Estimation of nonlinear psychophysical kernels

Peter Neri

Department of Zoology, University of Cambridge,
England, UK   ✉

Reverse correlation techniques have been extensively used in physiology (Marmarelis & Marmarelis 1978; Sakai, Naka, & Korenberg, 1988), allowing characterization of both linear and nonlinear aspects of neuronal processing (e.g., Emerson, Bergen, & Adelson, 1992; Emerson & Citron 1992). Over the past decades, Ahumada (1996) developed a psychophysical reverse correlation technique, termed noise image classification (NIC), for deriving the linear properties of sensory filters in the context of audition first (Ahumada, 1967; Ahumada, Marken, & Sandusky, 1975), and then vision (Ahumada, 1996, 2002; Beard & Ahumada, 1998). This work explores ways of characterizing nonlinear aspects of psychophysical filters. One approach consists of an extension of the NIC technique (ExtNIC), whereby second-order (rather than just first-order) statistics in the classified noise are used to derive sensory kernels. It is shown that, under some conditions, this procedure yields a good estimate of second-order kernels. A second, different approach is also considered. This method uses functional minimization (fMin) to generate kernels that best simulate psychophysical responses for a given set of stimuli. Advantages and disadvantages of the two approaches are discussed. A mathematical appendix shows some interesting facts: (1) that nonlinearities affect the linear estimate (particularly target-present averages) obtained from the NIC method, providing a rationale for some related observations made by Ahumada (1967); (2) that for a linear filter followed by a static nonlinearity (LN system), the ExtNIC estimate of the second-order nonlinear kernel is correctly null, provided the criterion is unbiased; (3) that for a biased criterion, such an estimate may contain predictable modulations related to the linear filter; and (4) that under certain assumptions and conditions, ExtNIC does return a correct estimate for the second-order nonlinear kernel.

Keywords: reverse correlation, triggered correlation, recursive least squares, linear-nonlinear model, system identification

Introduction

For the purpose of this work, psychophysical observers are conceived of as functionals that map an input function I, such as a visual stimulus, into a binary number:

$$o = thr_c \left(F \left(I(x_1, \ldots, x_n) \right) \right) \qquad (1)$$

where o is 0 for "no" and 1 for "yes," I is an input function defined over an n-dimensional space (e.g., x, y, t), F is a functional mapping I into a real number, and $thr_c(x)$ is a function that returns 0 when $x < c$, and 1 when $x \geq c$. This work considers only yes/no tasks (a brief explanation of how the considerations made here can be extended to two-alternative forced-choice [2AFC] tasks is given in the next section). To simplify notation, I is defined over one dimension only (x).

The functional F will, in general, be nonlinear. One way of representing a nonlinear system such as this one is by Volterra expansion, where:

$$V_i[\mathbf{L}_i(v_1, v_2, \ldots, v_i), I] =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{L}_i(v_1, \ldots, v_i) I(v_1) I(v_2) \ldots I(v_i) dv_1 dv_2 \ldots dv_i \qquad (2)$$

The objects $\mathbf{L}_n$ are the system's kernels, and V_n are outcomes of filtering stimulus I with these kernels.[1] Volterra kernels are intrinsically symmetric. Volterra showed that this representation applies to nonlinear systems that are time-invariant, with finite memory, and analytic (Volterra, 1959).

For a linear system (i.e., $F = V_1$), Ahumada's noise image classification (NIC) technique (Beard & Ahumada, 1998) returns the linear kernel $\mathbf{L}_1$ up to a multiplicative factor (Ahumada, 2002, and "Appendix").

When the system is nonlinear, one needs to determine higher-order kernels ($\mathbf{L}_2$, $\mathbf{L}_3$,..., $\mathbf{L}_n$) to characterize it. This work will focus only on second-order kernels, but can be easily extended to higher orders.

The second-order Volterra kernel L_2 dictates how pairs of input pulses interact in the system, that is whether pairs of inputs need to covary positively or negatively (or not at all) in order to drive a positive response from the system. The natural extension of the NIC technique to deriving this nonlinear kernel is to compute, instead of the first-order statistics associated with the classified noise, its second-order statistics (e.g., Marmarelis & Marmarelis, 1978), in terms of second-order moments or covariance matrices. In Section 2 and in the "Appendix," it is shown that this method does return a good estimate of psychophysical second-order kernels.

A different approach to nonlinear system characterization involves functional minimization (fMin). Section 3 describes in more detail how this method works. In short, this approach allows determination of the system's kernels by minimizing, for a given input sequence I_i (where i refers to the ith trial), the difference between the experimentally determined output sequence o_i^{exp} and one (o_i) computed using an equation similar to (1). The minimization is first

carried out for a linear system, determining the linear kernel W_1 that accounts for most of the output. The system in Equation 1 is then upgraded to second-order, and the best second-order kernel W_2 is similarly determined. In general (i.e., when input is arbitrary), every time the system is upgraded to a higher order, one has to correct for the fit deriving from lower orders (this is explained in Section 3).

One advantage of the fMin approach is that it has no specific requirements on input characteristics. The ExtNIC method works only with specific types of input (e.g., Gaussian white noise; see "Appendix" for details). However, fMin involves a minimization step that can be prohibitive in most practical applications, whereas ExtNIC is, under certain conditions, a simple and robust estimator.

2. Derivation of second-order kernels by computing second-order statistics (ExtNIC)

For the purpose of this section, all the investigator needs from a psychophysical experiment is four sets of noise images associated with the four stimulus-response classes (hits, false alarms, misses, and correct rejections). The type of noise used has to satisfy certain requirements (e.g., orthogonality). These requirements are listed in the "Appendix"; Gaussian white noise, for example, satisfies them all. $I_i^{[s,o]}$ is the ith stimulus image associated with stimulus s (where $s = 0$ is noise-only, $s = 1$ is target+noise) and response o.

Ahumada's estimate $\mathbf{L}_1^{est}$ of the system's linear kernel $\mathbf{L}_1$ is (Ahumada, 1996):

$$L_1^{est}(\nu) = L_1^{[1,1]}(\nu) + L_1^{[0,1]}(\nu) - L_1^{[1,0]}(\nu) - L_1^{[0,0]}(\nu) \tag{3}$$

where $L_1^{[s,o]} = E(I^{[s,o]}(\nu))$ ($E(x)$ being the expectation of x across trials). The "Appendix" provides a formal derivation of how this estimate (target-present averages in particular) is affected by a second-order kernel $\mathbf{L}_2$.

Similarly, what is proposed here is that one can estimate the system's second-order Volterra kernel, $\mathbf{L}_2$, by computing:

$$L_2^{est}(\nu,\xi) = L_2^{[1,1]}(\nu,\xi) + L_2^{[0,1]}(\nu,\xi)$$
$$\qquad\qquad - L_2^{[1,0]}(\nu,\xi) - L_2^{[0,0]}(\nu,\xi) \tag{4}$$

or, similarly,

$$L_2^{est}(\nu,\xi) = \mathrm{cov}^{[1,1]}(\nu,\xi) + \mathrm{cov}^{[0,1]}(\nu,\xi)$$
$$\qquad\qquad - \mathrm{cov}^{[1,0]}(\nu,\xi) - \mathrm{cov}^{[0,0]}(\nu,\xi) \tag{5}$$

where

$$L_2^{[s,o]}(\nu,\xi) = E\left(I^{[s,o]}(\nu) I^{[s,o]}(\xi)\right).$$

One of the main results in the "Appendix" (Equation 17) is that for a linear system followed by a static nonlinearity (LN system) with odd-symmetric decisional transducer at unbiased criterion, L_2^{est} is correctly null.

There are a few reasons for using covariance (Equation 5) rather than second-order moments (Equation 4); one reason is that for an LN system, $\mathbf{L}_1$ can create spurious modulations in L_2^{est} when the criterion is biased, and these modulations are of the type $\mathbf{L}_1(\nu) \cdot \mathbf{L}_1(\xi)$ (see "Appendix"). Covariance partly compensates for these effects by subtracting a similar term (this relates to the difference between solid squares and dashed line in the top right panel of Figure 3b; see Section 4).

Another reason is that for an LN system, filtering always reduces variance compared to baseline noise variance; this happens because the noise distribution is truncated at locations where the filter is applied (Ahumada, 2002). Inspection of the $L_2^{[s,o]}$ images for the different stimulus-response classes is, therefore, informative as to whether the final modulations observed in L_2^{est} are possibly due to LN filtering or not. For example, a positive modulation (with respect to baseline noise variance) along the diagonal in an individual response class (in cov$^{[s,o]}$) cannot be due to LN filtering, and would require further investigation.

Other important diagnostic tools involve identifying sizeable covariance modulations in L_2^{est} that are not localized at the level of $\mathbf{L}_1(\nu) \cdot \mathbf{L}_1(\xi)$, as those too cannot be due to repercussions from LN filtering. In general, it is important that the experimenter examines $L_2^{[s,o]}$'s and $L_1^{[s,o]}$'s carefully and decides whether, in the specific context of the experiment being carried out, these are informative.

Figure 1 provides an example that illustrates the outcome of this procedure. A psychophysical experiment is simulated for a system with known, randomly generated Volterra kernels (shown in a and d); this system maps input noise images into binary responses according to Equation 1 for a fixed, unbiased c. Noise image intensities were uniformly distributed around 0 (spanning –1 to 1), the size of images was 7, and the target was the vector [0 0 0 1 0 0 0] (added on half trials). A total of 5,000 trials were run.

In b, the function computed using Equation 3 is shown, and it can be seen that it returns the linear kernel quite well (it would be optimal if the system were linear). In e, the function obtained by computing covariance (Equation 5) is plotted, and it can be seen that this also provides a fairly good estimate of the second-order kernel in d. c and f plot real versus estimated values for the two kernels; in this example, correlation values were 0.93 and 0.96 (for first- and second-order kernels).

As mentioned in the "Introduction," this work focuses on yes/no tasks. The most trivial extension of these methods to 2AFC tasks is, for example, to classify each noise image in each interval separately, taking the response from the observer as a double statement on both intervals (e.g., if the target is in interval 1 and the observer responds "interval 2," this statement is taken as "interval 2, not interval 1," and the noise image in interval 1 is classified as a miss, the
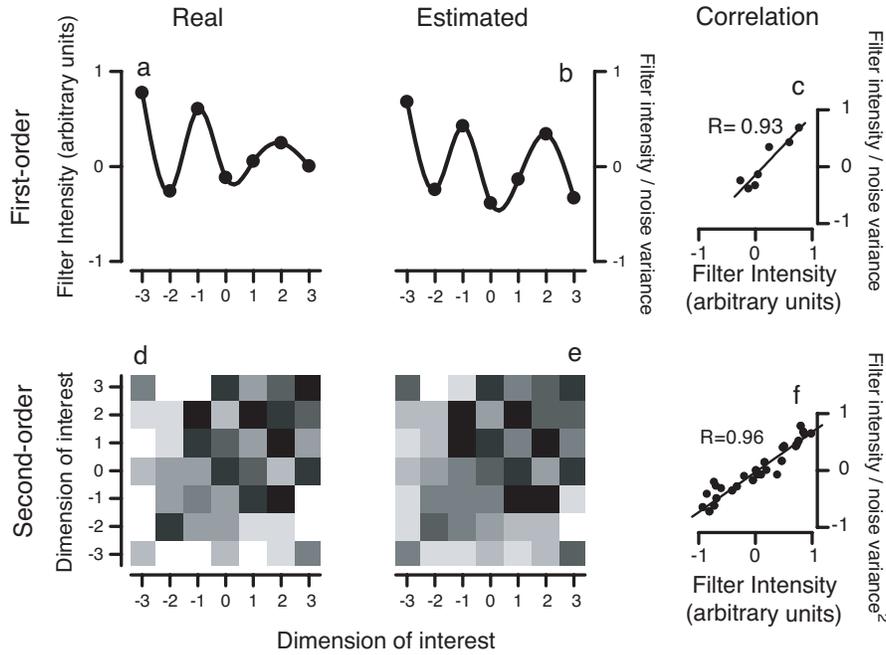
Figure 1. Outcome of a simulation involving kernel derivation as described in Section 2. a and d are two randomly generated kernels (linear and second-order nonlinear, respectively); b and e are the corresponding estimates. c and f show correlations between real and estimate: c plots real values (those in a) versus estimated values (those in b) for the linear kernel; f plots real values (those in d) versus estimated values (those in e) for the second-order nonlinear kernel.

noise image in interval 2 as a false alarm). This approach may be problematic in some experimental contexts, as it relies on the assumption that Equation 1 can be applied separately to each interval in a 2AFC task. For more details on how the NIC approach extends to AFC tasks, the reader should refer to Abbey and Eckstein (2002).

3. Derivation of linear and nonlinear kernels using functional minimization (fMin)

For this section, what we need from a psychophysical experiment is a full description of input and output, that is, a sequence of input images (including the target) and a sequence of binary responses.

There are no special requirements on the input, as long as it can be adequately described for use in Equation 1. I_i is the ith input image, and o_i^{exp} the ith response from the observer (0 for "no," 1 for "yes"). The fMin method works by computing the best linear kernel first, using a Volterra representation of the system that extends only up to the linear order:

$$o_i = thr_c \left(F\left(I_i(x)\right)\right) = thr_c \left(V_1\left[\mathbf{L}_1(\nu), I_i\right]\right)$$

$$= thr_c \left(\int_{-\infty}^{\infty} \mathbf{L}_1(\nu) I_i(\nu) d\nu \right) \qquad (6)$$

The estimated first-order kernel for the system, W_1, is the L_1 that minimizes the difference between o_i^{exp} and o_i:

$$W_1 \Rightarrow \min_{NoTrialsPerBlock}\left(D(o_i^{exp} - o_i)\right) \qquad (7)$$

where $min_{NoTrialsPerBlock}$ is the best minimization obtained by allowing criterion c in Equation 6 to vary every *NoTrialsPerBlock* number of trials, and $D(x_i, y_i)$ for i ranging from 1 to n (and x and y binary) is $\sum_{i=1}^{n}(x_i \neq y_i)$, where $(x_i \neq y_i)$ returns 1 if $x_i \neq y_i$ and 0 if $x_i = y_i$ (the choice of criterion variation and distance measure D offered here are not the only ones that are possible; e.g., one may allow c to vary according to a prespecified probability distribution without optimizing its value every *NoTrialsPerBlock* number of trials, and D could be chosen to be the sum of square differences).

The extent to which o_i matches o_i^{exp} will be affected by the choice of c in Equation 6. It seems reasonable that c should be allowed to vary, as is the case in the psychophysics (over a large number of trials, it is expected that observers' criteria fluctuate a bit). How often this should happen will depend on the specifics of the experiment. c should not be allowed to vary too often: in the extreme case of *NoTrialsPerBlock* = 1, minimizing Equation 6 becomes meaningless, as on each trial i there will always be a value for c that returns $o_i = o_i^{exp}$. A reasonable value for *NoTrialsPerBlock* could be, for example, 200. This means that the

minimization in Equation 7 would be carried out computing the threshold step in Equation 6 for an optimal choice of c every 200 trials.

The next step involves a similar procedure, applied to the second-order nonlinearity. The system is now represented as follows:

$$o_i = thr_c\left(F\left(I_i(x)\right)\right)$$

$$= thr_c\left(V_1\left[W_1(\nu), I_i\right] + V_1\left[\mathbf{L}_1^{adj}(\nu), I_i\right] + V_2\left[\mathbf{L}_2(\nu_1, \nu_2), I_i\right]\right)$$

The best approximating linear kernel W_1 has already been computed; what needs to be determined are an adjusting linear kernel W_1^{adj}, and a second-order kernel W_2 (Victor, 1992): these are obtained adopting the same minimization procedure as before. The best approximating second-order kernel W_2 is the second-order kernel L_2 that minimizes the difference between o_i^{exp} and o_i:

$$[W_1^{adj}, W_2] \Rightarrow \min_{NoTrialsPerBlock}\left(D(o_i^{exp} - o_i)\right).$$

In the process of minimizing for $\mathbf{L}_2$, one also has to minimize for L_1^{adj}, obtaining both W_2 and W_1^{adj}. The best approximating first-order kernel, however, is still the one computed before, W_1, not the correcting one (W_1^{adj}) introduced here (see below for an explanation).

The best way to understand this formulation is to think of it in relation to polynomial approximation. The problem analyzed in this section is similar to an attempt to approximate an unknown function $y = f(x)$ within a specified interval $[a,b]$ using a polynomial expansion of x, $f^*(x) = \sum_{n=0}^{\infty} a_n x^n$; say $f(x) = x^2$, and the interval [0,1]. If we approximate this with a zeroth-order expansion $f^* = a_0$, the best a_0 (in the least mean square difference sense) is $w_0 = 1/3$. At the first-order, $f^*(x) = a_0 + a_1 x$, we have that the best approximation is for $a_0 = -1/6$, and $a_1 = 1$. We write this as $f^*(x) = w_0 + w_{0,1}^{adj} + w_1 x$, with $w_{0,1}^{adj} = -1/2$ being the first-order-related correction for the zeroth order term (so that $w_0 + w_{0,1}^{adj} = -1/6$), and $w_1 = 1$. In other words, we need to adjust our best estimate for the zeroth order term as we introduce a linear term. At the second order, we can achieve perfect fit using $f^*(x) = x^2$, so we need to cancel out both zeroth- and first-order terms: $f^*(x) = w_0 + w_{0,1}^{adj} + w_{0,2}^{adj} + (w_1 + w_{1,2}^{adj}) + x + w_2 x^2$, with $w_{0,2}^{adj} = 1/6$, $w_{1,2}^{adj} = -w_1$, and $w_2 = 1$. From here, the reader can see that w_n is the best approximating coefficient of order n for the nth order approximation. This is also what the best approximating kernels W_n's are. The reader is referred to Victor,1992, where this method is presented as an abstract formulation of the Wiener approach to nonlinear system representation.

If the interval $[a,b]$ is symmetric ($a = -b$), then $w_{1,2}^{adj} = 0$, that is, there is no linear correction term when upgrading to second-order, as first and second order terms are orthogonal (i.e., $\int_{-\nu}^{\nu} a_1 x \cdot a_2 x^2 dx = 0$). In the examples considered in this work, input noise is indeed symmetric, so $W_1^{adj} = 0$ as in Figure 4. However, the formulation has been kept general to include cases in which the input is not constrained (e.g., one can think of applications in which the input consists of a limited set of images from medical reports).

Figure 2 provides an example of the outcome of this procedure (with *NoTrialsPerBlock* = 200). A psychophysical experiment was simulated as in the previous section, for a system of known Volterra kernels (shown in a and d). b shows W_1 for this system, and e plots W_2 (minimizations were carried out using standard Matlab routines); the thin trace in b shows W_1^{adj}. Both kernel estimates are quite good (as in the previous section, correlations for the two kernels [shown in c and f] are high). It must be pointed out that the validity of this procedure is independent of the particular structure that was selected for the nonlinear system used in these simulations (i.e., the system does not have to be literally implementing a Volterra expansion; this was done here only for illustrative purposes).

4. Brief quantitative comparison between these two approaches

Figure 3a plots correlations for the two kernels, such as those shown in Figures 1c and 1f and 2c and 2f (correlations between real and estimated kernel values), one (linear kernel, abscissa) against the other (second-order nonlinear kernel, ordinate). The two curves are for the two different methods described in the previous sections (squares: ExtNIC, Section 2; circles: fMin, Section 3). Moving along each curve following the arrow, the system goes from being very nonlinear (low [0.1] ratio between first- and second-order kernel amplitudes) to being very linear (high [10] ratio). Kernel estimates depend on the linearity/nonlinearity of the system: for a highly linear system (bottom right corner), estimates are best for linear kernels, but they get worse as the system becomes more nonlinear, and it is now estimates of the nonlinear kernels that improve (top left). This is, of course, what is expected. When the system has roughly equivalent linear and nonlinear components (top right), both kernel estimates are reasonably good. Overall, the two approaches perform very similarly (each open symbol refers to one estimate for one randomly generated system, each solid symbol to the average of 50 estimates for 50 different randomly generated systems).

Figure 3b plots more correlation values for both approaches (top row, ExtNIC; bottom row, fMin) and both kernels (open for first-order, solid for second-order) at three different values of linearity/nonlinearity ratio for the sys-
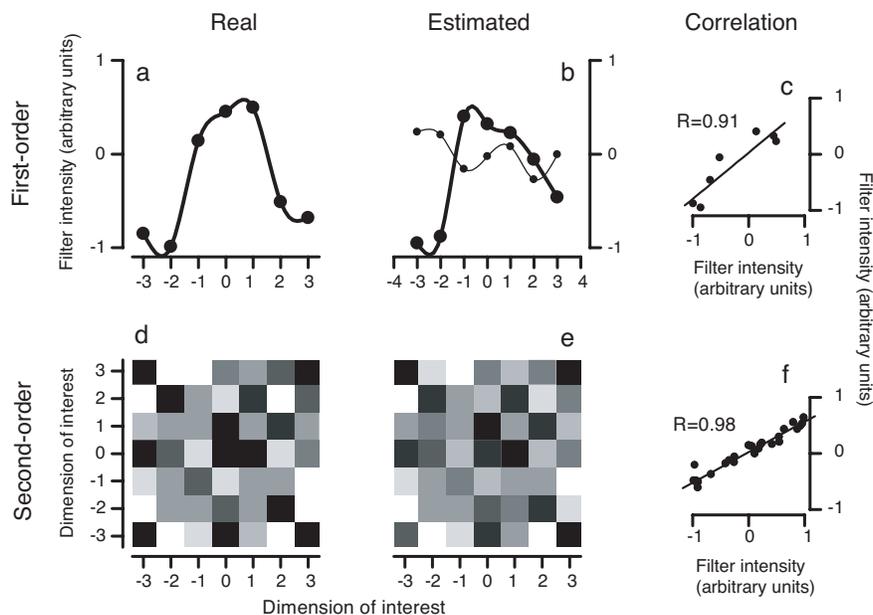
Figure 2. This figure is equivalent to Figure 1, except here kernels are estimated using the method described in Section 3 (functional minimization), rather than Section 2 (extended noise image classification). The thin line in panel b shows the adjusting linear kernel W_1^{adj} – this is expected to be 0 for the input used in this simulation.

tem (0.1 left, 1 middle, 10 right panels), as a function of criterion bias (x axis). Criterion bias is in units of half the difference between the mean response to signal+noise and that to noise alone; 0 is for a criterion that is halfway between the mean responses. Values greater than 1 and smaller than –1 refer to very conservative and very lax criteria, respectively. Again, the two approaches overall perform very similarly. The main difference is that for a very nonlinear system, fMin recovers the linear filter slightly better than ExtNIC at unbiased criterion, but performs more poorly for biased criteria. The linear estimate from ExtNIC is improved by using false alarms only ($L_1^{[0,1]}$, solid line). The extreme case of a very nonlinear system shown in the left panels is very unlikely; more likely scenarios are those in middle and right plots. Both techniques work well for these cases, and are reasonably robust to criterion bias. ExtNIC estimates obtained using second-order moments (Equation 4) were very similar to those obtained using co-variance (5), except in the estimate of the nonlinear kernel for a very linear system (dotted line). Using covariance is slightly more robust to criterion bias, as anticipated in Section 2.

Figure 3c (same plotting conventions as 3b) shows how correlation values improve with the number of trials for a system with linear/nonlinear ratio of 1, and 0 bias. Both techniques converge within 1,000 trials for the conditions explored here. It must be pointed out that no internal noise was used in these simulations; convergence will be slower in the presence of internal noise.

5. Application to real psycho-physical data

First-order and second-order kernels were estimated for a real dataset from a stereo-surface detection experiment (this dataset is the same used for Figure 2 of Neri et al., 1999). In Figure 4, left panels are kernels derived using ExtNIC, right panels are kernels obtained using fMin (with *NoTrialsPerBlock* = 200). Clearly, there is no way here to assess the goodness of derivation for these kernels as was done in Figure 3, because we do not have direct knowledge of the system's structure.

First-order kernels, as derived using the two different techniques, look different (compare panels a and b). A possible reason for this difference is that in this particular experiment, the target-absent stimulus was not just noise. For reasons of careful psychophysical design, it was necessary to have "signal" dots at zero disparity in the no-target condition (they were at 6 arc min [target disparity] in the target-present condition). This would require modification of some of the maths in the "Appendix," and could contribute to the difference between the linear estimates in Figure 4.

Second-order kernels look different too, although they share some features (such as the location of positive and negative peaks). The fMin estimate, however, does not reach statistical significance. The author has verified that fMin is rather unstable for this dataset: when, for example, 6,000 rather than 12,500 trials are used, the minimization procedure returns a flat estimate for W_2.
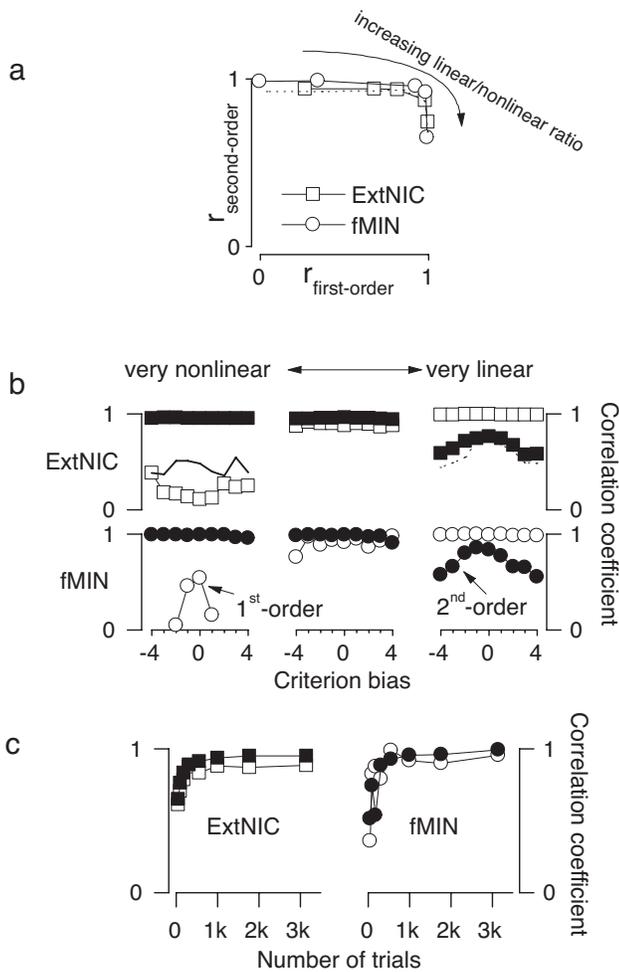
Figure 3. a. Correlations (r's) as computed in Figures 1 and 2 are plotted for both kernels, **L**₁ (linear, abscissa) against **L**₂ (nonlinear, ordinate). Squares (ExtNIC) show correlations obtained using Equation 5 (dashed curve using Equation 4), circles using the fMin method (Section 3). Different points along the curve refer to different amplitude ratios for the two kernels: points at the top left refer to systems for which the amplitude of **L**₂ was larger than the amplitude of **L**₁ (very nonlinear system), whereas points at the bottom right refer to systems for which the opposite was true. b. Correlations for both kernels (open for first-order; solid for second-order) as a function of criterion bias in systems ranging from very nonlinear (left panels) to very linear (right panels), obtained using both methods (top row for ExtNIC, bottom for fMIN). Dotted line is for estimates obtained using Equation 4 rather than 5; solid line is for estimates obtained using only false alarm trials. c. Same plotting conventions, as a function of number of trials.

It should be noted that adding nonlinear filtering only slightly improves simulation of observer's responses. For example, the linear kernel derived using the ExtNIC method, when run through Equation 1, predicts psychophysical responses on individual trials with an accuracy of 72% (this is when optimizing for criterion value every 200 trials). Adding the nonlinear kernel in c only increases this
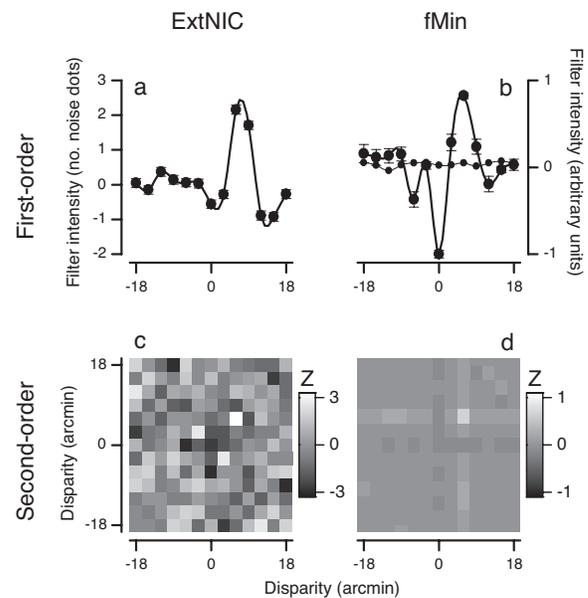

Figure 4. Linear (a and b) and second-order nonlinear (c and d) kernels as computed using ExtNIC (Section 2) and fMin (Section 3) for a dataset collected during a real psychophysical experiment involving surface detection and disparity noise (Neri, Parker, & Blakemore, 1999). The thin line in plot b shows W_1^{adj}, the correcting linear kernel (see Section 3); 12,500 trials were used. Errors and *Z* scores were computed using the bootstrap for ExtNIC, and the Hessian matrix for fMin.

value by about 1%, and the situation is similar for **fMin**. The exact figures depend on how often Equation 1 is allowed to adjust for criterion changes, but the improvement is small (this is the case even when avoiding overfit by predicting responses on trials that were not used to derive kernels).

The fact that adding nonlinear kernels brings about only a minor improvement in simulating observers' responses requires some discussion. First, this only applies to the specific example considered here; nonlinear kernels may play a greater role in other experimental contexts. The author's experience is, however, that although the role may be greater, it is never particularly sizeable. This leads to the following considerations (notice that these considerations relate to a small but nonetheless *statistically significant* contribution of second-order nonlinearity). The measure taken above (what percentage of psychophysical responses is correctly predicted by Equation 1) applies only to the threshold condition explored in the experiment. That is, one cannot predict, from the estimate at threshold, what the importance of nonlinear behavior in the system would be in suprathreshold conditions. It is possible that a nonlinear mechanism that is being exposed at threshold may play a more important role at suprathreshold – the fact that it is being studied at threshold has to do only with the technical requirements that are necessary to expose the mechanism, but this does not mean that the mechanism is being stud-

ied in its "ecological" regime for signal-to-noise ratio. In light of these considerations, it becomes very important to devote efforts to the characterization of nonlinear mechanisms, even if their impact on performance may be small (but nonetheless significant) within the limited threshold range explored in the experiment.

Even if second-order nonlinearities are relatively small, it should be interesting to compare them across conditions. For example, one could compare second-order kernels before and after perceptual learning; it may be that no difference is observed in the linear kernels, but that one is found for the nonlinear ones. This would be informative regardless of the absolute impact of these nonlinearities.

As hinted in the previous paragraph, whichever view is taken on this, it is necessary to assess the statistical reliability of the estimates for both linear and nonlinear kernels. For the extended NIC method, a suitable approach is the bootstrap (Efron & Tibshirani, 1993). For the fMin method, the experimenter needs to adopt standard techniques for estimating the spread of the minimum in search space. Because this is not an experimental work, these topics are not dealt with any further. The reader may refer to Neri et al. (1999) and Neri and Heeger (2002) for examples of applications of the bootstrap to linear and nonlinear kernel estimation. In Figure 4, errors and *Z* scores were computed using this technique for panels on the left (ExtNIC); for fMin kernels, they were estimated using the Hessian matrix at the minimum.

Conclusions

This work presents two different approaches to the computation of psychophysical (both linear and nonlinear) kernels. One approach consists of an extension of the NIC method developed by Ahumada (Beard & Ahumada, 1998) (termed ExtNIC), that takes into account second-order as well as first-order statistics in the classified noise (Section 2); the other method (fMin) uses a minimization approach (Section 3). Under the conditions explored by the simulations in Section 4, both methods work reasonably well.

A big disadvantage associated with the fMin method is that it brings in all problems associated with the solution of a minimization problem (e.g., local minima). If input space is large, kernels are also large, and the minimization step becomes prohibitively difficult (if not impossible) or very unreliable, as well as time-consuming. On the other hand, the ExtNIC method provides (within reasonable ranges of criterion bias and nonlinearity of the system) a fast and efficient route to kernel estimation. However, when the input is not under direct control of the experimenter and does not satisfy certain basic requirements (see "Appendix"), ExtNIC cannot be used at all; in this situation, fMin or similar approaches become a necessary choice.

In general, characterizing nonlinear kernels is an important step in understanding psychophysical systems. This is true regardless of the impact that such nonlinear process-

ing may have within the context of the experiment that was performed to characterize it. For example, it may be the case that nonlinear behavior observed at noise threshold plays a much more important role in suprathreshold conditions.

Appendix

This section shows that:

(1) for a linear filter ($\mathbf{L}_2 = 0$) followed by a static nonlinearity (LN system), L_1^{est} (Equation 3) provides a good estimate of the first-order kernel $\mathbf{L}_1$ (Equation 16; this is in line with Ahumada's [2002] demonstration of the classical Bussgang result [Bussgang, 1952]);

(2) the target-present average noise images $L_1^{[1,o]}$ in Equation 3 (i.e., those for hits and misses) are affected, for a system with nonzero second-order kernel $\mathbf{L}_2$, by "pollution" from this nonlinear term (Equation 15), providing an explanation for some observations made by Ahumada in relation to target-present averages (Ahumada, 1967);

(3) for an LN system and unbiased criterion, $L_2^{est} = 0$ (Equation 17); this result is central to this "Appendix" – however,

(4) when the criterion is biased, L_2^{est} (Equation 4) may contain modulations related to $\mathbf{L}_1$ by a $\mathbf{L}_1(\nu) \cdot \mathbf{L}_1(\xi)$ term (Equation 18);

(5) Equations 4 and 5 are, under some conditions, a reasonable estimate of $\mathbf{L}_2$ (Equation 19).

The approach used here is similar to Lee-Schetzen's (Schetzen, 1980). The system we will consider is nonlinear only in the second-order ($\mathbf{L}_i = 0$ for $i > 2$), and input space is one-dimensional (this is the system considered in all the simulations in this work). Its (unthresholded) output on trial i is:

$$r_i^{[s]} = \sum_{j=1}^{Dim} \mathbf{L}_1(j)I_i^{[s]}(j) + \sum_{j,k}^{Dim} \mathbf{L}_2(j,k)I_i^{[s]}(j)I_i^{[s]}(k), \qquad (8)$$

where s is the stimulus configuration ($s = 0$ is noise only, $s = 1$ is target+noise), Dim is the size of input space (equal to filter size), and:

$$I_i^{[0]}(j) = n_i(j)$$

$$I_i^{[1]}(j) = n_i(j) + t(j)$$

where $n_i(j)$ is the noise image on trial i, and $t(j)$ is the function defining the target. For Gaussian noise n:

$$\frac{1}{N}\sum_{i=1}^{N} n_i(\nu) = E\big(n(\nu)\big) = 0, \qquad \forall \nu$$

$$E\big(n(\nu)n(\xi)\big) = \begin{cases} 0 & \text{if } \nu \neq \xi \\ \mathrm{P} & \text{if } \nu = \xi \end{cases}$$

$$E\big(n(x_1)\ldots n(x_{2M+1})\big) = 0, \qquad M = 0,1,2,\ldots$$

$$E\big(n(x_1)\ldots n(x_{2M})\big) = \sum\prod E\big(n(x_j)n(x_k)\big), M = 0,1,2,\ldots$$

where the symbol $\sum\prod$ stands for summation over all distinct ways of partitioning the $2M$ random variables into products of averages of pairs (Schetzen, 1980).

We can now compute the average response of the system to stimulus types $s = 0$ and $s = 1$ (N_s is the total number of trials of type s; typically, $N_0 = N_1$):

$$\overline{r}^{[0]} = \frac{1}{N_0}\sum_{i=1}^{N_0}\sum_{j=1}^{Dim}\mathbf{L}_1(j)n_i(j) + \frac{1}{N_0}\sum_{i=1}^{N_0}\sum_{j,k}^{Dim}\mathbf{L}_2(j,k)n_i(j)n_i(k)$$
$$= P\sum_{j=1}^{Dim}\mathbf{L}_2(j,j) \tag{9}$$

$$\overline{r}^{[1]} = \overline{r}^{[0]} + r^{[t]} \tag{10}$$

where $r^{[t]}$ is the response of the system to target-alone.

Responses on individual trials are (from 8):

$$r_i^{[0]} = \sum_{j=1}^{Dim}\mathbf{L}_1(j)n_i(j) + \sum_{j,k}^{Dim}\mathbf{L}_2(j,k)n_i(j)n_i(k) \tag{11}$$

$$r_i^{[1]} = r_i^{[0]} + r^{[t]} + 2\cdot\sum_{j,k}^{Dim}\mathbf{L}_2(j,k)n_i(j)t(k) \tag{12}$$

For a linear system ($\mathbf{L}_2 = 0$), $r_i^{[1]} = r_i^{[0]} + r^{[t]}$, which makes sense. However, adding a nonlinear kernel $\mathbf{L}_2$ alters this simple relationship,

$$r_i^{[0]} = \sum_{j=1}^{Dim}\mathbf{L}_1(j)n_i(j) + \sum_{j,k}^{Dim}\mathbf{L}_2(j,k)n_i(j)n_i(k)$$

Let us derive L_1^{est}, the estimate for the linear kernel L_1. Average noise images for the four stimulus-response classes are indicated by $L_1^{[s,o]}$, where s is, as above, the target (0, absent; 1, present) and o is the response (e.g., the average noise image for the false alarms is $L_1^{[0,1]}$). $L_1^{[s,1]}$, as computed using Ahumada's method, can be written as:

$$L_1^{[s,1]}(v) = \frac{1}{N_s}\sum_{i=1}^{N_s} p_i^{[s,1]}n_i(v) \tag{13}$$

where $p_i^{[s,1]}$ is the probability that trial i will be of type $[s,1]$, and $p_i^{[s,1]} = g(r_i^{[s]})$ where g is a static nonlinear function that maps output response from the system onto probability of psychophysical response "yes." We now expand g up to its second-order Taylor term around mean response $\overline{r}^{[s]}$ (implying that we are assuming g to have continuous derivatives up to order 3). Equation 13 can then be written as:

$$L_1^{[s,1]}(v) = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(\begin{array}{c}g(\overline{r}^{[s]}) + g'(\overline{r}^{[s]})(r_i^{[s]} - \overline{r}^{[s]}) + \\ \frac{1}{2}g''(\overline{r}^{[s]})(r_i^{[s]} - \overline{r}^{[s]})^2\end{array}\right)n_i(v) \tag{14}$$

Substituting Equations 9 to 12 into 14, this becomes

$$L_1^{[s,1]}(v) = g'(\overline{r}^{[s]})P\left(L_1(v) + 2s\sum_{j=1}^{Dim}\mathbf{L}_2(v,j)t(j)\right)$$

$$+2P^2 g''(\overline{r}^{[s]})\left(\sum_{j=1}^{Dim}\mathbf{L}_1(j)\mathbf{L}_2(v,j) + 2s\sum_{j,k}^{Dim}\mathbf{L}_2(v,j)\mathbf{L}_2(j,k)t(k)\right) \tag{15}$$

and, from Equation 13, it is easy to show that $L^{[s,0]}(v) = -L_1^{[s,1]}(v)$. Equation 3 is then

$$L_1^{est}(v) = 2P\left(g'(\overline{r}^{[1]}) + g'(\overline{r}^{[0]})\right)\mathbf{L}_1(v) + 4Pg'(\overline{r}^{[1]})\sum_{j=1}^{Dim}\mathbf{L}_2(v,j)t(j)$$

$$+4P^2\left(g''(\overline{r}^{[1]}) + g''(\overline{r}^{[0]})\right)\sum_{j=1}^{Dim}\mathbf{L}_1(j)\mathbf{L}_2(v,j) + 8P^2 g''(\overline{r}^{[1]})\sum_{j,k}^{Dim}\mathbf{L}_2(v,j)\mathbf{L}_2(j,k)t(k)$$

Let us verify something very familiar: for $\mathbf{L}_2 = 0$, this reduces to

$$L_1^{est}(v) \propto \mathbf{L}_1(v) \tag{16}$$

Figure 5 is a useful tool for thinking about g. This nonlinearity is assumed to be approximately odd-symmetric around its midpoint. As a matter of fact, most decisional transducers enjoy this symmetry (e.g., a noisy threshold belongs to this category; see Nykamp & Ringach, 2002, for more examples). It is also the case that $\overline{p}^{[0,1]} \approx g(\overline{r}^{[0]})$, and $\overline{p}^{[1,1]} \approx g(\overline{r}^{[1]})$. When the criterion is unbiased (i.e., $\overline{p}^{[1,1]}_{[0,1]} = 1 - \overline{p}^{[1,1]}_{[0,1]}$), the two regions of g that map $\overline{r}^{[0]}$ and $\overline{r}^{[1]}$ onto $\overline{p}^{[0]}$ and $\overline{p}^{[1,1]}$, respectively, are mirror-symmetric, so that $g'(\overline{r}^{[0]}) = g'(\overline{r}^{[1]})$ and $g''(\overline{r}^{[0]}) = -g''(\overline{r}^{[1]})$. This means that, for an unbiased criterion,

$$L_1^{est}(v) = 4Pg'(\overline{r}^{[0]})\left(\mathbf{L}_1(v) + \sum_{j=1}^{Dim}\mathbf{L}_2(v,j)t(j)\right)$$

$$-8P^2 g''(\overline{r}^{[0]})\sum_{j,k}^{Dim}\mathbf{L}_2(v,j)\mathbf{L}_2(j,k)t(k)$$

Despite lack of bias in the criterion, L_1^{est} is still affected by terms that depend on the interaction between $\mathbf{L}_2$ and the target. These terms come, of course, from target-present averages. This result may be related to the observation made by Ahumada (1967) that for a system pooling nonlinearly (e.g., max rule) from a bank of linear filters, averages for noise-only trials would return the average of the bank (i.e., the linear part of the process), whereas aver-

ages from target+noise trials would return something heavily affected by the target shape. In fact, Equation 15 shows that when the target is present ($s = 1$), there are extra terms that involve $\mathbf{L}_2 \cdot t$; for a nonlinear system ($\mathbf{L}_2 \neq 0$), these terms affect the target-present averages $L_1^{[1,1]}$ and $L_1^{[1,0]}$, making L_1^{est} (equation above) depart from Equation 16.

Let us now turn to the nonlinear kernel $\mathbf{L}_2$:

$$L_2^{[s,o]}(\nu,\xi) = \frac{1}{N_s}\sum_{i=1}^{N_s} p_i^{[s,o]} n_i(\nu)n_i(\xi)$$

By adopting a procedure similar to that used for deriving $L_1^{[s,1]}$, one can show that:

$$L_2^{[s,1]}(\nu,\xi) = \delta(\nu-\xi)Pg(\overline{r}^{[s]}) + 2P^2 g'(\overline{r}^{[s]})\mathbf{L}_2(\nu,\xi)$$

$$+ \frac{P^2}{2}g''(\overline{r}^{[s]})(A + s4B)$$

where

$$A = 2\mathbf{L}_1(\nu)\mathbf{L}_1(\xi) - P\sum_{j=1}^{Dim}\mathbf{L}_2(j,j)\mathbf{L}_2(\nu,\xi)$$

$$+ 9P\sum_{j=1}^{Dim}\mathbf{L}_2(j,\nu)\mathbf{L}_2(j,\xi)$$

$$+ \delta(\nu-\xi)\left(\sum_{j=1}^{Dim}\mathbf{L}_1^2(j) + 2P\sum_{j,k}^{Dim}\mathbf{L}_2^2(j,k)\right)$$

$$B = 2\sum_{j,k}^{Dim}\mathbf{L}_2(\nu,j)\mathbf{L}_2(\xi,k)t(j)t(k)$$

$$+ \sum_{j=1}^{Dim}\mathbf{L}_1(\xi)\mathbf{L}_2(\nu,j)t(j) + \sum_{j=1}^{Dim}\mathbf{L}_1(\nu)\mathbf{L}_2(\xi,j)t(j)$$

$$+ \delta(\nu-\xi)\left(\sum_{j,k,l}^{Dim}\mathbf{L}_2(j,k)\mathbf{L}_2(j,l)t(k)t(l) + \sum_{j,k}^{Dim}\mathbf{L}_1(j)\mathbf{L}_2(j,k)t(k)\right)$$

and $\delta(x)$ is the Dirac delta function ($\delta(x) = 0$ for $x \neq 0$, $\int_{-\infty}^{\infty}\delta(x)dx = 1$). It is easy to show that $L_2^{[s,0]}(\nu,\xi) = \delta(\nu-\xi)P - L_2^{[s,1]}(\nu,\xi)$. We can derive the estimate for $\mathbf{L}_2$ according to Equation 4:

$$L_2^{est}(\nu,\xi) = 2\delta(\nu-\xi)P\left(g(\overline{r}^{[0]}) + g(\overline{r}^{[1]}) - 1\right)$$

$$+ 4P^2\mathbf{L}_2(\nu,\xi)\left(g'(\overline{r}^{[0]}) + g'(\overline{r}^{[1]})\right)$$

$$+ \left(g''(\overline{r}^{[0]}) + g''(\overline{r}^{[1]})\right)P^2 A + g''(\overline{r}^{[1]})P^2 4B$$
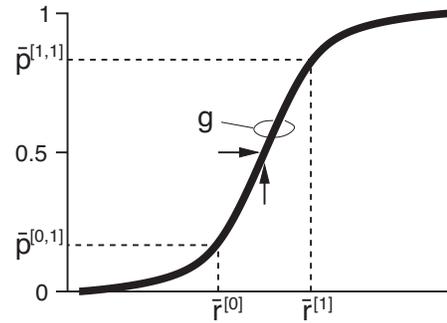

Figure 5. The function *g* maps the system output *r* (abscissa) onto probability of responding "yes" (ordinate). In the text, *g(r)* is approximated around $\overline{r}^{[0]}$ and $\overline{r}^{[1]}$, the average responses when the target is absent and present. *g* is assumed to be an odd function with respect to the shifted origin $[g^{-1}(0.5),0.5]$ (indicated by arrows) – this is a reasonable assumption, as most decisional nonlinearities satisfy this requirement. For an unbiased criterion, *g* is necessarily placed with respect to the $\overline{r}$'s so that $\overline{p}^{[0,1]} = 1 - \overline{p}^{[1,1]}$, and $\overline{r}^{[0]}$ and $\overline{r}^{[1]}$ map to symmetric regions of *g*. This is the condition depicted here. In this case, the symmetry of *g* means that odd derivatives of *g* are the same (e.g., $g'(\overline{r}^{[0]}) = g'(\overline{r}^{[1]})$) and even derivatives are of opposite sign (e.g., $g''(\overline{r}^{[0]}) = -g''(\overline{r}^{[1]})$) at $\overline{r}^{[0]}$ and $\overline{r}^{[1]}$. This is not true when the criterion is biased, and $\overline{r}^{[0]}$ and $\overline{r}^{[1]}$ map to nonsymmetric regions of *g*.

Let us focus on an LN system ($\mathbf{L}_2 = 0$). If the criterion is unbiased, we obtain

$$L_2^{est}(\nu,\xi) = 0 \tag{17}$$

Notice that this result applies even for expansions of *g* to orders higher than the second. This means that if an experiment is carried out at unbiased criterion and $L_2^{est} \neq 0$, then the system cannot be modeled as LN (which is the most widely used model in vision). In other words, the effect of the nonlinear decisional step *g* on L_2^{est} can be neutralized by having a balanced criterion (provided *g* enjoys approximate odd-symmetry).

If the criterion is biased, an LN system returns

$$L_2^{est}(\nu,\xi) = \delta(\nu-\xi) \cdot K_1 + K_2 \cdot \mathbf{L}_1(\nu)\mathbf{L}_1(\xi) \tag{18}$$

where the *K*'s are constants (for a given overall criterion bias). The last term is the reason for using Equation 5 rather than 4: computing covariance partly compensates for the $\mathbf{L}_1(\nu)\mathbf{L}_1(\xi)$ term in the estimate above.

If $\mathbf{L}_2 \neq 0$, then L_2^{est} is an appropriate estimator for $\mathbf{L}_2$ if the criterion is unbiased and *g* can be sufficiently well approximated as linear over the range spanned by $r_i^{[s]} - \overline{r}^{[s]}$ (i.e. $g''(\overline{r}^{[s]}) \approx 0$). For these conditions, one obtains:

$$L_2^{est}(\nu,\xi) \propto \mathbf{L}_2(\nu,\xi) \tag{19}$$

Criterion bias only slightly modifies this equation, adding $\delta(\nu-\xi) \cdot constant$. If the second-order term in the expan-

sion of g is sizeable, then the term B appears in L_2^{est}, whether the criterion is biased or not. However, the simulations in section 4 show that this term is not particularly disruptive of L_2^{est}.

Acknowledgments

Commercial relationships: none.
Corresponding author: Peter Neri.
Address: Department of Zoology
University of Cambridge
Downing Street
Cambridge CB2 3 EJ
England, UK
Email: pn232@hermes.cam.ac.uk.

Footnotes

[1]This is a simplified description of the Volterra expansion. The original formulation involves *convolution* between $\mathbf{L}_i$ and I to obtain V_i, so that the dimensionality of V_i is equal to that of the input, I:

$$V_i\big[\mathbf{L}_i(\nu_1,\nu_2,\ldots,\nu_i),I\big](x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathbf{L}_i(\nu_1,\ldots,\nu_i)$$

$$I(x-\nu_1)I(x-\nu_2)\ldots I(x-\nu_i)d\nu_1 d\nu_2 \ldots d\nu_i$$

In Equation 2, V_i are scalars, as it is simply the *cross-correlation* between $\mathbf{L}_i$ and I that is being computed. Whether this simplified representation is applicable or not depends on the specific context being studied. It incorporates the assumption that the nervous system is basing its behavioral decisions on outputs from the most sensitive mechanism(s). A convolution operation can be thought of as extended sampling by a bank of linear filters; of all these filters, only those sampling the region in the vicinity of the target will be used to drive behavior, as those provide useful information for the task at hand (i.e., they are the most sensitive to the target). Equation 2 refers only to these mechanisms. Another (more practical) reason for adopting this simplified representation is that it conforms to Ahumada's formulation.

References

Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for 2AFC experiments. *Journal of Vision, 2*(1), 66-78. [PubMed] [Article]

Ahumada, A. J. (1967). Detection of tones masked by noise: A comparison of human observers with digital-computer-simulated energy detectors of varying bandwidth. Doctoral dissertation (Technical Report No.29, Human Communications Laboratory; Department of Psychology), UCLA, Los Angeles, CA.

Ahumada, A. J., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *Journal of the Acoustical Society of America, 57*, 385-390.

Ahumada, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise [Abstract]. *Perception, 25*, 18.

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision, 2*(1), 121-131. [PubMed] [Article]

Beard, B. L., & Ahumada, A. J. (1998). A technique to extract relevant image features for visual tasks. *Proceedings of SPIE, 3299*, 79-85.

Bussgang, J. J. (1952). Crosscorrelation functions of amplitude distorted Gaussian signals (Tech. Rep. No. 216). Boston: MIT Research Laboratory of Electronics.

Efron, B., & Tibshirani, R. (1996). *An introduction to the bootstrap.* New York: Chapman and Hall.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*, 203-218. [PubMed]

Emerson, R. C., & Citron, M. C. (1992). Linear and nonlinear mechanisms of motion selectivity in simple cells of the cat's striate cortex. In R. B. Pinter & N. Nabet (Eds.), *Nonlinear vision* (pp. 75-89). New York: CRC Press.

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). *Analysis of physiological systems: The white-noise approach.* New York: Plenum Press.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*, 695-698. [PubMed]

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience, 5*, 812-816. [PubMed]

Nykamp, D. Q., & Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision, 2*(1), 1-11. [PubMed] [Article]

Sakai, H. M., Naka, K.-I., & Korenberg, M. J. (1988). White-noise analysis in visual neuroscience. *Visual Neuroscience, 1*, 287-296. [PubMed]

Schetzen, M. (1980). *The Volterra and Wiener theories of nonlinear systems.* New York: John Wiley & Sons.

Victor, J. D. (1992). Nonlinear systems analysis in vision: Overview of kernel methods. In R. B. Pinter & N. Nabet (Eds.), *Nonlinear vision* (pp. 1-37). New York: CRC Press.

Volterra, V. (1959). Theory of functionals and of integral and integrodifferential equations. New York: Dover Publications.