Perceptual decision making “Through the Eyes” of a large-scale neural model of V1

Jianing V. Shi1, Jim Wieland1, R. Theodore Smith1,2 and Paul Sajda1*

1 Department of Biomedical Engineering, Columbia University, New York, NY, USA
2 Department of Ophthalmology, New York Presbyterian Hospital, New York, NY, USA

INTRODUCTION

How the brain represents information and how such a representation is ultimately utilized to form decisions and mediate behavior are fundamental questions in systems neuroscience, being addressed at many different scales from single-unit recordings to neuroimaging across the brain. Work in theoretical neuroscience has argued that a useful information processing strategy for the brain is to map sensory information into a sparse representation – the sparse coding hypothesis (Olshausen and Field, 1996a). Sparse coding has been viewed as an optimal strategy for minimizing redundancy and has been experimentally observed and theoretically justified for a number of sensory systems including visual (Baddeley, 1996; Dan et al., 1996; Baddeley et al., 1997; Vinje and Gallant, 2000, 2002), auditory (Hahnloser et al., 2002; Hromadka et al., 2008; Greene et al., 2009), olfactory (Perez-Orive et al., 2002; Szymza et al., 2006; Brecht et al., 2004), and motor (Brecht et al., 2004) systems. It is also considered efficient from a metabolic and energy perspective.

Sparse coding has received substantial attention in the visual neurosciences (Baddeley, 1996; Dan et al., 1996; Olshausen and Field, 1996a,b, 1997, 2003; Rolls and Tovee, 1995; Vinje and Gallant, 2000, 2002; Simoncelli and Olshausen, 2001), particularly from the perspective of encoding. For example, it has been shown that imposing sparsity constraints on rate-based neuronal models yields receptive fields that resemble those of simple cells in V1 (Olshausen and Field, 1996a). Others have shown how sparse encoding can emerge temporally and is consistent with the spike time distributions of visual neurons (Baddeley, 1996). Some have looked more closely at the specific nature and degree of the sparse encoding, investigating differences between “soft” and “hard” sparseness constraints (Rehn and Sommer, 2007). Many conclude that the visual system maps the visual scene into a sparse representation. However given such a sparse encoding strategy, how does the rest of the visual system utilize it for robust and efficient object recognition? How is the sparse representation exploited downstream to yield the behavior we observe?

Sparse coding has been posited as an efficient information processing strategy employed by sensory systems, particularly visual cortex. Substantial theoretical and experimental work has focused on the issue of sparse encoding, namely how the early visual system maps the scene into a sparse representation. In this paper we investigate the complementary issue of sparse decoding, for example given activity generated by a realistic mapping of the visual scene to neuronal spike trains, how do downstream neurons best utilize this representation to generate a “decision.” Specifically we consider both sparse (L1-regularized) and non-sparse (L2 regularized) linear decoding for mapping the neural dynamics of a large-scale spiking neuron model of primary visual cortex (V1) to a two alternative forced choice (2-AFC) perceptual decision. We show that while both sparse and non-sparse linear decoding yield discrimination results quantitatively consistent with human psychophysics, sparse linear decoding is more efficient in terms of the number of selected informative dimension.

Keywords: computational modeling, decision making, neuronal network, sparse coding

ORIGINAL RESEARCH ARTICLE

doi: 10.3389/fpsyg.2013.00161

MATERIALS AND METHODS

PERCEPTUAL DECISION MAKING PARADIGM

We adapted a 2-AFC paradigm of face and car discrimination, where a set of 12 face (Max Plank Institute face database) and 12 car grayscale images were used. The car image database was the same used in Philiastides and Sajda (2006) and Philiastides...
et al. (2006), which was constructed by taking images from the internet, segmenting the car from the background, converting the image to grayscale, and then resizing to be comparable as the face images. The pose of the faces and cars was also matched across the entire database and was sampled at random (left, right, center) for the training and test cases. All the images (512 × 512 pixels, 8 bits/pixel) were equated for spatial frequency, luminance, and contrast. The phase spectra of the images were manipulated using the weighted mean phase method (Dakin et al., 2002) to introduce noise, resulting in a set of images graded by phase coherence. Specifically we computed the 2D Fourier transform of each image, and constructed the average magnitude spectra by averaging across all images. The phase spectra of an image were constructed by computing a weighted sum of the phase spectra of the original image (\(\Phi_{\text{image}}\)) and that of random noise (\(\Phi_{\text{noise}}\)).

Each image subtended 2° × 2° of visual angle, and the background screen was set to a mean luminance gray. The image size was set to match the size of the V1 model, which covered 4 mm² of cortical sheet. Figure 1 shows examples of the face and car images used in the experiment as well as the effect on the discriminability of the image class when varying the phase coherence.

The sequence of images was input to the model where an image was flashed for 50 ms, followed by a gray mean luminance image with an inter-stimulus-interval (ISI) of 200 ms (Figure 1C). Since simulating the model is computationally expensive, we minimized the simulation time by choosing an ISI which was as small as possible yet did not result in network dynamics leaking across trials. We conducted pilot experiments that showed that network activity settled to background levels approximately 200 ms after stimulus offset. We ran the simulation for each of the two classes, face and car, at different coherence levels (20, 25, 30, 35, 40, 45, 55%) respectively. Each image was repeated by 30 trials in the simulation, where the sequence of trials was randomly generated. In each simulation, we randomized the order of different images, making sure not to push the model into a periodic response pattern.

Parallel to simulating the model response, we conducted human psychophysics experiments. Ten volunteer subjects were recruited. All participants provided written informed consent, as approved by the Columbia University Institutional Review Board. All the subjects were healthy with corrected visual acuity of 20/20. Psychophysics testing was administered in a monocular manner. Images of different phase coherences were randomized in the psychophysics experiment. During the experiment subjects were instructed to fixate at the center of the images, and to make a decision on whether they saw a face or car, as soon as possible, by pressing one of two buttons with their right hand. The ISI for human psychophysics experiments was longer and randomized between 2500 and 3000 ms in order to provide for a comfortable reaction time and to reduce the subjects’ ability to predict the time of the next image. A Dell computer with an nVIDIA GeForce4 MX 440 AGP8X graphics card and E-Prime software controlled the stimulus presentation.

MODEL SUMMARY

An overview of the model architecture and decoding is illustrated in Figure 2. We modeled the early visual pathway with a feedforward lateral geniculate nucleus (LGN) input and a recurrent spiking neuron network of the input layers of (4Ca/B) of primary visual cortex (V1). We modeled the short-range connectivity within the V1 layer, without feedback from higher areas. We simulated a magnocellular version of the model, the details of which have been described previously (Wieland and Sajda, 2006a,b, 2007). Note our model is a variant of an earlier V1 model (McLaughlin et al., 2000; Wieland et al., 2001).

In brief, the model consists of a layer of \(N\) (4096) conductance-based integrate-and-fire point neurons (one compartment), representing about a 2 × 2 mm² piece of a V1 input layer (layer 4C). Our model of V1 consists of 75% excitatory neurons and 25% inhibitory neurons. In the model, 30% of both the excitatory and inhibitory cell populations receive LGN input. In agreement with experimental findings, the LGN neurons are modeled as rectified center-surround linear spatio-temporal filters. Sizes for center and surround were taken from experimental data (Hicks et al., 1983; Derrington and Lennie, 1984; Shapley, 1990; Spear et al., 1994; Croner and Kaplan, 1995; Benardete and Kaplan, 1999). Noise, cortical interactions, and LGN input are assumed to act additively in contributing to the total conductance of a cell. The noise term is modeled as Poisson spike train convolved with a kernel which comprises a fast AMPA component and a slow NMDA component (see Supplementary Materials in Wieland and Sajda, 2006a).
We used a linear decoder to map the spatio-temporal activity in our spike neuron model, fully exploring the spatio-temporal dynamics. The spike train for each neuron in the population is given by:

\[ s_{i,k}(t) = \sum_j S(t - t_{i,j,k}), \]

where \( t \in [0,250] \text{ ms} \), \( i = 1 \ldots N \) is the index for neurons, \( k = 1 \ldots M \) is the index for trials, \( l = 1 \ldots P \) is the index for spikes. Based on the population spike trains, we estimated the firing rate on each trial by counting the number of spikes within a time bin of width \( \tau \), resulting in a spike count matrix:

\[ r_{i,j,k} = \int_{(j-1)\tau}^{j\tau} s_{i,k}(t) dt, \]

where \( i = 1 \ldots N \) represents the \( i \)th neuron, \( j = 1 \ldots T/\tau \) represents the \( j \)th time bin, \( k = 1 \ldots M \) represents the \( k \)th trial. Note that we explored decoding using time bins of different length. When \( \tau = 25 \text{ ms} \), we assume that information is encoded in both neuron and time, since the firing rate is closer to instantaneous firing rate; when \( \tau = 250 \text{ ms} \), we integrate the spiking activity over the entire trial, leading to a rate-based representation of information.

A separate post hoc analysis showed that 25 ms was in fact the bin width that yielded the highest discrimination accuracy (bin width varied from 5 to 250 ms). The class label of each sample \( b_l \) takes the value of \{-1, +1\} representing either face or car with \( M \) being the number of trials. In order to explore the information within the spatio-temporal dynamics, we compute the weighted sum of firing rate over different neurons and time bins. This leads to seeking the solution of the following constrained minimization problem:

\[
\{w, v\} = \underset{w, v}{\arg \min} \frac{1}{M} \sum_{k=1}^{M} L((w^T x_k + v)b_k) + \lambda J(w),
\]

where the first term is the empirical logistic loss function, and the second term is the regularization function, with \( \lambda > 0 \) as the regularization parameter. We create a stacked version of the spike count matrix:

\[
x_{i,k} = r_{i,j,k} \quad \text{with} \quad l = (i-1)N + j, \text{i.e., stacking the neuron and time bin dimensions together.}
\]

The resulting linear decoder can be geometrically interpreted as a hyperplane that separates the classes of face and car, where \( w \) represents the weights for the linear decoder, and \( v \) is the offset. In the case of the sparse decoder, we use an L1 regularization term \( J(w) = \|w\|_1 \); alternatively for the non-sparse decoder, we use the L2 regularization \( J(w) = \|w\|_2^2 \). In the language of Bayesian analysis, the logistic loss term comes from maximum likelihood, L1 corresponds to the Laplacian prior, and L2 corresponds to the Gaussian prior. L1-regularized logistic regression results in a sparse solution for the weights (Krishnapuram et al., 2005; Koh et al., 2007; Meier et al., 2008). So-called “sparse logistic regression” serves as an approach for feature selection, where features that are most informative about the classification survive in the form of non-zero weights (Ng, 2004). We developed an efficient and accurate method to solve this optimization problem (Shi et al., 2010, 2011). Once we learn the hyperplane, for any new image, we can predict the image category via the sign of \( w^T x + v \).

Figure 3 provides a geometric intuition of why L1 and L2 regularization lead to sparse and non-sparse solutions, respectively. The paper we focus on the first approach, since we view our decoder as a tool for analyzing the information content in the neurodynamics and how downstream neurons might best decode this information for discrimination.

We constructed an optimal decoder to read out the information in our spike neuron model, fully exploring the spatio-temporal dynamics. Sparse decoding has been previously investigated for decoding real electrophysiological data, for instance by Chen et al. (2006), Palmer et al. (2007), and Quiroga et al. (2007).

Since a primary purpose of using the decoder is to identify informative dimensions in the neurodynamics, we estimate new decoder parameters at each stimulus noise level (coherence level) independently. Alternatively we could train a decoder at the highest coherence level and test the decoder at each coherence level. In this case we focus on the first approach, since we view our decoder as a tool for analyzing the information content in the neurodynamics and how downstream neurons might best decode this information for discrimination.

The LGN RF centers were organized on a square lattice. These lattice spacing and consequent LGN receptive field densities imply LGN cellular magnification factors that are in the range of the experimental data available for macaque (Malpeli et al., 1996). The connection structure between LGN cells and cortical cells is made so as to establish ocular dominance bands and a slight orientation preference which is organized in pinwheels (Blasdel, 1992). It is further constructed under the constraint that the LGN axonal arbor sizes in V1 do not exceed the anatomically established values of 1.2 mm (Blasdel and Lund, 1983; Freund et al., 1989).

In the construction of the model our objective was to keep the parameters deterministic and uniform as much as possible. This enhances the transparency of the model, while at the same time provides insight into what factors may be essential for the considerable diversity observed in the responses of V1 cells.

**SPARSE DECODING**

We used a linear decoder to map the spatio-temporal activity in the V1 model to a decision on whether the input stimulus is a face or a car. We employed a sparsity constraint on the decoder in order to control the dimension of the effective feature space. Sparse decoding has been previously investigated for decoding real electrophysiological data, for instance by Chen et al. (2006), Palmer et al. (2007), and Quiroga et al. (2007).

Since a primary purpose of using the decoder is to identify informative dimensions in the neurodynamics, we estimate new decoder parameters at each stimulus noise level (coherence level) independently. Alternatively we could train a decoder at the highest coherence level and test the decoder at each coherence level. In this paper we focus on the first approach, since we view our decoder as a tool for analyzing the information content in the neurodynamics and how downstream neurons might best decode this information for discrimination.

We constructed an optimal decoder to read out the information in our spike neuron model, fully exploring the spatio-temporal dynamics. The spike train for each neuron in the population is:

\[ s_{i,k}(t) = \sum_j S(t - t_{i,j,k}), \]

where \( t \in [0,250] \text{ ms} \), \( i = 1 \ldots N \) is the index for neurons, \( k = 1 \ldots M \) is the index for trials, \( l = 1 \ldots P \) is the index for spikes. Based on the population spike trains, we estimated the firing rate on each trial by counting the number of spikes within a time bin of width \( \tau \), resulting in a spike count matrix:

\[ r_{i,j,k} = \int_{(j-1)\tau}^{j\tau} s_{i,k}(t) dt, \]

where \( i = 1 \ldots N \) represents the \( i \)th neuron, \( j = 1 \ldots T/\tau \) represents the \( j \)th time bin, \( k = 1 \ldots M \) represents the \( k \)th trial. Note that we explored decoding using time bins of different length. When \( \tau = 25 \text{ ms} \), we assume that information is encoded in both neuron and time, since the firing rate is closer to instantaneous firing rate; when \( \tau = 250 \text{ ms} \), we integrate the spiking activity over the entire trial, leading to a rate-based representation of information. A separate post hoc analysis showed that 25 ms was in fact the bin width that yielded the highest discrimination accuracy (bin width varied from 5 to 250 ms). The class label of each sample \( b_l \) takes the value of \{-1, +1\} representing either face or car with \( M \) being the number of trials. In order to explore the information within the spatio-temporal dynamics, we compute the weighted sum of firing rate over different neurons and time bins. This leads to seeking the solution of the following constrained minimization problem:

\[
\{w, v\} = \underset{w, v}{\arg \min} \frac{1}{M} \sum_{k=1}^{M} L((w^T x_k + v)b_k) + \lambda J(w),
\]

where the first term is the empirical logistic loss function, and the second term is the regularization function, with \( \lambda > 0 \) as the regularization parameter. We create a stacked version of the spike count matrix:

\[ x_{i,k} = r_{i,j,k} \quad \text{with} \quad l = (i-1)N + j, \text{i.e., stacking the neuron and time bin dimensions together.}
\]

The resulting linear decoder can be geometrically interpreted as a hyperplane that separates the classes of face and car, where \( w \) represents the weights for the linear decoder, and \( v \) is the offset. In the case of the sparse decoder, we use an L1 regularization term \( J(w) = \|w\|_1 \); alternatively for the non-sparse decoder, we use the L2 regularization \( J(w) = \|w\|_2^2 \). In the language of Bayesian analysis, the logistic loss term comes from maximum likelihood, L1 corresponds to the Laplacian prior, and L2 corresponds to the Gaussian prior. L1-regularized logistic regression results in a sparse solution for the weights (Krishnapuram et al., 2005; Koh et al., 2007; Meier et al., 2008). So-called “sparse logistic regression” serves as an approach for feature selection, where features that are most informative about the classification survive in the form of non-zero weights (Ng, 2004). We developed an efficient and accurate method to solve this optimization problem (Shi et al., 2010, 2011). Once we learn the hyperplane, for any new image, we can predict the image category via the sign of \( w^T x + v \).

Figure 3 provides a geometric intuition of why L1 and L2 regularization lead to sparse and non-sparse solutions, respectively.
The V1 model, we used two approaches. One way to utilize dynamics was based on a heuristic approach, where we only consider the one that results in the highest \( \Delta z \) value. Az (area under ROC curve) values through cross validation. The best window is also measured sparseness in the spatial domain. Figure 4C shows the time course of the firing rate for one example neuron over 50 trials. The firing rate of this neuron is sparse over time. One can also see the increase in firing rate 50–100 ms shortly after the stimulus onset. Figure 4B plots the histogram of firing rates for this neuron, averaged over 50 trials, showing that the firing rates of all neurons at 50 ms during one trial. The distribution of firing rates is shown in Figure 4D, with a kurtosis of 10.26. Thus for the stimuli we consider for constructing our 2-AFC discrimination paradigm (faces vs. cars), the V1 model produced population responses which are relatively spatially sparse and marginally temporally sparse. However they are not highly sparse (e.g., kurtosis > 100), as has been seen in models which impose hard sparseness (Rehn and Sommer, 2007).

### RESULTS

**ENCODING BY THE V1 MODEL**

We first characterized the sparseness in the neuronal population activity given face and car stimuli (see Materials and Methods). We measured both the lifetime sparseness and population sparseness for the network using a flashed presentation of the sample images. Figure 4A shows the time course of the firing rate for one example neuron over 50 trials. The firing rate of this neuron is sparse over time. One can also see the increase in firing rate 50–100 ms shortly after the stimulus onset. Figure 4B plots the histogram of firing rates for this neuron, averaged over 50 trials, showing that the kurtosis is 1.65, where a positive kurtosis indicates sparseness and a kurtosis of zero indicates a Gaussian distribution. Similarly, we also measured sparseness in the spatial domain. Figure 4C shows the firing rates of all neurons at 50 ms during one trial. The distribution of firing rates is shown in Figure 4D, with a kurtosis of 10.26. Thus for the stimuli we consider for constructing our 2-AFC discrimination paradigm (faces vs. cars), the V1 model produced population responses which are relatively spatially sparse and marginally temporally sparse. However they are not highly sparse (e.g., kurtosis > 100), as has been seen in models which impose hard sparseness (Rehn and Sommer, 2007).

**DECODING SPATIALLY AVERAGED FIRING RATE**

Next we analyzed the spatially averaged firing rate from the V1 model to see whether it carried information for discriminating stimulus class. Specifically we constructed a proxy for multi-unit activity by averaging firing rates over all 4096 cortical neurons and all 30 trials for each phase coherence level. We first investigated whether spatially averaged firing rates were sensitive to phase

Population sparseness characterizes the activity of a population of neurons for a given time window. We estimate instantaneous firing rates using a Gaussian window 25 ms wide with a standard deviation of 5 ms. Sparseness in firing rates can be measured by kurtosis (Olshausen and Field, 2004), namely the forth moment relative to the variance squared.

Using the sparse decoding framework, we are able to identify the informative dimensions that are critical for our specific decision making task. We define “informative dimensions” as the number of non-zero weights in the decoder, which is equal to the cardinality of the weight vector. Informative dimensions thus reflect the number of non-zeros in the spatio-temporal “word.” Note one neuron can be selected by the decoder at multiple time bins, therefore, we define “informative neurons” as the number of neurons having at least one non-zero weight across different time bins.

**STATISTICAL TESTS**

We used a likelihood ratio test to evaluate the goodness of fit. We fit a single Weibull curve jointly to both the psychometric and neurometric dataset (dof = 4), as well as fitting two Weibull curves independently to both dataset (dof = 8). We computed the likelihood ratio using \( D = -2 \ln(l_j/p_{ln}) \). The null hypothesis is psychometric and neurometric data can be described by the same curve, and the decision rule is based on the Chi-square statistics \( \chi^2 \). If \( p > 0.05 \), do not reject null hypothesis; otherwise, reject null hypothesis.

We characterized the sparseness of the neural representation in the population spike trains, for both the temporal and spatial domains. According to Willmore and Tolhurst (2001), lifetime sparseness describes the activity of a single neuron over time, while
coherence levels in the stimulus. Figure 5 shows that the spatially averaged firing rate decreases as the coherence level decreases for face trials and car trials. This is consistent with the V1 neurons being selective to oriented edge energies, which are reduced as the phase coherence decreases. Although spatially averaged firing rates in V1 have never been directly observed for this experimental paradigm, evidence for graded responses, as a function of task difficulty (i.e., coherence level), have been observed in full scalp EEG, particularly in electrodes over visual cortex (Philiastides et al., 2006).

We then considered the difference in the spatially averaged firing rate between face and car trials. Figure 5C shows a graded response for the average difference, with the trend being toward smaller differences for lower coherence levels. To investigate how discriminating this difference was on a trial-to-trial basis, we constructed histograms of the spatially averaged firing rates for both face and car trials. We constructed a neurometric function by computing the area under the receiver operating curve (ROC) curve (i.e., area given as Az) at each coherence level and fitting a Weibull function to these points. The neurometric curve was constructed based on the time window from 50–150 ms post-stimulus, which is the time window with the maximum mean difference in the spatially average firing rates for the two stimulus classes. The resulting neurometric curve (black curve), shown in Figure 5D, was a poor match to the group psychometric curve (red curve), which was computed by averaging over the behavioral data of 10 human subjects performing the same two-class discrimination task. The finding that the stimulus class, for such complex stimuli as faces and cars, cannot be decoded from the spatially averaged firing rate is also consistent with previous findings measured by EEG (Philiastides and Sajda, 2006), where EEG in visual cortex is selective for coherence level but not selective for image class.

DECODING SPATIO-TEMPORAL DYNAMICS

The above observation prompted us to investigate whether a more fine-grained mesoscopic decoding strategy holds information for stimulus class discrimination. We investigated a sparse decoding strategy that exploits the spatio-temporal neurodynamics produced by the V1 model.

Specifically, we investigated whether the transient nature of the neurodynamics is critical for learning a linear discrimination boundary. If discrimination is driven by a sustained activation of neurons over time, then an optimal decoding strategy would be based on integrating the neuronal spiking activities over a substantial fraction of the trial. On the other hand, if the transient
dynamics are critical for discrimination, then a finer grain temporal integration of the dynamics would yield better discrimination. We evaluated the difference between these two cases by constructing a neural “word” using integration time bins of different lengths. A time bin of 25 ms captures the transient information in the neural activities while a time bin of 250 ms would treat the neural information in a sustained manner. Theoretical studies have investigated how information content in the neural activity can vary as a function of the length of the window used for estimation (Panzari et al., 1999). Here we empirically study this within the context of the neural activity generated by the model. Figure 6 illustrates how the spike count matrix for discrimination is constructed. Note that the dynamic word has a dimensionality that is 10 times larger than the sustained word.

The decoding results, shown in Figure 7A, reveal that the predicted discrimination accuracy is substantially higher when transient neurodynamics are utilized. The neurometric curve for the transient dynamics decoding (i.e., temporal coding model, black curve) lies within the variation of the human behavioral data, whereas the neurometric curve for the sustained decoding (gray curve) is clearly outside the range of human behavioral performance. Figure 7B shows the mean neurometric curve (black curve), constructed by averaging the discrimination accuracy across six random initializations of the model. The average psychometric curve (average of 10 subjects, red curve) is shown for comparison. The significance of the similarity between the neurometric and psychometric curves should not be overlooked, given that the model of V1 is not optimized/tuned to match the psychophysics nor is the decoder trained to match the psychophysics. Taken together with the fact that the V1 model also produces realistic classical and extraclassical physiological responses, at both the single cell and population level (Wielaard and Sajda, 2006a,b), this result provides validation for this modeling substrate.

We further investigated the role dynamics plays in decoding accuracy and the fit of the neurometric curve relative to human psychometric performance by comparing the decoding of activity generated by our V1 model to the activity generated by the first layer of the feedforward HMAX model (Serre et al., 2007). Shown in Figure 8, this version of HMAX, which has no dynamics, results in a neurometric curve that substantially deviates from the psychometric performance, overestimating decoding accuracy.
FIGURE 6 | (A) Spike trains from simulated neurons are aligned relative to stimulus onset. Words are constructed by binning spike trains using a given temporal bin width. (B) A spatio-temporal word represented as a matrix of size \(N\) neurons by 10 time bins, each time bin being 25 ms wide. Numbers indicate spike counts for bins with at least one spike. (C) A word represented as a vector of length \(N\), computed using a single bin of size 250 ms.

FIGURE 7 | (A) Simulated neurometric performance for one initialization of the V1 model. Shown are (black) neurometric curve constructed by decoding the full spatio-temporal word, \(D = 3.19, p = 0.53\); (gray) neurometric curve constructed from a decoder that ignores dynamics, \(D = 52.26, p < 0.01\). Also shown is (dashed red lines) the psychophysical performance of 10 human subjects, and (solid red curve) the group psychometric curve across 10 subjects. Error bars on both curves represent standard error. Likelihood ratio test yields \(D = 6.42, p = 0.17\).

FIGURE 8 | Decoding accuracy compared to a feedforward model. Comparison of psychometric and two neurometric curves, one for sparse linear decoding of activity generated by our dynamical V1 model (black curved – same as Figure 7A, \(D = 3.19, p = 0.53\)) and the other for a sparse linear decoding of activity generated by the first layer of the feedforward HMAX model (gray curve, \(D = 48.26 (p < 0.01)\). Code for simulating HMAX feedforward model is freely available from http://cbcl.mit.edu/jmutch/cns/hmax/doc/.

at low coherence levels and underestimating it at high coherence levels.

Finally we investigated how decoding accuracy is affected by the degree of sparsity in the neural activity. Table 1 was constructed by varying the initial conditions of the model in such a way that physiological response properties, such as orientation tuning and modulation ratios (i.e., fraction of simple and complex cells), were not significantly affected though there were observed changes in both spatial and temporal sparsity. As can be seen in Table 1, the decoding accuracy is strongly correlated with the spatial sparseness of the activity (Pearson’s \(r = 0.96, p = 0.009\) but not significantly correlated with temporal sparseness (Pearson’s \(r = -0.65, p = 0.24\) with the higher decoding accuracy associated with a greater degree of spatial sparseness.

ALTERNATIVE DECODING STRATEGIES

In this work, we use the decoder to evaluate the information content for varying level of stimulus noise. In this regard, we compared several decoding strategies. Figure 9A shows neurometric curves for two decoding scenarios: (black) train and test the decoder at each coherence independently (gray), train at the highest coherence and test at each coherence. Simulation results indicate training and testing the decoder on each coherence level independently, though suboptimal relative to training at the highest coherence, is a better match to the human psychophysics. This would suggest that not only information about the signal but also information about the noise distribution is potentially used by the subjects in the task.

Figure 9B compares sparse and non-sparse decoding strategies: (black) sparse decoding (gray), non-sparse decoding. In this

www.frontiersin.org
Table 1 | Five V1 models differing in their spatial and temporal sparsity, measured via kurtosis, and the resulting classification performance measured via the area under the ROC curve (Az).

| Model | Spatial Sparseness (Kurtosis) | Temporal Sparseness (Kurtosis) | Classification (Az) |
|-------|-------------------------------|-------------------------------|---------------------|
| Model 1 | 8.50                          | 2.51                          | 0.73                |
| Model 2 | 11.83                         | 1.42                          | 0.84                |
| Model 3 | 14.91                         | 0.08                          | 0.87                |
| Model 4 | 8.84                          | 0.60                          | 0.76                |
| Model 5 | 9.86                          | 1.15                          | 0.78                |

FIGURE 9 | Comparison of several decoding strategies. (A) Two decoding strategies: (black) train and test at each coherence independently, \( D = 3.19, P = 0.53 \); (gray) train at the highest coherence and test at each coherence, \( D = 12.88, P = 0.01 \). Both are trained as sparse decoders. (B) Sparse and non-sparse decoding strategies: (black) sparse decoder, \( D = 3.19, P = 0.53 \); (gray) non-sparse decoder each with the same amount of regularization, \( D = 3.66, P = 0.45 \). Both are trained on each coherence independently. Also shown are (dashed red lines) the psychophysical performance of 10 human subjects; (solid red curve) the group psychometric curve across 10 subjects.

computer vision and deep learning community have investigated the importance of sparsity in classification tasks, with sparse autoencoding being an example of how sparsity can improve the final classification result (Ranzato et al., 2007; Coates and Ng, 2011).

INFORMATIVE DIMENSIONS

The sparse decoder not only makes decoding of high-dimensional features feasible, but also facilitates an avenue for investigating feature selection and quantifying the informative dimensions. Figure 11A shows that the number of informative dimensions increases as the task becomes more difficult, i.e., the decoder will exploit additional dimensionality in the neural word to improve decoding performance. Realizing the fact that one neuron might participate in the decision process multiple times, we also computed the number of informative neurons, which is the number of uniquely chosen neurons in the neural word. Our results indicated that the number of informative neurons also decreases as the task gets easier, as shown in Figure 11B. Such an observation offers possible insight into the tradeoff between decoding accuracy and metabolic cost. Finally, Figure 11C shows the number of neurons in which more than one time bin is selected in the neural word. Thus most neurons participate only once in the decoding process.

“MELODY” WITHIN NEURAL DYNAMICS

We analyzed the spatial and temporal distributions of the informative dimensions arising from the sparse decoding framework. Figure 12A shows three coherence levels while Figure 12B shows the spatial distribution of neurons selected for these three levels. It is clear that more neurons are selected as the task becomes more difficult. Given that the stimuli are presented as a monocular simulation, one can see “banding” indicative of ocular dominance columns. We also investigated the temporal distribution of the selected neurons. Using an adaptive windowing technique
(described in the Materials and Methods) we found a pattern of activity that can best be described as a “melody.” See Figure 12C, which shows that the visual stimulus elicits a spatio-temporal pattern in the cortical activity that carries information about discrimination. Our finding is similar to the activation patterns that have been observed in vitro for spontaneous cortical activity (Cossart et al., 2003; Watson et al., 2008).

The adaptive windowing technique also allows us to investigate the optimal time window over which the decoder integrates information to form a decision. Figure 12D can be interpreted as the network’s chronometric function, namely the time needed for the network to form the decision. Though the network does not consider motor preparation or other possible sources of delay, and thus the chronometric function cannot be strictly interpreted as a measure of reaction time, this trend is very consistent with subjects’ reaction time curves, see Figure 12E. Thus we see that by reading out the neurodynamics of the V1 model using a sparse linear decoder, we observe accuracy and timing for the discrimination that are very much consistent with the results for human subjects.

**FIGURE 11 |** (A) Number of informative dimensions decrease as the task becomes easier. (B) Number of informative neurons decreases as the task becomes easier. (C) Difference between the number of informative dimensions and the number of informative neurons.

**FIGURE 12 |** (A) Example of images at three coherence levels. (B) Spatial distribution of neurons selected by the decoder. Shown are neurons in the cortical space, where selected neurons are indicated in black. (C) Temporal windows selected by the decoder together with their corresponding weights. (D) Chronometric function indicating the time needed by the decoder to accumulate evidence in the network dynamics. (E) Behavioral chronometric functions derived from human psychophysics.
We also find that short sampling intervals/windows (25 ms) yield a follow-up study using single-unit and multi-unit recordings in all angular discrimination tasks (Kang and Sompolinsky, 2001; Seung and Sompolinsky, 1990). Fisher information was utilized in later models to study the discrimination of tuning curves (Bradley et al., 1987; Hawken and Parker, 1990). Fisher information was utilized in later models to study the discrimination of tuning curves (Seung and Sompolinsky, 1993; Abbott and Dayan, 1999; Sompolinsky et al., 2001) under the assumption that information was not limited to a point on the tuning curve, but was subject to the limitation that Fisher information could only be measured for a very small orientation difference. Besides the population code, a further study using the Chernoff distance between two distributions of spike counts paid attention to the global information of a tuning curve, and found that narrow orientation tuning was not necessarily optimal for all angular discrimination tasks (Kang and Sompolinsky, 2001; Kang et al., 2004). These studies, however, all resorted to using the average responses of neurons across time, therefore ignoring the rich dynamics of the spiking activity of the neuronal population.

On the other hand, substantial work has been done to characterize the temporal characteristics of neural dynamics using statistical models. Bayesian methods (Bialek and Zee, 1990; Bialek et al., 1991; Sanger, 1996; Osborne et al., 2005) are used to characterize the spiking dynamics of neurons, providing the flexibility for statistical inference (Mendel, 1995; Brown et al., 1998). Point process likelihood-based generalized linear models (GLM) have been used to analyze history dependence in neural spiking activity. State-space models provide a flexible tool for neural spike train decoding (Brown et al., 1998; Barbieri et al., 2004; Brockwell et al., 2004; Paninski, 2004; Wu et al., 2006; Deneve et al., 2007; Czanner et al., 2008). These statistical methods typically model the neural dynamics with some assumptions on the hidden brain states, and use a generative approach to infer the stimuli based on the neural dynamics. Latency coding scheme (Shriki et al., 2012) and spike timing dependent plasticity (Masquelier and Thorpe, 2007) have been investigated for decoding V1 information, all pointing to the conclusion temporal dynamics is critical for object recognition. Our results have focused on a discriminative approach, mapping the neural dynamics into a (binary) decision.

### RELATIONSHIP TO OTHER DECODING METHODS

Several investigators have argued for population coding (Pouget et al., 2000, 2002; Averbeck et al., 2006) as a model of orientation discrimination, based on the assumption that orientation tuning, computed from average response over the period of sinusoidal drifting gratings, was responsible for discrimination. Models based on selecting the most informative point on a tuning curve, usually the steepest point, only utilized the local shape of the tuning curve (Bradley et al., 1987; Hawken and Parker, 1990). Fisher information was utilized in later models to study the discriminating information for tuning curves (Seung and Sompolinsky, 1993; Abbott and Dayan, 1999; Sompolinsky et al., 2001) under the assumption that information was not limited to a point on the tuning curve, but was subject to the limitation that Fisher information could only be measured for a very small orientation difference. Besides the population code, a further study using the Chernoff distance between two distributions of spike counts paid attention to the global information of a tuning curve, and found that narrow orientation tuning was not necessarily optimal for all angular discrimination tasks (Kang and Sompolinsky, 2001; Kang et al., 2004). These studies, however, all resorted to using the average responses of neurons across time, therefore ignoring the rich dynamics of the spiking activity of the neuronal population.

On the other hand, substantial work has been done to characterize the temporal characteristics of neural dynamics using statistical models. Bayesian methods (Bialek and Zee, 1990; Bialek et al., 1991; Sanger, 1996; Osborne et al., 2005) are used to characterize the spiking dynamics of neurons, providing the flexibility for statistical inference (Mendel, 1995; Brown et al., 1998). Point process likelihood-based generalized linear models (GLM) have been used to analyze history dependence in neural spiking activity. State-space models provide a flexible tool for neural spike train decoding (Brown et al., 1998; Barbieri et al., 2004; Brockwell et al., 2004; Paninski, 2004; Wu et al., 2006; Deneve et al., 2007; Czanner et al., 2008). These statistical methods typically model the neural dynamics with some assumptions on the hidden brain states, and use a generative approach to infer the stimuli based on the neural dynamics. Latency coding scheme (Shriki et al., 2012) and spike timing dependent plasticity (Masquelier and Thorpe, 2007) have been investigated for decoding V1 information, all pointing to the conclusion temporal dynamics is critical for object recognition. Our results have focused on a discriminative approach, mapping the neural dynamics into a (binary) decision.
CAVEATS OF A MODEL
A potential criticism of this work is that our results are based on decoding of activity generated from a computational model of V1 and not from experimental recordings. Our model was designed to investigate classical and extraclassical response properties of large population of neurons in early visual cortex. Thus though the neural activity we decode is simulated, our model is physiologically realistic, validated against experimental data across a wide range of responses.

Complex visual objects are likely not directly discriminated from V1 responses alone, and more complex non-linearities and dynamics, including feedback, likely play a role in how objects are detected, identified, and recognized. However we believe that the paradigm presented in this paper represents a first step in understanding how complex images might be encoded within the V1 activity and be ultimately decoded further downstream.

ACKNOWLEDGMENTS
This work was supported by NGA NURI Grant HM1582-07-1-2002, NIH Grant R01EY015520-01A2 a grant from the Army Research Office (W911NF-11-1-0219) and New York Community Trust.

REFERENCES
Abbott, L. F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. Neural Comput. 11, 91–101.
Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nat. Rev. Neurosci. 7, 358–366.
Baddeley, R., Abbott, L. F., Booth, M. C., www.frontiersin.org

Bialek, W., Rieke, F., de Ruyter van Koven, G., H. (2002). A new framework for neural spike train modeling. In The geometry of neuronal interactions and large-scale brain function, ed. C. Koch, pp. 457–507. MIT Press, Cambridge, MA.

Blakemore (Cambridge: Cambridge University Press), 103–116.

Brecht, M., Schneider, M., Sakmann, B., and Margrie, T. W. (2004). Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. Nature 427, 704–710.

Brown, E. N. (2008). Analysis of neural population responses in a reaction-time visual detection task. J. Neurosci. 28, 4958–4961.

Kang, K., Shapley, R. M., and Sompilin- sky, H. (2004). Information tuning of populations of neurons in primary visual cortex. J. Neurosci. 24, 3726–3735.

Kang, K., and Sompilinskii, H. (2001). Mutual information of population codes and distance measures in probability space. Phys. Rev. Lett. 86, 1519–1555.

Krishnapuram, B., Carin, L., Figueiredo, M. A. T., and Hartemink, A. J. (2005). Sparse multinomial logregression: Fast algorithms and generalization bounds. IEEE Trans. Pattern Anal. Mach. Intell. 957–968.

Legenstein, R., Pecevski, D., and Maass, W. (2008). A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. PLoS Comput. Biol. 4:e1000180. doi:10.1371/journal.pcbi.1000180

Maass, W., Natschlager, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput. 14, 2531–2560.

Malpeli, J. G., Lee, D., and Baker, F. H. (1996). Laminar and retinotopic organization of the macaque lateral geniculate nucleus: magnocellular and parvocellular magnification functions. J. Comp. Neurol. 375, 363–377.
Olshausen, B. A., and Field, D. J. (2004). Natural image statistics: a strategy employed by V1? Cereb. Cortex 16, 509–518.

Shi et al. Perceptual decision making in a V1 model

Mendel, J. M. (1995). Lessons in Estimation Theory for Signal Processing, Communications, and Control (Book). Englewood Cliffs, NJ: Prentice Hall PTR.

Ng, A. Y. (2004). “Feature selection, L 1 vs. L 2 regularization, and rotational invariance,” in Proceedings of the Twenty-First International Conference on Machine Learning (New York: ACM).

Olshausen, B. A., and Field, D. J. (1996a). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609.

Olshausen, B. A., and Field, D. J. (1996b). Natural image statistics and efficient coding. Network 7, 335–339.

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res. 37, 3311–3321.

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. Curr. Opin. Neurobiol. 14, 481–487.

Osborne, L. C., Lisberger, S. G., and Bialek, W. (2005). A sensory source for motor variation. Nature 437, 412–416.

Palmer, C., Cheng, S. Y., and Seidemann, E. (2007). Linking neuronal and behavioral performance in a reaction-time visual detection task. J. Neurosci. 27, 8122–8137.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. Network 15, 243–262.

Panzeri, S., Treves, A., Schultz, S., and Rolls, E. T. (1999). On decoding the responses of a population of neurons from short time windows. Neural Comput. 11, 1553–1577.

Peretz-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., and Laurent, G. (2002). Oscillations and sparsening of odor representations in the mushroom body. Science 297, 359–365.

Philastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J. Neurosci. 26, 8965–8975.

Philastides, M. G., and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb. Cortex 16, 509–518.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. Nat. Rev. Neurosci. 1, 125–132.

Pouget, A., Denève, S., and Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. Nat. Rev. Neurosci. 5, 741–747.

Quiroga, R. Q., Reddy, L., Koch, C., and Fried, I. (2007). Decoding visual inputs from multiple neurons in the human temporal lobe. J. Neurophysiol. 98, 1997–2007.

Ranzato, M., Boureau, Y., and LeCun, Y. (2007). “Sparse feature learning for deep belief networks,” in Advances in Neural Information Processing Systems, Vancouver, 2007.

Rehn, M., and Sommer, F. T. (2007). A network that uses few active neurons to code visual input predicts the diverse shapes of cortical receptive fields. J. Comput. Neurosci. 22, 135–146.

Rinberg, D., Koulakov, A., and Gelperin, A. (2006). Sparse odor coding in awake behaving mice. J. Neurosci. 26, 8857–8865.

Rolls, E. T., and Tovee, M. J. (1995). Sparserness of the neuronal representation of stimuli in the primate temporal visual cortex. J. Neurophysiol. 73, 713–726.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. J. Neurophysiol. 76, 2790–2793.

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. U.S.A. 104, 6424–6429.

Shi, J., Yin, W., Osher, S., and Sajda, P. (2010). A Fast Hybrid Algorithm for Large-Scale L 1-Regularized Logistic Regression. J. Mach. Learn. Res. 11, 713–741.

Shi, J. V., Wieland, J., Smith, R. T., and Sajda, P. (2011). Decoding simulated neurodynamics predicts the perceptual consequences of age-related macular degeneration. J. Vis. 11, 4.

Shriki, O., Kohn, A., and Shamir, M. (2012). Fast coding of orientation in primary visual cortex. PLoS Comput. Biol. 8:e1002536. doi:10.1371/journal.pcbi.1002536

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. Annu. Rev. Neurosci. 24, 1193–1216.

Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 64, 051904.

Spear, P. D., Moore, R. I., Kim, C. B., Xue, J. T., and Tumosa, N. (1994). Effects of aging on the primate visual system: spatial and temporal processing by lateral geniculate neurons in young adult and old rhesus monkeys. J. Neurophysiol. 72, 402–420.

Stryksza, P., Diizen, M., Galkin, A., Galizia, C. G., and Menzel, R. (2005). Sparsening and temporal sharpening of olfactory representations in the honeybee mushroom bodies. J. Neurophysiol. 94, 3303–3313.

Vinje, W. E., and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287, 1273–1276.

Vinje, W. E., and Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. J. Neurosci. 22, 2904–2915.

Watson, B. O., MacLean, J. N., and Yuste, R. (2008). UP states promote ongoing cortical activity from thalamic inputs. PLoS ONE 3:e3971. doi:10.1371/journal.pone.0003971

Wieland, D. J., Shelley, M., McLaughlin, D., and Shelley, R. (2001). How simple cells are made in a nonlinear network model of the visual cortex. J. Neurosci. 21, 5203.

Wieland, J., and Sajda, P. (2006a). Circuitry and the classification of simple and complex cells in V1. J. Neurophysiol. 96, 2739–2749.

Wieland, J., and Sajda, P. (2006b). Extraclassical receptive field phenomena and short-range connectivity in V1. Cereb. Cortex 16, 1531–1545.

Wiesel, T., and Takahashi, D. J. (2001). Characterizing the sparseness of neural codes. Network 12, 255–270.

Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P., and Black, M. J. (2006). Bayesian population decoding of motor cortical activity using a Kalman filter. Neural Comput. 18, 80–118.

Yamazaki, T., and Tanaka, S. (2007). A spiking network model for passage-of-time representation in the cerebellum. Eur. J. Neurosci. 26, 2279–2292.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 September 2012; paper pending published: 07 October 2012; accepted: 14 March 2013; published online: 19 April 2013.

Citation: Shi JV, Wieland J, Smith RT and Sajda P (2013) Perceptual decision making “Through the Eyes” of a large-scale neural model of V1. Front. Psychol. 4:161. doi: 10.3389/fpsyg.2013.00161

This article was submitted to Frontiers in Perception Science, a specialty of Frontiers in Psychology.

Copyright © 2013 Shi, Wieland, Smith and Sajda. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.