Complete Functional Characterization of Sensory Neurons by System Identification

Michael C.-K. Wu,[1] Stephen V. David,[2] and Jack L. Gallant[3,4]

[1]Biophysics Graduate Group, [3]Department of Psychology, [4]Program in Neuroscience,University of California, Berkeley, California 94720; email: gallant@berkeley.edu

[2]Institute for Systems Research, University of Maryland, College Park, Maryland 20742

Annu. Rev. Neurosci.
2006. 29:477–505

The *Annual Review of Neuroscience* is online at neuro.annualreviews.org

doi: 10.1146/annurev.neuro.29.051605.113024

0147-006X/06/0721-0477$20.00

Key Words

reverse correlation, receptive field, nonlinear functional model, regression, maximum a posteriori estimate, prediction

Abstract

System identification is a growing approach to sensory neurophysiology that facilitates the development of quantitative functional models of sensory processing. This approach provides a clear set of guidelines for combining experimental data with other knowledge about sensory function to obtain a description that optimally predicts the way that neurons process sensory information. This prediction paradigm provides an objective method for evaluating and comparing computational models. In this chapter we review many of the system identification algorithms that have been used in sensory neurophysiology, and we show how they can be viewed as variants of a single statistical inference problem. We then review many of the practical issues that arise when applying these methods to neurophysiological experiments: stimulus selection, behavioral control, model visualization, and validation. Finally we discuss several problems to which system identification has been applied recently, including one important long-term goal of sensory neuroscience: developing models of sensory systems that accurately predict neuronal responses under completely natural conditions.

Contents

1. INTRODUCTION 478
2. THE MAXIMUM A POSTERIORI ESTIMATION FRAMEWORK 480
2.1. The Model Class 482
2.2. The Noise Distribution (Loss Function) 484
2.3. The Prior (Regularizer) 484
3. COMPUTING THE STIMULUS-RESPONSE MAPPING FUNCTION 486
3.1. Properties of the MAP Objective Function 486
3.2. Number of Model Parameters 488
3.3. Number of Free Parameters .. 488
4. EXPERIMENTAL CONSIDERATIONS FOR SYSTEM IDENTIFICATION ... 489
4.1. Stimulus Selection 489
4.2. Behavioral Control 491
4.3. Visualization and Interpretation 491
4.4. Prediction and Validation 491
5. APPLICATIONS 493
5.1. Tuning Properties 493
5.2. Nonlinear Response Properties 494
5.3. Functional Anatomy 494
5.4. Natural Stimuli 495
5.5. Cognitive Factors 495
5.6. Optimal Stimuli 496
5.7. Predictive Power 496
6. FUTURE DIRECTIONS 497
7. CONCLUSIONS 497

1. INTRODUCTION

Sensory neuroscience is one of the most quantitative and computationally sophisticated areas of neuroscience. Sensory systems are particularly amenable to a computational approach because sensory input can be tightly controlled during experiments and quantified for analysis. Quantitative neurophysiological studies have produced many computational models that describe the functional relationship between sensory inputs and neuronal responses, and there is now broad agreement about the first-order computational description of many sensory systems. Quantitative and computational models are commonly used to describe sensory processing in many simple animals (Frye & Dickinson 2001), in the peripheral mammalian auditory (McAlpine 2005) and visual systems (Carandini et al. 2005), and in primary visual cortex (Carandini et al. 1997, Daugman 1988).

The success of computational modeling has inspired researchers to apply these methods to more complex systems. However, progress on these more difficult problems has been hampered by several interrelated factors. Much of the theoretical and experimental work in neurophysiology uses specialized synthetic stimuli optimized to probe specific stimulus attributes. The resulting models provide a good description of neuronal processing of these synthetic stimuli, but it is unclear how general the models are. Natural stimuli have a much richer and more complex statistical structure than those typically used in neurophysiological experiments (Simoncelli & Olshausen 2001). Do models developed using synthetic stimuli accurately describe neuronal function under realistic conditions? This question can be addressed by examining how well each model predicts responses to natural stimuli, during natural behavior (David et al. 2004, Theunissen et al. 2000). Only by testing models under natural conditions can we determine where they fail, and only by investigating these failures is it possible to construct more general and powerful models.

The computational modeling approach has been used effectively to characterize simple systems and the sensory periphery, but it has been less successful when applied to the more central sensory systems of mammals. For example, although there are good first-order computational models of primary

visual cortex (Carandini et al. 1997) and area MT (Simoncelli & Heeger 1998), there are no comparable models for other visual areas. Similarly, quantitative models have been developed for neurons of the mammalian auditory periphery (McAlpine 2005), yet there is no agreement on an appropriate model for primary auditory cortex. The failure to develop models of more central sensory areas suggests these areas are less amenable to computational approaches. This is likely a result of several key differences between neurons in more central sensory areas and those in peripheral areas. Central neurons are more nonlinear than peripheral neurons (Carandini et al. 2005, Ito et al. 1995); they represent more abstract, semantic, and behaviorally relevant stimulus attributes (Gallant et al. 1996, Kobatake & Tanaka 1994); and they are more highly modulated by attention, learning, and memory (Reynolds & Chelazzi 2004). All of these properties make it difficult to apply approaches developed for simple systems to more complex central sensory systems.

This chapter reviews recent efforts to address these problems using system identification (SI), an approach that is receiving increasing attention in sensory neuroscience. In the classical approach to sensory neurophysiology, each experiment is optimized to test a specific hypothesis about neuronal coding. In contrast, SI treats the problems of sensory receptive field estimation as a regression problem. SI aims to construct a quantitative model that describes how a neuron will respond to any potential stimulus.

Variants of the SI approach go by many names: SI (Marmarelis & Marmarelis 1978, Marmarelis 2004), receptive field estimation (Theunissen et al. 2001), spike-triggered analysis (Rust et al. 2005, Touryan et al. 2002), reverse correlation (Alonso et al. 2001, Jones & Palmer 1987, Ringach et al. 1997b, Smyth et al. 2003), and white-noise analysis (Chichilnisky 2001). In each case the goal is to estimate a function that describes the way stimuli are mapped onto neuronal responses. The function that relates stimulus to response also goes by many names: the spectrotemporal or spatiotemporal receptive field (Aertsen & Johannesma 1981b; David et al. 2004; DeAngelis et al. 1993; Klein et al. 2000; Ringach et al. 1997b; Theunissen et al. 2001), the kernel (Smyth et al. 2003, Willmore & Smyth 2003), the stimulus-response function (Sahani & Linden 2003a), the transducer (Victor & Shapley 1980), or the transfer function (Cook & Maunsell 2004, Marmarelis & Marmarelis 1978). In this chapter, we refer to the general strategy for understanding sensory processing as the SI approach and the function that maps stimulus to response as the mapping function.

The SI approach offers powerful tools for addressing the problems noted above. The core of the method focuses on quantitative modeling of experimental data and the evaluation of models in terms of predictive power (Marmarelis 2004). The approach can also be used to test an existing model by redefining it within the SI framework. Responses that are not predicted by the model can then be investigated to determine how the model might be improved (David et al. 2004). Predictions also provide a straightforward method for comparing models across studies (David & Gallant 2005). Finally, the SI approach can be used to reveal novel principles of sensory coding, even in the absence of a prior theoretical or quantitative model. In this case responses can be modeled using a more open-ended framework and a rich stimulus set. Visualization procedures can then used to interpret the estimated mapping function.

One complication of the SI approach is that there is a large diversity of SI algorithms, each making different assumptions about the underlying nature of the system. Here we show that virtually all of these algorithms can be interpreted as maximum a posteriori (MAP) estimates within the framework of Bayesian inference (Robert 2001). In essence, the MAP estimate describes the most probable model, given the available data and prior knowledge of the system. Apparent

SI: system identification

mapping function: the function that describes the relationship between a sensory stimulus and neuronal response

MAP: maximum a posteriori

MAP estimate (estimator): the model parameter values that specify a mapping function that is both plausible and fits the data well

differences between the various methods used to estimate the mapping function merely reflect differences in how the MAP estimator is constructed. Bayesian inference may be foreign to many readers and its exposition is somewhat complicated, but the effort is worthwhile: By viewing each algorithm as a MAP estimator, we can clearly compare various methods and determine which procedure is optimal for a given problem.

This review is divided into three parts. First, we describe the MAP estimation framework in detail (Section 2) and show how current SI algorithms fit within that framework (Section 3). These sections are necessarily quite technical, but we attempt to present the information in an approachable way. Second, we summarize the practical issues that must be considered for successful estimation of the mapping function in practice (Section 4). Finally, we summarize recent results obtained with the SI approach and discuss avenues for further research (Sections 5 and 6). Volterra and Wiener kernel analysis are described thoroughly in several excellent textbooks (e.g., Marmarelis & Marmarelis 1978, Marmarelis 2004). Therefore we focus on other methods that have been developed more recently.

model class: a restricted class of models used to fit the data; each model in the class is defined by a set of model parameters

noise distribution: a probability distribution that describes the assumed stochastic variability in the response

prior: a probability distribution that describes our subjective belief in the plausibility of each model by specifying the probability of the possible values of the model parameters

MAP objective function: the function that must be minimized to obtain the optimal MAP estimate

2. THE MAXIMUM A POSTERIORI ESTIMATION FRAMEWORK

According to the SI approach, functional characterization of sensory neurons is essentially a regression problem: Given a data set consisting of stimulus-response pairs, estimate the function that maps any stimulus to the appropriate response (Marmarelis 2004, Theunissen et al. 2001). Because neurons are both complicated and noisy, this estimate is usually data limited (Victor 2005). Therefore, it can be improved by incorporating reasonable assumptions and prior knowledge about the system studied into the regression procedure. The MAP estimation framework is a systematic procedure for combining these assumptions with experimental data to obtain the most probable estimate of the stimulus-response mapping function (Robert 2001). The MAP estimation framework is more general than the simple regression approach with which many readers may be familiar; simple regression merely involves the currently available data and does not explicitly consider other information that might be known about the system.

Here we characterize the three constituents of the MAP estimation framework: the model class, noise distribution, and prior, and we use them to classify current SI algorithms (**Figure 1**). This unified framework clarifies the implications of choosing one algorithm over another and facilitates the design and implementation of novel algorithms.

MAP estimation is an inference problem: Infer the most probable mapping function, $y = f_\theta(\mathbf{x})$, from the observed data. The input, $\mathbf{x}$, is a vector representing the sensory stimulus. It may span several points in time to account for the memory of the system. The output, y, is a scalar representing neuronal activity. This might be spike counts (Jones & Palmer 1987), membrane potential (Bringuier et al. 1999, Priebe & Ferster 2005), local field potential (Victor et al. 1994), or instantaneous firing rate (Theunissen et al. 2001). The observed data set consists of N stimulus-response sample pairs, $\{\mathbf{x}_i, y_i\}$, where $i = 1, 2, \ldots, N$. The model class determines the form of the mapping function, f_θ, and the parameters of the model are given by the vector, $\boldsymbol{\theta}$. The most probable value of $\boldsymbol{\theta}$ is the MAP estimate (denoted $\boldsymbol{\theta}^*$),

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left[\underbrace{-\sum_{i=1}^{N} \log p\left(y_i \mid f_{\boldsymbol{\theta}}(\mathbf{x}_i)\right)}_{\text{Negative-log likelihood (Loss function)}} \underbrace{- \log p(\boldsymbol{\theta})}_{\text{Negative-log prior (Regularizer)}} \right]. \quad 1.$$

Likelihood: Noise distribution, Model class. Prior.

The quantity in the square bracket of Equation 1 is the MAP objective function. This function contains three constituents: the model class, noise distribution, and prior.

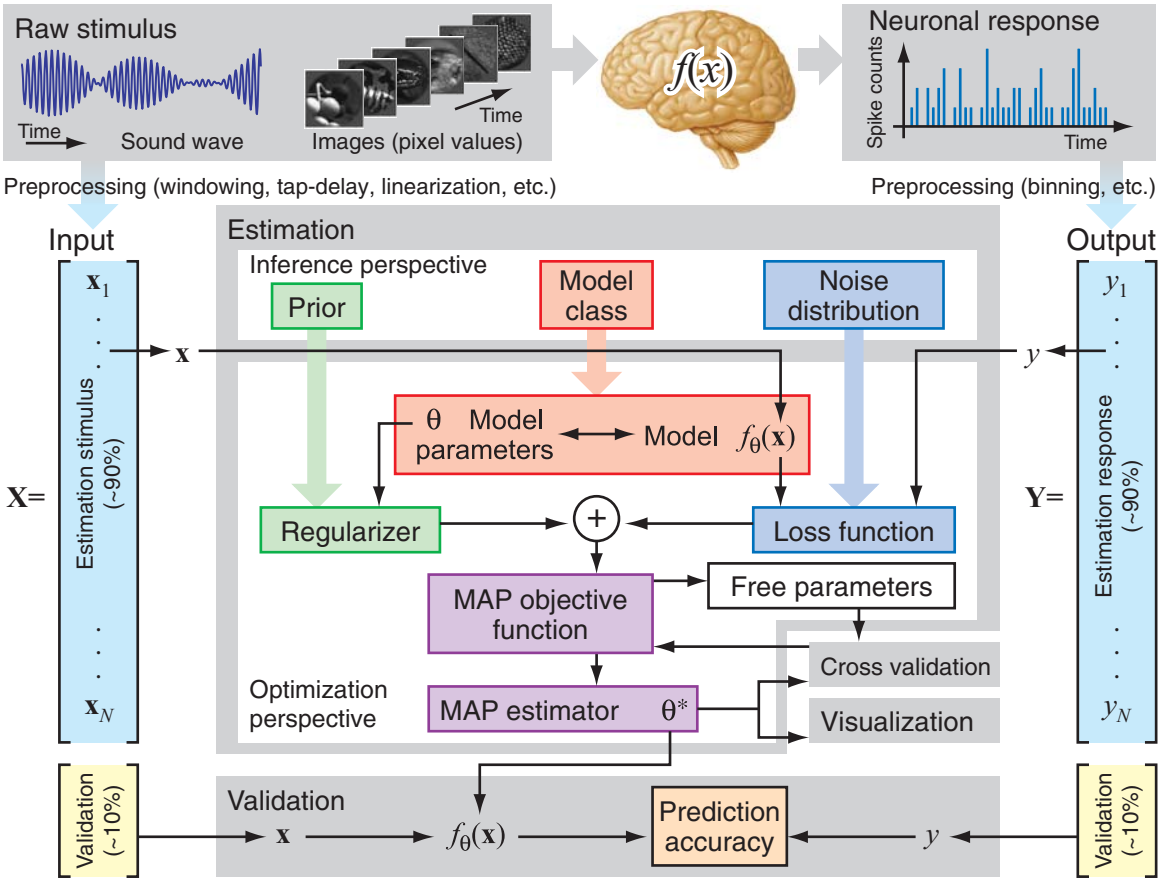

Figure 1

General overview of system identification in the maximum a posteriori (MAP) framework. (*Top row*) The system identification approach aims to estimate the stimulus-response mapping function for a single neuron. (*Left and right columns*) Stimulus and response data are first split into two independent sets. The estimation set (~90% of the data, *cyan*) is used to fit the model, and the validation set (~10% of the data, *yellow*) is used to validate model predictions. (*Center panel*) System identification can be viewed in two perspectives: as an inference problem (*top*) or as an optimization problem (*bottom*). From the inference perspective, the MAP estimator consists of three constituents: the model class (Section 2.1), noise distribution (Section 2.2), and prior (Section 2.3). From the optimization perspective, the model class determines the form of the mapping function, $f_\theta(\mathbf{x})$, and its model parameters, θ. The MAP estimate, θ^*, is obtained by combining the regularizer (specified by the prior) and the loss function (specified by the model class and noise distribution), and then finding the value of θ that minimizes the resulting MAP objective. Free parameters are optimized as necessary by cross-validation (Section 3.3). After estimation, θ^* can be visualized to aid interpretation (Section 4.3). (*Center bottom column*) Model performance is assessed by measuring how well the estimated mapping function predicts responses in the reserved validation set (Section 4.4).

[Each $p(\cdot)$ denotes a probability distribution.] The model class specifies the form of f_θ used to fit the data. The noise distribution describes the stochastic variability in the response. The prior describes the subjective belief in the plausibility of each possible model, independent of the estimation data.

The three constituents of the MAP objective are grouped into two terms, the likelihood and the prior. The likelihood term

posterior distribution: a probability distribution that describes the probability of a model by combining the likelihood and prior information

likelihood: a probability that describes the plausibility of each model based on the model's ability to fit the data under the assumed noise distribution

loss function: a function that measures the lack of fit between a particular model and the observed data

regularizer: a function that measures how well the model satisfies the constraints imposed by the prior

comprises two of the constituents: model class and noise distribution. When combined, the prior and the likelihood give the posterior distribution that describes the probability of a model (Kass et al 2005, Robert 2001). The MAP estimate is the set of parameter values that maximizes the posterior distribution. To find the MAP estimate, we minimize the negative log of posterior distribution (Robert 2001). (This is a computational trick: Minimization of the negative log is equivalent to maximization.) Minimizing the negative-log likelihood gives the best fit to the observed data regardless of its plausibility; minimizing the negative-log prior biases the estimate of θ toward models that are more plausible (Mackay 1995, Robert 2001). Thus the likelihood and prior work in tandem to give the most probable parameter estimates.

So far we have treated MAP estimation as an inference problem. In practice MAP estimation is implemented as an optimization problem (**Figure 1**). In the optimization view the negative-log likelihood and negative-log prior are called the loss function and regularizer, respectively (Equation 1). The loss function measures the residual error between f_θ and the data. The regularizer measures how much the model violates the constraints imposed by the prior. Note that the inference view and the optimization view offer two perspectives of the same problem. Although the terminology is different, the problem is the same.

2.1. The Model Class

The model class determines the form of the mapping function, f_θ. Each model class implies strong assumptions about the functional properties that a neuron might exhibit. To obtain a good estimate of the mapping function, it is therefore important to choose a model class that can provide a good description of the neuron.

Common model classes used for neuronal SI are listed in **Table 1**. The simplest model class is the linear model (De Boer 1967, Marmarelis & Marmarelis 1978). Linear models have been used to measure temporal, spatial, and spatiotemporal properties in the auditory (De Boer 1967, Eggermont 1993, Eggermont et al. 1983), visual (DeAngelis et al. 1993, Jones & Palmer 1987), and somatosensory systems (Arabzadeh et al. 2005, DiCarlo et al. 1998). Linear models are easy to fit, but they cannot capture the nonlinear properties of neurons found at more central stages of sensory processing (David & Gallant 2005, DeAngelis et al. 1993, Eggermont 1993).

The model class first used to describe nonlinear sensory neurons is the Wiener or Volterra series (Aertsen & Johannesma

Table 1 Model classes used in neuronal system identification*

Model class		Model form	Model parameter θ
Linear models		$f_\theta(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$	$\theta = \boldsymbol{\beta}$
Nonlinear models			
Parametric (data independent)	Nonlinear Wiener/Volterra	$f_\theta(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta} + \mathbf{x}^T\mathbf{B}\mathbf{x}$	$\theta = \{\boldsymbol{\beta}, \mathbf{B}\}$
	Linearized model	$f_\theta(\mathbf{x}) = \mathbb{L}(\mathbf{x})^T\boldsymbol{\beta}$	$\theta = \boldsymbol{\beta}$ and $\mathbb{L}$
Nonparametric (data dependent)	Neural networks	$f_\theta(\mathbf{x}) = a + \mathbf{w}^T \tanh(\mathbf{b} + \mathbf{U}^T\mathbf{x})$	$\theta = \{a, \mathbf{b}, \mathbf{w}, \mathbf{U}\}$
	Kernel regression	$f_\theta(\mathbf{x}) = \langle\Phi(\mathbf{x}), \boldsymbol{\beta}\rangle$	$\theta = \boldsymbol{\beta}$ and Φ

*Each model is specified by a general parameter vector, θ. The definition of θ for each model class is shown in column 3. $\boldsymbol{\beta}$ is a vector of linear coefficients, $\mathbf{B}$ is a symmetric matrix of quadratic coefficients. Nonlinear Wiener/Volterra models in general have higher-order terms but are limited to second order in practice. $\mathbb{L}$ denotes a linearizing transform. For artificial neural network models, $\mathbf{U}$ is a matrix of input weights, $\mathbf{b}$ is a vector of input biases, $\mathbf{w}$ is a vector of output weights, a is a scalar output bias, and **tanh** is a function that is applied component-wise to its argument. For kernel regression models, Φ is the feature map that nonlinearly transforms the stimulus into a high (possibly infinite) dimensional feature space, and $\langle\cdot,\cdot\rangle$ is a general inner product.

1981b, Emerson et al. 1987, Victor & Shapley 1980). In this case the linear model is considered to be the first-order term of a Wiener or Volterra series expansion of the mapping function. Therefore, a nonlinear model can be constructed by adding higher-order terms to the series (Marmarelis & Marmarelis 1978). Second-order Wiener/Volterra models have been used in the auditory (Eggermont 1993, Lewis & van Dijk 2004) and visual systems (Anzai et al. 2001, Emerson et al. 1987, Gaska et al. 1994, Mancini et al. 1990, Victor & Shapley 1980). One commonly used SI algorithm that employs a second-order model is the method of spike-triggered covariance (Brenner et al. 2000; Rust et al. 2005; Touryan et al. 2002, 2005). It is usually not feasible to estimate terms of the full Wiener/Volterra models beyond second order because the amount of data required to estimate each higher-order term scales exponentially with the order of the model (Victor 2005). However, if the stimulus is restricted to a relevant low-dimensional subspace, higher-order Volterra models can be computed by the Volterra relevant-space technique (Rapela et al. 2006).

Another strategy for modeling nonlinear responses is to use a linearized model. In this approach the stimulus is transformed nonlinearly into a new space where the relationship between the stimulus and response is more linear (Aertsen & Johannesma 1981b). This linearizing transform is often inspired by biophysical considerations or previous neurophysiological studies. Linearized models are efficient; a good linearized model can describe complex nonlinearities while requiring relatively few model parameters (David & Gallant 2005). A linearized model was first developed to describe functional properties in the auditory periphery (Aertsen & Johannesma 1981b). Since then, linearized models have been used in central auditory areas (Kowalski et al. 1996, Machens et al. 2004, Theunissen et al. 2000) and in primary visual cortex (Bredfeldt & Ringach 2002, David et al. 2004, Mazer et al. 2002, Nishimoto et al. 2005, Nykamp & Ringach 2002, Ringach et al. 1997a).

The nonlinear models described thus far are parametric: Their complexity is fixed and does not grow with the data. In theory the Wiener/Volterra series is nonparametric because the number of terms in the series could potentially increase as more data become available. In practice, however, Wiener/Volterra models are parametric because they are truncated at a fixed order (usually second order), and this order is maintained regardless of how much data is available. Because the fixed-order Wiener/Volterra model is parametric, it can be viewed as a linearized model whose linearizing transform is described by the nonlinear terms of the series expansion (David & Gallant 2005).

A final strategy for estimating a nonlinear stimulus-response mapping function is to use a nonparametric nonlinear model. In theory, nonparametric models can describe any stimulus-response mapping. For example, artificial neural networks (ANNs) have been used to describe area V1 complex cells (Lau et al. 2002, Lehky et al. 1992, Prenger et al. 2004), and Kernel regression has been applied to visual areas V1 and V4 (Wu & Gallant 2004). Nonparametric models can describe response properties that have not been modeled explicitly, but they tend to overfit to noise. Thus, when using these models, it is important to use appropriate regularization (Section 2.3) and to validate the estimated mapping function with an independent data set (Section 4.4).

Computational models of neuronal function are often described in terms of a nonlinear-linear-nonlinear (NLN) cascade. SI can also be used to fit an NLN cascade, though the fit process will depend on which SI algorithm is used. The parametric Wiener/Volterra and linearized models can be viewed as nonlinear-linear cascades, so they can be transformed to NLN form by fitting the output nonlinearity in a separate procedure (David et al. 2004, Rust et al.

ANN: artificial neural network

regularization: a computational procedure used to prevent overfitting by imposing constraints (by adding regularizers to the MAP objective) on the model

NLN: nonlinear-linear-nonlinear

2005). Many linearized models implement the initial nonlinearity in a preprocessing step; in these cases the subsequent linear-nonlinear stages are fit separately (Aertsen & Johannesma 1981b, Eggermont 1993). Nonparametric ANN and kernel regression models can fit the entire NLN cascade at once [i.e., they are universal approximators (Hammer & Gersmann 2003, Hornik et al. 1989)], so no additional fitting stages are required.

2.2. The Noise Distribution (Loss Function)

The second constituent of the MAP objective function is the noise distribution. Because noise is not predictable from the stimulus, the estimated mapping function should only model the deterministic variability of the response and not the noise. The noise distribution embodies assumptions about which aspects of the response are likely to reflect deterministic variability and which are likely to reflect noise. In the MAP objective function, the noise distribution is represented by the loss function (the negative log of the noise distribution). During the minimization of the MAP objective, the loss function will penalize any residuals that are unlikely to be a result of noise (i.e., residuals that fall in the tails of the noise distribution). For example, if the noise distribution is assumed to be Gaussian, then most noise variance will lie within two standard deviations of the mean. Hence the corresponding loss function, the square loss (Victor 2005) (**Table 2**), only lightly penalizes residuals within two standard deviations of the mean, while levying a large penalty on residuals beyond this range.

Selection of an appropriate noise distribution involves a trade-off between biophysical plausibility and computational tractability. Most SI algorithms in neurophysiology employ a square loss and therefore implicitly assume a Gaussian noise distribution. The square loss is computationally tractable, and the Gaussian noise distribution provides a good approximation to the real noise when the response is a continuous variable (e.g., instantaneous firing rate). When the response is a discrete variable, such as the spike count, a Poisson noise distribution may be more appropriate (**Table 2**) (Rieke et al. 1997).

2.3. The Prior (Regularizer)

The final constituent of the MAP objective function is the prior. The prior incorporates information about the plausibility of different model parameter values independent of the current data. A good prior will integrate useful knowledge about the system under study, thereby improving prediction accuracy (Smyth et al. 2003, Theunissen et al. 2001). In the MAP objective function, the prior is represented by the regularizer (the negative log of the prior). During minimization of the MAP objective, the regularizer will favor mapping functions that are more probable according to the prior. Because the MAP objective contains both the loss function and the regularizer, the optimal MAP estimate will reveal the model parameters that both fit the data well

Table 2 Possible noise distributions (loss functions) for neuronal system identification*

Noise distribution	$p(y\|f_\theta(\mathbf{x}))$	Loss function $L(f_\theta(\mathbf{x}), y) \equiv -\log p(y\|f_\theta(\mathbf{x}))$	Common names
Gaussian	$\exp\left\{\frac{-[y - f_\theta(\mathbf{x})]^2}{2\sigma^2}\right\}$	$\frac{1}{2\sigma^2}[y - f_\theta(\mathbf{x})]^2$	L_2 loss or square loss
Poisson	$\frac{e^{-f_\theta(\mathbf{x})}[f_\theta(\mathbf{x})]^y}{y!}$	$f_\theta(x) - y \log[f_\theta(\mathbf{x})]$	Poisson loss

*Each noise distribution is associated with a unique loss function, obtained by taking the negative log of the noise distribution. $f_\theta(\mathbf{x})$ represents the predicted mean neuronal response to stimulus $\mathbf{x}$ over repeated presentations. σ^2 is the noise variance of the neuronal response.

(by minimizing the loss function) and satisfy the constraints of the prior (by minimizing the regularizer). The MAP objective is formulated so that the prior will have less influence as the number of data samples, N, grows larger (Equation 1). For this reason the optimal MAP estimate will rely heavily on the prior only when few data are available, and it will give less weight to the prior when ample data are available. In practice, the relative influence between the loss function and the regularizer must be tuned by cross-validation to achieve an accurate estimate of the mapping function (**Figure 2**) (Section 3.3).

Many SI studies in neurophysiology use simple algorithms that do not include an explicit regularizer (Eggermont et al. 1983, Jones & Palmer 1987, Rust et al. 2005, Touryan et al. 2002). These algorithms assume implicitly that all values of the model parameters are equally likely (flat prior). Studies that specify an explicit prior often use a Gaussian prior. A Gaussian prior assumes that the model parameters (components of θ) are sampled from Gaussian distributions and that these components may be correlated among themselves (Mackay 1995). The Gaussian prior can be conveniently summarized by a covariance matrix, **A**, that describes the covariance among the components of θ. The Gaussian prior has several attractive features. It specifies a quadratic regularizer that can be minimized efficiently (Section 3.1) and it provides a good second-order approximation to more complex priors.

Table 3 summarizes several Gaussian priors used in neuronal SI. These priors are specified by different covariance matrices, **A**. Note that the elements of **A** are often specified by one or more free (regularization) parameters that must be tuned to achieve the mapping function that best predicts responses to novel stimuli (Section 3.3).

An independence prior is used in a regularization procedure called automatic relevancy determination ARD (Mackay 1995, Prenger et al. 2004, Sahani & Linden 2003a). ARD assumes that model parameters are independent, and it aims to remove irrelevant model parameters. ARD therefore tends to produce sparse models that make effective use of the available data. The spherical Gaussian prior has been used in ridge-regression algorithms for auditory SI (Machens et al. 2004).

The most effective priors are those that incorporate known properties of the mapping function. For example, a smooth prior is appropriate if the mapping function is expected

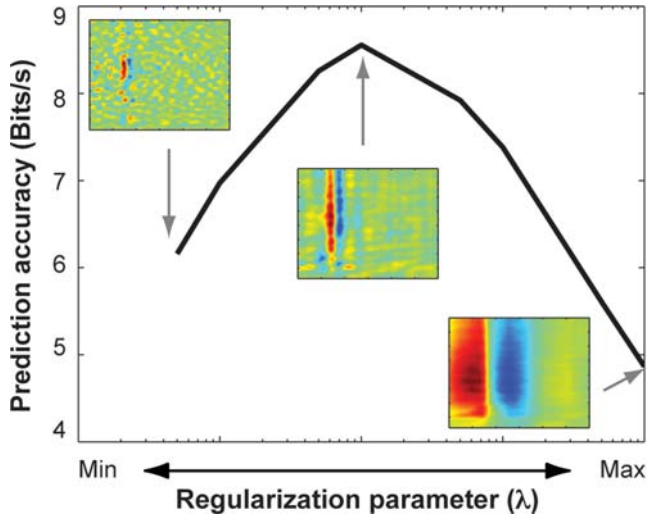

Figure 2

Regularization optimizes mapping functions estimated using natural stimuli. Natural stimuli contain strong correlations that can introduce noise into estimates of the mapping function. The effects of noise can be minimized by an appropriate regularizer. Data shown here were acquired from a single neuron in the zebra finch thalamus during stimulation by bird songs. The mapping function was estimated using a linearized spectrogram model, a Gaussian noise distribution, and a Gaussian (stimulus-subspace) prior. The stimulus-subspace prior (**Table 3**) requires specifying a regularization parameter, λ, that determines the high-frequency noise threshold. The optimal regularization parameter value was selected by cross-validation (Section 3.3). The x-axis shows the value of the regularization parameter and the y-axis shows the corresponding prediction accuracy estimated via cross-validation. Insets show mapping functions estimated using different regularization parameter values. The mapping functions are plotted as time- (*horizontal axis*) frequency (*vertical axis*) spectrograms. Excitatory tuning is shown in red and inhibitory tuning in blue. Low values of the regularization parameter do not remove much high-frequency noise, so predictions are poor. High values smooth the mapping function too much and also produce poor predictions. Prediction accuracy is maximized for intermediate values of the regularization parameter. Figure adapted from Woolley et al. (2006), by permission of the authors.

Table 3 Common Gaussian priors (regularizers) used in neuronal system identification*

Prior	Assumptions	Gaussian covariance	Regularizer
Flat prior	All values of the parameters vector are equally likely	$\mathbf{A}^{-1} = \mathbf{0}$	Constant
Spherical Gaussian	Parameters are independent with equal variance	$\mathbf{A} = \sigma_{\theta}^2 \mathbf{I}$	$\sigma_{\theta}^{-2} \|\theta\|_2^2 = \sigma_{\theta}^{-2} \sum \theta_k^2$
Independence prior	Parameters are completely independent	$\mathbf{A} = diag(\sigma_{\theta_k}^2)$	$\sum \sigma_{\theta_k}^{-2} \theta_k^2$
Smooth prior	Parameters vary smoothly over stimulus dimensions	$\mathbf{A} = \mathbb{D}^{-2}$	$\delta^{-1} \theta^T \mathbb{D}^2 \theta = \delta^{-1} \|\mathbb{D}\theta\|_2^2$
Stimulus subspace	Parameters lie within a stimulus principal component subspace	$\mathbf{A} = \mathbf{Q}\mathbf{D}_{\lambda}\mathbf{Q}^T$	$\sum_{k,d_k<\lambda} \infty (\mathbf{Q}_k^T \theta)^2$

*Each Gaussian prior is specified by a covariance matrix, **A** (column 3). The associated regularizer (column 4) is obtained by taking the negative log of the prior. σ_{θ}^2 is the variance of each model parameter, which is the same for all model parameters in the spherical prior. $\sigma_{\theta_k}^2$ is the variance of the k-th model parameter in the independence prior. $\mathbb{D}^2$ is the discretized Laplacian operator used in the smooth prior, and δ determines the scale over which θ is expected to be smooth. **Q** is the matrix of principle components (PC) of the stimulus, where $\mathbf{Q}_k$ is the k-th PC with associated eigenvalues, $d_k \cdot \mathbf{D}_{\lambda}$ is a diagonal matrix whose k-th diagonal element is infinite if $d_k \geq \lambda$ or zero if $d_k < \lambda$. The stimulus subspace prior applies a threshold operation to force parameter values to zero for those principle components with eigenvalues smaller than λ. Computationally, the regularizer of the stimulus subspace prior will levy infinite penalty on the MAP objective whenever a parameter associated with small eigenvalue is not zero.

to be smooth in space and/or time (Sahani & Linden 2003a, Smyth et al. 2003). The stimulus-subspace prior can be used to attenuate high-frequency noise when using natural stimuli to estimate the mapping functions. This prior requires specifying a regularization parameter for the noise threshold, which is usually set via cross-validation (David et al. 2004, Theunissen et al. 2001) (**Figure 2** and Section 3.3).

3. COMPUTING THE STIMULUS-RESPONSE MAPPING FUNCTION

Almost all SI algorithms can be described in terms of the MAP estimation framework (**Table 4**). Because the MAP constituents are defined independently, different algorithms can be created by combining the constituents in different ways. To date only a few of these potential algorithms have actually been used in neurophysiology.

When constructing the MAP estimator, one important consideration is how difficult it will be to minimize the MAP objective function. Three factors influence this process: the properties of the MAP objective itself, the number of model parameters that need to be fit, and the number of free parameters that must be tuned by cross-validation.

3.1. Properties of the MAP Objective Function

The difficulty of finding the global minimum of the MAP objective function is determined by two properties: whether the MAP objective is smooth and whether it has local minima. If the MAP objective is smooth and has no local minima, then the unique global minimum can be found easily with standard gradient-based optimization methods (Fletcher 1987). If the MAP objective function is not smooth or if it has local minima, then it may be difficult to find the global minimum even if one exists. If any one of the MAP constituents is not smooth or has local minima, then the entire MAP objective function will not be smooth or it will have local minima. For example, the MAP objective of an ANN is smooth, but it has local minima (because the ANN model class includes nonconvex functions). In contrast, the MAP objective of support-vector regression is not smooth (because its loss function is not smooth), but it has no local minima.

Table 4 Classification of system identification algorithms in neurophysiology

	Theoretical constituents of the MAP objective*				Minimization of MAP objective		
System identification algorithm and reference(s)	Model class	Loss function (noise distribution)	Regularizer	Assumptions and priors on parameters	Objective function	# of model parameters	# of free parameters
Spike-triggered average[a]	Linear	L_2 loss (Gaussian)	Constant	All values are equally likely: flat prior $\mathbf{A}^{-1} = \mathbf{0}$	No local minima and smooth	$d+1$	0
Normalized reverse correlation[b]	Linear	L_2 loss (Gaussian)	Constant	All values are equally likely: flat prior $\mathbf{A}^{-1} = \mathbf{0}$	No local minima and smooth	$d+1$	0
Ridge regression[c]	Linear	L_2 loss (Gaussian)	$\sigma_{\vec{\beta}}^{-2} \|\boldsymbol{\beta}\|_2^2$	Independent with equal variance: Gaussian $\mathbf{A} = \sigma_{\vec{\theta}}^2 \mathbf{I}$	No local minima and smooth	$\boldsymbol{d+1}$	1
Linear model (with automatic relevancy determination)[d]	Linear	L_2 loss (Gaussian)	$\sum_{k=1}^{d} \sigma_{\vec{\beta}_k}^{-2} \beta_k^2$	Independent: Gaussian $\mathbf{A} = diag(\sigma_{\vec{\theta}_k}^2)$	No local minima and smooth	$d+1$	d
Linear model with smooth prior[e]	Linear	L_2 loss (Gaussian)	$\delta^{-1} \|\mathbb{D}\boldsymbol{\beta}\|_2^2$	Smooth over input space: Gaussian $\mathbf{A} = \mathbb{D}^{-2}$	No local minima and smooth	$d+1$	1
Wiener/Volterra kernel method[f]	2nd Order Wiener/Volterra	L_2 loss (Gaussian)	Constant	All values are equally likely: flat prior $\mathbf{A}^{-1} = \mathbf{0}$	No local minima and smooth	d^2+d+1	0
Spike-triggered covariance[g]	2nd Order Wiener/Volterra	L_2 loss (Gaussian)	Constant	All values are equally likely: flat prior $\mathbf{A}^{-1} = \mathbf{0}$	No local minima and smooth	d^2+d+1	0
Linearized reverse correlation[h]	Linearized model	L_2 loss (Gaussian)	$\sum_{k,d_k<\lambda} \infty(\mathbf{Q}_k^T \boldsymbol{\theta})^2$	In a stimulus PC subspace: Gaussian $\mathbf{A} = \mathbf{Q}\mathbf{D}_\lambda\mathbf{Q}^T$	No local minima and smooth	$kd+1$	2
Neural network (with automatic relevancy determination)[i]	Neural network	L_2 loss (Gaussian)	$\sum_{k=1}^{(d+2)b+1} \sigma_{\vec{\theta}_k}^{-2} \theta_k^2$	Independent: Gaussian $\mathbf{A} = diag(\sigma_{\vec{\theta}_k}^2)$	Has local minima, but smooth	$(d+2)b+1$	$d+3\&b$
Kernel ridge regression[j]	Kernel regression	L_2 loss (Gaussian)	$\sigma_{\vec{\beta}}^{-2} \|\boldsymbol{\beta}\|_2^2$	Independent with equal variance: Gaussian $\mathbf{A} = \sigma_{\vec{\theta}}^2 \mathbf{I}$	No local minima and smooth	$\leq N$	$1\&\Phi$
Support-vector regression[j]	Kernel regression	ε-insensitive loss[k]	$\sigma_{\vec{\beta}}^{-2} \|\boldsymbol{\beta}\|_2^2$	Independent with equal variance: Gaussian $\mathbf{A} = \sigma_{\vec{\theta}}^2 \mathbf{I}$	No local minima, but not smooth	$<N$	$2\&\Phi$

*Maximum a posteriori (MAP) constituents for system identification algorithms commonly used in neurophysiology. All these algorithms assume a Gaussian prior, so only the covariance matrix of the relevant Gaussian prior is shown. All but one algorithm (support-vector regression) use the L_2 loss; the others assume a Gaussian noise distribution. Maximally informative dimension analysis is one algorithm that does not fit under the MAP estimation framework (Sharpee et al. 2004).

[a](Eggermont et al. 1983, Marmarelis & Marmarelis 1978).
[b](Theunissen et al. 2001).
[c](Machens et al. 2004).
[d](Sahani & Linden 2003).
[e](Sahani & Linden 2003, Smyth et al. 2003).
[f](Aertsen & Johannesma 1981b, Eggermont 1993, Emerson et al. 1987, Victor & Shapley 1980).
[g](Brenner et al. 2000, Rust et al. 2005, Touryan et al. 2005).
[h](Aertsen & Johannesma 1980b, David et al. 2004, Kowalski et al. 1996, Machens et al. 2004, Mazer et al. 2002, Ringach et al. 1997a, Theunissen et al. 2000).
[i](Lau et al. 2002, Lehky et al. 1992, Prenger et al. 2004).
[j](Wu & Gallant 2004)
[k](Vapnik 1995).

LGG: linear model class, Gaussian noise distribution, and Gaussian prior

To simplify computation, most SI studies in neurophysiology use algorithms that have a smooth MAP objective function with no local minima. The simplest of these consists of a linear model class, a Gaussian noise distribution, and a Gaussian prior (the LGG family, shaded entries in **Table 4**). The MAP objective of any algorithm in the LGG family can be minimized analytically,

$$\boldsymbol{\theta}^{*} = \underbrace{\left(\mathbf{X}^{T}\mathbf{X} + \lambda'\mathbf{A}^{-1}\right)^{-1}}_{\text{regularized inverse of the stimulus (feature) autocovariance}} \mathbf{X}^{T}\mathbf{Y}. \qquad 2.$$

Here $\lambda' = \sigma^2$ is the noise variance (**Table 2**), and $\mathbf{A}$ is the covariance matrix of the Gaussian prior. The matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T$ gives the concatenated stimulus samples, and the vector $\mathbf{Y} = [y_1, \ldots, y_N]^T$ represents the response samples. The object above the brace is the regularized inverse of the stimulus autocovariance matrix. When the stimulus is white noise, there are no correlations between stimulus channels and the stimulus autocovariance, $\mathbf{X}^T\mathbf{X}$, is trivially the identity matrix (Victor 2005). Equation 2 can be extended to nonlinear models that are linear in the model parameters [i.e., generalized linear models (Dobson 2002)]. Generalized linear models include all the model classes in **Table 1** except ANNs.

3.2. Number of Model Parameters

The difficulty of computing the MAP estimate also depends on the number of model parameters, which in turn depends on the choice of model class and the dimensionality of the stimulus (**Table 4**). With few exceptions (David & Gallant 2005), nonlinear models usually have more parameters than linear models and therefore require more data to fit.

When the stimulus is not white (e.g., natural stimuli) and Equation 2 is used to compute the MAP estimate, then additional data are required to correct for stimulus bias. This correction requires computing the regularized inverse, which is difficult when the model has many parameters (e.g., Wiener/Volterra models). The data limitation inherent in natural stimulus experiments can be mitigated in several ways. The MAP estimate can be computed without invoking Equation 2 by recursive least squares methods (Lesica et al. 2003, Ringach et al. 2002, Stanley 2002) or by gradient-based methods (Prenger et al. 2004). Alternatively, linearized models that require relatively few model parameters can be used (**Table 4**).

A model with few parameters makes efficient use of available data and facilitates the interpretation of the estimated mapping function (Section 3.3). Several procedures can be used to reduce the number of model parameters. If the stimulus dimensions are separable, a simpler model can be fit to each dimension independently (David et al. 2004). Regularization methods such as ARD (Mackay 1995, Prenger et al. 2004, Sahani & Linden 2003a) can be used to remove irrelevant model parameters (Section 2.3). Finally, alternative fit algorithms such as boosting can be used to fit sparse models with low model complexity (Buhlmann & Yu 2003, Willmore et al. 2005).

3.3. Number of Free Parameters

Many SI algorithms contain free parameters that are required to specify the MAP objective function before it can be minimized. Free parameters may be required for any of the three MAP constituents: the model class (e.g., the number of hidden units in an ANN); the noise distribution (e.g., the Gaussian noise variance); or most commonly, the prior (regularization parameters). Most priors require at least one regularization parameter to determine the trade-off between the loss function and the regularizer (Section 2.3). When the total number of free parameters is small, cross-validation is an objective method to find their optimal values (David et al. 2004, Woolley et al. 2005, Wu & Gallant 2004). When there are too many free parameters for

cross-validation, an iterative algorithm such as evidence optimization can be used (Mackay 1995, Prenger et al. 2004, Sahani & Linden 2003a).

4. EXPERIMENTAL CONSIDERATIONS FOR SYSTEM IDENTIFICATION

Several technical aspects of neurophysiological experiments place hard limits on the practical implementation of SI algorithms. In most cases recordings can only be made from a single sensory neuron in vivo for several hours. Neuronal responses are also quite noisy owing to the stochastic nature of spike generation (Johnson 1980) and uncontrolled effects of anesthesia or changes in cognitive state (Cook & Maunsell 2004, David et al. 2002, Fritz et al. 2003). These problems are exacerbated when studying sensory neurons at central stages of processing. Estimating the stimulus-response mapping function successfully requires that the experiment be optimized practically as well as theoretically. This section reviews several of the most important practical issues.

4.1. Stimulus Selection

The major stimulus classes used in neuronal system identification are summarized in **Table 5**. The simplest stimulus for SI is Gaussian white noise. In white noise, each stimulus channel varies randomly and independently, making it straightforward to estimate first- (i.e., linear) and second-order Wiener/Volterra models. White noise has been used to estimate linear mapping functions in the auditory (De Boer 1967, Eggermont 1993), visual (Chichilnisky 2001, Horwitz et al. 2005), and somatosensory (Arabzadeh et al. 2005) systems. However, white noise tends to be an inefficient stimulus for more central sensory neurons. Central neurons represent the higher-order features

Table 5 Major classes of stimuli used for neuronal system identification

Stimulus class	Advantages	Disadvantages	Examples
White noise	Computationally convenient, spans appropriate subspace	Huge space, inefficient	White-noise sound pressure waveform (De Boer 1967, Eggermont 1993) Spatiotemporal white noise (Chichilnisky 2001) High-contrast noise, m-sequences (Cottaris & De Valois 1998, DiCarlo et al. 1998; Sutter 1987) Sparse noise (DeAngelis et al. 1993, Jones & Palmer 1987)
Parametric noise	Computationally convenient, effective stimulus	Restricted subspace	Random tones (Aertsen & Johannesma 1981b, deCharms 1998) Ripples, TORCs[a] (Klein et al. 2000, Kowalski et al. 1996, Miller et al. 2001) Grating sequences (Mazer et al. 2002, Ringach et al. 1997b) Sum of sinusoids (Victor & Shapley 1980) Bars (Lau et al. 2002, Rust et al. 2005, Touryan et al. 2002) Random dot noise (Borghuis et al. 2003, Cook & Maunsell 2004)
Natural stimuli	Effective stimulus, spans appropriate subspace	Complex statistics	Animal vocalizations (Aertsen & Johannesma 1981a, Theunissen et al. 2000) Natural image sequences (David et al. 2004, Smyth et al. 2003, Touryan et al. 2005) Natural vision movies (David et al. 2004, Ringach et al. 2002)

[a] Temporally orthogonal ripple combination.

of natural stimuli (Dan et al. 1996, Woolley et al. 2005), and these structured patterns occur rarely in white noise. For this reason, most of the studies that have used white noise to estimate second-order Wiener/Volterra models have been conducted in simple organisms or in peripheral sensory systems (Brenner et al. 2000, Pece et al. 1990). Given data limitations, one practical way to use white noise to estimate nonlinear mapping functions in more complex systems is to modulate the noise along only a few dimensions, such as time (Mancini et al. 1990, Sakai & Naka 1992) or one dimension of space (Citron & Emerson 1983, Rust et al. 2005, Touryan et al. 2002).

Two variants of white noise have been used to increase estimation efficiency. In high-contrast (typically binary) noise, each input channel is randomly assigned one of several discrete power levels, thus increasing the effective contrast (Alonso et al. 2001, DiCarlo et al. 1998, Nishimoto et al. 2006). High-contrast noise is often used in conjunction with the m-sequence method to maximize the efficiency of stimulus sampling (Cottaris & De Valois 1998, Reid et al. 1997, Sutter 1987). In sparse noise, only one (or at most a few) random stimulus channel(s) is active in each sample. This increases local contrast in the stimulus and generally produces more robust responses (DeAngelis et al. 1993, Jones & Palmer 1987).

As alternatives to white noise, two classes of stimuli have been employed for mapping function estimation. Parametric noise is not noise in the classical sense. Rather, it is a random series of structured patterns that span a stimulus subspace known to be relevant for the sensory system under study. The parameters that define these stimuli can be a linear function of the stimulus (Yu & de Sa 2004) or a nonlinear function such as spectral power (Aertsen & Johannesma 1981b) or orientation (Ringach et al. 1997b). In the latter case a linearized model can be used to estimate the mapping function by fitting a linear model in the parameter space.

In the auditory system, typical parametric noise sources are random tone sequences (Aertsen & Johannesma 1981b, deCharms et al. 1998), narrowband noise bursts (Nelken et al. 1997), and harmonic stacks (i.e., ripples) (Klein et al. 2000, Kowalski et al. 1996, Theunissen et al. 2001). Because many neurons in primary visual cortex are sensitive to orientation and spatial frequency, parametric noise for the visual system often consists of oriented features such as bars (Lau et al. 2002, Rust et al. 2005, Touryan et al. 2002), sinusoidal gratings (Bredfeldt & Ringach 2002, Mazer et al. 2002, Ringach et al. 1997b), or a sum of sinusoids (Victor & Shapley 1980). Parametric moving dot noise has also been used in motion-sensitive area MT (Borghuis et al. 2003, Cook & Maunsell 2004).

Parametric noise presents features known to drive more peripheral neurons, but it is not necessarily an effective stimulus for more central neurons tuned to higher-order combinations of the parameterized features. Indeed, simple parametric stimuli often fail to elicit robust responses in more central sensory areas (Gallant et al. 1996, Grace et al. 2003), and parametric noise has not been particularly useful for characterizing neurons in these areas.

A second alternative to white noise is to use natural stimuli for mapping function estimation. Over the course of evolution, sensory systems should have been optimized for processing natural stimuli (Barlow 1961, Field 1987, Simoncelli & Olshausen 2001), and the space of potential natural stimuli is many orders of magnitude smaller than the entire white-noise space (Field 1994). Natural stimuli sampled at random can be used just like any other noise signal to estimate the mapping function of sensory neurons. However, natural stimuli contain strong correlations (Field 1987, Simoncelli & Olshausen 2001), and these must be taken into account during estimation to avoid systematic bias (Aertsen & Johannesma 1981a, David et al. 2004, Sharpee et al. 2004, Smyth et al. 2003, Theunissen

et al. 2001). Natural stimuli have been used to estimate the stimulus-response mapping function in auditory (Aertsen & Johannesma 1981a, Machens et al. 2004, Theunissen et al. 2001), visual (David et al. 2004, Felsen et al. 2005, Ringach et al. 2002, Smyth et al. 2003, Touryan et al. 2005), and somatosensory systems (Arabzadeh et al. 2005).

Although the space of natural stimuli is much smaller than the space of white noise, it is still far too large to be sampled completely in any real neurophysiology experiment. To enable denser sampling many experiments focus on restricted subsets of natural stimuli: species-specific vocalizations (Aertsen & Johannesma 1981a, Theunissen et al. 2000), stimuli with natural spatial structure but white temporal structure (David et al. 2004, Smyth et al. 2003), or vice versa (Yu et al. 2005). However, mapping functions estimated using these subsets do not necessarily generalize to the entire space of natural stimuli (David et al. 2004).

4.2. Behavioral Control

There is growing evidence that the stimulus-response mapping function of sensory neurons can be modified by top-down effects such as attention (David et al. 2002, Fritz et al. 2003) and learning (Polley et al. 2004, Yang & Maunsell 2004). These effects can be ignored in experiments involving anesthetized, untrained animals but must be considered in studies involving awake and behaving animals. One simple way to minimize uncontrolled variability in awake animals is to use a consistent behavioral task such as passive fixation. With the assumption that this task induces a consistent cognitive state, the mapping function can then be estimated directly without modeling top-down effects (David et al. 2004). Alternatively, these top-down effects can be investigated by comparing estimates of mapping functions obtained under different cognitive states (Section 5.5) (Cook & Maunsell 2004, David et al. 2002, Fritz et al. 2003).

4.3. Visualization and Interpretation

Mapping functions estimated using parametric models can usually be interpreted relatively easily. Linear and linearized models can be visualized by simply plotting the values of the model parameters (Jones & Palmer 1987, Mazer et al. 2002). Second-order Wiener/Volterra models can usually be visualized by plotting the relevant stimulus subspace (Brenner et al. 2000, Rust et al. 2005, Touryan et al. 2002). Nonparametric models cannot be visualized easily and are notoriously difficult to interpret. For this reason they have been rarely used for SI in neurophysiology (Lau et al. 2002, Lehky et al. 1992, Prenger et al. 2004, Wu & Gallant 2004). One effective method for visualizing such models is to identify the stimulus dimensions that most influence responses. For example, the parameters of an ANN define highly nonlinear interactions, but the subspace of stimuli that modulates responses may be relatively small. This subspace can be extracted from the model parameters by dimensionality reduction (e.g., singular value decomposition), and the relevant stimulus subspace can be visualized directly (**Figure 3**) (Lau et al. 2002, Prenger et al. 2004).

4.4. Prediction and Validation

In SI, a good estimate of the stimulus-response mapping function is not merely one that is statistically significant (i.e., unlikely to have occurred by chance). Rather, a good estimate is one that accurately predicts future responses, especially responses to stimuli that were not used for estimation (David & Gallant 2005, Machens et al. 2004, Nelken et al. 1997, Theunissen et al. 2000). A model that accurately describes the entire stimulus-response mapping function will predict responses to any stimulus; a bad model will not generalize and will predict poorly.

Mapping function estimates can be contaminated by bias stemming from stimulus statistics, experimental noise, and

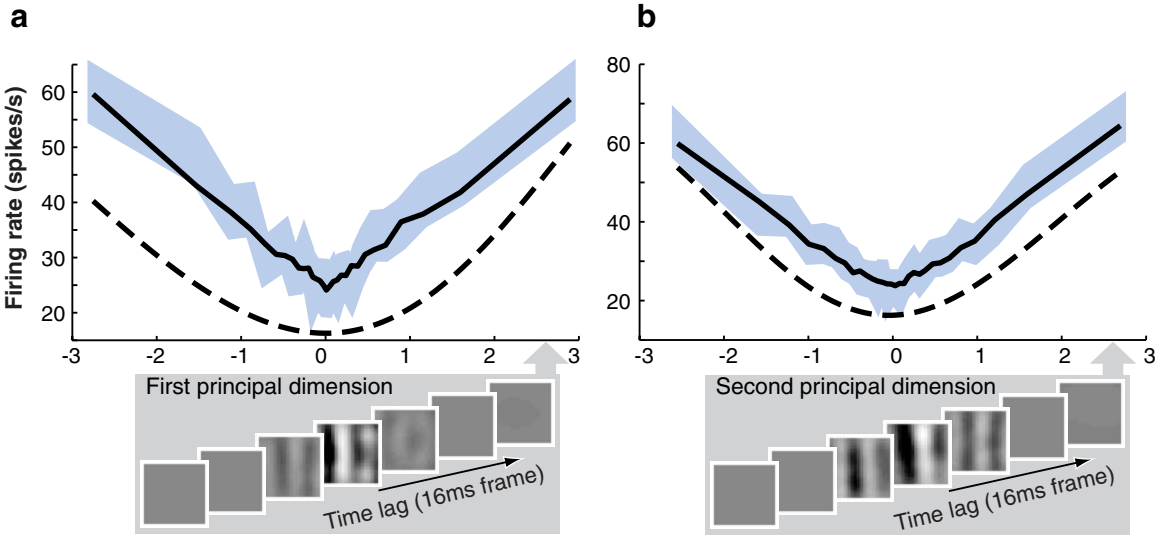

Figure 3

Effective visualization is important for interpreting nonparametric mapping functions. Nonparametric models have many parameters and can be difficult to interpret. Data shown here were acquired from a single V1 complex cell during stimulation with a sequence of natural images. The mapping function was estimated with a nonparametric artificial neural network. To visualize the network, the principle dimensions of the relevant input space were identified by singular value decomposition of the input weights (**U** in **Table 1**). (*a*) The x-axis shows the projection of the stimulus onto the first principle dimension, and the y-axis gives the corresponding response rate. The spatiotemporal filter shown beneath the x-axis illustrates the spatiotemporal pattern of the first principle dimension. This dimension represents tuning to vertical orientation, spatial frequency of about two cycles per receptive field, even spatial phase, and latency of 32 ms. The gray-shaded curve represents the mean (*center of curve*) and two standard errors (*boundaries of curve*) of the response evoked by stimuli differing along the principle dimension. The solid line shows the response predicted by the full neural network, and the dashed line shows the responses predicted by this principle dimension alone. The fact that the solid and dotted lines have the same shape indicates this dimension contributes significantly to the response. Because this neuron responds to both positive and negative projections onto the principle dimension, it is a complex cell. (*b*) The second principle dimension of the same network. This dimension is similar to the first except it has an odd spatial phase. Figure adapted from Prenger et al. (2004), by permission of the authors.

regularization procedures. This is a particular danger when comparing mapping functions estimated using different stimuli, different data sets, or different regularization schemes (David et al. 2004, Sharpee et al. 2004). Predictive power provides an objective method for evaluating the quality of an estimate and thus can help identify potential sources of bias. Prediction accuracy can also be used to compare competing models: If several models are used to predict responses to a single validation data set, the model that produces the best prediction is the better model (David & Gallant 2005). When two models yield virtually identical predictive power, the simpler model is usually favored. It may also be useful to compare the models directly to determine whether they describe different aspects of the data, or whether one of the models encompasses the other. As long as the validation data set is independent and distinct from the estimation set, then prediction can always be used to compare different models objectively.

The most common measure of prediction accuracy is the correlation between predicted and observed responses (Pearson's r). The correlation coefficient provides an intuitive measure of predictive power: The squared correlation coefficient gives the fraction of response variance explained by the model (David & Gallant 2005, Machens et al. 2004, Sahani & Linden 2003b). Other measures of predictive power that may be more appropriate for neuronal spiking data have been proposed [e.g., coherence (Hsu et al. 2004), mutual information (Sharpee et al. 2004)]. However, these measures have not been shown to be superior to correlation in practice. In any case, no single metric has been shown to be appropriate for all circumstances.

Because neurons are stochastic, it will never be possible to predict all of the variance in neuronal responses. The potentially explainable variance is the fraction of total response variance that could theoretically be predicted, given neuronal noise and a finite sample (David & Gallant 2005, Sahani & Linden 2003b). A perfect model will not explain all of the response variance, but it will explain all of the potentially explainable variance. Hence the best model is the one that predicts the largest fraction of the potentially explainable response variance.

5. APPLICATIONS

The SI approach has undergone continual development and refinement over the last 40 years. Early work focused on simple algorithms that are easy to compute. Most of these studies used white noise and a linear model to investigate peripheral sensory systems. Subsequent algorithmic advances and increases in computer power have dramatically expanded the range of problems that can be addressed. Modern SI studies select from a wide range of parametric and nonparametric models, use complex stimuli, and are conducted in many different sensory systems.

5.1. Tuning Properties

The most common application of SI in neurophysiology is to efficiently characterize spatial and temporal receptive fields. In more peripheral sensory systems, tuning can often be measured with a linear model (Chichilnisky 2001, DeAngelis et al. 1993, De Boer 1967, DiCarlo et al. 1998, Reid et al. 1997). If nonlinear response properties are known, an appropriate linearizing transformation can be used to transform the stimulus into a space where a linear model can be fit. Linearized models have been used to characterize nonphase locked neurons in A1 (Kowalski et al. 1996, Machens et al. 2004, Miller et al. 2001), complex cells in V1 (Bredfeldt & Ringach 2002; Mazer et al. 2002; Ringach et al. 1997a, 1997b), and velocity selective cells in MT (Cook & Maunsell 2004, Pack et al. 2003, Perge et al. 2005). Mapping functions that describe basic tuning properties can be interpreted easily and generally agree with tuning curves measured using more traditional approaches (DeAngelis et al. 1993, Jones & Palmer 1987, Kowalski et al. 1996, Nishimoto et al. 2005, Pack et al. 2003).

One common application of SI is to determine whether the tuning of stimulus dimensions is separable (i.e., independent) or whether there is an interaction between dimensions. Separability is often taken as a sign that the two dimensions are processed in distinct neuronal populations (Kowalski et al. 1996). Many peripheral sensory neurons are space-time separable (Jenison et al. 2001), whereas many central neurons are inseparable (DeAngelis et al. 1993, DiCarlo et al. 1998). Neurons in auditory cortex have inseparable spectral and temporal tuning, a property that may facilitate detection of up- and downsweeps (Depireux et al. 2001). In primary visual cortex, orientation and temporal tuning are predominantly separable (Mazer et al. 2002). In contrast, spatial frequency and temporal tuning are inseparable (Bredfeldt & Ringach 2002, Frazor et al. 2004, Mazer et al. 2002). Inseparable spatial frequency and

LGN: lateral geniculate nucleus

temporal tuning may reflect the integration of faster low spatial frequency magno- and slower high spatial frequency parvo-cellular inputs from the lateral geniculate nucleus (LGN) (Frazor et al. 2004, Mazer et al. 2002).

Several studies have examined the mapping functions of a large sample of neurons to make inferences about population coding. Analysis of mapping functions in LGN suggests the population of retinal ganglion cells is optimized to convey information about natural scenes (Dan et al. 1996). Similarly, neuronal tuning in the auditory thalamus of the songbird appears to be optimized for representing song (Woolley et al. 2005). Mapping functions estimated from simultaneously recorded pairs of neurons can be broken down into their simultaneous and asynchronous spiking components. In LGN, mapping functions estimated from simultaneous responses of neuron pairs tend to be more narrowly tuned than those estimated from individual neurons (Dan et al. 1998), suggesting simultaneous activity of adjacent neurons is narrowly selective for distinct stimulus features. Similar results have been reported in auditory cortex (Tomita & Eggermont 2005).

5.2. Nonlinear Response Properties

Sensory neurons have many nonlinear responses properties difficult to characterize using classical approaches. The SI approach provides a general procedure for fitting parametric and nonparametric nonlinear models that do not require an explicit theory about neural coding in a particular sensory area. The estimated mapping functions can then be used to develop theories that can be refined in classical experiments.

Second-order Wiener/Volterra models (Citron & Emerson 1983, Eggermont 1993, Mancini et al. 1990) have been extremely useful for characterizing neurons in many sensory systems. In primary visual cortex, a second-order model describes phase invariance (Gaska et al. 1994, Touryan et al. 2002), direction selectivity (Citron & Emerson 1983, Rust et al. 2005), and velocity selectivity (Emerson et al. 1987). These findings agree broadly with the predictions of the motion-energy model (Emerson et al. 1992, Gaska et al. 1994). Second-order models also account for the spatially tuned suppression (Citron & Emerson 1983, David et al. 2004, Rust et al. 2005, Touryan et al. 2005) and short-term adaptation (Mancini et al. 1990) of visual cortical neurons. In the auditory system, second-order models have been used to characterize frequency-specific facilitatory and inhibitory interactions in auditory nerve fibers (Young & Calhoun 2005). A second-order nonlinearity also accounts for the envelope responses of many cells in peripheral (Yamada & Lewis 1999) and central (Keller & Takahashi 2000, Kowalski et al. 1996, Sen et al. 2001) pathways.

It is difficult to characterize nonlinear responses beyond second order using Wiener/Volterra models (Sections 2.1 and 3.2). Therefore, more flexible nonparametric models have been used to investigate effects such as nonlinear spatial pooling (Prenger et al. 2005, Wu & Gallant 2004). Nonparametric models may be particularly valuable for investigating sensory neurons at more central stages of sensory processing.

5.3. Functional Anatomy

Sensory systems consist of several anatomically distinct stages of processing (Felleman & Van Essen 1991, Read et al. 2002). By mere virtue of their connectivity, neurons at each processing stage are likely to have different response properties and represent different stimulus attributes. The mapping function provides a common currency for comparing neuronal coding properties across sensory areas.

A few studies have estimated mapping functions for monosynaptically connected pairs of neurons in different anatomical areas. Comparing these mapping functions offers insight into the circuitry that transforms sensory representations between areas. One study

that estimated linear mapping functions for connected pairs of neurons in LGN and primary visual cortex neurons reported that thalamo-cortical connections are extremely specific (Alonso et al. 2001). The ON and OFF subregions of receptive field maps tend to overlap closely, suggesting feedforward connections play a dominant role in determining tuning in primary visual cortex. Mapping functions have also been estimated for connected pairs of neurons in auditory thalamus and cortex (Miller et al. 2001). Neurons in both areas tend to share the same characteristic frequency but differ in their temporal and spectral modulation properties. This suggests neurons in auditory cortex derive some of their basic tuning properties from intracortical rather than feedforward connections.

Several other studies have compared mapping functions of populations of neurons recorded in different auditory areas: A1 and the lateral belt (Barbour & Wang 2003), A1 and anterior auditory field (Linden et al. 2003), and several auditory nuclei in the songbird (Sen et al. 2001). Mapping functions differ across areas in many respects: latency (Linden et al. 2003, Sen et al. 2001), bandwidth (Linden et al. 2003), preferred modulation frequency (Sen et al. 2001), strength of inhibition (Sen et al. 2001), and degree of linearity (Linden et al. 2003, Sen et al. 2001). The application of SI throughout the auditory system has revealed the functional hierarchy across sensory areas that were originally defined only anatomically.

5.4. Natural Stimuli

The response properties of many sensory neurons are modulated by stimuli that do not evoke responses in isolation (Albright & Stoner 2002, Allman et al. 1985, Nelken 2004). These contextual effects have been described variously as contrast gain control (Albrecht et al. 1984, Bonds 1991, Ohzawa et al. 1985), nonclassical receptive field modulation (Gilbert & Wiesel 1990, Knierim & Van Essen 1992, Vinje & Gallant 2002), and side-band modulation (Bendor & Wang 2005, Bonds 1991). Natural stimuli contain rich combinations of features that may drive these context-dependent responses in a way that cannot be predicted with simple synthetic stimuli (David et al. 2004).

Several studies have addressed this issue by comparing mapping functions estimated for sensory neurons using parametric noise and natural stimuli [e.g., auditory system (Theunissen et al. 2000, Woolley et al. 2005), visual system (David et al. 2004, Yu et al. 2005)]. In all cases mapping functions estimated using the different stimulus classes are broadly similar, but they differ in several important and consistent respects. Broadband natural stimuli tend to elicit substantially more inhibition than narrowband parametric noise, and this inhibition tends to make tuning curves narrower in space (David et al. 2004) and time (David et al. 2004, Woolley et al. 2005). This modulation appears to reflect a dynamic process in which sensory neurons adapt to become more selective within the prevailing stimulus statistics. Moreover, V1 neurons will adapt to the statistics of natural stimuli to increase information transmission (Sharpee et al. 2006).

5.5. Cognitive Factors

Sensory neurons in more central areas are modulated by cognitive processes such as attention, learning, and memory (Desimone & Duncan 1995, Reynolds & Chelazzi 2004). These processes change the way sensory information is routed and pooled at successive stages of representation (Olshausen et al. 1993). SI offers a means for studying how these cognitive processes interact with sensory representation. To investigate cognitive factors from the SI perspective, it is convenient to view them as additional input channels (David et al. 2002). Because it is impossible to gather data under every possible cognitive state, a single study must instead estimate separate mapping functions under just a few specific states (Cook & Maunsell 2004,

David et al. 2002, Fritz et al. 2003). The mapping function acquired in each state represents a single stimulus-response slice through the complete stimulus-cognition-response mapping function.

SI studies of attention have focused on the separability of attention and sensory tuning. If these are separable, then attention merely modulates the response rate (Luck et al. 1997) or gain (McAdams & Maunsell 1999, Reynolds & Chelazzi 2004) of a neuron. If they are inseparable, then attention can actually change the structure of the stimulus-response mapping function. The data available thus far present a conflicting picture of this interaction. One study in area MT found that spatial attention only modulates response gain and thus is separable from sensory tuning (Cook & Maunsell 2004). In contrast, studies in areas V4 (David et al. 2002) and A1 (Fritz et al. 2003) reported that attention can also modulate tuning properties. This apparent conflict may reflect differences in experimental control of attention, the anatomical areas investigated, or the models used to characterize the mapping function. More exhaustive SI studies are required to clarify this issue.

5.6. Optimal Stimuli

Estimating the entire stimulus-response mapping function of a sensory neuron is an ambitious goal. Other studies have taken a simpler approach, seeking merely to identify the optimal stimulus for sensory neurons. The most straightforward procedure for finding an optimal stimulus is to search through the stimulus space and identify the pattern that elicits the largest response from a neuron (Foldiak 2001, Kobatake & Tanaka 1994, Machens et al. 2005, Tanaka et al. 1991). An exhaustive search works well in simple sensory systems where the relevant stimulus space is small (Machens et al. 2005), but it cannot be applied to more complex systems where the space of potential stimuli is huge. In such cases heuristic search methods can identify patterns that elicit strong neuronal responses (Kobatake & Tanaka 1994). However, it is difficult to conclude that a particular stimulus is truly optimal when the tested stimulus set is very small.

One factor complicating the optimal stimulus approach is that there is usually more than one optimal stimulus for a neuron. In cortex, sensory neurons are tuned to some features of a stimulus but invariant to others. For example, complex cells in visual cortex are insensitive to the spatial phase of their optimal oriented feature (De Valois et al. 1982, Movshon et al. 1978, Touryan et al. 2002), and envelope-sensitive cells in the central auditory pathway are insensitive to the phase of the carrier (Eggermont 1993). Invariant tuning becomes even more common at more central stages of visual processing [e.g., inferior temporal cortex (Ito et al. 1995)]. Because invariant neurons respond equally well to all stimuli along the invariant dimensions, there is no single optimal stimulus.

The mapping function obtained by SI can describe responses along all stimulus dimensions, including the invariant dimensions. By inverting the mapping function, all stimuli that elicit any specific level of response can be identified. In fact, one good way to validate an estimated mapping function is to use it to generate optimal stimuli and to determine whether those stimuli actually elicit the predicted response (deCharms et al. 1998, Touryan et al. 2002).

5.7. Predictive Power

The SI approach can be used to compare and evaluate models in terms of their predictive power. Many SI studies have assessed predictions between two independent sets of simple synthetic stimuli, one used for estimation and one for validation. When these estimation and validation stimuli have similar statistical properties, predictive power is often quite high, both for neurons near the sensory periphery (Arabzadeh et al. 2005, Citron et al. 1988, Golomb et al. 1994, Nelken et al. 1997) and in more central structures (DiCarlo et al. 1998, Lau et al. 2002). This indicates that

the mapping functions estimated using simple stimuli provide a good description of neuronal responses to those specific stimuli.

Simple synthetic stimuli are experimentally convenient, but the ultimate test of any model of sensory processing is how well it predicts neuronal responses to natural stimuli during natural sensory exploration (Carandini et al. 2005, David & Gallant 2005). Several studies have measured how well various nonlinear models can account for responses to natural stimuli (David & Gallant 2005, David et al. 2004, Grace et al. 2003, Keller & Takahashi 2000, Machens et al. 2004, Prenger et al. 2004, Touryan et al. 2005). These predictions are uniformly lower than those obtained for simple stimuli, reflecting both the complexity of natural signals and the fact that such signals often evoke nonlinear responses not observed with simpler stimuli (David et al. 2004, Theunissen et al. 2000, Woolley et al. 2005). No SI studies have assessed predictions during natural sensory exploration; this is a likely avenue for future research.

6. FUTURE DIRECTIONS

Applications of SI to sensory systems have exploded in the past decade, reflecting the successful efforts of early pioneers (Aertsen & Johannesma 1981b, Jones & Palmer 1987, Lehky et al. 1992, Marmarelis & McCann 1973). Current theoretical work is focused on several areas. Several groups are actively developing new SI algorithms (Rust et al. 2005, Sharpee et al. 2004, Theunissen et al. 2001, Tomita & Eggermont 2005). One avenue is to explore novel combinations of MAP constituents that have not previously been used. Another is to develop new nonparametric methods such as support-vector regression (Wu & Gallant 2004) or maximally informative dimensions (Sharpee et al. 2004). Nonparametric models require specialized tools for interpreting and visualizing estimated mapping functions (Prenger et al. 2004). Visualization is especially important when the sensory neuron is highly nonlinear and when the system is probed with spectrally complex or natural stimuli, so this is likely be a critical aspect of future work.

Long-term changes in neuronal response properties over timescales ranging from many seconds [e.g., adaptation (Bonds 1991, Ohzawa et al. 1985)] to many weeks [e.g., learning (Li et al. 2004, Polley et al. 2004, Yang & Maunsell 2004)] are common throughout all sensory systems. These nonstationary processes complicate the estimation of mapping functions. One clear avenue for future work is to develop SI algorithms that capture these nonstationary processes. One recent series of studies used a recursive least squares algorithm to model slow contrast adaptation in the mammalian visual system (Lesica et al. 2003). Adaptive algorithms such as this may also be useful for describing effects such as attention and learning.

Neurons at central stages of sensory processing have two important properties that cannot be modeled using current methods. First, these cells are likely involved in figure-ground segmentation (Albright & Stoner 2002). Scene segmentation is a difficult problem for which no theoretical solution exists, and little is known about how the brain accomplishes this task. Second, neurons in central areas are invariant to many stimulus dimensions [e.g., size, position (Gallant et al. 1996, Ito et al. 1995)]. There is currently no standard theoretical framework for representing invariances, and no systematic approach is available for identifying invariant dimensions in neurophysiological experiments. Of course, scene segmentation and invariant responses are also difficult to study using conventional neurophysiological methods.

7. CONCLUSIONS

SI is an objective and powerful approach for describing how sensory neurons encode information about stimuli. The approach is more general than classical hypothesis testing because it provides information about both the

statistical significance and the importance of a model. SI experiments show that current models provide a good description of the functional properties of peripheral sensory neurons. As one might expect, current models of central sensory neurons are less accurate. Because the SI approach reveals both the strengths and weaknesses of current models, it provides important information about the best directions for future research.

The SI approach is not limited to sensory systems. Many current studies of representation in motor cortex attempt to estimate a mapping function that describes how signals from a population of neurons are transformed into a motor response (Schwartz 2004). Likewise, many functional neuroimaging studies aim to describe the functional relationship between sensory stimuli and cortical activity (Hansen et al. 2004, Victor 2005). These are simple variants of the SI approach reviewed here. As further research clarifies the links between sensory processing, motor responses, cognitive states, and functional neuroimaging, the SI approach may also provide a quantitative framework for integrating findings across many stages of cortical processing.

ACKNOWLEDGMENTS

Preparation of this article was supported by grants from the NEI and the NIMH (J.L.G), a DOE Computational Science Graduate Fellowship (M.C.-K.W.), and the NSF (S.V.D.). We thank Jonathan Victor, Bin Yu, Frederic Theunissen, Ryan Prenger, Thomas Naselaris, Shinji Nishimoto, and Ben Willmore for technical advice, useful discussion, and comments on the manuscript.

LITERATURE CITED

Aertsen AMHJ, Johannesma PIM. 1981a. A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol. Cybern.* 42:145–56

Aertsen AMHJ, Johannesma PIM. 1981b. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol. Cybern.* 42:133–43

Albrecht DG, Farrar SB, Hamilton DB. 1984. Spatial contrast adaptation characteristics of neurons recorded in the cat's visual-cortex. *J. Physiol.* 347:713–39

Albright TD, Stoner GR. 2002. Contextual influences on visual processing. *Annu. Rev. Neurosci.* 25:339–79

Allman J, Miezin F, McGuinness E. 1985. Stimulus specific responses from beyond the classical receptive-field: neurophysiological mechanisms for local global comparisons in visual neurons. *Annu. Rev. Neurosci.* 8:407–30

Alonso JM, Usrey WM, Reid RC. 2001. Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J. Neurosci.* 21:4002–15

Anzai A, Ohzawa I, Freeman RD. 2001. Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. *Nat. Neurosci.* 4:513–18

Arabzadeh E, Zorzin E, Diamond ME. 2005. Neuronal encoding of texture in the whisker sensory pathway. *PLoS Biol.* 3:155–65

Barbour DL, Wang X. 2003. Auditory cortical responses elicited in awake primates by random spectrum stimuli. *J. Neurosci.* 23:7194–206

Barlow HB. 1961. The coding of sensory messages. In *Current Problems in Animal Behavior*, ed. WH Thorpe, OL Zangwill, pp. 331–60. Cambridge, UK: Cambridge Univ. Press

Bendor D, Wang XQ. 2005. The neuronal representation of pitch in primate auditory cortex. *Nature* 436:1161–65

Bonds AB. 1991. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Vis. Neurosci.* 6:239–55

Borghuis BG, Perge JA, Vajda I, van Wezel RJA, van de Grind WA, Lankheet MJM. 2003. The motion reverse correlation (MRC) method: a linear systems approach in the motion domain. *J. Neurosci. Methods* 123:153–66

Bredfeldt CE, Ringach DL. 2002. Dynamics of spatial frequency tuning in macaque V1. *J. Neurosci.* 22:1976–84

Brenner N, Bialek W, de Ruyter van Steveninck RR. 2000. Adaptive rescaling maximizes information transmission. *Neuron* 26:695–702

Bringuier V, Chavane F, Glaeser L, Fregnac Y. 1999. Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283:695–99

Buhlmann P, Yu B. 2003. Boosting with the L2 loss: regression and classification. *J. Am. Statist. Assoc.* 98:324–39

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, et al. 2005. Do we know what the early visual system does? *J. Neurosci.* 25:10577–97

Carandini M, Heeger DJ, Movshon JA. 1997. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17:8621–44

Chichilnisky EJ. 2001. A simple white noise analysis of neuronal light responses. *Network* 12:199–213

Citron MC, Emerson RC. 1983. White noise analysis of cortical directional selectivity in cat. *Brain Res.* 279:271–77

Citron MC, Emerson RC, Levick WR. 1988. Nonlinear measurement and classification of receptive fields in cat retinal ganglion cells. *Ann. Biomed. Eng.* 16:65–77

Cook EP, Maunsell JHR. 2004. Attentional modulation of motion integration of individual neurons in the middle temporal visual area. *J. Neurosci.* 24:7964–77

Cottaris NP, De Valois RL. 1998. Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature* 395:896–900

Dan Y, Alonso JM, Usrey WM, Reid RC. 1998. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat. Neurosci.* 1:501–7

Dan Y, Atick JJ, Reid RC. 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* 16:3351–62

Daugman JG. 1988. Complete discrete 2-D gabor transforms by neural networks for image-analysis and compression. *IEEE Trans. Acoust.* 36:1169–79

David SV, Gallant JL. 2005. Predicting neuronal responses during natural vision. *Network.* 16:239–60

David SV, Mazer JA, Gallant JL. 2002. Pattern specific attentional modulation of V4 spatiotemporal receptive fields during free-viewing visual search. *Soc. Neurosci. Abstr.* 559.2

David SV, Vinje WE, Gallant JL. 2004. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.* 24:6991–7006

DeAngelis GC, Ohzawa I, Freeman RD. 1993. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J. Neurophysiol.* 69:1118–35

De Boer E. 1967. Correlation studies applied to the frequency resolution of the cochlea. *J. Aud. Res.* 7:209–17

deCharms RC, Blake DT, Merzenich MM. 1998. Optimizing sound features for cortical neurons. *Science* 280:1439–43

Depireux DA, Simon JZ, Klein DJ, Shamma SA. 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.* 85:1220–34

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18:193–222

De Valois RL, Albrecht DG, Thorell LG. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vis. Res.* 22:545–59

DiCarlo JJ, Johnson KO, Hsiao SS. 1998. Structure of receptive fields in area 3b of primary somatosensory cortex in the alert monkey. *J. Neurosci.* 18:2626–45

Dobson AJ. 2002. *An Introduction to Generalized Linear Models*. Boca Raton, FL: Chapman & Hall/CRC. 225 pp.

Eggermont JJ. 1993. Wiener and Volterra analyses applied to the auditory system. *Hear. Res.* 66:177–201

Eggermont JJ, Aertsen AMHJ, Johannesma PIM. 1983. Quantitative characterization procedure for auditory neurons based on the spectro-temporal receptive field. *Hear. Res.* 10:167–90

Emerson RC, Bergen JR, Adelson EH. 1992. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vis. Res.* 32:203–18

Emerson RC, Citron MC, Vaughn WJ, Klein SA. 1987. Nonlinear directionally selective subunits in complex cells of cat striate cortex. *J. Neurophysiol.* 58:33–65

Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1:1–47

Felsen G, Touryan J, Han F, Dan Y. 2005. Cortical sensitivity to visual features in natural scenes. *PLoS Biol.* 3:1819–28

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4:2379–94

Field DJ. 1994. What is the goal of sensory coding? *Neural Comput.* 6:559–601

Fletcher R. 1987. *Practical Methods of Optimization*. Chichester, NY: Wiley

Foldiak P. 2001. Stimulus optimisation in primary visual cortex. *Neurocomputing* 38:1217–22

Frazor RA, Albrecht DG, Geisler WS, Crane AM. 2004. Visual cortex neurons of monkeys and cats: temporal dynamics of the spatial frequency response function. *J. Neurophysiol.* 91:2607–27

Fritz J, Shamma SA, Elhilali M, Klein DJ. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6:1216–23

Frye MA, Dickinson MH. 2001. Fly flight: a model for the neural control of complex behavior. *Neuron* 32:385–88

Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC. 1996. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* 76:2718–39

Gaska JP, Jacobson LD, Chen HW, Pollen DA. 1994. Space-time spectra of complex cell filters in the macaque monkey: a comparison of results obtained with pseudowhite noise and grating stimuli. *Vis. Neurosci.* 11:805–21

Gilbert CD, Wiesel TN. 1990. The influence of contextual stimuli on the orientation selectivity of cells in primary visual-cortex of the cat. *Vis. Res.* 30:1689–701

Golomb D, Kleinfeld D, Reid RC, Shapley RM, Shraiman BI. 1994. On temporal codes and the spatiotemporal response of neurons in the lateral geniculate nucleus. *J. Neurophysiol.* 72:2990–3003

Grace JA, Amin N, Singh NC, Theunissen FE. 2003. Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophysiol.* 89:472–87
Hammer B, Gersmann K. 2003. A note on the universal approximation capability of support vector machines. *Neural Process. Lett.* 17:43–53
Hansen KA, David SV, Gallant JL. 2004. Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *Neuroimage* 23:233–41
Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. *Neural Netw.* 2:359–66
Horwitz GD, Chichilnisky EJ, Albright TD. 2005. Blue-yellow signals are enhanced by spatiotemporal luminance contrast in macaque V1. *J. Neurophysiol.* 93:2263–78
Hsu A, Borst A, Theunissen FE. 2004. Quantifying variability in neural responses and its application for the validation of model predictions. *Network* 15:91–109
Ito M, Tamura H, Fujita I, Tanaka K. 1995. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73:218–26
Jenison RL, Schnupp JWH, Reale RA, Brugge JF. 2001. Auditory space-time receptive field dynamics revealed by spherical white-noise analysis. *J. Neurosci.* 21:4408–15
Johnson DH. 1980. Applicability of white-noise nonlinear system analysis to the peripheral auditory system. *J. Acoust. Soc. Am.* 68:876–84
Jones JP, Palmer LA. 1987. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58:1187–211
Kass RE, Ventura V, Brown EN. 2005. Statistical issues in the analysis of neuronal data. *J. Neurophysiol.* 94:8–25
Keller CH, Takahashi TT. 2000. Representation of temporal features of complex sounds by the discharge patterns of neurons in the owl's inferior colliculus. *J. Neurophysiol.* 84:2638–50
Klein DJ, Depireux DA, Simon JZ, Shamma SA. 2000. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J. Comput. Neurosci.* 9:85–111
Knierim JJ, Van Essen DC. 1992. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67:961–80
Kobatake E, Tanaka K. 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71:856–67
Kowalski N, Depireux DA, Shamma SA. 1996. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J. Neurophysiol.* 76:3503–23
Lau B, Stanley GB, Dan Y. 2002. Computational subunits of visual cortical neurons revealed by artificial neural networks. *Proc. Natl. Acad. Sci. USA* 99:8974–79
Lehky SR, Sejnowski TJ, Desimone R. 1992. Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. *J. Neurosci.* 12:3568–81
Lesica NA, Boloori AS, Stanley GB. 2003. Adaptive encoding in the visual pathway. *Network* 14:119–35
Lewis ER, van Dijk P. 2004. New variations on the derivation of spectro-temporal receptive fields for primary auditory afferent axons. *Hear. Res.* 189:120–36
Li W, Piech V, Gilbert CD. 2004. Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* 7:651–57
Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM. 2003. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J. Neurophysiol.* 90:2660–75

Luck SJ, Chelazzi L, Hillyard SA, Desimone R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77:24–42

Machens CK, Gollisch T, Kolesnikova O, Herz AVM. 2005. Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron* 47:447–56

Machens CK, Wehr MS, Zador AM. 2004. Linearity of cortical receptive fields measured with natural sounds. *J. Neurosci.* 24:1089–100

Mackay DJC. 1995. Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks. *Network* 6:469–505

Mancini M, Madden BC, Emerson RC. 1990. White noise analysis of temporal properties in simple receptive fields of cat cortex. *Biol. Cybern.* 63:209–19

Marmarelis PZ, Marmarelis VZ. 1978. *Analysis of Physiological Systems: The White-Noise Approach*. New York: Plenum. 487 pp.

Marmarelis PZ, McCann GD. 1973. Development and application of white-noise modeling techniques for studies of insect visual nervous system. *Kybernetik* 12:74–89

Marmarelis VZ. 2004. *Nonlinear Dynamic Modeling of Physiological Systems*. Hoboken, NJ: Wiley Interscience. 541 pp.

Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL. 2002. Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl. Acad. Sci. USA* 99:1645–50

McAdams CJ, Maunsell JHR. 1999. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19:431–41

McAlpine D. 2005. Creating a sense of auditory space. *J. Physiol.* 566:21–28

Miller LM, Escabi MA, Read HL, Schreiner CE. 2001. Functional convergence of response properties in the auditory thalamocortical system. *Neuron* 32:151–60

Movshon JA, Thompson ID, Tolhurst DJ. 1978. Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol.* 283:79–99

Nelken I. 2004. Processing of complex stimuli and natural scences in the auditory cortex. *Curr. Opin. Neurobiol.* 14:474–80

Nelken I, Kim PJ, Young ED. 1997. Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models. *J. Neurophysiol.* 78:800–11

Nishimoto S, Arai M, Ohzawa I. 2005. Accuracy of subspace mapping of spatiotemporal frequency domain visual receptive fields. *J. Neurophysiol.* 93:3524–36

Nishimoto S, Ishida T, Ohzawa I. 2006. Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *J. Neurosci.* 26:3269–80

Nykamp DQ, Ringach DL. 2002. Full identification of a linear-nonlinear system via cross-correlation analysis. *J. Vis.* 2:1–11

Ohzawa I, Sclar G, Freeman RD. 1985. Contrast gain control in the cat's visual system. *J. Neurophysiol.* 54:651–67

Olshausen BA, Anderson CH, Vanessen DC. 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* 13:4700–19

Pack CC, Born RT, Livingstone MS. 2003. Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. *Neuron* 37:525–35

Pece AEC, French AS, Korenberg MJ, Kuster JE. 1990. Nonlinear mechanisms for gain adaptation in locust photoreceptors. *Biophys. J.* 57:733–43

Perge JA, Borghuis BG, Bours RJE, Lankheet MJM, van Wezel RJA. 2005. Temporal dynamics of direction tuning in motion-sensitive macaque area MT. *J. Neurophysiol.* 93:2104–16

Polley DB, Heiser MA, Blake DT, Schreiner CE, Merzenich MM. 2004. Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. *Proc. Natl. Acad. Sci. USA* 101:16351–56

Prenger RJ, Willmore B, Gallant JL. 2005. Nonlinear pooling of orientation and spatial frequency channels in V2. *Soc. Neurosci. Abstr.* 618.17

Prenger RJ, Wu MCK, David SV, Gallant JL. 2004. Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Netw.* 17:663–79

Priebe NJ, Ferster D. 2005. Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron* 45:133–45

Rapela J, Mendel JM, Grzywacz NM. 2006. Estimating nonlinear receptive fields from natural images. *J. Vis.* In press

Read HL, Winer JA, Schreiner CE. 2002. Functional architecture of auditory cortex. *Curr. Opin. Neurobiol.* 12:433–40

Reid RC, Victor JD, Shapley RM. 1997. The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Vis. Neurosci.* 14:1015–27

Reynolds JH, Chelazzi L. 2004. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* 27:611–47

Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W. 1997. *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press. 395 pp.

Ringach DL, Hawken MJ, Shapley R. 1997a. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387:281–84

Ringach DL, Hawken MJ, Shapley R. 2002. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J. Vis.* 2:12–24

Ringach DL, Sapiro G, Shapley R. 1997b. A subspace reverse-correlation technique for the study of visual neurons. *Vis. Res.* 37:2455–64

Robert CP. 2001. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York: Springer. 604 pp.

Rust NC, Schwartz O, Movshon JA, Simoncelli EP. 2005. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46:945–56

Sahani M, Linden JF. 2003a. Evidence optimization techniques for estimating stimulus-response functions. In *Advances in Neural Information Processing Systems*, ed. S Becker, S Thrun, K Obermayer, pp. 301–8. Cambridge, MA: MIT Press

Sahani M, Linden JF. 2003b. How linear are auditory cortical responses? In *Advances in Neural Information Processing Systems*, ed. S Becker, S Thrun, K Obermayer, pp. 109–16. Cambridge, MA: MIT Press

Sakai HM, Naka KI. 1992. Response dynamics and receptive-field organization of catfish amacrine cells. *J. Neurophysiol.* 67:430–42

Schwartz AB. 2004. Cortical neural prosthetics. *Annu. Rev. Neurosci.* 27:487–507

Sen K, Theunissen FE, Doupe AJ. 2001. Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* 86:1445–58

Sharpee T, Rust NC, Bialek W. 2004. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural. Comput.* 16:223–50

Sharpee T, Sugihara H, Kurgansky A, Rebrik S, Stryker MP, Miller KD. 2006. Adaptive filtering enhances information transmission in visual cortex. *Nature* 439:936–42

Simoncelli EP, Heeger DJ. 1998. A model of neuronal responses in visual area MT. *Vis. Res.* 38:743–61

Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24:1193–216

Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ. 2003. The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J. Neurosci.* 23:4746–59

Stanley GB. 2002. Adaptive spatiotemporal receptive field estimation in the visual pathway. *Neural. Comput.* 14:2925–46

Sutter E. 1987. A practical nonstochastic approach to nonlinear time-domain analysis. In *Advanced Methods in Physiological System Modelling, Volume 1*, ed. VZ Marmarelis, pp. 303–15. Los Angeles, CA: Biomed. Simul. Resour. Univ. S. Calif.

Tanaka K, Saito H, Fukada Y, Moriya M. 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* 66:170–89

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. 2001. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12:289–316

Theunissen FE, Sen K, Doupe AJ. 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* 20:2315–31

Tomita M, Eggermont JJ. 2005. Cross-correlation and joint spectro-temporal receptive field properties in auditory cortex. *J. Neurophysiol.* 93:378–92

Touryan J, Felsen G, Dan Y. 2005. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 45:781–91

Touryan J, Lau B, Dan Y. 2002. Isolation of relevant visual features from random stimuli for cortical complex cells. *J. Neurosci.* 22:10811–18

Vapnik VN. 1995. *The Nature of Statistical Learning Theory*. New York: Springer. 188 pp.

Victor JD. 2005. Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy. *Nat. Neurosci.* 8:1651–56

Victor JD, Purpura K, Katz E, Mao BQ. 1994. Population encoding of spatial frequency, orientation, and color in macaque V1. *J. Neurophysiol.* 72:2151–66

Victor JD, Shapley RM. 1980. Method of nonlinear analysis in the frequency domain. *Biophys. J.* 29:459–83

Vinje WE, Gallant JL. 2002. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.* 22:2904–15

Willmore B, Prenger RJ, Gallant JL. 2005. Spatial and temporal receptive field properties of neurons in area V2. *Soc. Neurosci. Abstr.* 137.11

Willmore B, Smyth D. 2003. Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. *Network* 14:553–77

Woolley SMN, Fremouw TE, Hsu A, Theunissen FE. 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.* 8:1371–79

Woolley SMN, Gill PR, Theunissen FE. 2006. Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *J. Neurosci.* 26:2499–512

Wu MCK, Gallant JL. 2004. What's beyond second order? Kernel regression techniques for nonlinear functional characterization of visual neurons. *Soc. Neurosci. Abstr.* 984.20

Yamada WM, Lewis ER. 1999. Predicting the temporal responses of nonphase-locking bullfrog auditory units to complex acoustic waveforms. *Hear. Res.* 130:155–70

Yang T, Maunsell JHR. 2004. The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.* 24:1617–26

Young ED, Calhoun BM. 2005. Nonlinear modeling of auditory-nerve rate responses to wide-band stimuli. *J. Neurophysiol.* 94:4441–54

Yu HH, de Sa VR. 2004. Nonlinear reverse correlation with synthesized naturalistic noise. *Neurocomputing* 58–60:909–13

Yu Y, Romero R, Lee TS. 2005. Preference of sensory neural coding for 1/f signals. *Phys. Rev. Lett.* 94:108103

Annual Review
of Neuroscience

Volume 29, 2006

Contents

Adaptive Roles of Programmed Cell Death During Nervous System Development
Robert R. Buss, Woong Sun, and Ronald W. Oppenheim 1

Endocannabinoid-Mediated Synaptic Plasticity in the CNS
Vivien Chevaleyre, Kanji A. Takahashi, and Pablo E. Castillo 37

Noncoding RNAs in the Mammalian Central Nervous System
Xinwei Cao, Gene Yeo, Alysson R. Muotri, Tomoko Kuwabara, and Fred H. Gage 77

The Organization of Behavioral Repertoire in Motor Cortex
Michael Graziano 105

TRP Ion Channels and Temperature Sensation
Ajay Dhaka, Veena Viswanath, and Ardem Patapoutian 135

Early Events in Olfactory Processing
Rachel I. Wilson and Zachary F. Mainen 163

Cortical Algorithms for Perceptual Grouping
Pieter R. Roelfsema 203

Deep Brain Stimulation
Joel S. Perlmutter and Jonathan W. Mink 229

RNA-Mediated Neuromuscular Disorders
Laura P.W. Ranum and Thomas A. Cooper 259

Locomotor Circuits in the Mammalian Spinal Cord
Ole Kiehn 279

Homeostatic Control of Neural Activity: From Phenomenology to Molecular Design
Graeme W. Davis 307

Organelles and Trafficking Machinery for Postsynaptic Plasticity
Matthew J. Kennedy and Michael D. Ehlers 325

Noncanonical Wnt Signaling and Neural Polarity
Mireille Montcouquiol, E. Bryan Crenshaw, III, and Matthew W. Kelley 363

Pathomechanisms in Channelopathies of Skeletal Muscle and Brain
Stephen C. Cannon 387

Imaging Valuation Models in Human Choice
P. Read Montague, Brooks King-Casas, and Jonathan D. Cohen 417

Brain Work and Brain Imaging
Marcus E. Raichle and Mark A. Mintun 449

Complete Functional Characterization of Sensory Neurons by System Identification
Michael C.-K. Wu, Stephen V. David, and Jack L. Gallant 477

Neurotrophins: Mediators and Modulators of Pain
Sophie Pezet and Stephen B. McMahon 507

The Hedgehog Pathway and Neurological Disorders
Tammy Dellovade, Justyna T. Romer, Tom Curran, and Lee L. Rubin 539

Neural Mechanisms of Addiction: The Role of Reward-Related Learning and Memory
Steven E. Hyman, Robert C. Malenka, and Eric J. Nestler 565

INDEXES

Subject Index 599

Cumulative Index of Contributing Authors, Volumes 20–29 613

Cumulative Index of Chapter Titles, Volumes 20–29 617

ERRATA

An online log of corrections to *Annual Review of Neuroscience* chapters (if any, 1977 to the present) may be found at http://neuro.annualreviews.org/