Second-order receptive fields reveal multidigit interactions in area 3b of the macaque monkey

Pramodsingh H. Thakur, Paul J. Fitzgerald, and Steven S. Hsiao

Departments of Neuroscience and Biomedical Engineering, Zanvyl Krieger Mind/Brain Institute and Kennedy-Krieger Institute, Johns Hopkins University, Baltimore, Maryland

Submitted 29 November 2010; accepted in final form 28 March 2012

One of the earliest attempts at characterizing the RF structures of neurons in area 3b of SI involved indenting the RF with temporally and spatially separated paired probes (Gardner 1984; Gardner and Costanzo 1980a, 1980b). These studies revealed that the RF has two components, a central excitatory core with a latency of ~10 ms and a spatially overlapping inhibitory field called infield inhibition that is delayed with respect to the excitatory core. Pharmacological blockade of GABAergic transmission increases the size of the RF excitatory area, as well as the maximum response evoked at the RF center (Alloway and Burton 1991; Dykes et al. 1984). This suggests that infield inhibition is mediated by cortical interneurons, whereas the central excitatory core represents the thalamic feed-forward input. Thus understanding the respective contributions of excitatory core and infield inhibition to the neural responses may elucidate the roles that feed-forward inputs and local intracortical circuits play in imparting feature selectivity.

Infield inhibition has been also shown in intracellular recordings as delayed inhibitory postsynaptic potentials (IPSPs) (Andersson 1965; Innocenti and Manzoni 1972; Whitehorn and Towe 1968) and accounts for the decreased response to a test stimulus preceded by a conditioning stimulus within the RF (Gardner and Costanzo 1980a, 1980b; Laskin and Spencer 1979). Since it is only evoked by paired stimuli, and not revealed when single punctate probes are used as stimuli, which show area 3b RFs as simple homogeneous excitatory regions (Iwamura et al. 1983; Mountcastle and Powell 1959; Sur 1980, Sur et al. 1984), infield inhibition is an interactive second-order effect. Linear RF models estimated from responses to multiple simultaneous indentations, via either scanned embossed random dot patterns (DiCarlo et al. 1998) or random probe indentations (Sripati et al. 2006), capture inhibition via negative linear weights, but only in spatiotemporal regions beyond the excitatory core. These studies show two components of inhibition. The portion of infield inhibition that is temporally coincident with the excitatory core but spatially extends beyond the excitation is revealed as the temporally coincident component and gives the neurons a center-surround-like structure.

The portion of infield inhibition that is delayed with respect to the excitatory core. Pharmacological blockade of GABAergic transmission increases the size of the RF excitatory area, as well as the maximum response evoked at the RF center (Alloway and Burton 1991; Dykes et al. 1984). This suggests that infield inhibition is mediated by cortical interneurons, whereas the central excitatory core represents the thalamic feed-forward input. Thus understanding the respective contributions of excitatory core and infield inhibition to the neural responses may elucidate the roles that feed-forward inputs and local intracortical circuits play in imparting feature selectivity.

Infield inhibition has been also shown in intracellular recordings as delayed inhibitory postsynaptic potentials (IPSPs) (Andersson 1965; Innocenti and Manzoni 1972; Whitehorn and Towe 1968) and accounts for the decreased response to a test stimulus preceded by a conditioning stimulus within the RF (Gardner and Costanzo 1980a, 1980b; Laskin and Spencer 1979). Since it is only evoked by paired stimuli, and not revealed when single punctate probes are used as stimuli, which show area 3b RFs as simple homogeneous excitatory regions (Iwamura et al. 1983; Mountcastle and Powell 1959; Sur 1980, Sur et al. 1984), infield inhibition is an interactive second-order effect. Linear RF models estimated from responses to multiple simultaneous indentations, via either scanned embossed random dot patterns (DiCarlo et al. 1998) or random probe indentations (Sripati et al. 2006), capture inhibition via negative linear weights, but only in spatiotemporal regions beyond the excitatory core. These studies show two components of inhibition. The portion of infield inhibition that is temporally coincident with the excitatory core but spatially extends beyond the excitation is revealed as the temporally coincident component and gives the neurons a center-surround-like structure.

The portion of infield inhibition that is delayed with respect to the excitatory core. Pharmacological blockade of GABAergic transmission increases the size of the RF excitatory area, as well as the maximum response evoked at the RF center (Alloway and Burton 1991; Dykes et al. 1984). This suggests that infield inhibition is mediated by cortical interneurons, whereas the central excitatory core represents the thalamic feed-forward input. Thus understanding the respective contributions of excitatory core and infield inhibition to the neural responses may elucidate the roles that feed-forward inputs and local intracortical circuits play in imparting feature selectivity.

Infield inhibition has been also shown in intracellular recordings as delayed inhibitory postsynaptic potentials (IPSPs) (Andersson 1965; Innocenti and Manzoni 1972; Whitehorn and Towe 1968) and accounts for the decreased response to a test stimulus preceded by a conditioning stimulus within the RF (Gardner and Costanzo 1980a, 1980b; Laskin and Spencer 1979). Since it is only evoked by paired stimuli, and not revealed when single punctate probes are used as stimuli, which show area 3b RFs as simple homogeneous excitatory regions (Iwamura et al. 1983; Mountcastle and Powell 1959; Sur 1980, Sur et al. 1984), infield inhibition is an interactive second-order effect. Linear RF models estimated from responses to multiple simultaneous indentations, via either scanned embossed random dot patterns (DiCarlo et al. 1998) or random probe indentations (Sripati et al. 2006), capture inhibition via negative linear weights, but only in spatiotemporal regions beyond the excitatory core. These studies show two components of inhibition. The portion of infield inhibition that is temporally coincident with the excitatory core but spatially extends beyond the excitation is revealed as the temporally coincident component and gives the neurons a center-surround-like structure.

The portion of infield inhibition that is delayed with respect to the excitatory core. Pharmacological blockade of GABAergic transmission increases the size of the RF excitatory area, as well as the maximum response evoked at the RF center (Alloway and Burton 1991; Dykes et al. 1984). This suggests that infield inhibition is mediated by cortical interneurons, whereas the central excitatory core represents the thalamic feed-forward input. Thus understanding the respective contributions of excitatory core and infield inhibition to the neural responses may elucidate the roles that feed-forward inputs and local intracortical circuits play in imparting feature selectivity.

Infield inhibition has been also shown in intracellular recordings as delayed inhibitory postsynaptic potentials (IPSPs) (Andersson 1965; Innocenti and Manzoni 1972; Whitehorn and Towe 1968) and accounts for the decreased response to a test stimulus preceded by a conditioning stimulus within the RF (Gardner and Costanzo 1980a, 1980b; Laskin and Spencer 1979). Since it is only evoked by paired stimuli, and not revealed when single punctate probes are used as stimuli, which show area 3b RFs as simple homogeneous excitatory regions (Iwamura et al. 1983; Mountcastle and Powell 1959; Sur 1980, Sur et al. 1984), infield inhibition is an interactive second-order effect. Linear RF models estimated from responses to multiple simultaneous indentations, via either scanned embossed random dot patterns (DiCarlo et al. 1998) or random probe indentations (Sripati et al. 2006), capture inhibition via negative linear weights, but only in spatiotemporal regions beyond the excitatory core. These studies show two components of inhibition. The portion of infield inhibition that is temporally coincident with the excitatory core but spatially extends beyond the excitation is revealed as the temporally coincident component and gives the neurons a center-surround-like structure.

The portion of infield inhibition that is delayed with respect to the excitatory core. Pharmacological blockade of GABAergic transmission increases the size of the RF excitatory area, as well as the maximum response evoked at the RF center (Alloway and Burton 1991; Dykes et al. 1984). This suggests that infield inhibition is mediated by cortical interneurons, whereas the central excitatory core represents the thalamic feed-forward input. Thus understanding the respective contributions of excitatory core and infield inhibition to the neural responses may elucidate the roles that feed-forward inputs and local intracortical circuits play in imparting feature selectivity.

Infield inhibition has been also shown in intracellular recordings as delayed inhibitory postsynaptic potentials (IPSPs) (Andersson 1965; Innocenti and Manzoni 1972; Whitehorn and Towe 1968) and accounts for the decreased response to a test stimulus preceded by a conditioning stimulus within the RF (Gardner and Costanzo 1980a, 1980b; Laskin and Spencer 1979). Since it is only evoked by paired stimuli, and not revealed when single punctate probes are used as stimuli, which show area 3b RFs as simple homogeneous excitatory regions (Iwamura et al. 1983; Mountcastle and Powell 1959; Sur 1980, Sur et al. 1984), infield inhibition is an interactive second-order effect. Linear RF models estimated from responses to multiple simultaneous indentations, via either scanned embossed random dot patterns (DiCarlo et al. 1998) or random probe indentations (Sripati et al. 2006), capture inhibition via negative linear weights, but only in spatiotemporal regions beyond the excitatory core. These studies show two components of inhibition. The portion of infield inhibition that is temporally coincident with the excitatory core but spatially extends beyond the excitation is revealed as the temporally coincident component and gives the neurons a center-surround-like structure.

The portion of infield inhibition that is delayed with respect to the excitatory core. Pharmacological blockade of GABAergic transmission increases the size of the RF excitatory area, as well as the maximum response evoked at the RF center (Alloway and Burton 1991; Dykes et al. 1984). This suggests that infield inhibition is mediated by cortical interneurons, whereas the central excitatory core represents the thalamic feed-forward input. Thus understanding the respective contributions of excitatory core and infield inhibition to the neural responses may elucidate the roles that feed-forward inputs and local intracortical circuits play in imparting feature selectivity.
infield inhibition is lost. To fully disentangle the effects of excitation and inhibition, especially in regions of spatiotemporal overlap, it is essential to extend the linear models to include second-order terms to accurately capture interactive infield inhibitory effects.

The second aspect of somatosensory processing investigated here is multidigit interaction. Although neurons in the digit region of area 3b have traditionally been thought to be driven exclusively by inputs from single finger pads, optical imaging studies (Chen et al. 2001, 2005) indicate that subthreshold inputs to area 3b neurons may extend over an area substantially larger than suprathreshold inputs. Also, cortical plasticity studies (Wang et al. 1995) have shown that coincident presentation of stimuli on adjacent digits expands neuronal RFs in area 3b to span multiple digits, indicating that an underlying substrate for multidigit interactions may exist in area 3b. Recently, Reed et al. (2006) showed, using multiple probe stimuli, that responses of neurons with palmar RFs can be modulated by stimuli outside the classical RF, like nonclassical RF effects observed in the visual system (see Allman et al. 1985). The RF mechanisms that underlie these nonclassical effects in area 3b neurons are not known.

In this study, we estimate multidigit second-order RFs of area 3b neurons. Neural responses were obtained with methods used previously (DiCarlo et al. 1998) except that random dot patterns were scanned simultaneously across three distal pads (i.e., digits D2–4) of the awake macaque. RFs were then estimated by using a point process model that includes both linear and second-order (quadratic) terms. To circumvent the dimensional explosion due to inclusion of quadratic terms in the model, we used an iterative method based on matching pursuit (Mallat and Zhang 1993) to determine a subset of linear and quadratic terms that have the most salient effects on the neuronal responses. We then investigated how these nonlinear properties affect the neural responses.

METHODS

Animals and Surgery

Single-unit recordings were made with standard methods from three cortical hemispheres in two male rhesus monkeys (M. mulatta) weighing 5–7 kg. Briefly, before the recordings, animals were trained to sit quietly in a primate chair (see below). Surgery was then performed to secure a head-holding device and recording chambers (19-mm diameter) to the skull. Surgical anesthesia was induced with ketamine HCl (20 mg/kg im) and maintained with pentobarbital (10–25 mg/kg h−1 iv). All surgical procedures were done under sterile conditions and were reviewed and approved by the Johns Hopkins Animal Care and Use Committee and in accordance with the rules and regulations of the Society for Neuroscience.

General Recording Techniques

All recordings were performed in awake, behaving monkeys. The monkeys were trained to sit in a primate chair with their hands restrained and were rewarded with sips of water at random intervals (about once every 7 s) while tactile stimuli (see below) were presented to the distal pads of digits D2, D3, and D4. On each recording day, a multielectrode microdrive (Mountcastle et al. 1991) was loaded with seven quartz-coated platinum-tungsten (90:10) microelectrodes (shaft diameter 80 μm, tip diameter 4 μm, impedance 1–5 MΩ at 1,000 Hz). On approximately half of the recording days, microelectrodes were coated with fluorescent dye (DiI; Molecular Probes, Eugene, OR) to facilitate later histological localization of the recording sites (DiCarlo et al. 1996). For all recordings, the tip of the microdrive was inserted into the recording chamber and oriented with the microelectrodes normal to the dura mater.

The seven microelectrodes were oriented at an angle of 45° to the mediolateral axis (parallel to the central sulcus) with center-to-center spacings of 400 μm. On each recording day, the microelectrodes were positioned within the recording chamber with micrometer-controlled translation stages (model 430; Newport, Irvine, CA) (accuracy within 5 μm).

To record in SI, we used a stereotaxic instrument to first center the recording chamber over the central sulcus [Horsley-Clarke (HC) coordinates: anterior 6, lateral 21]. The central sulcus was then located within the recording chamber based on the responses encountered on the microelectrodes. Responses anterior to the sulcus are motor, whereas responses posterior to the sulcus are somatosensory. Once the location of the sulcus was determined, the microelectrode array was adjusted to span the digit representation, just posterior to the central sulcus.

Because SI lies in the postcentral gyrus, we recorded first superficially before driving the electrodes deeper into cortex. Thus, for each position of the array within the chamber, recordings were first made in area 1, then on the next day in area 3b, and then on the third day in area 3a. After sampling the responses from the three areas, the microelectrode array was moved systematically along the postcentral gyrus until neurons ceased to be driven by the digits. Determination of which area we were recording from was based on the progression of RFs from the distal pads to the proximal pads and then to the distal pads as the electrodes passed from area 1 to area 3b and confirmed postmortem. Results in this report are based on the data collected from area 3b.

Data Collection

The signals from the seven channels were band-pass filtered and fed into a spike sorter (Alpha Omega), which detected and classified action potentials based on their similarity to spike templates. Neurons that met the following criteria were selected for analysis: 1) the action potentials were well isolated from noise and from other spike templates; 2) the neuronal firing was stable throughout the stimulus protocol; and 3) the neuronal RF, as determined by handheld probes, was located on the distal pads.

Stimulus

The stimuli consisted of raised dot patterns fabricated from sheets of photosensitive plastic that are water-soluble until exposed to UV light (Toyoba Printing Plastics, EF series). The patterns are produced by laying a photographic negative of the pattern over the plastic sheet and exposing it to UV light, which polymerizes and hardens the exposed plastic. The portion of the surface layer not exposed to UV light is scrubbed off lightly in water, leaving a raised pattern whose thickness equals that of the original water-soluble layer (380 μm). The stimulus pattern was specified in Postscript and printed at 3,386 dots/in. After construction, the sheet containing the stimulus patterns was wrapped around and glued to a cylindrical drum, 640 mm in circumference, which was mounted on a rotating drum stimulator (Johnson and Phillips 1988).

The random dot patterns were designed to be similar to the patterns used in a previous study (DiCarlo et al. 1998). Each dot was 380 μm high (in relief) and 400 μm in diameter at its top, with sides that sloped away at 60° relative to the surface of the drum. The location of the center of each dot was determined by selecting two random numbers from a uniform random number generator and scaling them to the height (40 mm) and width (560 mm) of the pattern. The dot centers were specified to a precision of 1 μm, and dots were allowed to overlap when the random number generator placed them closer than
the dot diameter. Dots were added until the average dot density equalled 10 dots/cm².

After one or more neurons with RFs (as determined by qualitative mapping) on the distal pads of digit D2, D3, or D4 were isolated, the drum with the random dot pattern was lowered and positioned so that it simultaneously scanned across the distal pads of digits D2, D3, and D4. The pattern was scanned in the proximal to distal direction at a speed of 25 mm/s, while a torque motor maintained a contact force of 90g. After each complete rotation, the drum was shifted laterally by 200 μm, thus presenting a slightly different pattern during successive scans. Data were collected continuously for a minimum of 30 scans, or ∼13 min. The contact patch of the finger on the drum was monitored with a metallic “hot” pin (1-mm diameter) placed on the surface of the drum. Contact of the finger with this pin triggered a series of square wave pulses that were stored as events in the data stream. This pulse signal was then used to align the section of stimulus pattern that was in contact with the pads at any given time step with the neural response.

Linear and Quadratic Receptive Field Estimation

The recorded spike train was binned (32-ms bin width) to yield a spike count vector ($R_{\text{ab}} = [r_1, r_2, \ldots, r_{\text{bin}}]$) (see Fig. 1). Each stimulated pad was divided into subregions of size 800 μm × 800 μm, yielding 432 subregions (3 pads × 144 subregions each) that could potentially affect the neural response. For a linear model, each subregion was associated with a weight $w_i$ that defines its linear excitatory or inhibitory influence on the neural response. The neural firing rate at the $n$th step is related to the linear combination ($e_n$) of the corresponding weights at each subregion multiplied by the number of dots $x_i(n)$ in that subregion at the $n$th time step (see Fig. 1A). Specifically,

$$e_n = w_0 + \sum_{i=1}^{432} x_i(n)w_i$$

(1)

In this model, excitation and inhibition are captured by positive and negative weights, respectively.

The model was estimated in two steps. First, a subset of terms or pixels (usually <100) that had a salient contribution to the neural response was identified with a dimensionality reduction method based on matching pursuit (see APPENDIX). Weights were then estimated by fitting a generalized linear model (GLM) (function glmfit, MATLAB; MathWorks) with a log-link function between the selected subset of terms and observed spike count vector. This assumes a Poisson distribution of spike counts [i.e., $r_n \sim \text{Poisson} (e^n)\]$, yielding a log-linear model in input covariates or subregions ($x$). Poisson assumption has been shown to be reasonable with a sufficient number of repeated trials or at intervals beyond refractory periods (Kass et al. 2005; Ventura et al. 2002).

For the quadratic model, in addition to the linear weights ($w_i$) were defined to capture every pair of possible interactions between subregions ($i,j$) across the three pads (see Fig. 1B). Such weights denote the interaction between the specific pairs of subregions, and their values indicate the effect of stimulating the two subregions simultaneously on the neural response. Specifically, the model was defined as

$$e_n = w_0 + \sum_{i=1}^{432} w_i x_i(n) + \sum_{i=1}^{432} \sum_{j=1}^{432} w_{ij} x_i(n)x_j(n)$$

(2)

where $x_i$ and $x_j$ are the number of dots stimulating the $i$th and $j$th subregions during the $n$th time step. The second-order weight $w_{ij}$ captures the effect of simultaneously indenting subregions $i$ and $j$ on the neural response. A negative (positive) weight indicates suppressive (facilitative) interaction due to a simultaneous stimulation of subregions $i$ and $j$. On the other hand, the contribution of this term is zero if either of the two subregions $i$ and $j$ is not stimulated. Thus, under this model, any linear effect will be captured by the linear weights $w_i$, which we refer to as the linear component of the quadratic model, whereas the interactive effects between pairs of subregions are captured by the second-order weights $w_{ij}$. If the infold inhibition is a linear effect it will be captured by negative linear weights, but if it is an interactive effect it will be captured by the second-order weights $w_{ij}$ and will not be found in the linear component of the quadratic model.

Like the linear model, the quadratic model was estimated in two steps. First, a set of salient terms was identified with matching pursuit, followed by an estimate of the weights corresponding to those terms...
via fitting a GLM, yielding a log-quadratic model in input covariates or subregions (\(x\)) [i.e., \(r_n \sim \text{Poisson}(e^{x})\)].

**Linear and Quadratic Kernels Spanning Multiple Subregions**

With an average dot density of 10 dots/cm\(^2\) (see above), the probability of stimulating a subregion at a given instant of time is \(P = 10/(12.5 \times 12.5) = 0.064\). With 598 bins each from 30 scans (\(N = 598 \times 30 = 17,984\) unique equations for fitting), the expected number of nonzero occurrences of stimulus dots at any subregion over the period analyzed is \(N \times P = 1,148\). However, the number of simultaneous occurrences of nonzero dot counts over two subregions (occurrences of simultaneous stimulation to uncover quadratic response) is \(N \times P = 74\). Because of such low prevalence of paired stimulation, the errors in estimating quadratic weights are likely to be higher. To improve the reliability of quadratic weights, we expanded our dictionary to include “kernels” that span multiple subregions (up to 5 subregions \(\times 5\) subregions or 4 mm \(\times 4\) mm) and used matching pursuit to iteratively pull out the best matching linear or quadratic kernel at any stage. For example, a linear kernel spanning 2 \(\times 2\) subregions was defined as the total number of dots within all 4 subregions spanned. A quadratic kernel spanning 2 \(\times 2\) subregions was defined as the product of the number of dots within the two 2 \(\times 2\) regions represented by that kernel. Mathematically, using such multiple subregion spanning kernels translates into assuming that weights are similar for all subregions spanned (and hence combining them and estimating a single weight instead). For example, a linear kernel \(B_L\) spanning subregions \(x_1, x_2, x_3,\) and \(x_4\) involves the following approximation:

\[
w_{i1}x_1 + w_{i2}x_2 + w_{i3}x_3 + w_{i4}x_4 \approx w(x_1 + x_2 + x_3 + x_4) \approx wB_L
\]

Similarly, a 2 \(\times 2\) quadratic kernel \(B_Q\) representing a “combined” interaction between subregions \(x_1, x_2, x_3,\) and \(x_4\) and subregions \(x_5, x_6, x_7,\) and \(x_8\) involves the following approximation:

\[
w_{i1}x_1x_5 + w_{i2}x_1x_6 + w_{i3}x_1x_7 + w_{i4}x_1x_8 + \ldots + w_{i45}x_5x_6 + w_{i46}x_5x_7 + w_{i47}x_5x_8 + \ldots + w_{i456}x_5x_6x_7 + w_{i457}x_5x_6x_8 + \ldots = w(x_1 + x_3 + x_5 + x_7)(x_5 + x_6 + x_7 + x_8) \approx wB_Q
\]

Although the prevalence of paired stimulation of subregions is low, the prevalence increases dramatically as we consider kernels spanning multiple subregions. For example, the probability of finding at least one dot within 4 subregions is \(P = 10/(6.25 \times 6.25) = 0.256\). As a result, the expected number of occurrences of nonzero terms for a 2 \(\times 2\) quadratic kernel is \(N \times P = 1,174\), which is similar to a 1 \(\times 1\) linear kernel spanning just one subregion. For our analysis linear kernels spanning 1 \(\times 1\), 2 \(\times 2\), 3 \(\times 3\), 4 \(\times 4\), and 5 \(\times 5\) subregions and quadratic kernels spanning 2 \(\times 2\), 3 \(\times 3\), 4 \(\times 4\), and 5 \(\times 5\) subregions were included in the dictionary for matching pursuit; 1 \(\times 1\) quadratic kernels were excluded because of their low prevalence.

**Multiple Instances of Linear and Quadratic Models via Randomized Matching Pursuit**

Matching pursuit is an iterative algorithm that finds the best-fitting terms at any iteration from a dictionary of kernels. At the end of each iteration, it projects the remaining terms in the dictionary into a space orthogonal to the response vector, so that at subsequent iterations the algorithm picks up terms that best capture the portion of the response vector not accounted for up to the current iteration (see Appendix). Given the iterative nature of the algorithm, a noisy term picked up at any step may be noisy and may be selected only on a few runs since the noise component of the response is random. To minimize the effects of noisy terms being picked up at any given step, we obtained multiple instances of each type of model by using randomization. Rather than picking the kernel with the maximum projection at every iteration, we randomly selected a kernel with projection \(\geq 75\%\) of the maximum projection. Because of the randomization, the algorithm selects a different sequence of vectors at each instance. We hypothesize that the signal terms or kernels are selected more often across the different instances since they should depend on the systematic component of the response common across all data sets, whereas a given noisy kernel will be selected only on a few runs since the noise component of the response will be dependent on the current rate vector at that iteration. The final weights obtained for each type of model were averaged across 25 such instances.

**Expressing Second-Order Terms as Quadratic Filters**

The excitatory and inhibitory effects of a linear RF can be easily visualized by plotting the weights at the corresponding spatial locations. However, second-order interactions between different spatial locations cannot be visualized directly from the quadratic terms. To obtain a visual display of the pattern of interactions between various spatial locations, we obtained an equivalent representation of Eq. 2 in terms of 2D quadratic filters. Specifically, Eq. 2 can be rewritten in matrix form:

\[
e_n = w_0 + A_n^T W_L + A_n^T W Q A_n
\]

where

\[
A_n = \begin{bmatrix}
  x_1(n) \\
  x_2(n) \\
  : \\
  x_{432}(n)
\end{bmatrix},
W_L = \begin{bmatrix}
w_1 \\
  w_2 \\
  : \\
  w_{432}
\end{bmatrix}
\quad \text{and } W_Q = \begin{bmatrix}
w_{11} & w_{12}/2 & \cdots & w_{1,432}/2 \\
  w_{12}/2 & w_{22} & \cdots & w_{2,432}/2 \\
  : & : & \cdots & : \\
  w_{1,432}/2 & w_{2,432}/2 & \cdots & w_{432,432}
\end{bmatrix}
\]

\(W_Q\) is known as the Hessian matrix and represents the interactive effect between pairs of subregions in a compact matrix form.

To obtain a visual display of the pattern of interactions between various spatial locations, an equivalent representation of the Hessian matrix can be obtained in terms of its eigenvalues (\(\lambda_i\)) and eigenvectors (\(q_i\)):

\[
W_Q = \sum_{i=1}^{432} \lambda_i q_i q_i^T
\]

Substituting in Eq. 3,

\[
e_n = w_0 + A_n^T W_L + \sum_{i=1}^{432} \lambda_i (A_n^T q_i)^2 = w_0 + A_n^T W_L + \sum_{i=1}^{432} \frac{\lambda_i}{|\lambda_i|} [A_n^T (\sqrt{|\lambda_i|} q_i)]^2
\]

The term \(\frac{\lambda_i}{|\lambda_i|}\) is 1 if \(\lambda_i\) is positive and \(-1\) if negative and determines whether the contribution of a particular eigenvector is suppressive (negative) or facilitative (positive).

Thus the eigenvectors, scaled by the root of the absolute value of the corresponding eigenvalue (i.e., \(\sqrt{|\lambda_i|}\)), can be considered as quadratic filters that contribute to the response by the square of the convolution of these filters with the input stimulus pattern. Because of the eventual squaring operation, the overall sign of any quadratic filter is immaterial. (If a quadratic filter is multiplied by \(-1\), so that the signs of the weights of all subregions are reversed, the contribution of the filter to the response remains unchanged.) Thus the effect of a given quadratic filter depends on the sign and magnitude of the corresponding eigenvalue, as well as the relative signs of the weights.
of different subregions in the quadratic filter. If a pair of subregions in the quadratic filter has the same sign, then a negative eigenvalue implies suppressive interaction, whereas a positive eigenvalue implies facilitative interaction, whereas a positive eigenvalue implies a suppressive interaction (see APPENDIX for details on interpretation). These quadratic filters are analogous to second-order RF structures obtained via second-order methods such as spike-triggered covariance.

Separating Within-Digit and Cross-Digit Interactions

Instead of obtaining quadratic filters from the Hessian matrix directly, we transformed the matrix in such a way that it enables a direct and straightforward quantification, comparison, and interpretation of the different types of within-digit and cross-digit interactions. For every neuron we classified each of the three pads as (1) maximally responsive digit ($D_{\text{max}}$; defined as the pad containing the subregion giving the highest linear response), (2) digit(s) adjacent to the maximally responsive digit ($D_{\text{adj}}$), and (3) nonadjacent digit ($D_{\text{nonadj}}$). The quadratic weights representing the interaction between two subregions were sorted into six classes of within-digit or cross-digit interactions depending upon the locations of the two subregions on the digits, namely, 1) $D_{\text{max}} \times D_{\text{max}}$; both the subregions on $D_{\text{max}}$; 2) $D_{\text{adj}} \times D_{\text{adj}}$; both the subregions on $D_{\text{adj}}$; 3) $D_{\text{nonadj}} \times D_{\text{nonadj}}$; both the subregions on $D_{\text{nonadj}}$; 4) $D_{\text{max}} \times D_{\text{adj}}$; one subregion each on $D_{\text{max}}$ and $D_{\text{adj}}$; 5) $D_{\text{max}} \times D_{\text{nonadj}}$; one subregion each on $D_{\text{max}}$ and $D_{\text{nonadj}}$, and 6) $D_{\text{adj}} \times D_{\text{nonadj}}$; one subregion each on $D_{\text{adj}}$ and $D_{\text{nonadj}}$.

Next, the Hessian matrix was expressed as a combination of six component matrices that represent specific types of within-digit or cross-digit interaction. Specifically,

$$W_0 = W_{D_{\text{max}} \times D_{\text{max}}} + W_{D_{\text{adj}} \times D_{\text{adj}}} + W_{D_{\text{nonadj}} \times D_{\text{nonadj}}} + W_{D_{\text{max}} \times D_{\text{adj}}} + W_{D_{\text{adj}} \times D_{\text{nonadj}}} + W_{D_{\text{max}} \times D_{\text{nonadj}}}$$

The $(i,j)$th element of the component matrix $W_{D_{\text{max}} \times D_{\text{adj}}}$ (which represents the interaction between digits $y$ and $z$) is equal to the corresponding element of $W_0$ if subregion $i$ belongs to digit $y$ and subregion $j$ belongs to digit $z$ or subregion $i$ belongs to digit $y$ and subregion $j$ belongs to digit $y$ and is zero otherwise. Thus this is a simple redistribution or sorting of the various terms of $W_0$ into six groups (6 matrices) based on the location of the subregions that each term represents. Each component matrix, when subjected to an eigenvector/eigenvalue decomposition, yields quadratic filters that capture a specific type of interaction represented by the parent component matrix $W_{D_{\text{max}} \times D_{\text{adj}}}$.

The quadratic filters that correspond to within-pad interactions on the maximally responsive digit ($D_{\text{max}} \times D_{\text{max}}$) were obtained by setting all the quadratic weights, except the weights representing the $D_{\text{max}} \times D_{\text{max}}$ type of interactions, to zero before decomposing the Hessian matrix into eigenvalues and eigenvectors. Similarly, the quadratic filters representing the $D_{\text{max}} \times D_{\text{adj}}$ type of interactions were obtained by setting all the quadratic weights, except the weights representing the $D_{\text{max}} \times D_{\text{adj}}$ type of interactions, to zero before decomposing the Hessian matrix.

Comparing Strengths of Different Types of Interactions

The relative strength of the six classes of interactions was quantified by the ratio of the sum of squares of the quadratic weights belonging to a particular type of interaction to the sum of squares of all the quadratic weights (called fraction of quadratic energy captured by a particular type of interaction). The significance of each type of interaction was assessed by comparing the distribution with a null distribution obtained by repeating the entire model estimation procedure with the stimulus pattern on the adjacent digits shuffled along the time axis in the regression equations.

Strength of a Quadratic Filter

We again used measures based upon the sum of squares of quadratic weights to characterize the fraction of total quadratic strength accounted for by a given quadratic filter. If $W_{D_{\text{max}} \times D_{\text{adj}}}$ represents the Hessian matrix constructed from the weights corresponding to a particular type of interaction ($D_{\text{max}} \times D_{\text{adj}}$), then the sum of squares of all the weights corresponding to this particular interaction can be written as

$$\|W_{D_{\text{max}} \times D_{\text{adj}}}\|^2 = \|W_{D_{\text{max}} \times D_{\text{adj}}}^T W_{D_{\text{max}} \times D_{\text{adj}}} - \left( \sum_{i=1}^N \lambda_i q_i q_i^T \right) \|^2 = \sum_{i=1}^N \lambda_i^2$$

Also, the norm of the projection of a quadratic filter (defined as $QF_i^T QF_j$) is $\lambda_i^2$. Hence, the fraction of the quadratic strength of a component matrix accounted for by a given quadratic filter can be measured by the ratio $\frac{\sum_{i=1}^N \lambda_i^2}{\lambda_1^2}$. $\lambda_1$ indicates that all of the nonzero pixels of that quadratic filter have equal strengths, whereas $HI = 1$ indicates that all of the non-zero pixels have similar signs.

The strength of the contribution of a quadratic filter was measured as the ratio of the square of the corresponding eigenvalue to the sum of squares of all the quadratic weights representing that particular type of interaction.

Cross Validation

The data set was divided into two separate sets for the purposes of estimating free parameters (80%) and cross-validation (20%). We used the fraction of the variance in the 20% of data set aside that can be accounted for by a given model as a measure of predictive $r^2$. Specifically, if $y_i$ is the observed response at the $i$th time step and $\hat{y}_i$ is the response predicted by a given model, we obtained the total sum of squares (TSS) and the residual sum of squares (RSS) as

$$\text{TSS} = \sum_{i=1}^N \left( y_i - \frac{1}{N} \sum_{j=1}^N y_j \right)^2$$

$$\text{RSS} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

In addition, we obtained an estimate of the variance attributed to noise (NSS) by assuming successive scans as repeats (see DiCarlo et al. 1998 for details).

$$\text{NSS} = \sum_{i=1}^N \text{Var}(Y_i)$$

where $Y_i = [y_i, y_{i-p}, y_{i-p+1}, \ldots]$ Here $T$ is the number of steps in each scan. Predictive $r^2$ was defined as...
The procedure was repeated 150 times, using a different set of 20% of the data each time. For each combination of 80–20 (data estimation-cross validation), model response $\hat{y}_i$ was averaged across 25 instances of the model obtained by randomization (see above). The final predictive $r^2$, linear and quadratic weights, and the predicted response reported in this article were averaged across all 150 repeats.

Linear Model with Spiking History

To test whether the improvements observed with a quadratic model over a linear model could be attributable to known neural nonlinearities such as spiking history (Paninski et al. 2004; Pillow and Simoncelli 2003; Truccolo et al. 2005, 2010), we also evaluated the goodness of fit of the linear model (Eq. 2) with terms added to capture spiking history. Specifically, the linear models derived by matching pursuit as described above were expanded by including a linear combination of terms denoting the effects of spiking history at a delay $\tau$ as follows:

$$e_n = w_0 + \sum x_i(n)w_i + \sum_{i=1}^{\tau_{\text{max}}} w_{hi} \cdot r_{n-\tau}$$

Here $w_{hi}$ captured the weight/contribution of the neural firing at a history of $\tau$ ms.

The model was separately evaluated for multiple levels of spiking history such as $\tau_{\text{max}} = 32$ ms, $\tau_{\text{max}} = 160$ ms, and $\tau_{\text{max}} = 320$ ms by fitting a GLM, again yielding a log-linear model in linear terms and spiking history [i.e., $r_n \sim \text{Poisson}(e^n)$]. Note that unlike previous models, which let the algorithm iteratively pick salient terms from a dictionary of potential candidates, we simply appended terms for spiking history to the linear model obtained earlier (i.e., Eq. 1). As a result, if spiking history does not contribute to neuronal firing, it may lead to overfitting and hence poorer performance on cross validation.

RESULTS

Comparing Responses Predicted by Linear and Quadratic Receptive Fields

Figure 2 shows the responses of the linear model and the quadratic model (consisting of linear and quadratic terms as described in Eq. 2) of a typical 3b neuron. The linear RF (Fig. 2A) indicates that the major drive for this neuron is from D3. Consistent with previous studies (DiCarlo et al. 1998), the linear RF on this digit comprises an excitatory region (red) surrounded by an inhibitory region (blue). The inhibitory region is shifted distally on the finger pad with respect to the excitatory region, which could be due to spatial effects or due to inhibition that is delayed with respect to the excitation (DiCarlo et al. 1998; Gardner and Costanzo 1980b). Since the model fitting procedure correlates the responses with the instantaneous pattern of dots on the finger pad, this temporally trailing inhibition can only manifest in the linear RF by being shifted distally along the scan direction. However, trailing inhibition is an interactive effect between two dots separated along the temporal axis (Gardner and Costanzo 1980b). Since a linear RF, by definition, is restricted to just the linear terms, the fitting procedure accounts for this inhibitory effect by means of negative weights (DiCarlo et al. 1998) in regions on the pad surrounding the excitatory spot. When quadratic terms are included in the model, the linear component is composed solely of an excitatory area (Fig. 2B), with the trailing and infield inhibition being captured by the quadratic terms.

The first raster of Fig. 2C displays the observed response as a raw spatial event plot (SEP, consisting of 30 successive scans of the drum plotted as 30 rows). Note that the raw SEP reflects neuronal all-or-none firings, which are probabilistic events. To visualize the underlying firing rate, the second raster plot displays the same SEP convolved with a 2D Gaussian ($\sigma = 1$ pixel). The next two raster plots correspond to the responses of the linear (4th raster) and quadratic (3rd raster) RF over the data set. The bottom band (5th raster) displays the difference between the responses of the quadratic and the linear RFs. Regions where the quadratic model outputs a higher (lower) response than the linear model are indicated in red (blue). Overall, both the linear and quadratic responses qualitatively match well with the observed responses, suggesting that the linear models capture much of the primary responses of the neuron. However, at a finer scale, there are significant differences between the predictions of the two models. Examples of such differences are illustrated via enlarged images of three regions (namely, $A$, $B$, and $C$) in Fig. 2D.

The number of dots stimulating the excitatory region of the RF is higher at region $B$ than region $A$. Hence the linear RF outputs a proportionately higher response at bins around $B$ (compare the overall intensity of the top 2 panels in the 4th column of Fig. 2D). However, as indicated by the convolved SEP of Fig. 2D, the observed responses are unaffected by the dot density. This is because, unlike peripheral neural responses that increase their firing rate logarithmically as the average dot density within the RF increases, area 3b neurons are largely unaffected by the average dot density (Sripati et al. 2006). The quadratic terms correct for this difference by increasing the model response around $A$ (red region in the top panel in the last column of Fig. 2D) and lowering the model response around $B$ (blue region in the center of last column of Fig. 2D) such that the responses are similar around $A$ and $B$, which matches the similarity of observed responses around $A$ and $B$. As a result, the quadratic RF model provides a closer match to the observed responses and better mimics the invariance in the neural responses to changes in average dot density.

Region $C$ in Fig. 2D corresponds to a time step when the number of dots in the inhibitory region of the linear RF exceeds the number of dots in the excitatory region. As indicated by the bottom panel of the fourth column in Fig. 2D, the linear RF predicts no response around this region in the raster, although the observed response shows neuronal firing (1st and 2nd columns, last row). This suggests that the inhibition in the linear RF is a result of a linear approximation to trailing inhibitory effects, which is in fact an interactive or quadratic effect. A quadratic filter is better at capturing the inhibition by second-order weights, and hence the linear component of the quadratic RF (Fig. 2B) differs from the purely linear model and lacks the distal inhibitory region. The result is a better match to the observed data (bottom panel of the 3rd column in Fig. 2D), as the quadratic model predicts a nonzero response around $C$ similar to the observed response.

Characterization of Linear Component of Quadratic Field

The linear components were classified based on the strengths of the positive and negative weights within the three stimulated
pads, with respect to the maximum weight. For most neurons (65/80), the negative weights within the dominant digit or the weights on the nondominant digits did not exceed 10% of the maximum weight, indicating that the linear response is dominated by a single excitatory region exclusively on a single pad (Fig. 3A, left). A few neurons had an inhibitory region surrounding the excitatory region (5/80) (Fig. 3A, 2nd from left), another excitatory region on an adjacent pad (multidigit) (2/80) (Fig. 3A, center), distinct multiple regions of excitation on a single pad (1/80) (Fig. 3A, 2nd from right), an inhibitory region on some other pad (2/80) (Fig. 3A, right), or no linear excitatory component at all (5/80) (not shown). Figure 3B displays the
The relative strength of each type of interaction is shown in Fig. 4A. As expected, the fraction of quadratic strength within a single digit is largest for the maximally responsive digit (34%). The adjacent and nonadjacent digits by themselves account for a small fraction of the total quadratic strength (9% and 5%, respectively). This is reasonable since neurons in area 3b are classically thought to have RFs confined to a single digit. Surprisingly, a large fraction of the total quadratic strength is accounted for by interactions between digits (45%).

Testing Validity of Model Fits

In any system identification framework, the response obtained to any set of inputs can be characterized as having a systematic or driven component and a noise component. The aim of a system identification procedure is to capture as much of the systematic component without fitting the noise. Models can be tested for noise fitting with cross validation. However, given that we use matching pursuit with a very large dictionary of potential terms there is a possibility that the algorithm will find some combination of terms that fits the systematic component that does not have any functional meaning.

To test whether the quadratic terms that we observed are functionally relevant or are an artifact of blind curve-fitting, we obtained a null distribution of fractions of quadratic energies obtained to any set of inputs can be characterized as having a systematic or driven component and a noise component. The aim of a system identification procedure is to capture as much of the systematic component without fitting the noise. Models can be tested for noise fitting with cross validation. However, given that we use matching pursuit with a very large dictionary of potential terms there is a possibility that the algorithm will find some combination of terms that fits the systematic component that does not have any functional meaning.

Testing Validity of Model Fits

In any system identification framework, the response obtained to any set of inputs can be characterized as having a systematic or driven component and a noise component. The aim of a system identification procedure is to capture as much of the systematic component without fitting the noise. Models can be tested for noise fitting with cross validation. However, given that we use matching pursuit with a very large dictionary of potential terms there is a possibility that the algorithm will find some combination of terms that fits the systematic component that does not have any functional meaning.

Testing Validity of Model Fits

In any system identification framework, the response obtained to any set of inputs can be characterized as having a systematic or driven component and a noise component. The aim of a system identification procedure is to capture as much of the systematic component without fitting the noise. Models can be tested for noise fitting with cross validation. However, given that we use matching pursuit with a very large dictionary of potential terms there is a possibility that the algorithm will find some combination of terms that fits the systematic component that does not have any functional meaning.

Testing Validity of Model Fits

In any system identification framework, the response obtained to any set of inputs can be characterized as having a systematic or driven component and a noise component. The aim of a system identification procedure is to capture as much of the systematic component without fitting the noise. Models can be tested for noise fitting with cross validation. However, given that we use matching pursuit with a very large dictionary of potential terms there is a possibility that the algorithm will find some combination of terms that fits the systematic component that does not have any functional meaning.
terms related to $D_{adj}$ and $D_{nonadj}$ that are picked up after the shuffling are most likely a result of blind curve-fitting.

Figure 4B compares the null distribution (red curve) and the actual distributions (blue curves) of quadratic energies for the five types of interactions related to $D_{adj}$ and $D_{nonadj}$. The null and actual distributions for $D_{max} \times D_{adj}$, $D_{max} \times D_{nonadj}$, $D_{nonadj} \times D_{nonadj}$, and $D_{adj} \times D_{adj}$ types of interactions are significantly different ($P < 0.0001$; paired $t$-test), suggesting that these classes of quadratic interactions we observed are not fully attributable to noise. The null and actual distributions for the $D_{nonadj} \times D_{nonadj}$ type of interaction were not significantly different ($P > 0.4$; paired $t$-test) (Fig. 4B, top right), indicating that the within-digit interaction on the nonadjacent digits is small and is no better than the fit to the noise itself.

**Within-Digit Quadratic Component on the Maximally Responsive Digit**

We characterized the quadratic component on $D_{max}$ by estimating the quadratic filters corresponding to the most negative (termed *filter Q1*), and most positive (termed *filter
Filter Q1. Figure 5A shows a distribution of the RF homogeneity index (HI), which measures the relative strengths of regions with positive and negative signs within the quadratic filter (see METHODS) for filter Q1 across all the neurons. The distribution is significantly skewed toward 1 ($H_0$: median = 0.5, $P < 0.0001$, Wilcoxon signed-rank test), indicating that for a majority of the neurons the nonzero subregions of filter Q1 tend to have similar signs. This is further illustrated by plotting filter Q1 for three sample neurons from three different locations on the distribution. Neuron C, which is drawn from a region with high HI (HI = 1), is characterized by a single large region with a negative sign. Neuron B, drawn from the center (HI = 0.55), again has a dominant large region with a positive sign, flanked by a small region of negative sign on the proximal side. Neuron A, which represents a small population of neurons with low HI (HI = 0.17), has two regions with opposite signs aligned along the proximal to distal or scan direction.

Filter Q1 was further characterized by segmenting the largest contiguous lobe. Figure 5B, which plots the offset between the center of this lobe and the center of the corresponding linear component, indicates that for most neurons the center of this lobe spatially overlaps with the linear component (mean $x$ offset $0.33 \pm 0.94$ mm: mean $y$ offset $0.14 \pm 1.13$ mm). The mean area across the population (Fig. 5C) is 51.4 mm$^2$, which is significantly greater than the area spanned by the linear component (mean area 10.87 mm$^2$) ($P < 0.0001$; paired $t$-test). This is further illustrated in Fig. 5D, which plots the ratio of the two areas. The linear component spans only $\sim 30\%$ of the area spanned by the largest similarly signed contiguous region of Q1 across the population.

We characterized the shape of this region by fitting an ellipse and estimating the eccentricity (Fig. 5E) and orientation (Fig. 5F).

Fig. 5. Characterization of filter (Q1) corresponding to most negative eigenvalue of $D_{\text{max}} \times D_{\text{max}}$ interaction matrix. A: distribution of homogeneity index that measured the relative strengths of positive and negative regions within the filter on a scale of 0 (balance between positive and negative regions) to 1 (dominated by a single sign). Sample filters from 3 different ranges of the distribution are shown at bottom. B: difference between the centers of the single largest lobe of this quadratic filter and the center of the excitatory component of the linear component. C: distribution of area of the single largest lobe. D: distribution of the ratios of the area of linear component to the area of the single largest lobe of filter Q1. E: distribution of the eccentricities of ellipses fitted to the single largest lobe of filter Q1 across the population. F: distribution of orientation of the fitted ellipse. G: distribution of the fraction of $D_{\text{max}} \times D_{\text{max}}$ quadratic energy captured by filter Q1, measured as the ratio of the square of the eigenvalue associated with this filter (equals the Frobenius norm of the projection matrix associated with this filter) to the sum of squares of all $D_{\text{max}} \times D_{\text{max}}$ quadratic weights (see APPENDIX).
5F) of the fitted ellipses. The mean eccentricity across the population is 0.17, which is significantly higher than the mean eccentricity of the linear component ($P < 0.0001$; paired t-test). The distribution of orientation of the major axis had a peak at $90^\circ$, which corresponds to the distal-proximal axis, and the distribution is significantly different from a uniform circular distribution (Raleigh test, $P < 0.01$). Figure 5G shows a distribution of the fraction of the $D_{\text{max}} \times D_{\text{max}}$ type of quadratic strength accounted for by this filter (see METHODS). On average, this filter accounts for $\sim 59\%$ of the total within-pad quadratic strength on the maximally responsive digit.

Thus the salient characteristic of filter $Q1$ across the population was a single contiguous region, either positive or negative, that was larger in size, more elliptical, and spatially overlapped with the linear component of the quadratic RF.

Filter $Q2$. In contrast to $Q1$, the distribution of the homogeneity index for filter $Q2$ (Fig. 6A) is significantly skewed toward 0 ($H_{0.5}; \text{median} = 0.5, P < 0.0001$, Wilcoxon signed-rank test), indicating that for a majority of the neurons the strengths of regions with positive and negative signs are balanced. This is illustrated by plotting filter $Q2$ for three sample neurons from three different locations on the distribution.

Neurons $A$ and $B$, which are drawn from regions with low to moderate $HIs$ (0.14 and 0.53, respectively) and are representative of the majority of the neurons, are characterized by a positive and a negative lobe each aligned nearly along the distal-proximal axis. Neuron $B$, drawn from the center ($HI = 0.53$), has, in addition to the two lobes, a few small regions in the surrounding space. Neuron $C$, which represents the small population of neurons with high $HI$ ($HI = 0.78$), has a large positive region flanked by a few weak negative regions.

Filter $Q2$ was characterized by segmenting the largest contiguous positive and negative lobes. We measured the fraction of the total $Q2$ filter structure captured by these two lobes by taking the ratio of the sum of squares of weights of the pixels within the two lobes to the sum of squares of all nonzero weights within the filter (Fig. 6B). Across the population, the two lobes capture $\sim 86\%$ of the total structure of filter $Q2$. Figure 6C, which plots the offset between the mean of the centers of the two lobes and the center of the corresponding linear component, indicates that, like filter $Q1$, for most neurons the two lobes spatially overlap with the linear component (mean $x$ offset $0.58 \pm 1.4$ mm; mean $y$ offset $0.5 \pm 1.7$ mm).

Figure 6D shows the distribution of orientation of the line

---

**Fig. 6.** Characterization of filter ($Q2$) corresponding to most positive eigenvalue of $D_{\text{max}} \times D_{\text{max}}$ interaction matrix. $A$: distribution of homogeneity index that measured the relative strengths of positive and negative regions within the filter on a scale of 0 (balance between positive and negative regions) to 1 (dominated by a single sign). Sample filters from 3 different ranges of the distribution are shown at bottom. $B$: fraction of the total structure on this filter captured by the single largest positive and single largest negative lobe. $C$: difference between the mean of the centers of largest positive and largest negative lobes of filter $Q2$ and the center of the excitatory component of the linear component. $D$: orientations of the line joining the centers of the 2 lobes. $E$: distribution of the fraction of $D_{\text{max}} \times D_{\text{max}}$ quadratic energy captured by filter $Q2$. $F$: distribution of the fraction of $D_{\text{max}} \times D_{\text{max}}$ quadratic energy captured by both filters $Q1$ and $Q2$.
joining the centers of the two lobes. The distribution has a peak at 90°, which corresponds to the distal-proximal axis, and the distribution is significantly different from a uniform circular distribution (Raleigh test, \( P < 0.05 \)). On average, this filter accounts for \(~12\%\) of the total within-pad quadratic strength on the maximally responsive digit (Fig. 6E). Filters \( Q1 \) and \( Q2 \) together capture \(~71\%\) of the total quadratic energy on the maximally responsive digit (Fig. 6F).

Thus filter \( Q2 \) was characterized across the population by a pair of lobes, opposite in sign and aligned along the proximal-distal axis or the scan direction.

**Functional Roles of Quadratic Filters**

To illustrate the roles played by the filters characterized above, we compared output responses of the linear component and the two quadratic filters \( Q1 \) and \( Q2 \) with the observed responses for a sample neuron (Fig. 7). The outputs for the quadratic filters are obtained by squaring the convolution of the filters with the instantaneous stimulus patterns scaled by the corresponding eigenvalue (see METHODS). Since the squaring operation always results in a positive number, the sign of the output response (i.e., whether the effect is facilitative or
suppressive) is determined by the sign of the corresponding eigenvalue. \(Q1\) provides a suppressive interaction, which is evident by the negative contributions (blue) shown in Fig. 7 (4th raster). In contrast, the contribution of \(Q2\) lacks any negative predictions (blue color) because of the positive sign of the eigenvalue (Fig. 7, 5th raster).

The roles of the three filters in shaping the neural response are further illustrated via three time bins \(A\), \(B\), and \(C\). The panels on the top right and bottom of Fig. 7 show the stimulus patterns during these time bins overlaid with the linear component and the two quadratic filters for the sample neuron. The corresponding locations of these time points are highlighted by black circles on each raster. The neuron is stimulated by a single dot at \(A\) but multiple dots at \(B\). Since a linear filter linearly scales its output with the number of dots, the response of the linear component is significantly higher at \(B\) than \(A\) (compare the 2 left circles on the 3rd raster in Fig. 7).

However, the observed responses around these bins, as highlighted by the corresponding circles on the first raster, indicate that the neuron is slightly inhibited around \(B\) compared with \(A\). Filter \(Q1\) corrects for this discrepancy by contributing a strong suppressive response at \(B\) (compare the corresponding circles on the 3rd raster).

The example denoted by region \(C\) also corresponds to a time bin when the stimulus consisted of multiple dots in the responsive region. However, unlike \(B\), which corresponds to an instance when the multiple dots are spread evenly between the two lobes of filter \(Q2\), \(C\) corresponds to an instance when the multiple dots were located within one of the two lobes of filter \(Q2\). \(Q2\) provides no response at \(B\) but provides a strong facilitative contribution at \(C\) (last 2 panels in last column, Fig. 7). Thus \(Q2\) masks or reduces the suppression provided by \(Q1\) if a pair of dots is located within one of the two lobes of \(Q2\) but leaves the suppression unaffected if the two dots happen to be in the two lobes. For the majority of the neurons, the two lobes of \(Q2\) are nearly aligned along the proximal-distal direction (Fig. 7), which is also the scan direction. This shows that filter \(Q2\) does not mask the suppression provided by \(Q1\) if the pair of dots is aligned along the scan direction.

Thus filter \(Q1\) provides a nonspecific suppression in a region overlapping with the neuronal RF that scales quadratically with dot density within the RF and offsets the linear scaling with dot density provided by the linear component of the RF. Filter \(Q2\) modulates the suppression provided by \(Q1\) depending upon where the dots are located with respect to the scan axis. It reduces the suppression provided by \(Q1\) if the multiple dots are located orthogonal to the scan axis but leaves the suppression unaffected if the dots are located along the scan axis. In other words, filter \(Q1\) captures the spatial extent of infield inhibition, whereas filter \(Q2\) captures the temporal or trailing effects of inhibition along the scan direction.

**Characterizing Quadratic Interaction Between Digits**

As mentioned above, the interaction between the maximally responsive digit and its adjacent digits was the second largest quadratic component across the population. In this section, we characterize this component by looking at the quadratic filters corresponding to the most positive and most negative eigenvalues of the Hessian matrix that represents this type of interaction. These filters are complementary with respect to each other in the sense that they have identical spatial structure, except that the overall sign on one of the digits is reversed (see sample examples in Fig. 8A). Thus we will characterize the spatial properties of just one of the two filters.

Figure 8A shows the distribution of the homogeneity index for the components of this filter on the maximally responsive digit (top) and the adjacent digit (bottom). The two distributions are significantly skewed toward 1 \((P < 0.0001, \text{Wilcoxon rank sum test})\), indicating that for the majority of the neurons the components of these filters on each pad tend to have similar signs. If the components on the two digits are opposite for the filter corresponding to the most negative eigenvalue (or similar for the filter corresponding to the most positive eigenvalue, since the two filters have complementary structures), it indicates facilitative interaction (see **Appendix** for interpretation).

The majority of neurons with a linear excitatory component on the linear component to the area spanned by the largest lobe of this filter on the maximally responsive digit, indicates that the linear component spans only \(~33\%\) of the area spanned by the largest lobe. Figure 8C plots the difference between the center of the largest lobe on the maximally responsive digit and the center of the corresponding linear component. Compared with filters \(Q1\) and \(Q2\) described above, the largest lobe of this filter has a greater offset with respect to the linear component (mean \(x\) offset 0.66 ± 1.79 mm, mean \(y\) offset 0.82 ± 2.2 mm). Across the population, these filters capture \(~30\%\) of the total \(D_{\text{max}} \times D_{\text{adj}}\) type of interaction (Fig. 8D).

Figure 8E illustrates the contribution of the multidigit quadratic terms in the response SEP of a sample neuron. The top two rasters plot the raw SEP and an estimate of the underlying firing rate [obtained by convolving with a 2D Gaussian \((\sigma = 1 \text{ pixel})\)], while the third and fourth rasters plot the outputs of single-digit quadratic and multidigit quadratic models, respectively. The last raster we plot the difference between the outputs of the two models. Red spots (highlighted by blue circles) on the last raster indicate regions on the SEP where the multidigit RF response was significantly greater than the single-digit RF. The corresponding locations on the neuronal response (highlighted by red circles) with the difference plot shows that a majority of the differences correlate with regions with bursts of high firing. This indicates that the presence of coincident stimuli on the maximally responsive digit and the adjacent digit (activating the \(D_{\text{max}} \times D_{\text{adj}}\) interaction terms) tends to increase the local firing rate to specific features.

**Predictive Power of Models**

We used cross validation to compare the predictive power of the quadratic model with that of the linear model by predicting responses to 20% of the data based on weights estimated from 80% of the data. Because of the approximation built into the estimate of the portion of the total sum of squares attributable to noise (i.e., NSS; see **Methods**), individual estimates of predictive \(r^2\) may not be bounded between -1 and 1 like traditional regression. For example, individual samples in Fig. 9A, which shows the distribution of 150 samples of predictive \(r^2\) for the linear (blue) and quadratic (red) models for an example neuron, may
Fig. 9. Cross validation performance of the linear and quadratic RFs. The weights were estimated using 80% of the data. Cross validation was measured as the fraction of the variance of the remaining 20% of data that is accounted for by a given model. A shows the distribution of predictive $r^2$ for the linear (blue) and quadratic (red) models for a sample neuron across the 150 repeats using a unique set of 20% data for prediction over each repeat. B compares the average (median) predictive $r^2$ for the linear ($x$-axis) and quadratic ($y$-axis) models across all neurons. C compares the average number of terms (free parameters) in the linear ($x$-axis) and quadratic ($y$-axis) models. D compares the predictive $r^2$ between the original linear model with an extended version of the linear model with number of free parameters matched to the corresponding quadratic model for that neuron. E compares the predictive $r^2$ for the original linear model with versions of the linear model that incorporated spiking history over 32 ms (red), 160 ms (green), and 320 ms (blue).

Fig. 8. Characterization of the interaction between the maximally responsive digit and its adjacent digit. A: distribution of homogeneity index linear components of the quadratic RF for 2 sample neurons. For the 2 neurons, quadratic filters corresponding to the most negative eigenvalue (1st column) and the most positive eigenvalue (2nd column) of the $D_{max} \times D_{adj}$ component matrix are shown together with the corresponding linear component (3rd column). B: ratio of the area of the quadratic filters on the maximally responsive digit to the excitatory area of the linear component on the maximally responsive digit. C: difference between the centers of the part of the quadratic filter on the maximally responsive digit and the centers of the excitatory regions of the linear component. D: distribution of the quadratic strength of these filters, measured as the ratio of the sum of the squares of the eigenvalues associated with these filters for a neuron (equals the Frobenius norm of the projection matrix associated with this filter) to the square of the norm of the $D_{max} \times D_{adj}$ component matrix. E: raw SEP for a sample response showing observed response (1st raster), an estimate of the underlying firing rate obtained by convolving the SEP with a 2D Gaussian ($\sigma = 1$ pixel) (2nd raster), single-digit RF output (3rd raster), and multidigit RF output (4th raster). The last raster plots the difference between the outputs of the multidigit and single-digit RFs.
model cannot be attributed to differences in order of the models.

To test whether the quadratic effect is attributable to known neural nonlinearities such as spiking history, we compared the original linear model with variants of the linear model with terms explicitly accounting for spiking history (Fig. 9E) up to 32 ms (red), 160 ms (green), and 320 ms (blue). Adding terms to account for spiking history degrades the model, resulting in poorer fits and indicating that the interactive effects in area 3b neurons captured by the quadratic terms are not attributable to spiking history.

**DISCUSSION**

In this study, we investigated the nonlinear interactions in the responses of area 3b neurons by estimating quadratic RFs with linear and quadratic terms spanning the distal pads of digits D2–4. Unlike previous studies (DiCarlo et al. 1998; Sripati et al. 2006) that used a multiple regression framework between linear terms and a neuronal spike count to estimate linear RF, we used a GLM with a log-link function that yields a log-linear relationship between the input subregions and firing rate. A GLM fit may be better given the Poisson nature of spike counts; however, we obtained similar results when we used a multiple regression framework like the previous studies (results not shown). Since the inclusion of second-order terms in the quadratic model potentially causes a dimensional explosion in the number of terms in the model, we first estimated, using a method based on matching pursuit, a subset of the quadratic terms that contribute significantly to the observed response. We find that, for a majority of the neurons, the linear component of the RF consisted of a single excitatory region on the maximally responsive digit. Furthermore, we show that there are quadratic interactions on the maximally responsive digit, as well as quadratic interactions between the maximally responsive digit and its adjacent digit(s).

About 71% of the quadratic interactions on the maximally responsive digit were captured by two filters that correspond physiologically to previously reported nonlinear inhibitory fields, namely, infield suppression and trailing inhibition (DiCarlo and Johnson 2000; Gardner 1984). Furthermore, the interaction between the maximally responsive digit and its adjacent digit(s) was characterized by a pair of filters that together represented facilitative or suppressive modulations of the response on the dominant digit.

Since the method that we have derived from matching pursuit is statistical, it cannot completely eliminate picking up terms that fit noise. Thus at least a part of the observed multidigit interactions may still be attributable to noise. However, we believe that this contribution is small because 1) the strengths of different types of quadratic energies (Fig. 4A) show a systematic reduction in interactions as we move away from the center of the RF and terms fitting to noise would tend to distribute randomly across the RF and 2) the distributions of observed fractions of quadratic energies are significantly greater than a chance distribution obtained by fitting the response to arbitrary dot patterns obtained by shuffling the dot patterns on the adjacent and nonadjacent fingers (Fig. 4B).

**Previous Studies**

Previous studies in area 3b have revealed overlapping excitatory and inhibitory inputs throughout the RF of the neuron (Alloway and Burton 1991; Dykes et al. 1984). However, infield inhibition is rarely revealed by single punctate probes (Iwamura et al. 1983; Mountcastle and Powell 1959; Sur 1980; Sur et al. 1984), which show area 3b RFs as homogeneous excitatory regions (Sur 1980). This is in agreement with the predictions of our quadratic RFs. Since a single stimulus cannot evoke interactive components, the observed RF will be dominated by the linear component of the quadratic RF, which is a single excitatory region for most neurons.

Linear RFs estimated with a scanned random dot pattern (DiCarlo et al. 1998) or multiple probes randomized in space and time (Sripati et al. 2006) do reveal a significant amount of inhibition, but a linear RF fails to capture that part of inhibition that spatially and temporally overlaps with the excitatory core. Since the excitatory core is stronger than the overlapping inhibition, the weights are positive in the overlapping region. By including quadratic terms, we have disentangled the feed-forward excitatory drive (excitatory core), captured as the linear excitatory component, and the infield inhibition, which manifests as spatially overlapping suppressive quadratic filters. These filters span an area larger than the linear component across the population, which is consistent with previous reports suggesting that the inhibition extends over an area larger than the revealed excitatory region (Alloway et al. 1989; Alloway and Burton 1991; DiCarlo et al. 1998; Gardner and Costanzo 1980b). A linear RF reveals only parts of this infield inhibition as a temporally coincident but spatially offset inhibitory component or a spatially coincident but temporally lagging component (DiCarlo and Johnson 2000; Sripati et al. 2006). Since the interactive inhibition in the overlapping region is a systematic effect that the linear models fail to capture but is accounted for by the second-order model, quadratic models are better descriptions of the response than simple linear models (Fig. 9B).

Another factor that may contribute to the infield inhibition, especially in regions of small RF sizes such as the distal pads, is skin mechanics. The role of skin mechanics is evident in the periphery as surround inhibition (Sripati et al. 2006; Vega-Bermudez and Johnson 1999) in the RFs of SA1 afferents, although there are no known synaptic mechanisms in the periphery that could lead to lateral interactions. Since these very same afferents constitute the major input of area 3b neurons, they are likely to pass on the skin mechanical effects in their response.

In addition to spatial interactions, studies have revealed a temporal effect of infield inhibition in the form of a reduced response to the second of a pair of temporally separated stimuli (Gardner and Costanzo 1980b). Since we stimulated the skin with a constant scan velocity and direction (proximal-distal), effects along the temporal axis manifest as spatial effects along the distal-proximal axis. Thus the filter corresponding to the most positive eigenvalue, which overlaps with the linear component and consists of two lobes of opposite sign separated along the distal-proximal axis, captures the temporal aspect of infield inhibition in the form of suppressive interaction between the two regions of opposite sign separated along the distal-proximal direction. We hypothesize that if the scan
direction is varied the two lobes of filter $Q2$ will move along the scan direction, whereas if the scan velocity is increased (decreased) the distance between the two lobes will increase (decrease). Furthermore, if a randomized spatiotemporal stimulus is used instead of a scanned stimulus to obtain a spatiotemporal quadratic RF, the two lobes will lie along the time axis. This is analogous to the fact that the trailing component of infield inhibition manifests (though partly) as a “replacing inhibition” along the temporal axis if a spatiotemporal stimulus is used (Sripat et al. 2006) but manifests as a spatially offset component of inhibition along the scan direction if a scanned stimulus is used (DiCarlo and Johnson 2000).

Pharmacological blockade of GABAergic transmission reveals an underlying excitatory drive from adjacent digits. This drive is attributable to intracortical lateral connections due to greater response latencies (Alloway and Burton 1991). However, given that $<10\%$ of the neurons in area 3b show multidigit responses to punctate probes (Sur 1980), the excitatory drive from adjacent digits is balanced by the inhibitory drive potentially through GABAergic interneurons. Our results suggest that this balance between lateral excitation and inhibition may be slightly offset. Such an offset may be the result of a short delay due to the extra synapse involved with GABAergic interneurons. This offset may not be sufficient to drive the neuron to a stimulus on the adjacent digit but may modulate the membrane potential so that the neuron fires more vigorously to a coincident volley of excitatory postsynaptic potentials (EPSPs) arising from the stimulation of the maximally responsive digit. Such a modulatory effect may be manifested as the quadratic interaction between the maximally responsive digit and its adjacent digits.

**Functional Significance**

Our results suggest that the transformation performed by area 3b neurons on the input spatial pattern has a significant nonlinear component, which may play an important role in shaping the feature selectivity. The linear excitatory core in area 3b is weakly elliptical and may have a role in shaping feature selectivity such as orientation selectivity. Thus orientation selectivity observed in area 3b (Bensaia et al. 2008; Hyvarinen and Poranen 1978; Pubols and Leroy 1977; Warren et al. 1986) may arise because of the eccentricities observed in the spatially overlapping quadratic filters. A bar oriented along the major axis of this filter will result in a greater interactive suppression (and weaker response) than a bar oriented along the minor axis.

The presence of a significant interaction between the maximally responsive digit and its adjacent digits implies that integration of tactile information across pads begins as early as area 3b. For a stimulus spanning multiple digits, the “extension” of the stimulus on the adjacent digits provides a contextual modulation of the response to the part of the stimulus within the dominant digit. As studies in the visual system have shown, such a contextual modulation may be critical for the representation of extended contours (Fitzpatrick 2000; Gilbert and Wiesel 1990), corners (Silito et al. 1995), or local curvature (Krieger and Zetzsche 1996; Wilson and Richards 1992) and thus provides a first step of transforming the isomorphic input from the peripheral afferents to a nonisomorphic form that underlies perception.

This cross-digit interaction may also be the basis for the RF changes observed in cortical plasticity studies (Wang et al. 1995). Conditioning through a coincident presentation of stimuli on adjacent digits could strengthen the lateral excitatory inputs from adjacent digits. This long-term potentiation will lead to an increase in the drive from lateral inputs and cause the neuron to fire to any stimulus on the adjacent digit, causing the linear component of the RF to spread to adjacent digits.

We conclude that nonlinear mechanisms play a major role in shaping somatosensory neural responses as early as area 3b. The nonlinear effects function to shape the RFs to be selective to local spatial variations in form and give neurons in area 3b selectivity to local simple features such as orientation and motion. This underlies the perception of local spatial form and RFs that form the basis for roughness perception (Hsiao et al. 2003).

**APPENDIX**

**Matching Pursuit to Select Subset of Terms with Salient Contribution to Neural Response**

The analytical tool used in this paper to circumvent the problem of “dimensional explosion” due to the inclusion of quadratic terms is an iterative algorithm derived from matching pursuit (Mallat and Zhang 1993). We assume that underlying firing rate ($R^{true}$) of the neuron governing the spiking response is related to the input ($X$) through a linear combination of some kernels $B(X) = [B_1(X), B_2(X), \ldots, B_q(X)]$.

Specifically,

$$R^{true} = \sum_{i=1}^{n} B_i(X)w_i \quad (A1)$$

Given $n$ observations in an experiment, $R^{true}$ as well as the kernels $B_i(X)$ can be thought of as $n$-dimensional vectors [with $i$th entry of the kernel vector $B_i(X)$ computed from the input stimulus pattern during the $i$th time step]. These kernels can be thought of as nonlinear feature maps that transform the input into a feature space, in which neuronal response is linear (Wu and Gallant 2004). The nature of kernels depends upon the choice of basis set over which the user wishes to express the firing rate ($R^{true}$). For example, if we want to estimate the linear RF, then $B(X)$ is simply the linear terms $X$ spanning the distal pads. If we wish to express the function in terms of a Volterra series, then $B(X)$ consists of various powers and products of powers of $X$. The model is known exactly if we identify the set of kernels $B_i(X)$ and estimate the corresponding weights $w = [w_1, w_2, \ldots, w_q]$. In general, this is a hard problem since we usually have an infinite set of possible kernels and a finite amount of data. The proposed algorithm circumvents this problem by building the model in steps from a constant term. At every iteration, the algorithm picks up and adds to $B(X)$ a kernel vector from an overcomplete dictionary $D$ that maximizes the projection of the observed spike train. At the end of each iteration, the remaining kernel vectors in $D$ are orthogonalized with respect to the selected kernel. This ensures that the dictionary always spans a space orthogonal to the selected kernel. The use of an overcomplete dictionary allows for the selection of kernels that account for the current running model. The iterations are terminated when the maximum projection at a given step is not statistically significant.

**Note 1.** The idea of orthogonalizing the entire dictionary $D$ with respect to the selected kernel is a daunting proposition. However, it can be achieved by projecting the observation vector $R$ into the orthogonal space instead (see algorithm below for details).

**Algorithm.** The spike trains collected in the experiments are converted into an array of spike count vectors (observation vector) by binning at a suitable bin interval. Let $R$ ($n \times 1$) be the “current”
observation vector (given n observations) and \( D = \{ D_1, D_2, \ldots, D_p \} \) be the dictionary of all vectors that we wish to test the observed response against. Let \( N = [\| D_1 \|, \| D_2 \|, \ldots, \| D_p \|] \) be a vector holding norms of the dictionary elements. Let \( B \) be the set of selected vectors and \( w \) be the corresponding projections. (Note that this is not the same as weights \( w \) in Eq. 1. They converge only when dictionary elements are samples of Gaussian white noise, in which case the method resembles spike-triggered averaging). The steps involved in the algorithm are as follows:

1) Project \( R \) onto the dictionary elements and divide by the norm vector to obtain the current projections.

\[
p_i = \frac{R^T D_i}{N_i}
\]

2) Choose the vector \((D_i, p_i^2)\) with maximum projection energy \((p_i^2)\) as the current best estimate. Obtain the component of \( D_i \) orthogonal to the vectors in \( B \) (call it \( D_i^\perp \)) and include it in \( B \) (i.e., \( B = [B, D_i^\perp] \)). This ensures that all vectors in \( B \) are mutually orthogonal.

3) Update the projection vector, observation vector, and norm vector as follows:

\[
\begin{align*}
w_{\text{curr}} &= \frac{R^T D_i^\perp}{\| D_i^\perp \|} \\
w_i &= [w_{\text{prev}}, w_{\text{curr}}] \\
N_i &= \sqrt{N_i^2 - (\| D_i^\perp \|^2/\| D_i^\perp \|^2)}
\end{align*}
\]

During the update, the observation vector \((R)\) is replaced by the component of \( R \) that is orthogonal to the selected vector \( D_i^\perp \). The update in norm vector ensures that at the next step we compute the projection of \( R \) on the components of the dictionary elements orthogonal to \( D_i^\perp \), without explicitly computing the orthogonal component of each of the dictionary elements.

4) If the current projection energy \((p_i^2)\) is less than a predetermined threshold (see below) then stop; otherwise, repeat the procedure by going back to step 1.

At the first iteration, we choose \( D_{\text{opt}} \), a vector of ones, as the selected vector. This ensures that the observation vector \( R \) and all the elements of the dictionary have a mean of zero from the second iteration onwards and that the vectors that are selected by the algorithm capture the variability in the response not already accounted for by the mean. This also enables us to test directly our new model against the null model that the response can be explained simply by the mean firing rate.

Stop condition. At every step, the algorithm tries to find a basis function that best captures the response vector. If unchecked, the algorithm will continue until it finds \( n \) linearly independent vectors from the dictionary that span a space containing the response vector \( R \), thus providing an exact solution to the system of equations. In theory, if \( R \) is a vector of true firing rates it will have some nonzero projection on those kernels in the dictionary that do not modulate the response. Hence, the algorithm will never pick up such vectors. However, the observation vector \( R \) consists of samples of a random variable and is noisy. It will have some nonzero projection on vectors that do not contribute to the response. The following is a theoretical framework to optimally terminate the algorithm and determines when the projection on a particular vector is not significantly better than chance.

The projection of the observation vector \( R \) on \( D_i \) is \( p = D_i^T R/\| D_i \| \). If \( R \) is sufficiently big (large number of observations), successive stimulus presentations are random, and we assume a Poisson spike count, then by the central limit theorem \( p \) tends to Gaussian with mean \( \mu = D_i^T R/\| D_i \| \) and variance \( \sigma^2 = \sum_{j=1}^{n} d_{ij}^2/\| D_j \|^2 \), where \( R^{\text{true}} = [r_1, r_2, \ldots, r_n]^T \) is the underlying true rate vector. If \( D_i \) is orthogonal to \( R^{\text{true}} \) (\( D_i \) has no role in the neuronal response), then \( \mu = 0 \) and \( p^2/\sigma^2 \) follows a \( \chi^2 \)-distribution with 1 degree of freedom.

Thus at the end of every iteration we can formulate a null hypothesis that the current estimate of rate vector \( \hat{R} \) (obtained as the best-fitting linear combination of the kernels selected up to the current iteration passed through a static nonlinearity, i.e., \( \hat{R}^T = [B w_i^T]^T \)) is the best estimate of the firing rate underlying the observed neural response. Under this null hypothesis, projection energy on any other kernel \((D_j)\) in the dictionary is simply a sample of a zero-mean \( \chi^2 \)-distribution with variance \( \sigma^2 = \sum_{j=1}^{n} d_{ij}^2/\| D_j \|^2 \). Hence, to terminate the algorithm at a particular level \( \alpha \), the maximum projection energy at any iteration can be compared to the cumulative value of this \( \chi^2 \)-distribution at a level \( \alpha \). A cutoff of \( \alpha = 0.01 \) was used in this study.

Estimating weights. Once a suitable set of basis vectors consisting of linear and nonlinear kernels \( B(X) \) is identified, weights \( w \) are obtained by fitting a GLM with a log-link function. Usually, the number of kernels picked up by the algorithm is relatively small. Hence, fitting a GLM is tractable. Note that a GLM fit yields a model that is log-linear in the selected terms, which differs from the original linear framework between firing rate and kernels used to iteratively select the subset of salient terms (Eq. A1). Alternatively, a linear least-squares method could be used to fit the observed spike count vector and selected kernels with explicit static nonlinearity, which may maintain the linear relationship between the selected terms and neuronal firing rate but may be suboptimal given the Poisson nature of spike counts.

Interpreting Quadratic Filters

The interpretation of second-order interaction implied by a given quadratic filter can be understood by considering the following simple example. Assume that we have a 3-pixel or 3-term system given by \( X = [X_1, X_2, X_3]^T \). Let us assume that the interaction between these 3 pixels is given by the following Hessian matrix:

\[
H = \begin{bmatrix}
0 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

This Hessian implies that there is a positive or facilitative interaction between pixels \( X_1 \) and \( X_3 \). An eigenvector/eigenvalue decomposition of this matrix gives two eigenvectors with nonzero eigenvalues, namely,

1. \( E_1 = \begin{bmatrix} 1, -1 \end{bmatrix}^T \) (with eigenvalue \(-1\))
2. \( E_2 = \begin{bmatrix} -1, 1 \end{bmatrix}^T \) (with eigenvalue \(1\))

Thus a facilitative quadratic interaction between two regions can manifest in two ways in the quadratic filter. If the eigenvalue is negative, then the two regions in the filter have opposite signs. If the eigenvalue is positive, then the two regions have identical signs.

Similarly, the Hessian

\[
H = \begin{bmatrix}
0 & 0 & -1 \\
0 & 0 & 0 \\
-1 & 0 & 0
\end{bmatrix}
\]

which implies a suppressive interaction between pixels \( X_1 \) and \( X_3 \), has the following two eigenvalues:

1. \( E_1 = \begin{bmatrix} 1, 1 \end{bmatrix}^T \) (with eigenvalue \(-1\))

\( J \) Neurophysiol • doi:10.1152/jn.01022.2010 • www.jn.org
2. \( E_2 = \left[ \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right] \) (with eigenvalue 1)

Thus if two regions have suppressive quadratic interaction, they have identical signs if the corresponding eigenvalue is negative and opposite signs if the eigenvalue is positive.

ACKNOWLEDGMENTS

We thank Bill Nash and Bill Quinlan for designing the drum stimulus. We thank Justin Killebrew, Eric T. Carlson, Frank Dammann, Zhicheng Lai, and Sung Soo Kim for their help in setting up the experiments and analysis.

GRANTS

This work was supported by National Institute of Neurological Disorders and Stroke Grant NS-34086.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: P.H.T., P.J.F., and S.S.H. conception and design of research; P.H.T., P.J.F., and S.S.H. performed experiments; P.H.T. analyzed data; P.H.T., P.J.F., and S.S.H. interpreted results of experiments; P.H.T. and P.J.F. prepared figures; P.H.T. drafted manuscript; P.H.T., P.J.F., and S.S.H. edited and revised manuscript; P.H.T. and S.S.H. approved final version of manuscript.

REFERENCES


