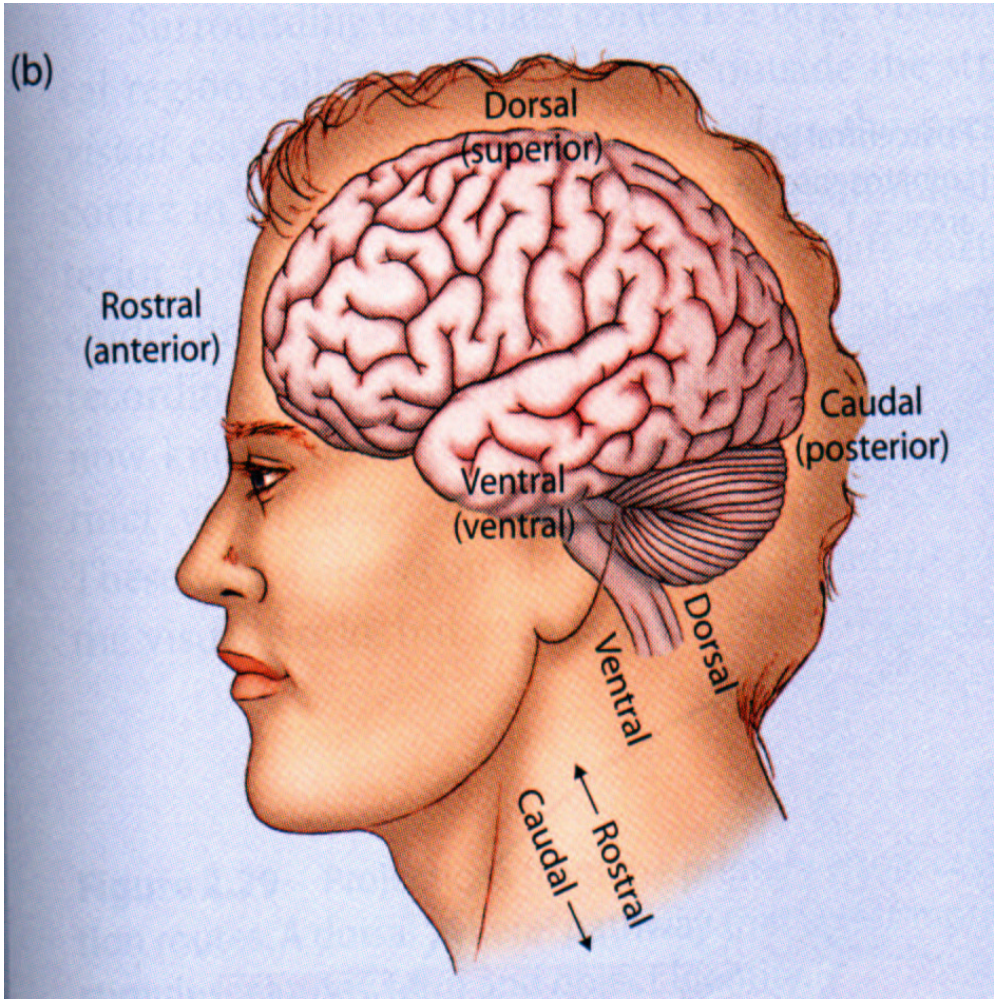
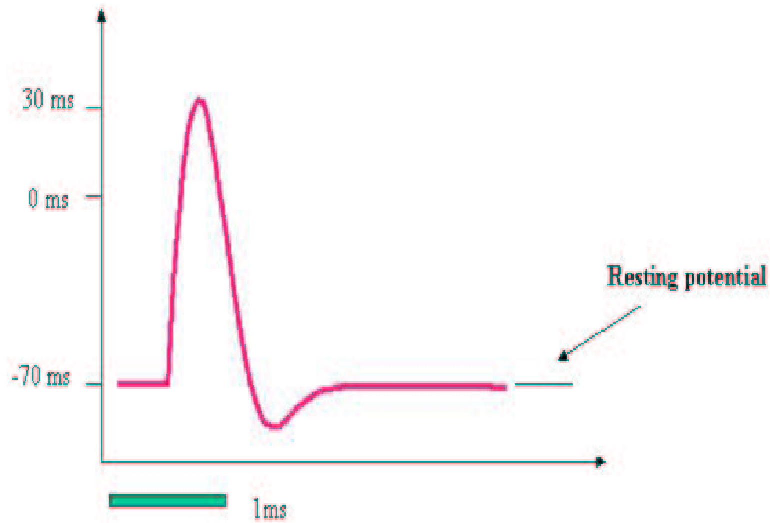
Bayesian Functional Data Analysis and its Application to Neuron Firing Patterns

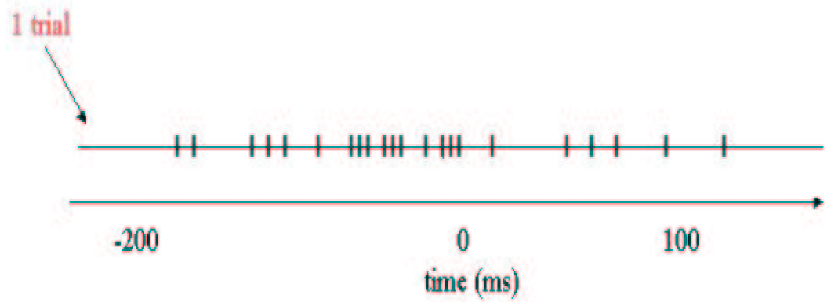
Rob Kass

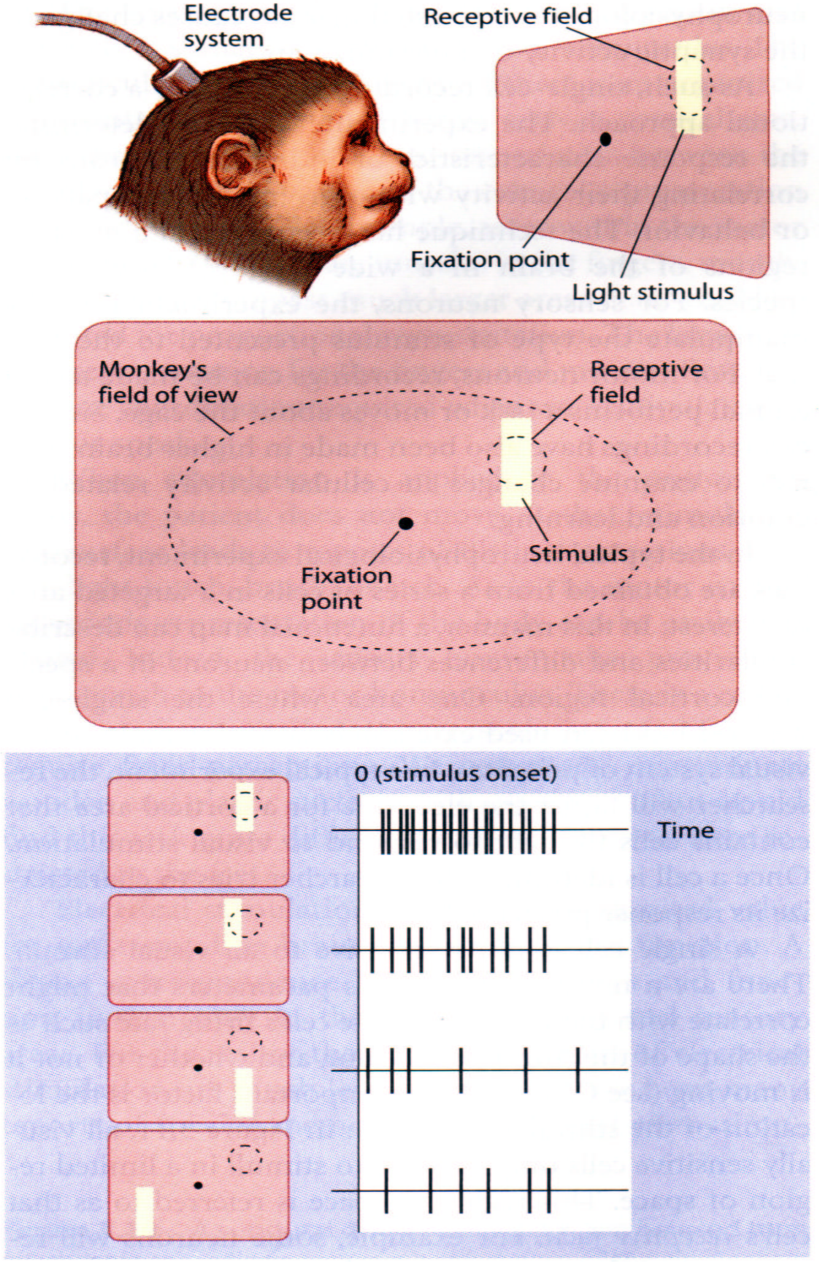
Department of Statistics
and
The Center for the Neural Basis of Cognition
Carnegie Mellon University

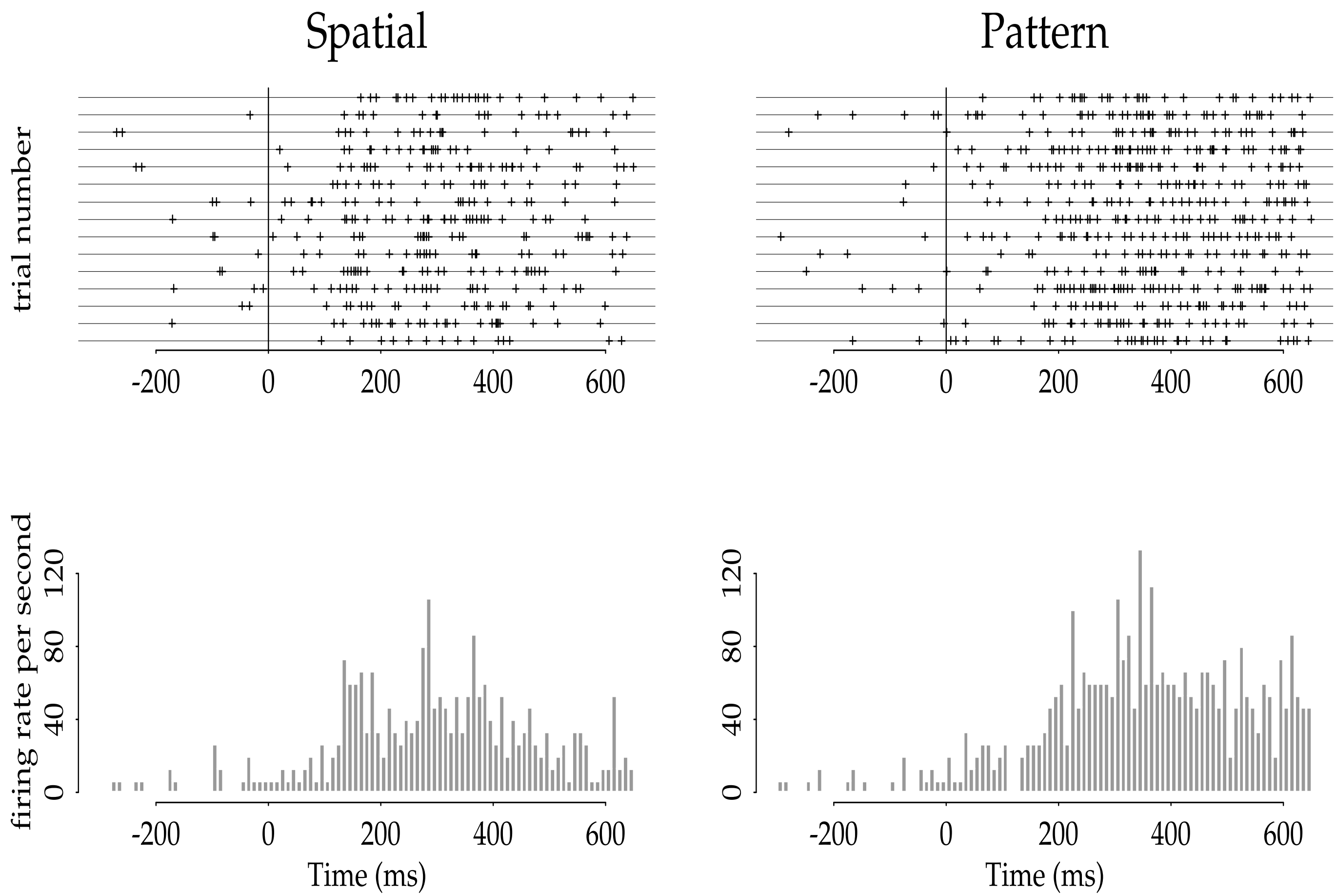


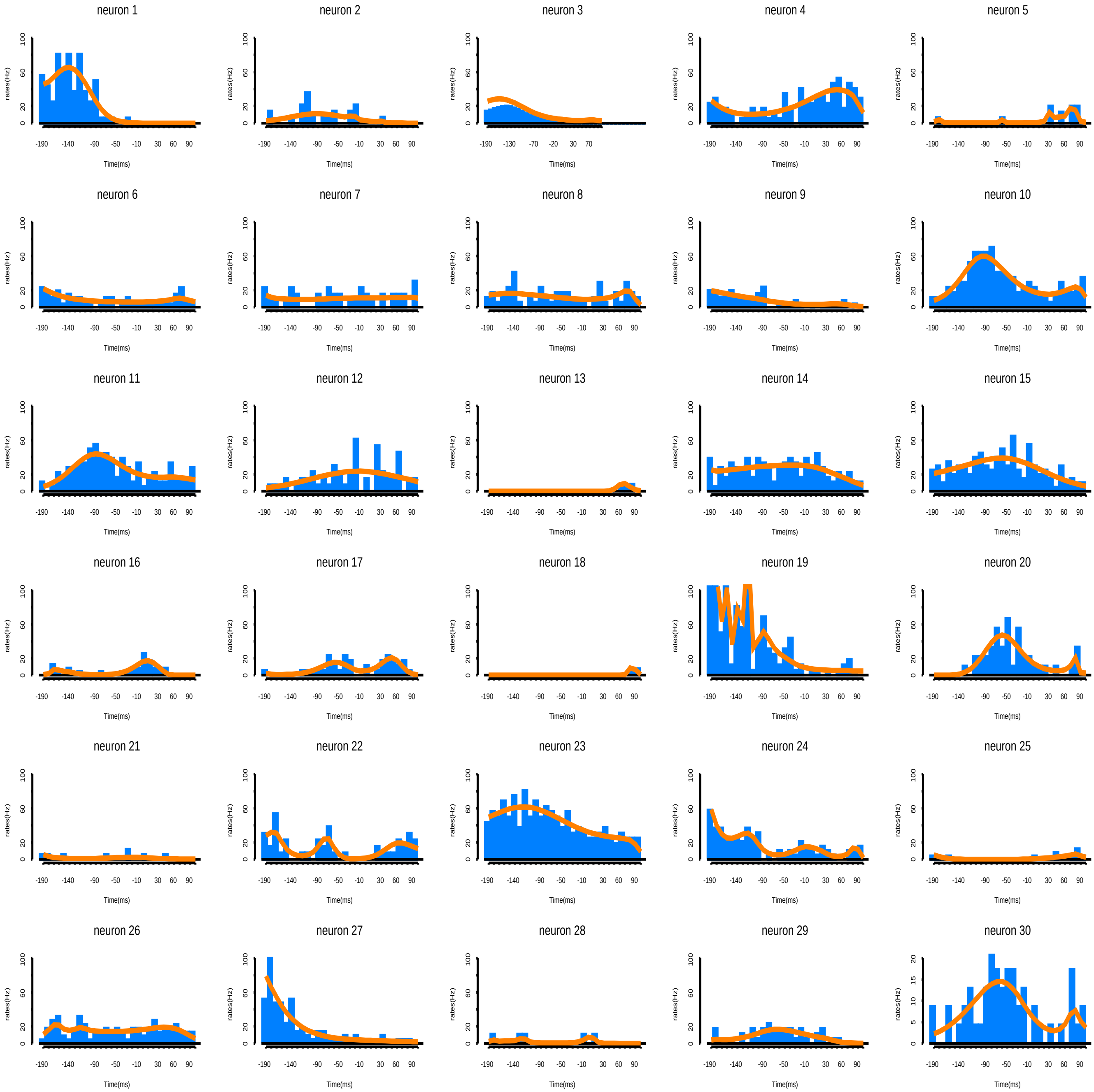
Source: Gazzaniga, Michael.S, Ivry, R.B. , Mangun, G.R., Congnitive Neuroscience, page:49, W.W.Norton, 1998.


30 ms
0 ms
-70 ms
Resting potential
1ms


1 trial
-200
0
100
time (ms)


Electrode system
Receptive field
Fixation point
Light stimulus
Monkey's field of view
Receptive field
Stimulus
Fixation point
0 (stimulus onset)
Time


Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)



neuron 1
neuron 2
neuron 3
neuron 4
neuron 5
neuron 6
neuron 7
neuron 8
neuron 9
neuron 10
neuron 11
neuron 12
neuron 13
neuron 14
neuron 15
neuron 16
neuron 17
neuron 18
neuron 19
neuron 20
neuron 21
neuron 22
neuron 23
neuron 24
neuron 25
neuron 26
neuron 27
neuron 28
neuron 29
neuron 30
rates(Hz)
0 20 60 100
0 5 10 15 20
Time(ms)
-190 -140 -90 -50 -10 30 60 90
-190 -130 -70 -20 30 70


neuron 1
neuron 2
neuron 3
neuron 4
neuron 5
neuron 6
neuron 7
neuron 8
neuron 9
neuron 10
neuron 11
neuron 12
neuron 13
neuron 14
neuron 15
neuron 16
neuron 17
neuron 18
neuron 19
neuron 20
neuron 21
neuron 22
neuron 23
neuron 24
neuron 25
neuron 26
neuron 27
neuron 28
neuron 29
neuron 30
rates (Hz)
0 20 40 60 80 100
-190 -140 -90 -50 -10 20 50 80
Time (msec)

Outline

- Background: brain is not simply modular and sequential.

Outline

- Background: brain is not simply modular and sequential.
- Timing involves instantaneous firing rate, i.e., intensity function $\lambda(t)$.

Outline

- Background: brain is not simply modular and sequential.
- Timing involves instantaneous firing rate, i.e., intensity function $\lambda(t)$.
- “Neural coding” is a rich application area for FDA. We have many projects that involve analysis of intensity functions in some form.

Some things I won't talk about (much)

- Inhomogeneous Markov Interval models for non-Poisson firing
- Evolving correlation structure of pairs of neurons
- Evolving correlation structure of multiple neurons
- Real-time prediction for neural prostheses

Outline (continued)

- We developed a Bayesian nonparametric regression method called BARS (for Bayesian Adaptive Regression Splines).
- The multiple-curve case of BARS provides a version of Bayesian Functional Data Analysis. This appears useful when
  - variability among multiple curves is of interest,
  - the curves are generally smooth except for some sharp changes, and
  - there is non-trivial noise in the data from which the curves will be estimated.

  Our main interest is in point process intensity functions.

Statistical Collaborators

Students: Sam Behseta, Can Cai, Roberto Carta, Ilaria DiMatteo, Cari Kaufman, Alex Rojas, Liuxia Wang

Faculty: Anthony Brockwell, Valérie Ventura, Garrick Wallstrom
and

Emery Brown, Satish Iyengar

REFERENCES
See www.stat.cmu.edu/∼kass

Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E., and Frank, L.M. (2002) The time-rescaling theorem and its application to neuronal spike train data analysis, *Neural Comput.*, 14: 325-346.

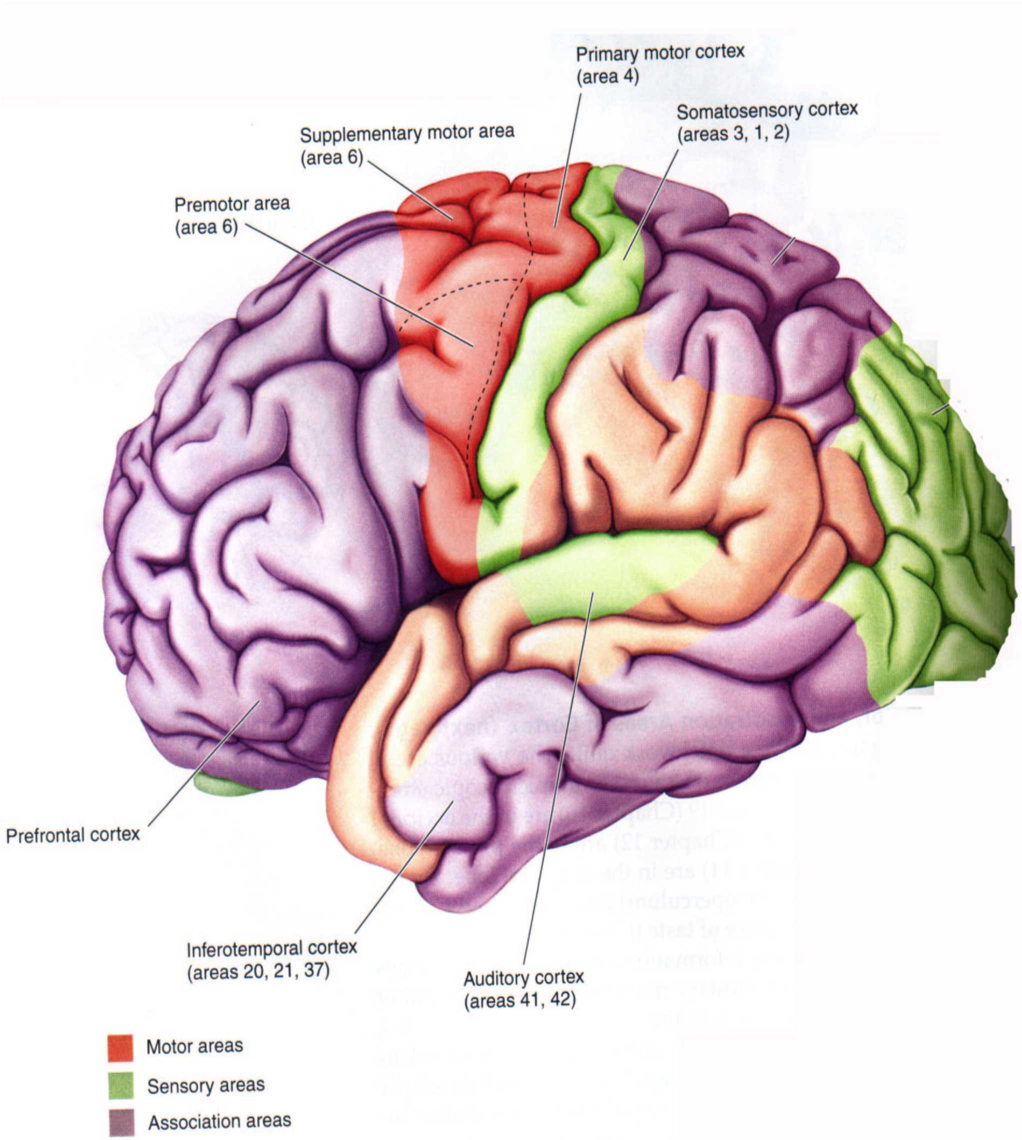
DiMatteo, I., Genovese, C.R., and Kass, R.E. (2001) Bayesian curve-fitting with free-knot splines. *Biometrika*, 88:1055-1071.

Kass, R.E., and Ventura, V. (2001) A spike-train probability model. *Neural Comput.*, 13: 1713-1720.

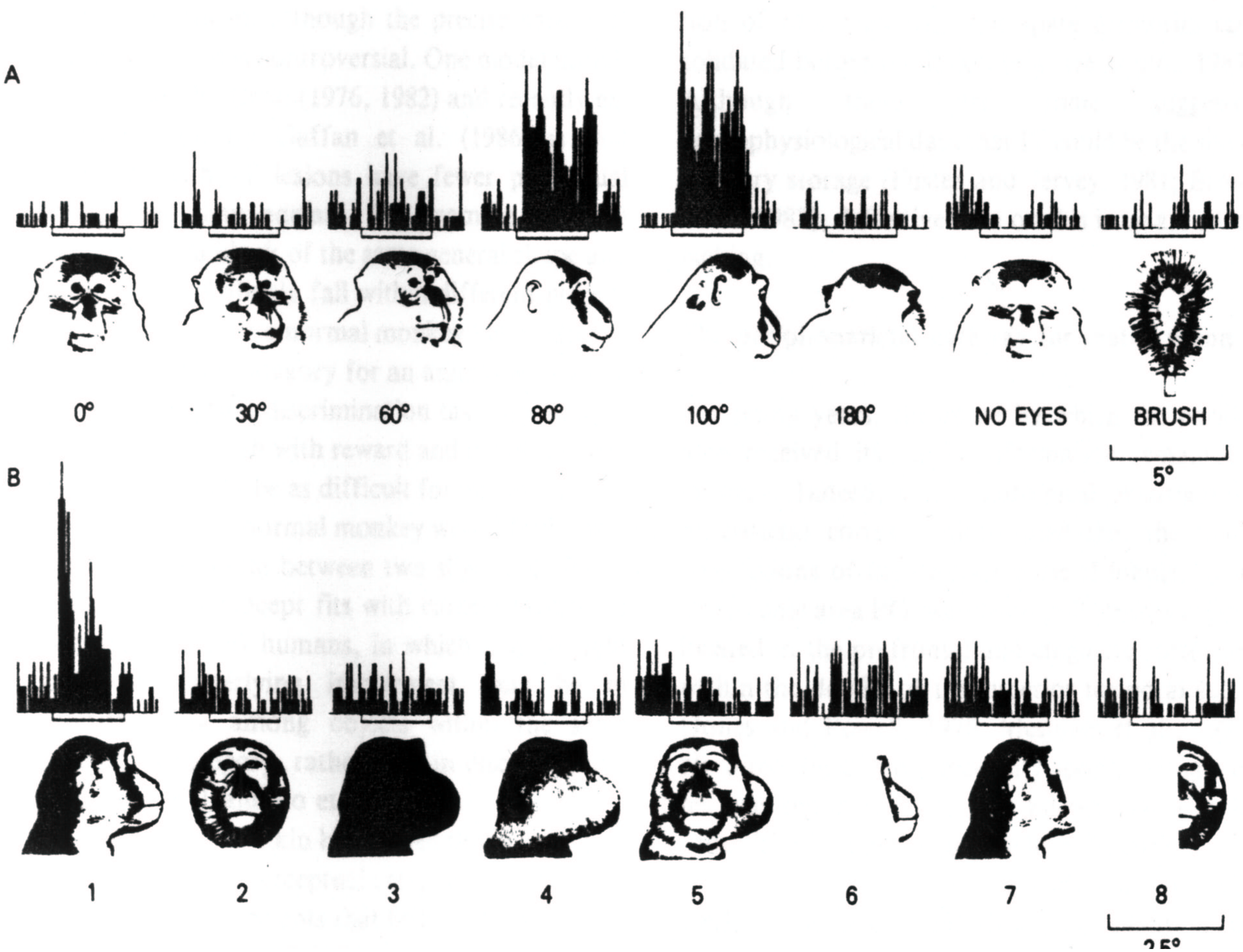
Olson, C.R., Gettner, S.N., Ventura, V., Carta, R. and Kass, R.E. (2000) Neuronal activity in macaque supplementary eye field during planning of saccades in response to pattern and spatial cues. *J. Neurophys.*, 84: 1369-1384.

Ventura, V., Carta, R., Kass, R.E., Gettner, S.N., and Olson, C.R. (2001) Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics*, 3: 1-20.

It is convenient to think of the brain as modular and sequential, e.g., for vision and movement.



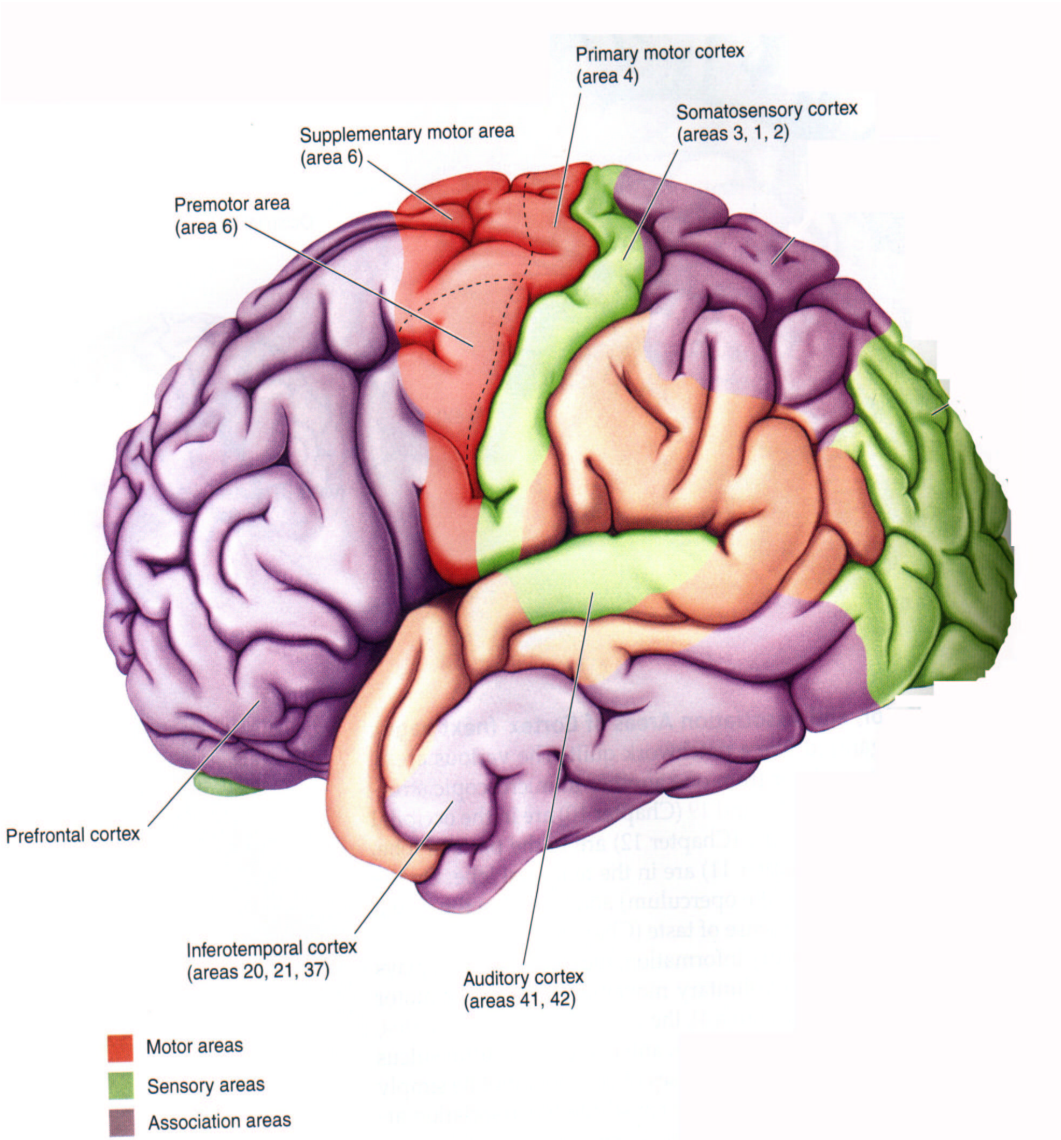
Source: Bear M.F., Connors, B.W., Paradiso, M.A., Neuroscience: Exploring the Brain, page:208, Lippincott, Williams and Wilkins, 2001.



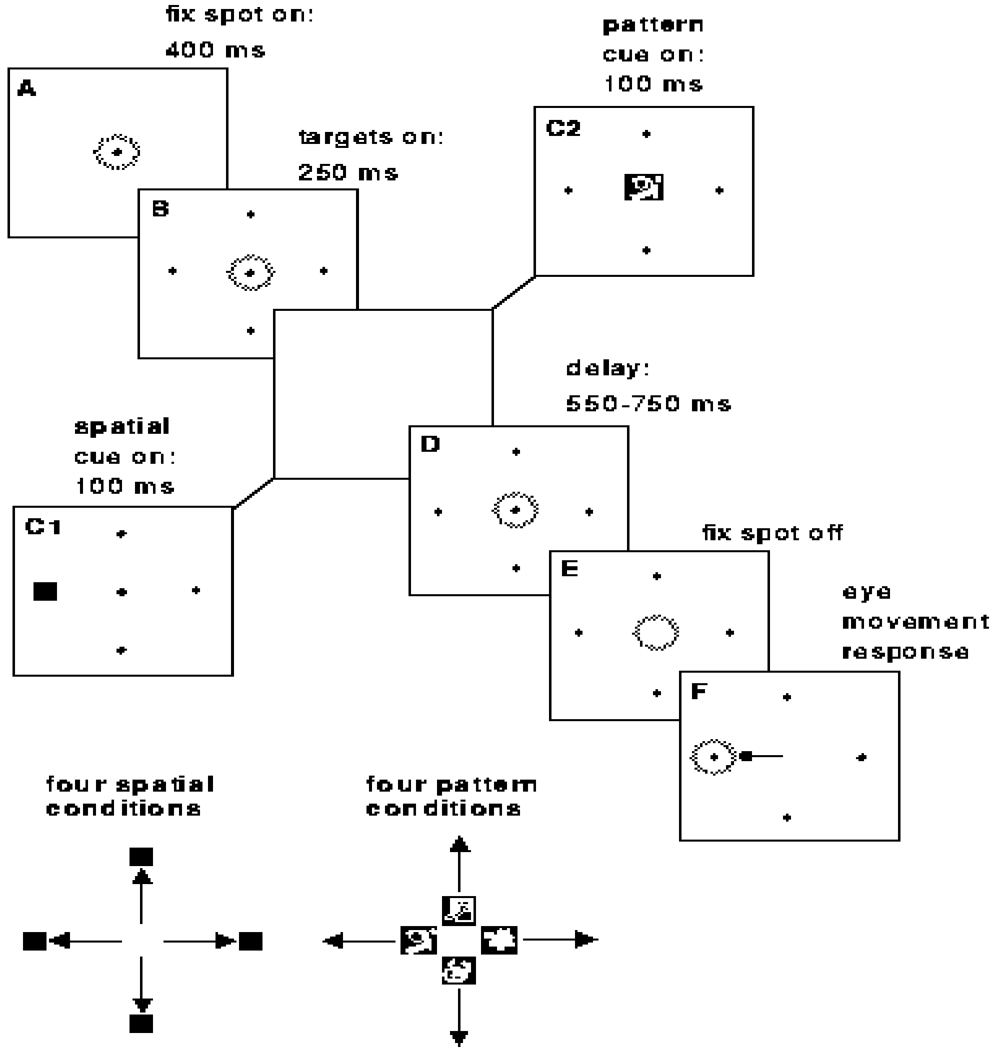
Source: Desimone, R., Ungerlieder, L.G., Handbook of Neurophysiology, page 287, Elsevier Science Publishers, 1989.

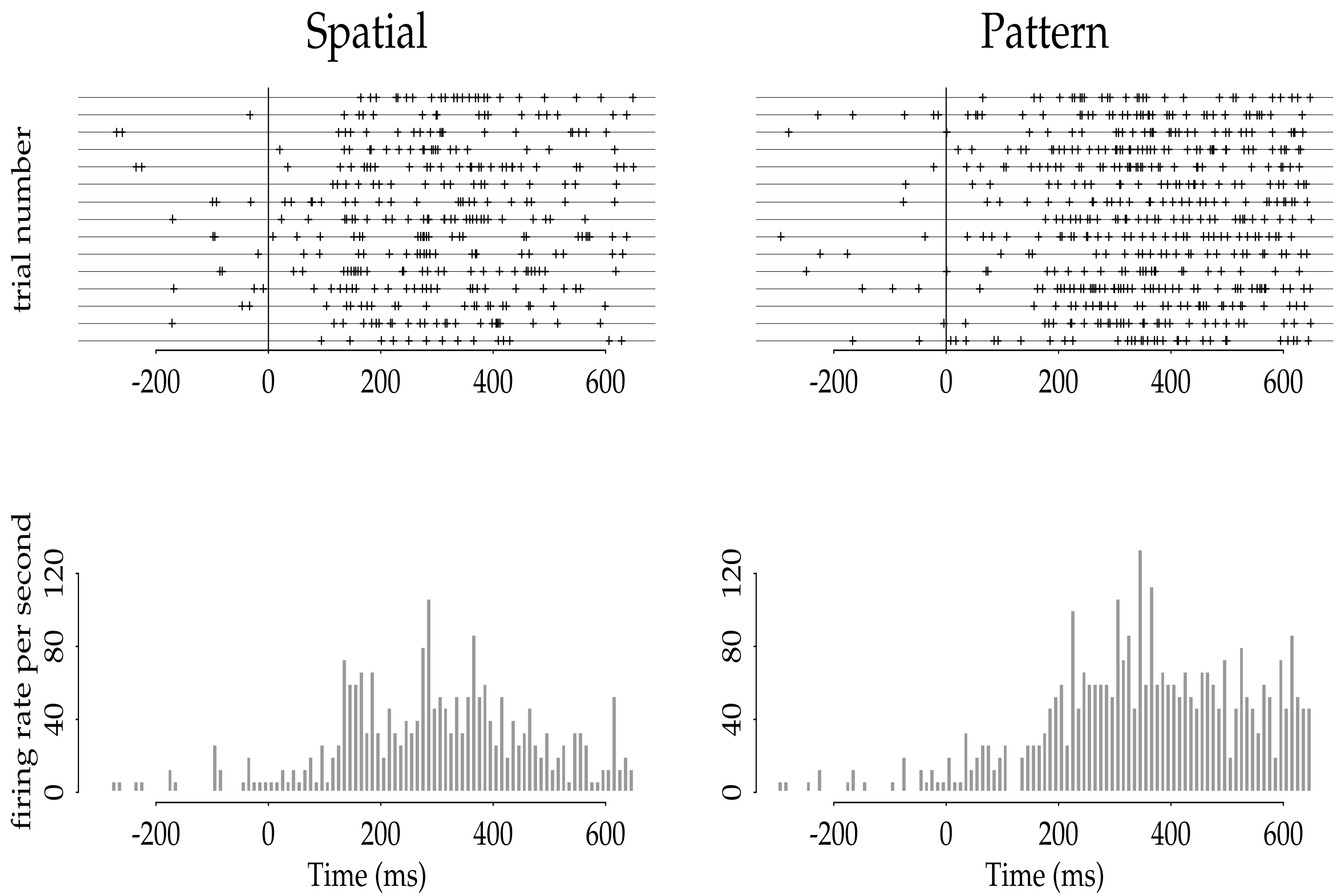
Not so simple: neuronal activity in areas that might be characterized as relay command centers has been shown to depend on higher-level cognition.

For example: Supplementary Eye Field (data shown earlier).

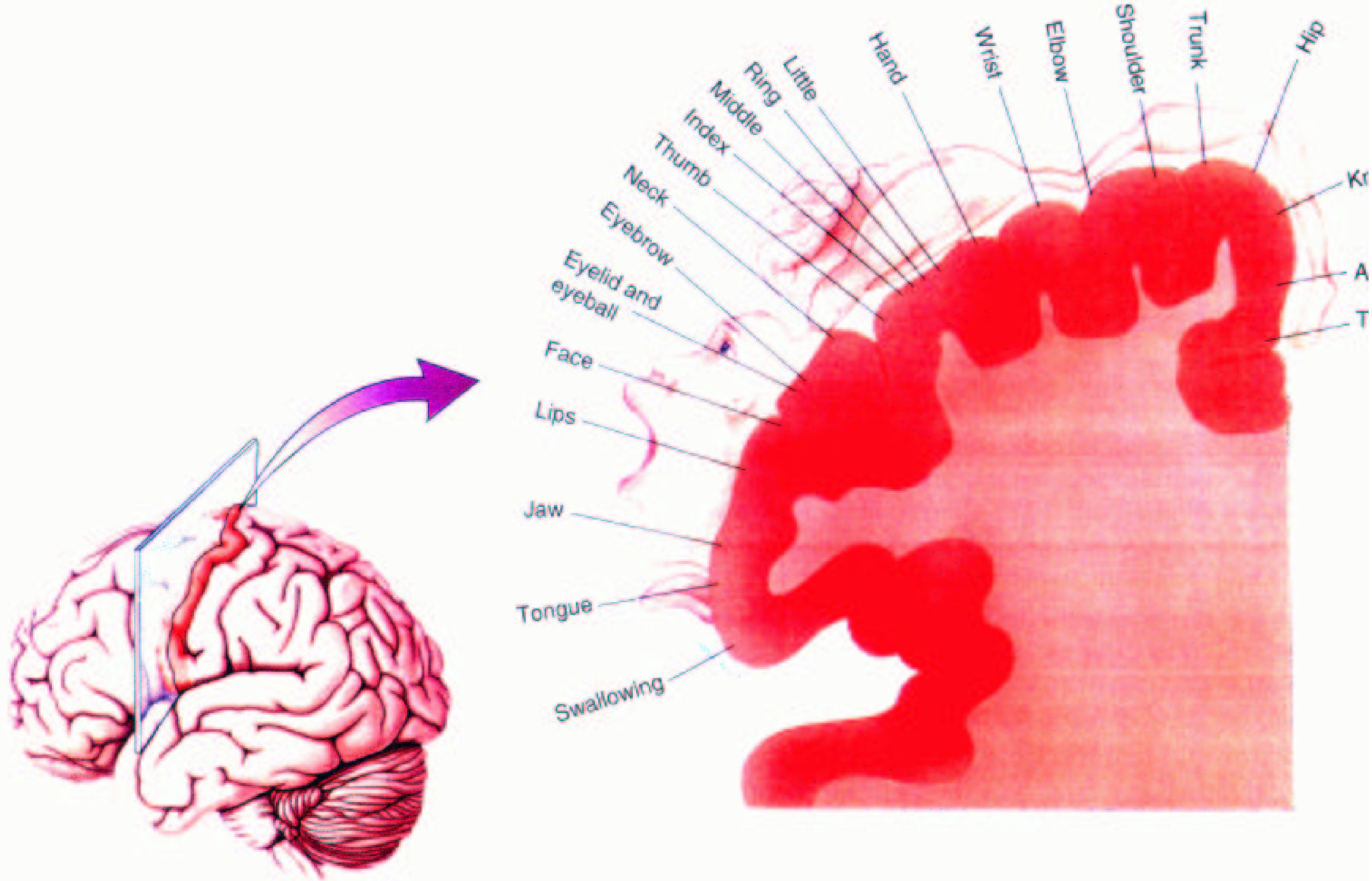


Source: Bear M.F., Connors, B.W., Paradis o, M.A., Neuroscience: Exploring the Brain, page:208, Lippincott, Williams and Wilkins, 2001.

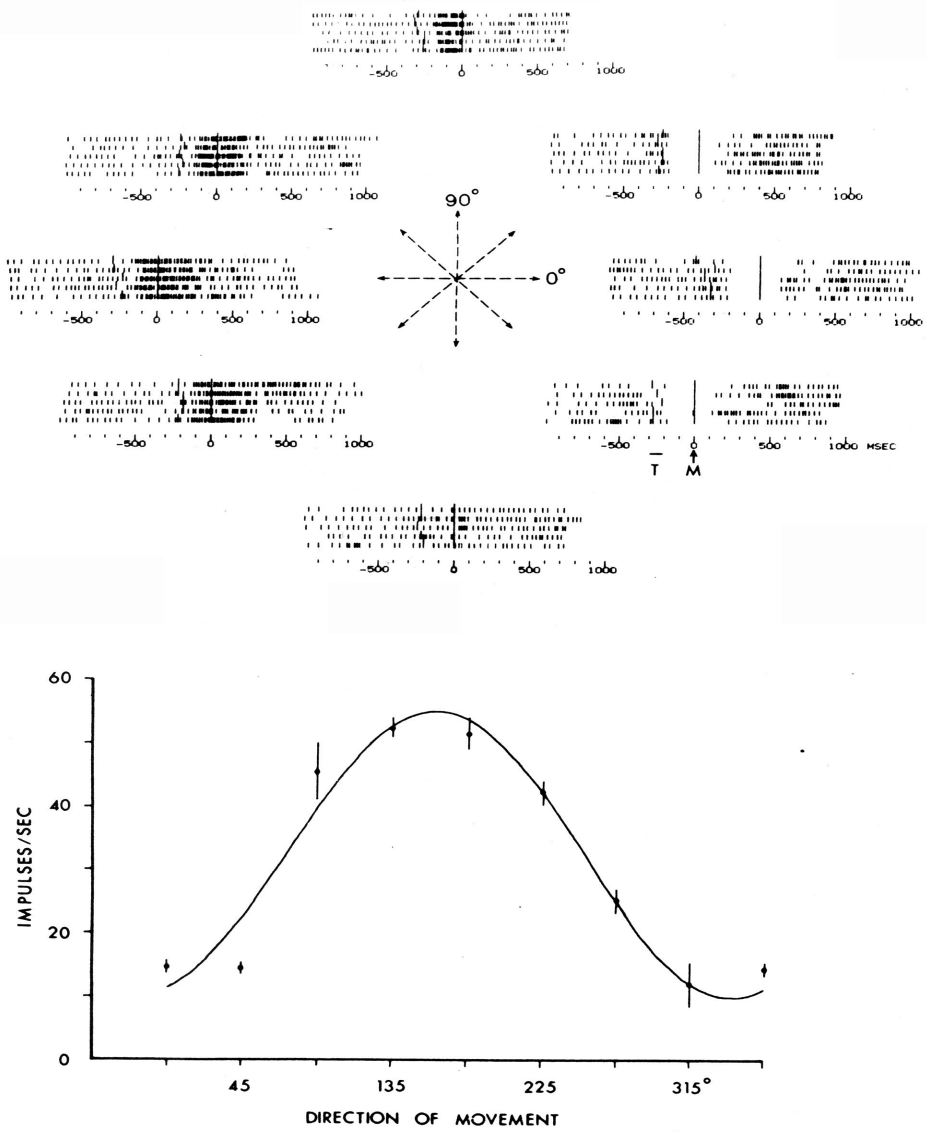

fix spot on:
400 ms
A
targets on:
250 ms
B
pattern
cue on:
100 ms
C2
delay:
550-750 ms
D
spatial
cue on:
100 ms
C1
fix spot off
E
eye
movement
response
F
four spatial
conditions
four pattern
conditions


Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)

Similarly, neurons in motor cortex not only drive muscles, they also respond in a manner that correlates with movements that are extrinsically defined.

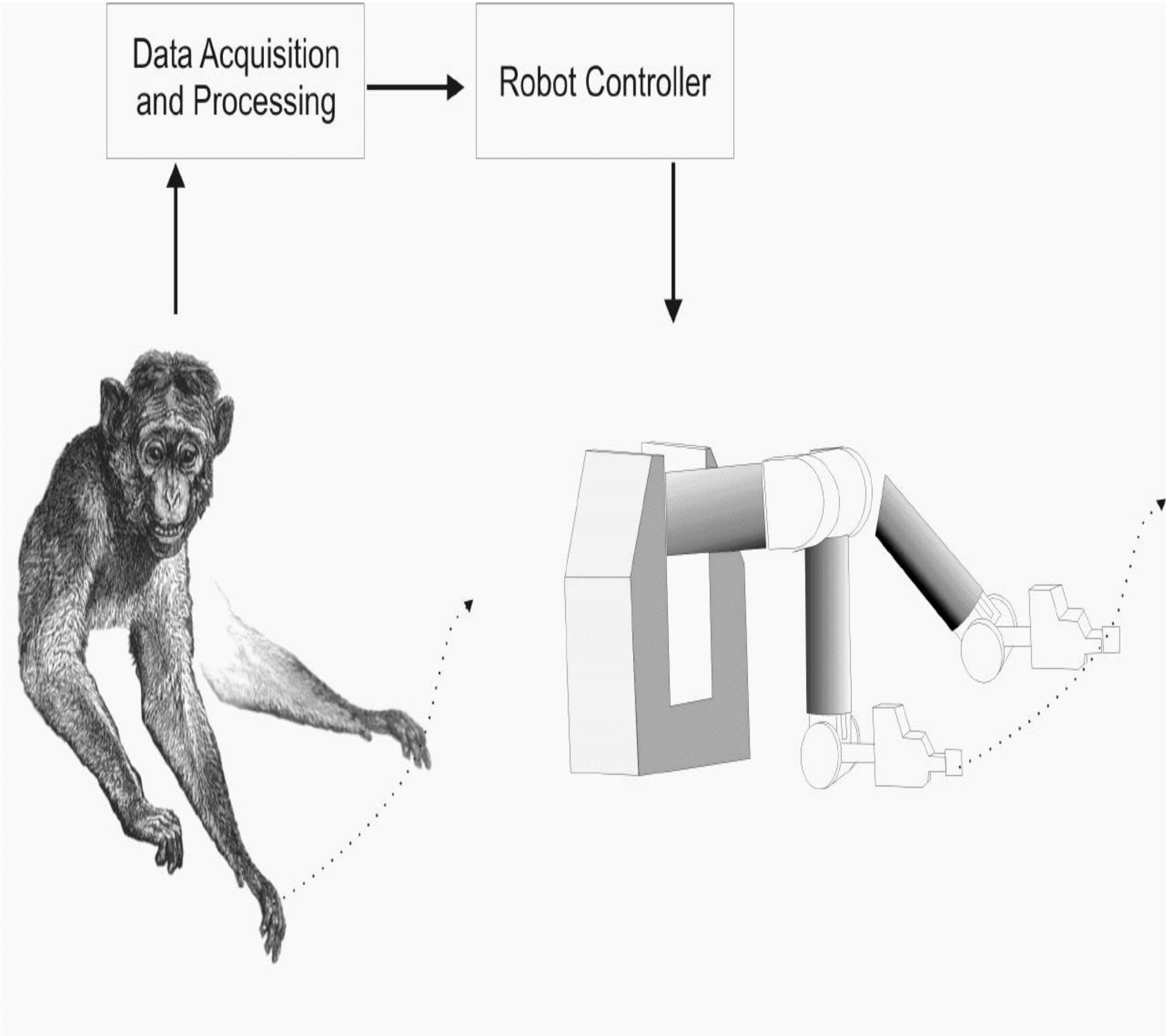


Source: Bear M.F., Connors, B.W., Paradiso, M.A., Neuroscience: Exploring the Brain, page:415, Lippincott, Williams and Wilkins, 2001.



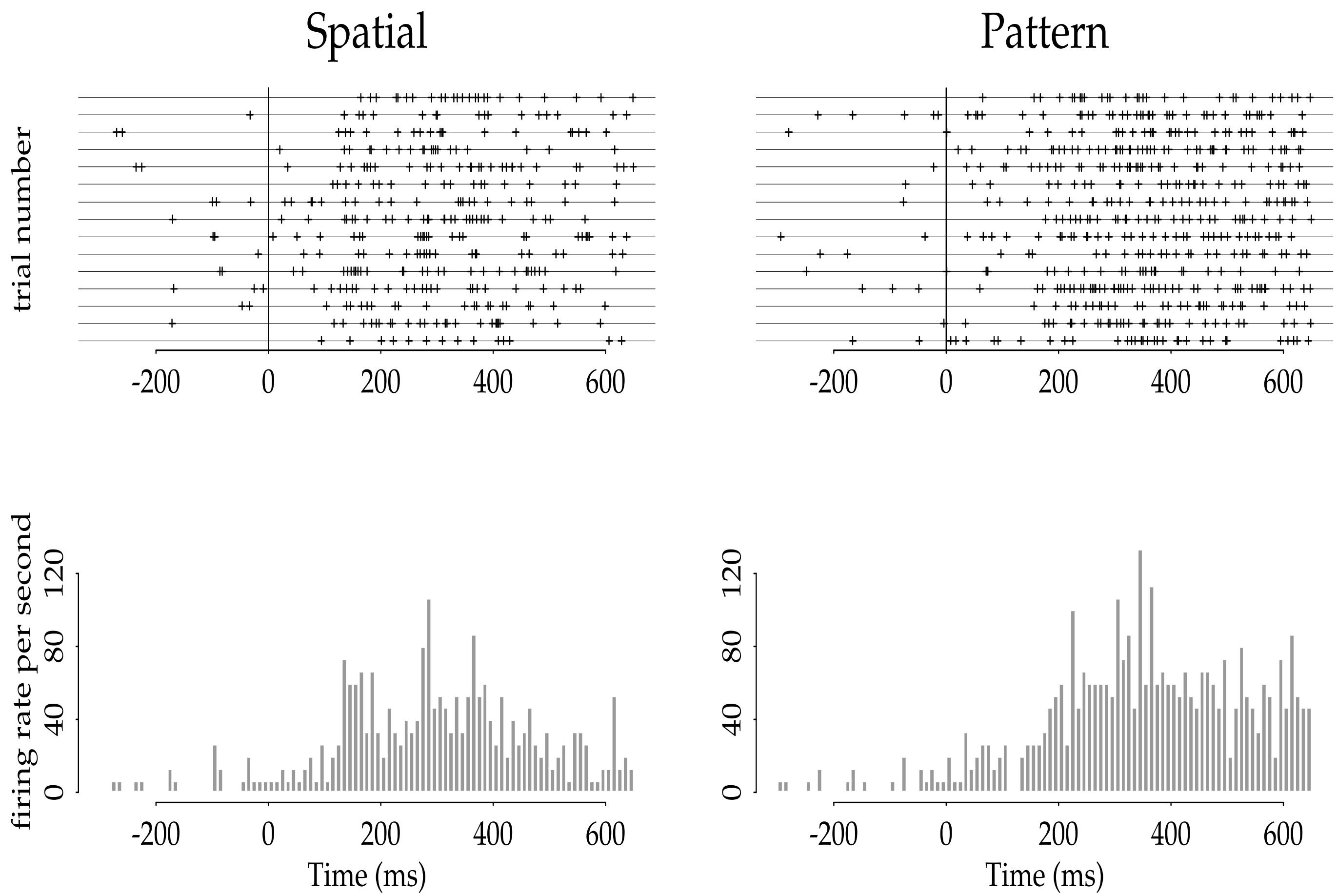
90°
0°
-500
0
500
1000
1000 MSEC
T
M
60
40
20
0
IMPULSES/SEC
45
135
225
315°
DIRECTION OF MOVEMENT

Directional tuning of motor cortex neurons may be captured to create a neural prosthetic device.


Data Acquisition
and Processing
Robot Controller

Summary of Background

- Electrophysiological studies identify conditions under which neurons in particular locations are active.
- Wonderful progress has been made by assuming the brain to be modular and sequential, but investigators are now trying to deal with greater complexity.
- This complexity often involves timing.

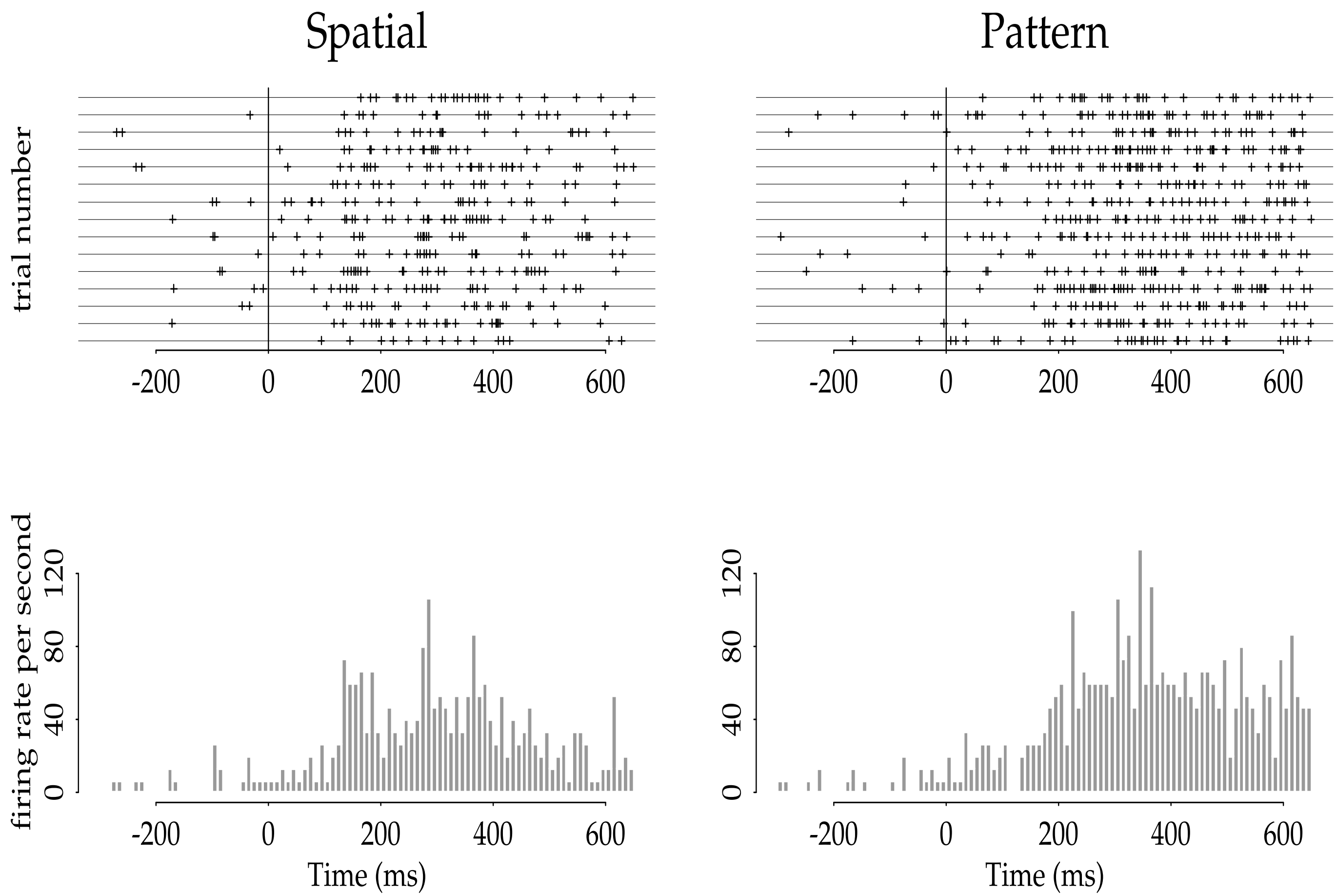

Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)

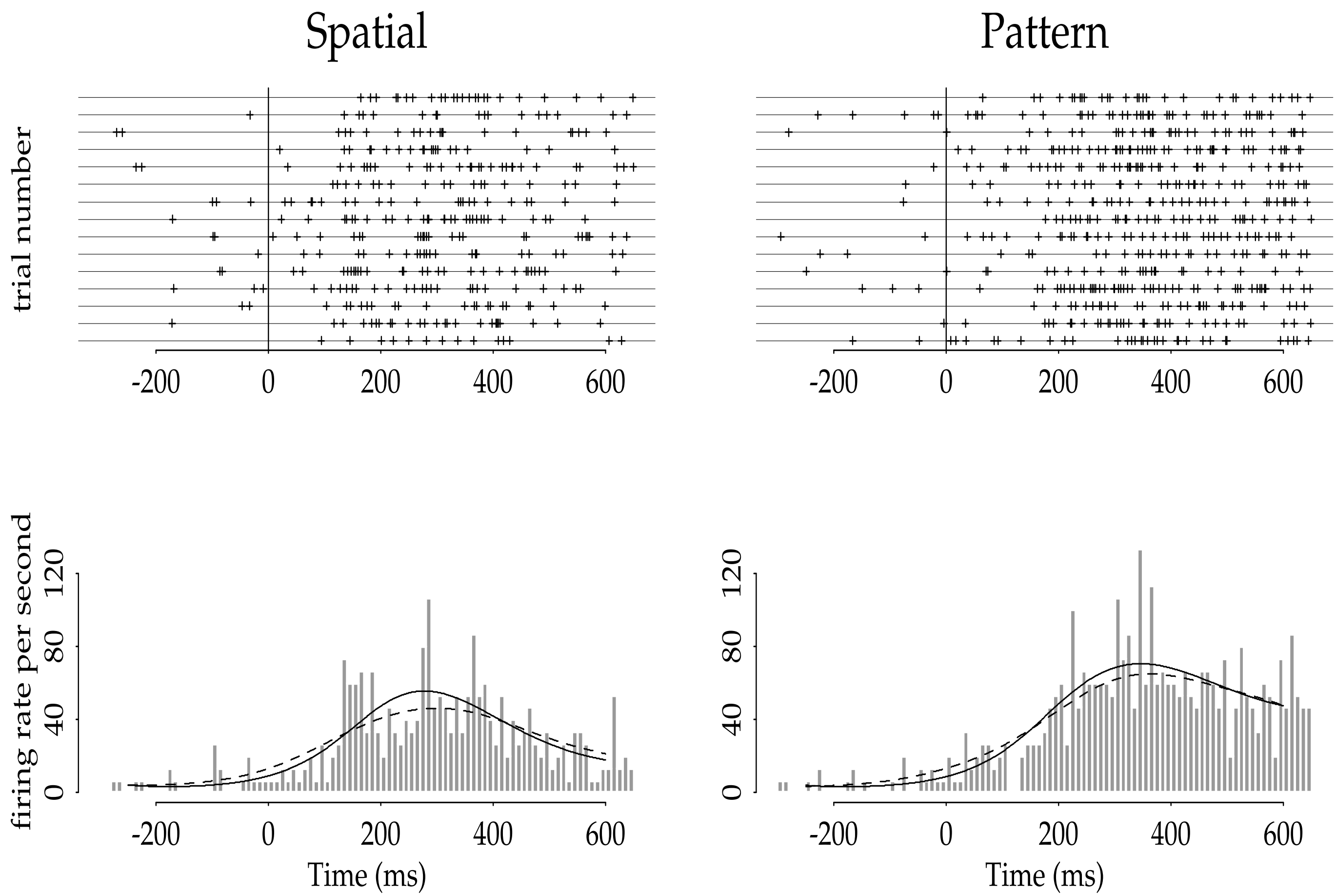
Data pooled across trials

Inhomogeneous Poisson

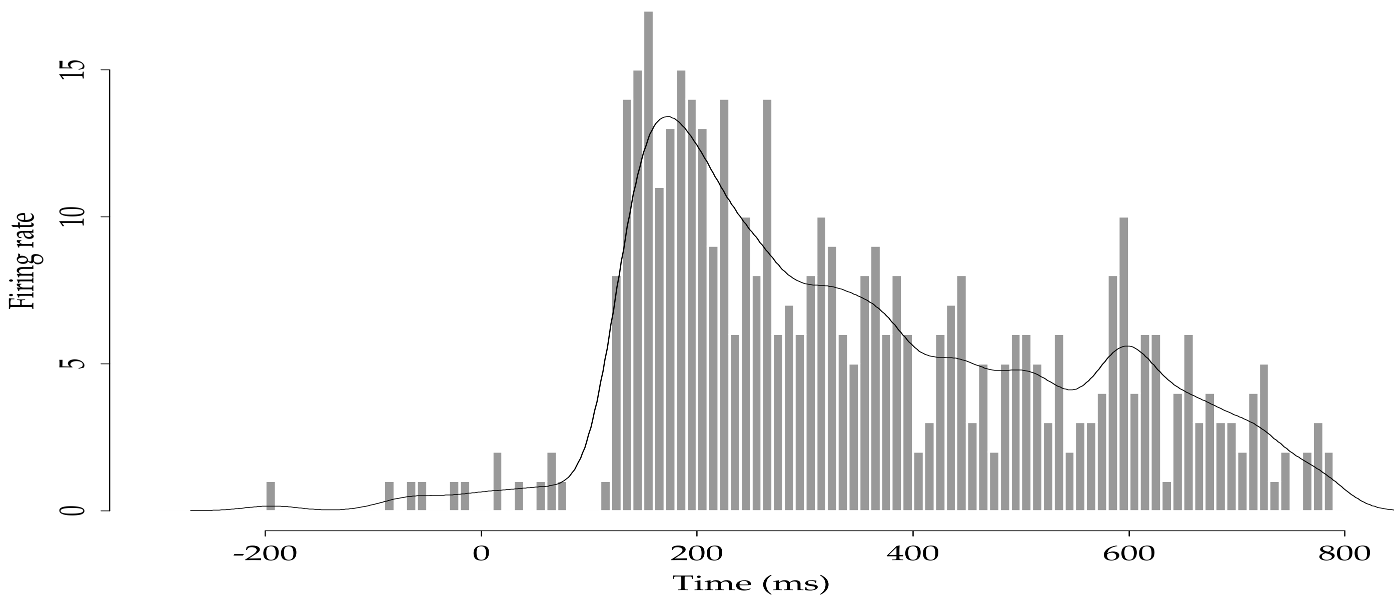
- In $(t, t + dt)$, Probability of spike $= \lambda(t)dt$
- Large number of trials $\Rightarrow$ approx. Poisson
- $\lambda(t)$ modeled using splines; apply Poisson regression

(Non-Poisson, within trials: we use Markov models)

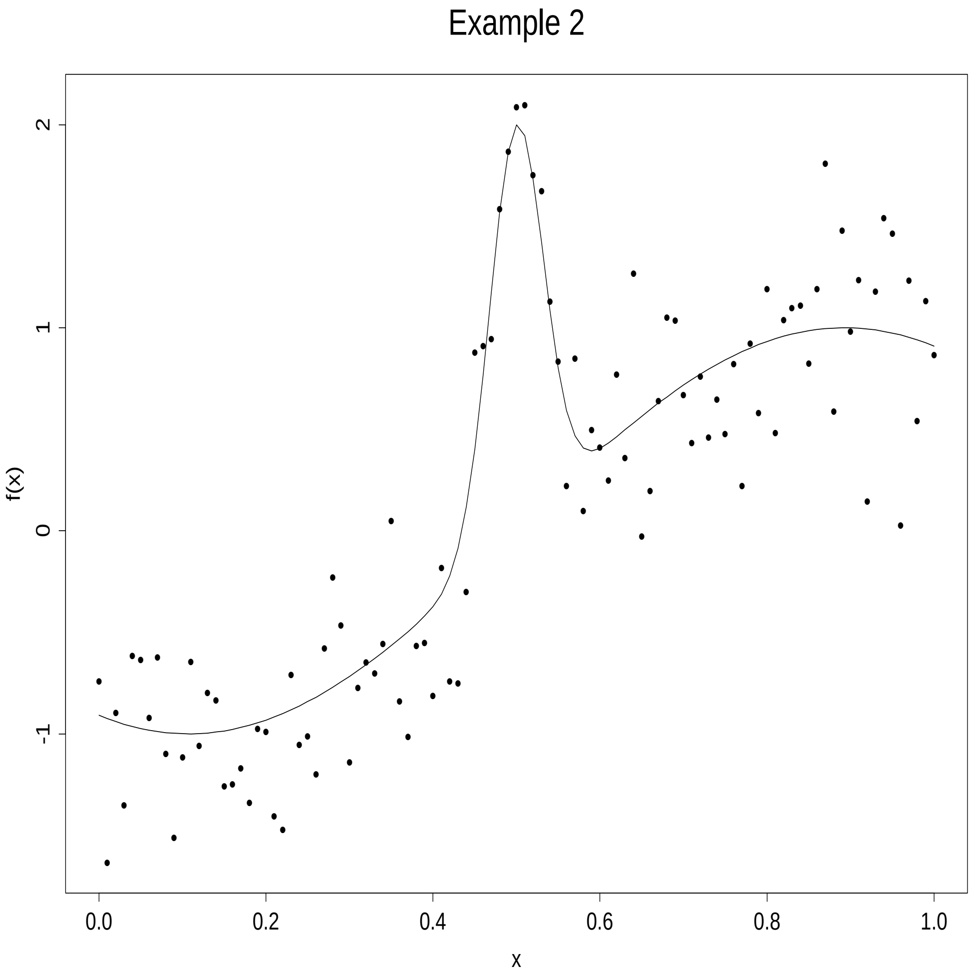

Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)


Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)

- Here: regression splines with two knots, and ML estimation; also Gaussian kernel density estimator.
- Sometimes fixed-bandwidth methods are not adequate.


Firing rate
0
5
10
15
-200
0
200
400
600
800
Time (ms)

Consider first the usual curve-fitting setting.



Usual curve-fitting framework With “free-knot” splines

$$Y_i = f(x_i) + \varepsilon_i$$

f is (approximated by) cubic spline with k knots at $\xi_1, \ldots, \xi_k$

$$f(x) = \sum_{j=1}^{k+2} b_j(x)\beta_j$$

- Conditionally on (ξ, k) would have a linear regression problem.
- Parameters: (β, σ, ξ, k).
- Algorithm: first integrate (β, σ).

Priors

$$\pi(\beta, \xi, k, \sigma) = \pi_\beta(\beta|\xi, k, \sigma)\pi_\xi(\xi|k)\pi_k(k)\pi_\sigma(\sigma)$$

where $\pi_\sigma(\sigma) = 1/\sigma$ and

$$\beta \,|\, \xi, k, \sigma \sim N_{k+2}\left(0, \sigma^2 n(B_\xi^T B_\xi)^{-1}\right)$$

with $B_{\xi,ij} = b_j(x_i)$. See Pauler (1998).

Priors

$$\pi(\beta, \xi, k, \sigma) = \pi_\beta(\beta|\xi, k, \sigma)\pi_\xi(\xi|k)\pi_k(k)\pi_\sigma(\sigma)$$

where $\pi_\sigma(\sigma) = 1/\sigma$ and

$$\beta \,|\, \xi, k, \sigma \sim N_{k+2}\left(0, \sigma^2 n(B_\xi^T B_\xi)^{-1}\right)$$

with $B_{\xi,ij} = b_j(x_i)$. See Pauler (1998).

Kass and Wasserman (1995) called this the “unit-information prior.”

Priors

$$\pi(\beta, \xi, k, \sigma) = \pi_\beta(\beta|\xi, k, \sigma)\pi_\xi(\xi|k)\pi_k(k)\pi_\sigma(\sigma)$$

where $\pi_\sigma(\sigma) = 1/\sigma$ and

$$\beta \,|\, \xi, k, \sigma \sim N_{k+2}\left(0, \sigma^2 n(B_\xi^T B_\xi)^{-1}\right)$$

with $B_{\xi,ij} = b_j(x_i)$. See Pauler (1998).

Kass and Wasserman (1995) called this the "unit-information prior."

We used Uniform prior for $\pi_\xi(\xi|k)$ and Uniform or Poisson for $\pi_k(k)$.

MCMC scheme

- Reversible-jump chain on (ξ, k) after integrating analytically

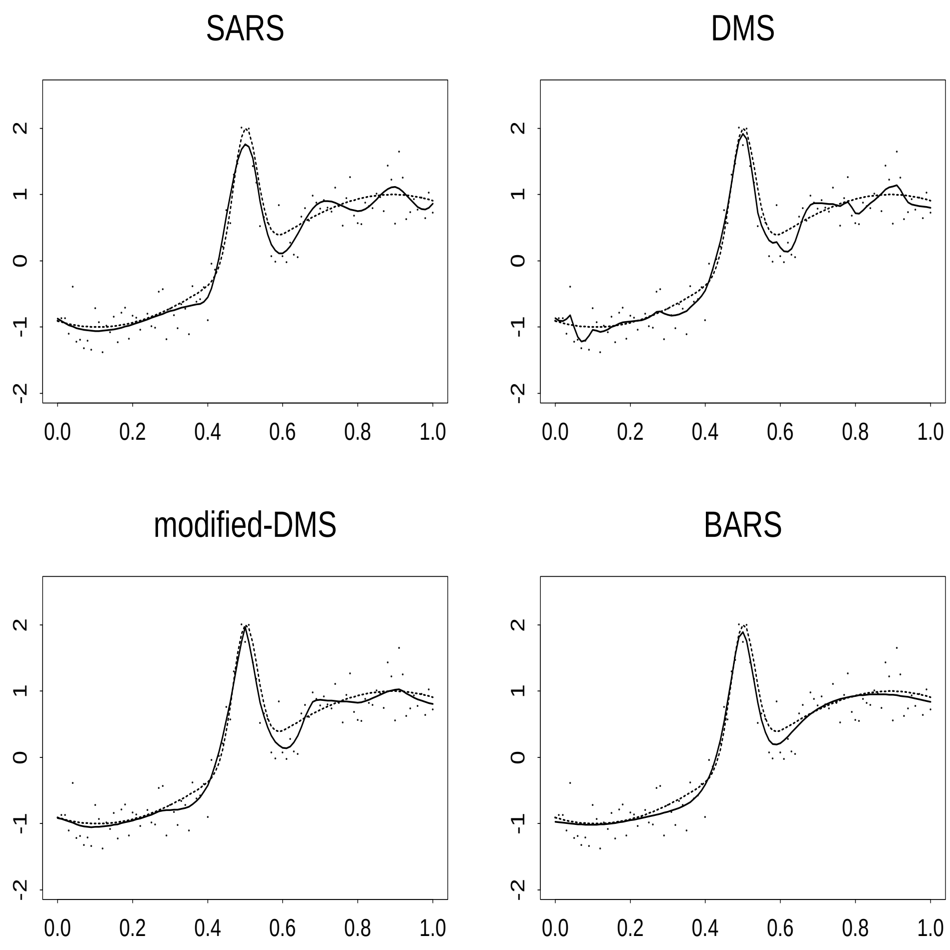
$$p(y|\xi, k) = \int p(y|\beta, \xi, k, \sigma)\pi(\beta, \sigma|\xi, k)d\beta d\sigma.$$

- Get: (i) estimate, from posterior mean $\hat{f}(x) = E(f(x)|y)$ (which is a mixture of splines, or could use mode) and (ii) uncertainty, from full posterior.

Average MSE (with simulation SE)

	SARS	DMS	BARS
Example 2	0.015 (0.001)	0.025 (0.002)	0.008 (0.001)

SARS: Zhou and Shen (*JASA*, 2001)
DMS: Denison, Mallick, and Smith (*JRSSB*, 1998)


SARS
DMS
modified-DMS
BARS
-2
-1
0
1
2
0.0
0.2
0.4
0.6
0.8
1.0

- Note: Many other Bayesian (e.g., Biller, 2000; Smith and Kohn, 1996) and frequentist approaches to spline-based fitting.

  See Hansen and Kooperberg (2002, *Statistical Science*) for review and discussion.

  My view:

- Note: Many other Bayesian (e.g., Biller, 2000; Smith and Kohn, 1996) and frequentist approaches to spline-based fitting.

  See Hansen and Kooperberg (2002, *Statistical Science*) for review and discussion.

  My view:
  - BARS is powerful, but
  - slow.
  - LOGSPLINE can help.

Generalization

$$\begin{aligned} Y_i &\sim p(y|\theta_i, \zeta) \quad \text{e.g., Poisson} \\ \theta_i &= f(x_i) \end{aligned}$$

Generalization

$$\begin{aligned} Y_i &\sim p(y|\theta_i, \zeta) \quad \text{e.g., Poisson} \\ \theta_i &= f(x_i) \end{aligned}$$

- Easy. Everything else stays the same, except:

  integral $p(y|\xi, k)$ can not be evaluated analytically; we use BIC, following Kass and Wasserman (1995) (Laplace's method).

Generalization

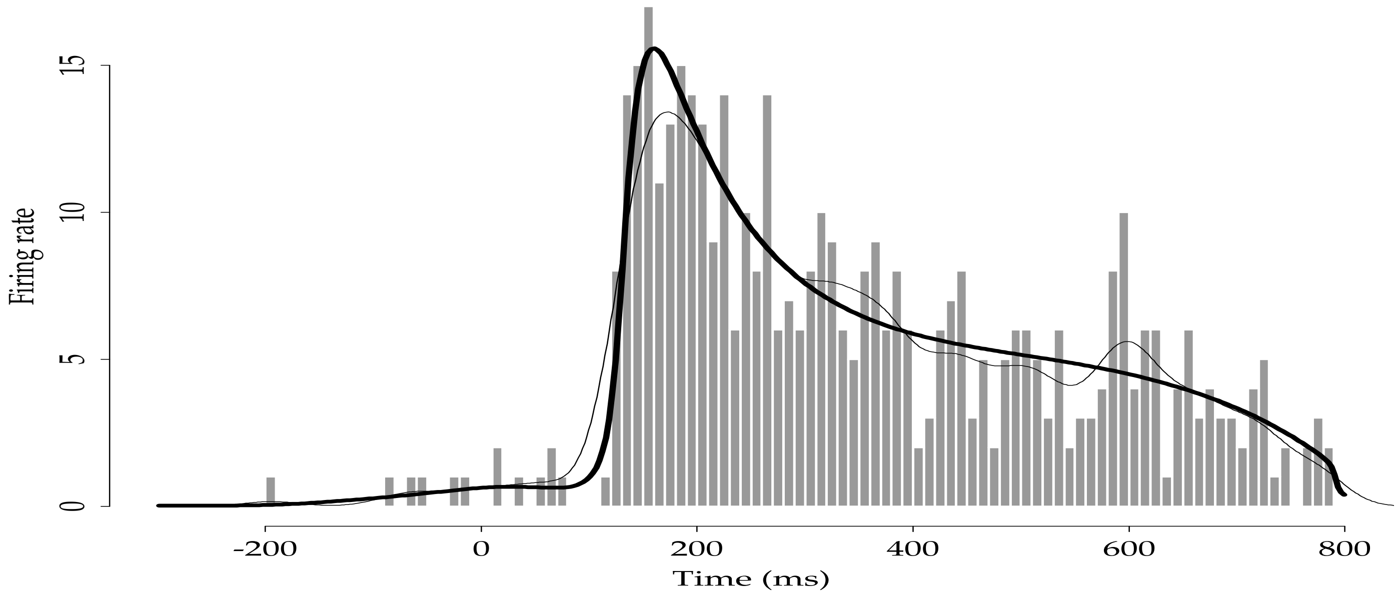
$$\begin{aligned} Y_i &\sim p(y|\theta_i, \zeta) \quad \text{e.g., Poisson} \\ \theta_i &= f(x_i) \end{aligned}$$

- Easy. Everything else stays the same, except:

  integral $p(y|\xi, k)$ can not be evaluated analytically;
  we use BIC, following Kass and Wasserman (1995)
  (Laplace's method).

- In neuron firing example:

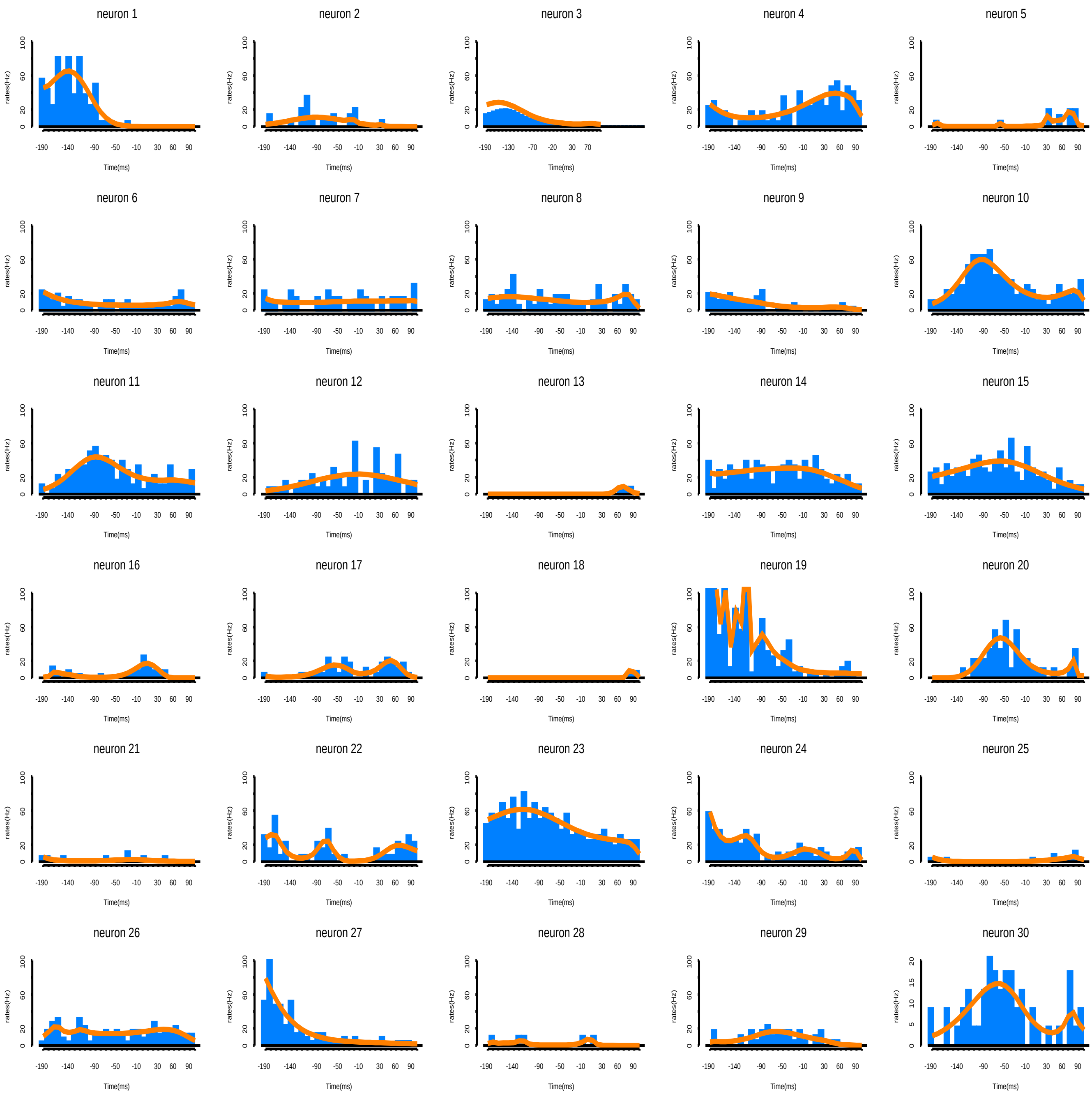
  for each (ξ, k) fit Poisson generalized linear model;
  evaluate likelihood; get BIC which (with prior on (ξ, k)) defines
  Markov chain.


Firing rate
0
5
10
15
-200
0
200
400
600
800
Time (ms)

Simultaneous Multiple Curve-Fitting

“Bayesian Functional Data Analysis with Free-Knot Splines”

Motivation



neuron 1
neuron 2
neuron 3
neuron 4
neuron 5
neuron 6
neuron 7
neuron 8
neuron 9
neuron 10
neuron 11
neuron 12
neuron 13
neuron 14
neuron 15
neuron 16
neuron 17
neuron 18
neuron 19
neuron 20
neuron 21
neuron 22
neuron 23
neuron 24
neuron 25
neuron 26
neuron 27
neuron 28
neuron 29
neuron 30
rates(Hz)
Time(ms)
0 20 60 100
-190 -140 -90 -50 -10 30 60 90
-190 -130 -70 -20 30 70
0 5 10 15 20


neuron 1
neuron 2
neuron 3
neuron 4
neuron 5
neuron 6
neuron 7
neuron 8
neuron 9
neuron 10
neuron 11
neuron 12
neuron 13
neuron 14
neuron 15
neuron 16
neuron 17
neuron 18
neuron 19
neuron 20
neuron 21
neuron 22
neuron 23
neuron 24
neuron 25
neuron 26
neuron 27
neuron 28
neuron 29
neuron 30
rates (Hz)
Time (msec)
0 20 40 60 80 100
-190 -140 -90 -50 -10 20 50 80

Assessing Population Variability

Viewing $f(t)$ as realization from random process, want to estimate Cov$(f(u), f(v))$.

Assessing Population Variability

Viewing $f(t)$ as realization from random process, want to estimate Cov$(f(u), f(v))$.

- Also want variability of functionals (e.g., location of maximum).

Assessing Population Variability

Viewing $f(t)$ as realization from random process, want to estimate Cov$(f(u), f(v))$.

- Also want variability of functionals (e.g., location of maximum).
- FDA (Multivariate Analysis) useful descriptively (Optican and Richmond, *J. Neurophys.*, 1987).

Assessing Population Variability

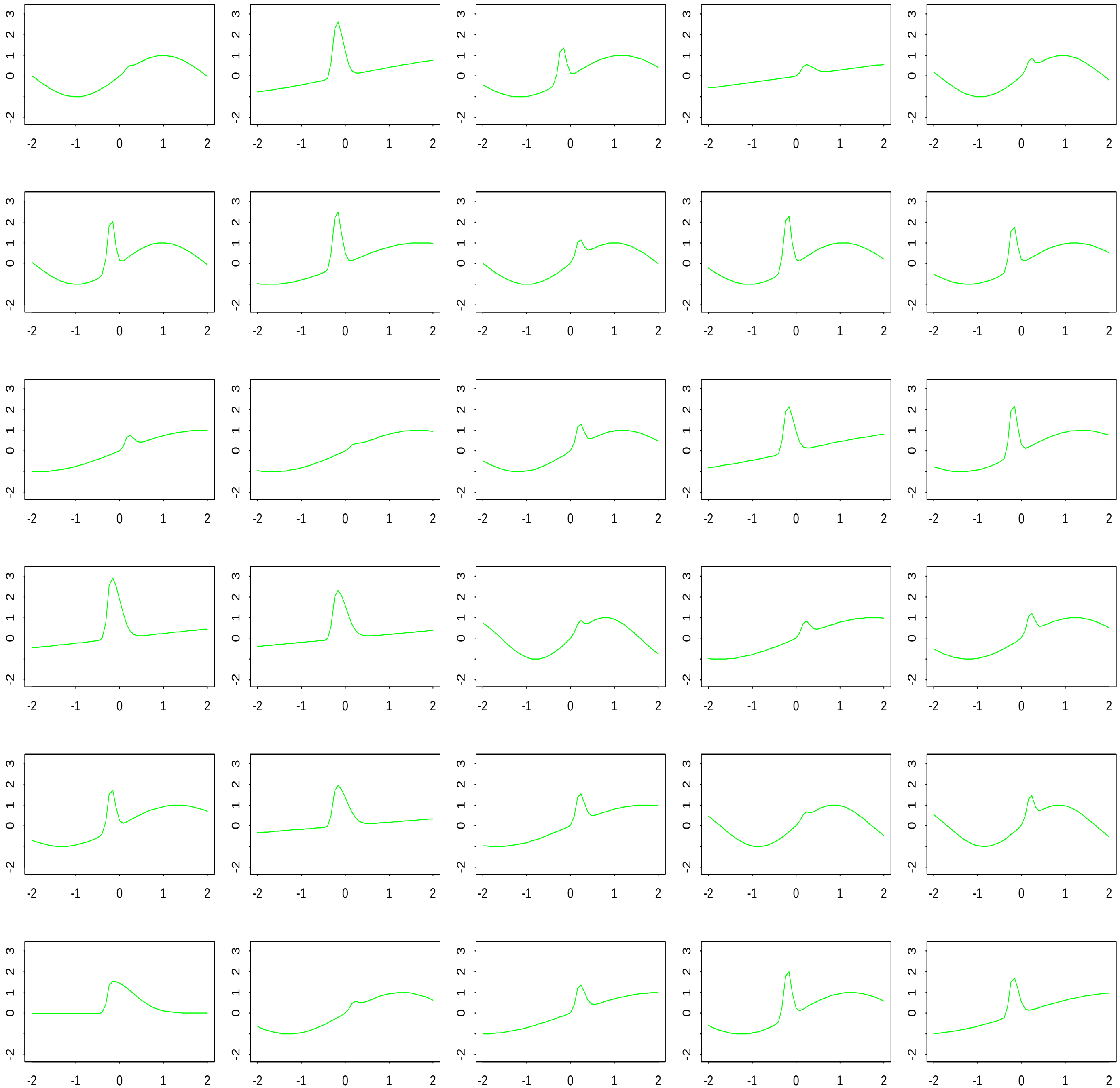
Viewing $f(t)$ as realization from random process, want to estimate Cov$(f(u), f(v))$.

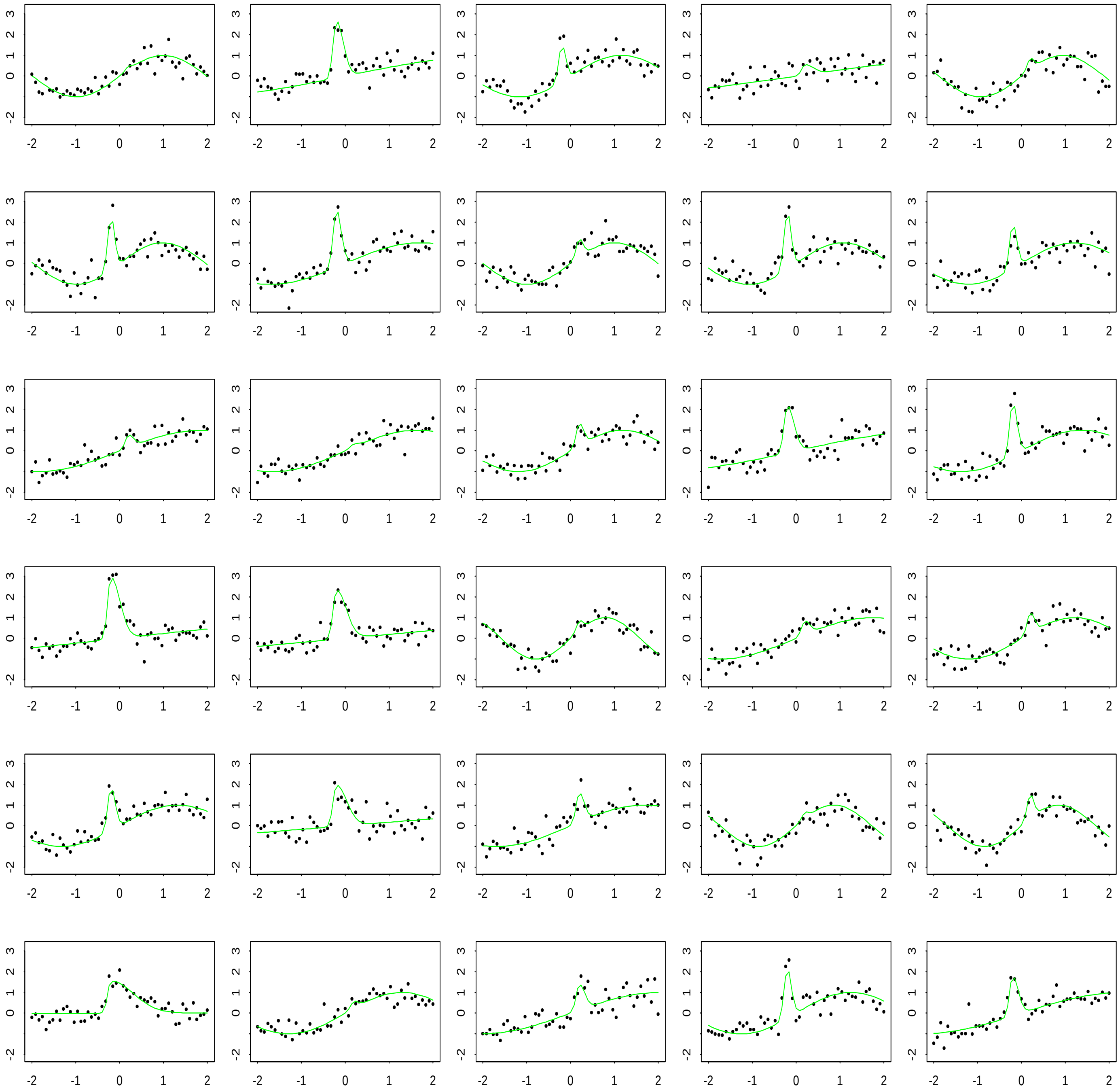
- Also want variability of functionals (e.g., location of maximum).
- FDA (Multivariate Analysis) useful descriptively (Optican and Richmond, *J. Neurophys.*, 1987).
- Multiple-curve version of BARS will
  (i) allow for occasional rapid changes in firing rate and
  (ii) take account of variability in estimating firing-rate curves.

Simultaneous Multiple Curve-Fitting

“Bayesian Functional Data Analysis with Free-Knot Splines”

Usual Curve-Fitting Case



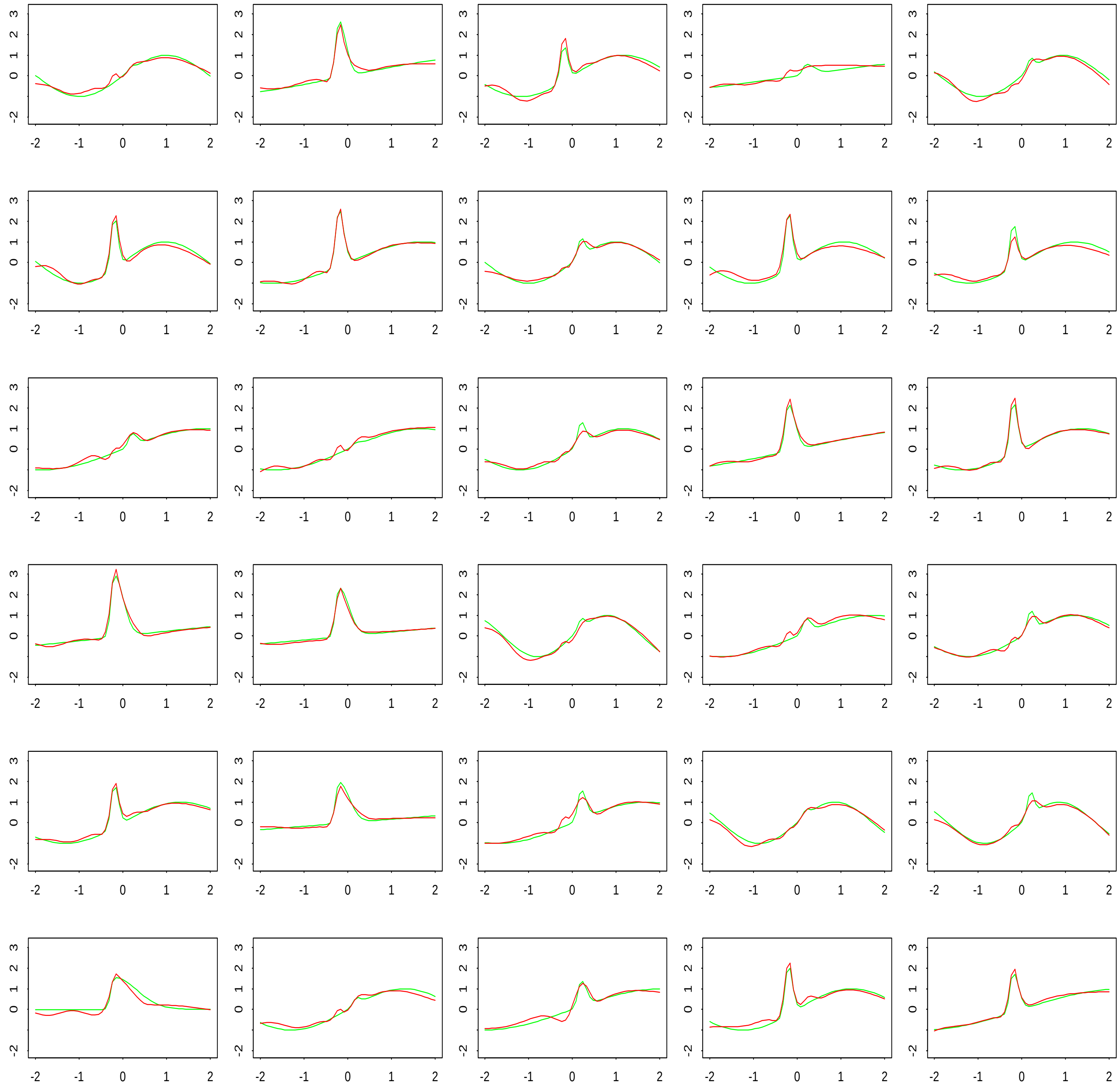

-2
-1
0
1
2
3

Usual curve-fitting case: linear mixed model *conditionally* on knots ξ.

$$\begin{aligned} Y_i|\xi, \alpha, \beta_i, \sigma, D &\sim N\left(X\alpha + X\beta_i, \sigma^2 I\right) \\ \beta_i|\xi, \alpha, \sigma, D &\sim N\left(0, \sigma^2 D\right) \\ (\xi, \alpha, \sigma, D) &\sim \pi\left(\xi, \alpha, \sigma, D\right) \end{aligned}$$

Because all curves have same knots, variation among curves becomes variation among spline coefficients.

Note: $Y_i|\xi, \alpha, \sigma, D \sim N\left(X\alpha, \sigma^2\left(I + XDX^T\right)\right)$



Simultaneous Multiple Curve-Fitting

“Bayesian Functional Data Analysis with Free-Knot Splines”

Poisson Case

Simple General Version, in Two Stages: Apply single-curve method,

$$
\begin{aligned}
Y_i &\sim p(y|\theta_i, \zeta) \quad \text{e.g., Poisson} \\
\theta_i &= \sum_{j=1}^{k+2} b_j(x)\beta_j .
\end{aligned}
$$

except now do so simultaneously across curves under *constraint*: all curves use *same* knots.

Get MLEs $\hat{\beta}_i$, information-based covariance matrices $\hat{\Sigma}_i$, then apply

$$\begin{aligned} \hat{\beta}_i &\sim N\left(\beta_i, \hat{\Sigma}_i\right) \\ \beta_i &\sim N\left(\alpha, D\right) \\ \left(\alpha, D\right) &\sim \pi\left(\alpha, D\right) \end{aligned}$$

Get MLEs $\hat{\beta}_i$, information-based covariance matrices $\hat{\Sigma}_i$, then apply

$$\begin{aligned} \hat{\beta}_i &\sim N\left(\beta_i, \hat{\Sigma}_i\right) \\ \beta_i &\sim N\left(\alpha, D\right) \\ (\alpha, D) &\sim \pi\left(\alpha, D\right) \end{aligned}$$

Note: Daniels and Kass (1998) discuss correction of this approximation by importance sampling.

Proportion of Variance Results

From 30 neurons

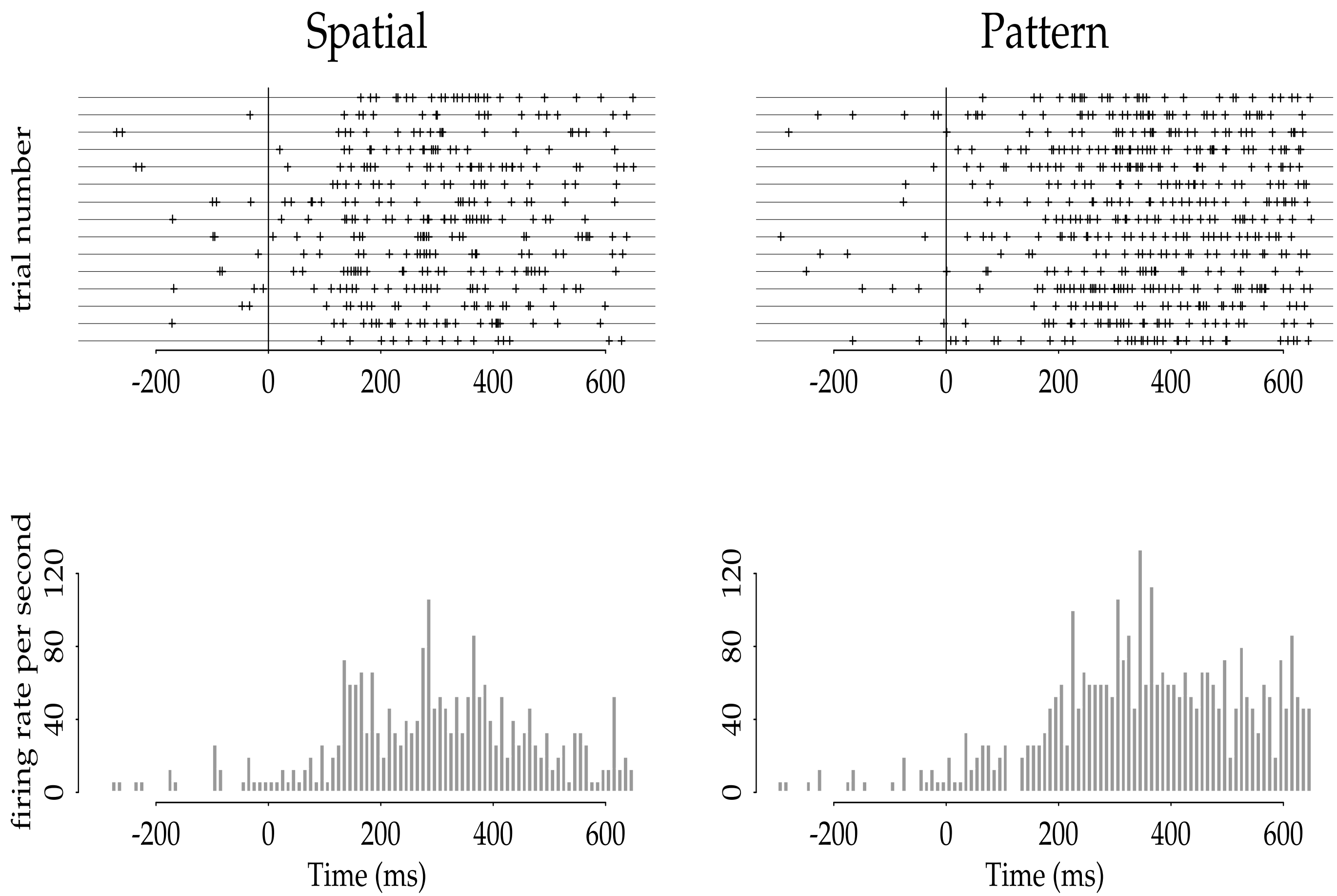
First principal component:

$$.697 \pm .092$$

Compare FDA: .78.

In some neuronal applications, variability due to estimation of curves matters enough to be worth accounting for.

Application (in progress): Trial-to-Trial Variability



Spatial
Pattern
trial number
-200
0
200
400
600
firing rate per second
0
40
80
120
Time (ms)

Application (in progress): Trial-to-Trial Variability

- Behavior (e.g., reaction time) varies from trial-to-trial, but each trial has limited data.
- A few studies have used large numbers of trials and grouped them by response variable (reaction time).
- We have begun to use splines with low d.f. and smoothed PSTH (from data pooled across trials) as offset.
- Latencies should be included, too.

Registration, generally

- Shifts may be incorporated; can be done simultaneously;
- time-warping could be done using free-knot splines, i.e., by transforming via

$$\alpha_0 + \alpha_1 \int_0^t \exp(\sum_{j=1}^{k+2} b_j(u)\beta_j)du$$

  as in Ramsay (1998).
- Here, as elsewhere, must consider benefit of BARS approach versus cost.

Summary and Discussion

- Neural timing relationships may be studied via inhomogeneous point processes, with intensities $\lambda(t)$.
- When the intensity varies rapidly in some interval(s) of time, Bayesian Adaptive Regression Splines (BARS) improves fits and inferences.
- (By the way, BARS has also been applied to fMRI, EEG, EMG data.)
- Our “Bayesian Functional Data Analysis via Free-knot Splines” refers to the extension of BARS, etc., to multiple curves.

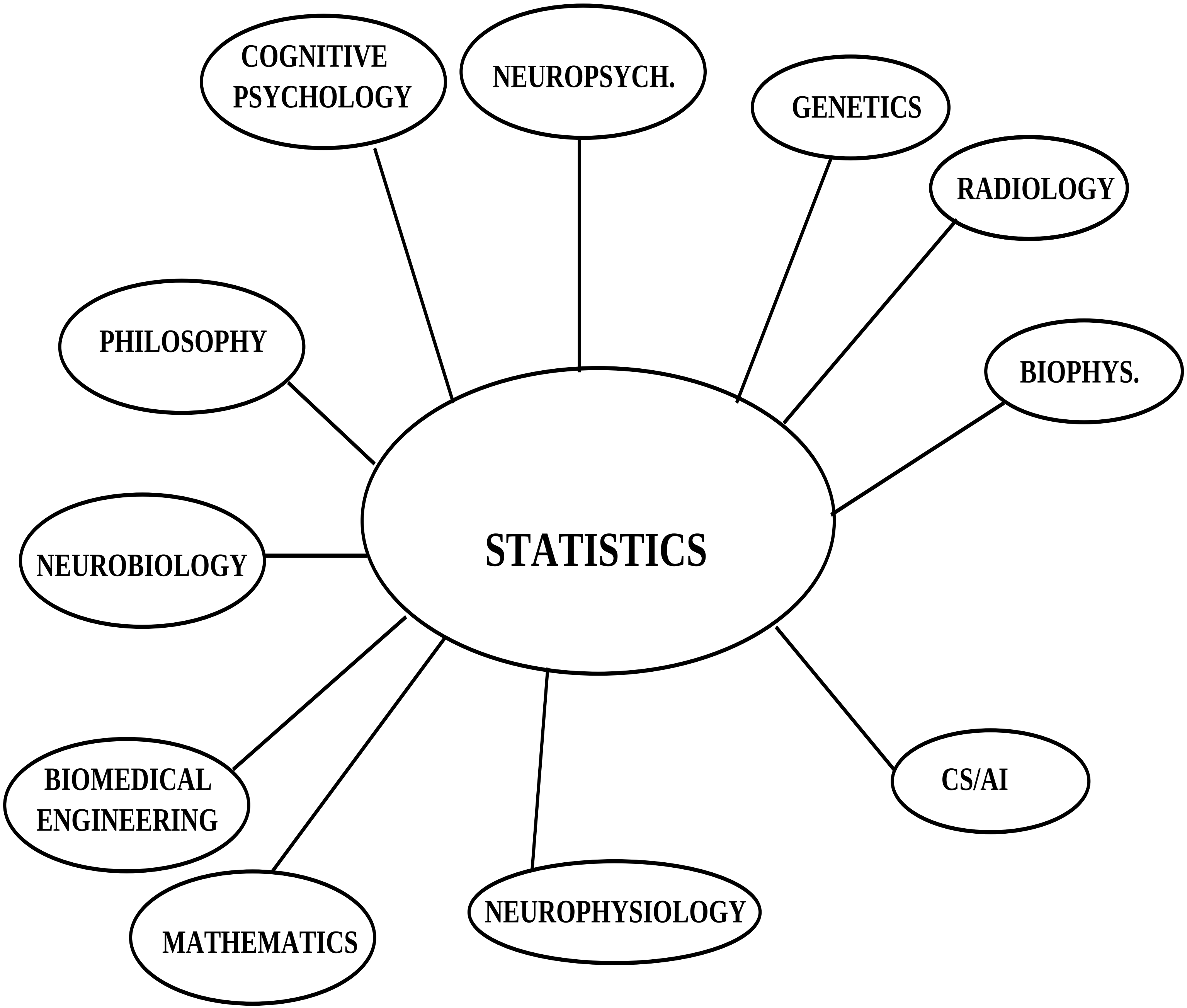
Ongoing Work

- We are investigating BARS vs. LOGSPLINE vs. BARS + LOGSPLINE (with Mark Hansen and Charles Kooperberg), and anticipate extension to multiple curves.
- We plan to work on various additional speed-up ideas and comparisons:
  - smoothing splines (e.g., Ke and Wang, 2001),
  - wavelets (Vannucci *et al.*, 2003; Genovese and Wasserman, 2002),
  - methods for large data sets (Komarek and Moore, 2003).
- Of course, all of this emphasizes the neural context.

(Continued ... A Final Thought)

A Statistician’s View of Cognitive Neuroscience

A Statistician’s View of Cognitive Neuroscience



A Statistician's View of Cognitive Neuroscience