Low dimensional modeling of localized neural and hemodynamic response with habituation from multimodal brain measurements

A Dissertation Presented

by

Srinivas Laxminarayan

to

The Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Northeastern University
Boston, Massachusetts

March 2011
To my Father and Mother
Acknowledgement

Foremost, I would like to express my sincere gratitude to my advisers Dr. Dana H. Brooks and Dr. Gilead Tadmor for their continuous support during my Ph.D research and study. I thank them for the long hours of discussion leading to a fruitful research in a completely new area to me, pointing out my flaws specifically in developing a scientific approach to research and research based writing. Their encouragements and time consuming criticisms of my research writing skills have helped me think and improve the way research work can be presented to a wide audience.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Solomon G. Diamond at Department of Biomedical Engineering at Dartmouth college with whom I mainly worked during the initial days of my PhD research. He provided me with references that helped me learn the field of neuroscience in general, and read papers related to our research work in particular. I thank Dr. Maria Angela Franceschini at the Athinoula A. Martinos research center of biomedical imaging for providing us with the data for my work, and also guiding me with respect to the biological aspects of my work. I also thank Dr. Deniz Erdogmus for his insightful comments and questions, which has helped improve the quality of my work.

My sincere thanks also goes to: Dr. Eric Miller at the Department of Electrical and Computer Engineering in Tufts university for guiding me along with my advisers during the initial days of my PhD research. Dr. David Boas, and Dr. Theodore Huppert for refering me to papers related to blood flow measurements in the brain neocortex.
Last but not the least, I would like to thank my family: my parents K. R. Laxminarayan and Hemalata for supporting and guiding me throughout my life. I thank them for enduring long periods of separation and difficulties to help me in my betterment. My room-mates through the years of my PhD, and specifically I reserve special thanks to my friend Arun K. Parthasarathy for being extremely patient to bear my idiosyncrasies and relieve me from cooking duties, and taking care of my wellbeing through the last two years of my PhD.
Abstract

Multimodality functional brain imaging has been gaining importance due to the complementary nature of many of the modalities. Dynamical systems based models of neural activity and local hemodynamics can offer enhanced spatiotemporal resolution and insight into physiological signals and mechanisms. One major tool for studying brain function is to provoke local "evoked responses" by repeated application of a stimulus.

Evoked responses to stimuli show complex habituation behavior as the stimulus repetition frequency increases. To gain insight into the relation between local neural activity and hemodynamics we propose a control structure that enables neural mass models to predict habituation as revealed in rat EEG under medial nerve stimulus. We report on the accuracy with which these models recreate the data waveform under stimuli at varying frequencies (from 1-8 Hz), as well as the accuracy with which they mimic complete inhibition of firing at higher stimulus rates. We also compare the predictive power of the models, demonstrating the capability of simplified representations to capture key features of the mass evoked neuronal response.

Equipped with such a predictive neuronal model we explore the relationship between the neural response and the local hemodynamics by reconstructing the inputs of a Windkessel-based hemodynamic model and relating the basis functions of the neural signal to those of the inputs to the hemodynamic model. The model that relates the neuronal responses to the hemodynamic inputs is termed as the Neurovascular model.
Contents

Acknowledgement iii

Abstract v

1 Introduction 1

1.1 Model based reconstruction for functional brain imaging 2

1.2 Focus of this work 3

1.3 Arrangement of the thesis 4

1.4 Papers resulting from the work 5

2 Neuron model 7

2.1 Introduction 7

2.2 Phenomenology and models 11

2.2.1 The experimental data 11

2.2.2 Waveforms of characterization of the evoked response, and the habituation phenomenon 12

2.2.3 The Riera model 17

2.3 Habituation control structure 21

2.4 Empirical model optimization 26

2.4.1 Model Variants 27

2.4.2 Best Fit Model 30

2.4.3 Closing the loop: Habituation-adjusted models 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>Model performance analysis</td>
<td>34</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Performance criteria</td>
<td>35</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Model performance</td>
<td>36</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Cross validation</td>
<td>40</td>
</tr>
<tr>
<td>2.6</td>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>2.6.1</td>
<td>The phenomena: Evoked response and habituation</td>
<td>44</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Isolating the evoked response from the EEG signal</td>
<td>45</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Labeling discernible and indiscernible evoked response</td>
<td>46</td>
</tr>
<tr>
<td>2.6.4</td>
<td>Simplifications of the Riera model</td>
<td>47</td>
</tr>
<tr>
<td>2.6.5</td>
<td>Additional simplifications: The model as a parameterized curve fit</td>
<td>48</td>
</tr>
<tr>
<td>2.6.6</td>
<td>Feature estimates vs. curve fit evaluation</td>
<td>48</td>
</tr>
<tr>
<td>2.7</td>
<td>Concluding Remarks</td>
<td>49</td>
</tr>
</tbody>
</table>

### 3 Neurovascular coupling

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>Experimental methods, data collection and description</td>
<td>53</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Experimental procedure and data collection</td>
<td>53</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Description of typical hemodynamic responses</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>Physiologically motivated hemodynamic model</td>
<td>58</td>
</tr>
<tr>
<td>3.3.1</td>
<td>The model structure</td>
<td>59</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Reconstruction of the inputs to the hemodynamic model from data</td>
<td>60</td>
</tr>
<tr>
<td>3.4</td>
<td>Development of the coupling model</td>
<td>63</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Selecting $V_D(t)$ and $V_{hi}(t)$</td>
<td>64</td>
</tr>
<tr>
<td>3.4.2</td>
<td>The temporal model</td>
<td>66</td>
</tr>
<tr>
<td>3.4.3</td>
<td>The static functional mapping</td>
<td>67</td>
</tr>
<tr>
<td>3.5</td>
<td>Results</td>
<td>69</td>
</tr>
</tbody>
</table>
3.6 Discussion .......................................................... 72
  3.6.1 Essence of the neurovascular model ..................... 72
  3.6.2 On the advantage that the model structure provides .... 74
  3.6.3 On the sensitivity of the hemodynamic model .......... 74
3.7 Conclusion ......................................................... 75

4 Model reduction .................................................. 77
  4.1 Introduction ...................................................... 77
  4.2 The Dynamic multimodal Framework ......................... 78
    4.2.1 Sensory block ............................................. 80
    4.2.2 The Neuronal mass model ............................... 80
    4.2.3 The neurovascular response ............................. 82
    4.2.4 Three compartment windkessel model ................. 82
    4.2.5 Combining all the models ............................... 83
  4.3 Dimension reduction methods ............................... 85
    4.3.1 Balanced Truncation method ............................ 86
    4.3.2 Proper Orthogonal Decomposition (POD) method .... 86
  4.4 Results and Discussion ....................................... 87
  4.5 Future directions ............................................ 89

5 Conclusions and future work .................................. 90

A EEG data preprocessing ....................................... 92

B Order reduction in the Riera model ......................... 95
  B.0.1 Riera model equations .................................. 96
  B.0.2 An overview of the proposed model reduction ........ 98
  B.0.3 Removing the feedback block ............................ 100
C Single compartment Windkessel model

C.1 Hemodynamic model equations ............................................ 110

C.2 Sensitivity analysis of the hemodynamic model ......................... 112

C.2.1 Sensitivity to baseline parameters .................................... 113

C.2.2 Sensitivity of the model to the $\Delta CMRO_2(t)$ input .......... 114
List of Figures

1.1 The Dynamic Multimodal Integrated (DMI) framework introduced in [17]. .................................................. 2

1.2 Sensory stimuli drive the habituation predictive neuron model whose output can be measured by EEG. The Coupling model relates the neuron model output to the inputs of the hemodynamic model which generates oxygenated/de-oxygenated hemoglobin concentration (HbO/HbR). Hemodynamic response can be measured by optical methods. ........... 3

2.1 Signal-averaged waveforms illustrating habituation to stimulus as the inter-stimulus interval shortens. Left panels show the EEG response to stimuli at frequencies of 1 – 8 Hz, each over a 4 second recording interval. Each right panel shows an expanded view of the boxed region in the corresponding left panel, covering a 1 second 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. .................................................. 10
2.2 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 \(\sim 50\) Hz, for the background signal, and higher than \(\sim 200\) 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 \(\leq 4\) dB, are negligible over these frequency ranges. 13

2.3 Riera’s Neural mass model consists of a lumped dynamic representations of Pyramidal Cells (PC) and of two InterNeuron groups (IN), 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 feed-forward IN is represented by \(I_3^+\). 18

2.4 Block diagram of the Riera model showing the connectivity between states. 19

2.5 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 non-linear, low-pass filter, driven by the EEG evoked response. 22
2.6 The rectified EEG signal, as predicted neural mass model (top panels) is the input to the ODE (2.3.1) that governs the evolution of the auxiliary state $s$, shown in the bottom panels. Results are shown for the first 2 sec of stimulus trains with ISI of 0.25 sec (4 Hz, left panels) and 0.125 sec (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 §2.4.3. .................................................. 23

2.7 The solid curve is one second trajectory of $s(t)$, from equation (2.3.1), 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 Hz 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). .................. 24

2.8 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 (2.4.2). 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 dataset. Essentially identical mappings were used for the nonlinear Riera model. .................................................. 29
2.9 The average EEG response to trains of 8 Hz stimuli (the most challenging case) from dataset 4L, and the predictions generated by three types of habituation models, using the linear simplification of Riera model: The Best Fit Model (BFM), a Single Cycle Predictive Model (PM1), and Predictive Model (PM). .......................................................... 30

2.10 Firing threshold detection using a normalized smoothed histogram of peak-to-peak amplitudes from the second cycle of 6 Hz to 8 Hz from dataset 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. . . . 36

2.11 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 percentile 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. ................................. 39

2.12 Comprehensive summary of linear Riera model accuracy as reflected by waveform features across all datasets. Each box represents a set of time (left half) or shape (right half) features for the particular dataset designated by the two-character code near the top of each box. The set of panels on the left show the mean error ($p_{data} - p_{model}$) in time features, with bars representing the standard deviation, in milliseconds. The set of panels on the right show the mean relative error ($\|p_{data} - p_{model}\|^2/\|p_{data}\|^2$) between the shape features collected from the data, and those collected from models. The top six panels on each side show the results for datasets from the left brain (right fore-paw) while the bottom six panels show the results for datasets from the right brain (left fore-paw). .......................................................... 41
2.13 Comprehensive results of waveform feature accuracy for the non-linear Riera model. The layout is the same as the previous figure. .......................... 42

2.14 Comparison of the Leave-One-Out Cross-Validation (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. ......................................................... 43

3.1 EEG evoked neural responses drive the Coupling model which in turn provides the inputs for the hemodynamic model. The hemodynamic model relates the signals from the Coupling model to changes in blood oxygen concentrations (HbO/HbR). The coupling signals are CMRO$_2$ (the metabolic rate of oxygen demand by the surrounding brain tissue), and $g$ (vascular conductance). .......................................................... 53

3.2 Head map of EEG electrodes and the optical source-detector arrangement for gathering concurrent measurements from medial nerve stimulation in rats. The EEG electrodes e1 and e3, are positioned near the somatosensory cortex. ................................................................. 54

3.3 The left panels show the EEG responses for different stimulus frequencies for a fixed duration of 4 seconds (dataset 4L). The top right panel shows the hemodynamic responses for varying stimulus frequencies (frequency experiment, dataset 4L). The bottom right panel shows the hemodynamic responses for a fixed stimulus frequency of 3 Hz and varying durations of stimulus (duration experiment, dataset 7L). The EEG responses for the duration experiments are identical to the responses for the 3 Hz stimulus (third left panel) over varying durations of stimulus, and hence are not shown. The notation used for the hemodynamic responses are $\Delta$HbO$=$HbO$-$HbO$_0$, where HbO$_0$ is the baseline oxy-hemoglobin, similarly $\Delta$HbR$=$HbR$-$HbR$_0$. .......................... 56
3.4 The block diagram of the single compartment Windkessel model. The state $v$ represents the vascular volume, the coupling signals $x_2$ and $x_4$ are states of the $O_2$ transport subsystem which represent the concentration of oxygen in the blood vessel, and the pial compartment respectively.

3.5 An example illustrating the reconstruction of the inputs of the hemodynamic model from $\Delta HbO$ and $\Delta HbR$. Known signals of $g$, and $\Delta CMRO_2$ (solid blue curves in the top left and bottom right panels) drive the hemodynamic model to generate the data ($\Delta HbO$, $\Delta HbR$ and $\Delta HbT$) in the top right panel. $\Delta HbT$ is used to drive equation (3.3.1) to recover $g$. $\Delta CMRO_2$ is recovered from the hemodynamic data by optimizing the weights in equation (3.3.2). The reconstructed signals are marked by red ‘+’ signs.

3.6 Block diagram of the coupling model. The rectified EEG response is filtered with a 1 second long averaging filter. The filtered response is represented by a scalar $a$ and temporal basis function $V_D(t)$. The temporal model relates $V_D$ to $\hat{V}_{hi}$ with $i = \{1, 2\}$, and static mappings $f(a)$ map $a$ to $a_j$, with $j = \{c, g\}$, to corresponds to the two inputs to the hemodynamic model. A linear combination of the temporal functions $\hat{V}_{hi}$, using $a_j$, results in the signals $g$, and $\Delta CMRO_2$, which drive the hemodynamic model.

3.7 The temporal model fits to the basis functions extracted from the duration datasets. Panel a) shows $V_D(t)$ for each stimulus duration. Panels b) through g) show $V_{hi}(t)$, where $i = \{1, 2\}$. The basis functions extracted from the duration datasets as described in $\S$3.4.1 are shown as solid curves. The solution of the temporal model in equation (3.4.1) driven by $V_D(t)$ are shown as dashed curves($\hat{V}_{hi}$) in panels b) though g).
3.8 Relation between the amplitudes $a$ of the filtered EEG responses and 
the amplitudes of $\Delta \text{CMRO}_2$ ($a_c$). The left panels show the two am-
plitudes $a_{c1}$ and $a_{c2}$ of the first two dominant basis (in blue), and 
$\tilde{a}_c = f_c(a)$ (in red) averaged over the frequency datasets 1L to 6R. 
The right panels show the same for the duration datasets 7L to 12R. 
The errorbars show the variation across datasets. 68

3.9 Sample of model fits to HbO, and HbR from the duration and frequency 
experiments. The blue circles follow the mean curve and the error bars 
show the standard deviation. 70

3.10 Relation between the amplitudes $a$ of the filtered EEG responses and 
the weight $w_{h1}$ obtained by projecting the HbO data onto the dominant 
temporal basis found via SVD analysis. Frequency experiments (mean 
of 1L to 6R) are on the left panels, and duration experiments (7L 
to 12R) on the right. The bottom panels show a nonlinear relation 
between $w_{h1}$ and $w_{h2}$, where $w_{h2}$ are the weights corresponding to the 
second dominant singular value. 73

4.1 The Dynamic Multimodal Framework 79

4.2 Normalized sensory response 80

4.3 The Neuronal Mass model from Riera’s paper [53] 81

4.4 Normalized neuronal response 81

4.5 Normalized neurovascular response 82

4.6 (top) circuit model for flow (bottom) compartment model for $O_2$ trans-
port to tissue 83

4.7 hemodynamic response 83

4.8 MEG data from 2 channels 84

4.9 EEG data from 2 channels 84

4.10 Optical data from 2 channels 85
LIST OF FIGURES

4.11 MEG channel 200: data from estimated states of the full and reduced order system .................................................. 88
4.12 EEG channel 354: data from estimated states of the full and reduced order system .................................................. 88
4.13 Optical channel 382: data from estimated states of the full and reduced order system .................................................. 88

B.1 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 B.2. $f_1(x_3^u)$ 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. ................................................. 104
# List of Tables

2.1 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 (2.2.1). .......................... 20

2.2 F-scores of firing detection. ........................................ 38

3.1 Performance of the neurovascular model across frequency and duration experiments. For each experiment type the Best Fit Error (BFE), and the Leave Two Out Cross Validation Error (LTOCVE) are tabulated. The LTOCV method involves removing the two datasets corresponding to a chosen rat. Estimating model parameters from the remaining datasets, and evaluating $E$ for the chosen dataset. .......................... 71

4.1 Closeness between simulated and estimated states .......................... 89

B.1 Relation between the state-space parameters and the Riera model parameters. ........................................ 99

B.2 Nominal values of the parameters of input currents $I^+_1$ and $I^+_2$ which are used for analyzing the dynamics of the state $x_3$. .......................... 100

B.3 The range of the varying parameters of input currents $I^+_1$, and $I^+_2$, collected from all rat datasets. .......................... 106
B.4 Features of the states $x_3$ and $x_2$ collected from numerical computation of equations (B.0.2b) through (B.0.2g) using 10,000 uniform random samples of the input current parameters with the range of values from Table B.3. .......................................................... 106

B.5 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 $y_{nfo}$. .................................................. 107

C.1 Values of constants in the hemodynamic model .............................................. 111
C.2 Initial conditions for the equation set (C.1.1) ................................................. 111
C.3 Relation between parameters in the equation set (C.1.1) and the baseline parameters. ................................................................. 111
C.4 Nominal values for the baseline parameters of the hemodynamic model. 112
C.5 Sensitivity of the hemodynamic model to the baseline parameters. The analysis was performed by using the nominal values of the baseline parameters from Table C.4 and using the sample problem which shows the reconstruction example in Figure 3.5 on page 62. .................. 114
Chapter 1

Introduction

Functional information from brain activity provides an insight into physiological processes that may help in understanding memory and cognition, and can provide diagnostic tools for brain diseases like epilepsy, Alzheimer, Parkinson etc [51, 48, 44]. These advantages have motivated a great interest in measuring functional brain activity using several noninvasive imaging modalities.

Perhaps one of the most common combinations measures both electrical response (e.g. EEG or MEG) and hemodynamic response (e.g. fMRI or diffuse optical measurements) [27, 52]. These modalities complement each other in their spatial and temporal resolution while providing information about functionally connected brain activity. However combining the information gained to reconstruct signals of interest in the brain presents many challenges. One such challenge is the inherent ill-posedness of the reconstruction problem not only due to less number of measurements, but also due to brain activity that spans multiple time scales.
Figure 1.1: The Dynamic Multimodal Integrated (DMI) framework introduced in [17].

1.1 Model based reconstruction for functional brain imaging

A systems approach of using physiologically motivated state-space models of brain function connected to forward models that relate brain function to the physics of the imaging modalities provides a logical framework to combining measurements from different modalities. Such a multimodal imaging framework is shown in Figure 1.1.
The advantage that such a framework provides are numerous in alleviating the illposedness of the inverse problem, i) Physiological models of brain function lead to an implicit reduction in the degrees of freedom alleviating the problem of over fitting the data, ii) more importantly such models reduce the degree to which one may simply find statistical associations between sets of measurements and thus increase the likelihood that physiologically meaningful associations are found, and, iii) updated versions of individual models could be directly plugged in and the overall system behavior can be analyzed. This approach might be helpful in understanding the global picture when measurement comes from multiple different modalities.

![Figure 1.2: Sensory stimuli drive the habituation predictive neuron model whose output can be measured by EEG. The Coupling model relates the neuron model output to the inputs of the hemodynamic model which generates oxygenated/deoxygenated hemoglobin concentration (HbO/HbR). Hemodynamic response can be measured by optical methods.](image)

### 1.2 Focus of this work

A major tool used in systems based modeling approaches for studying brain function is to provoke local “evoked responses” by repeated application of stimuli. It has been known from prior studies of brain function that under repeated stimuli, evoked responses adapt to the stimulus pattern in a complex manner. This habituation
phenomena is thought to be fundamental to many of the processes that underly learning, memory and cognition [7, 46]. Hence, modeling this habituation phenomena becomes an important challenge that has to be addressed if evoked response data has to be combined to gain functional information.

Indeed, previous studies using concurrent measurements of neural and hemodynamic activity developed systems based models of which we are aware, are only valid under fixed stimulus conditions [61, 54]. These models were not designed to predict the habituation to different stimulus patterns, and hence generalizing them to arbitrary stimulus conditions would involve re-estimation of model parameters to each condition. Here our goal is to contribute to a single generative model capable of predicting both EEG and optical measurements in a fashion that includes predicting the habituation to changes in stimulus application parameters. The model we present in this thesis includes an habituation component to predict the EEG, and adds to that a model for predicting the optical measurements via a coupling structure and a hemodynamic model. Figure 1.2 summarizes our modeling effort. The input is the stimulus train on the left. The measurements are the EEG and optical measurements on the bottom. The neural mass model includes an habituation control block that automatically modifies its own parameters to predict response to the types of stimulus patterns studied here. The remaining blocks predict the hemodynamic signals derived directly from the optical measurements in a fashion that tracks the changes in those signals that are driven by the changes in the neural mass model output.

1.3 Arrangement of the thesis

In Chapter 2 the habituation phenomena of the EEG in response to different stimulus application frequencies is used to develop a control structure that is applied to a neural mass model in order to enable habituation prediction. In Chapter 3 the relation between EEG and hemodynamic responses under variations in stimulus frequency
CHAPTER 1. INTRODUCTION

and duration is used to develop a single generative neurovascular model. The next logical step involves using the local model developed in Chapters 2, and 3 to combine functional imaging data. This process necessitates replicating the combined local model of brain activity at each brain voxel so as to enable dynamic reconstruction of the brain signals across the brain regions. But this process immediately leads to a problem of exponential increase in the number of states that need estimation, and hence computing them becomes an arduous task. If such approaches are to be made useful and feasible to the neuroscience community, then the issue of dimensionality reduction has to be addressed. Apart from offering computationally feasible models, identification of important states would also help in understanding hidden brain structures and relationships between the different structures. Chapter 4 deals with some methods to perform model reduction which, apart from offering computationally feasible models, would also help in understanding hidden brain structures and relationships between the different structures. This part of the work has not been fully addressed in this thesis and is a definite candidate for future research. Chapter 5 concludes this thesis by stating the contribution of this work and presents some directions for further research.

1.4 Papers resulting from the work

The chapters in this thesis have resulted in four papers under different stages of publication. The paper titles are given here as a quick reference.

1. A coupled model to jointly predict EEG and optical evoked response changes in rats under varying stimulus patterns, accepted for publication, NFSI & ICBEM conference, 2011.

3. Modeling habituation in rat EEG evoked responses via a neural mass model with feedback, Biological Cybernetics, in communication.

Chapter 2

Neuron model

2.1 Introduction

Neural modeling covers a wide range of scales, from coarse-scale dipole models [43] all the way to detailed biochemical computational models [6, 15, 16]. 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 [47, 30, 11, 35] and physiologically inspired models, based on a single or few functionally representative neurons, each representing a neural sub-population [53, 54]. 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 [13, 27, 32].

This chapter 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 [50]. It has been studied extensively over decades, covering diverse species, from aplysia [7, 49], through crayfish [75], crickets [40] and flies [66], snakes [2], mice and rats [58, 25], to humans [46]. Yet remarkably little is
known about the neural mechanisms underlying habituation [50], and the dynamics of habituation in a neuronal population [46]. Elucidating the nature of that dynamics is therefore a useful first step towards the understanding of mechanisms of learning and cognition at higher levels of complexity [50].

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 [22, 59].

To demonstrate this mechanism we employ the neural mass model structure introduced by Riera et al. [53, 54], 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 [60, pp. 67-71] and Jansen and Rit [29]). The Riera model, schematically shown in Figure 2.3, consists of a single Pyramidal Cell (PC) with feed-forward and feedback currents from two neighboring GABAergic InterNeurons (IN), each representing the respective sub-population 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 dataset used to identify and validate our model has been presented and discussed in detail in Franceschini et al. [19]. It consists of EEG measurements of the rat somatosensory cortex during episodes of medial nerve stimulation. Four second long trains of stimuli were applied to the animal, anesthetized with alpha chloralose, at frequencies varying from 1 Hz to 8 Hz. Generic examples of the ensemble-averaged

\footnote{GABA: \(\gamma\text{-Amino Butyric Acid}\) is a inhibitory neurotransmitter}
evoked response are illustrated in Figure 2.1, where each left panel depicts a complete 4 second record of the evoked response potential (ERP) and the corresponding right panel zooms in on a 1 second 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 chapter is organized as follows: §2.2 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 §2.3. 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 §2.4. Performance evaluation is presented in §2.5. Our approach and results are discussed in §2.6, 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 §2.7. Some technical aspects of the data processing method we use and of the analysis of Riera’s model and its simplifications, are deferred to the Appendices.
Figure 2.1: Signal-averaged waveforms illustrating habituation to stimulus as the inter-stimulus interval shortens. Left panels show the EEG response to stimuli at frequencies of 1 – 8 Hz, each over a 4 second recording interval. Each right panel shows an expanded view of the boxed region in the corresponding left panel, covering a 1 second 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.
2.2 Phenomenology and models

Following a brief description of the experimental data used here, in §2.2.1, we highlight properties and issues associated with the evoked response, in §2.2.2, and review the underpinning of Riera’s model and of its simplified versions, in §2.2.3. 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.

2.2.1 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. \[19\]. We detail below a few facts that are particularly pertinent to the present discussion.

Each stimulus consisted of a 200 \( \mu \text{s} \) current pulse, applied to the rat fore-paw. Twenty runs per paw comprised trains of stimuli, each over a 4 sec time interval, separated by random-length quiet periods, during which no stimulus is applied, averaging 12 seconds in duration. The length of the inter-stimulus intervals (ISI) between adjacent stimuli remained fixed during the 4 sec 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\} sec, and in the other 10, from \{1/2, 1/4, 1/6, 1/8\} sec\(^2\). Each of the 20 runs in each of the datasets 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

\(^2\) While 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.
single electrode with the maximal average response, which was thus the closest to the somatosensory area.

Using the index \( n=1,\ldots,6 \), 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 datasets, each referred to by the respective \( nS \) notations. Thus, dataset 3L comprises left hemisphere EEG records from rat #3.

### 2.2.2 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 Figure 2.1. 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 Figure 2.1. Details on data preprocessing procedure are provided in Appendix A.

Figure 2.1 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 sec, top two rows). The combined duration of these waveforms is typically well less than 125 ms, with P1 typically lasting \( \sim 10 \) ms, N1 lasting \( \sim 30 \) 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 \( \sim 330 \) ms (third and lower rows in Figure 2.1). For even shorter ISI (\( \leq 200 \) ms, rows 5-8 in Figure 2.1), attenuation may reach points where no coherent P1-N1-P2 waveform is discernible. Strong attenuation is invari-
Figure 2.2: 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 \( \sim 50 \) Hz, for the background signal, and higher than \( \sim 200 \) 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 \( \leq 4 \) dB, are negligible over these frequency ranges.

Ablably manifest in the response to the second stimulus in these 4 sec 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 [50]. 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 chapter.

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 Figure 2.1, 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 $L_{\infty}$ (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 Figure 2.2:

(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 Figure 2.2 were computed for dataset 4L, applying the averaged periodogram approach \([63]\) 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 sec signal obtained by averaging the 6 sec starting intervals of the twenty data records in dataset 4L, recorded immediately before the first stimulus of each respective run. This 6 sec 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 in\(^3\) §2.6.3. 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 Figure 2.2 include that:

- The amplitudes of the three spectra peak at about the same frequency, i.e., at \(\sim 10\) 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.

---

\(^3\)We 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 the ambiguity 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 §2.6.3.
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 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 towards 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 inter-stimulus interval (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 Figure 2.1. 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 Figure 2.2 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 pre-processed data used in Figure 2.2. 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.

2.2.3 The Riera model

Riera’s model, originally presented in [53, 54], 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 Figure 2.3, describes the model’s key structural components: Each of three mass neurons represents a neuronal population, including
Figure 2.3: Riera’s Neural mass model consists of a lumped dynamic representations of Pyramidal Cells (PC) and of two InterNeuron groups (IN), 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 feed-forward IN is represented by $I_3^+$. 

Pyramidal Cells (PC), and two InterNeuron (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 ($I_1^+$), and the excitatory innervations to the transmission GABAergic interneurons ($I_3^+$).

At a system block diagram level, each of the three system blocks in Figure 2.4 represents the dynamical system model of a neuronal group in Figure 2.3, with an indication of states participating in that block. In all the model comprises 9 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. [53, 54]. 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

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

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, \( \tau \), and to affect the shape of the generated waveform components, using the standard deviation \( \sigma \) 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 (2.2.1) is intentional and is the subject of the following cautionary note about a potential notational morass: To facilitate reference to the
Table 2.1: 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 (2.2.1).

<table>
<thead>
<tr>
<th>Wave</th>
<th>Current</th>
<th>Gain, g</th>
<th>Delay, τ</th>
<th>Std. σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>(I_3^+), (I_3^-)</td>
<td>(g_1)</td>
<td>(t_1)</td>
<td>(σ_1)</td>
</tr>
<tr>
<td>N1</td>
<td>(I_1^+)</td>
<td>(g_2)</td>
<td>(t_2)</td>
<td>(σ_2)</td>
</tr>
<tr>
<td>P2</td>
<td>(I_2^+)</td>
<td>(g_3)</td>
<td>(t_3)</td>
<td>(σ_3)</td>
</tr>
</tbody>
</table>

original presentation in Riera et al. [53, 54], we keep the original input current notations, \(I_i^+\), \(i = 1, 2, 3\), as in Figures 2.3 and 2.4. 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 2.1. (The reference to \(\tilde{I}_3^+\) will be clarified below.)

The habituation mechanism, which we describe in §2.3, 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 Figure 2.4 \(f(x_2) \approx 0\) under broad conditions, including the operating range of this study. This observation renders the (non-linear) feedback path in Figures 2.3 and 2.4 redundant, and reduces the original dynamical system from nine to three states. The signal \(\tilde{I}_3^+\), in Figure 2.4, 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 (B.0.12), 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 §2.4.

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 2.1 and to equation (2.2.1), 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, $\sigma_i$, remain constant throughout, and are given in the Appendix.

### 2.3 Habituation control structure

To enable a neural mass model to mimic the effect of habituation, we introduce the schematic control structure in Figure 2.5. 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 §2.2.2, 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 [45, 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 [23]; in particular, these ideas are funda-
mental in the closely related, low order models of firing patterns in single neurons
[22, 59], and in low order models in fluid dynamics [64, 65], 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 $s \geq 0$, 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 stimulus$^4$ at the time $t_s$, the value of $s(t_s + 0.5 \text{ sec})$ should

$^4$We 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 sec train, in our data, satisfies this requirement.
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 sec and 0.5 sec (1 and 2 Hz stimulus trains).

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 $\tau = 0.33$ sec 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 $\tau = 0.166$ sec or shorter. This is dictated by the lack of noticeable evoked response to the second stimulus,
Figure 2.7: The solid curve is one second trajectory of $s(t)$, from equation (2.3.1), 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 Hz to 8 Hz. The heavy arrows on the left vertical axis indicate levels of $s(t_{\text{stim}})$, 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).

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

$$\dot{s} = -\frac{1}{\tau_s}(0.5 + s)s + g_s u(t),$$ \tag{2.3.1}$$

where the input, $u(t)$, is the rectified (i.e., the absolute value of the) EEG signal. Indeed, the effective time constant, $\tau = \tau_s/(0.5 + s)$, in this formulation, decays as a function of $s$. Model parameters were selected as $g_s = 20$ and $\tau_s = 0.5$ sec. 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 Figures 2.6 and 2.7 to illustrate the required properties of $s$ and our considerations in making the choice of its governing dynamics.

Figure 2.6 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 (2.3.1), initiated at $s(0) = 0$. The P1-N1-P2 waveforms are extracted from response data for stimulus trains with ISI of 0.25 sec (4 Hz, left panels) and 0.125 sec (8 Hz, right panels), using a best-fit procedure we elaborate in §2.4. 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\textsuperscript{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, second 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 sec 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

\textsuperscript{5}Considering 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.
new stimulus, at the end of each response cycle.

The data used to produce Figure 2.7 were obtained by the same procedure as in Figure 2.6. The figure shows the continuous 1 sec long trajectory of $s$, following the application of the first, isolated stimulus in a 4 sec stimulus train (solid / red curve). The + signs mark the sampled $s(t_s)$ for eight 4 sec 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 (2.3.1): 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 Figure 2.7, 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 $\tau_s$, 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, $s$.

## 2.4 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 §2.4.2, model parameters are optimized for best data fit, 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 §2.4.3. To put the overall discussion in context, we begin, in §2.4.1, with brief descriptions of the three model variants that will be constructed in this section, and analyzed and discussed in subsequent sections.

2.4.1 Model Variants

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 sec signal, in response to a stimulus train. Model parameters are optimized for best fit (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 (2.3.1). Specifically, the input in (2.3.1), 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 sec 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 sec stretches, in our data, should be essentially identical. Yet, non-negligible variations between these first cycles, however limited, are conspicuous in our data. Since the average wait time between distinct 4 sec stimulus trains is of 12 sec, and since no consistent habituation is observed when the ISI lasts 1 sec, the residual habituation effect of one 4 sec 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 §2.2.2, 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.
Figure 2.8: 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 (2.4.2). 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 dataset. Essentially identical mappings were used for the nonlinear Riera model.

$$y_p(t) = \begin{cases} 
0, & \text{if } t_e \leq t < t_s1 \\
(p_{11}(t - t_{s1})^2 + p_{12}(t - t_{s1})), & \text{if } t_{s1} \leq t < t_{s2} \\
n_{11}(t - t_{s2})^3 + n_{12}(t - t_{s2})^2 + n_{13}(t - t_{s2}) + n_{14}, & \text{if } t_{s2} \leq t \leq t_{s3} \\
p_{21}(t - t_{s3})^3 + p_{22}(t - t_{s3})^2 + p_{23}(t - t_{s3}) + p_{24}, & \text{if } t_{s3} < t \leq t_e \\
0, & \text{if } t_e \leq t \leq 125 \text{ms}, 
\end{cases}$$

(2.4.1a)
2.4.2 Best Fit Model

Parameter estimation for the Best Fit Model (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

\[
\begin{align*}
n_{14} &= p_{11}(t_{s2} - t_{s1})^2 + p_{12}(t_{s2} - t_{s1}), \\
n_{11}(t_{s3} - t_{s2})^3 + n_{12}(t_{s3} - t_{s2})^2 + n_{13}(t_{s3} - t_{s2}) + n_{14} &= p_{24}, \\
p_{21}(t_e - t_{s3})^3 + p_{22}(t_e - t_{s3})^2 + p_{23}(t_e - t_{s3}) + p_{24} &= 0, \\
3p_{21}(t_e - t_{s3})^2 + 2p_{22}(t_e - t_{s3}) + p_{23} &= 0,
\end{align*}
\]

Continuity at P1-N1 connection
Continuity at N1-P2 connection
P2 vanishes at \( t_e \)
The derivative of P2 vanishes at \( t_e \)

Figure 2.9: The average EEG response to trains of 8 Hz stimuli (the most challenging case) from dataset 4L, and the predictions generated by three types of habituation models, using the linear simplification of Riera model: The Best Fit Model (BFM), a Single Cycle Predictive Model (PM1), and Predictive Model (PM).
stimulus cycle. This stage thus serves also the goal of producing a consistent, quantifiable \textit{ground truth} definition of features of that waveform, that will be used in evaluating model performance, later on. Leveraging the latitude, highlighted in §2.3, in making that definition, we select the parameterized family of piecewise polynomial curves, of the form \(2.4.1a\), subject to the continuity constraints \(2.4.1b\), 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 \((2.4.1)\).

The nontrivial part of this optimization procedure concerns finding the start/end times of the P1, N1 and P2 components; i.e. \(t_{si}, i = 1, 2, 3, \) 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 \(t_{si}, i = 1, 2, 3, \) 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.

\textit{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 Riveria’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 \((2.2.1)\) of the input currents \(I^+_i\). Good initial guesses are essential in nonlinear/non-convex optimization heuristics, such as the one employed here. The role of the piecewise polynomial fit, in the overall derivation of a BFM, is precisely to produces such guesses. Thus, the start times \(t_{si}, i = 1, 2, 3, \) 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 piecewise polynomial curve produce initial guesses of the gains,
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 minutes for the entire ensemble of response cycles, in our twelve datasets.

### 2.4.3 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 §2.4.2.

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 Figure 2.8, 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 Figure 2.8, 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 equation (2.4.2). 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 Figure 2.8 for the linear simplification of Riera’s model, were computed, using the MATLAB nonlinear least-squares routine.

\[
\begin{align*}
g_1/g_1(0) &= \begin{cases} 
a_1s + a_2s^{1/2} + a_3, & 0 \leq s \leq 1 \\
e^{-a_4(s-a_5)} & s > 1 \end{cases} \\
g_2/g_2(0) &= \begin{cases} 
b_1s + b_2s^{1/2} + b_3, & 0 \leq s \leq 1 \\
e^{-b_4(s-b_5)} & s > 1 \end{cases} \\
g_3/g_3(0) &= c_1g_2/g_2(0)
\end{align*}
\]

(2.4.2)

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 sec stretches of stimuli and response, excluding the response to the first stimulus in each of these 4 sec stretches. (The respective BFM response is always used during the first cycle.) Having fixed a parameter choice in (2.4.2), 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 sec 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 equation (2.4.2). The average run time of the optimization procedure (same hardware as in §2.4.2) was 160 sec for each of the twelve dataset. Figure 2.9 shows a sample PM1 and PM responses, compared with the BFM response for the linear Riera model.

2.5 Model performance analysis

We begin, in §2.5.1, 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 §2.5.2 we compare ground truth data, as defined in §2.5.1, to features of the response generated by the fully predictive model (PM) and to single-step-ahead predictions (PM1). To put these results in yet another perspective, comparisons are also made to feature extraction by best fit models (BFM), serving as alternative data filters. Finally, leave-one-out cross-validation (LOOCV), presented in §2.5.3, tested model-generated predictions in each of the twelve datasets, using models identified by the data in the remaining sets.
2.5.1 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:

<table>
<thead>
<tr>
<th>Timing Features</th>
<th>Shape Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 peak time ((pt))</td>
<td>P1 peak amplitude ((pp))</td>
</tr>
<tr>
<td>N1 peak time ((nt))</td>
<td>N1 peak amplitude ((np))</td>
</tr>
<tr>
<td>P2 start time ((p2t))</td>
<td>P2 area ((p2a))</td>
</tr>
</tbody>
</table>

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 Figures 2.12, 2.13 and 2.14.

Considering the quantitative criteria, the convention, introduced in §2.4.2, 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 piecewise polynomial curve of the form (2.4.1). The extraction of the six features listed above from the piecewise 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 §2.2.2. 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 piecewise 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-model nature of the PDF, seen in the example, in Figure 2.10, supports our use of the value that separates the two lobes as a natural threshold for the distinction between fire and no-fire cases.

![Figure 2.10: Firing threshold detection using a normalized smoothed histogram of peak-to-peak amplitudes from the second cycle of 6 Hz to 8 Hz from dataset 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.](image)

2.5.2 Model performance

To determine the accuracy of firing detection we computed the precision ($P$, complement of specificity), the recall ($R$, the sensitivity), and the F-score ($F$), which are
common measures in the information retrieval community [38, 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 –

$$
P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F = \frac{2}{1/P + 1/R}. \quad (2.5.1)$$

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

Table 2.2 reports the F-scores for firing predictions. Scores are given separately for each dataset, for each of the BFM, PM1, and PM, and for both the non-linear 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, dataset 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
Table 2.2: F-scores of firing detection.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Linear Riera</th>
<th>Nonlinear Riera</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BFM</td>
<td>PM1</td>
</tr>
<tr>
<td>1L</td>
<td>0.96</td>
<td>0.84</td>
</tr>
<tr>
<td>1R</td>
<td>0.93</td>
<td>0.81</td>
</tr>
<tr>
<td>2L</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>2R</td>
<td>0.92</td>
<td>0.76</td>
</tr>
<tr>
<td>3L</td>
<td>0.94</td>
<td>0.86</td>
</tr>
<tr>
<td>3R</td>
<td>0.97</td>
<td>0.86</td>
</tr>
<tr>
<td>4L</td>
<td>0.99</td>
<td>0.92</td>
</tr>
<tr>
<td>4R</td>
<td>0.96</td>
<td>0.79</td>
</tr>
<tr>
<td>5L</td>
<td>0.96</td>
<td>0.82</td>
</tr>
<tr>
<td>5R</td>
<td>0.93</td>
<td>0.79</td>
</tr>
<tr>
<td>6L</td>
<td>0.94</td>
<td>0.86</td>
</tr>
<tr>
<td>6R</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>mean</td>
<td>0.95</td>
<td>0.83</td>
</tr>
<tr>
<td>std</td>
<td>0.026</td>
<td>0.047</td>
</tr>
</tbody>
</table>

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.

The similarity seen in Table 2.2, between the performance of the linear and the nonlinear models, prevails throughout the selected set of criteria, and across all datasets. To simplify graphic presentations of the quality of feature predictions, these results are presented first, in Figure 2.11, only for the linear model variants, and for a single, representative dataset (4L). Ground truth values continue to be defined by the best piecewise polynomial fit (2.4.1) 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 dataset. 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 piecewise polynomial fit. We argue that this particular spread reflects the difference between the family of curves defined by the piecewise polynomial curves, (2.4.1), 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 (2.4.1), 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.

Figure 2.11: 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 percentile 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.
A summative representation of each of the performance criteria, across all datasets and models types, is provided in Figure 2.12, where BFM, PM1 and PM are based on the linear Riera model. and in a counterpart presentation, in Figure 2.13, 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 predictive models perform comparably well on all datasets. 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 piecewise polynomial curves (2.4.1) and by the family of response curves that can be generated by the Riera model.

2.5.3 Cross validation

Model performance was also tested, using the leave-one-out cross validation approach. These tests were applied to PM performance, as determined by the fourteen parameters of the functional dependencies (2.4.2) of model parameters on s. Specifically

1. Having chosen a dataset, the optimal values of each of the fourteen s-mapping parameters, from the remaining eleven datasets, were averaged. These average values were used in generating the model response for the chosen dataset.

2. Model errors were compared with those generated by using the optimal parameters, for the same dataset.
Figure 2.12: Comprehensive summary of linear Riera model accuracy as reflected by waveform features across all datasets. Each box represents a set of time (left half) or shape (right half) features for the particular dataset 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 datasets from the left brain (right fore-paw) while the bottom six panels show the results for datasets from the right brain (left fore-paw).
This procedure was performed for each of the twelve datasets. Figure 2.14 compares the error using the cross-validation approach with the error obtained by using the optimized model, for each dataset. 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 datasets.

2.6 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.
Figure 2.14: Comparison of the Leave-One-Out Cross-Validation (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.
2.6.1 The phenomena: Evoked response and habituation

Evoked response is characterized and analyzed in terms of the distinctive P1-N1-P2 waveforms. Figure 2.1 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 (Figure 2.1, 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 maintain 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 Figure 2.1, 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 (2.3.1). The stability of (2.3.1), 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 §2.4.2 and 2.4.3, and will be revisited in §2.6.2 and 2.6.3, 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 [14], the evidence in Figures 2.1 and 2.2 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, Figure 2.2 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 Figure 2.1, 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.

2.6.2 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 ensure numerical robustness of parameter estimation and support the objective of avoiding over-fitting. We believe that the constrained least-squares polynomial fit (2.4.1), in §2.4.2, 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 said, 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.

2.6.3 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 Figure 2.10 and to the description in §2.5.1, 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.
2.6.4 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 Figures 2.3 and 2.4, 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 chapter, including parameter variations due to habituation. The conceptual status of our parameterization of $\tilde{I}_3^+$ by a quadratic waveform (equation (B.0.12) 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.
2.6.5 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 §2.6.1, 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 piecewise-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 piecewise 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.

2.6.6 Feature estimates vs. curve fit evaluation

Our precision criteria for model-generated predictions concern six numerical features of the evoked response, i.e., the timing features $p_t$, $n_t$ and $p_{2t}$, and the shape features $p_{pp}$, $n_p$ and $p_{2a}$. 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 [47, 30]. 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.

### 2.7 Concluding Remarks

The chapter 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. [53, 54]. The model was validated using twelve datasets, containing the evoked response to over 1700 stimulus trains, applied to the medial nerve, each lasting 4 sec, with inter-stimulus periods ranging between 125 ms - 1 sec (1 - 8 Hz). Validation included both single-step-ahead predictions (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 neuro-vascular 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 [20, DCM], which uses neural mass models to predict or explain both event-related and steady-state neurophysiological responses. In that context, our predictive model 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 work.
Chapter 3

Neurovascular coupling

3.1 Introduction

Sensory stimulation results in a period of increased electrical activity of neurons and is accompanied by an increase in local blood flow leading to hemodynamic changes. Specifically, neuronal responses lead to an increase in oxygenated hemoglobin (HbO), and a decrease in deoxygenated hemoglobin (HbR), during sensory stimulation. Modeling this "neuro-vascular coupling" is the main topic of this chapter.

The physiology of neurovascular coupling involve complex bio-electro-chemical activity of the neurons and the glial cells (astrocytes) which activate the blood vasculature [26, 69, 31, 42]. Some previous studies modeled these relationships using simple dynamical models [9, 21, 61], but they were not directly validated on experimental data.

Other studies have used the concurrent recording of neural activity using EEG, while monitoring the hemodynamic activity using fMRI or optical imaging, in response to stimulation. Models developed in these studies were data driven and typically involved filtering of the neuronal response or a nonlinear mapping of the neuronal response using FIR (static) [36], or IIR (dynamic) [39, 19] filters. There were also a few data driven studies that developed models using state space representations...
All these models in essence identify a map from a parametrization of the neuronal response to a parametrization of the hemodynamic response under a fixed stimulus condition. Generalizing them to arbitrary stimulus conditions would involve fitting the model parameters to each condition, thus, creating a look up table of model parameters.

The purpose of the present study is to develop a neurovascular model that maintains the essence of the previous models, but at the same time captures non-trivial qualitative changes in neural and hemodynamic signals in response to changes in stimulus duration and frequency.

The approach we take for developing the model is to identify a gray box model that captures the mapping from a parametrization of the neuronal response to a parametrization of the hemodynamic response with a postulated model structure. Here we focus on a model structure which comprises of two components (refer Figure 3.1) that are briefly described below:

i) A physiologically motivated, first principle model of the blood vasculature (hemodynamic model in Figure 3.1) where the parameters have physiological meaning and hence are either known from empirical studies or can be estimated from baseline conditions.

ii) Another component (coupling model in Figure 3.1) that relates the neural responses to the inputs of the hemodynamic model. It turns out that this component can be constructed using a static nonlinear mapping along with a third order linear system.

The procedure used to develop the neurovascular model involves reconstructing the inputs of the hemodynamic model using the data (HbO and HbR), and then identifying the coupling model by using the EEG signals and the reconstructed inputs, for changes in stimulus duration and frequency.
Figure 3.1: EEG evoked neural responses drive the Coupling model which in turn provides the inputs for the hemodynamic model. The hemodynamic model relates the signals from the Coupling model to changes in blood oxygen concentrations (HbO/HbR). The coupling signals are $CMRO_2$ (the metabolic rate of oxygen demand by the surrounding brain tissue), and $g$ (vascular conductance).

The chapter is organized as follows: §3.2 describes the experimental procedure and data processing. §3.3 describes the hemodynamic model and details the procedure for reconstruction of its inputs. §3.4 describes the identification of the coupling model. §3.5 presents the model performance and validation over all the experimental conditions. §3.6 presents a discussion on some aspects of the modeling procedure, and §3.7 concludes the chapter.

3.2 Experimental methods, data collection and description

In this section we detail the experiments performed, the data collection and preprocessing. Then, we discuss the typical waveforms of the data which provides an insight for model development.

3.2.1 Experimental procedure and data collection

The data are concurrent recordings of EEG and optical measurements from the somatosensory cortex of rats, during episodes of medial nerve stimulation. The arrange-
ment of the optical source-detector configuration, and the EEG electrodes are shown in Figure 3.2.

The EEG measurements correspond to the average evoked potential of the synchronous electrical activity of many neurons firing in response to sensory stimulus. The electrodes e1 and e3 which are near the somatosensory cortex were used to collect the EEG data. The optical sources transmit near infrared light at two different wavelengths (690 and 830 nm). The photons from the sources, are scattered, absorbed, and reflected back to be collected by the detectors. Thus, the data obtained are the changes in light intensity.

![Figure 3.2: Head map of EEG electrodes and the optical source-detector arrangement for gathering concurrent measurements from medial nerve stimulation in rats. The EEG electrodes e1 and e3, are positioned near the somatosensory cortex.](image)

The experiments consisted of two stimulus conditions:

- **Frequency experiments**: Stimulus at a constant amplitude current was applied to the rat fore-paw for a fixed 4-second duration with stimulus frequency chosen from a random ordering from 1, 2, . . . , 8 Hz, with the time between the stimulus trains being 12 seconds long on average. The same procedure was repeated ten times for each fore-paw of each of the six rats, resulting in a total of 12 datasets.

- **Duration experiments**: Stimulus at a fixed frequency of 3 Hz was applied to
the rat fore-paw with stimulus duration chosen from a random ordering from 1, 2, . . . , 8 seconds. The same procedure was repeated ten times for each fore-paw of each of the six rats, resulting in a total of 12 datasets. The duration experiments were performed on a different set of rats than those of the frequency experiments.

The raw EEG data was filtered to remove extreme low frequency changes in the baseline, and further filtered to remove noise from the cardiac cycle and 60 Hz power signal. Further processing was performed to improve the signal-to-noise ratio, the details are in [34, 19]. Here, we describe the procedure to identify the source-detector pairs that respond to the sensory stimulus and use them as the hemodynamic data.

We are interested in identifying the source-detector pairs that may contain the hemodynamic response to sensory stimulation. To identify the relevant source-detector pairs we notice that there are 144 (9 sources x 16 detectors) source-detector pairs in total. It is a well established fact that on the cortex the evoked responses occur contralateral to the limb on which the stimulus is applied [3], hence the number of source-detector pairs that may contain any response are 48 (6 sources x 8 detectors) per half brain. Since we are interested in capturing the hemodynamic response in the somatosensory area we only use the data from source-detector pairs that are around this region of the cortex. There are 24 (4 sources x 6 detectors) such source-detector pairs. For each of these 24 source-detector pairs over both experimental conditions for all rats, the procedure followed for computing the hemodynamic data is enumerated below:

1. The optical measurements, which are temporal changes in the intensity were obtained by sampling at the rate of 10 samples per second. The temporal changes in the intensity were translated into temporal changes in the concentrations of oxy-hemoglobin (∆HbO) and doxy-hemoglobin (∆HbR) using the modified Beer Lambert law [12].
2. Spectral analysis of the hemodynamic response revealed that the second lobe of the magnitude spectrum occurred near 3.5 Hz and was 20 dB below the principle lobe. Hence, a low pass filter with a cutoff frequency of 3.5 Hz was applied to the $\Delta$HbO and $\Delta$HbR signals, thus improving the signal-to-noise ratio.

![Figure 3.3](image-url)

Figure 3.3: The left panels show the EEG responses for different stimulus frequencies for a fixed duration of 4 seconds (dataset 4L). The top right panel shows the hemodynamic responses for varying stimulus frequencies (frequency experiment, dataset 4L). The bottom right panel shows the hemodynamic responses for a fixed stimulus frequency of 3 Hz and varying durations of stimulus (duration experiment, dataset 7L). The EEG responses for the duration experiments are identical to the responses for the 3 Hz stimulus (third left panel) over varying durations of stimulus, and hence are not shown. The notation used for the hemodynamic responses are $\Delta$HbO=$\text{HbO} - \text{HbO}_0$, where $\text{HbO}_0$ is the baseline oxy-hemoglobin, similarly $\Delta$HbR=$\text{HbR} - \text{HbR}_0$.

It was also found that the detector that were geometrically nearest to a source had the highest signal strength. Signal strength in this discussion refers to the dominant
singular value of the matrix of the total hemoglobin ($\Delta HbT = \Delta HbO + \Delta HbR$), where the rows are the number of trials, and the columns are the time points. There were 12 such source-detector pairs per dataset.

The final step involved identifying the source-detector pairs among the 12 that could be deemed to have responded to the stimulus, for each dataset. Variations in amplitude and even in amplitude differences prevented the use of a simple amplitude or power threshold, so instead we used a template-based approach designed to identify hemodynamic waveform shapes that corresponded to meaningful physiological responses. For this purpose we took the ensemble mean of the $\Delta HbO$, and $\Delta HbR$ over all of the 12 source-detector pairs, for both halves of the brain, and all the 6 rats for each experiment type (frequency or duration) to represent the generic shape of the hemodynamic response. We then computed the correlation coefficient between the mean $\Delta HbO$, and $\Delta HbR$ of each source-detector pair and the ensemble mean, and selected those source-detector pairs for which the correlation-coefficient was greater than 0.7 for each dataset. The threshold of 0.7 was chosen so as to ensure that for each half brain per rat at least one source-detector pair gets selected for both experiments.

The datasets, thus obtained are denoted in the sequel by nS, where n runs from 1 to 6 for frequency experiments, and 7 to 12 for the duration experiments, to indicate a particular animal, and S is either L or R to denote the left or right brain region.

### 3.2.2 Description of typical hemodynamic responses

Typical waveforms of the EEG responses at different stimulus frequencies are shown in the left panels of Figure 3.3. It is clear from the figure that as the stimulus frequency increases beyond 2 Hz the amplitude of the EEG responses decrease. Further increase in stimulus frequency results in diminished responses, and in some stimulus cycles the responses are not even discernable. These observations suggest an increase in habituation to stimulus with increasing stimulus frequency.
The hemodynamic responses corresponding to the EEG are shown in the top right panel of Figure 3.3. It can be observed that the amplitude of the hemodynamic response increases from 1 to 2 Hz stimulus and then decreases as the stimulus frequency increases beyond 2 Hz.

When stimulus duration is varied (keeping the stimulus frequency fixed at 3Hz) the EEG responses essentially remain the same as the one shown in the 3rd left panel in Figure 3.3. This fact is expected due to the faster time constants involved in neural responses. On the other hand the hemodynamic response (bottom right panels of Figure 3.3) increases with increasing duration of stimulus, and the amplitude saturates for stimulus durations > 2 seconds.

It may be observed from the hemodynamic responses shown in Figure 3.3 that the temporal behavior of the hemodynamic responses are dependent only on the stimulus duration. This observation hints at a model structure with two components, one depending only on the temporal behavior, and the other on the amplitude. We keep this model structure in mind for the development of the coupling model in §3.4. At present we introduce a physiologically motivated hemodynamic model and then trace backwards (refer Figure 3.1) to describe the reconstruction of its inputs in §3.3.2, following the approach mentioned in §3.1.

3.3 Physiologically motivated hemodynamic model

Following the approach to developing a coupling model we introduce a physiologically motivated hemodynamic model, use it to reconstruct appropriate input signals, which are the desired outputs of the coupling model. These signals along with the EEG signals are then used to identify an appropriate coupling model in §3.4. In this section we describe the Windkessel hemodynamic model we adopt, as well as the procedure we used to reconstruct its inputs from the hemodynamic data ($\Delta$HbO and $\Delta$HbR).
3.3.1 The model structure

Figure 3.4: The block diagram of the single compartment Windkessel model. The state $v$ represents the vascular volume, the coupling signals $x_2$ and $x_4$ are states of the $O_2$ transport subsystem which represent the concentration of oxygen in the blood vessel, and the pial compartment respectively.

The hemodynamic model that we introduce in this section falls into the class of Windkessel based models which were first applied to study hemodynamic responses in [10, 37]. Following those initial papers several Windkessel based models of varying complexity single compartment [72], three compartment [71, 28], and compartmental network model [4] have been developed.

Windkessel models of blood vasculature constitute of two subsystems, i) the blood flow equations, and ii) the oxygen transport from blood vessel to tissue. The flow subsystem (refer Figure 3.4) is modeled by assuming that blood flow is laminar and that the blood vessel is cylindrical. The assumption of laminar flow enables us to relate the change in vessel volume to the flow in and out of the vessel. The assumption of the cylindrical shape enables us to relate the change in vessel volume to the change in diameter of the vessel via Poiseuille’s law. The oxygen transport on the other hand is driven by the metabolic rate of $O_2$ demand by the local brain tissue ($\Delta CMRO_2$).

The model we present here is a single compartment version with the following assumptions:

1. The volume of arterial compartment does not change.

2. A single state $v$ represents the vascular volume and combines the volumes of capillaries and vein.
3. The pial compartment is assumed to have constant volume.

4. An additional feature of the model is that the change in oxygen concentration of the blood vessel is driven by the difference in the partial pressures of oxygen between the blood vessel and the surrounding tissue using Fick’s law [4].

This last assumption leads to the model being highly sensitive to the $\Delta C_{MRO_2}$ input (refer Appendix C.2.2 for the sensitivity analysis), which is necessary in order to reconstruct it from hemodynamic data.

Figure 3.4 shows the block diagram of the single compartment Windkessel model. The model consists of nine parameters and four states. A single state represents the vascular volume $v$ (hence single compartment), and three states ($x_2$, $x_3$ and $x_4$) represent the oxygen concentration in the vessel, the number of free oxygen molecules in the tissue, and the oxygen concentration in the pial compartment. The state equations and details of model parameters can be found in Appendix C.1.

### 3.3.2 Reconstruction of the inputs to the hemodynamic model from data

The inputs of the hemodynamic model are required in order to identify the neurovascular coupling model. Hence, the reconstruction of the two inputs (vascular conductance $g$, and oxygen demand $\Delta C_{MRO_2}$), from optical data is described here.

**Reconstructing $g$ from hemodynamic data**

The $g$ input can be reconstructed by a direct inversion of the flow equation given in equation (C.1.1a) in the Appendix C.1.

\[
\dot{g} = \frac{2\sqrt{g}}{(1 - \phi_v)}(\dot{h} - \phi_v \dot{v})
\]

\[
\dot{v} = g(g_0 + (1 - g_0)v^\beta) - v^{\beta+2}
\]
The pair of equations (3.3.1) are driven by the time derivative of the normalized HbT \( h \). The procedure to compute the derivative of the normalized HbT is detailed here. First, we formed \( \Delta \text{HbT} = \Delta \text{HbO} + \Delta \text{HbR} \), and then formed the normalized HbT \( h = \Delta \text{HbT}/\text{HbT}_0 + 1 \), where \( \text{HbT}_0 \) is the baseline HbT value. To estimate the derivative of \( h \), we used a cubic B-spline representation of \( h \), and then differentiated the B-splines to obtain a smooth estimate of \( \dot{h} \).

**\( \Delta \text{CMRO}_2 \) reconstruction from the hemodynamic data**

![Diagram](image)

Figure 3.5: An example illustrating the reconstruction of the inputs of the hemodynamic model from \( \Delta \text{HbO} \) and \( \Delta \text{HbR} \). Known signals of \( g \) and \( \Delta \text{CMRO}_2 \) (solid blue curves in the top left and bottom right panels) drive the hemodynamic model to generate the data (\( \Delta \text{HbO}, \Delta \text{HbR} \) and \( \Delta \text{HbT} \)) in the top right panel. \( \Delta \text{HbT} \) is used to drive equation (3.3.1) to recover \( g \). \( \Delta \text{CMRO}_2 \) is recovered from the hemodynamic data by optimizing the weights in equation (3.3.2). The reconstructed signals are marked by red ‘+’ signs.
Recovering the $\Delta$CMRO$_2$ input is difficult due to the presence of nonlinearities which prevents direct inversion of the oxygen transport subsystem (see Figure 3.4). However, we can convert the problem into an optimization problem by using

$$\Delta\text{CMRO}_2 = \sum_{i=1}^{N} w_i B_i(t),$$

(3.3.2)

where $B_i(t)$ are any set of basis functions, for example cubic B-splines, and $w_i$ are the weights that are estimated by minimizing the l-2 norm of the error between the model and the data.

The temporal support of the B-splines and the order ($N$ in the equation above) depend on the following two criteria, i) the goodness of fit, and ii) the temporal resolution. Low temporal resolution leads to worse fits. On the other hand, a very high temporal resolution may lead to overfitting and oscillatory behavior of the waveshape. Thus, a tradeoff exists between the two extremes. For our datasets we found that the delay times involved were about 1 to 1.5 seconds following the stimulus, and hence we choose the cubic B-spline to have a support of 1.5 seconds, which resulted in $N = 28$ over a time span of 14 seconds. We reconstructed the $\Delta$CMRO$_2$ by estimating the weights of the B-spline basis using lsqnonlin from MATLAB which implements a nonlinear trust region based method for parameter optimization.

Figure 3.5 illustrates the reconstruction of $g$ and $\Delta$CMRO$_2$ for a simulated example. The hemodynamic data is generated by known signals of $g$, and $\Delta$CMRO$_2$, and in turn the data is used to reconstruct the signals using the procedure described here and in §3.3.2. The reconstructed curves (red ’+' signs) are overlaid on the true curves (solid blue curve). This procedure was used to estimate $g$ and $\Delta$CMRO$_2$ for all datasets.
3.4 Development of the coupling model

As outlined earlier, our overall strategy for developing a model to couple the EEG (or equivalently the output of a neural mass model) and the hemodynamic model just described is to work from "both ends" towards the middle. Specifically, on the input side we first put the EEG through a low-pass filter with a time constant long enough to match the time scales of the filtered EEG with the inputs of the hemodynamic model. The next step is to parameterize the range of filtered outputs by finding a suitable set of temporal basis functions and, on the output side, doing the same for the (reconstructed) inputs of the hemodynamic model. Finally, the temporal basis functions on both sides are related via a dynamical system, and based on that relationship the amplitudes at both ends are matched via static mappings. We elaborate on this procedure next.

![Block diagram of the coupling model](image)

Figure 3.6: Block diagram of the coupling model. The rectified EEG response is filtered with a 1 second long averaging filter. The filtered response is represented by a scalar $a$ and temporal basis function $V_D(t)$. The temporal model relates $V_D$ to $\hat{V}_{hi}$ with $i \in \{1, 2\}$, and static mappings $f(a)$ map $a$ to $a_j$, with $j \in \{c, g\}$, to correspond to the two inputs to the hemodynamic model. A linear combination of the temporal functions $\hat{V}_{hi}$, using $a_j$, results in the signals $g$, and $\Delta CMRO_2$, which drive the hemodynamic model.

Figure 3.6 shows the structure of the proposed coupling model. First, the EEG response is filtered and projected onto a suitable temporal basis function $V_D(t)$ which yields an amplitude parameter denoted by the scalar $a$. The model is now split into two components, namely:
• A temporal model, driven by $V_D(t)$ whose output is a pair of temporal basis functions $(\hat{V}_{hi}(t)$, where $i = \{1, 2\})$, and

• A static functional mapping, that takes scalar $a$ and produces $\hat{a}_j$, where $j = \{c, g\}$, which weight the $\hat{V}_{hi}(t)$ signals to form the $\Delta$CMRO$_2$ and $g$ inputs of the hemodynamic model respectively.

The identification of the temporal model requires the basis function $V_D(t)$ from the filtered EEG responses, and $V_{hi}(t)$, where $i = \{1, 2\}$ from the (reconstructed) inputs of the hemodynamic model. Selecting a suitable set of basis function $V_D(t)$, and $V_{hi}(t)$ is the first topic of description. Followed by it the structure of the temporal model is presented in §3.4.2. Finally, the static mapping functions are described in §3.4.3.

### 3.4.1 Selecting $V_D(t)$ and $V_{hi}(t)$

The hemodynamic responses shown in Figure 3.3 provides a guideline to develop a procedure for selecting a suitable set of temporal basis functions. Focusing on the stimulus frequency experiments (top right panel in the figure), the shape of the hemodynamic responses across all stimulus frequencies from 1 to 8 Hz suggest that a pair of sinusoidal temporal basis functions could be used to characterize all of them. In other words the plots suggest that a single pair of temporal basis functions could represent the hemodynamic responses for a fixed stimulus duration of 4 seconds. The plots for the stimulus duration experiments with a fixed frequency of 3 Hz (bottom right panel) suggest that the temporal basis expands as the duration of stimulus increases, i.e. for each stimulus duration we would require a pair of temporal basis functions.

These two observations suggest that a suitable set of temporal basis functions can be found from the duration experiments, and once found the temporal basis functions corresponding to the 4-second duration experiment can be used for all stimulus frequency experiments. Hence we only used the duration experiments to find the basis functions...
functions $V_D(t)$, and $V_{h1}(t)$, and from now on regard the suffix $D$ to stand for the duration length, for example $V_4(t)$ corresponds to the 4-second duration.

With the above description in mind the procedure followed to construct $V_D(t)$ is given below:

- The rectified EEG data was low pass filtered to match the slow dynamics of the vasculature. The low pass filter used was a 1-second long quadratic box-spline.
- The filtered EEG responses from the duration datasets were collected in a matrix, with rows representing the number of datasets and columns marking the time points. Each matrix corresponded to one stimulus duration from $\{1, \ldots, 8\}$ s. A singular value decomposition (SVD) was then performed on each matrix, and the right singular vector corresponding to the maximum singular value was taken to be $V_D(t)$.

The procedure followed to construct $V_{h1}(t)$, and $V_{h2}(t)$ is given below:

- The inputs of the hemodynamic model, $g(t)$ and $\Delta CMRO_2(t)$ were recovered from the hemodynamic data as per the procedure described in § 3.3.2.
- The $\Delta CMRO_2$ and $g$ data were collected in matrices for each stimulus duration from 1 second to 8 seconds.
- The SVD of the $\Delta CMRO_2$ and $g$ matrices were performed for each stimulus duration.

It was found that the first two dominant modes captured 90% of the energies for both signals for each stimulus duration. We also found that the inner product between the basis functions of $\Delta CMRO_2$ and $g$ exceeded 0.89 for all stimulus durations, meaning that they were highly correlated with each other and thus largely redundant. Hence we used a single set of basis functions $V_{h1}(t)$ and $V_{h2}(t)$ for both $\Delta CMRO_2(t)$ and $g(t)$. 
Figure 3.7: The temporal model fits to the basis functions extracted from the duration datasets. Panel a) shows $V_D(t)$ for each stimulus duration. Panels b) through g) show $V_{hi}(t)$, where $i = \{1, 2\}$. The basis functions extracted from the duration datasets as described in §3.4.1 are shown as solid curves. The solution of the temporal model in equation (3.4.1) driven by $V_D(t)$ are shown as dashed curves($\hat{V}_{hi}$) in panels b) through g).

### 3.4.2 The temporal model

The temporal model can be identified from the temporal basis functions of the filtered EEG data $V_D(t)$, and the temporal basis functions ($V_{h1}$, and $V_{h2}$) of $g$ and $\Delta CMRO_2$. The basis functions formed using the procedure in §3.4.1 are shown in Figure 3.7, which shows $V_D$ (panel a)), and $V_{h1}$, and $V_{h2}$ (solid curves in panels b) through g)), for different stimulus durations. The solid curves in the figure guide the development of the temporal model. The model consists of three states, the first two form a damped oscillator. A third state is used as a feedback path in order to sustain the output even
after the input has died out. Since the time scales of the filtered EEG responses and the vasculature are similar, the model consists of a direct path from input \( V_D(t) \) to output \( \hat{V}_{h1}(t) \) and \( \hat{V}_{h2}(t) \). The outputs of the temporal model \( \hat{V}_{h1} \) and \( \hat{V}_{h2} \) are shown as dashed curves in panels b) through g) in Figure 3.7. The third order linear system is given in equation (3.4.1). The system has a total of 17 parameters that were estimated from \( V_D \) and \( V_{hi} \) using lsqnonlin in MATLAB. Equation 3.4.1 is a stable system provided \( \alpha_1, \alpha_6, \) and \( \alpha_7 \) are positive. The structure of the state transition matrix allows for a complex eigenvalue pair and a negative real eigenvalue.

\[
\begin{bmatrix}
\dot{x}_1 \\
\dot{x}_2 \\
\dot{x}_3
\end{bmatrix} =
\begin{bmatrix}
-\alpha_1 & \alpha_2 & \alpha_3 \\
-\alpha_5 - \alpha_6 & 0 \\
0 & 0 & -\alpha_7
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} + 
\begin{bmatrix}
-\beta_1 \\
-\beta_2 \\
\beta_3
\end{bmatrix} V_D(t)
\]

\[
\begin{bmatrix}
\hat{V}_{h1}(t) \\
\hat{V}_{h2}(t)
\end{bmatrix} =
\begin{bmatrix}
c_{11} & -c_{12} & -c_{13} \\
-c_{21} & c_{22} & c_{23}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} + 
\begin{bmatrix}
d_1 \\
d_2
\end{bmatrix} V_D(t) \quad (3.4.1)
\]

3.4.3 The static functional mapping

Given the temporal basis functions we can now relate the amplitudes \( a \) obtained by projecting the filtered EEG signals on \( V_D(t) \) to the weights \( a_j \), with \( j = \{c, g\} \), obtained by projecting \( \Delta CMRO_2(t) \), and \( g(t) \) onto \( V_{hi} \), where \( i = \{1, 2\} \). We can relate \( a \) to \( a_j \) using a quadratic polynomial. The coefficients of the polynomial have to be estimated by taking the pseudo-inverse of the Vandermonde matrix whose rows are stimulus frequency or duration ordered from 1 to 8, and columns are \( a^2, a, \) and 1, where each \( a \) corresponds to the amplitude of that particular stimulus duration/frequency. In general Vandermonde matrices are known to suffer from high condition numbers especially when \( a \gg 1 \), which is the case with our datasets. Hence, to alleviate this
Figure 3.8: Relation between the amplitudes $a$ of the filtered EEG responses and the amplitudes of $\Delta$CMRO$_2$ ($a_c$). The left panels show the two amplitudes $a_{c1}$ and $a_{c2}$ of the first two dominant basis (in blue), and $\hat{a}_c = f_c(a)$ (in red) averaged over the frequency datasets 1L to 6R. The right panels show the same for the duration datasets 7L to 12R. The errorbars show the variation across datasets.

problem we used a set of orthonormal polynomials

$$
p_0 = \frac{1}{\sqrt{M}},
$$

$$
p_1 = \frac{\sqrt{3}}{M^{3/2}} \left( a - \frac{M}{2} \right), \text{ and}
$$

$$
p_2 = \frac{\sqrt{5}}{M^{5/2}} \left( a^2 - Ma + \frac{M^2}{6} \right),
$$

in the interval $(0, M)$, where $M = 250$ is chosen such that $M > \max a = 204$ over all datasets of both experiments. For ease of notation we denote $\hat{a}_{c1} = p_2a_i + p_1b_i + p_0c_i$ as $a_c = f_c(a)$, and similarly $a_g = f_g(a)$. The parameters $a_i$, $b_i$, and $c_i$, with $i = \{1, 2\}$, can be estimated from $a_c$ using least squares method. Similarly the six parameters of
the mapping \( f_g(a) \) can be estimated from \( a_g \). The relation \( \hat{a}_c = f_c(a) \) captures the variations in the amplitudes of the \( \Delta \text{CMRO}_2 \) as can be seen from Figure 3.8. The relation \( \hat{a}_g = f_g(a) \) similarly captures the variations in the amplitudes of \( g \).

## 3.5 Results

In this section the neurovascular model performance results for both the duration and frequency experiments are discussed. We present the model error using the optimum parameters, and then validate the model on all the datasets. In what follows the model error using the best parameter set is denoted as Best Fit Error (BFE), and the model error using the validation procedure discussed below is denoted as Leave Two Out Cross Validation Error (LTOCVE). The performance metric \( E \) used is given in equation (3.5.1). The measure \( E = 0 \) when the model output matches the mean data, and \( E > 0 \) when there is a mismatch.

\[
E = \frac{\text{Estimation error of the model}}{\text{Standard deviation of the data}} - 1 = \frac{\| H_{\text{trials}} - H_{\text{model}} \|}{\| H_{\text{trials}} - H_{\text{mean}} \|} - 1. \quad (3.5.1)
\]

We begin by showing the best model fits for a sample dataset each taken from the duration experiments (refer Figure 3.9(a)), and the frequency experiments (refer Figure 3.9(b)). Both figures show that the model output is close to the mean curves of both duration and frequency experiments.

In order to validate the model, we follow the method of Leave Two Out Cross Validation (LTOCV) over all the datasets.

- The procedure for LTOCV is to leave out the data from the chosen rat, for example if the model has to be validated for dataset 1R, then the model parameters are estimated by removing datasets 1L and 1R.
- The datasets had different amplitude scales due to inherent variations among animals, and hence to normalize them we determined a multiplying factor.
Figure 3.9: Sample of model fits to HbO, and HbR from the duration and frequency experiments. The blue circles follow the mean curve and the error bars show the standard deviation.
• The multiplying factor for each dataset is the reciprocal of the maximum singular value across the 8 stimulus frequency/duration conditions for the $\Delta HbT = \Delta HbO + \Delta HbR$ curves.

• This process eliminates the scaling effects of different datasets and allows cross validation.

• Finally we compute the error metric $E$ between the model and the dataset, adjusted by the multiplying factor. This error is denoted as LTOCVE.

This procedure is repeated for all datasets for both the duration and frequency experiments.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Frequency Exp.</th>
<th></th>
<th>Duration Exp.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BFE</td>
<td>LTOCVE</td>
<td>BFE</td>
<td>LTOCVE</td>
</tr>
<tr>
<td>1L</td>
<td>0.013</td>
<td>0.027</td>
<td>7L</td>
<td>0.005</td>
</tr>
<tr>
<td>1R</td>
<td>0.005</td>
<td>0.089</td>
<td>7R</td>
<td>0.014</td>
</tr>
<tr>
<td>2L</td>
<td>0.018</td>
<td>0.035</td>
<td>8L</td>
<td>0.020</td>
</tr>
<tr>
<td>2R</td>
<td>0.020</td>
<td>0.040</td>
<td>8R</td>
<td>0.007</td>
</tr>
<tr>
<td>3L</td>
<td>0.013</td>
<td>0.030</td>
<td>9L</td>
<td>0.009</td>
</tr>
<tr>
<td>3R</td>
<td>0.020</td>
<td>0.049</td>
<td>9R</td>
<td>0.031</td>
</tr>
<tr>
<td>4L</td>
<td>0.025</td>
<td>0.041</td>
<td>10L</td>
<td>0.007</td>
</tr>
<tr>
<td>4R</td>
<td>0.007</td>
<td>0.020</td>
<td>10R</td>
<td>0.016</td>
</tr>
<tr>
<td>5L</td>
<td>0.015</td>
<td>0.030</td>
<td>11L</td>
<td>0.005</td>
</tr>
<tr>
<td>5R</td>
<td>0.010</td>
<td>0.038</td>
<td>11R</td>
<td>0.009</td>
</tr>
<tr>
<td>6L</td>
<td>0.010</td>
<td>0.028</td>
<td>12L</td>
<td>0.016</td>
</tr>
<tr>
<td>6R</td>
<td>0.014</td>
<td>0.035</td>
<td>12R</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 3.1: Performance of the neurovascular model across frequency and duration experiments. For each experiment type the Best Fit Error (BFE), and the Leave Two Out Cross Validation Error (LTOCVE) are tabulated. The LTOCV method involves removing the two datasets corresponding to a chosen rat. Estimating model parameters from the remaining datasets, and evaluating $E$ for the chosen dataset.

Table 3.1 records $E$ for each dataset from both duration and frequency experiments. For each experiment type the first column shows the error for the best set of parameters per dataset, and the second column shows the error using the Leave Two Out Cross Validation (LTOCV) method, i.e. by removing the two datasets of the right and left hemispheres of the brain. From the table it is clear that the maximum
error is about 10% (dataset 1R) and so we conclude that model performs well across all datasets.

3.6 Discussion

We discuss two different issues in this section. The first one about the structure of the neurovascular model and its advantages. The second one dealing with the sensitivity of the hemodynamic model to inputs and baseline parameters.

3.6.1 Essence of the neurovascular model

The neurovascular model structure essentially uses a scalar functional of the neuronal response to characterize the hemodynamic response, via the inputs of the hemodynamic block. This characterization is performed by mapping a scalar representing the strength of the neuronal response ($a$) to the weights of the $V_h$ basis functions for $\Delta CMRO_2$ and $g$ (see Figure 3.8).

If we had taken the direct approach (i.e. chosen to drop the hemodynamic block) to relate the neuronal and hemodynamic responses, then a similar relation between the amplitudes would have emerged. In fact, analysis based on the SVD of HbO and HbR responses from all the datasets revealed this to be true (see the top panels of Figure 3.10). To strengthen our claim we point the reader to Figure 5 of [19] that shows the relation between the integrated Somatosensory Evoked Potential ($\sum \text{SEP}^2$) square and the integrated hemodynamic response ($\sum \text{HbO}$) (first and third figures in the right panel), which is similar to the Figure 3.8 in this study. We reiterate that this direct approach leads to a similar number of states of the neurovascular model with a disadvantage of not being motivated by physiology (refer §3.1).

Another interesting feature that we notice from Figure 3.10 and Figure 5 of [19] is that the amplitude relations over the frequency experiments show a gamma function type dependence across stimulus frequencies ($fe^{-f}$, where $f$ represents stimulus
Figure 3.10: Relation between the amplitudes $a$ of the filtered EEG responses and the weight $w_{h1}$ obtained by projecting the HbO data onto the dominant temporal basis found via SVD analysis. Frequency experiments (mean of 1L to 6R) are on the left panels, and duration experiments (7L to 12R) on the right. The bottom panels show a nonlinear relation between $w_{h1}$ and $w_{h2}$, where $w_{h2}$ are the weights corresponding to the second dominant singular value.

frequency), and $1 - e^{-t}$ (where $t$ represents stimulus duration) type relation across stimulus durations. This feature warrants further experimental investigation of habituation behavior across the time-frequency plane.

Overall we may summarize that irrespective of the modeling approach taken we find that a single scalar functional is sufficient to characterize the hemodynamic responses for both the frequency and duration experiments. This fact also suggests that the choice of the low pass filter (made in §3.4.2) does not affect the model performance, and may be replaced by any other integrator of the rectified neuronal
response.

3.6.2 On the advantage that the model structure provides

The neurovascular model in combination with a mass neuron model enabled with habituation prediction (such a model was discussed in [34]) would be capable of predicting both the EEG and hemodynamic responses for arbitrary stimulus patterns. The prediction ability relies on the static mapping in §3.4.3 remaining true for arbitrary stimulus patterns. An experimental investigation of the dependence of habituation behavior across the stimulus time-frequency plane would enable us to improve the static mappings in order to make the model predict EEG and hemodynamic responses under random stimulation.

Further experiments would also help improve the temporal model structure, while maintaining the same framework. Thus, the framework used for model development is sufficiently generic to capture complex habituation patterns, while maintaining a simple model structure.

3.6.3 On the sensitivity of the hemodynamic model

We now focus on the second issue of discussing the sensitivity of the hemodynamic model to the baseline parameters and inputs. We first describe the sensitivity to baseline parameters and then the sensitivity to inputs.

Sensitivity to baseline parameters

The hemodynamic model uses nine baseline parameters that depend on the conditions that exist prior to stimulation. Physiological variations in each individual rat may change the values of these parameters and could affect model performance. Hence, it is important to analyze the sensitivity of the model to these parameters.

We carry out this analysis in §C.2.1, which reveals that the model is most sensitive
to the saturation of oxygen in the artery and vein. Thus, if baseline information is available then the baseline values of SaO2 and SvO2 could be estimated using the model, else they could be fixed based on past studies since they have a small physiological range [70]. The model is relatively less sensitive to other parameters and hence empirical values from previous studies could be used for all other parameters.

**Sensitivity to inputs**

The method used to identify the coupling model requires reliable reconstructions of the inputs to the hemodynamic model. For this purpose we require that the hemodynamic model be sensitive to the inputs. The model is clearly sensitive to \( g \) input which can be reconstructed by direct inversion using equation (3.3.1). Hence, we only discuss the model sensitivity to the \( \Delta\text{CMRO}_2 \) input which can only be obtained via nonlinear optimization, as was mentioned in §3.3.2.

The sensitivity analysis to the \( \Delta\text{CMRO}_2 \) input carried out in §C.2.2 reveals that an error at the input results in twice that error in the output. Any least squares based optimization scheme attempts to reduce the error between model output and data. Since the error at the output is twice the error at the input, the optimization scheme will inherently compute a reliable estimate of the input. Hence, it could be concluded that the model is sensitive to the \( \Delta\text{CMRO}_2 \) input.

The sensitivity of the model to the \( \Delta\text{CMRO}_2 \) input is a direct consequence of the physiologically motivated feature of the model, that the partial pressure of oxygen between tissue and the vessel drives the oxygen transport (the term \( f(x_2) - \alpha_b x_3 \) in equations (C.1.1b), (C.1.1c), and (C.1.1d)).

### 3.7 Conclusion

In this chapter we developed a neurovascular model that relates habituated EEG data to hemodynamic responses under varying conditions of medial nerve stimulation
in rats. The neurovascular model comprised of a single compartment hemodynamic model, and a coupling model. The model development procedure required reconstruction of the hemodynamic model inputs and so we discussed the issue of sensitivity, and found that the hemodynamic model is sensitive to its inputs (a detailed analysis is presented in §C.2). The coupling model was identified from the EEG data and the reconstructed inputs. The results suggest that the model performance is good across both the experiments.
Chapter 4

Model reduction

4.1 Introduction

A unified system model of the human brain that incorporates morphological, neurochemical and functional variables is a pinnacle objective of neuroscience and a large-scale scientific initiative [67]. The human brain’s structure and function may be modeled as systems, networks, and single units at multiple timescales. A multimodal paradigm, where measurements from imaging modalities like EEG, MEG, fMRI etc. are used to estimate the relevant parameters involved has been of recent interest to the neuroscience community.

A state space model is well suited for an integrated multimodal approach to functional brain modeling. Dynamic analysis methods using the state space approach have been applied to EEG, fMRI [54], and DOT [18], demonstrating improved resolution, dynamic signal tracking, and physiological modeling. The widespread entry of modern control theory into functional neuroimaging is primarily impeded by the lack of a generally accepted state-variable model of the brain. Another factor which is crucial for progress is control over the dimension of the states that need to be estimated; more complex models for individual structures in the brain would result in enormous number of unknown states that need to be estimated.
Of particular practical interest is improving the spatiotemporal resolution and physiological accuracy of multimodal functional neuroimaging with concurrent measurements. The number of parameters involved in such a model increases exponentially, and hence computing them becomes an arduous task. If such approaches are to be made useful and feasible to the neuroscience community, then the issue of dimensionality reduction has to be addressed. Apart from offering computationally feasible models, identification of important states would also help in understanding hidden brain structures and relationships between the different structures.

Dimension reduction methods have been applied previously in other areas like weather pattern prediction based on oceanography [41], and tidal flow forecasting [68], based on subspace approaches. In this paper we explore some dimension reduction approaches to ameliorate the problems faced in a multimodal approach to brain imaging.

Section 4.2 gives a brief description of our dynamic multimodal framework. Section 4.3 describes the underlying theory of different dimension reduction methods. Section 4.4 describes the simulation experiment procedure, and order reduction results. Section 4.5 concludes the paper with some discussion and issues for future work on the topic.

4.2 The Dynamic multimodal Framework

The Dynamic Multimodal Framework is a multiscale-system framework for dynamic functional modeling and multimodal integration (Figure 4.1). The input to the system is a sensory stimulus, which activates the sensory response block (N), which in turn transfers this information to the neuronal block, in the form of neuronal currents. The neuronal block (M) processes this information and activates the vasculature in that part of the brain, via the neurovascular coupling block (J). The neuronal response is measured with EEG and MEG. Cerebral auto-regulation in response to
blood pressure changes etc, is modeled by the autonomic response block (V), and thus the Hemodynamics (H) is driven by both neuronal activation and cerebral autoregulation. The hemodynamic activity is observed by modalities like DOT and fMRI.

The biophysics of the imaging modalities is captured in the lead field functions \( F \). The \( F_{\text{EEG}} \) and \( F_{\text{MEG}} \) lead fields estimate electric and magnetic fields at the scalp surface generated by postsynaptic currents within the brain. This lead field computation requires knowledge of the probe positions, anatomical structure of the head, the electrical conductivity, and the magnetic permeability of the biological tissues. The \( F_{\text{DOT}} \) lead field estimates the photon fluence measured at the scalp surface from an arrangement of laser sources. Photon transport in biological tissue is characterized by the optical absorption and scattering properties of the tissue, which are functions of the dominant chromophore concentrations. The specifics of the \( F_{\text{MRI}} \) lead field are dependent on the particular scanning procedure used. Blood oxygen level dependent (BOLD) signal contrast derives largely from fractional venous blood
CHAPTER 4. MODEL REDUCTION

Figure 4.2: Normalized sensory response

volume and deoxyhemoglobin content.

Since this framework separates the different functions into interdependent blocks, each having its own model, we can replace a simple model with a complex one and vice versa without changing the structure of the system as a whole. Below we briefly describe what goes into each of the blocks in Figure 4.1. The autonomic response block (V) has not yet been implemented.

4.2.1 Sensory block

The sensory response is modeled as a cascade of four second order systems. A stimulus and the three neuronal currents are shown in Figure 4.2.

4.2.2 The Neuronal mass model

In the present paper we use the neuronal model developed by Riera et. al [53]. A basic micro-network composed of the layer V Pyramidal Cell (PC) in the cerebral cortex connected to two types of Gamma Amino-Butyric Acid (GABA) interneuron (IN) inhibitors summarizes a neuronal mass in the cortical circuitry (see Figure 4.3). These GABAergic INs differ in terms of the inhibitory effect they produce locally on the cortex, which could be due to feed-forward or feedback type connections.

Riera’s model is a nonlinear model which takes in three excitatory currents $I_1^+, I_2^+,$
and $I_3^+$, which model the nature of the connectivity patterns with other equivalent cortical units. These inputs/outputs, associated with large-scale synaptic currents, have dimensions comparable to Pico Amperes (pA), and they are always set up as excitatory, with values between 0 and 1. $I_1^+$ represents the excitatory inputs to the layer V PCs from deep within the brain, from areas like the thalamus. $I_2^+$ represents the neuron to neuron dendritic currents in the cortex, and $I_3^+$ represents the excitatory innervations to the transmission GABAergic IN. The stimulus and the voltages which drive the neurovascular block ($J$) are shown in Figure 4.4. The voltages are states in Riera’s model and are not shown in Figure 4.3.
4.2.3 The neurovascular response

The neurovascular block was modeled as a cascade of two second order systems. The outputs are the cerebral metabolic rate of oxygen (CMRO$_2$) and the rate change in arterial resistance ($dR_a/dt$), which serves as our flow inducing signal. A typical response is shown in Figure 4.5.

4.2.4 Three compartment windkessel model

Local vasculature in the brain has been modeled previously with the balloon [8] and windkessel [37] models. In this balloon model there is a compartment which has a compliance (similar to capacitance) and resistance. As flow enters the compartment the vessel expands. The earliest models were single compartment models; however, recently it has been suggested that single compartment models are insufficient in capturing the hemodynamics in response to brain activation [74]. Hence we use a recently developed three compartment model [28]. The model has two components (see the two blocks in Figure 4.6) one for blood flow and the other for oxygen transport. Both are modeled as having three compartments, the arteries, capillaries, and veins. Oxygenated blood flows through the arteries into capillaries, and deoxygenated blood flows out of the veins. The model is detailed in [28] (Figure 4.6). A typical response of change in oxygenated hemoglobin concentration (HbO$_2$) and deoxygenated
Figure 4.6: (top) circuit model for flow (bottom) compartment model for $O_2$ transport to tissue

Figure 4.7: hemodynamic response

hemoglobin concentration HbR is shown in Figure 4.7.

4.2.5 Combining all the models

Since the neuronal and hemodynamic models are nonlinear we have a nonlinear system in general. We linearize the system by taking sufficiently small time steps. This method is called local linearization. Thus we get a linear time varying state-space
model given in equation 4.2.1.

\[
\begin{align*}
\dot{x}(t) &= A(t)x(t) + B(t)u(t) + w(t) \\
y(t) &= C(t)x(t) + v(t)
\end{align*}
\] (4.2.1)

A sample dataset was simulated by adding Gaussian noise to the measurements generated by the forward models for EEG, MEG and DOT. Some channels are shown in figures 4.8, 4.9 and 4.10.
CHAPTER 4. MODEL REDUCTION

4.3 Dimension reduction methods

Dimension reduction can be performed in two steps. The first is to build physical models which have few redundancies, i.e. the states in the model are detailed enough to explain the data we observe. This step depends on the physics of the problem and the complexity needed for that application.

The second step involves using standard subspace based model reduction procedures to further reduce the redundancies in the first model. This also gives us information about the various interactions between different physical models which are combined in a state-space framework in the first step, and thus reveals any coherent structure in our model.

Subspace based approaches try to concentrate the state dynamics into a small number of dominant states. Mathematically we want to find a \( Z(t) \) and \( V(t) \) in equation 4.3.1, of dimensions \( n_x \times m_x \), where \( n_x \) is the number of states in the original system and \( m_x \) is the reduced number of states \( (m_x << n_x) \)

\[
\dot{x}(t) = Z^T(t)A(t)V(t)\dot{x}(t) + Z^T(t)w(t)
\]
\[
y(t) = C(t)V(t)\dot{x}(t) + \nu(t) \tag{4.3.1}
\]

To find a suitable \( Z(t) \) and \( V(t) \) we need some constraint on the solution. We discuss
two methods. The first one is a model based method, i.e. it does not use state information, and is based only on the $A$, $B$ and $C$ matrices. The second method uses the state information, i.e. its a data based method.

### 4.3.1 Balanced Truncation method

The Balanced Truncation method is widely used as a constraint which maintains system stability (if the original system is stable). It makes the system as observable as it is controllable. This can be achieved by transforming the original system by a transformation matrix that diagonalizes $G_c(t)G_o(t)$ [33], where $G_c(t)$ and $G_o(t)$ are controllability and observability Grammians of the system obtained by solving the Ricatti equations 4.3.2 and 4.3.3.

\[
A(t)G_c(t) + G_c(t)A^T(t) + B(t)B^T(t) = \dot{G}_c(t) \tag{4.3.2}
\]

\[
A^T(t)G_o(t) + G_o(t)A(t) + C^T(t)C(t) = -\dot{G}_o(t) \tag{4.3.3}
\]

We can then choose $V(t) = U_{n_x \times m_x}(t)\Sigma^{-1/2}_{m_x \times m_x}(t)$, so that $\tilde{G}_c(t)\tilde{G}_o(t) = \Sigma^2(t)$, where $\tilde{G}$’s are the Grammians of the reduced order system. Then we get $Z^T(t) = V^\dagger(t)$ (i.e. a pseudo-inverse). Solving equations 4.3.2 and 4.3.3 is in general computationally challenging [56].

### 4.3.2 Proper Orthogonal Decomposition (POD) method

If we have the snapshots in time of the states, i.e. if we have the state trajectories then we can construct a snapshot matrix, see equation 4.3.4.

\[
\mathcal{X} = \begin{bmatrix} x(t_0) & \ldots & x(t_N) \end{bmatrix} \tag{4.3.4}
\]

We perform a singular value decomposition and keep only the left dominant singular vectors to get the matrix $U$ of dimensions $n_x \times m_x$. Thus we get a reduced order
system. This method can work when the state trajectories lie in a linear manifold; if this is not true then we would have to linearize in time and construct a time varying $U$ matrix.

4.4 Results and Discussion

We used the POD method to reduce our model since we can easily simulate even a large system and so can construct the snapshot matrix from the simulated states. The details of the simulation and the results we present are enumerated below.

1. A visual stimulus applied at time $t = 5s$ activates the primary visual cortex in the back of the brain.

2. Two source locations are selected to represent the maximum stimulus response, and for this simulation they are treated as a single composite neural mass model.

3. Measurements are simulated for 305 MEG channels, 59 EEG channels, and 228 DOT channels (114 source-detector pairs, two wavelengths 690 and 830 nm).

4. The simulation model had 68 states and was run for 20 seconds. Gaussian noise at appropriate variances was added to each modality.

5. Approaches tested:
   - State estimation on the full system using the Kalman filter.
   - State estimation from the reduced order system (17 states) using the POD approach.

6. After state estimation, we solved the forward problem again to give an estimate of the data.

The measure of closeness between states used was $\frac{\|X_{sim}\| - \|X_{estim}\|}{\|X_{sim}\|}$. The closeness is given in Table 4.1. The reconstructed data for 3 different channels is shown in figures 4.11, 4.12, and 4.13.
CHAPTER 4. MODEL REDUCTION

Figure 4.11: MEG channel 200: data from estimated states of the full and reduced order system

Figure 4.12: EEG channel 354: data from estimated states of the full and reduced order system

Figure 4.13: Optical channel 382: data from estimated states of the full and reduced order system
Table 4.1: Closeness between simulated and estimated states

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation of all (68) states</td>
<td>$1.7 \times 10^{-6}$</td>
</tr>
<tr>
<td>Estimation of reduced (17) states</td>
<td>$4.2 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

4.5 Future directions

In the near future we will perform parameter estimation for our system (at this point the parameters are fixed \textit{a priori}), and optimize the computation of the time varying $A$, $B$ and $C$ matrices in the reduced order system. We will also focus on the problem of different time scales in real data. These advances will enable model based analysis of real experimental data that further incorporates dimension reduction. We will also try to understand the physical relationships between the reduced order system and the biophysical variables.
Chapter 5

Conclusions and future work

The models presented in Chapters 2, and 3 could be used to combining measurements from complementary modalities like EEG/MEG which measure average neuronal responses and fNIRS/fMRI which measure the changes in oxygen concentration in hemoglobin. These models capture some of the relevant characteristics of the response waveforms and the complex changes that result from variations in the stimulus patterns. The framework followed in this thesis enables a simple way of updating the models without essentially changing the model structure in order to fit experimental data. Furthermore, experimental data with different variation patterns of the applied stimulus might lead to a better understanding of the underlying relationships, thus improving the prediction ability of the model. An interesting direction for further research has sprung from recent evidence that the habituation behavior of evoked responses are heavily dependent on the anesthetic used while performing the experiments [1]. The use of the model structure developed in this thesis can be explored to gain insights into this dependence.

The process of reducing the model order for functional imaging was briefly addressed in 4, but the work is as yet incomplete and is a direction of research worth exploring in the future. If the functional connectivity between various brain regions of a specific task is known, then order reduction could be achieved by “turning on”
the local models in the active regions and “turning off” the models in other regions. Thus, understanding functional connectivity patterns, and enabling the local models to capture this behavior becomes an important area of research. Such research may also lead to understanding the changes in connectivity patterns from normal brain conditions to pathological conditions. Thus, combining functional data using the modeling framework used in this thesis can help in diagnosis. This is another area on which future research may be performed.
Appendix A

EEG data preprocessing

As noted in §2.2.2, 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 Figure 2.1, 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 piecewise linear curve from the original EEG signal. Successive 4 sec segments are each evaluated as the average values of the symmetric 4 second interval, about the respective vertex. (This applies to each of the twenty, multi-interval runs, in each dataset, 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 (author?) [19].

Stochastic variations in the 4 sec response intervals are averaged out in the corrected signal. Averaging is carried separately in each dataset, and for each stimulus frequency. Thus, each stimulus frequency in each of the twelve dataset is represented
by a single 4 sec 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 $\sim 30 - 35$ ms), is comparable to the generic width of the P1 phase of the evoked response, which is $\sim 10$ ms ($\equiv 100$ 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 (A.0.1), below. Exploiting this fact, our method to remove the higher frequency EEG component can be formalized as follows:

$$\hat{y}(t) = w(t) \cdot y(t) + (1 - w(t)) \cdot (F \ast y)(t),$$  \quad (A.0.1)

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 sec 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 Figure 2.2, 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 sec 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 Figure 2.2, 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 Figures
2.1, 2.6 and 2.9. Applying the piecewise polynomial fit (2.4.1) 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 on
Figure 2.2, 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.
Appendix B

Order reduction in the Riera model

The analysis in this chapter 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 §B.0.3, using the model parameter values given in [54].

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 §B.0.1. Referring to the block diagram in Figure 2.4 on page 19, §B.0.2 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 §B.0.3. 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 (B.0.1), and their simplification, in (B.0.2) and beyond.
B.0.1 Riera model equations

The differential equations of the Riera model are reproduced below.

\[
\begin{align*}
\frac{dV_{IN}^T}{dt} &= \frac{1}{\tau_m} \left[ -V_{IN}^T(t) + I_3^+(t)R_m^0 \right] \tag{B.0.1a} \\
\frac{dV_{IN}^F}{dt} &= \frac{1}{\tau_m} \left[ -V_{IN}^F(t) + I^+(t - \tau_{PC})R_m^0 \right] \tag{B.0.1b} \\
\frac{dV_{PC}}{dt} &= \frac{1}{\tau_m} \left[ -\left( \alpha_0 + \sum_{k=1}^{2} \frac{1}{\beta_k} \right) V_{PC}(t) \\ &\quad - \frac{\Omega(t)}{\beta_1 \beta_2} - R_m I^-(t - \tau_{IN}) + \sum_{k=1}^{2} \left( R_m v_k(t) \right) \left( \frac{R_k^i + R_k^e}{R_k^i + R_k^e} \right) \right. \\
&\quad \left. - \frac{v_-(t)}{\beta_k} \right] \tag{B.0.1c} \\
\frac{d\Omega}{dt} &= \frac{1}{\tau_m} \left[ -\Omega(t) + R_m \left( \sum_{k=1}^{2} \frac{\beta_k (V_{PC}(t) - v_k(t))}{R_k^i + R_k^e} \right) \right. \\
&\quad \left. + V_{PC}(t) + v_-(t) \right] \tag{B.0.1d} \\
\frac{dv_k}{dt} &= \frac{1}{\tau_m} \left[ -v_k(t) + R_m I_k^+(t) \right] \quad k = \{1, 2\} \tag{B.0.1e} \\
\frac{dv_-}{dt} &= \frac{1}{\tau_m} \left[ -v_-(t) + R_m I^-(t - \tau_{IN}) \right] \tag{B.0.1f} \\
\frac{d\Phi}{dt} &= \frac{1}{\tau_m} \left[ -\left( \alpha_0 + \sum_{k=1}^{2} \frac{1}{\beta_k} \right) \Phi(t) - \frac{\Theta(t)}{\beta_1 \beta_2} \right. \\
&\quad + \frac{R_k^e}{(R_k^e + R_k^i)} \left( R_m I^-(t - \tau_{IN}) + R_m I^+_m(t) \right) \left( R_l^m v_-(t) + v_2(t) + R_m (v_2(t) - v_1(t)) \right) \left( R_k^e + R_k^i \right) \right] \\
&\quad \left. \prod_{k=1}^{2} \left( R_k^e + R_k^i \right) \right] \tag{B.0.1g} \\
\frac{d\Theta}{dt} &= \frac{1}{\tau_m} \left[ -\Theta(t) + \left( 1 + R_m \sum_{k=1}^{2} \frac{1}{R_k^m} \right) \Phi(t) \right] \tag{B.0.1h} \\
y &= \Phi. \tag{B.0.1i}
\end{align*}
\]
In the above equations, the currents $I^+(t - \tau_{PC})$, and $I^-(t - \tau_{IN})$ are given by:

\[
I^+(t - \tau_{PC}) = \alpha_{PC} f_{PC}(V_{PC}(t - \tau_{PC}))
\]
\[
I^-(t - \tau_{IN}) = \alpha_{IN} \left( f_{IN}(V^T_{IN}(t - \tau_{IN})) + f_{IN}(V^F_{IN}(t - \tau_{IN})) \right).
\]

These currents are not to be confused with the input currents of the Riera model, namely $I^+_i(t)$, $i \in i = 1, 2, 3$. The time delays $\tau_{PC}$, and $\tau_{IN}$ are set to zero in [54]. We follow the same in the present analysis. The functions $f_{PC}$, and $f_{IN}$, in the equations above, are given by:

\[
f_{PC}(V) = A_l + \frac{A_u}{(1 + Te^{-\gamma_{PC}(V - V^0_{PC})})^{1/T}}
\]
\[
f_{IN}(V) = A_l + \frac{A_u}{(1 + Te^{-\gamma_{IN}(V - V^0_{IN})})^{1/T}}.
\]

The equations in (B.0.1) are rewritten below for notational convenience.

\[
\dot{x}_1 = -ax_1 + bI_3^+(B.0.2a)
\]
\[
\dot{x}_2 = -ax_2 + bf_1(x_3)(B.0.2b)
\]
\[
\dot{x}_3 = -cx_3 - dx_4 - g_1[f(x_1) + f(x_2)] + g_2x_5 + g_3x_6 - hx_7(B.0.2c)
\]
\[
\dot{x}_4 = sx_3 - ax_4 - k_1x_5 - k_2x_6 + ax_7(B.0.2d)
\]
\[
\dot{x}_5 = -ax_5 + d_1I_1^+(B.0.2e)
\]
\[
\dot{x}_6 = -ax_6 + d_1I_2^+(B.0.2f)
\]
\[
\dot{x}_7 = -ax_7 + g_1[f(x_1) + f(x_2)](B.0.2g)
\]
\[
\dot{x}_8 = -cx_8 - dx_9 - r_1x_5 + r_2x_6 + qx_7 + o[f(x_1) + f(x_2)] + pI_2^+(B.0.2h)
\]
\[ \dot{x}_9 = sx_8 - ax_9 \]  
\[ y = x_8. \]  

In these equations,

\[ f(x) = \frac{1}{(1 + Te^{-\gamma(x-x_0)})^T} = f_{IN}(V) \]
\[ f_1(x) = \frac{w_1}{(1 + Te^{-\gamma_1(x-x_{01})})^T} = \alpha_{PC}f_{PC}(V). \]

\( f(x) \) 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 B.1 shows the relations between the parameters in the state-space model to the parameters of the Riera model.

**B.0.2 An overview of the proposed model reduction**

The block diagram of the Riera model connectivity in terms of the states in equation (B.0.2) is shown in Figure 2.4 on page 19. As seen in Figure 2.4, 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 \((x_5 \text{ through } x_9)\). 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 equations (B.0.2c) and (B.0.2d)). Removal of states \( x_3 \) and \( x_4 \) enables us to define a new state \( x_n = -dx_9 - r_1x_5 + r_2x_6 + qx_7 \), 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 equation (B.0.12) on page 109 substitutes the nonlinear subsystem,
Table B.1: Relation between the state-space parameters and the Riera model parameters.

<table>
<thead>
<tr>
<th>State-space parameters</th>
<th>Relation to Riera model parameters</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>( \frac{1}{\tau_m} ) ( R_0^0 / \tau_m )</td>
<td>50</td>
</tr>
<tr>
<td>b</td>
<td>( \frac{1}{\tau_m \beta_1 \beta_2} ) ( R_m / \tau_m )</td>
<td>204</td>
</tr>
<tr>
<td>c</td>
<td>( \frac{1}{\tau_m} \left( \alpha_0 + \sum_{k=1}^{2} \frac{1}{\beta_k} \right) )</td>
<td>396</td>
</tr>
<tr>
<td>d</td>
<td>( \frac{1}{\tau_m \beta_1 \beta_2} ) ( R_m / \tau_m )</td>
<td>3</td>
</tr>
<tr>
<td>d(_1)</td>
<td>( \frac{R_m \alpha_{IN} / \tau_m}{\beta_1 R_m} )</td>
<td>144</td>
</tr>
<tr>
<td>g(_1)</td>
<td>( \frac{(R_i^1 + R_e^1) \tau_m}{R_m} )</td>
<td>288</td>
</tr>
<tr>
<td>g(_2)</td>
<td>( \frac{(R_i^2 + R_e^2) \tau_m}{R_m} )</td>
<td>29</td>
</tr>
<tr>
<td>g(_3)</td>
<td>( \frac{1}{\tau_m} \sum_{k=1}^{2} \frac{1}{\beta_k} )</td>
<td>29</td>
</tr>
<tr>
<td>h</td>
<td>( \frac{1}{\tau_m} \left( 1 + R_m \sum_{k=1}^{2} \frac{1}{R_m} \right) )</td>
<td>912</td>
</tr>
<tr>
<td>k(_1)</td>
<td>( \frac{\beta_1 R_m}{(R_i^1 + R_e^1) \tau_m} )</td>
<td>647</td>
</tr>
<tr>
<td>k(_2)</td>
<td>( \frac{\beta_1 R_m}{(R_i^2 + R_e^2) \tau_m} )</td>
<td>215</td>
</tr>
<tr>
<td>r(_1)</td>
<td>( \frac{\tau_m \prod_{k=1}^{2} (R_i^k + R_e^k)}{R_e^2 (R_m + R_m^1) \tau_m} )</td>
<td>157</td>
</tr>
<tr>
<td>r(_2)</td>
<td>( \frac{\tau_m \prod_{k=1}^{2} (R_i^k + R_e^k)}{R_e^2 R_m \tau_m} )</td>
<td>169</td>
</tr>
<tr>
<td>q</td>
<td>( \frac{\tau_m \prod_{k=1}^{2} (R_i^k + R_e^k)}{R_m \tau_m} )</td>
<td>12</td>
</tr>
<tr>
<td>o</td>
<td>( \frac{\tau_m (R_i^1 + R_e^1)}{R_e^2 R_m} )</td>
<td>78</td>
</tr>
<tr>
<td>p</td>
<td>( \frac{\tau_m (R_i^2 + R_e^2)}{R_e^1 R_m^2} )</td>
<td>18</td>
</tr>
<tr>
<td>w(_1)</td>
<td>( \alpha_{PC} )</td>
<td>0.4</td>
</tr>
<tr>
<td>T</td>
<td>( T )</td>
<td>0.03</td>
</tr>
<tr>
<td>α</td>
<td>( \alpha_{IN} )</td>
<td>0.3</td>
</tr>
<tr>
<td>γ</td>
<td>( \gamma_{IN} )</td>
<td>5</td>
</tr>
<tr>
<td>γ(_1)</td>
<td>( \gamma_{PC} )</td>
<td>6</td>
</tr>
<tr>
<td>x(_{01})</td>
<td>( V^0_{PC} )</td>
<td>0.6</td>
</tr>
<tr>
<td>x(_0)</td>
<td>( V^0_{IN} )</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Values of the state-space parameters are computed using the relations with the Riera model parameters, whose values are published in [54]. Riera’s parameter values in [54] imply the selection of \( a = 33 \). Our election of \( a = 50 \) reflects the difference between the human data used in [54] and the rat data used in the present study.
Table B.2: Nominal values of the parameters of input currents $I_1^+$ and $I_2^+$ which are used for analyzing the dynamics of the state $x_3$.

<table>
<thead>
<tr>
<th>Input current parameter</th>
<th>nominal value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.3</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>20 ms</td>
</tr>
<tr>
<td>$\tau_3$</td>
<td>50 ms</td>
</tr>
</tbody>
</table>

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.0.3 Removing the feedback block

In Figure 2.4 on page 19 we notice that the feedback block can be removed if $f(x_2) \approx 0$. We also notice that $f_1(x_3)$ drives (B.0.2b). 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 equations (B.0.2e), and (B.0.2f) when they are driven by Gaussian functions as inputs. Substituting the Gaussian functions for $I_1^+$, and $I_2^+$ explicitly into (B.0.2e), and (B.0.2f), we have:

\[
\dot{x}_5 = -ax_5 + d_1\alpha_1 e^{-s_2^2(t-\tau_2)^2} \quad (B.0.3a)
\]
\[
\dot{x}_6 = -ax_6 + d_1\alpha_2 e^{-s_3^2(t-\tau_3)^2}. \quad (B.0.3b)
\]

We continue the present analysis by using the nominal values of the input current parameters, given in Table B.2. A statistical analysis over a range of parameter values, is presented in § B.0.3.
The solution for equations (B.0.3a), and (B.0.3b), are given by:

\[
\begin{align*}
x_5(t) &= A_1 e^{-a(t-\tau_2)} \left[ 1 + \text{erf} \left( s_2(t - \tau_2) - \frac{a}{2s_2} \right) \right] \quad \text{(B.0.4a)} \\
x_6(t) &= A_2 e^{-a(t-\tau_2)} \left[ 1 + \text{erf} \left( s_3(t - \tau_3) - \frac{a}{2s_3} \right) \right], \quad \text{(B.0.4b)}
\end{align*}
\]

where

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

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

In the equations (B.0.4) the function \text{erf} is the standard error function, which is defined as the integral of the Gaussian function (see [62]). 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 \(x_6\):

\[
\begin{align*}
x_5^u(t) &= a_1 A_1 e^{-a(t-\tau_2)} \quad a_1 = \begin{cases} 
1 & 0 \leq t < \tau_2 \\
2 & \tau_2 \leq t \leq \tau_e 
\end{cases} \quad \text{(B.0.5a)} \\
x_6^u(t) &= a_2 A_2 e^{-a(t-\tau_3)}, \\
& a_2 = \begin{cases} 
1 & 0 \leq t < \tau_3 \\
2 & \tau_3 \leq t \leq \tau_e 
\end{cases}, \quad \text{(B.0.5b)}
\end{align*}
\]

where \(\tau_2, \tau_3\) and \(\tau_e\) 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 diago-
nalizing their joint evolution equation

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

Equation (B.0.6) is of the form

\[\dot{x} = Ax + Bu\]

This equation is diagonalized by the change of state \(x = V\ddot{x}\), where \(V\) is the eigenvector matrix of \(A\). For the values in Table B.1 \(V\) turns out to be

\[
V = \begin{bmatrix}
-0.35 & 0.008 \\
0.94 & -1 \\
\end{bmatrix}.
\]

Clearly \(V\) can be approximated by

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

A straightforward substitution of \(x \approx \tilde{V}\ddot{x}\) in (B.0.6), yields

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

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 equation (B.0.2c) (see Table B.1).
Step 3. Simplification of the equation (B.0.7)

In equation (B.0.7), the contributions from the states $x_7$, $x_1$ and $x_2$ are non-positive, because negative coefficients multiply non-negative 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 equation (B.0.7). Hence, we create an upper bound on the equation (B.0.7) by removing the contributions of the states $x_7$, $x_1$ and $x_2$.

$$
\dot{x}_3^u = -\tilde{c}x_3^u + \tilde{g}_2x_5 + \tilde{g}_3x_6, \quad (B.0.8)
$$

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. $\approx 0$.

**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 $x_5^u$ (equation (B.0.5a) with $a_1 = 1$) in equation (B.0.8), leading to:

$$
\dot{x}_3^u = -\tilde{c}x_3^u + \tilde{g}_2x_5^u, \quad (B.0.9)
$$
Solving equation (B.0.9), we get

\[
x_3^u = \frac{\tilde{g}_2 A_1}{\tilde{c} - a} e^{-a(t-\tau_2)}(1 - e^{-(\tilde{c} - a)t}) = 0.45 e^{-a(t-\tau_2)}(1 - e^{-(\tilde{c} - a)t}) \quad (B.0.10)
\]

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 equation (B.0.2b) we get:

\[
\dot{x}_2^u = -ax_2^u + bf_1(x_3^u), \quad (B.0.11)
\]

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.

Figure B.1: 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 B.2. \(f_1(x_3^u)\) 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.

For the discussion of phases 2 and 3, we proceed by solving equations (B.0.8) and (B.0.11) for the states \(x_3^u\) and \(x_2^u\), using the nominal parameter values. Figure B.1
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.

**Phase 2: $t \in [\tau_2, \tau_3)$**

From the numerical computation shown in Figure B.1, $x_3^u \leq 0.55$, $f_1(x_3^u) \leq 0.10$, $x_2^u \leq 0.43$, 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 equation (B.0.2h) which forms the output of the system.

**Phase 3: $t \in [\tau_3, \tau_e]$**

In this phase $x_3$ is a decreasing function. The numerical evaluation of Figure B.1 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 $f(x_2) \neq 0$, 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.

**Statistical evaluation of the set of equations affecting $x_2$ in all cases**

So far we examined the details of the nominal unhabituated case. Our next task is to understand the deviations from the nominal case for both unhabituated and the habituated responses. The data is obtained by simulations of equations (B.0.2b), through (B.0.2g) 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.
Table B.3: The range of the varying parameters of input currents $I_1^+$, and $I_2^+$, collected from all rat datasets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>$[0.1, 1.4]$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$[0, 0.3]$</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>$[18, 45]$ ms</td>
</tr>
<tr>
<td>$\tau_3$</td>
<td>$[40, 80]$ ms</td>
</tr>
</tbody>
</table>

Table B.4: Features of the states $x_3$ and $x_2$ collected from numerical computation of equations (B.0.2b) through (B.0.2g) using 10,000 uniform random samples of the input current parameters with the range of values from Table B.3.

<table>
<thead>
<tr>
<th>Feature</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_3$</td>
</tr>
<tr>
<td>Peak amplitude</td>
<td>$0.58 \pm 0.08$</td>
</tr>
<tr>
<td>Peak time</td>
<td>$43 \pm 7$ ms</td>
</tr>
<tr>
<td>Approx. support of $f(\cdot)$</td>
<td>$17 \pm 3$ ms</td>
</tr>
<tr>
<td>$f(\cdot)$ peak</td>
<td>$0.13 \pm 0.06$</td>
</tr>
</tbody>
</table>

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

1. The Best Fit Model (BFM) (estimated in §2.4.2) parameters of the input currents $I_1^+$, and $I_2^+$ that vary under habituation were collected from the entire dataset.

2. The absolute minimum and maximum values of parameters that vary under habituations were taken as the range of the parameters (see Table B.3).

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 equations (B.0.2b) through (B.0.2g), 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.
Table B.5: 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 $y_{nfo}$.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Relative difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1L</td>
<td>.020</td>
</tr>
<tr>
<td>1R</td>
<td>.030</td>
</tr>
<tr>
<td>2L</td>
<td>.020</td>
</tr>
<tr>
<td>2R</td>
<td>.021</td>
</tr>
<tr>
<td>3L</td>
<td>.021</td>
</tr>
<tr>
<td>3R</td>
<td>.020</td>
</tr>
<tr>
<td>4L</td>
<td>.017</td>
</tr>
<tr>
<td>4R</td>
<td>.020</td>
</tr>
<tr>
<td>5L</td>
<td>.020</td>
</tr>
<tr>
<td>5R</td>
<td>.023</td>
</tr>
<tr>
<td>6L</td>
<td>.023</td>
</tr>
<tr>
<td>6R</td>
<td>.019</td>
</tr>
</tbody>
</table>

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

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 B.4 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 B.4 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 B.4. In particular $f(x_2) \leq 0.02$ throughout, and non-negligible 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.

**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 B.5. 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:

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

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 dataset under consideration.

**Justification of the substitution of $f(x_1)$ by $I_{3}^+$**

The nonlinear function $f(x_1)$ appears as a feedforward input into the reduced order linear model. It was observed throughout the dataset 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 $I_{3}^+$, given in equation (B.0.12) was made to match this feature. The quadratic function for $I_{3}^+$ is parameterized to it to have a value of zero when $t - \bar{\tau}_1 = 0$, and $t - \bar{\tau}_1 = \beta$, and to have a peak at the center of the interval $(0, \beta)$.

$$I_{3}^+ = \alpha(t - \bar{\tau}_1) \left( 1 - \frac{(t - \bar{\tau}_1)}{\beta} \right), \quad 0 \leq t - \bar{\tau}_1 \leq \beta \quad (B.0.12)$$
A least squares fit to $f(x_1)$ over the rat datasets for the $\alpha$ and $\beta$ parameters led us to fix $\beta = 8$ ms in our work. The results summarized in Figures 2.12 and 2.13 demonstrate in particular that the match between the linear model and reduced three state model using $f(x_1)$ are similar.
Appendix C

Single compartment Windkessel model

C.1 Hemodynamic model equations

We present the detailed equations and parameters of the hemodynamic model here. We then present the sensitivity analysis done on the baseline parameters of the model in §C.2.1, and the sensitivity of the model to the $\Delta\text{CMRO}_2$ input in §C.2.2. We first show the equations of the model followed by the parameters and constants of the model.

The hemodynamic model consists of four state equations given below.

\[
\dot{v} = g(g_0 + (1 - g_0)v^\beta) - v^{\beta+2} \quad (C.1.1a)
\]
\[
\dot{x}_2 = \frac{1}{v} \left( a(C_{in} - x_2)(g(g_0 + (1 - g_0)v^\beta) + v^{\beta+2}) - 2Ka(f(x_2) - \alpha_b x_3) \right) \quad (C.1.1b)
\]
\[
\dot{x}_3 = K(f(x_2) - \alpha_b x_3) - \text{CMRO}_2(0) - \Delta\text{CMRO}_2(t) \quad (C.1.1c)
\]
\[
\dot{x}_4 = \frac{2}{3}av^{\beta+2}(x_2 - x_4) - \frac{1}{3}\left( (1 - a)(g(g_0 + (1 - g_0)v^\beta) - (1 - a)v^{\beta+2})x \right) + aC_{in}(g(g_0 + (1 - g_0)v^\beta) + v^{\beta+2}) - 2Ka(f(x_2) - \alpha_b x_3) \quad (C.1.1d)
\]

The state equations contain nine baseline parameters. The relation between the
parameters in the equation set (C.1.1) and the baseline parameters are given in Table C.3. The constants used in the model are given in Table C.1. The initial conditions for the equation set (C.1.1) are given in Table C.2. The nominal values of the baseline parameters are given in Table C.4.

<table>
<thead>
<tr>
<th>Constants</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_n$</td>
<td>$O_2$ carrying capacity</td>
<td>1.39 ml O2/g Hb</td>
</tr>
<tr>
<td>$H_{GB}$</td>
<td>Hematocrit</td>
<td>16 g Hb/dL blood</td>
</tr>
<tr>
<td>$\alpha_T$</td>
<td>tissue $O_2$ solubility</td>
<td>0.003051 ml $O_2$/dL blood/mm Hg</td>
</tr>
</tbody>
</table>

Table C.1: Values of constants in the hemodynamic model

<table>
<thead>
<tr>
<th>Initial conditions</th>
<th>Relation to baseline parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(0)$</td>
<td>$g_0/g_0 = 1$</td>
</tr>
<tr>
<td>$x_2(0)$</td>
<td>$H_nH_{GB}(2SvO_2(0) - SaO_2)$</td>
</tr>
<tr>
<td>$x_3(0)$</td>
<td>$pTO_2\alpha_T/b$</td>
</tr>
<tr>
<td>$x_4(0)$</td>
<td>$H_nH_{GB}(2SvO_2(0) - SaO_2)$</td>
</tr>
</tbody>
</table>

Table C.2: Initial conditions for the equation set (C.1.1)

<table>
<thead>
<tr>
<th>Parameter in equation set (C.1.1)</th>
<th>Relation to baseline parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{in}$</td>
<td>$H_nH_{GB}SaO_2$</td>
</tr>
<tr>
<td>$K$</td>
<td>$CMRO_2(0)$</td>
</tr>
<tr>
<td>$\alpha_b$</td>
<td>$(f(x_2(0)) - pTO_2)$</td>
</tr>
<tr>
<td></td>
<td>$b/\alpha_T$</td>
</tr>
</tbody>
</table>

Table C.3: Relation between parameters in the equation set (C.1.1) and the baseline parameters.

In equations (C.1.1b),(C.1.1c) and (C.1.1d), $f(x_2) = p_vO_2$ is the partial pressure of $O_2$ in the vein which is related to saturation of oxygen in the vein by equation (C.1.2) which is modified from the equation given in [57].

$$S_vO_2 = \left(\frac{23000}{p_vO_2^3 + 150p_vO_2} + 1\right)^{-1}, \quad (C.1.2)$$

also $S_vO_2 = \alpha_H(C_{in} + x_2)$. Hence, to solve for $f(x_2) = p_vO_2$ we have to find the real root of the cubic equation C.1.3, where $C(x_2) = 23000\left(\frac{S_vO_2}{1-S_vO_2}\right)$.

$$p_vO_2^3 + 150p_vO_2 - C(x_2) = 0 \quad (C.1.3)$$
### Table C.4: Nominal values for the baseline parameters of the hemodynamic model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Nominal value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_0$</td>
<td>baseline conductance</td>
<td>1/0.7 $\Omega^{-1}$</td>
<td>[37, 5]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>inverse of grub exponent</td>
<td>2.5</td>
<td>[37, 24]</td>
</tr>
<tr>
<td>$S_{aO_2}$</td>
<td>$O_2$ saturation in artery</td>
<td>0.95</td>
<td>[70]</td>
</tr>
<tr>
<td>$S_{vO_2}$</td>
<td>$O_2$ saturation in vein</td>
<td>0.7</td>
<td>[70]</td>
</tr>
<tr>
<td>$a$</td>
<td>inverse of vascular transit time</td>
<td>1/0.9 $s^{-1}$</td>
<td>[28, 72]</td>
</tr>
<tr>
<td>$b$</td>
<td>inverse of tissue volume</td>
<td>1/2</td>
<td></td>
</tr>
<tr>
<td>$HbT_0$</td>
<td>baseline HbT</td>
<td>75 $\mu$ Mol.</td>
<td>[28, 70]</td>
</tr>
<tr>
<td>$\phi_v$</td>
<td>vein volume fraction</td>
<td>0.6</td>
<td>[28, 71]</td>
</tr>
<tr>
<td>$pO_2$</td>
<td>partial pressure of $O_2$ in tissue</td>
<td>15</td>
<td>[70]</td>
</tr>
</tbody>
</table>

Using Cardano’s method for finding the root of equation C.1.3 we get equation C.1.4.

$$f(x_2) = p(x_2) - 50/p(x_2), \quad p(x_2) = \left( \frac{C(x_2) + \sqrt{C(x_2)^2 + c^2}}{2} \right)^{1/3}$$  \hspace{1cm} (C.1.4)

where $c = 500,000$.

The output equations for the hemodynamic model are given in equation C.1.5, where $S_{pO_2} = \alpha_H(x_2 + x_4)$

\[
\begin{align*}
HbO_T &= (\phi_v SvO_2 v + (1 - \phi_v)S_{aO_2}(0)\sqrt{g})HbT_0 \\
HbR_T &= (\phi_v(1 - SvO_2)v + (1 - \phi_v)(1 - S_{aO_2}(0))\sqrt{g})HbT_0 \\
HbO_p &= \frac{3}{a}S_{pO_2}HbT_0 \\
HbR_p &= \frac{3}{a}(1 - S_{pO_2})HbT_0
\end{align*}
\]

The total HbO = HbO$_T$ + HbO$_p$, and total HbR = HbR$_T$ + HbR$_p$. The data we have is $\Delta$HbO = HbO-HbO(0), and $\Delta$HbR = HbR-HbR(0).

### C.2 Sensitivity analysis of the hemodynamic model

In this paper we used the hemodynamic model with the baseline parameter set to their nominal values obtained from prior studies (refer Table C.4). Analysis of model
sensitivity to these parameter values becomes important because the datasets used in this paper lack information about baseline conditions. This analysis is carried out in §C.2.1. We then analyze the model sensitivity to the $\Delta$CMRO$_2$ input because reconstruction of the true $\Delta$CMRO$_2$ using the least squares method discussed in §3.3.2 depends on the model being sensitive to that input. Notice that the model is sensitive to the $g$ input because we can reconstruct $g$ via direct inversion, and hence it is not discussed here.

C.2.1 Sensitivity to baseline parameters

The hemodynamic model uses nine baseline parameters fixed to their nominal values (refer Table C.4). We performed a sensitivity analysis on the baseline parameters of the model using the sample problem which was used to show the reconstruction of inputs of the model in Figure 3.5 on page 62. The magnitude of the derivative of the output with respect to change in each parameter is a measure of sensitivity of the model for that parameter. We used sensitivity measures based on the maximum of the magnitude of the derivative with respect to the parameter. We approximated the derivative of the output with respect to the parameter by the difference of the parameter varied by 10% of its fixed value. We computed the maximum of the change in HbO and HbR using the formulae given below:

$$
sens_{HbO} = \frac{1}{2} \left( \max \left| \frac{HbO_m(p) - HbO(p - \delta p)}{\delta p} \right| + \max \left| \frac{HbO(p + \delta p) - HbO_m(p)}{\delta p} \right| \right)$$

$$
sens_{HbR} = \frac{1}{2} \left( \max \left| \frac{HbR_m(p) - HbR(p - \delta p)}{\delta p} \right| + \max \left| \frac{HbR(p + \delta p) - HbR_m(p)}{\delta p} \right| \right)
$$

where HbO$_m$ refers to the HbO computed with the nominal values of the baseline parameters given in Table C.4, HbR$_m$ refers to the same for HbR. Table C.5 records the sensitivity of the model to the nine baseline parameters from equations (C.2.1).

We infer from the table that the model is most sensitive to oxygen saturation in the
APPENDIX C. SINGLE COMPARTMENT WINDKESSEL MODEL

Table C.5: Sensitivity of the hemodynamic model to the baseline parameters. The analysis was performed by using the nominal values of the baseline parameters from Table C.4 and using the sample problem which shows the reconstruction example in Figure 3.5 on page 62.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$sens_{HbO}$</th>
<th>$sens_{HbR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_0$</td>
<td>6.65</td>
<td>5.72</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.84</td>
<td>0.46</td>
</tr>
<tr>
<td>$SaO_2$</td>
<td>45.42</td>
<td>45.42</td>
</tr>
<tr>
<td>$SvO_2$</td>
<td>42.27</td>
<td>42.27</td>
</tr>
<tr>
<td>$a$</td>
<td>4.41</td>
<td>2.66</td>
</tr>
<tr>
<td>$b$</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>$HbT_0$</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>$\phi_v$</td>
<td>4.74</td>
<td>1.23</td>
</tr>
<tr>
<td>$pTO_2$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

arteries and veins, i.e. $SaO_2$, and $SvO_2(0)$. Generally $SaO_2$ varies from around 0.85 to 0.98, and $SvO_2$ varies from 0.6 to 0.8 and hence a direct search in their ranges could be performed if needed. Thus, the model is found to have least dependence on most baseline parameters that can be empirically picked from the existing literature.

C.2.2 Sensitivity of the model to the $\Delta CMRO_2(t)$ input

For reconstruction of the $\Delta CMRO_2(t)$ input from $\Delta HbO$ and $\Delta HbR$ data we rely upon the sensitivity of those outputs to the $\Delta CMRO_2(t)$ input. To measure the sensitivity we again used the sample problem from Figure 3.5 on page 62, computed the B-spline weights for the true $\Delta CMRO_2$ input which we denote here as $C_t$. We also refer to the $\Delta HbO$ and $\Delta HbR$ generated by the model using the true inputs as $H_t$ in this section.

We then generated 100 different random $\Delta CMRO_2$ inputs by randomizing the weights of the B-splines, by first selecting a random sample from a uniform distribution of weights between one third and thrice of the original weights, and then randomizing the signs. We then computed the mean across the 100 trials of the
variation in output to the variation in input as shown below

\[ \text{sens}_{\text{CMRO}} = \frac{1}{100} \sum_{i=1}^{100} \frac{\| H_i - H_t \|^2_F}{\| C_i^t - C_t^t \|^2_2} = 2.1 \]

This crude sensitivity computation suggests that the error in output can be twice as much as the error in input, i.e. a wrong estimate of the input weights leads to twice the error in matching the output and so any least squares based algorithm would search in the direction of the true input. Thus we conclude that the model is sensitive to the \( \Delta \text{CMRO}_2(t) \) input.
Bibliography


