Quantitative theory of driven nonlinear brain dynamics

J.A. Roberts a,b,c,⁎, P.A. Robinson a,b

a School of Physics, University of Sydney, New South Wales 2006, Australia
b Brain Dynamics Center, Sydney Medical School — Western, University of Sydney, Westmead, New South Wales 2145, Australia
 c Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia

A R T I C L E   I N F O
Article history:
Accepted 21 May 2012
Available online 29 May 2012

Keywords:
Steady state visual evoked potential
Entrainment
Neural field model
Nonlinear dynamics
EEG
Periodic stimulus

A B S T R A C T
Strong periodic stimuli such as bright flashing lights evoke nonlinear responses in the brain and interact nonlinearly with ongoing cortical activity, but the underlying mechanisms for these phenomena are poorly understood at present. The dominant features of these experimentally observed dynamics are reproduced by the dynamics of a quantitative neural field model subject to periodic drive. Model power spectra over a range of drive frequencies show agreement with multiple features of experimental measurements, exhibiting nonlinear effects including entrainment over a range of frequencies around the natural alpha frequency fα, subharmonic entrainment near 2fα, and harmonic generation. Further analysis of the driven dynamics as a function of the drive parameters reveals rich nonlinear dynamics that is predicted to be observable in future experiments at high drive amplitude, including period doubling, bistable phase-locking, hysteresis, wave mixing, and chaos indicated by positive Lyapunov exponents. Moreover, photosensitive seizures are predicted for physiologically realistic model parameters yielding bistability between healthy and seizure dynamics. These results demonstrate the applicability of neural field models to the new regime of periodically driven nonlinear dynamics, enabling interpretation of experimental data in terms of specific generating mechanisms and providing new tests of the theory.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Periodic flashing light, also termed flicker or photic driving, evokes responses in the electroencephalogram (EEG) that have been studied since the 1930s. Despite being widely used to probe visual function and perception (Müller et al., 2006; Nunez, 1995; Regan, 1989), the underlying neurophysiological mechanisms of these responses remain poorly understood. Early studies identified a clear nonlinear contribution to the driven cortical dynamics (van der Tweel and Spekreijse, 1969; van der Tweel and Verduyn Lunel, 1965). Nonlinearity in large-scale brain activity has seen considerable recent interest (Stam et al., 1999; Lehnertz and Elger, 1998; Robinson et al., 2002; Wendling et al., 2000), and functional neuroimaging (Stephan et al., 2008a). Notable nonlinear features in the EEG response to flicker include entrainment and frequency mixing (Herrmann, 2001; Rager and Singer, 1998; Regan, 1989; Townsend et al., 1975), both of which involve interactions between stimuli and ongoing activity and are thus of central importance to understanding brain function (Engel et al., 2001).

Periodic stimuli can also provoke epileptic seizures (Parra et al., 2005), particularly for driving frequencies near the natural alpha rhythm fα (≈10 Hz in adult humans), the most prominent feature of healthy resting EEG (Nunez, 1995). The mechanisms of these seizures are not well understood, nor is there a clear relation to the brain’s response in cases when the stimuli do not cause seizure. Thus there is a need for modeling to unify these nonlinear phenomena within a single framework.

Studies of the nonlinear mechanisms responsible for flicker-induced EEG signals have primarily taken a signal analysis approach (Kelly, 1981; Liu et al., 2010; Regan, 1989; van der Tweel and Spekreijse, 1969; van der Tweel and Verduyn Lunel, 1965), modeling time series data as arising from abstract filters. Although successful in describing input–output relationships, such models are difficult to generalize to other phenomena; it is preferable to have a single systematic theory. Indeed there is a strong need in neuroscience to move beyond models of time series to models of the underlying brain structures and connectivity, from which the observed dynamics emerge (Breakspear et al., 2010; Stephan et al., 2008b). Such modeling is now possible given recent developments in neural field theory (Coombes, 2010; Deco et al., 2008), particularly as it relates to the corticothalamic system (Robinson et al., 1997, 2002, 2004).

Neural field models describe the aggregate activity of populations of neurons and are thus especially appropriate for describing large-scale brain dynamics. Building on early work (Amari, 1975; Ermentrout and Cowan, 1979; Freeman, 1975; Lopes da Silva et al., 1974; Nunez, 1974;
Wilson and Cowan, 1973), neural field models (Baker and Cowan, 2009; Bressloff et al., 2002; Coombes, 2005; Hutt and Atay, 2005; Jirsa, 2009; Jirsa and Haken, 1996; Pinotsis and Friston, 2011; Pinotsis et al., 2012; Robinson et al., 1997; Taylor and Baier, 2011; Wright and Liley, 1996) and their limiting case of neural mass models (Cosandier-Rimélé et al., 2008; David and Friston, 2003; David et al., 2006; Goodfellow et al., 2011; Jansen and Rit, 1995; Lopes da Silva et al., 2003; Moran et al., 2009; Wendling et al., 2000) have successfully described a broad range of healthy and pathological brain dynamics at the spatial and temporal scales probed by EEGs and magnetoencephalograms. Neural field models have been notably successful in describing spatiotemporal dynamics in visual cortex (Baker and Cowan, 2009; Bressloff et al., 2002; Coombes, 2005; Ermentrout and Cowan, 1979), in particular explaining pattern formation during hallucinations. Visual evoked responses to individually-flashed stimuli have also been studied using neural mass models (David et al., 2006; Jansen and Rit, 1995), and (Spiegler et al., 2011) very recently applied the (Jansen and Rit, 1995) model to study entrainment and other nonlinear features of photic-driven activity. Both classes of model have explained a range of seizure phenomena, relating this pathological activity to healthy dynamics generated by the same models (Breakspear et al., 2006; Cosandier-Rimélé et al., 2008; Goodfellow et al., 2011; Kim et al., 2009; Lopes da Silva et al., 2003; Nevado-Holgado et al., 2012; Roberts and Robinson, 2006; Robinson et al., 2002; Taylor and Baier, 2011; Wendling et al., 2000).

In this paper we explain experimentally-observed nonlinear interactions between brain activity and flicker by incorporating periodic driving into a corticothalamic model (Roberts, 2010; Robinson et al., 1997, 2002). The results reproduce multiple features of published data and constrain the physiologically allowable states. We further predict various other nonlinear phenomena whose existence can be tested in future experiments, including period doubling, bistable phase-locking, hysteresis, wave–wave interactions, and chaos, and show how to treat entrainment and seizures in the same framework. This extends neural field modeling into a new regime and enables new experimental tests of the theory.

We focus on EEG responses to sinusoidally modulated light, which are termed steady-state visual evoked potentials (SSVEPs) (Nunez, 1995; Regan, 1989), but our approach is general and could be applied to any periodic stimulus. The SSVEP is dominated by synchronous cortical activity concentrated at the drive frequency \(f_D\) and its harmonics. Experiments have measured the response over a wide range of \(f_D\) in cats (Rager and Singer, 1998) and humans (Herrmann, 2001), showing clear spectral peaks, as seen in Fig. 1(a) for spatially uniform square wave modulated stimuli. Here, as in the original figure in Herrmann (2001), the spectrum for each \(f_D\) is multiplied by \(f_D\) to enhance peaks at higher frequencies, and the peaks appear as discrete dots due to being sampled on a coarse grid. These peaks involve both responses at \(f_D\) (which can be linear or nonlinear) and inherently nonlinear effects.

The clearest nonlinear effect is the generation of \(N:1\) harmonics (N response oscillations phase-locked to each drive oscillation). Since the square wave input signal used here contains only odd harmonics, the even harmonics in the output (in particular the prominent \(2:1\) harmonic) must be generated nonlinearly; power at the odd harmonic frequencies likely contains both linear and nonlinear components. The position of this nonlinearity in the visual pathway is not yet well established, and could involve effects in retinal, thalamic, and/or cortical neuronal populations.

Suppression of background alpha activity over a range of \(f_D\) is another significant nonlinear effect seen in Fig. 1(a), where the dominant response tracks \(f_D\) or \(2f_D\), while the dominant frequency is unchanged outside this range. (These oscillations are network-level collective modes, whose frequencies are not individual neural firing rates in general.) This is an example of entrainment, characterized by a reduction of ongoing activity in favor of activity phase-locked to the drive, at frequencies harmonically related to the drive frequency. Fig. 1(a) thus shows entrainment to \(f_D\) (1:1 phase locking) and subharmonic entrainment to \(f_D/2\) (1:2 locking).

Quantitative modeling, such as that used in the present work, will be critical in analyzing the data to distinguish between the above possibilities. In this context, it is particularly important to use established models, rather than introducing a new one ad hoc for each new experiment. Established models enable new phenomena to be incorporated into wider, unified frameworks, thereby relating them to other measures and situations, and permitting parameters to be constrained a priori, rather than being treated as free. The wide use of the Hodgkin–Huxley model is perhaps the best known example of this approach in neuroscience.

This paper is organized as follows. The Methods section outlines the nonlinear corticothalamic model. In the Results section the periodically driven dynamics of this model system are shown to reproduce the main features of the experimental SSVEP spectra in Fig. 1(a) as well as predicting rich nonlinear dynamics whose existence and properties can be explored in future experiments.

Methods

The Corticothalamic model section presents the nonlinear corticothalamic model, the dynamics of which will be analyzed using the time series methods described in the Time series analysis section.
Corticothalamic model

Neural field theories have successfully reproduced many aspects of the dynamics of populations of neurons by averaging over small spatial and temporal scales to give a description at scales ≥0.1 mm; see Deco et al. (2008) for a recent review. We use a recent corticothalamic model (Breakspear et al., 2006; Robinson et al., 1997, 2002, 2004) to analyze nonlinear SSVEPs (Roberts, 2010; Robinson et al., 2002, 2008). We follow the exposition and notation of Robinson et al. (2002) here, except where otherwise noted. The neuronal populations included are (see Fig. 2) excitatory (e) and inhibitory (i) cortical neurons, and the specific relay (s) and reticular (r) thalamic neurons. The relay in the visual system is the lateral geniculate nucleus, which has reciprocal connections with both the cortex and the reticular nucleus as shown in Fig. 2 (Sherman and Guillery, 1996; Steriade et al., 1997). External input from the brainstem and sensory systems (n) drives the system via s. Unstructured stimuli presented to the entire visual field drive a large area of cortex, so we concentrate on spatially uniform activity.

A nonlinear sigmoidal function $S(V_b)$, given by (Freeman, 1975; Wilson and Cowan, 1973)

$$S(V_b) = \frac{Q_{\text{max}}}{1 + \exp \left( -\frac{V_b - \theta}{\sigma} \right)} ,$$

relates the mean firing rate $Q_a = S(V_a)$ of each neuronal population $a$ to its mean cell body potential $V_a$ (relative to resting), where $Q_{\text{max}}$ is the maximum firing rate, $\theta$ is the mean threshold voltage and $\sigma \pi / \sqrt{3}$ is its population standard deviation. Here $V_b = \sum_s b S_{ab}$, with

$$D_{ab} V_{ab}(r, t) = V_{ab}\phi_a(r, t - \tau_{ab}),$$

$$D_{ab} = \frac{1}{c_{ab}^2} \frac{d^2}{dt^2} + \left( \frac{1}{\alpha_{ab}} + \frac{1}{\beta_{ab}} \right) \frac{d}{dt} + 1, \tag{3}$$

where $V_{ab}$ is the contribution due to input $\phi_0$ from neurons in population $b$ arriving after mean axonal delay $\tau_{ab}$, $D_{ab}$ gives the soma potential response allowing for synaptodendritic dynamics and soma capacitance, connection strength $V_{ab} = N_{ab} s_{ab}$. $N_{ab}$ is the mean number of synapses per neuron of type $a$ from neurons of type $b$, $s_{ab}$ is the strength of response to a unit input from neurons of type $b$, $c_{ab}$ and $s_{ab}$ are the inverse decay and rise times of the soma response to input $b$, respectively. For the system in Fig. 2, the only nonzero delays $\tau_{ab}$ are $\tau_{ee} = \tau_{ig} = \tau_{re} = \tau_{se} = \tau_{ef} = \tau_{i} / 2$ where $\tau_{i}$ is the total corticothalamic loop delay.

Average afferent firing rates $\phi_0$ propagate as fields in the continuum limit, approximately governed by a damped wave equation with source $Q_b$ (Jirsa and Haken, 1996; Nunez, 1995; Robinson et al., 1997), giving

$$\left[ \frac{1}{Y_e} \frac{d^2}{dt^2} + \frac{2}{Y_s} \frac{d}{dt} + 1 - r^2 \nabla^2 \right] \phi_b(r, t) = S[V_b(r, t)].$$

where $Y_e = V_e / r_a$, $V_e$ is the axonal conduction speed, and $r_a$ is the characteristic axon length. We assume that the activity is spatially uniform over the cortical area of interest, consistent with a large-scale mode driven by flicker stimuli; thus the $\nabla^2$ term is zero and for notational simplicity we omit the $\nabla$ argument hereafter. We set $\gamma_e = \infty$ for $a = e, i, s$ owing to the short ranges of these axons (Robinson et al., 2002). We retain long-range excitatory intracortical connectivity via nonzero $r_e$ and finite $\gamma_e$. Note that, although we neglect position dependence of activity, we have not reduced our model to a neural mass model, since this would also require shrinking the cortical manifold to be zero-dimensional by setting $r_e = 0$ and $\gamma_e = \infty$.

We assume random intracortical connectivity, which implies $V_{ab} = V_{ba}$ for $b = e, i, s$ and hence $V_i = V_e$, as derived in detail previously (Robinson et al., 1998). The inhibitory population is modeled explicitly ($V_{re} V_{ie}$, $V_{se}$, and $V_{re}$ are all nonzero); random connectivity implies that its mean-field dynamics are closely related to (but not identical to) those of the excitatory population with which it is spatially commingled (the excitatory neurons project much further, giving rise to the parameter $\gamma_e$). For simplicity we assume that all the synaptodendritic time constants are equal except for the inhibitory intrathalamic $\sigma$ connection, which is mediated by GABA$_B$ dynamics slower than the other channels (Steriade et al., 1997). Hence we set $c_{ab} = \alpha$, $s_{ab} = \beta$, $D_{ab} = D_a$ for all $ab$ except for the $sr$ connection; $\alpha$ and $\beta$ are thus to be interpreted as effective values. We do not explicitly model the bursting mode of thalamic neurons, which tends to only be relevant to frequencies lower than those typically observed in SSVEPs (Steriade et al., 1997).

We set parameters at or near previously published values (see figure captions) (Breakspear et al., 2006), which enables unification of the findings here with those of previous studies. Here the parameters place the system near a stability boundary associated with the intrathalamic loop (Robinson et al., 2002). All model parameters correspond to physiological quantities and have been rigorously constrained across numerous independent experimental measures (Robinson et al., 2004). Thus for the connectivity in Fig. 2 and the above assumptions, the model equations are

$$\frac{1}{Y_e^2} \frac{d^2}{dt^2} + \frac{2}{Y_s} \frac{d}{dt} + 1 \phi_e(t) = S[V_e(t)], \tag{5}$$

$$D_a V_e(t) = V_{ee} \phi_e(t) + V_{sr} \phi_r(t) + V_{se} \phi_s(t - \tau_0 / 2), \tag{6}$$

$$D_a V_s(t) = V_{se} \phi_e(t) + V_{sr} \phi_r(t) + V_{ss} \phi_s(t - \tau_0 / 2), \tag{7}$$

$$V_{i} = V_{ie} \phi_e(t) + V_{ir} \phi_r(t) + V_{in} \phi_n(t), \tag{8}$$

$$D_a V_{ie}(t) = V_{i} \phi_e(t - \tau_0 / 2), \tag{9}$$

$$D_a V_{ir}(t) = V_{i} \phi_r(t), \tag{10}$$

$$D_a V_{in}(t) = V_{i} \phi_n(t). \tag{11}$$

with $\phi_b(t) = S[V_b(t)]$ for $b = i, r, s$.

External stimuli $\phi_0$ drive the brain via the relay nuclei. Resting EEG spectra are successfully generated when driving the model with white noise (Robinson et al., 1997, 2001a, 2001b, 2002, 2004), which assumes that the collective input from all sensory pathways is so complex that it does not favor any particular frequency in the range of interest (Engel et al., 2001; Lopes da Silva et al., 1974; Stam et al., 1999). Impulsive stimuli yield evoked response potentials (Kerr et al., 2008; Robinson et al., 2001a, 2004). Sinusoidally modulated flicker stimuli generate SSVEPs (Robinson et al., 2002, 2008) and in the linear regime the model yields cortical phase velocities for SSVEPs consistent with experiment (Robinson et al., 2008). Here we concentrate on the nonlinear regime, where the drive is strong.

Fig. 2. Connectivities of cortical (excitatory e, inhibitory i) and thalamic (relay s, reticular r) neural populations; arrowheads ab denote where fields $\phi_0$ project, with connection strength $V_{ab}$ for input to a from b; $\phi_0$ denotes sensory input.
To study interaction between flicker and background activity we model $\phi_n$ as
\[ \phi_n(t) = \phi_n^{\text{noise}}(t) + \phi_n^{D}(t), \]
where $\phi_n^{\text{noise}}$ is Gaussian white noise with mean $\phi_n^{(0)}$ and standard deviation $\sigma_n$, and $\phi_n^{D}$ is a periodic drive with zero mean. We use two forms for $\phi_n^{D}$ in the Results section. For comparison with Fig. 1(a) we use a square wave drive given by
\[ \phi_n^{D}(t) = \phi_e \text{sgn}(2nf_0 t), \]
where $\phi_e$ is the drive amplitude and sgn indicates the sign function. This square wave consists of a sum of sinusoids, such that the drive contains power at all odd multiples of $f_0$. System responses at even multiples are necessarily nonlinear. To study the simpler case of a single sinusoid we use the form
\[ \phi_n^{D}(t) = \phi_e \cos(2nf_0 t), \]
for which all higher harmonics in the model output are necessarily generated nonlinearly.

Our use of the square wave drive Eq. (13) matches the photic drive used in Herrmann (2001), but neglects possible nonlinearities in the retina that might have modified the latter drive before it arrived at the corticothalamic system. We thus use a square wave so as to introduce the fewest additional assumptions, but note that it would be straightforward to implement other more complicated forms for the input. For example, the model could be explicitly coupled to models of retinal dynamics, which would enable it to compare and evaluate putative nonlinearities in the earliest parts of the visual stream. However, these topics are beyond the scope of this paper.

To solve the model equations numerically we use a standard explicit forward step integration scheme with a timestep of 0.1 ms (Press et al., 1992), of the type used by Robinson et al. (2002), for example. We have checked that the results are numerically stable and that the timestep is sufficiently short that they do not change significantly when it is further reduced.

Time series analysis

To characterize the frequency content of the model’s response to a range of stimulus parameters, we calculate power spectra from model time series. We estimate spectra from windowed Fourier transforms via Welch’s method (as implemented in MATLAB R2008b), using 10 s windows overlapping by 50%. To display sharp peaks more clearly, we additionally smooth spectra with a 0.4 Hz moving window.

To elucidate the role of nonlinear frequency coupling in generating peaks in the model EEG spectra we calculate bicoherence from model time series (Kim and Powers, 1979). Bicoherence is a nonlinear coherence measure that is sensitive to phase locking between Fourier components at different frequencies, and is derived from the bispectrum, a higher-order spectrum that generalizes the usual power spectrum (Kim and Powers, 1979). The bicoherence $b^2(\omega_1, \omega_2)$ between frequencies $\omega_1$ and $\omega_2$ of a signal with Fourier transform $X(\omega)$ is given by (Kim and Powers, 1979)
\[ b^2(\omega_1, \omega_2) = |X(\omega_1)X(\omega_2)|^2|X(\omega_1 + \omega_2)|^2 |X(\omega_1)X(\omega_2)|^2 |X(\omega_1 + \omega_2)|^2, \]
where $X'$ is the complex conjugate of $X$. We estimate the $X(\omega)$ terms by averaging FFTs of windowed time series (16384-point Hanning window, mean subtracted from each).

To characterize nonlinear dynamics at large drive amplitudes we estimate the largest Lyapunov exponent (LLE) (Pralitz, 1998) from model time series using the software package OpenTSTool v1.2 (Merkwirth et al., 2009). A positive LLE implies that nearby trajectories diverge exponentially and is indicative of chaos; periodic orbits have an LLE of zero. We use the method of time-delay embedding on long (80 s) time series for $\phi_e$ sampled at 1000 Hz, with the minimum embedding dimension chosen using the method of Cao (1997) with three nearest neighbors and 1000 reference points, and the delay chosen using the first minimum of the auto-mutual information function (see OpenTSTool documentation for further details). A benefit of using this method for LLE estimation is that it is also applicable to experimental time series (i.e., does not require access to the full underlying dynamical system).

To analyze phase-locking regimes as a function of drive amplitude and frequency, we calculate for each point in parameter space the ratio of the drive’s period to that of time series simulated in the absence of noise. We use the sequence of turning points in $\phi_e$ to calculate both the time series period (thus identifying $1:1$ phase-locking regimes) and the number of oscillations per period (identifying $N:1$ regimes, where the drive and $\phi_e$ have the same period).

Results

We begin in the Results section by driving the system with a square wave photic input to compare the model output to the data of Herrmann (2001) in Fig. 1(a). We analyze the driven dynamics in more detail using sinusoidal input in the Sinusoidal drive section.

Square wave drive

Fig. 1(b) shows the model response spectrum vs. $f_0$ for a square wave drive given by Eq. (13), with each spectrum multiplied by $f$ for comparison with Fig. 1(a). Key features are peaks at $f_0 = 1:1$ and its harmonics $(N:1)$, in good agreement with Fig. 1(a) (Herrmann, 2001; Rager and Singer, 1998). Most significantly, entrainment to $f_0$ is evidenced by suppression of the background alpha activity $f_0 \approx 11$ Hz in favor of $f_0$ for $f_0 \leq 15$ Hz, consistent with Fig. 1(a) where no strong alpha peaks are seen off the 1:1 diagonal. The alpha peak is suppressed without any changes to cortical or thalamic parameters; only the drive frequency changes. Further striking agreement is shown by $1:2$ subharmonic entrainment for $f_0 \approx 15–24$ Hz, where peaks follow $f = f_0/2$. Suppression of background activity is necessarily nonlinear; in the linear regime the drive and background superpose without interacting. The suppression observed experimentally cannot be attributed to averaging over trials, or else the alpha peak would not be observed for any fixed drive frequency contradicting Fig. 1(a). The extent to which the alpha activity is suppressed by the drive in the experimental data is not as clear as in the model; peaks in Fig. 1(b) are narrower than in Fig. 1(a) due to the FFT used here yielding better resolution than the autoregressive spectrum in Fig. 1(a). The experimental data also appears “lumpier” than the model data due to being sampled on only a coarse grid of frequencies; similarly coarse sampling in the model would yield an even more similar appearance. Additional responses in the model at $f = 3f_0/2$ and $f = 5f_0/2$ are not seen in Fig. 1(a), likely due to Fig. 1(a) being an average over ten subjects, thus further blurring peaks via variability between subjects.

Sinusoidal drive

In this subsection we drive the system with a single sinusoid to avoid ambiguity between spectral peaks that would be present in a nonsinusoidal drive and peaks generated by nonlinear processes. Fig. 3(a) shows that a sinusoidal drive given by Eq. (14) reproduces the main features of the EEG response to a square wave drive, demonstrating that entrainment and harmonic generation are not specific to square wave input. Here and in Fig. 3(b) additional features are revealed by plotting a wider dynamic range than in Fig. 1(a) (here we
also show the actual spectrum without multiplication by $f_t$. The model parameters here are modified slightly from those in Fig. 1(b) to better maintain the resemblance to key features of the data, such as widths of entrainment regimes. This parameter adjustment is needed because removing harmonics from the drive in going from square to sine wave reduces power at higher frequencies. This modifies the EEG spectra (such as by changing the mix of subharmonics generated), which can be approximately compensated for by small adjustment of model parameters. We stress that this step is carried out solely to probe the underlying mechanisms with more clarity and does not affect the level of agreement between the square-wave case and experiment found in the Square wave drive section.

Entrainment of the alpha frequency to $2f_D$ occurs over a narrow range where $2f_D \approx f_\alpha$; just above this drive frequency there is an enhancement where the background activity is shifted slightly downwards. Sum and difference frequencies $f_{\pm} = |f_D \pm f_\alpha|$ are generated by nonlinear interactions (other than entrainment) between the drive and background activity. This phenomenon is well-known as

![Image 1](https://example.com/image1.png)

![Image 2](https://example.com/image2.png)

Fig. 3. Model results for $(\nu_{ee}, \nu_{ei}, \nu_{es}, \nu_{se}, \nu_{sr}, \nu_{sn}, \nu_{re}, \nu_{rs}) = (1.3, -2.9, 0.13, 2.9, -0.57, 1.0, 0.67, 2.9) \text{ mV s}, \Phi_n = 2.8 \text{ s}^{-1}$, other parameters as in Fig. 1(b). (a) Spectral response for $f_D = 0-50 \text{ Hz}$, labeled as in Fig. 1(b) but with grayscale range extended to show lower-level peaks, and without multiplying by $f_t$. (b) Detail of (a); Features A-C are discussed in the text. (c) Bicoherence for SSVEP at $f_D = 30 \text{ Hz}$ and parameters of (a). (d) Turning points (after transients) of noise-free time series vs. $\Phi_n$ for $f_D = 20 \text{ Hz}$ and parameters of (a). Green and red points show bistable solutions obtained by continuation up and down in $\Phi_n$, respectively; yellow shows where red would otherwise overplot green. (e) Phase-locking zones for noise-free driven system initially at a stable fixed point for parameters of (a). (f) Largest Lyapunov exponent (LLE) vs. $\Phi_n$ for time series corresponding to the forward continuation in (d). Dark points in (a)-(c) correspond to high values.
mixing in nonlinear wave physics (Robinson et al., 2001a; Whittam, 1974), and is seen in SSVEP studies using two stimulus frequencies (Regan, 1989). To briefly illustrate how mixing occurs in a general setting, consider a signal \( y(t) = \sin(\omega_1 t) + \sin(\omega_2 t) \) comprising two sinusoids with frequencies \( \omega_1 \) and \( \omega_2 \). If this signal is passed through a quadratic nonlinearity (i.e., is squared), the resulting signal (after using trigonometric identities) is \( y^2(t) = 1 - \cos(2\omega_1 t)/2 - \cos(2\omega_2 t)/2 + \cos[(\omega_1 - \omega_2)t] - \cos[(\omega_1 + \omega_2)t] \), and thus contains second harmonics \( 2\omega_1 \) and \( 2\omega_2 \) and sum and difference frequencies \( \omega_1 + \omega_2 \) and \( \omega_1 - \omega_2 \), respectively. To have high power at frequencies \( f_1 \) requires a sharp alpha peak, so they are not observed in Fig. 1 (a) [except possibly near \( (0, 10) \) Hz], where the alpha peak is broader and intersubject variability tends to obscure this feature in the group average. Coherence (Kim and Powers, 1979) is shown in Fig. 3 (c) for \( f_D = 30 \) Hz. Clear peaks at \( (f_1, f_2) = (f_{0A}, f_{0B}) \) signify high phase coherence between \( f_{0A} \) and \( f_{0B} \) and \( f_0 + f_{0C} \). Peaks around \( (f_{0D}, f_{0E}) \) and \( (0, f_D) \) show phase coherence between \( f_{0D} \) and \( f_{0E} + f_D \) (2:1 harmonic generation; phase coherence with \( f_{1,2} = 0 \) represents the limiting case as \( f_{1,2} \rightarrow 0 \)). This is seen also on diagonals where \( f_1 + f_2 = f_{0C} = 2f_0, 3f_0 \) showing coupling between frequencies that sum to \( f_0, 2f_0, \) and \( 3f_0 \).

Entrainment involves more than just suppression of background peaks; enhancements A, B, and C in Fig. 3 (b) are precursors to phase-locked regions entered by increasing the amplitude (the maximum minus minimum shown) depends linearly in intervals around \( f_0 \) showing that 1:1 entrainment properties. This is complementary to constraints previously elucidated for linear background EEG spectra (Robinson et al., 2004).

Another key feature is the existence of bistability between 1:1 and 1:2 entrainment, shown by coexistence of green and red points around \( f_D \approx 1 \) s\(^{-1}\). Thus both 1:1 and 1:2 entrainments (the latter always exhibits 1:1 activity concurrently) exist at the same point in parameter space. The implication is that experimental SSVEPs at given \( f_D \) and \( f_0 \) are not necessarily unique, yielding a new source of intertrial variability that warrants study in its own right. We also predict hysteresis in studies where \( f_D \) is ramped slowly up and down through a zone where bistable entrained states exist, which would give different spectra on the up and down paths. This should be relatively straightforward to test; one possible place to look is where the 1:1 and 1:2 entrainment zones meet in Fig. 1. Hysteresis effects would likely be obscured in studies where stimuli at the \( f_D \) of interest are presented in random order, as is commonly used to avoid habituation effects (Regan, 1989). Spontaneous transitions between bistable states during a long recording session might also contribute to intertrial variability.

Extending the analysis of period doubling bifurcations in Fig. 3 (d) across a range of drive frequencies reveals a rich pattern of entrainment zones. Fig. 3 (e) shows phase-locking regions \( f_{0C} = f_{0D} = f_0, \) and \( f_{0E} \) for dynamics initially at a stable fixed point. For comparison, Fig. 3 (d) lies at \( f_0 = 20 \) Hz, and Fig. 3 (a) lies at \( f_D = 2.8 \) s\(^{-1}\) showing that 1:1 entrainment of \( f_0 \) occurs near the 1:2 regime boundary. The 1:1 region is near-linear for small \( f_0 \) but increasingly nonlinear at larger \( f_0 \) for example yielding regions of N:1 phase-locking at \( f_{0n} \approx (N+1)f_0/N \). The 1:1 regime bifurcates to 1:N regimes over ranges of \( f_D \) around (and for minimal \( f_D \) at) \( Nf_0 \), consistent with observed 1:3 and 1:8 entrainments (Hermann, 2001) (in Figs. 4 and 5 therein, respectively). The 1:N regions are always nonlinear, showing nested and overlapping period doublings toward dark blue regions where dynamics are variously of high period (\( f_D > 80 \) f), quasiperiodic, or chaotic. Gaps in the 1:3 family of period doublings around \( f_D = 30 \) Hz and \( f_0 = 14 \) s\(^{-1}\) reveal the locking regions underneath, hinting at a complicated structure reminiscent of overlapping Arnold tongues (Glass, 2001). The layout of entrainment zones in parameter space is also a new source of parameter constraints, because a given parameter set can be ruled out if it does not exhibit the correct entrainment properties. This is complementary to previous constraints in the linear regime, because satisfying these does not guarantee the same for the nonlinear regime. For example, the requirement that the system lie near a period doubling in the nonlinear dynamics is complementary to constraints previously elucidated for linear background EEG spectra (Robinson et al., 2004).

The structure in Fig. 3 (d) is for a driven system initially at a stable fixed point, not a pre-existing limit cycle, and the entrainment in

![Fig. 4. Spectra vs. f:D for increasing drive amplitude f:D with other parameters as in Fig. 3(a). (a) f:D = 1 s\(^{-1}\). (b) f:D = 2 s\(^{-1}\). (c) f:D = 3 s\(^{-1}\). (d) f:D = 4 s\(^{-1}\). Dark shading corresponds to high values.](image-url)
Fig. 1(b) is for noisy perturbations of a fixed point. Bifurcations of the fixed point itself (under changes in parameters other than the drive) affect entrainment. Numerically we find that entrainment requires the system to lie near a linear stability boundary, in accord with evidence that the brain operates near marginal stability (Breakspear et al., 2006; Robinson et al., 1997; Stam, 2005; Stam et al., 1999), and also near a nonlinear stability boundary, as seen in Fig. 3(e). Entrainment like that in Fig. 1 also requires that no other attractor coexist near the fixed point; otherwise the spectrum would change markedly. For a bistable limit cycle associated with seizure dynamics, such as near the subcritical 10 Hz bifurcation in the model (Breakspear et al., 2006), a sinusoidal drive near resonance perturbs the system away from the stable state onto a large-amplitude attractor, even for relatively weak stimuli (Kim et al., 2009); a typical case in Fig. 5 shows a dramatic increase in power at all frequencies for drives near $f_D \approx 9$ Hz and $2f_D$, with a waveform similar to the 10 Hz seizure in Robinson et al. (2002) and Breakspear et al. (2006); such behavior is characteristic of photosensitive epilepsy (Kim et al., 2009; Parra et al., 2005). The seizure frequency $f \approx 10$ Hz is little affected by the drive, which hardly perturbs the large-amplitude attractor.

Numerical exploration reveals that entrainment of $f_D$ is favored for parameters yielding decreases (2- to 10-fold relative to typical alert resting values) in mean firing rates $\phi_v$, $\phi_d$, and $\phi_r$, with $\phi_r$ increased (possibly substantially) or unchanged. Such decreases are qualitatively consistent with decreased metabolic load in these structures during entrainment (Parkes et al., 2004) (assuming that background activity is analogous to the aperiodic stimulation used there for comparison), suggesting that subnetworks activated by stimuli likely differ from those underlying resting activity, agreeing with recent experiments (Birca et al., 2006; Kerr et al., 2008). A relative increase in thalamic inhibition is plausibly a form of automatic gain control (Schwartz and Simoncelli, 2001), attenuating strong sensory stimuli before they are relayed to the cortex.

**Discussion**

We have studied driven nonlinear brain dynamics in a physiologically-based model of the cortex and thalamus, have compared its predictions with experiments from the literature, and have made a series of new predictions that can be tested in future experimental work. The main results are as follows:

(i) Our theory reproduces the key features of entrainment of the alpha rhythm to periodic stimuli, including entrainment to subharmonics of the drive, showing extensive agreement with experiment (Herrmann, 2001). Further good agreement is shown by the presence of nonlinearly-generated harmonics of the drive. The comparison in Fig. 1 demonstrates that the model agrees with experiment over wide ranges of drive and response frequencies.

(ii) The model predicts additional nonlinear dynamics to be found in future experiments. We predict bistability between different entrained states having the same drive amplitude and drive frequency, which can be tested experimentally by looking for hysteresis when slowly ramping the drive frequency up and down. Spontaneous transitions between bistable states during a long recording session might also contribute to intertrial variability, a possibility that could be tested experimentally.

(iii) At large drive amplitudes we predict period doubling of phase-locked leading states to quasiperiodic and chaotic dynamics. The 1:1 state bifurcates to a 1:N regime for minimal drive amplitude when $f_D \approx N f_a$, consistent with experiment (Herrmann, 2001).

(iv) We predict nonlinear sum and difference frequency generation in cases where background activity is sharply peaked but not entrained by the drive. This has been observed in studies using two stimulus frequencies simultaneously (Regan, 1989).

(v) The model predicts that when a stable limit cycle coexists with the stable fixed point corresponding to the resting state, periodic stimuli near the alpha frequency and its harmonics can drive the system into a seizure state. This plausibly explains seizures induced by high amplitude flashing light (Kim et al., 2009; Parra et al., 2005), and is open to further test.

These findings provide new verifications of the model, and of neural field theory more generally, complementary to previous studies. Since the same model has previously successfully described many linear and nonlinear dynamics (Breakspear et al., 2006; Robinson et al., 1997, 2002, 2004), this paper unifies these phenomena with nonlinear SSVEPs within the same framework using compatible parameters. Moreover, this work opens new interdisciplinary research avenues by predicting a rich variety of nonlinear dynamics that should be experimentally verifiable in human subjects. Quantitative modeling is essential for distinguishing between potential physiological mechanisms in such studies, and a model that also covers multiple other phenomena is essential to enable unification and avoid ad hoc interpretations.

It is worth pointing out that the data sets of Herrmann (2001) and Rager and Singer (1998) were acquired under quite different experimental conditions, yet our model reproduces the main features of both. Herrmann (2001) recorded scalp EEGs from awake human subjects, while Rager and Singer (1998) recorded intracortical multi-unit activity (MUA) and local field potentials (LFPs) from anesthetized cats. While EEGs necessarily involve filtering through the scalp, this is predominantly linear (David et al., 2006; Nunez, 1995; Robinson et al., 2001a) and so does not change the observed frequencies. In any case, EEGs and cortical LFPs are closely related because the same brain structures (mainly pyramidal cells) are ultimately responsible for both (Nunez, 1995; Steriade et al., 1997), and indeed the results of Herrmann (2001) and Rager and Singer (1998) are in good agreement. Two differences are that the cat LFP data do not show evidence of the subharmonics and gamma-band resonances seen in the human data; however, preliminary investigations indicate that our model can also reproduce this situation for plausible parameter values, though the exploration of appropriate parameters for the anesthetized cat is beyond the scope of this paper, and any such
study should also account for the fact that the cat “alpha” resonance is near 25 Hz (Nunez, 1995; Rager and Singer, 1998).

Parameter dependences of features such as entrainment regions, coupled with experiment, potentially enable calibration of parameters in the real brain, extending the utility of SSVEPs in probing physiology. For example, experiments could map out the phase locking zones as a function of drive frequency and amplitude for direct comparison with Fig. 3(e). This would be a particularly sharp test of the theory. Another measure that could be mapped experimentally is LLE; indeed Spiegler et al. (2011) recently performed such an exploration in their model. Formal model inversion schemes may also enable parameter inference by fitting directly to EEG data, particularly when coupled to model selection criteria (David et al., 2006; Moran et al., 2009; Pinotis et al., 2012), or by using state estimation techniques (Valdes-Sosa et al., 2009) for real-time parameter estimation; such methods are highly nontrivial for nonlinear time series and an active area of research.

We analyzed the driven dynamics in the vicinity of a stable fixed point, whereas Spiegler et al. (2011) analyzed interactions between an external drive and a pre-existing limit cycle. Recent evidence points toward both types of dynamics occurring in the alpha band with erratic switching between low and high amplitude modes (Freyer et al., 2009, 2011). Thus there is the potential for entrainment properties to fluctuate on similar time scales, or perhaps for the mode-switching dynamics to themselves be altered by a periodic drive. Such possibilities could be explored experimentally by measuring the time dependence of entrainment and other nonlinear measures.

Another important avenue for future theoretical work is to study phase relations between the drive and the ongoing and driven dynamics, thus complementing the frequency properties studied here. Phase-resetting of ongoing activity is thought to contribute to some ERP phenomena (Klimesch, 1999), and phase coupling in the brain is an area of intense research interest (Canolty et al., 2006; Varela et al., 2001).

Retention of spatial variation in activity and intracortical connectivity will enable study of spatiotemporal SSVEP dynamics (Nunez, 1995; Robinson et al., 2008) and gamma activity (Robinson, 2007), which is known to interact with flicker (Herrmann, 2001), although we have shown that neither is necessary to reproduce the entrainment properties discussed here. Spatial variation in both the cortex and the stimuli will also enable exploration of the rich array of patterns associated with visual hallucinations (Bressloff et al., 2002). In addition, spatially-varying stimuli such as stationary (Muthukumaraswamy and Singh, 2008) and moving (Swettenham et al., 2009) gratings evoke beta and gamma band (≥20 Hz) activity, and visual evoked responses to gratings differ in photosensitive subjects at both alpha and lower frequencies (Porciatti et al., 2000). These phenomena could all be explored in the model.

Here we focused on visual stimulation, but the same model and analysis should be applicable to other periodic sensory stimuli that pass through the thalamus, with appropriate parameter changes to account for the different thalamic relay nuclei and cortical subnetworks activated. For example, cortical responses to auditory stimulation in schizophrenic subjects and healthy controls have been shown to differ in their frequency profiles (Vierling-Glaassen et al., 2008), and such differences could be explored and interpreted with the same techniques as used here. Another potential avenue lies in driving cortical populations directly. This would enable modeling of nonsensory stimulation methods such as transcranial alternating current stimulation (tACS), which has been shown to interact with ongoing cortical activity in a frequency-dependent manner (Kanai et al., 2008).

In summary, the model applied here exhibits the main features of observed EEG responses to periodic visual stimuli, reproducing key nonlinear features of entrainment and harmonic generation, and enabling quantitative analysis in terms of underlying physiology. It does this by using an established approach that immediately relates these phenomena to ones in other fields to which the model has been applied. This unified framework permits parameters determined elsewhere to be used here; this is far more parsimonious than introducing an ad hoc approach with free parameters. Additional readily-verifiable nonlinear dynamics are also predicted for strong drives, including period doubling, bistability, and chaos, thereby enabling the theory to be further tested in future experiments. Moreover, photosensitive seizures emerge within the same framework, illustrating the ability of neural field models to unify a wide range of phenomena and opening new avenues for future work.

Acknowledgments

We thank J. W. Kim for helpful comments on the manuscript. The Australian Research Council and the Westmead Millennium Institute supported this work.

References
