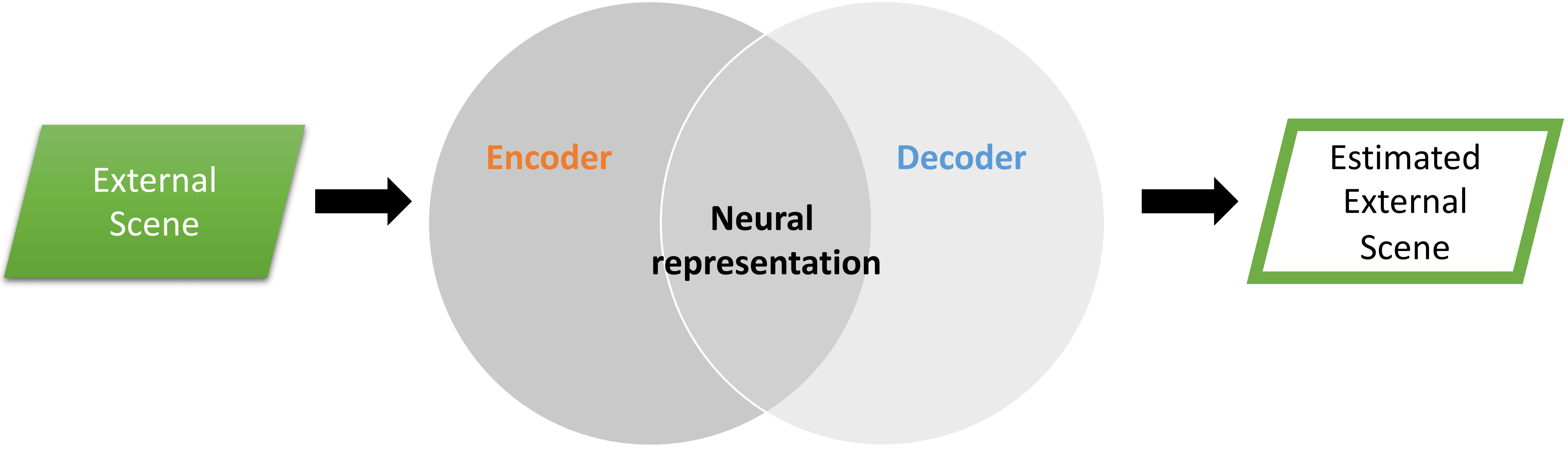
Encoding and decoding visual information in the presence of eye movements

Janis Intoy

April 1, 2022



External Scene

Encoder

Neural representation

Decoder

Estimated External Scene

Encoding Process

How is information represented in the visual system?

How much information is maintained? By what mechanisms is information efficiently encoded (compressed)?

Decoding Process

By what mechanisms are the external scenes estimated? How is a stable visual representation achieved?

What internal representations are needed for the decoding process? Is an internal estimate of retinal image motion required? Is retinal information enough?

Big Picture Questions

1. We study space-time encoding. What about decoding?
    - Current decoding models describe <u>spatial mechanisms</u> that disentangle position and image (luminance pattern)
    - Temporal mechanisms?

2. Active processes in 3D vision
    - Form (depth pattern) arises from position

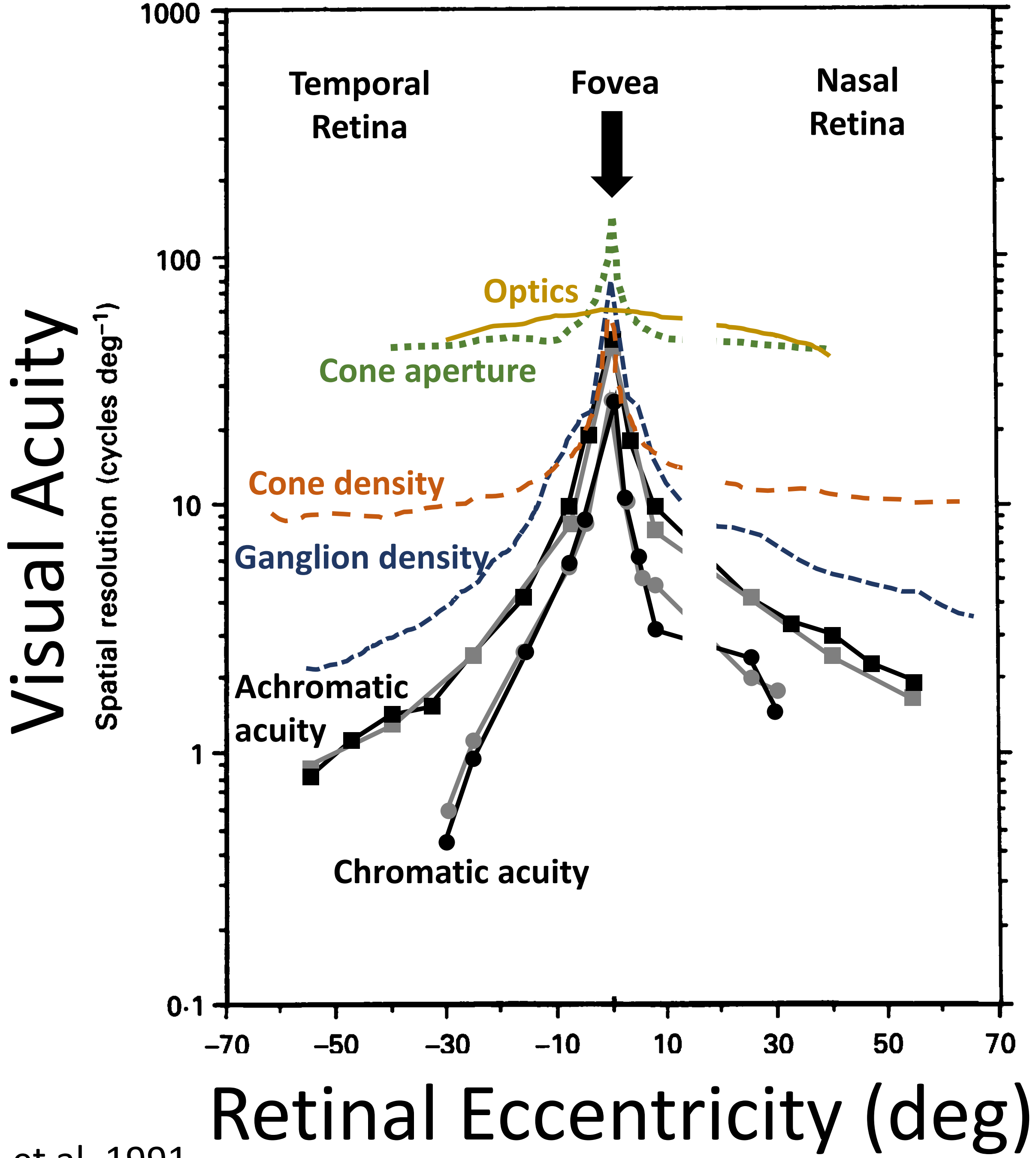
Encoding

Spatial Encoding in the Retina

Visual information is limited by the spatial components of the eye & retina.

The chromatic and achromatic acuity measured in two observers (Anderson et al, 1991) is compared to the the maximum spatial resolution afforded by:

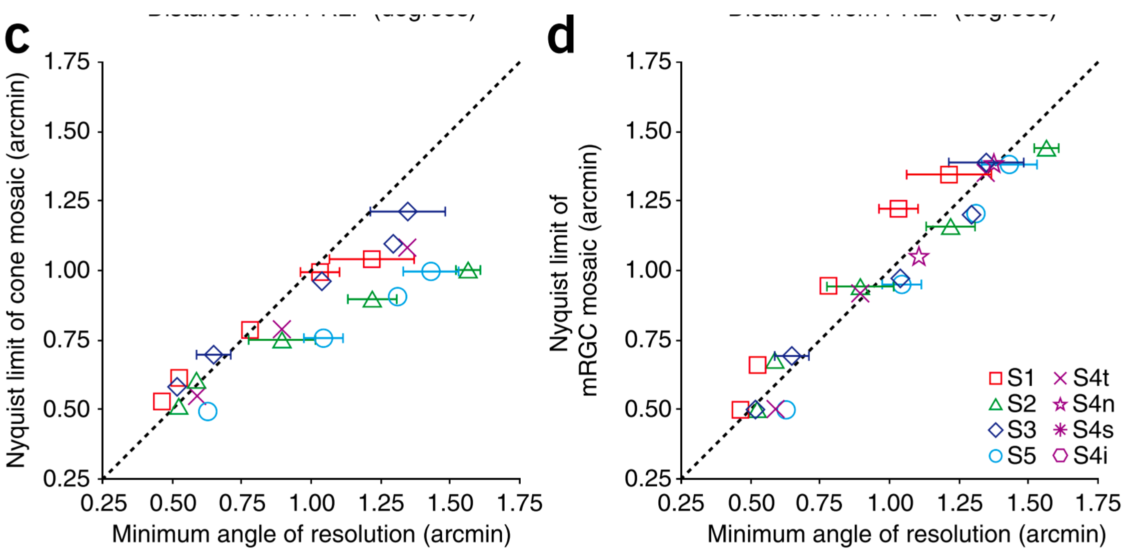
- the optical properties of the eye (Campbell & Gubisch, 1966; Jennings & Charman, 1981)
- aperture size of cones (Curcio)
- Nyquist limits of cone density (Curcio et al, 1990)
- Nyquist limits of of ganglion cell density (Curcio & Allen, 1990)



Adapted from Anderson et al, 1991

Spatial Encoding in the Retina

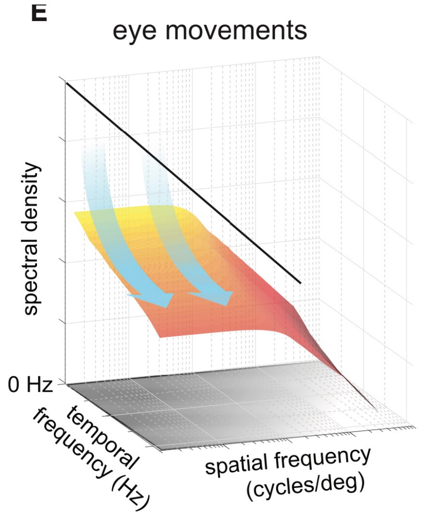
Rossi & Roorda (2010) compared acuity in the fovea to the Nyquist limits of the AO-imaged cone mosaic and the modelled mRGC mosaic.



Rossi & Roorda, 2010

Space-Time Encoding in APLab

Retinal Input

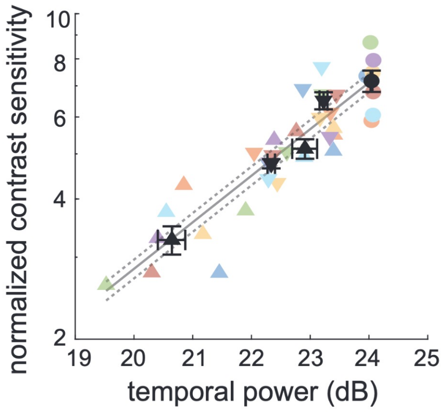


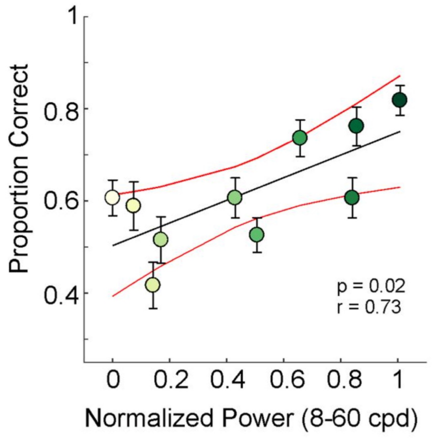
Visual information is encoded in **temporal signals** (e.g. input to neurons in time or neural responses).

Theory: (Kuang et al, 2012; Mostofi, Zhao et al, 2020)
- Reduces spatial redundancies in the retinal input
- Saccades and drift selectively encode low and high frequency information respectively

Experimental Support:
- Saccades and drift selectively enhance contrast sensitivity to low and high frequency stimuli (Rucci et al, 2007; Boi et al, 2017)

Visual performance is correlated with the power of temporal signals.





Changes in temporal power predicts:
- Changes in contrast sensitivity of gratings during drift [JI]
- Differences in individual acuity during drift [AMC]
- Changes in perceived contrast of gratings during saccades [HL]
- Visibility of gratings during blinks and smooth pursuit [BY]

The decoding process benefits from improvements in encoding.

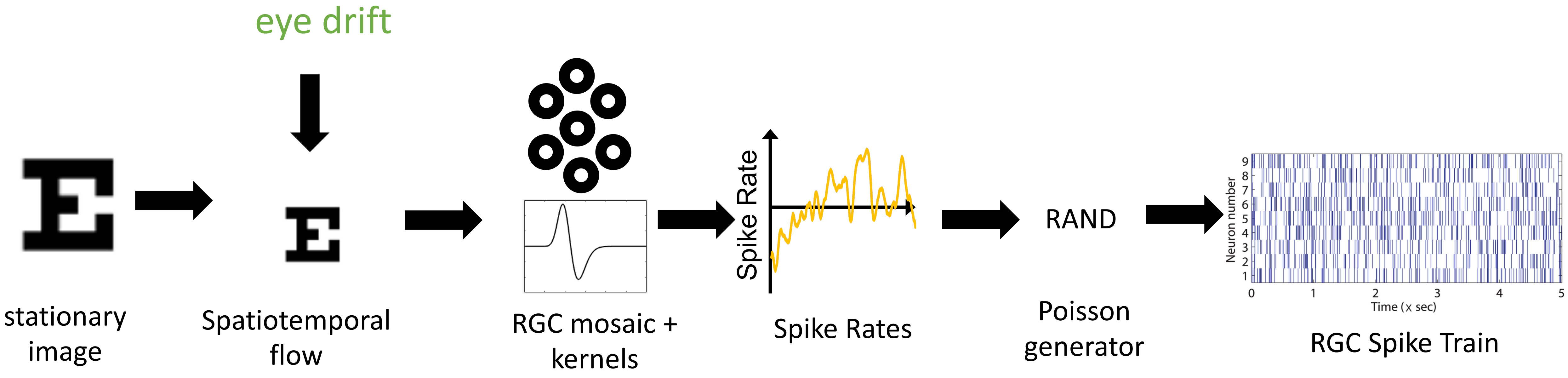
Decoding

Decoder Models

Models of high-acuity vision in the presence of ocular drift. What visual information is encoded? How is the image decoded from neural representations?

1. Pitkow, Sompolinsky, Meister (2007). **A neural computation for visual acuity in the presence of eye movements**. PLOS Biology 5(12):2898-2911.
2. Burak, Rokni, Meister, Sompolinsky (2010). **Bayesian model of dynamic image stabilization in the visual system**. PNAS 107(45):19525-19530.
3. Anderson, Ratnam, Roorda, Olshausen (2020). **High-acuity vision from retinal image motion**. JoV 20(7):1-19.

Simulation of the retina (encoding)



stationary image

Spatiotemporal flow

RGC mosaic + kernels

Spike Rates

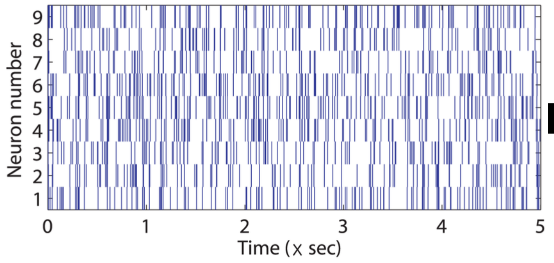
Poisson generator

RGC Spike Train

Key differences between studies:

	Input	RGC Mosaic	RGC kernels	Eye drifts
1. Pitkow et al	High-acuity image (horz or vert bar)	Discrete, square lattice Centered at image pixels	Spatial and temporal -effect of temporal kernel	Discrete random walk (pixel to pixel)
2. Burak et al	Black and white image (not necessarily high-acuity)	Discrete, square lattice Centered at image pixels	Spatial and temporal -effect of temporal kernel	Discrete random walk (pixel to pixel)
3. Anderson et al	High-acuity image (mostly Es)	Realistic hexagonal grid with spatial jitter	Spatial only	Continuous random walks and real drifts (Ratnam et al, 2017)

The encoding and decoding process



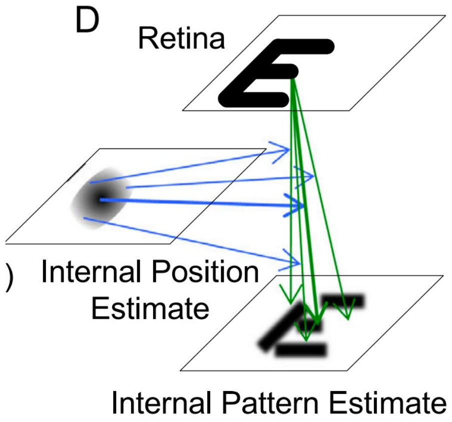
RGC Spike Train

Joint inference of image position on retina and image estimate over time

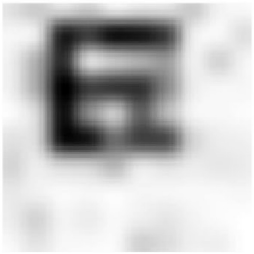
Time (new spikes)

Spike Train at single time

Where: internal representation of position



What: internal representation of image



Estimated Image

Where: Cross-correlate current image estimate with incoming spikes to determine where the image is now.

What: Shift the retinal image to some internal reference frame and update the internal pattern.

Decoder Models

Key differences between studies:

	Internal Estimates	**Position Assumptions**	**Decoder Output**
1. Pitkow et al	Stimulus position	Initial position unknown	2-AFC discrimination (only 2 possible images)
2. Burak et al	Stimulus position	Initial position known	Pixel-by-pixel estimate of binary image Same resolution as original image
3. Anderson et al	Stimulus position and Uncertainty of Stimulus Weights	Initial position known	Weights of sub-images (e.g. oriented edges) Higher resolution than original image

Decoder model details

X = eye position

A = image estimate

R = spike train

Infer position and image at each time:

1. Update position distribution

$$\underbrace{p(X_{t+1}|R, A_t)} \propto \underbrace{p(R_{t+1}|X_{t+1}, A_t)}_{\text{Likelihood of spikes}} \underbrace{\sum_{X_t} p(X_{t+1}|X_t)\, p(X_t|R_{0:t}, A_t)}_{\text{Position prior (Brownian)}}$$

Anderson et al (2020)

Decoder model details

X = eye position

A = image estimate

R = spike train

Infer position and image at each time:

1. Update position distribution

$$\underbrace{p(X_{t+1}|R, A_t)}_{} \propto \underbrace{p(R_{t+1}|X_{t+1}, A_t)}_{\text{Likelihood of spikes}} \underbrace{\sum_{X_t} p(X_{t+1}|X_t)p(X_t|R_{0:t}, A_t)}_{\text{Position prior (Brownian)}}$$

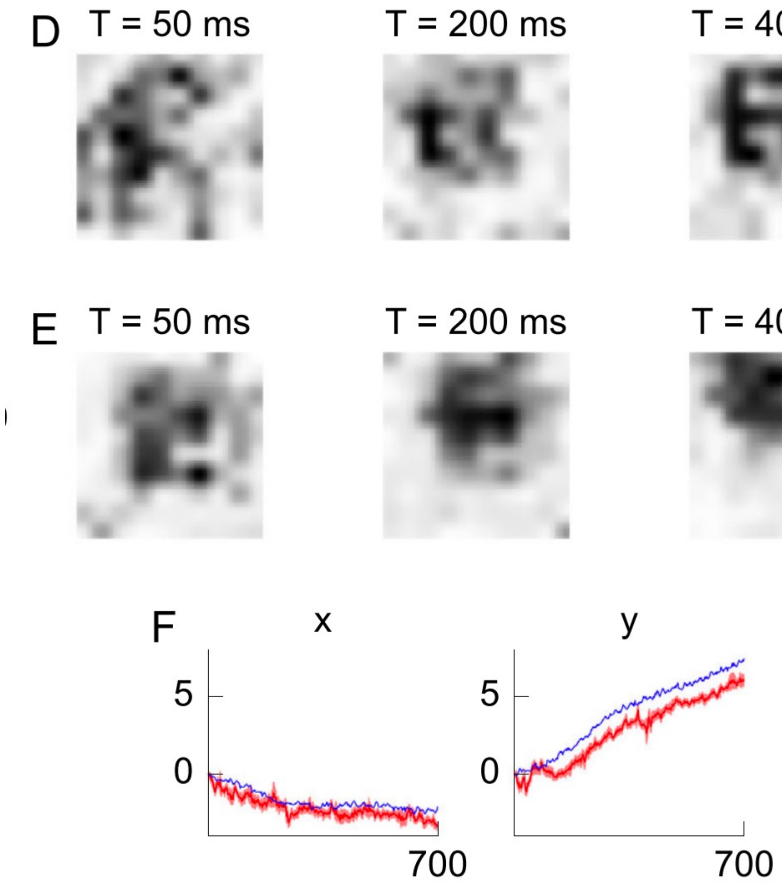
2. Update image (find A that maximizes the posterior probability)

$$p(A|R) \propto \underbrace{p(A)}_{\text{Image prior}} \sum_{X_{t+1}} \underbrace{p(R_{t+1}|X_{t+1}, A)}_{\text{Likelihood of spikes}} \underbrace{p(X_{t+1}|R, A_t)}_{}$$

$$A_{t+1} = \arg\max_{A} p(A|R)$$

Anderson et al (2020)

Decoder output



Decoder assumes motion

Decoder does NOT assume motion

Blue = true eye pos
Red = estimated eye position

Anderson et al (2020)

Results

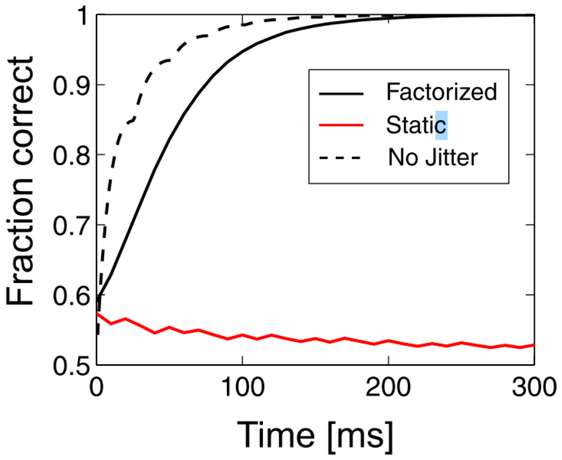
- When the stimulus is moving, a decoder that accounts for motion performs better than one that does not account for motion [1,2,3].
- Decoder performance improves for:
  - small D [1, 2, 3]
  - larger stroke widths [1, 3]
  - larger image size [2]
  - broader range of firing rates [1, 2]
- The decoder performs well despite its ignorance of the temporal filter [1, 2].

Conflicting Results

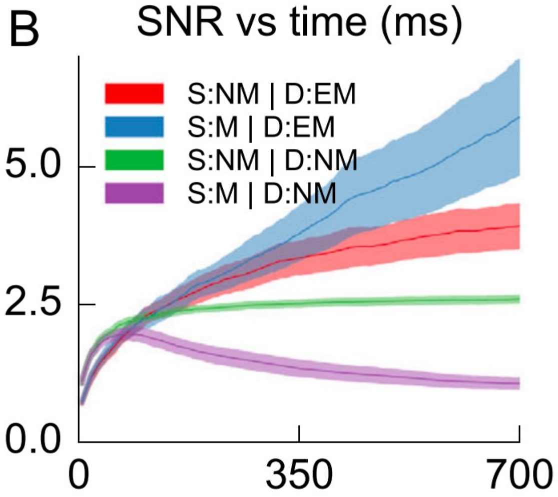
[2] Eye movements are a nuisance the visual system must overcome.

[3] Eye movements contribute to high-acuity vision.

- [2] A static decoder with a static image (“No Jitter”) performs better than an active decoder with a moving image (“Factorized”).
- [3] says the opposite (blue versus green)



[2] Burak et al, 2010



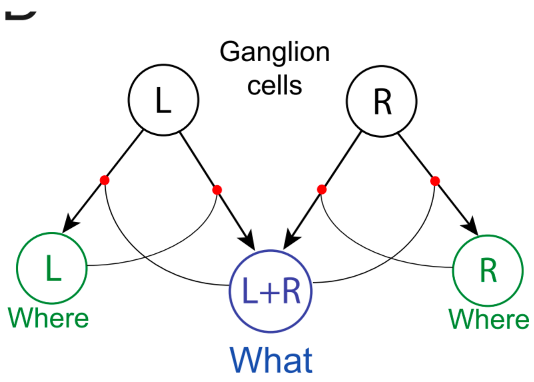
[3] Anderson et al, 2020

Temporal mechanisms?

- Temporal or correlation coding?
- Experiment that can differentiate between complementary theories of spatial averaging and temporal transients?
  - Ideal conditions:
  - spatial jitter with no temporal transients
  - temporal transients with no spatial jitter

3D and Binocular Vision

Combination of monocular image decoder with disparity-energy model (a spatial model of depth) would provide a spatial model of 3d vision in the presence of eye movements



[2] Burak et al, 2010

Luminance Patterns	Disparity Patterns
Luminance is driving input to photoreceptors and retinal neurons	Complex V1 neurons are tuned to (absolute) disparity
Retinal image motion modulates the luminance inputs to individual neurons	Differences in retinal image motion modulate disparity, stimulating a set of disparity-sensitive neurons with the same retinal locations