The relationship between optimal and biologically plausible decoding of stimulus velocity in the retina

Edmund C. Lalor,[1,*] Yashar Ahmadian,[2] and Liam Paninski[2]

[1]*Trinity Centre for Bioengineering and Institute of Neuroscience, Trinity College Dublin, College Green, Dublin 2, Ireland*
[2]*Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, New York 10027, USA*
**Corresponding author: edlalor@tcd.ie*

Received January 30, 2009; revised June 14, 2009; accepted July 23, 2009; posted August 7, 2009 (Doc. ID 106996); published September 11, 2009

A major open problem in systems neuroscience is to understand the relationship between behavior and the detailed spiking properties of neural populations. We assess how faithfully velocity information can be decoded from a population of spiking model retinal neurons whose spatiotemporal receptive fields and ensemble spike train dynamics are closely matched to real data. We describe how to compute the optimal Bayesian estimate of image velocity given the population spike train response and show that, in the case of global translation of an image with known intensity profile, on average the spike train ensemble signals speed with a fractional standard deviation of about 2% across a specific set of stimulus conditions. We further show how to compute the Bayesian velocity estimate in the case where we only have some *a priori* information about the (naturalistic) spatial correlation structure of the image but do not know the image explicitly. As expected, the performance of the Bayesian decoder is shown to be less accurate with decreasing prior image information. There turns out to be a close mathematical connection between a biologically plausible "motion energy" method for decoding the velocity and the Bayesian decoder in the case that the image is not known. Simulations using the motion energy method and the Bayesian decoder with unknown image reveal that they result in fractional standard deviations of 10% and 6%, respectively, across the same set of stimulus conditions. Estimation performance is rather insensitive to the details of the precise receptive field location, correlated activity between cells, and spike timing.

OCIS codes: 330.4060, 330.4150.

1. INTRODUCTION

The question of how different attributes of a visual stimulus are represented by populations of cells in the retina has been addressed in a number of recent studies [1–8]. This field has received a major boost with the advent of methods for obtaining large-scale simultaneous recordings from multiple retinal ganglion neurons that almost completely tile a substantial region of the visual field [9,10]. The utility of this new method for understanding the encoding of behaviorally relevant signals was exemplified by [4], where the authors examined the question of how reliably visual motion was encoded in the spiking activity of a population of macaque parasol cells. These authors used a simple moving stimulus and attempted to estimate the velocity of that stimulus from the resulting spike train ensemble; this analysis pointed to some important constraints on the visual system's ability to decode image velocity given noisy spike train responses. We will explore these issues in more depth in this paper.

In parallel to these advances in retinal recording technology, significant recent advances have also been made in our ability to model the statistical properties of populations of spiking neurons. For example, a statistical model of a complete population of primate parasol retinal ganglion cells (RGCs) was recently described [7]. This model was fit using data acquired by the array recording techniques mentioned above and includes spike-history effects and cross-coupling between cells of the same kind and of different kinds (i.e., ON and OFF cells). The authors demonstrated that the model accurately captures the stimulus dependence and spatiotemporal correlation structure of RGC population responses, and allows several insights to be made into the retinal neural code. One such insight concerns the role of correlated activity in preserving sensory information. Using pseudorandom binary stimuli and Bayesian inference, they reported that stimulus decoding based on the spiking output of the model preserved 20% more information when knowledge of the correlation structure was used than when the responses were considered independently [7].

At the psychophysical level, Bayesian inference has been established as an effective framework for understanding visual perception [11]; some recent notable applications to understanding visual velocity processing include [12–17]. In particular, [14] argued that a number of visual illusions actually arise naturally in a system that attempts to estimate local image velocity via Bayesian methods (though see also [18,19]).

Links between retinal coding and psychophysical behavior have also been recently examined using Bayesian methods; [20,21], for example, examine the contribution of turtle RGC responses to velocity and acceleration encoding. This study reported that the instantaneous firing rates of *individual* turtle RGCs contain information about

speed, direction, and acceleration of moving patterns. The firing-rate-based Bayesian stimulus reconstruction carried out in that study involved a couple of key approximations. These included the assumptions that RGCs generate spikes according to Poisson statistics and that they do so independently of each other. The work of [7] emphasizes that these assumptions are unrealistic, but the impact of detailed spike timing and correlation information on velocity decoding remains uncertain.

The primary goal of this paper is to investigate the fidelity with which the velocity of a visual stimulus may be estimated, given the detailed spiking responses of the primate RGC population model of [7], using Bayesian decoders, with and without full prior knowledge of the image. We begin by describing the mathematical construction of the Bayesian decoders, and then compare these estimates to those based on a biologically plausible "net motion signal" derived directly from the spike trains without any prior image information [4]. We derive a mathematical connection between these decoders and investigate the decoders' performance through a series of simulations.

2. METHODS

A. Model

The generalized linear model (GLM) [22,23] for the spiking responses of the sensory network used in this study was described in detail in [7]. It consists of an array of ON and OFF retinal ganglion cells (RGC) with specific baseline firing rates. Given the spatiotemporal image movie sequence, the model generates a mean firing rate for each cell, taking into account the temporal dynamics and the center-surround spatial stimulus filtering properties of the cells. Then, incorporating spike history effects and cross-coupling between cells of the same type and of the opposite type, it generates spikes for each cell as a stochastic point process.

In response to the visual stimulus $\mathbf{I}$, the ith cell in the observed population emits a spike train, which we represent by a response function

$$r_i(t) = \sum_{\alpha} \delta(t - t_{i,\alpha}), \tag{1}$$

where each spike is represented by a delta function, and $t_{i,\alpha}$ is the time of the αth spike of the ith neuron. We use the shorthand notation $\mathbf{r}_i$ and $\mathbf{r}$ for the response function of one neuron and the collective spike train responses of all neurons, respectively. The stimulus $\mathbf{I}$ represents the spatiotemporal luminance profile $I(\mathbf{n},t)$ of a movie as a function of the pixel position $\mathbf{n}$ and time t.

In the GLM framework, the intensity functions (instantaneous firing rate) of the responses $\mathbf{r}_i$ are given by [7,24–26]

$$\lambda_i(t) \equiv f\left(b_i + J_i(t) + \sum_{j,\beta} h_{ij}(t - t_{j,\beta})\right), \tag{2}$$

where $f(\cdot)$ is a positive, strictly increasing rectifying function. As in [7], we adopt the choice $f(\cdot)=\exp(\cdot)$. The b_i represents the log of the baseline firing rate of the cell, the coupling terms h_{ij} model the within- and between-neuron spike history effects noted above, and the stimulus input $J_i(t)$ is obtained from $\mathbf{I}$ by linearly filtering the spatiotemporal luminance,

$$J_i(t) = \int\int k_i(t - \tau, \mathbf{n}) I(\tau, \mathbf{n}) \mathrm{d}^2\mathbf{n}\mathrm{d}\tau, \tag{3}$$

where $k_i(t,\mathbf{n})$ is the spatiotemporal receptive field of the cell i. The parameters for each cell were fit using 7 min of spiking data recorded during the presentation of a nonrepeating stimulus, with the baseline log firing rate being a constant and the various filter parameters being fit using a basis of raised cosine "bumps" [7]. Given Eq. (2), we can write down the point process log-likelihood in the standard way [27]

$$\log p(\mathbf{r}|\mathbf{I}) \equiv \sum_{i,\alpha} \log \lambda(t_{i,\alpha}) - \sum_i \int_0^T \lambda_i(t)\mathrm{d}t. \tag{4}$$

For movies arising from images rigidly moving with constant velocity $\mathbf{v}$ we have

$$I(t,\mathbf{n}) = x(\mathbf{n} - \mathbf{v}t), \tag{5}$$

where $x(\mathbf{n})$ is the luminance profile of a fixed image. Substituting Eq. (5) into Eq. (3) and shifting the integration variable $\mathbf{n}$ by $\mathbf{v}\tau$, we obtain

$$J_i(t) = \int \mathcal{K}_{i,\mathbf{v}}(t;\mathbf{n}) x(\mathbf{n}) \mathrm{d}^2\mathbf{n}, \tag{6}$$

where we defined

$$\mathcal{K}_{i,\mathbf{v}}(t;\mathbf{n}) \equiv \int k_i(t - \tau, \mathbf{n} + \mathbf{v}\tau)\mathrm{d}\tau. \tag{7}$$

In the following we replace $p(\mathbf{r}|\mathbf{I})$ with its equivalent $p(\mathbf{r}|\mathbf{x},\mathbf{v})$ [since, via Eq. (5), $\mathbf{I}$ is given in terms of $\mathbf{x}$ and $\mathbf{v}$] and use the short-hand matrix notation $\mathbf{J}_i = \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}$ for Eq. (6). An important point is that in the case of a convex and log-concave GLM nonlinearity, $f(\cdot)$ [conditions that are true for our choice, $f(\cdot)=\exp(\cdot)$], the GLM log-likelihood, Eq. (4), is a concave function of $\mathbf{x}(\mathbf{n})$.

B. Decoding

In order to estimate the speed of the moving bar given the simulated output spike trains $\mathbf{r}$ of our RGC population, we employed three distinct methods. The first method involved a Bayesian decoder with full image information, the second method utilized a Bayesian decoder with less than full image information, while the third method involved an "energy-based" algorithm introduced by [4] that used no explicit prior knowledge of the image. For reasons that will become clear, these decoders will be hereafter known as the optimal decoder, the marginal decoder, and the energy method, respectively. Given a simulated output spike train ensemble, we use each of these methods to estimate the speed of the stimulus that evoked the ensemble by maximizing some function across a range of possible or "putative" speeds.

1. Bayesian Velocity Estimation

To compute the optimal Bayesian velocity decoder we need to evaluate the posterior probability for the velocity

$p(\mathbf{v}|\mathbf{r})$ conditional on the observed spike trains $\mathbf{r}$. Given a prior distribution $p_v(\mathbf{v})$, from Bayes' rule we obtain

$$p(\mathbf{v}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{v})\mathbf{p}_v(\mathbf{v})}{\sum_{\mathbf{v}'} p(\mathbf{r}|\mathbf{v}')\mathbf{p}_v(\mathbf{v}')}. \tag{8}$$

If the image $\mathbf{x}$ (e.g., a narrow bar with a luminance distinct from the background) is known to the decoder, then we can replace $p(\mathbf{r}|\mathbf{v})$ with the likelihood function $p(\mathbf{r}|\mathbf{x},\mathbf{v})$, obtaining

$$p(\mathbf{v}|\mathbf{r},\mathbf{x}) = \frac{p(\mathbf{r}|\mathbf{x},\mathbf{v})\mathbf{p}_v(\mathbf{v})}{\sum_{\mathbf{v}'} p(\mathbf{r}|\mathbf{x},\mathbf{v}')\mathbf{p}_v(\mathbf{v}')}. \tag{9}$$

$p(\mathbf{r}|\mathbf{x},\mathbf{v})$ is provided by the forward model Eq. (4), and therefore computation of the posterior probability is straightforward in this case.

Alternatively, if the image is not fully known, we represent the decoder's uncertain *a priori* knowledge regarding $\mathbf{x}$ with an image prior distribution $p_x(\mathbf{x})$. In this case, $p(\mathbf{r}|\mathbf{v})$ is obtained by marginalization over $\mathbf{x}$:

$$p(\mathbf{r}|\mathbf{v}) = \int p(\mathbf{r},\mathbf{x}|\mathbf{v})\mathrm{d}\mathbf{x} = \int p(\mathbf{r}|\mathbf{x},\mathbf{v})p_x(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{10}$$

Hence, we will refer to $p(\mathbf{r}|\mathbf{v})$ as the marginal likelihood. Given the marginal likelihood, Eq. (8) allows us to calculate Bayesian estimates for general velocity priors. The prior distribution $p_x(\mathbf{x})$ which describes the statistics of the image ensemble, can be chosen to have a naturalistic correlation structure. In our simulations in Section 3 we use a Gaussian image ensemble with power spectrum matched to observations in natural images [28,29].

In general, the calculation of the high-dimensional integral over $\mathbf{x}$ in Eq. (10) is a difficult task. However, when the integrand $p(\mathbf{r},\mathbf{x}|\mathbf{v})$ is sharply peaked around its maximum [which is the maximum *a posteriori* (MAP) estimate for $\mathbf{x}$—as the integrand is proportional to the posterior image distribution $p(\mathbf{x}|\mathbf{r},\mathbf{v})$ by Bayes' rule] the so-called "Laplace" approximation (also known as the "saddle-point" approximation) provides an accurate estimate for this integral [for applications of this approximation in the Bayesian setting, see e.g., [30]]. The Laplace approximation in the context of neural decoding is further discussed in, e.g., [31–35]. We briefly review this approximation here.

Following [29], we consider Gaussian image priors with zero mean and covariance $\mathcal{C}_x$ chosen to match the power spectrum of natural images [28]. Let us define the function

$$\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v}) \equiv \log p_x(\mathbf{x}) + \log p(\mathbf{r}|\mathbf{x},\mathbf{v}) + \tfrac{1}{2}\log(2\pi)^d|\mathcal{C}_x|, \tag{11}$$

where d represents the number of pixels in our simulated image, and rewrite Eq. (10) as

$$p(\mathbf{r}|\mathbf{v}) = \frac{1}{\sqrt{(2\pi)^d|\mathcal{C}_x|}} \int e^{\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})}\mathrm{d}\mathbf{x}. \tag{12}$$

Using Eq. (4) and $p_x(\mathbf{x})=\mathcal{N}(0,\mathcal{C}_x)$, we obtain the expression

$$\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v}) = -\tfrac{1}{2}\mathbf{x}^{\mathrm{T}}\mathcal{C}_x^{-1}\mathbf{x} + \sum_i \left[\sum_\alpha \log\lambda_i(t_{i,\alpha};\mathbf{x},\mathbf{r}) - \int \lambda_i(t;\mathbf{x},\mathbf{r})\mathrm{d}t\right], \tag{13}$$

where λ_i are given by Eqs. (2), (6), and (7), and we made their dependence on $\mathbf{x}$ and $\mathbf{r}$ manifest. Since both terms in Eq. (13) are concave (see the closing remarks in Subsection 2.A), the log-posterior $\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})$ is concave in $\mathbf{x}$. To obtain the Laplace approximation, for fixed $\mathbf{r}$, we first find the value of $\mathbf{x}$ that maximizes $\mathcal{L}$ (i.e., the image MAP, $\mathbf{x}_{\mathrm{MAP}}$). When the integrand is sharply concentrated around its maximum, we can Taylor expand $\mathcal{L}$ around $\mathbf{x}_{\mathrm{MAP}}$ to the first nonvanishing order beyond the zeroth-order (i.e., its maximum value) and neglect the rest of the expansion. Since at the maximum the gradient of $\mathcal{L}$ and hence the first-order term vanish, we obtain

$$\mathcal{L}(\mathbf{x},\mathbf{r}) \approx \mathcal{L}(\mathbf{x}_{\mathrm{MAP}},\mathbf{r},\mathbf{v}) - \tfrac{1}{2}(\mathbf{x}-\mathbf{x}_{\mathrm{MAP}})^{\mathrm{T}}H(\mathbf{r},\mathbf{v})(\mathbf{x}-\mathbf{x}_{\mathrm{MAP}}), \tag{14}$$

where the negative Hessian matrix

$$H(\mathbf{r},\mathbf{v}) \equiv -\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})|_{\mathbf{x}=\mathbf{x}_{\mathrm{MAP}}}, \tag{15}$$

is positive semidefinite due to the maximum condition. Exponentiating this yields the Gaussian approximation (up to normalization)

$$e^{\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})} \propto p(\mathbf{x}|\mathbf{r},\mathbf{v}) \approx \mathcal{N}(\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}),\mathcal{C}_x(\mathbf{r},\mathbf{v})), \tag{16}$$

where $\mathcal{N}(\mu,C)$ denotes a Gaussian density with mean μ and covariance C for the integrand of Eq. (12). [An important technical point here is that this Gaussian approximation is partially justified by the fact that the log-posterior (13) is a concave function of $\mathbf{x}$ [24,26,34] and therefore has a single global optimum, like the Gaussian (16).] Here, the posterior image covariance $\mathcal{C}_x(\mathbf{r},\mathbf{v})$ is given by the inverse of the negative Hessian matrix $H(\mathbf{r},\mathbf{v})$. (Note the dependence on both the observed responses $\mathbf{r}$ and the putative velocity $\mathbf{v}$.) The elementary Gaussian integration in Eq. (12) then yields

$$p(\mathbf{r}|\mathbf{v}) \approx \frac{e^{-\mathcal{L}(\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}),\mathbf{r},\mathbf{v})}}{\sqrt{|\mathcal{C}_x H(\mathbf{r},\mathbf{v})|}} \tag{17}$$

for the marginal likelihood, or its logarithm

$$\log p(\mathbf{r}|\mathbf{v}) \approx -\mathcal{L}(\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}),\mathbf{r},\mathbf{v}) - \tfrac{1}{2}\log|\mathcal{C}_x H(\mathbf{r},\mathbf{v})|. \tag{18}$$

The MAP itself is found from the condition $\nabla_x\mathcal{L}=0$, which in the case of exponential GLM nonlinearity $f(\cdot)=\exp(\cdot)$ yields the equation

$$\mathbf{x}_{\mathrm{MAP}}(\mathbf{n};\mathbf{r},\mathbf{v}) = \int \mathrm{d}^2\mathbf{n}'\mathcal{C}_x(\mathbf{n},\mathbf{n}')\sum_i \int \mathcal{K}_{i,\mathbf{v}}(t;\mathbf{n}')[r_i(t) - \lambda_i(t;\mathbf{x}_{\mathrm{MAP}},\mathbf{r})]\mathrm{d}t. \tag{19}$$

Note that this equation is nonlinear due to the appearance of $\mathbf{x}_{\mathrm{MAP}}$ inside the GLM nonlinearity on the right-hand side. As mentioned above, the objective function Eq. (11) is concave and can be efficiently optimized using

gradient-based optimization algorithms, such as the Newton–Raphson method. In particular, by exploiting the quasi-locality of the GLM likelihood we can implement the Newton–Raphson method such that, in cases where the image $\mathbf{x}(\mathbf{n})$ depends only on one component of $\mathbf{n}$, the MAP can be found in a computational time scaling only linearly with the spatial size of the image (see Appendix B for further elaboration on this point). Once $\mathbf{x}_{\mathrm{MAP}}$ is found, the Hessian at MAP and Eq. (17) can be calculated easily, and using Eq. (17), the approximate computation of $p(\mathbf{r}|\mathbf{v})$ is complete.

To recapitulate, in the case of an *a priori* uncertain image, given the observed spike trains $\mathbf{r}$, we numerically find $\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v})$ for a range of putative velocities $\mathbf{v}$ and using Eq. (17), we compute $p(\mathbf{r}|\mathbf{v})$, from which we may obtain $p(\mathbf{v}|\mathbf{r})$ via Eq. (8). We then take the value of velocity $\mathbf{v}_\star$ that maximizes $p(\mathbf{v}|\mathbf{r})$ as the estimate; i.e., we use the MAP estimate for the velocity.

As discussed in Section 1, our goal here was to critically examine the role of the detailed spiking structure of the GLM in constraining our estimates of the velocity. Since the spiking network model structure enters here only via the likelihood term $p(\mathbf{r}|\mathbf{v})$, we did not systematically examine the effect of strong *a priori* beliefs $p(\mathbf{v})$ on the resulting estimator (as discussed at further length, e.g., in [14]). Instead we used a simple uniform prior on velocity, which renders the MAP velocity estimate equivalent to the maximum (marginal) likelihood estimate, i.e., the value of $\mathbf{v}$ that maximizes $p(\mathbf{r}|\mathbf{v})$ given by the approximation Eq. (17) [or equivalently, its logarithm Eq. (18)]. Similarly, in the case of *a priori* known image $\mathbf{x}$ we chose the velocity $\mathbf{v}$ that maximizes the likelihood $p(\mathbf{r}|\mathbf{x},\mathbf{v})$.

2. Velocity Estimation Using the Energy Method

In order to assess the precision of our Bayesian estimates of velocity, we compared our estimates to those obtained using the correlation-based algorithm described in [4]. This algorithm closely resembles the spatiotemporal energy models for motion processing introduced by [36]. In order to understand the rationale behind this method, assume, hypothetically, that all the cells have exactly the same receptive fields up to the positioning of their centers and that they respond reliably and without noise to the stimulus. Then the RGCs' spike trains $\mathbf{r}_i$ in response to moving images would clearly be identical up to time translations. In other words, $r_i(t+n_i/v)$ would be equal for all i, where n_i is the center position of the ith cell's receptive field along the axis of motion, and v is the magnitude of $\mathbf{v}$. Thus even in the realistic, noisy situation, we expect the $\mathbf{r}_i$ for different i to have a large overlap if they are shifted in time as described, and in principle, we should be able to recover the true velocity by maximizing a smoothed version of this overlap. Inspired by this observation, an energy function is constructed as follows. First, the spike trains are convolved with a Gaussian filter $w(t)\propto\exp(-t^2/2\tau^2)$ (we chose τ to be 10 ms; see below and [4]). Let us define

$$\tilde{\mathbf{r}}_i(t) = w\star\mathbf{r}_i = \sum_\alpha e^{-(t-t_{i,\alpha})^2/2\tau^2}. \tag{20}$$

Then, the "energy" function for the entire population of cells is determined by the sum of the overlaps of the shifted and smoothed responses of all cells [37],

$$\mathcal{E}(\mathbf{v},\mathbf{r}) = \sum_{i,j}\int \tilde{r}_i\left(t+\frac{n_i}{v}\right)\tilde{r}_j\left(t+\frac{n_j}{v}\right)\mathrm{d}t = \int\left[\sum_i \tilde{r}_i\left(t+\frac{n_i}{v}\right)\right]^2\mathrm{d}t. \tag{21}$$

In order to cancel the effect of spontaneous activity of the cells, in [4] a "net motion signal" $N(\mathbf{v},\mathbf{r})$ is obtained by subtracting energy of the left-shifted spike trains from that of the right-shifted responses:

$$N(\mathbf{v},\mathbf{r}) \equiv \mathcal{E}(\mathbf{v},\mathbf{r}) - \mathcal{E}(-\mathbf{v},\mathbf{r}). \tag{22}$$

Finally, $N(\mathbf{v},\mathbf{r})$ is calculated for $\mathbf{v}$ across a range of putative velocities, and the value that maximizes the net motion signal is taken as the velocity estimate. Figure 1 illustrates the basic idea of this method for ON cells, although it should be noted that OFF cells are also included in our analysis.

3. Connection between the Bayesian and Energy-Based Methods

An interesting connection can be drawn between Bayesian velocity decoding and the method of Subsection 2.B.2 based on the energy function Eq. (21). For simplicity,

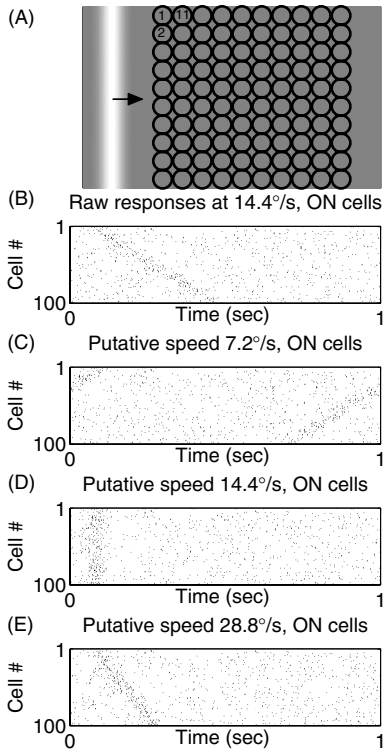

Fig. 1. Ensemble motion signals. (A) Moving bar stimulus and cell layout. Cells in the left most column are numbered 1–10 from top to bottom, cells in the second column are numbered 11–20 from top to bottom, etc. (B) Raw responses from the ON cells for a moving bar with speed 14.4°/s. Each tick represents one spike and each row represents the response of a different cell. (C)–(E) Same spike trains circularly shifted by an amount equal to the time required for a stimulus with the indicated putative speed to move from an arbitrary reference location to the receptive field center. Responses from OFF cells were also included in this procedure.

imagine that spike trains are generated not by the GLM, but rather by a simpler linear-Gaussian (LG) model. In this case, it turns out that the marginal likelihood method is closely related to the energy function method described above. Specifically, we model the output spike trains as

$$\mathbf{r}_i = b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x} + \boldsymbol{\epsilon}_i, \tag{23}$$

where the noise term is Gaussian $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \Sigma)$. In the case that the noise terms for different cells are independent, we have $p_{\mathrm{LG}}(\mathbf{r}|\mathbf{x},\mathbf{v}) = \Pi_i \mathcal{N}(b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}, \Sigma)$, though the generalization to correlated outputs is straightforward. We show in Appendix A that in a certain regime the logarithm of the LG marginal likelihood is given by [see Eqs. (A7) and (A8) and Eq. (A18)]

$$\log p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v}) = \frac{1}{2}\sum_{i,j} \int R_i\left(t + \frac{n_i}{v}\right) R_j\left(t + \frac{n_j}{v}\right) \mathrm{d}t + A(\mathbf{v}), \tag{24}$$

where $A(\mathbf{v})$ has no dependence on the observed spike trains and only a weak dependence on $\mathbf{v}$. We find empirically that the term $A(\mathbf{v})$ in Eq. (24) grows with velocity, and therefore its inclusion shifts value of the maximum likelihood estimate toward higher velocities. Conversely, its absence in the energy function Eq. (21) causes the energy method estimate to have a negative bias. See Fig. 5 for an illustration of this effect. The resemblance of the remaining term to Eq. (21) above is clear. Here, $\mathbf{R}_i$ are smoothed versions of the spike trains $\mathbf{r}_i$ (with the baseline log firing rate subtracted out) and are given, as in Eq. (20), by

$$\mathbf{R}_i = w_{\mathrm{LG}} * (\mathbf{r}_i - b_i), \tag{25}$$

where here the optimal smoothing filter w_{LG} is determined by the receptive fields k_i, the prior image correlation statistics, and the velocity [its explicit form is given in Eq. (A21) in Appendix A], as we discuss in more depth below.

Thus maximizing the marginal likelihood Eq. (24) is, to a good approximation, equivalent to maximizing the energy Eq. (21). The major difference between Eq. (21) and Eq. (24) is in the filter we apply to the spike trains: $\tilde{\mathbf{r}}_i$ has been replaced by $\mathbf{R}_i$. The key point is that $\mathbf{R}_i$ depends on the stimulus filters k_i, the velocity $\mathbf{v}$, and the image prior in an *optimal* manner, unlike the smoothing in Eq. (20). (Note that, while changes in optimal filters at differing light levels have been discussed in terms of motion estimation in fly vision [38], no account of varying light levels was taken here.) The dependence of this optimal filter as a function of v can be explained fairly intuitively, as we discuss at more length in Appendix A following Eq. (A21). We find that τ_w, the time scale of the smoothing filter w_{LG}, is dictated by three major time scales, some of which depend on the velocity v: τ_k, the width of the time window in which each RGC integrates its input; l_k/v, where l_k is the spatial width of the receptive field; and l_{corr}/v where l_{corr} is the correlation length of natural images. At low velocities, l_k/v and l_{corr}/v are large, and the smoothing time scale τ_w is also large, since in this case we gain more information about the underlying firing rates by averaging over a longer time window. At high velocities, on the other hand, τ_k dominates l_k/v and l_{corr}/v, and $\tau_w \sim \tau_k$. This setting of τ_w makes sense because although the image movie I can vary quite quickly here, the filtered input $J_i(t)$ induces a firing rate correlation time of order τ_k, and examining the responses at a temporal resolution finer than τ_k only decreases the effective signal-to-noise.

Figure 2 illustrates these effects by plotting the optimal smoothing filters w_{LG} for several different values of the velocity v. Interestingly, in the high-velocity limit, the analytically derived optimal temporal filter width τ_w is of the order of 10 ms, which was the value chosen empirically for the optimal Gaussian filter used in [4]. We recomputed the optimal empirical filter for our simulated data here by plotting the standard deviation of the velocity estimates obtained using the net motion signal [defined in Eq. (22)] against the filter width (Fig. 3). For this velocity (28.8°/s) the optimal filter is of the order of 10 ms; thus, we used a filter of width 10 ms when comparing the energy method to the Bayesian decoder.

To summarize, maximizing the likelihood marginalized over the unknown image is very closely related to maximizing the energy function introduced by [4], if we replace the GLM with the simpler linear Gaussian model. Since the actual spike train generation is much better modeled by the GLM than by the Gaussian model, we expect Bayesian velocity estimation (even with uncertain prior knowledge of the image) based on the correct GLM to be more accurate. This expectation is borne out by our simulations, though it is worth noting that the improvement is significantly smaller than when the Bayesian decoder has access to the exact image.

C. Simulations

We simulated the presentation of a bar moving across the gray background of a CRT monitor refreshing at 120 Hz.

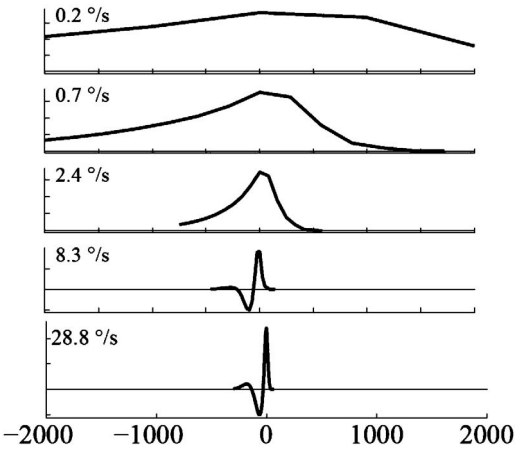

Fig. 2. Optimal linear spike train filter w_{LG} for a range of velocities from 0.2°/s to 28.8°/s (bottom) in exponential steps. The y axes are scaled in dimensionless units for clarity here. As discussed in Subsection 2.B.3, there are three time scales that determine the time scale of our filter w_{LG}. At low velocities, shown in the upper panels, the width of $w(t)$ is determined by the two scales x_k/v and x_{corr}/v and is thus quite large (since the denominator v is small). At the higher velocities shown in the lower panels, the optimal filter width is dominated by the time scale of the receptive field τ_k, and is of the order of τ_k, which is ~10–20 ms. For even higher velocities the shape of this filter remains essentially the same as in the bottom panel.

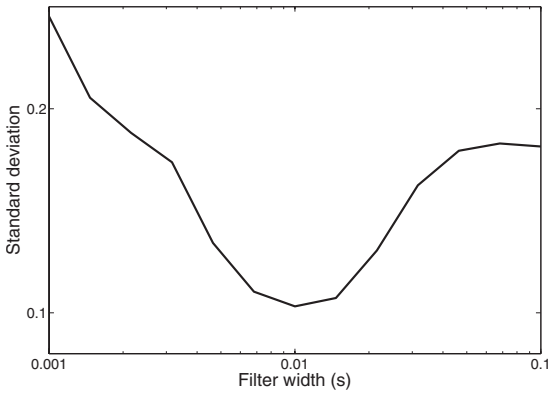

Fig. 3. Effect of filter width τ_w on the standard deviation of velocity estimates [obtained using the signal defined in Eq. (22)] across 100 presentations of a black bar moving at a speed of 28.8°/s across a gray background. Note that a filter width of $\tau_w \approx 10$ ms is optimal, in agreement with the findings of [4] and with the width of the optimal filter shown in Fig. 2.

The spatial profile of the bar in the direction of motion was a Gaussian function with a standard deviation (SD) of 96 μm. The visual field was represented by a grid of 100×100 pixels covering the receptive fields of two layers of cells each arranged in a uniform 10×10 grid. One layer consisted of ON cells, while the other represented OFF cells. The pixel resolution used was 10 times that used in [7] resulting in a pixel size of 12 μm. The bar moved across the visual field in discrete steps of $\mathbf{v}$ pixels/refresh, although $\mathbf{v}$ was not restricted to integer values. On each trial, the bar traversed the entire visual field once at a constant velocity. (Therefore, low-velocity trials lasted longer than high-velocity trials; this will affect some of our analyses below.) Stimulus dimensions and speeds were converted to degrees/second using the approximation 200 μm/° [39] with a pixel size of 12×12 μm. This meant that, with a refresh rate of 120 Hz, a speed of 1 pixel/refresh corresponded to a speed of 7.2°/s.

Then, to investigate the fidelity with which speed was encoded by our model, we ran simulations using a variety of stimulus parameter settings. Specifically, we conducted 100 trials at each of 48 stimulus conditions. These 48 conditions were made up of eight speeds (10.8, 14.4, 21.6, 28.8, 36.0, 43.2, 50.4 and 57.6°/s) by six luminance levels (0, 0.125, 0.25, 0.75, 0.875 and 1 on a grayscale level where 0 is black and 1 is white, and the background level was set at 0.5). We also refer to the six different luminance levels in terms of the contrast of the bar with respect to the background. More precisely we define the contrast as $(I_{\text{bar}}-I_{\text{background}})/I_{\text{background}}$, where I_{bar} and $I_{\text{background}}$ denote the bar and the background luminance, respectively. These six luminance levels thus become contrast levels −1, −0.75, −0.5, 0.5, 0.75, and 1.

For each of these trials, we obtained a set of spike trains $\mathbf{r}$. From these spike trains, it was possible to estimate the speed of the stimulus used as being one of a number of putative speeds. The putative speeds tested in our simulations ranged from 7.2 to 108°/s in steps of 0.36°/s. Thus, we could compare speed estimates across stimulus conditions by examining the SD of estimates across the 100 trials performed for each condition. As in [4], we focused on the fractional SD (SD of velocity estimate divided by the true stimulus speed) of estimates to assess the fidelity of retinal speed signals, as any systematic bias in speed estimate can in principle be compensated for by downstream processing. However, we will also present the dependence of the estimate bias on stimulus condition. As will be seen, the fractional bias and the fractional SD are roughly of the same order and thus both contribute to the total root-mean-square fractional error of the velocity estimate. The latter is given by the square root of the sum of the squared fractional bias and squared fractional SD. It should be noted that other contrast levels between −0.5 and 0.5 were also tested but are not presented, as for some combinations of decoder and speed, the velocity estimation performance at these low contrasts was not significantly above chance.

As outlined above, we used three different decoding methods to estimate the stimulus velocity from the simulated spike train ensembles; we compared Bayesian velocity decoding with (optimal decoder) and without (marginal decoder) complete prior information about the image with velocity estimation using the energy method. In particular, in Subsection 3.A.4 we discuss the effect of prior image uncertainty on the performance of the Bayesian decoder in more detail. In order to parametrically vary the prior information available to the decoder, in the simulations used in that section, the image was flashed a number of times to the cells while it was held fixed, and the image prior $p(I)$ was updated according to the observed spike train data elicited by the flashes (no preview flashes were used in the simulations discussed in other sections). See Fig. 6(b) below for an illustration of this procedure. Short flashes were used instead of a continuous uninterrupted presentation, because in the latter case, the cells rapidly filter out the fixed image contrast, and thus after a brief interval (~20–30 ms), the spike trains cease to carry extra information about the image. The more times the image is flashed, the smaller the decoder's uncertainty $\mathcal{C}_x$ when the image starts moving. This allows the decoder to better estimate the velocity when it finally sees the same image in motion.

3. RESULTS

A. Comparison of the Different Velocity Decoders

In this section we compare the performance of the energy model with the optimal and marginal decoders, as described in Section 2. Figure 4(a) plots the velocity posterior $p(\mathbf{v}|\mathbf{r},\mathbf{x})$ for the case of an *a priori* known image (the moving bar described above) given a specific observed population spike train $\mathbf{r}$ in response to the moving bar stimulus as a function of putative stimulus speed $\mathbf{v}$. Here, the true stimulus speed was 36.0°/s. Figure 4(b) shows the log of the marginal likelihood in the case where the image is not completely known, and Fig. 4(c) shows the value of the net motion signal N again as a function of putative speed and for the same stimulus. All three decoders successfully estimated the speed in the trial shown; however, it is clear from the figure that the net motion signal is much less sharply peaked around the stimulus speed than for the Bayesian decoders. The consequences of these findings are reflected in the lower panels of Fig. 4, which show that the distribution of speed estimates

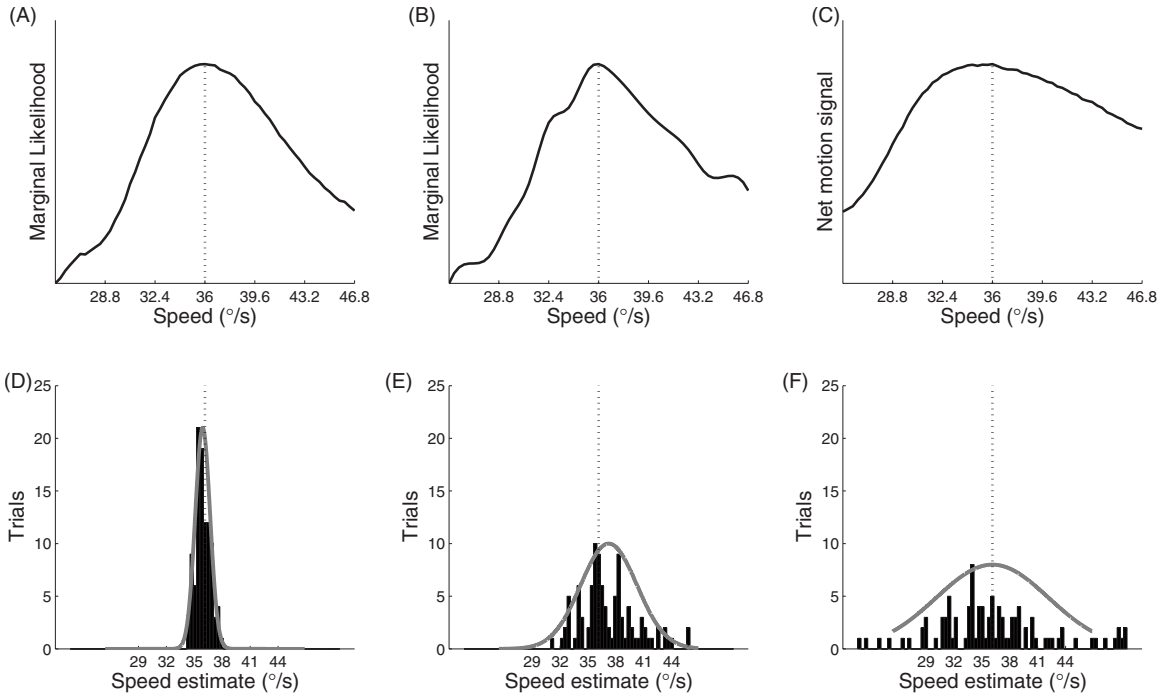

Fig. 4. Optimal decoder leads to the most precise velocity estimates, while the marginal decoder outperforms the energy-based "net motion signal" method, in terms of precision. (A) Posterior for the optimal decoder. (B) Log marginal likelihood for the marginal decoder. (C) Net motion signal N as a function of putative stimulus speed $\mathbf{v}$ for spike trains generated using a stimulus with speed 36.0°/s and contrast −0.5 for a trial where all methods successfully estimate the stimulus speed. It can be seen that the nonmarginal likelihood (A) is more sharply peaked around the stimulus speed than the marginal likelihood (B) and the net motion signal (C). Distribution of speed estimates across 100 presentations of a bar moving at a speed of 36.0°/s using (D) the optimal posterior probability, (E) the marginal likelihood, and (F) the net motion signal. Also plotted are Gaussian fits to the distributions with mean±SD of 35.76±0.81 for the optimal decoder, 37.1±3.05 for the marginal decoder, and 36±6.09 for the net motion signal.

across 100 presentations of a bar of contrast −0.5 moving at a speed of 36.0°/s is most precise using the optimal decoder [Fig. 4(d)], rather than either the marginal decoder (E) or the energy method (F). Also plotted are Gaussian fits to the distributions with a mean ±SD of 35.76±0.81°/s for the optimal decoder, 37.1±3.05°/s for the marginal decoder and 36±6.09°/s for the net motion signal. The fractional SD averaged across all conditions simulated in this study was 1.6% of the stimulus speed for the optimal decoder, 6.4% for the marginal decoder, and 10% of the stimulus speed for the energy method. Since the estimators are not unbiased, their root-mean-square error is larger than their SD, as the error receives a contribution from the bias as well. The root-mean-square fractional errors averaged across all stimulus conditions were 2%, 6.9%, and 11%, for the optimal decoder, marginal decoder and the energy method, respectively. Because the velocity estimation based on the energy method does not make use of the image profile at any stage, these results are as expected with the performance of the marginal decoder being intermediate between that of the optimal decoder and the energy method.

1. Accuracy as a Function of Stimulus Speed

Because in our simulations the moving bar stimulus makes only one pass over the visual field, more time is spent traversing the field and more spike train information is obtained for slower moving stimuli. Figure 5(a) illustrates the fractional SD of 100 speed estimates for both of the Bayesian methods and the energy method, at each of the eight stimulus speeds, averaged across the six contrast levels. As expected, performance declines with increasing speed for all three methods. The Bayesian decoders provide more precise estimates than the energy method at all speeds. Again, the advantage of the Bayesian decoder over the energy method is partly lost when its prior information about the image is uncertain.

2. Accuracy as a Function of Stimulus Contrast

Lowering the contrast of the moving bar causes a reduction in the number of stimulus-related spikes generated by the GLM model, according to Eqs. (2) and (3). As with increasing stimulus speed, this obviously results in a reduction in stimulus-related information with which to estimate the stimulus speed. (Note that the model of [7] lacks explicit luminance-or contrast-gain control effects; thus, these results should be interpreted in terms of local modifications around a fixed luminance pedestal that are sufficiently small to avoid engaging classical luminance gain-control mechanisms.) To examine this relationship, we averaged the SD of the 100 speed estimates at each of the six contrast levels across the eight stimulus speeds. The results are shown in Fig. 5(b) and illustrate the expected increase in performance with increasing stimulus contrast. Again, the Bayesian decoders clearly outperform the energy method at all levels.

3. Effect of Contrast and Speed on Mean Speed Estimate

While we were primarily concerned with the precision of speed estimates in the current study, a number of well-

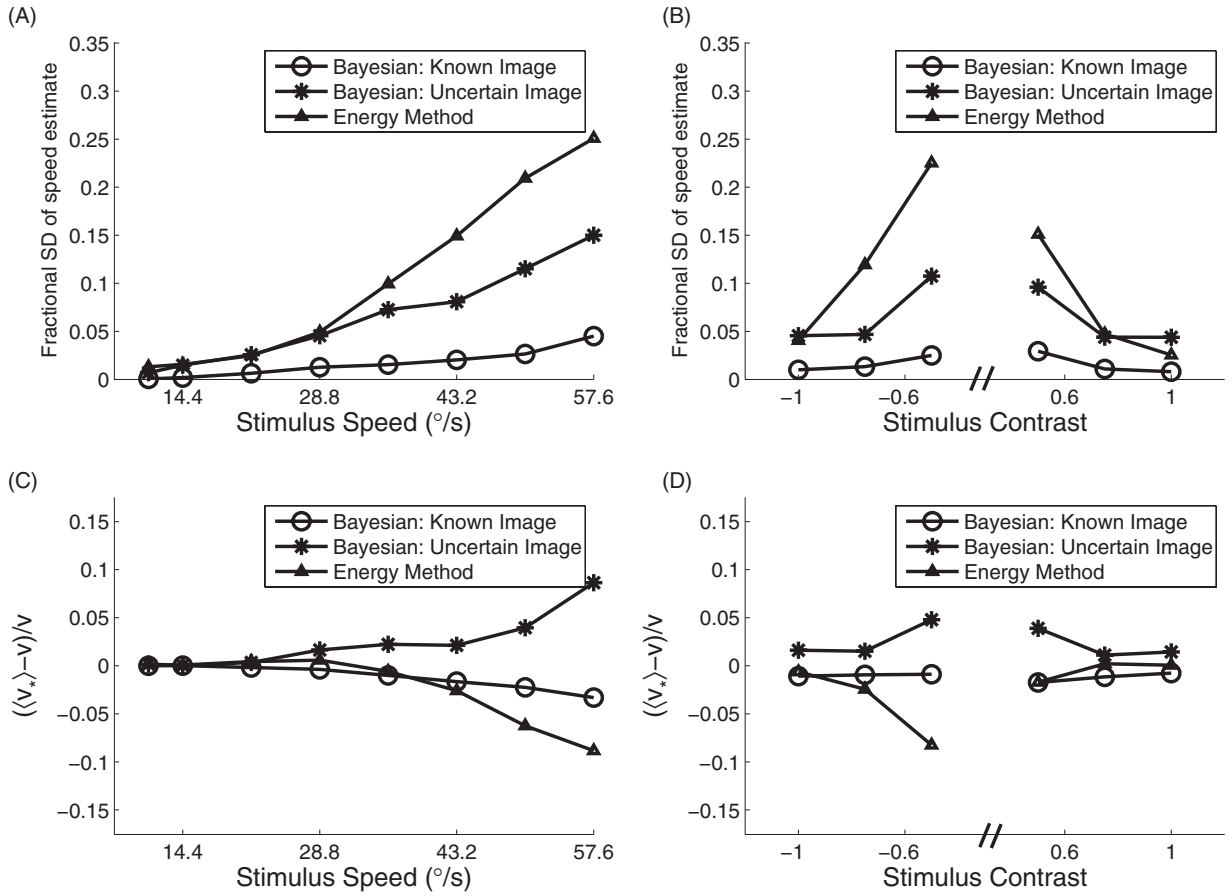

Fig. 5. Fractional standard deviation of speed estimates versus (A) stimulus speed and (B) stimulus contrast for the Bayesian decoder with full image information (optimal decoder), the Bayesian decoder with incomplete image information (marginal decoder), and the energy method. (C), (D) plot the difference between the mean estimated speed $\langle \mathbf{v}_\star \rangle$ and the true stimulus speed $\mathbf{v}$ normalized by $\mathbf{v}$ against the true stimulus speed and stimulus contrast, respectively. Note that the Bayesian decoder provides more precise estimates than the energy method at all levels, with performance improving with prior image information. Furthermore, it should be noted that for contrasts lower than ±0.5, particularly at high speeds, the dearth of information in the spike train ensemble resulted in the estimate from an inordinate number of trials being either the highest or lowest input putative speed. Accordingly, performance for s in this range were not calculated and not plotted.

researched visual phenomena concerning the relationship between the mean visual speed perceived, i.e., the bias, and the properties of the visual stimulus prompted us to investigate this in our simulations. The first phenomenon of interest was that in which humans tend to choose the slowest motion that explains the incoming information [40], i.e., we have a bias toward slower speeds. As can be seen in Fig. 5(c), the energy method is biased toward lower velocity estimates at higher stimulus speeds. The optimal decoder shows a very slight tendency in this direction also. On the other hand, the marginal decoder has a positive bias toward higher velocities. The second phenomenon of interest was that in which stimuli with low contrast are typically perceived as moving slower than those with high contrast [41,42]. Figure 5(d) plots the fractional bias of the speed estimate, i.e., the difference between the true stimulus speed $\mathbf{v}$ and the mean estimated speed $\langle \mathbf{v}_\star \rangle$ normalized by $\mathbf{v}$, versus the stimulus contrast for both the Bayesian decoder and the energy method against the stimulus contrast, averaged across all speeds tested in our simulations. There appears to be a slight trend toward greater bias at low contrast, although it should be noted that this is due to a strong bias at low negative contrast, while at low positive contrast, the bias is close to zero. The fact that the fractional SD of the speed estimate at this low negative contrast value is so large makes it difficult to say anything definitive about a relationship between stimulus contrast and speed estimate bias. It is important to remember that a uniform prior on velocity was used in this study when considering the lack of any clear effect.

4. Effect of Prior Image Information

Here we discuss the effect of preview flashes of the fixed image on the velocity decoding performance of the marginal decoder. As mentioned above, the more times the image is flashed or "shown" to the cells, the less will be the decoder's uncertainty about it and the better the velocity estimate made by the decoder when it finally sees the same image in motion. This effect is shown in Fig. 6 where panel (A) shows the decrease in the relative error of the velocity estimate as the number of flashes increases. For a large number of flashes the error asymptotically reaches the level for the fully known image

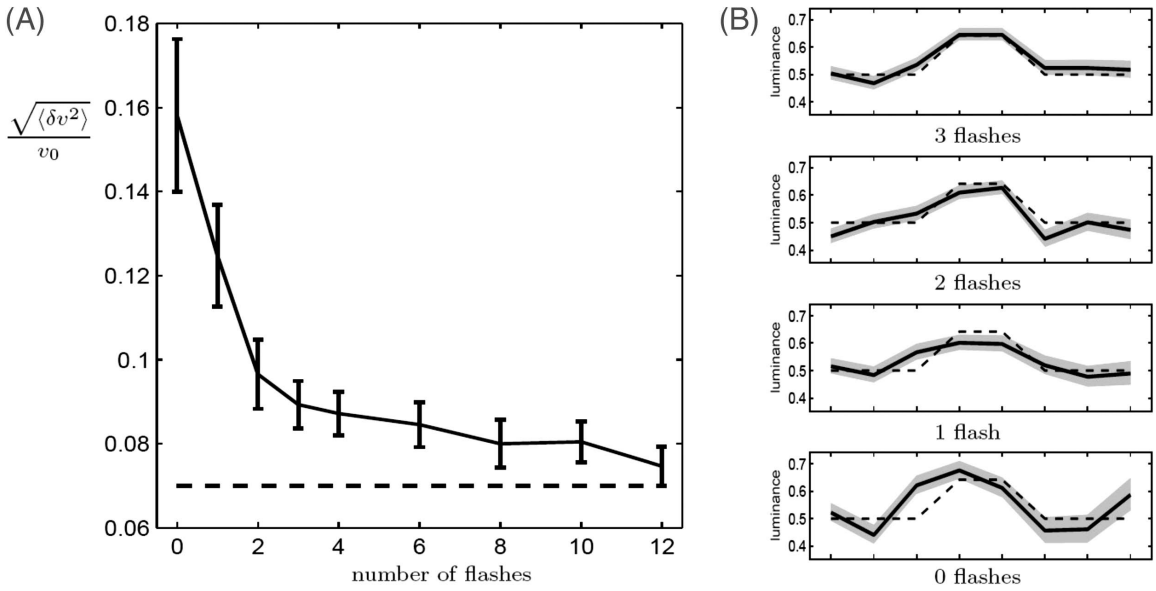

Fig. 6. Effect of decreasing image uncertainty on accuracy of Bayesian velocity estimation. See Subsection 2.C for a detailed description of this simulation. (A) The solid curve with error bars shows the drop in the fractional rms error of the velocity estimate for an *a priori* unknown image as the number of preview flashes increases. The dashed curve is the fractional error for the case of an *a priori* known image. The true velocity was 28.8°/sec and the bar contrast 0.6. (B) The plots show the maximum *a posteriori* estimate of the image luminance profile (solid curve) in four trials with different numbers of preview flashes (indicated below each plot). The gray areas indicate the marginal uncertainty of the estimated luminance, and the dashed curve shows the actual image profile.

(shown by dashed curves). Panel (B) shows the convergence of the estimated luminance profile $\mathbf{x}_{\mathrm{MAP}}$ to that of the actual bar image as the number of preview flashes increases. Figure 6 shows this effect for a particular stimulus velocity and contrast, but we note that the effect is qualitatively the same for all other values of these stimulus parameters; as the number of flashes increases beyond a few, the accuracy of the marginal decoder's estimate approaches that of the optimal decoder's.

As seen here and above, the efficiency of the GLM-based Bayesian decoder can be significantly deteriorated when the prior information about the image is too incomplete. As we showed in Subsection 2.B.3, Bayesian decoding with uncertain prior image information is, except for the replacement of the GLM with the LG model, closely related to the energy model. Indeed, in our simulations, the disparity between the performances of the energy model and those of the GLM-based Bayesian decoder was largely lost when the latter decoder's prior knowledge of the image became too uncertain.

B. Effects of Manipulating Model Parameters

1. Importance of Correlation Between Cells

In order to investigate the importance of correlated activity between cells, we wished to remove the interaction between neighboring spike trains without reducing the overall spiking rate. We used a straightforward trial-shuffling approach: we generated 200 individual spike trains, one for each cell, using 200 distinct presentations of the stimulus to the full model. We then constructed a single trial surrogate population spike train by serially assigning each independent spike train recorded on simulated trial i as the observed spike train in cell i. We repeated this 100 times to obtain spike ensembles representing 100 trials for each of the 48 conditions mentioned above (i.e., eight different speeds and six different contrast levels). This allowed us to determine the fractional standard deviation of the speed estimate for each of the 48 different stimulus conditions. It should be noted that this (somewhat involved) procedure was carried out in preference to simply removing the coupling between cells, as that would have resulted in a different average number of population wide spikes compared with the output from the full model, which would have had a confounding effect on the results.

The results are shown in Figs. 7(a)–7(c) for the optimal decoder, marginal decoder and the energy method, respectively, and are plotted versus the fractional standard deviation of the speed estimate for the same 48 conditions using the spike train ensembles obtained directly from the model. The diagonal lines indicate equality between the fractional SD of the speed estimates obtained using the shuffled responses and that obtained directly from the model. Somewhat surprisingly, given the significant correlations in this data (cf. Fig. 2 in [7]), this trial-shuffling procedure did not significantly hurt the performance of any of the three velocity estimators. For the marginal decoder, there is a noticeable reduction in performance for those stimulus conditions with more precise speed estimates, while for conditions with higher fractional SD, most points lie just below the line. However, for the other two decoders, if anything, there is a slight bias in Fig. 7(a) and 7(c), with data points tending to lie a bit below the identity line in both plots, indicating that the shuffling procedure happened to lead to velocity estimates with slightly reduced variability. These results are consistent with the conclusions of [2], that treating retinal ganglion cells as independent encoders leads only to a minor loss of information.

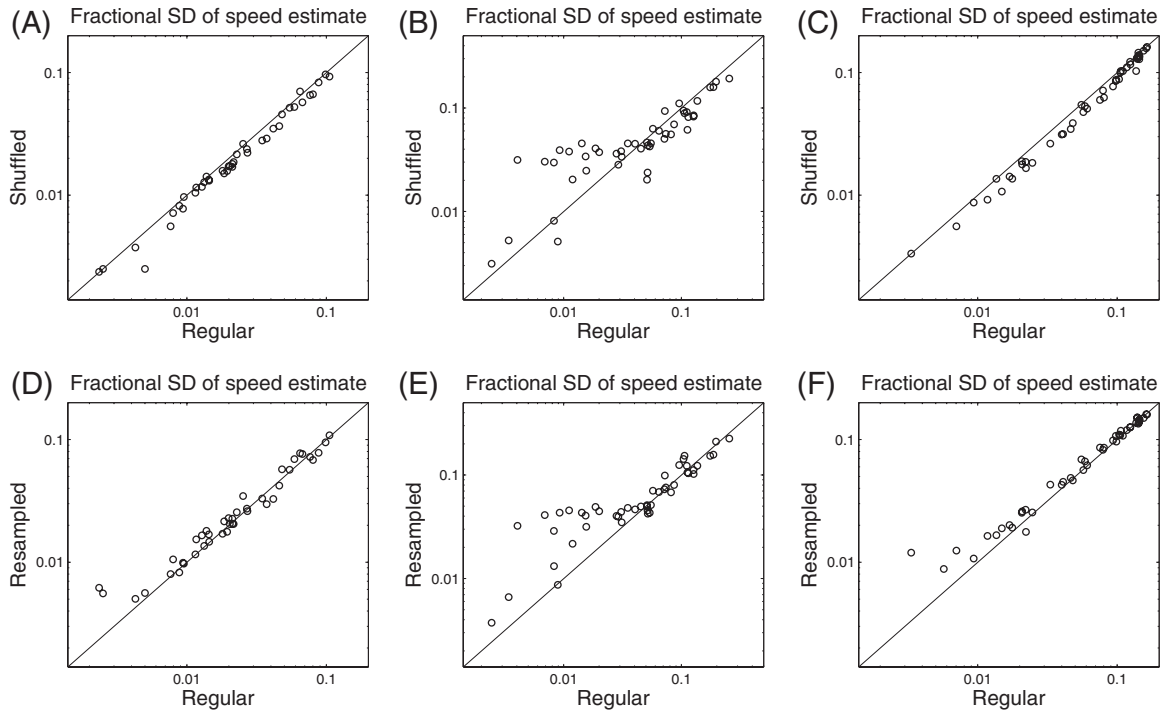

Fig. 7. Effect of correlated activity and spike timing structure on speed estimates. Fractional SD of speed estimates using shuffled responses plotted as a function of that obtained using regular simulated data for (A) the Bayesian decoder with full image information, (B) the Bayesian decoder with incomplete image information, and (C) the energy method. Fractional SD of speed estimate using resampled spike trains plotted as a function of that obtained using regular simulated data for (D) the Bayesian decoder with full image information, (E) the Bayesian decoder with incomplete image information, and (F) the energy method. Diagonal lines indicate equality. Each circle represents a different one of the 48 speed-by-contrast stimulus conditions. Note that the performance of the decoders is relatively unaffected by these rather drastic manipulations of spike timing.

2. Timing Structure of Spike Trains

The question of whether cell spiking activity can be accurately modeled as a simple Poisson process with a time-varying rate or whether the intrinsic temporal structure of retinal spike trains plays an important role in communication has a long history in systems neuroscience. Simulations with the retinal ganglion cell model used in this study have demonstrated that preserving the spike history and cross-coupling effects can increase stimulus decoding performance by up to 20% [7]. We wished to examine the effect of removing the specific timing information of the individual spike trains. This was carried out using the method of [4]. Specifically, we generated a spike train for each cell for 100 trials of the moving bar stimulus. We then randomly selected spike times for each cell, with replacement, from that cell's spike distribution [its peri-stimulus time histogram (PSTH)], such that the number of spikes in each resampled spike train was equal to the average number of spikes in the corresponding original spike trains. This results in a spike train for each cell where spikes occur according to the marginal mean firing rate only, with no consideration given to spike history effects such as action potential refractoriness. Note that this process is even more disruptive of spike timing information than the shuffling procedure described in the previous subsection, since now we are destroying spike train structure both between and within cells. Again, this convoluted process was carried out in preference to simply removing the spike history filters h_{ij} from the model before generating the spike trains, as removal of those filters would have resulted in a greater number of total spikes and would thus have resulted in a misleadingly good speed estimation performance. This process of generating a spike train ensemble through resampling was carried out for each of the 48 stimulus conditions mentioned above.

The results are shown in Figs. 7(d)–7(f) for the optimal decoder, marginal decoder and the energy method, respectively, and are plotted versus the fractional standard deviation of the speed estimate for the same 48 conditions using spike train ensembles obtained directly from the model. Once again, the effects of this spike timing disruption on the performance of the velocity estimators was fairly minimal, with the resampled spike trains appearing to give a marginally worse performance as indicated by the preponderance of data points slightly above the identity line. Again for the marginal decoder the decreased performace is more pronounced for stimulus conditions with better performance.

3. Parameters of Cell Population

In the simulations above, two simple assumptions were made about the parameters of the cell population. First, the cells were arranged in an oversimplistic grid as in Fig. 8(a). And second, all ON cells were given a baseline log firing rate [b_i in Eq. (2)] of 2 spikes/s and all OFF cells a baseline log firing rate of 3 spikes/s, corresponding to the mean values obtained when fitting the model [7]. In order to examine a somewhat more biologically realistic case we jittered the center location of the cells as in Fig. 8(b) and

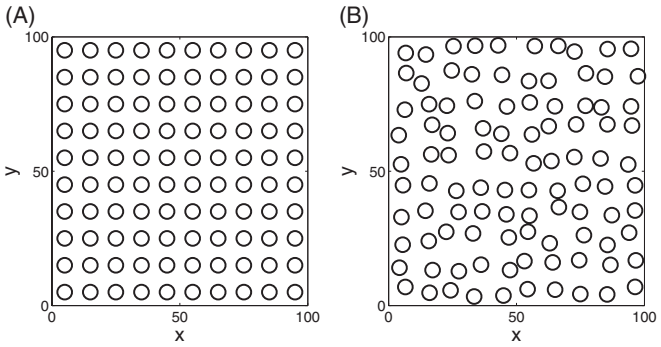

Fig. 8. (A) Simple rectangular grid cell arrangement. (B) Jittered cell arrangement.

randomly selected the baseline log firing rates of the ON and OFF cells from uniform distributions on intervals 1 to 3 spikes/s and 1.5 to 4.5 spikes/s, respectively.

Figure 9 illustrates the speed estimates over 100 trials for a stimulus with speed of 28.8°/s and contrast of −1 using the regular cell arrangement and uniform baseline log firing rates (left column) versus the jittered cell arrangement and random baseline log firing rates (middle column) for all three methods. The performances of the optimal decoder, marginal decoder and energy method are shown in the top, middle, and bottom row, respectively. No significant difference in performance between the regular and jittered arrangements is apparent for any method.

While randomly jittering the baseline log firing rates around the mean caused no obvious change in estimation accuracy, this does not allow us to comment on the pos-

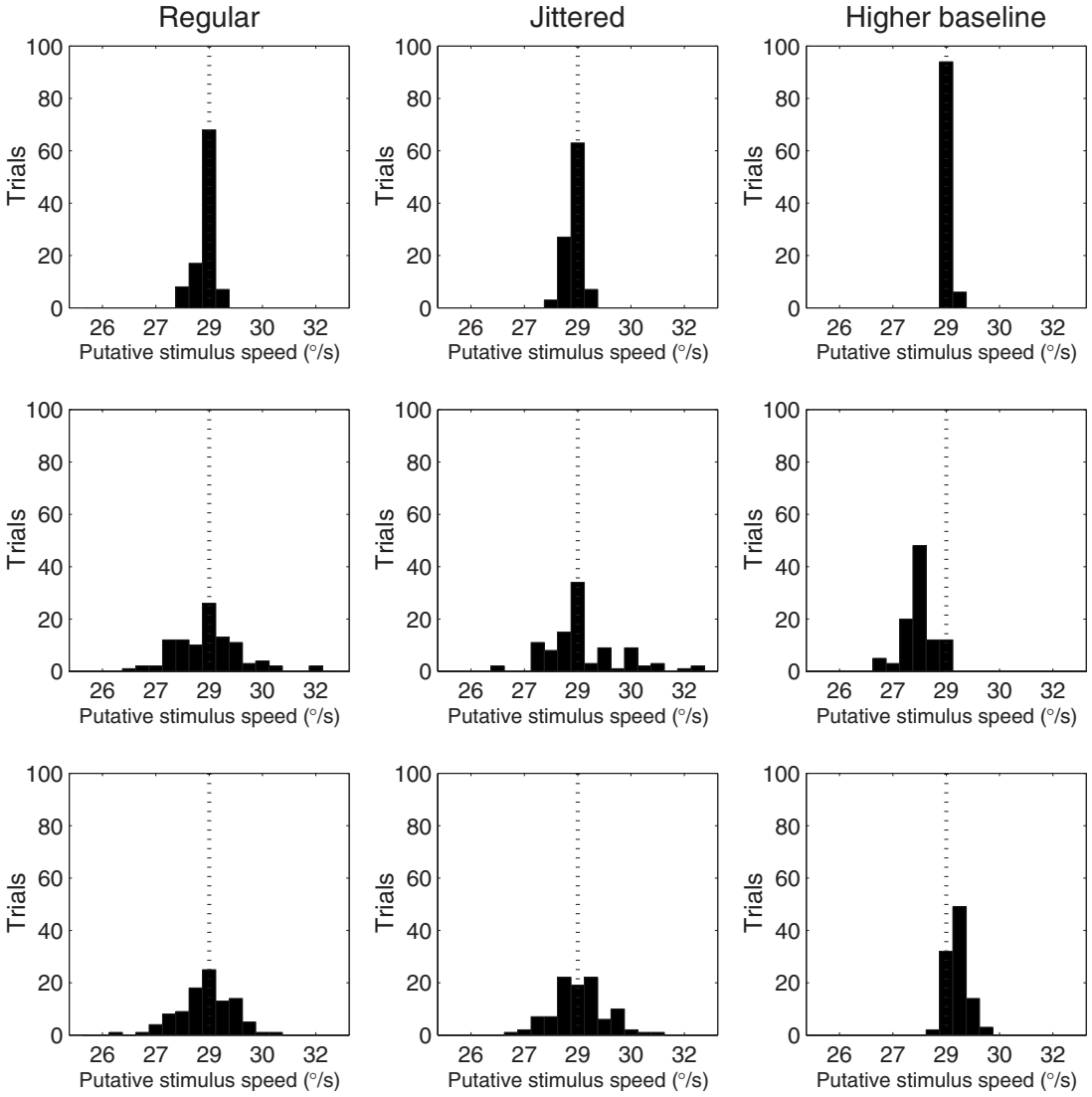

Fig. 9. Histograms illustrating the velocity estimates over 100 trials for a stimulus with velocity 28.8°/s and contrast of −1 using the regular cell arrangement and uniform baseline log firing rates (left column) and the jittered cell arrangement and random baseline log firing rates (middle column). The top row represents the performance of the optimal Bayesian decoder, the middle row represents that of the marginal decoder, and the bottom row presents that of the energy method. Similar performance was obtained with both the rectangular-grid and randomized spatial layouts for all three methods. The right column illustrates the improved estimation performance obtained for all three methods by doubling the baseline log firing rates from 2 and 3 spikes/s to 4 and 6 spikes/s for the ON and OFF cells respectively.

sible effects of changes in the *mean* baseline log firing rate. To assess this, we also carried out 100 simulations using a stimulus with speed of 28.8°/s and a contrast of −1, where the cells were arranged in the original simple grid and the ON and OFF cells were given baseline log firing rates of 4 and 6 spikes/s, respectively. The right column of Fig. 9 illustrates the significantly improved estimation performance obtained by inflating the baseline log firing rates compared to the fitted values used throughout the rest of this study for all three methods.

4. DISCUSSION

The model of [7] employed stochastic checkerboard stimuli in order to accurately capture both the stimulus dependence and detailed spatio temporal correlation structure of responses from a population of retinal ganglion cells. In this study, we have examined responses from this model to a more behaviorally relevant coherent velocity stimulus. Specifically, we have used these responses to assess how faithfully speed is encoded in a population of neurons using an optimal Bayesian decoder, with complete knowledge of the stimulus image. We have also shown how to compute the Bayesian velocity estimate in the case where we have only a limited amount of information about the stimulus image, and how the Bayesian estimate in this case is closely related to a biologically plausible motion-energy-based method [36,43].

A connection between Bayesian velocity estimation and the energy method of [36] has been noted before [12,15]. In that work, a Bayesian model of local motion information was described. It was shown that this model could be represented using a number of mathematical "building blocks" that qualitatively resembled direction-selective complex cells. Given that models of those cells have been based on the energy method of [36], a link was drawn between the two methods. Furthermore, previous work has sought to optimally estimate instantaneous motion from spike train ensembles in the fly [38,44]. However, to the best of our knowledge, in the case of non local estimation of rigid motion, the mathematical connection revealed here between the energy method and the Bayesian method based on the marginal image likelihood in the LG case has not been previously described.

In terms of biological plausibility, it is unlikely that the brain performs optimal Bayesian inference with full knowledge of the image in order to estimate velocity. This is supported by a recent study, which [8] employed the energy method (Subsection 2.B.2) to examine the efficiency of the code from a population of primate RGCs. They did this by comparing the estimate of the velocity of a stimulus using the spiking activity in the cell population with psychophysical estimates made by human observers. While the energy model consistently outperformed the human observers, it was shown that at very brief presentation times, i.e., <100 ms the difference in estimation performance between the energy method and the human behavior was much smaller than at longer presentation times. This suggests that readout of the retinal population code can be extremely efficient when exposure to the moving stimulus is very brief, but less efficient over long trials when storing information over a long time is required by the optimal Bayesian decoder. In this study, having used longer presentation times (125–675 ms), and given that the optimal Bayesian decoder significantly outperforms even the energy method, it seems clear that human observers do not decode using a known image in this task. Instead, given the relationship presented here between the marginal decoder and the energy method, it appears that a strategy equivalent to marginalization over the uncertain image seems to be more consistent with the available data.

A couple of factors in the relationship between the marginal decoder and the energy method are worthy of further discussion. First, the optimal filter width for velocity estimation from cell population responses when using the energy method on real data was reported to be of the order of 10 ms [4]. This implies that the elementary motion signal was conveyed with a time resolution comparable to the interspike interval of RGCs. A similar filter width was empirically shown using our simulated data (Fig. 3), which was not unexpected given the large amount of variance captured by the model in peristimulus time histograms in response to novel stimuli [7]. Of more interest, however, the optimal filter derived analytically for our (LG) marginal decoder is also shown to be of similar width, at least in the case where stimulus velocities are above about 5°/s (Fig. 2). This lends further weight to the biological plausibility of the marginal decoder. The notion that optimal filters based on stimulus filters, natural image prior, and velocity could have a biological instantiation seems reasonable. Second, in this study, we have assessed the performance of the energy method relative to the marginal decoder where the spike trains were generated not by a simple LG model, but by a GLM model. Because of the resulting improvement in spike train modeling we saw a significant improvement in velocity estimation for the marginal decoder relative to the energy method. Obviously the optimal decoder outperformed both other methods given the extra image information with which it was furnished.

In terms of performance specifics, the optimal Bayesian decoder achieved an average relative precision of 2% across all 48 stimulus conditions, with the marginal decoder achieving 6.4% and the energy method realizing only 10% relative precision. It is interesting to compare the estimation performance using our model to that obtained using similar stimuli with real cells in [4]. The authors of that study reported that the ensemble activity of around 100 RGCs signaled speed with a precision of the order of 1%. The precision of 10% obtained using the same decoder on our model output spike trains is higher than that result. One likely reason for this is that our stimulus range included much lower contrast stimuli. If we restrict our precision estimate to those conditions that most closely resemble those used by [4], i.e., speeds of 10.8, 14.4, 28.8, and 57.6°/s and contrast levels of −1 and 1, we obtain a value of 2.8% using the energy method which is of the same order as their result.

We examined the precision of our speed estimates as a function of both stimulus speed and stimulus contrast. As expected, decoding performance improves with increasing contrast and with decreasing speed (Fig. 5). Figure 5(a) illustrates that our model approximately followed a

Weber–Fechner law with visual speed discrimination being roughly proportional to speed [45]. As discussed in Subsection 3.A.1, the faster the moving bar traverses the retina, the less time spent stimulating the cells, and the smaller the total number of spikes we have with which to decode the stimulus speed. Similarly, the precision of the speed estimate improves with increasing absolute contrast, which increases the effective signal-to-noise of the retinal output [see Fig. 5(b)]. The nonlinear function $f(\cdot)$ used in Eq. (2) for this study was chosen to be $\exp(\cdot)$. Given that in determining the firing intensities $\lambda_i(t)$, this function operates on the stimulus input (as well as the baseline firing rates and spike history and cross-coupling effects), any increase in stimulus contrast would be expected to have a strong impact on the stimulus-related firing rates; similar conclusions may be drawn from an analysis of the Fisher information in this model [26].

As mentioned earlier, Bayesian modeling has been employed in a number of studies investigating how visual speed perception is affected by properties of the visual stimulus. In [16] an optimal Bayesian observer model was used to examine human psychophysical data in terms of stimulus noise characteristics and prior expectations. They reported that the perception that low-contrast stimuli move more slowly than high-contrast stimuli was well modeled by an ideal Bayesian observer. This was because the broader likelihood (based on psychophysical measurements), when multiplied by a prior favoring low speeds [46], resulted in a larger shift toward zero than multiplication by a narrower likelihood. In the present study, a uniform prior was used for the speed of the moving bar. Thus, we would not expect a widening of the likelihood distribution by lowering the stimulus contrast to shift the location of the posterior probability distribution. As such, we would not expect any relationship between stimulus contrast and the mean (or median) of the speed estimate distribution. This appeared to be the case, with no straightforward relationship seen to exist between speed estimate bias and contrast [Fig. 5(d)]. There did appear to be a very slight trend toward greater bias to low speeds at low contrasts for the energy method, but given the much higher variance in the speed estimate at this contrast [Fig. 5(b)], we are disinclined to draw any deeper conclusions from these results.

In terms of a relationship between speed estimate bias and stimulus speed, however, our results indicate a clear trend. Specifically, there appears to be a systematic bias in speed estimation tending to underestimate speed at high stimulus velocities for both the energy method and the Bayesian decoder with known image, while tending to overestimate speed at the same high stimulus velocities for the Bayesian decoder with uncertain image [Fig. 5(c)]. This can be explained by the well-known fact that likelihood-based estimators can display bias in low-information settings (as the high-speed setting is here, since effectively less time is available to observe spiking data during the stimulus presentation). In the low-speed, high-information setting, the bias of the likelihood-based estimator is negligible, as expected. The discrepancy between the biases of the marginal decoder and the energy-based estimate is clarified by the connection between these two methods as described in Subsection 2.B.3 and Appendix A. Specifically, see the discussion after Eqs. (24) and (25) of Subsection 2.B.3 and Eqs. (A7) and (A8) of Appendix A.

In [7] it was found that, when comparing the full RGC model with an uncoupled version (retaining spike history effects), Bayesian stimulus decoding recovered 20% more information (about the spatiotemporal light intensity profile) using pseudorandom stimuli. The authors also noted that additionally ignoring spike history effects further reduced the recovered information by 6%. Thus, we wished to examine the importance of correlations between cells and of the intrinsic timing structure of the spike trains for speed estimation precision. We followed the procedure employed in [4] and, as in that study, it appeared that, for most stimulus conditions, the shuffled, uncorrelated spike trains surprisingly resulted in a weak improvement in estimation precision. The one exception to this was in the case of the marginal decoder, where for high-information stimulus conditions (i.e., low speeds and high contrast), the shuffled spike trains resulted in a significant reduction in performance. We also replicated their test of how precise spike timing might affect speed estimation precision [4]. Again, as in their study, we found only a very slight reduction in performance for the optimal Bayesian decoder and the energy method. Similar to the reshuffled case, the deterioration in the performance of the marginal decoder was more pronounced. Given that we have completely abolished the intraneuronal and interneuronal nonstimulus-driven correlation structure here, these small decreases in performance indicate that velocity decoding does not depend strongly on the fine spike train structure—at least in the very simple case of a moving bar. It should be noted that for the results plotted in Fig. 7, all spike train ensembles were decoded using the full model. That is, coupling filters and spike history effects were assumed and accounted for when calculating λ_i in the decoding step. Given that coupling effects were removed by our shuffling procedure and that both coupling effects and spike history effects were removed by our resampling procedure, it is possible that decoding the spike trains with an appropriately reduced model might provide more accurate speed estimation for these manipulated spike train ensembles. To that end, we used a model without coupling filters to decode the speed of the shuffled spike train ensembles and a model with all h_{ij} set to zero to decode the speed of the resampled spike train ensembles. It is interesting to note that incorporating this knowledge about the presence or absence of cell coupling and spike history effects into the decoding made no qualitative difference to the accuracy of the estimated velocity (not shown).

5. CONCLUSION

Optimal Bayesian decoding with full image information has been shown to outperform a "motion energy" method that uses no prior image information. A method for performing Bayesian decoding without full image information has been described and has demonstrated performance intermediate between that of the optimal decoder and the energy method. All of these methods appear to outperform human psychophysical performance [8], par-

ticularly in experiments in which the motion stimulus was visible for an extended period of time. A mathematical description of the connection between the Bayesian decoder with less than full image information and the energy method indicates that, in addition to the extra information about the image used by the Bayesian estimator, information about the network's spatiotemporal stimulus filtering properties also plays an important role in optimal velocity estimation. The results of a number of simulations indicate a good correspondence between the speed encoding performance of the model and that of a population of real RGCs. This work thus provides a rigorous framework with which to explore the factors limiting the estimation of velocity in vision. Future work will seek to utilize these methods to investigate motion decoding using more complex stimuli moving in nontranslational ways, perhaps incorporating real-world issues such as occlusions and accelerations. Also, we aim to employ the methods to investigate real data.

APPENDIX A: MARGINAL LIKELIHOOD IN THE LINEAR GAUSSIAN MODEL

In this appendix we show that the logarithm of the marginal likelihood $p(\mathbf{r}|\mathbf{v})$ for a simple LG model of the RGCs is closely related to the energy function of [4], and thus for this model the Bayesian velocity decoding is nearly equivalent to the energy method. In the linear Gaussian model, the response of cell i, $\mathbf{r}_i$, is given linearly in terms of the image intensity profile $\mathbf{x}$ up to additive Gaussian noise with covariance Σ, as in Eq. (23). Thus we have

$$p_{\mathrm{LG}}(\mathbf{r}|\mathbf{x},\mathbf{v}) = \prod_i \mathcal{N}(b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}, \Sigma).$$

Using this and $p_x(\mathbf{x}) = \mathcal{N}(0, \mathcal{C}_x)$ as the Gaussian image prior, we repeat the steps in Eqs. (11)–(19) of Subsection 2.B.1. For the LG model, the log-posterior function is given by

$$\begin{aligned}\mathcal{L}_{\mathrm{LG}}(\mathbf{x},\mathbf{r},\mathbf{v}) \equiv \log[p_x(\mathbf{x})] + \log[p_{\mathrm{LG}}(\mathbf{r}|\mathbf{x},\mathbf{v})] &= -\tfrac{1}{2}\mathbf{x}^{\mathrm{T}}\mathcal{C}_x^{-1}\mathbf{x} \\ &\quad - \tfrac{1}{2}\sum_i (\mathbf{r}_i - b_i - \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x})^{\mathrm{T}}\Sigma^{-1}(\mathbf{r}_i - b_i - \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}) \\ &\quad + \text{const.} \end{aligned} \tag{A1}$$

instead of Eq. (11), and the marginal distribution $p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v})$ by

$$p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v}) = \int e^{\mathcal{L}_{\mathrm{LG}}(\mathbf{x},\mathbf{r},\mathbf{v})} \mathrm{d}\mathbf{x}, \tag{A2}$$

similar to Eq. (10). As before, setting $\nabla_x \mathcal{L}_{\mathrm{LG}} = 0$ yields the equation for $\mathbf{x}_{\mathrm{MAP}}$, which unlike Eq. (19) is linear, and can be easily solved to yield

$$\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}) = H(\mathbf{v})^{-1} \sum_i \mathcal{K}_{i,\mathbf{v}}^{\mathrm{T}} \cdot \Sigma^{-1} \cdot (\mathbf{r}_i - b_i). \tag{A3}$$

Here, the negative Hessian is given by

$$H(\mathbf{v}) = -\nabla_x \nabla_x \mathcal{L}_{\mathrm{LG}} = \mathcal{C}_x^{-1} + \sum_i \mathcal{K}_{i,\mathbf{v}}^{\mathrm{T}} \cdot \Sigma^{-1} \cdot \mathcal{K}_{i,\mathbf{v}}, \tag{A4}$$

which is now independent of the observed spike trains $\mathbf{r}$. Using Eqs. (A3) and (A4), we can rearrange the terms in Eq. (A1) to complete the square for $\mathbf{x}$, and obtain

$$\begin{aligned}\mathcal{L}_{\mathrm{LG}}(\mathbf{x},\mathbf{r},\mathbf{v}) &= -\tfrac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{MAP}})^{\mathrm{T}} H(\mathbf{v})(\mathbf{x} - \mathbf{x}_{\mathrm{MAP}}) \\ &\quad - \tfrac{1}{2}\sum_i \delta\mathbf{r}_i^{\mathrm{T}}\Sigma^{-1}\delta\mathbf{r}_i + \frac{1}{2}\sum_{ij} \mathbf{X}_i^{\mathrm{T}}\mathcal{C}_x(\mathbf{v})\mathbf{X}_j + \text{const.},\end{aligned} \tag{A5}$$

where $\mathcal{C}_x(\mathbf{v}) = H^{-1}(\mathbf{v})$ is the posterior covariance over the fixed image, and we defined the mean-adjusted response $\delta\mathbf{r}_i \equiv \mathbf{r}_i - b_i$ and the prefiltered response

$$\mathbf{X}_i \equiv \mathcal{K}_{i,\mathbf{v}}^{\mathrm{T}}\Sigma^{-1}\delta\mathbf{r}_i. \tag{A6}$$

The marginalization in Eq. (A2) is thus a standard Gaussian integration, which yields

$$\log p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v}) = \tfrac{1}{2}\sum_{ij} \mathbf{X}_i^{\mathrm{T}}\mathcal{C}_x(\mathbf{v})\mathbf{X}_j - \tfrac{1}{2}\log|\mathcal{C}_x H(\mathbf{v})| + \text{const.} \tag{A7}$$

(the constant term is independent of $\mathbf{v}$, and therefore irrelevant for estimating it). The decomposition into the two terms on the right-hand side of Eq. (A7) is similar to that in Eq. (18). In both equations the second term arose from a Gaussian integration over $\mathbf{x}$ [an approximation in the case of Eq. (18)], and the first was (up to a constant in $\mathbf{v}$) the value of the logarithm of the joint distribution of $\mathbf{x}$ and $\mathbf{r}$, given $\mathbf{v}$, at $\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v})$. Unlike Eq. (18), however, although the second term on the right-hand side of Eq. (A7) depends on $\mathbf{v}$, it is nevertheless independent of the observed response $\mathbf{r}$. The only term that modulates the velocity posterior depending on $\mathbf{r}$ (through the implicit dependence of $\mathbf{X}_i$) is the first, which we denote by $\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r})$. We will see that this term corresponds closely to the energy function introduced in [4]. More explicitly, we have

$$\begin{aligned}\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) &\equiv \frac{1}{2}\sum_{ij} \mathbf{X}_i^{\mathrm{T}}\mathcal{C}_x(\mathbf{v})\mathbf{X}_j \\ &= \frac{1}{2}\sum_{ij} \int\int X_i(\mathbf{n}_1)\mathcal{C}_x(\mathbf{n}_1,\mathbf{n}_2;\mathbf{v})X_j(\mathbf{n}_2)\mathrm{d}^2\mathbf{n}_1\mathrm{d}^2\mathbf{n}_2.\end{aligned} \tag{A8}$$

In the following we will rewrite Eq. (A8) in a form which is explicitly akin to Eq. (21). For simplicity, we assume that the noise covariance is white, i.e., $\Sigma = \sigma^2\mathbf{1}$. Physiologically, this implies that we are ignoring stimulus-conditional correlations and history dependences in the network (as, e.g., in the uncoupled model discussed in [7]). From Eq. (A6) and the definition of $\mathcal{K}_{i,\mathbf{v}}$, Eq. (7), we then obtain the explicit form

$$X_i(\mathbf{n}) = \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau k_i(t-\tau, \tau\mathbf{v} + \mathbf{n})\delta r_i(t). \tag{A9}$$

If we further assume that the spike train observation has not revealed much information about the identity of the fixed image (as happens, e.g., for low contrasts or short

presentation times), then the posterior distribution over $\mathbf{x}$ will not be very different from the prior $p_x(\mathbf{x})$. Therefore, we can use the approximation $\mathcal{C}_x(\mathbf{v}) \approx \mathcal{C}_x$. In the one-dimensional case, which we are studying in this paper, the image profile $\mathbf{x}(\mathbf{n})$, and hence the prior image covariance, depend only on the component of $\mathbf{n}$ parallel to the direction of motion $\hat{\mathbf{v}} = \mathbf{v}/|\mathbf{v}|$ and are constant in the perpendicular direction. Denoting the former component by $n(=\mathbf{n}\cdot\hat{\mathbf{v}})$ and the latter by $n_\perp(=\mathbf{n}-n\hat{\mathbf{v}})$, we can then perfom the integrals over $n_\perp$ in Eq. (A8) and rewrite it as

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2}\sum_{ij} \int\int \tilde{X}_i(n_1)\mathcal{C}_x(n_1,n_2)\tilde{X}_j(n_2)\mathrm{d}n_1\mathrm{d}n_2, \tag{A10}$$

$$\tilde{X}_i(n) \equiv \int X_i(\mathbf{n})\mathrm{d}n_\perp = \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau \tilde{k}_i(t-\tau, \tau v + n)\delta r_i(t), \tag{A11}$$

where $v \equiv |\mathbf{v}|$, and we defined $\tilde{k}_i(t,n) \equiv \int k_i(t,\mathbf{n})\mathrm{d}n_\perp$. For each cell i, we specify a fixed point $\mathbf{n}_i$ positioned at its receptive field center, so that $k_i(t,\mathbf{n}_i+\Delta\mathbf{n})$ vanishes when $|\Delta\mathbf{n}|$ gets considerably larger than the size of the receptive field surround ($\sim 1°$). Hence, if we define

$$q_i(t,n) \equiv \tilde{k}_i(t,n+n_i) = \int k_i(t,\mathbf{n}+\mathbf{n}_i)\mathrm{d}n_\perp, \tag{A12}$$

(where $n_i \equiv \mathbf{n}_i\cdot\hat{\mathbf{v}}$), $q_i(t,n)$ vanishes when $|n| \gg 1°$; for all cells, q_i are localized (up to the above scale) around the origin, as opposed to around the position of their respective receptive field centers along $\mathbf{v}$. In order to make the comparison with the energy model of Subsection 2.B.2 clearer, we also switch to the time domain (recalling that space n and time t are linked here via the velocity v); we define $\tilde{R}_i(t) \equiv \tilde{X}_i(n_i - vt)$ [equivalently, $\tilde{X}_i(n) = \tilde{R}_i((-n+n_i)/v)$] and rewrite Eq. (A10) by changing the integration variables from $n_{1(2)}$ to $vt_{1(2)}$:

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2v^2}\sum_{ij} \int\int \tilde{R}_i\left(-t_1+\frac{n_i}{v}\right)\mathcal{C}_x(vt_1,vt_2) \times \tilde{R}_j\left(-t_2+\frac{n_j}{v}\right)\mathrm{d}t_1\mathrm{d}t_2. \tag{A13}$$

Using Eq. (A11) and the definiton (A12), we write $R_i(t_1)$ explicitly as

$$\begin{aligned}\tilde{R}_i(t_1) &\equiv \tilde{X}_i(n_i - vt_1) \\ &= \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau \tilde{k}_i(t-\tau, v\tau - vt_1 + n_i)\delta r_i(t) \\ &= \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau q_i(t-\tau, v(\tau - t_1))\delta r_i(t).\end{aligned} \tag{A14}$$

Exploiting the translation invariance of the prior image ensemble that dictates $\mathcal{C}_x(n_1,n_2) = \mathcal{C}_x(n_1-n_2)$, we define $\mathcal{B}_x$ to be the operator square root of $\mathcal{C}_x$, in the sense that

$$\mathcal{C}_x(n_1-n_2) = \int \mathcal{B}_x(n_1-n)\mathcal{B}_x(n_2-n)\mathrm{d}n. \tag{A15}$$

In general, given an explicit form of $\mathcal{C}_x(n_1-n_2)$, $\mathcal{B}_x$ can be computed in the Fourier domain by taking the square root of the power spectrum [29]. In particular, for $\mathcal{C}_x(n_1-n_2) = c^2 e^{-|n_1-n_2|/l_{\mathrm{corr}}}$, we have $\mathcal{B}_x(n) = c\sqrt{2/l_{\mathrm{corr}}}\theta(n)e^{-n/l_{\mathrm{corr}}}$, where c is the image contrast, l_{corr} is the correlation length of typical images in the naturalistic prior ensemble, and $\theta(t)$ is the Heaviside step function. In the simulations of Subsection 3.A.4 we used this particular form of $\mathcal{C}_x$, as it yields (for spatial frequencies f larger than the inverse of the correlation length l_{corr} but smaller than the inverse image pixel size) a power spectrum $\propto 1/f^2$, as observed in natural images. Substituting definition (A15) (after renaming the integration variable n to vt) in Eq. (A13), we rewrite the latter as

$$\begin{aligned}\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) &= \frac{1}{2v}\sum_{ij} \int\int\int \tilde{R}_i\left(-t_1+\frac{n_i}{v}\right)\mathcal{B}_x(v(t_1-t)) \\ &\quad\times \mathcal{B}_x(v(t_2-t))\tilde{R}_j\left(-t_2+\frac{n_j}{v}\right)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t \\ &= \frac{1}{2v}\sum_{ij} \int\int\int \tilde{R}_i(t_1)\mathcal{B}_x\left(v\left(t+\frac{n_i}{v}-t_1\right)\right) \\ &\quad\times \mathcal{B}_x\left(v\left(t+\frac{n_j}{v}-t_2\right)\right)\tilde{R}_j(t_2)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t.\end{aligned} \tag{A16}$$

We derived the last line by renaming the integration variables as $t_1 \to n_i/v - t_1$, $t_2 \to n_j/v - t_2$, and $t \to -t$. Finally, defining

$$R_i(t) \equiv \frac{1}{\sqrt{v}}\int \mathcal{B}_x(v(t-t_1))\tilde{R}_i(t_1)\mathrm{d}t_1, \tag{A17}$$

we obtain

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2}\sum_{ij} \int R_i\left(t+\frac{n_i}{v}\right)R_j\left(t+\frac{n_j}{v}\right)\mathrm{d}t. \tag{A18}$$

Equation (A18) is akin to the energy function used in [4], and together with Eq. (A7) yields Eq. (24) of Subsection 2.B.3. To find the explicit form of the smoothing filter in Eq. (25), we compare that equation, in the form

$$R_i(t) = \int w_{\mathrm{LG}}(t-t')\delta r_i(t')\mathrm{d}t', \tag{A19}$$

with definition (A17):

$$\begin{aligned}R_i(t) &= \frac{1}{\sigma^2\sqrt{v}}\int \mathrm{d}t_1 \int \mathrm{d}t' \int \mathrm{d}\tau \mathcal{B}_x(v(t-t_1)) \\ &\quad\times q_i(t'-\tau, v(\tau-t_1))\delta r_i(t'), \\ &= \frac{1}{\sigma^2\sqrt{v}}\int \mathrm{d}t_1 \int \mathrm{d}t' \int \mathrm{d}\tau \mathcal{B}_x(v(t-t'-t_1)) \\ &\quad\times q_i(-\tau, v(\tau-t_1))\delta r_i(t'),\end{aligned} \tag{A20}$$

[where we used Eq. (A14) to write the first line, and

shifted τ and t_1 by t' to derive the second], and obtain

$$w_{\mathrm{LG}}(t) = \frac{1}{\sigma^2\sqrt{v}} \int \mathrm{d}t_1 \int \mathrm{d}\tau \mathcal{B}_x(v(t-t_1)) q_i(-\tau, v(\tau - t_1)). \tag{A21}$$

Thus, $R_i(t)$ is a version of the response function of the cell i offset by its baseline log firing rate b_i and smoothed out on the time scale dictated by the largest of the spatiotemporal scales of the receptive fields (via q_i) or the correlation length of typical images (via $\mathcal{B}_x$)—with spatial scales converted to time scales by dividing by v. To see this more precisely, let us define $\Delta\tau_1 \equiv \tau$, $\Delta\tau_2 \equiv t_1 - \tau$, and $\Delta\tau_3 \equiv t - t_1$, such that $t = \Delta\tau_1 + \Delta\tau_2 + \Delta\tau_3$. On the other hand, because of the finite support of the factors of its integrand, the double integral Eq. (A21) receives nonzero contributions only when $|\Delta\tau_1| \lesssim \tau_k$, $|\Delta\tau_2| \lesssim l_k/v$, and $|\Delta\tau_3| \lesssim l_{\mathrm{corr}}/v$ [where τ_k and l_k are the typical temporal and spatial size of the receptive field filters $k_i(t,n)$, respectively, and l_{corr} is the correlation length of typical images in the naturalistic prior ensemble]. Thus if $|t| = |\tau_1 + \Delta\tau_2 + \Delta\tau_3|$ is much larger than the sum of the three scales τ_k, l_k/v, and l_{corr}/v, the filter $w(t)$ is bound to vanish. This leads to the discussion of Subsection 2.B.3 following Eq. (25).

APPENDIX B: $\mathcal{O}(D)$ DECODING

Here, we discuss how to implement the Newton–Raphson optimization algorithm such that it finds the maximum *a posteriori* estimate for the image $\mathbf{x}_{\mathrm{MAP}}$ [satisfying Eq. (19)] in cases where the image depends only on one spatial dimension and in a computational time that scales only linearly with the spatial dimensionality of the image vector d. The Newton–Raphson algorithm for minimizing the function $\mathcal{L}(\mathbf{x})$ works as follows [we have in mind the objective function defined in Eq. (13), but for simplicity we drop $\mathbf{r}$ and $\mathbf{v}$ from its arguments in this appendix]. At each iteration of this algorithm, starting from the vector $\mathbf{x}$, we change this vector by an amount $\Delta\mathbf{x}$ which is found by solving the set of linear equations $H(\mathbf{x})\Delta\mathbf{x} = \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x})$. Here, the right-hand side is the gradient of $\mathcal{L}(\mathbf{x})$, and $H(\mathbf{x})$ is its negative Hessian matrix [as in Eq. (15)], both evaluated at $\mathbf{x}$. In general, the solution of a set of d linear equations can be calculated in $\mathcal{O}(d^3)$ elementary operations [47]. This would make the decoding of images with even moderate angular extension forbidding. Fortunately, as we will now explain, the quasi-locality of the GLM model allows us to overcome this limitation. The negative Hessian of $\mathcal{L}(\mathbf{x})$, Eq. (13), is given by

$$H(\mathbf{x}) = \mathcal{C}_x^{-1} + \sum_{i,t} J_{i,t}(\mathbf{x}), \tag{B1}$$

where the matrices $J_{i,t}(\mathbf{x})$ have the elements

$$J_{i,t}(n_1, n_2; \mathbf{x}) = \mathcal{K}_{i,\mathbf{v}}(t; n_1)\mathcal{K}_{i,\mathbf{v}}(t; n_2)\lambda_i(t; \mathbf{x})\mathrm{d}t, \tag{B2}$$

and $\mathcal{K}_{i,\mathbf{v}}(t;n)$ was defined in Eq. (7) in terms of the receptive field filter of the cell i, $k_i(\tau,n)$ [as we are considering the one-dimensional image case, $k_i(\tau,n)$ is understood to be the full receptive field integrated along the transverse spatial dimension]. Here, we turned the integral over t in Eq. (13) into a discrete sum in Eq. (B1), as is done in the numerical implementation, and for simplicity we wrote Eq. (B2) for the case of exponential GLM nonlinearity. Generalization of this equation and the rest of the argument to general nonlinearities is straightforward. Because of the finite spatial size of the receptive field and the finite duration of the temporal filter, $k_i(\tau,n)$ is nonzero only when $\tau \in [0, T_k]$ and $n \in [n^i_{\min}, n^i_{\max}]$, where T_k and $\Delta n \equiv n^i_{\max} - n^i_{\min}$ are upper bounds on the cells' temporal integration windows and the size of the cells' receptive field surrounds, respectively. It follows then from Eq. (7) that $\mathcal{K}_{i,\mathbf{v}}(t;n)$ vanishes unless $n^i_{\min} - vt \leq n \leq n^i_{\max} - vt + vT_k$ (we assumed $v > 0$, but generalization to $v < 0$ is straightforward). Thus $J_{i,t}(n_1, n_2; \mathbf{x})$ vanishes if $|n_1 - n_2| > \Delta n + vT_k$, regardless of i and t. In other words, for all (i,t), $J_{i,t}(\mathbf{x})$ are banded matrices with a band width of $\Delta n + vT_k$, and so is their sum. If we further use a prior covariance $\mathcal{C}_x$ with a banded inverse, then the full Hessian Eq. (B1) will be banded [e.g., the naturalistic prior covariance introduced after Eq. (A15) can be defined as the inverse of a tridiagonal matrix in the numerical implementation].

Unlike in the general case, the solution of a set of d linear equations with a banded equation matrix of band width B can be found in a computational time $\propto B^2 d$—i.e., in our case, in a computational time scaling only linearly (as opposed to cubically) with the image size d. On the other hand, we have observed empirically that the number of necessary Newton–Raphson iterations is more or less constant and does not scale with d. Hence the overall optimization procedure for finding $\mathbf{x}_{\mathrm{MAP}}$ can be performed in $\mathcal{O}(d)$ computational time. This allows us to decode the velocity of large moving images. Similar methods with $\mathcal{O}(d)$ computational cost have been used in inference and estimation problems involving state–space models [48,49], e.g., in applications to neural data analysis [50].

ACKNOWLEDGMENTS

Thanks to E. P. Simoncelli for very detailed comments on the manuscript, to J. Pillow for providing us with the parameters for the network model introduced in [7], and to D. Pfau and E. J. Chichilnisky for many useful comments. YA and LP are partially supported by NEI grant R01 EY018003 and by a McKnight Scholar award to LP. YA is additionally supported by a Patterson Trust Fellowship in Brain Circuitry. EL is supported by an Irish Research Council for Science, Engineering and Technology (IRCSET) Government of Ireland Postdoctoral Research Fellowship.

REFERENCES

1. M. Meister, L. Lagnado, and D. Baylor, "Concerted signaling by retinal ganglion cells," Science **270**, 1207–1210 (1995).
2. S. Nirenberg, S. Carcieri, A. Jacobs, and P. Latham, "Retinal ganglion cells act largely as independent encoders," Nature **411**, 698–701 (2002).
3. E. Chichilnisky and R. Kalmar, "Functional asymmetries in ON and OFF ganglion cells of primate retina," J. Neurosci. **22**, 2737–2747 (2002).

4. E. Frechette, A. Sher, M. Grivich, D. Petrusca, A. Litke, and E. Chichilnisky, "Fidelity of the ensemble code for visual motion in the primate retina," J. Neurophysiol. **94**, 119–135 (2005).
5. E. Schneidman, M. Berry, R. Segev, and W. Bialek, "Weak pairwise correlations imply strongly correlated network states in a neural population," Nature **440**, 1007–1012 (2006).
6. J. Shlens, G. Field, J. Gauthier, M. Grivich, D. Petrusca, A. Sher, A. Litke, and E. Chichilnisky, "The structure of multi-neuron firing patterns in primate retina," J. Neurosci. **26**, 8254–8266 (2006).
7. J. Pillow, J. Shlens, L. Paninski, A. Sher, A. Litke, E. Chichilnisky, and E. Simoncelli, "Spatio-temporal correlations and visual signalling in a complete neuronal population," Nature **454**, 995–999 (2008).
8. E. S. Frechette, M. I. Grivich, R. S. Kalmar, A. M. Litke, D. Petrusca, A. Sher, and E. J. Chichilnisky, "Retinal motion signals and limits on speed discrimination," J. Vision **4**, 570 (2004).
9. A. Litke, N. Bezayiff, E. Chichilnisky, W. Cunningham, W. Dabrowski, A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R. Kalmar, K. Mathieson, D. Petrusca, M. Rahman, and A. Sher, "What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity," IEEE Trans. Nucl. Sci. **51**, 1434–1440 (2004).
10. R. Segev, J. Goodhouse, J. Puchalla, and M. Berry, "Recording spikes from a large fraction of the ganglion cells in a retinal patch," Nat. Neurosci. **7**, 1154–1161 (2004).
11. D. Knill and W. Richards, eds., *Perception as Bayesian Inference* (Cambridge Univ. Press, 1996).
12. E. P. Simoncelli, "Distributed analysis and representation of visual motion," Ph.D. thesis (Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1993). Also available as MIT Media Laboratory Vision and Modeling Technical Report #209.
13. D. Ascher and N. Grzywacz, "A Bayesian model for the measurement of visual velocity," Vision Res. **40**, 3427–3434 (2000).
14. Y. Weiss, E. Simoncelli, and E. Adelson, "Motion illusions as optimal percepts," Nat. Neurosci. **5**, 598–604 (2002).
15. E. P. Simoncelli, "Local analysis of visual motion," in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, eds. (MIT Press, 2003), Chap. 109, pp. 1616–1623.
16. A. Stocker and E. Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," Nat. Neurosci. **9**, 578–585 (2006).
17. A. E. Welchman, J. M. Lam, and H. H. Bulthoff, "Bayesian motion estimation accounts for a surprising bias in 3D vision," Proc. Natl. Acad. Sci. U.S.A. **105**, 12087–12092 (2008).
18. F. Hurlimann, D. Kiper, and M. Carandini, "Testing the Bayesian model of perceived speed," Vision Res. **42**, 2253–2257 (2002).
19. P. Thompson, K. Brooks, and S. Hammett, "Speed can go up as well as down at low contrast: Implications for models of motion perception," Vision Res. **46**, 782–786 (2005).
20. A. Thiel, M. Greschner, C. Eurich, J. Ammermüller, and J. Kretzberg, "Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding," J. Neurophysiol. **98**, 2285–2296 (2007).
21. J. Kretzberg, I. Winzenborg, and A. Thiel, "Bayesian analysis of the encoding of constant and changing stimulus velocities by retinal ganglion cells," presented at Frontiers in Neuroinformatics 2008, Stockholm, September 7–9, 2008.
22. D. Brillinger, "Maximum likelihood analysis of spike trains of interacting nerve cells," Biol. Cybern. **59**, 189–200 (1988).
23. P. McCullagh and J. Nelder, *Generalized Linear Models* (Chapman & Hall, 1989).
24. L. Paninski, "Maximum likelihood estimation of cascade point-process neural encoding models," Network Comput. Neural Syst. **15**, 243–262 (2004).
25. W. Truccolo, U. Eden, M. Fellows, J. Donoghue, and E. Brown, "A point process frame-work for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects," J. Neurophysiol. **93**, 1074–1089 (2005).
26. L. Paninski, J. Pillow, and J. Lewi, "Statistical models for neural encoding, decoding, and optimal stimulus design," in *Computational Neuroscience: Progress in Brain Research*, P. Cisek, T. Drew, and J. Kalaska, eds. (Elsevier, 2007).
27. D. Snyder and M. Miller, *Random Point Processes in Time and Space* (Springer-Verlag, 1991).
28. D. Field, "Relations between the statistics of natural images and the response profiles of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).
29. D. H. Brainard, D. R. Williams, and H. Hofer, "Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots," J. Vision **8**, 1–23 (2008).
30. R. Kass and A. Raftery, "Bayes factors," J. Am. Stat. Assoc. **90**, 773–795 (1995).
31. E. Brown, L. Frank, D. Tang, M. Quirk, and M. Wilson, "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells," J. Neurosci. **18**, 7411–7425 (1998).
32. W. Bialek and A. Zee, "Coding and computation with neural spike trains," J. Stat. Phys. **59**, 103–115 (1990).
33. S. Koyama and S. Shinomoto, "Empirical Bayes interpretations of random point events," J. Phys. A **38**, 531–537 (2005).
34. J. Pillow, Y. Ahmadian, and L. Paninski, "Model-based decoding, information estimation, and change-point detection in multi-neuron spike trains," submitted to Neural Comput.
35. Y. Ahmadian, J. Pillow, and L. Paninski, "Efficient Markov chain Monte Carlo methods for decoding neural spike trains," submitted to Neural Comput.
36. E. Adelson and J. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Am. A **2**, 284–99 (1985).
37. E. Chichilnisky and R. Kalmar, "Temporal resolution of ensemble visual motion signals in primate retina," J. Neurosci. **23**, 6681–6689 (2003).
38. W. Bialek (Princeton University, bbrinker@princeton.edu) and R. de Ruyter van Steveninck (Indiana University, deruyter@indiana.edu) (personal communication, 2003).
39. V. Perry and A. Cowey, "The ganglion cell and cone distributions in the monkey's retina: implications for central magnification factors," Vision Res. **25**, 1795–1810 (1985).
40. S. Ullman, *The Interpretation of Visual Motion* (MIT Press, 1979).
41. P. Thompson, "Perceived rate of movement depends on contrast," Vision Res. **22**, 377–380 (1982).
42. L. Stone and P. Thompson, "Human speed perception is contrast dependent," Vision Res. **32**, 1535–1549 (1992).
43. D. C. Bradley and M. S. Goyal, "Velocity computation in the primate visual system," Nat. Rev. Neurosci. **9**, 686–695 (2008).
44. M. Potters and W. Bialek, "Statistical mechanics and visual signal processing," J. Phys. I France **4**, 1755–1775 (1994).
45. S. McKee, G. Silvermann, and K. Nakayama, "Precise velocity discrimination despite random variations in temporal frequency and contrast," Vision Res. **26**, 609–619 (1986).
46. M. Blakemore and R. Snowden, "The effect of contrast upon perceived speed: a general phenomenon?" Perception **28**, 33–48 (1999).
47. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C* (Cambridge Univ. Press, 1992).

48. N. Shephard and M. Pitt, "Likelihood analysis of non-Gaussian measurement time series," Biometrika **84**, 653–667 (1997).
49. R. Davis and G. Rodriguez-Yam, "Estimation for state-space models: an approximate likelihood approach," Stat. Sin. **15**, 381–406 (2005).
50. L. Paninski, Y. Ahmadian, D. Ferreira, S. Koyama, K. Rahnama, M. Vidne, J. Vogelstein, and W. Wu, "A new look at state-space models for neural data," J. Comput. Neurosci. (to be published). Epub ahead of print, doi 10.1007/s10827-009-0179-x.

Available online at www.sciencedirect.com
ScienceDirect
Schizophrenia Research 98 (2008) 256–264
SCHIZOPHRENIA RESEARCH
www.elsevier.com/locate/schres

Dissecting the cellular contributions to early visual sensory processing deficits in schizophrenia using the VESPA evoked response

Edmund C. Lalor [a,b,c], Sherlyn Yeap [a,b], Richard B. Reilly [a,c], Barak A. Pearlmutter [d], John J. Foxe [a,b,e,*]

[a] *The Cognitive Neurophysiology Laboratory, St Vincent's Hospital Fairview, Dublin, Ireland*
[b] *The Cognitive Neurophysiology Laboratory, Nathan S. Kline Institute for Psychiatric Research, Program in Cognitive Neuroscience and Schizophrenia, Orangeburg, New York 10962, USA*
[c] *School of Mechanical, Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland*
[d] *The Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland*
[e] *Program in Cognitive Neuroscience, Department of Psychology, City College of the City University of New York, 138th Street & Convent Avenue, New York, New York 10031, USA*

Received 8 March 2007; received in revised form 29 August 2007; accepted 6 September 2007
Available online 8 November 2007

Abstract

Electrophysiological research has shown clear dysfunction of early visual processing mechanisms in patients with schizophrenia. In particular, the P1 component of the visual evoked potential (VEP) is substantially reduced in amplitude in patients. A novel visual evoked response known as the VESPA (Visual Evoked Spread Spectrum Analysis) was recently described. This response has a notably different scalp topography from that of the traditional VEP, suggesting preferential activation of a distinct subpopulation of cells. As such, this method constitutes a potentially useful candidate for investigating cellular contributions to early visual processing deficits. In this paper we compare the VEP and VESPA responses between a group of healthy control subjects and a group of schizophrenia patients. We also introduce an extension of the VESPA method to incorporate nonlinear processing in the visual system. A significantly reduced P1 component was found in patients using the VEP (with a large effect size; Cohen's $d = 1.6$), while there was no difference whatsoever in amplitude between groups for either the linear or nonlinear VESPA. This pattern of results points to a highly specific cellular substrate of early visual processing deficits in schizophrenia, suggesting that these deficits are based on dysfunction of magnocellular pathways with parvocellular processing remaining largely intact.

Keywords: EEG; Visual evoked potential; VESPA; Schizophrenia; P1 component

1. Introduction

Visual evoked potential (VEP) studies have consistently demonstrated that patients with schizophrenia exhibit relatively severe deficits in early visual sensory processing, as indexed by a robust decrement in amplitude of the occipital P1 component (e.g., Foxe et al., 2001, 2005; Butler et al., 2001, 2007; Doniger et al., 2002;

* Corresponding author. The Cognitive Neurophysiology Laboratory, Nathan S. Kline Institute for Psychiatric Research, Program in Cognitive Neuroscience and Schizophrenia, 140 Old Orangeburg Road, Orangeburg, New York 10962, USA. Tel.: +1 845 398 6547; fax: +1 845 398 6545.
E-mail addresses: elalor@nki.rfmh.org (E.C. Lalor), richard.reilly@ucd.ie (R.B. Reilly), barak@cs.nuim.ie (B.A. Pearlmutter), sherlyn_yeap@hotmail.com (S. Yeap), foxe@nki.rfmh.org (J.J. Foxe).

0920-9964/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.schres.2007.09.037

Spencer et al., 2003; Schechter et al., 2005; Haenschel et al., in press). Concomitant structural deficits have also been shown in the visual sensory pathways (Butler et al., 2006). Scalp topographies and source analysis have suggested that these deficits may specifically reflect dysfunction of the dorsal visual stream while processing in the ventral stream remains relatively more intact (e.g. Foxe et al., 2001, 2005). It is also suggested that certain ventral stream processes are contingent on inputs from the dorsal stream and as a result failure in these 'higher-level' ventral stream processes may ultimately be a consequence of these underlying dorsal stream deficits (Doniger et al., 2002; Foxe et al., 2005).

Further to the above findings, a substantial decrement in the P1 component was recently demonstrated in clinically unaffected first-degree relatives of schizophrenia patients (Yeap et al., 2006), establishing a possible genetic basis for the observed effects (see also Donohoe et al., in press). This points to the potential use of P1 amplitude as an endophenotypic marker for schizophrenia and, as such, it may be a significant step in the quest for a diagnostic test facilitating early detection of schizophrenia in high-risk individuals.

It would be of great benefit to this line of research to have a more sensitive method for eliciting this deficit. One very promising candidate method, known as the VESPA technique (for Visual Evoked Spread Spectrum Analysis), was recently described (Lalor et al., 2006). This method uses stimuli, the luminance or contrast of which is rapidly and unobtrusively modulated by a stochastic signal, enabling the estimation of the linear impulse response of the visual system. The temporal profile of these VESPAs is highly correlated with that of transient VEPs evoked using standard, discrete stimuli. This includes a clearly defined and, hence, measurable P1 component. The rapidly estimable VESPA has been shown to be superior to the VEP in terms of the amount of time necessary to obtain a response with a specific signal-to-noise ratio. Furthermore, the method allows for a large degree of flexibility in design, not just in terms of the parameters of the stimuli, as in VEP studies, but also in the characteristics of the modulating signal.

The topography of the VESPA is notably different from that of the transient VEP. The abiding characteristic of the early VESPA maps is a persistently delimited focus over midline occipital scalp without any evidence for the characteristic early bilateral spread over lateral occipital scalp regions that is consistently seen for the standard VEP (e.g. Gomez-Gonzalez et al., 1994; Foxe and Simpson, 2002). This pattern suggests that the VESPA may well have a distinct cellular activation pattern from that of the VEP, favoring midline structures such as striate cortex and neighboring retinotopically mapped extrastriate regions, and perhaps also regions in the dorsal visual stream, activation of which are known to produce midline scalp topographies (Clark and Hillyard 1996; Foxe and Simpson 2002). This suggests the VESPA as an excellent candidate for further investigation of a dorsal stream based P1 deficit in schizophrenia.

For that reason, the aim of this paper is to compare VEPs and VESPA responses from schizophrenia patients and healthy controls. Specifically, we examine the relative magnitudes of the P1 components between groups for both types of response. A direct comparison between the VEP and VESPA is complicated by the assumption of linearity intrinsic to the VESPA estimation. In order to address this, we introduce a method for extending the VESPA analysis to higher orders and we expand our comparison between patients and controls to quadratic VESPA responses.

2. Materials and methods

2.1. Subjects

Written informed consent is obtained from 13 (1 female) patients with DSM-IV diagnosis of schizophrenia. The Ethics Committee of St. Vincent's Hospital approved the experimental procedures. Patients were aged 21 to 49 (mean±SD, 33.2±10.1 years) and had a mean illness duration of 12.3 years (SD±9.5). These patients had mean±SD scores on the Brief Psychiatric Rating Scale and SANS of 33.1±5.8 and 22.2±17.7, respectively. Twelve of the patients were receiving antipsychotic medication at the time of testing with a mean chlorpromazine equivalent dose of 406.71 mg/d (range, 50–1500 mg/d). The types of antipsychotics included atypicals, typicals or a combination of both. One patient had ceased taking medication 5 months prior to testing and was medication-free at the time of testing.

Control subjects were recruited from the St Vincent's Hospital staff community and through local recruitment efforts in the hospital catchment area. This group comprised 11 (2 female) paid volunteers aged 19 to 50 years (mean±SD, 26.5±8.7 years). The mean age of patients and controls did not differ significantly ($t_{50}=1.6$, $p=0.12$). All of the 11 controls, and 12 of the 13 patients were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). None of the controls were receiving any psychotropic medication at the time of testing. Also, all controls were free of any psychiatric illness or symptoms by self-report using criteria from the Structured Clinical Interview for DSM-III-R–Non-

Patient (SCID-NP), and all reported no history of alcohol or substance abuse.

2.2. Stimuli

VEPs were obtained during a study aimed at assessing schizophrenia patient deficits in visual binding processes using Kanizsa illusory figures. Accordingly, twelve different stimulus displays were presented, some containing Kanizsa illusions and others not. Specifically, the VEPs presented in the current study were obtained using a layout consisting of three centrally presented, black "pacmen" (disks with missing sectors) elements, whose arrangement was such as to *not* give any illusory effect, as in Fig. 1(a).[1] The array of elements subtended maximal visual angles of 2.28° horizontally and vertically and was presented for 400 ms. This allowed ERP analysis to be performed for the first 400 ms without contamination of any visual offset response. The mean inter-stimulus-interval (ISI) between trials was 900 ms (ranging from 600–1200 ms). Thus, each trial had a mean duration of 1300 ms. Note that this stimulus arrangement was chosen as it has previously been shown to elucidate a large VEP P1 deficit in patients with schizophrenia (Foxe et al., 2005; Spencer et al., 2003).

In the case of the VESPA, the stimulus consisted of a checkerboard pattern with equal numbers of black and white checks as in Fig. 1(b). Each check subtended a visual angle of 0.65° both horizontally and vertically, while the checkerboard as a whole subtended visual angles of 5.25° vertically and horizontally. The refresh rate of the monitor was set to 60 Hz and on every refresh the contrast of the checkerboard pattern was modulated by a stochastic signal with the mean luminance remaining constant. The stochastic signals used had their power distributed uniformly between 0 and 30 Hz. See Lalor et al. (2006) for details.

2.3. Experimental procedure

Each VEP experimental block consisted of, on average, 12.75 presentations of each of the 12 display types in a random order. Subjects underwent 20 blocks, resulting in 255 presentations of each display type. During VEP runs a small fixation point was present in the center of the screen, on which subject were instructed to maintain their gaze.

[1] Although only the non-illusion inducing arrangement was used, the reader should note that the P1 component is entirely insensitive to the presence or absence of illusory contours (Murray et al., 2002, 2004, 2006).

Fig. 1. Stimuli used to elicit (a) the VEP — non-illusory arrangement from Kanizsa study (b) the VESPA — single checkerboard the contrast of which is rapidly modulated as in Lalor et al. (2006).

Every subject underwent three VESPA runs of 200 s each. Subjects were instructed to maintain visual fixation on the center of the screen for the duration of each run. While abstaining from eye-blinks was not possible given the trial lengths, subjects were instructed to keep the number of eye-blinks to a minimum. A different modulating waveform was used for each run, although all waveforms had identical statistics.

2.4. EEG acquisition and analysis

EEG data were recorded from 72 electrode positions referenced to location Fz, filtered over the range 0–134 Hz and digitized at a rate of 512 Hz using the BioSemi Active Two system. Subsequently, the EEG was digitally filtered with a high-pass filter with passband above 2 Hz and −60 dB response at 1 Hz and a low-pass filter with 0–35 Hz passband and −50 dB response at 45 Hz.

VEPs were calculated by averaging time-locked responses to the presentations of the display type described earlier. A time window of 500 ms starting 100 ms pre-stimulus was used. Any epochs where the EEG exceeded +/−120 μV were rejected, resulting in a mean rejection rate of 11%.

The VESPA is an estimate of the linear impulse response of the visual system (Lalor et al., 2006). It is based on the assumption that the EEG response to a stimulus, whose luminance or contrast is rapidly modulated by a stochastic signal, consists of a convolution of that signal with an unknown impulse response. Given the known stimulus signal and the measured EEG, this impulse response, i.e., the VESPA, can be estimated using the method of linear least squares. In the present study VESPAs were measured using a sliding window of 500 ms of data starting 100 ms pre-stimulus.

It is possible that a VESPA founded on an assumption of linearity may not be sensitive to the deficits apparent in the VEP. The method can, however, easily be extended to higher orders. For example, in the case of a quadratic analysis, this is accomplished by including in the least squares estimation not only the

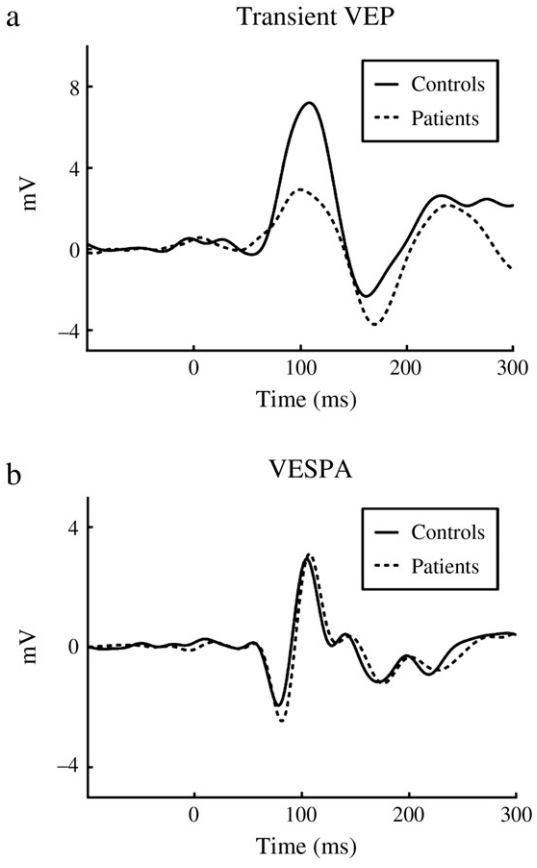

Fig. 2. Average time courses of the transient VEP and the VESPA for controls and patients for both methods at electrode location O2.

1st-order values of the modulating signal within the desired window but also all 2nd-order products of these values (see Appendix for details). This allows us to determine how the EEG depends, not only on the individual input signal values, but also on interactions between inputs at different time lags. In the present study, the quadratic VESPA response was measured using a sliding window of 120 ms of data starting 20 ms post-stimulus.

3. Results

Fig. 2(a) and (b) show the transient VEP and the average VESPA respectively for both the control group and the patients, at electrode location O2. Because the goal of this study was to examine the relative sensitivities of the VESPA and VEP methods to the P1 deficit in schizophrenia, we wished to determine the magnitudes of the P1 component for both methods and groups. We defined the P1 dependent measure as the average amplitude in the interval 90–115 ms, selected on the basis of peak latencies in group-average waveforms. First, an omnibus 2×2×9 ANOVA was carried out with factors of group (controls vs. patients), method (VEP vs. VESPA) and electrode (PO7, PO3, O1, Oz, POz, Pz, O2, PO4, PO8).

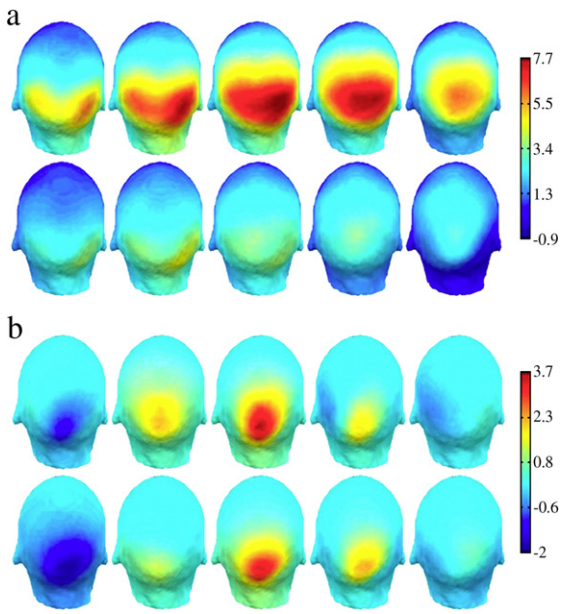

Fig. 3. (a): VEP in μV at 85, 95, 105, 115 and 125 ms. Controls (top) and patients (bottom). (b): VESPA in μV at 85, 95, 105, 115 and 125 ms. Controls (top) and patients (bottom).

A main effect of method ($F(1, 21)=32.24$, $p<0.001$) was found which simply reflects differences in response magnitudes between the two methods, either as a result of the methods themselves or of the specific stimuli used in each method. More importantly, an interaction was found between group and method ($F(1, 21)=7.89$, $p<0.05$). As can be seen in Fig. 2, this was driven by a much larger reduction in P1 amplitude for patients using the VEP method than the VESPA method. A main effect of

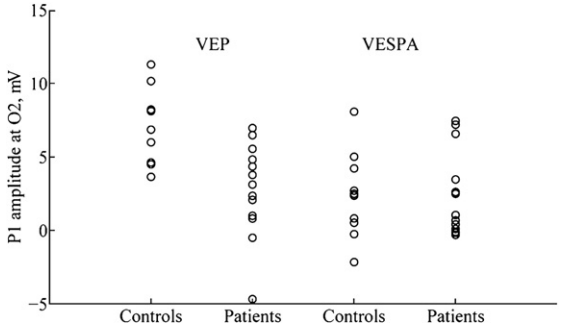

Fig. 4. Scatter plot, showing the distribution of the mean value of the P1 component in the interval 95–115 ms at electrode O2, for all control subjects and patients for both methods.

electrode ($F(8, 168)=10.29$, $p<0.001$) reflected the topographic specificity of the P1. A significant interaction between electrode and method ($F(8, 168)=3.11$, $p<0.05$) reflected the topographic differences between methods evident in Fig. 3. There was no three-way interaction between group, method and electrode ($F(8, 168)=1.57$, $p>0.1$).

To examine the interaction between group and method further, planned *t*-tests were carried out for each method separately. We used the P1, averaged across electrodes O1, Oz and O2, as the dependent measure. A significant difference was found between groups for the VEP method ($t=3.5$, $p<0.005$) whereas no difference was found between groups for the VESPA method ($t=0.08$, $p>0.9$). The Cohen's *d* effect size was calculated for the VEP P1 and found to be 1.57. Figs. 4 and 5 provide further illustration of the differing effects found using the VEP and VESPA methods. Fig. 4 is a scatter plot showing how the mean value of the P1 component in the interval 95–115 ms at electrode O2 is distributed within the control and patients groups for both methods. The standard deviations for the controls were 2.37 for the VEP and 2.76 for the VESPA and for the patients were 3.21 for the VEP and 2.83 for the VESPA respectively.

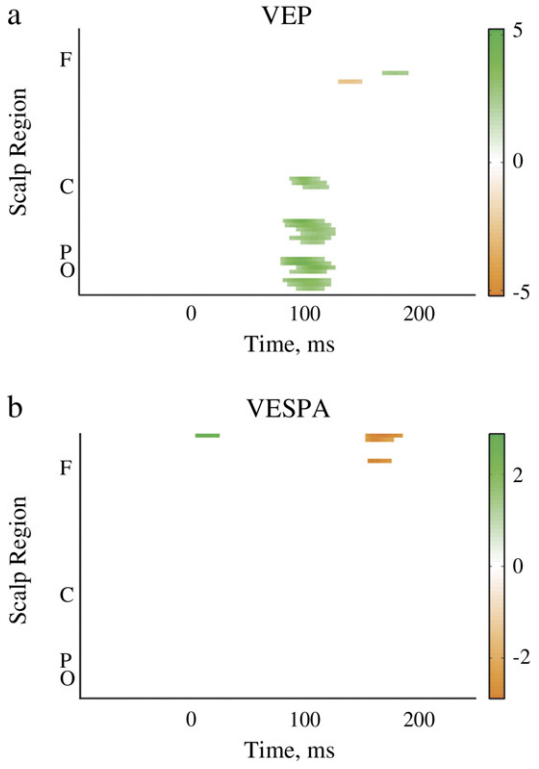

Fig. 5. Statistical cluster plot marking for all electrodes the time points at which the event-related potential differed significantly between groups on the basis of 2-tailed *t*-tests at an α level of 0.05. White denotes nonsignificance while positive *t* values (Controls>Patients) are marked on a green scale and negative *t* values (Patients>Controls) are marked in gold. Electrodes are ordered from the bottom, occipital (O), parietal (P), central (C), and frontal (F) proceeding in the anterior direction in rows from left to right. In the case of the VEP, a cluster is seen over posterior sites in the P1 interval 90 to 120 ms as expected from the results of the planned analysis of variance. No meaningful clusters are seen for the VESPA.

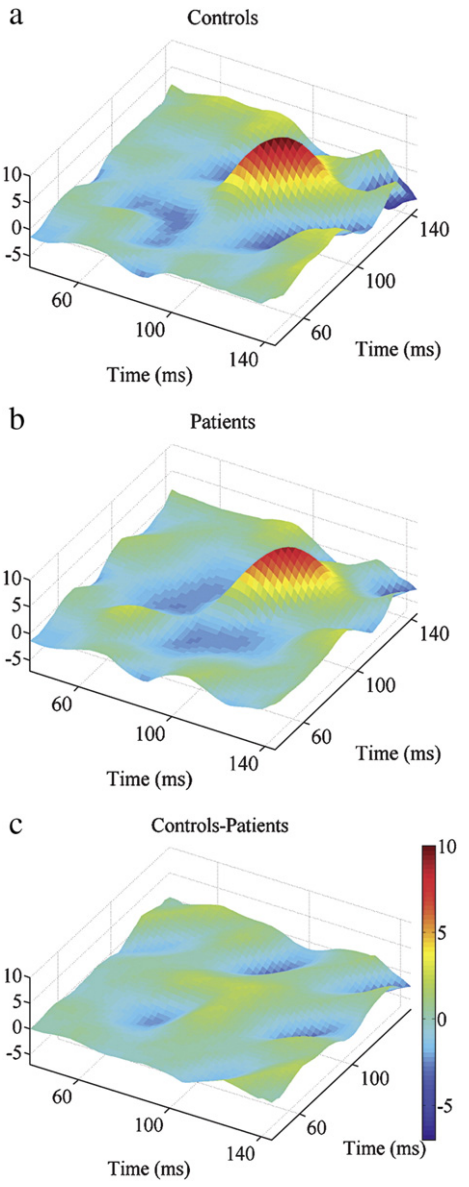

Fig. 6. Grand average quadratic VESPAs at electrode location Oz for controls, patients and the difference between controls and patients. These plots indicate how strongly the EEG at a particular time point depends on the interaction between inputs at two previous time points. Prominent "P1" activity can be see around 100×100 ms for both groups with a negligible difference between groups.

Fig. 5 illustrates a pair of statistical cluster plots marking the time points at which the VEP and VESPA respectively differ significantly between groups for all electrodes. A large cluster is evident in the VEP plot over posterior sites in the P1 interval 90–120 ms. No such cluster is evident in the VESPA plot.

Fig. 6 shows the average quadratic VESPA for both groups for electrode location Oz. Also plotted is the difference between controls and patients. Statistics confirmed the lack of a between-group difference evident from the figure. We defined the P1 dependent measure as the average amplitude in the interval 90–115 ms × 90–115 ms. A 2-way ANOVA was carried out using this P1 measure, with factors of group (controls vs. patients) and electrode (PO7, PO3, O1, Oz, POz, Pz, O2, PO4, PO8). No significant difference was found between groups ($F(1, 22) = 0.197$, $p > 0.6$).

4. Discussion

Replicating the findings of earlier studies (Foxe et al., 2001, 2005; Butler et al., 2001, 2007; Doniger et al., 2002; Spencer et al., 2003; Schechter et al., 2005; Haenschel et al., in press), a substantial reduction in the amplitude of the P1 component of the transient VEP was observed for schizophrenia patients compared with healthy controls. However, somewhat to our surprise, there was no difference whatsoever in VESPA P1 amplitude between patients and controls, in either the linear or quadratic case. This striking contrast in outcomes between two essentially similar methods points to a highly specific disparity between the early visual sensory processing systems of patients and controls. While the scalp topography of the VESPA (Fig. 3) suggests that it may preferentially activate the dorsal visual stream and given that a number of studies have proposed that the VEP P1 deficit shown by patients with schizophrenia may be due to specific dysfunction of this stream (Doniger et al., 2002; Foxe et al., 2005), we had originally expected that the midline-focused VESPA might prove to be more sensitive to this deficit. This is clearly not the case here. In what follows, we consider a number of possible reasons for the lack of a reduction in the amplitude of the VESPA P1.

One explanation concerns the subcortical source of the scalp VESPA and that of the P1 deficit in the VEP. The human visual system consists of discrete subcortical magnocellular and parvocellular pathways that project preferentially to dorsal and ventral cortical streams. In our previous studies, we have consistently posited a magnocellular basis for the observed VEP P1 deficits (e.g. Foxe et al., 2005; Butler et al., 2005, 2007; see also Kim et al., 2005a). The VESPA scalp maps of Fig. 3 suggest possible preferential stimulation of the dorsal stream and thus of magnocellular pathways. Therefore, the lack of a difference in the VESPA between controls and patients suggests that either there is no dysfunction of the magnocellular system in schizophrenia or that the VESPA does not actually reflect activity of the magnocellular system.

Magnocellular and parvocellular cells differ not only anatomically, but also functionally, in terms of preferred stimuli. Parvocellular cells with their spectrally opponent nature are know to be less sensitive to luminance contrast than magnocellular cells (Kaplan et al., 1990; Lee et al., 1990). While, the high contrast gain of cells in the magnocellular pathway might suggest that they may be more sensitive to the contrast modulations of the VESPA stimulus, their response saturates at fairly low contrasts (10–15%; e.g., Baseler and Sutter 1997). Parvocellular neurons, meanwhile, have lower contrast gain, but do not saturate (see Butler et al., 2007). Given that the stimulus described in this study spends less than 2% of its time below 15% contrast (Lalor et al., 2006), it seems reasonable to conclude that the VESPA may actually reflect mostly activity of parvocellular pathways.

The two pathways are also known to differ in their response characteristics to the temporal frequencies of stimuli. The commonly held belief is that magnocellular cells are more suited to high temporal frequency flicker (e.g., Kaplan and Benardete, 2001). As a result, it could again be concluded that the rapid modulation of the VESPA might preferentially activate that subsystem. However, both parvocellular and magnocellular cells in the lateral geniculate nucleus (LGN) of the macaque have been reported to respond best at temporal frequencies in the range of 10–20 Hz (Hicks et al., 1983). More specifically, Derrington and Lennie (1984) found that parvocellular units were most sensitive to stimuli modulated at temporal frequencies close to 10 Hz and magnocellular units to stimuli modulated at frequencies nearer 20 Hz. They also reported that the loss of sensitivity as temporal frequency fell below optimum was more marked in magnocellular than parvocellular units. These findings suggest that, while magnocellular cells are known to have a shorter latency and more transient response to stimuli (Marrocco et al., 1982; Maunsell et al., 1999), parvocellular cells should have no difficulty in following the 0–30 Hz frequency content of the VESPA stimulus.

A further property that differs between the two subsystems is the linearity of their temporal response. While the parvocellular system is approximately linear, the temporal responses of magnocellular cells are particularly nonlinear due to contrast gain control

(Kaplan and Benardete, 2001). The nonlinear nature of this system has been referred to in a recent study, which investigated early-stage visual processing deficits in patients with schizophrenia using the steady-state VEP (SSVEP; Kim et al., 2005b). This study has shown reductions in the second harmonics of the stimulus frequency. Given that second harmonics are thought to depend preferentially on magnocellular pathways, the reduced harmonics are attributed to deficits in those pathways. However, it was also pointed out that deficits in nonlinear mechanisms present in cortex, which are important in producing responses at higher harmonics or temporal frequencies, would also result in greater attenuation of higher harmonic responses in patients than controls. For these reasons, it is clear that a linear VESPA simply may not be sensitive to the nonlinear systems responsible for the generation of the P1 deficit in the transient VEP. In order to address this, we have extended the VESPA method to a quadratic analysis in this paper. The fact that no significant differences were found between patients and controls for the quadratic VESPA lends further support to the notion that, if indeed magnocellular dysfunction underlies the P1 deficit in the VEP, the stimulus used in this study was biased toward linear parvocellular cell populations.

Another potential reason for the dramatic dissociation between VEP and VESPA results stems from the debate over whether the ERP in response to a stimulus constitutes an evoked event or comes about through induced changes in ongoing brain dynamics. While most ERP studies assume the former, some studies have suggested that the ERP at least partly arises from the stimulus-induced phase-resetting of electrophysiological processes (e.g., Makeig et al., 2002; Hanslmayr et al., 2007). While the VESPA does not rule out the notion of an induced contribution to VEPs obtained using discrete stimuli, its continuous nature, which does not allow for any time-locked lower frequency phase-resetting of ongoing brain dynamics, clearly demonstrates that ERPs can be evoked. This leads to a confound in the comparison between VEP and VESPA in that it is at least possible that the reduced VEP P1 components displayed by the patients reflect dysfunction of phase-resetting processes or ongoing oscillatory activity. In support of this notion, one recent study proposed alpha band activity as the likely source of an early induced ERP contribution (Hanslmayr et al., 2007) while various characteristics of alpha oscillations have been shown to differ in patients with schizophrenia, including lower peak frequency (Javitt, 1997) and lower power (Sponheim et al., 1994). The inconclusive (and sometimes contradictory) nature of studies attempting to evaluate phase-resetting and the demonstration of the purely evoked VESPA ERP itself lend support to studies positing a predominant role for stimulus-evoked activity in sensory ERP generation (e.g., Shah et al., 2004). Therefore, attributing a divergence in results as dramatic as reported in this paper to deficiency in induced ERP generation seems, at best, speculative.

In summary, we have demonstrated a striking disparity in relative ERP responses between patients and controls using two different methods of visual stimulation. This points to the highly specific nature of early visual deficits in schizophrenia and speaks particularly to the notion that those deficits are based substantially on magnocellular stream dysfunction where activity of the parvocellular system is largely spared. While the VESPA as implemented in this study was not sensitive to the mechanisms responsible for a reduced P1 component in schizophrenia, the flexibility of the method, in terms of the characteristics of both the stimuli and the modulating signal, suggests its utility as a method for further investigation of those mechanisms.

Appendix A. Extension to quadratic VESPA estimation

As detailed in Lalor et al. (2006), we estimate the linear VESPA as an n-dimensional vector $\mathbf{w}$ consisting of the sampled points of the response function

$$(\mathbf{w}(\tau_0), \mathbf{w}(\tau_1), \ldots, \mathbf{w}(\tau_{n-1}))^T, \tag{1}$$

where n is the number of sampled points of the response function that we wish to estimate. This is done by first forming the n-dimensional vector $\mathbf{x}_t$ consisting of the sampled points of the modulating stimulus

$$(\mathbf{x}(t-t_0), \mathbf{x}(t-(t_0+1)), \ldots, \mathbf{x}(t-(t_0+n-1)))^T, \tag{2}$$

where t_0 is the estimation window offset. The values of $\mathbf{x}_t$ are simply the normalized luminance or contrast values of the displayed stimulus.

We then solve for $\mathbf{w}$ using the equation,

$$\mathbf{w} = \langle \mathbf{x}_t \mathbf{x}_t^T + \lambda M \rangle^{-1} \mathbf{x}_t \mathbf{y}_t \tag{3}$$

where λ is a regularization parameter and M is a near-diagonal matrix.

In this paper, we expand the VESPA estimation to a quadratic model of how the EEG depends on the input stimulus. This is accomplished by replacing Eq. (2) with a vector with $n+n(n+1)/2$ elements, where n is the window size, containing the n 1st-order elements as before, and the $n(n+1)/2$ 2nd-order elements (all products of the form $x(t-t_0-i)x(t-t_0-j)$ where

$0 \le i \le j \le n$). The quadratic VESPA **w**, of this same dimensionality can be solved using,

$$\mathbf{w} = \langle \mathbf{x}_t \mathbf{x}_t^T + \delta I \rangle^{-1} \mathbf{x}_t \mathbf{y}_t \quad (4)$$

where δ is a different regularization parameter and I is the identity matrix. In this study, $\delta = 5 \times 10^{-6}$ gave good reduction in estimation error.

Role of funding source

This work was supported in part by a National Institute of Mental Health (NIMH) Grant MH-65350 to J. J. Foxe, and by a Fund for Digital Research Programme grant from the Higher Educational Authority (HEA) of Ireland to R. B. Reilly. Neither the NIMH nor HEA had any further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Contributors

Dr. Lalor designed the stimulus sequences, programmed all paradigms, analyzed all data and wrote the first draft of the manuscript. Dr. Foxe designed the experimental protocol and edited multiple drafts of the manuscript. Dr. Yeap collected all data and tabulated patient demographics, performed the clinical ratings and aided in analyses. Drs. Reilly and Pearlmutter provided critical input regarding development of the VESPA technique and its implementation and provided comments on early drafts of the manuscript. All authors contributed to and have approved the final manuscript. The principle investigator, Dr. Foxe, takes responsibility for the integrity of the data and the accuracy of the data analysis, and attests that all authors had full access to all the data in the study.

Conflict of interest

Drs. Lalor, Pearlmutter, Reilly and Foxe are listed as inventors on a patent application for the VESPA method. Dr. Yeap declares no conflict of interest.

Acknowledgements

The authors are grateful to Dr. Simon P. Kelly for helpful comments and discussion. The authors would also like to express their sincere gratitude to the Chief Executive Officer at St. Vincent's Hospital, Mr. Edward Byrne and to the Director of Nursing, Mrs. Phil Bourke, for their ongoing and essential support of the Cognitive Neurophysiology Laboratory.

References

Baseler, H.A., Sutter, E.E., 1997. M and P components of the VEP and their visual field distribution. Vis. Res. 37 (6), 675–690.

Butler, P.D., Schechter, I., Zemon, V., Schwartz, S.G., Greenstein, V.C., Gordon, J., Schroeder, C.E., Javitt, D.C., 2001. Dysfunction of early-stage visual processing in schizophrenia. Am. J. Psychiatry 158 (7), 1126–1133.

Butler, P.D., Zemon, V., Schechter, I., Saperstein, A.M., Hoptman, M.J., Lim, K.O., Revheim, N., Silipo, G., Javitt, D.C., 2005. Early-stage visual processing and cortical amplification deficits in schizophrenia. Arch. Gen. Psychiatry 62 (5), 495–504.

Butler, P.D., Hoptman, M.J., Nierenberg, J., Foxe, J.J., Javitt, D.C., Lim, K.O., 2006. Visual white matter integrity in schizophrenia. Am. J. Psychiatry 163 (11), 2011–2013.

Butler, P.D., Martinez, A., Foxe, J.J., Kim, D., Zemon, V., Silipo, G., Mahoney, J., Shpaner, M., Jalbrzikowski, M., Javitt, D.C., 2007. Subcortical visual dysfunction in schizophrenia drives secondary cortical impairments. Brain 130 (Pt 2), 417–430.

Clark, V.P., Hillyard, S.A., 1996. Spatial selective attention affects extrastriate but not striate components of the visual evoked potential. J. Cogn. Neurosci. 8, 387–402.

Derrington, A.M., Lennie, P., 1984. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. J. Physiol. 357, 219–240.

Doniger, G.M., Foxe, J.J., Murray, M.M., Higgins, B.A., Javitt, D.C., 2002. Impaired visual object recognition and dorsal/ventral stream interaction in schizophrenia. Arch. Gen. Psychiatry 59 (11), 1011–1020.

Donohoe, G., Morris, D.W., De Sanctis, P., Magno, E., Montesi, J.L., Garavan, H.P., Robertson, I.H., Javitt, D.C., Gill, M., Corvin, A.P., Foxe, J.J., in press. Early visual processing deficits in dysbindin-associated schizophrenia. Biological Psychiatry. doi:10.1016/j.biopsych.2007.07.022.

Foxe, J.J., Simpson, G.V., 2002. Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. Exp. Brain Res. 142 (1), 139–150.

Foxe, J.J., Doniger, G.M., Javitt, D.C., 2001. Early visual processing deficits in schizophrenia: impaired P1 generation revealed by high-density electrical mapping. NeuroReport 12 (17), 3815–3820.

Foxe, J.J., Murray, M.M., Javitt, D.C., 2005. Filling-in in schizophrenia: a high density electrical mapping and source-analysis investigation of illusory contour processing. Cereb. Cortex 15 (12), 1914–1927.

Gomez-Gonzalez, C.M., Clark, V.P., Fan, S., Luck, S.J., Hillyard, S.A., 1994. Sources of attention-sensitive visual event-related potentials. Brain Topogr. 7 (1), 41–51.

Haenschel, C., Bittner, R.A., Haertling, F., Rotarska-Jagiela, A., Maurer, K., Singer, W., Linden, D.E.J., in press. Impaired early-stage visual processing contributes to working memory dysfunction in adolescents with schizophrenia — a study with event-related potentials and functional magnetic resonance imaging. Arch Gen Psychiatry.

Hanslmayr, S., Klimesch, W., Sauseng, P., Gruber, W., Doppelmayr, M., Freunberger, R., Pecherstorfer, T., Birbaumer, N., 2007. Alpha phase reset contributes to the generation of ERPs. Cereb. Cortex 17 (1), 1–8.

Hicks, T.P., Lee, B.B., Vidyasagar, T.R., 1983. The responses of cells in macaque lateral geniculate nucleus to sinusoidal gratings. J. Physiol. 337, 183–200.

Javitt, D.C., 1997. Psychophysiology of schizophrenia. Curr. Opin. Psychiatry 10 (1), 11–15.

Kaplan, E., Benardete, E., 2001. The dynamics of primate retinal ganglion cells. Prog. Brain Res. 134, 17–34.

Kaplan, E., Lee, B.B., Shapley, R.M., 1990. New views of primate retinal function. In: Osborne, N., Chader, G. (Eds.), Progress in Retinal Research, vol. 9. Pergamon Press, Oxford, pp. 273–336.

Kim, D., Wylie, G., Pasternak, R., Butler, P.D., Javitt, D.C., 2005a. Magnocellular contributions to impaired motion processing in schizophrenia. Schizophr. Res. 82 (1), 1–8.

Kim, D., Zemon, V., Saperstein, A., Butler, P.D., Javitt, D.C., 2005b. Dysfunction of early-stage visual processing in schizophrenia: harmonic analysis. Schizophr. Res. 76 (1), 55–65.

Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G., Foxe, J.J., 2006. The VESPA: a method for the rapid estimation of a visual evoked potential. NeuroImage 32 (4), 1549–1561.

Lee, B.B., Pokorny, J., Smith, V.C., Martin, P.R., Valberg, A., 1990. Luminance and chromatic modulation sensitivity of macaque

ganglion cells and human observers. J. Opt. Soc. Am. A 7 (12), 2223–2236.

Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. Science 295 (5555), 690–694.

Marrocco, R.T., McClurkin, J.W., Young, R.A., 1982. Spatial summation and conduction latency classification of cells of the lateral geniculate nucleus of macaques. J. Neurosci. 2 (9), 1275–1291.

Maunsell, J.H.R., Ghose, G.M., Assad, J.A., McAdams, C.J., Boudreau, C.E., Noerager, B.D., 1999. Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. Vis. Neurosci. 16 (1), 1–14.

Murray, M.M., Wylie, G.R., Higgins, B.A., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. The spatiotemporal dynamics of illusory contour processing: combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. J. Neurosci. 22 (12), 5055–5073.

Murray, M.M., Foxe, D.M., Javitt, D.C., Foxe, J.J., 2004. Setting boundaries: brain dynamics of modal and amodal illusory shape completion in humans. J. Neurosci. 24 (31), 6898–6903.

Murray, M.M., Imber, M.L., Javitt, D.C., Foxe, J.J., 2006. Boundary completion is automatic and dissociable from shape discrimination. J. Neurosci. 26 (46), 12043–12054.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.

Schechter, I., Butler, P.D., Zemon, V.M., Revheim, N., Saperstein, A.M., Jalbrzikowski, M., Pasternak, R., Silipo, G., Javitt, D.C., 2005. Impairments in generation of early-stage transient visual evoked potentials to magno- and parvocellular-selective stimuli in schizophrenia. Clin. Neurophysiol. 116 (9), 2204–2215.

Shah, A.S., Bressler, S.L., Knuth, K.H., Ding, M., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2004. Neural dynamics and the fundamental mechanisms of event-related brain potentials. Cereb. Cortex 14 (5), 476–483.

Spencer, K.M., Nestor, P.G., Niznikiewicz, M.A., Salisbury, D.F., Shenton, M.E., McCarley, R.W., 2003. Abnormal neural synchrony in schizophrenia. J. Neurosci. 23 (19), 7407–7411.

Sponheim, S.R., Clementz, B.A., Iacono, W.G., Beiser, M., 1994. Resting EEG in first-episode and chronic schizophrenia. Psychophysiology 31 (1), 37–43.

Yeap, S., Kelly, S.P., Sehatpour, P., Magno, E., Javitt, D.C., Garavan, H., Thakore, J.H., Foxe, J.J., 2006. Early visual sensory deficits as endophenotypes for schizophrenia: high-density electrical mapping in clinically unaffected first-degree relatives. Arch. Gen. Psychiatry 63 (11), 1180–1188.

EJN fens EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, pp. 1–7, 2012 doi:10.1111/j.1460-9568.2012.08060.x

At what time is the cocktail party? A late locus of selective attention to natural speech

Alan J. Power,[1,2] John J. Foxe,[3,4,5] Emma-Jane Forde,[5] Richard B. Reilly[1,2,3] and Edmund C. Lalor[1,2,3]
[1]School of Engineering, Trinity College Dublin, Dublin 2, Ireland
[2]Trinity Centre for Bioengineering, Trinity College Dublin, Dublin 2, Ireland
[3]Trinity College Institute of Neuroscience, Trinity College Dublin, 152-160 Pearse Street, Dublin 2, Ireland
[4]The Cognitive Neurophysiology Laboratory, Children's Evaluation and Rehabilitation Center, Departments of Pediatrics and Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA
[5]The Cognitive Neurophysiology Laboratory, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA

Keywords: AESPA, EEG, exogenous processing, multi-speaker environments

Abstract

Distinguishing between speakers and focusing attention on one speaker in multi-speaker environments is extremely important in everyday life. Exactly how the brain accomplishes this feat and, in particular, the precise temporal dynamics of this attentional deployment are as yet unknown. A long history of behavioral research using dichotic listening paradigms has debated whether selective attention to speech operates at an early stage of processing based on the physical characteristics of the stimulus or at a later stage during semantic processing. With its poor temporal resolution fMRI has contributed little to the debate, while EEG–ERP paradigms have been hampered by the need to average the EEG in response to discrete stimuli which are superimposed onto ongoing speech. This presents a number of problems, foremost among which is that early attention effects in the form of endogenously generated potentials can be so temporally broad as to mask later attention effects based on the higher level processing of the speech stream. Here we overcome this issue by utilizing the AESPA (auditory evoked spread spectrum analysis) method which allows us to extract temporally detailed responses to two concurrently presented speech streams in natural cocktail-party-like attentional conditions without the need for superimposed probes. We show attentional effects on exogenous stimulus processing in the 200–220 ms range in the left hemisphere. We discuss these effects within the context of research on auditory scene analysis and in terms of a flexible locus of attention that can be deployed at a particular processing stage depending on the task.

Introduction

Distinguishing between multiple speakers and focusing attention on a single speaker is integral to human communication. Our behavioral capacity to do this was first examined by Cherry (1953). Since then much debate has revolved around the issue of whether attention to speech operates at an early stage of processing based on the physical characteristics of the stimulus or at a later stage during semantic processing (Broadbent, 1958; Moray, 1959; Deutsch & Deutsch, 1963; Johnston & Wilson, 1980). In particular, while it has been shown that unattended speech is processed semantically even when it is not available to conscious recollection (Lewis, 1970; Bentin *et al.*, 1995), this semantic processing does not always take place (Treisman *et al.*, 1974). Thus, it seems plausible that selective attention to speech may operate as (at least) a two-stage process (Treisman, 1960, 1964).

Behavioral and electroencephalography (EEG) studies using non-speech stimuli have made important contributions to this debate by demonstrating that the temporal locus of selective attention is sensitive to task demands (Yantis & Johnston, 1990; Lavie *et al.*, 2004; Vogel *et al.*, 2005). These include event-related potential (ERP) studies of attention effects on auditory stream segregation which often use stimuli containing multiple harmonics or patterns of low and high tones to form multiple concurrent streams (e.g., Alain *et al.*, 2001; Snyder *et al.*, 2006; Sussman & Steinschneider, 2009). In fact several of these studies have pointed to distinct cortical mechanisms: a bottom-up automatic segregation of sounds based on their physical characteristics, and a top-down attention-dependent process that occurs at a later stage (Alain *et al.*, 2001; Snyder *et al.*, 2006). However, where the naturalistic deployment of attention to speech is concerned, the ability of the ERP technique to disentangle the effects of several putative attentional processes is hampered by the need to average the EEG in response to discrete events. Researchers have sought to address this by superimposing task-irrelevant probes on continuous speech (Hink & Hillyard, 1976; Woods *et al.*, 1984; Coch *et al.*, 2005; Nager *et al.*, 2008) or by averaging EEG around 'discrete' speech features such as hard consonants (Teder *et al.*, 1993). These studies have provided unequivocal evidence for early attention

Correspondence: Dr E. C. Lalor, [3]Trinity College Institute of Neuroscience, as above.
E-mail: edlalor@tcd.ie.

Received 26 October 2011, revised 30 December 2011, accepted 2 February 2012

effects in the form of a broad attentional potential known as the processing negativity (PN) or negative deflection wave (Nd) which onsets as early as 50–60 ms post-stimulus and can persist until up to 1000 ms (Näätänen, 1982). However, the temporally broad nature of the Nd, which suggests it is an endogenously generated process distinct from the exogenously generated componentry typical of the ERP, means that it overlaps the ERP in time. This complicates the use of the ERP's exquisite temporal resolution for precisely determining the temporal loci of attention effects on the exogenously driven processing of continuous speech.

Here we aim to overcome this complication using the AESPA method (auditory evoked spread spectrum analysis; Lalor *et al.*, 2009; Power *et al.*, 2011). The AESPA is an estimate of the impulse response of the auditory system which can be obtained using the amplitude envelope of a wide class of stimuli, including speech (Lalor & Foxe, 2010). This is particularly useful given that the envelope of speech is of overriding importance when it comes to speech recognition (Shannon *et al.*, 1995; Smith *et al.*, 2002). As such, the AESPA can be utilized as an index of the exogenous processing of natural speech and can facilitate the investigation of attentional effects on that processing without being obscured by temporally broad scalp potentials such as the Nd wave.

Materials and methods

Subjects

Forty subjects took part (mean ± standard deviation age, 27.3 ± 3.2 years; 32 male; seven left-handed). Twenty subjects (age 27.5 ± 3.44 years; 18 male; two left-handed) attended to the left ear and 20 subjects attended to the right ear (age 27.24 ± 3.03 years; 14 male; five left-handed). The experiment was undertaken in accordance with the Declaration of Helsinki. The Ethics Committees of the Nathan Kline Institute and the School of Psychology at Trinity College Dublin approved the experimental procedures and each subject provided written informed consent. Subjects reported no history of hearing impairment or neurological disorder.

EEG acquisition

Electroencephalography data were recorded for 34 of the subjects using 130 electrode positions (17 of these subjects attended to the left and the remaining 17 to the right). Data for the remaining six participants were collected using 162 electrode positions (three of these subjects attended to the left and the remaining three to the right). The data were filtered over the range 0–134 Hz and digitized at the rate of 512 Hz using a BioSemi Active Two system. EEG data were then digitally filtered off-line with a band-pass filter between 2 and 35 Hz. The data at each channel were re-referenced to the average of the responses at the left and right mastoids. Responses extracted from the data acquired using the 162-electrode system were mapped down to the same 130 electrode positions used for all other subjects using a spline interpolation algorithm (EEGLAB; http://sccn.ucsd.edu/eeglab/) resulting in a coherent dataset with identical channel configuration.

Stimuli and procedure

Two classic works of fiction were presented, one to the left ear and the other to the right ear. These works were segmented into 30 passages each approximately 1 min in length. Further to this, and in order to minimize the possibility of the unattended stream capturing the subjects' attention during silent periods in the attended stream, silent gaps exceeding 0.5 s were truncated to 0.5 s in duration. Subjects were divided into two groups of 20 with each group being instructed to attend to the story in either the left or right ear throughout all 30 passages (i.e. approximately 1800 s of data per subject). After each passage subjects were required to answer between four and six multiple choice questions on both stories (i.e. on the attended story and the unattended story). The questions had four possible answers. Each passage took up from where the previous passage left off in the story and stimulus amplitudes in each stream within each run were normalized to have the same root mean squared (RMS) intensity. We used a between-subjects design as we wanted each subject to follow just one story to make the experiment as natural as possible and because we wished to avoid any repeated presentation of stimuli. Figure 1 shows the experimental procedure.

AESPA estimation

We obtain the AESPA by performing a linear least-squares fit of the response model

$$y(t) = w(\tau) * x(t) + \text{noise}$$

where $y(t)$ is the measured EEG response, $x(t)$ is the amplitude envelope of the stimulus (see below), the symbol * indicates convolution, $w(\tau)$ is the impulse-response function to the amplitude of the stimulus, and the noise is assumed to be Gaussian (Lalor *et al.*, 2009). This impulse response function, known as the AESPA, is not equivalent to a standard ERP but shares a number of properties including detailed temporal precision (Lalor *et al.*, 2009) and sensitivity to attentional modulation (Power *et al.*, 2011).

Summarizing the method in qualitative terms, the AESPA response $w(\tau)$ is analogous to a filter which describes how the brain transforms the auditory input into the EEG output. Keeping this in mind, the time axis for the AESPA carries a different meaning than the time axes in

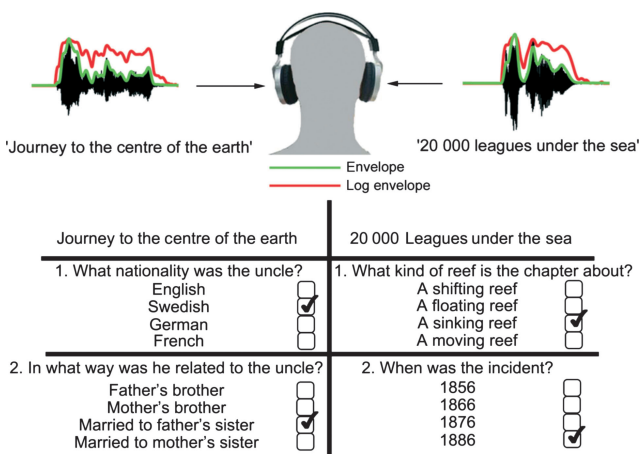

FIG. 1. The experimental procedure. In each trial, subjects listened to approximately 60 s of two stories presented concurrently and dichotically. After each 60-s trial, subjects were presented with between four and six multiple-choice questions on each story, with each question having four possible answers. Subjects were asked to preferentially attend to one story, but to attempt to answer all questions on both stories. Answers were indicated on screen using a mouse click. Each 60-s trial continued from the point in the stories at which the previous trial ended, with no repetition of stimuli. EEG data were not collected while subjects answered the questions. Responses were extracted using the log envelope of the speech stimuli.

traditional ERP studies. Each point on the time axis can be interpreted as being the relative time between the continuous EEG and the continuous input intensity signal. Therefore the AESPA at −100 ms, for example, indexes the relationship between the input intensity signal and the EEG 100 ms earlier; obviously this should be zero. As another example, the AESPA at +100 ms indexes how the input intensity signal affects the EEG 100 ms later (Lalor *et al.*, 2009). Furthermore, any activity that explicitly follows the stimulus envelope (including possible endogenous fluctuations) would be represented in the AESPA response. That said, however, slow endogenous potentials (such as the attention-related processing negativity PN/Nd) will not contribute to the AESPA.

In order to estimate the AESPA it was necessary to determine the amplitude envelope of the speech signals and to resample the speech signal to the same sampling rate as the EEG. To do this, we simply calculated the RMS value of an average of 83.133 neighboring audio samples around every EEG sample. Further to this, as envelope frequencies between 2 and 16 Hz contribute most to speech intelligibility (Drullman *et al.*, 1994a,b; van der Horst *et al.*, 1999), the envelope was then low-pass filtered with a corner frequency of 20 Hz. Because intensity processing varies by the log of stimulus intensity, each envelope was transformed by taking the log to the base 10 of the intensity and normalizing between zero and the maximum before mapping. A similar approach was employed by Aiken & Picton (2008) when investigating speech envelope processing in a passive paradigm. As our model assumes a linear relationship between the EEG and stimulus envelope the EEG was also low-pass filtered with a corner frequency of 20 Hz.

EEG analysis

To test for statistical differences we subjected the responses to statistical parametric mapping (SPM; Kiebel & Friston, 2004a,b). When using SPM for EEG the EEG signals are mapped from the electrode domain to a flattened two-dimensional scalp space, consisting of a 64 × 64 pixel grid, by way of a linear interpolation. This mapping is done at every time point of the ERP and results in a three-dimensional representation of the data (two dimensions in space and one in time). This representation consists of what is called a voxel-based representation of the EEG but the representations are still of the scalp activity and no underlying source configuration has been derived. Thus a voxel in this case refers to a position on the scalp (in 2-D space) at a particular time. We tested this representation for temporally and regionally specific effects using separate factorial ANOVAs for each story, with each ANOVA having two levels (attended vs. unattended). We controlled for multiple spatiotemporal comparisons using the family-wise error rate based on a significance level of 0.05.

Results

Behavioral results

The results of a 2 × 2 ANOVA with levels of story (left ear/right ear) and attention (attended/unattended) stream found a significant main effect of attention ($F = 1164.13$, $P < 0.001$), no effect of story ($F = 3.08$, $P = 0.084$) and no story × attention interaction ($F = 2.15$, $P = 0.147$). On average, subjects correctly answered 80.4 ± 7.3% of questions on the attended story and 27.1 ± 7.0% on the unattended story, which, consistent with previous reports on dichotic listening behavior, was not statistically greater than chance ($P = 0.77$).

EEG results

The above-mentioned behavioral effects were accompanied by clear differences in the AESPA responses to attended vs. unattended speech streams (Fig. 2). Specifically, attention had obviously affected the response to the left ear story in the timeframe of the positive component, approximately 195–230 ms. Similarly the right ear story showed differences in the same range, particularly over the left hemisphere. Importantly, these attention effects appear large relative to the SEM which is important given that this is a between-subjects design. Another noteworthy feature of Fig. 2 is that the responses in the 90- to 200-ms range also appear as if they may have been affected by attention for the right ear story only. Because this is a between-subjects design it is important not to read too much into the responses from one story in isolation, and we note again that the common feature of the two sets of responses is a difference in the amplitude of the positive component at approximately 200 ms.

In order to statistically test the responses for attentional effects, we employed SPM as mentioned above. Given the possible attention effect in both stories in the 195- to 230-ms range (Fig. 2), and the possibility that the right ear response was also affected from 90 to 195 ms, we included the data from across the entire 90- to 230-ms range in our SPM analysis. This analysis highlighted that the attention effects were located over the left hemisphere during the timeframe of the positive component at approximately 200 ms (left ear story, peak effect at 213 ms, $F_{1,38} = 33.51$, $P = 0.002$; right ear story, peak effect at 207 ms, $F_{1,38} = 22.47$, $P = 0.03$; see Fig. 3A). In addition there was a contralateral attentional effect on the response to the story in the left ear during the same timeframe (peak effect at 213 ms, $F_{1,38} = 27.65$, $P = 0.007$). Using family-wise error correction, we found no statistically significant effects on any of the earlier components.

Given the identification of an attention effect in the 195- to 230-ms range and the obvious inability of subjects to recall any of the content of the unattended stream, we wished to assess whether our attention effect may have been driven by suppression of the unattended stream at approximately 200 ms. As previous studies have clearly shown the existence of a positive component at approximately 200 ms in AESPA responses obtained during the non-selective processing of speech (Lalor & Foxe, 2010), we hypothesized that if the same positive component to the unattended speech stream in this study were not significantly greater than baseline, then that may be indicative of suppression. We tested this by comparing the RMS response power on the scalp during the interval 195–230 ms for both the attended and unattended stories with the RMS power from the baseline interval from −100 to 0 ms. Separate paired *t*-tests on this RMS measure did not find any statistical differences for either story when unattended (left ear story, $P = 0.36$; right ear story, $P = 0.07$), indicating that neural activity related to the unattended story was suppressed to baseline during this timeframe. Unsurprisingly, RMS power was different from baseline when attended (left ear story, $P << 0.01$; right ear story, $P = 0.01$).

Discussion

Our behavioral results clearly demonstrate that our paradigm, which used only continuous natural speech streams, was extremely effective in engaging subjects' attention. Subjects answered significantly more questions related to the attended story than the unattended story, with performance being at the level of chance for the unattended stories; this is consistent with previous reports on dichotic listening behavior (Cherry, 1953).

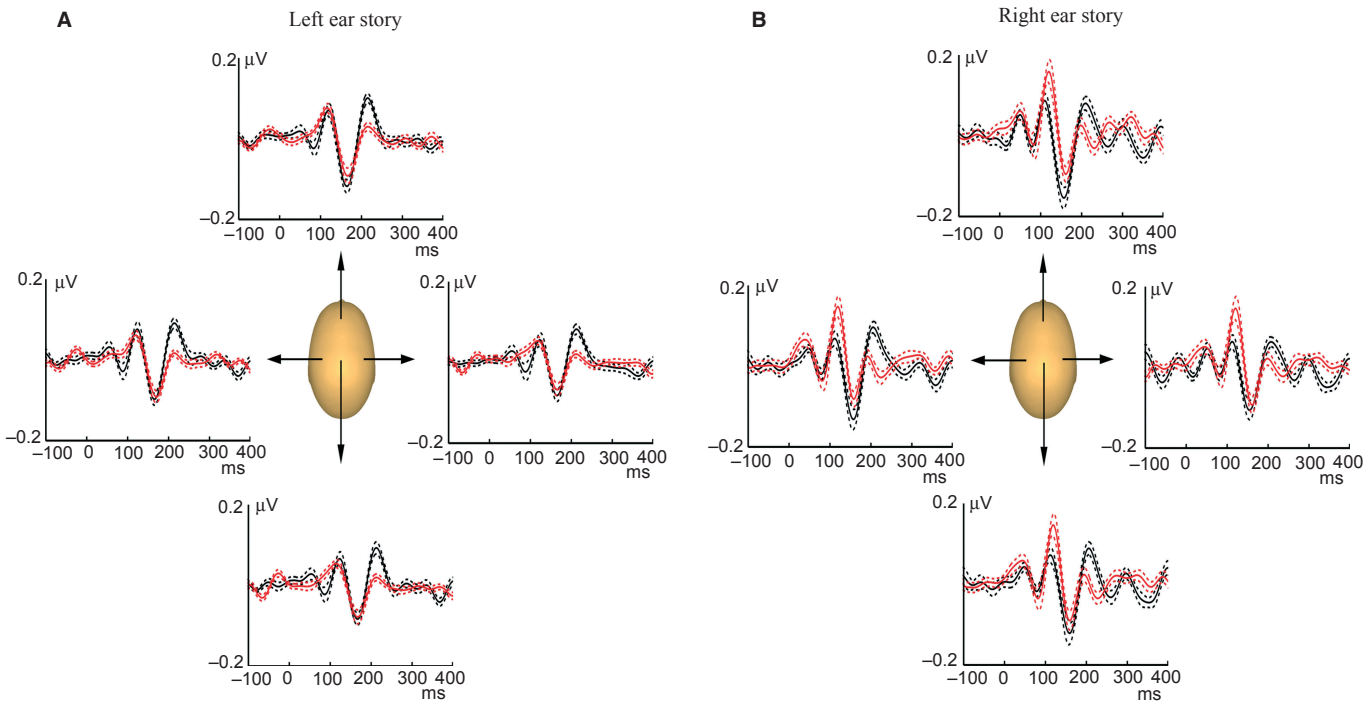

FIG. 2. AESPA responses at four representative electrodes for (A) the story presented to the left ear and (B) the story presented to the right ear when attended (black traces) and unattended (red traces). The dashed lines around the solid are plots of the SEM across subjects.

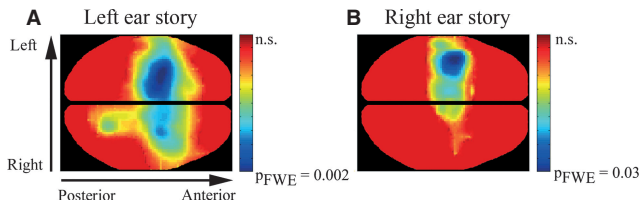

FIG. 3. The effect of attention on the scalp recordings. Spatial distribution of the effect on the positive AESPA component in the interval 195–230 ms identified by the SPM analysis for (A) the left and (B) the right stories. pFWE is the significance level when family-wise error (FWE) is taken into account.

In parallel with these behavioural effects, our AESPA responses show a robust, left-lateralized attention effect peaking at approximately 209 ms for both stories. This finding stands in contrast to a large number of dichotic speech studies using the average ERP method, which have typically shown much earlier effects, often over frontocentral locations (e.g., Hink & Hillyard, 1976; Woods *et al.*, 1984; Teder *et al.*, 1993; Coch *et al.*, 2005; Nager *et al.*, 2008). As mentioned above, one possible reason for this is that the average ERP method may result in temporally overlapping attention effects: those that impact on exogenously driven processing and those that are endogenously generated. For example, the endogenously generated Nd component has been shown to onset as early as approximately 50 ms with a frontocentral distribution, and to persist up to 1000 ms (e.g., Näätänen, 1982; Woods, 1990; Teder *et al.*, 1993). This component has been divided into two distinct phases: an early phase that is thought to index a flexible mechanism of rough selection between sounds based on their feature-based discriminability (Hansen & Hillyard, 1980), and a less well understood later phase that is thought to be related to the maintenance of an attentional trace (Näätänen, 1982; Woods, 1990). Neither of these phases shows any lateralization (Woods *et al.*, 1984; Woods, 1990). Furthermore, in studies where the ERP is derived in response to probe stimuli superimposed on speech, the extended temporal duration of the late phase has led to the suggestion that it actually relates to how relevant the superimposed probes are to the speech stream and not to the processing of the speech itself (Woods *et al.*, 1984). An important point here is that this endogenously generated component has been implicated in attentional modulation of the exogenously generated N1 component of the auditory evoked potential (AEP; Näätänen, 1982). Specifically, it has long been debated whether the N1 component of the AEP was actually enhanced by attention, or whether attentional effects during the N1 timeframe were actually a result of an overlapping Nd wave (Näätänen, 1982; Woods, 1990). The AESPA method recently contributed to this debate by showing that obligatory sensory processing was enhanced by attention during the timeframe of the N1 and that the entire attention effect was unlikely to be accounted for by just the Nd (Power *et al.*, 2011). In the present study we show no attention effects during the timeframe of the N1 (or earlier), perhaps indicative of our success in removing the influence of the Nd wave.

As such, the question that arises is: to what attentional mechanism does our late effect relate? One tentative proposal is that, unlike the early effects that probably relate to the physical characteristics of the stimulus (e.g., pitch, spatial location), our late effect may represent a filtering process located at the level of semantic analysis (Treisman, 1964; Lewis, 1970, Treisman *et al.*, 1974). This is consistent with previous reports on the timing of the onset of semantic processing (Helenius *et al.*, 2002; Bonte *et al.*, 2006; Salmelin, 2007) and with reports on left hemispheric specialization for the processing of semantic and linguistic information (e.g., Wernicke, 1874; Binder *et al.*, 1997). Previous studies using the AESPA have shown the positive component peaking at ~200 ms (known as the Pd component, see Lalor *et al.*, 2009) to be robust in the non-selective processing of speech (Lalor & Foxe, 2010). Accordingly, the fact that

Pd response power to the unattended speech stimuli in the present study was found to be not significantly greater than baseline noise power suggests suppression of irrelevant information at this stage. As previously mentioned, the earlier negative component of the AESPA (known as the Nc component), which has been shown to be sensitive to attentional modulation (Power *et al.*, 2011), did not show any robust or consistent attention effects under the present task for either the left or right ear story. Therefore, our result may be indicative of a task-related flexible temporal locus of selective attention (Lavie *et al.*, 2004), deployed here, given the higher level task, in order to maximize the ability to process attended speech.

Given previous reports of the semantic processing of unattended speech without the ability to consciously identify that speech (Lewis, 1970; Holender, 1986), the locus of the left hemispheric Pd suppression effect observed in our data may relate to the prevention of memory trace formation for the semantic information in the unattended speech stream. This accords reasonably well with EEG and magnetoencephalography data suggesting that semantic memory use during language comprehension can be indexed by the so-called N400 component (Kutas & Hillyard, 1980; Kutas & Federmeier, 2000), which has been purported to onset as early as 200–250 ms in the posterior half of the left superior temporal gyrus (Kutas & Federmeier, 2011). Furthermore, it has been shown in a dichotic listening word list task that, while both attended and unattended words are semantically processed and activate semantic representations, the N400 elicited by unattended words is insensitive to semantic manipulation (Bentin *et al.*, 1995). The attentional suppression of our Pd component at 200–220 ms, combined with the general lack of (or at least severe reduction in; see Kutas & Federmeier, 2011) any semantic manipulation effects on the N400 to unattended stimuli, may suggest a specific temporal locus of attentional suppression before which semantic processing occurs and after which semantic information would otherwise be encoded into working memory. Based on the differences between the current data and the previous AESPA attention study using non-speech stimuli (Power *et al.*, 2011), this tentative hypothesis appears at least somewhat plausible. However, further work needs to be done using controlled manipulation of the stimuli and task within one group of subjects in order to confirm this claim over other possible explanations.

These other possible explanations include a variety of attentional mechanisms that are likely to play a part in the observed effects. For example, low-level features such as space and frequency cannot be entirely ruled out. In terms of space, it is worth noting that the topographic distribution of our attentional effect depended somewhat on which ear was attended. However, given that the timing and lateralisation of our effects are so distinct from those early, non-lateralized effects typically reported for auditory spatial attention (Woods, 1990; Power *et al.*, 2011), we suggest that the role played by space is relatively minor. The notion that frequency might play a role arises as a result of the fact that, though the two speakers were male, their voices were clearly distinguishable on the basis of pitch. Attention to low-level features such as pitch, however, is often indexed by early ERP effects, including the Nd wave (Hansen & Hillyard, 1980), and thus it is unlikely to have played a dominant role in our effects.

Another potentially important factor in our results is the possible role of auditory stream segregation, i.e., the separation of the two speech streams into distinct auditory objects (Bregman, 1990). The perception of a separation of concurrent non-speech auditory stimuli into two distinct streams has previously been shown to be indexed by a number of ERP components, including the P1, N1 and P2 components (e.g., Gutschalk *et al.*, 2005; Snyder *et al.*, 2006), in addition to a negative component at approximately 160 ms (Snyder *et al.*, 2006) and a positive component in the 350- to 450-ms range (Alain *et al.*, 2001). Furthermore, a number of these segregation indices have been shown to be affected by attention (Alain *et al.*, 2001), including some during the timeframe of our AESPA results (Snyder *et al.*, 2006). In contrast to our results, however, these attentionally sensitive segregation effects are typically distributed over frontocentral and central scalp regions (Alain *et al.*, 2001; Snyder *et al.*, 2006), with some components being biased to the right hemisphere (Snyder *et al.*, 2006). While sound segregation undoubtedly took place in our study, the implications of these previous findings for our left hemisphere results are unclear. One obvious possible reason for the discrepancy is that our stimuli were speech streams, which are preferentially processed on the left (Wernicke, 1874). On that note, it has been shown that successful segregation of brief speech stimuli leads to enhanced activity in the left thalamus, Heschl's gyrus, the superior temporal gyrus and the planum temporale (Alain *et al.*, 2005), which would be more in line with our data. This issue of attention and how it interacts with stream segmentation is a complex one and it is difficult for us to comment on it further given our current data. As a final observation on this topic, it is worth noting that our experiment was designed to emphasize naturalness rather than to explicitly quantify the contributions of the various attentional processes at play in cocktail party listening. The naturalness of the experiment, demonstrated by the behavioural performance, undoubtedly involves the recruitment of the many attentional mechanisms that have evolved to maximize performance in multispeaker environments. While we contend that attention to low-level features is likely to have made a relatively minor contribution to our AESPA effects, the quantification of the interactions between attention, stream segregation and semantic processing will need to be investigated using a series of more controlled experiments. The AESPA method offers a very promising avenue for this future work.

A number of other recent studies on speech processing and the cocktail party paradigm have employed EEG methods other than the averaged ERP, including those examining the neural tracking of speech (Luo & Poeppel, 2007; Kerlin *et al.*, 2010; Ding & Simon, 2012). Kerlin *et al.* (2010) used a template matching algorithm on ERP N1-derived source waveforms and found gain control affecting areas in and around Heschl's gyrus. Because of the fact that they fit their source waveforms based on the N1, their results are likely to be biased to an earlier stage of processing than our later attention effects. A more direct comparison with our results might be possible by adapting their approach in order to bias a later processing stage, or even to discriminate multiple processing stages. Luo & Poeppel (2007) show that the phase pattern of theta band (4–8 Hz) activity tracks and can discriminate speech in auditory cortex. The authors suggest that the measured theta activity may reflect the interaction between auditory core and belt areas as well as possible contributions from para-belt areas. They also show that the discrimination ability of the phase tracking is correlated with speech intelligibility, which suggests that the method may have utility for examining higher order semantic processes. While these studies have been of great importance to the understanding of the mechanisms involved in speech processing and attentional selection in the cocktail party problem, it is arguably more difficult to determine what processing stage(s) contribute to their findings than is the case with our highly temporally resolved AESPA response.

Having said that, the advantages of the AESPA method come at a price. As already mentioned, the method is insensitive to cortical activity unrelated to the stimulus envelope. In addition, in its current form the AESPA simplistically assumes a linear relationship between the stimulus envelope and the EEG response. Given that information processing in the brain is conducted in a network-based manner

incorporating feedforward, feedback and recurrent activity, it is clear that a linear feedforward assumption would render the AESPA an incomplete measure. For example, the modulation of early afferent activity by efferent activity from higher order areas is likely to be a highly non-linear process that would not be well characterized by our method. Even so, on the basis of the data presented here we argue that the AESPA method has allowed us to determine an important temporal locus of selective attention under natural cocktail-party-like conditions. Further work remains to be done to quantify the contributions of low-level feature-based attention and of stream segregation to the AESPA effects. In addition we will aim to further evaluate the tentative hypothesis that our results point to a temporal locus for the suppression of irrelevant semantic information and the disruption of the encoding of this information into working memory.

Acknowledgements

This study was supported by a grant from the United States National Science Foundation to J.J.F. (BCS0642584) and a grant from Science Foundation Ireland to R.B.R. (09-RFP-NES2382). A.J.P. and E.C.L. were supported in part by the Irish Research Council for Science, Engineering & Technology. We thank Dr Robert Whelan for assistance with the statistical analysis.

Conflict of interest

None.

Abbreviations

AESPA, auditory evoked spread spectrum analysis; EEG, electroencephalography; ERP, event-related potential; Nd, negative difference wave; PN, processing negativity; RMS, root mean squared; SPM, statistical parametric mapping.

References

Aiken, S.J. & Picton, T.W. (2008) Human cortical responses to the speech envelope. *Ear Hear.*, **29**, 139–157.

Alain, C., Arnott, S.R. & Picton, T.W. (2001) Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol. Hum. Percept. Perform.*, **27**, 1072–1089.

Alain, C., Reinke, K., McDonald, K.L., Chau, W., Tam, F., Pacurar, A. & Graham, S. (2005) Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage*, **26**, 592–599.

Bentin, S., Kutas, M. & Hillyard, S.A. (1995) Semantic processing and memory for attended and unattended words in dichotic listening: behavioral and electrophysiological evidence. *J. Exp. Psychol. Hum. Percept. Perform.*, **21**, 54–67.

Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M. & Prieto, T. (1997) Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.*, **17**, 353–362.

Bonte, M., Parviainen, T., Hytönen, K. & Salmelin, R. (2006) Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cereb. Cortex*, **16**, 115–123.

Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge.

Broadbent, D. (1958) *Perception and Communication*. Pergamon Press, London, UK.

Cherry, E.C. (1953) Some experiments on the recognition of speech, with one and two ears. *J. Acoust. Soc. Am.*, **25**, 975–979.

Coch, D., Snaders, L.D. & Neville, H.J. (2005) An event-related potential study of selective auditory attention in children and adults. *J. Cog. Neuro.*, **17**, 605–622.

Deutsch, R.P. & Deutsch, D. (1963) Attention: some theoretical considerations. *Psychol. Rev.*, **70**, 80–90.

Ding, N. & Simon, J.Z. (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.*, **107**, 78–89.

Drullman, R., Festen, J.M. & Plomp, R. (1994a) Effects of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.*, **95**, 1053–1064.

Drullman, R., Festen, J.M. & Plomp, R. (1994b) Effects of reducing slow temporal modulation on speech reception. *J. Acoust. Soc. Am.*, **95**, 2670–2680.

Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M. & Oxenham, A.J. (2005) Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.*, **25**, 5382–5388.

Hansen, J.C. & Hillyard, S.A. (1980) Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr. Clin. Neurophysiol.*, **49**, 277–290.

Helenius, P., Salmelin, R., Service, E., Connolly, J.F., Leinonen, S. & Lyytinen, H. (2002) Cortical activation during spoken-word segmentation in nonreadingimpaired and dyslexic adults. *J. Neurosci.*, **22**, 2936–2944.

Hink, R.F. & Hillyard, S.A. (1976) Auditory evoked potentials during selective listening to dichotic speech messages. *Percep. & Psychophys.*, **20**, 236–242.

Holender, D. (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. *Behav. Brain. Sci.*, **9**, 1–66.

van der Horst, R., Leeuw, A.R. & Dreschler, W.A. (1999) Importance of temporal-envelope cues in consonant recognition. *J. Acoust. Soc. Am.*, **105**, 1801–1809.

Johnston, W.A. & Wilson, J. (1980) Perceptual processing of nontargets in an attention task. *Mem. Cognit.*, **8**, 372–377.

Kerlin, J.R., Shahin, A.J. & Miller, L.M. (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J. Neurosci.*, **30**, 620–628.

Kiebel, S.J. & Friston, K.J. (2004a) Statistical parametric mapping for event-related potentials I: generic considerations. *NeuroImage*, **22**, 492–502.

Kiebel, S.J. & Friston, K.J. (2004b) Statistical parametric mapping for event-related potentials II: a hierarchical temporal model. *NeuroImage*, **22**, 503–520.

Kutas, M. & Federmeier, K.D. (2000) Electrophysiology reveals semantic memory use in language comprehension. *Trends. Cogn. Sci.*, **4**, 463–470.

Kutas, M. & Federmeier, K.D. (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.*, **62**, 621–647.

Kutas, M. & Hillyard, S.A. (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, **207**, 203–205.

Lalor, E.C. & Foxe, J.J. (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.*, **31**, 189–193.

Lalor, E.C., Power, A.J., Reilly, R.B. & Foxe, J.J. (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.*, **102**, 349–359.

Lavie, N., Hirst, A., de Fockert, J. & Viding, E. (2004) Load theory of selective attention and cognitive control. *J. Exp. Psych.*, **133**, 339–354.

Lewis, J.L. (1970) Semantic processing of unattended messages using dichotic listening. *J. Exp. Psych.*, **85**, 225–228.

Luo, H. & Poeppel, D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, **54**, 1001–1010.

Moray, N. (1959) Attention in dichotic listening: affective cues and the influence of instructions. *Q. J. Exp. Psychology.*, **11**, 56–60.

Näätänen, R. (1982) Processing Negativity: an evoked-potential reflection of selective attention. *Psych. Bull.*, **92**, 605–640.

Nager, W., Dethlefsen, C. & Munte, T.F. (2008) Attention to human speakers in a virtual auditory environment: brain potential evidence. *Brain Res.*, **1220**, 164–170.

Power, A.J., Lalor, E.C. & Reilly, R.B. (2011) Endogenous auditory spatial attention modulates obligatory sensory activity in auditory cortex. *Cereb. Cortex*, **21**, 1223–1230.

Salmelin, R. (2007) Clinical neurophysiology of language: the MEG approach. *Clin. Neurophysiol.*, **118**, 237–254.

Shannon, R.V., Fan-Gang, Z., Kamath, V., Wygonski, J. & Ekelid, M. (1995) Speech recognition with primarily temporal cues. *Science*, **207**, 303–304.

Smith, Z.M., Delgutte, B. & Oxenham, A.J. (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, **416**, 87–90.

Snyder, J.S., Alain, C. & Picton, T.W. (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.*, **18**, 1–13.

Sussman, E. & Steinschneider, M. (2009) Attention effects on auditory scene analysis in children. *Neuropsychologia*, **47**, 771–785.

Teder, W., Kujala, T. & Näätänen, R. (1993) Selection of speech messages in free-field listening. *NeuroReport*, **5**, 307–309.

Treisman, A.M. (1960) Contextual cues in selective listening. *Q. J. Exp. Psych.*, **12**, 242–248.
Treisman, A.M. (1964) Verbal cues, language, and meaning in selective attention. *Am. J. Psychol.*, **77**, 206–219.
Treisman, A., Squire, R. & Green, J. (1974) Semantic processing in dichotic listening? *A replication. Mem. Cognit.*, **2**, 641–646.
Vogel, E.K., Woodman, G.F. & Luck, S.J. (2005) Pushing around the locus of selection: evidence for the flexible-selection hypothesis. *J. Cogn. Neurosci.*, **17**, 1907–1922.
Wernicke, C. (1874) *Der aphasische Symptomenkomplex*. Cohn and Weigert, Breslau.
Woods, D.L. (1990) The physiological basis of selective attention: Implications of event-related potential studies. In Rohrbaugh, J.W., Parasuraman, R. & Johnson, R. Jr (Eds), *Event-Related Brain Potentials: Basic Issues and Applications*. Oxford UP, New York, pp. 178–209.
Woods, D.L., Hillyard, S.A. & Hansen, J.C. (1984) Event-related brain potentials reveal similar attentional mechanisms during selective listening and shadowing. *J. Exp. Psych.*, **10**, 761–777.
Yantis, S. & Johnston, J.C. (1990) On the locus of visual selection, evidence from focused attention tasks. *J. Exp. Psychol. Hum. Percept. Perform.*, **16**, 135–149.