Network: Comput. Neural Syst. **12** (2001) 289–316 www.iop.org/Journals/ne PII: S0954-898X(01)23517-9

Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli

Frédéric E Theunissen[1,5], Stephen V David[2], Nandini C Singh[1], Anne Hsu[3], William E Vinje[4] and Jack L Gallant[1]

[1] Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA
[2] Graduate Group in Bioengineering, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA
[3] Department of Physics, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA
[4] Department of Molecular and Cell Biology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, USA

E-mail: fet@socrates.berkeley.edu

Received 4 December 2000

Abstract

We present a generalized reverse correlation technique that can be used to estimate the spatio-temporal receptive fields (STRFs) of sensory neurons from their responses to arbitrary stimuli such as auditory vocalizations or natural visual scenes. The general solution for STRF estimation requires normalization of the stimulus–response cross-correlation by the stimulus auto-correlation matrix. When the second-order stimulus statistics are stationary, normalization involves only the diagonal elements of the Fourier-transformed auto-correlation matrix (the power spectrum). In the non-stationary case normalization requires the entire auto-correlation matrix. We present modelling studies that demonstrate the feasibility and accuracy of this method as well as neurophysiological data comparing STRFs estimated using natural versus synthetic stimulus ensembles. For both auditory and visual neurons, STRFs obtained with these different stimuli are similar, but exhibit systematic differences that may be functionally significant. This method should be useful for determining what aspects of natural signals are represented by sensory neurons and may reveal novel response properties of these neurons.

1. Introduction

There is increasing evidence that sensory systems have evolved to optimally process behaviourally relevant natural stimuli. For example, the auditory systems of many species

[5] Corresponding author.

appear to be tuned for those species' vocalizations. Neurons in bat auditory cortex are selective for echolocation calls (Ohlemiller *et al* 1996, Suga *et al* 1978); neurons in songbirds are selective for the birds' own songs (Margoliash 1986) and some neurons in primate auditory cortex are selective for conspecific vocalizations (Rauschecker *et al* 1995). Furthermore the stimulus–response function of lower level sensory neurons might be matched to the statistics of natural stimuli. For example, neurons in the early auditory system of frogs are more efficient at encoding sounds with natural power spectra (Rieke *et al* 1995). Similarly, the tuning of neurons early in the mammalian visual system appears to be matched to the statistics of natural scenes (Dan *et al* 1996, Field 1987).

One important goal of sensory neurophysiology is to further understand how sensory systems encode natural stimuli. However, natural stimuli are complex and relatively little is currently known about their statistical and mathematical structure. Therefore, in order to describe the stimulus–response functions of sensory neurons, neurophysiologists have until very recently used synthetic stimuli. These stimuli are typically varied along one or two dimensions to estimate a neural tuning curve of theoretical interest. For example, spatial characteristics of the stimulus are varied to obtain the frequency tuning for auditory neurons or orientation and spatial frequency tuning for visual neurons. The responses to dynamic stimuli are studied using similar methods to obtain temporal tuning characteristics (e.g. amplitude modulation tuning in auditory neurons, motion tuning in visual neurons). When spatial and temporal tuning are considered simultaneously, one obtains a spatio-temporal receptive field (STRF). The STRF describes the linear transformation between a time varying spatial stimulus and the neural response. STRFs were first used to describe the properties of neurons in the auditory system (Aertsen and Johannesma 1981b) but are now commonly used in other sensory modalities (DeAngelis *et al* 1995, DiCarlo and Johnson 2000).

STRFs have been estimated predominantly by using white-noise stimuli and extracting the STRF by reverse correlation (DeBoer and Kuyper 1968). For white noise, the STRF is proportional to the correlation between the stimulus and the response, and the reverse correlation method relies on this fact. This technique is powerful and efficient, but has two serious limitations. First, some sensory neurons are particularly selective for natural stimuli and respond poorly to white noise. In this case the estimated STRF will be an unacceptably noisy descriptor of the neuron's transfer function (Aertsen and Johannesma 1981a, Nelken *et al* 1997). Second, even if white noise drives the neuron, the resulting STRF may not predict responses to natural stimuli. Complex stimuli can alter responses by pushing a neuron into a different operational range or by evoking nonlinear response properties (Theunissen *et al* 2000). If we can obtain STRFs from behaviourally relevant natural stimuli, we can obtain a rigorous description of a neuron's responses in its natural operating regime.

In this paper we derive a more general reverse correlation method that can be used to extract STRFs from neural responses to arbitrary stimuli, including natural stimuli. We show that the analytical solution for the STRF takes different forms, depending on the nature of the second-order statistics of the stimuli. We then demonstrate, both by modelling and in neurophysiological experiments, how this method can be used to successfully estimate STRFs using natural stimuli. The general solution to the reverse correlation problem is mathematically involved, and there is no single reference that offers a rigorous summary of this method. Therefore, we also include an appendix with a collection of the theorems and proofs that form the foundations of STRF estimation.

2. Methods

We present a general method for estimating the STRFs of sensory neurons from their responses to complex stimulus ensembles. The solution takes the second-order statistics of the stimulus ensemble into account. If the statistics are stationary, the solution has a simple analytical form in the Fourier domain: the STRF is obtained by normalizing cross-stimulus–response power spectral density by the power spectral density of the stimulus. We show that the discrete form of this solution is a particular case of the general solution of the linear minimum mean-square estimate (LMMSE) of the STRF. The general solution of the LMMSE involves normalizing the cross-covariance matrix between the stimulus and response by the auto-covariance matrix of the stimulus. The link between the general solution and the specific solution found in the stationary case is given by the fact that the eigenvectors of a stationary stimulus covariance matrix are the coefficients of the discrete Fourier transform (DFT).

We then demonstrate how the general solution can be applied when the spatial statistics are non-stationary, a case that is important for both audition and vision. In audition, the spectrographic representation of natural sounds does not have stationary spatial second-order statistics. In vision the two-point correlation function is translation invariant for ensembles of natural images, but this may not be true for data acquired in real neurophysiology experiments using relatively brief, spatially bounded stimulus ensembles.

The STRF is defined mathematically as a spatio-temporal impulse response, $h(t, x)$, that relates a spatio-temporal description of a stimulus, $s(t, x)$, to a neural response $r(t)$. For example, in the visual system, $s(t, x)$ can represent the light intensity as a function of time and position. In the auditory system, $s(t, x)$ can represent the time-varying amplitude of sound in a frequency band centred at x. If we assume that the neuronal system is time invariant, then a linear estimate of the stimulus–response transfer function is obtained by convolving the STRF, $h(t, x)$, with the stimulus

$$\hat{r}(t) = \iint h(\tau, x)s(t - \tau, x)\,\mathrm{d}\tau\,\mathrm{d}x + r_0, \tag{1}$$

where r_0 is the mean neural response and $\hat{r}(t)$ is the estimate of the actual neural response $r(t)$. The impulse response function $h(t, x)$ is the STRF of the neuron.

An analytical solution for $h(t, x)$ and r_0 can be found by minimizing the expectation value of the square difference between $r(t)$ and $\hat{r}(t)$, $E[(\hat{r} - r)^2]$. This expectation value is obtained by averaging across data samples and can be rewritten as $\left\langle\left(r(t) - \hat{r}(t)\right)^2\right\rangle$. It is easily shown that the best mean-square estimate for r_0 is simply $\langle r\rangle$ and, therefore, the response can be normalized to a mean of zero. We can also assume without loss of generality that the stimulus parameters have been re-scaled to have zero means.

2.1. Solution for one dimension in the stationary case: continuous formulation

The one-dimensional solution for h that satisfies equation (1) has been discussed elsewhere for the case of stationary temporal statistics where the stimulus and response are continuous functions of time (e.g. Marmeralis and Marmeralis 1978, Rieke *et al* 1997, Theunissen *et al* 1996). For completeness, a particular form of this solution is given in the appendix. The following section presents the solution for the multi-dimensional case. Finally, in later sections we demonstrate how the same answer can be obtained from the general derivation of the problem in the discrete case, where time and space are sampled and represented as vectors.

The one-dimensional version of equation (1) is written as

$$\hat{r}(t) = \int_{-\infty}^{\infty} h(\tau)s(t - \tau)\,\mathrm{d}\tau. \tag{2}$$

As long as the second-order statistics of $s(t)$ are stationary, the derivation can be obtained directly in the Fourier domain. Denoting by $R(\omega_t)$, $H(\omega_t)$ and $S(\omega_t)$ the Fourier transform of $r(t)$, $s(t)$ and $h(t)$, respectively, the convolution in equation (2) then becomes a product in the Fourier domain:

$$\hat{R}(\omega_t) = H(\omega_t)S(\omega_t).$$

The solution for $H(\omega_t)$ is then obtained with (appendix, proof 1)

$$H(\omega_t) = \frac{\langle S^*(\omega_t)R(\omega_t)\rangle}{\langle S^*(\omega_t)S(\omega_t)\rangle}, \tag{3}$$

where the averages are taken across data samples and the * indicates the complex conjugate.

These equations demonstrate that the impulse response of the linear estimator is found by normalizing the cross-correlation between the stimulus and the response (numerator) by the power spectrum of the stimulus (denominator). If the stimulus is spectrally white then the denominator is a constant and the impulse response of the linear estimator is proportional to the cross-correlation between the stimulus and the response. In this case the solution can be found directly in the time domain. With a zero-mean stimulus the cross-correlation can be obtained simply by averaging the stimulus before each spike and reversing the time axis (appendix, proof 2). In other words, the impulse response of the filter is proportional to the spike-triggered average (STA) stimulus. This particular solution is the well known reverse correlation method that was first applied in the auditory nerve (DeBoer and Kuyper 1968).

2.2. *Solution for multiple spatial dimensions in the stationary case: continuous formulation*

The derivation for the spatial dimension differs slightly from the derivation for the temporal dimension. Rewriting equation (1) and ignoring the temporal dimension for the moment, we have

$$\hat{r} = \int_{-\infty}^{\infty} h(x)s(x)\,\mathrm{d}x. \tag{4}$$

One can rewrite the previous equation as

$$\hat{r}(x) = \int_{-\infty}^{\infty} h(x+x')s(x')\,\mathrm{d}x', \qquad \text{with} \quad \hat{r}(x) = \delta(x)\hat{r},$$

so that the Fourier domain representation of (4) can be written as

$$\hat{R}(\omega_x) = H(\omega_x)S^*(\omega_x).$$

A similar derivation to that used for the time dimension then leads to

$$H(\omega_x) = \frac{\langle S(\omega_x)r\rangle}{\langle S^*(\omega_x)S(\omega_x)\rangle}. \tag{5}$$

Note that in this case the complex conjugate of $S(\omega_x)$ does not appear in the numerator.

We can now combine the derivation for the time and space domains to find the solution of equation (1) for the case of stationary spatial and temporal statistics. Taking the Fourier transform of both the spatial and temporal dimension and defining $R(\omega_t, x) = R(\omega_t)\delta(x)$, equation (1) can be rewritten in the Fourier domain as

$$\hat{R}(\omega_t, \omega_x) = H(\omega_t, \omega_x)S^*(\omega_t, \omega_x).$$

Therefore, for stationary temporal and spatial statistics

$$H\,(\omega_t, \omega_x) = \frac{\langle S(\omega_t, \omega_x)R(\omega_t)\rangle}{\langle S^*(\omega_t, \omega_x)S(\omega_t, \omega_x)\rangle}. \tag{6}$$

Just as for the spatial one-dimensional case, if the statistics of the stimulus are stationary the impulse response of the linear filter is obtained by normalizing the two-dimensional Fourier transform of the cross-correlation between the stimulus and response by the power spectrum of the stimulus. Here, the power spectrum of the stimulus is also obtained by the two-dimensional Fourier transform.

2.3. Solution in the general case: discrete formulation

If the second-order statistics of the stimulus ensemble vary in either time or space, the ensemble is non-stationary. In this case, the general solution of equation (1) must be solved with a discrete formulation. We therefore derive the general form of the solution in the discrete case and show that the discrete form of equations (6) can be recovered when stationarity is assumed. We then show that a similar solution can be found when the second-order stimulus statistics are not stationary. An important case for sensory neurophysiology occurs when the temporal statistics are stationary and the spatial statistics are not.

The discrete form of equation (1) can be written as

$$\hat{r}[t] = \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} h_t[i,k] s[t-i,k], \tag{7}$$

where the index i sums over N points in time (the memory of the system) and the index k sums over M spatial parameters. The number of time points, N, and the time interval between sampling points depends on both the memory of the system and the highest temporal frequency being encoded by the neuron (or present in the stimulus). Similarly, M depends on the spatial extent of the receptive field of the neuron and on the highest spatial frequency of interest.

Equation (7) can be written in vector form by using one column to represent both the spatial and temporal dimension of the stimulus

$$\hat{r}[t] = \sum_{i=0}^{M\times N-1} h[i] s_t[i] \qquad \text{or} \qquad \hat{r}[t] = \boldsymbol{h}^{\mathrm{T}} \cdot \boldsymbol{s}_t,$$

where $\boldsymbol{h} = [h_0, \ldots, h_{N\times M-1}]^{\mathrm{T}}$ are the linear coefficients describing the linear estimator, and both the stimulus and the response are sampled at particular points in time, t.

Minimizing $\langle(\hat{r} - r)^2\rangle$ gives the solution for the vector $\boldsymbol{h}$ (appendix, proof 3):

$$\boldsymbol{h} = \boldsymbol{C}_{\mathrm{ss}}^{-1} \boldsymbol{C}_{\mathrm{sr}}, \tag{8}$$

where $\boldsymbol{C}_{\mathrm{ss}}$ is the $(NM) \times (NM)$ auto-correlation matrix of the stimulus, $\boldsymbol{s}$, and $\boldsymbol{C}_{\mathrm{sr}}$ is the $1 \times (NM)$ cross-correlation vector of the stimulus, $\boldsymbol{s}$, and the neural response, $r[t]$ (see appendix, proof 3).

Equation (8) is the well known multi-dimensional solution of the linear regression problem (Kay 1993) and is similar in form to equations (3), (5) and (6). However, unlike those equations, equation (8) does not involve any assumptions about the statistical properties of the stimulus and therefore involves all possible cross-moments of the stimulus. To solve for $\boldsymbol{h}$, equation (8) requires the inversion of the stimulus auto-correlation matrix, $\boldsymbol{C}_{\mathrm{ss}}$ (whereas equations (3), (5) and (6) involve simple scalar divisions). If the statistics of the stimulus are not stationary in time, equation (8) gives an estimate for $\boldsymbol{h}$ at each time point, t. In this case multiple samples are obtained by repeating the experiment with identical initial conditions to generate multiple stimulus–response combinations at time t. However, if the linear response of the neuron is time invariant, then these estimates of $\boldsymbol{h}$ at different points in time can be averaged to obtain a single time-invariant linear estimate of the neuron's transfer function.

Equation (8) can be used to estimate the linear transfer function of a neuron regardless of whether the transfer function is time invariant; whether the second-order temporal statistics of the stimulus are stationary or whether the second-order spatial statistics of the stimulus are stationary. We now simplify the calculation of equation (8) by deriving analytical expressions for the case where the transfer function of the neuron is time invariant and the statistics of the stimulus are stationary in time.

We can diagonalize the symmetric correlation matrix $\boldsymbol{C}_{\text{ss}}$:

$$\boldsymbol{C}_{\text{ss}} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{-1},$$

where $\boldsymbol{Q}$ is the eigenvector matrix of $\boldsymbol{C}_{\text{SS}}$ and Λ is the diagonal matrix of eigenvalues. Equation (8) can then be written as

$$\boldsymbol{Q}^{-1}\boldsymbol{h} = \Lambda^{-1}\boldsymbol{Q}^{-1}\boldsymbol{C}_{\text{sr}}. \tag{9}$$

This indicates that a solution for $\boldsymbol{h}$ can be found by a scalar division along each eigenvector in the eigenspace of the stimulus auto-correlation matrix. In the following section, we show that when the statistics are stationary the eigenspace is the discrete Fourier space, so that equation (9) becomes equivalent to (3), (5) and (6).

2.4. *Solution for stationary statistics in the discrete case*

In this section we assume that the transfer function is time invariant and that the temporal statistics of the stimulus are stationary. Initially, we also assume that space is one dimensional. Equation (7) is then simply

$$\hat{r}[t] = \sum_{i=0}^{N-1} h[i]s[t-i]. \tag{10}$$

Because the stimulus auto-correlation matrix ($\boldsymbol{C}_{\text{ss}}$) is now independent of reference time, t, it can be written as

$$\boldsymbol{C}_{\text{ss}} = \begin{bmatrix} c_{\text{ss}}[0] & c_{\text{ss}}[1] & \cdots & c_{\text{ss}}[N-1] \\ c_{\text{ss}}[1] & c_{\text{ss}}[0] & \cdots & c_{\text{ss}}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ c_{\text{ss}}[N-1] & c_{\text{ss}}[N-2] & \cdots & c_{\text{ss}}[0] \end{bmatrix},$$

where $c_{\text{ss}}[j]$ is the correlation between two stimulus points separated by j time points,

$$c_{\text{ss}}[j] = \langle s[t]s[t-j]\rangle.$$

Each row of this matrix is a copy of the previous row shifted by one index. Such matrices are called symmetric Toeplitz. The eigenvalues and eigenvectors of the symmetric Toeplitz matrix are given by the DFT (Davis 1979, appendix, proof 4).

Calling the matrix containing the DFT eigenvectors $\boldsymbol{Q}_{\text{FT}}$, equation (9) can be rewritten as

$$\boldsymbol{Q}_{\text{FT}}^{-1}\boldsymbol{h} = \frac{\boldsymbol{Q}_{\text{FT}}^{-1}\boldsymbol{C}_{\text{sr}}}{\Lambda}. \tag{11}$$

Equation (11) is the discrete form of equation (3) and, as in the continuous formulation, it shows that the impulse response can be estimated by a scalar division in the Fourier domain. The impulse response (in the Fourier domain) is given by the Fourier transform of the cross-correlation between the stimulus and response, divided by the power of the stimulus. In equation (11), Λ is the diagonal matrix whose elements are the power spectrum as a function of temporal frequency (see appendix, proof 4). Note that in this section, the index of the cross-correlation between the stimulus and response goes backward in time. This inversion of the temporal axis results in the complex conjugate in the numerator of equation (3).

A similar situation arises when the stimulus has stationary spatial statistics that are independent of the temporal dimension. In equation (10) we simply replace $s[t-i]$ by $s[i]$, where the N spatial parameters of the stimulus are represented by the index i. Just as for the temporal parameters, the two-point correlation between $s[i]$ and $s[i+j]$ depends only on j, the distance between the points. The eigenvectors and eigenvalues of the stimulus auto-correlation matrix are again given by the DFT. The spatial filter coefficients, $\boldsymbol{h}$, are also given by equation (11), which corresponds to equation (5). Note that in contrast to the solution for the temporal dimension, the index in the cross-correlation between the stimulus and the response goes forward in space. Therefore, there is no complex conjugate in the numerator of equation (5).

2.5. General solution in the discrete case for stationary temporal statistics

In the more general case of equation (9), the stimulus auto-correlation matrix can be separated into its spatial and temporal dimensions:

$$\boldsymbol{C}_{\mathrm{ss}} = \begin{pmatrix} \boldsymbol{c}_{0,0} & \cdots & \boldsymbol{c}_{0,M-1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{c}_{M-1,0} & \cdots & \boldsymbol{c}_{M-1,M-1} \end{pmatrix},$$

where

$$\boldsymbol{c}_{i,j} = \begin{pmatrix} \langle s[t,i]s[t,j]\rangle & \langle s[t,i]s[t-1,j]\rangle & \cdots & \langle s[t,i]s[t-(N-1),j]\rangle \\ \langle s[t-1,i]s[t,j]\rangle & \langle s[t-1,i]s[t-1,j]\rangle & & \vdots \\ \vdots & & \ddots & \\ \langle s[t-(N-1),i]s[t,j]\rangle & \cdots & & \langle s[t-(N-1),i]s[t-(N-1),j]\rangle \end{pmatrix}.$$

The sub-matrix $\boldsymbol{c}_{i,j}$ describes the correlations between spatial dimensions i and j for all relevant time delays. When the system is linear and time invariant and the second-order temporal statistics are stationary, each $\boldsymbol{c}_{i,j}$ is symmetric Toeplitz and its eigenvectors are contained in $\boldsymbol{Q}_{\mathrm{FT}}$ defined above. Defining $\boldsymbol{Q}_t$ as

$$\boldsymbol{Q}_t = \begin{pmatrix} \boldsymbol{Q}_{\mathrm{FT}} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{Q}_{\mathrm{FT}} \end{pmatrix},$$

equation (9) can be written as

$$\Lambda_t \boldsymbol{H} = \boldsymbol{C}_{\mathrm{SR}}, \tag{12}$$

where

$$\boldsymbol{H} = \boldsymbol{Q}_t^{-1}\boldsymbol{h}, \qquad \boldsymbol{C}_{\mathrm{SR}} = \boldsymbol{Q}_t^{-1}\boldsymbol{C}_{\mathrm{sr}}$$

and

$$\Lambda_t = \begin{pmatrix} \Lambda_{0,0} & \cdots & \Lambda_{0,M-1} \\ \vdots & \ddots & \\ \Lambda_{M-1,0} & & \Lambda_{M,M} \end{pmatrix}.$$

Each $\Lambda_{j,k}$ is an $N \times N$ diagonal matrix describing the cross-correlation of stimulus spatial dimensions j and k, for all N temporal frequencies, $\omega[i]$. Because Λ_t is composed of diagonal matrices, equation (12) can be written as a set of linear equations for each frequency, $\omega[i]$ (appendix, proof 5):

$$\Lambda\left(\omega_t[i]\right)\boldsymbol{H}\left(\omega_t[i]\right) = \boldsymbol{C}_{\mathrm{SR}}\left(\omega_t[i]\right), \tag{13}$$

where $\Lambda\,(\omega_t[i])$ is the $M \times M$ matrix describing all the stimulus correlations across spatial dimensions for the temporal frequency $\omega_t[i]$ (see appendix, proof 5).

The set of equations (13) can be use to solve for $\boldsymbol{H}$, and $\boldsymbol{h}$ can then be obtained by taking the inverse DFT. To solve (13), we must invert each $\Lambda\,(\omega_t[i])$. These matrices are Hermitian and therefore have eigenvectors and real eigenvalues. Denoting by $\boldsymbol{Q}_x(\omega_t[i])$ the matrix of eigenvectors of $\Lambda\,(\omega_t[i])$ and by $\Lambda_x\,(\omega_t[i])$ the corresponding eigenvalues gives

$$\boldsymbol{H}\,(\omega_t[i]) = \boldsymbol{Q}_x\,(\omega_t[i])\,\Lambda_x^{-1}\,(\omega_t[i])\,\boldsymbol{Q}_x^{-1}\,(\omega_t[i])\,\boldsymbol{C}_{\mathrm{SR}}\,(\omega_t[i])\,. \tag{14}$$

Equation (14) can be used to estimate the STRF of a sensory neuron without making any assumptions about the second-order statistical properties of the spatial dimension.

In summary, the discrete solution for $\boldsymbol{h}$ with stationary temporal statistics is obtained by the following steps: (1) calculate the stimulus auto-correlation and stimulus–response cross-correlations; (2) take the Fourier transform along the temporal dimension of both the auto- and cross-correlation; (3) for each temporal frequency, find the eigenvectors and eigenvalues of the spatial auto-correlation matrix of the stimulus; (4) use the eigenvalues of this spatial auto-correlation function to normalize the cross-correlation in the basis set defined by the eigenvectors; (5) return to the original spatial basis set; (6) take the inverse temporal Fourier transform.

When the second-order spatial statistics are also stationary then all the $\boldsymbol{Q}_x(\omega_t[i])$ are equal to $\boldsymbol{Q}_{\mathrm{FT}}$. In this case steps (3) and (5) above consist of taking the spatial Fourier transform and its inverse, and several steps ((2) and (3), and (5) and (6)) can be merged by taking the two-dimensional (time–space) DFT. Step (4) then reduces to normalizing the stimulus–response cross-correlation by the two-dimensional power spectrum of the stimulus. Mathematically, we write

$$\boldsymbol{H}\,(\omega_t[i], \omega_x[k]) = \frac{\boldsymbol{C}_{\mathrm{SR}}\,(\omega_t[i], \omega_x[k])}{\boldsymbol{P}_{\mathrm{s}}\,(\omega_t[i], \omega_x[k])}, \tag{15}$$

where $\boldsymbol{H}\,(\omega_t[i], \omega_x[k]) = \boldsymbol{Q}_{\mathrm{FT}}^{-1}\boldsymbol{H}\,(\omega_t[i])$, $\boldsymbol{C}_{\mathrm{SR}}\,(\omega_t[i], \omega_x[k]) = \boldsymbol{Q}_{\mathrm{FT}}^{-1}\boldsymbol{C}_{\mathrm{SR}}\,(\omega_t[i])$ and $\boldsymbol{P}_{\mathrm{S}}\,(\omega_t[i], \omega_x[k]) = \Lambda_x\,(\omega_t[i])$. Equation (15) is the discrete equivalent of equation (6). The index i in these equations goes back in time.

In the results section we use these methods to estimate STRFs of both model and real neurons using natural and synthetic stimuli. We also compare the results obtained using equation (14), which does not assume spatial stationarity, with equation (15), where spatial stationarity is assumed.

2.6. Goodness of fit measure

To judge the quality of the estimated STRF we need to compare the predicted response, $\hat{r}(t)$, with the actual response, $r(t)$. To prevent over-fitting such comparisons are made with a set of response data (the validation set) that were not used to estimate the STRF.

We use two different measures to quantify the estimated STRF's goodness of fit. The first measure is based on the coherence between $r(t)$ and $\hat{r}(t)$. The coherence is a function of frequency and is given by Marmeralis and Marmeralis (1978):

$$\gamma^2(\omega) = \frac{\langle R(\omega)\hat{R}(\omega)^*\rangle\langle R(\omega)^*\hat{R}(\omega)\rangle}{\langle R(\omega)R(\omega)^*\rangle\langle\hat{R}(\omega)\hat{R}(\omega)^*\rangle}. \tag{16}$$

The coherence measures the correlation between the actual response and the estimator at each temporal frequency, ω. Coherence is a good measure because it can be calculated directly from a raw post-stimulus time histogram (PSTH) without any smoothing (i.e. for an arbitrarily small time window). However, when the signal-to-noise ratio is low, coherence is a positively biased measure and additional statistical techniques must be used to guarantee its validity (Thomson and Chave 1991). An overall goodness-of-fit estimate is obtained from the coherence function by the following integration:

$$I = -\int_0^\infty \log_2(1-\gamma^2)\,\mathrm{d}\omega. \tag{17}$$

Here I is expressed in bits/second and can be used to estimate a lower bound on the information transmitted by the neuron if the noise (defined as the difference between $r(t)$ obtained from a single trial and $\hat{r}(t)$) is Gaussian (Rieke *et al* 1997, Borst and Theunissen 1999). More generally I serves as a measure of the integrated coherence.

Because neurophysiological experiments often report simple correlation, it is also useful to calculate the correlation coefficient (cc) between $r(t)$ and $\hat{r}(t)$:

$$cc = \frac{\left\langle \left(r(t) - \overline{r(t)}\right)\left(\hat{r}(t) - \overline{\hat{r}(t)}\right)\right\rangle}{\sqrt{\left\langle (r(t) - \overline{r(t)})^2\right\rangle \left\langle (\hat{r}(t) - \overline{\hat{r}(t)})^2\right\rangle}}. \tag{18}$$

It is important to note that the cc depends on the time binning that is used to obtain $r(t)$ from the PSTH. Small time bins tend to produce low ccs while large time bins tends to artificially elevate ccs. Transmitted information is roughly proportional to the cc divided by the length of the time window, so ccs can only be compared between similar time windows.

2.7. Numerical issues

Equations (14) and (15) both require the division of the stimulus and response cross-correlation by the eigenvalues that describe the power of the stimulus ensemble in the appropriate basis. The spatial Fourier basis is used for equation (15) and a different spatial eigenspace is used in equation (14). When natural scenes or sounds are used as stimuli, or indeed for any band-limited stationary stimulus, stimulus power may be very small or even zero for a set of spatial and temporal frequencies. Similarly, the eigenvalues along specific temporal frequencies and spatial eigendimensions in equation (14) will be very close to zero. To prevent numerical errors, it is necessary to determine the subset of eigenvectors in equation (14) that have significant non-zero eigenvalues. Similarly, for equation (15) it is necessary to determine the spatio-temporal frequencies that have significant non-zero power.

To solve this problem, we include only those eigenvalues in the analysis that exceed a specific fraction of the largest eigenvalue or peak power. The eigenvalues or power levels below this threshold are considered to be insignificant. This is equivalent to performing a singular-value decomposition (SVD) of the stimulus auto-correlation. In practice, we find the threshold level maximizes the ability of estimated STRFs to predict responses obtained in the validation data set (recall that these data were not used to estimate the STRF). Prediction quality is determined by calculating the information value shown in equation (28). This cross-validation method effectively removes noise from both the stimulus auto-correlation estimate and the stimulus–response cross-correlation.

It is important to note that in this procedure we are, in effect, dividing the stimulus space into two subspaces: one subspace that is sampled sufficiently for analysis, and another subspace that is not sampled sufficiently. The estimated STRF will therefore only be representative of the neuron's true STRF if the experimental stimulus space adequately samples all the spatio-temporal or eigenvector components covered by the neuron's true STRF. We will return to this important point in the results and discussion sections.

3. Results

In the following sections, we first describe the second-order statistics of the natural sounds and natural vision movies that we use in our experiments. We then show results from modelling experiments that (1) demonstrate the validity of our methods, (2) compare the results obtained with equation (14) versus equation (15) and (3) illustrate some potential limitations in the use of bandpass and natural stimuli to estimate STRFs. Finally, we apply the methods to high-level auditory neurons in songbirds and to visual neurons in area V1.

3.1. Properties of the second-order statistics of natural stimuli used in our experiments

One of the challenges of using natural stimuli to probe sensory systems is to describe the sensory space itself. In order to estimate the linear stimulus response function, a complete description of the second-order statistics of the stimulus ensemble is required. When these second-order statistics are stationary they are given by the power spectrum.

3.1.1. Natural sounds. To characterize high-level auditory neurons, the one-dimensional sound pressure waveform, $s(t)$, is first transformed into a spectrographic representation, $s(t, x)$, where $s(t, x)$ is the amplitude envelope of the sound in a set of frequency bands centred at x (Klein *et al* 2000, Theunissen *et al* 2000). Figure 1 shows the second-order statistics of the spectrogram representation for two different stimulus ensembles, a collection of zebra finch songs and a collection of tone pips designed to have some of the same acoustical properties as zebra finch song. (The duration of the tone pips and tone gaps had the same mean and variance as found in zebra finch song syllables and gaps, and the power spectra of the stimuli were matched.) Each ensemble consisted of approximately 20 s of sound. The spectrographic representation used 31 frequency bands of 250 Hz bandwidth, spanning the frequency range 250–8000 Hz. The amplitude envelope in each band was sampled at 1 kHz (Theunissen *et al* 2000).

The left-hand panel of figure 1 shows the stimulus auto-correlation matrix, $\boldsymbol{C}_{\mathrm{ss}}$. Since the matrix is symmetric, only the upper diagonal is shown. Each element corresponds to the correlations between two spatial dimensions, $\boldsymbol{c}_{ij}$. In this case the $\boldsymbol{c}_{ij}$ represent correlations between the amplitude envelopes of different frequency bands. Since the temporal statistics are stationary, each submatrix, $\boldsymbol{c}_{ij}$, is symmetric Toeplitz and can be collapsed into a single row. This simplification is shown in the left-hand panels of figure 1.

A close look at the auto-correlation matrix reveals some basic properties of these two sound ensembles. First, because the rows are not identical to shifted versions of one another, the spatial statistics of these ensembles are not completely stationary. Second, the large off-diagonal terms indicate that the zebra finch song contains correlations between all frequency bands. This occurs because the syllables of zebra finch song are broad band and the amplitude envelopes are co-modulated across frequencies. In contrast, the tone pip ensemble shows only small off-diagonal terms, confirming that these stimuli have small correlations between frequency bands.

The middle panels of figure 1 show the two-dimensional power spectra of the two sound ensembles. $\boldsymbol{P}_{\mathrm{S}}\,(\omega_t[i], \omega_x[k])$, which is used in equation (15), is plotted as a function of temporal (ω_t, x-axis) and spatial (ω_x, y-axis) frequencies. If the second-order spatial statistics of these sounds were stationary, the power spectra would be as informative as the full-correlation matrix shown in the left-hand panels of figure 1. In this case, the two-dimensional power spectra could be obtained by averaging the auto-correlation matrix across rows and taking the two-dimensional Fourier transform.

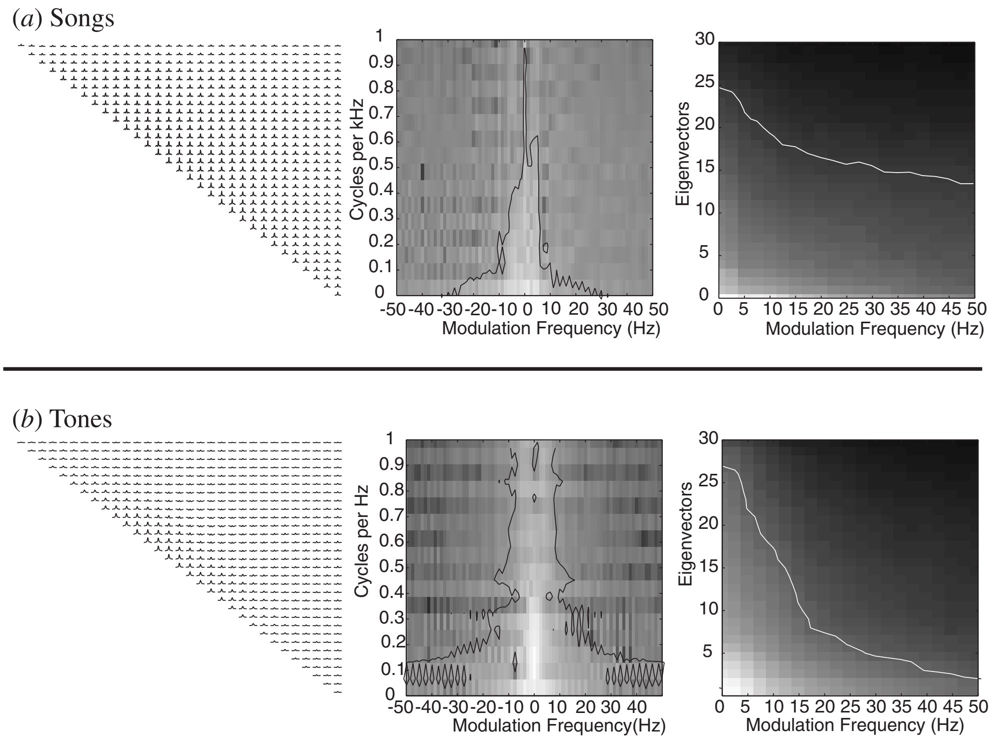

Figure 1. (*a*) Second-order statistics for zebra finch song. The left-hand panel shows the stimulus auto-correlation matrix in the space–time domain. Each entry in the matrix corresponds to the temporal cross-correlations of the sound amplitude in two different frequency bands. The time course covers a period of ±200 ms. The spectrographic representation of the sound is composed of 31 frequency bands spanning the frequency range between 250 and 8000 Hz. The frequency scale is linear and each frequency band has a bandwidth of 250 Hz. The matrix is organized with lowest centre frequency of the band at the top left corner and the highest frequency at the bottom right (see Theunissen *et al* 2000 for details). The middle panel shows the spatio-temporal power spectra of the song stimuli. A fit of the marginal power spectra shows that the power of the amplitude scales approximately as $1/\omega_t^{0.5}$ and that the power of the frequency ripples scales approximately as $1/\omega_x^{0.5}$. The right-hand panel shows the magnitude of the eigenvalues for the spatial eigenvectors (y axis) for each temporal frequency (x axis). The eigenvectors are different for each temporal frequency and have been sorted in rank order. The black and white lines bound the area of the stimulus space where the power is significantly different from zero (see methods and results). (*b*) Second-order statistics for tone pips designed to have statistics similar to those of zebra finch song (see results).

The two-dimensional power spectrum of the zebra finch song shows that most of the energy is in the low temporal frequencies (i.e. the low frequencies of amplitude modulation of the sound envelopes) and that for amplitude modulation frequencies below 50 Hz the power scaled as $1/\omega_t^{0.5}$. The spatial frequencies are also dominated by the low-frequency components, except at low temporal frequencies. The power spectrum of the songs is also slightly asymmetric, indicating a difference in the acoustics of downward (right quadrant) versus upward (left quadrant) frequency sweeps. In contrast, the two-dimensional power spectrum of the tone pips covers the same range of temporal and spatial frequencies but is completely symmetric.

The right-hand panels of figure 1 show the spatial eigenvalues, Λ_x $(\omega_t[i])$ of equation (14), for each temporal frequency, $\omega_t[i]$. In this figure each $\omega_t[i]$ (represented on the x-axis) is associated with a different set of eigenvectors (shown along the y-axis). The eigenvectors are ordered by decreasing eigenvalue. The black line in the middle and right-hand panels separates the values above threshold from those those below threshold for the specific neuron shown in the last section 3.3.1. In practice, only the region of acoustical space above this threshold is used to calculate $\boldsymbol{h}$ using (14) or (15).

3.1.2. Natural vision movies. Figure 2(*a*) shows the stimulus auto-correlation matrix for a natural vision movie which simulates natural viewing of a static natural scene. Natural vision movies were constructed by extracting image patches from a greyscale image along a simulated eye scan path. Eye scan paths were generated using a statistical model of natural eye movements made during free viewing (Vinje and Gallant 2000). In analogy with figure 1, each element in the matrix shown in figure 2(*a*) represents the correlation $\boldsymbol{c}_{ij}$ between two spatial dimensions. In the visual case these are the correlations between the greyscale luminance intensity of pixel pairs. Because the temporal statistics are stationary, the submatrix $\boldsymbol{c}_{ij}$ has been collapsed into a single function of time-lag (covering a range of ±840 ms). The stimulus ensemble consisted of a natural vision movie 50 s long and displayed at 72 Hz.

As noted earlier, the ensemble of all possible natural scenes has stationary statistics (Field 1987). However, inspection of figure 2(*a*) reveals that it is not truly Toeplitz, which indicates that the spatial statistics of this particular 50 s natural vision movie are non-stationary. In this sense these specific visual stimuli are similar to natural auditory stimuli, which are also non-stationary. The non-stationary statistics of natural vision movies is an unavoidable artifact of their use in neurophysiological experiments. First, natural vision movies used in experiments are often masked by a circular window and pixels on opposite sides of this boundary have different statistical properties. Second, neuro-physiological experiments have a finite length and therefore finite stimulus sampling. This finite sampling effectively introduces non-stationary statistics.

Another notable aspect of figure 2(*a*) is the temporal auto-correlation, which has a large peak around a time-lag of zero. This peak is caused by the saccadic structure of natural vision. During a single fixation each pixel is completely correlated in time, but uncorrelated across fixations. Because fixation durations are roughly equal, the stimulus auto-correlations decrease almost linearly up to the average fixation time.

Figure 2(*b*) shows the logarithm of the absolute value of the stimulus auto-correlation matrix in the spatial Fourier domain. This figure is complementary to the one shown in figure 2(*a*), but the Fourier domain is shown because analysis in the Fourier domain facilitates analysis of V1 complex cells that have spatially phase-invariant responses (David *et al* 1999). Each element in figure 2(*b*) represents the correlation between two spatial frequencies, but only one temporal frequency (3.6 Hz, the lowest sampled temporal frequency) is shown for each element. Here the non-stationarity of natural vision movies is clearly revealed by the significant off-diagonal terms.

Figure 2(*c*) shows the spatial Fourier domain power spectrum of the natural vision movie used in calculating the auto-correlation matrix in figure 2(*b*). Figure 2(*d*) shows power as a function of both spatial and temporal frequency (spatial frequencies correspond to a vertical slice through figure 2(*c*) at one cycle/frame). As in the auditory case, the power spectra are dominated by low frequencies. The spatial power spectrum of natural scenes falls off roughly as $1/\omega_t^2$ (Field 1987). In the temporal domain power falls off as $\sim 1/\omega_t^2$. Saccades resemble 'scene cut' transitions in movies, and such transitions are known to generate power-law behaviour (Dong and Atick 1995). In contrast to the structure in zebra finch song, these visual images are symmetric along the positive and negative frequencies, demonstrating that there is no preferred spatial orientation or direction of motion within a natural vision movie.

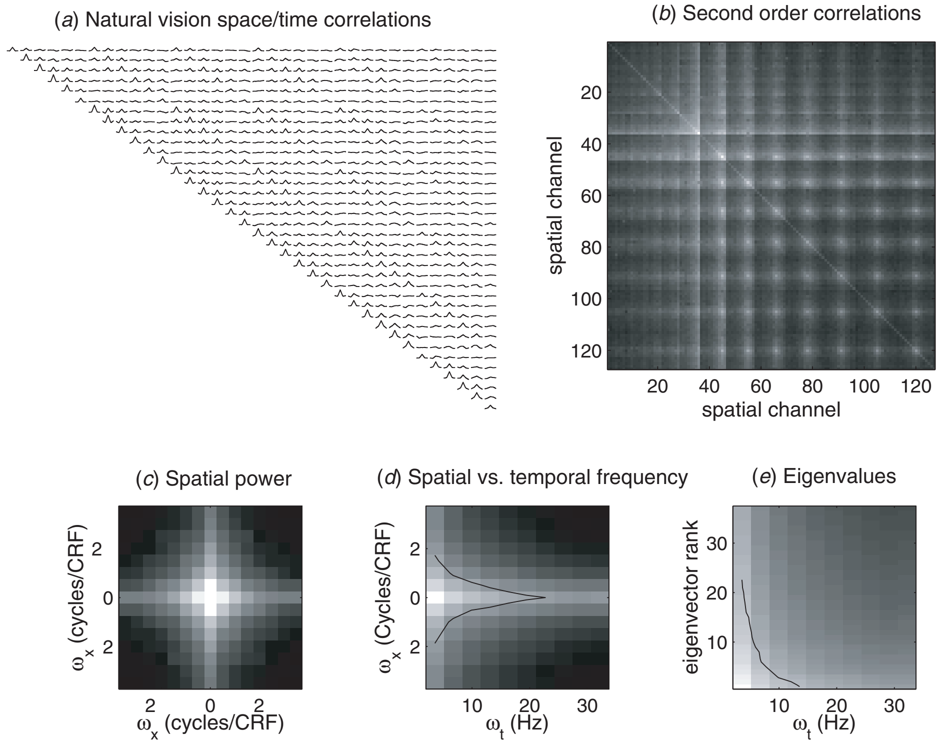

Figure 2. (*a*) Stimulus auto-correlation matrix for a natural vision movie in the space–time domain. Axes correspond to spatial locations of each pixel. Each curve shows the temporal correlation function between a pair of pixels. The time course covers a period of ±840 ms. The movie length was 50 s. The spatial plane has been downsampled from 60×60 to 6×6 pixels in order to facilitate display. The repeating pattern in the off-diagonal elements of the auto-correlation matrix does not reflect structure in the stimuli, but rather is caused by wrapping each ω_x and ω_y into a single vector before computing each $\boldsymbol{c}_{ij}$. (*b*) Stimulus auto-correlation matrix for a natural vision movie in the Fourier power domain. Axes correspond to spatial frequencies and each point represents the correlation between a pair of two-dimensional spatial frequencies. The auto-correlation matrix is shown for one temporal frequency, 3.6 Hz, and terms corresponding to dc spatial frequency are not shown. The spatial plane was down-sampled from 60×60 to 16×16 pixels to reveal the fine structure of the matrix. The rectangular texture is caused by reordering the two-dimensional spatial Fourier plane into a single vector for display. The strong off-diagonal terms indicate the presence of non-stationarity in the stimulus. (*c*) Spatial power spectrum of a natural vision movie. Axes correspond to points in the spatial Fourier domain. The spectrum is taken at 3.6 Hz temporal frequency. The dc term—zeroed for the analysis—has been set to the top of the colour scale in order to better show spectral structure. The $1/\omega_t^2$ structure of the power spectrum is readily apparent at all orientations. (*d*) Spatio-temporal power spectrum of a natural vision movie. The vertical axis corresponds to Fourier coefficients in the spatial domain; the horizontal axis corresponds to Fourier coefficients in the temporal domain. The spatial power spectrum corresponds to a vertical slice through figure 2(*c*) at 1 cyc/frame. The spatial domain again shows $1/\omega_t^2$ structure, while power in the temporal domain also shows $1/\omega_t^2$ structure due to the pattern of fixations and saccades. Those spectral components inside the contour line have stimulus power that is above the threshold used in the stationary analysis of the model cell discussed in section 3.2.2. (*e*) Eigenvalues of the stimulus auto-correlation matrix as a function of temporal frequency. The vertical axis corresponds to eigenvector rank, assigned in order of descending eigenvalue amplitude. The horizontal axis corresponds to coefficients in the temporal Fourier domain. The eigenvalues within the contour lie above the significance threshold used to analyse data from the model cell. The eigenvectors corresponding to these eigenvalues are retained for the normalization of STRF estimates in equation (14).

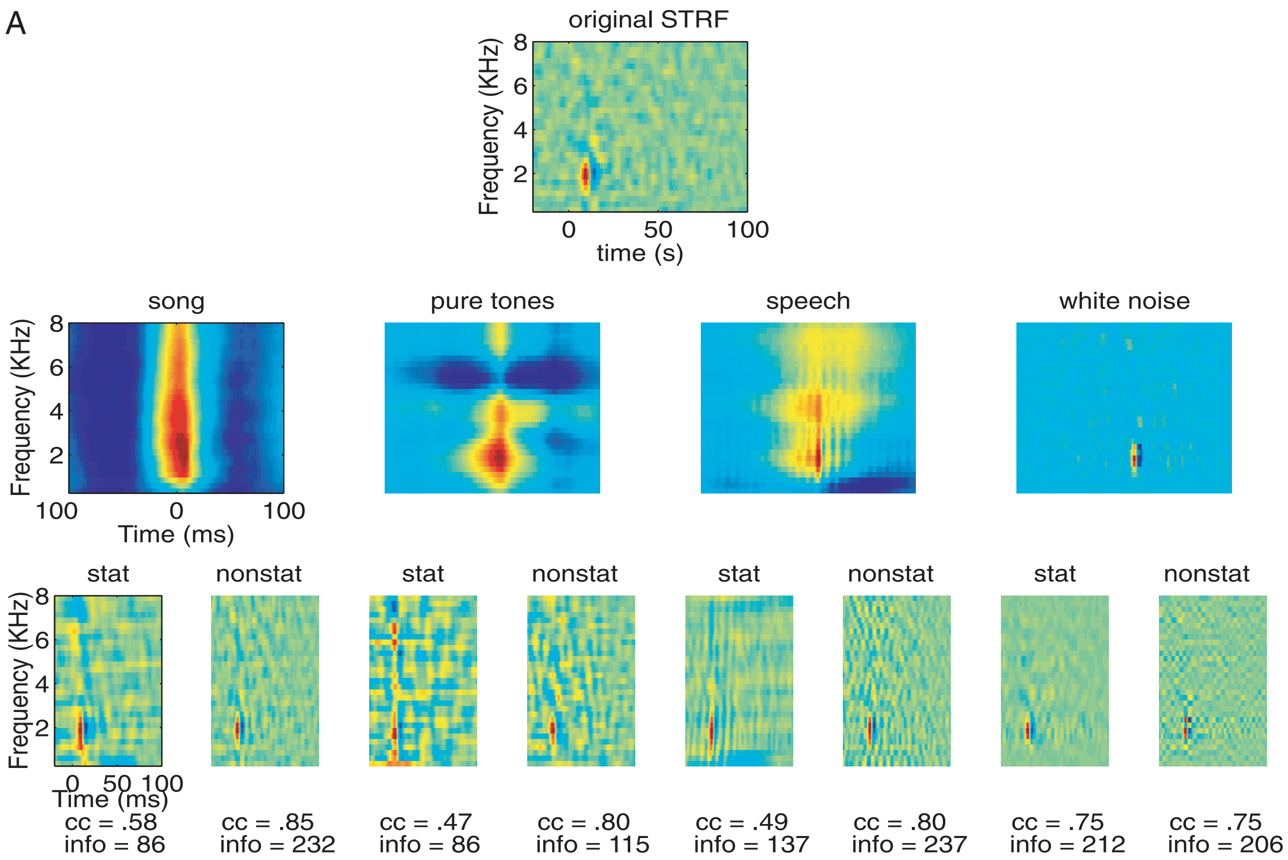

Figure 3. Analysis of two model auditory neurons. In panel A, the model neuron's STRF was obtained from a real auditory neuron in the auditory forebrain of a zebra finch. In panel B, the model neuron's STRF was a delta function centred at 0 ms and 2 kHz. Responses to four different stimulus ensembles, songs, tone pips, speech and white noise, were generated by using these STRFs as filters. The middle row in each panel shows the STA for each stimulus. The bottom row shows the estimated STRFs obtained assuming stationarity and non-stationarity. The information value and *cc*s are shown in the last rows. The *cc*s were calculated using a time bin of 5 ms for the song neuron STRF and a time bin of 3 ms for the delta function STRF.

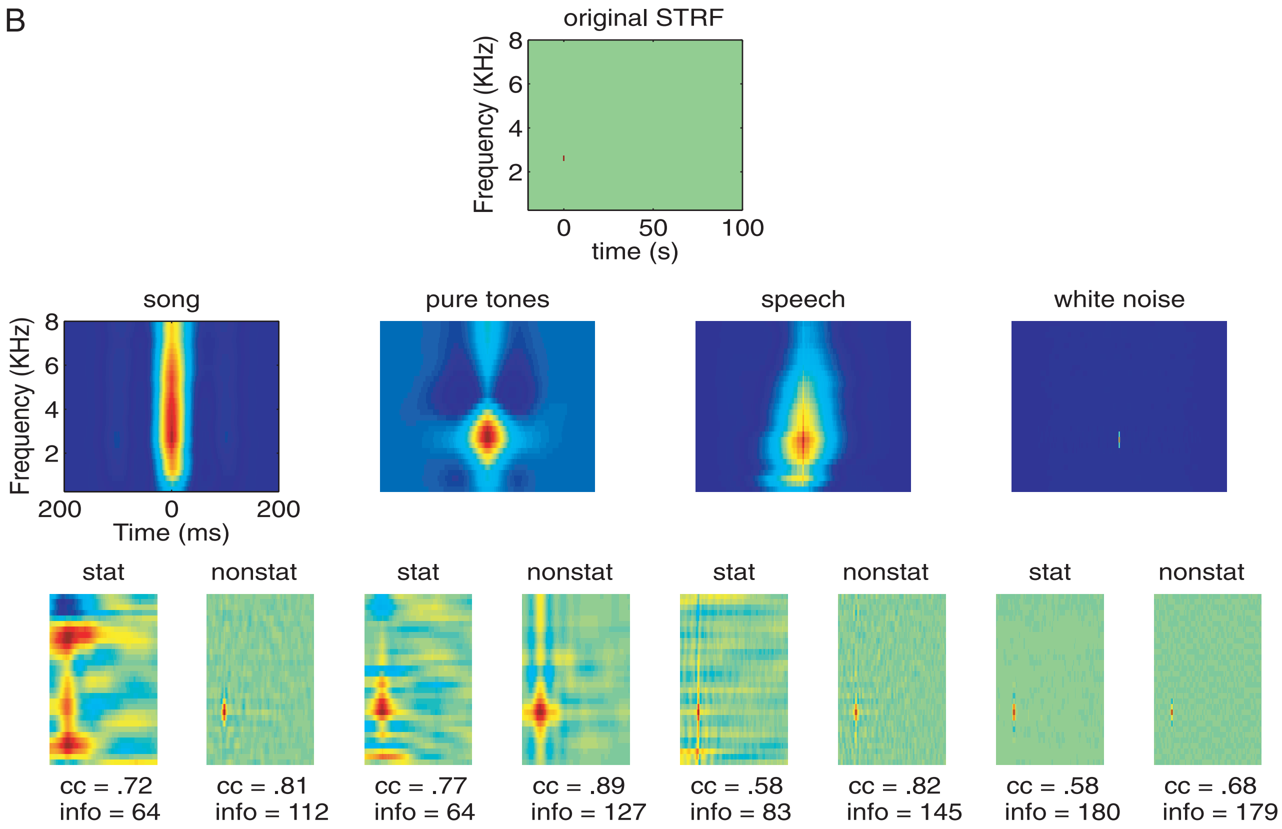

Figure 3. (Continued.)

Figure 2(*e*) shows the eigenvalues of the stimulus auto-correlation matrix as a function of temporal frequency. The y axis represents a different set of eigenvectors for each $\omega_t[i]$ and the eigenvectors are ordered by decreasing eigenvalue. The solid curve in figure 2(*d*) indicates the spatial and temporal frequency thresholds that are used when $\boldsymbol{h}$ is calculated using equation (15) for the model neuron presented in section 3.2.2. In figure 2(*e*) the solid line indicates the thresholds for the spatial eigenvectors at each temporal frequency when $\boldsymbol{h}$ is calculated for the same model neuron using equation (14).

3.2. Results of modelling experiments

To validate our methods and their implementation we estimated STRFs from data sets representing the responses of model neurons to various stimulus ensembles. In all cases the model neurons had linear STRFs but their output was governed by a stochastic (Poisson) spike generator.

3.2.1. Modelling experiments with auditory stimuli. Four stimulus ensembles were included in this analysis: zebra finch songs, tone pips, white noise (Gaussian and flat from 0 to 8000 Hz) and speech (20 short sentences from a database of 100 sentences, Tyler *et al* 1990). For each of the four classes, model responses were obtained to 100 repetitions of approximately 40 s each (20 songs, 22 s tone pip trains, 22 s noise bursts and 20 sentences). The mean firing rate of the model neurons was 20 spikes s^{-1} and the deviation from the mean was calculated by convolving the spectrographic representation of each sound with the model STRF. The gain of the STRF was adjusted so that its modulation also had a mean root square strength of 20 spikes s^{-1} (this ensured that the signal strength approximately matched noise strength).

Two model neurons were probed with these stimuli. The STRF of one model was obtained from a real auditory neuron in the forebrain of a zebra finch (top panel, figure 3(*a*)), while the second STRF was a simple delta function in space and time (single peak response at $t = 0$ ms and $x = 2000$ Hz, top panel figure 3(*b*)). The middle panels of figures 3(*a*) and (*b*) show the cross-correlation between the stimulus and the response for the four sound classes. As expected from theory, these cross-correlations are proportional to the original STRF only when the cell is probed with white noise. Correlations in the other stimulus classes spread the STRF in time and space. The bottom row of figures 3(*a*) and (*b*) shows the estimates of the model STRFs for the non-stationary (equation (14)) and stationary (equation (15)) cases. When the STRFs are derived from white noise the two estimates are similar and predict real model neural responses to novel stimuli equally well. In contrast, STRFs derived from non-white stimuli were always closer to the true model neural response if they did not assume stationarity. These non-stationary estimates were also best at predicting real neural responses to novel stimuli.

The STRF estimated for the delta function model (shown in figure 3(*b*)) illustrates another important point: if the true STRF has power in regions of the stimulus space that are not sampled in the experiment, then the estimated STRF may be a relatively poor model. The delta function STRF has high spatial and temporal frequencies that are not present in the song, tone-pip or speech ensembles. Therefore, STRFs estimated from responses to those three stimulus classes are low-pass versions of the true delta function STRF.

3.2.2. Modelling experiments with visual stimuli. Two stimulus classes were used in this analysis: movies simulating natural vision (described above and in Vinje and Gallant 2000) and dynamic grating sequences (Ringach *et al* 1997). In the dynamic grating sequences the orientation, spatial frequency and phase of the gratings were chosen randomly on each 72 Hz video frame. All gratings were shown at a contrast of 0.5. For both stimulus classes, model responses were obtained from 150 s of stimulation.

The model neuron consisted of a V1 complex cell implemented as a quadrature phase energy mechanism (Carandini *et al* 1997, Heeger 1992, Vinje and Gallant 1998). First, linear kernels were constructed at four spatial phases (each offset by 90°). Each kernel was a Gabor function in space multiplied by a weakly biphasic impulse response in time. Second, the inner product in space and the convolution in time was computed between each of the kernels and the stimulus sequence to form subunit outputs. Next, the four subunit outputs were half-wave rectified, averaged and normalized, to give the model's output. This output was used to modulate a Poisson spiking process, constrained so that the mean rate was 7 spikes s^{-1} and the standard deviation of the modulation signal was 8 spikes s^{-1}.

In order to recover tuning for both simple and complex cells, we transformed the stimulus into a spatial phase-separated Fourier domain representation before completing reverse correlation. Each frame of the movie was transformed into the spatial Fourier domain and then projected onto each of the positive and negative real and imaginary axes. This produced a representation of the stimulus which contained four phase channels at each spatial frequency, paralleling the quadrature phase structure of the complex cell model. Thus contributions from different spatial phases were not confounded during reverse correlation, and STRFs could be recovered regardless of a neuron's phase invariance properties (David *et al* 1999). For visualization, the visual STRFs shown in this paper have been collapsed over the phase dimension so that they indicate the relative power of the STRF at different spatial frequencies and latencies.

The results of the STRF estimation for the model neuron are shown in figure 4, in which each frame displays the Fourier domain STRF at one latency step. Figure 4(*a*) shows the actual STRF of the model cell used to generate data for the analysis. The Gabor structure of the STRF is clearly visible in the fourth and fifth frames. Figure 4(*b*) shows the STA estimated via simple reverse correlation of the model's responses to the dynamic grating sequence. Because the model is quasi-linear and the grating sequence is relatively white with respect to the cardinal dimensions of the model STRF (orientation and spatial frequency), the STA closely matches the true STRF. Figure 4(*c*) shows the STA estimated from the model's responses to natural vision movies. In this case the STA bears little resemblance to the true STRF, reflecting stimulus bias in the stimulus–response cross-correlation. Figures 4(*d*) and (*e*) illustrate the effect of correcting for the stimulus bias in the natural vision movies. In figure 4(*d*), the analysis assumed stationarity (equation (15)) by normalizing only by the diagonal elements of the stimulus auto-correlation matrix. This procedure removes much of the low-frequency bias but does not eliminate noise at high spatial frequencies. In figure 4(*e*), the analysis did not assume stationarity (equation (14)) by normalizing with the full stimulus auto-correlation matrix. This STRF estimate is very similar to the true STRF of the model neuron.

3.3. Results of real experiments

As a final check on the feasibility of our methods we used equations (14) and (15) to estimate STRFs of sensory neurons recorded in the avian forebrain and in primate area V1.

3.3.1. Real experiments on avian auditory neurons. We estimated the STRFs of seven auditory neurons in the auditory forebrain of three songbirds (details of physiological methods can be found in Theunissen *et al* 2000). Data were obtained from song and tone-pip stimuli like those described in section 3.1.1. We estimated two STRFs for each neuron using equations (14) and (15). Figure 5 shows the results obtained for one neuron from area L2a of the avian

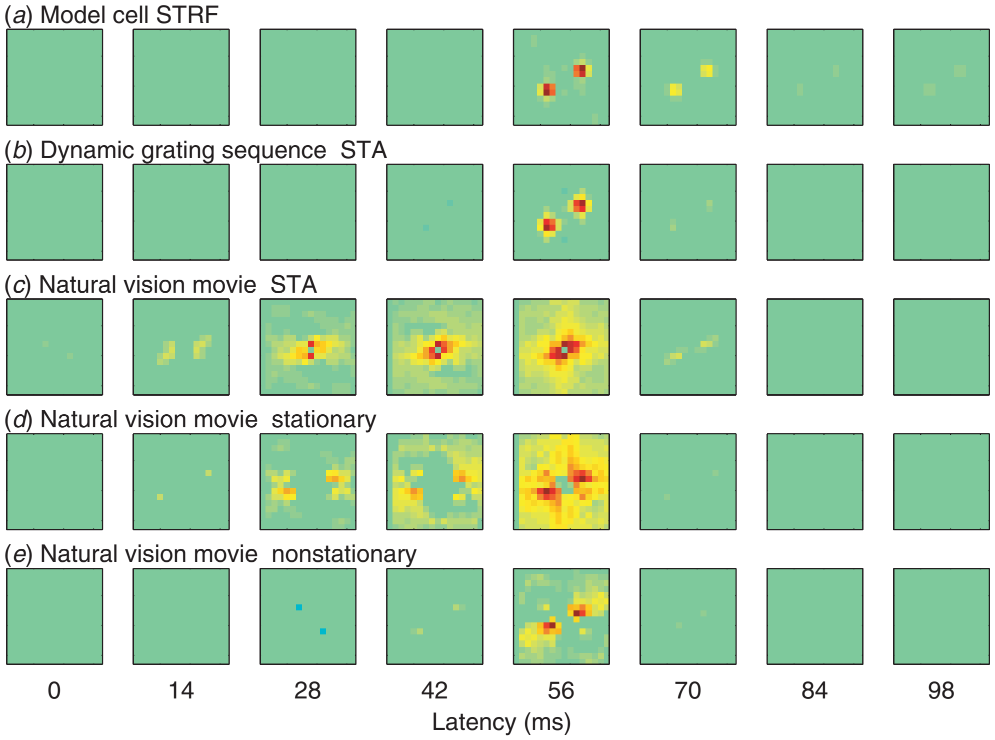

Figure 4. Analysis of a model complex cell constructed by summing the rectified output of four linear Gabor subunits in quadrature spatial phase. The model has a peak orientation of 60° from horizontal, a peak spatial frequency of 2 cycles per receptive field and a latency of 56 ms. Each panel displays the spatial Fourier power for one latency, and latency is plotted from left ($t = 0$) to right ($t = 7$ bins, 98 ms). Positive responses are shown by warm colours while negative responses are shown by cool colours. Stimuli spanned a region twice the diameter of the receptive field and were downsampled to 16×16 pixels before analysis, so that effective spatial frequency of the estimated STRF ranged from 0.5 to 4 cycles per receptive field. (*a*) STRF of the model neuron. In frames $t = 56$ and 70 ms the Gabor structure of the model neuron's STRF is clearly visible. This STRF was used as a linear filter to generate model responses to dynamic grating sequences and natural vision movies. (*b*) STA calculated from the model's responses to 150 s of stimulation by a dynamic grating sequence. The STA is similar to the model's true STRF. (*c*) STA calculated from the model's responses to 150 s of stimulation by a natural vision movie. The low-frequency correlations present in the stimulus corrupt the STA and make it impossible to estimate the model's STRF without normalization. (*d*) STRF estimated from the model's responses to a natural vision movie, assuming stationarity (equation (15)). Although the orientation of the estimated STRF is similar to the true STRF, the quality of the estimate is quite low. (*e*) STRF estimated from the model's responses to a natural vision movie after full normalization (equation (14)). This estimate corresponds closely to the true structure of the model STRF.

auditory forebrain. The left-hand panels show the cross-correlation between the stimulus and the response, the middle panels show the estimated STRF obtained assuming stationarity and the right-hand panels show the estimated STRF obtained without assuming stationarity. The estimated STRFs are clearly different. Most notably, the STRF estimated by assuming stationarity (equation (15)) is more distributed in space and time than the one obtained without this assumption (equation (14)). The distribution of energy in the STRF has a substantial effect on its predictive power. As shown in figure 6, STRFs estimated without assuming stationarity were always more predictive than those obtained under the more restrictive assumption.

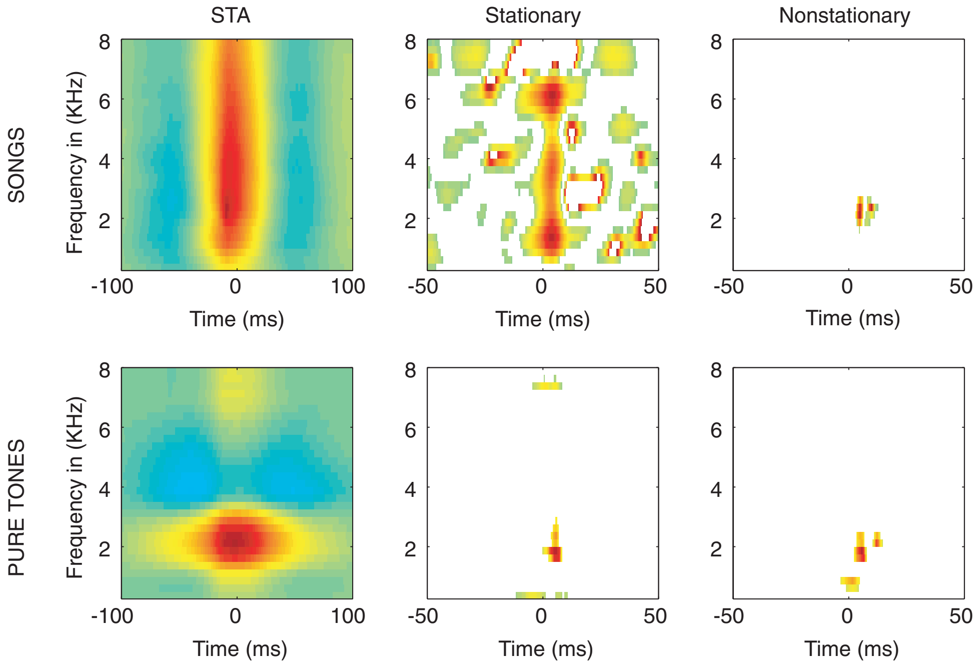

Figure 5. Results of STRF estimation for an auditory neuron recorded from area L2a of the zebra finch auditory forebrain. Data were obtained in response to both zebra finch song (top) and tone pips (bottom). The left-hand panels show the STA. The centre panels show the STRF obtained assuming stationarity (equation (15)). The right-hand panels show the STRF obtained with full normalization (equation (14)).

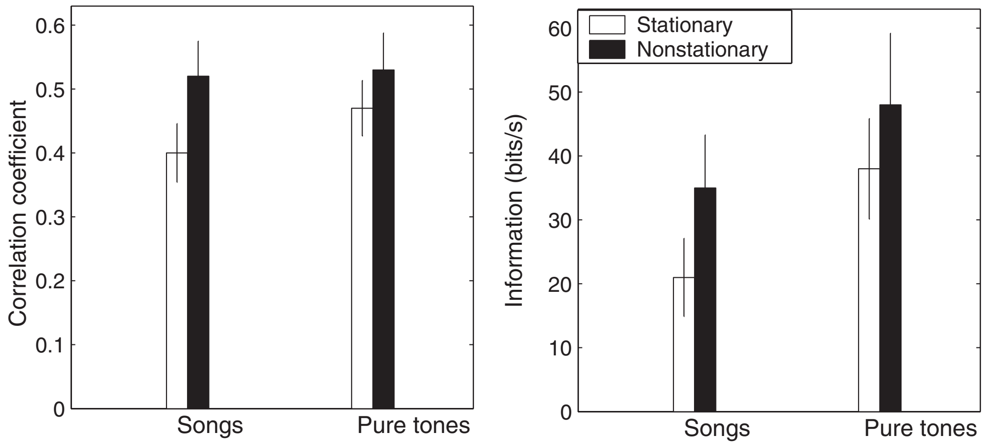

Figure 6. Predictive power of auditory STRFs to account for responses to novel song stimuli. Prediction quality is quantified in terms of the *cc* and the information. Prediction quality is shown for both stationary and non-stationary STRFs obtained from responses to songs and tone pips. Error bars correspond to one standard error. A two-tailed paired t-test shows that the difference between the stationary and non-stationary goodness of fit is significant both for songs and tones and in both for the information or correlation measures (info for songs $p = 0.0077$; info for pure tones $p = 0.033$; cc for songs $p = 0.0214$; cc for tones $p = 0.0004$).

There were also clear differences between STRF estimates obtained with songs versus tone pips. These stimulus classes have equivalent second-order statistics, the differences between the estimated STRFs. Therefore, these differences reflect nonlinear response properties, such as adaptation, that are elicited by the different higher-order statistics of the two stimulus ensembles (Theunissen *et al* 2000).

3.3.2. Real experiments on STRF for primate cortical visual neurons. We also estimated the STRFs of neurons in area V1 using both dynamic grating sequences (250 s, 1 repetition) and natural vision movies (40 s, 30 repetitions) like those described in section 3.1.2 (details of physiological methods can be found in Vinje and Gallant 2000). Figure 7 summarizes the results obtained from one V1 complex cell. The STA estimated from the neuron's responses to the dynamic grating sequence is shown in figure 7(*a*). For this stimulus, power is uniformly distributed throughout the Fourier domain, so the STA is proportional to the true STRF. The STRF shows a clear positive response beginning at a latency of 42 ms. The corresponding STA obtained with natural vision movies is shown in figure 7(*b*). This STA is clearly different from the STRF estimated using the dynamic grating sequence, and the difference is probably due to the large spatio-temporal correlations in the natural vision movie. Figures 7(*c*) and (*d*) show the estimated STRFs obtained via equations (15) and (14), respectively. The STRF obtained by assuming stationarity (figure 7(*c*)) produces an estimate of the STRF more similar than the STA alone to the STRF estimated from dynamic grating sequences. However, the correspondence between STRFs estimated from responses to the two stimulus ensembles appears to be substantially improved by use of equation (14) (figure 7(*d*)).

We also examined how well each STRF estimate could predict neural responses to novel natural vision movies. The *cc*s between STRF predictions and data were: 0.48 for the STA obtained from natural vision movies (figure 7(*b*)), 0.67 for the STRF estimated assuming stationarity (figure 7(*c*)) and 0.84 for the STRF estimated without assuming stationarity obtained with equation (14) (figure 7(*d*)). Thus, STRF estimates were substantially improved by assuming non-stationarity when normalizing the STA.

As in the auditory experiments, there are some pronounced differences between STRFs estimated from these two stimulus classes. For example, the STRF estimated from dynamic grating sequences (figure 7(*a*)) is tuned to a higher spatial frequency than is the STRF estimated with natural vision movies (figure 7(*d*)). Because the power spectrum of natural scenes is biased toward low spatial frequencies we should expect to see a low-frequency bias in the STA (figure 7(*b*)). However, this difference remains even after normalization and regardless of the stationarity assumption. The remaining difference could be an artifact of the thresholding procedure, which tends to exclude high-frequency eigenvectors that have relatively low signal power. Alternatively, it might reflect nonlinear response properties elicited by naturalistic stimuli.

The STRFs estimated from the two stimulus classes also have different temporal responses. The STRF estimated from dynamic grating sequences has a positive monophasic response, while that estimated from natural vision movies has a biphasic response (compare figures 7(*a*) and (*d*)). This effect is probably due to the temporal properties of natural vision. The dynamic grating stimuli were updated at 72 Hz, while natural vision movies changed at about 3 Hz to simulate fixations and saccades. The slower time course of the natural vision movies possibly elicits nonlinear adaptation and produces a biphasic STRF.

Taken together, the differences in spatial frequency tuning and temporal response profile can have a large effect on the ability of STRFs obtained with different stimulus ensembles to predict neural responses to novel natural vision movies. The dynamic grating STA achieved a correlation of 0.52 between predicted and observed responses, whereas the STRF estimated from natural vision movies using equation (14) achieved a correlation of 0.84.

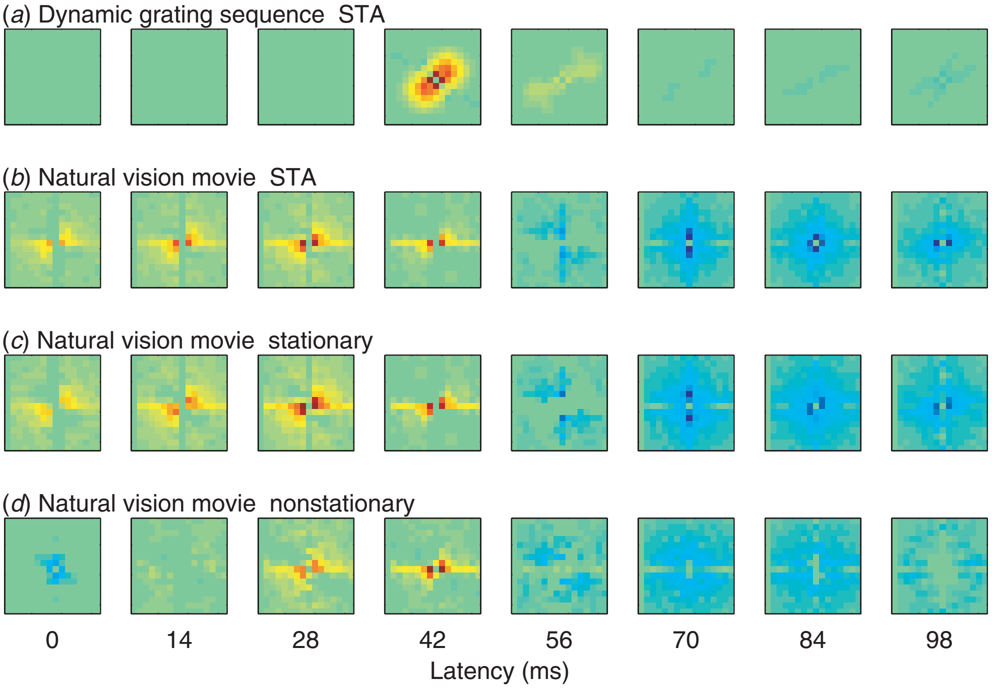

Figure 7. Results of STRF estimation for a V1 neuron. Axes and layout as in figure 3. The spatial receptive field of this neuron was 0.36° and the stimuli covered an area twice this size. Thus, each spatial frequency bin of the STRF corresponds to 1.4 cycles per degree. (*a*) STA obtained from the neuron's responses to a dynamic grating sequence (250 s, 1 repetition). Mostly positive response tuning is evident, and the temporal response is monophasic. (*b*) STA calculated from the neuron's responses to a natural vision movie (40 s, 30 repetitions). In this case the STA is corrupted by the low-frequency stimulus bias. (*c*) STRF estimated from the neuron's responses to a natural vision movie assuming stationarity (equation (15)). The tuning of the estimate is similar to the STA obtained with gratings, but it is still affected by stimulus bias. (*d*) STRF estimated with full normalization (equation (14)) from the neuron's responses to a natural vision movie. The spatial profile of this estimate is quite similar to the STA obtained with the dynamic grating sequence, though the temporal response profile is different. The ability of this STRF to predict neural responses to novel natural vision movies is substantially higher than the predictive power of the STA obtained with dynamic gratings. The prediction correlation is 0.84 for the STRF estimated using equation (14) versus only 0.52 for the grating STA.

4. Discussion

We have presented a generalized reverse correlation method that can be used to estimate STRFs from neural responses to arbitrary stimulus ensembles, including natural stimuli. The method involves normalizing the cross-correlation between the stimulus and the response by the stimulus auto-correlation matrix. Because the stimulus auto-correlation matrix is symmetric, this normalization can be done in the eigenspace of the matrix.

An important consideration involves whether the spatial and temporal statistics of the stimulus are stationary or non-stationary. When the second-order statistics of the stimulus are stationary, the eigenspace is the Fourier domain and the normalization is equivalent to dividing by the power of the stimulus at each frequency. When the statistics are non-stationary a different eigenspace is found, but the normalization is mathematically homologous. We emphasized a particular case in which the temporal statistics are stationary but the spatial statistics are not. In this case the eigenvectors are found by diagonalizing the matrix of the spatial correlations for each temporal frequency (equation (14)).

Normalization of non-stationary spatial statistics is particularly important when STRFs are to be estimated from natural stimuli. The spatial statistics of natural sounds such as birdsong or speech are inherently non-stationary, while natural vision movies are non-stationary due to experimental constraints on visual stimuli.

When stimuli have non-stationary spatial statistics, optimal STRF estimation requires a different normalization at each temporal frequency (equation (14)). This is a computationally demanding procedure because it requires the calculation of a different eigenspace at each temporal frequency. Thus, we compared the results obtained in this full normalization with those obtained with the approximate, stationary solution, as shown in equation (15). In this approximation the same eigenspace (the spatial Fourier domain) is used for all temporal frequencies, reducing the computation load of the analysis. Our model simulations, analyses of neuronal data and evaluation of predictive power all clearly demonstrate that normalization assuming non-stationary spatial statistics produces substantially better STRF estimates than when stationarity is assumed.

An important methodological question concerns the choice of threshold for the exclusion of some of the eigenvectors of the auto-correlation matrix prior to inversion. If the magnitude of a given eigenvector is relatively large then it is likely to be important for the normalization process. However, eigenvectors of small magnitude are more likely to reflect noise and sampling limitations; these will tend to corrupt STRF estimates if they are included in the normalization process. We set thresholds according to an iterative procedure that uses the predictive power of the estimated STRF as its success criterion. This approach works well whether the stimulus statistics are stationary or not.

The same procedure allows one to identify the spatial and temporal range of the experimental stimulus ensemble (solid lines in figures 1 and 2). Complex stimuli are necessarily band limited, regardless of whether they are natural or synthetic. In such cases a neuron's STRF may extend outside the spatio-temporal range being sampled in a particular experiment. This was demonstrated in section 3.2.1, where it was impossible to completely recover the delta-function STRF because of its high spatial and temporal frequency power. On the other hand, if the spatio-temporal power of the STRF is much more restricted than the stimulus space being sampled (as might occur when broad-band white noise is used), the power density of the stimulus could be insufficient to drive the neuron (Klein *et al* 2000). One argument in favour of natural stimuli in STRF estimation is that such stimuli are likely to closely match stimulus space sampled by a sensory neuron. It is therefore always important to describe the spatio-temporal range of experimental stimuli. One should be particularly careful when comparing STRFs obtained from different stimulus classes (such as natural versus synthetic), because apparent differences in the STRFs might be due to differences in the sampling range of the various stimulus classes.

Although this paper has emphasized methodological issues, we have successfully applied these techniques to the auditory system of songbirds (Theunissen *et al* 2000) and the visual system of primates (David *et al* 1999). In some cases the STRFs estimated from natural stimuli are similar to those those obtained with simpler synthetic stimuli (David *et al* 1999), while in other cases they may be quite different (Theunissen *et al* 2000). This suggests that natural stimuli are a critical complement to synthetic stimuli in sensory experiments. By combining the STRF estimation methods presented here with simpler methods appropriate for synthetic stimuli (e.g. Klein *et al* 2000, Ringach *et al* 1997), it should be possible to bridge the gap between qualitative neuro-ethological approaches and more quantitative neuro-physiological approaches.

Acknowledgments

This work was supported by grants from the Sloan and Searle Foundations and the NIMH to FET, and grants from the Sloan and Whitehall Foundations and the NEI to JLG. The authors would like to thank Achim Klug and two anonymous reviewers for their constructive comments on a previous version of the paper.

Appendix

Proof 1. Solution of equation (1) for the one-dimensional case and for stationary statistics. We wish to obtain an equation for the one-dimensional impulse response function $h(t)$ that yields the best prediction of the neural response, $\hat{r}(t)$, in the minimum-mean-square sense. $\hat{r}(t)$ is given by

$$\hat{r}(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)\,\mathrm{d}\tau.$$

Invoking the convolution theorem:

$$\hat{R}(\omega_t) = H(\omega_t)S(\omega_t).$$

Parseval's theorem asserts that it is equivalent to minimize

$$\left\langle \left(r(t)-\hat{r}(t)\right)^2 \right\rangle_t \qquad \text{and} \qquad \left\langle \left(R(\omega_t)-\hat{R}(\omega_t)\right)^2 \right\rangle_{\omega_t},$$

where the first estimate of the mean-square error is taken over time points, t, and the second estimate is taken over the Fourier frequencies, ω_t:

$$\left\langle \left(R(\omega_t)-\hat{R}(\omega_t)\right)^2 \right\rangle_{\omega_t} = \left\langle \left(R(\omega_t)-H(\omega_t)S(\omega_t)\right)^2 \right\rangle_{\omega_t}.$$

Taking the complex derivative with respect to $H(\omega_t)$ and setting it to zero:

$$\frac{\mathrm{d}}{\mathrm{d}H(\omega_t)} \left\langle \left(R(\omega_t)-H(\omega_t)S(\omega_t)\right)^2 \right\rangle_{\omega_t} = 0.$$

If we suppress the ω_t argument, the equation can be written as

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}H} \left\langle RR^* - R^*HS - RH^*S^* + HH^*SS^* \right\rangle_{\omega_t} = 0\\
&\left\langle H^*SS^* + H^*S^*S - SR^* - SR^* \right\rangle_{\omega_t} = 0\\
&\left\langle H(\omega_t)S(\omega_t)S^*(\omega_t) \right\rangle_{\omega_t} = \left\langle S^*(\omega_t)R(\omega_t) \right\rangle_{\omega_t}.
\end{aligned}$$

Here the complex derivative is defined as

$$\frac{\mathrm{d}}{\mathrm{d}H} = \frac{\mathrm{d}}{\mathrm{d}H_\mathrm{r}} + \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}H_\mathrm{i}},$$

where H_r and H_i are the real and complex part of H. If the second-order statistics of the stimulus are stationary, then the averages taken across ω_t for the entire data set are identical to averages taken over ω_t for segments of the data, denoted as D:

$$\left\langle \left\langle H(\omega_t)S^*(\omega_t)S(\omega_t) \right\rangle_{\omega_t} \right\rangle_D = \left\langle \left\langle S^*(\omega_t)R(\omega_t) \right\rangle_{\omega_t} \right\rangle_D.$$

The unique solution of the previous equation for arbitrary functions $S(\omega_t)$ and $R(\omega_t)$ is given by

$$H(\omega_t) = \frac{\left\langle S^*(\omega_t)R(\omega_t) \right\rangle_\mathrm{S}}{\left\langle S^*(\omega_t)S(\omega_t) \right\rangle_\mathrm{S}}.$$

Proof 2. Cross-covariance and spike-triggered average stimulus. We want to show that the cross-covariance between the stimulus and the response is equal to the STA stimulus, given that the stimulus, $s(t)$, has zero mean and the neural response consists of spiking events.

The cross-covariance between $s(t)$ and $r(t)$ is given by

$$C_{\mathrm{sr}}(\tau) = \frac{1}{2T}\int_{-T}^{\mathrm{T}} (r(t) - \bar{r})s(t+\tau)\,\mathrm{d}t,$$

or in discrete form

$$C_{\mathrm{sr}}(\mathrm{d}t) = \frac{1}{N}\sum_{i=0}^{N-1} (r(i) - \bar{r})s(\mathrm{i} + \mathrm{d}t).$$

In the above equations $\bar{r}$ is the mean of r and we assume that the stimulus has zero mean. The neural response, $r(i)$, consists a set of times $\{j\}$ where we recorded a spike, $r(i \in \{j\}) = 1$, and other times where we recorded no spikes, $r(i \notin \{j\}) = 0$. Therefore,

$$C_{\mathrm{sr}}(\mathrm{d}t) = \frac{1}{N}\left(\sum_{i\in\{j\}} (1-\bar{r})s(\mathrm{i}+\mathrm{d}t) + \sum_{i\notin\{j\}} -\bar{r}s(\mathrm{i}+\mathrm{d}t)\right).$$

Since $\sum_{i\in\{j\}} s(\mathrm{i}+\mathrm{d}t) + \sum_{i\notin\{j\}} s(\mathrm{i}+\mathrm{d}t) = 0$,

$$C_{\mathrm{sr}}(\mathrm{d}t) = \frac{1}{N}\left(\sum_{i\in\{j\}} s(\mathrm{i}+\mathrm{d}t)\right).$$

Therefore, in this case the cross-covariance between the stimulus and response is given simply by the STA stimulus.

Proof 3. Solution to the general LMMSE. We want a solution for a vector of coefficients, $\boldsymbol{h}$, that describes the linear transform between the stimulus $\boldsymbol{s}$ and the response $\boldsymbol{r}$:

$$\hat{r}[t] = \sum_{i=0}^{M\times N-1} h[i]s_t[i].$$

A solution for $\boldsymbol{h}$ is found by minimizing the mean square difference between the predicted response, $\hat{r}$, and the actual response, r. The derivation is the linear algebra equivalent of the derivation in proof 1 (see also Kay 1993). Expanding the mean-square difference

$$\begin{aligned}\langle(\hat{r} - r)^2\rangle &= \langle(\boldsymbol{h}^{\mathrm{T}}\boldsymbol{s} - r)^2\rangle \\ &= \langle \boldsymbol{h}^{\mathrm{T}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{h} - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{s}r - r\boldsymbol{s}^{\mathrm{T}}\boldsymbol{h} + r^2\rangle \\ &= \langle \boldsymbol{h}^{\mathrm{T}}\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{h}\rangle - \langle \boldsymbol{h}^{\mathrm{T}}\boldsymbol{s}r\rangle - \langle r\boldsymbol{s}^{\mathrm{T}}\boldsymbol{h}\rangle + \langle r^2\rangle \\ &= \boldsymbol{h}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{ss}}\boldsymbol{h} - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{sr}} - \boldsymbol{C}_{\mathrm{rs}}\boldsymbol{h} + \mathrm{var}(r)\end{aligned}$$

where $C_{\mathrm{ss}} = \left\langle ss^{\mathrm{T}}\right\rangle$ is the stimulus auto-correlation matrix and $C_{\mathrm{sr}} = \langle sr\rangle$ is the stimulus–response cross-correlation vector

$$\boldsymbol{C}_{\mathrm{ss}} = \left\langle \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\right\rangle = \begin{pmatrix} \langle s[t-0]s[t-0]\rangle & \cdots & \langle s[t-0]s[t-((NM)-1)]\rangle \\ \vdots & \ddots & \vdots \\ \langle s[t-((NM)-1)]s[t-0]\rangle & \cdots & \langle s[t-((NM)-1)]s[t-((NM)-1)]\rangle \end{pmatrix}$$

and

$$\boldsymbol{C}_{\mathrm{sr}} = \left\langle \boldsymbol{s}r\right\rangle = \begin{pmatrix} \langle s[t-0]r[t]\rangle \\ \vdots \\ \langle s[t-((NM)-1)]r[t]\rangle \end{pmatrix}.$$

To find the best $\boldsymbol{h}$, we the take the derivative of $\left\langle(\hat{r}-r)^2\right\rangle$ with respect to $\boldsymbol{h}$ and set it to zero:

$$\frac{\mathrm{d}\left\langle(\hat{r}-r)^2\right\rangle}{\mathrm{d}\boldsymbol{h}} = 2\boldsymbol{h}\boldsymbol{C}_{ss} - 2\boldsymbol{C}_{sr} = 0,$$

where we have used the following linear algebra equations:

$$\frac{\mathrm{d}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x} \qquad \text{and} \qquad \frac{\mathrm{d}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{b}.$$

Therefore,

$$\boldsymbol{h} = \boldsymbol{C}_{ss}^{-1}\boldsymbol{C}_{sr}.$$

Proof 4. Eigenvectors and eigenvalues of a symmetric Toeplitz matrix. When the stimulus statistics are stationary, the stimulus auto-correlation matrix can be written as

$$\boldsymbol{C}_{\mathrm{ss}} = \begin{bmatrix} c[0] & c[1] & \cdots & c[N-1] \\ c[1] & c[0] & \cdots & c[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ c[N-1] & c[N-2] & \cdots & c[0] \end{bmatrix}.$$

This matrix is symmetric Toeplitz. In addition the correlation data can be organized to have period N (i.e. $c[N-1]$ is set equal to $c[1]$ and so forth until $c[N-N/2] = c[N/2]$). In such cases, the memory of the system is given by $N/2$ data points in time and the data are repeated twice for mathematical convenience.

The N eigenvectors of a such a symmetric Toeplitz matrix of period N are then given by

$$\boldsymbol{q}_k = \frac{1}{\sqrt{N}}\left[1, \exp\left(\mathrm{j}\frac{2\pi k}{N}\right), \exp\left(\mathrm{j}\frac{2\pi k}{N}2\right), \ldots, \exp\left(\mathrm{j}\frac{2\pi k}{N}(N-1)\right)\right]^{\mathrm{T}},$$

where k goes from 0 to $N/2$ and

$$\boldsymbol{q}_k = \frac{1}{\sqrt{N}}\left[1, \exp\left(\mathrm{j}\frac{2\pi(k-N)}{N}\right),\right.$$
$$\left.\exp\left(\mathrm{j}\frac{2\pi(k-N)}{N}2\right), \ldots, \exp\left(\mathrm{j}\frac{2\pi(k-N)}{N}(N-1)\right)\right]^{\mathrm{T}}$$

for k from $N/2$ to $N-1$.

These $\boldsymbol{q}$ eigenvectors are the coefficients of the DFT and the corresponding eigenvalues give the power of the stimulus at each frequency $\omega_t[k] = \frac{2\pi k}{N}$:

$$\lambda_k = P_{\mathrm{s}}(\omega_t[k]) = \sum_{i=0}^{N-1} c[i]\exp(\mathrm{j}\omega_t[k]i).$$

This property can be shown by first verifying

$$\boldsymbol{C}_{\mathrm{ss}}\boldsymbol{q}_k = \lambda_k\boldsymbol{q}_k.$$

Expanding $\boldsymbol{C}_{\mathrm{ss}}\boldsymbol{q}_k$ for $k \in [0, N/2]$,

$$\boldsymbol{C}_{ss}\boldsymbol{q}_k = \frac{1}{\sqrt{N}}\begin{bmatrix} c[0] + c[1]\exp\left(\frac{\mathrm{j}2\pi k}{N}\right) + \cdots + c[N-1]\exp\left(\frac{\mathrm{j}2\pi k(N-1)}{N}\right) \\ c[1] + c[0]\exp\left(\frac{\mathrm{j}2\pi k}{N}\right) + \cdots + c[N-2]\exp\left(\frac{\mathrm{j}2\pi k(N-1)}{N}\right) \\ \vdots \\ c[N-1] + c[N-2]\exp\left(\frac{\mathrm{j}2\pi k}{N}\right) + \cdots + c[0]\exp\left(\frac{\mathrm{j}2\pi k(N-1)}{N}\right) \end{bmatrix}$$

$$\boldsymbol{C}_{ss}\boldsymbol{q}_k = \frac{\lambda_k}{\sqrt{N}}\begin{pmatrix} 1 \\ \exp\left(\frac{\mathrm{j}2\pi k}{N}\right) \\ \vdots \\ \exp\left(\frac{\mathrm{j}2\pi k(N-1)}{N}\right) \end{pmatrix}$$

where

$$\lambda_k = r[0] + r[1]\exp\left(\frac{\mathrm{j}2\pi k}{N}\right) + \cdots + r[N-1]\exp\left(\frac{\mathrm{j}2\pi k(N-1)}{N}\right).$$

A similar expansion can be performed for the negative frequencies (for $k \in [N/2, N]$). Therefore, we verified that

$$\boldsymbol{C}_{ss}\boldsymbol{q}_k = \lambda_k\boldsymbol{q}_k,$$

for the $\boldsymbol{q}_k$ corresponding to the DFT. Since these N vectors are orthonormal, we have also shown that they constitute a complete set of eigenvectors for $\boldsymbol{C}_{\mathrm{ss}}$.

Proof 5. Separation of spatial and temporal dimensions in equation (9) for stationary temporal statistics. For stationary temporal statistics equation (9) can be rewritten as equation

$$\Lambda_t\boldsymbol{H} = \boldsymbol{C}_{\mathrm{SR}}, \tag{A.5.1}$$

where

$$\boldsymbol{H} = \boldsymbol{Q}_t^{-1}\boldsymbol{h}, \qquad C_{\mathrm{SR}} = \boldsymbol{Q}_t^{-1}\boldsymbol{C}_{\mathrm{sr}}, \qquad \boldsymbol{Q}_t = \begin{pmatrix} \boldsymbol{Q}_{\mathrm{FT}} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{Q}_{\mathrm{FT}} \end{pmatrix}$$

and

$$\Lambda_t = \begin{pmatrix} \Lambda_{0,0} & \cdots & \Lambda_{0,M-1} \\ \vdots & \ddots & \\ \Lambda_{M-1,0} & & \Lambda_{M-1,M-1} \end{pmatrix}.$$

Λ_t is composed of $N\times N$ diagonal submatrices $\Lambda_{i,j}$ describing the cross-correlation of stimulus spatial dimension i with stimulus spatial dimension j, for all N temporal frequencies, $\omega[i]$:

$$\Lambda_t = \begin{pmatrix} \lambda_{0,0,\omega[0]} & 0 & 0 & & \lambda_{0,M-1,\omega[0]} & 0 & 0 \\ 0 & \ddots & 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \lambda_{0,0,\omega[N-1]} & & 0 & 0 & \lambda_{0,M-1,\omega[N-1]} \\ & \vdots & & \ddots & & \vdots & \\ \lambda_{M-1,0,\omega[0]} & 0 & 0 & & \lambda_{M-1,M-1,\omega[0]} & 0 & 0 \\ 0 & \ddots & 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \lambda_{M-1,0,\omega[N-1]} & & 0 & 0 & \lambda_{M-1,M-1,\omega[N-1]} \end{pmatrix}.$$

The form of Λ_t tells us that all the correlations between stimulus points with different temporal frequencies are 0, which is a direct consequence of the stationarity of the temporal statistics. By changing the order of the rows of Λ_t, equation (A.5.1) can be arranged along its temporal frequency components, $\omega_t[i]$. In this case Λ_t becomes block diagonal:

$$\Lambda_t = \begin{pmatrix} \lambda_{0,0,\omega[0]} & \cdots & \lambda_{0,M-1,\omega[0]} & & & & \\ \vdots & \ddots & \vdots & \cdots & & 0 & \\ \lambda_{M-1,0,\omega[0]} & \cdots & \lambda_{M-1,M-1,\omega[0]} & & & & \\ & \vdots & & \ddots & & \vdots & \\ & & & & \lambda_{0,0,\omega[N-1]} & \cdots & \lambda_{0,M-1,\omega[N-1]} \\ & 0 & & \cdots & \vdots & \ddots & \vdots \\ & & & & \lambda_{M-1,0,\omega[N-1]} & \cdots & \lambda_{M-1,M-1,\omega[N-1]} \end{pmatrix}$$

$$= \begin{pmatrix} \Lambda(\omega_t[0]) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Lambda(\omega_t[N-1]) \end{pmatrix}$$

and therefore equation (A.5.1) can be written as N independent linear equations:

$$\Lambda\left(\omega_t[k]\right) \boldsymbol{H}\left(\omega_t[k]\right) = \boldsymbol{C}_{\mathrm{SR}}\left(\omega_t[k]\right),$$

where

$$\Lambda(\omega_t[k]) = \begin{pmatrix} \lambda_{0,0,\omega[k]} & \cdots & \lambda_{0,M-1,\omega[k]} \\ \vdots & \ddots & \vdots \\ \lambda_{M-1,0,\omega[k]} & \cdots & \lambda_{M-1,M-1,\omega[k]} \end{pmatrix}.$$

References

Aertsen A M and Johannesma P I 1981a A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli *Biol. Cybern.* **42** 145–56

——1981b The spectro-temporal receptive field. A functional characteristic of auditory neurons *Biol. Cybern.* **42** 133–43

Borst A and Theunissen F E 1999 Information theory and neural coding *Nat. Neurosci.* **2** 947–57

Carandini M, Heeger D J and Movshon J A 1997 Linearity and normalization in simple cells of the macaque primary visual cortex *J. Neurosci.* **17** 8621–44

Dan Y, Atick J J and Reid R C 1996 Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory *J. Neurosci.* **16** 3351–62

David S V, Vinje W E and Gallant J L 1999 Natural image reverse correlation in awake behaving primates *Soc. Neurosci. Abstr.* **25** 1935

Davis P J 1979 *Circulant Matrices* (New York: Wiley)

DeAngelis G C, Ohzawa I and Freeman R D 1995 Receptive-field dynamics in the central visual pathways *Trends Neurosci.* **18** 451–8

DeBoer E and Kuyper P 1968 Triggered Correlation *IEEE Trans. Biomed. Eng.* **15** 159–79

DiCarlo J J and Johnson K O 2000 Spatial and temporal structure of receptive fields in primate somatosensory area 3b: effects of stimulus scanning direction and orientation *J. Neurosci.* **20** 495–510

Dong D W and Atick J J 1995 Statistics of natural time-varying images *Network* **6** 345–58

Field D J 1987 Relations between the statistics of natural images and the response properties of cortical cells *J. Opt. Soc. Am.* A **4** 2379–94

Heeger D J 1992 Normalization of cell responses in cat striate cortex *Vis. Neurosci.* **9** 181–97

Kay S 1993 *Fundamentals of Statistical Signal Processing: Estimation Theory* vol 1 (Upper Saddle River, NJ: Prentice-Hall)

Klein D J, Depireux D A, Simon J Z and Shamma S A 2000 Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design *J. Comput. Neurosci.* **9** 85–111

Margoliash D 1986 Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow *J. Neurosci.* **6** 1643–61

Marmeralis P and Marmeralis V 1978 *Analysis of Physiological Systems: The White Noise Approach* (New York: Plenum)

Nelken I, Kim P J and Young E D 1997 Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models *J. Neurophysiol.* **78** 800–11

Ohlemiller K K, Kanwal J S and Suga N 1996 Facilitative responses to species-specific calls in cortical FM–FM neurons of the mustached bat *Neuroreport* **7** 1749–55

Rauschecker J P, Tian B and Hauser M 1995 Processing of complex sounds in the macaque nonprimary auditory cortex *Science* **268** 111–4

Rieke F, Bodnar D A and Bialek W 1995 Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents *Proc. R. Soc.* B **262** 259–65

Rieke F, Warland D, de Ruyter van Steveninck R and Bialek W 1997 *Spikes: Exploring the Neural Code* (Cambridge, MA: MIT Press)

Ringach D L, Sapiro G and Shapley R 1997 A subspace reverse-correlation technique for the study of visual neurons *Vis. Res.* **37** 2455–64

Ruderman D L 1997 Origins of scaling in natural images *Vis. Res.* **37** 3385–98

Suga N, O'Neill W E and Manabe T 1978 Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the moustache bat *Science* **200** 778–81

Theunissen F, Roddey J C, Stufflebeam S, Clague H and Miller J P 1996 Information theoretic analysis of dynamical encoding by four identified primary sensory interneurons in the cricket cercal system *J. Neurophysiol.* **75** 1345–64

Theunissen F E, Sen K and Doupe A J 2000 Spectral–temporal receptive fields of nonlinear auditory neurons obtained using natural sounds *J. Neurosci.* **20** 2315–31

Thomson D J and Chave A D 1991 Jackknifed error estimates for spectra, coherences and transfer functions *Advances in Spectrum Analysis and Array Processing* vol 1, ed S Haykin (Upper Saddle River, NJ: Prentice-Hall)

Tyler, Preece and Tye-Murray 1990 *Iowa Audivisual Speech Perception Tests* (Iowa City, IA)

Vinje W E and Gallant J L 1998 Modeling complex cells in an awake macaque during natural image viewing *Advances in Neural Information Processing Systems 10* ed M I Jordan, M J Kearns and S A Solla (Cambridge, MA: MIT Press) pp 236–42

——2000 Sparse coding and decorrelation in primary visual cortex during natural vision *Science* **287** 1273–6