A Neurally Efficient Implementation of Sensory Population Decoding

Kris S. Chaisanguanthum^{2,3,4} and Stephen G. Lisberger^{1,2,3,4}

¹Howard Hughes Medical Institute, ²Department of Physiology, ³W. M. Keck Foundation Center for Integrative Neuroscience, ⁴Sloan-Swartz Center for Theoretical Neurobiology, University of California, San Francisco, San Francisco, California 94143

A sensory stimulus evokes activity in many neurons, creating a population response that must be “decoded” by the brain to estimate the parameters of that stimulus. Most decoding models have suggested complex neural circuits that compute optimal estimates of sensory parameters on the basis of responses in many sensory neurons. We propose a slightly suboptimal but practically simpler decoder. Decoding neurons integrate their inputs across 100 ms, incoming spikes are weighted by the preferred stimulus of the neuron of origin, and a local, cellular nonlinearity approximates divisive normalization without dividing explicitly. The suboptimal decoder includes two simplifying approximations. It uses estimates of firing rate across the population rather than computing the total population response, and it implements divisive normalization with local cellular mechanisms of single neurons rather than more complicated neural circuit mechanisms. When applied to the practical problem of estimating target speed from a realistic simulation of the population response in extrastriate visual area MT, the suboptimal decoder has almost the same accuracy and precision as traditional decoding models. It succeeds in predicting the precision and imprecision of motor behavior using a suboptimal decoding computation because it adds only a small amount of imprecision to the code for target speed in MT, which is itself imprecise.

Introduction

In the cerebral cortex, information about a sensory stimulus or an impending movement is represented by the profile of activity across a population of neurons. Cortical neurons tend to be tuned for specific properties of a stimulus or movement, responding most for a preferred stimulus or movement and less in relation to how much the stimulus or movement differs from the preferred one. Because the exact stimulus or movement is encoded by which neurons are the most active, rather than the level of activity in the full population, we call the cortical representation a “place code.” For motor activity and perception, the place code present in any cortical area is converted, either explicitly or implicitly, to a “rate code,” in which properties of the sensory stimulus or motor act are represented in terms of the firing rate of individual neurons (Groh, 2001). Sensorimotor integration provides examples of explicit conversions from place to rate codes, for example, the conversion from the place code in the superior colliculus into a rate code for the generation of saccadic eye movements (Lee et al., 1988) or from the place code for visual motion in visual area

MT into the cerebellar rate code that drives the initiation of smooth pursuit eye movements (Lisberger and Fuchs, 1978; Churchland and Lisberger, 2001).

Here, we ask how the brain decodes information in a place code to estimate a specific parameter of a sensory stimulus. Most of the decoding computations proposed in previous papers—such as “vector averaging” (Churchland and Lisberger, 2001), “maximum likelihood” (Jazayeri and Movshon, 2006), and more complicated statistical methods (Beck et al., 2008)—are motivated by considerations such as algorithmic elegance or theoretical optimality. They indicate how decoding could be performed in the brain. Our goal is to turn the conversation to implementation. Although it is possible that the brain extracts estimates optimally (Salinas and Abbott, 1994; Jazayeri and Movshon, 2006; Beck et al., 2008), our premise is that optimal computations may require neural circuits that are large and complex, posing a burden for biological implementation.

Divisive normalization is a special case of this broader concern. Normalization has been used to account for many neural responses (Heeger, 1992; Simoncelli and Heeger, 1998; Bonin et al., 2005; Zoccolan et al., 2005; Solomon et al., 2006) but has been implemented using extensive neural circuitry to sum the population response and an as yet unproven mechanism to accomplish division (Groh, 2001; Chance et al., 2002). In the present paper, we propose a decoding model that is qualitatively different from previous models and decidedly suboptimal, yet works well. It estimates the parameters of a sensory stimulus from the occurrence of a few spikes in each of many noisy sensory neurons bases normalization on an imperfect estimate of the total population activity and implements division through a local, cellular mechanism of gain modulation.

Received Dec. 28, 2010; revised Jan. 31, 2011; accepted Feb. 2, 2011.

This research was supported by National Institutes of Health Grant EY017210, including an American Recovery and Reinvestment Act supplement, the Howard Hughes Medical Institute, and the Swartz Foundation. We thank Philip Sabes for helpful discussions and Loren Frank, Michael Shadlen, and the members of the Lisberger laboratory for comments on a previous version of this manuscript.

Correspondence should be addressed to Kris Chaisanguanthum, Department of Physiology, Box 0444, University of California, San Francisco, 513 Parnassus Avenue, Room HSE-802A, San Francisco, CA 94143-0444. E-mail: chaisang@phy.ucsf.edu.

DOI:10.1523/JNEUROSCI.6776-10.2011

Copyright © 2011 the authors 0270-6474/11/314868-10\$15.00/0

Materials and Methods

Much of work here is analytical. We sketch the derivations in Results and fill in details in Appendix. For the part that is based on simulations, we simulated the response of a model MT population with parameters adapted from Huang and Lisberger (2009). For target motion at speed S (in degrees per second), the mean response μ_k of the k th cell (in number of spikes) is predicted by its tuning curve:

$$\mu_k = M \times \exp\left(-\frac{(\log_2 S - \log_2 S_k)^2}{2 \times 1.45^2}\right) \times T, \quad (1)$$

where $M = 100$ spikes/s is the maximum mean cell response (firing rate), and $T = 0.1$ s is the length of the time window under consideration. The n model neurons have a Fano factor equal to unity so that the variance of the response across trials is equal to its mean. The correlation C_{kl} between the output of the k th and l th cells is required to satisfy the following:

$$C_{kl} = R \times \exp\left(-\left(\frac{\log_2 S_k - \log_2 S_l}{\alpha \lambda}\right)^2\right) \quad (2)$$

where $r = 0.36$ is the maximum pairwise correlation among cells. Cells with similar speed tuning have larger noise correlations, which fall off with “distance” constant $\alpha \lambda$, expressed in terms of the full range of the cell tunings ($\lambda = \log_2 S_n - \log_2 S_1$); typically, we use $\alpha = 0.3$.

To impose the noise and correlation structure defined above, we computed the Cholesky decomposition D of the correlation matrix C . We defined ρ_k as a vector of $n = 1600$ pseudorandom numbers, normal, centered at zero and with unity variance, and we defined the noise

$$v_k = \sqrt{\mu_k} \times \sum_i D_{ki} \rho_i, \quad (3)$$

so that the respective spike counts of the cells in the time window T are $N_k = \mu_k + v_k$. Each N_k was rounded to the nearest non-negative integer, although we confirmed in simulations that allowing v_k to be continuous and/or negative does not materially affect our conclusions. This numerical method generates spike count noise as though drawn from a multivariate normal distribution satisfying the empirically measured statistical properties of MT responses (Huang and Lisberger, 2009): each unit has noise variance equal to its mean firing rate, and pairwise noise correlations are given by the correlation matrix C .

Once we calculated the spike counts N_k , we simulated the spiking activity of each model neuron as an independent Poisson process, with the constraint that the calculated number of events occurs within the time window. Because the decoder looks at interspike intervals (ISIs) of its full input spike train to a decoder rather than the input from any individual sensory neuron, the effect of adding an absolutely refractory period of 2 ms to the sensory cells should be negligible. We verified this prediction in simulations.

To decode the population response with maximum likelihood estimation, we calculated the likelihood $L(S'|N_k)$ of target motion at speed S' given spike counts N_k and assuming a uniform prior for speed. Under the assumption of a multivariate normal distribution, the likelihood is proportional to the following:

$$L(S'|N_k) \propto \frac{1}{\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(N - \mu(S'))\Sigma^{-1}(N - \mu(S'))\right), \quad (4)$$

where $\mu_k(S')$ are the expected mean firing rates given target motion at speed S' , $\Sigma(S') = \Delta(S')C\Delta(S')$ defines the expected covariance matrix, and $\Delta(S')$ is the diagonal matrix with terms $\sqrt{\mu_k(S')}$. In practice, Equation 4 provides a good approximation of the true likelihood but does not take into account that the N_k , measured experimentally as spike counts, are non-negative integers. Such approximations allow for tractable computations, and their effects are assumed to be small, as in the literature (Abbott and Dayan, 1999). Then, given a set of

Figure 1. Diagram of population decoding. The three sensory cells on the left are sensitive to a stimulus variable and are tuned for peak firing rates at different values of the stimulus variable. The population decoder aggregates the output of the sensory cells and estimates the sensory parameter, with the goal of an output that is linear in the stimulus variable and quite precise.

spike counts N_k , the maximum likelihood estimate of the underlying target speed S_{ML} is

$$S_{ML} = \arg \max_{S'} L(S'|N_k), \quad (5)$$

Results

Population decoding by vector averaging of firing rates

Consider a population of n neurons (called “sensory cells”) that are tuned for the value of a stimulus variable x that has the true value $x = X$. The k th cell, tuned to a value $x = X_k$, has a response quantified as $r_k = r(X; X_k)$; for example, $r(X; X_k)$ might a Gaussian function centered at X_k . As diagrammed in Figure 1, for an example population of $n = 3$ neurons that are tuned for sensory variable x , the brain must use the responses of multiple sensory neurons to decode an estimate X' that is reliably related to X .

The estimates of X by the brain have variation that we can estimate by making repeated measurements of perceptual or motor behaviors that are driven by those estimates. In cases in which we know that the behavioral variation arises from noise or errors in sensory processing (Osborne et al., 2005), the decoded estimates in our models should have means and variances of X' that match the measurements from our data. If the X_k tile the discernable range for x in a reasonably regular way, then it is intuitive to think of the population response as a probability density function (Ma et al., 2006). A simple and, under some assumptions, optimal estimator of X is provided by (Salinas and Abbott, 1994) the following:

$$X' = \frac{\sum_{k=1}^n r_k X_k}{\sum_{k=1}^n r_k}. \quad (6)$$

Equation 6 gives, by definition, the center of mass of a population response and is often called its “vector average.” The vector average is one of many equations that can estimate sensory parameters from a population response (Seung and Sompolinsky, 1993), and we choose it as a starting point because (1) it has been successful in accounting for the relationship between sensory population responses and motor behavior in the system we study (Churchland and Lisberger, 2001; Huang and Lisberger, 2009; Yang and Lisberger, 2009) and (2) it is a relatively simple and generic way to estimate a stimulus parameter from such a population. We regard vector averaging not as something the brain necessarily does but rather as a tool that we use for reading out the mean and variance of the representation of a sensory parameter in a population response. Thus, our goal is not to find a neural implementation of vector averaging per se but rather to find a neurally plausible decoding implementation that has the same biologically realistic performance as does vector averaging.

If the magnitude of the overall population response $\sum_k r_k$ is known and constant across all stimulus conditions, then Equation 6 has a straightforward interpretation as a neural circuit model (Groh, 2001). The quantity in the numerator can be computed by using static synaptic weights proportional to the preferred stimulus X_k of each neuron to create the weighted sum of the population response (Fig. 1). If the denominator is the same for all stimuli, then the sum would simply be proportional to X and divisive normalization would be unnecessary. In practice, however, $\sum_k r_k$ is a dynamic quantity of the population response: for example, the overall response level of motion-tuned neurons in visual area MT is sensitive to other parameters of target motion such as stimulus form and contrast (Sclar et al., 1990; Kretzelberg et al., 2006; Priebe et al., 2006). Decoding must include a mechanism to render the estimate of X independent of the summed activity of the population response, for example via the explicit divisive normalization in Equation 6.

Population decoding based on spike trains

One of the challenges of population decoding is to estimate the parameters of a sensory stimulus in a short temporal interval from the spike trains of many neurons. We start by creating a structure for doing so, keeping in mind the 100 ms interval of sensory inputs that drive pursuit eye movements (Lisberger and Westbrook, 1985). In that interval, the relevant sensory neurons in area MT will discharge just a few spikes each (Osborne et al., 2004).

Formally, the k th sensory cell in a population response has a spike train $b_k = (t_{k,1}, t_{k,2}, t_{k,3}, \dots, t_{k,N_k})$, defined as the ordered sequence of times at which the cell generated spikes between times 0 and T . Within the relevant time window T , the k th cell generates N_k spikes so that its firing rate r_k is estimated by N_k/T .

From the perspective of a decoding neuron such as the one illustrated in Figure 1, we can merge the incoming spike trains b_k from the k sensory neurons and sort them by their arrival time at the decoding neuron (Fig. 2) to compile the input stream to the decoder: $B = [(x_1, t_1), (x_2, t_2), (x_3, t_3), \dots, (x_N, t_N)]$. This is a formal description of the full population response converging onto the decoder: each pair (x_j, t_j) corresponds to one input spike that occurred at time t_j in the k th neuron, which has preferred stimulus X_k that becomes the value of x_j . Pairs are time ordered (i.e., $t_{j+1} > t_j$). The total number of spikes across all sensory cells over the full time window of length T is given by $N = \sum_{k=1}^n N_k$.

Figure 2. Aggregation of input spikes into the spike train seen by the decoding neuron. The top rows represent hypothetical spike rasters for a population of three sensory cells. The bottom row illustrates the population raster for all three sensory cells, as seen by the decoder. Colors label the three different sensory neurons.

Now, we can translate Equation 6 from a sum over sensory cells, indexed by k , into a sum over the input spikes to a decoder unit, indexed by j (detailed derivation in Appendix):

$$X' = \frac{1}{N} \sum_{j=1}^N x_j. \quad (7)$$

Normalization for different population response amplitudes is provided by counting the number of spikes N in the input stream B .

Our next step is to estimate N from the intervals between spikes in the sequence B . We define $\Delta t_j \equiv t_j - t_{j-1}$ (and $t_0 \equiv 0$) and make the reasonable assumption for a sufficiently dense population input that the last spike occurs at time t_N near the end of the time window (i.e., $T \approx t_N$). Then, we obtain $N \approx T / \langle \Delta t_j \rangle_j$, where the angle brackets denote the mean over all input spikes j . This expression is related to the statement that the average interspike interval Δt_j is simply the length of the time window divided by the number of spikes. Now, Equation 7 becomes

$$X' = \frac{\langle \Delta t_j \rangle_j}{T} \sum_{j=1}^N x_j. \quad (8)$$

The estimator in Equation 8 is equivalent to Equation 6. However, it yields a different interpretation as a model: it is based on weighting incoming spikes rather than estimating firing rates. The decoder is an accumulator: each spike input (j) causes an increment in the decoder output according to the preferred stimulus or “label” on the incoming axon. The assumption of accumulation, or temporal integration, is reasonable when the time intervals used to estimate stimulus properties are short, as they are for pursuit eye movements in primates (~ 100 ms). Neurons expressing much longer temporal integrations have been discussed in relation to a number of other neural functions (Shadlen and Newsome, 2001; Aksay et al., 2003).

Normalization without division

In population decoding, normalization compensates for dynamic population response amplitudes. Equation 8 shows that

normalization can be accomplished by using the mean interspike interval in a merged spike train seen by a decoder as a proxy for the overall population response N . The value of $\langle \Delta t_j \rangle$ estimates the total activity in the neural population, even as it is an exact assessment of the total input to a decoding neuron.

If a decoding neuron went one step further by estimating $\langle \Delta t_j \rangle$ and using that estimate to set the gain of neural processing in a biological way, then it would be possible to normalize for population response amplitude without either summing the full population activity or dividing. We suggest that such an approach would be much easier to implement with neurons, and we show later that estimating population activity in this way does not add very much noise to the decoded estimate X' , given the correlated noise already present in the population activity.

Suppose the overall gain of a decoding cell has a response nonlinearity that depends on the duration of the time interval (Δt_j) between the j th and $(j - 1)$ th spike in the incoming spike train B , without regard for the axon of origin of the two spikes. The response of the decoding neuron to a spike in a sensory neuron would be $\Delta t_j x_j$ and would accumulate the quantity

$$X' = \frac{1}{T} \sum_{j=1}^N \Delta t_j x_j. \quad (9)$$

In Equation 9, the effect of each incoming spike depends on (1) a static gain determined by the preferred speed of its neuron of origin, perhaps implemented as a synaptic strength, and (2) a dynamic gain factor that is related to the interval since the previous incoming spike. Because the dynamic gain factor depends on the most recent interspike interval in the merged input spike train, it estimates the total response of the population and can be used for normalization in place of the actual total population response. We show in Appendix that Equation 9 differs from Equations 6 and 8 by a term that averages to zero under reasonable assumptions. Therefore, Equation 9 is an unbiased estimator of the actual value of the stimulus variable, X .

A hypothetical decoding unit based on Equation 9 would respond to each input spike with a gain that is proportional to the time Δt_j since the prior spike in the collection of all the spikes in its inputs. It would normalize automatically the population response that serves as its input; its output would be related to the preferred stimulus parameter for the most active neurons in the population response. Stochastic behavior in firing across a large population tends to create a Δt_j distribution sharply peaked at zero (Fig. 3, dashed curve); in practice, it would be sufficient for a Δt_j -dependent response gain to be continuous and increasing for the range of interspike intervals that dominate the probability distribution of the Δt_j for the aggregated inputs to the decoding cell. To take into account the realities of a biological system, we have allowed the Δt_j -dependent gain to saturate (Fig. 3, solid curve).

If there were only a single decoding neuron with inputs from many sensory cells, then the decoder would need to operate on a very fast timescale to be able to discern differences between very short interspike intervals in the total incoming spike train. If, however, there were m decoding units, each receiving inputs from a small subset of the sensory cells, then the gain function could be based on longer interspike intervals. Once the place code of a population response is translated into noisy rate codes in a group of decoding units, the computationally simple step of averaging across decoding units could provide a better rate code. Furthermore, the number of decoding neurons could be “chosen” to keep the number of inputs to each decoding unit in a

Figure 3. Postulated nonlinear weighting of decoder inputs according to the previous interspike interval in the aggregate input spike train. The dashed curve shows the probability density of interspike intervals in the input spike train. The continuous curve shows a suitable nonlinear weighting function as a function of the previous interspike interval.

range that is on the rising phase of the Δt_j nonlinearity. The decoding stage also could be modularized for other reasons, such as robustness or redundancy.

A specific implementation

As a specific example of the general problem of sensory population decoding, we use smooth pursuit eye movements. Sensory signals for pursuit arise in extrastriate visual area MT, in which neurons respond selectively to motion across the visual field and are tuned for target speed and direction (Maunsell and Van Essen, 1983). We apply the decoder described in Equation 9 to a population of model neurons designed to mimic the tuning of MT neurons for the speed of visual motion. Our example is chosen to elucidate how a decoder of target direction and speed could be implemented to convert the population response in MT into estimates of target speed with means and variances defined by measurements of the pursuit response (Osborne et al., 2005, 2007). Pursuit latency is on the order of 100 ms with respect to the initiation of target motion, implying that an estimate of target velocity is made from the population response within a time window of that duration.

Our model MT population is explicitly constructed to mimic statistical features—both at the level of single cells and noise correlations—of the population response measured experimentally (Maunsell and Van Essen, 1983; Osborne et al., 2004; Huang and Lisberger, 2009). It consists of $n = 1600$ model neurons, each with Gaussian tuning as a function of the logarithm of the speed S (measured in degrees per second) of a moving visual target. The mean response of each model neuron is normal in $\log_2 S$ with a width of 1.45 log units and a peak response of 100 spikes/s at $\log_2 S_k$; the S_k are spaced uniformly on a $\log_2$ scale between $S_1 = 0.1^\circ/\text{s}$ and $S_{1600} = 512^\circ/\text{s}$; some example tuning curves are shown in Figure 4B. The variance of the spike count of each model sensory neuron is constrained to replicate the variance found empirically, yielding a population response like that in Figure 4C.

Figure 4. Features of the model MT population response. **A**, Noise correlation matrix, in which the colors indicate the correlation created between different model sensory neurons. **B**, Mean responses of sensory cells are Gaussian in the logarithm of target speed; three representative sensory cells are shown. **C**, One example model population response, for a target moving at 21.2°/s. Each purple marker represents the spike count of a sensory cell over a 100 ms window, plotted as a function of the preferred speed of that cell. Gray curve shows the mean responses across the population for the given target speed.

The variance in the model MT population also is constrained to have the correlation structure found empirically (Huang and Lisberger, 2009); in particular, the noise correlation between two cells is a function of the square of the difference between their preferred log speeds, peaking at 0.36 and falling exponentially with a “distance” constant that is 30% of the full (logarithmic) range tiled by the preferred speeds of sensory cells (i.e., 3.7 log units). The structure and amplitude of the neuron–neuron correlations, depicted in Figure 4A, are particularly important because they place an upper limit on the coding precision for target speed in the population response. Because of the correlations, the addition of more sensory cells after the first several hundred has negligible impact on the mean or variance of the decoded estimates of target speed. Thus our results are not sensitive to our specific choice for the size ($n = 1600$) of the model population. In reality, MT neurons are tuned for both speed and direction of motion. For simplicity, we only consider the speed component; our results generalize to the two dimensional case.

We simulated the response of the model MT population to 500 target motions at random speeds uniformly distributed be-

tween 2 and 64°/s. This corresponds approximately to the range of target motion velocities over which smooth pursuit is studied behaviorally and is within the range over which the distribution of the preferred speeds in our model population is sufficiently uniform that we can decode target speed directly, without nonlinear compensation for the fact that the population speed-tuning distribution and individual speed tuning are natively logarithmic (Lisberger and Movshon, 1999; Nover et al., 2005; Huang and Lisberger, 2009). For each target motion, we decoded the response of the population with the model we propose using ISI in Equation 9 and compare it with the true vector average shown by Equation 6 and a maximum likelihood decoder, which is asymptotically optimal. For the latter decoding model, we take the speed S' of the pair (S', M') that maximizes the likelihood in Equation 5. We quantified the performance of each decoder via the fractional decoding errors $[(S' - S)/S]$, where S' is the decoded speed, and S is the actual target speed. We measured decoder bias as the mean fractional decoding error and decoder variation as the variance of the fractional decoding error.

Figure 5 shows that our model and the vector average decoding methods yield nearly identical answers in terms of both the bias and the variation of the estimates of target speed; the additional error introduced by our model is small. The top row of Figure 5 shows the similar performance of the three decoders in terms of decoded speed as a function of target speed. For all three decoders, the variation of decoded speed increases as a function of the value of the decoded speed, as found in pursuit (Osborne et al., 2005) and motor control in general (Harris and Wolpert, 1998).

In agreement with empirical data (Nover et al., 2005), the fractional decoding error of all the decoders, rather the decoding error per se, remains approximately constant as a function of target speed. Pooling the responses to similar target speeds revealed that the vector average and ISI decoders showed a similar small bias at low target speeds, indicated by the nonzero values of the mean fractional coding errors (Fig. 5, bottom row). This slight bias is an artifact not of the decoders but rather of reading out linear speed from a model MT population that natively codes speed logarithmically. The bias disappears, and the performance of the decoders improves to 14.0 and 14.1%, respectively, if we decode the logarithm of the target speed directly, a conversion that requires a nonlinear transformation. The maximum likelihood decoder performs the logarithmic conversion intrinsically.

The variance of the fractional decoding errors indicated that the maximum likelihood decoder produces slightly less variable estimates of target speed than do the other two models. Although the maximum likelihood decoder takes into account the neuron–neuron correlations in the model MT population, it cannot eliminate decoding error altogether, and we conclude that its 11.4% variance is an estimate of the error inherent in the population response itself. Vector averaging and our proposed model are slightly suboptimal, in that their variances (15.5 and 15.6%) are larger than those of the optimal decoder. Given the variation of the decoding error inherent in the population, the marginal decrease in performance for the two suboptimal decoders is relatively small. The size of the errors predicted by all three decoders is comparable with data from behavioral experiments (Osborne et al., 2005), meaning that measurements of pursuit variation alone cannot distinguish among possible decoding schemes.

We verified that the normalization method in our proposed decoder does, indeed, decode target speed in a way that is independent of the amplitude of the population response. For values of population response amplitude (M in Eq. 1) spanning from

Figure 5. Performance of population decoding computations. **A–C**, Each point plots the estimate of target speed from one run of the model as a function of the actual (simulated) target speed used to create the model MT population response. **D–F**, The quality of the decoding computation as a function of target speed: for each decoder, the points show the mean fractional decoding error, and the vertical error bars indicate the variance of the fractional decoding error for multiple target speeds within bins of simulated target speed.

Figure 6. Decoder performance is independent of the size of the sensory population response and has fast dynamics. **A**, Performance of ISI decoder defined by Equation 9 is quantified as the mean and variance across trials of the fractional decoding error and plotted as a function of the maximum spike count in the population of sensory neurons. **B, C**, Evolution of estimate of target speed for target motion at 20 or 40°/s. Black dashed traces show target speed, and the red and blue traces show the estimates of the ISI decoder and vector averaging. **B** shows performance as a function of time and **C** shows performance as a function of the number of spikes in the input sensory neurons.

12.5 to 100 spikes/s, the mean and variance of the fractional decoding error remain nearly constant (Fig. 6A), with only a slight improvement as the population response size grows. However, even for population response amplitudes several times smaller than values measured experimentally (Huang and Lisberger, 2009), the quality of decoding approaches its limit.

The ISI decoder also produces estimates of target speed that evolve rapidly in time and track either increases or decrease in target speed. For target speeds of either 20 or 40°/s, the speed estimate from

both vector averaging and the ISI decoder can reach asymptotic performance within <2 ms (Fig. 6B) and after <100 spikes (Fig. 6C) have been produced by the population of model sensory neurons. If the model decoder neurons are implemented as “integrate-and-fire” units, then they also track changes in target speed on a timescale determined by the interspike intervals of the decoding neuron. The rapid response of the ISI decoder with instantaneous changes in the normalizing gain nonlinearity points out that decoder output would respond more slowly in proportion to the presence of any longer temporal dynamics within the nonlinearity.

Our results did not depend materially on the parameters of the model population, as long as it retained stochastic spike counts in short time intervals. The number m of independent decoding units was unimportant, as were the characteristics of the saturation (if any) of the gain modulation by Δt (Fig. 3), as long as each decoding unit received most of its input with Δt_i in the

nonsaturation range. The model also worked when we implemented decoding units as simulated integrate-and-fire neurons.

Figure 7 documents the similar effects on the decoding performance of all three models we tested of noise correlation parameters R and α in Equation 2, which determine the peak correlation amplitude and the decay of the correlation across the population. For the ISI decoder, vector averaging, and maximum likelihood, increasing the length constant for the decay of correlations decreased the fractional decoding error, whereas increasing the maximum correlation increased the fractional decoding error, as reported previously (Huang and Lisberger, 2009; Lisberger, 2010). All three decoders show similar performance for the parameters of the model that best reproduced the structure of noise correlations across the MT population response (black circle in all three graphs). In the biologically unrealistic regimen of very high peak noise correlations (in which R approaches unity), the difference in performance between maximum likelihood and the other two decoding algorithms becomes more pronounced, although all three decoders perform poorly under these conditions.

Finally, we verified that the ISI decoding algorithm is robust to possible non-uniformities in the distribution of the preferred speeds of sensory cells. Our generic model population contained 1600 sensory cells with preferred speeds spaced evenly along a logarithmic scale between $S_1 = 0.1^\circ/\text{s}$ and $S_{1600} = 512^\circ/\text{s}$. To test the effects of disrupting the uniformity, we jittered the preferred speeds by choosing them according to $\log S_{j,k} = \log(S_k) + Z_{j,k}$, where the $Z_{j,k}$ are randomly drawn from a normal distribution with a mean of zero and a width of $J\Delta(\log S)$. Thus, J is a measure of the deviation from even spacing of the distribution of the preferred speeds of sensory cells. Figure 8 shows that the variation of the decoding estimates was independent of the magnitude of the jitter in all three decoding models. The mean errors of the estimates of target speed retain the small deviations from zero illustrated in Figure 5D–F for all three decoding models.

We conclude that the imprecision inherent in encoding information in the finite spike counts of a correlated neuron population can be much larger than additional sources of error that are

introduced by simple population decoders. As a result, suboptimal decoders add relatively little estimation error above that inherent in the population and have performance that mimics biological observations to a satisfactory degree.

Discussion

It is common to assess neural processing models in terms of optimality in an information-theoretical sense. We have outlined a different approach, one that we think is more plausible in terms of implementation with neurons. Previous work has shown that the imprecision inherent in a population code of neurons with noise correlations is passed all the way to behavioral outputs, for both perception (Zohary et al., 1994; Bair et al., 2001) and action (Osborne et al., 2005; Huang and Lisberger, 2009). We propose that the inexorable, correlated noise in the sensory encoding creates inescapable imprecision in estimates of sensory parameters. This allows decoding mechanisms to be suboptimal in an information processing sense and permits increased efficiency of implementation without compromising behavioral precision. In the present paper, we have outlined our proposal analytically and demonstrated its successes through simulation of a specific example in which a sensory place code is known to drive movement after transformation into a rate code. Perhaps the brain, in decoding population responses, does not act as an ideal observer. Rather than striving for optimal information processing, it may settle for “good enough.”

Our proposed decoding mechanism provides a numerical example of a more general principle that non-optimal processing may be sufficient for a system already beset by a nontrivial amount of correlated noise. The principle of operation of our model is to estimate neural parameters rather than computing them. Instead of assuming knowledge of firing rate, it estimates firing rate from the Poisson spike trains of the sensory neurons. Instead of computing the total population response for divisive normalization, it estimates the total population response in real time from the interspike interval in the net incoming spike train to a decoding neuron. Instead of dividing, it uses a local, nonlinear modulation of neural gain. The philosophy of performing computations locally is shared by a previous model of contrast-invariant orientation tuning (Kayser et al., 2001).

Implementation of our model relies on three biologically plausible functions. (1) Inputs to decoding units are weighted according to the preferred speed of their neuron of origin. Weighting according to the preferred speed of the input neuron has been an accepted feature of population vector decoding (Salinas and Abbott, 1994; Churchland and Lisberger, 2001) and could be implemented by setting synaptic strengths. (2) Decoding units are integrators. Integrator neurons or circuits have been postulated to account for the oculomotor velocity-to-position integrator in the brainstem (Aksay et al., 2003) and the ramp up to threshold of neurons in the lateral intraparietal area (Shadlen and Newsome, 2001). The duration of temporal integration in our model is much shorter than in the previous models. (3) Decoding units are rendered insensitive to the overall response amplitude across the population through a local, nonlinear modulation of the gain of each input according to the recent history of spikes. The nonlinear gain modulation we propose suggests a postsynaptic analog to short-term synaptic depression (Chance et al., 1998). It depends, however, on the

Figure 7. Effect of correlation structure and amplitude on decoder performance. The color of each pixel shows the SD of the residuals of the decoded target speeds for the correlation parameters given on the x- and y-axes. The three panels show simulations using three different decoders: the ISI decoder, vector averaging, and maximum likelihood decoding. Black points show correlation parameters used in the other figures, which agree with the measurements of neuron–neuron correlations by Huang and Lisberger (2009).

Figure 8. Effect of population non-uniformity on decoder performance. Each symbol shows the variance of the fractional decoding error as a function of the amount of jitter of the preferred speeds of the population of model sensory neurons. Different symbols show the results for the three different decoders.

total spike train seen by the postsynaptic neuron rather than on the spikes of any individual presynaptic axon. The cellular mechanism we propose has not been demonstrated in analyses of neural electrophysiology but could be compatible with input nonlinearities that arise in dendritic spike attenuation (Polsky et al., 2004), input attenuation resulting from depletion of extracellular resources (Montague, 1996), or other local, mutually suppressive interactions between synaptic inputs. Evidence for time-dependent sub-linear input summation has been observed in pyramidal cells (Urban and Barrionuevo, 1998).

The use of a cellular mechanism to implement divisive normalization has the advantage of greatly reducing the amount of neural circuitry needed to render the decoded estimates of sensory parameters insensitive to the amplitude of the population response. Previous models have implemented divisive normalization by accumulating the total activity across the population response and then using those signals as the input to division on another group of neurons. In principle, division could be implemented to some degree by shunting inhibition (Carandini et al., 1997) or the interplay of noisy excitatory and

inhibitory inputs (Chance et al., 2002). In our formulation, a separate set of “normalizing neurons” could estimate the total population activity by using the postsynaptic gain mechanism in our model while sampling from a population of model sensory neurons. As long as the sampling is sparse, the estimate could be achieved with relatively little circuitry and then used to perform divisive normalization with the mechanisms already suggested by others (Carandini et al., 1997; Groh, 2001; Chance et al., 2002). Of course, the cellular mechanism we propose for normalization would be applicable only when normalization is controlled only by the same inputs that drive the responses of a neuron. Thus, either a different kind of cellular mechanism or a network normalization mechanism would be needed to explain phenomena such as contrast normalization from outside the receptive field in the primary visual cortex (Geisler and Albrecht, 1992). In addition, the specific mechanism we suggest is just one of many possible cellular mechanisms of normalization that would have similar properties.

A cellular mechanism of divisive normalization would allow a neuron to normalize its inputs automatically. Divisive normalization also is one of the key elements in Heeger’s original model (Heeger, 1993) of the generation of motion sensitive responses in primary visual cortex V1 and of other accounts of cortical responses (Simoncelli and Heeger, 1998; Bonin et al., 2005; Zoccolan et al., 2005; Solomon et al., 2006). It is also a common ingredient in neural networks (Deneve et al., 1999). Furthermore, a cellular mechanism of self-normalization might be adaptively regulated on a neuron-by-neuron basis according to sensible rules of plasticity. Our model shows the utility of the concept that important neural computations might be handled by local cellular mechanisms rather than by neural circuits. Indeed, the solution by the brain to a specific problem such as population decoding may be very different from the mechanisms implied by even the most intuitively appealing equations. Vector averaging and the other extant decoding computations are perhaps at best metaphors for what happens in the brain.

Our decoding model also has the feature that, like vector averaging, it is agnostic to many of the specific details of the population response it is decoding. It has a simple implementation but ignores features such as shapes of the tuning curves of individual sensory cells and the correlation structure within the population. Although these ignored features could, in principle, be exploited to improve the performance of a decoder (Latham and Nirenberg, 2005; Jazayeri and Movshon, 2006; Beck et al., 2008), recent results show that, at least in MT, tuning properties and neuron–neuron correlations can be dynamic (Cohen and Newsome, 2008; Yang and Lisberger, 2009). Any decoding algorithm seeking to extract information optimally must be able to update its data about neural response properties on a fast timescale, lest it be optimized for only an arbitrary subset of conditions. A maximum likelihood decoder with incorrect information about the correlation structure of its inputs can perform worse than a naive vector averaging model or the model we propose.

The mechanistic simplifications in our model allow sensory decoding to proceed with a relatively small number of decoding neurons and synaptic connections. In principle, it would be possible to apply the same simplifications to other decoding approaches, such as some maximum likelihood (Jazayeri and Movshon, 2006) or Bayesian (Beck et al., 2008) approaches. However, the need to do so by reducing the number of neurons

and synapses would make their estimates suboptimal, contradicting the philosophy of optimal estimation on which the models are based. The point of our model is perhaps not so much embodied in the exact implementation but rather in the demonstration that compact, suboptimal neuron circuitry can perform well enough to account for neural data. Simplifications of the decoding computation would work because the precision of sensory estimates is limited by the presence of substantial correlated noise in the sensory population response. In the system we study, smooth pursuit eye movements, sensory coding, and behavioral performance both are decidedly non-optimal, and these observations are linked (Osborne et al., 2005, 2007; Medina and Lisberger, 2007; Huang and Lisberger, 2009; Osborne and Lisberger, 2009; Lisberger, 2010). This allows sensory decoding to be suboptimal without compromising behavioral performance much further. We imagine that the same is true for many, although possibly not all, neural systems.

The organization of computations in the brain must be constrained by the limitations of computational properties of neurons and neural circuits and by the need to perform complex functions reliably with minimal elements to keep the volume and energy use of the brain manageable (Montague, 1996). Especially in situations in which the sensory code is already suboptimal, there may be a tradeoff between the constraints of implementation and the drive for optimal performance. A small decrement in optimality of information processing could allow a large improvement in the efficiency of implementation. To accomplish this efficiency, we imagine that the brain might co-opt any one of a plethora of possible cellular mechanisms to perform complex computations without using a ton of neural wire.

Appendix

Derivation of the decoding computation

In Results, we omitted some details about how to translate the vector average defined by Equation 1 from a sum over sensory cells (indexed by k) into a sum over population spikes (indexed by j). Those details are presented here.

We use the observation that the firing rate of each sensory neuron r_k is defined as N_k/T and multiply the top and bottom of Equation 6 by T :

$$X' = \frac{\sum_{k=1}^n r_k X_k}{\sum_{k=1}^n r_k} = \frac{\sum_{k=1}^n \frac{N_k}{T} X_k}{\sum_{k=1}^n \frac{N_k}{T}} = \frac{\sum_{k=1}^n N_k X_k}{\sum_{k=1}^n N_k}. \quad (10)$$

Recognizing that the denominator in Equation 10 is simply the total number of input spikes N and that summing over the preferred stimuli for the input neurons in the numerator is equivalent to summing the labels retained in the spike train B that describes the total input to the decoding neuron, we end up with Equation 7:

$$X' = \frac{1}{N} \sum_{j=1}^N x_j. \quad (11)$$

Inaccuracies in estimates of sensory parameters

The use of interspike intervals to estimate the total population response has the potential to introduce estimation errors. Here, we show that Equation 9 differs from Equations 6 and 8 by a term that averages to zero, under reasonable assumptions. Using the individual values of Δt_j in the estimator rather than the average $\langle \Delta t_j \rangle$ means that each term in the summand of Equation 9 con-

tains a term attributable to the residual $\delta t_j \equiv \Delta t_j - \langle \Delta t_j \rangle_{j'}$. If we substitute into Equation 9, we obtain the following:

$$X' = \frac{1}{T} \sum_{j=1}^N \Delta t_j x_j = \frac{1}{T} \sum_{j=1}^N \left(\langle \Delta t_j \rangle_{j'} + \delta t_j \right) x_j \\ = \frac{1}{T} \left(\langle \Delta t_j \rangle_{j'} \sum_{j=1}^N x_j + \sum_{j=1}^N \delta t_j x_j \right). \quad (12)$$

This estimator is equivalent to that in Equations 6 and 8, up to the extra $\delta X' \equiv \frac{1}{T} \sum_{j=1}^N \delta t_j x_j$ term, which defines additional error on the estimate of X attributable to the variance of Δt_j . If we define $\delta x_j \equiv x_j - \langle x_j \rangle_{j'}$ and substitute into the equation for $\delta X'$, we obtain

$$\delta X' = \frac{1}{T} \left(\sum_{j=1}^N \delta t_j \langle x_j \rangle_{j'} + \sum_{j=1}^N \delta t_j \delta x_j \right). \quad (13)$$

The first summation in Equation 13 is manifestly zero; the second vanishes on average if the δx_j and δt_j are statistically independent, as in the case of stochastically generated spikes in the sensory cells. We conclude that the estimator described in Equation 9 is an unbiased approximation of the vector average of the population response.

For the specific example of estimating target speed from a model MT population response, the imprecision of the true vector average is limited by the mismatch between spike counts in a finite time window and underlying firing rates in a finite time window and by the neuron–neuron correlations in the population noise. Given the noise correlations found in MT neurons, the error in the estimates of target speed from Equation 1 asymptotes at approximately the $\pm 11.4\%$ variance seen in Figure 5; this error is inherent in the population response itself (Huang and Lisberger, 2009). Thus, approximation of the vector average using Equation 9 differs from the vector average by $\delta X'$ (Eq. 13). The additional contribution to the error from $\delta X'$ scales as $1/\sqrt{N}$ when the Δt_j are independent. In this case, because of the relatively large error already present in the population response, the additional error from the $\delta X'$ term is almost negligible, worsening the precision only slightly. Potential biases introduced into this estimator from correlations between δx_j and δt_j from measured spike-timing correlations (Bair et al., 2001) are also negligible compared with the error inherent in the population response. It appears that Equation 9 extracts much (perhaps most) of the available information from a realistic population response.

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Aksay E, Baker R, Seung HS, Tank DW (2003) Correlated discharge among cell pairs within the oculomotor horizontal velocity-to-position integrator. *J Neurosci* 23:10852–10858.
- Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J Neurosci* 21:1676–1697.
- Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A (2008) Probabilistic population codes for Bayesian decision making. *Neuron* 60:1142–1152.
- Bonin V, Mante V, Carandini M (2005) The suppressive field of neurons in lateral geniculate nucleus. *J Neurosci* 25:10844–10856.
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644.
- Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. *J Neurosci* 18:4785–4799.
- Chance FS, Abbott LF, Reyes AD (2002) Gain modulation from background synaptic input. *Neuron* 35:773–782.
- Churchland MM, Lisberger SG (2001) Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. *J Neurosci* 21:9387–9402.
- Cohen MR, Newsome WT (2008) Context-dependent changes in functional circuitry in visual area MT. *Neuron* 60:162–173.
- Deneve S, Latham PE, Pouget A (1999) Reading population codes: a neural implementation of ideal observers. *Nat Neurosci* 2:740–745.
- Geisler WS, Albrecht DG (1992) Cortical neurons: isolation of contrast gain control. *Vision Res* 32:1409–1410.
- Groh JM (2001) Converting neural signals from place codes to rate codes. *Biol Cybern* 85:159–165.
- Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. *Nature* 394:780–784.
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197.
- Heeger DJ (1993) Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J Neurophysiol* 70:1885–1898.
- Huang X, Lisberger SG (2009) Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J Neurophysiol* 101:3012–3030.
- Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696.
- Kayser A, Priebe NJ, Miller KD (2001) Contrast-dependent nonlinearities arise locally in a model of contrast-invariant orientation tuning. *J Neurophysiol* 85:2130–2149.
- Krekelberg B, van Wezel RJ, Albright TD (2006) Interactions between speed and contrast tuning in the middle temporal area: implications for the neural code for speed. *J Neurosci* 26:8988–8998.
- Latham PE, Nirenberg S (2005) Synergy, redundancy, and independence in population codes, revisited. *J Neurosci* 25:5195–5206.
- Lee C, Rohrer WH, Sparks DL (1988) Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature* 332:357–360.
- Lisberger SG (2010) Visual guidance of smooth-pursuit eye movements: sensation, action, and what happens in between. *Neuron* 66:477–491.
- Lisberger SG, Fuchs AF (1978) Role of primate flocculus during rapid behavioral modification of vestibulo-ocular reflex. I. Purkinje cell activity during visually guided horizontal smooth pursuit eye movements and passive head rotation. *J Neurophysiol* 41:733–763.
- Lisberger SG, Movshon JA (1999) Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *J Neurosci* 19:2224–2246.
- Lisberger SG, Westbrook LE (1985) Properties of visual inputs that initiate horizontal smooth pursuit eye movements in monkeys. *J Neurosci* 5:1662–1673.
- Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.
- Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49:1127–1147.
- Medina JF, Lisberger SG (2007) Variation, signal, and noise in cerebellar sensory-motor processing for smooth pursuit eye movements. *J Neurosci* 27:6832–6842.
- Montague PR (1996) The resource consumption principle: attention and memory in volumes of neural tissue. *Proc Natl Acad Sci U S A* 93:3619–3623.
- Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci* 25:10049–10060.
- Osborne LC, Lisberger SG (2009) Spatial and temporal integration of visual motion signals for smooth pursuit eye movements in monkeys. *J Neurophysiol* 102:2013–2025.
- Osborne LC, Bialek W, Lisberger SG (2004) Time course of information about motion direction in visual area MT of macaque monkeys. *J Neurosci* 24:3210–3222.
- Osborne LC, Lisberger SG, Bialek W (2005) A sensory source for motor variation. *Nature* 437:412–416.
- Osborne LC, Hohl SS, Bialek W, Lisberger SG (2007) Time course of precision in smooth-pursuit eye movements of monkeys. *J Neurosci* 27:2987–2998.
- Polsky A, Mel BW, Schiller J (2004) Computational subunits in thin dendrites of pyramidal cells. *Nat Neurosci* 7:621–627.
- Priebe NJ, Lisberger SG, Movshon JA (2006) Tuning for spatiotemporal

- frequency and speed in directionally selective neurons of macaque striate cortex. *J Neurosci* 26:2941–2950.
- Salinas E, Abbott LF (1994) Vector reconstruction for firing rates. *J Comput Neurosci* 1:89–107.
- Sclar G, Maunsell JH, Lennie P (1990) Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res* 30:1–10.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753.
- Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86:1916–1936.
- Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. *Vision Res* 38:743–761.
- Solomon SG, Lee BB, Sun H (2006) Suppressive surrounds and contrast gain in magnocellular-pathway retinal ganglion cells of macaque. *J Neurosci* 26:8715–8726.
- Urban NN, Barrionuevo G (1998) Active summation of excitatory postsynaptic potentials in hippocampal CA3 pyramidal neurons. *Proc Natl Acad Sci U S A* 95:11450–11455.
- Yang J, Lisberger SG (2009) Relationship between adapted neural population responses in MT and motion adaptation in speed and direction of smooth-pursuit eye movements. *J Neurophysiol* 101:2693–2707.
- Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25:8150–8164.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.