Image-computable Bayesian model for 3D motion estimation with natural stimuli explains human biases

Daniel Herrera-Esposito, Johannes Burge

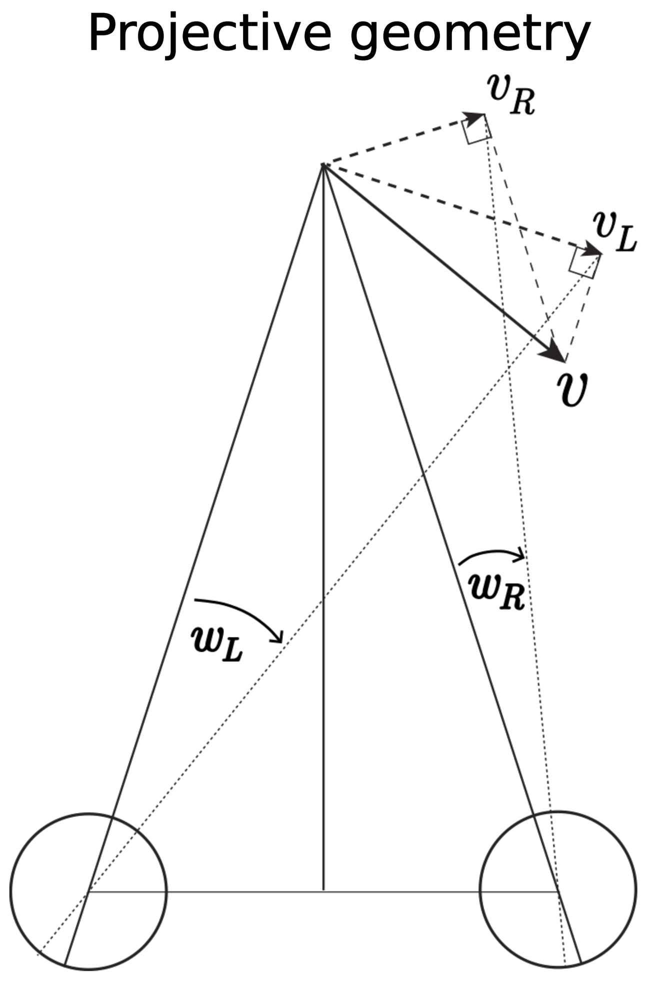
University of Pennsylvania

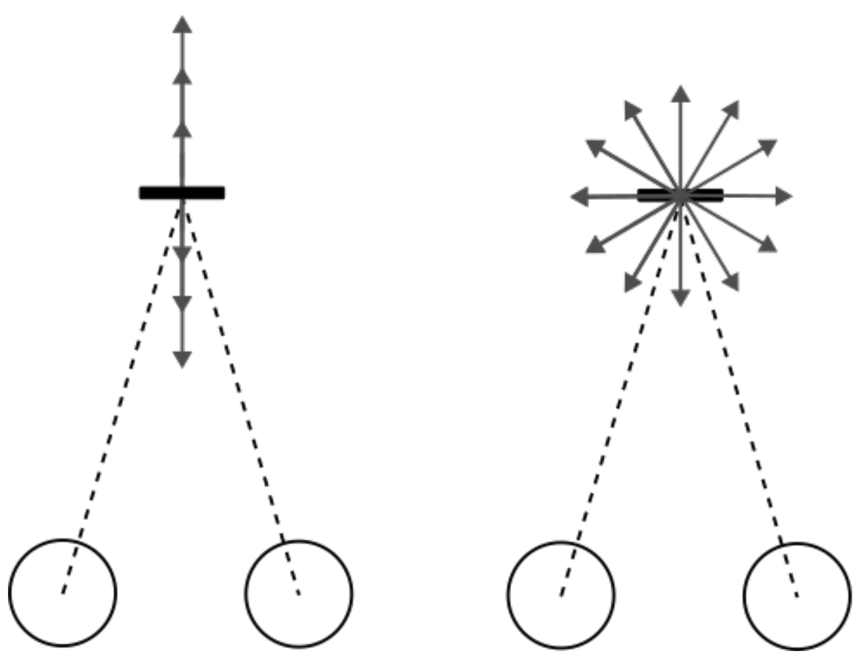
Aim

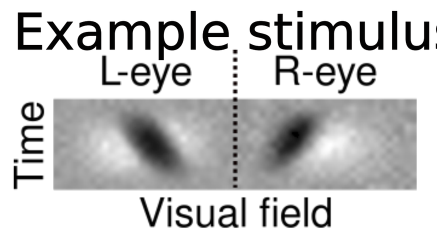
Develop ideal-observer model for local 3D motion estimation from naturalistic binocular videos, compare to physiology and behavior

Methods

- **Tasks:** 3D speed estimation, & 3D direction estimation
- Naturalistic dataset of 250 ms binocular clips







Model encoding stage: Burge, Jaini; PloS Comput Biol (2017)

1) Contrast normalize binocular video
2) Spatiotemporal filters f, plus noise:

$$\boldsymbol{c_{nrm}} = \frac{\boldsymbol{c}+\boldsymbol{n}}{\|\boldsymbol{c}+\boldsymbol{n}\|} \qquad \boldsymbol{R} = \boldsymbol{f}^T \boldsymbol{c_{nrm}} + \boldsymbol{\eta}$$

Model decoding stage:

3) Determine conditional response probabilities

$$P(\boldsymbol{R}|X = X_j)$$

4) Compute the likelihoods
5) Read out optimal estimate of 3D motion $\hat{X}$

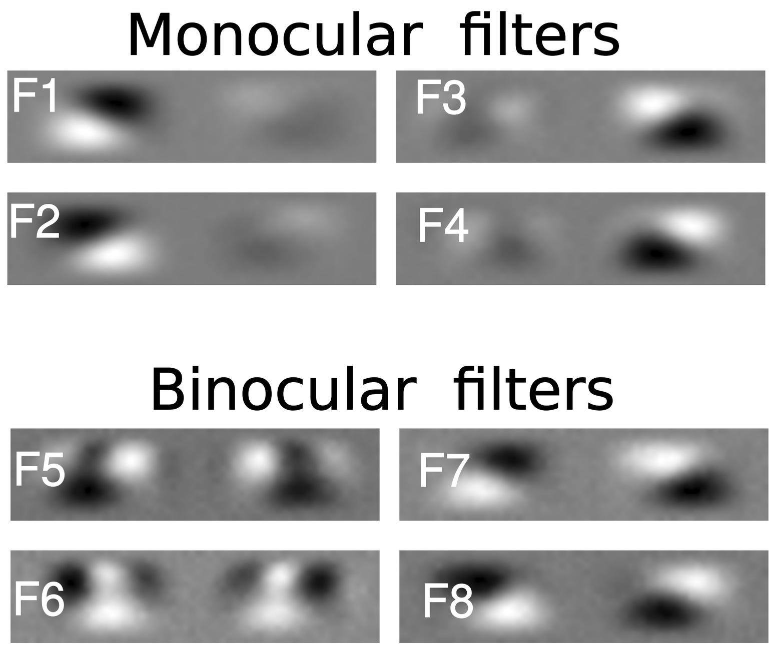
Training: Gradient descent using the expected cost (L0 norm) over the dataset to learn the filters f. Approximate conditional response distributions as Gaussian:

$$P(\boldsymbol{R}|X = X_j) \sim N(\boldsymbol{R}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

Results: Speed estimation

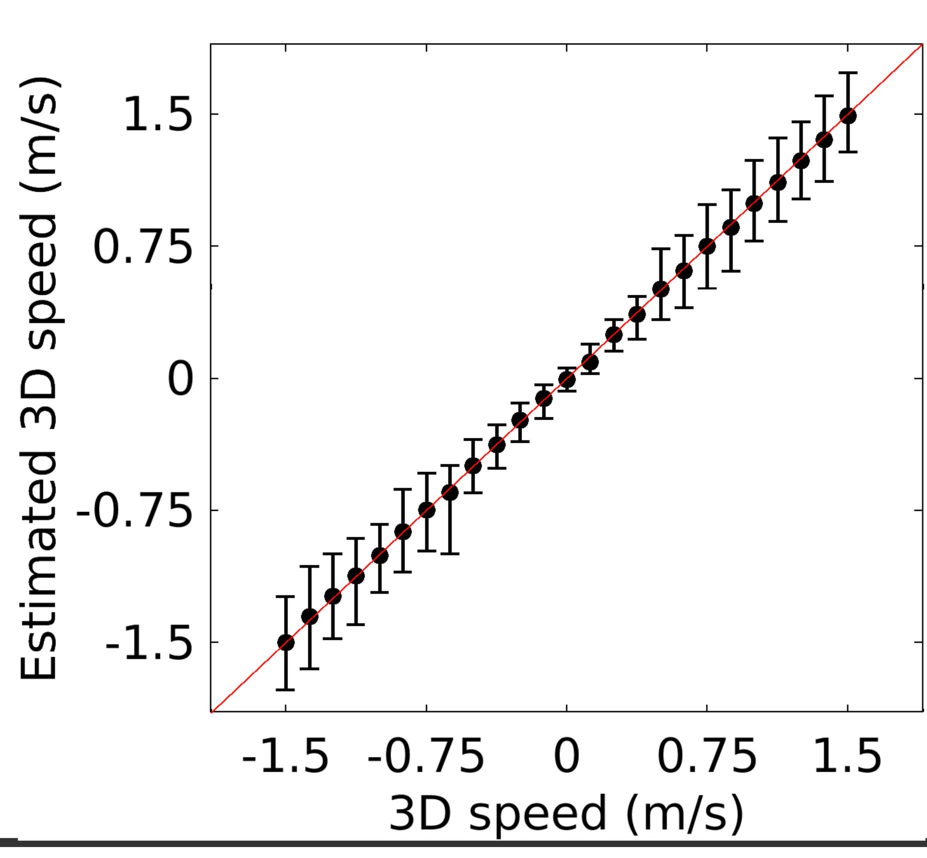
Filters: Two different mechanisms

Czuba et al.; J Neurophysiol, 2010

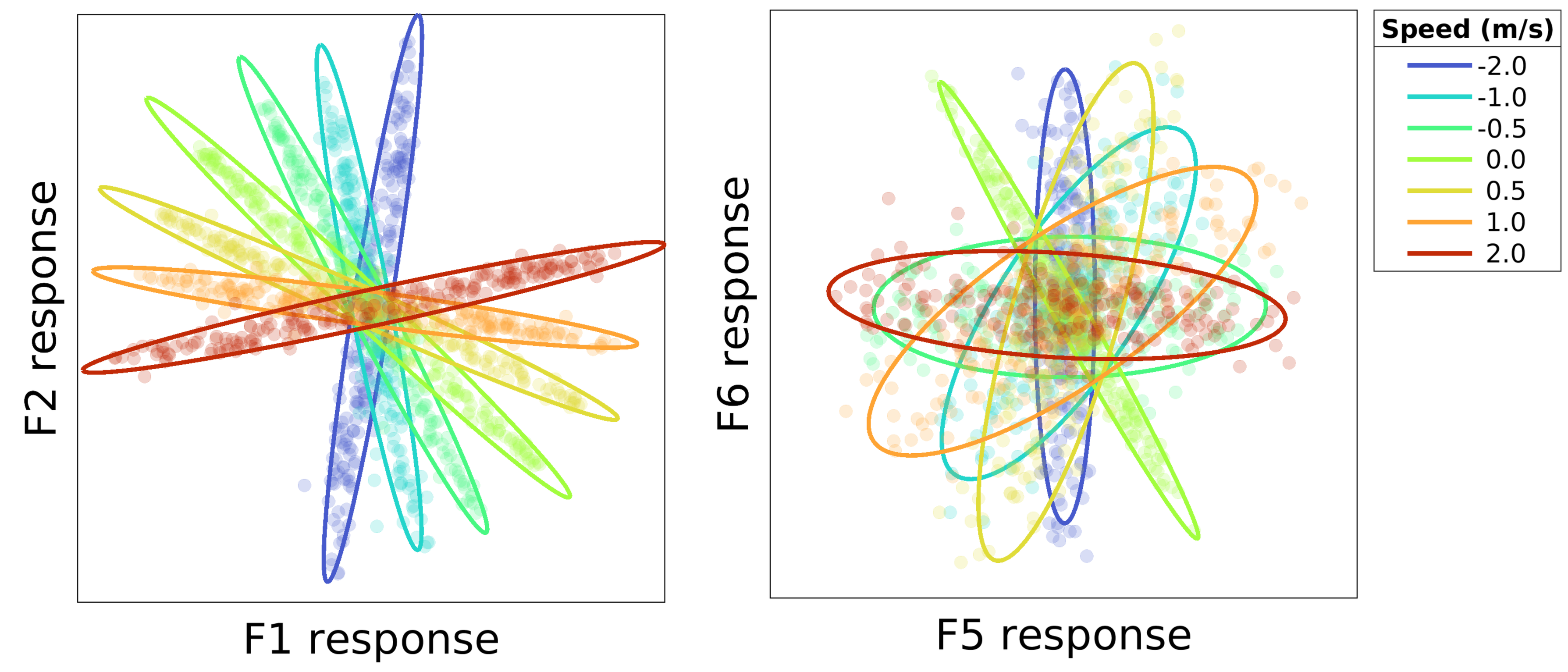


Inter-ocular velocity differences:
i) Time derivative of each eye (velocity)
ii) Binocular comparison

Changing disparity over time:
i) Binocular comparison (disparity)
ii) Time derivative

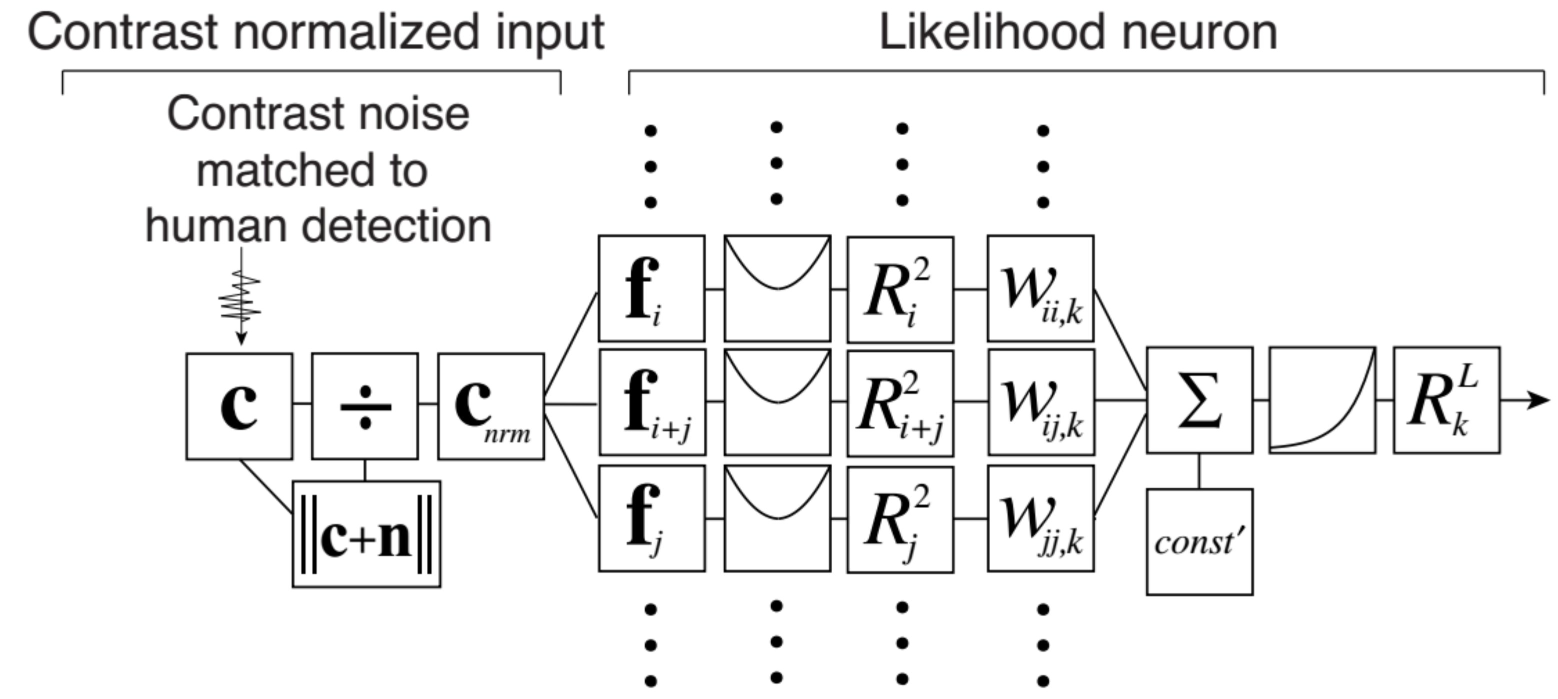


Conditional filter response distributions are Gaussian

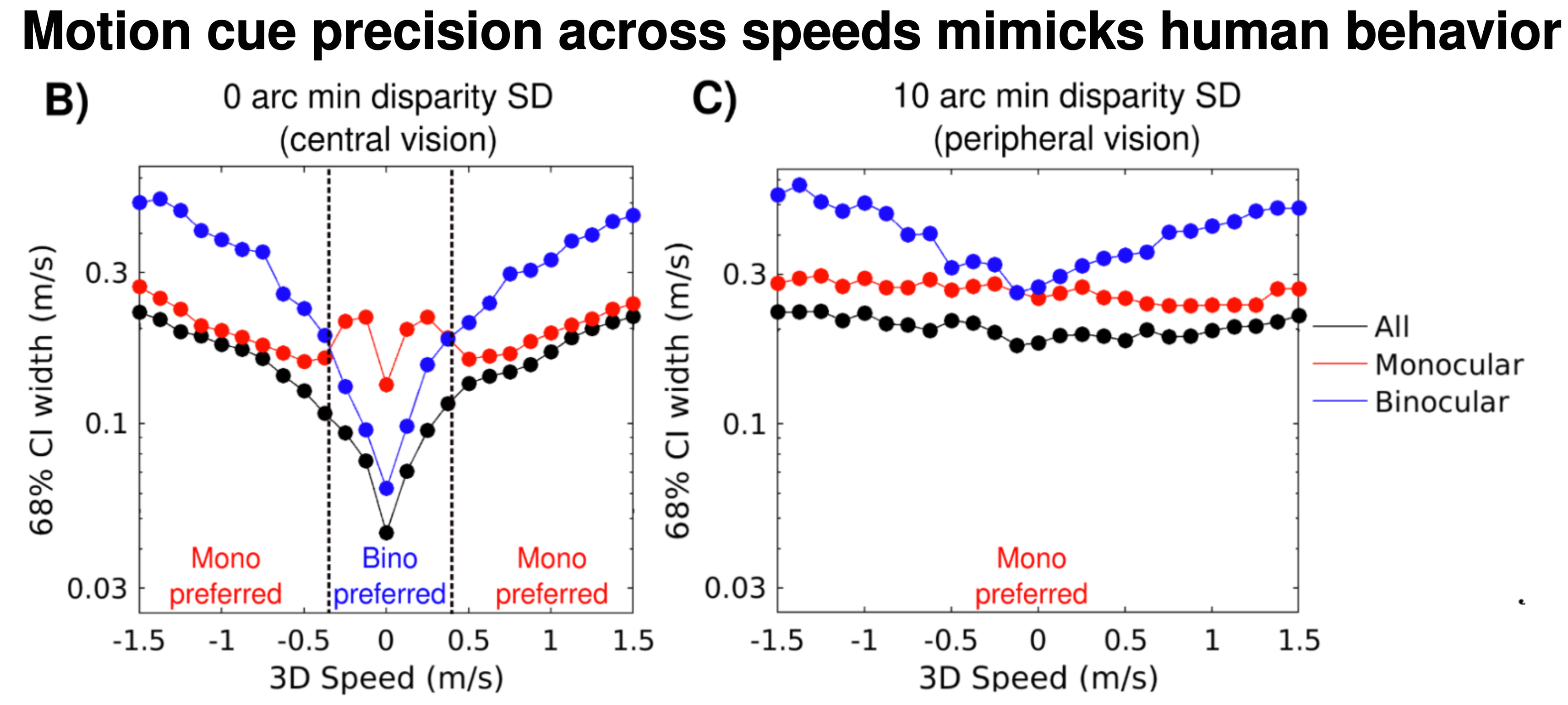


- Filter responses specify optimal pooling
- Gaussian response distributions specify 'energy-like' computations (Adelson, Bergen JOSA 1985)

$$ln\left[P(\boldsymbol{R}|X = X_j)\right] = -\frac{1}{2}\boldsymbol{R}^T\boldsymbol{\Sigma}_j^{-1}\boldsymbol{R} + C$$

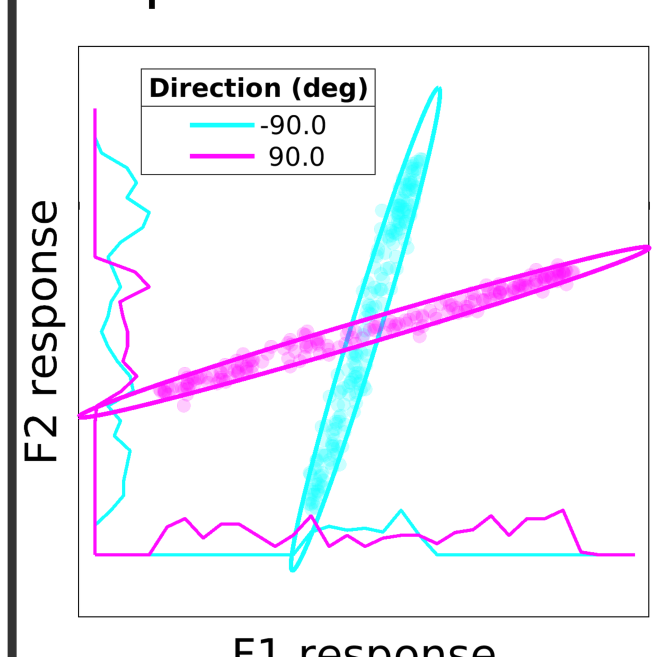


Motion cue precision across speeds mimicks human behavior

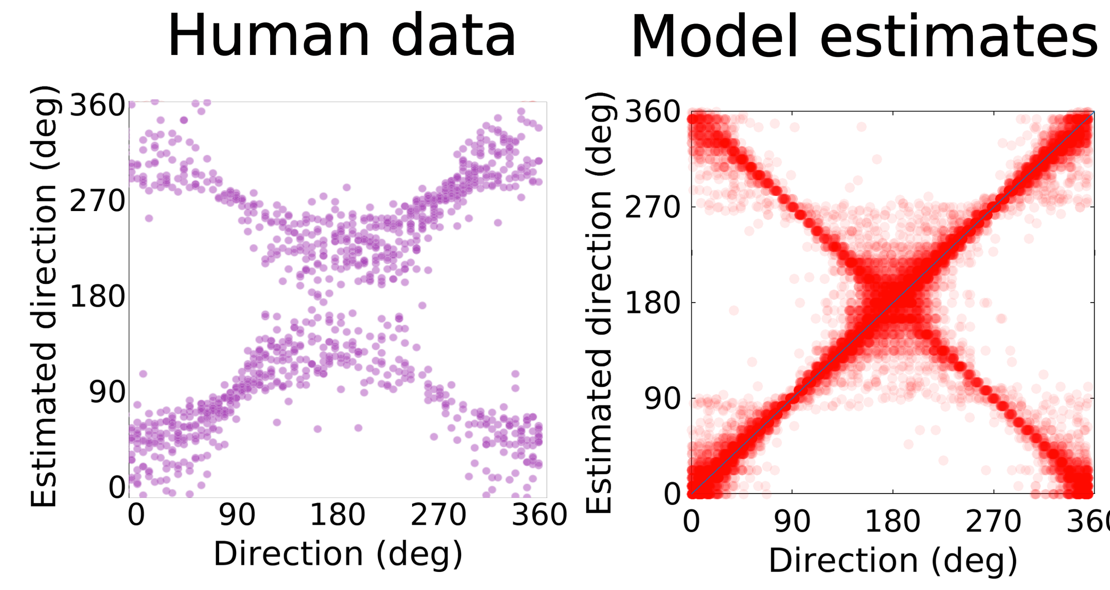


Results: Direction estimation

Filter response distribution: Some responses are bimodal



Model estimates: Sign confusions similar to humans



Bonnen et al. (2020) Nat Neuro

Conclusions

- Biologically plausible, energy-model-like computations optimal for local 3D speed estimation, as determined by natural statistics
- Two known computational mechanisms are learned by the ideal-observer. Their condition-dependent precision with natural images mimicks human behavior
- Ideal observer for direction estimation reproduces idyosincratic human biases