



Neural Field Theory of Corticothalamic Attention With Control System Analysis

Tara 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 analyze attention by extending an existing model of the large-scale activity in the corticothalamic system to incorporate local feedbacks that modulate the gains of neural connectivity as part of the response to incoming stimuli. Treatment of both activity changes and connectivity changes as part of a generalized response enables generalized linear transfer functions of the combined response to be derived. These are then analyzed and interpreted via control theory in terms of stimulus-driven changes in system resonances that were recently shown to implement data filtering and prediction of the inputs. Using simple visual stimuli as a test case, it is shown that the gain response can implement attention by evaluating two main features of the stimuli: the magnitude and the rate of change, by increasing the weight placed on the rate of change in response to sudden changes, while reducing the contribution of stimuli value in tandem. These changes of filter parameters are shown to improve the prediction of the upcoming stimuli based on its recent time course. This outcome is analogous to controller-parameter tuning for performance enhancement in engineering control theory.

Keywords: brain dynamics, attention, prediction, cortex, thalamus, neural field theory, control systems, data filter

OPEN ACCESS

Edited by:

Jesús Malo,
University of Valencia, Spain

Reviewed by:

Jordi Soriano-Fradera,
University of Barcelona, Spain
Ravinder Jerath,
Charitable Medical Healthcare
Foundation, United States

*Correspondence:

Tara Babaie-Janvier
tara.babaie-janvier@sydney.edu.au

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Neuroscience

Received: 29 March 2019

Accepted: 04 November 2019

Published: 28 November 2019

Citation:

Babaie-Janvier T and Robinson PA
(2019) Neural Field Theory of
Corticothalamic Attention With Control
System Analysis.
Front. Neurosci. 13:1240.
doi: 10.3389/fnins.2019.01240

1. INTRODUCTION

Attention is implemented in brain by dynamically selecting specific neural activities for further evaluation, in which irrelevant sensory information is filtered out in favor of the behaviorally relevant. It has been proposed that attention can be viewed as enhancing the weight given to relevant information in a competitive system (Noudoost et al., 2010). In theory, this can be done in a number of ways, including strengthening selected signals and reducing noise (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998; Barceló et al., 2000; Noudoost et al., 2010).

Morgan et al. (1996) showed that the magnitude of steady-state visual evoke potentials (SSEVPs) recorded from the scalp of the human subjects who were cued to attend to a stimulus were significantly enlarged. Luck et al. (2000) studied the same phenomenon for single unit cells in macaque monkeys and concluded that when several simultaneous stimuli are present, a neuron's sensory response to the attended signal is significantly enhanced. This amplification is analogous to improving the signal-to-noise ratio which enhances the extraction of pertinent features of the external stimulus (Morgan et al., 1996; Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998). The examples of such sensory gain control mechanisms involved in visual attention, were measured via event-related potentials (ERPs) and cerebral blood flow (positron emission tomography known

as PTE, and functional magnetic resonance imaging known as fMRI) (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998; Barceló et al., 2000; Kastner and Ungerleider, 2001; Vuilleumier et al., 2001; Pinsk et al., 2004). Furthermore, top-down effects are suggested to play an important role in the processing of stimuli in corticothalamic network where they contribute to prediction of forthcoming inputs (Pessoa et al., 2003; Pinsk et al., 2004; Friston, 2005, 2010; Feldman and Friston, 2010). Engel et al. (2001) and Engel and Fries (2010) proposed that making such predictions might involve the temporal structure of both stimulus-driven and background activity, which suggested that prediction and attention are processes interconnected through synchronous oscillations. In psychology, qualitative consciousness models such Integrated Information Theory (Jerath and Beveridge, 2019) and Default Space Theory (Jerath et al., 2015) advocate a central role of corticothalamic system in dynamic and global functions with oscillations in the corticothalamic system amplifying relevant stimuli (neural facilitation of attended signals) based on the current state of internal predictions.

Despite above qualitative proposals regarding the existence of a gain control mechanisms in attention and its influence in corticothalamic neural prediction, the basic mechanisms of such modulation are poorly understood. Examples include the temporal dynamics of activity modulation (i.e., temporal corticothalamic processing), the relationship of modulation to stimulus-driven activity, the relationship of modulations to brain rhythms (i.e., how slow waves or alpha oscillations change during attention), and above all, a plausible neuronal basis by which gain calibration can be implemented. One example of this point is the Bayesian-based models of prediction and attention in the brain, motivated by machine learning applications in neuroscience along with cortical neuroanatomy and neurophysiology (Feldman and Friston, 2010; Friston, 2010), including the Kalman-filter approach used by Rao and Ballard (1999) to explain visual information processing in the brain. This method assumes linear dynamics and that Gaussian distributions describe the activities of populations of neurons, but central steps (such as complex matrix operations) have not yet been shown to be physiologically realizable in the brain's tissues.

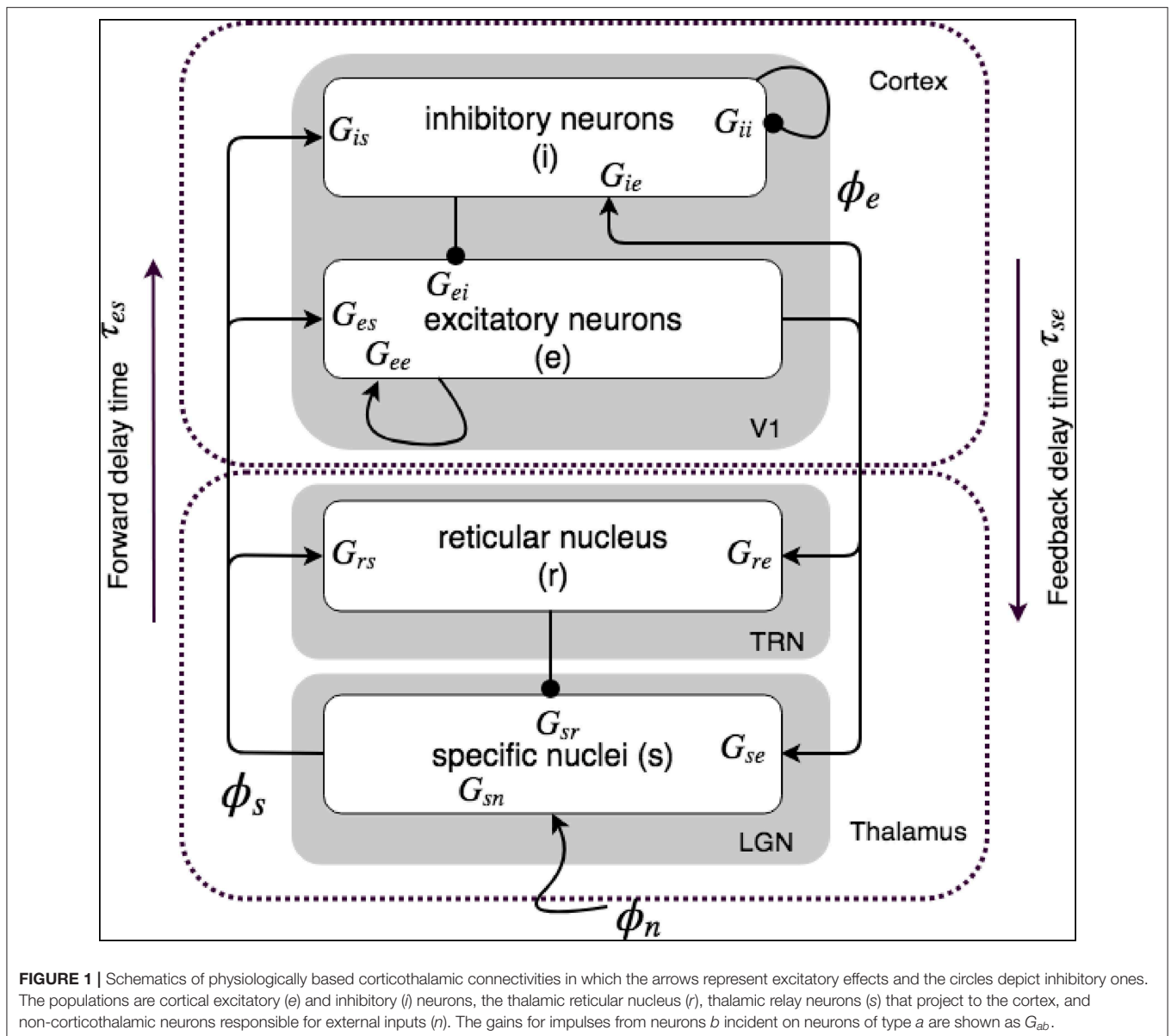
In previous work, neural field theory (NFT) was used to determine the dynamics of corticothalamic responses to visual stimuli and it was shown that their properties are analogous to those widely used in control systems for prediction, gain tuning, and control (Babaie-Janvier and Robinson, 2018). NFT of the corticothalamic system uses the physiological properties of the brain to describe the dynamics of, and interactions between, populations of neurons in the cortex and thalamus, including feedback mechanisms (Robinson et al., 1997, 2002, 2004; Rennie et al., 2002). The equations of the NFT model are non-linear in general and they have been successfully employed to explain highly non-linear phenomena, such as epileptic seizures (Breakspear et al., 2006). However, normal brain dynamics have been shown to be able to be approximated by linear perturbations from spatially uniform steady states (Robinson et al., 1997, 2002, 2004). This approximation enabled a plethora of experimental phenomena, in cohorts of up to 2,100 subjects, to be reproduced, including evoked responses (O'Connor and Robinson, 2004;

Robinson et al., 2005; Kerr et al., 2008; Van Albada et al., 2010; Roberts and Robinson, 2012; Abeysuriya et al., 2015). Neural field models of firing rate activity have been widely used to analyse the dynamics of multi-scale properties of spatio-temporal neural mechanisms (Nunez and Cuttillo, 1995; Jirsa and Haken, 1996; Coombes, 2005). Certainly, linear responses account for a host of phenomena and must be thoroughly understood before proceeding to non-linear cases.

In recent work (Babaie-Janvier and Robinson, 2018) we showed that corticothalamic dynamics are dominated by alpha, beta, and slow-wave resonances, each of which can be considered as a proportional-integral-derivative (PID) filter (Ogata and Yang, 1970) that are widely used in engineering control systems to predict the future course of inputs in their relevant frequency ranges. The best prediction of the stimuli was found to be obtained by summing the filter predictions after weighting them by means of separate gain adjustments (Babaie-Janvier and Robinson, 2018). The brain, therefore, may be implementing attention through control and adjustment of input gains, which suggests that using mismatches between features of internal models and external stimuli to drive gain changes can be analogous to implementing attention in this framework. The objective of these gain adjustments, and the mechanisms by which they may be implemented in the brain is the focus of the current study.

The present work uses NFT of the corticothalamic system, based on the methods developed in (Babaie-Janvier and Robinson, 2018), to incorporate feedbacks that change synaptic gains in response to stimuli, driven by changes in pre- and/or post-synaptic activity (Koch, 1999; Rennie et al., 1999, 2002). Here we consider them as local feedbacks that directly affect the local synaptic strength. These local feedbacks are formulated in a sufficiently general way that a broad range of specific biophysical mechanisms, such as plasticity, long-term potentiation/depression, facilitation, habituation, and sensitization, can be explained by them (Rennie et al., 1999, 2002). Such biophysical effects dynamically alter the synaptic gains and their influences can be represented by dynamical equations for the evolution of the gain parameters. This yields a tractable representation of gain modulations in which the gains are dynamically adjusted as part of the system's stimulus response. We then use control theory to determine when and if gain adjustments can be interpreted as implementing attention to salient information in the stimulus by increasing gains that correspond to relevant input streams. This is done for a simple spatially unstructured visual stimulus to avoid unnecessary complexity, and explore the response of the model to impulse and step stimuli and resulting attentional changes.

This paper is organized as follows. Section 2 outlines the NFT of the corticothalamic system, the resulting transfer functions and their interpretations in terms of resonance filters. A general form of biophysically plausible mechanism for gain modulation is also briefly outlined in this section. The effect of these modulations on dispersion relations are analyzed in section 3 and interpretation of attention by gain adjustment is presented. Finally, section 4 summarizes the main findings and discusses future directions.



2. MATERIALS AND METHODS

2.1. Corticothalamic Model

We first outline our original NFT model, which was developed and extensively discussed in Babaie-Janvier and Robinson (2018), and explain its key parameters. Here we briefly summarize the relevant parts of that work, including essentials of NFT and its application to a corticothalamic model that incorporates key anatomic connectivities in the cortex and the thalamus, and between them. More extensive details is found in Robinson et al. (2002), Robinson et al. (2004), and Babaie-Janvier and Robinson (2018).

Figure 1 shows our corticothalamic model which includes cortical excitatory (*e*) and inhibitory (*i*) neurons, the thalamic reticular nucleus (TRN) (*r*), thalamic relay neurons (*s*), and

non-corticothalamic neurons responsible for external inputs (*n*). In this study, we consider external inputs as visuals, and the relevant relay nucleus is the lateral geniculate nucleus (LGN), where projections are to primary visual cortex (V1). The model incorporates the visual projection system with reciprocal corticothalamic feedback projections, excitatory projections to TRN from LGN-V1 feedforward axons and then V1-LGN feedback axons, and inhibitory projections from TRN onto LGN relay neurons.

NFT finds equations for evolution of dynamical variables of neural populations *a*, which are the local mean cell-body potentials V_a , the mean rate of firing Q_a , and the propagating axonal pulse rate fields ϕ_a , by averaging over short spatial and temporal scales larger than ≈ 1 mm (Wilson and Cowan, 1973; Freeman, 1975). Mean firing rates Q_a are related to mean cell

body potentials $V_a(\mathbf{r}, t)$ through a sigmoid function, denoted by S , that increases from 0 to Q_{\max} as $V_a(\mathbf{r}, t)$ increases from $-\infty$ to ∞ and can be approximated by (Wilson and Cowan, 1973; Freeman, 1975)

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

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

where θ presents the mean threshold voltage and the standard deviation of the distribution of the threshold is given by $\sigma'\pi/\sqrt{3}$. The response of a single neuron is thus smeared out to yield a sigmoid on the population average.

Neurotransmitter at synapses of neurons type a are released by arrival of signals from neurons type b which causes the propagation of post-synaptic voltage changes along dendrites, and soma charging, and results in the spread of temporal profile of the signals. The cell body potential at population a , $V_a(\mathbf{r}, t)$, can thus be written as $\sum_b V_{ab}(\mathbf{r}, t)$, where the subscripts on V_{ab} determines the combinations of afferent neural types, and

$$D_{ab}(t)V_{ab}(\mathbf{r}, t) = \sum_b v_{ab}\phi_b(\mathbf{r}, t - \tau_{ab}), \quad (3)$$

$$D_{ab}(t) = \frac{1}{\alpha_{ab}\beta_{ab}} \frac{d^2}{dt^2} + \left(\frac{1}{\alpha_{ab}} + \frac{1}{\beta_{ab}}\right) \frac{d}{dt} + 1, \quad (4)$$

where the temporal differential operator D_{ab} governs the temporal response of V_{ab} to afferent axonal pulse rate fields ϕ_b , and β_{ab} and α_{ab} describe the rise and fall rate of the voltage at the cell body, respectively. On the right side of Equation (3), v_{ab} presents the overall connection strength between two neural populations, and is given by $v_{ab} = N_{ab}s_{ab}$, where N_{ab} is the average number of synapses on neurons of type a from neurons of type b , and 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 , delayed by τ_{ab} due to discrete anatomical separations between different populations.

In NFT model, a filed $\phi_a(\mathbf{r}, t)$ formed by the short scale average of the neural spikes in population a and propagates at a velocity v_a , obeys a damped wave equation whose source of spikes is $Q_a(\mathbf{r}, t)$, with

$$D_a(\mathbf{r}, t)\phi_a(\mathbf{r}, t) = Q_a(\mathbf{r}, t), \quad (5)$$

$$D_a(\mathbf{r}, t) = \frac{1}{\gamma_a^2} \frac{\partial^2}{\partial t^2} + \frac{2}{\gamma_a} \frac{\partial}{\partial t} + 1 - r_a^2 \nabla^2, \quad (6)$$

where the operator $D_a(\mathbf{r}, t)$ is a good approximation for the damping effect (Jirsa and Haken, 1996; Robinson et al., 1997), and γ_a is the damping rate defined by $\gamma_a = v_a/r_a$, in which r_a is the mean axon length of the population a , and v_a is the pulse velocity. In our model, only the axons of excitatory cortical neurons r_e are considered significantly long and cause propagation effects in Equation (6); in the other populations, we assume $r_a \approx 0$ and therefore $D_a \approx 1$ which leads to $\phi_a(\mathbf{r}, t) = Q_a(\mathbf{r}, t)$ in these populations.

In our model, $v_{ie} = v_{ee}$, $v_{ii} = v_{ei}$, and $v_{is} = v_{es}$ because in cortex, the number of synapses is closely proportional to the

TABLE 1 | Estimated brain parameters for normal adults in the alert, eyes-open state.

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
Firing rate			
$\phi_e^{(0)}$	Steady-state firing rate of e neurons	16	s^{-1}
$\phi_s^{(0)}$	Steady-state firing rate of s neurons	16	s^{-1}
$\phi_r^{(0)}$	Steady-state firing rate of r neurons	16	s^{-1}
$\phi_n^{(0)}$	Steady-state firing rate of n neurons	16	s^{-1}
Sigmoid slope			
ρ_e	For e neurons	4.2×10^3	$V^{-1} s^{-1}$
ρ_s	For s neurons	4.2×10^3	$V^{-1} s^{-1}$
ρ_r	For r neurons	6.3×10^3	$V^{-1} s^{-1}$
Synaptic gain (dimensionless)			
$G_{ee}^{(0)}$	Steady-state synaptic gain from e to e	6.8	—
$G_{se}^{(0)}$	Steady-state synaptic gain from e to s	2.5	—
$G_{ij}^{(0)}$	Steady-state synaptic gain from i to i	8.1	—
$G_{sr}^{(0)}$	Steady-state synaptic gain from r to s	-1.9	—
$G_{es}^{(0)}$	Steady-state synaptic gain from s to e	1.7	—
$G_{sn}^{(0)}$	Steady-state synaptic gain from n to s	0.8	—
$G_{ie}^{(0)}$	Steady-state synaptic gain from e to i	6.8	—
$G_{re}^{(0)}$	Steady-state synaptic gain from e to r	1.0	—
$G_{er}^{(0)}$	Steady-state synaptic gain from i to e	-8.1	—
$G_{rs}^{(0)}$	Steady-state synaptic gain from s to r	0.19	—
$G_{is}^{(0)}$	Steady-state synaptic gain from s to i	1.7	—

Adapted from Table 1 in Robinson et al. (2004), the estimated values are a self-consistent nominal set of parameters and concentrated on mean values.

numbers of source and target neurons (Robinson et al., 1997; Braitenberg and Schüz, 1998), assuming the strength of synapses is determined by the source neurons. With regards to time delays, the forward delays are $\tau_{es} = \tau_{is} \approx 20$ ms corresponding to thalamocortical propagation times and the backward delays are $\tau_{se} = \tau_{re} \approx 60$ ms, which correspond to corticothalamic propagation times, while the remained of τ_{ab} are set to zero. We use a single form of D_{ab} for all the populations, i.e., the same α_{ab} and β_{ab} are assigned to every filed, which corresponds to the approximation that the mean dendritic dynamics can be described by a single pair of time constants. Note that time delays in the long-range excitatory axons in the cortex are included via Equation (6).

Table 1 lists nominal values of model parameters (Robinson et al., 2004). These values were estimated for normal adults and they have been extensively used to generate a plethora of phenomena and numerical analysis within the field, some in cohorts of up to 2,100 subjects (Rennie et al., 2002; O'Connor and Robinson, 2004; Breakspear et al., 2006; Kerr et al., 2008; Abeysuriya et al., 2015).

2.2. Steady States and Transfer Functions

The equations in the NFT of the corticothalamic model are non-linear in general, and helped study highly non-linear phenomena like seizures (Robinson et al., 2002; Breakspear et al., 2006). By setting all derivatives of the NFT equations to zero, small perturbations from spatially uniform steady-states of the system is found which have been found to correspond to normal brain states (Robinson et al., 1997, 2002, 2004; Abeyesuriya et al., 2015). Stable steady-state solutions are interpreted as characterizing the baseline of normal activity, with firing rates that are in accord with experiments (Robinson et al., 2002, 2004). Linear perturbations from these neuronal steady-states represent time dependent brain activity by which numerous experimental phenomena, including evoked responses, have been reproduced (Robinson et al., 1997, 2002, 2004, 2005; O'Connor and Robinson, 2004; Kerr et al., 2008; Van Albada et al., 2010; Roberts and Robinson, 2012; Abeyesuriya et al., 2015).

Low ϕ_a steady-states have been characterized with normal states of brain neuronal activity (Robinson et al., 1997) and only in considerably strong stimulation condition non-linear terms are found in the brain activities (Herrmann, 2001; Roberts and Robinson, 2012; Abeyesuriya et al., 2015). We use $\phi_a^{(1)}$ and $V_a^{(1)}$ to denote linear perturbations of ϕ_a and V_a from their steady-state values $\phi_a^{(0)}$ and $V_a^{(0)}$. Thus, approximately,

$$D_a(\mathbf{r}, t)\phi_a^{(1)}(\mathbf{r}, t) = \rho_a V_a^{(1)}(\mathbf{r}, t), \quad (7)$$

where $\rho_a = dS(V_a^{(1)})/dV_a^{(1)}$, evaluated at the steady state value $V_a^{(0)}$. The stimulus signal ϕ_n also has a steady-state component $\phi_n^{(0)}$ and a time-varying part $\phi_n^{(1)}$ that causes the response. A more detailed study of the model with respect to its parameters can be found in Robinson et al. (2004).

Applying the operator D_a on both sides of Equation (7), plus use of Equation (3), yield

$$D_{ab}(t)D_a(\mathbf{r}, t)\phi_a(\mathbf{r}, t) = \sum_b G_{ab}\phi_b(\mathbf{r}, t - \tau_{ab}), \quad (8)$$

where linear gain $G_{ab} = \rho_a \nu_{ab} = \rho_a N_{ab} s_{ab}$ is the response in population of neurons a due to unit input from neurons b ; i.e., the number of extra spikes out for each additional one in. The net gain of more than two populations of neurons connected serially is denoted by $G_{abc} = G_{ab}G_{bc}$.

2.2.1. Transfer Functions

A linear transfer function is the ratio of the output of a linear system to its input. To obtain transfer functions one can apply either the Laplace or Fourier transform to both sides of Equation (8) for every population in the corticothalamic system and transform it from time domain t to frequency domain Ω or s . We use single-sided Laplace transform (Ogata and Yang, 1970)

$$\mathcal{L}[f(t)](s) = f(s) = \int_0^\infty f(t)e^{-st} dt, \quad (9)$$

where the response est is parameterized by the complex frequency $s = -i\omega = \Gamma - i\Omega$. We define the inverse of Laplace transformed operator in Equation (3) by

$$L_{ab}(s) = D_{ab}^{-1}(s) = \left(1 + \frac{s}{\alpha_{ab}}\right)^{-1} \left(1 + \frac{s}{\beta_{ab}}\right)^{-1}, \quad (10)$$

where the serial filters of connected populations are $L_{abc} = L_{ab}L_{bc}$. The Laplace transform of damping operator in Equation (6) is

$$D_a(\mathbf{k}, s) = \left(1 + \frac{s}{\gamma_a}\right)^2 + k^2 r_a^2, \quad (11)$$

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

In the Laplace domain, the transfer functions to thalamus, TRN, and cortex from retina, respectively, are given by Babaie-Janvier and Robinson (2018)

$$T_{sn}(\mathbf{k}, s) = \frac{\phi_s^{(1)}(\mathbf{k}, s)}{\phi_n^{(1)}(\mathbf{k}, s)} = \frac{G_{sn}L_{sn}\mathcal{M}_c(s)}{\mathcal{M}_c(s)\mathcal{P}_t(s) - \mathcal{P}_c(s)}, \quad (12)$$

$$T_{rn}(\mathbf{k}, s) = \frac{\phi_r^{(1)}(\mathbf{k}, s)}{\phi_n^{(1)}(\mathbf{k}, s)} = \frac{G_{sn}L_{sn}\mathcal{P}_r(s)}{\mathcal{M}_c(s)\mathcal{P}_t(s) - \mathcal{P}_c(s)}, \quad (13)$$

$$T_{en}(\mathbf{k}, s) = \frac{\phi_e^{(1)}(\mathbf{k}, s)}{\phi_n^{(1)}(\mathbf{k}, s)} = \frac{G_{sn}L_{sn}\mathcal{N}_s(s)}{\mathcal{M}_c(s)\mathcal{P}_t(s) - \mathcal{P}_c(s)}, \quad (14)$$

with

$$\mathcal{M}_c(s) = D_{ee}(1 - G_{ei}L_{ii}) - G_{ee}L_{ee}, \quad (15)$$

$$\mathcal{P}_t(s) = 1 - G_{srs}L_{srs}, \quad (16)$$

$$\mathcal{P}_c(s) = (G_{ese}L_{ese} + G_{esre}L_{esre}) \exp[-s(\tau_{es} + \tau_{es})], \quad (17)$$

$$\mathcal{N}_s(s) = G_{es}L_{es} \exp(-s\tau_{es}), \quad (18)$$

$$\mathcal{P}_r(s) = G_{res}L_{res} \exp[-s(\tau_{es} + \tau_{se})] + G_{rs}L_{rs}\mathcal{M}_c(s). \quad (19)$$

More details of the calculation is found in Babaie-Janvier and Robinson (2018).

Here, we calculated the transfer functions that relate corticothalamic activities to stimuli. These allow us to investigate the linear response of each population to any external input. One key feature of the transfer functions is the denominator which represents the characteristic dispersion equation of the system. Setting this equation to zero yields system's eigenvalues and mark the poles which determine the basic modes into which the system response can be decomposed. Furthermore, all transfer functions have their poles in common, as seen in Equations (12, 13), as a result of the interconnectedness of the system.

2.3. Corticothalamic Data Filters

The transfer functions can be decomposed into basic modes whose behaviors are shown to be associated with well-known data filters (Babaie-Janvier and Robinson, 2018). Here, we briefly outline the resulting data filters obtained for corticothalamic populations; the details are found in Babaie-Janvier and

TABLE 2 | Characteristic parameters estimated for slow, alpha, and beta filters of corticothalamic transfer functions T_{an} , for $a = s, r, e$, using parameters for normal healthy adults in alert, eyes-open state, adapted from Table 3 in Babaie-Janvier and Robinson (2018).

Quantity	Description	T_{en}			T_{rn}			T_{sn}			Unit
		T_{en}^{ℓ}	T_{en}^A	T_{en}^B	T_{rn}^{ℓ}	T_{rn}^A	T_{rn}^B	T_{sn}^{ℓ}	T_{sn}^A	T_{sn}^B	
$-\Gamma^-$	Damping rate 1st pole	9.3	14.1	26.9	10.2	13.5	27	15.3	23.8	22.3	–
$-\Gamma^+$	Damping rate 2nd pole	17.2	14.1	26.9	14.5	13.5	27	337	23.8	22.3	–
Ω_c	Characteristic frequency	0	57.4	143	0	56.2	101	0	57.9	84	s ⁻¹
$ K $	Gain	1.8	3.82	1.62	7.44	4.3	8.3	52.2	48.8	4.6	–
τ_p	Prediction time	16	28	6	73	7	15	18	23	2	ms

Robinson (2018). Note that only the spatially-uniform effects of perturbations (i.e., $\mathbf{k} = 0$), were explored in the study.

If we approximate each corticothalamic transfer function $T_{ab}(s)$, which is a ratio of exponential polynomials of s , by a rational function of s and decompose it into partial fractions, we find

$$T_{ab}(s) = \sum_{j=1}^n \frac{r_j}{s + p_j}, \tag{20}$$

where, the $p_j = \Gamma_j \pm i\Omega_j$ are all distinct poles of the system (we do not consider degenerate roots here), and the residues $r_j = r \pm i\Omega_r$ are given by

$$r_j = (s + p_j)T_{ab}(s)|_{s=-p_j}, \tag{21}$$

we estimate the reduced transfer functions T_{ab}^r by seeking the smallest n that preserves the main dynamics. Babaie-Janvier and Robinson (2018) showed that a 6-pole approximation ($n = 6$) is accurate to within an root-mean-square (rms) fractional error of 0.02 over the frequency range 0–150 Hz for the parameters in **Table 1**. These partial fractions then are summed in pairs that dominate in slow ($f \lesssim 5$ Hz), alpha ($5 \text{ Hz} \lesssim f \lesssim 15$ Hz) and beta ($15 \text{ Hz} \lesssim f$) frequency regimes, respectively. We thus write (for $b = s, r, e$)

$$T_{bn}^r(s) = T_{bn}^{\ell}(s) + T_{bn}^A + T_{bn}^B(s), \tag{22}$$

where T_{bn}^{ℓ} is the sum of the two real poles and represent low frequency responses while T_{bn}^A and T_{bn}^B are the sums over complex conjugate pairs of poles that represent oscillatory responses in the alpha and beta frequency ranges, respectively. 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 $T_{ab}^J(s)$ for $J = \ell, A, B$, with

$$T_{ab}^J(s) = (s + \tau_p^{-1}) \left[\frac{K}{(s + p_j)(s + p_{j+1})} \right], \tag{23}$$

$$= H_{ba}^J(s)I_a^J(s) \tag{24}$$

respectively, with $\tau_p = (r_j + r_{j+1}) / (r_j p_{j+1} + r_{j+1} p_j)$ and $K = r_j + r_{j+1}$. If we write $r_j = r \pm i\Omega_r$ for the residues at a conjugate pair of poles $p_j = \Gamma \pm i\Omega_j$, we have

$$\tau_p = \frac{r}{-r\Gamma + \Omega_r \Omega_j}, \tag{25}$$

$$K = 2r. \tag{26}$$

Similarly, writing $r_j = r, r'$ for the residues at real poles $p_j = \Gamma, \Gamma'$, we have

$$\tau_p = -\frac{r + r'}{r\Gamma' + r'\Gamma}, \tag{27}$$

$$K = r + r'. \tag{28}$$

Table 2 presents the parameters calculated for three transfer functions to specific nuclei, TRN, and cortex from retina. Each slow, alpha, and beta filter captures and evaluates part of the information coming from external world, which is relevant to their frequency limit, and summing up these parallel responses results in the total response to stimulus. Each population’s neural response contains an alpha resonance (7.5–12 Hz for present, adult human parameters), with peak frequencies ranging from ≈ 8.4 to ≈ 9.3 Hz, and a beta response in the beta band (12.5–30 Hz) with amplitudes significantly smaller than those of the alpha resonances (Babaie-Janvier and Robinson, 2018). Both filters present higher peak amplitudes in thalamus than cortex. Furthermore, the alpha waves persist longer than beta waves in populations r, e , and i , which is the direct result of the damping rates of alpha filters are approximately half of those of the beta filters. However, the results show that alpha and beta waves should last for approximately the same time in LGN structure where their damping rates are similar (Babaie-Janvier and Robinson, 2018).

2.4. Synaptic Gain Modulation

The NFT model described in the previous section treated all gains that encode physiology of coupling strengths as being fixed. However, numerous biophysical processes modulate neuronal gains, dependent on current or recent level of activity, including plasticity, long-term potentiation/depression, facilitation, habituation, and sensitization. Each is considered here as a form of feedback, whereby pre- or post-synaptic neuronal activity modulates model parameters that were previously considered constant. These time-dependent effects can be included in the NFT of corticothalamic model by introducing dynamical equations for evolution of the gains; gain changes are caused by local firing rates or voltages.

We choose a general form of modulatory response that can be applied to a wide variety of specific mechanisms (Koch, 1999; Rennie et al., 1999; Robinson et al., 2002), with

$$G_{ab}(\mathbf{r}, t) = G_{ab}^{(0)} + G_{ab}^{(1)}, \quad (29)$$

$$= G_{ab}^{(0)} + g_{ab}F_{ab}(t) \otimes [\phi_b(\mathbf{r}, t) - \phi_b^{(0)}], \quad (30)$$

where $G_{ab}^{(0)}$ is the steady-state value and $G_{ab}^{(1)}$ represents the perturbation of gain caused by the feedback, where $F_{ab}(t)$ describes the temporal form of the modulation and the constant g_{ab} is its strength. The formulation in Equation (30) assumes that: (i) the perturbations are small enough that a linear equation is a reasonable approximation, (ii) modulation is local in space, and (iii) g_{ab} and $F_{ab}(t)$ do not vary with position or time. An alternative possible set of drives are the cell body potentials V_a (Rennie et al., 1999, 2002), but these are also linearly related to mean firing rates Q_a for small amplitude perturbations, so we do not consider them separately here. For the temporal form of the modulation we propose $F_{ab}(t) = \eta_{ab} \exp(-\eta_{ab}t)$ when $t \geq 0$ and zero otherwise to enforce causality. The rate constant η_{ba} characterizes the response process and $\eta_{ab} > 0$. This form of $F_{ab}(t)$ implies a simple differential equation form of Equation (30):

$$\left(\frac{1}{\eta_{ab}} \frac{d}{dt} + 1 \right) [G_{ab}(\mathbf{r}, t) - G_{ab}^{(0)}] = g_{ab} [\phi_b(\mathbf{r}, t) - \phi_b^{(0)}]. \quad (31)$$

The Laplace transform of Equation (31) is

$$G_{ab}(\mathbf{k}, s) = G_{ab}^{(0)} + g_{ab}F_{ab}(s)\phi_a^{(1)}(\mathbf{k}, s), \quad (32)$$

where

$$F(s) = \frac{1}{\tau_{ab}^F s + 1}, \quad (33)$$

$$\tau_{ab}^F = \frac{1}{\eta_{ab}}. \quad (34)$$

The formulation developed above can be applied to different biophysical processes that change synaptic gains. For instance, Equation (32) can represent habituation if appropriate local feedback strengths g_{ab} are adapted so that a decrement in the neural activity is caused in response to stimuli sustained over several hundred milliseconds. On the other hand, if the local feedback strengths g_{ab} are tuned to increase the neuronal response due to current activity it can describe neuronal sensitisation. We further discuss the effect of feedbacks on local gains and how they modulate the neural responses in the Results section.

3. RESULTS

In this section we work on deriving a general neural activity formulation which involves gain modulations via local feedbacks. The resulted activity form is then interpreted in terms of attention to more salient input. Finally, we carry out simulation

using impulse and step stimuli to evaluate the corticothalamic response to various change in the information coming from external world.

3.1. Effect of Gain Modulation on Neural Activity

In a linear regime, the cell body potentials are related to mean firing rate through Equation (2) which gives rise to field pulses, obeying the damped wave equation in Equation (5). This results in the general firing rate equation, to first order in deviations of the variables from their steady-state values,

$$\begin{aligned} D_{ab}(t)\mathcal{D}_{ab}(t) [\phi_a^{(0)} + \phi_a^{(1)}(\mathbf{r}, t)] \\ = \sum_b G_{ab}(\mathbf{r}, t) [\phi_b^{(0)} + \phi_b^{(1)}(\mathbf{r}, t - \tau_{ab})]. \end{aligned} \quad (35)$$

The Laplace transform of Equation (35) is then

$$\begin{aligned} \mathcal{D}_{ab}(s) [\phi_a^{(0)} + \phi_a^{(1)}(\mathbf{k}, s)] = L(s) \sum_b [G_{ab}^{(0)} + g_{ab}F(s)\phi_b^{(1)}(\mathbf{k}, s)] \\ [\phi_b^{(0)} + \phi_b^{(1)}(\mathbf{k}, s) \exp(-s\tau_{ab})]. \end{aligned} \quad (36)$$

By equating orders on both left and right side of Equation (36), and omitting second order term, we find

$$\phi_a^{(0)} = \sum_b G_{ab}^{(0)}\phi_b^{(0)}, \quad (37)$$

$$\begin{aligned} \mathcal{D}_{ab}(s)\phi_a^{(1)}(\mathbf{k}, s) = L(s) \sum_b G_{ab}^{(0)}\phi_b^{(1)}(\mathbf{k}, s) \exp(-s\tau_{ab}) \\ + L(s) \sum_b g_{ab}\phi_b^{(0)}F(s)\phi_b^{(1)}(\mathbf{k}, s) \exp(-s\tau_{ab}), \end{aligned} \quad (38)$$

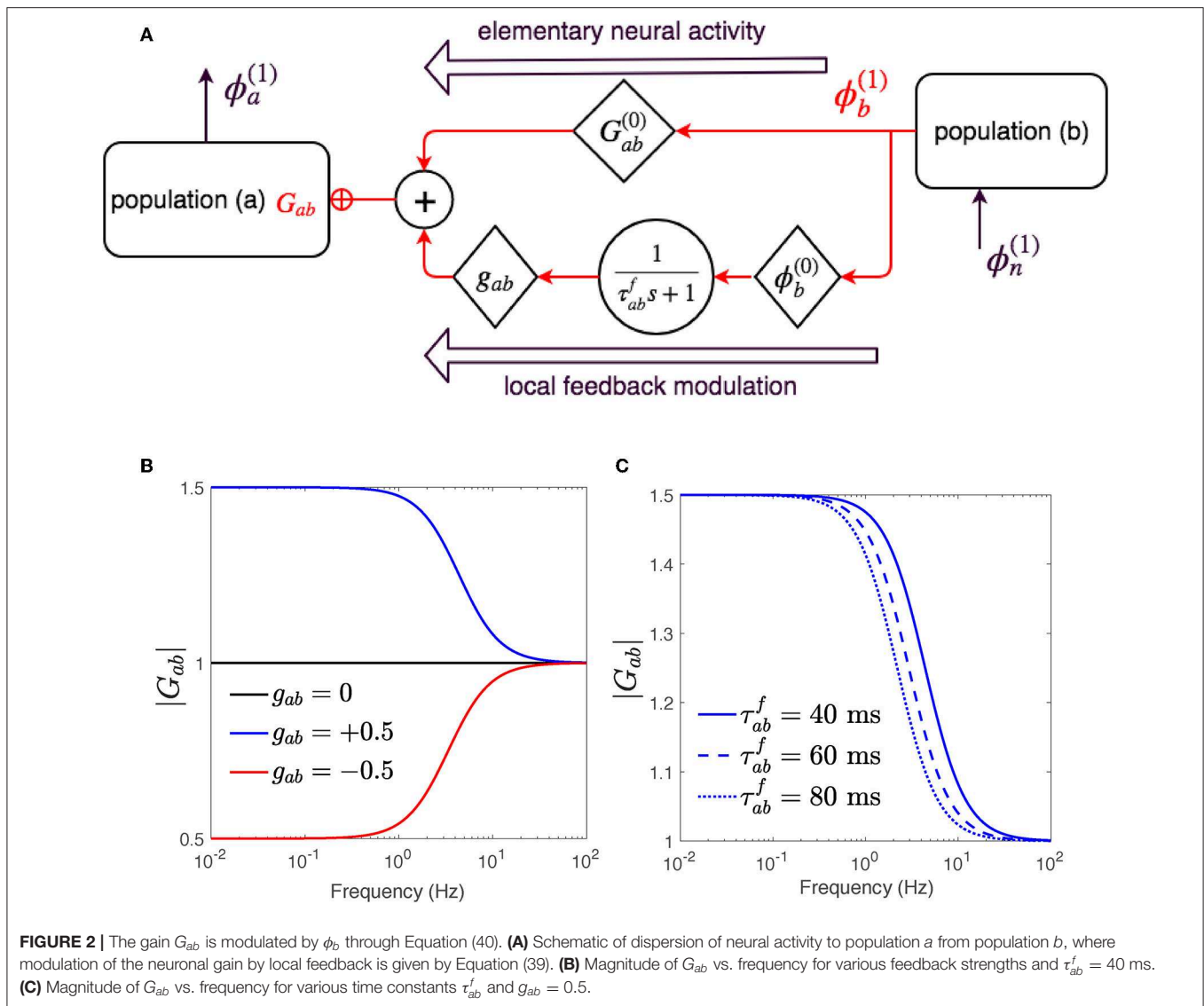
$$= L(s) \sum_b [G_{ab}^{(0)} + g_{ab}\phi_b^{(0)}F(s)] \phi_b^{(1)}(\mathbf{k}, s) \exp(-s\tau_{ab}). \quad (39)$$

Equation (37) represents the steady-state relations. Equation (38) expresses first order responses of two types: the first term represents the part of response that would occur without change to the steady-state synaptic gains, whereas the second term represents the response due to the effect of stimulus induced gain changes on steady-state activity. **Figure 2** presents a schematic of the gain modulation mechanism achieved in Equation (39).

The linear dispersion Equation (39) differs from Equation (9) only in their gain terms G_{ab} . While the gains in Equation (9) comprise only fixed synaptic effect terms; i.e., $G_{ab} = G_{ab}^{(0)}$, the synaptic coefficients in the new dispersion Equation (39) contain an added modulation term $g_{ab}\phi_b^{(0)}F_{ab}(s)$,

$$G_{ab} = G_{ab}^{(0)} + g_{ab}\phi_b^{(0)}F(s), \quad (40)$$

which results in transfer functions of corticothalamic system that represent adjustment of gains as part of the response. These new transfer functions are thus derived by substituting the term G_{ab}



for $a, b = n, s, r, e$ in Equations (18, 19) using their equivalent modulated form in Equation (40). Studying the response of the new transfer functions with two modulation terms: (i) feedback strength g_{ab} , and (ii) feedback time constant η_{ab} , requires these terms to be assigned to appropriate values. To determine a physiologically applicable range for these terms, we first focus on the process where gain is modulated by a local signal $\phi_b^{(1)}$, and explore how reduction or escalation of feedback strength g_{ab} and/or time evolution constant η_{ab} affect the gain. The applicable estimated values through these examinations can be tested by experiments, in future works.

To establish a characteristic scale of gain changes, we note that if a permanent change in $\phi_b^{(0)}$ occurs, the gain G_{ab} will eventually settle to a new steady-state $G_{ab}^{(new)}$, with the time evolution described by $F_{ab}(s)$,

$$G_{ab}^{(new)} = G_{ab}(s)|_{s \rightarrow 0}, \tag{41}$$

$$= G_{ab}^{(0)} + g_{ab}\phi_b^{(0)}. \tag{42}$$

Thus the new steady-state gain depends on the magnitude and sign of g_{ab} . We define a relative coefficient of modulation, denoted by Δ , which is the magnitude of the fractional change in $G_{ab}^{(new)}$ relative to $G_{ab}^{(0)}$

$$\Delta = \frac{G_{ab}^{(new)}}{G_{ab}^{(0)}} - 1, \tag{43}$$

to investigate a variety of changes. We then choose various Δ in Equation (42) and calculate corresponding values of g_{ab} needed to produce them in the steady state. **Figure 2B** shows the magnitude of gains G_{ab} vs. frequency as in Equation (40) for $G_{ab}^{(0)} = 1$, time constant $\eta^{-1} = 40$ ms and two positive and negative feedback strengths $g_{ab} = \pm 0.5$. As $\phi_b^{(1)}$ is elicited, the gain G_{ab} begins either increasing, for $g_{ab} = +0.5$, or

TABLE 3 | Gain modulation allocated to g_{ab} for various level Δ of increased or decreased gain modulation, using parameters in **Table 1** and $\eta_{ab}^{-1} = 40$ ms.

Quantity	Description	$\Delta = -0.5$	$\Delta = -0.2$	$\Delta = +0.2$	$\Delta = +0.5$	Unit								
Modulation gains														
g_{sn}	Feedback strength	-0.025	-0.010	+0.010	+0.025	-								
g_{sr}	Feedback strength	+0.039	+0.015	-0.015	-0.039	-								
g_{se}	Feedback strength	-0.078	-0.031	+0.031	+0.078	-								
		T_{sn}^{ℓ}	T_{sn}^A	T_{sn}^B	T_{sn}^{ℓ}	T_{sn}^A	T_{sn}^B	T_{sn}^{ℓ}	T_{sn}^A	T_{sn}^B	T_{sn}^{ℓ}	T_{sn}^A	T_{sn}^B	
Filter parameters														
$-\Gamma^-$	Damping coefficient	137	11.1	12.1	26.1	24.2	19.18	11.5	21.4	24.8	8.7	17.7	27.2	-
$-\Gamma^+$	Damping coefficient	259	11.1	12.1	334	24.2	19.18	338	21.4	24.8	339	17.7	27.2	-
Ω_c	Cut-off frequency	0	55.7	108	0	54	87.5	0	59	83	0	58.3	83.6	s ⁻¹
$ K $	Gain	11.5	363	3.9	45.9	44.8	2.21	51.9	39.1	14.1	49.9	25.7	25.6	-
τ_p	Prediction time	1	363	40	9	61.3	2	19.7	16	7	18.5	12.4	23	ms

Parameters obtained for slow, alpha, and beta filters of corticothalamic transfer functions T_{an} using the filter identification method developed in section 3.

decreasing for $g_{ab} = -0.5$ and within 40 ms reaches its new steady-state value $G_{ab}^{new} = +1.5$ or $G_{ab}^{new} = -1.5$, respectively. **Figure 2C** shows the magnitude of gain G_{ab} vs. frequency as in Equation (40) for $G_{ab}^{(0)} = 1$, $g_{ab} = 0.5$, and various time evolving constant η^{-1} . The smaller the η_{ab} the faster G_{ab} settles to G_{ab}^{new} .

In this study we select the transfer function to specific nuclei from retina, to investigate the effect of gain modulations. The same course can be used for studying other populations. We select a mild modulation $\Delta = \pm 0.2$ and a significant one $\Delta = \pm 0.5$ to cover a significant range of both increased and decreased modulation effects. **Table 3** presents the calculated feedback strengths g_{sn} , g_{sr} , and g_{se} (these are gains involved in transfer function T_{sn}), that cause these amount of modulation. It should be noted that the amount of modulation, Δ , is not necessarily equal for various neuronal populations during any particular gain adjustment process, but for simplicity, we assign the same Δ in the present work. A reasonable choice for the time constant η_{ab}^{-1} was made at 40 ms (Rennie et al., 1999; Robinson et al., 2002). **Figure 3A** shows the magnitude of transfer function T_{sn} vs. frequency for various values of involved local feedback gains as presented in **Table 3**, and using **Table 1** for the rest of parameters. **Figures 3B–D** show slow, alpha, and beta filters calculated for these transfer functions. We use the Control System Toolbox of Matlab 2018a to carry out the calculations.

3.2. Interpreting Attention as Gain Adjustment

We have incorporated gain modulation into the corticothalamic system and showed that systems' transfer functions, and consequently its dynamics, are influenced by local feedbacks that adjust the local gains. Here we discuss how this gain adjustment can be interpreted as implementing attention in the corticothalamic system. We can write

$$\phi_a^{(1)} = \phi_a^{\ell} + \phi_a^A + \phi_a^B, \tag{44}$$

$$= (T_{an}^{\ell} + T_{an}^A + T_{an}^B) \phi_n^{(1)}. \tag{45}$$

Considering the part of response associated with the filter ϕ_a^J for J of ℓ , A , and B , in Equation (45), and replacing the filter's transfer function from Equation (23), we have

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

$$= K (s + \tau_p^{-1}) I_a^J(s) \phi_n^{(1)}(s), \tag{47}$$

$$= k_1 s \phi_I^J(s) + k_0 \phi_I^J(s), \tag{48}$$

where

$$\phi_I^J(s) = I^J(s) \phi_n^{(1)}, \tag{49}$$

$$k_0 = K \tau_p^{-1}, \tag{50}$$

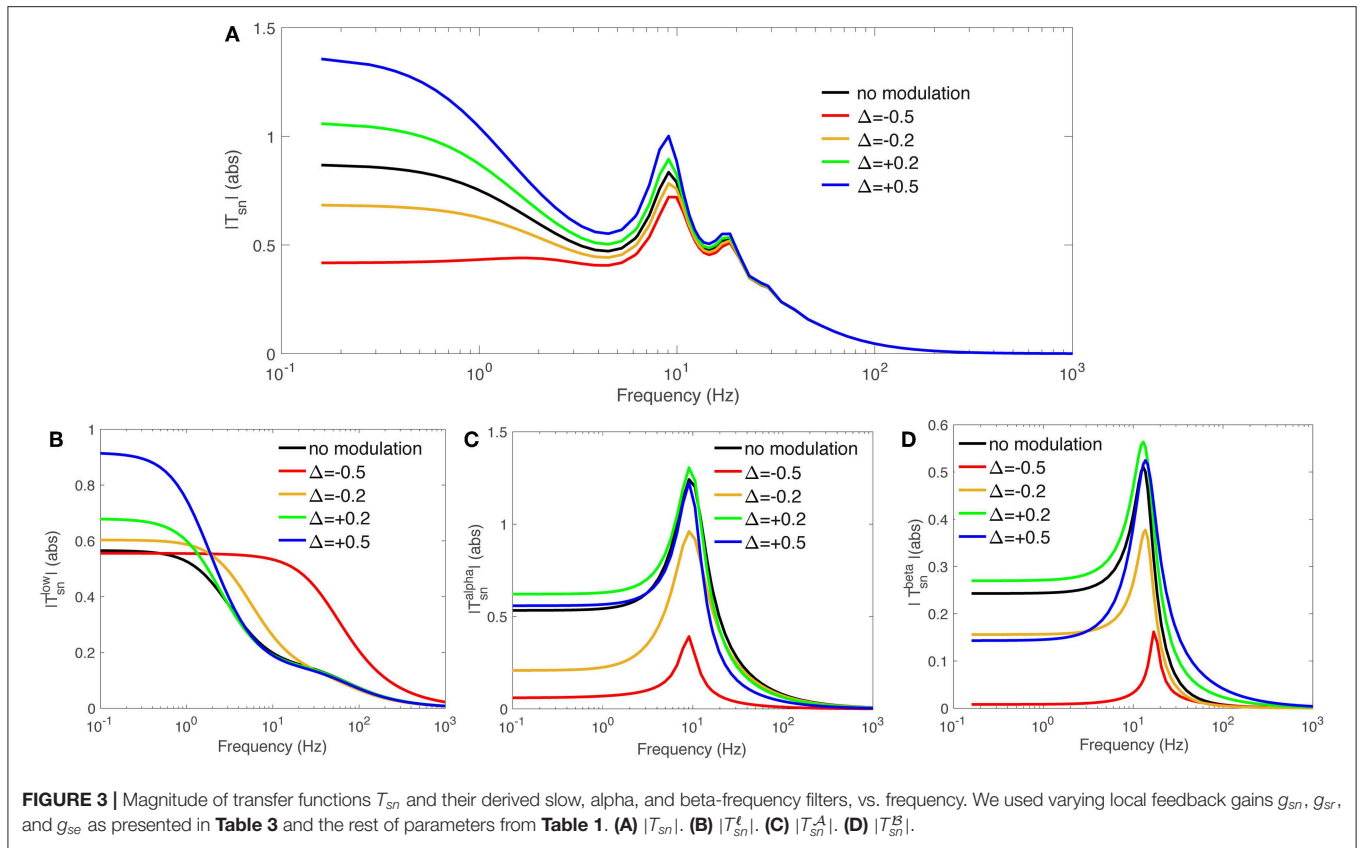
$$k_1 = K. \tag{51}$$

The term $I^J(s)$ is a second-order low-pass convolution filter whose function is to allow external stimuli of a given band of frequencies to pass while attenuating or weakening all others that are not favored. In Babaie-Janvier and Robinson (2018), it was shown that $I^{\ell}(s)$ is a low-pass filter while $I^A(s)$ and $I^B(s)$ are resonance filters exhibiting damped oscillations at alpha and beta rhythms, respectively.

Equation (48) shows that the convolved signal $\phi_a^J(s)$ is then modulated by two coefficients k_0 and $k_1 s$, independently, where each generates a partial response. The inverse Laplace transform of Equation (48) yields

$$\phi_b^J(t) = K \left[\frac{1}{\tau_p} + \frac{d}{dt} \right] \phi_I^J(t), \tag{52}$$

which shows the partial responses are the intensity and its rate of change, weighted by $k_0 = K \tau_p^{-1}$ and $k_1 = K$, respectively. The coefficients k_0 and k_1 determine which parts of the incoming information should be emphasized and which parts can be discarded by actively weighting the features of the stimulus, i.e.,



its value and rate of change. The modulation signal caused by filter gains k_0 and k_1 , are given by

$$K_0^J(t) = \mathcal{M} \left\{ k_0^J \phi_I^J(t) \right\}, \tag{53}$$

$$K_1^J(t) = \mathcal{M} \left\{ k_1^J \left[\frac{d}{dt} \phi_I^J(t) \right] \right\}, \tag{54}$$

where \mathcal{M} is defined as the envelop of the signals. This envelope generalizes the concept of a constant amplitude obtained for each data filter. Therefore, $K_0^J(t)$ and $K_1^J(t)$ are the modulation signals of the magnitude and the rate of change components.

Figure 4 shows the filter parameters τ_p and K and corresponding k_0 and k_1 of slow, alpha, and beta filter, calculated for various level of synaptic gain changes of $-0.5 \leq \Delta \leq +0.5$, using parameters in **Table 3**. The results show that high-frequency gains; i.e., the beta filter's k_0 and k_1 , show inverse behavior to those of the slow and alpha filters for most Δ . For example, k_1 is reduced for slow and alpha filters when Δ is increased from 0 to 0.5, whereas this weight increases for the beta filter; however, the k_0 values show the opposite behavior, increasing slightly for the alpha filter and significantly for the slow filter, while decreasing for the beta filter. This confirms that by increasing the synaptic gains, the emphasis on the high frequencies is increased. The effect of these changes are explored in more detail in the next section.

The sum of slow, alpha, and beta-range responses constitute the total response to the stimulus, with

$$\phi_a^{(1)} = \phi_a^{\ell} + \phi_a^A + \phi_a^B, \tag{55}$$

$$= \left(k_0^{\ell} I_a^{\ell}(s) + k_0^A I_a^A(s) + k_0^B I_a^B(s) \right) \phi_n^{(1)} \tag{56}$$

$$+ \left(k_1^{\ell} I_a^{\ell}(s) + k_1^A I_a^A(s) + k_1^B I_a^B(s) \right) \phi_n^{(1)}. \tag{57}$$

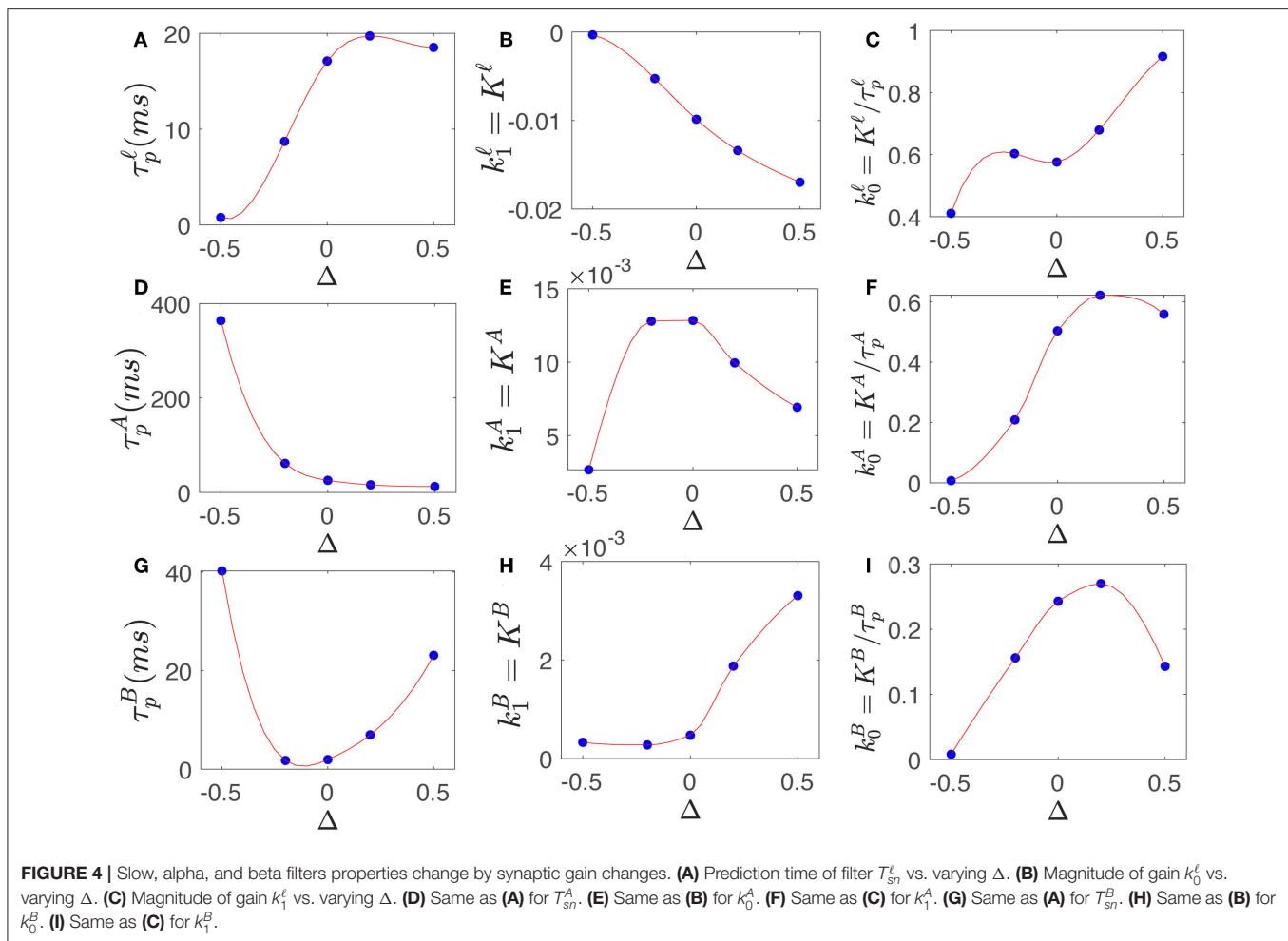
The inverse Laplace transform of Equations (56, 57) yields

$$\phi_a^{(1)}(t) = \sum_J K^J \left[\frac{1}{\tau_p^J} + \frac{d}{dt} \right] \phi_I^J(t). \tag{58}$$

Equation (58) shows the total response is a superposition of all partial responses each scaled by a particular gain. Once a perturbation, i.e., $\phi_n^{(1)}$, occurs in sensory input the rate of change, i.e., $\frac{d}{dt} \phi_I^J(t)$, is significant and therefore a large part of the total response is induced by this change, whereas when the rate of change declines, i.e., $\frac{d}{dt} \phi_I^J(t)$ approaches zero, the total response is largely proportional to the magnitude of sensory input.

3.3. Impulse and Step Responses

Aside from the issue of how to interpret corticothalamic dynamics in terms of gain adjustment *per se*, there is the central question of how well these adjustments enable the system to pay attention to more important information in its input signals



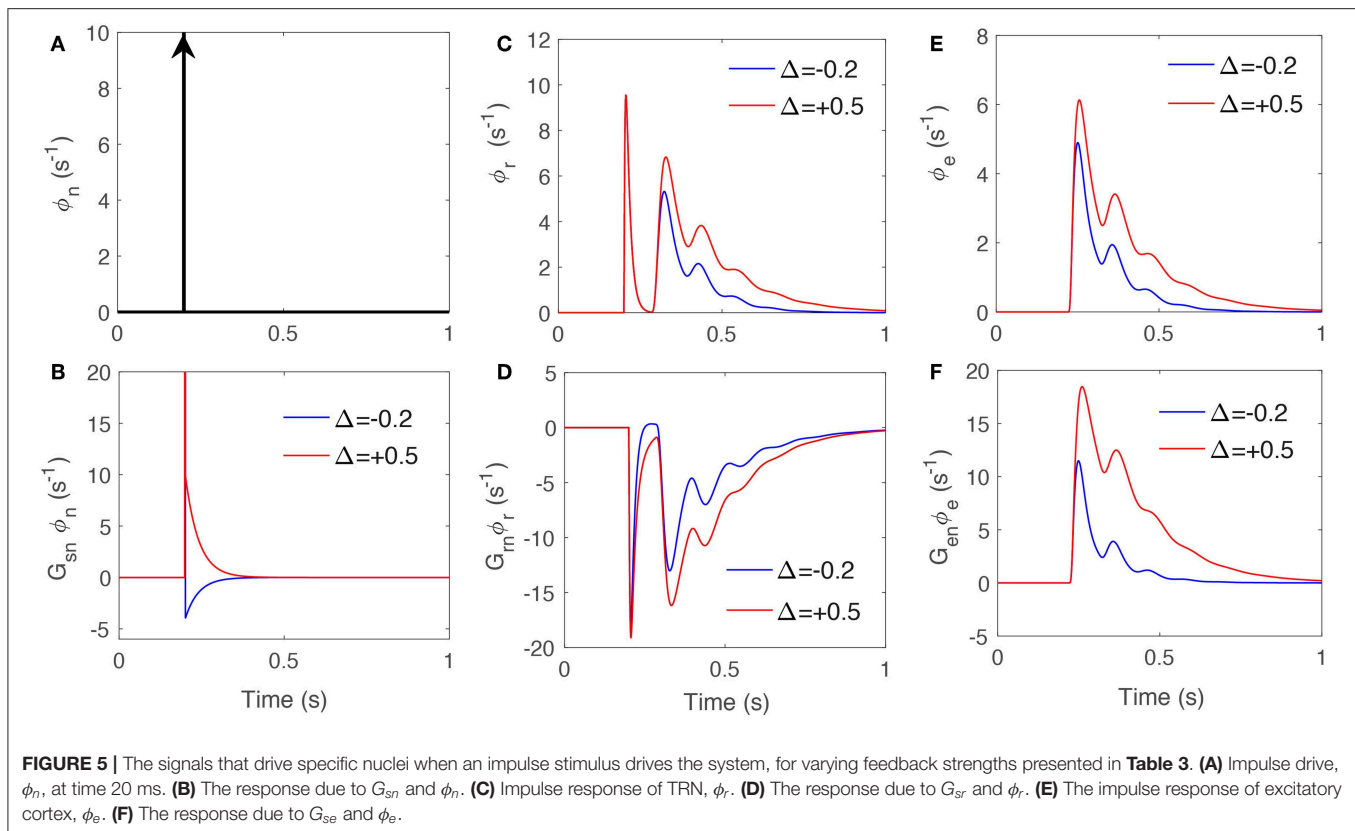
by increasing or decreasing the gains in the relevant frequency ranges. To investigate the time varying effect of local feedback in adjusting gains, we stimulate the system separately with impulse and step stimuli and track the specific nuclei's slow, alpha, and beta-frequency range responses and consequently, total firing ϕ_{sn} . In the visual system, the impulse response is neural activity provoked by a sudden change in either the intensity (magnitude) or the rate of change (frequency) of the external stimulus that only lasts for a minimal time. It therefore corresponds to Evoked Related Potentials. A step response is neural activity stimulated by a sudden change in visual field that stays in the new level; therefore, it enables the study of steady state properties of the error signal between the external stimulus and internal model's prediction. We use varying local feedback strengths from **Table 3**, the system parameters from **Table 1**, and the Control System Toolbox Matlab 2018a to carry out the calculations.

3.3.1. Impulse Response

We first drive the system with a unit impulse signal at time 200 ms. **Figures 5A–F** present the reticular and excitatory cortex responses when the corticothalamic system is stimulated by an impulse stimulus at time 20 ms and the signals projected

in LGN through G_{sn} , G_{sr} , and G_{se} . These show varying local feedback gains cause changes in both magnitude and latency of the evoked responses. In particular, the feedback signal from excitatory cortex shows more sensitivity to varying gains; with 50% increase in G_{se} causing the feedback signal ϕ_e to last 40 ms longer. Furthermore, the first peak is $\sim 50\%$ enlarged while the second peak is approximately tripled. This data supports theories in both physiological and behavioral studies (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998), which suggested that the amplitudes of stimulus-driven neural activity in sensory pathways are adjusted when tasks require attention. Gazzaley et al. (2005) provided evidence that attention not only modulates the magnitude of neural activity but also shortens the time to reach maximal response, which suggested that attention modifies the speed of neural processing.

In order to uncover the effect of local gain modulation, we compare the response for the fixed-gain T_{sn} with the response for the modulated-gain transfer function to a sudden change, an impulse signal. We choose the parameters for the fixed gain transfer function from **Table 2** and parameters of $\Delta = 0.5$ for the modulated transfer function from **Table 3**. We stimulate both systems with a unit impulse signal and in each case, we calculate



the contribution of the individual slow, alpha, and beta filters and the total response ϕ_s . **Figure 6** shows the result in which the black curve in every plot corresponds to the original filters without gain modulation and the red curve corresponds to the modulated filters.

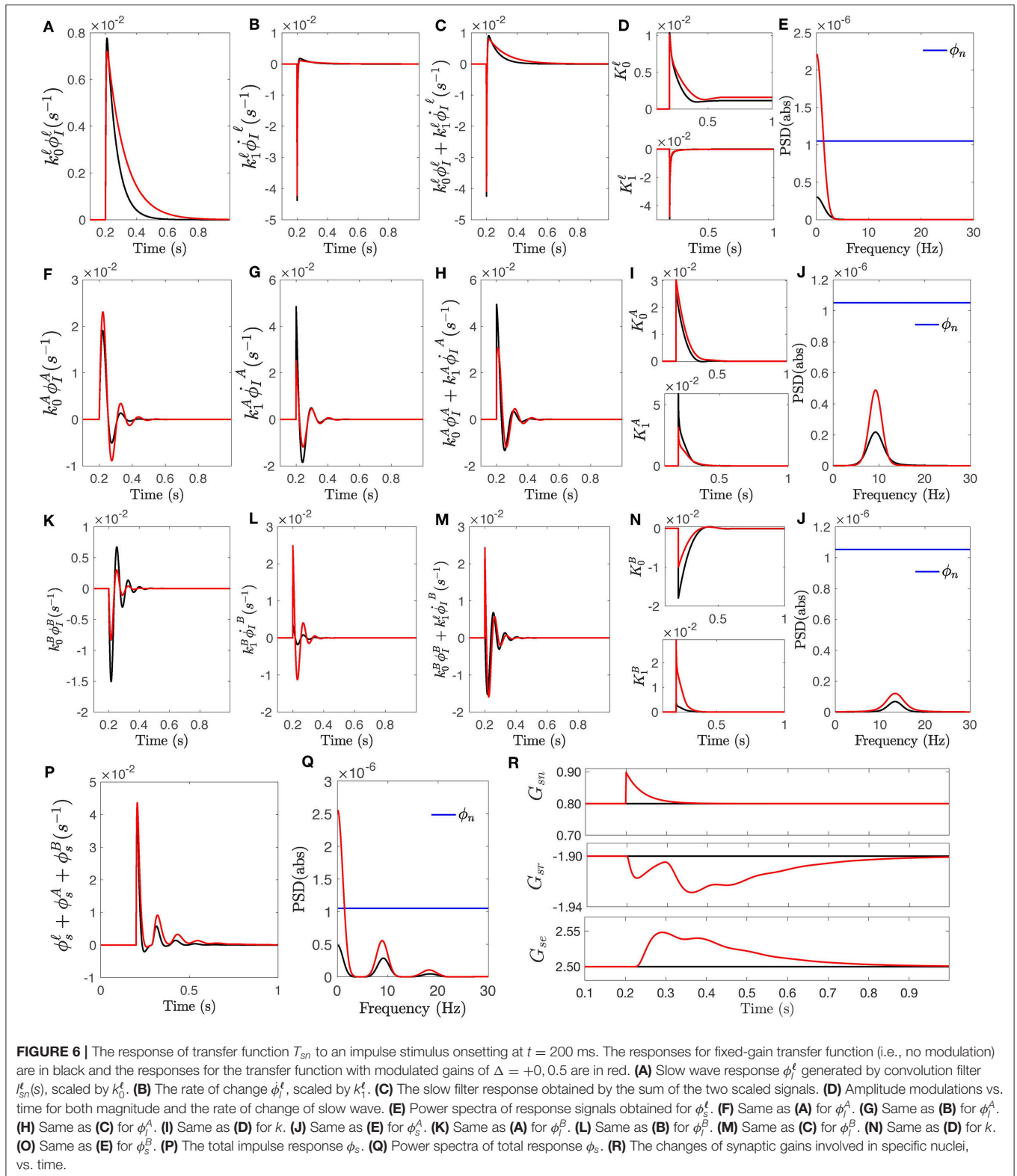
Once the stimulus onsets, second order convolution filters I^ℓ , I^A , and I^B generate specific frequency-range response; each exhibits their resonant properties. These are then scaled by k_0^ℓ , k_0^A , and k_0^B , respectively. **Figures 6A,E,D** show these signals for both fixed and modulated transfer functions. The rate of change of these specific frequency-range responses are also measured by the filters and weighted by k_1^ℓ , k_1^A , and k_1^B , shown in **Figures 6B,G,L**. Filter parameters plotted in **Figure 4** show that the high-frequency weight k_0^B significantly decreased in the modulated filter compared to the fixed filter while its weight for rate of change k_1^B has increased. The slow filter weights k_0^ℓ and k_1^ℓ have shown the reverse behavior; i.e., k_0^ℓ has increased while k_1^ℓ decreased. This results in a modulated response that has less contribution from the rate of change. Similar behavior is observed for alpha filter waves, albeit less significantly. The dynamics of weights of rate of change are caused by prediction time τ_p , which determine how in advance the filters estimate the future values. The prediction time of modulated high-frequency filter has increased to 23 ms from 2 ms for fixed gain filter, which shows the rapid changes are predicted sooner in advance. The alpha filter prediction time is similar in both cases. The magnitude and rate of change signals vs. time are plotted in **Figures 6D,I,N**.

The sum of adjusted magnitude and rate of change constitute the output of each filter, as shown in **Figures 6C,H,M**. As seen in these figures, once a sudden change happens (i.e., the onset of impulse function), the contribution from the rate of change component is significant at the early stage as it is seen at times between 20 and 25 ms. As the rate of change decreases, the contribution from its component also declines and the full responses of filters are mostly proportional to the value of its input. **Figures 6E,J,O** show the power spectra of the corresponding signals which show that modulation of both slow and alpha filters has improved the power of their corresponding rhythm bandwidth while the modulation of high-frequency (i.e., beta) filter has decreased the power of beta frequency range.

Ultimately, the total response of specific nuclei is obtained by summing the responses generated by each filter, shown in **Figure 6P**. The power spectra of the responses are plotted in **Figure 6Q** which confirms that the modulated gain filters contain more information about the stimulus. The synaptic gain changes vs. time, for those involved in specific nuclei, are plotted in **Figure 6R**.

3.3.2. Step Response

Another aspect of stimulus-driven responses and their relation to attention, is how visual evoked potentials are related to a sudden change from an initial uniform steady state input to a new level. Hence, we drive the system by a step-function stimulus at $t = 20$ ms and again plot the specific nuclei's slow, alpha, and beta-frequency filter responses to obtain the total response



ϕ_{sn} . This enables us to study how gain modulation can modify the prediction error when a sudden change in input occurs. The Control System Toolbox of Matlab 2018a is used to carry out the calculations of time responses for both fixed and modulated gain

transfer function T_{sn} using the equations and parameters from previous sections.

Figure 7 shows the detailed responses of modulated-gain transfer function T_{sn} and compares it with responses

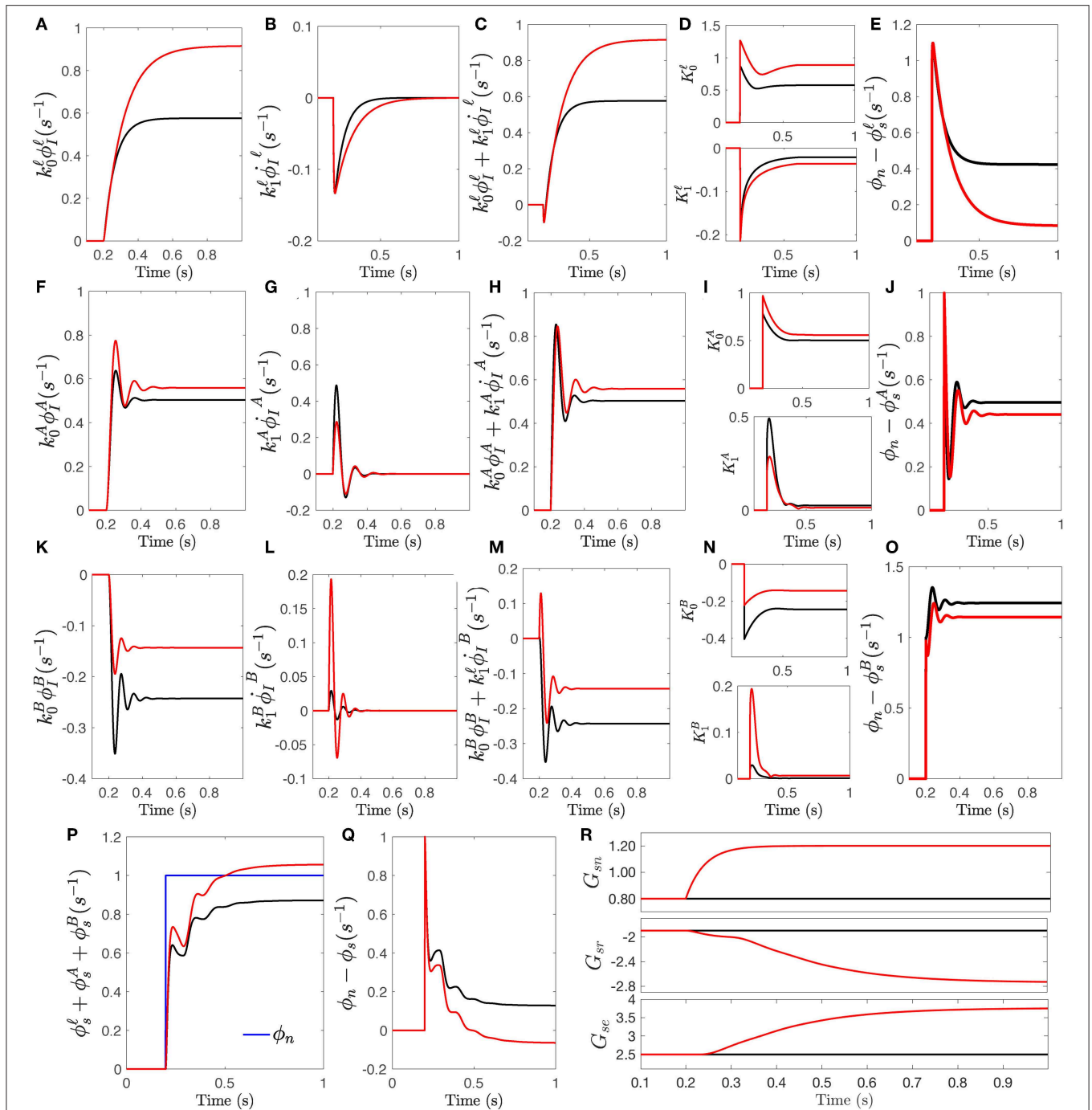


FIGURE 7 | The response of transfer function T_{sn} to a step stimulus onset at $t = 200$ ms. The responses for fixed-gain transfer function (i.e., no modulation) are in black and the responses for the transfer function with modulated gains of $\Delta = +0, 0.5$ are in red. **(A)** Slow wave response ϕ_s^l generated by convolution filter $l_{sn}^l(s)$, scaled by k_0^l . **(B)** The rate of change $\dot{\phi}_s^l$, scaled by k_1^l . **(C)** The slow filter response obtained by the sum of the two scaled signals. **(D)** Amplitude modulations vs. time for both magnitude and the rate of change of slow wave. **(E)** The residual signals obtained for ϕ_s^l . **(F)** Same as **(A)** for ϕ_s^A . **(G)** Same as **(B)** for ϕ_s^A . **(H)** Same as **(C)** for ϕ_s^A . **(I)** Same as **(D)** for k . **(J)** Same as **(E)** for ϕ_s^A . **(K)** Same as **(A)** for ϕ_s^B . **(L)** Same as **(B)** for ϕ_s^B . **(M)** Same as **(C)** for ϕ_s^B . **(N)** Same as **(D)** for k . **(O)** Same as **(E)** for ϕ_s^B . **(P)** The total step response ϕ_s . **(Q)** The residual signals obtained of total response ϕ_s . **(R)** The changes of synaptic gains vs. time.

calculated for fixed-gain transfer function. Similar to the impulse response, each resonance filter collects a specific range of stimulus and measures the amplitude and rate of

change of corresponding part of the stimulus. **Figures 7A,E,K** show the weighted amplitude and **Figures 7B,G,L** show the weighted rate of changes for each filter. The slow,

alpha, and beta responses are evoked in the way similar to the impulse responses; i.e., same weights are applied to the waves. The modulation signals as the envelop of the magnitude and rate of change signals vs. time are plotted in **Figures 7D,I,N**. The output of filters are plotted in **Figures 7C,H,M**. Although both impulse and step stimuli provide similar results in transient time, because the step drive does not return to zero but remains at a new intensity, it enables the analysis of the steady-state error of the response in time domain. The errors between the stimulus and each filter's output are plotted in **Figures 7E,J,O** which confirms that in all three filters the error is reduced by the modulated-gain transfer function compared to the fixed-gain transfer function. The individual error reductions result in the overall error between the total response ϕ_{sn} , plotted in **Figure 7P**, and stimulus being reduced, as seen in **Figure 7Q**. The synaptic gains changes induced by the step response are plotted in **Figure 7R**.

Overall, comparing the responses of the systems with and without modulation terms reveals that the key effect of modulation is to adjust the weights on features of input: its value (modulating amplification) and rate of change (modulating the latency by changing the prediction time) in such a way that new information from the external world is emphasized. These results support theories of attention that suggested that the magnitude and the processing time of stimulus-driven neural activity can be modified by attention and can provide a basis for mechanisms that explain attentional suppression and facilitation of unattended and attended inputs, respectively (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998; Gazzaley et al., 2005; Noudoost et al., 2010).

4. SUMMARY AND DISCUSSION

Starting with the brain's physical characteristics we have used a neural field corticothalamic model, with modulation of neuronal gains, to evaluate the linear responses of corticothalamic neural populations to stimuli and interpret the results in terms of signal prediction and attention via control theory. In particular, control theory allows stimulus-evoked synaptic gain modulations to be interpreted as implementing attention such that salient stimulus features are emphasized, especially by increasing high-frequency gains in response to sudden stimulus changes. The main results are:

- (i) We extended the NFT of corticothalamic system by incorporating an expression for neural gain modulation that is sufficiently general to encompass a variety of synaptic modulation mechanisms, such as long-term potentiation/depression, facilitation, habituation, and sensitization.
- (ii) Once gain modulation was incorporated as part of the system response, the system transfer functions were shown to differ from fixed term transfer functions only in their now-dynamic gain coefficients. These contain a fixed

part, as in previous analyses, plus a gain change that occurs in response to the stimulus.

- (iii) The modulated transfer functions were then used to calculate the dominant resonant response modes that are excited by external signals. As in recent work (Babaie-Janvier and Robinson, 2018), these corresponded to slow-wave, alpha, and beta responses each of which can be interpreted as being a standard second-order PID filter, but the gain parameters of these filters can change dynamically in response to stimuli.
- (iv) Notably, it was shown that each resonance can be interpreted as a filter that responds to two features of the incoming stimulus in its resonant frequency range: the stimulus value and rate of change. These responses are then separately scaled and summed to generate the filter output, with the scalings tuned by the gain changes that occur as part of the system response. Gain increases tend to increase the weight placed on the stimulus's rate of change at the expense of its value, which we interpret as a form of attention that focuses on the feature that will better enable the future course of the stimulus to be predicted. This type of attention is fundamental, in that it involves an implicit model of stimuli as following second-order differential equations, as embodied in the transfer function's poles; understanding it is essential before progressing to more complex circumstances and stimuli.
- (v) It was further found that weight changes of the slow and alpha filters tend to be anticorrelated with those of the beta filter, which further enhances the shift of attention from stimulus value to rate of change when there is a sudden change in the stimulus. This enables the system to respond even faster to changes in the input stimulus. These gain changes alter the strengths of the various resonances, an effect that can potentially be tested against correlations observed in evoked response experiments.
- (vi) Simulations of responses to impulse and step-function stimulus changes, with and without dynamic gain changes, verify the above points and demonstrate that attention shifts to emphasize rate of change when sudden stimulus changes occur, then back to emphasize stimulus value under static conditions. Errors in stimulus tracking and prediction were significantly reduced by dynamic gain responses relative to the fixed-gain case.

Overall, this study has demonstrated that corticothalamic responses to time-varying visual signals can be interpreted as implementing data filters that collectively span a wide frequency range, and whose parameters are dynamically adjusted, to enable estimation of incoming signals. Attention is interpreted as being due to dynamic gain changes that increase the weight attached to stimulus rate of change when sudden changes occur, and to stimulus value under static conditions. Future work will extend these insights to incorporate spatially varying stimuli, for example, and to the use of resulting stimulus predictions in decision processes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

TB-J and PR conceived the research and developed the theory. TB-J performed the computations. TB-J and PR verified the analytical methods. TB-J worked out the

technical details, performed the numerical calculations, produced the figures, and numerical results. TB-J and PR discussed all the results. TB-J drafted the first version of the manuscript and both authors contributed to the final manuscript.

FUNDING

This work was supported by the Australian Research Council under Center of Excellence grant CE140100007 and Laureate Fellowship grant FL140100025.

REFERENCES

- Abeyuraya, R. G., Rennie, C. J., and Robinson, P. A. (2015). Physiologically based arousal state estimation and dynamics. *J. Neurosci. Methods* 253, 55–69. doi: 10.1016/j.jneumeth.2015.06.002
- Babaie-Janvier, T., and Robinson, P. A. (2018). Neural field theory of corticothalamic prediction with control systems analysis. *Front. Hum. Neurosci.* 12:334. doi: 10.3389/fnhum.2018.00334
- Barceló, F., Suwazono, S., and Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nat. Neurosci.* 3:399. doi: 10.1038/73975
- Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*. 2nd Edn. Heidelberg: Springer-Verlag.
- Breakspear, M., Roberts, J. A., Terry, J. R., Rodrigues, S., Mahant, N., and Robinson, P. A. (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
- Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biol. Cybernet.* 93, 91–108. doi: 10.1007/s00422-005-0574-y
- Engel, A. K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2:704. doi: 10.1038/35094565
- 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. R. 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
- Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., and D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *J. Cogn. Neurosci.* 17, 507–517. doi: 10.1162/0898929053279522
- Herrmann, C. S. (2001). Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353. doi: 10.1007/s002210100682
- Hillyard, S. A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 95, 781–787. doi: 10.1073/pnas.95.3.781
- Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. B Biol. Sci.* 353, 1257–1270. doi: 10.1098/rstb.1998.0281
- Jerath, R., and Beveridge, C. (2019). Multimodal integration and phenomenal spatiotemporal binding: a perspective from the default space theory. *Front. Integr. Neurosci.* 13:2. doi: 10.3389/fnint.2019.00002
- Jerath, R., Crawford, M. W., and Barnes, V. A. (2015). A unified 3d default space consciousness model combining neurological and physiological processes that underlie conscious experience. *Front. Psychol.* 6:1204. doi: 10.3389/fpsyg.2015.01204
- 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
- Kastner, S., and Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39, 1263–1276. doi: 10.1016/S0028-3932(01)00116-6
- Kerr, C. C., Rennie, C. J., and Robinson, P. A. (2008). Physiology-based modeling of cortical auditory evoked potentials. *Biol. Cybernet.* 98, 171–184. doi: 10.1007/s00422-007-0201-1
- Koch, C. (1999). *Biophysics of Computation*. Oxford: Oxford University Press.
- Luck, S. J., Woodman, G. F., and Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn. Sci.* 4, 432–440. doi: 10.1016/S1364-6613(00)01545-X
- Morgan, S. T., Hansen, J. C., and Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4770–4774. doi: 10.1073/pnas.93.10.4770
- Noudoost, B., Chang, M. H., Steinmetz, N. A., and Moore, T. (2010). Top-down control of visual attention. *Curr. Opin. Neurobiol.* 20, 183–190. doi: 10.1016/j.conb.2010.02.003
- Nunez, P. L., and Cutillo, B. A. (1995). *Neocortical Dynamics and Human EEG Rhythms*. Oxford: Oxford University Press.
- O'Connor, S. C., and Robinson, P. A. (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.
- Pessoa, L., Kastner, S., and Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *J. Neurosci.* 23, 3990–3998. doi: 10.1523/JNEUROSCI.23-10-03990.2003
- Pinsk, M. A., Doniger, G. M., and Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *J. Neurophysiol.* 92, 622–629. doi: 10.1152/jn.00974.2003
- Rao, R. P., 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
- Rennie, C. J., Robinson, P. A., and Wright, J. J. (1999). Effects of local feedback on dispersion of electrical waves in the cerebral cortex. *Phys. Rev. E* 59:3320. doi: 10.1103/PhysRevE.59.3320
- Rennie, C. J., Robinson, P. A., and Wright, J. J. (2002). Unified neurophysical model of EEG spectra and evoked potentials. *Biol. Cybernet.* 86, 457–471. doi: 10.1007/s00422-002-0310-9
- 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., Rennie, C. J., and Rowe, D. L. (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., Rowe, D. L., O'Connor, S. C., 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 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
- Van Albada, S. J., Kerr, C. C., Chiang, A. K., Rennie, C. J., and Robinson, P. A. (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
- Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fmri study. *Neuron* 30, 829–841. doi: 10.1016/S0896-6273(01)00328-2
- Wilson, H. R., and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol. Cybernet.* 13, 55–80. doi: 10.1007/BF00288786

Conflict of Interest: 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 © 2019 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.