Stochastic characterization of small-scale algorithms for human sensory processing

Peter Neri[a)]

Aberdeen Medical School, Institute of Medical Sciences, Aberdeen, Scotland AB25 2ZD, United Kingdom

Human sensory processing can be viewed as a functional $\mathcal{H}$ mapping a stimulus vector $\mathbf{s}$ into a decisional variable r. We currently have no direct access to r; rather, the human makes a decision based on r in order to drive subsequent behavior. It is this (typically binary) decision that we can measure. For example, there may be two external stimuli $\mathbf{s}^{[0]}$ and $\mathbf{s}^{[1]}$, mapped onto $r^{[0]}$ and $r^{[1]}$ by the sensory apparatus $\mathcal{H}$; the human chooses the stimulus associated with largest r. This kind of decisional transduction poses a major challenge for an accurate characterization of $\mathcal{H}$. In this article, we explore a specific approach based on a behavioral variant of reverse correlation techniques, where the input $\mathbf{s}$ contains a target signal corrupted by a controlled noisy perturbation. The presence of the target signal poses an additional challenge because it distorts the otherwise unbiased nature of the noise source. We consider issues arising from both the decisional transducer and the target signal, their impact on system identification, and ways to handle them effectively for system characterizations that extend to second-order functional approximations with associated small-scale cascade models. © *2010 American Institute of Physics*. [doi:10.1063/1.3524305]

During the past 50 years, reverse correlation has become the elective methodology for the characterization of sensory neurons. Starting with the 1970s, similar tools have been developed for probing the sensory processes mediating vision/audition in humans. In both single-neuron electrophysiology and sensory psychophysics, the dominant model has been one where a linear filter is followed by a static nonlinearity (e.g., spike generation in single neurons, behavioral choice in human observers). For white-noise inputs and provided this simple model is adequate, results from reverse correlation experiments are relatively easy to interpret and the linear kernel is an appropriate descriptor for the process of interest. However, many processes operating within both neurons and observers are not adequately captured by the linear-nonlinear cascade model; in these instances, the linear kernel characterization must be augmented by additional nonlinear kernels. Methods for characterizing nonlinear kernels have been developed for application with single neurons, but the extension to human observers is not trivial due to significant differences between the two systems relating to the characteristics of the sensory input to the system, as well as the nature of the available output from the system. We tackle this problem at both theoretical and experimental levels, and show how some of the distortions in the kernel estimation procedure that are idiosyncratic to the human sensory process can be turned to the experimenter's advantage and exploited to gain additional information on the process at hand.

I. INTRODUCTION

A. Animal sensors as signal detection devices

The most important goal of sensory processing is to drive adequate behavior: the animal must process information about the environment in order to take action that will maximize its survival. Natural selection has produced sophisticated sensory systems, which have been conceptualized in the form of signal detection devices during the past 50 years;[17,18,20] the application of signal detection theory (SDT) to animal sensory processing still represents the most robust approach for interpreting quantitative measurements of this phenomenon.[38] We consequently adopt SDT here without questioning its applicability, while at the same time recognizing that it may not provide an adequate description for every aspect of animal sensory processing.

Implicit to SDT is the decision variable assumption:[64] the animal's decision comes in the form of a choice among a set of sensory stimuli, where the choice is based on one figure of merit for each stimulus. Regardless of the dimensionality of the incoming stimulation, each stimulus is therefore mapped to a single scalar value; this value is meant to reflect how likely the corresponding stimulus is to contain the signal of interest. The animal then chooses the stimulus associated with largest estimated likelihood. The estimated likelihood is the actual likelihood if the animal operates like a Bayesian device;[20] if not, then it is simply some figure of merit constrained by what the animal's neural hardware is capable of doing.

B. Output distortion: The decisional transducer

There is a fundamental distinction between the operator that maps input vectors into scalars on the one hand, and the operator that converts scalars into decisions on the other. The

[a)]Electronic mail: peter.neri@abdn.ac.uk.

former (which we call $\mathcal{H}$) is a functional that encompassesall processing layers believed to reflect general and physiologically relevant properties of how visual processing operates, while the latter (which we call Ψ) is dependent upon the specific question that is asked of observers and the form in which the response is acquired. Our primary interest is in characterizing $\mathcal{H}$, not Ψ—indeed the desired goal is to characterize $\mathcal{H}$ regardless of the specific Ψ that is instantiated by the chosen experimental protocol.

To provide a concrete example, we imagine that our interest is in how the human visual system processes the direction of moving objects. We select some stimulus specifics, say dots moving at a certain speed. One member of our laboratory runs experiment 1 in which the observer sees two such stimuli on each trial, and must choose which one moves better (where a precise definition of "better" is irrelevant for the current discussion). Another member runs experiment 2 in which the observer sees four such stimuli on every trial, and must choose one out of four. A third member of the laboratory runs experiment 3 in which the observer only sees one stimulus on every trial and must judge how well it moves on a rating scale from 1 to 10. The three experimenters are measuring outputs of different kinds: binary in experiment 1, quaternary in experiment 2, on a 10-point rating scale in experiment 3. These different outputs are necessarily associated with different decisional rules on the part of the observer, yet our hope is that the final characterization of the system will be one and the same: the neural circuit used by human observers to process moving stimuli. Not simply because we wish the ensuing characterization to be representative of the experiments carried out in the laboratory, but more importantly because we wish it to be representative of visual processing in the natural environment where humans must cope with a wide range of different tasks. We therefore split the description of the system into a general-purpose operator ($\mathcal{H}$) followed by an *ad hoc* decisional transducer (Ψ). We want $\mathcal{H}$; the question is whether we can bypass Ψ to get at it, which is one of the main concerns of this article.

C. Input distortion: The target signal

The other main concern relates to the unavoidable presence of a target signal. As we have detailed above, our ability to measure behavior relies on querying human participants and recording their response. We therefore need to pose a question, and it must be well defined in order to allow meaningful treatment of the subject. Within the context of SDT, the question involves selecting a stimulus that contains a certain signal, requiring us to define a signal-to-be-detected. This apparently innocuous requirement has fundamental implications for reverse correlation methodologies like the one described here. These approaches rely on sensory inputs consisting of noisy processes with unbiased characteristics, such as Gaussian white noise.[40] The introduction of a consistent signal source, in the form of a sensory target for the human to detect, represents a significant departure from these requirements.

It is legitimate to ask whether one could not simply do away with the target, and ask human participants to press

FIG. 1. The sensory process as conceptualized here. Gray items delineate the system of interest and the ideal setting for accessing and characterizing it: a controlled unbiased input with elliptical symmetry (Gaussian noise source) is mapped to a scalar output value by the functional $\mathcal{H}$ which captures the relevant aspects of the sensory apparatus. Our goal is to characterize $\mathcal{H}$. When applied to human sensory processing, this framework must be modified in two substantial ways, highlighted by the black items: (1) the input must contain a target signal, thus shifting and biasing the input distribution on the left (Sec. I C); (2) the scalar output must be converted into a (typically binary) choice by the decisional transducer Ψ (Sec. I B). The challenge is to characterize $\mathcal{H}$ despite the gross distortions introduced by these two transformations.

buttons while observing pure noise (e.g., Ref. [19]). This approach (which amounts to an extreme version of previous strategies where the analysis is restricted to target-absent trials in yes-no tasks[1,4,52,75]) presents insurmountable problems. First, it is very difficult to prompt a stable behavioral strategy on the part of the observer because, among other reasons, no meaningful correct/incorrect feedback can be provided (there is no correct or incorrect answer when there is no signal). Second, it is impossible to know whether the observer is performing the task at all: he/she may be pressing buttons randomly without any way for the experimenter to know, because there is no objective manner of establishing whether the system is detecting anything (there is nothing to be detected). Third, worst and most important of all, it is impossible to know whether the observer is performing a different task from the one specified by the experimenter: observers may perform some task and generate some form of measurable outcomes, but the experimenter cannot be certain as to what task that is. These and related issues exclude the possibility that robust results may be obtained in the total absence of a target signal. Furthermore, there are situations in which adding a target can actually increase the ability of the experimenter to characterize a certain class of mechanisms (see Sec. VI A). The impact of the target signal on a reliable characterization of $\mathcal{H}$ is the second main concern of this article. The two topics discussed above, introduction of a target signal and decisional transduction, are highlighted and summarized in Fig. 1. As emphasized in the figure, the former affects what goes into the system (input), the latter affects what comes out of the system (output).

D. Structure of this article

We start by fleshing out the conceptual framework outlined above, i.e., define the input, $\mathcal{H}$, Ψ, etc. This is done in Sec. II where we approximate $\mathcal{H}$ using a Volterra functional expansion (Sec. II B). This approximation recasts $\mathcal{H}$ in terms of system kernels; within Sec. II we therefore consider one specific way of estimating these kernels based on cross-correlation[47,68] (Sec. II C). Only a brief overview is given in Sec. II, to be followed by more detailed treatments in Secs. III–V. Section III introduces some basic results re-

lating to the handling of Ψ, i.e., how it can be bypassed to access the $\mathcal{H}$ kernels. We then consider issues relating to first-order kernel estimation in Sec. IV and second-order kernel estimation in Sec. V. Many of the difficulties that arise in connection with kernel estimation considered in these sections relate to the presence of the target signal (as already mentioned in Sec. I C). Finally, in Sec. VI we consider how kernel estimation can be used to infer/inform the structure of simple algorithms that attempt to simulate human sensory processing; we do this via a combination of both theoretical and experimental materials. Because the connection between kernel structure and potential underlying cascade model is currently best understood only for a standard set of simple models,[16,27,41,86] we show how well-established nonlinear models in the vision literature can be approximated using these simple cascades (Secs. VI A and VI B) and then consider relevant examples from real data (Secs. VI C–VI F). Of the four data sets which we evaluate, two are easily interpreted using the tools described in this article (Secs. VI C and VI D) while the remaining two present significant challenges for the proposed approach (Secs. VI E and VI F).

E. Notation

In treating these topics we adopt the following conventions: vectors/matrices are indicated in bold letters, e.g., $\mathbf{H}$; occasionally we refer to individual elements (e.g., k) via $\mathbf{H}(x_k)$ (we do not adopt $\mathbf{H}_k$ or H_k to avoid additional subscripting and to accommodate the : notation detailed in the next line). When we index using : we take the entire corresponding vector dimension, e.g., $\mathbf{H}_2(:,x_k)$ is a 1D vector consisting of the m elements $\mathbf{H}_2(x_j,x_k)$ for j from 1 to m for a fixed k. We denote the inner product by $\langle\cdot,\cdot\rangle$ and the outer product by $\otimes$ (we do not adopt $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{x}\mathbf{y}^{\mathrm{T}}$ to avoid additional superscripting). When matrices appear as arguments of $\langle\cdot,\cdot\rangle$, the Frobenius inner product is assumed [i.e., if $\mathbf{A}$ and $\mathbf{B}$ are two matrices, $\langle\mathbf{A},\mathbf{B}\rangle=\mathrm{tr}(\mathbf{A}\mathbf{B}^{\mathrm{T}})$]. Using this notation, $\langle\mathbf{H}_2(:,x_k),\mathbf{H}_1\rangle$ is the inner product between 1D vector $\mathbf{H}_2(:,x_k)$ and 1D vector $\mathbf{H}_1$. $\langle\ \rangle_i$ stands for (ensemble) average over an infinite number of trials m, i.e., $\lim_{m\to\infty}1/m\Sigma_m$ (also $\langle\mathbf{x}\rangle_i=\langle\mathbf{x},1/m\rangle$ if x_i are elements of $\mathbf{x}$). The following common symbols are adopted: $*$ for convolution, $\star$ for cross-correlation, and $\circ$ for Hadamard (entrywise) product. We use $\triangleq$ for matrix assignments $\mathbf{A}\triangleq\mathbf{B}$ to indicate $\mathbf{A}=(\mathbf{B}+\mathbf{B}^{\mathrm{T}})/2$ (assign symmetry by definition to $\mathbf{A}$ whenever not explicit in $\mathbf{B}$). We adopt the digital signal processing notation for Kronecker δ: $\delta[x]=1$ for $x=0$, $\delta[x]=0$ for $x\neq 0$, where x is integer.

II. GENERAL FRAMEWORK

A. Input, task, and decisional transducer

Our focus is on the two alternative forced choice (AFC) design (detailed below), arguably the most robust and best understood experimental protocol in sensory psychophysics.[20] We occasionally consider the implications of response bias (Sec. III B 2) for which the 2AFC design

approximates a yes-no protocol. This is undesirable for a number of reasons,[20] particularly in relation to nonlinear kernel estimation.[47,50]

In the 2AFC design, two stimuli are presented on every trial: stimulus $\mathbf{s}^{[0]}$ contains the nontarget $\mathbf{t}^{[0]}$, while stimulus $\mathbf{s}^{[1]}$ contains the target $\mathbf{t}^{[1]}$. Noise $\mathbf{n}$ is added to both so that, on trial i, $\mathbf{s}_i^{[q]}=\mathbf{n}_i^{[q]}+\mathbf{t}^{[q]}$ ($q=0$ for nontarget, 1 for target). The statistics governing every element of $\mathbf{n}$ is the same across stimuli and trials and is a Gaussian noise source. Some results (Sec. III A) are applicable to any circularly symmetric noise distribution (see also Ref. 61), but we restrict the specifics of the derivations here to Gaussian noise for which $\langle\mathbf{n}_i\rangle_i=0$, $\langle\mathbf{n}_i(x_j)\mathbf{n}_i(x_k)\rangle_i=0$ for $j\neq k$ and $=\sigma_N^2=\mathbf{w}$ for $j=k$, $\langle\mathbf{n}_i(x_1)\cdots\mathbf{n}_i(x_{2M+1})\rangle_i=0$ for $M=0,1,2\ldots$, $\langle\mathbf{n}_i(x_1)\cdots\mathbf{n}_i(x_{2M})\rangle_i=\Sigma\Pi\langle\mathbf{n}_i(x_j)\mathbf{n}_i(x_k)\rangle_i$, where $\Sigma\Pi$ stands for summation over all distinct ways of partitioning the $2M$ random variables into products of averages of pairs.[68]

In line with SDT, each stimulus $\mathbf{s}_i^{[q]}$ maps to a scalar response $r_i^{[q]}$. This is the transformation we are specifically interested in, i.e., the functional $\mathcal{H}:\mathbf{s}\to r$. However we do not have access to r, but rather to a binary response z_i (0 for incorrect and 1 for correct detection), which is generated based on a comparison between $r_i^{[1]}$ and $r_i^{[0]}$. Following SDT, we make the very reasonable assumption that the probability of making a correct response $\mathrm{p}(z_i=1)$ on trial i (which we abbreviate as p_i) is $\mathrm{p}_i=\Psi(r_i^{[1]}-r_i^{[0]})=\Psi(r_i)$. We can make a number of statements about Ψ: $\Psi(0)=0.5$ (the percentage of correct responses should be at chance if the output of the system is the same for target and nontarget), $\lim_{x\to-\infty}\Psi(x)=0$, $\lim_{x\to\infty}\Psi(x)=1$, and Ψ is monotonically increasing.[14] In this article, we use a polynomial approximation for Ψ, which is technically adequate only if Ψ is analytic. However, with a few exceptions our conclusions are not based on accurate approximations of Ψ, so they can be at least qualitatively extended to nonanalytic transducers too. In practice, Ψ almost certainly conforms to a well-behaved sigmoid curve so that our polynomial approximation poses no concern. In principle, the only candidate exception would be the unit step function [$u(x)=0$ for $x<0$ and $=1$ for $x>0$, not analytic at 0], but it is not even approximately realizable in human sensory processing due to the presence of a very significant amount of internal noise.[50] No deterministic transducer can therefore be operating in the human observer, requiring that Ψ be specified statistically to reflect the characteristics of the internal noise source. We can capture Ψ adequately using a relatively shallow Weibull function, cumulative Gaussian function, or hyperbolic tangent (any of these choices would be appropriate here). When we approximate Ψ near its operating point $\bar{r}=\langle r_i\rangle_i$ using

$$\hat{\Psi}_{\tilde{d}}(r_i)\approx\sum_{d=0}^{\tilde{d}}\Psi^{(d)}(r_i-\bar{r})^d,\qquad(1)$$

we do not necessarily use $\Psi^{(d)}$ to indicate local derivatives as in the Taylor series, but the coefficients that minimize
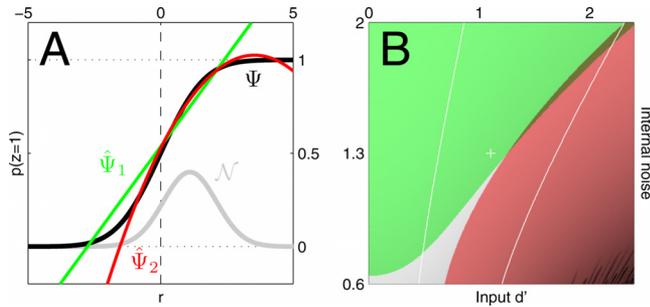
FIG. 2. (Color online) Realistic characterization of Ψ. (a) Gray trace shows a plausible distribution for r, the differential output generated by $\mathcal{H}$, assumed normal ($\mathcal{N}$) in line with SDT. Black trace presents a plausible characteristic for Ψ in the form of a cumulative Gaussian function with a standard deviation of 1.3; this corresponds to the most basic 2AFC decisional rule (choose stimulus with largest associated output) corrupted by a veridical internal noise source (Ref. [49]). Green trace shows first-order approximation $\hat{\Psi}_1$ to Ψ, red trace shows second-order approximation [see Eq. (1)]. (b) Explained variance for first-order approximation $\hat{\Psi}_1$ within the range spanned by $\mathcal{N}$ as a function of input d′ [before the addition of internal noise (Ref. [49])] and internal noise. Green region indicates acceptable range for $\hat{\Psi}_1$, defined as explained variance >90%. Red region indicates range for which second-order approximation $\hat{\Psi}_2$ is markedly superior to $\hat{\Psi}_1$, defined as explained variance for the former exceeding the latter by 10%. White lines delimit the typical region for humans corresponding to a 2AFC correct performance between 65% and 85%.

$$\int \mathcal{N}(r - \bar{r})(\Psi(r) - \hat{\Psi}_{\tilde{d}}(r))^2 dr \tag{2}$$

subject to the inequality constraint

$$\frac{\partial}{\partial r}\hat{\Psi}_{\tilde{d}}(\bar{r} + r, \theta) > 0, r \in (-2, 2) \tag{3}$$

for a specific choice of approximation order $\tilde{d}$, parametrization θ, and mean response $\bar{r}$, where r is in units of its expected standard deviation across trials and $\mathcal{N}(x)$ is the normal distribution. The minimizer in Eq. (2) is weighted by $\mathcal{N}$ to target the operating range for r, which we assume to be normally distributed following the standard practice in SDT;[20] Equation (2) is therefore $\langle(\Psi(r_i) - \hat{\Psi}_{\tilde{d}}(r_i))^2\rangle_i$ written as expectancy. The requirement in Eq. (3) enforces positive monotonicity on Ψ within the range of interest (± 2 standard deviations around $\bar{r}$), which is theoretically necessary to enable meaningful treatments of the mapping from stimulus to percept.[14]

Figure 2 uses the approximation in Eqs. (1)–(3) to demonstrate that given current experimental knowledge relating to Ψ, approximations of low-order (sometimes as low as linear) are often adequate for our purposes. Figure 2(a) shows plausible examples of what Ψ may look like and where $\mathcal{N}(r - \bar{r})$ would have to be located with relation to it in order to generate 75% correct responses, which is a typical choice for threshold performance;[20,46] Ψ is modeled as a cumulative Gaussian function centered at 0 with standard deviation θ_1. Corresponding first- and second-order Ψ approximations are shown in green and red, respectively. Figure 2(b) plots the explained variance for different values of internal noise (θ_1) and input d′ (Ref. [49]) ($\bar{r}$); the specific choice shown in (a) corresponds to the white cross in (b).

The green region indicates acceptable first-order approximations; it can be seen that this region extends to most of the "physiological" range (defined here as the range of 65%–85% correct performance and bracketed by the white contours). However, there is a substantial portion of this range for which a second-order approximation is preferable (red region). Throughout this study we therefore consider both first-order and second-order approximations to Ψ, depending on which one is necessary to demonstrate specific results.

B. Volterra expansion of $\mathcal{H}$

We approximate $\mathcal{H}$ using an adapted Volterra expansion,[40,50,68]

$$r^{[q]} = \sum_d \langle \mathbf{H}_d, \Theta_d(\mathbf{s}^{[q]})\rangle, \tag{4}$$

where $\Theta_d(j_1, \ldots, j_d, \mathbf{s}) = \Pi_d \mathbf{s}(x_{j_d})$ is the dth degree monomial matrix of $\mathbf{s}$ (same feature mapping used by polynomial classifiers in machine learning[15,70]). For example, $\Theta_1 = \mathbf{s}$ and $\Theta_2 = \mathbf{s} \otimes \mathbf{s}$. $\mathbf{H}_d$ therefore enjoys symmetry by definition. Equation (4) can be viewed as a generalization of the Taylor series to several variables.[41] From an experimental viewpoint, our interest is restricted to second-order approximations because higher-order kernel characterizations are impractical (previous work has shown that reliable characterizations of $\mathbf{H}_2$ require at least 6–$10k$ trials per human observer,[50] already approaching the upper limit for feasible projects on a sizeable number of participants). We occasionally approximate to third-order [e.g., Eq. (20)] to study the impact of third-order system kernels on lower-order kernel estimation.

Readers familiar with Volterra expansion will recognize that Eq. (4) should be expanded using convolution;[40,68] please refer to Sec. III B 3 for details on why the psychophysical variant can be equivalently expanded using inner products. Different from the output of the convolution-based expansion, which is ordinarily expressed as a time-varying function,[40] the output of Eq. (4) is a scalar (the decision variable): it generates a static snapshot from the underlying dynamic behavior of the system.[36] The lack of temporal dynamics for the output does not preclude characterization of the temporal dynamics associated with the sensory process that analyzes the input (see Fig. 9 for relevant examples); however, it represents an important limitation on the experimenter's ability to constrain the resulting characterization.[35] The issue is one of resolving power: because we only acquire a binary response from the human participant for every temporal segment of the stimulus that we present on every trial (and we assume that the process is static from trial to trial), our ability to resolve the temporal scale of the output is poorer than our ability to control the temporal scale of the input.

Alternative experimental protocols have been attempted in which a response is acquired from the participant at any time during a long stimulus presentation,[11,66,72] but these methods have not afforded higher temporal resolution of the resulting system characterization (in some cases the resolution seemed lower, see Ref. [50] for a discussion of this issue in relation to a comparison between Refs. [11] and [53]); even if

somewhat higher resolution is obtained by prompting fast responses, it is then difficult to know whether the observed effects are due to perceptual or motor preparation.[73] The limitation in resolving the decisional output at high temporal resolution derives from at least two factors: (1) the intrinsic timescale of the read-out process supporting behavioral decisions (the "psychological moment") is expected to lie in the ~ 200 ms range;[85] (2) the motor act by which humans convey a perceptual decision is inherently noisy on the range of 200–500 ms (e.g., Ref. [72]). Because of these and potentially other factors, there is an intrinsic limit to the temporal resolution at which behavioral decisions can be accessed experimentally; this limit is likely to exceed the temporal scale at which most sensory detectors operate. Nevertheless, the temporal dynamics of the perceptual process can be estimated at higher resolution by modulating the temporal properties of the input on a finer scale; we discuss a relevant example in Sec. VI E (see also Ref. [50]).

The above-detailed limitations become less relevant if we transition from low-level sensory processing on a scale of ~ 20–200 ms to higher-level cognitive processing on a scale of ~ 2–20 s (generically referred to as learning). The nonlinear dynamics (sometimes chaotic[23]) of this class of phenomena has been characterized using a variety of methods,[22] including nonlinear kernel estimation:[23] examples come from a wide range of psychological functions such as manual tracking,[24] handwriting dynamics,[23] associative memory,[45] visual illusions,[21,78,79] decision making,[28] prediction of event sequences,[23,43,56,74] and even cognitive development.[81] In these experiments, the output is sampled at a temporal resolution comparable to the modulation applied to the input, but the associated temporal scale is typically two orders of magnitude greater than the relevant scale for sensory signal detection. Furthermore, it is unclear whether the critical parameter for the dynamics of these phenomena is time *per se* or rather the occurrence of a cognitive event: first-order kernels for priming (a form of short-term memory) decay as a function of the number of sensory events,[37] to some extent independent of event duration, similar to other memory phenomena.[45] These cognitive processes are no doubt interesting but they fall outside the focus of interest for the present article, which is primarily on phenomena that approach the maximum affordable temporal resolution at which behavior can be accessed for empirical measurement. In Sec. VI F we consider an application to visual adaptation, a process that belongs to the class of dynamic phenomena which operate on a significantly longer timescale.

C. Kernel estimation via cross-correlation

Following established methods,[3] first-order kernel estimates can be expressed as

$$\hat{\mathbf{H}}_1 = \sum_{q,z} (2\delta[q-z]-1)\hat{\mathbf{H}}_1^{[q,z]},$$

$$\hat{\mathbf{H}}_1^{[q,1]} = \frac{1}{\bar{\mathrm{p}}}\langle \mathrm{p}_i \mathbf{n}_i^{[q]}\rangle_i, \tag{5}$$

$$\hat{\mathbf{H}}_1^{[q,0]} = \frac{1}{(1-\bar{\mathrm{p}})}\langle (1-\mathrm{p}_i)\mathbf{n}_i^{[q]}\rangle_i,$$

where $\Sigma_{q,z}=\Sigma_{q=0}^1\Sigma_{z=0}^1$ and $\bar{\mathrm{p}}=\langle \mathrm{p}_i\rangle_i$. Expressed this way, $\hat{\mathbf{H}}_1^{[q,z]}$ is the expected value of $\mathbf{n}^{[q]}$ when restricted to trials on which the observer responded z (see Ref. [3]). Second-order kernel estimates can be similarly expressed as[47,50]

$$\hat{\mathbf{H}}_2 = \sum_{q,z} (2\delta[q-z]-1)(\hat{\mathbf{H}}_2^{[q,z]} - \hat{\mathbf{H}}_1^{[q,z]}\otimes\hat{\mathbf{H}}_1^{[q,z]}), \tag{6}$$

where

$$\hat{\mathbf{H}}_2^{[q,1]} = \frac{1}{\bar{\mathrm{p}}}\langle \mathrm{p}_i\mathbf{n}_i^{[q]}\otimes\mathbf{n}_i^{[q]}\rangle_i,$$

$$\hat{\mathbf{H}}_2^{[q,0]} = \frac{1}{1-\bar{\mathrm{p}}}\langle (1-\mathrm{p}_i)\mathbf{n}_i^{[q]}\otimes\mathbf{n}_i^{[q]}\rangle_i. \tag{7}$$

It is clear from the subtraction term in Eq. (6) (involving $\hat{\mathbf{H}}_1$) that $\hat{\mathbf{H}}_2$ represents a differential covariance matrix. In principle, there appears to be no obvious reason why one should not simply compute the differential second-order moment $\Sigma_{q,z}(2\delta[q-z]-1)\hat{\mathbf{H}}_2^{[q,z]}$; this would seem a more natural choice to keep in line with established practice.[40] We show later (Sec. V B) that covariance is a more robust estimator of $\mathbf{H}_2$ under certain conditions, thus motivating the correction term in Eq. (6). We also introduce the following notation for estimates obtained from target versus nontarget stimuli, respectively:

$$\hat{\mathbf{H}}_1^{[1]} = \sum_z (2\delta[z-1]-1)\hat{\mathbf{H}}_1^{[1,z]},$$

$$\hat{\mathbf{H}}_1^{[0]} = \sum_z (2\delta[z]-1)\hat{\mathbf{H}}_1^{[0,z]}.$$

Clearly, $\hat{\mathbf{H}}_1=\hat{\mathbf{H}}_1^{[1]}+\hat{\mathbf{H}}_1^{[0]}$. We define $\hat{\mathbf{H}}_2^{[1]}$ and $\hat{\mathbf{H}}_2^{[0]}$ similarly, i.e., using Eq. (6) restricted to $q=1$ or $q=0$ so that $\hat{\mathbf{H}}_2=\hat{\mathbf{H}}_2^{[1]}+\hat{\mathbf{H}}_2^{[0]}$.

The following expressions are immediately derived from those listed above:

$$\hat{\mathbf{H}}_1^{[q,0]} = -\frac{\bar{\mathrm{p}}}{1-\bar{\mathrm{p}}}\hat{\mathbf{H}}_1^{[q,1]}, \tag{8}$$

$$\hat{\mathbf{H}}_2^{[q,0]} = \frac{\mathbf{Iw}}{1-\bar{\mathrm{p}}} - \frac{\bar{\mathrm{p}}}{1-\bar{\mathrm{p}}}\hat{\mathbf{H}}_2^{[q,1]}, \tag{9}$$

$$\hat{\mathbf{H}}_1^{[q,0]}\otimes\hat{\mathbf{H}}_1^{[q,0]} = -\frac{\bar{\mathrm{p}}^2}{(1-\bar{\mathrm{p}})^2}\hat{\mathbf{H}}_1^{[q,1]}\otimes\hat{\mathbf{H}}_1^{[q,1]},$$

where $\mathbf{I}$ is the identity matrix. From Eq. (8) we expect first-order subclass estimates from correct and incorrect trials to be scaled versions of each other, where the scaling factor is determined by the relative percentage of the two trial types. This result is known.[3] Equation (9) shows that a similar scaling factor applies to second-order subclass estimates, although with an additional diagonal term. From these equations, we can readily derive the following useful expressions:

$$\hat{\mathbf{H}}_1 = \frac{1}{1-\bar{p}}(\hat{\mathbf{H}}_1^{[1,1]} - \hat{\mathbf{H}}_1^{[0,1]}), \tag{10}$$

$$\sum_{q,z}(2\delta[q-z]-1)\hat{\mathbf{H}}_2^{[q,z]} = \frac{1}{1-\bar{p}}(\hat{\mathbf{H}}_2^{[1,1]} - \hat{\mathbf{H}}_2^{[0,1]}), \tag{11}$$

$$\sum_{q,z}(2\delta[q-z]-1)\hat{\mathbf{H}}_2^{[q,z]} \otimes \hat{\mathbf{H}}_1^{[q,z]} = \frac{1-2\bar{p}}{(1-\bar{p})^2}(\hat{\mathbf{H}}_1^{[1,1]}$$
$$\otimes \hat{\mathbf{H}}_1^{[1,1]} - \hat{\mathbf{H}}_1^{[0,1]} \otimes \hat{\mathbf{H}}_1^{[0,1]}), \tag{12}$$

where Eq. (12) refers to the covariance correction term in Eq. (6). It will be relevant for further treatment of this term (Sec. V B) that the factor $(1-2\bar{p})/(1-\bar{p})^2$ is negative because $\bar{p} > \frac{1}{2}$.

D. Experimental cross-check

We need to check that the framework detailed above works well in real applications, because it forms the basis for all subsequent analytical treatments of behavioral kernel estimation presented here. To appreciate the importance of providing a cross-check, consider Eq. (5): this is different from what is actually done when computing $\hat{\mathbf{H}}_1^{[q,1]}$ in the laboratory. Suppose we are computing $\hat{\mathbf{H}}_1^{[1,1]}$; in the laboratory, we collect say $10k$ trials on a given observer, and select the subset of those trials on which the observer responded correctly ($z=1$). We then select the subset of noise fields on which we presented the target ($q=1$), and average only these noise fields. This is not a probabilistic procedure; if we were to write it down, it would look more like

$$\hat{\mathbf{H}}_1^{[q,z]} = \langle \mathbf{n}_{i_k}^{[q]} \rangle_{i_k}, i_k \in A_z, \tag{13}$$

where A_z is the indexing subset for trials corresponding to response z given a specific finite collection of trials out of all possible trials otherwise indexed by i. In Eq. (5) the average is taken over this infinite set of all possible trials, weighted by p_i. We cannot actually compute Eq. (5), not even approximately for a finite number of trials, because we do not have direct access to a description of p_i (i.e., Ψ). We therefore need to check that we are on solid grounds when abstracting from Eq. (13) to Eq. (5) to ensure that our analytical results can be transferred back to the laboratory.

We can attempt a cross-check of this kind by verifying the empirical applicability of Eqs. (8) and (9), which are directly derived from Eq. (5), to data sets from our laboratory (detailed in Refs. 50 and 51). These equations involve kernel estimates for incorrect ($\hat{\mathbf{H}}_d^{[q,0]}$) and correct ($\hat{\mathbf{H}}_d^{[q,1]}$) trials, which we can measure from data using Eq. (13), and the percentage of correct responses $\bar{p}$, which we can easily estimate for a given experiment. We write Eq. (8) as $\hat{\mathbf{H}}_1^{[q,0]} = k\hat{\mathbf{H}}_1^{[q,1]}$, and compute the scaling factor k either by linear regression between $\hat{\mathbf{H}}_1^{[q,0]}$ and $\hat{\mathbf{H}}_1^{[q,1]}$, or from $-\bar{p}/(1-\bar{p})$ as prescribed by Eq. (8). If this equation is correct, we expect the two estimates to match. Figure 3(a) plots the scaling factor computed using method 1 on the abscissa versus method 2 on the ordinate. Figure 3(b) plots the same for $\hat{\mathbf{H}}_2$
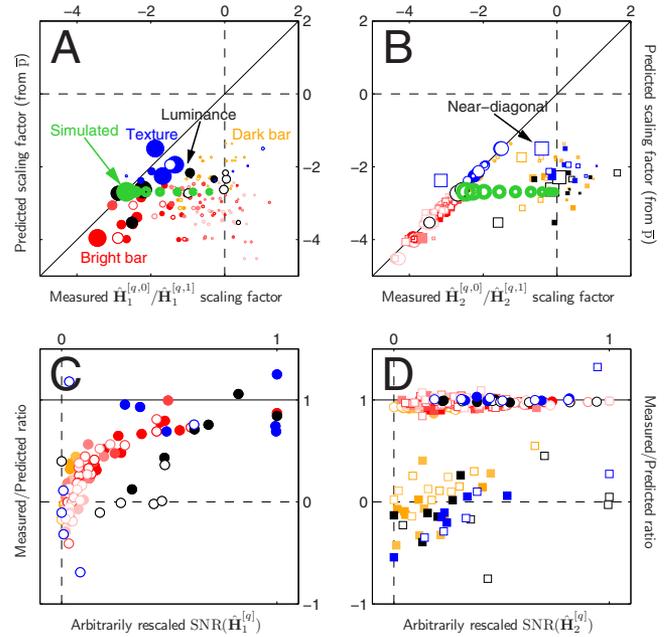

FIG. 3. (Color online) Experimental cross-check for Eqs. (8) and (9). Abscissa plots the linear regression coefficient (scaling factor) for kernel estimates from correct vs incorrect trials for first-order kernels in (a) and second-order kernels in (b). Ordinate plots the corresponding predictions computed from the percentage of correct responses (estimate for $\bar{p}$) using Eqs. (8) and (9). Solid symbols refer to target estimates ($q=1$), open symbols refer to nontarget ($q=0$); for these experiments, $\mathbf{t}^{[0]}=0$. Black and blue symbols show data for the luminance and texture experiments detailed in Ref. 50; red and yellow symbols show data for the spatial uncertainty experiments involving bright and dark-bar detection detailed in Ref. 51. Color saturation for the latter data set reflects spatial uncertainty (more saturated, less uncertain). Symbol size scales with kernel SNR [Eq. (14)]. Green symbols refer to the average of 100 iterations of a simulated L(N) model; increasing symbol sizes (correlations) were obtained by increasing the number of simulated trials per iteration from 100 to 102 400 in logarithmic steps. In (b), circles refer to coefficients computed from the entire second-order kernel, squares only to near-diagonal region (see Sec. II D). (c) and (d) replot the data from (a) and (b) with SNR on the abscissa and the ratio between the two values in (a) and (b) on the ordinate.

after accounting for the additional term $\mathbf{I}\mathbf{w}/(1-\bar{p})$ in Eq. (9). For both first-order [panel (a)] and second-order [panel (b)] kernel estimates, we have that the measured scaling factor on the abscissa falls between the expected value on the ordinate (when points fall on the solid unity line) and 0 (when points fall on the vertical dashed line). Does the latter trend reflect a failure of Eqs. (8) and (9)? Not necessarily, because a value of 0 on the abscissa is expected for noisy kernel measurements (see green symbols showing simulated results where less trials were used as proxy for less reliable measurements). In particular, according to this interpretation we expect that scaling factors near 0 would correspond to lower kernel signal-to-noise ratios (SNR) defined as

$$\text{SNR}(\hat{\mathbf{H}}_d) = \frac{dN\langle\hat{\mathbf{H}}_d^2, 1/m\rangle}{\mathbf{w}^d}, \tag{14}$$

where d can be 1 or 2, m is the full dimensionality of $\hat{\mathbf{H}}_d$ (i.e., the inner product in the numerator is simply the mean square), and N is the number of noise fields that went into computing $\hat{\mathbf{H}}_d$. Equation (14) extends the standard definitions of SNR (e.g., Ref. 46) to second-order kernels. Symbol size

scales with SNR in Figs. 3(a) and 3(b); particularly in the case of first-order kernels (a), it is clear that smaller symbols are closer to the vertical dashed line, while bigger ones are closer to the solid unity line (in conformity with the expectation detailed above). This trend can be exposed more readily by replotting the data as done in (c), where the ordinate plots the ratio between the two quantities in (a) and the abscissa plots the arbitrarily rescaled SNR. Equation (8) predicts that points should fall on the horizontal solid line; panel (c) shows that they asymptote to this value as SNR grows [all data sets show a significant positive correlation at p < 0.01 with the exception of the texture data set (blue), which is only marginally significant at p = 0.06].

Results for second-order kernels [panels (b) and (d)] require a slightly more involved analysis. When the entire $\hat{\mathbf{H}}_2$ kernel is considered (circles), Eq. (9) holds for all data sets and kernel subtypes [$\hat{\mathbf{H}}_2^{[1]}$ (solid) and $\hat{\mathbf{H}}_2^{[0]}$ (open)]. There is also no clear dependence on SNR (p > 0.05 for all data sets). This analysis treats diagonal and nondiagonal regions of the kernel in the same way, but there are good reasons for inspecting them separately: (1) the diagonal region corresponds to differential variance, rather than covariance; (2) Eq. (9) contains a term that specifically affects the diagonal alone; (3) one of the simplest cascade models used in neuroscience applications, the Hammerstein model[26] (see Sec. IV B), predicts second-order modulations only within the diagonal region.[26,41,86] These issues are highlighted by the square symbols in Fig. 3(b), which refer to the kernel region immediately adjacent to the diagonal (corresponding to 1's in the matrix $\mathbf{M}(x_\nu, x_\xi) \triangleq \delta[\nu - \xi - 1]$). For this region, the measured scaling factor is often close to 0 (several square symbols cluster around vertical dashed line in b) and there is a dependence on SNR similar to that observed for first-order kernels (square symbols in d), although only for dark-bar (yellow) and texture (blue) data sets (p < 0.005). The bright-bar data set (red symbols) is particularly interesting, in that Eq. (9) holds closely for both diagonal and near-diagonal regions (all red symbols fall on solid lines in b and d) and there is no detectable dependence on SNR (p > 0.05).

To summarize the results from Fig. 3, in general (i.e., across the data sets presented here) Eqs. (8) and (9) apply well. Most departures are at least in part attributable to limitations imposed by measurement noise, which curtails our ability to verify the applicability of the predictions.

III. THE DECISIONAL BOTTLENECK AND ITS IMPLICATIONS

A. Benchmark result: Bypassing Ψ for Wiener systems

In this section, we show that for the simple LN (linear-nonlinear) cascade ($\mathbf{H}_d = 0$ for $d > 1$ and N referring to Ψ) we have $\hat{\mathbf{H}}_1 \propto \mathbf{H}_1$ and $\hat{\mathbf{H}}_2 = 0$. The former result is well-established (generally referred to as Bussgang theorem[12]), the latter is specific to the context explored here and differs from that obtained in standard applications of Lee–Schetzen cross-correlation for which $\hat{\mathbf{H}}_2 \propto \mathbf{H}_1 \otimes \mathbf{H}_1$.[41,86] For the system

of interest $r_i - \bar{r} = \langle \mathbf{H}_1, \mathbf{n}_i^{[1]} - \mathbf{n}_i^{[0]} \rangle$ [from Eq. (4)]. We substitute it into Eq. (1) and insert the resulting Ψ approximation into Eq. (5),

$$\hat{\mathbf{H}}_1^{[1,1]}(x_\nu) = \frac{1}{\mathrm{p}} \left\langle \sum_{d=1}^{\infty} \Psi^{(d)} \left[\sum_j \mathbf{H}_1(x_j)(\mathbf{n}_i^{[1]}(x_j) - \mathbf{n}_i^{[0]}(x_j)) \right]^d \right.$$
$$\left. \times \mathbf{n}_i^{[1]}(x_\nu) \right\rangle_i, \tag{15}$$

where the dot product term within [] has been expanded into a sum for clarity. For each value of d, the term under Σ_d consists of a sum of terms of the form

$$\left[\prod_{k=1}^{a} \mathbf{H}_1(x_{y_k})\mathbf{n}_i^{[1]}(x_{y_k}) \prod_{k=1}^{b} - \mathbf{H}_1(x_{v_k})\mathbf{n}_i^{[0]}(x_{v_k}) \right] \mathbf{n}_i^{[1]}(x_\nu),$$

where y and v are indexing sets. These terms mostly vanish, e.g., whenever a is even and/or b is odd. When they do not vanish, they can be written as $k\mathbf{H}_1$, where $k > 0$ because b must be even. Equation (15) can therefore be rewritten compactly as

$$\hat{\mathbf{H}}_1^{[1,1]}(x_\nu) = \omega_1 \mathbf{H}_1(x_\nu),$$

where ω_1 does not depend on ν (note that we are implicitly exploiting the circular symmetry of the input noise source here; more generally, this result depends on elliptical symmetry[61]). $\hat{\mathbf{H}}_1^{[0,1]}$ follows a similar expression except $\mathbf{n}_i^{[1]}$ and $-\mathbf{n}_i^{[0]}$ are swapped, leading to a sign inversion,

$$\hat{\mathbf{H}}_1^{[0,1]}(x_\nu) = - \omega_1 \mathbf{H}_1(x_\nu).$$

We then have from Eq. (10) that $\hat{\mathbf{H}}_1 \propto \mathbf{H}_1$.

Following a similar procedure for $\hat{\mathbf{H}}_2$, we have

$$\hat{\mathbf{H}}_2^{[1,1]}(x_\nu, x_\xi) = \frac{1}{\mathrm{p}} \left\langle \sum_{d=1}^{\infty} \Psi^{(d)} \left[\sum_j \mathbf{H}_1(x_j)(\mathbf{n}_i^{[1]}(x_j) - \mathbf{n}_i^{[0]}(x_j)) \right]^d \right.$$
$$\left. \times \mathbf{n}_i^{[1]}(x_\nu)\mathbf{n}_i^{[1]}(x_\xi) \right\rangle_i,$$

where the terms summed under Σ_d are of the form

$$\left[\prod_{k=1}^{a} \mathbf{H}_1(x_{y_k})\mathbf{n}_i^{[1]}(x_{y_k}) \prod_{k=1}^{b} - \mathbf{H}_1(x_{v_k})\mathbf{n}_i^{[0]}(x_{v_k}) \right] \mathbf{n}_i^{[1]}(x_\nu)\mathbf{n}_i^{[1]}(x_\xi). \tag{16}$$

Using the same logic adopted for $\hat{\mathbf{H}}_1^{[1,1]}$, we obtain

$$\hat{\mathbf{H}}_2^{[1,1]} = \hat{\mathbf{H}}_2^{[0,1]} = \omega_2 \mathbf{H}_1 \otimes \mathbf{H}_1, \tag{17}$$

where there is no sign inversion between $\hat{\mathbf{H}}_2^{[1,1]}$ and $\hat{\mathbf{H}}_2^{[0,1]}$ because both a and b must be even for terms not to vanish in Eq. (16). Substituting into Eqs. (11) and (12) and then Eq. (6) leads to $\hat{\mathbf{H}}_2 = 0$. This result applies for any polynomial expansion of Ψ and here it is referred to as the "benchmark" result.

Later in this article, we demonstrate that kernel subclass estimates show strong target-related modulations, a result that has been experimentally observed on numerous occasions for first-order kernels.[1,4,50,52,75,82] This observation

leads one to question whether it may not be advisable to rely exclusively on target-absent estimates $\hat{\mathbf{H}}_1^{[0]}$ and $\hat{\mathbf{H}}_2^{[0]}$. This option may be available (depending on whether $\mathbf{t}^{[0]}=0$ and the amount of data is sufficient to afford curtailing it) for $\hat{\mathbf{H}}_1$ because it does not violate the benchmark result; however, it is not viable for $\hat{\mathbf{H}}_2$ because in this case $\hat{\mathbf{H}}_2^{[0]}\neq 0$, even though $\mathbf{H}_2=0$ (thus violating the benchmark result), as apparent from the above demonstration [Eq. (17)] that $\hat{\mathbf{H}}_2=0$ results from $\hat{\mathbf{H}}_2^{[1,1]}$ and $\hat{\mathbf{H}}_2^{[0,1]}$ canceling each other out in Eq. (11). If $\hat{\mathbf{H}}_2$ is computed by omitting either one (as in $\hat{\mathbf{H}}_2^{[0]}$), the benchmark result does not hold any longer. $\hat{\mathbf{H}}_2^{[0]}$ is therefore inadequate for the purposes examined here because it is largely affected by the decisional nonlinearity which (as discussed in Sec. I B) we are trying to circumvent in order to characterize $\mathcal{H}$.

B. Handling peridecisional operators (those applied right before choice)

1. Assimilation of nonlinearities into Ψ

Given the framework $p_i=\Psi(\mathcal{H}(\mathbf{s}_i^{[1]})-\mathcal{H}(\mathbf{s}_i^{[0]}))$ (Sec. II A) and associated benchmark result (Sec. III A), there are conditions under which $\mathcal{H}$ can be augmented by applying an operator Φ to its output r_i without altering our analysis in any significant way. The theoretical interest in this question stems from a number of considerations (see below), but chiefly from the possibility that peridecisional (i.e., immediately preceding the psychophysical choice) static nonlinearities may be incorporated into Ψ so as to extend the applicability of the benchmark result to $\mathcal{H}$ systems that are not strictly linear. To provide an example of how this analysis may be applied, we start with a noiseless system, i.e., $\Psi(x)=u(x)$ (deterministic choice transducer—see Sec. II A), and the application of a plausible static nonlinearity Φ; by plausible we mean that it is strictly monotonic and continuous within the operating range of r (see also Ref. 14). We can then easily incorporate Φ into Ψ: $\Psi(\Phi(\mathcal{H}(\mathbf{s}_i^{[1]}))-\Phi(\mathcal{H}(\mathbf{s}_i^{[0]})))=\Psi(\mathcal{H}(\mathbf{s}_i^{[1]})-\mathcal{H}(\mathbf{s}_i^{[0]}))$. It is easy to see why this simple result, also known as Birdsall's theorem,[33,80] applies in the noiseless case: $\Phi(y)-\Phi(x)$ preserves the sign of $y-x$, which is the only determinant of choice under a noiseless Ψ. We then have that the benchmark result applies even though the system was not written as linear-Ψ because its output is identical to the linear-Ψ version.

The extension to the noisy case is not trivial, unless Φ is applied after the internal noise source in which case Birdsall's theorem still applies. If applied before, our goal is $\Psi(\Phi(\mathcal{H}(\mathbf{s}_i^{[1]}))-\Phi(\mathcal{H}(\mathbf{s}_i^{[0]})))=\tilde{\Psi}(\mathcal{H}(\mathbf{s}_i^{[1]})-\mathcal{H}(\mathbf{s}_i^{[0]}))$, where $\tilde{\Psi}$ hopefully retains the same overall characteristics as Ψ [$\tilde{\Psi}(0)=0.5$, $\lim_{x\to-\infty}\tilde{\Psi}(x)=0$, $\lim_{x\to\infty}\tilde{\Psi}(x)=1$, monotonically increasing, see Sec. II A]. This is rarely the case; consider for example $\Phi(x)=e^x$. The map $(r^{[1]}-r^{[0]})\to(e^{r^{[1]}}-e^{r^{[0]}})$ is not injective, so it is not possible to specify a unique $\tilde{\Psi}$ on $r^{[1]}-r^{[0]}$; this, in turn, means that the benchmark result is not applicable and $\hat{\mathbf{H}}_2\neq 0$ for sizeable internal noise.
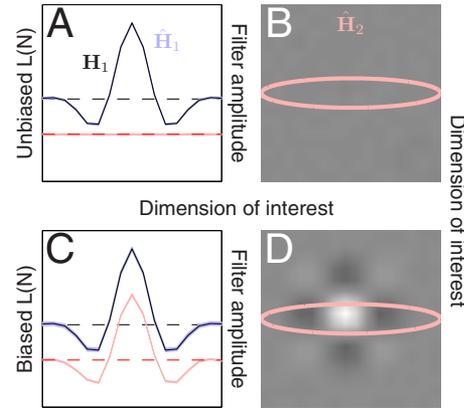

FIG. 4. (Color online) Response bias violates the benchmark result. The latter is demonstrated in (a) and (b): for a simulated system output $\langle \mathbf{H}_1,\mathbf{s}\rangle$ subjected to the decisional transducer Ψ [modeled using the unit step function $u(x)$], we obtain $\hat{\mathbf{H}}_1\propto\mathbf{H}_1$ (compare blue shading, which shows ± 1 standard deviation on $\hat{\mathbf{H}}_1$ for 100 simulations of 100k trials each, to black trace for $\mathbf{H}_1$) and $\hat{\mathbf{H}}_2=0$ [panel (b)]. Black/blue traces in (a) are plotted to the zero line indicated by the black dashed horizontal line, while light-red shading [corresponding to $\hat{\mathbf{H}}_2$ slice indicated by oval in (b)] is plotted to the zero line indicated by the red dashed horizontal line. In the presence of response bias [panels (c) and (d)], modeled as the application of a static nonlinearity (e^x) to the output from one of the two stimuli (randomly choosing between $\mathbf{s}^{[0]}$ and $\mathbf{s}^{[1]}$ on each trial), the benchmark result no longer holds for $\hat{\mathbf{H}}_2$, which is now $\neq 0$ (d) and more specifically $\propto\mathbf{H}_1\otimes\mathbf{H}_1$ as expected from theoretical considerations (see Sec. III B 2). Filter amplitude has been arbitrarily rescaled for each curve/surface, but using the same scaling factor for a given curve/surface between the top and bottom rows, so as to allow direct comparison between biased and unbiased kernels.

2. A notable example: Response bias

The critical requirement specified in Sec. III B 1 is that a unique $\tilde{\Psi}$ can be identified. We now consider its implications for what is perhaps the most significant source of artifactual results in human psychophysics: response bias (see Ref. 69 for a recent example). Response bias can introduce nonlinear distortions, potentially leading to $\hat{\mathbf{H}}_2\neq 0$ even though $\mathbf{H}_2=0$. We can model response bias as the application of a nonlinearity to the output from one of the two intervals before both outputs are combined and submitted to Ψ for choice behavior. This procedure warps the L(N) layer into a mixed L$\tilde{\text{N}}$(N) plus L(N) cascade, where $\tilde{\text{N}}$ cannot be assimilated into (N) using the result in Sec. III B 1 (see below). For this combined cascade (which stands for $\mathcal{H}$+bias rather than $\mathcal{H}$ alone), we know that the effective $\mathbf{H}_2\neq 0$ and therefore $\hat{\mathbf{H}}_2\neq 0$.

To demonstrate this result, it is sufficient to rewrite the decision variable as $r_{i_1}=\Phi(r_{i_1}^{[1]})-r_{i_1}^{[0]}$ on half the trials i_1 (when $\mathbf{t}^{[1]}$ is presented in the distorted interval) and $r_{i_2}=r_{i_2}^{[1]}-\Phi(r_{i_2}^{[0]})$ on the remaining half. We cannot assimilate Φ into Ψ because the map $(r^{[1]}-r^{[0]})\to(\Phi(r^{[1]})-r^{[0]})$ is not injective, i.e., it is not possible (in general) to rewrite the process using only one $\tilde{\Psi}$ (Sec. III B 1). If we then compute $\hat{\mathbf{H}}_2$ from the i_1 trials, it will expose the $\mathbf{H}_2$ kernel from the Wiener system $\Phi(r)$, which is of the form $\mathbf{H}_1\otimes\mathbf{H}_1$ [see Ref. 41 and Eq. (17)]. The same structure is exposed when computing from the i_2 trials, thus biasing $\hat{\mathbf{H}}_2$. This result is easily demonstrated via simulation, as shown in Fig. 4; notice the clear

$\mathbf{H}_1 \otimes \mathbf{H}_1$ structure for $\hat{\mathbf{H}}_2$ [panel (d)].

It is important to emphasize that the result detailed in the previous paragraph applies regardless of the presence/absence of internal noise [i.e., even if $\Psi(x) = u(x)$], because the map $(r^{[1]} - r^{[0]}) \rightarrow (\Phi(r^{[1]}) - r^{[0]})$ will in general not be homomorphic with respect to the order relation, i.e., $y > x$ does not necessarily mean that $y > \Phi(x)$ [$\Phi(x) = e^x$ provides a useful example], leading to a different choice from $r^{[1]} - r^{[0]}$ depending on the specific values of $r^{[1]}$ and $r^{[0]}$. In this sense, response bias implements a more pervasive distortion than the symmetrically applied nonlinearity discussed in Sec. III B 1. The simulated results in Fig. 4 refer to a noiseless system.

Response bias is a well-known issue in psychophysical experiments where only one stimulus is presented on every trial (either $\mathbf{s}^{[0]}$ or $\mathbf{s}^{[1]}$ randomly), and the observer is required to report on which one he/she perceived.[20] This class of experimental protocols, known as yes-no, should be avoided whenever possible. AFC protocols, when implemented effectively, eliminate bias. This goal is achieved well in the case of spatial AFC designs, where, for example, $\mathbf{s}^{[0]}$ or $\mathbf{s}^{[1]}$ appear simultaneously on opposite sides of the screen and the observer is asked to indicate which side contains $\mathbf{s}^{[1]}$. There are situations, however, when spatial AFC designs are not viable (e.g., auditory experiments); in this case, the choice is between first and second intervals, for which some bias may be expected.[89] A powerful method for reducing bias in these instances is to provide the participant with trial-by-trial feedback (correct/incorrect), which tends to make the observer/listener converge toward a near-optimal (bias-free) strategy.

3. Formulation of end linearity as inner product

When the last linear stage in $\mathcal{H}$ is linear, it can be written as an inner product regardless of how it is formulated.[50] This result is a direct consequence of the decision variable assumption (Sec. II A): $\mathcal{H}$ must return a single scalar value as argument to Ψ, whatever the dimensionality of the vector returned by the previous stage. If we call this vector $\mathbf{r}$, and we map it to scalar r via a linear transformation, we can always write $\mathbf{r} \rightarrow r$ as $\langle \mathbf{g}, \mathbf{r} \rangle$ for some appropriate choice of $\mathbf{g}$. Any preceding linear stage is similarly assimilated into a dot product, which allows us to rewrite the Volterra expansion in Eq. (4) using $\langle \cdot, \cdot \rangle$ rather than $*$ (as originally formulated by Volterra[41,68]). To see this explicitly for a second-order Volterra operator, first we can write its output using convolution,[40]

$$\mathbf{r} = \check{\mathbf{H}}_1 * \mathbf{s} + \check{\mathbf{H}}_2 * (\mathbf{s} \otimes \mathbf{s}). \tag{18}$$

We must now apply $\mathbf{r} \rightarrow r$ so that we can submit the output to Ψ for choice behavior (see above). When we express $\mathcal{H}$ this way, we can always rewrite its output either using Eq. (18) or equivalently Eq. (4),

$$\langle \mathbf{g}, \check{\mathbf{H}}_1 * \mathbf{s} + \check{\mathbf{H}}_2 * (\mathbf{s} \otimes \mathbf{s}) \rangle = \langle \mathbf{H}_1, \mathbf{s} \rangle + \langle \mathbf{H}_2, (\mathbf{s} \otimes \mathbf{s}) \rangle, \tag{19}$$

where $\mathbf{H}_d = \check{\mathbf{H}}_d \star \mathbf{g}$ [for $d = 2$ this is $\mathbf{H}_2(x_\nu, x_\xi) = \sum_j \check{\mathbf{H}}_2(x_{j-\nu}, x_{j-\xi}) \mathbf{g}(x_j)$] and $*$ has been replaced by $\langle \cdot \rangle$. Because in general we have no direct access to $\mathbf{g}$, the two for-

mulations are indistinguishable; our access is restricted to $\mathbf{H}_d$ in the form of $\hat{\mathbf{H}}_d$ (with associated limitations, e.g., Sec. IV A).

C. The -(N) notation

Taken together, the two simple results detailed above (Secs. III A and III B 1) allow us to bypass the decisional transducer Ψ and access $\mathcal{H}$ using established tools from nonlinear systems analysis. First we focus on the most basic system: linear ($\mathbf{H}_d = 0$ for $d > 1$). If we were reading off $\mathcal{H}$ directly, we would expect $\hat{\mathbf{H}}_2 = 0$ (as well as $\hat{\mathbf{H}}_1 \propto \mathbf{H}_1$). Due to the presence of Ψ, we are effectively monitoring the output of a Wiener linear-nonlinear cascade where the static nonlinearity is introduced by Ψ. As mentioned earlier, the standard applications of nonlinear kernel analysis return $\hat{\mathbf{H}}_2 \propto \mathbf{H}_1 \otimes \mathbf{H}_1$ for this cascade system.[41] Our goal is to develop an estimation strategy for $\mathbf{H}_2$ that bypasses Ψ, i.e., one in which we can treat the linear $+ \Psi$ system as simply linear and obtain $\hat{\mathbf{H}}_2 = 0$. This benchmark result is achieved by Eq. (6) (see Sec. III A). Later in the article (Sec. V), we show that Eq. (6) also approximates $\mathbf{H}_2$ when $\neq 0$, although distortions may be present. From Sec. III B 1 we also have that under restricted conditions, the application of any number of plausible Φ late transducers does not impact our discussion, thus extending the generality of the benchmark result. We indicate the bypassing of Ψ by bracketing the end-nonlinearity in cascade model formulations of $\mathcal{H}$: for example, if we assume a Korenberg LNL model (a cascade consisting of a linear filter, a static nonlinearity, and an additional linear filter, see Sec. IV C) for $\mathcal{H}$, we indicate this model as LNL(N) where the end decisional nonlinearity is bracketed to indicate that it does not affect our treatment of the subject. This approach allows us to apply known results for LNL cascade models[40,41,86] directly to $\mathcal{H}$ (Ref. 50) (see, for example, Sec. VI D).

IV. FIRST-ORDER KERNEL ESTIMATE

A. Nonlinear distortions

For $\mathcal{H}$ linear, we know that $\hat{\mathbf{H}}_1 \propto \mathbf{H}_1$ (Sec. III A). For a nonlinear system, the main distortion on $\hat{\mathbf{H}}_1$ is introduced by the presence of the target signal, regardless of Ψ; we can therefore demonstrate this result analytically using a simple linear approximation to Ψ. We also ignore kernels of order higher than 3 ($\mathbf{H}_d = 0$ for $d > 3$). Given these approximations, we have

$$\hat{\mathbf{H}}_1^{[q,1]}(x_\nu) = \frac{\mathbf{w}}{\bar{\mathsf{p}}} \Psi^{(1)}(2q-1)[\mathbf{H}_1 + 2\langle \mathbf{H}_2(x_\nu, :), \mathbf{t}^{[q]} \rangle$$
$$+ 3w \langle \mathbf{H}_3(x_\nu, :, :), \mathbf{I} \rangle + 3\langle \mathbf{H}_3(x_\nu, :, :), \mathbf{t}^{[q]}$$
$$\otimes \mathbf{t}^{[q]} \rangle]. \tag{20}$$

Readers familiar with Wiener orthogonalization will recognize that the term $3\mathbf{w} \langle \mathbf{H}_3(x_\nu, :, :), \mathbf{I} \rangle$ can be avoided by redefining $\hat{\mathbf{H}}_1$ as a Wiener (rather than Volterra) kernel estimate.[40] This distinction is not interesting in the present context mainly for two reasons. First, in most applications we as-

sume $\mathbf{H}_d=0$ for $d>2$;[47,50] in this case, there is no distinction between Volterra and Wiener first- and second-order kernels.[40] Second, even if we approximate to third-order as in Eq. (20), only the first-order kernel differs between Volterra and Wiener representations. The difference, represented by the term highlighted above, is negligible compared to the distortions introduced by terms containing $\mathbf{t}$. The distinction between target-present and target-absent kernels is far more relevant to $\hat{\mathbf{H}}_1$, so here we focus on this aspect of the estimation procedure rather than adopting a Wiener kernel representation. This representation is equivalent to the Volterra representation[40,86] and amounts to little more than a change in notation for the purposes examined here.

Further distortions can be demonstrated by approximating Ψ to second-order (we set $\mathbf{H}_3=0$ to simplify calculations),

$$
\hat{\mathbf{H}}_1^{[q,1]}(x_\nu) = \frac{\mathbf{w}}{\bar{\mathrm{p}}}\Psi^{(1)}(2q-1)[\mathbf{H}_1(x_\nu) + 2\langle\mathbf{H}_2(x_\nu,:),\mathbf{t}^{[q]}\rangle]
$$
$$
+ \frac{4\mathbf{w}^2}{\bar{\mathrm{p}}}\Psi^{(2)}(2\langle\mathbf{H}_2,\mathbf{H}_2(x_\nu,:)\otimes\mathbf{t}^{[q]}\rangle
$$
$$
+ \langle\mathbf{H}_2(x_\nu,:),\mathbf{H}_1\rangle). \tag{21}
$$

The above expression also demonstrates that even if we rely on target-absent estimates alone for an experiment where $\mathbf{t}^{[0]}=0$ (e.g., Ref. 50), we obtain

$$
\hat{\mathbf{H}}_1^{[0,1]}(x_\nu) = -\frac{\mathbf{w}}{\bar{\mathrm{p}}}\Psi^{(1)}\mathbf{H}_1(x_\nu) + \frac{4\mathbf{w}^2}{\bar{\mathrm{p}}}\Psi^{(2)}\langle\mathbf{H}_2(x_\nu,:),\mathbf{H}_1\rangle, \tag{22}
$$

exposing a residual departure from $\hat{\mathbf{H}}_1^{[0]}\propto\mathbf{H}_1$.

B. Special case: Hammerstein NL(N) cascade

The output of a Hammerstein cascade is defined as $\mathbf{f}*\Phi(\mathbf{s})$:[26] the underlying block-model consists of a front-end static nonlinearity followed by a linear filter (dynamic linearity). As explained in Sec. III B 1, the $*$ operation can be replaced by an inner product because $\mathcal{H}$ must return a scalar decision variable to obtain $\langle\mathbf{f},\Phi(\mathbf{s})\rangle$. For this cascade, we have $\mathbf{H}_2(x_\nu,x_\xi)=\delta[\nu-\xi](\Phi^{(2)}/\Phi^{(1)})\mathbf{H}_1(x_\nu)$ and $\mathbf{H}_3(x_\nu,x_\xi,x_\iota)=\delta[\nu-\xi]\delta[\nu-\iota](\Phi^{(3)}/\Phi^{(1)})\mathbf{H}_1(x_\nu)$, where $\Phi^{(d)}$ is the dth coefficient in the Taylor expansion of Φ (see, for example, Ref. 86). We consider the case $\mathbf{t}^{[1]}=\rho\delta[\tau]$ and $\mathbf{t}^{[0]}=0$. Equation (20) then reduces to

$$
\hat{\mathbf{H}}_1^{[1,1]}(x_\nu) = \frac{\mathbf{w}}{\bar{\mathrm{p}}}\Psi^{(1)}\bigg(\mathbf{H}_1(x_\nu)(1+3\mathbf{w}\Phi^{(3)})
$$
$$
+ \frac{\rho}{\Phi^{(1)}}\mathbf{H}_1(x_\nu)\delta[\nu-\tau_0](\Phi^{(2)}+3\rho\Phi^{(3)})\bigg),
$$

$$
\hat{\mathbf{H}}_1^{[1,1]}(x_\nu) = (\omega_3 + \rho\omega_4\delta[\nu-\tau_0])\mathbf{H}_1(x_\nu), \tag{23}
$$

where ω_3 and ω_4 do not depend on ν. Equation (23) shows that the $\hat{\mathbf{H}}_1^{[1,1]}$ estimate is a signal-distorted image of $\mathbf{H}_1$. From Eq. (20) we also have $\hat{\mathbf{H}}_1^{[0,1]}\propto\mathbf{H}_1$.
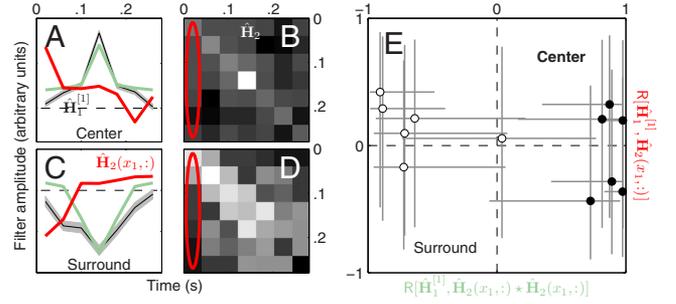

FIG. 5. (Color online) Target-present first-order estimates ($\hat{\mathbf{H}}_1^{[1]}$) for Korenberg systems [black traces in panels (a) and (c)] do not resemble the underlying front-end filter $\mathbf{f}$ but $\mathbf{f}\star\mathbf{f}$. Data set from Ref. 50; observers were required to perform an orientation discrimination task for a stimulus divided into center and surround. (b) and (d) show the corresponding second-order kernel estimates. If we assume a Korenberg cascade $\langle\mathbf{g},\Phi(\mathbf{f}*\mathbf{s})\rangle$ (see Sec. IV C), $\mathbf{f}$ can be estimated via $\hat{\mathbf{H}}_2(x_1,:)$ indicated by red oval and replotted in (a) and (c). Black traces show target-present first-order kernel estimates ($\hat{\mathbf{H}}_1^{[1]}$), green traces show the autocorrelation of the red traces. (e) plots the correlation coefficient between red and black traces on the ordinate vs the correlation coefficient between green and black traces on the abscissa for each subject separately (different data points) and for both center (solid) and surround (open). Error bars show ± 1 SEM. Traces in (a) and (c) have been independently and arbitrarily rescaled, including sign inversion for red and green in (c).

C. Special case: Korenberg LNL(N) cascade

The output of a Korenberg (sometimes termed "sandwich") cascade is defined as $\mathbf{g}*\Phi(\mathbf{f}*\mathbf{s})$:[31] the underlying block-model consists of a front-end linear filter, a static nonlinearity, and an additional linear stage (the nonlinearity is therefore "sandwiched" in between the two linear stages). We replace the end $*$ with inner product $\mathcal{H}$ (Sec. III B 1), and for $\langle\mathbf{g},\Phi(\mathbf{f}*\mathbf{s})\rangle$ we have[41,86]

$$
\mathbf{H}_d(x_{\nu_1},\ldots,x_{\nu_d}) = \Phi^{(d)}\sum_k \mathbf{g}(x_k)\mathbf{f}(x_k-x_{\nu_1})\cdots\mathbf{f}(x_k-x_{\nu_d}). \tag{24}
$$

We consider the special case $\mathbf{g}=1$ and $\mathbf{t}^{[1]}=\rho\delta$ ($\mathbf{t}^{[0]}=0$),

$$
\hat{\mathbf{H}}_1^{[1,1]} = \frac{\mathbf{w}}{\bar{\mathrm{p}}}\Psi^{(1)}[\Phi^{(1)}\langle\mathbf{f},1\rangle + 2\rho\Phi^{(2)}\mathbf{f}\star\mathbf{f} + 3\Phi^{(3)}
$$
$$
\times(\rho^2\mathbf{f}\star\mathbf{f}^2 + \mathbf{w}\langle\mathbf{f}\star\mathbf{f}^2,1\rangle)]. \tag{25}
$$

The inner products do not depend on ν, so we can rewrite this compactly as $\mathbf{f}\star\mathbf{f}+\rho\omega_5\mathbf{f}\star\mathbf{f}^2$. At low SNR (small ρ), we therefore expect $\hat{\mathbf{H}}_1^{[1,1]}\propto\mathbf{f}\star\mathbf{f}$ (see also Ref. 51).

We can attempt to verify whether this analytical result is born out by data for measurements of human texture processing; we previously established that the relevant data set is consistent with a Korenberg cascade[50] (see Sec. VI D for details). If we assume that this was the underlying structure in the human observers, we can estimate $\mathbf{f}$ via $\hat{\mathbf{H}}_2(x_1,:)$ (see Refs. 41 and 86 for details relating to this well-established result) shown in red in Fig. 5(a); $\hat{\mathbf{H}}_1^{[1]}$ is shown in black. We expect that the autocorrelation of the red trace, shown in green, is proportional to the black trace; this prediction seems approximately correct for both stimulus center [panel (a)] and surround [panel (c)]. To quantify the relationship

across observers, panel (e) plots the correlation coefficients for each observer separately; the abscissa plots the correlation between $\hat{\mathbf{H}}_1^{[1]}$ and the green trace, while the ordinate plots the correlation between $\hat{\mathbf{H}}_1^{[1]}$ and the red trace. If $\hat{\mathbf{H}}_1^{[1]}$ returned an estimate of the underlying impulse response $\mathbf{f}$, we would expect symbols to fall either below or above the horizontal dashed line (0 correlation on ordinate); instead, the ordinate values are no different from 0 (t-test across observers, p=0.68 for center and p=0.13 for surround). From Eq. (25) we expect $\hat{\mathbf{H}}_1^{[1]}$ to return an estimate of the autocorrelation of $\mathbf{f}$, not $\mathbf{f}$ itself; indeed the abscissa values are significantly different from 0 (points lie away from vertical dashed line at $p < 10^{-5}$ for center and $p < 0.01$ for surround). This result provides indirect confirmation of the applicability of Eq. (25) to real data.

V. SECOND-ORDER KERNEL ESTIMATE

A. Basic expressions

By combining Eqs. (1), (4), and (7) and adopting the same procedure used for $\hat{\mathbf{H}}_1$, we have the corresponding expressions for $\hat{\mathbf{H}}_2$,

$$\hat{\mathbf{H}}_2^{[q,1]} = \frac{\mathbf{wI}}{\bar{p}}\Psi^{(0)} + \frac{2\mathbf{w}^2}{\bar{p}}\Psi^{(1)}(2q-1)$$
$$\times\left(\mathbf{H}_2 + 3\sum_j \mathbf{H}_3(x_j,:,:)\mathbf{t}^{[q]}(x_j)\right) \quad (26)$$

for a third-order approximation of $\mathcal{H}$ and first-order approximation of Ψ, and

$$\hat{\mathbf{H}}_2^{[q,1]} = \frac{\mathbf{wI}}{\bar{p}}\Psi^{(0)} + \frac{2\mathbf{w}^2}{\bar{p}}\Psi^{(1)}(2q-1)\mathbf{H}_2$$
$$+ \frac{2\mathbf{w}^2}{\bar{p}}\Psi^{(2)}(4\mathbf{A}^{[q]} + \mathbf{B} + \mathbf{C}) \quad (27)$$

for a second-order approximation of $\mathcal{H}$ and second-order approximation of Ψ, where

$$\mathbf{A}^{[q]} \triangleq \mathbf{H}_1 \otimes \sum_j \mathbf{H}_2(x_j,:)\mathbf{t}^{[q]}(x_j) + \sum_{j,k}(\mathbf{H}_2(x_j,:)$$
$$\otimes \mathbf{H}_2(x_k,:))\mathbf{t}^{[q]}(x_j)\mathbf{t}^{[q]}(x_k), \quad (28)$$

$$\mathbf{B} = \mathbf{H}_1 \otimes \mathbf{H}_1 + 2\mathbf{w}\sum_j(\mathbf{H}_2(:,j) \otimes \mathbf{H}_2(:,j)),$$

$$\mathbf{C} = \mathbf{I}\left[\langle\mathbf{H}_1^2,1\rangle + 2\sum_j\mathbf{H}_1(x_j)\langle\mathbf{H}_2(j,:),\mathbf{t}^{[1]}+\mathbf{t}^{[0]}\rangle + 2\mathbf{w}\langle\mathbf{H}_2^2,1\rangle\right.$$
$$\left. + \sum_j 2\langle(\mathbf{H}_2(:,x_j) \otimes \mathbf{H}_2(:,x_j)),\mathbf{t}^{[1]}\otimes\mathbf{t}^{[1]}+\mathbf{t}^{[0]}\otimes\mathbf{t}^{[0]}\rangle\right].$$

These three terms have been kept separate to emphasize that $\mathbf{A}^{[q]}$ depends on q (i.e., it differs for target and nontarget subestimates), whereas $\mathbf{B}$ and $\mathbf{C}$ do not [i.e., they cancel out in Eq. (11)], and that $\mathbf{C} \propto \mathbf{I}$ (the quantity within [] is a scalar).

B. Why covariance and not second-order moment?

If we use a second-order moment estimate $\tilde{\mathbf{H}}_2$ for $\mathbf{H}_2$, we have from Eq. (27),

$$\tilde{\mathbf{H}}_2 = \sum_{q,z}(2\delta[q-z]-1)\hat{\mathbf{H}}_2^{[q,z]}$$
$$= \frac{1}{1-\bar{p}}\left(\frac{4\mathbf{w}^2}{\bar{p}}\Psi^{(1)}\mathbf{H}_2 + \frac{8\mathbf{w}^2}{\bar{p}}\Psi^{(2)}\mathbf{A}\right),$$

where $\mathbf{A} = \mathbf{A}^{[1]} - \mathbf{A}^{[0]}$ is the distorting term, i.e., causing departure from the desired relationship $\hat{\mathbf{H}}_2 \propto \mathbf{H}_2$. We rewrite this term below in a slightly different format from Eq. (28) to emphasize the structure induced by $\mathbf{t}$,

$$\mathbf{A}(x_\nu,x_\xi) \triangleq \mathbf{H}_1(x_\nu)\langle\mathbf{H}_2(x_\xi,:),\mathbf{t}^{[1]}-\mathbf{t}^{[0]}\rangle + \langle\mathbf{H}_2(x_\nu,:)$$
$$\otimes \mathbf{H}_2(x_\xi,:),\mathbf{t}^{[1]}\otimes\mathbf{t}^{[1]}-\mathbf{t}^{[0]}\otimes\mathbf{t}^{[0]}\rangle.$$

This term is proportional to the covariance subtraction term [from Eq. (21)],

$$\sum_{q,z}(2\delta[q-z]-1)\hat{\mathbf{H}}_1^{[q,z]} \otimes \hat{\mathbf{H}}_1^{[q,z]} = \left(\frac{2\mathbf{w}\Psi^{(1)}}{\bar{p}(1-\bar{p})}\right)^2(1-2\bar{p})\mathbf{A}.$$

When we estimate $\mathbf{H}_2$ using covariance [Eq. (6)], we therefore obtain

$$\hat{\mathbf{H}}_2 = \frac{4\mathbf{w}^2}{\bar{p}(1-\bar{p})}\left(\Psi^{(1)}\mathbf{H}_2 + \mathbf{A}\left[2\Psi^{(2)} - (\Psi^{(1)})^2\frac{1-2\bar{p}}{\bar{p}(1-\bar{p})}\right]\right).$$

Our goal is $\hat{\mathbf{H}}_2 \propto \mathbf{H}_2$ so we want the expression within [] to vanish. If we adopt a Taylor expansion for Ψ, we can rewrite this condition as

$$\ddot{\Psi} = \dot{\Psi}^2\frac{1-2\bar{p}}{\bar{p}(1-\bar{p})}.$$

At threshold $\left(\bar{p} \sim \frac{3}{4}\right)$ this translates into the requirement $\ddot{\Psi}/\dot{\Psi}^2 \sim -\frac{8}{3}$. For a realistic internal noise source and threshold performance [Fig. 2(a)], we have $\ddot{\Psi}/\dot{\Psi}^2 \sim -3$ at $\bar{r}$, which is within 15% of the above target value. Therefore, by computing $\hat{\mathbf{H}}_2$ in the form of a differential covariance matrix, we may compensate for 85% of the distortions introduced into the second-order moment estimate.

We demonstrate this result in Fig. 6 for simulated Hammerstein (top row) and Korenberg (bottom row) cascades (Secs. IV B and IV C). First-order kernels [black traces in (a) and (e)] are best estimated via $\hat{\mathbf{H}}_1^{[0]}$ (gray traces); when target-present trials are included, the resulting estimates $\hat{\mathbf{H}}_1$ (blue traces) are distorted, as expected from Eqs. (23)–(25). Expected second-order kernels are shown in Figs. 6(b) and 6(f); these were not simulated, but computed as predicted in the absence of Ψ (see Secs. IV B and IV C). The last two columns show estimates obtained from differential second-order moments (third column) and covariance (fourth column). As further emphasized by replotting diagonals in (a) and (e), (see captions) second-order moment estimates $\tilde{\mathbf{H}}_2$ present clear distortions near target position (negative dip in green trace, compare with red trace). These distortions are largely eliminated when computing differential covariance
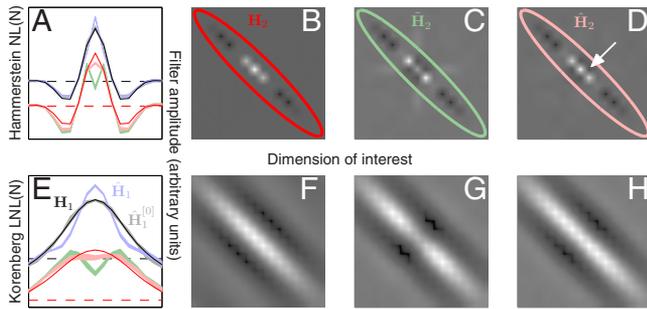
FIG. 6. (Color online) Second-order kernel estimates from differential second-order moments ($\tilde{\mathbf{H}}_2$) vs covariance ($\hat{\mathbf{H}}_2$). Top row shows the results for a simulated Hammerstein NL(N) cascade $\langle \mathbf{f}, e^s \rangle$ (Sec. IV B), bottom row for Korenberg LNL(N) $\langle e^{\mathbf{f}*s}, \mathbf{g} \rangle$ (Sec. IV C), where $\mathbf{g}$ is a Gaussian envelope with standard deviation equal to $\approx \frac{1}{2}$ the period of the $\mathbf{f}$ carrier. (a) Black trace shows $\mathbf{f}$ (which is $\propto \mathbf{H}_1$), blue trace shows the first-order kernel estimate, and gray trace shows the target-absent first-order kernel estimate [$\mathbf{t}^{[0]}$ was set to 0 for these simulations, and $\mathbf{t}^{[1]} = \rho\delta$ (peak at center of dimension of interest, zero elsewhere) with ρ selected to yield 75% correct responses]. Red/green traces plot the diagonals of kernels shown in (b), (c), and (d). Black trace in (e) shows $\mathbf{H}_1$ (not $\mathbf{f}$) for Korenberg, i.e., $\mathbf{f} \star \mathbf{g}$. [(b) and (f)] Expected second-order kernel structure if Ψ is bypassed successfully (see Secs. IV B and IV C). [(c) and (g)] Second-order kernel estimates using differential second-order moments. [(d) and (h)] Second-order kernel estimates using differential covariance. Plotting conventions similar to Fig. 4.

$\hat{\mathbf{H}}_2$ (light red), as we predict from the equations detailed above. Notice, however, that some distortions persist, in particular the negative flanks outside the diagonal in (d) (indicated by an arrow). For a Hammerstein cascade $\mathbf{H}_2(x_\nu, x_\xi) \propto \delta[\nu - \xi]\mathbf{H}_1(x_\nu)$ (see Refs. 26, 41, and 86 and Sec. IV B): no modulation is present outside the diagonal in $\mathbf{H}_2$ [see panel (b)], a characteristic that $\hat{\mathbf{H}}_2$ fails to estimate correctly.

VI. SMALL-SCALE ALGORITHMS

Cascade models like those considered in Secs. IV B and IV C consist of a few processing components assembled into a relatively simple circuit: they are small-scale networks. A significant advantage of considering sensory processing at the level of small-scale circuitry is that it affords scalability across systems: Korenberg cascades can be successfully exploited to describe processes within the fly brain[27,48] as well as the human brain,[50] despite obvious macroscopic anatomical differences between these two neural structures. Although their behavior can be unexpectedly complex and give rise to somewhat unexpected results (e.g., Ref. 9), it is legitimate to ask whether they can provide a reasonable account of a phenomenon like perceptual processing, which originates from a large-scale neuronal assembly (the brain). A related concern can be raised in connection with the Volterra cascade [Eq. (4)]: given that it is empirically feasible to obtain kernel characterizations only up to second-order (Sec. II B), is it reasonable to expect that this approach may yield a useful description of complex processes like human vision?

There is really no way to answer this question outside the laboratory: the ultimate and most relevant test of any approach for the characterization of human sensory processing is whether the approach is able to provide an accurate reconstruction of the sensory process as conveyed to us via empirical measurements. Lacking extensive experimental

evidence, there is no reason to prefer other approaches over the one outlined above; in fact, the knowledge accumulated so far suggests that this may be the most fruitful framework for understanding sensory processing within limited contexts, such as relatively simple detection tasks. Starting with pioneering work in the 1960s by a number of investigators (e.g., Barlow,[5] Mountcastle[44]), quantitative studies of sensory processing in humans have adopted the working hypothesis known as the lower envelope principle (LEP). According to this notion, humans presented with a specific sensory task monitor the output of only a restricted subset of their available neural circuitry; this subset consists of the neural components that are most sensitive for the task at hand.[62] Sensitivity is defined within the context of SDT (Ref. 20) and is therefore a well-characterized concept.

If we were to provide a concrete example of what the LEP translates to, we could imagine an observer who is presented with moving dots and asked to detect drift in a specified direction. According to the LEP, the observer bases his/her choice not on the output of his/her entire brain from occipital to frontal cortex, but only on the output from visual cortex. Further, the LEP postulates that the observer relies not on the entirety of visual cortex, but only on the area within visual cortex that is most responsive to motion; current knowledge would indicate MT+ as being the relevant area.[8] Even further, the LEP would restrict the relevant signals only to the subset of MT+ neurons with spatial receptive fields located within the region of visual space spanned by the stimulus, not on neurons responsive to other regions. And finally, the LEP may be interpreted to indicate that of these neurons only those responsive to the target direction would be monitored by the observer.[59] Those neurons and the associated circuitry would represent the bulk of $\mathcal{H}$.

The view expressed above is no doubt simplistic and open to a number of criticisms: to mention one, there is not even consensus over the notion that motion processing may be restricted to the MT (middle-temporal) region.[34] However, there is convincing experimental evidence that it may be applicable at least within the context of specific sensory stimuli, tasks, and the experimenter's ability to measure the quantities of interest.[62] The most compelling evidence in this respect dates back to the classic work by Newsome and collaborators in the 1990s,[59] who demonstrated that the behavioral choice expressed by primates in response to a visual motion task could be biased in a systematic and predictable manner by applying electrical stimulation to a restricted portion of visual cortex. For example, if the animal is asked to indicate the direction of a moving stimulus by pressing button 1 for leftward motion and button 2 for rightward motion, the proportion of button 1 presses can be increased by electrically stimulating the subregion of MT cortex containing neurons responsive to leftward motion.[58] Subsequent work has shown that similar results can be obtained for more complex stimuli and tasks such as face recognition.[2] These studies do not provide conclusive evidence for the LEP in its most extreme formulation, but they strongly suggest that weak variants may be applicable,[62] for example, versions of the LEP in which it is hypothesized that the subset of neural resources monitored by the observer may only roughly cor-

respond to the sensitive neurons with the inclusion of a number of irrelevant components, and furthermore that the extent of the envelope may depend on attention.[57]

Even if one were to accept the LEP in some form, it is not clear that the relevant neural circuitry would be adequately captured by small-scale cascade models: again this may seem oversimplistic. However, the question of interest here is not whether the relevant circuitry actually conforms to such descriptors, but rather whether its output in the form of a behavioral choice is consistent with such descriptors within the precision afforded by the empirical measurements. When the problem is viewed from the standpoint of an algorithmic solution, the latter specification is critical: there is an intrinsic limit to how detailed any characterization of human sensory processing can be for detection/discrimination at threshold; beyond this limit further algorithmic distinctions are irrelevant. Currently, this limit is determined primarily by the number of single-trial choices that can be feasibly recorded in the laboratory from a given participant, as well as the presence of a sizeable noise source that is intrinsic to the participant and therefore not under direct experimental control.[49] Both factors are integral to the process of recording behavior regardless of available technology, so it is unlikely that their impact will be reduced in the future (although it is possible to envisage, for example, that drugs may be developed to specifically reduce internal sensory noise). It is possible to obtain indirect estimates of internal noise in individual participants,[10] and use these to establish an upper limit on how well different models can be expected to predict human choices on individual trials.[55] It then becomes an empirical question whether small-scale algorithms can approach this limit or not. In previous work,[50,51] we have shown that they often do.

Below we consider two examples of models that have been successful in a vast number of applications in perception science, with the goal of demonstrating how they can be approximated via Korenberg cascades (already popular in neuroscience[76,88]) and subsequently analyzed using some of the tools described in this article. We use general formulations to emphasize their algorithmic potential for the solution of simple processing problems on a small scale. In both cases, a nonstatic nonlinearity is approximated by a combination of static nonlinearities and linear operators,[30,60] the advantage of the latter being that it forms the backbone of well-studied cascade systems for which kernel interpretation is relatively straightforward (Secs. IV B and IV C). Figure 7 summarizes these approximations; below we refer to them as "toy" examples to emphasize that they are model-based, not data based. We then consider four "real" examples from data. The first two examples (Secs. VI C and VI D) consist of kernel measurements which could be interpreted based on the kind of approximation presented in Fig. 7; the first example is for a Hammerstein cascade, the second one for a Korenberg cascade.

The third and fourth examples are in a sense counterexamples, in that the approach outlined in Fig. 7 did not offer a satisfactory interpretation for the data. In the third example (Sec. VI E), we consider a data set for which there was no straightforward, simple approximation of the kind presented
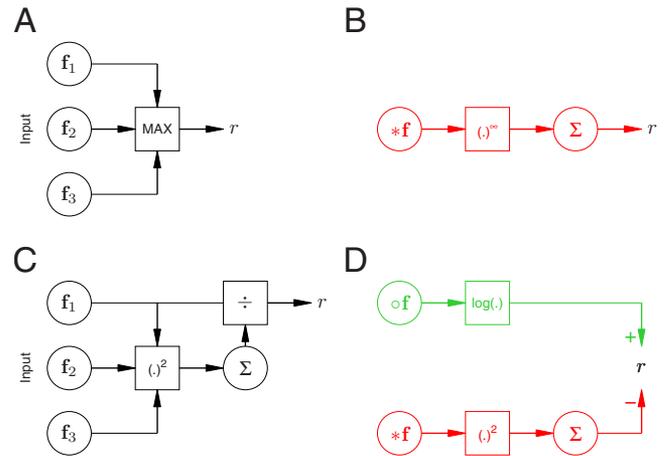

FIG. 7. (Color online) Approximation of MAX uncertainty (a) and normalization (c) models using Wiener (green) and Korenberg (red) cascades (Refs. 26 and 31). (a) is approximated in (b); (c) in (d). All symbols correspond to those used in the main text; see in particular Secs. VI A and VI B.

in Fig. 7; in this case, we resorted to a dynamically nonlinear model [Fig. 9(f)] that was able to capture the data satisfactorily while at the same time retaining physiological plausibility. This example in particular demonstrates how the utility of kernel characterization is not restricted to small-scale static approximations of dynamic systems; the information it provides can be effectively used to constrain plausible dynamic models. The fourth example (Sec. VI F) makes a related point but from a different perspective: it shows how dynamic nonlinearities that occur on a relatively long timescale (perceptual adaptation in this case) can be approached by treating each dynamic state as static on the much shorter timescale of the experimental characterization, and by inducing and characterizing different dynamic states separately one at a time. In the specific case of stereo motion adaptation considered in Sec. VI F, we used reverse correlation to estimate $\hat{\mathbf{H}}_1$ kernels after prolonged viewing of an adapting stimulus containing moving elements at specific three-dimensional positions, and compared them to kernels obtained in the absence of adaptation. The adaptor caused the system to shift onto a new perceptual state; we were able to measure specific signatures of this shift at the level of the $\hat{\mathbf{H}}_1$ kernels. Although these measurements in isolation are not able to capture the highly nonlinear and nonstatic nature of perceptual adaptation, they can be combined to generate an informative picture of the overall phenomenon at a greater level of detail than afforded by measurements that do not rely on kernel estimation (see Ref. 54 for the specific rationale behind this statement).

A. Toy example 1: MAX uncertainty model

The MAX uncertainty model[65] [Fig. 7(a)] can be approximated using a Korenberg cascade [Fig. 7(b)] by expressing the max operation as an ℓ^∞ norm.[51] For this model, we have

$$\mathbf{H}_d(x_{\nu_1}, \ldots, x_{\nu_d}) = \Phi^{(d)} \sum_k \mathbf{g}(x_k)^d \mathbf{f}(x_k - x_{\nu_1}) \cdots \mathbf{f}(x_k - x_{\nu_d}), \tag{29}$$

where $\mathbf{g}$ is raised to the d power because the uncertainty model is written as $\langle \Phi(\mathbf{g} \circ (\mathbf{f} * \mathbf{s})), 1 \rangle$, i.e., the sandwiched nonlinearity is applied after weighting by $\mathbf{g}$ (see Ref. 51); see Sec. IV C and compare Eq. (29) with Eq. (24).

If we restrict our analysis to the special case considered in Sec. IV C, i.e., $\mathbf{g} = 1$ and $\mathbf{t}^{[1]} = \rho\delta$ ($\mathbf{t}^{[0]} = 0$), we have (at low SNR) $\hat{\mathbf{H}}_1^{[1]} \propto \mathbf{f} \star \mathbf{f}$ (as detailed in Sec. IV C). From Eq. (20) we also have $\hat{\mathbf{H}}_1^{[0]} = \omega$, i.e., the target-absent first-order kernel is essentially featureless. More generally, it is $\propto \mathbf{g} \star \mathbf{f}$ and therefore reflects the extent of the uncertainty window, while $\hat{\mathbf{H}}_1^{[1]}$ provides an indirect image of the front-end filter $\mathbf{f}$ (see Sec. IV C for a tentative experimental validation of this result and Ref. 83 for a discussion of related notions). The uncertainty model therefore provides an interesting example of how both target-present and target-absent first-order estimates can deliver useful information in a complementary fashion.

B. Toy example 2: Divisive normalization

The other model we consider here is the normalization model commonly used to implement contrast gain control[25] [Fig. 7(c)] with output,

$$r = \frac{\langle \mathbf{f}, \mathbf{s} \rangle}{k + \langle (\mathbf{f} * \mathbf{s})^2, 1 \rangle}. \tag{30}$$

If internal noise is relatively small, we can log this expression without affecting our discussion (Sec. III B 1),

$$r = \log(\langle \mathbf{f}, \mathbf{s} \rangle) - \log\left(1 + \frac{\langle (\mathbf{f} * \mathbf{s})^2, 1 \rangle}{k}\right) + \log(k). \tag{31}$$

We have assumed $r > 0$ in Eq. (30) (necessary to log it). We can ignore the last additive term in Eq. (31) (it does not affect the final psychophysical response). If k is large compared to the normalizing term (its numerator), we can approximate the above expression to

$$r = \log(\langle \mathbf{f}, \mathbf{s} \rangle) - \frac{1}{k}\langle (\mathbf{f} * \mathbf{s})^2, 1 \rangle.$$

This is now the combined output of a Wiener LN system $\log(\langle \mathbf{f}, \mathbf{s} \rangle)$, where N is the log operation, and a Korenberg LNL system $-1/k\langle (\mathbf{f} * \mathbf{s})^2, 1 \rangle$, where N is the squaring operation [Fig. 7(d)]. We emphasize that several assumptions (e.g., small internal noise) and approximations were involved in this derivation, therefore its utility is primarily qualitative.

C. Real example 1: Spatial uncertainty approximated by Hammerstein cascade

This section reproduces results from a more extensive publication[51] on the topic of spatial uncertainty. Figure 8(a) shows a set of $\hat{\mathbf{H}}_1$ kernels for detecting a luminance bar appearing within a specified spatial range; the range could be very narrow (no spatial uncertainty), coded by the blue symbols, or very wide (large spatial uncertainty), coded by the red symbols. Cyan and magenta colors refer to intermediate
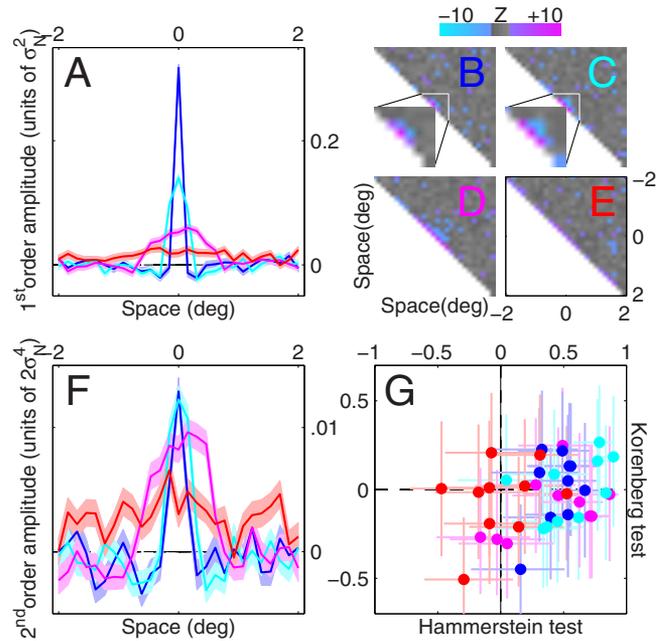

FIG. 8. (Color online) First-order and second-order kernels for visual detection under spatial uncertainty (Ref. 51) (data from >110K trials). (a) Aggregate $\hat{\mathbf{H}}_1$ kernels for four different uncertainty levels; uncertainty increases from blue to cyan, magenta, and red. [(b)–(e)] Aggregate second-order kernels for four different uncertainty levels, color-coded for $|Z| > 2$ (red for positive, blue for negative). For the two smaller uncertainty levels [(b) and (c)], the central regions of the kernels are magnified for ease of inspection. (f) Aggregate $\hat{\mathbf{H}}_2$ diagonals. (g) Correlation between $\hat{\mathbf{H}}_1$ kernel and $\hat{\mathbf{H}}_2$ diagonal (Hammerstein test) is plotted on the x axis vs correlation between $\hat{\mathbf{H}}_1$ and $\Sigma_k\hat{\mathbf{H}}_2(:,k)$ on the y axis for each observer separately. The former correlation tests for the applicability of a Hammerstein NL(N) cascade (Sec. IV B), the latter for the applicability of a Korenberg LNL(N) cascade (Sec. IV C); if the phenomenon of interest is well approximated by either cascade, the corresponding correlation is expected to differ from 0 (see Sec. VI C for details on these two tests). Error bars and shading show ±1 SEM.

uncertainty levels. As expected, the kernels spread over a progressively wider spatial range with increasing uncertainty [blue kernel peaks at target location (center of plot), red kernel spreads across x axis]. Figures 8(b)–8(e) show corresponding $\hat{\mathbf{H}}_2$ kernels; because we observed modulations primarily within the diagonal region, we replot $\hat{\mathbf{H}}_2$ diagonals in Fig. 8(f). It is apparent that $\hat{\mathbf{H}}_1$ kernels [Fig. 8(a)] and $\hat{\mathbf{H}}_2$ diagonals [Fig. 8(f)] share similar characteristics, a potential signature of the Hammerstein cascade (see below).

As a first check of whether the general strategy outlined in Fig. 7 was applicable to this data set, we carried out two standard tests in nonlinear system identification based on the following two results:[41,86] for a Hammerstein cascade $\mathbf{H}_1 \propto \text{diag}(\mathbf{H}_2)$; for a Korenberg cascade $\mathbf{H}_1 \propto \Sigma_k\mathbf{H}_2(:,k)$. We therefore expect a positive correlation between the first-order kernel and the diagonal of the second-order kernel for a Hammerstein cascade; failing that, we test for a positive correlation between the first-order kernel and the second-order marginal to check for compatibility with a Korenberg cascade. This strategy is a standard practice in neuroscience applications (e.g., Refs. 16 and 27) and it essentially relies

on the logic of Fig. 7, whereby the problem is simplified by approximating the biological system using a well-understood model with tractable nonlinear operators.

Figure 8(g) shows the result of both tests, Hammerstein on the x axis versus Korenberg on the y axis. With the only exception of the largest uncertainty level (red), all other kernels show significant correlations between $\hat{\mathbf{H}}_1$ and diag($\hat{\mathbf{H}}_2$) (blue, cyan, and magenta data points fall to the right of the vertical dashed line at $p < 0.01$), supporting the applicability of the Hammerstein model. Correlations between $\hat{\mathbf{H}}_1$ and $\Sigma_k \hat{\mathbf{H}}_2(:,k)$ (Korenberg test) were no different from 0 (at $p > 0.05$) for any of the four uncertainty levels (data points fall on the horizontal dashed line), indicating that modulations outside the diagonal region contained primarily noise (which eliminated the diagonal correlation). From this analysis, we conclude that the observed kernel structure is consistent with a Hammerstein model (Sec. IV B), with no clear evidence that this model needs further elaboration into a Korenberg cascade.[51]

D. Real example 2: Texture processing approximated by Korenberg cascade

We have already used this data set in Fig. 5 when discussing $\hat{\mathbf{H}}_1^{[1]}$ estimates and their relation to the system front-end filter (Sec. IV C). Figure 9(a) uses plotting conventions similar to Figs. 5(a) and 5(c), except both center and surround data are plotted in the same panel (black and blue traces, respectively) and red is used for the $\hat{\mathbf{H}}_2$ diagonal as opposed to $\hat{\mathbf{H}}_2(x_1,:)$ (yellow refers to data for the surround). As in the previous example, $\hat{\mathbf{H}}_1$ kernels and $\hat{\mathbf{H}}_2$ diagonals share similar characteristics; however, notice that there is a sign inversion for the surround kernels (compare blue trace with yellow trace), which is not observed for the center kernels (compare black with red). This sign inversion is not accommodated by a Hammerstein cascade with a single static nonlinearity (see Sec. IV B and Ref. 50 for details). Figure 9(b) demonstrates the sign inversion across observers: correlations between $\hat{\mathbf{H}}_1$ and diag($\hat{\mathbf{H}}_2$) [plotted on the x axis as in Fig. 8(g)] are significantly positive for center data and significantly negative for surround data ($p < 0.01$). A similar result is obtained for the Korenberg test (y axis), but in this case the underlying cascade is able to account for the sign inversion by applying a linear filter before the static nonlinearity with opposite sign for center and surround.[50] From this analysis, we conclude that the observed kernel structure is consistent with a Korenberg LNL(L) cascade; this result is in line with existing models of human texture processing, which are mostly of the LNL type,[7,32] also termed filter-rectify-filter models.[88]

E. Real example 3: Nonlinear dynamics on neuronal timescale

In a previous publication,[50] we documented the temporal dynamics of a simple visual process involving the detection of a luminance increment within a circular region surrounded by a large annulus. Figure 9(c) [plotted to the same conventions used for Fig. 9(a)] shows $\hat{\mathbf{H}}_1$ (cold colors) and $\hat{\mathbf{H}}_2$ di-
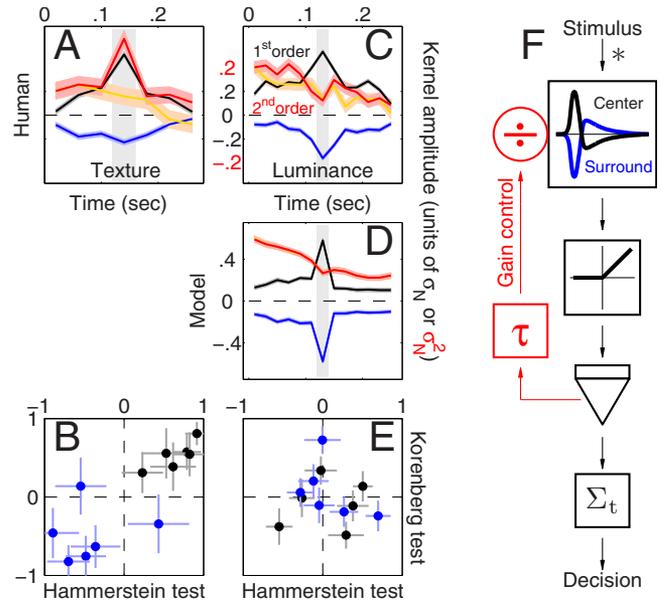

FIG. 9. (Color online) Temporal dynamics of texture (a) and luminance (c) processing (Ref. 50) (data from >56K trials). (a) shows aggregate (across observers) first-order (black/red) and second-order (red/yellow) kernels for both center and surround of an annular stimulus defined by texture modulations (see Ref. 50 for details). Temporal dynamics was captured over a range of $\sim\frac{1}{4}$ s (see x axis). (c) is plotted to the same conventions as (a) but shows data for a stimulus defined by luminance modulations. (d) Luminance kernels generated by the model in (f) (traces for center and surround second-order kernels are not visible separately because of overlapping). (b) and (e) plot the results from the Hammerstein test on the x axis vs the Korenberg test on the y axis (see caption to Fig. 8 and Sec. VI C) for texture and luminance, respectively. (f) shows the luminance model structure (constructed around knowledge from retinal circuitry): input stimulus is convolved ($*$) with center-surround receptive field (black and blue smooth traces) using first-order filter from primate retinal ganglion cells (Ref. 6). The output from this stage is half-wave rectified and fed onto an accumulator (inverted triangular symbol) implemented by differential equation (see Ref. 50 for details), then integrated across time (Σ_t) to yield scalar decision variable further converted into psychophysical decision. Time-varying signal from accumulator stage is carried back into circuit with temporal delay $\tau = 60$ ms to control gain of linear stage via divisive inhibition (refer to Sec. VI E for more details). Gain-control loop is highlighted in red. The model was used to simulate a psychophysical experiment adopting same parameters selected for human observers. Shaded regions in (a) and (c) show ± 1 SEM, in (d) ± 1 SD across repeated Monte Carlo simulations. Gray rectangular regions indicate the time of occurrence of the target signal.

agonal (warm colors) for both center (black/red) and surround (blue/yellow). As expected from the antagonistic nature of center-surround interactions in human vision,[77] the surround $\hat{\mathbf{H}}_1$ takes opposite polarity with respect to the center $\hat{\mathbf{H}}_1$ (compare blue and black traces). The $\hat{\mathbf{H}}_2$ diagonal is essentially identical for center and surround (compare red and yellow traces) and its time-course differs significantly from the $\hat{\mathbf{H}}_1$ profile, a feature that proved challenging for subsequent modeling efforts (see below).

As with the data sets discussed above, we carried out the two tests that were able to account for those previous results [see Figs. 8(g) and 9(b)]. However, as shown in Fig. 9(e), both tests failed: correlation coefficients were not statistically different from 0 (at $p > 0.05$) for either model (Hammerstein/Korenberg) or stimulus region (center/surround), as evident from the clustering of data points around 0 (intersection of

dashed lines). This lack of correlation is primarily a consequence of the different temporal dynamics we consistently observed between first- and second-order kernels [these differences were not observed for texture processing, see Fig. 9(a)].

Failing the two basic tests detailed above [Fig. 9(e)], there is no obvious way to proceed in terms of attempting an approximation using a simple block-model of the kind outline in Fig. 7. This is not to say that such an approximation is impossible: a sufficient number of Korenberg cascades can approximate virtually any nonlinear system of physiological interest via variants of the Wiener–Bose model[30,60,86] and methods have been developed to identify the relevant structure for specific applications (e.g., Refs. 13 and 63). However, it may then become more parsimonious to explore sensible and simpler models containing a plausible dynamic nonlinearity. Figure 9(f) shows an example from this class of models; its time-varying response $\mathbf{r}$ is determined by the following differential equation:

$$\frac{d\mathbf{r}}{dt} = \left[\frac{\mathbf{r}(t)}{b\mathbf{r}(t-\tau)+1} \right]_+,$$

where $\mathbf{r} = \mathbf{f} * \mathbf{s}_{ctr} - \mathbf{f} * \mathbf{s}_{sur}$, $\mathbf{f}$ is the front-end impulse response of a primate retinal ganglion cell [shown by the black trace in Fig. 9(f)], $\mathbf{s}_{ctr}$ is the center stimulus, $\mathbf{s}_{sur}$ the surround stimulus, $b = \frac{1}{5}$, $[\ ]_+$ is half-wave rectification, and $\tau = 60$ ms. For the purpose of generating a decision variable, we simply summed $\mathbf{r}$ across time ($r = \langle \mathbf{r}, 1 \rangle$). The critical component is the delayed gain-control feedback loop [red in Fig. 9(f)], which is able to generate dynamics similar to those observed experimentally [Fig. 9(d)]. The choice of a servo-mechanism was prompted primarily by the damped profile of the $\hat{\mathbf{H}}_2$ diagonal [red trace in Fig. 9(c)], which is common in nonlinear feedback systems,[39,87] and by the knowledge that gain control is a well-characterized phenomenon in both vertebrate[67,71] and invertebrate[84] retinas. Interestingly, our earlier work on the dynamics of directional processing in humans[53] required the introduction of a 90 ms temporal delay in the divisive feedback loop of the motion detector model, a figure comparable with the 60 ms used here.

F. Real example 4: Nonlinear dynamics on cognitive timescale

Visual adaptation is a complex phenomenon that operates on a variety of timescales,[29] typically much longer than those we considered in Secs. VI D and VI E, and provides a clear example of nonstatic behavior. If we look at a motionless object at a given time A, it will look motionless; if we then observe a moving pattern for a prolonged length of time (~ 1 min) and subsequently look back to the same motionless object at a given time B, it will appear to move in a direction opposite to that of the moving pattern we had seen previously. This motion after-effect is robust and can be easily experienced even in natural vision.[42] Because the visual input is the same at time A as it is at time B, and yet our visual system delivers an entirely different percept, the underlying phenomenon is clearly dynamic. However, the relevant dynamics is slow (order of several seconds) compared
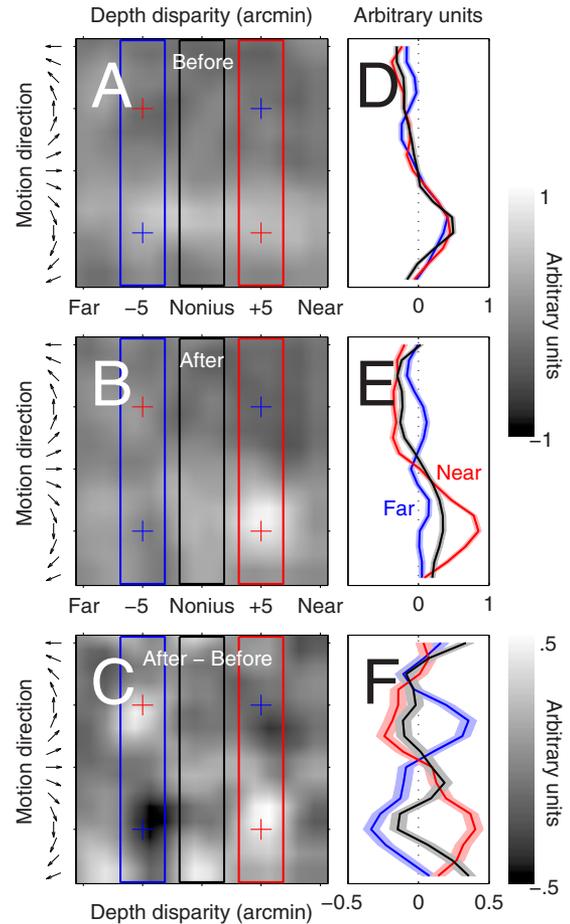

FIG. 10. (Color online) Effect of visual adaptation on first-order kernels for processing motion in depth (Ref. 54) (data from ~20K trials). [(a) and (b)] Smoothed (using Gaussian filter with SD equal to 1 surface pixel) $\hat{\mathbf{H}}_1$ kernels (average across observers) obtained before (a) and after (b) adaptation; (c) the differential filter [(b) minus (a)]. The stimulus consisted of moving elements presented at different depths; motion direction is plotted on the y axis, depth (as delivered via binocular disparity) on the x axis. Red crosses indicate signal locations for the target, blue crosses for the nontarget [see Ref. 54 (particularly Fig. 1) for details on the stimulus and experimental protocol]. (a) and (b) are plotted to the gray scale shown by the top right legend; (c) is plotted to the scale shown by the bottom right legend. [(d)–(f)] Slices along direction of motion averaged across the disparity range indicated by the corresponding dashed rectangular regions in (a)–(c) (red for near, blue for far, black for nonius fixation, i.e., the depth level corresponding to where the eyes fixate). Shaded regions show ±1 SEM. Units are arbitrary because averaging was performed after observer-by-observer normalization (Ref. 54).

to the dynamics we have been interested in so far (Sec. VI E); it is therefore reasonable to expect that we may be able to capture the effect of adaptation on kernel structure by probing the system with brief stimuli under different states of adaptation, a strategy that has been successfully exploited in physiology (e.g., Ref. 16).

Figure 10(a) shows an $\hat{\mathbf{H}}_1$ kernel defined across the dimensions of depth (stereoscopic disparity on the x axis) and motion direction (y axis); the reader is referred to Ref. 54 for details on the stimulus and experimental design. The target signal was delivered to the locations indicated by the red crosses; some kernel modulations colocalize with the signal, although only to a coarse extent [see Fig. 10(d) for cross-sections]. Figure 10(b) shows the same measurement per-

formed after prolonged adaptation to the target signal, and Fig. 10(c) shows the difference between Figs. 10(b) and 10(a). It is clear that adaptation induced specific changes within the kernels; these changes were meaningfully related to the target signal structure [Fig. 10(f)]. The characterization afforded by this approach is of course very limited: we only obtain two snapshots of the adaptation process, not a detailed time-course of its dynamics. However, each snapshot is remarkably detailed; this level of detail allows the experimenter to draw a number of significant conclusions about the underlying mechanism (see Ref. 54 for an extended treatment of this topic).

VII. CONCLUSIONS

We have examined a methodology for characterizing second-order functional approximations to a general-purpose sensory apparatus $\mathcal{H}$ embedded within realistic experimental constraints. From a general theoretical viewpoint, the two main constraints are those summarized in Fig. 1: a distortion of the input caused by the target signal $\mathbf{t}$ and a distortion of the output caused by the decisional transducer Ψ. Our goal is to estimate a first-order kernel $\mathbf{H}_1$ and a second-order kernel $\mathbf{H}_2$ that characterize $\mathcal{H}$ in the sense of a Volterra approximation [Eq. (4)]. These estimates can then be used to build and constrain processing algorithms operating at small scales (Sec. VI), i.e., via a few components connected in relatively simple ways.

For the first-order kernel estimate $\hat{\mathbf{H}}_1$, there are important differences between estimates obtained from noise fields containing the target ($\hat{\mathbf{H}}_1^{[1]}$) and those not containing the target ($\hat{\mathbf{H}}_1^{[0]}$), whenever this distinction is applicable (i.e., $\mathbf{t}^{[0]}=0$). In general, $\hat{\mathbf{H}}_1^{[0]}$ provides a less distorted estimate of $\hat{\mathbf{H}}_1$ [although still potentially distorted by higher-order kernels, see Eq. (22)]. There are instances, however, when $\hat{\mathbf{H}}_1^{[0]}$ is not particularly informative, e.g., uncertainty models (see Sec. IV C). In these instances, the front-end filter $\mathbf{f}$ may be estimated more effectively using $\hat{\mathbf{H}}_1^{[1]}$, with the caveat that this estimator reflects $\mathbf{f} \star \mathbf{f}$ not $\mathbf{f}$ (see Fig. 5 and Ref. 51).

For the second-order kernel estimate $\hat{\mathbf{H}}_2$, the distinction between target-present and target-absent subestimates is not relevant, because these do not allow reliable measurements of $\mathbf{H}_2$ when considered separately. We have shown in Sec. III A that it is possible to bypass Ψ for a benchmark linear system ($\mathbf{H}_d=0$ for $d>1$) to obtain $\hat{\mathbf{H}}_2=0$. This result ensures that any modulation within $\hat{\mathbf{H}}_2$ must reflect departures from linearity, and not simply artifacts introduced by Ψ; any estimator for $\mathbf{H}_2$ must satisfy this benchmark result. $\hat{\mathbf{H}}_2^{[1]}$ and $\hat{\mathbf{H}}_2^{[0]}$ do not, making them unsuitable for the application of interest. We have further demonstrated that the second-order estimator based on differential covariance $\hat{\mathbf{H}}_2$ is more robust than that based on differential second-order moments $\widetilde{\mathbf{H}}_2$ (see Sec. V B). A more effective estimation strategy for $\mathbf{H}_2$ is therefore achieved via differential covariance, as expressed in Eq. (6).

A major issue, which we have chosen to ignore in this article, is whether it is at all possible to reduce the treatment of animal sensory processing to a functional of the kind exemplified by $\mathcal{H}$ here. This is of course a critical issue for the applicability of the tools we have considered so far. Whether the framework adopted here is applicable or not will surely depend on the specific aspect of sensory processing that is being considered, on the animal system which implements it, and on the kinds of stimuli and measurements that are being carried out; ultimately, it is an empirical question that can only be answered effectively by transferring the theoretical concepts explored here to the laboratory.

ACKNOWLEDGMENTS

[1]Abbey, C. K. and Eckstein, M. P., "Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments," J. Vision **2**, 66–78 (2002).

[2]Afraz, S. R., Kiani, R., and Esteky, H., "Microstimulation of inferotemporal cortex influences face categorization," Nature (London) **442**, 692–695 (2006).

[3]Ahumada, A. J., "Classification image weights and internal noise level estimation," J. Vision **2**, 121–131 (2002).

[4]Ahumada, A. J., Marken, R., and Sandusky, A., "Time and frequency analyses of auditory signal detection," J. Opt. Soc. Am. A Opt. Image Sci. Vis **57**, 385–390 (1975).

[5]Barlow, H. B., Levick, W. R., and Yoon, M., "Responses to single quanta of light in retinal ganglion cells of the cat," Vision Res. **11**, 87–101 (1971).

[6]Benardete, E. A. and Kaplan, E., "The dynamics of primate M retinal ganglion cells," Visual Neurosci. **16**, 355–368 (1999).

[7]Bergen, J. R. and Landy, M. S., "Computational models of visual texture segregation," in *Computational Models of Visual Processing*, edited by M. Landy and J. Movshon (MIT, Cambridge, MA, 1991), pp. 253–271.

[8]Born, R. T. and Bradley, D. C., "Structure and function of visual area MT," Annu. Rev. Neurosci. **28**, 157–189 (2005).

[9]Borst, A., Flanagin, V. L., and Sompolinsky, H., "Adaptation without parameter change: Dynamic gain control in motion detection," Proc. Natl. Acad. Sci. U.S.A. **102**, 6172–6176 (2005).

[10]Burgess, A. E. and Colborne, B., "Visual signal detection. IV. Observer inconsistency," J. Opt. Soc. Am. A **5**, 617–627 (1988).

[11]Busse, L., Katzner, S., Tillmann, C., and Treue, S., "Effects of attention on perceptual direction tuning curves in the human visual system," J. Vision **8**, 1–13 (2008).

[12]Bussgang, J. J., "Cross-correlation functions of amplitude-distorted Gaussian signals," MIT Research Laboratory Electricity Technical Report No. 216, August 1952.

[13]Citron, M. C. and Marmarelis, V. Z., "Applications of minimum-order Wiener modeling to retinal ganglion cell spatiotemporal dynamics," Biol. Cybern. **57**, 241–247 (1987).

[14]Falmagne, J.-C., *Elements of Psychophysical Theory* (Oxford University Press, New York, 1985).

[15]Franz, M. O. and Schölkopf, B., "A unifying view of Wiener and Volterra theory and polynomial kernel regression," Neural Comput. **18**, 3097–3118 (2006).

[16]French, A. S., Korenberg, M. J., Jarvilehto, M., Kouvalainen, E., Juusola, M., and Weckstrom, M., "The dynamic nonlinear behavior of fly photoreceptors evoked by a wide range of light intensities," Biophys. J. **65**, 832–839 (1993).

[17]Geisler, W. S., "Ideal observer theory in psychophysics and physiology," Phys. Scr. **39**, 153–160 (1989).

[18]Gold, J. I. and Shadlen, M. N., "The neural basis of decision making," Annu. Rev. Neurosci. **30**, 535–574 (2007).

[19]Gosselin, F. and Schyns, P. G., "Superstitious perceptions reveal properties of internal representations," Psychol. Sci. **14**, 505–509 (2003).

[20]Green, D. M. and Swets, J. A., *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).

[21]Gregson, R. A. and Britton, L. A., "The size-weight illusion in 2-D nonlinear psychophysics," Percept. Psychophys. **48**, 343–356 (1990).

[22] Guastello, S. J., "Nonlinear dynamics in psychology," Discrete Dyn. Nat. Soc. **6**, 11–29 (2001).

[23] Heath, R. A., *Nonlinear Dynamics: Techniques and Applications in Psychology* (Lawrence Erlbaum Associates, New Jersey, 2000).

[24] Heath, R. A. and Fulham, R., "Applications of system identification and adaptive filtering techniques in human information processing," in *Cognition, Information Processing and Motivation*, edited by G. d'Ydewalle (Elsevier Science, Amsterdam, 1985), pp. 117–147.

[25] Heeger, D. J., Simoncelli, E. P., and Movshon, J. A., "Computational models of cortical visual processing," Proc. Natl. Acad. Sci. U.S.A. **93**, 623–627 (1996).

[26] Hunter, I. W. and Korenberg, M. J., "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models," Biol. Cybern. **55**, 135–144 (1986).

[27] James, A. C., "Nonlinear operator network models of processing in the fly lamina," in *Nonlinear Vision*, edited by R. B. Pinter and B. Nabet (CRC, Boca Raton, FL, 1992), pp. 39–73.

[28] Kelly, A., Heathcote, A., Heath, R., and Longstaff, M., "Response-time dynamics: Evidence for linear and low-dimensional nonlinear structure in human choice sequences," Q. J. Exp. Psychol. A **54**, 805–840 (2001).

[29] Kohn, A., "Visual adaptation: Physiology, mechanisms, and functional benefits," J. Neurophysiol. **97**, 3155–3164 (2007).

[30] Korenberg, M. J., "Parallel cascade identification and kernel estimation for nonlinear systems," Ann. Biomed. Eng. **19**, 429–455 (1991).

[31] Korenberg, M. J. and Hunter, I. W., "The identification of nonlinear biological systems: LNL cascade models," Biol. Cybern. **55**, 125–134 (1986).

[32] Landy, M. S. and Graham, N., "Visual perception of texture," in *The Visual Neurosciences*, edited by L. M. Chalupa and J. S. Werner (MIT, Cambridge, MA, 2004), pp. 1106–1118.

[33] Lasley, D. J. and Cohn, T. E., "Why luminance discrimination may be better than detection," Vision Res. **21**, 273–278 (1981).

[34] Lennie, P., "Single units and visual cortical organization," Perception **27**, 889–935 (1998).

[35] Luce, R. D., "Four tensions concerning mathematical modeling in psychology," Annu. Rev. Psychol. **46**, 1–27 (1995).

[36] Luce, R. D., *Response Times: Their Role in Inferring Elementary Mental Organization* (Oxford University Press, New York, 1986).

[37] Maljkovic, V. and Martini, P., "Implicit short-term memory and event frequency effects in visual search," Vision Res. **45**, 2831–2846 (2005).

[38] Maloney, L. T. and Zhang, H., "Decision-theoretic models of visual perception and action," Vision Res. **50**, 2362–2374 (2010).

[39] Marmarelis, P. Z., "Identification of nonlinear biological systems using Laguerre expansions of kernels," Ann. Biomed. Eng. **21**, 573–589 (1993).

[40] Marmarelis, P. Z. and Marmarelis, V. Z., *Analysis of Physiological Systems: The White-Noise Approach* (Plenum, New York, 1978).

[41] Marmarelis, V. Z., *Nonlinear Dynamic Modeling of Physiological Systems* (Wiley IEEE, Piscataway, NJ, 2004).

[42] Mather, G., Pavan, A., Campana, G., and Casco, C., "The motion aftereffect reloaded," Trends Cogn. Sci. **12**, 481–487, 2008.

[43] Metzger, M. A. and Theisz, M. F., "Forecast: Program to obtain forecasts from subjects for successive values of chaotic time series," Behav. Res. Methods Instrum. Comput. **26**, 387–394 (1994).

[44] Mountcastle, V. B., LaMotte, R. H., and Carli, G., "Detection thresholds for stimuli in humans and monkeys: Comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand," J. Neurophysiol. **35**, 122–136 (1972).

[45] Murdock, B. B., *Human Memory: Theory and Data* (Lawrence Erlbaum Associates, MD, 1974).

[46] Murray, R. F., Bennett, P. J., and Sekuler, A. B., "Optimal methods for calculating classification images: Weighted sums," J. Vision **2**, 79–104 (2002).

[47] Neri, P., "Estimation of nonlinear psychophysical kernels," J. Vision **4**, 82–91 (2004).

[48] Neri, P., "Fast-scale adaptive changes of directional tuning in fly tangential cells are explained by a static nonlinearity," J. Exp. Biol. **210**, 3199–3208 (2007).

[49] Neri, P., "How inherently noisy is human sensory processing?," Psychon. Bull. Rev. **17**, 802–808 (2010).

[50] Neri, P., "Nonlinear characterization of a simple process in human vision," J. Vision **9**, 1–29 (2009).

[51] Neri, P., "Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity," Front. Comput. Neurosci. **4**, 151 (2010).

[52] Neri, P. and Heeger, D. J., "Spatiotemporal mechanisms for detecting and identifying image features in human vision," Nat. Neurosci. **5**, 812–816 (2002).

[53] Neri, P. and Levi, D., "Temporal dynamics of directional selectivity in human vision," J. Vision **8**, 1–11 (2008).

[54] Neri, P. and Levi, D. M., "Evidence for joint encoding of motion and disparity in human visual perception," J. Neurophysiol. **100**, 3117–3133 (2008).

[55] Neri, P. and Levi, D. M., "Receptive versus perceptive fields from the reverse-correlation viewpoint," Vision Res. **46**, 2465–2474 (2006).

[56] Neuringer, A. and Voss, C., "Approximating chaotic behavior," Psychol. Sci. **4**, 113–119 (1993).

[57] Newsome, W. T., "Visual attention: Spotlights, highlights and visual awareness," Curr. Biol. **6**, 357–360 (1996).

[58] Newsome, W. T., Britten, K. H., and Movshon, J. A., "Neuronal correlates of a perceptual decision," Nature (London) **341**, 52–54 (1989).

[59] Newsome, W. T., Shadlen, M. N., Zohary, E., and Britten, K. H., "Visual motion: Linking neuronal activity to psychophysical performance," in *The Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge, 1995), pp. 401–414.

[60] Palm, G., "On representation and approximation of nonlinear systems. Part II: Discrete time," Biol. Cybern. **34**, 49–52 (1979).

[61] Paninski, L., "Convergence properties of three spike-triggered analysis techniques," Network **14**, 437–464 (2003).

[62] Parker, A. J. and Newsome, W. T., "Sense and the single neuron: Probing the physiology of perception," Annu. Rev. Neurosci. **21**, 227–277 (1998).

[63] Paulin, M. G., "A method for constructing data-based models of spiking neurons using a dynamic linear-static nonlinear cascade," Biol. Cybern. **69**, 67–76 (1993).

[64] Pelli, D. G., "Noise in the visual system may be early," in *Computational Models of Visual Processing*, edited by M. Landy and A. J. Movshon (MIT, Cambridge, 1991), pp. 147–152.

[65] Pelli, D. G., "Uncertainty explains many aspects of visual contrast detection and discrimination," J. Opt. Soc. Am. A **2**, 1508–1532 (1985).

[66] Ringach, D. L., "Tuning of orientation detectors in human vision," Vision Res. **38**, 963–972 (1998).

[67] Sakai, H. M., Wang, J. L., and Naka, K., "Contrast gain control in the lower vertebrate retinas," J. Gen. Physiol. **105**, 815–835 (1995).

[68] Schetzen, M., *The Volterra and Wiener Theories of Nonlinear Systems* (Wiley, New York, 1980).

[69] Schneider, K. A. and Komlos, M., "Attention biases decisions but does not alter appearance," J. Vision **8**, 1–10 (2008).

[70] Schölkopf, B. and Smola, A. J., *Learning with Kernels* (MIT, Cambridge, MA, 2002).

[71] Shapley, R. and Enroth-Cugell, C., "Visual adaptation and retinal gain controls," Prog. Retinal Res. **3**, 263–346 (1984).

[72] Shub, D. E. and Richards, V. M., "Psychophysical spectro-temporal receptive fields in an auditory task," Hear. Res. **251**, 1–9 (2009).

[73] Simpson, W. A., Braun, W. J., Bargen, C., and Newman, A. J., "Identification of the eye-brain-hand system with point processes: A new approach to simple reaction time," J. Exp. Psychol. Hum. Percept. Perform. **26**, 1675–1690 (2000).

[74] Smithson, M., "Judgment under chaos," Org. Behav. Hum. Decis. Process **69**, 58–66 (1997).

[75] Solomon, J. A., "Noise reveals visual mechanisms of detection and discrimination," J. Vision **2**, 105–120 (2002).

[76] Spekreijse, H. and Oosting, H., "Linearizing: A method for analysing and synthesizing nonlinear systems," Kybernetik **7**, 22–31 (1970).

[77] Spillmann, L., "Foveal perceptive fields in the human visual system measured with simultaneous contrast in grids and bars," Pfluegers Arch. **326**, 281–299 (1971).

[78] Stewart, I. N. and Peregoy, P. L., "Catastrophe theory modeling in psychology," Psychol. Bull. **94**, 336–362 (1983).

[79] Ta'edd, L. K., Ta'eed, O., and Wright, J. E., "Determinants involved in the perception of the Necker cube: An application of catastrophe theory," Behav. Sci. **33**, 97–115 (1988).

[80] Tanner, W. P., "Physiological implications of psychophysical data," Ann. N.Y. Acad. Sci. **89**, 752–765 (1961).

[81] Thelen, E. and Smith, L. B., *H Dynamic Systems Approach to the Development of Cognition and Action* (MIT, Cambridge, MA, 1995).

[82] Thomas, J. P. and Knoblauch, K., "Frequency and phase contributions to the detection of temporal luminance modulation," J. Opt. Soc. Am. A Opt. Image Sci. Vis **22**, 2257–2261 (2005).

[83]Tjan, B. S. and Nandy, A. S., "Classification images with uncertainty," J. Vision **6**, 387–413 (2006).

[84]van Hateren, J. H. and Snippe, H. P., "Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells," Vision Res. **41**, 1851–1865 (2001).

[85]von der Malsburg, C., "The what and why of binding: The modeler's perspective," Neuron **24**, 95–104 (1999).

[86]Westwick, D. T. and Kearney, R. E., *Identification of Nonlinear Physiological Systems* (Wiley IEEE, Piscataway, NJ, 2003).

[87]Wiener, N., *Cybernetics or Control and Communication in the Animal and the Machine* (Wiley, New York, 1948).

[88]Wilson, H. R., "Non-Fourier cortical processes in texture, form, and motion perception," in *Cerebral Cortex*, edited by P. S. Ulinski, E. G. Jones, and A. Peters (Plenum, New York, 1999), Vol. 14, pp. 445–477.

[89]Yeshurun, Y., Carrasco, M., and Maloney, L. T., "Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model," Vision Res. **48**, 1837–1851 (2008).