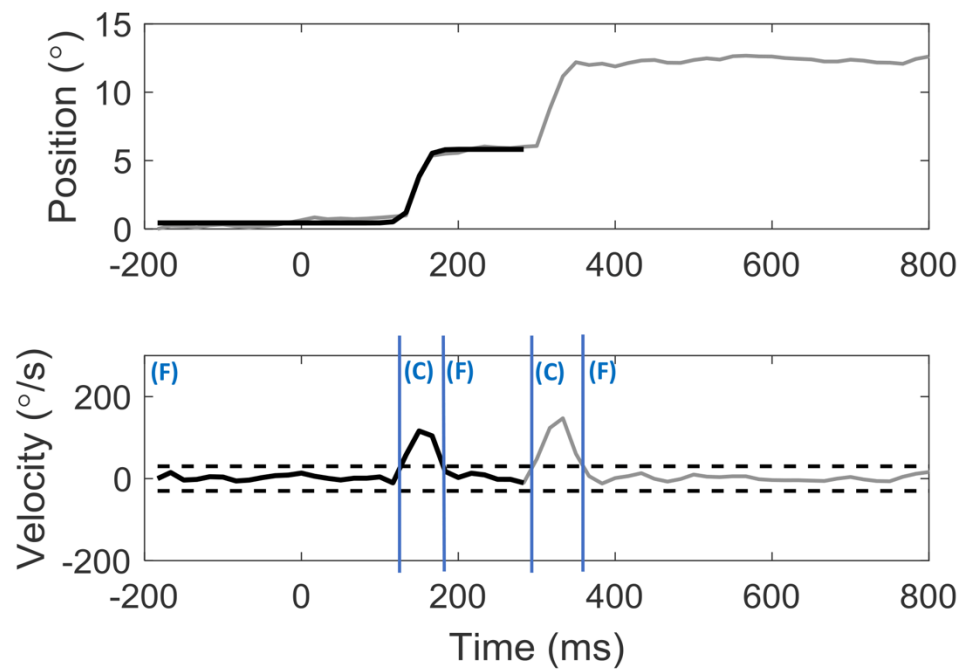




Previously, we fit a tanh on the entire trace. Now we need to define the window of fit => the first eye movement after stimulus presentation i.e. A->D

Velocity can be used to determine the window of fit





The window of fit is the first fixation-correct-fixation sequence.

With an adaptive window of fit, the proportion of bad saccades: 17% => 3%

How do mobile-device recordings compare to recordings made with a research-grade camera?

We find a proper error rate definition

- Assumptions:
  - $P(dE|C) \approx 0, P(dC|E) \approx 0,$
  - $P(LS|E) \approx P(LS|C),$
- Then

False positives and false negatives are rare.

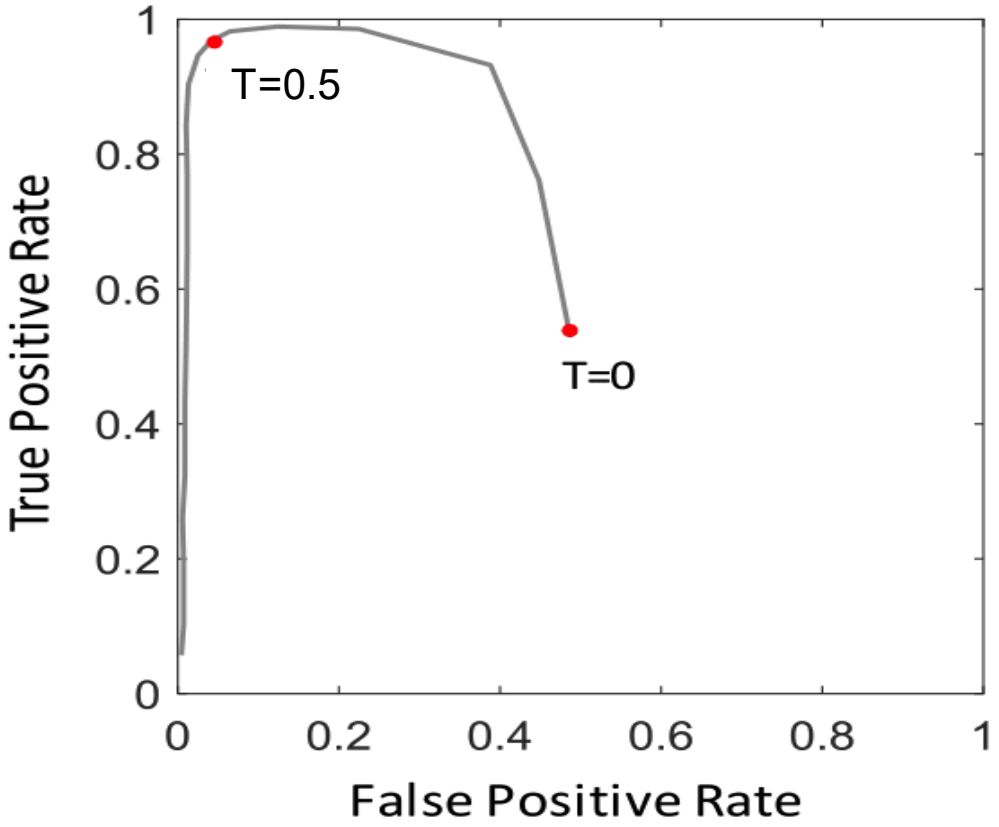
LS does not bias towards correct saccades or errors.

$$
\begin{aligned}
&\frac{P(dE)}{1-P(LS)} \\
=&\frac{P(dE|E)P(E)+P(dE|C)P(C)}{1-P(LS|E)P(E)-P(LS|C)P(C)} \\
\approx&\frac{P(dE|E)P(E)}{P(E)[1-P(LS|E)]+P(C)[1-P(LS|C)]} \\
\approx&\frac{[1-P(LS|E)]P(E)}{[1-P(LS|E)]P(E)+[1-P(LS|C)]P(C)} \\
\approx&\frac{P(E)}{P(E)+P(C)} \\
=&P(E)
\end{aligned}
$$

Error rate = dE/(#saccades-#LS)

We thus completed an automated measurement pipeline.

Our error detection achieves high sensitivity and specificity



- As T goes to 0, sn_t and sp_t are both likely to cross the threshold due to noise.
- If we choose T=0.5, we achieve 0.97 sensitivity and 0.97 specificity.