Published in final edited form as:

Biol Cybern. 2011 December ; 105(5-6): 371–397. doi:10.1007/s00422-012-0472-z.

Modeling habituation in rat EEG-evoked responses via a neural mass model with feedback

Srinivas Laxminarayan,

Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA

Gilead Tadmor,

Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA

Solomon G. Diamond,

Thayer School of Engineering at Dartmouth, 8000 Cummings Hall, Hanover, NH 03755, USA

Eric Miller,

Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155, USA

Maria Angela Franceschini, and

Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA

Dana H. Brooks

Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA

Srinivas Laxminarayan: slaxmina@ece.neu.edu

Abstract

Habituation is a generic property of the neural response to repeated stimuli. Its strength often increases as inter-stimuli relaxation periods decrease. We propose a simple, broadly applicable control structure that enables a neural mass model of the evoked EEG response to exhibit habituated behavior. A key motivation for this investigation is the ongoing effort to develop model-based reconstruction of multimodal functional neuroimaging data. The control structure proposed here is illustrated and validated in the context of a biophysical neural mass model, developed by Riera et al. (Hum Brain Mapp 27(11):896–914, 2006; 28(4):335–354, 2007), and of simplifications thereof, using data from rat EEG response to medial nerve stimuli presented at frequencies from 1 to 8 Hz. Performance was tested by predictions of both the response to the next stimulus based on the current one, and also of continued stimuli trains over 4-s time intervals based on the first stimulus in the interval, with similar success statistics. These tests demonstrate the ability of simple generative models to capture key features of the evoked response, including habituation.

Keywords

Neural mass model; Evoked response; Habituation prediction; Low order model

1 Introduction

Neural modeling covers a wide range of scales, from coarse-scale dipole models (Mosher and Leahy 1998) all the way to detailed biochemical computational models (Bower and Beeman 1998; de Schutter and Bower 1994a,b). This article concerns neural mass models that provide simplified representations of the average, synchronized behavior of ensembles of neurons. Examples include models based on mean field theory (David and Friston 2003; Kiebel et al. 2006; Coombes et al. 2007; Liley et al. 2002) and physiologically inspired models, based on a single or few functionally representative neurons, each representing a neural subpopulation (Riera et al. 2006, 2007). Parsimonious models that still include a sufficient level of representation of physiology to be open to physiological interpretation, and to manipulation by such factors as biochemical or sensory stimuli, have the potential to be key components of joint representations of electrical and hemodynamic activities, relating EEG and fMRI or optical measurements in a unified set of source models (Daunizeau et al. 2007; Horwitz and Poeppel 2002; Lahaye et al. 2004).

This article concerns neural mass models that can replicate the phenomenon of habituation in the evoked EEG response to repeated stimuli. Habituation is viewed as the simplest form of learning (Rankin et al. 2009). It has been studied extensively over decades, covering diverse species, from aplysia (Bristol and Carew 2005; Purves 2004), through crayfish (Zuker 1972), crickets (May and Hoy 1991) and flies (Tully and Koss 1992), snakes (Andry and Luttges 1972), mice and rats (Shea et al. 2008; Hall 1968), to humans (Noguchi et al. 2004). Yet remarkably little is known about the neural mechanisms underlying habituation (Rankin et al. 2009; Garrido et al. 2009), and the dynamics of habituation in a neuronal population (Noguchi et al. 2004). Elucidating the nature of that dynamics is therefore a useful first step toward the understanding of mechanisms of learning and cognition at higher levels of complexity (Rankin et al. 2009).

Here we do not directly address underlying mechanisms; however, we do propose a methodology to model habituation dynamics using a simplified version of an existing neural mass model. In particular, to enable that neural mass model to predict habituation we propose a control structure in which model parameters are defined as functions of an auxiliary state, whose slow dynamics represents the post-firing relaxation of the neural mass. The use of slow states to regulate fast or bursty dynamics is commonplace, including low order models of the firing patterns of single neurons (Ghigliazza and Holmes 2004; Shilnikov and Rulkov 2004).

To demonstrate this mechanism, we employ the neural mass model structure introduced by Riera et al. (2006, 2007), which will be referred to as the *Riera model*. We chose this model as a contemporary representative which follows the classical framework of nonlinear lumped parameter models [see Sornmo and Laguna (2005, pp. 67–71) and Jansen and Rit (1995)]. The Riera model, schematically shown in Fig. 3, consists of a single pyramidal cell (PC) with feedforward and feedback currents from two neighboring GABAergic¹ Inter-Neurons (IN), each representing the respective subpopulation of the neuronal ensembles. The presentation includes an analysis and a simplification of Riera's model, the optimization of that model's parameters at varying levels of habituation, and their parametrization as functions of the control state.

The data set used to identify and validate our model has been presented and discussed in detail in Franceschini et al. (2008). It consists of EEG measurements of the rat somato-sensory cortex during episodes of medial nerve stimulation. 4-s long trains of stimuli were

¹GABA: γ -Amino butyric acid is an inhibitory neurotransmitter.

applied to the animal, anesthetized with alpha chloralose, at frequencies varying from 1 to 8 Hz. Generic examples of the ensemble-averaged evoked response are illustrated in Fig. 1, where each left panel depicts a complete 4-s record of the evoked response potential and the corresponding right panel zooms in on a 1-s portion. Features of the response waveform that are targeted by the Riera model are the first three half-cycles of a time varying, large scale, post-stimulus oscillations, starting with a sharp positive peak (P1), followed by increasingly diffused negative (N1) and positive (P2) peaks. The figure illustrates the diminishing amplitude of the response, as the inter-stimulus wait time reduces, including cycles with no discernible response. To illustrate the need for signal averaging and the essentially statistical nature of meaningful model predictions, the top right panel also shows the record of a single response (the dashed/blue line) where the coherence of the targeted features is lost.

The article is organized as follows: Sect. 2 presents the data, the targeted waveform features, and a detailed analysis of the Riera model, including reduced complexity versions of the model that will be subsequently used. Observations concerning habituation and the dynamic structure of the control state, leading to the control structure we propose, are presented in Sect. 3. Model parameter estimation, including the optimization of the mass neural model under varying habituation conditions, and of the functional dependence of model parameters on the control state, are the subject of Sect. 4. Performance evaluation is presented in Sect. 5. Our approach and results are discussed in Sect. 6, including comments on the methods and on the difficulty of characterizing ground truth in these evaluations, the advantages and disadvantages of the linear approximation to the Riera model that we introduced, and some comments on the implications of our study. Concluding remarks are presented in Sect. 7. Appendices are used to detail technical aspects of the data processing method that we use, in Sect. A, and of the analysis of Riera's model and its simplifications, in Sect. B. The Appendix also lists, in Sect. C, those consensus habituation criteria, from Rankin et al. (2009), that are pertinent to this study, and reviews the way these criteria are met by the proposed model and by the EEG data used in its validation.

2 Phenomenology and models

Following a brief description of the experimental data used here, in Sect. 2.1, we highlight properties and issues associated with the evoked response, in Sect. 2.2, and review the underpinning of Riera's model and of its simplified versions, in Sect. 2.3. Mathematical details associated with data preprocessing and with model analysis are deferred to the Appendices. The quantitative working definitions of the evoked response waveform, and their use in model parameter identification and in model evaluation, are the subject of subsequent sections.

2.1 The experimental data

Records of EEG-evoked response to medial nerve stimulation were collected from six rats at the Martinos Center for Biomedical Imaging of the Massachusetts General Hospital. Data acquisition is described in details, in Franceschini et al. (2008). We detail below a few facts that are particularly pertinent to the present discussion.

Each stimulus consisted of a 200- μ s current pulse, applied to the rat fore-paw. Twenty runs per paw comprised trains of stimuli, each over a 4-s time interval, separated by random-length quiet periods, during which no stimulus is applied, averaging 12 s in duration. The length of the inter-stimulus intervals (ISIs) between adjacent stimuli remained fixed during the 4-s period, but were changed randomly from one stimulus train to the next. Specifically, in 10 of the runs, the ISI was chosen from a random ordering of $\{1, 1/3, 1/5, 1/7\}$ s, and in the other 10, from $\{1/2, 1/4, 1/6, 1/8\}$ s.² Each of the 20 runs in each of the data sets provided an average of 70 stimulus intervals at each of the four (even or odd) pertinent

frequencies. The EEG data sampling rate was 1 kHz. While multiple electrodes were used in data collection, the analysis presented here is based on the EEG records from the single electrode with the maximal average response, which was thus the closest to the somatosensory area.

Using the index $n = 1, \dots, 6$, to identify an individual rat, and $S = L$ (left) or $S = R$ (right) to indicate the brain hemisphere containing the stimulated region, the data will be organized and analyzed in 12 data sets, each referred to by the respective nS notations. Thus, data set 3L comprises left hemisphere EEG records from rat #3.

2.2 Waveforms of characterization of the evoked response, and the habituation phenomenon

We are interested in the generic P1–N1–P2 waveform component of the post-stimulus EEG signal, illustrated in Fig. 1. Preprocessing of raw experimental data is required to isolate these waveforms. This includes the removal of a slow baseline drift, of the ambient 60 Hz and cardiac cycle interference, and of a substantial high frequency component of the EEG signal. Finally, ensemble signal averaging is needed to remove the significant stochastic distribution of individual response cycles, illustrated by the dashed/blue curve in the top right panel in Fig. 1. Details on data preprocessing procedure are provided in Appendix A.

Figure 1 illustrates some basic observations: the P1–N1–P2 waveform is clearly prominent in the post-stimulus EEG signal, when stimuli are separated by a sufficiently long ISI (0.5–1 s, top two rows). The combined duration of these waveforms is typically well less than 125 ms, with P1 typically lasting ~10 ms, N1 lasting ~30 ms, and P2 about the same or longer.

The intensity of the P1–N1–P2 waveform begins to attenuate when shorter ISI's are applied. This phenomenon is already observed at the response to the second stimulus in a train, when the ISI reduces to ~330 ms (third and lower rows in Fig. 1). For even shorter ISI (200 ms, rows 5–8 in Fig. 1), attenuation may reach points where no coherent P1–N1–P2 waveform is discernible. Strong attenuation is invariably manifest in the response to the second stimulus in these 4-s trains, and near complete attenuation is typically observed in an alternating fashion in subsequent stimulus cycles. These latter observations reveal an additional (so far, qualitative) dependency of the observed attenuation: The level of attenuation of the evoked response to a stimulus becomes more pronounced when the response to a preceding stimulus has been more intense. Since the response to the first stimulus in a train is the highest, attenuation is most pronounced in the second stimulus. Furthermore, the level of attenuation seems to depend on the response of multiple previous stimuli, in a cumulative manner, as evidenced by an increasing occurrence of cycles with no response at shorter ISI's, even when the previous response, when discernible, is of relatively low intensity.

Summing up, these qualitative observations suggest that the attenuation seen in the evoked response, in our data, reflects a balance between a cumulative effect of past responses, and a *forgetting* process that diminishes that cumulative effect. This phenomenon of adaptation to short term changes in stimulus patterns is commonly referred to as *habituation* and viewed as the simplest form of a learning mechanism (Rankin et al. 2009). This qualitative conclusion is at the core of the modeling approach proposed in this article. It will be revisited, stated in concrete mathematical terms, and validated with quantitative rigor, through the remainder of the article.

²While we shall often simplify notations by use frequency terms (i.e., refer to 1–8 Hz stimuli), we bear in mind that the analysis presented here is focused on the temporal evolution of the evoked response, and the adaptation in the response to varying ISI lengths.

The goal of developing a habituation model of the evoked response, as captured by changes in quantifiable attributes of the P1–N1–P2 waveform, and the rigorous evaluation of that model, both require consistent, quantitative definitions. Yet even following preprocessing, these attributes remain loose qualitative idealizations which are difficult to quantify. In particular, residual high frequency oscillations, conspicuous in the 3–8 Hz zoomed plots in Fig. 1, obscure both the starting and terminal times of each of the three phases, and meaningful measures of intensity of these components, such as their L_1 (area), L_2 (RMS) or L_∞ (peak) norms.

To further elucidate the causes of an inherent ambiguity in a definition of this waveform which clearly distinguishes it from pre-stimulus EEG signal characteristics, we studied average power spectra of three categories of subintervals of the EEG signal, as shown in Fig. 2:

- i. Pre-stimulus, 125 ms intervals of the EEG signal, termed *background* (dash-dot/blue curve).
- ii. The 125 ms intervals, each immediately following a stimulus, for those cases where a discernible P1–N1–P2 waveform is found, referred to as *response* intervals (solid/green curve).
- iii. Post-stimulus signals over intervals of the same length, during which no discernible P1–N1–P2 waveform is detected (found here when the ISI is <166 ms), presumably as a result of habituation. Those are referred to as *no-response* intervals (dashed/red curve).

The spectra in Fig. 2 were computed for data set 4L, applying the averaged periodogram approach (Stoica and Moses 1997) to collections of time intervals (always 125 ms long as noted) representing the three categories. Specifically, the spectrum of the *background* signal was estimated from a 6-s signal obtained by averaging the 6-s starting intervals of the twenty data records in data set 4L, recorded immediately before the first stimulus of each respective run. This 6-s signal was partitioned into 48 blocks of 125 ms each. The estimated *background* signal spectrum is the average of the spectra computed separately for each of block.

Each 125 ms interval of the EEG signal, immediately following a stimulus, was labeled as either a *response* (102 blocks) or a *no-response* (42 blocks) interval, using the criterion described in³ Sect. 6.3. The *response* spectrum is estimated as the average of the power spectra computed over each of the 125-ms *response* intervals. The *no-response* spectrum was computed analogously.

Observations concerning Fig. 2 include that

- The amplitudes of the three spectra peak at about the same frequency, i.e., at ~10 Hz, which is at the low end of the frequencies associated with components of the P1–N1–P2 waveform (10–50 Hz). This suggests the possibility that signals of all three categories are variants of common neural mass dynamics.
- The spectrum of the *response* signal has the highest amplitude over the low-to-middle frequency range. It is higher with a significant margin (note the logarithmic scale) at the frequency range pertinent to the rapidly varying oscillation period of the combined half-cycles P1, N1, and P2 (10–50 Hz). It is also significantly higher

³We are aware of the “chicken and egg” issue, of using a quantitative distinction between cycles with and without a discernible evoked response in a discussion of that distinction. We ask for the reader’s indulgence till we revisit this issue in a rigorous and quantitative manner, culminating in the derivation of this labeling, in Sect. 6.3.

than the other two spectra over the frequency range of 5–15 Hz, near the peak of the background signal spectrum.

- The amplitude of the spectrum of post-stimulus *no-response* cycles is the lowest of the three spectra at lower frequencies. It becomes larger than that of the background signal at about 40 Hz, and it gradually closes toward the response signal at higher frequencies.
- Post-stimulus spectra, with and without response, are significantly lower than the background signal spectrum at higher frequencies.

A possible interpretation, suggested by these properties, is that the P1–N1–P2 evoke response pattern is an expression of a gradual modification of normal background activity, rather than a distinct phenomenon. According to such an interpretation, the higher amplitude and narrower bandwidth, during the P1 phase, are the result of an elevated synchrony of normal oscillatory behavior, along with a slight shortening of the characteristic wavelength. This synchrony is gradually eroded during the N1 and P2 phases, and beyond, resulting with lower ensemble average amplitudes, lengthening periods and widening bandwidth. Following a short ISI, the new stimulus is insufficient to cause the same level of synchronization, hence the lower averaged oscillation amplitudes and wider bandwidth. These reductions makes it increasingly difficult to distinguish the P1–N1–P2 waveform from similar amplitude but higher frequency oscillations, as observed in the 8-Hz response in Fig. 1. Yet even in that case, habituation does not altogether eliminate the post-stimulus synchronization; the EEG bandwidth remains narrower than in the background signal, and the high frequency component of the habituated signal remains less intense than that of the background signal.

Adopting this interpretation as a reasonable working hypothesis, our conclusion is that any consistent quantitative definition of the P1–N1–P2 waveform, and of features thereof, must necessarily be based on a somewhat arbitrary choice. Likewise, any categorical distinction between response and no-response stimulus cycles is necessarily based on a threshold simplification of a gradual change over a continuous range. We bear these conclusions in mind, along with their implications of the inherent ambiguity about concepts of ground truth, as we proceed to make and use such definitions and distinctions, through the remainder of this article.

In closing, we comment on the fact that the spectra plotted in Fig. 2 are used as a generic example. The response spectra would change somewhat if, for example, they were computed separately for data records associated with different stimulus ISI lengths (i.e., for different stimulus frequencies), or if they were computed for the raw data, rather than the preprocessed data used in Fig. 2. That said, we note that our studies of such changes showed that they do not affect the qualitative shapes of the plots, and hence that the observations made above will remain intact.

2.3 The Riera model

Riera's model, originally presented in Riera et al. (2006, 2007), will be used in this study as a benchmark for the proposed, broadly applicable habituation mechanism. The purpose of this section is to highlight details and conclusions regarding Riera's model that will be needed during the main body of this article.

The precise mathematical formulation of the model, its analysis, and a detailed discussion, leading to its substantial simplification, are deferred to Appendix B. In lieu of these details, we use here two schematic representations. The first, already mentioned above and shown in Fig. 3, describes the model's key structural components. Each of three *mass neurons*

represents a neuronal population, including PCs, and two IN groups, acting in inhibitory feedforward and feedback roles. Three input currents represent the arrival of the stimulus signal at the somatosensory cortex, along three different paths. Those include signals received from the thalamus through the basal dendrites (I_1^+), signals from surrounding neurons, via the apical dendrites (I_2^+), and the excitatory innervations to the transmission GABAergic IN (I_3^+).

At a system block diagram level, each of the three system blocks in Fig. 4 represents the dynamical system model of a neuronal group in Fig. 3, with an indication of states participating in that block. In all the model comprises nine states. The diagram also shows the saturation nonlinearity, $f(\cdot)$, which governs interconnections between the neuronal groups. The structure of the differential equations, in Riera's model, and the values of the parameters that serve as coefficients in these equations, are based on physiological representations of single neurons, as elaborated in Riera et al. (2006, 2007). This has the advantage of direct connection to measurable or generally known physiological data, regarding a specific organism or process.

In contrast, Riera's definitions of the three input currents are generic, using delayed temporal Gaussian curves of the general form

$$I^+(t) = g e^{\left(\frac{t-\tau}{\sigma}\right)^2}. \quad (1)$$

This simple choice enables the parameterized current inputs to play a key role in scheduling the predicted arrivals of P1, N1, and P2, using the respective delay coefficient, τ , and to affect the shape of the generated waveform components, using the standard deviation σ and the gain g . Indeed, model analysis reveals that each of I_i^+ , $i = 1, 2, 3$, is used in Riera's model as the primary driver of one of the three waveform components: I_3^+ drives P1, I_1^+ drives N1, and I_2^+ drives P2 (these statements are sharpened by the mathematical analysis of the model in Appendix B).

The absence of indices in (1) is intentional and is the subject of the following cautionary note about a potential notational morass. To facilitate reference to the original presentation in Riera et al. (2006, 2007), we keep the original input current notations, I_i^+ , $i = 1, 2, 3$, as in Figs. 3 and 4. Yet, in order to facilitate the ensuing discussion of this article, we chose to index the Gaussian formulations according to the order of arrival of the respective inputs. The two sets of indices are reconciled in Table 1 (the reference to $\tilde{I}_3^+$ will be clarified below).

The habituation mechanism, which we describe in Sect. 3, substitutes constant model parameters by parameters that are functionally dependent on an auxiliary saturation state, s . Finding such dependencies is an inverse problem. A well-posed formulation of that problem requires a careful dimensionality analysis of model parameterization, to avoid excessive degrees of freedom which would lead to an ill-posed problem. Two complementary aspects of that analysis, in the current study, are (a) model structure simplification and order reduction, and (b) the identification of that subset of the reduced order model parameters that need to be functionally dependent on the saturation state.

Considering model order reduction, it is established in Appendix B that the feedback signal shown in Fig. 4 $f(x_2) \approx 0$ under broad conditions, including the operating range of this study. This observation renders the (nonlinear) feedback path in Figs. 3 and 4 redundant, and reduces the original dynamical system from nine to three states. The signal $\tilde{I}_3^+$, in Fig. 4, is then defined as the saturated output of a first-order linear filter, applied to the generic

(Gaussian) representation of I_3^+ . There will thus be neither a loss of generality, nor a loss of a physiological foundation or, for that matter, of technical facility, if the original parameterization of I_3^+ will simply be substituted by a direct parameterization of $\tilde{I}_3^+$ [by (18), in the Appendix]. This reduces the three state nonlinear model to a linear, two state model, with only negligible changes of the predicted response, when compared with the original, nine state model. This simplification will become an asset in facilitating the laborious, nonlinear optimization problem, used in the course of setting and solving system identification problems, in Sect. 4.

An analysis of the dependence of model output on its parameters revealed that only six parameters are habituation-dependent, in both the three state, nonlinear model and the two state, linear model. Referring to Table 1 and to Eq. 1, the parameters we shall use to account for habituation will be the three gains, g_i , and three delays, t_i . With the exception of the need to compensate for the linear filter phase-lag between the arrival of I_3^+ and that of $\tilde{I}_3^+$, these parameters are identical in the two and the three states models. For simplicity, we shall thus use the same nomenclature for both models, referring to g_i and t_i , $i = 1, 2, 3$, in both models. The values of the dynamical system coefficients and of the standard deviations, σ_i , remain constant throughout, and are given in the Appendix.

3 Habituation control structure

To enable a neural mass model to mimic the effect of habituation, we introduce the schematic control structure in Fig. 5. This structure stipulates the existence of an auxiliary, slow control variable, denoted s , which emulates a saturation (hence the notational choice) and relaxation dynamics in the neural mass. Recalling the intuition highlighted in Sect. 2.2, the purpose of the control variable s is thus to capture both the cumulative contribution to habituation of previous response cycles, and the dynamic *forgetting* of that contribution while no new stimulus is applied. The variable s is therefore defined as the output of a stable dynamic system that integrates the amplitude of the response to past stimuli. In turn, as a new stimulus is applied, at a time t_s , parameters of the neural mass model are adjusted as functions of $s(t_s)$. The class of parameter scheduling methods to which this mechanism belongs is common in nonlinear system identification (Nelles 2001, p. 195). It builds on classical ideas of time constant separation and inertial manifold embedding that are common in models of myriad physical and chemical phenomena (Gorban et al. 2007); in particular, these ideas are fundamental in the closely related, low order models of firing patterns in single neurons (Ghigliazza and Holmes 2004; Shilnikov and Rulkov 2004), and in low order models in fluid dynamics (Stuart 1971; Tadmor et al. 2011), the second author's other area of interest.

The following discussion is used to motivate our choice of the structure of the ordinary differential equation (ODE) governing the auxiliary state, s . The convention we use, in what follows, is that $s = 0$, that $s(t_s) = 0$ is associated with no habituation, and that habituation increases as $s(t_s)$ grows. To be able to serve in the role we envision, the spread of values of $s(t_s)$ needs to match the spread of levels of evoked response, observed in our data. Specific examples include:

1. Following an isolated stimulus⁴ at the time t_s , the value of $s(t_s + 0.5 \text{ s})$ should be sufficiently small to be associated with no habituation. This will match the absence of noticeable habituation in our EEG response data, when the ISI is 1 and 0.5 s (1 and 2 Hz stimulus trains).

⁴We use the term "an isolated stimulus" to refer to one that is separated from any preceding stimulus by an ISI that guarantees no residual habituation effects; presumably, the first stimulus in any 4 s train, in our data, satisfies this requirement.

2. The value of $s(t_s + \tau)$, following an isolated stimulus, should be sufficiently high to cause, at least a moderate habituation, when the ISI is $\tau = 0.33$ s or shorter. This is dictated by the observed habituation in the response to the second stimulus, at stimulus frequencies of 3 Hz and above.
3. The value of $s(t_s + \tau)$, following an isolated stimulus, should be sufficiently high to prevent any discernible response, when the ISI is $\tau = 0.166$ s or shorter. This is dictated by the lack of noticeable evoked response to the next stimulus, for stimulus frequencies of 6 Hz and above.
4. Beyond these, easily stated requirements, s should be able to integrate the effect of frequent, small responses, to generate the alternating response amplitude observed in the data for the shorter range of ISI lengths.

These requirements translate into constraints on the decay time constant of a hypothesized first-order ODE, governing the evolution of s . A detailed examination of our data revealed that they could not be matched by the response to a first order, linear time invariant ODE. Rather, the data suggest that the desired dynamics should provide for a far shorter decline time constant, at large values of s , than at small values of s . Our choice of a nonlinear system structure that meets these characteristics is

$$\dot{s} = -\frac{1}{\tau_s}(0.5+s)s + g_s u(t), \quad (2)$$

where the input, $u(t)$, is the rectified (i.e., the absolute value of the) “P1–N1–P2” portion of the EEG signal. Indeed, the effective time constant, $\tau = \tau_s/(0.5+s)$, in this formulation, decays as a function of s (for details about how this control structure meets the qualitative behavior of habituation refer Sect. C.2). Model parameters were selected as $g_s = 20$ and $\tau_s = 0.5$ s. We shall motivate the guidelines for this selection shortly.

Before we turn to our next task, the determination of model parameter dependencies on values of s , we use Figs. 6 and 7 to illustrate the required properties of s and our considerations in making the choice of its governing dynamics.

Figure 6 shows trajectories of the rectified P1–N1–P2 waveforms (top panels), and of the auxiliary state s , (bottom panels), when the former are used as inputs in (2), initiated at $s(0) = 0$. The P1–N1–P2 waveforms are extracted from response data for stimulus trains with ISI of 0.25 s (4 Hz, left panels) and 0.125 s (8 Hz, right panels), using a best-fit procedure we elaborate in Sect. 4. In essence, during each response cycle, the parameters of the neural mass model are optimized for best fit of the predicted P1–N1–P2 waveform to the data signal.⁵ Having no preceding stimulus, the response to the first stimulus is large, driving the value of s , sampled at the end of that cycle, to be the largest of the values sampled along the respective train. By the same token, the habituation effects of the first response cycle is the most pronounced, and the response to the subsequent stimulus, in each of the two cases, is the smallest. Indeed, no response is detected during the second stimulus cycle, when an ISI of 0.125 s is used (8 Hz case). Subsequent variations in the response amplitudes continue to correlate with variations of $s(t_s)$, the evaluations of $s(t)$ at the time of a new stimulus, at the end of each response cycle.

⁵Considering the difference seen between the two examples in the response amplitudes along the first stimulus cycle, note that causality bars that difference from being indicative of the subsequent stimulus properties, such as frequency. Rather, the observed difference is an example of the spread in our data, which may be attributed, e.g., to the effect of stochastic variations in pre-stimulus background activity or the mild habituation caused by the background signal.

The data used to produce Fig. 7 were obtained by the same procedure as in Fig. 6. The figure shows the continuous 1-s long trajectory of s , following the application of the first, isolated stimulus in a 4-s stimulus train (solid/red curve). The + signs mark the sampled $s(t_s)$ for eight 4-s response data records, corresponding to each of the stimulus frequencies, from 1 to 8 Hz. (The dotted rectilinear lines are included as visualization aids, and are *not* continuous time trajectories of s .) This figure illustrates the requirements from a successful parameter choice in (2): The sampled value of s are intended to guide the adjustment of the neural mass model parameters, and thus, to determine the level of habituation in the response waveform, generated by the model, following each stimulus. The vertical separation of the sampled values, in Fig. 7, should therefore be ample to robustly distinguish between the corresponding response intensities, in subsequent cycles. The faster-than-exponential decay of the response to the initial stimulus (the continuous solid curve) is needed to achieve that separation. While, the selection of the gain, g_s , would have been immaterial to the desired stretch in a linear equation, the implied nonlinearity makes a successful choice dependent on both g_s and of τ_s , which were found by a parameter scan.

Our next task is to determine the precise functional dependence of the neural mass model parameters on the auxiliary, saturation state, s .

4 Empirical model optimization

The ultimate goal of computing functional dependence of neural mass model parameters on the control variable s will be derived in two steps. Starting in Sect. 4.2, model parameters are optimized for best-fit data, separately for each distinct stimulus cycle. The results of this step will be instrumental in initiating the optimization procedure by which functional dependencies of neural mass model parameters on s are derived as optimized curve-fits, in Sect. 4.3. To put the overall discussion in context, we begin, in Sect. 4.1, with brief descriptions of the three model variants that will be constructed in this section, and analyzed and discussed in subsequent sections.

4.1 Model Variants

4.1.1 We consider three model variants

BFM: The first variant is not a truly predictive model, and may be viewed as a data filter of sort. Its use in subsequent sections will be as a performance benchmark and a computational tool. In this variant, model parameters are optimized for best fit between the P1–N1–P2 waveform, generated by the model, and the same waveform, as found in the 125-ms stretch of the post-stimulus EEG data signal *in a specific stimulus cycle*. This variant will be termed the *best-fit model*, and referred to by the acronym BFM.

The remaining two model types are predictive and require an explicit formulation of functional dependencies of model parameters on the control state s .

PM1: Here, model parameter values that are used to generate the predicted response during one stimulus cycle, are computed as functions of an evaluation of s , at the stimulus that initiates that cycle. Here s has been driven by the BFM response during the immediately preceding stimulus. This *single-step-ahead predictive model* will be referred to by the abbreviation PM1.

PM: The last variant predicts an entire 4-s signal, in response to a stimulus train. Model parameters are optimized for BFM of the first cycle, and predictions through the remaining response cycles are produced by a simulation of the two-way coupling between the neural mass model (with s -adjusted coefficients) and the auxiliary dynamics of s , according to (2).

Specifically, the input in (2), during one response cycle, is the contemporary model-prediction of the EEG waveform. In turn, the evaluation of s at the arrival of the next stimulus, determines the coefficients of the neural mass model that is used to predict the EEG response, during that subsequent response cycle. This, *long term predictive model* is denoted PM.

The following comment concerns both the PM1 and PM variants. In both cases, the auxiliary state s is initiated with the value $s = 0$, at the beginning of the first response cycle of each of the studied 4-s stimulus trains. This reflects the idealization by which the response to the first stimulus is un-habituated. By the same rationale, one may expect that the EEG responses to all first stimuli in these 4-s stretches, in our data, should be essentially identical. Yet, nonnegligible variations between these first cycles, however limited, are conspicuous in our data. Since the average wait time between distinct 4-s stimulus trains is of 12 s, and since no consistent habituation is observed when the ISI lasts 1 s, the residual habituation effect of one 4-s response train on the first response cycle of the next, seems to be ruled out as the main cause for observed response amplitude variations. A plausible, albeit speculative alternative, attributes these variations to the stochastic variability in pre-stimulus background activity, and to the small habituating effects of that activity. This explanation continuous the line of thought suggested in Sect. 2.2, whereby the change from background activity to a fully pronounced evoked response, occur along a continuous scale, including the varying levels of neural mass synchronization and of bandwidth narrowing and shifting. Thus, just as the evoked response, sustained background activity has an habituation effect. That effect would vary with both the slow intensity variations of the background activity, and with the point along the oscillatory cycle of the background waveform, at which a new stimulus is applied.

4.2 BFM

4.2.1 Parameter estimation for the BFM is broken into a two-stage optimization procedure

Step 1: The first stage is a procedure that extracts the P1–N1–P2 waveform component from the 125-ms stretch of the post-stimulus EEG data signal, in a given stimulus cycle. This stage thus serves also the goal of producing a consistent, quantifiable *ground truth* definition of features of that waveform, that will be used in evaluating model performance, later on. Leveraging the latitude, highlighted in Sect. 3, in making that definition, we select the parameterized family of piece-wise polynomial curves, of the form (3a), subject to the continuity constraints (3b), as a relatively simple and minimally restrictive representation of P1–N1–P2 waveforms. The waveform present in a particular response cycle is selected as the least mean squares (L_2) optimal approximation of the EEG signal by a curve in the family described by (3).

$$y_p(t) = \begin{cases} 0, & \text{if } t_s \leq t < t_{s1} \\ p_{11}(t-t_{s1})^2 + p_{12}(t-t_{s1}), & \text{if } t_{s1} \leq t < t_{s2} \\ n_{11}(t-t_{s2})^3 + n_{12}(t-t_{s2})^2 + n_{13}(t-t_{s2}) + n_{14}, & \text{if } t_{s2} \leq t \leq t_{s3} \\ p_{21}(t-t_{s3})^3 + p_{22}(t-t_{s3})^2 + p_{23}(t-t_{s3}) + p_{24}, & \text{if } t_{s3} \leq t \leq t_e \\ 0, & \text{if } t_e \leq t < 125\text{ms}, \end{cases} \quad (3a)$$

$$\begin{aligned}
n_{14} &= p_{11}(t_{s2} - t_{s1})^2 + p_{12}(t_{s2} - t_{s1}), \\
&\text{Continuity at P1-N1 connection} \\
n_{11}(t_{s3} - t_{s2})^3 + n_{12}(t_{s3} - t_{s2})^2 + n_{13}(t_{s3} - t_{s2}) + n_{14} &= p_{24}, \\
&\text{Continuity at N1-P2 connection} \\
p_{21}(t_e - t_{s3})^3 + p_{22}(t_e - t_{s3})^2 + p_{23}(t_e - t_{s3}) + p_{24} &= 0, \\
&\text{P2 vanishes at } t_e \\
3p_{21}(t_e - t_{s3})^2 + 2p_{22}(t_e - t_{s3}) + p_{23} &= 0, \\
&\text{The derivative of P2 vanishes at } t_e
\end{aligned} \tag{3b}$$

The nontrivial part of this optimization procedure concerns finding the start/end times of the P1, N1, and P2 components; i.e., $t_{s,i}$ $i = 1, 2, 3$, and t_e . Once those are set, the problem becomes a turn-of-the mill linear-quadratic (LQ) projection. Since the range of feasible values of $t_{s,i}$ $i = 1, 2, 3$, and t_e , is known and limited, the entire problem is easily solved, combining an exhaustive parametric search for the start and end times, with an LQ solution for the respective coefficients, for each selection of the said times.

Step 2: The second stage concerns optimizing the BFM coefficients, where we employed a bounded simplex line search for each parameter, iteratively cycling over the entire parameter list. As noted earlier, a preliminary sensitivity analysis indicated that the parameters of the dynamical system remain fixed, and are based on Riera's original selection (cf. the discussion in Appendix B). BFM optimization thus concerns only the delays t_i and gains g_i $i = 1, 2, 3$, in the delay Gaussian parameterization (1) of the input currents I_i^+ . Good initial guesses are essential in nonlinear/nonconvex optimization heuristics, such as the one employed here. The role of the piece-wise polynomial fit, in the overall derivation of a BFM, is precisely to produce such guesses. Thus, the start times $t_{s,i}$ $i = 1, 2, 3$, are used as natural initial guesses for the delays t_i $i = 1, 2, 3$, and the respective peak values of the three components of the piece-wise polynomial curve produce initial guesses of the gains, g_i $i = 1, 2, 3$. The iterative optimization procedure is stopped when the relative improvement in the norm approximation, from one iteration to the next, was less than 10^{-8} .

The procedure was implemented in MATLAB on a 64-bit Linux system with a 3-GHz processor. The average run time of the entire procedure was 277.8 ms for each stimulus cycle, totaling about 8-min for the entire ensemble of response cycles, in our 12 data sets.

4.3 Closing the loop: habituation-adjusted models

Much like BFM derivation, formulating functional dependencies of adjustable model parameters on the control variable s involves a challenging optimization task, that we address by a two stage procedure:

Step 1: Functional dependencies are first sought as optimal least-squares fit of curves to the scatter of BFM parameters, as defined in Sect. 4.2.

For each stimulus cycle in our data, BFM computations thus provide both the best-fit neural mass model parameters and the evaluation of $s(t_s)$, at the time of the stimulus that initiated that cycle. Together, these optimal coefficients and evaluations of s form the data used in deriving *initial* formulations of the functional dependencies of adjustable model parameters, on s . The scatter plots of each of the six parameters, $\Delta t_i = t_i - t_i(0)$ and of $g_i/g_i(0)$, $i = 1, 2, 3$, against the corresponding values of s , are shown in Fig. 8, where $g_i(0)$ and $t_i(0)$ are the gains and timing parameter values for the un-habituated response. In relations to the high scatter of the BFM values of the delay parameters, for high values of s , we note that the diminished values of the corresponding amplitudes, g_i , reduce the significance of poor delay

parameterization, for these values of s . We also note that g_3 is linearly proportional to g_2 , as shown in the lower left panel of Fig. 8, reducing the effective number of adjustable parameters to five.

The formal expressions we used for the functional dependencies $t_i(s)$ and $g_i(s)$, $i = 1, 2, 3$, are given in Eq. 4. An optimal fit of these expressions to the BFM scatter data amounts to an optimal selection of the fourteen free parameters, in these expressions. The optimal solutions, shown in Fig. 8 for the linear simplification of Riera's model, were computed, using the MATLAB nonlinear least-squares routine.

$$\begin{aligned}
 g_1/g_1(0) &= \begin{cases} a_1 s + a_2 s^{1/2} + a_3, & 0 \leq s \leq 1 \\ e^{-a_4(s-a_5)} & s > 1 \end{cases} \\
 g_2/g_2(0) &= \begin{cases} b_1 s + b_2 s^{1/2} + b_3, & 0 \leq s \leq 1 \\ e^{-b_4(s-b_5)} & s > 1 \end{cases} \\
 g_3/g_3(0) &= c_1 g_2/g_2(0) \\
 t_1 &= t_1(0) + d_1 s^2 \\
 t_2 &= t_2(0) + d_2 s^2 \\
 t_3 &= t_3(0) + \begin{cases} d_3 s^2 & s \leq 3 \\ 9 d_3 & \text{else.} \end{cases}
 \end{aligned} \tag{4}$$

Step 2: A second stage is required since the BFM response does not perfectly match either the data, or even the response of the neural mass model, when using the respective cycle's curve fit parameters, from Step 1. Here parameters are optimized with respect to either the PM1 or the PM objective: In PM optimization, the cost functional that is minimized is the sum of the squared relative L_2 error norms, computed for the entire length of each of the 4-s stretches of stimuli and response, excluding the response to the first stimulus in each of these 4-s stretches. (The respective BFM response is always used during the first cycle.) Having fixed a parameter choice in (4), the neural mass model parameters, in each cycle, are defined by the sampled value of s , at the stimulus time, where s has been driven by the model-predicted EEG response of the previous cycle. The difference in PM1 optimization is twofold: First, in accordance with the PM1 definition, the neural mass model parameters are defined by the sampled value of s , where s has been driven by the BFM-approximated EEG response of the previous cycle. Second, the relative estimation error is computed separately, for each response cycle, and not for an entire 4-s stretch.

This simulation based optimization is computationally demanding. The parameters found by optimal fit to the BFM scatter data facilitated that process, serving as initial guesses in this optimization. Here too we employed the simplex optimization procedure, in iterative cycles through the fourteen parameters of Eq. 4. The average run time of the optimization procedure (same hardware as in Sect. 4.2) was 160 s for each of the twelve data set. Figure 9 shows a sample PM1 and PM responses, compared with the BFM response for the linear Riera model.

5 Model performance analysis

We begin, in Sect. 5.1, with the list and quantitative definitions of features of the P1–N1–P2 evoked response waveform that will be used in comparisons of model-generated predictions to empirical data. In particular we shall state and justify the criterion used to determine the existence or absence of a discernible evoked response (termed *firing/no-firing*) in EEG response data, over a stimulus cycle. In view of the inherent stochasticity of lumped characteristics of neural mass activity, evaluation of model predictions will be based on

probabilistic criteria. In Sect. 5.2, we compare ground truth data, as defined in Sect. 5.1, to features of the response generated by the fully PM and to PM1. To put these results in yet another perspective, comparisons are also made to feature extraction by BFM, serving as alternative data filters. Finally, leave-one-out cross-validation (LOOCV), presented in Sect. 5.3, tested model-generated predictions in each of the twelve data sets, using models identified by the data in the remaining sets.

5.1 Performance criteria

Model performance was evaluated in terms of the match between model-generated response and empirical data. specific criteria include the binary prediction of firing or no-firing, and the quantitative predictions of the following selected set of characteristic features:

Timing features	Shape features
P1 peak time (pt)	P1 peak amplitude (pp)
N1 peak time (nt)	N1 peak amplitude (np)
P2 start time (p2t)	P2 area (p2a)

The timing features, in this list, measure the latency of the response. Shape features are indicators of response intensity. The parenthesized symbols are used as abbreviated references in the visualization of our results, in Figs. 12, 13, and 14.

Considering the quantitative criteria, the convention, introduced in Sect. 4.2, is that the *ground truth* evaluation of the P1–N1–P2 response waveform is consistently *defined* as the best fit of the data signal by a piece-wise polynomial curve of the form (3). The extraction of the six features listed above from the piece-wise polynomial is straightforward.

A meaningful, quantitative means to determine the presence or absence of an evoked response during a particular stimulus cycle, is needed in order to enable evaluating model performance in making such predictions. The inherent ambiguity of response detection has been elaborated in Sect. 2.2. As a starting point we thus recognize that what we look for is a threshold classification criterion, which we want to be both objective and robust.

The P1–N1 peak-to-peak amplitude, derived from the piece-wise polynomial fit, was chosen as a distinctive quantity, to be used in that a classification. This choice is supported by the empirical probability density function (PDF) of this feature. The empirical PDF is defined by the smoothed and normalized histogram of each of the twelve histograms. (An FIR filter, defined by a Gaussian kernel with $\sigma = 0.03$, was used to smooth the empirical histogram.) The conspicuous bi-modal nature of the PDF, seen in the example, in Fig. 10, supports our use of the value that separates the two lobes as a natural threshold for the distinction between fire and no-fire cases.

5.2 Model performance

To determine the accuracy of firing detection we computed the *precision* (P , complement of *specificity*), the *recall* (R , the *sensitivity*), and the *Fscore* (F), which are common measures in the information retrieval community (Manning et al. 2008, p. 155). The precision, P , measures the probability that the prediction of firing by the model is correct. The recall, R , measures the probability that the model predicts firing, when firing is found in the data. Formalizing these probabilistic indicators, we introduce the following quantities–

True positives (TP): The number of response cycles in which a response was found in the data and predicted by a model.

False positives (FP): The number of response cycles in which no response was found in the data but one was predicted by a model.

False negatives (FN): The number of response cycles in which response was found in the data but was not predicted by a model.

In these terms, the precision, recall, and *F* score, are defined as–

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F = \frac{2}{1/P+1/R}. \quad (5)$$

In a perfect model, both *P* and *R* attain their maximal value of 1.0, and so does the *F* score.

Table 2 reports the *F* scores for firing predictions. Scores are given separately for each data set, for each of the BFM, PM1, and PM, and for both the nonlinear and linear approximations of the Riera model. The table also includes the mean and standard deviation of the twelve scores in each column. As seen, data set *F* scores, their mean values and statistical spread are similar for the linear and the nonlinear models, and when the PM and PM1 scores are compared. As expected, BFM scores are higher and less spread, reflecting the fact that these scores measure the agreement of two curve fitting methods [the model vs. the piece-wise polynomial fit (3)], rather than the predictive power of a model.

The similarity seen in Table 2, between the performance of the linear and the nonlinear models, prevails throughout the selected set of criteria, and across all data sets. To simplify graphic presentations of the quality of feature predictions, these results are presented first, in Fig. 11, only for the linear model variants, and for a single, representative data set (4L). Ground truth values continue to be defined by the best piecewise polynomial fit (3) of EEG data. Comparisons model predictions to ground truth values are shown for the BFM, PM1 and PM. The figure includes the mean of the actual and predicted values of each feature, and the 10–90% spread of these values, across the data set. It illustrates the fact that all the features are qualitatively captured by the predictive models (PM1 and PM), including key response trends, such as a decrease in the peak values of P1 and N1, and in the P2 area, and greater latency of the P1 and N1 peaks, as the ISI shrinks. The BFM, PM1 and PM predictions of the (inherently ambiguous) P2 start time are very similar, and display the largest joint discrepancy from the “ground truth” value, as defined by the piece-wise polynomial fit. We argue that this particular spread reflects the difference between the family of curves defined by the piece-wise polynomial curves (3), and those that can be generated by the Riera model, rather than an objective gap between model predictions and an objective reality. That is, the P2 start time results illustrate the inherent limitation of the ground truth definition we use. Had we leveraged the latitude in defining ground truth values, and used the family of feasible P1–N1–P2 waveforms generated by the (simplified) Riera model instead of (3), then the BFM response would have been identified as ground truth. In that case, the mean PM1 and PM predictions of the P2 start time would have been deemed near perfect.

A summative representation of each of the performance criteria, across all data sets and models types, is provided in Fig. 12, where BFM, PM1, and PM are based on the linear Riera model and in a counterpart presentation, in Fig. 13, where the nonlinear formulation is used. The six features are divided, in these figures, between timing features (left panels), and shape features (right panels). Relative errors are naturally used in the evaluation of the prediction of shape parameters, and these results are represented by their mean values. Absolute errors (in ms) are used for time parameters, and results include both mean values

and the corresponding standard deviations. The figures clearly show that the PMs perform comparably well on all data sets.

Once again, the models perform most poorly in capturing the P2 start time, and thus, the P2 area. As explained earlier, these poor indicators can be equally viewed as reflecting the difference in the respective definitions of these, inherently ambiguous features, by two, synthetic waveform parameterizations, i.e., by the piece-wise polynomial curves (3) and by the family of response curves that can be generated by the Riera model.

5.3 Cross-validation

Model performance was also tested, using the *LOOCV* approach. These tests were applied to PM performance, as determined by the fourteen parameters of the functional dependencies (4) of model parameters on s . Specifically—

1. Having chosen a data set, the optimal values of each of the fourteen s -mapping parameters, from the remaining eleven data sets, were averaged. These average values were used in generating the model response for the chosen data set.
2. Model errors were compared with those generated by using the optimal parameters, for the same data set.

This procedure was performed for each of the 12 data sets. Figure 14 compares the error using the cross-validation approach with the error obtained by using the optimized model, for each data set. These results are shown only for the linear Riera PM. As seen, errors using the cross-validation approach are comparable to those achievable with models optimized for each of the data sets.

6 Discussion

This article highlights the issue of habituation in EEG-evoked response and proposes a simple mathematical formulation for modeling this phenomenon. We use this section to recap few key technical points concerning problem identification, formulation, and the proposed solution path.

6.1 The phenomena: evoked response and habituation

Evoked response is characterized and analyzed in terms of the distinctive P1–N1–P2 waveforms. Figure 1 illustrates the challenge to any quantitative analysis of these waveforms, in general, and of their habituation, in particular. The P1–N1–P2 waveform cannot be reliably identified in the response to a single stimulus (Fig. 1, top right panel, dashed curve). It is possible to isolate and quantify P1–N1–P2 features only in the ensemble mean of multiple experiments. Even then, these waveforms maintains only a subtle and often elusive presence in the EEG signal, and as habituation builds up, so does the ambiguity of their exact definition.

Considering habituation at the qualitative level and the examples in Fig. 1, we observed that the evoked response is increasingly attenuated at shorter ISI's, and as the cumulative intensity of the response to a succession of previous stimuli is higher. Habituation thus seems to reflect the balance of two competing temporal processes. An integrative process that sums the cumulative effect of past responses, and a forgetting process, that diminishes the effect of past responses as the elapsed time grows. The habituation control mechanism proposed in this article formalizes this intuitive and qualitative description, by the first-order, nonlinear, low-pass filter (2). The stability of (2), hence the decay of the auxiliary state $s(t)$, represents the forgetting processes. The growth of $s(t)$ due to a nonzero input represents the cumulative aspect of past responses. Clearly, such a representation cannot,

and is not intended to, capture the intricate neuronal processes of cross-membrane ion balance during polarization and depolarization. Rather, like other mass neural model components, it is used to capture the dynamics of changes in the averaged neural mass response to successive stimuli.

A quantitative study of the success of the proposed model necessitates two critical ingredients: (i) An objective means to reliably isolate the P1–N1–P2 waveform by which we characterize the ensemble mean response, and to evaluate features used to measure the precision of model predictions, and (ii) an objective labeling of cycles in which habituation makes the chosen filtering scheme unreliable. These issues have been addressed in Sects. 4.2 and 4.3, and will be revisited in Sects. 6.2 and 6.3, below.

Pertinent to the development of these tools is the very distinction between the evoked response and background neuronal activity. In line with the observations by David et al. (2006), the evidence in Figs. 1 and 2 supports the view that this distinction reflects a continuous quantitative shift, rather than a qualitative difference. Arguably, a qualitative distinction between background activity and the P1–N1–P2 evoked response, and the association of background activity and the evoked response with distinct neuronal populations, would imply that a habituated evoked response be associated with EEG measurements at the characteristic background level. Instead, Fig. 2 shows that the power spectrum amplitude of the habituated response over the P1–N1–P2 frequency band is unambiguously lower than the EEG power spectrum without stimulus. Considering the time traces, in Fig. 1, we see that the P1–N1–P2 waveform is typically followed by a succession of later peaks and troughs, within a similar frequency range, whose coherence and amplitude gradually wear off. These observations suggest that the same type of neuronal activity is involved in the evoked response, the phenomenon of habituation, and what we term *background activity*. A stimulus leads to elevated synchronization, hence to higher coherence and amplitude of the first few half-periods of the EEG signal, which we term P1, N1 and P2. The gradual reduction of that synchronization leads to the gradual return to the unstimulated ensemble behavior. Habituation leads to an attenuation of activity in the participating neuronal populations, hence to a level of activity that is lower than background.

6.2 Isolating the evoked response from the EEG signal

The use of a parameterized curve fitting is a natural approach for the isolation of the P1–N1–P2 waveform. The requirements from such a parameterization include: (i) Being generic, to avoid a bias toward any stipulated model structure; (ii) a focus on the most salient features of the targeted phenomenology, to avoid superfluous over-fitting of background signal idiosyncrasies, (iii) minimum feasible complexity, to insure numerical robustness of parameter estimation and support the objective of avoiding over-fitting. We believe that the constrained least-squares polynomial fit (3), in Sect. 4.2, strikes a balance among these objectives. All references to the evoked response hence, including the labeling of discernible and indiscernible response, right below, and the definition of what ground truth means, are based on these choices and definitions. That is, we continue to bear in mind the intrinsic ambiguity regarding the meaning of a specific features, and that the larger that ambiguity is, so is the ambiguity concerning the value of performance evaluation, for the prediction of that particular feature, whether in comparison to polynomial fit, or otherwise. In the investigated data, this comment applies, in particular, to features pertinent to the P2 portion of the evoked response.

6.3 Labeling discernible and indiscernible-evoked response

Since habituation occurs over a continuous range, both the conceptual and the quantitative distinctions between discernible and indiscernible response are necessarily artificial.

Referring to Fig. 10 and to the description in Sect. 5.1, our choice of a threshold labeling method is motivated by the strongly bimodal nature of the distribution of the P1–N1, peak-to-peak amplitudes of the evoked response. The bimodal distribution enables a clear and objective definition of a threshold for the parameterization of response cycles by the P1–N1, peak-to-peak amplitude, designating cycles with lower parameter values as fully habituated.

6.4 Simplifications of the Riera model

Riera's neural mass model aims at a balance between physiological fidelity and simplicity. The dynamical system is defined in terms of a few, lumped, physiological features of three neurons, representing distinct neural populations. This way the model maintains its physiological foundations, even as it reaches for an extreme level of simplification, when compared with detailed neural models. In contrast, the Gaussian descriptions of the input currents, I_i^+ , $i = 1, 2, 3$, provide purely phenomenological parameterizations of neural activity that are not intended to be included in the model's focal area. The proposed three state nonlinear and two state linear variants of Riera's model were derived by further restricting the level of detail resolved. As such, these models belong to the same class as the original model, and represent merely a moderate shift in the trade off between simplicity and physiological fidelity. In fact, as demonstrated in the Appendix, the fidelity loss due to this simplification is negligible.

Considering the block diagrams in Figs. 3 and 4, the simplifications considered here leave only the linear block in the dynamical system. The justification for severing the feedback link is in the observation, established in the Appendix, is that $f(x_2) \approx 0$, persistently, when model parameters are chosen within a physiologically meaningful range. This includes Riera's choice of parameters and the choices made in this article, including parameter variations due to habituation. The conceptual status of our parameterization of I_3^+ by a quadratic waveform (Eq. 18 in the Appendix) is identical to the parameterization of I_3^+ by a Gaussian in the original model. That is, it represents a simplified phenomenological representation of neural activity not resolved by the dynamical system. The ultimate justification of the use of the simplified model is in the nearly indistinguishable response of the two models.

6.5 Additional simplifications: the model as a parameterized curve fit

Using only the parameterized impulse response, at the level of simplification of the neural mass model, one may well argue that the role of the model is merely that of a physiologically motivated parameterized curve fitting tool, targeting a very narrow range of phenomena. This view is further supported if one subscribes to the opinion, discussed in Sect. 6.1, that the very concept of an evoked response is a simplification of a more complex reality. The question therefore arises whether an even simpler, purely phenomenological, parameterization would not suffice in an investigation of the evoked response. The piece-wise polynomial curve fit that we use as a preliminary filtering tool is a case in point. Indeed, we have explored that possibility. As seen in the previous section, the match between the physiologically motivated and the polynomial fit is high. In the absence of an unambiguous ground truth, the difference may attest more to the effect of using different filtering methods than to the advantage of one or the other. Sensitivity analysis reveals that the number of free parameters required by a piece-wise polynomial fit is actually identical to that required by the Riera model. In the long run, however, the potential advantage of a physiologically motivated model is in its possible use to predict neural activity effects of abnormal changes in physiological parameters.

6.6 Feature estimates versus curve fit evaluation

Our precision criteria for model-generated predictions concern six numerical features of the evoked response, i.e., the timing features pt , nt , and $p2t$, and the shape features pp , np , and $p2a$. We are cognizant of the very common use of curve fit quality rather than feature predictions. Indeed, that has been our choice in the computations leading to the extraction of the selected features from data streams, as well as the criterion used in related studies cited earlier, such as David and Friston (2003); Kiebel et al. (2006). Yet since the model and its predicted response are determined by a few—indeed only six—varying parameters, performance evaluation is necessarily restricted to a family of curves parameterized by this number of degrees of freedom, restricting the true meaning of evaluation criteria accordingly. With this in mind, matching feature values, such as those we use, is a curve fit measure, much like L_p norms. Our chosen features have the advantage of being motivated by physiologically pertinent quantities, such as latency, intensity, etc.

7 Concluding remarks

The article introduces a general dynamic control mechanism that enables a neural mass model of the evoked EEG response to account for habituation—the physiological adaptation to variations in the inter-stimulus relaxation period. The proposed control structures requires the augmentation of the original model by a first (hence, least) order auxiliary ODE. The auxiliary state represents the slow dynamics of post-stimulus relaxation. The proposed mechanism was applied to the physiologically motivated neural mass model of Riera et al. (2006, 2007). The model was validated using twelve data sets, containing the evoked response to over 1,700 stimulus trains, applied to the medial nerve, each lasting 4 s, with inter-stimulus periods ranging between 125 ms and 1 s (1–8 Hz). Validation included both PM1 and long term predictions (PM), with comparable, good performance.

Careful structural analysis, carried under broad, physiologically pertinent conditions, enabled the simplification of Riera's original, nonlinear, nine state model, to a linear, two state system, with a nearly identical output. In all, these results thus demonstrate how the combination of a simple augmentation and a systematic examination of model structure can lead to simplified (reduced) models that can be efficiently optimized, enabling the correct representation of habituation-induced changes in key features of the evoked neuronal response.

In a subsequent study, currently underway, the model we have presented here is an enabler for an investigation of the relations between changes in neural mass EEG response to various stimuli, and the resulting changes in hemodynamic activity in the brain. That investigation links the saturation control mechanism, which encodes the post-stimulus relaxation, to a neurovascular coupling model that modifies the inputs to a hemodynamic model, replicating habituation-induced variations in the hemodynamic response.

These advances can be viewed within a range of low order dynamic model development perspectives. The authors backgrounds, in areas ranging from 4D medical imaging to fluid flow systems, bring forth a set of intuitions and tools, rooted in the realm of reduced order models on nonlinear inertial manifolds, and the coupling of fast and slow dynamics in the representation of bursty and modulated oscillations. The inherently stochastic nature of the problem, and of the performance measures and cross-validation tests we used, fits comfortably within the Bayesian framework of dynamic causal modeling (Friston et al. 2003), which uses neural mass models to predict or explain both event-related and steady-state neurophysiological responses. In that context, our PM becomes a generative model, our bounds or constraints on parameters are formulated in terms of (continuous) shrinkage priors, and our objective is to find a parsimonious model that explains data features

accurately. Our evaluation and cross-validation tests have the same role as that of evidence based Bayesian model comparisons in adjudicating between different hypotheses about neural mass models generating data. Following along this path, the parameterization and structure of adaptation dynamics can be optimized in terms of log odds ratios or Bayes factors. Exploring this important and pragmatic application is another important direction of future study.

Acknowledgments

The authors would like to thank David Boas for his helpful suggestions, and Ilkka Nissila and Weicheng Wu, for collecting the data sets. The authors would also like to thank the reviewers for their suggestions on the context of our study with respect to the mechanism of habituation in neuronal responses, and the tools used for model estimation. Their suggestions have been very helpful in driving and guiding us to prepare what we believe is a better presentation of our study. The study of the first, second, and last authors was supported in part by the Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems (Gordon-CenSSIS) under the Engineering Research Centers Program of the National Science Foundation (NSF) (award EEC-9986821), and by RO1EB001954. The authors declare that they have no conflict of interest. The study was approved by the Animal Care Committee of Massachusetts General Hospital.

References

- Andry ML, Luttges MW. Neural habituation in garter snakes. *Physiol Behav.* 1972; 9:107–111. [PubMed: 5080147]
- Bower, JM.; Beeman, D. The book of GENESIS: exploring realistic neural models with the GEneral NEural SIMulation System. Springer; New York: 1998.
- Bristol AS, Carew TJ. Differential role of inhibition in habituation of two independent afferent pathways to a common motor output. *Learn Mem.* 2005; 12:52–60. [PubMed: 15647595]
- Coombes S, Venkov N, Shiao L, Bojak I, Liley DTJ, Laing C. Modeling electrocortical activity through improved local approximations of integral neural field equations. *Phys Rev E.* 2007; 76:051901–1051908.
- Daunizeau J, Grova C, Marrelec G, Mattout J, Jbabdi S, Plgrini-Issac M, Lina J, Benali H. Symmetrical event-related eeg/fmri information fusion in a variational Bayesian framework. *Neuroimage.* 2007; 3:69–87. [PubMed: 17408972]
- David O, Friston KJ. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage.* 2003; 20:1743–1755. [PubMed: 14642484]
- David O, Kilner JM, Friston KJ. Mechanisms of evoked and induced responses in MEG/EEG. *Neuroimage.* 2006; 31(4):1580–1591.10.1016/j.neuroimage.2006.02.034 [PubMed: 16632378]
- de Schutter E, Bower JM. An active membrane model of the cerebellar purkinje cell. I. Simulation of current clamps in slice. *J Neurophysiol.* 1994a; 71:375–400. [PubMed: 7512629]
- de Schutter E, Bower JM. An active membrane model of the cerebellar purkinje cell. II. Simulation of synaptic responses. *J Neurophysiol.* 1994b; 71:401–419. [PubMed: 8158238]
- Franceschini MA, Nissil I, Wu W, Diamond SG, Bonmassar G, Boas D. Coupling between somatosensory evoked potentials and hemodynamic response in the rat. *Neuroimage.* 2008; 41(2): 189–203. [PubMed: 18420425]
- Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage.* 2003; 19:1273–1302. [PubMed: 12948688]
- Garrido MI, Kilner JM, Kiebel SJ, Stephan KE, Baldeweg T, Friston KJ. Repetition suppression and plasticity in the human brain. *Neuroimage.* 2009; 48(1):269–279. [PubMed: 19540921]
- Ghigliazza RM, Holmes P. Minimal models of bursting neurons: How multiple currents, conductances, and timescales affect bifurcation diagrams. *SIAM J Appl Dyn Syst.* 2004; 3:636–670.
- Gorban, AN.; Kégl, B.; Zinovyev, AY. Principal manifolds for data visualization and dimension reduction. Springer; Berlin: 2007.
- Hall R. Habituation of evoked potentials in the rat under conditions of behavioral control. *Electroencephalogr Clin Neurophysiol.* 1968; 24(2):155–165. [PubMed: 4170481]

- Horwitz B, Poeppel D. How can eeg/meg and fmri/pet data be combined. *Hum Brain Mapp.* 2002; 17(1):1–3. [PubMed: 12203682]
- Jansen BH, Rit VG. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern.* 1995; 73:357–366. [PubMed: 7578475]
- Kiebel SJ, David O, Friston KJ. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *Neuroimage.* 2006; 30(4):1273–1284.10.1016/j.neuroimage.2005.12.055 [PubMed: 16490364]
- Lahaye, PJ.; Baillet, S.; Poline, JB.; Garnero, L. Fusion of simultaneous fmri/eeg data based on the electro-metabolic coupling. 2nd Proc IEEE, ISBI; 2004. p. 864-867.
- Liley DTJ, Cadusch PJ, Dafilis MP. A spatially continuous mean field theory of electrocortical activity. *Comput Neural Syst.* 2002; 13:67–113.
- Manning, CD.; Raghavan, P.; Schtze, H. An introduction to information retrieval. Cambridge University press; Cambridge: 2008.
- May ML, Hoy RR. Habituation of the ultrasound-induced acoustic startle response in flying crickets. *J Exp Biol.* 1991; 159:489–499. [PubMed: 1940771]
- Mosher JJ, Leahy R. Recursive music: a framework for EEG and MEG source localization. *IEEE Trans Biomed Eng.* 1998; 45(11):1342. [PubMed: 9805833]
- Nelles, O. Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Springer; Heidelberg: 2001.
- Noguchi Y, Inui K, Kakigi R. Temporal dynamics of neural adaptation effect in the human visual ventral stream. *J Neurosci.* 2004; 24(28):6283–6290. [PubMed: 15254083]
- Purves, D. Neuroscience. Sinauer; Maryland: 2004.
- Rankin CH, Abrams T, Barry RJ, Bhatnagar S, Clayton D, Colombo J, Coppola G, Geyer MA, Glanzman DL, Marsland S, McSweeney F, Wilson DA, Wu CF, Thompson RF. Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiol Learn Mem.* 2009; 92(2):135–138. [PubMed: 18854219]
- Riera JJ, Xiaohong W, Jimenez JC, Kawashima R. Nonlinear local electrovascular coupling. I: a theoretical model. *Hum Brain Mapp.* 2006; 27(11):896–914. [PubMed: 16729288]
- Riera JJ, Xiaohong W, Jimenez JC, Kawashima R, Ozaki T. Nonlinear local electrovascular coupling. II: from data to neuronal masses. *Hum Brain Mapp.* 2007; 28(4):335–354. [PubMed: 16933303]
- Shea SD, Katz LC, Mooney R. Noradrenergic induction of odor-specific neural habituation and olfactory memories. *J Neuroscience.* 2008; 28:10711–10719.
- Shilnikov AL, Rulkov NF. Subthreshold oscillations in a map-based neuron model. *Phys Lett A.* 2004; 328:177–184.
- Sornmo, L.; Laguna, P. Bioelectric signal processing in cardiac and neurological applications. Elsevier Academic press; Boston: 2005.
- Stark, H.; Woods, J. Probability and random processes with applications to signal processing. Prentice hall; Englewood Cliffs: 2001.
- Stoica, P.; Moses, RL. Introduction to spectral analysis. Prentice hall; Upper Saddle River: 1997.
- Stuart J. Nonlinear stability theory. *Ann Rev Fluid Mech.* 1971; 3:347–370.
- Tadmor, G.; Lehmann, O.; Noack, BR.; Morzyński, M. Galerkin models enhancements for flow control. In: Noack, BR.; Morzyński, M.; Tadmor, G., editors. Reduced-order modelling for flow control, CISM courses and lectures. Vol. 528. Springer; Vienna: 2011.
- Thompson R, Spencer W. Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychol Rev.* 1966; 73:16–43. [PubMed: 5324565]
- Tully T, Koss S. Habituation of the jump reflex to olfactory cues in normal and mutant *Drosophila*. *Soc Neurosci Abstr.* 1992; 8:942.
- Zuker RS. Crayfish escape behavior and central synapses. II. physiological mechanisms underlying behavioral habituation. *J Neurophysiol.* 1972; 35:621–637. [PubMed: 5054507]

Appendices

A Data preprocessing

As noted in Sect. 2.2, preprocessing of raw experimental data is required to enable the isolation of evoked response waveforms. Tasks include the removal of a slow baseline drift, of an interference caused by the 60-Hz AC power and by the cardiac cycle, at 3–5 Hz. In addition, the presence of EEG signal oscillations at frequencies that are higher than those of the targeted waveform, and the stochastic fluctuations of individual response cycles, illustrated by the dashed/blue curve in the top right panel, in Fig. 1, need to be mitigated. Elements of the procedures used to achieve this goal are reviewed below.

The slow baseline drift was corrected by subtracting a piece-wise linear curve from the original EEG signal. Successive 4-s segments are each evaluated as the average values of the symmetric 4-s interval, about the respective vertex. (This applies to each of the twenty, multi-interval runs, in each data set, including stimulus and relaxation intervals alike.) The 60-Hz power supply interference was removed by a simple notch filter. Cardiac cycle interference was removed by a linear regression model, described in Franceschini et al. (2008).

Stochastic variations in the 4-s response intervals are averaged out in the corrected signal. Averaging is carried separately in each data set, and for each stimulus frequency. Thus, each stimulus frequency in each of the twelve data set is represented by a single 4-s stretch of averaged response to a stimulus train.

Removal of the higher frequency oscillations in the evoked response is challenged by the fact that the wavelength of fluctuations found along the N1 component, and even more so, along the P2 component (each lasting ~30–35 ms), is comparable to the generic width of the P1 phase of the evoked response, which is ~10 ms (≈ 100 Hz). A facilitating factor is found in the fact that the disturbance is sufficiently small during the P1 phase (presumably, a manifestation of the strong early synchronization) to be able to identify the beginning and end times of that phase, as described following (6), below. Exploiting this fact, our method to remove the higher frequency EEG component can be formalized as follows:

$$\hat{y}(t) = w(t) \cdot y(t) + (1 - w(t)) \cdot (F * y)(t), \quad (6)$$

where y and $\hat{y}$ are the original and the filtered EEG signal, w is a trapezoidal boxcar weight function, supported over the identified time interval of the P1 phase and F is a Hanning windowed FIR low-pass filter of order 20, with a cutoff set to 50 Hz. In these notations, the determination of the start and end times of P1, $t_b \in [t_s + 8, t_s + 12]$ ms and $t_e \in [t_s + 18, t_s + 24]$ ms, which define the support interval of w , was done by an exhaustive, sample-wise, two-dimensional bounded search, minimizing the relative L_2 error norm $\|\hat{y} - y\|_2 / \|y\|_2$. This procedure is applied to the entire 4 s of each response stretch, attenuating the signal power above 50 Hz, with the exception of the short intervals supporting P1 phases. To maintain a meaningful comparison, in Fig. 2, the Hanning FIR filter was applied to the entire background signal in the data used to produce the background spectrum.

In closing, it is important to note the possible impact—or lack thereof—of the selected preprocessing steps on the findings of this study. The removal of the 60-Hz power frequency with a narrow, linear notch filter, can effect only a narrow, well defined frequency band, with no qualitative effects on the P1 phase, at a higher band, or the N1 and P2 phases, at a lower band. With 4-s segments, and with vertices scattered randomly, relative to the start times of the studied stimulus trains, the impact of baseline removal on the spectra plots, in Fig. 2, is essentially limited to the [0, 0.25] Hz range, and is therefore inconsequential. The

impact of cardiac interference removal is likewise limited to a frequency range that is substantially lower than that of the phenomena shown here. The nonlinearity of the procedure to remove higher frequency components of the EEG signal, the last to be described, makes it a “natural suspect.” As it turns out, the effect of skipping this procedure is essentially limited to the aesthetics (and clarity) of waveform plots, such as in Figs. 1, 6, and 9. Applying the piece-wise polynomial fit (3) and BFM optimization to the data, prior to this step, has no significant impact on the quantitative and qualitative model evaluation results. The obvious quantitative impact of that step in Fig. 2, lowering amplitudes of all three spectra at frequencies above 50 Hz, does not change the qualitative properties highlighted in the discussion of that figure, and the intuition drawn from these properties.

B Order reduction in the Riera model

The analysis in this article is based on reduced order versions of Riera’s original, nine state, nonlinear model. The purpose of this appendix is to show that the nine state Riera model can be reduced to a three state nonlinear model under reasonable approximations. This three state nonlinear model is further reduced to a two state linear model that was used as the primary representation in this work. We validate this model order reduction under the conditions explained in Sect. B.3, using the model parameter values given in Riera et al. (2007).

The appendix is organized as follows. The differential equations for the Riera model are reproduced for ease of reference, and rewritten in standard state-space form in Sect. B.1. Referring to the block diagram in Fig. 4, Sect. B.2 discusses a method that leads to a reduction in model order. The mathematical details of the reduction procedure and comparison of the reduced model to the original Riera model are presented in Sect. B.3. To simplify the presentation, the nomenclature of the appendix is self-contained and may be different from simplified notations in the body of the article; example include Riera’s original notations, in (7), and their simplification, in (8) and beyond.

B.1 Riera model equations

The differential equations of the Riera model are reproduced below.

$$\frac{dV_{IN}^T}{dt} = \frac{1}{\tau_m} \left[-V_{IN}^T(t) + I_3^+(t) R_m^0 \right] \quad (7a)$$

$$\frac{dV_{IN}^F}{dt} = \frac{1}{\tau_m} \left[-V_{IN}^F(t) + I^+(t - \tau_{PC}) R_m^0 \right] \quad (7b)$$

$$\frac{dV_{PC}}{dt} = \frac{1}{\tau_m} \left[-\left(\alpha_0 + \sum_{k=1}^2 \frac{1}{\beta_k} \right) V_{PC}(t) - \frac{\Omega(t)}{\beta_1 \beta_2} - R_m I^-(t - \tau_{IN}) + \sum_{k=1}^2 \left(\frac{R_m v_k(t)}{(R_i^k + R_e^k)} - \frac{v_-(t)}{\beta_k} \right) \right] \quad (7c)$$

$$\frac{d\Omega}{dt} = \frac{1}{\tau_m} \left[-\Omega(t) + R_m \left(\sum_{k=1}^2 \frac{\beta_k (V_{PC}(t) - v_k(t))}{(R_i^k + R_e^k)} \right) + V_{PC}(t) + v_-(t) \right] \quad (7d)$$

$$\frac{dv_k}{dt} = \frac{1}{\tau_m} [-v_k(t) + R_m I_k^+(t)] \quad k=\{1, 2\} \quad (7e)$$

$$\frac{dv_-}{dt} = \frac{1}{\tau_m} [-v_-(t) + R_m I^-(t - \tau_{IN})] \quad (7f)$$

$$\begin{aligned} \frac{d\Phi}{dt} = \frac{1}{\tau_m} \left[- \left(\alpha_0 + \sum_{k=1}^2 \frac{1}{\beta_k} \right) \Phi(t) - \frac{\Theta(t)}{\beta_1 \beta_2} \right. \\ \left. + \frac{R_e^2}{(R_1^2 + R_e^2)} (R_m I^-(t - \tau_{IN}) + R_m^2 I_2^+(t)) + \frac{(R_m^1 (v_-(t) + v_2(t)) + R_m (v_2(t) - v_1(t))) R_e^2}{\prod_{k=1}^2 (R_1^k + R_e^k)} \right] \quad (7g) \end{aligned}$$

$$\frac{d\Theta}{dt} = \frac{1}{\tau_m} \left[-\Theta(t) + \left(1 + R_m \sum_{k=1}^2 \frac{1}{R_m^k} \right) \Phi(t) \right] \quad (7h)$$

$$y = \Phi. \quad (7i)$$

In the above equations, the currents $I^-(t - \tau_{PC})$, and $I^-(t - \tau_{IN})$ are given by:

$$\begin{aligned} I^+(t - \tau_{PC}) &= \alpha_{PC} f_{PC}(V_{PC}(t - \tau_{PC})) \\ I^-(t - \tau_{IN}) &= \alpha_{IN} (f_{IN}^T(V_{IN}^T(t - \tau_{IN})) + f_{IN}^R(V_{IN}^R(t - \tau_{IN}))). \end{aligned}$$

These currents are not to be confused with the input currents of the Riera model, namely $I_i^+(t)$, $i \in \{1, 2, 3\}$. The time delays τ_{PC} , and τ_{IN} are set to zero in Riera et al. (2007). We follow the same in the present analysis. The functions f_{PC} , and f_{IN} , in the equations above, are given by:

$$\begin{aligned} f_{PC}(V) &= A_I + \frac{A_u}{(1 + T e^{-\gamma_{PC}(V - V_0^{PC})})^{1/T}} \\ f_{IN}(V) &= A_I + \frac{A_u}{(1 + T e^{-\gamma_{IN}(V - V_0^{IN})})^{1/T}}. \end{aligned}$$

The equations in (7) are rewritten below for notational convenience.

$$\dot{x}_1 = -ax_1 + bI_3^+ \quad (8a)$$

$$\dot{x}_2 = -ax_2 + bf_1(x_3) \quad (8b)$$

$$\dot{x}_3 = -cx_3 - dx_4 - g_1[f(x_1) + f(x_2)] + g_2x_5 + g_3x_6 - hx_7 \quad (8c)$$

$$\dot{x}_4 = sx_3 - ax_4 - k_1x_5 - k_2x_6 + ax_7 \quad (8d)$$

$$\dot{x}_5 = -ax_5 + d_1 I_1^+ \quad (8e)$$

$$\dot{x}_6 = -ax_6 + d_1 I_2^+ \quad (8f)$$

$$\dot{x}_7 = -ax_7 + g_1 [f(x_1) + f(x_2)] \quad (8g)$$

$$\dot{x}_8 = -cx_8 - dx_9 - r_1 x_5 + r_2 x_6 + q x_7 + o[f(x_1) + f(x_2)] + p I_2^+ \quad (8h)$$

$$\dot{x}_9 = s x_8 - a x_9 \quad (8i)$$

$$y = x_8. \quad (8j)$$

In these equations,

$$f(x) = \frac{1}{(1 + T e^{-\gamma(x-x_0)})^{\frac{1}{T}}} = f_{\text{IN}}(V)$$

$$f_1(x) = \frac{w_1}{(1 + T e^{-\gamma_1(x-x_{01})})^{\frac{1}{T}}} = \alpha_{\text{PC}} f_{\text{PC}}(V).$$

$f(x)$ is a sigmoid-shaped function whose range is the interval $[0, 1]$, with $f(x_0) = 0.37$. $f_1(x)$ is also a sigmoid-shaped function whose range is the interval $[0, 0.4]$, with $f_1(x_{01}) = 0.15$.

Table 3 shows the relations between the parameters in the state-space model to the parameters of the Riera model.

B.2 An overview of the proposed model reduction

The block diagram of the Riera model connectivity in terms of the states in Eq. 8 is shown in Fig. 4. As seen in Fig. 4, if the feedback block consisting of states x_i , $i = 2, 3, 4$, is removed, then the Riera model reduces from nine to six states. The six state model consists of the feedforward subsystem (x_1), with nonlinear output $f(x_1)$, and a linear subsystem (x_5 through x_9). Justification for this removal is given below.

Different linear combinations of the states x_5 , x_6 and x_7 , drive the states x_3 and x_4 (refer to Eqs. 8c and 8d). Removal of states x_3 and x_4 enables us to define a new state $x_n = -d x_9 - r_1 x_5 + r_2 x_6 + q x_7$, thus, reducing the linear subsystem to a two state system with states x_8 and x_n . Hence, the original Riera model reduces to a three state model with x_1 forming the nonlinear subsystem, and x_n and x_8 forming the linear subsystem. The simplified waveform parameterization of the current $\tilde{I}_3^+$ in Eq. 18 substitutes the nonlinear subsystem, removing the state x_1 , and hence we obtain a linear system with two states, x_n and x_8 , and three inputs, I_1^+ , I_2^+ , and $\tilde{I}_3^+$, defined as parameterized waveforms.

B.3 Removing the feedback block

In Fig. 4, we notice that the feedback block can be removed if $f(x_2) \approx 0$. We also notice that $f_1(x_3)$ drives (8b). Hence, an analysis of the bound on the solution for the state x_3 would help in analyzing the effect of $f(x_2)$ on the system, over the time duration from the arrival of

a stimulus till the end of the P1–N1–P2 waveform. Below, we follow a series of steps to derive the equation for the upper bound of the state x_3 .

Step 1: explicit solutions for the states x_5 and x_6 —We find explicit solutions to Eqs. 8e, and 8f when they are driven by Gaussian functions as inputs. Substituting the Gaussian functions for I_1^+ and I_2^+ explicitly into (8e), and (8f), we have:

$$\dot{x}_5 = -ax_5 + d_1\alpha_1 e^{-s_2^2(t-\tau_2)^2} \quad (9a)$$

$$\dot{x}_6 = -ax_6 + d_1\alpha_2 e^{-s_3^2(t-\tau_3)^2}. \quad (9b)$$

We continue the present analysis by using the nominal values of the input current parameters, given in Table 4.

A statistical analysis over a range of parameter values, is presented in Sect. B.3.4.

The solution for Eqs. 9a and 9b, are given by:

$$x_5(t) = A_1 e^{-a(t-\tau_2)} \left[1 + \operatorname{erf} \left(s_2(t-\tau_2) - \frac{a}{2s_2} \right) \right] \quad (10a)$$

$$x_6(t) = A_2 e^{-a(t-\tau_2)} \left[1 + \operatorname{erf} \left(s_3(t-\tau_3) - \frac{a}{2s_3} \right) \right], \quad (10b)$$

where

$$A_1 = \frac{\sqrt{\pi}}{2} \frac{d_1\alpha_1 e^{a^2/4s_2^2}}{s_2} = 0.52,$$

and

$$A_2 = \frac{\sqrt{\pi}}{2} \frac{d_1\alpha_2 e^{a^2/4s_3^2}}{s_3} = 0.48.$$

In the Eq. 10, the function erf is the standard error function, which is defined as the integral of the Gaussian function (see, Stark and Woods 2001). It is a sigmoid-shaped function whose range is the interval $[-1, +1]$, with a center at zero.

We also define the following upper bounds ($x_5^u(t)$ and $x_6^u(t)$) on the solution of the states x_5 and x_6 :

$$x_5^u(t) = a_1 A_1 e^{-a(t-\tau_2)} \quad a_1 = \begin{cases} 1 & 0 \leq t < \tau_2 \\ 2 & \tau_2 \leq t \leq \tau_e \end{cases} \quad (11a)$$

$$x_6^u(t) = a_2 A_2 e^{-a(t-\tau_3)}, \quad a_2 = \begin{cases} 1 & 0 \leq t < \tau_3 \\ 2 & \tau_3 \leq t \leq \tau_e \end{cases}, \quad (11b)$$

where τ_2 , τ_3 , and τ_e divide the time interval of the P1–N1–P2 waveform into consecutive segments, as will be described below in *Step 3*.

Step 2: decoupling the states x_3 and x_4 —The state x_3 is almost decoupled from the state x_4 , which is revealed by diagonalizing their joint evolution equation

$$\begin{bmatrix} \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} -c & -d \\ s & -a \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} g_2 & g_3 & -h & -g_1 \\ -k_1 & -k_2 & a & 0 \end{bmatrix} \begin{bmatrix} x_5 \\ x_6 \\ x_7 \\ f(x_1) + f(x_2) \end{bmatrix}. \quad (12)$$

Equation 12 is of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

This equation is diagonalized by the change of state $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$, where $\mathbf{V}$ is the eigenvector matrix of $\mathbf{A}$. For the values in Table 3, $\mathbf{V}$ turns out to be

$$\mathbf{V} = \begin{bmatrix} -0.35 & 0.008 \\ 0.94 & -1 \end{bmatrix}.$$

Clearly $\mathbf{V}$ can be approximated by

$$\tilde{\mathbf{V}} = \begin{bmatrix} -0.35 & 0 \\ 0.94 & -1 \end{bmatrix}.$$

A straightforward substitution of $\mathbf{x} \approx \tilde{\mathbf{V}}\tilde{\mathbf{x}}$ in (12), yields

$$\dot{x}_3 = -\tilde{c}x_3 + \tilde{g}_2x_5 + \tilde{g}_3x_6 - \tilde{h}x_7 - \tilde{g}_1[f(x_1) + f(x_2)], \quad (13)$$

where $\tilde{c} = 388$, $\tilde{g}_1 = 43$, $\tilde{g}_2 = 292$, $\tilde{g}_3 = 28$, and $\tilde{h} = 29$. Note that these values are not much different from the original coefficient values, namely c , g_i for $i = 1, 2, 3$ and h in the Eq. 8c (see Table 3).

Step 3: simplification of the Eq. 13—In Eq. 13, the contributions from the states x_7 , x_1 , and x_2 are nonpositive, because negative coefficients multiply nonnegative functions. Only positive values of x_3 contribute to $f_1(x_3)$, and the values of x_3 will only increase if we ignore negative input terms in Eq. 13. Hence, we create an upper bound on the Eq. 13 by removing the contributions of the states x_7 , x_1 , and x_2 .

$$\dot{x}_3^u = -\tilde{c}x_3^u + \tilde{g}_2x_5 + \tilde{g}_3x_6, \quad (14)$$

Where x_3^u is an upper bound on x_3 .

We use the upper bound on the state x_3 , namely x_3^u , for our analysis from this point onward. We divide the analysis of the equations into the following three phases, which roughly correspond to the time intervals of P1, N1, and P2:

Phase 1: $t \in [0, \tau_2]$; this phase roughly corresponds to P1 being active.

Phase 2: $t \in [\tau_2, \tau_3]$; this phase roughly corresponds to N1 being active.

Phase 3: $t \in [\tau_3, \tau_e]$; the time interval during which P2 is active.

The time $\tau_e = 125$ ms, which is considered to be the end of the P1–N1–P2 waveform. For the analysis, we consider a value below 10^{-3} as being negligible, i.e., ≈ 0 .

B.3.1 Phase 1: $t \in [0, \tau_2]$

During this phase, the solution of the states x_5 and x_6 are increasing functions. Here $x_5 < 0.52$ for $t < \tau_2$, and $x_6 \approx 0$. Hence, we can ignore x_6 , and replace x_5 by x_5^u (Eq. 11a with $a_1 = 1$) in Eq. 14, leading to:

$$\dot{x}_3^u = -\tilde{c}x_3^u + \tilde{g}_2x_5^u. \quad (15)$$

Solving Eq. 15, we get

$$\begin{aligned} x_3^u &= \frac{\tilde{g}_2A_1}{\tilde{c}-a}e^{-a(t-\tau_2)}(1-e^{-(\tilde{c}-a)t}) \\ &= 0.45e^{-a(t-\tau_2)}(1-e^{-(\tilde{c}-a)t}) \end{aligned} \quad (16)$$

Therefore, $x_3^u \leq 0.45$, which gives $f_1(x_3^u) \leq f_1(0.45) = 0.037$. Substituting $f_1(x_3^u)$ in Eq. 8b, we get:

$$\dot{x}_2^u = -ax_2^u + bf_1(x_3^u), \quad (17)$$

Where $x_2^u \leq \frac{b}{a}f_1(0.45) = 0.15$. Subsequently $f(x_2^u) \leq f(0.15) \approx 0$. Thus, the feedback block may be removed during this phase.

For the discussion of phases 2 and 3, we proceed by solving Eqs. 14 and 17 for the states x_3^u and x_2^u , using the nominal parameter values. Figure 15 shows the plots of the states x_5 and x_6 on the upper left panel. The plot of the states x_3^u and x_2^u are on the upper right panel. $f_1(x_3^u)$ and $f(x_2^u)$ are plotted in the bottom left and right panels respectively.

B.3.2 Phase 2: $t \in [\tau_2, \tau_3]$

From the numerical computation shown in Fig. 15, $x_3^u \leq 0.55$, $f_1(x_3^u) \leq 0.10$, $x_2^u \leq 0.43$, and subsequently $f(x_2^u) \leq 0.019$. In summary even where $f(x_2)$ is non zero but its contribution is very small relative to the contributions from the other terms in Eq. 8h which forms the output of the system.

B.3.3 Phase 3: $t \in [\tau_3, \tau_e]$

In this phase x_3 is a decreasing function. The numerical evaluation of Fig. 15 revealed $x_3 < x_3^u|_{t=\tau_3} = 0.23$. Hence, $f_1(x_3^u) \leq f_1(0.23) \approx 0$, $x_2^u \approx 0$, and finally $f(x_2) \approx 0$ in this interval. Thus, here again the feedback block can be neglected.

In summary, in phases 1 and 3 the contribution of the feedback block is negligible. During phase 2 $f(x_2) = 0$, but its relative contribution to the state x_8 remains negligible. Thus, we can remove the feedback block with no ill effects on the model performance. The numerical simulations in the next section reiterate this point.

B.3.4 Statistical evaluation of the set of equations affecting x_2 in all cases

So far we examined the details of the nominal un-habituated case. Our next task is to understand the deviations from the nominal case for both un-habituated and the habituated responses. The data is obtained by simulations of Eqs. 8b through 8g since those are the only equations affecting the computation of $f(x_2)$. Note that here we deal with the actual states x_2 through x_7 , and not the upper bounds.

The following procedures were used in running and analyzing the numerical simulations.

1. The BFM (estimated in Sect. 4.2) parameters of the input currents I_2^+ , and I_2^+ that vary under habituation were collected from the entire data set.
2. The absolute minimum and maximum values of parameters that vary under habituations were taken as the range of the parameters (see Table 5).
3. A set of 10,000 samples were obtained by postulating a uniform distribution of the parameters between their minimum and maximum respectively, were generated.
4. The states x_2 through x_7 , were computed using Eqs. 8b through 8g, for each sample point.
5. The values of $f_1(x_3)$ and $f(x_2)$ were computed for each sample.
6. The set of samples for which the event $f_1(x_3) > 0.04$ was non empty was identified, and pertinent features were evaluated.
7. The set of samples where $f(x_2) > 0.01$ was true over a nonvanishing interval were identified, and pertinent features of this set were collected.

The features mentioned in (vi) and (vii) found in Table 6 were the following:

1. The mean and standard deviation of the peak amplitude for the states x_2 , and x_3 .
2. The mean and standard deviation of peak time for the states x_2 , and x_3 .
3. The approximate support interval of $f_1(x_3)$ and $f(x_2)$ (both are denoted in Table 6 as $f(\cdot)$). The approximate support is defined as the time interval over which each $f(\cdot)$ is greater than 1% of its respective maximum value.

The following was observed from the numerical simulations.

1. In 32.39% of the cases, $f_1(x_3) > 0.04$.
2. In only 0.1% of the cases, $f(x_2) > 0.01$.

The pertinent features of the states x_3 and x_2 are shown in Table 6. In particular $f(x_2) = 0.02$ throughout, and nonnegligible values of $f(x_2)$ last typically for 2–6 ms which is relatively short compared to the time interval in which it occurs. Thus, the conclusion is indeed that f

(x_2) can be ignored and the feedback block can be removed, and the impact is negligible on the model performance.

B.3.5 Comparison of the reduced Riera model with the original Riera model

Complementing the conclusions of the preceding analysis we compare the reduced three state model with the original nine state Riera model. This comparison is summarized in Table 7. We denote the output of the three state Riera model by y_{nro} , and the output of the original nine state Riera model, denoted by y_{nfo} . The performance metric we use is the relative L_1 norm of the error:

$$\frac{\|y_{\text{nfo}} - y_{\text{nro}}\|_1}{\|y_{\text{nro}}\|_1}.$$

With relative error less than 2.3%, the table shows that the three states reduced order model produces essentially the same predictions as the original, nine state Riera model, throughout the data set under consideration.

B.3.6 Justification of the substitution of $f(x_1)$ by $\tilde{I}_3^+$

The nonlinear function $f(x_1)$ appears as a feed forward input into the reduced order linear model. It was observed throughout the data set that $f(x_1)$ had the approximate shape of a rounded pulse much like the inputs I_i^+ , where $i = 1, 2, 3$. The parameterization of $\tilde{I}_3^+$, given in Eq. 18 was made to match this feature. The quadratic function for $\tilde{I}_3^+$ is parameterized to it to have a value of zero when $t - \tilde{\tau}_1 = 0$, and $t - \tilde{\tau}_1 = \beta$, and to have a peak at the center of the interval $(0, \beta)$.

$$\tilde{I}_3^+ = \alpha(t - \tilde{\tau}_1) \left(1 - \frac{(t - \tilde{\tau}_1)}{\beta} \right), \quad 0 \leq t - \tilde{\tau}_1 \leq \beta \quad (18)$$

A least-squares fit to $f(x_1)$ over the rat data sets for the α and β parameters led us to fix $\beta = 8$ ms in our work. The results summarized in Figs. 12 and 13 demonstrate in particular that the match between the linear model and reduced three state model using $f(x_1)$ are similar.

C Habituation characteristics

The term *habituation* is subject to a broad range of interpretations in a large body of published literature. Occasional cases of conflicted or opaque interpretations motivated the distillation of nine, well defined attributes of habituation phenomena, in the seminal article by Thompson and Spencer (1966). Thirty three years later, that list was revised and updated in the excellent review by Rankin et al. (2009), now elaborating ten, widely accepted descriptors. Positioning this study within that consensus, this appendix summarizes those descriptors from Rankin et al. (2009) that are pertinent to our study, in Sect. C.1. In Sect. C. 2 we explain how the proposed model meets these criteria, and in Sect. C.3, demonstrate that the same applies to the EEG data we use.

C.1 Characteristics pertinent to the present study

Referring to the numbering in Rankin et al. (2009), of the ten characteristics listed in that review, those pertinent to the data used in the present study are #1, #2, #4, #6, and #10. For ease of reference, the summary descriptions of those characteristics are provided below. The other descriptors in Rankin et al. (2009) relate to phenomena involving, e.g., stimuli with

varying intensity or duration, and are therefore not applicable to the present study (cf. the brief description in Sect. 2.1 and the detailed original, in Franceschini et al. (2008)).

- 1 Repeated application of a stimulus results in a progressive decrease in some parameter of a response (habituation).
- 2 If the stimulus is withheld after response decrement, the response recovers over time (spontaneous recovery).
- 4 Other things being equal, more frequent stimulation results in more rapid and/or more pronounced response decrement, and more rapid spontaneous recovery.
- 6 The effects of repeated stimulation may continue to accumulate even after the response has reached an asymptotic level (which may or may not be zero, or no response).
- 10 Some stimulus repetition protocols may result in (long lasting changes in) properties of the response decrement (e.g., more rapid rehabilitation than the baseline, smaller initial responses than baseline, smaller mean responses than baseline, less frequent responses than baseline).

C.2 Applicability to the proposed model

The proposed habituation model is encapsulated in (2). It provides a quantitative realization of the qualitative descriptors of habituation stated above, and is the central contribution of this article. For ease of reference we rewrite (2) as (19a), right below, and recall the definitions and role of its ingredients:

$$\dot{s} = -\frac{1}{\tau_s}(0.5+s)s + g_s u(t) \quad (19a)$$

The input, “ $u(t)$ ”, in (19a), is the absolute value of the post-stimulus “P1–N1–P2” waveform, in the EEG signal. The system’s state, $s(t)$, serves as a quantitative measure of the instantaneous level of habituation. The sampled value of s at the time a new stimulus arrives is used to tune the coefficients of the mass neuronal model that will be used during the response to that stimulus, to appropriately account for habituation. Figure 8 illustrates that dependence. The parameter τ_s scales the state-dependent time constant, in (19a):

$$\tau(s) := \frac{\tau_s}{0.5+s}. \quad (19b)$$

As we see, $\tau(s)$ is monotonously decreasing function of s , representing the instantaneous level of habituation, a point we shall revisit shortly.

Equation 19a is a nonlinear, stable, low-pass filter. Interpreting the dependence of $\tau(s(t))$ as a time function, we observe that the solution, $s(t)$, must satisfy the *variation of parameters* solution of the equivalent linear time varying system:

$$s(t) = e^{-\int_0^t \frac{1}{\tau(s(\xi))} d\xi} s(0) + g_s \int_0^t e^{-\int_\xi^t \frac{1}{\tau(s(\zeta))} d\zeta} u(\xi) d\xi \quad (19c)$$

Note that the selection of the zero time, $t = 0$, is arbitrary. An ideal, “infinite time” expression, thus, precludes the first term on the right hand side of (19c) and substitutes $-\infty$ for zero, as the lower bound of integration.

Dynamic properties of (19), integrated with the modified Riera model, insure that the model-predicted response agrees with the listed characteristics:

Characteristic #1

As illustrated by Fig. 8 (left column), the dependence of the amplitude-parameters of the mass neuronal model, on the value of s , is monotonously decreasing. A larger value of s at the time a stimulus arrives, results in smaller values of these parameters and, in turn, greater attenuation of the model-predicted evoked EEG response to that stimulus.

The contribution of the rectified evoked response, u , to the value of s is by the (time-dependent) convolution with a positive kernel, in (19c). At time t , this contribution is cumulative, over the response to all prior stimuli. All other factors being equal, each past stimulus leads to a positive contribution to $s(t)$, and thus, to an increase in the attenuation of the evoked response to a subsequent stimulus. This cumulative effect establishes the basic aspect of #1.

The exponential decay of the convolution kernel becomes more accentuated for larger s . This easily translates to the observation that a bounded input will lead to a bounded output in s . When past evoked stimuli are repetitive, but the ISI allows enough time for recovery the sampled values of s , at the arrival of each new stimulus, may be (nearly) constant. That is the case in the lower stimulus frequencies in our study (1–2 Hz). With shorter ISI, the response may be an oscillatory attractor: an attenuated evoked response, or a sequence of several attenuated responses periods, enables some recovery in s , and thus, lead to a larger response to the next stimulus, causing s to rise, and vice versa. These oscillations may be periodic, but more likely, display a chaotic nature, whose statistics depend on the ISI in the periodic signal. This is illustrated in the bottom two panels of in Fig. 6, each representing the first 2 s of the response of s to inputs at 4 Hz and at 8 Hz. Figure 7 illustrates these phenomena, showing the sampled values of s at stimulus times over an entire 4-s stimulus train, for the full set of 1–8 Hz stimuli. The dependence of statistical characteristics of the response on the stimulus frequency is discussed in detail in Sect. 5 and in the figures and tables therein. Whether in a deterministic or a statistical sense, constant or periodic, the level of attenuation of the model-predicted response, reaches an asymptotic attractor in all cases. This asymptotic behavior is attained, early on, when attenuation is minimal. Yet as attenuation levels increase, it may take a substantial portion of the stimulus period. For example, the asymptotic behavior of s , in response to an 8 Hz. response, seems to requires some 2.5 s to be reached. Certainly, the sequence of the first three responses is not repeated anywhere later on. The existence of this convergence to an asymptotic behavior is the complementary aspect of the key characteristic #1.

Characteristic #2

Absent an input following a time $t = t_1$, the variation of parameters formula (19c) reduces to a decaying exponential term:

$$s(t) = e^{-\int_{t_1}^t \frac{1}{\tau(s(\zeta))} d\zeta} s(t_1), \quad t > t_1.$$

Thus, $s(t)$ will decline as long as a new stimulus is not introduced, the amplitude-parameters g_i will grow, and with them, the intensity of the evoked response.

Characteristic #4

A decrease in the ISI (equivalently, an increase in the frequency of stimuli) causes windowed time averages of the rectified evoked responses, i.e., the input u , to increase. The impact will be a growth of s , as shown in Fig. 7. Higher values of s are reflected by stronger attenuation of the evoked response, as noted right above. Additionally, higher levels of attenuation, associated with higher values of s , translate to smaller values of the instantaneous time constant $\tau(s)$, in (19), hence to faster rates of both a decline in s and recovery, absent a stimulus, and of growth of s and subsequent attenuation, following the introduction of a new stimulus.

Characteristic #6

As illustrated in Fig. 8, the amplitude-parameters g_i saturate, reaching a small, “no discernible response level”, at $s = 3$. Nonetheless Eq. 19a continues to integrate the positive rectified response signal, u , small as it might be. Thus, not only is it impossible for s to then decline to zero, but it may actually be the case that s will rise to a level higher than the saturation level, a fact illustrated by Fig. 7.

Characteristic #10

The meaning of “long term” is restricted by the nature of our data to stimulus trains lasting 4 s, with up to 32 consecutive stimuli. Within this limitation, both Figs. 7 and 9 illustrate the fact that once the properties mentioned above, the attenuation of the evoked response and the shortening response time constant, as predicted by the model, last throughout the said time stretches. [This fact is confirmed by the response to longer stimulus trains, lasting up to 8 s, which are discussed in Franceschini et al. (2008)].

In summary, what we have established so far is that, by design, model predictions are consistent with the listed characteristics of habituation from Rankin et al. (2009). Our next task is to use this observation to establish the same conclusion with respect to the EEG data used in this study. That is our next topic.

C.3 Habituation is present in our EEG data

To make the discussion precise, this article quantifies the evoked response in terms of six quantifiable properties of the P1–N1–P2 waveform: Three amplitude quantities (g_i for $i = \{1, 2, 3\}$) that provide a clear connection to attenuation, and three delay times (τ_i) that are perhaps less crisply associated with attenuation, but that are nonetheless demonstrators of a quantifiable cumulative effect of recent responses. Our argument that the EEG response, measured by these quantifiers, exhibits habituation, is by the following simple chain of deductions:

The rigorous analysis, reviewed in Sect. 5 and discussed in Sect. 6, demonstrates unequivocally that model predictions fit the EEG data. These discussions also demonstrate the dependence of waveform quantifiers on values of s , as determined by our EEG data, via (19). The previous section explained how the response of s meets the listed criteria of habituation from Rankin et al. (2009). Adding these observations up, we conclude that habituation, as characterized in Rankin et al. (2009), is present in the data used in our study.

In closing we comment on the particular issue of alternating presence and absence of a discernible P1–N1–P2 pattern in the evoked response. Difficult as it might be to make that

observation by an inspection of an individual response period, the sharp dichotomy of a bimodal distribution, in Fig. 10, is a demonstration that, statistically, that distinction is clear. The capability to predict this statistical behavior by the proposed model, and the fact that s meets the habituation criteria, together provide a strong support to the claim that the presence or absence of a discernible response, as defined here, is a feature that meets the habituation criteria, as well.

Fig. 1. Signal-averaged waveforms illustrating habituation to stimulus as the ISI shortens. *Left panels* show the EEG response to stimuli at frequencies of 1–8 Hz, each over a 4-s recording interval. Each *right panel* shows an expanded view of the boxed region in the corresponding *left panel*, covering a 1-s interval. The *right panels* clearly reveal the phenomenological components of our interest, the P1–N1–P2 waveform. The *top right panel* (1-Hz response) also shows the response from a single trial (*dashed curve*) to illustrate the effect of the signal averaging

Fig. 2.

Spectrum of the background EEG signal (pre-stimulus, *dashed-dot/blue*), compared to the spectrum of the post-stimulus EEG signal, in cycles where a P1–N1–P2 waveform is conspicuously present (*solid/green*), and in cycles where that waveform is not discernible (*dashed/red*). The ripple in the plotted spectra, noticeable at frequencies higher than ~50 Hz, for the background signal, and higher than ~200 Hz, for the response and no-response signals, is an artifact of the combination of small ensemble averaging and short time intervals. The ripple stroke amplitudes, ranging at 4 dB, are negligible over these frequency ranges. (Color figure online)

Fig. 3.

Riera's Neural mass model consists of a lumped dynamic representations of PCs and of two IN groups, where one acts as a feed-forward connection, and the other, as a feedback mechanism, with a nonlinear (saturation) interconnection. The PC receives signals from the surrounding neurons via the apical dendrites, represented by I_2^+ , and from the thalamus through the basal dendrites, represented by I_1^+ . The signal driving the feedforward IN is represented by I_3^+

Fig. 4.
Block diagram of the Riera model showing the connectivity between states

Fig. 5.

A generic block diagram of habituation control. The predicted habituation in the response of a neural mass model to a sensory stimulus is determined by adjustments of model parameters, as functions of the habituation control variable s . That parameter is the state of a nonlinear, low-pass filter, driven by the EEG-evoked response

Fig. 6.

The rectified EEG signal, as predicted neural mass model (*top panels*) is the input to the ODE (2) that governs the evolution of the auxiliary state s , shown in the *bottom panels*. Results are shown for the first 2 s of stimulus trains with ISI of 0.25 s (4 Hz, *left panels*) and 0.125 s (8 Hz, *right panels*). The + signs mark the respective samples of $s(t)$ at stimulus times, denoted $s(t_s)$. The neural mass model parameters during a response cycle are functions of these sampled values, and determine the prediction of habituation effects, as described in Sect. 4.3

Fig. 7.

The *solid curve* is a 1-s trajectory of s , from Eq. 2, driven by the un-habituated EEG response to an isolated stimulus. The $+$ marker indicates samples of $s(t)$ at the time of arrival of subsequent stimuli. Eight data records were used, each characterized by a different stimulus frequency, from 1 to 8 Hz. The *heavy arrows* on the *left* vertical axis indicate levels of $s(t_s)$, where noticeable habituation begins (*dark (black) arrow*), and above which the neural mass does not show a discernible P1–N1–P2 respond (*light (red) arrow*). (Color figure online)

Fig. 8.

BFM scatter plots of the parameters of the linear Riera model, marked by +, shown as functions of the corresponding values of the saturation control parameter s . The *solid curves* are obtained by optimal least-squares of the functional dependencies in (4). Note that low expected values of g_i make high scatter of corresponding values of the time delay parameters (say for $s > 3$) immaterial. The dependence curves in these mappings were used subsequently as the initial estimates for optimization of the PM1 and PM models over the entire data set. Essentially identical mappings were used for the nonlinear Riera model

Fig. 9.

The average EEG response to trains of 8 Hz stimuli (the most challenging case) from data set 4L, and the predictions generated by three types of habituation models, using the linear simplification of Riera model: The BFM, a Single Cycle Predictive Model (PM1), and Predictive Model (PM)

Fig. 10.

Firing threshold detection using a normalized smoothed histogram of peak-to-peak amplitudes from the second cycle of 6–8 Hz from data set 4L. The cycles in the portion marked by *dark (red) bar* were designated as no-response, and those in the portion marked by a *light (green) bar* were designated as indicating presence of a response. (Color figure online)

Fig. 11.

A sample plot of the values of features collected from the data, BFM, PM1, and PM, for the linear Riera model. The *bars* show the range between 10 and 90 percentile. The predictive models (PM1 and PM) qualitatively capture all the features. The P2 start time feature similar across BFM, PM1, and PM

Fig. 12.

Comprehensive summary of linear Riera model accuracy as reflected by waveform features across all data sets. Each *box* represents a set of time (*left half*) or shape (*right half*) features for the particular data set designated by the two-character code near the *top* of each *box*. The set of *panels on the left* show the mean error ($p_{\text{data}} - p_{\text{model}}$) in time features, with *bars* representing the standard deviation, in milliseconds. The set of *panels on the right* show the mean relative error ($\|p_{\text{data}} - p_{\text{model}}\|_2^2 / \|p_{\text{data}}\|_2^2$) between the shape features collected from the data, and those collected from models. The *top six panels* on each side show the results for data sets from the left brain (right fore-paw) while the *bottom six panels* show the results for data sets from the right brain (left fore-paw)

Fig. 13. Comprehensive results of waveform feature accuracy for the nonlinear Riera model. The layout is the same as the previous figure

Fig. 14.

Comparison of the LOOCV linear PM errors, across all features, with errors generated by using the optimal model parameters for each set. The layout is the same as in the previous two figures

Fig. 15.

Plot of the states x_5 , and x_6 (*top left panel*), and x_3^u and x_2^u (*top right panel*) for the parameters of the input currents I_1^+ , and I_2^+ , given in Table 4. $f_1(x_3^u)$ and $f(x_2^u)$ are plotted in the *bottom left* and *right panels*, respectively. $\max f(x_2^u)=0.02$ suggesting that the feedback block may be removed

Table 1

The correspondence between the response waveform components, P1, N1, and P2, the current inputs that serve as the primary drive for each of these components, and the parameters used in the Gaussian representation of each of these current inputs, as in (1)

Wave	Current	Gain, g	Delay, τ	Std. σ
P1	$I_3^+, \tilde{I}_3^+$	g_1	t_1	σ_1
N1	I_1^+	g_2	t_2	σ_2
P2	I_2^+	g_3	t_3	σ_3

Table 2

F scores of firing detection

Data set	Linear Riera			Nonlinear Riera		
	BFM	PM1	PM	BFM	PM1	PM
1L	0.96	0.84	0.80	0.95	0.84	0.81
1R	0.93	0.81	0.83	0.93	0.81	0.85
2L	0.91	0.76	0.75	0.91	0.78	0.80
2R	0.92	0.76	0.79	0.85	0.76	0.74
3L	0.94	0.86	0.85	0.91	0.85	0.89
3R	0.97	0.86	0.80	0.92	0.84	0.78
4L	0.99	0.92	0.98	0.97	0.92	0.97
4R	0.96	0.79	0.87	0.92	0.79	0.86
5L	0.96	0.82	0.82	0.95	0.83	0.85
5R	0.93	0.79	0.81	0.92	0.77	0.79
6L	0.94	0.86	0.88	0.95	0.86	0.89
6R	1.00	0.88	0.85	0.98	0.84	0.96
Mean	0.95	0.83	0.84	0.93	0.82	0.85
Std	0.026	0.047	0.056	0.033	0.043	0.067

The PM is the only model that truly detects a firing but the numbers for BFM and PM1 are given for comparison.
Predictions by the linear and nonlinear Riera models are similar

Table 3

Relation between the state-space parameters and the Riera model parameters.

State-space parameters	Relation to Riera model parameters	Value
a	$1/\tau_m$	50
b	R_m^0/τ_m	204
c	$\frac{1}{\tau_m} \left(\alpha_0 + \sum_{k=1}^2 \frac{1}{\beta_k} \right)$	396
d	$\frac{1}{\tau_m \beta_1 \beta_2}$	3
d_1	R_m/t_m	144
g_1	$R_m \alpha_{IN}/t_m$	44
g_2	$\frac{R_m}{(R_1^1 + R_c^1)\tau_m}$	288
g_3	$\frac{R_m}{(R_1^2 + R_c^2)\tau_m}$	29
h	$\frac{1}{\tau_m} \sum_{k=1}^2 \frac{1}{\beta_k}$	29
s	$\frac{1}{\tau_m} \left(1 + R_m \sum_{k=1}^2 \frac{1}{R_m^k} \right)$	912
K_1	$\frac{\beta_1 R_m}{(R_1^1 + R_c^1)\tau_m}$	647
k_2	$\frac{\beta_1 R_m}{(R_1^1 + R_c^1)\tau_m}$	215
R_1	$\frac{R_c^2 R_m}{\tau_m \prod_{k=1}^2 (R_1^k + R_c^k)}$	157
r_2	$\frac{R_c^2 (R_m + R_m^1)}{\tau_m \prod_{k=1}^2 (R_1^k + R_c^k)}$	169
q	$\frac{R_c^2 R_m^1}{\tau_m \prod_{k=1}^2 (R_1^k + R_c^k)}$	12
o	$\frac{R_m R_c^2}{\tau_m (R_1^2 + R_c^2)}$	78
p	$\frac{R_c^2 R_m^2}{\tau_m (R_1^2 + R_c^2)}$	18
w_1	α_{PC}	0.4
T	T	0.03
a	α_{IN}	0.3
γ	γ_{IN}	5
γ_1	γ_{PC}	6

State-space parameters	Relation to Riera model parameters	Value
x_{01}	V_0^{PC}	0.6
x_0	V_0^{IN}	0.7

Values of the state-space parameters are computed using the relations with the Riera model parameters, whose values are published in Riera et al. (2007). Riera's parameter values in Riera et al. (2007) imply the selection of $a = 33$. Our election of $a = 50$ reflects the difference between the human data used in Riera et al. (2007) and the rat data used in this study

Table 4

Nominal values of the parameters of input currents I_1^+ and I_2^+ which are used for analyzing the dynamics of the state x_3

Input current parameter	Nominal value
a_1	1
a_2	0.3
τ_2	20 ms
τ_3	50 ms

Table 5

The range of the varying parameters of input currents I_1^+ and I_2^+ collected from all rat data sets

Parameter	Range
α_1	[0.1, 1.4]
α_2	[0, 0.3]
τ_2	[18, 45] ms
τ_3	[40, 80] ms

Table 6

Features of the states x_3 and x_2 collected from numerical computation of Eqs. 8b through 8g using 10,000 uniform random samples of the input current parameters with the range of values from Table 5

Feature	State x_3	x_2
Peak amplitude	0.58 ± 0.08	0.41 ± 0.01
Peak time	43 ± 7 ms	60 ± 3 ms
Approx. support of $f(\cdot)$	17 ± 3 ms	4 ± 2 ms
$f(\cdot)$ peak	0.13 ± 0.06	0.02 ± 0.002

The $f(\cdot)$ refers to both $f_1(x_3)$ and $f(x_2)$ in the table

Table 7

Comparison between the output of the nonlinear three state Riera model, denoted by y_{nro} , and the output of the original nine state Riera model, denoted by y_{nfo}

Data set	Relative difference
1L	0.020
1R	0.030
2L	0.020
2R	0.021
3L	0.021
3R	0.020
4L	0.017
4R	0.020
5L	0.020
5R	0.023
6L	0.023
6R	0.019

The performance metric we use is the relative L_1 norm of the error $\frac{\|y_{\text{nfo}} - y_{\text{nro}}\|_1}{\|y_{\text{nro}}\|_1}$