НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

# Modeling temporal dynamics of attention with Leaky Integrate-and-Fire model


Liya Merzon

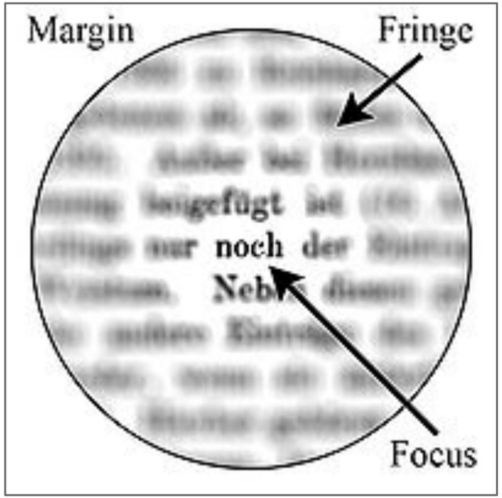
Georgii Zhulikov

# Visual attention

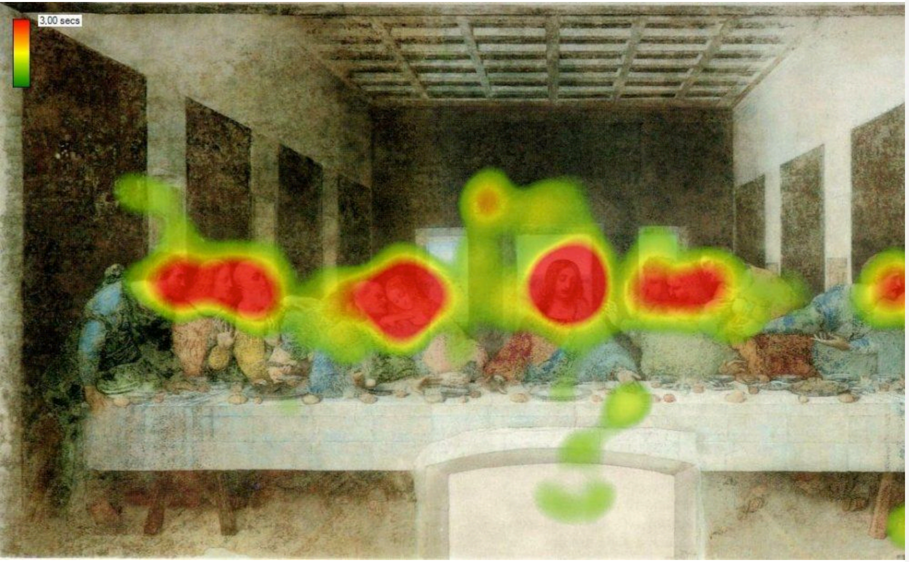
Despite limits to the processing capacity of the human visual system, we can process incoming visual information in real time.

Attentional shift occurs when directing attention to a point to increase the efficiency of processing. When an object or area is attended, processing operates more efficiently

(Posner, 1980; Carrasco, 2011)




# Different way to look at the data




To model both spatial and temporal will be the best

# Saliency model (Itti & Koch, 2000; toolbox Walther & Koch, 2006)

Pros:

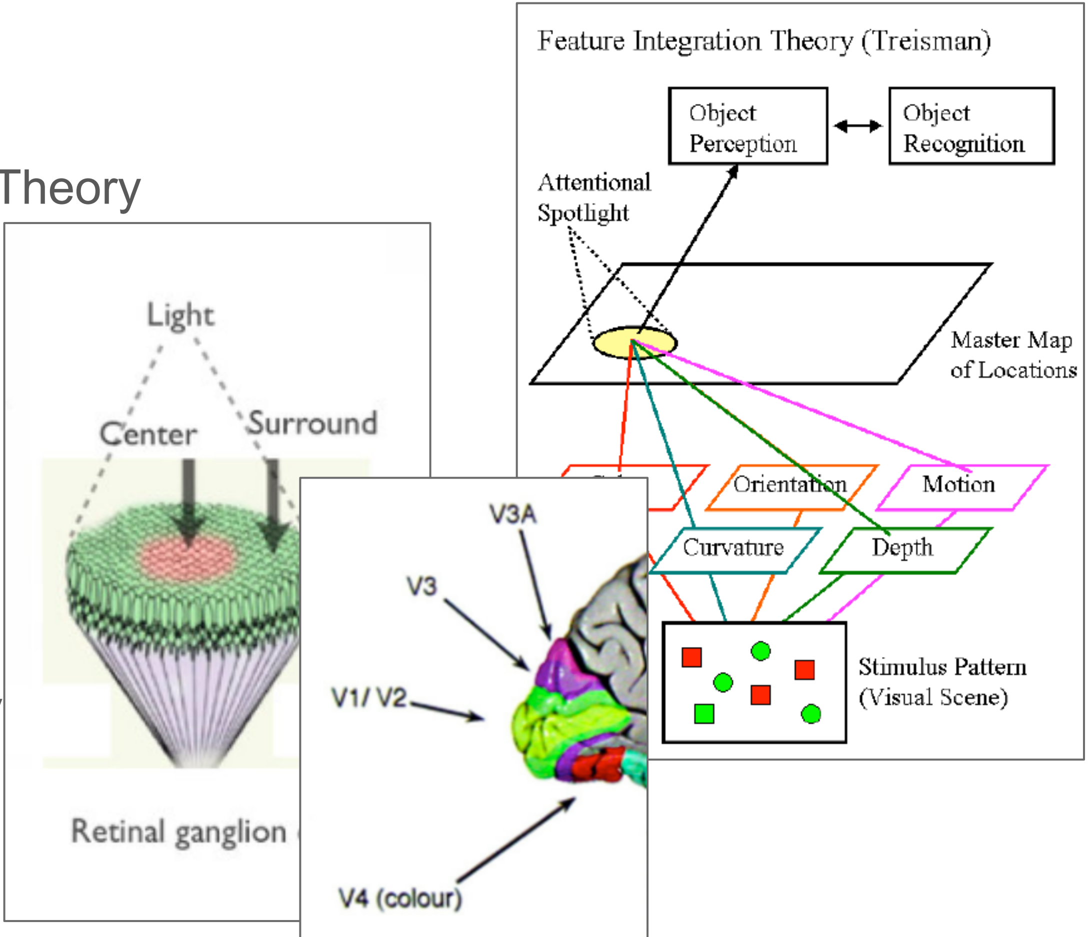
- Based on the Feature Integration Theory

- Neurally plausible

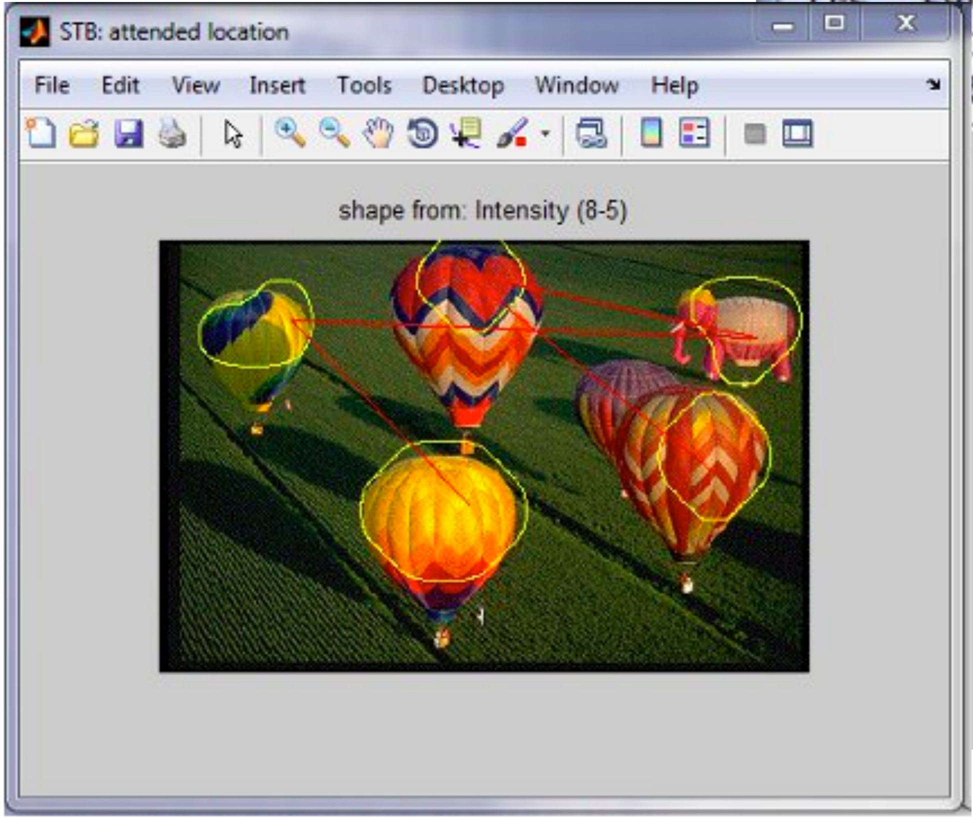
- Testable

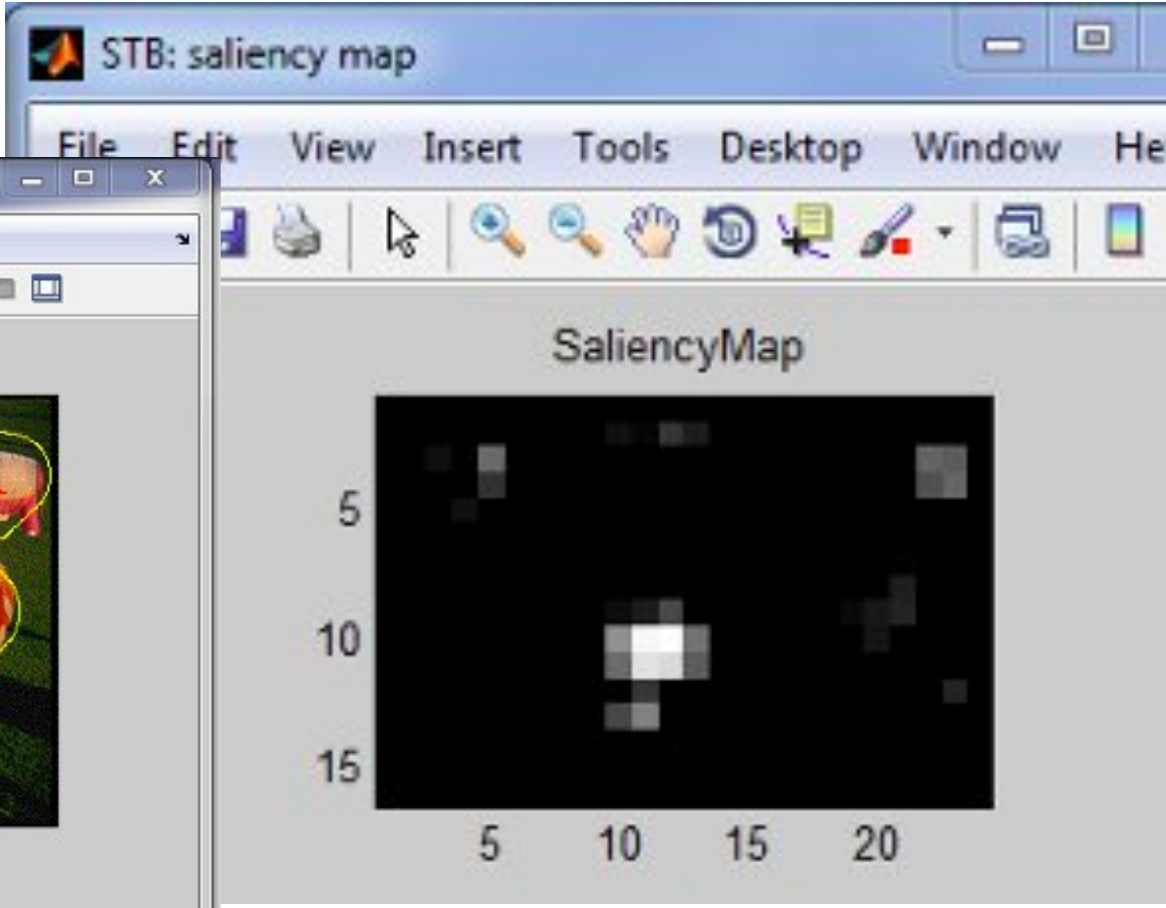
Cons:

- Less accurate than modern deep neural nets (see MIT Saliency Benchmark)

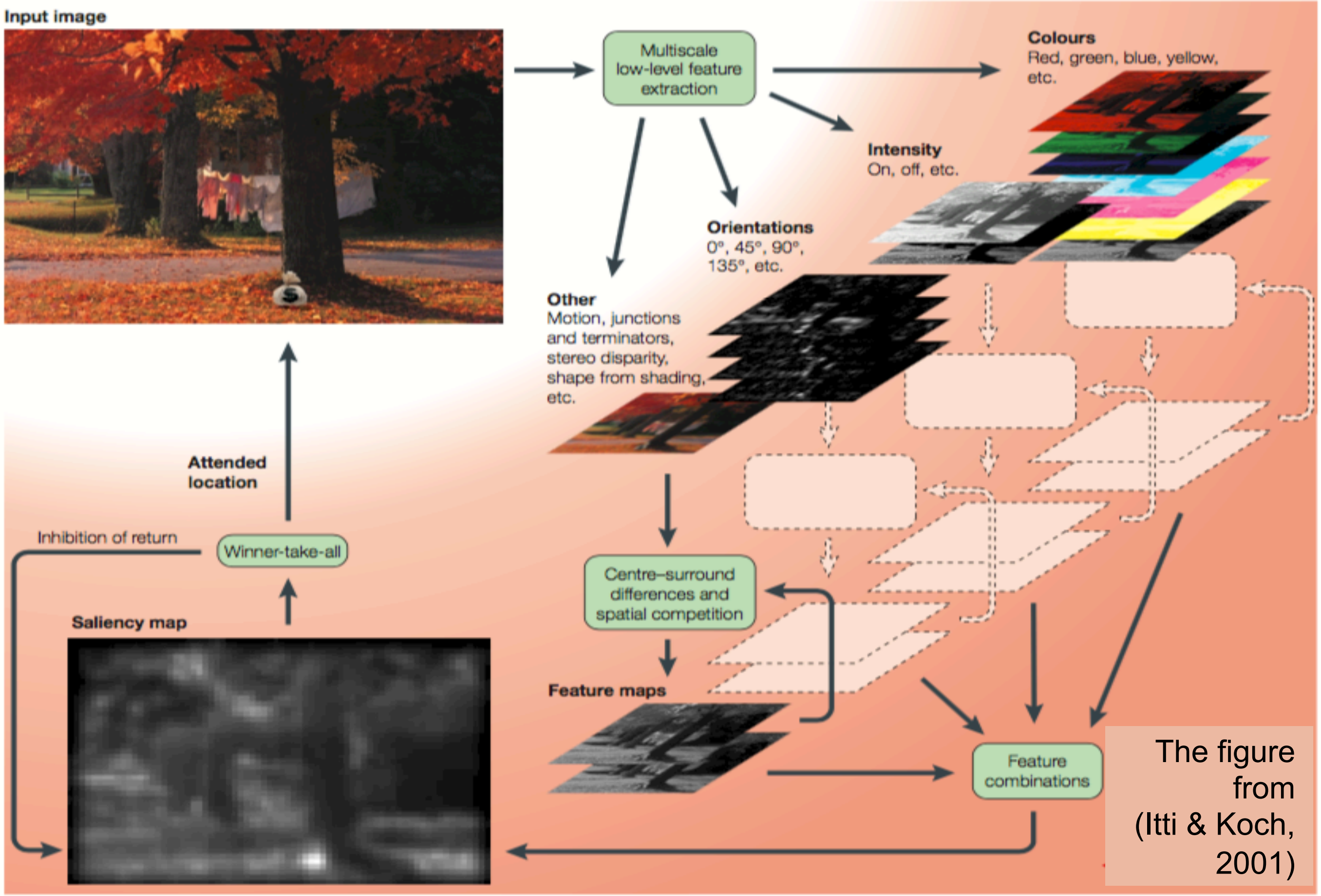


# What does it do?



# How is it done?




The figure from (Itti & Koch, 2001)

But
the model also allows
temporal predictions

# Temporal prediction of Itti & Koch (2000) model

- The most salient location is marked as a fixation
- Inhibition of return is applied to that location
- The process produces an ordered set of fixations
- A model of reaction times is needed to turn “an ordered set of fixations” into “a gaze simulation”
- Saliency toolbox uses Leaky integrate-and-fire model

# LIF

Leaky integrate-and-fire is a simplified model for spiking neurons

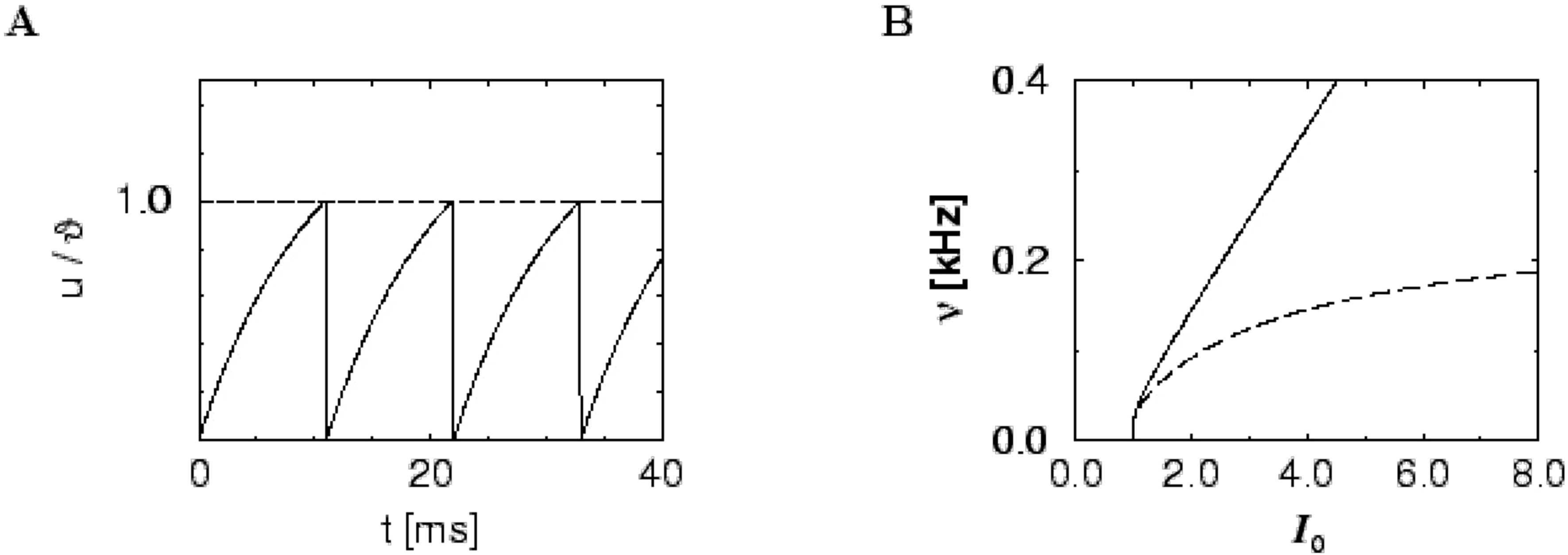
$$C_m \frac{dV_m(t)}{dt} = -\frac{1}{R_m}\left[V_m(t) - V_{rest}\right] + I_{ext}(t)$$

$$\text{if } V > V_{thr} \text{ then Spike and } V = V_{reset}$$

Spikes are events and the model controls only the timing of these events

For the spikes to occur, the input current has to exceed a threshold

# LIF




$$C_m \frac{dV_m(t)}{dt} = -\frac{1}{R_m}[V_m(t) - V_{rest}] + I_{ext}(t)$$

$$\text{if } V > V_{thr} \text{ then Spike and } V = V_{reset}$$

# Aim of the study

There are over 20 parameters of LIF layer (leak potential, potential for excitatory and inhibitory channels, conductivity of the channels, etc.)

These parameters are influence reaction time generated by the model

Goals:

1) Check the default parameters on temporal accuracy
2) Find the parameter space, that allows reproduce temporal dynamics in human data

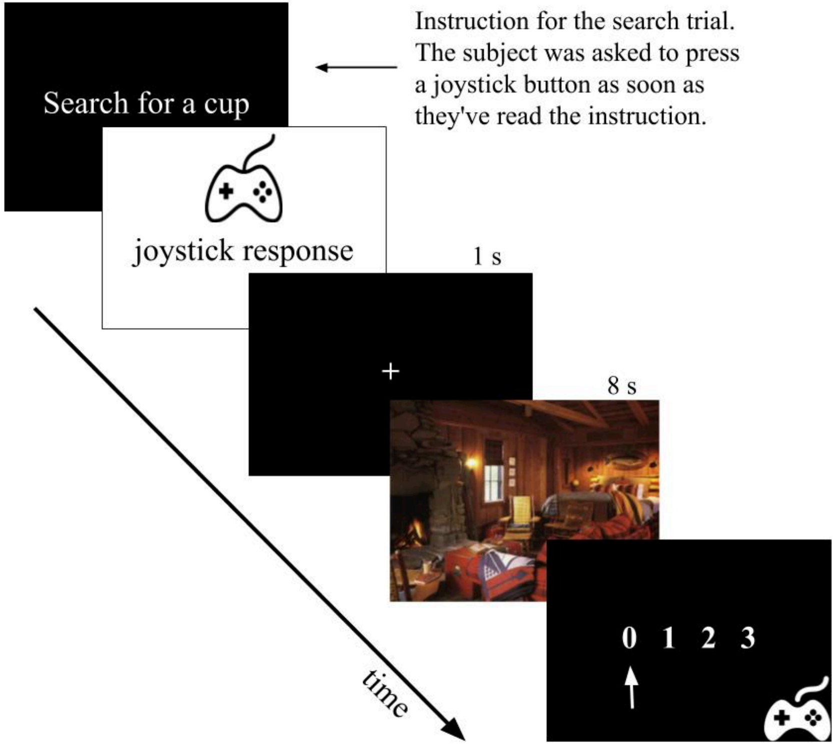
# Data for training: LabelMe database

The open database (Russell, Torralba, Murphy, & Freeman, 2008) with images of natural scenes.

# Ground truth: human data

Visual search experiment (Gordienko, 2016),

Only duration of the first fixation was used in the current study (as the model was tested without IOR)



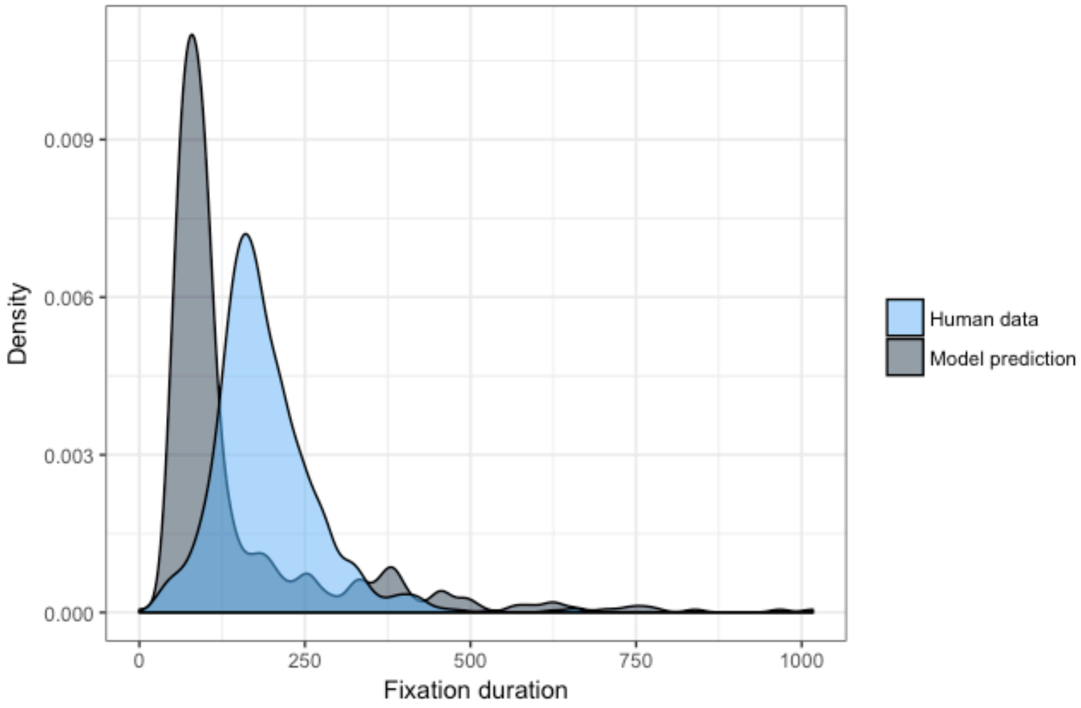
# Experiment #1. Default Saliency toolbox

The default parameter space from Walther and Koch (2006) against the human data.

Model vs human

Mean: 150 ms vs 190.8 ms

SD: 146 vs 77.7



# Experiment #2. Optimization of parameters

Genetic algorithm

- Very good for complex functions and processes
- Slower
- Unpredictable
- Hard to look into and debug

Nelder-Mead algorithm

- Good when there is no access to gradients
- Faster
- More predictable
- Can’t debug (standard MATLAB function)
- Bad at avoiding local minima

Optimized values: Kolmogorov-Smirnov statistic, Z-test statistic

# Results on optimized parameters

The best parameters were found by by Nelder-Mead algorithm.

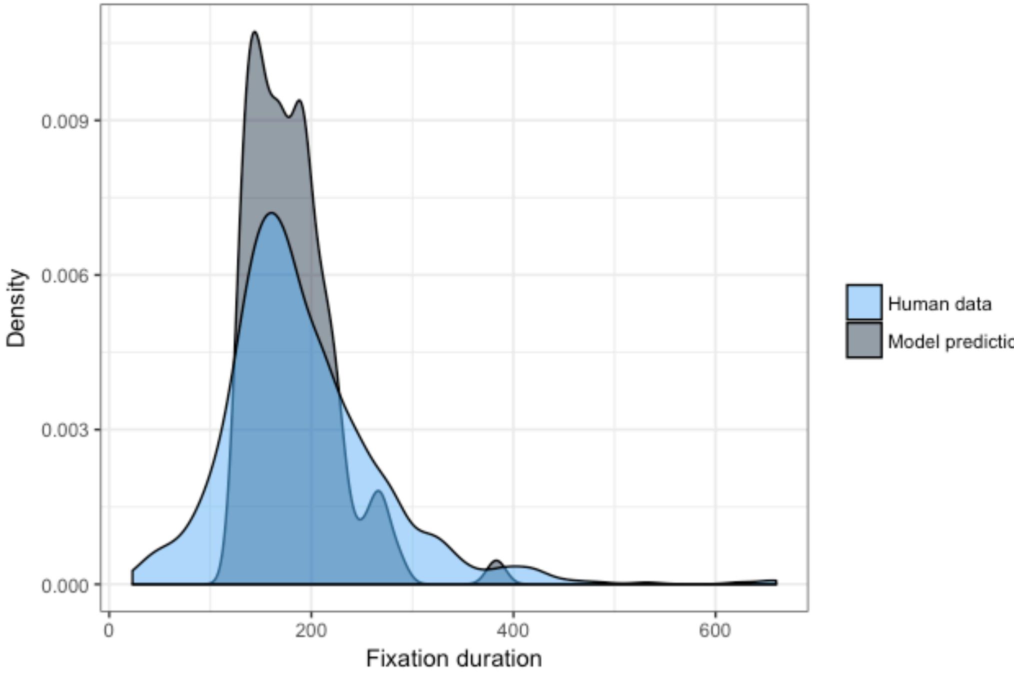
Optimization function: z-test + ks-test

Z-test statistics:

z = 3.1117, p = 0.00186

Ks-statistics:

D = 0.15857, p = 1.305e-09.



# Discussion of the result

In the previous experiments we looked at the whole distribution at once.

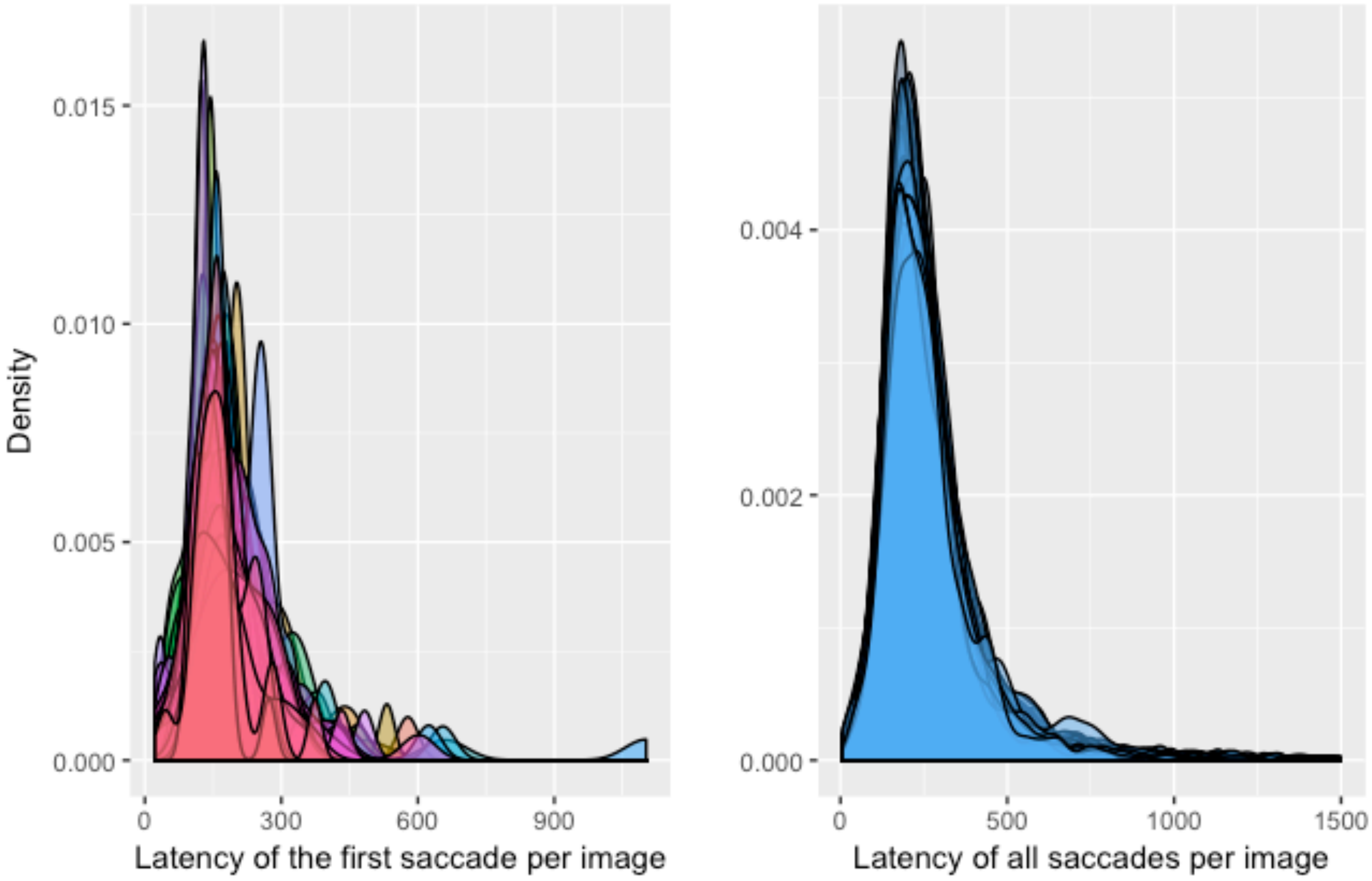
But human data on one separate image produce a reaction time **distribution.**

The model produce one the same value in any run on the same image. The model has noise components, which theoretically allows to produce a distribution, but doesn’t learn to use the noise.

The model use only differences between images.

Let’s force it to learn within image variability of fixation durations

# Human data grouped by image



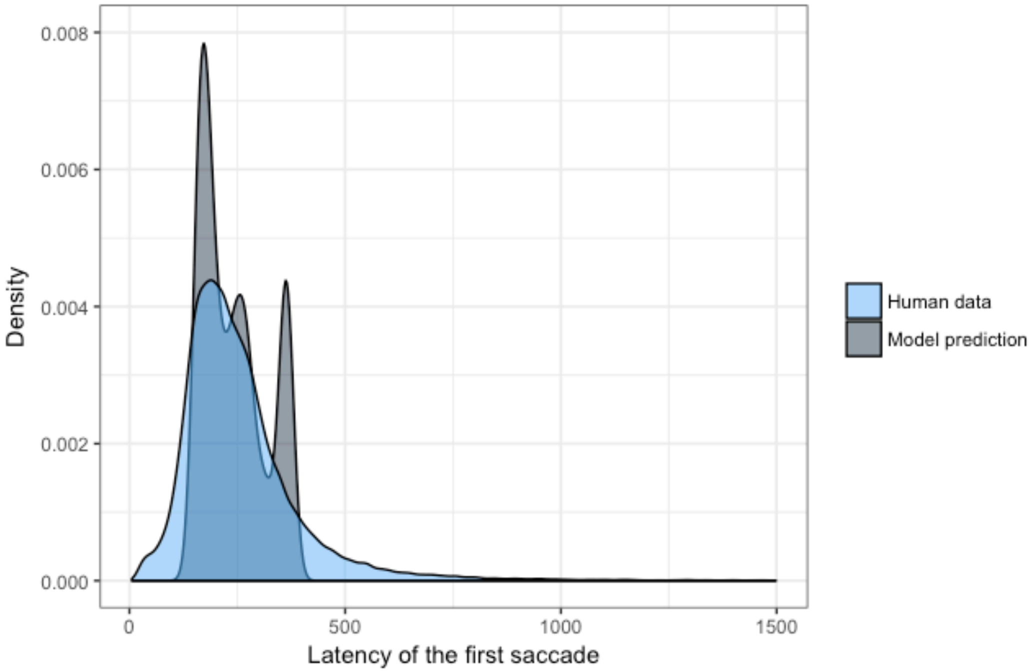
# Result of Experiment #3 (on separated images)

Z-test statistics:
$z = 6.5611$, $p = 5.341e-11$.

Ks-statistics:
$D = 0.19327$, $p < 2.2e-16$

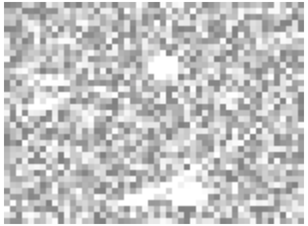
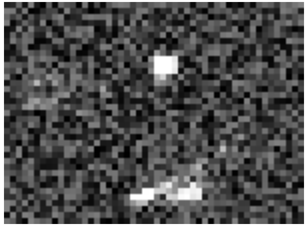
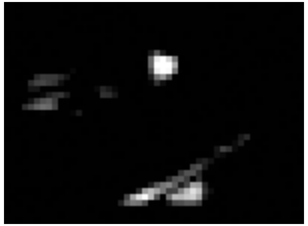
Not more than 2 different time per image

The best noise constants were ($10^{-14}$ and $10^{-11}$ for random and constant noise respectively)



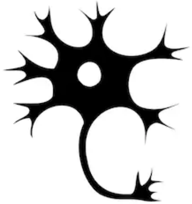
# Investigation of randomness sources in LIF

- Distributions are only possible with random values
- The source of randomness: noise in saliency maps

| Noise amplitude | Noise constant | |
|---|---|---|
| | $10^{-9}$ | $10^{-11}$ |
| $10^{-9}$ |  |  |
| $10^{-11}$ |  |  |

# Discussion

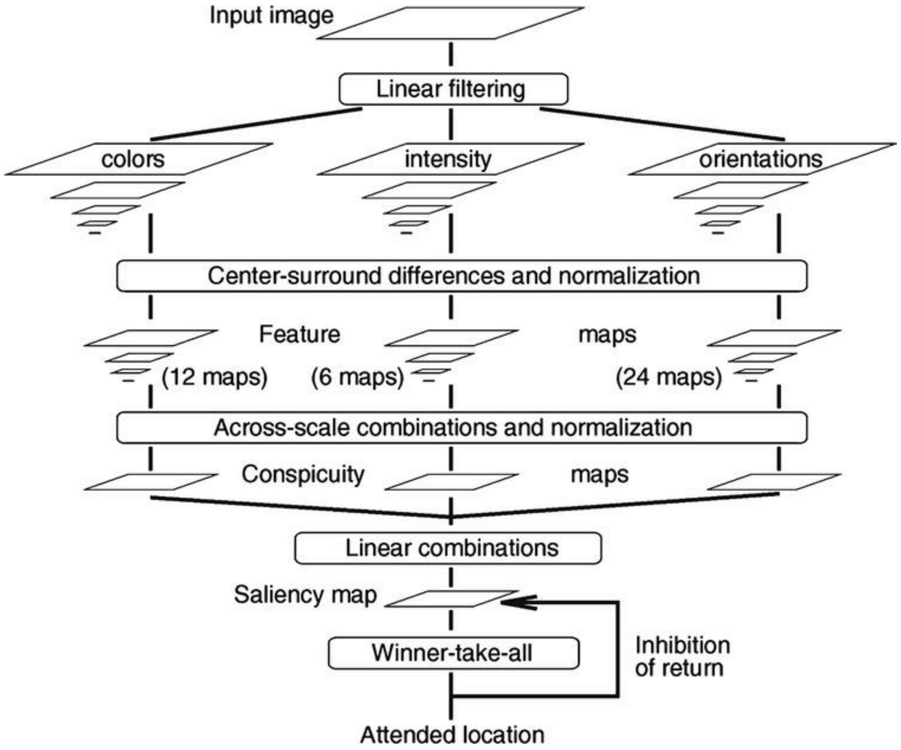
- The model doesn’t have enough random components to produce a distribution of reaction times
- The generated distributions are the result of differences between images
- Adding random components may be a good idea
    - Drift diffusion component
    - A more complex spiking model



Thank
you
for
your
attention!

# Additional slides

# Saliency model

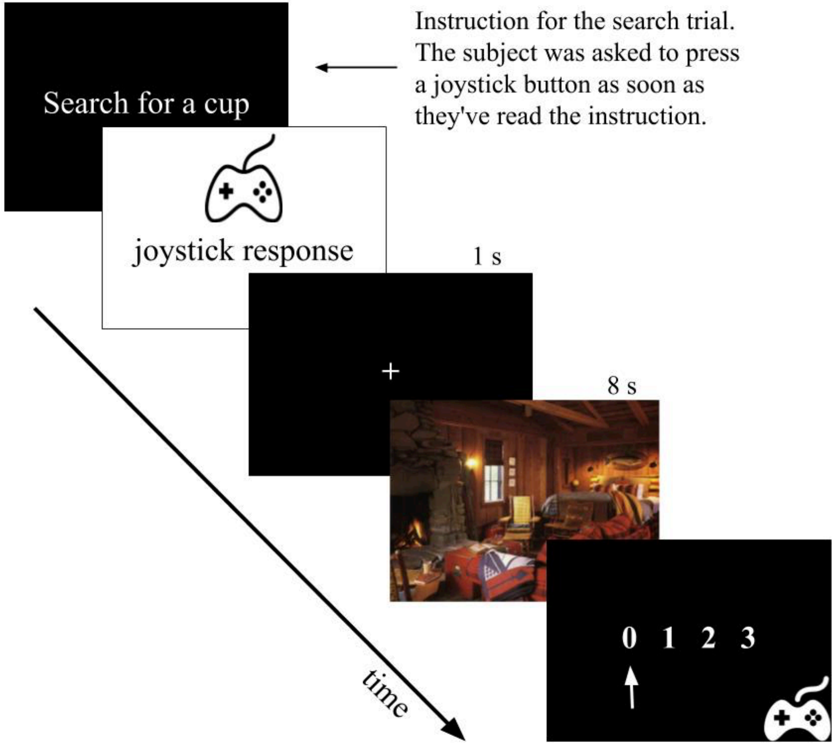


# Ground truth: human data

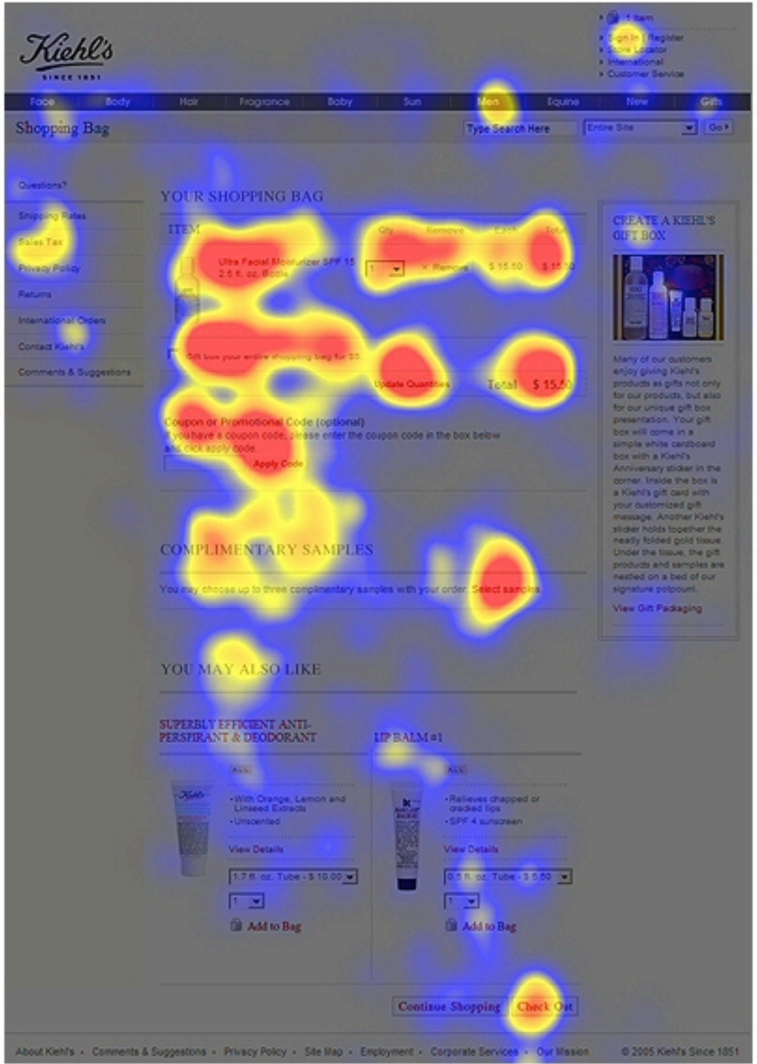
Visual search experiment (Gordienko, 2016), 16 participants, search for a cup or for a painting.

Duration of the first fixation was used in the current study.

The first, smaller dataset, includes data, collected on 44 pictures. Information includes 782 first fixations (29528 fixations in total). This dataset was used in Experiment #1. The second, larger, dataset includes data, collected on 91 pictures: 1593 first fixations



# Eye movements and overt attention



# Temporal distribution of reaction times

- Not only where a fixation happens, but **when** it happens
- Distribution of modeled reaction times compared to the distribution of human reaction times
- Only taking the first saccade into account, to minimize IOR’s effect

# Modified Saliency toolbox

- Removed IOR
- Separated LIF layer
- Added comparison with human data
- Added possibility to run optimization algorithms
- Added genetic optimization algorithm

# Drift-diffusion model

- A model for the two-alternative forced choice task
- Fits saccadic reaction times

$$dx = vdt + cdW$$
$$x(0) = x_0$$

- At each step a highly stochastic value is added to the current state