



Neural Field Theory of Corticothalamic Prediction With Control Systems Analysis

Tahereh Babaie Janvier^{1,2*} and Peter A. Robinson^{1,2}

¹ School of Physics, University of Sydney, Sydney, NSW, Australia, ² Center of Excellence for Integrative Brain Function, University of Sydney, Sydney, NSW, Australia

Neural field theory is used to model and analyze realistic corticothalamic responses to simple visual stimuli. This yields system transfer functions that embody key features in common with those of engineering control systems, which enables interpretation of brain dynamics in terms of data filters. In particular, these features assist in finding internal signals that represent input stimuli and their changes, which are exactly the types of quantities used in control systems to enable prediction of future input signals, and adjustment of gains which is argued to be the analog of attention in control theory. Corticothalamic dynamics are shown to be analogous to the classical proportional-integral-derivative (PID) filters that are widely used in engineering.

Keywords: brain dynamics, cortex, thalamus, neuronal prediction, neural field theory, neuro-dynamic, control systems, filter

OPEN ACCESS

Edited by:

Tamer Demiralp,
Istanbul University, Turkey

Reviewed by:

Ahmet Ademoglu,
Boğaziçi University, Turkey
Murat Okatan,
Cumhuriyet University, Turkey
Karl Friston,
University College London,
United Kingdom

*Correspondence:

Tahereh Babaie Janvier
tahereh.babaie@sydney.edu.au

Received: 21 May 2018

Accepted: 02 August 2018

Published: 10 September 2018

Citation:

Babaie Janvier T and Robinson PA
(2018) Neural Field Theory of
Corticothalamic Prediction With
Control Systems Analysis.
Front. Hum. Neurosci. 12:334.
doi: 10.3389/fnhum.2018.00334

1. INTRODUCTION

The brain must carry out functions that implement attention to external stimuli, prediction of their future course, and decision with regard to actions to take in response, including interventions to control aspects of the environment. In the natural environment, a large fraction of visual processing deals with the prediction of spatiotemporal trajectories of objects and the taking of actions to either intercept or avoid them. For example, an organism may seek to capture and eat an object in its environment or to avoid being captured and eaten. These decisions thus differ markedly from the binary decision and discrete classification tasks widely studied in cognitive neuroscience.

With regard to attention to multiple sensory streams, there is extensive evidence that attention paid to individual streams is apportioned on an approximately Bayesian basis (Feldman and Friston, 2010; Friston, 2010). Notably, the weight placed on a particular stream is approximately inversely proportional to its variance (Feldman and Friston, 2010; Friston, 2010). These biological observations, along with advances in machine learning and neuroscience, have motivated a plethora of models of attention and prediction in the brain, each with Bayesian features, with most motivated by cortical neuroanatomy and neurophysiology. However, none of the proposed frameworks has yet been shown to be fully implementable in the brain's tissues.

One example of this point is the model of Rao and Ballard (1999), who used the Kalman filter to model visual information processing in the brain. A Kalman filter is a Bayesian approach that applies when all the distributions of internal states, external data, and their uncertainties, are Gaussian distributed and the underlying internal dynamics are linear. Rao and Ballard showed that a Kalman filter could accomplish some prediction outcomes, and demonstrated neuronal implementation of some of its steps, but did not explain how its all complex matrix operations would actually be implemented in the brain.

Helmholtz (1867) suggested that the brain predicts its inputs and adjusts an internal model to minimize mismatches between these predictions and external subsequent inputs. Friston and coworkers elaborated on this idea to develop Bayesian estimation schemes with both bottom-up and top-down signaling, and proposed that the brain employs a hierarchical Bayesian approach for perception and, crucially, action (Lee and Mumford, 2003; Friston, 2005, 2010; Hawkins and Blakeslee, 2007; Daunizeau et al., 2010b,a). The common theme of these approaches rests upon the minimization of the mismatch between top-down predictions and bottom-up expectations. Subsequent work Mathys et al. (2011, 2014) expanded classic Bayesian inference to optimize the precision of prediction errors which enabled them to include reinforcement learning into the ensuing *hierarchical Gaussian filter* (HGF) for perception, attention, and action. This family of models has been shown to be able to carry out a number of sophisticated prediction tasks, but this raises a number of questions. Specifically, they rely on the brain being able to determine, store, and update multivariate probability distributions of Bayesian priors, to carry out multidimensional integrations over these distributions, and to compute large scale matrix operations, all in real time. Although these problems have been recognized, and suggestions have been made to simplify the evaluations by means of reduced moment-based representations, for example (Bastos et al., 2012; Pouget et al., 2013; Mathys et al., 2014; Friston et al., 2017), it has yet to be established that the brain can carry out the necessary calculations.

The current state of affairs is thus that each proposal in the literature is motivated by the real brain, but relies in places on mathematical steps which have no established implementation in neural tissue. Hence, although these schemes have many plausible features, and many interesting applications have been demonstrated, none has been shown to be fully realizable in the brain.

Motivated by the need for a formulation of brain dynamics that is physiologically realizable, analytically and numerically tractable, and experimentally testable, we take a different approach. Instead of deciding on a favored mathematical formulation and assuming that it works in the brain, the present work takes the physically motivated reverse approach of first analyzing realistic corticothalamic responses to simple visual stimuli using neural field theory (NFT). This enables us determine what filter properties they exhibit, rather than advancing a predetermined model. We thus analyze the corticothalamic system, focusing on the response of primary visual cortex V1 to spatially unstructured stimuli which has been widely used in visual flicker experiments to probe steady state visual evoked potentials (SSVEPs) (Spekreijse, 1966; Spekreijse et al., 1973; Herrmann, 2001; VanRullen and Macdonald, 2012). In doing this we postpone application to more complex stimuli in order to focus on establishing the first example of a fully neurally implementable scheme for prediction, followed in a subsequent paper by attention and decision. This involves treating the brain first and foremost as a physical system that is responding to its environment, rather than as a computer or abstract information processing engine. Some of the dynamical processes that we

uncover may be able to be fruitfully interpreted as information processing or computation, but such interpretations are subject to the fact that the brain is a physical object.

We first employ NFT to model the corticothalamic system and determine the extent to which its dynamics can be interpreted within a control systems framework which has the potential to encompass prediction, gain tuning, and control. Most control systems and Bayesian schemes such as Kalman filtering are reduced or constraint form of Bayesian learning under optimal estimation theory (Chen, 2003). In carrying out our analysis, we determine which signals within the system represent input stimuli and their rates of change, because these are the classes of quantities used in control systems to enable prediction of future input signals, and adjustment of gains which is argued to be the analog of attention in control theory. The finding of analogous quantities in the corticothalamic system enables interpretation of its dynamics in terms of control systems, and assists in localizing the structures in which gain control is possible in principle. In a forthcoming paper, we will use these findings to propose physiologically-based mechanisms for attention and control via feedbacks. At each stage, we are explicit about the neural implementation of the mechanisms.

This paper is structured as follows. Section 2 briefly outlines the corticothalamic model of the brain using the neural field theory and then we obtain transfer functions for the corticothalamic model. The transfer functions are analyzed and data filter interpretations of them are presented at Section 3. Section 4 concludes and discusses future work.

2. MATERIALS AND METHODS

2.1. Corticothalamic Neural Field Theory

In this section we outline the essentials of NFT, then apply it to a generalized corticothalamic model. For more details see Robinson et al. (2002, 2004).

2.1.1. Neural Field Theory

NFT averages over short spatial and temporal scales to obtain equations for the evolution of dynamical variables within each neural population a , which are the local mean cell-body potentials V_a , the mean rate of firing at the cell body Q_a , and the propagating axonal pulse rate fields ϕ_a .

The mean firing rates Q_a are related to the cell body potentials $V_a(\mathbf{r}, t)$, relative to resting, by

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

where S is a sigmoid function that increases smoothly from 0 to Q_{\max} as $V_a(\mathbf{r}, t)$ increases from $-\infty$ to ∞ . An approximation of this function is (Wilson and Cowan, 1973; Freeman, 1975)

$$S[V_a(\mathbf{r}, t)] = \frac{Q_{\max}}{1 + \exp\{-[V_a(\mathbf{r}, t) - \theta]/\sigma'\}}, \quad (2)$$

where θ is the mean neural firing threshold and $\sigma'\pi/\sqrt{3}$ is the standard deviation of the difference between the steady state depolarization of individual neurons and their thresholds.

Effectively, the threshold response of a single neuron is smeared out to yield a sigmoid when averaged over the population.

Signals arriving at neurons of type a stimulate neurotransmitter release at synapses. This is followed by propagation of voltage changes along dendrites and soma charging, with dynamics that spread the temporal profile of the signals. The total cell body potential can thus be written

$$V_a(\mathbf{r}, t) = \sum_b V_{ab}(\mathbf{r}, t), \tag{3}$$

where the subscripts on V_{ab} distinguish the different combinations of afferent neural type and synaptic receptor, and

$$D_{ab}(t)V_{ab}(\mathbf{r}, t) = \sum_b N_{ab}s_{ab}\phi_b(\mathbf{r}, t - \tau_{ab}), \tag{4}$$

where the differential operator D_{ab} that governs the temporal response of V_{ab} to afferent pulse rate fields ϕ_b is

$$D_{ab}(t) = \frac{1}{\alpha_{ab}\beta_{ab}} \frac{d^2}{dt^2} + \left(\frac{1}{\alpha_{ab}} + \frac{1}{\beta_{ab}} \right) \frac{d}{dt} + 1. \tag{5}$$

The operator D_{ab} encapsulates the rates β_{ab} and α_{ab} of the rise and fall, respectively, of the response at the cell body. On the right of Equation (4), N_{ab} is the mean number of synapses on neurons a from neurons of type b , s_{ab} is the mean time-integrated strength of soma response per incoming spike, and $\phi_b(\mathbf{r}, t - \tau_{ab})$ is the mean spike arrival rate from neurons b , allowing for a time delay τ_{ab} due to anatomic separations between discrete structures. The overall connection strength between two neural populations is $v_{ab} = N_{ab}s_{ab}$.

Each part of the corticothalamic system gives rise to neural pulses, whose values averaged over short scales form a field

$\phi_a(\mathbf{r}, t)$ in our model that propagates at a velocity v_a . To a good approximation, $\phi_a(\mathbf{r}, t)$ obeys a damped wave equation whose source of pulses is $Q_a(\mathbf{r}, t)$ (Jirsa and Haken, 1996; Robinson et al., 1997), with

$$\mathcal{D}_a(\mathbf{r}, t)\phi_a(\mathbf{r}, t) = Q_a(\mathbf{r}, t), \tag{6}$$

where the spatiotemporal differential operator $\mathcal{D}_a(\mathbf{r}, t)$ is

$$\mathcal{D}_a(\mathbf{r}, t) = \frac{1}{\gamma_a^2} \frac{\partial^2}{\partial t^2} + \frac{2}{\gamma_a} \frac{\partial}{\partial t} + 1 - r_a^2 \nabla^2. \tag{7}$$

Here the damping rate γ_{ab} satisfies $\gamma_a = v_a/r_a$, where r_a and v_a are the characteristic range and conduction velocity of axons of type a .

2.1.2. Corticothalamic Model

Corticothalamic NFT incorporates key anatomic connectivities as shown in **Figure 1**. The neural populations included are cortical excitatory (e) and inhibitory (i) neurons, the thalamic reticular nucleus (r), thalamic relay neurons (s) that project to the cortex, thalamic interneurons (j), and noncorticothalamic neurons responsible for external inputs (n). In the present case, external inputs are visual, the relevant relay nucleus is the lateral geniculate nucleus (LGN), and its projections are to the primary visual cortex (V1).

The only nonzero values of τ_{ab} in our model are the forward delays $\tau_{es} = \tau_{is} \approx 20$ ms and the backward delays including $\tau_{se} = \tau_{je} = \tau_{re} \approx 60$ ms, which correspond to thalamocortical and corticothalamic propagation times, respectively. The use of a single form of D_{ab} corresponds to the approximation that the mean dendritic dynamics can be described by a single pair of time constants.

In our model, only the axons of excitatory cortical neurons $a = e$ are long enough to yield significant propagation

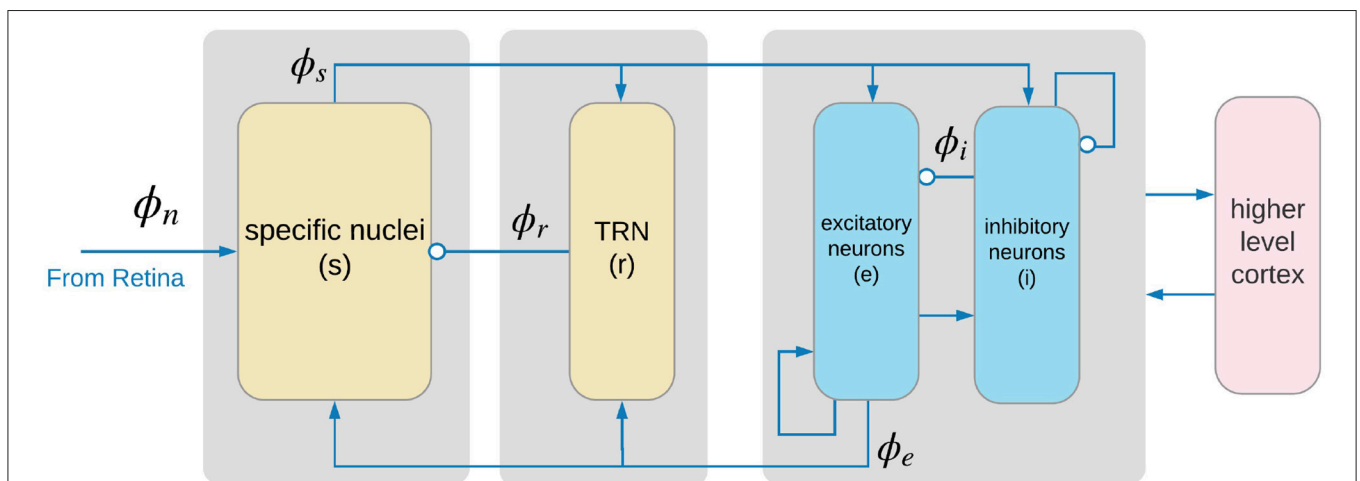


FIGURE 1 | Physiologically based corticothalamic model in which the arrows represent excitatory effects and the circles depict inhibitory ones. The populations are cortical excitatory (e) and inhibitory (i) neurons, the thalamic reticular nucleus (r), thalamic relay neurons (s) that project to the cortex, and non-corticothalamic neurons responsible for external inputs (n). Gray boxes depict the lateral geniculate nucleus (LGN) (left), the thalamic reticular nucleus (TRN) (middle), and primary visual cortex (right).

TABLE 1 | Physiologically estimated model parameters for normal adults in the alert, eyes-open state (Robinson et al., 2004).

Quantity	Description	Value	Unit
Q_{\max}	Max firing rate	250	s^{-1}
θ	Firing threshold	15	mV
σ'	Threshold spread	3.3	mV
γ_e	Cortical damping rate	100	s^{-1}
α_{ab}	Inverse decay time	80	s^{-1}
β_{ab}	Inverse rise time	320	s^{-1}
τ_{es}	Forward delay time	20	ms
τ_{se}	Feedback delay time	60	ms

effects in Equation (7); in the other cases, r_a is so small that the solution of equations above can be approximated by $\phi_a(\mathbf{r}, t) = S[V_a(\mathbf{r}, t)] = Q_a(\mathbf{r}, t)$. In the cortex, the number of synapses is closely proportional to the numbers of source and target neurons (Robinson et al., 1997, 1998; Braitenberg and Schüz, 1998), which implies that $v_{ee} = v_{ie}$, $v_{es} = v_{is}$, and $v_{ei} = v_{ij}$. **Table 1** lists the nominal values estimated for the corticothalamic model parameters; the details can be found in Robinson et al. (2004).

2.2. Steady States, Linearity, and Transfer Functions

The NFT equations are nonlinear in general, and highly nonlinear phenomena like epileptic seizures have been studied with them (Breakspear et al., 2006). Normal brain states have been shown to correspond to spatially uniform steady states of corticothalamic NFT, and are obtained by setting all time and space derivatives to zero (Robinson et al., 2002, 2004; Abeysuriya et al., 2015). Stable steady state solutions are interpreted as representing the baseline of normal activity, which yields firing rates in accord with experiment (Robinson et al., 2002, 2004). Time dependent brain activity is then represented by linear perturbations from the steady states—an approximation that has reproduced a host of experimental phenomena, including evoked responses (Robinson et al., 1997, 1998, 2002, 2004, 2005, 2008; O'Connor and Robinson, 2004; Rowe et al., 2004; Kerr et al., 2008; van Albada et al., 2010; Roberts and Robinson, 2012; Abeysuriya et al., 2015; Sanz-Leon and Robinson, 2017). In this section, we calculate the linear transfer functions of stimuli to corticothalamic populations.

2.2.1. Linear Dynamics

The low firing rate of steady-states have been identified with normal states of brain activity (Robinson et al., 1997) and nonlinear terms are only found to be significant in very strong stimulation conditions (Herrmann, 2001; Roberts and Robinson, 2012; Abeysuriya et al., 2015). The criterion for linear approximation to be valid is that voltage perturbations should be significantly less than $\sigma'/3$. Linear perturbations relative to uniform steady-state values $\phi_a^{(0)}$, where $a = e, i, r, s$, and $V_a^{(0)}$

approximately obey the damped wave equation

$$\mathcal{D}_{ab}(\mathbf{r}, t)\phi(\mathbf{r}, t) = \rho_a V_a(\mathbf{r}, t), \quad (8)$$

where we henceforth use the symbols ϕ_a and V_a to denote the linear perturbations of these quantities relative to their steady-state values, unless otherwise indicated, and $\rho_a = dS(V_a)/dV_a$, evaluated at V_a^0 . The external field ϕ_n which drives the brain via the relay nuclei also comprises a steady-state component $\phi_n^{(0)}$ plus a time-varying signal that causes the response, denoted by ϕ_n . The differences from threshold voltage variations of about $\pm\sigma'/3$ are therefore needed before nonlinear terms become appreciable relative to the linear ones. This yields the variations of order 1 mV, or slightly larger, which corresponds to approximately a 2-fold firing rate bound. Detailed analysis of the model with respect to these parameters can be found in Robinson et al. (2004).

Operation with \mathcal{D}_{ab} on both sides of Equation (8), plus use of Equation (4), yield

$$\mathcal{D}_{ab}(t)\mathcal{D}_{ab}(\mathbf{r}, t)\phi_a(\mathbf{r}, t) = \sum_b G_{ab}\phi_b(\mathbf{r}, t - \tau_{ab}), \quad (9)$$

with

$$G_{ab} = \rho_a v_{ab} = \rho_a N_{ab} s_{ab}. \quad (10)$$

The gain G_{ab} is the response in neurons a due to unit input from neurons b ; i.e., the number of additional pulses out for each additional pulse in.

A transfer function is the ratio of the output of a system to its input in the linear regime. Either the Laplace or Fourier transform can be used to determine transfer functions, but the former is more widely used in engineering control theory, particularly to analyze responses to impulses. To derive transfer functions one may apply the Laplace transform to both sides of Equation (9) to transform it from time t to complex frequency s . The unilateral Laplace transform is defined by Ogata and Yang (1970)

$$\mathcal{L}[f(t)](s) = f(s) = \int_0^{\infty} f(t)e^{-st} dt, \quad (11)$$

where $s = -i\omega = \Gamma - i\Omega$ is the complex frequency that parametrizes the response e^{st} . Here, the real quantities Ω and Γ denote the oscillation frequency and growth rate of the response, respectively, so stable responses correspond to s lying in the left half of the complex plane, with neutrally stable responses having s on the imaginary axis. Alternatively, one may use the continuous Fourier transform \mathcal{F} , which is equivalent to evaluating the bilateral Laplace operator with imaginary argument

$$\mathcal{F}[f(t)](\omega) = \mathcal{L}[f(t)]|_{s=-i\omega} \quad (12)$$

$$= \int_{-\infty}^{\infty} f(t)e^{i\omega t} dt. \quad (13)$$

Before we calculate system transfer functions, we note that the operator in Equation (4) has the Laplace transform

$$D_{ab}(s) = \left(1 + \frac{s}{\alpha_{ab}}\right) \left(1 + \frac{s}{\beta_{ab}}\right). \quad (14)$$

We define the corresponding filter function by

$$L_{ab}(s) = D_{ab}^{-1}(s) = \left(1 + \frac{s}{\alpha_{ab}}\right)^{-1} \left(1 + \frac{s}{\beta_{ab}}\right)^{-1}. \quad (15)$$

The Laplace transform of Equation (7) is

$$D_a(\mathbf{k}, s) = \left(1 + \frac{s}{\gamma_a}\right)^2 + k^2 r_a^2, \quad (16)$$

where we have also Fourier transformed the spatial Laplacian operator via $\nabla^2 \rightarrow -k^2$ where k is the wave number.

2.2.2. Transfer Function to Thalamus From Retina

The firing rate (in the spatial-Fourier, temporal-Laplace domain henceforth) in the reticular nucleus is

$$\phi_r = G_{re} L_{re} e^{-s\tau_{se}} \phi_e + G_{rs} L_{rs} \phi_s. \quad (17)$$

For relay neurons in the LGN, the firing rate ϕ_s is

$$\begin{aligned} \phi_s = & G_{se} L_{se} e^{-s\tau_{se}} \phi_e + G_{sj} L_{sj} \phi_j \\ & + G_{sr} L_{sr} \phi_r + G_{sn} L_{sn} \phi_n. \end{aligned} \quad (18)$$

By substituting ϕ_r from Equation (17) into Equation (18), the transfer function to thalamus from retina is found to be

$$T_{sn}(\mathbf{k}, s) = \frac{\phi_s(\mathbf{k}, s)}{\phi_n(\mathbf{k}, s)}, \quad (19)$$

$$= \frac{M(s)U(s)}{M(s)R(s) - N(s)P(s) \exp[-s(\tau_{es} + \tau_{se})]}, \quad (20)$$

where

$$M(s) = D_{ee}(1 - G_{ei}L_{ii}) - G_{ee}L_{ee}, \quad (21)$$

$$U(s) = 1 - G_{srs}L_{srs}, \quad (22)$$

$$R(s) = G_{sn}L_{sn}, \quad (23)$$

$$P(s) = G_{se}L_{se} + G_{sre}L_{sre}, \quad (24)$$

$$N(s) = G_{es}L_{es}, \quad (25)$$

where $G_{abc} = G_{ab}G_{bc}$, and $L_{abc} = L_{ab}L_{bc}$.

2.2.3. Transfer Functions to Cortex From Retina

By setting $a = e$ and $a = i$ in Equation (9), the axonal fields of V1 cortical cells are found to obey

$$D_{ee}\phi_e = L_{ee}G_{ee}\phi_e + L_{ei}G_{ei}\phi_i + L_{es}G_{es}e^{-s\tau_{es}}\phi_s, \quad (26)$$

$$\phi_i = L_{ii}G_{ii}\phi_i + L_{ie}G_{ie}\phi_e + L_{is}G_{is}e^{-s\tau_{es}}\phi_s. \quad (27)$$

Replacement of ϕ_i in Equation (26) by means of Equation (27), yields the transfer function to the cortex from the thalamus,

$$T_{es}(\mathbf{k}, s) = \frac{\phi_e(\mathbf{k}, s)}{\phi_s(\mathbf{k}, s)} = \frac{N(s)}{M(s)} \exp(-s\tau_{es}), \quad (28)$$

where $M(s)$ and $N(s)$ are as in Equations (21–25). Multiplying Equations (19) and (28) yields the overall transfer function to cortex from retina:

$$T_{en}(\mathbf{k}, s) = \frac{\phi_e(\mathbf{k}, s)}{\phi_n(\mathbf{k}, s)}, \quad (29)$$

$$= T_{es}(\mathbf{k}, s)T_{sn}(\mathbf{k}, s), \quad (30)$$

$$= \frac{U(s)N(s) \exp(-s\tau_{es})}{M(s)R(s) - N(s)P(s) \exp[-s(\tau_{es} + \tau_{se})]}. \quad (31)$$

Replacement of ϕ_s and ϕ_e in Equation (27) by means of Equations (19) and (31) yields the transfer function to the inhibitory population from the retina,

$$T_{in}(\mathbf{k}, s) = \frac{\phi_i(\mathbf{k}, s)}{\phi_n(\mathbf{k}, s)}, \quad (32)$$

$$= \frac{V(s)U(s) \exp(-s\tau_{es})}{O(s) [M(s)R(s) - N(s)P(s) \exp\{-s(\tau_{es} + \tau_{se})\}]}. \quad (33)$$

where

$$O(s) = 1 - G_{ii}L_{ii}, \quad (34)$$

$$V(s) = G_{ie}L_{ie}(s)N(s) + G_{is}L_{is}M(s). \quad (35)$$

2.2.4. Transfer Function to TRN From Retina

Replacement of ϕ_s and ϕ_e in Equation (17) by means of Equations (19) and (31), yields the transfer function to the TRN from retina,

$$T_{rn}(\mathbf{k}, s) = \frac{\phi_r(\mathbf{k}, s)}{\phi_n(\mathbf{k}, s)}, \quad (36)$$

$$= \frac{W(s)U(s)}{M(s)R(s) - N(s)P(s) \exp[-s(\tau_{es} + \tau_{se})]}, \quad (37)$$

where

$$W(s) = G_{re}L_{re}N(s) \exp[-s(\tau_{es} + \tau_{se})] + G_{rs}L_{rs}M(s). \quad (38)$$

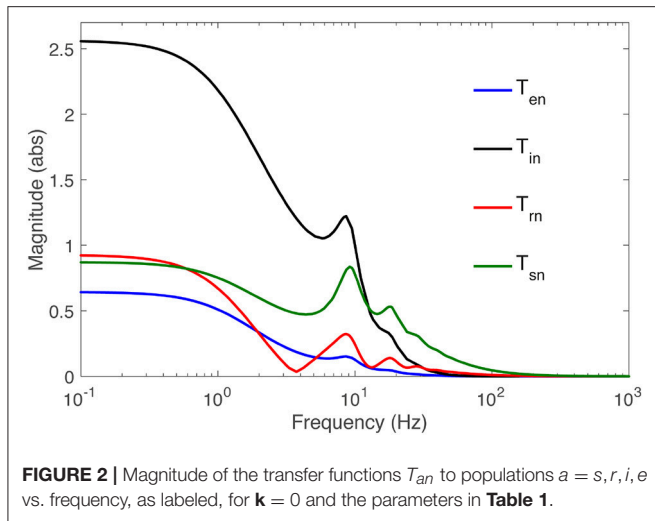


FIGURE 2 | Magnitude of the transfer functions T_{an} to populations $a = s, r, i, e$ vs. frequency, as labeled, for $\mathbf{k} = 0$ and the parameters in **Table 1**.

TABLE 2 | Estimated (dimensionless) synaptic gains for normal adults in the alert, eyes-open state (Robinson et al., 2004).

Gain	Value	Gain	Value
G_{ee}	6.8	G_{ie}	6.8
G_{se}	2.5	G_{re}	1.0
G_{ij}	8.1	G_{ei}	-8.1
G_{sr}	-1.9	G_{rs}	0.19
G_{es}	1.7	G_{is}	1.7
G_{sn}	0.8	G_{sn}	0.8

2.2.5. Corticothalamic Transfer Function Characteristics

The frequency response of the transfer functions is the transfer function evaluated on the imaginary axis of the s -plane, where $s = -i\Omega$. **Figure 2** shows the frequency responses of all populations for the nominal parameter values in **Table 2**, using the Control System Toolbox of Matlab 2017a to carry out the calculations. More detailed analysis of the model with respect to these parameters can be found in Robinson et al. (2002, 2004) and Abey Suriya et al. (2015). Only the the spatially-uniform effects of perturbations, i.e., $\mathbf{k} = 0$, is explored in this study. Low frequencies are passed, while high frequencies are attenuated, and for input signals with small frequency, each transfer function represents an amplifier with constant gain. At higher frequencies, pronounced resonances at 9 and 18 Hz are present in all transfer functions, which can be associated with the alpha and beta peaks in the brain’s wake state. The functions become less resonant as the signal gets further away from retina to the cortex: the thalamic functions T_{sn} and T_{rn} have higher amplitude and wider bandwidth than the cortical ones T_{en} and T_{in} .

The transfer function fully describes the linear system properties and enables us to investigate its response to any external signal. Setting the denominator of the transfer function to zero yields the characteristic equation of the

system, whose roots are its eigenvalues and mark the poles; these poles determine the basic modes into which the system response can be decomposed. Furthermore, all corticothalamic transfer functions calculated above, have some or all of their poles (basic modes) in common, which is a direct result of the interconnectedness of the system. Roots of the numerator of the transfer function are the zeros of the system; signals at these frequencies are not transferred through the system.

3. RESULTS

3.1. Control Systems Interpretation of Corticothalamic Transfer Functions

In this section, we decompose the transfer functions into elementary modes whose behaviors we associate with data filters whose control system properties are well understood. Only the spatially-uniform effects of perturbations, i.e., the simpler the description of the system, at $\mathbf{k} = 0$, have been explored in this study.

3.1.1. Reduced Model

The corticothalamic transfer functions, are ratios of exponential polynomials of s . We approximate each transfer function $T_{ab}(s)$ by a rational function, whose properties can be interpreted in terms of data filters (Ogata and Yang, 1970; Kwakernaak and Sivan, 1972), with

$$T_{ab}(s) = \frac{q(s)}{p(s)} = \frac{q(s)}{\prod_{j=1}^n (s + p_j)}, \quad (39)$$

where $q(s)$ and $p(s)$ are polynomials of degree m and n , respectively, with $m < n$. Therefore, when the p_j are all distinct (we do not consider degenerate roots here), one has the partial fraction decomposition

$$T_{ab}(s) = \sum_{j=1}^n \frac{r_j}{s + p_j}, \quad (40)$$

where the residues r_j are

$$r_j = (s + p_j)T_{ab}(s)|_{s=-p_j}. \quad (41)$$

The smaller n is, the simpler the description of the system, although accuracy is lost if n is made too small. In subsequent sections, we seek the smallest n that retains the main dynamics. Generally, this leads to the most heavily damped modes (poles with the largest negative real parts) being discarded.

3.1.2. Filter Identification

Once we have a few-pole approximation of the system transfer function, we examine it from a control-systems perspective to determine its predictive properties. Using Equation (40), the

response ϕ_b of population b to an input signal ϕ_a can be written as

$$\phi_b(s) = \sum_j \phi_b^j(s) = \sum_j \frac{r_j}{s + p_j} \phi_a(s), \quad (42)$$

where Equation (42) defines $\phi_b^j(s)$.

A general transfer function will have one or more pairs of complex conjugate poles in the Laplace domain, in addition to one or more real poles. Therefore, each pair of conjugate poles generates a real response mode. We also pair up real poles in the next part of the analysis to conveniently treat both cases together as second order filters whose functions are well known.

Hence, we consider the partial transfer function of the sum of two fractions associated with poles p_j and p_{j+1} either both real or conjugate pair, which we denote by $T_{ab}^j(s)$, with

$$T_{ab}^j(s) = K \frac{s + \tau_p^{-1}}{(s + p_j)(s + p_{j+1})}, \quad (43)$$

$$\tau_p = \frac{r_j + r_{j+1}}{r_j p_{j+1} + r_{j+1} p_j}, \quad (44)$$

$$K = r_j + r_{j+1}. \quad (45)$$

Equation (43) yields

$$\phi_b^j(s) = (s + \tau_p^{-1}) \left[\frac{K}{(s + p_j)(s + p_{j+1})} \right] \phi_a(s), \quad (46)$$

$$= H_{ba}^j(s) G_a^j(s) \phi_a(s). \quad (47)$$

By defining

$$\phi_a^I(s) = G_a^j(s) \phi_a(s), \quad (48)$$

we can write

$$\phi_b^j(s) = H_{ba}^j(s) \phi_a^I(s), \quad (49)$$

$$= H_{ba}^j(s) G_a^j(s) \phi_a(s). \quad (50)$$

Equations (48) and (49) express how the input $\phi_a(s)$ first passes through $G_a^j(s)$ to generate $\phi_a^I(s)$, which is transferred via $H_{ba}^j(s)$ to generate the response $\phi_b^j(s)$.

The input $\phi_a(t)$ passes through the filter $G_a^j(s)$,

$$G_a^j(s) = \frac{K}{(s + p_j)(s + p_{j+1})}. \quad (51)$$

Transforming Equation (51) to the time domain yields

$$\phi_a^I(t) = \int_0^t g(t - \tau) \phi_a(\tau) d\tau, \quad (52)$$

where

$$g(t) = ke^{-p_j t} - ke^{-p_{j+1} t}, \quad (53)$$

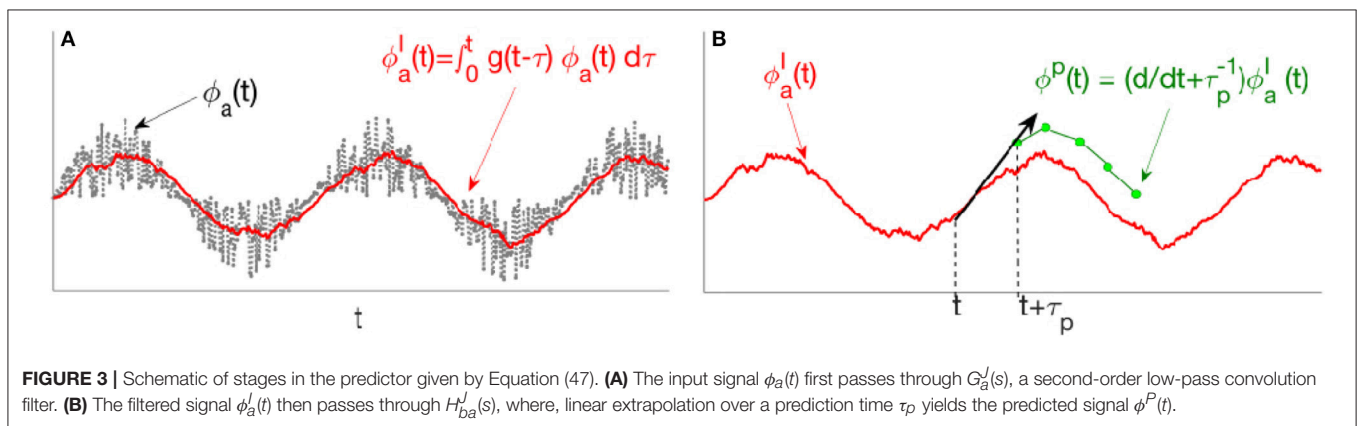
with

$$k = \frac{K}{|p_{j+1} - p_j|}; \quad (54)$$

hence this filter acts as an integrator in the form of a convolution; **Figure 3A** shows a schematic illustration of this convolution. This filter can be rewritten as a standard second-order filter, as

$$G_a^j(s) = \frac{K\Omega_0^2}{s^2 + 2\zeta\Omega_0 s + \Omega_0^2}, \quad (55)$$

where K is the low frequency gain, Ω_0 is the natural frequency, ζ is the damping coefficient, and the filter's bandwidth termed as B and is $2\zeta\Omega_0$. At $s \ll \Omega_0$ the right of Equation (55) approaches K , while in the opposite limit it approaches K/s^2 . When $\zeta > 1$, the



filter is overdamped and exponentially decays to the steady state without oscillating, while larger values of ζ yield a slower return to equilibrium. A critically damped filter has $\zeta = 1$ and returns to the steady state as quickly as possible without oscillating. When $\zeta < 1$, the filter is underdamped and exhibits damped oscillations. The peak frequency response occurs at $\pm\Omega_{peak}$, with

$$\Omega_{peak} = \begin{cases} 0, & \zeta \geq 1/\sqrt{2} \\ \Omega_0\sqrt{1 - 2\zeta^2}, & 0 \leq \zeta < 1/\sqrt{2} \end{cases} \quad (56)$$

The peak magnitude of $G_a^J(s)$ satisfies

$$G_a^J(i\Omega_{peak}) = \begin{cases} 1, & \zeta \geq 1/\sqrt{2} \\ \frac{1}{2\zeta\sqrt{1 - \zeta^2}}, & 0 \leq \zeta < 1/\sqrt{2} \end{cases} \quad (57)$$

Figure 4A shows the behavior of a nominal $G_a^J(i\Omega)$ when ζ changes, for fixed Ω_0 and K . **Figure 4B** shows the response of this filter when the input $\phi_a(t)$ is a Dirac delta function for a the same range of ζ .

The convolved signal ϕ_a^I is transferred through the filter $H_{ba}^J(s)$. By transforming Equation (49) into the time domain one then finds

$$\phi_b^J(t) = \tau_p^{-1}\phi^P(t), \quad (58)$$

$$\phi^P(t) = \tau_p \frac{d}{dt}\phi_a^I(t) + \phi_a^I(t), \quad (59)$$

where τ_p is termed the prediction time and ϕ_a^P the predicted signal. Hence, the filter H_{ba}^J predicts its own input a time τ_p in

the future, which is illustrated in **Figure 3B**. **Figure 4C** shows the magnitude of the T_{ba}^J for the same nominal $G_a^J(i\Omega)$, $\zeta = 0.3$, and $H_{ba}^J(s) = (s + \tau_p^{-1})$ when τ_p varies. **Figure 4D** shows how T_{ba} predicts the Dirac delta function when τ_p varies. At larger τ_p the response signal attains its final steady state value more slowly with a smaller resonant peak and less oscillation.

The above convolution plus prediction processes can be interpreted as a Proportional-Integral-Derivative (PID) control scheme used in engineering control systems (Ogata and Yang, 1970; Kwakernaak and Sivan, 1972). Specifically, in the time domain Equation (46) becomes

$$\phi_b^J(t) = \left(\frac{d}{dt} + \frac{1}{\tau_p} \right) \int_0^t g(t - \tau)\phi_a(t)d\tau, \quad (60)$$

which is a continuous-time serial PID. In this equation the integral (I) smooths the input signal to reduce the effects of noise. The smoothed signal is then extrapolated a time τ_p into the future by the combined effects of the proportional (P) and derivative (D) operators in the first parentheses on the right. Hence, each pair of poles in Equation (40) yields a partial transfer function that can be interpreted as a PID controller.

3.2. Control System Interpretation of Corticothalamic Dynamics

We now examine each corticothalamic transfer function from Section 2.2 to determine its control-systems characteristics.

3.2.1. Corticothalamic Filters

We find that a 16-pole approximation of T_{en} is accurate to within a root-mean-square (rms) fractional error of 0.01

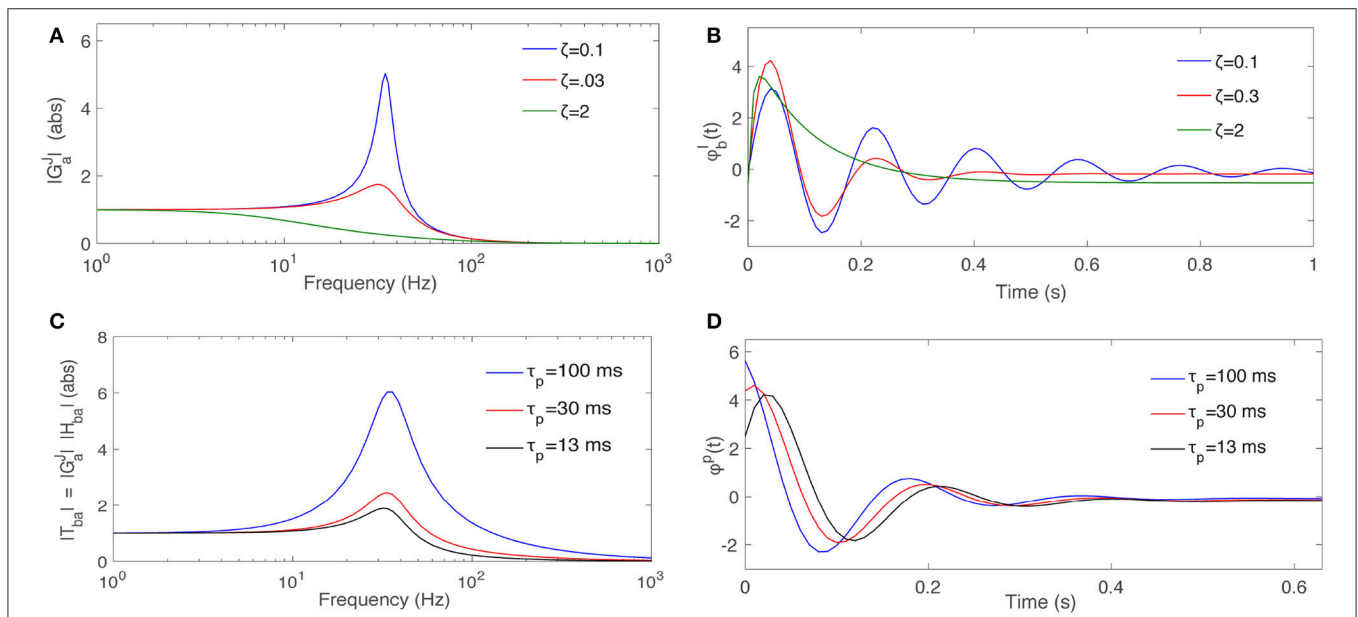
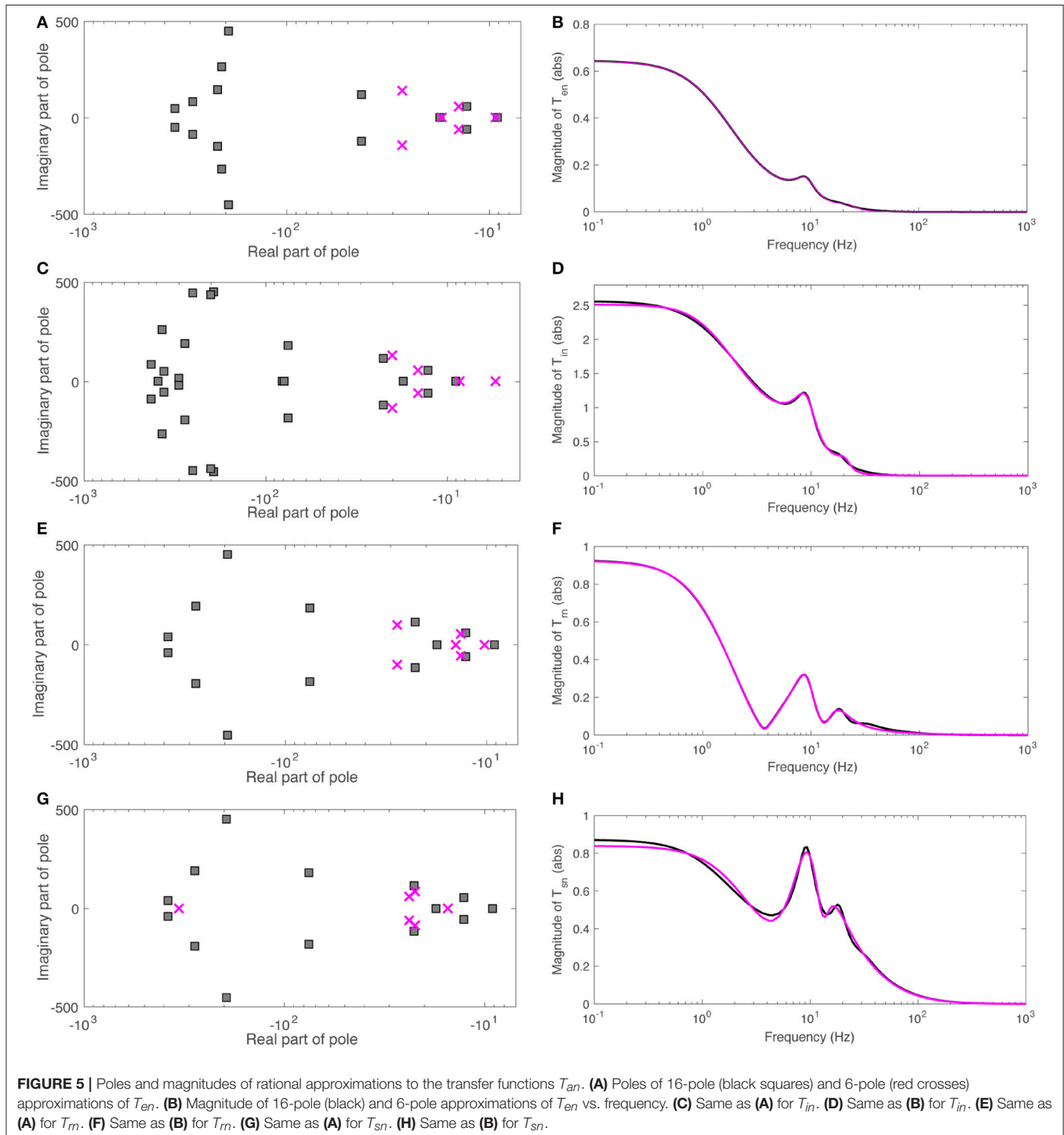


FIGURE 4 | Dependence of filter properties on parameters. **(A)** Magnitude of a typical second order filter $G_a^J(s)$ from Equation (55) for $\Omega_0 = 35 \text{ s}^{-1}$, $K = 1$, and various ζ , as labeled. **(B)** Response of $G_a^J(s)$ to an impulse input signal for various ζ . **(C)** Magnitude of the total $T_{ba}^J = H_{ba}^J(s)G_a^J(s)$, for fixed $\zeta = 0.3$ and $H_{ba}^J(s) = (s + \tau_p^{-1})$ for various τ_p . **(D)** Response of T_{ba}^J when the input signal is an impulse function for various τ_p .

over the frequency range 0 to 150 Hz for the parameters in **Table 2**, while a 6-pole approximation suffices for most purposes, accurate to within an rms fractional error of 0.02 over the same range. **Figure 5A** shows the poles of the 16- and 6-pole approximations of T_{en} , while **Figure 5B** shows the magnitudes of both functions. These results show that the 6-pole approximation retains the main features of the dynamics

and is sufficient for analyzing its effects; in most cases, it exhibits slightly shifted versions of the least-damped poles of the 16-pole approximation. Similar observations apply to the other transfer functions T_{in} , T_m , and T_{sn} whose 6-pole approximations are accurate to within an rms fractional error of 0.02. The pole maps and the frequency responses for these approximations are plotted in **Figures 5C–H**. We used the



Control System Toolbox of Matlab 2017a to carry out these approximations.

Turning to filter analysis, six poles can be summed in pairs that dominate at low ($f \lesssim 5$ Hz), alpha ($5 \text{ Hz} \lesssim f \lesssim 15$ Hz) and beta ($15 \text{ Hz} \lesssim f$) frequency ranges, respectively. We thus write

$$T_{bn}^r(s) = \sum_{j=1}^6 \frac{r_j}{s + p_j}, \tag{61}$$

$$= T_{bn}^{low}(s) + T_{bn}^{alpha}(s) + T_{bn}^{beta}(s), \tag{62}$$

where $b = s, r, i, e$ and T_{bn}^{low} is the sum of the two real poles while T_{bn}^{alpha} and T_{bn}^{beta} are the sums over the complex conjugate pairs of poles that represent oscillatory responses in the alpha and beta frequency ranges, respectively; in each case $p_j = \Gamma_j \pm i\Omega_j$.

3.2.2. Filter Properties of T_{bn}^{low}

This filter has two poles on the negative real axis at $p_1 = \Gamma^-$ and $p_2 = \Gamma^+$, with

$$\Gamma^\pm = -\left(\zeta \pm \sqrt{\zeta^2 - 1}\right) \Omega_0. \tag{63}$$

This corresponds to all $\zeta > 1$ and determines that G_n^{low} is an overdamped (no oscillation) low-pass filter. The cut-off frequency of G_n^{low} is $\Omega_0^{low} = \sqrt{\Gamma^- \Gamma^+} \text{ s}^{-1}$. The parameters of T_{bn}^{low} calculated for populations $b = s, r, i, e$ are listed in the **Table 3**. The impulse response takes the form

$$\phi_{bn}^{low}(t) = r^- \exp(\Gamma^- t) + r^+ \exp(\Gamma^+ t). \tag{64}$$

Because the second exponential decays faster than the first, the response is approximately first-order integrator with a time constant Γ^- for large ζ . Thus, poles closest to the positive half of the s -plane dominate the response.

3.2.3. Filter Properties of T_{bn}^{alpha} and T_{bn}^{beta}

Both filters share the same main features because each comprises a complex conjugate pair of poles. Therefore, we focus on the

alpha filter's properties. The poles are complex conjugates lying in the left half of the s -plane with $p_j, p_{j+1} = \Gamma \pm i\Omega_c$,

$$\Gamma = -\zeta \Omega_0, \tag{65}$$

$$\Omega_c = \Omega_0 \sqrt{1 - \zeta^2}, \tag{66}$$

where Ω_c is termed the cut-off frequency and $0 \leq \zeta < 1$. The frequency response of G_n^{alpha} has a resonance with

$$\Omega_{peak} = \Omega_0 \sqrt{1 - 2\zeta^2}, \tag{67}$$

$$|G_n^{alpha}(\Omega_{peak})| = \frac{1}{2\zeta \sqrt{1 - \zeta^2}}, \tag{68}$$

as the resonance frequency and magnitude, respectively. Note that $\Omega_{peak} \neq \Omega_c$ and that the two frequencies are equal only for $\zeta = 0$, which corresponds to a pure oscillator with $\Gamma = 0$, more generally, the resonance bandwidth is 2Γ .

Regarding the temporal response, we must consider the nature of the residues at the poles as well as the pole locations. Since $p_j = p_{j+1}^*$, these residues are conjugates, so the impulse response of G_n^{alpha} is

$$\begin{aligned} \phi_n^I(t) &= \frac{\Omega_0}{\sqrt{1 - \zeta^2}} \exp(-\Omega_0 \zeta t) \\ &\times \cos\left[\left(\Omega_0 \sqrt{1 - \zeta^2}\right) t - \frac{\pi}{2}\right]. \end{aligned} \tag{69}$$

and the predicted signal passed through H_{bn}^{alpha} is

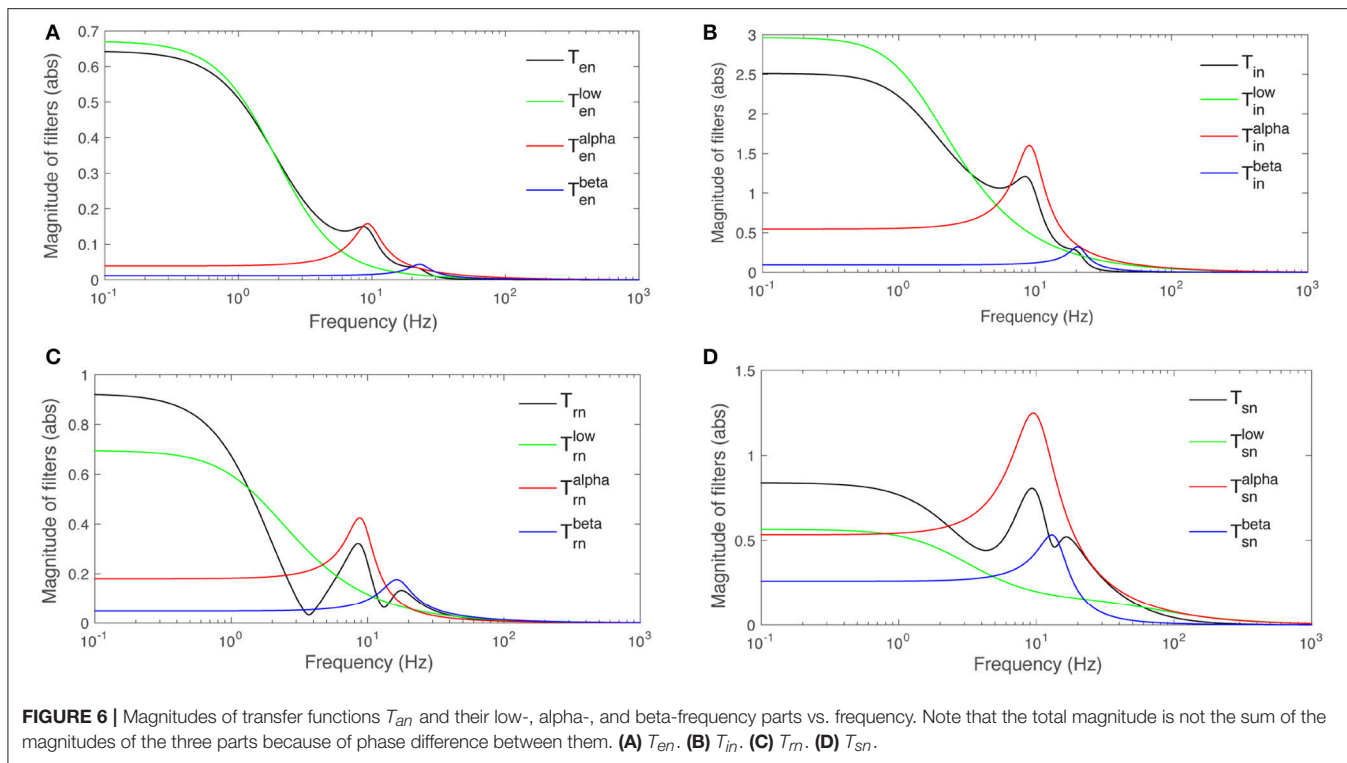
$$\phi_n^{alpha}(t) = \tau_p^{-1} \phi_n^I(t), \tag{70}$$

$$= 2|r| \exp(-\Omega_0 \zeta t) \cos\left[\left(\Omega_0 \sqrt{1 - \zeta^2}\right) t - \arg(r)\right]. \tag{71}$$

The parameters calculated for this filter are listed in **Table 3**. The corresponding frequency responses for $b = s, r, i, e$ are plotted in **Figure 6**. An alpha rhythm (7.5–12 Hz) is detectable in every population's alpha response, with peak frequencies ranging from

TABLE 3 | Parameters obtained for low, alpha, and beta filters of corticothalamic transfer functions T_{an} using filter identification method developed in Section 3.

Quantity	Description	T_{en}			T_{in}			T_{rn}			T_{sn}			Unit
		Low	Alpha	Beta	Low	Alpha	Beta	Low	Alpha	beta	Low	Alpha	Beta	
$-\Gamma^-$	damping rate 1 st pole	9.3	14.1	26.9	8.6	10.2	20	10.2	13.5	27	15.3	23.8	22.3	-
$-\Gamma^+$	damping rate 2 nd pole	17.2	14.1	26.9	5.5	14.5	20	14.5	13.5	27	337	23.8	22.3	-
Ω_c	cut-off frequency	0	57.4	143	0	56.8	130	0	56.2	101	0	57.9	84	s ⁻¹
$ K $	gain	1.8	3.82	1.62	28.7	33.1	3.4	7.44	4.3	8.3	52.2	48.8	4.6	-
ζ	damping coefficient	1.04	0.23	0.18	1.02	0.24	0.15	1.01	0.23	0.26	2.45	0.38	0.25	-
Ω_0	natural frequency	12.6	59.1	145	6.8	58.6	131	12	57.8	104	71.8	62.6	86.8	s ⁻¹
B	bandwidth	26.5	28.3	53.9	14.2	29	40.2	24.5	27	55.4	352	47.8	44.6	s ⁻¹
Ω_{peak}	resonance frequency	0	55.6	140	0	54.9	128	0	54.5	97.2	0	52.7	90	s ⁻¹
M_{peak}	resonance magnitude	1	2.1	2.7	1	2.1	3.3	1	2.2	2.0	1	1.4	2.0	-
τ_p	prediction time	16	28	6	20	17	2	73	7	15	18	23	2	ms



≈ 8.4 Hz for the specific nuclei to ≈ 9.3 Hz for inhibitory neurons in the cortex.

Similar results for beta filters are also listed in **Table 3** and plotted in **Figure 6**. The beta responses exhibit resonances in the beta band (12.5 – 30 Hz) where amplitudes are significantly smaller than the alpha resonances.

Both alpha and beta filters have higher peaks in thalamus than cortex. Furthermore, in every population except s , the damping rate of alpha filters are approximately half of the beta filters’, meaning the alpha waves live longer than beta waves in these populations. But, the result shows that alpha and beta waves would last for the same time in the response of population s where their damping rates are relatively close. This suggests that beta waves live longer in the LGN response to stimuli than in the rest of the system.

Calculating τ_p from the filter parameters yields

$$\tau_p \Omega_0 = \left[\sqrt{1 - \zeta^2} \frac{\text{Im}(r)}{\text{Re}(r)} - \zeta \right]^{-1}, \tag{72}$$

and shows that τ_p is governed by both the filters’ poles and their corresponding residues. Alpha and beta filters of the corticothalamic functions have small damping rates, $\zeta \ll 1$, and therefore the the prediction time in Equation (72) for these filters can be approximated as

$$\tau_p \Omega_0 \approx \frac{\text{Re}(r)}{\text{Im}(r)}. \tag{73}$$

The quantity $\tau_p \Omega_0$ represents what portion of its resonance cycle the filter predicts in advance. One significant observation is that alpha and beta filters of populations e , i and s present very similar patterns of having τ_p of alpha filters longer than those of their beta filters. In contrast, the population r has the opposite relation, which means its beta filter predicts further in advance.

3.3. Simulation With Random Input Signal

Aside from the issue of how to interpret corticothalamic dynamics in terms of data filters *per se*, there is the central question of how well these filters enable the system to predict its complex input signals out to some horizon in the future.

The corticothalamic system can only respond significantly to signals out to approximately the flicker fusion frequency of around 20 Hz. Hence, as a particularly severe test of its prediction capabilities, we simulate the system response to white noise, bandwidth limited to 30 Hz, with total power P_n . We are interested in time-varying, but spatially unstructured stimuli. The perturbation analysis then corresponds to presenting the entire field of view with a stimulus that consists of a sequence of luminances that fluctuates according to a small amplitude (perturbation) around a base level of intensity (steady-state). Such stimuli are widely employed in visual flicker experiments to probe SSVEPs (Herrmann, 2001; VanRullen and Macdonald, 2012). We used Control System Toolbox of Matlab 2017a to carry out the calculations of time responses for transfer function T_{en} using the relevant equations and parameters in the **Tables 1, 2**, and to generate a random seed for ϕ_n to stimulate the system. Note that the random noise does not affect spectra or other conclusions; it only changes the specific realization of the time

TABLE 4 | Optimal parameters computed for T_{en} filters, when stimulated with white noise of bandwidth 30 Hz and total power P_n .

Optimization Scheme	k_l	k_a	k_b	P^R/P_n	τ_d (ms)
Optimize the individual filters	4.4	0	0	0.53	49
	0	11.2	0	0.39	12
	0	0	28	0.33	6
Optimize the total of sum of filters through a single gain, as in Equation (76)	3.6	3.6	3.6	0.46	51
Optimize the total of sum of filters through separate gains, as in Equation (77)	1.9	8.9	28	0.29	50

series. We estimate the time offset τ_d between the measured output and the stimulus and denote the shifted signal by $\phi_e(t - \tau_d)$. We then define the difference between the stimulus and the output of each filter as

$$\phi^R(t) = \phi_n(t) - K\phi_e(t - \tau_d), \quad (74)$$

$$\approx \phi_n(t) - [k_l \quad k_a \quad k_b] \begin{bmatrix} \phi_e^{low}(t - \tau_d^l) \\ \phi_e^{alpha}(t - \tau_d^a) \\ \phi_e^{beta}(t - \tau_d^b) \end{bmatrix}, \quad (75)$$

which we term the residual signal. For each filter, we find the optimal K that minimizes the power of residual signal, termed as P^R . The power of the residuals is minimized to reduce the error between the predicted values and the real coming information subject to temporal frequencies that would normally be encountered in the signal part of sensory input. The resulting parameters are listed in the **Table 4**.

Figure 7A shows the optimized outputs of the slow filter contribution to T_{en} , and its residual signal compared to the stimuli. This filter predicts the slow trends of the external stimulus, corresponding to its large prediction horizon. **Figure 7B** shows the power spectra of the corresponding signals, which show that the narrow bandwidth of the filter limits it to predicting slow changes because it suppresses P^R at low frequencies. The filter has a relatively narrow bandwidth which results in an optimal k_l that reduces P^R/P_n to 0.53.

Figure 7C shows the optimally normalized outputs of the alpha filter contribution to T_{en} , and its residual signal compared to the stimuli. The alpha filter requires larger gain k_a than the low frequency filter and has a short delay time. **Figure 7D** shows the power spectra of the corresponding signals which show that this filter acts as an alpha band predictor with $P^R/P_n = 0.39$ that cannot forecast higher frequencies or slow trends. Similar results for the beta filter are plotted in **Figures 7E,F**. The results show a need for a significantly higher gain k_b to achieve the closest fit to

the stimuli; however, this filter can reduce $P^R/P_n = 0.33$ because of its significant tail at lower frequencies.

Slow, alpha, and beta filters detect and predict different frequency bands and generate responses, each of which focuses on information in the relevant band. This explains the need for a parallel set of filters normalized so that information about slow trends, and alpha and beta-band oscillations can be retrieved; the best response is obtained by summing the filters. We thus sum the filters in forms in which the gains can be controlled so that the full response is adjusted to optimally track the stimuli. In the first case, we normalize the total sum of the filters through a single factor, k , which might correspond to an overall neuromodulation, for example. This yields

$$\phi_{en}^{(I)} \approx k \left(\phi_{bn}^{low} + \phi_{bn}^{alpha} + \phi_{bn}^{beta} \right), \quad (76)$$

and we adjust k so that P^R is minimized. The parameters listed in **Table 4** then result in a minimum of $P^R/P_n = 0.46$. **Figure 7G** plots the normalized-shifted $\phi_{en}^{(I)}$ and the residual signal compared to the stimuli for these parameters. **Figure 7H** shows the corresponding power spectrum plots. The results are quite close to the results of individual low-frequency filter in terms of residual power and time delay, because T_{en}^{low} has the highest magnitude, and normalizing the sum of the filters by a single gain causes it to retain its dominance over the total response.

More accurate tracking of the stimuli could be achieved by tuning the parameters separately for each filter. If the gains k_m could be separately adjusted for $m = l, a, b$, denoting the low-frequency, alpha, and beta filters, respectively, then

$$\phi_{en}^{(II)} \approx k_l \phi_{bn}^{low} + k_a \phi_{bn}^{alpha} + k_b \phi_{bn}^{beta}. \quad (77)$$

Such an outcome might well be achievable by the real brain because we know that the strengths of the slow, alpha and beta peaks do not vary in lockstep. **Figure 7I** shows the response $\phi_{en}^{(II)}$ and the residual signal compared to the stimuli for the parameters listed in **Table 4**, yielding $P^R/P_n = 0.29$. **Figure 7J** confirms that this response contains information about slow trends of the stimulus, while alpha and beta band waves are predicted with more emphasis than in $\phi_{en}^{(I)}$.

Overall, these results show that the response of population e to the stimulus can be approximated by the sum of predicted values. To improve the performance, various normalization constants can potentially be allocated to different filters through gain adjustments. Although we focus on T_{en} here, the same approach can be used to reveal the mechanisms behind other populations' dynamics and prediction capabilities.

4. DISCUSSION

Motivated by the need for a formulation of brain's global dynamics that starts from its physical characteristics (rather than a predetermined formalism or endpoint) and is analytically,

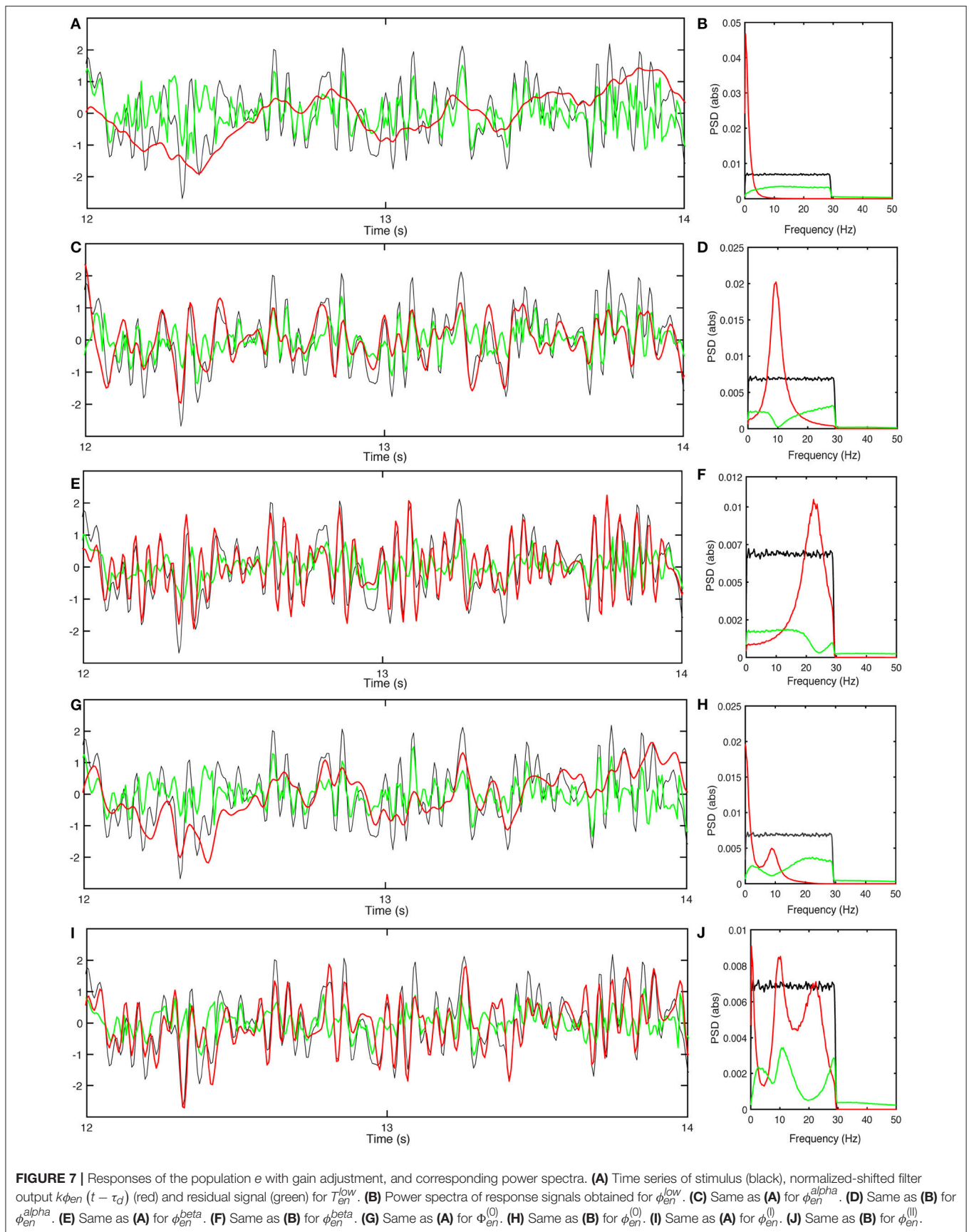


FIGURE 7 | Responses of the population e with gain adjustment, and corresponding power spectra. **(A)** Time series of stimulus (black), normalized-shifted filter output $k\phi_{en}(t - \tau_d)$ (red) and residual signal (green) for T_{en}^{low} . **(B)** Power spectra of response signals obtained for ϕ_{en}^{low} . **(C)** Same as **(A)** for ϕ_{en}^{alpha} . **(D)** Same as **(B)** for ϕ_{en}^{alpha} . **(E)** Same as **(A)** for ϕ_{en}^{beta} . **(F)** Same as **(B)** for ϕ_{en}^{beta} . **(G)** Same as **(A)** for $\Phi_{en}^{(0)}$. **(H)** Same as **(B)** for $\phi_{en}^{(0)}$. **(I)** Same as **(A)** for $\phi_{en}^{(I)}$. **(J)** Same as **(B)** for $\phi_{en}^{(II)}$.

numerically, and experimentally tractable, we have used a neural field corticothalamic model to evaluate the transfer functions for various population of neurons. Particular attention has been paid to the understanding of the transfer functions from a control systems perspective. The main results are:

(i) NFT transfer functions for each corticothalamic population were derived and approximated as rational polynomials. These were explored to determine the linear response of each population, and its rate of change, to any stimulus, which are exactly the types of quantities that are used in control systems to enable prediction of future states, and adjustment of gains which is argued to be the analog of implementing attention.

(ii) We approximated transfer functions, while preserving accuracy over the dominant frequency range. These were then used to uncover basic modes by which each population in the corticothalamic system responds to external signals. All corticothalamic transfer functions were then shown to be dominantly governed by a few basic modes and to have similar frequency responses, with different gains.

(iii) Notably, it was shown that each pair of basic modes yields a sub-response that can be expressed in terms of a standard second order PID filter with a low-frequency, alpha, or beta resonance.

(iv) Slow, alpha, and beta filters operate in parallel, with each filter capturing and processing part of the information coming from the external world. The total response to stimuli is obtained by summing these independent responses. This corresponds to the common practice in control systems of using a set of controllers to improve the performance of a tracking system.

(v) Using a random white noise covering the bandwidth of human vision as the stimulus, the responses of corticothalamic filters were simulated. We explored the tracking performance of each filter as well as their sum for the excitatory population in the

cortex. The results showed each filter successfully tracks part of the stimulus, while a better prediction is obtained by summing them after separate optimal gain adjustments. This showed how the filters' gains can be adjusted to improve prediction of the future input signal based on its recent time course (value and derives, implicitly) and internal corticothalamic dynamics. Attention, in the other words, may be implemented in the brain through control and adjustment of input gains. This suggested directions for modeling attention in our framework by using mismatches between internal models and external stimuli to drive gain changes. This study will lead to modeling decision and control, and generalization to encompass more general stimuli and sensory systems.

Overall, the results show and interpret the utility of control theory schemes in understanding brain's dynamics, yielding insights into dynamic processes that underlie cognition and action without imposing an outcome in advance.

AUTHOR CONTRIBUTIONS

TB and PR conceived of the presented idea. TB developed the theory and performed the computations. TB and PR verified the analytical methods. TB worked out the technical details, performed numerical calculations, produced figures and numerical results. TB and PR discussed all the results. TB drafted the first manuscript and both authors contributed to the final manuscript.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council under Center of Excellence grant no. CE140100007 and Laureate Fellowship grant no. FL140100025.

REFERENCES

- Abeyururiya, R. G., Rennie, C., and Robinson, P. (2015). Physiologically based arousal state estimation and dynamics. *J. Neurosci. Methods* 253, 55–69. doi: 10.1016/j.jneumeth.2015.06.002
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Braitenberg, V., and Schüz. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*, (2nd edn.). Heidelberg: Springer-Verlag.
- Breakspear, M., Roberts, J., Terry, J. R., Rodrigues, S., Mahant, N., and Robinson, P. (2006). A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. *Cereb. Cortex* 16, 1296–1313. doi: 10.1093/cercor/bhj072
- Chen, Z. (2003). Bayesian filtering: From kalman filters to particle filters, and beyond. *Statistics* 182, 1–69. doi: 10.1080/02331880309257
- Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Friston, K., and Stephan, K. E. (2010a). Observing the observer (ii): deciding when to decide. *PLoS ONE* 5:e15555. doi: 10.1371/journal.pone.0015555
- Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Stephan, K. E., and Friston, K. (2010b). Observing the observer (i): meta-bayesian models of learning and decision-making. *PLoS ONE* 5:e15554. doi: 10.1371/journal.pone.0015554
- Feldman, H., and Friston, K. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215
- Freeman, W. J. (1975). *Mass Action in the Nervous System*. New York, NY: Academic Press.
- Friston, K. (2005). A theory of cortical responses. *Phil. Trans. Roy. Soc. B* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). Active inference: a process theory. *Neural Comput.* 29, 1–49. doi: 10.1162/NECO_a_00912
- Hawkins, J., and Blakeslee, S. (2007). *On Intelligence*. New York, NY: Macmillan.
- Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*, Vol. 9. Hamburg: Voss.
- Herrmann, C. (2001). Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exper. Brain Res.* 137, 346–353. doi: 10.1007/s002210100682
- Jirsa, V. K., and Haken, H. (1996). Field theory of electromagnetic brain activity. *Phys. Rev. Lett.* 77, 960–963. doi: 10.1103/PhysRevLett.77.960
- Kerr, C. C., Rennie, C. J., and Robinson, P. A. (2008). Physiology-based modeling of cortical auditory evoked potentials. *Biol. Cyber.* 98, 171–184. doi: 10.1007/s00422-007-0201-1
- Kwakernaak, H., and Sivan, R. (1972). *Linear Optimal Control Systems*, Vol. 1. New York, NY: Wiley-Interscience.
- Lee, T. S., and Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *J Opt. Soc. Am. A* 20, 1434–1448. doi: 10.1364/JOSAA.20.001434

- Mathys, C., Daunizeau, J., Friston, K., and Stephan, K. (2011). A bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* 5:39. doi: 10.3389/fnhum.2011.00039
- Mathys, C. D., Lomakina, E., Daunizeau, J., Iglesias, S., Brodersen, K., Friston, K., et al. (2014). Uncertainty in perception and the hierarchical gaussian filter. *Front. Hum. Neurosci.* 8:825. doi: 10.3389/fnhum.2014.00825
- O'Connor, S. C., and Robinson, P. (2004). Spatially uniform and nonuniform analyses of electroencephalographic dynamics, with application to the topography of the alpha rhythm. *Phys. Rev. E* 70:011911. doi: 10.1103/PhysRevE.70.011911
- Ogata, K., and Yang, Y. (1970). *Modern Control Engineering*. Englewood Cliffs, NJ: Prentice-Hall.
- Pouget, A., Beck, J., Ma, W., and Latham, P. (2013). Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 16, 1170–1178. doi: 10.1038/nn.3495
- Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Roberts, J. A., and Robinson, P. A. (2012). Quantitative theory of driven nonlinear brain dynamics. *Neuroimage* 62, 1947–1955. doi: 10.1016/j.neuroimage.2012.05.054
- Robinson, P. A., Chen, P.-C., and Yang, L. (2008). Physiologically based calculation of steady-state evoked potentials and cortical wave velocities. *Biol. Cyber.* 98, 1–10. doi: 10.1007/s00422-007-0191-z
- Robinson, P. A., Rennie, C., Rowe, D., O'Connor, S., and Gordon, E. (2005). Multiscale brain modelling. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1043–1050. doi: 10.1098/rstb.2005.1638
- Robinson, P. A., Rennie, C. J., and Rowe, D. (2002). Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. *Phys. Rev. E.* 65:041924. doi: 10.1103/PhysRevE.65.041924
- Robinson, P. A., Rennie, C. J., Rowe, D. L., and O'Connor, S. C. (2004). Estimation of multiscale neurophysiologic parameters by electroencephalographic means. *Hum. Brain Mapp.* 23, 53–72. doi: 10.1002/hbm.20032
- Robinson, P. A., Rennie, C. J., and Wright, J. J. (1997). Propagation and stability of waves of electrical activity in the cerebral cortex. *Phys. Rev. E* 56:826. doi: 10.1103/PhysRevE.56.826
- Robinson, P. A., Rennie, C. J., Wright, J. J., and Bourke, P. D. (1998). Steady states and global dynamics of electrical activity in the cerebral cortex. *Phys. Rev. E* 58:3557. doi: 10.1103/PhysRevE.58.3557
- Rowe, D. L., Robinson, P. A., and Rennie, C. J. (2004). Estimation of neurophysiological parameters from the waking eeg using a biophysical model of brain dynamics. *J. Theor. Biol.* 231, 413–433. doi: 10.1016/j.jtbi.2004.07.004
- Sanz-Leon, P., and Robinson, P. (2017). Multistability in the corticothalamic system. *J. Theor. Biol.* 432, 141–156. doi: 10.1016/j.jtbi.2017.07.015
- Spekreijse, H. (1966). *Analysis of EEG-Responses in Man. Evoked by Sine Wave Modulated Light*. Thesis, University of Amsterdam, Amsterdam; W. Junk Publ. The Hague, The Netherlands.
- Spekreijse, H., Van der Tweel, L., and Zuidema, T. (1973). Contrast evoked responses in man. *Vision Res.* 13, 1577–1601. doi: 10.1016/0042-6989(73)90016-3
- van Albada, S., Kerr, C., Chiang, A., Rennie, C., and Robinson, P. (2010). Neurophysiological changes with age probed by inverse modeling of eeg spectra. *Clin. Neurophysiol.* 121, 21–38. doi: 10.1016/j.clinph.2009.09.021
- VanRullen, R., and Macdonald, J. S. (2012). Perceptual echoes at 10 hz in the human brain. *Curr. Biol.* 22, 995–999. doi: 10.1016/j.cub.2012.03.050
- Wilson, H. R., and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol. Cybern.* 13, 55–80. doi: 10.1007/BF00288786

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

Copyright © 2018 Babaie Janvier and Robinson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.