Neural Engineering: Unraveling The Complexities Of Neural Systems

by Chris Eliasmith
University of Waterloo, Waterloo, ON

Abstract

The incredible and often subtle complexity of neural systems may seem like an engineer's nightmare. But, when we examine such systems carefully, they can turn out to be an engineer's dream - a way to learn about robust, complex systems. Using techniques from information theory, control theory, and signals and systems analysis, it is possible to formulate a framework for constructing large-scale, biologically plausible simulations of neural systems. Such simulations help us learn both about how the underlying neural systems work, and about good solutions to the problems faced by such systems.

Sommaire

L'incroyable, souvent subtile complexité des systèmes neuronaux semble être le cauchemar de l'ingénieur. Mais quand on examine ces systèmes soigneusement, ils peuvent devenir le rêve de l'ingénieur - un moyen pour étudier des systèmes robustes et complexes. En utilisant des techniques de la théorie de l'information, théorie du contrôle, et l'analyse des signaux et systèmes, il est possible de formuler un cadre pour construire de grandes et biologiquement plausibles simulations de systèmes neuronaux. Ces simulations nous aident à apprendre comment fonctionnent les systèmes neuronaux sous-jacent et comment obtenir de bonnes solutions aux complexes problèmes faisant face à ces systèmes.

1.0 An engineering approach to neuroscience

Recently, several leaders in neuroscientific research have made independent calls for the development of a theoretical framework that can help unify the field [1]. Such a framework, it is hoped, can provide structure to the chaotic landscape of experimental and theoretical results in contemporary neuroscience. Essentially, these neuroscientists have come to the realization that neuroscience is 'data rich, but theory poor.' That is, generating results regarding neural systems, their components, and their behaviour has become routine. But, there is no established method for integrating such results into a consistent, informative theory of neural systems. As a result, it is difficult to make predictions regarding such systems, and to determine what the most needed experiments on a neural system are.

There are reasons to think that engineering is not going to help provide such a unifying framework. This is because engineers typically deal with systems composed of identical, well-characterized, sparsely interconnected, and often digital parts. None of these constraints apply to neural systems. Neurons are wildly diverse in size (10^{-4}-5 m), transmission speed (2-400 km/h), response curves (orders of magnitude gain for the same stimuli), connectivity (500-200,000 inputs), and temporal dynamics (5-100 ms for synapses alone). But, of course, engineers do not have to focus on digital, identical component systems.

In this article, I briefly describe the results of work that I have been doing with Charles H. Anderson from Washington University School of Medicine. We attempt to show how the complexities of neural systems can be systemmatically understood using quantitative tools standard in engineering. A more comprehensive discussion can be found in our recent book *Neural engineering: Computation, representation, and dynamics in neurobiological systems* [2].

2.0 Three principles of neural engineering

Our research has built on the important contributions of a number of others to understanding neural coding and dynamics [3-6]. Our contribution has been to synthesize these results, extend them to characterize neural computation, and incorporate them into a neurally-relevant version of control theory. The resulting framework is effectively summarized by the following three principles:

1. Neural representations are defined by the combination of nonlinear encoding (exemplified by neuron tuning curves, and neural spiking) and weighted linear decoding (over populations of neurons and over time).
2. Transformations of neural representations are functions of the variables represented by neural populations. Transformations are determined using an alternately weighted linear decoding.
3. Neural dynamics are characterized by considering neural representations as control theoretic state variables. Thus, the dynamics of neurobiological systems can be analyzed using control theory.

In addition to these main principles, we take the following addendum to be important for analyzing neural systems:

- Neural systems are subject to significant amounts of noise. Therefore, any analysis of such systems must account for the effects of noise.

I do not discuss the addendum in detail here, but it is important to note how central it is for properly characterizing real-world, biological systems. Let us consider each of the main principles in more detail.

2.1 Principle 1 - Representation

There are two obvious nonlinearities in neural systems. The first is the neural action potential. This is a rapid, stereotypical depolarization of the neuron that results when the current in the cell body goes over some threshold. These neural 'spikes' effectively convert an analog voltage inside the cell body into a series of delta-function like responses (creating a 'spike train') that are then sent down the axon to subsequent cells. Despite this highly nonlinear temporal encoding, it has been shown that about 95% of the information carried by the action potentials can be recovered using an optimal linear filter (i.e., a first-order Weiner filter) [6].

The second nonlinearity in neural systems is evident in the neuron response function. This function describes the increase in spike rate as a function of input voltage to the cell body. While these functions are generally monotonically increasing, they go to zero abruptly (mathematically a singularity) and saturate. It would be very difficult to encode a stimulus with one such response function. However, neurons generally work in concert to encode any given stimulus. That is, different neurons are sensitive to over-lapping but non-identical parts of the stimulus space. As a result, different neurons provide different, though partially redundant information regarding a stimulus (i.e., the neurons form an overcomplete encoding of the stimulus space). As a result, we can show that the optimal linear population filter provides a good decoding, whose error decreases at a rate of 1/(Number of Neurons), even under noise.

As is well-known from information theory, to properly define a code we must specify both an encoding and decoding. Given the above characterizations of neural responses, we can define such a code over both time and populations of neurons. Mathematically, we can express the following 'population-temporal' code:

$$\sum_n \delta_i(t-t_n) = G_i[\tilde{\varphi}_i \cdot \mathbf{x}(t)] \qquad \text{Encoding} \quad (1)$$

$$\hat{\mathbf{x}}(t) = \sum_{in} \delta_i(t-t_n) * \varphi_i^{\mathbf{x}}(t) \qquad \text{Decoding} \quad (2)$$

where $\sum_n \delta_i(t-t_n)$ is the spike train resulting from $G_i[\cdot]$ which defines the (spiking) neuron response function, $\tilde{\varphi}_i$ are encoders (or 'preferred' stimulus) vectors observable in neurons, and $\varphi_i^{\mathbf{x}}(t)$ are the product of the linear population decoders and the temporal filter, giving a combined 'population-temporal filter.'

2.2 Principle 2 - Transformation

In order to make these neural representations useful, we need to be able to define transformations of these representations (i.e. functions of the encoded variables). In fact, this turns out to be rather straightforward: rather than finding the optimal representational decoders (i.e., those that extract the original variable from the information encoded by the neural spikes), we can find optimal transformational decoders to extract some function of the encoded variable from those spikes:

$$\hat{f}(\mathbf{x}) = \sum_{in} \delta_i(t-t_n) * \varphi_i^{f(\mathbf{x})}(t) \qquad (3)$$

These decoders, $\varphi_i^{f(\mathbf{x})}(t)$, tell us how to implement computations using biologically plausible networks. Notably, both linear and nonlinear functions of the encoded variable can be computed in this manner. In fact, we can use this approach to determine precisely what set of functions can be computed with this linear decoding given a particular population of neurons (i.e. neurons with a particular distribution of nonlinear tuning curves). This can be extremely useful for limiting the set of possible functions a given neural population can compute.

2.3 Principle 3 - Dynamics

Traditional approaches to artificial intelligence, including work on artificial neural networks, have often been criticized for ignoring the importance of temporal constraints for determining successful behaviour. In the biological world, not performing certain behaviours (i.e., computing certain functions) fast enough can mean the difference between life and death. As a result, many contemporary approaches to modeling cognitive systems have elevated dynamics to be of as much importance as representation and computation when trying to understand neural systems.

Adopting this philosophy, we have integrated our characterization of representation and transformation with modern control theory. Control theory is the engineer's tool for describing the dynamics of physical systems and is thus a natural choice for describing neural dynamics. Our central suggestion, which makes the marriage between standard control theory and our description of neural systems possible, is that the variables represented by a neural population are control theoretic state variables.

However, we must do some work to show how this can occur. Standard control theory takes the basic transfer function to be integration. However, neural systems do not perform integration easily, but have their own intrinsic dynamics. Fortunately, we can show that the dynamics of the post-synaptic current (PSC) which results in the dendrite of a neuron that receives a spike is likely to dominate the dynamics of the cellular response as a whole. As a result, it is reasonable to characterize the dynamics of neural populations based on synaptic dynamics. Assuming a simple but plausible model for synaptic dynamics, we can effect a 'translation' between traditional control theory and a 'neural' control theory[1].

For example, in the linear case, the dynamics of a standard control structure are written as:

$$\dot{\mathbf{x}}(t) = \mathbf{Ax}(t) + \mathbf{Bu}(t) \qquad (4)$$

where $\mathbf{x}(t)$ is the vector of state variables, $\mathbf{u}(t)$ is an input or control signal, and $\mathbf{A}$ and $\mathbf{B}$ determine the dynamics. In the neural case, assuming our simple PSC model, we can find two neural dynamics matrices:

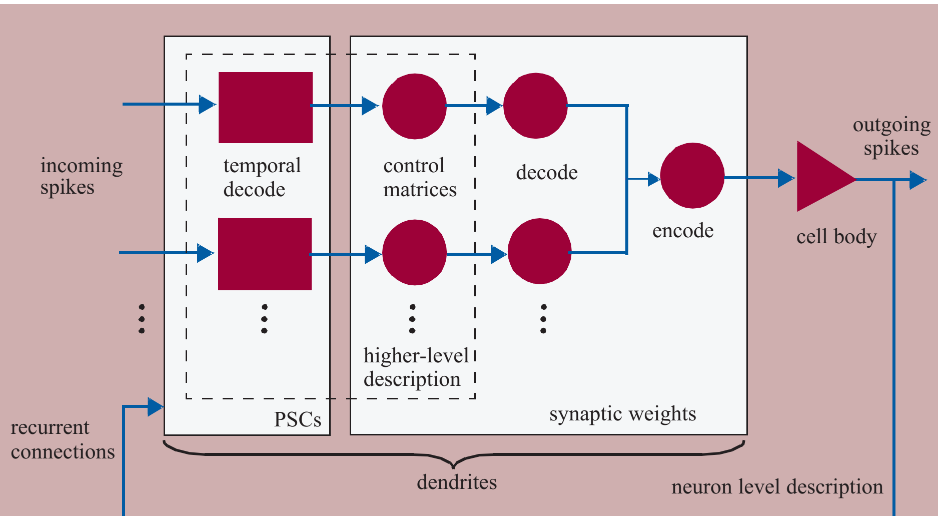
$$\begin{aligned} \mathbf{A}' &= \tau\mathbf{A} + \mathbf{I} \\ \mathbf{B}' &= \tau\mathbf{B} \end{aligned} \qquad (5)$$

Importantly, using these new matrices results in the same dynamics in a neural system $\mathbf{A}$ and $\mathbf{B}$ did in the standard control system. This means that all of the tools of control theory, including known control structures, stability analyses, and so on, can be brought to bear on understanding neural systems.

To complete the synthesis, we can now embed this description of the

1: Specifically, we assume $h(t) = \frac{1}{\tau}e^{-t/\tau}$ where τ is the synaptic time constant. More accurate models of PSC generation can be used, but this one makes the analysis more tractable.

Figure 1: A generic neural subsystem. The labels provide a qualitative description of the function of each of the interior elements (defined in equations (1)-(4)). See text for further discussion.



dynamics of a neural representation directly into our characterization of neural encoding in equation (1). This results in the following general, quantitative description of neural spiking:

$$\sum_n \delta_i(t - t_n) = G_i\left[\tilde{\varphi}_i \cdot h_i(t) * \left[\mathbf{A'x}(t) + \mathbf{B'u}(t)\right]\right] \qquad (6)$$

where $h_i(t)$ is our PSC model. The functions needed to implement such a neurally embedded control system can, of course, be implemented using principle 2.

2.4 Synthesis - A generic neural subsystem

These three principles quite clearly come together in equation (6). The result can be expressed as defining the 'generic neural subsystem' shown in Figure 1.

In this figure, the interior dotted line indicates the higher-level description of the overall neural dynamics. Separating out these elements can be practically useful because it allows for a computationally cheap means of characterizing the system's behaviour. The exterior dotted line bounds elements usually referred to in a description of neural function: i.e., spikes impact dendrites that give rise to PSCs weighted by a synaptic weight whose effects at the cell body result in the generation of outgoing spikes. The grey boxes indicate the elements that we can now quantify given these principles. Most importantly, we can determine the synaptic connection weights that need to connect this population to its predecessors in order to exhibit the desired high-level behaviour. This alleviates the need for (though does not exclude the possibility of) including learning in constructing neurally plausible simulations.

3.0 Discussion

Taken together, these three principles can serve to direct the construction of large-scale, biologically plausible simulations. This is no more evident than in the numerous simulations we have constructed. Here is a brief description of three:

1. Vestibular system (sensory): This is a large-scale, spiking neuron model that solves a nonlinear control problem: namely, estimating inertial acceleration given linear acceleration (from the otolith organs) and angular velocity (from the semi-circular canals). The model maps well to known vestibular nucleus physiology, and provides predictions regarding the distribution of receptors and tuning curves in the relevant neural populations.

2. Working memory (cognitive): This recurrent spiking model accounts for two phenomena previously observed but that remained unexplained by simulations (parametric variation and multiple target representation in working memory). It simulates parts of the lateral intraparietal cortex, involved in remembering the location of external targets.

3. Lamprey swimming (motor): This spiking model demonstrates the synthesis of top-down and bottom-up data in a model. The result is a novel model that guarantees certain high-level behavior (e.g. swimming stability over a range of frequencies), unlike past models.

In addition, these principles help to unify several central concepts in neuroscience - going some way to systematizing neuroscientific results. For instance, principle 1 unifies population and temporal representation in neural systems via the definition of a population-temporal decoder. As well, principles 1 and 2 taken together provide a unified characterization of representation and transformation (or computation) as optimal linear decoding. Considering all three principles together, we can see how this approach can be used to unify top-down and bottom-up evidence on a single neural system. High-level hypotheses, which inform the control theoretic description of the overall system is integrated with the evidence regarding individual neurons, such as tuning curves and response properties.

As well, the principles are general. While I have here characterized the principles in terms of vector representation, there are equivalent characterizations for scalar and function representations, and their combinations (e.g., vector fields). As well, the characterization generalizes over modeling assumptions made regarding individual neuron behavior (i.e., how neural spikes are generated), transformations (i.e., linear and nonlinear), and dynamics (i.e., time invariant, time varying, linear, nonlinear, or stochastic control).

4.0 Conclusion

While we hope that these principles can provide some needed structure to the many results being generated by neuroscientists, we are careful to remind others (and ourselves) that this is, at best, a 'first guess.' However, because there are not a lot of other theories of this kind on offer, and because a first guess paves the way for better guesses, we think that this framework can play a valuable role in the development of neuroscience. We have greatly benefitted from adopting this view because of the new and important issues and insights it has generated regarding the organization of neural systems. Many of the seeming complexities of neural behaviour (e.g., the heterogeneity of neural responses) become expected once we understand these behaviours as a certain kind of neural representation. Of course, many such questions remain unanswered, but we, and others, have found this framework useful for determining which questions are most worth asking.

5.0 References

[1]. Nature Neuroscience (2000). 3: Special issue on Computational Neuroscience. (See especially the series of 'Viewpoints' articles.)

[2]. Eliasmith, C. and C. H. Anderson (2003). Neural engineering: Computation, representation, and dynamics in neurobiological systems. Cambridge, MA: MIT Press.

[3]. Georgopoulos, A. P., A. Schwartz, and R. E. Kettner (1986). Neuronal population coding of movement direction. Science. 233: 1416-1419.

[4]. Salinas, E. and L. F. Abbott (1994). Vector reconstruction from firing rates. Journal of Computational Neuroscience. 1: 89-107.

[5]. Seung, H. S. (1996). How the brain keeps the eyes still. Proceedings of the National Academy of Sciences. 93: 13339-13344.

[6]. Rieke, F., D. Warland, R. R. de Ruyter van Steveninick, and W. Bialek (1997). Spikes: Exploring the neural code. Cambridge, MA: MIT Press.

About the author

Chris Eliasmith received his Ph.D. from the Philosophy-Neuroscience-Psychology program at Washington University in St. Louis in 2000. Since 2001 he has been an Assistant Professor in Philosophy, and is cross-appointed to Systems Design Engineering, at the University of Waterloo. Prior to that he held a post-doctoral position at Washington University Medical School. His research interests in neuroscience include working memory, motor control, vision and audition. He also works on theories of mental representation and aspects of philosophy of science. He has recently received support from the Canadian Foundation for Innovation to develop a parallel simulation environment for building large-scale models using the methods described here. He can be reached at eliasmith@uwaterloo.ca.