Ultra-fast detection of salient contours through horizontal connections in the primary visual cortex

2011 EPL 93 64001

(http://iopscience.iop.org/0295-5075/93/6/64001)

A Letters Journal Exploring the Frontiers of Physics

March 2011

EPL, **93** (2011) 64001
doi: 10.1209/0295-5075/93/64001

www.epljournal.org

Ultra-fast detection of salient contours through horizontal connections in the primary visual cortex

P. N. Loxley[1], L. M. Bettencourt[1] and G. T. Kenyon[2]

[1] *CNLS and T-5, Theoretical Division, Los Alamos National Laboratory - Los Alamos, NM 87545, USA*
[2] *P-21, Physics, Los Alamos National Laboratory - Los Alamos, NM 87545, USA*

received 6 January 2011; accepted in final form 18 February 2011
published online 17 March 2011

PACS 42.66.Si – Psychophysics of vision, visual perception; binocular vision
PACS 87.18.Sn – Neural networks and synaptic communication
PACS 07.05.Pj – Image processing

Abstract – Salient features instantly attract visual attention to their location and are crucial for object recognition. Experiments in ultra-fast visual perception have shown that object recognition can be surprisingly accurate given only ~ 20 ms of observation. Such short times exclude neural dynamics of top-down feedback and require fast mechanisms of low-level feature detection. We derive a neural model of the primary visual cortex with physiologically parameterized horizontal connections that reinforce salient features, and apply it to detect salient contours on ultra-fast time scales. Model performance qualitatively matches experimental results for human perception of contours, suggesting rapid neural mechanisms involving feedforward horizontal connections can be used to distinguish low-level objects.

The human visual system can rapidly make sense of objects in cluttered environments filled with distracting information. A key step in this process is transforming visual input into a representation where useful information is made explicit [1]. *Salient* features [2–5] instantly attract visual attention to their location. In this sense they are features which "pop out", making whole objects stand out from cluttered visual scenes. In evolutionary terms a rapid response to salient objects would be crucial for enabling biological organisms to respond quickly to predators, prey, or mates. Salient features also form the most distinctive set of visual prototypes for any class of objects, making them essential in object and facial recognition tasks [6,7]. Salient features can be made explicit early in the visual system by transforming image intensity values detected at the retina, into regions of high neural activity in the primary visual cortex (V1). In one class of models used to describe fast object recognition, neural activity is assumed to proceed along the visual pathway in a bottom-up feedforward manner, from lower to higher visual areas, before an object is identified with a particular object class [8]. The justification is that on fast time scales ~ 50 ms top-down feedback connections do not have time to play an active role in recognition. However, horizontal connections within a visual area [9,10] are active over fast time scales. These connections modulate neural activity from stimuli at different locations in the visual field to generate contextual effects such as saliency [11]. Several models using horizontal connections to perform salient contour detection have been proposed [12–14]. Closed contours are known to be highly salient in the human visual system [15], and may also be necessary for ultra-fast detection of animals in natural scenes. It is also known that the human visual system can distinguish between certain types of complex stimuli on ultra-fast (~ 20) time scales [16]. Recently, it was found that a shape cue consisting of human-segmented outlines of natural images allowed animal scenes to be discriminated from non-animal scenes more rapidly than cues involving luminance, color, or texture [17]. In this case, task-relevant information was extracted within 12–17 ms. However, none of the models [12–14] has yet been shown to describe salient contour detection on such ultra-fast (~ 20 ms) time scales. It is likely that similar neural mechanisms are involved in processing the shape cue necessary for rapid animal detection.

In this letter, we show how horizontal connections in V1 can form a representation on ultra-fast time scales that distinguishes closed contours from background clutter. Our model and approach differ from those in [12–14] and [18] in several important respects. The model is based on known neural mechanisms up to the level of V1, while

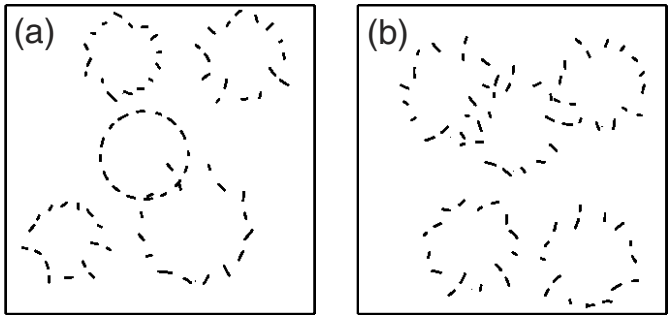

Fig. 1: Images in the contour detection task. (a) An image from ω_1 that contains a single closed contour (center) and 4 distractors (surround). (b) An image from ω_2 that is statistically similar to (a) and contains 5 distractors.

non-neural components were used in refs. [12,13] and [18]. The model proposed here only requires a single excitatory neural population to perform contour detection, while the model due to Li includes both excitatory and inhibitory populations [14]. In addition, our approach takes into account simple-cell receptive field structure, and uses a large dataset of contour and distractor images in order to generate performance statistics on contour detection. Our model does not require repeated iterations for the visual system to eventually relax into a final state as in [18], which is incompatible with ultra-fast contour detection.

The contour detection task we consider requires two image sets ω_1 and ω_2. In ω_1, each image contains a fragmented closed contour similar to that shown in fig. 1(a), which varies in its size and location from image to image. To include the effects of a cluttered background we add n *distractors*, each made from a contour with its fragments (or sequences of fragments) rotated at random so that no closed contour is formed (see fig. 1(a)). Contours and distractors are statistically similar; however, in the human visual system contours will have greater saliency. Images in ω_2 contain only distractors, no closed contours are present. To make ω_2 statistically similar to ω_1 we include $n+1$ distractors, as shown in fig. 1(b). The task is to determine if an image (selected from ω_1 or ω_2) contains a closed contour. An image is presented for 20 ms before visual processing is interrupted by replacing the first (target) image with a second (masking) image in a paradigm known as backward masking [16]. Due to short-term memory being present in the visual system it is likely some combination of the two images is then processed, and task performance typically decreases drastically if a suitable masking image is chosen. In this case the effect of a masking image is to limit the maximum time available for visual processing of the target image to approximately 20 ms. Any processing stages taking longer than 20 ms will be strongly influenced by the masking image (see supplementary information in ref. [8] for further details).

We now propose a V1 neural model to describe visual processing of the target image that takes place over a 20 ms duration. Time delays experienced by a signal traveling along the visual pathway mean this duration will not begin when the target image is first presented, but will begin some time later. The first stage of the model takes an image represented as a set of intensity values $I(\mathbf{r})$ at each point $\mathbf{r}=(x,y)$, and returns $I_\phi(\mathbf{r})$, the regions of $I(\mathbf{r})$ that activate V1 neurons with orientation preference ϕ. This is done by convolving $I(\mathbf{r})$ with the Gabor function $g_\phi(\mathbf{r})$, then thresholding:

$$I_\phi(\mathbf{r}) = H\left(\int g_\phi(\mathbf{r}-\mathbf{r}')I(\mathbf{r}')\mathrm{d}\mathbf{r}'\right), \tag{1}$$

where the Gabor function [19] is

$$g_\phi(\mathbf{r}) = g(R_\phi\mathbf{r}), \tag{2}$$

$$R_\phi = \begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix}, \tag{3}$$

$$g(\mathbf{r}) = \exp\left(-\frac{y^2+0.25x^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}y\right), \tag{4}$$

and depends on the spatial variance σ^2, and spatial frequency λ. The Gabor function includes the important physiological characteristics of simple-cell receptive field structure [19]. The thresholding function H satisfies $H(x)=1$, if $x\geqslant\kappa$, and $H(x)=0$, if $x<\kappa$, and acts to sharpen V1 orientation tuning. The initial activity $I_\phi(\mathbf{r})$ is then modulated by long-range horizontal connections $w_{\phi\phi'}(\mathbf{r}-\mathbf{r}')$, that link neural populations at $\mathbf{r}$ and ϕ, to populations at $\mathbf{r}'$ and ϕ'. The neural-population activity dynamics $u_\phi(\mathbf{r},t)$ is then found by solving [20,21]

$$\left(\tau\frac{\partial}{\partial t}+1\right)u_\phi(\mathbf{r},t) = \sum_{\phi'}\int w_{\phi\phi'}(\mathbf{r}-\mathbf{r}')S(u_{\phi'}(\mathbf{r}',t))\mathrm{d}\mathbf{r}' + I_\phi(\mathbf{r}), \tag{5}$$

where S is the population spiking rate,

$$S(u) = \frac{1}{1+\exp(-(u-\theta)/\sigma')}, \tag{6}$$

which is a sigmoid-shaped function with spiking-rate threshold θ and width σ', and where τ is the time constant giving the duration of smoothed action potential spikes. We assume a value of $\tau\approx 20$ ms which accounts for mean synaptic and dendritic delays of V1 horizontal connections.

We now derive an equilibrium solution of eq. (5) as follows. Setting the time derivative in eq. (5) to zero, and taking the limit $\sigma'\to 0$ so that S reduces to a unit step function H satisfying $H(x)=1$, if $x\geqslant\theta$, and $H(x)=0$, if $x<\theta$, leads to

$$u_\phi(\mathbf{r}) = \sum_{\phi'}\int w_{\phi\phi'}(\mathbf{r}-\mathbf{r}')H(u_{\phi'}(\mathbf{r}'))\mathrm{d}\mathbf{r}' + I_\phi(\mathbf{r}). \tag{7}$$

Since output from eq. (1) is binary-valued, setting θ so that all non-zero regions of input are above threshold

means no subthreshold input is present in (7). Next, we physiologically constrain parameters in eq. (7) so that neural activity cannot reach threshold in regions of zero input: $u_\phi(\mathbf{r},t) < \theta$ where $I_\phi(\mathbf{r}) = 0$, which implies

$$\sum_{\phi'} \int w_{\phi\phi'}(\mathbf{r}-\mathbf{r}') H(u_{\phi'}(\mathbf{r}'))\mathrm{d}\mathbf{r}' < \theta, \tag{8}$$

which is satisfied if $\sum_{\phi'} \int w_{\phi\phi'}(\mathbf{r})\mathrm{d}\mathbf{r} < \theta$. Provided this constraint holds, we now have *input-dominated* activity: $u_\phi(\mathbf{r},t) \geqslant \theta$ only where $I_\phi(\mathbf{r}) \geqslant \theta$. This allows us to replace $u_\phi(\mathbf{r},t)$ with $I_\phi(\mathbf{r})$ in the argument of H with no change to the dynamics, making eq. (5) linear in $u_\phi(\mathbf{r},t)$. The dynamics given by (5) then converges as $\exp(-t/\tau)$ to an equilibrium given by

$$u_\phi(\mathbf{r}) = \sum_{\phi'} \int w_{\phi\phi'}(\mathbf{r}-\mathbf{r}') H(I_{\phi'}(\mathbf{r}'))\mathrm{d}\mathbf{r}' + I_\phi(\mathbf{r}) \tag{9}$$

—meaning that local feedback reduces to feedforward activity *within* the V1 layer. In this case V1 neurons are driven directly by visual input, both locally (within their classical receptive field) and non-locally via horizontal connections from surrounding neurons. More generally, if subthreshold input was present in eq. (5) then local feedback might result in slower convergence to equilibrium. Long-range horizontal connections are predominantly excitatory [9,10], and highly anisotropic —neurons are linked most strongly to other neurons that lie in the direction of their orientation preference [10]. We therefore assume $w_{\phi\phi'}(\mathbf{r}) \geqslant 0$, and decompose connections into orientation-dependent and spatially dependent parts as

$$w_{\phi\phi'}(\mathbf{r}) = m_{\phi\phi'} w(R_{\phi'}\mathbf{r}). \tag{10}$$

For each orientation preference ϕ —given by applying R_ϕ to $\mathbf{r}$— the spatially dependent part is given by $w(\mathbf{r}) \in \{0,1\}$ shown in fig. 2, where 2ψ is the angular dispersion of connections about the orientation preference axis (the x-axis in fig. 2) and r is the characteristic connection range. The orientation-dependent part is

$$m_{\phi\phi'} = \begin{cases} 1, & \text{if } |\phi \pmod{\pi} - \phi'| \leqslant \Delta\phi_{\max}, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $\Delta\phi_{\max}$ is chosen so that only populations with similar orientation preferences are connected [9,10]. This form of connectivity is also consistent with experimental studies on human perception of contours [22].

The V1 model comprising eqs. (1)–(4) and (9)–(11) is now used to assign saliency to the images in fig. 1. In order to allow horizontal connections to have long-range effects we do not break each 256×256 pixel image into smaller "image patches" for processing, but rather we process each image whole. The first step is finding $I_\phi(\mathbf{r})$ from eqs. (1)–(4). To do this we discretize ϕ into $K = 8$ elements between 0 and π as $\phi = k\pi/K$ for $k = 0$ to $K-1$. A non-zero $I_\phi(\mathbf{r})$ indicates a set of contour or distractor fragments

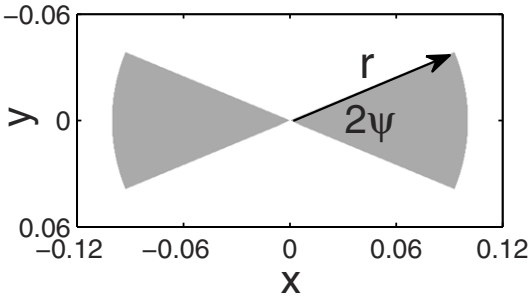

Fig. 2: Plot of function $w(\mathbf{r})$ in eq. (10) showing its dependence on the angle 2ψ, and range r.

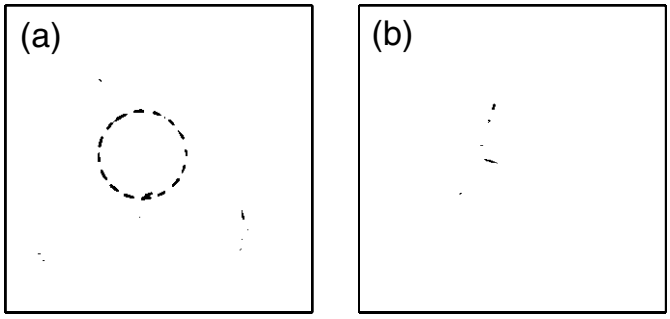

Fig. 3: Neural activity from applying eqs. (1)–(4) and (9)–(11) to fig. 1. (a) $u_\phi(\mathbf{r}) \geqslant 1.5$ using fig. 1(a) for $I(\mathbf{r})$. (b) $u_\phi(\mathbf{r}) \geqslant 1.5$ using fig. 1(b) for $I(\mathbf{r})$. For 1×1 image dimensions, parameters are: $\sigma = 0.004$, $\lambda = 0.02$, $\kappa = 0.016$, $\theta = 0.98$, $r = 0.1$, $\psi = \pi/8$, and $\Delta\phi_{\max} = \pi/6$.

that trigger a neural response corresponding to ϕ. Next, after time τ the equilibrium given by eq. (9) is established, and we can apply eqs. (9)–(11) to yield the neural activity $u_\phi(\mathbf{r}) \geqslant 1.5$ shown in fig. 3. To test the constraint following from eq. (8) is satisfied we can approximate it using fig. 2: $\sum_\phi \int w(\mathbf{r})\mathrm{d}\mathbf{r} \approx 8 \times 1 \times 0.24 \times 0.12 < \theta = 0.98$. Neural activity is modulated by horizontal connections so that, on average, nearest-neighbor contour fragments reinforce each other through two-body interactions while distractor fragments do not. Most of the high neural activity is seen to be due to the the closed contour in fig. 1(a), suggesting it as the most salient feature in fig. 1.

The contour detection task for an ensemble of images is generally more difficult than shown in fig. 1 due to strong intermixing of contour and distractor fragments in many images —there are no constraints on the relative positions of contours and distractors and they often overlap. Although $u_\phi(\mathbf{r})$ contains all the information available from the V1 model, to show this representation is sufficient for contour detection we need to find some functional of $u_\phi(\mathbf{r})$ (called a feature vector [23]) which can be used to statistically discriminate between images in ω_1 and ω_2. The feature vector $X[u_\phi(\mathbf{r})] = \sum_\phi \int_R u_\phi(\mathbf{r})\mathrm{d}\mathbf{r}$ with $R = \{x, y | u_\phi(\mathbf{r}) \geqslant \theta + \Delta\theta\}$, corresponds to the total V1 neural activity at level $\theta + \Delta\theta$ or greater. In the case of fig. 3, X will assign a large value to (a), and a smaller value to (b). More generally, for an ensemble of images

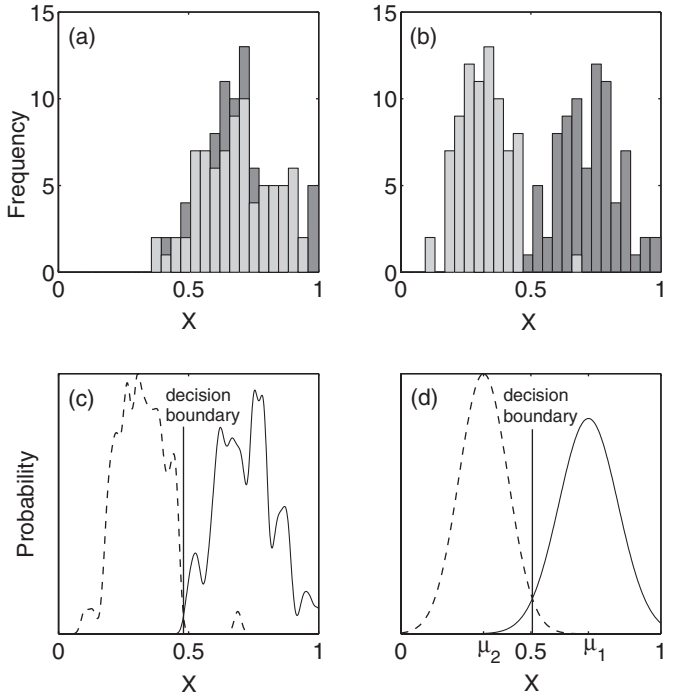

Fig. 4: Distributions for ω_1 and ω_2. (a) Using $X[I(\mathbf{r})]$ on raw images from ω_1 (dark gray) and ω_2 (light gray). (b) Using $X[u_\phi(\mathbf{r})]$ after applying eqs. (1)–(4) and (9)–(11) to each image. (c) Parzen window estimate of distributions in (b) for ω_1 (solid) and ω_2 (dashed). Gaussian estimate of distributions in (b).

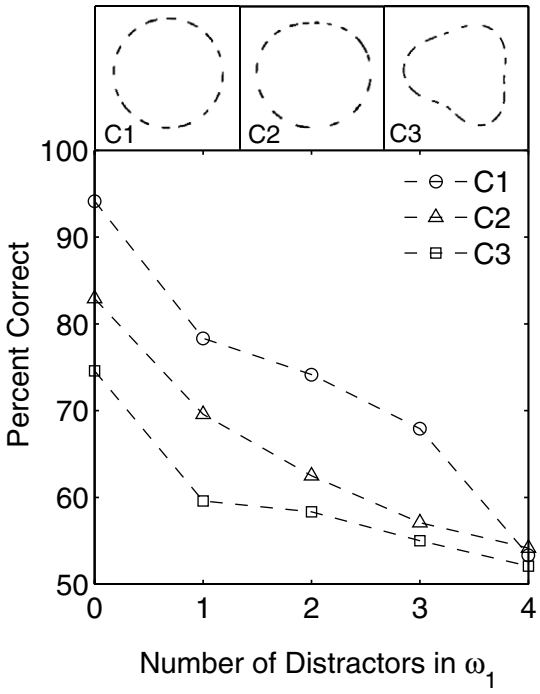

Fig. 5: Performance of model on contour detection task. Top: contour shapes for C1–C3. Bottom: each data point is the percent of correctly classified images for a particular contour shape and number of distractors.

its mean $\mu = \langle X \rangle$ takes two different values μ_1 and μ_2 for ω_1 and ω_2, respectively. A Bayes classifier [23] consisting of two Gaussian probability distributions $p(X, \omega_i)$: one centered at $X = \mu_1$, and another at $X = \mu_2$, leads to a decision boundary where $p(X, \omega_1) = p(X, \omega_2)$, shown as the vertical line in fig. 4(d). The Bayes decision rule states we should decide ω_1 for values of X where $p(X, \omega_1) > p(X, \omega_2)$, otherwise decide ω_2. In fig. 4 we show distributions for ω_1 and ω_2 for image sets containing single closed contours and single distractors. Using a feature vector $X[I(\mathbf{r})] = \int I(\mathbf{r}) \mathrm{d}\mathbf{r}$ outputting the total gray value of each image yields the distributions in fig. 4(a). Due to the statistical similarity of ω_1 and ω_2 these distributions are virtually identical and no decision boundary can be drawn. In fig. 4(b) we apply the V1 model to each image, then use $X[u_\phi(\mathbf{r})]$ to obtain the distributions shown. The ω_1 and ω_2 distributions are now distinct. In figs. 4(c) and (d), the distributions in fig. 4(b) are estimated using Parzen window (non-parametric) and Gaussian (parametric) methods [23]. Both decision boundaries lead to similar classification performance, so we use the Gaussian classifier in the following.

We now consider a contour detection task where contour saliency is varied by changing the contour shape, as well as the number of distractors present. The model results are shown in fig. 5. Data points account for 3 different contour shapes (shown in the top panel, fig. 5), and between 0 and 4 distractors in ω_1. Results converged after 400 images were used for each data point: the classifier was trained on 80 images from each image set, and tested on 120. For each data point the feature vector parameter $\Delta\theta$ was found by maximizing classification performance on the training set, while ψ, r, and $\Delta\phi_{\max}$ were found in the same way using only the first 2 data points for C1. The model is translation invariant and therefore insensitive to contour position. However, it is not completely scale invariant. There is a trade-off between specificity (having connections that are contour-specific) and invariance (having tolerance to contour size variation). In fig. 5, it is seen that model performance varies from very good at 94.2% for circular contours with no distractors, to just above chance at 52.1% for non-circular contours with 4 distractors. Performance generally decreases with increase in contour curvature, and increase in the number of distractors. This is in agreement with [12,15], where one or two sudden changes in local curvature rapidly degraded contour visibility. Results in fig. 5 also qualitatively match experimental results in [12] for human perception of contours (see figs. 5 and 6 in [12]). In both cases performance decreases as shown in fig. 5 when contours become less circular, or as the level of background clutter increases. However, in [12] the task appears to be more difficult than ours and visual processing took 150 ms instead of our hypothesized 20 ms.

We now propose a simple model framework that explains experimental results for human contour perception. The V1 model and connectivity is optimal for

classification performance on circular contours. As contours become less circular, each contour fragment receives less reinforcement from its neighbors and neural activity over a contour decreases. Alternatively, as the number of distractors increases, there is a greater chance that contour and distractor fragments overlap so that both species are reinforced by neural interactions. In both cases contour saliency decreases and it becomes more difficult to distinguish contours from distractors. Model performance improves when contour fragments are reinforced over several iterations using a relaxation-type algorithm as in [18]. However, the existence of such an algorithm in the visual system is hard to justify [1], and the resulting speed of detection would be much slower. The rapid detection performance found here suggests low-level object recognition (in this case contour detection) can take place much faster than has previously been realized. This hypothesis could be tested in human perception experiments by using easier contour detection tasks than those proposed in [11,12,15,22], combined with masking images as in [16] in order to measure the speed of detection.

* * *

We gratefully acknowledge the support of the U.S. Department of Energy through the LANL/LDRD Program project 20090006DR for this work. PNL gratefully acknowledges support from The Center for Nonlinear Studies (CNLS).

REFERENCES

[1] Marr D., *Vision* (W. H. Freeman, New York) 1982.
[2] Koch C. and Ullman S., *Hum. Neurobiol.*, **4** (1985) 219.
[3] Itti L., Koch C. and Niebur E., *IEEE Trans. Pattern Anal. Mach. Intell.*, **20** (1998) 1254.
[4] Li Z., *Proc. Natl. Acad. Sci. U.S.A.*, **96** (1999) 10530.
[5] Li Z., *Trends Cogn. Sci.*, **6** (2002) 9.
[6] Moosmann F., Larlus D. and Jurie F., *ECCV International Workshop on the Representation and Use of Prior Knowledge in Vision*, 2006.
[7] Walker K. N., Cootes T. F. and Taylor C. J., *Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.
[8] Serre T., Oliva S. and Poggio T., *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 6424.
[9] Gilbert C. D. and Wiesel T. N., *J. Neurosci.*, **9** (1989) 2432.
[10] Bosking W. H., Zhang Y., Schofield B. and Fitzpatrick D., *J. Neurosci.*, **17** (1997) 2112.
[11] Kapadia M. K., Ito M., Gilbert C. D. and Westheimer G., *Neuron*, **15** (1995) 843.
[12] Pettet M. W., Mckee S. P. and Grzywacz N. M., *Vision Res.*, **38** (1997) 865.
[13] Yen S. C. and Finkel L. H., *Vision Res.*, **38** (1998) 719.
[14] Li Z., *Neural Comput.*, **10** (1998) 903.
[15] Kovács I. and Julesz B., *Proc. Natl. Acad. Sci. U.S.A.*, **90** (1993) 7495.
[16] Rolls E. and Tovee M. J., *Proc. R. Soc. London, Ser. B*, **257** (1994) 9.
[17] Elder J. H. and Velisavljević L., *J. Vision*, **9** (2009) 1.
[18] Sha'ashua A. and Ullman S., *Proceedings of the International Conference on Computer Vision*, 1988, p. 321.
[19] Jones J. P. and Palmer L. A., *J. Neurophys.*, **58** (1987) 1233.
[20] Coombes S. and Owen M. R., *Phys. Rev. Lett.*, **94** (2005) 148102.
[21] Loxley P. N. and Robinson P. A., *Phys. Rev. Lett.*, **102** (2009) 258701.
[22] Field D. J., Hayes A. and Hess R. F., *Vision Res.*, **33** (1992) 173.
[23] Duda R. O., Hart P. E. and Stork D. G., *Pattern Classification* (John Wiley & Sons) 2001.

Modal analysis of corticothalamic dynamics, electroencephalographic spectra, and evoked potentials

P. A. Robinson,[1,*] P. N. Loxley,[1,†] S. C. O'Connor,[1] and C. J. Rennie[1,2,3]
[1]*School of Physics, University of Sydney, New South Wales 2006, Australia*
[2]*Department of Medical Physics, Westmead Hospital, Westmead, New South Wales 2145, Australia*
[3]*Brain Dynamics Center, Department of Psychological Medicine, Westmead Hospital and University of Sydney, Westmead, New South Wales 2145, Australia*
(Received 3 October 2000; published 29 March 2001)

The effects of cortical boundary conditions and resulting modal aspects of continuum corticothalamic electrodynamics are explored, including feedbacks. Dispersion relations, electroencephalographic spectra, and stimulus response functions are calculated from the underlying physiology, and the effects of discrete mode structure are determined. Conditions under which modal effects are important are obtained, along with estimates of the point at which modal series can be truncated, and the limit in which only a single globally uniform mode need be retained. It is found that for physiologically plausible parameters only the lowest cortical spatial eigenmode together with the set of next-lowest modes can produce distinct modal structure in spectra and response functions, and then only at frequencies where corticothalamic resonances reduce dissipation to the point where the spatial eigenmodes are weakly damped. The continuum limit is found to be a good approximation, except at very low frequencies and, under some circumstances, near the alpha resonance. It is argued that the major electroencephalographic rhythms result from corticothalamic feedback resonances, but that cortical modal effects can contribute to weak substructure in the alpha resonance. This mechanism is compared and contrasted with purely cortical and pacemaker-based alternatives and testable predictions are formulated to enable experimental discrimination between these possibilities.

DOI: 10.1103/PhysRevE.63.041909 PACS number(s): 87.10.+e, 87.19.La, 87.18.−h, 87.19.Nn

I. INTRODUCTION

In recent work, we developed a physiologically based continuum model of corticothalamic electrodynamics that is able to reproduce and unify the main features of observed EEGs, including the discrete spectral peaks, or rhythms, seen in waking and sleeping states [1–6]. In most of these papers we argued that the typical damping rate of cortical waves is sufficiently large that boundary conditions make little difference to their properties. However, it has long been recognized that, if frequency ranges exist in which damping is small, the effects of discrete eigenmode structure in the finite cortex will be important [7,8]. Our recent work on corticothalamic feedback indicates that damping is indeed weakened by such feedback at frequencies close to the spectral rhythms, especially the alpha and beta rhythms near 10 and 20 Hz, respectively, in the waking state, and theta rhythm and sleep spindles near 5 and 15 Hz, respectively, in sleep [9].

In our model, the above rhythms result from resonances in a corticothalamic feedback loop, rather than lying at the frequencies of purely cortical eigenmodes; however, their weak damping opens the possibility that cortical eigenmode effects may also be important near these resonances, even if not elsewhere. Hence, a key aim of this paper is to reconcile these two views of the production of EEG resonances by incorporating modal effects explicitly into our model and determining their influence semiquantitatively. A further motivation for the present work is the reasonably common observation of split-band alpha activity, displaying two discrete alpha frequencies in a single individual, typically separated by 1–2 Hz. It is plausible that these peaks may represent nondegenerate eigenfrequencies that happen to occur in the frequency range in which damping is weakened by corticothalamic feedback effects. By unifying treatments of corticothalamic feedbacks and cortical modal effects, we will determine the feasibility of such a mechanism and contrast it with other possible explanations such as the existence of multiple pacemakers or subcortical loops with different resonant frequencies.

A further motivation for our work arises from the fact that scalp EEGs are spatially large scale because (i) the cortical signal is spatially low-pass filtered by the effects of volume conduction in overlying tissues [7,8,10], (ii) electrodes are relatively widely spaced in practice, leading to coarse spatial resolution, (iii) some rhythms are intrinsically spatially extended, and (iv) the least damped modes are the largest scale ones [1,2]. In previous work we used these features to justify exploring spatially uniform, or global, cortical dynamics as a first approximation to the overall cerebral electrodynamics [2]. Here we extend these ideas to the corticothalamic system and explore how many modes are needed to represent the dynamics well enough to predict spectra and the potentials evoked by discrete stimuli [4].

Recent work has showed that some seizure EEGs have simple structures that imply that the underlying dynamics is low dimensional [11], possibly following a strange attractor or limit cycle in a relatively simple parameter space. There

*Electronic address: robinson@physics.usyd.edu.au
†Present address: Department of Physics, University of Western Australia, Nedlands, Western Australia 6009, Australia.

has also been considerable interest in possible nonlinear aspects of the alpha rhythm [12]. By verifying the dominance of a few low-order modes, the analysis carried out here provides a natural and systematic means of obtaining truncated, low-dimensional systems of equations with the potential of reproducing the nonlinear dynamics of seizures.

In Sec. II we briefly review and generalize the corticothalamic model developed in our previous work, including both intracortical and corticothalamic feedbacks. We then impose boundary conditions and expand the resulting equations in a series of spatial eigenmodes in Sec. III and find equations for the time evolution of the expansion coefficients. Section IV is concerned with modal predictions for spectra and the circumstances under which modal effects are important. A similar discussion of response functions applicable to steady state evoked potentials and evoked response potentials is presented in Sec. V. In Sec. VI we critically discuss a range of mechanisms for the production of spectral resonances, especially split-band alpha rhythms, including new possibilities implied by the results obtained here, and formulate testable predictions to enable experiments to discriminate between them.

II. THEORY

In this section we outline the main relevant results of our neurophysical continuum model of the corticothalamic system, and its predictions of EEG spectra and evoked potentials [5], generalizing them where relevant. Readers should see Ref. [5] for further details and additional references. In this section we consider the case of an infinite cortex in which boundary conditions play no role.

A. Basic model

The mean firing rates (or *pulse densities*) Q_a of excitatory $(a=e)$ and inhibitory $(a=i)$ neurons are approximately related to the cell-body potentials V_a by

$$Q_a(\mathbf{r},t)=\Sigma[V_a(\mathbf{r},t)], \tag{1}$$

where the sigmoidal function Σ increases monotonically from 0 to a maximum $Q_a^{\max}$ as V_a increases from $-\infty$ to ∞. Two such forms of Σ used below are

$$\Sigma_1(V_a)=\frac{Q_a^{\max}}{1+\exp(-\pi z/\sqrt{3})}, \tag{2}$$

$$\Sigma_2(V_a)=Q_a^{\max}\frac{z-1+(z^2+1)^{1/2}}{2z}, \tag{3}$$

$$z=(V_a-\theta_a)/\sigma_a, \tag{4}$$

where θ_a is the mean threshold of neurons, σ_a is the standard deviation of this threshold, $Q_a^{\max}$ is the maximum firing rate, and $\Sigma(\theta_a)=Q_a^{\max}/2$. The coordinate $\mathbf{r}$ in (1) refers to position on the cortex, modeled as a two-dimensional sheet. For later reference, the inverses of (2) and (3) are

$$\Sigma_1^{-1}(y)=\theta_a+\frac{\sigma_a\sqrt{3}}{\pi}\ln\left(\frac{y/Q_a^{\max}}{1-y/Q_a^{\max}}\right), \tag{5}$$

$$\Sigma_2^{-1}(y)=\theta_a+\sigma_a\frac{\dfrac{y}{Q_a^{\max}}-\dfrac{1}{2}}{\dfrac{y}{Q_a^{\max}}\left(1-\dfrac{y}{Q_a^{\max}}\right)}. \tag{6}$$

The potential V_a can be written [5]

$$V_a(\mathbf{r},t)=\int_{-\infty}^{\infty}L(t-t')P_a(\mathbf{r},t')dt', \tag{7}$$

$$L(u)=\alpha^2 u e^{-\alpha u}\Theta(u), \tag{8}$$

where P_a is the mean potential generated by action potentials arriving from other neurons, Θ is the unit step function, and α is a rate constant. Hence,

$$D_\alpha V_a=P_a, \tag{9}$$

$$D_\alpha=\frac{1}{\alpha^2}\frac{d^2}{dt^2}+\frac{2}{\alpha}\frac{d}{dt}+1. \tag{10}$$

The Fourier transform of $L(u)$, is

$$L(\omega)=(1-i\omega/\alpha)^{-2}, \tag{11}$$

which implies that the dendrites act as a low-pass filter with cutoff frequency α.

The potential P_a comprises contributions $\phi_{e,i}$ from other cortical neurons, and subcortical inputs ϕ_s:

$$P_a=N_{ae}s_e\phi_e+N_{ai}s_i\phi_i+N_{as}s_s\phi_s. \tag{12}$$

Here, N_{ab} is the mean number of couplings from neurons of type $b=e,i,s$ to those of type a, and s_b is the size of the response to a unit signal from neurons of type b.

The field ϕ_a of outgoing pulses propagates at $v=5$–10 m s^{-1} and obeys the damped wave equation

$$D_a\phi_a(\mathbf{r},t)=Q_a(\mathbf{r},t), \tag{13}$$

$$D_a=\frac{1}{\gamma_a^2}\left[\frac{\partial^2}{\partial t^2}+2\gamma_a\frac{\partial}{\partial t}+\gamma_a^2-v^2\nabla^2\right], \tag{14}$$

where $\gamma_a=v/r_a$ and r_a is the range of axons a.

Our model incorporated corticothalamic (CT) feedback [13–16], by assuming that ϕ_s is the sum of a non-CT part ϕ_N and a feedback ϕ_T, which originates where part of the excitatory field ϕ_e projects to the thalamus, then returns to the cortex. This adds a propagation time delay t_0 and $n\approx 1$ extra stages of dendritic filtering with rate constant $\eta\approx\alpha$. Our previous work showed that the approximation $n=1$ was adequate, and we assume it henceforth [5]. We also allowed for the possibility of both direct feedback, and feedbacks that emphasize changes in cortical signals by differentiating them in the loop. These features yield

$$D_\alpha \phi_T(\mathbf{r},t) = \frac{G_{ee}}{G_{es}} \int_0^\infty dt_0 \int d^D\mathbf{r}' \Big[\psi(\mathbf{r},\mathbf{r}',t_0) + \psi'(\mathbf{r},\mathbf{r}',t_0) t_0 \frac{\partial}{\partial t} \Big] \phi_e(\mathbf{r}',t-t_0), \quad (15)$$

where the prefactor on the right is separated out to simplify later algebra, and ψ and ψ' measure the strengths of direct and differential feedbacks. Equation (15) generalizes our earlier results by including possible dependences of the feedback on $\mathbf{r}$, $\mathbf{r}'$, and t_0, whereas these were previously treated as being spatially constant and delta function in time [5]. More generally, there are likely to be relatively slow dependences of ψ and ψ' on t itself—in changing between states of arousal, for example—but we ignore these here, simply adopting the appropriate values for the given state.

Local intracortical feedbacks are also possible. Previous analysis showed that a broad class of such feedbacks can be written [3]

$$D_{xy}^j [x(\mathbf{r},t) - x^{(0)}] = \chi_{xy} [y(\mathbf{r},t) - y^{(0)}], \quad (16)$$

$$D_{xy} = \frac{1}{\eta_{xy}} \frac{d}{dt} + 1, \quad (17)$$

where η_{xy} is a time constant, j is a small non-negative integer, χ_{xy} is the linear susceptibility of a quantity x to changes in another quantity y (χ_{xy} could more generally be position-dependent and/or nonlinear), x is a feedback-dependent variable with steady-state value $x^{(0)}$, and y is a variable of steady-state value $y^{(0)}$ that drives the feedback. Typically, $x = s_b$ or θ_a and $y = \phi_e$ or V_e. We found that the resulting wave dispersion relations fell into only four distinct classes, greatly simplifying their analysis [3]. In places below we use feedback of ϕ_e on s_e as an illustrative example, for which the relevant feedback equation can be written [4]

$$D_{s\phi} [s_b(\mathbf{r},t) - s_b^{(0)}] = \chi_{s\phi} [\phi_e(\mathbf{r},t) - \phi_e^{(0)}]. \quad (18)$$

B. Steady states

Upon setting all the spatial and temporal derivatives to zero in (1)–(18), these equations determine the steady states of cortical activation, when the cortex is driven by a constant, spatially uniform non-CT stimulus ϕ_N. By analogy with Ref. [2], one finds

$$\Sigma^{-1}(\phi_e) = [N_{ee} s_e (1+\psi) + N_{ei} s_i] \phi_e + N_{es} s_s \phi_N, \quad (19)$$

in the steady state.

The structure of the solutions of (19) is easily seen in the case in which $\chi_{sV} = 0$, implying $s_b = s_b^{(0)}$. In this case, the left-hand side of (19) is monotonic increasing with a downward curvature for $\phi_e < Q_e^{\max}/2$ and upward curvature for larger ϕ_e, as illustrated in Fig. 1, while the right-hand side of (19) is linear in ϕ_e. Hence, either one or three solutions exist [2]. When three solutions are found, the middle one represents an unstable equilibrium, the lower corresponds to normal activity, and the upper to a high firing rate seizurelike state [2].

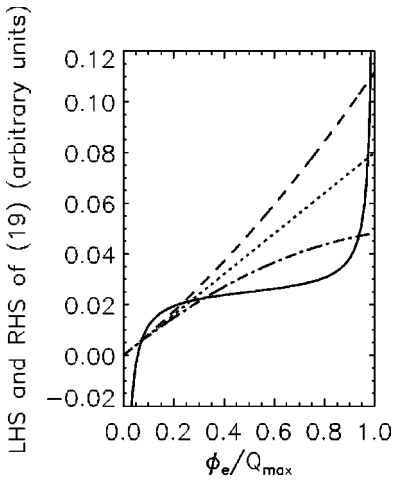

FIG. 1. Determination of steady states. Solid and broken lines show schematic forms of left-hand and right-hand sides of (19), respectively, in cases with three roots, with the dotted line for $\chi_{s\phi} = 0$, the dashed line for $\chi_{s\phi} > 0$, and the dotted–dashed line for $\chi_{s\phi} < 0$. In drawing this figure it is assumed that $N_{ee} s_e (1+\phi) + N_{ei} s_i > 0$ is satisfied, so that the straight line has a positive slope.

If $\chi_{s\phi} \neq 0$, the form of the right-hand side of (19) is modified by the replacement

$$s_b = s_b^{(0)} + \chi_{s\phi} [\phi_e - \phi_e^{(0)}]. \quad (20)$$

Equation (20) yields a quadratic form for the right-hand side of (19), as illustrated in Fig. 1. If $\Sigma = \Sigma_2$, the steady state equation (19) becomes a quartic in ϕ_e. The topology of the loci of the left-hand and right-hand sides of (19) in this case is such that an odd number of solutions must occur between $\phi_e = 0$ and $\phi_e = \phi_e^{\max} = Q_e^{\max}$, a conclusion that also follows from the requirement that stable and unstable steady states alternate, with stable ones at both ends of the sequence [2]. Hence, at least one of the four roots lies outside the physical range, and at least one lies in it. The result that either 1 or 3 roots lie in the physical range can be assumed to apply to any other forms of Σ that incorporate robust features of the physics—further roots might be possible if Σ had a specially chosen form, but would not be robust (although they might correspond to pathological states).

The effect of increasing the external stimulus ϕ_N is to shift the straight line and quadratic curves up in Fig. 1. Hence, as has been discussed previously [2], there should be one low-ϕ_e root at very low (or perhaps negative) ϕ_N, with two more roots appearing at intermediate ϕ_N, and a single high-ϕ_e root at very high ϕ_N.

C. Linear waves

Small perturbations relative to the steady states of the previous subsection obey a linear wave equation. For constant ψ and ψ', with a single value of t_0 and a thalamic

dendritic rate constant equal to α, this yields the transfer functions [3,5]

$$\frac{\phi_e(\mathbf{k},\omega)}{\phi_N(\mathbf{k},\omega)} = G_{es}L(\omega)F_{\theta V}\{(D_e - F_{\theta\phi})[1-G_{ii}L(\omega)]F_{sV} - G_{ee}[1+\Psi(\omega)\tau(\omega)]L(\omega)F_{s\phi}F_{\theta V}\}^{-1} \quad (21)$$

$$= \frac{G_{es}L(\omega)F_{\theta V}}{[1-G_{ii}L(\omega)]F_{sV}} \frac{1}{k^2r_e^2 + q^2(\omega)r_e^2}, \quad (22)$$

$$D_e(\mathbf{k},\omega) = k^2r_e^2 + (1 - i\omega/\gamma_e)^2, \quad (23)$$

$$\Psi = \psi - i\omega t_0\psi', \quad (24)$$

$$\tau(\omega) = \frac{e^{i\omega t_0}}{(1-i\omega/\alpha)^2}, \quad (25)$$

$$F_{s\phi} = 1 + \chi_{s\phi}\phi_e^{(0)}D_{s\phi}(\omega)/s_e^{(0)}, \quad (26)$$

$$F_{sV} = 1 - N_{ee}\chi_{sV}\phi_e^{(0)}D_{sV}(\omega)L(\omega), \quad (27)$$

$$F_{\theta\phi} = -\rho_e\chi_{\theta\phi}D_{\theta\phi}(\omega), \quad (28)$$

$$F_{\theta V} = 1 - \chi_{\theta V}D_{\theta V}(\omega), \quad (29)$$

$$D_{xy}(\omega) = 1 - i\omega/\eta_{xy}. \quad (30)$$

$$q^2(\omega)r_e^2 = (1 - i\omega/\gamma_e)^2 - F_{\theta\phi} - \frac{G_{ee}[1+\Psi(\omega)\tau(\omega)]L(\omega)F_{s\phi}F_{\theta V}}{[1-G_{ii}L(\omega)]F_{sV}} \quad (31)$$

in Fourier space for feedbacks of excitatory quantities on excitatory ones, with analogous equations for inhibitory ones. In (21), (22), and (31) the gains $G_{ab} = \rho_a N_{ab} s_b$ express the response of neurons a to a unit signal from neurons b. The parameter $\rho_a = dQ_a^{(0)}/dV_a$ is evaluated in the steady state where $Q_a^{(0)} = 5 - 10\ \mathrm{s}^{-1} \ll Q_a^{\max}$ is the steady-state firing rate. One has

$$\rho_a = \frac{\pi Q_a^{(0)}}{\sigma_a\sqrt{3}}\left(1 - \frac{Q_a^{(0)}}{Q_a^{\max}}\right), \quad (32)$$

$$\rho_a = \frac{Q_a^{\max}}{2\sigma_a}\frac{(z^2+1)^{1/2}-1}{z^2(z^2+1)^{1/2}}, \quad (33)$$

for Σ_1 and Σ_2, respectively.

The dispersion relation of waves in our model system is given by setting the denominator of (22) to zero, giving

$$k^2 + q^2(\omega) = 0. \quad (34)$$

For the system to be stable, q^2 must not cross the negative real axis [5]. We defined a stability parameter

$$S = 1 - \frac{G_{ee}(1+\psi)}{1-G_{ii}}, \quad (35)$$

that must satisfy $S \geqslant 0$ for the system to avoid instability at $\omega = 0$ [5], although further conditions must be satisfied for stability at all frequencies. Comparison with data has showed that $S \ll 1$ and that the approximation $S = 0$ can be made for most purposes [5].

In general, the dispersion relation (34) must be solved numerically to obtain ω in terms of k. The special case with $\Psi = 0$, $\alpha \gg \omega$, and no intracortical feedbacks can be solved analytically (along with a few other special cases). In this case, one finds that the system supports waves that are purely damped and nonpropagating for $kr_e < 1$, but which approach damped plane waves at large k, with

$$\omega = -i\gamma_e \pm \gamma_e(k^2r_e^2 - 1)^{1/2}, \quad (36)$$

in general [1].

For nonzero feedback Ψ we previously found that waves are least damped at even multiples of π/t_0 if $\psi > 0$ and $\psi'/\psi > 0$ or if both these inequalities are reversed, and at odd multiples if only one is reversed, as appears to be the case in sleep [5]. We will discuss the consequences for spectra and instabilities in the next sections.

D. Spectra

In previous work we used the complexity of cortical inputs to approximate fluctuations in ϕ_N relative to its mean as white noise in space and time. In calculating scalp EEG spectra we also included filtering via volume conduction in intervening tissues [5,7,8,10], as fitted by the spatial filter function

$$F(k) = e^{-k^2/k_0^2}, \quad (37)$$

where $F(k)$ is the square of the ratio of scalp to cortical voltage and $k_0 \approx 30\ \mathrm{m}^{-1}$. The resulting spectrum was

$$P(\omega) = \int |\phi_e(\mathbf{k},\omega)|^2 F(k) d^2\mathbf{k} \quad (38)$$

$$= P_N\left|\frac{L(\omega)}{[1-G_{ii}L(\omega)]F_{sV}}\right|^2 \times \frac{\mathrm{Im}[\exp(q^{*2}/k_0^2)E_1(q^{*2}/k_0^2)]}{|q^2r_e^2|\sin\theta}, \quad (39)$$

$$P_N = \pi|\phi_N G_{es}F_{\theta V}|^2/r_e^2, \quad (40)$$

where $\theta = \mathrm{Arg}(q^2)$, $|\phi_N^2|$ is the white-noise power level in Fourier space, and E_1 is the exponential integral function [17]. This result generalizes an earlier one [5] to include F_{xy}.

The limit $k_0 \to \infty$ corresponds to the absence of volume conduction in which case one has [5]

$$P(\omega) = P_N \left| \frac{L(\omega)}{[1 - G_{ii} L(\omega)] F_{sV}} \right|^2 \frac{\theta}{|q^2 r_e^2| \sin\theta}. \tag{41}$$

Realistic values of k_0 yielded low-pass frequency filtering with a cutoff of around 30 Hz [5]. For simplicity, we ignore the factor $F(k)$ from now on.

The shape of the spectrum depends strongly on the locus of q^2 in the complex plane, with instability occurring if this locus intersects the negative real axis. We have shown that (39) can reproduce both the peaks and the underlying spectrum seen in EEGs for physiologically reasonable values of the input parameters [5].

If $S \approx 0$, $q^2(0) \approx 0$, the behavior of $P(\omega)$ at small ω depends on the leading terms in the expansion of $q^2(\omega)$ in powers of ω, and the effects of volume conduction can be ignored in this frequency range. This gives

$$q^2(\omega) r_e^2 = \sum_{j=0}^{\infty} A_j (-i\omega)^j, \tag{42}$$

where the A_j are real [5].

Momentarily ignoring the spectral peaks and examining the smooth, underlying spectrum for $S = 0$, one finds a small-ω regime in which [5]

$$P(\omega) \approx \frac{P_N}{G_0^2} \frac{\pi}{2\omega |A_1|}, \tag{43}$$

with $G_0 = (1 - G_{ii})(1 - N_{ee} \chi_{sV} \phi_e^{(0)})$. If A_1 is very small, this is modified to

$$P(\omega) = \frac{P_N}{G_0^2} \frac{\pi}{|A_3| \omega^3}, \tag{44}$$

and $A_1 = 0$ defines a stability boundary [5]. At large ω

$$P(\omega) = \frac{P_N \pi \alpha^2 \beta^2 \gamma_e}{2\omega^5}. \tag{45}$$

Assuming $\psi'/\psi > 0$, we found that the frequencies ω_m of spectral peaks are given approximately by [5]

$$\omega_m t_0 \approx x_m + \sin^{-1}(\psi' x_m / |\Psi_m|) \mathrm{sign}(\psi), \tag{46}$$

$$x_m = (m - 1/2)\pi, \tag{47}$$

$$|\Psi_m| = (\psi^2 + \psi'^2 x_m^2)^{1/2}, \tag{48}$$

with $m = 2,4, \ldots$ for $\psi > 0$, $m = 1,3,5, \ldots$ for $\psi < 0$, and $\mathrm{sign}(u) = 0$ for $u = 0$ here (the families of m values are reversed if ψ'/ψ is negative). The positive ψ peaks correspond to waking states, and negative ψ to sleep [5]. Alpha and beta rhythms correspond to $m = 2$ and $m = 4$, respectively, while theta and sleep spindles have $m = 1$ and $m = 3$.

E. Green functions and evoked potentials

An evoked response potential (ERP; also termed an event related potential) is the transient response of the brain to an impulsive stimulus. We have argued that an ERP can be represented by the impulse response to a delta-function input—the Green function of the system [18,19]. Closely related to ERPs are steady state evoked potentials (SSEPs), which are responses to monochromatic sinusoidal inputs.

In Fourier space, the Green function $G(\mathbf{k}, \omega)$ is simply the ratio ϕ_e / ϕ_N given by (21) or (22), its poles define the linear dispersion relation, and its squared modulus yields the spectral power at $\mathbf{k}$ and ω. When analyzing SSEPs, one is interested in the response $G(r, \omega)$ a distance r from an input point. This is given by

$$G(r,\omega) = \frac{G_{es} L(\omega) F_{\theta V}}{[1 - G_{ii} L(\omega)] F_{sV}} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{k^2 r_e^2 + q^2(\omega) r_e^2}, \tag{49}$$

$$= \frac{G_{es} L(\omega) F_{\theta V}}{[1 - G_{ii} L(\omega)] F_{sV}} \int_0^{\infty} dk \frac{k J_0(kr)}{2\pi r_e^2 (k^2 + q^2)}, \tag{50}$$

$$= \frac{G_{es} L(\omega) F_{\theta V}}{[1 - G_{ii} L(\omega)] F_{sV}} \frac{K_0[q(\omega) r]}{2\pi r_e^2}, \tag{51}$$

where K_0 is a Macdonald function (a modified Bessel function of the second kind) [17] and $\mathrm{Re}\, q > 0$ for stable solutions. This result generalizes one obtained previously for purely cortical waves and $\chi_{xy} = 0$ [18]. The dominant behavior of (51) at large r is $\exp[-\mathrm{Re}(qr)]$, implying that wave intensities fall off rapidly with distance unless $\mathrm{Re}\, q$ is small or, equivalently, unless q^2 lies near the negative real axis.

The time dependence of ERPs is of great interest in applications, requiring the calculation of $G(r,t)$. The Fourier transform of (51) to the time domain cannot be evaluated in closed form in general, but is straightforward to calculate numerically and does not lead to modal aspects beyond those involved in $G(r, \omega)$; hence, we do not consider it further here.

III. MODAL FORM OF CORTICOTHALAMIC EQUATIONS

In this section we expand the dynamic equations from Sec. II in series of spatial eigenmodes with time varying coefficients. For definiteness, we consider only Fourier modes of a one-dimensional (1D) or 2D rectangular cortex here, since they incorporate the main physical feature of discreteness due to the imposition of boundary conditions. A spherical cortex can be treated using spherical-harmonic eigenmodes with modest additional effort, while spheroidal eigenmodes are considerably more complex [8], and the actual convoluted geometry of the cortex is amenable only to numerical treatment. Before any attempt to proceed to the full cortical geometry, our aim here is to determine the qualitative effects produced by discrete modal structure and to find the conditions under which they become important. Use of a rectangular system in 2D enables this to be done, while

removing some of the degeneracy implicit in a square one, while 1D systems are used in some illustrations in later sections.

If we consider moderate (but not necessarily linear) perturbations relative to a steady state with $V_a = V_a^{(0)}$, the Fourier transform of (1) can be expanded in either of two equivalent series:

$$Q_a(\mathbf{k},t) = \Sigma^{(0)}\delta(\mathbf{k}) + \Sigma^{(1)}[V_a(\mathbf{k},t) - V_a\delta(\mathbf{k})] + \frac{\Sigma^{(2)}}{2}[V_a^2(\mathbf{k},t) - 2V_a^{(0)}V_a(\mathbf{k},t) + V_a^{(0)2}\delta(\mathbf{k})] + \cdots, \tag{52}$$

$$= \lambda_a^{(0)}\delta(\mathbf{k}) + \sum_{n=1}^{\infty} \lambda_a^{(n)} V_a^n(\mathbf{k},t), \tag{53}$$

where $\Sigma^{(n)}$ is the nth derivative of Σ evaluated at $V_a^{(0)}$ and the $\lambda^{(n)}$, which are *not* derivatives, are easily obtained by comparing (55) and (56). In Fourier space the quantity V_a^2 is expressible as the convolution

$$V_a^2(\mathbf{k},t) = \int \frac{d^D\mathbf{p}}{(2\pi)^D} V_a(\mathbf{p},t) V_a(\mathbf{k}-\mathbf{p},t), \tag{54}$$

in a D-dimensional system, with similar expressions for higher-order terms. Equation (54) embodies three-wave interactions in which waves of wave vector $\mathbf{p}$ and $\mathbf{k}-\mathbf{p}$ interact to produce a response at $\mathbf{k}$. Higher order terms in the series (52) and (53) represent four-wave and more complex interactions.

The dynamical equations (9) and (10) are unchanged in Fourier space, except that the arguments of V_a and P_a are $\mathbf{k}$ and ω. Likewise, Eqs. (13), (16), and (17) are only altered by the use of Fourier arguments, provided χ_{xy} is constant, while (14) becomes

$$D_a = \frac{1}{\gamma_a^2}\left[\frac{d^2}{dt^2} + 2\gamma_a\frac{d}{dt} + \gamma_a^2 + k^2v^2\right]. \tag{55}$$

If feedbacks on the s_b are incorporated, the Fourier form of (12) is the nonlinear equation

$$P_a(\mathbf{k},t) = \int \frac{d^D\mathbf{p}}{(2\pi)^D}[N_{ae}s_e(\mathbf{p},t)\phi_e(\mathbf{k}-\mathbf{p},t) + N_{ai}s_i(\mathbf{p},t)\phi_i(\mathbf{k}-\mathbf{p},t) + N_{as}s_s(\mathbf{p},t) \times\{\phi_N(\mathbf{k}-\mathbf{p},t) + \phi_T(\mathbf{k}-\mathbf{p},t)\}], \tag{56}$$

where $\phi_s = \phi_N + \phi_T$.

Finally, turning to the thalamic feedback equation (15), we note that virtually all neurons entering the cortex are excitatory, implying $G_{ee}/G_{es} \approx N_{ee}/N_{es}$. If we further assume a single value for t_0, as in previous work, and omit this argument from ψ and ψ', Eq. (15) becomes

$$D_\alpha\phi_T(\mathbf{k},t) = \frac{N_{ee}}{N_{es}}\int \frac{d^D\mathbf{p}}{(2\pi)^D}\Psi(\mathbf{k},\mathbf{p})\phi_e(-\mathbf{p},t-t_0), \tag{57}$$

$$\Psi(\mathbf{k},\mathbf{p}) = \psi(\mathbf{k},\mathbf{p}) + \psi'(\mathbf{k},\mathbf{p})t_0\frac{d}{dt}, \tag{58}$$

in Fourier space, where ψ and ψ' have each been expanded in double Fourier series in $\mathbf{r}$ and $\mathbf{r}'$ to obtain (56), with

$$\Psi(\mathbf{k},\mathbf{p}) = \int d^D\mathbf{r}\, d^D\mathbf{r}'\, e^{i\mathbf{k}\cdot\mathbf{r}} e^{i\mathbf{p}\cdot\mathbf{r}'}\Psi(\mathbf{r},\mathbf{r}'). \tag{59}$$

In the special case in which $\Psi(\mathbf{r},\mathbf{r}')$ depends only on $\mathbf{r}-\mathbf{r}'$, Eq. (15) is a convolution and one finds

$$D_\alpha\phi_T(\mathbf{k},t) = \frac{N_{ee}}{N_{es}}\Psi(\mathbf{k})\phi_e(\mathbf{k},t-t_0), \tag{60}$$

$$\Psi(\mathbf{k}) = \psi(\mathbf{k}) + \psi'(\mathbf{k})t_0\frac{d}{dt}. \tag{61}$$

When boundary conditions are imposed, the values of $\mathbf{k}$ and $\mathbf{p}$ are restricted in our modal equations (9), (10), (13), (14), (16), (17), and (52)–(61), with

$$\mathbf{p} = \left(\frac{2\pi m}{L_x}, \frac{2\pi n}{L_y}\right), \tag{62}$$

for a 2D rectangular cortex of size $L_x \times L_y$, with an analogous equation for $\mathbf{k}$ and an obvious simplification for a 1D system. In consequence, the integrals over $\mathbf{p}$ are replaced by sums over the allowed values, with

$$\int \frac{d^2\mathbf{p}}{(2\pi)^2} \to \frac{1}{L_xL_y}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}, \tag{63}$$

in 2D. The correspondence (63) yields the power spectrum per unit area in the 2D discrete case, which corresponds directly to what is calculated for continuous $\mathbf{k}$.

IV. MODAL EFFECTS ON SPECTRA

There has been significant recent progress in calculating EEG spectra from the underlying physiology. One issue that remains contentious is whether the discrete spectral peaks are due in part or whole to discrete cortical resonances whose frequencies are set by spatial boundary conditions. Our work has stressed the contrasting role of corticothalamic resonances in producing discrete peaks, with resonances induced via time delays, not spatial boundary conditions. However, at frequencies where damping is small, it is possible that nondegenerate spatial eigenmodes might give rise to peak splitting, possibly including the production of split-band alpha rhythms seen in a significant percentage of subjects.

In this section we explore the effects of discrete eigenmode structure on spectra, retaining corticothalamic feedback Ψ, but setting the intracortical feedback susceptibilities $\chi_{xy} = 0$. Moreover, we neglect filtering by the skull. We de-

termine the number of modes that must be retained in a modal analysis to reproduce various features of the spectra, the circumstances under which a continuum approximation may be made, and the role of the cortical size in determining the spectrum. In Sec. IV A we illustrate many of the essential points using a 1D cortex, whose discrete spectrum can be evaluated in closed form, before analyzing the 2D case in Sec. IV B.

A. 1D cortex

In a 1D cortex of linear size L_x, with periodic boundary conditions, the power spectrum is given by

$$P(\omega)=A(\omega)\frac{1}{L_x}\sum_{m=-\infty}^{\infty}\frac{1}{\left|\left(\frac{2\pi m}{L_x}\right)^2+q^2\right|^2}, \tag{64}$$

$$=A(\omega)\frac{\mathrm{Im}[q\coth(q^*L_x/2)]}{2|q^2|\mathrm{Im}(q^2)}, \tag{65}$$

$$A(\omega)=\frac{|\phi_N|^2G_{es}^2}{r_e^4}\left|\frac{L(\omega)}{1-G_{ii}L(\omega)}\right|^2 \tag{66}$$

[20]. The corresponding result in the continuum case is

$$P(\omega)=A(\omega)\int\frac{dk}{2\pi}\frac{1}{|k^2+q^2|^2} \tag{67}$$

$$=\frac{A(\omega)}{4|q^2|\mathrm{Re}\,q}. \tag{68}$$

We require $\mathrm{Re}\,q>0$ for stability. If $\mathrm{Re}(qL_x/2)\gtrsim 1$, one finds $\coth(qL_x/2)\approx\coth(q^*L_x/2)\approx 1$ and the result (65) approaches (68), implying that the system can be treated as a continuum. This corresponds to waves with $\mathrm{Im}\,k$, which measures the linewidth of the mode in k, exceeding the separation $2\pi/L_x$ between modes. At sufficiently high frequencies, the condition for the continuum limit is fulfilled if $L\gtrsim 2\pi r_e$, which is marginally satisfied in the human cortex, according to the parameters in Table I. (Incidentally, to avoid the physiologically wasteful phenomenon of corticocortical fibers that wrap more than half way around the cortex, rather than taking a shorter route, $L\gg r_e$ must be satisfied, which implies that the continuum limit will also be at least marginally valid in other species.) If $\mathrm{Re}(qL_x/2)\ll 1$, the waves are weakly damped and one finds that the spectrum (68) is dominated by a series of resonances where $\mathrm{Im}(qL_x/2)=m$. In this case, one can approximate (68) by

$$P(\omega)\approx\frac{A(\omega)L_x^4}{(4\pi m)^2|qL_x-2m\pi i|^2}, \tag{69}$$

for $\mathrm{Im}\,q$ close to $m\pi/L_x$. The dispersion relation (34) then implies that such waves satisfy $k\approx\pm 2\pi m/L_x$; i.e., they are weakly damped standing waves of the system. Such a resonance mechanism has been discussed extensively by Nunez [7,8]; the difference here is that cortical modal resonances can only become apparent at frequencies where the corticothalamic loop already has a resonance that leads to weak wave damping. This means that modal resonances may lead to substructure in the CT resonances (e.g., the split-band alpha structure discussed in Sec. VI), but not to the major resonances at the alpha, beta, and other rhythms.

TABLE I. Physiologically realistic values of some corticothalamic quantities, in the ranges used in Ref. [5]. Also included for illustrative purposes are numerically determined values of ψ and ψ' appropriate to a marginally stable waking state with a strong alpha peak; these have not been estimated physiologically.

Quantity	Value	Unit
$Q_{e,i}^{\max}$	200	s^{-1}
$\theta_{e,i}$	15	mV
$\sigma_{e,i}$	5	mV
α	100	s^{-1}
N_{ae}	4000	
N_{ai}	600	
N_{as}	60	
$s_{e,s}$	1	μV s
$-s_i$	5	μV s
v	10	m s^{-1}
r_e	0.1	m
r_i	0.1	mm
γ_e	100	s^{-1}
γ_i	10^5	s^{-1}
G_{ee}	1	
G_{ii}	-1	
G_{es}	0.5	
t_0	70	ms
ψ	1.0	
ψ'	0.8	
L_x, L_y	0.5	m

Figure 2(a) shows a series of spectra calculated for the parameters in Table I, except that L_x is varied. Rapid convergence to the continuum limit is seen for $L_x\gtrsim 2r_e=0.2$ m, in accord with the above discussion. The only significant difference is that at small L_x, there is a large enhancement in the low-frequency part of the spectrum, reflecting the strong role of the uniform ($k=0$) mode in this case because other modal resonances occur at large negative q^2 for small L_x, with exactly resonant values satisfying

$$q^2r_e^2=-k_m^2r_e^2=-(2\pi r_e/L_x)^2m^2, \tag{70}$$

for the mth resonance. The low frequency enhancement has $P(f)\sim f^{-2}$, as opposed to f^{-1} for the continuum limit in this marginally stable case [see (43)]. For stable systems, both spectra level off as $f\to 0$, but the modal one remains enhanced.

For modal resonance to produce discrete peaks in a 1D cortex, the locus of q^2 must pass near more than one of the poles given by (70). If many poles are comparably close to the q^2 locus at a particular frequency, as is the case at large

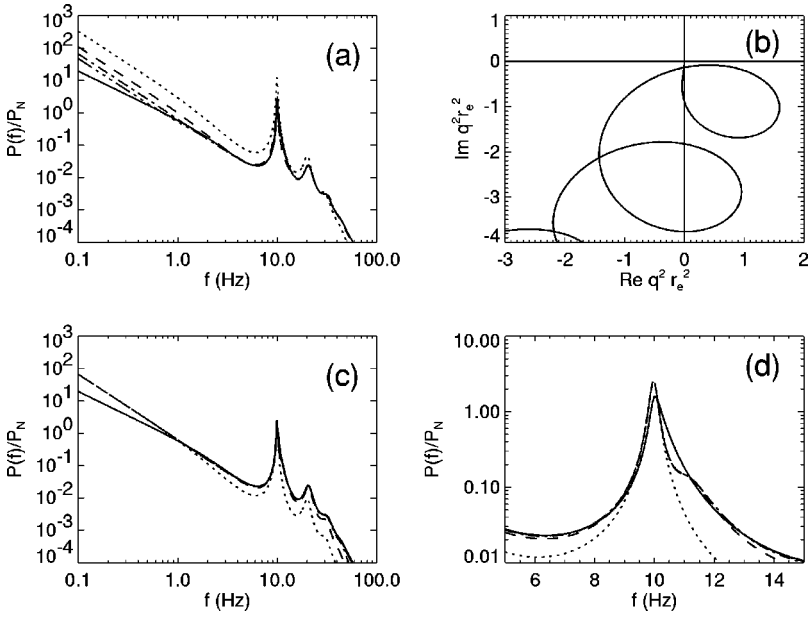

FIG. 2. Spectra for the parameters in Table I, but various values of L_x and $m_{\max}$ in a 1D cortex. In each case the solid line shows the continuum limit, while the dotted, dashed, dotted-dashed, and triple-dotted–dashed lines correspond to increasing values of the parameter being varied. (a) $L_x=0.1$, 0.3, 0.5, 0.7 m. (b) Locus of $q^2(\omega)r_e^2$ (same for all parameter sets in other frames and figures), with $q^2(0)=0$. (c) $m_{\max}=0$, 1, 2, 3. (d) Expanded view of the alpha peak in (c).

ω, for example, the continuum limit will provide a good approximation and discrete modal structure will not be seen. Hence, visible modal structure corresponds to cases where ≈ 2 poles are involved. The most prominent case is predicted to correspond to the $m=0$ and $m=\pm 1$ poles, since these lie relatively close together and to the low frequency part of the q^2 locus. Higher poles are less relevant to discrete modal effects since they also lie on the negative real q^2 axis and the overall trend is for the imaginary part of q^2 to increase in magnitude with frequency, as illustrated in Fig. 2(b).

The number of modes contributing significantly to the spectrum at a given frequency can be estimated from (67), with the fractional contribution from wave numbers above k decreasing as k^{-3} for $k \gtrsim |q|$. The number of strongly active modes is thus at most a few times $|q|L_x/\pi$ if $\mathrm{Re}\, q^2 > 0$. If $\mathrm{Re}\, q^2 < 0$, we write (67) as

$$P(\omega)=A(\omega)\int \frac{dk}{2\pi}\,\frac{1}{(k^2+\mathrm{Re}\, q^2)^2+(\mathrm{Im}\, q^2)^2}, \qquad (71)$$

which implies that the number of active modes is of order $2L_x\,\mathrm{Re}\, q\,\mathrm{Im}\, q/\pi|q| \leqslant |q|L_x/\pi$, which yields the same estimate as for positive $\mathrm{Re}\, q^2$. At high frequencies, $|q| \approx \omega/v$, implying of order $2fL_x/v$ major modes, or $\sim 0.1f$ modes for the human parameters in Table I. Hence, only a modest number of modes contribute strongly to observed spectra for typical EEG frequencies of $\lesssim 50$ Hz. In the frequency range of interest, Fig. 2(b) implies that the modes of relevance extend from $m=0$ to a maximal value $\pm m_{\max}$. Figure 2(c) shows spectra calculated using various $m_{\max}$. Rapid convergence is seen, with just three modes ($m_{\max}=1$) giving a good approximation up to 30 Hz, in accord with the above estimate. The single global mode gives a reasonable representation of the spectrum up to the vicinity of the alpha frequency, although the alpha resonance is sharper and stronger in this approximation than in the full calculation.

Figure 2(d) shows an expanded view of the vicinity of the alpha peak in Fig. 2(c), revealing a shoulder at around 11 Hz, which appears at $m_{\max}=1$ and hardly changes for larger m. This is due to the contribution from the $m=0$ and $m=\pm 1$ resonances, whose poles lie at $q^2r_e^2=(0,0)$ and $(-1.6,0)$, respectively, in Fig. 2(b) and are successively approached by the q^2 locus as f increases past the nominal alpha frequency. Our earlier work on variations in the form of the q^2 locus with physiological changes [5] implies that it is very difficult, if not impossible, to obtain a locus that passes close enough to the $m=0$ and $m=\pm 1$ poles to produce strong peak structure in 1D without encountering an instability. In any event, if such a situation were attainable, it would be realized only in a very narrow parameter range, whereas split alpha peaks are seen in several percent of subjects.

B. 2D cortex

It is possible to reduce the 2D discrete summation corresponding to (38) to a single sum, giving

$$P(\omega)=\frac{A(\omega)}{L_y}\sum_{n=-\infty}^{\infty}\frac{\mathrm{Im}[q_n\coth(q_n^*L_x/2)]}{2|q_n^2|\mathrm{Im}(q_n^2)}, \qquad (72)$$

$$q_n^2=q^2+(2\pi n/L_y)^2, \qquad (73)$$

but it does not appear to be possible to evaluate (72) in closed form. Even so, the insights obtained in the 1D case above remain valid. In particular, the continuum limit is valid for $\mathrm{Im}\, k \gtrsim 2\pi \max\{L_x^{-1},L_y^{-1}\}$ or, equivalently at high f, $\min\{L_x,L_y\} \gtrsim 2\pi r_e$. Thus, the smallest overall dimension of the cortex governs the applicability of the 2D continuum limit, with the 1D continuum limit being approached as this dimension shrinks to zero, with approximately $2n_{\max}=2m_{\max}L_y/L_x$ terms needing to be retained if $2m_{\max}$ terms contribute significantly to the sum over m (without loss of generality, we may assume $L_y \leqslant L_x$). This is borne out by the

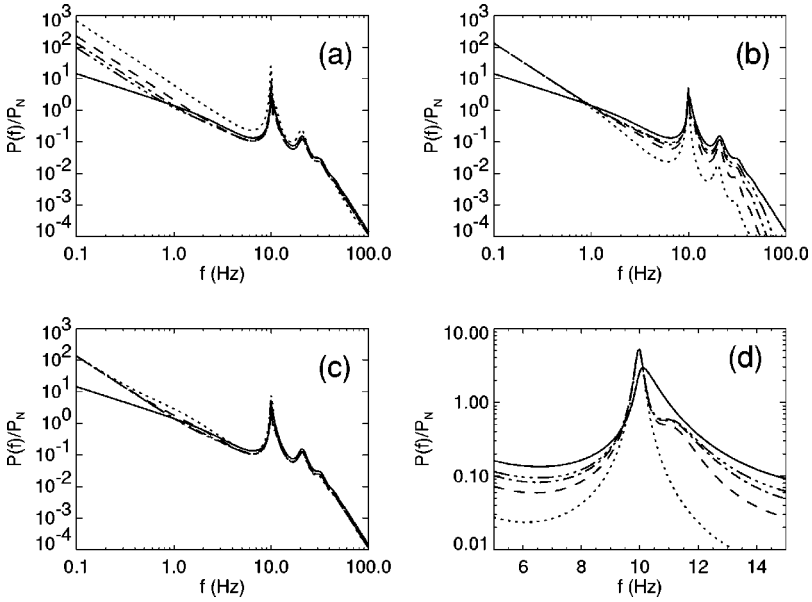

FIG. 3. Spectra for the parameters in Table I, but various values of L_x, L_y, and $m_{\max}$ in a 2D cortex. In each case the solid line shows the continuum limit, while the dotted, dashed, dotted-dashed, and triple-dotted–dashed lines correspond to increasing values of the parameter being varied. (a) $L_x=0.1, 0.3, 0.5, 0.7$ m. (b) $m_{\max}=0, 1, 2, 3$. (c) $L_y=0.2, 0.3, 0.4, 0.5$ m with $L_xL_y=0.25$ m^2 fixed. (d) Expanded view of the alpha peak in (b).

results in Fig. 3(a) where rapid convergence is seen to the continuum limit for frequencies above about 1 Hz in the case where $L_x=L_y\gtrsim 0.2$ m. Likewise, Fig. 3(c) shows that if the cortical area is held constant, cases with $L_x\approx L_y$ have spectra closer to the 2D continuum limit than those with very small L_y, which approach the 1D version for $L_y\lesssim 0.2$ m. The 1D spectrum never differs from the 2D one by more than a factor of 2 in this example.

The number of strongly active modes remains restricted in 2D, with the analog of (71) implying that at most a few times $|q^2|L_xL_y/2\pi$ modes contribute strongly. This is confirmed by the results in Fig. 3(b), which show rapid convergence as $m_{\max}$ increases beyond about 2, where modes in a circular region of radius $m_{\max}$ are included, with

$$(m^2+n^2)^{1/2}\leqslant m_{\max}. \tag{74}$$

The number of modes required is greater than in 1D because of the larger high-k weighting in 2D Fourier space.

If they are weakly damped, resonant modes have

$$q^2r_e^2=-k_{mn}^2r_e^2=-(2\pi r_e)^2\left(\frac{m^2}{L_x^2}+\frac{n^2}{L_y^2}\right), \tag{75}$$

for integers m and n. This form allows multiple modes with similar k_{mn}, leading to the possibility of multiple resonances at nearby frequencies. However, for these resonances to give rise to distinct peaks in the spectrum, they must not be too close together, implying that m and n must be of order unity. Evidence of such modal structure is seen in the alpha peak in Fig. 3, where a secondary peak lies at about 11 Hz, on the flank of the main peak at 10 Hz. The expanded view in Fig. 3(d) shows this most clearly, demonstrating that this is due to the $(m,n)=(0,0),(0,\pm 1),(\pm 1,0)$ modes, with little change as $m_{\max}$ increases beyond 1. The reason that the peak is stronger in this case is simply the larger number of modes with the same $|\mathbf{k}|$ but, as in 1D, a discrete peak is difficult or impossible to obtain, except perhaps in a very restricted parameter regime. It is even more difficult to produce subpeaks within the beta CT resonance because of the tendency of the q^2 locus to move away from the negative real axis (where the resonances lie) at higher frequencies. The separation between the alpha subpeaks is also restricted to 1–2 Hz at most, since they must correspond to a relatively small arc of one of the loops in Fig. 2(b), while a circuit of such a loop occurs in only about 10 Hz (the alpha frequency). We return to these issues in Sec. VI.

V. MODAL EFFECTS ON GREEN FUNCTIONS AND EVOKED POTENTIALS

Impulse responses of the brain, in the form of evoked response potentials, are commonly used to probe cognitive processes. Connections between prestimulus EEGs and subsequent ERPs are also widely explored, and will be treated in detail by us elsewhere. In this section, we concentrate on modal effects on the Green function $G(r,\omega)$ of the corticothalamic system, which is closely related to the steady state evoked potential (SSEP) produced by a sinusoidally varying input at a frequency ω and distance r from the detecting electrode. As in Sec. IV we begin with a 1D cortex, for which many of the modal expressions can be evaluated in closed form, then turn to the 2D case.

A. 1D cortex

If one wishes to evaluate the steady state evoked potential a distance x from a sinusoidally modulated point stimulus, one must evaluate $G(x,\omega)$. For a real sinusoidal perturbation, a linear combination of the real quantities $[G(x,\omega)+G(x,-\omega)]/2$ and $[G(x,\omega)-G(x,-\omega)]/2i$ is actually what is relevant.

The discrete 1D analog of (49) for $\chi_{xy}=0$ is [20]

$$G(x,\omega)=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\frac{1}{L_x}\sum_{m=-\infty}^{\infty}\frac{e^{i2\pi mx/L_x}}{\left(\frac{2\pi mr_e}{L_x}\right)^2+q^2r_e^2} \tag{76}$$

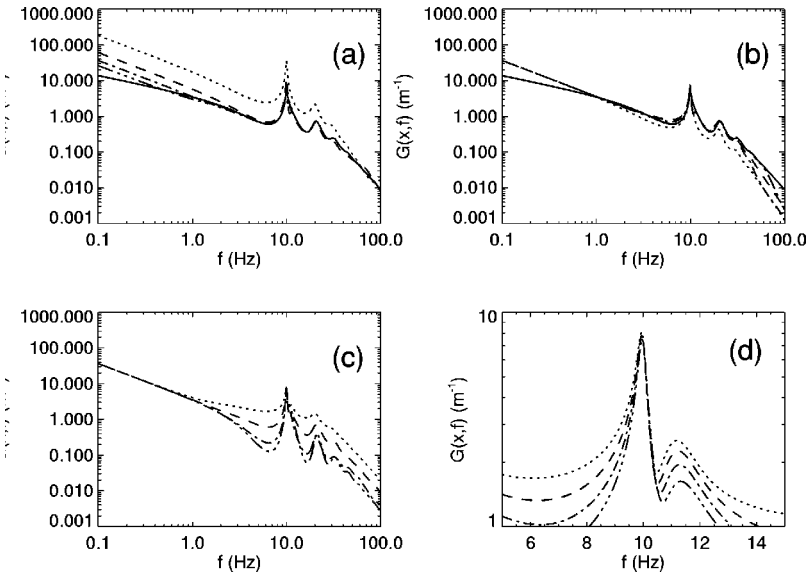

FIG. 4. Green functions $|G(x,f)|$ vs f for $x=0.08$ m and the parameters in Table I, but various values of L_x, $m_{\max}$, and x in a 1D cortex. In frames (a) and (b) the solid line shows the continuum limit, while the dotted, dashed, dotted–dashed, and triple-dotted–dashed lines correspond to increasing values of the parameter being varied. (a) $L_x=0.1, 0.3, 0.5, 0.7$ m, (b) $m_{\max}=0, 1, 2, 3$, (c) $x=0, 0.08, 0.16, 0.24$ m. (d) Expanded view of the alpha peak in (c) for $x=0, 0.02, 0.04, 0.06$ m.

$$=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\frac{\cosh[q(L_x-2|x|)/2]}{2qr_e^2\sinh(qL_x/2)}, \tag{77}$$

which holds for $|x|\leqslant L_x$ and can also be written

$$G(x,\omega)=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\frac{1}{2qr_e^2}\times\{\tanh(qL_x/2)\cosh(qx)-\sinh(q|x|)\} \tag{78}$$

$$=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\frac{1}{2qr_e^2}\{e^{-q|x|}+\cosh(qx)\times[\tanh(qL_x/2)-1]\}. \tag{79}$$

The 1D continuum result is

$$G(x,\omega)=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\int\frac{dk}{2\pi r_e}\frac{e^{ikx}}{k^2r_e^2+q^2(\omega)r_e^2} \tag{80}$$

$$=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\frac{e^{-q|x|}}{2qr_e^2}. \tag{81}$$

The discrete expression (79) rapidly approaches the continuum limit (81) as $\mathrm{Re}(qL_x)$ increases beyond about 2. For smaller values of $\mathrm{Re}\,q$ there are resonances at the same locations in the q^2 plane as for the spectra discussed in Sec. IV.

Figure 4 shows variations in $|G(x,f)|$ as a function of frequency f as L_x, $m_{\max}$, and x are varied in a 1D cortex whose parameters are otherwise those of Table I. In Fig. 4(a) it is seen that the continuum limit is rapidly approached as L_x increases, with significant differences only below 1 Hz for $L_x\gtrsim 0.5$ m where the discrete and continuous versions scale as f^{-1} and $f^{-1/2}$, respectively. Rapid, but nonmonotonic, convergence to the $m_{\max}=\infty$ limit is seen in Fig. 4(b), with $m_{\max}=1$ sufficing to obtain excellent agreement up to circa 30 Hz. In Fig. 4(c) we see that the high-frequency components of $G(x,f)$ weaken and the $m_{\max}=0$ form from Fig. 4(b) is approached as r increases, reflecting the rapid damping of these waves. In contrast, the low-f part is almost independent of r, reflecting the dominance of the $m=0$ global mode in this case. Apart from the substructure near the alpha resonance, the continuum limit is found to be a much better approximation at small r, which reflects the requisite inclusion of a large number of modes up to at least $k\sim 1/r$ in order to resolve the spatial scales involved. Figure 4(d) shows that the modal Green function also exhibits two subpeaks in the alpha rhythm. These correspond to the $m=0$ (≈ 10 Hz) and $m=\pm 1$ (≈ 11 Hz) resonances and appear stronger than in the spectra, because only a single r value is involved: when sources at various r contribute to a spectrum, the large-r contributions, which involve primarily the $m=0$ mode, dominate the more local ones, which contribute to the secondary peak.

B. 2D cortex

The continuum limit yields the explicit 2D result (51), and the double sum involved in the 2D analog of (76) can be reduced to the 1D sum

$$G(x,\omega)=\frac{G_{es}L(\omega)}{1-G_{ii}L(\omega)}\frac{1}{2r_e^2L_y}\sum_{n=-\infty}^{\infty}\frac{1}{q_n}\exp(i2\pi ny/L_y)\times\{\tanh(q_nL_x/2)\cosh(q_nx)-\sinh(q_n|x|)\}, \tag{82}$$

but (82) does not seem to be able to be evaluated in closed form. We thus proceed numerically in this section, apart from noting that $2n_{\max}\approx 2m_{\max}L_y/L_x$ terms are required in (82) if $2m_{\max}$ terms are significant in the sum over m.

Figure 5 shows variations in $|G(x,f)|$ vs f as $L_x=L_y$, $m_{\max}$, and $\mathbf{r}=(x,0)$ are varied in a 2D cortex whose parameters are otherwise those of Table I. The conclusions are very similar to those for the 1D case seen in Fig. 4. One difference is that the second alpha subpeak is slightly stronger in 2D, as seen in Fig. 5(d), because of the larger number of modes contributing to it.

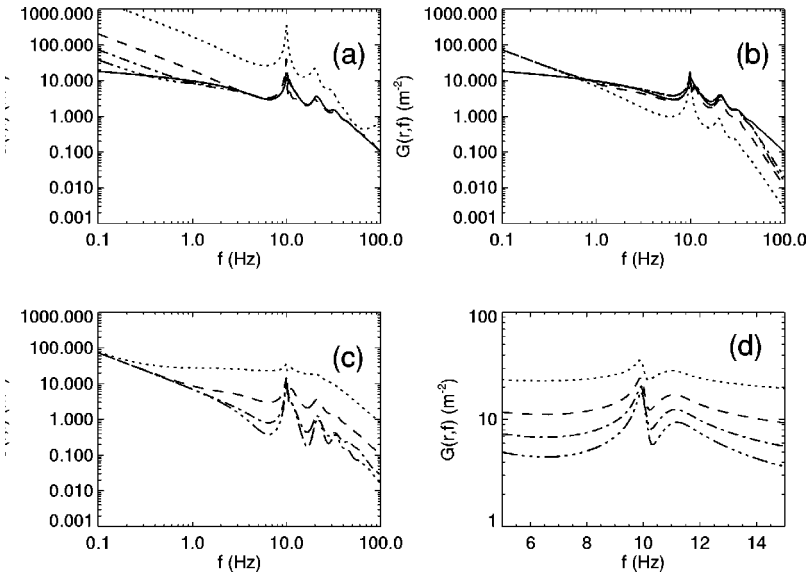

FIG. 5. Green functions $|G(\mathbf{r},f)|$ vs f for $\mathbf{r}=(0.08,0)$ m and the parameters in Table I, but various values of L_x, $m_{\max}$, and r in a 2D cortex. In frames (a) and (b) the solid line shows the continuum limit, while the dotted, dashed, dotted-dashed, and triple-dotted–dashed lines correspond to increasing values of the parameter being varied. (a) $L_x=L_y=0.1, 0.3, 0.5, 0.7$ m, (b) $m_{\max}=0, 1, 2, 3$, (c) $r=0, 0.08, 0.16, 0.24$ m. (d) Expanded view of the alpha peak in (c) for $r=0, 0.02, 0.04, 0.06$ m.

VI. RESONANCE MECHANISMS

A key aim of EEG theories has been to explain the occurrence of the major rhythms seen in waking and sleeping states, particularly the alpha ($\approx$10 Hz) and beta ($\approx$20 Hz) waking rhythms, and the theta ($\approx$5 Hz), spindle ($\approx$14 Hz), and delta ($<$3.5 Hz) rhythms most prominent in sleep. The recent prediction and tentative detection of additional rhythms at nominal frequencies of around 25 Hz in sleep and 30 Hz waking further extends this task [5]. A comprehensive theory should also account for the splitting of the alpha rhythm into two subpeaks, as seen in a few percent of normal individuals, and the occurence of Rolandic mu rhythm, an alphalike rhythm most prominent centrally on the head and which also possesses a betalike component [9]. Several mechanisms have been proposed, including ones based on modal resonances. We briefly recapitulate these here, including new ones based on the present work, and outline some experimental tests that can be used to distinguish between them.

One mechanism that could account for any number of peaks is that of pacemakers located in the thalamus or elsewhere in the brain, each of which comprises neurons with a characteristic frequency. It is argued that these neurons mutually entrain one another via nonlinear couplings, leading to a linewidth less than that of the frequency response curve of any single neuron [9]. Although neurons with resonant frequencies certainly exist, the pacemaker theory has a number of problems. First, it does not explain why the resonant frequencies are approximately in a harmonic relationship in the waking state, nor why sleep resonances occur almost exactly midway between the waking ones. The transition from waking to sleep would involve deactivation of the pacemakers for resonances at 2, 4, and 6 times a base frequency of around 5 Hz, and activation of those at 1, 3, and 5 times, and this pattern is not explained by the theory. Nor does pacemaker theory account for the strong enhancement of delta waves in sleep or the splitting of the alpha peak seen in some subjects (except perhaps via the ad hoc addition of another pacemaker at the requisite frequency). Moreover, from the point of view of parsimony, postulation of a different pacemaker for each peak is unconvincing.

A second widely discussed mechanism is that resonances may be the result of spatial cortical eigenmodes, with substructure caused by the breaking of modal degeneracy by the complicated cortical geometry [7,8]. In a series of papers, we have shown that purely cortical waves appear to be too heavily damped to have sharp resonances and that such resonances are certainly not possible for the cortical part of the model discussed here for physiologically realistic parameters [1,5]. Nunez has developed a different model in which there are additional degrees of freedom in choosing the relative coefficients in a wave equation analogous to (14) [7,8]. In his model, it *is* possible in principle to obtain weakly damped standing waves at any frequency, with the actual resonant frequencies selected by the cortical geometry. One prediction of this model is that the frequencies of the global resonances should not depend on location in the cortex, although their amplitudes may vary due to differences in overlying tissues, etc. Under this mechanism, splitting in peaks is predicted to be due to breaking of initial degeneracy as a result of geometric asymmetries in the real cortex; the globally uniform mode is nondegenerate and cannot be split by this mechanism. Relative amplitudes of split-band peaks should be similar at all locations as long as the splitting is not so large that the two peaks undergo significantly different spatial filtering, and provided that the eigenmodes have similar spatial structure. If the latter proviso is not fulfilled, the higher frequency mode of an initially degenerate pair will tend to have higher amplitudes in frontal regions because the posterior lobes of the brain are larger, implying smaller k (and hence ω) is favored in this region. This argument is exactly analogous to standard textbook ones used to infer the qualitative spatial structure and frequencies of pairs of initially degenerate quantum mechanical eigenmodes in the presence of a perturbing potential that breaks that degeneracy.

We have previously argued that the major EEG peaks are due to resonances in a corticothalamic feedback loop, demonstrating that this mechanism can account for the entire

series of resonances previously seen, their inversion in sleep along with simultaneous enhancement of delta, and the occurrence of additional resonances near 25 Hz in sleep and 30 Hz waking [5]. In the present paper, we have shown that this mechanism can also account for weak substructure in the alpha peak arising from the effect of purely cortical eigenmodes in frequency ranges where damping has been reduced by the CT resonance, particularly if the measuring electrode is close to a localized source at the frequency involved. This mechanism predicts that there should be little or no structure in the beta resonance, because its part of the q^2 locus is relatively remote from the resonance poles. It also implies that there should be an f^{-2} enhancement in the delta range if the cortex is very near marginal stability. As with the previous mechanism, this one predicts that neighboring peaks should exhibit no spatial variation of resonance frequencies and little variation in relative amplitudes, although absolute amplitudes may change. Again, one would expect the higher frequency mode of a split pair to be predominantly frontal (this does not apply to resonances due to different modes such as the $m=0$ and $m=1$ resonances, since these are nondegenerate to begin with).

A second alternative involving CT loops is that cortical eigenmodes may be largely irrelevant to submode structure, but that there may be two different values of t_0 in split-alpha subjects. In this crude form, this suggestion suffers from similar objections to the pacemaker idea (on the grounds of being ad hoc), although each t_0 value gives rise to an entire family of peaks, not just one. However, a bimodal distribution of t_0 values can arise very naturally and robustly, since loops from the thalamus to various parts of the cortex and back are not all of the same length, and there will thus be a spread in the value of $t_0=t_0(\mathbf{r})$. In particular, there must be at least one maximum and one minimum of t_0 corresponding to the longest and shortest CT loops, most likely at the front and back of the head or vice versa. If the spatial second derivatives of t_0 are increasing in magnitude at these extremums, it is straightforward to show that there will be an enhanced probability of observing these values of t_0, relative to nearby values, with a reduced probability in the opposite case. In the first case, if the maximum and minimum t_0 values differ sufficiently they may give rise to distinguishable peaks in the spectrum. A key observable consequence of this mechanism would be that all CT resonances would be expected to be affected in the same way, including the beta rhythm in particular. This contrasts with the cortical submode mechanism which would affect the alpha peak preferentially and would lead to different (and probably unobservably weak) submode structure within other CT resonances. One would also expect systematic shifts in the resonant frequencies and relative amplitudes of alpha subpeaks across the scalp, although each value of t_0 would be observable at an attenuated level over much of the scalp because of the spread of corticocortical fibers arising from each point. Under this mechanism the power spectrum is due to a superposition of the effects of various values of t_0, and at low to moderate frequencies all peaks are dominated by the response of the global mode in the presence of the overall feedback via the thalamus—peaks do not generally correspond to different spatial eigenmodes. Rolandic mu rhythm fits naturally into this picture as the alpha rhythm corresponding to corticothalamic loops in the central part of the cortex, particularly involving the sensorimotor cortex which is known to be involved in mu [9]. This mechanism predicts that the spatial attenuation of a given spectral peak will result largely from its intracortical damping and, hence, that sharper peaks (which are necessarily weakly damped) should have more uniform spatial distribution than broader ones.

Another possibility is that pacemaker cells may operate in conjunction with cortical or CT resonances, effectively sharpening the resonance beyond what could be achieved by either mechanism alone. For example, in the CT context, the q^2 curve could be distorted toward a nearby pole even by weakly resonant pacemaker effects, producing a significant enhancement in the spectrum without a large absolute change in q^2. Such compound mechanisms are quite possible, particularly given that weakly resonant neurons have existed in the thalamus, for example, for millions of years. If any evolutionary advantage accrues to sharp resonances, adaptation of these cells to enhance them is likely to have occured, but investigation of such possibilities is beyond the scope of the present work.

VII. SUMMARY AND DISCUSSION

We have examined the effect of boundary conditions and resulting discrete spatial eigenmodes on the predictions of our corticothalamic model of EEG generation, generalizing it in the process. This work yields equations for modal dynamics, spectra, and the response functions corresponding to steady state evoked potentials and evoked response potentials. These equations incorporate both modal and corticothalamic resonances, enabling a unified treatment of these potential resonance mechanisms, contributing to determining their relative effects on observations, and laying the groundwork for future nonlinear EEG studies.

Our results for modal effects on white noise-driven spectra showed that, for human parameters, the continuum limit (no modal effects) is an excellent approximation at most frequencies for systems of linear cortical sizes exceeding roughly 0.2 m, which is well fulfilled for humans ($L_x \approx 0.5$ m). The exceptions are at very low frequencies ($f \lesssim 1$ Hz) and near the alpha resonance. At low f in a marginally stable system, the uniform mode produces a spectral enhancement, with $P(f) \sim f^{-2}$, rather than the continuum behavior f^{-1}. Near the alpha frequency, some substructure may be seen if the resonance is strong, but distinct subpeaks were not found for parameters near the physiologically realistic ones. The substructure seen is due almost entirely to the effects of the global mode and the modes nearest to it in Fourier space.

It was further found that the modal spectrum was reproduced semiquantitatively up to the vicinity of the alpha frequency by the contribution from the global mode alone, although the alpha peak was somewhat narrower and more pronounced in this approximation. Inclusion of further modes led to rapid convergence toward the limiting result.

Green functions $G(r,\omega)$ were calculated in Sec. V, including modal and corticothalamic effects. There it was ar-

gued that these functions represent steady state potentials evoked by sinusoidal stimuli (SSEPs), while their Fourier transforms $G(r,t)$ represent responses evoked by impulsive stimuli (ERPs). Our results again showed that modal effects were most prominent at very low frequencies and near the alpha peak, especially close to a localized source. In other respects, the continuum approximation is best at short ranges, where many modes must be included to resolve the spatial scales involved, while the global-mode approximation is best at large scales.

Several possible mechanisms for the production of major EEG resonances, substructure within them, and related rhythms such as Rolandic mu, were critically discussed in Sec. VI, and experimentally testable predictions were listed for each. It was argued that pacemaker mechanisms have trouble in accounting convincingly for the relative frequencies of major rhythms, and sleep–wake variations. Purely cortical resonances have difficulty avoiding strong damping for physiologically realistic parameters, and we argue that they predict higher frequencies for frontally concentrated members of modal pairs whose degeneracy has been broken by geometric irregularities. Corticothalamic resonances can weaken damping sufficiently for cortical eigenmode structure to become significant near major rhythms, primarily the alpha peak; again, higher-f members of initially degenerate pairs should be frontally concentrated. The most promising mechanism overall relies on corticothalamic resonances to produce the major rhythms, and a distribution of CT delays t_0 to produce substructure in these resonances. This mechanism predicts correlated substructure in both the alpha and beta peaks, accounts for relative frequencies of peaks, and sleep–wake differences. It also predicts a inverse relationship between peak sharpness and spatial attenuation.

Overall, we thus conclude that the effects of spatial cortical eigenmodes are relatively weak for physiologically realistic parameters, except perhaps in very narrow parameter regimes. In any event, they only appear to be significant in frequency ranges in which corticothalamic resonances have weakened the wave dissipation to the point that spatial eigenmodes are weakly damped. The most robust signatures of cortical modes are expected to be the f^{-2} enhancement at very low frequencies in marginally stable systems, and possible substructure in the alpha peak with specific spatial properties that contrast with those expected from other mechanisms. From the perspective of numerical modeling, we conclude that convolutions and other cortical irregularities and detailed boundary conditions are not very important in determining the form of the spectrum under most circumstances.

ACKNOWLEDGMENT

The authors thank J.J. Wright for his constructive comments on the paper.

[1] P.A. Robinson, C.J. Rennie, and J.J. Wright, Phys. Rev. E **56**, 826 (1997).
[2] P.A. Robinson, C.J. Rennie, J.J. Wright, and P.D. Bourke, Phys. Rev. E **58**, 3557 (1998).
[3] C.J. Rennie, P.A. Robinson, and J.J. Wright, Phys. Rev. E **59**, 3320 (1999).
[4] C.J. Rennie, J.J. Wright, and P.A. Robinson, J. Theor. Biol. **205**, 17 (2000).
[5] P.A. Robinson, C.J. Rennie, J.J. Wright, H. Bahramali, E. Gordon, and D.L. Rowe, Phys. Rev. E (to be published).
[6] J.J. Wright and D.T.L. Liley, Behav. Brain Sci. **19**, 285 (1997).
[7] P.L. Nunez, *Electric Fields of the Brain* (Oxford University Press, Oxford, 1981).
[8] P.L. Nunez, in *Neocortical Dynamics and Human EEG Rhythms*, edited by P.L. Nunez (Oxford University Press, Oxford, 1995), Chaps. 1 and 9.
[9] E. Niedermeyer, in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 4th ed., edited by E. Niedermeyer and F.H. Lopes da Silva (Williams and Wilkins, Baltimore, 1999), Chaps. 9, 13, and 27.
[10] R. Srinivasan, P.L. Nunez, and R.B. Silberstein, IEEE Trans. Biomed. Eng. **45**, 814 (1998).
[11] M. Feucht, U. Moller, H. Witte, K. Schmidt, M. Arnold, F. Benninger, K. Steinberger, and M.H. Friedrich, Cereb. Cortex **8**, 524 (1998).
[12] C.J. Stam, J.P.M. Pijn, P. Suffczynski, and F.H. Lopes da Silva, Clin. Neurophysiol. **110**, 1801 (1999).
[13] F.H. Lopes da Silva, J.E. Vos, J. Mooibroek, and A. van Rotterdam, Electroencephalogr. Clin. Neurophysiol. **50**, 449 (1980).
[14] D.L. Robinson, Int. J. Neurosci. **22**, 81 (1983).
[15] M. Steriade, P. Gloor, R.R. Llinás, F.H. Lopes da Silva, and M.-M. Mesulam, Electroencephalogr. Clin. Neurophysiol. **76**, 481 (1990).
[16] G. Buzsáki, Neuroscience (Oxford) **41**, 351 (1991).
[17] *Handbook of Mathematical Functions* edited by M. Abramowitz and I.A. Stegun (Dover, New York, 1970).
[18] P.A. Robinson, J.J. Wright, and C.J. Rennie, Phys. Rev. E **57**, 4578 (1998).
[19] J.J. Wright, A.A. Sergejew, and H.G. Stampfer, Brain Topogr. **2**, 293 (1990).
[20] A.P. Prudnikov, Y.A. Brychkov, and O.I. Marichev, *Integrals and Series* (Gordon and Breach, New York, 1986), Vols. 1 and 2.

Soliton Model of Competitive Neural Dynamics during Binocular Rivalry

P. N. Loxley[1,2] and P. A. Robinson[1,2,3]

[1]School of Physics, The University of Sydney, Sydney, New South Wales 2006, Australia
[2]Brain Dynamics Center, Westmead Hospital, Westmead, New South Wales 2145, Australia
[3]Faculty of Medicine, The University of Sydney, Sydney, New South Wales 2006, Australia

(Received 23 September 2008; revised manuscript received 6 May 2009; published 23 June 2009)

Binocular rivalry is investigated in a continuum model of the primary visual cortex that includes neural excitation and inhibition, stimulus orientation preference, and spike-rate adaptation. Visual stimuli consisting of bars or edges result in localized states of neural activity described by solitons. Stability analysis shows binocular fusion gives way to binocular rivalry when the orientation difference between left-eye and right-eye stimuli destabilizes one or more solitons. The model yields conditions for binocular rivalry, and two types of competitive dynamics are found: either one soliton oscillates between two stimulus regions or two solitons fixed in position at the stimulus regions oscillate out of phase with each other.

DOI: 10.1103/PhysRevLett.102.258701 PACS numbers: 42.66.Si, 05.45.Yv, 87.10.Ed, 87.19.lj

Binocular rivalry occurs when we view certain dissimilar monocular stimuli, such as a series of horizontal bars with one eye and a series of vertical bars with the other. This leads to alternations in visual perception, where one eye's input dominates, then the other's [1,2]. A similar phenomenon occurs when viewing ambiguous figures such as the Necker cube [2], where two perceptual interpretations alternate over time. Binocular rivalry is linked to neural activity in the primary visual cortex (V1) [3,4], where neurons respond most strongly to visual stimuli such as a bar or edge of particular orientation [5]. This orientation preference (OP) changes smoothly across V1, except at isolated locations [5]. Visual stimuli consisting of bars or edges of specific orientation therefore activate specific groups of neurons in V1. Key elements in most neural models of visual rivalry include reciprocal inhibition and adaptation [6–9]: reciprocal inhibition takes place over fast time scales and involves neurons coupled so that when one set is active, another set is suppressed, while adaptation causes the active neurons to fatigue over slow time scales, allowing the suppressed neurons to become active. While current neural models successfully reproduce some key properties of binocular rivalry, such as the distribution of perceptual dominance durations [6,8,9], and the propagation of dominance changes across V1 [8], there are no clear predictions of what conditions trigger the breakdown of binocular fusion (where inputs to left and right eyes form a stable fused percept) and the onset of rivalry. It is also unclear what types of dynamics take place during rivalry, with some findings suggesting neural activity consists of traveling waves [4,8], and others that alternations in activity remain fixed in position [7,9].

In this Letter, we present a physiologically based continuum model of binocular rivalry that allows detailed investigation of the fusion-rivalry transition as stimulus parameters are varied and exhibits two types of competitive neural dynamics consistent with rivalry. For simplicity we consider a one-dimensional (1D) slice through V1, and calculate the activity $u(x, t)$ of neural populations coupled through a synaptic connectivity function $w(R)$ which depends on the distance R between neurons, and subject to a visual stimulus $h(x)$ that drives neural activity. The dynamical equations are

$$\mathcal{D}_f u(x,t) = \int_{-\infty}^{\infty} w(x-x')H(u(x',t)-\theta(x',t))dx' + h(x), \tag{1}$$

$$\mathcal{D}_s \theta(x,t) = \theta_0 + \kappa u(x,t) H(u(x,t)), \tag{2}$$

where $\mathcal{D}_a \equiv \tau_a \partial/\partial t + 1$, τ_a is a time constant, and H is the unit step function: satisfying $H(x) = 0$ if $x < 0$, and $H(x) = 1$ if $x \geq 0$. Equation (1) describes the fast process where incoming spikes are smoothed into pulses of duration τ_f at the synapses of a neuron [7,8,10]. The spiking rate of a neural population is assumed to follow a threshold response $H(u - \theta)$. Equation (2) describes slow processes which cause the spiking rate to adapt over time [7,8,11,12], and changes the spiking threshold $\theta(x, t)$ over a time scale $\tau_s \gg \tau_f$; $H(u)$ is used in (2) so the coupling is linear in u for $u > 0$, and zero for $u < 0$.

Visual stimuli consisting of bars or edges activate groups of neurons in V1 with OP near the stimulus orientation—some sensitive to left-eye stimuli, others to right-eye stimuli. Since OP varies across V1, stimuli of different orientations are mapped to different parts of V1 even when they are from the same point in the visual field. A typical 1D slice through V1 (see Fig. 1) that includes the full OP range 0°–180°, plus left-eye and right-eye stimulus preferences, contains the essential elements for a fusion-rivalry transition, and is consistent with the known structure of V1 [5]. Monocular stimuli are included in (1) using $h(x)/h_{\max} = e^{-(x+d/2)^2/\delta} + e^{-(x-d/2)^2/\delta}$, where d is the distance between stimulus regions in V1 and depends on the orienta-

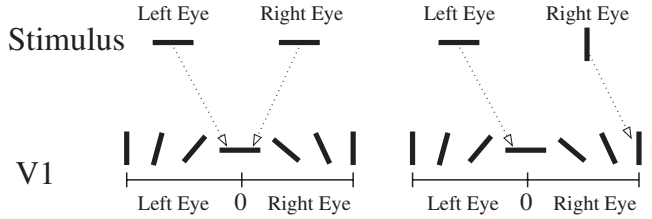

FIG. 1. Mapping of visual stimulus to V1 slice. A bar viewed by the left or right eye drives activity in V1 at a position where the bar orientation matches neural OP (bars) and eye preference. Examples are shown where both eyes view horizontal bars (left) and where the left eye views a horizontal bar and the right eye views a vertical bar (right).

tion difference between left-eye and right-eye stimuli as in Fig. 1, δ is the OP width, and $h_{\max}$ is determined by stimulus strength. The only conditions we impose on $w(R)$ for rivalry are $w(-x) = w(x)$, and $w(R) < 0$ for large R. These are satisfied by a Mexican-hat form commonly used in cortical models: $w(R) = ce^{-R^2/r_1^2} - (r_1/r_2)e^{-R^2/r_2^2}$ with $r_1 < r_2$ and $c > 0$—although excitatory connections in $w(R)$ are unnecessary for rivalry as neural activity can be driven above threshold by the stimulus term in (1).

Extending the analysis in [10] to nonuniform $h(x)$ and $\theta(x, t)$, we set time derivatives in (1) and (2) to zero, and consider an equilibrium comprising N_s spatially localized regions each with $u > \theta$ over width m centered at a_i:

$$u_s(x) = \sum_{i=1}^{N_s} \int_{-m/2}^{m/2} w(x - x' - a_i)dx' + h(x), \qquad (3)$$

$$\theta_s(x) = \theta_0 + \kappa u(x)H(u(x)). \qquad (4)$$

We term these localized states *solitons* [13]: in [14] states similar to those given by (3) were shown to have several particlelike properties, and despite the nonlinearity, dissipation, and external driving present, perturbed equilibrium states always remained localized in numerical simulations of (1) and (2). Consistency requires $u(x) = \theta(x)$ at the boundaries of each localized region, and is implemented using (3) and (4) with $u_s(a_i \pm m/2) > 0$, yielding $2N_s$ simultaneous equations $u_s(a_i \pm m/2) = \theta_0/(1 - \kappa)$. For $N_s = 1$ and $a_1 = 0$, we get $u_s(\pm m/2) = \theta_0/(1 - \kappa)$. Upon using (3) and the symmetry $h(-m/2) = h(m/2)$, this becomes

$$\int_0^m w(x)dx + h\left(-\frac{m}{2}\right) = \frac{\theta_0}{1 - \kappa}, \qquad (5)$$

which is used to determine m. An equilibrium from (3)–(5) is shown in Fig. 2(a).

To determine equilibrium stability we apply an Evans function technique [11], in which linear deviations given by $u(x, t) = u_s(x) + \delta u(x)e^{\lambda t}$ and $\theta(x, t) = \theta_s(x) + \delta\theta(x)e^{\lambda t}$ are substituted into (1) and (2), and an eigenvalue equation is generated by expanding to first order in $\delta u(x)$ and $\delta\theta(x)$. Defining $(x_1, x_2) = (-m, m)/2$ for $N_s = 1$, this

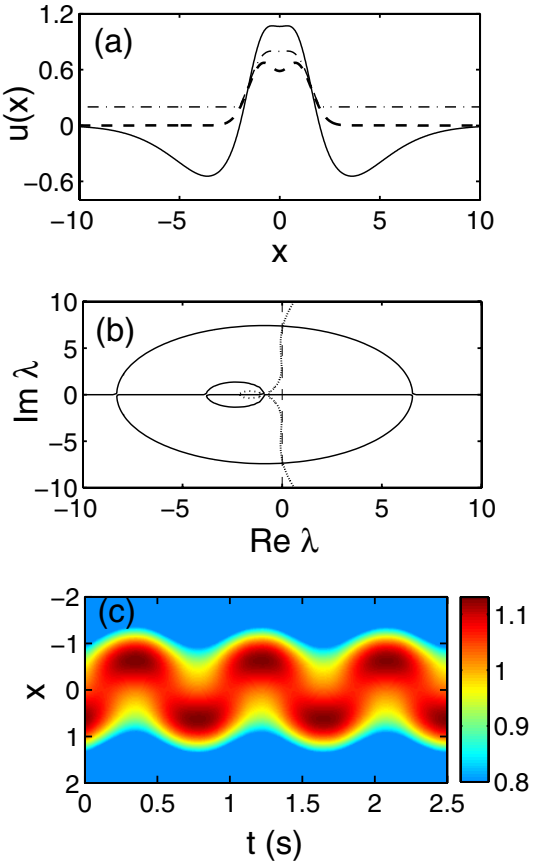

FIG. 2 (color online). Fusion-rivalry transition for $N_s = 1$. (a) Equilibrium u_s (solid line) and θ_s (dash-dotted line) from (3)–(5) for a stimulus (dashed line), and parameters $h_{\max} = 0.64$, $\delta = 1$, $d = 1.77$, $c = 0.7$, $r_1 = 2$, $r_2 = 4.5$, $\theta_0 = 0.2$, and $\kappa = 0.56$. (b) Zeros of $\mathcal{E}(\lambda)$ from (6) with $\tau_f = 20$ ms and $\tau_s = 500$ ms occur at intersections of the solid line and the dotted line. (c) Numerical solution for $u(x, t) > 0.8$ from (1) and (2).

eigenvalue equation can be written as $\mathcal{E}(\lambda) = 0$, with

$$\mathcal{E}(\lambda) = \det\left(\frac{\mathcal{D}_f(\lambda)\mathcal{D}_s(\lambda)}{\mathcal{D}_s(\lambda) - \kappa}\mathcal{I} - \mathcal{A}\right), \qquad (6)$$

where $\mathcal{A}$ is a $2N_s \times 2N_s$ matrix with elements $\mathcal{A}_{ij} = w(x_i - x_j)/|u_s'(x_j) - \theta_s'(x_j)|$, $\mathcal{I}$ is an identity matrix, and $\mathcal{D}_a(\lambda) = \tau_a\lambda + 1$. The derivatives are found from (3) and (4), giving $|u_s'(x_i) - \theta_s'(x_i)| = (1 - \kappa)|w(x_i + m/2) - w(x_i - m/2) + h'(x_i)|$ for $N_s = 1$. Contours of $\mathrm{Re}\mathcal{E}(\lambda) = 0$ and $\mathrm{Im}\mathcal{E}(\lambda) = 0$ are shown in Fig. 2(b) for the $N_s = 1$ equilibrium, and $\mathcal{E}(\lambda) = 0$ where they intersect, corresponding to an eigenvalue of the linear stability problem. For $d < d_c$ ($\simeq 1.77$ in Fig. 2), where d depends on the orientation difference between left-eye and right-eye stimuli as in Fig. 1, we find $\mathrm{Re}\lambda < 0$ so the one-soliton equilibrium is stable. We interpret this as binocular fusion: two monocular stimuli form a stable nonoscillating percept [6]. As d increases, the orientation difference between the two monocular stimuli reaches a critical level at $d = d_c$, where a pair of complex-conjugate eigenvalues cross $\mathrm{Re}\lambda = 0$:

the one-soliton equilibrium is now unstable to a Hopf bifurcation. The numerical solution for $u(x, t)$ is shown in Fig. 2(c), where high neural activity localizes to one stimulus region for approximately 0.5 s before moving to the other stimulus region for approximately 0.5 s, and so on. We interpret this as binocular rivalry: two monocular stimuli lead to oscillations between states where one stimulus is dominant and the other is suppressed [6].

The origin of the Hopf instability for $N_s = 1$ is the following: the coupling term in (2) causes the spiking threshold to increase over time τ_s in the soliton region, resulting in a translational instability as the soliton moves to avoid a region of high threshold. However, a stimulus breaks translational symmetry and tends to pin the soliton. For $d < d_c$, one soliton can span both stimulus regions and so remains centered midway between them. At $d = d_c$, this is no longer the case just off equilibrium, and the soliton must choose one region or the other. As the spiking threshold increases in one region, the soliton eventually moves to the other, as in Fig. 2(c). This is consistent with rivalry as a single soliton that oscillates between two stimulus regions.

To find the $N_s = 2$ equilibrium we assume $(a_1, a_2) = (-a, a)/2$ for $a > 0$, leading to $u_s(\pm a/2 \pm m/2) = \theta_0/(1 - \kappa)$. The stimulus symmetry obeys $h(-a/2 - m/2) = h(a/2 + m/2)$ and $h(-a/2 + m/2) = h(a/2 - m/2)$, reducing the four equations to two. One is

$$\int_0^m [w(x) + w(x + a)]dx + h\left(-\frac{m}{2} - \frac{a}{2}\right) = \frac{\theta_0}{1 - \kappa}, \quad (7)$$

which determines m when a is fixed. Subtracting the remaining equation from (7), yields

$$\int_0^m [w(x + a) - w(x - a)]dx = h\left(\frac{m}{2} - \frac{a}{2}\right) - h\left(-\frac{m}{2} - \frac{a}{2}\right), \quad (8)$$

which determines a when m is fixed. Both (7) and (8) are solved self-consistently. Solitons usually repel due to long-range inhibition, and $a \to \infty$. However, inclusion of a stimulus allows (8) to be solved for finite a, and a resulting equilibrium is shown in Fig. 3(a). Defining $(x_1, x_2, x_3, x_4) = (-m - a, m - a, -m + a, m + a)/2$ generates the previous eigenvalue equation, with $\mathcal{E}(\lambda)$ from (6) and $|u_s'(x_i) - \theta_s'(x_i)| = (1 - \kappa)|w(x_i + m/2 + a/2) + w(x_i + m/2 - a/2) - w(x_i - m/2 + a/2) - w(x_i - m/2 - a/2) + h'(x_i)|$. The two-soliton equilibrium is stable for $d > d_c$ ($\simeq$6.1 in Fig. 3), and we interpret this as binocular fusion. As d decreases, at $d = d_c$ a pair of complex-conjugate eigenvalues cross $\mathrm{Re}\lambda = 0$ in Fig. 3(b) and the two-soliton equilibrium becomes unstable to a Hopf bifurcation. The numerical solution for $u(x, t)$ is shown in Fig. 3(c), where neural activity grows in one stimulus region for approximately 1.5 s, before it decays there and grows in the other stimulus region for approximately 1.5 s, and so on. We interpret this as binocular rivalry.

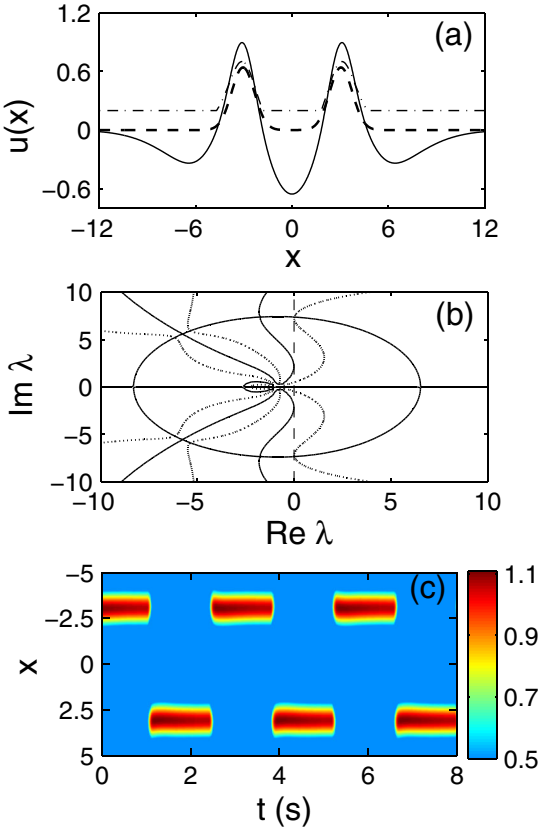

FIG. 3 (color online). Fusion-rivalry transition for $N_s = 2$. (a) Equilibrium u_s (solid line) and θ_s (dash-dotted line) for a stimulus (dashed line) with $d = 6.1$ and other values as in Fig. 2. (b) Zeros of $\mathcal{E}(\lambda)$. (c) Numerical solution for $u(x, t) > 0.5$ from (1) and (2).

The Hopf instability for $N_s = 2$ has a different origin to that for $N_s = 1$. Both stimulus regions are now occupied, so each soliton remains fixed in position. For $d > d_c$, there is only weak reciprocal inhibition between two solitons and both remain above threshold. At $d = d_c$, each soliton inhibits the other more strongly, and just off equilibrium only one can remain above threshold. However, this dominant soliton falls below threshold after τ_s, releasing the suppressed soliton from inhibition. The latter then grows above threshold and becomes dominant, as in Fig. 3(c). This is consistent with rivalry as two solitons, one fixed at each stimulus region, that oscillate out of phase with one another.

Complete dominance of either monocular stimulus is less likely as both stimuli extend over a larger part of V1; instead, rivalry breaks up into a patchwork of zones of alternating dominance [1,2,6]. This can be described in our model by considering an extended 1D slice with spatially periodic OP and eye preference. We take the basic unit cell shown in Fig. 1 and repeat it N times end to end: setting the unit cell length to $L \simeq 2r_2$ based on estimates of $\sim$1 mm unit cell size [5,15], and the $\sim$0.5 mm range of nonspecific lateral connections in V1 [15,16]. As the orientation difference $\Delta\phi$ between two monocular stimuli increases, binocular fusion gives way to binocular rivalry

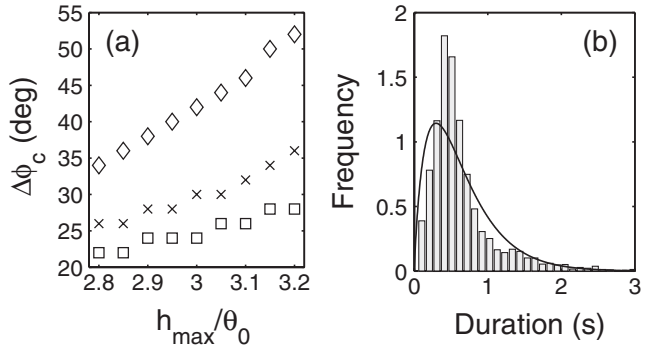

FIG. 4. Model predictions for binocular rivalry with parameters as in Fig. 2. (a) $\Delta\phi_c$ vs $h_{\max}$ for $\delta = 0.8$ (diamonds), $\delta = 1$ (crosses), and $\delta = 1.2$ (squares). (b) Dominance durations from (1) and (2) with added noise and fitted gamma distribution (solid line).

at $\Delta\phi_c = 180°(d_c/L)$, where d_c is the critical value of d for $N_s = 1$. For small $\Delta\phi \geq \Delta\phi_c$, interactions between solitons in neighboring unit cells are weak, so each soliton alternates independently on short time scales. A patchwork forms unless neighboring alternations all begin precisely in phase and the system is noise-free. As $\Delta\phi$ increases, the interactions become stronger, until at $\Delta\phi = 90°$ ($d = L/2$) there is equal reciprocal inhibition between all solitons and the activity in each unit cell is far from the $N_s = 1$ equilibrium. Fusion eventually sets in again as $\Delta\phi \to 180°$ and the orientations are again similar. In Fig. 4(a), $\Delta\phi_c$ is plotted for different values of the stimulus parameters. The results agree with [17], where $\Delta\phi_c$ for fusion-rivalry and rivalry-fusion transitions was 17.1°–30.4°. Figure 4(a) also shows $\Delta\phi_c$ increasing approximately linearly with increase in $h_{\max}$ and decrease in δ. The reason for the δ dependence is that it is less likely one soliton can span two stimulus regions as each region widens, so the soliton becomes unstable at smaller $\Delta\phi_c$. The h dependence is due to the above tendency of a stimulus to pin a soliton. This behavior can be interpreted using the energy (i.e., Lyapunov) function $\mathcal{F}$ of Eq. (1), which has a stimulus-dependent part given by $\mathcal{F}_{\text{stim}} = -\langle h, H(u_s - \theta_s)\rangle$, where $\langle f, g\rangle$ is the inner product of functions f and g [14]. This means that energy is minimized in regions of soliton-stimulus overlap where h is maximal, so increasing h stabilizes a soliton—raising $\Delta\phi_c$ in Fig. 4(a). In [17], a lower stimulus contrast led to larger $\Delta\phi_c$, meaning that $h_{\max}$ should be inversely related to stimulus contrast. This interpretation is supported by [18], where decreasing the contrast of one stimulus corresponded to a decrease in $\mathcal{F}$ (i.e., by increasing h) of the opposite stimulus, thereby lengthening its dominance duration—as found experimentally [1,2,18]. In our model, $\Delta\phi_c$ is therefore determined by OP periodicity and the range of lateral connections in V1, as well as specific stimulus attributes.

A histogram of stimulus dominance durations simulated from (1) and (2) with added noise is shown in Fig. 4(b). Specifically, we added $+1.5$ or -1.5 with equal probability to the right-hand side of (2) at each time step and each grid point. Each time u crossed $\theta \pm \Delta\theta$ (where $\pm\Delta\theta$ was used to ensure a definite switch in dominance had taken place) at a point chosen near the stimulus maximum, the time since the previous crossing was recorded. A total of 3137 crossings were simulated. A gamma distribution $\beta^\alpha x^{\alpha-1} e^{-\beta x}/\Gamma(\alpha)$ was fitted to the histogram in Fig. 4(b) by matching the mean and variance of both, giving $\alpha = 1.9$ and $\beta = 2.9$. These values are well within the ranges $\alpha = 1.3$–6.7, and $\beta = 0.46$–4.0 from fits to experimental and simulated distributions [6,7,9].

In summary, we have developed a soliton model that gives the first theoretical prediction of binocular rivalry onset. It also predicts two types of competitive neural dynamics consistent with rivalry and matches experimental results, including stimulus conditions for rivalry onset and the distribution of dominance durations.

The Australian Research Council supported this work.

[1] R. Blake, Psychol. Rev. **96**, 145 (1989).
[2] R. Blake and N. K. Logothetis, Nat. Rev. Neurosci. **3**, 13 (2002).
[3] A. Polonsky, R. Blake, J. Braun, and D. J. Heeger, Nat. Neurosci. **3**, 1153 (2000).
[4] S. H. Lee, R. Blake, and D. J. Heeger, Nat. Neurosci. **8**, 22 (2005).
[5] N. V. Swindale, Netw., Comput. Neural Syst. **7**, 161 (1996).
[6] S. R. Lehky, Perception **17**, 215 (1988).
[7] H. R. Wilson, B. Krupa, and F. Wilkinson, Nat. Neurosci. **3**, 170 (2000).
[8] H. R. Wilson, R. Blake, and S. H. Lee, Nature (London) **412**, 907 (2001).
[9] C. R. Laing and C. C. Chow, J. Comput. Neurosci. **12**, 39 (2002).
[10] S. Amari, Biol. Cybern. **27**, 77 (1977).
[11] S. Coombes and M. R. Owen, Phys. Rev. Lett. **94**, 148102 (2005).
[12] P. N. Loxley and P. A. Robinson, Biol. Cybern. **97**, 113 (2007).
[13] R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).
[14] P. N. Loxley and P. A. Robinson, Phys. Rev. E **76**, 046224 (2007).
[15] W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick, J. Neurosci. **17**, 2112 (1997).
[16] R. Malach, Y. Amir, M. Harel, and A. Grinvald, Proc. Natl. Acad. Sci. U.S.A. **90**, 10 469 (1993).
[17] A. Buckthought, J. Kim, and H. R. Wilson, Vision Res. **48**, 819 (2008).
[18] Y. J. Kim, M. Grabowecky, and S. Suzuki, Vision Res. **46**, 392 (2006).