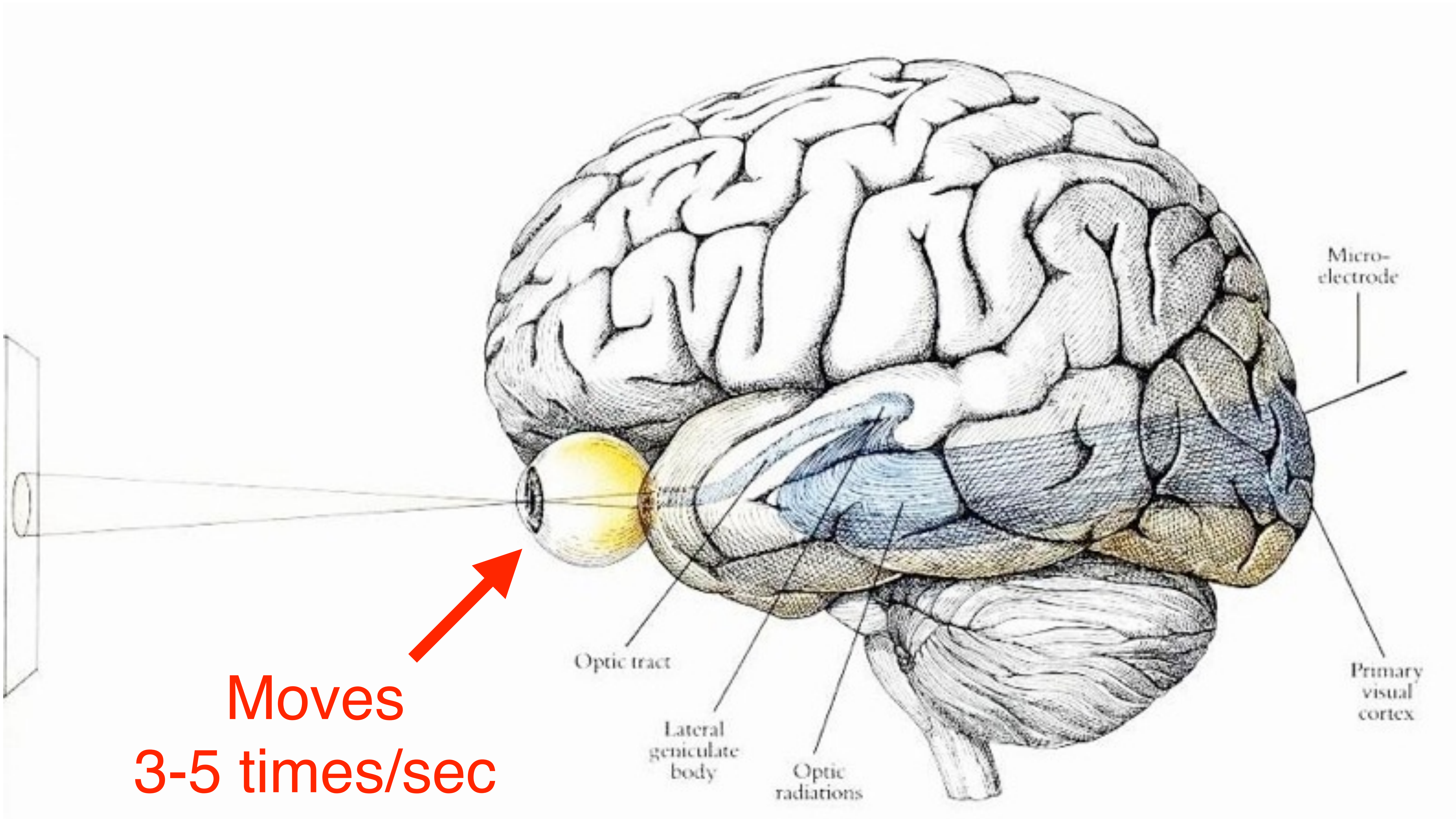
What are the principles governing information processing in this system?



Moves
3-5 times/sec

'AlexNet'

(Krizhevsky, Sutskever & Hinton 2012)

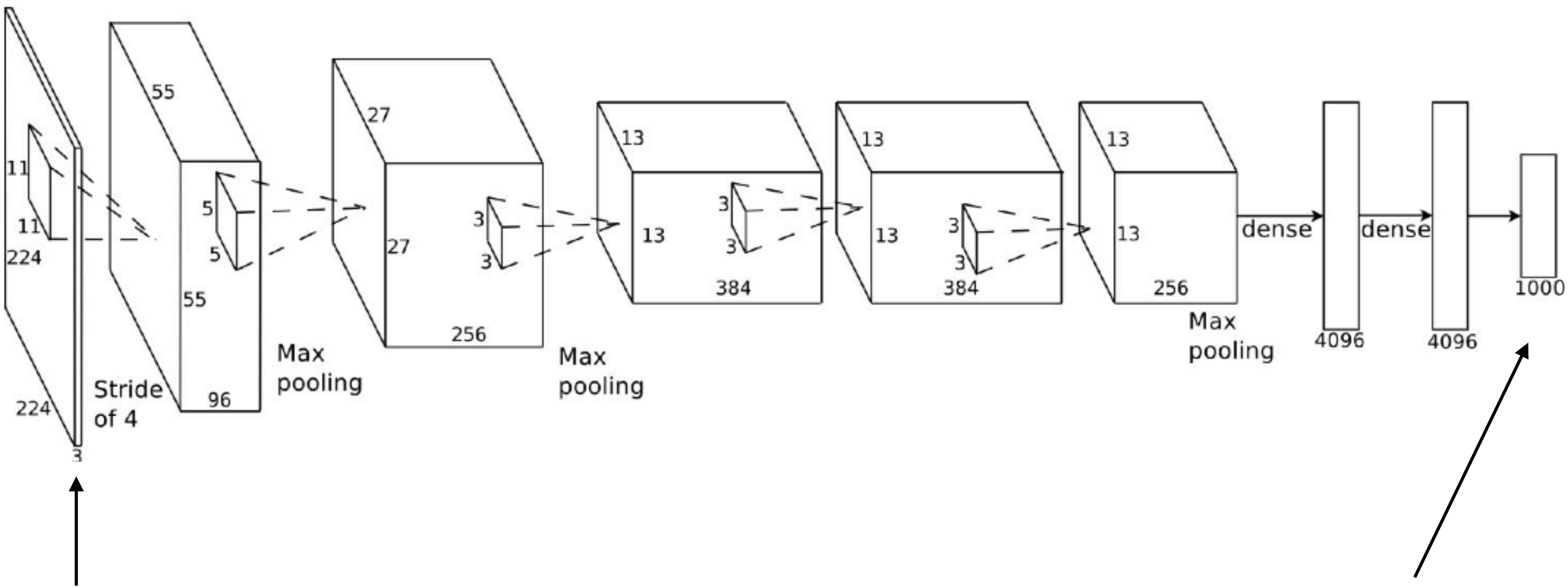
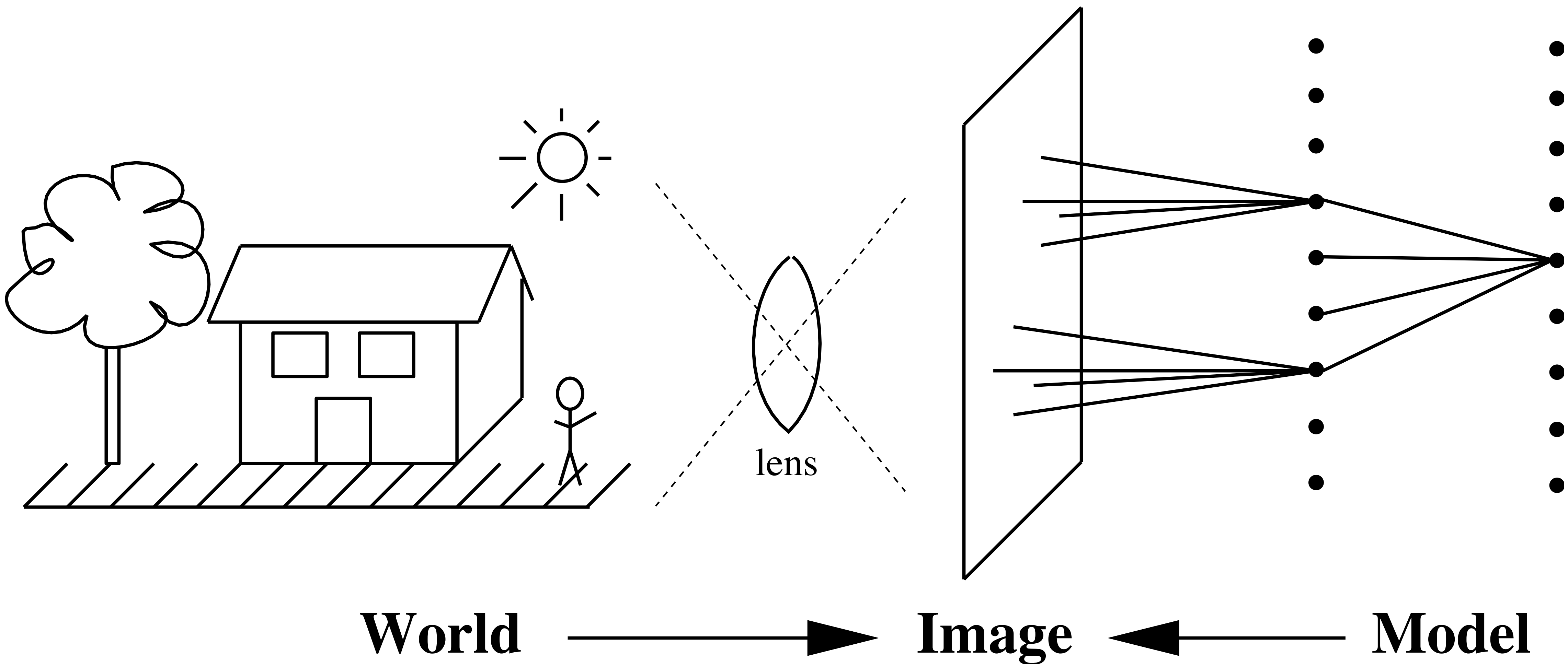
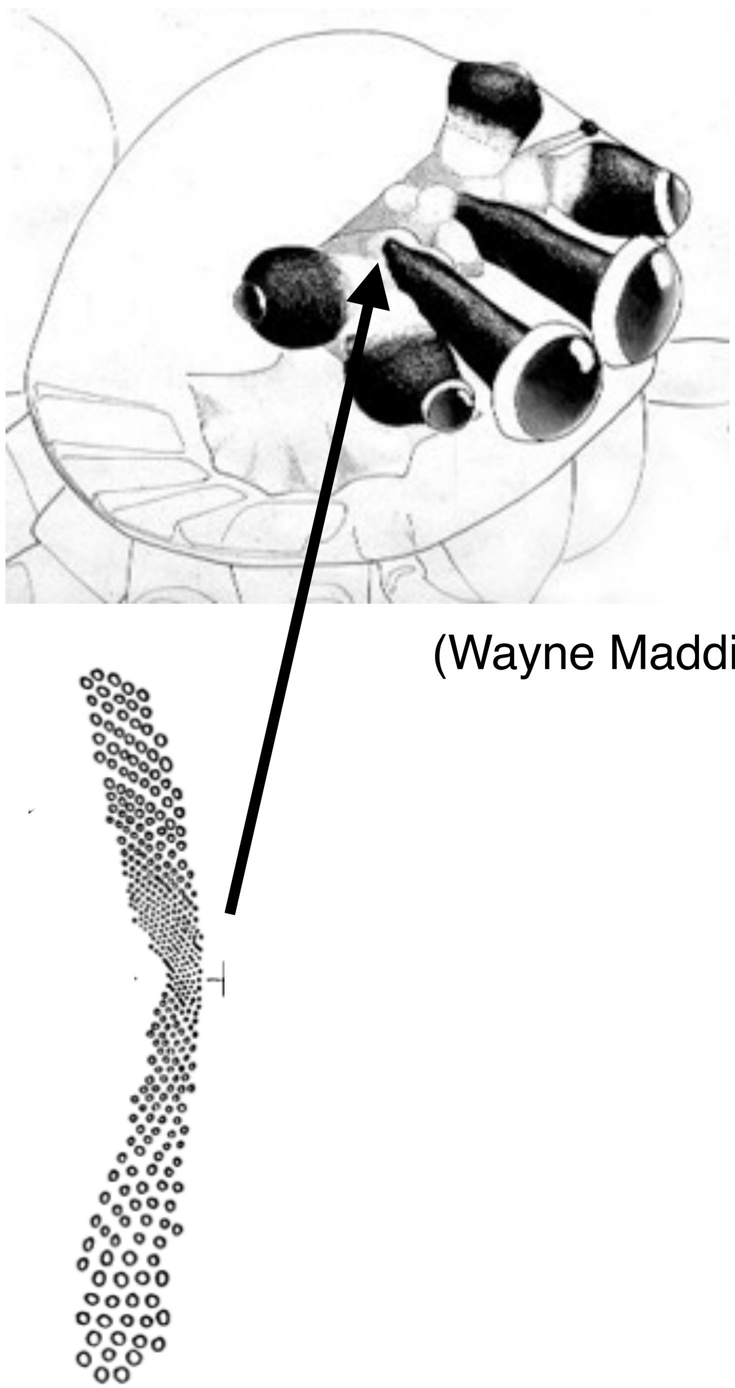


image feature extraction and pooling classification

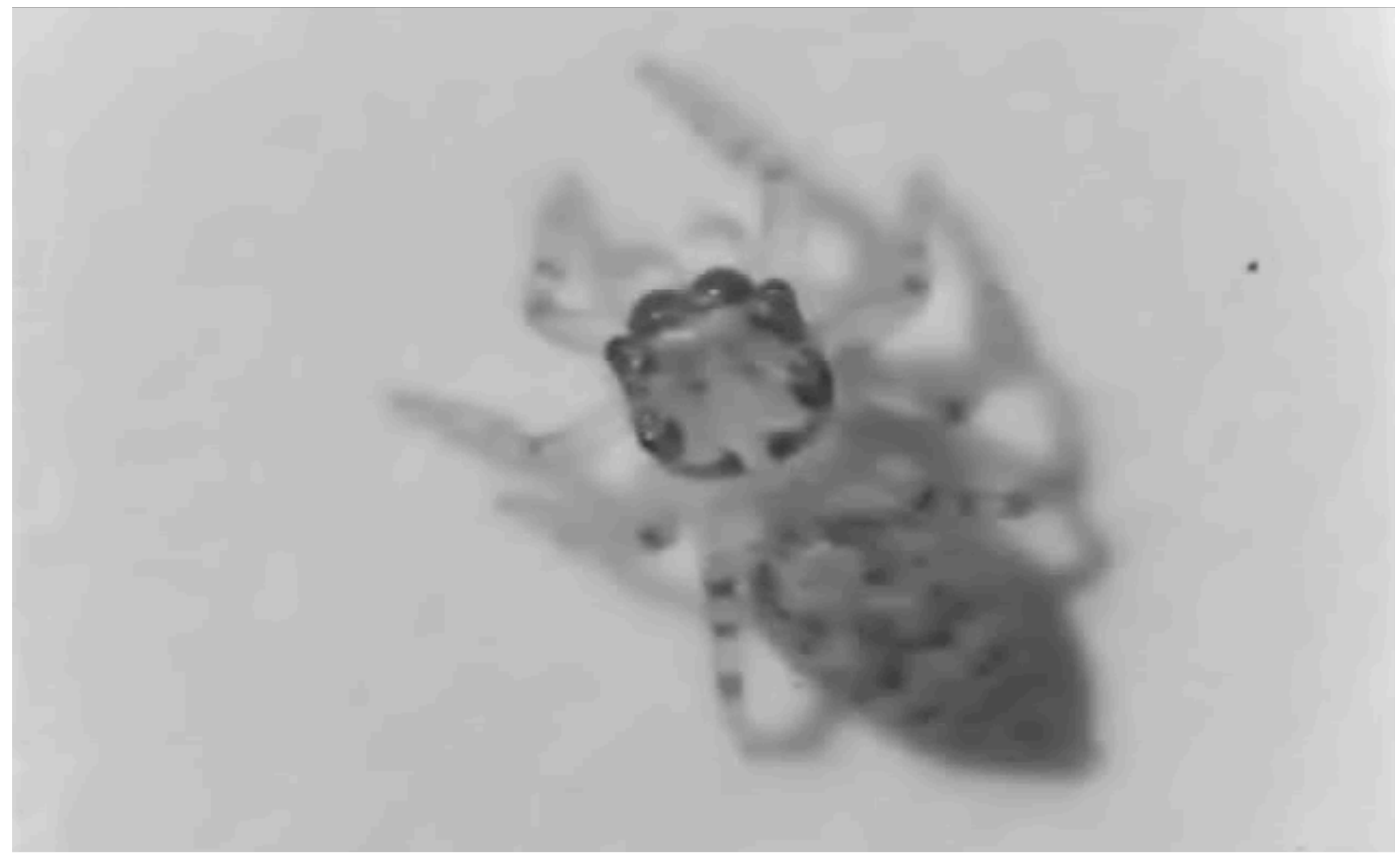
Vision as inference



Active vision in jumping spiders



(Wayne Maddison)



(Bair & Olshausen, 1991)

One-day old jumping spider
(filmed in the Bower lab, Caltech 1991)

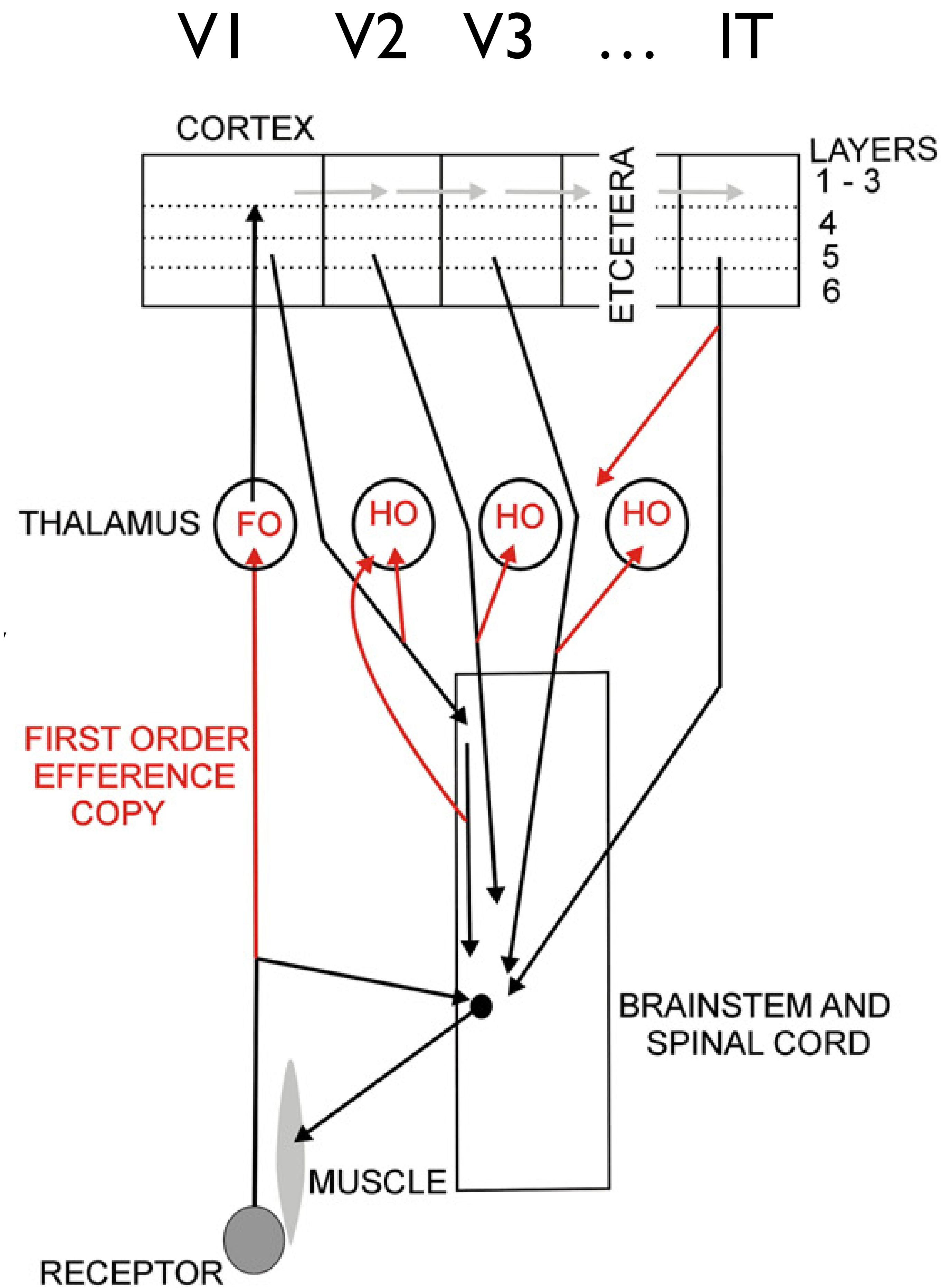
One-day old jumping spider
(filmed in the Bower lab, Caltech 1991)

…problem solving behavior, language, expert knowledge and application, and reason, are all pretty simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction. This part of intelligence is where evolution has concentrated its time--it is much harder.

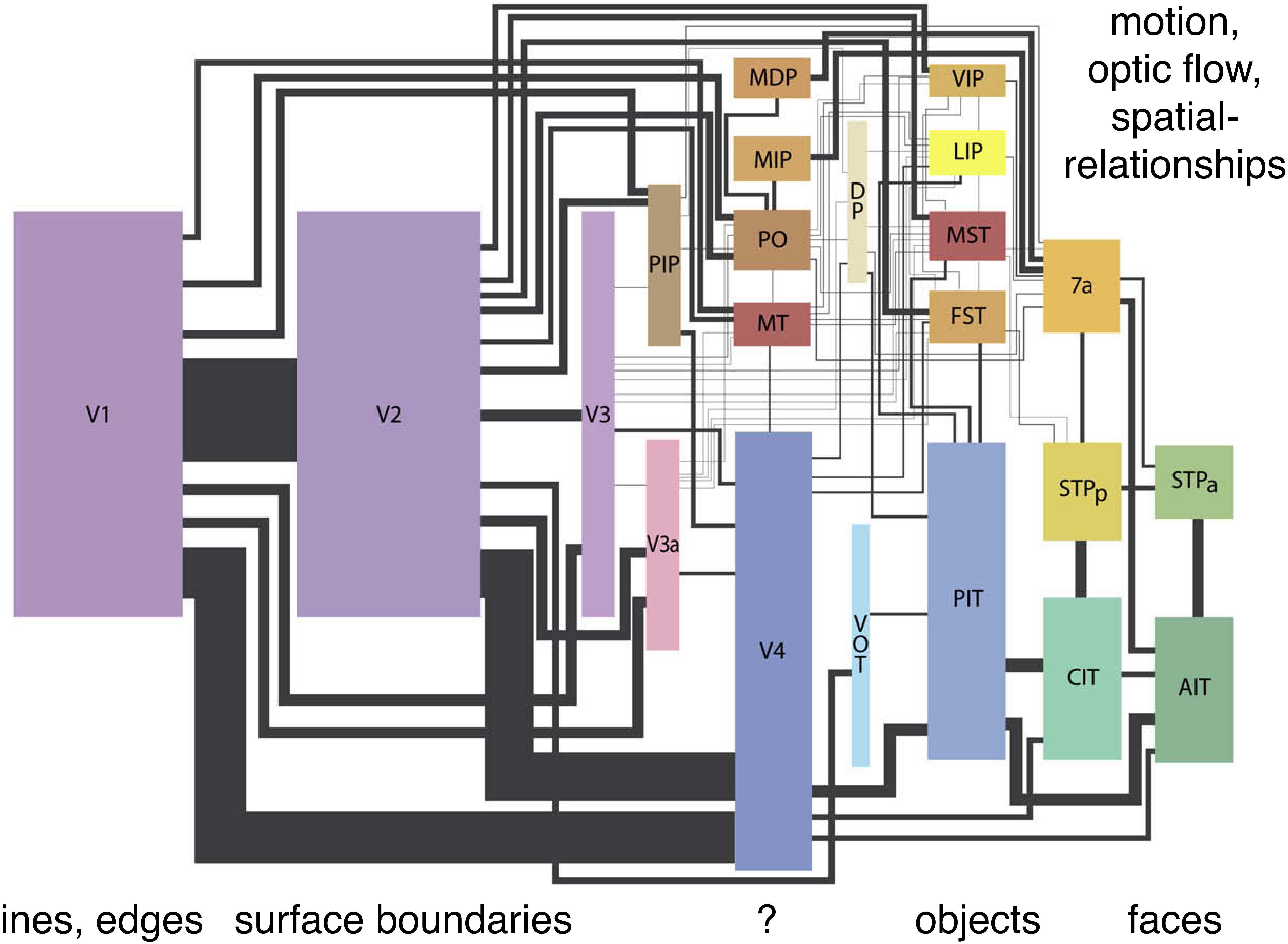
— Rodney Brooks, "Intelligence without representation," *Artificial Intelligence* (1991)

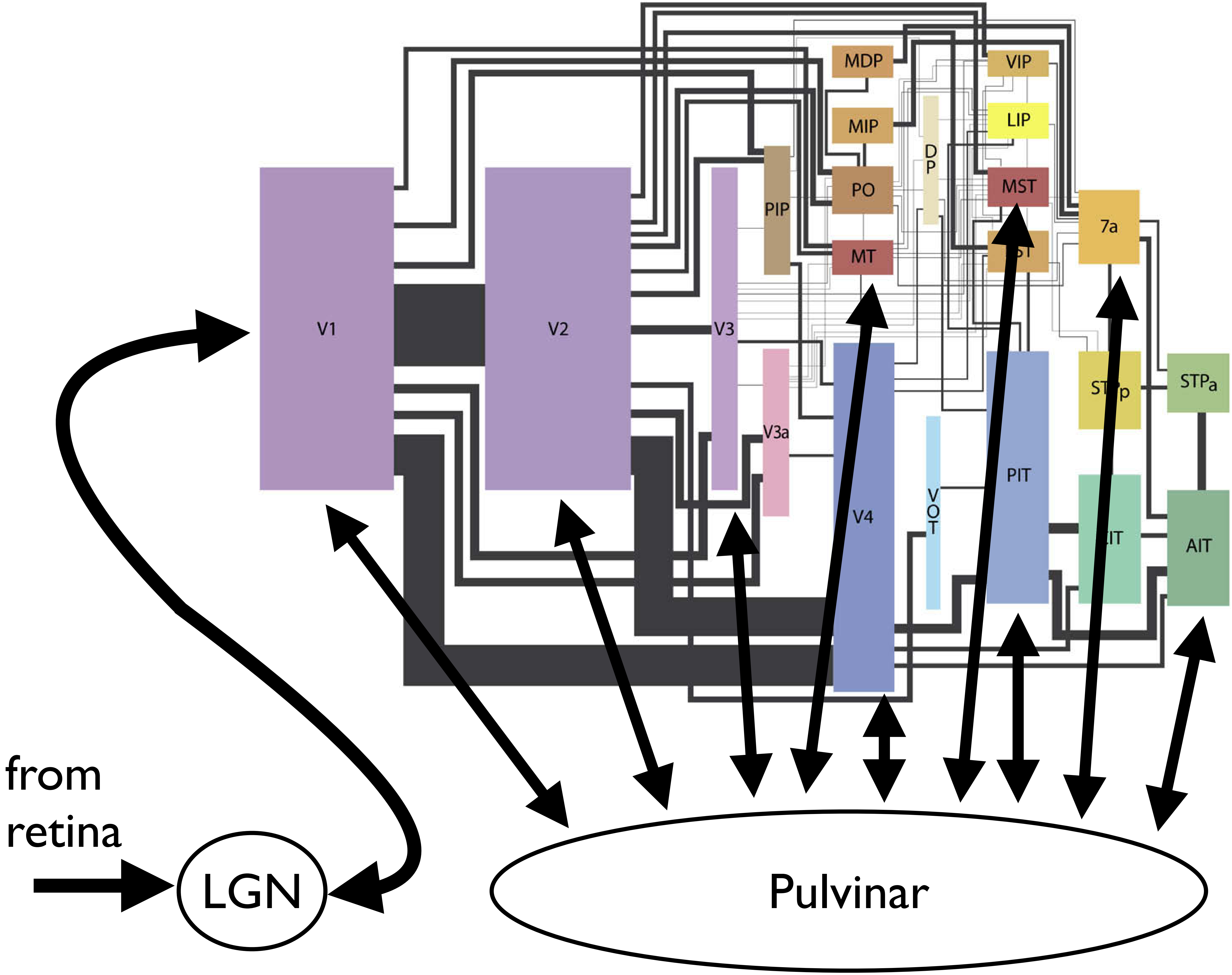
An alternative view

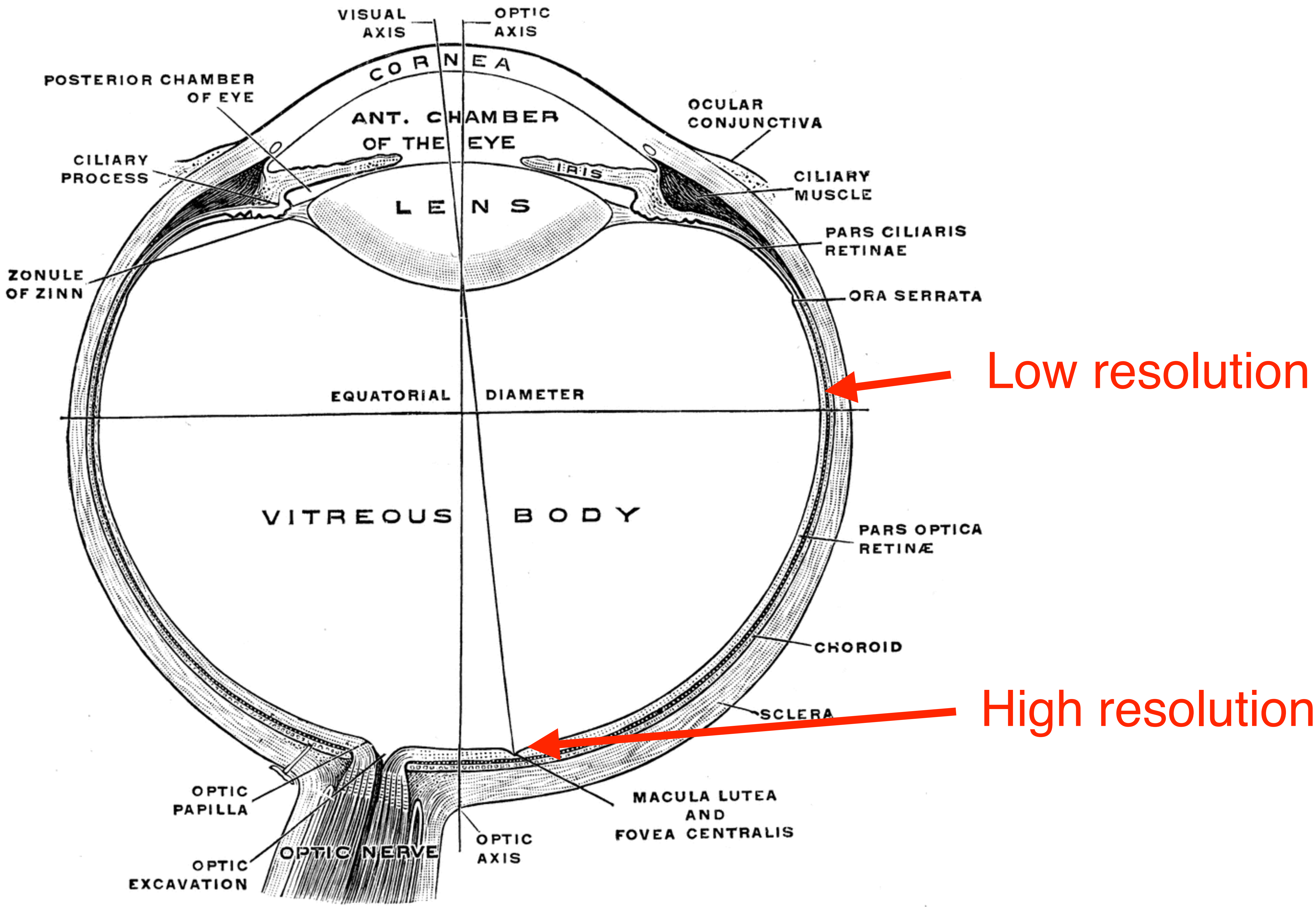
(Guillery & Sherman, 2011)



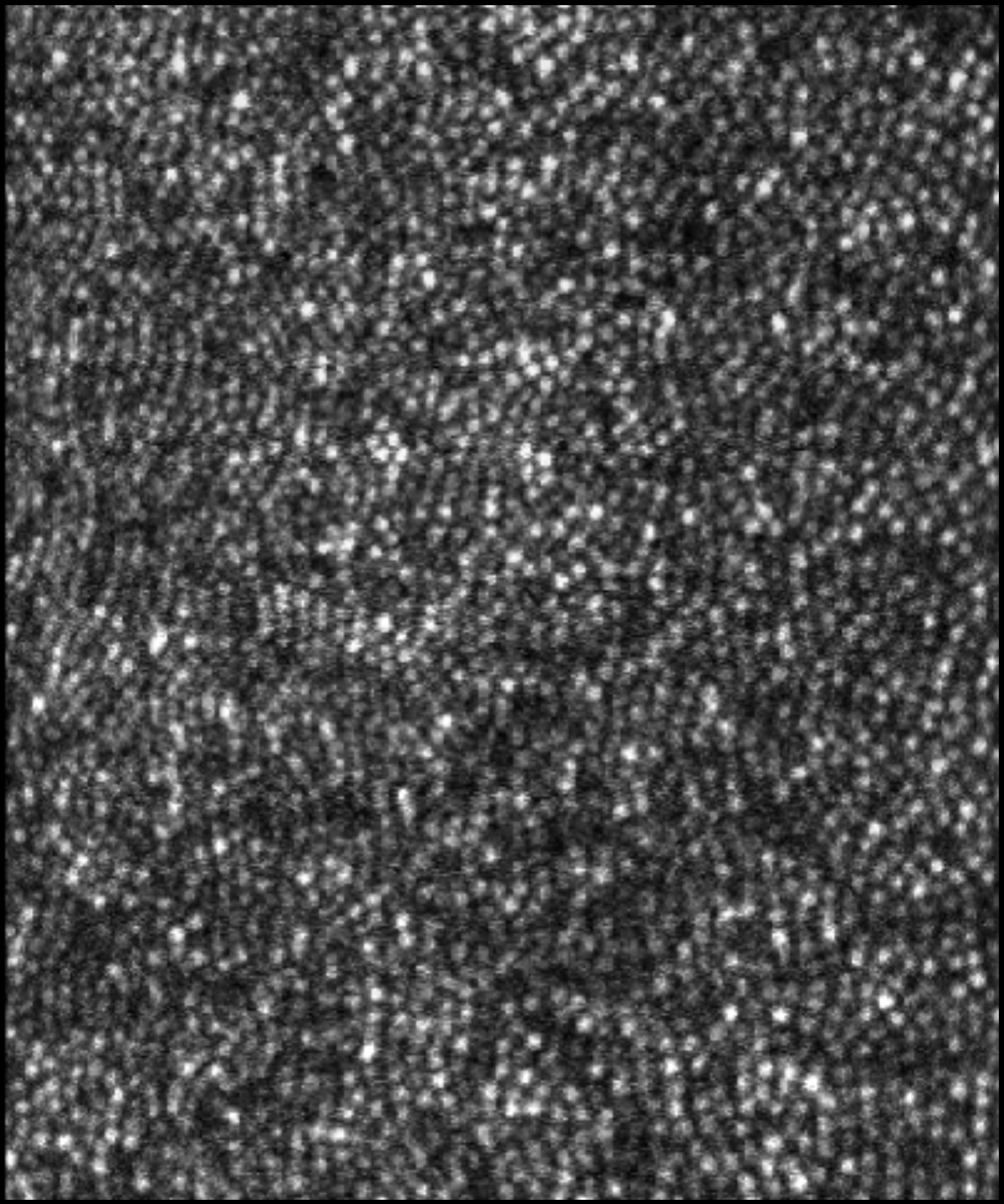
Wallisch & Movshon (2008)




MDP
VIP
MIP
LIP
DP
PO
MST
PIP
MT
7a
V1
V2
V3
STPa
V3a
V4
PIT
VOT
AIT
from retina
LGN
Pulvinar


VISUAL AXIS
OPTIC AXIS
CORNEA
POSTERIOR CHAMBER OF EYE
ANT. CHAMBER OF THE EYE
OCULAR CONJUNCTIVA
CILIARY PROCESS
IRIS
CILIARY MUSCLE
LENS
PARS CILIARIS RETINAE
ZONULE OF ZINN
ORA SERRATA
Low resolution
EQUATORIAL DIAMETER
VITREOUS BODY
PARS OPTICA RETINÆ
CHOROID
SCLERA
High resolution
OPTIC PAPILLA
MACULA LUTEA AND FOVEA CENTRALIS
OPTIC NERVE
OPTIC AXIS
OPTIC EXCAVATION

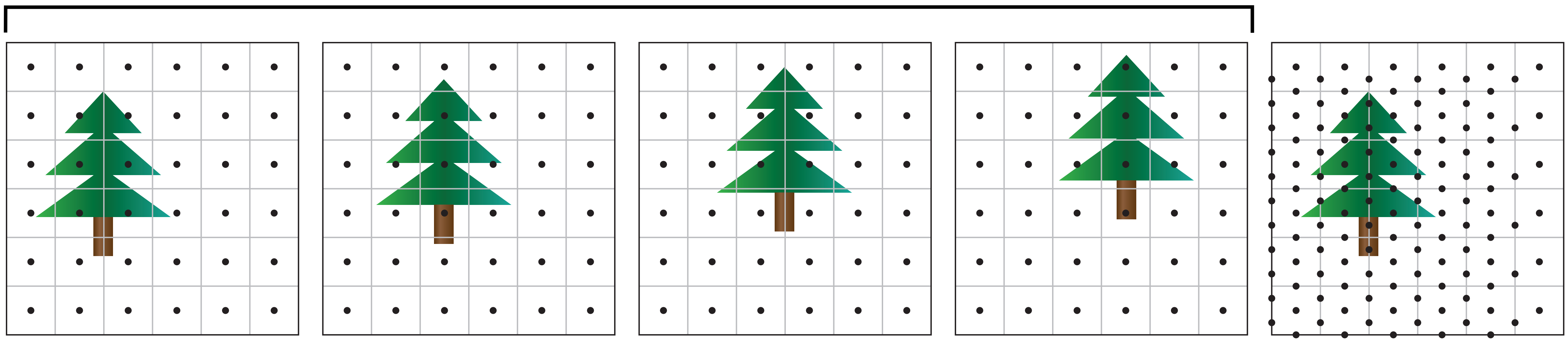
Fixational eye movements (drift)



(from Austin Roorda, UC Berkeley)

Super-resolution in the Google Pixel camera

image displacements from natural hand tremor

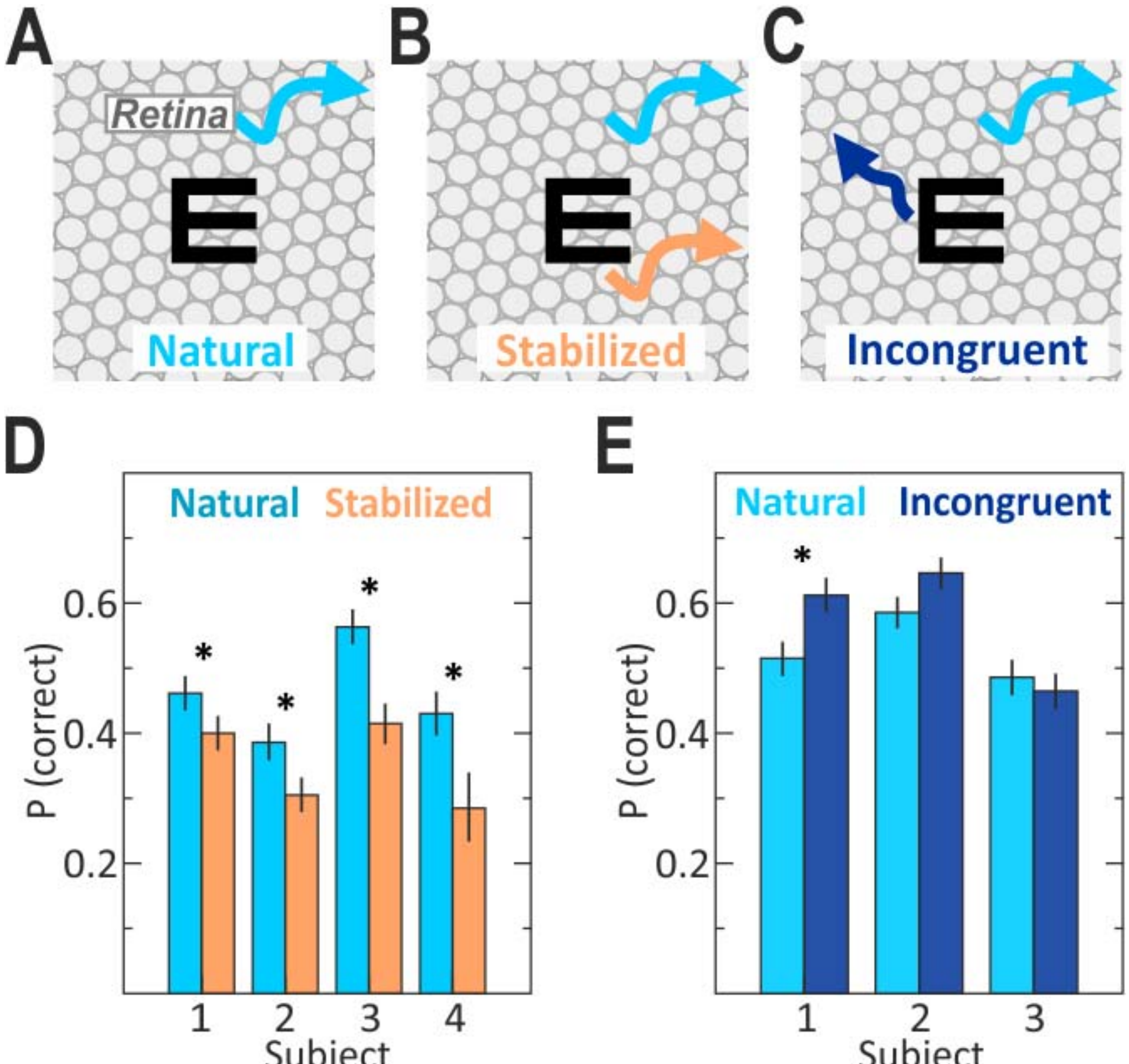


1st frame (base frame) | 2nd frame | 3rd frame | 4th frame | All frames aligned to base frame

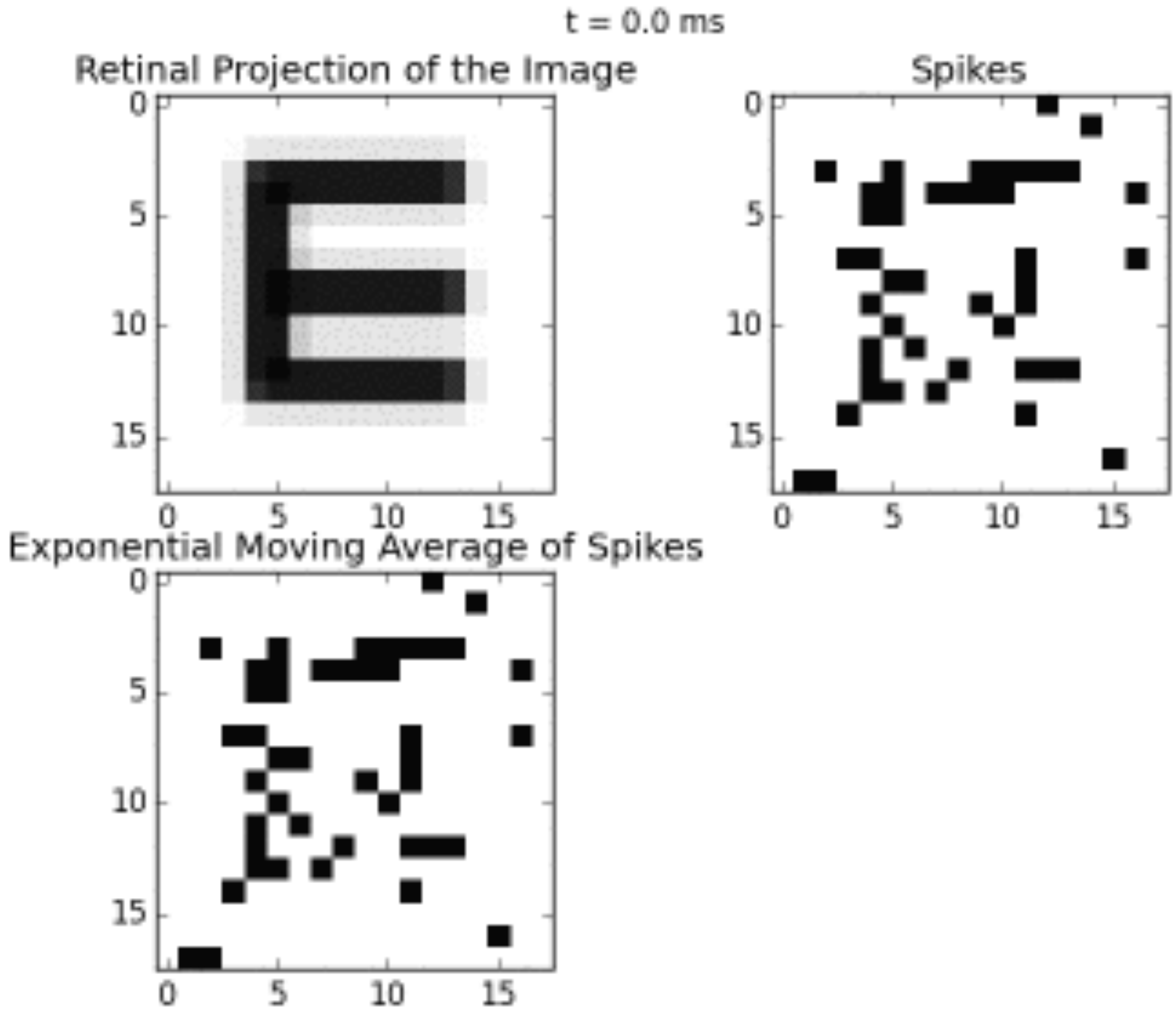
Wronski, Garcia-Dorado, Ernst, Kelly, Krainin, Liang, Levoy, Milanfar (2019)

Retinal image motion helps pattern discrimination

(Ratnam, K., Domdei, N., Harmening, W. M., & Roorda, A., *Journal of Vision, 2017)*



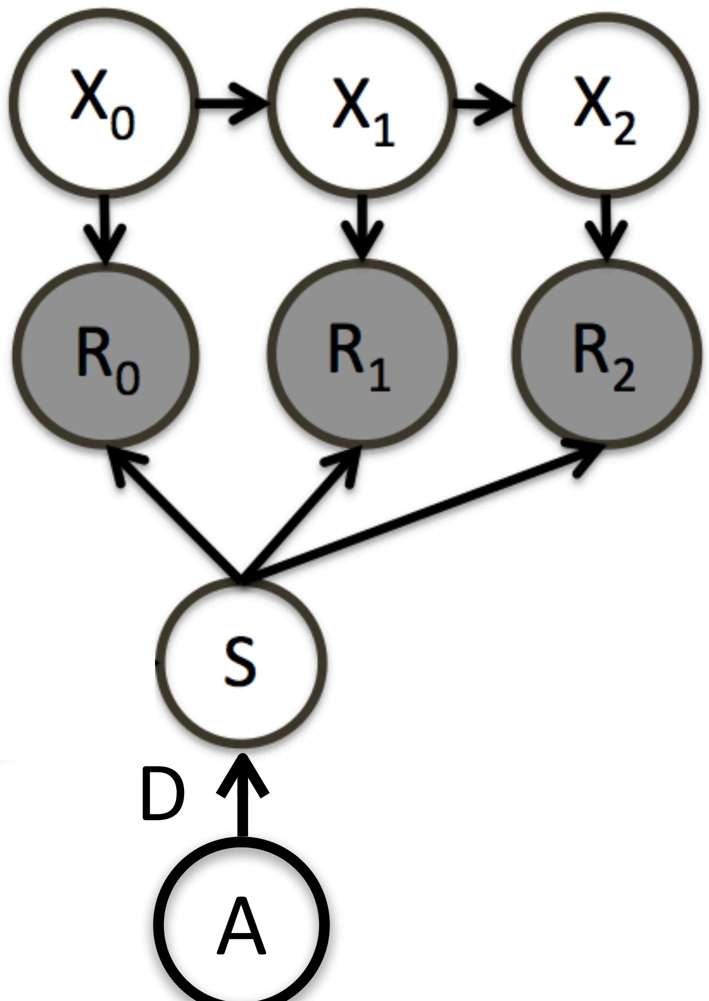
Simple averaging by cortex is not sufficient



Bayesian model for inferring form and motion

(Anderson, Ratnam, Roorda & Olshausen, 2020)

Alex Anderson



Eye position

Spikes
(from LGN afferents)

Pattern $S = D\,A$

'Shape'

$$\hat{A} \quad = \quad \arg\max_{A} \, \log P(A|R)$$

Joint estimation of pattern (S) and position (X) from retinal spike trains (R)

(Anderson, Ratnam, Roorda & Olshausen, 2020)

$$
\begin{aligned}
\hat{A} &= \arg\max_A \log P(A|R) \\
&= \arg\max_A \log P(R|A)\, P(A) \\
&= \arg\max_A \log \sum_X P(R|X, S = D\, A)\, P(X) + \log P(A)
\end{aligned}
$$

$$
\begin{aligned}
\hat{A} &\leftarrow \arg\max_A \sum_X q(X)\, \log P(R|S = DA, X) + \log P(A) \\
q(X) &\leftarrow P(X|R, S = D\hat{A})
\end{aligned}
$$

Joint estimation of pattern (S) and position (X) from retinal spike trains (R)

(Anderson, Ratnam, Roorda & Olshausen, 2020)

Given current estimate of position (X_t), update A

$$\hat{A}^{t+1} = \arg\max_A \sum_{t'=0}^{t} \sum_j \langle \log P(R_{j,t'} | X_{t'}, S = D\,A) \rangle_{P(X_{t'}|S^t, R_{0:t})} + \log P(A)$$

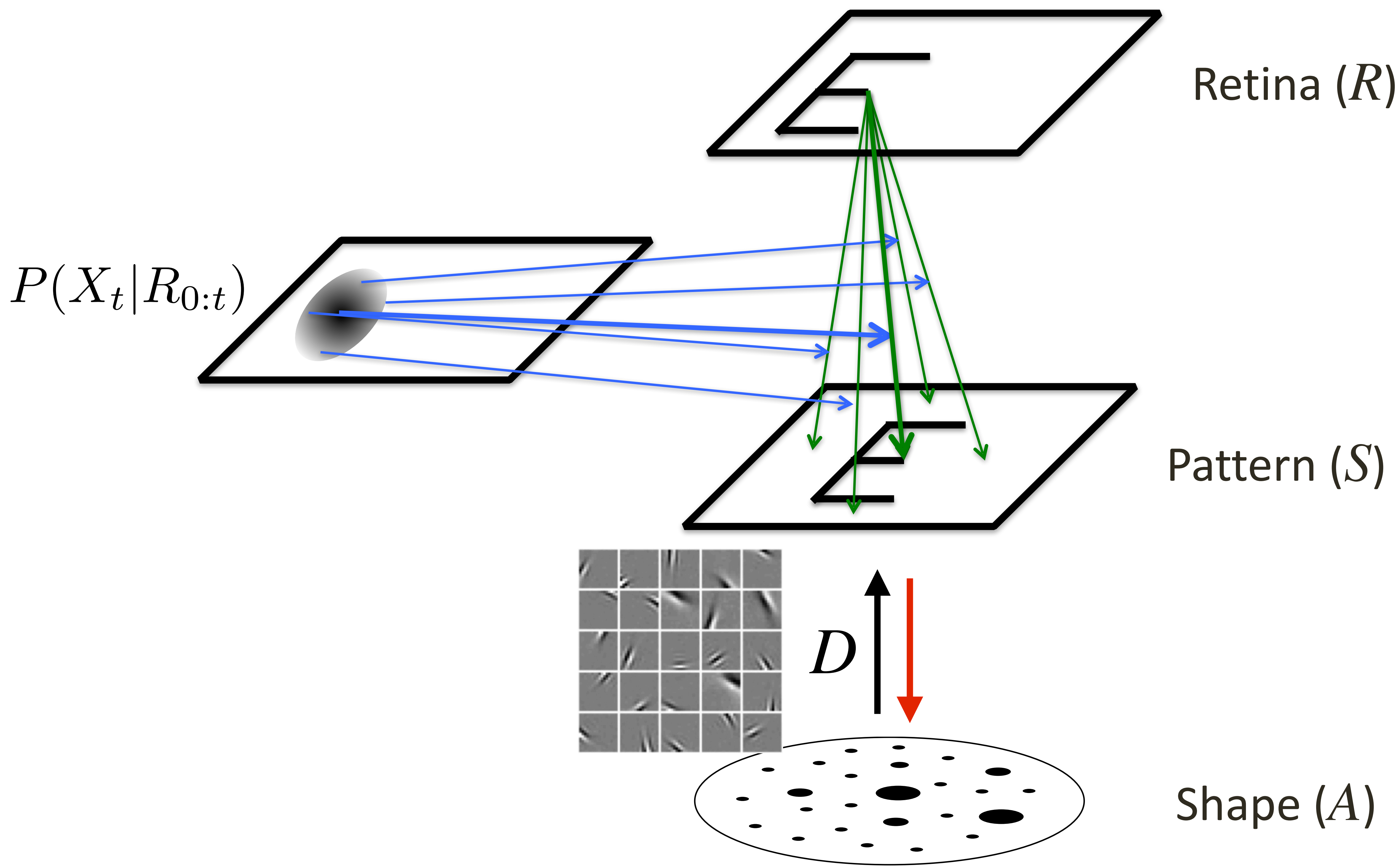
$$\log P(R_{j,t}|X_t, S) = R_{j,t}\, \log \lambda_j - \lambda_j\, dt$$

$$\log \lambda_j = \sum_{\vec{x}} g_j(\vec{x})\, S(\vec{x} - \vec{X}_t)$$

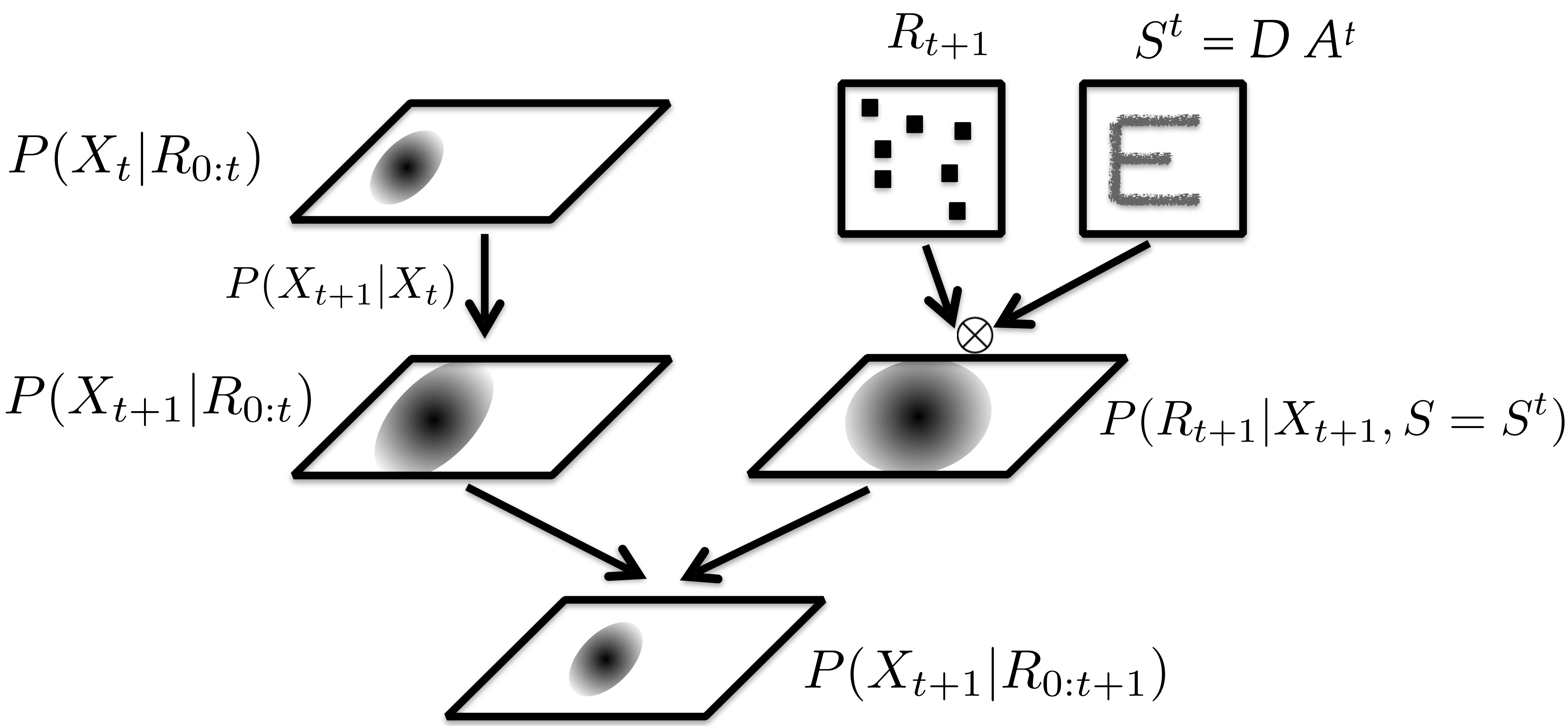
Given current estimate of shape ($\hat{A}^t$), update X

$$P(X_t|S^t, R_{0:t}) \propto P(R_t|X_t, S^t = D\,\hat{A}^t) \sum_{X_{t-1}} P(X_t|X_{t-1})\, P(X_{t-1}|S^{t-1}, R_{0:t-1})$$

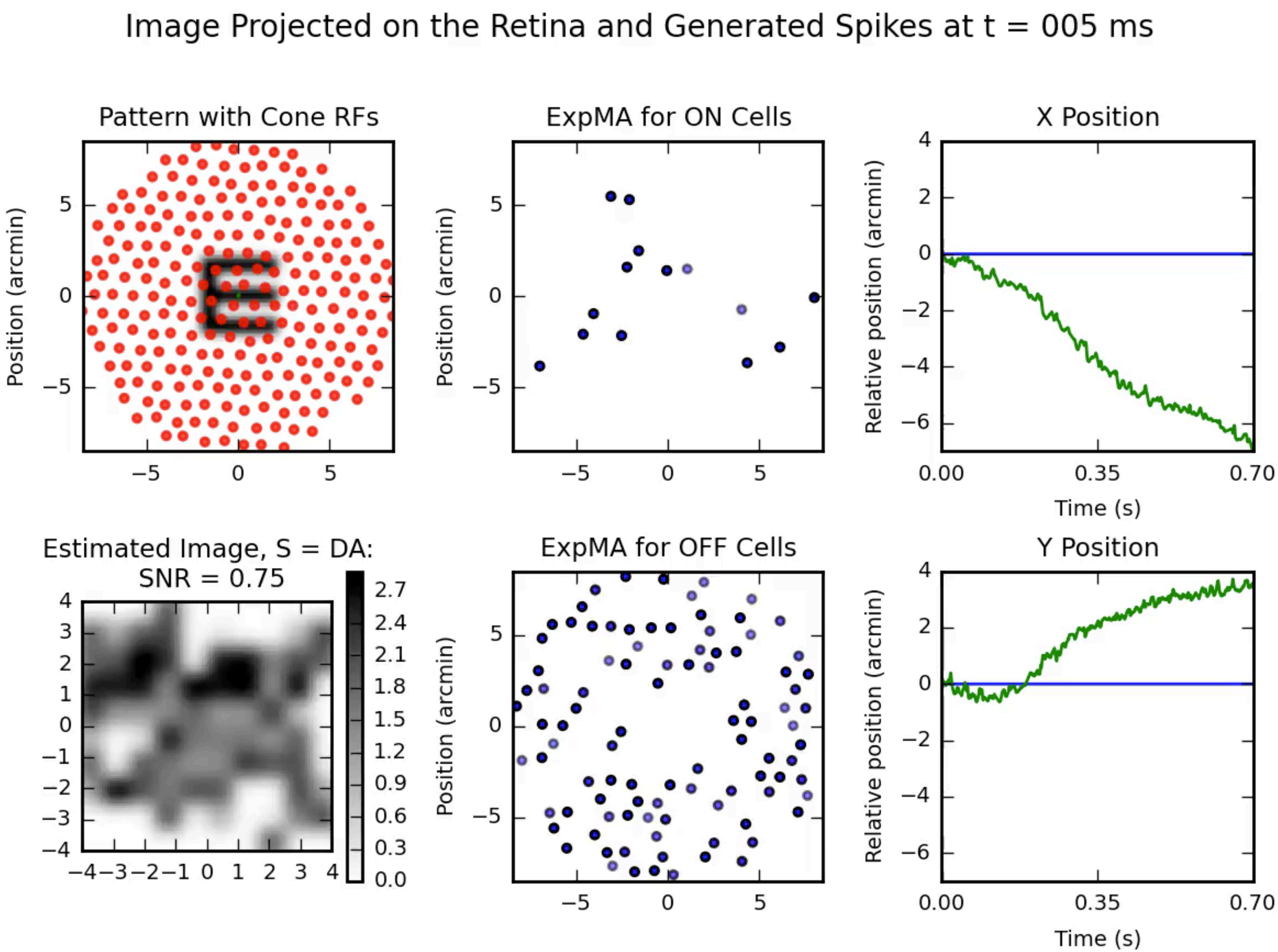
Given current estimate of position (X), update A



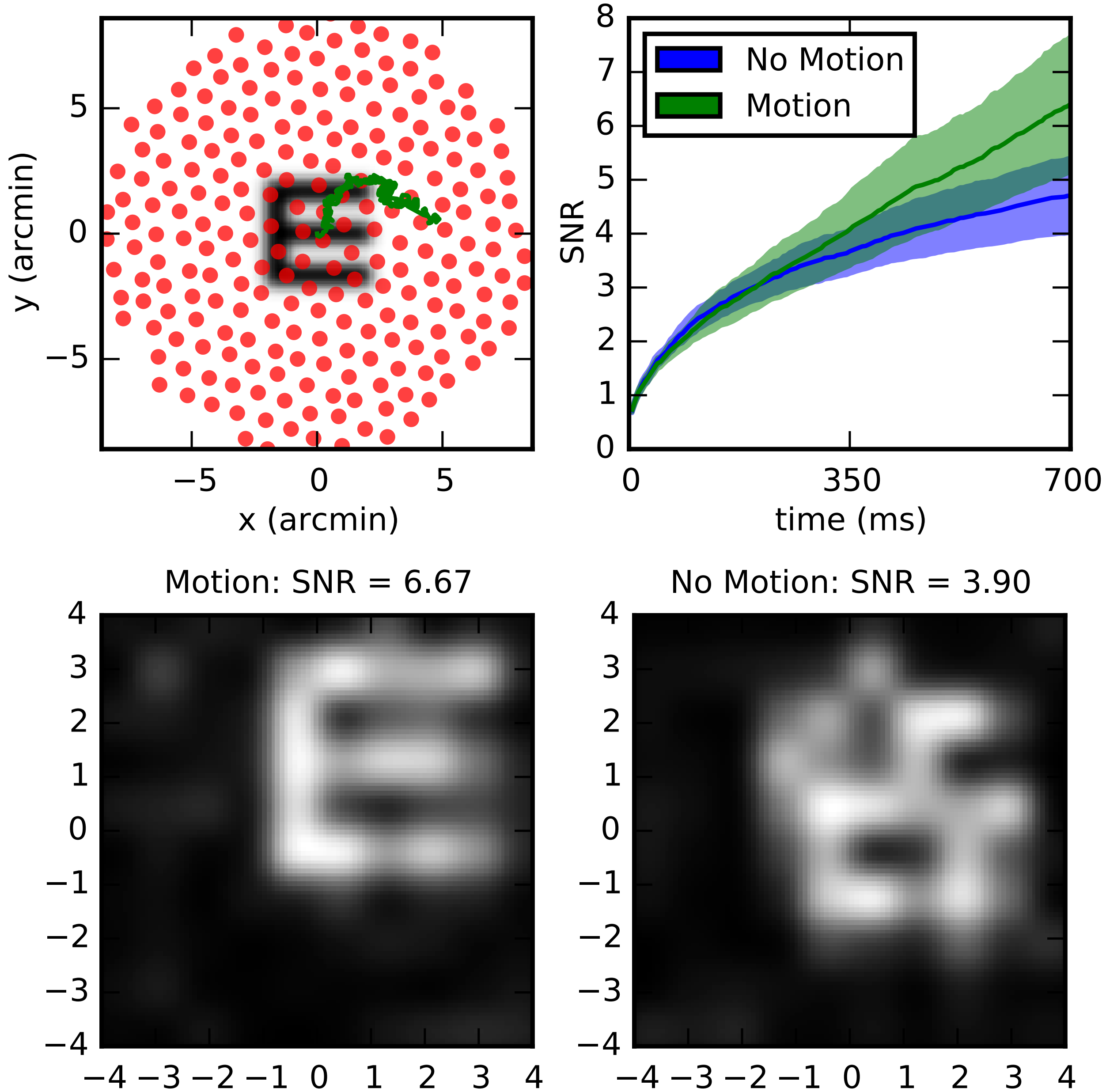
Given current estimate of pattern (A), update X



Joint estimation of pattern (S) and position (X)

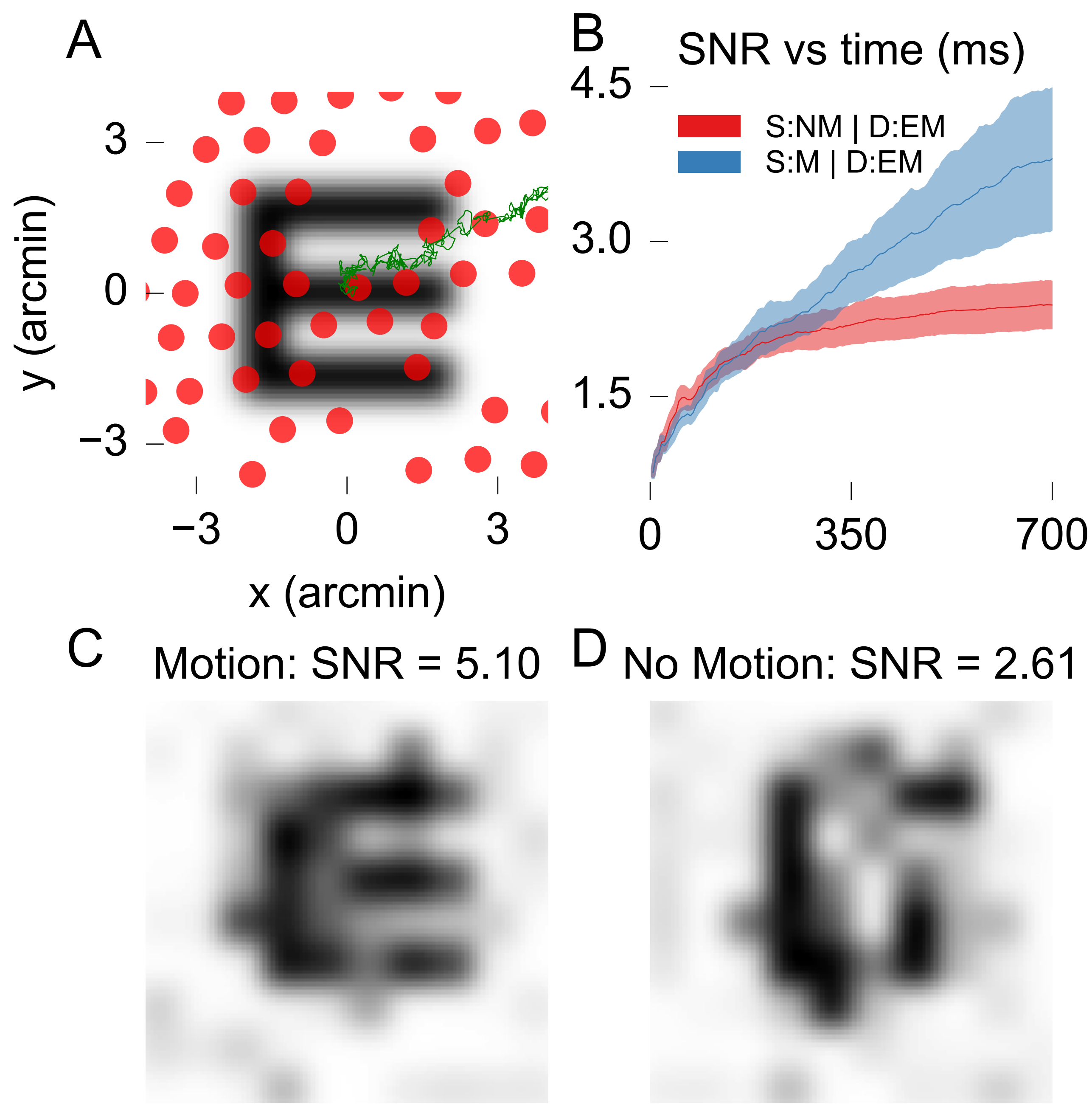


Motion helps estimation of pattern *S*

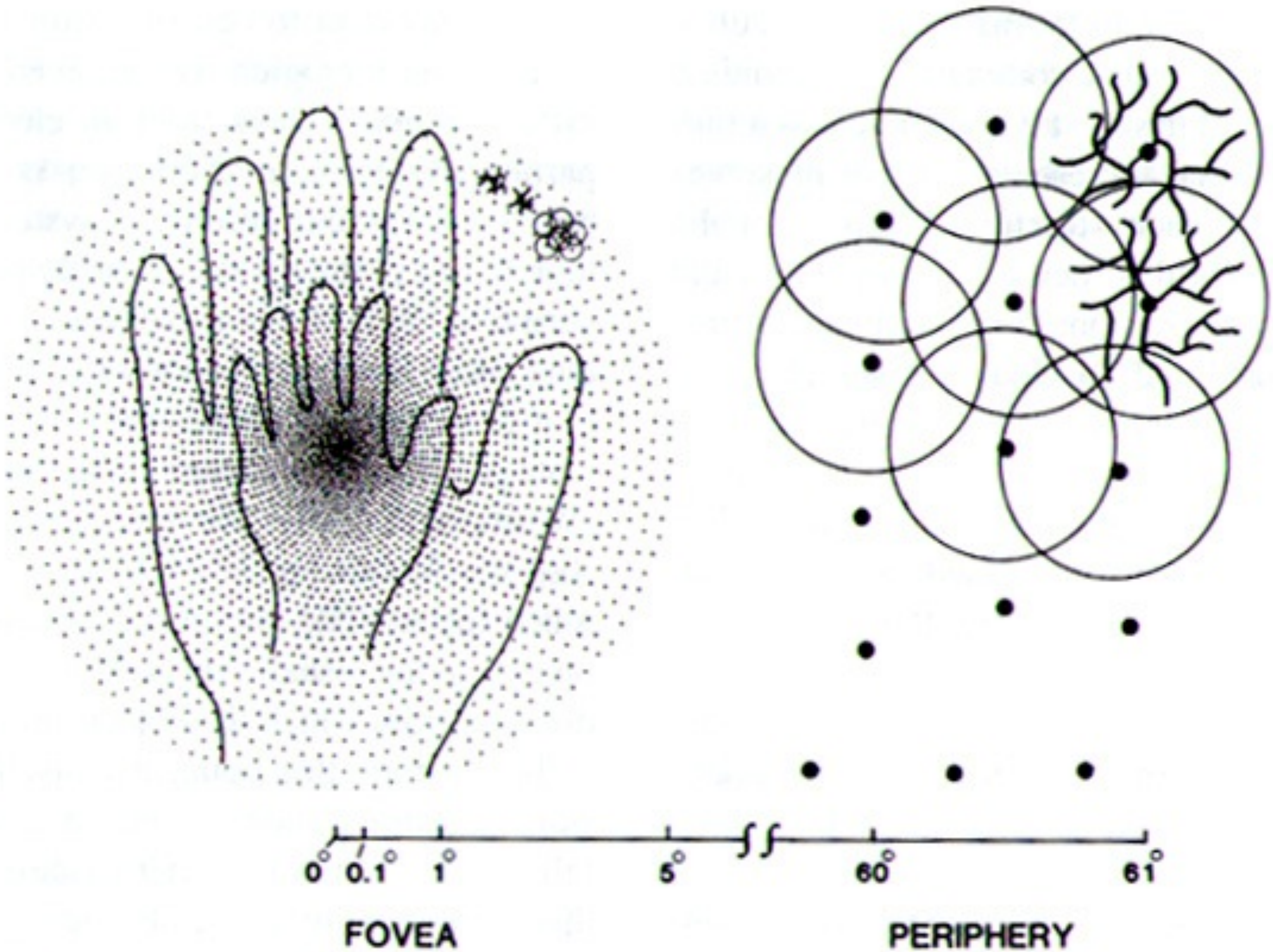


...especially under conditions of cone loss.

simulated
30% dropout

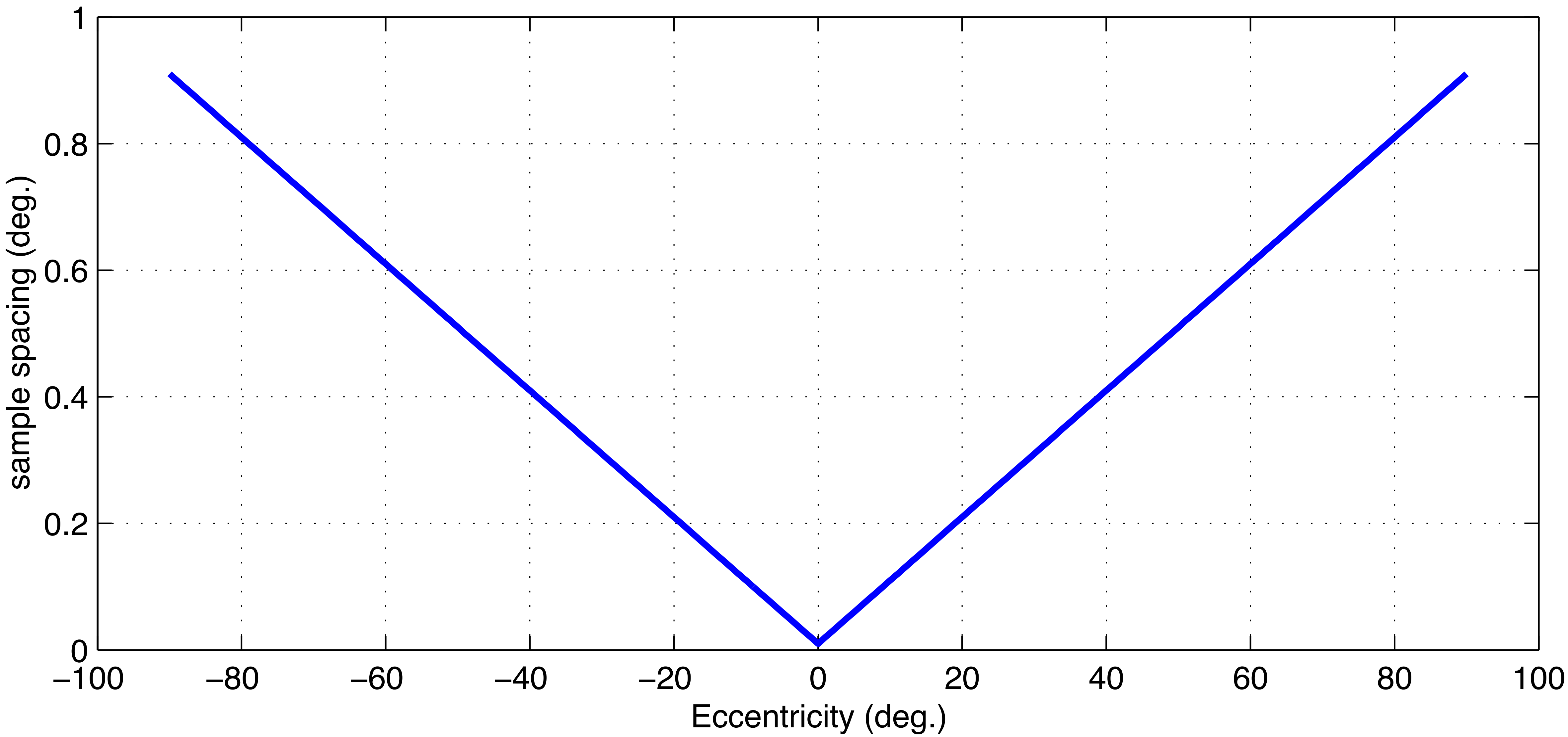


Foveated sampling

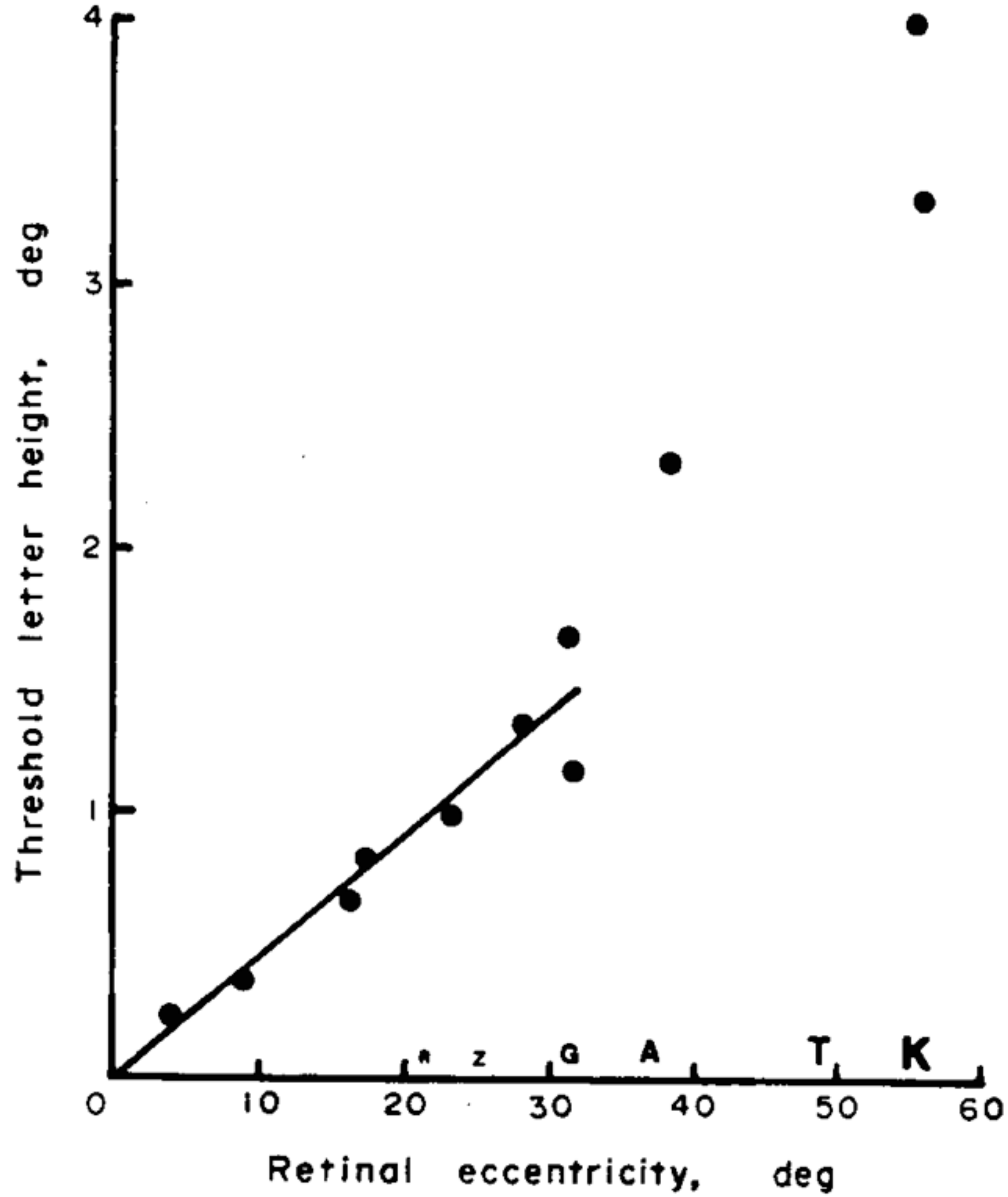


Retinal ganglion cell spacing as a function of eccentricity

$$\Delta E \approx .01(|E| + 1)$$



Letter size vs. eccentricity
(Anstis, 1974)



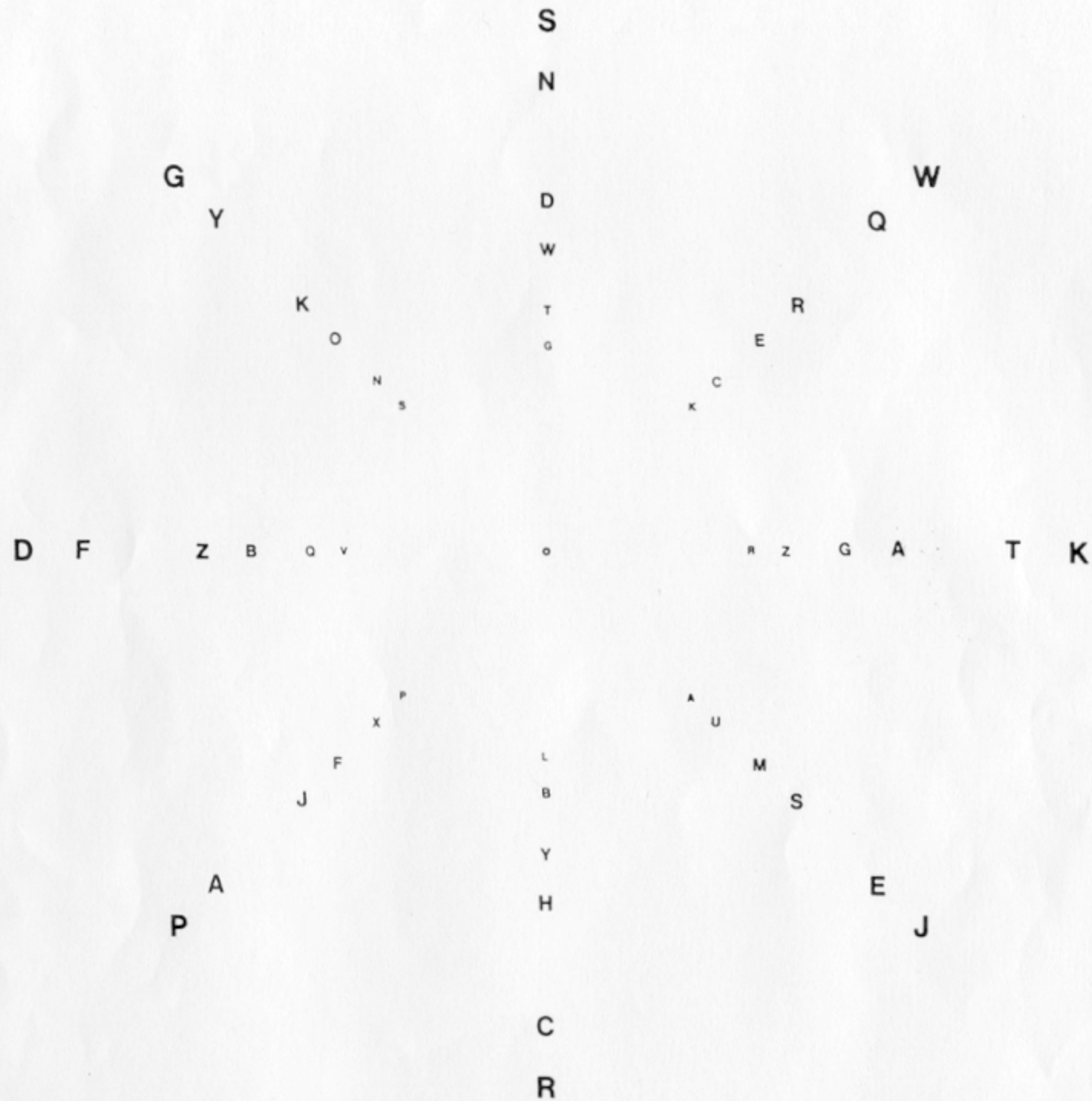


Fig. 2. All letters should lie at threshold when centre of this chart is fixated. Threshold letter size increases linearly with increasing distance from fixation point.

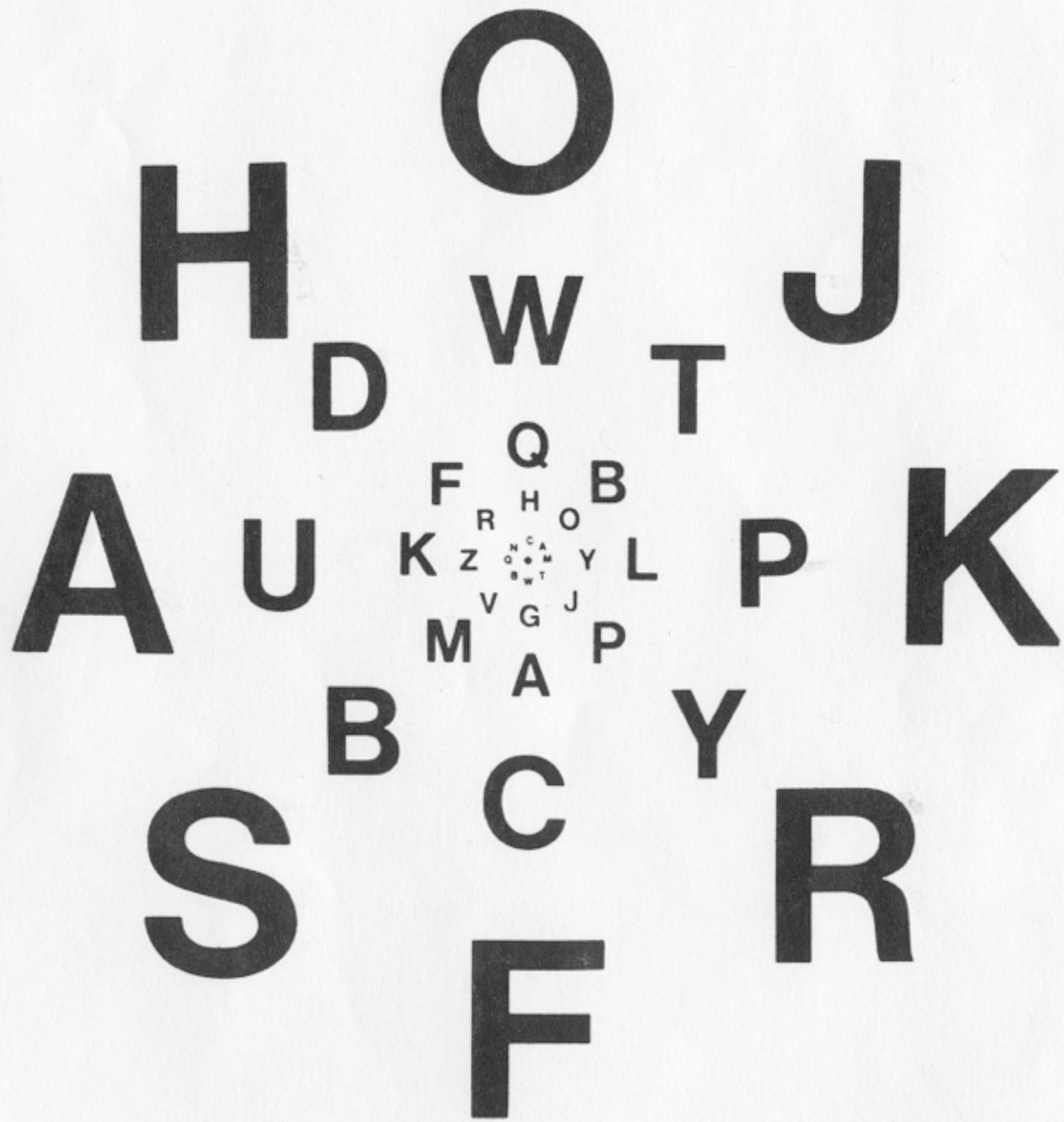


Fig. 3. All letters should be equally readable when centre of this chart is fixated, since each letter is ten times its threshold height.

A FOVEATED RETINA-LIKE SENSOR USING CCD TECHNOLOGY

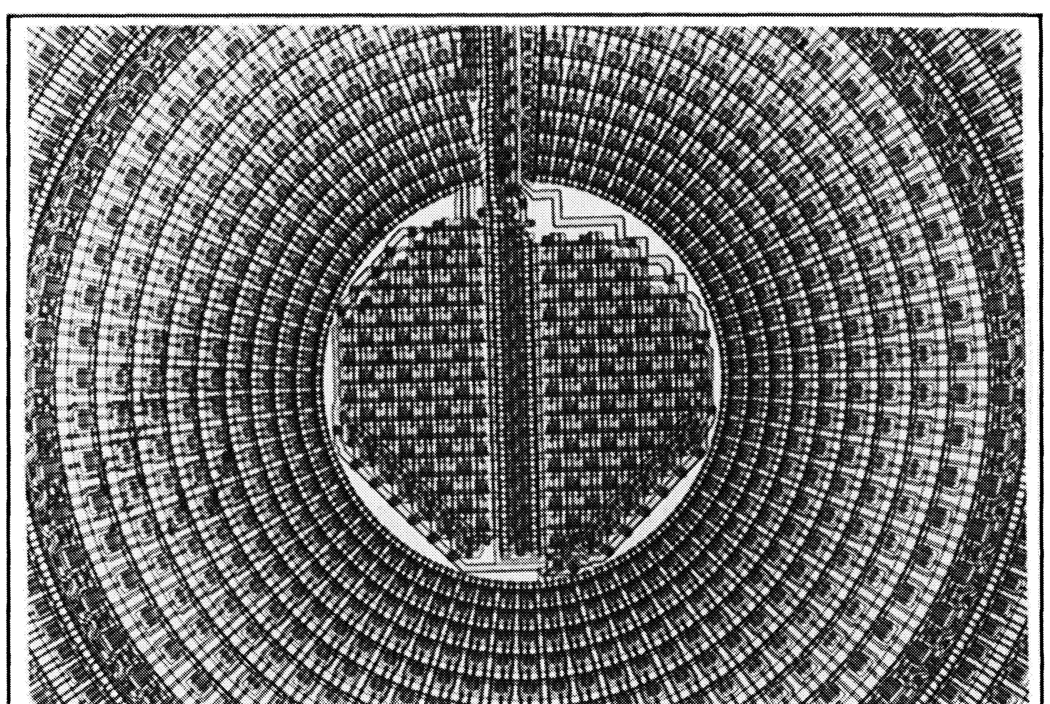
J. Van der Spiegel, G. Kreider
Univ. of Pennsylvania, Dept. of Electrical Engineering
Philadelphia, PA 19104-6390

C. Claeys, I. Debusschere
IMEC, Leuven, Belgium

G. Sandini
University of Genova, DIST, Genova, Italy

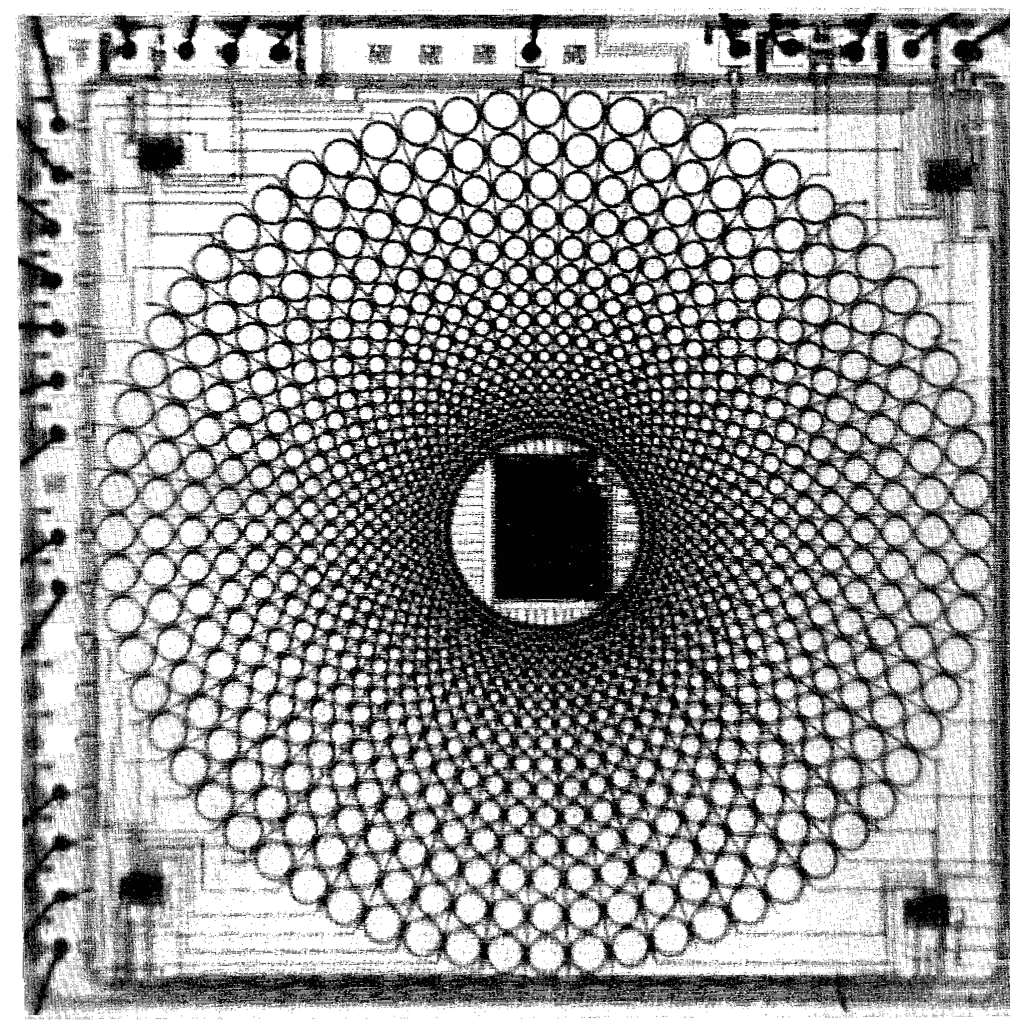
P. Dario, F. Fantini
Scuola Superiore S. Anna, Pisa, Italy

P. Bellutti, G. Soncini
IRST, Trento, Italy

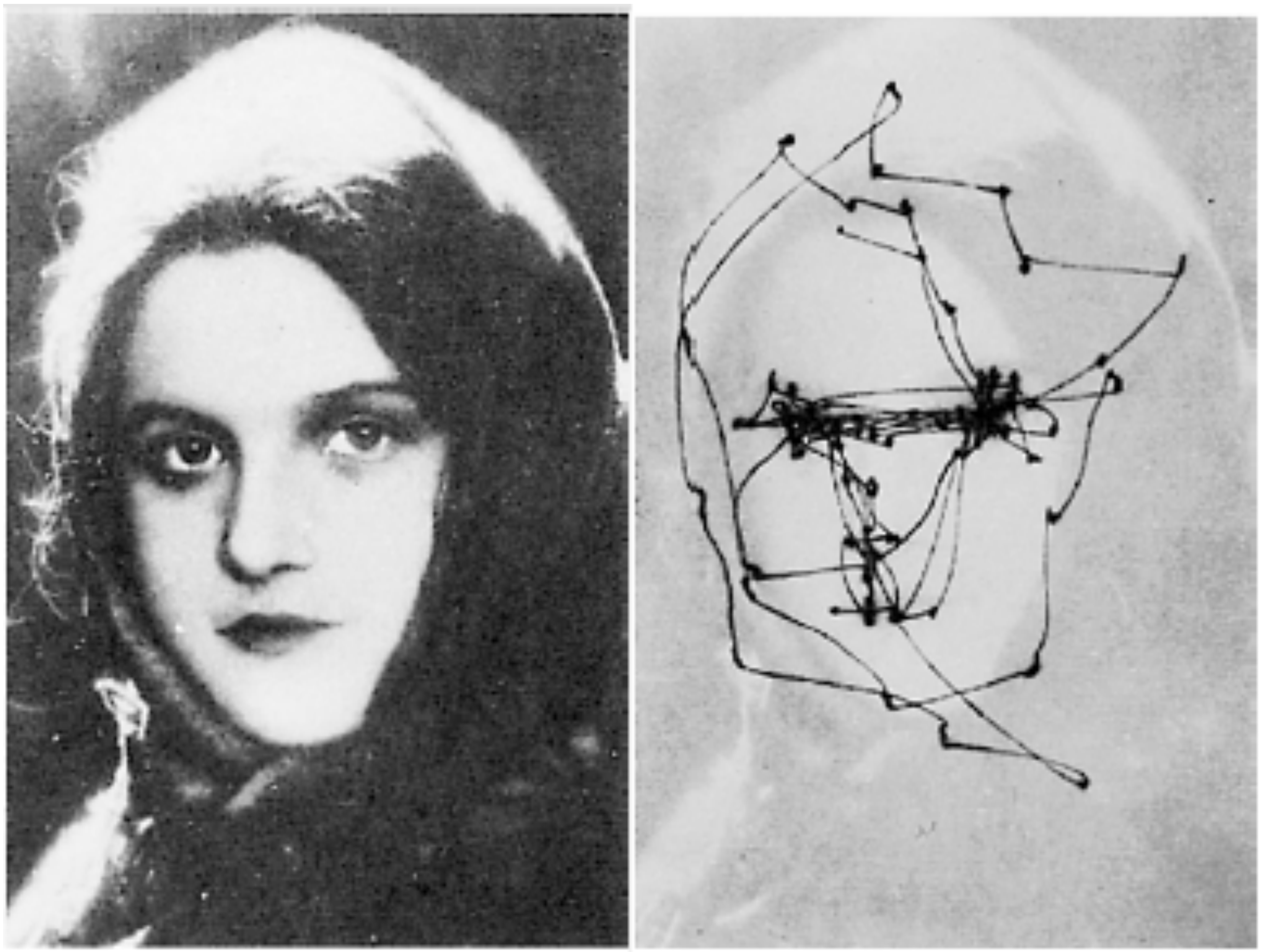


A Foveated Image Sensor in Standard CMOS Technology

Robert Wodnicki, Gordon W. Roberts, Martin D. Levine
Department of Electrical Engineering, McGill University,
Montréal, Québec, CANADA, H3A 2A7



Human eye movements during viewing of an image



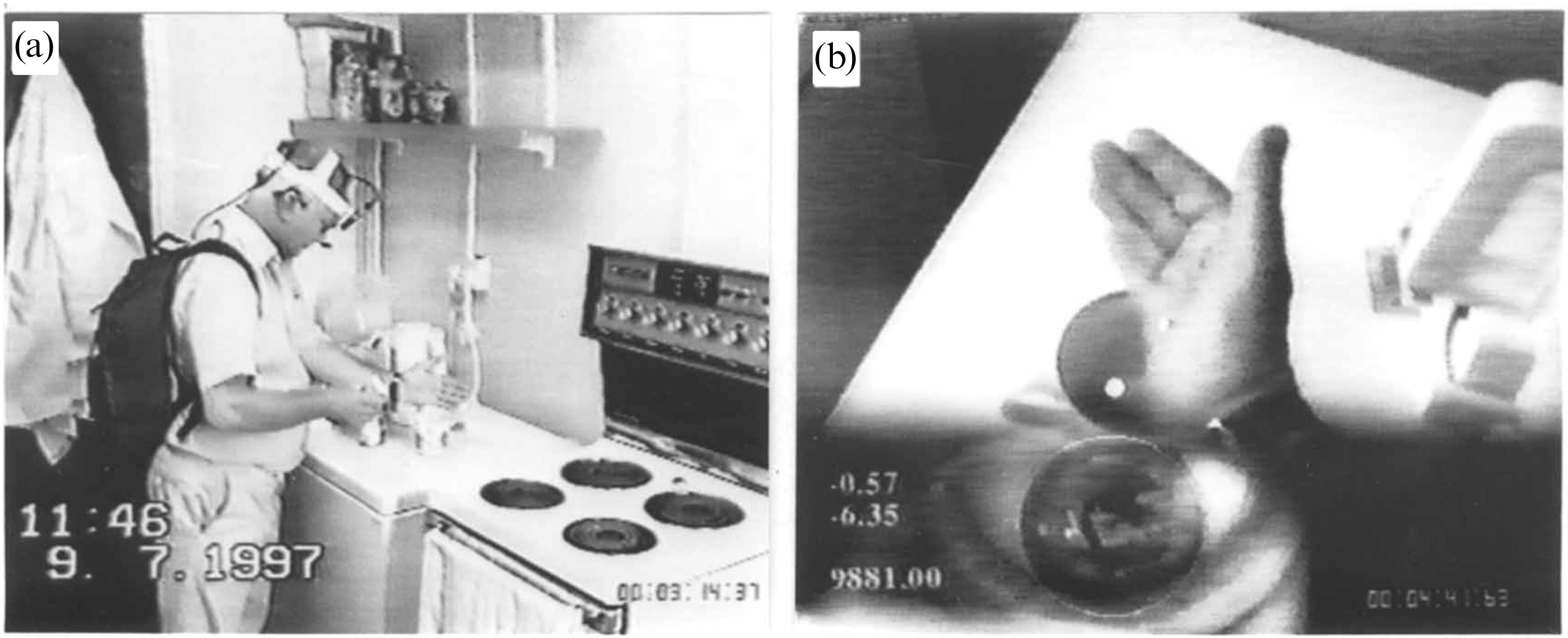
Yarbus (1967)

Perception, 1999, volume 28, pages 1311 – 1328

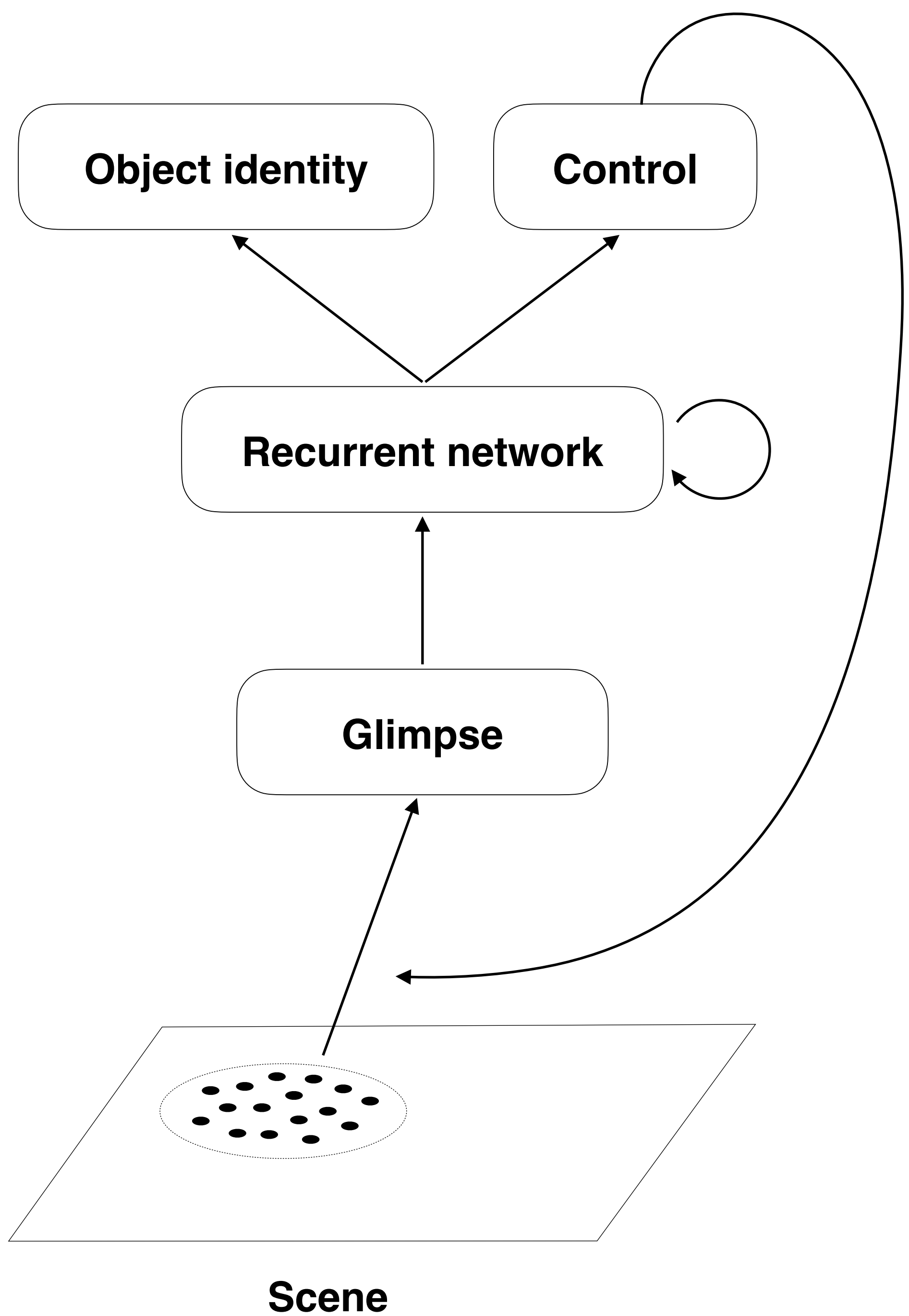
DOI:10.1068/p2935

The roles of vision and eye movements in the control of activities of daily living

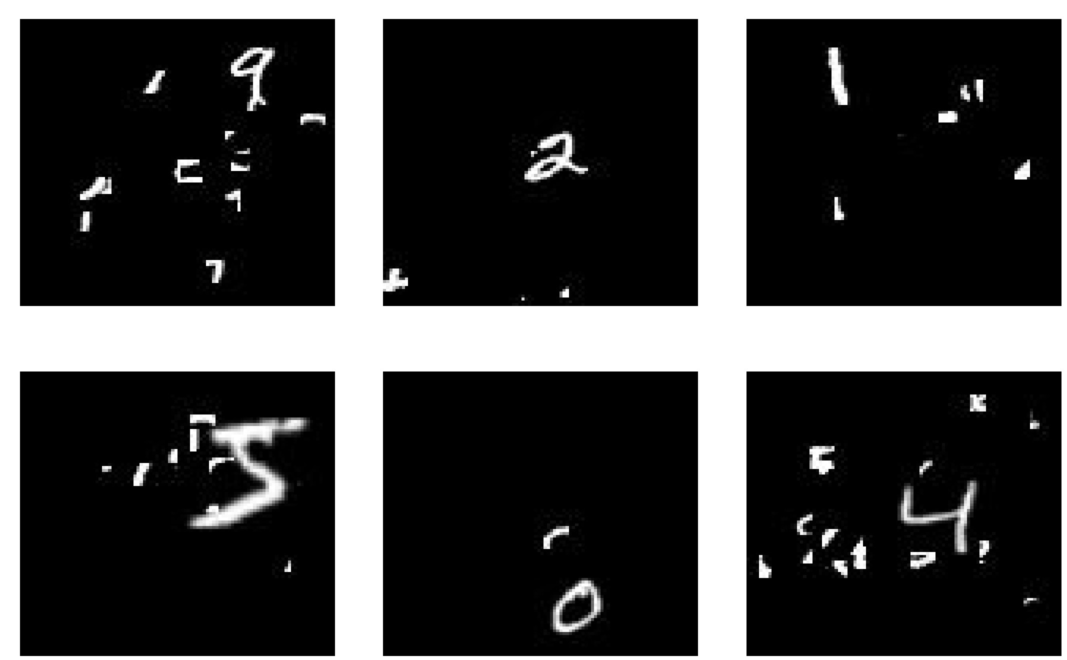
Michael Land, Neil Mennie, Jennifer Rusted
Sussex Centre for Neuroscience and Laboratory of Experimental Psychology, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK; e-mail: M.F.Land@sussex.ac.uk
Received 4 May 1999, in revised form 9 August 1999



Learning the glimpse window sampling array
(Cheung, Weiss & Olshausen, 2017)

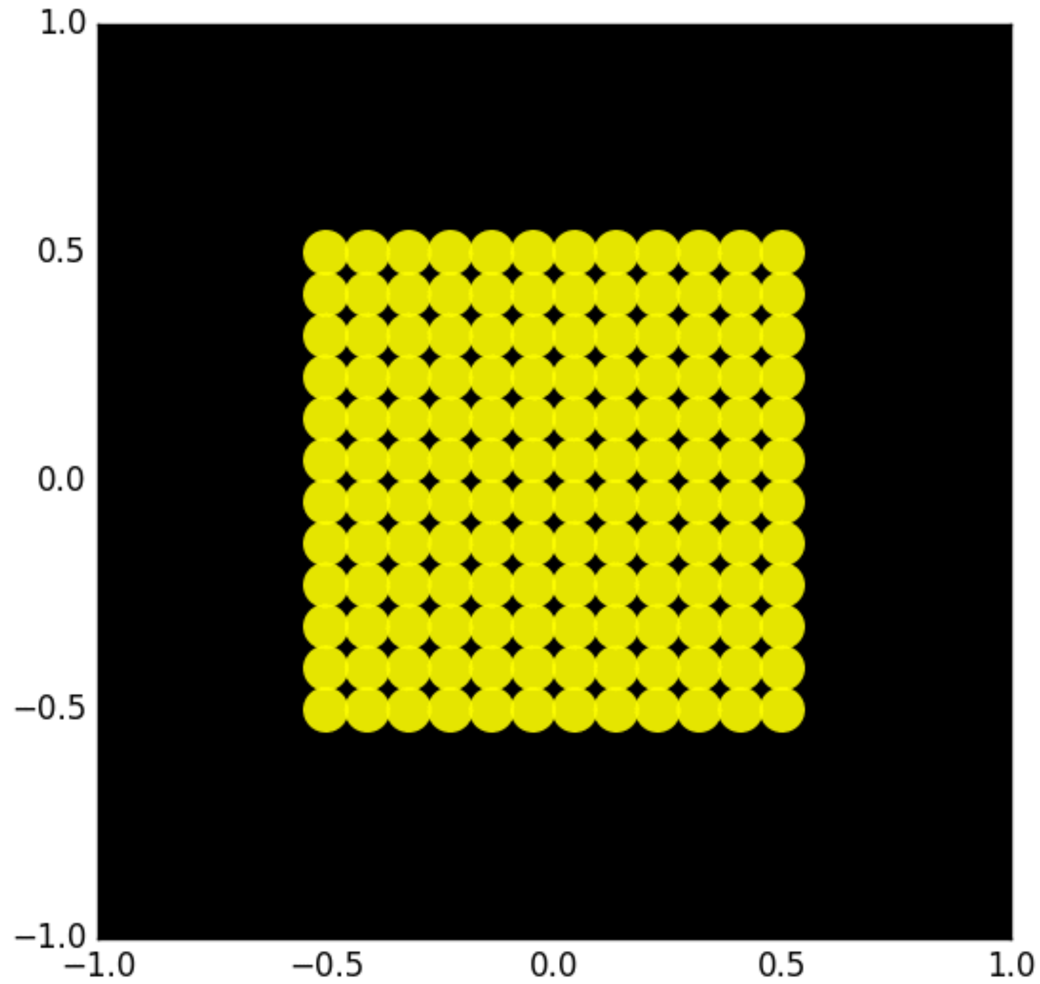


- Network is trained to correctly classify the digit in the scene.
- To do this it must find a digit and move its glimpse window to that location.

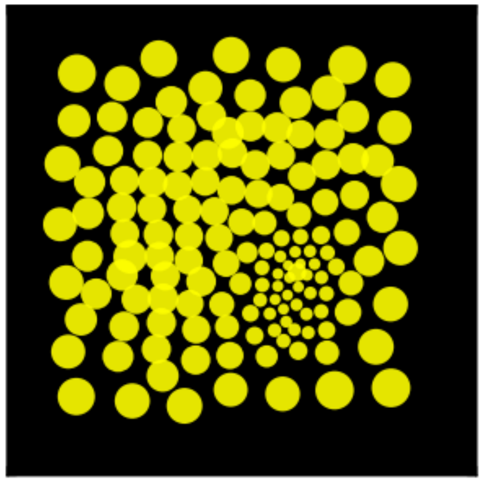


Example MNIST scenes

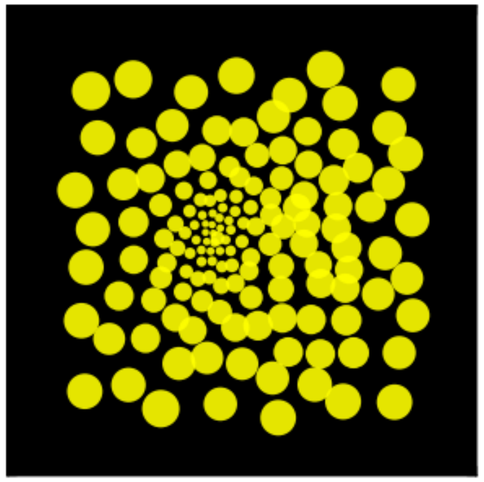
Evolution of the sampling array during training



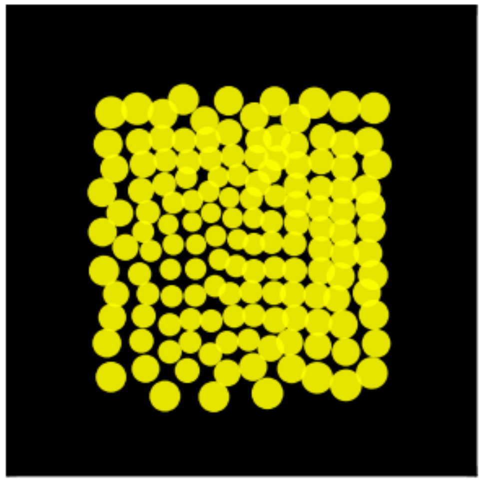
Learned sampling arrays for different conditions



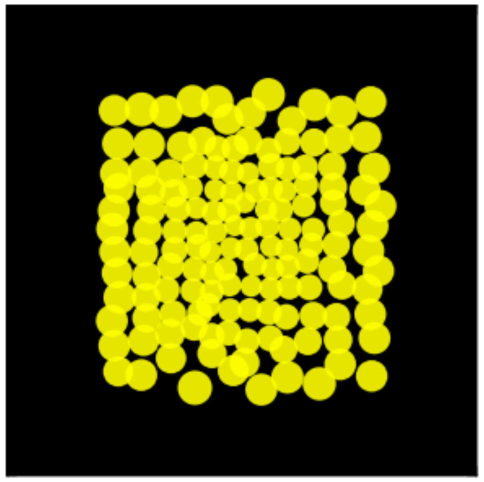
Translation only
(Dataset 1)



Translation only
(Dataset 2)



Translation & zoom
(Dataset 1)

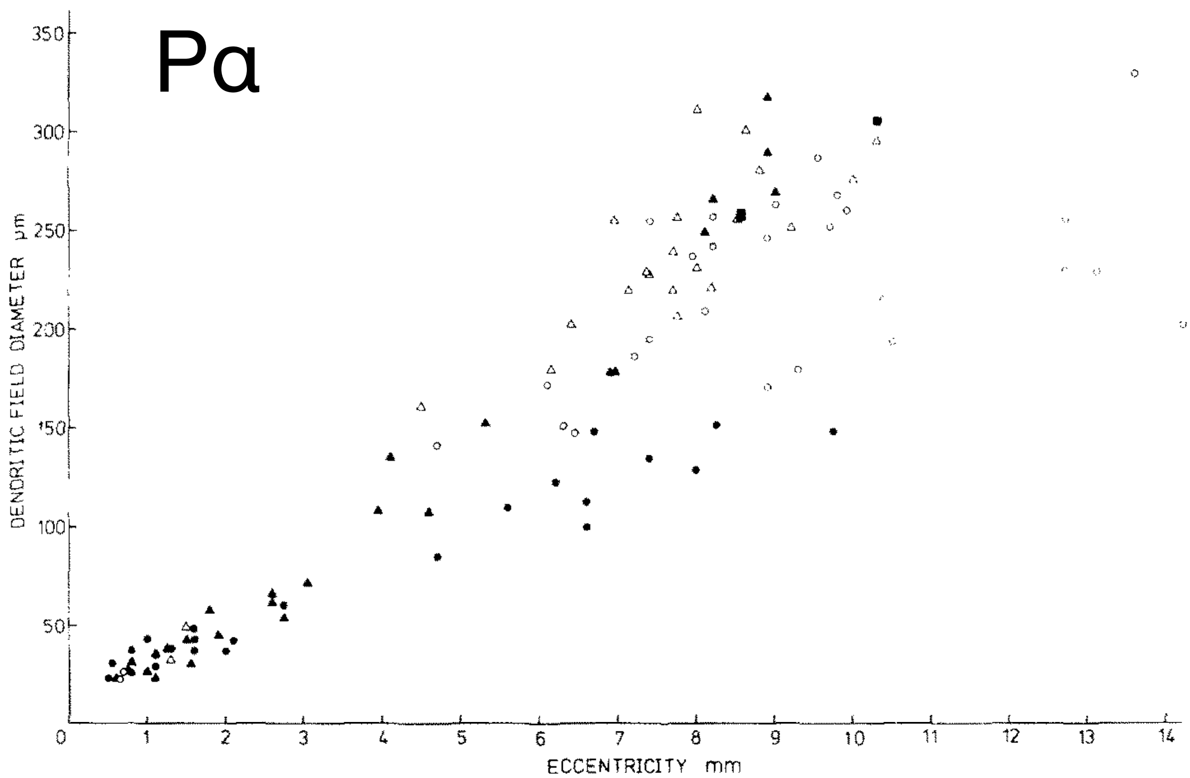


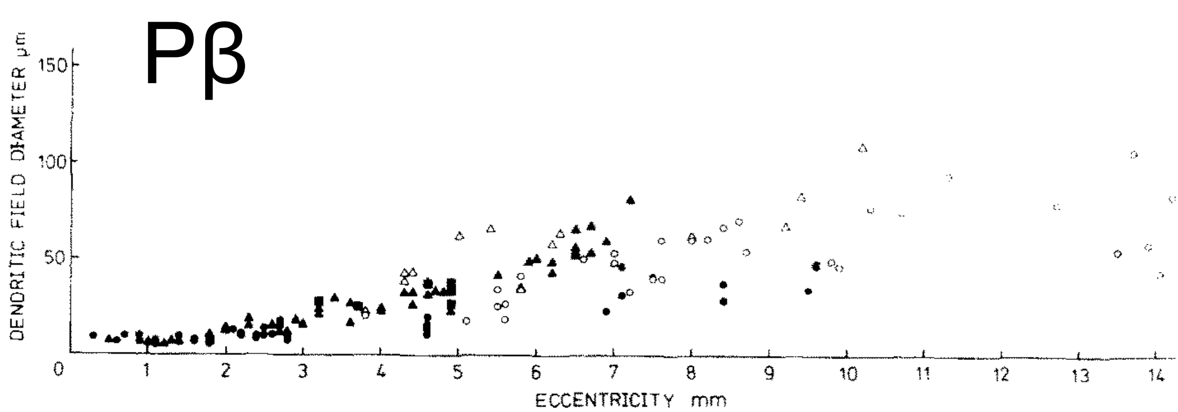
Translation & zoom
(Dataset 2)

Comparison to primate retina

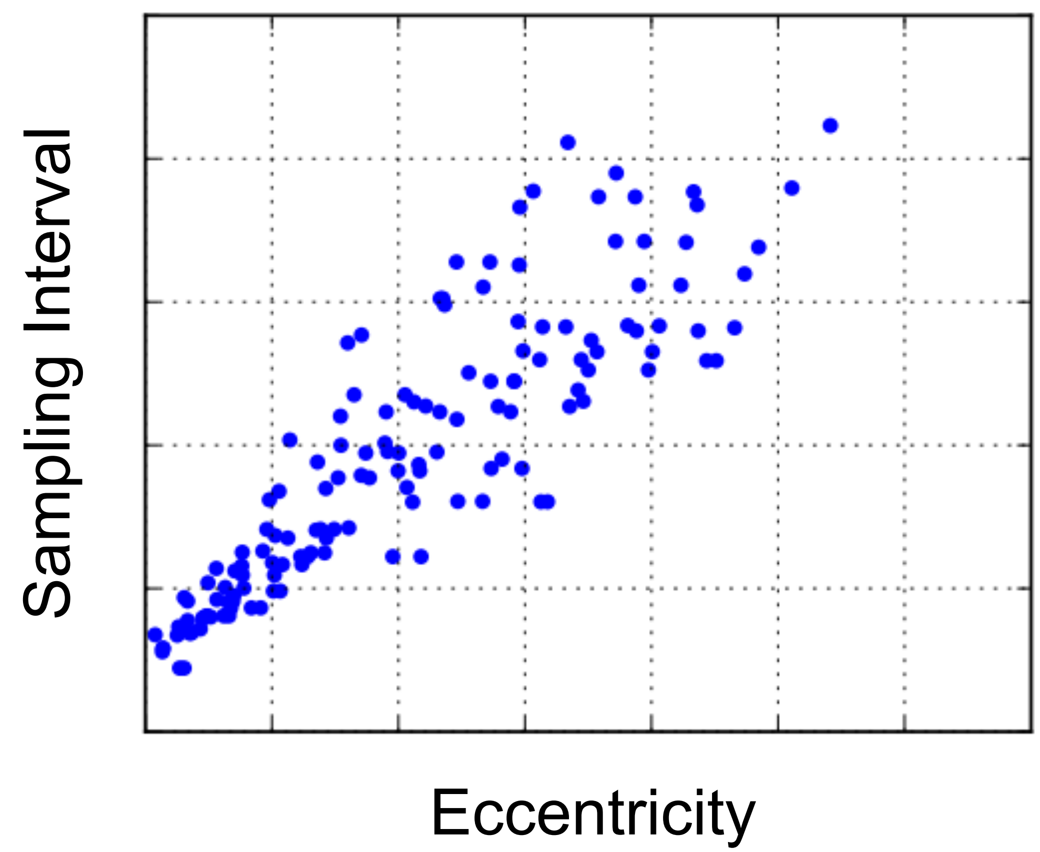
Macaque Retina

(Perry, Oehler & Cowey, 1984)



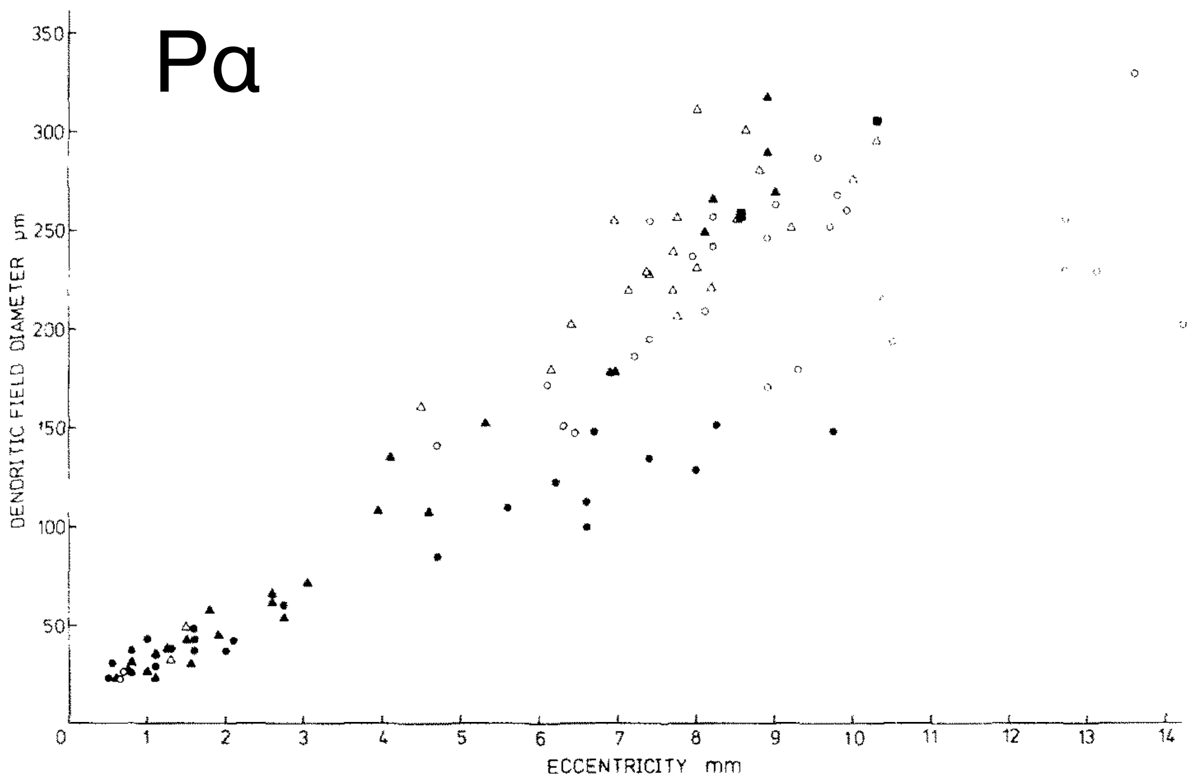


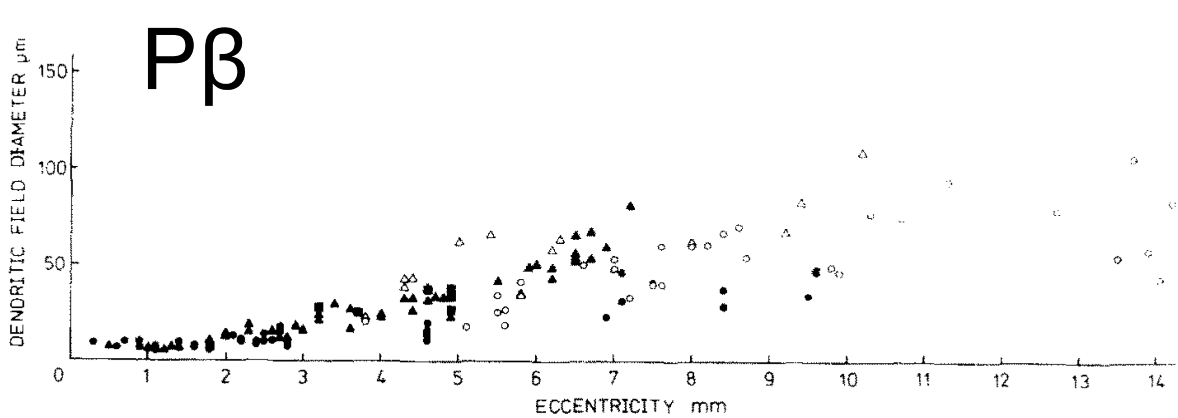
Model



Comparison to primate retina

Macaque Retina
(Perry, Oehler & Cowey, 1984)





Model