Bernstein Center for
Computational Neuroscience
Berlin

Kernel Methods and Perceptual Classification

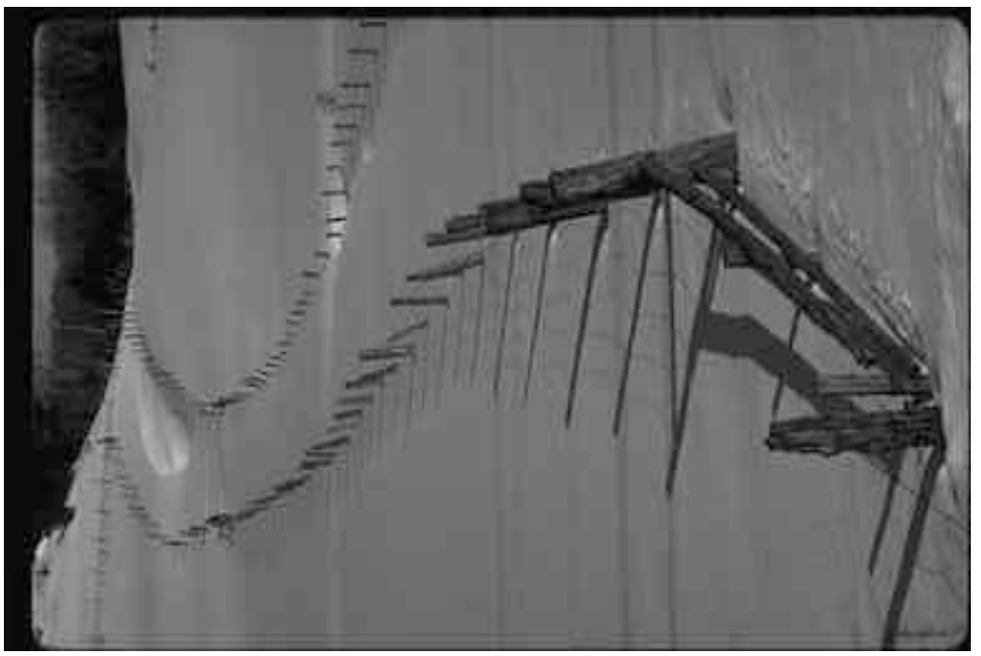
Felix A. Wichmann

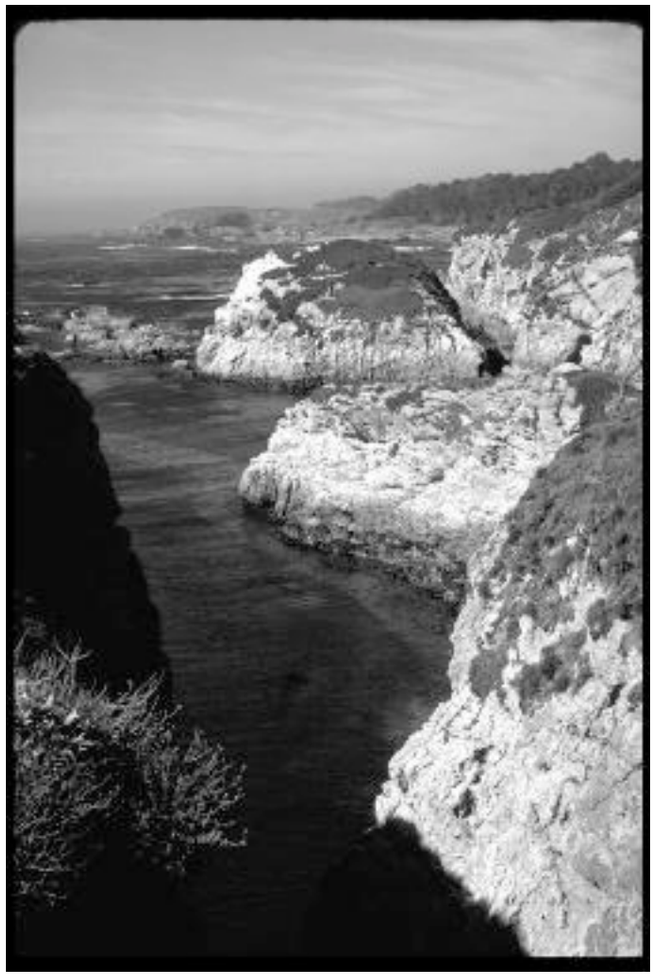
Modelling of Cognitive Processes Group
Bernstein Center for Computational Neuroscience
and
Technische Universität Berlin

felix.wichmann@tu-berlin.de

Bundesministerium
für Bildung
und Forschung


RAIL
ROAD
CROSSING






The Steamship Brand
Molasses Candy

How many animals?

Animal detection in natural scenes: Critical features revisited

Felix A. Wichmann — Modelling of Cognitive Processes, Berlin Institute of Technology & Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany

Jan Drewes — Abteilung Allgemeine Psychologie, Universität Giessen, Giessen, Germany

Pedro Rosas — Centro de Neurociencias Integradas, Facultad de Medicina, Universidad de Chile, Santiago, Chile

Karl R. Gegenfurtner — Abteilung Allgemeine Psychologie, Universität Giessen, Giessen, Germany

S. J. Thorpe, D. Fize, and C. Marlot (1996) showed how rapidly observers can detect animals in images of natural scenes, but it is still unclear which image features support this rapid detection. A. B. Torralba and A. Oliva (2003) suggested that a simple image statistic based on the power spectrum allows the absence or presence of objects in natural scenes to be predicted. We tested whether human observers make use of power spectral differences between image categories when detecting animals in natural scenes. In Experiments 1 and 2 we found performance to be essentially independent of the power spectrum. Computational analysis revealed that the ease of classification correlates with the proposed spectral cue without being caused by it. This result is consistent with the hypothesis that in commercial stock photo databases a majority of animal images are pre-segmented from the background by the photographers and this pre-segmentation causes the power spectral differences between image categories and may, furthermore, help rapid animal detection. Data from a third experiment are consistent with this hypothesis. Together, our results make it exceedingly unlikely that human observers make use of power spectral differences between animal- and no-animal images during rapid animal detection. In addition, our results point to potential confounds in the commercially available "natural image" databases whose statistics may be less natural than commonly presumed.

Keywords: rapid animal detection, natural scenes, power spectrum, amplitude spectrum, scene gist, local features, natural image statistics

Citation: Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision, 10*(4):6, 1–27, http://journalofvision.org/10/4/6/, doi:10.1167/10.4.6.

Critical Features: System Identification

- "Determining the features of natural stimuli that are most useful for specific natural tasks is critical for understanding perceptual systems" (Geisler, Najemnik & Ing, *Journal of Vision*, 2009, 9(13)17: 1-16.

- In neurophysiology, we want to determine what features of a stimulus make a neuron spike.

- In psychophyics, we want to find the features that determine the decisions of an observer.

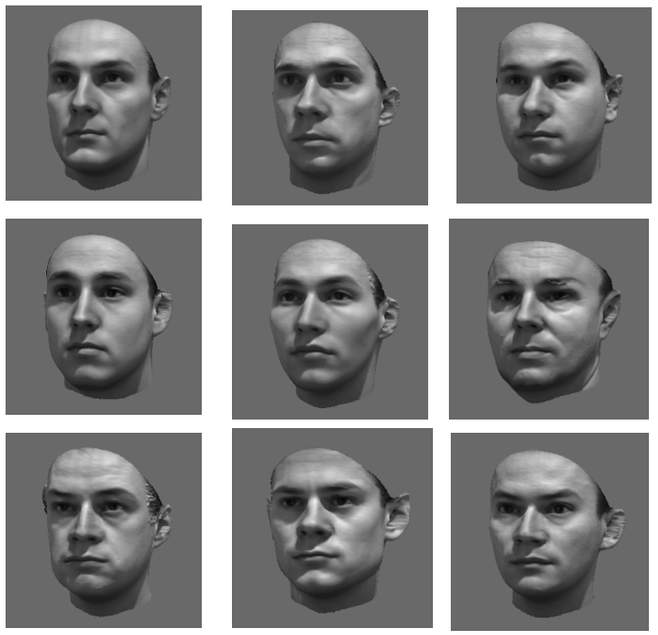
- *Approach: Reverse engineering an algorithm mimicking human behaviour—inverse machine learning!*

- First: Demonstrate how regression techniques can be used to extract the features which are predictive of the decisions of human observers in a classification task.

- Second: Use non-linear kernel extension to find the features which are predictive of human fixation target selection in a free viewing task (visual saliency).

- Third: Show the importance of sparse regularization in a human auditory task.

Gender Categorization of Human Faces



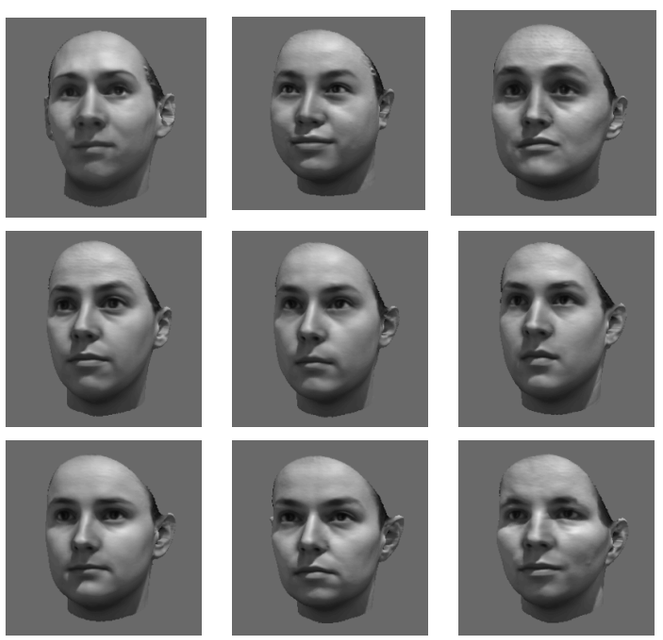
- We eliminated "obvious" cues such as

mean and variance

size of faces

texture (i.e. facial hair)

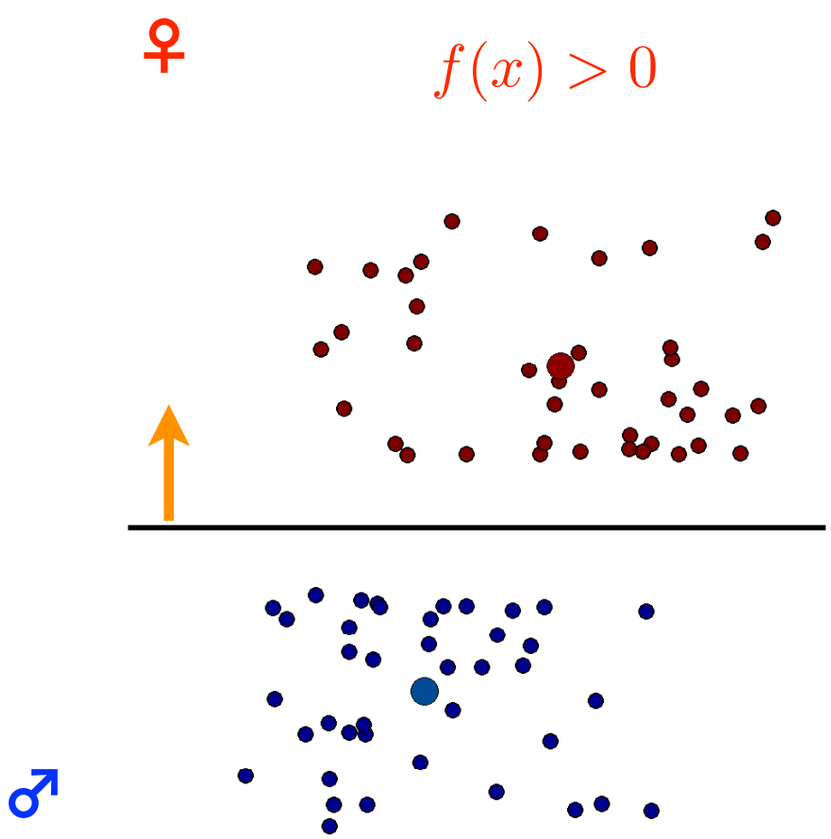
- In what ways are (perceived) "female" faces different to "male" faces?



- Can we find statistical quantities that differentiate one class of images from the other class?

$$
\begin{aligned}
f : \text{all images} &\rightarrow \mathbb{R} \\
f(\text{image}) &> 0 \text{ if female} \\
f(\text{image}) &< 0 \text{ if male}
\end{aligned}
$$

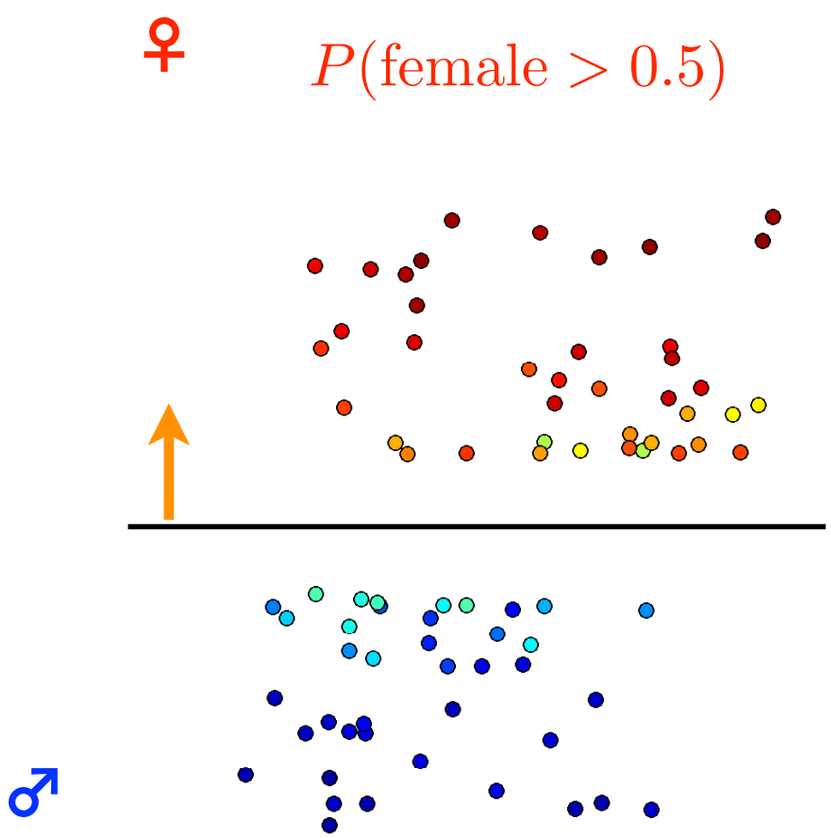
Linear Decision Rules



- We restrict ourselves to linear functions:
$$f(x) = \omega^\top x + b$$

- ω, the normal to the decision hyperplane, is called the decision image.

Linear Decision Rules

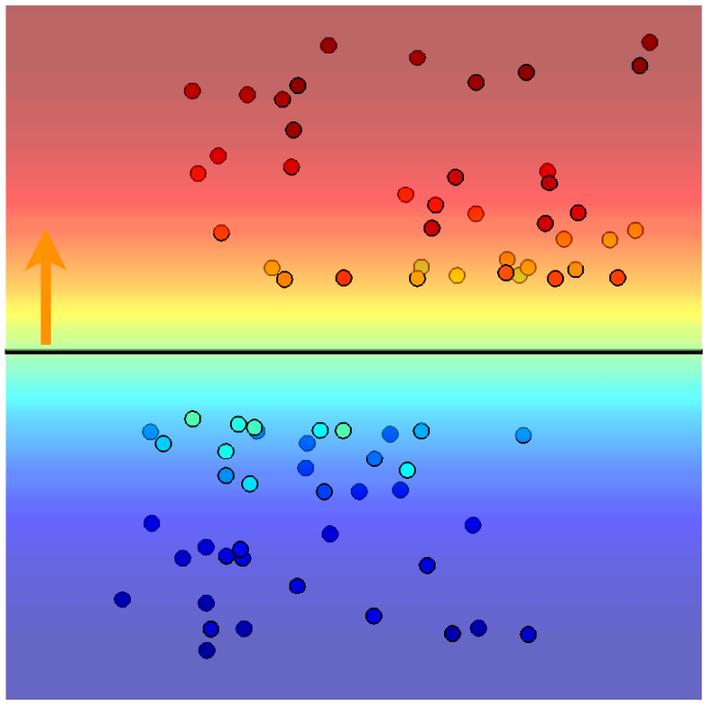


- We restrict ourselves to linear functions:
$$f(x) = \omega^\top x + b$$

- ω, the normal to the decision hyperplane, is called the decision image.

Linear Decision Rules

♀ $P(\text{female} > 0.5)$



♂ $P(\text{female} < 0.5)$

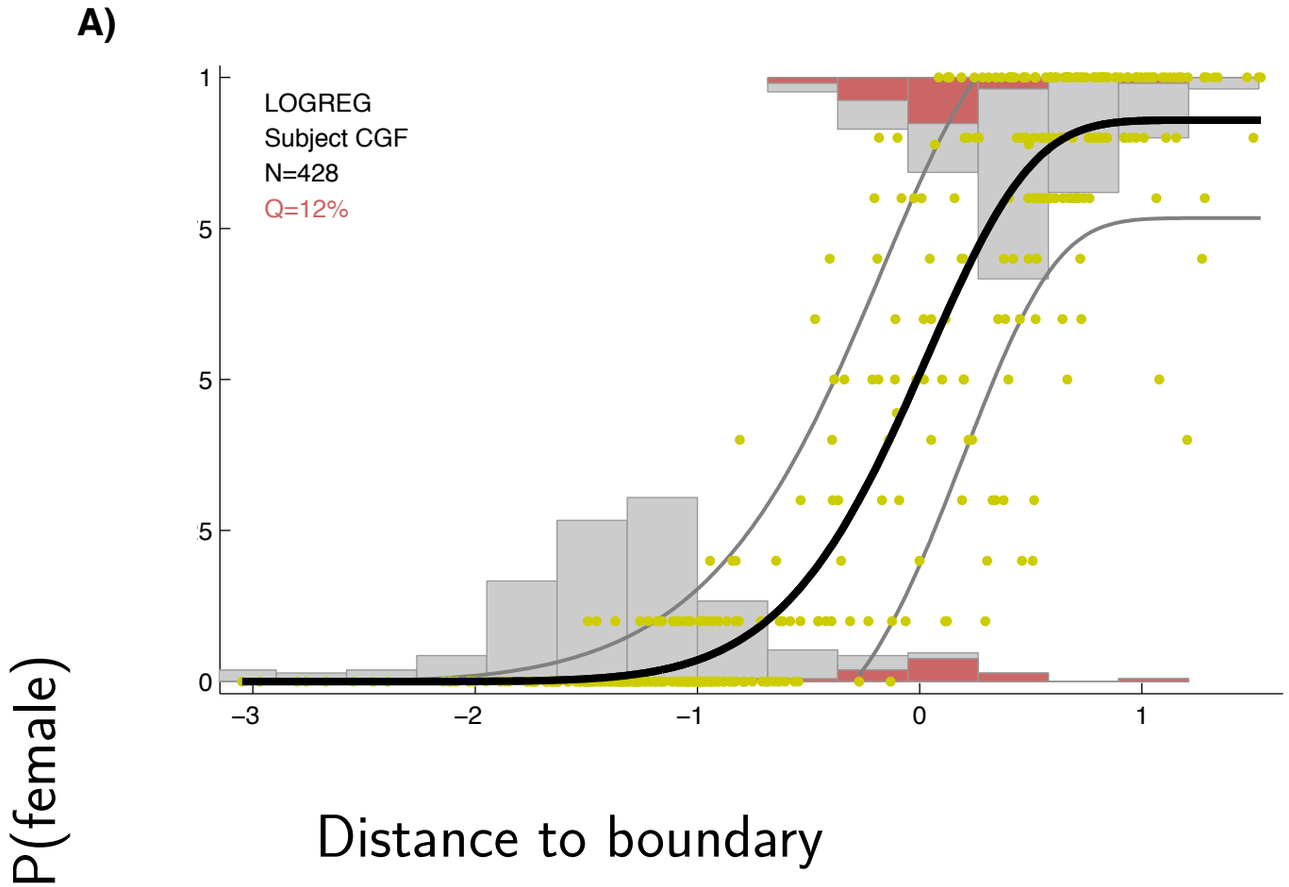
- We restrict ourselves to linear functions:
$$f(x) = \omega^\top x + b$$

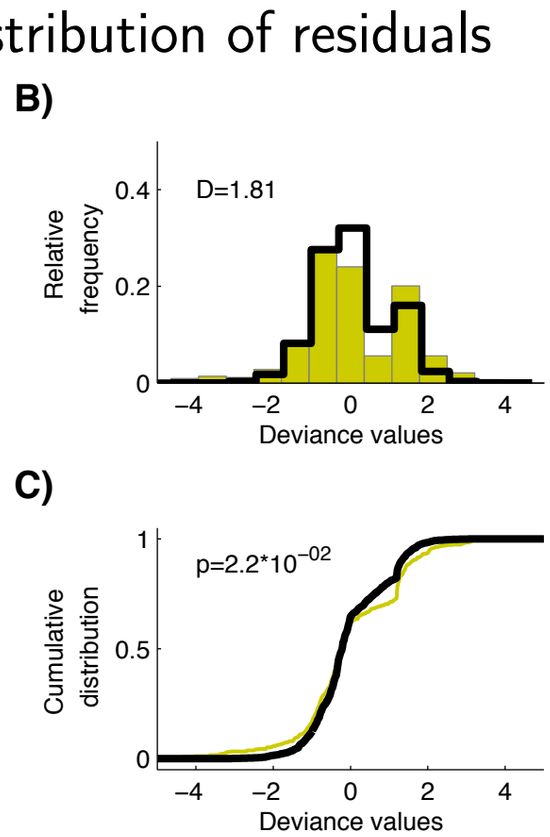
- ω, the normal to the decision hyperplane, is called the decision image.

- By modelling decision probabilities, we get additional information about the location of the boundary:
$$P(\text{female}|x) = g(f(x))$$

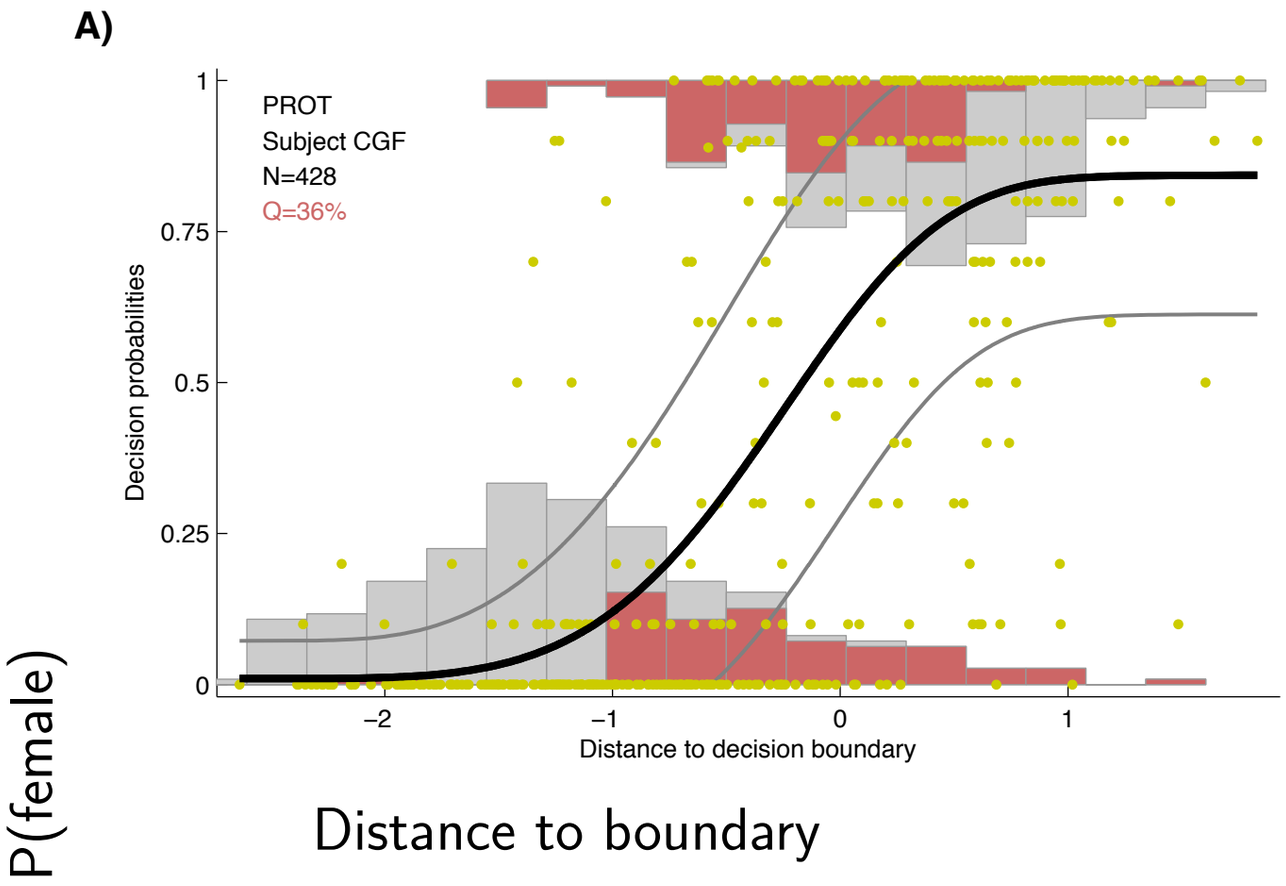
- ω is found by likelihood optimization: regularized logistic regression

Psychometric Function Along LogReg-ω

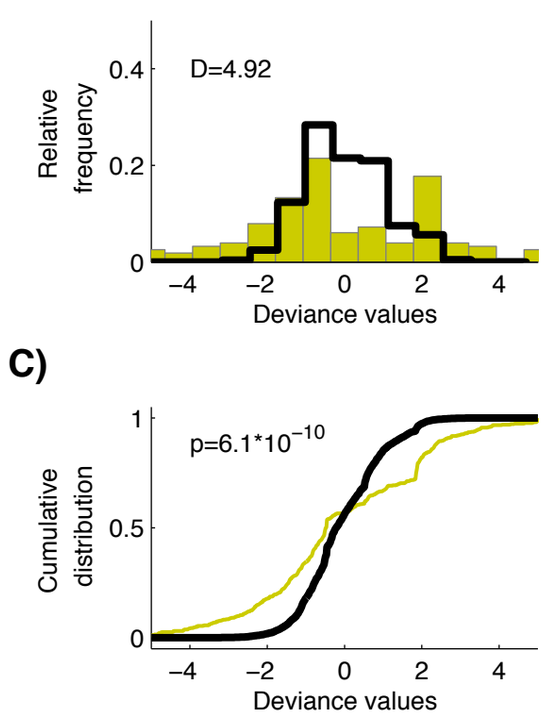




Psychometric Function Along Prototype-ω

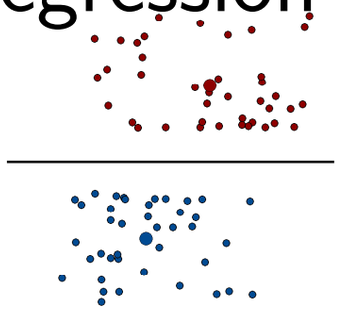


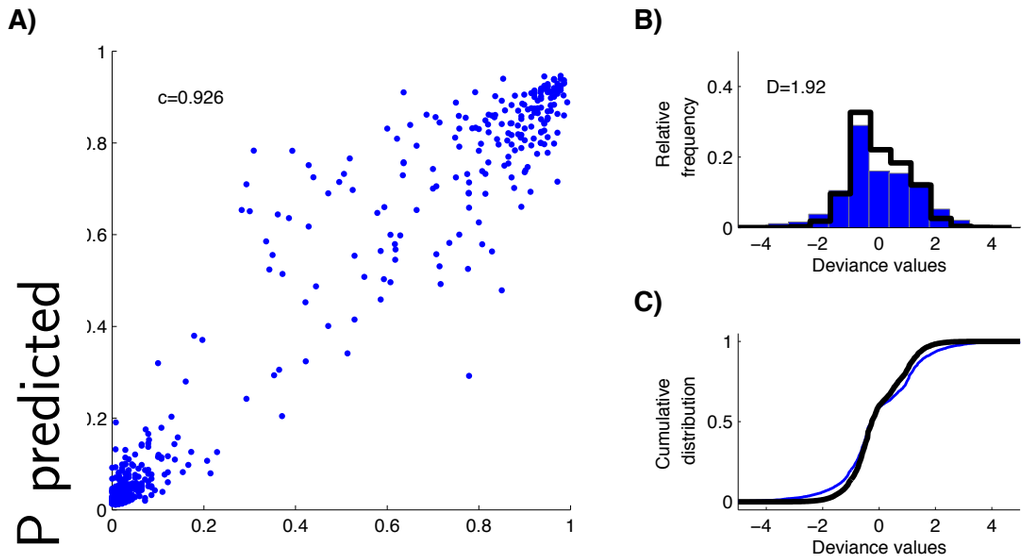
Distribution of residuals



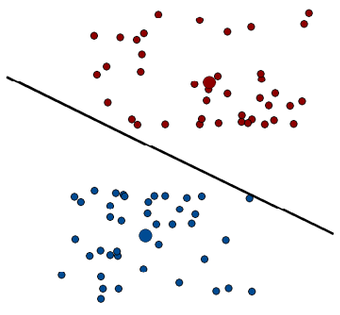
Summary Statistics across Observers

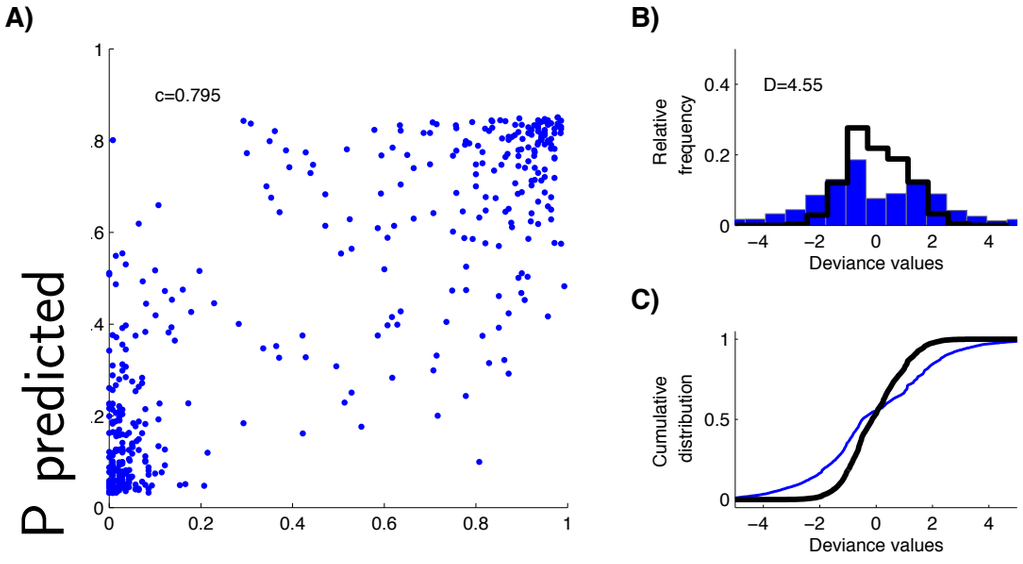
Logistic regression





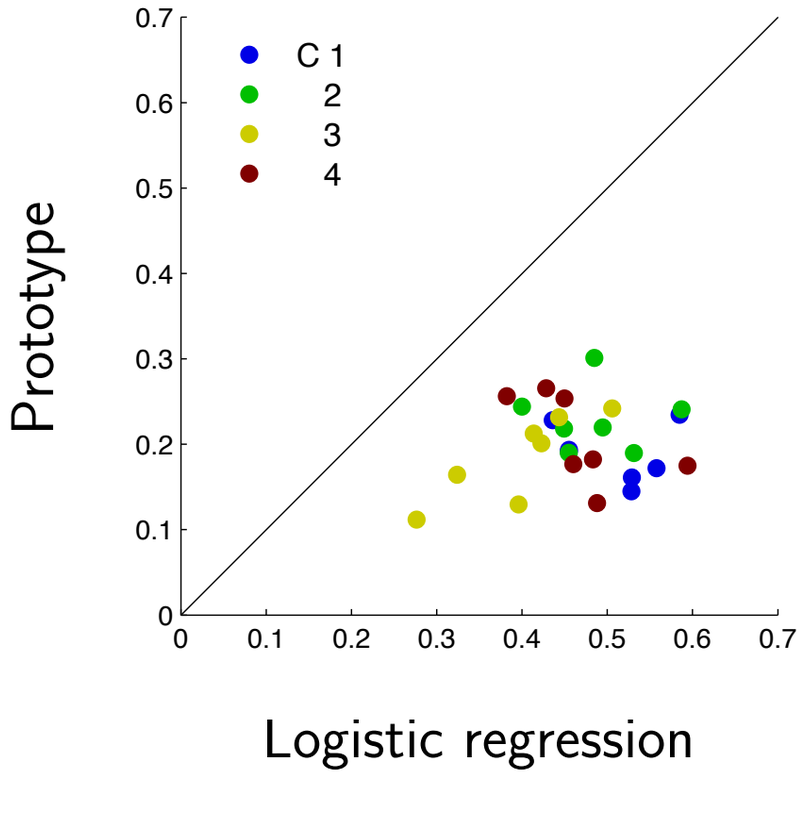
Prototype



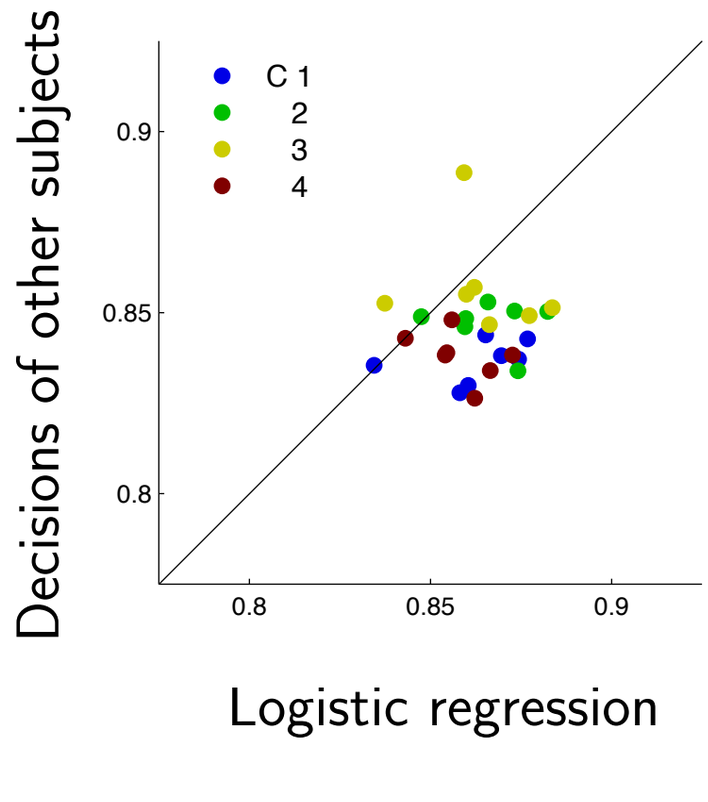


How Good is the Prediction?

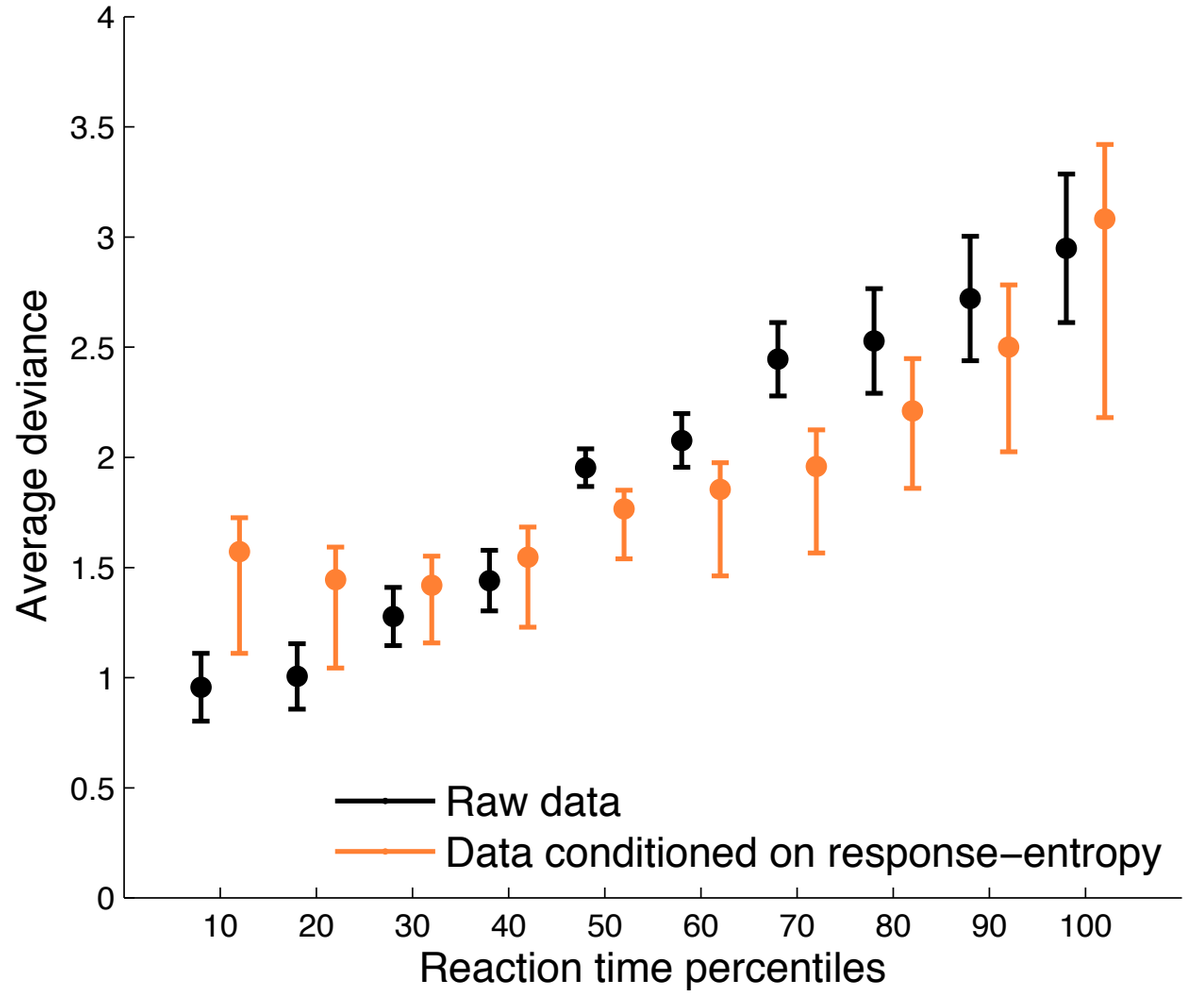
How good are predictions conditioned on real gender?



Are the algorithms sensitive to inter-observer differences?



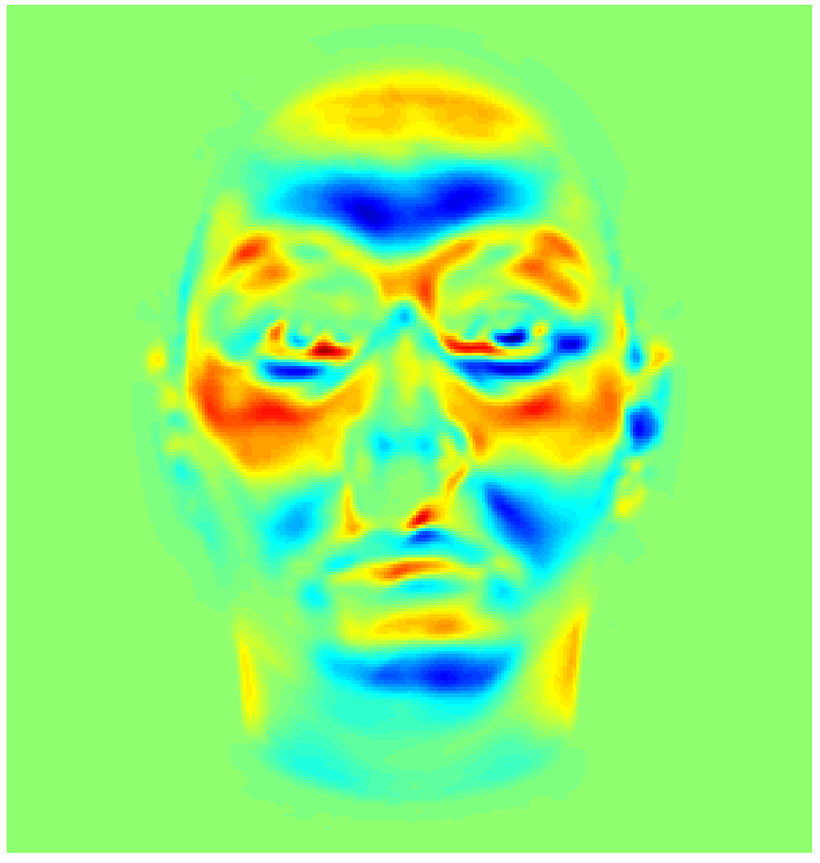
Predictability and Reaction Times



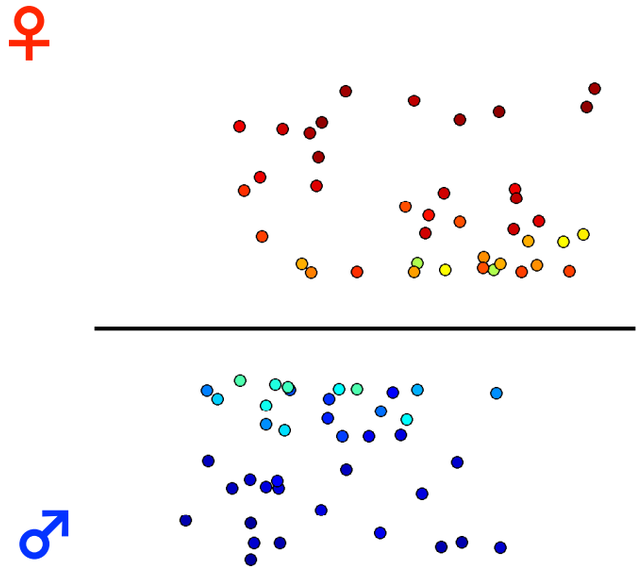
The Decision Images ω

Logistic regression

Prototype classifier



Evaluating Decision Images with Optimized Stimuli



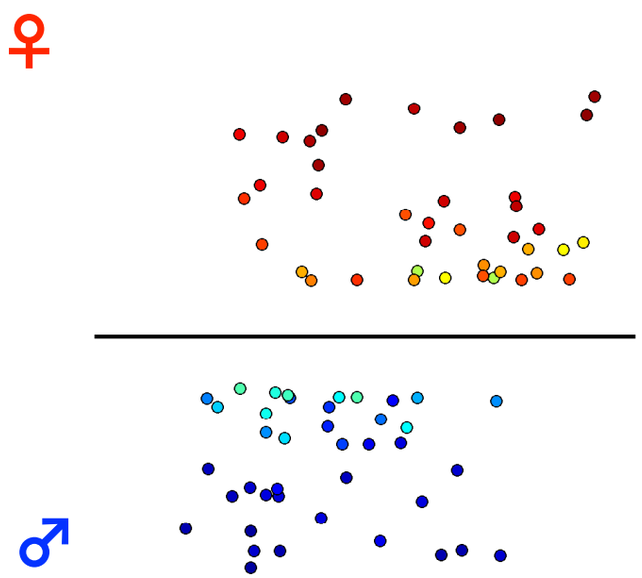
Decision probabilities

i. change quickly **orthogonal** to boundary.

ii. do not change **within** boundary.

- Recipe for generating optimized stimuli!

Evaluating Decision Images with Optimized Stimuli

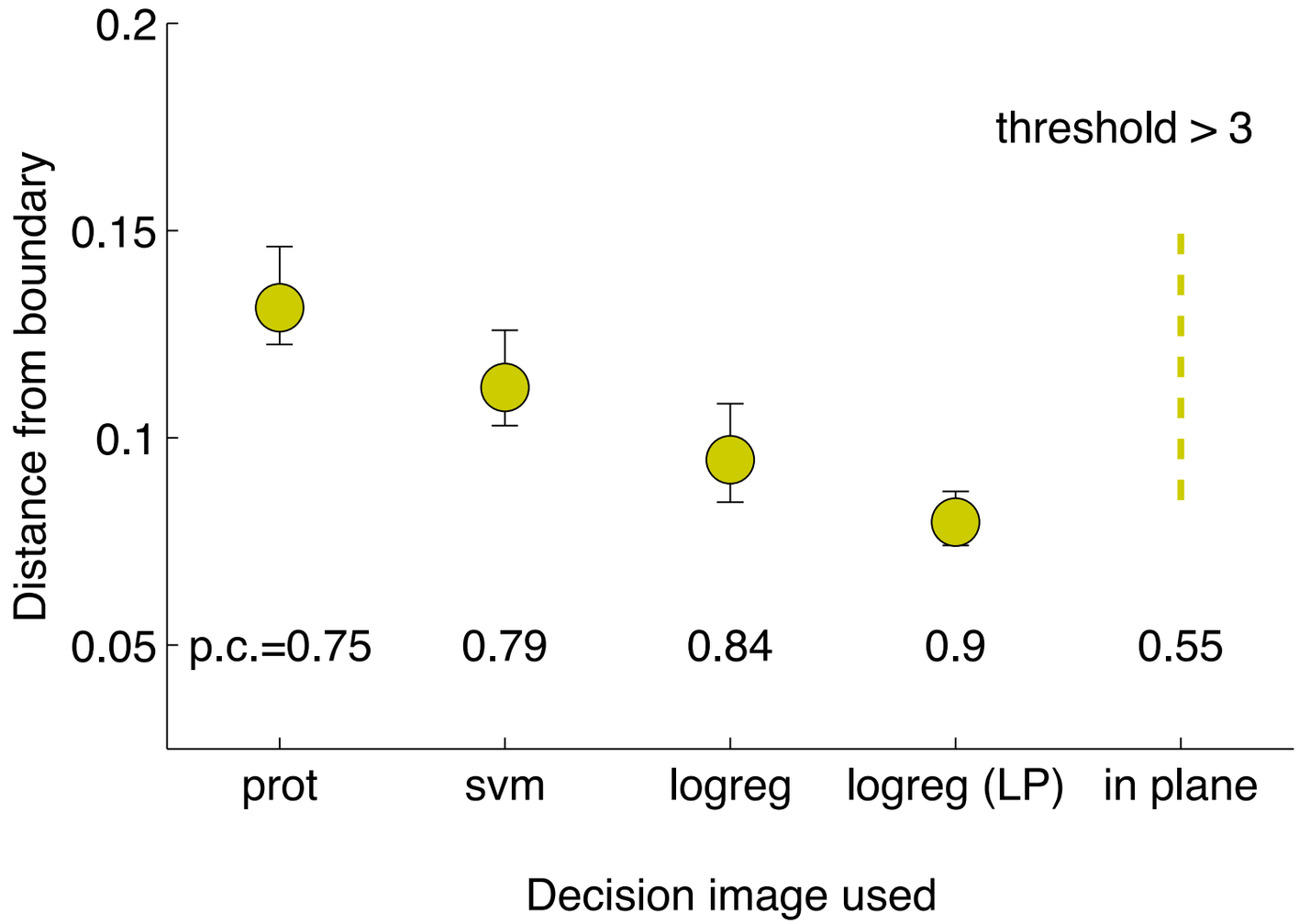


Decision probabilities

i. change quickly **orthogonal** to boundary.

ii. do not change **within** boundary.

- Recipe for generating optimized stimuli!



Interim Conclusions (1)

- Machine learning techniques (CV, regularization) can be used to fit predictive models with minimal assumptions about the stimulus statistics.

- It is possible to predict response probabilities in classification tasks from stimulus features using natural images.

- Need to test whether extracted features go beyond rediscovering the class-structure of the stimulus!

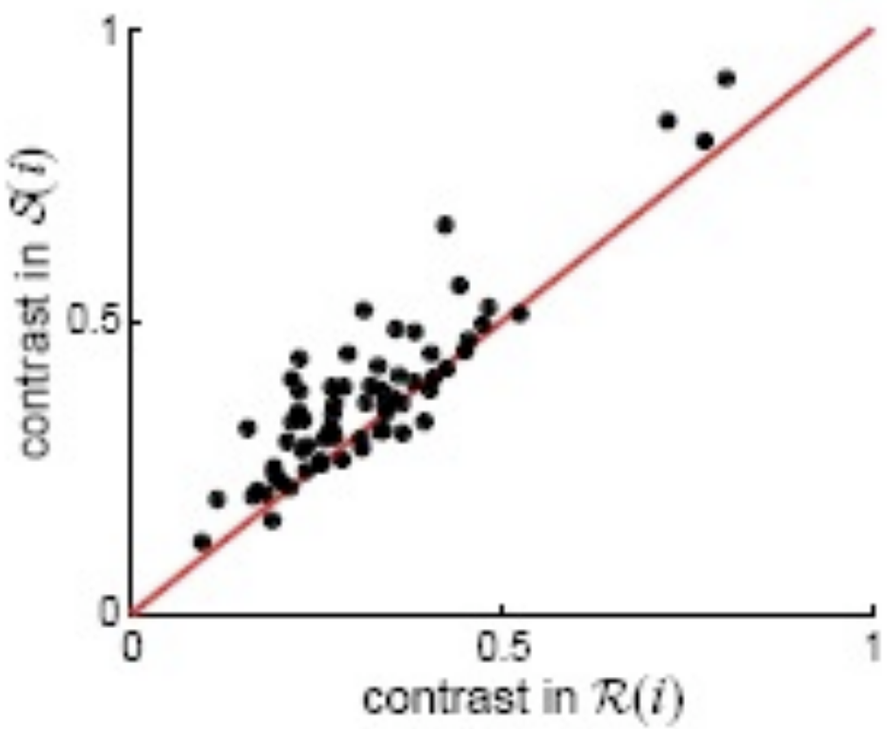
- The obtained *decision images* can be used to generated optimized stimuli for subsequent experiments.

- While the methods used here were linear, the approach can be extended to nonlinear decision images using *kernels*.

Scientific Question

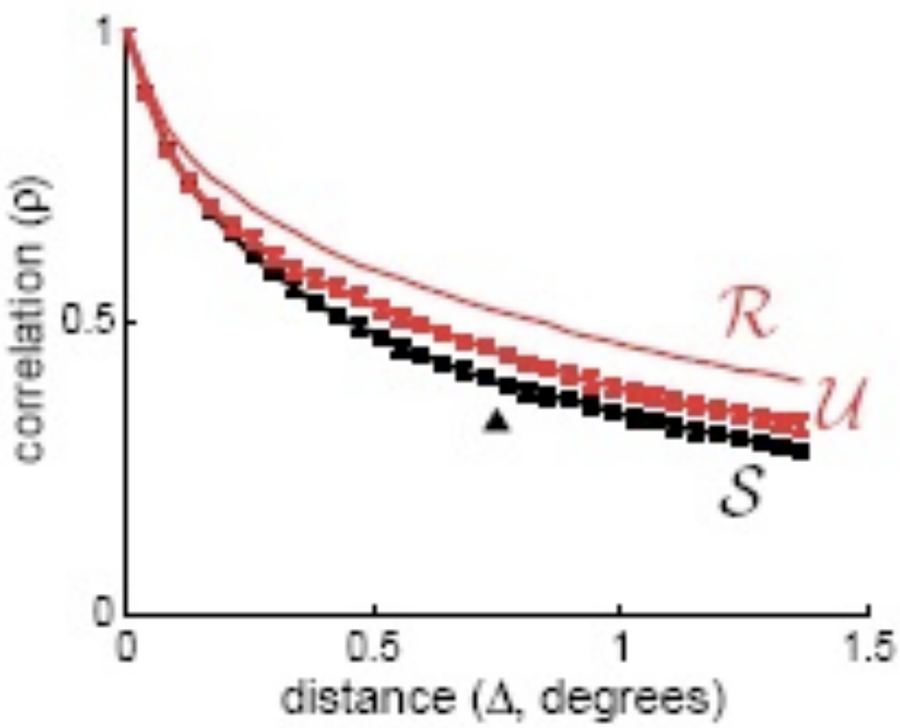
- What is special about the local image structure at fixation points?
- Does p(fixation) depend on local image statistics? (Bottom-up **visual saliency**)

Previous Work (1)



Correlation coefficient of RMS and model output: 0.69

B



Center pixel "more different" to surrounding pixels in fixation patches (Reinagel & Zador, 1998)

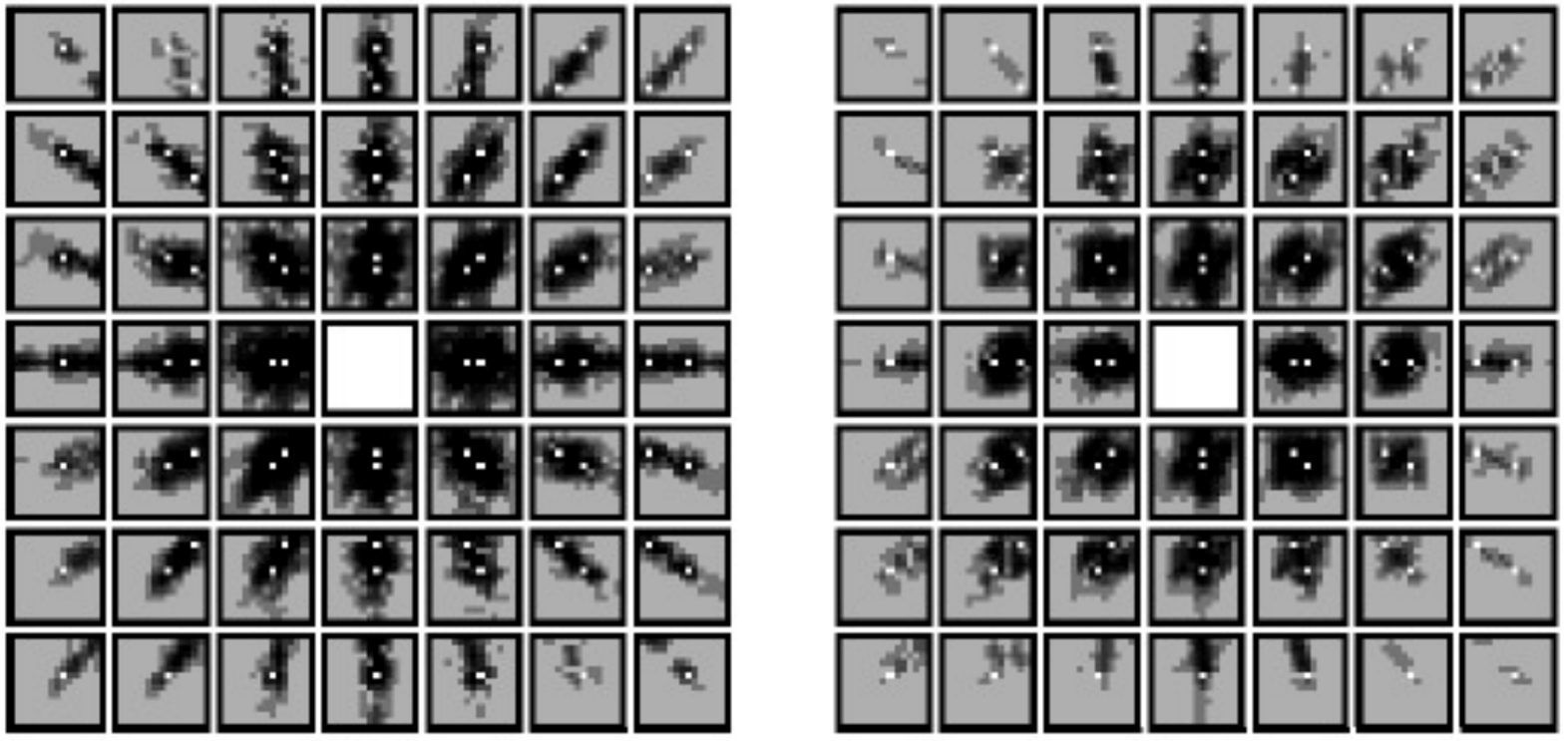
Previous Work (2)

- "The saccadic selection system avoids image regions which are dominated by a single oriented structure. Instead, it selects regions containing different orientations, like occlusions, corners, etc." (Krieger et al., 2001)

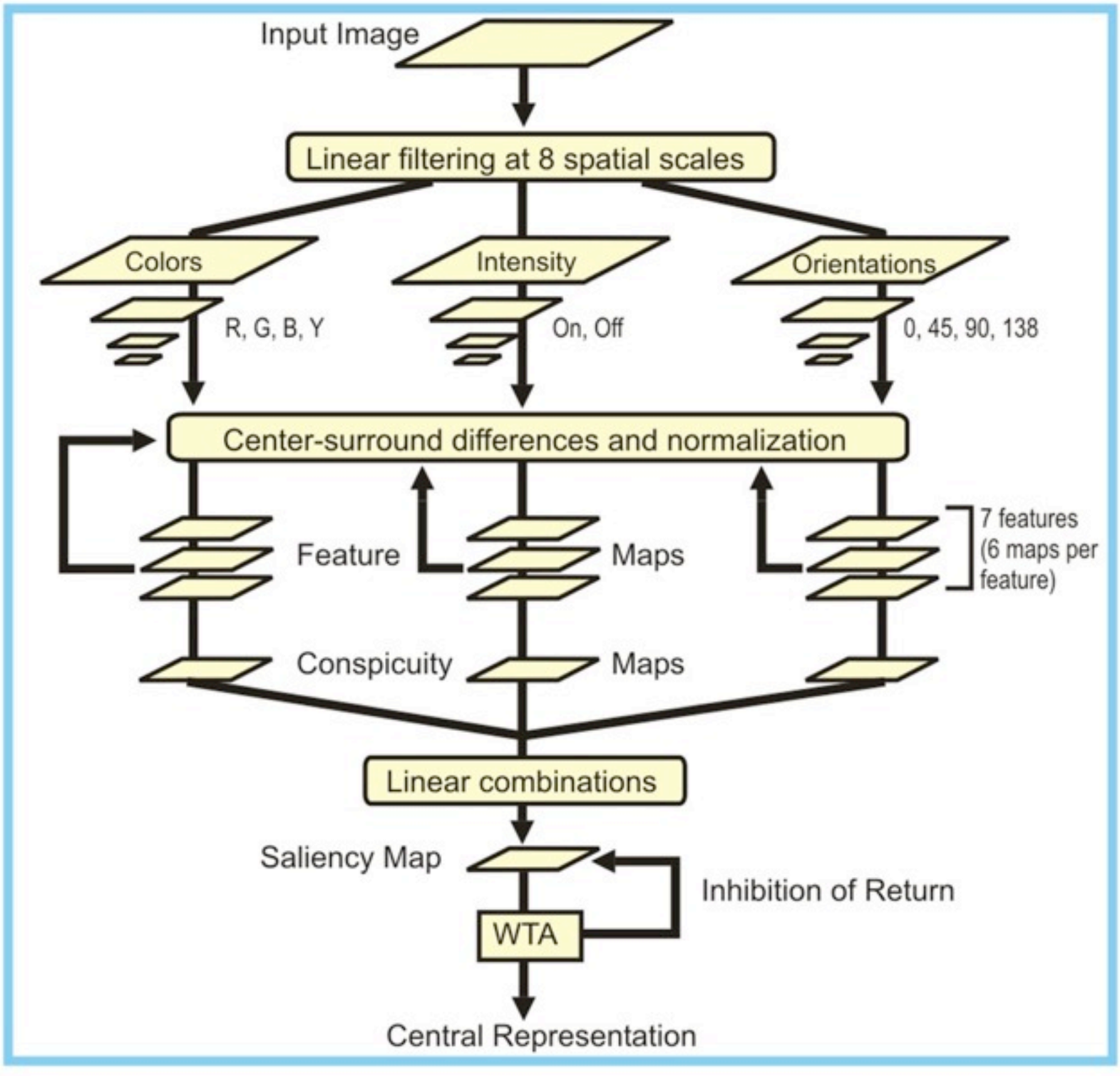
Third order statistics, "energy distribution is more circular":

$$C_3^{U_{rand}}(f_{x1}, f_{y1}, f_{x2}, f_{y2})$$

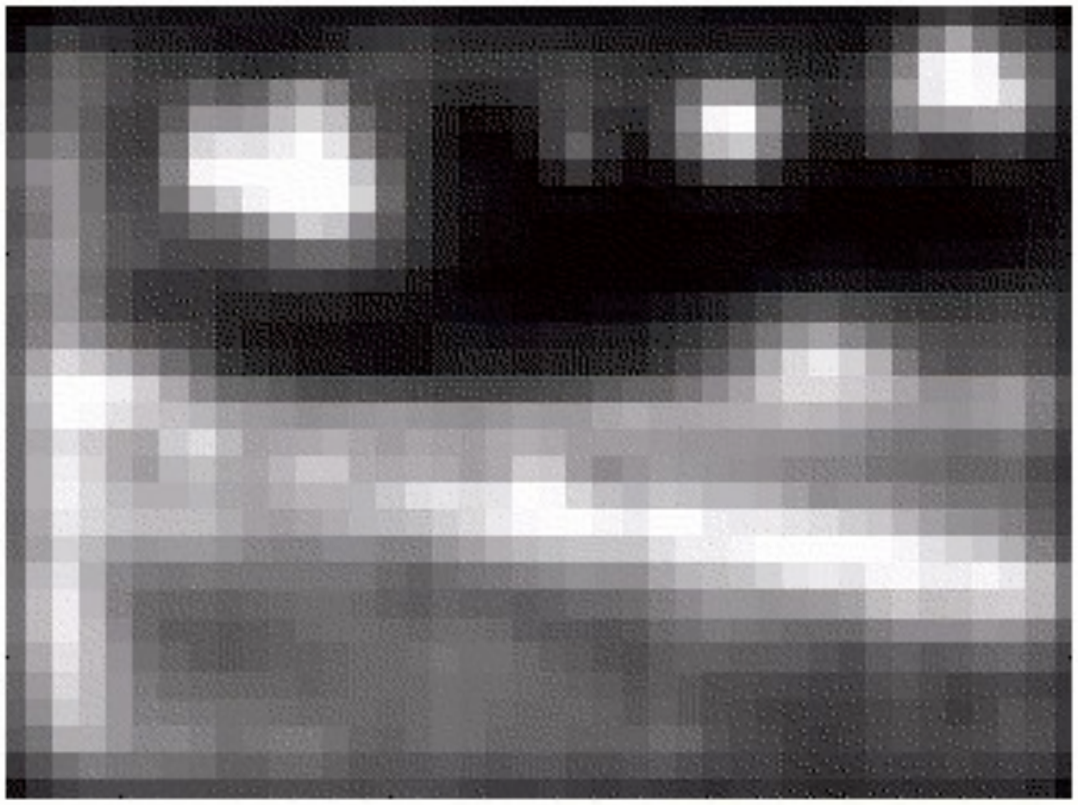
$$C_3^{U_{eye}}(f_{x1}, f_{y1}, f_{x2}, f_{y2})$$



Previous Work (3)



Saliency Maps



Machine Learning Approach

Previously: Top-Down modeling approach developing "biologically inspired" models built using "neurophysiological-hardware" like Gabor filters, ...

Many more or less ad-hoc choices have to be made, e.g. exact filter types, sizes, numbers, combination strategies, ...

Machine Learning approach: construct a model from the data, i.e. ...

i. Use a very general model class that does not "know" about the problem, but can adapt very well to a large class of problems.

ii. Numerically optimize (= learn) its parameters such that data is explained best.

Data Representation

For each data point (i=1...36,000), store local pixel values in a feature vector $\mathbf{x}_i$ and associate a label $\mathbf{y}_i$=1/-1 (fixation/bg)

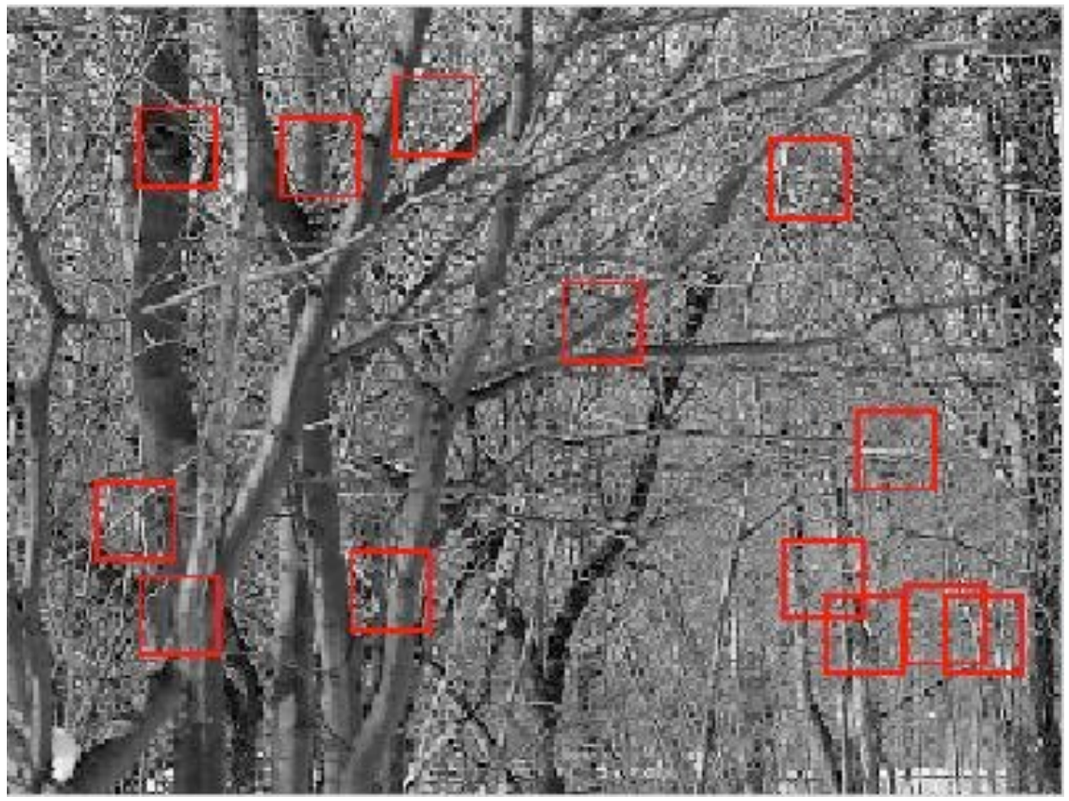
$\mathbf{x}_i$

$\mathbf{x}_{i+1}$

Background Examples

Generate background examples with same spatial distribution as fixations (Reinagel & Zador 1998).

Fixations



Background

Machine Learning Method (1)

Overall strategy: make the model class as general as possible

The model is a radial basis function (RBF) network with one basis function centered on each training example. ("Nonparametric" as its complexity grows with the number of data points.)

General? Universal approximation property, no preference for any image structure, no knowledge about shape or size of receptive fields.

Machine Learning Method (2)

We compute the weights (α_i) using hinge loss + L2-regularizer (= SVM)—finding α_i is convex, i.e. efficient and guaranteed to find the global optimum.

We use accuracy (0/1-loss function) to analyze our results—every misclassified image patch gets error "1".

We find the *design parameters* λ, γ, and patch size d via exhaustive grid-search, using cross-validation estimates of accuracy—feasible, as problem only 3D (and you have access to Bernhard Schölkopf's MPI Compute Cluster in Tübingen!).

Radial-Basis-Function Support Vector Machine (RBF-SVM)

Kernel bandwidth

$$f(\mathbf{x}) = \sum_{i=1}^{m} \alpha_i \exp\left(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2\right)$$

Weights

Patch size: d

$$\lambda \|f\|^2 + \sum_{i=1}^{m} \max(0, 1 - y_i f(\mathbf{x}_i))$$

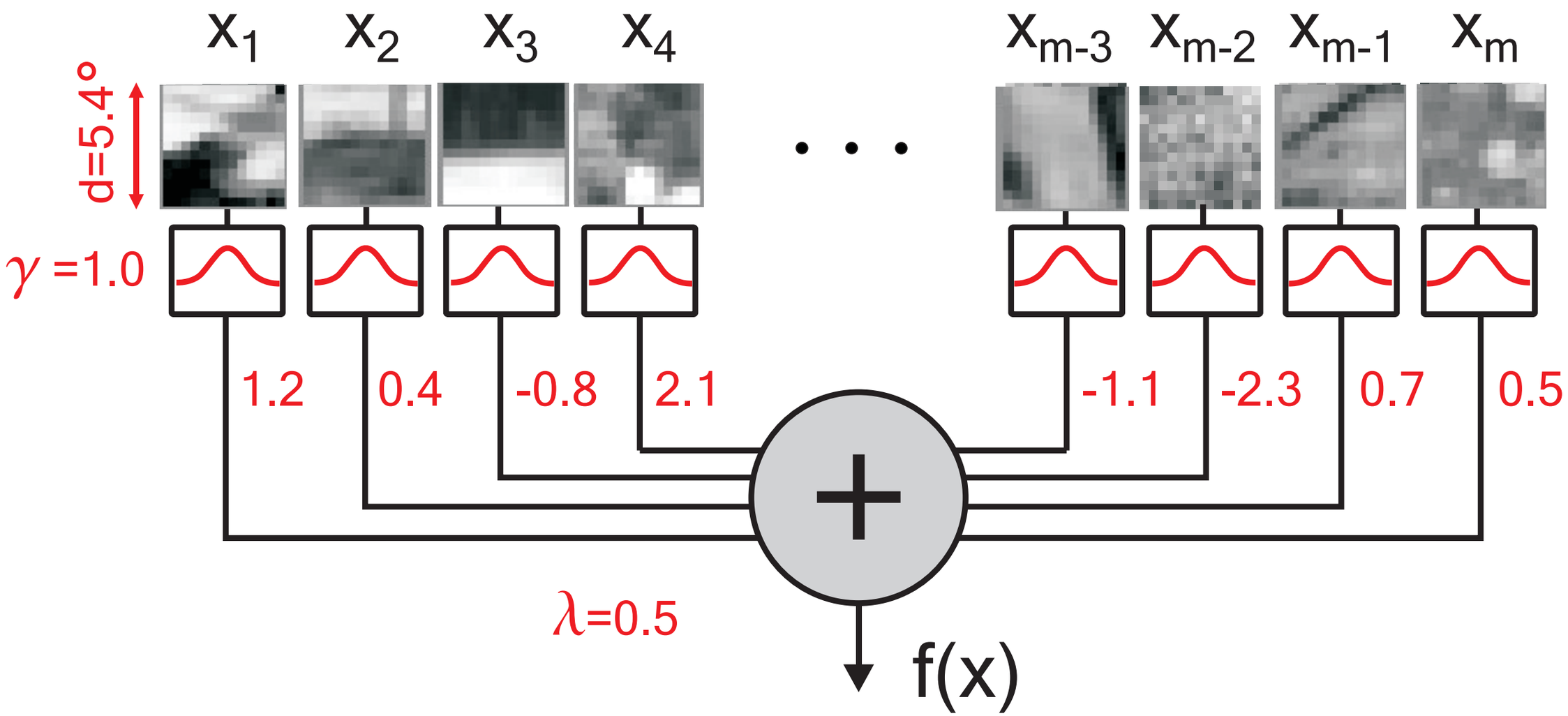
Smoothness

>24,000 weights

and

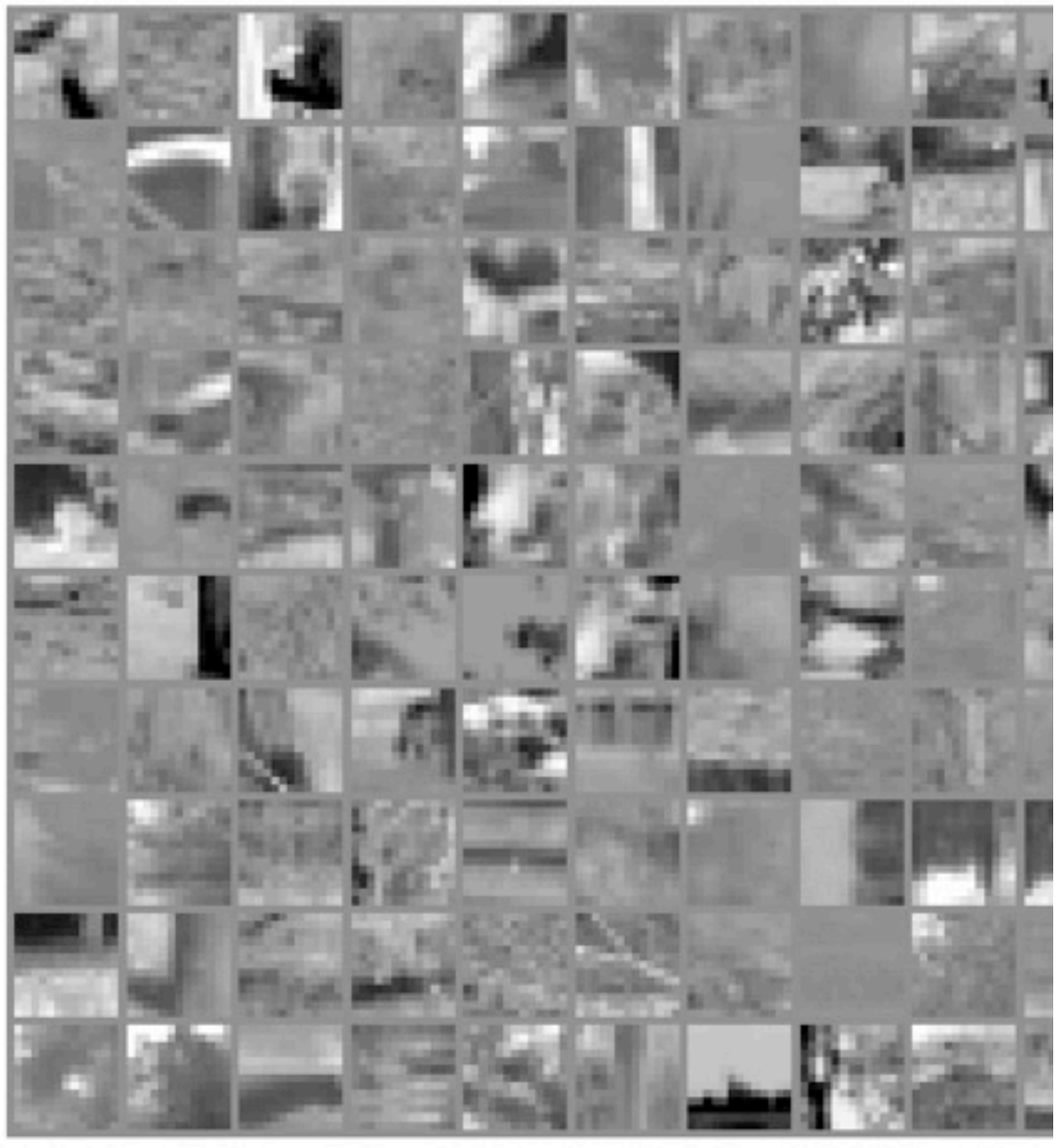
3 design parameters

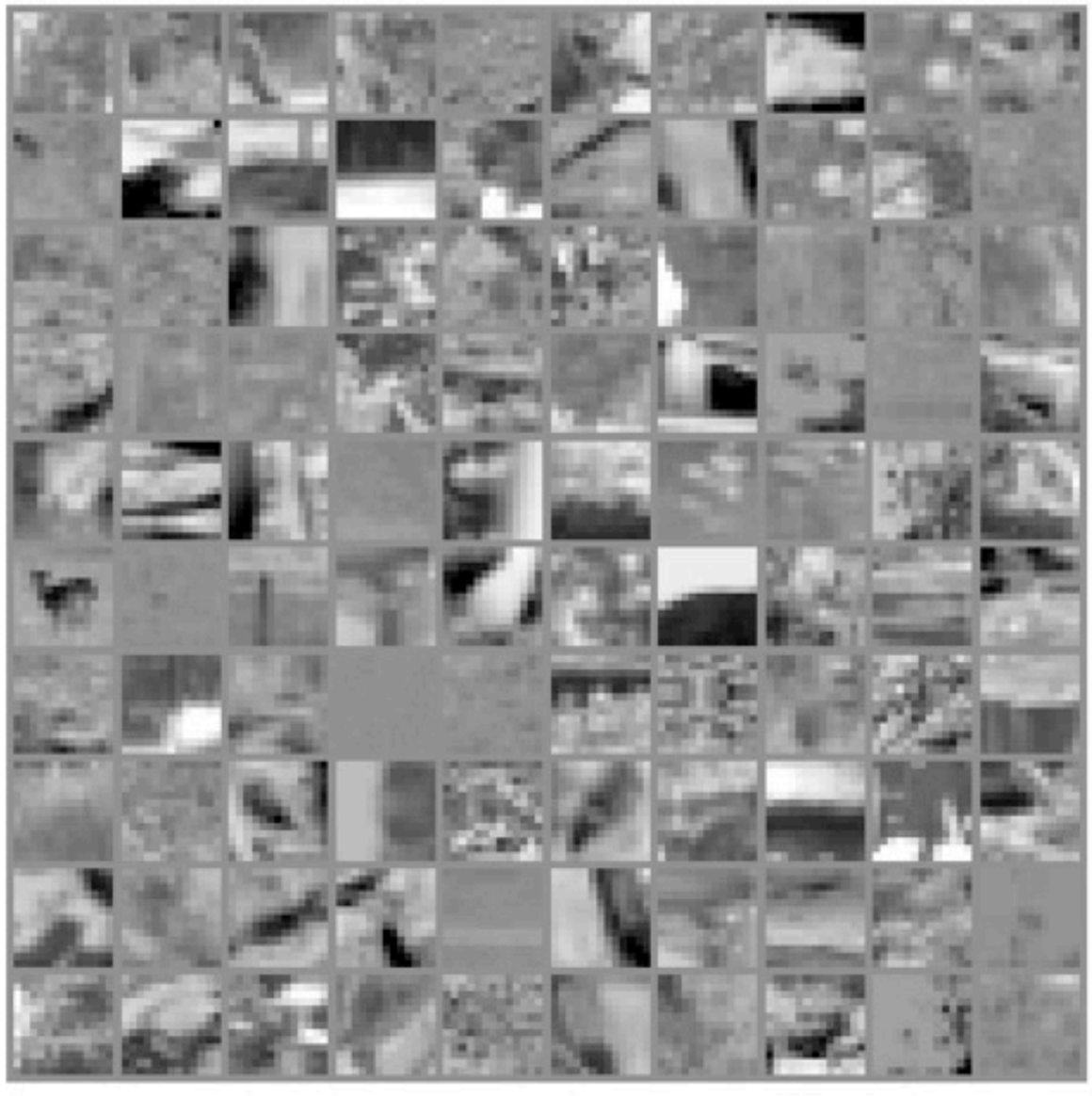
RBF-SVM after Optimization ("Learning")



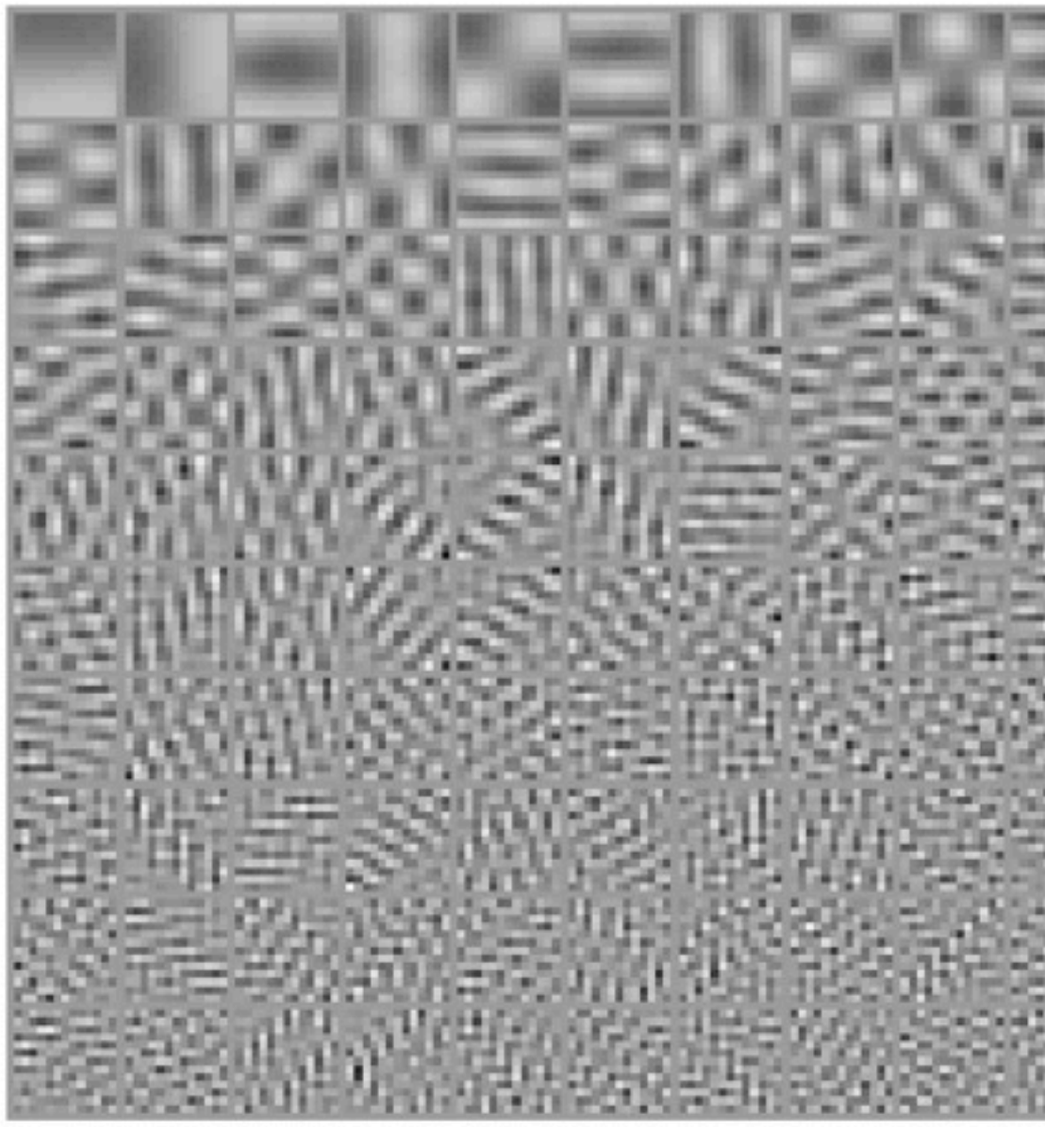
Predicitivity (area under ROC): 0.64±0.010

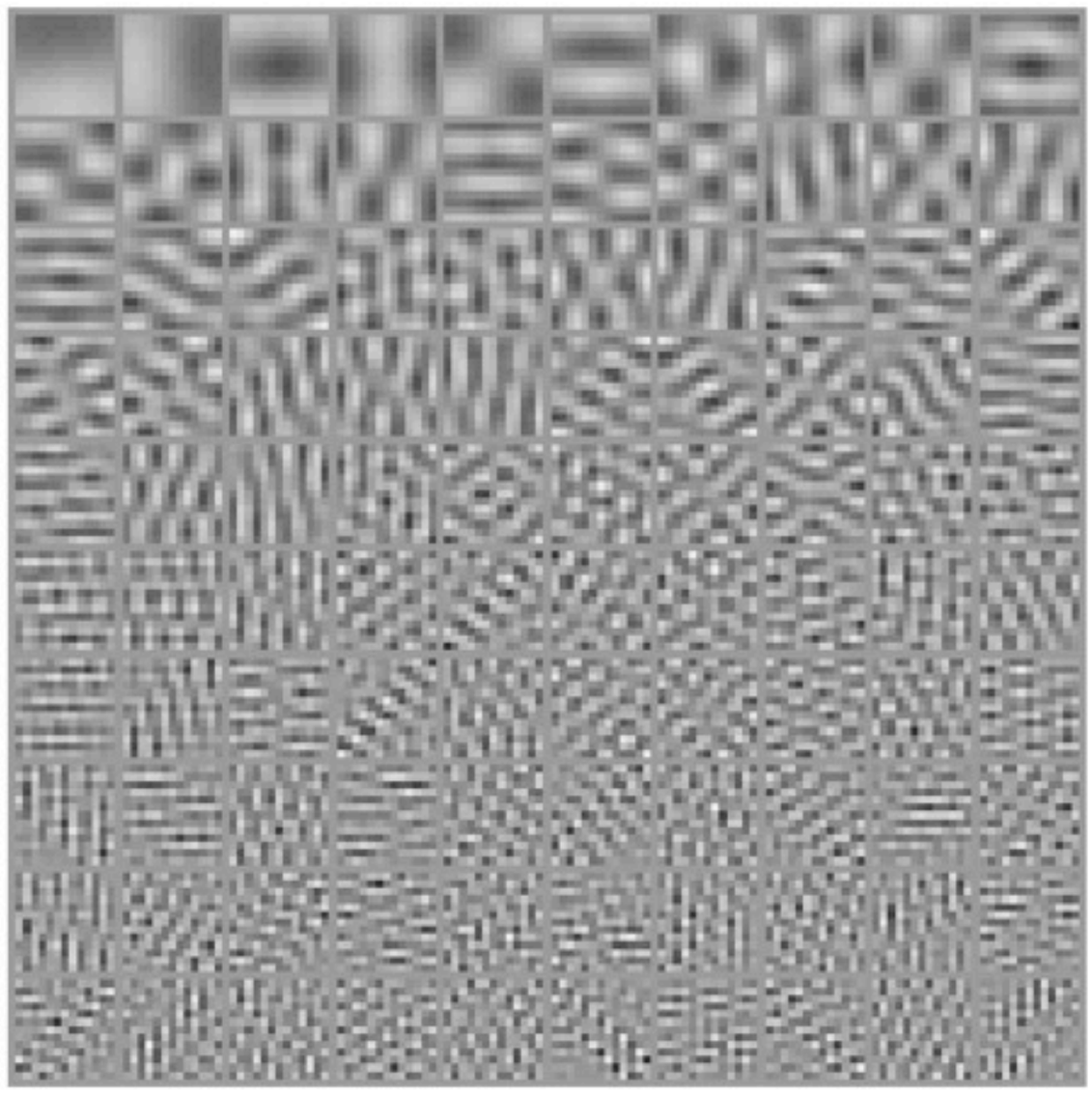
Randomly Selected vs. Fixated Image Patches



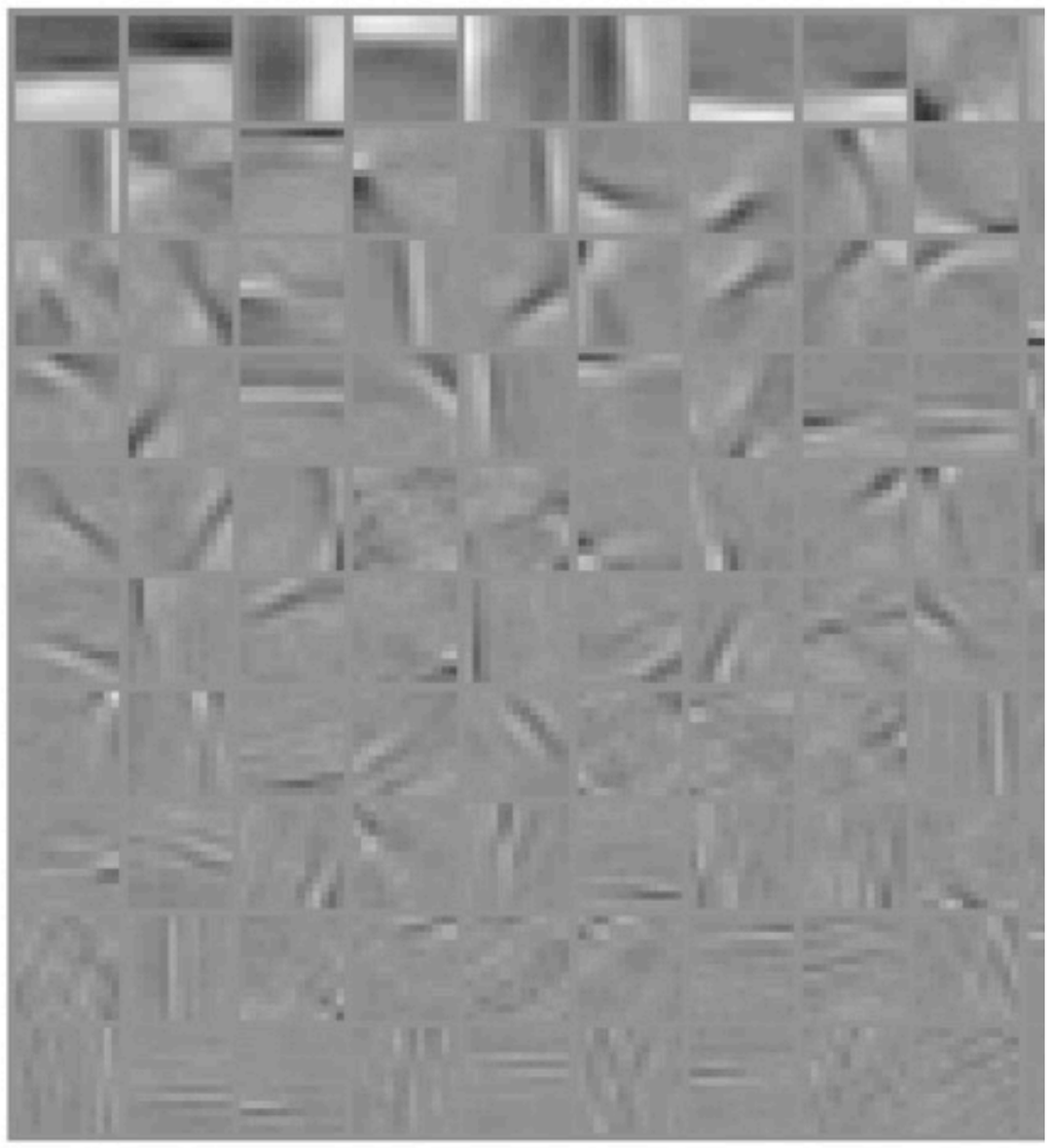


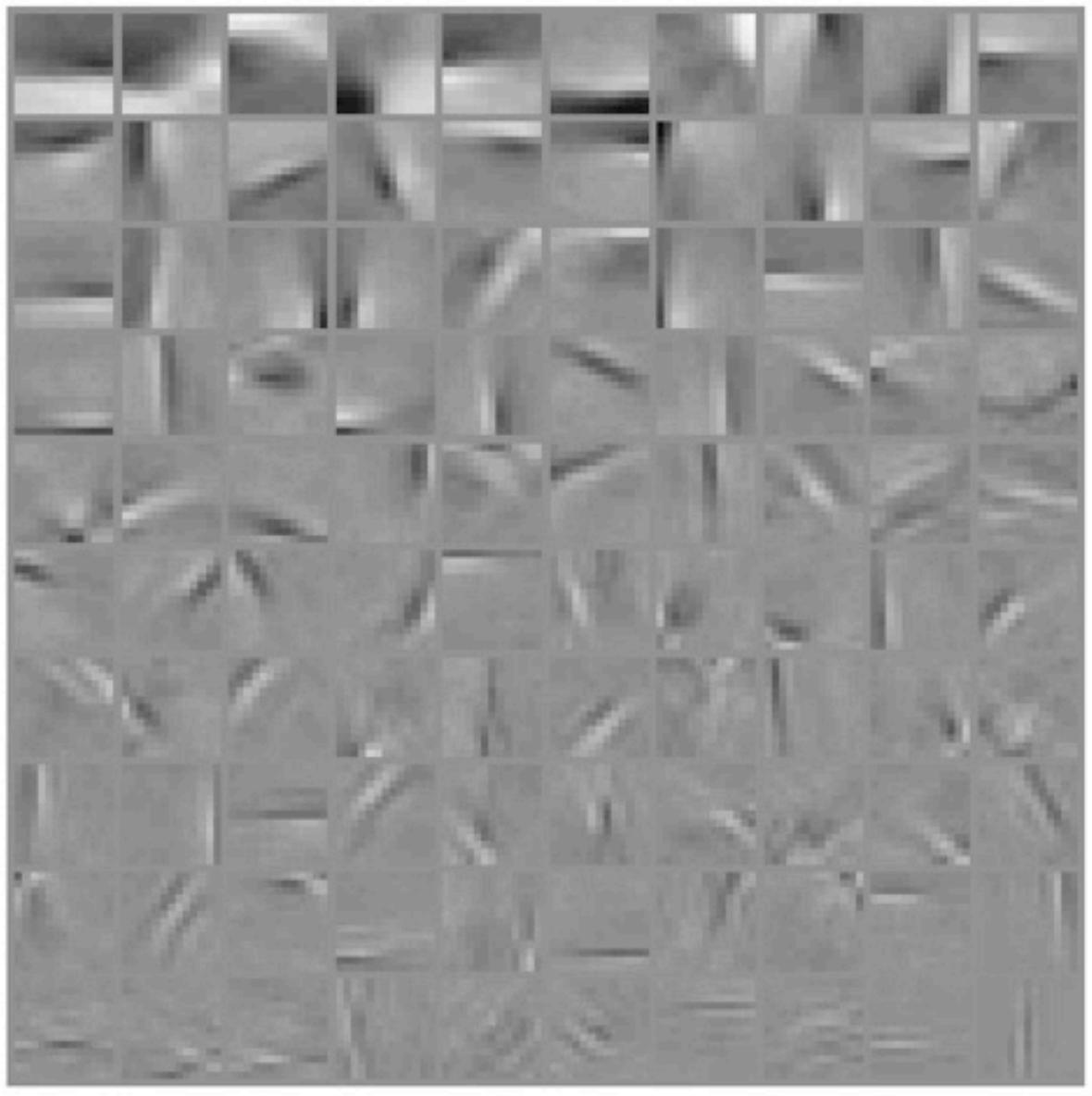
Randomly Selected vs. Fixated Patches: PCA Basis



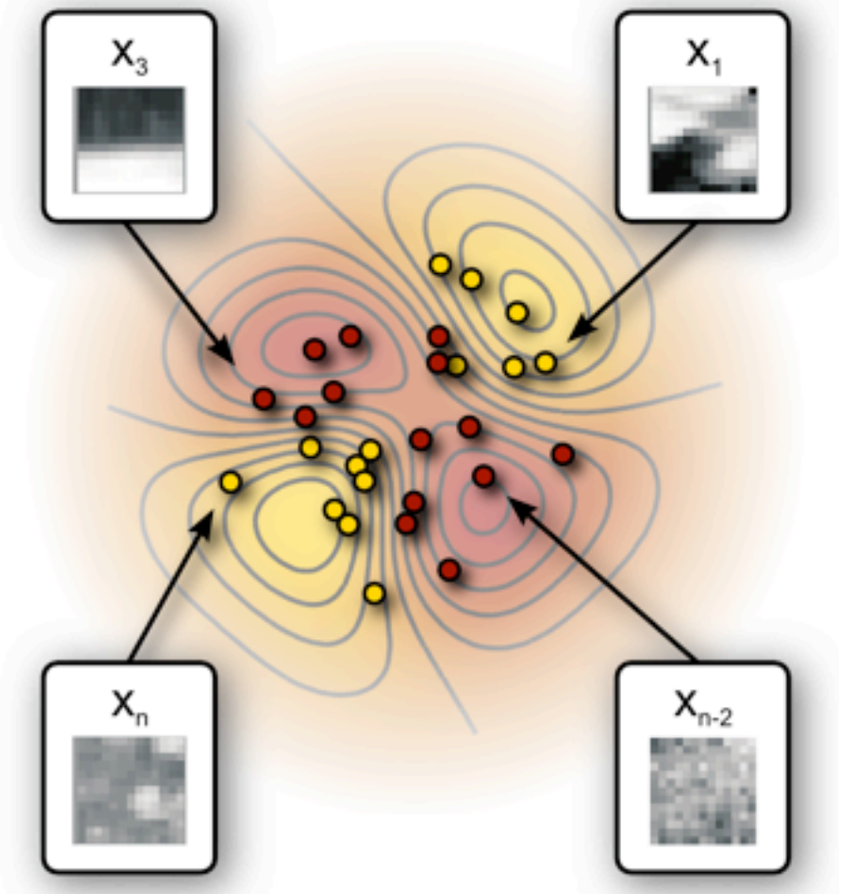


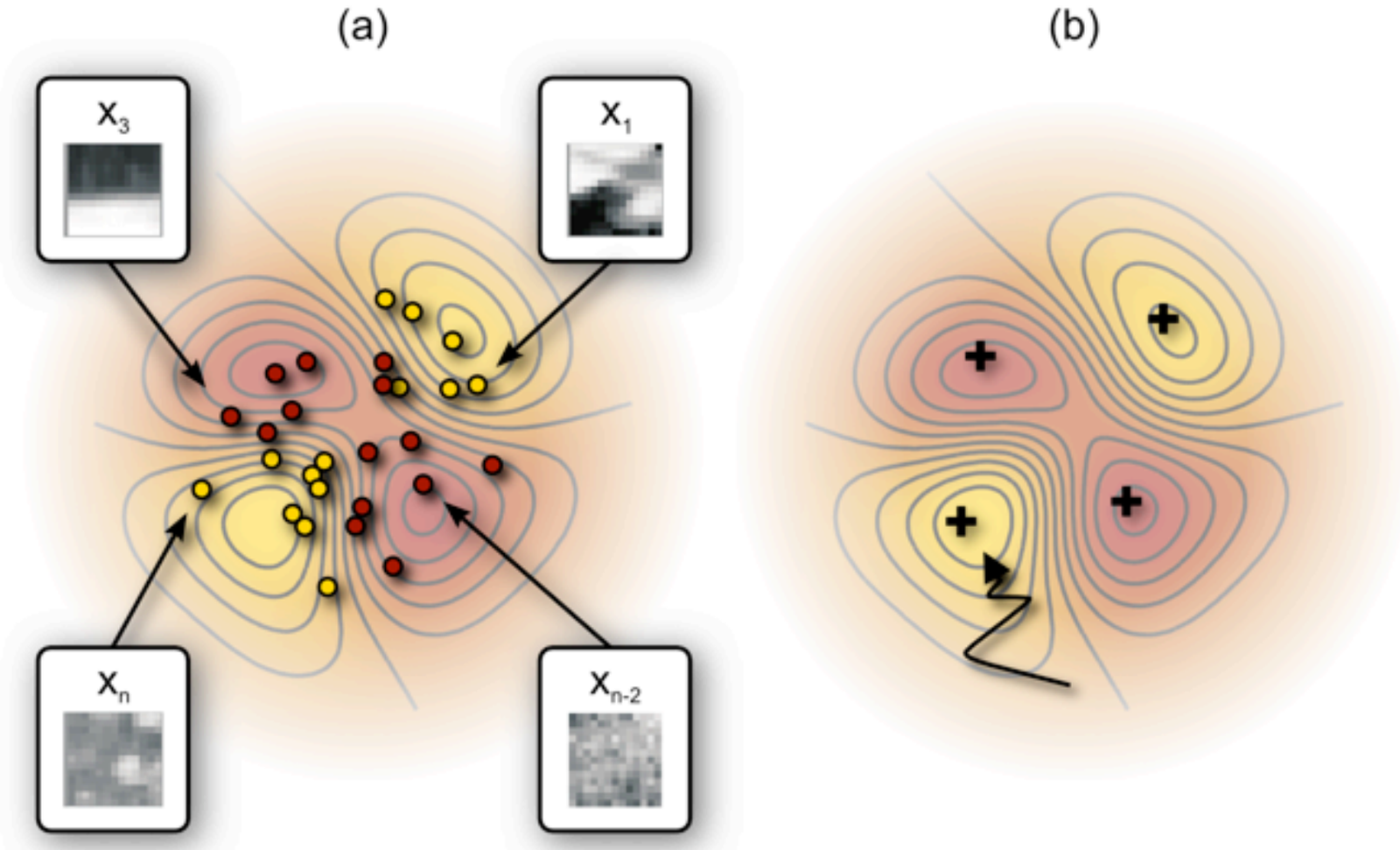
Randomly Selected vs. Fixated Patches: ICA Basis

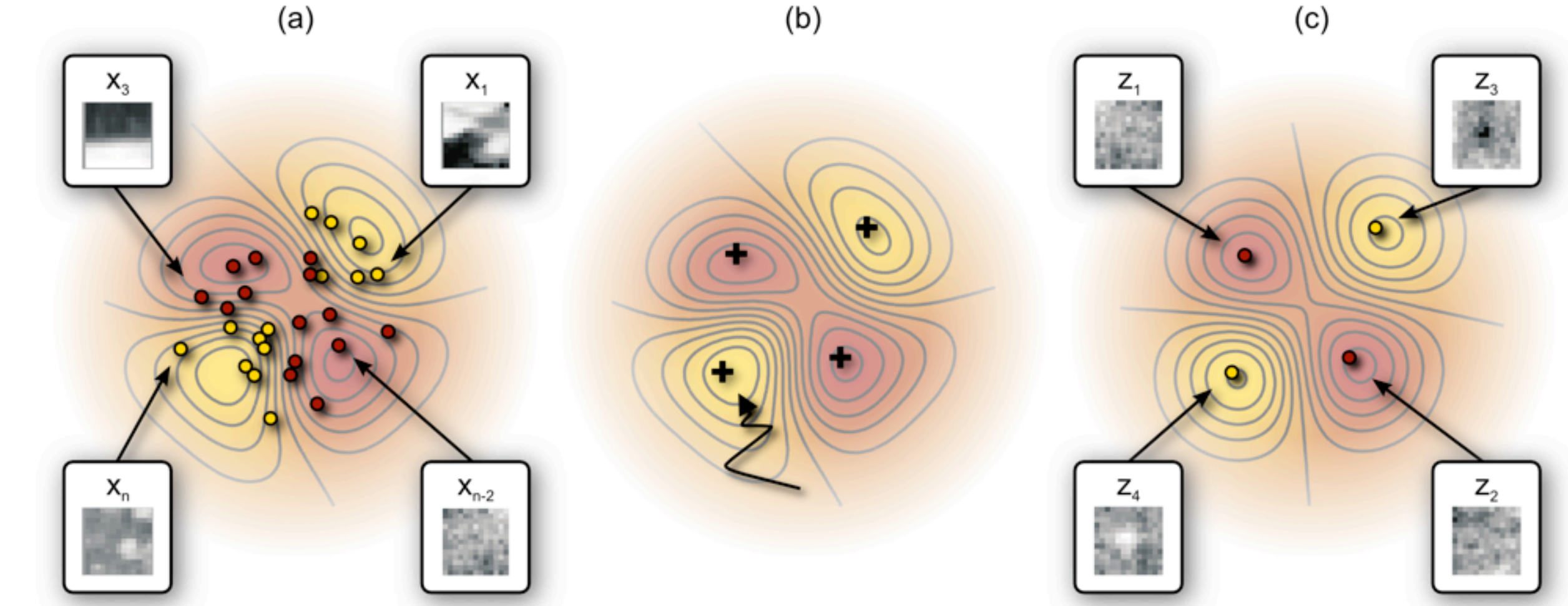




(a)




(a)
(b)
x_3
x_1
x_n
x_{n-2}


(a)
(b)
(c)
x_3
x_1
x_n
x_{n-2}
z_1
z_3
z_4
z_2

(a)

(b)

(c)

x_3

x_1

x_n

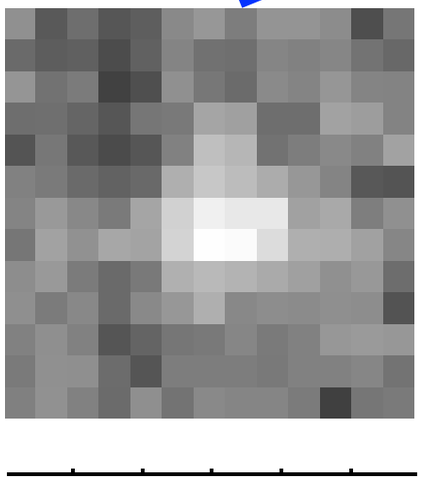
x_{n-2}

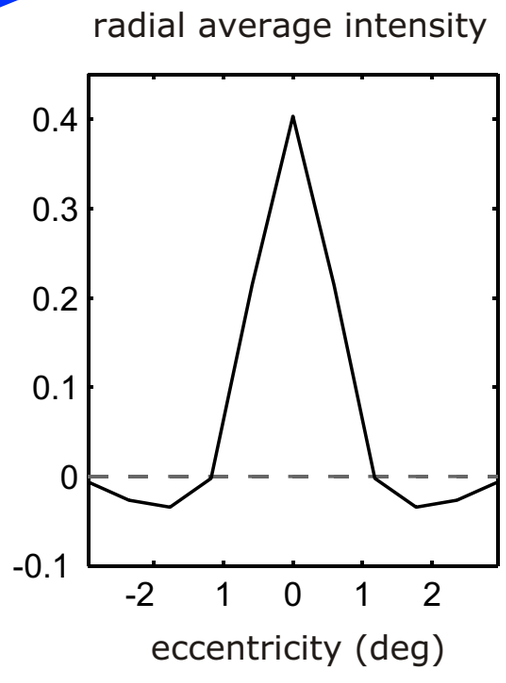
z_1

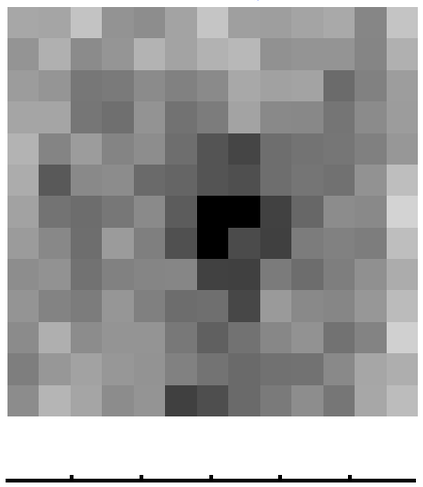
z_3

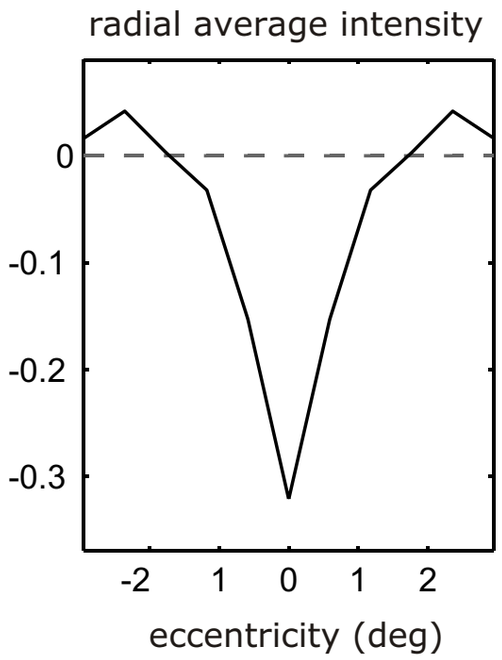
z_4

z_2





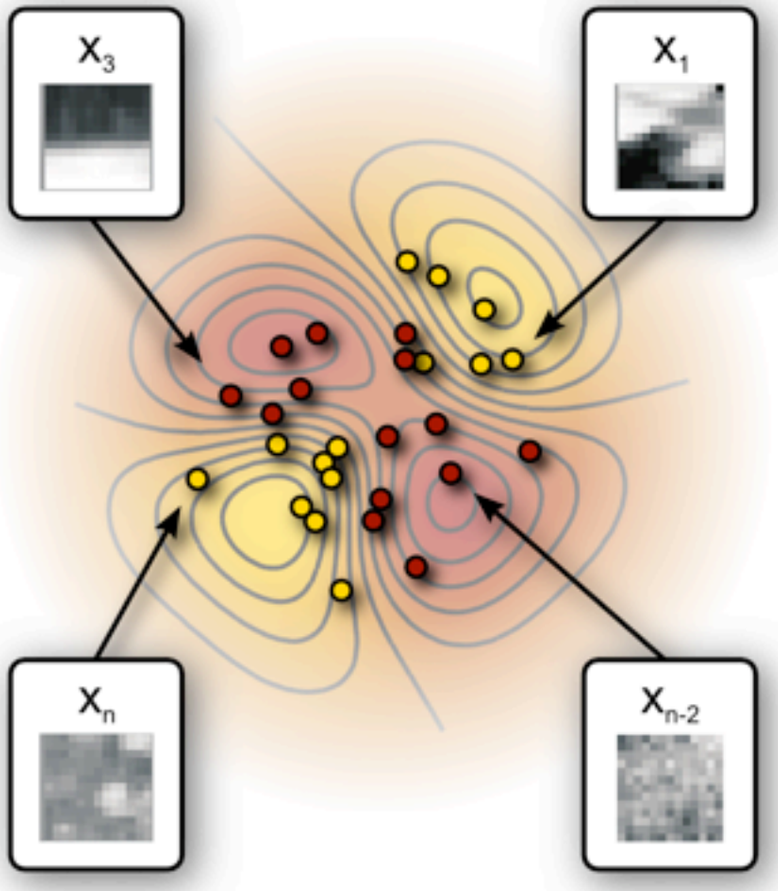


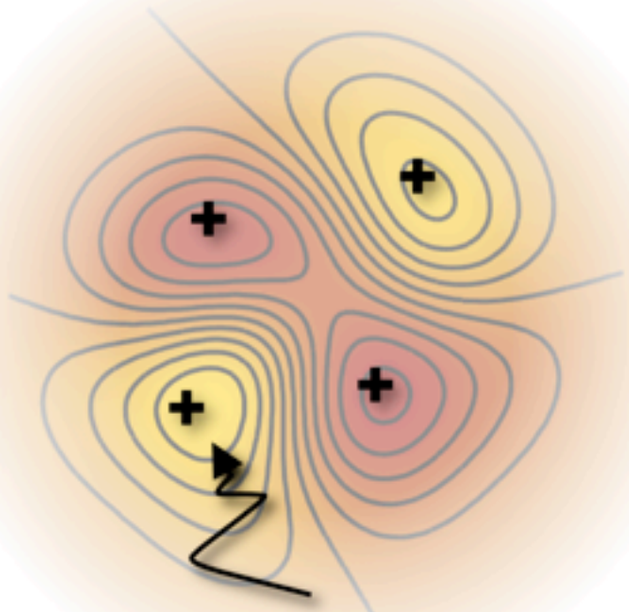


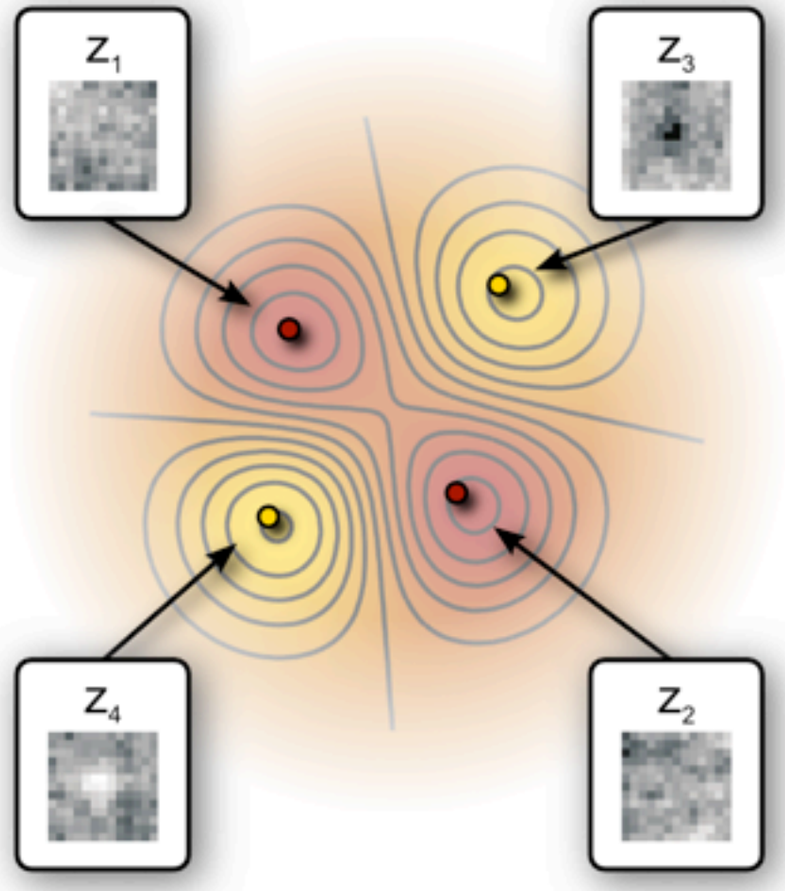
(a)

(b)

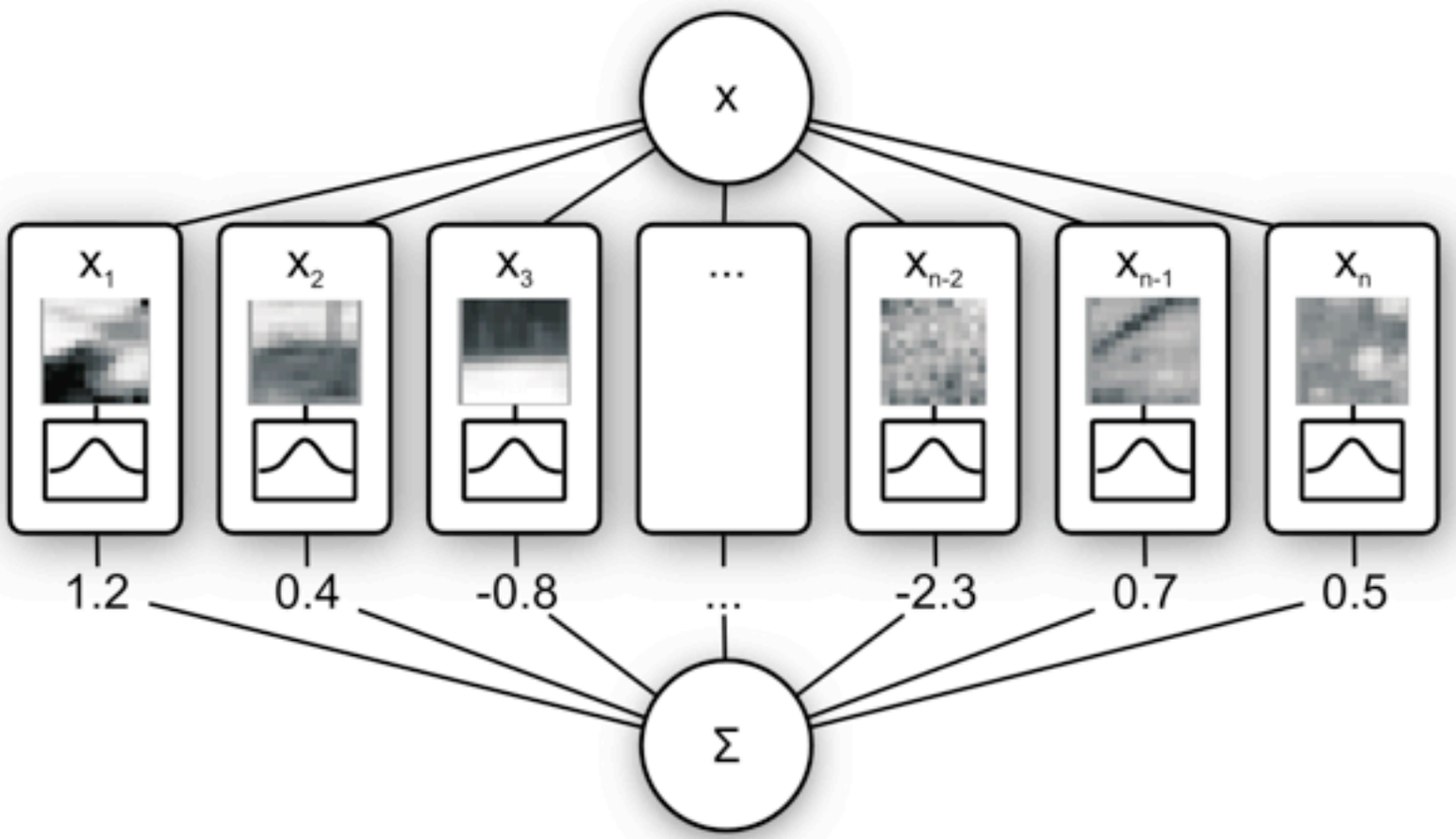
(c)


x_3
x_1
x_n
x_{n-2}

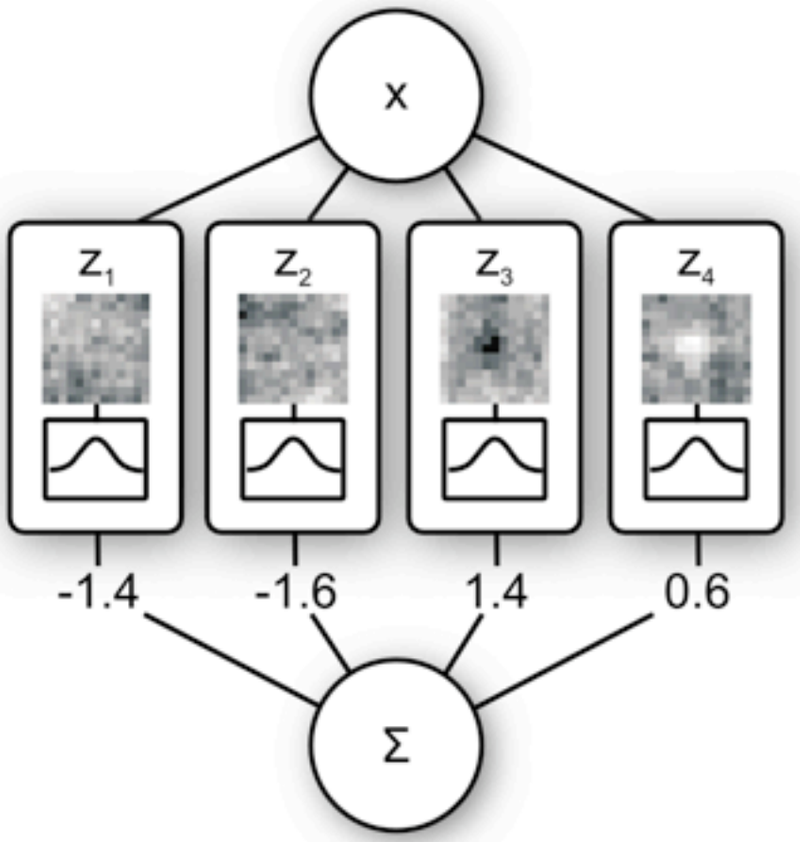



z_1
z_3
z_4
z_2

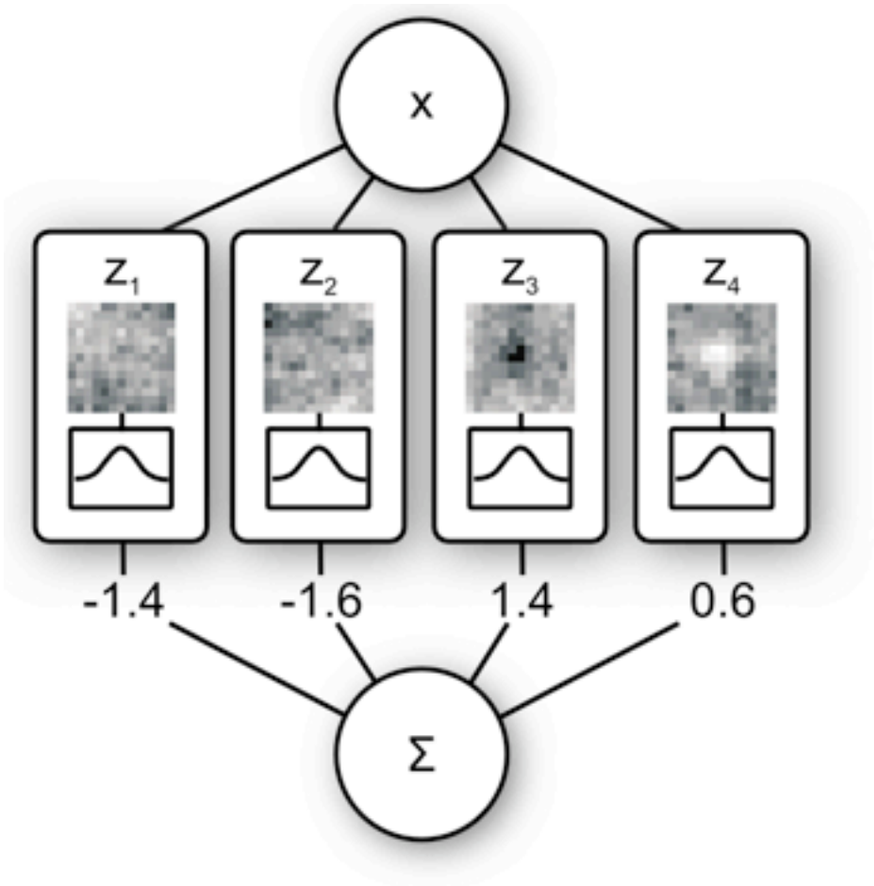
(d)

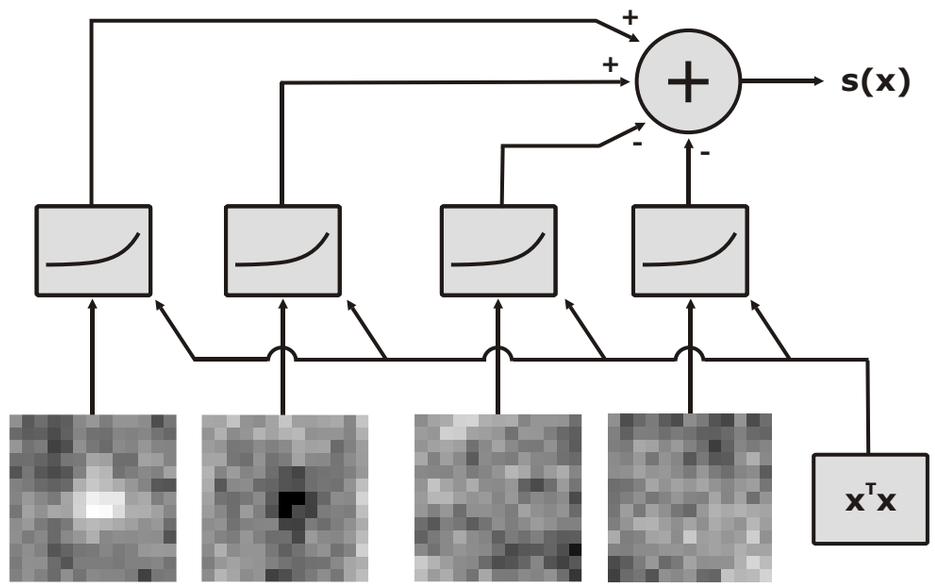

x
x_1
x_2
x_3
...
x_{n-2}
x_{n-1}
x_n
1.2
0.4
-0.8
...
-2.3
0.7
0.5
Σ

≈

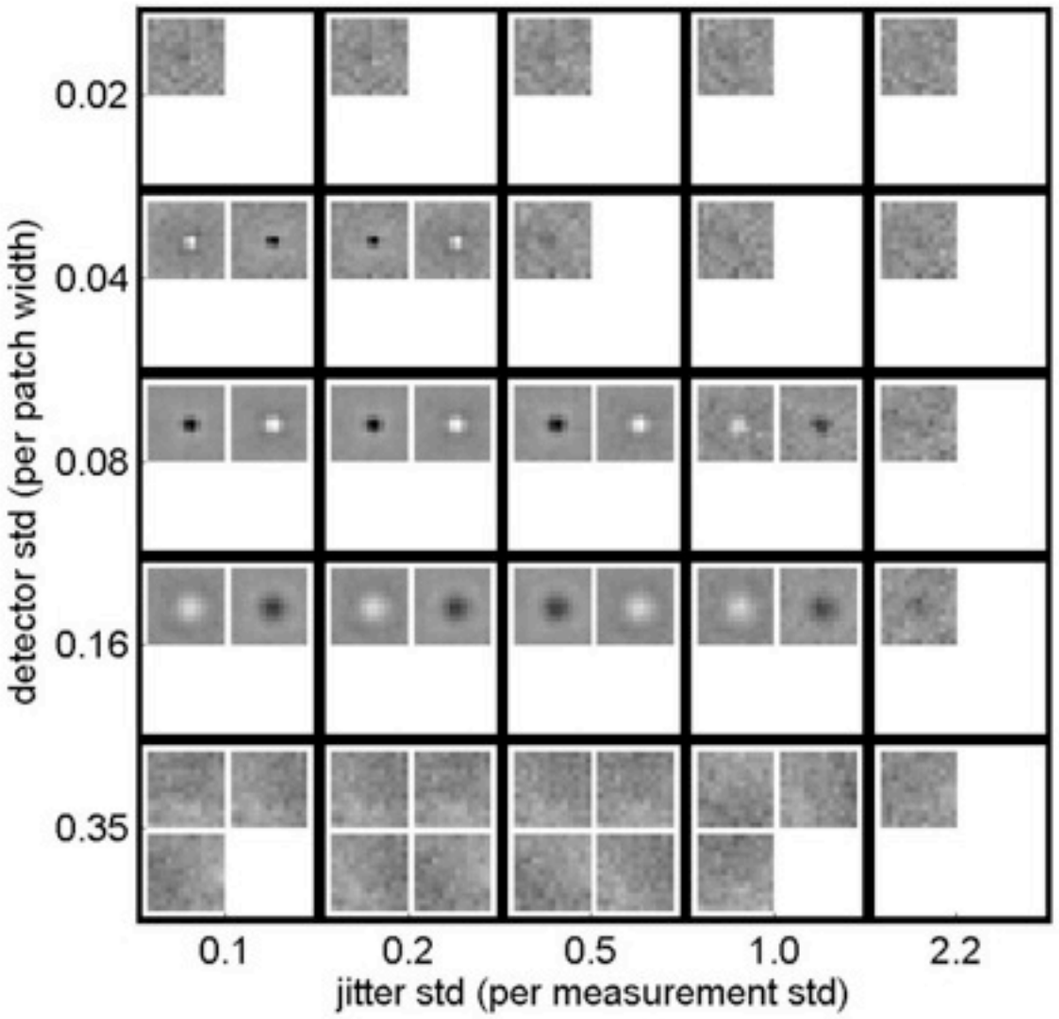

x
z_1
z_2
z_3
z_4
-1.4
-1.6
1.4
0.6
Σ

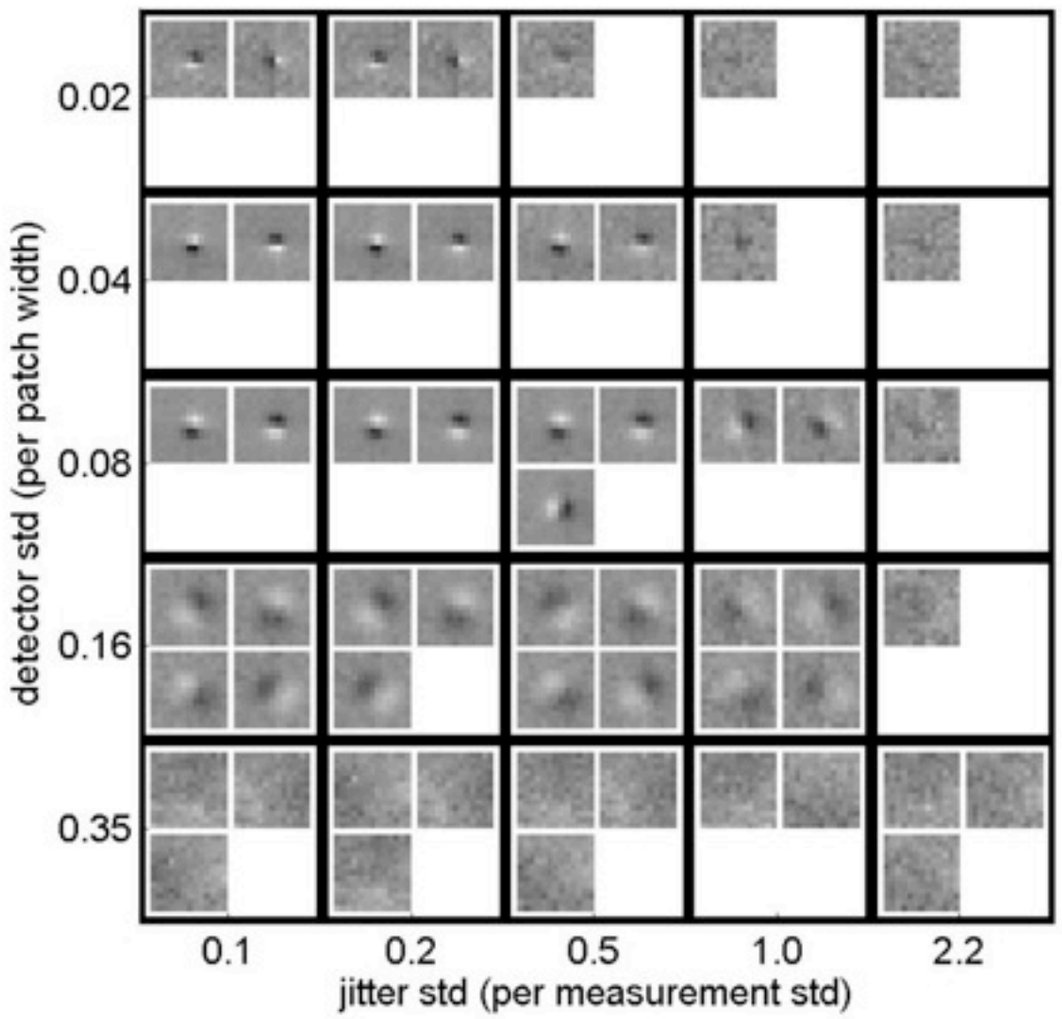
Non-linear Decision-Image Network for Visual Saliency





Critical Control 1: Ground Truth Test

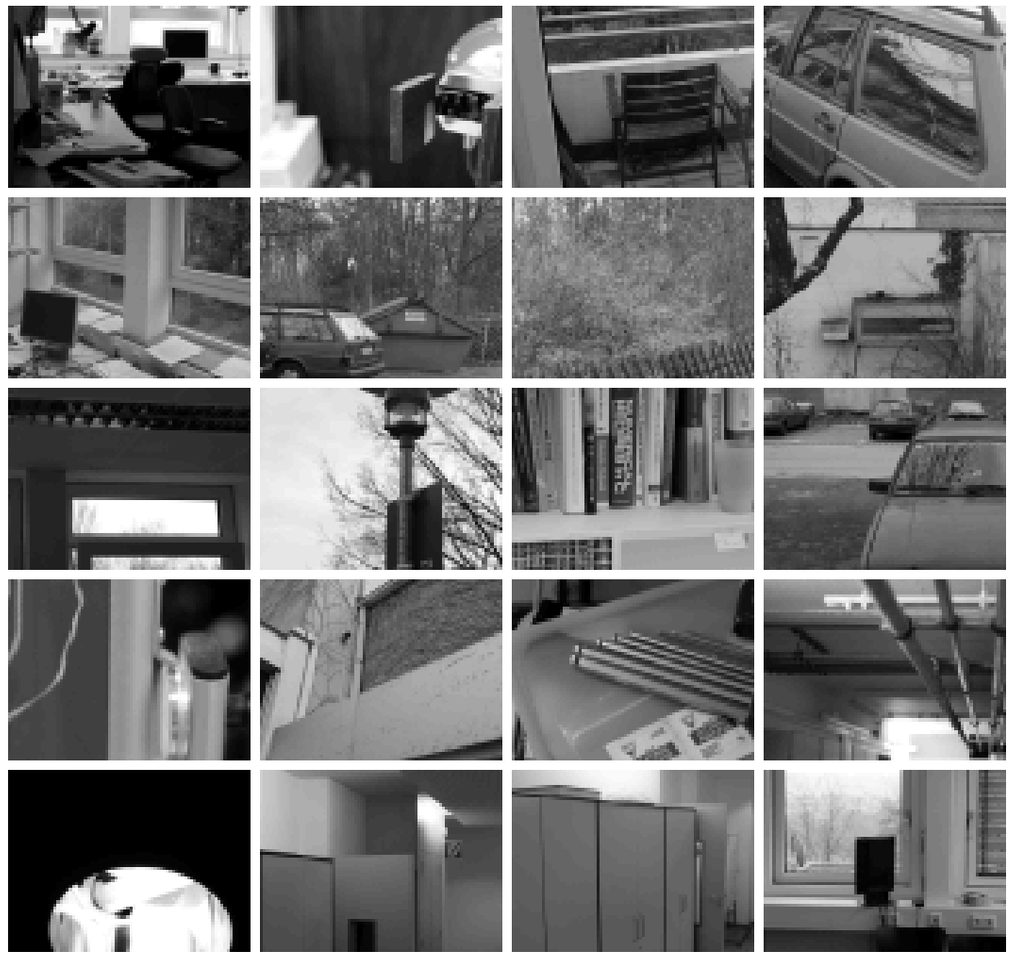




Critical Control 2: Generalization to Novel Data Set

ML-model: 0.64 ± 0.010 s.e.m.

Itti-Koch: 0.62 ± 0.020 s.e.m.



ML-model: 0.62 ± 0.012 s.e.m.

Itti-Koch: 0.57 ± 0.020 s.e.m.

Interim Conclusions (2)

- Bottom-up saliency can be inferred from data, without prior assumptions regarding the computational architecture.

- The most relevant regularity in local image structure at fixation is a simple center-surround configuration. (Biologically plausible but *learned from the data* not assumed!)

- Assembled into a small network with only four standard, linear receptive fields followed by a static nonlinearity and contrast gain-control, the prediction performance of the full RBF-SVM is obtained—this model is very simple compared to previously suggested ones.

- *System identification via reverse-engineering a non-linear kernel machine!*

- This analysis can be seen as an extended psychophysical receptive or perceptive field analysis, recovering *perceptive field networks.*

- Unlike *classification images* or the *bubbles technique* this method can be used under natural viewing conditions, i.e. no image distortion is needed (noise, "bubbles").

Literature (Heavily Biased Sample!)

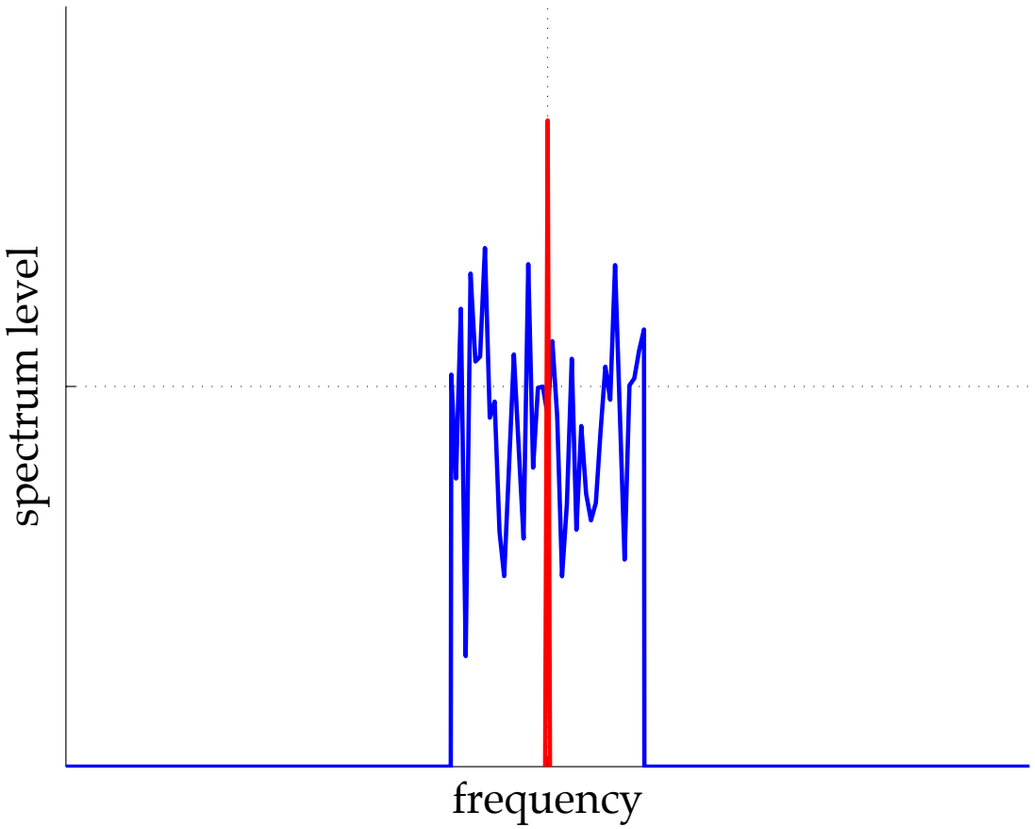
Macke, J.H. & Wichmann, F.A. (2010). Estimating predictive stimulus features from psychophysical data: The decision-image technique applied to human faces. *Journal of Vision* (in press).

Kienzle, W., Franz, M.O., Schölkopf, B. & Wichmann, F.A.. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision*, 9(5):7, 1-15.

Kienzle, K., Wichmann, F.A., Schölkopf, B. & Franz, M.O. (2007). A nonparametric approach to bottom-up visual saliency. *Advances in Neural Information Processing Systems*, 19, 689-696.

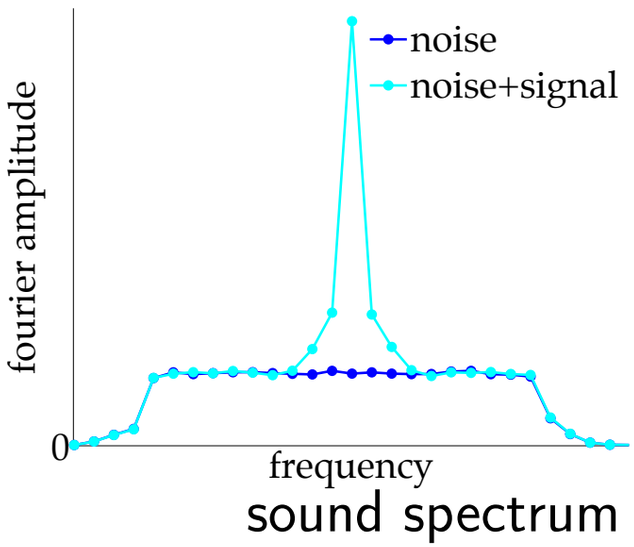
Wichmann, F.A., Graf, A.B.A., Simoncelli, E.P., Bülthoff, H.H. & Schölkopf, B. (2005). Machine learning applied to perception: decision-images for gender classification. *Advances in Neural Information Processing Systems*, 17, 1489-1496.

Tone-in-Noise Detection

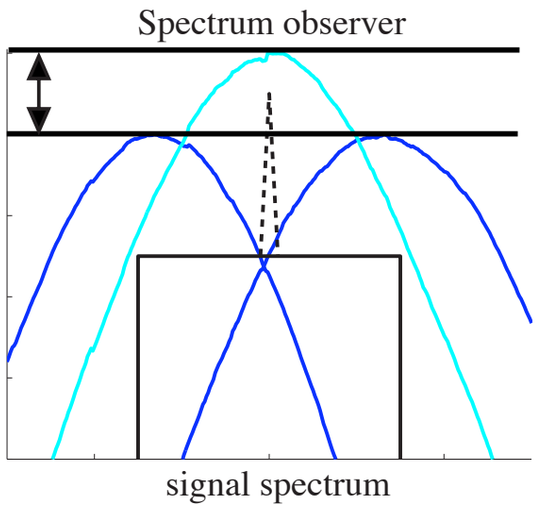


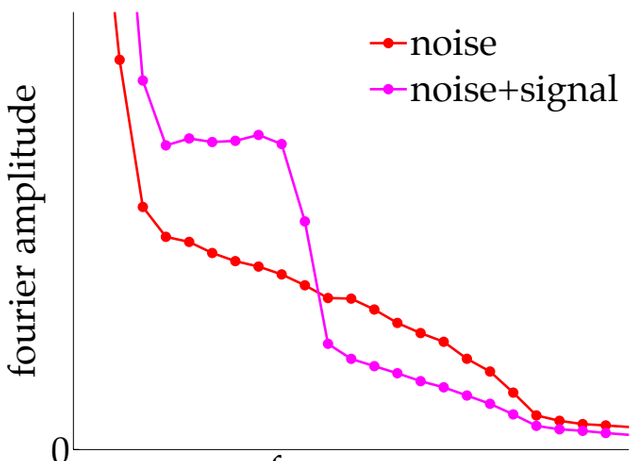
Harvey Fletcher (left) at Bell Telephone Labs in NYC

Synthetic Observers (i.e. Simulated Features)

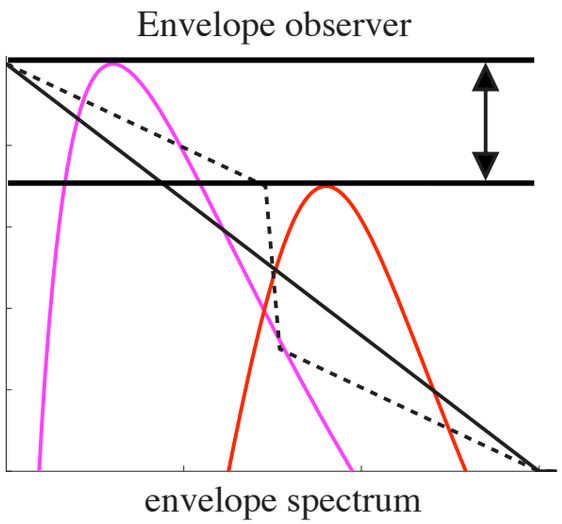


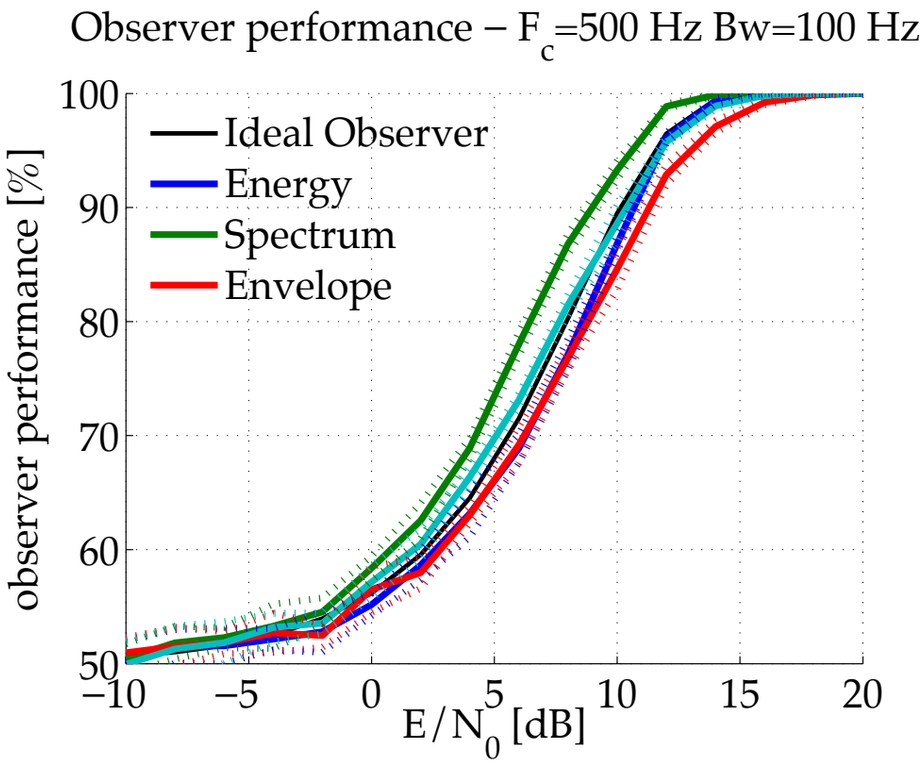
sound spectrum



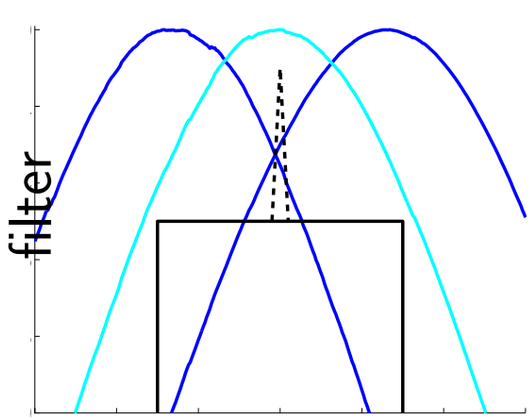


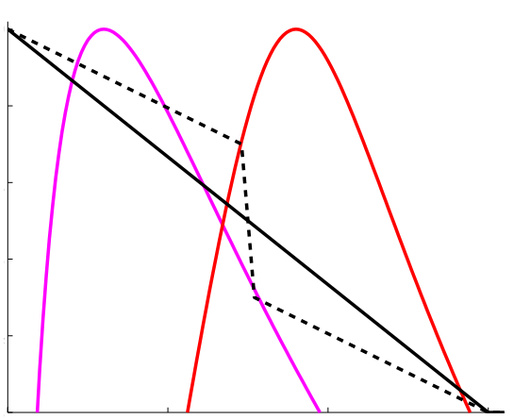
envelope spectrum





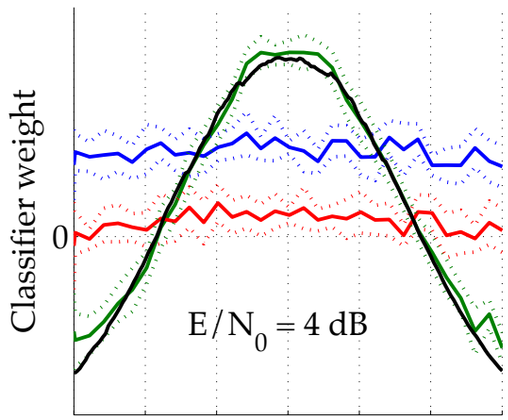
Observer Reconstruction

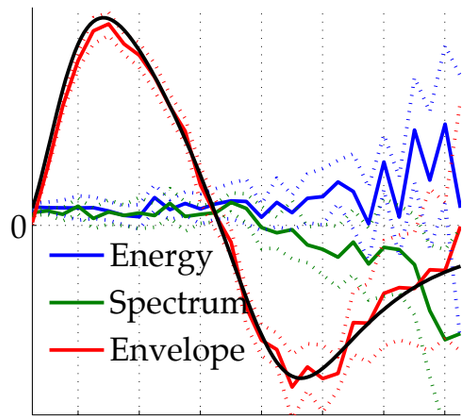




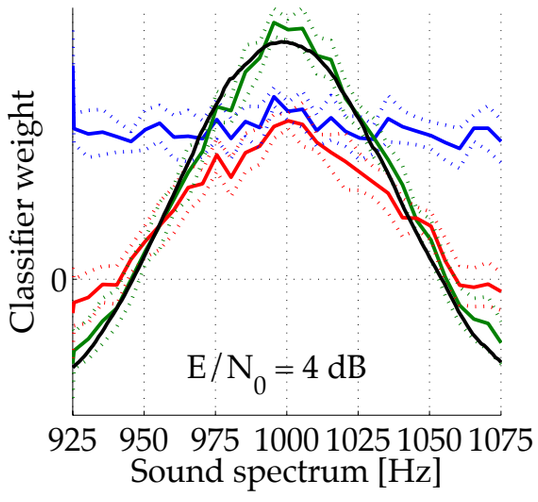
sound spectrum envelope spectrum

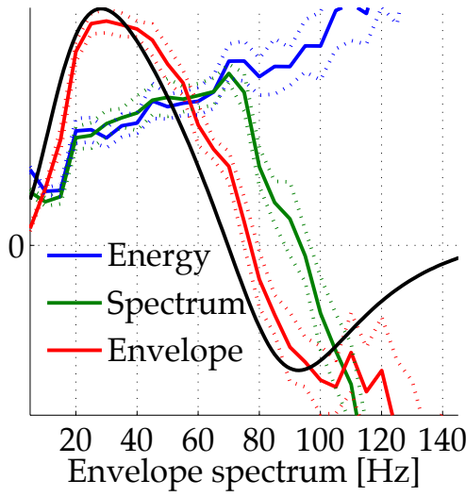
Ground Truth



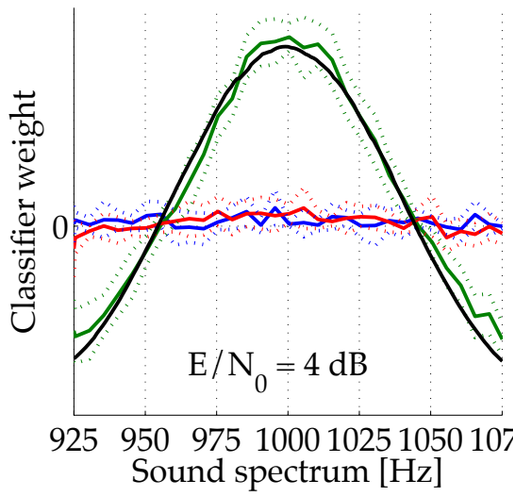


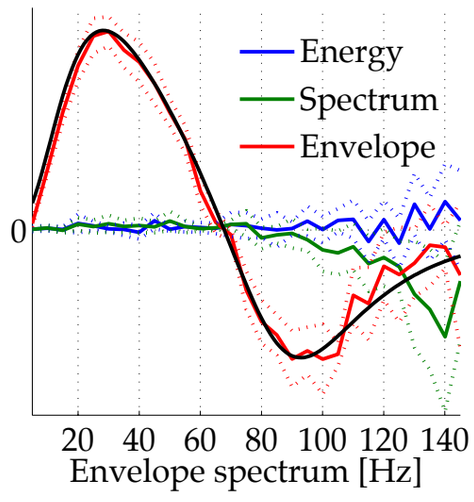
2-norm Classifier





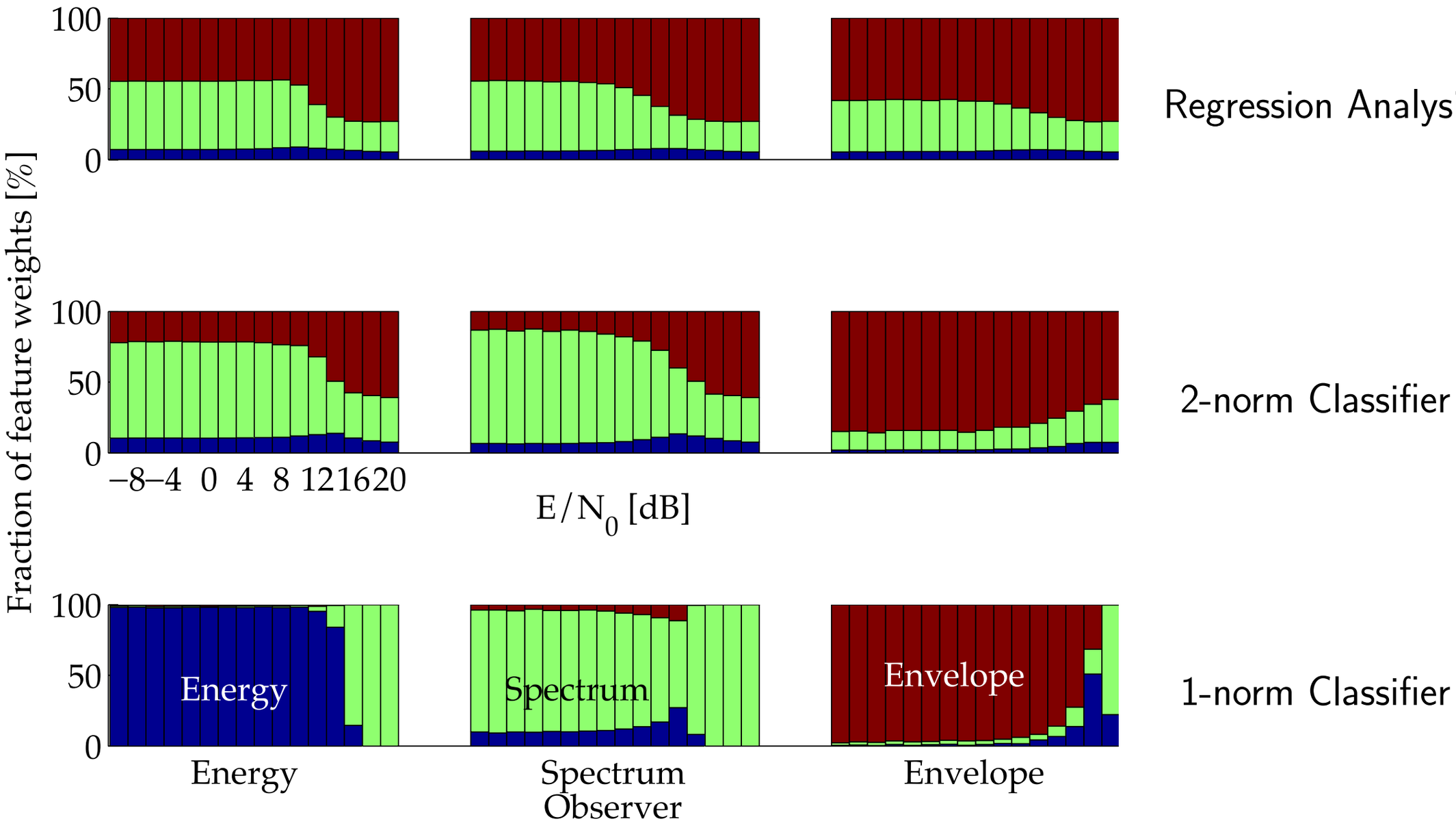
Regression Analysis

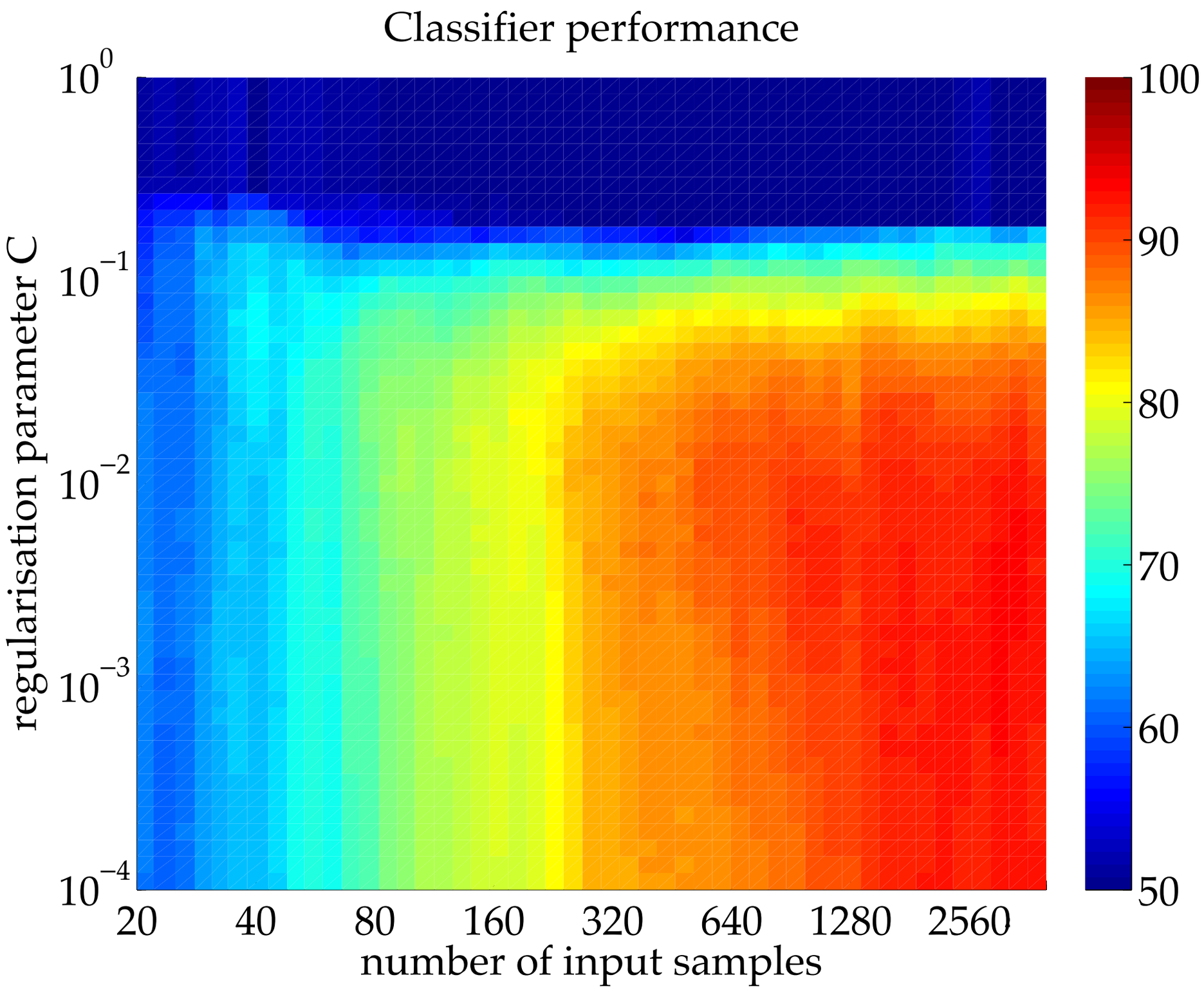




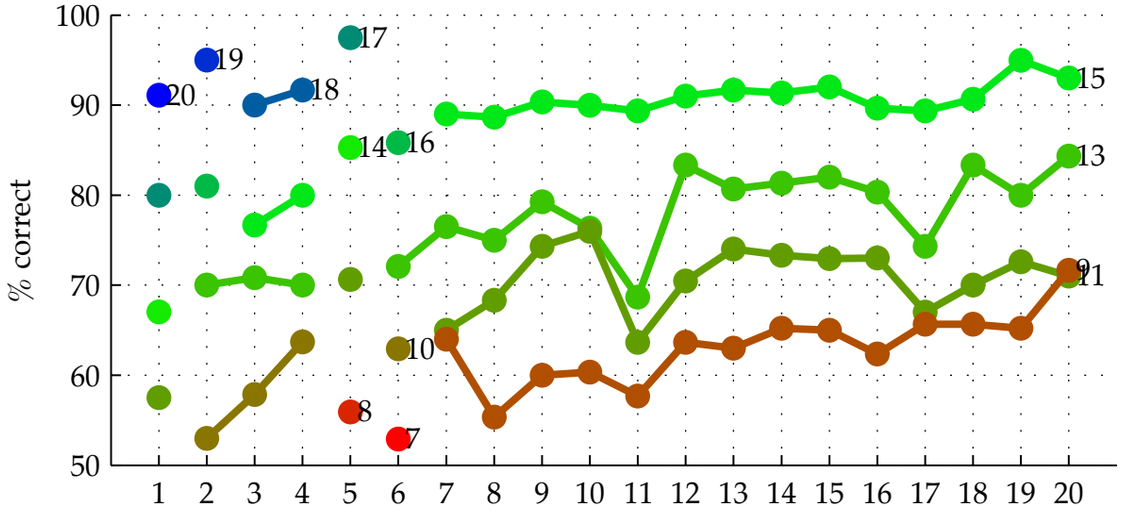
1-norm Classifier

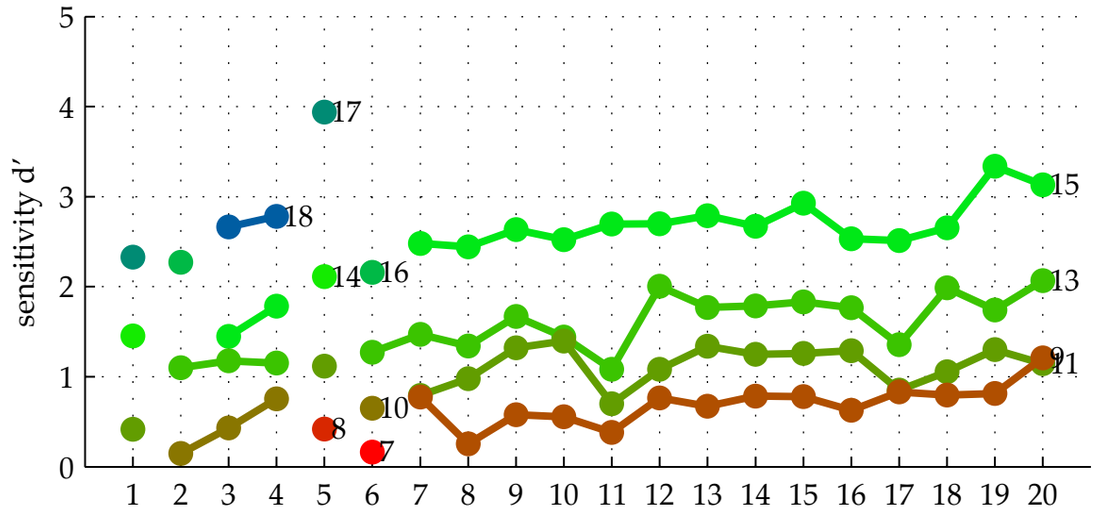
Observer Reconstruction Continued—Feature Weights

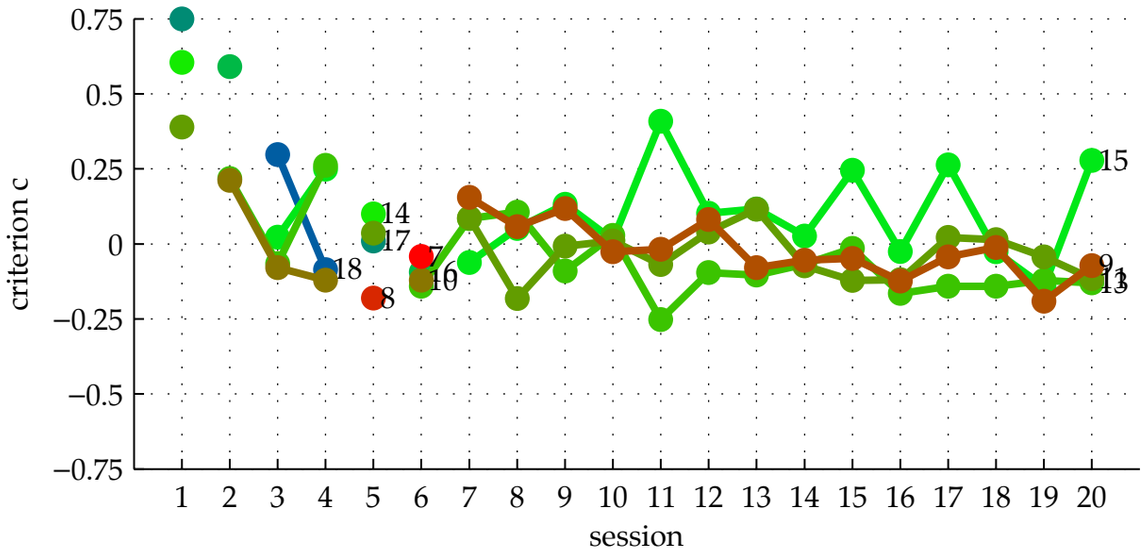



Classifier performance
regularisation parameter C
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
20
40
80
160
320
640
1280
2560
number of input samples
100
90
80
70
60
50

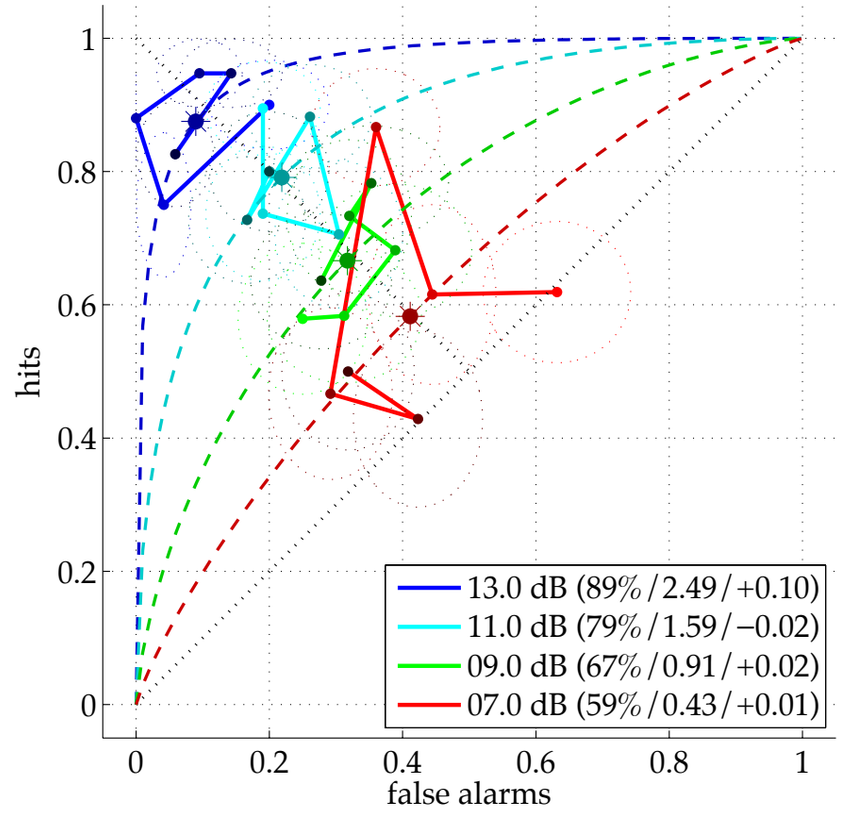
observer JR



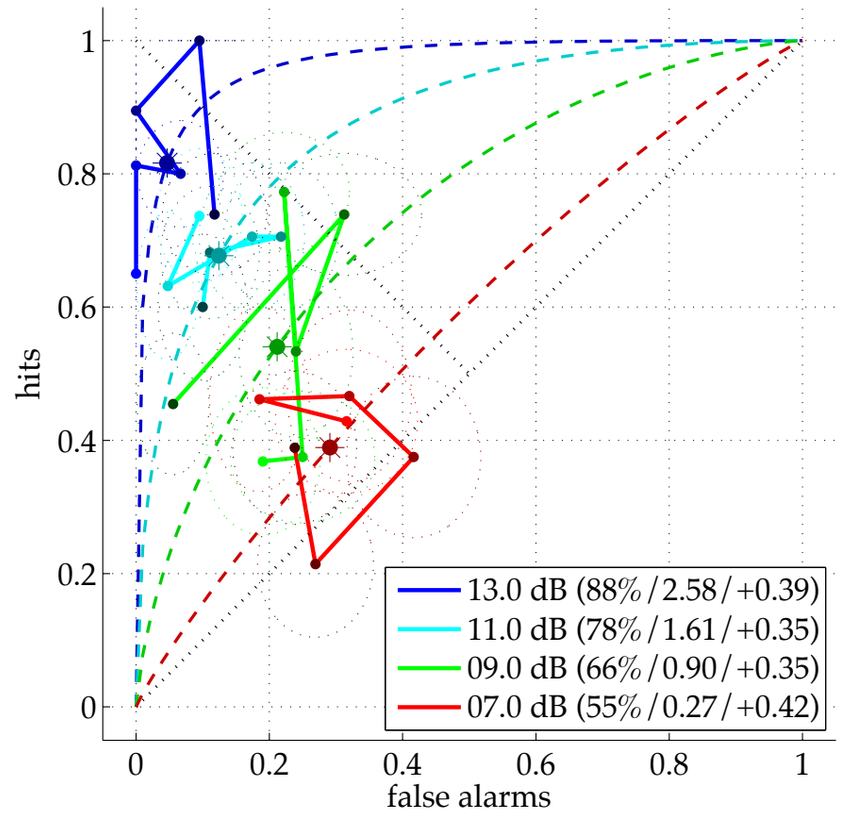


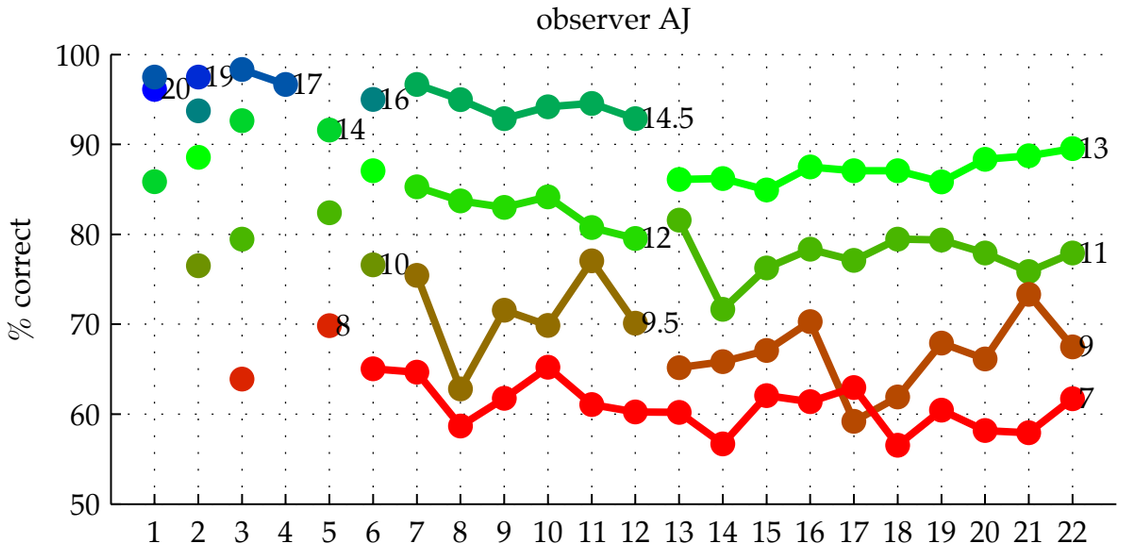


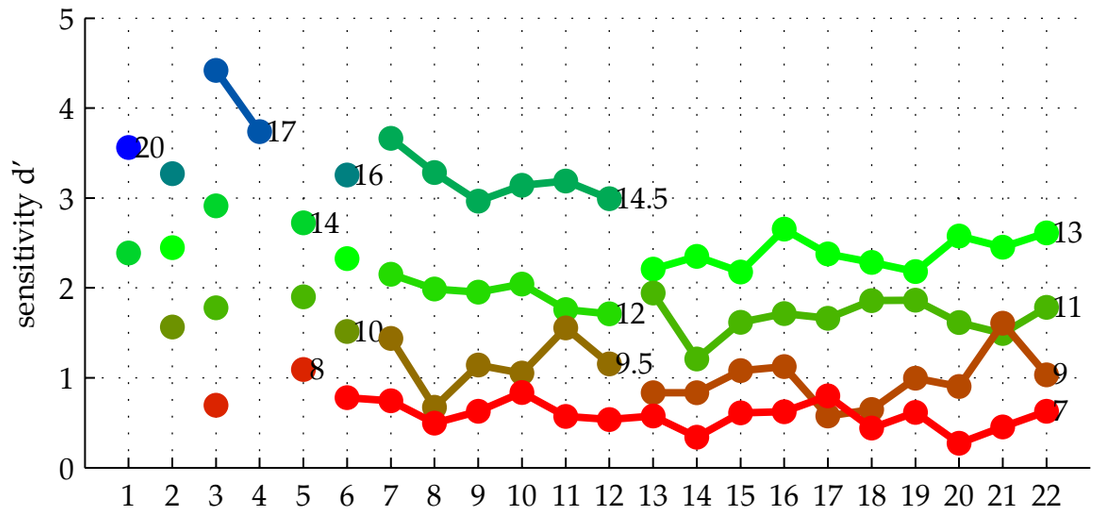
AB – session 20 – 40 samples

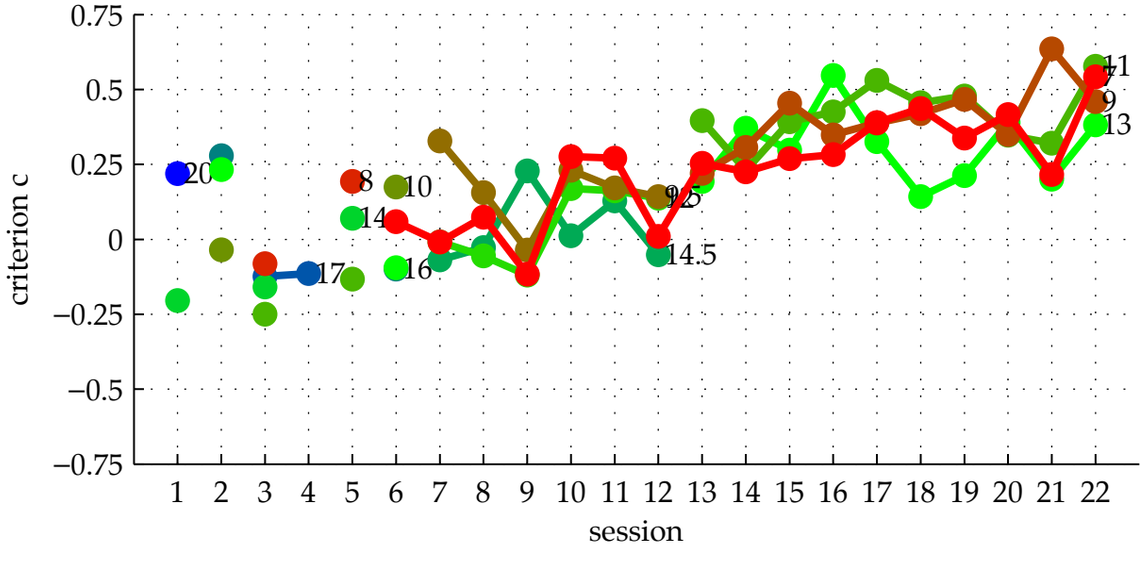


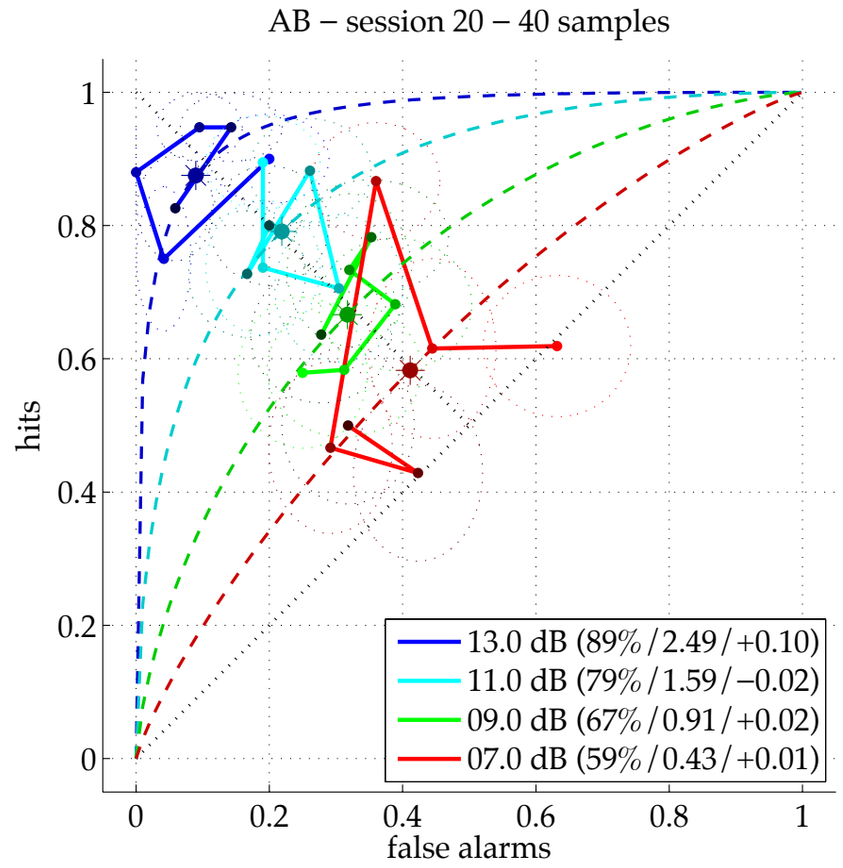
AJ – session 20 – 40 samples

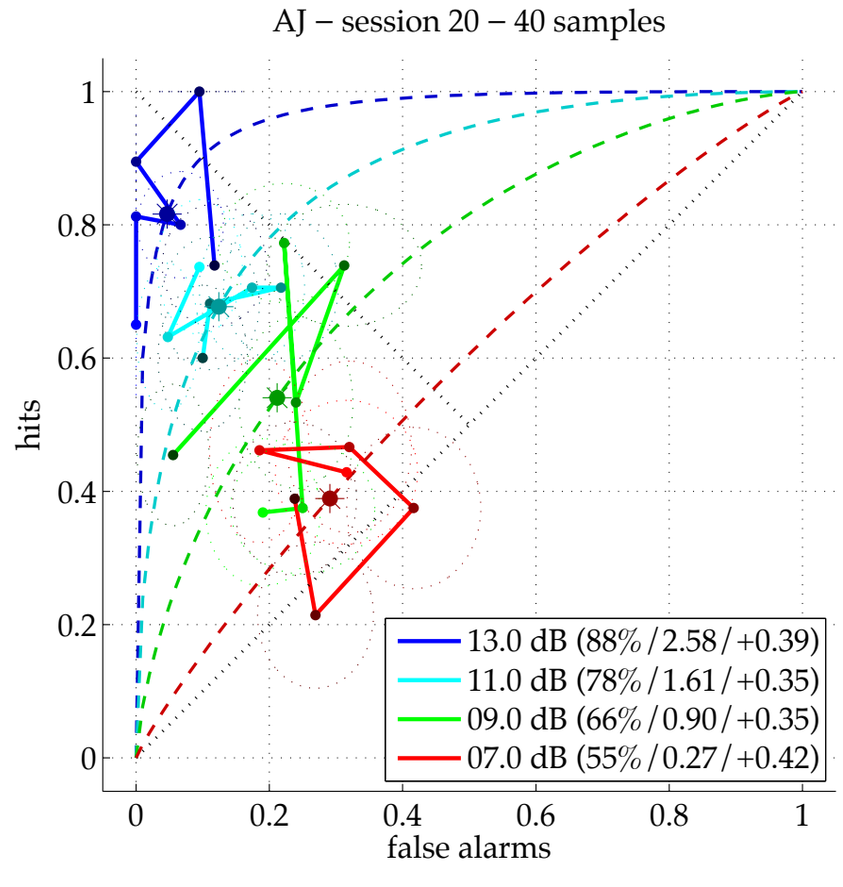



observer AJ
% correct
100
90
80
70
60
50
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
20
19
17
16
14
14.5
13
12
11
10
9.5
9
8
7


sensitivity d'
5
4
3
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
20
17
16
14
14.5
13
12
11
10
9.5
9
8
7


criterion c
0.75
0.5
0.25
0
−0.25
−0.5
−0.75
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
session
20
17
16
14
8
10
9.5
14.5
11
9
13


AB – session 20 – 40 samples
hits
false alarms
0
0.2
0.4
0.6
0.8
1
13.0 dB (89% / 2.49 / +0.10)
11.0 dB (79% / 1.59 / −0.02)
09.0 dB (67% / 0.91 / +0.02)
07.0 dB (59% / 0.43 / +0.01)


AJ – session 20 – 40 samples
hits
false alarms
0
0.2
0.4
0.6
0.8
1
13.0 dB (88% / 2.58 / +0.39)
11.0 dB (78% / 1.61 / +0.35)
09.0 dB (66% / 0.90 / +0.35)
07.0 dB (55% / 0.27 / +0.42)

Bernstein Center for
Computational Neuroscience
Berlin

Thank you very much!

Felix A. Wichmann

Modelling of Cognitive Processes Group
Bernstein Center for Computational Neuroscience
and
Technische Universität Berlin

felix.wichmann@tu-berlin.de

Bundesministerium
für Bildung
und Forschung